Routledge Encyclopedia of Philosophy

A posteriori

A prominent term in theory of knowledge since the seventeenth century, 'a posteriori' signifies a kind of knowledge or justification that depends on evidence, or warrant, from sensory experience. A posteriori truth is truth that cannot be known or justified independently of evidence from sensory experience, and a posteriori concepts are concepts that cannot be understood independently of reference to sensory experience. A posteriori knowledge contrasts with a priori knowledge, knowledge that does not require evidence from sensory experience. A posteriori knowledge is empirical, experience-based knowledge, whereas a priori knowledge is non-empirical knowledge. Standard examples of a posteriori truths are the truths of ordinary perceptual experience and the natural sciences; standard examples of a priori truths are the truths of logic and mathematics. The common understanding of the distinction between a posteriori and a priori knowledge as the distinction between empirical and non-empirical knowledge comes from Kant's Critique of Pure Reason (1781/1787).

1 Empirical warrant

Kant (1781/1787: A2/B3) notes that 'opposed to [a priori knowledge] is empirical knowledge, which is knowledge possible only a posteriori, that is, through experience' (see Kant, I. §4). Empirical knowledge is a posteriori in virtue of the kind of warrant, or justification, it requires for the proposition known: a kind of warrant somehow grounded in sensory experience. The standard approach to knowledge claims that propositional knowledge requires justified true belief. The belief condition for knowledge, according to this approach, must be appropriately related to the satisfaction of the truth condition, thereby excluding true groundless conjectures from the category of knowledge. This requirement involves a justification condition for knowledge, and this condition typically receives most of the attention from contemporary accounts of empirical knowledge.

Contemporary accounts of empirical, or a posteriorif justification ordinarily seek to explain what sorts of processes (vision, memory, introspection, inference and so on) can, and perhaps standardly do, yield empirical justification for beliefs. These accounts typically assume *fallibilism* about empirical justification: an empirically justified contingent belief can be false (see Fallibilism). These accounts also typically assume that evidence providing empirical justification for a belief need not logically entail that belief, but can be inductive or probabilistic. Although contemporary epistemologists do not share a single account of the kind of probability appropriate to empirical justification, they largely agree that empirical evidence is *defeasible*, that it can cease to be justifying upon one's acquiring broader evidence. Upon approaching an apparent large pool of water on the road, for example, one might lose one's justification for thinking that there actually is such a pool on the road. An account of empirical justification must, in any case, identify a suitable role for sensory experience in the conferring of justification. Otherwise, the distinction between a posteriori and a priori warrant and knowledge will be unclear.

Contemporary philosophers debate whether any necessarily true proposition is knowable a posteriori. Saul Kripke, for instance, has argued that some necessarily true propositions must be known a posteriori if they are to be known at all. His main examples are true identity statements involving names: for example, 'Hesperus is Phosphorus'. Kripke holds that such statements are necessarily true, but that they cannot be known a priori (Kripke 1980). Other philosophers have challenged the view that such identity statements are necessarily true.

2 Coherentism about empirical warrant

Coherentism about empirical justification identifies particular 'coherence relations' among beliefs as constituting empirical justification. Coherence relations can include logical entailment relations, relations of explanation, and various probabilistic, or inductive, relations. Whatever they include, coherentism about empirical justification must identify what makes justification empirical; it must acknowledge a role for sensory experience.

Coherentism about empirical justification faces an objection if empirical evidence can come from a kind of sensory experience that does not require corresponding beliefs descriptive of the relevant sensory experience. Suppose, for instance, that empirical evidence can come from the non-propositional contents (such as visual images) of one's non-belief sensory awareness-states (for example, non-belief visual states). Not all sensory states are, or even require, belief states. One might believe that one has a certain visual image, but this does not entail that the image in question is a proposition one believes; nor does this entail that every visual image requires a corresponding

belief descriptive of that image. Coherentism makes justification depend just on coherence relations among propositions one believes. It thus neglects the evidential role of non-belief sensory experiences in empirical justification. In particular, coherentism apparently permits that what is warranted by coherence relations among one's beliefs can be isolated, or divorced, from the perceptual contents of one's non-belief sensory experiences. This would be a serious defect in an account of empirical justification, justification dependent on sensory experience. Proponents of coherentism have not reached agreement on how to treat the previous objection (see Knowledge and justification, coherence theory of).

Coherentism about justification is often called 'epistemic holism'. According to Quine and others, such holism recommends that we dispense with (1) the analytic-synthetic distinction and (2) the view that our beliefs are tested individually against the input of sensory experience (see Quine, W.V. §3). Our beliefs, given epistemic holism, are confirmed or disconfirmed as a system, collectively rather than individually. In addition, Quine holds, any of our beliefs can be held true come what may, if we make drastic enough changes elsewhere in our system of beliefs; by the same token, none of our beliefs is immune to revision. On this basis, Quine opposes any distinction between the a priori and the a posteriori due to considerations about irrevisability of beliefs. This is not, however, a general argument against the distinction between the a priori and the a posteriori, because some versions of the distinction do not appeal to considerations about irrevisability. Some versions appeal, instead, to considerations about the dependence of relevant evidence on sensory experience.

3 Foundationalism about empirical warrant

The leading alternative to coherentism is *foundationalism* about empirical justification: empirical justification has a two-tier structure in that some instances of such justification are non-inferential, or foundational, and all other instances of such justification are inferential, or non-foundational, owing to their dependence on foundational justification. Bertrand Russell, A.J. Ayer, and C.I. Lewis advanced influential twentieth-century versions of foundationalism about empirical knowledge and justification. Foundationalists differ among themselves on two key matters: the explanation of what precisely constitutes non-inferential, foundational empirical justification, and the explanation of how empirical justification can be transmitted from foundational beliefs to non-foundational beliefs. Opposing the radical foundationalism of Descartes, contemporary foundationalists typically endorse *modest* foundationalism, implying that foundational beliefs need not possess or yield certainty and need not deductively support justified non-foundational beliefs. A foundational belief, minimally characterized, is non-inferentially justified in that its justification does not derive from other beliefs.

Traditionally, empiricist proponents of foundationalism have countenanced foundational justification by non-belief sensory experiences. These empiricists, notably represented by C. I. Lewis, hold that foundational empirical beliefs can be justified by non-belief sensory experiences (for example, your non-belief experience involving your 'seeming to see' a book page) that either make true, are best explained by, or otherwise support those foundational beliefs (for example, the belief that there is, or at least appears to be, a book page here, before you). More recently, proponents of foundational empirical justification by reliable empirical origins have proposed that non-inferential empirical justification derives from a belief's origin in a non-belief empirical belief-forming process (for example, perception, introspection) that is truth-conducive to a certain extent, in virtue of tending to produce true rather than false beliefs. The latter proposal, a species of 'process reliabilism', cites the reliability of a belief's non-belief origin, whereas the previous view invokes, as justifiers, the particular sensory experiences that underlie a foundational belief. Both approaches can, however, accommodate the view that foundational empirical justification is defeasible, or overridable given the acquisition of new evidence (see Foundationalism).

A comprehensive account of empirical justification will explain the nature of sensory experience. If sensory experience does not have a non-conceptual component, it will be ill-suited to serve the purposes of traditional foundationalism about empirical justification. In that case, sensory experience will fail to provide a basis for justified empirical beliefs that does not itself require evidential support. If sensory experience is identical to conceptualization, judgment or belief, then it will itself need evidential support if it is to confer justification on some beliefs; and if this needed support is to be genuinely empirical, it will have to involve more than just propositional relations among beliefs. At least, various proponents of traditional foundationalism have, on this basis, sought a non-conceptual foundation for empirical justification in sensory experience, a foundation that yields a straightforward distinction between a posteriori and a priori warrant.

See also: A priori; Empiricism; Justification, epistemic; Knowledge, concept of; Rational beliefs; Reasons for belief

PAUL K. MOSER

References and further reading

Ayer, A.J. (1936) *Language, Truth, and Logic*, New York: Dover, 2nd edn, 1946.(An accessible defence of extreme empiricism about meaning and knowledge.)

BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.(A defence of coherentism about empirical justification; see §2 above.)

Carruthers, P. (1992) *Human Knowledge and Human Nature*, Oxford: Oxford University Press.(An accessible treatment of issues about the sources of empirical knowledge.)

Goldman, A.I. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(Defends an account of justification and knowledge in terms of reliable belief-forming processes, discussed in §3 above.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1963.(Classic statement of the a priori-a posteriori and analytic-synthetic distinctions; see especially the introduction, §§I-IV.)

Kripke, S.A. (1980) *Naming and Necessity*, Oxford: Blackwell, and Cambridge, MA: Harvard University Press. (An influential treatment of the a priori-a posteriori distinction.)

Lewis, C. I. (1946) *An Analysis of Knowledge and Valuation*, LaSalle, IL: Open Court.(A classic statement of foundationalism that acknowledges a central role for a ‘given’ element in empirical justification.)

Moser, P.K. (1989) *Knowledge and Evidence*, Cambridge: Cambridge University Press.(A defence of a foundationalist approach to empirical knowledge that acknowledges empirical justification by non-belief sensory experiences; see §3 above.)

Moser, P.K. (1993) *Philosophy After Objectivity*, New York: Oxford University Press.(An examination of the kinds of warrant available for claims to objective knowledge, knowledge of the external world.)

Moser, P.K. and Vander Nat, A. (eds) (1987) *Human Knowledge: Classical and Contemporary Approaches*, New York: Oxford University Press, 2nd edn, 1995. (Classical and contemporary selections bearing on the conditions for a priori and a posteriori knowledge. Includes selections from Ayer, Kripke, Lewis, Quine and Russell, among others.)

Quine, W.V. (1953) *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 2nd edn, 1961. (Includes ‘Two Dogmas of Empiricism’, an influential defence of epistemic holism and challenge to the analytic-synthetic distinction.)

Yolton, J. (1984) *Perceptual Acquaintance from Descartes to Reid*, Oxford: Blackwell.(A survey of classical modern approaches to the nature of perceptual experience.)

A priori

An important term in epistemology since the seventeenth century, 'a priori' typically connotes a kind of knowledge or justification that does not depend on evidence, or warrant, from sensory experience. Talk of a priori truth is ordinarily shorthand for talk of truth knowable or justifiable independently of evidence from sensory experience; and talk of a priori concepts is usually talk of concepts that can be understood independently of reference to sensory experience. A priori knowledge contrasts with a posteriori knowledge, knowledge requiring evidence from sensory experience. Broadly characterized, a posteriori knowledge is empirical, experience-based knowledge, and a priori knowledge is non-empirical knowledge. Standard examples of a priori truths are the truths of mathematics, whereas standard examples of a posteriori truths are the truths of the natural sciences.

1 Necessity, analyticity and the a priori

Contemporary understanding of the distinction between the a posteriori and the a priori, as the distinction between the empirical and the non-empirical, derives mainly from Kant's *Critique of Pure Reason* (1781/1787), although versions of it precede Kant in the writings of Leibniz and Hume (see Kant, I. §4). The epistemological distinction between a priori and a posteriori knowledge differs from the logical or metaphysical distinction between necessary and contingent truth, and from the semantical distinction between analytic and synthetic truth (see Analyticity). In particular, the concept of a priori knowledge is not the same as either the concept of what is (logically or metaphysically) necessarily true or the concept of what is true analytically, just in virtue of the meanings of a proposition's constituent terms. Kant's talk of a priori 'modes of knowledge' suggests an epistemological, knowledge-oriented characterization of what is a priori. As standardly characterized, a priori knowledge is knowledge that does not depend on evidence from sensory experience. The previous considerations do not, however, settle the issue of whether every proposition knowable a priori is either necessarily true or analytically true.

A necessarily true proposition is not possibly false, or in Leibniz's words, is true in 'all possible worlds'. Contingently true propositions are possibly false, that is, false in some possible worlds. Traditionally, many philosophers have assumed that a proposition is knowable a priori only if it is necessarily true, presumably on the ground that if a proposition is possibly false, then it requires for its justification supporting evidence from sensory experience. Contingent truths, according to this traditional view, are not candidates for a priori knowledge.

Saul Kripke (1980) has argued that some contingently true propositions are knowable a priori. He cites the knowledge that stick *S* is one metre long at a certain time, where stick *S* is the standard metre-bar in Paris. If one uses stick *S* to 'fix the reference' of the term 'one metre', then, according to Kripke, one can know a priori that stick *S* is one metre long. The truth that stick *S* is one metre long is contingent rather than necessary; for *S* might not have been one metre long. (Application of sufficient heat to *S*, for instance, would have changed its length.) It seems arguable, then, that some contingent truths are knowable a priori, contrary to what many philosophers have assumed. This matter has prompted considerable discussion among contemporary philosophers, with some still contending that no contingently true proposition is knowable a priori. Some of the latter philosophers have noted, with regard to Kripke's example, that 'one metre' can be used either as (1) the name of the length of *S* whatever that length may be, or as (2) the name of a particular length singled out by a speaker. Given option (1), these philosophers hold, the claim that stick *S* is one metre long will be necessary and knowable a priori, and given option (2), the claim that stick *S* is one metre long will be contingent and knowable only a posteriori.

Many philosophers have held that a priori knowledge is restricted to such analytic truths as 'All rectangles have four sides' and 'All bodies are extended'. If such truths are analytic, they are true just in virtue of the meanings of their constituent terms. Such truths differ from synthetic truths, which are true in virtue of something other than just the meanings of their constituent terms (for example, in virtue of observable situations in the world). Synthetic judgments, according to Kant, are 'ampliative' in that they 'add to the concept of the subject a predicate which has not been in any wise thought in it, and which no analysis could possibly extract from it' (1781/1787: A7/B11). Some philosophers, notably W.V. Quine, have contested the viability of any philosophically important distinction between analytic and synthetic truths (see Quine, W.V. §3).

One issue of philosophical controversy is whether any synthetic truth is knowable a priori. Kripke's

aforementioned metre example offers, according to some philosophers, a synthetic truth knowable a priori. Kant held that some synthetic truths, for example, those of geometry, have a kind of necessity that cannot be derived from experience, and can be known a priori. Such synthetic truths, Kant argued, can be known independently of evidence from sensory experience. Kant's doctrine of synthetic a priori truths still generates controversy among philosophers, specifically in connection with such apparently synthetic propositions as 'Nothing can be green and red all over' and 'A straight line is the shortest path between two points'. The later Wittgenstein, for example, proposed that propositions of the latter sort are actually conventional 'rules of grammar', or non-synthetic normative standards for representation (see Necessary truth and convention).

2 Innate concepts, certainty and the a priori

Many philosophers deny that having a priori knowledge requires having innate concepts, concepts that do not derive from, or depend for their being understood on, sensory experience. (Some theorists, in the tradition of Platonism, hold that mathematical concepts, among others, are innate.) Propositions, one might suppose, consist of concepts, perhaps analogously to the way in which sentences consist of terms. Propositions knowable a priori, according to the philosophers in question, need not consist of innate concepts. The notion of a priori knowledge depends on a notion of a priori *warrant*, not on a notion of a non-empirical origin of the concepts constituting the known proposition. A notion involving special conditions for the justification of a believed proposition is not automatically a notion involving special conditions for either the origin or one's understanding of the belief in question.

The notion of a priori knowledge, construed as a notion of non-empirically grounded knowledge, is not the same as a notion of epistemic certainty. Philosophers have understood 'epistemic certainty' in various ways: for instance, as epistemically indubitable belief or as self-evident belief. A belief is epistemically indubitable if and only if it would not be epistemically justifiable to doubt that belief under any circumstance. It is not obvious that a priori warrant for a proposition requires epistemic indubitability of this proposition. A proposition's being warranted a priori for someone seemingly allows for an expansion of their relevant evidence, whereby a proposition justified on the original evidence ceases to be justified on the expanded evidence. (One might, for example, come to appreciate further implications of a proposition that was justified a priori.) A priori justification for a proposition apparently can be subject to 'epistemic defeat' given a change in a priori evidence (see Certainty; Doubt).

Philosophical talk of 'self-evidence' is often unclear. On a literal construal, a self-evident proposition is justified but does not depend on anything else for its justification. The problem in linking a priori warrant to such self-evidence is that a priori warrant is compatible with *inferential* warrant, wherein a proposition owes its warrant to inferential relations with other propositions, as might a theorem in a mathematical system. (It is a separate issue whether *all* a priori warrant might be inferential.) The notion of a priori knowledge should thus be explained independently of a literal construal of self-evidence. Other construals of self-evidence will contribute here only if they elucidate a notion of non-empirical warrant that differs from notions of necessary truth, analyticity and certainty as epistemic indubitability.

3 Prominent explanations of the a priori

Philosophers have long sought an account of the defining feature of truths that humans can know a priori. One result is a variety of accounts of the a priori in circulation. *Psychologism* about the a priori, advanced initially but later opposed by Husserl, claims that a true proposition is knowable a priori by humans if and only if our psychological constitution precludes our regarding that proposition as false. *Linguisticism* about the a priori, endorsed by A.J. Ayer and various other twentieth-century empiricists, states that a true proposition is knowable a priori if and only if our denying that proposition would violate rules of coherent language-use; this view denies the existence of synthetic a priori truths. *Pragmatism* about the a priori, advanced by C.I. Lewis, claims that a true proposition is knowable a priori by a person if and only if it describes their pragmatically guided intention to use a certain conceptual scheme of classification for the organizing of experiences. Lewis argued that pragmatic considerations regarding what suits one's needs guide the way in which one formulates a conceptual scheme. A different view, supported by Roderick Chisholm and many others, affirms that a true proposition is knowable a priori by us if and only if our understanding that proposition is all the evidence we need to see that the proposition in question is true. Yet another view about the a priori is suggested by the later writings of Wittgenstein: A

proposition is knowable a priori by us if and only if our 'forms of life' (that is, human nature as determined by our biology and cultural history) preclude the intelligibility for us of the denial of that proposition. (Wittgenstein did not offer a detailed account of 'forms of life' or of their role in determining what is a priori.) These are the most influential, but not the only, accounts of the a priori in circulation.

A theory of a priori knowledge should identify the strengths and weakness of the aforementioned accounts of the a priori. It should also identify the feature of a priori justification that requires limitation of the set of propositions knowable a priori to the distinctive kind of propositions specified by that theory. Such a theory must avoid confusing the notion of what is a priori with the notions of what is necessarily true, what is analytically true, what is innate, and what is certain. It must also draw a clear distinction between what is a priori and what is a posteriori.

See also: A posteriori; Justification, epistemic; Innate knowledge; Knowledge, concept of; Rational beliefs; Rationalism

PAUL K. MOSER

References and further reading

Carruthers, P. (1992) *Human Knowledge and Human Nature*, Oxford: Oxford University Press.(An accessible treatment of issues about the sources of knowledge, including a priori knowledge.)

Coffa, J.A. (1991) *The Semantic Tradition from Kant to Carnap*, Cambridge: Cambridge University Press.(A history of philosophical reaction to Kant's doctrine of the a priori.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1963.(Classic statement of the distinctions a priori-a posteriori and analytic-synthetic; see especially the introduction, §§I-IV.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.(Referred to in §1 above. Challenges the view that only necessarily true propositions are knowable a priori.)

Moser, P.K. (ed.) (1987) *A Priori Knowledge*, Oxford: Oxford University Press. (Contains ten of the most important recent essays on a priori knowledge and a bibliography of recent work on the topic. The selections in this book treat the positions identified in §3 above.)

Moser, P.K. and Vander Nat, A. (eds) (1987) *Human Knowledge: Classical and Contemporary Approaches*, New York: Oxford University Press, 2nd edn, 1995.(A wide range of classical and contemporary selections bearing on the conditions for a priori and a posteriori knowledge.)

Pap, A. (1958) *Semantics and Necessary Truth*, New Haven, CT: Yale University Press.(A detailed survey and assessment of many prominent seventeenth through twentieth century views on necessity, analyticity and the a priori, including the views of Leibniz, Kant, Locke and Hume.)

Quine, W.V. (1953) *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 2nd edn, 1961. (Includes 'Two Dogmas of Empiricism', an influential challenge to the analytic-synthetic distinction.)

Shanker, S.G. (1987) *Wittgenstein and the Turning-Point in the Philosophy of Mathematics*, London: Croom Helm.(Expounds Wittgenstein's views on mathematical truth and the a priori.)

‘Abduh, Muhammad (1849-1905)

The Egyptian reformer and Muslim apologist Muhammad ‘Abduh was a pupil and friend of al-Afghani. Although deeply influenced by him, ‘Abduh was less inclined to political activism and concentrated on religious, legal and educational reform. His best-known writings are a theological treatise, Risalat al-tawhid (translated into English as The Theology of Unity), and an unfinished Qur’anic commentary, Tafsir al-manar (The Manar Commentary), on which he collaborated with Rashid Rida. One of the key themes of these works is that since modernity is based on reason, Islam must be compatible with it. But ‘Abduh’s ‘modernism’ went hand in hand with returning to an idealized past, and his ‘rationalism’ was tempered by a belief in divine transcendence which limits the scope of intellectual inquiry. In ethics as in theology, he regarded the classical debates as arid and divisive, although on the issues of free will and moral law his position was in fact similar to that of the Mu‘tazila.

1 Faith and reason

‘Abduh trained as an *‘alim* (religious scholar) at al-Azhar where, under al-Afghani’s influence, he developed an interest in Islamic philosophy and a revulsion for traditional teaching methods which encouraged *taqlid*, the unquestioning acceptance of received opinion. The rational liberalism which he imbibed from al-Afghani was, however, only one facet of his thought. In his youth he was drawn to Sufism and, despite his subsequent attacks on popular superstition, he seems never to have lost his respect for those who in some conditions ‘have access in part to the ultimate mysteries and true insights into the visionary world’ (*Risalat al-tawhid*, in Musa‘ad and Cragg 1966: 97) (see Mystical philosophy in Islam). A third influence - the one which is dominant in the *Risalat al-tawhid (The Theology of Unity)* and the *Tafsir al-manar (The Manar Commentary)* - is that of the fourteenth-century Hanbalite jurist Ibn Taymiyya, who fuelled his desire to purify Islam of later accretions and return to the essentials of the faith as practised by the first generations of Muslims.

‘Abduh believed that Islam was the one true religion based on reason and revelation, but that in the course of time it had become distorted by various extrinsic factors. For instance, whereas the Qur’an fosters the scientific spirit by directing man to inquire rationally into the workings of the universe, the Islamic philosophers had uncritically accepted the theories of matter and physics propounded by Plato and Aristotle, with the result that the Islamic world had come to lag behind Europe in science and technology. His rejection of Greek philosophy in favour of modern science was, however, only partial. He accepted the distinction between necessary being, possible things and impossible things, using it to prove the existence of God. He also accepted the distinction between essences and accidents, arguing that reason gives us knowledge of the latter but not of the former. A corollary of this is that it is pointless for theologians to argue about the divine attributes because we cannot know their nature (see Islamic theology).

2 Ethics

On the issue of free will versus predestination, ‘Abduh’s starting point is the recognition that the man of sound mind is conscious of acts which stem from his volition:

> He weighs them and their consequences in his mind and evaluates them in his will, and then effectuates them by an inward power. To deny any of this would be tantamount to a denial of his existence itself, so opposed would it be to rational evidence.
>
> (*Risalat al-tawhid*, in Musa‘ad and Cragg 1966: 62)

However, ‘Abduh is equally insistent that all events in the world are ordered by God in accordance with his knowledge and will. He rejects further inquiry into how human freedom and divine prescience can be reconciled, on the grounds that such speculation is forbidden.

In discussing the moral law, ‘Abduh again begins with an appeal to common sense, arguing that we have no difficulty in recognizing our voluntary actions as good or bad in themselves or by reference to their particular or general consequences. If actions are self-evidently good or bad in the absolute way in which ‘Abduh alleges, however, it might be thought that religion is unnecessary. On the contrary, in matters of right and wrong, rational proof will not obviate conflict because people differ in intelligence, the vast majority being unable to understand

Platonic philosophy or Aristotelian logic. Moreover, because of its stress on God’s pleasure and wrath, religion has a greater impact on ordinary folk than the moralist’s claim that some acts are beneficial and others harmful. In any case there are some elements of the Qur’anic revelation which could not be known by unaided reason. These include the certainty of the afterlife, and the various ritual prescriptions.

All this is far-removed from the traditional Ash‘arite position. It is possible that here ‘Abduh was influenced by Mu‘tazilism as mediated by al-Afghani’s Shi‘ism, or less probably that we should detect the influence of Kantian philosophy. There seems little doubt, however, that his ethical thinking was moulded by the needs of apologetics. This is particularly clear in his essay on Islam and Christianity, in which he replied to Hanotaux, a French cabinet minister who had contrasted the Semitic mentality of Islam - with its transcendentalism, predestinarianism and contempt for individuals - with the Aryan humanism of Christianity, which through the Trinity raised human dignity to that of God.

See also: al-Afghani; Islamic philosophy, modern

NEAL ROBINSON

List of works

‘Abduh, M. (1874) *Risalat al-waridat (Treatise of Mystical Inspirations)*, Cairo.

‘Abduh, M. (1876) *Hashiyya ‘ala sharh al-Dawwani li ’l-‘aqa’id al-adudiyya (Gloss on Dawwani’s Commentary on the Sentences of Adud al-Din al-Iji)*, Cairo.

‘Abduh, M. (1897) *Risalat al-tawhid (The Theology of Unity)*, Cairo; trans. I. Musa‘ad and K. Cragg, *The Theology of Unity*, London: George Allen & Unwin, 1966. (One of ‘Abduh’s major works on philosophical theology.)

‘Abduh, M. (1902) *al-Islam wa-’l-nasraniyya ma‘a al-‘ilm wa ’l-madaniyya (Islam and Christianity in Relation to Science and Civilization)*, Cairo.

‘Abduh, M. and Rida, M.R. (1927-36) *Tafsir al-Qur’an al-hakim (Commentary on the Wise Qur’an)*, Cairo. (Commentary on the Wise Qur’an usually referred to as *Tafsir al-manar* because it originally appeared in instalments in the journal *al-Manar*).

References and further reading

Abu Rabi, I. (1996) ‘The Arab World’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 64, 1082-1114.(Detailed description of the various ways in which modern Arabic philosophy has responded to the issues of modernity and Westernization.)

Adams, C.C. (1933) *Islam and Modernism in Egypt*, London: Oxford University Press.(Classic work, unsurpassed despite its early date.)

Amin, O. (1944) *Muhammad Abduh essai sur les idées philosophiques et religieuses (Muhammad ‘Abduh, Essay on his Philosophical and Religious Ideas)*, Washington, DC.(English translation of the standard Arabic biography.)

Badawi, M.A.Z. (1978) *The Reformers of Egypt*, London: Croom Helm.(Critical analysis by an Egyptian scholar of the views of al-Afghani, ‘Abduh and Rida.)

Hourani, A. (1983) *Arabic Thought in the Liberal Age 1798-1939*, Cambridge: Cambridge University Press. (Devotes only thirty pages specifically to ‘Abduh, but strongly recommended for anyone who wishes to situate his work in its historical context.)

Jomier, J. (1954) *Le commentaire coranique du Manar: tendances modernes de l’exégèse coranique en Égypte (The Manar Commentary on the Qur’an: Modern Trends in Qur’anic Exegesis in Egypt)*, Paris: Maisonneuve. (Painstaking thematic analysis of the Qur’anic commentary by a French Dominican; particularly valuable because of the way in which it distinguishes between the views of ‘Abduh and Rida.)

Abelard, Peter (1079-1142)

Among the many scholars who promoted the revival of learning in western Europe in the early twelfth century, Abelard stands out as a consummate logician, a formidable polemicist and a champion of the value of ancient pagan wisdom for Christian thought. Although he worked within the Aristotelian tradition, his logic deviates significantly from that of Aristotle, particularly in its emphasis on propositions and what propositions say. According to Abelard, the subject matter of logic, including universals such as genera and species, consists of linguistic expressions, not of the things these expressions talk about. However, the objective grounds for logical relationships lie in what these expressions signify, even though they cannot be said to signify any things. Abelard is, then, one of a number of medieval thinkers, often referred to in later times as 'nominalists', who argued against turning logic and semantics into some sort of science of the 'real', a kind of metaphysics. It was Abelard's view that logic was, along with grammar and rhetoric, one of the sciences of language.

In ethics, Abelard defended a view in which moral merit and moral sin depend entirely on whether one's intentions express respect for the good or contempt for it, and not at all on one's desires, whether the deed is actually carried out, or even whether the deed is in fact something that ought or ought not be done.

Abelard did not believe that the doctrines of Christian faith could be proved by logically compelling arguments, but rational argumentation, he thought, could be used both to refute attacks on Christian doctrine and to provide arguments that would appeal to those who were attracted to high moral ideals. With arguments of this latter sort, he defended the rationalist positions that nothing occurs without a reason and that God cannot do anything other than what he does do.

1 Life

In comparison with most other medieval thinkers we know a great deal about Abelard's career, since he left an autobiographical essay entitled *Historia Calamitatum (The Story of My Misfortunes)*, which in effect recounts his life up to about 1132. He was born to a noble and very religious family in Le Pallet, near Nantes in Brittany, in 1079. Roscelin of Compiègne was one of his earliest teachers, and later in Paris he studied under William of Champeaux. In these years Abelard's main interest was logic, or dialectic as it was then called, which also included metaphysical issues such as the status of universals as well as psychological topics such as the role of images in thought. It was over the subject of universals that Abelard came into acrimonious conflict with William of Champeaux, and eventually he left Paris to set up his own schools nearby, first in Corbeil and later in Melun.

Having achieved a reputation as a subtle logician, Abelard moved on to study theology with Anselm of Laon, but here again he fell into competition with his teacher and even lured away some of his students. The stay at Laon was short, and by 1114 Abelard was back teaching logic in Paris. However, theology was to move more and more to the centre of his interests as his career progressed.

It was at this juncture that Abelard's famous romance with Heloise began. Heloise was only 17 but was well educated, thanks to the devotion of her guardian and uncle, Fulbert. It was by securing from Fulbert the job of teaching her that Abelard was able to pursue his amorous ends. Abelard confesses that during the height of their passion he was more given to writing love poems than studying logic. Heloise became pregnant, and Abelard took her away to his home in Brittany without Fulbert's permission. There she gave birth to their only child, Astrolabe. In order to reconcile himself to Fulbert, Abelard married Heloise and brought her back to Paris where she again lived with her uncle. The marriage was to have been kept secret in order not to endanger Abelard's career as a cleric, but Fulbert broadcast the news, with the result that Abelard had Heloise removed to a convent. Fulbert was so enraged by this that he hired thugs who attacked Abelard and castrated him. Subsequently, Heloise remained a nun for the rest of her life and became herself well known in this vocation. Abelard resumed his scholarly career after first entering the abbey of St Denis.

The next period of Abelard's life was intensely productive. His most important logical works, as well as the first version of his *Theologia (Theology)*, were probably produced while he was at St Denis. Some of his enemies, well placed in the church hierarchy, found in the *Theologia* what they took to be errors, and at a church council summoned in April 1121 in Soissons, Abelard's work was condemned (this condemnation was later revoked).

However, Abelard thrived on controversy, and from 1122 to 1127 he continued his teaching and work at a retreat in the country near Quincy. His reputation became even greater, and students from all over Europe flocked to his lectures on both logic and theology.

In 1127 Abelard accepted an appointment as the abbot of a monastery in Brittany known for its moral laxity. By this time he had become a strong proponent of church reform, especially in moral matters. His efforts to establish discipline at the monastery only angered the monks there, to the point where they tried several times to murder him. Eventually Abelard was relieved of these duties, and returned to Paris to teach at the school on Mont Ste-Geneviève.

Following the council at Soissons Abelard had revised his *Theologia* several times, but his doctrine of the Trinity as well as his view of sin again angered the church hierarchy, including the now very powerful Bernard of Clairvaux. In June of 1140, a council of bishops at Sens condemned several of his positions. Abelard's appeal to the pope was countered by Bernard, but eventually a reconciliation was arranged. Abelard died at Cluny on 21 April 1142.

Although in his own day Abelard was famous and influential, his impact on later generations was less than might have been expected. Peter Lombard's *Sentences*, which became the standard theological textbook in the thirteenth century, owed much to Abelard's *Sic et Non (Yes and No)*, and strong realism about universals was never an option after Abelard's attack; but many of Abelard's logical innovations were forgotten once the corpus of Aristotle's logic became available in the West. Abelard was hardly ever referred to after 1200, and a fellow nominalist like William of Ockham, writing around 1317, seems totally unaware of his work.

2 Works

In order to understand and assess Abelard as a philosopher, it is important to consider not only his works on logic but also his writings on theology. At risk of considerable over-simplification, his works may be divided into those composed before his stay in the community near Quincy (1122-7) and those written during and after that stay. As C.J. Mews (1985) has pointed out, this break seems to correspond with certain revisions Abelard made in both his logical and theological teachings.

In the earlier group lie his two great logical works, *Dialectica* and the *Logica ingredientibus (Logic for Beginners)*. The latter is really four works, consisting of glosses on four earlier and important logical works, Porphyry's *Isagōgē*, Aristotle's *Categories*, Aristotle's *Peri hermenias* or *De Interpretatione*, and on Boethius' *De differentiis topicis* (see Aristotle; Boethius, A.M.S.; Porphyry). These were interpretative commentaries on what was later called the 'old logic'. The *Dialectica* is an independent treatise roughly covering the same topics treated in the old logic, but without direct commentary. In theology, Abelard produced in the period before Quincy the version of his *Theologia* known as the *Theologia 'summi boni' (Theology of the Highest Good)* and the collection of conflicting passages from scriptures and earlier church fathers and doctors known as the *Sic et Non*.

After he moved to Quincy, Abelard's *Theologia* went through several revisions, resulting in a long work divided into five books now referred to as the *Theologia Christiana (Christian Theology)*. A final reworking of his theological ideas in shorter form is found in the *Theologia Scolarium (Theology for Students)*. Also in the area of theology from the later period is his *Dialogus inter philosophum, Iudaeum et Christianum (Dialogue of a Philosopher with a Jew and a Christian)* and his *Ethica* or *Scito te ipsum (Ethics, or Know Thyself)*. Abelard's work in logic seems to have diminished in the post-Quincy period, but he did compose a new and important gloss on Porphyry's *Isagōgē* called the *Glossule super Porphyrium* or *Logica 'nostrorum petitioni sociorum' (Logic in Response to the Request of Our Comrades)*, probably while at Quincy. Because of this diminishing interest, Abelard never made much use of the 'new logic', that of Aristotle's *Analytics*, *Topics* and *Sophistical Refutations*, which became available in western Europe in the 1130s (see Translators).

3 Pagan wisdom and Christian faith

Abelard was, at least in his later life, a devout Christian committed to the moral reform of the church and the defence of traditional orthodox Christian beliefs. However, he was also well read in the literature of pagan Greece and Rome, and was convinced that among the works of the philosophers of the ancient world could be found useful models of moral life and more precise understandings of dogmas such as the Trinity than were to be found

even in the Old Testament. His *Theologia* evidences an overriding concern to defend the use of pagan learning in explicating and defending the Christian faith, while at the same time refuting 'pseudo-philosophers' who in his own day were creating difficulties for belief by misusing that most important legacy of the ancients, logic.

Abelard found that the ancient philosophers, although they had lived before the time of Christ, had already taught some of the most important doctrines of the Christian religion, namely the immortality of the soul, the existence of a creator, the supreme importance of living virtuously, the likelihood of retribution in the next life for sins in this one and even, in Plato's case, the doctrine of the Trinity (see Plato; Trinity). It was not Abelard's view that the ancient sages came to these insights simply by the use of reason and other natural cognitive faculties; rather, he held that God had given them a certain revelation as a reward for their exemplary lives.

4 Logic

Abelard's actual knowledge of ancient philosophy outside of ethics was very limited and was often reliant on the fragmentary accounts of it given by early Christian thinkers such as Augustine. In the field of logic he at least had translations of a few primary sources, such as Aristotle's *Categories* and *Peri hermenias* and Porphyry's *Isagōgē*. Also in his possession were the commentaries of Boethius on these sources, as well as a few of Boethius' own treatises. The rest of Aristotle's logical works became available only late in Abelard's life, after his own interest in logic had waned. In this legacy Abelard's genius discovered many difficulties and omissions, and his reflection on these led him to develop a quite original logical theory, even though he always presented it as an extension and modification of the traditional Aristotelian framework.

Perhaps the most un-Aristotelian feature of Abelard's logical theory is the central role he gives to sentences and what sentences say, rather than to terms. Any sentence, even a question or an imperative, says or proposes something called a *dictum*, although only assertions commit the speaker to the truth of what is said. These *dicta* are the primary subjects of truth and falsity, with assertions being true or false only insofar as their *dicta* are. Abelard recognizes that the genuine contradictory of an affirmative sentence is one in which the negation operates on the whole *dictum* of the latter, not just on the predicate. His elaborate theory of conditionals rests on the view that a conditional is true only when the truth of the antecedent's *dictum* requires the truth of the consequent's *dictum*. We have here the makings of a genuinely 'propositional' logic akin to that developed in ancient times by the Stoics, that is, a logic in which the basic logical relationships of entailment, opposition and so on are seen as holding between propositions (see Stoicism).

Abelard's appreciation of the saying-function of sentences leads him to a very original analysis of what is involved in verbs, and in the verb 'to be' used as a copula. The tendency of Aristotelian logic is to view simple categorical sentences as having three parts: two noun phrases linked by some form of the copula. Thus 'Socrates is a human' has as its parts the noun phrases 'Socrates' and 'a human', plus the copula 'is'. Abelard, on the other hand, sees such sentences as falling into two parts, a subject noun phrase and a predicate verb phrase. The peculiar function of the verb is thus to provide the saying-force, without which the string of words would be at best a list rather than an assertion. The copula really has no signification on its own, but merely acts to transform a noun phrase into a verb.

Nouns, he thinks, turn out to be implicit verbs, as can be inferred from their covert importation of tense. In 'Some man is a philosopher' 'man' extends over only presently existing men, so that the whole sentence really means 'Something which is now a man is a philosopher', where the verb implicit in the subject is made explicit. Although nouns often have 'appellation' - for example, in many contexts they are taken as applying to or naming certain individual things - these things are nevertheless not what the noun signifies. Signification has both a psychological and more properly a semantic aspect. The former will be discussed later, but as for the latter, Abelard refers to *status* or 'natures' as what is signified by both verbs and nouns. These are apparently referred to by verbal nominalizations such as 'being a man' or 'walking'. Perhaps the best way to view a *status* is as that which, in a *dictum*, corresponds to the predicate of the corresponding sentence, for Abelard talks in much the same way about *status* as he does about *dicta*. Thus a *status* or nature associated with a predicate is what might be said of some thing by using that predicate in a sentence in which the subject noun named that thing.

Status and *dicta* turn out to be important both for Abelard's analysis of modal propositions and for his treatment of conditionals. A modal proposition, such as 'It is possible for this man, who is sitting, to be standing', turns out to

have a false sense in which the whole *dictum* of the sentence ‘This man, who is standing, is sitting’ is the apparent subject of the modality, and another true sense in which what is being asserted is both that standing is compatible with the *status* of being a man and that the man in question is in fact sitting. Conditional propositions are necessarily true only if the *dictum* of the antecedent explicitly or implicitly contains the dictum of the consequent. Necessarily true conditionals such as, ‘If there is a rose, there is a flower’, the truth of which depends simply on the logical relationship between the natures of being a rose and being a flower, are sometimes called ‘laws of nature’ by Abelard and are fundamental for science (see Language, medieval theories of; Natural philosophy, medieval §§8-9).

5 Ontology

Although Abelard never had a kind word for his old teacher Roscelin and the two quarrelled bitterly over matters of theology, it is clear that the student in large measure took over his master’s view that logic is about vocally produced sounds qua signifiers, not about the things those sounds may be used to refer to. Such a view sharply separates logic from sciences such as physics, which talk directly about things without regard to any use of them as signifiers. Instead, physics treats logic as a tool which can be of help in any inquiry that employs language. Abelard conducted a vigorous attack on ‘realist’ views of logic, such as that held by William of Champeaux. This dispute focused on ‘universals’, for example, items such as genera and species (that is, the items which Aristotle claimed were each predicated of many subjects by true affirmations), since these were crucial to Aristotelian logic and science. As viewed in Abelard’s day, the issue was whether universals were things existing independently of language (realism) or whether they were only the words that referred to things (nominalism). Abelard came down clearly on the latter side (see Nominalism; Realism and anti-realism; Universals).

William’s version of the realist position treated genera and species as something like materials common to many things. The different things they are common to are differentiated by opposed forms which also exist in these materials. Abelard argued that the end result of this view is that one and the same individual thing may have opposed characteristics at the same time. Abelard’s own positive view shifted somewhat from that of his *Logica ingredientibus*, where it is simply the vocally produced physical sound (the *vox*) which is a universal, and that of the *Logica ‘nostrorum petitioni sociorum’*, where it is words (*sermones*) that are universals and different words might share the same physical sound. However, the basic idea remains: universality results from some thing being used as a sign of many.

Where Abelard’s view differs radically from that of his teacher Roscelin is in the role of the *status*. Abelard claims that the *status* ‘being a human’ is the reason why the term ‘human’ applies to the many things to which it does apply; but, he says enigmatically, this is ‘no thing’. Nevertheless it is that which verbs, and nouns, signify. Do they signify, then, without signifying anything? In the *Logica ‘nostrorum petitioni sociorum’*, Abelard explains that through the idea associated with a universal word there does not have to be some particular thing that one thinks of when one uses the word. For example, in just the way that there does not have to be some thing that I want when I want a hood - in other words, I can want a hood without there being any particular thing that I want - similarly I can think of something through the idea associated with ‘human’ even though there is no particular thing (or human) that I think of.

Dicta, too, turn out not to be things. Abelard thinks it would make the necessity of any conditional proposition incomprehensible if we thought of the *dicta* of the antecedent and consequent as things. Nevertheless, his whole semantic theory is deeply dependent on talk about both *status* and *dicta*. Even in his theory of the Trinity, without *status* he has no grounds on which to defend the objective differentiation of the three divine persons. Apparently Abelard thought it was possible to evade the embarrassing ontological consequences of a ‘realist’ approach to logic while retaining all the advantages of an approach which bases logic in objective facts, for, as discussed below, he eschews the idea of founding logic on relations between mental entities such as ideas or concepts. Perhaps the chief conundrum for interpreters of Abelard is how this is supposed to be possible.

Abelard also introduces here the concept of ‘impersonal’ propositions wherever there is a statement of grammatical subject-predicate form but the subject is the nominalization of a verb phrase or a sentence (see above). Since neither verb phrases nor sentences name anything, Abelard thinks it is a mistake to take their nominalizations as nonetheless naming something. It is not just that ‘being a human’ fails to name a thing; it doesn’t name at all. Nouns like *status* and *dictum*, since their use is based in their being predicated of items like

being a human and Socrates' being a human, likewise do not really name at all. Is this tantamount on Abelard's part to saying that the language in which we talk about the foundations of logic necessarily escapes the scope of the very logic about which it is designed to talk? A full study of the implications of Abelard's logic has yet to be written.

Another area of ontology in which Abelard was very inventive concerns the doctrine of sameness and distinctions. While defending the doctrine of the Trinity, Abelard was led to develop the idea that items might be the same 'essentially' even though some predicates true of one are not true of the other, because the items are distinct 'in property'. An example is the way that a wax statue is the same 'essentially' as the wax that composes it; but the statue is not the matter for the statue, even though the wax is. This same idea lies behind the distinction of words from the physical sounds in which they are realized. The word is the same 'essentially' as the sound, but nevertheless two words might be the same as one vocal sound and yet not the same as each other.

6 Psychology of signification

Although logic is about words that are realized in sound, Abelard acknowledges that those sounds constitute words and thus have properties significant for the logician only because they express ideas in the mind of the speaker. His views on the mental side of signification are fairly elaborate and reflect his careful reading of Aristotle's *Peri hermenias*.

The fundamental notion here is that of an 'idea' (*intellectus*), which is a genuine mental thing, an act or a disposition toward a mental act. The idea is not the content of such an act, but rather the idea itself always has some content. Only beings with reason can have ideas and thus ideas must be distinguished from both sensings and imaginings. Abelard locates this distinction in the way an idea isolates some property for attention, while sensings and imaginings grasp something without separating out some single property or complex of properties of that thing from its other properties. In other words, any idea reflects a certain abstractive thought process. What is isolated and thought of through the idea is some *dictum* or *status*: in other words, what is signified by the word or words which express the idea in question.

Abelard often talks of linguistic expressions as signifying the ideas they express, but he makes it clear that this sense of 'signification' is not one which permits inferring that language necessarily talks about ideas. He is very insistent that we must avoid a semantic theory which would have all language ultimately be about something in the mind. This applies also to the images which the mind creates in order to give itself something to focus on when it thinks. Abelard does not think that these images are mental things in the way ideas are; indeed, they are not 'things' at all. My image of a four-sided tower is not a four-sided thing which exists, either in my mind or in external reality. Such images cannot be equated with what words signify, since, Abelard claims, we can use exactly the same image when we use words with different meanings. The 'attention' of the mind allows us to focus on one property at one time and another at another time, in both cases using the very same image. The image is not the content of the idea that uses it.

The contents of ideas, then, are the *status* or *dicta* which are the 'things' (as opposed to the ideas) which words signify. Since, as noted above, *status* and *dicta* are not themselves things, Abelard's position here teeters on the brink of self-contradiction. However, he argues vigorously that we need not always find some thing for our terms to signify. When I am thinking of a man, I need not be thinking of any particular man. In the sentence 'A man is in the house' 'a man' does not signify any particular man, not even the one which happens to be in the house, if such a person exists. Abelard is keenly aware that noun phrases in certain contexts are used 'non appellatively', in other words in such a way that one cannot point to any particular thing and say that it is the referent of the phrase. His observation here clearly dovetails with the view mentioned in §5, that an idea can be of something even though there is no particular thing it is of, and is thus part of Abelard's perhaps dubious way of avoiding the apparent ontological implications of his semantics.

7 Ethics

Abelard's positions on ethical questions were heavily influenced on the one hand by what he knew of Stoic ethics, and on the other hand by the Christian doctrine of rewards and punishments in the next life for righteousness or sinfulness in this one. Stoic ethical doctrines were accessible to Abelard through the writings of Cicero and

Seneca, as well as through the detailed accounts of Stoic ethics in the writings of Augustine and other church fathers. His views largely amount to an attempt to take the Stoic doctrine of moral virtue and vice and make it the basis of how God judges souls to be meritorious (deserving of eternal blessedness) or guilty (deserving of damnation). In the course of this effort, however, a number of important notions are drastically redefined.

Whereas in the Stoic view virtues and vices were settled practices of deliberately choosing good and bad courses of action, for Abelard they are inclinations which manifest themselves prior to choice and over which we have little control. In his view, a tendency to get angry too easily is a vice, but having that vice does not of itself make one sinful and deserving of punishment. It is deciding to yield to that tendency that is sinful. In Abelard's view, then, there is no logical inconsistency in supposing that a person has many vices but is in fact a very righteous person deserving of God's rewards. In fact, Abelard says, having vices to fight against is required in order to attain the highest merit. The person who has nothing but virtues easily does the right thing and thus merits reward less.

Neither does Abelard think that sin or righteousness lies merely in willing something. Here his view changed somewhat over time, but in his *Ethica*, probably a late work, he adopts a conception of a will which equates it with a desire, not with any deliberate decision to undertake some course of action nor with a mere faculty for such decisions. On this point Abelard was clearly not in accord with earlier writers such as Augustine and Anselm. Given such a definition of will, Abelard has no trouble showing that a person can have a will to do something they should not and yet not sin, and also sin while not having any will for doing something they should not. The first case is illustrated by a man who lusts after a married woman but restrains himself and never adopts the intention to seduce her. The second is exemplified by a servant who kills his master in self-defence; he never willed the death of his master but rather only desired his own life, a perfectly legitimate will. And yet (on Abelard's very rigorous moral standard) he sinned in not allowing himself to be killed rather than engage in a killing.

Once the firm intention to do wrong is formed, however, the guilt exists, and it is not increased by the actual performance of the deed. Thus Christ can rightly condemn men for having committed adultery in their hearts even if they have not actually seduced a woman. Sin, in Abelard's view, lies entirely in the consent to do what one knows one should not, or to omit doing what one knows one should. (The men Christ condemned, Abelard believes, had consented to such a seduction; they had not merely lusted after the woman.) This alone makes one guilty and deserving of retribution from God. That sinners must know they are wrong is important for Abelard. One of the theses which the bishops at the council of Sens objected to was Abelard's claim that, because they thought they were doing what they were supposed to do, the men who crucified Christ were not guilty. The consent or intention to do something may be objectively bad - what one consents to may in fact be something one ought not consent to - and yet, because of the ignorance of the person who consents, no guilt be incurred. The consent is sinful when and only when it amounts to contempt for what is good and right, which Abelard often equates with contempt for God. Likewise, a consent is meritorious only when it shows respect and love for what is good, in other words, for God (see Sin).

Although sin is the only thing that makes one deserving of punishment, both divine and human punishment can in certain circumstances be justifiably visited on those who do not deserve it. Abelard describes a case where a woman, who was only trying to keep her baby warm, accidentally smothered it to death. The judge, he says, is justified in punishing her in order to set an example that will remind others to be more careful. Human justice, he claims, is legitimately more concerned with outward deeds than with inner intentions, since it is the former, not the latter, that directly affect the peace and welfare of the community. Similarly, Abelard allows the doctrine that, for obscure reasons of his own, God is justified in condemning to eternal death some unbaptized infants who have never sinned.

8 Theodicy

Abelard argues for God's existence as follows. We recognize that we ourselves are not made by ourselves but by something else; however, we as rational beings are certainly superior to the world of non-rational things. This would not be so if that world were something that caused itself, because anything which relies only on itself for its subsistence is superior to anything that relies on something other than itself. Hence that world is brought into existence by something else, a maker or ruler, and this we call God (see God, arguments for the existence of).

Abelard recognizes that logically this argument is less than completely compelling, but he thinks that in this area

we should accept those arguments which make an appeal to our better sentiments. His attitude here reveals how far Abelard was from developing theology as a branch of rational philosophy or metaphysics in the way that would be attempted later, in the thirteenth century (see Aquinas, T.).

One conclusion that Abelard reached on the basis of such arguments was that everything that is or occurs has a reason for being or occurring. To allow otherwise would be to claim that there are some things that God either makes occur or allows to occur, without having a full reason for why these rather than other things should occur. This Abelard thinks would detract from the divine goodness. On this basis, Abelard draws the conclusion that God could not do or omit anything other than what he at some time does do or omit. He will not allow that God might be faced with several alternatives which are equally best, so that there could be no reason for him to choose one rather than the other. And, of course, God cannot choose an alternative to which there is a better alternative. From God's point of view, then, all that he does or omits he must do or omit (see Omnipotence). This view too was among those condemned at Sens, and Abelard admitted that few agreed with him on this subject. Although he cleverly used his theory of modalities to argue otherwise, the position seems clearly headed toward fatalism of the Stoic variety. That Abelard proposed such a view at all shows to what extent he was attracted by the extreme rationalism represented in ancient Stoicism.

See also: Aristotelianism, medieval; Bernard of Clairvaux; Language, medieval theories of; Logic, medieval; Nominalism; Platonism, medieval; William of Champeaux

MARTIN M. TWEEDALE

List of works

Abelard, P. (*c.*1120?) *Logica ingredientibus (Logic for Beginners)*, ed. B. Geyer in *Peter Abaelards Philosophische Schriften (Peter Abelard's Philosophical Writings)*, Beiträge zur Geschichte der Philosophie des Mittelalters 21, Heft 1-3, Münster: Aschendorffschen Verlagbuchhandlung, 1919-27.(The most important Latin text for Abelard's logic. The *Gloss on Porphyry* has been translated by Paul Spade in *Five Texts on the Medieval Problem of Universals: Porphyry, Boethius, Abelard, Duns Scotus, Ockham*, Indianapolis, IN: Hackett, 1994, pages 26-56.)

Abelard, P. (*c.*1120?) *Dialectica*, ed. L.M. de Rijk, Assen: Van Gorcum, 1970.(After the previous work, the most significant of Abelard's philosophical writings.)

Abelard, P. (*c.*1124?) *Logica 'nostrorum petitioni sociorum' (Logic in Response to the Request of our Comrades)*, ed. B.Geyer in *Peter Abaelards Philosophische Schriften (Peter Abelard's Philosophical Writings)*, Beiträge zur Geschichte der Philosophie des Mittelalters 21, Heft 1-3, Münster: Aschendorffschen Verlagbuchhandlung, 1919-27.(A short piece on Porphyry's *Isagōgē*, that reviews the earlier gloss in the *Logica ingredientibus*.)

Abelard, P. (*c.*1120?) *Theologia 'summi boni' (Theology of the Highest Good)*, ed. E.M. Buytaert and C.J. Mews in *Petri Abaelardi Opera Theologica* III, Corpus Christianorum, Continutatio Mediaevalis XIII, Turnholt: Brepols, 1987, 39-201.(The first version of Abelard's *Theologia*.)

Abelard, P. (*c.*1125?) *Theologia Christiana (Christian Theology)*, ed. E.M. Buytaert in *Petri Abaelardi Opera Theologica* II, Corpus Christianorum, Continuatio Mediaevalis XII, Turnholt: Brepols, 1969, 5-372.(The second and most complete version of Abelard's *Theologia*, including his doctrine on the Trinity.)

Abelard, P. (*c.*1135?) *Theologia Scolarium (Theology for Students)*, ed. E.M. Buytaert and C.J. Mews in *Petri Abaelardi Opera Theologica* III, Corpus Christianorum, Continuatio Mediaevalis XIII, Turnholt: Brepols, 1987, 203-613.(The third version of Abelard's *Theologia*.)

Abelard, P. (*c.*1125?) *Dialogus inter philosophum, Iudaeum et Christianum (Dialogue of a Philosopher with a Jew and a Christian)*, ed. R. Thomas, Stuttgart-Bad Cannstatt: Frommann, 1970; trans. P.J. Payer, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1979. (Abelard's attempt to view Christianity as much more in line with ancient philosophy than was Judaism.)

Abelard, P. (*c.*1138?) *Ethica (Ethics)* or *Scito te ipsum (Know Thyself)*, ed. and trans. D.E. Luscombe in *Peter Abelard's Ethics*, Oxford: Clarendon Press, 1971.(Contains Abelard's fullest explanation of the doctrine of sin.)

References and further reading

Jolivet, J. (1969) *Arts du language et théologie chez Abailard (Arts of Language and Theology in Abelard)*, Études de philosophie médiévale 58, Paris: Vrin.(The most detailed account of how Abelard used his logic in his theology.)

Luscombe, D.E. (1988) 'Peter Abelard', in P. Dronke (ed.) *A History of Twelfth Century Western Philosophy*, Cambridge: Cambridge University Press, 279-307.(A brief survey of Abelard's philosophical views.)

Luscombe, D.E. (1969) *The School of Peter Abelard*, London: Cambridge University Press. (Describes the influence of Abelard's thought in the early scholastic period.)

Marenbon, J. (1997) *The Philosophy of Peter Abelard*, Cambridge: Cambridge University Press.(The most recent biography of Abelard.)

Mews, C.J. (1985) 'On dating the works of Peter Abelard', *Archives d'histoire doctrinale et littéraire du moyen âge* 52: 73-134.(The most thorough attempt so far to order and date all of Abelard's works.)

Mews, C.J. and Jolivet, J. (1990) 'Peter Abelard and His Influence', in G. Floistad (ed.) *Contemporary Philosophy*, Dordrecht: Kluwer, vol. 6, ed. R. Klibansky.(A complete bibliography of recent Abelard studies.)

Tweedale, M.M. (1976) *Abailard on Universals*, Amsterdam: North Holland. (Translates and analyses from a modern perspective the main texts relating to Abelard's ontology.)

Aberdeen Philosophical Society

The Aberdeen Philosophical Society (1758-73) played a formative role in the genesis of Scottish common sense philosophy. Its founder members included the philosopher Thomas Reid and the theologian George Campbell. Its discussions favoured the natural and human sciences, particularly the science of the mind, and one of its central concerns was the refutation of the work of David Hume.

Popularly known as the ’Wise Club’, the Aberdeen Philosophical Society was founded in January 1758 by a core group of six men that included the philosopher Thomas Reid and the theologian George Campbell. The Society initially brought together individuals who were either associated with the two Aberdeen colleges or connected with the local political magnate, Lord Deskford. During the next decade, figures such as Alexander Gerard and James Beattie joined the club, and the Society became a respected body within the European republic of letters because of the growing reputation its leading members had achieved through their publications. However, by the late 1760s the Society was in decline, and it finally dissolved in 1773 due to internal divisions caused by college politics.

Unlike other prominent Scottish groups of the period, the Wise Club carefully circumscribed the scope of its proceedings. According to its constitution the Society was to exclude the discussion of the traditional scholarly fields of grammar, philology and history, and limit itself to ’philosophical’ subjects. Under this rubric the founding members included reports on matters of fact, inductive generalizations concerning the phenomena of the material and mental realms, disquistions on false theoretical systems and erroneous methods of philosophizing, and the exploration of the relations between philosophy and the practical arts. Consequently, the Society focused largely on topics drawn from the natural and human sciences, and rarely touched on the literary and artistic questions which interested many of the other polite clubs of the Scottish Enlightenment (see Enlightenment, Scottish). The most popular subject for discussion was the anatomy of the mind, particularly as it related to morals. Other areas canvassed at meetings included politics, education, political economy, mathematics, natural philosophy, natural history and agricultural improvement. Thus, like most provincial clubs of the eighteenth century, the Aberdeen Philosophical Society covered a broad spectrum of topics, but the scientistic approach to moral philosophy which it espoused was rooted in the local philosophical tradition initiated by the moralist George Turnbull.

Although their debates ranged widely, the members of the Society were centrally concerned with the refutation of the writings of David Hume. One of the club’s founders, Robert Trail, had earlier criticized Hume in a sermon published in 1755, and his fellow members subsequently took up the charge. Alexander Gerard and George Campbell attacked Hume’s critique of religion; while Thomas Reid, John Farquhar and James Beattie among others challenged various aspects of his epistemology. The Society therefore functioned as a forum for the articulation of a common-sense reponse to Humean scepticism, and Thomas Reid did not exaggerate when he wrote to Hume in 1763 that ’If you write no more in morals, politics, or metaphysics, I am afraid we [that is, the Wise Club] shall be at a loss for subjects’ (Ulman 1990: 57).

See also: Common Sense School

PAUL WOOD

References and further reading

Ulman, H.L. (ed.) (1990) *The Minutes of the Aberdeen Philosophical Society 1758-1773*, Aberdeen: Aberdeen University Press.(Includes a transcription of the minute books of the Wise Club, along with biographical entries on the members, a detailed analysis of the proceedings and a comprehensive bibliography.)

Abhinavagupta (*c*.975-1025)

Abhinavagupta was a Kashmiri philosopher, theologian and early exponent of the Hindu Tantra, often counted as the most illustrious representative of the nondual Śaivism of Kashmir. Author of influential Sanskrit works dealing with the philosophy of recognition and the theological interpretation of the Śaivite scriptures, including the encyclopedic Tantrāloka (Light on the Tantras), he also wrote definitively on Indian aesthetic theory. The tradition of the nondual Śaivism of Kashmir gathers up the teachings of several related lineages of northern Śaivite philosophers which developed in Kashmir between the ninth and thirteenth centuries, such as Vasugupta, Kallaṭa, Somānanda and Utpala. Basing his writings on these authors as well as on revealed texts, Abhinavagupta propounds a tantric alternative to the restrictive and orthodox Mīmāṃsā and elaborates a challenge to the later mainline Vedānta. He offers the earliest theoretical bases for a complex and sophisticated Hindu Tantra based on the notion of Śiva as the nondual and all-pervading consciousness. His writings highlight the centrality of the Goddess as the Śakti *or power of consciousness. Abhinavagupta elaborates the ritual and meditative methods for the experiential and blissful recognition of Śiva as the intrinsic self-identity of the practitioner.*

1 General context and life

Abhinavagupta is considered to be one of the most sophisticated and enduringly definitive theoreticians of the medieval Hindu Tantra. His work as a theoretician of the Hindu Tantra and as an interpreter of the revealed scriptures of Śaivism is intertwined with his life as a devotee of Śiva. Abhinavagupta's contributions to the philosophy of language and as a philosophical interpreter of aesthetic theory remain definitive. His tantric synthesis is termed the Trika-Kaula because it skillfully melds together doctrinal and ritual elements drawn from these two Śaivite preceptorial lineages.

Abhinavagupta (*c*.975-*c*.1025) was born to a rich and noble Brahman family residing in the city of Śrinagar. His mother, Vimalā, died when he was still young and her death affected him greatly. The family were devotees of Śiva and Abhinavagupta matured in an atmosphere charged with religious devotion and dedication to learning. He began his studies with his learned father, Narasimhagupta, but quickly began visiting teachers in Kashmir and elsewhere. While he was studying literature and poetry, he was overcome with an intoxicating devotion to Śiva. After this, he studied everything he could: traditional texts of the dualistic and monistic Śaivism, literature, drama and aesthetic theory, Indian philosophical thought and the various branches of Śaivism. He also studied with Buddhist and Jaina teachers. His love of learning and spiritual search led him to Jālandhara where he encountered the tantric master, Śambhunātha, who initiated him into the practices of the Kaula tradition.

This early period of study and spiritual practice lead to a mature life dedicated to the absorption of knowledge in an atmosphere of extreme religious fervour. He never married, and spent his life living in the homes of his many teachers. At the height of his fame, he was revered as an authoritative and charismatic tantric master whose authority as guru or teacher was enhanced by the fact that he was considered to be a *mahāsiddha* (perfected and accomplished mystic). Twenty-one of his works are extant and there are references to titles of twenty-three others now apparently lost. His primary disciple was Kṣemarāja who wrote works applying his master's teachings. We have no definitive information about Abhinavagupta's death. Local Kashmiri legend has it that the great master walked into a cave with 1,200 of his disciples and simply disappeared.

Given his deep erudition, the intellectual context of Abhinavagupta's writings is very broad, in part due to his relatively late date with regard to the earlier traditions of Indian philosophy and religion. In addition to the revealed Hindu literature and to the texts produced by his predecessors in the Śaivism of Kashmir, Abhinavagupta's work resonates with practically all that precedes him in Indian thought, including the traditional brahmanical or Vedic literature, the debates of the philosophical traditions, the mysticism of the Yogis, the various Bhakti traditions and the varieties of Buddhist and Jaina philosophical discourse.

2 Śiva's nonduality

From a doctrinal point of view, Abhinavagupta's articulation of the nondual Śaivism of Kashmir is to be distinguished philosophically by its assertion that what is termed 'Śiva', the absolute and primordial consciousness, is nondual. Moreover, this nondualism differs in important ways from the Vedāntic *advaita*. For the

Kashmiri nondual Śaivites, the nondualism or Śiva does not in any way imply that the world and all who dwell in it are illusory. On the contrary, Abhinavagupta asserts that this world is real precisely because it is only Śiva, the absolute consciousness. However, this assertion of the reality of the world does not fall into a position of naive realism. Rather, it seeks to articulate the enlightened and transformed vision of the mystic for whom the paradoxical omnipresence of Śiva has become a tangible experience. Thus, Śiva's nonduality allows without contradiction for the arising of duality and diversity within it. In addition, Abhinavagupta's expression of nondual Śaivism places emphasis on the ultimate power (*śakti*) that abides inseparably with the absolute consciousness. Alhough one, this transcendent power of consciousness has many faces: the powers of freedom, of grace or revelation and of manifestation or concealment.

This doctrine resides at the very core of Abhinavagupta's entire intellectual production, which is large in scope and has not been fully explored by scholarship. Often, facets of his intellectual production have been characterized in a piecemeal fashion, text by text, or in terms of the various initiatory lineages (rather than schools), thought to determine the content of individual texts. However, there is another, more general way to approach Abhinavagupta's work. In philosophical terms, we find in it three connected although differentiable intellectual agendas which are best characterized in terms of the mode of knowledge that predominates in each of them: inference, revealed scripture and enlightened knowledge. Each of these agendas articulates an aspect of this core doctrine of Śiva's nonduality.

3 Philosophy of recognition

Abhinavagupta's first intellectual agenda revolves around the philosophical interpretation of the doctrine of recognition. His writings on the 'school' of recognition, primarily in the *Īśvarapratyabhijñāvimarśinī (Commentary on the Recognition of the Lord)*, are systematic philosophical exposition and argument. An appeal is made to the authority of good reasoning, and arguments through inference are employed as the primary mode of philosophical discourse. The major audience seems to have been external, for these are primarily polemical works of intellectual debate aimed at philosophical opponents.

Abhinavagupta's writings on the philosophy of recognition are continuous with the debates put forward by the philosophers of Nyāya-Vaiśeṣika, the Mīmāṃsā, the Advaita Vedānta, the schools of Buddhist logic and the traditional Grammarians (see Mīmāṃsā; Nyāya-Vaiśeṣika; Vedānta). Buddhism was a vigorous presence in the Kashmir of Abhinavagupta's time. While his opponents in these texts include proponents of Sāōkhya, as well as Vaiṣṇavas and Jainas, his primary philosophical arguments were aimed at the Buddhist logicians (see Sāōkhya). The Buddhists marshalled arguments against the primary categories of the Śaiva philosophy including the notion of the self, the idea of the Lord, the *śakti* and the manifested universe (see Buddhist philosophy, Indian).

In his works on recognition, Abhinavagupta elaborates closely reasoned arguments to counter the Buddhists' criticisms. He points to the notion of Śiva as the absolute consciousness and posits a reality that is totally free, completely unbounded and blissful in its intrinsic nature. This he says is the reality of consciousness, saturated with power (*śakti*) of at least five different kinds: the powers of consciousness, bliss, will, knowledge and action. It is this absolute consciousness surcharged with power that both dwells within all beings as their true nature and mutates to appear as the forms composing the manifest universe.

Abhinavagupta contends that it is because of doubt that the recognition of this omnipresent consciousness does not take place. Consequently, his philosophy of recognition is meant to present the good reasoning that will remove doubt. This precipitates the process of recognition of what is already the case, namely, that the self is in fact the powerful Lord, filled with *śakti* and creative of the visible universe.

This initial description of an absolute consciousness reveals the shape of a problematic in Abhinavagupta's work that bridges philosophy, theology and mysticism. Ordinary human awareness and perception do not reveal that an absolute consciousness is operative either as the inner reality of the self, or as the underlying reality of the visible and manifest universe. Therefore, Abhinavagupta is concerned to show why this is the case: that is, how and why the obscuring of the absolute takes place and how and why the absolute comes to be revealed. The first of these questions involves Abhinavagupta in descriptions of the nature of bondage and contraction, involving a series of limiting sheaths and impurities. The second leads him to a discussion of tantric metaphysics and mysticism.

4 Tantric metaphysics and mysticism

The second major agenda found in Abhinavagupta's works is what might be called a systematic theology of the Hindu Tantra. This is best exemplified by his *Tantrāloka (Light on the Tantras)* where he interprets the tantric metaphysics and esotericism of the revealed scriptures, particularly the *Mālinīvijayottara Tantra*. In this aspect of his works he writes as an authoritative religious and theological exponent of Śaivism. His appeal is to the authority of revealed scripture, which must be systematically interpreted in the light of reason and spiritual insight so that apparent contradictions may be rationalized. The major audience for this aspect of his work were the varied schools of Śaivism prevalent in the Kashmir of his day. Thus, his mode of discourse is that of religious theology and the systematic exposition of the doctrines of Śaivism.

The themes taken up in the *Tantrāloka (Light on the Tantras)* (and elsewhere in Abhinavagupta's Śaivite theology) are complex and detailed. One strand of discourse explores the contraction of the ultimate consciousness and the manifestation of the world in terms of an emanationist scheme composed of thirty-six progressively more manifest principles of reality (*tattvas*). Another important strand centres on a mysticism of the Word, filled with vibration and dwelling in unity with the absolute consciousness at the transcendent level of reality. This supreme Word of consciousness is termed the supreme mantra *aham* (I am) and it is understood as the perfectly full and blissful egoity.

The third major intellectual agenda in his works might be called a practical theology or mysticism of Śaivism, encountered primarily in his elaboration of Kaula mysticism in his *Parātriṃśikāvivaraṇa (Longer Commentary on the Thirty Verses on the Supreme)*. Here, Abhinavagupta advances interpretations of the many Śaivite lineages into which he had been initiated as well as offering interpretations of the meanings of tantric ritual and practice. The appeal is to the authority of his own enlightened experience, and the audience seems to have been primarily his own initiated disciples whom he was addressing in the role of spiritual teacher.

A central concern in this dimension of his work is the elaboration of the meaning of *jīvanmukti* (liberation while still alive) in terms of the mystical experiences of the yogin, who, by penetrating the core of that absolute consciousness, achieves identity with Śiva and experiences the universe as continuously unfolding and being reabsorbed into his own consciousness.

Abhinavagupta's discussion of tantric initiation includes a long clarification of the descent of the energy of Śiva's grace. In addition, he elaborates the intrinsic meanings of the great variety of rituals in the Hindu Tantra linked to an elaborate pantheon of tantric deities. The primary hermeneutical move throughout is one of interiorization and of interpretation in terms of the mechanics of an overarching consciousness. Thus, while the earlier tantric tradition will speak of external goddesses who are to be appeased by a variety of transgressive offerings in the tantric ritual, Abhinavagupta will consistently recast these goddesses as the forces of the absolute consciousness present in the depths of the practitioner's own being.

Abhinavagupta's core statements about Śiva modulate into an ambitious and multivalent intellectual and theological enterprise. This enterprise will also allow him to put forward an innovative interpretation of aesthetic theory that posits an analogy between the absorption of the yogin and the refined aesthete's delectation of the experience of art. Abhinavagupta's work will have manifold direct and indirect influences on all that follows him in the subsequent evolution of the Hindu Tantra.

See also: Hindu philosophy; Mysticism, history of; Mysticism, nature of

PAUL E. MULLER-ORTEGA

List of works

Abhinavagupta (*c*.990-*c*.1014) *Īśvarapratyabhijñāvimarśnī (Commentary on the Recognition of the Lord)*, ed. M. Rāma, Srinagar: Research Department, Jammu and Kashmir Government, 1918; trans. K.C. Pandey, *Bhāskarī: The Īśvara Pratyabhijñā Vimarśinī*, The Princess of Wales Saraswati Bhavana Texts, Lucknow: Superintendent, Printing and Stationery, 1954.(The *Commentary on the Recognition of the Lord* sets out Abhinavagupta's arguments about the nature of the process of recognition of Śiva.)

Abhinavagupta (*c*.990-*c*.1014) *Śrī Mālinīvijayavārtika*, ed. M. Kaul, Srinagar: Research Department, Jammu and Kashmir Government, 1921.(A commentary on the first verse of the *Mālinīvijayottara Tantra* which is considered by Abhinavagupta to be the most authoritative scripture on which his Trika-Kaula synthesis is

based.)

Abhinavagupta (*c*.990-*c*.1014) *Paramārthasāra (Essence of the Highest Truth)*, Srinagar: Research Department, Jammu and Kashmir State, 1916; French translation by L. Silburn, *Le Paramārthasāra*, Publications de l'Institut de Civilisation Indienne, Paris: Éditions E. de Boccard, 1957.(The *Essence of the Highest Truth* is a summary text of Abhinavagupta's essential teachings.)

Abhinavagupta (*c*.990-*c*.1014) *Parātriṃśikālaghuvṛtti (The Short Gloss on the Supreme, the Queen of the Three)*, ed. J. Zādoo, Srinagar: Research Department, Jammu and Kashmir Government, 1947; Italian translation by R. Gnoli, *La Trentina della Suprema*, Turin: Boringhieri, 1978.(*The Short Gloss on the Supreme, the Queen of the Three* is a concise statement on the nature of tantric *sadhana* dealing especially with mantra.)

Abhinavagupta (*c*.990-*c*.1014) *Parātriṃśikāvivaraṇa (Longer Commentary on the Thirty Verses on the Supreme)*, ed. M. Rāma, Srinagar: Research Department, Jammu and Kashmir Government, 1918; trans. J. Singh, *A Trident of Wisdom*, Albany, NY: State University of New York Press, 1989.(The *Longer Commentary on the Thirty Verses on the Supreme* comments on the same āgamic verses as the previous text, but here Abhinavagupta places great emphasis on the emanationist system of the *tattvas*, the theology of the Word and the esoteric mysticism of the Trika-Kaula practices.)

Abhinavagupta (*c*.990-*c*.1014) *Tantrāloka (Light on the Tantras)*, M. Rāma and M. Kaul (eds), Srinagar: Research Department, Jammu and Kashmir Government, 1918-38.(The *Light on the Tantras* presents Abhinavagupta's major statement of scriptural interpretation and exegesis.)

Abhinavagupta (*c*.990-*c*.1014) *Tantrasāra (Essence of the Tantras)*, ed. M. Rāma, Srinagar: Research Department, Jammu and Kashmir Government, 1918.(The *Essence of the Tantras* is Abhinavagupta's epitome of his longer *Tantrāloka*.)

References and further reading

Dyczkowski, M. (1992) *The Stanzas on Vibration*, Albany, NY: State University of New York Press.(An excellent introduction to the Spanda lineages of Kashmiri Śaivism.)

Masson, J.L. and Patwardhan, M.V. (1969) *Śāntarasa and Abhinavagupta's Philosophy of Aesthetics*, Bhandarkar Oriental Series 9, Poona: Bhandarkar Oriental Research Institute.(A detailed discussion of aspects of the aesthetic dimension of Abhinavagupta's work.)

Muller-Ortega, P.E. (1989) *The Triadic Heart of Śiva: Kaula Tantricism of Abhinavagupta in the Non-Dual Shaivism of Kashmir*, Albany, NY: State University of New York Press.(An introduction to the Kaula mysticism of the heart as elaborated by Abhinavagupta.)

Padoux, A. (1990) *Vāc: The Concept of the Word in Selected Hindu Tantras*, Albany, NY: State University of New York Press.(Treats in detail the philosophy of language and the topic of mantra.)

Pandey, K.C. (1963) *Abhinavagupta: An Historical and Philosophical Study*, Chowkhamba Sanskrit Studies, vol. 1, Varanasi: Chowkhamba Sanskrit Series Office. (The only existing in-depth study of Abhinavagupta's life and work. Considerably dated but still contains much that is useful.)

Sanderson, A. (1986a) 'Mandala and Agamic Identity in the Trika of Kashmir', in *Mantra et Diagrammes Rituels dans l'Hindouisme*, ed. A. Padoux, Paris: CNRS. (Offers solutions to many historical and interpretive problems surrounding the Śaivism of Kashmir.)

Sanderson, A. (1986b) 'Purity and Power among the Brahmans of Kashmir', *The Category of the Person: Anthropology, Philosophy, History*, M. Carrithers, S. Collins, S. Lukes (eds), Cambridge: Cambridge University Press, 1986, 190-216.(Deals with many historical and interpretive problems surrounding Śaivism.)

Silburn, L. (1970) *Hymnes de Abhinavagupta*, Publications de l'Institut de Civilisation Indienne, Paris: Institut de Civilisation Indienne avec le concours du Centre National de la Recherche Scientifique.(Silburn presents the tantric dimensions of Abhinavagupta's thought.)

Silburn, L. (1975) *Hymnes aux Kālī: la Roue des Energies Divines (Hymn to Kālī: the Wheel of Divine Energies)*, Publications de l'Institut de Civilisation Indienne, Paris: Editions E. de Boccard.(Another dimension of Abhinavagupta's thought.)

Singh, J. (trans. and ed.) (1991) *The Yoga of Delight, Wonder, and Astonishment: A Translation of the Vijñānabhairava*, Albany, NY: State University of New York Press. (A useful translation of one of the definitive tantras often referred to by Abhinavagupta.)

Abravanel, Isaac (1437-1509)

Abravanel is often seen as having a unique position in Jewish philosophy, between the end of the Middle Ages and the beginning of the Renaissance. His ideas point both to the past - especially to Maimonides - and to the future, in his approach to the questions of history and of authority in the state. His defence of what he takes to be religious orthodoxy is carried out with serious attention to the arguments of his predecessors. Abravanel takes great pains to understand their reasoning. He even supplies them with additional arguments, before he presents what he takes to be a decisive objection. In particular he expounds Maimonides' thought in considerable detail, defending him from his critics, while also insisting that Maimonides misrepresented the religious notions he analyses.

Abravanel's most original work lies in his view of history as either natural or artificial. Most human history is artificial, since it represents life in rebellion against God. The best form of government is not a monarchy, despite the views of most Jewish philosophers. For a monarchy does not essentially replicate the relationship of God with his subjects, and other forms of government can produce relatively successful societies. The Messiah, who will eventually transform artificial into natural history, is not a king but more a judge and prophet. He will establish the perfect society through a divine miracle. As long as the state is an absolute monarchy, however, its citizens owe absolute obedience to its ruler.

1 Life and works

Isaac Abravanel lived a politically active life during a particularly difficult time for the Jews in the Iberian Peninsula. Born in Lisbon, he received an education which was not concerned exclusively with Jewish subjects and from an early age he was writing philosophy and theology. He worked as the treasurer of the ruler of Portugal, and then in Spain, until the expulsion of the Jews in 1492 drove him to Naples. He then moved to Venice, where he did important political work for the authorities. His time in Italy seems to have been his most productive as a writer, and by his death he had composed a large number of works. Many of these are mainly theological, commentaries on the books of the Bible and other Jewish writings, but he also wrote more strictly philosophical works dealing with the thought of Maimonides (§3) and the particular theological and philosophical topics that Maimonides had made so controversial in Jewish philosophy. These include prophecy, providence, the creation of the world and the principles of Judaism. Abravanel is remarkable for his careful descriptions of the arguments of Maimonides and his opponents; indeed he lavishes so much time on exposition it is often difficult to discriminate his own opinion among those of the thinkers he discusses. He tends to argue that both Maimonides and his leading critics are mistaken in their arguments. Powerful though those arguments may appear, they need to be replaced by better arguments which accord more surely with the principles of Judaism.

2 The principles of religion

The cornerstone of Abravanel's critique of Maimonides lies in his response to Maimonides' theory of prophecy (see Prophecy §4). Maimonides provides what seemed to Abravanel to be an excessively naturalistic account of prophecy in his *Guide to the Perplexed*, and Abravanel was determined to establish that this could not be the Jewish position. Abravanel argued that prophecy is not a natural phenomenon but a miracle established by God. An individual requires no special characteristics, such as the wisdom that Maimonides had supposed was a minimal prerequisite of any authentic prophecy. On the contrary, Abravanel argues, God can make anyone prophesy whom he wishes. It is not true, Abravanel reminds his readers, that only Moses received prophecy at Sinai. On the contrary, Scripture makes it clear that all Israel participated directly in the communication from God. Nor is it the case, as Maimonides supposed, that only Moses could achieve prophecy through his intellect alone. Other prophets too surpass reliance on imagination when prophesying. Nor is prophecy a matter of dreaming; it is a miracle and unique. Abravanel was concerned that the stature of Maimonides as a rabbinic jurist would mislead Jews into thinking that the doctrine of the *Guide to the Perplexed* is an accurate guide to the correct understanding of the Jewish religion. In that book, Maimonides argues that (apart from the vexed question of creation and a few related issues) there is no essential incompatibility between Aristotelian philosophy (of the Neoplatonic variety then widely accepted) and the principles of religion. True, Maimonides had argued that ordinary believers should not have their faith challenged by having to work out how the philosophical and scriptural systems of thought are to be connected. Abravanel seeks to show that the biblical and philosophical theories are incompatible, and that the

latter cannot be seen as a valid distillation of the former.

He argues further that the view that prophecy is a miracle is the only view that is compatible with the religious texts of Judaism. The prophet's natural abilities are insufficient to produce prophecy of themselves, and the only necessary condition of being a prophet that human beings can achieve is moral excellence. Since all prophecy is miraculous, there is no essential distinction in this regard between Mosaic and non-Mosaic prophecy. Moses' prophecy is certainly superior to that of the other prophets in degree, but there is no qualitative difference between it and the revelations received by other prophets. The difference is that the prophecy of Moses came from God directly without going through the 'active intellect', the imagination or material apprehensions. Moses, moreover, prophesied while fully conscious. The message he received was entirely conceptual; it was grasped wholly by the intellect. A less perfect form of prophecy, intellectual prophecy, comes from God via the active intellect. It reaches the intellect of the prophet and not his imagination. At a still lesser level, imaginative prophecy consists of images, parables and mysteries. These may seem to resemble dreams, but in fact they are quite different. Imaginative prophecy receives the objective truth from God, albeit in a rather garbled form; dreams are merely subjective ideas produced by the imagination when not under any rational control. The least perfect form of prophecy is that experienced by the Israelites at Sinai through sight and sound, without going through either the intellect or the imagination. The prophet is the passive recipient of what flows from God, and the words and images of the prophecy itself are a direct result of God's will. The real distinction between Mosaic and non-Mosaic prophecy is simply that in that the former is received directly from God. The active intellect, moreover, is not to be understood in the Maimonidean way as a more or less free-standing source of natural events, but rather as a completely determined instrument of the divine will.

Although Abravanel seeks to counter Maimonides' theory of prophecy and to replace it with a theory which in his view more accurately coheres with the religious texts of Judaism, it is noticeable that he stays very close to the Maimonidean methodology in *Perush le-Moreh Nevukhim*, his commentary on the *Guide to the Perplexed*. In *Rosh Amanah (Principles of Faith)*, Abravanel takes up the defence of Maimonides' 'Thirteen Principles', which had led to much controversy in medieval Jewish philosophy. Maimonides had identified thirteen principles of belief that form the basis of the immortality promised to all Israel in the Talmud. The singling out of these thirteen beliefs seemed to claim for them a special status as the core expression of the theological essence of Judaism. But that would seem to imply that there are many other aspects of the religion that are not basic dogmas and so may be questioned without offending orthodoxy. Maimonides' approach to formulating a Jewish creed was criticized on this basis, and Abravanel stoutly defends the importance of Maimonides' principles against his critics, especially Crescas (§3) and Albo (§2), while finally arguing that these principles should not be regarded as the axioms of Judaism but just as very important beliefs, among many other doctrines and practices. The point of the principles, he suggests, is to impress upon ordinary believers what they should believe, since they are unlikely to have studied enough of the Torah to have understood the whole breadth of the 613 commandments. Maimonides is seen as presenting his principles as a religious aid for the intellectually unsophisticated. The more sophisticated understand that there are no basic dogmas in Judaism, and that all the commandments constitute the essence of the religion.

3 Political philosophy

The state arises as a necessity only through the expulsion of Adam from Eden. It will survive until the coming of the Messiah. Political life reflects our spiritual exile. It can never be perfect, although some states are more successful than others at fulfilling the spiritual as well as the political needs of citizens. In his commentaries on the Bible, Abravanel provides an account of what he takes to be the best form of society. This proves largely to be a recapitulation of Mosaic society. He sees this ideal state as consisting of lower courts, a high court and a ruler, perhaps a king. The officials of the lower courts are chosen by the people and handle their local affairs for themselves, while the high court or Sanhedrin is appointed by the ruler and establishes the juridical structure of the state as a whole. The members of this court are to be selected from the priests and levites. What is interesting in this notion of a mixed constitution is that it radically reduces the religious role of the king. There had been something of a tradition in Jewish philosophy of emphasizing the role of the king, given biblical passages which seem to suggest that choosing a king was a duty imposed by God.

Abravanel deals with these passages by arguing that what they mean is that choosing a king in Israel was not a duty, but simply a practice that was allowed. If a king is to be chosen, the biblical account tells the Jews what

characteristics he is to have. But there is no necessity to select a king for the state to be properly organized. Abravanel gives the examples of republican Rome and Venice to suggest that the desirable features of a monarchy can be easily replicated in a very different form of society. Israel does not require a king, because the judges can do whatever a king could do. Indeed, even judges are not essential, since God and the divine law, the Torah, can organize society correctly. Judges, of course, are required to carry out sanctions for disobedience to divine law. But in a society that is truly regulated by God, there would be no need for sanctions. Gentiles, since they do not have the Torah, do require political organization, and a king might be the right person to carry out this task for them. But from the experience of the Scriptures, there is no reason to think that a monarchy represents the best form of government for Israel. For the nation can rely on divine support and leadership for its welfare.

In states where a king disregards the laws and becomes a tyrant, his subjects have no right of rebellion. As subjects they must obey the monarch unconditionally; otherwise they are not subjects. The king in the state replicates God in the world. He has the right to take action not simply as regulated by the law but as is required by the particular facts of the case. As far as Israel is concerned, the choice of the king is a matter for God, and no one else has the right to remove or challenge him. Gentiles may rebel against monarchs who are not behaving as they ought, but Jews should not. Even a Gentile king should not be overthrown, since even he is in the place of God as far as his Jewish subjects are concerned, carrying out whatever punishment God thinks appropriate for his people. There is no point in Jews actively trying to set up a government which accords with Messianic rule. They must wait for God to bring this about when he decides the time is right. In a sense, then, Gentile rulers rule by divine right and with some degree of arbitrary authority, but Jewish rulers do not; they rule within the context of Jewish law. There are strong religious arguments, Abravanel argues, forbidding rebellion against either type of monarch. Within the context of the perfect society, however, in Messianic times, there is no need for a king. It would be better to be ruled by a group of judges guided by the will of God. It is not unlikely that this somewhat muffled but still perceptible anti-monarchistic line has its origins in Abravanel’s knowledge of Christian political thinkers, whose reflections he applies to Jewish political problems (see Political philosophy, history of).

4 Philosophy of history

History is the result of God’s decisions, and he is entirely free to do whatsoever he wishes, having created the world *ex nihilo* (see History, philosophy of). Both form and matter are divine creations, so there are no limitations on God’s power. God has not set up a natural mechanism which then operates independently of him, as the thinking of Maimonides might seem to suggest. Rather, the events that befall humanity are due either to the direct influence of God, the actions of another supernatural being, or the choices made by the exercise of human freedom. History starts with the creation of the universe. It consists of a divinely organized pattern of events. The end of history is salvation, which will be established by the Messiah and which will see Judaism triumph over its enemies. Adam was a perfect being who could have stayed in the Garden of Eden and developed his spiritual potentialities, but he freely chose to act otherwise. He disobeyed God through eating from the forbidden tree of knowledge, and in consequence he became subject to death and life in a hostile environment. Yet the possibility of salvation remained, and the Jews were established to represent the continuity of salvation. Their role is identical to the role of history itself, to attain final salvation for the whole of humanity. To make this possible God has helped the Jews, providing them with prophecy, which finally led up to the revelation at Sinai. He also took them to the land of Israel, which was entirely appropriate for their spiritual perfection and prophecy. Like Adam, though, they sinned and were punished. The Temple was destroyed, and they were sent into exile, a state which will continue until the coming of the Messiah, when history will come to an end, the Jews will be redeemed and perfect peace will prevail. At that time, humanity will be transformed into a final state of perfection and fulfilment. The sovereignty of the Messiah will be universal, the distortion of humanity since the exile from Eden will be ended, and human beings will once again be in a position to realize their own potentialities. The physical universe will be replaced by a spiritual realm in which human souls become immortal as they contemplate eternally the nature of the deity.

Abravanel’s general philosophical approach is entirely coherent, in that he sets out to link what he takes to be orthodox rabbinic Judaism with rational understanding. Where Maimonides goes awry, he argues, is in seeking religious correlates for his philosophical views. The sort of philosophy that attracted Maimonides is completely incompatible with Judaism, something which the extraordinary status of Maimonides in Jewish law tends to disguise. Abravanel has his finger on a vital issue here, in that it is the constant argument of Maimonides that if

one looks at scriptural and legal texts in the right sort of way one comes to realize that they are just another way of expressing philosophical truths. Abravanel seeks to establish an understanding of Judaism close to its literal formulation, and he argues that the apparent contradictions which Maimonides highlights between religion and philosophy are in fact real contradictions which show that Aristotelian philosophy cannot do justice to the Jewish religion. The thought of Maimonides is worthy of detailed study, since as a system in itself it represents an impressive attempt at a rational grasp of Judaism. All the same the effort eventually comes to nothing.

While Maimonides is a far more radical thinker than Abravanel when dealing with metaphysics, the position is reversed when it comes to political philosophy. Abravanel is prepared to interpret scriptural passages far more freely than is Maimonides. And he arrives does so quite uninhibitedly in his critique of monarchy as a system of government. His motives here arise from his attempt at establishing the genuine position of rabbinic Judaism as he sees it, but they are also mixed with his experience as a politician serving under a wide variety of rulers. The remarkable nature of his thought has led to its widespread study in both Jewish and Christian contexts.

See also: History, philosophy of; Maimonides, M.; Political philosophy, history of; Prophecy

OLIVER LEAMAN

List of works

Abravanel, I. (1494) *Rosh Amanah (Principles of Faith)*, trans. M. Kellner, London: Associated University Presses, 1982.(A detailed introduction and notes of one of Abravanel's most accessible works.)

Abravanel, I. (1498) *Shamayim Chadashim (New Skies)*, Roedelheim, 1829; repr. in *Opera Minora*, London: Gregg International, 1972.(An account of creation, where Abravanel argues that belief in creation is an axiom for Judaism.)

Abravanel, I. (1501) *Mif'alot Elohim (Works of God)*, Venice, 1592; repr. in *Opera Minora*, London: Gregg International, 1972.(Another account of his views on the creation of the world.)

Abravanel, I. (before 1509) *Perush le-Moreh Nevukhim (Commentary on the Guide to the Perplexed)*, trans. A. Reines, *Maimonides and Abrabanel on Prophecy*, Cincinatti, OH: Hebrew Union College, 1970.(His analysis of Maimonides' arguments, critical but always respectful.)

Abravanel, I. (before 1509) *Perush 'al-ha-Torah (Commentary on the Torah)*, Jerusalem: Torah we-Da'at, 1969. (An interesting approach to the Bible, in which Abravanel introduces many contemporary political concerns.)

References and further reading

Kellner, M. (1980) 'Rabbi Isaac Abravanel on Maimonides' Principles of Faith', *Tradition* 18: 343-56.(A useful summary of the topic.)

Kimmelman, R. (1995) 'Abravanel and the Jewish Republican Ethics', in D. Frank (ed.) *Commandment and Community: New Essays in Jewish Legal and Political Philosophy*, Albany, NY: State University of New York Press, 195-216.(An excellent analysis of the precise reasons for Abravanel's lukewarm attitude towards monarchy as a form of government.)

Netanyahu, B. (1972) *Don Isaac Abravanel: Statesman and Philosopher*, Philadelphia, PA: Jewish Publication Society.(The standard work on the life and thought of Abravanel.)

Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Cambridge: Cambridge University Press, 393-7.(A short but useful summary of his thought.)

Trend, J. and Loewe, R. (eds) (1937) *Isaac Abravanel: Six Lectures*, Cambridge: Cambridge University Press. (Some important contributions to a general understanding of Abravanel.)

Weiler, G. (1988) 'Antipolitics Articulate and Triumphant in Abravanel', in *Jewish Theocracy*, Leiden: Brill, 69-85.(An interesting approach to the understanding of Abravanel's political thought.)

Abravanel, Judah ben Isaac (*c*.1460/5-*c*.1520/5)

Judah ben Isaac Abravanel was born in Lisbon. After the expulsion of the Jews from Spain in 1492, Leone, as he was known, and his family migrated to Naples, but fled two years later following the French invasion. After brief residences in various Italian cities, Leone returned to Naples where he served as court physician to the Spanish Viceroy. Well-versed in the sciences of his day, including physics, medicine and philosophy, whether Jewish, Islamic or Christian, he composed his major work, Dialoghi d'amore (Dialogues of Love), in 1501-2. Although the work influenced such important thinkers as Montaigne, Bruno and Spinoza, its main influence was in literature rather than philosophy. Its style resembles that of other Renaissance works in the ambit of Ficino's commentary on Plato's Symposium but, unlike these works, it is neither philosophical commentary nor courtly literature. Adopting the idiom of courtly love and drawing on Platonic and Neoplatonic sources, it complements them with mythological, biblical and Aristotelian sources to produce a novel synthesis of Plato and Aristotle with ideas drawn from the pagan and the revealed traditions, aiming to demonstrate that love is the animating principle of the universe and the cause of all existence, divine as well as material.

The three dialogues between Philo, the poetic lover, and his beloved Sophia address the relations between love and desire, the universality of love and the origin of love. Each discussion pivots on an apparent opposition between Philo's Aristotelian and Sophia's Platonic views. The discussion of the relations between love and desire raises fundamental questions about the relations of soul and body.

1 Life

Son of the well-known Jewish thinker and statesman Isaac Abravanel, Judah ben Isaac Abravanel, known as Leone Ebreo, was born in Lisbon (the Italian Leone rendering the Hebrew Judah, in accordance with custom). Despite the family's fame, our knowledge of Leone's life is scant and rife with rumour and surmise. The following is restricted to what is relatively certain.

After serving for years as treasurer to the Portuguese King Alfonso V, Don Isaac Abravanel and his family fled Lisbon for Spain in 1483 when, after Alfonso's death, Don Isaac was accused of conspiring against the new king. He was soon summoned to the service of Ferdinand and Isabella and raised funds needed in their consolidation of power. The monarchs none the less decreed the tragic expulsion of the Jews from Spain in 1492. The Abravanel family migrated to Naples but were forced to flee again by the wars following the French invasion of 1494. After brief residences in various Italian cities, including Genoa, Barletta and Venice, Leone returned to Naples and became court physician to the Spanish Viceroy, Don Gonsalvo de Cordoba. The last reliable evidence about him is a document dated 1520 exempting 'Master Leon Abarbanel, the physician' and his family from all tribute in recognition of his services to the Viceroy. He died in Naples at some time between 1520 and 1525.

Leone was well-versed in the sciences of his day, including physics, astronomy, medicine and philosophy from the Presocratics to the Renaissance, spanning the Jewish, Christian and Islamic philosophical traditions. During his sojourns in various Italian cities, he visited the Italian academies, met celebrated Renaissance thinkers and wrote a treatise *De Coeli Harmonia (On the Harmony of the Heavens)*, at the request of Pico della Mirandola (probably the elder Pico, Giovanni, rather than his nephew Gianfracesco, who studied Hebrew under another well-known Jewish thinker, Yohanan Alemmano) (see Pico della Mirandola, G.). Leone also composed poetry, including an autobiographical Hebrew poem, *T'lumah 'al ha-Z'man (A Plaint on Time)*.

2 History and structure of *Dialoghi d'amore*

Uncertainties and controversies surround Leone's famous work, but by the author's own testimony, it was written in 5262 of the Hebrew calendar, that is, 1501/2. It was published in 1535 at Rome, when Leone's friend Mariano Lenzi 'rescued the work from the shades in which it was buried'. It first appeared in Italian, but the language of its original composition is disputed. Spanish, *Ladino* and Hebrew have their advocates, all more or less nationalistic in their motives and all prepared to argue that so erudite an author as Leone Ebreo should exhibit a more elegant style, even in an acquired language, than the published text presents. The only tangible evidence for a Hispanic original is a single late *Ladino* manuscript extant in the British Museum. Arguments for a Hebrew original are based on the presumed Jewish audience of the work, the long delay in its publication, and the survival of different

Italian versions.

The desire to claim Leone extends to his religious affiliation. The title page of the second and third editions of *Dialoghi d'amore*, describes Leone as Hebrew by nation but Christian by faith, prompting claims that he converted late in life. But there is no evidence of such a conversion, and the inscription, published in a period of religious intolerance (1541 and 1545), may well be bogus or wishful.

Ironically, in view of the efforts to claim them, *Dialoghi d'amore* is often dismissed as derivative of the Renaissance Platonist tradition that began with Marsilio Ficino's commentary on Plato's *Symposium* (see Ficino, M.). Yet the influence of *Dialoghi d'amore* extended far wider than Ficino's work. In the twenty years following its appearance the work had five Italian editions, three Spanish translations, two French translations, and translations into Latin and Hebrew. It influenced thinkers from Montaigne and Burton to Bruno and Spinoza, whose library contained a Spanish edition. Its poetic, dialogical style and what may superficially appear as an indiscriminate blending of sources focused its abiding acknowledged influence more in literature than in philosophy.

Dialoghi d'amore comprises three discussions of love as the animating principle of the universe. Philo is the poetic lover; Sophia, his beloved. The first dialogue discusses the relations between love and desire; the second the universality of love; the third the origin of love. The theoretical discussions in each dialogue are framed by a brief preliminary dialogue represented as an actual exchange between lovers, in which Philo voices his desire for Sophia and she critically refuses his seductive attempts to unite with her. The dialogical structure of the text arises from a fundamental difference between two philosophical views about love, broadly stated: (1) that love and desire are essentially the same, since we desire what we love, since love and desire are always for the good, and since genuine desire is based upon knowledge; (2) that love and desire are opposites, since love originates in knowledge of what is and is good, desire in knowledge of what is lacking in being and in goodness. The first opinion is Philo's and can loosely be called Aristotelian; the second is Sophia's and can be identified as Platonic.

The entire exchange, of course, is an allegory of philosophy. Philo's desire for Sophia is the philosopher's quest for wisdom, which is human perfection. The three successive dialogical attempts to resolve an apparently fundamental disagreement between Plato and Aristotle are also attempts to delineate the relations between the human and the divine. The dialogues progress from the more to the less evident, gaining in abstraction and complexity. Each later discussion develops the conclusions of what has gone before. The second discussion is twice as long as the first; the third, over twice as long as the second. No resolution of the question is presented, prompting the received opinion that Leone meant to compose a concluding fourth dialogue. This inference is unwarranted. The twenty years between the completion of *Dialoghi d'amore* and Leone's death suggests that he had ample time to complete the work had he considered it either possible or necessary. Rather, the lack of an explicit resolution mirrors the structure of the Platonic dialogues, especially the *Symposium*, the model for all Renaissance writings on love.

3 Philosophical significance of *Dialoghi d'amore*

The Platonic form that situates *Dialoghi d'amore* among other Renaissance discussions of love unfortunately obscures as much about its philosophical lineage and import as it discloses. Western scholars typically read Renaissance Platonic texts in Christian perspective, ignoring or misconstruing the influence of Islamic and Jewish philosophy, traditions that are seen as predominantly Aristotelian and thus fundamentally at odds with Renaissance Platonism, especially as regards the central issues of the *Dialoghi d'amore*. As a result, the influence here of such thinkers as Maimonides, Avicenna and Averroes tends to be judged rather superficially and on the basis of the few explicit references to them. For example, the possible influence of Avicenna's *Risalah fi'l-'ishq (Treatise on Love)*, or of Maimonides' focus on the intellectual love of God are rarely mentioned, and the Jewish aspect of Leone's work is reduced to its biblical allusions, for example Leone's claim that the Platonic ideas have a Mosaic origin (see Maimonides, M.; Ibn Sina).

Presuming that the literary genre of *Dialoghi d'amore* reflects sheer Christian Platonism, moreover, occludes the subtlety of Leone's resolution of the tension between the Platonic and Aristotelian approaches and covers over the contributions which the Aristotelian view of the soul can make to an understanding of desire, love and knowledge. Indeed, Leone's literary style may reflect a deliberate intention to occlude the radical thesis of the dialogue, a

deistic conclusion that identifies God with the totality of the world.

Rather than contribute further to the obscuring of Leone’s subtle and original contribution to Renaissance philosophy by offering an overview of *Dialoghi d’amore*, with its numerous, circuitous, subordinate discussions, lengthy and popularizing disquisitions on ancient mythology, astronomy and astrology, we shall focus on Leone’s nuanced resolution of the primary disagreement about the relations between desire and love, that is, the apparent contradiction between the claims that desire and love are the same, and that they are opposites. But we must note at the start that, while the first claim does not require a strict identity, the second assumes that any opposition implies a contradiction.

From the outset, Sophia’s resistance to Philo depends on interpreting the Platonic position as requiring a strict division between body and soul. It also requires real divisions within the soul, making its most noble part, the intellect, a distinct entity, absolutely independent of embodiment. Thus Sophia, the personification of wisdom, repeatedly and severely insists that if Philo truly loves her he should desire to satisfy that alone which pleases her, her mind.

There is a poignant irony here in the inverted relation between the male and female personae. Philo the male, or formal principle of the dialogue denies the real distinction between body and soul and, hence, between desire and love, whereas Sophia, the female, traditionally material, principle insists on such distinctions. But the resolution of the apparent contradiction between the Platonic and Aristotelian positions will establish body and soul, desire and love, male and female as natural correlatives: neither could be without the other. The same natural correlation is found between God and the world, as is mentioned briefly in passing, in the philosophical parts of the dialogues, thus protecting its radical theses from ‘vulgar’ view.

In the third, as in the preceding dialogues, Philo’s philosophical analyses (as distinct from the long poetic digressions) are distinctly Aristotelian in form, and often in content. Sophia’s objections continuously challenge the Aristotelian position with a Platonic one. In view of Philo’s pedagogic role in *Dialoghi d’amore*, the text can be read as the re-education of Sophia, whose Platonism reflects the dogma of Christian Neoplatonism. Thus, before he reverts to the definition of love, Leone makes amply evident his Aristotelian view of the soul:

> The soul is in itself one and indivisible, but by distributing its powers throughout the body and permeating even its surface and extremities, it branches out to certain activities pertaining to perception, movement and nutrition among various organs and divides itself among many diverse faculties.

This view of the relationship between body and soul will be reflected in Leone’s view of the relationship between the One and the many, that is, God and the world. Again, in the discussion of beauty as the form of the object which originates the motion of desire, Leone ‘corrects’ the Platonic doctrine of knowledge as recollection with an Aristotelian view of knowledge as arising from sensation. Forms do not exist independently of their corporeal manifestations but are embodied. They are abstracted by the intellect, which is initially mere receptivity to form. The ‘correction’ here is in fact a reconciliation of Plato with Aristotle, transposing the Platonic myth of anamnesis into a philosophical mode while simultaneously retrieving it from Christian interpretations:

> You must know, therefore, that all forms and Ideas do not spring from bodies into our souls, because to migrate from one subject to another is impossible; but their representation by the senses makes these same forms and essences to shine forth which before were latent in our soul. This enlightenment Aristotle calls the act of understanding and Plato memory, but their meaning is the same, although differently expressed.

Plato’s and Aristotle’s views are not only compatible, they are interdependent. Their harmony makes evident the insufficiency of either position taken in isolation. Plato’s teachings may be divinely inspired, but his mode of presentation lacks philosophical precision and so might lead to error. The resolution of the tension between Plato and Aristotle, fully and finally articulated in the third dialogue, makes clear that the opposition is only apparent and reflects terminological differences and a failure to recall that natural opposites belong to a single motion from potentiality to act.

Returning to the discussion of love in the third dialogue, Philo proceeds in an exemplary Aristotelian manner, pointing out that the questions ‘What is love?’ and ‘What is its first cause or origin?’ presuppose that love exists. He repeats the definition of love as desire and answers Sophia’s insistent objections that love and desire are not the

same, since we love what is and what we actually possess, but desire what we lack and what may or may not exist, by pointing out that reason demonstrates that love and desire are the same, although ‘in the vulgar tongue each has its own significance’. Sophia has taken a mere linguistic, conventional distinction for a real one, a confusion that Philo finds to be common among ‘certain modern theologians’.

Love and desire are different words denoting a single affection of the soul. Desire is a motion towards a desired object, love, a motion towards the beloved, which is the desired object. The cause of this motion, that is, the origin of love, is the desire for the pleasure of union with the beloved. Furthermore, *pace* Plato, love and desire are found in God. Indeed, God is their origin; in God they are found most eminently, as the desire for the perfection of all that is.

By following Aristotle closely, Philo demonstrates the proximity of truth and desire characteristic of true friendship, especially a friendship oriented by and toward God. Echoing the *Nicomachaean Ethics*, Philo explains to Sophia that he disagrees with Plato ‘because, as Aristotle, his disciple, said of him, although I am the friend of Plato, I am a greater friend of truth’. It is in virtue of their common love of truth that philosophers are friends and are also friends of God, a friendship, Philo insists, that cannot be one-sided.

To appreciate the radical force of the definition of love that emerges from the resolution of the tension between the Platonic and Aristotelian positions, we must turn to the cosmology underlying *Dialoghi d’amore*. According to Philo, the universe is a single individual composed of perfectly proportioned parts, all of which, whether eternal or perishable, are constitutive of its perfection. The perfection proper to each part, its final end or good, is the perfection of the whole, not just its own perfection. More precisely, since the final end or good is the perfection that all things by nature desire, all strive in proportion to their capacity to attain the perfection of the universe. Thus the natural perfection desired by each part is simultaneously a desire for perfect harmony or union with the whole. That is why love is the animating principle of the entire universe, of being as well as becoming. As the animating principle of the universe, love is its efficient cause; as the desire for the perfection of the whole, it is its final cause. The universe as a whole is a perfect circle, its final cause identical with its other causes, first, formal and efficient. Conceived as unified, this cause is identical with God.

The identity of God and the whole of nature cautiously alluded to in the philosophical parts of *Dialoghi d’amore* was later developed explicitly and in detail by Spinoza. Nowhere is the kinship between the two Portuguese Jewish thinkers more evident than in Leone’s description of the union between the human and divine intellects:

> The pure intellect which shines forth in us is likewise the copy of the pure divine intellect, and is stamped with the unity of all the Ideas; and this it is which, crowning the discourse of reason, reveals to us those ideal essences in intuitive, single, and abstract knowledge, when our well schooled reason merits such knowledge. So that with its eyes, we may behold in one intuition the highest beauty of the first intellect and of the divine ideas.

See also: Aristotle; Arisotelianism, Renaissance; Love; Plato; Platonism, Renaissance

IDIT DOBBS-WEINSTEIN

List of works

Abranavel, Judah ben Isaac (1501/2) *Dialoghi d’amore (Dialogues of Love)*, ed. S. Caramella, Bari: Laterza e figli, 1929; ed. C. Gebhardt, *Dialoghi d’Amore: hebraeische Gedichte*, Frankfurt am Main: Curis Societatis Spinozanae, 1929; trans. F. Friedberg-Seeley and J. H. Barnes, *The Philosophy of Love*, London: Soncino Press, 1937; trans. M. Dorman, *Sihot ‘al ha-Ahavah*, Jerusalem: Mossad Bialik, 1983.(Comprises three discussions of love as the animating principle of the universe. The three dialogues are between Philo, the poetic lover, whose beliefs about love can loosely be called Aristotelian, and Sophia, Philo’s beloved, whose views can be identified as Platonic. The entire exchange is an allegory of philosophy. Philo’s desire for Sophia is the philosopher’s quest for wisdom, which is human perfection. The three successive dialogical attempts to resolve an apparently fundamental disagreement between Plato and Aristotle are also attempts to delineate the relations between the human and the divine.)

References and further reading

Cooperman, D. (ed.) (1983) *Jewish Thought in the Sixteenth Century*, Cambridge, MA: Harvard University Press.

(Of particular importance are the papers by Herbert Davidson, Alfred Ivry and Shlomo Pines.)

Damiens, S. (1971) *Amour et Intellect chez Leon l'Hebreu*, Toulouse: Privat.(The only recent sustained philosophical study of Leone Ebreo.)

Martins, J.V. de Pina (1989) *Humanisme et Renaissance de l'Italie au Portugal: Les deux regards de Janus*, Lisbon and Paris: Fondation Calouste Gulbenkian, vol. 1, 357-81.(Devoted to the *Dialoghi*.)

Ruderman, D. (ed.) (1992) *Essential Papers on Jewish Culture in Renaissance and Baroque Italy*, New York: New York University Press.(Of special importance is 'The Place of the *Dialoghi d'amore* in Contemporaneous Jewish Thought', by A.M. Lesley.)

Ruderman, D. (1988) 'The Italian Renaissance and Jewish Thought', in A. Rabil (ed.) *Renaissance Humanism: Foundations, Form, and Legacy*, vol. 1, Philadelphia, PA: University of Pennsylvania Press, 382-433.(An essay on the current understanding of the impact of the Italian Renaissance movement on major Jewish thinkers of the time.)

Absolute, the

The expression 'the Absolute' stands for that (supposed) unconditioned reality which is either the spiritual ground of all being or the whole of things considered as a spiritual unity. This use derives especially from F.W.J. Schelling and G.W.F. Hegel, prefigured by J.G. Fichte's talk of an absolute self which lives its life through all finite persons. In English-language philosophy it is associated with the monistic idealism of such thinkers as F.H. Bradley and Josiah Royce, the first distinguishing the Absolute from God, the second identifying them.

1 Theories of the Absolute

With Fichte, Kant's transcendental ego (the relation of which to individual persons was somewhat vague) became a single absolute cosmic self which lives its life through each of us and posits the natural world as that through struggle with which it can develop itself morally (see Fichte, J.G. §3). With Schelling, the absolute self became the Absolute as the mysterious ground of all being, revealing itself in parallel both in *nature* (the creation of our shared unconscious) and in *finite thought*, of which it is somehow the identity-in-difference (see Schelling, F.W.J. §1). Later, Schelling held that the ordinary world of finite things springs from their somehow falling away from their proper eternal place within the Absolute in an arrogant attempt at independent existence. With Hegel, the Absolute became the Absolute Idea, somehow both the culmination and the self-differentiating unity of the cosmic dialectical sequence or circle (see Hegel, G.W.F. §6). This may be most conveniently conceived as starting with the concept of pure being, moving through a series of concepts towards their exhibition in empirical nature, and finally through a series of levels of human life, until it becomes self conscious as Absolute Spirit in the higher forms of this, especially in the civilized nation state and in the philosophic mind. Whether Hegel also ascribes to it a more cosmic self-consciousness is controversial - certainly it is somehow a self-differentiating spiritual unity.

More conspicuous in the English-speaking philosophical world were the conceptions of the Absolute of F.H. Bradley, Josiah Royce, J.M.E. McTaggart and other absolute idealists in a tradition not yet dead (and still representing, arguably, our best grasp of how things really are). For Bradley and Royce, the Absolute was a single spiritual individual of which all finite things are, if not exactly parts, at least part-like aspects. For McTaggart, in contrast, it was the Universe conceived as a system of selves linked directly or indirectly by bonds of affection and appearing to each other much of the (apparent) time as the physical world.

British and US absolutists tended to treat time more bluntly as an illusion than did their German predecessors mentioned above. For the former, the Absolute was a *Nunc Stans* containing all finite experiences which, although they appear to themselves to be in time, are in truth its eternal components. For the German absolutists (at least for Fichte and Schelling) free will, fairly much as it is commonly conceived, was among the most stressed features of the Absolute, and derivatively of humans. Indeed, their chief quarrel with Spinoza, whose monism they were trying to capture in more Kantian terms, was his determinism. The British and US absolutists were more Spinozistic on this matter, believing frankly in a 'block universe'.

2 Bradley and Royce

Arguably the best case for, and the best characterization of, the Absolute was that of Bradley and Royce. For Bradley, the Absolute was 'a single Experience, superior to relations and containing in the fullest sense everything which is' (Bradley 1914: 246). For Royce it was a universal self whose life is composed of the lives of all conscious beings (there being no others) experienced in unity.

Their argument was somewhat as follows. First comes the well-argued idealist claim that there is no such thing as unexperienced reality - in fact there is nothing other than lived experience itself. Now, on the face of it, experiences occur as aspects or components of finite centres of experience, such as we think of as the consciousness of men and animals, so it seems that the world must be the totality of these. But how are they in any real relation to one another rather than isolated monads, each altogether wrapped up in itself? The common-sense view is that they are variously juxtaposed in a single space and time, just as are the brains which underpin them. But for idealism, space and time and all their contents exist only as presentations within, or constructions wrought within, centres of experience so that it is these that contain them rather than vice versa. There must, then, be some kind of whole other than that of space and time within which finite centres of experience exist, and this can only be

some more comprehensive, indeed 'infinite', centre of experience within which they all occur together in a mutually necessitating manner. This will be timeless, yet will contain each centre of experience in every one of its temporal phases, and it will be a state of perfect understanding of, and satisfaction with, itself and all that it contains (for ignorance and dissatisfaction can have no place when nothing from outside can block knowledge or desire).

Royce added his own special twist to the argument by contending that there could not be a relation of *aboutness* between one's thoughts and external objects (and that therefore one's thoughts could only be about the present contents of one's own mind) unless the thoughts and the external objects were all contents of one all-embracing mind which used the former as a correct or incorrect way of characterizing the latter.

One problem with the Absolute, as conceived by Bradley and Royce, was a special version of the problem of evil: How can the whole be perfect when so many of its parts or aspects are so bad? This objection was presented in its most lively form by William James. Royce gave the most elaborate solution. The highest goods consist in the overcoming of evils, so that what is mere evil from a finite point of view is, from the Absolute's point of view, that which is eternally overcome by the good which keeps it down. The true image of the Absolute is Saint George, with the dragon struggling perpetually beneath him, the whole thereby possessing a perfection much greater than the merely saccharine goodness of a world without evil (see Evil, problem of §1).

Many of these absolutists, though decidedly not all, saw the Absolute as the Christian God properly conceived; however, in some respects the Absolute of Western absolute idealism is much closer to Brahman as conceived by Advaita Vedānta, though a more active Brahman than the West wrongly supposes to be the Hindu one (see Vedānta).

See also: Idealism; Kant, I.

T.L.S. SPRIGGE

References and further reading

Bradley, F.H. (1897) *Appearance and Reality*, Oxford: Oxford University Press, 1930. (Main statement of Bradley's metaphysics, one of the most important original works of metaphysics written in English.)

Bradley, F.H. (1914) *Essays on Truth and Reality*, Oxford: Clarendon.(Major essays containing developments of Bradley's logic and metaphysics. Source of definition cited in §2 above.)

Fichte, J.G. (1794) *The Science of Knowledge*, trans.P. Heath and J. Lachs with the 1797 introduction, Cambridge: Cambridge University Press, 1982.(This is the main statement of Fichte's extremely difficult philosophy.)

Findlay, J.N. (1970) *Ascent to the Absolute*, London: George Allen & Unwin.(Argues for a revival of 'absolute theory', that is, the study of the unconditioned basis of all things.)

Hegel, G.W.F. (1817) *The Encylopedia of the Philosophical Sciences*, translated in volumes as: Vol. 1 *The Logic of Hegel*, trans. W. Wallace, Oxford: Oxford University Press, 1974; Vol. 2 *Hegel's Philosophy of Nature*, trans. M.J. Petry, Oxford: Oxford University Press, 1970; Vol. 3 *Hegel's Philosophy of Mind*, trans. W. Wallace and A.V. Miller, Oxford: Oxford University Press, 1971.(Vols 1 and 3 are those of most interest to the modern reader. These works give a summary statement of Hegel's view of the world as the gradual growth to self-knowledge, through the human mind, of the Absolute Idea as infinite spirit.)

McTaggart, J.M.E. (1921, 1927) *The Nature of Existence*, Cambridge: Cambridge University Press, 2 vols.(One of the most closely argued works of constructive metaphysics ever written, though its conclusions are somewhat fantastic.)

Royce, J. (1885) *The Religious Aspect of Philosophy*, Gloucester, MA: Peter Smith, 1965. (The first and perhaps finest of Royce's philosophical works. As well as arguing for the existence of the Absolute as an all-including consciousness, it is outstanding for its treatment of how *ought* relates to *is*.)

Schelling, F.W.J. (1797) *System of Transcendental Idealism*, trans. P. Heath, Charlottesville, VA: University Press of Virginia, 1978.(Schelling changes his philosophical position several times. This is the main statement of it in its earliest form.)

Sprigge, T.L.S. (1983) *The Vindication of Absolute Idealism*, Edinburgh: Edinburgh University Press.(This argues first for the impossibility of an unexperienced reality, deducing therefrom a panpsychic view of nature and the existence of an absolute, all-containing consciousness.)

Absolutism

The term 'absolutism' describes a form of government in which the authority of the ruler is subject to no theoretical or legal constraints. In the language of Roman law - which played a central role in all theories of absolutism - the ruler was legibus solutus, *or 'unfettered legislator'. Absolutism is generally, although not exclusively, used to describe the European monarchies, and in particular those of France, Spain, Russia and Prussia, between the middle of the sixteenth century and the end of the eighteenth. But some form of absolutism existed in nearly every European state until the late eighteenth century. There have also been recognizable forms of absolute rule in both China and Japan.*

As a theory absolutism emerged in Europe, and in particular in France, in the late sixteenth and early seventeenth centuries, in response to the long Civil Wars between the Crown and the nobility known as the Wars of Religion. In the late eighteenth century, as the reform movement associated with the Enlightenment began to influence most European rulers, a form of so-called 'enlightened absolutism' (or sometimes 'enlightened despotism') emerged. In this the absolute authority of the ruler was directed not towards enhancing the power of the state, but was employed instead for advancing the welfare of his subjects.

1 The legal definition

Like most such terms, absolutism is a nineteenth-century coinage. But the term 'royal power absolute' was employed, in one context or another, by nearly all the monarchs of Europe from the late fifteenth to the mid-eighteenth century.

Although it was sustained by, and has become associated with, a number of social and cultural developments - lavish courts, an extensive bureaucracy, standing armies and the creation of a new noble class - absolutism was primarily a theory of legislative authority. This maintained that all rulers possessed exclusive executive and legislative authority, and consequently that the laws they made constituted an expression of their will. In the formula used by the jurist Ulpian, and repeated in one form or another by all would-be absolute monarchs, 'that which pleases the prince has the force of law'. All theories of absolutism claimed, in effect, that modern kings possessed the same authority - frequently described by the term *imperium* - as that once exercised by the Roman emperors. To sustain this claim common law, based as it was upon the legislative will of the community, was gradually replaced throughout Europe during the sixteenth and seventeenth centuries by Roman law. (The sole major exception was England, the ambition of whose monarchs was checked by the execution of Charles I in 1649.) The only theoretical limit to the monarch's legislative authority was the Divine and Natural Laws, which, since there was no one other than God to execute them, provided only the flimsiest of constraints. The kings' positive laws might be interpreted within certain narrowly defined limits (although all absolute rulers from Justinian to Napoleon attempted by codification to eliminate interpretation), but they were mandatory (see Roman law).

Absolutism was thus a radical attempt by increasingly centralized, increasingly modern, states to overturn the broadly contractual theories of authority which had grown up after the collapse of the Roman Empire. These had maintained that the king derived his authority from a contract with the people, who had been granted power directly by God himself (see Contractarianism). The laws were made by the representative assemblies of the people and administered by the king. He was the *servant* of his people, and his role was one of magistrate not judge. And although in practice the people had few means by which they might rid themselves of a ruler, there did exist quite powerful safeguards for their collective rights. Absolutism, in contrast, denied the representative assemblies any legislative or executive power. In Spain, the Cortes had been reduced to little more than a tax-voting body by the mid-sixteenth century. In France, the Estates General fell into disuse after 1612, and in 1673 the Parlement of Paris, a body which modelled itself on the Roman Senate, was deprived of even its hallowed right to remonstrate against royal edicts before registering them.

2 Two theories of absolutism

Despite its essentially legal core, absolutism was by no means a single or unitary theory. In its most extreme form, it maintained that the power of the ruler derived directly from God, as the source of all created things. This, known

as Divine Right, found its most powerful exponent in Jacques-Bénigne Bossuet. Bossuet claimed (an argument he derived from Hobbes) that as the people had had no historical sovereign before the institution of kingship, kings must have been the creation of God, and were, therefore, accountable to him alone. Most theories, however, including Hobbes's own, were dependent upon some version of an original contract theory (see Hobbes, T. §7). These fell into two broad categories.

The first maintained that royal authority derived, as the contractualists had maintained, from an agreement between the ruler and his people, but that since this was a contract, it involved an irreversible transference of power. The people might be entitled under certain conditions to resist their rulers, but they no longer possessed the authority to replace him. Perhaps the most persuasive, and certainly the earliest exponent of these claims was Jean Bodin (§3) whose *Les six livres de la république (The Six Books of a Commonwealth)* first appeared in 1576. For Bodin the state was the 'lawful government of many families'. The ruler held the same position *vis-à-vis* society as the father did towards the members of his family. This granted him absolute and undisputed power over them. (Bodin even wished to see the right granted to fathers by Roman law to execute their children reintroduced into France.) Sovereignty, which was to become the key term in all subsequent discussions of royal authority, could thus be defined as 'the giving of laws to their subjects in general *without their consent*'. Legal authority and the state itself thus became an expression of the king's legislative will (see Sovereignty).

The second theory, whose most powerful exponent was the Spanish Jesuit theologian and jurist Francisco Suárez (§4), maintained that the authority of all rulers derived from a delegation of power from the people (Suárez 1612). Although, like Divine Right, this granted the ruler absolute legislative and executive authority, it meant that the people could, *in extremis* (or if the throne fell vacant), reclaim that authority. For this reason Suárez, like most Jesuits, accepted that tyrannicide might be an acceptable means of disposing of an unjust ruler, an opinion which led to the public burning of his writings in both London and Paris.

On both accounts, the people retained certain rights as individuals; crucially, that of self-defence and (for Bodin) the right of property. Having surrendered, in Hobbes' formulation, both their will and their judgment to their rulers, they were wholly subject to them, but they remained free agents. Furthermore, although his person and the state became a single entity, the ruler did not have property rights in the state and could not, therefore, alienate any part of it. There therefore existed a distinction, albeit one which subsequent legal theorists such as Montesquieu would reject as illusory, between 'absolutism' and 'despotism'. As even Bossuet was prepared to concede, a king's rule should be absolute, but never (as the Ottoman Sultan's was thought to be) arbitrary. In practical terms this meant that the king should, wherever possible, consult the representatives of his subjects, but he was not obliged to follow their advice. The king should also abide by his own laws, and respect local custom. But all absolutists allowed that such constraints could be set aside in cases of necessity. The concept of 'necessity' as a device whereby the rule of law might be suspended - closely associated as it was with theories of 'reason of state' - thus became a key term in the vocabulary of absolutism.

3 Absolutism and the modern state

For all its archaic trappings, absolutism was essentially a modern theory of state authority. Writers like Bodin, Hobbes, Suárez and Bossuet possessed a recognizably modern conception of legislative authority as the means to protect the interests of the entire society. All, in different idioms, argued that unless such authority was clearly vested in one source, no society could ultimately escape internal conflict of the kind which had afflicted France in the sixteenth and England in the seventeenth century.

Numerous historians have argued that, despite the vast body of theoretical literature endorsing absolute rule, the power of early-modern monarchs was never in fact complete. But no state power ever is. The theory of absolute sovereignty was about *de jure* not *de facto* power. By the end of the seventeenth century, moreover, the rulers of Spain, France, Sweden, Russia and large parts of Germany had gained effective control over all the armed forces and communications within their territories. Their subjects may have defied them, but they never consistently challenged their right to rule.

Absolutism sought to establish in this way the undisputed power of the state over the numerous factional and regional interests which had governed medieval societies. In the sixteenth and seventeenth centuries this power was embodied in kings. But, as Hobbes said, it did not really matter if the ruler was one or many. The final

beneficiaries of this insight were the French Revolutionaries. Far from being the end of absolutism, the Revolution was its fulfilment. For it was the Revolutionaries who, by replacing the monarch with an assembly, assured the final transformation of French society into a modern - and absolute - parliamentary state. By the end of the nineteenth century all the major states in Europe, with the exception of Britain and Russia, had followed their example.

See also: Filmer, Sir Robert

ANTHONY PAGDEN

References and further reading

Anderson, P. (1974) *Lineages of the Absolutist State*, London: Verso.(A wide-ranging (England to Japan) and highly intelligent, if now somewhat dated, Marxist interpretation.)

Bodin, J. (1576) *Les six livres de la république*, Paris: Jacques du Puy, 1583; repr. Aalen: Scientia, 1961; trans. J. Bodin, *De republica libri sex*, Paris: Jacques du Puy, 1586; trans. R. Knolles, *The Six Books of a Commonwealth*, London: B. Bishop, 1606; repr. ed. K.D. McRae, Cambridge, MA: Harvard University Press, 1962; ed. and partial trans. J.H. Franklin, as *Bodin: On Sovereignty*, Cambridge: Cambridge University Press, 1992.

Bossuet, J.-B. (1709) *Politique tirée des propres paroles de l'Écriture-Sainte*, ed. J. Le Brun, Geneva, 1967; ed. and trans. P. Riley, *Politics drawn from the very words of Holy Scripture*, Cambridge: Cambridge University Press, 1991.

Franklin, J.H. (1969) *Constitutionalism and Resistance in the Sixteenth Century*, Cambridge: Cambridge University Press.(Provides the background for the early development of absolutism.)

Franklin, J.H. (1973) *Jean Bodin and the Rise of Absolutist Theory*, Cambridge: Cambridge University Press. (The best detailed study of Bodin's political writings.)

Fulbrook, M. (1983) *Piety and Politics*, Cambridge: Cambridge University Press.(Mostly concerned with England and Prussia.)

Giesey, R.E. (1961) *The Juristic Basis of Dynastic Right to the French Throne*, Philadelphia, PA.(A detailed account of the legal arguments for sovereignty.)

Gilmore, M. (1941) *Argument from Roman Law in Political Thought 1200-1600*, Cambridge: Cambridge University Press.(Still the only comprehensive study of the role of Roman law in the formation of the absolutist state.)

Harth, E. (1983) *Ideology and Culture in Seventeenth-Century France*, Ithaca, NY: Cornell University Press. (Studies the uses made of culture by the French monarchy to enforce the power of the state.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.

Keohane, N. (1980) *Philosophy and the State in France from the Renaissance to the Enlightenment*, Princeton, NJ: Princeton University Press.(An outstanding account of the transformation of political thinking in this period.)

MacFarlane, L.J. (1977) 'Absolutism, Tyranny and the Minimum Conditions for Constitutional Rule', *Government and Opposition* XII: 212-33.

Skinner, Q. (1978) *The Foundations of Modern Political Thought*, vol. 2, *The Age of Reformation*, Cambridge: Cambridge University Press.(Provides the most succinct account of the political theory of absolutism.)

Sommerville, J.P. (1991) 'Absolutism and Royalism', in *The Cambridge History of Political Thought 1450-1700*, eds.J.H. Burns and M. Goldie, Cambridge: Cambridge University Press, 347-73.(A useful general survey of the literature.)

Suárez, F. (1612) *Tractatus de legibus ac deo legislatore (On Laws)*, in *Corpus Hispanorum de Pace*, ed. L. Perena, V. Abril and P. Suner, Madrid: Consejo Superior de Investigaciones Cientificas, 1971-81, vols 11-17, 21-2.

Abstract objects

The central philosophical question about abstract objects is: Are there any? An affirmative answer - given by Platonists or Realists - draws support from the fact that while much of our talk and thought concerns concrete (roughly, spatiotemporally extended) objects, significant parts of it appear to be about objects which lie outside space and time, and are therefore incapable of figuring in causal relationships. The suggestion that there really are such further non-spatial, atemporal and acausal objects as numbers and sets often strikes Nominalist opponents as contrary to common sense. But precisely because our apparent talk and thought of abstracta encompasses much - including virtually the whole of mathematics - that seems indispensable to our best attempts to make scientific sense of the world, it cannot be simply dismissed as confused gibberish. For this reason Nominalists have commonly adopted a programme of reductive paraphrase, aimed at eliminating all apparent reference to and quantification over abstract objects. In spite of impressively ingenious efforts, the programme appears to run into insuperable obstacles.

The simplicity of our initial question is deceptive. Understanding and progress are unlikely without further clarification of the relations between ontological questions and questions about the logical analysis of language, and of the key distinction between abstract and concrete objects. There are both affinities and, more importantly, contrasts between traditional approaches to ontological questions and more recent discussions shaped by ground-breaking work in the philosophy of language initiated by Frege. The importance of Frege's work lies principally in two insights: first, that questions about what kinds of entity there are cannot sensibly be tackled independently of the logical analysis of language; and second, that the question whether or not certain expressions should be taken to have reference cannot properly be separated from the question whether complete sentences in which those expressions occur are true or false.

1 Logical and ontological categories

Although what most obviously needs explaining is the abstract-concrete distinction, the relevant notion of *object* also calls for elucidation. There is a familiar, everyday use of the term 'object' in which we may speak of the objects found in the accused's pockets, for example. There are probably no very precise rules governing this use, but it seems clear that being extended in space and time is at least a necessary - but probably not a sufficient - condition for its application. If 'object' is so understood, the term 'abstract object' is straightforwardly self-contradictory. We should infer, not that Nominalism wins by default, but that some other more general, less restrictive notion of object is in play in philosophical discussions. But if so, how should it be characterized? To avoid begging questions, it might be proposed that anything should be reckoned an object to which we may make reference. Arguably, however, this goes too far the other way - we may as well be said to refer to fiddling as to Nero, when we assert that Nero fiddled, but should be loathe to count fiddling an *object*.

A way forward which preserves this general approach is to take objects to be the referents of expressions of a certain restricted class - what are usually called 'singular terms'. To take this step is to follow Frege in viewing the ontological categorization of entities as dependent upon a prior logical categorization of expressions. Objects, properties and relations, for example, are essentially the non-linguistic correlates of, respectively, singular terms (for example 'Nero', 'this lake', 'the dome of St. Peter's' and so on), one-place predicates ('… fiddles', '… is deep'), and two- or more-place predicates ('… loves… ', '… is taller than… '). An object, on this account, is the referent of an actual or possible singular term.

When *object* is so understood, our opening question is, in an important sense, a distinctively *modern* one. It is not that we can discern no significant common concerns underlying ancient disagreements over the status of Plato's Forms and the great medieval battle between realists and nominalists over the existence of universals on the one hand, and modern ontological disputes on the other. Traditional and modern discussions share a general concern with the relations between language and the world. At bottom, disagreement over abstract entities is disagreement over whether an adequate account of language-world relations can be provided without reference to any such entities. It remains the case that a fundamental shift has taken place in the way very many philosophers conceive and argue about ontological issues in general, and issues about abstract entities especially. Ancient and medieval disputes focused on the existence of universals as opposed to particulars, with the former thought of as abstract

entities which both predicates ('is red', 'is wise') and corresponding abstract nouns ('redness', 'wisdom') stand for. But on the Fregean approach, it makes no sense to suppose some one kind of thing to be the common referent of expressions of completely different logical types. This need not mean that there is no significant disagreement between medieval realists and nominalists; but it does mean that they misconceived the issue, or at least ran together questions we should separate. For it is one question whether abstract nouns are to be conceived as genuine singular terms, standing for objects, and a quite distinct question whether the corresponding predicates have reference - if so, then they stand, not for objects, but for properties (concepts, in Frege's sense) (see Universals).

2 The abstract-concrete distinction

Abstract objects can be neither seen nor heard, nor can they be tasted, felt or smelled. But for several reasons it would be unsatisfactory to take inaccessibility to sense-perception as the basis of our distinction. Besides importing an unwanted relativity to human sensory faculties, it would fail to draw the distinction clearly, there being room for dispute over what should count as perceiving something. If the range of sense-perception is taken as including only what can be discerned with the naked organ, as it were, the condition for being concrete is clearly too restrictive. The range might be extended to allow for detection via more or less remote effects, but once the criterion is loosened in this way, the proposal slides into taking capacity for involvement in causal interactions as the mark of the concrete. This suggestion avoids the difficulties with a sensory-access criterion but, even if extensionally correct, does not go to the heart of the matter. We expect capacities in general to have some categorical basis. Why are concrete objects capable of causal interaction but abstract objects not? The answer, it would seem, should yield a more illuminating account of the distinction. Partly for this reason, a more promising account of the distinction sees lack of location in space or time as distinctive of the abstract - what cannot be anywhere, anywhen, cannot be a factor in the causal nexus.

Although it is widely endorsed and gives intuitively correct results in the cases to which philosophers have attended, this account is nevertheless flawed. This is because there are candidates for abstract status which, though plainly lacking spatial properties, are not wholly atemporal. In the sense in which two pairs of players at different chessboards may be said to be engaged in one and the same game, the game of chess is plausibly taken to be an abstract object; but while it is not located anywhere, it has not always existed, but was devised at a certain time. Other examples are natural languages, many if not all works of art, and words and letters in the type- as opposed to token-sense (roughly, the sense in which there are just six, not eight, distinct letters in the word 'abstract') (see Type/token distinction). Thus while the abstract-concrete distinction undoubtedly has much to do with spatiality and temporality, it does not seem straightforwardly identifiable with the distinction between what has spatial or temporal position and what has neither. An alternative proposal of considerable interest is that concrete objects are those which are, in principle, capable of being picked out ostensively, while abstract objects are those to which we can refer only by means of some functional expression (Dummett 1973: ch. 14). Thus we may pick out a particular tree by the words 'That beech', perhaps accompanying our utterance with a pointing gesture; but we cannot, for example, literally point to a certain shape or number - rather, we must refer to them as, say, the shape of such and such a vase or the number of eggs in the carton (Noonan 1976; Hale 1987: ch. 3).

3 Grounds for belief in abstract objects

Many philosophers, appealing to Ockham's Razor - the principle that entities should not be multiplied beyond necessity - deem it mortally sinful to believe in abstract objects unless such belief is unavoidable, but disagree about whether it is actually avoidable. Orthodox nominalists hope to avoid it by carrying through a programme of reductive paraphrase. However, in view of the resistance of various kinds of apparent reference to/quantification over abstract objects to elimination by reductive paraphrase or re-interpretation in concrete terms, this does not appear feasible as a completely general means of escaping commitment to abstract objects (see Ontological commitment). This has led some philosophers to conclude that reference to and quantification over domains including abstract objects is indispensable to a fully adequate account of the world. There is a strong appearance that this is the case with reference to mathematical entities - numbers of various kinds, functions and more generally, sets. On the face of it, the natural sciences, and physics especially, require substantial use of arithmetic and analysis, and the latter in turn draws fairly heavily on set theory. This argument - known as the Quine-Putnam Indispensability Argument - provides, if accepted, a strong *indirect* reason for believing in numbers and sets at least: scientific theories require acceptance of mathematical theories, so that whatever reasons we have to believe

that our best scientific theories are true is reason to accept mathematical theories, and so to believe in the abstract objects of which they speak.

This argument has been vigorously contested, particularly by Field (1980), who argues - in support of a new and highly unorthodox brand of nominalism - that there is, contrary to appearances, no need for mathematical theories to be true for their use in science to be justified. It is enough that such theories should have a certain strong kind of consistency property, which he calls ‘conservativeness’. Since a nominalist can accept mathematical theories as having this property without believing them to be true, they have no need to engage in any kind of reductive translation programme of the sort previously mentioned - they can simply use mathematical theories while denying that they are literally true, thereby avoiding commitment to the abstract objects their truth requires. Among the difficulties confronting this approach, one important assumption Field makes is worth highlighting. Field takes the Quine-Putnam argument to offer the only ground worth taking seriously for holding mathematical theories to be true, so that if he is able to undermine it, there remains no pressure to take on the ontological commitments they import. If Field’s assessment were correct, the best grounds we could have for believing maths and so on, would be indirect and a posteriori. But this assessment rests upon the challengeable assumption that the only statements we may justifiably accept on other-than-indirect a posteriori grounds are those directly ascertainable as true by observation. Perhaps we should take seriously, as he does not, the possibility that belief in the truth of mathematical statements and acceptance of their ontology may be warranted a priori.

4 Grounds for disbelief

Unquestionably the most important arguments against abstract objects are epistemological. One is that - in view of the presumed causal inertia of abstract objects - to construe statements of some given kind as having their truth-conditions constituted by states of affairs essentially involving such objects, puts those statements irretrievably beyond the reach of our knowledge. Crudely, if mathematical statements have Platonistic truth-conditions, we could not possibly know them to be true; since we do have mathematical knowledge, Platonism is false. In its simplest and earliest versions, this argument relies upon a very exacting form of causal theory of knowledge, which takes it to be an invariably necessary condition for a thinker *X* to know that *p*, that *X*’s true belief that *p* should itself be caused by, or otherwise suitably causally related to, the fact that *p* (see Knowledge, causal theory of). A problem with this argument is that while such a strong condition (just how strong depends on how precisely the vague phrase ‘suitable causal relation’ is understood) may be satisfied in standard cases of perceptual and memory knowledge, it is very hard to see how it could be quite generally met, even when restricted in scope to ordinary empirical knowledge concerning perfectly concrete matters. Our inductively grounded belief that all aardvarks have bugs is, we may suppose, causally induced by inspection of a large and suitably varied contingent of bug-infested aardvarks - but there is no sort of causal relation, however complicated or attenuated, of which it may with any plausibility be claimed both that it holds between our general belief and the fact that all past, present or future aardvarks have bugs and that its holding is epistemically significant. If knowledge does not demand a suitable causal link in every case, the argument against Platonism collapses, at least in its present form.

A related argument alleges that no satisfactory sense can be made of the idea that we are capable of identifying reference to or thought about abstract objects. And once again, the argument in its simplest form rests upon an eminently challengeable assumption - in this case, that identifying reference or thought about a particular object *always* requires a suitable causal link between the speaker/thinker (or their utterance/thought) and the object in question (see Reference). Opponents of Platonism may hope to fashion more sophisticated causal analyses of knowledge and reference which are strong enough to sustain versions of these objections without being so strong as to be independently objectionable, but none has yet come forth.

A more powerful epistemological objection appeals to the thought that, even if knowledge is not to be analysed in specifically causal terms, we should expect to be able to provide a naturalistic explanation of our tendency to get things right significantly more often than not, in any area where we are disposed to credit ourselves with a capacity for knowledge (see Reliabilism). In the absence of causal or other natural relations between ourselves and abstract objects, it is hard to see how any such credible explanation might run for any region of discourse whose statements are supposed to carry Platonistic truth-conditions. The argument relies on the assumption that ontological views are tenable only to the extent that they leave space for a credible epistemology. The arguments reviewed here

confront Platonism with a strong challenge, even if they could not, by their very nature, tell decisively against it.

See also: Nominalism; Ontology; Realism and antirealism; Universals

BOB HALE

References and further reading

Dummett, M. (1973) *Frege: Philosophy of Language*, London: Duckworth.(Hard going, but the best introduction to Frege's approach to the analysis of language. Chapters 2, 4 and 14 are especially relevant.)

Field, H. (1980) *Science without Numbers*, Oxford: Blackwell.(Introduction and chapter 1 give a clear outline of the approach discussed in §3 above.)

Field, H. (1989) *Realism, Mathematics and Modality*, Oxford: Blackwell.(Occasionally technically difficult, but generally very readable. Chapters 1 and 2 provide an excellent account of Field's overall position; chapter 7.2 develops the epistemological argument against Platonism mentioned in §4 above.)

Frege, G. (1884) *The Foundations of Arithmetic*, trans. J.L. Austin, Oxford: Blackwell, 1953.(Essential reading for anyone wishing to pursue the present subject in depth. See especially §§60-8.)

Frege, G. (1892) 'On concept and object' in P. Geach and M. Black (eds) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 1970. (Especially relevant to §1 above. Provides a very readable informal elucidation of Frege's contrasted notions of object and concept, indicating their connection with his fundamental distinction, at the level of language, between proper names and predicates.)

Goodman, N. and Quine, W.V.O. (1947) 'Steps towards a constructive nominalism', *Journal of Symbolic Logic* 12: 105-22; repr. in N. Goodman *Problems and Projects*, Indianapolis, In and New York: Bobbs-Merrill, 1972, 173-98.(Classic nominalist manifesto. Technically quite difficult after opening sections.)

Hale, R. (1987) *Abstract Objects*, Oxford: Blackwell.(Along with Noonan (1976), chapter 3 develops a proposal to base the abstract-concrete distinction on the different kinds of identity criterion appropriate to the different kinds of object. Chapters 1-3 are relevant to §§1-2 above, chapter 5 to §3, and chapters 4, and 7 to §4.)

Hale, R. (1994) 'Is Platonism epistemologically bankrupt?', *Philosophical Review* 103 (2): 299-325.(Defends Platonism against arguments discussed in §4 above.)

Hale, R. and Wright, C. (1993) 'Nominalism & the Contingency of Abstract Objects' *Journal of Philosophy* 89 (3): 111-35.(Criticizes Field's version of nominalism.)

Lewis, D. (1987) *On the Plurality of Worlds*, Oxford: Blackwell.(Very readable. Chapter 2 discusses problems with the abstract-concrete distinction.)

Noonan, H. (1976) 'Dummett on abstract objects', *Analysis* 36 (2): 49-54. (Difficult but rewarding discussion, relevant to §2 above.)

Putnam, H. (1971) 'Philosophy of logic', in *Mathematics, Matter and Method: Philosophical Papers*, vol. 1, Cambridge: Cambridge University Press, 1979, 2nd edn. (Technically difficult in places, but contains important material relevant to §3 above, including presentation of indispensability arguments.)

Quine, W.V.O. (1948) 'On what there is', in *From a Logical Point of View*, New York: Harper Torchbooks, 1961, 1-19.(Classic paper on ontology, with strong nominalist sympathies.)

Wright, C. (1983) *Frege's Conception of Numbers as Objects*, Aberdeen: Aberdeen University Press.(Excellent exposition and vigorous defence of Frege's position. Chapters 1 and 2 are especially relevant.)

Academy

The Academy was a public gymnasium in northwest Athens. Plato taught there, and the Academy remained the centre of Platonic philosophizing until the first century BC*. Hence the term 'Academy' came to be used to designate Plato's school; members of the school were called 'Academics'. (And hence, ultimately, the modern use of the words to describe intellectual institutions and their members.)*

The word 'Academy' originally had a topographical reference. A mile and a half northwest of the Athenian agora, along the Ceramicus road, there was a public gymnasium and wrestling square set in a spacious park. Like most gymnasia, the Academy contained an exedra - a sort of open-air lecture theatre. Here Plato talked and taught philosophy. He set up a shrine to the Muses in the park; and he acquired a house, with a little garden, in the neighbourhood, where his friends and pupils congregated. A contemporary comic poet imagines a group of students assembled in the garden earnestly attempting to produce a definition of the pumpkin.

Plato's house was used by his successors until the time of Polemo (in the late fourth century BC), whose pupils lived in huts in the garden; and Platonists continued to teach in the Academy until the beginning of the first century BC. But after that time there seems to have been no special relationship between Platonism and the geographical Academy.

The word 'Academy' was readily transferred from the concrete to the abstract: it came to designate the school or institution which Plato established and which his successors conserved. The nature of the institution is imperfectly known; but it is clear that there was a head, or 'scholarch', who was (at least sometimes) elected to office; that there were senior and junior members; and that there were discussions, lectures and dinners. Yet the Academy was not an embryonic university: there were no degrees and no administration block.

On Plato's death in 347 BC, his nephew Speusippus led the school. He was followed by Xenocrates Polemo and Crates. In about 265 BC Arcesilaus assumed the scholarchate and turned Platonism down the sceptical path which it followed for almost two centuries. Of later scholarchs the most engaging and the most celebrated was Carneades. In 88 BC, when Athens was in the grip of war, the scholarch Philo of Larissa decamped to Rome. It seems likely that Philo was the last Platonist geographically connected to the Academy. But philosophy is above mere geography, and Platonism survived and flourished until the end of the ancient world. Modern authors will refer to later Platonists as Academics: the nomenclature is inaccurate, the inaccuracy venial.

Ancient writers, remarking upon apparent changes in the intellectual drift of the school, would speak of a plurality of Academies. The most generous listed five: the Old Academy, which lasted from Plato to Polemo; the Middle Academy, founded by Arcesilaus; the New Academy, inaugurated by Carneades; a fourth Academy under Philo; and a fifth under Antiochus. The Academics did not necessarily endorse these divisions. Thus Cicero, himself professing an Academic scepticism, simply distinguished between the Old Academy from Plato to Polemo and the New Academy from Arcesilaus onwards; and Philo notoriously maintained that there had only ever been one Academy.

See also: Platonism, Early and Middle §1

JONATHAN BARNES

References and further reading

Billot, F.-M. (1989) 'Académie', in R. Goulet (ed.) *Dictionnaire des philosophes antiques*, vol. 1, Paris: Éditions du CNRS.(Documents the archaeological and literary evidence.)

Cicero, M.T. (early 45 BC) *Academics*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press, and London: Heinemann, 1933.(A philosophical retrospect on the Academy by one of its best-known adherents.)

Glucker, J. (1978) *Antiochus and the Late Academy*, Hypomnemata 56, Göttingen: Vandenhoeck & Ruprecht. (Includes an exhaustive account of all non-philosophical aspects of the Academy.)

Philodemus (*c*.50 BC) *History of the Academy*, trans. T. Dorandi, *Filodemo: storia dei filosofi - Platone e l'Academia*, Naples: Bibliopolis, 1991.(A biographical account of the school's history; includes Greek text, and Italian translation and commentary.)

Action

Philosophical study of human action owes its importance to concerns of two sorts. There are concerns addressed in metaphysics and philosophy of mind about the status of reasoning beings who make their impact in the natural causal world, and concerns addressed in ethics and legal philosophy about human freedom and responsibility. 'Action theory' springs from concerns of both sorts; but in the first instance it attempts only to provide a detailed account that may help with answering the metaphysical questions.

Action theorists usually start by asking 'How are actions distinguished from other events?'. For there to be an action, a person has to do something. But the ordinary 'do something' does not capture just the actions, since we can say (for instance) that breathing is something that everyone does, although we don't think that breathing in the ordinary way is an action. It seems that purposiveness has to be introduced - that someone's intentionally *doing something is required.*

People often do the things they intentionally do by moving bits of their bodies. This has led to the idea that 'actions are bodily movements'. The force of the idea may be appreciated by thinking about what is involved in doing one thing by doing another. A man piloting a plane might have shut down the engines by depressing a lever, for example; and there is only one action here if the depressing of the lever was (identical with) the shutting down of the engines. It is when identities of this sort are accepted that an action may be seen as an event of a person's moving their body: the pilot's depressing of the lever was (also) his moving of his arm, because he depressed the lever by moving his arm.

But how do bodies' movings - such events now as his arm's moving - relate to actions? According to one traditional empiricist account, these are caused *by* volitions *when there are actions, and a volition and a body's moving are* alike *parts of the action. But there are many rival accounts of the causes and parts of actions and of movements. And volitional notions feature not only in a general account of the events surrounding actions, but also in accounts that aim to accommodate the* experience *that is characteristic of agency.*

1 Actions, events and individuation

'Action' and 'act' are often used interchangeably. But 'action' is given a definite meaning when actions are taken to be a species of events: it denotes particulars of a certain sort - concrete items in the spatiotemporal world. It is useful then to give a different meaning to 'act': acts are things people do (which are sometimes called act-types). Such things are not particulars, since one person can do the same thing as another. If Mary and John both voted for Smith, for instance, then Mary's act (voting for Smith) was the same as John's; but obviously the event that was Mary's voting for Smith was not the same action as John's voting for Smith.

Is Mary's voting for Smith ever the same action as *Mary's* doing some *other* act? This is a question now about the individuation of actions. Proponents of a fine-grained account (for example Goldman 1970) say that there are as many actions as there are acts exemplified by the agent on occasion, so that, in our pilot example, the man's depressing of the lever is one action and his shutting down of the engines is another action. Proponents of a coarse-grained account, on the other hand (for example Davidson 1971), assert the identity of his depressing of the lever with his closing down of the engine. They think that actions are often described in terms of their effects. When the description 'Paul's depressing of the lever' is used, an event is seen to have resulted in the lever's being depressed; when 'Paul's shutting down the engines' is used, an event is seen to have resulted in the engines' shutting down. But if Paul shut down the engines *by* depressing the lever, then the lever's coming to be down in its turn caused the engines' cutting out, so that in fact an action of Paul's is spoken of twice over here: it is described now in terms of one effect, now in terms of another.

This coarse-grained account can be made plausible by thinking of *Paul's* part in the engines' coming to be shut down as his moving of his arm. The thought is that for the engines to shut down, nothing was required of Paul after his arm had moved. Since he is the only relevant agent, Paul's moving his arm is the only action that there is, and 'his moving of his arm' is just one of its descriptions.

Various objections are made to the coarse-grained account, and these have led some philosophers (for example Ginet 1990) to intermediate accounts of actions' individuation. (An intermediate account might have it that Paul's

moving his arm is the same as his moving his arm against the lever (which would be denied by the fine-grained theorist), but is different from his shutting down of the engines (which would be denied by the coarse-grained theorist).) One objection to the coarse-grained account would say about Paul's case that his shutting down of the engines (*s*) took place later than his moving his arm (*m*), so that these have different properties: *m* occurred at *t*, *s* at *t* + *n*, so that $m \neq s$. A proponent of the coarse-grained account of course maintains that *s* really did occur at *t* if *m* did. They claim that our tendency to suppose otherwise is explained by our confusing s with the effect in terms of which we think of it. (The engines' shutting down is an effect which indeed occurred later than the action, they say.) A related objection relies on intuitions about when tensed sentences such as 'He has shut down the engines' are first true. Here the coarse-grained account's proponent allows the relevant intuition: this sentence is not true until some time later than 'He has moved his arm' is first true. But, they say, the truth of 'He has shut down the engines' requires more than the past occurrence of the event *s*: it also requires the past occurrence of the engines' shutting down.

The coarse-grained account joins with a definition of 'an action' as 'someone's intentionally doing something' to give intuitively right results. Suppose that Paul's shutting down of the engines was a terrible mistake, and that he was responsible for an accident. We understand how we can think of Paul as an agent in respect of something he did quite unintentionally, when we appreciate that his doing one thing was (the same as) his doing another. One thing he did was to shut down the engines, and he did this accidentally; but his doing this was his depressing of the lever, and depressing the lever was something that he did intentionally. For an event actually to have been an action of some person's according to the definition, it has to be true only that at least one of the things they did was something they intentionally did.

This criterion of actionhood is sometimes put by saying that an action has to be 'intentional under one of its descriptions'. 'Under a description' has wide philosophical currency. But it can be misleading. When 'doing things under descriptions' is employed, it is made to seem as if there were these things people do, and that *they* have various descriptions. Things people do do not have various descriptions, however: people's *actions* have various descriptions, and these descriptions correspond one:one with the things, or acts, that they then do.

2 'Basic acts'

Acts are done by doing other acts. But not *every* act someone does could be done by doing some different one, or nothing would ever get done. Among the things that a person does on occasion, then, there must be something which is *simply* done - not done by doing something else. This has been called the basic act. Where someone Φs by Ψ-ing, Ψ-ing is said to be *more basic than* Φ-ing; and the basic act is defined as the one than which no other was more basic.

Moving the body (that is, moving a bit of it in one or another way) is usually a basic act. When Mary raises her right arm directly - in order to vote at the meeting - raising the right arm is the basic act. But in the unusual case in which someone raises their right arm by lifting it with their left arm, raising the right arm, although a bodily act, is not the basic one. (What is basic here is moving the left arm.) Such an example shows that in order to say what was basic in a particular case, one has to know not only what bodily movements occurred in that case but also what was actually done by doing what else. Acts, then, are not basic *tout court*. Relative to Mary's action, but not relative to the action that was someone's raising their right arm 'indirectly', raising the right arm is basic.

The need to think about applications of basicness in connection with particular cases has sometimes given rise to talk of basic *actions*. But such talk conflicts with the coarse-grained account of actions' individuation. Where a person's raising their arm is considered to be identical to their voting, it cannot be supposed that their raising their arm is basic and their voting is not. If a notion of a basic act that is not relativized to particular cases is wanted, we have to think about what someone *can* do directly. (The person who used their left arm to raise their right arm might, or might not, have been *able* to simply raise their right arm.) Using a notion of a basic *ability*, we could speak of things as basic for a person with a particular repertoire of motions (not relative to any particular action now).

We encounter relations of dependence when we go through a list of more and more basic acts. Considering Paul's action, and going through his various acts - causing an accident, shutting off the engine, depressing the lever, moving the arm - it is natural to think of what is less basic as depending on what is more basic. We may think of

all the dependencies as causal ones in the particular example. But there are different kinds of dependence, and when the different kinds are distinguished, different relations of 'more basic than' can be distinguished. For example if we take it that a convention must obtain for someone's raising their arm to count as their voting, we could say about Mary's action that voting was *conventionally* more basic than raising the arm.

The thought that moving the body is basic seems now to be the thought that moving the body does not usually depend upon anything else - neither causally, nor in any other way. And yet physiologists tell us that, in fact, our bodily movements depend upon our muscles' contractions - that we move our bodies by contracting our muscles. It seems, then, that even where someone simply moved their arm we have a candidate for a more basic act than moving the arm - namely, contracting muscles. In fact, what this shows is that the perspective of an agent is ordinarily assumed in thinking about what is done; when moving the body is taken as basic, the focus is on things that the people might think of themselves as doing. A different notion of *basicness* is needed to accommodate the facts that the shift to a physiologists' perspective reminds us of. To allow for the fact that moving the body depends upon other things being done, a 'purely causal' notion of basicness may be introduced. This is not the intuitive, central notion that recapitulates the idea of what someone 'simply does' or 'does directly'. Philosophers have meant a variety of different things by 'more basic than'.

3 Volitional theories: actions, parts and causes

Events like muscles' contractions, which occur beneath the body's surface, come to notice not only in our thinking about different ideas of basicness: they may be prominent also when we enquire what precise causal story should be told about any action. And it is not only physiological thinking which makes philosophers want a precise causal story: a definition of 'an action' as 'someone's doing something intentionally' belongs with a view which distinguishes actions from other events by reference to a particular sort of psychologically specifiable causal history. On this view, a person who does something intentionally does the thing because they have a reason to. Saying what their reason was requires knowing what their relevant beliefs and desires were (see Belief; Desire; Intention); and it provides a distinctive kind of explanation of why they did the thing (see Reasons and causes). But it may be asked whether there is not a more immediate causal story to be told about an action than that which shows up in a reason explanation. Do actions have immediate mental antecedents of a certain sort?

It has sometimes seemed that actions must have such antecedents, because wanting, believing and intending all seem inadequate to explain actually doing something. Suppose you want to move your arm. Your arm doesn't move until … what? 'Until there is a volition' was an answer often given in the eighteenth and nineteenth centuries: philosophers often posited volitions, or acts of will (as they are alternatively called), as events which initiate the causal process of acting, bridging the gap between wanting and doing. This was a gap between the mind and the body in the thinking of dualists (see Dualism; Will, the).

Volitions fell out of philosophical favour when Ryle objected to them as spurious (Ryle 1949). Ryle asked why 'the Will' has to be exercised in action at all, thinking that the postulation of volitions was a hangover from the idea of a 'ghost in the machine'. One of his arguments against volitions posed a dilemma: either volitions are themselves 'active', or they are not. If they are 'active', then if a volition really were required for a genuine action, we should always have to posit a new volition as cause of any volition, and we should be led to an infinite regress. If, on the other hand, volitions are not themselves 'active', but are mere causes of actions, it is hard to see why anyone should think that their introduction helps in recording what is special to action.

Some volitional theories rather obviously escape this objection. John Stuart Mill, for instance, thought that an action was 'a series of two things; the state of mind called a volition followed by an effect' (Mill 1843: I 3.5). In its twentieth-century guise, the Millean theory takes an action to be composed from (a) a volition (b) a movement of a bit of the body of the person whose volition it is. On this theory, Ryle's question as to whether a volition itself, or only its effect, is 'active' has no simple answer, since each of these things is a *part* of an action. But Ryle's underlying question may still be pressed: 'Why posit a sort of mental item such that actions are present only when an item of that sort is a cause?'

The account of volitions as parts of actions draws attention to the distinction between *actions*, each one of which is someone's moving a bit of their body, and *bodies' movements*, each one of which is a bit of someone's body's moving. (The thesis which is often used to summarize the coarse-grained view of actions' individuation - that

actions are bodily movements - is now seen to be crucially ambiguous at best.) When this distinction is made, there are two other views about bodies' movements, both different from the Millean, componential view. (A) Actions are identical with bodies' movements, so that, for instance, a person's raising their arm is their arm's rising. (B) Bodies' movements are not even parts of actions, so that a person's arm's rising is wholly distinct from their raising it. (A) is implausible inasmuch as it seems to sever the connection between acting and doing something; unless a person's arm's rising is itself the person's doing something, that connection is broken when movements are identified with actions. (B) is a more plausible view - at least for the philosopher who think that actions are described in terms of their effects; for the latter, a person's arm's going up can be the effect of their raising it, just as a flag's going up is the effect of someone's raising the flag.

According as (A) or (B) is accepted, the doctrine that volitions cause bodies' movings turns out differently. When bodies' movings are thought to be actions, volitions are conceived in the manner Ryle found objectionable - they are thought of as the last item in a mental causal chain leading outward to something physical. But when bodies' movings are thought to be *no parts* of actions, the theorist can say that a volition causes a body's movement and is itself an action. When that is said, an item is recognized the status of which is ineluctably *psychophysical*, being both a volition and an action; the theorist may refuse any picture in which the mental can be marked off from the physical on a causal chain. It remains a good question why one should suppose that there is a faculty of the Will the products of which, volitions, have to be brought into an account. But when volitions are identified with actions, we can be certain at least that there is nothing mythical or 'ghostly' about them.

The claim that physical actions are redescribable in recognizably psychological terms is made not only by philosophers who say that actions are volitions, but also by others who have no truck with volitions. Some philosophers argue that anyone who does something intentionally *tries to* do it. (They allow that one need not *think of* oneself as trying to do the things one does intentionally, and they allow that 'They tried to do it' is not usually a natural thing to say about someone who encountered no difficulties and who did not need to make any special effort.) If that is correct, then, given a coarse-grained view of individuation, each action *is* someone's trying to do something. One may arrive at an account in which a person's having a reason to do something leads to their trying to do it; when their trying to do it has the effects they want (as usually it does), it is their doing the thing. To the question 'Your arm doesn't move until … what?', the answer now could just as well be 'Until you move your arm' or 'Until you try to move it' (see Mental causation).

4 Agents' experience and knowledge

When the Will features in accounts of action, it may be thought of, in Cartesian spirit, as something the operations of which are available only to introspection (see Introspection, psychology of; Private states and language). One of the ideas to which Ryle was objecting in his attack on volitions was the idea that for each visible action of a person there is an inner item accessible to them alone. Now, although we may not wish to describe our experience of agency in Cartesian terms, it seems undeniable that there is consciousness of voluntary agency. If volition is thought of as action's conscious aspect, then it is not an invention of philosophers but a feature of everyone's experience. So it could be correct to suppose that volition is a part of the phenomenon of action, even if it should be denied that *each* action has *a* volition as a part or as a cause.

A person who is doing something intentionally knows what they are doing - or, if they do not know this, they know at least what they are trying to do; and they know this without making observations of themselves of a kind that others can make. The experience of acting, which is the basis of such knowledge, is not just proprioceptive experience (it is not just experience got from information fed back from the body about the body when the body is moved). So the idea of a distinctive conscious state contemporaneous with an action seems correct. There is little agreement about how such an idea should be recorded in an account. Sometimes the content of the experience of acting is spoken of in terms of 'exertion', which can make it seem as though some actual effort were always required to move a bit of the body. But if the idea of experienced exertion is meant only to capture the fact that it would feel very differently to us if we did not move our bodies voluntarily, it is acceptable. Suppose that you were wired up in such a way that your efferent neural pathways could be so stimulated that your muscles would contract and your finger move when some other person determined that this should happen. Of course there would not be an action of your moving your finger in that case; but also, we think, the characteristic experience of agency would be missing.

Such experience, it might be thought, is present in all conscious creatures who do things - whether or not they are rational agents who do things intentionally. If that is right, it may be necessary also to record another kind of experience, which is peculiarly human now, and which may be called the experience of freedom, or the sense of alternative possibilities (see Free will §5). Thinking of agents as conscious subjects can remind us of how narrowly focused the philosophy of action becomes when it is concerned exclusively with questions of actions, events and individuation, basic acts and volitional theories (see Consciousness).

5 Philosophy of action applied to ethics and law

There is another sort of narrowness in accounts of action that confine themselves to a conception of actions as a species of events. In marking out a class of actions, and investigating how these relate to volitions, movements, etc., the philosopher does not start to address many of the conceptual questions that are asked about action by the legal, or moral, philosopher. An ordinary concern with human action is not a concern with which events occur, but is a concern with what people do. (A fine-grained ontology, of ‘act tokens’ or ‘tropes’ is sometimes introduced in order to reflect this ordinary concern.)

Arguably ‘intentionally’ is the only piece of psychological vocabulary needed to characterize the actions, when actions are taken to be a species of events (see above). But resources beyond ‘intentionally’ are certainly required to make the many distinctions needed to understand lawyers’ accounts of *mens rea* and moral philosophers’ accounts of responsibility. In considering someone’s culpability, it is not enough to consider those of their attitudes which constitute their reason for doing what they do and which connect with what they do *intentionally*. One might need to know, for instance, whether some effect that was not intended by them as a means was or was not one which they *foresaw* as resulting from their choice. (Did Paul know what he was doing when he unintentionally shut down the engines?)

For a person to be praised or blamed as an agent there need not be any event which is an action of theirs. There are cases, for example, where someone intentionally does not do something (Jane intentionally did not answer the question, say), and where although ‘intentionally’ applies, its application is not to any *event* (there is no *event* which is Jane’s not answering the question). There are other cases where an agent is held responsible for some effect in the world, but where none of the things they did or did not do was something they *intentionally* did, or did not, do. Legal conceptions of recklessness, negligence, or strict liability would all introduce examples in this category (see Moral agents; Responsibility).

Questions asked in legal and moral philosophy require a more broadly-based conception of the phenomena of agency than the action/theoretic account on its own can provide.

See also: Rationality, practical

JENNIFER HORNSBY

References and further reading

Brand, M. and Walton, D. (eds) (1976) *Action Theory*, Dordrecht: Reidel.(Varied collection of accessible articles.)

Danto, A. (1965) ‘Basic Actions’, *American Philosophical Quarterly* 2: 141-48. (Original source for the idea of basicness: see above §2.)

Davidson, D. (1971) ‘Agency’, in R. Binkley *et al*. (eds) *Agent, Action and Reason*, Toronto, Ont.: Toronto University Press; repr. in Davidson (1980). (Defends a coarse-grained account of actions’ individuation, working with the idea of actions as a species of events which are ‘intentional under some description’.)

Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Oxford University Press.(The author’s ideas in the philosophy of action are presented (in the five essays in part 2) in the context of an account of events (part 1) and an account of the psychological (part 3).)

Donagan, A. (1987) *Choice: The Essential Element in Human Action*, London: Routledge & Kegan Paul. (Readable introduction, introducing more background than many standard works on ‘action theory’.)

Ginet, C. (1990) *On Action*, Cambridge: Cambridge University Press.(Good, elementary, if not easy, introduction, defending a ‘medium-grained account of actions’ individuation and a phenomenologically-based version of volitionalism. Also addresses some questions about human freedom.)

Goldman, A. (1970) *A Theory of Human Action*, Princeton, NJ: Princeton University Press.(Defends a fine-grained account of the individuation of act-tokens, or tropes. Considers some issues connected with

responsibility.)
Hornsby, J. (1980) *Actions*, London: Routledge & Kegan Paul.(Attends to distinctions between notions of basicness (in chapters 5 and 6), and to different versions of doctrines about actions' parts and causes (in chapters 3 and 4).)
Mill, J.S. (1843) *A System of Logic*, London: Routledge, vols 7 and 8, 1991.(One source of a componential volitionalist view.)
O'Shaughnessy, B. (1980) *The Will*, 2 vols, Cambridge: Cambridge University Press.(Detailed and sustained defence of a psychophysical conception of willed behaviour, paying close attention to phenomenology.)
Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson, ch.3.(Source of objections to dualistically conceived volitionalism.)
Searle, J. (1983) *Intentionality*, Cambridge: Cambridge University Press.(Chapter 3 explores the relation between intentions and actions using a conceptual apparatus developed for analysing problems of intentionality.)

Adorno, Theodor Wiesengrund (1903-69)

Philosopher, musicologist and social theorist, Theodor Adorno was the philosophical architect of the first generation of Critical Theory emanating from the Institute for Social Research in Frankfurt, Germany. Departing from the perspective of more orthodox Marxists, Adorno believed the twin dilemmas of modernity - injustice and nihilism - derived from the abstractive character of Enlightenment rationality. In consequence, he argued that the critique of political economy must give way to a critique of Enlightenment, instrumental reason.

Identity thinking, as Adorno termed instrumental rationality, abstracts from the sensory, linguistic and social mediations which connect knowing subjects to objects known. In so doing, it represses what is contingent, sensuous and particular in persons and nature. Adorno's method of negative dialectics was designed to rescue these elements from the claims of instrumental reason. Adorno conceded, however, that all this method could demonstrate was that an abstract concept did not exhaust its object. For a model of an alternative grammar of reason and cognition Adorno turned to the accomplishments of artistic modernism. There, where each new work tests and transforms the very idea of something being a work of art, Adorno saw a model for the kind of dynamic interdependence between mind and its objects that was required for a renewed conception of knowing and acting.

1 Life

Theodor Wiesengrund-Adorno (Wiesengrund, his father's name, shrank to the initial W. during his exile in California in 1943) was born in 1903 in Frankfurt. From his mother and sister the young Adorno derived his lifelong passion for music. Near the end of the First World War, Adorno began spending his Saturday afternoons studying Kant's *Critique of Pure Reason* with the social critic and film theorist Siegfried Kracauer. Under Kracauer's guidance, Adorno came to experience the first *Critique* not as mere epistemology, but ' as a kind of coded text from which the historical situation of spirit could be read' (1992: vol. 2, 58). This method of reading and thinking, entwining epistemology with social physiognomy, became the constitutive gesture of Adorno's philosophy.

After completing a dissertation on Edmund Husserl's phenomenology, Adorno received his doctorate from the Johann Wolfgang Goethe University in 1924. In the following year, he travelled to Vienna to study composition with Alban Berg and involve himself with the circle of composers and musicians gathered around Arnold Schoenberg. His Vienna interlude was to have a lasting impact; not only did he become a leading advocate of the 'new music', but his philosophical style can be traced to the 'atonal' compositional techniques of Schoenberg and Berg.

Returning to his studies in Frankfurt, Adorno took his habilitation with a thesis published as *Kierkegaard: Konstruktion des Ästhetischen (Kierkegaard: The Construction of the Aesthetic)* (1933). In this difficult work, three themes that were to remain decisive emerge: (1) the criticism of existentialism as betraying its desire for concreteness by transforming existential elements into abstract categories, such as that of subjectivity in Kierkegaard (Adorno continued to make an analogous criticism of Heidegger's notion of 'being'); (2) a reading of the social world as reified, that is, a world in which institutions indifferent to the claims of subjectivity dominate over persons; (3) the attempt to provide a historical and materialist concretization of theological ideas.

Adorno fled Hitler's Germany in 1934 to Merton College, Oxford. During his three and a half years in England, Adorno wrote articles for the house journal of the *Institut für Sozialforschung* (Institute for Social Research), which was then under the direction of his friend Max Horkheimer (see Frankfurt School), and worked on a book on Husserl, which was eventually published in 1956. Adorno spent the war years in the USA. During that time he collaborated with Horkheimer on *Dialektik der Aufklärung (Dialectic of Enlightenment)* (1947), often regarded (not altogether accurately) as *the* statement of first-generation critical theory (see Critical theory).

After the war, Adorno returned to Frankfurt to help rebuild the Institute. Over the next twenty years he produced a stream of works of musical and literary criticism, social theory and philosophy. His 1957 article 'Sociology and Empirical Research' is now regarded as the initiator of the 'positivist dispute' that raged in Germany in the 1960s, with Adorno and Karl Popper as the main combatants. Adorno's two major philosophical works, *Negative Dialektik (Negative Dialectics)* (1966) and *Ästhetische Theorie (Aesthetic Theory)* (1970), were written during this

period, the latter published a year after Adorno's death.

2 For and against Marx

Adorno's philosophy is a response to his understanding of the social world he inhabited. Adorno never doubted that advanced, Western societies were structured by capitalist relations of production as analysed by Marx. In particular, he accepted Marx's account of commodity fetishism and the domination of use values by exchange value. Adorno also accepted the proposal that the same mechanisms structuring the economy were effective in structuring cultural practices. While domination and poverty (broadly speaking, injustice) are the central consequences of capital's rationalization of the economy, alienation and meaninglessness (broadly speaking, nihilism) are the central consequences of its rationalization of culture (see Alienation; Nihilism).

However, against the background of the rise of fascism in Europe and the dissolution of workers' movements, later augmented by the events of the Holocaust, Adorno came to doubt that there really were significantly progressive tendencies latent in the economic and social fabric of the modern world. On the contrary, he came to believe that the rationalization of modern societies was all but complete, and hence came to view Marx's theory of history, with its commitment to an intrinsically progressive developmental sequence of social formations, as drawing on the same structures of rationality as those governing capitalist processes of production. If it is those structures of reason and rationality that are at the roots of the deepest dilemmas of modernity, then the crisis of modernity is primarily a crisis of reason. What is thus required before all else is a critical diagnosis of modern reason; in criticizing this formation of reason Adorno is simultaneously criticizing the world it engenders and providing the terms for a radical transformation of that world.

3 A genealogy of reason

It is modern scientific rationality, with its commitment to the primacy of method, analysis, subsumption, universality and logical systematicity, that Adorno believes is at the centre of the modern crisis of reason. He contends that knowing and its objects become deformed or distorted when reason is defined in terms radically independent of the objects to which it applies, where by 'objects' Adorno means not just objects known, but equally the sensory images of those objects, the articulation of those images in language, the entanglement of natural languages in social practices and the complex histories of those practices. Each of these items could be regarded as a systematic source of error (and in the course of the emergence of modern, enlightened rationality was so regarded), from sceptical worries about the deliverances of the senses to concern about collective prejudices sedimented in linguistic and social practice (see Descartes, R.; Bacon, F.). With respect to the theory of rationality, anxieties about these sources of error led to the view that reason must be fully autonomous, and not determined by anything external to it. It is this thought that underlies the primacy of method. In the theory of language, the same project is pursued in the attempt to eliminate opacity, indeterminacy and vagueness from the meaning of concepts; this is the project of positivism and the analytic tradition generally.

Dialectic of Enlightenment aims to provide a genealogy of enlightened rationality. Enlightenment opposes myth, the enchantment of the natural world through the projection onto it of human fears and hopes. The presumed superiority of reason over myth is hence its freedom from anthropomorphic projections; reason depicts the world objectively rather than through subjective projections. Horkheimer and Adorno contend that this flattering self-image of reason is both formally and substantively fallacious. Both myth and reason emerge in the course of humankind's struggle to free itself from bondage to mythic powers (themselves projections of primordial fear of the natural world in which humankind was immersed) and to gain control over the natural world in order to satisfy human needs and desires. Both myth and reason employ the principle of immanence, the explanation of every event as the repetition of a given pattern or law (what Adorno elsewhere calls 'identity thinking'), in order to combat fear of the natural world by bringing it into an explicable order. Repetition, 'the new is the old', originally provides for conceptual control over the natural world by revealing an intelligible order and eventually, through the technological application of modern science, for actual control over the natural world. Hence the formal features which provide for the supposed autonomy of enlightened reason are in fact grounded in the anthropogenesis of human reason in its struggle with nature. Enlightened reason is not objective, but subservient to the human desire to control nature; such reason can be construed as the discursive embodiment of the human drive for self-preservation, and hence as instrumental.

4 Nonidentity and negative dialectics

Enlightened reason is premised on a false inference: because some false beliefs (myths, superstitions and the like) are subjective projections, then the medium of those projections (sensory images, language, social practices and history) must themselves be systematic sources of error. Complete independence from these mediums is thus taken to be a condition for true knowledge. This drive for independence is most fully elaborated in the writings of the German Idealists, above all Kant and Fichte, where the autonomy of reason and the meaning-independence of concepts become explicitly identified with the spontaneity of the 'transcendental' subject. Unknown to itself, this subject and the philosophical concept of system it subtends are still in the throe of the drive to self-preservation, their abstract conceptuality still harbouring both fear and rage against their objects. The conception of idealism as rationalized rage is Adorno's appropriation and transformation of Nietzsche's notion of *ressentiment*. Idealist rage is directed at anything that refuses to fit or, in Adorno's terminology, is nonidentical with the demands of autonomous reason. Because the autonomy of reason is secured through the meaning-independence of concepts from concrete experience and its mediums, then what is incommensurable with this reason is whatever is irredeemably particular and contingent. The goal of Adorno's philosophy is the 'rescue' of nonidentity - the thing in itself in its concrete, historically mediated sensuous particularity.

Adorno's method of rescue is the use of dialectic. The point of dialectical analysis is to demonstrate that the rationalized concept of an object does not exhaust the thing conceived. It attains this end by showing that what were conceived to be extrinsic encumbrances on reason (sensory images and so on) that could be stripped away in its attainment of autonomy are in fact the *necessary mediations* through which knowing subjects come into relation to objects known. Adorno borrowed this conception of dialectic from Hegel. Adorno construes his dialectic as 'negative', in opposition to Hegel, because, on the one hand, he believes that Hegel's 'system' collapses back into the kind of identity thinking that dialectic opposes; and, on the other hand, because he believes that dialectical analysis only works under conditions in which the mediations it elaborates are systematically, in theory and in practice, denied.

Because an alternative conception of reason is not currently available, despite being a real historical possibility, Adorno's philosophy is utopian. Cognitively and practically, utopia is conceived of by Adorno 'as above identity and above contradiction; it would be a togetherness of diversity' ([1966] 1973: 150). An image of such 'togetherness of diversity' is provided by modernist works of art.

5 Aesthetic theory

Adorno argues that distinctly modernist works of art exemplify the possibility of an alternative grammar of reason and cognition. He focuses on modernist works - atonal music, abstract painting, 'absurdist' literature (particularly Kafka and Beckett) - because these works self-consciously attempt to establish their aesthetic validity, and hence their objectivity, in explicit opposition to all existing norms for artistic production and all established criteria in accordance with which art works have been judged. Existing norms and established criteria are the equivalents in art to the demands of method in science. Enlightened reason has it that such norms and criteria, in science and art, are spontaneous products of reason itself. Success for a modernist work is for it to be compelling, demanding aesthetic attention and assent, in excess of established criteria of aesthetic value and, even more radically, in excess of all criteria which heretofore have constituted what it is for an item to be a work of art.

'The falsehood opposed by art,' Adorno argues, ' is not rationality *per se* but the fixed opposition of rationality to particularity' ([1970] 1997: 144). The binding of rationality to what occurs in particular cases refutes the thesis of the meaning-independence of concepts from their objects and the autonomy of reason, and hence the principle of immanence. That this refutation occurs in art works entails that such binding is only a semblance or image of an alternative grammar of reason, since in modernity art is no longer a rationally legitimated vehicle of representation; art works now are 'meaningful' wholes without external purpose. That what happens in art can none the less matter to rationality generally derives from the hypothesis that the language of art and the discourse of rationality outside the artworld are not mutually indifferent language games. Rather, art picks up the debris of nonidentity left over from rationalization processes outside art; it is the refuge of the nonidentical. Further, art is driven to its modernist extremes of atonality, abstraction and absurdity in order to sustain itself as *art*, unique works of contemplation, in opposition to the recurrent demands of the principle of immanence.

Adorno's philosophical practice explicitly binds itself to the practices and fate of artistic modernism, and in this he is being self-consistent. Adorno aims to expose philosophy, the attempt to ground rationality and cognition, to its nonidentical other, forcing philosophy to surrender its claim to autonomy and meaning-independence. This is an avowedly peculiar terminus for a radical philosopher: defending the claims of the victims of history by forging an alliance between philosophy and high modernist art.

This state of affairs links together with the three dominant lines of criticism of Adorno's thought: (1) it is unduly pessimistic about the emancipatory potential of modern liberal societies; (2) it turns its face against the call for praxis indigenous to the Marxist tradition; (3) it provides only an aesthetic alternative to current problems and conceptions of reason. Although it will remain a matter of dispute, it can be argued that these objections simply bypass Adorno's original insight, namely that the dilemmas of injustice and nihilism have a common root in the abstractive achievements of autonomous reason. Traditional Marxism focuses on the question of injustice, while ignoring the problem of nihilism; conversely, existentialists such as Nietzsche and Heidegger aim to overcome nihilism while they remain insensitive to the claims of justice. If Adorno is correct in maintaining that these dilemmas are interconnected, then his philosophy has something to say to us. The fragile hope of his philosophizing lies in the belief that the claims of justice are best served through the defence of the claims of the rationality inherent in modernist works of art.

See also: Enlightenment, Continental

J.M. BERNSTEIN

List of works

Adorno, T.W. (1933) *Kierkegaard. Konstruktion des Ästhetischen*, Tübingen: J.C.B. Mohr; trans. R. Hullot-Kentor, *Kierkegaard: Construction of the Aesthetic*, Minneapolis, MN: University of Minnesota Press, 1989.(This is Adorno's first major work, and it includes all the major themes - the critique of existentialism as abstract, the role of aesthetics, the thematics of sacrifice - that will come to dominate his thought. The introduction to the translation by Hullot-Kentor is helpful.)

Adorno, T.W. and Horkheimer, M. (1947) *Dialektik der Aufklärung*, Amsterdam: Querido; trans. J. Cumming, *Dialectic of Enlightenment*, London: Allen Lane and New York: Herder & Herder, 1972.(This is the founding document of first-generation Critical Theory in which the critique of instrumental reason comes to displace the critique of political economy. It includes Adorno's famous treatment of Odysseus as already enacting the Enlightenment sacrifice of the particular to the universal, and his analysis of the culture industry.)

Adorno, T.W. (1949) *Philosophie der neuen Musik*, Tübingen: J.C.B. Mohr; trans. A.G. Mitchell and W. Blomster, *Philosophy of Modern Music*, Sheed & Ward, 1973.(Adorno's classic defence of Arnold Schoenberg's twelve-tone system as the high point of musical modernism. Its conception of modern music was the crucial source for Thomas Mann's *Doctor Faustus*.)

Adorno, T.W. (1951) *Minima Moralia. Reflexion aus dem Beschädigten Leben*, Frankfurt: Suhrkamp; trans. E.F.N. Jephcott, *Minima Moralia: Reflections from Damaged Life*, London: New Left Books, 1974.(One hundred and fifty-three dazzling aphorisms, in which Adorno reflects on the vanishing of concrete, individual experience in modern, bourgeois society. Its fluent mixture of philosophy and cultural criticism makes it the most accessible of Adorno's works.)

Adorno, T.W. (1956) *Zur Metakritik der Erkenntnistheorie. Studien über Husserl und die phänomenologischen Antinomien*, Stuttgart: Kohlhammer; trans. W. Domingo, *Against Epistemology: A Metacritique, Studies in Husserl and the Phenomenological Antinomies*, Oxford: Blackwell, 1982.(A dense reading of Husserl's phenomenology, with the emphasis on the inevitable abstractness of the phenomenological method, and hence its loss of the very concreteness it seeks.)

Adorno, T.W. (1957) 'Sociology and Empirical Research', in T.W. Adorno *et al.*, *Der Positivismusstreit in der deutschen Soziologie*, Neuwied and Berlin: Luchterhand, 1969; trans. G. Adey and D. Frisby, *The Positivist Dispute in German Sociology*, London: Heinemann, 1970.(Includes essays by Popper, Habermas, Dahrendorf, Harald Pilot and Hans Albert, among others. The translation also includes a review of the original by Popper, and a helpful introduction by David Frisby.)

Adorno, T.W. (1963) *Drei Studien zu Hegel*, Frankfurt: Suhrkamp; trans. S.W. Nicholsen, *Hegel: Three Studies*, Cambridge, MA, and London: MIT Press, 1993.(These very essayistic explorations of Hegel elaborate the competing ideals of rationality in dialectical and deductive thinking.)

Adorno, T.W. (1964) *Jargon der Eigentlichkeit. zur deutschen Ideologie*, Frankfurt: Suhrkamp; trans. K. Tarnowski and F. Will, *Jargon of Authenticity*, London: Routledge & Kegan Paul, 1973. (Adorno's fiercely

critical account of Martin Heidegger's existentialism as abstract and ahistorical.)

Adorno, T.W. (1966) *Negative Dialektik*, Frankfurt: Suhrkamp; trans. E.B. Ashton, *Negative Dialectics*, London: Routledge & Kegan Paul, 1973.(This work pursues an immanent critique of the idealism of Kant and Hegel as the vehicle for a critique of modern instrumental reason. It contains Adorno's most sustained arguments concerning the nature of human conceptuality, and his famous reflections on the meaning of philosophy 'after Auschwitz'.)

Adorno, T.W. (1970) *Ästhetische Theorie*, Frankfurt: Suhrkamp; trans. R. Hullot-Kentor, *Aesthetic Theory*, London: Athlone Press, and Minneapolis, MN: University of Minnesota Press, 1997.(Arguably this is the premier work of twentieth-century philosophical aesthetics and the philosophy of art. Transforming the central concepts of modern aesthetics accordingly, Adorno contends that the works of high modernism model a suppressed conception of human rationality that challenges that of Enlightenment rationalism.)

Adorno, T.W. (1992) 'The Curious Realist: On Siegfried Kracauer', in *Notes to Literature*, trans. S.W. Nicholson, New York: Columbia University Press.(Adorno's account of the thought and influence of his early tutor.)

References and further reading

Bernstein, J.M. (1992) *The Fate of Art: Aesthetic Alienation from Kant to Derrida and Adorno*, Oxford: Polity Press and State College, PA: Penn State Press.(Chapters 4 and 5 detail the argument of §5.)

Bernstein, J.M. (ed.) (1994) *The Frankfurt School, vol. 2: Horkheimer and Adorno*, London: Routledge.(Contains twenty of the best journal articles on the whole range of Adorno's output, including influential pieces by Jürgen Habermas on *Dialectic of Enlightenment* and Albrecht Wellmer on *Aesthetic Theory*.)

Bernstein, J.M. (1997) *Adorno: Of Ethics and Disenchantment*, Cambridge: Cambridge University Press. (Expounds the analysis of concepts and rationality in §3, and relates it to the ethical vision animating Adorno's philosophy.)

Buck-Morss, S. (1977) *The Origin of Negative Dialectics: Theodor W. Adorno, Walter Benjamin and The Frankfurt Institute*, Hassocks, Sussex: Harvester Press.(An intellectual biography of Adorno that usefully focuses on the deep influence of his friend Walter Benjamin in the formation of several of the key concepts in Adorno's theory.)

Jarvis, S. (1986) *Adorno: A Critical Introduction*, Cambridge: Polity Press.(An illuminating study that tracks Adorno's Hegelianism and his use of the idea of determinate negation through the breadth of his work.)

Jay, M. (1984) *Adorno*, London: Fontana.(A clear introduction for the general reader, but without a sharp, philosophical focus.)

Rose, G. (1978) *The Melancholy Science: An Introduction to the Thought of Theodor W. Adorno*, London: Macmillan.(An advanced introduction, particularly strong on reification and the concept of dialectic.)

Sacks, M. (1990) 'Through a Glass Darkly: Vagueness in the Metaphysics of the Analytic Tradition', in D. Bell and N. Cooper (eds) *The Analytic Tradition: Meaning, Thought and Knowledge*, Oxford: Blackwell.(Elaborates the suggestion about language made in §3 in Wittgensteinian terms.)

Zuidervaart, L. (1991) *Adorno's Aesthetic Theory: The Redemption of Illusion*, Cambridge, MA, and London: MIT Press.(A sound study of Adorno's aesthetics, especially useful on the central terms of Adorno's argument in *Aesthetic Theory*. Zuidervaart is critical of Adorno's strong cognitivist approach to art.)

Adverbs

Adverbs are so named from their role in modifying verbs and other non-nominal expressions. For example, in 'John ran slowly', the adverb 'slowly' modifies 'ran' by characterizing the manner of John's running. The debate on the semantic contribution of adverbs centres on two approaches. On the first approach, adverbs are understood as predicate operators: for example, in 'John ran slowly', 'ran' would be taken to be a predicate and 'slowly' an operator affecting its meaning. Working this out in detail requires the resources of higher-order logic. On the second approach, adverbs are understood as predicates of 'objects' such as events and states, reference to which is revealed in logical form. For example, 'John ran slowly' would be construed along the lines of 'there was a running by John and it was slow', in which the adverb 'slowly' has become a predicate 'slow' applied to the event that was John's running.

Since adverbs are exclusively modifiers, they are classed among the syncategorematic words of terminist logic, the investigation of which carried the subject forward from Aristotle in the thirteenth century. (The contrasting 'categoremata' - grammatical subjects and predicates - are those words which have meaning independently.) They are of contemporary interest for philosophical logic and semantic theory, because particular accounts of them carry implications for the nature of combinatorial semantics and language understanding, and for ontology.

1 Syntactic types and semantic combination

There are several types of adverbial constructions, of which we distinguish the following classes: (a) 'manner' adverbs, which intuitively function as simple modifiers of verbs; (b) 'thematic' adverbs, of which some and possibly all function as (at least) two-place predicates in their own right; (c) adverbs of quantification, which express generality applying to whole sentences; and (d) discourse particles, whose meaning evidently derives from their role in linking clauses or independent sentences. (These categories are not exhaustive.) 'Adverbs', especially manner 'adverbs', are not in fact confined to single words. The general category is therefore not that of adverbs, but of adverbial phrases or adverbials (for example, 'more quickly than Mary', 'very frequently').

Typical manner adverbials are as in (1) below, thematic adverbs as in (2) and adverbs of quantification as in (3):

(1) John walked slowly/quietly/more quickly than Mary.
(2) Mary apparently/reluctantly went to New York.
(3) Mary occasionally/always walks to work.

Discourse particles, considered briefly below, include 'but', 'anyway' and several others. We discuss these cases in turn.

The essential logical problem of manner adverbials is already apparent in the simplest examples. A verb combines with a manner adverb to form a complex verbal construction of the same type. Thus 'walk' and 'walk slowly' are both predicates, and the syntax of the combination may be depicted as follows:

$[_V [_V \text{walk}][_{Adv} \text{slowly}]]$

If (disregarding tense) we take 'walk' as a one-place predicate, then the semantics of this combination might be given by positing that 'slowly' is interpreted as a predicate operator; that is, as a function that maps one-place predicate interpretations onto other one-place predicate interpretations. Alternatively, it may be suggested that 'slowly' and the other manner adverbials are, logically speaking, predicates in their own right, specifically predicates of actions. The adjectives to which they are related do seem to play this role. Corresponding to (1), for instance, we have the adjectival predications

(4) John's walk was slow/quiet/quicker than Mary's.

If we take the further step of supposing that the verb 'walk' is in fact a two-place predicate, with a position for actions, then the combination 'walk slowly' can be interpreted as

walk(x, e) & slow(e),

where *e* ranges over actions. Comparing this account with the first alternative,

(slowly(walk))(*x*),

we see a trade-off: where predicates are taken to have a simple structure, the adverbial must be understood as an operator; but where extra structure, in the form of a place for actions, is posited, the semantic combination is truth-functional and predicate operators are not required.

The alternatives just sketched each have their defenders in the literature on adverbials. Adverbs are construed as predicate operators in formal theories of linguistic structure, including those of Montague (1974) and Lewis (1975). The predicative alternative was first advanced at length by Davidson (1967), and is elaborated by Parsons (1990). Semantic and metaphysical issues arise for each account; we take up some of these below.

Thematic adverbs are intuitively distinguished from manner adverbs in as much as they yield constructions adverb + verb which cannot be treated simply as new verbs: apparently going to New York is not a way of going to New York; and reluctantly going to New York is not a manner of travel, but an instance of travel whose agent was reluctant to undertake it. For the examples in (2) the following paraphrases suggest themselves:

It was apparent that Mary went to New York.
Mary went to New York and she was reluctant to go to New York.

The discourse particles, traditionally and appropriately called adverbs, have come under relatively formal study only in recent years. The following examples are representative:

(5) He was poor *but* honest.
(6) *Anyway*, I’m going to New York.

They resist analysis in terms of their contribution to truth-conditions, but carry implications for the discourses in which they occur, with (5) involving some presumptive contrast with what might have been expected and (6), as an assertion, functioning to indicate a return to a superordinate topic of conversation. See Levinson (1983) for examples and discussion.

2 Ambiguities

Many adverbs are ambiguous between thematic-adverbial and manner-adverbial interpretation. In an example such as

(7) Mary quickly objected,

we may have either the interpretation ‘Mary’s objection was delivered in a quick manner’ or the interpretation ‘Mary’s objection came a short time after the enunciation of the proposition to which she objected’. The second, thematic-adverbial interpretation shows up in the corresponding adjective ‘quick’ in a construction such as

Mary was quick to object.

These examples suggest that grammatical appearance belies logical structure, since the thematic adverbial in (7) functions, logically speaking, as the main predicate of that sentence. Austin (1956) observed that adverbial position often disambiguates, with post-verb adverbs favouring the manner-interpretation, and pre-verb adverbs the thematic, as in the pair

He trod on the snail clumsily.
Clumsily, he trod on the snail.

Adverbs of quantification pick up arguments including, but not restricted to, the temporal. Sentences such as (8) are ambiguous, depending upon whether the quantification is over occasions, or over the subject:

(8) Travel books are seldom worth reading.

Where the adverb quantifies over time, (8) means that the occasions are few when it is worth reading travel books. But there is another salient interpretation:

Few travel books are worth reading.

Where quantification over time would be ridiculous in view of the subject matter, construal of the adverb with the subject is particularly salient, for example:

Quadratic equations seldom have real solutions.

The above reflections have it that thematic adverbs and adverbs of quantification are not modifiers at all, except in a purely grammatical sense. Inversely, there have been suggestions that what appear in language as if they were arguments to a predicate actually function as adverbial modifiers. Perhaps the best known account of this type is Roderick Chisholm's discussion of a certain class of statements about appearances. Chisholm suggests that a man who 'sees spots before his eyes' should be thought of as 'sensing in a spotty manner' or as someone whom things 'appeared to spottily' (1957). This philosophical move is designed to rid the locution of any implication that in sensing spots before his eyes the man is sensing a mysterious object, an appearance, which is before him, his eyes or his mind (see Mental states, adverbial theory of). In a similar vein, Goodman ([1968] 1976) construed the locution '*x* represents an *F*', under the condition where there is no implication that such an *F* exists, as 'is an *F*-representation', effectively treating the predicate *F* adverbially.

The suggestions of Chisholm and Goodman, left as they are, become problematic if the project of giving a combinatorial semantics for language is taken seriously (see Compositionality). Suppose *x* is a picture and (9) is true, with no existential implications:

(9) *x* represents two unicorns galloping in a field.

If we write, with Goodman,

x is a two-unicorns-galloping-in-a-field representation,

then we have yet to provide a semantic structure to go along with the syntax. But there must be some such structure, since, for example, (9) obviously implies

x represents more than one unicorn in motion.

Similar issues arise for the predicate-operator theory of manner adverbs. For instance, that theory does not immediately deliver the obvious implication

(10) John walked slowly; therefore, John walked.

In Montague (1974) and much subsequent literature, this and similar implications are the consequences of semantic postulates in the sense of Carnap (1956). Even with such postulates, the relation between the adverbial constructions in (1) and their adjectival paraphrases (4) requires clarification. By contrast, the predicative theory proposed by Davidson is specifically designed to capture such implications and paraphrases. The premise of (10) is rendered as

(11) $(\exists e)$(walk(John, *e*) & slow(*e*)),

that is, 'There was a walk by John and it was slow'; and the conclusion has the form

(12) $(\exists e)$walk(John, *e*),

'There was a walk by John', a trivial implicate of the premise. For the paraphrase relation between (1) and (4), we have only to note that the complex noun 'John's walk' would be understood as a definite description of an action; that is, as '(the *e*) walk(John, *e*)'. 'John's walk was slow' then becomes

(13) slow((the *e*) walk(John, *e*)).

Given any standard treatment of the definite description, (13) will imply (11). However, the Davidsonian view is committed to supplying extra, hidden structure in all cases, and to taking events as individuals, a step that has often been considered metaphysically dubious.

Ensembles of adverbs show ambiguity of scope (see Scope). For example, one interpretation of (14) has John being clever in that he made a stupid response:

(14) John cleverly responded stupidly.

Modal adverbs such as 'necessarily' and 'contingently', which modify whole sentences, allow singular terms and quantifiers within their superficial scope to be interpreted as outside it (a point known to the terminist logicians). Tracking the relative scopes of modalities is part of the contemporary application of modal logic. As in the cases of manner and thematic adverbs, there are both operator-theories and predicate-theories of these expressions; that is, the modalities may be developed as one-place modal operators, with the same syntax as negation; or as predicates of sentences in a first-order formulation of the logical syntax of natural language (see Modal operators). Montague (1963) argued that the latter was unacceptable since the normal laws of modal logic could not all be maintained, on pain of paradox. The argument has subsequently been developed and discussed by others: see Koons (1992) for a survey and response.

3 Extensionality

Adverbial constructions of both the manner and thematic types, together with others, show a kind of superficial but persistent non-extensionality: superficial, because it may disappear under analysis; and persistent, because it may show up in places that the analysis itself uncovers. Predicate-operator theories can accept non-extensionality as the norm. Supposing, for example, that those who breathe are exactly those who perceive, it is absurd to infer that *x* perceives rapidly from the premise that *x* breathes rapidly. The predicate-operator theory, taking all operations to be in intension, has no such consequence. Or supposing that those who went to New York were exactly those who visited Times Square, it does not follow that if Mary reluctantly went to New York she also reluctantly visited Times Square.

For theories of Davidson's type, the constructions

x breathed rapidly
x perceived rapidly

are of the same logical type, but the events on which the adverb is predicated are different. The coextensiveness of 'breathe' and 'perceive' amounts to the coextensiveness of

(for some *e*) breathe(*x*, *e*)
(for some *e*) perceive(*x*, *e*).

Such coextensiveness no more implies the equivalence of '*x* breathed rapidly' to '*x* perceived rapidly' than the coextensiveness of '*x* kicked something' and '*x* saw something' would imply the equivalence of '*x* saw something red' and '*x* kicked something red'.

Non-extensionality is more troublesome when one considers events related as genus and species. For example, 'Mary flew slowly across Spain' does not imply 'Mary travelled slowly across Spain', even though any event of flying is an event of travelling. It follows that it is inadequate to represent 'Mary flew slowly across Spain' simply as

(for some *e*) (fly(Mary, *e*) & across Spain(*e*) & slow(*e*)).

Rather, we must add that the event *e*, which was a flying and therefore a travelling, was slow *for a flying*. These examples show at least that manner adverbials are relative to the sets of events against which a given event is evaluated. Other examples may show that it is not merely sets but also properties, not extensionally individuated, that form the background.

In any case, non-extensionality is evident in constructions with thematic adverbs. Davidson, following a suggestion by Hector-Neri Castañeda, considered examples such as that of

Oedipus intentionally married Jocasta,

from which it does not follow that Oedipus intentionally married his mother. In this case, the locus of non-extensionality, although due to the presence of the adverb 'intentionally', is the predicate 'intend', from which the adverb is derived. In this sense, it is independent of any peculiarities of adverbs, as in the Davidsonian paraphrase

Oedipus married Jocasta, and he intended to marry Jocasta.

Higginbotham (1989) explores a number of similar examples.

See also: Logical and mathematical terms, glossary of; Syntax §6

JAMES HIGGINBOTHAM

References and further reading

Austin, J.L. (1956) 'A Plea for Excuses', *Proceedings of the Aristotelian Society*; repr. in *Philosophical Papers*, Oxford: Clarendon Press, 1961, 123-52.(A famous and highly readable paper in which Austin tries to clarify the nature of action from the way in which we use excuses to escape responsibility for the consequences of what we do.)

Carnap, R. (1956) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Develops the idea that many apparently logical inferences are to be accounted for by semantic postulates that relate the meanings of the words in premises and conclusion; quite technical.)

Chisholm, R. (1957) *Perceiving: A Philosophical Study*, Ithaca, NY: Cornell University Press, esp. ch. 8. (Proposes an adverbial analysis of perception statements to avoid commitment to such entities as 'sense-data'.)

Davidson, D. (1967) 'The Logical Form of Action Sentences', in N. Rescher (ed.) *The Logic of Decision and Action*, Pittsburgh, PA: University of Pittsburgh Press; repr. in *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, 105-48.(An analysis of adverbial modification as predication of events; readable and influential.)

Goodman, N. (1968) *Languages of Art*, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1976.(Proposes an adverbial analysis of the likes of 'is a picture of an *F*' to prevent the existence of a picture of an *F* entailing the existence of a pictured *F*.)

Higginbotham, J. (1989) 'Elucidations of Meaning', *Linguistics and Philosophy* 12: 465-517. (§2 defends a Davidsonian analysis in the context of a discussion of how lexical analysis interacts with permissible modes of syntactic composition; quite technical.)

Koons, R.C. (1992) *Paradoxes of Belief and Strategic Rationality*, Cambridge: Cambridge University Press. (Mainly concerns the role of common knowledge or mutual belief in generating liar-like paradoxes about rational action; very technical.)

Kretzmann, N. (1968) *William of Sherwood's Treatise on Syncategorematic Words*, Minneapolis, MN: University of Minnesota Press.(Elucidates the categorematic/syncategorematic distinction mentioned in the introduction to this entry.)

Levinson, S.C. (1983) *Pragmatics*, Cambridge: Cambridge University Press. (Includes discussion of words such as 'but' and 'anyway' which are traditionally regarded as adverbs but whose contribution to meaning is primarily at the level of pragmatics.)

Lewis, D.K. (1975) 'Adverbs of Quantification', in E. Keenan (ed.) *Formal Semantics of Natural Language*, Cambridge: Cambridge University Press, 3-15. (Semantics for adverbs construed as predicate operators.)

Montague, R. (1963) 'Syntactical Treatment of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability', *Acta Philosophica Fennica* 16; repr. in *Formal Philosophy*, ed. R. Thomason, New Haven, CT: Yale University Press, 1974, 286-302.(Shows that straightforward analysis of modal adverbs ('necessarily', 'possibly') as predicates of sentences ('is necessarily true', 'is possibly true') leads to the generation of a paradox similar to Tarski's in his proof of the 'undefinability of truth'; very technical.)

Montague, R. (1974) *Formal Philosophy*, ed. R. Thomason, New Haven, CT: Yale University Press.(Montague's collected papers in natural language semantics; extremely technical. For adverbs, see especially 'English as a Formal Language'.)

Parsons, T. (1990) *Events in the Semantics of English*, Cambridge, MA: MIT Press. (Extended development and discussion of Davidson's analysis of adverbs; clear and, for the most part, non-technical; good bibliography.)

Aenesidemus (1st century BC)

Aenesidemus was a Greek philosopher of the first century BC who revived Pyrrhonian Scepticism, formulating the basic Ten Modes of Scepticism, or tropoi*, and demonstrating that concepts such as cause, explanation, goodness and the goal of life engendered endemic and undecidable dispute; faced with this the Sceptic suspends judgment - and tranquillity follows.*

Aenesidemus was probably active around the middle of the first century BC. He considered that Academic scepticism under Philo of Larissa had so far abandoned its original, uncompromising attitude to knowledge that he described the dispute between Philo and Antiochus as 'Stoics fighting with Stoics' (fr. 71C9). In response, he turned to the Scepticism of Pyrrho for sustenance, effectively re-founding Pyrrhonism and determining the broad outlines it was to follow (see Pyrrhonism). The fundamental arguments of the Ten Modes of Scepticism are attributed to him. He wrote an eight volume *Pyrrhonian Discourses*, a summary of which survives in the ninth-century patriarch Photius' library catalogue (frs 71C, 72L), and is said to have composed an *On Inquiry*, an *Against Wisdom* and a *First Introduction*.

The first book of *Pyrrhonian Discourses* argued that the Academics were in fact dogmatists, committed to beliefs both positive ('some things are plausible') and negative ('nothing is apprehensible'). Pyrrhonists, by contrast, 'determine absolutely nothing, not even this claim that nothing is determined' (fr. 71C8). They will not assert dogmatically (that is, with strong commitment to the truth) that something either is or is not the case, saying only that it no more is than is not, or that it sometimes is and sometimes is not, or that it is for one person and not for another (fr. 71C6-7) (see Pyrrhonism §§1, 3). This, according to Aenesidemus, amounts simply to following the appearances, reacting to the way things seem to be; about reality the Pyrrhonist suspends judgment, the result of such suspension being tranquillity (Diogenes Laertius, IX 106-7).

The remainder of *The Pyrrhonian Discourses* cast doubt on the concepts of dogmatist physics (cause, principle, generation, motion, and so on) (see Pyrrhonism §5) and the veridicality of perception, as well as dealing sceptically with signs, the gods, scientific explanation and various topics in ethics, the aim being to emphasize the dubiousness of all dogmatic positions on these topics and the extent of their differences, with a view to promoting suspension of judgment (and hence tranquillity). Some of these arguments are preserved elsewhere: signs should be evident as signs (that is in what they signify) to everybody if they are to function as signs, but they are not (Sextus Empiricus, *Against the Professors* VIII 215); the concept of cause is incoherent (IX 219-26); scientific 'explanations' are underdetermined (Sextus Empiricus, *Outlines of Pyrrhonism* I 180-5). Aenesidemus' Scepticism appears both consistent and complete.

Elsewhere, however, Sextus reports that 'Aenesidemus…says that the Sceptic way is a road leading to Heraclitean philosophy, since saying that opposites appear to hold of the same thing precedes saying that they actually do hold of the same thing' (*Outlines of Pyrrhonism* I 210) (see Heraclitus §3). Thus, apparently, indeterminacy in appearances is grounds for belief in indeterminacy in the objects, which is unsceptical, involving a commitment to the way things actually are. Perhaps Aenesidemus simply offered such arguments dialectically, to emphasize dogmatic disagreement; perhaps he thought (possibly following Pyrrho - see Pyrrhonism §1, and compare Pyrrho §3) that such statements were coherently Sceptical; or perhaps he simply changed his mind. The evidence is insufficient - and we can only suspend judgment on the issue.

R.J. HANKINSON

References and further reading

Aenesidemus (1st century BC) Fragments, in A.A. Long and D.N. Sedley, *The Hellenistic Philosophers*, vol. 1, Cambridge: Cambridge University Press, 1987. (Invaluable collection of fragments in translation with philosophical commentary; contains Photius' report of Aenesidemus at §§71-2.)

Diogenes Laertius (*c.*early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(Greek text with facing translation; IX 61-116 is devoted to Pyrrhonism.)

Hankinson, R.J. (1995) *The Sceptics*, London: Routledge.(Chapter VII deals with Aenesidemean Scepticism.)

Sextus Empiricus (*c.* AD 200) *Outlines of Pyrrhonism*, trans. J. Annas and J. Barnes, *Outlines of Scepticism*,

Cambridge: Cambridge University Press, 1994.(Fine translation with introduction and notes.)

Sextus Empiricus (*c.* AD 200) *Against the Professors*, trans. R.G. Bury, *Against the Logicians*, *Against the Physicists*, *Against the Ethicists* and *Against the Professors*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 3 vols, 1935, 1936, 1949.(Parallel Greek text with English translation and minimal notes.)

Woodruff, P. (1988) 'Aporetic Pyrrhonism', *Oxford Studies in Ancient Philosophy* 6: 139-68.(Stimulating if unorthodox treatment of Aenesidemus' Pyrrhonism.)

Aesthetic attitude

It is undeniable that there are aesthetic and non-aesthetic attitudes. But is there such a thing as the *aesthetic attitude? What is meant by the aesthetic attitude is the particular way in which we regard something when and only when we take an aesthetic interest in it. This assumes that on all occasions of aesthetic interest the object attended to is regarded in an identical fashion, unique to such occasions; and this assumption is problematic. If an attitude's identity is determined by the features it is directed towards; if an aesthetic interest in an object is (by definition) an interest in its aesthetic qualities; and if the notion of aesthetic qualities can be explained in a uniform manner; then there is a unitary aesthetic attitude, namely an interest in an item's aesthetic qualities. But this conception of the aesthetic attitude would be unsuitable for achieving the main aim of those who have posited the aesthetic attitude. This aim is to provide a definition of the aesthetic, but the aesthetic attitude, understood as any attitude focused upon an object's aesthetic qualities, presupposes the idea of the aesthetic, and cannot be used to analyse it. So the question is whether there is a characterization of the aesthetic attitude that describes its nature without explicitly or implicitly relying on the concept of the aesthetic. There is no good reason to suppose so. Accordingly, there is no such thing as the aesthetic attitude, if this is an attitude that is both necessary and sufficient for aesthetic interest and that can be characterized independently of the aesthetic.*

1 The idea of the aesthetic attitude

What is meant by the aesthetic attitude? Although it has often been supposed that there is such a thing, its existence has also been denied. Those who believe in the aesthetic attitude embrace one of two conceptions: either a certain conception of the essence of *aesthetic* appreciation or interest, or a certain conception of the essence of *artistic* appreciation or interest, that is, appreciation of or interest in something *as* a work of art. Each maintains that there is a particular attitude we adopt towards an item when and only when our interest in the item is of a certain kind, and each calls this attitude 'the aesthetic attitude'; but whereas the first alternative specifies this interest as appreciation of the item from the aesthetic point of view, the second identifies it as appreciation of the item as a work of art. The first is usually understood as being the wider conception, including the second as a special case. For it is often believed that (1) aesthetic interest can be directed towards either nature or art, and (2) to be interested in a work of art *as* a work of art is to be interested in it aesthetically. But the first conception is not always thought of as wider than the second. Some thinkers deny (2), on the ground that interest in a work of art as a work of art requires more than, or something different from, an aesthetic interest in the item. And some thinkers maintain that the concept of regarding something as a work of art is basic, and that to regard nature aesthetically *is* to regard it as, or as if it were, a work of art. Without prejudging any of the issues, it will simplify exposition to focus on the first conception.

2 Aesthetic and non-aesthetic attitudes

One source of the idea that there is such a thing as the aesthetic attitude is undoubtedly the thought that any item that we can treat or regard aesthetically we can also treat or regard non-aesthetically. It is indeed true that we can adopt towards any item - even something that is a work of art - an attitude that is not concerned with its aesthetic appeal, for there certainly are non-aesthetic or non-artistic attitudes and no item necessarily precludes having an attitude of one of these kinds directed towards it. You can look at a hibiscus, not to delight in the beauty of its form and colours, but merely to see whether it needs water; and even if you are bored by Bach's *The Art of Fugue* you might listen to it in order to fall asleep. Now from the fact that for any item it is possible to adopt an attitude towards it that is not concerned with its aesthetic or artistic appeal, it is easy to jump to the conclusion that there is an attitude we adopt towards any item when and only when we take an aesthetic interest in it. This would be *the* aesthetic attitude. However, the conclusion makes a much stronger claim than the premise from which it is derived. The premise claims only that whatever you may be taking an interest in, your interest might not be in its aesthetic or artistic appeal. This does not imply the double-barrelled thesis that (1) when your interest in any item is aesthetic or artistic, your attitude towards the item must always be of the same unitary kind, and (2) whenever your attitude towards anything is of this kind, you are taking an aesthetic or artistic interest in the item. So the conclusion follows from the premise only if the premise is supplemented.

The strongest basis for the conclusion would be, first, an exhaustive specification of different kinds of attitude and,

second, a demonstration that one of these attitudes is such that (a) if any other attitude is adopted towards an item, but this attitude is not, then the item is not being regarded aesthetically (and so the attitude is necessary for aesthetic appreciation), and (b) if an item is the object of this attitude, then it is being treated aesthetically, no matter what other kinds of attitude may also be directed towards it (and so the attitude is sufficient for aesthetic appreciation). The provision of such a basis is clearly the intention of those who offer, first, a positive characterization of a certain attitude, all other attitudes being defined by contrast with this attitude (thus exhausting the possibilities), and, second, a number of examples for which, supposedly, the adoption of this attitude is both necessary and sufficient for the item to be the object of aesthetic appreciation.

If there is such a thing as the aesthetic attitude, can it be given a positive characterization that specifies its nature? Nearly all adherents of the aesthetic attitude have believed that it is susceptible of a helpful analysis, but it has also been argued that it is not.

The principal interest of the concept of the aesthetic attitude derives from a concern to define or circumscribe the notion of the aesthetic or (more usually) of art. One leading idea has been that of defining the notion of a work of art in terms of the aesthetic attitude, perhaps in such a fashion as this: a work of art is an artefact solely, mainly or at least partly designed to give satisfaction to the aesthetic attitude. But a definition of art in terms of the aesthetic attitude will be illuminating only if an understanding of the concept of the aesthetic attitude does not presuppose an understanding of the aesthetic. If the aesthetic attitude is identified merely as the attitude of concern for or interest in an item’s aesthetic properties or aesthetic value, and if a work’s artistic properties or artistic value are constituted (partly or wholly) by its aesthetic properties or value, then the suggested definition will be unhelpfully circular; and an analogous conclusion will follow if the aesthetic attitude is defined as the attitude necessary for or most conducive to the derivation of the aesthetic pleasure an object merits, or that is necessary to ensure that an aesthetic judgment about an object is as well-founded as it can be. So the question arises of whether it is possible to define the aesthetic attitude independently of any idea of the aesthetic.

I shall consider, first, the project of defining the aesthetic attitude independently of any prior concept of the aesthetic; second, the idea that although there is such a thing as the aesthetic attitude, it resists definition; and third, the claim that the aesthetic attitude is a myth.

3 Characterizing the aesthetic attitude

A recognition of truth, beauty and goodness as the principal concerns of the human mind has given rise to the idea that the aesthetic attitude must be distinguished from, on the one hand, cognitive attitudes, and on the other, practical ones. Whereas a cognitive attitude towards an object is concerned with the acquisition of knowledge from it and a practical attitude with its utility, the aesthetic attitude has a different focus. Or so it is claimed. How might this focus be defined?

In fact, only a small number of locutions have been used to characterize the aesthetic attitude. Perhaps the most common has been in terms of the notion of disinterestedness. That your attitude towards an object is disinterested does not mean that you are not interested in it. What does it mean? If ‘disinterested’ means no more than ‘unbiased’, this would not serve by itself to mark off the aesthetic attitude from paradigms of cognitive or practical attitudes, which can also be unbiased. If it means that your attitude is not aesthetic if you are interested in determining what the object is, what to do with it or its suitability for some purpose, it is wide of the mark; for, leaving other considerations aside, it rules as unaesthetic those cases in which you consider whether an object is suitable for an *aesthetic* end, such as when you consider whether a vase is the right shape, size and colour for the collection of flowers you propose to arrange.

It is sometimes suggested that something is an object of aesthetic appreciation only if it is being attended to for its own sake. This idea can be understood in stronger and weaker forms. On the one hand, ‘for its own sake’ might mean ‘for no further reason but just for the sake of it’. But this would yield a mistaken account of aesthetic appreciation, for you can have a variety of reasons for attending to something that you are appreciating aesthetically. On the other hand, ‘for its own sake’ might mean ‘not solely as a means to an end’, thus allowing for the possibility that an object is being attended to for its own sake, or as an end in itself, even when it is being attended to for ulterior purposes. If this requires a spectator’s interest in the object not to be solely concerned with some *non-aesthetic* end, it would render an account of the essence of the aesthetic in terms of the aesthetic attitude

viciously circular. But if the requirement is only that the interest should not be solely as a means to *some* end, then even if this were to be a necessary condition it would not be a sufficient one for the interest to be aesthetic. Your interest in a mathematical proof or a game of soccer is not aesthetic simply because you are uninterested in any use to which these might be put.

Another suggestion is that your attitude towards an object is aesthetic if and only if, in interacting with it, you consider just the object itself - its elements and the relations amongst its elements - not any relations in which it stands to anything other than itself. Accordingly, the aesthetic attitude is thought of as being an attitude of contemplative detachment from all considerations of utility, which focuses only on what the object is 'in itself' (its shapes and colours, for example). But, apart from any other considerations, this overlooks the fact that works of art can be designed to serve non-artistic functions and that aesthetic admiration can encompass the appearance that an object presents of its suitability for discharging these functions. This is especially pertinent to works of architecture, for which harmony of form with function is an aesthetic merit.

So none of these suggestions, either in itself or combined with another, provides a definition of an attitude that satisfies the conditions required by the concept of the aesthetic attitude.

4 Transitive and intransitive particularity

Wittgenstein drew a distinction between a 'transitive' and an 'intransitive' use of the word 'particular'. You use the word in the transitive fashion if you use it as a prelude to a description, comparison or specification of the nature of the phenomenon you are referring to - a description that you intend to produce or, perhaps, that you wish to have produced by the person you are addressing. When you use the word intransitively it is not thereby your intention to follow it up with a specification of anything. Rather, you are using it, Wittgenstein says, as what might be called an 'emphasis': either it has some such force as 'strong', 'striking' or 'impressive', or it merely gives expression to the fact that your attention is taken up with the phenomenon you are indicating.

Richard Wollheim (1980) has argued that although there is such a thing as the aesthetic attitude, which he equates with the attitude of treating or regarding something as a work of art, philosophers of art who refer to the aesthetic attitude as a particular attitude are systematically ambiguous about whether they intend a 'particular attitude' in the transitive or the intransitive sense. And he claims that despite the many attempts to give a positive characterization of the aesthetic attitude, it can be conceived of as a particular attitude only in the intransitive sense. It might seem that this is tantamount to denying the existence of the aesthetic attitude or to asserting that there is nothing distinctive of it. But, Wollheim maintains, the point is rather that 'there need not be any comprehensive way of referring to what is distinctive of it other than as the aesthetic attitude'.

But Wollheim's position receives no support from Wittgenstein's distinction. For consider saying 'Jack has a particular way of asking a favour.' If you are using the word 'particular' transitively, then you will continue in some such fashion as 'namely, he drops his eyes and then looks to see how his request has been received.' If you are using the word in the intransitive sense, you will not continue in this fashion, for you are merely expressing the fixity of your attention on the way John asks a favour. However, it does not follow from the fact that your use is intransitive that there is no particular way in the transitive sense in which Jack asks a favour. Clearly, there is: Jack's way of asking a favour can be characterized in other terms. The same holds for the remark that the aesthetic attitude is a particular attitude. The fact that the word 'particular' is here being used intransitively does not preclude a positive characterization of the nature of the attitude. Hence no reason has been given for believing that what is distinctive of the aesthetic attitude cannot be captured other than by referring to it as the aesthetic attitude.

5 Myth or reality?

Is the aesthetic attitude a myth? That depends, first, on the difference between a single attitude and a motley collection of attitudes, and, second, on what work the aesthetic attitude is supposed to do.

An attitude towards an object is a disposition to think and feel about and to behave towards it in characteristic ways. You have a hostile attitude towards someone if you are liable to think hostile thoughts about them, to experience hostility when meeting them, to avoid their company, and, perhaps, to behave in ways that are harmful to them. Now your attitude towards an object can be such that you are disposed to thoughts about its aesthetic

character, to feelings aroused by its aesthetic qualities, and to aesthetically relevant behaviour. But there is no hope of circumscribing aesthetically relevant behaviour without using the concept of the aesthetic, as can be easily seen from the fact that it might on occasion consist in nothing more than walking away from an aesthetically uninteresting picture. And neither aesthetic thoughts nor aesthetic feelings have a nature that can be specified independently of the idea of the aesthetic. Accordingly, even if the diversity of aesthetic thoughts, feelings and behaviour does not imply that aesthetic attitudes are irreducibly diverse, their unity is achieved only by bringing them under the concept of the aesthetic. So the idea of the aesthetic attitude explicated in this way - as an attitude that must be thought of as a disposition to aesthetic thoughts, feelings and behaviour - is not suited to play the foundational role it is often assigned. It does not enable a penetrating analysis of the aesthetic, but must be defined in terms of it.

The prospects are bleak indeed for specifying in non-aesthetic terms an attitude that is both necessary and sufficient for aesthetic appreciation, as my critique of the standard definitions indicates; and the idea that there is an attitude that is distinctive of aesthetic interest but that is resistant to analysis is baseless. Hence, the aesthetic attitude is either a myth or of little interest.

See also: Aesthetic concepts

MALCOLM BUDD

References and further reading

Bullough, E. (1912) '"Psychical Distance" as a Factor in Art and an Aesthetic Principle', *British Journal of Psychology* 5: 87-118, reprinted in E.M. Wilkinson (ed.) *Aesthetics: Lectures and Essays*, London: Bowes & Bowes, 1957.(A celebrated account of the aesthetic attitude in terms of contemplation free from all practical concerns. Criticized by Dickie.)

Dickie, G. (1964) 'The Myth of the Aesthetic Attitude', *American Philosophical Quarterly* 1 (1): 54-65.(Criticizes a number of attempts to characterize the aesthetic attitude and concludes that it is a myth.)

Stolnitz, J. (1961) 'On the Origins of "Aesthetic Disinterestedness"', *Journal of Aesthetics and Art Criticism* 20: 131-43.(An interesting account of how the idea of disinterestedness, mentioned in §3, was introduced into aesthetic theory.)

Wittgenstein, L. (1960) *The Blue and Brown Books*, Oxford: Blackwell.(On page 158, the source of the distinction referred to in §4 between transitive and intransitive uses of 'particular'.)

Wollheim, R. (1980) *Art and Its Objects*, 2nd edn, Cambridge: Cambridge University Press, esp. §§40-2.(Presents the argument of §4 for the view that the aesthetic attitude is a particular attitude only in the intransitive sense.)

Aesthetic concepts

Aesthetic concepts are the concepts associated with the terms that pick out aesthetic properties referred to in descriptions and evaluations of experiences involving artistic and aesthetic objects and events. The questions (epistemological, psychological, logical and metaphysical) that have been raised about these properties are analogous to those raised about the concepts.

In the eighteenth century, philosophers such as Edmund Burke and David Hume attempted to explain aesthetic concepts such as beauty empirically, by connecting them with physical and psychological responses that typify individuals' experiences of different kinds of objects and events. Thus they sought a basis for an objectivity of personal reactions. Immanuel Kant insisted that aesthetic concepts are essentially subjective (rooted in personal feelings of pleasure and pain), but argued that they have a kind of objectivity on the grounds that, at the purely aesthetic level, feelings of pleasure and pain are universal responses.

In the twentieth century, philosophers have sometimes returned to a Humean analysis of aesthetic concepts via the human faculty of taste, and have extended this psychological account to try to establish an epistemological or logical uniqueness for aesthetic concepts. Many have argued that although there are no aesthetic laws (for example, 'All roses are beautiful,' or 'If a symphony has four movements and is constructed according to rules of Baroque harmony, it will be pleasing') aesthetic concepts none the less play a meaningful role in discussion and disputation. Others have argued that aesthetic concepts are not essentially distinguishable from other types of concepts.

Recently theorists have been interested in ways that aesthetic concepts are context-dependent - constructed out of social mores and practices, for example. Their theories often deny that aesthetic concepts can be universal. For example, not only is there no guarantee that the term 'harmony' will have the same meaning in different cultures: it may not be used at all.

1 Eighteenth-century views

Although questions about the origin and nature of our ideas of the beautiful, the proportionate, the harmonious, and so forth, can be found in ancient and medieval writings, the seat of aesthetic concepts developed in the eighteenth century was increasingly located primarily in human experience rather than in the objects of those experiences *per se*. Edmund Burke, for example, set out to explain where our ideas (and hence concepts) of the sublime and beautiful come from, and answered that they come not from objects *per se* but from objects as they are experienced (see Sublime, the §2). Thus one and the same object can be described as 'fearful' or 'sublime' depending upon the circumstances of the viewer. A building in flames is frightful if one is in the building and fears for one's life; it is sublime if one is standing across the street from it and has no fear that oneself or one's loved ones are endangered, and one can thus take pleasure (though Burke called it 'relative pleasure') in the shapes, colours and sounds, and so on. (see Burke, E.).

David Hume asserted that the existence of what are now generally referred to as aesthetic concepts depends upon human beings having the requisite psychophysical machinery for experiencing the world in certain ways. The locus of aesthetic experiences and ideas, he argued, is taste (see Artistic taste §1). This faculty not only accounts for the fact that one can claim that, say, a flower is beautiful (just as the faculty of sight accounts for the fact that one can claim that the flower is red), it also provides an empirical basis for what he calls the 'standard of taste'. Just as people with adequate colour vision and proper training will correctly apply colour concepts, that is, apply them in such a way that others with adequate vision and training will agree, so people whose taste is sufficiently sensitive and who have been correctly trained will competently apply aesthetic concepts, and will agree with other similarly competent judges.

Hume's approach exemplifies an urge to reconcile two counterintuitive forces that continue to appeal to both theorists and non-theorists in aesthetics. On the one hand there is the belief that aesthetic concepts are somehow subjective - that, as the saying goes, beauty is in the eye of the beholder. There is no guarantee that my concept of beauty will match yours, indeed there is ample evidence that one should not even hope for such agreement. Some like it cold, some like it hot, and any basis for a standard of application for evaluative concepts like

‘aesthetically good’ or ‘aesthetically bad’ seems ineluctably illusive. On the other hand, there does at least sometimes seem to be an objective basis for aesthetic concepts. We do seem to mean the same things when we talk about a beautiful sunset or suspenseful movie. If someone denied the grandeur of the Rocky Mountains, we would think less of the person, and not less of the mountains (see Hume, D.).

The drive to account for both subjectivity and objectivity is at the heart of perhaps the most systematic and influential treatment of the nature of aesthetic concepts in the history of western philosophy, namely that of Immanuel Kant. In his *The Critique of Judgment*, he provides a many-faceted definition of ‘the beautiful’ (as an exemplar of aesthetic concepts) that incorporates metaphysical, epistemological, psychological and logical analyses and which he believes reconciles the attractions of both subjectivity and objectivity (see Beauty §4).Kant, like Hume, agrees that aesthetic concepts are ‘taste concepts’, whose existence depends upon human experience. In particular, he believes that aesthetic concepts are grounded in pleasure and pain. Hence they are subjective, and any agreement between persons cannot be empirically based, as Hume had hoped. But,Kant argues, aesthetic concepts are pure in the sense that they are neither related to nor dependent upon cognitive or ethical concepts. Our feelings of pleasure or pain in the presence of an object have nothing to do with our scientific conception of the object (and are even independent of our belief that the object exists) nor with our belief that it is useful or morally good (see Art and morality §3). Since aesthetic concepts are not connected to anything special about me (what I as an individual believe about the nature of the object or its relation to my preferences or duties), I must believe that all human beings who similarly respond as human beings, and not as individuals with special histories, will react as I do - that they will similarly feel pleasure or pain in the presence of this object. Thus an aesthetic concept has a universal aspect. If I attribute beauty to a flower, I expect everyone else (to the extent that they respond aesthetically) to do so, too. Thus aesthetic concepts are subjectively located, but universally applicable, according to Kant.

2 Frank Sibley and his critics

In the twentieth century, the role of taste and the Kantian drive to distinguish aesthetic concepts from other kinds of concepts is exemplified in the writings of the influential British aesthetician, Frank Sibley. Taste, according toSibley, is a special mental faculty that enables some people to have certain perceptions and to form concepts on the basis of those perceptions. This special psychological faculty explains, he further believes, a special logic that characterizes the use of aesthetic concepts in discussion and argument.

If we observe people describing an event or object, according to Sibley, we notice that there are some features that everyone with normal eyes, ears and intelligence perceives - shape or loudness, for example. But there are also features that are perceived only by people with a special sensitivity - balance or unity, for example. These latter people are the ones who have taste. If a vase is gracefully curved, either one sees the gracefulness or one does not.

Sibley believes that this explains why aesthetic concepts are not condition-governed. That is, no list of non-aesthetic features (those perceivable by everyone) is logically sufficient for deducing that an object or event has any particular aesthetic features (those perceivable only by people with taste). Told that a vase is pink, made of glass, and fifty centimetres high, one is unable to conclude that the vase must be gracefully curved. And this is true, Sibley argues, no matter how long a list of non-aesthetic properties is provided.

None the less, Sibley argues, aesthetic concepts are objective; that is, ‘The vase is gracefully curved’ is either true or false. In this respect, aesthetic concepts are like colour concepts. Only people with adequate colour vision can see pinkness, but nevertheless we believe that the sentence, ‘The vase is pink,’ is either true or false. This is because people with normal colour vision agree about it. With respect to aesthetic concepts, agreement also provides the foundation of objectivity. People of taste agree that the vase is gracefully curved, and this is all we need to support the claim that aesthetic judgments have truth values.

Sibley’s view combines psychological, logical, epistemological and metaphysical components, and it has been criticized in all of these fields. It is often objected that ‘taste’ is a very unclear notion - certainly not adequate to support a unique logic of aesthetic concepts. Nor is it so easy to distinguish aesthetic from non-aesthetic properties (and hence the corresponding concepts). And many writers have argued that the analogy of aesthetic and colour concepts does not really support the sought-for objectivity of the former; there is nothing like the widespread

agreement that can be found for non-aesthetic qualities (on the assumption that they can be distinguished from aesthetic qualities) in the aesthetic realm. Disagreement among recognized experts in discussions of works of art is notorious.

None the less, many theorists agree with Sibley that there is something special about aesthetic concepts and that there are no 'aesthetic laws', that is, that there is no way of defining an aesthetic concept in terms of non-aesthetic concepts. Isabel Hungerland (1968) has described a distinction that she believes marks non-aesthetic concepts from aesthetic concepts: a seeming/being distinction. One can say that a person looks (seems) healthy but is not healthy, or that a house looks pink but is not pink. But this difference is absent in aesthetic attributions, she believes. If a vase looks gracefully curved, it is gracefully curved; if a voice sounds sweet, it is sweet.

Peter Kivy has objected to this way of distinguishing aesthetic from non-aesthetic concepts (1968). 'Unified', he argues, is surely an aesthetic concept; but it fails Hungerland's seeming/being test. A symphony may seem unified but may not really be unified. Kivy is also sceptical that there are no aesthetic concepts that can be reduced to non-aesthetic concepts. A very full description of a piece of music in non-aesthetic terms may lead one to conclude that the piece is unified. Thus there does not seem to be, for Kivy and others, a definitive way to distinguish aesthetic from non-aesthetic concepts.

3 Recent attempts to establish aesthetic realism

The view that there are no aesthetic laws connecting aesthetic and non-aesthetic concepts remains prevalent. Mary Mothersill, for example, argues that there are no principles or laws of taste and that everyone admits this when pressed (1984). None the less, aesthetic judgments are genuine judgments, that is, they have truth values and play a role in inference. They are open to serious question and debate (unlike, say, judgments about whether raw oysters taste good). They can be confirmed or disconfirmed by pointing to features of an object or event that are believed to cause pleasure, not just in the individual making the judgment but in other individuals who similarly investigate the object or event. Discussions that include aesthetic concepts are undertaken with the same hope of eventual agreement among persons of normal intelligence and interest as are discussions about many non-aesthetic concepts. There may be no reason to expect, as Kant believed, that everyone ought to find the same things beautiful; none the less we do expect that when we point to features or objects or events that we find beautiful, at least like-minded individuals will concur, according to Mothersill.

Another way theorists have tried to establish aesthetic realism is by arguing that aesthetic concepts connect to objective features of the world via a relationship known as 'supervenience' (see Supervenience). Even if aesthetic concepts may not be definable in terms of non-aesthetic concepts, if one can establish that the properties associated with these aesthetic concepts supervene on physical properties that exist in the world, then there is a foundation for the objectivity of aesthetic judgments. Philosophers have characterized supervenience in various ways; however, the important feature they all try to capture is the connection between a stable set of base properties and some property that seems to depend upon them, even though that dependence cannot be captured via a strict definition. It is possible to imagine two houses that share all base properties except that one house is yellow, the other not. 'Yellow' is not supervenient. But if both houses share all base properties then it is not possible for one to be beautiful, the other not beautiful. 'Beauty' is supervenient. As long as the base properties are stable and the meaning of 'beauty' remains the same, then it will be either true or false that anything with those base properties is beautiful. Thus a vase's graceful curves cannot be defined in terms of its angles, size, colour, material, and so forth, but its gracefulness does depend upon the particular physical properties that it possesses. If all the physical properties stay the same, so will all of its aesthetic properties, in this view. Thus aesthetic concepts are connected to the real world, and individual attributions of an aesthetic term can be explained or justified in terms of 'real world' properties.

But not everyone agrees that aesthetic concepts are illuminated via the notion of supervenience. For one thing, it is difficult to specify clearly what counts as a base property. Sibley, for example, thinks that colour is not an aesthetic property (for anyone with ordinary perceptive powers can see the colour of an object); others think that colour is obviously an aesthetic property. Furthermore, an essential assumption is that the meaning of 'beauty' be stable - and this is an assumption that has been seriously questioned. A building considered beautiful in one culture may not be considered beautiful elsewhere. Thus it appears that beauty is not supervenient, for it is possible for two houses to have all the same properties except that one is beautiful, the other not beautiful.

4 Aesthetic concepts as contextual constructs

If taste or some similar universal human propensity to have certain sorts of experiences in the presence of some objects or events is the foundation of aesthetic concepts, then one would expect that all people (at least all who share a certain degree of sensitivity) will form similar concepts. Beauty and ugliness, for example, should be concepts formed and sustained by persons across time and space. Just as colour concepts or concepts of heat and cold exist across cultures because human beings are physically and mentally constructed to experience variations in colour and temperature, so aesthetic concepts should exhibit cross-cultural similarity if we are physically and mentally constructed so as to experience variations, say, in proportion or rhythm.

Recently, however, a growing number of theorists have discussed ways in which aesthetic concepts may be socially constructed. The widely shared common-sense attitude that there is no fact of the matter with respect to aesthetic judgments (some people like opera, some rock music), and the relativism that often accompanies this attitude are more systematically articulated in aesthetic theories that ground aesthetic concepts in contextual features of the circumstances in which they arise.

Not only do people from different communities disagree about what is beautiful, moving or harmonious, communities do not universally share the same aesthetic concepts at all. In Japanese poetry, for example, *makoto* is a very important concern. No direct English translation of the word, however, is possible. At best one can give a rough gloss: sincere or genuine expression of an appropriate emotion. The word is most commonly used in discussions about haiku - a poetic form that also lacks a direct equivalent in English literature. We lack the word and hence the concept. R.G. Collingwood warned that we cannot describe African art in English without making it seem like English art and vice versa. Some ethnomusicologists observe that even as basic a concept as rhythm does not travel from Europe to Africa without significant alteration. One’s concepts are determined by the language that one speaks - and this language is shaped and determined by interests and attitudes that are culturally specific.

If aesthetic concepts are culturally determined, then understanding any particular concept will demand fluency in the culture in which it is operative. This is exactly what several contemporary theorists urge. Many feminists theoreticians, for example, criticize aestheticians like Kant and Sibley for assuming that they could speak with a ‘universal voice’. If Kant did not see a relation between moral and aesthetic concepts, it does not follow that they are separate in the experiences of all human beings. One’s gender, class, religion, economic or political status do often affect the way one forms and applies aesthetic concepts, they insist. If a person does not see what Frank Sibley sees, Sibley is not necessarily a more sensitive person.

Aesthetic concepts are learned in contexts where roles of performer, creator, audience, critic, tourist, and so forth are learned. These roles are culture-bound; indeed, some of them do not even exist in some communities. Some aesthetic responses may seem ‘natural’ (for example, all people seem to like watching sunsets), but many are the consequence of social prescription and proscription. Often sharing non-aesthetic values is a prerequisite of sharing aesthetic values, and hence of sharing aesthetic concepts.

See also: African aesthetics; Art criticism; Feminist aesthetics

MARCIA EATON

References and further reading

Brand, P. and Korsmeyer, C. (eds) (1990) ‘Feminism and Traditional Aesthetics’, special issue of *Journal of Aesthetics and Art Criticism* 48: 4.(Examples of feminist approaches to understanding aesthetic concepts; advanced reading.)

Burke, E. (1756) *A Philosophical Inquiry into the Ideas of the Sublime and the Beautiful*, Oxford: Blackwell, 1987.(Classic historical text.)

Cohen, T. (1973) ‘Aesthetic/Nonaesthetic and the Concept of Taste’, *Theoria* 39: 113-152. (A response to Sibley, for the advanced reader.)

Collingwood, R.G. (1938) *The Principles of Art*, Oxford: Clarendon Press.(Influential and readable work on the role of expression in explaining aesthetic concepts.)

Eaton, M.M. (1995) ‘The Social Construction of Aesthetic Response’, *British Journal of Aesthetics* 35 (2): 95-107.(Contextualist approach to aesthetic concepts, for the advanced reader.)

Hume, D. (1757) *Of the Standard of Taste and Other Essays*, ed. J.W. Lenz, Indianapolis, IN: Bobbs-Merrill,

1965.(A classic and highly readable historical text.)
Hungerland, I. (1968) ‘Once Again, Aesthetic and Non-Aesthetic’, *Journal of Aesthetics and Art Criticism* 27: 285-95.(Written as a response toSibley; contains advanced material.)
Kant, I. (1790) *The Critique of Judgment*, trans. J.C. Meredith, Oxford: Clarendon Press, 1973.(A classic historical text.)
Kivy, P. (1968) ‘Aesthetic Aspects and Aesthetic Qualities’, *Journal of Philosophy* 65 (4): 85-93.(One response toSibley’s view; an advanced text.)
Levinson, J. (1990) ‘Aesthetic Supervenience’, in *Music, Art, and Metaphysics*, Ithaca, NY: Cornell University Press.(An advanced discussion of supervenience of aesthetic properties.)
Mothersill, M. (1984) *Beauty Restored*, Oxford: Clarendon Press.(This advanced book presents a twentieth-century Kantian view.)
Petit, P. (1983) ‘The Possibility of Aesthetic Realism’, in E. Schaper (ed.) *Pleasure, Preference and Value: Studies in Philosophical Aesthetics*, Cambridge: Cambridge University Press, esp. 17-38.(Good discussion of problems in aesthetic realism; an advanced text.)
Sibley, F. (1959) ‘Aesthetic Concepts’, *Philosophical Review* 68: 421-50.(An influential and clearly written article shaping twentieth-century discussion of aesthetic concepts.)
Stahl, G. (1971) ‘Sibley’s "Aesthetic Concepts": An Ontological Mistake’, *Journal of Aesthetics and Art Criticism* 29: 385-90.(A response to Sibley’s 1959 article; for the advanced reader.)

Aesthetics

Aesthetics owes its name to Alexander Baumgarten who derived it from the Greek *aisthanomai*, which means perception by means of the senses (see Baumgarten, A.G.). As the subject is now understood, it consists of two parts: the philosophy of art, and the philosophy of the aesthetic experience and character of objects or phenomena that are not art. Non-art items include both artefacts that possess aspects susceptible of aesthetic appreciation, and phenomena that lack any traces of human design in virtue of being products of nature, not humanity. How are the two sides of the subject related: is one part of aesthetics more fundamental than the other? There are two obvious possibilities. The first is that the philosophy of art is basic, since the aesthetic appreciation of anything that is not art is the appreciation of it as if it were art. The second is that there is a unitary notion of the aesthetic that applies to both art and non-art; this notion defines the idea of aesthetic appreciation as disinterested delight in the immediately perceptible properties of an object for their own sake; and artistic appreciation is just aesthetic appreciation of works of art. But neither of these possibilities is plausible.

The first represents the aesthetic appreciation of nature as essentially informed by ideas intrinsic to the appreciation of art, such as style, reference and the expression of psychological states. But in order for that curious feeling, the experience of the sublime - invoked, perhaps, by the immensity of the universe as disclosed by the magnitude of stars visible in the night sky (see Sublime, the) - to be aesthetic, or for you to delight in the beauty of a flower, it is unnecessary for you to imagine these natural objects as being works of art. In fact, your appreciation of them is determined by their lack of features specific to works of art and perhaps also by their possession of features available only to aspects of nature (see Nature, aesthetic appreciation of).

The second fails to do justice to the significance for artistic appreciation of various features of works of art that are not immediately perceptible, such as a work's provenance (see Artistic forgery) and its position in the artist's oeuvre. A more accurate view represents the two parts of the subject as being related to each other in a looser fashion than either of these positions recognizes, each part exhibiting variety in itself, the two being united by a number of common issues or counterpart problems, but nevertheless manifesting considerable differences in virtue of the topics that are specific to them. In fact, although some issues are common to the two parts, many are specific to the philosophy of art and a few specific to the aesthetics of non-art objects.

Both works of art and other objects can possess specifically aesthetic properties, such as beauty and gracefulness. If they do possess properties of this sort, they will also possess properties that are not specifically aesthetic, such as size and shape. And they will be susceptible of aesthetic and non-aesthetic appreciation, and subject to aesthetic and non-aesthetic judgments. What distinguishes an item's aesthetic from its non-aesthetic properties and what faculties are essential to detecting aesthetic properties (see Aesthetic concepts)? What is the nature of aesthetic appreciation? It has often been thought that there is a particular attitude that is distinctive of aesthetic appreciation: you must adopt this attitude in order for the item's aesthetic properties to be manifest to you, and if you are in this attitude you are in a state of aesthetic contemplation (see Aesthetic attitude). This suppositious attitude has often been thought of as one of disinterested contemplation focused on an item's intrinsic, non-relational, immediately perceptible properties. But perhaps this view of aesthetic interest as disinterested attention is the product of masculine bias, involving the assumption of a position of power over the observed object, a reflection of masculine privilege, an expression of the 'male gaze' (see Feminist aesthetics §3). Another idea is that awareness of an object's aesthetic properties is the product of a particular species of perception, an idea which stands in opposition to the claim that this awareness is nothing but the projection of the observer's response onto the object (see Artistic taste).

An object's beauty would appear to be a relational, mind-dependent property - a property it possesses in virtue of its capacity to affect observers in a certain manner. But which observers and what manner? And can attributions of beauty, which often aspire to universal interpersonal validity, ever attain that status (see Beauty)? The great German philosopher Immanuel Kant presented a conception of an aesthetic judgment as a judgment that must be founded on a feeling of pleasure or displeasure; he insisted that a pure aesthetic judgment about an object is one that is unaffected by any concepts under which the object might be seen; and he tried to show that the implicit claim of such a judgment to be valid for everyone is justified. But how acceptable is his conception of an aesthetic judgment and how successful is his attempted justification of the claims of pure aesthetic judgments (see Kant, I.

§12)?

1 Aesthetics of art

Those questions that are specific to the philosophy of art are of three kinds: ones that arise only within a particular art form or set of related arts (perhaps arts addressed to the same sense), ones that arise across a number of arts of heterogeneous natures, and ones that are entirely general, necessarily applying to anything falling under the mantle of art.

Here are some of the most salient facts about art. Not everything is art. Artists create works of art, which reflect the skills, knowledge and personalities of their makers, and succeed or fail in realizing their aims. Works of art can be interpreted in different ways, understood, misunderstood or baffle the mind, subjected to analysis, and praised or criticized. Although there are many kinds of value that works of art may possess, their distinctive value is their value as art. The character of a work of art endows it with a greater or lesser degree of this distinctive value.

Accordingly, the most fundamental general question about art would seem to be: what is art? Is it possible to distinguish art from non-art by means of an account that it is definitive of the nature of art, or are the arts too loosely related to one another for them to possess an essence that can be captured in a definition (see Art, definition of)? Whatever the answer to this question may be, another entirely general issue follows hard on its heels. It concerns the ontology of art, the kind of thing a work of art is. Do some works of art fall into one ontological category (particulars) and some into another (types) or do they all fall within the same category (see Art works, ontology of)? And a number of other important general questions quickly arise. What is a work's artistic value and which aspects of a work are relevant to or determine this value? Is the value of a work of art, considered as art, an intrinsic or an extrinsic feature of it? Is it determined solely by the work's form or by certain aspects of its content - its truth or its moral sensitivity, for example? Can judgments about a work's artistic value justifiably lay claim to universal agreement or are they merely expressions of subjective preferences? And how is a work's artistic value related to, and how important is it in comparison with, other kinds of value it may possess (see Art, value of; Formalism in art; Art and truth; Art and morality; Schiller, J.C.F.)? What is required to detect the critically relevant properties of artworks, over and above normal perceptual and intellectual powers, and how can judgments that attribute such properties be supported (see Art criticism)? What kinds of understanding are involved in artistic appreciation, and must an acceptable interpretation of a work be compatible with any other acceptable interpretation (see Art, understanding of; Artistic interpretation; Structuralism in literary theory)? In what way, if any, does the artist's intention determine the meaning or their work (see Artist's intention)? What is an artist's style and what is its significance in the appreciation of the artist's work (see Artistic style)?

2 Aesthetics and the arts

One question that arises only for a small set of art forms concerns the nature of depiction. It might be thought that the analysis of the nature of depiction has no special importance within the philosophy of art, for pictorial representation is just as frequent outside as inside art. But this overlooks the fact that real clarity about the ways in which pictures can acquire value as art must be founded on a sophisticated understanding of what a picture is and the psychological resources needed to grasp what it depicts. So what is it for a surface to be or contain a picture of an object or state of affairs? Must the design on the surface be such as to elicit a certain species of visual experience, and must the function of the means by which the pattern was produced, or the intention of the person who created it, be to replicate features of the visible world? Or is a picture a member of a distinctive kind of symbol system, which can be defined without making use of any specifically visual concepts (see Depiction; Goodman, N. §2)? Another question that has a limited application concerns the distinctive nature and value of a particular artistic genre, the response it encourages from us, and the insight into human life it displays and imparts. For example, whereas a comedy exploits our capacity to find something funny, a tragedy engages our capacity to be moved by the fate of other individuals, and erotic art aims to evoke a sexual reaction; and this difference in the emotional responses at the hearts of the genres goes hand in hand with the different aspects of human life they illuminate (see Comedy; Emotion in response to art; Erotic art; Humour; Tragedy).

Questions about the individual natures and possibilities of the various arts include some that are specific to the particular art and some that apply also to other arts. On the one hand, relatively few art forms (architecture and pottery, for example) are directed to the production of works that are intended to perform non-artistic functions, or

are of a kind standardly used for utilitarian purposes, and, accordingly, the issue of the relevance to its artistic value of a work's performing, or presenting the appearance of performing, its intended non-artistic function satisfactorily is confined to such arts (see Architecture, aesthetics of). Again, only in some arts does a spectator witness a performance of a work, so that issues about a performer's contribution to the interpretation of a work or about the evaluation of different performances of the same work are limited to such arts (see Art, performing). And since only some works of art (novels, plays and films, for example) tell a story, and only some refer to fictional persons or events, questions about the means by which a story is told or how references to fictional objects should be understood have a restricted application within the arts (see Narrative; Fictional entities). On the other hand, most, if not all, arts allow of works within their domain being correctly perceived as being expressive of psychological states, and, accordingly, give rise to the question of what it is for a work to be expressive of such a condition (see Artistic expression). But the means available within the different arts for the expression of psychological states are various: poetry consists of words, dance exploits the human body, and instrumental music uses nothing other than sounds. And these different artistic media impose different limits on the kinds of state that can be expressed by works of art, the specificity of the states, and the significance within an art of the expressive aspects of its products (see Gurney, E. §2). Furthermore, it is a general truth about the various arts, rather than one special to expression, that what can be achieved within an art is determined by the nature of the medium the art is based on. Accordingly, an adequate philosophy of art must investigate the variety of such media and elucidate the peculiar advantages they offer and the limitations they impose (see Abstract art; Dance, aesthetics of; Film, aesthetics of; Hanslick, E.; Langer, S.K.K>; Lessing, G.E. §2; Music, aesthetics of; Opera, aesthetics of; Photography, aesthetics of; Poetry).

See also: Aesthetics, African; Aesthetics and ethics; Aesthetics, Chinese; Aesthetics in Islamic philosophy; Aesthetics, Japanese; Belinskii, V.G.; Metaphor; Rhetoric; Russian literary Formalism; Tolstoi, Count L.N.

MALCOLM BUDD

References and further reading

Hegel, G. (1835) *Aesthetics: Lectures on Fine Art*, trans. T.M. Knox, Oxford: The Clarendon Press, 1975. (Hegel's lectures on aesthetics, delivered in Berlin in the 1820s, are a classic introduction to the subject.)

Kant, I. (1790) *Critique of Aesthetic Judgement*, trans. W.S. Pluhar, Indianapolis, IN and Cambridge: Hackett Publishing Company, 1987.

Wollheim, R. (1980) *Art and Its Objects*, Cambridge: Cambridge University Press.(Essay considering the philosophical issues of works of art.)

Aesthetics and ethics

The contrast between ethical and aesthetic judgments, which has provided a good deal of the subject-matter of aesthetics, stems largely from Immanuel Kant's idiosyncratic view of morality as a series of imperatives issued in accordance with the dictates of practical reason, while for him judgments of taste are based on no principles. This has led even non-Kantians to argue that aesthetic judgments are primarily concerned, as is art itself, with uniqueness, while morality has mainly to do with repeatable actions. This tends to separate art from other human activities, a separation which was encouraged by the collection of useless items by 'connoisseurs', who took over as their vocabulary of appreciation the traditional language of religious contemplation. This viewpoint has been attacked passionately by idealist aestheticians, who claim that art is a heightening of the common human activity of expressing emotions, to the point where they are experienced and rendered lucidly, as they rarely are in everyday life. Marxist aestheticians, whose roots lie in the same tradition as idealists, argue that art is inherently political, and that the realm of 'pure aesthetic experience' is chimerical. Meanwhile the analytic tradition in aesthetics has spent much effort amplifying Kant-style positions, without taking into account their historical conditioning. There is a tendency to contrast the activities of the moralist, prescribing courses of action, with that of the critic, whose only job can be to point to the unrepeatable features which constitute a work of art.

1 Origins of the discussion

Ethics is the study of what people ought to do and the concepts employed in giving an account of why they should behave in certain ways; this study includes consideration of such factors as the nature of the judgments made about people's actions (are they objective or subjective?). Aesthetics is the study of a certain way of responding to nature and to some artefacts, pre-eminently works of art, and an investigation of the status of the judgments we make about them. At least, that is roughly how the subjects have been conceived traditionally. It is impossible to give a noncontroversial definition of either of them. That has consequences for the comparisons and contrasts between them which have been a feature of, at least, aesthetic discussion since Kant (see Kant, I. §12). He was the first great philosopher who gave extended and elaborate treatment to the nature of aesthetic judgment, though unfortunately that was as part of his attempt to overcome crippling difficulties in the rest of his system. Despite that, he largely set the agenda for subsequent discussion of aesthetic judgment, both its form and its content.

Though he did not place his thoughts on the subject in a historical context, it aids understanding if we do. Kant was writing in an intellectual climate in which natural science was increasingly taken to be the paradigm of human knowledge, with its picture of a wholly causally determined order (see Determinism and indeterminism §1). Kant accepted that account of the world of appearances (nature as we normally conceive it, including human nature), but argued that if we are to be moral agents we must be free in some strong (metaphysical) sense. For Kant, to be free is to be subject to a set of universal laws different to the descriptive laws of nature, namely those of practical reason, which are prescriptive. So as inhabitants of the phenomenal world we are intelligible according to the laws which govern everything else in it. As inhabitants of the noumenal world (the world of things-in-themselves) we are answerable to the moral law, alongside all other rational beings (see Kantian ethics).

This oppressively well-regulated view of things, in which both as objects and as agents we are to be regarded simply as law-governed items, not only has its own difficulties in reconciling the two kinds of being, but also leaves no room for taking an interest in something without seeing it as a member of a class (see Universalism in ethics §3). It is at this point that the aesthetic makes its debut, as a kind of interest in certain things, expressed in a kind of judgment, which is concerned with those things quite apart from any use to which they may be put, or even, Kant says, independently of whether they actually exist: an interest which is without, in the appropriate sense, interest.

2 Development of a contrast

It might be felt that, given the importance which Kant attached to the contrast between an aesthetic and any other kind of interest in things, it is surprising that it had not been made before. In fact it had, but in a different context, as we shall see. Furthermore, a set of pressures of a kind quite remote from anything Kant had in mind made such a contrast not only intelligible but urgent. With the growth of a merchant class intent on displaying its wealth, the purchase of useless objects which could be collected and displayed in private exhibitions became a major form of

consumption. Taste, in a new and extended sense, was necessary to make fine discriminations between superior and inferior specimens of the same kind of thing, where what made the crucial difference between good and bad became ever more recondite, requiring the activity of people called ‘virtuosi’ or, in subsequent gallicized form, ‘connoisseurs’. They employed a fairly rich evaluative vocabulary, but the justification for its use was to be found not in general rules, but in the possession by individual things of certain perceptual properties, properties discerned through taste. Thus on account of the colours which constitute them, and the balances between those colours, one judges whether paintings are unified, harmonious, and so on. The proper judgment of taste, as Kant was to put it, is always singular: ‘This is beautiful’.

The peculiar set of forces which drove Kant’s system, especially by the time he came to deal with beauty in the *Critik der Urteilskraft (Critique of Judgment)* (1790), can be ignored. Despite them he managed to exert a huge influence on the young subject of aesthetics, and to establish a contrast or set of contrasts with ethics which is still going strong, though there have been many opposing voices. As much as the welcome stress on the particular which aesthetic contemplation was alleged to certify, there was a vocabulary at hand which had seemed in danger of losing its employment: the vocabulary of theology, or rather that part of it which was concerned to stress the unique value of the object of devotion. Hence the striking similarity between the traditional terminology of worship and the new one of aesthetic absorption. It is no coincidence that for the first time we come across the phrase ‘the religion of art’. The phenomenon itself would have been inconceivable at an earlier date.

3 Aesthetics and history

Though philosophers are notoriously prone to self-consciousness about the nature of their subject, they have tended to regard the relationship between it and other disciplines in a less historically aware way than might have been expected. So the Kantian stress on the uniqueness of the aesthetic object has not often been subjected to historical scrutiny. This failure has resulted in a set of contrasts between the aesthetic and the ethical which have been taken as timeless truths about the two areas, rather than as the consequences of the triumphant progress of science and the decline of religion, among other factors. During the nineteenth century the conflicting, rather than complementary, tendencies to see art and aesthetic experience as *sui generis*, or as socially significant and morally and politically committed, simply confronted one another, strongly-worded manifestos appearing on each side, with little recognition of the elaborate background which had led to this polarization of attitudes.

The most striking contributions to aesthetics have, as a consequence, come from, on the one hand, idealist philosophers, such as R.G. Collingwood (1938), who have stressed the relationship between aesthetic and moral experience, and in general the connections between art and life, to the point where we are all seen as potential artists, though we rarely actualize our capacities (see Collingwood, R.G.; Art and morality §1). For a school of thinkers which lays great moral stress on self-realization, and sees the artistic enterprise as a matter of clarifying feelings, it is not surprising that morality finally comes to be seen as an art, if not as art (see Self-realization). On the other hand, major statements of a quite different hue have come from Marxist-oriented philosophers, above all some members of the Frankfurt School, such as Theodor Adorno, who have been at pains to point out both ‘the ideology of the aesthetic’, that is, the extent to which insisting that aesthetic experience is unrelated to other kinds is itself a political statement, and the general pervasiveness of politics, so that a major function of aesthetic theory is to bring to light the hidden ethical and political allegiances of varying aesthetic stances (see Frankfurt School).

Both these opposing schools derive in large measure from the inspiration of Hegel (§8); that is not surprising, given the stress in his own writings on aesthetics on the primacy of art rather than nature, the historical dimensions of artistic production and understanding, and the different functions which art serves in different societies, or the same society at different stages.

4 Aesthetics and ethics in analytic philosophy

The least rewarding incursions into aesthetic theory and its relationship to ethics have been by Anglo-American philosophers of the last forty years. They have tended to produce, no doubt unwittingly, parodies of Kantianism, in which the contrast between our responses to art and our moral judgments is rendered so extreme that it has been alleged by, for example, Hampshire (1952) and Strawson (1966) that the uniqueness of art-objects is such as to render them ineffable. These philosophers are so filled with abhorrence at the idea of using works of art for any purpose at all that they claim that their glory is their purposelessness, except perhaps for their unique rendering of

uniqueness. On this view, morality assimilates, art differentiates. Whereas the moralist has to recommend and to judge, the critic has only to point out the specific features of objects under contemplation. Aesthetics has as a large part of its point the indication of its own limited range: mainly to stress that it is not, at any crucial juncture, comparable to ethics. Unfortunately the tendency of this tradition of aesthetics to refrain from giving examples means that the reader is likely, when encountering a reference to ‘moral actions’, to think of keeping a promise, or something equally banal; whereas when the reference is to ‘a work of art’ the natural tendency will be to think of the Sistine Chapel ceiling, or a Shakespeare play. Hence it is not surprising that morality and art, ethics and aesthetics, are thought to be sharply divergent, and that the critic’s role is taken to be utterly dissimilar to the moralist’s.

Analytic aesthetics has been written almost exclusively from the spectator’s viewpoint. The same may be true of Hegel, but it did not influence the way he saw the subject as it has done recent aesthetics. For most of its history, writing in general terms about art (though it was not then called ‘aesthetics’) in fact consisted of discussions of the rules which, if not sufficient for the production of great art, were thought to be certainly necessary. Equally, a large amount of moralizing did not take the form of giving rules, but of holding up examples of the good life, of preaching ways of being rather than doing. In such a different climate of thought, the comparisons and contrasts to be drawn between the disciplines of aesthetics and ethics would be startlingly unlike those we have grown familiar with in recent thinking.

See also: Aesthetics, Chinese; Art and morality; Art, value of

MICHAEL TANNER

References and further reading

Abrams, M.H. (1985) ‘Art-As-Such: The Sociology of Modern Aesthetics’, *Bulletin of the American Academy of Arts and Sciences* 38: 8-33; repr. in *Doing Things with Texts*, New York: W.W. Norton, 1989.(An invaluable essay for showing the effect of general historical factors on the production of philosophical theory.)

Collingwood, R.G. (1938) *The Principles of Art*, Oxford: Oxford University Press. (The most impressive, though extreme, statement of an idealist aesthetic.)

Eagleton, T. (1990) *The Ideology of the Aesthetic*, Oxford: Blackwell.(A useful conspectus, from a Marxist standpoint, of the development of the notion of the aesthetic during the last two centuries.)

Hampshire, S. (1952) ‘Logic and Appreciation’, *World Review*; repr. in W. Elton (ed.) *Aesthetics and Language*, Oxford: Blackwell, 1959.(A classic statement of the starkest possible opposition between moral and aesthetic judgments.)

Kant, I. (1790) *Critik der Urteilskraft*, trans. W.S. Pluhar, *Critique of Judgment*, Indianapolis, IN: Hackett Publishing Company, 1987.(The basic work in the history of aesthetics, which has determined the way in which the issues have been discussed, as well as what they are.)

Mothersill, M. (1984) *Beauty Restored*, Oxford: Oxford University Press.(A full-length Kantian account of aesthetic judgment, full of resourceful and stimulating argument.)

Strawson, P.F. (1966) ‘Aesthetic Appraisal and Works of Art’, *Oxford Review* 3; repr. in *Freedom and Resentment*, London: Methuen, 1974.(A position very similar to Hampshire’s, with arguments supplied: it raises the question in the reader’s mind of why we should take any interest at all in works of art.)

Aesthetics in Islamic philosophy

The major Islamic philosophers produced no works dedicated to aesthetics, although their writings do address issues that contemporary philosophers might study under that heading. The nature of beauty was addressed by Islamic philosophers in the course of discussions about God and his attributes in relation to his creation, under the inspiration of Neoplatonic sources such as the pseudo-Aristotelian Theology of Aristotle*, a compilation based upon the* Enneads *of Plotinus. Considerations of artistic beauty and creativity were also addressed in works inspired by Aristotle's* Rhetoric *and* Poetics*, and Islamic philosophers also adapted some of Plato's views on literature and imitation, particularly those expressed in the* Republic.

On the whole, Islamic philosophers did not view artistic and literary creativity as ends in themselves. Rather, their interest was in explaining the relations of these activities to purely intellectual ends. In the case of poetics and rhetoric in particular, the emphasis in Islamic philosophy was pragmatic and political: poetics and rhetoric were viewed as instruments for communicating the demonstrated truths of philosophy to the populace, whose intellectual abilities were presumed to be limited. The medium of such communication was usually, although not necessarily, that of religious discourse. Islamic philosophers also devoted considerable attention to explaining the psychological and cognitive foundations of aesthetic judgment and artistic production within the spectrum of human knowledge. They argued that rhetoric and poetics were in some important respects non-intellectual arts, and that poetics in particular was distinctive in so far as it addressed the imaginative faculties of its audience rather than their intellects.

1 Beauty

Plotinus' *Ennead* V.8, 'On Intelligible Beauty', was the basis for the fourth chapter of the Arabic compilation known as the *Theology of Aristotle* (see Plotinus §§1, 7). Against the background of the discussion of beauty in this text, Islamic philosophers developed the theme of the differences between sensible and intelligible beauty; and the love and pleasure associated with each.

The notion of intelligible beauty is included in the discussion of the names and attributes of God contained in al-Farabi's *al-Madina al-fadila (The Virtuous City)* (see al-Farabi §2). Among the divine names al-Farabi lists 'beauty' (*al-jamal*), 'brilliance' (*al-baha'*), and 'splendour' (*al-zina*). Although the connotations of these terms are principally visual and hence sensible, al-Farabi argues that beauty in all things is primarily ontological: the more any being attains its final perfection, the more beautiful it is. From this he reasons that God, whose existence is most excellent, is the most beautiful of beings. Moreover, God's beauty surpasses all other beauty because it is essential, not accidental: the source of God's beauty is his own substance as defined by his self-contemplation, whereas created beauty derives from accidental and corporeal qualities that are not one with their own substances. Finally, al-Farabi argues that pleasure and beauty are intimately related, and that consequently God's pleasure, like his beauty, is beyond our comprehension. Pleasure is attendant upon the perception or apprehension (*idrak*) of beauty, and it increases in proportion to the beauty of what is perceived. Since God is the most beautiful of beings, and since his proper activity consists in an act of self-contemplation in which knower and known are completely one, the intensity and certitude of God's perception of his own beauty, al-Farabi reasons, must yield a pleasure of equal intensity. Moreover, since God's perception of his own beauty is the function of an eternal and uninterrupted act of contemplation, his pleasure, unlike ours, is continual rather than intermittent.

While al-Farabi's treatment of beauty in this context is principally an extension of his general account of divine transcendence and perfection along standard Neoplatonic lines, the development of the connection between beauty, perception and pleasure introduces a more properly aesthetic element into his account. Beauty in God, like beauty in the sublunar world, is found principally in things in so far as they achieve their proper perfection; when that beauty, be it sensible or intelligible, becomes an object of contemplation, it becomes in turn a source of pleasure for the one beholding it.

The contrast between sensible and intelligible beauty and the affective pleasures proper to each is developed in more detail in the *Risala fi al-'ishq (Treatise on Love)* by Ibn Sina. In the fifth chapter of this work, Ibn Sina discusses the youthful love of external, bodily beauty. He opens his discussion of the love of beauty with a consideration of four principles, three of which pertain to the psychology of the human soul. The first is based

upon Ibn Sina's characteristic view of the soul as a single substantial unity comprising a hierarchy of distinct powers. Either these powers can work together in harmony, in which case the lower will be ennobled by their cooperation with the highest faculty, that of reason, or the lower powers can rebel. These two possibilities are especially evident in the relations between reason and imagination (*al-takhayyul*) and the desires attendant upon them. The second principle is an elaboration upon the first: there are some human actions which pertain only to the bodily, 'animal' faculties within this hierarchy, including sensation, imagination, sexual intercourse, desire and aggression. Either these actions can be pursued in a purely animal fashion, or they can be transformed into something uniquely human under the guidance of reason.

Ibn Sina's third principle is that everything ordained by God has its own proper goodness and hence is the object of some legitimate desire; nonetheless, the lower desires can interfere with the higher, and thus their unlimited pursuit is to be avoided. Finally, Ibn Sina's fourth principle presents his definition of beauty in so far as it is the object of love for both the rational and animal souls: beauty (*al-husn*) consists in order (*al-nazm*), composition, (*al-ta'lif*) and symmetry (*al-i'tidal*). In the animal soul, this love of beauty is purely natural, arising either from instinct or from the simple pleasure of sensible perceptions. In the rational soul, however, love of beauty is more reflective, ultimately resting upon the recognition of the proximity of the beloved object to God, the First Beloved.

In applying these principles, Ibn Sina argues that there is what we might call an innate aesthetic sense implanted in every intellectual being (*al-'aqil*) which kindles in it a passionate desire for what is beautiful to behold (*al-manzar al-husn*). Despite the overall orientation of his discussion to the desire for the supra-sensible and purely intelligible beauty of God, Ibn Sina's remark here clearly pertains to the realm of sensible judgments. In fact, Ibn Sina even argues that such a desire for sensible beauty on the part of an intellectual being can be a noble thing, so long as the purely animal aspects of the desire are subordinated and the intelligible allowed to influence the sensible: such a purified aesthetic desire, according to Ibn Sina, results in a partnership (*al-shirka*) between the animal and rational souls. As evidence of this more general claim, Ibn Sina notes that even the most wise of humans can be preoccupied by a 'beautiful human form', and he implies that such a preoccupation is justified not only by the intrinsic aesthetic principles he has outlined, but also on the assumption that internal and external beauty and harmony mirror one another, unless the external beauty has been accidentally harmed or the internal character has been altered (for better or worse) by habituation. Finally, Ibn Sina also defends the desire for some sort of physical union with such a beloved, through kissing and caressing, although the expression of such an aesthetic impulse through sexual union is considered inappropriate except for the purpose of procreation, and where sanctioned by religious law.

2 Rhetoric and poetics

Most discussions of aesthetic themes by Islamic philosophers occur in the context of their considerations of the arts or rhetoric and poetics and the Aristotelian treatises devoted to these topics (see Aristotle §29). Following a practice established by the sixth-century Greek commentators on Aristotle, these treatises were classified by the Islamic philosophers as parts of Aristotle's logical corpus, the Organon (see Aristotelianism in Islamic philosophy). Thus the approach to these arts was not primarily aesthetic, but was focused on linguistic issues and the cognitive functions of rhetorical and poetic language. Rhetoric and poetics were classified as popular methods of instruction which produced less than certain states of belief in their audiences, who were assumed to be incapable of grasping the finer points of truly philosophical demonstration.

The Islamic philosophers did not explicitly limit the use of rhetoric and poetics to the spheres of religious discourse and political communication, however, and in their commentaries on Aristotle's *Poetics* some effort was spent on explaining the linguistic mechanisms whereby speech becomes figurative and metaphorical. Ibn Rushd in particular attempted to apply his understanding of Aristotle's views on poetics to the interpretation and criticism of Arabic poetry, and his *Talkhis kitab al-shi'r (Middle Commentary on the Poetics)* is full of citations of the works of well-known Arabic poets. Nonetheless, most of the interest taken by the Islamic philosophers in the arts of rhetoric and poetics stemmed from the foundations provided by these arts for explaining the relationship between philosophy and religion. The central books of al-Farabi's *Kitab al-huruf (The Book of Letters)*, along with Ibn Rushd's *Fasl al-maqal (Decisive Treatise)*, are devoted to this theme, which is nicely summed up in the following passage from al-Farabi:

> And since religion only teaches theoretical things by evoking imaginings and by persuasion, and its followers

> are acquainted with these two modes of instruction alone, it is clear that the art of theology which follows religion is not aware of anything that is not persuasive, and it does not verify anything at all except by persuasive methods and statements.
>
> (*Kitab al-huruf*: 132)

The use of the language of 'imaginings' and 'persuasions' indicates a reference to the cognitive aims that the Islamic philosophers traditionally ascribed to the arts of rhetoric and poetics. Religion is a reflection of and handmaiden to philosophy, dependent upon philosophy as a copy is dependent upon its original. In understanding religion as an imitation of philosophy, the Islamic philosophers were consciously evoking the background of Aristotle's *Poetics* and Plato's *Republic* and the aesthetic theories which they developed through a creative blending of the respective views of their two ancient sources on the nature of imitation.

3 Imitation and imagination

Ibn Sina's *Risala fi al-'ishq*, discussed in §1, contains elements of a theory of aesthetic judgment that is also developed, from a somewhat different perspective, in his discussions of the psychological underpinnings of the art of poetics. In these discussions, aesthetic judgments are attributed to the faculty of imagination (*al-mutakhayyila*) and the related internal sense faculties that formed a part of the Islamic Aristotelians' development of the concept of imagination (*phantasia*) found in Aristotle's *On the Soul* and *Parva naturalia*. In turn the notion of imitation or *mimēsis*, as found in Plato's *Republic* as well as in Aristotle's *Poetics*, was interpreted in terms of the functions of the imaginative faculty.

Al-Farabi, Ibn Sina and Ibn Rushd all identify the imagination as the faculty by which poets produce the figurative discourses proper to their art, and to which they appeal in their audience. These authors all contrast this use of and appeal to the imagination with the strictly intellectual and rational aim proper to all other modes of discourse and forms of reasoning. Al-Farabi's *Ihsa' al-'ulum (The Book of the Enumeration of the Sciences)* provides one of the most extensive descriptions of the character of poetic imagination. Two aspects of poetic statements are emphasized by al-Farabi: their representation of their subjects in terms 'more noble or more debased' than they actually are, and their ability to bring about an appetitive, as well as a cognitive, movement in the audience. That is, by depicting a subject in terms of images that evoke a loathsome object, the poet is able to make the hearers feel aversion to the thing depicted, 'even if we are certain that it is not in fact as we imagine it to be' (*Ihsa' al-'ulum*: 84). The reason for this aversion is directly linked to the poet's appeal to the imaginative faculty: 'for the actions of a human being frequently follow his imagination, more than they follow his opinion and his knowledge, because often his opinion or his knowledge are contrary to his imagination, whereas his doing of something is proportional to his imagining of it, and not to his knowledge or his opinion about it' (*Ihsa' al-'ulum*: 85).

A similar point is made by Ibn Sina in a number of texts. Ibn Sina frequently contrasts poetics with other modes of discourse by distinguishing the poet's attempt to produce an act of imagination (*takhyil*) in the audience with the more intellectual goal of seeking to produce an act of assent (*tasdiq*) to the truth or falsity of some claim. Ibn Sina, like al-Farabi, emphasizes the fact that such acts of imagination may often be contrary to what we know or believe to be the case, and he has a favourite example to illustrate this point: if someone tells us that 'honey is vomited bile', we are likely to lose our appetite for the honey before us, even if we are quite certain that the metaphor is literally false. Ibn Sina also echoes al-Farabi's claim that this ability of the imagination to affect our action is owing to the close link between the imaginative faculty and the appetitive motions of the soul.

The emphasis upon the imagination's ability to intervene in the soul's intellectual assent appears to have been directly linked by the Islamic philosophers to the theme of imitation. Al-Farabi, for example, appears to have made this connection in his *Ihsa' al-'ulum*, since he concludes his remarks on the poetic statement's ability to influence behaviour with the observation that this is 'what happens when we see likenesses imitative of the thing, or things resembling something else'. By the same token, throughout his *Talkhis kitab al-shi'r*, Ibn Rushd consistently interprets the Arabic term for *mimēsis* (*muhaka*) as equivalent to *takhyil*, the evoking of an image. And in several passages, Ibn Sina contrasts imaginative utterances which 'imitate one thing by another' with imaginative utterances that happen to be literally true as well. Generally, then, for the Islamic philosophers 'imitation' appears to refer to those specific acts of imaginative representation in which the object is depicted in terms not proper to it, or more specifically, which portray it as better or worse than its actual state. In this way, imitation is linked not

only or even principally to Aristotelian *mimēsis*, but rather to Plato's notion of imitation as it relates to the theory of the Forms found in the *Republic* (see Mimēsis; Plato §14).

This emerges clearly from a discussion in a little treatise by al-Farabi known simply as the *Kitab al-shi'r (Book on Poetics)*. In this treatise, al-Farabi identifies imitation, along with metric composition, as constitutive of the very substance of poetry, with imitation the most crucial of the two elements. In order to explain the nature of poetic imitation, which occurs through language, al-Farabi draws heavily upon its similarities to imitation through action, for example, in the making of statues or in performative imitations. Here too imitation is said to have as its end to 'cause an imagining' of the imitated object, either directly or indirectly. The difference between direct and indirect imitation refers to the distance that separates the representation of the object from the reality itself, as illustrated in the example of a statue. For if an artist wished to imitate a person named Zayd:

> … he might make a statue which resembles him, and along with this make a mirror in which he sees the statue of Zayd. And it might be that we would not see the statue itself, but rather the form of his statue in the mirror. And then we would know him through what imitates an imitation of him, and thus be two degrees removed from him in reality.

(*Kitab al-shi'r*: 94-95)

The possibility of degrees of removal from the original is highly evocative of Plato's description of the possible states of removal from the Forms in the myth of the cave. Al-Farabi believes this possibility holds not only for artistic imitation, but also for linguistic imitation in poetry. While these associations are sometimes viewed pejoratively by the Islamic philosophers, as one might expect in the light of their Platonic resonances, this attitude is not universal. Al-Farabi himself reports noncommittally that many people consider the more remote imitation to be the more perfect and artistic, and here as in his other works he admits the power of imitative utterances for inciting humans to actions to which intellectual opinion or knowledge fail to move them.

It is Ibn Sina (§8), however, who goes furthest in eliminating the negative overtones of these descriptions of poetic speech. In all but his most youthful writings, Ibn Sina emphasizes that the poet's concern with the imagination requires that his work be judged on its own terms and not on the level of intellectual judgments. Strictly speaking, poetic imaginings are neither true nor false; but in so far as poetic statements may imply corresponding intelligible propositions, they may possess a truth-value incidentally and secondarily. For this reason, although many will remain literally false, this need not be universally the case:

> And in general poetic [syllogisms] are composed of premises which evoke images… be they true or false. Generally they are composed of premises to the extent that they possess a figure and a composition which the soul receives by means of what is in them of imitation and even of truth; for nothing prevents this [that is, their being true].

(*al-Isharat wa-'l-tanbihat*: 80-1)

By the same token, Ibn Sina also allows for the use of poetic and imaginative discourse that is ethically neutral, seeking neither to ennoble nor to debase what is imitated, but rather merely aiming to 'provoke wonder through the beauty of the comparison' and thus to fulfil what could be termed a purely aesthetic end.

See also: Aesthetics; Aristotle §29; Beauty; al-Farabi; Ibn Rushd; Ibn Sina; Imagination; Mimēsis; Plato §14; Platonism in Islamic philosophy; Poetry; Rhetoric

DEBORAH L. BLACK

References and further reading

Black, D.L. (1990) *Logic and Aristotle's 'Rhetoric' and 'Poetics' in Medieval Arabic Philosophy*, Leiden: Brill. (Discusses the interpretation of these Aristotelian texts as works of logic; includes considerations of the themes of imagination and imitation.)

al-Farabi (*c*.870-950) *al-Madina al-fadila (The Virtuous City)*, ed. and trans. R. Walzer, *Al-Farabi on the Perfect State*, Oxford: Clarendon Press, 1985.(Text with facing translation of *al-Madina al-fadila*; includes detailed notes regarding al-Farabi's Greek sources and antecedents.)

al-Farabi (*c*.870-950) *Kitab al-shi'r (Book on Poetics)*, ed. and trans. A.J. Arberry, 'Farabi's Canons of Poetry', *Rivista degli Studi Orientale* 17 (1938): 267-78; ed. M. Mahdi, *Shi'r* 3 (1959): 91-6.(A curious little text

presenting al-Farabi's understanding of Greek poetics.)
al-Farabi (*c*.870-950) *Ihsa' al-'ulum (The Book of the Enumeration of the Sciences)*, ed. U. Amin, Cairo: Librairie Anglo-Égyptienne, 3rd edn, 1968.(Al-Farabi's discussion of different kinds of knowledge.)
al-Farabi (*c*.870-950) *Kitab al-huruf (The Book of Letters)*, ed. M. Mahdi, Beirut: Dar el-Mashreq, 1969. (Al-Farabi's account of the nature of logic and languages.)
Heath, P. (1992) *Allegory and Philosophy in Avicenna (Ibn Sina)*, Philadelphia, PA: University of Pennsylvania Press.(Ibn Sina's theories on allegory in the context of his philosophy as a whole; aimed at the non-specialist in philosophy and useful for audiences with primarily literary interests.)
Ibn Rushd (*c*.1174) *Talkhis kitab al-shi'r (Middle Commentary on the Poetics)*, trans. C.E. Butterworth, *Averroes' Middle Commentary on Aristotle's 'Poetics'*, Princeton, NJ: Princeton University Press, 1986.(A translation of Ibn Rushd's major work on this topic, with a helpful introduction.)
Ibn Rushd (*c*.1179-80) *Fasl al-maqal (Decisive Treatise)*, trans. G.F. Hourani, *Averroes on the Harmony of Religion and Philosophy*, London: Luzac, 1961.(Translation of Ibn Rushd's analysis of the links between religion and philosophy.)
Ibn Sina (980-1037) *al-Isharat wa-'l-tanbihat (Remarks and Admonitions)*, ed. J. Forget, Leiden: Brill, 1892; part translated by S.C. Inati, *Remarks and Admonitions, Part One: Logic*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1984.(The sixth and ninth 'methods' of this text discuss rhetoric and poetics.)
Ibn Sina (980-1037) *al-Shifa' (Healing)*, *Kitab al-shi'r*, trans. I.M. Dahiyat, *Avicenna's Commentary on the 'Poetics' of Aristotle*, Leiden: Brill, 1974.(Translation of the *Poetics* section of Ibn Sina's encyclopedic work, *al-Shifa'*, with excellent introductory essays; aimed at students of literary theory.)
Ibn Sina (980-1037) *Risala fi al-'ishq (Treatise on Love)*, trans. E. Fackenheim, 'A Treatise on Love by Ibn Sina', *Mediaeval Studies* 7 (1945): 211-28.(A translation of the *Risala fi al-'ishq*.)
Kemal, S. (1991) *The Poetics of Alfarabi and Avicenna*, Leiden: Brill.(Various aspects of these two philosophers' views on poetics.)
Kemal, S. (1996) 'Aesthetics', in S.H. Nasr and O. Leaman (eds), *History of Islamic Philosophy*, London: Routledge, ch. 56, 969-78.(Account of some of the main concepts of aesthetics, along with the leading controversies of the classical period.)

Aesthetics, African

The study and analysis of African art and aesthetics have been dominated by Western culture. Initially the aesthetic sensitivities of African cultures were characterized as 'primitive' and of low intellectual calibre. Africans reacted to such negative stereotyping by articulating their own, deliberately non-Western aesthetic theories. The best known of these is négritude.

With its emphasis on learning about a culture by living in it for a period of time, anthropology encouraged scholars to relate African art directly to the aesthetic values of the cultures that produced it. This kind of contextual approach has also become the special concern of African art historians.

One can exemplify the exploration of the aesthetic conceptions of a particular culture in this way by considering the case of the Yoruba peoples of southwestern Nigeria. The Yoruba have a detailed and refined aesthetic vocabulary that has been subjected to extended description and analysis. Where human beings are concerned the highest form of beauty is attributed to a person's good moral character. Where objects are concerned beauty is influenced by their utility or, in the case of figurative carvings, by the intelligence and ability of the artist.

1 Primitive cultures and *négritude*

The history of Western art and aesthetics might have been different if the founding fathers of these disciplines had been natives of a non-Western culture. The artistic genres and aesthetic standards imputed to Western culture by these experts might have been spurned as mistaken, alien, or pejorative. They might be said to have been prime examples of an ethnocentrism which imposes the aesthetic standards of one culture upon others. They might have succeeded in crossing cultures in order to identify some Western aesthetic sensitivities, about which they may have voiced some derogatory remarks and attributed their origins to something like a 'primitive' or 'child-like' mentality.

Each of these possibilities evokes unpleasant memories of scholarly theses that were enunciated when African and Western cultures interacted over the issue of aesthetic sensitivities, in particular those said to inform the genres of artefact that Western art collectors and museums began to define as constituting African art. The bulk of material published in this field is the product of Western scholarship. This scholarship began by denying African intellectual capabilities the same level of aesthetic consciousness that was embraced by the West. African aesthetic sensitivities were said to be primal, inarticulate, collective and instinctive. Africans were thought to be emotional rather than intellectual. As a result the artefacts produced by their cultures were said to express and represent semiconscious hopes and fears about the vaguely understood forces that controlled tribal society and the natural world. 'Primitive' art was conceived of as the product of 'primitive' understanding.

Eventually this emotional and expressive, as opposed to rational and intellectual, characterization of African personality and culture was adapted and reformulated as a positive, lyrical and intellectually refined aesthetic by indigenous African scholars. Termed *négritude*, this school of thought whose most prominent exponent was the sometime president of Senegal Leopold Sédar Senghor, rebuked Western scholarship for failing to assign African aesthetic sensitivities separate but equal status *vis-à-vis* their Western equivalents. It also inspired a renewed sense of black artistic merit that led to a flowering of modern African art in such diverse fields as drama, poetry, fiction, fashion and dance. Theoretical elements of *négritude* have also been adopted by the cultural and intellectual movements identified with Afrocentrism (see African philosophy, Francophone §4).

Nevertheless, as far as the Western-engendered collection and study of African art was concerned, the predominant interest and emphasis remained with mainly figurative, carved sculpture produced prior to the European colonization of Africa during the nineteenth century. Such pieces were viewed as comparatively 'authentic'. However, this term became associated with exaggerated Western notions of the semibarbaric, isolated, expressive tribal cultures that worshipped them, made sacrifices to them and created them out of an imagination that had not distinguished the rational from the emotional, the scientific from the superstitious.

2 Anthropology and African art history

It was modern twentieth-century Western anthropology which renewed interest in the identification, study and

appreciation of the intellectual dimension to African aesthetic values. African art objects, whose aesthetic properties had previously been assessed in virtual ignorance of the specific cultures that produced them (as expressions of a generalized, primitivized mentality) began to be culturally recontextualized on the basis of tribal attributions. Even if African art and aesthetics were not anthropological priorities, the general emphasis upon reintegrating objects and beliefs with the social and cultural contexts that originally produced them eventually reignited interest in indigenous artistic traditions and aesthetic values. Such renewed interest was underscored by the importance of doing fieldwork in Africa.

One controversial issue that remained unresolved was the degree to which traditional Africans were capable of articulating whatever aesthetic values informed their cultures. The point was whether Africans themselves could be relied upon to articulate their aesthetic values or whether (usually foreign) anthropologists would have to intuit them on their behalf. Some scholars continued to argue that much of whatever constituted African culture remained at a preverbal level. Whatever aesthetic standards there were implicit in inherited traditions and traditions governed cultures as rules rather than as reasons. Discussion and innovation were not priorities. The forms of figurative sculpture and patterns of textiles were inherited from a tribal past, replicated in the tribal present and passed on unaltered to the tribal future. In this intellectual atmosphere of elementary discursiveness there was no need to articulate or discuss fundamental aesthetic values. Aesthetic standards need not involve more complex considerations than how to polish an object, or colour a textile.

Increased interest in the social and cultural forces responsible for African art and in the aesthetic values that governed its creation has led to the emergence of a separate and specialized discipline known as African art history. In the process of disengaging itself from traditional anthropological interests and methods, this discipline has come to focus on the artistic and aesthetic in African cultures. Further social and cultural research has been encouraged to illuminate different African artistic and aesthetic elements. The scope of African art and aesthetics has been broadened beyond the objects found in private collections and museums so as to include architecture, bodily adornment and jewellery, as well as craftwork and oil and watercolour paintings. The technical aesthetic vocabulary and standards applied by Africans to the artistic elements of their cultures have been detailed. The peoples of Western cultures have been reminded that their aesthetic sensitivities and artistic preferences are products of Western acculturation and may therefore cause them to perceive the aesthetic in a different manner from individuals who are the product of an African culture. In addition, the level of cross-cultural critical appreciation of African art and aesthetics is in the process of being improved so that its technical, formal and intellectual properties may be better interrelated and appreciated. In this way African art history as subject matter can be accorded separate but equal status alongside the aesthetic traditions and heritages of other geographical and art historical groupings.

3 Yoruba moral discourse

To the philosopher of language the extent to which members of African cultures can articulate their aesthetic values is an important consideration. It seems an unnecessarily indirect and complicated strategy to arrive at the exegesis of any African aesthetic vocabulary by approaching it via the eclectic variety of objects that Western collectors and museums have made into African art. It would make better sense to begin the study of the aesthetic of any African culture with the terminology used by people of that culture. Such an approach may not provide information about whatever technical vocabulary might be employed by professional artists and critics in such a society. But it might be an essential foundational exercise preliminary to the identification and analysis of any more technical terminology.

The Yoruba of southwestern Nigeria have a subtle and systematic aesthetic vocabulary evidenced by ordinary discourse (see Yoruba epistemology). Yoruba aesthetic discourse is intimately related to Yoruba moral discourse. Moral discourse is grounded in epistemological, or cognitive, considerations as one of the fundamental issues is how a person can know the moral character, or *iwa*, of another. The Yoruba argue that since consciousness - *okon* - is private, behaviour - *isesi* - is the only basis on which the moral character of other people can be judged.

The Yoruba define behaviour by the actions, or *ise*, and words, or *oro enu*, of a person. In philosophical terminology this is conventionally referred to as nonverbal and verbal behaviour. Since the Yoruba equate knowledge, or *imo*, with what a person witnesses at first hand, there are severe strictures governing the attribution of either words or actions not experienced in this way.

Such strictures are predictable in an oral culture where literacy cannot be taken for granted. Their point seems to be as epistemic as it is moral. Just as people in Western cultures are concerned with exercising control over the media, members of Yoruba culture wish to exercise control over the mouths constituting the media in oral culture. False claims of knowledge of technical problems, or the moral character of people can endanger lives and destroy reputations.

4 Yoruba aesthetic discourse

Morality is linked with the aesthetic in everyday Yoruba discourse. This is because the Yoruba believe that the purest or highest form of beauty, or *ewa*, in humans is a good moral character, or *iwa rere*. They appreciate that in purely physical terms one person may be more attractive than others. As this is a matter of chance rather than choice, it is of superficial importance when it is a question of a person whose moral character can be relied upon in any situation. This means that a person who happens to be physically unattractive may still be deeply admired and praised for the beauty of their words and actions. Conversely a person of remarkable physical beauty may come to be regarded as viciously immoral on the basis of their words and actions. This preference for a beauty, or *ewa*, that is 'moral' or 'inner' is summed up by the Yoruba aphorism '*Iwa l'ewa*', meaning 'good moral character is beauty'.

Viewing Yoruba aesthetic discourse from this perspective reminds us that the Yoruba, like many other peoples, spend most of their time talking about 'beauty' with regard to humans rather than to the kinds of African art objects of special interest to collectors and museums. The Yoruba also attribute beauty to the natural world and to human artefacts, although different criteria are employed for measuring it.

With reference to the natural world a thing may be appreciated solely for its physical beauty, as might be the case with a glorious sunset. A natural object may be described as having both beauty (*ewa*) and character (*iwa*) if it is also useful to humankind in the sense of an attractive tree that also provides edible fruit, or a splendidly feathered chicken that also provides eggs and the main course for supper. In other words both beauty and character are most importantly assigned to natural objects by measuring their utility and usefulness to humankind.

With reference to human artefacts, the Yoruba may refer to a well-maintained agricultural farm, a new piece of furniture, or an attractive woven textile as possessing beauty. As with natural objects, the human artefacts may also be said to have character arising from their utility in the sense of a new chair that is comfortable and sturdy and the cloth that is durable as well as attractive.

Figurative sculptures, or carvings of human beings, also fall within the category of human artefacts. However, a sculpture of a human is not human. It cannot say or do and does not behave in any conventional human sense. This means that it cannot have moral character as this arises from verbal and nonverbal behaviour, or *isesi*. Yoruba appreciation of this is indicated by the fact that such objects are referred to as having exterior or bodily (*ara*) beauty (*ewa*). Such beauty takes into account factors such as form, polishing and colouring.

In recent years scholars have sought to identify the more specialized vocabulary by means of which professional artists and connoisseurs in Yoruba society articulate the formal principles and values that define and govern art work in the indigenous culture. The main aims are to elaborate whatever indigenous aesthetic consciousness may inform many of the objects the West has come to regard as African art, as well as to illuminate the many other decorative and performative practices, such as fashion, masquerade, poetry and song that are often of greater aesthetic importance and more enduring value to indigenous Yoruba culture. What is most striking about this more specialized vocabulary is the degree to which it replicates values affirmed by the epistemology and morality. Unlike Western culture, where the importance and relevance of aesthetic theory to artistic practice has declined, the relationship between the aesthetic and the artistic in Yoruba culture appears to be fundamental.

There is an emphasis upon the artist being calm, controlled and reasonable (*ifarabale*). The artist must also possess the aptitude for clear observation, understanding and expression (*iluti*). These essential foundational elements of the aesthetic consciousness must be intact in order to show sensitivity to the subjects portrayed and the audience reached. The Yoruba emphasize that the more narrow and conventional (at least in Western terms) 'artistic imaginative insight', or *oju inu*, which enables artists to 'picture' the forms they seek to instantiate, along with design components, or *oju ona*, such as colour, substance, rhythm, outline or harmony, can be exemplified in a superior manner only if based upon these foundational elements.

Previously published accounts of a Yoruba aesthetic by professional art historians do not present it as a single component belonging to an integrated system of epistemic, moral and aesthetic values. This may be attributed to a variety of influences, such as the vested interests of dealers and private collectors of African art whose focus is on the objects, the vestiges of the false stereotype that Africans are less capable of expressing themselves in rational and systematic terms, and the lack of attention paid by academic philosophers to semantic networks of meaning which underlie African discourse.

There is no obvious reason why the methodological approach illustrated by this account of a Yoruba aesthetic could not be applied to the discourse of other African cultures. Relatively few systematic studies of the values underlying, informing and interrelating epistemological, moral and aesthetic priorities and practices in African cultures have so far been undertaken. Even if their discourse embraces different epistemic standards and alternative moral virtues leading to other aesthetic values, the results could be of interest. In addition, such narratives would help to demonstrate that the museum objects which for many constitute African art may only occupy a minor and subsidiary role in the overall aesthetic that defines such cultures.

See also: Aesthetic concepts; Aesthetics, Chinese; Aesthetics, Japanese

BARRY HALLEN

References and further reading

Abiodun, R., Drewal, H. and Pemberton, J. (eds) (1990) *African Art Studies: the State of the Discipline*, Washington, DC: Smithsonian Institution.(A collection of clear assessments by African art historians, detailing the accomplishments and shortcomings of their discipline. The contribution by R. Abiodun on Yoruba aesthetic terminology is noteworthy.)

Abiodun, R., Drewal, H. and Pemberton, J. (eds) (1994a) *Yoruba Art and Aesthetics*, Zurich: The Centre for African Art and the Reitburg Museum.(A well-illustrated introductory discussion of Yoruba art and aesthetics, with some discussion of their relationships to Yoruba religion.)

Abiodun, R., Drewal, H. and Pemberton, J. (eds) (1994b) *The Yoruba Artist: New Theoretical Perspectives on African Arts*, Washington, DC and London: Smithsonian Institution Press. (A critical, innovative and multidisciplinary review of the verbal and visual arts as interpreted by academic disciplines.)

Appiah, K.A. (1982) 'The Post-Colonial and the Postmodern', in *In My Father's House: Africa in the Philosophy of Culture*, London: Methuen; 2nd edn, Oxford: Oxford University Press, 1992, 137-57.(A study of the dangers inherent in discussing art as a cross-cultural phenomenon with particular reference to Africa.)

Appiah, K.A. (1995) 'Why Africa? Why Art?', in T. Phillips (ed.) *Africa: the Art of a Continent*, Munich and New York: Prestel, 21-6.(This brief discussion makes it apparent why even the most obvious artistic cross-cultural generalizations may be misguided.)

Blier, S.P. (1987) *The Ontology of Architecture: Ontology and Metaphor in Batammaliba Architectural Expression*, Cambridge: Cambridge University Press.(A detailed and well-illustrated study of the aesthetic values operative in a subject area largely ignored by canonical African art history.)

Boone, S.A. (1986) *Radiance from the Waters: Ideals of Feminine Beauty in Mende Art*, New Haven, CT and London: Yale University Press.(A study of the interplay between human and artistic aesthetic priorities and values.)

Clifford, J. (1988) *The Predicament of Culture: Twentieth Century Ethnography, Literature and Art*, Cambridge, MA: Harvard University Press.(A useful critique of Western cultural values and their consequences for the appreciation of African art and aesthetic values.)

Hallen, B. (1997) 'African Meanings, Western Words', *The African Studies Review* 40 (1): 1-11. (Reflections on the difficulties facing African art historians when they translate African aesthetic terminology into Western languages.)

Irele, A. (1990) *The African Experience in Literature and Ideology*, Bloomington, IN: Indiana University Press.(A cultural and historical analysis of *négritude* as an African ideological movement.)

Lawal, B. (1974) 'Some Aspects of Yoruba Aesthetics', *The British Journal of Aesthetics* 14 (3): 239-49.(An introductory synopsis of everyday Yoruba vocabulary and social conventions relevant to aesthetics.)

Roberts, M. and Roberts, A. (eds) (1996) *Memory: Luba Art and the Making of History*, New York: The Museum for African Art; Munich: Prestel.(An innovative collection that stresses the flexible role of the arts in both recording and interpreting the history of an African ethnic group.)

Thompson, R.F. (1983) *Flash of the Spirit: African and Afro-American Art and Philosophy*, New York: Random House.(An exploratory study of African aesthetics and the diaspora.)

Van Damme, W. (1987) *A Comparative Analysis Concerning Beauty and Ugliness in Sub-Saharan Africa*, Ghent: Rijksuniversiteit.(A pioneering first step towards comparing aesthetic priorities from diverse west, central and southern African languages.)

Aesthetics, Chinese

*In China, poetry, painting and calligraphy are traditionally known as the 'Three Perfections' of the cultivated scholar. They are construed as ethico-aesthetic acts of self-signification and are evaluated as to their efficacy in fostering harmonious relations of social exchange within the concrete circumstances of particular social contexts. In contrast to Western notions of mimesis, the Chinese poetic tradition assumes the existence of fundamental, mutually implicating correlations between the patterns (*wen*) immanent in nature and those of human culture.*

*This gives rise to two traditions of Chinese poetics. First, there is the canonical tradition of Confucian exegesis, in which a poem was assumed to invoke a network of pre-established categorical correlations (*lei*) between poet and world, which enabled the imagery to be read as verbal indices of both personal feeling and the relative stability of the social and natural order. Second, there is the non-canonical tradition of neo-Daoist and Buddhist-inspired poetics which represented a shift from the didactic to the affective power of natural imagery to make reference to the poet's state of mind.*

Calligraphy and painting were adopted by the gentleman-scholar as ethico-aesthetic practices of xiushen *(self-cultivation) and self-expression, and for promotion of social exchange. Early writings describing calligraphy and painting deploy metaphorical imagery that makes reference to both nature and the body. This imagery invoked the indigenous correlative rhetoric that sought consonance between the patterns immanent within the natural order and those of the human realm. The embodiment of tradition, through the practice of making artistic references to the past, was fundamental to the art of the scholar-painter, for it served to establish one's artistic lineage and to sanction or authorize one's own self-presentation within a particular historical situation.*

1 The ethico-aesthetic Way of Chinese art

In China, poetry, painting, and calligraphy are traditionally known as the 'Three Perfections' (*sanjue*) of the cultivated scholar. They were so designated because each of these modes of aesthetic practice came to serve as an important means of 'self-cultivation' (*xiushen*) as well as self-expression, and often appear combined in a single work of the scholar-artist.

According to tradition, poetry, painting and calligraphy originally stem from a common root-metaphor, namely *wen*. In the first Chinese etymological dictionary, Xu Shen's *Shuowen jiezi (Explanations of Simple and Compound Characters)*, *c.* AD 100, the character *wen* is said to consist of 'intersecting strokes, representing a criss-cross pattern'. *Wen* is a semantically multivalent term that can refer to physical markings, patterns on coloured woven silk and painted designs of carriages as well as writing, literature and culture. The aesthetic conception of *wen* is given by Liu Xi (early third century AD) in his lexicographic work *Shiming (Explanations of Names)*: 'Wen means assembling various colours to form brocade or embroidery, assembling words to form phrases and meanings like patterned embroidery' (Liu 1975: 101).

The Western word that comes closest in meaning to *wen* is *textus*, Latin for 'woven', from which the words 'text' and 'textile' are derived. However, *textus* does not connote the cosmological, and thus ethical, implications of *wen*. These can be traced back to the legendary accounts of the sage-king Bao Xi or Fu Xi, recorded in the *Dazhuan (Great Commentary)* to the early divination text, the *Yijing (Book of Changes)*. Contemplating the images (*xiang*) in the heavens above and the patterns (*wen*) on earth below, Bao Xi invented the *bagua* or eight trigrams, traditionally regarded as the prototypes of the Chinese scripts, the very foundation of Chinese civilization.

It is precisely through a patterning of human relations in correlation with the configurations immanent in natural phenomena that the ethical implications of *wen* are realized: 'Contemplate the configurations [*wen*] of heaven to observe the changes of seasons; contemplate the configurations of man to accomplish the [cultural] transformation of the world' (Liu 1975: 18-19). It is in light of this that we can appreciate the Chinese word for 'civilization', *wenhua*, which is literally 'transformation through patterning'.

2 Poetry

The earliest writings on Chinese poetry can be found in the *Daxue (Great Preface)* and commentaries to the first anthology of poetry, the *Shijing (Book of Songs)* in the sixth century BC, and commentaries to the *Lisao*

(Encountering Sorrow), the longest elegy in the *Chuci (Songs of Chu)* anthology, attributed to Qu Yuan (?343-278 BC). These commentaries, written by Confucian scholars of the Han dynasty (206 BC-AD 220), establish a way of construing poetry as an 'ethico-aesthetic practice' that dominated literary exegesis up until the Tang dynasty (618-906).

Chinese poetry does not take its start, as does Western poetry, from the act of positing the subject and object of cognition (that is, the objectification and independent, self-subsistent existence of self and world). In the dualistic thinking predominant in the West, the poetic representation of objective things and events is a mimetic, and hence fictive, representation of the non-subjective things and events in external reality (see Poetry). The Western conception of mimesis is essentially predicated on the dualistic notion of metaphor as 'identity in difference'. This binary logic has led to two fundamentally opposing positions concerning the image (*imago*), be it poetic or pictorial. Classical views of mimesis treat the image as an imitation (*imitatio*) of a pre-existing phenomenal world (see Mimēsis). Advocates of this position tend to embrace a 'perceptualist' theory of art. The romanticist, however, construes the image as the pure artifice (*inventio*) of the artist, representing, yet distinctly different from, the sensible world.

In contradistinction to the mimetic practices and subject-object duality of the West, the Chinese poetic tradition assumes the existence of fundamental, mutually implicating correlations between the patterns (*wen*) immanent in nature and those of human culture. The Chinese poem was thus understood to invoke a network of pre-established categorical correlations (*lei*) between poet and world, which receives systematic formulation in such thinkers as Dong Zhongshu, the leading Confucian of the Western Han period.

The correlation between poetic natural imagery and human situations is achieved through three master tropes: *fu* or descriptive 'exposition', *bi* or metonymic 'comparison', and *xing* or metaphoric substitution or 'stimulus'. What differentiates these poetic tropes from their Western rhetorical counterparts is that they were not construed as the invention of the individual poet. Rather, links between poetic imagery, be they metonymic juxtapositions (as in the *Shijing*) or metaphoric substitutions (as in the *Lisao*), were traditionally considered to draw upon shared affiliations in a priori categorical correlations (*lei*) antecedent to the poem. It is an expressive-affective conception of poetry that assumes an internal (emotion, *qing*) and external (nature, *jing*) correlation, based on relations of causality, which enable the imagery to be read as verbal indices of personal feeling and the stability of the political and cosmic order. It is a Confucian notion of poetry which fulfils a crucial function of moral, social and political critique.

Lyric poetry, according to *Lu Ji* (AD 261-303), 'originates in emotion' (*shi yuan qing*), a response of the poet to the stimulus of nature. This is made clear in his *Wen fu (Exposition on Literature)*:

> He moves with the four seasons, to sigh at transience,
>
> And looks at the myriad objects, contemplating their complexity.
>
> He laments the falling leaves during autumn's vigour,
>
> And delights in the tender branches of fragrant spring.

(*Wen fu*, in Yu 1987: 33)

Nature is thus conceived as the stimulus or semantic evocator and source of poetic imagery.

It is a stimulus/expression/affect conception of *poesis* in which natural images draw forth pre-established categorical correlations (*ganlei*) which stimulate (*ying*) or arouse (*qi*) in the heart-and-mind (*xin*) of the reader a morally didactic and emotionally affective response (*gan*). Then, through a process of contextualization, the poet's work is interpreted as a didactic response to the particular set of circumstances that occasion the poetic response. In this way, Chinese poetry is understood as ultimately addressing an actual historical condition in the world around the poet. This tradition persisted in later years in attempting to read poetry as a political commentary or moralistic purpose, referring, no matter how obliquely, to the life history of the poet.

With the collapse of the Han Confucian order, China entered into a period of political division and social instability that enabled neo-Daoism and eventually Buddhism to exert an influence on the intelligentsia in southern China. It was during this period that Chinese scholars turned away from the discredited affairs of the court and

engaged in practices of 'self-cultivation' and spontaneous acts called 'self-so-ing' (*ziran*) within natural settings (see Self-cultivation in Chinese philosophy). It is also at this time that we witness the rise of a tradition counter to that of the Confucian exegetes, beginning with the non-canonical landscape-poetry of Tao Qian (AD 365-427) and Xie Lingyun (AD 385-433) of the Six Dynasties Period and culminating in the classical poetry of Wang Wei, Li Bo and Du Fu of the Tang dynasty (AD 618-906). This represents a shift from the didactic to the affective power of natural imagery. Scenic elements of an ostensibly observed natural scene are not only described in vivid detail, but also integrated into the coherent order of the represented scene and establish correlative relationship with the emotional situation of the poet.

Tao Qian is credited with transforming the expectations of an occasional poem so that it reads less as an allusion to the political realm than as a reference to the life of the individual poet (for example, retirement). While he often employs natural images that carry long established conventional associations (such as the pine tree) by embedding them within passages of detailed descriptions, they are naturalized by the reader as integral parts of the visualized world of the poem. The natural images are read not only as a description of an ostensibly perceived scene, but also as making personal reference to the poet's state of mind.

In the landscape poetry of Xie Lingyun, nature is even more precisely described than in the poetry of Tao Qian. However, it is not a detailed description of nature for its own sake, nor is it meant to convey symbolic or philosophical overtones. In the absence of natural images replete with conventional associations, Xie Lingyun's landscape poems are read as autobiographical narratives whose references are largely personal. The detailing of his perceptions and reactions to apparently observed scenery and his engagement with it carry meanings that establish a relation to the poet's state of mind. This is explicitly brought out in summary statements at the end of his poems.

In the Tang dynasty, this led to a significant reformulation of aesthetic theory, one in which the poet adopts a more receptive attitude and responds spontaneously to the world around him. This is a position informed by Daoist and Buddhist ideals, an embodied subjectivity that enables the poet to lodge his mind in the objects of the world in such a way that they can be imbued with emotional and intellectual content.

In the poetry of Wang Wei and Li Bo, informed by Daoist and Buddhist thought, the agency of the poet as authorial or poetic subject yields to the active presence of the objects of the natural world and defers to natural images to suggest emotion. In Du Fu's poetry, images of self and images of the world are so intertwined as to present, at one level, coherent scenes with a vividness and perceptual accuracy and, at another level, imagery that evokes an unstated emotional and/or intellectual 'meaning beyond the words'. *Ershisi shipin (The Twenty-four Types of Poetry)*, by the late Tang poet-critic Sikong Tu (837-908), exemplifies the move toward a poetry that employs natural images of the object world 'to suggest something ineffable and intangible'. The description of nature in his poems embodies the notion of going beyond the words of the poem, pointing to that which lies beyond words.

Although Confucian canonical exegetes focused on the moral implications and didactic force of poetic images of nature, and non-canonical critics focused on the way natural images come to embody human feelings, their premises were nevertheless identical. In the words of Fang Hui (1227-96), 'the profound meaning of the comparison and stimulus is to establish the secret links that hold together all things in the universe' (Yu 1987: 217). Lastly, there was an unstated assumption that all poetry, be it canonical or non-canonical, makes ultimate reference to the world of the poet.

3 Calligraphy and painting

The ethico-aesthetic practice of Chinese calligraphy and, by extension, Chinese painting has its genesis in the ancient ritual practices (*li*) of the Zhou dynasty (1122 BC-221 BC), a formative period in Chinese civilization, which 'witnessed a transition from spirit-centred to human-centred ritual, from shaman-counsellors to sage-counsellors, from authority by virtue of one's position to authority of one's person' (Hall and Ames 1987: 87). One of the ways of seeking moral perfection, of becoming an authoritative person (*ren*), is through the practice of art as a means of self-cultivation (*xiushen*) (see Confucian philosophy, Chinese).

Calligraphy and music, two of the 'Six Arts' (*liuyi*) in classical Confucian thought, emerge from this formative period of Chinese culture as ritual aesthetic practices for the upper class, as disciplines of the body (*ti*) and mind/heart (*xin*), which engage the gentleman-scholar in the cultivation of the self. *Xiushen* ('self-cultivation',

literally 'cultivation of the body') speaks to the 'importance of taking care of one's body as a necessary condition for learning to be human' (Tu 1985: 96) (see Self-cultivation, Chinese theories of). In this respect it is interesting to note that, 'Etymologically the character *yi*, which is commonly rendered as "art", signifies the activity of planting of [sic] cultivating fields' (Tu 1983: 60). Noting the cognate relation between the characters for 'ritual action' (*li*) and 'body' (*ti*), it is interesting to observe that 'li actions are embodiments or formalizations of meaning and value that accumulate to constitute a cultural tradition' (Hall and Ames 1987: 88).

The idea that art can serve as a means of self-cultivation finds its earliest expression in a passage on the meaning of music in the *Liji (Book of Rites)*: 'The perfection of virtue is primary, and the perfection of art follows afterward' (*Liji* 17, 3, §5, in Cahill 1966: 122). An affirmation of the practice of art can also be found in the following statement by Confucius: 'The Master said, Set your heart upon the Way, support yourself by its power, lean upon Goodness, seek delight in the arts' (*Lunyu* 7, 6, §§1-4, in Ledderose 1979: 29).

The concept of the Six Arts did not survive the fall of the Han Dynasty. During the Six Dynasties period, under the influence neo-Daoism and Buddhism, the perfection of selfhood came to be conceived as a 'dynamic process of spiritual development' whose internal generative force was often said to be heavenly endowed in nature. It was during this time that members of China's cultured scholarly elite adopted the practice of calligraphy and the playing of the lute (van Gulick 1969) as the specific means to pursue aesthetic self-expression and self-cultivation. As Tu Wei-ming has noted, 'One learns to play the lute or to sing lyric songs in order to communicate with others and, more importantly perhaps, to experience the internal resonance one shares with nature' (Tu 1983: 62).

In neo-Daoist inspired calligraphy of the Eastern Jin dynasty (AD 317-420), for example, the metaphysical principle *ziran* ('naturalness' or 'self-so-ing'), an impersonal creative potential, is cited as one of the most important aesthetic principles. The Northern Song calligrapher/connoisseur Mi Fu (1052-1107) reserved the aesthetic ideal of *tianzhen* (natural perfection) to praise the calligraphy of the Eastern Jin master Wang Xianzhi (AD 344-88). Lothar Ledderose has observed that: 'plain tranquillity (*pingdan*) and natural perfection (*tianzhen*) were not only stylistic and aesthetic concepts which could be used to describe and evaluate works of calligraphy, but these terms also described the ideal state of mind of the artist' (Ledderose 1979: 58). According to Mi Fu: 'the movement of the brush should come swiftly with a natural perfection and emerge unintentionally' (Ledderose 1979: 64). An appeal to an egoistic source of creativity would simply be unthinkable.

One of the distinguishing features of early Chinese aesthetic discourse is a predominance of physiological and nature imagery. The following passage from the *Bizhen tu (Diagram of the Battle Formation of the Brush)*, an early text attributed variously to Wei Furen (AD 272-349) and Wang Xizhi (AD 321-379), exemplifies the way in which the aesthetic discourse on Chinese calligraphy is framed in the terminology of human physiology:

> Calligraphy by those good in brush strength has much bone; that by those not good in brush strength has much flesh. Calligraphy that has much bone but slight flesh is called sinew-writing; that with much flesh but slight bone is called ink-pig. Calligraphy with much strength and rich in sinew is of sage-like quality; that with neither strength nor sinew is sick. Every writer proceeds in accordance with the manifestation of their digestion and respiration of energy, *hsiao-hsi* [*xaoxi*].

(*Bizhen tu*, in Hay 1983: 85)

In another passage in the *Bizhen tu*, images from nature are used to characterize the ideal rendering of the seven strokes that represent the so-called 'diagram of the battle formation of the brush':

> First stroke - like a cloud formation stretching a thousand *li*; indistinct, but not without form.
> Second stroke - like a stone falling from a high peak, bouncing and crashing, about to shatter.
> Third stroke - the tusk of an elephant or rhinoceros (thrust into and) broken by the ground.
> Fourth stroke - fired from a three-thousand pound crossbow.
> Fifth stroke - a withered vine, ten thousand years old.
> Sixth stroke - crashing waves or rolling thunder.
> Seventh stroke - the sinews and joints of a mighty bow.

(*Bizhen tu*, in Barnhart 1964: 16)

The deployment of metaphorical imagery, referencing the human body and nature in Chinese aesthetic theory, is

not simply a rhetorical flourish but, in fact, serves a specific epistemological function. It constitutes an indigenous correlative rhetoric stemming from the Chinese view of spiritual development that sought within the ritual aesthetic acts of self-cultivation to embody patterns of behaviour deemed consonant with the immanent patterns perceived within the natural order of things.

The concepts of *lei* ('kind' or 'categorical correlation') and *ganlei* 'responding according to categorical correlation', which were fundamental to early Chinese poetic theory, play a prominent role in Zong Bing's (375-443) essay *Hua shanshui xu (A Preface to the Painting of Mountains and Rivers)*, the earliest extant philosophical treatise on painting written in China. This can be illustrated in the following passage:

> Now the Sage, with his spirit realizes the Way; thus the worthy can pass through it. Mountains and rivers (likewise), with their forms, relish the way; thus the virtuous can enjoy it. How similar they are to each other!
> (*Hua shanshui xu*, in Munakata 1983: 118)

The reference to 'mountains and rivers' alludes to the Chinese term for landscape painting, *shanshui hua* (literally, 'mountain-water painting'). Zong Bing's categorical correlation of the 'Sage' and 'mountains and rivers' brings to mind a passage from the Confucian *Lunyu (Analects)*:

> The Master said, 'The wise [*zhi*] find joy in water [*shui*]; the benevolent [*ren*] find joy in mountains. The wise are active [*dong*]; the benevolent are still [*jing*]. The wise are joyful [*le*]; the benevolent are long-lived [*shou*]'.
> (*Lunyu*, in Lau 1992: 53)

The mountains and waters thus come to symbolize, respectively, the dimensions of constancy and change and, by metaphorical extension, tradition and its creative adaptation to the conditions of an ever-changing present.

The correlative rhetoric of *ganlei* was soon eclipsed by such physiological concepts as *qi* (vital force or energy flow) or *qiyun* (resonance of vital force,) and *xue* (blood) or *xuemo* (blood-pulse). *Qi* is variously translated as 'breath', 'spirit' or 'energy': the vital force that animates life (see Qi). *Xue* or *xuemo*, when it appears in discussions on calligraphy, refers to the energy functioning through the rhythmic flow of the ink within and between the characters. The quintessential use of *qi* as an aesthetic term appears in the *Liufa* or 'Six Laws' of painting by the portrait painter and theorist Xie He in the early sixth century AD. The first, and thus most important, of Xie He's laws is '*Qiyun shengdong*' which can be translated as 'life-movement [is achieved through] spirit resonance (or resonance of vital force)'. *Xue*, or the energy functioning through the rhythmic flow of the ink, is implied in the second of Xie He's laws of painting, '*Gufa yongbi*' or 'bone-method (that is, inner structure) when wielding the brush'. This can be interpreted as indicating the precise way in which to achieve 'spirit resonance'. Painting and calligraphy are thus conceived as configurations of energy, materializing through the brush into the traces of ink.

Chinese medical treatises have been shown to be another important source for the deployment of metaphorical imagery referencing the human body and nature in Chinese aesthetic theory. Previously, it was noted that ritual aesthetic acts of self-cultivation embody patterns of behaviour deemed consonant with the immanent patterns perceived within the natural order of things. In traditional Chinese medical treatises, the body is conceived as a system or network of patterned energy flow and transformation, that is, as a microcosmic correlative of the macrocosmic world of nature. Painting and calligraphy, conceived as configurations of energy materializing through movements of the brush into the traces of ink, thus came to be seen as ways of capturing the patterns of *shengdong* or 'life movement' of the phenomenal world.

The *Bifaji (A Note on the Art of the Brush)* of Jing Hao in the early tenth century is a reformulation of Xie He's 'Six Laws' for the purpose of representing the landscape. Jing Hao was a Confucian scholar-painter who, during the social and political turmoil of the Five Dynasties Period, retired to the Taihang mountains of Southern Shenxi. The centrepiece of the *Bifaji* is the Six Essentials in painting a landscape, reported to have been conveyed to the author/narrator by a rustic old man whom he came upon while painting in the Stone Drum Cave:

> Spirit (*qi*) is obtained when your mind moves along with the movement of the brush and does not hesitate in delineating images. Resonance (*yun*) is obtained when you establish forms while hiding [obvious] traces of the brush, and perfect them by observing the proprieties and avoiding vulgarity. Thought (*si*) is obtained when you grasp essential forms eliminating unnecessary details [in your observation of nature], and let your ideas

crystallize into forms to be represented.

(*Bifaji*, in Munakata 1974: 12)

These first three essentials of landscape painting prescribe artistic norms and conventions for the use of brush and ink that are self-effacing, concealing all traces of the material or formal process of representation and thus, by implication, all traces of personal expression in order to give transparent access to that which is represented. Jing Hao's 'Six Essentials of Painting' exhibits the influence of neo-Confucian values when it emphasizes the disclosure and transmission, through the receptive mind and the responsive hand of the painter, of the immanent patterns of nature in terms of the rhythmic patterned relations of the painted landscape forms.

4 Literati aesthetics

The re-establishment of national unity and order under the Song dynasty (960-1279) ushered in social and political conditions conducive to the formulation of a new literati aesthetic. The feudal aristocracy of landed gentry, prominent during the Tang dynasty, gave way in the Song to an 'aristocracy of merit' (Bush 1971: 4), a 'meritocracy', as civil service examinations provided truly talented scholars with access to high government office. Towards the end of the Northern Song (960-1126), a new and distinctive literati style of painting and calligraphy began to develop among a small circle of scholar-officials. A literati aesthetic theory was also formulated in an attempt to define the artistic and social identity of what the great Song poet and calligrapher Su Shi (1037-1101) referred to as *shiren hua* (scholar's art), in contradistinction to that of the professional painter and calligrapher. Where professional artists were dependent upon and sought the patronage of others for their livelihood, the literatus engaged in the practice of painting and calligraphy as means of self-cultivation, self-expression and social exchange with other, like-minded scholars. To this end, the literati style placed less emphasis on the descriptive depiction of nature, choosing rather to foreground the expressive potentialities of a more 'calligraphic' handling of the brush. A key tenet of literati aesthetics, the claim of equivalence between literati painting and poetry, appears in an inscription written by Su Shi on a painting by the great Tang poet Wang Wei: 'When one savors Mo-chieh's [Mojie's] poems, there are paintings in them, /When one looks at Mo-chieh's pictures, there are poems' (Bush 1971: 25), became one of the defining features of the literati aesthetic.

Song literati aesthetic theory discounted the mere technical skill of the professional painter to represent nature, in favour of what would come to be termed *xieyi*, 'to sketch, or paint ideas'. Terms such as *chu* (mood or flavour) and *pingdan* (plain tranquility) figure prominently in the writings of Mi Fu (1052-1107) and his son, Mi Yuren (1086-1165). These aesthetic terms identify the emotionally nuanced *yi* ('quality', or 'idea') of a scene, which can only come to artistic expression through the cultivated sensibilities of the literatus. For example, in Mi Fu's opinion: 'When Chu-jan [Zhuran] was young, he made many [forms like] 'alum lumps'; when he was older, in his tranquility (*pingdan*) the flavor (*chu*) was lofty' (Bush 1971: 68). *Pingdan* is 'a simplicity with underlying depth' (Bush 1971: 72).

During the succeeding Mongol Yuan dynasty (1260-1368), literati theory consciously stressed the non-professional status of the scholar-painter and the expressive, non-representational style of literati painting. There were two schools of literati aesthetics: classicist and individualist. Zhao Mengfu (1254-1322), a brilliant painter and calligrapher, is identified with the classicist position. A traditionalist, Zhao Mengfu stressed the importance of *guyi*, or the 'sense of antiquity':

> A sense of antiquity is essential in painting. If there is no sense of antiquity, then although a work is skillful, it is without value. Modern painters only know how to use the brush in a detailed manner and apply colours abundantly, and then think that they are competent artists. The fact is that if a sense of antiquity is lacking, all types of faults appear throughout a work, and why should one look at it? What I paint seems to be summary and rough, but connoisseurs realize that it is close to the ancients, and so consider it beautiful.

(Bush 1971: 121-2)

A more individualist position is asserted by such reclusive scholar-painters as Wu Zhen and Ni Zan (1301-74). Often, as in the following colophon by Ni Can (dated 1368), there is a self-conscious affirmation of the status of literati painting as merely 'ink-play' (*moxi*) in which great liberty is taken in the rendering of motifs in the interest of expressing the artist's mood:

> Chang I-chung [Zhang Yizhong] always likes my bamboo paintings. I do bamboo simply to express the untrammeled spirit (*yiqi*) in my breast. Then how can I judge whether it is like something or not; where its leaves are luxuriant or sparse, its branches slanting or straight? Often when I have daubed and rubbed awhile, others seeing this take it to be hemp or rushes. Since I cannot bring myself to argue that it is truly bamboo, then what of the onlookers? I simply do not know what sort of things I-chung is seeing.
>
> (Bush and Shih 1985: 280)

In the Ming dynasty (1368-1644), China is once again under native rule. Aesthetic theory as well as artistic practice come to take on an art historical dimension, as scholar-painters explicitly re-assert their social and artistic identity within a lineage of literati painters and calligraphers that is traced back through the Yuan and Song to the patriarchs of the tradition, Wang Wei, Dong Yuan and Zhuran. Towards the end of the Ming, this tendency culminates in the formulation of the theory of the 'Northern and Southern Schools' of painting (*nanbei pai*). Attributed to Dong Qichang (1555-1636), this theory systematically establishes the canon of literati painting and calligraphy (the Southern School) as the orthodox tradition for future generations of scholar-painters.

Dong Qichang formulates an aesthetic theory and artistic practice that synthesizes the classicist and individualist tendencies of the preceding Yuan dynasty. For the late Ming master, the proper approach to the canonical art of the past is one of a 'creative imitation' (*fang*) and 'transformation' (*bian*) within one's own personal style in a way that will allow one to speak with authority to the historical and art historical conditions of the present. Dong Qichang writes:

> Chu-jan [Zhuran] followed [imitated] Tung Yuan [Dong Yuan], Mi Fu followed Tung Yuan, Huang Kung-wang [Huang Gongwang] and Ni Tsan [Ni Can] both followed Tung Yuan. It was all the same Tung Yuan, but there several [versions of his style] did not resemble each other. If another kind of painter had done it, it would have been just like a copy. How could anything done that way be transmitted down through the ages?
>
> (Cahill 1982: 123)

It is precisely through an interpretive re-inscription or 'embodiment' of the 'orthodox' tradition within his own body of artistic expression, that the scholar-painter, sanctioned by the past, comes to signify himself in the present. Dong Qichang's notions of *shenhui* (communion of the spirit), *fang* (creative imitation) and *bian* (transformation) established the basis for both the Orthodox and Individualist Schools of literati painting in the Qing dynasty (1645-1912).

See also: Aesthetics; Aesthetics, Japanese; Art works, ontology of; Artistic expression; Poetry

STEPHEN J. GOLDBERG

References and further reading

Acker, W.R.B. (1954) *Some T'ang and Pre-T'ang Texts on Chinese Painting*, Leiden: Brill. (A translation of early Chinese texts on painting.)

Barnhart, R. M. (1964) 'Wei Fu-jen's *Pi Chen T'u* and the Early Texts on Calligraphy', *Archives of the Chinese Art Society of America* 18: 13-25.(Important discussion of early writings on calligraphy.)

Bush, S. (1971) *The Chinese Literati on Painting: Su Shih (1037-1101) to Tung Ch'i-ch'ang (1555-1636),* Cambridge, MA: Harvard University Press.(Systematic survey of literati theories of painting.)

Bush, S. and Shih Hsiao-yen (eds) (1985) *Early Chinese Texts of Painting*, Cambridge, MA: Harvard University Press.(Thematically arranged compilation of translations of Chinese writings on art.)

Cahill, J. (1966) 'Confucian Elements in the Theory of Painting', in A. Wright (ed.) *The Confucian Persuasion*, Stanford, CA: Stanford University Press, 115-40. (Discussion of Confucian aspects of Chinese painting.)

Cahill, J. (1982) *The Distant Mountains: Chinese Painting of the Late Ming Dynasty, 1570-1644*, New York: Weatherhill.(Thorough discussion of late Ming painting.)

Driscoll, L. and Toda, K. (1935) *Chinese Calligraphy*, Chicago, IL: The University of Chicago Press.(The first systematic discussion of Chinese calligraphy in English.)

Hall, D.L. and Ames R.T. (1987) *Thinking Through Confucius*, Albany, NY: State University of New York.(A most important discussion of Confucian ways of thought.)

Hay, J. (1983) 'The Human Body as a Microcosmic Source of Macrocosmic Values in Calligraphy', in S. Bush and C. Murck (eds) *Theories of the Arts in China*, Princeton, NJ: Princeton University Press, 74-102.(Important

discussion of the relation of Chinese medical treatises and early theories of calligraphy.)

Lau, D.C. (trans.) (1992) *The Analects*, Hong Kong: The Chinese University Press. (A standard translation of this Confucian classic.)

Ledderose, L. (1979) *Mi Fu and the Classical Tradition of Chinese Calligraphy*, Princeton, NJ: Princeton University Press.(A thorough discussion of the formation and transmission of the Classical tradition of Chinese calligraphy.)

Liu, J.J.Y. (1975) *Chinese Theories of Literature*, Chicago, IL: University of Chicago Press. (A thorough study of the different theories of Chinese literature.)

Munakata, K. (1974) *Ching Hao's Pi-fa-chi: A Note on the Art of Brush*, Ascona: Artibus Asiae Publishers.(A translation and discussion of an important early treatise on Chinese landscape painting.)

Munakata, K. (1983) 'Concepts of Lei and Kan-lei in Early Chinese Art Theory', in S. Bush and C. Murck (eds) *Theories of the Arts in China*, Princeton, NJ: Princeton University Press, 105-31.(A study of the role of the concept of 'categorical correlation' in early Chinese theories of art.)

Owens, S. (1992) *Readings in Chinese Literary Thought*, Cambridge, MA: Harvard University Press.(A meticulously annotated translation of seven major works in Chinese literary thought.)

Tu Wei-ming (1983) 'The Idea of the Human in Mencian Thought: An Approach to Chinese Aesthetics', in S. Bush and C. Merck (eds) *Theories of the Arts in China*, Princeton, NJ: Princeton University Press, 57-73.(An informative discussion of key terms in the philosophy of Mencius and their implications for Chinese aesthetics.)

Tu Wei-ming (1985) *Confucian Thought: Selfhood as Creative Transformation*, Albany, NY: State University of New York Press.(An important discussion of the concept of selfhood in Confucian thought.)

Tu Wei-ming (1979) 'Ultimate Self-Transformation as a Communal Act: Comments on Modes of Self-Cultivation in Traditional China', *Journal of Chinese Philosophy* 6: 237-46.(A discussion of the concept of 'self-cultivation' in Confucian thought.)

van Gulick, R.H. (1969) *The Lore of the Chinese Lute*, 2nd revised edn, Tokyo: Tuttle.(A study of the role of the lute in Chinese culture.)

Yu, P. (1987) *The Reading of Imagery in Chinese Poetic Tradition*, Princeton, NJ: Princeton University Press.(A thorough study of the role of imagery in Chinese poetry.)

Aesthetics, Japanese

While the terms 'aesthetics' and 'philosophy' were only introduced into Japan during the Meiji Period (post 1868), Japanese culture has nevertheless witnessed the proliferation of various arts and theories of art for over a millenium. Given that 'aesthetics' generally connotes a scientific, often taxonomic approach to the inquiry into beauty and art, it may be preferable to consider Japanese art and theories of art from the perspective of different ways of artistry, rather than impose on it alien categories and assumptions. Even our understanding about what constitutes art must alter when we consider such arts as the production of incense, the tea ceremony, the martial arts or flower arrangement, most of which do not have precise analogues in the West; or if they do, are not considered arts alongside poetry, drama, music and painting.

One of the hallmarks of Japanese art is the emphasis on an awareness of nature. Not only is the natural world a rich storehouse of images and metaphors for use as subject matter, but it is also the means whereby the practices, values and aspirations of the art are defined. Significantly, art itself is seen to be catalysed directly by an encounter with the natural world. All living beings, we are told, are given to song. Yet the natural world also came to be a shibboleth in society among the members of the Japanese court, where a finely honed seasonal awareness came to attest to the refinement and sensibility of the individual. Of all the arts, poetry was seen as pre-eminent, in part because of poetry's powers to influence the spirits inherent in the natural world. Even the emphasis on place and place-names in Japanese art may be traced to an understanding of the Japanese landscape and language as sacredly imbued.

Another feature of Japanese art and theories of art is its orientation toward the human. In other words, we may define Japanese art as 'expressive-affective' in its configuration, stressing the experience of the artist as well as the response of the audience in encountering such a work. In fact, the two roles of artist and audience are related through the focus of the work of art, which usually frames a single moment and its quintessential significance, hon-i, *which is unchanging. The quality which ideally characterizes both artist and audience is* makoto *or sincerity, underlining the point that the function of most Japanese art is to make us feel, rather than think.*

As in a number of other traditions, Japanese ways of art are bound up inextricably with issues of religion and religious practice. Not only did Shintō animatism have a profound impact on how Japanese viewed their landscape as well as their own lives, but other imported systems of belief also influenced the course of artistic development, especially Buddhism. Buddhism darkened the hues of classical Japanese art by introducing ideas such as mappō *(Latter Days of the Law), which saw the present as degraded and corrupt with respect to the past, and* mujō *(inconstancy), or the awareness of the ephemerality of this phenomenal world. In Mahāyāna Buddhism, art was perceived as a means of religious awakening, both in the case of poetry viewed as a form of intense meditation (*shikan*) and as parables whereby the truth could be disseminated obliquely (*hōben*). This paved the way for the pursuit of various forms of art to become a path (*michi*) to spiritual awareness. The relation of teacher and student in an art form closely resembled the relation of spiritual master to disciple, a feature which is echoed in the various 'secret' artistic treatises whose form, approach and significance suggest esoteric Buddhist manuals setting forth precepts for future generations.*

Japanese theories of art also concerned themselves with various aesthetic ideals, distillations of the changing notion of beauty in each era. From aware *(the beauty inherent in transience) and* miyabi *(courtly beauty) during the Heian Period (784-1185), to* yūgen *(the beauty of mystery and overtones) and* sabi *(the beauty of desolation and loneliness) in the medieval period, finally to* wabi *(the beauty of dearth and the humble) and* karumi *(the beauty of playful lightness) during the Edo Period (1600-1868), to mention only a few of the many ideals, we see an evolution of ideals as a response to cultural and historical change.*

What becomes evident in any survey is the assumption of an underlying unity, as in the notions that the impulse toward art is natural and universal; that art functions as a bridge mediating the experience of artist and audience; that sincerity and heart are to be privileged above all other qualities; and that the discipline of art can be a means of spiritual awakening. But we also discover that ideas, such as play, are critical to all forms of art in Japan. Other issues have surfaced periodically in various art forms in the course of Japanese history, such as the struggle between tradition and innovation or the debate about art as spontaneous versus art as the product of careful cultivation (that is, the question of artifice in art), or the question of the singularity of Japanese art.

1 Art, poetry and the natural world

Any sustained encounter with Japanese art or theories of art reveals the importance of the natural world, in the form of images of seasonal beauty, organic metaphors derived from natural objects and processes, and the identification of the internal psychological landscape with the natural world. Further, within the notion of the natural world lies embedded the importance of specific places within the landscape of Japan, which serve as touchstones to which art frequently alludes.

The twenty-one imperially commissioned anthologies of classical Japanese poetry (*chokusenshū*), compiled from the tenth to the fifteenth centuries, illustrate well the ubiquity of natural images and topics; each anthology devotes the first six books specifically to seasonal poems (two each for spring and autumn with one for summer and one for winter) and the remaining books, which focus on travel, love, sorrow and miscellaneous topics, are also replete with natural images. Similarly, most visual art focuses on scenes of natural beauty or wonder, ranging from the three classical beauty spots in Japan known as *nihon sankei* (Matsushima, Amanohashidate and Itsukushima) to typical scenes of enduring beauty (such as cranes amidst the pines) or seasonal beauty (brilliantly coloured autumn foliage) to a single object, such as a fish, in the natural world. In fact, the Edo period court painter Tosa Mitsuoki in the *Honchō Gahō Taiden (The Authoritative Summary of the Rules of Japanese Painting)* stresses: 'Anyone who wishes to learn the art of painting should first study the way of things in nature' (Ueda 1967: 31).

Other arts, such as flower arrangement or garden design, have as their explicit object the representation in miniature of the natural world. For example, in the *Sakuteiki (Notes on Garden Design)*, an eleventh-century manual by the courtier Tachibana no Toshitsuna, we are instructed from the outset:

> You should design each part of the garden tastefully, recalling your memories of how nature presented itself for each feature…. Think over the famous places of scenic beauty throughout the land, and… design your garden with the mood of harmony, modelling after the general air of such places.
>
> (Slawson 1987: 57)

Even arts which seem somewhat removed from the natural world (music, calligraphy, incense, swordsmanship) view themselves in some measure as representational arts and use natural images and metaphors taken from life to describe elements of their art, especially the epiphanic culmination of beauty in their art, as in the notion of the Flower (*hana*).

Indeed, while it becomes imperative for the student of any art in Japan to begin by studying nature, it is clearly not enough to have a general feel of the natural world. Instead we are commended to study nature precisely and minutely. The greatest *haiku* poet, Matsuo Bashō, according to his disciple Hattori Dohō, once commented, 'If you wish to learn about a pine, go to a pine and if you wish to learn about bamboo, go to bamboo.' This notion was often echoed by later writers such as the modern poet Masaoka Shiki, who advocated a theory of poetry known as *shasei* (sketch from life) as embodied in his statement: 'If you have the time to sit at a desk and read a book on *tanka* [classical poetry], you should instead pick up a cane and go for a leisurely walk along a path in the woods' (Ueda 1983: 19).

In the course of Japanese cultural history, however, the natural world as embodied in art did not simply realistically reflect external reality. Instead, beginning especially in the courtly Heian Period (784-1185), certain elements of the natural world were deemed to be worthy of inclusion into art, such as cherry blossoms, maple leaves, warblers and frogs, while other natural beings and objects, such as mudsnails, onions, the flowers of maple trees and foliage of cherry trees, were either discarded or ignored entirely. These elements obviously do exist in nature, but courtly poetic decorum eliminated them from the vocabulary of art. Not until the radical rejection of canonical poetic vocabulary by the Kyōgoku-Reizei school in the fourteenth century do we see dogs appear in formal poetry, for example, and it is really with the advent of *haikai* (comic linked-verse poetry) that we see such lowly elements as fleas and lice included as a matter of course.

If the study of the natural world, however circumscribed, seems necessary to the production of good art, it is also true that the genesis of art itself traditionally emerges from a direct encounter with the natural world. For example, Ki no Tsurayuki's [Japanese] Preface to the *Kokinwakashū (The Anthology of Japanese Poetry Ancient and Modern)* (*circa* 905), the *locus classicus* of theories of art in Japan, begins:

> Japanese poetry has the human heart for its seed and burgeons forth into myriad words as its leaves. Human beings in this world confront a plethora of experiences, and, hence, they give utterance to the concerns in their hearts through the sights and sounds around them. When hearing the call of the warbler amidst the blossoms, and the cries of the frog dwelling in the waters, among all living beings, is there any who does not burst forth in song? It is poetry that effortlessly moves both heaven and earth, inspires pity in the unseen demons and gods around us, makes tender the connections between men and women, and consoles the hearts of raging warriors.
> (*Kokinwakashū*, in Rodd and Henkenius 1984)

For Tsurayuki, it is the natural world that inspires in us the desire to compose verse and in fact the poetic process itself is posited by him in organic terms, utilizing the metaphor of a plant sprouting.

Echoing the Chinese courtly predilection for appreciating the natural world seasonally (emphasizing autumn and spring), in part derived from the Daoist love of the natural world, Japanese poets typically have associated each season with particular flora and fauna (for example, plum blossoms with very early spring, bell crickets with autumn). Further, each season was paradigmatically embodied in a particular time of day as attested by the opening of the well-known *Makura no Sōshi (Pillow Book)* of the court lady Sei Shōnagon, which begins elliptically, 'In spring the dawn, in autumn the dusk… ' An appreciation and understanding of seasonality became a kind of shibboleth among the members of the court, in which the truly refined and courtly sensibility could be discriminated from the uncouth and insensitive through the tacit acknowledgement of or subtle allusion to the modalities of the calendar. In *renga* (linked verse) poetry, flora and fauna became catalogued quite strictly according to the months of the year. For example, we are told that in verse associated with the fifth month, one may mention nightingales, early summer showers, orange flowers and irises. In *haiku* (or *hokku*) poetry, which developed out of *renga*, this seasonal association was mandated in the form of *kigo* or seasonal words, in which through allusion to the month, season, flora, fauna, climatic phenomena, or seasonal occurrence such as a festival, one would indicate the season.

In addition to the widespread use of natural images and topics in various forms of art, the natural world also functioned as a kind of dynamic and resonant mirror to the internal landscape of human feeling. While the notion of 'pathetic fallacy', or the according of human emotion to the natural world, has been derided by some in the Western tradition, notably John Ruskin in the nineteenth century, in Japan it has held a place of overarching importance. The 'pathetic fallacy' in the West implies a heavy-handed anthropomorphizing of nature, for example, 'the heavens weeping rain'. In the Japanese tradition, by contrast, we discover that the natural world and human world serve to mirror one another as well as catalyse parallel feeling in one another, in a much more nuanced fashion. Hence, while autumn with its inevitable desolation of life, withered leaves, and cries of insects comes to embody melancholy, as it often has in Western forms of art, it also serves as an externalized psychological mirror for the human heart, as attested in this illustration from *Genji Monogatari (The Tale of Genji)*, when we are informed that following the death of the Emperor's beloved and his rejection of any substitute, 'His serving women were plunged into dew-drenched autumn'. It is both literally autumn and metaphorically autumn, and it is less that one catalyses the other than a parallel accord between the internal and external worlds.

The proliferation of natural imagery in Japanese art also incorporates a preoccupation with place - both generic, as in Tsurayuki's opening, and specific, as in Hiroshige Andō's series of woodblock prints (*ukiyo-e*) in the Edo Period (1600-1868) entitled 'Thirty-six Views of Mount Fuji'. Another illustration of the emphasis on specific place names may be found in the classical poetic rhetorical technique of *utamakura* (the citing of one or more traditional place names from the poetic catalogue, such as Mount Yoshino or Tatsuta River).

Beyond the superficial understanding of Japanese culture as intrinsically enamored of natural beauty, commentators have ascribed this concern with the natural world and with place to either spiritual or functionalist causes, in other words, as the result of the claims of the indigenous system of beliefs known as Shintō and its myriad divinities of the natural world (*kami*), or as the outcome of inhabiting a densely populated island nation, insular and keenly aware of land as precious. These surely have had an impact on the emergence of Japanese consciousness, but what remains obscure is what deeper relation art, especially verbal art, has with the natural world and the idea of place.

Of all the arts, poetry emerges as the most valued form, and one whose assumptions underlie each of the other arts.

Further, as we consider the various forms of art that emerged - painting, the study and production of different kinds of incense, flower arranging or sculpture - we detect the importance of the textual and verbal components of each of these arts. The supremacy of poetry among the arts derives not only from its influence on and presence in each of the arts, but also from its extraordinary efficacy, as suggested by the far-reaching claims of the quotation from the preface of the *Kokinwakashū* above.

We may of course ascribe the ubiquity of natural imagery and the pre-eminence of poetry among the different arts to the influence of Chinese culture on Japan, and indeed many of the images that Tsurayuki uses are derived from the Great Preface to the Chinese classic the *Shijing (Book of Songs)*, but even in the earliest poems as represented in the oldest texts (*Man'yōshū*, *Kojiki*, *Engi Shiki*) which betray relatively little trace of sinification, we may detect a preoccupation with poetry itself and its manifold powers. Tsurayuki notes in his preface the universality of poetry and how the impulse to sing is shared by all life, but alludes only offhandedly to how poetry is able to accomplish these things, as he speaks of poetry moving the hearts of gods and demons.

Apart from the hierarchy among arts in the course of Japanese cultural development to which we have already alluded, it is generally serious art, which demonstrates a concern with public conventions, that is privileged over comic or aberrant art, which is personal or idiosyncratic in nature. The folklorist-literary scholar Orikuchi Shinobu deduces from all of this that the genesis of poetry arose indeed from the intersection between art and nature, though not simply from lyrical rapture amidst the blossoms, but from ancient word-charms and spells, thought to have been bequeathed by the gods to human society to effect certain results in the natural world (prosperous harvests, purification, exorcism).

In support of this view, many have cited the famous passage in the *Kojiki*, which discusses the origin of the performing arts (drama-dance-mime) in the form of *Nō*, as illustrated by the gods themselves. In this episode, the Sun Goddess, Amaterasu Ōmikami, having been perturbed, has shut herself into a cave thereby relegating the world to darkness. The gods having tried various manoeuvres to persuade her to come out to no avail, another divinity, Ama no Uzume, manages to lure her out by means of a song and dance she performs while in a state of abandoned divine possession. This distracts the gods from their predicament, inspires curiosity in the Sun Goddess and causes all to become engrossed in her performance. It is through the clever and beguiling use of art that order is restored to the world and that the gods of the land themselves have been moved and reconciled. But while this narrative illustrates the placative function of verbal art in dealing with *kami* (spirits inhabiting the natural world), in other words its rhetorical ability to persuade the gods to alter their course of action, as well as suggesting song as a medium familiar to the gods, it does not account for the emphasis on place.

According to the literary critic Konishi Jin'ichi (1984), the problem of place may be resolved through an understanding of the idea of *kotodama* or 'word-spirit'. Various utterances, which could only be pronounced in particular contexts and particular places that were called *kotoage*, unleashed the power of the *kotodama* which could be either malign or benevolent. It was not their semantic or rhetorical capacity that produced these effects, but rather simply the utterance of those syllables by a human voice. He argues that the concept of *kotodama* (and hence the construction of such a term), as opposed to the living enactment of it in archaic Japan, only came much later after Japanese poets in their intercourse with the continent became aware of the singularity of their beliefs about language and the natural world. Since *kotodama* were only efficacious when released by the utterance of *kotoage* in purely Japanese language (with no foreign loan-words) and within Japan itself, there comes about a new understanding of the natural world and the places within it, whose names alone can activate the mighty powers inherent in language.

2 Art in its expressive-affective capacity

A number of critics have pointed out that virtually all Japanese art and theories of art betoken an expressive-affective orientation. In other words, rather than privileging the relation of a work to nature (art as representational or mimetic) or as an elaborate structure whose significance inheres principally in its formal properties, Japanese art emphasizes the relationship between the artist and the work of art as well as the relationship between the work of art and its audience.

More often than not, we are told, artists produce works, not because of a desire to assert themselves but because of an inability to restrain themselves. Tsurayuki's female narrator comments in his fictionalized diary *Tosa Nikki*

(Tosa Diary):

> I do not set down these words, nor did I compose the poem, out of mere love of writing. Surely both in China and Japan art is that which is created when we are unable to suppress our feelings.
>
> (Miner *et al.* 1985: 7)

We sense an urgency about art: as humans we are filled with emotions generated by experiences in this world, and we must express them in the form of art.

At its base, we may say that Japanese art is lyrical in nature; it highlights the emotional responses of an individual, or the group in the voice or person of an individual, to the natural world around. Hence we see comparatively little in the way of abstract, didactic or gnomic art, other than explicitly religious or heuristic works such as visual representations of Buddhas or neo-Confucian homilies. The principal function of most art is not to instruct directly (though the very cultivation of art may be seen as uplifting to character or, put more precisely, the pursuit of art necessitates the cultivation of a high moral discipline), but rather to allow the individual to manifest an outpouring of feelings in a resonant fashion that serves as gratifying release.

The focus of many works of art is the revelation of a single moment of being, to use Virginia Woolf's phrase, in which the intensity and purity of feeling and experience are transmuted into a thing of beauty. There is a spontaneity and singularity about the artistic moment, which of course is the culmination of much concentrated study of the art form, and which at particular moments of artistic inception allows the artist to produce a masterpiece without hesitation or conscious forethought. What a work of art offers us ideally is not the opportunity to uncover or unearth the mind or personality of the artist but the quintessential distillation of an authentic human experience amidst the natural world. It is not validity but sincerity (*makoto*) that we prize. Analogously, the affective relation between a work and its audience is intended not so much to make us think, but to make us feel. As a result, a tacit connection is effected between artist and audience which is mediated by the work of art.

Some critics and theorists structure their texts specifically towards the production of art such as Ki no Tsurayuki while others, such as Zeami, the great playwright and theorist of *Nō* theatre, stress the role of the audience in its appreciation of art as well as the development of the artist. The more cultivated the audience, the greater the inducement for artists to express themselves well. Thus, Zeami insists: 'It is crucial for the actor to perform in such a fashion so as to harmonize with the feelings of the nobility' (Rimer and Yamazaki 1984: 19). What is idealized is the reciprocal relation between the artist of impeccable sensibility and training and the audience, equally trained and possessed of a discriminating awareness, a relation which allows for the intuited sharing of fundamental experience.

While the focus of the artistic work is understood to be expressive-affective in its orientation, this does not preclude the notion of art as imitation (*monomane*). For example, Ōkura Toraakira in his *Warambegusa (For My Young Successors)* articulates the ideals of *kyōgen* theatre (comic interludes performed in between the more serious and weighty *Nō* dramas), asserting: 'More than anything else *kyōgen* is an art of imitation…. It imitates all kinds of things in this world' (Ueda 1967: 102). However, while *monomane* sounds very much like the idea of mimesis, in fact it suggests less a connection with the object as it exists in the world, than the idealized essence of the object as realized in the realm of art. Zeami mentions the principle, 'first truly become the thing you are performing' (Rimer and Yamazaki 1984: 77), but he does not mean a performance along the lines of method-acting. By 'the thing', he means the reified and transmuted kernel of the role, what is sometimes called *hon-i*, or essential character as codified by artistic conventions, not the role as it is embodied in an actual being in society. Though *hon-i*, derived from a Chinese term *pen-i*, originally meant something like 'individual will or aspiration', it came to connote 'poetic intent', and thence 'fundamental character or nature'. As examples, the *hon-i* of cherry blossoms is embodied by beautiful transience, of travel by misery and loneliness; of age by regret and nostalgia. The artist in encountering the outside world does not in fact confront the real world, but rather the world of *hon-i* or codified essences, those very essences recognized as such by the audience. This is what Chikamatsu Monzaemon, the supreme playwright and arbiter of the Edo-period puppet theatre, *ningyō jōruri*, meant when he asserts, 'Art is that which exists in the slender gap between truth and falsehood.'

The notion of *hon-i* of course represents in part the world of accreted tradition against and atop of which artists inscribe their own work, and poses a critical dilemma about how artists can be authentic about their experience

even as their vision is filtered through the screen of tradition. However, *hon-i* is understood to be more than just convention; it is what Motoori Norinaga referred to as ‘the heart of things’ (*koto no kokoro* or *mono no kokoro*), that essential nature that is not subject to the vicissitudes of time (see Kokoro). Nevertheless, throughout the course of Japanese artistic development we may detect a fruitful tension in the form of tradition versus innovation that is crucial to the expressive-affective orientation, given the inevitable changes that each new generation must confront. Tsurayuki notes at the end of his preface to the *Kokinwakashū*, ‘Those who know poetry and who understand the heart of things will look up to the old and admire the new as they look up to and admire the moon in the broad sky’ (*Kokinwakashū*: 47). The medieval poets and critics, Fujiwara Shunzei and his son Teika, also address this issue in their well-known exhortation ‘old words, new heart’. According to Bashō’s disciple Kyorai:

> Of the *haiku* there is a style that remains unchanged for thousands of years. There is also a style that prospers only for a time. These are the two poles of the late Master’s teaching, and they are really the same in essence. They are the same, because they both resort to a single source, the poetic spirit.
>
> (Ueda 1967: 147)

Another implication of the emphasis on the expressive-affective orientation, the human aspect of the work of art, is the overriding importance placed on process as opposed to telos, goal or end. Almost all treatises on art in Japan discuss in great detail how artists can go about cultivating their art, rather than enumerate precisely the constituent features of a great work of art in the abstract. The twentieth-century philosopher Kuki Shūzō speaks of the traditional Japanese emphasis on process when he discusses the notion of *bitai* (coquetry) as a cultural ideal and how its evocation of *kanōsei* (possibility) makes it endlessly alluring. He speaks of how after the devastating 1923 Kanto earthquake, the Japanese quickly went about rebuilding the subway in the city of Tokyo, knowing of course that it could be destroyed at any point in time, to the amazement of Europeans for whom it seemed a doomed effort. Kuki explains that it is the process of construction and cultivation, not the end result, that sustains the Japanese will. Hence, he argues that from a Japanese perspective, the position of Sisyphus in Greek myth is quite a desirable one, since he will always be in the midst of a task, never having to confront the bleakness of completion or perfection.

3 Art and its relation to religion

The indigenous system of beliefs known as Shintō had a significant impact on the development of the arts in Japan (see Shintō) and, as some believe, literally gave rise to them in the form of the *kotodama*. Other forms of art associated with Shintō animatism (which has been described as a pre-animistic system which features belief in spiritual entities with non-human traits) include *kagura* (god’s music), *kamiasobi* (god’s dances), both of which contributed to the development of *Nō* theatre, *norito* (poem-like liturgical texts) and *setsuwa*, or short mythic narratives dealing with the deities and various creation themes, often orally related by professional narrators known as *kataribe*.

As an aside, the *kataribe*, who were often female, were thought to represent a transitional phase in verbal art between the stage of oral transmission and the advent of writing. In a world where memory was being displaced by written text, the *kataribe* functioned as a repository of group memories of ancient narratives until finally they were almost entirely displaced by written accounts in the highly literate late ninth century, when their only remaining function was limited to recitation of specific works at ceremonial times. According to Orikuchi Shinobu, these narratives, as well as poetry and proverbs, originated in the incantatory utterances of shamanesses who were possessed by the spirits of the *marebito*, or the visitor gods, who would descend at regular intervals into the human realm from the Land of the Eternal. Eventually, of the incantations divided into *ji* and *kotoba* sections (third-person and first-person assertions by the god respectively), the former developed into narrative art and the latter into poetry and proverbs. The narratives, no longer seen as the actual words of the gods, came to be related by female *kataribe*, presumably the secularized descendents of and successors to shamanesses who left their ancestral villages, and thus was born narrative art from religious beliefs.

Even more important was the impact of Buddhism on the arts of Japan, in terms of form, practice and understanding (see Buddhist philosophy, Japanese). It is noteworthy to remember that Buddhism was also the principal vehicle whereby writing was introduced into and popularized in Japan. Further, both the Tendai and Shingon sects, but especially Shingon with its emphasis on the symbolic understanding of essential truth as

conveyed in art, had a formidable impact on the development of courtly arts from the latter part of the ninth century onwards. Various arts were associated specifically with the religious practices of Buddhism, such as carved wooden statues of Buddhas and Bodhisattvas, the *bussokusekika* (the Buddha's Footprint Poems at Yakushiji in Nara), painted *mandaras* (cosmic diagrams of Buddhist figures), *jigoku-hen* (frightening hell screens designed to chasten unbelievers), and the texts of *Gozan bungaku* (the heavily Chinese-influenced literature of the Zen Buddhist monks of Kyoto and Kamakura during the Kamakura-Muromachi Period). While Jōdo (Pure Land) and Nichiren Buddhism held considerable sway over the populace, it was the esoteric sects as well as Zen that had the greatest influence on the practice of poetics and the development of the arts.

The two Buddhist concepts *mappō* and *mujō* substantially altered the outlook of poets and artists in the medieval period. *Mappō* (the Latter Days of the Law), known in Sanskrit as *Kaliyug* (the Dark Age), represents in the cyclical time of Mahāyāna Buddhism the final corrupt, decadent stage of life in which a Buddha is needed to be born on earth to enlighten the misguided and transmit the Law for the cycle to begin anew. In poetry and poetics, this resulted in the glorification of the past as a kind of Golden Age and the dismissal of the present as a barbarous era in a dismal state of decline. Hence, throughout poetic treatises and works of art we see evidence of this idea in the disdain shown for the present state of affairs and the nostalgia and reverence for the past. *Mujō*, the inconstancy of the phenomenal realm, like the earlier notion of *hakanasa* (ephemerality), underscored change as ineluctable in this world, but presented it in a much darker fashion. It no longer functions as an emblem of the beautiful, but that which we as deluded beings refuse to acknowledge, thereby bringing misery and suffering on ourselves (see Mujō). Both *mappō* and *mujō* discomfit us with our place in the present world and foreshadow the need for transcendence.

The paradoxical question of how art, itself rooted in the phenomenal world, can help us to transcend the unrelenting realities of this realm, was thought to be resolved through the evocation of the ideal of *hōben* (skill in means or expediency). In the *Hokkekyō (Lotus Sutra)*, the central scripture of Tendai Buddhism, Buddha is asked how one goes about teaching the truth to those ignorant of it and hence unable to recognize truth as truth. Buddha responds with the idea of *hōben* or heuristic aids, such as parables. Even though these are fictions, he explains, ironically they can help the ignorant to grasp some notion of truth, which eventually will lead to complete understanding. Thus, writers in the middle ages were fond of quoting the Tang Chinese poet Bo Zhuyi's famous phrase, *kyōgen kigo* (wild words and fancy phrases) as alluding to the possibilities of art as *hōben*: 'For many years have I hoped one thing, that my actions in this world and the sins resulting from my wild words and fancy phrases shall become ever after a means of paying homage to the Law and be allied to the dissemination of Buddha's truth. Let the countless Buddhas of the Three Realms hearken unto this.'

Thus even art, which does not seem to promulgate religious truth explicitly, may be seen to further the teachings of Buddha. Even a genre perceived to be as frivolous as *monogatari* or narrative may at heart be an instrument for Buddha's teachings. Art, though fictive, can serve to guide an audience indirectly to the higher truth and often with greater efficacy than any tract or scripture. This offers an interesting counterpoint to Sir Philip Sidney's *An Apology for Poetry*, in which he defends art against the criticism that as fiction it lies, by pointing out that since fiction avers nothing, it cannot be equated with falsehood.

Beyond its function as a conduit to truth, art may also serve as a direct means of religious awakening as addressed in the idea of poetry as a form of *shikan*, or the Tendai practice of intense and concentrated meditation. According to the *Maka Shikan (Great Cessation and Insight)*, one of the three great texts of the Tendai sect, by desisting from the impulse towards discrimination between objects in the world and concentrating on the integrity of each, we can gain insight into our own inner nature. Fujiwara Shunzei applies the insights garnered from this text to the composition of poetry. Poetry in this form of meditation allows us to dissolve polarities and distinctions, recognizing them as having significance only in the illusory phenomenal world, and in so doing, allows for a fundamental unity to be realized between the poet and the object. Thus, the very act of poetic composition as *shikan* enables us to achieve a higher spiritual awareness of self.

Just as Tendai laid the groundwork for Zen, so too did the notion of *shikan* shape the notion of *michi* (path or way) in its implications for the arts. By the fourteenth century, we see a number of arts subscribing to the Zen notion that any vocation, if pursued fervidly and with a purity of heart, can be a means to enlightenment in the form of *satori* or sudden awakening in this life. These include poetry, tea, flower arrangement, *Nō* theatre, martial arts such

as the art of the sword, and a host of others. What is significant about a *michi* is that it requires a singular devotion to the art form, the eschewal of base behaviour, a humble, almost monastic lifestyle, and the realization that the purpose of such a pursuit is neither power nor fame, nor any other mere gratification of the senses, but rather a lofty and transcendent goal which requires us to see into ourselves.

Following a *michi*, whether it be the way of *renga* or calligraphy, does not imply a solitary pursuit, but, more often than not, the attaching of oneself as a student to a teacher or master. The organization of most arts in Japan revolves around specific schools or lines of artistic inheritance known as *ie*. In many ways, this is the legacy of esoteric Buddhism with its emphasis on secret transmission of the innermost truths strictly from master to chief disciple. Just as Saichō (Dengyō Daishi) dispensed precepts for spiritual awakening to his chief disciples (*endonkai*), so too are artistic precepts passed through the *ie* system from master to disciple. Thus, we encounter the practice of *hidensho* or secret treatises, which purport to divulge the ultimate truth that can only be understood by those few chosen initiates. These were often written in a dense and elliptical style that precluded the ignorant from understanding the hidden implications. The *hidensho* were not substitutes for the master's teachings, but instead, since they consisted of generalized truths that could only be understood only by those already initiated into the arcane mysteries of the art, were meant to be passed down through the generations ensuring that the art form continued. In addition, they were proof in themselves of one's having received the mantle of authority in the particular art form. Thus, during the decades that followed the death of the great Fujiwara Teika, his descendents continued to dispute fiercely among themselves about the ownership of his writings and documents, which were simultaneously both a poetic and political inheritance.

The relation between teacher and student in the arts represents an almost sacred bond in which students abase themselves before the master as proof of their readiness to learn, paralleling the relation of monk to spiritual master. To belong to an *ie* headed by a lineal master means to subscribe to its rules, its practices, its dicta, and moreover to learn, not by reading or imagining or experimenting, but rather by direct emulation of the master's teachings. This teaching invariably takes the form of *kuden* or oral transmission, which is seen as far superior to anything transmittable in writing. Further, within esoteric traditions, *kuden* referred to those hermetic teachings, secret in nature, which revealed the ultimate truths. Given the importance placed on *kuden*, many traditional treatises in Japan take the form of *mondo* or catechistic question and answer, as if representing as closely as possible the actual oral transmission.

Though evident in certain periods, in certain genres and in the work of particular artists, both Confucianism and Daoism had much less direct influence than did Buddhism on the course of Japanese poetics and the arts. The didactic strain characteristic of Confucianism does surface on occasion, as in Edo Period literati homilies designed to 'praise virtue, castigate vice' (*kanzen choaku*), a phrase deriving from Confucianism that appears as early as Shōtoku Taishi's *Jūshichijō Kempō (Seventeen Article Constitution)* of 604 (see Shōtoku Constitution). Daoist ideas, while especially evident in certain of the writings of the Man'yō poets, become amalgamated with other forms of Japanese thought such as Zen, and thus manifests itself only indirectly in the general esteem for the natural world, and the borrowing of terms, such as the Chinese *feng-liu* as *fūryū*, or elegance (see below).

4 The historical development of traditional aesthetic ideals

When we look at pre-ninth century Japan, although we encounter a great body of art and poetry, we see little in the way of contemporary commentary on aesthetics and poetics. While later eras detect certain ideals implicit in this period, such as *makoto* (sincerity), *masuraoburi* (ideal of virility) or *man'yōgokoro* (the Man'yō spirit), in fact these ideals are anachronistic and represent an attempt to instill conceptions of art somewhat removed from the period. What the idealization of these elements reveals is an Edo-period nostalgia for an imagined simpler, more straightforward time, which precluded any outside influence and any notion of transcendence.

One of the earliest commentators on art was the Buddhist monk Kūkai (Kōbō Daishi), who in his *Bunkyōhifuron (Secret Treasury of Poetic Mirrors)* and *Bumpitsu genjinshō (Essentials of Poetry and Prose)* sought to introduce Chinese theories about verbal art into Japan. It is after this point that we detect the emergence of indigenous artistic ideals that, though indebted at some level to Chinese culture, represent singularly Japanese notions about art. The construction of aesthetic ideals arose mainly in the Heian Period as courtiers sought to codify matters of art into a coherent form that both set itself off from the domain of Chinese culture and art (principally of the Six Dynasties) and within Japan delineated the boundary line between those within and those outside the court.

Apart from critical terms, such as *kokoro* (heart), *kotoba* (words) and *sama* (style), one of the earliest ideals evident in Japan is *aware*, a term signifying poignant beauty that recurs a number of times in works such as *Genji monogatari* (a repository in fictional form for Heian responses to and understanding of art) and the imperially commissioned anthologies of poetry. In Heian Period usage, *aware* seems closely aligned with the idea of *hakanasa* or transience, suggesting the truism that what is sad is necessarily beautiful, and what is beautiful is sad. The notion of *kokoro* is central to the ideal of *aware*, since the capacity to be moved is a function of sensibility. In its approbation of *Genji* as the supreme work of fiction, the *Mumyōzōshi (Nameless Writings)*, *circa* 1200, a treatise in the form of a dialogue on *monogatari* or narrative, cites numerous instances of *aware* such as Yūgao's death, Genji's exile and Fujitsubo's becoming a nun and taking the tonsure.

Our modern understanding of *aware* depends in large measure on the Edo commentator Motoori Norinaga's conceptualizing of it in the form of *mono no aware* (the pathos of things), often seen as analogous to the expression *lacrimae rerum*. He explains *aware* as originating in two cries of wonderment at the world, 'Ah', and 'Hare'. While he notes that originally *aware* was used of anything that moved the human heart, later it came to be associated exclusively with the miserable or wretched. For Norinaga, *aware* symbolized the sensitivity to temporal beauty that defines the experience of creating or appreciating art.

The overall aesthetic of the court, especially in the Heian period, was referred to as *miyabi* (also read sometimes as *ga* in its Sino-Japanese reading) or courtly beauty, a beauty that was ornate, brilliant, and characterized by elegance. This is the kind of beauty personified by the title character Genji himself: handsome, refined, well-versed in all the arts, graceful in all his actions. It implies what the Renaissance writer Castiglione referred to as *sprezzatura*, an elegant nonchalance, as well as a peerless sensibility. Closely allied with it was the notion of *fūryū*, from the Chinese Daoist term *fengliu*, which originally connoted the elegant world enjoyed by the Immortals consisting of music, poetry, wine, and the world of desire. *Fūryū*, whose characters were read also as *miyabi* in Japanese, emphasized a fashionable chicness, brilliantly realized as well by the figure of Hikaru Genji, the Shining One. *Miyabi*, needless to say, eschewed anything touched by ugliness, poverty, or corroded by age.

In the late Heian and early part of the Kamakura Period, we see a new set of ideals emerge, the most important of these being *sabi* and *yūgen*, as propounded by Fujiwara Shunzei and his son Teika among others. This new era was heralded by active civil strife from 1183-85 in which the imperial family, various noble clans and the major Buddhist monasteries fractured internally and, in the ensuing conflict, many were eventually decimated. Hence the tone of these new ideals was much darker and more tinctured by the Buddhist notion of the inconstancy (*mujō*) of this phenomenal and deluded world. *Sabi*, or the ideal of loneliness or desolation, remains one of the most enduring ideals in the course of Japanese cultural development, playing a significant role in various arts such as the tea ceremony, flower arrangement and brushed ink painting, as well as later forms of poetry such as the *haikai* of Bashō. *Sabi*, in sharp contrast to *miyabi*, is subdued, monochromatic in hue, and melancholic in tone. We see the opposition of these two ideals in this poem by Teika:

> Miwataseba
>
> hana mo momiji mo
>
> nakarikeri
>
> ura no tomaya no
>
> aki no yugure
>
> (As I gaze out, both cherry blossoms and maple leaves are absent, instead grass-thatched huts in the autumn gloaming.)

(*Shinkokinwakashū* 4: 363)

Instead of the gorgeous and conventional images of vernal and autumnal beauty, we are presented with an alternate image of isolated, unadorned, and barely discernible beauty. *Sabi* implicitly acknowledges the darkness of life, even as it reconstructs the misery into a thing of quiet beauty. While Shunzei also favored the ideal of *aware*, similar in content if not in tone or affect, the austerity of *sabi* became the hallmark of his work and this period in Japan. In the hands of later poets, such as Matsuo Bashō, *sabi* or loneliness became conflated with *sabi*, a

homophonous word for rust, and thus is transformed into the antithesis of *miyabi*, the beauty of the extraordinary giving way to the beauty of the ordinary, the sere, the solitary.

Another ideal developed by Shunzei's son Teika in his early formulations was the notion of *yōen* or ethereal charm. This, like many other concepts, had its origin in Chinese poetics and signified a dreamy, feminine, winsome beauty of a sort associated with the delicacy of a fragrant blossom or the romance and magic of a spring evening. In its emphasis on transcendent possibility and the resonance of *yojō* or overtones, *yōen* is not unrelated to *yūgen*, perhaps the most profound and ineffable of all Japanese aesthetic ideals. *Yūgen*, the style of mysterious beauty or alternately mystery and depth, was originally a Buddhist term meaning 'obscure, dim, or deep', but was elevated by Shunzei into an aesthetic ideal suggesting great subtlety, complexity and reverberation. *Yūgen* portends an otherworldly atmosphere that hints at but never elucidates fully the possibility of transcendent vision.

Zeami also borrows the notion of *yūgen* from poetry and employs it in his discussions on drama, noting: 'Particularly in the *Nō*, *yūgen* can be regarded as the highest principle.' However, his understanding of it suggests more the mastery of elegant beauty and grace than what is suggested by the use of *yūgen* in poetry. Still, even in *Nō*, *yūgen* continues to represent an elusive ideal that is the culmination of artistic endeavour.

The cultural shifts of the fifteenth and sixteenth centuries foreshadowed the aesthetic ideals of the Edo period, with their emphasis on quotidian life amidst the newly urbanized lower social strata as embodied in the townsmen or *chōnin* class. Sen no Rikyū, the acknowledged master of *sadō* (or *chadō*), the tea ceremony, according to his student Nambo Sokei, preferred above all other ideals *wabi* or the beauty of impoverishment. He remarks, 'There should be a dearth of tea implements in the room', and advises that these few objects should be arrayed simply. Among the many revealing anecdotes related about him is the story of his deliberating mutilating a vase to use in tea ceremony as a reminder to us of the beauty inherent in the imperfect and the shabby.

Bashō, the *haiku* master, embraced the ideals of earlier ages, especially the notion of *sabi*, but sought to mitigate *sabi* first through the humanizing notion of *wabi* and then through the ideal of *karumi* (lightness). *Sabi*, with its unrelieved austerity and detachment, was in some ways as alien to the common people as the earlier *miyabi* had been, and hence Bashō sought to include the commonplace, the humble and the destitute, as well as the simple, the playful and the light-hearted, in his work in order to suggest a more encompassing aesthetic for his age. While *wabi* helped to convey a more humanistic and egalitarian ideal, in that by elevating the disfigured and the discarded we attest indirectly to the resilience and beauty of imperfect humanity, the notion of *karumi* or lightness proved to be central to his poetry in relieving the weight and darkness implied by *sabi*. Both *wabi* and *karumi* are evident in a verse such as this excerpted from a *haikai* sequence:

> Te no hira ni
>
> shirami hawasuru
>
> hana no kage
>
> (In the palm of my hand, the lice crawl forth, in the shade of cherry blossoms)

The humble image of lice, juxtaposed with the traditionally exalted cherry blossoms, is framed by the human observer's eye and hand to produce a poem of considerable whimsy and poignancy.

One last major aesthetic ideal known as *iki* (the ideal of the chic or stylish) emerged in the late Edo period among the female entertainers of the cities of Edo. *Iki* paralleled the ideal of *sui* (written with the same character signifying essence), which dominated in western Japan, but was much less constrained, ornate and traditional as a fashion. *Iki* in its casual nonchalance and disdain for convention suited well the character of Edo, unburdened by many centuries of tradition, and came to symbolize the pragmatic surface coolness belying the sentiment welling up within, characteristic of this period dominated by urban, mercantile values.

5 The unity of the arts and the identity of the nation

Far more than the art and aesthetic theories of other cultures and traditions, those of Japan betray a remarkable uniformity in their underlying assumptions regarding the nature of art, its provenance and its aims. For example, one shared quality (not discussed above), the idea of playfulness, appears in virtually every art form from the

comic narrative about the plump, middle-aged goddess Ama no Uzume luring the Sun Goddess from her cave by means of a lascivious primordial striptease, to the courtly game competitions known as *monoawase*, which involved the wagering of iris roots, pictures, poetry or virtually any object, to the various kinds of playful poetry including the courtly *oriku* or acrostic, the *darumauta* or nonsense poetry associated with Zen practices in the medieval period, and the satiric, often bawdy poetry known as *senryū* produced in the Edo period. While art is not always ironic or comic, the notion of art as serious play or pasttime prevails across the various genres and modalities of artistic expression in Japan. Even the preference shown by the aristocratic elite for the gifted amateur in the arts, as opposed to the professional, while indeed betraying the influence of Chinese literati notions, also demonstrates this principle of play as being central to art in Japan. Play does not stand in opposition to *makoto* (sincerity); if anything, it attests to it.

Many schemas have been proposed for understanding Japanese art as an integrated unity. The literary critic Donald Keene, for example, offers us a formalist grid in which four qualities recapitulate the notion of beauty within the boundaries of Japanese art and aesthetics: suggestion, irregularity, simplicity and perishability. These can be associated with the ideals that arose in specific historic contexts (that is, *yojō* or *yūgen*; *wabi*; *sabi*; and *hakanasa* or*mujō*), but represent as well enduring concepts that characterize the landscape of Japanese art from tea ceremony to calligraphy to garden design to poetry.

By contrast, other schemas that have been proposed imply an ideological difference that functions to distinguish Japanese art from any other. Konishi Jin'ichi invokes a polar struggle in the tension between what he terms the poles of 'consummation' and 'infinity' in the ubiquitous human 'longing for the eternal'. *Ga*, or the aesthetic of the high and the refined which seeks perfection of form in the flowering of ideals already in existence (in other words, tradition), reflects this impulse toward consummation, whereas *zoku*, or the impulse toward infinity, is characterized by a preoccupation with the low, the vulgar and the popular, embracing innovation freely and with energy. During much of Japan's cultural history, *ga* was associated with the legacy, direct and indirect, of Chinese aristocratic culture whereas *zoku* frequently was associated with indigenous notions emerging from the common people. For Konishi, *ga* and *zoku* function not merely as aesthetic ideals but ideological poles, which have been naturalized in the context of Japanese society and thought and whose dynamic interaction in different eras is evident in the forms of cultural production.

The novelist Tanizaki Jun'ichirō, in his essay 'In'ei Raisan' (In Praise of Shadows), also argues for a polarity, but in his case this polarity derives not so much from ideological differences within a society as from differences between Western art and Japanese art. In his wide-ranging analysis which discusses various arts from theatre to domestic architecture, with an especially appealing comparative section on bathrooms, he presents Western art as revolving around the worship of light and illumination, and Japanese art around darkness and obscurity as central metaphors. In many ways, what Tanizaki presents is an apology or defence of Japanese culture and art, suggesting that the very terms of Western approbation are inadequate for understanding Japanese notions of beauty. In doing so, he echoes the ideas of other commentators, such as Okakura Kakuzōwho in his *Ideals of the East* and *The Book of Tea* sought to define an aesthetic of Japan that stood aloof from Western paradigms.

Lest this be mistaken as a solely modern impulse to see the ethnic or national character embodied in cultural production, such as art and aesthetic theories, it is useful to recall Tsurayuki's Preface in which the very first words suggest his insistence on and preoccupation with Japanese poetry, attesting to what we might term 'the anxiety of influence' resulting from the penumbra cast by China's looming cultural shadows. We see this same preoccupation with Japanese art and its implications throughout *Genji Monogatari*, where China functions simultaneously as a yardstick and as that cultural monolith which must be transcended. Without a doubt, we see this same urge to identify artistic production with a peculiarly Japanese understanding of the world in the work of the *kokugakusha* of the eighteenth century, including Kamo no Mabuchi, whose work especially on the *Man'yōshū* asserts the centrality of indigenous thinking, and Motoori Norinaga, perhaps the greatest of all Japanese literary critics, who champions a Japanese artistic perspective that stands apart from Chinese notions of didacticism and morality, stressing both the intrinsic essence of things in the natural world as well as their effect on an observer.

One of the resulting tensions produced by these views is the paradoxical irony of asserting on the one hand the ubiquity of art and its universal nature as exemplified by Japanese art, while simultaneously arguing for its singular Japanese composition and tenor. We encounter this tension in the work of a number of modern

philosophers who were engaged by questions of aesthetics in the context of phenomenology and hermeneutics, including Kuki Shūzō (in *Iki no kōzō(The Structure of Iki)*) and Watsuji Tetsurō (in *Fūdo (Climate and Environment)*). Kuki saw in the Edo ideal of *iki* a configuration, not merely aesthetic or ideological, which quintessentially mapped out Japanese being along the continuum of possibilities. For Watsuji, whose work also pays homage to the thought of Martin Heidegger, the resolution of this irony depends on understanding all cultural production as emanating from a particular geographical context; in short, he argues that place is as key as time in understanding human being. In other words, we are shaped ineluctably by the land into which we are born and by our society, a view clearly in accord with those articulated in the early works of Japanese art.

See also: Aesthetics; Aesthetics, Chinese; Kokoro; Nature, aesthetic appreciation of

MEERA VISWANATHAN

References and further reading

Brower, R. and Miner, E. (1961) *Japanese Court Poetry*, Stanford, CA: Stanford University Press.(Seminal study of classical poetry approached historically.)

Keene, D. (1988) *The Pleasures of Japanese Literature*, New York: Columbia University Press, 1-22.(A collection of thoughtful essays based on lectures on Japanese traditional aesthetics and pre-modern poetry, fiction and theatre.)

Konishi Jin'ichi (1984) *A History of Japanese Literature*, volume 1, *The Archaic and Ancient Ages*, trans. A. Gatten and N. Teele, Princeton, NJ: Princeton University Press.(Brilliant and provocative literary history of early Japan.)

Konishi Jin'ichi (1986) *A History of Japanese Literature*, volume 2, *The Early Middle Ages*, trans. A. Gatten, Princeton, NJ: Princeton University Press.(Continuation of Konishi (1984).)

LaFleur, W.R. (1983) *The Karma of Words: Buddhism and the Literary Arts in Medieval Japan*, Berkeley, CA: University of California Press.(Focuses on Chōmei's *Hōjōki*, *Nō* and *kyōgen* theatre and poetry.)

Miner, E., Hiroko Odagiri and Morrell, R.E. (1985) *The Princeton Companion to Classical Japanese Literature*, Princeton, NJ: Princeton University Press.(Extremely useful compendium.)

Okakura Kakuzō (1903) *Ideals of the East*, London: John Murray.(In English, based on Okakura's extensive travels in Asia, deals with traditional Asian art and thought beginning in China and India and reaching its ultimate flowering, he argues, in Japan. Okakura is also commonly known as Tenshin.)

Okakura Kakuzō (1906) *The Book of Tea*, London: Penguin.(In English, this is Okakura's most famous work. He begins with a comparative history of tea-drinking and preparation, and goes on to elucidate the philosophy of tea culture as it developed in Japan, which in his view underlies traditional Japanese aesthetics.)

Rimer, J.T. and Yamazaki Masakazu (1984) *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press.(The introduction is a useful guide to Zeami's principal ideas.)

Slawson, D.A. (1987) *Secret Teachings in the Art of Japanese Gardens*, Tokyo: Kodansha International.(A how-to manual of Japanese garden design philosophy for contemporary readers, derived from a variety of pre-modern treatises including the Heian period *Sakuleiki*.)

Suzuki, D.T. (1959) *Zen and Japanese Culture*, Princeton, NJ: Bollingen-Princeton University Press.(An imaginative and insightful explanation of how Zen principles permeate Japanese art and thought by one of the great proponents of our century; the section on 'Zen and Swordsmanship' is especially pertinent.)

Tanizaki Jun'ichirō (1933-4) 'In'ei Raisan' (In Praise of Shadows), *Keizi Orai*; trans. T.J. Harper and E.G. Seidensticker, Rutland, VT, and Tokyo: Charles E. Tuttle, 1984.(A fascinating and highly idiosyncratic essay on the aesthetic polarities of Japan and the West. He argues that whereas the West prizes light and clarity, Japan privileges darkness, shadows and the elliptical.)

Tsurayuki, Ki no (ed.) (905) *Kokinwakashū*, ed. Saeki Umetomo, Nihon Koten Bungaku Taikei, Tokyo: Iwanami Shoten, 1958; trans. L.R. Rodd and M.C. Henkenius, *Kokinshū*, Princeton, NJ: Princeton University Press, 1984.(First imperially commissioned anthology of Japanese poetry whose chief compiler was Tsurayuki; his well-known preface raises critical concerns alluded to by later theorists.)

Ueda Makoto (1967) *Literary and Art Theories in Japan*, Cleveland, OH: Press of Western Reserve University. (Comprehensive overview of the theories of a number of different theorists from a variety of periods and disciplines.)

Ueda Makoto (1983) *Modern Japanese Poets and the Nature of Literature*, Stanford, CA: Stanford University Press.(Insightful study of eight major poets and their theories of art.)

Affirmative action

The term 'affirmative action' originated in the USA under President Kennedy. Originally it was designed to ensure that employees and applicants for jobs with government contractors did not suffer discrimination. Within a year, however, 'affirmative action' was used to refer to policies aimed at compensating African-Americans for unjust racial discrimination, and at improving their opportunities to gain employment. An important implication of this shift was that affirmative action came to mean preferential treatment.

Preferential treatment was later extended to include women as well as other disadvantaged racial and ethnic groups. The arguments in favour of preferential treatment can be usefully classified as backward-looking and forward-looking. Backward-looking arguments rely on the claim that preferential treatment of women and disadvantaged racial minorities compensates these groups or the members for the discrimination and injustices they have suffered. Forward-looking arguments rely on their claim that preferential treatment of women and disadvantaged racial minorities will help to bring about a better society.

There has been much criticism of both types of argument. The most common accusation is that preferential treatment is reverse discrimination. Other criticisms are based around who exactly should be compensated, by what means and to what extent, and at whose cost. Finally, there is the fear of the unknown consequences of such action. Arguments have been forwarded to try and solve such difficulties, but the future of preferential treatment seems to lie in a combination of the two arguments.

1 Backward-looking arguments

In the USA, Native Americans and African-Americans are the best examples of disadvantaged racial minorities that have been treated unjustly. One kind of backward-looking argument claims that the members of these groups suffer from ongoing racial discrimination practised by the white majority, and also from the effects of past injustices that the nation practised against their parents and ancestors; that they therefore deserve to be compensated; and that according them some preference over whites in the competition for jobs, promotions and places at universities is an appropriate way to give them the compensation they deserve. A second kind of backward-looking argument adds that preferential treatment of Native Americans and African-Americans is also a way to compensate the groups to which they belong. Both kinds of arguments are extended with somewhat diminished force to justify preferential treatment of women and other racial minorities that have been unjustly discriminated against.

2 Forward-looking arguments

Forward-looking arguments for preferential treatment do not require that preferentially treated individuals be themselves the victims of injustice. They justify treating some individuals preferentially if this will help make society more efficient, and more likely to give equal consideration to the common interests of its members. For example, preferential treatment of women in fields like engineering may make society more efficient by encouraging women with engineering talent to develop and use it. Similarly, preferential admission of African-Americans to medical school may help enable society if African-American doctors are more likely than white doctors to practise medicine in black ghettos where medical care is relatively scarce.

3 Criticisms of backward-looking arguments

The most general criticism of preferential treatment is that it is reverse discrimination. The implication is that preferential treatment is morally similar to the discrimination it is meant to compensate for. This implication is false. The discrimination preferential treatment is meant to compensate for is based on contempt for the interests or abilities of those discriminated against (see Discrimination). Preferential treatment is not based on such contempt. More specific criticisms of preferential treatment are directed at the backward-looking and forward-looking arguments.

Since backward-looking arguments for preferential treatment claim that it is required by compensatory justice, they presuppose a view of what compensatory justice requires. The most elementary requirement of compensatory justice is that those that deserve compensation must have been wrongly injured. Some critics of preferential

treatment object that it cannot be justified by compensatory justice because this is not usually the case. When directed against preferential treatment of African-Americans this objection is based on two grounds: an underestimation of the effects of racial discrimination and prejudice; and a false inference based on the fact that the black middle-class beneficiaries of preferential treatment are probably less injured than blacks in the lower and under classes to the conclusion that the former are only slightly injured or not injured at all.

A more serious question concerns who should bear the costs of compensating the beneficiaries of preferential treatment. A plausible view is that the costs of compensating the victims of wrongful injury should be borne by those responsible for the wrongful injury. Many critics object that even if practically everyone in the society is at least indirectly responsible for wrongly injuring the beneficiaries of preferential treatment, the policies implemented in its name usually seem to impose the heaviest costs on those least responsible. A similar objection can be raised if compensatory justice allows that the costs of compensating the victims of wrongful injury may have to be borne by the beneficiaries of the wrongful injury. The most troubling difficulty, however, is that we usually cannot know whether the beneficiaries of preferential treatment are getting the compensation they deserve. A plausible view of compensatory justice is that it requires that the wrongly injured persons be brought up to 'the level of wealth and welfare that they would now have if they had not been disadvantaged' (Nickel 1975: 536). On this account, beneficiaries are compensated presumably if compensatory justice secures them jobs and positions roughly similar to those they would have secured in the absence of injury. But preferential treatment does not obviously secure its beneficiaries the jobs and positions they would have recovered in the absense of injury. Consider, for example, the black beneficiaries of preferential treatment. If unjust racial discrimination had never happened, the conditions and prospects for blacks would be very different from what they are. For example, many more blacks, probably including the beneficiaries of preferential treatment, would be better qualified, and certainly many would be chosen for the jobs and places which preferential treatment secures. What is uncertain is that the black beneficiaries of preferential treatment would be the very ones chosen for these jobs and positions.

This difficulty is not insurmountable, but it has moved some advocates of backward-looking arguments of preferential treatment to stress that the larger aim is to compensate unjustly treated groups. On this account, it does not matter that the individuals who get jobs and places as a result of preferential treatment may not be the ones who would get these jobs and places in the absence of injustice directed at the groups they belong to. What matters is that their getting such jobs and places is a way to compensate these groups. This shift in the backward-looking argument raises a number of questions. Even if the typical beneficiaries of preferential treatment have been unjustly treated, it does not follow that the groups to which they belong are owed compensation; not every group of unjustly treated individuals need be owed compensation over and above the compensation owed the individuals that compose it. Indeed, not every group of unjustly treated individuals is the kind of group that can claim compensation. It is widely acknowledged that certain kinds of groups can claim compensation; nation states, firms and families are examples. But it is not obvious that the groups whose members benefit from typical programmes of preferential treatment are those kinds of groups. This difficulty can probably be resolved with respect to groups like African-Americans and Native Americans, groups that most resemble those that theorists acknowledge can meaningfully be owed compensation. Supposing this to be the case, this still leaves the difficulty of establishing the conditions that have to be satisfied to compensate such groups. Some theorists doubt that preferential treatment as standardly practised satisfies these conditions, given that it seems to contribute to the growing gap between the black middle-class and the black underclass (Wilson 1987). Presumably this objection can be met by redesigning present policies of preferential treatment. More general questions have been raised about the level of wealth and wellbeing preferentially treated groups would have to be brought to in order to be compensated. The assumption that this level is the level of flourishing groups in the society has been challenged on the ground that cultural differences between groups may explain most of the inequalities in their levels of wealth and wellbeing. However, even if this challenge is generally sound it does not defeat the claim that some preferentially treated groups deserve compensation.

4 Criticisms of forward-looking arguments

Forward-looking arguments do not have to resolve the difficult counterfactual problems that beset the backward-looking arguments. A common criticism of forward-looking arguments is that since they justify discriminating in favour of minorities and women when doing so maximizes utility, they must be committed to discriminating against women and minorities when doing so maximizes utility. This criticism is, however, largely

irrelevant because most advocates of the forward-looking arguments do not rely heavily on the supposition that preferential treatment maximizes utility; their more favoured supposition is that preferential treatment will enable society to give more equal consideration to the common interests of its members. The example given earlier concerning the treatment of women in fields such as engineering suggests the plausibility of this supposition. A frequent objection is that even if this supposition is plausible, preferential treatment is unjustified because it violates colour-blind or gender-blind principles. These principles forbid denying an individual a place, job or promotion on account of their colour or gender. They are plausibly implied by the equal opportunity principle, assuming that individuals have not previously been denied opportunities to acquire qualifications for places, jobs and promotions on account of their colour or gender. That assumption is false where preferential treatment is urged. However, although preferential treatment violates colour-blind and gender-blind principles it need not violate the equal opportunity principle. On the contrary it may help to implement that principle. For example, preferential admission of women in fields such as engineering may help society equalize opportunities by helping to break down stereotypes that falsely suggest that women cannot perform well in such fields. A deeper weakness of the objection is that it falsely assumes that the purpose of providing opportunities for places and jobs is to reward merit. In fact, the purpose of providing opportunities for places and jobs is to satisfy the needs and give equal consideration to the interests of members of the society. Meeting that purpose may require violating the colour-blind and gender-blind principles.

Further objections to the forward-looking arguments focus on the consequences of preferential treatment. One set of objections denies that it will have the good consequences its advocates predict. A crude example of this kind of objection is that we do not equalize the interests of the black poor in medical treatment by certifying unqualified blacks as doctors to treat them. A more serious objection denies that there is any good reason to suppose that black doctors are more likely than white doctors to work among the black poor. Another set of objections maintain that preferential treatment is likely to have some untoward consequences. Favourite arguments are that it will create the stereotype that women and minorities cannot succeed in competition with white males without special help, and that it will undermine the self-esteem and self-respect of those it sets out to benefit. These dangers seem most likely where the beneficiaries of preferential treatment misunderstand its rationale.

5 Conclusion

The case for preferential treatment remains highly controversial. Philosophers disagree about what its consequences are likely to be, and about who has been injured enough to deserve it. More importantly, they also disagree about the requirements of compensatory justice, the demands of equality and the nature of the good society. The case for preferential treatment of such groups can only be strengthened by grounding it explicitly on well-argued answers to these philosophical questions.

See also: Equality; Justice

BERNARD BOXILL

References and further reading

Boxill, B. (1992) *Blacks and Social Justice*, Lanham, MD: Rowman & Littlefield, revised edn.(Chapters 7 and 11 contain extended defences of affirmative action.)

Cahn, S. (1993) *Affirmative Action and the University*, Philadelphia, PA: Temple University Press.(A collection of recent, mainly critical, discussions of affirmative action in higher education.)

Chapman, J. (ed.) (1991) *Compensatory Justice*, New York: New York University Press.(Contains sophisticated discussions of compensatory justice.)

Dworkin, R. (1978) 'Reverse Discrimination', in *Taking Rights Seriously*, Cambridge, MA: Harvard University Press.(This essay is a famous interpretation of the forward-looking argument for affirmative action.)

Ezorsky, G. (1991) *Racism and Justice*, Ithaca, NY: Cornell University Press.(A recent persuasive defence of affirmative action.)

Hill, E., Jr (1991) 'The Message of Affirmative Action', in E.F. Paul, F.D. Miller and J. Paul (eds) *Reassessing Civil Rights*, Cambridge, MA: Harvard University Press.(Defends affirmative action in the university and argues that it communicates a message to its beneficiaries, acknowledging that they have been wronged and welcoming them into the university.)

Lawson, P. (1992) *The Underclass Question*, Philadelphia, PA: Temple University Press.(In this collection, several African-American philosophers discuss a variety of recent criticisms of affirmative action.)

McGary, H. (1989) ‘Reparations, Self-Respect and Public Policy’, in D. Goldberg (ed.) *Ethical Theory and Social Issues*, Fort Worth, TX: Holt.(An excellent defence of backward-looking arguments for affirmative action.)

Nickel, J.W. (1975) ‘Preferential Policies in Hiring and Admissions: A Jurisprudential Approach’, *Columbia Law Review* 75 (April): 534-58.(Assesses the affirmative action debate from the point of view of the law.)

Wilson, W.J. (1987) *The Truly Disadvantaged*, Chicago, IL: University of Chicago Press.(Contains arguments that affirmative action rewards African-Americans who are already advantaged, and adds to the isolation of those who are truly disadvantaged.)

African philosophy

In order to indicate the range of some of the kinds of material that must be included in a discussion of philosophy in Africa, it is as well to begin by recalling some of the history of Western philosophy. It is something of an irony that Socrates, the first major philosopher in the Western tradition, is known to us entirely for oral arguments imputed to him by his student Plato. For the Western philosophical tradition is, above all else, a tradition of texts. While there are some important ancient philosophers, like Socrates, who are largely known to us through the reports of others, the tradition has developed increasingly as one which pays careful attention to written arguments. However, many of those arguments - in ethics and politics, metaphysics and epistemology, aesthetics and the whole host of other major subdivisions of the subject - concern questions about which many people in many cultures have talked and many, although substantially fewer, have written outside of the broad tradition of Western philosophy. The result is that while those methods of philosophy that have developed in the West through thoughtful analysis of texts are not found everywhere, we are likely to find in every human culture opinions about some of the major questions of Western philosophy. On these important questions there have been discussions in most cultures since the earliest human societies. These constitute what has sometimes been called a ‘folk-philosophy’. It is hard to say much about those opinions and discussions in places where they have not been written down. However, we are able to find some evidence of the character of these views in such areas as parts of sub-Saharan Africa where writing was introduced into oral cultures over the last few centuries.

As a result, discussions of African philosophy should include both material on some oral cultures and rather more on the philosophical work that has been done in literate traditions on the African continent, including those that have developed since the introduction of Western philosophical training there.

1 Oral cultures

Two areas of folk-philosophy have been the object of extended scholarly investigation in the late twentieth century: the philosophical psychology of people who speak the Akan languages of the west African littoral (now Ghana) (see Akan philosophical psychology) and the epistemological thought of Yoruba-speaking people of western Nigeria (see Yoruba epistemology). In both cases the folk ideas of the tradition have been addressed by contemporary speakers of the language with Western philosophical training. This is probably the most philosophically sophisticated work that has been carried out in the general field of the philosophical study of folk-philosophy in Africa. It also offers some insight into ways of thinking about both the mind and human cognition that are different from those that are most familiar within the Western tradition.

One can also learn a great deal by looking more generally at ethical and aesthetic thought, since in all parts of the continent, philosophical issues concerning evaluation were discussed and views developed before writing (see Aesthetics, African; Ethical systems, African). Philosophical work on ethics is more developed than in aesthetics and some of the most interesting recent work in African aesthetics also focuses on Yoruba concepts which have been explored in some detail by Western philosophers. The discussion of the status of such work has largely proceeded under the rubric of the debate about ethnophilosophy, a term intended to cover philosophical work that aims to explore folk philosophies in a systematic manner (see Ethnophilosophy, African). Finally, there has also been an important philosophical debate about the character of traditional religious thought in Africa (see African traditional religions).

2 Older literate traditions

Although these oral traditions represent old forms of thought, the actual traditions under discussion are not as old as the remaining African literate traditions. The earliest of these is in the writings associated with the ancient civilizations of Egypt, which substantially predate the pre-Socratic philosophers who inhabit the earliest official history of Western philosophy (see Egyptian cosmology, ancient). The relationship between these Egyptian traditions and the beginnings of Western philosophy have been in some dispute and there is much recent scholarship on the influence of Egyptian on classical Greek thought (see Egyptian philosophy: influence on ancient Greek thought).

Later African philosophy looks more familiar to those who have studied the conventional history of Western

philosophy: the literate traditions of Ethiopia, for example, which can be seen in the context of a long (if modest) tradition of philosophical writing in the horn of Africa. The high point of such writing has been the work of the seventeenth-century philosopher, Zar'a Ya'ecob. His work has been compared to that of Descartes (see Ethiopia, philosophy in).

It is also worth observing that many of the traditions of Islamic philosophy were either the product of, or were subject to the influence of scholars born or working in the African continent in centres of learning such as Cairo and Timbuktu (see Islamic philosophy). Similarly, the work of some of the most important philosophers among the Christian Church Fathers, was the product of scholars born in Africa, like St Augustine, and some was written in the African provinces of Rome.

In considering African-born philosophers, there is Anton Wilhelm Amo, who was born in what is now Ghana and received, as the result of an extraordinary sequence of events, philosophical training during the period of German Enlightenment, before returning to the Guinea coast to die in the place he was born. Amo's considerable intellectual achievements played an important part in eighteenth- and nineteenth-century polemics relating to the 'capacity of the negro'. Unfortunately, only a portion of his work has survived.

3 Recent philosophy

Most work in African philosophy in the twentieth century has been carried out by African intellectuals (often interacting with scholars outside Africa) under the influence of philosophical traditions from the European countries that colonized Africa and created her modern system of education. As the colonial systems of education were different, it is helpful to think of this work as belonging to two broadly differentiated traditions, one Francophone and the other Anglophone. While it is true that philosophers in the areas influenced by French (and Francophone Belgian) colonization developed separately from those areas under British colonial control, a comparison of their work reveals that there has been a substantial cross-flow between them (as there generally has been between philosophy in the French- and English-speaking worlds). The other important colonial power in Africa was Portugal whose commitment to colonial education was less developed. The sole Portuguese-speaking African intellectual who made a significant philosophical contribution is Amílcar Cabral, whose leadership in the independence movement of Guinea Bissau and the Cape Verde islands was guided by philosophical training influenced by Portuguese Marxism. Cabral's influence has not been as great as that of Frantz Fanon. He was born in the French Antilles, but later became an Algerian. He was a very important figure in the development of political philosophy in Africa (and much of the Third World).

Among the most important political thinkers influenced by philosophy are Kwame Nkrumah, Kenneth Kaunda and Julius Nyerere (see African philosophy, Anglophone). Out of all the intellectual movements in Africa in this century, the two most important ones of philosophical interest have been *négritude* and pan-Africanism (see African philosophy, Francophone; Pan-Africanism).

Philosophy in Africa has changed greatly in the decades since the Second World War and, even more, as African states have gained their independence. Given the significance of the colonial legacy in shaping modern philosophical education in Africa it is not surprising that there have been serious debates about the proper understanding of what it is for a philosophy to be African. These lively debates, prevalent in the areas of African epistemology, ethics and aesthetics, are found in both Francophone and Anglophone philosophy (see Aesthetics, African).

See also: Marginality; Postcolonialism

K. ANTHONY APPIAH

References and further reading

Abraham, W.E. (1962) *The Mind of Africa*, Chicago: University of Chicago Press.(A classical discussion of the place of philosophical ideas in African life.)

Eze, E. (1997) *Postcolonial African Philosophy: a Critical Reader*, Cambridge, MA: Basil Blackwell.(A useful collection of important work in contemporary African philosophy.)

Hountondji, P.J. (1983) *African Philosophy: Myth and Reality*, London: Hutchinson. (The most important Francophone discussion of the project of modern African philosophy.)

Masolo, D.A. (1994) *African Philosophy in Search of Identity*, Edinburgh: Edinburgh University Press.(A critical history of modern African philosophy.)

African philosophy, Anglophone

Contemporary African philosophy is in a state of flux, but the flow is not without some watersheds. The chief reason for the flux lies in the fact that Africa, in most part, is in a state of transition from a traditional condition to a modernized one. Philosophically and in other ways, the achievement of independence was the most significant landmark in this transition. Independence from European rule (which began in Libya in 1951, followed by Sudan in 1956, Ghana in 1957 and continued to be won at a rapid pace in other parts of Africa in the 1960s) did not come without a struggle. That struggle was, of necessity, both political and cultural. Colonialism involved not only political subjection but also cultural depersonalization. Accordingly, at independence it was strongly felt that plans for political and economic reconstruction should reflect the needs not only for modernization but also for cultural regeneration. These are desiderata which, while not incompatible in principle, are difficult to harmonize in practice. The philosophical basis of the project had first to be worked out and this was attempted by the first wave of post-independence leaders. The task of devising technical philosophies cognizant of Africa's past and present and oriented to her long-term future has been in the hands of a crop of professional philosophers trained in Western-style educational institutions. Philosophical results have not been as dramatic as in the case of the political, but the process is ongoing.

The political figures that led African states to independence were not all philosophers by original inclination or training. To start with only the best known, such as Leopold Senghor of Senegal, or Kwame Nkrumah of Ghana, were trained philosophers, but others, such as Kenneth Kaunda of Zambia, brought only an educated intelligence and a good sense of their national situations to the enterprise. In all cases they were rulers enthusiastically anointed by their people to chart the new course and lead them to the promised land. An example of how practical urgency can inspire philosophical productivity can be found in the way that all these philosophers propounded blueprints for reconstruction with clearly articulated philosophical underpinnings. Circumstantial necessity, then, rather than Platonic selection made these leaders philosopher-kings. It is significant, also, to note that all the leaders mentioned (and the majority of their peers) argued for a system of socialism deriving from their understandings of African traditional thought and practice, and from their perceptions of the imperatives generated by industrialization, such as it had been. Concern with this latter aspect of the situation led to some flirtation and even outright marriage with Marxism. But, according to the leaders concerned, the outcome of this fertilization of thought had enough African input to be regarded as an African progeny. Accordingly, practically all of them proffered their theories and prescriptions under the rubric of African socialism. No such labelling is possible in the work of African philosophers, but there are some patterns of preoccupation.

1 The epistemological anthropology of *négritude*

Of all the African leaders under study, Senghor, poet and man of letters, is perhaps the most learned and most remarkable in his views. He is also the most famous champion of *négritude*. This term refers to both a literary movement and its defining outlook. *Négritude* was focused on restoring in black people the pride in their being and culture that had been eclipsed by colonization. The pioneer of *négritude* was Aimé Césaire, the black poet and playwright from Martinique. In the hands of Senghor, *négritude* also became a kind of epistemological anthropology and a political philosophy.

According to Senghor, *négritude* designates 'the whole complex of civilized values - cultural, economic, social and political which characterize the black peoples or, more precisely, the Negro-African world' (Mutiso and Rohio 1975: 83). The character of these values emerges at the social, and more fundamentally, epistemological levels. Socially, the key to these values was held to consist in what Senghor calls the 'communal' characteristic of African society. Countries exhibiting that kind of social formation are ones in which 'the group takes precedence over the individual'; they are 'above all, religious countries where money is not king' (Senghor 1965: 58). Senghor stresses the importance of the institution of the family, which he says is the 'microcosm' of this kind of society. As he points out, 'family' in this context has to be understood in a non-Western sense as referring to a kinship unit including 'all persons, living and dead, who acknowledge a common ancestor' (1965: 48). Senghor suggests that this unit is better called a clan following anthropological usage. In a typical traditional village or town it would be more numerous and at the level of a nation, innumerable. In its smallest proportions, then, a 'family' in this sense is a substantial community and provides, as the immediate context of early socialization, a natural school for the

cultivation of a broad sense of social belonging and obligation. ‘The African’, says Senghor, ‘is thus held in a tight network of vertical and horizontal communities which bind and at the same time support him’ (1965: 43). More significantly, Senghor adds, ‘ He is the fullest illustration of the truth, honored in our time by socialism, that man can only live and realise himself in and through society’ (1965: 43).

However, Senghor does not suggest that Africans devalue individuality: the African ‘claims his autonomy… to affirm himself as a being’ (1964: 94). But, the logical point to be noted here is that in the African scheme of things individuality is defined in terms of community, not vice versa. In consequence, the African approach to self-consciousness is non-Cartesian. Instead of ‘I think, therefore I am’, he or she, according to Senghor, would say (dispensing with ‘the logician’s conjunction "therefore"’ as a mere distraction), ‘I feel, I dance the Other; I am’ (1964: 73). This axiom of communalist self-consciousness became even better known about a decade later in Mbiti’s formulation as ‘I am because we are, and since we are, therefore I am’ ([1969] 1990: 106). However, Senghor’s epistemological reading of this mode of self-affirmation remains uniquely his own. To him it was typical of the ‘emotive’ and ‘participatory’ character of African cognition, which he considered to be in marked contrast with Western ways of knowing. He wrote:

> Let us consider the negro-African as he faces the object to be known, as he faces the Other: God, man, animal, tree or pebble, natural or social phenomenon. In contrast to the classic European, the African negro does not draw a line between himself and the object; he does not hold it at a distance, nor does he merely look at it and analyse it. More exactly, after having held it at a distance, after having scanned it without analysing it, he takes it vibrant in his hands, careful not to kill or fix it. He touches it, feels it, smells it … Thus the negro African sympathises, abandons his personality to become identified with the other.
>
> (Senghor 1964: 72)

Senghor actually maintained that these ‘modes of knowledge’ or ‘forms of thought (are) different and linked to the psycho-physiology of each race’ (1965: 33).

Senghor attached great importance to this account of African cognition because he thought that it explained ‘the cultural values of the Africans… their religion and social structure, their art and literature, above all the genius of their languages’ (1965: 35). Thus the communalist cast of African society is a social manifestation of the sense of community which the African feels with the whole of creation. This manifestation traditionally took the form of social arrangements of mutual caring and support which ensured for the individual a reasonable amount of wellbeing. A key feature of this system was the combined individual and clan ownership of the means of production (mainly land) and the products of labour, which, in the opinion of Senghor, made African society ‘collectivist’ or ‘socialist’. To Senghor, the task facing contemporary African political thought was how to capture this pristine socialism in modern social and political institutions. In the event he found little rationale for extensive nationalization, and his prescriptions and their implementation were only remotely analogous to socialism customarily understood.

From a philosophical point of view the most interesting questions relate to Senghor’s theory of ‘forms of thought’. Recalling his claim that: ‘European reasoning is analytical, discursive by utilization; African reasoning is intuitive by participation’ (1964: 73), the following questions may be asked: Do these characterizations represent distinctive cognitive categories? Are they, can they be, physiologically ingrained? And is the racial apportionment justified? In response to outcries from some African intellectuals scandalized by this apparent attribution of a constitutional incapacity for analysis to the African psyche, Senghor, notwithstanding protestations to the contrary, changed his position substantially: ‘In truth every ethnic group possesses, along with different aspects of Reason, all the virtues of man, but each has stressed only one aspect of Reason, only certain virtues’ (1964: 75). He even called for the integration of cognitive methods.

A striking feature of Senghor’s discussions of the philosophical bases of his ideological recommendations is his frequent grappling with Marxism. He is highly impressed by the intellectual power of Marx and much taken with Marx’s dialectical method, which he somehow believed was in harmony with African ways of thinking: ‘Negro African reason is traditionally dialectical, transcending the principles of identity, non-contradiction and the "excluded middle"’ (1964: 75). But there is a pronounced ambivalence with regard to certain elements of the Marxist construct. These were the materialistic aspect of dialectical materialism, its atheism and what he perceived to be its determinism, all of which, in themselves, he regarded as objectionable and incompatible with the

traditional African worldview. Moreover, he did not think that the class struggle was a necessary factor in the African quest for a contemporary form of socialism.

2 Metaphysics and African socialism

These rejections of Marxist materialism, atheism and determinism generally came to be taken as the marks by which to distinguish the philosophy of African socialism from the Marxist variety. However, these marks of distinction were not always applicable. For example, the philosophy of Kwame Nkrumah of Ghana, the major architect of Africa's victories in her anti-colonial struggles of this century, was an eminent counterexample in at least two respects. First, although Nkrumah was not an atheist, materialism appealed to him. Second, although he was not initially enthusiastic about the necessity for a class struggle in Africa, he later changed his mind. These differences are symptomatic of a deeper difference in philosophical outlook and ideological commitment. Senghor's appreciation of Marx was theoretical rather than ideological. Nkrumah's, on the other hand, was far-reaching in both its theoretical and ideological aspects. Furthermore, Nkrumah had an emphatically neo-Marxist notion, not apparent in Senghor, of the intimate relationship of abstract philosophy with political practice. This is seen in his book *Consciencism: Philosophy and Ideology for De-Colonization* ([1964] 1970) in assertions like 'Idealism favours an oligarchy, materialism favours an egalitarianism' ([1964] 1970: 75), which illustrate the predominant tendency of the interpretations of the history of Western philosophy with which he prefaced his ideological affirmations. But by far the most philosophically interesting difference between the two philosopher-kings is the fact that, while Senghor believed that materialism (dialectical or otherwise) was incompatible with religion, Nkrumah did not (see Dialectical materialism).

Thus, very early in his life as an anti-colonial leader in Ghana Nkrumah proclaimed in a public lecture, 'I am a Marxist Socialist and a nondenominational Christian, and I see no contradiction in this'. In view of his claims of African authenticity for his theory, he also might have added, 'Moreover, I believe in the essentials of the African traditional worldview, and I see no contradiction in this either'. The way in which he defended his belief that the dialectical materialism of Marxism is consistent with the theism of Christianity and the metaphysics of the traditional African worldview was to deploy an ingenious distinction between materialism as the theory of the *sole* reality of matter and the *primary* reality of matter. In Nkrumah's view, the first variety of materialism is injudicious as it conflicts with both fact of mind and the spiritual aspects of human experience. On the other hand, a materialism of the second persuasion can accommodate these previously recalcitrant facts, provided it has an intelligible and valid account of the emergence of mind from matter. Dialectical materialism is, according to Nkrumah, of the second type and he invented for it just such an account: 'The key to the solution of the problem', he explained, 'lies in categorial convertibility' ([1964] 1970: 20). Categorial conversion was defined not by direct specification, but almost recursively, by cases: 'By categorial conversion, I mean such a thing as the emergence of self-consciousness from that which is not self-conscious: such a thing as the emergence of mind from matter, of quality from quantity' ([1964] 1970: 20). As an aid to the understanding of these category transitions, *Consciencism: Philosophy and Ideology for De-Colonization* says that philosophy turns to science for 'models' and discovers exemplars in 'the inter-reducibility of matter and energy' and in chemical change wherein 'physical quantities give rise to emergent qualities' ([1964] 1970: 21).

For further enlightenment as to the nature of categorial conversion, the reader was directed to the following explanatory comparison: 'The average man belongs to a category distinct from that of the men and women of flesh and blood; but the concept of the average man is obtained by a certain conceptual conversion of information about individual men and women' ([1964] 1970: 22). Here the category of 'living men and women' is primary and that of the average man derivative. Similarly, from the primary category of matter we can arrive at the derivative category of mind by a logical processing of data about 'nervous' matter. Propositions about the mind are then seen to be 'materially equivalent' to propositions about 'a critical organization of matter'. In this way the categorial differences are revealed as *façons de parler*.

The problems afflicting this account are difficult to minimize. For example, in so far as the average-man illustration is germane, the convertibility involved is a logical relationship. Consequently, phenomena such as the emergence of chemical properties from 'physical quantities' held up as a model must fall beyond the pale of categorial conversion. More gravely, the objective of reconciling dialectical materialism with Christianity, however nondenominational, must entail conceiving of the eternal, supreme, spiritual being of that religion as a

kind of emergence from matter, which would mean reconceiving that being out of all recognition. Consistency, then, cannot be achieved. However, given Nkrumah's faith in categorial conversion, it is easy to understand this problematic catholicity on his part. The same faith in the idea of categorial convertibility, which he held was to be found among African traditional conceptions, enabled him to combine Marxist materialism with his background in African thought. That some very transformative 'conversion' was necessary is apparent from the fact that he believed that 'man is regarded in Africa as primarily a spiritual being' ([1964] 1970: 68).

This African connection was as important to Nkrumah as it was to Senghor. To Nkrumah it showed that his appropriation of the Marxist philosophy (and in general his explorations of Western philosophy), did not compromise his African authenticity. With the attainment of independence, Nkrumah felt there was need to articulate a philosophy that could harmonize the competing segments that have come to inhabit the African conscience through historical circumstances. These segments derived from the presence in contemporary Africa of influences from African traditional culture and from Islamic and Euro-Christian sources. The synthesis of Marxist philosophy with some Christian and African traditional conceptions, which Nkrumah called philosophical consciencism, was he said, exactly such a philosophy.

Nkrumah gave the following indications of the traditional African ingredients in this philosophical compound: philosophical consciencism 'agrees with the traditional African idea of the absolute and independent existence of matter, the idea of its powers of self-motion… the idea of categorial convertibility, and the idea of the grounding of… ethics in the nature of man' ([1964] 1970: 97). Also, as far as socialism is concerned, 'the traditional face of Africa includes an attitude towards man which can only be described, in its social manifestations, as being socialist' ([1964] 1970: 68). The implied reference is to the communalism of traditional Africa, which according to *Consciencism: Philosophy and Ideology for De-Colonization*, is 'the sociopolitical ancestor of socialism' ([1964] 1970: 73). On the question of the socialistic complexion of traditional communalism Nkrumah and Senghor are at one. They are also in agreement with Julius Nyerere of Tanzania, one of the most respected of Africa's statesman-thinkers. However, they differ with Nyerere in other ways. Both Senghor and Nkrumah seem to make a veritable conceptual linkage between their metaphysics and their socialist ideology. This is not true of Nyerere: 'There is not the slightest necessity for people to study metaphysics and decide whether there is one God or many Gods or no God before they can be socialists. These questions are important to man, but irrelevant to socialism' (1969: 39).

3 An analysis of socialism and a reflection on violence

There is an absence of metaphysical learning in Nyerere's pages. However, the reader does find a highly philosophical approach to the socialist ideology based on an analysis of the traditional communalism of his society and its contemporary condition. Nyerere traces the idea of socialism to its foundation, which he finds in the principle of equality. 'Socialism', he says, 'is, in fact, the application of the principle of human equality to the social, economic and political organization of society' (1968: 79). Broadly construed, this is an equality of benefits. Consequently, for Nyerere, socialism is a distributive dispensation and not primarily a system of production. But he concedes that certain forms of production can lead to the unequal acquisition of wealth on a scale which makes it possible for some people to gain exploitative dominance over others. Therefore a rationally selective public ownership of the means of production can become a means for achieving the basic aim of socialism. However, social ownership ought to be distinguished from social control as the former can be combined with all kinds of despotisms, while the latter, if genuine, cannot. Furthermore, the latter can conceivably be had without the former. Hence, what socialism requires is the social control of certain means of production. This has important political implications. Social control is not conducive to socialism (as the social embodiment of human equality) unless it is exercised by a citizenry enjoying equality of freedom and participation. Therefore, 'Democracy is another essential characteristic of a socialist society', and the rule of law is a part of it: 'until it prevails socialism does not prevail' (1969: 31, 34). Besides, a socialist society in Nyerere's view must be suffused with an ethos of cooperativeness as opposed to personal competitiveness.

There have been expositions of social and political testaments with significant philosophical components by other African political leaders, such as Sekou Touré of Guinea, Obafemi Awolowo of Nigeria, Felix Houphouët-Boigny of the Ivory Coast, Kenneth Kaunda of Zambia and Amílcar Cabral of Guinea-Bissau. They all, except Houphouët-Boigny, advocated varieties of socialism, but none surpassed the conceptual clarity of Nyerere.

Without exception, however, all the socialisms have come under a cloud on account of the uniform failure of the socialist experiments in Africa and elsewhere. Their philosophical components, nevertheless, challenge a properly philosophical evaluation.

Worthy of special mention in connection with the literature produced by Africa's statesman-philosophers is Kaunda's book of meditations on the theme of violence (*Kaunda on Violence* (1980)). The question of the moral legitimacy of violence in the liberation struggles of Africa was the cause of much earnest soul-searching among sensitive people in Africa (Mazrui (1978), Wiredu (1986), Serequeberhan (1991)). The book by Kaunda is an agonized philosophical soliloquy on his personal evolution from an adherent of Gandhian nonviolence to a principal supporter and sustainer of the armed struggle in southern Africa. His considered judgment, after extended reflections articulated with singular lucidity, is that violence is never justifiable morally, but it may be forgiven (by the Lord Almighty) if it is in reaction to the violence of an oppressor. This thoughtful conclusion raises quite subtle issues in moral philosophy of relevance in Africa and everywhere else. But in Africa it also provides a reasoned contrast to the relatively untroubled advocacy of anti-colonial violence in Fanon's *The Wretched of the Earth* (1961), a book which deservedly has a considerable African following (see Fanon, F.).

4 The question of African philosophy in our time

If the challenges of independence provided the direct stimuli of the philosophical enterprises of Africa's political leaders, they also provide the indirect cause of the character and present state of academic philosophy in Africa. This brings us to our second watershed. Pre-independence curricula in philosophy in African universities in the British colonial orbit were unmodified importations from the UK, not to say impositions, without any African admixtures. If mention was made of any African philosophical conception bordering on the philosophical inside those universities, it was likely to emanate from a department of anthropology or religion. Naturally, not long after independence, the general movement towards the reclamation of the African identity made itself felt in philosophy departments in the form of a search for an African orientation in teaching and research in philosophy. Two of the subprojects that have received some attention in the pursuit of this objective are the study of the philosophical ideas embedded in African oral traditions and the utilization of insights from that study in combination with insights from other sources in the contemporary world for the construction of philosophies for modern existence.

To take the first project first: although its necessity is widely recognized, its modalities are enveloped in deepest controversy. Debate has raged principally around the criticisms by Paulin Hountondji (1983) of contemporary studies of traditional African thought that construe it as a continental monolith of philosophical unanimity. Hountondji, the Francophone African philosopher who has most influenced contemporary philosophical discussions in Anglophone Africa, has used the term 'ethnophilosophy' to designate studies of this kind, which he regards not as nonphilosophy (as it is sometimes supposed), but rather as bad philosophy. They are apt, he says, to be 'the description of an implicit, unexpressed worldview, which never existed anywhere but in the anthropologist's imagination' (1983: 63). Among the works Hountondji places in this category are *Bantu Philosophy* (1959) by Placide Tempels, a Belgian missionary and *African Religions and Philosophy* (1969) by John Mbiti, an African theologian. African traditional thought itself, however, according to Hountondji, 'possesses a complexity, a richness and a depth with which we have as yet scant acquaintance, and which we must now recover' (1983: 280). Nevertheless, not being articulated in the form of explicit and systematic expositions, traditional African thought cannot be said to contain the discipline of philosophy, even though its ancestral originators may have been philosophers in their own right. At best, we 'can probably recover philosophical fragments from our oral literature' (1976: 106-7), but the forging of a discipline from this material worthy of the name of philosophy remains a challenge to contemporary Africans that can only be met in close alliance with the quest for scientific knowledge (see Ethnophilosophy, African).

These sentiments have fallen harshly on the sensibilities of the traditionalists among contemporary African philosophers, and the ensuing controversy has seemed interminable to both onlookers and some insiders (Serequeberhan (1991) is an excellent sampling of the exchanges). Nevertheless, some significant work has been done because of that controversy and alongside it, although it is somewhat scattered. Among the most accessible of such work are Gyekye (1987) and Gbadegesin (1991). These books contain detailed interpretative expositions of African traditional philosophy, focusing on the Akan of Ghana and the Yoruba of Nigeria respectively. Both philosophers find in their traditional heritage dualistic but richly stratified conceptions of human personality. They

also find, like the philosopher-kings before them, a communalistic ethic and, contrary to a longstanding orthodoxy, a rationalistic, rather than a supernaturalistic ethics. In philosophical theology they call attention to unaided indigenous postulations of a supreme being, although in Gyekye's interpretation, but not in Gbadegesin's, the conception is substantially similar to the God of Christianity. In either case, their interpretations stand in contrast to the account of Luo traditional religion advanced by Okot P'Bitek, the Ugandan poet, novelist and philosopher, in an earlier phase of contemporary African philosophy. In two books of notable conceptual sophistication, *African Religions in Western Scholarship* (1970) and *Religion of the Central Luo* (1971), P'Bitek argued that the conceptual framework of the Luo has no place for the notion of a supreme being or even for the beginning of the world and that the widely received notions of a Luo supreme being are only thanks to a Western missionary superimposition.

Whatever the truth in this matter, it cannot be doubted that the years of Western leadership in the literature on African thought have left encrustations on indigenous conceptual structures that are due for systematic unscrambling. It is arguable, for example, that the dualistic conception of body and mind, which is often attributed to Africans, in fact, presupposes a mode of conceptualization that ill-coheres with African traditional thought habits, which are frequently empirical, as distinct from empiricist (see Akan philosophical psychology). This suggests a need for conceptual decolonization which, although not in itself a touchstone of philosophical truth, is a necessary preliminary to the accomplishment of the historic tasks facing contemporary African philosophy.

5 Typology of current trends

In addition to traditional thought there have been other objects of attention. In his now famous typology of trends in contemporary African philosophy, the Kenyan philosopher Odera Oruka (1990) discriminates four trends which he lists as (1) ethnophilosophy, (2) philosophic sagacity, (3) nationalist-ideological philosophy and (4) professional philosophy. Subsequently in *Sage Philosophy* (1991) he increased the number to six, adding 'hermeneutic philosophy and artistic or literary philosophy'. By 'artistic and literary philosophy' Oruka means not only the explicit philosophical reflections volunteered from time to time by the creative spirits of contemporary Africa, such as in Wole Soyinka's *Myth, Literature and the African World* (1976), but also the philosophies implicit in their poems, plays, novels and other artistic productions. Oruka's act of inclusion betokens a strong sense of the value of intensive interaction between professional philosophy, on the one hand, and art and literature, on the other, a source of intellectual riches in other traditions.

According to Oruka's scheme of classification the work of the statesman-thinkers falls under nationalist-ideological philosophy while the studies of traditional thought may be said to fall under ethnophilosophy, provided this term is divested of any pejorative connotations. But ethnophilosophy is also generally part of the work of the professionals. The same is true of what Oruka calls 'hermeneutical philosophy'. This he defines as consisting of 'philosophical analysis of concepts in a given African language to help clarify meaning and logical implications' (1990: 11). In illustration of the 'hermeneutical category' he cites Gyekye's *An Essay on African Philosophical Thought* (1987), Wiredu's 'The Concept of Mind with Particular Reference to the Language and Thought of the Akans' (1987) and Hallen's and Sodipo's *Knowledge, Belief and Witchcraft: Analytic Experiments in African Philosophy* (1986).

Of special interest is Gyekye's chapter on 'Philosophy, Logic and the Akan Language'. Disavowing any a priori relativization of philosophical theses to particular languages, Gyekye points out how travel across languages can affect the fate of a philosophical thesis or problem. In one of his examples, he argues that in Akan (the language of one of the ethnic groups of Ghana), the expression *wōhō*, by means of which the concept of existence may be translated, has an irreducibly locative component, carried by the particle *hō*, which means 'at some place'. Given this conceptual situation, he explains, something like St Anselm's ontological argument for the existence of God would be unlikely to enjoy an appearance of plausibility in Akan philosophical discourse. Nor could the controversy, historically precipitated by the Saint's argumentation as to whether existence is a predicate, tempt Akan curiosity, for in Akan terms, the question would reduce to 'something like "Is that something is there an attribute?"', which is 'bizarre' (1987: 179-81) (see Existence). In a similar vein, Kwasi Wiredu in his article, 'The Concept of Mind with Particular Reference to the Language and Thought of the Akans' (1987) argued the point that any conception of mind as some kind of an entity must jar severely on the conceptual framework embedded in Akan thought and talk about things mental. There is no unanimity among contemporary Akan philosophers on

these interpretations, but in matters philosophical this should catch no one by surprise. Furthermore, in neither case are the conceptual disparities revealed taken as proof of philosophical wisdom or its opposite. What they might establish is that certain ways of thought and puzzlement entrenched in Western philosophy are not humanly ineluctable. While it would be premature to invest this reflection with a relativistic significance, the need for a serious examination of relativism is certainly collateral to these types of inquiries.

The book by Hallen and Sodipo (1986) represents something of a methodological innovation in contemporary research in African philosophy (see Yoruba epistemology). The two academic philosophers sought the expertise of a class of Nigerian traditional healers (known as *onisegun*) recognized for their mastery of Yoruba thought and language, by interacting with them on terms of respectful collegiality. They regarded the *onisegun* not as informants, as was customary in academic research, but as fellow philosophers and held long discussions with them on the semantics of certain philosophically-sensitive Yoruba concepts. They then drew upon the results, which they published in faithful transcription, to make comparisons with approximate conceptual counterparts in English. They discovered that the approximations in every case had a significant roughness at their philosophical edges. The best analogy of the knowledge-belief distinction in the Yoruba language is that between *mò* and *gbàgbó*. But *mò* exacts more stringent conditions than knowledge, as the following observation by Hallen and Sodipo indicates:

> *Gbàgbó* that may be verified is *gbàgbó* that may become *mò*. *Gbàgbó* that is not open to verification (testing) and must therefore be evaluated on the basis of justification (*àlàyé*, *papò*, etc.) cannot become *mò* and consequently its *òótó* [truth] must remain indeterminate.
>
> (Hallen and Sodipo 1986: 81)

The contrast with well-known analyses of knowledge in English-speaking philosophy is so striking that Hallen and Sodipo are moved to conclude that ‘propositional attitudes are not universal’. As a brief illustration of one implication of the contrast, it is important to consider the following. In contemporary Anglo-American philosophy it is generally agreed that s knows that p if and only if p is true, s believes that p and s is justified in believing that p under some Gettier-chastened condition(s) (see Justification, epistemic §3). But, however chastened, the justification condition must be too weak for the Yoruba *mò*, since it requires, as is clear from the context, a first-person experiential verification. More interestingly, the belief condition does not survive a Yoruba conversion either. If a person is acknowledged to *mò* p, then to say that they *gbàgbó* p is not just an understatement, but a self-contradictory misstatement, for the use of *gbàgbó* implies that *mò* is unattained. Thus, though *gbàgbó* may ascend to *mò*, the transformation is qualitative and *mò* at the cognitive apex categorially scorns the base degrees by which it did ascend (with apologies to Shakespeare). All this is, of course, on the assumption that Hallen and Sodipo are right in their interpretation, quoted above, of what the *onisegun* say and furthermore that the *onisegun* are right in their account of the conceptual situation in the African vernacular. Without mentioning the complications introduced by the comparative dimension, this already suggests the multifaceted character of the problem of evaluating all ‘hermeneutical studies’ that have a transcultural effect. Nevertheless, the importance of such investigations in contemporary African philosophy can hardly be exaggerated.

6 Indigenous African philosophers

Equally important are investigations into what Oruka calls ‘philosophic sagacity’ or, better, ‘sage philosophy’. Oruka himself is responsible for the path-breaking publication in this area of research. In his *Sage Philosophy* (1991) he gives extensive exposure to the philosophical views of indigenous thinkers in contemporary Kenyan society practically uninfluenced by Western ideas. Particularly noteworthy about their thinking is the fact that they are aware of the traditional thought of their community, but not overawed by it. They put forward their views, like most philosophic thinkers, as their own, not as the community’s. In this they differ from Hallen and Sodipo’s *onisegun*. Although those indigenous experts on Yoruba thought may be bright philosophers, they prefer to throw light on the Yoruba conceptual framework rather than on their own in a personal sense. They abjure intellectual individualism, going as far as to forbid the mention of their names in the published text. Both types of thinkers are known in African society.

To return to Oruka’s philosophic sages: in sketches of their philosophical positions in response to their academic interlocutors, they express with force and lucidity a variety of views on many subjects, including God, religion,

witchcraft, body and mind, virtue, good and evil, truth and falsehood, happiness, life and death, justice, equality, freedom, law, crime, punishment, human suffering, man and woman, ethnicity, and communalism. Perhaps the most intriguing variation of views is to be found in their conceptions of God. Sage Akoko maintains that the nature of God in unknowable, but his existence can be inferred from the uniformity of nature (1991: 37). Chaungo, on the other hand, is convinced that God is the sun. It 'heats the land all day, and its absence cools the land all night. It dries things: plants use it to grow. Surely, it must be the God we talk about' (1991: 115-6). To Kithanje, God is of the nature of a process rather than of an object. His thought about God revolves round 'the mixture of heat and cold'. When these merge 'there comes life… The act of fusion which brings forth life is what we call God. And that is what we mean when we say that God created the universe' (1991: 134). For his part, Oruka Rang'inya contends that it is wrong to personalize God. He is simply 'an idea, the idea which represents goodness itself' (1991: 119). With Njeru wa Kinyenje, outright atheism is finally reached: 'Both religion and witchcraft… have no truths in them' (1991: 38).

In the literature of African thought south of the Sahara such philosophic individuality has, by a long-standing mistake, rarely been ascribed to individual Africans in traditional life. M. Griaule's *Conversations with Ogotemmeli* (1965) is an apparent exception. Ogotemmeli, a Dogon sage from the west African nation of Mali, is clearly shown to be an individual of outstanding speculative abilities. But the intricate web of thought he weaves belongs to his ethnic group, not to him personally. However, since communal thought is a kind of pooling together of the thought of individual thinkers, the Ogotemmeli phenomenon should have given pause to those who were tempted to suppose that traditional Africa was bereft of individuals of philosophic initiative.

With respect to Africa north of the Sahara, it has not been possible for anyone to harbour a like misapprehension. The existence of individual philosophical thinkers in ancient Egypt has never been in doubt, even if their exact racial identity has been the subject of debate. Moreover, the Arab portions of Africa are heir to a time-honoured Islamic tradition of written philosophy. Even further south, it is widely known, through the industry of Claude Sumner of the University of Addis Ababa, that in Ethiopia there is a historical heritage of written philosophy (see Ethiopia, philosophy in §1). The high point of that heritage is *The Treatise of Zar'a Ya'ecob* ([*c*.1599-*c*.1692] 1993). Living and meditating contemporaneously with Descartes, although independent of him, and initially in seclusion (by reason of Catholic persecution), Zar'a Ya'ecob developed a philosophy which affirmed belief in God, but subordinated religious creeds and their moral prescriptions to the dictates of reason. He held that in the search for truth, especially in religious matters, no one should passively depend on the determinations of other people 'for all men are plaintiffs and defendants between themselves' ([*c*.1599-*c*.1692] 1993: 17). Everyone should be guided by the light of their own intelligence. That intelligence is capable of conceiving of God as 'a creator, greater than all creatures' and of 'seeing him mentally'. Out of 'the abundance of his intelligence' God has created a world of natural law and order and given us the intelligence to grasp it if we would but inquire rationally. Accordingly, in his evaluations of religious teachings he proceeds on the principle that something accords with the will of God only if it accords with the deliverances of reason. With this intellectual weapon he makes short work of Christian 'stories of miracles that they claim had been wrought in Egypt and on Mount Sinai' by Moses. Nor is he any less scathing of certain Mohammedan tenets. For example, he believes that Mohammed's teaching 'that a man could marry many wives' could not possibly come from God as it 'ruins the usefulness of marriage' ([*c*.1599-*c*.1692] 1993: 18-19).

In general Ya'ecob's treatment of morality is uncompromisingly rationalistic. For him the golden rule is a directive of reason. Since the 'prohibitions of killing, stealing, lying, adultery' are derivable from it, he says of them: 'our reason teaches us these and similar ones'. In respect of these, therefore, 'the decalogue of the Pentateuch expresses the will of the creator'. But he makes an exception in the case of the commandment of the Sabbath because 'reason says nothing of the observance of the Sabbath' ([*c*.1599-*c*.1692] 1993: 21). Altogether, Ya'ecob sets up a stringent regime of reason in the face of which Descartes might conceivably have balked. Undoubtedly, the work of Ya'ecob is eminent in what might be called the classical heritage of African philosophy.

7 The quest for a synthesis

The study and evaluation of African traditional philosophy and the classical heritage of African philosophy are two of the more straightforward components of the agenda of contemporary African philosophy. Another, infinitely trickier, is the project of synthesizing any insights that might be had from the study of Western philosophy (and

any other philosophical traditions) with those gained from indigenous sources in the construction of African philosophies for contemporary existence (from which open-minded non-Africans might have something to learn). One potent source of ambivalence among some Africans towards this type of effort, as it pertains to Western philosophy, is the fact that it is apt to look like a mindless imitation of the philosophical ways of one's erstwhile colonizers. The best antidote to such qualms is for African researchers to be conscious of their African purposes and always to scrutinize critically the conceptual pertinence - recall the need for conceptual decolonization - and the theoretical soundness of any appropriations of ideas or adaptations of technique. They do not need to approach the enterprise in a spirit of passivity. In principle, they can make original and enriching contributions, at one and the same time, to the given foreign tradition as well as to their own: a fact which is sometimes forgotten. There are, moreover, existential reasons for African interest in Western philosophy. Through the twin historical facts of Western colonization and Christian evangelization, African cultures have been profoundly impregnated with ethical, metaphysical and epistemological ideas of a Western provenance. These ideas cry out for critical examination in Africa as much as in their places of origin. (The principle is applicable, *mutatis mutandis*, to Islamic ideas in Africa.)

Whether in full realization of these considerations or from a semiconscious attunement to the needs and possibilities of the contemporary African situation, many African philosophers have devoted considerable attention to topics that historically have been grist for the Western philosophical mill. The unspoken principle of this practice is that they can be turned to African purposes too. The 1980s and 1990s have seen a great deal of African philosophical output of this sort.

Africans working towards the advancement of African philosophy can ponder the historic precedent of such famous sons of Africa as Tertullian, St Cyprian and St Augustine. These were among the Graeco-Romanized indigenes of colonized North Africa whose thought has influenced Western philosophy, leaving in some cases permanent imprints. They do not seem in their intellectual efforts to have been motivated by the quest for a synthesis of the African and non-African elements in their experience, although it cannot be assumed that their thinking was altogether without African traces. On the contrary, it may well be, as John Ferguson suggests in his essay 'Aspects of Early Christianity in North Africa' (1969), that the character of the Christianity they advocated owed something to their African background. Nevertheless, forging an African tradition of thought does not seem to have been a priority in their concerns. In pursuit of this objective Africans will have to bend any foriegn influences to African purposes. These purposes will sometimes be common to humankind, but sometimes contingent upon the distinctive circumstances of African existence.

See also: Aesthetics, African; African philosophy, Francophone; Ethical systems, African §1; Ethnophilosohy, African

KWASI WIREDU

References and further reading

Appiah, A.K. (1989) *Necessary Questions: An Introduction to Philosophy*, Engelwood Cliffs, NJ: Prentice Hall. (Lucid introduction to most of the principal areas of philosophy with a concluding chapter that is of special African interest.)

Appiah, A.K. (1982) *In My Father's House: Africa in the Philosophy of Culture*, London: Methuen; 2nd edn, Oxford: Oxford University Press, 1992.(Significant, among many things, for the interplay between literature and philosophy in the African intellectual milieu.)

Bonevac, D. and Phillips, S. (eds) (1993) *Understanding Non-Western Philosophy: Introductory Readings*, Mountain View, CA: Mayfield Publishing Company.(An interesting read.)

Danquah, J.B. (1944) *The Akan Doctrine of God: A Fragment of Gold Coast Ethics and Religion*, London: Frank Cass, 1968.(A classic of Akan philosophy expounding an interpretation of Akan perceptions of deity and morality.)

Fanon, F. (1961) *The Wretched of the Earth*, London: Penguin, 1983.(An articulation of a philosophy of liberation and national reconstruction by an African-descended thinker.)

Ferguson, J. (1969) 'Aspects of Early Christianity in North Africa', in J. Ferguson and L. Thompson (eds) *Africa in Classical Antiquity*, Nigeria: Ibadan University Press.(Exploration of the indigenous elements in the thought of prominent African Church Fathers, such as Tertullian and St Augustine in Roman north Africa.)

Fløistad, G. (1987) *Contemporary Philosophy*, vol. 5: *African Philosophy*, Boston, MA: Kluwer.(A collection of

some samples of current work in African philosophy.)
Forde, D. (ed.) (1954) *African Worlds: Studies in the Cosmological Ideas and Social Values of African Peoples*, Oxford: Oxford University Press.(Anthropological discussions of the religion, ethos and ethics of a wide range of African peoples.)
Gbadegesin, S. (1991) *African Philosophy: Traditional Yoruba Philosophy and Contemporary African Realities*, New York: Peter Lang.(In addition to a reflective treatment of Yoruba traditional philosophy, there is a philosophical discussion of cultural, ethical and socioeconomic problems of contemporary Nigerian life.)
Griaule, M. (1965) *Conversations with Ogotemmeli: An Introduction to Dogon Religious Ideas*, London: Oxford University Press.(An historic exposition of the cosmology of the Dogon people of the west African nation of Mali.)
Gyekye, K. (1987) *An Essay on African Philosophical Thought: The Akan Conceptual Scheme*, New York: Cambridge University Press; 2nd edn, Philadelphia, PA: Temple University Press, 1995.(An expository as well as analytical account of the philosophical thought of the Akans of Ghana with some characterization of African philosophy in general.)
Hallen, B. and Sodipo, J.O. (1986) *Knowledge, Belief and Witchcraft: Analytic Experiments in African Philosophy*, London: Ethnographica Ltd.(Inquiries into Yoruba conceptions of belief and knowledge together with clarifications of Yoruba ideas about witchcraft in contrast with some Western ones.)
Hountondji, P. (1983) *African Philosophy: Myth and Reality*, Bloomington, IN: Indiana University Press, 2nd edn, 1996.(Frequently the definitive point of departure in contemporary controversies on the right conception of African philosophy. Stresses the necessity of writing in the discipline of philosophy. Also contains an interesting discussion of A.-W. Amo, the eighteenth-century Ghanaian philosophy professor in Germany, plus an analysis and evaluation of the philosophy and ideology in Nkrumah’s *Consciencism: Philosophy and Ideology for De-Colonization* (1964).)
Idowu, E.B. (1962) *Olodumare: God in Yoruba Belief*, London: Longman.(Famous study of traditional Yoruba perceptions of God, personhood and morality by a Yoruba academic.)
Idowu, E.B. (1973) *African Traditional Religion: A Definition*, London: SCM Press. (Attempts a definition of religion in general and African traditional religion in particular, confronting various philosophical questions.)
Kaunda, K.D. (1980) *Kaunda on Violence*, London: Collins.(Rational soul-searching on the question of the moral legitimacy of violence in the African liberation struggle and in human affairs in general.)
Lienhardt, G. (1961) *Divinity and Experience: The Religion of the Dinka*, Oxford: Oxford University Press.(A conceptually sensitive and philosophically useful study of the religion of the Dinka people of Sudan.)
Mazrui, A.A. (1978) ‘Mahatma Gandhi and Black Nationalism’, in *Political Values and the Educated Class in Africa*, London: Heinemann.(A discussion of the influence of Gandhi’s philosophy of nonviolence on African liberation leaders, such as Nkrumah and Kaunda.)
Mbiti, J. (1969) *African Religions and Philosophy*, London: Heinemann, 2nd edn, 1990.(A survey of religious and philosophical ideas from many parts of Africa. Chapter three gives an account of the concept of time in African thought.)
Mutiso, G.-C.M. and Rohio, S.W. (1975) *Readings in African Political Thought*, London: Heinemann.(A comprehensive and extremely useful anthology of the political thought of Africa’s first wave of post-independence leaders.)
Nkrumah, K. (1964) *Consciencism: Philosophy and Ideology for De-Colonization*, London: Panaf Books, 1970. (The most severely technical of the philosophical writings by African statesmen in support of socialism.)
Nyerere, J.K. (1968) *Ujamaa: Essays on Socialism*, New York: Oxford University Press.(Prescriptive discussions of socialism as a basis of development in Tanzania.)
Nyerere, J.K. (1969) *Nyerere on Socialism*, Dar es Salaam, Tanzania: Oxford University Press.(A combination of the introductions to two previous books on aspects of socialism. Provides abstract analyses of the socialist idea.)
Okere, T. (1983) *African Philosophy: A Historico-Hermeneutical Investigation of the Conditions of its Possibility*, New York: University Press of America.(A hermeneutically-oriented discussion of the nature of philosophy with illustrative explorations of Hegel, Heidegger, Ricoeur and others as a preliminary to hints as to how best African philosophy might be done.)
Oruka, O.H. (1990) *Trends in Contemporary African Philosophy*, Nairobi: Shirikon Publishers.(Analyses and criticisms of trends in contemporary African philosophy by one of the most active contributors to the discipline.)
Oruka, O.H. (1991) *Sage Philosophy: Indigenous Thinkers and the Modern Debate on African Philosophy*, Nairobi: African Center for Technology Studies.(Presents in translation the personal philosophical views of

some thinkers of Kenya traditional society together with discussions of reprinted criticisms of his methodology.)

P'Bitek, O. (1970) *African Religions in Western Scholarship, Prose and Poetry*, London: Oxford University Press. (Briefer extracts in the prose section from writings dealing with a variety of topics, including the ideological and the philosophical.).

P'Bitek, O. (1971) *Religion of the Central Luo*, east African Literature Bureau.(An account of the Luo religion.)

Sawyer, H. (1970) *God, Ancestor or Creator?*, London: Longman.(An inquiry into traditional perceptions of God among the Akans of Ghana, the Yoruba of Nigeria and the Mende of Sierra Leone.)

Second Order (1972-1985), Nigeria: University of Ife Press; new series, Nigeria: Obafemi Awolowo University Press, 1985-.(Served as a vehicle for African philosophical output at a historically critical time.)

Senghor, L.S. (1964) *On African Socialism*, New York: Praeger.(Senghor's ideological prescriptions for Senegal together with their philosophical underpinnings.)

Senghor, L.S. (1965) *Prose and Poetry*, London: Oxford University Press.(Briefer extracts, in the prose section, from writings dealing with a variety of topics, including the ideological and the philosophical.)

Serequeberhan, T. (1991) *African Philosophy: The Essential Readings*, New York: Paragon House.(Anthology of contributions to the debate on the meaning and method of African philosophy from Bodunrin, Hountondji, Keita, Okolo, Onyewuenyi, Oruka, Owomoyela, Serequeberhan, Towa, Wamba-Dia-Wamba and Wiredu.)

Soyinka, W. (1976) *Myth, Literature and the African World*, Cambridge: Cambridge University Press.(Literary and philosophical reflections of one of Africa's foremost writers.)

Sumner, C. (1994) *Classical Ethiopian Philosophy*, Los Angeles, CA: Adey.(A discussion of traditional philosophical systems in Ethiopia.)

Tempels, P. (1959) *Bantu Philosophy*, Paris: Présence Africaine.(Historic study of an African philosophy from the point of view of a Catholic missionary. Subject of heated discussions in contemporary African philosophy.)

UNESCO (1984) *Teaching and Research in Philosophy: Africa*, Paris: UNESCO. (Report on the role of philosophy in the educational systems of most African nations from secondary school and beyond. It was the first UNESCO survey of teaching and research in philosophy in various parts of the world.)

Wiredu, K. (1980) *Philosophy and An African Culture*, Cambridge: Cambridge University Press.(Considerations on the tasks of African philosophy and on the concepts of truth and existence.)

Wiredu, K. (1986) 'The Question of Violence in Contemporary African Political Thought', *Praxis International* 6 (3).(Study of contemporary African thought regarding violence and nonviolence as options in the anti-colonial struggle.)

Wiredu, K. (1987) 'The Concept of Mind with Particular Reference to the Language and Thought of the Akans', in G. Fløistad (ed.) *Contemporary Philosophy*, vol. 5: *African Philosophy*, Boston, MA: Kluwer, 1987. (Discussion of the concept of mind in general and Akan thought in particular.)

Ya'ecob, Z. (*c*.1599-*c*.1692) *The Treatise of Zar'a Ya'ecob*, in D. Bonevac and S. Phillips (eds) *Understanding Non-Western Philosophy: Introductory Readings*, Mountain View, CA: Mayfield Publishing Company, 1993. (The definitive English-language editions of the *The Treatise* by Zar'a Ya'ecob, the seventeenth-century indigenous Ethiopian philosopher and other sources of Ethiopian philosophy are those of Professor Claude Sumner.)

African philosophy, Francophone

The imaginative and intellectual writings that have come out of French-speaking Africa have tended to be associated exclusively with the négritude *movement and its global postulation of a black racial identity founded upon an original African essence. Beyond its polemical stance with regard to colonialism, the movement generated a theoretical discourse which served both as a means of self-validation for the African in particular and the black race in general. This discourse developed further as the elaboration of a new worldview derived from the African cultural inheritance of a new humanism that lays claim to universal significance.*

Despite its prominence in the intellectual history of Francophone Africa and in the black world generally, négritude *does not account for the full range of intellectual activity among the French-speaking African intelligentsia. The terms of its formulation have been challenged since its inception, leading to ongoing controversy. This challenge concerns the validity of the concept itself and its functional significance in contemporary African thought and collective life. It has involved a debate regarding the essential nature of the African, as well as the possibility of constructing a rigorous and coherent structure of ideas (with an indisputable philosophical status) derived from the belief systems and normative concepts implicit in the institutions and cultural practices subsisting from Africa's precolonial past.*

The postcolonial situation has enlarged the terms of this debate in French-speaking Africa. It has come to cover a more diverse range of issues touching upon the African experience of modernity. As an extension of the 'indigenist' theme which is its point of departure, the cultural and philosophical arguments initiated by the adherents of négritude *encompass a critical reappraisal of the Western tradition of philosophy and its historical consequences, as well as a consideration of its transforming potential in the African context. Beyond the essentialism implied by the concept of* négritude *and related theories of Africanism, the problem at the centre of French-African intellectual preoccupations relates to the modalities of African existence in the modern world.*

From this perspective, the movement of ideas of the French-speaking African intelligentsia demonstrates the plurality of African discourse, as shaped by a continuing crisis of African consciousness provoked by the momentous process of transition to modernity. A convergence can be discerned between the themes and styles of philosophical discourse and inquiry in Francophone Africa and some of the significant currents of twentieth-century European philosophy and social thought engaged with the fundamental human issues raised by the impact of modern technological civilization.

Two dominant perspectives frame the evolution of contemporary thought and philosophical discourse in French-speaking Africa: the first is related to the question of identity and involves the reclamation of a cultural and spiritual heritage considered to be imperilled; the second relates to what has been called 'the dilemma of modernity' experienced as a problematic dimension of contemporary African life and consciousness.

1 The French colonial context

The development of philosophical discourse as a distinct current of intellectual activity in Francophone Africa has run parallel to that of an innovative imaginative expression. Such development is bound up with the ideological project of an assertive cultural nationalism. The movement of thought that informs the process of self-reflection on the part of French-speaking African intellectuals, culminating in the idea of *négritude*, derives its impulse from an affective response to the colonial situation. It reflects an effort to grapple with the multiple implications of the collective predicament that forms the larger historical context of the colonial experience, namely the violent encounter between Africa and Europe and its concomitant ideological devaluation of the black race. These factors and the inherent discomforts of the immense process of social and cultural change have been determinants in the origin and evolution of what Robert July (1968) has called 'modern African thought'.

If the general circumstances of the historic conflict between Africa and Europe provide the sentimental hinterland from which the energy of intellectual activity in Africa derives, the specific orientation of contemporary thought in Francophone Africa has been further conditioned by the sustained contact of its intellectual elite with the literary and philosophical traditions to which their French education gave them access. It is worthy of note that the cultural tenets of colonial administration in the areas of Africa under French and Belgian rule, and the educational system

they inspired, were given coherence as functional elements of what was termed the policy of assimilation. The notion of the civilizing mission of European colonialism central to this policy was premised on the idea of the basic inferiority of African culture, which was in need of the redeeming function of Western civilizing values. Constraints of assimilation account for the centrality in Francophone African literature of the theme of alienation, which was given expression as a sense of dissociation from the moral and psychological security of defining origins. The imaginative exploration of this theme found its parallel in a conceptual engagement on the part of the Francophone black (African and Caribbean) intellectual elite with the question of identity. The force of lived experience lent urgency to the thought-provoking question of existence. For the Francophone African elite who were 'assimilated' but none the less preoccupied with interpreting and coming to terms with the colonial experience, intellectual activity could only proceed as a meditation upon the self in relation to a singular historicity.

Associated with the cultural malaise of assimilation was the negative image of Africa that was constantly projected by the Western texts on which was based much of the education of the Francophone African elite. The ideological thrust of these texts is exemplified by the work of Pierre Loti (1888) and other writers associated with the so-called colonial novel. Their perspective helped to propagate the idea of Africa as a landscape whose inhospitable nature was reflected in the savage disposition of its indigenous populations (Fanoudh-Siefer 1968). This literature was the symbolic expression of a European ethnocentrism that had been given philosophical respectability by Hegel, who excluded the African continent from his conception of the world historical process and the unfolding of the universal mind, the foundations of his philosophical system. Arthur de Gobineau's *Essai sur l'inégalité des races humaines* (1884) gave systematic form to the hierarchy of the races established as commonplace to European thought in his time within which African and black races occupied the lowest level. However, it was left to Lucien Lévy-Bruhl to lend the authority of learned discourse to the great divide between the West and the rest of humanity affirmed in de Gobineau's essay. In the series of studies beginning with *Les fonctions mentales dans les sociétés inférieures* (1912) and culminating in *La mentalité primitive* (1922), Lévy-Bruhl undertook to establish the disparity between Western and non-Western cultures at the level of the mental operations by which both were regulated. The term 'prelogical mentality' which he proposed to describe the quality of mind of non-Western peoples was to have resonance beyond the discipline of anthropology. These and other works of the same tenor composed an articulated Western discourse on Africa, which emerged as the antithesis of Europe in the structure of ideas and images by which the colonial ideology was sustained.

2 Intellectual resistance to colonial discourse

The counterdiscourse that was articulated by the Francophone African elite in the 1930s was called into being by the demoralizing effect and egregious nature of this discourse of imperial hegemony. Their response was facilitated by the crisis of European civilization in the early twentieth century after the First World War. The disenchantment with the traditional humanism in Europe reflected in the literature and philosophy of the period provided an appropriate context for the note of dissidence voiced in the ideological writings of the colonized Francophone black intellectual (Kesteloot 1965). Marxism and Surrealism were primary influences, but more pertinent were the formative roles played by French thinkers in the interwar years, which added a particular tone to the expression of some of the leading figures in Francophone African intellectual movements. Of special interest in this respect is the organic nationalism of Maurice Barrès and the anti-intellectualist philosophy of Henri-Louis Bergson, both of whom bequeathed an ambiguous legacy of attitudes and ideas to the cultural nationalism of France's colonial subjects. While the conflation of race and culture provided an anchor in Barrès (1897) for an exclusive vision of the national community, Bergson promoted a special reverence for those noncognitive modes of experience embodied in forms of artistic expression in reaction against the dominant rationalist tradition. Both laid the foundation for Senghor's later celebration of *négritude* as a black racial endowment and provided the language for its formulation.

Paradoxically, the discipline of anthropology, in which a new spirit of cultural relativism had begun to prevail, provided the immediate source of intellectual armoury of the Francophone African response to colonial ideology. The efforts of French scholars Robert Delavignette and Maurice Delafosse to explicate African forms of social and cultural expression and accord them recognition culminated in Marcel Griaule's *Dieux d'eau: entretiens avec Ogotemmeli* (1948). The articulation in this work of the elaborate cosmology of the Dogon, as related by the African sage Ogotommeli, revealed an evident symbolic architecture and conceptual organization in an African

culture that advanced the case for a revaluation of the continent and its peoples.

3 Placide Tempels's *Bantu Philosophy*

Placide Tempels's *Bantu Philosophy* (1945) was decisive in giving a philosophical orientation to the emerging discourse of cultural nationalism in Francophone African. Tempels's objective was to reveal the existence of a reflective disposition among the Luba, an ethnic group in the then Belgian Congo. He ascribed to them a collective philosophy distinguished by an ontology summed up in the following quotation:

> I believe that we should most faithfully render Bantu thought in the European language by saying that the Bantu speak, act, live as if, for them, beings were forces. Force is not for them an adventitious accidental reality. Force is even more than a necessary attribute of beings: Force is the nature of being, force is being, being is force.
>
> (1945: 35)

The passage makes obvious the derivation of Tempels's work from Bergson: the notion of 'vital force' by which he sought to characterize Bantu thought recalled the French philosopher's *élan vital*. Tempels's reconstruction of mental structure from 'collective representations' dear to Durkheim and his disciples in the French school of anthropology was an application of Lévy-Bruhl's method, although a reversal of its theoretical import and ideological implications. *Bantu Philosophy* provided the model and conceptual framework for the construction of an original African philosophy and has remained a central reference of philosophical debate in Africa.

4 *Négritude*

It is against this historical and intellectual background that the concept of *négritude* took form. It was the eminent French philosopher Jean-Paul Sartre who was the first to give the concept extended philosophical formulation. His essay 'Orphée noir' (Black Orpheus) (1949) was an expansive reflection on the term which had been coined by the Martinican poet Aimé Césaire in the context of his poem *Cahier d'un retour au pays natal (Notebook of a Return to my Native Land)* (1939) to denote the advent of a liberated black consciousness. In the essay Sartre offered a definition of *négritude* in Heideggerian/Existentialist terms, as 'the-being-in-the-world-of-the-Negro'. Extending this definition by reference to the orthodoxies of Marxism, he situated the racial consciousness designated by *négritude* and the project of collective freedom it proclaimed in an historical perspective as a stage in a dialectic destined to be transcended by the advent of a classless and raceless world society.

Senghor's conception of *négritude* both enlarges upon Sartre's definition and gives it a new orientation. Rather than a contingent factor of black collective existence and consciousness as with Sartre (for Senghor this aspect corresponds to what he calls 'subjective *négritude*'), the concept denotes for Senghor an enduring quality of being constitutive of the black race and exempt from the exigencies of the historical process. The term further signifies a complex of objective factors that shape the African experience, embodied in forms of life on the continent and manifested in the modes of thought and feeling of its people, hence Senghor's definition of *négritude* as 'the sum total of African cultural values' (1970). His theory of *négritude* takes the form of an exposition of the African's distinctive manner of relating to the world. Appropriating Lévy-Bruhl's notion of 'participation', Senghor accords primacy to emotion as distinctive of an African mode of access to the world. Emotion is accorded special signification by Senghor; it is no longer merely a psychological state, but a mode of apprehension, a 'capturing of integral being - body and consciousness - by the indeterminate world' (1962: 15). Senghor's thinking concerns itself with the opposition between both the mystical approach to reality that the developed emotion determines in the African, as well as the pure intellection that is held to be characteristic of the West and historically enshrined in the *cogito* of Descartes. According to Senghor, emotion is governed by intentionality and thus presents itself as a valid mode of cognition.

We have here the epistemological foundation of the African worldview and collective ethos as interpreted by Senghor, who posits in the African a total grasp of reality embracing the continuum from the realm of nature to the supernatural. The informing principle of this *Weltanschauung* and system of social organization emanating from it amounts to a spiritualism that invests all phenomena with a sacred character. Senghor has extended this idea into his theory of African socialism, presented as the social philosophy entailed by the theory of *négritude*. Although commanded by practical considerations, African socialism as enunciated by him is a strategy for reconciling the

imperatives of modernity - social and economic development in Western terms -with an African ethos. For Senghor (1961) this socialist ideal is governed by the need to infuse the humanizing values of traditional Africa into the new structures of collective life in the modern dispensation. Therefore, African socialism presents itself less as the construction of a concrete social progamme than as an axiology.

Senghor's theory of *négritude* developed as a function of his poetic vocation. Although in later works (Ndaw 1983) he restated his system of ideas to align it more closely with the classical epistemology codified by Aristotle, the theory bears a close affinity with the various continental forms of Lebensphilosophie that have sprung up as a reaction to the instrumental reason of modern social organization (see Lebensphilosophie). There is a sense in which Senghor's *négritude* may be interpreted as an African version of Bergsonism: a verification in African form of the cultural expression of the idea of intuition as the sign of experience at the most profound level of consciousness.

5 Ethnophilosophy

Senghor's *négritude* represents an effort to provide a comprehensive elucidation of African being. Despite its limitations and disputed status as philosophy, it marks, as D.A. Masolo has observed, 'the legitimate origin of philosophical discussion in Africa' (1994: 10). The movement of self-definition it initiated led to the effort in Francophone Africa to generate an African philosophy from anthropological literature pertaining to the traditional cultures on the continent. The school of thought spawned by this effort, known as ethnophilosophy, is represented by Alexis Kagamé's *La Philosophie bantu-rwandaise de l'être (Bantu-Rwandan Philosophy of Being)* (1956), a work conceived as a verification and reformulation of Tempels's propositions in more rigorous analytical terms. Kagamé appealed to his native Rwandan language to reconstruct the philosophy underlying his people's worldview. From the root stem, *ntu*, signifying essence in general, Kagamé has deduced four fundamental categories of Bantu thought: man, being endowed with intelligence, or *muntu*; being without intelligence, such as animals, plants, minerals, or *kintu*; the space-time continuum, or *hantu* and modality, or *kuntu*. According to Kagamé these terms function both as markers of implicit thought processes and vehicles of an explicit philosophical discourse demonstrable by reference to Rwandan oral tradition.

Kagamé's exposition is not intended as a reconstruction but as a description, *stricto sensu*, of an authentic system of Bantu thought, which corresponds with Aristotle's system for its translation into a non-African language and frame of reference. For this reason the work raises the question of language in African philosophy and the problem that Benveniste has identified as the relation between 'categories of language and categories of thought' (1966). Kagamé's pioneering effort was followed up by explorations of traditional systems of thought in the work of scholars who form what V.Y. Mudimbe has designated (1986) as Tempels's philosophical school. Composed mainly of central Africans and dominated by clerics, the major preoccupation of this school has been to identify those elements of the African personality compatible with Christian doctrine. Their endeavour has fostered the emergence of a theology that reconciles the West and Africa through a shared spirituality.

6 Cheikh Anta Diop

Ethnophilosophy, as a direct tributary of *négritude*, seeks to define African identity in terms of an ontology. Another current of cultural nationalism, the historical school associated with the work and personality of the Senegalese scholar Cheikh Anta Diop, discovers this identity in what may be called an African *longue durée*. Diop is best known for his book *Nations nègres et culture (Black Nations and Culture)* (1956), which advanced the thesis of ancient Egypt as an integral part of a black African civilization. The real significance of Diop's work resides less in the validity of his arguments and conclusions than in the development he gave to the thesis in subsequent works. In *L'Unité culturelle de l'Afrique noire (Cultural Unity of Black Africa)* (1959), Diop considered Africa as a single, unified cultural area on the basis of the continuity of cultural forms and value systems between ancient Egypt and indigenous civilizations throughout Africa. This argument was summarized in 'Egypte ancienne et Afrique noire' (1962). The philosophical implications of Diop's work emerge from the comprehensive vision of Africa's historical personality by which it is informed and its spirit of confrontation with Hegel's philosophy of history. The erudition and methodological effort he invested in constructing an 'historical sociology' aimed to restore Africa to an honourable place in universal history. As he says, 'Historical science cannot shed all the light one might expect it to cast upon the past until it integrates the African component of humanity, in proportion to the role it has actually played in history, into its synthesis' (1962: 11).

Diop's work established a line of historical reflection and research in Francophone Africa, as exemplified in the writings of Joseph Ki-Zerbo (1972) and especially Théophile Obenga, Diop's most accomplished disciple. His *L'Afrique dans l'Antiquité* (1970) represents a summation of the ideas and methods of the school spawned by Diop (see Egyptian philosophy: influence on ancient Greek thought).

7 The critique of *négritude*

A reaction set in against the theory of a black racial self and the creation of an African collective identity propounded by *négritude* and endorsed by ethnophilosophy. The critique of *négritude*, which began in the 1950s with attacks on Sartre's definition by Albert Franklin (1953) and Gabriel D'Arboussier (1959), developed into controversy that has not subsided. The radical spirit of this critique was embodied in the work of Frantz Fanon, beginning with his analysis of the pathology of colonialism in *Black Skin, White Masks* (1952). This analysis took the form of a Hegelian enactment of the black subject's drama of consciousness, that of the struggle for recognition involved in the master/slave dialectic. Fanon's clinical perspective focused on the inward psychological depredations of colonial domination. The ethics of violence elaborated in *The Wretched of the Earth* (1961) springs from his conception of its restorative value for the colonized native. His uncompromising radicalism with its repudiation of mere culturalism endows violence with a transcendent significance: 'African culture will take concrete shape around the struggle of the people, not around songs, poems or folklore' (1961: 164).

The critique of Senghor undertaken by Stanilas Adotevi (1972) owes its force to Fanon's example and to his disposal of identity as an issue worthy of moral concern and theoretical interest. Fanon's influence also accounts for the break with the spirit of cultural nationalism embodied in *négritude* by the philosopher Marcien Towa (1971). His intransigence is displayed in the following terms: 'The transformation of one's present condition signifies at the same time the transformation of one's essence, of what is particular to the self, of what is original and unique about it; it is to enter into a negative relationship with the self' (1971: 41). This growing disaffection towards *négritude* developed into a theoretical attack on ethnophilosophy as its outgrowth, marking a significant phase in the evolution of Francophone African philosophy. Eboussi-Boulaga's initial objection to Tempels, whose philosophy he described as 'an ontological system that is totally unconscious, and given expression in an inadequate and incoherent vocabulary' (1968) extended in Towa's essay into a critical reappraisal of ethnophilosophy, culminating in an effort to demolish its conceptual edifice in Paulin Hountondji's *African Philosophy: Myth or Reality* (1983). Hountondji's focus on the methodological procedures of the ethnophilosophers led him to discern a 'confusion of genres' in their attempts to construct a philosophical discourse from material with an ethnological interest. For him ethnophilosophy was 'a hybrid ideological discipline without a status in the world of theory' (1983: 52). To the unanimism implicit in the conception of philosophy as a collective system of thought immanent in a people's culture, Hountondji opposed the criterion of philosophy as an explicit discourse and its rigorous character as a critical activity. He represented philosophy as a reflection on science considered as a significant component of modern culture and equated the philosophical enterprise with the development of science. The lack of scientific culture in Africa forced him to reach the conclusion that the continent is a long way from fulfilling the conditions necessary for philosophical practice.

Hountondji progressed from a narrow conception of philosophy to a broader view amounting to a form of pragmatism (see Pragmatism), involving an interrogation of the possible function of philosophy in the African context. A reappraisal of modes of scientific thought and practice in traditional Africa and a concern for their modernization and expansion in contemporary Africa have come to provide the principal orientation of his reflection, inspired by a sharper sense of the possible relation of philosophy to public policy and social practice. Therefore, the role of philosophy has come to include for Hountondji 'the analysis of the collective experience with a view toward a critique of everyday life' (1992: 359). The political implications of such a critique, suggested by the work of Henri Lefebvre in France after the precedent of the Frankfurt school, are made clear.

The political dimension of Hountondji's critique is fully actualized in Achille Mbembe's 'Provisional Notes on the Postcolony' (1992: 3-37). A phenomenology of political life in contemporary Africa, the essay emphasizes the introspective and critical character of intellectual activity in French-speaking Africa in the postindependence period as a function of the existential problems inherent in the process of transition in contemporary Africa. Beyond what has been called 'the crisis of relevance' in African philosophy (Oladipo 1992), this activity aims to

lay the philosophical foundation for social development in Africa in pursuit of a new order of collective life, which Hountondji termed ‘the Utopia of another society’ (Mudimbe 1992: 360).

8 V.Y. Mudimbe and the critique of Africanist discourses

Valentin Mudimbe’s work is significant in terms of the question of the relationship of discourse and constitution of thought with the ambiguous modernity of Africa. He delineated, after Foucault, an ‘archaeology of African knowledge’ motivated by the ambition to found a new African philosophy with an original register of enunciation, able to underwrite Africa’s conceptual autonomy. In *L’autre face du royaume* (1973), he criticizes the discourse of ethnology as an aberrant language.

The Invention of Africa (1988), Mudimbe’s best known work, is a development of this judgment and an examination of its implications for African expression in the modern world. In his view the homology between the political and economic imperialism of the West on one hand and its ‘epistemological imperialism’ on the other, constitutes Africa as a province of a Western epistemological territory. The function of anthropology developed through the nineteenth century was to ‘account for the normality, creative dynamism and achievements of the "civilized world" against the abnormality, deviance and primitiveness of the non-literate world’ (1988: 24). African studies formed part of this development. It has been so fully integrated into the Western order of discourse that the entry of Africans served to amplify the conceptual scope of this order in what Mudimbe calls a ‘discourse of succession’. Mudimbe remarked that ‘the main problem concerning the being of African discourse remains one of the transference of methods and their cultural integration in Africa’ (1988: 182). His solution was to adapt structuralism to the project of reconstruction in African philosophy (see Structuralism). The structuralist method permits an escape from the constraints of a systematized rationality while affording an entry into the truth of the world: ‘empirical categories can be used as keys to a silent code, leading to universals’ (1988: 35). It is unclear how this approach yields the ‘absolute’ or ‘transhistoric discourse’ that Mudimbe claims as the alternative to Western rationality. Despite what a commentator has called ‘the ambiguous nature of the project suggested by Mudimbe’ (Masolo 1991: 109), the interest in Mudimbe’s work resides in its account of the African intellectual adventure, which amounts to a vision of the African mind in its encounter with the Western world system.

9 Summary

The themes and positions reviewed provide the main lines of French African thought which have inspired a current of philosophical activity in Africa with its own style of discourse. This has prompted the view that the academic practice of philosophy in Africa is divided between the analytical tradition in Anglophone Africa and the continental tradition in Francophone Africa. Philosophical inquiry in both parts of Africa exhibits the three modes that Richard Rorty has identified in contemporary Western philosophy as ‘science, metaphor, politics’ (1991: 9-26). Although French-speaking African philosophers do not employ the vocabulary of Anglo-American analytical philosophy, the debate on the epistemological status of traditional thought in Africa has involved them in a sustained reflection on the nature and scope of philosophy itself. Both sides in the debate have been obliged to undertake a clarification of the terms of their discourse, as with Kagamé, whose categories also receive some close technical scrutiny by Hountondji (1983: 188-9). The debate has generated a metaphilosophy concerned with issues such as the relation of myth to metaphysics and the procedural questions touching upon the proper order of terms and concepts as well as the conditions of philosophy as both a discipline and cultural practice. The debate assumes significance by reason of the comparative perspective it projects on the discipline, covering such questions as the meaning of concepts across cultures, leading ultimately to the problem of universalism.

Francophone African thought provides an African perspective on the relation between ‘Thought and Change’ (Gellner 1965) demonstrated in the West by the progressive imbrication of social science with philosophy since Weber: a development that points to a critical engagement with the whole range of political, social, cultural, moral and aesthetic issues posed to modern awareness by the triumph of rationalism and the scientific revolution. The critical thrust of current debates associated with postmodernism concerning the philosophical legacy of the Enlightenment reflects a sustained effort of internal reassessment in the West, a process in which the reappraisal of Western rationalism by Senghor and other French-speaking African intellectuals is profoundly implicated. As a ‘strategy of differentiation’ (Irele 1995: 15-34), *négritude* seeks to redefine the terms of the relationship between peoples and cultures within a comprehensive intelligence of the world. The metaphoric allure of a certain style of philosophical discourse identified by Rorty is captured in *négritude*, whose speculative mode offers a challenge to

the Western paradigm in rejection of its 'master narratives' (Lyotard 1979).

Beyond this polemical aspect of *négritude*, which also informs Mudimbe's work, Francophone African philosophy assumed a theoretical and historical interest in a global assessment of the dominant trends in modern philosophy and social thought. The commonality between such developments in Western thought exemplified by the Frankfurt school's critique of culture in modern industrial society (see Frankfurt School), the Neo-Marxism of Henri Lefebvre, North American neopragmatism and 'communitarianism' bears witness to a renewed focus on first order questions and on concrete issues of existence in the 'lifeworld' (Habermas 1985). The intersection between these trends in modern Western philosophy and intellectual activity in French-speaking Africa assumes a broad contemporary significance in this light, as under the pressure of historical experience, French-speaking African intellectuals have forced philosophy to confront anew the problems that presided at its origins in the West and which seem to govern its future direction.

See also: African philosophy, Anglophone;Cultural identity; Marginality

F. ABIOLA IRELE

References and further reading

Adotevi, S. (1972) *Négritude et Négrologues*, Paris: Union Générale d'Editions.(A comprehensive critique of Senghor's concept of *négritude* from a sociological and political standpoint.)

Barrès, M. (1897) *Le Roman de l'energie nationale (Book of National Energy)*, vol. 1, *Les Déracinés (The Uprooted People)*, Paris: Fasquelle; repr. Paris: Livre de Poche, 1967.(An early fictional expression of French conservative nationalism, this work exerted a significant influence on Senghor's intellectual development.)

Benveniste, E. (1966) 'Catégories de pensée et catégories de langue' (Categories of Thought and Categories of Language), *Problèmes de Linguistique Générale*, Paris: Gallimard, 63-74.(Analyses the relation between the concept of being in ancient Greek philosophy and the linguistic structure of classical Greek.)

Carroll, D. (1995) *French Literary Fascism: Nationalism, Anti-Semitism and the Ideology of Culture*, Princeton, NJ: Princeton University Press.(A study of the relationship between literature and right-wing movements in France from the late nineteenth century up to the outbreak of the Second World War.)

Césaire, A. (1939) *Cahier d'un retour au pays natal*; repr. Paris: Présence Africaine, 1956; trans. M. Rosello with A. Pritchard, *Notebook of a Return to my Native Land*, Newcastle-upon-Tyne: Bloodaxe Press, 1995.(The long poem in Surrealist imagery in which the word *négritude* first appeared in print.)

D'Arboussier, G. (1959) 'Une dangereuse mystification: la théorie de la Négritude' (A Dangerous Mystification: the Theory of Négritude), *Nouvelle Critique* (June): 34-47.(An early attack on *négritude* from a radical Marxist perspective.)

Diop, C.A. (1956) *Nations nègres et culture (Black Nations and Culture)*, Paris: Présence Africaine.(With this major reference of African historiography, Diop initiated the debate on the racial belonging of the ancient Egyptians.)

Diop, C.A. (1959) *L'Unité culturelle de l'Afrique noire (Cultural Unity of Black Africa)*, Paris: Présence Africaine.(Posits an opposition between the nomadic culture and aggressive disposition of the early Europeans and the sedentary civilization that emerged in the Nile valley and spread to the rest of Africa.)

Diop, C.A. (1962) 'Egypte ancienne et Afrique noire' (Ancient Egypt and Black Africa), *Bulletin de l'Institut Fondamental de l'Afrique Noire* 24 (3-4): 449-574; repr. Dakar: IFAN, 1989.(Argues for the continuity between ancient Egyptian civilization and indigenous cultures in sub-Saharan Africa.)

Eboussi-Boulaga, F. (1968) 'Le Bantou problématique' (Bantu Problematic), *Présence Africaine* (April-June): 3-40.(A critical appraisal of Placide Tempels's exposition of Bantu philosophy.)

Fanon, F. (1952) *Black Skin, White Masks*, London: Pluto Press, 1986.(Explores, in the style of French phenomenology, black subjectivity through its complex mode of responses to colonial domination and racism.)

Fanon, F. (1961) *The Wretched of the Earth*, London: Penguin, 1983.(An examination of the revolutionary consciousness in Africa and the Third World, grounded in an ethics of violence that is counterpoised to the psychological and moral ravages of the colonial situation.)

Fanoudh-Siefer, L. (1968) *Le Mythe du nègre et de l'Afrique noire dans la littérature française de 1800 à la deuxième guerre mondiale (The Myth of the Negro and of Sub-Saharan Africa in French Literature from 1800 to the Second World War)*, Paris: Klincksieck.(The most complete study of the construction of the African image in French literature.)

Franklin, A. (1953) 'La Négritude: réalité ou mystification? Reflections sur ''Orphée noir"' (Négritude: Reality or Mystification? Reflections on 'Black Orpheus'), *Présence Africaine* 14: 287-303.(The first text to present

theoretical and practical objections to Sartre's definition of *négritude*.)
Gellner, E. (1965) *Thought and Change*, London: Weidenfeld & Nicholson.(Opposes the analytical school to argue for an engagement of philosophy with the phenomenon of modernity.)
Gobineau, A. de (1853-5) *Essai sur l'inégalité des races humaines*, Paris: Didot, 3 vols; trans. A. Collins, *The Inequality of Human Races*, Torrance, CA: Noontide Press, 1983.(Elaborates a hierarchy of races based on the supposed genetic traits and endowments of the different branches of humanity; a central text of European ethnocentrism.)
Griaule, M. (1948) *Dieux d'eau: entretiens avec Ogotemmeli*, Paris: Editions du Chêne; 1st edn trans. R. Butler, 2nd edn trans. A.I. Richards and B. Hooke, *Conversations with Ogotemmeli: An Introduction to Dogon Religious Ideas*, London: Oxford University Press, 1965.(An extended account of the cosmology of the Dogon people in west Africa, narrated to a French anthropologist by a traditional sage.)
Habermas, J. (1985) *The Philosophical Discourse of Modernity: Twelve Lectures*, Cambridge, MA: MIT Press.(A review of the project of modernity as formulated in Western philosophy since Hegel in order to restate the theory of 'communicative reason' developed by Habermas in earlier works.)
Horton, R. (1993) 'African Traditional Thought and Western Science', in *Patterns of Thought in Africa and the West*, Cambridge: Cambridge University Press, 197-258.(Horton's essay proceeds from a reformulation of classical British anthropology to the postulation of a formal correspondence between African systems of thought and Western scientific methods.)
Hountondji, P. (1983) *African Philosophy: Myth or Reality*, Bloomington, IN: Indiana University Press, 2nd edn, with preface by author, 1996.(A systematic critique of the effort in ethnophilosophy to derive an African philosophical system from the traditional background of belief.)
Hountondji, P. (1981) 'Que peut la philosophie?' (What Can Philosophy Do?), *Présence Africaine* 119: 47-71.(A discussion of the practical possibilities of philosophy in the African context.)
Hountondji, P. (1992) 'Daily Life in Black Africa: Elements for a Critique', in V.Y. Mudimbe (ed.) *The Surreptitious Speech*, Chicago, IL: Chicago University Press, 344-64.(An argument for a moral engagement of philosophy with the political and economic crisis in post-independence Africa.)
Hountondji, P. (1994) 'Démarginaliser' (Demarginalize), intro. to *Les savoirs endogènes (Endogenous Knowledge)*, Dakar: CODESRIA, 1-34.(Outlines the possible relation of indigenous forms of scientific thought and technologies to Western theory and practice.)
Irele, A. (1995) 'Dimensions of African Discourse', in K. Myrsiades and J. McGuire (eds) *Order and Partialities: Theory, Pedagogy and the 'Postcolonial'*, New York: State University of New York Press, 1995: 15- 34. (Includes a discussion of *négritude* as counter-discourse.)
July, R. (1968) *The Origins of Modern African Thought*, London: Faber & Faber.(An intellectual history of west Africa from the eighteenth century to the era of independence in the early 1960s.)
Kagamé, A. (1956) *La Philosophie bantu-rwandaise de l'être (Bantu-Rwandan Philosophy of Being)*, Brussels: Académie Royale des Sciences Coloniales; repr. New York: Johnson Reprint Corporation, 1966.(An exposition of the conceptual categories of Bantu philosophy based upon terms in the Rwandan language.)
Kagamé, A. (1976) *La philosophie bantu comparée (Bantu Philosophy Compared)*, Paris: Présence Africaine.(A comparison of Bantu philosophy.)
Kesteloot, L. (1965) *Les écrivains noirs de langue française*, Brussels: Institut Solvay; trans. E. Conroy Kennedy, *Black Writers in French: A Literary History of Négritude*, Washington, DC: Howard University Press, 1991. (Provides an exhaustive account of the genesis of French-speaking African and Caribbean literature and thought.)
Ki-Zerbo, J. (1972) *Histoire de l'Afrique noire (History of Black Africa)*, Paris: Hatier.(A celebration of the medieval states of west African empires as part of a revaluation of the precolonial history of Africa.)
Lévy-Bruhl, L. (1912) *Les fonctions mentales dans les sociétés inférieures*, Paris: F. Alcan; trans. L. Clare, *How Natives Think*, Princeton, NJ: Princeton University Press, 1985.(The work in which the author first proposed the concept of 'participation' as the dominant mode of experience in non-Western cultures.)
Lévy-Bruhl, L. (1922) *La mentalité primitive*, Paris: Presses Universitaires, 4th edn, 1960; trans. L. Clare, *Primitive Mentality*, New York: AMS Press, 1978.(Develops the notion of 'prelogical mentality'.)
Loti, P. (1888) *Le Roman d'un Spahi*, Paris: Calman-Levy; trans. M.L. Watkins, *The Romance of Spahi*, New York: Rand, McNaly and Co., 1890.(Set in Africa, this work is the outstanding example of the colonial novel in French literature.)
Lyotard, J.-F. (1979) *La Condition postmoderne*, Paris: Editions de Minuit; trans. G. Bennington and B. Massumi, *The Postmodern Condition*, Manchester: Manchester University Press, 1984.(The classic statement of the postmodern consciousness and sensibility.)

Masolo, D.A. (1987) ‘Alexis Kagamé and African socio-linguistics’, in G. Fløistad (ed.) *Contemporary Philosophy*: vol. 5, *African Philosophy*, Boston, MA: Kluwer: 181-205.(A critical discussion of Kagamé’s use of Rwanda terms in developing Bantu philosophy.)
Masolo, D.A. (1991) ‘An Archaeology of African Knowledge’, *Callaloo* 14 (4).(A critical discussion of the ideas of Mudimbe.)
Masolo, D.A. (1994) *African Philosophy in Search of Identity*, Bloomington, IN: Indiana University Press.(The most comprehensive study of African philosophy published to date.)
Mbembe, A. (1992) ‘Provisional Notes on the Postcolony’, *Africa* 62 (1): 3-37.(A trenchant analysis of the social and mortal consequences of authoritarian rule in postcolonial Africa.)
Mudimbe, V.Y. (1973) *L’autre face du royaume (The Other Side of the Kingdom)*, Paris: L’Age d’Homme.(A critique of anthropology as a form of discourse concerned with African ‘otherness’.)
Mudimbe, V.Y. (1986) ‘On the Question of African Philosophy: the Case of French Speaking Africa’, in I. Mowoe and R. Bjornson (eds) *Africa and the West: The Legacies of Empire*, Westport, CT: Greenwood Press, 1986: 89-120.(A review of the development of philosophical discourse in French-speaking central Africa.)
Mudimbe, V.Y. (1988) *The Invention of Africa*, Bloomington, IN: Indiana University Press.(An effort to reconstruct an African philosophy as gnosis - a system that is not answerable to the logic of Western rationality.)
Ndaw, A. (1983) *La Pensée africaine: recherche sur les fondements de la pensée négro-africaine (African Thought: Research into the Basis of Negro-African Thought)*, with preface by Senghor, Dakar: Nouvelles Editions Africaines.(An attempt to provide a synthesis of the various indigenous systems of thought in Africa; Senghor’s preface is a restatement of his position on modes of thought in African philosophy.)
Obenga, T. (1970) *L’Afrique dans l’Antiquité (Africa in Ancient times)*, Paris: Présence Africaine.(Expands upon the historiography of Diop in its investigation of the Egyptian and Middle Eastern component of the African past.)
Oladipo, O. (1992) *The Idea of African Philosophy*, Ibadan: Molecular Publishers.(A wide-ranging discussion of the situation and role of philosophy in Africa.)
Rorty, R. (1991) *Essays on Heidegger and Others: Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press, 9-26.(Reviews three conceptions of philosophy in the twentieth century.)
Sartre, J.-P. (1949) ‘Orphée noir’ in *Situations III*, Paris: Gallimard; trans. S. Allen, ‘Black Orpheus’, Paris: Présence Africaine.(Contains Sartre’s philosophical formulation of *négritude*.)
Senghor, L.S. (1961) *Nation et voies africaines du socialisme*, Paris: Présence Africaine; trans. M. Cook, *On African Socialism*, New York: Praeger, 1964.(An attempt to derive a social programme in Africa from the tenets of *négritude*.)
Senghor, L.S. (1962) ‘De la Négritude: psychologie du Négro-africain’ (About négritude: the Psychology of the African Negro), *Diogène* 37.(An interesting discussion.)
Senghor, L.S. (1964) *Liberté I: Négritude et humanisme (Liberty 1: négritude and humanism)*, Paris: Editions de Seuil.(This volume contains Senghor’s principal essays on the theory of *négritude*.)
Senghor, L.S. (1970) ‘Négritude: A Humanism of the Twentieth Century’, in W. Cartey and M. Kilson (eds) *The African Reader: Independent Africa*, New York: Random House, 1970.(Senghor defends the relevance of *négritude* to contemporary concerns in Africa.)
Tempels, P. (1945) *Bantu Philosophy*, Paris: Présence Africaine, 2nd edn, 1956.(The first systematic exposition of an indigenous African system of philosophical thought, characterized as an ontology of force rather than of being.)
Towa, M. (1971) *Essai sur la problématique philosophique dans l’Afrique actuelle (Essay on the Problem of Philosophy in Contemporary Africa)*, Yaoundé: CLE.(Proceeds from a critique of *négritude* and ethnophilosophy to a statement of the transformative role of philosophy in Africa.)
Wiredu, K. (1996) *Cultural Universals and Particulars: An African Perspective*, Bloomington, IN: Indiana University Press.(A discussion of the significance of cultural particularities in philosophical discourse and their role in communication between different cultures and peoples.)

African traditional religions

Religion has been at the centre of recent philosophical debate in Africa for two major reasons. The first is that the answers to many central canonical philosophical questions in precolonial African societies take a religious form. As a result any attempt to construct an African philosophy begs attention to the epistemological and ontological standing of claims of this general sort. The second reason religion has been central to African philosophy is that one of the major issues in modern African philosophy is whether distinctively African modes of thought exist. Within this debate influential positions have been argued by reflecting on the character of traditional religious thought and practice and contrasting it with modes of thought purportedly associated with Western science.

1 Religion

Religion is a term whose definition is seen as controversial. Beliefs, institutions and practices can be said to be religious, but the relative importance of belief as opposed to ritual practice, or of ethical belief as opposed to the metaphysical, varies greatly among the belief systems with which Westerners might be familiar - Christianty, Buddhism, Judaism and Islam. Questions of creed, reflecting both the centrality of such questions to Western philosophy and the crucial role of creeds in Christian thinking are at the heart of the literature under discussion in this entry, which derives mainly from philosophers educated in Christian cultures. It will be necessary to draw attention to matters obscured by the focus on matters of propositional belief.

2 ‘African traditional religion and Western science’

Philosopher-anthropologist Robin Horton(1967) wrote a paper with this title in which he argued that the religious ideas of precolonial Africa maintained by many postcolonial Africans were best understood as constituting a body of theory whose fundamental aim, like that of Western science, was explanation, prediction and control of the phenomena of everyday life. Horton made the claim that traditional African religion is like modern Western science. Horton begins with the idea that anthropology’s first task is to provide for one culture, ‘the West’, an understanding of the concepts of another. Translation is the first step of this task. In the preface of *Patterns of Thought in Africa and the West* (1993) Horton argued that, ‘since translation involved finding equivalences of intention and structure between source-language and target-language, it followed that the scholar in quest of the appropriate translation instruments for African religious thought must be prepared to inquire deeply into the intentions and structures embodied in various areas of Western discourse’ (1993: 2). At the time of writing, ethnographic studies of African religion were dominated by a sort of Durkheimian consensus which held that religious notions were fundamentally symbolic of social relations. Taken to extremes this led to the ‘symbolist’ view that religious practices, far from being attempts to mobilize the world of spirits in the pursuit of mortal ends, were, like art, fundamentally expressive with the aim of representing social norms and ideas. Symbolists wanted to deny in particular that religious appeal to spirits presupposes literal belief in them any more than, for example, the function of *Hamlet* as drama requires belief in the literal existence of a Danish prince.

One motive for this view, Horton suggested, was the urge to escape the ethnocentrism of Victorian anthropology with its image of childlike primitives. In their rush to avoid ethnocentrism Horton thought that some symbolists, faced with the irrationality of traditional beliefs, insist that these beliefs are both rational and ‘symbolically’ true. Horton shares the urge to avoid ethnocentrism, but argues that the assumption underlying this argument that the beliefs are irrational is equally ethnocentric. False beliefs, simply put, do not have to be irrational.

Against the symbolists Horton argued that if any area of discourse in the West provided a model for the central purposes of African religious thought, it was not art but science. He sketched a picture of the role of science in the West as the development of theories which seek to place events in a wider causal context than that provided by common sense. At the heart of this process, in Horton’s account, was the development of a structure of belief in invisible entities whose behaviour accounted for the manifest behaviour of the visible world. He felt a key element of theory building was the development of analogies between invisible entities and visible ones. In the natural sciences these invisible entities - atoms and molecules - were modelled on everyday inanimate objects like tiny billiard balls speeding about in the vast spaces of the microcosm, crashing into each other from time to time. (All this at his time of writing amounted to mainstream philosophy of science.) In African religions, invisible entities - gods and spirits - were modelled not on objects but on human beings.

Horton pointed to this difference and offered to explain it, suggesting that it arose from the fundamental nature of explanation as the reduction of the unfamiliar to the familiar. In traditional cultures nature is untamed, alien and a source of puzzlement and fear. Social relations and people, on the contrary, are familiar and well understood. Thus, explaining the behaviour of nature in terms of agency is reducing the unfamiliar forces of the wild to the familiar explanatory categories of personal relations. In the West, on the other hand, 'alienated man' finds social relations puzzling and problematic and the physical world seems stable and familiar.

Horton went on to argue for a further difference. His summary was that African thought, unlike Western science, operated in a 'closed' predicament. His use of language derived from Karl Popper's distinction between 'open' and 'closed' societies (see Popper, K.R.), but Horton reduced Popper's connected set of oppositions to a simpler contrast: that between closed cultures 'characterized by a lack of awareness of alternatives, sacredness of beliefs, and anxiety about threats to them' (1993: 223) and open cultures, aware of alternatives and less threatened by the possibility of intellectual change.

These arguments have been subjected over the years to a great deal of scrutiny. The most controversial claims presuppose that precolonial African societies have remain unchanged. The stability of social relations implicit in his explanation of why people are appealed to as models, is belied by the turbulent history of many regions of west Africa since the 1700s. The very same wars and migrations must have made people extremely aware of alternative theoretical possibilities. This question of the openness of traditional cultures will be examined later.

3 Initial criticism

The centrality of Horton's argument about religion to the debate in Anglophone African philosophy is evident in the quantity of papers devoted to this question in the leading Anglophone African journal *Second Order*. In the first issue in 1972 Vernon Pratt argued, after Wittgenstein, that Horton had understated the significance of the fact that it was agents not objects that were central to traditional religious theory. Explanations in terms of agency, he argued, differ from causal explanations in two crucial ways. First, agency is intrinsically unpredictable: someone is predictable only if his choice 'has been made for him'. Second, 'it is of the nature of an action to *break in* on a course of events'. Horton (1967) replied that the first of these claims was mistaken and the second claim, although true, did not distinguish agent explanation from the general causal explanation.

A similar debate occurred between Horton and Beattie, the symbolist-anthropologist, in which Beattie (1973) offered a number of arguments in defence of the view that traditional religion does not involve literal belief in spirits. He argued that the reason why spirits are perceived as unobservable is that the practitioners of traditional religion understand that they do not exist. He suggested that religious entities are invoked at the point where a problem cannot be dealt with by 'available empirically-grounded techniques' and that they must therefore be dealt with 'in terms of expressive symbolism' (1993: 4). Horton responded with the view that there is no reason to suppose either that traditional believers are unconvinced of the existence of the spirits to which they refer or that the only possible response to the failure of 'available empirically-grounded techniques' is symbolic, since scientific theory is also a response to such failures.

An interesting philosophical exchange between Horton and John Skorupski (1976) concerned itself with Horton's proposed explanation of the ethnographic observation that in many traditional religions an object is said to belong to a kind to which, as far as an observer is concerned, it obviously does not belong. (Examples are the Nuer identification of twins with birds, reported by Evans-Pritchard in *Nuer Religion* (1956) and the Dinka claim that some men 'are' lions.) Horton proposed that these should be seen as theoretical identifications, such as the famous identification proposed by Eddington (1928) of the 'hard, solid table of common sense thought and action' with the 'largely empty space, peopled by minuscule planetary systems, of theoretical physics' (Horton 1993: 84).

Skorupski argued that Horton misidentified the character of the theoretical identifications of the sciences, believing instead that they are not inherently paradoxical. He went on to suggest that the right Western analogy can be found in certain Christian biblical tales, such as the identification of Christ's body and blood with the bread and wine of the Eucharist.

4 Critiques of the analogy between science and traditional religion

The Ghanaian philosopher Kwasi Wiredu (1980) has indicated that it is *prima facie* very odd to equate traditional

religious belief in west Africa with modern Western scientific theory when the obvious analogy is traditional Western religious belief. Kwame Anthony Appiah (1982), beginning with Wiredu's observation, has argued that the reason the parallel between science and religion is misleading is not, as the symbolists held, that religious appeals to spirits are not meant literally. Rather, Appiah suggested that religion has changed a great deal in the modern West over the centuries, particularly the religious life of intellectuals which has turned increasingly towards 'the contemplative, conceived of as spiritual intercourse with God'. Technical questions have 'remained recalcitrant to scientific investigation - questions about one's relations with others - and questions that could not even in principle be addressed by science - questions of value' (1982: 186). This change makes for substantial differences between the religious life of intellectuals in the industrialized world and that of traditional cultures.

There is a further crucial change, Appiah argues, in the nature of contemplative religion in the West. As interpersonal relations have become less ceremonious, so have private religious acts. Since the reformation, Christian prayer has become more like intimate conversation. The ceremoniousness, or ritual character of religious activity in traditional cultures is not analogous in the world or the practice of science. Appiah also argues that there is more of a fundamental reason why the equation of religion and science is misleading, stating that the social organization of inquiry in modern cultures is radically different from its traditional counterparts. Horton had acknowledged this in his initial discussions of the contrast between 'open' and 'closed' cultures. Much attention has been devoted to criticizing Horton's account and trying to alter it.

5 Critiques of the open-closed dichotomy

Barry Hallen (1977) based his work on his experience of philosophical discussions with Yoruba diviners and healers. He argued that there are certainly African religious traditions that show an awareness of other traditions. Hallen takes as his model Karl Popper's characterization of critical reflection on tradition. This is a significant gesture considering the Popperian provenance of the open-closed dichotomy which identifies the tradition as a tradition, displays an awareness of its consequences and is aware of at least one alternative and might choose to affirm or reject it. These tests show that the Yoruba diviners are critically appreciative of their tradition (see Yoruba epistemology).

In response to Hallen's critique, Horton chose to speak not of the closed nature of traditional belief systems but, borrowing a term from Wole Soyinka (1976), of their being 'accommodative'. He discussed work by students of UK anthropologist E.E. Evans-Pritchard (1937), such as G. Lienhardt's discussion of Dinka religion in *Divinity and Experience* and J. Middleton's *Lugbara Religion*, which not only addressed the kind of static body of belief captured in Evans-Pritchard's picture of the Azande thought world, but also stressed the dynamic and, as Horton came to admit, 'open' way in which they 'devise explanations for novel elements in… experience' and 'their capacity to borrow, re-work and integrate alien ideas in the course of elaborating such explanations'. He claims that it is this '"openness" that has given the traditional cosmologies such tremendous durability in the face of the immense changes that the 20th century has brought to the African scene' (Appiah 1987: 226).

Horton contrasted this accommodative style with the 'adversary' style of scientific theory, characterized by the way in which the main stimulus to change of belief is not 'novel experience but rival theory' (Appiah 1987: 226). This change from the Popperian terminology of 'open' and 'closed' allows Horton to reframe the difference between traditional religion and science as not being related to the individual cognitive strategies, but with social ones.

Evans-Pritchard (1937) argued in his classic work that the Azande were not 'experimentally inclined'; that although inclined to scepticism, their scepticism never reached to the level of general theory; that they did not share their experiences and that their beliefs were 'generally vaguely formulated' (1937: 202-4). However, none of this is true of science: it is centrally experimental; scepticism about general theories is one hallmark of great scientists; information is widely disseminated and precision of formulation is regarded as crucial. Each of these differences, as Appiah argued, is central to the social organization of inquiry.

Anthropologist Jack Goody (1977) argued that a key precondition of these forms of social organization of knowledge is the development of distributed literacy. He claimed that oral cultures have been limited by their inability to reaffirm what other theorists have written against experience. It is this fact, Horton went on to argue(1993: 161-93), that accounts for the possibility of the accommodative style. Literacy makes possible the

precise formulation of the issues under discussion as being characteristic of scientific theory. Precise formulation allows inconsistencies to be recognized.

6 The devout opposition

Perhaps the least known of Horton's controversies outside Africa is the best known among students of African religion because he has taken on a consensus view of African traditional religion developed by a group he dubbed 'the devout opposition'. This group included many Christian professors working in such places as Nigeria. For these scholars 'the focal object' (1993: 165) of African religion is the Christian God. Other spirits are regarded as his agencies and the attitude of believers towards God is one of awe. The group also thought that the aim of religious life was to achieve communion with God. Horton's views are summarized in the form of a debate with this group in his later work (1993) in which he argued, first, that the ethnographic evidence does not support these claims and that it is the desire not to denigrate traditional belief, combined with Christian theology, that leads the devout opposition to their views. He says that:

> 'It is not surprising that the clearest indications of the ideological character of the "devout" position should come from African rather than Western scholars. After all, it is *their* non-Christian kith and kin whose status is at stake… it is clear that there is a strong link… between establishing that African religions show the essential characteristics of True Religion and establishing the human worth and dignity of Africans'.
>
> (1993: 191)

Horton wanted to argue that scholarly discussion of religion can proceed while suspending the issue of whether monotheism is true. However, Appiah (1993: 7) suggested that if there is a God who makes himself known, albeit obscurely, in African religious experience, then his existence and these experiences may be as relevant to understanding the beliefs of Africans as any other facts about the world in which they live. Horton does not pretend to be a theist, however, his argument with the devout opposition is both anthropological and theological. Appiah felt that Horton was wrong in his contention that the question of God's existence is irrelevant to the philosophical study of religion.

See also: Latin America, Pre-Columbian and indigenous thought in; Religion and science

K. ANTHONY APPIAH

References and further reading

Appiah, K.A. (1987) 'Old Gods, New Worlds: Some Recent Work in the Philosophy of African Traditional Religion', in G. Fløistad (ed.) *Contemporary Philosophy*, vol. 5: *African Philosophy*, Boston, MA: Kluwer, 1987: 207-34.(A review of African debates about Horton's work.)

Appiah, K.A. (1982) *In My Father's House: Africa in the Philosophy of Culture*, London: Methuen; 2nd edn, Oxford: Oxford University Press, 1992.(Chapter 6 includes an extended discussion of the character of traditional religious thought.)

Appiah, K.A. (1993) 'Invisible Entities', *Times Literary Supplement*, 17 July: 7.(A review of Horton's *Patterns of Thought in Africa and the West*, 1993.)

Beattie, J. (1973) 'Understanding traditional African religion: a comment on Horton', *Second Order* 2 (2): 3-11. (A symbolist anthropologist's response to Horton's original claims.)

Eddington, A. (1928) *The Nature of the Physical World*, Cambridge: Cambridge University Press.(An examination of Eddington's view that everyday objects have duplicates that can be described in the language of physical theory.)

Evans-Pritchard, E.E. (1937) *Witchcraft, Oracles and Magic among the Azande*, abridged with intro. by E. Gillies, Oxford: Oxford University Press, 1976.(An anthropological classic that has been the focus of much philosophical discussion.)

Evans-Pritchard, E.E. (1956) *Nuer Religion*, Oxford: Oxford University Press.(This work discusses aspects of Nuer ritual and belief that are closer than Zande magic to what is ordinarily thought of as religion in the West.)

Gjertsen, D. (1980) 'Closed and open belief systems', *Second Order* 7 (1): 5-69.(A critique of Horton based on more recent history of science.)

Goody, J. (1977) *The domestication of the savage mind*, Cambridge: Cambridge University Press.(An important discussion of the significance of literacy for intellectual life.)

Hallen, B. (1977) 'Robin Horton on Critical Philosophy and Traditional Thought', *Second Order* 6 (1): 81-92.

(Hallen criticizes Horton's claims about the uncritical character of traditional thought by reference to Yoruba diviners.)

Horton, R. (1967) 'African traditional religion and Western science', *Africa* 37 (1-2): 50-71; 155-87.(Horton's classic paper.)

Horton, R. (1976) 'Understanding traditional African religion: a reply to Professor Beattie', *Second Order* 3 (1): 3-29.(A response to the Beattie paper, 1973.)

Horton, R. (1987) 'Traditional Thought and the Emerging African Philosophy Department: A Reply to Dr Hallen', in G. Fløistad (ed.) *Contemporary Philosophy*, vol. 5: *African Philosophy*, Boston, MA: Kluwer, 1987: 207-34. (An unpublished paper in which Horton responded to Hallen's discussion of traditionalism. The paper appeared in K.A. Appiah's 'Old Gods, New Worlds: Some Recent Work in the Philosophy of African Traditional Religion'.)

Horton, R. (1993) *Patterns of Thought in Africa and the West*, Cambridge: Cambridge University Press. (Summarizes the most significant debates in African philosophy of religion from the point of view of a leading partisan.)

Horton, R. and Finnegan, R. (eds) (1973) *Modes of Thought*, London: Faber & Faber.(Contains many important articles relevant to evaluating Horton's position.)

Popper, K. (1962) 'Towards a rational theory of tradition', in *Conjectures and Refutations*, New York: Basic Books.(The account of a critical tradition taken up by Hallen in his critique of Horton.)

Pratt, V. (1972) 'Science and traditional religion. A discussion of some of Robin Horton's views', *Second Order* 1 (1): 7-20.

Skorupski, J. (1976) *Symbol and theory*, Cambridge: Cambridge University Press.(A philosophical discussion of symbolism, including a review of the debate between Horton and Skorupski about religious and theoretical identifications.)

Soyinka, W. (1976) *Myth, literature and the African world*, Cambridge: Cambridge University Press.(A major discussion of contemporary African intellectual life, especially as it relates to drama and the novel, in which Soyinka introduces the contrast between accommodative and adverserial intellectual styles.)

Wiredu, K. (1980) *Philosophy and an African Culture*, Cambridge and London: Cambridge University Press. (This book begins with important discussions of the role of philosophy in contemporary Africa and takes up questions raised by Horton in that context.)

Agnosticism

In the popular sense, an agnostic is someone who neither believes nor disbelieves in God, whereas an atheist disbelieves in God. In the strict sense, however, agnosticism is the view that human reason is incapable of providing sufficient rational grounds to justify either the belief that God exists or the belief that God does not exist. In so far as one holds that our beliefs are rational only if they are sufficiently supported by human reason, the person who accepts the philosophical position of agnosticism will hold that neither the belief that God exists nor the belief that God does not exist is rational. In the modern period, agnostics have appealed largely to the philosophies of Hume and Kant as providing the justification for agnosticism as a philosophical position.

1 Degrees of agnosticism

Although the philosophical position described as ‘agnosticism’ - scepticism with respect to the existence or nonexistence of a supernatural divine being - has a long history, the term itself was introduced by Thomas H. Huxley in 1869 in order to provide an ‘ism’ that described his own intellectual outlook on matters of religion.Huxley held that neither belief nor disbelief in the existence of God or some supernatural divine reality is warranted, because in his judgment we are simply unable to discover sufficient rational grounds to support either belief or disbelief. Thus agnosticism as a philosophical position is best understood as the view that it is beyond our cognitive powers to determine the existence or nonexistence of God or some divine reality responsible for the existence of the natural universe.

We can distinguish three sorts of agnostic - weak, moderate and strong:

(1) An agnostic (weak sense) is one who understands the concept of God and/or the concept of a supernatural divine reality, but who neither believes nor disbelieves in the existence of God or in the existence of a divine reality responsible for the existence of the natural universe. This is the popular sense of the term.
(2) An agnostic (moderate sense) is one who is an agnostic in the weak sense, but who also holds that human beings are unable to discover sufficient reason to believe or disbelieve in the existence of such a God or divine reality.
(3) An agnostic (strong sense) is one who is an agnostic in the moderate sense, but who also holds that it is wrong or in some way improper to believe or disbelieve in the existence of such a God or divine reality unless one has a sufficient reason to believe or disbelieve.

A person may be an agnostic (weak sense) without holding that it is beyond human cognitive abilities to discover sufficient reason to believe in God. For one may make a correct assessment of one’s own lack of sufficient reason for belief or disbelief without affirming that no human being is without such reason. An agnostic in the weak sense need not hold the philosophical position of agnosticism, for that position requires that one hold that human cognitive powers are inadequate to justify belief with respect to the existence or nonexistence of God. On the other hand, one can hold the philosophical position of agnosticism without being an agnostic at all, whether weak, moderate or strong. Some religious thinkers (Kierkegaard, for example) have held that it is proper to believe in God by faith even though reason cannot provide sufficient rational grounds for or against the existence of God. Thus a religious believer may accept the philosophical position of agnosticism without being an agnostic in the sense of someone who neither believes nor disbelieves in God.

The moderate agnostic and the strong agnostic hold the philosophical position of agnosticism, although the moderate agnostic need not be critical of the religious believer or disbeliever. For a moderate agnostic may allow that it is not improper to believe or disbelieve without adequate grounds provided by human reason. Only the strong agnostic must be critical of the believer and disbeliever. It is fair to say that most philosophers who have been agnostics have been strong agnostics.

2 Justifications for agnosticism

Justifying agnosticism as a philosophical position requires a careful investigation of the limits of our cognitive powers. Specifically, it must be shown that human reason is simply incapable of reaching either affirmative or negative judgments concerning the existence of the God of traditional theism or any sort of divine reality

responsible for the existence of the natural universe. Huxley and other agnostics professed to find support for their view in the philosophies of Hume §6 and Kant §8. Hume had presented devastating critiques of traditional natural theology in his *Enquiry Concerning Human Understanding* (1748) and his *Dialogues Concerning Natural Religion* (1779). And although Kant, in his *Critique of Pure Reason* (1787), sought to counter what he took to be Humean scepticism regarding scientific knowledge, he repudiated all attempts at speculative metaphysics and natural theology, arguing that such efforts exceed the limits of pure reason.

In the nineteenth century the legacy of Hume and Kant flourished in the agnosticism of Sir William Hamilton, Herbert Spencer, Thomas Huxley, Leslie Stephen and John Stuart Mill. In his influential essay 'Theism' (1874), Mill contended that only the argument from design remained as a potential source for rational support for some form of divine reality responsible for the order, if not the existence, of the natural universe (see God, arguments for the existence of §§4-5). He noted Darwin's competing explanation involving the struggle for survival and natural selection. But since Mill regarded Darwin's theory as a legitimate but unproven hypothesis, he argued that the apparent design in animal life gives some degree of probability to the view that there is an intelligent being which, though finite in power, is responsible for the order in (but not the existence of) the natural world. However, this degree of probability is for Mill apparently insufficient to warrant belief, for he concludes his essay with the judgment that the rational position with respect to the supernatural is agnosticism as distinguished from either belief or disbelief.

In the twentieth century, two philosophical movements gave at least indirect support to agnosticism: logical positivism and naturalism. Logical positivists held that a statement is cognitively meaningful (asserts something true or false) only if it is either analytic or empirically verifiable in principle. They also maintained that statements about the God of traditional theism (or other supernatural entities) are neither analytic nor empirically verifiable in principle. The result is that logical positivism denies the assumption of agnosticism that theism and atheism are intelligible positions, while affirming its conclusion that neither belief nor disbelief in God is rational. Naturalism's basic thesis is that the only things about which reliable knowledge can be obtained are things that can be investigated by the methods of science. This thesis implies agnosticism concerning the supernatural in so far as the supernatural eludes investigation by the methods of science. Since naturalists tend to hold that the only individual things of whose existence we have reliable knowledge are physical things, the conclusion is drawn that the existence or nonexistence of God is unknowable. (Some naturalists have gone on to argue that we do have sufficient evidence for atheism.)

With the collapse of logical positivism and the resurgence of philosophy of religion as a discipline in the latter half of the twentieth century, significant challenges to agnosticism and naturalism have been advanced by philosophers seeking either to establish the truth of traditional theism or to establish that belief in theism is rational. The effort to establish that belief in God is rational has taken three directions. First, against the legacy of Hume and Kant it has been argued that there are truth-conducive reasons or evidence for belief in God. Second, following William James' classic essay 'The Will to Believe' (1897), it has been argued that pragmatic (but non-truth-conducive) reasons for belief in God are sufficient to render belief rational (see James, William §§4-5). Third, some philosophers (such as Plantinga 1981) have argued that belief in God is rational because it is a justified basic belief, rather than a belief justified by evidence. Agnostics, for the most part, have been concerned only with the issue of whether there are sufficient truth-conducive reasons to support belief or disbelief in a supernatural, divine reality.

See also: Atheism; Natural Theology; Religion, history of philosophy of §8; Religion and Epistemology; Religious Experience §§2-3

WILLIAM L. ROWE

References and further reading

Hume, D. (1748) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press, 1975.(Presents Hume's mature philosophy concerning human knowledge and conduct. The first of the two Enquiries contains his important essay attacking the rationality of belief in miracles.)

Hume, D. (1779) *Dialogues Concerning Natural Religion*, ed. with intro. by N. Kemp Smith, Indianapolis, IN: Bobbs-Merrill, 1962.(Hume's major critique of the argument from design. The cosmological argument is also

criticized.)
Huxley, T. (1893-95) *Collected Essays*, vol. 5, London: Macmillan.(In ‘Agnosticism’ and ‘Agnosticism and Christianity’ Huxley develops his views about agnosticism in relation to belief in religion, particularly Christianity.)
James, W. (1897) ‘The Will to Believe’, in *The Will to Believe and Other Essays*, New York: Dover, 1956. (Presents James’ famous argument for the right to believe in God in the absence of sufficient evidence to confirm that God exists.)
Kant, I. (1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 2nd edn, 1933.(Kant’s major work on the powers and limitations of human reason. Difficult reading.)
Krikorian, Y. (ed.) (1944) *Naturalism and the Human Spirit*, New York: Columbia University Press.(Contains a number of classic essays by American philosophers in defence of philosophical naturalism.)
Mill, J.S. (1874) *Three Essays on Religion*, London: Longmans, Green & Co. (Contains Mill’s important essays relevant to religion: ‘Nature’, ‘Utility of Religion’ and ‘Theism’.)
Plantinga, A. (1981) ‘Is Belief in God Properly Basic?’, *Nous* 15 (1): 41-51. (Argues that some beliefs are rational in the absence of evidence and suggests that belief in God is one of them.)
Quinn, P. (1985) ‘On Finding the Foundations of Theism’, *Faith and Philosophy* 2 (4): 469-86.(Critically discusses the view that belief in theism is rational because it is properly basic.)

Agricola, Rudolph (1444-85)

Rudolph Agricola was one of the leading humanists of northern Europe in the late fifteenth century. His polished Latin style, his Greek learning and his knowledge of classical literature made him a hero to Erasmus, More, Vives, Melanchthon and Ramus. His major work, De inventione dialectica (On Dialectical Invention) (1479), provides an original account of practical argumentation by combining elements from the established teachings of rhetoric and dialectic with analysis of passages from classical literature. It includes a new version of the topics of invention, based on Cicero's method of devising arguments, outlined in his Topics. Agricola's letter De formando studio (On Shaping Studies) (1484), which circulated widely in the sixteenth century, outlines a plan of knowledge and discusses methods of study. Although his approach was strongly humanist and the Roman rhetorician Quintilian was his favourite author, his logic remained firmly Aristotelian, unlike that of his predecessor Lorenzo Valla. He remained aware of the achievements of scholasticism, expressing admiration for Duns Scotus and adopting an extreme realist position in metaphysics.

1 Life

Rudolph Agricola (born Roeloff Huusman, at Baflo, near Groningen in the north east Netherlands) received a scholastic education in Belgium at Louvain (MA 1465) before moving on like many other northerners to study law in Italy at Pavia. He neglected his legal studies to concentrate on reading classical Latin literature and improving his Latin style. In 1475 he transferred to Ferrara in order to improve his knowledge of Greek. He served Duke Ercole I d'Este as organist for a time and made the acquaintance of prominent humanists such as Battista Guarini and Ermolao Barbaro. In 1479 he returned to Germany with his friend Dietrich von Plieningen, pausing in Dillingen to complete *De inventione dialectica (On Dialectical Invention)* (1479), before returning to Groningen. He was *secretarius* of Groningen from 1482-4, undertaking diplomatic missions on behalf of the town. He was able to return to full-time study, and to begin learning Hebrew, in 1484 when he went to Heidelberg to assist his friend Johann von Dalberg, who had recently become Bishop of Worms. He taught on the fringes of the university and twice preached to the clergy of the diocese. He died on 27 October 1485 shortly after returning from a journey to Rome to congratulate the newly elected Pope Innocent VIII.

Agricola was fortunate to have spent almost ten years in Italy studying Latin and Greek. In Pavia and Ferrara, he lived at the centre of a group of northern students whose unofficial humanistic studies he supervised and with whom he shared his discoveries. He looked on himself as a perpetual beginner, always learning languages while others went on to higher studies. He planned to devote his old age to biblical studies. For the generation of Erasmus he was most important as an example, the Frisian who had equalled Italians in humanistic learning, since his major work did not appear in print until 1515 (see Humanism, Renaissance §§5-6).

2 *De inventione dialectica*

Dialectical invention for Agricola involves not only finding persuasive material for a composition, but also putting it into an effective order. The core of his *De inventione dialectica* is the topics, the subject of Book 1. Agricola's new version provides a full and clear account of how the topics are used, a better organized list of topics, and a carefully exemplified account of the variety within each topic and the uses of each.

The topics are a list of headings (for example, definition, genus, cause, and effect) which can be applied to any subject. Agricola explains that while the things in the world and the connections between them are infinite in number, nevertheless there are certain common characteristics in the links between things, which can therefore be divided into classes. These classes of relations are the topics. In other words the topics are a useful way of investigating the things which are connected to particular subjects because they correspond to connections which exist in the world. This fits in with Agricola's realist position in metaphysics, and it provides (as no one had before) an explanation for the effectiveness of the topics, but it evidently suits some topics (genus, cause) better than others (etymology, similars).

None of his predecessors had attempted to explain how the (rather heterogeneous) list of topics works or why it is complete, whereas Agricola divides the topics into seven groups in line with the degree of closeness to the subject, ranging from 'within the substance' to 'opposites'. Although this list improves on previous accounts by Aristotle

(§7) (whose conception of the topic is rather different), Cicero (§2) and Boethius (§3), it remains inconsistent in places. Agricola's new topics entries are much longer than those of his predecessors. His approach is practical (showing how a definition is built up), inquiring (thinking about the different relations between causes, intermediate ends and effects) and literary (taking apart a simile by Lucan and exploring the implications of alternative objects of comparison).

In Book 2 Agricola shows how arguments derived from the topics can be used in real argumentative situations, looking at how material can be prepared for topical invention, analysing argumentative and expository writing, discussing the forms of argumentation appropriate in practical use, and providing a number of drills to familiarize the student with the use of the topics. One of these exercises involves analysing the argumentative structures which underlie passages of writing. Agricola himself used this method in his dialectical commentary on Cicero's oration *On the Manilian Law*, which in turn inspired Melanchthon's commentaries (see Melanchthon §§1-2) and the analyses of Ramus (§2) and his followers.

Book 3 shows how the topics can be used to produce material which will move and please an audience as well as teach them. It includes a discussion of emotion which relies on, and refers the reader to, Aristotle's *Rhetoric*. Agricola also revitalizes the rhetorical art of disposition by illustrating the range of possible forms of organization (both overall and local) available to a writer, and by showing how decisions about ordering involve considering subject-matter, audience and the writer's own persona and intention.

Agricola's work has often been linked with Lorenzo Valla's attack on Aristotelian logic in his *Dialectica* (see Valla, L. §3), but in fact Agricola endorses many Aristotelian doctrines which Valla rejects. The two men share an interest in the topics and the aim of writing a logic suitable for practical arguing in neoclassical Latin.

Manuscripts of *De inventione dialectica* were hard to find in the twenty years after Agricola's death, but after 1515 it became one of the most influential dialectic books of the sixteenth century, with forty-four editions of the (very long) text, and thirty-two editions of various epitomes. Praise of Agricola and details from his work can be found in many Renaissance dialectic books, notably in the works of Erasmus, Vives (§4), Melanchthon and Ramus.

3 Other works

Agricola produced translations, orations, letters, poems and treatises. Many of his works are pedagogic in their aim, notably his Latin translation of the *Progymnasmata* (short writing exercises) of Aphthonius, a Greek rhetorician of the third century AD, which in its adaptation by Lorichius became one of the most printed school textbooks of the sixteenth century. His minor works also offer two discussions of the nature and organization of philosophy. The *Oratio in laudem philosophiae (Oration in Praise of Philosophy)* was delivered in autumn 1476 to inaugurate the academic year in Ferrara. Agricola praises philosophy because it investigates everything, because it can turn everything to good, because it enables a person to rise above the vicissitudes of fortune and because the soul knows itself happy in self-contemplation.

Philosophy is divided into rational (grammar, dialectic and rhetoric), natural (physics, which includes medicine, mathematical arts and theology) and moral (politics, justice), according to Agricola. The general scheme seems to derive from the division of philosophy found in Isidore of Seville's *Etymologiae* (see Encyclopedists §7), but Agricola's placing of theology within natural philosophy is original. Agricola's letter *De formando studio (On Shaping Studies)* (1484) chooses philosophy as the subject of study for the addressee, the Antwerp musician Jacques Barbiriau, because of his ability and because he does not require that his studies provide him with income. Philosophy in other words is the highest of studies, but is only open to people with some independence. After setting aside the *trivium* (grammar, rhetoric and dialectic) as the entrance to the arts, not the thing itself, he provides a new division of philosophy into two parts, natural and moral, both of which must be studied using the best authors. Natural philosophy, which includes geography, agriculture, medicine, architecture, painting and sculpture, as well as physics and biology, is valuable in forming the mind because it shows the worthlessness of wealth and the fragility of human bodies. Moral philosophy, the more important part, involves the study of poetry, oratory, history, ethics and sacred writings (both the scriptures and Christian authors).

Agricola discusses three methods of study: reading, the compilation of commonplace books, and the sharing and development of knowledge through composition. In a sense, rather in the manner of Augustine's *De doctrina christiana (On Christian Doctrine)*, a literary approach, and a choice of classical and Christian authors, have taken

over the whole of philosophy (though it should be remembered that the elder Pliny's *Natural History* was one of Agricola's favourite books). Agricola also wrote two small treatises, *Singulares aliquot de universalibus quaestiones (Some Questions about the Universals)* (1539) and the unpublished *De universali singulari et uno (On the Universal, the Singular and the One)*, in which he maintains that all universal terms correspond to real things in the world. These treatises await thorough study, as does Agricola's understanding of Ramon Llull, to whom he makes a tantalizing reference in *De inventione dialectica*.

See also: Humanism, Renaissance; Logic, Renaissance

PETER MACK

List of works

See Akkerman and Vanderjagt (1988) in References and further reading, for details of Agricola's works (some of them unpublished) that are not included in the List of works.

Agricola, R. (1476) *Oratio in laudem philosophiae (Oration in Praise of Philosophy)*, in Alardus (ed.) Lucubrationes, Cologne, 1539; repr. Nieuwkoop: de Graaf, 1967, 144-59.(Oration outlining Agricola's earlier view of the shape and contents of philosophy.)

Agricola, R. (1479) *De inventione dialectica (On Dialectical Invention)*, ed. Alardus, Cologne, 1539; repr. Nieuwkoop: de Graaf, 1967; ed. and German trans. L. Mundt, Tübingen: Niemeyer, 1992; English trans. J.R. McNally, 'Rudolph Agricola's *De inventione dialectica*: A Translation of Selected Chapters', *Quarterly Journal of Speech* 34: 393-422.(Agricola's major work, describing the discovery of persuasive arguments and the composition of texts.)

Agricola, R. (1484) *De formando studio (On Shaping Studies)*, in Alardus (ed.) *Lucubrationes*, Cologne, 1539; repr. Nieuwkoop: de Graaf, 1967, 193-201.(Agricola's later outline of an idealized system of education.)

References and further reading

Akkerman, F. and Vanderjagt, A.J. (eds) (1988) *Rodolphus Agricola Phrisius (1444-1485)*, Leiden: Brill.(A valuable collection of essays with a full secondary bibliography and a list of Agricola's works.)

Green-Pedersen, N.J. (1984) *The Tradition of the Topics in the Middle Ages*, Munich: Philosophia.(A fine treatment of the classical background and medieval use of the topics.)

Huisman, G.C. (1985) *Rudolph Agricola: A Bibliography of Printed Works and Translations*, Nieuwkoop: de Graaf.(The standard bibliography of early printed editions.)

Mack, P. (1993) *Renaissance Argument: Valla and Agricola in the Traditions of Rhetoric and Dialectic*, Leiden: Brill.(An analysis of *De inventione dialectica* with an account of Agricola's sources and influence.)

Vasoli, C. (1967) *La dialettica e la retorica dell'Umanesimo*, Milan: Feltrinelli.(A good account of humanist rhetoric and dialectic.)

Agricultural ethics

Agricultural ethics is the study of moral issues raised by farming. These include: human interference with the course of nature; the effects of certain agricultural practices on present social conditions, and on the conditions under which future generations will live; the treatment of animals, especially when its aim is human advantage; and the value of farming as a human activity in itself.

1 Basic justification

Some anti-agriculturalists defend a return to the life of hunting and gathering. Few philosophers have explicitly defended such a view, but it seems a logical consequence of some positions in environmental ethics. (Taylor 1986), for example, holds that all living things including plants have a *telos*, or 'goal', and that we have at least a corresponding *prima facie* duty not to interfere with them. Most humans could survive, and many could flourish, eating only nuts, berries and vegetable products taken from dead or dying plants. If all living things deserve respect then agriculture, the implements and practices of which are expressly designed to kill targeted plants and animals, might be unjustifiable. Callicott (1989) believes it is our duty genuinely to share the earth with other species, an impossibility when farmers plough up wildlife habitat (see Environmental ethics).

Two popular presentations of the anti-agricultural ideal make explicit its practical implications. In Edward Abbey's novel, *Desert Solitaire*, a character laments the oppressive presence of humans in the United States' Southwest, and opines, 'I'd rather shoot a man than a snake.' In Daniel Quinn's novel, *Ishmael*, a gorilla explains that the majority of humans are 'Takers', who have deprived the world of its wildness and diversity. The preferred form of human life from the gorilla's perspective is that of hunting and gathering in which 'Leavers', eschewing the arts of cultivation, ensure the integrity of nonhuman planetary life.

A more anti-humanistic philosophy seems hardly imaginable when, as Callicott puts it, the measure of a truly ecocentric ethic is the extent of its misanthropy. It would seem to be one of our basic duties, commensurate with others' basic moral rights, to endeavour to feed the world's hungry (see Development ethics; Justice, international). To abandon the arts of cultivation would result in our failing to meet this duty. The justification of the practice of agriculture is secured by whatever arguments justify the existence of the most basic of duties.

Many Jewish and Christian theologians formulate duties to nature in terms of stewardship, holding that the earth is a gift of God to humans so that we may use but not abuse soil, water, air and animals. Similarly, many secular philosophers believe that we are justified in cultivating the earth and breeding plants and animals selectively, if we do so in a sustainable way: the entitlement to treat plants and animals as things of instrumental value only is circumscribed by duties to future generations, humans who will need adequate natural resources to grow crops (see Future generations, obligations to).

2 Social justice

Concerns about fairness in the distribution of food and farmland have been raised in both developing and developed countries. Most of the world's poor are small tenant farmers. In order to increase the standard of living of these farmers, the governments of many developing countries adopted in the 1970s the policy of 'industrializing' agriculture; urging their farmers to copy the model of large successful farmers in developed countries. During the green revolution of the 1960s and 1970s, countries such as India, Costa Rica and Nigeria increased the efficiency of farmers' yields by borrowing money from international lending agencies such as the World Bank. The funds were used to extend credit to farmers, who were taught to buy high yielding varieties of seeds (such as rice, wheat, and maize) and to use the necessary accompanying technologies: mechanical implements (tractors) and synthetic chemicals (herbicides and pesticides). Many farmers flourished and nations that once imported grain became self-sufficient in certain crops.

Questions were raised, however, about the equity of the strategy. Critics alleged that industrial farming benefited larger farmers unfairly because they had easier access than small farmers to credit and expanded landholding. As crops were grown in greater abundance, the price farmers received for each bushel decreased and producers were forced to try to spread their costs over more acres. Were the poor and hungry actually disadvantaged by the industrialization of agriculture? Were small tenants dispossessed of land unjustly when larger farmers,

beneficiaries of the new technologies, bought up their smallholdings? Some argue that they were (Lappe and Collins 1979), others that they were not (Ruttan and Hayami 1984). The debate turns on the resolution not only of important empirical questions (for instance, did industrial agriculture reduce opportunities for labour employment and earnings?) but also of significant philosophical questions (for example, is it obligatory or supererogatory to aid unfortunates in other nations?).

In developed countries debate about social justice in agriculture sometimes takes as a focus the structure of the agricultural industry. In the United States, for example, the question has been expressed in terms of the desirability of 'saving the family farm'. Family farms are medium-sized businesses owned, worked and loved by families, the kind of farm being displaced by smaller hobby farms, on which the majority of income derives from off-farm activities, and by large super farms, often worked by hourly employees who are not stakeholders.

Questions to be addressed here include: Do family farmers practice better stewardship of the land than other farmers? Are rural communities better places to live if they are surrounded by many medium-sized farms rather than a few large farms? Are farm animals treated more humanely on family farms? Can smaller farms take advantage of economies of scale and produce food as efficiently as larger farms?

Another issue concerns the role of governments in agriculture. Should public policy target benefits and subsidies at medium-sized farms, and not at hobby or super farms? Or are such policies inherently unfair in so far as they do not benefit all farms equally?

Finally there are social justice questions related to pesticides and farmworker and consumer health. It has been argued on deontological grounds, for example, that farmers are morally unjustified in using chemicals that are carcinogenic to consumers (see Risk).

3 Nonhuman animals

Perhaps the most controversial matter in contemporary agricultural ethics concerns the moral standing of nonhuman animals. Some, such as Regan (1983), argue that it is morally wrong to raise and slaughter animals for food because farm animals typically are 'subjects of a life' with intrinsic value and basic moral rights of their own. Others argue that animals lack moral rights because they lack conscious experiences, moral autonomy and a sense of justice, and that it is therefore permissible to use cows and chickens in humane ways (see Moral standing; Animals and ethics). Utilitarians generally believe that animal pain counts morally, but they differ over whether the benefits of using animals in agriculture outweigh the costs. The issue gains urgency with the development of powerful new scientific techniques to manipulate the animal genome (see Genetics and ethics §3). As subjects of genetic engineering, farm animals have suffered from unintended deleterious effects, while research animals have suffered the consequences of being intentionally bred for propensity to develop debilitating diseases.

4 Virtue

If we believe Xenophon in the *Oeconomicus*, Socrates once said that 'the best kind of work and the best kind of knowledge is farming, by which human beings supply themselves with necessary things'. While some believe that the past ten thousand years of agriculture has led inevitably to irreversible catastrophic environmental degradation, many affirm with Xenophon's Socrates that there is no better work or knowledge than farming. What did Socrates mean by the idea that farming provides the best kind of knowledge? Perhaps he meant what Wendell Berry meant when he wrote that it is 'a law' that

> land that is in human use must be lovingly used; it requires intimate knowledge, attention, and care…. A family that has farmed a farm through two or three generations will possess not just the land but a remembered history of its mistakes and of the remedies of those mistakes.
>
> (1987: 349)

Why should such knowledge be 'the best kind'? Perhaps because in it the intellect is uniquely connected with the body, and spirituality to physicality. As Berry puts it, those who farm 'gain the means of life; … they gain the longevity and dependability of sources of food, both natural and cultural. [On a farm] the proper answer to the spiritual calling becomes, in turn, the proper fulfilment of physical need' (1987: 351).

To farm may be to practice a virtuous calling, an art with its own intrinsic rewards (see Virtue ethics). For a people

to become landless, or to become utterly dissociated from the means by which their most basic physical needs are met, may mean they are destined to become bereft not only of the best kind of work, but of the best kind of knowledge as well.

See also: Applied ethics; Business ethics; Technology and ethics

GARY L. COMSTOCK

References and further reading

Abbey, E. (1968) *Desert Solitaire*, New York: Ballantine.(A novel, set in the author's home, the American Southwest. Very influential for the ecology movement in the United States.)

Aiken, W. and LaFollette, H. (eds) (1977) *World Hunger and Moral Obligation*, Englewood Cliffs, NJ: Prentice Hall.(Widely used collection, including seminal essays by G. Hardin, P. Singer, J. Arthur, J. Narveson, W. Frankena, O. O'Neill, and J. Rachels.)

Berry, W. (1978) *The Unsettling of America: Culture and Agriculture*, New York: Avon.(Very influential book that argues culture in the United States has declined as more and more of its population has moved off the farm.)

Berry, W. (1987) 'A Defense of the Family Farm', in *Home Economics: Fourteen Essays*, San Francisco, CA: North Point Press; repr. in G. Comstock (ed.) *Is There a Moral Obligation to Save the Family Farm?*, Ames, IA: Iowa State University Press, 1987, 347-60.(Berry, a farmer, poet, and essayist, defends family farms on cultural, moral and spiritual grounds.)

Blatz, C. (1991) *Ethics and Agriculture: An Anthology on Current Issues in World Context*, Moscow, ID: University of Idaho Press.(Sections on agriculture's aims, practitioners, conduct and development. Includes a useful bibliography.)

Callicott, J.B. (1989) *In Defense of the Land Ethic: Essays in Environmental Philosophy*, Albany, NY: State University of New York Press.(Argues that animal liberation and environmental ethics are incompatible with the land ethic and attempts to develop a non-anthropocentric holistic ethic.)

Comstock, G. (ed.) (1987) *Is There a Moral Obligation to Save the Family Farm?*, Ames, IA: Iowa State University Press.(A collection of essays from theologians, sociologists, economists and farmers focused on the social, economic, and moral dimensions of the 'farm crisis' in the United States during the 1980s.)

Comstock, G. (1992) 'Should We Genetically Engineer Hogs?', *Between the Species* 8: 196-202.(Argues that we have a *prima facie* duty not to interfere with the interests of animals, and that most genetic engineering of animals is morally indefensible in so far as it leads to the frustration of their desires.)

Frey, R.G. (1980) *Interests and Rights: the Case Against Animals*, Oxford: Clarendon Press.(Argues that animals lack moral rights because they lack language and the ability to take an interest in things.)

Gewirth, A. (1980) 'Human Rights and the Prevention of Cancer', *American Philosophical Quarterly* 17: 117-26. (Argues that the use of carcinogenic chemicals in farming is indefensible.)

Lappe, F.M. and Collins, J. (1979) *Food First: Beyond the Myth of Scarcity*, New York: Ballantine.(An early and seminal statement of the case against viewing industrial agriculture as the best way to feed the world's hungry.)

Leopold, A. (1949) *A Sand County Almanac: With Essays on Conservation from Round River*, Oxford: Oxford University Press; repr., New York: Ballantine, 1966.(An environmental classic containing Leopold's statement of the land ethic.)

O'Neill, O. (1986) *Faces of Hunger: An Essay on Poverty, Justice and Development*, London: Allen & Unwin. (Very useful Kantian approach to the problems of world hunger. Includes discussion of the distinction between special obligations owed to particular persons and general obligations owed to others but not to anyone in particular.)

Quinn, D. (1992) *Ishmael*, New York: Bantam.(A novel in which a gorilla distinguishes between different types of humans - 'Takers', comprising the majority, who have deprived the world of its wildness and diversity, and 'Leavers', who, by hunting and gathering, eschew the arts of cultivation and thus ensure the integrity of nonhuman planetary life. The novel won the Turner Tomorrow Fellowship prize.)

Regan, T. (1983) *The Case for Animal Rights*, Berkeley, CA: University of California Press.(The first and best account of the case for attributing moral rights to nonhuman animals based on their ability to be what Regan calls 'subjects of a life'.)

Ruttan, V. and Hayami, Y. (1984) 'The Green Revolution: Inducement and Distribution', *The Pakistan Development Review* 23: 37-63.(Argues that developing countries must develop more productive agricultural technologies to offset growing populations and to achieve both growth and equity.)

Taylor, P. (1986) *Respect for Nature: A Theory of Environmental Ethics*, Princeton, NJ: Princeton University

Press.(Articulates and defends a biocentric theory of environmental ethics in which all living things possess inherent worth.)

Xenophon (*c*.360s BC) *Oeconomicus*, trans. E.C. Marchant, Cambridge, MA: Harvard University Press, 1925. (Discusses proper management of a household and presents didactic material on agriculture, within a Socratic dialogue.)

Agrippa (1st/2nd century AD)

Agrippa, a Sceptic of the first or second century AD, compiled five general modes of Sceptical argument: the views of positive theorists are subject to endemic disagreement due to the relativity of appearances, and adjudication cannot succeed, since it will either be mere assertion (and hence will not command assent) or appeal to further considerations, which process will either be infinitely regressive or circular, or terminate in unfounded assumption.

Agrippa is mentioned only once in our sources (Diogenes Laertius, IX 88), where no information is given about him beyond the attribution to him (and his associates) of a set of five modes with which to commend Pyrrhonian Scepticism (see Pyrrhonism). Sextus Empiricus ascribes these modes only to 'the more recent Sceptics' (*Outlines of Pyrrhonism* I 164), with no mention of Agrippa at all; nor does Agrippa's name figure in the list of prominent Pyrrhonians with which Diogenes Laertius closes his life of Pyrrho (IX 115-16).

The basic argument schemata of Pyrrhonism - Aenesidemus' ten modes (see Aenesidemus; Pyrrhonism §2) - collect different types of cases of opposing appearances of one sort or another, and move from the relativity of appearances (and the impossibility of favouring one set over another) to suspension of judgment about the natures of things, although they are less clear as to quite how suspension is to be achieved. The Agrippan modes organize the Sceptical material rather differently, and remedy that last deficiency. The first is that 'according to which we find that an undecidable conflict has arisen among both lay people and philosophers concerning the matter in hand as a result of which, being unable either to accept or reject it, we end up suspending judgment' (*Outlines of Pyrrhonism* I 165). The third asserts that appearances are in general irremediably relative, which is responsible for the dispute in the first place.

By contrast, the remaining modes are general and methodological, designed to undermine the attempts of dogmatists to offer reasoned defences of their positions. The second mode is that of regress: each supposed justification of a position will itself require justification, and so on *ad infinitum*; but such a sequence offers no ultimate justification at all, and again suspension follows. Dogmatists may sometimes simply offer unargued assumptions, or hypotheses, but these are not compelling since in any case we might equally assume the opposite; this forms the basis of the fourth mode (*Outlines of Pyrrhonism* I 173-4). Finally dogmatists will sometimes attempt, whether wittingly or not, to show that their suppositions are reciprocally supporting; but, as the fifth mode demonstrates, if *p* rests on *q* and *q* upon *p*, then neither rests upon anything.

These modes may be (and generally are) deployed in combination against any dogmatist's claim to have established a criterion which might distinguish truth from falsity and thus provide foundations for knowledge (see Pyrrhonism §4). Suppose a dogmatist asserts that *p*. If that is a mere assertion, then the Sceptic produces the fourth mode. If *p* is in turn supported by *q*, and *q* by *r*, and so on, either that procedure terminates somewhere (in which case the fourth mode again becomes operable); or it does not, committing the dogmatist to regress; or, eventually one of the supporting propositions is itself shown to rest on *p*, in which case, as the fifth mode has it, the whole structure is built on sand.

The Agrippan modes are weapons of great scope and power; and Sextus' presentation of the Ten Modes of Scepticism was evidently influenced by them. Indeed, the whole subsequent history of the epistemology of justification may be seen as a series of attempts to evade their purportedly all embracing grasp.

R.J. HANKINSON

References and further reading

Barnes, J. (1990) *The Toils of Scepticism*, Cambridge: Cambridge University Press.(A detailed and subtle philosophical treatment of Agrippa's modes.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(Greek text with facing translation; IX 61-116 is devoted to Pyrrhonism.)

Hankinson, R.J. (1994) *The Sceptics*, London: Routledge.(Chapter X deals with the Agrippan modes.)

Sextus Empiricus (*c*.AD 200) *Outlines of Pyrrhonism*, trans. J. Annas and J. Barnes, *Outlines of Scepticism*,

Cambridge: Cambridge University Press, 1994.(Fine translation with introduction and notes.)

Agrippa von Nettesheim, Henricus Cornelius (1486-1535)

Famous in the sixteenth century for writings in which he steps forward variously as magician, occultist, evangelical humanist and philosopher, Agrippa shared with other humanist writers a thoroughgoing contempt for the philosophy of the scholastics. In his more evangelical moods Agrippa could be taken for a radical exponent of the philosophia Christi *of his older contemporary Erasmus, or mistaken for a follower of Luther, whose early writings he actively disseminated in humanist circles. However, his deepest affinities are with magically inflected philosophies: the Neoplatonism and Hermetism of Marsilio Ficino, and the syncretic Christian Kabbalah of Giovanni Pico della Mirandola, Johannes Reuchlin and Johannes Trithemius.*

As well as expounding an influential magical view of language, Agrippa contributed to the sixteenth-century revival of scepticism, denounced the 'tyranny' of those who obstructed a free search for truth, criticized the subjection of women and (with a courage unusual in his time) resisted and mocked the instigators of the witch-craze. Finding in Hermetic-Kabbalistic doctrines the inner truth both of religion and of philosophy, Agrippa was also aware of parallels between these magical doctrines and the Gnostic heresies. His heterodoxy made him a target for pious slanders: within several decades of his death he became the protagonist of demonological fictions which were soon absorbed into the legend of Dr Faustus.

1 Life

Born to a family of the lesser nobility in Cologne (from whose Latin name, Colonia Agrippina, he drew his humanist cognomen), Agrippa took his first degree at Cologne in 1502; after further studies in Paris and elsewhere, he claimed to have doctorates in canon law, civil law and medicine - and also to have been knighted in recognition of military service.

In 1508 he took part in an unsuccessful military adventure which a secret occultist society, of which he was a member, undertook in Spain, possibly at the behest of Emperor Maximilian I. Members of this society subsequently became prominent in French humanist and court circles, providing Agrippa with a network of supporters upon whom, as his reputation for encyclopedic learning grew, he was able to draw in his searches for patronage. When in 1509 he lectured on Reuchlin's Kabbalist philosophy at the University of Dôle in Franche-Comté and wrote *De nobilitate* (published in 1532), Agrippa had hopes of preferment in the court of Margaret of Austria, Regent of Franche-Comté and the Low Countries. These were dashed when he was denounced at court by a prominent Franciscan as a 'judaizing heretic'. Returning to Germany, in 1510 he completed the first version of *De occulta philosophia* (published in 1533), and in the same year travelled to England, apparently in the service of Maximilian I.

For the first several years of his Italian sojourn, which lasted from 1511 to 1518, Agrippa continued to serve the emperor both as diplomat and soldier. But by 1515 he was lecturing on the *Hermetica* at the University of Pavia - a position which he promptly lost, along with his library and other possessions, after the French victory at Marignano. In 1518 Agrippa moved north again, taking up a position as city orator and advocate in Metz. Intervening there in the case of a woman accused of witchcraft, he secured her freedom, recovered her property, and accused the inquisitor responsible for torturing her of heresy. But this and other instances of resistance to tyranny and obscurantism made him unpopular with the orthodox. He returned to Cologne in 1520, lived from 1521 to 1523 in Geneva (where he was at the centre of a group of reforming tendencies), and then moved to Fribourg (also in Switzerland), where he practised medicine.

In 1524 Agrippa secured a place in the French royal court at Lyons as personal physician to the queen mother, Louise de Savoy. But by 1526 he was in trouble, having rashly revealed his sympathy for the rebellious Duc de Bourbon and Emperor Charles V, who was at war with King Francis I. During the same year Agrippa wrote *De vanitate* (published in 1530), which includes a vehement critique of the corruption and venality of court life. Perhaps as a result, his salary was withheld, while at the same time he was refused permission to leave the court.

Dismissed at last in 1528, Agrippa obtained a place in the court of Margaret of Austria at Antwerp as historiographer to Emperor Charles V. But when Margaret died in late 1530 he was again unable to secure payment for his services. Furthermore, the printing of *De vanitate* in 1530 had earned him condemnation from the

theological faculties of Paris and Louvain, which led to difficulties with the imperial privy council. In 1531 the printing of a much expanded version of *De occulta philosophia* in Antwerp was blocked after the first of its three books had been printed; two years later, thanks to the patronage of the reform-minded Archbishop of Cologne, Agrippa was able to see this book and several others, including *De nobilitate* and a commentary on the art of Ramon Llull, through the press.

Returning in 1535 to Lyons, Agrippa was imprisoned by Francis I for having written against Louise de Savoy. Released through the intervention of friends, he died shortly afterwards in Grenoble.

2 Verbal magic

Agrippa derived from the Neoplatonists (and ultimately of course from Plato's *Cratylus*) the view that the power inherent in natural things lives on and is latent in 'the form of the signification' (1533: I.lxx). Because the hidden powers of things proceed in the first place from celestial causes, and because the celestial powers which move the elemental world, acting from circumference to centre, originate with 'the word of God, which word the wise Chaldeans of Babylon call the cause of causes' (1533: II.lx), it follows that the philosopher or magician whose words can draw upon the power of this originary creative Word should be able to intervene powerfully in the natural order. Agrippa's insistence on the purely natural quality of verbal magic cannot disguise the heterodox implication of this view of language, which is that the magician can get in at the top of the hierarchical structure of the cosmos because his 'mysterious words' and 'ingenious speech' draw upon the power contained within God's Word - a term which refers to the canonical scriptures as well as to Christ, the creative Logos (see Language, Renaissance philosophy of).

3 Agrippa as sceptic and free-thinker

The main purpose of *De vanitate* (1530) is to bring the reader to a position of Christian fideism (though one in which Christian faith is thoroughly infused with Hermetic and Kabbalistic motifs). To this end Agrippa's chapter on logic makes a brief but effective deployment of sceptical arguments. Aristotle's principles of demonstration (see Aristotle §6), he argues, require an understanding of causes and principles to which we give our assent on the basis either of authority or of sense-based experience (for knowledge is agreed to arise from the senses, and Averroes (Ibn Rushd) makes agreement with sensible things a criterion of truth) (see Ibn Rushd §6). But the senses are often deceived, and furthermore cannot attain the intellectual level at which we encounter the causes of lower things. It is therefore manifest that 'the way of the truth is shut up from the senses', and that sciences rooted in them are 'uncertain, erroneous and deceitful' (1530: cap. 7). Appeals to authority are no more acceptable, since the final recourse of the scholastics against those who deny the first principles of their sciences is to violence, 'so that of philosophers they are made torturers and hangmen, since they will compel us by force to confess that which they should teach by reason' (1530: cap. 1) (see Scepticism, Renaissance §2).

4 Agrippa as feminist

In *De nobilitate* (1532) Agrippa argues that 'between man and woman by the substance of the soul one has no higher pre-eminence of nobility above the other, but both have by nature equal liberty of worthiness. Yet in all other respects, apart from the divine substance of the soul, the excellence and nobility of womankind surpasses beyond limit the rude gross nature of men' (*Opera*: sig. Ii 4v). Some of the examples with which he develops this claim are deliberately frivolous, and yet he does insistently challenge the misogynist legal culture by which women, 'being subdued as it were by force of arms, are constrained to give place to men, and to obey their subduers, not by any natural or divine necessity or reason, but by custom, education, fortune, and a certain tyrannical occasion' (*Opera*: sig. Ll 3v). François Rabelais' portrait of Agrippa as Her Trippa, an occultist who is ready to predict Panurge's cuckoldry by all the magical arts at his disposal, while remaining unaware that the court lackeys are lining up to frolic with his own wife, can be understood as a sardonic response to Agrippa's feminism (see Rabelais, F. §3). A more positive response is evident in Johannes Wier's *De praestigiis daemonum (Of Demonic Deceptions)* (1563), a book which in some parts of western Europe had a moderating effect upon the witch-hunts of the time: Wier, who had been Agrippa's student, adopted his opinion that the elderly women who were the prime targets of the witch-hunters were suffering from melancholia rather than demonic possession, and that Christians should give them spiritual and material comfort rather than persecuting and torturing them (see Feminism §2).

5 Agrippa's philosophical influence

The apparent contradiction between the sceptical fideism of *De vanitate* and the encyclopedic syncretism of *De occulta philosophia* is to some extent dissipated by Agrippa's reliance in both books upon a Hermetic-Kabbalistic doctrine of spiritual rebirth and deification. However, Agrippa is neither a coherent nor in most respects an original thinker. His most strongly voiced opinions are often taken verbatim from the works of Marsilio Ficino, Giovanni Pico della Mirandola and Johannes Reuchlin, and since he typically appears more interested in assembling diverse opinions on a subject than in assessing their relative truth, his own impulses may seem more antiquarian than philosophical. (Given the hostility he encountered from theologians of the mendicant orders from 1509 onwards, one may suspect that he was content to allow the material he had assembled to work within the reader's mind, without himself taking the risk of underlining its heterodox implications.)

Agrippa was widely read for well over a century after his death. He was, on the one hand, denounced by John Calvin in *De scandalis (On Scandal)* (1550) as a mocker of sacred truths in the vein of Lucian of Samosata, by Jean Bodin in *De la démonomanie des sorciers (On the Devil-mania of Sorcerers)* (1581) as the leading sorcerer of his age, and by André Thevet in *Les vrais pourtraits et vies des hommes illustres (The True Portraits and Lives of Illustrious Men)* (1584) as having spawned hordes both of scoffers and magicians. On the other hand, his works were cited and echoed by literary figures ranging from Jean de la Taille to Sir Philip Sidney, Fulke Greville, Christopher Marlowe and Thomas Nashe, as well as by occult philosophers from John Dee and Giordano Bruno to Thomas Vaughan. Moreover, Michel de Montaigne's scepticism, which represents man as 'naked and empty, acknowledging his natural weakness, apt to receive from above some strange power, disfurnished of human knowledge, and so much the more fit to harbour divine understanding, nullifying his judgment so as to give more place to faith' ([1580] 1962 (1): 562), is clearly indebted to Agrippa's *De vanitate* (see Montaigne, M. de §3). Perhaps more significantly, it has recently been argued that René Descartes' writings, from the early *Olympica (Revelation from Olympus)* and *Cogitationes privatae (Private Thoughts)* (1619-21) to the *Meditations* (1641), make sustained use of motifs derived from the philosophical *Hermetica*, and that Descartes' understanding of the Hermetic writings was conditioned by his early reading of Agrippa (see Keefer 1996).

See also: Hermetism; Humanism, Renaissance; Kabbalah; Platonism, Renaissance §5; Scepticism, Renaissance §2

MICHAEL H. KEEFER

List of works

Agrippa, H.C. (1486-1535) *Opera (Works)*, Lyons, *c.*1600, 2 vols; facsimile repr. ed. R.H. Popkin, Hildesheim and New York: Olms, 1970.(The most easily accessible collection of Agrippa's writings - though in Latin.)

Agrippa, H.C. (1530) *De incertitudine et vanitate scientiarum et artium atque excellentia verbi dei declamatio*, Cologne; trans. J. Sanford, *Of the Vanitie and Uncertaintie of Artes and Sciences*, London, 1569; repr. ed. C.M. Dunn, Northridge, CA: California State University Press, 1974.(Agrippa's most sustained piece of satirical and polemical writing, this work was frequently reprinted and widely quoted - by, among others, later sixteenth-century religious radicals.)

Agrippa, H.C. (1532) *De nobilitate et praecellentia foeminei sexus*, Cologne; repr. in R.H. Popkin (ed.) *Opera*, vol. 2, Hildesheim and New York: Olms, 1970; trans. T. Clapham, *Of the Nobilitie and Excellencie of Womankynde*, London, 1542.(An important early expression of male feminism; the English translation of this book has been described as 'the most explicitly feminist text to be published in England in the first half of the [sixteenth] century' (Jordan 1990: 122).)

Agrippa, H.C. (1533) *De occulta philosophia libri tres*, Cologne; trans. J. Freake, *Three Books of Occult Philosophy*, London, 1651; repr., ed. D. Tyson, St Paul, MN: Llewellyn Publications, 1993.(The best-known Renaissance encyclopedia of learned magic, this book effectively de-centres orthodox Christianity through its explorations of parallels with Judaic, Muslim and pagan traditions.)

References and further reading

Hermes Trismegistus (pseud.) (*c.*100-300) *Hermetica*, ed. B.P. Copenhaver, Cambridge: Cambridge University Press, 1992.(The best English translation of writings in which Agrippa was obsessively interested.)

Jordan, C. (1990) *Renaissance Feminism: Literary Texts and Political Models*, Ithaca, NY, and London: Cornell University Press.(Includes an analysis of Agrippa's place on the feminist side of Renaissance debates on the status of women.)

Keefer, M.H. (1988) ‘Agrippa’s Dilemma: Hermetic "Rebirth" and the Ambivalences of *De vanitate* and *De occulta philosophia*’, *Renaissance Quarterly* 41: 614-53. (Explores the commonalities and divergences of Agrippa’s two major works in terms of their deployment of motifs derived from the Hermetica and their author’s awareness of parallels between Hermetic texts and Gnostic heresies.)

Keefer, M.H. (1996) ‘The Dreamer’s Path: Descartes and the Sixteenth Century’, *Renaissance Quarterly* 49: 30-76.(Referred to in §5, this essay argues that Agrippa’s writings led Descartes to Hermetic texts which were decisive in shaping his philosophical project.)

Montaigne, M. de (1580) *Essais de Montaigne*, ed. M. Rat, Paris: Éditions Garnier Frères, 1962, 2 vols.(Referred to in §5. Montaigne is both the most engaging and the most influential of sixteenth-century sceptical thinkers; the influence of Agrippa can be detected in the most extended statement of his sceptical philosophy, his ‘Apology for Raymond Sebond’. The first edition of Montaigne’s essays was published in 1580, although the final, most complete version appeared posthumously in 1595.)

Nauert, C.G., Jr (1965) *Agrippa and the Crisis of Renaissance Thought*, Urbana, IL: University of Illinois Press. (The standard biography and the most detailed study of Agrippa’s thought.)

Tomlinson, G. (1993) *Music in Renaissance Magic: Toward a Historiography of Others*, Chicago, IL, and London: University of Chicago Press, 44-66.(The best summary account available of Agrippa’s magic.)

Yates, F.A. (1964) *Giordano Bruno and the Hermetic Tradition*, London: Routledge & Kegan Paul.(This compulsively readable account of Renaissance magical traditions includes a rather dismissive account of *De occulta philosophia*.)

Yates, F.A. (1979) *The Occult Philosophy in the Elizabethan Age*, London: Routledge & Kegan Paul.(Gives greater emphasis than Yates’ previous books to the influence of Kabbalah on Renaissance occult philosophy.)

Zambelli, P. (1969) ‘Cornelio Agrippa, Erasmo e la teologia umanistica’ (Agrippa, Erasmus and Humanist Theology), *Rinascimento* 21 (2nd series, 10): 29-88.(See annotation to Zambelli 1985 for details.)

Zambelli, P. (1976) ‘Magic and Radical Reformation in Agrippa of Nettesheim’, *Journal of the Warburg and Courtauld Institutes* 39: 69-103.(See annotation to Zambelli 1985 for details.)

Zambelli, P. (1985) ‘Scholastiker und Humanisten: Agrippa und Trithemius zur Hexerei’ (Scholastics and Humanists: Agrippa and Trithemius on Witchcraft), *Archiv für Kulturgeschichte* 67: 41-79.(These and other articles by the leading Agrippa scholar have been crucial in situating his writings in relation to Erasmian humanism, the radical Reformation, and overlapping views of magic and witchcraft.)

Ailly, Pierre d’ (1350-1420)

D’Ailly was a prolific writer on a number of subjects. His best known philosophical works concentrate on logic and on faith and reason, with strong influences from Ockham in particular. He also wrote influential works on the nature of the soul. He was one of the most eminent partisans of the late medieval nominalist movement and was numbered among the foremost doctores renovatores *by King Louis XI in his decree against the nominalists. His works continued to be highly influential as late as the Reformation period.*

Pierre d’Ailly, rector of the Collège de Navarre, chancellor of the University of Paris, bishop of Le Puy and Cambrai, cardinal and papal legate to Germany, was a man of wide interests and inexhaustible energy. He wrote a great number of treatises on the most varied subjects, besides performing the many duties associated with his ecclesiastical and secular posts. He left behind over one hundred and seventy works; those devoted to purely philosophical matters are few in number and were all written in the early years of his academic career, while he was teaching philosophy at the Collège de Navarre. The largest group of his works is devoted to matters relating to the Great Schism (during which there were two, or even three, popes) which endured for almost forty years. After the beginning of the Schism in 1378, d’Ailly concentrated on finding a way to terminate this *horrenda monstruosaque divisio* (abhorrent and monstrous division) and was completely distracted from philosophy.

His influences were widespread and lasting. Martin Luther was substantially influenced by d’Ailly’s theory of the Eucharist (developed in the fourth book of his *Commentary on the Sentences*, written in 1376-7). In his own treatise *De captivitate Babylonica ecclesiae (The Babylonian Captivity of the Church)*, Luther recalls:

> Once, during my study of scholastic theology, Pierre d’Ailly gave me occasion to think (while I was reading his fourth book on the Sentences, where he argues most acutely) that it would be much more probable, and one would need fewer of those superfluous miracles, if one would affirm that on the altar there were real bread and real wine, not just their attributes - if the Church had not determined the opposite. When I later realized which Church it was that had determined this - the Thomistic Church, to be sure - I became more courageous.
>
> (*De captivitate Babylonica ecclesiae*)

Christopher Columbus owned d’Ailly’s geographical works *Imago mundi (The Appearance of the World)* and *Epilogus mappae mundi (Epilogue on the Map of the World)*, both written in 1410. They influenced Columbus in his search for a shorter sea route to India and consequent discovery of America; of the authors Columbus had studied, it was d’Ailly whom he preferred to quote. Nicholas of Cusa used d’Ailly’s *Exhortatio super kalendarii correctione (On Corrections to the Calendar)*, written in 1411 for Pope John XIII and later publicly read at the Council of Constance, for his own treatise *De correctione kalendarii*. In his *Exhortatio*, d’Ailly advocated - in vain - the reform of the calendar that was later successfully promulgated by Pope Gregory XIII. Johannes Kepler, himself an apologist for a reformed astrology, explicitly refers to d’Ailly’s astrological works in his treatise *De stella nova in pede Serpentarii (The New Star in the Foot of the Serpent)*; in one of those works, the *Elucidarium astronomicae concordiae cum theologica et historica veritate (On the Concordance of Astronomy with Theology and Historical Truth)*, written in 1414, one finds the first mention of d’Ailly’s famous prediction of the French Revolution, which he repeated in his treatise *De persecutionibus ecclesiae (The Persecutions of the Church)*.

Much of fourteenth-century thought is characterized by a desire to disengage faith from reason and to build upon empirical facts rather than metaphysical assumptions. Thus, it is not by chance that two of d’Ailly’s favourite phrases in his philosophical writings are *docet experientia* (experience teaches) and *patet inductive* (this is clear on the basis of induction). His main sources are William of Ockham, Thomas Bradwardine, Gregory of Rimini and John Buridan, among whom Ockham is clearly the foremost authority: ‘a few things said by him I value more highly than many volumes by certain others’ (*Tractatus de consolatione philosophiae Boethii*, q.1, art.4: 132).

Pierre d’Ailly’s logical writings - *Conceptus (Concepts)*, *Insolubilia (Insolubles)*, *Exponibilia (Exponible Propositions)* and *Destructiones modorum significandi (Attacks on the Modes of Signifying)* (of dubious attribution) - were very influential in the later Middle Ages and through the fifteenth century. D’Ailly adheres to the Ockhamist tradition in basing his logical theories on the notion of mental language. Concerning paradoxes of self-reference (*insolubilia*), for instance, he holds that there are none in mental language, and that spoken or

written paradoxes of this sort are ambiguous sentences in so far as they correspond to two distinct mental sentences, one true and the other false (see Language, medieval theories of §§2, 3, 14). It is worth noting that the term *vitalis immutatio* (vital change) plays an important role in d'Ailly's definitions of central logical terms such as *conceptus* and *significatio*, which are thus closely connected with his theory of the soul and its powers.

His *Tractatus de anima (Treatise on the Soul)*, certainly one of the most important systematic works on philosophical psychology written in the fourteenth century, was widely read throughout the fourteenth, fifteenth and sixteenth centuries. Gabriel Biel wrote a valuable commentary on it. The first chapter deals with the definition of the soul and its tripartite division, chapters 2-8 deal with the powers of the vegetative, sensitive and intellective soul, while the concluding seven chapters deal with the accidents of the soul: species, the acts of sensation, intellection, volition and passion, and habits. D'Ailly's definition of the soul deviates from Aristotle's formulation: 'The soul is the substantial form of a living body that has within it the capacity to carry out vital activities'. This definition of the soul was adopted by other late medieval authors such as Symphorien Champier (see Soul, nature and immortality of).

D'Ailly's *Tractatus super De consolatione philosophiae Boethii (Treatise on Boethius' Consolation of Philosophy)* is mainly devoted to the theory of human happiness. Whether a human being can be called 'truly happy in this life' was one of the central ethical questions in the Buridanist school, where the term *homo felicitabilis* (the human being with a capacity for happiness) was coined. In this context, d'Ailly also discusses the question of immortality. In *Tractatus de anima*, he had followed Buridan in affirming that if one follows human reason alone, then Alexander of Aphrodisias' theory of the soul - that the human soul is 'drawn from matter's potentiality' and hence mortal - is the most probable (see Alexander of Aphrodisias). But in *Tractatus de anima*, he strongly opposes Alexander and, while relying heavily on arguments taken from ancient authorities like Cicero and Seneca, sides with Nicole Oresme in declaring that even 'in accord with the light of nature', the immortality of the soul is more probable.

See also: Nicholas of Cusa; Soul, nature and immortality of the; William of Ockham

OLAF PLUTA

List of works

D'Ailly, Pierre (1350-1420) *Conceptus (Concepts)*, ed. J. Biard, L. Kaczmarek and O. Pluta, forthcoming; trans. P.V. Spade, *Peter of Ailly: Concepts and Insolubles*, Dordrecht: Reidel, 1980.(D'Ailly's writings on logic.)

D'Ailly, Pierre (1350-1420) *Insolubilia (Insolubles)*, ed. J. Biard, F. Del Punta, L. Kaczmarek and O. Pluta, forthcoming; trans. P.V. Spade, *Peter of Ailly: Concepts and Insolubles*, Dordrecht: Reidel, 1980.(D'Ailly's writings on logic.)

D'Ailly, Pierre (1350-1420) *Exponibilia (Exponible Propositions)*, ed. J. Biard, L. Kaczmarek and O. Pluta, forthcoming.(D'Ailly's writings on logic.)

D'Ailly, Pierre (1350-1420) *Destructiones modorum significandi (Attacks on the Modes of Signifying)*, ed. L. Kaczmarek, Amsterdam: Grüner, 1994.(Dubious attribution.)

D'Ailly, Pierre (1350-1420) *Tractatus de anima (Treatise on the Soul)*, in O. Pluta (ed.) *Die philosophische Psychologie des Peter von Ailly*, Amsterdam: Grüner, 1987.(Follows Buridan in supporting the views of Alexander of Aphrodisias.)

D'Ailly, Pierre (1350-1420) *Tractatus super De consolatione philosophiae Boethii*, in M. Chappuis (ed.) *Le Traité de Pierre d'Ailly sur la Consolation de Boéce*, Qu. 1, Amsterdam: Grüner, 1993.(Discusses the theory of human happiness.)

References and further reading

Biard, J. (1992) 'Présence et représentation chez Pierre d'Ailly: Quelques problèmes de théorie de la connaissance au XIVe siècle' (Presence and Representation in the Work of Pierre d'Ailly: Problems of the Theory of Knowledge in the Fourteenth Century), *Dialogue* 31: 459-74.(Expounds and discusses d'Ailly's theory of knowledge as found in his *Commentary on the Sentences* and his philosophical writings.)

Chappuis, M. (1984) 'Notice sur le Traité de Pierre d'Ailly sur la Consolation de Boèce' (Review of Pierre d'Ailly's treatise on the Consolation of Boethius), *Freiburger Zeitschrift für Philosophie und Theologie* 31: 89-107.(Examines all the extant manuscripts of the *Tractatus super De consolatione philosophiae Boethii* and gives a summary of its contents.)

Chappuis, M., Kaczmarek, L. and Pluta, O. (1986) 'Die philosophischen Schriften des Peter von Ailly. Authentizität und Chronologie' (The Philosophical Writings of Pierre d'Ailly: Authenticity and Chronology), *Freiburger Zeitschrift für Philosophie und Theologie* 33: 593-615.(Disputes the authenticity and chronology of d'Ailly's philosophical writings, including the *Destructiones modorum significandi*.)

Kaczmarek, L. (1990) 'Vitalis immutatio. Erkundungen zur erkenntnispsychologischen Terminologie der Spätscholastik' (Lively Changelessness: Explanations of the Epistemological Terminology of Late Scholasticism), in A. Heinekamp, W. Lenzen and M. Schneider (eds) *Festschrift für Heinrich Scheper*, Münster: Nodus, 189-206.(Analyses the use of the term *vitalis immutatio* in John of Ripa's *Commentary on the Sentences* and argues that d'Ailly has adopted this term from him.)

Nuchelmans, G. (1973) *Theories of the Proposition: Ancient and Medieval Conceptions of the Bearers of Truth and Falsity*, Amsterdam: North Holland, 259-65.(Gives an outline of d'Ailly's theory of the proposition in his *Commentary on the Sentences*, *Conceptus*, *Insolubilia* and the (dubious) *Destructiones modorum significandi*.)

Ouy, G. (1975) 'Le Collège de Navarre, berceau de l'humanisme français' (The Collège de Navarre, Cradle of French Humanism), *Actes du 95e Congrès National des Sociétés Savantes de Paris et des Départements*, Sect. de Philologie et d'Histoire, Reims 1970, vol. 1, Paris: Bibliothèque nationale, 275-99.(Shows that the College of Navarre was the cradle of French humanism, its foremost representatives being the 'Navarristes' Pierre d'Ailly and Jean de Montreuil, and later Jean Gerson and Nicolas de Clamanges.)

Pluta, O. (1990) 'Die Diskussion der Frage nach der Unsterblichkeit bei Nikolaus Oresme und Peter von Ailly' (The Discussion of the Question of Immortality in Nicole Oresme and Pierre d'Ailly), *Studia Mediewistyczne* 27 (2): 115-30.(Expounds and discusses the question of immortality as found in Nicole Oresme's *Quaestiones de anima* and d'Ailly's *Tractatus de anima* and *Tractatus super De consolatione philosophiae Boethii*.)

Spade, P.V. (1980) *Peter of Ailly: Concepts and Insolubles, An Annotated Translation*, Dordrecht: Reidel.(The introduction (1-15) presents an introduction to the main topics of *Conceptus* and *Insolubilia* with special emphasis on d'Ailly's sources.)

Ajdukiewicz, Kazimierz (1890-1963)

Ajdukiewicz, like other typical members of the Lwów-Warsaw School, the main Polish analytic movement, was basically interested in logic, philosophy of language, epistemology, and philosophy of science. In the 1930s, he proposed a form of radical conventionalism, an extension of the conventionalism of Duhem and Poincaré. Later, he rejected this radical conventionalism in favour of a semantic epistemology. In the philosophy of science he tried to build a general theory of fallible inferences based on decision theory. Ajdukiewicz's most important contribution to logic is his formal notation for syntactic categories.

1 Life

Ajdukiewicz was born on 12 December 1890 in Tarnopol, a town in the Austro-Hungarian Empire (now in Ukraine). In 1908-12 he studied philosophy at the University of Lwów, mainly under Kazimierz Twardowski. He was also trained by Jan Łukasiewicz in logic and Wacław Sierpiński in mathematics. In 1912 he obtained his Ph.D.; his dissertation concerned Kant's theory of space. In 1913 Ajdukiewicz studied in Göttingen where he attended courses given by David Hilbert and Edmund Husserl. He obtained his *Habilitation* degree from the University of Warsaw with a dissertation on the foundations of mathematics. In 1921-6 he was an associate professor (docent) at the University of Lwów. In 1926 he was appointed as professor of philosophy at the University of Warsaw. In 1928 he returned to Lwów, where he was given a professorship. During the Second World War he lived in Lwów and taught in clandestine Polish schools. In 1945 Ajdukiewicz accepted a professorship at the University of Poznan. In 1955 he moved to the University of Warsaw. He died on 12 April 1963 in Warsaw.

2 Radical conventionalism

Radical conventionalism is closely related to Ajdukiewicz's theory of language and meaning. The meanings of expressions in a language generate rules for accepting sentences of *L*. Ajdukiewicz singles out three kinds of meaning-rules: axiomatic (they require the unconditional acceptance of certain sentences, for example '*A* is *A*'), deductive (for example, *B* follows from 'if *A* then *B*' and *A*), and empirical (the sentence 'snow is white' is asserted in a situation in which a person asserting this sentence perceives that snow is white).

It follows from the foregoing explanations that meanings determine meaning-rules. But in general, meaning-rules do not determine the meanings of expressions; this holds, for example, for ordinary language. However, the situation radically changes when we pass to closed and connected languages. Roughly speaking, a language *L* is open if it can be extended to a new language L' without changes in the meanings of the expressions of *L*; otherwise, *L* is closed. A language *L* is disconnected if there is a non-empty subset *X* of expressions of *L* such that no element of *X* is related by meaning-rules of *L* to its remaining expressions; otherwise, *L* is connected. An important consequence of the theory of closed and connected languages is this: if *L* is a closed and connected language, it is impossible to enrich *L* by new expressions in such a way that old meanings are preserved.

For Ajdukiewicz, mature, particularly scientific, knowledge is expressed in closed and connected languages. The set of meanings of a closed and connected language *L* is its conceptual apparatus. From general theorems on closed and connected languages, one can infer that two conceptual apparatuses are either identical or mutually non-translatable. The acceptance or rejection of sentences is always related to a definite language *L*. If *L* is closed and connected, empirical situations do not force us either to accept or reject any sentence, because we can always change our conceptual apparatus. This is an essential strengthening of usual conventionalism. For Poincaré and Duhem, we are free to change our theoretical principles, because they are hidden conventions. For Ajdukiewicz, experiential reports are also closely related to conceptual apparatus, and since every conceptual apparatus produces a world-perspective, we can say that theories and observational reports are accepted not absolutely but relative to world-perspectives. This is why Ajdukiewicz called his conventionalism 'radical', contrary to the moderate view of the Frenchmen (see Conventionalism).

3 Semantic epistemology

In the middle 1930s Ajdukiewicz rejected radical conventionalism, because he came to the view that his idea of

connected and closed languages was a ‘paper fiction’. The change was also strongly motivated by the work of Tarski which convinced many philosophers that semantics has important applications in philosophy. When he was a radical conventionalist, Ajdukiewicz did not draw any ontological theses from his epistemological considerations; but his semantic epistemology is an attempt to bring together epistemology and ontology. If we speak about the world, we use an object-language. Since epistemology intends to say something about the world and our knowledge of it, an epistemologist must use a meta-language in order to capture knowledge and its object. Ajdukiewicz, employing metalogic and semantics, gave a rigorous analysis of Rickert’s transcendental idealism and Berkeley’s subjective idealism (see Berkeley, G.). For Ajdukiewicz, both kinds of idealism are incorrect, because they neglect basic results of metalogic and semantics. Ajdukiewicz rejects Rickert’s idealism, because truth, contrary to Rickert, cannot be established exclusively by purely deductive procedures; the incompleteness of arithmetic is an essential premise of Ajdukiewicz’s argument. Berkeley’s thesis that ordinary objects are complexes of our ideas is rejected, because it conflates syntax and semantics. According to Ajdukiewicz, Berkeley uses a language which is very similar to the language of syntax and offers a syntactic-like definition of existence. However, since existence is basically a semantic concept and semantics is not fully definable in syntax, Berkeley’s argument fails. Thus, semantic epistemology leads to a realist account of existence.

4 Philosophy of science

In addition to his discussions of radical conventionalism, which implies that there is no absolute gap between theories and experiential reports, Ajdukiewicz also worked on concrete problems in the philosophy of science. In particular, he was interested in the logic of fallible inferences. His approach was based on concepts borrowed from decision theory. In general, acceptances (rejections) of sentences are actions which are associated with profits and losses. Assume that A is a sentence to be accepted and that Z is the minimal acceptable profit for the agent, when A is true, and S is the minimal acceptable loss, when A is false. According to Ajdukiewicz the ratio $S/(S+Z)$ expresses the degree of certainty which an agent accepting A can ascribe to this sentence. This relates degrees of certainty to mathematical probabilities. Having this framework, Ajdukiewicz tried to establish the degree of conclusiveness of a fallible scheme of inference. Assume that K is background knowledge. We are interested in the degree of conclusiveness of a fallible (for example, inductive) inference from premises P to a hypothesis H. This inference is conclusive if the degree of certainty of H does not exceed the ratio of its initial probability to the initial probability of premises, relative to K. Ajdukiewicz’s analysis of fallible inferences is a typical example of his pragmatic approach to methodological problems, which consists in relating analysed concepts to attitudes of epistemic agents.

5 Contributions to logic

Ajdukiewicz’s notation for syntactic categories is his main contribution to logic. The following example shows how this notation works. We have two basic categories: sentences (s) and terms (n). Now, consider an expression ‘is tall’. It forms a sentence together with a term, for example ‘Tom’. We ascribe to ‘is tall’ the symbol s/n which informs us that ‘is tall’ is a function forming a sentence with a term as an argument. Now we build a sequence of symbols for the sentence ‘Tom is tall’. The sequence is this: n, sn. It is obtained by writing the symbols for the categories of all the expressions occurring in the considered sentence. We can simplify the sequence by performing ‘arithmetical’ operations on symbols by analogy with operations on ratios. Thus, we can ‘shorten’ $\{n, s/n\}$ by dividing both members by n; we assume that n/n can be cancelled. Thus, we obtain s as the sole member. A general rule is this: if the simplification ends with n or s, the original expression is syntactically correct; otherwise not. This idea gave rise to constructions known as Leśniewski-Ajdukiewicz-Lambek grammar, originated with Leśniewski, continued by Ajdukiewicz, and fully developed by Lambek (see Syntax).

See also: Poland, philosophy in

JAN WOLEŃSKI

List of works

Ajdukiewicz, K. (1921) *Contributions to the Methodology of Deductive Sciences* (in Polish), Lwów: Polskie Towarzystwo Filozoficzne; partially trans. J. Giedymin (1966), *Studia Logica* 19: 9-46.(Ajdukiewicz’s *Habilitation*.)

Ajdukiewicz, K. (1934a) ‘Sprache und Sinn’, *Erkenntnis* 4: 130-8; trans. J. Wilkinson, ‘Language and Meaning’

in *The Scientific World-Perspective and Other Essays 1931-1963*, Dordrecht: Reidel, 1978: 35-66.(A presentation of the theory of closed and connected languages.)

Ajdukiewicz, K. (1934b) 'Das Weltbild und die Begriffsapparatur', *Erkenntnis* 4: 259-87; trans. J. Wilkinson, 'The World-Picture and the Conceptual Apparatus' in *The Scientific World-Perspective and Other Essays 1931-1963*, Dordrecht: Reidel, 1978: 67-89. (A basic paper on radical conventionalism.)

Ajdukiewicz, K. (1936) 'Die syntaktische Konnexität', *Studia Philosophica* I: 1-276; trans. H. Weber, 'Syntactic Connexion' in *The Scientific World-Perspective and Other Essays 1931-1963,* Dordrecht: Reidel, 1978: 118-39.(Notation for syntactic categories.)

Ajdukiewicz, K. (1948) 'Epistemologia i semiotyka', *Przeglad Filozoficzny* 44: 336-47; trans. J. Giedymin, 'Epistemology and Semiotics' in *The Scientific World-Perspective and Other Essays 1931-1963*, Dordrecht: Reidel, 1978: 182-91.(Semantic epistemology.)

Ajdukiewicz, K. (1958) 'Zagadnienie racjonalnosci zawodnych sposobow wnioskowanis', *Studia Filozoficzne* 4 (7): 14-19; trans. D. Pearce, 'The Problem of Rationality of Non-Deductive Types of Inference' in *The Scientific World-Perspective and Other Essays 1931-1963*, Dordrecht: Reidel, 1978: 239-53.(Ajdukiewicz's analysis of fallible modes of inference.)

Ajdukiewicz, K. (1973) *Problems and Theories of Philosophy*, trans. H. Skolimowski and A. Quinton, Cambridge: Cambridge University Press.(An elementary introduction to philosophy; this book was originally published in Polish in 1949.)

Ajdukiewicz, K. (1974) *Pragmatic Logic*, trans. O. Wojtasiewicz, Dordrecht: Reidel.(An extensive treatise on methodology; this book was originally published in Polish in 1965.)

Ajdukiewicz, K. (1978) *The Scientific World-Perspective and Other Essays 1931-1963*, Dordrecht: Reidel.(A collection of basic papers by Ajdukiewicz.)

References and further reading

Sinisi, V. and Woleński, J. (eds) (1995) *The Heritage of Kazimierz Ajdukiewicz*, Amsterdam: Rodopi. (A collection of essays on Ajdukiewicz's philosophy; includes the complete bibliography of Ajdukiewicz's writings and a selected bibliography on Ajdukiewicz.)

Woleński, J. (1989) *Logic and Philosophy in the Lvov-Warsaw School*, Dordrecht: Kluwer.(Chapter X gives a presentation of Ajdukiewicz's epistemology.)

Akan philosophical psychology

The word Akan *refers to the Twi-speaking people of southern and central Ghana. Akan traditional philosophy is essentially a philosophy of the person. It has cosmological ramifications, but the basic concepts emerge from the analysis of the human personality. That analysis is extremely sensitive to the complexity of the human psyche and the social dimensions of individual consciousness. These considerations explain and justify the prominent position occupied by the concept of a person in contemporary Akan philosophy.*

This emphasis on the person is a time-honoured feature of the written tradition of Akan philosophy and reflects the priorities of the oral tradition, the mainspring of that system of thought. With respect to the nature of personhood, however, there has been a divergence of preoccupation between the oral and the written traditions. The predominant motivation of the written tradition has been metaphysical, as writers have sought to clarify the ontological status of the constituents of human personality, while the main orientation of the oral tradition is social and ethical. Within this tradition, a person is defined by their social relations more crucially than by their ontological essence.

In the Akan tradition the concept of a person has a normative as well as a descriptive component. The word *onipa*, the Akan equivalent of 'person', is often used normatively to designate a person who has largely attained a desirable social standing. To achieve a high degree of personhood is therefore an inherent part of the Akan ethic. To be a person in the fullest sense is to be an adult who works hard, thinks judiciously and is able to support a conjugal household as well as fulfil a range of obligations to an extended group of kinfolk and to the civic community at large. Such an individual, also known as an *obadwenma*, meaning someone of ethical and cerebral maturity, must also listen to and act in accordance with their conscience, known as *tiboa*, or literally, animal in the head (*ti* meaning head and (*a*)*boa* meaning animal).

In Akan circles to talk or act unintelligently is to risk unflattering descriptions of the *adwene* in your head. *Adwene*, or mind, is the noun form of the verb *dwen*, to think. The term denotes mental processes and emotional dispositions. The Akans delight in metaphor and often locate the mind in the head as if it is identical with the brain, or *amene*. Indeed, there is little temptation to identify the mind with the brain or with any other kind of substance, physical, abstract or spiritual.

Because the Akans traditionally do not conceive of the mind as a kind of substance, they do not include it in their ontological inventory of the elements which constitute an *onipa*, or person, in the minimal sense of a human being. Within the oral and written traditions there is near unanimity that a person in Akan traditional thought consists of *ornipadua*, literally the tree of a person, deriving from the *mogya*, or blood of the mother in combination with the *okra*, the animating element which emanates directly from the supreme being. The factor that completes the equation that yields a person is the *sunsum* which originates from the father's input in conception and is held to be responsible for the unique personal presence of every individual. All *akra* (plural of okra) are identical in nature in as much as they are all particles of the divine substance, but they are individuated by the unique destinies prenatally assigned to them by the supreme being. In view of this common divine element, all human beings whatever their social achievements or failures, are regarded as children of God each with the gift of immortality. (A famous Akan adage holds that everyone is the child of God; no one is a child of the earth.)

The *okra* and the *sunsum* cannot be identified with the soul or spirit contrary to frequent practice in Akan literature, as in the general cosmology of the Akans, all these entities are conceived in quasi-material terms. However, this and other issues in the interpretation and evaluation of the Akan concept of a person, remain matters of controversy among Akan philosophers.

See also: Persons

KWASI WIREDU

References and further reading

Abraham, W.E. (1962) *The Mind of Africa*, Chicago, IL: University of Chicago Press. (Chapter 2, especially the section on the Akan 'Theory of Man and Society'.)

Appiah, K.A. (1982) *In My Father's House: Africa in the Philosophy of Culture*, London: Methuen; 2nd edn, Oxford: Oxford University Press, 1992.(Chapter 5: 'Ethnophilosophy and its Critics', especially pages 96-100.)

Busia, K.A. (1954) 'The Ashanti of the Gold Coast', in D. Forde (ed.) *African Worlds: Studies in the Cosmological Ideas and Social Values of African Peoples*, Oxford: Oxford University Press.

Danquah, J.B. (1944) *The Akan Doctrine of God: A Fragment of Gold Coast Ethics and Religion*, London: Frank Cass, 1968.(Second edition with introduction by K.A. Dickson, 1968. See §§4 and 5.)

Engmann, J. (1992) 'Immortality and the Nature of Man in Ga Thought', in K. Wiredu and K. Gyekye (eds) *Person and Community: Ghanaian Philosophical Studies*, Washington, DC: Council for Research in Values and Philosophy.(An interesting read.)

Gyekye, K. (1987) *An Essay on African Philosophical Thought: The Akan Conceptual Scheme*, New York: Cambridge University Press.(Chapter 6: 'The Concept of a Person' and pages 163-8, 198-9 contain references to the work of scholars such as Rattray and Meyerowitz on the Akan conception of a person.)

Oguah, B.E. (1977) 'African and Western Philosophy: A Comparative Study', in R.A. Wright (ed.) *African Philosophy: An Introduction*, New York: University Press of America.(The third edition of this book appeared in 1984.)

Opoku, K.A. (1978) *West African Traditional Religion*, London: FEP International Private Limited.(Chapter 5: 'West African Conception of Man'.)

Parrinder, E.G. (1951) *West African Psychology: A Comparative Study of Psychological and Religious Thought*, London: Lutterworth Press.(Especially interesting are chapters 1-6, pages 1-84.)

Wiredu, K. (1987) 'The Concept of Mind with Particular Reference to the Language and Thought of the Akans', in G. Flöstad (ed.) *Contemporary Philosophy*, vol. 5: *African Philosophy*, Boston, MA: Kluwer, 1987.

Akrasia

The Greek word 'akrasia' *is usually said to translate literally as 'lack of self-control', but it has come to be used as a general term for the phenomenon known as weakness of will, or incontinence, the disposition to act contrary to one's own considered judgment about what it is best to do. Since one variety of* akrasia *is the inability to act as one thinks right,* akrasia *is obviously important to the moral philosopher, but it is also frequently discussed in the context of philosophy of action.* Akrasia *is of interest to philosophers of action because although it seems clear that it does occur - that people often do act in ways which they believe to be contrary to their own best interests, moral principles or long-term goals - it also seems to follow from certain apparently plausible views about intentional action that* akrasia *is simply not possible. A famous version of the suggestion that genuine* akrasia *cannot exist is found in Socrates, as portrayed by Plato in the* Protagoras*. Socrates argues that it is impossible for a person's knowledge of what is best to be overcome by such things as the desire for pleasure - that one cannot choose a course of action which one knows full well to be less good than some alternative known to be available. Anyone who chooses to do something which is in fact worse than something they know they could have done instead, must, according to Socrates, have wrongly judged the relative values of the actions.*

1 The Socratic view

The Socratic view that 'no one does wrong willingly' receives its most detailed elaboration in the *Protagoras*. The context is a discussion between Socrates and Protagoras about the nature of the virtues. It is in the course of defending the suggestion that the virtues form a kind of unity that Socrates maintains that no one can knowingly choose the worse of two available alternatives.

There is much controversy about how, exactly, Socrates' argument for his conclusion is to be understood. The argument certainly appears to invoke some highly questionable assumptions about the relationship between pleasure and pain and goodness and evil, and between all of these and human motivation; and there has been much debate about whether we ought to regard these assumptions as ones which Socrates himself accepted, or whether he is rather arguing *ad hominem* against those who do accept the hedonistic views on which his argument seems to be based. Protagoras and Socrates declare themselves straightforwardly agreed on the point that knowledge cannot be 'pushed around by all the other affections' (352c1) (that is, that if someone knows which is the best course of action, then nothing can persuade them to act otherwise than as that knowledge dictates). The argument is thus presented as an attempt to convince not Protagoras, but rather those ordinary people who have not realised that there is really no such phenomenon as being induced to act contrary to one's knowledge by the desire for pleasure. This may leave room, therefore, for the view that the premises on which Socrates' argument are based are not ones he himself believed, but rather are assumptions that he attributes to these hypothetical interlocutors.

With this caution in mind, we can say that the structure of the argument is something like this:

(1) The evaluation of actions is always based, ultimately, on judgments about the total amounts of pleasure and pain which will result, overall, from performing them.
(2) Since this is so, the good and the pleasant are (in some sense or other) the same, and similarly for the bad and the painful.
(3) This means that any description of an agent which involves their being said to be 'overcome by pleasure' can be replaced by a description of that agent as 'overcome by the good'. But the good by which the agent is overcome cannot, by hypothesis, outweigh the evil which will result from the action (otherwise the action would be virtuous and the agent would have done nothing wrong).
(4) The agent who is 'overcome by pleasure', therefore, must be described as someone who chooses evil (pain) in exchange for a good (pleasure) which does not adequately compensate for this evil.
(5) But, in weighing pleasures and pains against one another, one must always choose what one believes to be the greater pleasure or the lesser pain - and one must therefore take whichever course of action it is that one believes will result, overall, in the most pleasant or least painful outcome, all things considered.
(6) So, the only possible explanation for someone's choosing an action which will bring more pain, overall, or less pleasure, than some alternative known to be available is that they have misjudged the total amounts of pleasures and pains to be got from the alternative actions.

Obviously, this argument is vulnerable to attack at many points. In particular, the psychological hedonism invoked in premise (5) might be challenged. There may be other reasons for thinking that there is something right about the Socratic view, but it seems fair to say that what we are offered in the *Protagoras* amounts to a less than conclusive case.

2 Aristotle on *akrasia*

Aristotle's views on *akrasia* receive their most extensive airing in Book VII of his *Nicomachean Ethics*, and the bulk of what he says is concentrated in a single chapter of that book, chapter 3. The text of this chapter, however, is dense and complex, and there is considerable disagreement amongst commentators even about the very basic question of whether Aristotle is to be seen as an apologist for the Socratic position or rather as a critic of its over-intellectualized view of the causes of practical error. Some translators (for example Ross 1980) have credited Aristotle with the straightforward assertion, early in his discussion at VII 2, that the Socratic view 'contradicts the plain phenomena' (1145b27), which clearly supports the second hypothesis; but since the Greek word 'phainomena' can also be translated 'common opinions or beliefs', this interpretation is not uncontroversial. And it must be admitted that a good deal of what Aristotle says in VII 3 appears to associate the occurrence of *akrasia* with a special kind of ignorance, an association reminiscent of the Socratic view that if one really knows what is best, one can do nothing else. Furthermore, the chapter ends with the claim that, in the light of what has been said, 'the position that Socrates sought to establish actually seems to result' (1147b14).

On the other hand, there are numerous passages in other of Aristotle's works where Aristotle appears to imply that it is perfectly possible for an agent to pursue an undesirable course of action while knowing full well that it is not the best thing to do. In the *Eudemian Ethics*, for example, it is said that to act incontinently 'is to act through appetite contrary to what the man thinks best' (1223b8-9), and that the *akrates* has 'a pain of expectation, thinking that he is doing ill' (1224b20-21). Then again, in *Magna Moralia*, we have the incontinent agent characterised as one 'who knows indeed from reason that he ought not [to do the wrong thing], but gives in to pleasure and succumbs to it' (1203b5-6). These quotations seem to suggest that Aristotle is perfectly content with the description of *akrasia* that Socrates sought to reject, the condition of one whose knowledge is vanquished, but not necessarily clouded or annihilated, by the desire for pleasure. How, then, are we to square these passages with the appearance that NE VII 3 endorses some version of the Socratic view?

There is not space here to review the many attempts that have been made to resolve the apparent contradiction. Certainly, though, VII 3 does make it seem as though Aristotle is inclined to deny that there can be such a thing as utterly clear-eyed *akrasia* - the calm, deliberate and intentional performance of an action known not to be in one's own best interests. The chapter is largely concerned with the application to the problem of *akrasia* of two distinctions, one rooted in Aristotle's doctrine of the practical syllogism, the other concerned with a contrast between the mere possession of knowledge and its *use* or *exercise*, and it is hard to see how these epistemic considerations are intended if not as qualified support for the Socratic view that knowledge cannot be 'dragged around like a slave'. The suggestion seems to be that passion or desire makes it impossible for the *akrates* to reach full-fledged knowledge of the conclusion of the practical syllogism which would lead him to perform the right action. But if so, Aristotle would seem to be in agreement with Socrates that weakness of will is never really just that - that it is always bound up with some form of self-deception, delusion or other epistemic error (see Aristotle §23).

3 Davidson on weakness of will

In a famous article entitled 'How is Weakness of the Will Possible?' (1970), Donald Davidson locates the problem of *akrasia* firmly within the philosophy of action, discarding its traditional connections with morality and the defeat of moral judgment by passions of various kinds. According to Davidson, an agent acts incontinently if and only if: (a) the agent does *x* intentionally; (b) the agent believes there is an alternative action *y* open to them; and (c) the agent judges that, all things considered, it would be better to do *y* than to do *x*. It follows that, for Davidson, there is not necessarily anything iniquitous about incontinence. He gives the following example. Suppose that I suddenly remember, having already gone to bed, that I have not brushed my teeth. It is clear to me that missing one night's brushing won't make much difference to my dental health and that getting up may result in my having a disturbed and fitful night's sleep, so I conclude that, all things considered, it would be better for me to stay in bed. Nevertheless, I reluctantly get up and plod to the bathroom. On Davidson's definition, my action counts as

incontinent - though it is more plausible to say here that pleasure has been worsted by duty rather than the other way round.

The problem of incontinence is represented by Davidson as an apparent inconsistency between three principles, all of which, according to Davidson, seem 'self-evident', or at any rate can be made so, given suitable interpretation. The three principles (from Davidson [1970] 1980: 23) are the following:

'P1: If an agent wants to do *x* more than they want to do *y* and they believe themselves free to do either *x* or *y*, then they will intentionally do *x* if they do either *x* or *y* intentionally.
P2: If an agent judges that it would be better to do *x* than to do *y*, then they want to do *x* more than they want to do *y*.
P3: There are incontinent actions.'

Davidson rejects what he calls 'the most common way of dealing with the problem of incontinence', that is, the abandonment of P2. Though he admits that 'wanting' and 'judging better' can readily be interpreted so as to render P2 false, he insists that there is also a natural reading that makes it true, which leaves the apparent inconsistency intact. 'Judging better', he suggests, cannot be totally divorced both from behaviour and from desire - if one sincerely believes that *x* is a better course of action, all things considered, than *y*, then one must, in some sense or other, want to do *x* more than one wants to do *y*. And provided one concedes this, the problem about weakness of will will remain.

Davidson's solution to the problem is an attempt to show that P1-P3 are not, after all, inconsistent. The argument turns on a distinction made by Davidson between unconditional evaluative judgments (for example, the straightforward judgment that it is better to do *x* than to do *y*) and conditional, or *prima facie*, evaluative judgments, which are relative to some body of evidence or other (such as the judgment that it is better to do *x* than to do *y*, given all the relevant factors known to the agent). Davidson offers a treatment of the latter variety of judgment from which it follows that no conditional evaluative judgment can be in any logical conflict with an unconditional one. This enables him to reconcile P1-P3 by saying that the incontinent agent who does *x* rather than *y* does indeed want to do *x* more than *y* and makes an unconditional judgment that the incontinent course of action, *x*, is better than *y*. The awareness that *y* is better than *x*, on the other hand, is relative to a body of reasons - the total body of reasons available to the agent - and therefore does not contradict the unconditional judgment in accordance with which the agent acts (see Intention §3).

4 Moral weakness

Moral weakness is that particular form of *akrasia* which consists in failing to live up to one's sincerely expressed beliefs about what it would be morally best to do. Nothing is more obvious, one might think, than that people often do things they genuinely consider to be morally wrong; but the existence of this type of weakness has sometimes been thought to present problems for certain moral theories. In particular, any theory which makes action the test or criterion of the sincerity of moral beliefs may have difficulty accounting for moral weakness. For such theories will tend to suggest that those who do not act on the principles they profess cannot really be said to hold those principles at all - thus their failing will turn out to be not *akrasia*, but hypocrisy.

R.M. Hare's ethical prescriptivism is one theory in particular which has been thought to conflict with the existence of moral weakness (see Prescriptivism). Hare's view (1963) is that it is part of the meaning of moral judgments that they *prescribe*; that is, they are intended as guides to conduct. The acceptance of some particular moral judgment, then, is the acceptance of a recommendation to act in some particular way in some given circumstance or type of circumstance - but how could one be said sincerely to have accepted such a recommendation if one fails to act in accordance with it? Hare's response, roughly, is to insist that typically, cases of 'moral weakness' are cases where the agents concerned are psychologically incapable of doing what they think they ought. Hare's view is that 'I ought but I can't' is not a contradiction; the agent here is 'prescribing in general terms, but exempting himself because of the impossibility, in his case, of obeying this general prescription' (1963: 53). One might wonder though, whether this is really a satisfactory account of *all* cases of moral weakness. Sometimes, no doubt, moral agents are prevented from acting as they think they ought by compulsions and irresistible forces, but it seems wrong to think of this as the general, or even the typical case.

See also: Moral agents; Moral psychology; Rationality: practical; Self-deception; Will, the

HELEN STEWARD

References and further reading

Aristotle (384-322 BC) *Eudemian Ethics*, trans. J. Solomon, in *The Complete Works of Aristotle*, vol. 2, Oxford: Oxford University Press, 1984.(Probably an earlier work than the *Nicomachean Ethics*, with which it has three books in common.)

Aristotle (384-322 BC) *Nicomachean Ethics*, trans. and notes T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985; or trans. Sir David Ross, Oxford: Oxford University Press, 1980.(The main discussion of *akrasia* is contained in book VII, chapters 2- 4, and is the topic of §2 above.)

Aristotle (384-322 BC) *Magna Moralia*, trans. St. G. Stock, in *The Complete Works of Aristotle*, vol. 2, Oxford: Oxford University Press, 1984.(No agreed view on date of composition, which may well be a handbook written after Aristotle's time.)

Davidson, D. (1970) 'How is Weakness of the Will Possible?', in J. Feinberg (ed.) *Moral Concepts*, Oxford: Oxford University Press, repr. in D. Davidson *Essays on Actions and Events*, Oxford: Clarendon Press, 1980. (A classic article, a little technical towards the end; discussed in §3 above.)

Gosling, J. (1990) *Weakness of the Will*, London: Routledge.(A useful overview, which contains a good bibliography.)

Hare, R.M. (1963) *Freedom and Reason*, Oxford: Oxford University Press.(Chapter 5, 'Backsliding', contains Hare's discussion of moral weakness.)

Lemmon, E.J. (1962) 'Moral Dilemmas', *Philosophical Review*, 71: 139-58. (Contains a stalwart defence of the existence of moral weakness.)

Plato (*c*.386-380 BC) *Protagoras*, trans. and notes C.C.W. Taylor, Oxford: Clarendon, 2nd edn, 1991.(Contains text of the dialogue together with a helpful and detailed commentary. This is discussed in §1 above.)

Santas, G. (1966) 'Plato's Protagoras and Explanations of Weakness', *The Philosophical Review* 75: 3-33.(Good both on the *Protagoras* in particular and on incontinence in general.)

al-Afghani, Jamal al-Din (1838-97)

Al-Afghani is often described as one of the most prominent Islamic political leaders and philosophers of the nineteenth century. He was concerned with the subjection of the Muslim world by Western colonial powers, and he made the liberation, independence and unity of the Islamic world one of the major aims of his life. He provided a theoretical explanation for the relative decline of the Islamic world, and a philosophical theory of history which sought to establish a form of modernism appropriate to Islam.

1 Life and times

Jamal al-Din al-Afghani was born in 1838 about 180 miles from Kabul, of a distinguished family. He received a thorough training in a variety of languages of Islamic countries and the religious sciences. When he was eighteen years old he began the constant travels which were to mark his life. He visited much of the Islamic world as well as Europe, and set up a political organization which called on Muslims to fight injustice and the imposition of imperialism. He had a great impact upon Muhammad ʻAbduh and reactions by intellectual Egyptians to the incursion of the Europeans. He eventually sided with the Ottoman empire but soon became disillusioned with the Sultan, and died in suspicious circumstances in Turkey in 1897.

2 Philosophy of history

Al-Afghani's philosophical contributions are to be found in his book *ar-Radd ʻalal-dahriyyin (Refutation of the Materialists)*. Citing philosophers such as Democritus and Darwin, he criticized the naturalist and materialist philosophers for their denial, either directly or indirectly, of the existence of God. He then went on to elaborate at great length on religion's contribution to civilization and progress. According to al-Afghani, religion has taught humanity three fundamental beliefs: its angelical or spiritual nature, the belief of every religious community in its superiority over other groups, and the assertion that our existence in this world is but a prelude to a higher life in a world entirely free from sorrow and suffering. Our angelic nature urges us to rise above our bestial proclivities and live in peace with our fellow human beings. The feeling of competitive superiority on the part of the various religious groups generates competitiveness, whereby the various communities will strive to improve their lot and persist in their quest for knowledge and progress. Finally, the third truth provides an incentive to be constantly aware of the higher and eternal world that awaits us. This in turn will motivate human beings to refrain from the evil and malice to which they may be tempted, and live a life of love, peace and justice.

Al-Afghani mentions that religion implants in its believers the three traits of honesty, modesty and truthfulness. He further maintains that the greatness of the major nations of the world has always been entailed by their cultivation of these traits. Through these virtues the Greeks were able to confront and destroy the Persian empire. However, when the Greeks adopted the materialism and hedonism of Epicurus, the result was decay and subjection by the Romans. Likewise the ancient Persians, a very noble people, began with the rise of Mazdaism the same downward journey as the Greeks, which resulted in their moral erosion and subjection by the Arabs. Similarly, the Muslim empire, which rose on the same solid moral and religious foundation as did both the Greeks and Persians, became so weakened that a small band of Franks (that is, the crusaders), was able to score significant victories against them. Subsequently, the hordes of Genghis Khan were able to trample the whole land of Islam, sack its cities and massacre its people.

Al-Afghani bases his philosophy on a theory of history in which religion is portrayed as a catalytic force in the progress of humanity. Interestingly, he stresses that religious beliefs must be founded upon sound demonstration and valid proof without any supernatural aspect. This rationalism manifests an important element of modernity in al-Afghani's thinking. However, such modernity does not diminish his strong belief in religion as an integral component and fundamental force behind humanity's quest for morality, truthfulness and integrity.

Al-Afghani's philosophical views revealed a great deal of faith in the human mind and its capacity for innovations based on knowledge rather than ignorance. He expressed great faith in humanity as being one of the greatest miracles of the universe, and believed that there are no areas which can remain forever closed to the human mind. Surprisingly, he predicted that people would reach the moon as a step in a series of strides by mankind, as he believed that nature and the universe were created so that we could continue the challenge of unravelling their

secrets.

In his criticism of Darwin's theory of evolution (see Evolution, theory of), al-Afghani presents a philosophical theory about nature in response to Darwin's theory. He believes in the nature of what he termed 'natural selection', whereby survival in nature will be for the strongest and the fittest. Thus if a number of plants are planted in a single space of earth which does not have food for all these plants, it will be noticed that the plants will compete among themselves for food. In due course, some of the plants will become more developed than the others, which will wither. He applies the same theory to the world of animals, including human beings, where the influence of power is more noticeable than elsewhere. He even goes further than Darwin by applying the theory to the area of ideas, maintaining that ideas are born out of other ideas and may be greater than those ideas; this explains why posterity may sometimes excel and be superior to its ancestry. Al-Afghani believes that these developments are due to the impact of nature's aspects and not necessarily the result of human effort. His criticism of Darwin's theory lessened gradually as he began to express views similar to those of Darwin. He cites earlier Muslim scholars such as Ibn Bashroun who had talked about the evolution from dust of plants and animals. Al-Afghani, however, continued to maintain strong disagreement with Darwin on one fundamental issue, that of the creation of life; this al-Afghani unequivocally ascribes to God.

See also: 'Abduh, M.; Darwin, C.R.; Evolution, theory of; Islamic philosophy, modern

ELSAYED M.H. OMRAN
OLIVER LEAMAN

List of works

al-Afghani (1838-97) *ar-Radd 'alal-dahriyyin (Refutation of the Materialists)*, Cairo, 1955.(The main philosophical contribution of al-Afghani.)

References and further reading

Keddie, N. (1968) *An Islamic Response to Imperialism: Political and Religious Writings of Sayyid Jamal al-Din al-Afghani*, Berkeley, CA: University of California Press.(An useful series of essays linking al-Afghani's philosophical and political views.)

Keddie, N. (1972) *Sayyid Jamal ad-Din al-Afghani: A Political Biography*, Berkeley, CA: University of California Press.(An important study of al-Afghani's politics, with useful material on his general philosophical views.)

Kedourie, E. (1966) *Afghani and 'Abduh*, London: Cass.(Deals extensively with Afghani's political philosophy and its influence on 'Abduh.)

al-‘Amiri, Abu’l Hasan Muhammad ibn Yusuf (d. 992)

Although al-‘Amiri had only a limited long-term impact, his extant works provide useful insights into an extremely creative period in Islamic philosophy in the tenth century AD. He attempted to reconcile philosophy with religion by showing that the genuine conclusions of philosophy could not contradict the revealed truths of Islam, and attempted to build consensus within Islam. He argued for the individual immortality and the punishment or reward of the soul. His analysis of the soul is largely Neoplatonic. The reward of the afterlife is determined by the actualization of the intellect in this life, aided primarily by right actions which moderate the physical faculties and turn the intellect toward the Divine.

Abu’l Hasan Muhammad ibn Yusuf al-‘Amiri was born in Khurasan (in modern Iran) in the early fourth century AH (tenth century AD) and died in Nishapur in AH 381/AD 992. He began his career in Khurasan, where he studied under Abu Zayd al-Balkhi, and moved to Rayy and Baghdad, where he met and was discussed by such substantial intellectuals as al-Tawhidi and Ibn Miskawayh. He ended his career in Bukhara, where he had access to the Samani library (in which Ibn Sina studied shortly thereafter), and in Nishapur.

Al-‘Amiri’s main concern was the rational defence of Islam against a form of philosophy regarded as independent of revelation, and against competing religious traditions. In the tradition of al-Kindi, he attempted to reconcile philosophy with religion by showing that the real conclusions of philosophy could not contradict the revealed truths of Islam. Unlike his contemporary al-Farabi, however, al-‘Amiri argued that revealed truth must be superior to philosophy, since revelation was necessary for the completion of the human intellect and as the indubitable guide to right action. The Greeks possessed useful wisdom, but they could not be considered final authorities because they lacked a prophet.

In spite of his attacks on, for example, the Mu‘tazila and the Batiniya esotericists, al-‘Amiri’s approach was generally conciliatory toward philosophy, the *mutakallimun* (theologians) and a wide variety of Islamic sects. His respected treatise on Sufism, for example, provided both a rational, Aristotelian interpretation of Sufism and a reconciliation of Sufism with more conventional Islam. He preferred to emphasize areas of agreement between philosophers and Islamic sects, perhaps because he perceived the dangers of sectarianism in the diverse environment of Khurasan and perhaps also because Islam had not fully consolidated its position relative to pre-Islamic traditions. He had a marked preference for religious, rather than philosophical, terminology (for example, *ruh* rather than *nafs* for the soul), indicating that his probable primary audience was the Islamic religious elite.

In *al-I‘lam bi manaqib al-Islam (An Exposition on the Merits of Islam)* and *Inqadh al-bashar min al jahr wa’l-qadar (Deliverance of Mankind from the Problem of Predestination and Free Will)*, al-‘Amiri attempted a rational justification of the moral superiority of Islam to other religions, especially to Zoroastrianism and Manicheism. In the latter work, he also attempted a resolution of the theological problem of free will by the application of Aristotelian principles, a project which he repeated with greater philosophical subtlety in his *al-Taqrir li-awjuh al-taqdir (The Determination of the Various Aspects of Predestination)*.

His resolution of the problem of predestination required a distinction between necessary, contingent and possible beings. Only God is necessary existence (*wajib al-wujud*), whose essence is identical with his existence. Human use of multiple terms for divine attributes is thus figurative, since God is essentially a unity. All other existents are contingent and, in so far as they require the support of necessary existence, are preordained. The relations of contingent things to each other, however, are of a different order, in which individual responsibility is possible. Significantly, al-‘Amiri’s use of the term *wajib al-wujud* is one of only two extant examples (the other is Ibn Miskawayh) of the use of this term prior to Ibn Sina, who adopted the term into the very heart of his thought. Al-‘Amiri’s interpretation of Empedocles suggests the possible existence of a pseudo-Empedoclean text or tradition extant in his time which might have been a significant precursor of some important Avicennan arguments.

Al-‘Amiri’s list of the five ‘sages’ of Greek philosophy is unusual, since Empedocles is first in a line which progresses through Pythagoras, Socrates, Plato and Aristotle. In keeping with al-‘Amiri’s conciliatory method, each was given a means of contact with a prophetic tradition, even though each spoke from the perspective of

reason alone. Empedocles was said to have studied with Luqman in Syria, and Pythagoras with the companions of Solomon in Egypt. Socrates, Plato and Aristotle then preserved and developed the wisdom of Pythagoras. Al-‘Amiri’s sources for philosophical history are primarily Neoplatonic, especially pseudo-Ammonius (see Neoplatonism).

One or more fragmented translations of the *Phaedo* were especially important for al-‘Amiri’s*Kitab al-amad ‘ala’l-abad (On the Afterlife)*, in which he argued for the individual immortality and punishment or reward of the soul. His analysis of the soul is largely Neoplatonic, and the reward of the afterlife is determined by the actualization of the intellect in this life, aided primarily by right actions which moderate the physical faculties and turn the intellect toward the Divine.

Al-‘Amiri’s work was soon eclipsed by the philosophical revolution brought about by Ibn Sina. Nevertheless, his work provides a window into the philosophical and religious debates which formed the background of that revolution and into the sources upon which the participants in those debates drew. Although partisans of various schools may find his interpretations problematic, his emphasis on the importance of good action over particulars of doctrine and his synthesizing interpretations represent an important attempt to build consensus within Islam during a turbulent, fractious, creative period in its history.

See also: Ibn Sina; Islamic theology; Neoplatonism in Islamic philosophy; Predestination; Soul in Islamic philosophy

TOM GASKILL

List of works

Al-‘Amiri (before 992) *al-I‘lam bi manaqib al-Islam (An Exposition on the Merits of Islam)*. (A translation of most of Chapter 1, ‘The Quiddity of Knowledge and the Appurtenances of its Species’, can be found in F. Rosenthal, *The Classical Heritage of Islam*, Berkeley, CA: University of California Press, 1973, 63-70. A translation by F. Rosenthal of Chapter 7, ‘The Excellences of Islam in Relation to Royal Authority’, appears in ‘State and Religion According to Abu l-Hasan al-‘Amiri’, *Islamic Quarterly* 3: 42-52.)

Al-‘Amiri (before 992) *Inqadh al-bashar min al jahr wa’l-qadar (Deliverance of Mankind from the Problem of Predestination and Free Will)*.(There is at present no modern edition of this work.)

Al-‘Amiri (before 992) *al-Taqrir li-awjuh al-taqdir (The Determination of the Various Aspects of Predestination)*. (There is at present no modern edition of this work.)

Al-‘Amiri (before 992) *Kitab al-amad ‘ala’l-abad (On the Afterlife)*, ed. and trans. E.K. Rowson, *A Muslim Philosopher on the Soul and Its Fate: Al-‘Amiri’s Kitab al Amad ‘ala l-abad*, New Haven, CT: American Oriental Society, 1988.(Rowson contains a critical edition and translation with a commentary on al-‘Amiri’s most influential work, with thorough background and bibliographic material.)

References and further reading

Biesterfeldt, H.H. (1977) ‘Abu’l-Hasan al-‘Amiri und die Wissenschaften’ (Abu’l Hasan al-‘Amiri and the Sciences), *Zeitschrift der Deutschen Morgenländischen Gesellschaft*, Supplement III (1), Wiesbaden.(A useful discussion of the division of the sciences in the *I‘lam*.)

Rowson, E.K. (1996) ‘Al-‘Amiri’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 14, 216-21.(Concise evaluation of the career of this influential thinker.)

Vadet, J.C. (1974-5) ‘Une Défense philosophique de la sunna: les *Manaqib al-islam* d’al-‘Amiri’, *Revue des études islamiques* 42: 245-76 and 43: 77-96.(An overview of the *I‘lam*.)

al-Baghdadi, Abu ’l-Barakat (*fl. c.*1200-50)

A maverick philosopher, respected medical authority, and seemingly somewhat tempestuous individual, Abu ‘l-Barakat al-Baghdadi produced one voluminous work (the Kitab al-mu‘tabar) in which the philosophical views current in his day - principally associated with the name of Ibn Sina - were subjected to a penetrating analysis, and many interesting alternatives suggested. His most provocative ideas concern self-awareness, the physics of motion and the idea of time.

Hibat Allah ‘Ali ibn Malka Abu ’l-Barakat al-Baghdadi was an idiosyncratic, highly original philosopher who flourished in the first half of the twelfth century. Precise biographical information is unavailable. We know that he was born into a Jewish family (his Hebrew name was Nathanel) and, as a Jew, was refused entry to the lectures of Abu ’l-Hasan ibn Hibat Allah, a famous physician. Other slights, real or imagined, that he suffered on account of his faith seem to have contributed to his decision, very late in life, to convert to Islam. The appellation *awhad al-zaman*, ‘the singular [personage] of his time’, probably reflects his medical rather than philosophical achievements. His formal teaching seems to have been limited to medicine, in which he had a number of students. Ibn Khalliqan’s biographical dictionary describes him as ‘very presumptuous’, his hauteur being revealed in his many disputes with the humbler physician Ibn al-Talmidh, which display the ‘jealousy and rivalry that typically prevail between men who are eminent in the same profession’. His involvement in philosophy seems to have been informal (even by the standards of the time) and tentative. He had one notable disciple in the polymath Fakhr al-Din Al-Razi.

Al-Baghdadi’s mature views are found in his comprehensive *Kitab al-mu‘tabar*, the title of which should be translated, according to Shlomo Pines (1979), ‘the book of what has been established by personal reflection’. Al-Baghdadi characteristically begins his investigations with a pellucid statement of each problem he considers. He then surveys earlier opinions in detail, rarely naming his sources. He takes great pains to ferret out the reasoning underlying each claim, following closely the development of the theory, the objections raised against it and the adjustments made by its proponents. Each issue is approached freely and independently, without much reliance on any over-arching methodological or philosophical commitments. The interplay between words and concepts is given particular attention. For example, al-Baghdadi developed his strikingly innovative theory of time after reaching the conclusion that the word ‘time’ as used in everyday speech stands for a very fundamental concept, the true nature of which has been obscured by scholastic analysis. Again, he lambasts aspects of the sulphur-mercury theory of metals as ‘words that denote unreal fancies’.

Perhaps most interesting among al-Baghdadi’s achievements is his reappraisal of the idea of time (Pines 1979). Dissatisfied with the regnant approach, which treated time as an accident of the cosmos, al-Baghdadi drew the conclusion that time is an entity whose conception (*ma‘qul al-zaman*) is *a priori* and almost as general as that of being, encompassing the sensible and the non-sensible, that which moves and that which is at rest. Our idea of time results not from abstraction, stripping accidents from perceived objects, but from a mental representation based on an innate idea. Al-Baghdadi stops short of offering a precise definition of time, stating only that ‘were it to be said that time is the measure of being (*miqdar al-wujud*), that would be better than saying [as Aristotle does] that it is the measure of motion’. His reclassification of time as a subject for metaphysics rather than for physics represents a major conceptual shift, not a mere formalistic correction. It also breaks the traditional linkage between time and space. Concerning space, al-Baghdadi held unconventional views as well, but he did not remove its investigation from the domain of physics (see Space; Time).

Al-Baghdadi’s most significant departure in psychology concerns human self-awareness. Ibn Sina had raised the issue of our consciousness of our own psychic activities, but he had not fully pursued the implications for Aristotelian psychology of his approach. Al-Baghdadi took the matter much further, dispensing with the traditional psycholgical faculties and pressing his investigations in the direction of what we would call the unconscious.

Al-Baghdadi had many new ideas concerning the physics of motion. He seems to have adumbrated the notion of acceleration as an increase in the velocity of a moving body attributable to the application to it of a constant force. He also seems to suggest that motion is relative, that is, that there is motion only if the relative positions of the bodies in question change. These ideas are highlighted here because of their resemblance to modern thoughts on

the same subjects. The *Kitab al-mu‘tabar* contains many other, no less innovative ideas that have no modern counterpart; for example, the claim that each type of body has a characteristic velocity that reaches its maximum when its motion encounters no resistance. Although *al-Mu‘tabar* is not a systematic work, comprising instead notes on various subjects that al-Baghdadi wrote for himself over the years, Pines showed that the paramountcy of *a priori* knowledge underlies many of the work’s criticisms and innovations.

The impact of *al-Mu‘tabar* on Islamic thought seems to have been limited to the Ishraqi (illuminationist) tradition, broadly defined (see Illuminationist philosophy). Indeed, the work’s tripartite structure (logic, physics, metaphysics), the pride of place given to *a priori* knowledge, and the consequent primacy given to the author’s own speculations, are the distinguishing features of Ibn Sina’s *Isharat wa-’l-tanbihat*, the earliest prototype of the genre, and of al-Suhrawardi’s *al-Talwihat*, its most important representative. However, the spiritual tone of the latter two books is far less prominent in al-Baghdadi’s work, although perhaps not entirely absent. As Pines showed, the Ibn Sina of *al-Shifa’ (Healing)* is the target of many of al-Baghdadi’s criticisms, strictures often pursued in Fakhr al-Din al-Razi’s commentary on *al-Isharat*, and answered in Nasir al-Din al-Tusi’s glosses to the same work. Al-Tusi usually refers to ‘al-Baghdadi and the other later [philosophers],’ giving the impression that al-Baghdadi was the outstanding representative, or even founder, of a whole school of thinkers who challenged some of Ibn Sina’s views. Al-Suhrawardi’s *al-Talwihat* refers to al-Baghdadi obliquely as ‘one of the Jewish philosophers’. The target of that reference is made clear by Ibn Kammūna, whose commentary on *al-Talwihat* cites al-Baghdadi several times. Al-Baghdadi’s views were known to al-Shahrazuri, and he is mentioned some half-dozen times in Sadr al-Din al-Shirazi’s *al-Asfar al-arba‘a*.

Despite his conversion to Islam, al-Baghdadi’s works continued to be studied at the *yeshivah* of Baghdad, then the centre of Jewish conservatism, into the thirteenth century. Al-Baghdadi’s commentary to the Book of Ecclesiastes continued to be copied at the same *yeshivah*, with full acknowledgement of its authorship. Shmu’el ben Eli, head of the *yeshivah* and archrival of Moses Maimonides, cites the *Mu‘tabar* in support of his contention that even ‘the philosophers’ are forced to admit the possibility of bodily resurrection. Ben Eli does not reveal his source; it appears to have been Maimonides’ disciple, Yosef ben Yehudah, who tracked down the reference.

Ibn Khalliqan reports that al-Baghdadi had a high reputation in the field of medicine, and al-Suhrawardi refers to him as ‘a physician who sought to do philosophy’. He attended some of the Seljuq sultans and their families and is reported to have cured himself of leprosy, in the process causing himself a period of blindness. Ibn Abi Usaybi‘a’s biographical compendium relates some anecdotes and sayings and lists several of al-Baghdadi’s medical works. However, few if any of these survive, and none have been studied. Al-Baghdadi also wrote a short treatise ‘On the Reason that the Stars appear at Night but are invisible during the Day’, and another small tract on the intellect.

See also: Ibn Sina; Illuminationist philosophy

Y. TZVI LANGERMANN

List of works

al-Baghdadi (early 12th century) *Kitab al-Mu‘tabar*, Hyderabad: Osmania Publication Bureau, 1938-9, 3 vols.(A usable edition, one of the better of the Hyderabad printings.)

al-Baghdadi (early 12th century) Commentaries on Ecclesiastes, ed. S. Pines, ‘Towards the Study of Abu al-Barakat’s Commentary to Ecclesiastes: Four Texts’, *Tarbiz* 33, 1964: 198-213; repr. in *Bein Mahshevet Yisrael li-Mahshevet ha-‘Amim*, Jerusalem, 1977.(Four selected passages with translation and analysis, in Hebrew.)

References and further reading

Davidson, H.A. (1992) *Alfarabi, Avicenna, and Averroes, on Intellect*, New York: Oxford University Press, 154-61.(Includes a critical appraisal of Abu ’l-Barakat’s ideas on the intellect.)

Pines, S. (1979) *Studies in Abu’l-Barakat al-Baghdadi: Physics and Metaphysics*, The Collected Works of Shlomo Pines, vol. 1, Jerusalem and Leiden: Magnes & Brill.(A collection of Pines’ most important studies on al-Baghdadi, some not published elsewhere.)

Stroumsa, S. (1993) ‘On the Maimonidean Controversy in the East: the Role of Abu ’l-Barakat al-Baghdadi’, in H. Ben-Shammai (ed.) *Hebrew and Arabic Studies in Honour of Joshua Blau*, Tel Aviv and Jerusalem.(On the role of Abu ’l-Barakat’s writings in the resurrection controversy of the twelfth century; in Hebrew.)

Albert of Saxony (*c.*1316-90)

Albert of Saxony, active in the middle and late fourteenth century, taught at the University of Paris and was later instrumental in founding the University of Vienna. He is best known for his works on logic and natural philosophy. In the latter field he was influenced by John Buridan, but he was also influenced by the English logicians. His thought is rather typical of the sort that followed Buridan, combining critical analysis of language with epistemological realism. He was important in the diffusion of terminist logic in central Europe, and of the new physics in northern Italy.

Albert of Saxony, or Albert of Rickmersdorf (sometimes called Albertus Parvus (Little Albert) to distinguish him from Albert the Great) was born in Helmstedt, Germany, around 1316. He studied first in his native region before going on to Erfurt (although this is not known for certain) and then to Paris, where he became Master of Arts in 1351. In 1353, he became rector of the university, and taught in the arts faculty there for a decade; he also studied theology, though apparently without receiving a degree. After a period during which he was involved in diplomatic mediation between the pope and the Duke of Austria, Albert was put in charge of founding the University of Vienna, and in 1365 he became its first rector. In 1366 he was made a canon of Hildesheim and was named bishop of Halberstadt. He remained in the latter position until his death on 8 July 1390.

Since none of his theological writings survives, Albert of Saxony is known especially for his work in logic and natural philosophy. However, he also wrote commentaries on Aristotle's *Ethics* and *Economics*, as well as a few small works on mathematics (*Tractatus proportionum (Treatise on Proportions)*, *Quaestio de quadratura circuli (Question on the Quadrature of the Circle)*). In the field of logic, his main work is the summa entitled *Perutilis logica (Very Useful Logic)*. He also wrote a voluminous collection of *Sophismata* (examining various difficulties of interpretation due to the presence of syncategorematic words in sentences), a set of *Quaestiones logicales (Logical Questions)* that deal with semantical problems and the status of logic, and commentaries on Aristotle's Organon.

Although Buridan was then very popular in the arts faculty at Paris, Albert's work attests also to the influence of English ideas in Paris. The *Perutilis logica*, while developing treatises on obligations, insolubles and consequences, which were becoming more and more important at the time, is organized on the model of William of Ockham's *Summa logicae* (see William of Ockham §6). Albert accepts Ockham's conception of the nature of a sign. He believes that signification rests on a referential relation of the sign to the individual thing, and that the spoken sign depends for its signification on the conceptual sign. He follows Ockham again in his conception of universals and, for the most part, in his theory of supposition. In particular, he preserves Ockham's notion of simple supposition, the direct reference of a term to the concept on which it depends when it signifies an extra-mental thing. Finally, Albert follows Ockham in his theory of categories: contrary to Buridan, he refuses to treat quantity as a feature of reality in its own right, but rather reduces it to a disposition of substance and quality.

On a few points, however, Albert distances himself from Ockham. For instance, he denies that in disputation an equivocal proposition must be the object of a distinction through which it is assigned multiple senses: in disputation, even equivocal propositions can only be granted, denied or doubted. In his *Sophismata*, Albert often follows William Heytesbury (for example, in the analysis of epistemic verbs or of infinity). He admits that a proposition has its own signification, which is not that of its terms: just like a syncategorematic word, a proposition signifies a 'mode of being'. Albert avoids concretizing such modes of being and, in the final analysis, traces them back to relations among the things to which the terms refer. Nevertheless, he makes use of the idea of the distinguishable signification of the proposition in defining truth and in dealing with 'insolubles' (paradoxes of self-reference). Since every proposition, by its very form, signifies that it is true, an insoluble proposition will turn out to be false because it will signify at once both that it is true and that it is false (see Language, medieval theories of).

Albert's analysis of language is combined with an epistemological realism that emerges, for instance, from his analysis of the vacuum. One could imagine that, through divine omnipotence, a vacuum exists. However, no science of nature can countenance the hypothesis of the existence of a vacuum. Albert refuses to extend the reference of the terms of physics to supernatural possibilities; thus in his view, physics cannot be developed

through thought experiments or the study of imaginary cases, contrary to what was being done at Oxford at the time. Instead, physics must report on the natural course of things.

Aside from his commentaries on Aristotle's *libri naturales*, Albert wrote a commentary on John of Sacrobosco's *De sphera (Treatise on the Sphere)*, and a *Tractatus proportionum (Treatise on Proportions)* inspired by Thomas Bradwardine. Following the works of the Oxford Calculators (see Oxford Calculators), and of Nicole Oresme in Paris, he tried to calculate the acceleration of the fall of bodies, but without succeeding in determining whether it is proportionate to the time elapsed or to the distance covered. He was interested in many natural phenomena and studied the movements of the earth and the phenomena of tides and geology.

It is probably in the field of dynamics, however, that Albert's role is most important. In order to account for the motion of projectiles and for the acceleration in the fall of weights, he adopted Buridan's impetus theory, according to which impetus is a property acquired by bodies. With great clarity, he draws the consequences of extending this theory to celestial movements, and is thus able to reject the traditional notion of intelligences moving the spheres. He studies the motions of terrestrial and celestial bodies according to the same principles. Many of his contributions to the field of dynamics are in his commentary on Aristotle's *On the Heavens*, which was very influential in northern Italy. Albert of Saxony thus took part in the development of a vision of the cosmos which broke with Aristotelian views.

See also: Buridan, J.; Language, medieval theories of; Logic, medieval; Natural philosophy, medieval; William of Ockham

JOËL BIARD

List of works

Few texts are available in full in modern editions.

Albert of Saxony (before 1390) *Perutilis logica (Very Useful Logic)*, ed. and Spanish trans. A. Muñoz de García, Maracaibo: Universidad del Zulia, 1988.(This edition is after the incunable edition, first published in Venice (1522). An edition of the second treatise, *De proprietatibus terminorum (On the Properties of Terms)*, was published in C. Kann, *Die Eigenschaften der termini*, Amsterdam: Brill, 1993.)

Albert of Saxony (before 1390) *Quaestiones in artem veterem (Questions on the Old Logic)*, ed. A. Muñoz de García, Maracaibo: Universidad del Zulia, 1988.(This is a critical edition.)

References and further reading

Biard, J. (1989) 'Les sophismes du savoir: Albert de Saxe entre Jean Buridan et Guillaume Heytesbury' (The Sophismata on Knowledge: Albert of Saxony between John Buridan and William Heytesbury), *Vivarium* 27: 36-50.(Discussion of one of Albert's sophismata.)

Biard, J. (ed.) (1991) *Paris-Vienne au XIVe siècle. Itinéraires d'Albert de Saxe (From Paris to Vienna in the Fourteenth Century: The Journeys of Albert of Saxony)*, Actes de la table ronde internationale, Paris, 19-22 juin 1990, Paris: Vrin.(Twenty-one contributions showing the recent state of knowledge about logic and natural philosophy in the fourteenth century.)

Biard, J. (1993) 'Albert de Saxe et les sophismes de l'infini' (Albert of Saxony and the Sophismata on Infinity), in S. Read (ed.) *Sophisms in Medieval Logic and Grammar*, Boston: Kluwer, 288-303.(Discussion of Albert's sophismata.)

Gonzales, A. (1958) 'The Theory of Assertoric Consequences in Albert of Saxony', *Franciscan Studies* 18: 290-354, and 19: 13-114.(A formal study of the theory of inferences.)

Braakhuis, H.A.G. (1993) 'Albert of Saxony's *De obligationibus*: Its Place in the Development of Fourteenth Century Obligational Theory', in K. Jacobi (ed.) *Argumentationstheorie: scholastische Forschung zur den logischen und semantischen Regeln korrekten Folgerns*, Leiden: Brill, 323-41.(Discussion of Albert's treatise on obligation.)

Berger, H. (1994) 'Albert von Sachsen (1316[?]-1390). Bibliographie der Sekundärliteratur' (Albert of Saxony: Bibliography of Secondary Literature), *Bulletin de philosophie médiévale* 36: 146-85.(Bibliography of works relevant to Albert.)

Heidingsfelder, G. (1927) *Albert von Sachsen. Sein Lebensgang und sein Kommentar zur Nikomachischen Ethik des Aristoteles (Albert of Saxony: His Life and Commentary on Aristotle's Nicomachean Ethics)*, Beiträge zur Geschichte der philosophie des Mittelalters XXII, 3-4, Münster: Aschendorff.(The biographical section is now dated, but this remains the only study of Albert's ethics.)

Kann, C. (1993) *Die Eigenschaften der termini. Eine Untersuchung zur 'Perutilis Logica' des Alberts von Sachsen (The Properties of Terms: A Study of Albert of Saxony's Perutilis logica),* Amsterdam: Brill.(A study of the theory of the properties of terms, including the theory of supposition, with an edition of the second treatise of the *Perutilis logica*.)

Muñoz García, A. (1990) 'Albert of Saxony, Bibliographie', *Bulletin de philosophie médiévale* 32: 161-98. (Contains lists of manuscripts and editions, most of them in incunabula, of the texts of Albert of Saxony.)

Sarnowsky, J. (1989) *Die aristotelisch-scholastiche Theorie der Bewegung. Studien zum Kommentar Alberts von Sachsen zur Physik des Aristoteles (The Aristotelian-Scholastic Theory of Motion: Studies on Albert of Saxony's Commentary on Aristotle's Physics),* Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters XXXII, Münster: Aschendorff.(A complete and documented study of Albert's physics.)

Albert the Great (1200-80)

Albert the Great was the first scholastic interpreter of Aristotle's work in its entirety, as well as being a theologian and preacher. He left an encyclopedic body of work covering all areas of medieval knowledge, both in philosophy (logic, ethics, metaphysics, sciences of nature, meteorology, mineralogy, psychology, anthropology, physiology, biology, natural sciences and zoology) and in theology (biblical commentaries, systematic theology, liturgy and sermons). His philosophical work is based on both Arabic sources (including Alfarabi, Avicenna and Averroes) and Greek and Byzantine sources (such as Eustratius of Nicaea and Michael of Ephesus). Its aim is to insure that the Latin world was properly introduced to philosophy by providing a systematic exposition of Aristotelian positions.

Albert's method of exposition (paraphrase in the style of Avicenna rather than literal commentary in the style of Averroes), the relative heterogeneity of his sources and his own avowed general intention 'to list the opinions of the philosophers without asserting anything about the truth' of the opinions listed, all contribute to making his work seem eclectic or even theoretically inconsistent. This was compounded by the nature and number of spurious writings which, beginning in the fourteenth century, were traditionally attributed to him in the fields of alchemy, obstetrics, magic and necromancy, such as The Great and the Little Albert, The Secrets of Women and The Secrets of the Egyptians. This impression fades, however, when one examines the authentic works in the light of the history of medieval Aristotelianism and of the reception of the philosophical sources of late antiquity in the context of the thirteenth-century university.

1 Introduction of philosophy to the Latins

After studying in Padua and Cologne, Albert entered the Dominican order around 1220. He was the first German to become master of theology at the University of Paris (1245-8). He then taught at the Dominican *studium* at Cologne (where his students included Thomas Aquinas (until 1252) and Ulrich of Strasbourg). Between 1254 and 1257 he was the Dominican Provincial of Teutonia (Germany). As bishop of Ratisbon (Regensburg) in 1260, at the express request of Pope Urban IV, he preached the crusade 'in Germany, Bohemia and other Germanic countries'. After various visits to Würzburg (1264) and Strasbourg (1267), he lived in Cologne until his death in 1280.

Albert's teaching in Paris was dominated by his writing the *Summa de creaturis (Book of the Creatures)* before 1246. Despite the censure imposed on the study of Aristotle during the preceding decade, Albert made extensive use of Greek-Arabic Aristotelianism in his theology (see Aristotelianism, medieval). The same tendency can be seen in his commentary on the *Sentences*, begun in Paris and finished in Cologne in 1249. It was also in Cologne, while at the *studium generale* of the Dominican Order, that Albert wrote most of his works in natural philosophy, including the *Physics*, the commentary on *On the Heavens*, the *Liber de natura locorum (The Nature of Places)* and the *De causis et proprietatibus elementorum (The Causes and Properties of the Elements)*. In 1250-2, he presented in lectures his first commentary on the *Nicomachean Ethics* (the *Super Ethica*, a question-commentary). He returned to the *Nicomachean Ethics* in 1262-3, this time producing a paraphrastic commentary, the *Ethica*.

Albert's large-scale paraphrases on the Organon were written between 1252 and 1256, based on Arabic works by Avicenna (see Ibn Sina) and Alfarabi (see al-Farabi) which are for the most part lost today, and also on Latin works (the *Logica modernorum* and commentaries by Robert Kilwardby, with which Albert became familiar in Paris). The works on botany (*De vegetabilibus et plantis libri VII*) and on mineralogy (*De mineralibus*) were written in 1256-7; the treatises on biology and zoology (*Quaestiones super De animalibus*) are drawn from disputed questions held in 1258. In 1262-3, Albert wrote his commentary on Euclid (*Super Euclidem*). The last years of his life were devoted to metaphysics and theology; his paraphrase of the *Metaphysics* was written in 1263-7. At the same time he wrote the *De causis et processu universitatis (The Causes and Development of the Universe)*, a general exposition of an Aristotelian natural theology in which Albert brings together all the intellectual themes of late antiquity that were available in the second half of the thirteenth century.

At this time, the mendicant orders vigorously denounced the intrusion of philosophy into theology. For instance, Bonaventure in the *Collationes de Decem praeceptis (Discourses on the Ten Commandments)* in 1267 attacked the 'arrogant presumption of philosophical investigation' that 'corrupts all of Holy Scripture' and denounced not only those who 'create' the philosophical 'fictions' but also those who 'sustain and reproduce them'. In that context,

Albert's project to 'bring Aristotle to the Latins' constitutes a genuine defence of philosophical endeavour, not only in the medieval university but also in Christian society in general. In Albert's view, the enemies of philosophy who 'killed Socrates, threw Plato out of Athens… and forced Aristotle into exile' (he openly criticizes them in his paraphrase of the *Politics* VIII, 6) are comparable to the 'brute beasts' of his time who 'blaspheme against what they don't know', university 'preachers' who in their sermons 'attack the use of philosophy with all possible means', 'without anyone's being able to answer them' (commentary on the VIIth Letter of Dionysius).

In opposition to these critics, Albert asserts the need to know and assimilate the philosophy of the ancients. His insistence on the need for philosophy might seem ambiguous in so far as it gives rise to a distinction between two disciplines 'distinct in their principles': theology, which is 'founded on revelation and prophecy', and philosophy, which is 'founded on reason' (*Metaphysica* XI, 3, 7). This distinction, however, corresponds to a deeply rooted tendency in the thirteenth century. The condemnations of 1277 at Paris are evidence of its strength and efficacy. *Nihil ad me de Dei miraculis, cum ego de naturalibus disseram* (God's miracles mean nothing to me, since I am discussing natural things and events), the rallying cry of the 'Latin Averroists' popularized by Siger of Brabant (*De anima intellectiva (The Intellective Soul)* III), was originally Albert's (*De generatione et corruptione* I, 1, 22). He borrowed it consciously and simultaneously from two authorities, one philosophical (Averroes (see Ibn Rushd)) and the other theological (Bernard of Clairvaux, reformulating a passage from Augustine). Albert wrote all his philosophical paraphrases in order to develop fully the discipline of philosophical research, a discipline that is autonomous in its own domain, the domain of rational argumentation.

The nature of Albert's commentaries might also seem ambiguous, in so far as his avowed Aristotelianism covers a complex mix of Aristotelian and Neoplatonic theses. This ambiguity, however, cannot be blamed on Albert as it is present in his own Arabic 'Aristotelian' sources, which were for the most part permeated with the syncretic view of Aristotle inherited from the Neoplatonic commentators of the fifth and sixth centuries (see Aristotelianism in Islamic philosophy; Neoplatonism in Islamic philosophy). In asserting that 'philosophical perfection' can only be attained with both Aristotle and Plato as its foundation (*Metaphysica* I, 5, 15), Albert, who knows little of Plato, is really taking up the 'harmonizing' reading of the Neoplatonic philosophers of late antiquity (see Neoplatonism), which was adopted by the Arabic Aristotelians. If his vision of Aristotle seems more Neoplatonic than Aristotelian, it is precisely because it is based as much on the philosophies of Arabic commentators on Aristotle (Alfarabi, Avicenna, the *Liber de causis* and Averroes) as it is on the philosophy of Aristotle himself.

As the principal engineer of the introduction of philosophy to the Latins, Albert tried to portray as homogeneous a philosophy that is not and cannot be homogeneous in the eyes of the philologist. It is, however, this Arabic-Latin version of Aristotelianism which was successfully installed in the Schools and was opposed to other versions of Aristotle in the fifteenth century, Thomistic Aristotelianism and the Aristotelianism of the school of John Buridan, which present Aristotle in a form more recognizable to us (see Aristotelianism, medieval).

2 Logic

Although he paraphrased all of the Organon, in expositions that were used frequently until the end of the fifteenth century, it is not on account of its contribution to the development of the *logica modernorum* that Albert's logical work is most noteworthy. His principal contribution is in connection with the problem of universals (see Universals). Porphyry, in the *Isagōgē*, had wanted to keep this problem separate from the thought of the logician, and to make it the domain of metaphysics and theology. It was Albert who first systematically formulated the theory of universals that prevailed in scholastic and neoscholastic thought, the doctrine that there are three types of universals (*ante rem*, *in re* and *post rem*). This doctrine is characteristic of the harmonizing tendency that dominates all of Albert's thought. It is also evidence of the continuity which, thanks to this doctrine, was established between the philosophy of late antiquity and the philosophy of the late Middle Ages.

In answer to Porphyry's problem (whether genera or species exist in themselves or reside in mere concepts alone; whether, if they exist, they are corporeal or incorporeal; and whether they exist apart from or in and dependent on sense objects), Albert does not repeat the arguments and theses of the realists and nominalists of the twelfth century (though he knows them), and he does not choose between realism, conceptualism and nominalism. Instead, he takes up and develops a distinction between types of universals which allows him to give a three-part answer to the problem posed by Porphyry, an answer that is neither realist nor nominalist in the sense of the twelfth century but is meant to transcend the conflict itself. Universals, then, are neither universal extramental 'things' (as the

realists believed), nor simple words or concepts (as the nominalists believed). Rather, a universal is one entity with three different aspects, three *modi essendi* (modes of being), which differ depending on whether the universal is considered in itself (in divine thought or the separated Intellects), in natural things or in human thought. Albert draws this three-part distinction from Avicenna's *Logica*, reinforced by certain remarks from Eustratius of Nicaea's *Commentary on the Nicomachean Ethics*. This view of universals makes the Platonic notion of separated Forms compatible with the Aristotelian notions of immanent forms and abstract concepts, and makes it possible to preserve both notions within the same theory.

In reviving this doctrine, Albert unwittingly takes up the Neoplatonic solution to the problem of universals, the distinction between universals that are before particulars (*pro tōn pollōn*), after particulars (*epi tois pollois*) and in particulars (*en tois pollois*), a distinction that was systematized most notably by Ammonius, David and Elias (see Neoplatonism). Albert's extensive use of Arabic and Byzantine sources to illuminate Latin knowledge of the twelfth century had at least two clear consequences: the subordination of logic to metaphysics and, in metaphysics itself, the subordination of Aristotle's 'Aristotelianism' to a Greek-Arabic version of Aristotelianism (which became the foundation of early neoscholasticism). Indeed, it is on this structure that Albert's disciples (the 'neo-Albertians') built their school's most characteristic positions as early as the fifteenth century. Jean de Maisonneuve, Albert's principal disciple in Paris in the early 1400s, criticized the nominalist followers of Buridan (the *epicuri litterales*) for reducing the universal to the simple abstract concept, a criticism which Heimeric of Campo levelled against Thomists as well. On the other hand, Heimeric also criticized the *formalizantes* (such as Jerome of Prague) for believing in the existence of separated universals such as Plato's Ideas. In short, Albert's disciples criticized other views of universals for being unilateral philosophies that consider only one aspect of the being of universals.

3 Psychology

In the field of psychology, Albert worked primarily on the exposition of the fundamental concepts of Aristotle's theory of the soul, especially of the theory of the intellect. Albert tried to correct and contribute to the exposition of Aristotle in two areas. He fought against the doctrine of the unity of the intellect (or 'monopsychism' as Leibniz called it), which tradition attributed to Averroes but which Albert attributed to 'all of the Arabs'; and he followed Averroes in criticizing the materialism of Alexander of Aphrodisias. In addition to his critical work, Albert also tried to integrate the essence of the Greek-Arabic theory of the intellect, beginning with Averroes' version, from which he 'dissents little' ('*in paucis dissentimus*', *De anima* III, 3, 11). Indeed, it is clear that the monopsychism Albert criticizes is the same as that which Averroes already criticized in his 'Great' commentary on *On the Soul*: Avempace's thesis that there is only one intellect for all men, which is joined to the human soul 'by means of images' (*phantasmata*) (see Ibn Bajja; Ibn Rushd). This view is unacceptable, says Averroes, because it reduces the material intellect to a simple 'faculty of imagination'.

Albert only occasionally extends his criticism of monopsychism to Averroes. There are two reasons for this reticence: first, Albert is too dependent on Averroes to criticize his theory of intellect without making his own theories incoherent; and second, he does not interpret Averroes in the light of the Averroists' extrapolations, as Aquinas does. Far from seeing Averroes as the 'debaser' and the 'corrupter of Aristotelianism', Albert on the contrary wants to show that Averroes is the only one to have successfully opposed Alexander's materialist theory on the grounds that it is a 'very grave error which entails the denial of all the nobility and even of the immortality of the human soul' (*De anima* III, 2, 5). Therefore, Albert holds the view that the possible intellect is external to man, not because he understands it as a monopsychist thesis, as Aquinas does, but because, like Averroes, he understands it in the more precise and limited sense that the possible intellect is not the perishable 'form' of a perishable body, nor a 'corporeal power caused by the elements' that constitute the body.

What Albert judges to be fundamental in Averroes, then, is the criticism of Alexander's theory of the 'eduction' of the intellect, not his thesis of the unity of the possible intellect. That is why Albert often asserts his perfect agreement with Averroes who, better than any other, was able to prove the thesis which 'from antiquity, all Aristotelians have held', excepting Alexander, that 'the intellect enters the soul from the outside, it does not arise from the composite or the mix of the elements, and does not pre-exist in potentiality in them'. To apply the *eductio formarum* (eduction of forms) to the intellect implies denying its 'divinity' and 'depriving man of his nature', and that poses a greater danger than giving an unsatisfactory explanation of the manner in which the possible intellect

is united to human beings.

The essence of Albert's criticism of Averroes thus centres on the difficult notion of the 'acquired intellect' (*intellectus adeptus*). Averroes reduces the acquired intellect to a momentary union of the human soul and the separated intellect (each time there is an act of intellection), when instead one should think of it as a real power and part of the soul, emanated in it, which is developed and strengthened by the acquiring of more intelligibles. In Aristotelian language, one should think of the acquired intellect as a stable disposition (*habitus*) and not as a transitory state (*qualitas passibilis*). Thus, Albert does not reject Averroes' view of the union of the soul with the separated intellect: he adapts and perfects it. It is also on the basis of this view that Albert reorganizes the entirety of the Aristotelian doctrine of the intellect; and that he establishes, as early as the *Summa de creaturis*, a correspondence table for the different classifications found in tradition. In this, Albert and Aquinas are fundamentally opposed. Albert takes Averroes as his guide to reinterpret Aristotelianism, while Aquinas tries to distinguish Averroes from the Arabic sources (Avicenna and al-Ghazali) and the 'Greek' sources (Alexander, Themistius and Theophrastus) in order to interpret Aristotle anew against him (see Averroism).

4 Metaphysics

In the field of metaphysics, Albert's work takes an original direction, again characterized by a certain syncretism. Drawing from Aristotle, Avicenna (see Ibn Sina), the *Liber de causis* (see Liber de causis) and Pseudo-Dionysius, he formulates a system which again places him very much at odds with Aquinas. Albert's philosophy is set in the context of Aristotelian cosmology, which he claims is valid from the point of view of natural reason. As a consequence, Albert admits in his philosophy the system of Intelligences, which he carefully distinguishes from angels. (He considers that identifying the two is a theological error, though not a specifically Christian one since he denounces it primarily in two Jewish thinkers, Moses of Egypt, also called Maimonides, and Isaac Israeli.) This distinction was taken up and made more rigid by Albert's German Dominican followers, including Dietrich of Freiberg and Berthold of Moosburg, together with the distinction between the order of 'natural providence' (the order to which Aristotle, the Arabic philosophers and Proclus refer) and the order of 'voluntary providence' (the order to which the Bible and theologians refer). This is a strictly philosophical way of contrasting the world of nature and that of miracles, the world of the natural and that of the supernatural, a contrast which theologians capture by the strictly theological division between the 'ordered' and the 'absolute' power of God.

However, in his formulation of an Aristotelian metaphysics, Albert adds two important corrections to the elements he adopts from the Arabic sources. First, while adopting Avicenna's metaphysics of emanation, he radically modifies the theory of formal emanation (*fluxus* or *influentia*). The *fluxus formae* from which the intelligible and material universe arises does not consist in an 'overflowing' (*infusio*, *effusio*) of the forms from an emanating Principle (Avicenna's and al-Ghazali's *dator formarum*). Rather, it consists in an anagogical process, the final causality of the 'appeal of the good' (*advocatio boni*), a view Albert draws from Pseudo-Dionysius, John Scottus Eriugena and Maximus the Confessor, all of whom he would have read in the Dionysian corpus of the University of Paris. As a partisan of the unity of substantial form, who holds at the same time that the forms of things are contained in matter in an inchoate state ('the inchoation of forms'), Albert views all form-generating processes as ruled by the celestial spheres and their movers by means of a causality of attraction rather than strict emanation. The first principle or 'First Cause' does not infuse forms into matter: form and matter are con-created. The function of the first principle is to call to itself all forms that are contained in matter, to unify them and to assemble them by means of the 'final attraction' that it exerts on everything that is.

This notion of 'attraction' by the Good presupposes an identification of the First Cause of Arabic Aristotelianism with the Platonic Good. Albert makes this identification consciously by showing that the Latins who reduce emanation to a simple metaphysical mechanism corrupt all of philosophy. The true theory of emanation is the one that makes the Good into a principle that is 'diffusive of itself and of being' (*diffusivum sui et esse*), not by overflowing but by attracting. Albert explains that the supreme Good, the First Cause, 'calls all things to be', that is, to 'resemble' it, because the nature of goodness is to call (*bonum* comes from *boo*, *boas*, that is, *voco*, *vocas*, to call), and its diffusion is nothing other than its calling (*diffusio* and *boatus* are synonyms). He supports this view, not without paradox, by appealing to Paul's letter to the Romans 4: 17: 'God calls those things that are not as well as those that are'.

Second, in order to certify that this assimilation of the First Cause with the final cause of being qua being is

authentically Aristotelian, Albert claims that Aristotle's metaphysics is not the last word of Aristotelianism. Metaphysics must be completed by theology. Albert claims that he finds this theology in the *Liber de causis*. Far from seeing in the *Liber de causis* an Arabic adaptation of Proclus' *Elements of Theology* (as his student Aquinas did), Albert asserts that it is a work compiled by the mysterious 'David the Jew' (possibly Ibn Daud?) on the basis of one of Aristotle's letters *On the Principle of the Universe* (really a work of Alexander of Aphrodisias, which has survived only in Arabic translation) and other elements borrowed from Aristotelian philosophers. Albert's belief in the Aristotelian authenticity of the *Liber* (widely accepted in Paris in the 1250s), allows him to spell out a complete Aristotelian system that comprises metaphysics (the theory of being as being) and theology (the theory of the cause of being) and goes further than the rudiments available in Book XII of Aristotle's *Metaphysics*. The supreme principle is not a mere first mover (*primum movens*), it is also a first producer (*primum agens*) that produces (emanates) all things and draws back (attracts) all things into one.

In his *De causis et processu universitatis*, the height of his speculative philosophy, Albert offers us the most important 'reconstruction' of Aristotelian theology we have received from the middle ages. This complete theology is composed of a theory of the First Cause which draws from all the Arabic sources, (in particular al-Ghazali), and a theory of emanation drawn from the *Liber de causis*. It is thus clear that, unlike his contemporaries, Albert does not merely contrast revealed theology and rational philosophy, but within philosophy itself he contrasts natural theology and simple metaphysics. Aristotle's metaphysics is therefore completed twice.

Book I of *De causis et processu universitatis*, titled *De proprietatibus primae causae et eorum quae a prima causa procedunt* (On the Properties of the First Cause and of the Beings That Emanate From It), reveals the character of Albert's general inquiry by characterizing the three philosophical positions of antiquity (Epicurean, Stoic and Aristotelian) on the basis of their relation to the fundamental problem of theology: *de primo omnium principio* (the first principle of all things) or *de universi esse principio* (the universal principle of all things). Then comes the analysis of the fundamental themes of Aristotelian theology: *De scientia primi* (the knowledge of the first principle), *De libertate, voluntate et omnipotentia primi* (the freedom, will and omnipotence of the first principle) and *De fluxu causatorum a causa prima et causatorum ordine* (the outpouring of causal things from the first cause and their ordered relationship), all of which are profoundly influenced by al-Ghazali's views. In fact, Treatise IV, which constitutes the transition between Books I and II, corresponds exactly to the programme of metaphysics that al-Ghazali formulates at the beginning of Book I, V, of his own *Metaphysica*: *Quomodo omnia habent esse a primo principio et quomodo omnia perveniunt ad unum qui est causa causarum* (How all things have their being from the first principle, and how all things return to the One that is the cause of causes).

Albert gives his paraphrase of the *Liber de Causis* in Book II, which he devotes to the analysis of the elements of the noetic cosmos (Intelligences and 'noble souls', movers of the heavens), the description of the outpouring of beings (*de fluxu entis*) and the theory of the government of the universe by the First Cause. On almost every point, he holds views opposed to those of Aquinas in the latter's commentary. The strength and appeal of Albert's reading of the completed 'Aristotelian' system is such that it resists Aquinas' philological discovery. Until his death, Albert continued to assert the Aristotelian authenticity of the *Liber de causis*, and the necessity of it for completing the *Metaphysics* with a theology.

The distance between the philosophical positions of Albert and Aquinas can be seen even more clearly in their more specific and detailed theories. In the field of ontology, for example, Albert holds Avicenna's theory of the 'indifference of the essence' (the essence itself is neither universal, in the way empirical abstract concepts are, nor particular, in the way particular beings existing outside the soul are), and he draws a connection between this theory and the theory of the three states of a universal. Thus he provides a picture of the process of abstraction that is entirely different from Aquinas' Aristotelian view. This distance between the two was later reinforced by the neo-Albertians: neither Jean de Maisonneuve nor Heimeric of Campo, for whom 'in its essence, the universal is one, though it can occur in the soul, in the thing and in itself' according to three modes of being, viewed the universal as it is in the soul using nominalist and/or Thomistic models of abstraction. The formation of the universal *in anima* is not the result of abstractive induction on the basis of particulars, but rather the result of a complex illumination of the human soul by the Intelligences, according to the process of mental unification and simplification described in the theory of attractive causality. The formation of the concept called abstract is the result of the fact that the 'human soul is the instrument of the light of the First Intelligence' and that the First Intelligence uses it in order to draw back everything into one.

Similarly, while Albert, like Aquinas, is a resolute supporter of the analogy of being, his view, described long before that of Aquinas, combines in an original way the ‘focal’ analogy of Averroes (*analogia attributionis or analogia accidentis*, the coordination of the different ways in which being is accepted by the category of substance), and Pseudo-Dionysius’ analogy ‘of reception’ (*analogia recipientium*, the defining of each being by its ‘measure’ or ‘receptive capacity’ which places it in a hierarchy). The problems which Albert’s and Aquinas’ theories of analogy set out to solve are entirely different. Albert’s understanding of analogy is primarily meant to correct the static version of emanationism that dominates the Latin interpretation of the ‘Aristotelian’ cosmology and noetic. He rejects the Latin analogy between the overflowing of the first cause and the way in which the light of the sun is incorporated into different bodies: God’s communication with beings, according to the *analogia recipientium*, is not the simple overflowing of the ‘giver of forms’ into the universe of beings subordinated to him. For the Latin disciples of Avicenna, the light (*lux*) of the First Cause, unique and identical in itself, applies indifferently and uniformly to all beings, shining (*superlucens*) the same light (*lumen*) on all. It gets differentiated within them, according to their receptive capacity, which is determined by their nature or essence. For Albert, the axiom according to which the ‘received’ (*receptum*) is found in the ‘receiver’ (*recipiens*) according to its analogy or receptive capacity, is not sufficient to characterize the analogical communication of the Principle (as long as this capacity is conceived as a mere passive reception, and not rather as an active assimilation performed for all beings by the intellectual beings only). The communication of the Principle is fully realized only in the intellectual conversion of the entirety of being. In turn, this conversion occurs through the mediation of those beings that are capable of intellective activity and who insure the anagogical assimilation of the universe to the principle from which it emanates. The diffusion of the Good or the First Principle is not a simple ‘exit’, it is a double movement of exit (*exitus*) and return (*reditus*), of descending and rising, to which all thinking beings contribute.

For Albert, the theory of analogy is thus not primarily meant to answer the problem of the ‘multiple meanings of being’. It is not an ontological or semantic theory meant to solve the aporias of the ‘problem of being’ formulated by Aristotle in Book IV of the *Metaphysics*. Rather, it is a theological doctrine which, under the name of ‘Aristotelianism’, sets out a peculiar version of the Neoplatonic theory of the intellectual procession of the universe. Despite the fact that Albert and Aquinas share a certain language (indifference of essences, the analogy of being, the analogy of reception), they are answering different problems and their philosophical intentions are not congruent. On all important points of metaphysics, therefore, the historiographical notion of an ‘Alberto-Thomist Aristotelianism’ seems quite fragile, if not unfounded.

5 Ethics

In moral philosophy, Albert is a resolute supporter of Aristotle’s view that the ‘contemplative’ or ‘speculative life’ surpasses all other forms of life. Albert describes philosophical contemplation as the height and end of human life. These ideas on ‘intellectual happiness’ were later taken up both by the Latin Averroists (from John of Jandun to Nicoletto Vernia) and by Dante in his *Convivio*. Here again, Albert is close to Averroes and far from Aquinas. Paradoxically, Averroes’ claim that the philosophical life is necessary and pre-eminent (see Ibn Rushd §4) finds its clearest exposition in Albert’s Aristotelianism. It is in Albert’s work that psychology, the science of animate life, manages to grow naturally into a theology, in so far as psychology is in its highest branch a science of human beings, or more precisely in Aristotle’s terms, ‘a science of the most fundamental and best part of the being of men’.

The achievement of Albert’s Aristotelianism, then, is that it naturally links psychology, ethics and philosophical theology. Aristotle’s definition of the humanity of man receives an essentially practical interpretation: what defines man is his aspiration to ‘live according to the noblest part of himself’ (*secundum optimum eorum quae in ipso*). Since this ‘noblest’ (*principale et melius*) part is the intellect (considered both as a ‘divine element present in human beings’ and as ‘what is in the highest degree man himself’) it is by giving a new interpretation of the doctrine of the acquired intellect (*intellectus adeptus*) that Albert builds an ethical system. This ethical system, though it is set ‘against the contemplation of love’ described by Aquinas and theologians, and despite the condemnations of 1277, imposed itself as a kind of corporate ideal to the masters of arts, both Averroists and others.

The new doctrine of the acquired intellect can be easily summarized. The acquired intellect designates the state of the human soul when it is joined (*conjunctio*, *connexio*) to the separated agent intellect. This union can be in

potentiality - since the agent intellect is naturally joined to us as a power and faculty of the soul - or it can be causal - since the agent intellect is the efficient cause of the actualizing of intelligibles in the soul and since, in the acquired intellect, the agent intellect becomes the form of the soul. Their union produces in the soul the state of contemplative wisdom described by Aristotle as ‘the speculative life’. The state of union or speculative life is the state which philosophers define as the ‘supreme end’ of human life, the object of a specific longing (*fiducia philosophantium*). Thus, there is here on earth a form of life which, while it anticipates the happiness promised to the elect in the next world, is none the less self-sufficient (it is in this sense that Alfarabi defines the ‘other life’ as the union of the philosopher with the separated Intellect, in *De intellectu et intellecto* (see al-Farabi)). In the most literal sense of the term, this form of life is ‘acquired’: it is the result of work and implies a progression (*moveri ad continuationem*). The content of this form of life is precisely what Aristotle defined as the object of philosophical theology, the contemplation of the separated beings. The kind of life that is characterized by philosophical contemplation can be called ‘intellectual happiness’.

This conception of philosophy as a contemplative form of life is indissolubly speculative and ethical. Albert arrives at it by drawing from diverse sources, not only the Islamic Aristotelians (Alfarabi, Avicenna and Averroes), but also the Byzantine commentators on the *Nicomachean Ethics*, Eustratius of Nicaea and Michael of Ephesus, whom he was one of the first to read (see Byzantine philosophy). Albert’s often repeated identification of the Arabic doctrine of the acquired intellect (*intellectus adeptus*) and the ‘Greek’ doctrine of the ‘possessed intellect’ (*intellectus possessus*) is meant to establish the authentically Aristotelian character of Albert’s reformulation of the goal of philosophical endeavour. Albert makes his view of this goal an ethical view, by describing philosophy’s culmination as a state which he is not afraid to characterize, with Aristotle, as ‘divine’ (*intellectus divinus*). In comparison with authentic Aristotelianism, however, Albert’s position is marked by something remarkably novel: the idea of an ascetic progression of the human soul, rising progressively from the knowledge of the sublunar world to the intellectual intuition of the separated realities.

The central idea in Albert’s thought is that here on earth, there is a happiness that rewards a philosophical effort understood as a progressive detachment of the human soul from sensible things and the ‘acquisition of the intellect’. Albert’s metaphysics, his psychology, his ethics and his natural theology all converge in this central thought. It is this idea that the so-called ‘Latin Averroists’ in Paris inherited from Albert. It is also, probably, by this aspect of his work that the master of Cologne exerted his most long-lived and varied influence.

See also: Aristotle; Aristotelianism, medieval; Aquinas, T.; Averroism; Ibn Rushd; Ibn Sina; Islamic philosophy: transmission into Western Europe; Language, medieval theories of; Liber de causis; Logic, medieval; Neoplatonism; Platonism, medieval

Translated from the original French by Claudia Eisen Murphy
ALAIN DE LIBERA

List of works

Albert the Great (1200-80) Works, ed. P. Jammy, *Alberti Magni Opera omnia*, Lyon, 1651, 21 vols; ed. A. Borgnet and E. Borgnet, *Alberti Magni Opera omnia*, Paris, 1890-9, 38 vols; *Alberti Magni Opera omnia edenda curavit Institutum Alberti Magni Coloniense Bernhardo Geyer praeside*, Münster: Aschendorff, 1951-. (The latter is a critical edition, publication of which is still underway.)

Albert the Great (1200-80) *De anima (On the Soul)*, ed. C. Stroik, *Opera omnia* VII, 1, Münster: Aschendorff, 1968.(Albert’s Aristotelian writing on the soul.)

Albert the Great (1200-80) *Metaphysica*, ed. B. Geyer, *Opera omnia* XVI, 1-2, Münster: Aschendorff, 1960-4. (Albert’s paraphrase of Aristotle’s *Metaphysics*.)

Albert the Great (1200-80) *De causis et processu universitatis prima causa (The Causes and Development of the Universe)*, ed. W. Fauser, *Opera omnia* XVII, 2, Münster: Aschendorff, 1993.(Exposition of Aristotelian natural theology.)

References and further reading

Bianchi, L. (1990) *Il Vescovo e i Filosofi. La condanna pariginia del 1277 e l’evoluzione dell’aristotelismo scolastico (The Bishop and the Philosophers: The Parisian Condemnation of 1277 and the Evolution of Scholastic Aristotelianism)*, Bergamo: Lubrina.(On the influence of the theory of ‘intellectual happiness’ on the ethics of the masters of arts.)

D'Ancona Costa, C. (1995) *Recherches sur le Liber de causis (Research on the Liber de causis)*, Études de philosophie médiévale LXXII, Paris: Vrin.(A precise analysis of the theoretical universe of the *Liber de causis* and of its scholastic reception, with bibliography.)

Hoenen M.J.F.M. and Libera, A. de (1995) *Albertus Magnus und der Albertismus. Deutsche philosophische Kultur des Mittelalters (Albert the Great and Albertism: German Philosophical Culture of the Middle Ages)*, Studien und Texte zur Geistesgeschichte des Mittelalters XLVIII, Leiden: Brill.(On the influence of Albert in Germany, and the neo-Albertism of the fifteenth century, particularly Jean de Maisonneuve and Heimeric of Campo.)

Kaluza, Z. (1986) 'Le *De universali reali* de Jean de Maisonneuve et les epicuri litterales' (Jean de Maisonneuve's *De universali reali* and the Followers of Buridan), *Freiburger Zeitschrift für Philosophie und Theologie* 33: 469-516. (Standard exposition of the neo-Albertian doctrine of universals and its criticism of Buridanism.)

Libera, A. de (1990) *Albert le Grand et la Philosophie (Albert the Great and Philosophy)*, Paris: Vrin.(A general exposition of the broad themes in Albert's philosophy.)

Libera, A. de (1994) *La Mystique Rhenane. D'Albert le Grand à Maître Eckhart (Rhineland Mysticism from Albert the Great to Meister Eckhart)*, Paris: Éditions du Seuil.(An overview of the German Dominican school beginning with Albert the Great's metaphysics and psychology.)

Weisheipl, J.A. (1980) *Albertus Magnus and the Sciences: Commemorative Essays 1980*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies. (Excellent collective synthesis of Albert's contribution to medieval science.)

Zimmermann, A. (1981) *Albert der Große. Seine Zeit, Sein Werk, Seine Wirkung (Albert the Great: His Times, His Work and His Influence)*, Miscellanea Mediaevalia 14, Berlin: de Gruyter.(Collective work covering most of the areas of philosophy to which Albert contributed.)

Albo, Joseph (*c*.1380-*c*.1444)

Writing in the early fifteenth century, in times of extreme urgency for Spanish Jewry, Joseph Albo presented Judaism as an axiomatic system founded on three primary principles and eight secondary ones. His Sefer ha-'Iqqarim (Book of Principles), sought to defend Judaism against Christian attacks by laying out the basic presuppositions of the Mosaic law.

Albo's theology belongs to a tradition of theorizing going back to Maimonides in the twelfth century. But his approach was grounded in a non-Maimonidean moral psychology. Responding to the Aristotelian intellectualism of the Maimonidean philosophy, which held true belief to be essential to human virtue and salvation, Albo focused on practice, fulfilment of the commandments. His act-centred view, grounded in the premise that beliefs cannot be commanded, allowed for a certain latitude in faith, which had both intra-communal and inter-communal advantages. If controversial doctrines such as ex nihilo *creation could be made less prominent, acrimonious internal debates could be avoided, and the community could be somewhat less exposed to external attack.*

1 Life

We know little about Joseph Albo's life. He lived in Christian Spain, in Castile and Aragon, at a particularly troubled time for Spanish Jewry, a time of religious persecution and forced conversion. There were terrible anti-Jewish massacres in 1391 at Barcelona and elsewhere. Albo represented the Jewish community of Daroca (near Saragossa) at the last of the great public disputations, held in 1413-14, in Tortosa. His Christian adversary, Geronimo de Santa Fé, was born a Jew, Joshua Lorki. This *disputatio*, like others held since the mid-thirteenth century, was not a free-ranging debate between disinterested parties, but an occasion staged by the Church for the public humiliation of Jews. In representing the Jewish side, Albo was defending both Judaism and the Jewish people against a prejudiced, or even predetermined, response. Arraying his arguments in what seemed a hopeless cause, he wrote the *Sefer ha-'Iqqarim (Book of Principles)* in four parts, discussing divine law in general and the Mosaic law in particular. The book was hugely popular in Jewish circles and has never been out of print. Translated into Latin, it was highly esteemed by such Christian thinkers as Grotius for its clear differentiation of natural, conventional, and divine law, a distinction Albo probably learned from his study of Aquinas.

2 Thought and action

Albo's contribution to legal theory finds its context in his engagement with the problems of Jewish dogmatics, the quest to define the fundamental principles of Judaism and to lay out the further principles that derive from them. Maimonides was the first thinker in the rabbinic tradition to posit a set of fundamental doctrines or beliefs incumbent upon every Jew. For Maimonides, these beliefs, thirteen in number, included God's existence, unity, incorporeality and ontic primacy, the uniqueness and irreplaceability of the Mosaic revelation, the coming of a messianic redemption, and bodily resurrection. Belief in these ideas defined the Jew and were necessary conditions for the salvation promised to Israel in the world to come. Disbelief in these principles amounted to heresy.

Superficially, Albo may seem to follow Maimonides in outlining a set of necessary beliefs, differing only in that he distinguishes fundamental from derivative principles. But appearances here are deceptive, for Albo does not adopt Maimonides' premise that beliefs are criteria for identifying a Jew or that nonbelief amounts to heresy. This is not to say that Albo thinks an atheist has a place in the community, but for him, like his immediate predecessors Simeon ben Tzemach Duran and Hasdai Crescas, Albo's teacher, the focus is on acts. Acts make us what we are, and acts, unlike beliefs, can be commanded. The point is not that one can believe just anything and still be a member of the community, but that membership is consequent on acts, fulfilment of the commandments. Albo's account of belief is shorn of the normative weighting of Maimonides' account. He does not lay out what must be believed but treats the articles of faith as the presuppositions underpinning the practical life of a professing Jew.

Albo's act-centred theology rests on an original approach to moral psychology. Maimonides, like Aristotle, thought of human virtue as an expression of reason. It requires a rationally attuned disposition and knowledge suitable to one's endeavour. But this account came under fire in the centuries between Maimonides and Albo. Forced to defend themselves in disputations in terms that would have weight with their opponents, Jews in the fourteenth century came under the influence of Christian scholasticism, which was moving increasingly away from

its Aristotelian and Averroist moorings. With the condemnation of Averroism in 1277, Aristotelian moral psychology gave way to a greater voluntarism like that found in Duns Scotus and William of Ockham. Jewish philosophy mirrored the Christian trend, notably in the case of Crescas. The rejection of intellectualism meant that human felicity was no longer seen as the fruit of intellectual perfection or true belief. As a result, for Albo, as for Crescas, access to salvation takes on a less intellectually elitist cast. A further result of Albo's approach to understanding the principles of Judaism - and indeed of any revealed religion, once these principles are no longer conceived as mandatory dogmas but as necessary presuppositions - is the opening of the way for objective inquiry into revealed religion generally and Judaism in particular.

The *Sefer ha-'Iqqarim* has four parts, an introduction and three books. Its stated purpose is 'to explain those principles which pertain to a divine law in general, the principles without which a divine law cannot be conceived'. Before outlining these principles, Albo explains the need for such a law, relying on arguments as old as Jewish (and Islamic) philosophy itself: 'it is not possible for the human intellect alone to attain a proper knowledge of the true and the good. For human reason is not capable of comprehending things as they are in reality…. There must therefore be something higher than the human intellect by means of which the good can be defined and the truth comprehended so as to leave no doubt. This can be done only through divine guidance. It is incumbent, then, on every person to know that among all laws there is one divine law which gives this guidance' (I.1-2). Only a divine law can take us beyond the objectives of merely utilitarian legislation, which seeks no more than social order and stability, and can open up to us the possibility of our genuine felicity.

All divine law, Albo argues, presupposes three principles: (1) that God exists, (2) that the Torah is divine, and (3) that reward and punishment, both now and in the hereafter, attend our actions. Why does Albo urge that *all* divine legislation presupposes belief in the Torah? Albo's response would be that in fact all monotheistic faiths agree on the divine origin of the Mosaic law, although Christianity and Islam hold that the Mosaic legislation has been superseded. The polemical context that provides the background of Albo's writing is strikingly clear here. But it is important to Albo (partly because of his defensive posture) not to confine his argument suppositiously to parochial assumptions. His intention is to move from the more general to the more particular by specifying the implications of the basic principles. Thus, from his three general or 'root' principles (*'iqqarim*), he derives eight derivative principles, also called roots (*shorashim*). The denial of any one of these latter principles 'is tantamount to a denial of the fundamental principle from which it is derived'. From God's existence, he derives God's (1) unity, (2) incorporeality, (3) atemporality, and (4) perfection. From the divine origins of the Torah, he derives (5) God's wisdom, (6) the possibility of prophecy, and (7) the authenticity of the mission of the historic prophets. From the reality of reward and punishment, he derives (8) divine providence and knowledge of human actions and events.

Even these derivative principles are quite generic. None is distinctive to Judaism. But their derivation from the first tier allows Albo to deny the consistency of Christianity: if God's unity follows from his existence, Catholics should deny the Trinity, lest their view call into question the very existence of God.

A third set of beliefs are called branches (*'anafim*). They are (1) *creatio ex nihilo*, (2) the superiority of the Mosaic prophecy, (3) the irreplaceability of the Torah, (4) the possibility for a human being to attain perfection by fulfilling even a single commandment, (5) resurrection of the body, and (6) messianic redemption. For Albo these six doctrines are not derived from the others but are peculiar to Judaism, traditionally taught and accepted. A divine law can exist without these six 'branches'.

The secondary status of these beliefs seems to allow Albo to defuse the traditionally acrimonious intra-communal debates over such issues as *creatio ex nihilo*. This is not to say that Albo supposes that the tradition does not teach *creatio ex nihilo*, but it does mean that he does not take God's perfection to be inconsistent with the creation of the world from pre-existent matter, as Plato was held to have taught. Those who held to the Platonic doctrine of *formatio mundi* would not, as a result, have to be construed as implying the denial of a fundamental principle - as Trinitarians were. Again, the downplaying of messianism may reflect a reaction against the centrality of that theme in Christianity. Albo's special principles, then, both ground the particularity of Judaism and mitigate the intensity of intra-confessional theological debates.

3 Contribution to Jewish philosophy

Albo comes late in the history of Jewish dogmatics. His contribution lies in the architectonic he constructs for

Jewish faith and practice. Maimonides and others present the principles of Judaism on an equal footing. Albo offers a hierarchically graded, logically structured schematization. The structure is designed both for internal strength and for defence against external attack. The scheme allows Albo to contextualize Judaism generically as a divine law but also to render less vexed the internal debates over the niceties of those beliefs that are distinctive to Judaism. Albo's treatment of core beliefs not as axioms but more as themes whose appeal is traditional, but whose nexus to the canonical axioms is not one of entailment, is appreciated by contemporary students of Judaism for fostering a certain openness with respect to belief without departing from the larger axiomatic structure that Albo used in helping to define his faith and to defend it.

See also: Aristotelianism, medieval; Averroism, Jewish; Crescas, H.; Maimonides, M.

DANIEL H. FRANK

List of works

Albo, Joseph (*c*.1425) *Sefer ha-'Iqqarim (Book of Principles)*, ed. and trans. I. Husik, *Sefer ha-'Iqqarim: Book of Roots*, Philadelphia, PA: Jewish Publications Society, 1946, 5 vols.(The standard critical edition, with an English translation and notes.)

References and further reading

Guttmann, J. (1933) *Die Philosophie des Judentums*, trans. D. Silverman, *Philosophies of Judaism*, New York: Schocken Books, 1973, esp. 275-86.(An account of Albo and his relation to his immediate predecessors, Hasdai Crescas and Simeon ben Tzemach Duran.)

Husik, I. (1916) *A History of Medieval Jewish Philosophy*, New York: Macmillan; repr. New York: Atheneum, 1976, esp. 406-27.(A brief account in a standard history, treating Albo as the last Jewish philosopher, not merely the last medieval Jewish philosopher.)

Hyman, A. (1967) 'Maimonides' "Thirteen Principles"', in A. Altmann (ed.) *Jewish Medieval and Renaissance Studies*, Cambridge, MA: Harvard University Press, 119-44.(A presentation of Maimonides' dogmatics with discussion of various modern accounts of how best to interpret it.)

Kellner, M. (1986) *Dogma in Medieval Jewish Thought: From Maimonides to Abravanel*, Oxford: Oxford University Press, esp. 140-56.(The most recent full-length study of medieval Jewish dogmatics. Philosophically acute.)

Manekin, C. (1997) 'Hebrew Philosophy in the Fourteenth and Fifteenth Centuries: An Overview', in D.H. Frank and O. Leaman (eds*) History of Jewish Philosophy*, London and New York: Routledge, 350-78.(A fairly detailed discussion of major philosophical topics in the post-Maimonidean period.)

Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Cambridge: Cambridge University Press, esp. 374-81.(A brief account of Albo and his work in the context of Jewish-Christian polemics in late fourteenth- and early fifteenth-century Spain.)

Alchemy

*Alchemy is the quest for an agent of material perfection, produced through a creative activity (*opus*), in which humans and nature collaborate. It exists in many cultures (China, India, Islam; in the Western world since Hellenistic times) under different specifications: aiming at the production of gold and/or other perfect substances from baser ones, or of the elixir that prolongs life, or even of life itself. Because of its purpose, the alchemists' quest is always strictly linked to the religious doctrine of redemption current in each civilization where alchemy is practised.*

In the Western world alchemy presented itself at its advent as a sacred art. But when, after a long detour via Byzantium and Islamic culture, it came back again to Europe in the twelfth century, adepts designated themselves philosophers. Since then alchemy has confronted natural philosophy for several centuries.

In contemporary thought the memory of alchemy was scarcely regarded, save as protochemistry or as a branch of esotericism, until interest in it was revived by C.G. Jung. Recent research is increasingly showing the complexity of alchemy and its multiple relation to Western thought.

1 Name and definition

The name 'alchemy' appeared in Islamic culture, whence it passed to Latin. It evolved (apparently) from the Greek 'chemèia' (art of melting metals) or 'chymos' (juice). An alternative etymology, supported by the Hermetic tradition, indicates 'kemi' (black clay), the ancient name of Egypt, pointing to the mythological link with the god Hermes-Thoth.

Initially alchemy denoted both the art and its product; this latter use, however, is rare in the Western tradition. As a name indicating the art or *opus*, its meaning varies depending on the period referred to: originally it designated the practical and theoretical search for transmutation, whereas in contemporary esotericism it indicates the concrete achievement(s) associated with a pre-eminently spiritual quest.

Accordingly, the decision about what is an alchemical text or an alchemical image may differ: we distinguish a historical, an esoteric and a psychological approach. Historians consider written tradition (manuscript and printed texts and images) the one and only testimony of a doctrine evolving in time. For esotericists, this same tradition is nothing but the surface of a secret, immutable knowledge, often deliberately disguising its truth. For depth psychology alchemy encompasses virtually every kind of symbolic production.

The definition given by H.J. Sheppard (1986), currently the most widely accepted, takes into account all approaches: ' Alchemy is the art of liberating parts of the Cosmos from temporal existence and achieving perfection which, for metals, is gold, and for man, longevity, then immortality and, finally, redemption'.

2 Epistemological structure

The association of practice and theory characterizes alchemy from its very beginning and distinguishes it from other symbolic lore. Archaic metallurgy, in its connection with religious rites, is generally considered the cradle of alchemy: this opinion has accounted for the religious and even mystical elements in alchemy since the work of Zosimus of Panopolis (*c*.3-4 AD). Yet the earliest acknowledged alchemical text (Bolus of Mende, pseud. Democritus, *Physikà kai mystikà c*.1 AD) shows that just those metallurgical practices imbued with religious significance were the basis on which a theory of matter was beginning to be built: a theory derived from practice, not the contrary.

Indeed, the dependence of theory on practice marks the whole history of alchemy. For its theoretical content, Western alchemy was called 'the child of Greek philosophy', but we cannot accept such a genealogy without remembering that there was also another 'parent', namely, concrete work on matter. In the Middle Ages, as practical alchemy interacted with the development of techniques and craftmanships (metallurgy, goldsmith, dyeing, pharmacology), it was first considered an 'ars mechanica', but soon its theoretical meaning became clear to philosophers like Albert the Great and Roger Bacon, and the close connection between alchemical practice and religious-philosophical speculation continued with Paracelsus and the Rosicrucians, with Newton and Goethe (see Paracelsus §3).

Therefore we cannot speak of alchemy proper where we find only a practice, be it metallurgy, distillation or whatever else: the dyeing recipes of medieval painters, or distillation among Renaissance physicians. However, we must be equally careful not to speak of alchemy too readily whenever alchemical symbolism is used for other purposes: for example, by mystics.

The connection between practice and theory accounts for, and delimits, the contribution of alchemy to the birth of chemistry (see Chemistry, philosophical aspects of §1). Like ancient artisans, who owned secret techniques transmitted through apprenticeship, alchemists were secretely initiated to the *opus*. The secrecy of alchemy is a major point of divergence from chemistry, which, like all modern sciences, is characterized by public discussion and teaching. Moreover, although alchemists for centuries worked with metals and minerals, invented techniques (for example, distillation), designed and used laboratory apparatus that chemists would inherit, they never relinquished their original religious attitude; as a consequence of this, theoretical developments were radically different from those of modern chemistry. Basically, for alchemists, matter was no inert object but the body of their own Mother Nature. So there is no epistemological continuity between alchemy and chemistry; protochemical features may be disentangled only a posteriori from a doctrinal whole owing its orientation to totally different ideas and purposes.

3 Features of the alchemical literature

The initiatory character of alchemical teaching accounts for the most striking feature of the language of alchemists, that is, the use of a rich symbolism. Metaphorical names for substances and processes were used from the beginning, an attitude reinforced by Arab alchemists and complicated by the obvious difficulties of Latin translators. The use of metaphors met the need for secrecy and facilitated the merging of operative and religious meaning; but it also prevented the creation of a technical vocabulary (another major difference from chemistry) and fostered the transformation of alchemy into an occult art.

This retreat into the occult, accompanied by the development of alchemical imagery, began at the end of the Middle Ages, when metallurgical alchemy was defeated by the denunciation of alchemists as forgers and the idea of the medical elixir began to be associated with the prophetic and visionary mood of the Spirituals. So, while many medieval texts were written in a clear language and even, sometimes, in a truly philosophical style, the number of obscure writings playing with symbols and visions increased steadily from the fourteenth century. Later alchemists went further, explicitly linking their art to ancient mythology and eventually wholly replacing words with images.

Thus the main difference inside alchemical literature is between clear and obscure texts. Various genera belong to the first group: recipes, practical treatises, theory and practice texts, commentaries, veritable summae; rarely does their 'clear' character match our modern demands for clarity. Obscure texts comprise mainly visions, riddles, and poems.

A relevant feature of the alchemical literature, since its very origin, is its pseudoepigraphical character, often connected to the creation of legends concerning the supposed authors of alchemical writings. By means of pseudoepigraphy alchemists clearly attempted to enrol themselves in the philosophical tradition, albeit awkwardly. Texts were attributed to pagan gods, mythological and biblical figures, ancient and medieval philosophers. Such attributions assured secrecy, while raising the prestige of writings of obscure authors; they might even be a subtle indication of affiliation.

4 Alchemical doctrines

The basic idea of alchemy is the identity of nature and first matter as a dynamic unity: elements can pass one into another, in a circular movement that alchemists reproduced in their vessels. No theory of natural *loci* (low and high are interchangeable, according to the *Tabula smaragdina c*.9 AD), no dualism of matter and spirit exists, as first matter is the all-embracing source of change. The alchemist, who can obtain first matter by means of the dissolution of natural bodies, is almost a new creator who makes a new reality come out of the artificially produced chaos 'putting nature into nature', that is, cultivating the seeds of perfection existing in nature (perfect metals) according to natural rules, and 'awaiting nature's time of delivery'. (R. Llull, *Testamentum c*.14 AD).

This structure is first seen as continuity inside the inanimate field of metals, and as analogy between metals and planets: all metals are nothing but imperfect gold (like embryos at various stages), and the alchemist accomplishes nature's work outside the womb of earth in a shortened time, possibly within an astrological framework. Some alchemists viewed the process as a victory over nature and time, foreshadowing the Promethean developments of modern science and technology: there are some hints that medieval theologians rebuked alchemy for this claim. Yet the relation between alchemical art and nature's work was generally considered in a more subtle and complex way, especially in the theoretical attempts made by fourteenth-century alchemists who developed the idea of elixir. Continuity from inanimate matter to human beings was explicitly or implicitly affirmed, and the alchemists were conscious of themselves as a part of the matter/nature that they manipulated in order to perfect, not to dominate. This consciousness preserved their attitude of religious reverence for nature, whose abandonment was a major feature of modern science.

5 Alchemy and Western philosophy

Discussing matter and its transformations, alchemists encountered philosophical themes from the beginning. There have even been attempts to trace back alchemy to Aristotle's idea of change in material substances (see Matter §1), but actually alchemy (practice plus theory) was not yet born. In later Antiquity an especial relationship existed between alchemy and Hermetic thought (see Hermetism): the unity of first matter, the principle of sympathy, the doctrine of occult virtues, all are behind Bolus' axiom that 'Nature is charmed by nature, nature prevails over nature, nature rules nature'. The Stoic doctrine of *pneuma* lingers on in the search for material essences through distillation, a practice that goes back to Maria the Jewess (*c*.3 AD) (see Stoicism §4).

Medieval developments were considerable, as scholastic philosophers and alchemists compared alchemy to the Aristotelian philosophical concepts. According to Albert the Great (*De mineralibus c*.13 AD) alchemy helped to complete the Aristotelian science of metals. Roger Bacon showed a broader concern, viewing alchemy as the general theory of generation and corruption of all natural beings. Some alchemists even tried to translate into Aristotelian language their experience, identifying form with the purest and thinnest substance (quintessence) resulting from sublimation or distillation. How much of Stoic natural philosophy intermingled with Aristotelian ideas in this attempt is unclear. Form was also identified with the soul, so that all material bodies, metals included, were considered endowed with a soul; body and soul were kept united by spirit, an idea which the alchemists could also find in medical literature, and developed into that of the universal 'medium' that gives unity and life to the created world.

The Hermetic elements had never disappeared from Western alchemy, as the central role of the *Tabula smaragdina* shows; during the Renaissance, they became prevalent. Alchemical doctrines were known to virtually every Renaissance philosopher, discussed by most of them, accepted by many. The relation between Renaissance Platonism and alchemical thought might be considered afresh, as alchemy is a project to obtain on earth the stability and perfection that characterize the Platonic world of ideas, manipulating the universal spirit that mediates between matter and the divine world. The most significant development, however, can be found in Paracelsus, whose idea of 'making visible the invisible' rests on the alchemical assumption that the quintessence of material bodies can be revealed through the *opus*. So, Paracelsian alchemy aimed at revealing the secret of life and putting it to work for the spiritual and bodily health of humans. More definitely, the idea of an alchemical remedy or elixir crystallized in that of potable gold, which interested Ficino and Francis Bacon among others.

In the seventeenth century, Francis Mercurius van Helmont turned the idea of the universal spirit into that of alkahest, the basis for subsequent chemical developments which ultimately led to the discovery of oxygen. Newton's alchemical and cosmological speculation about the creative, non-mechanical spirit animating matter is at the core of the debate about the role of Hermetism in the Scientific Revolution. The re-emergence of Stoic ideas concerning first matter and mixed bodies in seventeenth-century alchemy has recently been considered to establish it as part of the normal science of that epoch. Even after the birth of modern chemistry, alchemy maintained its appeal to speculative spirits: Goethe apart, we have the clear instance of the subsequent development of *Naturphilosophie* in nineteenth-century Germany (see Naturphilosophie), with alchemical doctrines flowing into the mainstream of vitalism (see Vitalism). More surprising perhaps is to find alchemy defined in the *Encyclopédie* as chemistry brought to the highest degree of perfection and therefore capable of operating marvellous effects, 'la chimie sublime, la chimie par excellence'.

6 Alchemy and the present

Nineteenth-century scholarly research on protochemistry overlapped with the latest development of esoterical alchemy (hyperchemistry). Historians of chemistry judged alchemy a mix of positive empirical data about chemical matter with obscure mystical speculation. An echo of their attitude is still felt in the consideration of alchemy as an error in the history of science, indeed, according to Bachelard, 'the first error' in the scientific approach to the problem of matter.

Alchemy attracted the attention of C.G. Jung as a historical testimony of the dynamics of the unconscious. Jung's deep study of alchemy led him first to conceive of it as the projection upon matter of the unconscious tendency to individuation; but he also saw in alchemy the expression of a more complex relation between humanity and nature, where matter is recognized as the feminine counterpart of the divine, and human knowledge is fostered by the very light of nature (a Paracelsian idea), comparable to the light of Revelation. Thus Jung gave a positive value to the link between religious attitude and empirical research in alchemy. On the other hand, the idea of the alchemists as forerunners of the modern ideal of the scientist who overcomes nature and time is at the core of M. Eliade's (1956) view of alchemy as an intermediate stage between archaic metallurgy and modern technology.

Recent proposals from within French esotericism bear on epistemology and aesthetics: A. Faivre's (1971a) conception of the non-dualistic logic of alchemy links it to the most advanced results in the epistemology of physics, while F. Bonardel (1993) defines the alchemist's attitude as taking charge of the created world. She opposes the Promethean-Faustian view of alchemy, and deplores its gradual fall from the original position of the art of Hermes, proposing its identification with poetry. Another contemporary way of stressing the Hermetic meaning of alchemy is that of focusing on its figurative symbolism, not only in the images linked to alchemical texts or in the alchemical interpretation of artists of the past, but even in defining the creative process of contemporary art as alchemy.

MICHELA PEREIRA

References and further reading

Bachelard, G. (1937) *La formation de l' esprit scientifique*, Paris: Vrin.(Discusses alchemy in the context of scientific epistemology; in his further writings, known as 'quadrilogy of the elements', Bachelard approached alchemy psychologically and poetically.)

Bonardel, F. (1993) *La philosophie de l'alchimie. Grand Oeuvre et Modernité*, Paris: Presses Universitaires de France.(On alchemy, esotericism and contemporary culture.)

Coudert, A. (1980) *The Philosopher's Stone*, Boulder, CO: Shambala Publications.(A historical survey, not strictly limited to the Western world, paying attention to the religious contents of alchemy.)

Crosland, M.P. (1962) *Historical Studies in the Language of Chemistry*, Cambridge, MA: Harvard University Press.(Discusses the problems of the alchemical language and its relation with chemistry.)

Dobbs, B.J. (1975) *The Hunting of the Green Lion. The Foundation of Newton's Alchemy*, Cambridge: Cambridge University Press.(Connecting alchemical mentality and research to scientific creativity.)

Eliade, M. (1956) *Forgerons et alchimistes*, Paris: Flammarion.(The origin of alchemy and its relation with ancient metallurgy and religion.)

Faivre, A. (1971a) 'Alchimie occidentale et logique Aristotélicienne', *Revue d'Histoire des Religions*, new series, XX: 105-10.(The triadic structure of alchemy versus the dualism of matter and form.)

Faivre, A. (1971b) 'Pour un approche figurative de l'alchimie', *Annales ESC* XXVI: 841-53.(The logical structure of the alchemical *opus*.)

Festugière, A.J. (1942) *La révélation d' Hermès Trismegiste. I. L'Astrologie et les sciences occultes*, Paris: Gabalda.(Reprinted by Les Belles Lettres, 1980; chapter VII presents the texts referring to Hermetic alchemy in the Hellenistic age.)

Halleux, R. (1979) *Les textes alchimiques*, Turnhout: Brepols.(A historical survey of the medieval alchemical literature, with bibliography.)

Hopkins, A.J. (1934) *Alchemy the Child of Greek Philosophy*, New York: Columbia University Press.(A study about the relation of alchemy and classical philosophy.)

Joly, B. (1992) *Rationalité de l'alchimie au XVII^e siècle*, Paris: Vrin.(The Stoic doctrine of matter and the rational ground of the alchemical *opus*.)

Jung, C.G. (1944) *Psychologie und Alchemie*, Olten: Walter Verlag; trans. R. Hull, *The Collected Works of C.G. Jung*, vol. XII, Princeton, NJ: Princeton University Press, and London: Routledge, 1953-91.(Alchemy as

historical parallel to psychological individuation.)

Jung, C.G. (1954-6) *Mysterium Coniunctionis*, Olten: Walter Verlag; trans. R. Hull, *The Collected Works of C.G. Jung*, vol. XIV, Princeton, NJ: Princeton University Press, and London: Routledge, 1953-91.(Jung's basic studies give a psychological interpretation of alchemy, and present its historical development as providing a link between ancient Gnosticism and the modern psychology of the unconscious.)

Kren, C. (1990) *Alchemy in Europe. A Guide to Research*, New York and London: Garland Publishing Company. (Bibliographical survey.)

Multhauf, R.P. (1966) *The Origins of Chemistry*, London: Oldbourne.(A history of alchemy as science of matter, with bibliography.)

Newman, W.R. (1989) 'Technology and Alchemical Debate in the Late Middle Ages', *Isis* LXXX: 423-45. (Artificial versus natural in the writings of alchemists and theologians.)

Obrist, B. (1982) *Les débuts de l'imagerie alchimique*, Paris: Le Sycomore.(A historical approach to alchemical iconography.)

Sheppard, H.J. (1986) 'European Alchemy in the Context of a Universal Definition', in Ch. Meinel (ed.) *Die Alchemie in der europäischen Kultur - und Wissenschaftsgeschichte*, Wiesbaden: Harrassowitz.(A general definition of alchemy based on comparative research.)

Alcinous (*c.* 2nd century AD)

Long misidentified with the Middle Platonist philosopher Albinus, Alcinous is author of a 'handbook of Platonism', which gives a good survey of Platonist doctrine as it was understood in the second century AD. The work covers logic, physics and ethics, and shows considerable influence from both Stoicism and Aristotelianism, in both terminology and doctrine, while remaining in all essentials Platonic.

A Middle Platonist philosopher, Alcinous was the author of the *Didaskalikos tōn Platōnos dogmatōn* or 'instruction manual of Platonic doctrine'. He was long identified with the second-century Platonist Albinus, author of an extant introduction to Plato's dialogues, but this identification has been recently abandoned. In consequence, we know strictly nothing of the life or times of this author, nor even whether the name may not be pseudonym. None the less, he seems to fit best within the environment of second-century AD Platonism.

The work *Didaskalikos*, purporting to be a summary of Plato's doctrines, possesses its real value as a summation of the doctrines of at least one school of later Platonists. It presents a concise survey of Platonist doctrine, in thirty-six chapters. After three introductory chapters, concerned respectively with the definitions of philosophy, the laying down of requirements for the successful philosopher and an enumeration of the 'parts' of philosophy (logic, physics and ethics), Alcinous proceeds to take these topics in order, beginning with logic in chapters 4 to 6. Chapters 7 to 26 deal with 'physics', and comprise both an account of first principles, 'matter, form and god' (7-11), and one of the physical world; the latter is very closely based on the *Timaeus* (12-26), although Alcinous holds to a non-literal interpretation of the demiurgic creation myth. The final chapters, 27 to 34 are concerned with ethics.

A good deal of both Peripatetic and Stoic doctrine and formulation is incorporated into the exposition, although normally supported by the adducing of Platonic texts. However, all of Peripatetic logic is claimed for Plato, as well as such Aristotelian ethical principles as the mean, and *metriopatheia* (moderation of the passions). On the other hand, such Stoic concepts as the self-sufficiency of virtue and the concepts of *euphyiai*, or 'good natural dispositions', and *prokopē*, or 'moral progress', are also adopted. Alcinous' position on free will and determinism (ch. 26) also owes a good deal to Stoic theorizing, although Stoic determinism is firmly rejected. On the whole, Alcinous inclines to the Peripatetic rather than to the Stoic wing of Middle Platonism.

Distinctive features of his doctrine are his theology (ch. 10) and his views on the reasons for the embodiment of the soul (ch. 25). In Chapter 10 we find a hierarchy of principles set out, consisting of a supreme god, who is a transcendent intellect, an intellect of the world-soul, and the world-soul itself, which seems to be only rational by participation, after being 'roused up' by the supreme god. This is comparable to other Middle Platonic systems observable in Plutarch and Numenius, but the relationship between the principles is distinctive to Alcinous.

In Chapter 25 we find an interesting list of possible reasons for embodiment, giving evidence of considerable debate on this question, and a theory of three faculties of disembodied souls not found elsewhere. In general, a frustrating aspect of the *Didaskalikos* is the evidence it gives of active philosophical debate within Platonism in this period, while not preserving the actual arguments; but that is inherent in the nature of the work.

See also: Platonism, Early and Middle

JOHN DILLON

List of works

Alcinous (*c.* 2nd century AD) *Didaskalikos tōn Platōnos dogmatōn*, in J. Whittaker (ed.) *Alcinoos, Enseignement des Doctrines de Platon (Alcinous, Teachings on Plato's Doctrines)*, Paris: Les Belles Lettres, 1990; ed. J. Dillon, *Alcinous, The Handbook of Platonism*, Oxford: Clarendon Press, 1993.(The former offers text, French translation and notes, the latter an English translation, with introduction and philosophical commentary.)

References and further reading

Dillon, J. (1977) *The Middle Platonists*, London: Duckworth, 2nd edn 1996, ch. 6.(Here Alcinous is still identified as Albinus; offers a detailed comparison with the handbook of Apuleius.)

Witt, R.E. (1937) *Albinus and the History of Middle Platonism*, Cambridge: Cambridge University Press.(Still useful, but with reservations, in that it connects ’Albinus’ too closely with Antiochus.)

Alcmaeon (*c.* early to mid 5th century BC)

Alcmaeon of Croton was a Greek thinker with philosophical and medical interests. His work focused on the nature of man. Health was the outcome of 'equal rights' between, for example, hot and cold, moist and dry, disease that of the 'monarchy' of one of them. 'Passages' linked the sense organs to the brain, which Alcmaeon took to be the seat of sensation and understanding. Plato followed him in this view, as also in his proof of the immortality of the soul from its continual motion.

Was Alcmaeon a philosopher or a doctor? The interests revealed in our information about his views, some controversial evidence that he practised dissection and the existence in the southern Italian city of Croton of a medical tradition, famous from the sixth century BC, all suggest a doctor. But Croton was also the centre of Pythagoras' activities; and Aristotle and Theophrastus present Alcmaeon as a philosopher. The book Alcmaeon wrote was probably 'on the nature of man' (as later authors might have titled it). Its opening words survive. They include a dedication to three persons believed in later antiquity to be adherents of Pythagoreanism, although Aristotle implies that, despite some similarities, Alcmaeon was not a Pythagorean himself. In sum, Alcmaeon has the look of an independent thinker who responded creatively to a number of the intellectual currents of his time and place.

The doxographer Aëtius gives a striking exposition of Alcmaeon's theory of health and disease:

> Alcmaeon maintains that what sustains health is the 'equal rights' of the powers, moist and dry, cold and hot, bitter and sweet and the rest, while 'monarchy' among them is what causes disease; for the monarchy of either one of a pair is destructive.
>
> (fr. 4)

The use of political metaphor was presumably Alcmaeon's own. Explanatory appeal to opposite powers is a pervasive and fundamental feature of Presocratic thought, but it was here that Aristotle perceived a particular similarity - if also a dissimilarity - with the Pythagoreans' systematic listing of key contrarieties (see Pythagoreanism §2). He ascribes to Alcmaeon the generalizing remark that most things to do with human beings come in twos, yet apparently found in him no attempt to construct a single system of opposites (*Metaphysics* I 5).

Alcmaeon is credited with three notable contributions to the psychology and physiology of the senses. First, humans differ from other animals in that they alone have understanding as well as sense perception (A5). Most Presocratic thinkers treat animals, too, as exercising intelligence. Second, the senses, as Alcmaeon argued case by case, are all connected by 'passages' to the brain, which is conceived of as the seat of sensation (A5). Other Presocratics (for example, Empedocles) identify this seat with the heart. Third, he is said to have been the first person to have performed an *exsectio* (A10). Probably this Latin term means not that he initiated a general medical or scientific practice of dissection, but merely that once he cut out the eyeball of a dead animal in an attempt to verify his theory of 'passages', and thereby revealed the existence of what we call the optic nerve. Some scholars reject the authority of the report. If accepted it confirms the importance Alcmaeon says he attaches to using signs to interpret what is unapparent (fr. 1): which is not, however, the same as subscribing to a methodology of rigorous empiricism.

What sets Alcmaeon apart from the general run of early medical writers is his willingness to theorize about the soul. Aristotle reports his view in these terms:

> He says that the soul is immortal owing to its similarity to the immortals, and that this is true of it because it is always in motion - for everything divine is always in continuous motion: moon, sun, the stars and the whole heaven.
>
> (*On the Soul* I 2)

One testimony suggests in addition that Alcmaeon derived the continual motion of the soul from the capacity of living things for *self*-movement (A12). Another assures us that according to Alcmaeon humans die because they cannot join the beginning to the end (fr. 2). This obscure saying perhaps indicates the way the irreversible process of ageing makes human beings - as opposed to their souls - unlike the heavenly bodies, which continue in motion

by everlasting repetition of their revolution in the heavens. This group of texts furnishes evidence of the first attempt we know of to *argue* the Pythagorean doctrine of the soul's immortality.

Alcmaeon's argument for the immortality of the soul is clearly what supplied the inspiration for Plato's proof in the *Phaedrus* (245c). His identification of the brain as the seat of sensation was also accepted by Plato, after previous development by Diogenes of Apollonia (§3) and in the Hippocratic treatise *On the Sacred Disease*. Aristotle notoriously took a different view, but even he borrowed from Alcmaeon's ideas; for example, on the cause of sleep. As for Alcmaeon's theory of health and disease, there are reflections of it in various early Hippocratic writings such as *On Ancient Medicine*.

MALCOLM SCHOFIELD

References and further reading

Alcmaeon (*c.* early to mid 5th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 1952, 6th edn, vol. 1, 210-16.(The standard collection of the ancient sources, both fragments and testimonia, the latter designated by 'A'; includes Greek texts with translations in German.)

Aristotle (*c.* mid 4th century BC) *On the Soul (De Anima)*, ed. R.D. Hicks, Cambridge: Cambridge University Press, 1907.(Source of Aristotle's views on Alcmaeon quoted above.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press.(The most detailed and comprehensive English-language history of early Greek thought; the full treatment of Alcmaeon, in volume 1 pages 341-59, is the best available account of his philosophy in English; translates most of the main texts that are relevant.)

Lloyd, G.E.R. (1975) 'Alcmaeon and the Early History of Dissection', *Sudhoffs Archiv* 59: 113-47, repr. in *Methods and Problems in Greek Science*, Cambridge: Cambridge University Press, 1991, 164-98.(A full and searching analysis of the problem, including a substantial bibliography; the reprint includes a new introduction.)

al-Dawani, Jalal al-Din (1426-1502)

Jalal al-Din al-Dawani was a prominent philosopher and theologian from Shiraz, who came to the note of Western scholars through an English translation of his ethical treatise theAkhlaq-e Jalali (Jalalean Ethics), published in 1839. Although the larger part of his work written in Arabic has been little studied, he did write extensively and engaged in a famous and lengthy philosophical dispute with another leading philosopher, Sadr al-Din al-Dashtaki. His metaphysical views were quoted, and refuted, by Mulla Sadra. He emerges as a thinker who combined elements of illuminationist and Peripatetic philosophy (and possibly also interests in Ibn al-'Arabi) to confront theological, ethical, political and mystical concerns.

Jalal al-Din Muhammad ibn As'ad al-Dawani (or Dawwani) was born near Kazarun, southern Iran, in the village of Davan in AH 830 (AD 1426). He first studied there with his father, who had been taught by the Sayyid al-Sharif al-Jurjani (d. AH 816/AD 1413), before going on to further and complete his education in philosophy, theology and law in Shiraz. In common with the other leading religious scholars of his time and place, he was directly caught up in the turbulent politics of Iran in the second half of the ninth century AH (fifteenth century AD). He was inducted into various religious offices, and many of his works were dedicated to Aq Qoyunlu and other Timurid rulers and princes. He also achieved fame as a teacher in the Begum *madrasa* (Dar al-Aytam) in Shiraz. The question of his religious allegiance, whether Sunni or Shi'i (he wrote theological works of both persuasions), has always been the subject of debate and of many fanciful stories, but it may be of comparatively slight significance given the situation in the Iran of his time, which was marked by a Sunnism with a strong Shi'i colouring. He died in AH 908/AD 1502 near Kazarun, a year or so before the Safavid capture of Shiraz, and is buried in his home town.

Al-Dawani first came to the attention of Western scholarship through the 1839 English translation of his Persian ethical work, the *Akhlaq-e Jalali (Jalalean Ethics)*, more correctly known under its original title of *Lawami' al-ishraq fi makarim al-akhlaq (Lustres of Illumination on the Noble Virtues)*. Al-Dawani's text marks a third stage in the development of the ethical strand of writing begun by Ibn Miskawayh with the *Tahdhib al-akhlaq (Cultivation of Morals)* and continued by Nasir al-Din al-Tusi with his *Akhlaq-e Nasiri (Nasirean Ethics)*, on which al-Dawani's work is closely modelled. Al-Dawani retains al-Tusi's division of the text into three sections - ethics, economics and politics - and subdivides his work similarly, although significantly he entirely omits al-Tusi's theoretical first section of the ethics. The title, *Lawami' al-ishraq (Lustres of Illumination)*, may indicate the author's *ishraqi* (illuminationist) and mystical concerns. The political content of the work has been of some interest to historians, as regards both its descriptions of the ideal ruler and the titles used for its dedicatee, the Aq Qoyunlu Uzun Hasan, which betray a possible *ishraqi* influence and seem to foreshadow the extravagant claims of Isma'il, the first Safavid monarch of Iran.

The *Akhlaq-e Jalali* is generally acknowledged to be a less satisfactory work than al-Tusi's, being weaker in argument and encumbered with anecdotal material (following the literary taste of the period) from both Greek (indirectly) and Islamic sources, being more 'Ciceronian', as its 1839 translator, W.F. Thompson, apologetically expressed it. It is therefore easier to admire the work for its style than for its intellectual rigour. Thompson's translation does not improve matters, thanks to its baroque literary style and ponderous sentiments.

Apart from the *Akhlaq-e Jalali*, over seventy-five works by al-Dawani are recorded, covering the fields of philosophy, mysticism, theology and exegesis. Of particular interest to subsequent philosophers were his commentary on al-Suhrawardi's *Hayakil al-nur (The Temples of Light)*, *Shawakil al-hur fi sharh Hayakil al-nur (The Houri's Haunches in Commentary of the Temple of Light)* and his series of glosses on the commentary by 'Ala' al-Din al-Qushji (d. AH 879/AD 1474) on al-Tusi's *Tajrid al-kalam (Abstract of Theology)*. In both works he engaged with his contemporary Sadr al-Din al-Dashtaki (and subsequently the latter's son Ghiyath al-Din). All three were greatly influenced by al-Suhrawardi, although the Dashtakis perhaps more than al-Dawani. Sadr al-Din denied any reality to existence, either mental or extramental, and could thus be described as an extreme essentialist. Al-Dawani, on the other hand, held a view which harks back to Fakhr al-Din al-Razi. Existence in the outside world, for al-Dawani, is a single necessary reality, absolutely devoid of multiplicity, and is thus equal to God. Everything else is contingent: 'entities' whose existence is not real but only various 'portions of existence' (*hisas*) conceived by the mind. The reality of the external world is established solely through quiddities.

Al-Dawani's illuminationism is thus a modified one, but it proved more influential than al-Dashtaki's extreme form, for it was adopted by Mir Damad and initially by the latter's pupil Mulla Sadra before he turned to his radical existentialism.

See also: Ibn al-'Arabi; Illuminationist philosophy; Mystical philosophy in Islam

JOHN COOPER

List of works

al-Dawani (1467-77?) *Akhlaq-e Jalali (Jalalean Ethics)*, ed. M.K. Shirazi, Calcutta: printed at Habl-ul-Matin Press, 1911; trans. W.F. Thompson, *Practical philosophy of the Mohammadan people…, being a translation of the Akhlaq-i-Jalaly… from the Persian of Fakir Jany Muhammad Asaad…*, London: Oriental Translation Fund of Great Britain and Ireland, 1839; repr. Karachi, 1977.(No critical edition has yet been done of this work and the Shirazi edition is only one of several lithographs and printings. Thompson's translation, which is only partial, suffers from an overweight style, and the translator's attempts to stress in his notes the common Greek ancestry of Islamic and European philosophical ethics at the expense of the Islamic content of the work is now probably of little interest except to students of early nineteenth-century British thought.)

(1426-1502) *Shawakil-al-hur fi sharh-i-Hayakil-al-nur (The Houri's Haunches in Commentary on the Temple of Light)*, ed. M. Abdul Haq and M.Y. Kokan, Madras: Government Oriental Manuscripts Library, 1953.(Arabic text only, though there is a brief English introduction.)

References and further reading

Newman, A.J. (1994) 'Davani, Jalal al-Din Mohammad', *Encyclopaedia Iranica*, vol. VII, fasc. 2, Costa Meza, CA: Mazda Publishers, 132-3.(A mainly biographical article giving most of the relevant sources.)

Rosenthal, E.I.J. (1962) *Political Thought in Medieval Islam*, Cambridge: Cambridge University Press, 210-23.(A chapter in this work examines the political section of the *Akhlaq-e Jalali*.)

Wickens, G.M. (1984) 'Aklaq-e Jalali', *Encyclopaedia Iranica*, vol. I, fasc. 7, London: Routledge & Kegan Paul, 724.(A brief article on al-Dawani's best known work.)

Alemanno, Yohanan ben Isaac (1433/4-after 1503/4)

An outstanding Jewish thinker of the Italian Renaissance, Alemanno combined an eclectic Jewish philosophic rationalism, steeped in the medieval sources - Maimonidean, Averroist and Kabbalistic - with Renaissance humanism and Neoplatonism. He was an Aristotelian and Maimonidean in ethics, a Platonist and Averroist in political philosophy and a Neoplatonist and Kabbalist in metaphysics. His fusing of Aristotelian rationalism with Platonizing mysticism is striking but not atypical for the period. Influenced by Renaissance thought after he settled in Italy, he was active in Christian as well as Jewish circles in Florence, Padua and Mantua. Pico della Mirandola learned Hebrew under his instruction and relied on him for access to medieval Jewish texts in philosophy and Kabbalah. Both Christian Kabbalah and Renaissance Hebraism were products of the interactions in which Alemanno was a chief participant. His ties to the Florentine Academy of the late 1480s are evident in his adaptations to Jewish thinking of the ideas current among its members as to the unity of truth, the immortality of the soul and the dignity of man.

1 Life and background

Like many of his philosophical contemporaries, Alemanno was an immigrant to Italy. His name is an Italianized rendering of the surname Ashkenazi. He was born in Paris and died probably in Mantua. The work of such immigrants had a profound intellectual impact on the Italian Jewish communities of the latter Middle Ages and the Renaissance. Their influence was also felt on Renaissance culture at large. Like most of his scholarly Jewish contemporaries, Alemanno was a wandering scholar who travelled in search of a livelihood as a private teacher, preacher or secretary, always seeking the patronage of influential Jewish financiers. While wandering among such cities as Florence, Mantua, Padua and Bologna, he met many of the leading Jewish scholars of his time, including Judah Messer Leon, author of the *Nofet Tzufim (The Book of the Honeycomb's Flow)*, an important rhetorical treatise that aimed to integrate biblical rhetoric with the revived Ciceronian tradition of the Renaissance. Messer Leon's work profoundly influenced Alemanno, leading to his discovery of the full gamut of Renaissance humanist and Neoplatonic ideas and to his contacts with such leading exponents of Renaissance humanism as Giovanni Pico della Mirandola, his nephew Alberto Pico and Girolamo Benivieni. These scholars, especially Pico della Mirandola, relied on Jewish scholars like Alemanno, Elijah Delmedigo and Abraham Farissol, to learn Hebrew and to gain access to the sources of Jewish philosophy, the fascinating materials of Kabbalah, and the works of such Islamic thinkers as Averroes (see Ibn Rushd), which their Jewish guides translated for them into Latin from the medieval Hebrew and Arabic texts. The collaboration of Jewish with Christian scholars led to the creation of a Christian Kabbalah and was the foundation of Renaissance Hebraism.

Alemanno and other Jewish scholars cultivated their contacts with the humanists not only as an avenue to patronage, but also for the knowledge it gave them of the latest developments in literature, philosophy, medicine, politics and magic (see Humanism, Renaissance). Alemanno was particularly interested in such current learning and in transmitting it to his Jewish students and audiences. While Averroists like Delmedigo held suspect the mystical tendencies of Christian Neoplatonism, Alemanno eagerly pursued them. Commissioned by Pico in the late 1480s, the heyday of the Platonic Academy which Marsilio Ficino had established at Florence (see Platonism, Renaissance), Alemanno wrote his *Heshek Shelomo (The Passion of Solomon)*, an allegory on the Song of Songs and one of the first expressions of the Renaissance ideal of Platonic love, enunciated in Ficino's contemporaneous commentary on Plato's *Symposium*. Jewish exegetes had traditionally read the Song of Songs as an allegory of the spiritual love between God and Israel. That reading was now re-visioned through a Neoplatonic prism, in many ways anticipating the approach of Judah Abravanel (Leone Ebreo) in the *Dialoghi d'Amore (Dialogues of Love)*. In the foreword to his long introductory essay, *Shir ha-Ma'alot le-Shlomo (The Song of Solomon's Ascents)*, Alemanno describes his contacts with Pico at length, and portrayed Lorenzo di Medici, the patron of the Platonic Academy, as the living embodiment of the Platonic-Averroist philosopher king. In the text itself, King Solomon is made the prototype of the ideal philosophical ruler.

2 Writings

Most of Alemanno's writings survive today only in manuscript. *Shir ha-Ma'alot le-Shlomo* has been printed, but the early editions are partial and inaccurate. Of Alemanno's other writings, only a few extracts have appeared in

scholarly papers.

His most important work is *Hai ha-Olamim (Immortal Life)*, a work influenced in its construction by Ibn Tufayl's *Hayy Ibn Yaqzan (The Living Son of the Vigilant)* (see Ibn Tufayl). Alemanno in fact wrote a supercommentary on Moses of Narbonne's commentary on this work, most probably for a translation of the work from Hebrew to Latin commissioned by Pico. *Hai ha-Olamim* takes the form of a Platonic dialogue between a plain speaker and a philosopher, which unfolds in painstaking detail the development of the perfect individual, from his creation in the womb, through his physical, moral, political and intellectual development, until he reaches mystical union with God. As the title suggests, the soul of this perfect individual will become an immortal, *Hai ha-Olamim*. The description of his physical and moral perfection follows the Aristotelian tradition; that of his political perfection, mainly the Platonic tradition - all, of course, as seen through the eyes of medieval Muslim and Jewish intermediaries. With the approach to spiritual perfection, the Neoplatonic and Kabbalistic elements intensify, and the discussion ends on a powerful mystical chord.

Alemanno's other major work, *'Enei ha-'Edah (The Eyes of the Community)*, is a commentary on the Pentateuch through Genesis 5:1 and was probably linked originally to Pico's *Heptaplus de opere sex dierum Geneseos (On the Sevenfold Narration of the Six Days of Genesis)*, which deals with the same text and issues and was composed at the same time, in the late 1480s. His *Liqqutim (Compendia)*, comprises notebooks containing early drafts of his mature writings and important data on his life and intellectual background.

3 Philosophy

Rejecting what he saw as a medieval dichotomy between faith and reason, revelation and philosophy, and reacting against the widely bruited Christian Averroist notion of a double truth, Alemanno, like Pico, proclaimed the unity of truth and strove to harmonize philosophy, *halakhah*, Kabbalah, alchemy and astrology (see Halakhah; Kabbalah; Alchemy; Averroism, Jewish). His synthesis depended on ranking the various disciplines, spheres of existence and virtues in a hierarchy that acknowledged a dynamic relationship of emanation and love between the Creator and creation, and on making frequent analogies between matter and spirit, animal and man, microcosm and macrocosm, the Neoplatonic Intelligences and the Kabbalistic Sefirot (mystic Neopythagorean hypostases that mediate between the Infinite and creation). The heavy reliance on hierarchy and analogy generates a thick amalgam of ideas, in which Alemanno tried, often without success, to validate Neoplatonic theses by way of Aristotelian methods and typologies.

In the face of the esotericism fostered by medieval Jewish philosophers and Kabbalists, and insisted upon by many irate contemporaries, Alemanno addressed his writings to the widest possible audience, as is shown not only by his own declarations but also by his typically humanist attention to style and exposition. His model was the prophetic programme of addressing the entire community, each member in accordance with his highest understanding. For thinkers imbued with the values and practices of medieval Jewish philosophy, such exposure of the higher reaches of tradition to anyone who would listen was palpably subversive, but the aura of revealing long hidden secrets may explain the special interest that Pico and other Christian humanists took in Alemanno.

Central among the mysteries to be made known was the Platonic or Neoplatonic theory of the immortality of the soul, a core topic for the Florentine Academy. Following the teachings of medieval Jewish thought, both philosophical and Kabbalistic, Alemanno identified the soul's immortality with its knowledge of the one eternal truth that puts us in contact with the 'active intellect' and allows us to share the eternity of the Platonic Forms. But unlike many medieval thinkers who tended to limit this possibility to a handful of philosophers, Alemanno strove to widen it to the whole community, giving them access not only to physical resurrection, with the coming of the messiah, but also to spiritual immortality. Alemanno's teaching, preaching and lecturing was not only a quest for a livelihood, but also a vocation, a quest to disseminate knowledge, and so immortality, as widely as possible.

Not surprisingly, Alemanno adopted the Florentine theory of the dignity of man. He dovetailed Ficino's assignment of man to the mid-rank of being, between the bestial and the Divine, with Pico's insistence that while each created being has its proper place in the chain of being, only humans hold all the possibilities of existence. They can choose to descend into bestiality or raise themselves towards the Ideas and God. Alemanno, however, found limits to this freedom in the play of astral influences, and he was less sanguine than Pico as to our ability to make sound choices. In addressing this problem of human moral weakness, which a Christian thinker might have

interpreted in terms of original sin, Alemanno fell back on Jewish tradition: following Maimonides, he argued that the Commandments can steady our irresolution. But this meant that only Jews can properly use their freedom to attain the moral perfection that human beings need to achieve intellectual perfection and so be linked to the Sefirot and gain immortality.

See also: Averroism, Jewish; Ficino, M.; Humanism, Renaissance; Kabbalah; Pico della Mirandola, G.; Platonism, Renaissance

ABRAHAM MELAMED

List of works

Alemanno, Yohanan ben Isaac (*c*.1488-92) *Shir ha-Ma'alot le-Shlomo (The Song of Solomon's Ascents)*, ed. A.M. Lesley, 'The Song of Solomon's Ascents by Yohanan Alemanno: Love and Human Perfection According to a Jewish Colleague of Giovanni Pico della Mirandola', unpublished Ph.D. dissertation, Berkeley, CA: University of California, 1976, 2 vols.(Lesley's dissertation provides the Hebrew text with English summary and detailed introduction. Alemanno's long introduction to *Heshek Shelomo*, describing the various human virtues embodied by King Solomon.)

Alemanno, Yohanan ben Isaac (*c*.1488-92) *Heshek Shelomo (The Passion of Solomon)*, MS Oxford-Bodleian 1535/2 (Laud 103), MS London-Mentefiori 227, MS Berlin 143, and MS Moscow-Ginzburg 140.(An allegorical commentary on the Song of Songs.)

Alemanno, Yohanan ben Isaac (*c*.1470-1503) *Hai ha-Olamim (Immortal Life)*.(A detailed description of the various stages of human life, culminating with the metaphysical knowledge of God.)

Alemanno, Yohanan ben Isaac (*c*.1478-1504) *Liqqutim (Compendia)*.(An eclectic and fragmentary collection of notes and ideas, accumulated over a long period of time.)

Alemanno, Yohanan ben Isaac (before 1504) *'Enei ha-'Edah (The Eyes of the Community)*.(A philosophical commentary on the Torah. We have only the commentary on Genesis 1-5: 31. It is not clear whether it is incomplete or the later parts of the manuscript did not survive.)

References and further reading

Cassuto, U. (1918) *Gli Ebrei in Firenze, Nell' Eta del Rinascimento*, Florence: Tipografia Galletti e cocci, part 3, ch. 3.(Contains a description of Alemanno's life and writings.)

Idel, M. (1983) 'The Magical and Neoplatonic Interpretation of the Kabbalah in the Renaissance', in B.D. Cooperman (ed.) *Jewish Thought in the Sixteenth Century*, Cambridge, MA: Harvard University Press, 186-242.(Examines Alemanno's Kabbalistic ideas and Renaissance Kabbalism.)

Melamed, A. (1988) 'The Hebrew Laudatio of Yohanan Alemanno - In Praise of Lorenzo il Magnifico and the Florentine Constitution', in H. Beinart (ed.) *The Jews in Italy, Studies Dedicated to the Memory of U. Cassuto on the 100th Anniversary of His Birth*, Jerusalem: Magnes.(Discusses Alemanno's political philosophy and Renaissance humanism.)

Novak, B.C. (1982) 'Giovanni Pico della Mirandola and Jochanan Alemanno', *Journal of the Warburg and Courtauld Institutes* 45: 125-47.(A detailed account of their relationship.)

Alexander of Aphrodisias (*fl. c.* AD 200)

The Peripatetic philosopher Alexander was known to posterity as the *commentator on Aristotle, until Averroes took over this title. His commentaries eclipsed most of those of his predecessors, which now survive only in scattered quotations. Used by Plotinus, Alexander's commentaries were the basis for subsequent work on Aristotle by Neoplatonist commentators, and even though some themselves survive only in quotations by these later writers, Alexander's interpretations of particular passages are still helpful and are cited by commentators today.*

In addition to Alexander's commentaries we have a number of monographs, and also collections of short discussions which are connected with themes in his writings, though some are probably by pupils rather than by Alexander himself. Alexander's most influential and controversial doctrine has been his interpretation of Aristotle's theory of soul and intellect; regarding the soul as the product of the mixture of the bodily elements, he has been seen as subordinating form to matter and as thereby misinterpreting Aristotle. Certainly his view excludes any immortality for individuals, but even if Aristotle himself allowed this it is arguable that to do so was incompatible with his definition of soul as the form of potentially living body. Alexander himself interpreted Aristotle's 'active intellect' not as an immortal element in each individual, but as god, the unmoved mover, apprehended by our own intellects. Both on the question of soul and on that of the status of universals, Alexander gives a non-Platonizing reading of Aristotle, which accounts for some of the criticism to which he has been subjected by successors both ancient and modern. His treatment of the problem of free will has also been influential, though his criticisms of determinism are more telling than his own positive solution.

Seeing his task as interpreting Aristotle's writings with the aid of one another and explaining apparent inconsistencies, Alexander contributed to the growth of Aristotelianism as a system; he does not criticize nor challenge Aristotle, and regards his own innovations as Aristotelian doctrine, developed in the context of new questions which Aristotle himself had not confronted in the same form. He was better at seeing the details than at comprehending the global picture, and the potential of some of his doctrinal contributions is most apparent in what they suggested to others; but there is still much to interest philosophers in his detailed argumentation on particular points and passages.

1 Life, works and relation to Aristotle

Alexander's treatise *On Fate* is dedicated - with some elaborate rhetorical flourishes, and a request to consult him if further clarification is needed - to the emperors Septimius Severus and Caracalla, in gratitude for his appointment as a publicly recognized teacher of Aristotelian philosophy. Since Caracalla was made Augustus as Septimius Severus' colleague in AD 198, and Geta joined them as a third Augustus in AD 209, the date is fixed as between these two points; but we do not know at what stage in Alexander's career the appointment was made. Nor do we know for certain where the post in question was, though it is likely enough that it was the chair at Athens established by Marcus Aurelius (§1) in 176; Alexander's use of Aristotle's statue in Athens as an example in *On Aristotle's Metaphysics* (415.29-31) has been seen as supporting this.

Some of the general characteristics of Alexander's writings have been indicated above. His surviving commentaries are those on *Metaphysics* I-V (that on the remainder of the *Metaphysics*, like that on the *Sophistical Refutations*, is not by Alexander but by the twelfth-century Michael of Ephesus), *Prior Analytics* I, the *Topics*, the *Meteorology* and *On Sensation*. They are characterized by the frequent inclusion of alternative explanations, and by an absence of the formal organization, reflecting the programme of teaching in a school context, that is found in the later Neoplatonic commentaries on Aristotle. Alexander also wrote commentaries, now known only from later quotations, on the other logical and physical works of Aristotle. Whether he produced a full-scale commentary on the *Ethics* is debated, and he shows little or no interest in the zoological, political and rhetorical works.

Alexander's monographs include, surviving in Greek, *On the Soul* (as distinct from his commentary, now lost, on Aristotle's *On the Soul*), *On Fate*, *On Mixture*, and, surviving only in Arabic translation, *On the Principles of the Universe* (the authenticity of which has been questioned), *On Providence*, a work on differentiae, and *Refutation of Galen's Attack on Aristotle's Doctrine That Everything That Moves is Set In Motion by a Mover* (the actual connection of this treatise with Galen's views, like much else in the Arabic tradition concerning the relations between Alexander and Galen, is doubtful). There were other monographs, now lost (see below). In addition, the

extant collections of short discussions include, in Greek, the so-called second book of Alexander's *On the Soul*, better known by the name *Mantissa* or 'makeweight' given it by its modern editor, Ivo Bruns; three books of *Quaestiones (School-Puzzles and Solutions Concerning Nature)*; and one book of *Ethical Problems*. (Another collection, of *Medical Puzzles and Physical Problems*, has nothing to do with Alexander.) These collections were put together, often ineptly, by editors later than Alexander himself. Other similar material has been preserved in compendia in Greek manuscripts or in Arabic translation. Study of the relative dating of Alexander's works, and on the relation between the commentaries and the short discussions, is still in its infancy.

Both the relationships among Alexander's works and his loyalty to Aristotle can be illustrated by two particular topics. Both in the *Prior Analytics* commentary and in a separate monograph, now lost (see Alexander, *On Aristotle's Prior Analytics* 125.30- 1; Philoponus, *On Aristotle's Prior Analytics* 126.20), Alexander discussed Aristotle's modal logic; Alexander's writings are a major source for the controversy between Aristotle himself and his immediate followers, Theophrastus (§2) and Eudemus, over the conversion of contingent premises and the modality of the conclusions of syllogisms with 'mixed' premises (for example, one necessary and one assertoric). However, while many would hold that there is more logical elegance in Theophrastus' and Eudemus' view that the conclusion is in every case only as strong as the weakest premise (the medieval rule *peiorem semper conclusio sequitur partem*) Alexander remains loyal to Aristotle (*On Aristotle's Prior Analytics* 125.3-127.16). Second, Alexander answers Aristotle's problem in *Physics* VIII 4 254b33-, 'What is it that causes the natural movement of a falling heavy body?', by an analogy between the soul, as the form of a living creature and cause of its movement, and heaviness, as the form of a heavy body and the internal cause of *its* movement. This analogy - and it is presented only as an analogy - is put forward not only in the *Refutation of Galen on Motion*, but also in *On the Soul* (22.7-) and *On the Principles of the Universe*; and it has been seen by Pines (1961) as a possible ancestor of Philoponus' explanation of the motion of a projectile forced - in Aristotelian terms, rather than natural - by an internal impetus imparted to it by the thrower (see Philoponus §2). Alexander, as quoted by Simplicius, *On Aristotle's Physics* 1346.37-, remains loyal to Aristotle's implausible explanation of the continued motion of a projectile by movement imparted to the air behind the projectile as well as to it.

2 Soul and intellect

Aristotle defines soul as the first actuality of a natural body potentially possessing life, or, more shortly, of an organic body, and regards the soul of a living creature as its form (see Aristotle §17; Psychē). But it is controversial how this is to be understood. Some have interpreted Aristotle's notion of soul as a functionalist one; but this view has been criticized on the grounds that it does not do justice to the close connection in Aristotle between the performance of a given function and the particular arrangement needed for it. This close connection between form and matter in Aristotle's theory of them has caused major difficulties for interpreters, because it is not clear how soul and body can be logically distinguished, if a lifeless hand or eye is a hand or eye only in name, and only an already living, 'ensouled' body is to count as an organic body. It seems that either body must be defined in terms of soul, which raises the question whether there is any level at which the matter of a living body is specifiable without reference to its soul, or else that soul must be accounted for in terms of the arrangement of the body and its parts.

Interpreters of Aristotle have favoured the former approach. Aristotle himself (in *On the Soul* I 4) rejects - though with some hesitation - the notion that soul can be a 'harmony' or arrangement of the bodily elements, partly because such an arrangement cannot itself be a cause of movement as the soul is. However, Alexander not only defines the soul as the product of the mixture of the bodily elements (*On the Soul* 24.21-3), apparently following Andronicus, but sets out his exposition of the nature of soul by starting with the simple bodies, earth, air, fire and water, and working upwards through progressively more elaborate compounds until he arrives at living creatures and finally at human beings. It is therefore hardly surprising that his account of the soul has often been criticized as materialist, reductivist and un-Aristotelian. However, these criticisms may to some extent reflect the critics' own standpoints, and their own interpretations of Aristotle. It is scarcely un-Aristotelian to suggest that a given form requires a given arrangement of given types of matter, and Alexander's order of exposition need not indicate that he regards more complex forms as posterior to less complex ones so far as explanatory or ontological dependence is concerned. Indeed he derives the substantiality of the form-matter composite from that of the form and the matter (6.2-4), and insists that it is the form of each thing that determines its nature (7.4-8). Moreover, texts attributed to Alexander insist that form is not in matter, or soul in body, in the way that a quality can be in a

substrate, because it is by the form and the soul that the matter and the body are characterized in the first place (*Quaestiones* I 8, 17, 26; *Mantissa* 119-22.)

It is true that Alexander's treatment of soul excludes any individual immortality; indeed this was his chief source of popularity in the Renaissance (see §5). But Aristotle's view itself arguably encounters difficulties where personal immortality is concerned. Attribution to Aristotle, in his mature period, of belief in personal immortality turns on interpretation of his remarks concerning intellect, and especially the so-called 'active intellect' of his own *On the Soul* III.5 (see Aristotle §19; Nous). Alexander, however, identifies the active intellect not with an element peculiar to the soul of each individual but with god, the 'unmoved mover' of *Metaphysics* XII. The theory of intellect is discussed both in Alexander's *On the Soul* and in a section of the *Mantissa* which is of doubtful authenticity and seems itself to be a combination of several different texts, but which circulated independently in the Middle Ages first in Arabic translation and then in Latin, and was more influential than Alexander's *On the Soul* itself. Common to both works is the view that the individual human's intellect at birth is purely potential; it is therefore referred to as 'material intellect', by analogy with the potentiality of matter in the ordinary sense of the latter term. Since, however, it must be receptive to all forms, it has no nature of its own (see Aristotle, *On the Soul* III 4, 429b10-22); and indeed in *On the Soul* its state at birth is likened not so much to a blank writing-tablet (Aristotle, *On the Soul* III 4, 429a31-) as to the blankness of the tablet (84.24-7). As a person grows to adulthood the 'material' intellect develops, by the acquisition of concepts through the abstraction of matter from the forms in substances composed of form and matter, until it becomes intellect 'in disposition' (*en hexei*, later Latinized as *in habitu*), capable of independent thought.

What is less clear is the part which the active intellect is supposed to play in this process. Alexander's *On the Soul*, characteristically, simply presents two arguments *that* the unmoved mover, as pure self-thinking intellect and intelligible in its own right, is responsible for our thinking too, without explaining very adequately *how* this comes about. First, as supremely intelligible it must be the cause of other things' intelligibility (88.24-89.8) - an argument which sounds more Platonist than Aristotelian, though it is not indeed being used here to *establish* the existence of intelligible pure form. And second, it is the cause of being for all other things, and thus for all the objects of intellect (89.9-19). This is probably to be understood in terms of the movement of the heavens, caused by the unmoved mover, being the cause of sublunary coming-to-be (see §4); but as an explanation of how our intellects become able to think it scarcely seems adequate. In the short text *On Intellect* (107.31-4, 108.19-22), on the other hand, the active intellect appears to act directly upon our intellect, apparently by providing it with a paradigm of pure form and thus enabling it to separate other, 'enmattered' forms from the matter in which they are embodied - which apparently has the rather implausible implication that we must apprehend god, in order to possess this paradigm, before we can think of anything else in general terms. In *On the Soul* 90.11- 20 it is argued that, since intellect is identical with its object at any given time, immortality can be present in us when we think of god; but it is not our own 'material' intellect that then becomes immortal. This is the only immortality open to us as individuals (but see below on the eternity of species).

On Intellect (whether itself by Alexander or not) indicates that Alexander's treatment of the topic built upon earlier Peripatetic discussions, and that the identification of the active intellect with god rather than with an element in the individual soul had already been connected with Aristotle's reference to 'intellect from outside' (*Generation of Animals* II 3 736b27-). But that in fact relates to the origin of intellect in the context of the generation of individual human beings, with no explicit identification of the source from which such intellect comes, and no apparent reference to its entering into us through acts of intellectual apprehension, as in Alexander's view.

3 Universals

Aristotle rejects the Platonist view that forms of material objects can exist even in the absence of any material instances, and holds that the form of human being exists only in individual human beings and, in a different way, in the minds that think of them. What is much less clear is Aristotle's view of the ontological status of such forms, and in particular whether they are to be regarded as individual or universal. Alexander regards universals as posterior to individuals. He has therefore been criticized for adopting an un-Aristotelian nominalism; but his position is in fact more subtle. If we accept the evidence of the *Quaestiones* (I 3), Alexander draws a distinction between the nature, as such, of a species, and that nature as a universal. Definition is of what is common to the members of a species, as opposed to the individual accidents due to matter (*On the Soul* 85.15-18, where

Alexander seems to adopt a doctrine of numerically distinct forms in different members of the same species; see Aristotle). But the definition of the nature of the species would still be the same even if only one member of the species existed. Definition is thus of what is common, but not of what is common as common, and it is purely accidental to a specific nature whether it is universal, in the sense of having more than one instantiation, or not. However, while the individual is prior to the universal in the sense that an individual can exist without there being a universal, the universal is none the less, in cases where there is more than one instance, prior to any particular individual; the existence of 'human being' does not depend on the existence of Socrates or of any other particular named individual (*On Providence*, Ruland 1976: 89; compare Alexander reported by Dexippus, *On Aristotle's Categories* 45.16 and by Simplicius, *On Aristotle's Physics* 19.5-11). Similarly, the genus is prior to the species (*Quaestiones* I 11-). There can be animals without there being horses, but not horses without there being animals; on the other hand, this horse would still be an animal even if there were no other animals and no other horses at all.

Alexander does say (*On the Soul* 90.2-11) that forms embodied in matter depend for their existence on being intelligized, and that (*Quaestiones* II 28 78.18-20) genus *as genus* - that is, as including several different species - is just a name, its existence depending on its being thought of. But it is not clear that this involves nominalism, if by nominalism is meant the view that common natures are *arbitrary* thought-constructs or that their reality derives purely from our giving a common name to a particular collection of individuals (see Nominalism). The point of the statement at *On the Soul* 90.2-11 is to contrast them with pure intelligible forms (the unmoved movers), and, given the part played by specific natures in Alexander's theory of providence (see §4), it seems that, far from being nominalist, Alexander's theory of species is essentialist, involving a rigid distinction between the nature common to the species and individual accidents.

4 Providence and fate

On both providence and fate Alexander adapts Aristotelian materials to the discussion of new issues, presenting the resulting account as 'Aristotelian'. In the case of providence, discussed in the treatise *On Providence* and in several of the *Quaestiones*, especially the unfinished dialogue II 21, he is concerned to mediate between, on the one hand, interpretations of Aristotle (especially but not only by the hostile Platonist Atticus) as making divine influence on the sublunary world purely accidental and so not providence at all, and, on the other hand, the pantheistic doctrine of the Stoics (see Stoicism §5), which he regards as unworthy of the divine dignity by involving god directly in every detail of the world, however humble, and also as incompatible with the perceived existence of evils. His solution makes use of the Aristotelian theory (*Generation and Corruption* II 10) that the motion of the heavens and especially of the sun on the ecliptic, caused by desire for the unmoved mover, is responsible for the cycle of the seasons and thus for the continuity of coming-to-be and passing-away and the perpetuation of natural kinds. Alexander interprets this as providence, but a providence concerned with the eternity of species rather than with the fortunes of individuals. The charge that providence involves the divine existing for the sake of what is inferior to it was apparently answered by the argument (*Quaestiones* I 23 36.22-3; compare with I 25 41.1-2; *On the Principles of the Universe*, Badawi 1968: 127-8) that the continuation of the sublunary world benefits the heavens by giving them a centre around which to revolve; and Alexander apparently accepted that non-accidental providence must involve some awareness of its objects on the part of what exercises it, the divine presumably being aware of sublunary beings as species but not as individuals. Some of the details remain obscure, especially as concerns the identity of the being or beings exercising providence and the relation here between the divine heavenly spheres on the one hand and the unmoved mover(s) on the other; moreover, Alexander's theory of providence amounts to little more than an upholding of the general ordering of the world, as opposed to the concern with its complex history, and especially with the fortunes of individual human beings, characteristic of both the Stoic and the Judaeo-Christian traditions. At the end of the twelfth century Alexander's treatise *On Providence* was used as a source for ancient Greek theories - and its denial of divine concern for individuals rejected - by Moses Maimonides (*Guide to the Perplexed* III 16-17).

The notion of a general ordering of the universe which is unaffected by variations in detail also appears in Alexander's treatise *On Fate* (ch. 25), where it is used to counter the argument (Stoic, though the determinists attacked in the treatise are never actually named there as Stoics) that a nexus of causes and effects admitting of no exceptions is essential if the unity of the universe is to be preserved (see Stoicism §20). Unfortunately - and characteristically - Alexander in this treatise is more concerned to attack the determinist account of human agency,

as conflicting with common experience and detrimental to morality, than to explain how in his own view human agency fits into the world as a whole; this problem is more pressing for him than for Aristotle, because he is concerned to reject determinism while also claiming to avoid the introduction of any 'uncaused motion'. There is a similarity between Alexander's philosophical position here and the way in which Carneades had sought to escape determinism while rejecting the Epicureans' uncaused atomic 'swerve' (see Epicureanism §§4, 12; Carneades §3; and further below), though the question of possible historical influence of Carneades on Alexander is undecided, and Alexander makes no explicit reference to the atomic swerve or to the problems it involves. An attempt to locate human agency in the context of a general worldview is made by one of the short texts attributed to Alexander (*Mantissa*, Bruns 1887-92: 169-172), which thus reveals the limitations of Alexander's own treatment. However, by linking responsible choice to uncaused motion, contrary to Alexander's own view, it succeeds only in demonstrating the difficulties of a radical indeterminism.

In the absence of any single and systematic account of exceptions to determinism in *On Fate*, we are left with a series of separate claims:

(1) Alexander begins by setting out (chaps 3-6) an anti-determinist doctrine of fate - perhaps taken over, indeed, from earlier Peripatetic sources, basing it on the Aristotelian doctrine of nature as what applies for the most part but not always. Our actions are for the most part in accordance with our individual character, but not inevitably so.
(2) The occurrence of chance events disproves determinism (ch. 8; compare ch. 24). The stock Aristotelian examples of coincidences (finding buried treasure, and the like) which Alexander uses can, however, be accommodated within the Stoic system; true, if everything that happens is part of a single providential plan they will not really be coincidences, but Alexander's claim that they are rests, like much of his anti-Stoic argumentation in *On Fate*, on appeals to a common opinion which turns out to be Aristotelian school doctrine.
(3) Alexander argues (ch. 15) that our action can be free from being predetermined, and yet not be uncaused, because we ourselves, as agents, are the cause, this indeed being what it means to be human. This argument resembles, in general character though not in the details of its expression, that of Carneades on the same issue (Cicero, *On Fate* 25) which Richard Taylor has seen as anticipating modern agent causation theory. It is not, however, clear that introducing non-physical causes can provide a way out of the dilemma that either everything that occurs is predetermined on the physical level, or else there must be some break in the continuity of such physical causation.
(4) Following Aristotle (*Nicomachean Ethics* III 5), Alexander argues (chaps 27-9) that even if an agent cannot act contrary to a developed character, the development of that character is itself the agent's responsibility. This argument is unsatisfactory in itself - as Alexander apparently realizes; see (5) - simply pushing the problem back into the past; it also involves a view of the relation between natural endowment, developed character and action which is at least on the face of it different from that in (1). (Other texts attributed to Alexander take up this point, reconciling the two approaches by arguing that, while innate proclivities vary, everyone who is not morally deformed has the capacity to become virtuous: *Mantissa* 175.25-32; *Ethical Problems* 161.15-29.)
(5) To meet the difficulty of reconciling responsible choice, understood as requiring that the agent be able (and 'able' not just in a counterfactual sense) either to perform or not to perform the act in question, with the argument that for any given agent with a given perspective on a given situation only one course of action will be reasonable, Alexander makes three points. First, our actions are aimed not towards one goal but towards three: the noble, the advantageous and the pleasant (ch.15). The identification of these three ends is Aristotelian (*Nicomachean Ethics* II 3, 1104b30-), but not their treatment as equally valid alternatives for a single individual, and in the context of Alexander's argument it raises the question of what account we are to give of an agent's choice between them. Second, there is a certain degree of latitude, more than one possible action expressing a given character-trait or goal (ch. 29). This seems not to capture morally significant choices. Third, we may sometimes act otherwise than we normally would in a given situation, just to prove that we have the capacity to do so, and especially to confound a prophet (ch. 29). (A modern version might put it in terms of confounding a psychologist.) But this is entirely compatible with determinism, the abnormal action itself being a theoretically predictable reaction to an unusual situation.

Although Alexander fails to clarify his position, to such an extent that D. Frede has classified his position as compatibilist rather than libertarian (see Free will §1), the detailed arguments of the treatise *On Fate* anticipate many of the moves made in the free-will debate subsequently, and show considerable ingenuity. Alexander's

strongest anti-determinist argument rests on the inability of determinism to make sense of our experiences of choice and deliberation; and, in an argument analogous to Pascal's wager (see Pascal §6), he claims (ch. 21) that there is less danger in believing our actions are not predetermined when in fact they are, than in believing that they are when in fact they are not.

5 Influence

Alexander's commentaries were read in the school of Plotinus, the founder of Neoplatonism (Porphyry, *Life of Plotinus* 14). Studies of Alexander's influence on Plotinus have tended to find numerous correspondences in points of detail rather than conclusive evidence of influence on the major features of Plotinus' system. Plotinus used Alexander essentially as a guide to understanding Aristotle; the way in which Plotinus formulates the Aristotelian principle of the identity of intellect and its object shows Alexander's influence, but the doctrine itself had been adopted by Platonism earlier. Some have seen a connection between Alexander's discussion of a plurality of pure forms without matter (the plurality of unmoved movers) and Plotinus' doctrine of the unity of 'forms in intellect'; but Plotinian, and Platonic, forms are related to sensible objects in a different way from the Aristotelian unmoved movers, and in any case Alexander refers to pure forms sometimes in the singular and sometimes in the plural without seeming to attach much significance to the difference between the two. If Alexander's remarks did influence the Plotinian theory, they probably did so because of Plotinus' reflecting upon them rather than because of any awareness by Alexander of their possible significance. Similarly, too, with the suggestion that for Alexander our intellect, by apprehending the divine intellect, also apprehends its eternal objects; such a theory can be seen as a logical extension of Alexander's views, but it goes beyond anything that he actually says.

This point can be generalized. The limitations of Alexander's discussion of the active intellect have already been mentioned; his relatively terse reference to our achieving temporary immortality through contemplation of god is probably to be interpreted in terms not so much of mystical experience as of a desire to develop the logical implications of his account. Nevertheless, Alexander's doctrine of a single suprapersonal active intellect was immensely influential; it was later adopted by Averroes, though he - influenced by Neoplatonism - regarded the 'passive' or potential intellect too as one for all human beings. Aquinas was therefore able to cite Alexander for the individuality of the passive intellect in his controversy with the Averroists (see Aquinas §7; Averroism §1, 2), even though differing from him concerning the active intellect and the immortality of the individual. In the sixteenth century Alexander's view, rejecting personal immortality, was advocated notably by Petro Pomponazzi, in successful defiance of the decree of the Lateran Council in 1513 that 'individual immortality could be demonstrated philosophically and consequently had to be defended by all philosophers' (Kessler 1988: 495, 507), and by Jacopo Zabarella.

See also: Peripatetics; Aristotle Commentators

R.W. SHARPLES

List of works

Apart from *On Fate* (dated between AD 198 and 209; see §1), none of Alexander's works can be dated absolutely, and the study even of relative dates is still in its infancy. For texts preserved in Arabic see the listing and bibliography in Sharples (1987), to be corrected in the light of Hasnawi (1994) and Zimmermann (1994); Moraux *et al.* (forthcoming).

Alexander (*fl. c.* AD 200) Commentaries on Aristotle, in H. Diels (ed.) *Commentaria in Aristotelem Graeca* I-III, Berlin: Reimer, 1883-1901.(Greek text; for individual commentaries, see §1.)

Alexander (*fl. c.* AD 200) Other writings, in I. Bruns (ed.) *Supplementum Aristotelicum* II.1- 2, Berlin: Reimer, 1887-92.(Greek text; see §1.)

Alexander (*fl. c.* AD 200) Various works, trans. in R. Sorabji (ed.) *The Aristotelian Commentators*, London: Duckworth, 1987-.(Annotated translations of various works.)

Alexander (*fl. c.* AD 200) *On the Soul* and *On Intellect*, trans. in A.P. Fotinis, *The De Anima of Alexander of Aphrodisias*, Washington, DC: University Press of America, 1979.(English translation, with commentary.)

Alexander (*fl. c.* AD 200) *On the Soul*, trans. P. Accattino and P.L. Donini, *Alessandro di Afrodisia: L'anima*, Rome and Bari: Laterza, 1996.(Italian translation, with commentary.)

Alexander (*fl. c.* AD 200) *On Intellect*, trans. in F.M. Schroeder and R.B. Todd, *Two Aristotelian Greek Commentators on the Intellect*, Toronto, Ont.: Pontifical Institute of Medieval Studies, 1990.(English

translation, with commentary.)

Alexander (*fl. c.* AD 200) *On Fate*, in R.W. Sharples, *Alexander of Aphrodisias On Fate*, London: Duckworth, 1983; and in P. Thillet, *Alexandre d'Aphrodise: Traité du Destin*, Paris: Les Belle Lettres, 1984.(Sharples has Greek text reprinted from Bruns (1887- 92), with English translation and commentary; Thillet has Greek text with French translation and introduction.)

Alexander (*fl. c.* AD 200) *On Mixture*, in R.B. Todd, *Alexander of Aphrodisias on Stoic Physics*, Leiden: Brill, 1976.(Greek text, with English translation and commentary.)

Alexander (*fl. c.* AD 200) *On Providence*, in H.-J. Ruland, *Die arabischen Fassungen zweier Schriften des Alexander von Aphrodisias (Arabic Versions of Two Texts of Alexander of Aphrodisias)*, Saarbrücken: dissertation, 1976; and trans. S. Fazzo and M. Zonta, *Alessandro d'Afrodisia*, *Sulla Provvidenza*, Milan: Rizzoli, 1998.(Ruland has Arabic text, with German translation, introduction and appendices; Fazzo and Zonta, Italian translation with commentary.)

Alexander (*fl. c.* AD 200) *On the Principles of the Universe* and other works, trans. A. Badawi, in *La transmission de la philosophie grecque au monde arabe (The Transmission of Greek Philosophy to the Arab World)*, Paris: Vrin, 1968.(French translations from the Arabic.)

References and further reading

Donini, P.L. (1974) *Tre studi sull' aristotelismo nel II secolo d.C. (Three Studies of Aristotelianism in the Second Century ad)*, Turin: Paravia.(Discusses, in part 1, Alexander in relation to Neoplatonism and, in Part 3, Alexander on free will and determinism in the context especially of his psychology.)

Frede, D. (1984) 'Could Paris (Son of Priam) Have Chosen Otherwise?', *Oxford Studies in Ancient Philosophy* 2: 279-92.(Compatibilist interpretation of Alexander; see §4.)

Hasnawi, A. (1994) 'Alexandre d'Aphrodise vs Jean Philopon: notes sur quelques traités d'Alexandre "perdus" en grec, conservés en arabe' (Alexander of Aphrodisias v. John Philoponus: Notes on Some of Alexander's Treatises, 'Lost' in Greek and Preserved in Arabic), *Arabic Sciences and Philosophy* 4: 53-109.(Corrects misattributions to Alexander.)

Kessler, E. (1988) 'The Intellective Soul', in C.B. Schmitt and Q. Skinner (eds) *The Cambridge History of Renaissance Philosophy*, Cambridge: Cambridge University Press, 485-534.(Account of the Renaissance controversies referred to in §5.)

Lloyd, A.C. (1976) 'The Principle that the Cause is Greater Than its Effect', *Phronesis* 21: 146-56.(Discusses Alexander's use of this principle in connection with the active intellect; see §2.)

Merlan, P. (1967) 'Greek Philosophy from Plato to Plotinus', in A.H. Armstrong (ed.) *The Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press, 117-23.(Sees Alexander very much as a precursor of Neoplatonism; see §5.)

Moraux, P. ***et al.*** (forthcoming) *Der Aristotelismus bei den Griechen (Aristotelianism among the Greeks)*, vol. 3, ed. J. Wiesner, Berlin: de Gruyter.(Completed by others after Moraux's death, the definitive treatment of all aspects of Alexander's philosophy.)

Pines, S. (1961) 'Omne quod movetur necesse est ab aliquo moveri' (Everything that is moved must necessarily be moved by something), *Isis* 52: 21-54; repr. in *Studies in Arabic Versions of Greek Texts and in Medieval Science*, Jerusalem: Magnes Press, and Leiden: Brill, 1986, 218-51.(In English. Definitive treatment of Alexander's contributions to dynamics and possible influence; see §1.)

Robinson, H. (1991) 'Form and the Immateriality of the Intellect from Aristotle to Aquinas', in H. Blumenthal and H. Robinson (eds) *Aristotle and the Later Tradition*, Oxford: Clarendon Press, 207-26.(Forceful criticism of both Alexander and modern scholars for alleged reductivist misinterpretation of Aristotle's theory of soul; see §2.)

Sharples, R.W. (1987) 'Alexander of Aphrodisias: Scholasticism and Innovation', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36 2: 1, 176-243.(Full list of works attributed to Alexander, and more comprehensive discussion of the topics of the present article, with complete bibliography to 1986.)

Sharples, R.W. (1990) 'The School of Alexander', in R. Sorabji (ed.) *Aristotle Transformed: The Ancient Commentators and Their Influence*, London: Duckworth, 83-111.(Discusses Alexander's relation to earlier philosophical tradition, the didactic context of the shorter works and their relation to the commentaries.)

Taylor, R. (1967) 'Determinism', in P. Edwards (ed.) *Encyclopedia of Philosophy*, New York: Macmillan, vol. 2, 369.(A comprehensive general treatment of the free-will problem in all its aspects.)

Tweedale, M.M. (1984) 'Alexander of Aphrodisias' Views on Universals', *Phronesis* 29: 279-303.(A definitive discussion of the subject.)

Zimmermann, F.W. (1994) ’Proclus Arabus Rides Again’, *Arabic Sciences and Philosophy* 4: 9-51.(Corrects misattributions to Alexander, and discusses the methods of the Arabic translators and adaptors.)

Alexander of Hales (*c*.1185-1245)

Alexander's emphasis on speculative theology initiated the golden age of scholasticism. His philosophy was influenced by that of Aristotle, particularly in the field of ethics, and also by Augustine, Boethius and Peter Lombard. He believed that philosophy, based on natural reason, and theology, based on divine revelation, were two different disciplines and that philosophy ought to be independent of theology. He himself was primarily a theologian, and the colossal Summa Halesiana, most of which was compiled under his direction, constitutes the first complete theological synthesis in the West.

1 Life and influences

Alexander of Hales, called *doctor irrefragabilis* (the invincible doctor), was born in the village of Hales in the county of Shropshire (possibly the present Halesowen, now in neighbouring Worcestershire) *circa* 1180-85. The son of a rich rural family, he studied at the University of Paris, where he became regent master first in the faculty of Arts and later in the faculty of Theology. In 1230, he represented the University as procurator at the Papal Curia. In 1231, he returned for a short while to England, where he was made canon of Lichfield and soon after archdeacon of Coventry. In 1235, he was one of King Henry III's deputies, charged with renewing the peace treaty between England and France.

During this period, however, he retained his chair at the University of Paris. In 1236, he renounced honours and riches and entered the Order of St Francis, but he remained a master of theology and continued teaching as regent master in the Franciscan friary in Paris, which was an integral part of the University. Alexander was the first Franciscan Master of Theology and the first to teach theology by lecturing on the *Sentences* of Peter Lombard. His most prominent disciples include Bonaventure, Richard Rufus of Cornwall and John of La Rochelle, to whom he resigned his chair in theology near the end of his life. Alexander died in Paris on 21 August 1245.

As a master of arts, Alexander was familiar with the logical works of Aristotle, and with Books II and III of the *Nicomachean Ethics*, the so-called *ethica vetus*. He accepted Aristotle's definition of virtue and some of his views on the passions. Again following Aristotle, Alexander distinguished voluntary acts, involuntary acts and acts resulting from ignorance, and held that certain acts are indubitably evil.

By contrast, Alexander's knowledge of the *Metaphysics*, *Physics* and *On the Soul* was scant and superficial. At the time he was studying and teaching in the faculty of Arts, the so-called *libri naturales* were not available in Latin translation, and when they became available their reading and teaching were forbidden in Paris for several years. Philosophically, Alexander is more closely related to Augustine and Boethius than to Aristotle. He took from Augustine the distinction between natural, rational and moral philosophy, and from Boethius the distinction between nature and person and the view that all creatures are composed of essence (*quod est*) and existence (*quo est*). Most of the philosophical speculations in the *Summa Halesiana* must be attributed to Alexander's younger collaborators, John of La Rochelle, William of Militona and Odo Rigaldi, who had more thorough training in the philosophy of Aristotle.

2 Works

Alexander was, above all, a theologian. Though he believed that all philosophy begins with the principle of non-contradiction, he maintained that we must humble our minds in obedience to Christ and accept what appear to be contradictory statements if that is what faith requires (*Glossa* III d.24 FBS 14: 295). However, Alexander also supported the disciplinary separation of philosophy and theology which was to allow the independence of philosophy. He held that since philosophy is based on natural reason and theology on divine revelation, the two disciplines follow different paths and arrive at different levels of certitude.

Before 1945, none of Alexander's certainly authentic works were known. His fame and reputation rested entirely on the very extensive *Summa theologica* or *Summa Halesiana*, originally called the *Summa Fratris Alexandri*. However, the authenticity of this great work was a matter of debate among medievalists. This is partly because the Franciscan Roger Bacon, Alexander's contemporary, suggested that others wrote the *Summa* and attributed it to Alexander out of reverence. Also, different parts seem to be by different redactors; they are not always consistent. Finally, the Summa borrows from the writings of contemporary or earlier authors. Nevertheless, it is now an

established fact that the first three books (with the exception of some later additions) were composed before Alexander's death in 1245, very likely under his supervision, and the principal sources used were Alexander's own earlier writings. Therefore, even though the work is not, strictly speaking, a *Summa Fratris Alexandri*, it is correct to call it *Summa Halesiana*. The fourth book was finished after Alexander's death by William of Militona, who incorporated into it his own extensive and influential questions on the sacraments.

Later in his career as a master of theology, Alexander held many disputations on a variety of theological problems. The 68 questions he disputed before he became a Franciscan friar in 1236 have been edited; a further 78 questions, disputed later, still await editing.

Alexander's earliest known work, his commentary (*Glossa*) on books of the *Sentences* of Peter Lombard, was composed between 1223 and 1227. In the first book of the *Glossa*, there is a remarkable analysis of the mystery of the Trinity with subtle distinctions between properties, relations and notions (see Trinity). In Book II, Alexander teaches how we can see the vestiges of the Creator in the created universe and the image of the Trinity in the rational soul, which is one simple substance endowed with intellect, will, and memory (Fornaro 1985) (see Soul, nature and immortality of). In Book III, the doctrine of the hypostatic union is impressive for both its range of interest and its originality. It surpasses earlier authors in the variety of topics it treats, the sophistication of its theological method and the profundity of its thought (Principe 1967). In the fourth book, on the seven sacraments of the Church, he reveals himself an accomplished canonist, quoting Gratian and the *Decretals* of Gregory IX hundreds of times.

According to Roger Bacon, Alexander was the first to introduce the four books of the *Sentences* of Peter Lombard as a textbook for the faculty of theology. It was Alexander who divided the books into 'distinctions', according to the principal problems treated. As divided by Alexander, the *Sentences* remained the basis of theological instruction for three centuries; hundreds of commentaries were written on it. Alexander of Hales fully deserves his reputation as one of the greatest theologians of the thirteenth century.

See also: Aristotelianism, medieval; Augustinianism; John of La Rochelle; Lombard, P.; Trinity

GEDEON GÁL

List of works

Alexander of Hales (1223-7) *Glossa in Quattuor Libros Sententiarum Petri Lombardi (Commentary on the Four Books of Sentences of Peter Lombard)*, Bibliotheca Franciscana Scholastica Medii Aevi XII-XV, Quaracchi: Collegii S. Bonaventurae, 1951-57.(V. Doucet's prolegomenon to this edition (XII: 7-130) contains all the information available about Alexander of Hales and his works.)

Alexander of Hales (before 1236) *Quaestiones disputatae (Disputed Questions)*, Bibliotheca Franciscana Scholastica Medii Aevi IXI-XXI, Quaracchi: Collegii S. Bonaventurae, 1960. (Directed by V. Doucet, covers a variety of theological questions.)

Alexander of Hales (1241-42) *Expositio quattuor magistrorum super regulam Fratrum Minorum (1241-2) (Commentary by Four Masters on the Rule of the Friars Minor)*, ed. L. Oliger, Rome: Edizione de 'Storia e Letteratura', 1950.(A commentary on the Franciscan Rule, written with others.)

Alexander of Hales (before 1245) *Summa Halesiana*, ed. as *Doctoris Irrefragabilis Alexandri de Hales Ordinis Minorum Summa Theologica*, vols I-III, Quaracchi: Collegii S. Bonaveturae, 1924-48.(The vast prolegomenon in Librum III necnon in Libros I et II *Summae Fratris Alexandri* is a work of extraordinary erudition, a mine of information about scholastic authors in the first half of the thirteenth century.)

References and further reading

Bacon, Roger (1268) *Opus Minus*, ed. J.S. Brewer, Rerum Britannicarum Medii Aevi Scriptores 15, London: Longman, Green, Longman & Roberts, 1859: 325-29.(An important source for Alexander's life and works).

Boehner, P. (1945) 'The System of Metaphysics of Alexander of Hales', *Franciscan Studies* 5: 366-414.(Based on the *Summa Halesiana*.)

Doucet, V. (1947) 'The History of the Problem of the Authenticity of the Summa', *Franciscan Studies* 7: 26-41, 274-312.(See especially pages 310-12. Doucet was the foremost recent authority on Alexander and chief editor of his works.)

Fornaro, I. (1985) *La Teologia dell'Imagine nella Glossa di Alessandro d'Hales (The Image of God in Creation According to Alexander of Hales' Glossa)*, Vicenza: Edizione L.I.E.F.(The possibility of learning about God on

the basis of creation.)

Gössmann, E. (1964) *Metaphysik und Heilsgeschichte. Eine theologische Untersuchung der Summa Halesiana (Metaphysics and the Doctrine of Salvation: A Theological Examination of the Summa Halesiana)*, Munich: Max Huber Verlag.(A modern version of Boehner's study, which includes a full biblography.)

William of Militona [Guillelmi de Militona] (c.1245-60) *Quaestiones de Sacramentis (Questions on the Sacraments)*, ed. C. Piana and G. Gál, Bibliotheca Franciscana Scholastica Medii Aevi XXII-XXIII, Quaracchi: Collegii S. Bonaventurae, 1961.(Militona finished the fourth part of the *Summa Halesiana*, using his own *Quaestiones de Sacramentis*.)

Herscher, I. (1945) 'A Bibliography of Alexander of Hales', *Franciscan Studies* 5: 434-54.(An extensive bibliography of editions, books and articles in different languages.)

Osborne, K.B. (1994) 'Alexander of Hales: Precursor and Promoter of Franciscan Theology', in *The History of Franciscan Theology*, St Bonaventure, NY: The Franciscan Institute, 1-38.(A survey of Alexander's life, works and contribution to the field of theology.)

Principe, W.H. (1967) *The Theology of Hypostatic Union in the Early Thirteenth Century*, Vol. II, *Alexander of Hales' Theology of Hypostatic Union*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(An excellent exposition of Alexander's Christology).

Alexander, Samuel (1859-1938)

Alexander propounded a metaphysical system based on a view of Space-Time differentiated into 'motions' from which new qualities emerged at certain levels of organization; matter, life and mind being those qualities so far realized. Space-Time is a process with a 'nisus' (that is, an internal drive) towards a quality, as yet unrealized, called 'Deity'.

1 Space, time and deity

Alexander was born in Sydney, Australia, came to Balliol College, Oxford, and was subsequently Fellow of Lincoln College, Oxford, and Professor of Philosophy at the University of Manchester. His metaphysics rested on the concept of a Space-Time continuum ordered in four-dimensional perspectives from 'point-instants', which are not extended events, but limiting cases of 'motions', his term for the actual differentiations within Space-Time. These motions form patterns, the most general of which, notably causation, are all-pervasive categories. As a direct realist, he claimed that these categories were discerned, not imposed, by conceptual schemes. More specific motions in Space-Time displayed patterns in which distinctive qualities emerged at certain levels of organization. 'Matter', with its inertial properties, is the lowest level; certain of its complexes display the new quality 'life', and some of these display a further quality 'mind'. This view is a form of evolutionary naturalism, but not materialism, since 'matter' is the name for certain qualities from which further qualities of life and mind emerge, and these latter are not reducible to the former. Moreover, the basic reality is not matter but Space-Time.

Space-Time is not a closed system within which there can be redistributions of spatiotemporal coefficients. Its temporal aspect makes it an ongoing process in which there is a 'nisus' towards the production of new qualities. Beyond those known to us, there may be one, as yet unrealized, called 'Deity'. 'Deity' is not God as existent, but a quality towards which we can aspire.

2 Knowledge and values

The time dimension of Space-Time gives things an internal aspect as going through a process, while the space dimension sets them in an external relation called 'compresence'. A subject's inner experience of knowing is called 'enjoyment' and its relation of compresence to an object is called 'contemplation'. This is a direct realism, which raises difficulties over questions of error and counterfactual conditions. This realism extends to secondary qualities, such as colours. Besides secondary qualities there are 'tertiary' qualities called values, which arise in situations where one compresent factor is a mind. Chief of these are Truth, Beauty and Goodness. Here Alexander's main interest, expressed in a number of discussions of pieces of prose, poetry, art and architecture, was aesthetics. He held that there is a constructive impulse to manipulate materials which can be disengaged from practical ends and become contemplative, so leading to aesthetic appreciation. In stressing the need for a material medium as the carrier of the value of Beauty,Alexander's aesthetics are of a piece with his view of neurophysiological processes as bearers of non-reducible mental qualities shown in the capacity for conscious enjoyment and contemplation.

See also: Aesthetic concepts; Beauty; Value, ontological status of

DOROTHY EMMET

List of works

Alexander, S. (1889) *Moral Order and Progress: An Analysis of Ethical Concepts*, London: Trüber and Co. (Written from an idealist position that was later abandoned.)

Alexander, S. (1914) 'The Basis of Realism', *Proceedings of the British Academy* 6: 279-314; repr. R.M. Chisholm (ed.) *Realism and The Basis of Phenomenology*, Glencoe, IL: The Free Press, 1960, 186-222.(This relates Alexander's theory of knowledge to that of other philosophers of the time.)

Alexander, S. (1920) *Space, Time and Deity*, London: Macmillan; 2nd edn, 1927.(The second impression contains an important new preface. A paperback edition appeared in 1966.)

Alexander, S. (1933) *Beauty and the Other Forms of Value*, London: Macmillan.(Studies of Valuation, especially of Beauty, where practical impulses become contemplative, and of Goodness, where social impulses become subject to approbation and disapprobation.)

Alexander, S. (1939) *Philosophical and Literary Pieces*, London: Macmillan.(This is a posthumous collection edited with a memoir byJohn Laird. It contains a full bibliography of Alexander's writings.)

References and further reading

Broad, C.D. (1921) *Mind* 30: 23-39 and 129-50.(Review articles on *Space, Time and Deity*.)

Broad, C.D. (1929) *The Mind and Its Place in Nature*, London: Routledge & Kegan Paul. (Chapter 2 contains a discussion of the concept of emergence.)

Emmet, D. (1967) 'Alexander, Samuel', *Encyclopedia of Philosophy*, ed. P. Edwards, London and New York: Macmillan, vol. 1, 69-73.(A fuller treatment of Alexander's philosophy.)

Emmet, D. (1991) 'Whitehead und Alexander' in *Die Gifford Lectures und Ihre Deutung*, ed. M. Hampe and H. Massen, Frankfurt-am-Main: Suhrkamp, 100-20.(Translation by M. Hampe of an article comparing the philosophies of Whitehead and Alexander. This was republished (1992) in its original English version in *Process Studies* 21 (3): 137-48.)

al-Farabi, Abu Nasr (*c*.870-950)

Al-Farabi was known to the Arabs as the 'Second Master' (after Aristotle), and with good reason. It is unfortunate that his name has been overshadowed by those of later philosophers such as Ibn Sina, for al-Farabi was one of the world's great philosophers and much more original than many of his Islamic successors. A philosopher, logician and musician, he was also a major political scientist.

Al-Farabi has left us no autobiography and consequently, relatively little is known for certain about his life. His philosophical legacy, however, is large. In the arena of metaphysics he has been designated the 'Father of Islamic Neoplatonism', and while he was also saturated with Aristotelianism and certainly deploys the vocabulary of Aristotle, it is this Neoplatonic dimension which dominates much of his corpus. This is apparent in his most famous work, al-Madina al-fadila (The Virtuous City) which, far from being a copy or a clone of Plato's Republic, is imbued with the Neoplatonic concept of God. Of course, al-Madina al-fadila has undeniable Platonic elements but its theology, as opposed to its politics, places it outside the mainstream of pure Platonism.

In his admittedly complex theories of epistemology, al-Farabi has both an Aristotelian and Neoplatonic dimension, neither of which is totally integrated with the other. His influence was wide and extended not only to major Islamic philosophers such as Ibn Sina who came after him, and to lesser mortals such as Yahya ibn 'Adi, al-Sijistani, al-'Amiri and al-Tawhidi, but also to major thinkers of Christian medieval Europe including Thomas Aquinas.

1 Life and works

Abu Nasr Muhammad ibn Muhammad ibn Tarkhan ibn Awzalagh al-Farabi was born in approximately AH 257/AD 870. He may rightly be acclaimed as one of the greatest of Islamic philosophers of all time. While his name tends to be overshadowed by that of Ibn Sina, it is worth bearing in mind that the latter was less original than the former. Indeed, a well-known story tells how Ibn Sina sought in vain to understand Aristotle's *Metaphysics*, and it was only through a book by al-Farabi on the intentions of the *Metaphysics* that understanding finally came to him. However, unlike Ibn Sina, al-Farabi has left us no autobiography and we know far less about his life in consequence. Considerable myth has become attached to the man: it is unlikely, for example, that he really spoke more than seventy languages, and we may also query his alleged ascetic lifestyle. We do know that he was born in Turkestan and later studied Arabic in Baghdad; it has been claimed that most of his books were written here. He travelled to Damascus, Egypt, Harran and Aleppo, and in the latter city the Hamdanid ruler Sayf al-Dawla became his patron. Even the circumstances of his death are not clear: some accounts portray him dying naturally in Damascus while at least one holds that he was mugged and killed on the road from Damascus to Ascalon.

Al-Farabi became an expert in philosophy and logic, and also in music: one of his works is entitled *Kitab al-musiqa al-kabir (The Great Book of Music)*. However, perhaps the book for which he is best known is that whose title is abbreviated to *al-Madina al-fadila (The Virtuous City)*, and which is often compared, misleadingly in view of its Neoplatonic orientation, to Plato's *Republic*. Other major titles from al-Farabi's voluminous corpus included the *Risala fi'l-'aql (Epistle on the Intellect)*, *Kitab al-huruf (The Book of Letters)* and *Kitab ihsa' al-'ulum (The Book of the Enumeration of the Sciences)*.

2 Metaphysics

Majid Fakhry (1983) has described al-Farabi as 'the founder of Arab Neo-Platonism and the first major figure in the history of that philosophical movement since Proclus'. This should be borne in mind as we survey the metaphysics of the philosopher whom the Latin Middle Ages knew as Abunaser and whom the Arabs designated the 'Second Master' (after Aristotle). It should be noted that al-Farabi was an Aristotelian as well as a Neoplatonist: he is said, for example, to have read *On the Soul* two hundred times and even the *Physics* forty times. It should then come as no surprise that he deploys Aristotelian terminology, and indeed there are areas of his writings that are quite untouched by Neoplatonism. Furthermore, al-Farabi tried to demonstrate the basic agreement between Aristotle and Plato on such matters as the creation of the world, the survival of the soul and reward and punishment in the afterlife. In al-Farabi's conception of God, essence and existence fuse absolutely with no possible separation between the two. However, there is no getting away from the fact that it is the Neoplatonic element which dominates so much else of al-Farabi's work. We see this, for example, in the powerful

picture of the transcendent God of Neoplatonism which dominates *al-Madina al-fadila*. We see this too in al-Farabi's references to God in a negative mode, describing the deity by what he is not: he has no partner, he is indivisible and indefinable. And perhaps we see the Neoplatonic element most of all in the doctrine of emanation as it is deployed in al-Farabi's hierarchy of being.

At the top of this hierarchy is the Divine Being whom al-Farabi characterizes as 'the First'. From this emanates a second being which is the First Intellect. (This is termed, logically, 'the Second', that is, the Second Being). Like God, this being is an immaterial substance. A total of ten intellects emanate from the First Being. The First Intellect comprehends God and, in consequence of that comprehension, produces a third being, which is the Second Intellect. The First Intellect also comprehends its own essence, and the result of this comprehension is the production of the body and soul of *al-sama' al-ula*, the First Heaven. Each of the following emanated intellects are associated with the generation of similar astral phenomena, including the fixed stars, Saturn, Jupiter, Mars, the Sun, Venus, Mercury and the Moon. Of particular significance in the emanationist hierarchy is the Tenth Intellect: it is this intellect which constitutes the real bridge between the heavenly and terrestrial worlds. This Tenth Intellect (variously called by the philosophers the active or agent intellect in English, the *nous poiétikos* in Greek, the *dator formarum* in Latin and the *'aql al-fa''al* in Arabic) was responsible both for actualizing the potentiality for thought in man's intellect and emanating form to man and the sublunary world. With regard to the latter activity, it has been pointed out that here the active intellect takes on the role of Plotinus' Universal Soul (see Plotinus).

In Farabian metaphysics, then, the concept of Neoplatonic emanation replaces that of Qur'anic creation *ex nihilo* (see Neoplatonism in Islamic philosophy §2). Furthermore, the Deity at the top of the Neoplatonic hierarchy is portrayed in a very remote fashion. Al-Farabi's philosophers' God does not act directly on the sublunary world: much is delegated to the Active Intellect. However, God for al-Farabi certainly has an indirect 'responsibility' for everything, in that all things emanate from him. Yet we must also note, in order to present a fully rounded picture, that while it is the Neoplatonic portrait of God which dominates al-Farabi's writings, this is not the only picture. In some of his writings the philosopher *does* address God traditionally, Qur'anically and Islamically: he *does* invoke God as 'Lord of the Worlds' and 'God of the Easts and the Wests', and he asks God to robe him in splendid clothes, wisdom and humility and deliver him from misfortune. Yet the overwhelming Neoplatonic substratum of so much else of what he writes fully justifies Fakhry's characterization of al-Farabi, cited earlier, as 'the founder of Arab Neo-Platonism'.

3 Epistemology

Farabian epistemology has both a Neoplatonic and an Aristotelian dimension. Much of the former has already been surveyed in our examination of al-Farabi's metaphysics, and thus our attention turns now to the Aristotelian dimension. Our three primary Arabic sources for this are al-Farabi's *Kitab ihsa' al-'ulum*, *Risala fi'l-'aql* and *Kitab al-huruf*.

It is the second of these works, *Risala fi'l-'aql*, which provides perhaps the most useful key to al-Farabi's complex theories of intellection. In this work he divides *'aql* (intellect or reason) into six major categories in an attempt to elaborate the various meanings of the Arabic word *'aql*. First, there is what might be termed discernment or prudence; the individual who acts for the good is characterized by this faculty, and there is clearly some overlap with the fourth kind of intellect, described below. The second of al-Farabi's intellects is that which has been identified with common sense; this intellect has connotations of 'obviousness' and 'immediate recognition' associated with it. Al-Farabi's third intellect is natural perception. He traces its source to Aristotle's *Posterior Analytics*, and it is this intellect which allows us to be certain about fundamental truths. It is not a skill derived from the study of logic, but it may well be inborn. The fourth of the six intellects may be characterized as 'conscience': this is drawn by the philosopher from Book VI of Aristotle's *Nicomachean Ethics*. It is a quality whereby good might be distinguished from evil and results from considerable experience of life (see Aristotle §§18-21).

Al-Farabi's fifth intellect is both the most difficult and the most important. He gives most space to its description in his *Risala fi'l-'aql* and considers it to be of four different types: potential intellect, actual intellect, acquired intellect and agent or active intellect. *'Aql bi'l-quwwa* (potential intellect) is the intellect which, in Fakhry's words, has the capacity 'of abstracting the forms of existing entities with which it is ultimately identified' (Fakhry 1983: 121). Potential intellect can thus become *'aql bi'l-fi'l* (actual intellect). In its relationship to the actual intellect, the

third sub-species of intellect, *‘aql mustafad* (acquired intellect) is, to use Fakhry’s words again, the ‘the agent of actualization’ to the actualized object. Finally, there is the *‘aql al-fa‘‘al* (agent or active intellect), which was described in §2 above and need not be elaborated upon again.

The sixth and last of the major intellects is Divine Reason or God himself, the source of all intellectual energy and power. Even this brief presentation of Farabian intellection must appear complex; however, given the complexity of the subject itself, there is little option.

The best source for al-Farabi’s classification of knowledge is his *Kitab ihsa’ al-‘ulum*. This work illustrates neatly al-Farabi’s beliefs both about what can be known and the sheer range of that knowledge. Here he leaves aside the division into theological and philosophical sciences which other Islamic thinkers would use, and divides his material instead into five major chapters. Through all of them runs a primary Aristotelian stress on the importance of knowledge. Chapter 1 deals with the ‘science of language’, Chapter 2 formally covers the ‘science of logic’, Chapter 3 is devoted to the ‘mathematical sciences’, Chapter 4 surveys physics and metaphysics, and the final chapter encompasses ‘civil science’ (some prefer the term ‘political science’), jurisprudence and scholastic theology. A brief examination of these chapter headings shows that a total of eight main subjects are covered; not surprisingly, there are further subdivisions as well. To give just one example, the third chapter on the mathematical sciences embraces the seven subdivisions of arithmetic, geometry, optics, astronomy, music, weights and ‘mechanical artifices’; these subdivisions in turn have their own subdivisions. Thus al-Farabi’s epistemology, from what has been described both in this section and §2 above, may be said to be encyclopedic in range and complex in articulation, with that articulation using both a Neoplatonic and an Aristotelian voice.

4 Political philosophy

The best known Arabic source for al-Farabi’s political philosophy is *al-Madina al-fadila*. While this work undoubtedly embraces Platonic themes, it is in no way an Arabic clone of Plato’s *Republic*. This becomes very clear right at the beginning of al-Farabi’s work, with its description of the First Cause (Chapters 1-2) and the emanation of ‘the Second’ from ‘The First’ (Chapter 3). Later in the work, however, al-Farabi lays down in Platonic fashion the qualities necessary for the ruler: he should be predisposed to rule by virtue of an innate disposition and exhibit the right attitude for such rule. He will have perfected himself and be a good orator, and his soul will be, as it were, united to the active intellect (see §3). He will have a strong physique, a good understanding and memory, love learning and truth and be above the materialism of this world. Other qualities are enumerated by al-Farabi as well, and it is clear that here his ideal ruler is akin to Plato’s classical philosopher-king (see Plato §14).

Al-Farabi has a number of political divisions for his world. He identifies, for example, three types of society which are perfect and grades these according to size. His ideal virtuous city, which gives its name to the whole volume, is that which wholeheartedly embraces the pursuit of goodness and happiness and where the virtues will clearly abound. This virtuous city is compared in its function to the limbs of a perfectly healthy body. By stark contrast, al-Farabi identifies four different types of corrupt city: these are the ignorant city (*al-madina al-jahiliyya*), the dissolute city (*al-madina al-fasiqa*), the turncoat city (*al-madina al-mubaddala*) and the straying city (*al-madina al-dalla*). The souls of many of the inhabitants of such cities face ultimate extinction, while those who have been the cause of their fall face eternal torment. In itemizing four corrupt societies, al-Farabi was surely aware of Plato’s own fourfold division of imperfect societies in the Republic into timarchy, oligarchy, democracy and tyranny. The resemblance, however, is more one of structure (four divisions) rather than of content.

At the heart of al-Farabi’s political philosophy is the concept of happiness (*sa‘ada*). The virtuous society (*al-ijtima‘ al-fadil*) is defined as that in which people cooperate to gain happiness. The virtuous city (*al-madina al-fadila*) is one where there is cooperation in achieving happiness. The virtuous world (*al-ma‘mura al-fadila*) will only occur when all its constituent nations collaborate to achieve happiness. Walzer reminds us that both Plato and Aristotle held that supreme happiness was only to be gained by those who philosophized in the right manner. Al-Farabi followed the Greek paradigm and the highest rank of happiness was allocated to his ideal sovereign whose soul was ‘united as it were with the Active Intellect’. But Walzer goes on to stress that al-Farabi ‘does not confine his interest to the felicity of the first ruler: he is equally concerned with the felicity of all the five classes which make up the perfect state’ (Walzer, in introduction to *al-Madina al-fadila* (1985: 409-10)). Farabian political philosophy, then, sits astride the saddle of Greek *eudaimonia*, and a soteriological dimension may easily

be deduced from this emphasis on happiness. For if salvation in some form is reserved for the inhabitants of the virtuous city, and if the essence of that city is happiness, then it is no exaggeration to say that salvation is the reward of those who cooperate in the achievement of human happiness. *Eudaimonia*/*sa'ada* becomes a soteriological raft or steed.

5 Influence

The impact of al-Farabi's work on Ibn Sina was not limited merely to illuminating Aristotle's *Metaphysics*. It was with good reason that al-Farabi was designated the 'Second Master' (after Aristotle). One modern scholar recently acknowledged the dependence of Ibn Sina on al-Farabi in a book dealing with both which he entitled *The Two Farabis* (Farrukh 1944). And if Aquinas (§9) did not derive his essence-existence doctrine from al-Farabi but from the Latinized Ibn Sina, as is generally assumed, there is no doubt that Farabian concepts of essence and existence provided a base for the elaborated metaphysics of Ibn Sina and thence of Aquinas. Finally, the briefest of comparisons between the tenfold hierarchy of intellection produced by al-Farabi and the similar hierarchy espoused by Ibn Sina, each of which gives a key role to the Tenth Intellect, shows that in matters of emanation, hierarchy and Neoplatonic intellection, Ibn Sina owes a considerable intellectual debt to his predecessor.

Al-Farabi influenced many other thinkers as well. A glance at the period between AH 256/AD 870 and AH 414/AD 1023 and at four of the major thinkers who flourished in this period serves to confirm this: Yahya ibn 'Adi, Abu Sulayman al-Sijistani, Abu 'l-Hasan Muhammad ibn Yusuf al-'Amiri and Abu Hayyan al-Tawhidi may all be said to constitute in one form or another a 'Farabian School'. The Christian Monophysite Yahya ibn 'Adi studied in Baghdad under al-Farabi and others. Like his master, Yahya was devoted to the study of logic; like his master also, Yahya held that there was a real link between reason, ethics and politics. Al-Sijistani was a pupil of Yahya's and thus at one remove from al-Farabi; nonetheless, he shared in both his master's and al-Farabi's devotion to logic, and indeed was known as al-Sijistani al-Mantiqi (The Logician). In his use of Platonic classification and thought, al-Sijistani reveals himself as a true disciple of al-Farabi. Although al-'Amiri appears to speak disparagingly of al-Farabi at one point, there can be no doubt about al-Farabi's impact on him. Indeed, al-'Amiri's works combine the Platonic, the Aristotelian and the Neoplatonic. Finally, Abu Hayyan al-Tawhidi, a pupil of both Yahya and al-Sijistani, stressed, for example, the primacy of reason and the necessity of using logic. Like others of the Farabian School outlined above, al-Tawhidi contributed towards a body of thought the primary constituents of which were the soteriological, the ethical and the noetic.

See also: Aristotelianism in Islamic philosophy; Greek philosophy: impact on Islamic philosophy; Ibn Sina; Logic in Islamic philosophy; Neoplatonism in Islamic philosophy; Political philosophy in classical Islam

IAN RICHARD NETTON

List of works

al-Farabi (*c*.870-950) *al-Madina al-fadila (The Virtuous City)*, trans. R. Walzer, *Al-Farabi on the Perfect State: Abu Nasr al-Farabi's Mabadi' Ara Ahl al-Madina al-Fadila*, Oxford: Clarendon Press, 1985.(Revised with introduction and commentary by the translator.)

al-Farabi (*c*.870-950) *Risala fi'l-'aql (Epistle on the Intellect)*, ed. M. Bouyges, Beirut: Imprimerie Catholique, 1938.(A seminal text for the understanding of Farabian epistemology.)

al-Farabi (*c*.870-950) *Kitab al-huruf (The Book of Letters)*, ed. M. Mahdi, Beirut: Dar al-Mashriq, 1969. (Modelled on Aristotle's *Metaphysics*, but of interest to students of linguistics as well as of philosophy.)

al-Farabi(*c*. 870-950) *Kitab ihsa' al-'ulum (The Book of the Enumeration of the Sciences)*, ed. and trans. A. González Palencia, *Catálogo de las Ciencias*, Arabic text with Latin and Spanish translation, Madrid: Imprenta y Editorial Maestre, 1953.(A survey of the learned sciences of the day, of encyclopedic range.)

al-Farabi (*c*.870-950) *Kitab al-musiqa al-kabir (The Great Book of Music)*, ed. G.A. Khashab and M.A. al-Hafni, Cairo: Dar al-Katib al-'Arabi, 1967.(Al-Farabi's major contribution to musicology.)

References and further reading

Alon, I. (1990) 'Farabi's Funny Flora: Al-Nawabit as Opposition', *Arabica* 37: 56-90.(Highly creative discussion of the links between the philosophical terminology of Ibn Bajja and al-Farabi, which brings out the complexity of the theological and political ramifications of such language.)

Black, D. (1996) 'Al-Farabi', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London:

Routledge, ch. 12, 178-97.(Account of the thought and main works of al-Farabi.)
Fakhry, M. (1983) *A History of Islamic Philosophy*, London: Longman; New York: Columbia University Press, 2nd edn.(An excellent standard introduction to the field. See especially pages 107-128.)
Farrukh, U. (1944) *Al-Farabiyyan (The Two Farabis),* Beirut.(Ibn Sina's dependence on al-Farabi, as mentioned in §5.)
Galston, M. (1990) *Politics and Excellence: The Political Philosophy of Alfarabi*, Princeton, NJ: Princeton University Press.(A major analysis of an important aspect of Farabian philosophy.)
Netton, I.R. (1989) *Allah Transcendent: Studies in the Structure and Semiotics of Islamic Philosophy, Theology and Cosmology*, London and New York: Routledge. (Contains a wide-ranging chapter on al-Farabi, see pages 99-148. This volume was later published in paperback by Curzon Press in 1994.)
Netton, I.R. (1992) *Al-Farabi and His School*, Arabic Thought and Culture Series, London and New York: Routledge.(Assesses the philosopher through an epistemological lens.)

al-Ghazali, Abu Hamid (1058-1111)

Al-Ghazali is one of the greatest Islamic jurists, theologians and mystical thinkers. He learned various branches of the traditional Islamic religious sciences in his home town of Tus, Gurgan and Nishapur in the northern part of Iran. He was also involved in Sufi practices from an early age. Being recognized by Nizam al-Mulk, the vizir of the Seljuq sultans, he was appointed head of the Nizamiyyah College at Baghdad in AH 484/AD 1091. As the intellectual head of the Islamic community, he was busy lecturing on Islamic jurisprudence at the College, and also refuting heresies and responding to questions from all segments of the community. Four years later, however, al-Ghazali fell into a serious spiritual crisis and finally left Baghdad, renouncing his career and the world. After wandering in Syria and Palestine for about two years and finishing the pilgrimage to Mecca, he returned to Tus, where he was engaged in writing, Sufi practices and teaching his disciples until his death. In the meantime he resumed teaching for a few years at the Nizamiyyah College in Nishapur.

*Al-Ghazali explained in his autobiography why he renounced his brilliant career and turned to Sufism. It was, he says, due to his realization that there was no way to certain knowledge or the conviction of revelatory truth except through Sufism. (This means that the traditional form of Islamic faith was in a very critical condition at the time.) This realization is possibly related to his criticism of Islamic philosophy. In fact, his refutation of philosophy is not a mere criticism from a certain (orthodox) theological viewpoint. First of all, his attitude towards philosophy was ambivalent; it was both an object and criticism and an object of learning (for example, logic and the natural sciences). He mastered philosophy and then criticized it in order to Islamicize it. The importance of his criticism lies in his philosophical demonstration that the philosophers' metaphysical arguments cannot stand the test of reason. However, he was also forced to admit that the certainty of revelatory truth, for which he was so desperately searching, cannot be obtained by reason. It was only later that he finally attained to that truth in the ecstatic state (*fana'*) of the Sufi. Through his own religious experience, he worked to revive the faith of Islam by reconstructing the religious sciences upon the basis of Sufism, and to give a theoretical foundation to the latter under the influence of philosophy. Thus Sufism came to be generally recognized in the Islamic community. Though Islamic philosophy did not long survive al-Ghazali's criticism, he contributed greatly to the subsequent philosophization of Islamic theology and Sufism.*

1 Life

The eventful life of Abu Hamid Muhammad ibn Muhammad al-Ghazali (or al-Ghazzali) can be divided into three major periods. The first is the period of learning, first in his home town of Tus in Persia, then in Gurgan and finally in Nishapur. After the death of his teacher, Imam al-Haramayn al-Juwayni, Ghazali moved to the court of Nizam al-Mulk, the powerful vizir of the Seljuq Sultans, who eventually appointed him head of the Nizamiyyah College at Baghdad in AH 484/AD 1091.

The second period of al-Ghazali's life was his brilliant career as the highest-ranking orthodox 'doctor' of the Islamic community in Baghdad (AH 484-8/AD 1091-5). This period was short but significant. During this time, as well as lecturing on Islamic jurisprudence at the College, he was also busy refuting heresies and responding to questions from all segments of the community. In the political confusion following the assassination of Nizam al-Mulk and the subsequent violent death of Sultan Malikshah, al-Ghazali himself fell into a serious spiritual crisis and finally left Baghdad, renouncing his career and the world.

This event marks the beginning of the third period of his life, that of retirement (AH 488-505/AD 1095-1111), but which also included a short period of teaching at the Nizamiyyah College in Nishapur. After leaving Baghdad, he wandered as a Sufi in Syria and Palestine before returning to Tus, where he was engaged in writing, Sufi practices and teaching his disciples until his death.

The inner development leading to his conversion is explained in his autobiography, *al-Munqidh min al-dalal (The Deliverer from Error)*, written late in his life. It was his habit from an early age, he says, to search for the true reality of things. In the process he came to doubt the senses and even reason itself as the means to 'certain knowledge', and fell into a deep scepticism. However, he was eventually delivered from this with the aid of the divine light, and thus recovered his trust in reason. Using reason, he then set out to examine the teachings of 'the seekers after truth': the theologians, philosophers, Isma'ilis and Sufis. As a result of these studies, he came to the

realization that there was no way to certain knowledge except through Sufism. In order to reach this ultimate truth of the Sufis, however, it is first necessary to renounce the world and to devote oneself to mystical practice. Al-Ghazali came to this realization through an agonising process of decision, which led to a nervous breakdown and finally to his departure from Baghdad.

The schematic presentation of *al-Munqidh* has allowed various interpretations, but it is irrelevant to question the main line of the story. Though certain knowledge is explained in *al-Munqidh* as something logically necessary, it is also religious conviction (*yaqin*) as mentioned in the *Ihya' 'ulum al-din (The Revival of the Religious Sciences)*. Thus when he says that the traditional teachings did not grip him in his adolescence, he means to say that he lost his conviction of their truth, which he only later regained through his Sufi mystical experiences. He worked to generalize this experience to cure 'the disease' of his time.

The life of al-Ghazali has been thus far examined mostly as the development of his individual personality. However, since the 1950s there have appeared some new attempts to understand his life in its wider political and historical context (Watt 1963). If we accept his religious confession as sincere, then we should be careful not to reduce his thought and work entirely to non-religious factors. It may well be that al-Ghazali's conversion from the life of an orthodox doctor to Sufism was not merely the outcome of his personal development but also a manifestation of a new stage in the understanding of faith in the historical development of Islam, from the traditional form of faith expressed in the effort to establish the kingdom of God on Earth through the *shari'a* to a faith expressed as direct communion with God in Sufi mystical experience. This may be a reflection of a development in which the former type of faith had lost its relevance and become a mere formality due to the political and social confusion of the community. Al-Ghazali experienced this change during his life, and tried to revive the entire structure of the religious sciences on the basis of Sufism, while at the same time arguing for the official recognition of the latter and providing it with solid philosophical foundations.

2 Theological conceptions

Al-Ghazali wrote at least two works on theology, *al-Iqtisad fi'l-i'tiqad (The Middle Path in Theology)* and *al-Risala al-Qudsiyya (The Jerusalem Epistle)*. The former was composed towards the end of his stay in Baghdad and after his critique of philosophy, the latter soon afterwards in Jerusalem. The theological position expressed in both works is Ash'arite, and there is no fundamental difference between al-Ghazali and the Ash'arite school (see Ash'ariyya and Mu'tazila). However, some changes can be seen in the theological thought of his later works, written under the influence of philosophy and Sufism (see §4).

As Ash'arite theology came into being out of criticism of Mu'tazilite rationalistic theology, the two schools have much in common but they are also not without their differences. There is no essential difference between them as to God's essence (*dhat Allah*); al-Ghazali proves the existence of God (the Creator) from the createdness (*hadath*) of the world according to the traditional Ash'arite proof. An atomistic ontology is presupposed here, and yet there are also philosophical arguments to refute the criticism of the philosophers. As for God's attributes (*sifat Allah*), however, al-Ghazali regards them as 'something different from, yet added to, God's essence' (*al-Iqtisad*: 65), while the Mu'tazilites deny the existence of the attributes and reduce them to God's essence and acts. According to al-Ghazali, God has attributes such as knowledge, life, will, hearing, seeing and speech, which are included in God's essence and coeternal with it. Concerning the relationship between God's essence and his attributes, both are said to be 'not identical, but not different' (*al-Iqtisad*: 65). The creation of the world and its subsequent changes are produced by God's eternal knowledge, will and power, but this does not necessarily mean any change in God's attributes in accordance with these changes in the empirical world.

One of the main issues of theological debate was the relationship between God's power and human acts. The Mu'tazilites, admitting the continuation of an accident (*'arad*) of human power, asserted that human acts were decided and produced (or even created) by people themselves; thus they justified human responsibility for acts and maintained divine justice. In contrast, assuming that all the events in the world and human acts are caused by God's knowledge, will and power, al-Ghazali admits two powers in human acts, God's power and human power. Human power and act are both created by God, and so human action is God's creation (*khalq*), but it is also human acquisition (*kasb*) of God's action, which is reflected in human volition. Thus al-Ghazali tries to harmonize God's omnipotence and our own responsibility for our actions (see Omnipotence).

As for God's acts, the Mu'tazilites, emphasizing divine justice, assert that God cannot place any obligation on people that is beyond their ability; God must do what is best for humans and must give rewards and punishments according to their obedience and disobedience. They also assert that it is obligatory for people to know God through reason even before revelation. Al-Ghazali denies these views. God, he says, can place any obligations he wishes upon us; it is not incumbent on him to do what is best for us, nor to give rewards and punishments according to our obedience and disobedience. All this is unimaginable for God, since he is absolutely free and is under no obligation at all. Obligation (*wujub*), says al-Ghazali, means something that produces serious harm unless performed, but nothing does harm to God. Furthermore, good (*hasan*) and evil (*qabih*) mean respectively congruity and incongruity with a purpose, but God has no purpose at all. Therefore, God's acts are beyond human ethical judgment. Besides, says al-Ghazali, injustice (*zulm*) means an encroachment on others' rights, but all creatures belong to God; therefore, whatever he may do to his creatures, he cannot be considered unjust.

The Mu'tazilites, inferring the hereafter from the nature of this world, deny the punishment of unbelievers in the grave from their death until the resurrection, and also the reality of the various eschatological events such as the passing of the narrow bridge and the weighing on the balance of human deeds (see Eschatology). Al-Ghazali, on the other hand, rejecting the principle of analogy between the two worlds, approves the reality of all these events as transmitted traditionally, since it cannot be proven that they are rationally or logically impossible. Another important eschatological event is the seeing of God (*ru'ya Allah*). While the Mu'tazilites deny its reality, asserting that God cannot be the object of human vision, al-Ghazali approves it as a kind of knowledge which is beyond corporeality; in fact, he later gives the vision of God deep mystical and philosophical meaning. In short, the Mu'tazilites discuss the unity of God and his acts from the viewpoint of human reason, but al-Ghazali does so on the presupposition that God is personal and an absolute reality beyond human reason.

3 Refutation of philosophy

Al-Ghazali's relationship with philosophy is subtle and complicated. The philosophy represented by al-Farabi and Ibn Sina (Avicenna) is, for al-Ghazali, not simply an object of criticism but also an important component of his own learning. He studied philosophy intensively while in Baghdad, composing *Maqasid al-falasifa (The Intentions of the Philosophers)*, and then criticizing it in his *Tahafut al-falasifa (The Incoherence of the Philosophers)*. The *Maqasid* is a precise summary of philosophy (it is said to be an Arabic version of Ibn Sina's Persian *Danashnamah-yi ala'i (Book of Scientific Knowledge)* though a close comparative study of the two works has yet to be made). In the medieval Latin world, however, the content of the *Maqasid* was believed to be al-Ghazali's own thought, due to textual defects in the Latin manuscripts. As a result, the image of the 'Philosopher Algazel' was created. It was only in the middle of the nineteenth century that Munk corrected this mistake by making use of the complete manuscripts of the Hebrew translation. More works by al-Ghazali began to be published thereafter, but some contained philosophical ideas he himself had once rejected. This made al-Ghazali's relation to philosophy once again obscure. Did he turn back to philosophy late in life? Was he a secret philosopher? From the middle of the twentieth century there were several attempts to verify al-Ghazali's authentic works through textual criticism, and as a result of these works the image of al-Ghazali as an orthodox Ash'arite theologian began to prevail. The new trend in the study of al-Ghazali is to re-examine his relation to philosophy and to traditional Ash'arism while at the same time recognizing his basic distance from philosophy.

Al-Ghazali composed three works on Aristotelian logic, *Mi'yar al-'ilm (The Standard Measure of Knowledge)*, *Mihakk al-nazar fi'l-mantiq (The Touchstone of Proof in Logic)* and *al-Qistas al-mustaqim (The Just Balance)*. The first two were written immediately after the *Tahafut* 'in order to help understanding of the latter', and the third was composed after his retirement. He also gave a detailed account of logic in the long introduction of his writing on legal theory, *al-Mustasfa min 'ilm al-usul (The Essentials of Islamic Legal Theory)*. Al-Ghazali's great interest in logic is unusual, particularly when most Muslim theologians were antagonistic to it, and can be attributed not only to the usefulness of logic in refuting heretical views (*al-Qistas* is also a work of refutation of the Isma'ilis), but also to his being fascinated by the exactness of logic and its effectiveness for reconstructing the religious sciences on a solid basis.

There is a fundamental disparity between al-Ghazali's theological view and the Neoplatonic-Aristotelian philosophy of emanationism. Al-Ghazali epitomizes this view in twenty points, three of which are especially prominent: (1) the philosophers' belief in the eternity of the world, (2) their doctrine that God does not know

particulars, and (3) their denial of the resurrection of bodies. These theses are ultimately reducible to differing conceptions of God and ontology. Interestingly, al-Ghazali's criticism of philosophy is philosophical rather than theological, and is undertaken from the viewpoint of reason.

First, as for the eternity of the world, the philosophers claim that the emanation of the First Intellect and other beings is the result of the necessary causality of God's essence, and therefore the world as a whole is concomitant and coeternal with his existence (see Creation and conservation, religious doctrine of). Suppose, say the philosophers, that God created the world at a certain moment in time; that would presuppose a change in God, which is impossible. Further, since each moment of time is perfectly similar, it is impossible, even for God, to choose a particular moment in time for creation. Al-Ghazali retorts that God's creation of the world was decided in the eternal past, and therefore it does not mean any change in God; indeed, time itself is God's creation (this is also an argument based on the Aristotelian concept of time as a function of change). Even though the current of time is similar in every part, it is the nature of God's will to choose a particular out of similar ones.

Second, the philosophers deny God's knowledge of particulars or confine it to his self-knowledge, since they suppose that to connect God's knowledge with particulars means a change and plurality in God's essence. Al-Ghazali denies this. If God has complete knowledge of a person from birth to death, there will be no change in God's eternal knowledge, even though the person's life changes from moment to moment.

Third, the philosophers deny bodily resurrection, asserting that 'the resurrection' means in reality the separation of the soul from the body after death. Al-Ghazali criticizes this argument, and also attacks the theory of causality presupposed in the philosophers' arguments (see Causality and necessity in Islamic thought). The so-called necessity of causality is, says al-Ghazali, simply based on the mere fact that an event A has so far occurred concomitantly with an event B. There is no guarantee of the continuation of that relationship in the future, since the connection of A and B lacks logical necessity. In fact, according to Ash'arite atomistic occasionalism, the direct cause of both A and B is God; God simply creates A when he creates B. Thus theoretically he can change his custom (*sunna*, *'ada*) at any moment, and resurrect the dead: in fact, this is 'a second creation'.

Al-Ghazali thus claims that the philosophers' arguments cannot survive philosophical criticism, and Aristotelian logic served as a powerful weapon for this purpose. However, if the conclusions of philosophy cannot be proved by reason, is not the same true of theological principles or the teachings of revelation? How then can the truth of the latter be demonstrated? Herein lies the force of al-Ghazali's critique of reason.

4 Relation to philosophy

Philosophy declined in the Sunni world after al-Ghazali, and his criticism of philosophy certainly accelerated this decline. Nearly a century later, Ibn Rushd (Averroes) made desperate efforts to resist the trend by refuting al-Ghazali's *Tahafut* in his *Tahafut al-tahafut (The Incoherence of the Incoherence)* and *Fasl al-maqal (The Decisive Treatise)*, but he could not stop it. Philosophy was gradually absorbed into Sufism and was further developed in the form of mystical philosophy, particularly in the Shi'ite world (see Mystical philosophy in Islam). In the Sunni world also, Aristotelian logic was incorporated into theology and Sufism was partially represented philosophically. In all this, al-Ghazali's influence was significant.

Ghazali committed himself seriously to Sufism in his later life, during which time he produced a series of unique works on Sufism and ethics including *Mizan al-'amal (The Balance of Action)*, composed just before retirement, *Ihya' 'ulum al-din*, his *magnum opus* written after retirement, *Kitab al-arba'in fi usul al-din (The Forty Chapters on the Principles of Religion)*, *Kimiya'-yi sa'adat* (The Alchemy of Happiness), *Mishkat al-anwar (The Niche of the Lights)* and others. The ultimate goal of humankind according to Islam is salvation in paradise, which is depicted in the Qur'an and Traditions as various sensuous pleasures and joy at the vision of God. The greatest joy for al-Ghazali, however, is the seeing of God in the intellectual or spiritual sense of the beatific vision. In comparison with this, sensuous pleasures are nothing. However, they remain necessary for the masses who cannot reach such a vision.

Resurrection for Ibn Sina means each person's death - the separation of the soul from the body - and the rewards and punishments after the 'resurrection' mean the pleasures and pains which the soul tastes after death. The soul, which is in contact with the active intellect through intellectual and ethical training during life, is liberated from the body by death and comes to enjoy the bliss of complete unity with the active intellect. On the other hand, the

soul that has become accustomed to sensual pleasures while alive suffers from the pains of unfulfilled desires, since the instrumental organs for that purpose are now lost. Al-Ghazali calls death ‘the small resurrection’ and accepts the state of the soul after death as Ibn Sina describes. On the other hand, the beatific vision of God by the elite after the quickening of the bodies, or ‘the great resurrection’, is intellectual as in the view of the philosophers. The mystical experience (*fana’*) of the Sufi is a foretaste of the real vision of God in the hereafter.

A similar influence of philosophy is also apparent in al-Ghazali’s view of human beings. Human beings consist of soul and body, but their essence is the soul. The human soul is a spiritual substance totally different from the body. It is something divine (*amr ilahi*), which makes possible human knowledge of God. If the soul according to al-Ghazali is an incorporeal substance occupying no space (as Ibn Sina implies, though he carefully avoids making a direct statement to that effect), then al-Ghazali’s concept of the soul is quite different from the soul as ‘a subtle body’ as conceived by theologians at large. According to al-Ghazali, the body is a vehicle or an instrument of the soul on the way to the hereafter and has various faculties to maintain the bodily activities. When the main faculties of appetite, anger and intellect are moderate, harmonious and well-balanced, then we find the virtues of temperance, courage, wisdom and justice. In reality, however, there is excess or deficiency in each faculty, and so we find various vicious characteristics. The fundamental cause for all this is love of the world (see Soul in Islamic philosophy).

The purpose of religious exercises is to rectify these evil dispositions, and to come near to God by ‘transforming them in imitation of God’s characteristics’ (*takhalluq bi-akhlaq Allah*). This means transforming the evil traits of the soul through bodily exercises by utilizing the inner relationship between the soul and the body. Al-Ghazali here makes full use of the Aristotelian theory of the golden mean, which he took mainly from Ibn Miskawayh. In order to maintain the earthly existence of the body as a vehicle or an instrument of the soul, the mundane order and society are necessary. In this framework, the traditional system of Islamic law, community and society are reconsidered and reconstructed.

The same is also true of al-Ghazali’s cosmology. He divides the cosmos into three realms: the world of *mulk* (the phenomenal world), the world of *malakut* (the invisible world) and the world of *jabarut* (the intermediate world). He takes this division from the Sufi theorist Abu Talib al-Makki, although he reverses the meanings of *malakut* and *jabarut*. The world of malakut is that of God’s determination, a world of angels free from change, increase and decrease, as created once spontaneously by God. This is the world of the Preserved Tablet in heaven where God’s decree is inscribed. The phenomenal world is the incomplete replica of the world of *malakut*, which is the world of reality, of the essence of things. The latter is in some respects similar to the Platonic world of Ideas, or Ibn Sina’s world of intelligibles. The only difference is that the world of *malakut* is created once and for all by God, who thereafter continues to create moment by moment the phenomenal world according to his determination. This is a major difference from the emanationist deterministic world of philosophy. Once the divine determination is freely made, however, the phenomenal world changes and evolves according to a determined sequence of causes and effects. The difference between this relationship and the philosophers’ causality lies in whether or not the relation of cause and effect is necessary. This emphasis on causal relationship by al-Ghazali differs from the traditional Ash‘arite occasionalism.

The Sufis in their mystical experience, and ordinary people in their dreams, are allowed to glimpse the world of the Preserved Tablet in heaven, when the veil between that world and the soul is lifted momentarily. Thus they are given foreknowledge and other forms of supernatural knowledge. The revelation transmitted by the angel to the prophets is essentially the same; the only difference is that the prophets do not need any special preparation. From the viewpoint of those given such special knowledge of the invisible world, says al-Ghazali, the world is the most perfect and best possible world. This optimism gave rise to arguments and criticism even in his lifetime, alleging that he was proposing a Mu‘tazilite or philosophical teaching against orthodox Ash‘arism. He certainly says in his theological works that it is not incumbent upon God to do the best for humans; however, this does not mean that God will not in fact do the best of his own free will. Even so, behind al-Ghazali’s saying that God does so in actuality, we can see the influence of philosophy and Sufism.

Al-Ghazali’s criticism of philosophy and his mystical thought are often compared to the philosophical and theological thought of Thomas Aquinas, Nicholas of Autrecourt, and even Descartes and Pascal. In the medieval

world, where he was widely believed to be a philosopher, he had an influence through the Latin and Hebrew translations of his writings and through such thinkers as Yehuda Halevi, Moses Maimonides and Raymond Martin of Spain.

See also: Ash'ariyya and Mu'tazila; Causality and necessity in Islamic thought; Ibn Sina; Ibn Rushd; Islam, concept of philosophy in; Mystical philosophy in Islam; Neoplatonism in Islamic philosophy

KOJIRO NAKAMURA

List of works

Al-Ghazali (1094) *Maqasid al-falasifa (The Intentions of the Philosophers)*, ed. S. Dunya, Cairo: Dar al-Ma'arif, 1961.(A precise summary of Islamic philosophy as represented by Ibn Sina.)

Al-Ghazali (1095) *Tahafut al-falasifa (The Incoherence of the Philosophers)*, ed. M. Bouyges, Beirut: Imprimerie Catholique, 1927; trans. S.A. Kamali, *Al-Ghazali's Tahafut al-Falasifah*, Lahore: Pakistan Philosophical Congress, 1963. (Al-Ghazali's refutation of Islamic philosophy.)

Al-Ghazali (1095) *Mi'yar al-'ilm (The Standard Measure of Knowledge)*, ed. S. Dunya, Cairo: Dar al-Ma'arif, 1961.(A summary account of Aristotelian logic.)

Al-Ghazali (1095) *Mihakk al-nazar fi'l-mantiq (The Touchstone of Proof in Logic)*, ed. M. al-Nu'mani, Beirut: Dar al-Nahdah al-Hadithah, 1966.(A summary of Aristotelian logic.)

Al-Ghazali (1095) *al-Iqtisad fi'l-i'tiqad (The Middle Path in Theology)*, ed. I.A. Çubukçu and H. Atay, Ankara: Nur Matbaasi, 1962; partial trans. A.-R. Abu Zayd, *Al-Ghazali on Divine Predicates and Their Properties*, Lahore: Shaykh Muhammad Ashraf, 1970; trans. M. Asín Palacios, *El justo medio en la creencia*, Madrid, 1929.(An exposition of al-Ghazali's Ash'arite theological system.)

Al-Ghazali (1095) *Mizan al-'amal (The Balance of Action)*, ed. S. Dunya, Cairo: Dar al-Ma'arif, 1964; trans. H. Hachem, *Ghazali: Critère de l'action*, Paris: Maisonneuve, 1945.(An exposition of al-Ghazali's ethical theory.)

Al-Ghazali (1095-6) *al-Qistas al-mustaqim (The Just Balance)*, ed. V. Chelhot, Beirut: Imprimerie Catholique, 1959; trans. V. Chelhot, 'Al-Qistas al-Mustaqim et la connaissance rationnelle chez Ghazali', *Bulletin d'Études Orientales* 15, 1955-7: 7-98; trans. D.P. Brewster, *Al-Ghazali: The Just Balance*, Lahore: Shaykh Muhammad Ashraf, 1978.(An attempt to deduce logical rules from the Qur'an and to refute the Isma'ilis.)

Al-Ghazali (1096-7) *Ihya' 'ulum al-din (The Revival of the Religious Sciences)*, Cairo: Matba'ah Lajnah Nashr al-Thaqafah al-Islamiyyah, 1937-8, 5 vols; partial translations can be found in E.E. Calverley, *Worship in Islam: al-Ghazali's Book of the Ihya' on the Worship*, London: Luzac, 1957; N.A. Faris, The *Book of Knowledge, Being a Translation with Notes of the Kitab al-'Ilm of al-Ghazzali's Ihya' 'Ulum al-Din*, Lahore: Shaykh Muhammad Ashraf, 1962; N.A. Faris, *The Foundation of the Articles of Faith: Being a Translation with Notes of the Kitab Qawa'id al-'Aqa'id of al-Ghazzali's Ihya' 'Ulum al-Din*, Lahore: Shaykh Muhammad Ashraf, 1963; L. Zolondek, *Book XX of al-Ghazali's Ihya' 'Ulum al-Din*, Leiden: Brill, 1963; T.J. Winter, *The Remembrance of Death and the Afterlife: Book XL of the Revival of Religious Sciences*, Cambridge: The Islamic Text Society, 1989; K. Nakamura, *Invocations and Supplications: Book IX of the Revival of the Religious Sciences*, Cambridge: The Islamic Text Society, 1990; M. Bousquet, *Ih'ya 'ouloûm ed-din ou vivification de la foi, analyse et index*, Paris: Max Besson, 1951.(Al-Ghazali's summa of the religious sciences of Islam.)

Al-Ghazali (1097) *al-Risala al-Qudsiyya (The Jerusalem Epistle)*, ed. and trans. A.L. Tibawi, 'Al-Ghazali's Tract on Dogmatic Theology', *The Islamic Quarterly* 9 (3/4), 1965: 62-122.(A summary of al-Ghazali's theological system, later incorporated into the *Ihya'*.)

Al-Ghazali (1106-7) *Mishkat al-anwar (The Niche of the Lights)*, ed. A. Afifi, Cairo, 1964; trans. W.H.T. Gairdner, *Al-Ghazzali's Mishkat al-Anwar*, London: The Royal Asiatic Society, 1924; repr. Lahore: Shaykh Muhammad Ashraf, 1952; R. Deladrière, *Le Tabernacle des lumières*, Paris: Éditions du Seuil, 1981; A.-E. Elschazli, *Die Nische der Lichter*, Hamburg: Felix Meiner, 1987.(An exposition of al-Ghazali's mystical philosophy in its last phase.)

Al-Ghazali (1109) *al-Mustasfa min 'ilm al-usul (The Essentials of the Islamic Legal Theory)*, Cairo: al-Matba'ah al-Amiriyyah, 1322-4 AH.(An exposition and standard work of the Islamic legal theory of the Shafi'ite school.)

Al-Ghazali (*c.*1108) *al-Munqidh min al-dalal (The Deliverer from Error)*, ed. J. Saliba and K. Ayyad, Damascus: Maktab al-Nashr al-'Arabi, 1934; trans. W.M. Watt, *The Faith and Practice of al-Ghazali*, London: Allen & Unwin, 1953; trans. R.J. McCarthy, *Freedom and Fulfillment: An Annotated Translation of al-Ghazali's al-Munqidh min al-Dalal and Other Relevant Works of al-Ghazali*, Boston, MA: Twayne, 1980.(Al-Ghazali's spiritual autobiography.)

References and further reading

Abu Ridah, M. (1952) *Al-Ghazali und seine Widerlegung der griechischen Philosophie (Al-Ghazali and His Refutation of Greek Philosophy)*, Madrid: S.A. Blass.(An analysis of al-Ghazali's refutation of philosophy in the framework of his religious thought.)

Campanini, M. (1996) 'Al-Ghazzali', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 19, 258-74.(The life and thought of al-Ghazali is discussed in detail, with a conspectus of his thought through his very varied career.)

Frank, R. (1992) *Creation and the Cosmic System: al-Ghazali and Avicenna*, Heidelberg: Carl Winter Universitätsverlag.(One of the recent works clarifying the philosophical influence upon al-Ghazali, representing a new trend in the study of al-Ghazali.)

Frank, R. (1994) *Al-Ghazali and the Ash'arite School*, Durham, NC: Duke University Press.(A new attempt to prove al-Ghazali's commitment to philosophy and his alienation from traditional Ash'arism.)

Ibn Rushd (*c.*1180) *Tahafut al-tahafut (The Incoherence of Incoherence)*, trans. S. Van den Bergh, *Averroes' Tahafut al-Tahafut*, 2 vols, London: Luzac, 1969.(A translation with detailed annotations of Ibn Rushd's refutation of al-Ghazali's criticism of philosophy.)

Jabre, F. (1958a) *La notion de certitude selon Ghazali dans ses origines psychologiques et historiques (The Notion of Certitude According to al-Ghazali and Its Psychological and Historical Origins)*, Paris: Vrin.(A comprehensive analysis of al-Ghazali's important concept of certitude.)

Jabre, F. (1958b), *La notion de la ma'rifa chez Ghazali (The Notion of Gnosis in al-Ghazali)*, Beirut: Librairie Orientale.(An analysis of the various aspects of the notion of mystical knowledge.)

Laoust, H. (1970) *La politique de Gazali (The Political Thought of al-Ghazali)*, Paris: Paul Geuthner.(An exposition of al-Ghazali's political thought, showing him as an orthodox jurist.)

Lazarus-Yafeh, H. (1975) *Studies in al-Ghazali*, Jerusalem: The Magnes Press.(Literary stylistic analyses applied to al-Ghazali's works.)

Leaman, O. (1985) *An Introduction to Medieval Islamic Philosophy*, Cambridge: Cambridge University Press.(A good introduction to al-Ghazali's philosophical arguments against the historical background of medieval Islamic philosophy.)

Leaman, O. (1996) 'Ghazali and the Ash'arites', *Asian Philosophy* 6 (1): 17-27.(Argues that the thesis of al-Ghazali's distance from Ash'arism has been overdone.)

Macdonald, D.B. (1899) 'The Life of al-Ghazzali, with Especial Reference to His Religious Experiences and Opinions', *Journal of the American Oriental Society* 20: 71-132.(A classic biography, still useful.)

Marmura, M.E. (1995) 'Ghazalian Causes and Intermediaries', *Journal of the American Oriental Society* 115: 89-100.(Admitting the great influence of philosophy on al-Ghazali, the author tries to demonstrate al-Ghazali's commitment to Sufism.)

Nakamura Kojiro (1985) 'An Approach to Ghazali's Conversion', *Orient* 21: 46-59.(An attempt to clarify what Watt (1963) calls 'a crisis of civilization' as the background of al-Ghazali's conversion.)

Nakamura Kojiro (1993) 'Was Ghazali an Ash'arite?', *Memoirs of Research Department of the Toyo Bunko* 51: 1-24.(Al-Ghazali was still an Ash'arite, but his Ash'arism was quite different from the traditional form.)

Ormsby, E.L. (1984) *Theodicy in Islamic Thought: The Dispute over al-Ghazali's 'Best of All Possible Worlds'*, Princeton, NJ: Princeton University Press.(A study of the controversies over al-Ghazali's 'optimistic' remarks in his later works.)

Shehadi, F. (1964) *Ghazali's Unique Unknowable God: A Philosophical Critical Analysis of Some of the Problems Raised by Ghazali's View of God as Utterly Unique and Unknowable*, Leiden: Brill.(A careful philosophical analysis of al-Ghazali's religious thought.)

Sherif, M. (1975) *Ghazali's Theory of Virtue*, Albany, NY: State University of New York Press. (A careful analysis of al-Ghazali's ethical theory in his Mizan and the philosophical influence on it.)

Smith, M. (1944) *Al-Ghazali the Mystic*, London: Luzac.(A little dated, but still a useful comprehensive study of al-Ghazali as a mystic and his influence in both the Islamic and Christian worlds.)

Watt, W.M. (1963) *Muslim Intellectual: A Study of al-Ghazali*, Edinburgh: Edinburgh University Press.(An analysis of al-Ghazali's life and thought in the historical and social context from the viewpoint of sociology of knowledge.)

Zakzouk, M. (1992) *Al-Ghazali's Philosophie im Vergleich mit Descartes (Al-Ghazali's Philosophy Compared with Descartes)*, Frankfurt: Peter Lang.(A philosophical analysis of al-Ghazali's thought in comparison with Descartes with reference to philosophical doubt.)

Alienation

'Alienation' is a prominent term in twentieth-century social theory and social criticism, referring to any of various social or psychological evils which are characterized by a harmful separation, disruption or fragmentation which sunders things that properly belong together. People are alienated from one another when there is an interruption in their mutual affection or reciprocal understanding; they are alienated from political processes when they feel separated from them and powerless in relation to them. Reflection on your beliefs or values can also alienate you from them by undermining your attachment to them or your identification with them; they remain your beliefs or values faute de mieux*, but are no longer yours in the way they should be. Alienation translates two distinct German terms:* Entfremdung *('estrangement') and* Entaußerung *('externalization'). Both terms originated in the philosophy of Hegel, specifically in his Phenomenology of Spirit (1807). Their influence, however, has come chiefly from their use by Karl Marx in his manuscripts of 1844 (first published in 1930). Marx's fundamental concern was with the alienation of wage labourers from their product, the grounds of which he sought in the alienated form of their labouring activity. In both Hegel and Marx, alienation refers fundamentally to a kind of activity in which the essence of the agent is posited as something external or alien, assuming the form of hostile domination over the agent.*

1 Hegel

Hegel's philosophy regards all reality as *Geist*, or 'spirit' or 'mind'; Hegel's concept of spirit is at the same time a model of the human mind and human agency. It views mind or spirit as an activity which posits a reality or object distinct from itself and then for the first time achieves actuality as spirit by knowing this object as itself. Spirit is therefore a twofold activity, of creation or self-expression and of the reconciling self-interpretation of what it has created. The process through which spirit actualizes itself therefore involves an intermediate moment of 'division' (*Entzweiung*) in which its objectivity has been posited as external to it but has not yet been reconciled or taken back into it. The immediate positing of the object is 'externalization' (*Entaußerung*); the experience of it as an alien reality is 'estrangement' (*Entfremdung*). Thus spirit achieves full actuality only through becoming alienated and then overcoming its alienation. Spirit is this process not only at the level of individual activity but even more at a social level. Alienation and its overcoming are a process through which whole societies, peoples and historical traditions actualize freedom by expressing themselves in otherness, losing themselves through alienation and then regaining themselves through spiritual reconciliation (see Hegel, G.W.F. §8).

Hegel's chief use of the theme of alienation in *Phenomenology of Spirit* (1807) is to represent the rise of Christian culture out of classical antiquity as 'spirit in self-alienation' and to portray modernity as the process of overcoming this alienation. For Hegel, the paradigm of alienation is the 'unhappy consciousness', the finite, transitory, individual consciousness which feels itself cut off from its essence, which it posits in a perfect, unchanging deity residing in a 'beyond'. Or in other words, it is a form of misunderstood Christian religiosity which still has not achieved the form of Hegelian speculative pantheism. But the unhappy consciousness is merely the expression at the level of self-consciousness of a social process. 'Spirit in self-alienation' refers primarily to the loss of the social harmony which was present in the Greek *polis*, and its replacement by the supranational despotism of the Roman Empire. Because individuals no longer feel at home in their earthly society, they regard their true home as lying in a 'beyond', a 'kingdom of God'. The positive outcome of this process, however, is a deepening and transformation of individual self-consciousness through the rise of 'subjectivity': the modern conception of each individual as possessing a dignity, and of the social order as having to respect its sacred rights. Modern subjectivity also involves a transformation of the social order. As the religious alienation of the Christian Middle Ages gives way to modern moral consciousness, the right of the individual is reconciled with rational community in the form of the modern state. For Hegel, the decisive event on the spiritual level was the Lutheran Reformation; on the political level it was the French Revolution.

2 Feuerbach

Hegel's concept of alienation (although less often the term itself) is employed in Ludwig Feuerbach's critique of religion (see Feuerbach, L.A. §2). According to Feuerbach, the idea of God is really no more than the idea of our own human essence or species essence (*Gattungswesen*) projected as a supernatural entity distinct from and

opposed to us. Thus for Feuerbach, religion is the 'self-alienation of the human being, the division of the human being from himself'. The real appeal of religion is the appeal of our own self-affirmation, especially our collective or species affirmation, the appeal of a true human community and human love. But in religion this love and affirmation are actually subverted and denied, because they are misdirected toward an imaginary being alien to us. 'To enrich God, man must become poor; that God may be all, man must be nothing.' Feuerbach's critique of religious alienation is also aimed at its harmful moral and social consequences: the devaluation of our earthly wellbeing and of earthly (especially sexual) love in favour of its alienated religious counterpart, and the separation of men and women from each other in the strife of religious sectarianism and the oppressive class divisions in society. The overcoming of religious alienation is for Feuerbach the prerequisite for a transformation of real life.

3 Marx

Karl Marx used the concepts of *Entaußerung* and *Entfremdung* to portray the situation of modern individuals - especially modern wage labourers - who are deprived of a fulfilling mode of life because their life activity as socially productive agents is devoid of any sense of communal action or satisfaction and gives them no ownership over their own lives or their products. In modern society, individuals are alienated in so far as their common human essence, the actual cooperative activity which naturally unites them, is powerless in their lives, which are subject to an inhuman power - created by them, but separating and dominating them instead of being subject to their united will. This is the power of the market, which is 'free' in the sense that it is an autonomous power beyond the control of its human creators, enslaving them by separating them from one another, from their activity and from their products.

4 The alienation of labour

In the manuscript *Alienated Labour* (1844), Marx attempts to portray the social existence of human beings found in modern capitalism, and theorized by political economy, as a form of practical alienation. He distinguishes four aspects of this alienation:

(1) alienation of workers from the product of their labour;
(2) alienation of workers from their own labouring activity;
(3) alienation of individual human beings from their species essence;
(4) alienation of one human being from another.

(1) The natural relation between labour and its product is one of appropriation: labour appropriates its product for the labourer. Alienated labour, however, delivers the worker's product over to an alien power, the power of capital, which not only appropriates the worker's product but also turns it into an alien power over him. 'The object which labour produces stands opposed to it as an *alien thing*, as a *power independent* of the producer.' Further, the more workers produce, the larger grows the power of capital over them, and the more value accrues to this inhuman power, the more valueless the worker becomes. 'The worker becomes poorer the more wealth he produces…. The *increase* in the *value* of the world of things is directly proportional to the *decrease in value* of the human world.'

(2) Marx seeks the root of this relationship and locates it in the labouring activity itself. 'How could the worker stand in an alien relationship to the product of his activity if he did not alienate himself from it in the very act of production?… If the product of work is externalization, production itself must be active externalization, externalization of activity, activity of externalization.' This may be seen as the first anticipation of Marx's materialist conception of history: social and economic relationships always depend ultimately on the productive powers people employ, hence on the kind of labour they perform. But for Marx, alienated activity refers also to the fact that modern labour in factories is dehumanizing to workers, not actualizing their humanity but stunting and warping it. The lives of wage labourers are without meaning.

(3) The objective possibility for meaningful life present in all forms of social labour is that individuals confirm their humanity by contributing to the satisfaction of human needs, not only their own but even more the needs of the species. Alienated labour is meaningless labour because its conscious aim takes the egoistic form of satisfying individual needs, which for the worker are reduced to the barest conditions of physical survival. 'Life itself appears only as a *means of life*'; workers labour only to sustain the absurd cycle of their existence. By 'making species life the means of individual life' alienated labour reverses the natural relationship between the two and

hence alienates individuals from their own species essence. Since Marx regards the relation of each human being to the human species as its relation to nature, he also includes under this heading the alienation of human beings from nature, 'the inorganic body of the human species'.

(4) Finally, alienated labour alienates human beings from one another by making another human being the owner of their product, hence constituting one human being as an alien power over another human being. At the point where the *Alienated Labour* breaks off, Marx is in the process of comparing this relationship to Hegel's famous 'master-servant' dialectic in *Phenomenology*.

5 Alienation in the later Marx

The 'theory' of alienation sketched in this early fragment is Marx's first recognizable attempt to articulate the complex interconnections he sees between the various ills and irrationalities characteristic of modern industrial society. Like Hegel, the early Marx views alienation as a phase of a developmental process leading to human self-actualization. 'Alienation is founded in the essence of human development.' Yet Marx was already breaking away from Hegel and the entire 'Young Hegelian' movement (represented by Feuerbach and Marx's teacher Bruno Bauer (see Bauer, B. §1). In his view, they located the essence of alienation in a certain kind of *consciousness*, and believed that a practical reform of the social world must follow a reform in ideas, led by philosophers. In contrast to this, Marx maintained that alienation was fundamentally a matter of social practice, and that alienated forms of consciousness (such as religion) are merely a symptom of alienated activity. They can be abolished only by abolishing the alienation in real life.

Both the term 'alienation' (*Entfremdung*) and the concept (expressed in other terms) continue to play a prominent role in Marx's mature writings, especially in the *Grundrisse* (1857-8) and *Capital* (1867, 1884, 1894). But their function has changed. The philosophical concept of alienation is no longer viewed as fundamental to a systematic explanation of capitalism; it is now employed descriptively or diagnostically to characterize some of the evils and irrationalities of the system.

6 Alienation and meaningful life

The German verbs *entaußern* and *entfremden* are often reflexive, and for Marx, as for Hegel and Feuerbach, alienation is fundamentally *self*-alienation. To be alienated is to be separated from one's own essence or nature; it is to be forced to lead a life in which that nature has no opportunity to be fulfilled or actualized. In this way, the experience of alienation involves a sense of a lack of self-worth and an absence of meaning in one's life.

Alienation in this sense is not fundamentally a matter of whether your conscious desires are satisfied, or how you experience your life, but rather whether your life objectively actualizes your nature, especially your life with others as a social being on the basis of a determinate course of historical development. The view that alienation, so conceived, can nevertheless have historical consequences, and even be a lever for social change, clearly involves some sort of realism about the human good: it makes a difference, psychologically and socially, whether people actualize their nature, and when they do not, this fact explains what they think, feel and do, and can play a decisive role in historical change.

See also: Marxism, Western §2

ALLEN W. WOOD

References and further reading

Feuerbach, L. (1841) *The Essence of Christianity*, trans. M. Evans, New York: Harper & Row, 1957. (Feuerbach's famous critique of human alienation in the Christian religion.)

Feuerbach, L. (1843) *Principles of the Philosophy of the Future*, trans. M. Vogel, Indianapolis, IN: Bobbs-Merrill, 1966.(The chief text of Feuerbach in which the term alienation is explicitly employed.)

Geuss, R. (1981) *The Idea of a Critical Theory*, Cambridge: Cambridge University Press.(An excellent critical discussion of the kind of social theory in which the concept of alienation figures.)

Hegel, G.W.F. (1807) *Phenomenology of Spirit*, trans. A.V. Miller, Oxford: Oxford University Press, 1977. (Chapters IV, VI B and VII C contain the chief historical sources of the concept of alienation.)

Marx, K. (1857-8) *Grundrisse der Kritik der politischen Ökonomie*, trans. M. Nicholas, *Grundrisse*, Harmondsworth: Penguin, 1973.(An abandoned draft on political economy - superseded by *Das Kapital*.)

Marx, K. (1867, 1884, 1894) *Capital*, 3 vols, trans. B. Foulkes, Harmondsworth: Penguin, 1977.(Marx's mature masterpiece.)

Marx, K. and Engels, F. (1975-) *Collected Works*, New York: International Publishing Co. (Volume 3 contains the principal texts which discuss alienation.)

Ollman, B. (1976) *Alienation*, Cambridge: Cambridge University Press, 2nd edn.(A widely-known philosophical discussion of Marx's concept of alienation.)

Meszaros, I. (1972) *Marx's Theory of Alienation*, New York: Harper & Row.(An attempt to reconstruct the early Marx's theory of alienation as the foundation of his thought.)

Plamenatz, J. (1975) *Karl Marx's Philosophy of Man*, Oxford: Oxford University Press.(A thoughtful critical appraisal of Marx, emphasizing the theme of alienation.)

Wood, A. (1981) *Karl Marx*, London: Routledge.(Chapters I-IV contain an account of alienation.)

Alighieri, Dante (1265-1321)

Although Dante never received a systematic training in philosophy, he tackled some of the most controversial philosophical problems of his time. In his theory of science, he asked how we are to explain the fact that science is a unified, strictly ordered system of knowledge. He answered by comparing the scientific disciplines with the celestial spheres, claiming that the system of knowledge mirrors the cosmological order. In his political philosophy, he asked why all humans want to live in a peaceful society. All humans seek full use of their cognitive capacity, was his answer, and they can achieve it only if they interact socially. In his philosophy of nature, Dante asked what brings about the order of the elements, and suggested that the elements obey the laws of a universal nature in a strictly ordered cosmos. He elaborated all his answers in a scholastic framework that made use of both Aristotelian and Neoplatonic traditions.

1 Philosophical works and intellectual background

Dante Alighieri, best known as the author of the poetic work *Divina Commedia (Divine Comedy)*, was not only one of the most influential Italian poets but was also a scholastic philosopher, perceived and praised by his earliest commentators such as Marsilio Ficino as a 'poetic philosopher'. Three distinctively philosophical works of Dante's have been transmitted: the *Convivio (The Banquet)*, a philosophical symposium that addresses all who strive for knowledge but lack higher education; *Monarchia (The Monarchy)*, a treatise on political philosophy that aims to prove the need for a universal monarchy; and the *Questio de aqua et terra (Question on Water and Earth)*, a short work in natural philosophy that discusses the order among the elements. Since no medieval manuscript of this work is known (it has been preserved in an early printed edition), its authenticity has been disputed. In addition, the first part of his treatise *De vulgari eloquentia (On Eloquence in the Vernacular)*, dealing with the origin and function of language, includes important semantic theses, and in a long letter addressed to Cangrande della Scala, Dante expounds his hermeneutic principles.

Unlike the overwhelming majority of medieval philosophers, Dante was never a member of a monastery or a university and never received a systematic training in philosophy. Rather, he was a layman (*laicus*) in the medieval dual sense of the word: neither a clergyman nor a professional intellectual. He acquired his philosophical knowledge, as he confesses in an autobiographical passage (*Convivio* II, xii, 5), by frequenting the 'disputations of the philosophizing people' in Florence (probably in the Franciscan school of Santa Croce and in the Dominican school of Santa Maria Novella) between 1291 and 1295. He may have furthered his education during a stay in Paris (1310). However, his knowledge remained eclectic. He appears to have been more influenced by a reading of anthologies and texts by popular philosophical authors (such as Brunetto Latini) than by a systematic study of Aristotle and contemporary philosophers.

2 Scope and division of philosophy

Dante confesses (*Convivio* II, xii, 1-10) that after the death of Beatrice (the beloved woman who became Dante's guide in the Paradise, the third part of the *Divina Commedia*) in 1290, he could not find consolation until he started reading philosophical books. Cicero, Boethius and other classical authors helped him overcome his grief. He discovered that philosophy is a 'noble lady' (*donna gentile*):

> Only in glancing at her does one acquire human perfection, that is the perfection of reason, on which depends our essence as on its most fundamental part. And all our other activities, such as perceiving and nourishing, are founded upon reason. But reason exists through herself and not through something else. If reason is perfect, she is so perfect that a human being qua human being realizes that his desire is satisfied in her. And so a human is happy.
>
> (*Convivio* III, xv, 4)

This passage is not just a literary *topos* that alludes to Boethius' presentation of Lady Philosophy in the *De consolatione philosophiae (Consolation of Philosophy)*. It is also deeply rooted in the Aristotelian dictum that all human beings naturally seek knowledge - Dante opens both the *Convivio* and *Monarchia* with this famous statement - and that humans find happiness only if they satisfy their intellectual needs. Moreover, Dante's praise of intellectual activity betrays his acquaintance with certain late thirteenth-century authors sometimes labelled

‘radical Aristotelians’, such as Boethius of Dacia and Aubry of Reims (see Averroism). These Parisian masters claimed that philosophy is autonomous and should not be subordinated to any other discipline, in particular not to theology, because it provides humans with all the knowledge required for obtaining happiness. Since humans are essentially rational animals, they fully realize their capacities if they dedicate themselves to the most rational activity, philosophy.

Dante conceives of philosophy - ‘philosophy’ taken synonymously with ‘science’ (*scientia*) - as a unified system of knowledge that can be hierarchically divided. He explains the order by comparing these disciplines with the celestial spheres. There are at least three reasons, he claims, for making such a comparison (*Convivio* II, xiii, 3-6). First, every science moves around its subject, just as a celestial sphere turns around its immovable centre. Second, every science makes its subject clear and understandable, just as a celestial sphere illumines the things around which it turns. Third, every science leads to a certain perfection by structuring and explaining things, just as a celestial sphere improves things by keeping them in order and moving them. Not only is there, according to Dante, an analogy between science as a whole and the celestial spheres, but one can even attribute a specific science to each of the nine celestial spheres. There is a correspondence between each of the first seven spheres (Moon, Mercury, Venus, Sun, Mars, Jupiter and Saturn) and each of the seven liberal arts (grammar, logic, rhetoric, arithmetic, music, geometry and astrology). Physics and mathematics, dealing with innumerably many entities, correspond to the eighth sphere that consists of innumerable fixed stars. Finally, moral philosophy (including both ethics and political philosophy) corresponds to the ninth, so-called ‘crystalline sphere’, because moral philosophy enables humans to study all the other sciences just as this ninth unmoved sphere moves all the other spheres.

Not only does Dante believe, as did many medieval philosophers of science (for example, Robert Kilwardby), that philosophy (or science) is a unified, strictly ordered system of knowledge, but he also claims that the scientific system mirrors the cosmological system (see Natural philosophy, medieval). Whereas he follows the traditional scheme in his ordering of the seven liberal arts, he deliberately deviates from tradition in his claim that moral philosophy and not metaphysics, usually taken as ‘first philosophy’, is the highest scientific discipline. Although it is important to know, in the Aristotelian framework of metaphysics, what being *qua* being and its principles are, this is not the highest goal in philosophy. One obtains perfect knowledge, Dante claims, only if one is aware of the principles leading to a good life. That is why ‘morality is the beauty of philosophy’ (*Convivio* III, xv, 11). When one takes this important claim into account, it becomes clear why Dante emphatically holds that human happiness is to be sought through philosophy. Since moral philosophy, the culmination of every scientific activity, enables humans to recognize the principles governing a good life in a peaceful community, humans must philosophize if they want to attain such a life (see Happiness).

3 Political philosophy

In the *Monarchia*, a treatise that is highly scholastic in both style and content, Dante aims to show first, that a universal monarchy is necessary for a peaceful community of humans, second, that the Roman nation legitimately acquired its hegemony over all other nations, and third, that the power of the Roman emperor is not subordinate to that of the pope. Thus the main purpose of the work is clearly political and has to be understood in the light of the late medieval controversies over mundane and spiritual power (notably those between King Philip the Fair of France and Pope Boniface VIII). Yet in his defence of the Roman emperor, Dante makes a rich use of philosophical arguments, often going beyond the concrete political context. This becomes clear from his discussion of the principle governing human behaviour, a principle that underlies all his political theses, as he explicitly says (*Monarchia*, I, ii, 8).

Dante starts his search for such a principle by asking why humans want to live in a community. Following the Aristotelian tradition, he gives a teleological answer; all humans have certain goals they want to obtain. However, unlike most medieval authors in the Aristotelian tradition, he claims that there is not only a goal for each individual, each household and each city, but also a goal for the entire human species. This goal is to be attained only through an activity of the entire species. Such an activity consists in all realization of the ‘possible’ intellect (*Monarchia* I, iii, 8), that is, in an exhaustive development of one’s cognitive capacity. An isolated person, limited in activity, is never able to achieve such development, which requires one’s integration into a peaceful community in which all humans participate.

It is remarkable and original in the medieval context that Dante uses an epistemological argument to justify his

political theory. Neither a natural order established by God, nor a need for protection, nor a need for exchanging goods and making contracts is the main reason for which humans want to live in a peaceful community. Rather, the main reason is an intellectual need; humans want to use their cognitive capacity. Following Aristotle, Dante claims that this capacity is nothing other than the possible intellect that is 'realized', or put to active use, when it receives the forms of the cognized things. But, unlike Aristotle, he does not take the use of this intellect to be a mere individual operation; and, contrary to some radical Aristotelians (for example, Boethius of Dacia or Siger of Brabant), he does not hold that every person is an autonomous cognizer, capable of acquiring all possible knowledge through an individual contemplation of the highest principles. Nor does he make the controversial Averroistic claim, as his first critics (such as Guido Vernani) complained, that there is just one possible intellect for all humans, so that every individual would be dependent upon this all-embracing intellect. Rather, Dante holds that every person has an individual possible intellect; but one can make full use of one's own possible intellect only if one is a member of a peaceful community. Thus, social interaction is a *sine qua non* for successful cognitive activity (see Political philosophy, history of).

4 Natural philosophy

In the *Questio*, a written record of a scholastic debate, Dante (if indeed the text was written by him) raises the question of how the elements water and earth are related to each other. Which element has a higher location, water or earth? At first sight, this question seems puzzling or even pointless. Is it not obvious that there are some places (such as the mountains) where earth is higher, and others (such as the sea) where water is higher? But considered from a medieval point of view, the question points to a serious problem. According to the geocentric model that was strongly influenced by the Euclidean theory of spheres, the terrestrial world is located in the middle of a spherical cosmos and consists of four spherically ordered elements: earth, water, air and fire. The element water completely surrounds and covers the element earth, located in the middle of all the spheres. Thus we are confronted with a conflict between the common sense view that there is no fixed order among the elements, and the theoretical view that there must be a strict order according to the spherical model.

In light of this conflict, it is not surprising that the question raised by Dante was a prominent puzzle in medieval philosophy of nature (Duhem 1958: 79-235). Dante tried to resolve this dilemma by introducing a distinction between two kinds of nature. The 'universal nature' is present in the celestial spheres and determines the movement of all terrestrial things. The 'particular nature', on the other hand, is present in each terrestrial thing and makes it tend toward the place it is suited for. So, according to the particular nature, inherent in the elements, earth tends toward the middle of all elements and is completely covered by water. However, the universal nature is able to influence and change this tendency so that there can be some dry places (the so-called 'habitable parts') where earth emerges from water and, therefore, has a higher location than water (*Questio* 49). Dante claims that this general nature is to be found in the eighth sphere, the 'crystalline sphere' (*Questio* 69).

This explanation allows Dante to maintain the geocentric spherical model without rejecting the common sense view. It also allows him to avoid the solution suggested by earlier authors such as Andalò del Negro, that the two spheres (earth and water) may not be concentric. Such a solution would threaten the entire cosmological model. If there is no concentricity, neither is there any guarantee that all the spheres - and not just one element - will revolve around the earth. While insisting on the concentric order, Dante introduces a distinction between two kinds of nature that betrays a Neoplatonic background: God, the first cause of everything, created the celestial spheres, which act as secondary causes. Thus the cosmos is taken to be a strictly ordered, closed system in which the place and movement of every singular thing or element can be explained in terms of its dependency on certain causes (see Natural philosophy, medieval).

See also: Cosmology; Language, medieval theories of (§17); Natural philosophy, medieval; Political philosophy, history of

DOMINIK PERLER

List of works

Alighieri, Dante (1265-1321) *Opere minori (Minor Work)*, ed. C. Vasoli *et al.*, Milan: R. Ricciardi, 1979-84. (Authoritative edition of all Dante's philosophical works with extensive commentaries. *Philosophische Werke*, ed. R. Imbach, Hamburg: F. Meiner, 1993-, is a bilingual edition with extensive commentaries focusing on

philosophical aspects.)
Alighieri, Dante (1304-7) *Convivio (The Banquet)*, ed. C. Vasoli and D. de Robertis, *Opere minori* I.2, Milan: R. Ricciardi, 1988.(Philosophical symposium including an extensive discussion of problems in the theory of science.)
Alighieri, Dante (*c*.1305) *De vulgari eloquentia (On Eloquence in the Vernacular)*, ed. P.V. Mengaldo, *Opere minori* II, Milan: R. Ricciardi, 1979, 26-237.(Treatise on the origin and function of natural languages.)
Alighieri, Dante (1310-7) *Monarchia (The Monarchy)*, ed. B. Nardi, *Opere minori* II, Milan: R. Ricciardi, 1979, 280-503.(Treatise on political philosophy that makes original use of Aristotelian psychology.)
Alighieri, Dante (before 1321) *Questio de aqua et terra (Question on Water and Earth)*, ed. F. Mazzoni, *Opere minori* II, Milan: R. Ricciardi, 1979, 744-73.(Based on a disputation held in Verona in 1320; short scholastic text that discusses the order among the elements.)

References and further reading

Bosco, U. (1970-8) *Enciclopedia Dantesca (Dante Encyclopedia)*, Rome: Istituto della Enciclopedia Italiana, 6 vols.(Most comprehensive presentation, including many articles dealing with philosophical aspects of Dante's work.)
Boyde, P. (1981) *Dante Philomythes and Philosopher, Man in the Cosmos*, Cambridge: Cambridge University Press.(Detailed study of Dante's natural philosophy and cosmology with an extensive analysis of the scholastic background.)
Boyde, P. (1993) *Perception and Passion in Dante's Comedy*, Cambridge: Cambridge University Press.(Study focusing on Dante's philosophy of mind and epistemology).
Corti, M. (1983) *La felicità mentale. Nuove prospettive per Cavalcanti e Dante (Intellectual Happiness: New Perspectives on Cavalcanti and Dante)*, Turin: Einaudi.(Study focusing on Dante's praise of the philosopher's life; challenging theses about the origin and medieval background of this attitude.)
Duhem, P. (1958) *Le système du monde. Histoire des doctrines cosmologiques de Platon à Copernic (The System of the World: A History of Cosmological Doctrines from Plato to Copernicus)*, vol. 9, Paris: Hermann.(Pioneer study of the ancient and medieval background of the problem discussed in the *Questio*.)
Gilson, E. (1939) *Dante et la philosophie (Dante and Philosophy)*, Paris: J. Vrin.(Pioneering study that provides a detailed interpretation of all philosophical works.)
Imbach, R. (1989) *Laien in der Philosophie des Mittelalters (Laymen in Medieval Philosophy)*, Amsterdam: B.R. Grüner.(Important study of Dante's social and intellectual context and his definition of philosophy.)
Ficino, M. (1433-99) *Prohemia sopra la Monarchia (Commentary on the Monarchia)*, ed. P.A. Shaw in 'La versione ficiniana della "Monarchia"', *Studi danteschi* 51, 1978: 277-408.(Late fifteenth-century commenary on the *Monarchia*, presenting Dante as a 'poetic philosopher'.)
Nardi, B. (1967) *Saggi di filosofia dantesca (Essays in Medieval Philosophy)*, Florence: La Nuova Italia. (Provocative essays dealing with a variety of philosophical problems, taking into account both the distinctively philosophical and the literary works.)
Nardi, B. (1985) *Dante e la cultura medievale (Dante and Medieval Culture)*, ed. P. Mazantini, Bari/Rome: Laterza.(A collection of important essays focusing on epistemology, theories of language and the soul.)

Alison, Archibald (1757-1839)

Archibald Alison was born in Edinburgh but was educated at Balliol and ordained in the Church of England. He returned to Edinburgh in 1800 as an Anglican clergyman and served there until his death. His published works included collections of sermons, but he is best known for his Essays on the Nature and Principles of Taste. This work was published in 1790, the same year as Immanuel Kant's third Critique; but it became popular only after a second edition appeared in 1811.

Archibald Alison, Anglican clergyman, born in 1757 in Edinburgh, broke with earlier eighteenth-century theorists of taste in two respects. He denied that taste is a product of an internal sense, and he described the emotion of taste as complex rather than simple. Earlier theorists had developed taste using the analogy of sense perception. The exact nature of this sense varied. In some cases it was taken quite literally; in others, it was little more than a convenient analogy. In general, however, an internal sense was a reflexive, immediate response of the mind to qualities presented by objects. Alison abandoned such a sense altogether. Instead, the emotions of taste are the product of mental operations to which the mind contributes and in which the mind discovers its own qualities. Such emotions are inherently complex. Alison reasoned that if a simple perceptual quality such as colour were the source of an emotion of taste, that emotion would always accompany perceptions of colour. Such is not the case. Only when the mind operates on the perception in a certain way does the emotion of pleasure identified with taste occur.

In place of a simple emotion and an internal sense, Alison introduced expression, imagination, and association as the key aesthetic terms. Imagination is a faculty which acts upon simple emotions. It suggests other images which are not directly present but which share the emotional qualities of the perception. The emotional links themselves are formed by association. So both imagination and association are necessary conditions for emotions of taste. When, in addition, the emotions are expressive of qualities of mind, one has emotions of taste - what Immanuel Kant and later theorists came to call aesthetic emotions. Alison projected the following schema: natural objects are suited to produce simple emotions; those simple emotions are extended by association and take on qualities of mind (for example tenderness); the imagination is the faculty which accomplishes this extension. When the imaginative associations are unified throughout an occasion, the result will be a special emotion of taste - either beauty or sublimity, depending on the emotion. Pleasure always accompanies this extension.

Alison went substantially beyond earlier theorists in the way that he developed expressiveness as a quality of mind and in the way that he used association. Earlier eighteenth-century theorists, including Joseph Priestley, Alexander Gerard, and Thomas Reid, all spoke of a certain kind of mental exertion as intrinsically pleasurable and thus as a key element in taste. They were thinking of a physiological phenomenon, however. Strong, violent mental exertion could not produce the calm passions of beauty, but neither could a too languid mental operation. The emotions of taste were conceived of as belonging to the middle range - neither too strong, nor too mild. Alison, on the other hand, thought not of the exertion of the mind, but of qualities which, through association, became expressive of the mind's own powers. A natural quality becomes beautiful by acquiring mental associations.

Alison's theory of association is also substantially different as a consequence. Earlier theories of association, particularly that of David Hartley, sought to account for how the mind could produce ideas in the absence of immediate experience. The function of association was thus to extend experience and to provide a mechanism for the recall and production of mental images. Hartley's account is strongly mechanistic, for example. Alison used association differently. By itself, a colour would produce no emotion, he reasoned. But association connects perception with other experiences so that the emotional qualities of the complex are produced by the perception as well. So red can be exciting, white pure, and so forth. Association brings together disparate images and ideas into a complex which has emotional consequences. Its faculty is imagination, and its consequence is expression. Beauty is the complex result of a kind of associative network. These associations give rise to a multitude of predicates, and Alison made extensive use of language as an indicator of what emotional associations a particular kind of perception had acquired.

Alison distinguished natural from relative beauty. Natural beauty arises from the associations the mind forms in its direct encounters with the world. Alison's paradigms for natural beauty are scenery and gardens. Relative beauty

depends on associations which suggest a fitness and utility; it includes design, skill, and art. Thus Alison established an implicit hierarchy, in which art is secondary to nature as an aesthetic source. Design, fitness, and utility presuppose a mental order and a mind, so they are expressive and beautiful. But their expressiveness depends on complex emotions such as tenderness, grandeur, and majesty, which are first inspired by the imagination acting on natural scenes. The common feature is that both produce a special form of pleasure, which Alison, like Kant, called delight.

Alison continued a century-old tradition of speculation about taste and beauty along empiricist lines. He shared with that tradition a reliance on experience and an attempt to classify and codify elements of experience into a theory of beauty. Alison extended that empiricism in a significant way, however. His analysis relies on language, particularly emotive predicates. Association is based on similarity - which may be perceptual or metaphorical, but is always created by the mind itself. The imagination becomes a creative faculty, and aesthetic theory is less concerned with standards and epistemological questions than with the production of emotion. Alison remained a moralist. Nature, not art, is the source of emotion, and nature is God's handiwork, not ours. But it is a short step from Alison's forms of imagination and expressiveness to that of the Romantics who use many of the same terms. The aesthetic ground has shifted substantially.

See also: Emotion in response to art; Taste, artistic

DABNEY TOWNSEND

List of works

Alison, A. (1790) *Essays on the Nature and Principles of Taste*, London.(Alison's principal work went through a number of editions in the nineteenth century, but there is no modern reprint.)

References and further reading

Dickie, G. (1974) *Art and the Aesthetic: An Institutional Analysis*, Ithaca, NY: Cornell University Press.(In the first part of his theory, Dickie traces the rise of the concept of aesthetic experience and rejects Stolnitz's interpretation of Alison.)

Hipple, W. (1957) *The Beautiful, the Sublime, and the Picturesque in Eighteenth-century British Aesthetic Theory*, Carbondale, IL: Southern Illinois University Press.

Kant, I. (1790) *Kritik der Urteilskraft*, Berlin: Lagarde; trans. J.C. Meredith, as *Critique of Aesthetic Judgement*, Oxford: Clarendon Press, 1911; repr. as *Kant's Critique of Judgement*, trans. J.C.Meredith, Oxford: Clarendon Press, 1952.(Kant's *Critique of Judgement* is his primary critical treatment of aesthetic issues. It is the third and final of Kant's *Critiques*, in which he integrates the realms of theory and value, both aesthetic and moral.)

Kivy, P. (1976) *The Seventh Sense*, New York: Franklin.(Traces the concept of internal sense through eighteenth-century aesthetics.)

McCosh, J. (1875) *The Scottish Philosophy* New York: Carter; repr. Bristol: Thoemmes, 1990; New York: AMS Press, 1980.(Brief biographical sketch and commentary.)

Stolnitz, J. (1961) 'On the Origins of Aesthetic Disinterestedness', *Journal of Aesthetics and Art Criticism* 20: 131-43.(Stolnitz identifies Alison as the first aesthetic attitude theorist.)

Townsend, D. (1988) 'Archibald Alison: Aesthetic Experience and Emotion', *British Journal of Aesthetics* 28: 132-44.(Places Alison's theory in the context of shifting empiricist problems.)

al-Juwayni, Abu’l Ma‘ali (1028-85)

Al-Juwayni rose to great prominence as a theologian in the Islamic world, and his theoretical discussions of philosophical issues played a significant role in the development of Islamic philosophy. He provided a stout defence of the Ash‘arite theory that emphasizes the power of God and the insignificance of human beings. His work on the meaning of scriptural texts provided Muslims with a sophisticated and productive series of concepts with which to discuss issues of interpretation.

1 Life and works

Born in Juwayn in Persia, al-Juwayni spent his life defending the principles of Ash‘arism (see Ash‘ariyya and Mu‘tazila). By the time of his death he was widely known as the Imam al-Haramayn, the imam of the two great mosques of Mecca and Medina. This gives some indication of the influence of his thinking in the Islamic world. He worked in an interesting time of transition between the original Ash‘arite *kalam* (theology) (see Islamic theology) and the more radical developments which were to be undertaken by al-Ghazali.

Al-Juwayni argued that there are some forms of knowledge which are available to contingent beings such as ourselves, yet this knowledge is itself irretrievably contingent and should be distinguished from the sort of knowledge which God has. Although God is not a body and is neither a spatial nor temporal being, it is nonetheless possible for him to be seen, in the next life, through beatific vision. God is completely free, acting for no reason other than that which he gives himself. There is nothing necessary about causality, and the possibility of miracles is based upon the fact that there is nothing fixed about nature. God is not only the creator of the universe in the sense of being the first cause, but he is also the agent who is the cause of its continuous existence. The existence of the world at every moment depends upon God’s will. God is the sole creator, and even our actions do not really originate with us but are acquired from God.

The route to an understanding of the Qur’an is through a thorough grounding in the Arabic language. Al-Juwayni distinguishes between different kinds of text. Some texts are obvious and clear, some are accurate, some are concealed and yet others are obscure. Texts which are clear do not change their sense, whatever context they appear in. Those which are accurate have a sense which is clearly linked to a particular state of affairs which the text describes, and present no difficulties. Concealed texts have two sorts of meaning, one which requires interpretation by a prophet and his followers, or one which is capable of explanation by a body of readers who really understand the difficulties of what is before them. Obscure texts require *ta’wil* or analogical interpretation, in accordance with which the correct meaning will be carefully extended from the actual forms of words which are used. This form of interpretation should only be used as a last resort, and it is replete with dangers in that it can lead to a loose and undisciplined approach to understanding the meaning of scripture. Al-Juwayni presents in his work a highly organized system of hermeneutics designed to make scripture accessible and yet also restricted within particular theological boundaries.

2 Ethics

Al-Juwayni was a staunch defender of the Ash‘arite view of the basis of value judgments, which is entirely scriptural. What is good is what is said to be good in scripture, and what is bad is what scripture condemns, and there is no other basis to such judgments. Any attempt at finding a rational foundation is flawed, and it can be assumed that here he had the Mu‘tazilites in mind. He suggests that the Mu‘tazilites are wrong in thinking that there are basic rational moral truths, since if this were the case there would be no possibility of widespread moral disagreement, something which quite clearly does arise. Similarly, the idea that particular forms of behaviour are absolutely right or wrong is difficult to establish, given that we often base our judgments here on the context surrounding those actions and the precise nature of the agent (see Ethics in Islamic philosophy). For example, an adult and a child may perform a similar action; the former action may be called evil, but this is not appropriate as a description of the child’s action.

In his account of what it means for God to be obliged to act in certain ways, al-Juwayni totally rejects such language. He argues that it is mistaken to talk about God being under any obligation to his creatures at all. He often pokes fun at the very idea of explaining the sometimes tragic events of this world as part of an objective

divine plan. God does not need to operate in accordance with such a plan; this would be to deny the uniqueness of God and his radical separation from his creatures. Al-Juwayni argues that his thesis has the advantage of not needing to provide implausible explanations of why things are as they are in the world. The world is as it is because of the decisions of the deity, but we cannot fathom his reasons for those decisions and it is entirely inappropriate to hold him liable to adhere to an essentially human system of justice.

See also: Ash'ariyya and Mu'tazila; Omnipotence

OLIVER LEAMAN
SALMAN ALBDOUR

List of works

al-Juwayni (1028-85) *Kitab al-irshad ila qawati 'al-adilla fi usul al-i'tiqad (The Guide to the Cogent Proofs of the Principles of Faith)*, ed. M. Musa and A. 'Abd al-Hamid, Cairo, 1950.(A detailed account of al-Juwayni's theological principles.)

l-Juwayni (1028-85) *al-Burhan fi usul al-fiqh (The Proofs of Jurisprudence)*, ed. A. al-Dib, Cairo: Dar al-Ansar, 1980.(Al-Juwayni's doctrine of legal judgement.)

References and further reading

Allard, M. (1965) *Le problème des attributs divins dans la doctrine d'al-Ash'ari et de ses premiers grands disciples (The Problem of Divine Attributes in the Doctrine of al-Ash'ari and His First Major Disciples)*, Beirut: Recherches publiées sous la direction de l'Institut de Lettres Orientales de Beyrouth.(A classic introduction to Ash'arite thought.)

Gardet, L. and Anawati, G. (1948) *Introduction à la théologie musulmane (Introduction to Muslim Theology)*, Paris: Études de philosophie médiévale.(Excellent account of al-Juwayni's place in Ash'arite thought.)

Hourani, G. (1985) 'Juwayni's criticisms of Mu'tazilite ethics', in G. Hourani (ed.) *Reason and Tradition in Islamic Ethics*, Cambridge: Cambridge University Press, 124-34.(Al-Juwayni's critique of Mu'tazilite ethics.)

al-Kindi, Abu Yusuf Ya‘qub ibn Ishaq (d. *c*.866-73)

Practically unknown in the Western world, al-Kindi has an honoured place in the Islamic world as the 'philosopher of the Arabs'. Today he might be viewed as a bridge between Greek philosophers and Islamic philosophy. Part of the brilliant ninth-century 'Abbasid court at Baghdad, composed of literati of all types, he served as tutor for the caliph's son. He gained insights into the thought of Greek philosophers, especially Aristotle, through the translation movement; although he did not make translations himself, he corrected them and used them advantageously in his own thought.

Al-Kindi is notable for his work on philosophical terminology and for developing a vocabulary for philosophical thought in Arabic, although his ideas were superseded by Ibn Sina in the eleventh century. The debate about the allowability of philosophy in terms of orthodox Islam also began with al-Kindi, a battle that is usually considered to have been won for religion by al-Ghazali. Like other innovators, his ideas may no longer appear revolutionary, but in his own day, to push for the supremacy of reason and for the importance of a 'foreign science' - philosophy - as opposed to an 'Arab science' - grammar, Qur'anic studies - was quite astonishing. When the Khalif al-Mutawwakil came to power and sought to restore traditionalism, al-Kindi suffered a reversal of fortunes.

1 Logic and translation

Abu Yusuf Ya‘qub ibn Ishaq al-Kindi was an ethnic Arab (died in Baghdad between AH 252-60/AD 866-73), with an illustrious lineage going back to such near-mythic Arabian families as Qays. Al-Kindi was known as 'the philosopher of the Arabs' in contrast to the later Islamic philosophers who, though Muslim, were not Arabs and often learned Arabic as a second language. The early bio-bibliographers gave his ancestry and a long list of works, many of which are no longer extant, but his personal life remains unknown. Although he is remembered for introducing philosophy to the 'Abbasid court, his skills covered many fields including medicine, mathematics, music, astrology and optics. He also served as tutor to the son of the Khalif al-Mu‘tasim. Al-Qifti, one of the medieval Islamic bio-bibliographers, pointedly asserted that al-Kindi was skilled in the arts of the Greeks, the Persians and the Hindus.

Al-Kindi used early, Arabic-language translations of Greek philosophy, which enabled him to add part of the Hellenistic tradition to his programme. The founding of the *bayt al-hikma* (house of wisdom), for the large-scale translation of documents from Greek, in the early ninth century meant both that the 'foreign sciences' were available wholesale to Arabophone scholars and that there was serious interest in the knowledge they contained. Al-Kindi was occasionally credited (in the title inscription) with correcting the translation, but it is generally accepted that he did not read Greek himself. The pursuit of 'foreign sciences' was also politically acceptable at this juncture, which ceased to be the case later. A study of his terminology shows that al-Kindi was aware of particular terms used in Hellenistic philosophy, and of which Arabic word best expressed the same idea.

Al-Kindi may be thought of as a stage-setter for philosophy in the Islamic world, laying out terms *qua* terms and redirecting the metaphysical concerns suggested by the *mutakallimun* (theologians) from the realm of religion to that of philosophy. His lack of interest in religious argument can be seen in the topics on which he wrote. These topics were ontological, but he generally refrained from eschatological discussions on topics such as the resurrection, the last day and the last judgment. Even in his ethical treatise he dealt with the disciplined life in which a person might find interior serenity in their current life, rather than an emphasis on reward in the hereafter. Scholars have sometimes thought of al-Kindi as a Mu‘tazili sympathizer, but this has not been proved; he appears rather to coexist with the worldview of orthodox Islam.

Al-Kindi's work on definition is *Fi hudud al-ashya' wa-rusumiha (On the Definitions of Things and their Descriptions)*. Through the terms he chose to define - finitude, creation, the first cause - we can see where the constructs of Islamic philosophy diverged from their Greek predecessors. In the eleventh century the *Kitab al-hudud (Book of Definitions)* of Ibn Sina replaced al-Kindi's work; this was considerably more advanced, both in its definitions and in its organization of the world into a concise ontological schema.

2 Metaphysics

Al-Kindi's best known treatise is the metaphysical study, *Fi al-falsafa al-ula (On First Philosophy)*. Aristotelian

influence can be seen in certain elements, such as the four causes. However he is Aristotelian only up to a point. The point of divergence is reached over the question of the origin of the world. Aristotle teaches the eternity of the world; Al-Kindi propounds creation *ex nihilo*. The later philosophers, such as al-Farabi, are usually considered to understand Aristotle more accurately; they had the advantage of better translations and a greater number of works. In *Fi al-falsafa al-ula*, al-Kindi described the first philosophy, which is also the most noble and highest philosophy, as the knowledge of the first truth, including the cause of every truth (the first cause). The first cause is prior in time because it is the cause of time. By the study of philosophy, people will learn the knowledge of things in reality, and through this the knowledge of the divinity of God and his unity. They will also learn human virtue. Throughout many of his treatises, al-Kindi emphasizes the importance of the intellect (*‘aql*) and contrasts it with matter.

He also discusses the One Truth, which is another name for God, and states that it does not have any attributes, predicates or characteristics. This view is consonant with the Mu‘tazili declaration of the unity of God as being strictly without attributes, and consequently al-Kindi has sometimes been deemed to be a Mu‘tazili by scholars.

Other aspects of his position include emphasis on the absolute unity of God, his power - particularly as creator - and creation *ex nihilo*. The Eternal, that is God, is not due to another; he has no cause and has neither genus nor species. There is no ‘before’ for the Eternal. The Eternal is unchanging, immutable and imperishable. In human terms, death is the soul’s taking leave of the body, which it employed during life. For al-Kindi, the intellect continues. Perhaps the soul is primarily the locus of the intellect. He reiterated in his ethical treatise the idea that humans must choose the world of the intellect over the material world (see §3).

Al-Kindi differs from the Hellenistic philosophical tradition primarily in espousing the belief that the world was created *ex nihilo*. In Aristotelian metaphysics the Prime Mover set the world in motion, but in the Hellenistic tradition, time and motion are intrinsically linked. Matter set in motion is eternally existing, since it exists before motion (and therefore before time). In this system, time is defined as the extension of the series of movements. Thus time begins with movement. In al-Kindi’s system, matter, time and movement are all finite, with a beginning and a cessation at some future point. Other subjects that concern al-Kindi can be seen from his titles, including *Fi wahdaniya Allah wa tunahiy jirm al-‘alam (On the Unity of God and the Limitation of the Body of the World)*, and *Fi kammiya kutub Aristutalis wa ma yahtaj ilahi fi tahsil al-falsafa (The Quantity of the Books of Aristotle and What is Required for the Acquisition of Philosophy)*.

In his philosophical writings, al-Kindi does not so much direct arguments to the concerns of religion as avoid them altogether, instead describing a parallel universe of philosophy. He consistently tries to show that the pursuit of philosophy is compatible with orthodox Islam. The *mutakallimun* had previously speculated on questions about matter, atoms and substance, which he also considers. Another reason for the claim that he was a Mu‘tazili was his persecution by the Khalif al-Mutawwakil, who instigated a reactionary policy against the Mu‘tazili and a return to traditionalism (see Ash‘ariyya and Mu‘tazila). Al-Kindi was caught in the general net of the Khalif’s anti-intellectualism; the Kindian emphasis is always on rationalism, an attitude which the orthodox establishment of a revealed religion is bound to find inimical.

3 Ethics

Al-Kindi’s ethics and practical philosophy are most discussed in a treatise *Fi al-hila li-daf‘ al-ahzan (On the Art of Averting Sorrows)*, of questionable authenticity. Fehmi Jadaane (1968) argues that al-Kindi was strongly influenced by the Stoic tradition, particularly the thought of Epictetus, which was known throughout the Islamic world at the time through contact with Syriac Christian scholars, if not through specific texts. Epictetus emphasized the importance of freedom from the world and human beings’ status as agents, who through their ultimate independence were responsible for their own happiness and independent of others. His last logical step, however, was that suicide was permissible if life was no longer worth living. This last idea is not repeated in al-Kindi.

Like the writings of the Stoics, al-Kindi’s treatise, which is of the ‘consolation of philosophy’ type, exhorts readers to concentrate on the life of the mind and the soul, not of the body (see Stoicism). Al-Kindi says that human beings are what they truly are in the soul, not in the body. Again, on the futility of looking for eternities in the visible world, he says that whoever wishes for what is not in nature wishes for what does not exist. The reader is

admonished that unhappiness follows such an attitude. In this treatise, al-Kindi advocates maintaining an internal balance through the mechanism of the individual's interior autonomy. If worldly property becomes a concern and is then lost or damaged, this will upset an individual's mental equilibrium. Stoic ideas about the ephemeral nature of earthly goods are recalled; al-Kindi warns against attachment to favourite worldly goods, using an example from Plutarch's *On Moral Virtue*. In that story, Nero receives a gift of a gorgeous, elaborate crystal tent, with which he is obviously smitten. A philosopher who is present in the crowd advises him that he has already been impoverished through his keen attachment to this object. If Nero were to lose it, the philosopher says, he will suffer because it is irreplaceable. Later when the rare object is lost at sea during transport, Nero is devastated. Scholars have argued that this treatise appears to be a mélange of wisdom literature from various Hellenistic sources, with no ideas that sound Kindian. Ibn Miskawayh refers to the ideas of al-Kindi in his treatise on ethics, *Tahdib al-akhlaq*.

Some ethical remarks are contained in other treatises. The virtues discussed in the treatise on definitions are wisdom, courage and temperance. A reflection of each virtue which exists in the soul is seen in the body. Virtue exists as a focal point between two extremes. Bravery, for example, is both mental and physical; it is midway between rashness and timidity.

Some reverberations of al-Kindi's thought also continued in the twelfth-century Christian Latin West, as certain of his treatises were translated into Latin by the Scholastics, notably *De intellectu (On the Intellect)*. In the thirteenth century Giles of Rome criticized ‘Alkindus’ with other philosophers in his work *Errores philosophorum (Errors of the Philosophers)*. Only a portion of al-Kindi's work survives, so judgment of him must necessarily be imperfect. However, al-Kindi's influence endured longer in the Western Islamic tradition than in the Eastern, as reflected in the writings of the twelfth-century mystic Ibn al-‘Arabi. With al-Kindi, who pursued reason against the background of revealed religion, begins the Islamic philosophical tradition which continues with the works of Ibn Sina and Ibn Rushd.

See also: Aristotelianism in Islamic philosophy; Ethics in Islamic philosophy; Islam, concept of philosophy in; Logic in Islamic philosophy

KIKI KENNEDY-DAY

List of works

al-Kindi (before 873) *Rasa'il al-Kindi al-falsafiya (Philosophical Treatises of al-Kindi)*, ed. M.A. Abu Ridah, 2 vols in 1, Cairo, 1953.(The standard collection of al-Kindi's treatises, with introductory notes in Arabic.

al-Kindi (before 873) *Fi al-falsafa al-ula (On First Philosophy)*, ed. and trans. A.L. Ivry, *Al-Kindi's Metaphysics: A translation of Ya‘qub ibn Ishaq al-Kindi's Treatise ‘On First Philosophy’*, Albany, NY: State University of New York Press, 1974.(A clear account of al-Kindi's metaphysics with English translation.)

al-Kindi (before 873) *Risalah fi al-hilah li-daf‘ al-ahzan (On the Art of Averting Sorrows)*, ed. and trans. H. Ritter and R. Walzer, ‘Uno scritto morale inedito di al-Kindi’, *Memorie della Reale Accademia nazionale dei Lincei*, Rome, Series VI, 8 (1), 1938, 47-62.(Text and Italian translation.)

al-Kindi (before 873) *Fi hudud al-ashya' wa-rusumiha (On the Definitions of Things and their Descriptions)*, ed. M.A. Abu Ridah in *Rasa'il al-Kindi al-falsafiya*, Cairo, 1953; trans. D. Gimaret in *Cinq épîtres*, Paris: Centre National de la Recherche Scientifique, 1976.(Al-Kindi's treatise on definitions.)

al-Kindi (before 873) *Fi wahdaniya allah wa tunahiy jirm al-‘alam (On the Unity of God and the Limitation of the Body of the World)*, ed. M.A. Abu Ridah in *Rasa'il al-Kindi al-falsafiya*, Cairo, 1953.(Al-Kindi on the nature of God.)

al-Kindi (before 873) *Fi kammiya kutub Aristutalis wa ma yahtaj ilahi fi tahsil al-falsafa (The Quantity of the Books of Aristotle and What is Required for the Acquisition of Philosophy)*, ed. M.A. Abu Ridah in *Rasa'il al-Kindi al-falsafiya*, Cairo, 1953.(Writings on Aristotle.)

References and further reading

Gimaret, D. (1976) *Cinq Épîtres (Five Treatises)*, Paris: Centre National de la Recherche Scientifique. (An excellent French translation with commentary of five treatises by al-Kindi. There are unfortunately very few English translations of al-Kindi's works.)

Jadaane, F. (1968) *L'Influence du stoïcisme sur la pensée musulmane (The Influence of Stoicism on Muslim Thought)*, Beirut: Dar el-Machreq.(An interesting argument for Stoic, rather than merely Aristotelian and

Neoplatonic influence, on the Islamic philosophers.)
Jolivet, J. (1971) *L'Intellect selon Kindi*, Leiden: Brill.(A classic work: extensive commentary and French translation of al-Kindi's treatise on the intellect.)
Klein-Franke, F. (1996) 'Al-Kindi', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 11, 165-77.(Account of the role of al-Kindi as the first Muslim philosopher, and in particular the links between his philosophy and contemporary theology and understanding of Greek thought.)
Moosa, M. (1967) 'Al-Kindi's Role in the Transmission of Greek Knowledge to the Arabs', *Journal of the Pakistan Historical Society* 15 (1): 3-18.(Good discussion of 'The Quantity of the Books of Aristotle'.)
Rosenthal, F. (1940) Review article of 'Uno scritto morale', *Orientalia* IX: 182-91.(An interesting review of the Ritter-Walzer treatise, still important despite its age.)
Stern, S.M. (1959) 'Notes on al-Kindi's Treatise on Definitions', *Journal of the Royal Asiatic Society* parts 1 and 2: 32-43.(Considered a classic.)

al-Muqammas, Daud (*fl.* 9th century)

*Daud ibn Marwan, called al-Muqammas, is the first Jewish thinker known to have written in Arabic and one of the earliest Arabic speaking theologians whose work is extant. He also wrote on logic, biblical exegesis, doxography and polemical matters. His pioneering efforts toward a systematic Jewish philosophy show the influence both of Muslim theology (*kalam*) and the Aristotelian philosophy taught in the Syriac Christian academies.*

The main sources for al-Muqammas' biography are the tenth century Karaite author al-Qirqisani and al-Muqammas' own works. He was called al-Raqqi, suggesting that he came from the city of Raqqa, located in what is today northern Syria; he was also called ha-Bavli, 'the Mesopotamian'. The sobriquet 'al-Muqammas' ('he who wears a *qamis*', the latter referring to a shirt or shift) reflects his position at the crossroads between the early Arabicized Jewish commmunity and the Syriac-speaking Christian community; an Arabic life of St Anthony applies this term to an Arab or one who dressed like an Arab. By his own testimony, he took part in a theological debate with a Muslim in Damascus. At one point he converted to Christianity and studied philosophy for many years in Nisibis with one Nana (probably Nonnus of Nisibis, d. *circa* 860), but his anti-Christian polemics show that he returned to Judaism, and all his extant works were written after that point. Although he was claimed by later Karaite authors there is no independent evidence to date that he was a Karaite (see Karaism).

Al-Muqammas' writings reflect the intellectual curiosity that may have been a factor in his conversion and reconversion. In a *summa* entitled *'Ishrun maqala (Twenty Chapters)*, almost sixteen chapters of which are extant, al-Muqammas sought answers to basic philosophical questions about knowledge, the world, God, humanity, and revelation. His questions were about the existence and character (the what, how and why of each of his subjects). In dealing with these matters, he was prepared to weigh the merits of rival religious traditions. After a short introduction to Aristotelian logic, the *'Ishrun maqala* follows the pattern well known to us in works of *kalam*, arguing each of its points against the background of a refutation of rival opinions. Thus al-Muqammas accompanies his arguments for *ex nihilo* creation with polemics against sceptics and materialists. His proofs of God's existence and unity include lengthy polemics against dualism, Christianity, and anthropomorphism. The same polemical or apologetic posture is adopted in the discussions of free will, prophecy (and the veracity of the Mosaic revelation) and accountability in the hereafter. The final chapters, which discussed other religions, are lost. Although clearly meant to vindicate Judaism, the work notably lacks quotations in Hebrew and other features that might have limited its audience to a Jewish readership. The Christian influence is highly visible throughout, not least in the extended polemics against Christianity.

Al-Muqammas' anti-Christian polemics also include a short, aggressive work entitled *al-Masa'il al-khamsin radd 'ala al-nasara min tariq al-qiyas (Fifty Questions for the Logical Refutation of Christianity)*, fragments of which were found in the Cairo Geniza, and a lost longer work, the *Kitab al-dara'a (Book of Urging on to Attack)*, which apparently discussed the history of Christianity. Al-Muqammas translated from Syriac and adapted to Judaism commentaries on Genesis and Ecclesiastes. A surviving fragment of the commentary on Genesis reveals that this earliest example of Judaeo-Arabic exegesis follows the rationalistic hermeneutical method of the Syriac Christian exegetes. Al-Muqammas also wrote a commentary on Aristotle's *Categories* and a refutation of Indian religions, both known to us only by title.

Al-Muqammas' writings preserve an early stage of Judaeo-Arabic philosophy, which in turn reflects the early stages of Islamic philosophy and theology. Thus we see the formulaic and apologetic character of *kalam* (see Islamic theology) alongside the fusion of Neoplatonic and Aristotelian ideas characteristic of *falsafa* (see Aristotelianism in Islamic philosophy; Neoplatonism in Islamic philosophy). The boundaries between these disciplines were then much less pronounced than the later became, but the hybridization may also reflect al-Muqammas' unique position rather than the state of the disciplines in his time. As a Jewish author uncommitted to Muslim or Christian traditions, he could freely borrow from both. As one of the first Jews to write in Arabic, he was not constrained by set patterns of borrowing.

Al-Muqammas' polemics against predestination and anthropormorphism have close Islamic parallels, especially in Mu'tazilite writings (see Ash'ariyya and Mu'tazila). In discussions of physical theory, however, the Christian influence is more pronounced. Thus, in a manner that was typically Christian but was later widespread among

Jewish thinkers, he defines 'substance' and 'accident', as 'that which exists in itself' and 'that which exists in another'. These definitions echo the Aristotelian approach and break sharply with the atomism of the occasionalist Muslim *mutakallimun* (theologians or dialecticians, practitioners of *kalam*) (see Epistemology in Islamic philosophy); but al-Muqammas does not develop a systematic Aristotelian cosmology, and he hardly ever uses the concepts of matter and form.

Often, al-Muqammas fuses Muslim and Christian ideas. For example, in subordinating God's attributes to his absolute unity, he uses formulae found in later *kalam* works ('God lives with a life that is identical with Him', or 'God lives but not by life'). To elucidate these statements, however, he uses similes found in Christological treatises as illustrations of trinitarianism.

Al-Muqammas' originality is hard to gauge. A certain stiffness in his style may be the price he paid for venturing into new areas. Today his ideas and expositions look familiar or trite, but it is easy to forget that their familiarity to us comes from later authors. Al-Muqammas' openness to neighbouring cultures and his ability to adapt their materials are clearly the basis of his most original contributions to Jewish philosophy.

His influence is most noticeable in the works of al-Qirqisani and the later Karaites. In Rabbanite circles he is overshadowed by the towering figure of Saadiah Gaon, but he is cited by Rabbanite authors including Bahya ibn Paquda, Moses ibn Ezra, Judah ben Barzillai and Yedaya Ha-Penini, in a tradition extending down to the fourteenth century. That his work was apparently never translated *in toto* into Hebrew both expresses and contributes to the loss of his influence among later thinkers.

See also: Islamic theology; Islam, concept of philosophy in

SARAH STROUMSA

List of works

al-Muqammas, Daud (9th century) *'Ishrun maqala (Twenty Chapters)*, ed. S. Stroumsa, *Dawud Ibn Marwan al-Muqammis's Twenty Chapters ['Ishrun Maqala]*, Leiden: Brill, 1989.(An edition and annotated English translation of al-Muqammas' *Summa*.)

References and further reading

Vajda, G. (1959) 'À propos de la perpétuité de la rétribution d'outre-tombe en théologie musulmanne' ('On the Eternal Retribution in the Hereafter in Islamic Theology'), *Studia Islamica* 11: 29-38.(A paraphrase and analysis of the sixteenth chapter of al-Muqammas' *'Ishrun maqala*.)

Vajda, G. (1967a) 'Le problème de l'unité de Dieu d'après Dawud ibn Marwan al-Muqammis' ('The Problem of the Unity of God According to Daud ibn Marwan al-Muqammas'), in A. Altmann (ed.) *Jewish Medieval and Renaissance Studies*, Cambridge, MA: Harvard University Press, 49-73.(A paraphrase and analysis of the eighth chapter of al-Muqammas' *'Ishrun maqala*.)

Vajda, G. (1967b) 'La prophétologie de Dawud ibn Marwan al-Raqqi al-Muqammis' ('The Prophetology of Daud ibn Marwan al-Raqqi al-Muqammas'), *Journal Asiatique* 265: 227-35.(A paraphrase and analysis of the fourteenth chapter of al-Muqammas' *'Ishrun maqala*.)

Vajda, G. (1974) 'Le pari de Pascal dans un texte judéo-arabe du IX^{e} siécle' ('Pascal's Wager in a Judaeo-Arabic text from the Ninth Century'), in *Mélanges d'histoire des religions offerts à Henri-Charles Puech*, Paris: PUF, 569-71.(A translation and analysis of a surviving fragment from the twentieth chapter of al-Muqammas' *'Ishrun maqala*.)

Stroumsa, S. (1991) 'The Impact of Syriac Tradition on Early Judaeo-Arabic Bible Exegesis', *Aram* 3: 83-96.(On al-Muqammas' exegetical work.)

al-Razi, Abu Bakr Muhammad ibn Zakariyya' (d. 925)

Perhaps the most famous and widely respected Islamic authority on medicine in the medieval period, al-Razi also aspired to a comparable achievement in philosophy and the other sciences such as alchemy. His success in these other subjects, however, was seldom recognized either in his own time or later; in philosophy, for example, more writers cite him for purposes of rejection and refutation than for admiration and emulation. However, his ideas were and are important. Chief among his positive contributions is his advocacy of a doctrine of equal aptitude in all humans, which grants no special role for unique and divinely favoured prophets and which recognizes the possibility of future progress in the advancement of knowledge. Philosophically, al-Razi was by his own admission a disciple of Socrates and Plato, much of whose teaching he knew on the basis of the latter's Timaeus. *Accordingly, he was noted for upholding the eternity of five primary principles, God, soul, time, matter and space, and for a concept of pleasure that sees it as the return to a normal harmony following a serious deviation or disruption which is itself pain.*

1 Life and work

Al-Razi's main career was that of a physician, and in that field he earned great respect and wide acclaim even from his most vociferous detractors in other matters. He directed two major hospitals, one in Baghdad, the capital of the Islamic empire at that time, and another in his native city of Rayy in northern Iran. His voluminous writings on medicine were universally admired. Despite advancing infirmities, he continued his research and writing into old age, still surrounded by students and assistants when he died in AH 313/AD 925. In Christian Europe he was known as Rhazes, and his works on medicine were highly respected.

The whole of al-Razi's work, both in medicine and the physical sciences and in philosophy, derives its central concern from his naturalistic view of the universe as like a 'visible animal' which, in contrast to almost all of the other philosophers in his era, he regarded as a subject of empirical scrutiny. At the heart of his philosophy lies Plato's *Timaeus*, which seems to have been al-Razi's ultimate inspiration (see Plato). He displayed almost no interest in the rest of Plato, although curiously he was the most avowedly loyal follower of Plato in Arabic and Islamic literature. His other idol was, as might be expected, Socrates, whose way of life he attempted to emulate, advocate and defend; although al-Razi's Socrates was not the extreme ascetic of an earlier period in Socrates' life but rather a later, fully participating social being. For Aristotle he had little use, rejecting outright commonly accepted doctrines that had deeply influenced his philosophical contemporaries (see Aristotelianism in Islamic philosophy). Unusually for his time, al-Razi boldly claimed that he could and had moved beyond his ancient philosophical predecessors and that neither the religious prophets nor the Greek masters (nor even al-Razi himself) possessed the final word or ultimate truth. Others in the future can and will surpass the achievements already realized by the great minds of the past and present.

Although denying prophetic religion, al-Razi's own ethical philosophy suggests a kind of religion that encompasses God and the universal soul, and hints at the salvation of particular souls. Nevertheless, writings of his on the falsity of prophets were broadly and specifically condemned and he was branded an arch-heretic for them. An opponent quotes him as claiming that the Qur'an yields no information of particular value in comparison with the books of Ptolemy, Euclid, Hippocrates, Galen, Plato or even Aristotle. It is clear that al-Razi would not accommodate revealed religion, which he saw as both particularistic and divisive. There is thus no harmony to be found between such religion and philosophy - a dangerous and certainly radical stance to take in his day. Al-Razi's metaphysical doctrines received no approval from later scholars. As a consequence, with the exception of two treatises on ethics, few of his non-medical works survive, thereby making a fair and detailed judgment of his ideas now difficult and often impossible.

2 Metaphysics

The metaphysical doctrine of al-Razi, insofar as it can be reconstructed, derives from his concept of the five eternal principles. God, for him, does not 'create' the world from nothing but rather arranges a universe out of pre-existing principles. His account of the soul features a mythic origin of the world in which God out of pity fashions a physical playground for the soul in response to its own desires; the soul, once fallen into the new realm God has made for it, requires God's further gift of intellect in order to find its way once more to salvation and

freedom (see Soul in Islamic philosophy).

In this scheme, intellect does not appear as a separate principle but is rather a later grace of God to the soul; the soul becomes intelligent, possessed of reason and therefore able to discern the relative value of the other four principles. Whereas the five principles are eternal, intellect as such is apparently not. Such a doctrine of intellect is sharply at odds with that of all of al-Razi's philosophical contemporaries, who are in general either adherents of some form of Neoplatonism or of Aristotelianism.

The remaining three principles, space, matter and time, serve as the non-animate components of the natural world. Space is defined by the relationship between the individual particles of matter, or atoms, and the void that surrounds them. The greater the density of material atoms, the heavier and more solid the resulting object; conversely, the larger the portion of void, the lighter and less solid. Time and matter have both an absolute, unqualified form and a limited form (see Matter; Time). Thus there is an absolute matter - pure extent - that does not depend in any way on place, just as there is a time, in this sense, that is not defined or limited by motion. The absolute time of al-Razi is, like matter, infinite; it thus transcends the time which Aristotle confined to the measurement of motion. Al-Razi, in the cases of both time and matter, knew well how he differed from Aristotle and also fully accepted and intended the consequences inherent in his anti-Peripatetic positions.

3 Ethical and moral philosophy

More can be said about al-Razi's ethical doctrines because two of his treatises that contain elements of a moral philosophy - *al-Tibb al-ruhani (The Spiritual Physick)* and *al-Sira al-falsafiyya (The Philosophical Life)* - have survived. On the one hand, al-Razi saw ethics as a kind of psychological medicine. The restoration of equilibrium following upon dislocation is the goal of spiritual or psychic healing, and preventing such disruptions is ethics. For him, pleasure is not a positive or cumulative affection but instead the result of a prior pain that was itself caused by a rupture or departure from the normal state and which thereafter ceases as the normal condition returns or is restored. Passion and appetites will occur naturally but they must be restrained by reason from growing to excess; they should be neither served nor encouraged. True virtue lies in satisfying every need only so far as is indispensable. Al-Razi was against all forms of asceticism, specifically those practised by Muslims. In comparison to the ascetic model of Socrates, contemporaries faulted him for leading a public existence, marrying and having children, earning a living and enjoying the company of princes. Al-Razi, however, vigorously denied that such asceticism was true of his ancient master; Socrates, he insisted, eventually did return to public life and thereafter avoided the extremes of his earlier position.

Equally al-Razi denied excesses of commission as well as abstinence. The merciful Lord, he said, does not approve the causing of pain and injustice; hence inflicting hurt either on oneself or on any other being is wrong unless necessary or inevitable. All must be in accord with nature and thus, on occasion, the greater good or benefit may require it. In this way al-Razi warned against the needless slaughter of animals (except in the case of wild, carnivorous creatures whose own extinction may spare their victims' death and also provide for the release of the beast's own soul).

On the other hand, al-Razi, in line with his emphasis on the controlling role of reason, believed that philosophy and the philosophic life yields the only salvation that is ultimately possible (see Salvation). Pursuit of that life is to imitate God in the way possible for humankind. The ultimate end for which humans were created does not comprise a physical existence but another world, one without death and pain. The human soul will achieve its hold on that world in proportion to the quality of its previous life while in the body. Those who practise justice and seek to acquire knowledge - that is, to lead the philosophical life - become habituated to reason, living free of the body and accustomed to unceasing joy. Upon death this will be their permanent state. In contrast, those who cling to a physical existence will, as Plato said, fail to depart this world of generation and corruption and will therefore continue to suffer pain and unending distress.

See also: Ethics in Islamic philosophy; Platonism in Islamic philosophy; Soul in Islamic philosophy

PAUL E. WALKER

List of works

al-Razi (before 925) *al-Tibb al-ruhani (The Spiritual Physick)*, ed. P. Kraus in *Rasa'il falsafiyya li-Abu Bakr*

Muhammad b. Zakariyya’ al-Razi, Cairo: Fouad I University Faculty of Letters, 1939; repr. Beirut: Dar al-Afaq al-Jadida, 1973; trans A. Arberry, *The Spiritual Physick of Rhazes*, London: John Murray, 1950.(An account of al-Razi’s philosophical and ethical psychology.)

al-Razi (before 925) *al-Sira al-falsafiyya (The Philosophical Life)*, ed. P. Kraus in *Rasa’il falsafiyya li-Abu Bakr Muhammad b. Zakariyya’ al-Razi*, Cairo: Fouad I University Faculty of Letters, 1939; repr. Beirut: Dar al-Afaq al-Jadida, 1973; trans. A. Arberry, ‘Apologia Pro Vita Sua’, in *Aspects of Islamic Civilization*, London: George Allen & Unwin, 1964.(Al-Razi’s own philosophical autobiography.)

References and further reading

Goodman, L. (1971) ‘The Epicurean Ethic of M. b. Zakariya’ al-Razi’, *Studia Islamica* 34: 5-26.(An account of the links between Epicureanism and al-Razi.)

Goodman, L. (1972) ‘Razi’s Psychology’, *Philosophical Forum* 4: 26-48.(An explanation of the implications and background of his psychological views.)

Goodman, L. (1975) ‘Razi’s Myth of the Fall of the Soul: Its Function in His Philosophy’, in G. Hourani (ed.) *Essays in Islamic Philosophy and Science*, Albany, NY: State University of New York Press, 25-40.(A detailed account of the doctrine of the soul.)

Goodman, L. (1994) ‘Al-Razi, Abu Bakr Muhammad b. Zakariya’’, *Encyclopaedia of Islam*, new edn. vol. VIII: 474-77.(An excellent summary of all that is known about al-Razi and his work.)

Goodman, L. (1996) ‘Muhammad ibn Zakariyya’ al-Razi’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 13, 198-215.(Description of the life and thought of the thinker, and the wider relevance of his ideas.)

Walker, P. (1992) ‘The Political Implications of al-Razi’s Philosophy’, in C. Butterworth (ed.) *The Political Aspects of Islamic Philosophy*, Cambridge, MA: Harvard University Press, 61-94.(The standard work on al-Razi’s political theory.)

al-Razi, Fakhr al-Din (1149-1209)

Imam Fakhr al-Din al-Razi was one of the outstanding figures in Islamic theology. Living in the second half of the sixth century AH *(twelfth century* AD*), he also wrote on history, grammar, rhetoric, literature, law, the natural sciences and philosophy, and composed one of the major works of Qur'anic exegesis, the only remarkable gap in his output being politics. He travelled widely in the eastern lands of Islam, often engaging in heated polemical confrontations. His disputatious character, intolerant of intellectual weakness, frequently surfaces in his writings, but these are also marked by a spirit of synthesis and a profound desire to uncover the truth, whatever its source. A number of his metaphysical positions became well known in subsequent philosophical literature, being cited more often than not for the purposes of refutation. His prolixity and pedantic argumentation were often criticized, but he was widely considered the reviver of Islam in his century.*

1 Theology and philosophy

Fakhr al-Din al-Razi was born in Rayy near present-day Tehran in AH 543 or 544/AD 1149-50. Like his predecessor al-Ghazali, he was an adherent of the Shafi'i school in law and of the theology of Ash'arism (see Ash'ariyya and Mu'tazila). He was attracted at an early age to the study of philosophy, in which he soon became proficient. In his late twenties, he visited Khwarazm and Transoxania, where he came in contact with some of the last theologians in the Mu'tazilite tradition. Although he endured hardship and poverty at the beginning of his career, on returning to Rayy from Transoxania he entered into the first of a series of patronage relations with rulers in the east which contributed to his reputedly considerable wealth and authority.

Al-Razi's skill in polemic ensured that controversy followed him in his subsequent sojourns in Khurasan, Bukhara, Samarqand and elsewhere (he is said to have visited India). He consequently made several dangerous enemies, including among them the Karramiyyah (an activist ascetic sect, staunch defenders of a literal interpretation of scripture and of anthropomorphism), the Isma'ilis, and the Hanbalites, each of whom apparently threatened his life at various points. Al-Razi settled finally in Herat, where he had a teaching *madrasa* built for him, and where he died in AH 606/AD 1209.

In the religious sciences, al-Ghazali had legitimized the use of logic, while at the same time attacking those key metaphysical doctrines of the philosophers which most offended against orthodox doctrine. This move prepared the ground for the subsequent incorporation of philosophical argumentation into theology. It was through al-Razi that this marriage was most completely effected in the Sunni world. His major theological works all begin with a section on metaphysics, and this was to become the pattern for most later writers.

The problem of how far al-Razi should be considered a philosopher (rather than a theologian) is complicated by changes of view during the course of his life, and by his highly disputatious and often intemperate personality, which he himself acknowledged. His style is marked by an extensively ramifying dialectic, often ending in highly artificial subtleties, and is not easy to follow. The relentlessness and sometimes obvious delight with which al-Razi used this method to home in on his victims earned him among philosophers the sobriquet of Iman al-Mushakkikin (Leader of the Doubters). Nevertheless, al-Razi was scrupulous in representing the views he set out to criticize, manifesting his concern to lay out a rigorous dialectic in which theological ideas could be debated before the arbitration of reason. This predictably brought him under subsequent attack from those who believed that upholding orthodox doctrine was the primary task of theology, one of whom remarked that in al-Razi's works 'the heresy is in cash, the refutation on credit'.

One of al-Razi's major concerns was the self-sufficiency of the intellect. His strongest statements show that he believed proofs based on Tradition (*hadith*) could never lead to certainty (*yaqin*) but only to presumption (*zann*), a key distinction in Islamic thought. On the other hand, his acknowledgement of the primacy of the Qur'an grew with his years. A detailed examination of al-Razi's rationalism has never been undertaken, but he undoubtedly holds an important place in the debate in the Islamic tradition on the harmonization of reason and revelation. In his later years he seems to have shown some interest in mysticism, although this never formed a significant part of his thought.

Al-Razi's most important philosophical writings were two works of his younger days, a commentary (*sharh*) on

the physics and metaphysics of Ibn Sina’s *Kitab al-isharat wa-’l-tanbihat (Remarks and Admonitions)* (see Ibn Sina) and another work on the same subject, *al-Mabahith al-mashriqiyya (Eastern Studies)*, which is based in large part on the latter’s *al-Shifa’* and *al-Najat* as well as *al-Isharat*, but in which al-Razi frequently preferred the views of Abu ‘l-Barakat al-Baghdadi (d. after AH 560/AD 1164-5). Also of great philosophical interest is his theological text *Muhassal al-afkar (The Harvest of Thought)*. Perhaps al-Razi’s greatest work, however, is the *Mafatih al-ghayb (The Keys to the Unknown)*, one of the most extensive commentaries on the Qur’an, running to eight volumes in quarto and known more popularly as simply *al-Tafsir al-kabir (The Great Commentary)*. As its more orthodox detractors have been happy to point out, this work, which occupied al-Razi to the end of his life and was completed by a pupil, contains much of philosophical interest.

The person who did the most to defend Ibn Sina, and philosophy in general, against the criticisms of al-Razi was Nasir al-Din al-Tusi, whose commentary on the *Kitab al-isharat* was in large measure a refutation of al-Razi’s opinions. Al-Tusi also wrote a *Talkhis al-muhassal al-afkar (Abridgement of the Muhassal al-afkar)*, where he likewise undertook a criticism of many of the philosophical criticisms in the *Muhassal al-afkar*.

2 Metaphysics

Al-Razi was associated by later authors with the view that existence is distinct from, and additional to, essence, both in the case of creation and in the case of God, and that pure existence is merely a concept (see Existence). This view is at variance with the Ash‘arite and Mu‘tazilite positions, as well as with that of Ibn Sina and his followers. Al-Razi only departed from this view in his commentary on the Qur’an, where he went back to a more traditional view that in God essence and existence are one.

Another challenge to the philosophers for which al-Razi achieved fame was his refutation of the emanationist principle *ex uno non fit nisi unum* (only one can come from one.) In Ibn Sina’s formulation, if an indivisible single thing were to give rise to two things, *a* and *b*, this would result in a contradiction, for the same single thing would be the source of both *a* and of not-*a* ($= b$). Al-Razi’s refutation was based on the claim that the contradictory of ‘the emanation of *a*’ is ‘the non-emanation of *a*’, not ‘the emanation of not-*a*’. On a related point, he originally denied the possibility of a vacuum, but in his *Mafatih* he argues for its existence, and for the power of the Almighty to fill it with an infinity of universes.

The philosophers, following Ibn Sina, held knowledge to be an inhering in the knower of the form of the thing known, and that consequently God knew only universals and not particulars, knowledge of the latter implying inadmissible changes in God’s essence as particulars changed (see Immutability). For the most part theologians were opposed to thus restricting God’s knowledge, on the grounds that he was omniscient (see Omniscience). Al-Razi upheld the theological side of the debate through postulating that knowledge involved a relation between the knower and the thing known, so that a change in the thing known would produce a change in the relation but not in the essence of the knower. This notion of a relation involved the substitution of a philosophical term, *idafa* (relation), for a theological one, *ta‘alluq* (connection), in an argument about the attribute of knowledge which belonged essentially to Abu ’l-Husayn al-Basri’s Mu‘tazilite school.

In ethics, al-Razi held that God alone, through revelation, determines moral values for man, it being these which give rise to praise and blame. God himself was beyond the moral realm and acted from no purpose extraneous to himself, be it out of pure goodness or for the benefit of his creation. Following al-Ghazali, and before him al-Juwayni, al-Razi’s solution to the problem posed for divine subjectivists by God’s threats of punishment and reward was to acknowledge a subjective rational capacity within man allowing him to understand what causes him pleasure and pain and thus enabling him to perceive where his advantage lies. In his *‘Ilm al-akhlaq (Science of Ethics)* al-Razi built upon al-Ghazali’s ethical writings, particularly from the *Ihya’ ‘ulum al-din*, providing a systematic framework based on psychology, again under the influence of al-Baghdadi (see Ethics in Islamic philosophy).

On the question of free will, al-Razi took a radical determinist position and rejected outright the Ash‘arite doctrine of *kasb* (acquisition). Al-Razi postulated two factors necessary for the production of an action: the power to do it or not to do it, and a preponderating factor, the motivation, which leads to the action being performed or not. Once the preponderating factor exists together with the power, either the act comes about necessarily or else it becomes impossible. Al-Razi pushed this essentially Mu‘tazilite thesis, which is also similar to Ibn Sina’s thinking, to its

logical conclusion, arguing that both the power and the preponderating factor had to be created by God for the result to exist necessarily, and hence that all human actions have been produced through God's determination. We thus appear to be free agents because we act according to our motives, but in reality we are constrained. A consequence of this theory when it is applied to God's own acts is that since God acts through his power, he must himself either act through constraint (if there is a preponderating factor in this case) or else by chance (if there is not), both of which conclusions violate the central Sunnite position that God is a totally free agent. Those who came after al-Razi felt that he had never adequately solved this difficulty, and he himself confessed that, whether from the point of view of reason or of tradition, there was in the end no satisfactory solution to the free will problem (see Free will).

Al-Razi held the Ash'arite position that God could re-create what had been made inexistent, and this formed the basis of his literal understanding of bodily resurrection. However, he also expressed views which were influenced by the theory of the late Mu'tazili Ibn al-Malahimi, who held the contrary position on the restoration of non-existence, that the world did not pass into non-existence but its parts were dissociated, and that the essential of these parts were reassembled on the resurrection. This ambivalence on al-Razi's part perhaps reflects the changes in his position on atomism, which he vehemently denied in his earlier purely philosophical works but of which he was more supportive towards the end of his life.

See also: al-Ghazali; Ibn Sina; Islamic theology; al-Tusi

JOHN COOPER

List of works

al-Razi, Fakhr al-Din (before 1185) *al-Mabahith al-mashriqiyya fi 'ilm al-ilahiyyat wa-'l-tabi'iyyat (Eastern Studies in Metaphysics and Physics)*, Hyderabad: Da'irat al-Ma'arif al-Nizamiyyah, 1923-4, 2 vols; repr. Tehran, 1966.(One of al-Razi's most important philosophical texts.)

al-Razi, Fakhr al-Din (before 1239) *al-Tafsir al-kabir (The Great Commentary)*, Cairo: al-Matba'ah al Bahiyyah al-Misriyyah, 1938, 32 vols in 16; several reprints.(Al-Razi's commentary on the Qur'an, completed by his pupil al-Khuwayyi; useful in many places as in indication of his later philosophical positions.)

al-Razi, Fakhr al-Din (before 1209) *Muhassal afkar al-mutaqaddimin wa-'l-muta'akhkhirin min al-'ulama' wa-'l-hukama' wa-'l-mutakallimin (The Harvest of the Thought of the Ancients and Moderns)*, Cairo: al-Matba'ah al Bahiyyah al-Misriyyah, 1905.(Printed with al-Tusi's *Talkhis al-Muhassal* at the bottom of the page and al-Razi's *al-Ma'alim fi usul al-din (The Waymarks and Principles of Religion)* in the margin.)

al-Razi, Fakhr al-Din (before 1209) *Kitab al-nafs wa-'l-ruh wa sharh quwa-huma (Book on the Soul and the Spirit and their Faculties)*, ed. M.S.H. al-Ma'sumi, Islamabad: Islamic Research Institute, 1968; trans. M.S.H. al-Ma'sumi, *Imam Razi's 'Ilm al-akhlaq*, Islamabad: Islamic Research Institute, 1969.(Al-Razi's work on ethics.)

al-Razi, Fakhr al-Din (before 1209) *Sharh al-Isharat (Commentary on the Isharat)*.(No critical edition of al-Razi's commentary on Ibn Sina's Kitab al-isharat has appeared. Portions can be found in S. Dunya (ed.) *al-Isharat wa-'l-tanbihat*, Cairo: Dar al-Ma'arif, 1957-60, 4 parts, 3 vols in 2; also in *al-Isharat wa-'l-tanbihat*, Tehran: Matba'at al-Haydari, 1957-9, 3 vols. Both these editions contain al-Tusi's commentary as well as parts of Fakhr al-Din al-Razi's commentary, to which al-Tusi is responding. The Tehran edition also contains Qutb al-Din al-Razi's commentary, which set out to adjudicate between al-Tusi and al-Razi.)

al-Razi, Fakhr al-Din (before 1209) *Lubab al-Isharat (The Pith of the Isharat)*, ed. M. Shihabi in *al-Tanbihat wa-'l isharat*, Tehran: Tehran University Press, 1960; ed. A. 'Atiyah, Cairo: Maktabat al-Kharji, 1936/7. (Al-Razi's epitome of Ibn Sina's work, written after he had completed his commentary.)

References and further reading

Abrahamov, B. (1992) 'Fakhr al-Din al-Razi on God's Knowledge of Particulars', *Oriens* 33: 133-55.(Discussion of a key point of difference between Islamic theologians and philosophers.)

Arnaldez, R. (1960) 'L'oeuvre de Fakhr al-Din al-Razi commentateur du Coran et philosophe' (The Works of Fakhr al-Din al-Razi, Qu'ranic Commentator and Philosopher), *Cahiers du Civilization médiévale, Xe-XIIe siècles* 3: 307-23.(In this article, Arnaldez has dug into al-Razi's enormous commentary on the Qur'an to come up with his mature philosophical ideas. Can be compared with McAuliffe (1990) and Mahdi's response to McAuliffe.)

Arnaldez, R. (1989) 'Trouvailles philosophiques dans le commentaire coranique de Fakhr al-Dîn al-Râzî', *Études*

Orientales 4: 17-26. (A follow-up to Arnaldez (1960).)

Ibn Sina (980-1037) *Kitab al-Isharat wa-'l-tanbihat (Remarks and Admonitions)*, trans. A.-M. Goichon, *Livre des directives et remarques*, Beirut and Paris, 1951.(Introduction and notes by the translator. Contained in the notes are a number of al-Razi's comments from his commentary on this work, as well as some of al-Tusi's criticisms of al-Razi.)

Kholeif, F. (1966) *A Study on Fakhr al-Din al-Razi and His Controversies in Transoxania*, Pensée Arabe et Musulmane 31, Beirut: Dar al-Machreq Éditeurs.(Arabic text and English translation of al-Razi's text of sixteen questions (philosophical, logical, legal) broached with scholars in Transoxania; gives a good idea of al-Razi's style. Also contains a list of al-Razi's works.)

Kraus, P. (1936-7) 'Les "Controverses" de Fakhr al-Din Razi' (The 'Controversies' of Fakhr al-Din al-Razi), *Bulletin de l'Institut d'Egypte* 19: 187-214.(An important early study of the 'controversies' translated in Kholeif (1966). An English translation appears in 'The controversies of Fakhr al-Din Razi', *Islamic Culture* 12, 1938: 131-53.)

McAuliffe, J.D. (1990) 'Fakhr al-Din al-Razi on God as al-Khaliq', in D.B. Burrell and B. McGinn (eds) *God and Creation: An Ecumenical Symposium*, Notre Dame, IN: University of Notre Dame Press, 276-96.(An examination of al-Razi's late philosophical theology, with particular reference to the problem of creation; see also M. Mahdi's response in the same volume (297-303) on the general question of al-Razi as philosopher.)

al-Sabzawari, al-Hajj Mulla Hadi (1797/8-1873)

*Al-Sabzawari was the most influential nineteenth-century Iranian philosopher. His reputation rests in part on his Sharh al-manzuma, a commentary on his own Ghurar al-fara'id (The Blazes of the Gems), a didactic poem (*manzuma*) encapsulating in a systematic fashion an exposition of the existentialist philosophy of Mulla Sadra. He was also the most sought-after teacher of philosophy in his day, and many students travelled to Sabzavar to be taught by him. Famous for his saintliness as well as his erudition, he set the tone for much of twentieth-century Iranian philosophy.*

Al-Hajj Mulla Hadi al-Sabzawari, the most famous of the philosophers of the Qajar period in Iran, was born in AH 1212/AD 1797-8 in Sabzavar in northeastern Iran. He studied logic, mathematics, law and metaphysics in Mashhad, where he moved at the age of ten after completing his preliminary education in Sabzavar. He pursued his interests in philosophy by moving to Isfahan to study for seven years with, among others, Mulla 'Ali Nuri (d. AH 1246/AD 1830-1), the foremost interpreter of his day of the philosophy of Mulla Sadra. He returned to Mashhad to teach for five years, and then accomplished the pilgrimage to Mecca (*hajj*). On his way back from pilgrimage he spent a year in Kirman, where he married, before returning to Sabzavar where he spent the rest of his life devoted to teaching and writing. A remarkable number of students of philosophy came to study with him, not only from Iran but from Arab countries and India as well. So great was his reputation that Nasir al-Din Shah, for whom he wrote his *Asrar al-hikam (Secrets of the Wisdoms)*, came to visit him in AH 1284/AD 1867, but al-Sabzawari's pious and ascetic way of life (several minor miracles are attributed to him), led him to refuse direct royal patronage. He died in AH 1289/AD 1873, having turned Mulla Sadra's legacy into the predominant philosophical school of the nineteenth and twentieth centuries.

Al-Sabzawari's fame rests primarily on one work, his *Ghurar al-fara'id (The Blazes of the Gems)*, a poem in which he gives a systematic and complete presentation of the philosophy of the school of Mulla Sadra, together with the *Sharh al-manzuma*, his own commentary on this poem, which he composed despairing of the philosophical ignorance of his contemporaries. The merit of this work lies not so much in any radically new theories, but in its plan and organization, which have made it the standard text for students of philosophy in Shi'i *madrasa*s until the present day. The situation is now changing and new teaching texts are appearing, but most of these are still influenced by the *Sharh al-manzuma* in both structure and content.

In the centuries after Mulla Sadra, philosophers were on the whole inclined to write on specific topics, thus leaving a gap in so far as there was no text that treated the whole of post-Sadrian philosophy in a systematic and assimilable fashion to which students could turn. Al-Sabzawari filled this gap, first with his didactic poem, which was to be memorized, and second with his commentary, which elaborates the poem in the manner of a traditional teacher in the Islamic religious sciences. The completeness of the work can be gauged by its contents, which give a good idea of what subjects philosophy encompasses for the contemporary religious student in a Shi'i *madrasa*.

The first part is on logic, with the commentary separately subtitled *al-La'ali al-muntazima (The Well-Ordered Nights)*. The second part (with commentary subtitled *Ghurar al-fara'id*), is divided into seven sections: (1) general principles (*al-umur al-'amma*), covering existence and related matters (unity, systematic ambiguity, modality, actuality and so on), quiddity and causality; (2) substance and accident; (3) metaphysical theology (*al-ilahiyyat bi-'l-ma'na al-akhass*), God's essence, attributes and acts; (4) natural philosophy, including discussions of matter and motion and a section on psychology; (5) certain supernatural phenomena, including dreams, miracles and prophecy; (6) the resurrection (*ma'ad*); and (7) ethics (*akhlaq*), with a brief treatment of spiritual values. It should be noted that (2) is essentially part of (1), that (1-4) form the core of the work, and that (5-7) are relatively short sections.

The commentary is amply provided with proofs and arguments, but there is also a marked emphasis on intuitive and mystical perception. This aspect of al-Sabzawari's thought is even more evident in his other works, which included commentaries on two of the key supplications in Shi'i devotional literature and the above-mentioned *Asrar al-hikam*, in Persian, in which eschatology is elucidated through metaphysical theology, psychology, ethics and the law (*shari'a*). This latter work is threaded through with poetic quotations. Al-Sabzawari was not only a poet in his own right (he has a Persian *diwan*, or collection of poems), but he also produced an elaborate

metaphysical commentary on passages from Jalal al-Din al-Rumi's *Mathnavi* which amply reflects his own spiritual preoccupations.

Because of the elegance of his exposition of the entire scope of the Sadrian philosophy of his time, al-Sabzawari has attracted the attention of a number of modern scholars who, under the influence of his manifestly esoteric outlook, have tended to emphasize the mystical approach in nineteenth-century Iranian philosophy. Although this influence has tended to obscure other philosophical currents of the time and their legacy to twentieth-century Iranian philosophy, it cannot be denied that al-Sabzawari was the most significant philosopher of this period and the one who, through the large number of his students, exerted the most powerful effect on later generations.

See also: Islamic philosophy, modern; Mulla Sadra; Mystical philosophy in Islam

JOHN COOPER

List of works

In addition to the works listed below, al-Sabzawari wrote glosses on several works by Mulla Sadra. Those on the *Asfar* can be found in the footnotes of the printed edition of that work (see Mulla Sadra).

al-Sabzawari (1826-45) *Sharh al-manzuma* (commentary on the didactic poem (*manzuma*), the *Ghurar al-fara'id*), ed. H.H. al-Amuli and M. Talibi, Tehran: Nab, 1995, 3 vols; *Sharh-i manzuma*, lithograph editions in Arabic.(This edition is so far incomplete. The only complete text is provided by numerous printings of the lithographed texts, the most common being the so-called 'Nasiri' lithograph. In the published edition, Volume 1 contains the logic, *al-La'ali al-muntazima*; Volume 2 is sections 1 and 2 of the metaphysics of the *Ghurar al-fara'id*; Volume 3 contains the remainder of the metaphysics. The text is in Arabic with glosses by H.H. al-Amuli, but there is an English introduction to Volume 3. An earlier Arabic edition of the metaphysics is *Sharh ghurar al-fara'id*, ed. M. Mohaghegh and T. Izutsu, Tehran: McGill University Institute of Islamic Studies, Tehran Branch, 1969. This edition also contains extracts from the commentaries of Muhammad Taqi Amuli and Akhund-e Hidaji, and has a useful Arabic-English glossary of al-Sabzawari's philosophical terminology.)

al-Sabzawari (1867-70) *Asrar al-hikam (Secrets of the Wisdoms)*, ed. H.M. Farzad, Tehran: Intisharat-e Mawla, 1982.(Arabic text, not a critical edition.)

References and further reading

Akhtar, W. (1984) 'Sabzawari's Analysis of Being', *al-Tawhid* 2 (1): 29-65.(Makes a strong case for Sabzawari's originality as a thinker, and situates the *Sharh al-manzuma* within a history of ontology; useful comparative analysis.)

Izutsu Toshihiko (1971) 'The Fundamental Structure of Sabzawari's Metaphysics', in Izutsu Toshihiko (ed.) *The Concept and Reality of Existence*, Tokyo: Keio Institute of Cultural and Linguistic Studies, 57-149.(A discussion of existence and quiddity in peripatetic and illuminationist philosophy, as the background to understanding al-Sabzawari's thought.)

Izutsu Toshihiko and Mohaghegh, M. (1983) *The Metaphysics of Sabzawari*, Delmar, NY: Caravan.(An English translation of the first section of the *Ghurar al-fara'id*.)

Nasr, S.H. (1966) 'Renaissance in Iran (continued): Haji Mulla Hadi Sabziwari', in M.M. Sharif (ed.) *A History of Muslim Philosophy*, Wiesbaden: Harrasowitz; repr. Karachi, 1983. (A good introduction to the content of al-Sabzawari's *Sharh al-manzuma*.)

al-Sijistani, Abu Sulayman Muhammad (*c.*932-*c.*1000)

Al-Sijistani was one of the great figures of Baghdad in the fourth century AH (tenth century AD). He assembled around him a circle of philosophers and litterateurs who met regularly in sessions to discuss topics related to philosophy, religion and language. As a philosopher with a humanistic orientation, his concerns went beyond subjects of strictly philosophical nature. His philosophical ideas displayed Aristotelian and Neoplatonic motifs. He considered philosophy and religion to be totally different in nature and method, so that the two could not be reconciled. God is only prior to the world in essence, rank and nobility, not in time. Al-Sijistani insisted that in no way should one attribute to God the imperfections of created things. According to him, the soul is simple by nature and natural reason is capable of attaining a state of pure knowledge that enables one to distinguish between good and evil. Reason, if taken as a guide, could ensure happiness.

1 Life and works

Abu Sulayman Muhammad al-Sijistani, known as al-Mantiqi (the Logician), was born, *c.*AH 320/AD 932. His formative years were spent in Sijistan (now Sistan in Iran) but the mature phase of his career took place in Baghdad. He became one of the great figures of the Islamic humanist movement that flourished during the fourth century AH (tenth century AD) in Baghdad. He dominated his generation by the enlightenment of his judgments and the breadth of his knowledge. He was especially interested in ancient philosophy and its transmission into the world of Islam. He assembled around him a circle of friends, philosophers, scientists and litterateurs of various ethnic and religious affiliations. The circle met in regular sessions (*majalis*) and with open and critical minds discussed topics related to philosophy, religion, science and language. Al-Sijistani died *c.*AH 391/AD 1000.

Those of al-Sijsitani's works that have reached us are not numerous. One of the most important attributed to him is *Siwan al-hikma (Vessel of Wisdom)*, from which only selections have survived. The attribution to al-Sijistani of this collection of dicta and anecdotes of Greek and Islamic philosophers has been challenged by W. al-Qadi (1981), but Joel Kraemer (1986) believes that the work emanated from al-Sijistani's school, based upon classroom notes and texts. In addition to *Siwan al-hikma*, al-Sijistani wrote short treatises on the first mover, the fifth nature of the celestial spheres, the perfection peculiar to the human species, principles of being, dream omens and logic.

In a Socratic fashion, al-Sijistani preferred oral instruction over writing. We owe to his student and protégé Abu Hayyan al-Tawhidi most of the information concerning his philosophical ideas and doctrines. Al-Tawhidi recorded the sessions of al-Sijistani's circle in his works *al-Muqabasat (Conversations)* and *al-Imta' wa-al-mu'anasa (Book of Pleasure and Conviviality)*, which remain the major sources of information on the life and thought of al-Sijistani. Consequently we have to look at his teachings through the window of his pupil's writings.

2 Doctrines

As a philosopher with a humanistic orientation, al-Sijistani's concerns went beyond subjects of strictly philosophical nature. His philosophical ideas displayed Aristotelian and Neoplatonic motifs and touched on a wide range of subjects such as politics, aesthetics and friendship, among others. However, his chief preoccupation, and that of the circle members as reflected in al-Tawhidi's recordings, centred on the relations between philosophy and religion, the mind-body problem - why the soul was susceptible to virtues and vices, good and evil - the question of God's relation with and action in the universe, and finally the individual and society.

Al-Tawhidi represents his mentor as a man of deep religious sentiments but who regarded both religion and philosophy as true and valid. The two are independent and should not and cannot be reconciled. They differ in method and substance. In religion there are things which cannot be fathomed or understood but are to be accepted and assented to; the end of religion is proximity to God, whereas the aim of philosophy is contemplation. Al-Sijistani objected to the attempt by the Brethren of Purity to harmonize religion and philosophy, and lashed out at the Islamic theologians (*al-mutakallimun*) who claimed that their methodology was rationalist when in reality it was false rationalism (see Ikhwan al-safa'; Islamic theology).

For al-Sijistani the universe is divided into the terrestrial and the intelligible according to the Platonic system, but God acts in accordance with the Aristotelian concept of first mover. Having accepted the view that matter is eternal, he held that this does not detract from the perfection of God since in the final analysis everything depends

on his will. One should not attribute to God the imperfections of the created world. God is prior to the world in essence, rank and nobility, but not in time.

Knowledge implies two types: natural and supernatural. There are four degrees of knowledge: sensible knowledge possessed by non-reasoning animals; absolutely and exclusively intelligible knowledge possessed by the celestial bodies; the sensible-intelligible knowledge tied up with the imagination of those who have not reached perfect purity; and the intelligible-sensible knowledge which has been arrived at through rational and speculative investigation. This is the highest knowledge humans, including such persons as philosophers, augurs (*kahins*) and prophets, can acquire. Intuition, however, is the noblest kind of knowledge because it presents itself by itself in the soul and is not subject to generation and corruption. Through reason we overcome all obstacles to reach God through the intellect, which is the medium between human beings and the supernatural world. Reason has the power to contact the super-sensible beings until it reaches the First Being.

Having been greatly interested in the body-soul relationship, al-Sijistani distinguished between soul (*nafs*) and spirit (*ruh*), considering the soul to be a simple substance imperceptible to the senses and not subject to change or corruption. According to him, human beings are so by virtue of having a soul and not by the possession of a body, although the soul cannot make a human being by itself. The soul is the principle of knowledge, the body the principle of action. Because of the rival elements, nature versus reason, that constitute a human being and pull him in opposite directions, it is important to take reason as a guide. It alone can ensure our ultimate happiness, namely the knowledge of God and the good which he has reserved for the virtuous. One should aim high towards the celestial world in order to reach eternal life (see Soul in Islamic philosophy).

See also: Aristotelianism in Islamic philosophy; Greek philosophy: impact on Islamic philosophy; Neoplatonism in Islamic philosophy; al-Tawhidi

GEORGE N. ATIYEH

List of works

al-Sijistani (*c*.932-*c*.1000) *Siwan al-hikmah wa-thalath rasa'il (Vessel of Wisdom and the Three Treatises)*, ed. A. Badawi, Teheran: Bunyad Farhang, 1974; ed. D.M. Dunlop, *The Muntakhab Siwan al-Hikmah of Abu Sulaiman as-Sijistani, Arabic Text, Introduction and Indices*, The Hague: Mouton, 1979.(The Badawi edition includes, in addition to introductions in Arabic and French, the following treatises: *Fi anna al-ajram al- ulwiya tabi'atuha tabi'a khamisa (The Nature of the Celestial Bodies is a Fifth Nature)*; *Fi al-muharrik al-awwal (On the First Mover)*; and *Fi al-kamal al-khass bi-naw' al-insan (On the Perfection Peculiar to the Human Species)*. Dunlop contains fragments, especially on medicine, that are not in the Badawi edition. He believes that sections dealing with the Greek philosophers are derived from Porphyry's *History of Philosophy*.)

al-Sijistani (*c*.932-*c*.1000) *Fi mabadi' al-mawjudat (On the Principles of Beings)*, ed. and trans. G. Troupeau, *Pensamiento* 25, 1969: 259-70.(Discusses the ranks of the first beings and the attributes by which the first essence is qualified.)

al-Sijistani (*c*.932-*c*.1000) *Fi al-kamal al-khass bi-naw' al-insan (The Perfection Peculiar to the Human Species)*, ed. and trans. into French by M. Kügel-Turker, *Pensamiento* 25, 1969: 207-24.(By using the intellect, man reaches perfection.)

References and further reading

Jadaane, F. (1971) 'La philosophie de Sijistani', *Studia Islamica* 33: 67-95.(A general introduction to al-Sijistani's philosophical ideas and their Greek sources.)

Kraemer, J.L. (1986) *Philosophy in the Renaissance of Islam, Abu Sulayman al-Sijistani and His Circle*, Leiden: Brill.(Major work which examines the extent to which al-Sijistani and his circle were conversant with technical discussions of philosophical issues treated in Greek and late antiquity.)

Leaman, O. (1996) 'Islamic Humanism in the Fourth/Tenth Century)', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 10, 155-64.(Analysis of the nature of this important period of Islamic philosophy, concentrating on the ethical thought of the period.)

al-Qadi, W. (1981) '*Kitab Siwan al-Hikmah*: Structure, Composition, Authorship and Sources', *Der Islam* 58: 87-124.(Based on some internal evidence, concludes that the real author is the little known Abu al-Qasim al-Katib.)

al-Tawhidi (*c*.930-1023) *al-Imta' wa-al-mu'anasa (Book of Pleasure and Conviviality)*, ed. A. Amin and A. al-Zayn, Beirut: al-Maktaba al-'Asriyya, 2nd edn, 1953, 3 vols.(A record of night sessions in Baghdad in which

al-Sijistani participated in discussing philosophical questions.)

al-Tawhidi (*c*.930-1023) *al-Muqabasat (Conversations)*, ed. T. Husayn, Baghdad: Matba‘at al-Irshad, 1970.(A collection of 106 conversations on various philosophical subjects.)

al-Suhrawardi, Shihab al-Din Yahya (1154-91)

Al-Suhrawardi, whose life spanned a period of less than forty years in the middle of the twelfth century AD, *produced a series of highly assured works which established him as the founder of a new school of philosophy in the Muslim world, the school of Illuminationist philosophy (*hikmat al-ishraq*). Although arising out of the peripatetic philosophy developed by Ibn Sina, al-Suhrawardi's Illuminationist philosophy is critical of several of the positions taken by Ibn Sina, and radically departs from the latter through the creation of a symbolic language to give expression to his metaphysics and cosmology, his 'science of lights'. The fundamental constituent of reality for al-Suhrawardi is pure, immaterial light, than which nothing is more manifest, and which unfolds from the Light of Lights in emanationist fashion through a descending order of lights of ever diminishing intensity; through complex interactions, these in turn give rise to horizontal arrays of lights, similar in concept to the Platonic Forms, which govern the species of mundane reality. Al-Suhrawardi also elaborated the idea of an independent, intermediary world, the imaginal world (*alam al-mithal*). His views have exerted a powerful influence down to this day, particularly through Mulla Sadra's adaptation of his concept of intensity and gradation to existence, wherein he combined Peripatetic and Illuminationist descriptions of reality.*

1 Al-Suhrawardi and the philosophy of *ishraq*

Shihab al-Din Yahya ibn Habash ibn Amirak Abu 'l-Futuh al-Suhrawardi, known as al-Maqtul (the Slain) in reference to his execution, and usually referred to as Shaykh al-Ishraq after the Illuminationist philosophy (*hikmat al-ishraq*) which he espoused, was born in AH 549/AD 1154 in the village of Suhraward in northwest Iran. After studying in Maraghah (with Majd al-Din al-Jili, who also taught Fakhr al-Din Al-Razi) and Isfahan, he passed several years in southwest Anatolia, associating with Seljuq rulers and princes, before moving to Aleppo in AH 579/AD 1183. Here he taught and became a friend of the governor, al-Malik al-Zahir al-Ghazi (son of the Ayyubid Salah al-Din, famous in European literature as Saladin), who later also befriended Ibn al-'Arabi. However, he fell foul of the religious authorities, and was executed in AH 587/AD 1191 on the orders of Salah al-Din, in circumstances which remain unclear but which involved charges of corrupting the religion and allegations of claims to prophecy, and may also have had a political dimension.

Al-Suhrawardi clearly intended his philosophy to make a distinctive break with the previous peripatetic tradition of Ibn Sina, but the significance of this break has been interpreted in a number of ways. For subsequent Islamic philosophy, he was above all the conceiver and main proponent of the theory of the primacy of quiddity. While the predominant trend in Western scholarship has been to depict him as the originator of a distinctive mystical and esoteric philosophy, recent Western scholarship has emphasized his critique of peripatetic logic and epistemology and his own theories in these fields (see for example Ziai 1990).

Ibn Sina famously tackled the question of mystical knowledge in the last section of his *Kitab al-Isharat wa-'l-tanbihat (Remarks and Admonitions)*, thus assuring a place for this area of knowledge within the domain of *hikma* (wisdom). It was al-Suhrawardi, however, who turned mystical and intuitive knowledge into a paradigm of knowledge in general. This epistemology then served as a basis on which to construct both a critique of peripatetic philosophy and an original philosophy of lights, or Illumination (*ishraq*). Yet, however important it was for al-Suhrawardi to stress his radical departure from peripatetic philosophy, he also emphasized the necessity for those who would follow his method to study the peripatetic method closely.

Al-Suhrawardi's writings fall into several categories. First, there are his four major philosophical works, written in Arabic: *Kitab al-talwihat (The Intimations)*, *Kitab al-muqawamat (The Oppositions)*, *Kitab al-mashari' wa-'l-mutarahat (The Paths and Heavens)* and *Kitab hikmat al-ishraq (The Philosophy of Illumination)*. These were apparently intended by al-Suhrawardi to be studied in this order, and roughly follow a progression from a more or less conventionally peripatetic style to one in which the 'science of lights' is expressed through its own technical vocabulary and method, a progression described by al-Suhrawardi as a movement from a discursive philosophy (*hikma bahthiyya*) to an intuitive philosophy (*hikma dhawqiyya*). The second group of works contains a set of symbolic narratives, mostly in Persian but a few in Arabic, expounding the journey of the soul through the stages of self-realization and offering striking images of some of the notions of Illuminationism while seeking to cultivate the kind of intuitive vision at its heart. The remaining works consist of a number of shorter treatises in

Arabic, such as the *Hayakil al-nur (The Temples of Light)*, and others in Persian expounding Illuminationist philosophy in a simpler form, a collection of prayers and invocations, and some miscellaneous translations (or versions) and commentaries.

2 Epistemology

By basing his philosophy on light, al-Suhrawardi was able to introduce two important notions which may be thought of as the seeds of the entire system: that of intensity and gradation, and that of presence and self-manifestation. It is possible to see his philosophy as experiential, although his notion of experience was not confined to that obtained through the senses but embraced other forms including that of mystical experience. Ibn Sina's explanation of knowledge is based on the inhering of the form of the thing known in the mind of the knower, but for al-Suhrawardi such knowledge only guarantees certainty and the correspondence of knowledge with reality, because there exists a more fundamental kind of knowledge that does not depend on form and which is, like the experience of pain, unmediated and undeniable. The prime mode of this presential knowledge (*al-'ilm al-huduri*) is self-awareness, and every being existing in itself which is capable of self-awareness is a pure and simple light, as evinced by the pellucid clarity with which it is manifest to itself. In fact, being a pure and simple light is precisely the same as having self-awareness, and this is true of all self-aware entities up to and including God, the Light of Lights, the intensity of whose illumination and self-awareness encompasses everything else. The main constituent of reality is the hierarchies of such pure lights, differing solely in the intensity of their Illumination, and thus of self-awareness (see Illuminationist philosophy).

How then is the philosopher to realize this self-awareness? The prospective Illuminationist must engage in a variety of recommended ascetic practices (including forty-day retreats and abstaining from meat) to detach himself from the darknesses of this world and prepare himself for the experiences of the world of lights. The heightened pleasure afforded by this latter kind of experience is emphasized. Having spiritually purified himself, the philosopher is ready to receive the Divine Light and is rewarded with visions of 'apocalyptic' lights which form the basis for real knowledge. At this point the Illuminationist must employ discursive philosophy to analyse the experience and systematize it, in the same way as with sensory experience.

The relation between this direct intuitive knowledge and the philosophy of Illumination is compared to that between observation of the heavens and astronomy. The major portion of al-Suhrawardi's writings is devoted to this last stage of rational analysis and systematization, although he sometimes relates his visions. His symbolic narratives in Persian are in some sense a record of these, although in them al-Suhrawardi, the author, is never explicitly the first person. The narratives have a pedagogic function, and are guides to the kind of experiences to be encountered by the seeker and to their interpretation; indeed a central figure in these narratives is often a guide, the lord of the human species, sometimes though not exclusively identified as Gabriel.

3 Logic, physics, metaphysics and cosmology

The unfolding of reality in Illuminationism is governed by the different ways in which the pure lights interact to produce further levels of lights and darknesses, and by the subsequent interaction of all these different levels with each other, resulting eventually in a densely populated universe. The pure lights are the causes of three other categories of entities: accidental lights (actual physical light, and certain accidents of intellects and souls), dark modes (accidental categories in bodies excluding accidental lights) and intermediary isthmuses (*barzakhs*) or boundaries (bodies). The luminous properties of these degrees are also properties of self-awareness; thus for example, an accidental light subsists in something other than itself, and is also in need of something else to be aware of itself.

Existence as such does not perform much more than an explanatory role in Illuminationism, quite different from the central position it occupied in peripatetic philosophy, a major question for which was the nature of the relationship between existence and quiddity. However, it is important to notice that light does not merely act as a substitute for existence: existence, and its explanatory function, is rendered totally redundant. The lights (and darknesses and *barzakhs* caused by lights) in al-Suhrawardi's system are discrete entities whose interactions in turn bring about other lights. There is thus a primacy of the entity, and al-Suhrawardi regards existence as such to be no more than a mental abstraction having no external reality. Furthermore, although lights differ in intensity, there is nothing in this system to correspond with Ibn al-'Arabi's *wahdat al-wujud* (unity of being) (see Ibn

al-‘Arabi); al-Suhrawardi would not have said that all reality is light, but that it is lights. It is for these reasons that he was subsequently held responsible for the idea of the primacy of quiddity (*asalat al-mahiya*), although he did not use this expression himself. It was Mulla Sadra who, four centuries later, built upon the insight that reality was in effect a continuum of graded intensities, but a continuum of existence, not of light. He was thus able to fuse al-Suhrawardi’s system with those of the peripatetics and Ibn al- ‘Arabi into a metaphysical theory in which reality was nothing more than existence itself, and to turn quiddity into the purely mental abstraction which existence was for al-Suhrawardi.

His insight concerning presential knowledge (which al-Suhrawardi himself declared was vouchsafed to him by Aristotle in a dream) suggested solutions to weaknesses which al-Suhrawardi had detected in Ibn Sina’s philosophical system. The most important of these concerned the theory of definition, and the problem of definition as the basis of scientific knowledge. First, he objects that it is impossible to give a complete definition, for a complete definition should contain all the constituents of the definiendum, and such an enumeration is impossible. Second, the peripatetics held that definition is a means of proceeding from the known to the unknown; but the essential constituents, al-Suhrawardi asserts, are just as unknown as the definiendum, so this cannot be so. Contained within this is also an objection against induction: how can one know if the collection of essential elements of a thing is complete merely by enumerating them? His conclusion is that prior knowledge is always necessary and presupposed.

Another area of disagreement with peripatetic philosophy was the categories, which were treated by al-Suhrawardi not in his logic, but in his physics. He reduces the accidental categories to four (quality, quantity, relation and motion), and holds intensity to be a property of substances as well as of accidents. With change in intensity, there is no change in the essence of an accident (a colour, for example) or a substance (such as cause and effect); the only difference is the degree of perfection.

As is to be expected, al-Suhrawardi’s physics also contained a new theory of vision. He not only rejected the idea that the forms of objects were imprinted in the eye, but also the other current theory that light was emitted from the eye and fell onto the object. Vision is only possible, according to al-Suhrawardi, when the soul is illuminated by the light, substantial or accidental, of the object, and thus he brings vision within the compass of his illuminative theory of knowledge.

The physical or elemental world as depicted by al-Suhrawardi rejects the peripatetic division of matter and form, and substitutes for it a world of bodies composed of varying mixtures of light and darkness, which permit the passage of light to different degrees. Above the physical world the lights are arranged in a vertical array, corresponding to the emanationist scheme of Ibn Sina. However, these pure, immaterial lights are not restricted to ten as are the intellects of the peripatetic scheme. Al-Suhrawardi says only that they are limited to the number of stars in the fixed heavens; thus they are indefinite in number, but not infinite. Moreover, these vertically arrayed ‘triumphal’ (*qahira*) lights interact with each other to produce a horizontal array of similarly immaterial, ‘regent’ (*mudabbira*) lights. Each of these horizontally arrayed lights is the lord of a species, analogous to the Platonic Forms, but with the important difference that they are lights which ‘govern’ the species under them rather than universals. The species are depicted as ‘idols’ (*asnam*) of their archetypes. It is the interactions of both the vertical and the horizontal lights which give rise to the bodies of the lower world, which are also classified into degrees depending on the extent to which they receive and transmit light, each being a boundary (*barzakh*) between light and darkness. Al-Suhrawardi also elaborated the idea of the immaterial imaginal world (*‘alam al-mithal*), situated between the physical world and the world of the lords of species. This is the locus for the kinds of veridical experiences recounted in his symbolic narratives, an unmediated account of which can only be given in this way and not through discursive reason. Al-Suhrawardi’s cosmology is a good deal more complicated than this survey has suggested, employing a detailed terminology for the divisions of lights which classifies them in a variety of different ways.

4 The language of *ishraq*

The integrity of al-Suhrawardi’s complex philosophy is achieved in no small measure by the elegance and refinement of his means of expression. His original Illuminationist vocabulary - the Islamic roots of which are sometimes overlooked - is one aspect of this. Al-Ghazali had set a precedent with his *Mishkat al-anwar (The Niche of the Lights)*, which commented on the Light verse in the Qur’an (24: 35). However, al-Suhrawardi also uses a

number of other devices to stretch the reader's conceptual boundaries and to convey further dimensions of his total vision of reality. All the lights are related to each other in a downward sense by their being 'triumphal' or 'exalted', but cohesion is further maintained by the 'desire' or 'love' which the lower degrees feel for the upper, and by the explanation which this affords of such things as the joy we experience in the presence of the sun and our fear in the presence of darkness, and the delight which we take in certain minerals such as gold and rubies.

Al-Suhrawardi also chose to describe the horizontal lights as angels, using names of the *Anshaspands* of Zoroastrian mythology to denote them (Khordad, Murdad, Urdibihisht and so on), and he taps the vocabulary of Pahlawi for further terms. In various places in his works he also traces a genealogy for the transmission of illuminationist wisdom which goes back simultaneously through a Greek/Western line (including Pythagoras and Plato) and an Iranian/Eastern line including Zoroaster (see Zoroastrianism) to Hermes Trismegistus, and asserts that there have been illuminationists (*ishraqiyyun*) throughout time. All of this raises the question of precedents for, and influences on, al-Suhrawardi's thought, a subject which has caused some controversy in the Western literature on this subject. It is not necessary to go into the details of Corbin's largely phenomenological argument for the existence of a Persian *ishraqi* philosophical tradition independent from the peripatetic (Corbin 1971); it is sufficient to point out the paucity of historical evidence for such a thesis, and indeed the paucity of textual evidence for any specific conclusions about influences on al-Suhrawardi. The more economical approach is to regard his use of ancient Persian mythology and his genealogy as a means of expressing his overwhelming conviction that he had restored the original foundation of philosophy in the certainty of intuitive experience, a foundation which he believed had been undermined by the excessive discursiveness of such philosophers as Ibn Sina. He saw the traces of this foundation in the writings of Plato and Aristotle (who in the Islamic tradition was also the author of the famous *Theology*), in the remnants of the Zoroastrian religion which he encountered, and in the utterances and writings attributed to certain Sufis.

The influence of al-Suhrawardi on Mulla Sadra has been mentioned above, but there is in addition a long and lively tradition of commentaries on several of his texts. In the philosophical tradition which continued after the Mongol period in Iran and further east in India, al-Suhrawardi stands second only to Ibn Sina. Perhaps the greatest testimony to his lasting importance is the fact that to this day in Iran philosophers are still informally classified as either *mashsha'i* (peripatetic) or *ishraqi*, depending on their leaning towards rationality or mysticism.

See also: Ibn Sina; Illuminationist philosophy; Mulla Sadra; Mystical philosophy in Islam; Neoplatonism in Islamic philosophy; Platonism in Islamic philosophy

JOHN COOPER

List of works

al-Suhrawardi [Sohravardi, Shihaboddin Yahya] (1180?-91) *Œuvres philosophiques et mystiques*, vol. I: *La métaphysique: I. Kitab al-talwihat. 2. Kitab al-moqawamat. 3. Kitab al-mashari' wa'l-motarahat*, Arabic texts edited with introduction in French by H. Corbin, Tehran: Imperial Iranian Academy of Philosophy, and Paris: Adrien Maisonneuve, 1976; vol II: *I. Le Livre de la Théosophie oriental (Kitab Hikmat al-ishraq). 2. Le Symbole de foi des philosophes. 3. Le Récit de l'Exil occidental*, Arabic texts edited with introduction in French by H. Corbin, Tehran: Imperial Iranian Academy of Philosophy, and Paris: Adrien Maisonneuve, 1977; vol III: *Œuvres en persan*, Persian texts edited with introduction in Persian by S.H. Nasr, introduction in French by H. Corbin, Tehran: Imperial Iranian Academy of Philosophy, and Paris: Adrien Maisonneuve, 1977.(Only the metaphysics of the three texts in Vol. I were published. Vol. III contains a Persian version of the *Hayakil al-nur*, ed. and trans. H. Corbin, *L'Archange empourpré: quinze traités et récits mystiques*, Paris: Fayard, 1976, contains translations of most of the texts in vol. III of *Œuvres philosophiques et mystiques*, plus four others. Corbin provides introductions to each treatise, and includes several extracts from commentaries on the texts. W.M. Thackston, Jr, *The mystical and visionary treatises of Shihabuddin Yahya Suhrawardi*, London: Octagon Press, 1982, provides an English translation of most of the treatises in vol. III of *Œuvres philosophiques et mystiques*, which eschews all but the most basic annotation; it is therefore less useful than Corbin's translation from a philosophical point of view.)

al-Suhrawardi [Sohravardi, Shihaboddin Yahya] (1154-91) *Hayakil al-nur (The Temples of Light)*, ed. M.A. Abu Rayyan, Cairo: al-Maktaba al-Tijariyyah al-Kubra, 1957.(The Persian version appears in *Œuvres* vol. III.)

al-Suhrawardi [Sohravardi, Shihaboddin Yahya] (1180?-91) *Mantiq al-talwihat*, ed. A.A. Fayyaz, Tehran: Tehran University Press, 1955.(The logic of the *Kitab al-talwihat (The Intimations)*.)

al-Suhrawardi [Sohravardi, Shihaboddin Yahya] (1186-91) *Kitab hikmat al-ishraq (The Philosophy of Illumination)*, trans H. Corbin, ed. and intro. C. Jambet, *Le livre de la sagesse orientale: Kitab Hikmat al-Ishraq*, Lagrasse: Verdier, 1986.(Corbin's translation of the Prologue and the Second Part (The Divine Lights), together with the introduction of Shams al-Din al-Shahrazuri and liberal extracts from the commentaries of Qutb al-Din al-Shirazi and Mulla Sadra. Published after Corbin's death, this copiously annotated translation gives to the reader without Arabic immediate access to al-Suhrawardi's illuminationist method and language.)

References and further reading

Amin Razavi, M. (1997) *Suhrawardi and the School of Illumination*, Richmond: Curzon.(Clear and intelligent account of the main principles of his thought.)

Corbin, H. (1971) *En Islam iranien: aspects spirituels et philosophiques*, vol. II: *Sohrawardi et les Platoniciens de Perse*, Paris: Gallimard.(Corbin devoted more of his time to the study of al-Suhrawardi than to any other figure, and this volume represents the essence of his research.)

Ha'iri Yazdi, M. (1992) *The Principles of Epistemology in Islamic Philosophy: Knowledge by Presence*, Albany, NY: State University of New York Press.(An original work on epistemology by a contemporary Iranian philosopher drawing critical comparisons between certain Islamic and Western philosophers; incorporates the best exposition in a Western language of al-Suhrawardi's theory of knowledge.)

Nasr, S.H. (1983) 'Shihab al-Din Suhrawardi Maqtul', in M.M. Sharif (ed.) *A History of Muslim Philosophy*, vol. I, Wiesbaden: Otto Harrassowitz, 1963; repr. Karachi, no date. (Still one of the best short introductions to al-Suhrawardi, particularly useful on the cosmology.)

al-Shahrazuri, Shams al-Din (*c.*1288) *Sharh hikmat al-ishraq (Commentary on the Philosophy of Illumination)*, ed. H. Ziai, Tehran: Institute for Cultural Studies and Research, 1993.(Critical edition of the thirteenth-century original; Arabic text only, but a useful short introduction in English.)

Walbridge, J. (1992) *The Science of Mystic Lights: Qutb al-Din Shirazi and the Illuminationist Tradition in Islamic Philosophy*, Cambridge, MA: Harvard University Press, for the Centre for Middle Eastern Studies of Harvard University.(A study of one of al-Suhrawardi's principal commentators, with a useful introduction on the philosophy of illumination.)

Ziai, H. (1990) *Knowledge and Illumination: a Study of Suhrawardi's Hikmat al-ishraq*, Atlanta, GA: Scholars Press.(A pioneering study of al-Suhrawardi's logic and epistemology, particularly his criticism of the peripatetic theory of definition; unfortunately this work suffers from sloppy production.)

Ziai, H. (1996a) 'Shihab al-Din Suhrawardi: Founder of the Illuminationist School', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 434-64.(Biography of al-Suhrawardi.)

Ziai, H. (1996b) 'The Illuminationist Tradition', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 465-96.(General description of the Illuminationist tradition.)

al-Tawhidi, Abu Hayyan (*c.*930-1023)

Al-Tawhidi was an Arabic litterateur and philosopher, probably of Persian origin, and author of numerous books which reflect all the main themes of debate and reflection in the cultivated circles of his time. His basic outlook could be defined as a kind of simplified and vulgarized Neoplatonism, influenced by Gnostic elements, with four hypostases: God, Intellect, Soul and Nature. He also has a strong interest in moral questions on both the individual and social level.

‘Ali ibn Muhammad Abu Hayyan al-Tawhidi was probably of Persian origin. However, Arabic is the only language he is known to have used, and most of his life was spent in Baghdad and in Rayy (Tehran) at the court of the Buyid princes and their ministers, in particular the famous Ibn Sa‘dan. It is in the latter’s presence that the discussions recorded in *al-Imta‘ wa’-mu’anasa (Enjoyment and Conviviality)* took place. His last years were spent in Shiraz, where he died in AH 414/AD 1023.

Al-Tawhidi is a representative of Arabic belles-lettres (*adab*) rather than a philosopher in the strict sense. However, some of his main works report discussions devoted to philosophical themes and shed interesting sidelights on questions dealt with in a more systematic fashion by the great Arab philosophers. It goes with the genre adopted by al-Tawhidi that he rarely expresses his own opinions; his main authority is his master, Abu Sulayman al-Sijistani. He also appears to make extensive use of the *Rasa’il Ikhwan al-Safa’ (Epistles of the Brethren of Purity)*, although their name is rarely cited (see Ikhwan al-Safa’). Another source of inspiration is the ethical thinker Ibn Miskawayh, with whom he exchanged a philosophical correspondence, *al-Hawamil wal-shawamil (Rambling and Comprehensive Questions)*. Among Greek philosophers, Aristotle is by far the most commonly invoked authority.

Al-Tawhidi’s main philosophical work is *al-Muqabasat (Borrowed Lights)*. *Al-Imta‘ wa’-mu’anasa* also contains some philosophical material, besides some which is predominantly literary or grammatical. In metaphysics, he follows the basic Neoplatonic scheme of emanation (see Neoplatonism in Islamic philosophy). The First, frequently called the Creator, is the source of the world of nature which emanates continuously from him; God is thus also characterized by his generosity. Intellect, Soul and Nature are the three main levels of being, or hypostases, emanating from the First. The process is sometimes expressed in terms of illumination; the Intellect receives its light from the First, the Soul from the Intellect and Nature from the Soul. Elsewhere, the soul is considered as being pure light. Conversely, the First is said to encompass the Intellect, which in turn encompasses the Soul and so forth.

Many paragraphs are devoted to the human soul, concerning which al-Tawhidi takes up positions that can be defined as Platonic. The soul subsists by itself and is not tied down to the body; on the contrary, it uses the body as an instrument. The soul does not arise from the mixture of the elements; thus the Galenic theory is implicitly rejected, although it is ascribed in one passage to Zeno (probably meaning the Stoic) (see Galen; Zeno of Citium). The union with the body is described as a kind of fall in a way which has clear antecedents in some soteriological conceptions of Gnosticism and Neoplatonism (see Gnosticism; Neoplatonism). In the course of the soul’s descent from the heavenly realm it became covered in scales or veils, which it will cast off after physical death, that is, when it relinquishes the body. The soul becomes like a rusty mirror; just as the latter is no longer capable of reflecting external objects, the soul forgot what it knew in the intelligible world. Its true nature is also more fully active in sleep. In our ordinary waking life, we do not remember the world where our soul originated because we have been overcome by matter. These two states of the soul, incarnate and immaterial, correspond to the two realms of intellection and sense-perception. Intellection is an immediate form of apprehension, devoid of reflection and deliberation, whereas sense-perception is linked to discursive and inductive modes of thought, such as syllogism.

Humanity is thus in an intermediate position between the world of intellect and the world of nature. The latter is integrated into the emanationist scheme more neatly than is the case in Greek Neoplatonism; nature is a life force which emanates from the First Principle and penetrates all bodies, giving them their forms and linking them together. The Aristotelian definition of nature as principle of motion and rest is also quoted more than once.

Al-Tawhidi evinces a keen interest in linguistic questions. He predictably maintains the superiority of Arabic over other languages, but also discusses such topics as the respective place and function of prose and verse (see Aesthetics in Islamic philosophy). It is thanks to him that we have a report of the controversy between the partisans of logic and Greek culture and those of traditional Arabic grammar. Another discussion tackles the sensitive problem of the relationship between philosophy and religion. Among the current themes of his time, that of the characters and respective merits of the nations is taken up on several occasions. In the field of ethics, he devoted an entire epistle to friendship, and this is one of his more personal and interesting works.

See also: Ibn Miskawayh; Neoplatonism in Islamic philosophy; al-Sijistani

CHARLES GENEQUAND

List of works

al-Tawhidi (*c*.977-82) *al-Hawamil wa'-shawamil (Rambling and Comprehensive Questions)*, ed. A. Amin and A. Saqr, Cairo, 1951.(A correspondence with Ibn Miskawayh on a variety of topics.)

al-Tawhidi (*c*.983-5) *al-Imta' wa'-mu'anasa (Enjoyment and Conviviality)*, ed. A. Amin and A. al-Zayn, Cairo, 1953.(A collection of forty 'nights' or gatherings devoted to the discussion of literary, philosophical, religious, grammatical and other subjects.)

al-Tawhidi (*c*.996) *al-Muqabasat (Borrowed Lights)*, ed. M.T. Husayn, Baghdad, 1970.(A collection of 106 discussions, mainly theological and philosophical.)

al-Tawhidi (*c*.1010) *al-Sadaqa wa'-sadiq (Friendship and Friend)*, ed. I. Keilani, Damascus, 1964.(A short personal essay on friendship.)

References and further reading

Berge, M. (1979) *Pour un humanisme vécu: Abu Hayyan al-Tawhidi (Towards a Living Humanism: Abu Hayyan al-Tawhidi)*, Damascus: Institut Français de Damas, 1979.(A general study of Tawhidi's life, works and thought, the only such in a European language.)

Leaman, O. (1996) 'Islamic Humanism in the Fourth/Tenth Century', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 10, 155-64.(Analysis of the nature of this important period of Islamic philosophy, concentrating on the ethical thought of the period.)

Rowson, E.K. (1990) 'The Philosopher as Littérateur: al-Tawhidi and his Predecessors', *Zeitschrift für Geschichte der Arabisch-Islamischen Wissenschaften* 6: 50-92.(Short study of al-Tawhidi in English.)

Alterity and identity, postmodern theories of

Theories of alterity and identity can be said to be 'postmodern' if they challenge at least two key features of modern philosophy: (1) the Cartesian attempt to secure the legitimacy of knowledge on the basis of a subject that immediately knows itself and (2) the Hegelian attempt to secure self-knowledge and self-recognition by showing that knowledge and recognition are mediated by the whole. Postmodern thought does not necessarily champion a wholly other, but it generally conceives of self-identity in terms of a radical alterity.

1 Features of postmodern thought

Postmodern theories of identity and alterity refer to the use of these terms in modern philosophy. But the relation between modernity and postmodernity is not a simple matter, for much of what has come to be known as postmodern thought seeks to discover a 'postmodern moment' at the inception of modernity. According to Jean-François Lyotard, whose *Postmodern Condition* (1984) ignited many of the philosophical debates concerning the status of postmodernity, the various manifestations of modernity - in architecture, literature, art and philosophy - can be seen as retreats from these postmodern moments, and one of the tasks of postmodern thought is to retrieve the gestures and motifs that modernity has been compelled to erase in order to institute itself as an ever renewable project or method. Postmodernity appears in its own right once these projects and methods can no longer *guarantee* their own legitimacy.

In order to understand the 'postmodern condition' it is therefore necessary to pay attention to the breakdown of the various 'meta-narratives' by which modernity has tried to legitimize itself. To the extent that modern philosophy from its inception in Descartes' *Meditations* sets out to secure the validity of knowledge, postmodern thought can be understood as a wide-ranging effort to come to terms with - and not simply denounce or repair - the failure of all philosophical attempts to secure the legitimacy of knowledge. The concepts of identity and alterity are of particular importance for this effort because the Cartesian attempt to secure the legitimacy of knowledge finds its principle point of reference in the identity of the self-conscious subject. This subject can serve as the source of legitimation to the extent that it can immediately identify itself and can treat its act of self-identification as knowledge. Postmodern theories of identity and alterity are concerned for the most part with the nature of self-identity and with the relation between the self and whatever presents itself as other than the self. The following discussion does not seek to cover all of the writers who have been classified as postmodern but first defines one version of postmodern philosophy and then considers certain exemplary thinkers for whom the nature of self-identity and the relation of the self to the other are of paramount importance.

If modern philosophy rests on the principle of self-consciousness, then one criterion for a postmodern philosophy would be its contesting of this principle. Yet postmodern philosophy cannot simply demonstrate that the self-conscious subject is unable to secure the legitimacy of knowledge by an act of immediate self-identification: not only are such demonstrations part and parcel of many modern philosophical projects, Hegel's attempt to complete modern philosophy begins by pointing out the abstract character of the Cartesian ego. In order for postmodern thought to distinguish itself from its modern counterparts it cannot simply assert that the identity of the self is derived from - or, in the Hegelian version, mediated by - something else; the other that allows the self to identify itself cannot be, as for Hegelians, the whole; it must be the *wholly* other, not the other *of* the self but an other that no longer pertains to the self. Even if this other grants the subject its identity, it must retreat from all attempts on the part of philosophical methods or systems to grasp it *as* something, even as something 'other'. The importance of Nietzsche and Freud for the development of postmodern thought can be seen from this perspective, for they provide intricate analyses of consciousness as a secondary process and also propose something - the will, the unconscious - that withdraws from consciousness but cannot then be reinterpreted in terms of a metaphysical substance (as can Schopenhauer's will or Hartmann's unconscious). Postmodern thought does not necessarily champion a 'wholly other', but it declines the Hegelian proposal that the legitimacy of knowledge, including self-knowledge and self-consciousness, should be sought in the concept of a self-mediating whole.

2 The dialectics of desire

The rejection of Hegelian accounts of the relation between the self and the other can be counted among the criteria of postmodern thought. A decisive moment in the development of such thought - and a reason it has developed

most conspicuously in France - is found in Alexandre Kojève's striking interpretation of Hegel's 'master-slave dialectic' as a logic of human desire (see Kojève, a.). According to Kojève, the object of desire is never a tangible thing; every desire is a desire for recognition by the other. The slave wants to be recognized by the master; but since the master cannot be recognized by the slave without ceasing to be a master, he must look toward the 'absolute master' - death - to secure recognition. Self-negating (suicidal) desire propels history toward its completion, at which point everyone recognizes everyone else; but this stage of satisfied desire, 'the Sunday of History', is itself deeply problematic because all the categories by which someone can be recognized as something have been exhausted. By taking seriously Hegel's talk of the end of history and the end of philosophy, Kojève set the stage not only for postmodern thought but also for a host of 'post-philosophical' and 'post-historical' pronouncements.

Some of those who heard Kojève's lectures on Hegel, especially Georges Bataille and Jacques Lacan, developed new concepts of the subject from his dialectics of desire. Bataille (1985) wrote extensively of a sovereign subject which is always other than itself precisely because it never recognizes itself in any other, and Lacan found in Freudian psychoanalysis a subject of dreams and slips of the tongue, a subject which is therefore no longer in control of, or even present to, itself. According to Lacan's famous theory of the 'mirror stage', the self does not constitute itself by recognizing itself in its other but by identifying itself with its specular image: its identity is thus based on misrecognition (*méconnaissance*). The Other, who is not to be confused with another self or an *alter ego*, 'mediates' between the self and its specular counterpart. According to Lacan, Freud discovers a 'radical heteronomy… gaping within man' and responds to the question 'who, then, is this other to whom I am more attached than to myself, since at the heart of my assent to my identity it is still he who agitates me?' ('Agency of the Letter', in Lacan 1977) by conceiving of the other side of consciousness in a radically new way. Without 'the discourse of the Other' - which is one of Lacan's names for the unconscious - there would be neither the desire for recognition nor recognition of desire, for no object is adequate to desire. The locus of such inadequation is language, more exactly, 'the signifier' understood as that which forever slips away from signification. Lacan then defines desire as 'the desire of the Other', but this definition does not mean that the ego wants what an *alter ego* desires; rather, it indicates that the self desires only *as* the Other and thus cannot recognize the object of its desire. The self is, as it were, a detour in the trajectory of the desire of the Other.

Lacan's theory of desire arrives at a conundrum: how can we understand the fact that subjects desire very specific objects? He invented the term 'object a' (where the 'a' stands for *autre*, 'other') in response to this question. Such objects are distinguished by the fact that 'they have no specular image, or, in other words, alterity' ('Subversion of the Subject', in Lacan 1977). When Lacan counts the subject of consciousness among these 'other objects', he makes the question of self-identity considerably more complicated, for, on the one hand, the self is constituted by misidentifying itself with its specular image, and, on the other, it has no specular image but is only, as Lacan says, a 'shadow'. One trenchant response to the complications generated by Lacan's attempt to understand the specificity of objects of desire can be found in the work of Julia Kristeva, especially in her *Powers of Horror* (1982). Kristeva articulates the Lacanian analysis of alterity into three distinct moments - the Other, the *alter ego*, and the 'other' as object of desire - by concentrating on what cannot be captured by a dialectics of desire: the utterly undesirable or 'abject'. The self experiences abjection when it senses an undefinable 'something' that precedes, inhabits, and threatens to engulf it. For this reason, the abject is violently - but also only incompletely - expelled:

> I experience abjection only if an Other has settled in place and stead of what will be 'me'. Not at all an other with whom I can identify and incorporate, but an Other who precedes and possesses me, and through such possession causes me to be.
>
> (Kristeva 1982: 10)

Certain categories of religious discourse such as defilement, abomination and purification can be understood, according to Kristeva, in terms of abjection. Once the abject is excluded, it serves as the foundation of an always precarious culture within which objects of desire can be separated from one another.

3 The experience of alterity

The writings of Lacan and Kristeva are as much contributions to psychoanalytic practice as independent

theoretical exercises. Yet the questions to which they are addressed - the dialectics of desire, the experience of the alterity of the self - have also been posed by Emmanuel Levinas from a very different perspective. While studying the work of Husserl and Heidegger, Levinas came to realize that the phenomenology of the other cannot be accomplished in the same manner as the phenomenology of consciousness or the hermeneutics of existence. Levinas opens his most extensive work, *Totality and Infinity* (1969), by defining desire as 'desire for the absolutely other'. Because every object is inadequate to desire, its meaning must lie in the 'alterity of the Other [*autrui*]'. The phenomenological elucidation of the *autrui* (the 'personal' other) cannot simply be a matter of theoretical attitudes and descriptions because the *autrui* is never a definable theme. Instead of grounding philosophy on the *cogito*, Levinas returns to Descartes' discussion of 'the idea of infinity' as that which 'overflows' every intentional state. *Totality and Infinity* then presents subjectivity 'as welcoming the Other, as hospitality; in it the idea of infinity is consummated' (1969). Not only is the totality of entities to be distinguished from the totally other, so too is the phenomenon from the face: in the naked face of the *autrui* the idea of infinity overflows consciousness and its objects. The other can never be reduced to the same - not to the identity of the *cogito* (epistemology) nor to the sameness of being (ontology) - and so ethics becomes, for Levinas, 'first philosophy'.

In his second major work, *Otherwise than Being* (1981), Levinas no longer speaks of subjectivity as hospitality but as hostage: the self is not only held hostage by the other, but, as a hostage, it also 'substitutes' - and thus takes responsibility - for the other who holds it hostage. This paradoxical responsibility cannot be represented in terms of a self-positing subject, for, as the term 'hostage' indicates, the self is from the start sheer passivity:

> The uniqueness of the ego, overwhelmed by the other in proximity, is the other in the same, the psyche. But it is I, I and no one else, who am hostage for the others. In substitution my being that belongs to me and not to another is undone, and it is through this substitution that I am not 'another', but me.
>
> (1981: 116)

It is only because the self is assigned to the other before it acts on its own that it can be itself, that is, singular. Throughout his writing Levinas discovers an unmediated alterity in every identity: the vulnerability, susceptibility and 'nudity' of the self is evidence of such alterity. For philosophy to come to terms with evidence of this kind it must abandon idealism as well as empiricism and revise its notions of experience and sensibility. Experience does not consist in subsuming mental representations under general terms but in taking responsibility for the other and exposing oneself to one's own alterity.

Although the work of Gilles Deleuze could hardly be more different in tone and texture than that of Levinas, he too seeks to revise philosophical concepts of experience and sensibility, and, like Levinas, his programme takes its point of departure from a critique of all claims to identity. But unlike Levinas, he does not undertake this critique as an advocate, so to speak, of the transcendence of the other but as a champion of entirely immanent 'differential forces'. Whereas Kant treats identity and difference as concepts of reflection, Deleuze tries to develop a concept of difference in which it is no longer a term of reflection and can no longer be seen as the opposite of identity. Difference, for Deleuze, always implies a multiplicity of relations among positive forces, each of which expresses itself but none of which opposes any other. According to the terms set forth in *Difference and Repetition* (1981), every 'philosophy of representation' rests on a principle of identity. The reduction of difference to opposition serves to make singularities - and, for Deleuze, everything is a singularity - into representations of generic types. Every philosophy of representation denies difference in favour of identity and thus turns into a philosophy of negation. Difference must be negated in order to save the self-identical subject, even if - as in the case of Hegelianism - this negation is doubled, and the subject appears only as the negation of everything that opposes it.

All of Deleuze's writings emphasize the positivity of difference. *Anti-Oedipus* (Deleuze and Guattari 1977), perhaps his best-known book, not only attacks psychoanalysis for conceiving of desire in terms of castration (lack of the phallus) but seeks to undermine every conception of desire as absence or negativity. The question then arises: how can one account for alterity without using concepts such as lack, absence or negativity? The concluding sections of *Difference and Repetition* respond to this question. Sensible differences, according to Deleuze, are always differences of intensity: 'Intensity is the form of difference in so far as this is the reason of the sensible' (1994: 222). In order for an intensity to be experienced, it must first be developed or 'explicated'. But an

intensity (a *quale*) cannot be explicated unless it is transformed into an extension (a quantity) in accordance with a specific principle of identity. The ‘psychic system I-Self’ is one particularly complicated version of explication. The other enters into Deleuze’s philosophical scenario as the site of still unexplicated intensities. Instead of using words like ‘negativity’ or ‘absence’ to conceive of alterity, Deleuze employs a modal version of these terms - possibility - and he, like Levinas, finds in the human face the exemplary experience of alterity: ‘Consider a terrified face…. This face expresses a possible world: the terrifying world…. In every psychic system there is a swarm of possibilities around reality, but our possibles are always Others’ (1994: 260). The self, in sum, is the explication of implied intensities in accordance with a principle of identity, whereas the other, as an expression of a possible world, ‘represents’ the otherwise inaccessible intensities on which the ‘psychic system’ is founded.

4 Aporias of the wholly other

Jacques Derrida, like Deleuze, has developed a philosophy of difference, but unlike Deleuze, he does not present it as an alternative to the philosophy of identity. Derrida’s philosophical enquiries respond to a double exigency: they seek to show, on the one hand, that the operation of the principle of identity always rests on an unacknowledged play of difference and, on the other, that neither difference nor alterity can serve as principles on which a new philosophical project can be built. Derrida begins one of his earliest writings, ‘Descartes and the History of Madness’, by showing that the project Foucault undertakes in *Madness and Civilization* (1965) cannot be accomplished: it is not only impossible to let madness ‘itself’ speak, but it is also impossible to write an ‘archaeology’ of the silence into which the mad are driven without relying on the resources of reason (*logos*). Derrida then interrogates the very opposition between madness and reason by pointing toward a madness (or being *aliéné*) at the heart of modernity’s first philosophical manifesto, Descartes’ *Meditations*: in the process of constituting itself, according to Derrida’s reading, the self-reflective subject exposes itself to a doubt so hyperbolic that it exceeds every effort on the part of the self-reflective subject to bring it under control.

Because it is impossible to find a position from which to criticize the principle of identity without defining this position in terms of self-identity, Derrida refrains from describing his work as a ‘critique’ and prefers the word ‘deconstruction’ (see Deconstruction). The deconstruction of the principle of identity not only demonstrates the impossibility of any critique of identity in the name of alterity but welcomes this impossibility, for it implies that the self can never be entirely separated from the other. Philosophy cannot disclose a pure self or a pure other, but this impossibility cannot be resolved in a Hegelian manner by representing the subject as the process of self-mediation in the other. Yet it is *also* impossible, according to Derrida, simply to refuse this Hegelian move and once again presuppose, posit or postulate a wholly other. There is, for Derrida, such an other, but it does not exist, if ‘existence’ is defined by self-identity and presence-to-oneself. A philosophical enquiry cannot set out from or arrive at the wholly other; it *comes* - without being destined and without destination. Insofar as the future (*avenir*) is opened up by this coming (*venir*), it is the venue of alterity, and insofar as justice (in contrast to legal norms) always implies a singular relation to a singularity, according to Derrida, this future is also the site of justice.

All of the aporias surrounding the idea of a radical alterity are captured in one of Derrida’s most suggestive sentences: *tout autre est tout autre*, ‘every other is wholly other’ (1994: 82). This assertion of the radical and irreducible alterity of every other is at the same time a radical assertion of identity, for the sentence can also be understood as a tautology: ‘every other is every other’. In a single sentence Derrida thus captures one of the great challenges of all postmodern thought: by championing a radical alterity, it runs the risk of turning into a tautological affirmation of identity. Derrida emphasized the impossibility of safeguarding any thought of the wholly other from this risk, but he also insists that thought can never run away from such risks.

See also: Postmodernism; Subject, postmodern critique of

PETER FENVES

References and further reading

Bataille, G. (1985) *Visions of Excess*, trans. A. Stoekl, Minneapolis, MN: University of Minnesota Press. (Important, difficult, but often readable essays on many of the topics that have come to occupy postmodern thought.)

Baudrillard, J. (1988) *Selected Writings*, ed. M. Poster, Stanford, CA: Stanford University Press.(A good, very

accessible introduction to an influential postmodern sociologist.)
Bertens, J. (1995) *The Idea of the Postmodern*, New York: Routledge.(A readable introduction to the history of postmodernism.)
Borch-Jacobsen, M. (1991) *Lacan: The Absolute Master*, trans. D. Brick, Stanford, CA: Stanford University Press.(An advanced introduction to Lacan and the philosophical atmosphere in which he developed his version of psychoanalysis; an excellent, well-written, although at times difficult book.)
Deleuze, G. (1994) *Difference and Repetition*, trans. P. Patton, New York: Columbia University Press.(A major philosophical statement; the most accessible of Deleuze's independent philosophical works.)
Deleuze, G. and Guattari, F. (1977) *Anti-Oedipus*, trans. R. Hurley, M. Seem and H. Lane, New York: Viking. (One of the principal points of reference for postmodern philosophy.)
Derrida, J. (1978) *Writing and Difference*, trans. A. Bass, Chicago, IL: University of Chicago Press.(Important, often difficult essays in which Derrida works out his relation to contemporaneous attempts to rethink the relation of the self to other; especially significant are the essays on Foucault, Levinas, Bataille, and Freud.)
Derrida, J. (1993) *Aporias*, trans. T. Dutoit, Stanford, CA: Stanford University Press.(A good introduction to Derrida's thought, especially his way of coming to terms with the paradoxes of alterity.)
Derrida, J. (1994) *The Gift of Death*, trans. D. Wills, Chicago, IL: University of Chicago Press.(A highly readable exploration of the relation between giving, dying, sacrificing and secrecy.)
Descombes, V. (1980) *Modern French Philosophy*, trans. L. Scott-Fox and J. Harding, Cambridge: Cambridge University Press.(Originally titled '*The Same and the Other*', this is a good, critical introduction to many of the debates surrounding postmodern philosophy.)
Foucault, M. (1965) *Madness and Civilization*, trans. R. Howard, New York: Pantheon.(One of Foucault's most accessible books. As an 'archaeology' of the silence to which the mad were condemned in the 'age of reason', it is the point of departure for many re-evaluations of modern rationality.)
Habermas, J. (1987) *The Philosophical Discourse of Modernity*, trans. F. Lawrence, Cambridge, MA: MIT Press. (Important and influential essays attacking postmodern thought as a retreat from the unending project of modernity and enlightenment. Important for understanding all of the philosophical debates around the significance and validity of postmodern theory.)
Jameson, F. (1991) *Postmodernism, or, The Cultural Logic of Late Capitalism*, Durham, NC: Duke University Press.(A wide-ranging, easily accessible description and evaluation of various versions of postmodernism from a Marxist perspective.)
Hoesterey, I. (1991) *Zeitgeist in Babel: The Postmodernist Controversy*, Bloomington, IN: Indiana University Press.(A collection of essays on postmodernism in a variety of fields with an informative section on the political and social stakes of postmodern theory.)
Kojève, A. (1969) *Introduction to the Reading of Hegel*, trans. J. Nichols, Ithaca, NY: Cornell University Press. (One of the seminal books of twentieth century French thought. Familiarity with Hegel and Marx is indispensable for understanding this book.)
Kristeva, J. (1982) *Powers of Horror*, trans. L. Roudiez, New York: Columbia University Press.(A series of essays on abjection, selfhood, religious discourses and practices; includes a lengthy confrontation with abjection and horror in the writings of Céline.)
Lacan, J. (1977) *Écrits*, trans. A. Sheridan. New York: W.W. Norton.(Very difficult essays, but indispensable for an understanding of postmodern discourse on identity and alterity.)
Levinas, E. (1969) *Totality and Infinity*, trans. A. Lingis, Pittsburgh, PA: Duquesne University Press.(Levinas's most extensive treatise; it is generally quite readable, and it is important for all subsequent attempts to conceive of radical alterity.)
Levinas, E.(1981) *Otherwise than Being, or, Beyond Essence*, trans. A. Lingis, The Hague: Martinus Nijhoff, 1981.(One of Levinas's most important philosophical statements. Familiarity with phenomenological discourse helpful but not indispensable.)
Levinas, E.(1989) 'Substitution', in S. (ed.) *The Levinas Reader*, Oxford: Blackwell.(One of the preliminary drafts of *Otherwise than Being*, this essay argues for the idea of a responsibility that is justified by no prior commitment.)
Lyotard, J.-F. (1984) *The Postmodern Condition: A Report on Knowledge*, trans. G. Bennington and B. Massumi, Minneapolis, MN: University of Minnesota Press.(A very readable book that is the starting points for many of the debates about postmodern thought and culture.)
Lyotard, J.-F. (1988) *The Differend*, trans. G. van den Abbeele, Minneapolis, MN: University of Minnesota Press. (Lyotard's most thorough philosophical work; often difficult, with allusions to a wide variety of philosophical theories.)

Norris, C. (1990) *What's Wrong with Postmodernism: Critical Theory and the Ends of Philosophy*, Baltimore, MD: Johns Hopkins University Press.(A series of essays that attempt to stake out a version of postmodern philosophy and criticism that are heirs to enlightenment thought. Polemical and generally accessible.)

Pefanis, J. (1991) *Heterology and the Postmodern: Bataille, Baudrillard, and Lyotard*, Durham, NC: Duke University Press.(A readable introduction to postmodern theory which pays particular attention to the question of alterity.)

Rorty, R. (1989) *Irony, Contingency, Solidarity*, Cambridge: Cambridge University Press.(A collection of essays that have served as reference points for defining an influential version of postmodern thought.)

Vattimo, G. (1988) *The End of Modernity: Nihilism and Hermeneutics in Postmodern Culture*, trans. J. Snyder, Baltimore, MD: Johns Hopkins University Press.(Accessible meditations on both the general problem of postmodernity and the nature of postmodern philosophy.)

Althusser, Louis Pierre (1918-90)

Louis Althusser was the most influential philosopher to emerge in the revival of Marxist theory occasioned by the radical movements of the 1960s. His influence is, on the face of it, surprising, since Althusser's Marx is not the theorist of revolutionary self-emancipation celebrated by the early Lukács. According to Althusser, Marx, along with Freud, was responsible for a 'decentring' of the human subject. History is 'a process without a subject'. Its movement is beyond the comprehension of individual or collective subjects, and can only be grasped by a scientific 'theoretical practice' which keeps its distance from everyday experience. This austere version of Marxism nevertheless captured the imagination of many young intellectuals by calling for a 'return to Marx', with the implication that his writings had been distorted by the official communist movement. In fact, Althusser later conceded, his was an 'imaginary Marxism', a reconstruction of historical materialism reflecting the same philosophical climate that produced the post-structuralist appropriations of Nietzsche and Heidegger by Deleuze, Derrida and Foucault. Most of the philosophical difficulties in which Althusser found himself can be traced back to the impossibility of fusing Marx's and Nietzsche's thought into a new synthesis.

1 Life

On a pessimistic view of his influence, Althusser may be chiefly remembered for strangling his wife, Helene Rytman, on 16 November 1980, in their flat at the École Normale Superieure in Paris. Deemed unfit to stand trial for the murder, Althusser spent the last decade of his life in and out of mental hospitals. His attempt to explain this disaster - in *L'Avenir dure longtemps (The Future Lasts A Long Time)* (1992), an extraordinary confessional autobiography, posthumously published, which traced the manic depressive outbursts which dogged his adult life back to a Freudian family drama which began even before his birth in Algeria - has been shown by Yann Moulier Boutang (1992) to be as much fiction as fact.

Thanks, however, to his own memoirs and Moulier Boutang's biography, the outline of Althusser's public life is now becoming clear. From his Lyons schooldays in the 1930s until the early 1950s he was active in Catholic political circles. But, whereas before the Second World War Althusser was associated with the anti-republican right ('We were more or less… royalists then', he later wrote), the four years he spent in a German prisoner-of-war camp after France's defeat in 1940 pushed him to the left. Like many young French intellectuals of his generation, he rallied to the communist party after the war, joining in 1948. Until the Church's suppression of the worker-priest movement in the early 1950s Althusser sought to reconcile Marxism and Catholicism. In these years he espoused a messianic Hegelian Marxism, in some ways reminiscent of the early Georg Lukács (see Lukács, G. §2). The thesis for which Althusser was to become famous in the 1960s - that Marx had had to break with his own youthful Hegelian past - therefore recapitulated his own personal development.

In July 1948 Althusser was appointed *caïman* at the École Normale Superieure, a post he was to hold until the tragedy of November 1980. He was responsible for preparing philosophy students for the *agrégation*, or final examination. Althusser was an outstanding teacher, and the central role which the École Normale played in the teaching of philosophy in France (virtually every major twentieth-century French philosopher has been a *normalien*) meant that he had an exceptional influence on French intellectuals for a generation. But it was his writings of the 1960s - notably those in *Pour Marx (For Marx)* and *Lire le Capital (Reading Capital)*, both published in 1965 - which made Althusser a major figure on the intellectual scene. His reinterpretation of Marx, developed in collaboration with an exceptionally talented group of pupils, served to articulate young left-wing intellectuals' impatience with the caution and conservatism (as they saw it) of the communist party leadership. Unlike many of his followers, Althusser did not break with the party, even after what he regarded as its failure to seize the revolutionary opportunity presented by the events of May-June 1968. He found himself, however, increasingly at odds with the official apparatus. His last major public act was 'What Must Change in the Party', a root-and-branch denunciation of the leadership published in *Le Monde* in April 1978. By then, however, the Parisian intelligentsia had fallen out of love with Marxism, and Althusser's last years were spent in intellectual as well as personal isolation.

2 Rereading Marx

Althusser's reinterpretation of Marx involves, in the first place, a theory of interpretation. 'There is no such thing

as an innocent reading', he says. Every approach to a text brings with it certain theoretical presuppositions. Therefore 'a new theory of ''reading"' is required, one that is concerned to identify the theoretical framework implicit in the text itself, or what Althusser calls its 'problematic'. The problematic, he claims, is typically concealed within the text rather than visible on its surface. It can be detected, rather like the repressed desires which psychoanalysis supposes it uncovers, in the text's silences, gaps, ambiguities and inconsistencies. Texts are thus complex, and this complexity is, as we shall see, one aspect of the complexity of history itself.

When applied to Marx, this theory of reading reveals a discontinuity. Marx's writings up to and including the *Economic and Philosophic Manuscripts of 1844* belong to one problematic which Althusser regards as ideological. Their leading characteristic is their humanism: history is depicted as the unfolding of a human essence which, alienated under capitalism, will find fulfilment under communism. These early works of Marx are separated from his later writings, above all *Capital*, by an 'epistemological break' marking the formation of a new science of history. The mature Marx is a 'theoretical anti-humanist', for whom history is 'a process without a subject or goal'. It is not the development of the productive forces, or the self-realization of the working class conceived (as it was by the early Lukács) as the 'identical subject-object' of history, nor is socialist revolution the inevitable culmination of the historical process. Althusser liked to say that 'the materialist… is a man who catches a moving train without knowing where it has come from or where it is going' (1992: 210). Human beings are the 'bearers' of a process they neither create nor control.

The break from which Althusser's scientific, anti-humanist Marx emerged was above all one with Hegel. Althusser's most celebrated essay, 'Contradiction and Overdetermination' (1962; repr. in *For Marx*), was a direct attack on the idea, hitherto basic to Marxist orthodoxy, that Marx, while rejecting Hegel's idealist system, had taken over his dialectical method. Althusser pointed out that this contrast between method and system involved a separation of form and content that was itself 'pre-dialectical', contradicting Hegel's own account of his method. The formulation of a materialist dialectic required not the application of Hegel's method to a different object but 'the transformation of its structures'. Althusser sought to bring out the difference between Marx and Hegel by contrasting their conceptions of totality. As Lukács had argued, both might view society as an integrated whole, but the way in which they conceived this whole was very different. Hegel's was an 'expressive totality', in which all the different aspects of social life reflected a single informing centre. Taken over in a Marxist framework, this totality leads to economic reductionism; everything becomes an expression of the fundamental economic contradiction, in the way that, according to Lukács, reification pervades the whole of capitalist society (see Marxism, Western §2).

For Marx, by contrast, each social formation is a complex, structured totality composed of a plurality of practices irreducible to one another. The economy is 'determinant in the last instance', that is, its primacy consists not so much in its directly shaping the course of historical development, as in its selecting some particular 'instance' (or practice) to play the dominant role (politics, for example, in feudal society). Economic causality thus operates indirectly, and always in combination with non-economic practices. As Althusser famously put it, 'the lonely hour of the "last instance" never comes' ([1965] 1969: 113). What Marx called the superstructure - politics, law and ideology - is 'relatively autonomous': each practice develops according to its own specific laws, within the limits set for it by the economy. Thus major historical events, for example, the Russian Revolution of February 1917, are 'overdetermined': they are not simply expressions of an underlying contradiction between the forces and relations of production, but involve a combination of heterogeneous factors - economic, political, ideological - which accumulate until they produce a rupture.

This complexity of the social whole is one reason why Althusser does not believe that individuals can understand the history in which they are caught up. Ideology, he believes, is a necessary feature of any society, including the classless communism of the future. It serves as a factor of social cohesion, adapting human beings to the roles required of them as bearers of the prevailing relations of production. The very form of subjectivity - individuals' conception of themselves as coherent and autonomous persons - is the means through which they are subsumed under ideological social relations. The only way out of this imaginary relationship lies in scientific theory. The formation of a scientific problematic - for example, Marx's epistemological break - allows an escape from the repetition of a few stereotyped ideological themes into a potentially infinite process of theoretical self-development. Here too, however, individuals figure only as the bearers of an impersonal social process, since each science develops according to the conceptual patterns specific to it.

3 Conclusion

Though carried out under the slogan of a ‘return to Marx’, his rescue from the misinterpretations perpetrated both by orthodox Communists and by Hegelians such as the early Lukács, Althusser’s reading of Marx was on his own admission a ‘guilty’ one, reflecting a variety of influences. He himself paid explicit tribute to Spinoza and Freud in particular, and the debt his philosophy of science owed to his teacher Bachelard is evident. More generally, Althusser’s anti-humanism bears a strong resemblance to the attack on the subject as a unified and sovereign entity which, initiated in the 1950s by Lacan and Levi-Strauss, was to be radicalized in the late 1960s by Deleuze, Derrida and Foucault (see Post-structuralism §§2-4). But post-structuralism derives ultimately from Nietzsche’s critique of modernity, with its tendency (never fully realized in Nietzsche himself, but taken much further in his own way by Heidegger) to reject the Enlightenment *tout court*. How could Marxism, deeply embedded in which are preoccupations with progress and emancipation, place itself in the camp of the counter-Enlightenment? More specifically, in rejecting economically reductionist versions of Marxism, was not Althusser in danger of developing a pluralist conception of the social whole in which the economy had lost any causal primacy? Again, if each ‘theoretical practice’ developed autonomously according to its own internal norms, how could Althusser maintain a distinction between science and ideology and thereby resist Foucault’s Nietzschean reduction of all knowledge to expressions of the will to power? Althusser’s philosophical enterprise ultimately ran aground on these and similar difficulties. Nevertheless, his insistence on a careful conceptual analysis of Marx’s writings as the basis of a reconstruction of historical materialism, his demonstration of the methodological gulf separating Marx and Hegel, and his sensitivity to developments in the philosophy of science that were ignored by the Frankfurt School, for example, make his work of lasting value.

See also: Dialectical materialism §2

ALEX CALLINICOS

List of works

Althusser, L. (1959) *Montesquieu, La politique et l’historie*, Paris: Presses Universitaires de France; trans. B. Brewster in *Politics and History*, later renamed *Montesquieu, Rousseau, Marx*, London: New Left Books, 1972.(A stimulating interpretation of Montesquieu as both a theorist of absolutism and one of the first historical sociologists; the English edition contains important essays on Rousseau’s *Social Contract*, and on Marx’s relationship with Hegel.)

Althusser, L. (1965) *Pour Marx*, Paris: Maspero; trans. B. Brewster, *For Marx*, London: Allen Lane, 1969. (Althusser’s most famous book, a collection of essays that outline his theory of overdetermination and his account of the ‘epistemological break’ separating the young Marx from the old.)

Althusser, L., Balibar, E., Establet, R., Macherey, P. and Ranciere, J. (1965) *Lire le Capital*, Paris: Maspero; partly trans. B Brewster in Althusser, L. and Balibar, E., *Reading Capital*, London: New Left Books, 1970.(The most sustained attempt by Althusser and his pupils to characterize what distinguishes Marx’s method from the Hegelian dialectic; it helped to stimulate close philosophical scrutiny of the conceptual structure of *Capital*.)

Althusser, L. (1969) *Lénine et La philosophie*, Paris: Maspero; trans. B. Brewster in *Lenin and Philosophy and Other Essays*, London: New Left Books, 1972.(Advances the notoroius thesis that ‘philosophy is, in the last instance, the class struggle in theory’; the English edition also includes two of Althusser’s most influential essays: ‘Freud and Lacan’ and ‘Ideology and the Ideological State Apparatuses’.)

Althusser, L. (1970) ‘Sur le Rapport de Marx à Hegel’, in J. D’Hondt (ed.) *Hegel et la pensée moderne*, Paris: Presses Universitaires de France; trans. B. Brewster, ‘Marx’s Relation to Hegel’, in *Politics and History*, London: New Left Books, and New York: Verso, 1972.(Argues that Marx understood Hegel’s dialectic in terms of process.)

Althusser, L. (1973) *Réponse à John Lewis*, Paris: Maspero; trans. G. Locke in *Essays in Self-Criticism*, London: New Left Books, 1976.(Althusser’s polemical defence of his ‘theoretical anti-humanism’ at the height of his Maoist phase.)

Althusser, L. (1974a) *Éléments d’autocritique*, Paris: Hachette; trans. G. Locke in *Essays in Self-Criticism*, London: New Left Books, 1976.(A self-criticism which takes *For Marx* and *Reading Capital* to task for ‘theoreticism’ and ‘flirting with strucuralism’.)

Althusser, L. (1974b) *Philosophie et philosophie spontanée des savants*, Paris: Maspero; trans. W. Montag in *Philosophy and the Spontaneous Philosophy of the Scientists and Other Essays*, London: Verso, 1990. (Althusser’s main discussion of the relationship between philosophy and the sciences; the English edition also

reprints ‘Lenin and Philosophy’ and other important texts of the 1960s and 1970s.)

Althusser, L. (1978) *Ce qui ne peut pas plus durer dans le parti communiste*, Paris: Maspero; trans. P. Camiller, ‘What Must Change in the Party’, *New Left Review* 109: 19-45. (First published in *Le Monde*, 24-7 April 1978, in the wake of the left’s defeat in the French legislative elections; a fierce indictment of the communist party’s Stalinist internal regime.)

Althusser, L. (1988) ‘Machiavelli’s Solitude’, trans. T. O’Hagan, *Economy and Society* 17 (4): 468-79.(A lecture on Machiavelli, a thinker who was a major influence on Althusser.)

Althusser, L. (1992) *L’avenir dure longtemps suivi de les faits*, Paris: Stock/IMEC; trans. R. Veasey, *The Future Lasts a Long Time*, London: Chatto & Windus, 1993.(Two autobiographical texts; the first and longer one, written after Althusser’s disgrace, is a remarkable human document, and contains important reflections on his philosophical and political career. The second edition, published by Stock/IMEC in 1994, contains some important additional material, including a text in which Althusser discusses Machiavelli and Spinoza as materialist precursors of Marx.)

Althusser, L. (1993) *Écrits sur la psychanalyse (Writings on Psychoanalysis)*, Paris: Stock/IMEC.(Brings together all Althusser’s writings on Freud, thereby helping to clarify the significant influence that Jacques Lacan had on him.)

Althusser, L. (1994) *Sur la philosophie (On Philosophy)*, Paris: Gallimard.(An interview and correspondence with Latin American admirers, dating from the mid-1980s, and chiefly notable as an instance of the late Althusser’s stress on ‘the true materialist tradition’ (Macchiavelli, Spinoza and Rousseau) that antedated Marx.)

Althusser, L. (1994, 1995) *Écrits philosophiques et politiques*, Paris: Stock/IMEC, 2 vols. (Volume 1 contains Althusser’s previously unpublished early Hegelian writings; Volume 2 offers a wealth of hitherto unpublished texts from Althusser’s middle and late periods.)

References and further reading

Anderson, P. (1980) *Arguments within English Marxism*, London: Verso.(A measured assessment of Althusser’s contribution to Marxism.)

Benton, T. (1984) *The Rise and Fall of Structural Marxism*, London: Macmillan.(A sympathetic critical assessment of Althusser and some of those influenced by him.)

Callinicos, A. (1976) *Althusser’s Marxism*, London: Pluto.(Concentrates on the tensions internal to Althusser’s philosophical project.)

Elliott, G. (1987) *Althusser - the Detour of Theory*, London: Verso.(A meticulous historical reconstruction of Althusser’s political and intellectual thought, with the best available bibliography of his writings.)

Elliott, G. (ed.) (1994) *Althusser: A Critical Reader*, Oxford: Blackwell.(Includes contemporary critical reactions to Althusser by Eric Hobsbawm, Axel Honneth, Paul Ricoeur and Pierre Vilar, and a discussion by the editor of the autobiographical writings.)

Geras, N. (1972) ‘Althusser’s Marxism’, *New Left Review* 71: 57-86.(A careful and probing critique, written from the standpoint of classical Marxism.)

Glucksmann, A. (1972) ‘A Ventriloquist Structuralism’, *New Left Review* 72: 68-92. (Virtuoso exposition of the conceptual flaws in *For Marx* and *Reading Capital*.)

Lazarus, S. (ed.) (1993) *Politique et philosophie dans l’oeuvre de Louis Althusser (Politics and Philosophy in the Work of Louis Althusser)*, Paris: Presses Universitaires de France.(Collection of essays, many by some of Althusser’s leading pupils.)

Moulier Boutang, Y. (1992) *Louis Althusser: Une biographie: la formation du mythe (1918-1956), tome I*, Paris: Bernard Grasset.(Indispensable corrective to Althusser’s myth-making in his autobiography; of great value on Althusser’s early political and philosophical development.)

Sprinker, M. (ed.) (1993) *The Althusserian Legacy*, London: Verso.(Collection of essays perhaps most notable for a lengthy interview with Derrida on Althusser and Marxism, and for the speech he delivered at Althusser’s graveside.)

Thompson, E.P. (1978) *The Poverty of Theory and Other Essays*, London: Merlin.(A great historian’s brilliantly written damnation of Althusser and all his works.)

al-Tusi, Khwajah Nasir (1201-74)

While philosophical activity in the Islamic west virtually ceased after Ibn Rushd at the close of the sixth century AH (twelfth century AD), it experienced renewed vigour in the east through the intellectual efforts and political involvement of Nasir al-Din al-Tusi. Although primarily a reviver of the peripatetic tradition of Ibn Sina, he was also possibly influenced by the ideas of al-Suhrawardi. He defended Ibn Sina from the criticisms levelled against him from the direction of theology, notably by Fakhr al-Din al-Razi, made a significant contribution to the acceptance of metaphysical argumentation and terminology in Twelver Shi'i theology, brought the ethical tradition of Ibn Miskawayh and the philosophers into the centre of Islamic ethical discourse, and had a lasting effect on the study of the exact sciences in Islam through both his original contributions to mathematics and astronomy and the observatory at Maraghah which the Mongol Khan Hülegü established for him.

1 Intellectual development

Al-Tusi's intellectual development cannot be divorced from the drama of his own life, and the catastrophe of the Mongol invasion of the Islamic east. From his birth in Tus in Khurasan in northwest Iran in AH 597/AD 1201 up to his middle or late twenties, al-Tusi lived in a Twelver Shi'i milieu, in a family whose idea of learning was, according to his own account, the study of the religious law, and whose behaviour was measured by its practice. His jurist father, however, was sufficiently broad-minded to encourage him beyond scholastic studies to the philosophical and natural sciences, and to acquaint himself with the doctrines of other schools and sects. To study philosophy, al-Tusi went to nearby Nishapur, where he was taught by a scholar whose teaching lineage went back to Ibn Sina. Early in his career, as al-Tusi himself later wrote, he was not convinced that the intellect could answer the ultimate metaphysical questions, since it would thereby be inquiring into its own origins, something of which it would be incapable. Perhaps as a way out of his perplexity, and quite possibly as the result of sectarian connections through an uncle, he turned to the Isma'ilism of his day, which had been influenced by the Neoplatonic speculations of Isma'ili thinkers in the third and fourth centuries AH (ninth and tenth centuries AD). Isma'ili doctrine turned on the concept of an infallible Imam, without whose guidance, it claimed, the unaided intellect was unable to reach the truth.

From his late twenties or early thirties, al-Tusi was in the service of the local Isma'ili leaders of northern Iran, writing a number of theological and philosophical works for them in both Persian and Arabic and beginning his contribution to a major revival of Peripatetic philosophy in eastern Islamic lands. With the Mongol invasion of Iran in the middle of the thirteenth century the Isma'ili strongholds were destroyed, and al-Tusi found himself involved in the negotiations leading to the surrender of the Grand Isma'ili Master to the invaders. His efforts were appreciated by the Mongol conqueror Hülegü, who took him on as an advisor, in which capacity he assisted at the sacking of Baghdad in AH 656/AD 1258. Later al-Tusi was put in charge of religious endowments and affairs. Hülegü also had the great observatory and library at Maraghah built for al-Tusi, where he led a team of scientists and mathematicians from as far away as China. It is clear that immense resources were put at his disposal for this project, where the teaching and study of philosophy went on hand in hand with that of the exact sciences.

The end of his Isma'ili period also marked al-Tusi's turn (or return) to Twelver Shi'ism, and the last period of his life witnessed not only a remarkable output of scientific works but also a reformulation of Imami theology in philosophical terms which was as influential in the Shi'i world as was that of Fakhr al-Din al-Razi in the Sunni. Al-Tusi died in AH 672/AD 1274 in Baghdad, in the same year as Thomas Aquinas. He was buried according to his last wishes beside the shrine of the seventh Twelver Imam, Musa ibn Ja'far, in Kazimayn just outside Baghdad. Among his most remarkable students were the philosopher Qutb al-Din al-Shirazi (d. AH 710/AD 1310) and the Imami jurist and theologian, the 'Allamah al-Hilli (d. AH 726/AD 1325).

If in his early life al-Tusi believed in the need for reason to be sustained by a non-rational (or supra-rational) guarantor, his move to Twelver Shi'ism, with its doctrine of the hidden, inaccessible Imam, indicates a growing strength in his convictions about the ability of the intellect. His sectarian shifts have given rise to much argument concerning his genuine doctrinal loyalties, but throughout his life there runs a consistent philosophical thread whose main characteristic was the defence, rehabilitation and elaboration of Ibn Sina's method and theories. It is through his interpretations in texts, epitomes, commentaries and refutations that subsequent generations in the

Islamic east have approached their understanding of Ibn Sina. His output as an author in both Arabic and Persian was prodigious, including lasting contributions to logic, metaphysics, ethics, mathematics and astronomy.

2 Logic, metaphysics and theology

Al-Tusi's main contribution to logic is contained in his Persian *Asas al-iqtibas (The Ground for the Acquisition of Knowledge)*, written during his Isma'ili period. It is divided according to the habitual ten sections of the organon of the Islamic world, and its discussion of substance has recently attracted attention (Morewedge 1975). It also stands as a testimony to al-Tusi's ability to write about technical subjects in Persian, by incorporating Arabic terminology into a fluent and graceful style. An Arabic manual of logic by al-Tusi, the *Tajrid al-mantiq (Abstract of Logic)*, was commented on by his pupil the 'Allamah al-Hilli.

However, the major text on which al-Tusi's reputation as an interpreter of Ibn Sina's philosophy rests is his commentary on the *Kitab al-Isharat (Remarks and Admonitions)*, written towards the end of his stay with the Isma'ilis. Al-Tusi's work was in part, at least, intended as a response to the commentary of Fakhr al-Din al-Razi on the same text. This work shows one of al-Tusi's continuing complaints, that those who attack Ibn Sina are generally ill-equipped as philosophers and that analysis can reveal their weaknesses. He is not averse to the use of polemic himself, but his insistence that a philosopher be evaluated in terms of the soundness of his argumentation and not according to preconceived ideas about conclusions demonstrates his faithfulness to Ibn Sina's own method. Thus, although on the whole defending Ibn Sina's theories, he disagrees with the latter when he sees fit. In his interpretation of emanation, for example, al-Tusi is partly in agreement with al-Suhrawardi concerning the nature of God's knowledge, although he does not entirely reject Ibn Sina's theory of identity of form between knower and known as being applicable to other existents (see Ibn Sina). God's knowledge of the First Intellect, and consequently of the entire universe, is identical with its existence through a kind of presential knowledge, while lower entities in the emanative chain derive their knowledge through forms and representations as well as through presence. Al-Tusi's commentary is clear and systematic, and is still studied with the original text because of his lucid explanations of Ibn Sina's often difficult and dense prose.

The texts discussed so far belong to al-Tusi's Isma'ili period, but the break with this past was decisively accomplished in *Masari' al-musari' (The Floorings of the Wrestler)*, a refutation of an Isma'ili Neoplatonic text by the crypto-Isma'ili al-Shahrastani (d. AH 548/AD 1153) which attacked Ibn Sina for deviating from 'prophetic theology'. Al-Tusi's vehemently anti-Isma'ili defence of Ibn Sina is unreservedly polemical, using the same tactic of accusations of weak logic and feeble-mindedness which he had employed against al-Razi. Perhaps also a rejection of al-Tusi's own past, this text marks another stage in the development of his conviction of the superiority of philosophical thinking over religious dialectics.

Al-Tusi made several contributions to the field of metaphysical theology. The first attempt in this direction was an exposition of Isma'ili *qiyama* (resurrection) theology in his *Rawdat al-taslim (The Garden of Submission)*, but of more enduring consequence was his later Twelver work, the *Tajrid al-kalam (Abstract of Theology)*. This has been the subject of numerous commentaries down to the present century, the most important of which is the 'Allamah al-Hilli's *Kashf al-murad (Disclosing the Intention)*. After the *Tajrid*, practically all Imami theological works would be expressed in the terminology of metaphysics, with Mulla Sadra eventually achieving a comprehensive and lasting fusion.

3 Ethics, mathematics and the natural sciences

There are two main works in al-Tusi's ethical output, the *Akhlaq-i Muhtashami (Muhtashamean Ethics)* and the *Akhlaq-i Nasiri (The Nasirean Ethics)*, both written in Persian. The first of these was commissioned by the Isma'ili ruler (*muhtasham*) of Quhistan, Nasir al-Din 'Abd al-Rahman, who provided the outline and approved its contents but called in al-Tusi to do the major work because of the demands of his own political duties. This is scarcely more than a manual of ethical precepts, amply illustrated with quotations from the Qur'an, the Shi'i Imams and Greek sources. The *Akhlaq-i Nasiri*, the first 'edition' of which was dedicated to the same Nasir al-Din, is arranged as a work of philosophical ethics. Its divisions into three parts - ethics (*akhlaq*), domestic economics (*tadbir-e manzil*), politics (*siyasat-e mudun*) - set the pattern for subsequent works on practical philosophy in the Islamic tradition (see al-Dawani).

The first part on ethics is modelled on Ibn Miskawayh's *Tahdhib al-akhlaq (Cultivation of Morals)*, of which the work was initially commissioned to be merely a Persian translation (see Ibn Miskawayh). However, al-Tusi expands on Ibn Miskawayh both in the initial section on principles, mainly a theoretical treatment of psychology (the soul), and in his subsequent treatment of character and the virtues. This first part finishes with the addition of a section on the treatment of the vices as sicknesses of the soul, and of the cures to remedy them. The sources of the second part on domestic economics are the Arabic translation of Bryson's *Oikonomikos* and a text by Ibn Sina, his *Kitab al-siyasa (Book of Politics)*, while the third part, on politics, goes back to al-Farabi's *Kitab al-siyasa al-madaniyya (The Political Regime)* and *Fusul al-madani (Aphorisms of the Statesman)* (see al-Farabi). The last part contains an important section on the virtue of love (*mahabba*) as the cement of societies.

After al-Tusi joined Hülegü, he changed the introduction and conclusion to this work, excusing his previous praise of the Isma'ili leadership as the product of exigency. Moreover, he added that this was strictly a work of philosophy which transcended sectarian differences and was available to all. The work made available to Persian readers the Islamic ethical tradition taken from Greek philosophy but now incorporating Qur'anic material alongside the opinions of Plato and Aristotle. Justice explicitly comes to the fore as the supreme virtue running through all three parts of the book, implicitly linking it with Shi'i theology and the priority given in the latter to justice among the divine attributes; and philosophical ethics and the religious law are stated to be concerned with the same subject matter, thus affirming the intellect's capacity to view normative values in a way which could only have been acceptable at that time within Shi'i circles (see Ethics in Islamic philosophy).

Throughout his life al-Tusi was a prolific writer in mathematics and the natural sciences, and made advances in trigonometry, mathematics and astronomy. This aspect of his intellectual endeavour was eventually rewarded with the foundation of the Maraghah observatory. The result of the astronomical observations and calculations made there was the famous tables of the *Zij-e Ilkhani* (in Persian, but also translated into Arabic). Prior to Maraghah, the rational sciences had been cultivated by individuals with (or without) private patronage, the schools in Islam being devoted almost entirely to the law and dismissive of, if not actually hostile to philosophical activity. The setting up of the observatory and the institutionalization of the rational sciences created a demand for teaching materials, and al-Tusi was himself the author of a number of recensions (*tahrir*) of scientific texts as well as summaries and abridgements of theological, logical, and philosophical texts, clearly intended to supply this teaching need. Al-Tusi's lasting influence can be seen in the subsequent surge of activity in the rational sciences in the Islamic east, as well as in their gradual absorption into religious education, which in turn affected the development of theology, particularly among Shi'i scholars (see Science in Islamic philosophy).

See also: Ibn Sina; al-Razi, Fakhr al-Din; Science in Islamic philosophy

JOHN COOPER

List of works

(1235, 1265) *Akhlaq-i Nasiri (The Nasirean Ethics)*, trans. G.M. Wickens, London: George Allen & Unwin, 1964. (An excellent, meticulous translation of the *Akhlaq-e Nasiri*, with a brief introduction and notes.)

(1242) *Rawdat al-taslim (The Garden of Submission)*, trans. C. Jambet, *La convocation d'Alamût: somme de philosophie ismaélienne (Radat al-taslîm: Le jardin de vraie foi)*, Lagrasse and Paris: Éditions Verdier and Éditions UNESCO, 1996.(A work of Isma'li theology.)

(1244-5) *Asas al-iqtibas (The Ground for the Acquisition of Knowledge)*, ed. M. Radawi, Tehran: Tehran University Press, 1947.(Al-Tusi's major logical text, in Persian.)

(probably after 1246) *Sayr wa suluk (Contemplation and Action)*, ed. and trans. S.J.H. Badakhchani, London: Institute for Ismaili Studies, 1997.(This is the autobiography which al-Tusi wrote during his stay with the Isma'ilis, and is untinged with the complaints which he later made of this period in his life.)

(before 1258) *Sharh al-Isharat (Commentary on the Isharat)*, ed. S. Dunya in Ibn Sina, *al-Isharat wa-'l-tanbihat*, Cairo: Dar al-Ma'arif, 1957-60, 4 parts, 3 vols in 2; also in Ibn Sina, *al-Isharat wa-'l-tanbihat*, Tehran: Matba'at al-Haydari, 1957-9, 3 vols.(Both these editions contain al-Tusi's commentary as parts of Fakhr al-Din al-Razi's commentary, to which al-Tusi is responding. The Tehran edition also contains Qutb al-Din al-Razi's commentary, which sets out to adjudicate between al-Tusi and al-Razi.)

al-Tusi (probably before 1270-1) *Tajrid al-kalam (Abstract of Theology)*.(The text of this, al-Tusi's major theological work (also known as *Tajrid al-'aqa'id* and *Tajrid al-i'tiqad*), can be found in the commentary by his pupil Hasan ibn Yusuf ibn al-Mutahhar al-Hilli, *Kashf al-murad fi sharh tajrid al-i'tiqad*, Qum: Jama'at al-Mudarrisin, no date.)

References and further reading

Dabashi, H. (1995) ‘Khwajah Nasir al-Din al-Tusi: The Philosopher/Vizier and the Intellectual Climate of His Times’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge: 527-84.(Good introduction to al-Tusi, his work and his times.)

Ibn Sina (980-1037) *Kitab al-isharat wa-’l-tanbihat (Remarks and Admonitions)*, trans. A.-M. Goichon, *Livre des directives et remarques*, Beirut: Commission Internationale pour la Traduction des Chefs d’Oeuvres, and Paris: Vrin, 1951.(A useful French translation of Ibn Sina’s text with introduction and notes by the translator; also contains many of al-Tusi’s explanations in the notes, as well as some of his criticisms of al-Razi.)

Madelung, W. (1985) ‘Nasir al-Din Tusi’s Ethics Between Philosophy, Shi‘ism, and Sufism’, in R.G. Hovannisian (ed.) *Ethics in Islam*, Malibu, CA: Undena, 85-101.(Discusses the developments of al-Tusi’s ethical ideas in relation to his political involvement and religious allegiances.)

Morewedge, P. (1975) ‘The Analysis of "Substance" in Tusi’s *Logic* and in the Ibn Sinian tradition’, in G. Hourani (ed.) *Essays on Islamic philosophy and science*, Albany, NY: State University of New York Press. (Morewedge is one of the few scholars to have made specific studies of al-Tusi’s metaphysics, and this article discusses substance in the *Asas al-iqtibas*.)

Mudarris Radawi, M.T. (1975) *Ahwal wa athar-e… Abu Ja‘far Muhammad… al-Tusi… (Life and Works of… Abu Ja‘far Muhammad… al-Tusi…)*, Tehran: Bunyad-e Farhang-e Iran.(In the absence of any comprehensive work on al-Tusi in a European language, it is necessary to turn to this Persian work which contains a comprehensive inventory of al-Tusi’s works.)

Ambedkar, Bhimrao Ramji (1891-1956)

Bhimrao Ramji Ambedkar was a statesman, scholar, human rights advocate, educator, barrister, first law minister of the Republic of India and architect of its constitution. Born into the untouchable Mahar subcaste, he became the widely revered leader of India's more than 100 million hereditary outcastes, the social and psychological emancipation of whom remained his lifelong objective. Strongly influenced by Anglo-American liberalism and pragmatism, Ambedkar was a staunch constitutionalist and social democrat. Locating the source of untouchability within the caste system itself, he became a militant critic of Hinduism, eventually affirming Buddhism as the universal ethical teaching that he felt could lead all of India into modernity.

Sponsored by a socially progressive Indian mahārāja, Ambedkar became the first of his outcaste background to complete a full course of postgraduate study, receiving a Ph.D. from Columbia University and a D.Sc. from the University of London in addition to qualifying for the bar from Gray's Inn, London.

Ambedkar drew on both Indian and Western sources to formulate his social philosophy. Early in life he identified the Buddha, Kabīr and JyotibāPhule as his principal preceptors: the Buddha (sixth-fifth century BC) because of his programme of ethical and spiritual cultivation including an explicit critique of the caste system; the radical poet-saint and mystic Kabīr (fifteenth century) for his anti-Brahmanic, egalitarian spirituality; and the militant educationalist Phule (1828-90) as the Shudra founder of the ninteenth-century non-Brahman movement in Maharashtra that first coupled Western rationalism and liberalism with more traditional social and spiritual critiques of casteism and untouchability. Ambedkar was widely read in the modern classics of Western social and political theory and he was especially influenced by the faculty at Columbia University during his graduate period there from 1913-16. Through his contact with John Dewey, James Shotwell, Edwin Seligman, James Harvey Robinson, Franklin Giddings and Alexander Goldenweiser in particular, he acquired a pragmatic, even optimistic, conviction in the potential of democratic institutions to bring about social equality. Given Ambedkar's consequent commitment to constitutional democracy, it was inevitable that he would clash with Gandhi, rejecting both the latter's willingness to move outside the bounds of the legal process and also his patriarchal belief that emancipation of the Untouchables would result from a change of heart on the part of caste Hindus rather than from the political and ethical transformation of the Untouchables themselves (see Political philosophy, Indian).

A jurist and statesman more than a political philosopher, Ambedkar's chief intellectual contributions lay in the realm of social theory. His MA thesis revised prevailing sociological views regarding the origins of caste in India. It argued that caste discrimination arose not from race, colour, or occupation, but rather from the Brahmanic concept of endogamy-based ritual purity and pollution, a worldview subsequently adopted by other caste communities in imitation of the Brahmans. Later in his life he expanded this position with the provocative thesis linking the development of untouchability to Brahmanic oppression of Buddhist minorities beginning in the fifth century.

In assessing Ambedkar's social philosophy and his career as a statesman one must recognize three distinct but interlinked spheres of concern. He felt that India would not be fully liberated until its outcastes were emancipated and he further asserted that this emancipation must be legal, material and spiritual. Each of these three dimensions became the focus of successive, overlapping phases of his activities. His first object was to secure the legal status of equal rights for Untouchables and the illegality of traditional practices of caste discrimination that restricted access to water sources and other public facilities. This phase began with his early efforts in the 1920s and 1930s to organize *satyagraha* demonstrations and legal suits and culminated with his decisive role in bringing the new Indian constitution to ratification in 1949.

As an economist and sociologist he recognized that legal protection alone was not sufficient. Thus, his second sphere of concern focused on the material improvement of the Untouchables. Efforts to establish legally mandated reservations guaranteeing Untouchables access to the political and educational systems, the civil service and many public sector professions were crucial to this phase of Ambedkar's activity, but at its heart lay his recognition that change would come only when the Untouchables were able to take full advantage of these opportunities. This realization led him in 1945 to establish the People's Education Society. Its success has led to the introduction of over thirty institutions of higher education open to all castes, but was especially intended to help the

ex-Untouchables gain the education necessary for social advancement. In retrospect Ambedkar appears more successful as an educator than a politician: a fact evidenced by the recent emergence of a strong ex-Untouchable urban middle class and the flourishing of a vigorous and nationally influential *Dalit* (oppressed) literature movement in Marathi, Gujarati and other Indian languages.

The third concern that Ambedkar thought should be addressed went beyond his efforts to secure legal and material advancement for his people. He believed that educational economic advancement would be meaningless unless it was coupled with ethical and spiritual development. It was in Buddhism that he found an indigenous Indian source for completing his vision of emancipation, a source he felt was compatible with his Western rationalism and egalitarianism. The importance of this third element in his thought has been minimized by the more politically minded of his successors, some of whom feel that renouncing Hinduism is sufficient to break the stigma of caste. This was not Ambedkar's view. Attracted to the Buddha's compassionate injunction to work for the welfare of the many as early as 1908, Ambedkar swore in 1935 that although he had been born a Hindu he would not die a Hindu. By 1950 he had come to believe that it was only through Buddhism that the Untouchables would gain the self-respect and self-reliance necessary to realize fully their own advancement. In 1956 just weeks before his untimely death, he inaugurated a mass conversion movement that has subsequently attracted more than 10 million new Indian Buddhists.

See also: Political philosophy

ALAN SPONBERG

List of works

Ambedkar, B.R. (1919-57) *Dr. Babasaheb Ambedkar Writings and Speeches*, ed. V. Moon, Bombay: Education Department, Government of Maharashtra, 1979, 13 vols. (The most complete collection of Ambedkar's works available, including the three important works listed separately below.)

Ambedkar, B.R. (1945) *What Congress and Gandhi Have Done to the Untouchables*, Bombay: Thacker, 2nd edn, 1946.(Ambedkar's critique of the Gandhian solution to untouchability as both insufficient in scope and paternalistic in attitude.)

Ambedkar, B.R. (1948) *The Untouchables*, New Delhi: Amrit Book Co.(A historical and sociological monograph developing Ambedkar's thesis that the Untouchables were the remnants of India's original Buddhist population, the broken men oppressed and marginalized after a resurgence of Brahmanic Hinduism beginning in the fifth century.)

Ambedkar, B.R. (1957) *The Buddha and His Dhamma*, Bombay: People's Education Society. (Ambedkar's critical assessment of the Buddha's teaching, considered especially with regard to the contemporary needs of the ex-Untouchables.)

References and further reading

Keer, D. (1954) *Dr Ambedkar: Life and Mission*, Bombay: Popular Prakashan; revised 3rd edn, 1971.(The most popular and complete biography of Ambedkar.)

Sangharakshita (1986) *Ambedkar and Buddhism*, Glasgow: Windhorse Publications. (The best critical monograph available on Ambedkar to date, written by an English-born Buddhist monk who has worked with the Ambedkarite Buddhist movement since 1952.)

Zelliot, E. (1992) *From Untouchable to Dalit: Essays on the Ambedkar Movement*, New Delhi: Manohar.(An excellent collection of essays comprehensively covering many aspects of Ambedkar and the ex-Untouchable Buddhist movement he initiated. The volume includes Zelliot's especially seminal essay on Gandhi and Ambedkar and also extensive bibliographic references.)

Ambiguity

A word, phrase or sentence is ambiguous if it has more than one meaning. The word 'light', for example, can mean not very heavy or not very dark. Words like 'light', 'note', 'bear' and 'over' are lexically *ambiguous. They induce ambiguity in phrases or sentences in which they occur, such as 'light suit' and 'The duchess can't bear children'. However, phrases and sentences can be ambiguous even if none of their constituents is. The phrase 'porcelain egg container' is* structurally *ambiguous, as is the sentence 'The police shot the rioters with guns'. Ambiguity can have both a lexical and a structural basis, as with sentences like 'I left her behind for you' and 'He saw her duck'.*

The notion of ambiguity has philosophical applications. For example, identifying an ambiguity can aid in solving a philosophical problem. Suppose one wonders how two people can have the same idea, say of a unicorn. This can seem puzzling until one distinguishes 'idea' in the sense of a particular psychological occurrence, a mental representation, from 'idea' in the sense of an abstract, shareable concept. On the other hand, gratuitous claims of ambiguity can make for overly simple solutions. Accordingly, the question arises of how genuine ambiguities can be distinguished from spurious ones. Part of the answer consists in identifying phenomena with which ambiguity may be confused, such as vagueness, unclarity, inexplicitness and indexicality.

1 Types of ambiguity

Ambiguity is a property of linguistic expressions. A word, phrase or sentence is ambiguous if it has more than one meaning. Obviously this definition does not say what meanings are or what it is for an expression to have one (or more than one). For a particular language, this information is provided by a grammar, which systematically pairs forms with meanings, ambiguous forms with more than one meaning (see Semantics).

There are two types of ambiguity, *lexical* and *structural*. Lexical ambiguity is by far the more common. Everyday examples include nouns like 'chip', 'pen' and 'suit', verbs like 'call', 'draw' and 'run' and adjectives like 'deep', 'dry' and 'hard'. There are various tests for lexical ambiguity. One test is having two unrelated antonyms, as with 'hard', which has both 'soft' and 'easy' as opposites. Another is the conjunction reduction test. Consider the sentence, 'The tailor pressed one suit in his shop and one in the municipal court'. Evidence that the word 'suit' (not to mention 'press') is ambiguous is provided by the anomaly of the 'crossed interpretation' of the sentence, on which 'suit' is used to refer to an article of clothing and 'one' to a legal action.

The above examples of ambiguity are each a case of one word with more than one meaning. However, it is not always clear when we have only one word. The verb 'desert' and the noun 'dessert', which sound the same but are spelled differently, count as distinct words (they are homonyms). So do the noun 'bear' and the verb 'bear', even though they not only sound the same but are spelled the same. These examples may be clear cases of homonymy, but what about the noun 'respect' and the verb 'respect' or the preposition 'over' and the adjective 'over'? Are the members of these pairs homonyms or different forms of the same word? There is no general consensus on how to draw the line between cases of one ambiguous word and cases of two homonymous words. Perhaps the difference is ultimately arbitrary.

Sometimes one meaning of a word is derived from another. For example, the cognitive sense of 'see' (to see that something is so) seems derived from its visual sense. The sense of 'weigh' in 'He weighed the package' is derived from its sense in 'The package weighed two pounds'. Similarly, the transitive senses of 'burn', 'fly' and 'walk' are derived from their intransitive senses. Now it could be argued that in each of these cases the derived sense does not really qualify as a second meaning of the word but is actually the result of a lexical operation on the underived sense. This argument is plausible to the extent that the phenomenon is systematic and general, rather than peculiar to particular words. Lexical semantics has the task of identifying and characterizing such systematic phenomena. It is also concerned to explain the rich and subtle semantic behaviour of common and highly flexible words like the verbs 'do' and 'put' and the prepositions 'at', 'in' and 'to'. Each of these words has uses which are so numerous yet so closely related that they are often described as 'polysemous' rather than ambiguous.

Structural ambiguity occurs when a phrase or sentence has more than one underlying structure, such as the phrases 'Tibetan history teacher', 'a student of high moral principles' and 'short men and women', and the sentences 'The

girl hit the boy with a book' and 'Visiting relatives can be boring'. These ambiguities are said to be structural because each such phrase can be represented in two structurally different ways, for example '[Tibetan history] teacher' and 'Tibetan [history teacher]'. Indeed, the existence of such ambiguities provides strong evidence for a level of underlying syntactic structure (see Syntax). Consider the structurally ambiguous sentence, 'The chicken is ready to eat', which could be used to describe either a hungry chicken or a cooked chicken. It is arguable that the operative reading depends on whether or not the implicit subject of the infinitive clause 'to eat' is tied anaphorically to the subject ('the chicken') of the main clause.

It is not always clear when we have a case of structural ambiguity. Consider the elliptical sentence, 'Perot knows a richer man than Trump'. It has two meanings, that Perot knows a man who is richer than Trump and that Perot knows a man who is richer than any man Trump knows, and is therefore ambiguous. But what about the sentence 'John loves his mother and so does Bill'? It can be used to say either that John loves John's mother and Bill loves Bill's mother or that John loves John's mother and Bill loves John's mother. But is it really ambiguous? One might argue that the clause 'so does Bill' is unambiguous and may be read unequivocally as saying in the context that Bill does the same thing that John does, and although there are two different possibilities for what counts as doing the same thing, these alternatives are not fixed semantically. Hence the ambiguity is merely apparent and better described as semantic underdetermination.

Although ambiguity is fundamentally a property of linguistic expressions, people are also said to be ambiguous on occasion in how they use language. This can occur if, even when their words are unambiguous, their words do not make what they mean uniquely determinable. Strictly speaking, however, ambiguity is a semantic phenomenon, involving linguistic meaning rather than speaker meaning (see Meaning and communication). Generally when one uses ambiguous words or sentences, one does not consciously entertain their unintended meanings, although there is psycholinguistic evidence that when one hears ambiguous words one momentarily accesses and then rules out their irrelevant senses. When people use ambiguous language, generally its ambiguity is not intended. Occasionally, however, ambiguity is deliberate, as with an utterance of 'I'd like to see more of you' when intended to be taken in more than one way in the very same context of utterance.

2 Ambiguity contrasted

It is a platitude that what your words convey 'depends on what you mean'. This suggests that one can mean different things by what one says, but it says nothing about the variety of ways in which this is possible. Semantic ambiguity is one such way, but there are others: homonymy (mentioned in §1), vagueness, relativity, indexicality, nonliterality, indirection and inexplicitness. All these other phenomena illustrate something distinct from multiplicity of linguistic meaning.

An expression is vague if it admits of borderline cases (see Vagueness). Terms like 'bald', 'heavy' and 'old' are obvious examples, and their vagueness is explained by the fact that they apply to items on fuzzy regions of a scale. Terms that express cluster concepts, like 'intelligent', 'athletic' and 'just', are vague because their instances are determined by the application of several criteria, no one of which is decisive.

Relativity is illustrated by the words 'heavy' and 'old' (these are vague as well). Heavy people are lighter than nonheavy elephants, and old cats can be younger than some young people. A different sort of relativity occurs with sentences like 'Jane is finished' and 'John will be late'. Obviously one cannot be finished or late *simpliciter* but only finished with something or late for something. This does not show that the words 'finished' and 'late' are ambiguous (if they were, they would be ambiguous in as many ways as there are things one can be finished with or things one can be late for), but only that such a sentence is semantically underdeterminate - it must be used to mean more than what the sentence means.

Indexical terms, like 'you', 'here' and 'tomorrow', have fixed meaning but variable reference. For example, the meaning of the word 'tomorrow' does not change from one day to the next, though of course its reference does (see Demonstratives and indexicals).

Nonliterality, indirection and inexplicitness are further ways in which what a speaker means is not uniquely determined by what their words mean (see Speech acts §4). They can give rise to unclarity in communication, as might happen with utterances of 'You're the icing on my cake', 'I wish you could sing longer and louder', and 'Nothing is on television tonight'. These are not cases of linguistic ambiguity but can be confused with it because

speakers are often said to be ambiguous.

3 Philosophical relevance

Philosophical distinctions can be obscured by unnoticed ambiguities. So it is important to identify terms that do double duty. For example, there is a kind of ambiguity, often described as the ‘act/object’ or the ‘process/product’ ambiguity, exhibited by everyday terms like ‘building’, ‘shot’ and ‘writing’. Confusions in philosophy of language and mind can result from overlooking this ambiguity in terms like ‘inference’, ‘statement’ and ‘thought’. Another common philosophical ambiguity is the type/token distinction. Everyday terms like ‘animal’, ‘book’ and ‘car’ apply both to types and to instances (tokens) of those types. The same is true of linguistic terms like ‘sentence’, ‘word’ and ‘letter’ and of philosophically important terms like ‘concept’, ‘event’ and ‘mental state’ (see Type/token distinction).

Although unnoticed ambiguities can create philosophical problems, ambiguity is philosophically important also because philosophers often make spurious claims of it. Indeed, the linguist Charles Ruhl (1989) has argued that certain ostensible ambiguities, including act/object and type/token, are really cases of lexical underdetermination. Kripke (1977) laments the common stratagem, which he calls ‘the lazy man’s approach in philosophy’, of appealing to ambiguity to escape from a philosophical quandary, and Grice (1967) urges philosophers to hone ‘Modified Occam’s Razor: senses are not to be multiplied beyond necessity’. He illustrates its value by shaving a sense off the logical connective ‘or’, often thought to have both an inclusive and exclusive sense. Grice argues that, given its inclusive meaning, its exclusive use can be explained entirely on pragmatic grounds (see Implicature §6; Pragmatics §12). Another example, prominent in modern philosophy of language, is the ambiguity alleged to arise from the distinction between referential and attributive uses of definite descriptions (see Descriptions §5). Less prominent but not uncommon is the suggestion that pronouns are ambiguous as between their anaphoric and deictic use. So, for example, it is suggested that a sentence like ‘Oedipus loves his mother’ has two ‘readings’ - that is, it is ambiguous - because it can be used to mean either that Oedipus loves his own mother or that Oedipus loves the mother of some contextually specified male. However, this seems to be an insufficient basis for the claim of ambiguity. After all, being previously mentioned is just another way of being contextually specified. Accordingly, there is nothing semantically special in this example about the use of ‘his’ to refer to Oedipus.

Claims of structural ambiguity can also be controversial. Of particular importance are claims of scope ambiguity, which are commonly made but rarely defended (see Scope). A sentence like ‘Everybody loves somebody’ is said to exhibit a scope ambiguity because it can be used to mean either that for each person, there is somebody that that person loves or (however unlikely) that there is somebody that everybody loves. These uses may be represented, respectively, by the logical formulas ‘ $(\forall x)(\exists y)(\mathrm{L}xy)$’ and ‘ $(\exists y)(\forall x)(\mathrm{L}xy)$’. It is generally assumed that, because different logical formulas are needed to represent the different ways in which an utterance of such a sentence can be taken, the sentence itself has two distinct logical forms (see Logical form). Sustaining this claim of ambiguity requires identifying a level of linguistic description at which the sentence can be assigned two distinct structures. Some grammarians have posited a level of LF, corresponding to what philosophers call logical form, at which relative scope of quantified noun phrases may be represented. However, LF of this kind does not explain scope ambiguities that philosophers attribute to sentences containing modal operators and psychological verbs, such as ‘The next president might be a woman’ and ‘Ralph wants a sloop’. An utterance of such a sentence can be taken in either of two ways, but it is arguable that the sentence is not ambiguous but merely semantically underdeterminate with respect to its two alleged ‘readings’.

Notwithstanding the frequency in philosophy of unwarranted and often arbitrary claims of ambiguity, it cannot be denied that some terms really are ambiguous. The nouns ‘bank’ and ‘suit’ are clear examples and so are the verbs ‘bank’ and ‘file’. Philosophers sometimes lament the prevalence of ambiguity in natural languages and yearn for an ideal language in which it is absent. But ambiguity is a fact of linguistic life. Despite the potentially endless supply of words, many words do double duty or more. And despite the unlimited number of sentences, many have several meanings and their utterance must be disambiguated in the light of the speaker’s likely intentions.

See also: Language, philosophy of; Semantics

KENT BACH

References and further reading

Atlas, J.D. (1989) *Philosophy Without Ambiguity: A Logico-Linguistic Essay*, Oxford: Oxford University Press. (Examines ambiguity tests and questions certain philosophical appeals to ambiguity.)

Bach, K. (1994) ‘Conversational Impliciture’, *Mind and Language* 9: 124-62. (Identifies ways, distinct from those identified by Grice, in which linguistic meaning can underdetermine speaker meaning and, in particular, contrasts semantic underdetermination with ambiguity.)

Cruse, D.A. (1986) *Lexical Semantics*, Cambridge: Cambridge University Press. (Chapter 3 discusses linguistic features of ambiguity and examines tests for it.)

Grice, H.P. (1967) ‘Logic and Conversation’, *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, 1989.(Chapter 2 of this collection explains how multiple uses of an expression can often be explained without appealing to ambiguity.)

Kripke, S.A. (1977) ‘Speaker’s Reference and Semantic Reference’, *Midwest Studies in Philosophy* 2: 255-76.(A case study, focusing on the distinction between referential and attributive uses of definite descriptions, in how to expose a gratuitous philosophical claim of ambiguity.)

May, R. (1985) *Logical Form: Its Structure and Derivation*, Cambridge, MA: MIT Press.(Investigates the hypothesis within generative grammar that there is a level of representation, Logical Form.)

Ruhl, C. (1989) *On Monosemy: A Study in Linguistic Semantics*, Albany, NY: State University of New York Press.(Argues for the presumption that a word has a single, though highly abstract, meaning rather than the multiplicity of meanings commonly attributed, as by the list of definitions in its dictionary entry.)

Zwicky, A. and Sadock, J. (1975) ‘Ambiguity Tests and How to Fail Them’, in J. Kimball (ed.) *Syntax and Semantics*, vol. 4, New York: Academic Press.(Presents and assesses various linguistic tests of ambiguity.)

American philosophy in the 18th and 19th centuries

Jonathan Edwards, the first great American philosopher, interpreted Calvinist theology within the newer framework of Newtonian physics and Lockean empiricism in his Freedom of the Will (1754). However, he was all but forgotten by the end of the eighteenth century, when political rather than theological issues held centre stage. In the years leading up to the American Revolution, the moral sense theory of Shaftesbury and Hutcheson, Lockean liberalism and classical republican theory all contributed to the thought of Thomas Jefferson, James Madison and others who saw themselves as parties to a contract with a monarch, defenders of the rights of humans, and members of a new and virtuous republic.

In the early nineteenth century, Scottish common sense realism prevailed in the universities, but the most original and influential philosophical writing came from the communities of the transcendentalists. Emerson and Thoreau developed philosophies of life, language, knowledge and being in writings drawing on the Greek and Roman classics, English and German Romanticism, Christianity, and non-Western thought. After the Civil War (1861-5), a series of clubs in the East and Midwest, and the new Journal of Speculative Philosophy made Hegel more accessible to Americans; while in Cambridge, Massachusetts, the 'Metaphysical Club' of William James, Charles Peirce, Chauncey Wright and Oliver Wendell Holmes Jr became the birthplace of pragmatism.

The last quarter of the nineteenth century saw the professionalization of American philosophy: new graduate departments at Harvard and Johns Hopkins, professional journals, and state-supported universities in the Midwest building non-denominational departments of philosophy. By the end of the century, James had published his vast Principles of Psychology (1890) and enunciated a version of pragmatism; Peirce had produced an outpouring of writing on pragmatism, scientific method, logic, semiotics and metaphysics; and Josiah Royce and John Dewey were launched on influential academic careers.

1 Colonial America

The early Puritan communities were sustained by an intense and continuous involvement with abstract ideas. To the Harvard undergraduate studying Ramus' *Dialectica*, as to the townsman poring over Calvin's *Institutes*, questions concerning conversion and sanctification were understood to lie at the heart of New England covenant theology, and these in turn were inseparable from a set of problems in what we now call epistemology, ontology and ethics or moral psychology.

In the immediate background of New England Puritan divinity lay an unstable synthesis of medieval scholasticism and Calvinist theology, with what Calvin himself called the 'awful decree' of predestination at its centre. This synthesis would be exploded by Newton's *Principia* and Locke's *Essay Concerning the Human Understanding*, the two works together regarded in England's American colonies, as in England and Europe, as heralding the advent of a New Science and a new philosophical empiricism.

The great monument of the encounter between covenant theology and the new empiricism is Jonathan Edwards' *Freedom of the Will* (1754), deservedly famous for its apparently effortless reinterpretation of Calvinist doctrine within the newer framework of Newtonian physics - especially the new atomic or 'corpuscular' theory of matter - and Lockean sensationalism. This was the first significant work of philosophy produced in America, and the first American work in any category to have an important influence on European thought. Yet, although certain elements of his metaphysics were absorbed into Concord Transcendentalism, Edwards' influence on American philosophy was otherwise virtually extinct by the end of the eighteenth century (see Edwards, J.).

After Edwards, 'abstract' discourse in America - that is, discourse concerned with ideas and principles - shifted from a theological to a political register, as it did also in Europe (for example, in Montesquieu's *Lettres Persanes* and Voltaire's *Lettres Philosophiques*). In European political theory, a new sense of cultural relativity or 'climate of contingency' is reflected in the contract theories of Hobbes, Locke and Rousseau: consent is the necessary basis for the polity, and the particular form of a society may vary according to cultural and historical contingencies. According to 'Lockean liberal' interpretations of the political philosophy of the Founders (Hartz, Boorstin), the Continental Congress applied the principles of Locke's *Two Treatises on Civil Government* to their own case in declaring their independence from Britain: the monarch had violated his contract with the people, so the

arrangement between them was dissolved.

The Federalist Papers (1787-8) of John Madison and Alexander Hamilton (with some assistance from John Jay) constitute an extraordinary intervention of philosophy in the historical process, as they were written for a New York newspaper in the period when the Constitution was being voted on in the state legislatures. In the background of this and other American political documents of the late eighteenth century lies a sea of European ideas and their American inflections, including not just Lockean liberalism, but the moral sense theory of Shaftesbury and Hutcheson, and classical republican theory (deriving ultimately from Aristotle and Polybius via Machiavelli, Montesquieu and the English 'Country party' of Bailyn, Pocock and Dowling). Classical republican theory, with its vocabulary of 'luxury' and 'corruption' as opposed to 'virtue', allowed the American colonists to think of themselves as returning to something like the 'virtuous' state of ancient Rome. Moral sense theory provided an idea central to many revolutionary documents: that all human beings are possessed of a 'moral sense' that discerns right from wrong in the same way as the ear hears a dissonance in music (see Moral sense theories).

It was in fact the transmutation of moral sense theory into the common sense epistemology of Thomas Reid, Dugald Stewart and Sir William Hamilton that became the dominant philosophy taught at American universities from the late eighteenth century and through much of the nineteenth century. These writers offered a defence of direct perception against the scepticism of Hume that, in the hands of such teachers as John Witherspoon of Princeton (appointed in 1766), or Levi Hedge, the first professor of philosophy at Harvard (1792), could be seen as 'deist' or 'Christian'. God created a material world, these writers held, which we by our ingenuity and careful observation can know and improve. Witherspoon, a conservative Scottish Presbyterian who became a signatory of the Declaration of Independence, taught that moral questions could be investigated scientifically, and argued against radical scepticism on the ground that we know our experiential errors by means of other experiences. His *Lectures on Moral Philosophy and Eloquence* (1800) became a standard college text. Scots common sense theory helped make empiricism and science orthodox within the universities, while in its realism and insistence that relations are perceived, it anticipated doctrines of Peirce and James (see Common-sense ethics; Common Sense School).

2 19th century philosophy

The most original and influential early nineteenth century philosophical writers arose not in the universities, however, but among the Concord Transcendentalists. This group included Amos Bronson Alcott (1799-1888), Frederick Henry Hedge (1805-90), George Ripley (1802-80), Ralph Waldo Emerson (1803-82), Margaret Fuller (1810-50) and Henry David Thoreau (1817-62). Among these, Emerson and Thoreau stand out for their power as writers, and for their influence on such subsequent philosophers as James, Dewey, Nietzsche, and Ghandi.

Emerson enjoyed a highly visible career as a lecturer and writer. His sources include the classical philosophy he studied at Harvard, English and German Romantic poetry and philosophy, Hinduism and other non-Western philosophies and, of course, Christianity. Emerson's first book, *Nature* calls for a new 'original relation to the universe' (Emerson 1836). His controversial *'Divinity School Address'* (1838) condemns the 'Monster' of historical Christianity and urges the divinity graduates to find their own original natures, without which they can offer nothing to others. One makes the most sense to others, Emerson holds, by diving deeply into one's own heart. Emerson's *First Series* (1841) and *Second Series* (1844) of essays offer striking aphorisms and powerful paragraphs advocating a life of 'self-reliance', expanding 'circles', deep-seeing 'intellect', and balanced 'experience'. *Representative Men* (1850) and *The Conduct of Life* (1860) are important later works.

Thoreau thought of philosophy as a practice: a life of 'simplicity, independence, magnanimity, and trust' (Thoreau 1854). *Walden* is a record of that practice, based on two years spent living in the woods near Concord, Massachusetts, and offers a series of reflections on nature and human life. Thoreau finds the mass of men and women living 'lives of quiet desperation', driving themselves like slaves. In *Walden*'s long opening chapter on Economy, Thoreau construes his life at Walden as an 'experiment' to show how little is really necessary for life and, by contrast, how needlessly complex most people's lives happen to be. Later chapters blend descriptions of Walden Pond with reflections on the peculiar power of literature - 'the work of art nearest to life itself' (Thoreau 1854), on reading, vegetarianism, spring, ice, living in the present and neighbourliness. Thoreau's other works include his essays 'Walking' (1862), and the influential 'Civil Disobedience' (1849).

After the Civil War (1861-5), two of the many philosophical clubs scattered throughout the East and Midwest played a special role in the development of American philosophy. The 'St. Louis Hegelians' were led by William Torrey Harris (1835-1909) and Hans Conrad Brokmeyer (1826-1906). Brokmeyer emigrated to the US from Prussia in 1844, practised law, and eventually became lieutenant governor of Missouri. A leader in the German community, he worked on a translation of Hegel's *Logic*, which circulated in manuscript. Harris, a native of Connecticut who left Yale in his junior year, taught school in St. Louis and eventually became United States Commissioner of Education. He studied Bronson Alcott and Emerson, Goethe and Victor Cousin; with Brokmeyer, he founded the St. Louis Philosophical Club in 1866 and *The Journal of Speculative Philosophy* in 1867. The latter was the first technical philosophical journal in the USA or England, and published papers not only by US and English Hegelians such as Harris and Edward Caird, but by Peirce, Dewey, and William James (parts of *The Principles of Psychology* were first published in the journal). A few weeks of joint philosophical efforts among the Midwest and Eastern 'idealists' and the university professors of philosophy occurred during the summers of 1879-83, when the Concord School, founded by Emerson and Alcott, enlisted Harris, William James, Benjamin Peirce (Charles' father, a Harvard professor of mathematics), James McCosh (last of the Princeton Scottish realists), George Sylvester Morris (the Hegelian teacher of Dewey and Royce at Hopkins), and Emerson himself as lecturers.

The Cambridge Metaphysical Club had its origins in James' 1868 proposal to Oliver Wendell Holmes Jr (1841-1935) that they should establish 'a philosophical society to have regular meetings and discuss none but the very tallest and broadest questions' (Kuklick 1977: 47). By 1871 the club centred around six men, all with Harvard degrees: James and Holmes, Charles Peirce, Chauncey Wright, Nicholas St. John Green, and Joseph Bangs Warner. Green, a Boston attorney, introduced the thought of the British psychologist and philosopher Alexander Bain (1818-1903), particularly his definition of belief as 'that upon which a man is prepared to act'. Wright was a mathematician employed by the *Nautical Almanac* as a 'calculator', and an occasional lecturer in psychology and physics at Harvard. He applied Darwin's evolutionary theory to the development of consciousness in such publications as 'Evolution of Consciousness' (1873), maintaining that consciousness comes about not from any new capacity but from using an old capacity - forming images - in a new way (see Evolution, theory of).

3 Classical American philosophy

Although Wright was regarded as the leader of the Metaphysical Club, Peirce and then James proved to be its most significant members. Peirce seemed destined for intellectual achievement from an early age, and he began publishing papers on logic and semiotics in the 1860s. 'Some Consequences of Four Incapacities' (1868) contains the first published statement of his view that all thought is in signs, and 'On a New List of Categories' (1867) a first statement of his categorial scheme. Peirce presented what came to be called 'the pragmatic maxim' to the Metaphysical Club in an 1872 version of his paper 'How to Make Our Ideas Clear' (1878): 'Consider what effects, which might conceivably have practical bearing, we conceive the object of our conception to have. Then our conception of these effects is the whole of our conception of the object'. In 'The Fixation of Belief' (1877) Peirce considers four ways in which we come to form beliefs: by authority, tenacity (holding on to the beliefs one already has), rationality, or science. Only science, Peirce argues, has the integrity that comes from allowing itself to be determined by 'some external permanency—by something upon which our thinking has no effect' (Peirce 1877).

Peirce worked at the US Coast and Geodetic Survey in the 1860s and 1870s, and was appointed to a lectureship in logic in the new Graduate School at Johns Hopkins in 1879; but he was dismissed in 1884 and, despite occasional lectures at Harvard arranged by William James, never taught regularly again. In a series of papers in *The Monist* in the early 1890s he developed a system of metaphysics according to which absolute chance operates in the universe, but so does 'evolutionary love'; and matter is 'effete mind'. Central to Peirce's many writings was the idea of three categories, Firstness, Secondness, and Thirdness. He held that all signs are 'thirds': besides a purely linguistic element and an object of reference, they contain an irreducible element of interpretation.

William James studied chemistry in the Lawrence Scientific School at Harvard in the 1860s, and biology with Louis Agassiz (including fifteen months in Brazil), receiving his degree in medicine in 1869. He began teaching anatomy and physiology in 1872, and became an assistant professor of psychology in 1875, when he established the first psychological laboratory in America. James' earliest publications did not report research in physiology or the new psychophysics, however, but were a series of critiques of books on science, philosophy and culture. He

argues in ‘The Sentiment of Rationality’ (1879), for example, that reason is a passion, and at the end of ‘Remarks on Spencer’s Definition of Mind as Correspondence’ (1878) he anticipates the voluntaristic pragmatism of his later works:

> The knower is not simply a mirror floating with no foot-hold anywhere, and passively reflecting an order that he comes upon and finds simply existing. The knower is an actor, and co-efficient of the truth on one side, whilst on the other he registers the truth which he helps to create.
>
> (James 1878: 21)

James’ masterpiece, *The Principles of Psychology* (1890) gathers and integrates his writings of the 1870s and 1880s in a one-thousand-page work of physiology, psychology, and philosophy. The book became a standard text in newly established psychology programmes (especially in its shortened form), and influenced philosophers as diverse as Edmund Husserl (by its phenomenological description) and Bertrand Russell (by its distinction between knowledge by acquaintance and by description - see Knowledge by acquaintance and description). James introduces the ideas of the stream of thought and the ‘vague’ or ‘fringe’ areas of consciousness, in opposition to the discrete atomic sensations of traditional British empiricism. He stresses the importance of attention and habit in our mental life, and offers a theory of the emotions as responses to, rather than causes of, emotional behaviour. James’ moral outlook appears throughout the *Principles* and indeed throughout his philosophy, but is particularly explicit and prominent in the collections of papers, some from as early as the 1870s, that he published as *The Will to Believe and Other Essays in Popular Philosophy* (1896).

Although he credited Peirce with originating pragmatism, a lecture James gave at the University of California at Berkeley in 1898 entitled ‘Philosophical Conceptions and Practical Results’, contains the first published use of the term. Pragmatism, for James, is the view that ‘the effective meaning of any philosophic proposition can always be brought down to some particular consequence, in our future practical experience, whether active or passive’. He credits ‘English-speaking philosophers’ such as Locke and Berkeley with introducing the pragmatic ‘custom of interpreting the meaning of conceptions by asking what difference they make for life’, as Berkeley did when he found the ‘cash-value’ of matter to lie solely in our sensations.

Josiah Royce was brought up in the California goldrush town of Grass Mountain, studied English at Berkeley and philosophy in Germany. At Johns Hopkins from 1876 to 1878, he studied with George Sylvester Morris, a scholar of German philosophy and a proponent of T.H. Green. Receiving his Ph.D. in 1878, Royce taught English at Berkeley, then philosophy at Harvard, where he became a mainstay of the department. Royce introduced formal logic into the curriculum, and was a respected idealist opponent of James’ more naturalistic, open-ended pragmatism.

Royce’s early philosophical writing is in accord with his lifelong interests both in the history of philosophy, and in developing his own version of metaphysical idealism. His first book, *The Religious Aspect of Philosophy* (1885) argues for an Absolute Mind that contains all thoughts and their objects. In *The Spirit of Modern Philosophy: An Essay in the Form of Lectures* (1892), Royce traces ‘the rediscovery of the inner life’ from Spinoza to Kant, with special emphasis on Fichte - praised for his ‘beautiful waywardness’, the Romantic School, including Goethe, Novalis and Schelling, and Hegel. Royce argues, however, that the inner life is essentially public: that we live in our coherence or relationships with other people.

The third great pragmatist to emerge in the late nineteenth century, John Dewey, had neither the scientific background of Peirce and James, nor their association with Harvard. Dewey attended the University of Vermont in his home town of Burlington from 1875 to 1879. He studied not only the Scottish school but Kant and Hegel with the university’s philosophy professor, H.A.P. Torrey (1837-1902). According to his own testimony, Dewey found in Hegel’s philosophy ‘an immense release, a liberation’ from a sense of divisions between self and world, soul and body, nature and God. Enrolling in the new graduate school at Johns Hopkins in 1882, he studied Hegel and Green with Morris, logic with Charles Peirce, and the newly emerging experimental psychology with G. Stanlely Hall (1844-1924). Appointed to a post at the University of Michigan in 1884, he taught there, with the exception of a year at Minnesota, till 1894 when he began teaching at the University of Chicago.

Dewey’s early papers argue for a reconciliation of Darwinism, Hegelian idealism and religion. Intelligence, Dewey asserts, is latent in evolving matter. In the 1890s Dewey called his synthesis of Hegelianism and science

'experimental idealism', but he gradually moved - as he says in the title of his autobiography - 'from absolutism to experimentalism'. His paper 'The Reflex Arc Concept in Psychology' (1896), presages his future instrumentalism and pragmatism in its attacks on the prevailing stimulus-response theory, which Dewey sees as preserving a sharp metaphysical and epistemological distinction between sensory stimulation and motor response. Stimulus and response are, Dewey argues, aspects of a basic 'sensorimotor coordination', a 'circuit' or 'continual reconstitution'. The sensorimotor coordination, like Dewey's later 'problem situation', shares with Hegelian logic the idea of a progression of temporally evolving wholes.

Dewey's educational philosophy also took shape in the 1890s, when he was a professor not only of philosophy but also of psychology and pedagogy. He worked with high school faculty in Michigan, and with the Laboratory School at Chicago. In 'Interest in Relation to the Training of the Will' (1896), Dewey argues that because interest is a complex of felt worth and incipient action, when we are genuinely interested in something, we do not have to will to do it. Only through such genuine interest, which 'marks the annihilation of the distance between the person and the materials and results of his action', can the will be effectively trained (Dewey 1896). In 'My Pedagogic Creed' (1897), Dewey maintains that education is 'a process of living and not a preparation for future living', and that therefore it must seek 'forms of life that are worth living for their own sake' (Dewey 1897).

See also: Hegelianism §5

RUSSELL B. GOODMAN
WILLIAM C. DOWLING

References and further reading

Works and English translations of philosophers mentioned in the text and not included here can be found in the individual biographical entries throughout the *Encyclopedia*.

Appleby, J. (1984) *Capitalism and a New Social Order: The Republican Vision of the 1790s*, New York: New York University Press.(Argues that Jeffersonian democracy was the triumph in America of an economic individualism rooted in the political thought of John Locke.)

Bailyn, B. (1973) *The Ideological Origins of the American Revolution*, Cambridge, MA: Belknap Press of Harvard University Press.(The book that discovered the roots of Revolutionary thought in 'country ideology' - the local variant of classical republicanism that the colonists imported from England.)

Boorstin, D. (1964) *The Americans: The Colonial Experience*, New York: Vintage Books.(The American Revolution as the ideological triumph of Lockean liberal individualism.)

Cameron, S. (1985) *Writing Nature: Henry Thoreau's Journal*, New York and Oxford: Oxford University Press.(A fine literary study, discussing the relation of Thoreau's immense journal to nature, to his audience and to Walden.)

Cavell, S. (1980) *The Senses of Walden, An Expanded Edition*, Chicago, IL: University of Chicago Press.(A pioneering study of Thoreau as a philosopher of language and knowledge. The essays 'Thinking of Emerson' and 'An Emerson Mood' explore Emerson's epistemology and his relation to Heidegger.)

Cavell, S. (1988) *In Quest of the Ordinary: Lines of Skepticism and Romanticism*, Chicago, IL: University of Chicago Press.(Contains 'Being Odd, Getting Even,' an existentialist reading of Emerson's 'Self-Reliance', and 'The Philosopher in American Life', discussing the 'repression' of Emerson and Thoreau in accounts of American philosophy and culture.)

Cavell, S. (1990) *Conditions Handsome and Unhandsome: The Constitution of Emersonian Perfectionism*, Chicago, IL: University of Chicago Press.(Moral and political aspects of Emerson's thought.)

Clebsch, W. (1973) *American Religious Thought*, Chicago, IL: University of Chicago Press.(Readable and sympathetic account, emphasizing continuities among Edwards, Emerson and James.)

Conkin, P.K. (1968) *Puritans and Pragmatists: Eight Eminent American Thinkers*, Bloomington, IN: Indiana University Press.(Fine study of Edwards, Franklin, Adams, Emerson, Peirce, James, Dewey and Santayana by a distinguished historian.)

Coughlan, N. (1975) *Young John Dewey*, Chicago, IL: University of Chicago Press.(Superb short account of Dewey's development and career to 1894, when he began teaching at the University of Chicago.)

Dewey, J. (1897) 'My Pedagogic Creed', *School Journal* 54: 77-80; repr. in *The Early Works of John Dewey*, ed. J.A. Boydston, Carbondale, IL: Southern Illinois University Press, 1969-90, vol. 5, 84-95.(Dewey maintains that education must proceed through 'forms of life' that are 'worth living for their own sake'.)

Dowling, W.C. (1990) *Poetry and Ideology in Revolutionary Connecticut*, Athens, GA and London: University of Georgia Press.(Excellent survey of relations between classical republican political theory and poetry during the

Revolution and early republic.)

Dunn, J. (1969) 'The Politics of Locke in England and America in the Eighteenth Century', in J.W. Yolton (ed.) *John Locke: Problems and Perspectives*, Cambridge: Cambridge University Press, 45-80.(First important allegation of Locke's non-importance in eighteenth-century political theory in both England and America.)

Fiering, N. (1981) *Moral Philosophy at Seventeenth-Century Harvard*, Chapel Hill, NC: University of North Carolina Press.(Discusses the yielding of scholastic premises to Hutchesonian moral sense theory between 1650 and 1725.)

Flower, E. and Murphey, M.G. (1977) *A History of Philosophy in America*, New York: Capricorn Books, 2 vols. (Volume 1 contains important chapters on Jonathan Edwards and the development of Scottish common sense philosophy in the American context. Volume 2 discusses the St. Louis Hegelians, Peirce, James, Royce, Santayana and Dewey.)

Goodman, R.B. (1990a) *American Philosophy and the Romantic Tradition*, Cambridge and New York: Cambridge University Press.(Challenges the view that American philosophy originates with the Puritans by considering the influence of Romantic poetry and philosophy on Emerson, James and Dewey.)

Goodman, R.B. (1990b) 'East-West Philosophy in Nineteenth Century America: Emerson and Hinduism', *Journal of the History of Ideas* 1990: 625-45.(Discussion of Emerson's interest in and use of Hindu texts in such essays as 'Plato, or the Philosopher'.)

Guelzo, A.C. (1989) *Edwards on the Will: A Century of American Theological Debate*, Middletown, CT: Wesleyan University Press.(Summary account of Edwards' New Divinity and its Historiography.)

Hamilton, A. (and Madison, J.) (1787-8) *The Federalist Papers*, ed. C. Rossiter, New York: New American Library, 1961.(A defence of the new United States Constitution and the federal government it represents.)

Hartz, L. (1955) *The Liberal Tradition in American History: An Interpretation of American Political Thought Since the Revolution*, New York: Harcourt Brace.(The *locus classicus* of the 'Lockean liberal' interpretation of the American Revolution.)

James, W. (1878) 'Remarks on Spencer's Definition of Mind as Correspondence'; repr. in *The Will to Believe and Other Essays in Popular Philosophy*, Cambridge, MA: Harvard University Press, 1979.

(1879) 'The Sentiment of Rationality', *Mind* 4: 317-46; repr. in *The Will to Believe and Other Essays in Popular Philosophy*, Cambridge, MA: Harvard University Press, 1979.(Characterizes the sentiment of rationality as a feeling of ease, peace and rest, the result of a transition from a state of perplexity to one of lively relief and pleasure.)

Kuklick, B. (1977) *The Rise of American Philosophy: Cambridge, Massachusetts, 1860-1930*, New Haven, CT: Yale University Press.(A history of the great Harvard department of James, Royce and others; excellent bibliography.)

Kuklick, B. (1985) *Churchmen and Philosophers*, New Haven, CT: Yale University Press.(Traces continuing religious strains in American thought from Edwards through Dewey; contains a bibliographic essay.)

McCoy, D. (1980) *The Elusive Republic: Political Economy in Jeffersonian America*, Chapel Hill, NC: University of North Carolina Press.(Traces the influence of Rousseau and classical republican political theory on Jefferson's idealized agrarianism.)

Marsh, J. (1976) *Selected Works of James Marsh*, intro. P.C. Carafiol, Delmar, NY: Scholars Facsimiles and Reprints, 3 vols.(Volume 1 contains the 1829 edition of S.T. Coleridge's *Aids to Reflection*, with Marsh's preliminary essay - an important source for Emerson and Dewey.)

Miller, P. (1939) *The New England Mind*, New York: Macmillan.(Classic account of Covenant theology as it developed in New England between the first and second Puritan generations.)

Perry, R.B. (1935) *The Thought and Character of William James*, Boston, MA: Little Brown & Company; repr. Nashville, TN: Vanderbilt University Press, 1995.(Affectionate and philosophically sophisticated account of James and his work).

Pocock, J.G.A. (1975) *The Machiavellian Moment: Florentine Political Thought and the Atlantic Republican Tradition*, Princeton, NJ: Princeton University Press.(Magisterial summary of the classical republic political tradition from Aristotle to Jefferson. Claims that English Opposition thought - Bolingbroke, Trenchard and Gordon - rather than Locke, was the real source of revolutionary theory in the colonies.)

Rockefeller, S.C. (1991) *John Dewey: Religious Faith and Democratic Humanism*, New York and Oxford: Columbia University Press.(Contains an account of Dewey's early development as a Hegelian idealist and Christian social activist.)

Schneider, H. (1963) *A History of American Philosophy*, New York: Columbia University Press, 2nd edn.(Classic study, containing a good bibliography of primary sources.)

West, C. (1989) *The American Evasion of Philosophy: A Genealogy of Pragmatism*, Madison, WI: University of

Wisconsin Press.(Political and prophetic interpretation of American philosophy, beginning with Emerson.)
Wills, G. (1978) *Inventing America: Jefferson's Declaration of Independence*, Garden City, NY: Doubleday. (Scottish moral sense theory as the origin of the language of rights in the American constitution.)
Wright, C. (1877) *Philosophical Discussions*, ed. C.E. Norton; repr. New York: Burt Franklin, 1971.(Contains 'Evolution of Consciousness' (199-266) and other writings on Darwinian theory.)
Zuckert, M. (1994) *Natural Rights and the New Republicanism*, Princeton, NJ: Princeton University Press. (Reasserts the importance of Locke's influence on American political thought.)

Ammonius, son of Hermeas (*c.* AD 440-521)

The Greek philosopher Ammonius, 'son of Hermeas' was an Alexandrian Neoplatonist. Educated by Proclus in Athens, he succeeded his father as head of the school in Alexandria, where he cultivated the tradition of learned commentary on Aristotle. Simplicius, Philoponus, Asclepius, Damascius and Olympiodorus ranked among his pupils.

Ammonius' calibre is hard to assess: important works like commentaries on Aristotle's *On the Heavens* and *Meteorology* and on Plato's *Gorgias*, as well as monographs on logic and theology, are lost. Moreover, with the exception of the commentary on Aristotle's *De interpretatione*, Ammonius left the task of writing out his lectures in full to his students. The case of Philoponus (§1) teaches us that they were not always entirely faithful to their master's ideas. Ancient sources praise Ammonius' abilities as a commentator, mathematician and astronomer. The extant works bearing his name suggest that he dedicated himself to sober, instructive and learned exegeses of Aristotelian treatises. Ammonius held strong views on the order in which these had to be studied, starting with logic, followed by ethics, physics, mathematics and finally theology (*On Aristotle's Categories* 5.31-6.8). In preparation for Aristotle's *Categories*, freshmen were exposed to Porphyry's *Introduction (Isagōgē)* to Aristotle's *Categories*, a logical treatise in which Porphyry discusses the five predicables (genus, species, differentia, property, and accident). Ammonius begins by telling his class what philosophy means to a Neoplatonist (*On Porphyry's Introduction* 2.22-5.27), emphasizing that philosophy assimilates the soul to the godhead, separates it from the body, and turns it towards the source and highest Neoplatonic principle, the One.

Ammonius accepts the common Neoplatonist contention that Plato and Aristotle substantially agree with one another. The *Categories* commentary begins with ten questions on Aristotle's philosophy, concerning the division and order of his works, their format and purpose, and, finally, the requisite character of a commentator. According to Ammonius, a commentator ought to adopt a position of critical independence: the truth must be preferred to what Aristotle says (8. 15-18).

In his commentary on Aristotle's *De interpretatione* Ammonius dedicates an excursus on Chapter 9 to the problem of future contingents. Although he upholds God's full knowledge of events past, present and future, he nevertheless rejects determinism. God sees all things in the way appropriate to him, that is, at once in a single, eternal now, but this does not entail that future events are determined. The whole commentary seems to owe much to Proclus' lectures.

In a lost theological treatise Ammonius argued that Aristotle's prime mover is both the final and the efficient cause of the universe. The suggestion that the ideas put forward in the treatise were motivated by the influence of Christianity and amounted to a substantial deviation from orthodox Neoplatonism has been plausibly rejected.

The *Metaphysics* commentary of Asclepius of Tralles as well as the early commentaries of Philoponus are greatly indebted to Ammonius, while the *De interpretatione* commentary influenced Aquinas by way of a Latin translation.

See also: Aquinas, T.; Aristotle Commentators; Neoplatonism §1; Simplicius §1

CHRISTIAN WILDBERG

List of works

It is impossible to date Ammonius' works with any accuracy, although it is likely that they were composed after the death of Proclus in 485 and before Philoponus assumed responsibility for publication in the decade beginning 510. Below they are listed in the order in which Ammonius intended them to be read.

Ammonius Hermeae (*c*.485-510) *On Porphyry's Introduction*, ed. A. Busse, *Commentaria in Aristotelem Graeca* IV 3, Berlin: Reimer, 1891.(A commentary on Porphyry's immensely influential introductory treatise on Aristotle's *Categories*.)

— (*c*.485-510) *On Aristotle's Categories*, ed. A. Busse, *Commentaria in Aristotelem Graeca* IV 4, Berlin: Reimer, 1895; trans. S.M. Cohen and G.B. Matthews, *Ammonius On Aristotle's Categories*, London: Duckworth, 1991. (The latter contains a concise introduction to Ammonius by Richard Sorabji.)

— (*c*.485-510) *On Aristotle's De interpretatione*, ed. A. Busse, *Commentaria in Aristotelem Graeca* IV 5, Berlin:

Reimer, 1897; partial trans. by D. Blank, *Ammonius' Commentary on On Interpretation*, London: Duckworth, 1995.(The commentary contains, on pages 129-39, an excursus on the divine knowledge of future contingents.)

— (*c*.485-510) *On Aristotle's Prior Analytics I*, ed. M. Wallies, *Commentaria in Aristotelem Graeca* IV 6, Berlin: Reimer, 1899.(Ammonius' lectures on Aristotle's theory of the syllogism.)

References and further reading

Barnes, J. (1991) 'Ammonius and Adverbs', *Oxford Studies in Ancient Philosophy*, suppl. vol.: 145-63.(An ideosyncratic discussion of aspects of Ammonius' modal logic.)

Merlan, P. (1968) 'Ammonius Hermiae, Zacharias Scholasticus and Boethius', *Greek, Roman and Byzantine Studies* 9: 193-203.(A brief discussion of the Ammonius, a dialogue written by the Christian Zacharias Scholasticus and focusing on the pagan/Christian controversy about the eternity of the world.)

Obertello, L. (1981) 'Proclus, Ammonius and Boethius on Divine Knowledge', *Dionysius* V: 127-64.(This article includes a lucid outline of Ammonius' solution to the problem of future contingents and divine foreknowledge.)

Verrycken, K. (1990) 'The Metaphysics of Ammonius Son of Hermeias', in R.R.K. Sorabji (ed.) *Aristotle Transformed*, London: Duckworth, 199-231.(A rejection of the notion that Ammonius substantially modified the Neoplatonist metaphysical scheme.)

Westerink, L.G. (1990) 'The Alexandrian Commentators and the Introductions to their Commentaries', in R.R.K. Sorabji (ed.) *Aristotle Transformed*, London: Duckworth, 325-48.(This whole volume is an excellent guide to philosophy in late antiquity, offering an extensive bibliography.)

Amo, Anton Wilhelm (*c*.1703-56)

The first European-trained African philosopher, Amo pursued a scholarly career in jurisprudence and then in rationalist psychology, logic, and metaphysics. He trained at Halle, Wittenberg and Jena universities, and was influenced by the systems of Gottfried Wilhelm Leibniz and Christian von Wolff. While at Halle university, he wrote a pioneering legal dissertation on the application of Roman laws of slavery to Africans in Europe. Subsequently drawn to classical, biblical, and hermetic traditions that apotheosized a cultural continuity with ancient Africa, Amo focused his theoretical and practical concerns on the exterior world of international law and the interior world of deliberative intellectual acts.

The first African philosopher to study and teach in European universities, Anton Wilhelm Amo was born at Axim in Ghana to Nzima parents who were converted to Christianity by Dutch missionaries. Sent to The Netherlands at the age of four for a religious education, Amo was then transferred by representatives of the Dutch West India Company to the service of Anton Ulrich, Duke of Brunswick-Wolfenbüttel, whose Hapsburg court was the intellectual centre of early Enlightenment Germany. Through the Duke's patronage, which also supported G.W. Leibniz, Amo became proficient in Dutch, Hebrew, Greek, Latin, French and German and subsequently trained at the universities of Halle, Wittenberg and Jena in philosophy and jurisprudence.

As a student at Halle of Johann Peter von Ludewig, the Prussian diplomat and legal scholar, Amo prepared a dissertation, *De Jure Maurorum in Europa (On the Rights of Moors in Europe)* (1729), which utilized Roman law, hermeneutical and historiographical traditions to assert ancient prohibitions against enslaving Africans in Europe. He based his argument on legal rights ostensibly inherited from the early Christian era compacts between the Emperor Justinian and the indigenous kings of Roman north Africa. Amo's dissertation, extant only in digest form, represents one of the earliest scholarly responses to the growing legal dispute over slavery and the status of Africans as aliens in Europe during a period of expanding overseas slave colonies. It reflected the impact of rising African populations in metropolitan Europe and the Holy Roman Empire's disintegrating hegemony. The natural law principles and imperial precedents he employed paralleled those developed in Spain by the legal philosopher and theologian, Francisco de Vitoria, to oppose the 'Astral Empire' that Charles V planned to build, through slavery, at the outset of the New World conquest.

In philosophy, where Amo's interests inclined towards the Enlightenment rationalism of Leibniz and the related system of Christian Wolff, he specialized in pneumatology (rational psychology), logic and metaphysics. In rational psychology, he attempted to reconcile the tensions between Thomistic faculty psychology (see Aquinas, T. §1; Thomism §1) as it focused on the 'free' operations of intellect or will, and the more deterministic empirical psychology rooted in Lockean sensationalism (see Locke, J.) and medical physiology. In logic, he explored the nature of intellective acts such as contemplation, deliberation and reflection. Applying the mathematico-deductive method, Amo adopted precepts from the Port Royal semioticians, from modal logic and syllogistic methods, and from Christian Thomasius's 'practical logic' (see Thomasius, C.). In metaphysics, a religious attraction to the Cistercian, Carmelite and Franciscan spiritual orders immersed him in the study of Aristotle's *On the Soul* and the church fathers Tertullian and St Augustine, all of whom Amo traced through such modern thinkers as P. Melanchthon, R. Descartes, G.W. Leibniz, J. Le Clerc, and C. Wolff.

As a professor at Halle, Wittenberg and Jena, Amo lectured on topics that included Leibniz's principle of sufficient reason, the political thought of C. Wolff, J. Lorenz Fleischer's theories of the law of nature and the law of nations, and the decimal system. However, his interests were not confined to the formal university curriculum: as a private tutor, he conducted classes on several of the *ars hermetica* (the occult arts and sciences), including physiognomy, chiromancy, geomancy, natural astrology and decipherment. In 1738 as a member of various learned societies in Europe, including the academy of Flushing in The Netherlands, he published in Latin a compendium of his selected university lectures known as *Tractatus de Arte Sobrie et Accurate Philosophandi (A Treatise on the Art of Philosophizing Soberly and Accurately)*.

After the death of his academic patron, von Ludewig, Amo found his position in German academic and social circles increasingly tenuous and himself subject to racial rebuff. In 1747 he returned to Ghana and changed his profession to the honoured Ashanti vocations of goldsmith and seer. He died there in 1756. Since his philosophical

writings remained largely inaccessible until their recovery and dissemination in the twentieth century, Amo's historical influence lies primarily in his posthumous role in the international antislavery movement, where his example served to vindicate African moral and intellectual capacities. He is also known as a moral and intellectual vindicationist.

See also: Slavery

JOHN S. WRIGHT

List of works

Amo, A.W. (*c*.1703-56) *Antonius Guilielmus Amo Afer of Axim in Ghana: Translation of his Works*, ed. D. Siegmund-Schultze; trans. L.A. Jones and W.E. Abraham, Halle: Martin Luther University Halle-Wittenberg, 1968.(A collection of all of Amo's works.)

Amo, A.W. (1729) *De Jure Maurorum in Europa (On the Rights of Moors in Europe)*, unpublished.(Amo's dissertation against the enslavement of Africans in Europe.)

Amo, A.W. (1738) *Tractatus de Arte Sobrie et Accurate Philosophandi (A Treatise on the Art of Philosophizing Soberly and Accurately)*, unpublished.(A compendium of Amo's university lectures.)

References and further reading

Abraham, W.E. (1964) 'The Life and Times of Anton Wilhelm Amo', *Transactions of the Historical Society of Ghana* 7: 60-81.(Accessible biographical discussion.)

Brentjes, B. (1977) 'Anton Wilhelm Amo in Halle, Wittenberg and Jena', *Universitas* 6 (1): 39-55.(Accessible discussion of Amo's period in these three universities.)

Hountondji, P.J. (1983) 'An African Philosopher in Germany in the Eighteenth Century: Anton Wilhelm Amo', *African Philosophy: Myth and Reality*, Bloomington, IN: Indiana University Press, 111-30.(Reflections on Amo by a major contemporary African philosopher.)

Lochner, N. (1958) 'Anton Wilhelm Amo: a Ghana Scholar in Eighteenth Century Germany', *Transactions of the Historical Society of Ghana* 3: 169-79.(First significant modern discussion of Amo's life.)

Analysis, nonstandard

Nonstandard analysis is an important application of mathematical logic to the rest of mathematics. Invented in 1960, it provided a long-sought-for rigorous justification for the use of infinitely large and infinitely small (infinitesimal) quantities in the differential and integral calculus, and the first sound canon for manipulating such quantities.

Consider the structure $\Re$ *of real numbers, that is, the set of real numbers together with operations and relations on them. We specify a formal language* $\mathcal{L}$*, which has names for all individual real numbers as well as for the operations and relations of* $\Re$*. We can then obtain an enlargement (a special kind of extension)* ${}^*\Re$ *('pseudo-* $\Re$*') of* $\Re$ *such that* ${}^*\Re$ *and* $\Re$ *have exactly the same formal properties: that is, properties expressible in* $\mathcal{L}$*. But, far from being merely a copy of* $\Re$*,* ${}^*\Re$ *has convenient properties not expressible in* $\mathcal{L}$*, which are not shared by* $\Re$*. In particular,* ${}^*\Re$ *has additional 'nonstandard' objects (which have no names in* $\mathcal{L}$*), some of which behave as infinite or infinitesimal quantities. If, using these novel objects and properties of* ${}^*\Re$*, we prove a proposition about* ${}^*\Re$ *that can be expressed in* $\mathcal{L}$*, then this proposition holds automatically also for* $\Re$*. Such 'nonstandard' proofs of results about* $\Re$ *are often easier and more intuitive than conventional proofs, which operate wholly within* $\Re$*.*

More generally, this method is applied to other structures, in virtually every branch of mathematics - including algebraic number theory, various branches of classical and modern analysis, probability theory and mathematical physics, to yield highly intuitive characterizations of various infinitary concepts, to simplify proofs and to provide new, efficient ways of generating mathematical constructs.

1 Historical background

The lineage of nonstandard analysis (NSA) goes back to Leibniz in two ways: first, his advocacy of the use of infinite and infinitesimal quantities in analysis was remarkably sophisticated for his time (see Analysis, philosophical issues in §1); and second, his vision of *lingua characteristica* and *calculus ratiocinator* foreshadowed the invention of mathematical logic, which was to provide the means for a rigorous justification of the use of those quantities.

Leibniz regarded infinite and infinitesimal quantities as 'useful fictions'; ideal (rather than metaphysically real) entities, whose status was similar to that of imaginary roots such as $\sqrt{-2}$, and whose adjunction to the standard real numbers serves to shorten proofs and aid mathematical invention in problems involving continuity.

Like complex (and, in particular, imaginary) numbers, infinite and infinitesimal quantities were freely used by mathematicians during the eighteenth century, despite known inconsistencies apparently inherent in this practice, which provoked protests by philosophical critics such as Berkeley.

During the nineteenth century, with the development of a more critical attitude to the foundations of analysis, the notions of the real and complex number structures were greatly clarified, and consistent canons were established for operating with these entities (see Numbers). However, no rigorous canon was found for the use of infinite and infinitesimal quantities. (Cantor, the inventor of set theory, thought that he could actually *prove* that no such canon was possible, and his view was accepted by Peano and others.) Instead, work on the foundations of analysis - by Bolzano, Weierstrass, Dedekind and others - followed a different route (whose viability had also been asserted by Leibniz): the 'delta-epsilon' method, whereby 'instead of the infinite or infinitely small, one takes [standard] quantities as large or as small as required to make the error smaller than [any] given error' (see Analysis, philosophical issues in §2).

By the twentieth century, the success of this method (in effect, an extension of Archimedes' method of exhaustion) had been so spectacular and its domination in pure mathematics so total, that Russell was merely expressing received opinion when he asserted that 'infinitesimals as explaining continuity must be regarded as unnecessary, erroneous, and self-contradictory' ([1903] 1937: 345). Nevertheless, the older, now officially discredited, method survived in 'lowbrow' (mainly applied) mathematics and - as a powerfully creative heuristic aid - in the 'underground' private deliberations of 'highbrow' mathematicians.

Before the advent of NSA, some isolated attempts were still made, in the face of received opinion, to provide a

rigorous justification for the use of infinite and infinitesimal quantities in analysis. None of these was fully successful, although Schmieden and Laugwitz (1958) came quite close. The successful solution of this problem, achieved by NSA, was to use tools produced by mathematical logic during the first half of the twentieth century, in particular, the notion of a nonstandard model. Skolem showed the existence of nonstandard models of axiomatic set theory (1923) and of arithmetic (1934), but for a long time such models were regarded by logicians as pathological 'monsters' rather than a serious object of study - let alone a useful tool. This attitude began to change in about 1949, and during the 1950s there was a growing body of research into the properties and inner make-up of nonstandard models of arithmetic. This set the stage for Abraham Robinson's invention, in the autumn of 1960, of NSA, which uses nonstandard models as an instrument.

2 Enlargements

The method consists in systematically exploiting the peculiar properties of an enlargement (a special kind of nonstandard extension) $^*\mathrm{U}$ (pronounced 'pseudo- U') of a mathematical structure U, in order to study U itself.

As a paradigmatic example, consider the structure $\Re$ whose domain of individuals is the set $\mathbb{R}$ of real numbers. Apart from the domain $\mathbb{R}$, $\Re$ has as constituents all relations and operations on this domain. Let $\mathcal{L}$ be a first-order formal language which can be interpreted in $\Re$: so that the domain $\mathbb{R}$ can be taken as the range of the variables of $\mathcal{L}$, and for each entity of $\Re$ - an individual real, or a relation among reals, or an operation on reals - $\mathcal{L}$ has a symbol intended as a name for that entity.

Using the 'compactness theorem', we can prove the existence of a structure $^*\Re$, satisfying two conditions. First, $^*\Re$ is an 'elementary extension' of $\Re$. Second ('weak saturation condition'), if Φ is any family of $\mathcal{L}$-formulas having one free variable x, which is finitely satisfiable in $\Re$ (that is, every finite subset of Φ is satisfied by some individual real), then Φ as a whole is satisfiable in $^*\Re$.

The elementary extension condition is tantamount to the following 'transfer principle': any $\mathcal{L}$-sentence is true (when interpreted) in $\Re$ if and only if it is true (when re-interpreted) in $^*\Re$.

Thus any property of $\Re$ that can be formalized by means of an $\mathcal{L}$-sentence is true also of $^*\Re$, and conversely. For example, many well-known arithmetical laws (such as the fact that $x + y = y + x$ for any reals *x* and *y*) are automatically transferred to the extended domain $^*\mathbb{R}$ of the structure $^*\Re$. Similarly, the fact that the relation denoted by '$<$' is a total (that is, linear) ordering of $\mathbb{R}$ also transfers to $^*\mathbb{R}$. In this sense, $\Re$ and $^*\Re$ have exactly the same formal properties.

The members of $^*\mathbb{R}$ are called '*reals' ('pseudo-reals'). Note that every (standard) real (that is, member of $\mathbb{R}$) is also a *real, but not conversely. A *real which does not belong to $\mathbb{R}$ is said to be 'nonstandard'.

As an illustration of the (weak) saturation condition, consider the family of all $\mathcal{L}$-formulas of the form $\mathbf{r} < \mathbf{x}$, where x is a variable (common to all these formulas) and r is any $\mathcal{L}$-constant denoting a real *r*. Clearly, this family of formulas is finitely satisfiable in $\mathbb{R}$, because for any finitely many reals $r_1, r_2, \ldots, r_n$ there is always another real greater than all of them. Hence by saturation there exists some - necessarily nonstandard - *real *s* such that $r < s$ for every real *r*. Such an *s*, being greater than all reals, is to be regarded as a 'positive infinite' quantity. Similarly, there are nonstandard 'negative infinite' *reals. The reciprocals of infinite *reals, as well as the real 0, are 'infinitesimals', which are smaller than all positive reals and greater than all negative ones.

If *s* and *t* are *reals such that $s - t$ is infinitesimal, we write '$s \approx t$' and say that *s* and *t* are 'infinitely close' to each other. In whimsical homage to Leibniz, the collection $\mu(r)$ of all *reals infinitely close to a given real *r* was termed by Robinson the 'monad' of *r*.

By referring to the entities of an enlargement $^*\Re$, many standard concepts of analysis receive intuitively appealing characterizations which, while being rigorous, are often strikingly similar to the definitions used (without rigorous justification) before the advent of the 'delta-epsilon' method. For example, let *f* be a real-valued function defined in the neighbourhood of a real *r*. Then *f* is continuous at *r* if and only if $f(s) \approx f(r)$ whenever $s \approx r$. Also, *f* is differentiable at *r* if and only if there is a unique real *d* such that

$\frac{f(r+\delta)-f(r)}{\delta} \approx d$

for every non-zero infinitesimal δ. This *d* is then the derivative of *f* at *r*.

Such nonstandard characterizations of standard concepts can be used to obtain nonstandard proofs of standard results concerning $\Re$ itself. NSA is ‘conservative’, in the sense that any standard result that can be proved using NSA can, at least in principle, be proved also by purely standard means. However, the nonstandard proof is often considerably shorter and more transparent - and therefore easier to construct - than a conventional standard proof. Moreover, as shown by Henson and Keisler (1986), NSA is essentially more powerful than traditional standard analysis. (See also Keisler 1994.)

The ideas outlined so far in connection with the first-order structure $\Re$ of reals apply quite generally to any first-order mathematical structure U. Any such U can be taken as the ‘ground’ structure: we can prove the existence of an enlargement ${}^{*}\mathrm{U}$ of U satisfying the same two conditions (elementary extension and weak saturation).

3 Higher-order and special enlargements

Modern analysis is largely concerned with higher-order entities such as arbitrary sets of reals, sets of real-valued functions, sets of sets of functions, sets of sets of reals, functions from sets of reals to sets of reals, and so on.

To apply NSA in this wider context, Robinson used (from 1962) a type-theoretic approach: he considered a ‘layered’ type-theoretic structure which, in addition to the ground-level individuals, say, the reals, also contains all entities of finite orders obtainable from these individuals (such as those mentioned in the preceding paragraph). By a well-known procedure, a layered type-theoretic structure can be replaced by a first-order structure. Essentially, this amounts to considering the higher-order entities as ground-level individuals of different sorts. The membership relation between entities of different orders now becomes a first-order relation connecting individuals of different sorts. This many-sorted first-order structure can be taken as the ground structure U whose enlargements are obtained as outlined in §2 above.

An equivalent but technically simpler approach, first proposed by Machover (1967), is to embed the original set of individuals in a model of Zermelo’s set theory - a sufficiently large part of the so-called ‘cumulative hierarchy of sets’ (see Set theory §1). This constitutes a first-order structure in which all the entities of analysis are present as individuals (of the same sort). The basic relations in such a Zermelo structure are the first-order relations of equality ($=$) and set membership ($\in$). Zermelo structures can be enlarged in the usual way.

Another set-theoretic approach, invented by E. Nelson (1977), proceeds axiomatically. Nelson’s internal set theory is a conservative extension of Zermelo-Fraenkel set theory with the axiom of choice (see Set theory §1). Its universe of discourse contains all standard sets as well as nonstandard *sets. Using a special formal primitive predicate, ‘is standard’, counterparts of the transfer principle and saturation conditions are decreed as additional postulates.

For many applications of NSA it is irrelevant which of the infinitely many enlargements of a given ground structure is used. But special enlargements are sometimes singled out either for convenience of exposition or because their special properties are needed for a particular application.

One special kind of enlargement is obtained by taking an ‘ultrapower’ of the ground structure - an infinitary construction foreshadowed by Skolem in 1934 for the natural numbers and Hewitt (1948) for the reals. Luxemburg used an ultrapower enlargement of $\Re$ in one of the earliest broad expositions of NSA (1962). The advantage of this lucid presentation is that it does not presuppose acquaintance with logical technicalities, and was therefore particularly suited for an audience of mathematicians unfamiliar with them. However, it paid for this convenience by having to prove separately each instance of the transfer principle.

Some applications of NSA need enlargements that satisfy certain stronger saturation conditions. Notably, the extremely useful procedure for obtaining a Loeb measure (see below) requires the enlargement to satisfy a condition known as ‘$\aleph_1$-saturation’.

4 Range of applications

Robinson himself initiated a formidable array of applications of NSA in a wide spectrum of mathematical topics,

ranging from algebraic number theory, through various branches of classical and modern analysis and probability theory, to mathematical physics. Here we can mention only a few applications.

Generally speaking, NSA is a particularly powerful tool in problems involving the topological notion of compactness, because the nonstandard characterization of compactness is extremely simple and transparent compared with the standard definition.

Indeed, one of the earliest impressive successes of NSA was the solution by Robinson (1966) with A.R. Bernstein of an open problem concerning Hilbert spaces, which involved topological compactness in an essential way. Since then there have been many important applications to linear spaces and functional analysis.

NSA also provides illuminating insights into structures in which continuity or analyticity and algebraic notions interact; for example, in topological and analytic groups, whose nonstandard study was initiated by Robinson. In such structures, the whole of the continuous or analytic aspect is 'algebraicized' and encapsulated in a single group of infinitesimal *members. Exploiting these ideas, Hirschfeld (1990) was able to simplify very significantly the solution of one of Hilbert's celebrated problems.

Some of the most spectacular achievements of NSA are in probability theory. In this connection, E. Nelson's work (1987) is noteworthy in providing a far-reaching simplification of difficult parts of the theory. Arguably the most important breakthrough in NSA published after Robinson's death was the invention by Loeb (1975) of a powerful nonstandard procedure for obtaining standard probability measures (as well as more general unbounded measures) with prescribed desirable properties.

Robinson himself was - unusually for a logician - an expert in applied mathematics and mathematical physics (particularly in fluid dynamics). Applications of NSA in these areas, initiated by him and continued by others, have been of very great importance.

5 Status of NSA

If revolutions have ever occurred in mathematics, then NSA may be described as one of them (Dauben 1992). Logicians have in general welcomed it, in some notable cases with great enthusiasm. Thus, in Gödel's view, 'there are good reasons to believe that nonstandard analysis, in some version or other, will be the analysis of the future' (see Robinson [1966] 1974). Some logicians have written elementary calculus text-books for beginners using NSA (for example, Keisler 1976).

Among working mathematicians generally, the spread of NSA techniques has been relatively slow. As pointed out in §2, any nonstandard proof of a standard theorem can theoretically be replaced by a standard proof. Although nonstandard techniques are often significantly more efficient, they require familiarity with logic, which many mathematicians do not possess. However, the acceptance of NSA among mathematicians has been boosted by the successes of its practitioners. In this respect the results initiated by Loeb's work (see §4) have played a pivotal role.

But the view that NSA is destined to supersede standard analysis altogether should be taken with a grain of salt. This is because a given structure has infinitely many enlargements, none of which can be singled out as in any way canonical or 'best', and a standard definition is needed in order to verify that a nonstandard characterization of a concept is independent of the choice of enlargement (see Machover 1993).

See also: Logical and mathematical terms, glossary of

MOSHÉ MACHOVER

References and further reading

Capiński, M. and Cutland, N. (1992) 'Stochastic Navier-Stokes Equations', *Acta Applicanda Mathematicae* 25: 59-85.(Powerful use of Loeb measure and other NSA tools to solve a long-standing existence problem in applied mathematics.)

Cutland, N. (ed.) (1988) *Nonstandard Analysis and its Applications*, Cambridge: Cambridge University Press. (Collection of papers on NSA, including a wide range of applications to pure and applied mathematics. See especially S. Albeverio, 'Applications of Nonstandard Analysis in Mathematical Physics', and C.W. Henson, 'Infinitesimals in Functional Analysis'.)

Dauben, J.W. (1992) 'Revolutions Revisited', appendix to 'Conceptual Revolutions and the History of Mathematics: Two Studies in the Growth of Knowledge', in D. Gillies (ed.) *Revolutions in Mathematics*, Oxford: Clarendon Press.(Assessment of NSA as 'a contemporary revolution in mathematics'; mentioned in §5.)

Dauben, J.W. (1994) *Abraham Robinson: The Creation of Nonstandard Analysis, a Personal and Mathematical Odyssey*, Princeton, NJ: Princeton University Press.(An authoritative scientific biography.)

Henson, C.W. and Keisler, H.J. (1986) 'The Strength of Nonstandard Analysis', *Journal of Symbolic Logic* 51: 377-86.(Explores the strength of NSA compared to conventional analysis; mentioned in §2.)

Hewitt, E. (1948) 'Rings of Real-Valued Continuous Functions I', *Transactions of the American Mathematical Society* 64: 54-99.(Construction of a structure that, in retrospect, is seen to be essentially an ultrapower enlargement of $\Re$; mentioned in §3.)

Hirschfeld, J. (1990) 'The Nonstandard Treatment of Hilbert's Fifth Problem', *Transactions of the American Mathematical Society* 321: 379-400.(Great simplification compared to the original 1955 solution by D. Montgomery and L. Zippin; mentioned in §4.)

Keisler, H.J. (1976) *Elementary Calculus*, Boston, MA: Prindle, Weber & Schmidt.(A beginner's calculus course using NSA; mentioned in §5.)

Keisler, H.J. (1994) 'The Hyperreal Line', in P. Ehrlich (ed.) *Real Numbers, Generalizations of the Reals, and Theories of Continua*, Dordrecht: Kluwer.(A masterly survey of NSA as applied to the real line. Discusses the philosophical status of the reals. Mentioned in §2.)

Lindstrøm, T. (1988) 'An Invitation to Nonstandard Analysis', in N. Cutland (ed.) *Nonstandard Analysis and its Applications*, Cambridge: Cambridge University Press.(A very user-friendly introduction to NSA. Extensive bibliography up to the late 1980s.)

Loeb, P.A. (1975) 'Conversion from Nonstandard to Standard Measure and Applications in Probability Theory', *Transactions of the American Mathematical Society* 211: 113-22.(Construction of the Loeb measure. Pivotal advance in NSA. Mentioned in §4.)

Luxemburg, W.A.J. (1962) *Non-Standard Analysis: Lectures on A. Robinson's Theory of Infinitesimals and Infinitely Large Numbers*, Pasadena, CA: California Institute of Technology.(Early extensive presentation of NSA; uses ultrapower construction of enlargement but no formal logic; mentioned in §3.)

Machover, M. (1967) *Nonstandard Analysis Without Tears: An Easy Introduction to A. Robinson's Theory of Infinitesimals*, Technical Report no. 27, US Office of Naval Research, Jerusalem: The Hebrew University.(Introduces a set-theoretic framework for applying NSA to higher-order entities; mentioned in §3.)

Machover, M. (1993) 'The Place of Nonstandard Analysis in Mathematics and in Mathematics Teaching', *British Journal for the Philosophy of Science* 44: 205-12.(Expands on §5.)

Nelson, E. (1977) 'Internal Set Theory: A New Approach to Nonstandard Analysis', *Bulletin of the American Mathematical Society* 83: 1,165-98.(Elegant axiomatic framework for NSA; mentioned in §3.)

Nelson, E. (1987) *Radically Elementary Probability Theory*, Annals of Mathematics Studies 117, Princeton, NJ: Princeton University Press.(Tour de force of simplification: advanced probability theory brilliantly made easy by NSA; mentioned in §4.)

Robinson, A. (1961) 'Non-Standard Analysis', *Proceedings of the Royal Academy of Sciences, Amsterdam*, series A, 64: 432-40.(The first publication on NSA.)

Robinson, A. (1962) *Complex Function Theory over Non-Archimedean Fields*, Technical-Scientific Note no. 30, US Air Force Office of Scientific Research, Jerusalem: The Hebrew University.(NSA applied for the first time extensively to higher-order entities, using a type-theoretic framework, to prove new results on the zeros of complex polynomials and on the behaviour of an analytic function near an essential singularity; mentioned in §3.)

Robinson, A. (1966) *Non-Standard Analysis*, Amsterdam: North Holland; 2nd revised edn, 1974.(Uses a type-theoretic framework for higher-order entities; includes a wealth of applications; and presents the result on Hilbert spaces, mentioned in §4, proved with A.R. Bernstein. A historical chapter discusses in some detail topics touched upon in §1 and quotes extensively from Leibniz and others. Gödel's preface to the 1974 edition gives his assessment of NSA quoted in §5.)

Robinson, A. (1979) *Selected Papers of Abraham Robinson*, vol. 2, *Nonstandard Analysis and Philosophy*, ed. H.J. Keisler, W.A.J. Luxemburg and S. Körner, New Haven, CT: Yale University Press.(Includes Robinson's papers on NSA, including seminal applications in diverse fields.)

Russell, B.A.W. (1903) 'Philosophical Arguments concerning the Infinitesimal', in *The Principles of Mathematics*, London: Cambridge University Press, 2nd edn, 1937; repr. London: Routledge, 1992. (Quoted in §1.)

Schmieden, C. and Laugwitz, D. (1958) ‘Eine Erweiterung der Infinitesimalrechnung’ (An Extension of the Infinitesimal Calculus), *Mathematische Zeitschrift* 69: 1-39.(The last, partly successful, pre-NSA attempt to justify infinitesimals.)

Skolem, T. (1923) ‘Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre’, in *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, Helsinki: Akademiska Bokhandeln, 217-32; trans. S. Bauer-Mengelberg, ‘Some Remarks on Axiomatized Set Theory’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 291-301.(Proves the existence of nonstandard models of axiomatic set theory and hints at the existence of nonstandard models of arithmetic; mentioned in §1.)

Skolem, T. (1934) ‘Über die nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen’ (On the Non-Characterizability of the Number Sequence by Finitely or Denumerably Many Statements with Number Variables Only), *Fundamenta Mathematicae* 23: 150-61.(Existence of nonstandard models of arithmetic proved using a construction foreshadowing ultrapowers.)

Analysis, philosophical issues in

The term 'mathematical analysis' refers to the major branch of mathematics which is concerned with the theory of functions and includes the differential and integral calculus. Analysis and the calculus began as the study of curves, calculus being concerned with tangents to and areas under curves. The focus was shifted to functions following the insight, due to Leibniz and Isaac Newton in the second half of the seventeenth century, that a curve is the graph of a function. Algebraic foundations were proposed by Lagrange in the late eighteenth century; assuming that any function always took an expansion in a power series, he defined the derivatives from the coefficients of the terms. In the 1820s his assumption was refuted by Cauchy, who had already launched a fourth approach, like Newton's based on limits, but formulated much more carefully. It was refined further by Weierstrass, by means which helped to create set theory. Analysis also encompasses the theory of limits and of the convergence and divergence of infinite series; modern versions also use point set topology. It has taken various forms over the centuries, of which the older ones are still represented in some notations and terms. Philosophical issues include the status of infinitesimals, the place of logic in the articulation of proofs, types of definition, and the (non-) relationship to analytic proof methods.

1 The calculus: Newton, Leibniz and Lagrange

To begin in geometric terms, the differential calculus provides methods of determining the slope of a tangent to a curve, while the integral calculus finds the area between the curve and the *x*-axis. The study of tangents and areas goes back to ancient times, but prior to Isaac Newton and G.W. Leibniz tangents were determined at specific points and areas were always definite ones, that is, taken between a specific starting point and end point on the *x*-axis. Their new insight - treating a curve as being the graph of a function $f(x)$ - was to determine the tangents and areas as new *functions*. Thus the derivative function f' gives the rate of change of the function *f*, and the indefinite integral $\int f$ is a function showing how the area under *f* (taken from some arbitrary starting point) varies with the end point. The two calculi are inverse: the 'fundamental theorem' states that

(1) $(\int f)' = f$ and also $\int(f') = f$ (+ constant).

In Newton's version (created 1660s, published 1700s; see Newton (1967-81); Newton, I. §6) the differential was a 'fluxion', the rate of change of a variable quantity relative to time (understood conventionally); conversely, that variable was the 'fluent' of its fluxion. His definitions involved changes in dimension: a point as the limit of a sequence of lines of decreasing length, and so on. He defined his functions in terms of limits, but in a very naïve way.

Leibniz formed his differential and integral calculus in the 1670s (published 1680s; see Leibniz, G.W. §§1, 10) independently of Newton, and in a quite different form. It was based on the idea that a 'differential' d*x* could be created as an infinitesimal increment on *x*, with the same dimension as *x*; the slope of the tangent to the curve corresponding to $f(x)$ was $\mathrm{d}f(x)/\mathrm{d}x$. For linear variables *x* and $f(x)$ the 'integral' $\int f(x)\mathrm{d}x$ was the sum of the areas of rectangles $f(x)$ high and d*x* wide. Thus his version of the principal relationship (1) is

(2) $\mathrm{d}(\int f(x)\mathrm{d}x)/\mathrm{d}x = \int(\mathrm{d}f(x)/\mathrm{d}x)\mathrm{d}x = f(x).$

Another point of superiority of these theories over their predecessors was that the defined functions were themselves variable; thus they admitted of higher-order fluxions (differentials) and fluents (integrals). Physical phenomena of certain kinds could be expressed as 'differential' (or 'fluxional') equations of some order; finding their solution(s) became a major concern in mathematics, with many important applications in mechanics and geometry.

Of the two forms, Leibniz's gradually gained a considerable ascendancy during the eighteenth century. In one respect his use of the word 'function' was misleading, in that often no functional dependence obtained: variables could vary together. But during the 1750s Leonhard Euler realized the advantage of systematically specifying one variable (*x*, say) as a basis in a problem, relative to which the others were functionally dependent. His criterion was that d*x* be constant, so that further derivatives $\mathrm{dd}x, \ldots$ were all zero; then the differentials of other variables (*y*, say) were related to d*x* via 'differential coefficients' $p(x), q(x), \ldots$ of $f(x)$ with respect to *x*:

(3) $\mathrm{d}f(x) = p(x)\mathrm{d}x,$
$\mathrm{dd}f(x) = q(x)(\mathrm{d}x)^2, \ldots.$

Another major extension due then to Euler and others was to allow several variables in a problem to be independent; this move greatly increased the range of phenomena which could be represented, in ‘partial differential equations’.

During the late eighteenth century J.L. Lagrange proposed, as an alternative to limits and infinitesimals, that the differential calculus was obtainable from the so-called ‘Taylor series’, the expansion of a function *f* about any value of *x* with $h > 0$:

(4) $f(x+h) = f(x) + hf'(x) + h^2 f''(x)/2! + \ldots.$

(This important result had been named after Brook Taylor, although it had already been found by his master Newton and by Leibniz.) The ‘derived functions’ $f'(x),\ f''(x), \ldots$ (to use Lagrange’s name and notation for them) were *defined* as the coefficients of powers of *h* in the expansion (4); in his opinion they could be determined solely from (4) and algebraic manipulations. The integral calculus was based on reversing the operations; from f'' to f', and so on.

Lagrange’s approach became most widely circulated through the textbook *Théorie des fonctions analytiques* (1797), which was based upon his teaching at the recently founded École Polytechnique in Paris. He also extended his approach and associated with it an allied subject, the calculus of variations, which produced from a given function neighbouring functions which varied slightly from it. This theory, which became of major importance in applications, is also of philosophical interest, in that it requires possible worlds in which, for example, a particle may move along curves defined by functions varying slightly from the function whose curve it actually followed and which the theory was intended to determine.

Lagrange’s approach also came to have profitable influences on algebra in the early nineteenth century; in particular, some of its consequences provided powerful analogies for Boole and De Morgan in their contributions to algebraic logic. But the assumptions behind (4) seemed too good to be true; and in fact (4) was refuted by means of counter-examples, in the 1820s.

2 Mathematical analysis: Cauchy and Weierstrass

These torpedoes came from A.L. Cauchy, who had already begun to develop a fourth approach - ‘mathematical analysis’ (his name) - as part of his own teaching at the École Polytechnique. As with Newton’s calculus, the theory of limits served as his basis, but he treated them in a far more careful way, with general definitions and an exploration of properties. In addition, the theory of functions and the convergence of infinite series joined the calculus under limit theory. He raised the level of rigour in various ways, of which some involved points of logic. He gave independent definitions of the derivative and the integral:

$$f'(x) := \lim[(f(x+h) - f(x))/h] \text{ as } h \to 0;$$

$$\int f(x)\mathrm{d}x := \lim \sum_r [(x_r - x_{r-1})f(x_{r-1})], \text{ with}$$

$$x_0 \le x_{r-1} < x_r \text{ and } x_0 \le x \le \lim_{r\to\infty} x_r,$$

$$\text{as } \max_r |x_r - x_{r-1}| \to 0.$$

Cauchy retained Lagrange’s notation and name for $f'(x)$, but otherwise things were quite different. In particular, for the first time, (1) became the fundamental *theorem* of the calculus, concerning the inverse relationship between differentiation and integration under the sufficient conditions that $f(x)$ was continuous (in his sense); previously the switch between the two calculi had been more or less automatic. He showed the merit of systematically specifying necessary and/or sufficient conditions under which theorems were held to be true, and of enhancing theory by seeking weaker versions of the former and stronger versions of the latter. He even pioneered the methodical numbering of expressions in a paper or book.

Although Cauchy’s theory is not very intuitive (his students and colleagues hated it), it gradually became accepted,

especially after its adoption and refinement by Karl Weierstrass (at his widely influential lecture courses at Berlin from the 1860s) as the now standard ‘epsilon-delta’ treatment of limits. He stressed the need for extra care when several variables were moving towards their respective limits simultaneously, for example, in distinguishing between the following:

$$to \lim_{x\to a, r\to\infty} \sum_r u_r(x), (5) \lim_{x\to a} \lim_{r\to\infty} \sum_r u_r(x), \lim_{r\to\infty} \lim_{x\to a} \sum_r u_r(x),$$

where in the first case the two limits move together. Weierstrass and his followers also realized the importance of distinguishing the least upper bound of a collection of values from its upper limit. For example, the upper limit of the sequence 5, 7, -4, $\frac{3}{2}$, $\frac{5}{3}$, $\frac{7}{4}, \ldots (2n-1)/n, \ldots$ is 2, but the least upper bound is 7.

Among related developments, Georg Cantor was inspired from the early 1870s to create both his point set theory and transfinite arithmetic from solving certain technical questions in analysis by Weierstrassian methods. The further refinement of conditions on theorems such as (1) included the need to define irrational numbers in terms of rationals in order to prove theorems on their existence without begging the question posed. Cantor and Weierstrass had versions, but the best remembered is the ‘cut’ definition of Richard Dedekind in 1872. Associate each rational number with a point on the continuous straight line by natural order. Cut the line at any point, then the numbers are divided into two classes: to left (L) and to right (R). If L has no greatest member and R no smallest, then the cut does not correspond to any rational number, and thereby defines an irrational one (see Numbers §6).

The generally rising levels of rigour in proof - especially the increasing numbers of variables in problems, with ever more intricate functional relationships between them - inspired Giuseppe Peano in the late 1880s to create elements of ‘mathematical logic’ (his phrase), in order that the language itself of mathematical analysis would be made more rigorous, by relying less on prose. He formulated the propositional calculus in quite an axiomatic manner (not a common way of doing mathematics at that time), and introduced ‘propositional functions’ and quantification over their argument places. The latter technique was especially significant in clarifying the order of quantifiers for several variables; distinguishing the expressions in (5) is an important example.

The Weierstrassian tradition, together with Cantorian set theory, became standard fare in pure mathematics and some (but by no means all) applications by the beginning of the twentieth century. Cauchy’s theory of the integral was extended into the measure theory of Henri Lebesgue and others in the 1900s. At the same time various forms of the axiom of choice were found, revealing unexpected gaps and assumptions in proof methods; but gradually their place in analysis (and set theory) was evaluated and (for most figures) their non-constructive character was tolerated. Various generalizations were made: for example, defining the calculus over groups of some kind rather than over real numbers. In addition, the mathematical analysis of complex variables, which Cauchy had initiated along with his creation of the real-variable theory, rose to major status in mathematics - in his version and in alternatives due to Bernhard Riemann and to Weierstrass.

In one respect, reaction to Cauchy was negative. He had declared that divergent series were normally illegitimate for mathematical use; however, in the late nineteenth century, it became understood that the sum of a series could only be defined relative to a manner of summing its terms, and that the normal method (first term $+$ second term $+\ldots$) was not the only one. Thus a sum such as

(6) $1-1+1-1+\ldots=\frac{1}{2}$

was true under certain methods of summation, although the series had no sum under the usual method of adding term by term. This discovery rendered still more complicated the relationship between series and functions, which had already suffered a jolt from Cauchy’s refutation of Lagrange’s assumption (4).

3 Some philosophical issues

Several points of philosophical interest have already arisen. A few others will be made here.

Various types of infinitesimal had been used in the calculus and analysis over the centuries. The three main kinds, not always clearly distinguished, were: (a) the ‘infinitely small’, a variable quantity with limiting value zero (Cauchy proposed a new form of this kind); (b) ‘nulls’, such as the d*x* in (2), which are smaller than normal quantities; and (c) ‘pseudo-infinitesimals’ - quantities different from zero and yet smaller than any assignable

quantity (Berkeley's criticism of the foundations of the calculus in 1734 were based upon this view). They all became unacceptable in Weierstrass' reign; but even then they gained some adherents, and since the 1960s theories have been produced which vindicate their use. The best known is the 'nonstandard analysis' of Abraham Robinson (see Analysis, nonstandard).

One aspect which splits limit theory apart from infinitesimals is that of dimensions. For Leibniz, if *x* is a line, d*x* is a very short line, dd*x* is a very, very short line, and so on; thus dimensions are preserved under the operations of differentiation and integration. By contrast, a theory which uses limits changes dimensions; for example, the limit of a sequence of lines is a point. Intellectually this difference is vast, with profound consequences for applications; it is one source of the non-intuitiveness of mathematical analysis, and applied mathematicians have often preferred to work with differentials, which allow the geometry of a given problem, controlled by dimensions, to be represented and manipulated with incomparable clarity and flexibility.

Once Euler had introduced the differential coefficient (alongside infinitesimals) with (3) above, the notion of function was at centre stage. A major philosophical consideration was that of *generality* of the notion of function. Euler himself had said that variables *x* and *y* may be related 'in any way' by a function, but he had in mind only the kinds of functions normally used in his day. The same remark applies to successors in the nineteenth century, such as J.P.G. Dirichlet, though with a somewhat wider realm in mind. Only in the era of measure theory do we find the set-theoretic version which is now standard; that *x* and *y* are functionally related without assuming any details about the means. A striking example is the 'characteristic function of the rationals', a type of totally discontinuous function which takes one constant value when *x* is rational and a different one when *x* is irrational. Dirichlet had held it to be pathological, as it had no integral in any sense that he could imagine; by contrast, it played a central role in Lebesgue's formulation of measure theory.

It followed from Cauchy's theory of functions that a continuous function may be without a derivative at certain points. Riemann extended this aspect by producing an exotic range of functions with corners and/or discontinuities; and Weierstrass capped it in 1872, with an example of a continuous function which never had a derivative. Results of this kind raised new questions in the traditional discourse of continuity and discontinuity (see Intuitionism).

The axiomatic treatment of analysis is often based on the completeness axiom - every non-empty bounded set of real numbers has a least upper bound (equivalent to Dedekind's cut principle mentioned in §2 above). The completeness axiom is impredicative and therefore violates Russell's vicious circle principle (see Constructivism in mathematics).

Cauchy's rigour was based in part upon broad definitions of basic concepts; definabilty and modes of definitions remained an important topic. In particular, Peano treated them as a central part of his logico-analysis: rules for forming them correctly, and the role in theories of definition under hypothesis. In addition, Cauchy had insisted that a mathematical function always be single-valued (so that, for example, $\sqrt{x}$ would have to be split into $+\sqrt{x}$ and $-\sqrt{x}$); in expressing this feature Peano came in 1897 to formulate the criteria for the existence of a referent to a definite description in mathematics that Bertrand Russell was to propose in 1905 for natural language in his famous paper 'On Denoting' (see Descriptions; Russell, B.A.W. §9).

The long-suffering word 'analysis' was forced to fulfil still more duties by this subject. The traditional link with algebra was already becoming tenuous by the late eighteenth century, and was weakened further by Cauchy's refutation of (4) as a foundation. In addition, proofs in mathematical analysis from Cauchy onwards are usually synthetic, in that they start out from definitions of basic notions (continuity of a function, say) and/or proven results and finish with the theorem required.

Weierstrass introduced a methodology, with particular bearing upon mathematical analysis: examine all the exceptional and special cases, and do not be content with theorems about 'in general… '. Cantor was led to his set theory by obeying such a stricture (see Cantor, G. §1).

Do foundations have foundations of their own? Mathematical analysis shows beautiful examples to illustrate this question: Weierstrass underpinned the foundations which Cauchy had provided by refining single-limit theory into multiple-limit theory and the other aspects mentioned above, which Cauchy and his contemporaries had handled freely; but he and his generation took for granted making an infinite selection of members from a set, which was

recognized in the axiom of choice to be an additional question.

See also: Axiom of choice; Continuum hypothesis; Logical and mathematical terms, glossary of

I. GRATTAN-GUINNESS

References and further reading

Bos, H.J.M. (1974) 'Differentials, Higher-Order Differentials and the Derivative in the Leibnizian Calculus', *Archive for History of Exact Sciences* 14: 1-90.(On Leibniz and Euler.)

Bottazzini, U. (1986) *The Higher Calculus*, Heidelberg: Springer.(Covers both real and complex analysis, 1750-1900.)

Cohen, H. (1883) *Das Princip der Infinitesimal-methode und seine Geschichte (The Principle of the Infinitesimal Method and its History)*, Berlin: Dummler.(A fine historico-philosophical survey.)

Grattan-Guinness, I. (ed.) (1980) *From the Calculus to Set Theory, 1630-1910*, London: Duckworth. (Covers only real-variable analysis.)

Hardy, G.H. (1949) *Divergent Series*, Oxford: Clarendon Press.(Fine coverage, with historical notes.)

Hawkins, T.W. (1970) *Lebesgue's Theory of Integration: Its Origins and Development*, Madison, WI: University of Wisconsin Press; repr. New York: Chelsea, 1975.(On measure theory.)

Landau, L. (1930) *Grundlagen der Analysis (Foundations of Analysis)*, Leipzig: Teubner; repr. and trans. New York: Chelsea, 1951.(Classic introductory text.)

Newton, I. (1967-81) *The Mathematical Papers of Isaac Newton*, ed. D.T. Whiteside, Cambridge: Cambridge University Press, 8 vols.(Extensive selection of manuscripts and editorial commentary.)

Yushkevich, A.P. (1976) 'The Concept of Function up to the Middle of the 19th Century', *Archive for History of Exact Sciences* 16: 37-85.(On the development of functions and real variables.)

Analytic ethics

*Moral philosophy has traditionally been divided into normative ethics and meta-ethics. Normative ethics concerns judgments about what is good and how we should act. Meta-ethics, with which 'analytic ethics' is typically identified, seeks to understand such judgments. Are they factual statements capable of being literally true or false (*cognitivism*)? Or are they commands or expressions of attitude, capable only of greater or lesser appropriateness or efficacy (*noncognitivism*)? Cognitivists focus on whether the facts to which they claim moral judgments correspond are discovered from experience, or whether they occupy a different realm, as do mathematical facts. Noncognitivists, in contrast, arguing that moral judgments are not fact-stating, ask if they signal our feelings or commitments, or are imperatives of conduct.*

Other questions concerning moral judgments include whether they are subjective or objective, and how they are connected to motivation. Analytic ethics therefore not only concerns the meaning of moral terms, but ranges over such areas as epistemology, metaphysics and the theory of action. As a field it remains full of controversy. It has developed approaches that afford specific insights into morality, and contributed to our understanding of the functions of thought and language.

1 From intuitionism to noncognitivism

Prior to the twentieth century, questions about the nature of moral judgment were framed primarily in substantive terms: Which 'faculties' are involved? What sorts of acts or traits of character are morally good or bad, right or wrong? Some philosophers (such as Hutcheson and Hume) emphasized the role of sentiment, while others (such as Kant) stressed reason (see Kantian ethics; Moral sense theories). Some (Hume, Kant) took ordinary moral practice as a touchstone, while others (Bentham, Mill) attempted to defend an 'external standard' that could reform moral opinion.

In the twentieth century, G.E. Moore succeeded in founding a new tradition that came to insist upon a rigorous separation of questions of meaning from substantive questions. In *Principia Ethica* (1903), he introduced the 'open question' argument: for any purported analysis *A* of a moral concept in nonmoral terms (for example, 'Good' = 'Conducive to happiness'), consider the question, 'Yes, I see that *x* is *A* (for instance, *x* conduces to happiness), but is *x* good?'; if this question is intelligible, and not as trivial as, 'Is *x*, which is *A*, also *A*?', then even if we come to agree that *A* is good, *A* cannot simply be what we *mean* in calling something 'good'. Moore used this argument to advance his claim that 'good' names a *sui generis*, non-natural property, known by a kind of rational intuition (see Intuitionism in ethics). Moore also believed that this property supervenes on natural properties: any two things with just the same natural properties would necessarily be equally good (or bad) (see Supervenience). Natural properties were thus seen as 'good-making', but Moore thought it a 'naturalistic fallacy' to equate good with them.

Subsequent philosophers tended to agree that good cannot be analytically reduced to a natural property, but rejected Moore's rational intuition. Logical positivists, for example, divided cognitively significant propositions into two categories, the analytic, knowable a priori because tautological, and the synthetic, knowable a posteriori by empirical means (see Logical positivism §3; Analyticity). Intuitionist claims of synthetic a priori moral knowledge fit neither category: the open question argument showed that substantive moral theories are not analytic, but the lack of any empirical procedure for resolving fundamental moral disputes showed that they could not be synthetic either. Loyal to the positivist bifurcation, A.J. Ayer (1936) concluded that moral judgments expressed not cognitively significant propositions, but emotions (see Ayer, A.J.).

C.L. Stevenson (1944) endorsed this conclusion, though not for positivistic reasons, and moreover showed how emotivism could explain something intuitionism had made very mysterious - the seeming 'magnetism' or 'action-guidingness' of moral judgments. Intuitionists see moral judgments as descriptions of a special realm of abstract qualities, such as moral rightness. But description seems motivationally neutral. Why then is it odd for someone to say, 'This is the right thing to do, but I'm in no way for it'? (Contrast this with the remark, 'This is the tidiest thing to do, but I'm in no way for it.') Emotivists, on the other hand, see judgments of moral rightness as expressions of one's approval of a course of action, establishing a necessary or 'internal' connection to the speaker's motivations. Moral disagreement among speakers is seen as conflict in attitudes or commitments among speakers, given its special interest by the need for resolving the question, 'How should I (or we) act?' (see

Emotivism; Moral motivation §§1-2).

Emotivism is one form of noncognitivism, which interprets moral judgments as expressions or imperatives, not as descriptions of ‘objective reality’. Must noncognitivists therefore see moral judgments as ‘merely subjective’, immune to rational argument and liable to revision whenever the mood strikes? Stevenson, R.M. Hare (1952) and others have argued that nothing in noncognitivism prevented speakers from seeing their attitudes as based upon straightforwardly cognitive, objective considerations, so long as this relation is not seen as an analytic derivation (see Hare, R.M.). Perhaps it is part of what makes an attitude moral that one is committed to universalizing the judgments it supports or giving certain kinds of impersonal reasons in its defence. By the 1950s, noncognitivism’s combination of explanatory power with freedom from metaphysical baggage enabled it thoroughly to supplant intuitionism as the dominant position in analytic ethics.

2 Challenges to noncognitivism

Not everyone was convinced that noncognitivism could adequately accommodate the cognitive aspects of moral thought and language. Peter Geach (1965) pointed out that noncognitivists failed to show how moral claims, construed as essentially non-propositional, could display the full logic and grammar of propositions, as actual moral judgments unquestionably do. If, for example, ‘Stealing is wrong’ expresses one’s attitude of disapproval of stealing, then how are we to understand the conditional, ‘If stealing is wrong, then stealing without being caught is still wrong’, which could be asserted sincerely by someone who does not disapprove of stealing at all? Recent noncognitivists, notably Simon Blackburn and Allan Gibbard, have taken up Geach’s challenge, but it is fair to say that no fully satisfactory solution has yet been given.

Critics like Philippa Foot (1978) and Geoffrey Warnock followed a different strategy against noncognitivism, in part by noting that the open question argument has narrower scope than previously supposed. Understanding moral language, they argued, involves grasping certain contentful relations of relevance as well as knowing how to use its expressive force. Consider the remark, ‘This new petrochemical project is an ethical marvel; it will make our firm famous throughout the world! Of course, it will cause human dislocation and suffering - but what could that possibly have to do with ethics?’ Such a remark would betray a very imperfect understanding of moral language, even on the part of someone indifferent to human suffering who nevertheless cared unconditionally about the fame of his company. A plausible noncognitivism must at least allow that substantive norms of relevance govern moral discourse and that descriptive content can thereby accrue to moral language in a secondary way. But once substantive content has been admitted, the challenge can be pushed further. Why say that content is secondary and expressive force primary? Noncognitivists will point to the success of their ‘internalist’ account of the relation of moral judgment and motivation. But Foot challenges this. If a person calls their project an ethical marvel because it publicizes their company, we do not know what they are talking about. But someone who says, ‘Sure, a progressive tax on incomes would be more just, and that is fine if you care about justice for its own sake, but what is it to people like me?’, will strike us as more antipathetic than linguistically confused. Perhaps, as Foot and William Frankena have suggested, it would accord better with actual use to see the connection between moral judgment and motivation as a matter of what is typical rather than essential. Typically we care very much, individually and socially, about the kinds of things morality is concerned with - such as wellbeing and even-handedness - and we tend to dislike and distrust those who show no interest in such matters. Moreover, we tend to think they ought to show more interest. But we can understand this ‘ought’ as signalling the unconditional character of moral norms, without seeing it as marking a special realm of rational or motivational necessity. After all, we apply norms of etiquette even to the obdurately rude.

Critics have also questioned some of the contrasts upon which noncognitivism has been based. Some have followed Quine in challenging the logical positivist’s version of the analytic/synthetic distinction (see Quine, W.V. §§3, 8). Thomas Nagel (1970) and John McDowell (1985) have considered alternatives to the Humean thesis that cognitive states cannot suffice for action except when combined with desires. McDowell and David Wiggins (1987) have challenged us to rethink the subjective/objective distinction, by exploring a possible analogy between value and secondary qualities such as colour, which seem both to involve a human response and to be objective (see Secondary qualities). Hilary Putnam (1981) has argued that philosophy of science no longer supports the fact/value contrast as traditionally drawn, since rationally optional norms suffuse the process of scientific ‘fact-finding’ (see Putnam, H. §8).

At the same time, various philosophers have sought to develop a positive conception of the factual status of moral judgments. Some have done so ‘on the cheap’, by urging a minimalist conception of truth according to which a discourse is apt for truth or falsity merely if it displays the full grammar and logic of assertion, as moral discourse uncontroversially does. Others have retained a substantive notion of truth, but have claimed that moral statements can earn truth-evaluabilty in this more robust sense. For example, one group of nonreductionist moral realists, including Richard Boyd and Nicholas Sturgeon, have argued that moral predicates can meet naturalistic criteria of property-attribution by playing a role in the best explanatory theories of empirical science. Other philosophers who take naturalistic concerns seriously, such as Richard Brandt, Gilbert Harman and David Lewis, have sought to provide reductions or reforming definitions of moral notions in terms of idealizations of psychological properties. Yet another group, partly inspired by a non-naturalist Kantian tradition, have argued that the objectivity of moral judgment cannot be impugned by invidious comparisons with empirical science. Morality’s subject matter, they argue, is a realm of practical reasons - reasons for action, not belief - and whether this domain can be a source of truth and knowledge depends not upon vindicating an ontology of ‘moral facts’, but upon the existence of objective reasons for agents. Kurt Baier, Thomas Nagel and John Rawls, among others, have attempted to show how we might understand the possibility of such reasons.

Other critics of the dialectic set in motion by noncognitivism have taken a more radically questioning stance. Bernard Williams (1985), Alasdair MacIntyre and various feminist philosophers have in different ways asked where the demand for objective or rational validation in ethics comes from in the first place, and whether, on reflection, it has sufficient credibility or interest to sustain the project of analytic ethics or ‘moral theory’ (see Williams, B.A.O.; Feminist ethics).

3 Leading issues

Although noncognitivism has been widely criticized, no cognitivist explanation of moral phenomena has yet been developed with anything like the systematic power found in the work of Stevenson, Hare or Gibbard. Nor have those who call for an end to ethical theorizing managed to show how we could do without it. The main questions one might expect to figure in continuing debates include: In what sense is it a substantial achievement for a discourse to be truth-evaluable, or is this the modest and inevitable outcome of an assertoric grammar? If the former, then what does that achievement require? If the latter, then what notions, if any, of objectivity and factuality remain to distinguish among assertoric practices, and how does ethics stand with these? Is motivation necessarily related to making or being subject to moral judgment? Is a noncognitivist approach to moral discourse ultimately self-destabilizing, since it construes normativity in non-factualist terms and yet appears to presuppose for its own purposes facts of meaning, arguably a normative domain? Should historicist and feminist criticism of analytic ethics be understood as offering an alternative, or as showing how a more genuinely objective and worthwhile analysis of ethics might be possible?

See also: Moral judgment; Moral knowledge; Moral realism; Moral scepticism; Morality and ethics; Naturalism in ethics

PETER RAILTON

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz; 2nd edn, 1946, esp. ch. 6.(Ground-breaking discussion of the status of moral judgments within the logical empiricist programme.)

Brandt, R.B. (1979) *A Theory of the Good and the Right*, New York: Oxford University Press.(Systematic presentation of a contemporary naturalism based upon ‘reforming definitions’ and psychological theory.)

Darwall, S., Gibbard, A. and Railton, P. (1997) *Moral Discourse and Practice*, New York: Oxford University Press.(Anthology containing a selection of papers by authors discussed in this entry. Also contains an introductory overview of twentieth-century analytic ethics and bibliography.)

Foot, P. (1978) *Virtues and Vices*, Berkeley, CA: University of California Press.(A collection of essays that traces the evolution of Foot’s discontent with noncognitivist accounts of moral language, and with categorical conceptions of moral imperatives. Her views have since evolved yet further.)

Geach, P. (1965) ‘Assertion’, *Philosophical Review* 74: 449-65.(Seminal criticism of noncognitivism; Geach attributes the basic insight to Frege.)

Gibbard, A. (1990) *Wise Choices, Apt Feelings*, Cambridge, MA: Harvard University Press.(Powerful

presentation of a norm-expressivist analysis of our discourse about rationality, with extensions to ethics.)

Hare, R.M. (1952) *The Language of Morals*, Oxford: Oxford University Press.(The first of Hare's series of highly influential books developing and defending his 'non-descriptivist' interpretation of moral language - he rejects the term 'noncognitivism' because it fails to reflect how his account makes moral judgment reason-giving.)

McDowell. J. (1985) 'Values and Secondary Qualities', in T. Honderich (ed.) *Morality and Objectivity*, London: Routledge & Kegan Paul.(Influential revival of the analogy between values and secondary qualities, such as colour.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press. (Original presentation of the 'open question' argument; set the terms for much of twentieth-century analytic ethics.)

Nagel, T. (1970) *The Possibility of Altruism*, Oxford: Clarendon Press.(Powerful revival of the Kantian idea that moral objectivity consists in rational requirements on action.)

Pojman, L. (ed.) (1989) *Ethical Theory: Classical and Contemporary Readings*, Belmont, CA: Wadsworth. (Useful anthology containing many influential writings in analytic ethics.)

Putnam, H. (1981) *Reason, Truth, and History*, New York: Cambridge University Press.(Critical re-examination of orthodox assumptions about the fact/value distinction and the notion of rationality.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Highly influential discussion not only of the concept of justice but also of coherence and contractarian methods of justification in ethics.)

Sayre-McCord, G. (ed.) (1988) *Moral Realism*, Ithaca, NY: Cornell University Press.(Anthology containing relevant essays by Boyd, McDowell, Sturgeon and others.)

Stevenson, C.L. (1944) *Ethics and Language*, New Haven, CT: Yale University Press.(Seminal work in developing noncognitivism.)

Wiggins, D. (1987) *Needs, Values, Truth*, Oxford: Blackwell.(Contains several important essays discussing truth in ethics and the possibility of a sensible, cognitivist subjectivism.)

Williams, B. (1985) *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press and London: Fontana.(Influential criticism of various trends in moral theory.)

Analytical philosophy

Philosophical analysis is a method of inquiry in which one seeks to assess complex systems of thought by 'analysing' them into simpler elements whose relationships are thereby brought into focus. This method has a long history, but became especially prominent at the start of the twentieth century and, by becoming integrated into Russell's development of logical theory, acquired a greater degree of sophistication than before. The logical positivists developed the method further during the 1930s and, in the context of their anti-metaphysical programme, held that analysis was the only legitimate philosophical inquiry. Thus for them philosophy could only be 'analytical philosophy'.

After 1945 those philosophers who wanted to expand philosophical inquiries beyond the limits prescribed by the positivists extended the understanding of analysis to include accounts of the general structures of language and thought without the earlier commitment to the identification of 'simple' elements of thought. Hence there developed a more relaxed conception of 'linguistic analysis' and the understanding of 'analytical philosophy' was modified in such a way that a critical concern with language and meaning was taken to be central to it, leading, indeed, to a retrospective re-evaluation of the role of Frege as a founder of analytical philosophy. At the same time, however, Quine propounded influential arguments which suggest that methods of analysis can have no deep significance because there is no determinate structure to systems of thought or language for the analytical philosopher to analyse and assess. Hence some contemporary philosophers proclaim that we have now reached 'the end of analytical philosophy'. But others, who find Quine's arguments unpersuasive, hold that analytical philosophy has virtues quite sufficient to ensure it a role as a central philosophical method for the foreseeable future.

1 The method of analysis

The term 'analysis' has its origins in a classical Greek word denoting the activity of taking something apart; and the thought that such an activity might be a model for explanations of complex structures by reference to their simpler parts is itself a Greek thought, exemplified by Socrates' dream in Plato's *Theaetetus* (see Plato §15). In the early modern period, the idea of analysis reappears in Descartes' injunction that one should identify the 'simple natures' characteristic of one's subject-matter, which Arnauld explicitly describes in *La logique, ou l'art de penser (The Art of Thinking)* (1662) as the adoption of a 'method of analysis' (see Arnauld, A.; Descartes, R.). This method then can be found in much of the philosophy of the period, as in Locke's account of 'complex ideas' in his *Essay Concerning Human Understanding* (1689), which is conducted in terms of an analysis of them into their constituent 'simple ideas' (see Locke, J. §4).

Kant's 'Transcendental Analytic' in his *Critique of Pure Reason* moves away from the Cartesian analysis of 'ideas' to the analysis of our capacities for understanding and judgment (see Kant, I. §§6-7); and as the idealist tradition develops, especially in the work of Hegel, there is a further shift away from the method of analysis to the 'method of dialectic'. Hence one aspect of the break with idealism initiated by G.E Moore (§2) is a call for a return to the method of analysis: thus in the course of arguing against idealist accounts of judgment, he maintained that 'a thing becomes intelligible first when it is analysed into its constituent concepts' (Moore 1899: 182). As Russell always acknowledged, it was this Moorean conception of analysis which initially inspired his own analytical programme, although in also accepting that there is a sense in which 'analysis is falsification' (Russell 1903: 141), he equally recognized its limitations.

Although Russell had a decisive role in the emergence of the self-consciously 'analytical' conception of philosophy, many other philosophers at the end of the nineteenth century sought to return to a method of analysis. Brentano's approach to psychology was explicitly analytical, and there is a direct route from Brentano's analytical psychology to Husserl's programme of phenomenological analysis (see Brentano, F.; Husserl, E.; Phenomenology, epistemic issues in). Similarly, from among the American pragmatists, C.S. Peirce wrote that that 'the only thing I have striven to do in philosophy has been to analyse sundry concepts with exactitude' (Passmore [1957] 1968: 104).

2 From philosophical analysis to analytical philosophy

The core thought behind this conception of analysis is that of the explanation of a whole by reference to its parts. But since philosophical analysis involves no physical decomposition of a whole into its parts, it needs to be made clear how this talk of ‘analysis’ makes sense.

In the case of logical analysis the basic idea is that one can explain the inferential significance of a statement through an account of its ‘logical form’, which, by identifying the presence of certain simple ‘logical constants’ in the statement, enables one to fit it into a general logical theory that shows how to argue for it and from it. In Russell’s writings the scope of this idea was much extended by allowing that the logical analysis of a statement can lead one to an account of its logical form which discerns the presence of logical constants that were certainly not apparent in the surface structure of the statement. His theory of descriptions, which assigns to the statement ‘The present King of France is bald’ the logical form ‘ $(\exists x)(Fx \mathbin{\&} (\forall y)(Fy \rightarrow y = x) \mathbin{\&} Bx)$’ is a classic case of a logical analysis of this kind (see Descriptions). Russell further enhanced the significance of logical analysis by taking it that a statement’s logical analysis revealed the ‘constituents’ of the proposition expressed by the statement. Thus he took it that his theory of descriptions showed that it was a mistake to regard the phrase ‘The present King of France’ as identifying a constituent of the propositions expressed by statements in which it occurs, even when they are true, since the description is eliminated by the logical analysis (see Logical constants; Logical form).

In the case of epistemological analysis, the guiding assumption is that complex claims to knowledge are justified by reference to simpler items of evidence, typically observations concerning which a high degree of certainty can be accepted. Hence the classic context for epistemological analysis is the empiricist theory that all evidence is, in one way or another, perceptual evidence and many empiricists have held that, in so far as beliefs about the world are warranted at all, it should be possible to provide an epistemological analysis of them which shows just how they can be supported by perceptual evidence (see Empiricism).

Both logical and epistemological analyses are normative, and for this reason can also be subversive. Russell develops his theory of descriptions into a theory of ‘logical fictions’ which implies that our ontological commitments are less extensive than we might at first suppose. Likewise, epistemological analyses imply that where a putative belief (a belief in immortality, perhaps) is one for which analysis reveals no satisfactory evidence, the credentials of the belief are called into question. This normativity is an aspect of all properly philosophical analyses: phenomenological analyses, properly understood, are not just introspective descriptions of appearances; they are supposed to elucidate the priorities within different modes of consciousness.

I shall not attempt to characterize here all the different kinds of analysis that are properly philosophical. For ‘analytical philosophy’ can be provisionally defined by the priority it assigns to logical and epistemological analysis. I have already mentioned the work of Moore and Russell, though neither of them held that philosophy is just analysis. Their most famous pupil was Wittgenstein, whose *Tractatus Logico-Philosophicus* (1922) is a paradigmatic exercise in logical analysis, resting on the assumption that ‘A proposition has one and only one complete analysis’ (1922: 3.25) (see Logical atomism). A little later, the members of the Vienna Circle sought to develop the analyses of Russell and Wittgenstein within the context of their positivist programme (see Vienna Circle). Despite their many disagreements, one common feature of their programme was the belief that ‘what is left over for *philosophy*… is only a *method*: the method of logical analysis’ (Carnap 1932: 77). They held that the only proper task for the philosopher is to engage in logico-epistemological analysis which clarifies the sense of questions about the world in such a way that they can be answered on the basis of scientific observation and experiment. It is, then, in the explicitly anti-metaphysical context of logical positivism that there occurs the transition from ‘philosophical analysis’, conceived of as *an* important method of inquiry, to ‘analytical philosophy’, which restricts genuine philosophy to analysis (as in Bergmann (1945: 194), which is, I think, the first explicit use of the term ‘analytical philosophy’).

3 Linguistic analysis

In the early post-1945 period this positivist conception of analytical philosophy was regarded as unnecessarily restrictive by many philosophers, especially those not enamoured of the scientific context of the positivist programme. Not wanting to return directly to old-fashioned metaphysics, they sought instead to extend the range of analytical techniques to cover a general concern with the normative aspects of language, while shedding a commitment to the identification of simple meanings or basic certainties. In Wittgenstein’s writings of this period

the underlying hope was the 'therapeutic' aspiration that philosophical perplexities would thereby be set to rest; whereas in the writings of Ryle, Austin, Strawson and other 'ordinary language' philosophers there was still the presumption that answers to old metaphysical questions such as those concerning the status of the mind (Ryle), appearances (Austin) and universals (Strawson) could be extracted from these inquiries (see Ordinary language philosophy, school of).

This broader conception of 'analytical philosophy' came to self-consciousness in the 1961 Anglo-French conference at Royaumont (Montefiore and Williams 1966), and it is, I think, only after this conference that use of the term 'analytical philosophy' became widespread (as in the title of the characteristic collections *Analytical Philosophy*, edited by R.J. Butler - Butler 1962, 1965). The sense of the term was, therefore, no longer confined to the logico-epistemological analyses of Russell and the logical positivists; instead it embraced a much broader critical concern with language, still resting on the presumption that this concern could somehow be put to work to solve, or dissolve, important philosophical problems.

This conception of analytical philosophy has been further refined by Michael Dummett, who has maintained that its distinctive feature is the priority it assigns within philosophy to the philosophy of language and that the founder of analytical philosophy, so conceived, is not Russell, but Frege (Dummett 1993). Both these claims, however, need some qualification.

As Dummett acknowledges, his emphasis on Frege's work is in part a retrospective re-evaluation. For although familiarity with Frege's work in logic and mathematics played an essential part in the development of Russell's work, Russell took his basic philosophical assumptions from Moore. Doubtless it would have been better if Russell had appreciated the significance of Frege's sense/reference distinction, but in fact he famously thought that his theory of descriptions rendered it unnecessary (Russell 1905). However, Rudolf Carnap - who, as we have seen, is one of the first to propound the thesis that philosophy can only be analytical - explicitly acknowledges his indebtedness to Frege, whose lectures he had attended (Carnap 1937: xvi); and it was, therefore, through Carnap's writings that Frege's philosophy of language was brought into the mainstream of analytical philosophy. Carnap also here acknowledges his debt to the great Polish logicians Ajdukiewicz, Leśniewski, Łukasiewicz and Tarski, whose works constitute in their own right major contributions to analytical philosophy, though ones that there is no space to describe here in detail (a full account should also include an account of the work of Hägerström and the other members of the early Swedish school of analytical philosophers) (see Polish logic; Scandinavia, philosophy in).

Although Dummett's claim that the distinctive feature of analytical philosophy is the priority it assigns to the philosophy of language fits well the conception of analytical philosophy that emerged from the early work of Wittgenstein and Carnap, it is certainly not a feature of Russell's early work nor, in fact, a self-conscious feature of Frege's writings, which famously include instead the thesis that sentences are true only in virtue of the nonlinguistic thoughts they express. Further, although Dummett's explanation of this priority in terms of the basic role within philosophy of a 'theory of meaning' fits his own writings and some of those of Donald Davidson, one has only to think of the hostility to 'theorizing' that is characteristic of Wittgenstein's later writings to see that Dummett's account is distinctly tendentious. Indeed, recent work in the philosophy of language and of mind increasingly calls into doubt Dummett's priority thesis (Stalnaker 1984) (see Evans, G.). Yet since those who argue for the alternative priority of mind over language employ the methods of logical and epistemological analysis characteristic of previous analytical philosophers - though applied now to the structure of mental, rather than just linguistic, representations - it would be quite inappropriate to locate here a movement that threatens to bring about the 'end of analytical philosophy'. That threat certainly exists, but it comes from another direction.

4 The end of analytical philosophy?

Just when analytical philosophy was coming to self-consciousness in the early post-war period, two of the assumptions of traditional methods of logical and epistemological analysis were called into question by W.V. Quine: (1) the assumption that there is a clear distinction (the 'analytic/synthetic' distinction) between logic and other disciplines which enables one to conduct the logical analysis of a statement without reference to these other disciplines; (2) the assumption that there is a single chain of justification from observation to more speculative claims about the world which enables one to construct an epistemological analysis of the latter in terms of the former (Quine 1953). In both cases Quine argued that in fact we find only a complex network of interdependent

relationships which undermines the prospect of establishing definitive logical or epistemological analyses. Further, he argued later (Quine 1960), our understanding of each other, and in particular of each other's utterances, is generally underdetermined by our observations of each other, and, as a consequence, the meaning of most of our utterances must be intrinsically indeterminate since there cannot be anything more to their meaning than is available to an intelligent observer (see Analyticity; Radical translation and radical interpretation).

Despite the fact that Quine's writings employ the standard techniques of logical analysis, some hold that his conclusions signal the end of analytical philosophy (Rorty (1980) gives an influential statement of this thesis). In thinking about this, however, one must bear in mind that already by 1945 most analytical philosophers had abandoned any commitment to simple meanings and basic certainties, and that the positivist thesis that philosophy could only be analytical philosophy was also soon rejected. Instead, the practice of analytical philosophers rested only on the assumption that methods of analysis can clarify conceptual and epistemological relationships in a way which contributes to the resolution or dissipation of philosophical problems. Do Quine's claims show that this assumption is unfounded? His indeterminacy thesis implies that the results of analytic inquiries can only be relative to one of an indefinite number of alternative systems of 'analytical hypotheses' between which there is nothing to choose (Quine 1960: 68), and thus that little of intrinsic significance can be detached from them. But Quine's arguments for the indeterminacy thesis are problematic since he allows only a narrow, behaviourist characterization of the evidence that is admissible to determine questions about the meaning of utterances. In order to question this restriction one need only invoke his own criticisms of 'foundationalist' approaches to questions of justification which imply that no such restriction on admissable evidence is legitimate. Quine's earlier arguments, by contrast, point to the holistic structure of our language and beliefs, and few analytical philosophers would want to dispute this (though there are exceptions, such as Dummett). But they will argue that indeterminacy does not follow from holism; on the contrary, talk of holism implies a system of normative relationships that is itself a legitimate subject for analytical inquiry. Admittedly, if there is no absolute analytic/synthetic distinction, the significance of these inquiries must itself remain open to question in the light of the broader theoretical concerns that influence one's choice of logic and epistemology. But this, so far from being the end of analytical philosophy, just reiterates the anti-positivist thesis that philosophy is not just analysis.

Analytical philosophy can, therefore, survive the contemporary prophets of doom by retreating to the pre-positivist thesis of the merits of philosophical analysis as an ingredient of philosophical inquiry, with a core commitment to the *explicit* articulation of the normative relationships, involving inference and justification, that connect concepts, beliefs and statements. This may seem too minimal a commitment to be worth a distinctive characterization as a kind of philosophy, but one has only to sample contemporary philosophical writings that do not share this commitment to recognize its value. To say this is not to say, however, that the methods of philosophical analysis are equally valuable in all areas of philosophy. There are those, such as Bernard Williams, who hold that in ethics and aesthetics the characteristic virtues of philosophical analysis may actually obstruct the proper exercise of a philosophical inquiry that aims to be truthful or imaginative (Williams 1995).

None the less, the robustness of analytical philosophy thus understood is best seen in the remarkable expansion in the acceptance and use of its methods that has taken place during the last half-century. This expansion has been both geographical and disciplinary. There has been an explosion of interest in analytical philosophy within non-anglophone countries, both in Europe and elsewhere (for instance, South America), leading to new dialogues between philosophers who had previously remained isolated in their own traditions. At the same time the ideas and methods of analytical philosophy have been applied to advance debates in areas of philosophy which had previously appeared remote from analytic concerns, such as the study of ancient philosophy (see Owen, G.E.L.; Vlastos, G.) and Marxism (Cohen 1978). By putting down roots around the globe and throughout the academy, analytical philosophy has shown that it is far too early to write its obituary (see Analytic ethics; Analytical philosophy in Latin America).

See also: Logical positivism

THOMAS BALDWIN

References and further reading

Analysis (1933-)(A journal founded to promote analytical philosophy, whose style remains a characteristic expression of this kind of philosophy.)

Arnauld A. (1662) *La logique, ou l'art de penser*, Paris; trans. J. Dickoff and P. James, *The Art of Thinking*, Indianapolis, IN: Bobbs-Merrill, 1964.(In chapter 4 Arnauld commends Descartes' method of inquiry as the 'method of analysis'.)

Austin, J.L. (1962) *Sense and Sensibilia*, ed. G. Warnock, Oxford: Oxford University Press.(Austin uses his analysis of the language of perception to criticize sense-datum theories of perception.)

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz.(A classic statement of logical positivism: chapter 2 proclaims that philosophy can only be analysis).

Ayer, A.J. (ed.) (1959) *Logical Positivism*, Glencoe, IL: Free Press.(A useful collection of writings from the logical positivists.)

Baldwin, T.R. (1990) *G.E. Moore*, London: Routledge.(Chapters 1 and 2 discuss Moore's break with idealism and early conception of analysis; in chapter 7 his later conception of philosophical analysis is discussed critically.)

Bergmann, G. (1945) 'A Positivistic Metaphysics of Consciousness', *Mind* new series 54: 193-226.(While criticizing the logical positivist account of the intentionality of consciousness, Bergmann describes the positivists as 'analytical philosophers'.)

Butler, R.J. (ed.) (1962, 1965) *Analytical Philosophy*, Oxford: Blackwell, 2 vols.(A characteristic expression of the kind of linguistic analysis that was practised in Oxford in the early 1960s.)

Carnap, R. (1932) 'Überwindung der Metaphysik durch logische Analyse der Sprache', *Erkenntnis* 2: 219-41; trans. A. Pap, 'The Elimination of Metaphysics through Logical Analysis of Language', in Ayer 1959, 60-81. (Carnap's statement of the logical positivist thesis that philosophy can only be the logical analysis of language.)

Carnap, R. (1937) *The Logical Syntax of Language*, trans. A. Smeaton, London: Routledge & Kegan Paul. (Carnap's further development of his logical positivist programme for the analysis of language.)

Cohen, G.A. (1978) *Karl Marx's Theory of History: A Defence*, Oxford: Clarendon Press.(An application of the methods of analytical philosophy to the study of Marxism; the starting-point of 'Analytical Marxism'.)

Descartes, R. (1701) *Rules for the Direction of the Mind*, Amsterdam: P. & J. Blaeu; trans. J. Cottingham, R. Stoothof and D. Murdoch, in *The Philosophical Writings of Descartes*, vol. 1, Cambridge: Cambridge University Press, 1985.(Rules 6 and 12 prescribe the discerning intuition of 'simple natures'.)

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press. (A collection of Davidson's papers in which he develops his conception of a 'theory of meaning' and its place in philosophy.)

Davidson, D. (1986) 'A Nice Derangement of Epitaphs', in E. LePore (ed.) *Truth and Interpretation*, Oxford: Blackwell.(A paper in which Davidson substantially qualifies the conception of a theory of meaning advanced in Davidson 1984.)

Dummett, M.A.E. (1993) *The Origins of Analytical Philosophy*, London: Duckworth.(Dummett's most extended discussion of his theses that the mark of analytical philosophy is the priority it assigns to the philosophy of language, and that Frege is the founder of analytical philosophy so conceived.)

Dummett, M.A.E. (1976) 'What is a Theory of Meaning? II' in G. Evans and J. McDowell (eds) *Truth and Meaning*, Oxford: Clarendon Press, 67-115.(Dummett's fullest statement of his conception of a 'theory of meaning' which, for him, is the core of a philosophy of language.)

Evans, G. (1982) *The Varieties of Reference*, Oxford: Clarendon Press.(Evans here argues that the theory of reference for language needs to be situated within a broader understanding of mental content, thus challenging Dummett's 'priority of language' thesis.)

Frege, G. (1892) 'Über Sinn und Bedeutung', Zeitschrift fur Philosophie und philosophische Kritik 100: 25-50; trans. M. Black as 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. M. Black and P. Geach, Blackwell: Oxford, 1952.(Frege's classic statement of his sense/reference distinction which has become a central feature of analytical philosophy.)

Gibson R. (1988) *Enlightened Empiricism*, Tampa, FL: University of Southern Florida Press.(A robust defence of Quine's sceptical arguments.)

Hacking, I. (1975) *Why Does Language Matter to Philosophy?*, Cambridge: Cambridge University Press.(A discussion of the idea of linguistic analysis and its development in the period 1950-75; the 1992 edition of Rorty (1967) contains as an afterword a critical discussion ('Ten Years After') of this book by Rorty.)

Hylton, P. (1990) *Russell, Idealism and the Emergence of Analytical Philosophy*, Oxford: Clarendon Press.(An account of the development of Russell's philosophy over the period 1900-12.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 2nd edn, 1933.(The standard translation of the *Critique*.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P. Nidditch, Oxford: Oxford University Press, 1975.(A classic application of the method of analysis in the context of an empiricist theory of ideas.)

Montefiore, A. and Williams, B. (1966) *British Analytical Philosophy*, London: Routledge.(An influential collection of papers in which the post-war Oxford conception of linguistic analysis is propounded and discussed.)

Moore G.E. (1899) 'The Nature of Judgment', *Mind* new series 8: 176-93.(Moore's critique of the idealist theory of judgment, in which he propounds his analytic programme.)

Moore G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press.(Moore's presentation of analytical ethics, resting on the thesis that the concept of intrinsic value is 'unanalysable'.)

Moore G.E. (1925) *'A Defence of Common Sense'*, in J.H. Muirhead (ed.) *Contemporary British Philosophy* (second series), London: Allen & Unwin, 193-223; repr. in T. Baldwin (ed.) *G.E. Moore: Selected Writings*, London: Routledge, 1993, 106-33.(Contains Moore's later view that the difficult task for philosophy is not to defend common sense, but to analyse it.)

Passmore, J. (1957) *A Hundred Years of Philosophy*, London: Duckworth; 2nd edn repr. Harmondsworth: Penguin, 1968.(Much of the second half of the book provides a survey of the development of analytical philosophy, which is carried further in the 1966 Penguin edition.)

Plato (*c*.380-367 BC) *Theaetetus*, trans. J.H. McDowell, Oxford: Clarendon Press, 1973.(Socrates' dream (201d-202c) marks the first appearance of the method of analysis.)

Quine, W.V. (1953) *From a Logical Point of View*, Cambridge, MA: Harvard University Press.(A collection of Quine's early papers including 'Two Dogmas of Empiricism' in which he launches his radical critique of the analytic/synthetic distinction.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(In chapter 2 Quine launches his notorious thesis of the indeterminacy of translation.)

Quine, W.V. (1990) *Pursuit of Truth*, Cambridge, MA: Harvard University Press.(A lucid recent statement by Quine of his views about meaning and evidence.)

Rorty, R.M. (1967) *The Linguistic Turn*, Chicago, IL: University of Chicago Press.(A collection of essays, expository and critical, on philosophy as 'linguistic analysis' with an excellent bibliography; the 1992 edition has two retrospective essays by Rorty in which he looks back critically at the pretensions of linguistic analysis.)

Rorty, R.M. (1980) *Philosophy and the Mirror of Nature*, Princeton, NJ: Princeton University Press.(In chapter 6 Rorty argues that the time has come to move beyond analytical philosophy, and in chapter 8 he describes post-analytical philosophy as the 'conversation of mankind'.)

Russell, B.A.W. (1903) *The Principles of Mathematics*, London: Allen & Unwin(Russell's first systematic attempt at a programme of logical analysis.)

Russell, B.A.W. (1905) 'On Denoting', *Mind* 14: 479-93; repr. in *Logic and Knowledge*, ed. R. Marsh, London: Allen & Unwin, 1956, 41-56. (Russell's presentation of his theory of descriptions.)

Russell, B.A.W. (1921) *The Analysis of Mind*, London: Allen & Unwin.(A characteristic expression of Russell's mature analytic programme in philosophy, especially notable for the priority given to the philosophy of mind over the philosophy of language.)

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.(Ryle uses the techniques of linguistic analysis to criticize Descartes' 'Myth of the Mind'.)

Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press.(Chapters 1 and 2 are an excellent critical discussion of the 'priority of language' thesis.)

Strawson, P.F. (1959) *Individuals*, London: Methuen.(An account of universals and particulars by reference to an account of the roles of subject and predicate in language; but other chapters, especially chapter 2 on 'Sounds', show Strawsonventuring well beyond analytic inquiries.)

Urmson, J.O. (1956) *Philosophical Analysis*, Oxford: Clarendon Press.(A description of the development of twentieth-century philosophical analysis from the perspective of Oxford philosophy of the 1950s.)

Williams, B.A.O. (1995) 'Contemporary Philosophy: A Second Look', in N. Bunnin and E.P. Tsui-James (eds) *The Blackwell Companion to Philosophy*, Oxford: Blackwell, 25-37.(A sceptical discussion of the limited virtues of the methods of analytical philosophy especially as applied within ethics.)

Wittgenstein, L.J.J. (1922) *Tractatus Logico-Philosophicus*, trans. C.K. Ogden and F.P. Ramsey, London: Routledge & Kegan Paul.(Wittgenstein's classic statement of the logical atomist position.)

Analytical philosophy in Latin America

In Latin America, philosophical analysis has been portrayed as an intellectual revolution. Its avowed goal has been to replace the abstruseness and obscurantism of scholastic and metaphysical jargon perceived as typical of much of Latin American philosophy with the clarity and rigour of mathematical and scientific discourse. Arriving around the mid-1940s analytic philosophy at first met with little interest because of a shortage of its classics in translation, cultural obstacles and opposition from the more traditional, entrenched philosophies. By the 1960s it had overcome many obstacles and stimulated considerable philosophical activity, mainly in Mexico and Argentina. By the 1980s, in spite of political opposition, analysis had created an international forum for the discussion of philosophical problems. It attracted to its ranks several distinguished philosophers and scientists with philosophical interests, among them Mario Bunge (Argentina and Canada), Héctor-Neri Castañeda (Guatemala and the USA) and Francisco Miró Quesada (Peru). An ambitious translation effort was launched and several important journals were founded, such as Análisis filosófico (Philosophical Analysis), Revista latinoamericana de filosofía (Latin American Journal of Philosophy) (Argentina), Crítica. Revista hispanoamericana de filosofía (Criticism. Hispanoamerican Journal of Philosophy) (Mexico), Manuscrito (Manuscript) (Brazil) and Diálogos (Dialogues) (Puerto Rico).

1 Early development

Broadly understood, philosophical analysis is a collection of methods for uncovering the presuppositions, implicit meanings and logical implications of statements in philosophical, scientific, or ordinary language. Some Latin American contemporaries, such as Rabossi (1977), have portrayed analysis as a set of common concerns, an attitude of caution towards metaphysics, a demand for logical rigour and conceptual clarity and an interest in scientific and linguistic problems. Others, such as Salmerón (1992) considered it to be a collective task of philosophical reconstruction which leaves little space for local or personal peculiarities and makes international life and communications easier. Still others viewed it as 'an intellectual revolution… It's object is relatively clear: to steer philosophy in the direction of a coherent and scientific investigation' (Nuño 1965). The inspiration for analysis was the historical movement that emerged from the philosophical activities of Frege, Russell,Moore and the Vienna Circle (see Ordinary language philosophy, school of; Vienna Circle).

In Latin America, analysis did not receive much attention until the mid-1940s. A lack of translations accounts for this late start. Prior to 1945 only two works on analysis were available in Spanish: *Ethics* (Moore 1912) and *Problems of Philosophy* (Russell 1912), both of which were published in Buenos Aires in 1929, with a Portuguese edition of the latter at São Paulo in 1939. From 1945 a stream of translations began to appear of the work of such philosophers as Wittgenstein, Carnap, Ayer and Max Black. Another reason for the late start of analysis was a lack of awareness in Latin America of modern logical techniques so important to analytic work. Up to the 1940s most logic studies were about syllogisms (Romero and Pucciarelli 1939), or fallacies (Vaz Ferreira 1910). After the 1940s the modern logic texts of Ferreira da Silva (1940), Quine (1944) and Miró Quesada (1963) stand out. Later, the texts of Ferrater Mora and Leblanc (1955) and Hegenberg (1966) were prominent.

Other factors impeded the diffusion of analysis. Links to logical positivism brought to mind the outdated views of Comte, Spencer and the nineteenth-century Latin American positivists (see Anti-positivist thought in Latin America). The popularity of opposing approaches, such as well-entrenched idealism, Marxism, with its careful socioeconomic analysis of reality, or neo-Thomism, with its rigorous criticisms of science and modern formalism, hindered the growth of analysis. As a further hindrance, Gaos and other refugees from the Spanish Civil War brought with them well-articulated forms of existentialism and phenomenology. Several other Latin American philosophers expressed reservations about analysis, including Caso (1939; 1941) and Bunge (1944), both of whom were critical of logical positivism, and Frondizi (1945).

Some early support for analysis came from personal contacts with visiting scholars, such as Quine, Strawson and von Wright and from local philosophers who had direct acquaintance with logical positivists, such as Schajowicz (1964) and Lindemann (1946). It was not until the 1950s that analysis began to flourish, particularly in Mexico and Argentina.

2 Mexico

In Mexico, the earliest discussion of analysis can be found in Caso's works (1939; 1941). Caso, an opponent of nineteenth-century positivism, formulated a phenomenalist critique of what he called neopositivism. Philosophical analysis remained taboo until it was defended by Molina Flores (1954). However, it took six years for the Centro de Estudios Filosóficos, renamed in 1964 the Instituto de Investigaciones Filosóficas, to publish his translations of Moore (1912), Carnap and Ayer in its series *Cuadernos*.

With a growing body of translations and an increase in interest, the centre became a hub of analytic work and sponsored the journal *Crítica. Revista hispanoamericana de filosofía (Criticism. Hispanoamerican Journal of Philosophy)*, under the editorial supervision of Villoro, Salmerón and Rossi. The journal's policy was to display a critical attitude towards metaphysical speculation, traditional philosophical systems and problems, *Weltanschauung* philosophies and cultural or anthropological reflections on Latin American cultures. Committed to establishing a body of rigorous professional and technical materials, *Crítica. Revista hispanoamericana de filosofía (Criticism. Hispanoamerican Journal of Philosophy)*'s editors expressed a preference for verifiable explanations which referred to description and analysis, linked philosophy with methodological and scientific problems in the physical or social sciences, and which required rigour, precision and clarity instead of oratory or literary elegance. *Crítica. Revista hispanoamericana de filosofía (Criticism. Hispanoamerican Journal of Philosophy)*'s editors had originally trained in phenomenology under Gaos, but shifted towards analysis over a twenty-year period.

During the 1970s and 1980s, analysis became well established and a generation of students who grew up from there began to work at the institute, for instance, Trejo (1977), Margain (1978) and Olivé (1985a, 1985b). The institute brought to Mexico many well-known analytic philosophers. It also drew from the growing ranks of Latin American philosophers doing analysis, such as Bunge (1980) and Moulines (1982), as well as Castañeda, Coffa, Gracia and Sosa (see Mexico, philosophy in).

3 Argentina

One of the first Argentinian philosophers to be identified with analysis was Mario Bunge. A graduate from the University de la Plata, with a doctorate in physics from Simon Fraser University, Bunge was attracted to analysis after his initial rejection of it. In 1944 he became editor of the short-lived Minerva: Revista continental de filosofía (Minerva: Continental Magazine of Philosophy) (1944-5), the first Latin American journal dedicated exclusively to philosophy. Opting for an open, pluralistic approach, the journal published several early discussions on logical positivism, including Lindemann's, 'El "Círculo de Viena" y la filosofía crítica' (The Vienna Circle and Critical Philosophy) (1944).

In 1945 the members of the Grupo Argentino de la Academia Internacional de la Historia y Filosofía de la Ciencia began to meet periodically to examine the works of Russell and Moore. By 1952 the number of philosophers interested in analysis was large enough to found the Círculo Filosófico de Buenos Aires, with Bunge as its first chair. It was at the Círculo in 1954 that a draft of Bunge's important analysis of causality was first discussed (Bunge 1959). In 1956 the Agrupación Rioplatense de Lógica y Filosofía Céntifica was founded as a forum for Argentinian and Uruguayan philosophers and scientists interested in logic and the philosophy of science. Greater interest and specialization led in 1957 to the formation of two complementary groups. One, led by Gioja, the new director of the Instituto de Filosofía del Derecho y Sociología at the University of Buenos Aires, focused on problems in the philosophy of law, ethics, society and politics. The institute became a gathering place for international figures like Strawson, von Wright, Castañeda and younger Argentinian philosophers of law with an analytic orientation, among them Carrió (1965; 1970) and Rabossi (1977). The institute also sponsored *Notas de filosofía del derecho (Notes on Legal Philosophy)* (1964-9), a journal which published essays with an analytic orientation. The other, a more informal group, was interested in problems related to the philosophy of science, logic and language. It was led by Klimovsky and García, both of whom had joined Bunge in 1957 at the University of Buenos Aires. Of this group, Simpson distinguished himself with *Formas lógicas, realidad y significado (Logical Forms, Reality and Significance)* (1964), one of the earliest books in Spanish that dealt with such philosophers as Russell and Frege from an analytic perspective (see Argentina, philosophy in).

Between 1966 and 1983 political events decimated Argentinian intellectual life and many, including several analytic philosophers, were forced into exile - some permanently (Bunge went to Canada, Coffa and Gómez to the

USA).

4 Elsewhere in Latin America

In Brazil, the reaction against nineteenth-century positivism shifted attention to other approaches, such as Marxism and existentialism at the expense of the philosophies of science and logic. However, in São Paulo two of the earliest works in formal logic to appear in Latin America were published: Ferreira da Silva's *Elementos (Foundations of Mathematical Logic)* (1940) and Quine's *O Sentido da Nova Lógica (The Meaning of Modern Logic)* (1944). Quine's presence at the University of São Paulo in the early 1940s helped to make formal logic part of the philosophy curriculum and allowed Granger (1955), French philosopher of science, to teach logic and philosophy of science there from 1947 onwards. However, the 1964 military coup, in an effort to purge progressive elements at the University of São Paulo, eliminated logic and philosophical analysis from the curriculum. Philosophers found themselves without a livelihood or means of disseminating their knowledge until the late 1970s, when J.P. Monteiro, a specialist in Hume and contemporary epistemology, founded *Ciência e Filosofia* (1979-86), an interdisciplinary journal which published some analytic works. In 1975 Porchat Pereira, an Aristotelian scholar and logician, left the University of São Paulo to found the Centro de Lógica, Epistemología e História da Ciência at the University of Campinas. This new centre attracted many philosophers concerned with analysis, among them, da Costa, a specialist in nonstandard logic, and Dascal. It began to publish its journal, *Manuscrito* in 1977 (see Brazil, philosophy in).

In Peru two figures are important, Salazar Bondy, who died young and Miró Quesada, both members of the editorial board of *Crítica. Revista hispanoamericana de filosofía (Criticism. Hispanoamerican Journal of Philosophy)*. Salazar Bondy's essays were published in *Para una filosofía del valor (Towards a Philosophy of value)* (1971), a book on the analysis of the language of values in which he blends logical positivism and critical idealism to form a normative axiology. Miró Quesada, aware of the shortcomings of the traditional views of logic and science, found that formal logic and conceptual analysis provide both a critical attitude and a superior methodology, as they study the foundations of philosophy, or metatheory: 'a theory of theories… Its subject matter encompasses the study of principles, the analysis of points of departure, and the possibilities of deriving consequences of their fundamental presuppositions' (1963).

In Chile, where phenomenalism and neo-Thomism were popular, Rivano's translation of Ayer (1926) and Stahl's work on formal logic (1956; 1964) provided a theoretical base for later works by such philosophers as Torretti (1967), who made a study of Kant and later (1983) turned to problems of the history of science and Gómez-Lobo (1989), who translated Frege and applied the methods of analysis to ancient philosophy.

In Venezuela, Nuño turned to analysis after becoming interested in contemporary semantics. A polemical writer, his *Sentido de la filosofí contemporánea (Directions in Contemporary Philosophy)* (1965) is a general view of contemporary philosophy, especially logical empiricism and analytic philosophy. Elsewhere in Latin America, there has been little analytic activity. However, the wide dissemination of journals and books on analysis (Cuba being the exception) and the large number of international congresses and visiting philosophers make the scope of analysis (and of other Latin American philosophic movements as well) intercontinental. Contributing to that dissemination are the many philosophers who have migrated to the USA and Canada, such as Frondizi, Bunge, Castañeda, Gómez, Gómez-Lobo, Torretti, Sosa and Gracia. Of these, some, like Castañeda and Sosa, identify themselves with the English-speaking tradition: others, like Frondizi, with Latin America. While others, like Gracia, have succeeded in bridging the two continents.

See also: Analytical philosophy

OSCAR R. MARTÍ

References and further reading

Análisis filosófico (Philosophical Analysis) (1981-), Buenos Aires, Sociedad Argentina de Análisis Filosófico. (Argentinian journal dedicated to philosophical analysis.)

Ayer, A. (1926) 'Eliminación de la metafísica' (Elimination of metaphysics), trans. J. Rivano and C. Narvarte, *Revista de filosofía* 4 (2-3): 55-69, 1957.(One of many translations that appeared in this Chilean journal deriving from the department of philosophy in at the University of Chile.)

Bunge, M. (1944) 'Čiquest;Qué es la epistemología?' (What is Epistemology?), *Minerva* 1: 27-43.(An early

criticism of logical positivism.)
Bunge, M. (1959) *Causality*, Cambridge, MA: Harvard University Press.(An important analysis of causality in the physical sciences.)
Bunge, M. (1980) *Epistemología (Epistemology)*, Barcelona: Ariel.(An essay on epistemology and the philosophy of science.)
Carnap, R. (1932) 'Überwindung der Metaphysik durch logische Analyse der Sprache', *Erkenntnis* 2; trans. A. Pap, 'The Elimination of Metaphysics Through the Logical Analysis of Language', in A.J. Ayer (ed.) *Logical Positivism*, Glencoe, IL: Free Press, 1959, 60-81.(This essay was translated into Spanish by C.N. Molina Flores and appeared in *Cuadernos* 10, 1961.)
Carnap. R. (1935) 'Filosof ía y sintaxis lógica' (Philosophy and Logical Syntax), trans. C.N. Molina Flores, *Cuadernos* 12, 1963.(Another translation in the journal published by Centre for philosophical studies at the National Autonomous University of Mexico.)
Carrió, G.R. (1965) *Notas sobre derecho y lenguaje (Notes on Rights and Language)*, Buenos Aires: Abeledo-Perrot.(An ordinary language analysis of rights.)
Carrió, G.R. (1970) *Principos jurídicos y positivismo (Legal Principles and Legal Positivism)*, Buenos Aires: Abeledo-Perrot.(An essay on legal positivism.)
Caso, A. (1939) *Meyerson y la física moderna (Meyerson and Modern Physics)*, Mexico: La Casa de España en Mexico.(A negative assessment of logical positivism.)
Caso, A. (1941) *Positivismo, neopositivismo y fenomenología (Positivism, Neopositivism and Phenomenology)*, Mexico: Centro de Estudios Filosóficos de la Facultad de Filosofía y Letras.(A comparison of nineteenth-century and logical positivisms.)
Crítica. Revista hispanoamericana de filosofía (Criticism. Hispanoamerican Journal of Philosophy) (1967-), Mexico, Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México.(Mexican journal dedicated to philosophical analysis.)
Cuadernos de Ética (Notebook on Ethics) (1986-) Centro de Investigaciones Éticas 'Dr Frondizi', Buenos Aires: Facultad de Filosofía y Letras de la Universidad de Buenos Aires.(Argentinian journal of ethics that publishes articles on analysis.)
Diálogos (Dialogues) (1961-), Puerto Rico, Departamento de Filosofía de la Facultad de Humanidades de la Universidad de Puerto Rico.(Puerto Rican journal which publishes articles on analysis.)
Ferrater Mora, J. and Leblanc, H. (1955) *Lógica matemática (Mathematical Logic)*, Mexico: Fondo de Cultura Económica.(A textbook of mathematical logic.)
Ferreira da Silva, V. (1940) *Elementos da lógica matemática (Foundations of Mathematical Logic)*, São Paulo.(A basic logic textbook.)
Frondizi, R. (1945) *El punto de partida del filosofar (The Starting Point in Philosophy)*, Buenos Aires: Losada.(A critical assessment of logical positivism.)
Gómez-Lobo, A. (1989) *La ética de Socrates (Socratic Ethics)*, Mexico: Fondo de Cultura Económica.(A discussion of Socratic ethics.)
Gracia, J.J.E (1984) 'Philosophical Analysis in Latin America', *History of Philosophy Quarterly* 1 (1): 111-12. (An overview of philosophical analysis in Latin America.)
Gracia, J.J.E. and Frondizi, R. (eds) (1986) *Contemporary Latin American Philosophy*, Buffalo, NY: Prometheus.(A collection of contemporary Latin American philosophers, including several analytic philosophers.)
Gracia, J.J.E., Rabossi, E., Villanueva, E. and Dascal, M. (eds) (1984) *Philosophical Analysis in Latin America*, Synthese Library, Studies in Epistemology, Logic, Methodology and Philosophy of Science 172, Dordrecht: D. Reidel.(The best collection in English of contemporary Latin American analytic philosophers.)
Granger, G.G. (1955) *Lógica e filosofia das ciências (Logic and Philosophy of Science)*, São Paulo: Edições Malhoramentos.(An introduction to logic and the philosophy of science.)
Hegenberg, L. (1966) *Lógica simbólica (Symbolic Logic)*, São Paulo: Editora Herder.(A text on symbolic logic.)
Lindemann, H. (1944) 'El "Círculo de Viena" y la filosofía crítica' (The Vienna Circle and Critical Philosophy), *Minerva* 1: 123-51.(An early sketch in Latin America of logical positivism and of the Vienna Circle by a participant.)
Lindemann, H. (1946) *Lenguaje y filosofía (Language and Philosophy)*, Buenos Aires: Edición Problemas de América.(An analysis of contemporary language by a participant in the Vienna Circle.)
Manuscrito (Manuscript) (1977-), Brazil, Universidade Estadual de Campinas, Centro de Lógica, Epistemólogia e História da Ciência.(Brazilian journal on logic, epistemology and the history of science.)
Margain, H. (1978) *Racionalidad, lenguaje y filosofía (Rationality, Language and Philosophy)*, Mexico: Fondo

de Cultura Económica.(An analysis of metaphysical and epistemological problems.)
Martí, O.R. (1986) 'Review of Philosophical Analysis in Latin America', in J. Gracia, E. Rabossi, E. Villanueva and M. Dascal (eds) *Metaphilosophy* 17: 351-7.(A critical review of the first major English anthology of Latin American analytic philosophers.)
Martí, O.R. (1989) 'Mexican Philosophy in the 80's', in J.J.E. Gracia (ed.) *Philosophy and Literature in Latin America*, Buffalo, NY: State University of New York Press, 36-63.(An account of philosophical developments in Mexico since the 1970s.)
Martí, O.R. (1992) 'Gigantes y Cabezudos: la querella entre los historicistas y los analíticos' (Giants and Headstrong: The Quarrel Between Historicists and Analytic Philosophers), in *América Latina, historia y destino. Homenaje a Leopoldo Zea (Latin America, History and Destiny: Homage to Leopoldo Zea)* 2: 161-8. (A critical appraisal of the debate between analytical philosophers and historians of philosophy.)
Miró Quesada, F. (1963) *Apuntes para una teoría de la razón (Notes Towards a Theory of Reason)*, Lima: Universidad Nacional Mayor de San Marcos, Facultad de Letras.(A logical analysis of reasoning, logic and classical epistemological problems.)
Miró Quesada, F. (1980) *Filosofía de las matemáticas (Philosophy of Mathematics)*, Lima: I. Prado Pastor.(A discussion of the logical foundations of mathematics.)
Molina Flores, C. (1954) *Matemática y filosofía (Mathematics and Philosophy)*, Mexico.(An essay on the demarcation between mathematics and philosophy.)
Moore, G.E. (1912) *Ethics*, London: Williams & Norgate.(A translation of Moore's book by C. Molina Flores appeared in 1959.)
Moulines, C.U. (1982) *Exploraciones metacientíficas (Metascientific Explorations)*, Madrid: Alianza Editorial. (An essay on methodology and the philosophy of science.)
Nuño, J.A. (1965) *Sentido de la filosofía contemporánea (Directions in Contemporary Philosophy)*, Caracas: Universidad Central de Venezuela.(A discussion of logical empiricism and analytical philosophy.)
Olivé, L. (1985a) *Estado, legitimación y crisis (The State, Legitimization and Crisis)*, Mexico: Siglo XXI.(An analysis of the epistemological presuppositions of capitalism.)
Olivé, L. (1985b) *La explicación social del conocimiento (The Social Explanation of Knowledge)*, Mexico: Universidad Nacional Autonóma de México, Instituto de Investigaciones Filosóficas.(A study of the sociology of knowledge.)
Quine, W.V. (1944) *O Sentido da Nova Lógica (The Meaning of Modern Logic)*, São Paulo: Livraria Martins Editora.(An essay on trends in modern logic.)
Rabossi, E. (1977) *Análisis filosófico, lenguaje y metafísica (Philosophical Analysis, Language and Metaphysics)*, Caracas: Monte Ávila.(Essays on analytical philosophy.)
Revista latinoamericana de filosofía (Latin American Journal of Philosophy) (1975-), Centro de Investigaciones Filosóficas.(Argentinian philosophy journal with an analytical perspective.)
Romero, F. and Pucciarelli, E. (1939) *Lógica y nociones de teoría del conocimiento (Logic and Foundations of a Theory of Knowledge)*, Buenos Aires: Espasa-Calpé Argentina, 2nd edn.(A basic logic text.)
Rossi, A. (1989) *Lenguaje y singificado (Language and Meaning)*, Mexico: Fondo de Cultura Económica.(An essay on semantics, analysis and theory of description.)
Russell, B. (1912) *The Problems of Philosophy*, Oxford: Oxford University Press, 1974.(An introduction to philosophy, with an emphasis on epistemology.)
Salazar Bondy, A. (1971) *Para una filosofía del valor (Towards a Philosophy of Value)*, Santiago: Editorial Universitaria.(A discussion on axiology.)
Salmerón, F. (1992) 'Nota sobre la recepción del análisis filosófico en América Latina' (Notes on the Reception of Philosophical Analysis in Latin America), in *América Latina, historia y destino. Homenaje a Leopoldo Zea (Latin America, History and Destiny: Homage to Leopoldo Zea)* 2: 305-22, Mexico: Universidad Nacional Autónoma de México.(A recollection of the development of analysis in Latin America.)
Sanchez Reulet, A. (1956) *Contemporary Latin American Philosophy*, Albuquerque, NM: University of New Mexico Press.(A classic collection of contemporary Latin American philosophers.)
Schajowicz, L. (1964) *La filosofía y el mundo contemporáneo (Philosophy and the Contemporary World)*, Rio Piedras: Editorial Universities, Universidad de Puerto Rico.(An analysis of contemporary philosophical problems.)
Simpson, T.M. (1964) *Formas lógicas, realidad y significado (Logical Forms, Reality and Significance)*, Buenos Aires: Editorial Universitaria de Buenos Aires.(With a prologue by Gregorio Klimovsky, this is one of the earliest Latin American books to approach philosophy from an analytic perspective.)
Stahl, G. (1956) *Introducción a la lógica simbólica (Introduction to Symbolic Logic)*, Santiago de Chile:

Ediciones de la Universidad de Chile.(An introductory textbook to symbolic logic.)
Stahl, G. (1964) *Elementos de la metalógica y metamatemática (Elements of Metalogic and Metamathematics)*, Santiago de Chile: Editorial Universitaria.(A text on logical theory.)
Torretti, R. (1967) *Manuel Kant (Immanuel Kant)*, Santiago de Chile: Ediciones de la Universidad de Chile.(A study of Kant's philosophy.)
Torretti, R. (1983) *Relativity and Geometry*, New York: Pergamon Press.(A study of the problems of relativity and geometry.)
Trejo, W. (1977) *Fenomenalismo y realismo (Phenomenalism and Realism)*, Mexico: Universidad Nacional Autónoma de México.(A logical analysis of realism and phenomenology.)
Vaz Ferreira, C. (1910) *Lógica viva (Living Logic)*, Montevideo: Talleres Gráficos A. Barreiro y Ramos.(An examination of fallacies.)
Wittgenstein, L. (1957) *Tractatus Logico Philosophicus (Treatise on Logic and Philosophy)*, Madrid: Revista de Occidente.(Translation of the English version dated 1922.)

Analyticity

In Critique of Pure Reason Kant introduced the term 'analytic' for judgments whose truth is guaranteed by a certain relation of 'containment' between the constituent concepts, and 'synthetic' for judgments which are not like this. Closely related terms were found in earlier writings of Locke, Hume and Leibniz. In Kant's definition, an analytic judgment is one in which 'the predicate B belongs to the subject A, as something which is (covertly) contained in this concept A' ([1781/1787] 1965: 48). Kant called such judgments 'explicative', contrasting them with synthetic judgments which are 'ampliative'. A paradigmatic analyticity would be: bachelors are unmarried. Kant assumed that knowledge of analytic necessities has a uniquely transparent sort of explanation. In the succeeding two centuries the terms 'analytic' and 'synthetic' have been used in a variety of closely related but not strictly equivalent ways. In the early 1950s Morton White (1950) and W.V. Quine (1951) argued that the terms were fundamentally unclear and should be eschewed. Although a number of prominent philosophers have rejected their arguments, there prevails a scepticism about 'analytic' and the idea that there is an associated category of necessary truths having privileged epistemic status.

1 The attack on 'analytic'

'Analytic' has been used in a wide variety of ways: truth by conceptual containment and truth whose denial is contradictory (Kant 1781/1787); logical truth (Bolzano 1837; Feigl 1949); truth by definition and logical derivation (Frege 1884; Pap 1958); truth in virtue of form (Schlick 1930-1); truth by definition and logical truth (Carnap 1937, 1947); truth by definition (Ayer 1936); truth based on meaning (Ayer 1936; C.I. Lewis 1944); truth by semantical rule (Carnap 1947); truth in all possible worlds (C.I. Lewis 1944; D.K. Lewis 1969); convertibility into logical truth by substitution of synonyms (Quine 1951); truth by implicit convention (Putnam 1962); and so on. Although related, not all of these uses are equivalent. For example, logical truths are not true by definition (in the sense of explicit definition), but they are trivially true by definition *plus* logic. Furthermore, Gödel's incompleteness result shows that logical derivability and logical truth are not equivalent (see Gödel's theorems §3). Likewise various principles (for example, supervenience principles; see Supervenience) which are true in all possible worlds seem not to be true by definition plus logic (if 'definition' does not include 'implicit definitions'; see Definition). Similarly, it may be doubted that correct definitions provide exact synonyms (see §4). Little care has been taken to distinguish these disparate uses and needless confusions have resulted. But, as Strawson and Grice (1956) note, observations of this sort 'would scarcely amount to a rejection of the distinction [as Quine urges]. They would, rather, be a prelude to clarification'.

In 'Two Dogmas of Empiricism' (1951), Quine went far beyond the call for clarification; he argued that there simply was no distinction between analytic and synthetic truths. Quine's argument is an enthymeme with roughly the following form: there is no non-circular, purely *empiricist* clarification of the distinction, and therefore there is no distinction at all: 'That there is such a distinction to be drawn at all is an unempirical dogma of empiricists, a metaphysical article of faith' (see Quine, W.V. §8).

2 Extending the attack to 'necessary' and 'a priori'

Quine's attack did not stop with 'analytic'. Following the lead of logical positivists, he (wrongly) held that, if there were any such distinctions, the analytic/synthetic distinction would be the same as the necessary/contingent distinction, which in turn would be the same as the a priori/a posteriori distinction. So, on his view, since there is no analytic/synthetic distinction, there is no necessary/contingent distinction and no a priori/a posteriori distinction either. It is an error, however, to equate these distinctions, so these further conclusions do not follow.

Quine's attack on the necessary/contingent distinction has not convinced many. Most philosophers accept the distinction and invoke it in their work; modal logic and modal metaphysics are thriving subjects (see Modal logic; Modal logic, philosophical issues in). Do modal notions have non-circular definitions? Although definitions have been suggested, such as Church (1951) and Bealer (1982), there would be nothing unreasonable in holding them to be primitive (Prior and Fine 1977). After all, everyone must take *some* notions to be primitive.

What *epistemic justification* do we have for the necessary/contingent distinction? No doubt Quine was right that it cannot be justified on purely empirical grounds (that is, using only phenomenal experience and/or observation).

But justification is not always purely empirical: logicians, mathematicians, linguists and philosophers rely heavily on *intuitions* in the justification of their theories. (Quine himself relies on intuition in defending his formulations of set theory over others.) And there are no convincing arguments that such use of intuitions is illegitimate. Once one acknowledges intuitive evidence, the necessary/contingent distinction has a straightforward justification: we have a very wide range of robust modal intuitions (for example, the intuition that it is contingent that the number of planets is greater than seven - there could have been fewer), and when such intuitions are taken as evidence, the best theory is one which accepts the distinction at face value.

The necessary/contingent distinction is a metaphysical distinction. The a priori/a posteriori distinction, by contrast, is epistemological. Indeed, Kripke and Putnam have convincingly argued that there are necessary truths (Hesperus is Phosphorus, water is H_2O, and so on) which are impossible to justify a priori (see Kripke, S.A. §3; Putnam, H. §3). While the former distinction might well lack a non-circular analysis, the latter certainly has one: *p* is a piece of a posteriori evidence if and only if *p* is the content of an experience (a phenomenal experience, observation, memory or testimony); *p* is a piece of a priori evidence if and only if *p* is a piece of evidence which is not the content of an experience in the relevant sense. (On the assumption that one's intuitions are evidence, the contents of those intuitions would constitute one's a priori evidence.) A theory has an a posteriori justification if and only if the evidence on which that justification is based is a posteriori; a theory has an a priori justification if and only if the evidence on which that justification is based is a priori. Note that this sort of analysis does not presuppose that any theory has an a priori justification; one could deny that there is any a priori evidence. The point is that, contrary to Quine's allegation, the *notion* of a priori justification has a straightforward analysis (see A posteriori; A priori).

Quine seems to assume that a priori judgments would need to be infallible and unrevisable. But this is mistaken, as is evident from our ongoing a priori theorizing about the logical paradoxes. One of the main traditional lines of thought on the a priori - from Plato to Gödel - recognizes that a priori justification is fallible and holistic, relying on dialectic and/or a priori theory construction.

3 Defending 'analytic' against Quine

Unlike 'necessary', 'analytic' is a technical term. Accordingly, it is legitimate to demand, as Quine (1951) does, that its use be explained. This could be done with examples, but none of the (nonequivalent) notions listed above fits the standard examples perfectly, nor does any stand out as *the* most salient. This leaves the alternative of giving a definition. Quine's view is that *not one* of the historically prominent uses of 'analytic' (except the 'logical truth' use) has a satisfactory definition. But this radical scepticism is unwarranted.

Consider the following definition: a necessarily true sentence is analytic if and only if it may be converted into a logically true sentence by replacing its syntactically simple predicates with predicates which mean exactly the same thing. This seems to respect the idea that analytic truths are a priori and moreover are justifiable in a specially simple way. Quine (1960) would object to this definition by appeal to his thesis of the indeterminacy of translation - the thesis that 'there is no fact of the matter' concerning claims about identity of meaning (see Radical translation and radical interpretation §§1-3). His arguments for this thesis have not convinced many philosophers, however, for they depend on quasi-verificationist or behaviourist premises. Most philosophers reject Quine's scepticism about meaning, realism having become the dominant view. One reason for this is the advent of the broadly Gricean picture (Grice 1989) according to which meaning is analysable in terms of the propositional attitudes. Accordingly, if there is a fact of the matter about the latter, there is about the former (see Communication and intention). And, since the cognitivist revolution in psychology and philosophy of mind, nearly everyone is a realist about propositional attitudes. Thus, at least one use of 'analytic sentence' ought to be acceptable to these realist philosophers.

The same moral holds for at least certain uses of 'analytic' as it applies to propositions. Despite Quine's scepticism, most philosophers have become convinced that in logic, psychology and semantics there is need for structured propositions, that is, propositions which have a logical form (or sense structure). This makes possible a definition of 'analytic' in another one of its standard uses (Katz 1986; Bealer 1982): *p* is analytic iff_{def} every proposition having the same form (structure) as *p* is necessary.

4 More damaging problems with ‘analytic’

Although the above definition of ‘analytic’ is cogent, the term so-defined fails to apply to a number of examples which traditionally would have been deemed ‘analytic’; the definition is too narrow. For example, the defined use does not cover Kant’s paradigm example of an analyticity, namely, that bodies are extended. To accommodate this and a wide array of other examples (that circles are curves and so forth; see below), one must turn to a wider definition of ‘analytic’ - one relying on some philosophically robust notion such as definition, conceptual analysis, or the kind of meaning relations which hold between a definiendum and a definiens or between an analysandum and an analysans. For example, it is at least credible that there is a definition of ‘body’ in one of its senses according to which ‘Bodies are extended’ would be true by definition. Unfortunately, these wider accounts of ‘analytic’ give rise to a complementary problem: they let in too much.

The following familiar definitions illustrate the problem: *x* is a circle if and only if *x* is a closed plane figure every point on which is equidistant from a common point; *x* is a circle if and only if *x* is a closed plane figure every arc of which has equal curvature. There seems to be nothing to recommend one over the other; if either is a correct definition, both are. In that case it seems that the following would be true by definition plus logic: *x* is a closed plane figure every point on which is equidistant from a common point if and only if every arc of *x* has equal curvature. But Kant would deem this biconditional ‘ampliative’: in any standard axiomatic formulation of geometry, the proof of it would require axioms and axioms - as opposed to definitions - are supposed to be synthetic. Evidently, this argument can be adapted to many other a priori necessities traditionally thought to be paradigmatically synthetic.

A related kind of problem arises in connection with conceptual analysis. One of the most celebrated conceptual analyses in mathematical philosophy is the classical analysis of effective calculability, or computability. On Church’s version, a function is effectively calculable if and only if it is lambda-calculable. On Turing’s version, a function is effectively calculable if and only if it is Turing computable (see Church’s thesis). Most philosophers deem each version to be a successful conceptual analysis. But when the two analyses are combined, it follows immediately that a function is recursive if and only if it is Turing computable. But this biconditional - which is an important ‘ampliative’ theorem of formal number theory - would then turn out to be analytic (in the sense of being true by conceptual analysis plus logic) even in the event that logicism is false.

These problems suggest that there is no coherent way to draw an analytic/synthetic distinction along the lines Kant thought. One response is to settle for a severely restricted use of ‘analytic’ which concerns only concepts with unique structures (that is, unique ‘decompositions’). The price of this move, however, is high: the vast majority of our concepts - including nearly all of the concepts philosophers have sought to define or analyse (good, true, valid, number, meaning, knowledge, and so on) - are in this sense unstructured and so would not give rise to new analyticities. At best, rather uninteresting concepts (such as bachelor) are of this sort. In consequence, even if knowledge of analyticities (in the narrow sense) had a transparent epistemic explanation as Kant assumed, the sort of knowledge one seeks in typical philosophical definitions or analyses would need quite another sort of explanation.

5 Epistemological problems with ‘analytic’

Kant and his successors simply assumed that knowledge of analyticities has a transparent sort of explanation (often linked to a simplistic ‘pictorial’ or ‘mereological’ view of concepts). Just what that explanation is supposed to be has never been satisfactorily stated. It cannot be that analytic propositions are those whose truth is recognized just by virtue of possessing the constituent concepts, for *no* proposition is like this. For example, it is in principle possible that someone who possesses the relevant concepts but who is in sufficiently defective cognitive conditions (deficient intelligence, attentiveness, and so on) might fail to recognize that all and only bachelors are unmarried men. It does no good to relax this account by holding that analytic propositions are those whose truth would be recognized by anyone in sufficiently good cognitive conditions just by virtue of possessing the constituent concepts, for this lets in too much: anyone in sufficiently good cognitive conditions could not fail to recognize, say, that figure *A* is congruent with figure *B* if and only if *B* is congruent with *A*. But such propositions are the very paradigms of what Kant would have deemed synthetic a priori and requiring a different sort of explanation. (For these same reasons, purely epistemic definitions of ‘analytic’ are problematic.)

Another line of explanation is to liken our knowledge, say, that bachelors are men to our knowledge that unmarried men are men, or more generally, that *AB*s are *B*s - that is, to liken this knowledge to our knowledge of a certain kind of *logical* truth. But how do we know the logical truth that *AB*s are *B*s? Is the explanation fundamentally different from the explanation of our knowledge of other kinds of logical truths - for example, that *B*s are *A*s or *B*s? It is hard to see why it should be. This raises the question of how we know logical truths generally. For instance, is the explanation of our (logical) knowledge that identity is a symmetric relation ($A = B$ iff $B = A$) really different from the explanation of our (nonlogical) knowledge that congruence is a symmetric relation? From a phenomenological point of view, both instances of knowing (logical and nonlogical) arise from a priori intuitions and these intuitions, phenomenologically, are not relevantly different. On this score, therefore, there is no reason to think that our a priori knowledge divides neatly into two kinds having quite different explanations.

On the contrary, there is a promising unified explanation of a priori knowledge generally, which goes roughly as follows. In every case a priori knowledge is based evidentially on a priori intuition. The evidential force of a priori intuition is to be explained in terms of a general analysis of concept possession: it is constitutive of concept possession that in suitably good cognitive conditions intuitions regarding the behaviour of the concept need to be largely correct. If in suitably good cognitive conditions one did not have such intuitions, one would not be said to possess the concept. If something like this is right, then, although they mark cogent logical and metaphysical distinctions, all the listed uses of ‘analytic’ - even the narrow uses - fail to mark an epistemically significant category of knowledge.

The picture that results is complicated somewhat by Kripke and Putnam’s doctrine that there are essentially a posteriori necessities - for example, water = H_2O. Among these are some which may plausibly be deemed *scientific definitions*. If ‘analytic’ is used in the sense of truth by definition plus logic, where ‘definition’ is understood to include scientific definitions, then there would be necessities which would be both analytic and essentially a posteriori. Evidently the Kripke-Putnam doctrine applies only to ‘semantically unstable’ expressions - that is, expressions (‘water’, ‘gold’, ‘heat’, and so on) whose meaning could be different in some population of speakers whose epistemic situation is qualitatively identical to ours. These are expressions to whose meaning the external environment makes some contribution (see Content: wide and narrow). The above picture, however, holds straightforwardly for semantically stable expressions (‘conscious’, ‘know’, ‘good’, and so on) which loom large in philosophical analysis.

6 Whither ‘analytic’?

At this stage, one may reasonably ask whether continued use of ‘analytic’ serves any purpose in philosophy. Although the term evidently lacks the epistemological significance once attributed to it, the wider use of ‘analytic’ in the sense of true by definition plus logic still has utility, namely, in posing an important question: Are there necessary truths (supervenience principles, the incompatibility of colours, and so on) which are not analytic in this sense? The answer appears to be affirmative if ‘definition’ is understood straightforwardly as ordinary explicit definition. But if, as some have proposed, ‘definition’ is understood to include ‘implicit definitions’, the answer is controversial and depends on what information may legitimately be loaded into ‘implicit definitions’. On pain of trivializing significant traditional questions, however, surely not everything is admissible. Plainly there still are unanswered questions here. But they are really about the nature of definitions; ‘analytic’ does no work.

See also: Carnap, R.; Concepts; Conceptual analysis; Intensional entities; Kant, I.; Logical positivism; Necessary truth and convention

GEORGE BEALER

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz.(Uses of ‘analytic’ mentioned in §1.)

Bealer, G. (1982) *Quality and Concept*, Oxford: Clarendon Press.(Definition of ‘analytic’ in terms of propositions and concepts discussed in §3.)

Bealer, G. (1996) ‘A Priori Knowledge and the Scope of Philosophy’, *Philosophical Studies* 81: 121-42.(Account of a priori justification and knowledge in terms of intuition and concept possession described in §§2 and 5.)

Boghossian, P. (1997, forthcoming) ’Analyticity’, in C. Wright and B. Hale (eds) *A Companion to the Philosophy*

of Language, Oxford: Blackwell.(Epistemic account of analyticity.)

Bolzano, B. (1837) *Wissenschaftslehre*, Sulzbach: Seidel, 4 vols; trans. and ed. R. George, in *Theory of Science*, Oxford: Blackwell, 1972; trans. B. Terrell, ed. J. Berg, in *Theory of Science*, Dordrecht: Reidel, 1973.(Use of ‘analytic’ mentioned in §1. The George and Terrell translations provide selections from Bolzano’s work.)

Carnap, R. (1937) *Logical Syntax of Language*, London: Routledge.(Uses of ‘analytic’ mentioned in §1.)

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Uses of ‘analytic’ mentioned in §1.)

Church, A. (1951) ‘A Formulation of the Logic of Sense and Denotation’, in P. Henle, H.H. Kallen and S.K. Langer (eds) *Structure, Method and Meaning: Essays in Honor of Henry M. Scheffer*, New York: Liberal Arts Press, 3-24.(Offers non-circular definitions of modal notions.)

Feigl, H. (1949) ‘Logical Empiricism’, in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, 3-26.(Use of ‘analytic’ mentioned in §1.)

Fine, K. (1994) ‘Essence and Modality’, *Philosophical Perspectives* 8: 1-16.(Account of analyticity and definition in terms of the notion of essence.)

Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, 2nd edn, 1980.(Use of ‘analytic’ mentioned in §1.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.(Analysis of meaning in terms of propositional attitudes discussed in §3.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, New York: St Martin’s Press, 1965. (Provides the original definition of ‘analytic’ and ‘synthetic’ and an account of the possibility of synthetic a priori knowledge.)

Katz, J.J. (1986) *Cogitations*, Oxford: Oxford University Press.(Account of analyticity in terms of sense structure mentioned in §§3 and 5.)

Lewis, C.I. (1944) ‘The Modes of Meaning’, *Philosophy and Phenomenological Research* 4: 236-49; repr. L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University if Illinois Press, 1952, 50-63. (Uses of ‘analytic’ mentioned in §1.)

Lewis, D.K. (1969) *Convention*, Cambridge, MA: Harvard University Press.(Use of ‘analytic’ mentioned in §1.)

Pap, A. (1958) *Semantics and Necessary Truth*, New Haven, CT: Yale University Press.(Use of ‘analytic’ mentioned in §1 and extended discussion of analyticity and related issues.)

Parsons, C. (1997, forthcoming) *Quine and Gödel on Analyticity*, in P. Leonardi and M. Santambrogio (eds) *On Quine*, Cambridge: Cambridge University Press.(Discussion of Gödel’s uses of ‘analytic’ in relation to Quine.)

Prior, A.N. and Fine, K. (1977) *Worlds, Times and Selves*, London: Duckworth. (Defends thesis that modal notions are primitive.)

Putnam, H. (1962) ‘The Analytic and the Synthetic’, in H. Feigl and G. Maxwell (eds) *Minnesota Studies in the Philosophy of Science*, vol. 3, Minneapolis, MN: University of Minnesota Press; repr. H. Putnam, *Mind, Language and Reality: Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press, 1975, 33-69. (Use of ‘analytic’ mentioned in §1.)

Quine, W.V. (1951) ‘Two Dogmas of Empiricism’, *Philosophical Review* 60: 20-43; repr. in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1953, 20-46. (Classic attack on the analytic/synthetic distinction.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Defence of indeterminacy of translation and attack on the ontology of propositions discussed in §3.)

Quine, W.V. (1970) *Philosophy of Logic*, Englewood Cliffs, NJ: Prentice Hall.(Attack on the ontology of propositions discussed in §3.)

Schlick, M. (1930-1) *Wissenschftlicher Jahresbericht der Philosophischen Gesellschaft an der Universität zu Wein für das Vereinsjahr 1930/1*, repr. as ‘Is There a Factual A Priori?’, in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, 1949, 277-94.(Use of ‘analytic’ mentioned in §1.)

Strawson, P.F. and Grice, H.P. (1956) ‘In Defense of a Dogma’, *Philosophical Review* 65:141-58; repr. in H.P. Grice, *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, 1989, 196-212.(Early response to Quine mentioned in §1.)

White, M. (1950) *The Analytic and the Synthetic: An Untenable Dualism*, in S. Hook (ed.) *John Dewey: Philosopher of Science and Freedom*, New York: The Dial Press, 316-30; repr. in L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University of Illinois Press, 1951, 272-86.(Classic attack on the analytic/synthetic distinction.)

Anaphora

Anaphora describes a dependence of the interpretation of one natural language expression on the interpretation of another natural language expression. For example, the pronoun 'her' in (1) below is anaphorically dependent for its interpretation on the interpretation of the noun phrase 'Sally' because 'her' refers to the same person 'Sally' refers to.

(1) Sally likes her car.

As (2) below illustrates, anaphoric dependencies also occur across sentences, making anaphora a 'discourse phenomenon':

(2) A farmer owned a donkey. He beat it.

The analysis of anaphoric dependence has been the focus of a great deal of study in linguistics and philosophy. Anaphoric dependencies are difficult to accommodate within the traditional conception of compositional semantics of Tarski and Montague precisely because the meaning of anaphoric elements is dependent on other elements of the discourse.

Many expressions can be used anaphorically. For instance, anaphoric dependencies hold between the expression 'one' and the indefinite noun phrase 'a labrador' in (3) below; between the verb phrase 'loves his mother' and a 'null' anaphor (or verbal auxiliary) in (4); between the prepositional phrase 'to Paris' and the lexical item 'there' in (5); and between a segment of text and the pronoun 'it' in (6).

(3) Susan has a labrador. I want one too.
(4) John loves his mother. Fred does too.
(5) I didn't go to Paris last year. I don't go there very often.
(6) One plaintiff was passed over for promotion. Another didn't get a pay increase for five years. A third received a lower wage than men doing the same work. But the jury didn't believe any of it.

Some philosophers and linguists have also argued that verb tenses generate anaphoric dependencies.

1 Syntactic and static semantic approaches

An expression is 'anaphoric' if it depends for its interpretation on the interpretation of another expression. For example, the pronoun 'her' in (1) below is anaphorically dependent on the noun phrase 'Sally'.

(1) Sally likes her car.

Anaphoric dependencies also occur across sentences, making anaphora a discourse phenomenon:

(2) A farmer owned a donkey. He beat it.

Though anaphoric dependencies are a semantic phenomenon, syntax has a very important role to play in their analysis. In the early sixties, syntacticians noticed that anaphoric dependencies were subject to peculiar constraints that syntax could explain. For instance, consider (3) and (3′):

(3) John hit himself.
(3′) John hit him.

In (3) 'John' and 'himself' must refer to the same thing (they must be 'coreferential'), while in (3′) 'John' and 'him' cannot do so. Such observations prompted Chomsky (1981) to devise a theory of binding for anaphoric expressions, whereby all noun phrases in the syntactic representation are assigned an index. Reflexive pronouns are subject to a constraint that forces 'himself' in (3) to be co-indexed with and so bound to some noun phrase (NP) that occurs within the same clause, while nonreflexive pronouns obey another constraint that implies that 'him' in (3′) cannot be bound by an NP within the same clause in which it occurred.

While Chomsky and other syntacticians recognized that there is a semantic component to anaphora, they never

clearly articulated the semantic interpretation of co-indexed and non-co-indexed NPs. Semantics furnishes a translation of a syntactic structure into a logical representation, a ‘logical form’, that can then be assigned a model-theoretic interpretation using the formal semantics for the logical representation language, typically the language of first-order or higher-order logic (see Model theory). The syntactically generated indices on NPs are interpreted as indices on variables. Pronouns introduce an indexed variable in the logical form, while noun phrases introduce a generalized quantifier, which in turn introduces a bound variable with the index of the NP. The quantifier binds all occurrences of the same variable in its scope. So a sentence such as (1) can be represented as follows:

(1′) $\exists x_j \exists x_i(x_i =$ Sally & car(x_j) & owns(x_i, x_j) & likes(x_i, x_j)).

This approach works well enough for the simple cases of intra-sentential anaphora. But anaphoric links between pronouns and their antecedents also occur across sentences. For example, in giving the logical form of (2) above, anaphoric pronouns cause a problem. Formally, the first sentence yields (ignoring tense):

(2a) $\exists x$(farmer(x) & $\exists y$(donkey(y) & owns(x, y))).

We now tackle the second sentence. The pronouns in the second sentence are anaphorically bound to the noun phrases in the first sentence, and so we should translate the pronouns as variables that are to be bound by the quantifiers introduced by the noun phrases in the first sentence. But we have already finished the translation of the first sentence and so closed off the rightward scope of the quantifiers. Conjoining the translation of the second sentence does not produce a bound variable reading of the variables introduced by the pronouns. It produces an open sentence:

(2′) $\exists x$(farmer(x) & $\exists y$(donkey(y) & owns(x, y))) & beats(x, y).

Montague Grammar (see Montague, R.M.) seems to solve this problem with the procedure of ‘quantifying in’ (Gamut 1991). We first postulate a syntactic rule on which the noun phrase is replaced in the syntactic tree by a pronoun with an appropriate index, and the original noun phrase is adjoined with the sentence. We then apply the Montague Grammar translation procedure. Now if we interpret a sequence of two sentences as a conjunction, we can then use the quantifying-in procedure to interpret (2). First, we quantify in the two noun phrases ‘a farmer’ and ‘a donkey’. The syntactic rule for quantifying in produces the following structure for (2):

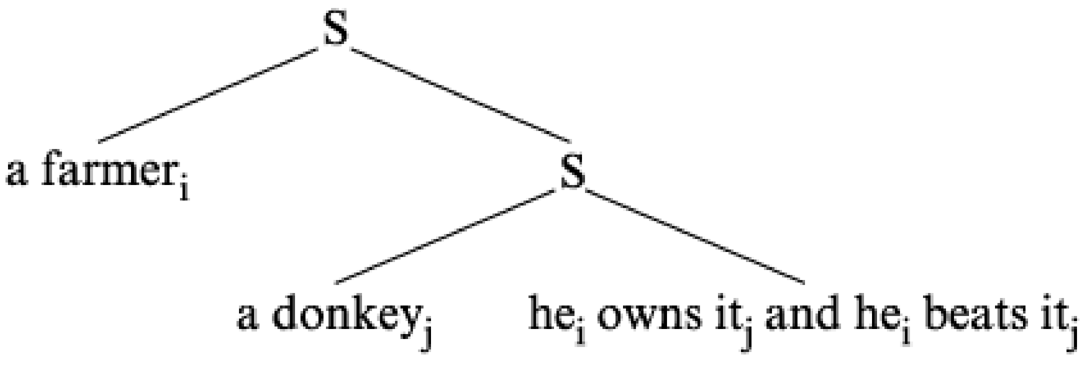

This tree now yields in Montague Grammar the correct logical form for (2):

(2*) $\exists x_j \exists x_i$(farmer(x_i) & donkey(x_j) & owns(x_i, x_j) & beats(x_i, x_j)).

However, if we wish to continue to use the procedure of quantifying in to deal with discourses in which anaphoric linkage to an antecedently occurring noun phrase exists over multiple sentences as in (4), then we must suppose a complete syntactic analysis of the discourse prior to the interpretation of any of its constituent sentences.

(4) A farmer owned a donkey. He beat it. It ran away.

This conclusion is not cognitively plausible and suggests that the Montague approach to anaphora is wrong. There are other difficulties too. Anaphoric dependencies across attitude contexts such as that between ‘a witch’ and ‘she’ in (5) below are also difficult to treat for Montague Grammar without assuming a *de re* reading of the attitude

reports (see further Geach 1963).

(5) Hob believes that a witch blighted his mare. Nob believes she killed his cow.

2 Dynamic semantics

Kamp (1981) and Heim (1982) independently proposed a formal and rigorous solution to the problem of anaphoric pronoun interpretation that has come to be called ‘dynamic semantics’. The solution redefines the semantic contribution of a sentence to the content of a discourse. In Montague Grammar, the contribution of a sentence is a proposition - formally, a set of possible worlds in which the sentence is true (see Semantics, possible worlds). Such a proposition contributes to the content of a discourse in a simple way: the meaning of a discourse is just those possible worlds that are in the intersection of all the propositions that are the meanings of the discourse’s constituent sentences. In dynamic semantics, the interpretation of a sentence *S* is a function from a discourse context to another discourse context. Such contexts may be understood in different ways (see Barwise 1987, Groenendijk and Stokhof 1991, Webber 1978). Sketched below is Kamp’s approach to discourse semantics, ‘Discourse Representation Theory’ or DRT.

A discourse context in DRT contains a set of discourse entities or ‘discourse referents’ to which elements of subsequent discourse may refer. DRT assigns a truth-conditional meaning to a natural language discourse in two steps. First, we construct a representation of the content of the discourse known as a ‘discourse representation structure’ or DRS. Second, we embed a DRS in a Tarskian model of the sort familiar from first-order logic to provide truth-conditions. A DRS is naturally construed as a partial model and conceptual representation of the discourse. It consists of a pair of sets: a set of discourse referents, or universe; and a set of conditions. The universe of a DRS is analogous to the domain of a partial model; it contains the objects (discourse referents) talked about in the discourse. Conditions are property ascriptions to these discourse referents (Asher 1993). DRSs as partial models should be distinguished from the language used to describe them. The DRS language uses a box notation - the upper part of the box lists the discourse referents in the universe of the DRS, while the lower part of the box describes the conditions. For instance, the DRS for ‘a farmer owns a donkey’ is given below. It tells us that the discourse speaks of two entities, one a farmer (*x*), the other a donkey (*y*), and that the farmer owns the donkey.

(K2a)

$$\begin{array}{|c|}\hline x \quad y \\ \hline \text{farmer}(x) \\ \text{donkey}(y) \\ \text{owns}(x, y) \\ \hline \end{array}$$

(Wada and Asher (1986), Zeevat (1989) and Asher (1993) develop a construction procedure whereby each lexical element contributes some sort of DR-theoretic structure. These then combine together compositionally following the syntactic structure of the sentence to produce a DRS for the sentence.) A noun phrase such as ‘a farmer’ introduces a novel discourse referent into the DRS as well as a condition on that discourse referent, while verbs introduce conditions on the discourse referents introduced by the noun phrases that constitute their syntactic arguments. Anaphoric pronouns, following the original Kamp treatment of DRT in 1981, are treated analogously to bound variables; they introduce occurrences of discourse referents that have been introduced previously by the processing of noun phrases in prior discourse.

The construction of a DRS for a discourse proceeds incrementally, exploiting the syntactic parse of each sentence. If $\tilde{\mathbf{K}}$ is the DRS derived from the first *n* sentences and K_{n+1} is the DRS derived from the sentence $n+1$, then the DRS for the $n+1$ sentence discourse is just the DRS that combines the universes of $\tilde{\mathbf{K}}$ and K_{n+1} and their condition sets:

$$\langle(\mathrm{U}_{\tilde{\mathbf{K}}} \cup \mathrm{U}_{\mathrm{K}_{n+1}}), (\mathrm{Con}_{\tilde{\mathbf{K}}} \cup \mathrm{Con}_{\mathrm{K}_{n+1}})\rangle.$$

So in constructing a DRS for (4), for example, we would add to the DRS K2a above the conditions ‘beats(*x*, *y*)’ and ‘runs away(*y*)’ to get a DRS for (4):

(K4)

$x \quad y$
$\text{farmer}(x)$
$\text{donkey}(y)$
$\text{owns}(x, y)$
$\text{beats}(x, y)$
$\text{runs away}(x, y)$

DRT solves another problem for Montague Grammar, having to do with indefinites, pronouns and conditionals. In Montague Grammar (as for Frege), the noun phrase ‘a man’ contributes a property of properties - that is, the property of some property *P* that there is some man that has *P* or, in symbols, $\lambda P \exists x(\text{man}(x) \ \& \ P(x))$. But then how do we account for the contribution of ‘a farmer’ in (6)?

(6) If a farmer owns a donkey, he beats it.

If we translate ‘a farmer’ with the existential quantifier, we must assign it wide scope over the conditional in order to bind the variable introduced by ‘he’ which is intuitively linked to ‘a farmer’. But this gives (6) incorrect truth-conditions. (6) requires the following translation for ‘a farmer’:
$\lambda P \forall x(\text{farmer}(x) \rightarrow P(x))$.

Montague Grammar has no uniform translation of the indefinite determiner that yields a correct treatment of (2) and (6), and also fails to explain why its translation of indefinites is context sensitive in this way. DRT, however, has a uniform translation. DRT uses its analysis of indefinite noun phrases and pronouns, together with construction procedures for conditionals and universally quantified NPs. Both constructions introduce what is known as a ‘complex condition’ in a DRS; complex conditions consist of DRSs as arguments to some operator. For instance, the grammatical conjunction ‘If…, then… ’ introduces the following relation on DRSs:

$\lambda K \lambda K'$

$K \quad \Rightarrow \quad K'$

When we have two clauses linked by ‘If…, then… ’, the first clause gives us a DRS replacing the variable *K*, while the second gives us a DRS replacing *K′*. For example, if we choose the right discourse referents for the pronouns ‘he’ and ‘it’, we translate (6) as in the DRS below, using a uniform treatment of indefinites, conditionals and pronouns.

(K6)

$x \quad y$		
$\text{farmer}(x)$ $\text{donkey}(y)$ $\text{owns}(x, y)$	$\Rightarrow$	$\text{beats}(x, y)$

As K6 shows, DRSs can occur within DRSs. DRT postulates a constraint on anaphora called ‘accessibility’ which exploits these complex structures. A discourse referent ‘ α’ is accessible from a condition ‘ β’ if and only if ‘ α’ is declared in the universe of a DRS that either contains ‘ β’ or a DRS containing ‘ β’, or is the antecedent of a conditional ‘ $\Rightarrow$’ of which the DRS containing ‘ β’ is the consequent. We may translate a pronoun with the discourse referent ‘ α’ only if ‘ α’ is accessible to the conditions of which the pronoun is the argument. This means that DRT predicts some anaphoric connections to be semantically incoherent - for example, as in ‘If a farmer $_i$ owns a donkey, he $_i$ beats it. *He $_i$ is fined’.

3 Dynamic semantics (cont.)

Let us now briefly discuss the correctness definition for DRSs. The correctness definition tells us what conditions must obtain in order for a DRS to be properly embedded in a 'DRS model'. These are the truth-conditions for the discourse the DRS represents. A DRS model is a pair $\langle D, I \rangle$, where *D* is a non-empty set (a domain of individuals) and *I* is a function that assigns to atomic *n*-ary conditions of DRSs *n*-tuples from D^n. Atomic conditions are those conditions that are derived from natural language nouns, verbs and some adjectives - the sort that are contained in the DRSs for (2) and (6). A DRS *K* is properly embedded in a DRS model *M* if and only if there is a function from U_K into the universe of *M* such that all the conditions of *K* are satisfied in *M*. What we have to do now is to define satisfaction of a condition in *K*. For atomic DRS conditions, the definition of satisfaction is completely analogous to the satisfaction of an atomic formula of first-order logic relative to a model and an assignment to free variables. The recursive definition for, for example, complex conditions of the form $K \Rightarrow K'$, however, is novel, reflecting the quantificational force of '$\Rightarrow$'. $K \Rightarrow K'$ is satisfied in a DRS model *M* relative to an embedding function *f* if and only if for every function *g* that extends *f* to a proper embedding of *K* in *M* there is an extension *h* of *g* that is a proper embedding of K' in *M*.

In applying the correctness definition to K4, we get the intuitively right truth-conditions for the sentence. The function *f* is a proper embedding of K4 just in case there is an object *a* and an object *b* such that *a* is a farmer, *b* is a donkey, *a* owns *b*, *a* beats *b*, and *b* runs away. The correctness definition also captures the truth-conditions of a conditional sentence such as (6) - for more details see Kamp and Reyle (1993).

Since its introduction, DRT and other theories of dynamic semantics have been used to analyse various anaphoric phenomena - for example, plural anaphora (Kamp and Reyle 1993), anaphora across attitude contexts (Asher 1987), temporal anaphora (Kamp and Rohrer 1983, Kamp and Reyle 1993) and verb phrase (VP) ellipsis (Klein 1986).

DRT makes an important contribution to our understanding of anaphora. But there are problems. DRT's approach to anaphors and quantification sometimes fails to get the right readings in examples such as the following:

(12) If I have a dime in my pocket, I'll put it in the meter.

DRT predicts that every dime I have in my pocket will be put in the meter, a prediction that must somehow be blocked. Sentences involving pronouns and other quantifiers yield similar difficulties. (13) casts doubt on a basic principle of both Montague Grammar and DRT; the 'it' in (13) does not seem to function as a bound variable but rather as a definite description (Evans 1980).

(13) Either there's no bathroom in this house, or it's in a funny place.

Yet another difficulty for the DRT account of anaphora is that it cannot handle many cases of propositional anaphora. DRT does not offer for (14), for instance, any entities of the appropriate type to which 'it' can refer.

(14) One plaintiff was passed over for promotion. Another didn't get a pay increase for five years. A third received a lower wage than men doing the same work. But the jury didn't believe any of it.

One approach to propositional anaphora such as that in (14) has been to analyse in greater detail the discourse structure of a text. Interpreters naturally understand a text as divided into meaningful segments related in various ways. Some parts give a narrative; others furnish a background for a narrative; still others may elaborate or explain other parts. Such discourse structure is completely missing from DRT, but Asher (1993) and Webber (1991) argue that these parts of the text can serve as referents to anaphoric pronouns. Notions of discourse structure have been incorporated into a dynamic-semantic framework (Asher 1993) to analyse propositional anaphora, VP ellipsis and temporal anaphora (Lascarides and Asher 1993).

See also: Discourse; Logical and mathematical terms, glossary of

NICHOLAS ASHER

References and further reading

Asher, N. (1987) 'A Typology for Attitude Verbs and their Anaphoric Properties', *Linguistics and Philosophy* 10:

125-97.(This research paper gives a detailed study of anaphora across attitude contexts within DRT.)

Asher, N. (1993) *Reference to Abstract Objects in Discourse*, Dordrecht: Kluwer.(This book offers an extension of DRT for the purpose of analysing anaphoric reference to propositions and other abstract objects.)

Barwise, J. (1987) 'Noun Phrases, Generalized Quantifiers, and Anaphora', in E. Engdahl and P. Gärdenfors (eds), *Generalized Quantifiers: Linguistic and Logical Approaches*, Dordrecht: Reidel.(This article offers an alternative formulation of dynamic semantics. It does not use semantic representations such as DRSs.)

Chomsky, N. (1981) *Lectures on Government and Binding: The Pisa Lectures*, Dordrecht: Foris.(This is one of the central works on syntactic constraints on anaphora.)

Evans, G. (1980) 'Pronouns', *Linguistic Inquiry* 11: 337-62.(This is one of the classic articles in the philosophy of language on anaphora. It introduces an alternative to the bound variable approach.)

Gamut, L.T.F. (1991) *Logic, Language and Meaning*, esp. vol. 2, *Intensional Logic and Logical Grammar*, Chicago, IL: University of Chicago Press.(This book is a very good introduction to semantics. Dynamic semantics is presented in volume 2.)

Geach, P.T. (1963) *Reference and Generality*, Ithaca, NY: Cornell University Press. (This classic text includes many puzzles for semantics.)

Groenendijk, J. and Stokhof, M. (1991) 'Dynamic Predicate Logic', *Linguistics and Philosophy* 14: 39-100. (This article presents an alternative to DRT's treatment of dynamic semantics. It is similar to Barwise's approach but better worked out.)

Heim, I. (1982) *The Semantics of Indefinite and Definite Noun Phrases* (Ph.D. thesis), Amherst, MA: University of Massachusetts Press.(Besides Kamp's 1981 paper, this is the principal source for dynamic semantics. It is devoted to the study of indefinite and definite noun phrases and their behaviour as anaphoric referents.)

Kamp, H. (1981) 'A Theory of Truth and Semantic Representation', in J. Groenendijk, T. Janssen and M. Stokhof (eds) *Formal Methods in the Study of Language*, Amsterdam: Mathematisch Centrum Tracts, 277-322; repr. in J. Groenendijk *et al*. (eds) *Truth, Interpretation and Information*, Dordrecht: Foris, 1984, 1-41.(This is Kamp's first paper on DRT. It is difficult to read but is the principal source for many if not most of the developments in dynamic semantics.)

Kamp, H. and Reyle, U. (1993) *From Discourse to Logic: Introduction to Model-Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Dordrecht: Kluwer.(This book is a lengthy but readable introduction to DRT. It is written for the non-specialist but also includes valuable material for the specialist.)

Kamp, H. and Rohrer, C. (1983) 'Tense in Texts', in R. Bäuerle, C. Schwarze and A. von Stechow (eds) *Meaning, Use and Interpretation of Language*, Berlin: de Gruyter, 250-69.(This article describes a dynamic-semantic approach to tense and temporal anaphora. The study is carried out in DRT.)

Klein, E. (1986) 'VP Ellipsis in DR Theory', in J. Groenendijk (ed.) *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, Dordrecht: Foris.(This paper shows how dynamic semantics can be used to analyse VP ellipsis.)

Lascarides, A. and Asher, N. (1993) 'Temporal Interpretation, Discourse Relations, and Commonsense Entailment', *Linguistics and Philosophy* 16: 437-93.(This article criticizes the dynamic-semantic treatment of tense and temporal anaphora.)

Wada, H. and Asher, N. (1986) 'BLDRS: A Prolog Implementation of LFG and DR Theory', *Proceedings of the 11th International Conference on Computational Linguistics*, Bonn, 540-5.(This was one early implementation of DRT and suggested a compositional treatment of dynamic semantics. Some other such implementations are mentioned in the bibliography.)

Webber, B.L. (1978) *A Formal Approach to Discourse Anaphora*, report 3761, Boston, MA: Bolt, Beranek & Newman Research.(This is a computational approach to anaphora. It has some similarities to dynamic semantics.)

Webber, B.L. (1991) 'Structure and Ostension in the Interpretation of Discourse Deixis', *Language and Cognitive Processes* 6: 107-36. (This article investigates a variety of anaphoric phenomena in discourse.)

Zeevat, H. (1989) 'A Compositional Approach to Discourse Representations', *Linguistics and Philosophy* 12: 95-131.(This article gives the first compositional treatment of DRT that brings out the similarities with dynamic predicate logic.)

Anarchism

Anarchism is the view that a society without the state, or government, is both possible and desirable. Although there have been intimations of the anarchist outlook throughout history, anarchist ideas emerged in their modern form in the late eighteenth and early nineteenth centuries in the wake of the French and Industrial Revolutions.

All anarchists support some version of each of the following broad claims: (1) people have no general obligation to obey the commands of the state; (2) the state ought to be abolished; (3) some kind of stateless society is possible and desirable; (4) the transition from state to anarchy is a realistic prospect

Within this broad framework there is a rich variety of anarchist thought. The main political division is between the 'classical' or socialist school, which tends to reject or restrict private property, and the 'individualist' or libertarian tradition, which defends private acquisition and looks to free market exchange as a model for the desirable society. Philosophical differences follow this division to some extent, the classical school appealing principally to natural law and perfectionist ethics, and the individualists to natural rights and egoism. Another possible distinction is between the 'old' anarchism of the nineteenth century (including both the classical and individualist traditions) and the 'new' anarchist thought that has developed since the Second World War, which applies the insights of such recent ethical currents as feminism, ecology and postmodernism.

Anarchists have produced powerful arguments denying any general obligation to obey the state and pointing out the ill effects of state power. More open to question are their claims that states ought to be abolished, that social order is possible without the state and that a transition to anarchy is a realistic possibility.

1 Philosophical anarchism

A common starting point for anarchists is 'philosophical anarchism', the denial of any general duty in the individual to obey the laws of the state. The best-known formulation of this view is that of R.P. Wolff (1970), who argues that any such claim of duty comes into conflict with the individual's moral autonomy (see Autonomy, ethical). People who accept a duty to act merely because the state has so commanded thereby surrender their capacity to judge the rightness of actions for themselves. Assuming that moral autonomy should never be surrendered, a general duty to obey the state (regardless of the content of its command) is inadmissible. This argument is a strong one, but from a genuinely anarchist point of view its implications are limited. Even if one should not act merely because the action has been commanded by the state, there may be good reason, acceptable to an autonomous agent, to comply with the state's command in particular cases. Philosophical anarchism is consistent with accepting the state's continued existence. It is at best a necessary, not a sufficient, element of a full anarchist case.

2 Abolition of the state

Full-blooded anarchism involves the claim that the state should be abolished. Anarchists have advanced several lines of argument to this end.

First, a utilitarian case might be made, in the manner of Godwin, the earliest systematic writer in the 'classical' anarchist tradition (Crowder 1991) (see Godwin, W. §3). For Godwin, the disutility of government results, first, from its support for 'the established administration of property', which diminishes the sum of human happiness by dividing society into unequal and antagonistic classes; second, from its tendency to overreach the limits of its competence (Godwin 1798). Utility requires that goods be distributed so as to bring about the greatest happiness, and that people's affairs be left to those most familiar with them, namely the individuals themselves (see Utilitarianism).

A second line of argument is also found in Godwin: perfectionism (see Perfectionism). Anticipating J.S. Mill, Godwin sees 'the perfection of which human nature is capable' as involving the development of a robustly autonomous personality. Government, however, by its very nature tends to crush or undermine this kind of autonomy, either by coercion, or by encouraging an attitude of servile obedience or by upholding relations of economic inequality that place one class at the mercy of another. Government must thus be abolished, as an obstacle to the realization of true humanity. A similar element of perfectionism is present in all the classical

anarchists. In Bakunin, Kropotkin and others, the influence of the romantic movement is apparent in their insistence that perfection requires the development of individual uniqueness. This is an emphasis popular among contemporary anarchists, but in the classical texts the promotion of personal idiosyncrasy is usually circumscribed by a prior commitment to some conception of a universal moral law.

This latter insistence suggests a third ground of opposition to the state, the assumption, again characteristic of the classical anarchists, that there is an objective moral law somehow immanent in or deducible from nature. This group thus subscribes to the notion of natural law (see Natural law §3). As to the precise form and content of natural law they offer different accounts, but all agree that it is violated by the state. Closest to the traditional Thomistic notion of natural law as the command of God is the view found put forward by Lev Tolstoi (Tolstoi 1936). For Tolstoi, the content of the moral law is Christian, centring upon the idea of universal brotherhood and the command to love one’s neighbour. These imperatives, Tolstoi believes, are incompatible with the existence of states, since states set up barriers of enmity between people by creating false national identities and shoring up class distinctions.

Tolstoi is the only major anarchist thinker to adopt an explicitly theistic and Christian foundation; the general self-image of the anarchists is militantly antireligious, an outlook exemplified by Bakunin’s (1882). The most fully worked out account of natural law along these secular lines is found in the work of Kropotkin, who takes his cue from Darwinian evolution. In opposition to ‘social Darwinists’ like Herbert Spencer, Kropotkin (1902) argues that the dominant factor of evolutionary development is not competition but ‘mutual aid’ within species (see Evolution and ethics). The state must be abolished because, divisive and hierarchical, it constitutes an obstacle to humanity’s conforming with this law of nature. Once government is removed, nature can be expected to take its course, ushering in a new era of universal cooperation and brotherhood.

In the individualist tradition, the natural law style of argument against the state takes the form of claims to natural rights. Murray Rothbard (1973), for example, argues that the kind of rights defended by John Locke - rights, rooted in human nature, to security of person and possessions - are incompatible with the state. Government, by its very nature, violates people’s rights by coercing them in various ways, including the taxation of their property. Although it might be replied that government is also capable of protecting the individual’s rights, the individualist anarchist response is that such protection is provided only at the cost of invading rights in other respects. This is impermissible, since natural rights are absolute. (A similar argument is made by ‘minimal state’ liberals like Robert Nozick (see Libertarianism §3; Nozick, R.).)

Another line of individualist argument appeals to self-interest. Ethical egoism, the view that people ought to prefer their own self-interest to any other principle of conduct, was first developed in an anarchist direction by the Young Hegelian Stirner. Extending Feuerbach’s argument against religion to strike at all notions of ‘the sacred’, Stirner (1845) argues that people’s potential for self-realization is impeded by the very idea of ethical obligation. Among the various ‘spooks’ that restrict our lives at present, the state, with its demand for obedience, is one of the most oppressive. The state is not the sacred institution we suppose it to be, but merely an especially pervasive agency of violence and coercion. Self-interest demands that we dismantle it.

These arguments may be challenged on many points. For example, the anarchist arguments based on natural law and natural rights may be thought to illustrate some familiar problems with those doctrines, such as the apparent arbitrariness of the content attributed to them (see Natural law §5). Kropotkin does not adequately justify his emphasis on mutual aid rather than competition; Rothbard’s natural rights are narrowly defined to exclude the possibility of ‘positive’ rights to welfare assistance of the kind historically associated with government intervention. More generally, all these anarchist arguments may be accused of a certain one-sidedness. Taken by themselves, the anarchists’ criticisms of the state are often powerful, but their attacks are seldom balanced by any fair consideration of points in government’s favour. The tendency is to set out a vivid catalogue of the evils of which the state is capable, and to leave the argument at that.

3 Types of anarchy

After the attack on the state, the next step in the anarchist case is the presentation of some account of the stateless society as a desirable alternative. The general picture anarchists present varies from case to case; their different grounds of opposition to government imply different values to be promoted in anarchy. At one extreme is

Kropotkin's vision of a society characterized by voluntary mutual aid, at the other Stirner's conception of a 'union of egoists' in which the only constraints on people's conduct will be the limits of their personal power.

A major source of contrast between different anarchies is the issue of how goods should be distributed. Roughly speaking, the classical tradition argues for the rejection or restriction of private property; the individualist school defends it, looking to market exchange as its central model of social relations. The dominant view in the classical tradition is Kropotkin's 'communist' anarchism, which requires distribution according to need (see Communism). At the other end of the spectrum, contemporary individualists like Rothbard reject virtually all restrictions on the operation of the market, and are happy to accept the label 'anarcho-capitalist'. On this view, even services traditionally performed by governments, such as law enforcement and defence, should be voluntarily purchased from private suppliers. An intermediate position is occupied by Proudhon's 'mutualism' (see Proudhon, P.-J.), under which goods are to be distributed by free contract, with the strong qualification that prices be fixed according to units of labour-time. The good society would thus be one in which all would work, and each (assuming roughly equal abilities and industry) would receive roughly similar rewards (Proudhon 1923).

The many difficulties faced by all of these schemes lead to a general problem with conceptions of anarchy, that of how to ensure social order in the absence of the familiar enforcement machinery of the state. Here again there are several different proposals. Confronted by the problem of crime, most classical anarchists first suggest that there will be very much less crime in the absence of private property and its associated inequalities. But assuming the likelihood of at least occasional criminal acts and, more generally, interpersonal disputes, they part company over the best response. For some, such problems will be solved by an eventual universal convergence of ethical judgment brought about by either or both of the following: (1) the release (especially through revolutionary comradeship) of natural feelings of solidarity hitherto suppressed by artificial state-imposed divisions; (2) an improvement in rational understanding of ethical obligations resulting from advances in the natural and social sciences. A second response to the problem of order is the device of public censure, under which people will be brought into line by the pressure of public opinion. Third, some individualists are happy to endorse a system of courts and coercive enforcement, as long as this is maintained by voluntarily paid subscription.

4 Theories of transition

A perennial difficulty in anarchist thought is the question of how society can pass from the existing state system, apparently well entrenched, to the future stateless order. It was principally on this point that the anarchist and Marxist streams of nineteenth-century socialism eventually diverged (see Socialism). Anarchists like Bakunin rejected the Marxist notions of a vanguard party and a transitional period of revolutionary government on the ground that these devices are incipiently authoritarian, and in this they may appear prescient. But the price of their consistency is that they are left to formulate a plausible alternative.

Some anarchists insist on strictly nonviolent means, relying on rational enlightenment (Godwin) or religious reawakening (Tolstoi). Bakunin and Kropotkin emphasize the role of science in promoting progress through rational agreement. Individualists like Tucker (1972) are generally committed to nonviolence, allowing at most passive forms of civil disobedience such as refusing to serve in the armed forces or to pay taxes.

Alongside these irenic views, and sometimes awkwardly combined with them, there is also a strong commitment in some anarchists (Proudhon, Bakunin, Kropotkin) to revolutionary action (see Revolution). All these thinkers agree in principle that the anarchist revolution must be spontaneous or 'bottom-up' rather than subject to any kind of leadership that could evolve into a government. There are, however, various proposals for ways of stimulating the revolution without actually leading it. 'Anarcho-syndicalists' believe that the revolution is to be advanced by direct industrial action culminating in a general strike (Rocker 1938). The school of 'propaganda by deed' urges that the insurrection be inspired by sudden, symbolic acts of anti-bourgeois terrorism. The resulting popular association of anarchism with arbitrary violence is, however, unfair and misleading. It is true that some anarchists, like Bakunin, place an emphasis on the violent destruction of the existing order that sometimes goes beyond the recommendation of a necessary means to an end and becomes a form of self-affirmation held to be valuable in itself. But this tendency is not distinctively anarchist, and neither follows from nor fits comfortably with the most fundamental anarchist principles, which give priority to autonomous judgment and ethical restraint.

5 The 'new' anarchism

The anarchists' principled repudiation of governmental organization even for the purposes of revolution is one of several factors that contributed to the eclipse of anarchism as a political force after the Second World War. Against a background of expanded welfare states in the West and state socialist hegemony in the East, the inheritors of the classical tradition in particular were challenged to revise their thinking. The results have been described as a 'new' generation of anarchist thought (Runkle 1972).

Not surprisingly, some post-war anarchists (see, for example, Ward 1973) have retreated from theories of comprehensive social transformation, concentrating instead on developing spheres of individual and small-group autonomy in the interstices of existing state structures. In this connection, existentialist approaches to anarchism have been developed (see, for example, Read 1954). More ambitious claims continue to be advanced, however. In the 1960s some anarchists were attracted to the apparently revolutionary potential of the 'situationist' ethic that emerged from the amorphous New Left movement (Debord 1983). However, current interest is for the most part focused on three main lines of thought: feminism, ecology and postmodernism.

Feminist anarchism, pioneered by Emma Goldman (1917), is the view that the state is a typical institution of patriarchy, an expression of the characteristic male ethic of impersonal 'justice' backed by violence. Women, in contrast, are said to be anarchists by nature, since female instincts and ideals - of sexuality and motherhood, for example - are non-hierarchical and anti-authoritarian. This claim is vulnerable to the same objections that attach to any 'essentialist' form of feminism, namely that evidence for distinctively male and female 'natures', and hence male and female ethics, is inconclusive (see Feminist political philosophy §5). That is not to rule out the more moderate claim that the historical experience of women gives them an interest in dismantling the state over and above the interest in doing so possessed by men.

The leading ecological anarchist, Murray Bookchin (1982), argues that the state expresses and reinforces a principle of hierarchy which, when applied to people's dealings with nature, encourages the view that humanity's proper relation to the natural world is one of control and conquest. The result is the devastation of our environment. The solution is the 'ecological society', which will be an anarchist society, since in repudiating the hierarchical values that threaten nature it will abandon the hierarchies that have oppressed human beings. To this it might be replied that it is far from obvious that a society that considers itself an equal partner with nature must be internally egalitarian; ecological goals might well be achieved - perhaps more easily achieved - through non-democratic institutions (see Green political philosophy §§1, 3).

Finally, the loosely defined tendency in contemporary thought known as 'postmodernism' may be thought suggestive from an anarchist point of view (May 1994). Michel Foucault's technique of undermining dominant value systems by laying bare the contingent and power-seeking genealogies underlying them implies a kind of liberation for those people previously controlled or marginalized by such systems. It is easy to see that notions of government, state and authority could be dismantled in this way. Less obvious is the possibility of passing from here to a positive recommendation of anarchism, since this too could be undermined by the same genealogical method. An anarchism based on postmodernism is conceivable, but it would probably be negative or destructive in tone (see Postmodernism and political philosophy).

Short of massive and unforseeable changes in world political and economic conditions, it is hard to see any of these more recent developments leading to a revival of anarchist influence at the level it enjoyed in the late nineteenth century. The present power of states, the pervasive influence of international capitalism and the widespread demand for the goods and services these institutions provide, all combine to make anarchism an unlikely political option. The new formulations testify, nevertheless, to the resilience of anarchist ideas and to their continuing role as correctives or counterweights to more conventional views.

GEORGE CROWDER

References and further reading

Bakunin, M. (1961-80) *Archives Bakounines*, ed. A. Lehning, Leiden: Brill, 7 vols in 8.(The best scholarly edition, although not yet complete. Another standard edition is the *Oeuvres de Bakounine*, ed. J. Guillaume, Paris: Stock, 6 vols, 1895-1914.)

Bakunin, M. (1973) *Selected Writings*, trans. S. Cox and O. Stevens, ed. A. Lehning, London: Jonathan Cape.(The best selection of Bakunin's writings available in English.)

Bakunin, M. (1882) *God and the State*, trans. B. Tucker, New York: Dover, 1970.(The best known work by Bakunin available in English. Contains a characteristic anarchist attack on religion as an ally of government; see §2.)
Bakunin, M. (1990) *Bakunin: Statism and Anarchy*, ed. M. Shatz, Cambridge: Cambridge University Press.(The leading anarchist critique of Marxism: see §4.)
Bookchin, M. (1982) *The Ecology of Freedom*, Palo Alto, CA: Cheshire.(Argues for a link between anarchist and ecological ideas; see §5.)
Crowder, G. (1991) *Classical Anarchism: The Political Thought of Godwin, Proudhon, Bakunin, and Kropotkin*, Oxford: Oxford University Press.(A critical study, arguing that the 'classical' tradition of anarchism is more coherent and persuasive, in historical context, than usually believed.)
Debord, G. (1983) *Society of the Spectacle*, Detroit, MI: Black & Red.(A manifesto of New Left 'situationism'; see §5.)
Godwin, W. (1798) *Enquiry Concerning Political Justice*, ed. I. Kramnick, Harmondsworth: Penguin, 1976.(The 1976 edition is the most accessible but there is also the standard modern edition (ed. F.E.L. Priestley, Toronto: University of Toronto Press, 3 vols, 1946) which includes a facsimile of the 3rd edition (London: Robinson, 3 vols) and variant readings among the three editions.)
Godwin, W. (1993) *Political and Philosophical Writings of William Godwin*, ed. M. Philp, M. Fitzpatrick and W. St Clair, London: Pickering & Chatto.(A selection of the most significant of Godwin's writings in politics and philosophy.)
Goldman, E. (1917) *Anarchism and Other Essays*, New York: Mother Earth Publishing Association, 3rd edn; repr. New York: Dover, 1969.(A collection of essays combining anarchist and feminist insights.)
Kropotkin, P. (1902) *Mutual Aid*, London: Heinemann.(Argues that the 'dominant factor' of evolution is cooperation rather than competition.)
Kropotkin, P. (1912) *Modern Science and Anarchism*, London: Freedom Press.(Sees anarchist principles as demonstrable by the methods of the natural sciences.)
Kropotkin, P. (1927) *Revolutionary Pamphlets*, ed. R. Baldwin, New York: Vanguard.(A collection of influential essays, including Kropotkin's 1905 article on anarchism for the *Encyclopedia Brittanica*.)
May, T. (1994) *The Political Philosophy of Poststructuralist Anarchism*, University Park, PA: Pennsylvania State University Press.(Explores the affinities between anarchism and postmodernism.)
Miller, D. (1984) *Anarchism*, London: Dent.(The best analytical and critical introduction.)
Proudhon, P.-J. (1851) *General Idea of the Revolution in the Nineteenth Century*, trans. J.B. Robinson, London: Freedom Press, 1923.(Includes the idea of contract as 'mutualist' exchange; see §3.)
Proudhon, P.-J. (1923-59) *Oeuvres complètes*, ed. C.C.A. Bouglé and H. Moysset, Paris: Rivière, 15 vols in 19. (The standard scholarly edition.)
Proudhon, P.-J. (1840) *What is Property?*, ed. and trans. D. Kelley and B. Smith, Cambridge: Cambridge University Press, 1994.(The famous answer is 'theft'.)
Read, H. (1954) *Anarchy and Order*, London: Faber.(A collection of essays drawing connections between anarchism and existentialism; see §5.)
Ritter, A. (1980) *Anarchism: A Theoretical Analysis*, Cambridge: Cambridge University Press. (Argues that 'communal individuality' rather than freedom is the paramount value for anarchists.)
Rocker, R. (1938) *Anarcho-Syndicalism*, London: Secker & Warburg.(Anarchist goals can be pursued through the trade union movement; see §4.)
Rothbard, M. (1973) *For a New Liberty*, New York: Macmillan.(A defence of 'individualist' anarchism in its anarcho-capitalist form; see §2.)
Runkle, G. (1972) *Anarchism: Old and New*, New York: Delta.(A useful survey, proposing a distinction between two phases of anarchism; see §5.)
Stirner, M. (1845) *The Ego and its Own*, trans. S. Byington, London: Rebel, 1982.(An 'egoist' critique of the state; see §2.)
Tolstoi, L. (1936) *The Kingdom of God and Peace Essays*, trans. A. Maude, London: Oxford University Press. (Anarchist conclusions reached from a Christian point of view; see §2.)
Tucker, B. (1972) *Instead of a Book*, New York: Arno.(Essays by a leading US individualist anarchist of the nineteenth century, influenced by Bakunin, Proudhon and Stirner.)
Ward, C. (1973) *Anarchy in Action*, London: Freedom Press.(A pragmatic from of anarchism, operating alongside the modern state; see §5.)
Wolff, R.P. (1970) *In Defense of Anarchism*, New York: Harper Torchbooks.(An argument for philosophical anarchism; see §1.)

Woodcock, G. (1963) *Anarchism*, Harmondsworth: Penguin.(The classic general introduction.)

Anaxagoras (500-428 BC)

Anaxagoras of Clazomenae was a major Greek philosopher of the Presocratic period, who worked in the Ionian tradition of inquiry into nature. While his cosmology largely recasts the sixth-century system of Anaximenes, the focus of the surviving fragments is on ontological questions. The often quoted opening of his book - 'all things were together' - echoes the Eleatic Parmenides' characterization of true being, but signals recognition of time, change and plurality. Even so, Anaxagoras is deeply committed to the Eleatic notions that, strictly speaking, there can be no coming into being or going out of existence, nor any separation of one part of reality from any other. His main object is to show how the variety of the world about us is somehow already contained in the primordial mixture, and is explicable only on the assumption that latent within each substance are portions of every other. Whether or not he owed his conception of unlimited smallness to Zeno of Elea, he held that there could be no such thing as a magnitude of least size; and he claimed that there was accordingly no difference in complexity between the large and the small.

Mind, however, is a distinct principle; unlimited, autonomous, free from the admixture of any other substance. Hence Anaxagoras' decision to make it the first cause of the ordered universe we now inhabit. Mind initiates and controls a vortex, which from small beginnings sucks in an ever-increasing expanse of the surrounding envelope. The vortex brings about an incomplete separation of the ingredients of the original mixture: hot from cold, dry from wet, bright from dark, and so on, with a flat earth compacted at the centre and surrounded by misty air and clearer ether above and below. Contemporaries were scandalized by Anaxagoras' claim that sun, moon and stars were nothing but incandescent stones caught up in the revolving ether.

Later fifth-century physicists - notably Archelaus and Diogenes of Apollonia - developed revised versions of Anaxagoras' system, but abandoned his dualism. His conception of mind excited but disappointed Socrates, and exercised a profound influence on Plato's cosmology and Aristotle's psychology. Aristotle was also fascinated by the complexities of the remarkable theory of 'everything in everything'. Anaxagoras' philosophy was never subsequently revived, but he was remembered as the mentor of the statesman Pericles and the poet Euripides. His reputation as a rationalist critic of religion persisted throughout antiquity.

1 Life and work

Anaxagoras, son of Hegesibulus, was a native of Clazomenae (a coastal town in what is now Turkey). He was the first major philosopher to spend time in Athens, where he was an associate of the statesman Pericles. The evidence for his residence there is confused although quite extensive. He is said to have arrived in Athens in the archonship of Callias (456/5 BC). Knowledge of his ideas probably preceded his arrival, to judge from echoes of his explanation of physical phenomena in Aeschylus' tragedies *Supplices* (*c.*463 BC) and *Eumenides* (458 BC). He was believed to have predicted the fall of a large meteorite in Thrace that is dated *c.*467 BC. Anaxagoras' stay in Athens may have lasted for about twenty years (so Mansfeld 1979-80), until his prosecution and trial on a charge of impiety (dated by Mansfeld to 437/6 BC). He died in Lampsacus on the Hellespont, where the funeral honours accorded to him suggest a high reputation.

Diogenes Laertius (I 16) counts Anaxagoras among those philosophers who wrote only one work. But some of his views, for example those on perspective and geometry, may have been the subjects of separate memoranda (A38-40). The surviving fragments all appear to come from a general work on the nature and origins of the physical world. The fact of their survival makes for a more direct impression of Anaxagoras' style of thought and his major theses than is possible in the case of Presocratics such as Anaximenes. However, there remain considerable problems of interpretation. One reason is that Anaxagoras' prose, while capable of striking and elevated effect, is too dense and imprecise to allow determinate formulation of the subtle and ingenious distinctions which scholarship has seen as fundamental to his system. Another is that the Aristotelian commentator Simplicius, who is responsible for preserving most of the fragments, quotes them not in the course of a straightforward exposition of Anaxagoras' philosophy, but in order to illustrate various points in either Aristotle's account of Anaxagorean ontology or his own Neoplatonist interpretation of the cosmogony. So while possession of actual extracts from Anaxagoras' book is a huge bonus, their function in his argumentative or expository strategy often eludes us.

2 The original condition

The first few sentences of Anaxagoras' book ran as follows:

> All things were together, unlimited both in quantity and in smallness. For the small was indeed unlimited. And with all things being together nothing was manifest on account of smallness. For air and ether contained all things, both being unlimited. For these are the greatest present among the totality of things, both in quantity and in magnitude.

(fr. 1)

Much in this speculative story is mysterious. What range or category of items count as 'things' (*chrēmata*)? How is their 'smallness' to be understood? In what sense did air and ether 'contain' everything? What is the difference intended between their greatness in quantity and their greatness in magnitude? It was a characteristic trait of Presocratic writing first to grab the reader's attention with a memorable opening, and then to whet the appetite with a pregnant development whose full meaning and justification would emerge only gradually. Anaxagoras provides a copybook example of the technique.

Only one other fragment describes the original condition:

> Before these things were separated off, when all things were together, not even any colour was manifest. For the mingling together of all things prevented it - of the wet and the dry and the hot and the cold and the bright and the dark, much earth being present there also and seeds unlimited in quantity, in no way like each other. For of the others no one is at all like another. Since this is so, one must suppose that all things were present in the totality.

(fr. 4, second part)

Putting the two passages together, we get the following lists of ingredients of the mixture: (1) air and ether (the predominant constituents), a lot of earth and seeds unlimited in number; (2) the wet and the dry, the hot and the cold, and so on. The relation of lists (1) and (2) is obscure. It may be that the wet and the dry and so on, simply *are* earth, air, ether, and so on, under another description, one particularly important for understanding the process of separation which initiates cosmogony. There is also the question of which form of description Anaxagoras conceived as the more fundamental for ontology.

The major items listed in (1) are presumably the actual quantities of the material stuffs which dominate the universe as it is now. The identity and ontological status of Anaxagoras' seeds has been much debated. The final sentences of fragment 4 suggest that he is once again extrapolating from what the universe at present contains to what it must be supposed to have contained originally. A best guess is that from the irreducible present variety of an unlimited number of living species (plants as well as animals), he infers unlimited numbers of biological seeds in the original condition. These he may have conceived as small particles - that would make sense of the explanation in fragment 1 that nothing was manifest 'on account of smallness'. So interpreted, Anaxagoras is making the general assumption that instances of whatever clearly differentiated species of living things now exist must also have existed at least in seminal form before cosmogony.

The thesis in fragment 1 - that there is no lower limit on how small something could be in the original condition - has as its counterpart the thesis that there is no upper limit on how large things can be. From fragment 2, which speaks of what surrounds our differentiated universe as unlimited in quantity, we can infer that the totality was and is infinite in size. Anaxagoras appears to argue the point about largeness by parity of reasoning:

> For of the small there is no least but always a lesser (for what is cannot not be) - but also there is always a larger than the large. And it is equal in quantity to the small. But with respect to itself each thing is both large and small.

(fr. 3)

Readers of fragment 3 have often been reminded of Zeno's paradoxes (see Zeno of Elea §§4-6), although it is not known which philosopher wrote first. Anaxagoras evidently differs from Zeno in finding the notion of infinite divisibility basically unproblematic. The similarity lies not just in the denial that there is a least, but in the metaphysical status of that claim, sustained as it is not by physical considerations but by reflection on our

conceptions of magnitude, and here particularly the logic of 'large' and 'small'.

3 In everything a portion of everything

The evidence suggests that Anaxagoras next introduced the theme of the separation of the ingredients of the primordial mixture. It looks as though this section of the book emphasized not the process itself nor its outcome in cosmic order, but a very general feature of reality as it is in its separated state: in everything there is a portion of everything. Despite separation, there is a sense in which mixture remains the general condition of things.

This paradoxical theory dominates the presentation of Anaxagoras' views in Aristotle and in what we can recover of other ancient writers who report them. Anaxagoras apparently appealed to several different lines of argument, perhaps designated 'signs' (*sēmeia*) or 'evidences' (*tekmēria*), in recommending the theory to his readers. For one of these we have his own statement (fr. 6). Others are mentioned by Aristotle and the subsequent tradition; here only one sentence attributable to Anaxagoras himself survives (fr. 10).

Most weight was accorded to an argument from growth (for example, A46). Bread and water must contain hair, flesh, nails, bone, and so on. Otherwise the ingestion of food would not make them grow: 'how could hair come from non-hair or flesh from non-flesh?' (fr. 10). As Aristotle puts it, Anaxagoras said that everything is in everything because he saw everything coming to be from everything (*Physics* I 4). A subsidiary argument developed a parallel story about opposites. White has black in it: otherwise the brilliance of snow could not turn into the darkness of water (A97). Anaxagoras made a similar point (details unknown) about heavy and light. Finally, only if there were indivisible minima would it be possible for microscopic pure stuffs to be isolated (fr. 6). But as it is there is no lower limit on smallness (fr. 3), and indeed the small (that is, the latent) must be regarded as no less complex than the large (that is, the manifest):

> And since, too, there are portions equal in number of both the large and the small, in this way too all things will be in everything; nor can they exist separately, but all participate in a portion of everything. Since the least cannot be, none of them could be separated, nor come to be on its own, but as in the beginning so too now all things must be together. And in all things there are many even of those that are separating off, equal in number in both the larger and the smaller.
>
> (fr. 6)

Ancient and modern interpreters alike have been intrigued by the doctrine that there is something of everything in everything. It has accordingly been subject to post-Anaxagorean theoretical refinement and elaboration. Aristotle, followed by the doxographical tradition, inferred from Anaxagoras' arguments for it that the basic building blocks of his ontology were 'homoeomerous', that is, things whose parts were all the same in character as the whole (fr. 6). And he identified these building blocks as the bone, flesh and so on, introduced in the main argument (see, for example, A43, 46). On two occasions he went so far as to claim that air and fire are not themselves elements but composites of 'these and all the other seeds' (see, for example, A43).

These interpretative moves are probably all mistaken. Seeds certainly look to be fundamental items in Anaxagoras' system. But if it is right to suppose that they are conceived in biological terms, as containing what we would call the genetic code for the various living forms which will emerge from them, none like any other, then they cannot be identified with homoeomerous substances such as bone. Indeed, in containing something of everything they will have bone, flesh and so on, as actual or potential ingredients. Whether biological seeds are themselves homoeomerous seems irrelevant to their role in Anaxagoras' system. Again, while Anaxagoras undoubtedly treated bone, flesh and so on, as ingredients also of bread and water, it is unlikely that he regarded them as more fundamental than the animal seeds from which ultimately they emerge. Lastly, to treat air and ether as composites of other more elemental items is to reverse the whole direction of Anaxagoras' explanatory enterprise, which insists on the irreducibility of the variety of the world, and is guided by the maxim: 'The appearances are a sight of what is not manifest' (fr. 21a).

Although Aristotle gave pride of place to homoeomerous stuffs in his account of Anaxagoras' ontology, he also acknowledged the elemental role of opposites. Some modern scholarship has argued that this relative emphasis should be reversed. In the surviving fragments it is certainly the opposites dry and wet, hot and cold, dark and bright and so on, which figure most prominently in Anaxagoras' most general statements about ingredients of

mixtures and separation from mixtures. They also recur as the key factors in his explanation of perception (see §5): we discern the cold by the hot, the drinkable by the brackish, the sweet by the pungent. This evidence has suggested that the claim that there is a portion of everything in everything is to be understood primarily as the thesis that every substance or seed of an organism contains opposite powers of every sort - as fragment 4 puts it: 'forms and colours and savours of every kind'. On this view, Anaxagoras may well have conceived such powers as more fundamental than air, ether, earth and living species, at least to the extent of holding that such things derive their distinctive characters and causal properties from the particular combinations of opposite powers which are inherent in them as portions.

The word 'portion' (*moira*) looks as if it may be an Anaxagorean technical term. Modern scholarship has argued that it should be interpreted as equivalent to '*pro*portion'. On this view the point of its introduction can be elucidated by confronting an objection to Anaxagoras' thesis that 'each thing is or was most manifestly the things it contains or contained most of' (fr. 12, end). Take the convenient if probably non-Anaxagorean example of gold. If to count as gold a substance has to contain a predominance of gold, even though it has in it something of everything, must not the predominant gold itself contain a predominance of gold, and so *ad infinitum*? This potentially vicious regress succumbs to a distinction between the discrete substances apparent in the world about us, which contain portions of all things, and the unmixed elements that in different proportions make up these discrete substances. Gold is most manifestly gold because it contains a predominant proportion of *pure* gold. So construed, participation (*metechein*) in portions begins to look like an anticipation of Plato's notion of participation in Forms.

Some distinction of this kind is undoubtedly needed if Anaxagoras' doctrine of a portion of everything in everything is to sustain a claim to intelligibility. And the proposal that the expression 'portion' is its vehicle is unquestionably attractive. But it is not without difficulties. Consider for example the beginning of fragment 4: 'These things being so, it is right to think that in all the things that were being put together there were many things of all kinds, and seeds of all things - having forms and colours and savours of every kind'.

This statement is clearly a version of the doctrine. The words 'these things being so' might even be a reference to the *tekmēria* we have just been reviewing. In any event, Anaxagoras is here applying the theory to emerging compounds, and highlighting the omnipresence of seeds within them. But are ingredient seeds pure elemental portions? Not if as argued above they are complex particles, which mention of forms, colours and so on, tends to confirm. Perhaps seeds are ingredients which are themselves made up of elemental portions; or perhaps there is a degree of indeterminacy in the notion of a portion.

4 Mind

At least once in his exposition of the doctrine of 'everything in everything' Anaxagoras announced an exception to the general rule: 'In everything there is a portion of everything except mind, but there are some things in which there is mind too' (fr. 11). Fragment 12 issues a symmetrical disclaimer: just as there are things whose ingredients do not include mind, so mind is something which itself has no ingredients. This is less obviously true, so Anaxagoras argues out the proposition at some length:

> The other things participate in a portion of everything, but mind is unlimited and self-controlling and is not mixed with any thing, but exists alone itself by itself. For if it were not by itself but had been mixed with something else, it would participate in all things, if it had been mixed with any (for in everything there is a portion of everything, as I have said earlier); and the things mixed together with it would be preventing it, so that it would not control any thing in the same way as it actually does being alone by itself. For it is finest of all things and purest; moreover it has all knowledge about everything and harbours greatest strength; moreover all the things that have soul, both the larger and the smaller, all of them mind controls.
>
> (fr.12)

This passage establishes Anaxagoras as a dualist. It is an open question whether in predicating fineness and purity of mind he means to characterize it as incorporeal, or as a uniquely pure body unlike any other material substance - or perhaps simply as something with no specific features of its own other than the capacity for knowledge and control, as seems to be Aristotle's interpretation. Some scholars, noting the hymnic quality of some of the prose in fragment 12, write 'Mind', not 'mind'. But Anaxagoras is talking in general terms, even though the agent of the

cosmogony he is about to describe must be a supreme mind. In seeking to identify a cause adequate to account for an ordered universe, he hits on two properties of mind which qualify it alone for the role: its freedom from the (unspecified) limitations which inhibit all other substances, and its power of controlling movement (presumably other things are subject to the forces exercised upon them by other bodies). No earlier thinker enunciated either the concept of a first cause or its identity as mind with anything approximating Anaxagoras' explicitness.

Having established mind's aptness for the task, Anaxagoras went on to sketch his hypothesis about how the world was formed. He writes as though there *is* only one world, and the ancient doxographers confirm that this was his view, despite some now obscure discussion by him of a separation 'elsewhere' (fr. 4, first part). The account runs as follows:

> And mind controlled the whole rotation, so that it began to rotate at the beginning. And first it began to rotate in a small way, but it is rotating more, and it will rotate more. And the things which were being mixed together and separated off and distinguished, mind knew them all. And whatever things were to be - both those which were and those which are now and those which will be - all these mind ordered, and also this rotation in which now rotate the stars and the sun and the moon and the air and the ether which are being separated off. And the rotation itself caused them to be separated off. And the dense is separated off from the rare, and the hot from the cold, and the clear from the murky, and the dry from the wet.
>
> (fr. 12, continued)

Aristotle echoed the Platonic Socrates' disappointment with this account (A47). Anaxagoras' introduction of mind as principal cause seemed to them to promise a demonstration of how things were ordered for the best. But its actual role in his cosmogony seemed like a mere device to trigger the vortex, which is what does all the real explanatory work. It is indeed quite unclear how mind exercises any 'control' over the process. Should we suppose that, like a design engineer, it could have designed a different mechanistic system from the one it actually produced? Another complaint was that in accounting for particular phenomena Anaxagoras seldom invoked mind, but appealed to 'airs and ethers and waters and many other absurdities' (Plato, *Phaedo* 98c). Post-Newtonian and post-Darwinian readers may think this a point in his favour.

The principal function of the vortex is clearly to 'separate off' (*apokrinein*) concentrations of air and ether, or again of heavier matter; for example, in the form of the earth and the heavenly bodies. But something else mind achieves is 'dispersal' (*diakrinein*):

> And when mind had begun to cause motion, it was separated off from everything moved by it. And whatever mind set in motion, all this was dispersed. But once such things were being moved and dispersed the rotation caused them to be dispersed much more.
>
> (fr. 13)

The idea of dispersal is less clear than that of separation. Perhaps it is the process whereby 'seeds' are disentangled from one another, in contrast with what is necessary for stuffs like air and ether. 'Mixture' (*symmisgesthai*), the other main process referred to in fragment 12 and elsewhere, is not attributed to mind's activity. Doubtless it is the mechanism by which humans and animals and plants are 'compounded' (fr. 4): fragment 17 asserts that what the Greeks call coming into being is really mixture. The extant fragments say nothing about how mixture works. Do seeds somehow control a process whereby portions of the appropriate stuffs are absorbed from the earth and elsewhere into the growing organism? On this view they would exercise one of the key functions carried out in Empedocles' system by Love, with mind playing a role analogous to his Strife (see Empedocles §5).

5 Conclusion

Anaxagoras evidently included in his book discussion of the full range of particular physical problems that had become mandatory topics for a treatise on nature. These include details of the formation of the earth and the heavenly bodies, accounts of meteorological phenomena, a treatment of the origins of animal life and an exploration of the fundamentals of human physiology and psychology. By comparison with his ontology, Anaxagoras' views on these issues are mostly less original. None the less some specific points deserve a mention.

Anaxagoras knows, for example, that the moon has plains and ravines, borrows its light from the sun and is eclipsed when screened by the earth. His explanation of the rising of the Nile in summer, as due to snows melting

to the south, is eminently reasonable, albeit probably wholly speculative (A42). His most interesting remark about humans is the claim that 'it is the possession of hands which makes them the wisest of living things' (A102). Theophrastus tells us at some length about Anaxagoras' theory of sensation. The main idea is that like is not affected by like: something that is as warm or cold as us does not warm or cool us. We are aware of warmth because we are cold, or rather because the warmth in us is deficient relative to the cold (A92).

Anaxagoras' physical system, like the atomism of Leucippus and Democritus, is developed within a framework that is at once Eleatic and Ionian. But he offers an opposite approach to most of the problems he attacks. He is what would nowadays be described as a mind-matter dualist: they are materialists. In place of discrete atoms in eternal motion through a void, he thinks of matter as intrinsically inert,and constituting an infinitely divisible continuum. And the powers inherent in matter are without effect unless triggered by the purposive decision of mind.

MALCOLM SCHOFIELD

References and further reading

Anaxagoras (500-428 BC) Fragments in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 2, 5-44.(The standard collection of the ancient sources, including fragments and testimonia, the latter designated by 'A'; includes Greek texts with translations in German.)

Furth, M. (1991) 'A "Philosophical Hero"? Anaxagoras and the Eleatics', *Oxford Studies in Ancient Philosophy* 9: 95-129.(One of a sequence of excellent philosophical articles on Anaxagoras' theory of matter; comments on many of the earlier classic essays on this subject.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols.(The most detailed and comprehensive English-language history of early Greek thought; a serviceable treatment of Anaxagoras, in vol. 2 pages 266-338, which includes a useful appendix offering translations of selected passages on his theory of matter.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(A valuable survey of Presocratic philosophy, including texts and translations; contains a useful study of Anaxagoras.)

Lanza, D. (1966) *Anassagora: Testimonianze e frammenti*, Florence: La Nuova Italia. (Reproduces the fragments and testimonia as in Diels and Kranz (1952), with Italian translation and commentary.)

Mansfeld, J. (1979-80) 'The Chronology of Anaxagoras' Athenian Period and the Date of his Trial', *Mnemosyne* 32: 39-60 and 33: 17-95; repr. in part in *Studies in the Historiography of Greek Philosophy*, Assen and Maastricht: Van Gorcum, 1990, 264-306. (The most complete study of Anaxagoras' chronology.)

Schofield, M. (1980) *An Essay on Anaxagoras*, Cambridge: Cambridge University Press.(A monograph on Anaxagoras' theories of mind and matter based on detailed study of the principal fragments.)

Sider, D. (1981) *The Fragments of Anaxagoras*, Meisenheim am Glan: Verlag Anton Hain. (A useful edition of the Greek text of the fragments, with translation and commentary.)

Anaxarchus (*c*.380-*c*.330 BC)

The Greek philosopher Anaxarchus of Abdera was a friend of Alexander the Great, teacher and friend of Pyrrho, and heroic victim of a tyrant. More a court philosopher than a school one, and an ambiguous personality, he seems to have mixed a highly original philosophical cocktail: a primarily ethical, cynically inclined outlook, combined with certain elements of Democritean ethics, epistemology and physics. His only attested work was a treatise On Kingship, his dominant interest perhaps being the theory and practice of relations between intellectual and ruler.

Little is known of Anaxarchus' early life before meeting Alexander. His background lay in the writings of Democritus, like him a native of Abdera, and in the teaching of the obscure Diogenes of Smyrna, a pupil of the sceptical Democritean, Metrodorus of Chios. The ancient writers of philosophical 'successions' placed Anaxarchus in a chain running from the Eleatics and Democritus to Pyrrho, to give Pyrrhonian scepticism a distinguished ancestry. That chain underlies the life of Anaxarchus recounted by Diogenes Laertius (Book IX).

After somehow joining Alexander's inner circle, Anaxarchus followed him on his eastern campaign, and brought Pyrrho with him. Alexander's biographers frequently cite his witticisms and sayings. He comes over mainly as caustic and arrogant - strongly enough placed to have no fear of making enemies, notably Nicocreon, who later took cruel vengeance on him, and Callisthenes, a relative of Aristotle who was eventually put to death by Alexander. This feud with Callisthenes explains the striking hostility of the Peripatetic tradition towards Anaxarchus, described in it as a voluptuary and an odious flatterer of Alexander (A4, A9). His attitude towards the king seems actually to have been the much more complex one of a sort of philosopher-jester. Boldly exercising his 'freedom of speech' (*parrhēsia*) in sometimes risky jokes, he pretends to encourage Alexander's ambitions of conquering the world, of being hailed in the oriental fashion (*proskynēsis*), or of being acknowledged as a god, using doubled-edged arguments which are implicitly derisive. He mocks the tears which Alexander sheds after killing his friend Clitus, with the put-down remark that 'everything done by those in power is just and lawful'. In these different situations, Anaxarchus 'skilfully mixes sweetness and acidity' (A7), and perhaps shows how we may interpret his ambition of 'bringing people back to their senses in the easiest way' (A1). The two preserved fragments of his treatise *On Kingship* happen to bear on the two areas in which the relations between the intellectual and power can pose thorny problems: the spoken word (fr. 1) and money (fr. 2). In these areas, one should show the kind of wisdom which Anaxarchus (fr. 1) defines as 'knowledge of the measures of *kairos* [the right moment]'.

The tales of Anaxarchus' death, however, confirm that he was not merely an opportunistic courtier. Having accidentally fallen into the hands of his old enemy, the Cypriot tyrant Nicocreon, he was thrown into a mortar and crushed with iron pestles. Retaining his calm and his irony even under torture, he said: 'Crush, crush Anaxarchus' container; Anaxarchus himself you will not crush.' This heroic death, like Socrates', long remained a paradigm of philosophical martyrdom.

According to Plutarch, Anaxarchus 'from the beginning beat an original path in philosophy'. This frustratingly brief statement is hard to flesh out. On the basis of his nickname, 'the happiness man' (*eudaimonikos*), earned by 'his impassivity and his even temper in life' (A1), late sources (A14) implausibly make him the founder of a 'happiness school'. In reality, 'impassivity' was a Cynic trait and 'even temper' a Democritean one. Both passed, in time, into Pyrrhonism (see Pyrrhonism §6).

According to Sextus Empiricus (A16), many ranked Anaxarchus with the philosophers who 'abolished the criterion of truth'. However, it is unlikely that he maintained a very sophisticated scepticism. In one of his preserved fragments (fr. 1) he presents himself implicitly as a 'polymath', someone of multiple knowledge, and one anecdote (A11) has him, in conversation with Alexander, preaching the Democritean doctrine of infinite worlds. If there is any scepticism in him, it is rather a practical, lived scepticism, albeit influenced by Democritus' famous assault on the reality of the sensible world (see Democritus §3). It is expressed in the ironic yet despairing saying attributed to Anaxarchus, as also to the Cynic Monimus: 'They compared existing things to a stage-painting (*skēnographia*), and thought that they resembled the images seen in dreaming or insanity' (A16).

Just as an old world was vanishing and a new one being born amidst the din, the blood and the fury, Anaxarchus knew how to give striking expression to that feeling of unreality which accompanies such turnings in the tide of history. In that troubled age, which lost its bearings without finding new ones, Anaxarchus remained happy both in pleasure and under torture. He won and kept the friendship of people like Alexander and Pyrrho, and aroused the lasting hatred and resentment of someone like Nicocreon. No doubt he was, as Timon suggests (A10), an ambiguous man. Certainly he was no ordinary man.

JACQUES BRUNSCHWIG

References and further reading

Anaxarchus (*c*.380-*c*.330 BC) Fragments and testimonia in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker*, Berlin: Weidemann, 6th edn, 1952, vol. 2, 235-40.(Standard collection of the ancient sources, both fragments and testimonia, the latter designated by ’A’; includes Greek texts with translations in German; now superseded by Dorandi (1994).)

Anaxarchus (*c*.380-*c*.330 BC) Fragments and testimonia in T. Dorandi, ‘Anaxarchi Democritei Fragmenta’, *Atti e Memorie dell’Accademia Toscana di Scienze e Lettere ‘La Colombaria’* 59, 1994, 9-59.(More complete and accurate than Diels and Kranz; with Italian translation.)

Bernard, P. (1984) ‘Le philosophe Anaxarque et le roi Nicocréon de Salamine’, *Journal des Savants*: 3-49. (Searching analysis of Anaxarchus’ jokes and of the material circumstances of his death.)

Brunschwig, J. (1993) ‘The Anaxarchus Case: An Essay on Survival’, *Proceedings of the British Academy* 82: 59-88.(General portrayal of the person and his ambiguities as an intellectual in the face of absolute power; comparison with Curzio Malaparte.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols. (Book IX, 58-60 is a life of Anaxarchus.)

Goulet, R. and Queyrel, F. (1989) ‘Anaxarque d’Abdère’, in R. Goulet (ed.) *Dictionnaire des philosophes antiques*, Paris: Éditions du CNRS, vol. 1, 188-91. (General portrayal, bibliography and iconography.)

Anaximander (*c*.610-after 546 BC)

The Greek philosopher Anaximander of Miletus followed Thales in his philosophical and scientific interests. He wrote a book, of which one fragment survives, and is the first Presocratic philosopher about whom we have enough information to reconstruct his theories in any detail. He was principally concerned with the origin, structure and workings of the world, and attempted to account for them consistently, through a small number of principles and mechanisms. Like other thinkers of his tradition, he gave the Olympian gods no role in creating the world or controlling events. Instead, he held that the world originated from a vast, eternal, moving material of no definite nature, which he called apeiron *('boundless' or 'unlimited'). From this, through obscure processes including one called 'separation off', arose the world as we know it. Anaximander described the* kosmos *(world) and stated the distances of the celestial bodies from the earth. He accounted for the origin of animal life and explained how humans first emerged. He pictured the world as a battleground in which opposite natures, such as hot and cold, constantly encroach upon one another, and described this process as taking place with order and regularity.*

1 Life and work

Very little is known of Anaximander's life. His dates (*c*.610 BC-shortly after 546 BC) are not certain, but make him a generation younger than Thales, whose pupil, successor and associate he is variously called. Like Thales he was a Milesian. He is said to have travelled to Sparta, where he predicted an earthquake and set up a *gnōmōn* (a Babylonian invention for marking the length of the sun's shadow, which he is credited with discovering) on the sundials there to mark solstices, equinoxes and the hours of the day. He is also said to have led a Milesian colony to Apollonia on the Black Sea and, it is reported, made the first map and the first 'sphere' or celestial globe. He wrote at least one work, known as *On Nature* (the title the Alexandrian scholars later gave to the works of most of the Presocratics; it is not Anaximander's title) in which he presented his views on the *kosmos*. (We hear of several other works - *Circuit of the Earth*, *On the Fixed Stars*, *Celestial Globe* - but these are dubious.)

As with other reported discoveries of the Presocratics, this evidence demands cautious treatment. The map is possible, although it will have been extremely crude and founded more on principles of symmetry than on measurement. (See Herodotus, IV 36 for a critical assessment of early maps.) And, since Anaximander had views about the size and shape of the *kosmos*, he may have constructed a model of it. If he foretold an earthquake however, it was just a lucky guess. Alternatively, later authors could have invented the prediction to give Anaximander something comparable to Thales' prediction of an eclipse. The report about the *gnōmōn* is usually accepted as likely, as it agrees with Anaximander's undoubted interest in astronomy.

2 The *apeiron*

Anaximander is best known for his physical theory, which described the original material of the universe as *apeiron*: 'boundless' or 'unlimited' or, possibly, 'indefinite' (the word later acquired the more technical meaning of 'infinite'). What this material is like, how it is related to the *kosmos* around us, and how Anaximander justified his view are basic and controversial issues in understanding his thought. To begin with, what kinds of bounds or limits does it lack? The description 'eternal and unageing' indicates that it is unlimited in time, and since it 'surrounds all the *kosmoi* [plural of *kosmos*]' (Hippolytus, *Refutation* I 6.1, A11) it is vast in extent, and if not infinite, at least unlimited in that there is nothing outside it that limits or determines its size. Furthermore, since it is 'neither water nor any other of the things called elements, but some different *apeiron* nature' (Simplicius, *On Aristotle's Physics* 24.16, A9), it is without any definite character or qualities. Hence some argue that it was called unlimited because it lacks internal boundaries or distinctions. It is not clear that the word *apeiron* can bear this meaning, but Anaximander's original substance is nevertheless indefinite in this way too. The original substance was also the originative substance 'out of which came to be all the heavens and the *kosmoi* in them' (*On Aristotle's Physics* 24.17, A9). Anaximander presented a cosmogony in which the *kosmos* arose from the *apeiron* in a series of developmental stages. Thus, the *apeiron* is the ancestor of all that exists.

Aristotle and his followers present another view: the *apeiron* is the element or substance out of which everything is composed, an Aristotelian material cause (see Aristotle §9). Thus, everything is made of *apeiron* in the way coal and diamonds are made of carbon. Aristotle occasionally identifies the *apeiron* instead as a mixture of the four

elements that he recognized, and also as a substance intermediate between fire and air or between air and water. These ’mixture’ and ’intermediate’ interpretations must be discarded as guesswork, and the idea that it is a material cause must be rejected as Aristotelian invention too, since it does not fit the rest of the evidence on the role of the *apeiron* in Anaximander’s system.

Not only is the *apeiron* our ancestor, it is divine. Anything that is ’eternal and ageless’ and also ’in motion’ (*On Aristotle’s Physics* 24.13, A9), which is ’immortal and imperishable’ and which ’surrounds all and steers all’ counts, for the Greeks, as a divine being (Aristotle, *Physics* 203b11-13, A15). It is disputed how many of these words were Anaximander’s, but the ideas they represent seem authentic. Just what ‘divine’ means in this context is of critical importance. In some sense the *apeiron* is the Creator, but it is remote from the Greeks’ anthropomorphic conception of the Olympian gods, who demand worship, intervene in human affairs and are motivated by pride, anger and favouritism. Like Xenophanes’ god, the *apeiron* is ’not similar to mortals in form or thought’ (Xenophanes, fr. 23) (see Xenophanes §3). Unlike Xenophanes’ god, the *apeiron* lacks perceptive and cognitive capacities (Xenophanes, fr. 24). It seems to have generated the *kosmos* not through any conscious purpose, but somehow as the result of its eternal motion, and the sense in which it ’steers all’ seems to be simply that the way the *kosmos* was generated guarantees that the events that take place in it are governed by an immutable, impersonal, universal law.

Why make the originative substance *apeiron*? The sources attribute two arguments for this thesis to Anaximander (although other ancient arguments are sometimes thought to go back to him as well). The first, which argues that it must be *apeiron* in the sense of ’unlimited in extent’, goes as follows: ’it must be unlimited lest generation fail’ (Aristotle, *Physics* 208a8, A14; Aëtius, I 3.3, A14). Aristotle criticizes the argument on the grounds that ’the destruction of one thing can be the origin of another, the total being limited’. If Anaximander is assumed to be referring either to our own finite *kosmos* or to a succession of finite *kosmoi*, one after another (see §3), it is indeed a bad argument. But if he held that there are an unlimited number of *kosmoi* at the same time, as Aëtius’ text suggests, the argument succeeds as far as the vagueness of the term ’unlimited’ permits.

The second argument concludes that the *apeiron* is qualitatively indefinite: ’The elements have opposite qualities. Air is cold, water wet, fire hot. If any of them were infinite, the others would have been destroyed. Therefore, the elements arise from something different’ (Aristotle, *Physics* 204b26, A16). As it stands, the argument is laden with Aristotelian terminology, and does not prove that the *apeiron* is qualitatively indefinite, only that it is different from the four Aristotelian elements. Its authenticity has been questioned, but it probably has an Anaximandrian kernel, attacking Thales’ conception of water as the basic material of the universe using the argument: ’If everything were made of, or originated from, water, then everything would be wet; but some things are wet, others dry (the opposite of wet); therefore, the basic substance cannot be wet, or dry either’ (see Thales §2). The argument can be generalized to show that the basic substance is unlike any definite substance and lacks all perceptible qualities.

3 The *kosmos*

The point of having an indeterminate originative substance is to allow to arise the wide variety of things (including opposites) that we find in the world around us. Anaximander’s account of the origin of the *kosmos* confirms this interpretation. From the eternal *apeiron*, ’something capable of producing hot and cold’ separated off. This gave rise to hot in the form of a sphere of flame, which surrounded cold, in the form of dark mist, ’like bark around a tree’. The sphere of flame subsequently broke up to form the sun, moon and stars, while the mist afterwards became earth and sea.

This is not a creation myth where a divinity creates or acts on matter separate from itself, but an essentially biological account of generation or development which takes place because of the nature of the material that generates the *kosmos*. On this account, the *apeiron* plays no active role after the obscure initial process, which is described as ’separation off’. Once started, the world goes its own way. Anaximander speaks of other events that involve ’separation off’, and to judge by them the process does not require purposeful activity or the intervention of an agent. It need not involve any more than some part or amount of an existing thing coming together and being isolated from the rest in such a way as to take on a distinct identity from the rest and behave differently.

Several details of Anaximander’s astronomy deserve mention. The heavenly bodies are rings of fire enclosed in

tubes of mist pierced with holes for the fire to shine through. Eclipses and the phases of the moon are caused when the holes are blocked. The sun is the same size as the earth. The sun is farthest from the earth, followed by the moon, with the stars nearest. The earth is a cylinder, ’like a column drum’, whose depth is one-third its breadth. Anaximander gave the distances of the heavenly bodies from the earth, and although the evidence is untidy, most scholars agree that he made the distances of the stars, moon and sun respectively nine, eighteen and twenty-seven times the size (?breadth) of the earth. The earth is in the centre of the *kosmos* and remains there without support.

The absence of the *apeiron* and of the Olympian gods is notable, as is the boldness of Anaximander’s ideas. His figures for the size of the *kosmos* are fanciful, but his conviction that the *kosmos* is symmetrical and based on proportion, and the tenet that important facts about the physical world can be expressed numerically, are a priori principles worthy of a scientist. The odd picture of the celestial bodies as rings rather than points of fire is doubtless the answer to the question where all the fire in the cosmogonic sphere of flame went to. Anaximander may (although this is speculation) have located the stars nearest the earth through the following reasoning: their light is dimmest and they give the least heat because the fire they are made of is less pure than that of the sun; celestial fire is purer the less affected it is - and therefore the more remote it is - from the cool region of the earth; therefore the stars are nearest the earth and the sun farthest.

Anaximander believed the earth stays in the centre without support ’because it stays put on account of its similar distance from all things’ (Hippolytus, *Refutation* I 6.3, A11). Also, ’it is no more appropriate for what is located in the middle and similarly related to the extremes to move up, down or sideways; and it is impossible to move in opposite directions simultaneously; and so by necessity it is at rest’ (Aristotle, *On the Heaven* 295b12, A26). Although not all scholars accept this reasoning as Anaximander’s, there is no reason why he could not have said something similar to the former of these reports. Aristotle could then have supplied the missing steps in the argument, which amounts to the first recorded use of the Principle of Sufficient Reason. The picture of the earth hanging free in the middle of the *kosmos* contrasts with Greek mythology and with Thales’ belief that the earth rests on water, as well as with that of Anaximander’s successor, Anaximenes (§2), who believed that it is supported by a column of air. It is based on our experience that the earth does not seem to move, along with Anaximander’s theory of the origin and structure of the *kosmos*, which leaves no room for a prop.

Like other Presocratics, Anaximander gave accounts of meteorological phenomena agreeing with his cosmogony and cosmology. Fire and dark mist are again prominent, as is ’separation off’, this time caused by the sun’s heat, which accounts for wind, clouds and (if ’separation off’ of dry from wet involves ’exhalation’ and ’drying’) rain, as well as other effects, including the eventual desiccation of the earth. Wind in turn is the cause of thunder, lightning, thunderbolts, waterspouts and hurricanes.

Anaximander took the same approach to the problem of the origin of life. ’The first animals were born in moisture, surrounded by thorny barks, and when they grew older they went forth onto drier regions, and when the bark broke off they lived a different life for a brief time’ (Aëtius, V 19.4, A30). Animals have a similar origin to the *kosmos* (the repetition of the image of a tree’s bark is hardly coincidental) and one linked to the progressive drying out of the sea. Regarding the beginning of human life, ’from water and earth that had been heated arose fishes or animals very similar to fishes; humans grew in these and remained inside until puberty; then at last they burst and men and women came forth already able to feed themselves’ (Censorinus, 4.7, A30). Although this is not an anticipation of the theory of evolution (as some have thought), it gives a typically bold solution to the problem of how the first generation of humans could have survived until they could take care of themselves and reproduce.

Two other issues are whether our *kosmos* will last forever and whether it is unique. Regarding the first, Anaximander’s fragment can be taken to imply that the inconclusive war between opposites will continue forever, so that our *kosmos* has a beginning but no end. However, there is no trace of such a view in the sources and, since later Greek philosophers strongly rejected such asymmetries, the silence indicates that Anaximander did not maintain the view explicitly. On the other hand, since it is unclear how the *kosmos* would come to an end, it is likely that he did not discuss the matter in any detail.

Regarding the second issue, most scholars hold either that Anaximander considered the *kosmos* unique or that he posited an infinite succession of *kosmoi*, one after another. However, several sources mention a limitless number of coexistent *kosmoi*. The evidence is not unanimous and can be taken in different ways, but it would arguably be consistent with Anaximander’s system and with his use of the Principle of Sufficient Reason for him to maintain

that the process which formed our world need not occur only at one time or in one place.

4 Anaximander's fragment

Aside from a few words in the doxography which may be original, we possess one fragment less than a sentence long. Even here it is disputed which words are original and which are paraphrase. The passage reads as follows (the material in italics is widely thought to be genuine): 'Things that are perish into the things from which they arise, *according to necessity. For they pay penalty and retribution to one another for their injustice, according to the ordinance of time*'. Simplicius, who quotes the fragment in his text *On Aristotle's Physics* (24.18, fr.1), says it describes the relation between things in the world and the *apeiron*, but on grammatical as well as systematic grounds most scholars think it gives Anaximander's view of events in the world without reference to the *apeiron*. Day and night, summer and winter, and many other phenomena involve the regular alternation of the preponderance of one opposite over another. The opposites here are hot and cold, wet and dry, light and dark, etc, conceived not as properties of substances but as 'powers' (*dynameis*) which are capable of affecting things and are embodied or contained in the substances characterized by them. The *kosmos*, then, is a battleground in a war between opposites, a dynamic equilibrium in which one invades the other's territory but is repulsed and loses some of its own ground in turn. On this interpretation the fragment accords with the view favoured above, that the *apeiron* plays no part in the ongoing functioning of the *kosmos*; it was needed at the beginning of the *kosmos*, but, because of the way the *kosmos* was generated, things go on without further need of it.

Noteworthy is the legal language, which Anaximander probably meant literally, since he and other Presocratics did not contemplate any radical difference between humans and the rest of the *kosmos* that would make certain concepts and vocabulary appropriate to one but not the other. Also significant is the idea that events in the world are governed by necessity and occur in an inevitable sequence - a revision of traditional beliefs about the powers of the Olympian gods.

Anaximander brilliantly extended Thales' approach to the understanding of the world to explain a wide range of phenomena. He saw the *kosmos* as a place of order, balance and symmetry; also of change and conflict, subordinated to larger-scale patterns of stability. The testimonia reveal him as a pioneer in several fields of science, but at least as important is his unprecedented use of abstract and general considerations in his reasoning, together with his observations of the world, to articulate a largely unified and coherent understanding of it. His most striking theories were too bold for his immediate successors, but his goal of constructing an intelligible account of the history and workings of the *kosmos* remained as his legacy to philosophy and science.

See also: Archē; Presocratic philosophy

RICHARD McKIRAHAN

References and further reading

Anaximander (*c.*610-after 546 BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 1, 67-81.(The standard collection of the ancient sources, both fragments and testimonia, the latter designated by 'A'; includes Greek texts with translations in German.)

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge & Kegan Paul.(Chapter 2 contains an examination of Anaximander's arguments; good bibliography.)

Classen, C.J. (1965) 'Anaximander', in A. Pauly, G. Wissowa and W. Kroll (eds) *Realencyclopädie der Altertumswissenschaft*, Stuttgart: Druckenmueller, suppl. vol. 12, cols 30-69.(Detailed and exhaustive discussion of the evidence; German text.)

Furley, D.J. (1987) *The Greek Cosmologists* vol. 1, Cambridge: Cambridge University Press.(Pages 16-30 give a general treatment of Milesian cosmology with special reference to Anaximander; argues for some non-traditional interpretations.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols. (Volume 1 pages 72-115 offer a thorough discussion of Anaximander; good bibliography.)

Hölscher, U. (1953) 'Anaximander und die Anfänge der Philosophie', Hermes 81: 257-77 and 385-418; trans. D.J. Furley and R.E. Allen, 'Anaximander and the Beginnings of Greek Philosophy', in D.J. Furley and R.E. Allen (eds) *Studies in Presocratic Philosophy*, 1970, vol. 1, 281-323.(Considers the source materials on Anaximander and locates him in relation to Hesiod and to the other Milesians.)

Kahn, C.H. (1960) *Anaximander and the Origins of Greek Cosmology*, New York: Columbia University Press; repr. Indianapolis, IN: Hackett, 1994.(The best book on the Milesians.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1982) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(Chapter 3 presents a general treatment of Anaximander, based on texts which are given in the original and in English translation.)

McKirahan, R.D. (1994) *Philosophy before Socrates*, Indianapolis, IN: Hackett.(Translation and discussion of source materials; Chapter 5 covers Anaximander.)

Anaximenes (6th century BC)

The Greek philosopher Anaximenes of Miletus followed Anaximander in his philosophical and scientific interests. Only a few words survive from his book, but there is enough other information to give us a picture of his most important theories. Like the other early Presocratic philosophers he was interested in the origin, structure and composition of the universe, as well as the principles on which it operates. Anaximenes held that the primary substance - both the source of everything else and the material out of which it is made - is air. When rarefied and condensed it becomes other materials, such as fire, water and earth. The primordial air is infinite in extent and without beginning or end. It is in motion and divine. Air generated the universe through its motion, and continues to govern it. The human soul is composed of air and it is likely that Anaximenes believed the entire kosmos *(world) to be alive, with air functioning as its soul. Like other Presocratics, he proposed theories of the nature of the heavenly bodies and their motions, and of meteorological and other natural phenomena.*

1 Air, the cosmic principle

Practically nothing is known of Anaximenes' life. He is called the associate and the pupil of Anaximander, on which inadequate basis the ancient testimony makes him a generation younger than his teacher. He wrote a book in 'plain and simple Ionian dialect', but like the works of the other Presocratics it has not survived, although it must be the source of the few words the sources assign to Anaximenes.

From Aristotle onwards, Thales, Anaximander and Anaximenes have been accepted as working within a single tradition: they tackled a single set of problems and used the same methods to solve them (see Thales; Anaximander). These three Milesian philosophers initiated the Greek philosophical tradition by investigating the origin, structure and workings of the *kosmos* by rational means, abandoning traditional religious and mythological accounts in favour of observation and reason. It is commonly accepted that Anaximander designed important elements of his system to avoid objections he found to Thales' views, and that Anaximenes returned to a theory closer to Thales' but one that both avoided Anaximander's objections and did not suffer from faults that can be found in Anaximander's theory. On this standard view, Milesian philosophy made progress from philosopher to philosopher, and the progress was due to the use of rational criticism. While this line of interpretation is plausible for certain central elements in the thought of these three thinkers, most obviously for their views on the nature of the basic substance, it does not give an adequate picture of other facets of their wide-ranging theories. Also, it is an interpretative construct, not one based on original texts. The sources tell to some extent what the three men's theories were, but not why they adopted them.

Where Anaximander postulated a substance of indeterminate nature, unlike anything found in the world around us, as the ultimate source of the diversity of materials found in the *kosmos*, Anaximenes returned to Thales' idea of basing the *kosmos* on a single familiar substance. However, where Thales selected water, Anaximenes chose air. According to the standard interpretation, Anaximander rejected water because there are materials in the world, such as fire, that lack the properties of water (indeed, they have opposite properties), and so could not have arisen from or be made of water. Thus, the basic material must lack all perceptible properties. Anaximenes objected to Anaximander's idea of founding the world on something imperceptible, indefinite and unfamiliar, and returned to a form of material monism, but one not open to Anaximander's objection. Anaximenes held that air in its most 'even' state (as on a clear, dry, windless day when the air seems neither hot nor cold) is imperceptible, but when it becomes more rare or more dense it turns into other kinds of materials. 'When thinned out it becomes fire, and when thickened it becomes wind, then cloud, and when thickened still more, water, then earth, then stones, and the rest come from these' (Theophrastus, quoted by Simplicius, *On Aristotle's Physics* 24.29-31, A5). The mechanisms of rarefaction and condensation engender a more sophisticated monism, according to which fire, for example, is not really different from air, but *is* air in a certain condition, just as ice is not really different from water, but is water in a certain state. 'And the rest come from these' suggests (although this issue is disputed) that air occurs in certain basic forms (the ones listed above), and that other substances, such as wood or wheat, arise through some combination of them. If Anaximenes held this 'proto-chemical' view of the composition of entities, the absence of further discussion in the sources makes it unlikely that he developed the idea through analysis of particular kinds of materials or illustrative examples of other sorts.

Anaximenes was the first reductionist. From the familiar phenomenon that when we exhale through pursed lips our breath feels cold, but when we exhale with the mouth wide open it feels warm, he concluded generally that compression is the cause of cooling and rarefaction the cause of heating, so that hot and cold depend on rare and dense (Plutarch, *The Principle of Cold* 947F, fr.1). The observation on which this conclusion is based is hardly a scientific experiment (although it has sometimes been called the first one), but there are many elements of scientific reasoning in the way Anaximenes begins from a repeatable result and generalizes it to an explanatory principle related to a larger theory, and also in the way he accounts for one range of phenomena in terms of another which he considers more basic. The fact that his conclusion is exactly wrong is less important than these other considerations.

Density and rarity are quantitative concepts - more or less of something in a given volume - while hot and cold are qualities. Anaximenes is thus credited with being the first to subscribe to the scientific objective (also pursued shortly afterwards by the Pythagoreans) of reducing qualities to quantities (see Pythagoreanism §2).This is probably a misinterpretation however. For although we define density in quantitative terms, Anaximenes need not have seen it this way. There is no reason to suppose he thought there were fixed proportions governing the changes of air, so that, for example, so much fire contains the same amount of air as so much water. He held that ice is denser than water (Aëtius, III 4.1, fr. 3), but is there more water in a cup full of ice than in a cup full of water? (Indeed, water at 4° Celsius is denser than ice, but Anaximenes was unaware of this fact.) The evidence suggests that he did not take a quantitative view of these matters, but considered air, clouds, rain, etc. as more or less dense, as if rarity and density were qualities apparent to the senses and not dependent on measurement.

Like Anaximander’s originative substance, Anaximenes’ air is described as eternally in motion and also as *apeiron* (’boundless’ or ’unlimited’;the word later meant ’infinite’, but in Anaximenes’ time did not yet have any precise mathematical sense), meaning that it has no limits in space or time. Not that all air is in motion all the time, since air in motion is wind, and wind is denser than air in its most ’even’ state. Thus, air is mobile, and at any given time, much of it is in motion. Moving through its own nature, it is alive, and being everlasting, it is divine. (These were also Anaximander’s reasons for regarding his basic substance as divine.)

2 The *kosmos*

Like other Presocratic philosophers, Anaximenes proposed a cosmogony and a cosmology that fitted his basic principles. Little information on these topics survives and some of it seems to be inconsistent. Our *kosmos* originated out of the self-moving air which became so dense in one region that it was ’felted’ (the word may be Anaximenes’ own), or compressed, to form earth. The earth is a flat ’table-like’ disc and so rides on the air beneath it like a leaf. There are different accounts of the heavenly bodies. According to the most plausible one, as the earth grew more dense there arose from it moisture that became rarefied and turned into fire; the fire went aloft and became the heavenly bodies. The stars do not give heat because of their great distance from the earth. Like the earth, these fiery bodies stay aloft because of their flatness. There are also earthen bodies in the heavens, which may be posited to account for eclipses. The heavenly bodies do not move underneath the earth, but around it, as when a felt cap (shaped like a yarmulke) is turned about its centre while on someone’s head. The geometry of this model is difficult to work out, but the simple and homespun analogy is characteristic of Anaximenes.

Also typical of the Presocratics is Anaximenes’ interest in meteorological phenomena. Wind, clouds, rain, etc. are condensed air. When rain freezes as it falls it becomes hail, and it becomes snow when ’something of the nature of breath’ is combined with the moisture. Anaximenes explains the rainbow, traditionally the goddess Iris, as the effect of the sun’s rays striking condensed air, and offers an account of why it has different colours. Lightning and earthquakes (the work of Zeus and Poseidon) likewise receive naturalistic explanations.

To our knowledge, Anaximenes did not discuss the origin and conditions of animal life, although this may be due to the meagre evidence we possess about his ideas. Certainly these issues were prominent with Anaximander. Anaximenes did declare that the gods are made of air, thus decisively subordinating the Olympian religion to physics. A purported quotation bears on the nature of the soul, a topic which as far as we know Anaximander did not discuss: ’As our soul, being air, holds us together and controls us, also breath and air surrounds the whole *kosmos*’ (fr. 2). This fascinating statement (some of whose language is almost certainly a rewording of the original) starts from the traditional, non-philosophical idea that the soul, or principle of life, is the breath. The soul maintains us in existence (’holds us together’) as living beings and governs our activities (’controls’), although

how it governs and what activities it controls are unclear. (So, for example, we cannot tell whether the soul has cognitive or moral aspects, or whether creatures other than humans possess souls.) The point of the analogy must be that 'breath and air' has (note the singular verb, as in the Greek of the fragment; 'breath and air' is a single idea, perhaps equivalent to 'breath in the form of air') similar functions in the *kosmos*. As it stands, the fragment is not an argument. It is an analogy, which despite its lack of probative force may reveal one of the insights that led Anaximenes to choose air as his basic substance or one that confirmed his choice. The fragment contains the first explicit use of the microcosm-macrocosm analogy in Greek thought. It suggests (although it is not certain that Anaximenes held this view) that air is the soul of the *kosmos*, and therefore that the entire universe is alive. This idea resembles the view attributed to Thales that all things possess soul through their connection with the divine originative substance water. Anaximenes, then, may have found air a more plausible basic substance than water because of the animating functions it possesses according to the traditional conception of the 'breath-soul'.

The last of the great Milesian thinkers, Anaximenes shared with his predecessors an interest in the natural world and the goal of explaining its principal features acceptably to human reason. His cosmology influenced many of the later Presocratics. More conservative than Anaximander in some ways, he devised a theory of the *kosmos* that was intellectually justifiable by the standards of his time and also closer to common experience.

See also: Archē; Doxography

RICHARD McKIRAHAN

References and further reading

Anaximenes (6th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 1, 90-6.(The standard collection of the ancient sources; includes Greek texts with translations in German.)

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge & Kegan Paul. (Chapter 3 contains an examination of Anaximenes' arguments; good bibliography.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols. (Volume 1 pages 115-40 offers a thorough discussion of Anaximenes; good bibliography.)

Kahn, C.H. (1960) *Anaximander and the Origins of Greek Cosmology*, New York: Columbia University Press; repr. Indianapolis, IN: Hackett, 1994.(The best book on the Milesians.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1982) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(Chapter 4 offers a general treatment of Anaximenes based on texts which are given in the original and in English translation.)

McKirahan, R.D. (1994) *Philosophy before Socrates*, Indianapolis, IN: Hackett.(Translation and discussion of source materials; Chapter 6 covers Anaximenes.)

Ancient philosophy

The philosophy of the Greco-Roman world from the sixth century BC to the sixth century AD laid the foundations for all subsequent Western philosophy. Its greatest figures are Socrates (fifth century BC) and Plato and Aristotle (fourth century BC). But the enormously diverse range of further important thinkers who populated the period includes the Presocratics and Sophists of the sixth and fifth centuries BC; the Stoics, Epicureans and sceptics of the Hellenistic age; and the many Aristotelian and (especially) Platonist philosophers who wrote under the Roman Empire, including the great Neoplatonist Plotinus. Ancient philosophy was principally pagan, and was finally eclipsed by Christianity in the sixth century AD, but it was so comprehensively annexed by its conqueror that it came, through Christianity, to dominate medieval and Renaissance philosophy. This eventual symbiosis between ancient philosophy and Christianity may reflect the fact that philosophical creeds in late antiquity fulfilled much the same role as religious movements, with which they shared many of their aims and practices.

Only a small fraction of ancient philosophical writings have come down to us intact. The remainder can be recovered, to a greater or lesser extent, by piecing together fragmentary evidence from sources which refer to them.

1 Main features

'Ancient' philosophy is that of classical antiquity, which not only inaugurated the entire European philosophical tradition but has exercised an unparalleled influence on its style and content. It is conventionally considered to start with Thales in the mid sixth century BC, although the Greeks themselves frequently made Homer (*c*.700 BC) its true originator. Officially it is often regarded as ending in 529 AD, when the Christian emperor Justinian is believed to have banned the teaching of pagan philosophy at Athens. However, this was no abrupt termination, and the work of Platonist philosophers continued for some time in self-imposed exile (see Aristotle commentators; Neoplatonism §1; Simplicius §1).

Down to and including Plato (in the first half of the fourth century BC), philosophy did not develop a significant technical terminology of its own - unlike such contemporary disciplines as mathematics and medicine. It was Plato's pupil Aristotle, and after him the Stoics (see Stoicism), who made truly decisive contributions to the philosophical vocabulary of the ancient world.

Ancient philosophy was above all a product of Greece and the Greek-speaking parts of the Mediterranean, which came to include southern Italy, Sicily, western Asia, and large parts of North Africa, notably Egypt. From the first century BC, a number of Romans became actively engaged in one or other of the Greek philosophical systems, and some of them wrote their own works in Latin (see Lucretius; Cicero; Seneca; Apuleius). But Greek remained the *lingua franca* of philosophy. Although much modern philosophical terminology derives from Latinized versions of Greek technical concepts, most of these stem from the Latin vocabulary of medieval Aristotelianism, not directly from ancient Roman philosophical writers.

2 The sixth and fifth centuries BC

The first phase, occupying most of the sixth and fifth centuries BC, is generally known as Presocratic philosophy (see Presocratic philosophy). Its earliest practitioners (Thales; Anaximander; Anaximenes) came from Miletus, on the west coast of modern Turkey. The dominant concern of the Presocratic thinkers was to explain the origin and regularities of the physical world and the place of the human soul within it (see especially Pythagoreanism; Heraclitus; Anaxagoras; Empedocles; Democritus), although the period also produced such rebels as the Eleatic philosophers (Parmenides; Zeno of Elea; Melissus), whose radical monism sought to undermine the very basis of cosmology by reliance on a priori reasoning.

The label 'Presocratic' acknowledges the traditional view that Socrates (469-399 BC) was the first philosopher to shift the focus away from the natural world to human values. In fact, however, this shift to a large extent coincides with the concerns of his contemporaries the Sophists, who professed to teach the fundamentals of political and social success and consequently were also much concerned with moral issues (see Sophists). But the persona of Socrates became, and has remained ever since, so powerful an icon for the life of moral scrutiny that it is his name that is used to mark this watershed in the history of philosophy. In the century or so following his death, many

schools looked back to him as the living embodiment of philosophy and sought the principles of his life and thought in philosophical theory (see especially Socratic schools).

3 The fourth century BC

Socrates and the Sophists helped to make Athens the philosophical centre of the Greek world, and it was there, in the fourth century, that the two greatest philosophers of antiquity lived and taught, namely Plato and Aristotle. Plato, Socrates' pupil, set up his school the Academy in Athens (see Academy). Plato's published dialogues are literary masterpieces as well as philosophical classics, and develop, albeit unsystematically, a global philosophy which embraces ethics, politics, physics, metaphysics, epistemology, aesthetics and psychology.

The Academy's most eminent alumnus was Aristotle, whose own school the Lyceum came for a time to rival its importance as an educational centre. Aristotle's highly technical but also often provisional and exploratory school treatises may not have been intended for publication; at all events, they did not become widely disseminated and discussed until the late first century BC. The main philosophical treatises (leaving aside his important zoological works) include seminal studies in all the areas covered by Plato, plus logic, a branch of philosophy which Aristotle pioneered. These treatises are, like Plato's, among the leading classics of Western philosophy.

Platonism and Aristotelianism were to become the dominant philosophies of the Western tradition from the second century AD at least until the end of the Renaissance, and the legacy of both remains central to Western philosophy today.

4 Hellenistic philosophy

Down to the late fourth century BC, philosophy was widely seen as a search for universal understanding, so that in the major schools its activities could comfortably include, for example, biological and historical research. In the ensuing era of Hellenistic philosophy, however, a geographical split helped to demarcate philosophy more sharply as a self-contained discipline (see Hellenistic philosophy). Alexandria, with its magnificent library and royal patronage, became the new centre of scientific, literary and historical research, while the philosophical schools at Athens concentrated on those areas which correspond more closely to philosophy as it has since come to be understood. The following features were to characterize philosophy not only in the Hellenistic age but also for the remainder of antiquity.

The three main parts of philosophy were most commonly labelled 'physics' (a primarily speculative discipline, concerned with such concepts as causation, change, god and matter, and virtually devoid of empirical research), 'logic' (which sometimes included epistemology), and 'ethics'. Ethics was agreed to be the ultimate focus of philosophy, which was thus in essence a systematized route to personal virtue (see Aretē) and happiness (see Eudaimonia).There was also a strong spiritual dimension. One's religious beliefs - that is, the way one rationalized and elaborated one's own (normally pagan) beliefs and practices concerning the divine - were themselves an integral part of both physics and ethics, never a mere adjunct of philosophy.

The dominant philosophical creeds of the Hellenistic age (officially 323-31 BC) were Stoicism (founded by Zeno of Citium) and Epicureanism (founded by Epicurus) (see Stoicism; Epicureanism). Scepticism was also a powerful force, largely through the Academy (see Arcesilaus; Carneades), which in this period functioned as a critical rather than a doctrinal school, and also, starting from the last decades of the era, through Pyrrhonism (see Pyrrhonism)

5 The imperial era

The crucial watershed belongs, however, not at the very end of the Hellenistic age (31 BC, when the Roman empire officially begins), but half a century earlier in the 80s BC. Political and military upheavals at Athens drove most of the philosophers out of the city, to cultural havens such as Alexandria and Rome. The philosophical institutions of Athens never fully recovered, so that this decentralization amounted to a permanent redrawing of philosophical map. (The chairs of Platonism, Aristotelianism, Stoicism and Epicureanism which the philosopher-emperor Marcus Aurelius (§1) established at Athens in AD 176 were a significant gesture, but did not fully restore Athens' former philosophical pre-eminence.) Philosophy was no longer, for most of its adherents, a living activity within the Athenian school founded by Plato, Aristotle, Zeno or Epicurus. Instead it was a subject pursued in small study groups led by professional teachers all over the Greco-Roman world. To a large extent, it was felt that the history

of philosophy had now come to an end, and that the job was to seek the correct interpretation of the ‘ancients’ by close study of their texts. One symptom of this feeling is that doxography - the systematic cataloguing of philosophical and scientific opinions (see Doxography) - concentrated largely on the period down to about 80 BC, as did the biographical history of philosophy written *circa* AD 300 by Diogenes Laertius.

Another such symptom is that a huge part of the philosophical activity of late antiquity went into the composition of commentaries on classic philosophical texts. In this final phase of ancient philosophy, conveniently called ‘imperial’ because it more or less coincides with the era of the Roman empire, the Hellenistic creeds were gradually eclipsed by the revival of doctrinal Platonism, based on the close study of Plato’s texts, out of which it developed a massively elaborate metaphysical scheme. Aristotle was usually regarded as an ally by these Platonists, and became therefore himself the focus of many commentaries (see Platonism, Early and Middle; Peripatetics; Neoplatonism; Aristotle Commentators). Despite its formal concern with recovering the wisdom of the ancients, however, this age produced many powerfully original thinkers, of whom the greatest is Plotinus.

6 Schools and movements

The early Pythagoreans constituted the first philosophical group that can be called even approximately a ‘school’. They acquired a reputation for secrecy, as well as for virtually religious devotion to the word of their founder Pythagoras. ‘He himself said it’ (best known in its Latin form ‘*ipse dixit*’) was alleged to be their watchword. In some ways it is more accurate to consider them a sect than a school, and their beliefs and practices were certainly intimately bound up in religious teachings about the soul’s purification.

It is no longer accepted, as it long was, that the Athenian philosophical schools had the status of formal religious institutions for the worship of the muses. Their legal and institutional standing is in fact quite obscure. Both the Academy and the Lyceum were so named after public groves just outside the walls of Athens, in which their public activities were held. The Stoics too got their name from the public portico, or ‘stoa’, in which they met, alongside the Athenian agora. Although these schools undoubtedly also conducted classes and discussions on private premises too, it was their public profile that was crucial to their identity as schools. In the last four centuries BC, prospective philosophy students flocked to Athens from all over the Greek world, and the high public visibility of the schools there was undoubtedly cultivated partly with an eye to recruitment. Only the Epicurean school kept its activities out of the public gaze, in line with Epicurus’ policy of minimal civic involvement.

A school normally started as an informal grouping of philosophers with a shared set of interests and commitments, under the nominal leadership of some individual, but without a strong party line to which all members owed unquestioning allegiance. In the first generation of the Academy, for example, many of Plato’s own leading colleagues dissented from his views on central issues. The same openness is discernible in the first generations of the other schools, even (if to a much lesser extent) the Epicurean. However, after the death of the founder the picture usually changed. His word thereafter became largely beyond challenge, and further progress was presented as the supplementation or reinterpretation of the founder’s pronouncements, rather than as their replacement.

To this extent, the allegiance which in the long term bound a school together usually depended on a virtually religious reverence for the movement’s foundational texts, which provided the framework within which its discussions were conducted. The resemblance to the structure of religious sects is no accident. In later antiquity, philosophical and religious movements constituted in effect a single cultural phenomenon, and competed for the same spiritual and intellectual high ground. This includes Christianity, which became a serious rival to pagan philosophy (primarily Platonism) from the third century onwards, and eventually triumphed over it. In seeking to understand such spiritual movements of late antiquity as Hermetism, Gnosticism, Neo-Pythagoreanism, Cynicism and even Neoplatonism itself, and their concern with such values as asceticism, self-purificaton and self-divinization, it is inappropriate to insist on a sharp division between philosophy and religion (see Hermetism; Gnosticism; Neo-Pythagoreanism; Cynics §4; Neoplatonism).

‘Ancient philosophy’ is traditionally understood as pagan and distinguished from the Christian Patristic philosophy of late antiquity (see Patristic philosophy). But it was possible to put pagan philosophy at the service of Judaism (see Philo of Alexandria), or Christianity (see for example Clement of Alexandria; Origen; Augustine; Boethius; Philoponus), and it was indeed largely in this latter capacity that the major systems of ancient philosophy eventually became incorporated into medieval philosophy and Renaissance philosophy, which they proceeded to

dominate (see Medieval philosophy; Renaissance philosophy).

This extensive overlap between philosophy and religion also reflects to some extent the pervasive influence of philosophy on the entire culture of the ancient world. Rarely regarded as a detached academic discipline, philosophy frequently carried high political prestige, and its modes of discourse came to infect disciplines as diverse as medicine, rhetoric, astrology, history, grammar and law. The work of two of the greatest scientists of the ancient world, the doctor Galen and the astronomer Ptolemy, was deeply indebted to their respective philosophical backgrounds.

7 Survival

A very substantial body of works by ancient philosophical writers has survived in manuscript. These are somewhat weighted towards those philosophers - above all Plato, Aristotle and the Neoplatonists - who were of most immediate interest to the Christian culture which preserved them throughout the Middle Ages, mainly in the monasteries, where manuscripts were assiduously copied and stored. Some further ancient philosophical writings have been recovered through translations into Arabic and other languages, or on excavated scraps of papyrus. The task of reconstituting the original texts of these works has been a major preoccupation of modern scholarship.

For the vast majority of ancient philosophers, however, our knowledge of them depends on secondary reports of their words and ideas in other writers, of whom some are genuinely interested in recording the history of philosophy, but others bent on discrediting the views they attribute to them. In such cases of secondary attestation, strictly a ‘fragment’ is a verbatim quotation, while indirect reports are called ‘testimonia’. However, this distinction is not always rigidly maintained, and indeed the sources on which we rely rarely operate with any explicit distinction between quotation and paraphrase.

It is a tribute to the philosophical genius of the ancient world that, despite the suppression and distortion which its contributions have suffered over two millennia, they remain central to any modern conspectus of what philosophy is and can be.

See also: Egyptian philosophy: influence on Greek thought; Atomism, ancient; Archē; Greek philosophy: impact on Islamic philosophy; Language, ancient philosophy of; Logic, ancient; Owen, G.E.L.; Vlastos, G.; Logos; Nous; Pneuma; Psychē.

DAVID SEDLEY

References and further reading

Algra, K., Barnes, J., Mansfeld, J. and Schofield, M. (eds) (1998) *The Cambridge History of Hellenistic Philosophy*, Cambridge: Cambridge University Press.(Bridges the gap between Guthrie (1962-81) and Armstrong (1967).)

Armstrong, A.H. (ed.) (1967) *The Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press.(Continuation of Guthrie (1962-81), but skipping Hellenistic philosophy.)

Guthrie, W.K.C. (1962-81) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols.(The major work of its kind in English, but does not go beyond Aristotle.)

Zeyl, D. (1997) *Encyclopedia of Ancient Philosophy*, Westport, CT: Greenwood Press.(An comprehensive guide to the Greek and Roman philosophers.)

Anderson, John (1893-1962)

Arguing against metaphysical 'ultimates' (that is, supposed unconditioned conditions of things), relative truth, appeals to subjective experience, and opposing some of the main tendencies of twentieth-century philosophy, Anderson developed a wide-ranging realist and empiricist philosophy. Highly critical of religion, he was much concerned with other cultural values and advanced views (influential in Australia) on freedom of thought, education, ethics and aesthetics. In ethics, for example, his view is that objective good is not good because it is approved of by certain people; rather those who approve of good (or have other relations to it) do so because it is *good. He carefully distinguished questions about the intrinsic character of good from those about relations social groups may have to it, and goes on to develop an account of intrinsic goods as certain socio-mental activities: enterprise or freedom, objective inquiry, artistic production and appreciation, love and courage.*

Similarly, in aesthetics he distinguishes characteristics of works of art from possible relations between artists, works, appreciators and critics, such as a work's relation to a writer's intentions. In Anderson's view the character and structure of the work itself alone provides an aesthetic criterion for assessing the merit of works of art.

1 Life

John Anderson, Scottish-born philosopher, studied at Glasgow University and lectured there and in Cardiff and Edinburgh before becoming professor of philosophy at Sydney University (1927-58) where he became the leading Australian philosopher of his time, and also a leading controversialist. He was twice censured by the New South Wales Parliament, but was also named as one of the official 200 Australian 'Greats' at the 1988 Bicentenary.

Anderson was influenced by idealist teachers at Glasgow and subsequently in realist ways by William James, G.E. Moore, the New Realists, Kemp Smith, and above all Samuel Alexander. He was also influenced by Heraclitus, Plato, Sorel, Freud, Marx, Vico and Croce, and on wider issues by *The New Age*, a journal of critical thought, during its editorship by A.R. Orage.

A charismatic teacher, Anderson influenced people in a remarkable number of fields, including about thirty who became professional philosophers. These varied in the extent to which they used or agreed with his views. For example, J.A. Passmore, J.L. Mackie and D.M. Armstrong produced well-known work of their own (although Mackie did expand Anderson's views on causes and on hypotheticals), while others, such as A.R. Walker, T.A. Rose, A.J. Anderson and W.V. Doniela, taught and developed views more in the Andersonian tradition. In Britain, Anderson influenced Rush Rhees at Edinburgh and through him his students at Swansea.

Anderson's political orientations, though not his main social theory, changed over the years. At first a communist sympathizer (he mistakenly believed Soviet workers were exhibiting Sorel's 'ethic of the producer'), he was then a Trotskyist between 1933 and 1937, and finally a critical oppositionist concerned to defend pluralist values and 'expose illusions' wherever they are found. In the 1950s he criticized communism and the welfare 'servile state' while tenaciously defending learning and thinking values in education against practicalist and supposedly 'egalitarian' ones.

2 Systematic realism

In Anderson's view there is a certain dogmatism about advancing any philosophic position; nevertheless, support can be offered by showing confusions and inconsistencies in rival views and, as Socrates suggested, by using a hypothetical method to reveal the consequences of your position. Criticism is integral to Anderson's philosophizing; not someone who relies on announcing his 'intuitions', he is one of the leading philosophical 'arguers'.

For brevity, Anderson's position is best stated in terms of the '-isms' he sometimes employed. It is mainly one of realism, objectivism, empiricism and pluralism - interlocking views which may be summed up as 'systematic realism'. Conceptions of relative truth are rejected. There may be much illusory thinking, but attempts to deny the absolute or objective truth of *all* propositions are self-refuting - ' If I say "X is true for me", then I am saying that X's being true for me is an absolute fact' (1962: 294). Anderson likewise rejects, for instance, antirealist claims

that unverified propositions about the past which are unverifiable today are not true or false.

Contrary to idealists and phenomenalists, he holds that the knower and the known have independent status. Opposing views, as in conceptions of ‘dependent existences’ such as ideas, perceptions or sense-data, involve confusions of qualities with relations, and Anderson calls such views ‘relativist’. In a relational situation in which *A* has *r* to *B*, neither *A* nor *r* nor *B* ontologically constitutes any of the others, even when, say, *B* depends on *A* for its origin or continued existence. But in the case of knowledge, relativist confusions are common and are facilitated by cognate accusatives, talk of intentional objects and other ambiguities in the use of such words as ‘perception’ and ‘experience’. Confusions about qualities and relations are also notably rife in ethics and aesthetics (see Realism and antirealism; Relativism).

Anderson’s realism and empiricism (which differs from traditional empiricism) rejects overt and covert rationalistic attempts to set up ‘levels of reality’ in the shape of philosophical ‘ultimates’ such as necessary or divine beings, monism’s one, dualism’s two, or the units of atomistic philosophers (for example, Hume’s perceptions and the simples of Russell and the early Wittgenstein). Arguing against such views - that ‘monism explodes into dualism’, though dualism cannot account for the relations between its two sides, and that atomism fails to reduce complexity to unitary elements - Anderson is an ontological egalitarian: whatever exists is on *the same level of existence* as anything else, and is open to objective, empirical investigation, but there are no underlying ultimate, final or purposive explanations.

In Anderson’s empiricism all knowledge is obtained by observation (and experiment), including introspection, but observation is not indubitable. In any field there is the possibility of discovery and also of error, which arises because of the ‘pluralist complexity’ both of mind and of non-mental things. Mind is a network of varied and conflicting tendencies: because of this and because of Freudian mechanisms we are prone to *mis*observe as well as observe. He sums up his pluralism thus: ‘ There is not only an unlimited multiplicity of things to which the single logic of events applies but anything whatever is infinitely complex so that we can never cover its characters in a single formula or say that we know "all about it"’ (1958: 55) (see Pluralism).

3 Logic, categories

Although the different parts of his position are separably arguable for or against, Anderson places special emphasis on a realist ‘logic of events’, including a realist formal logic. While his broad conception of logic is of the conditions of existence, his formal logic is traditional logic developed and made consistently realist. It deals with AEIO propositional forms, their implications and other logical relations. Just as there is one level of existence, there is one level of discourse and nothing is above or below the proposition whose copula is the ‘is’ of occurrence coupling two terms, one of which, the subject, locates an actual or possible situation, and the other of which, the predicate, characterizes it. Identity, existential and hypothetical statements are analysed in AEIO ways, any term is a ‘real term’ - or presupposed to exist in Strawson’s sense - and more precisely is a ‘complex place-character’ which is the subject of some true propositions and predicate of others. The force of the last point - in regard to the problem of universals - is that there are not separate entities, particular or universal; every situation has both particularity and universality (see Ancient logic §3; Strawson, P.F. §§2-3; Universals §3).

Anderson’s acceptance of subject-predicate logic thus puts him at variance with the assumptions of modern logic. But Anderson considers that modern logic and its refinements are derivative from realist logic, in that without prior recognition of the truths of that logic, modern logic would have no objective basis. In any case his logic provides a thoroughgoing realist version of syllogistic logic by requiring all propositions to be contingent and all terms, including their ‘logical opposites’, to be non-empty, and by paying considerable attention to complex terms and to arguments other than syllogistic ones.

The logic is also closely connected with Anderson’s theory of space, time and the categories, which is a revised version of Alexander’s theory, and ties in the categories with the form of the proposition as a kind of realistic parallel to Kant’s ‘metaphysical deduction’ of the categories (see Alexander, S.). In Anderson’s theory, the categories - particularity, universality, causality and so on - are not compartmentalized, but as pervasive conditions of existence they are all present in all situations.

4 Theory of mind, determinism

Anderson takes a materialist view of mind, maintaining that mental phenomena have spatial, temporal, chemical and suchlike properties. He disagrees, however, with the identification of mind and brain (as made, for instance, by Armstrong 1968) on the grounds that this position is excessively physicalist. His own view is that there is a qualitative difference between mental and non-mental brain processes, consisting in the fact that mind is feeling or emotion.

Mental occurrences, he argues, are all caused, but we can distinguish certain *free* mental activities - ones which have a spontaneous, unstrained way of working - from activities which have a forced or compulsive character. Anderson is a 'pluralist determinist', criticizing conceptions of a single transitive chain of causes and the like. There are, he holds, separate, complex series of causal factors, interacting in various ways, as a result of which complete prediction is impossible. That is why his social theory is critical of social planners and would-be social engineers. Though every social occurrence has causal conditions, we cannot often predict what is going to happen, let alone bring it about; and furthermore social planning very often has undesired, unanticipated consequences.

One of his original conceptions concerns the causal 'field' - what the cause acts on to produce the effect, as when a virus causes a person to have influenza. Attention to the field helps clear up problems about necessary and sufficient conditions, causal series, and 'immediate', 'underlying', 'principal' causes and so on, since differing fields are involved.

5 Social thought

Anderson is a pluralist both in holding that society's way of working is pluralist and in espousing pluralist values. Against the claims of social voluntarism and atomism ('methodological individualism') that unitary individuals and subordinate relations between them are the main determinants of social affairs, he argues that 'forms of association are the primary social fact' (Baker 1979: 17) (see Individualism, methodological). These associations, such as movements and institutions, involve 'forms of psycho-social activity' passing through individuals: that is, these are forms of activity that exist in their own right. He likewise argues that solidarist or totalistic conceptions are falsified by the variety of factors which are present, including conflicts of group interests, some of which are irreconcilable. Appeals to the 'good of all', for example, disguise actual conflicts and have the covert effect of confusing issues and furthering certain special interests at the expense of other special interests. Social unity and harmony theorists, in Anderson's view, would be wise to heed what Heraclitus said long ago: 'Homer was wrong in saying: "Would that strife might perish from among gods and men!" He did not see that he was praying for the destruction of the universe; for if his prayer were heard, all things would pass away' (Burnet 1908: Fr. 43, 150).

In the case of Marx and Marxists, Anderson rejects their main totalistic conception of society and history and its associated 'philanthropic, servile and salvationist' ethic, but argues that there is a different, scientific and pluralistic strain in Marx's though which lays the foundation of a 'materialism of a non-totalistic kind' with its 'doctrine of social struggles throughout history'. With regard to freedom and democracy, Anderson argues that for people to be more than 'mere voting machines' what is needed is their active concern with the affairs of society at large and with those institutions to which they belong. Such activity is rarely widespread - all the less so with increasing government bureaucratization - but has enough strength to keep alive a permanent struggle of positive freedom against servility, which has its ups and downs (he here agrees with Vico and Croce) and is gauged by 'the extent of opposition to the ruling order, or criticism of ruling ideas' as carried on by resistant social groups (see Democracy).

Partly because Anderson was little known outside Australia and partly because of the range of his thought, much of his position was neither subjected to informed criticism, nor refined or worked out in detail.

See also: Australia, philosophy in §3

A.J. BAKER

List of works

Anderson, J. (1958) 'Realism', *The Australian Highway* September: 53-6.(An inaccessible journal, but a good short summation of Anderson's position.)

Anderson, J. (1962) *Studies in Empirical Philosophy*, Sydney: Angus & Robertson.(Contains most of Anderson's philosophical articles.)

Anderson, J. (1980) *John Anderson, Education and Inquiry*, ed. D.Z. Phillips, Oxford: Blackwell. (Articles and lecture notes on education.)

Anderson, J. (1982) *John Anderson, Art and Reality*, ed. Janet Anderson *et al.*, Sydney: Hale & Iremonger. (Writings on aesthetics and literature.)

References and further reading

Anderson, A.J. (1987) 'Following John Anderson', *Dialectic* (University of Newcastle Philosophy Club) 30: 129-43.(Deals especially with ethics.)

Armstrong, D.M. (1963) 'Is Introspective Knowledge Incorrigible?', *The Philosophical Review* 72: 417-32. (Defends empiricism of Anderson's kind.)

Armstrong, D.M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul. (Argues for the identification of brain and mind.)

Baker, A.J. (1966) 'Non-empty Complex Terms', *Notre Dame Journal of Formal Logic* 7: 48-56.(Develops Anderson's logic in regard to these terms.)

Baker, A.J. (1977) 'Classical Logical Relations', *Notre Dame Journal of Formal Logic* 18: 164-8.(Presents a consistent, Andersonian analysis.)

Baker, A.J. (1979) *Anderson's Social Philosophy: The Social Thought and Political Life of Professor John Anderson*, Sydney: Angus & Robertson.(As the sub-title indicates, the book elucidates Anderson's social (and ethical) theory, and his political history and public controversies.)

Baker, A.J. (1986) *Australian Realism, The Systematic Philosophy of John Anderson*, Cambridge: Cambridge University Press.(Analyses the main features of Anderson's overall philosophy.)

Burnet, J. (1908) *Early Greek Philosophy*, London: Adam & Charles Black, especially 143-91.(Anderson was much influenced by Burnet's treatment of Heraclitus.)

Mackie, J.L. (1962) 'Counterfactuals and Causal Laws', in R.J. Butler (ed.) *Analytical Philosophy*, Oxford: Blackwell, 66-80.(Develops Anderson's views on hypotheticals.)

Mackie, J.L. (1963) 'Are There Any Incorrigible Empirical Statements?', *Australasian Journal of Philosophy* 41: 12-28.(Defends empiricism of Anderson's kind.)

Mackie, J.L. (1965) 'Causes and Conditions', *American Philosophical Quarterly* 2: 245-64. (Develops views on causation of Anderson's kind.)

McMullen, T. (1996) 'John Anderson on Mind as Feeling', *Theory and Psychology* 6 (1): 153-68.(Defends and develops Anderson's view that feeling or emotion is the distinctive quality of mind.)

Rhees, R. (1947) 'Social Engineering', *Mind* 56: 317-31.(Endorses Anderson's social theory.)

Rybak, J. and Rybak, J. (1984) 'Mechanizing Logic', *Notre Dame Journal of Formal Logic* 25: 250-82.(Shows in an extended Andersonian way how complex and relational arguments can be dealt with by map logic and computer.)

Ryle, G. (1950) 'Logic and Professor Anderson', *Australasian Journal of Philosophy* 28: 137-53.(Gives a lively assessment of Anderson's philosophy, although parts of it are misunderstood.)

Animal language and thought

The question of animal language and thought has been debated since ancient times. Some have held that humans are exceptional in these respects, others that humans and animals are continuous with respect to language and thought. The issue is important because our self-image as a species is at stake.

Arguments for human exceptionalism can be classified as Cartesian, Wittgensteinian and behaviourist. What these arguments have in common is the view that language and thought are closely associated, and animals do not have language. The ape language experiments of the 1960s and 1970s were especially important against this background: if apes could learn language then even the advocates of human exceptionalism would have to admit that they have thoughts. It is now generally believed that whatever linguistic abilities apes have shown have been quite rudimentary. Yet many sceptics are willing to grant that in some cases apes did develop linguistic skills to some extent, and clearly evidenced thought. Studies of other animals in captivity and various animals in the wild have provided evidence of highly sophisticated communicative behaviour. Cognitive ethology and comparative psychology have emerged as the fields that study animal thought. While there are conceptual difficulties in grounding these fields, it appears plausible that many animals have thoughts and these can be scientifically investigated.

1 Human exceptionalism versus continuity across species

Richard Sorabji (1993) has argued that debates about animal language and mind go to the core of the western philosophical tradition. Aristotelians and Stoics argued that only humans have reason or belief; some Platonists and Pythagoreans argued that these are shared by many kinds of animal. Indeed, Plato himself challenged the very framework presupposed by the debate. He thought that it was just as sensible to divide the world into cranes and non-cranes as humans and non-humans (see *Statesman* 263d).

Both human exceptionalism (HE) and continuity across species (CAS) have had strong supporters. HE was defended by Aquinas, Descartes and many twentieth-century linguistic philosophers, CAS by Voltaire, Hume and Darwin.

Although it is easy to characterize HE and CAS generally, it is difficult to do so precisely. Roughly, those who espouse HE believe that humans are unique in having language and sophisticated thought, and that there is a deep chasm between these human capacities and whatever thoughts and communication systems other animals may have. Those who embrace CAS hold that humans share the capacity for language or thought with at least some other animals and that the differences with respect to these capacities between humans and other animals are gradual and incremental.

However advocates of HE need not hold that humans are exceptional with respect to every capacity and every animal. They may contemplate honorary humanhood for dolphins (for example), or grant that a few other animals have thoughts but insist that these thoughts are always first-order or nonconscious, and thus very different from the thoughts that humans are capable of having. Defenders of HE may allow that many animals have communication systems, but then go on to claim that these are vastly weaker and less sophisticated than human language.

Supporters of CAS may grant that humans are capable of having some thoughts that no other animal can have. But they will generally see this as an evolutionary fact about humans that is not importantly different from other evolutionary facts about other animals. Perhaps only humans can ponder the neurophysiology of the wildebeest, but lions can think thoughts about wildebeests that humans cannot conceive. Defenders of CAS sometimes grant that only humans use language, but they often see this as a matter of definition or otherwise trivial. They are inclined to see complex communication systems as similar to languages or as sophisticated and important. Although the difference between HE and CAS may be vague, proponents of HE assert that there are enormous differences between humans and animals that centre on language and thought while advocates of CAS deny this.

The dispute between HE and CAS is important for several reasons. If animals lack central features of language and thought that humans have, then a profound gulf separates us from them. The existence of this gulf may have implications about the relations between the natural sciences and the human sciences. It may also justify discontinuous moralities with respect to humans and animals (see Animals and ethics; Moral standing). If HE is

correct, then we may be justified in seeing ourselves as special - perhaps even as ‘the crown of creation’. On the other hand, if CAS is correct, this may mean that action theory and philosophy of language should be seen as branches of ethology, and that our treatment of animals is a moral scandal. We may have to give up the view of ourselves as morally and metaphysically privileged, and instead see ourselves as one animal species among many. What is potentially at stake in arguments about animal language and thought is our human self-image - who we are, what we are like and what constitutes our proper relations with the rest of nature.

2 Arguments for human exceptionalism

Many philosophers have defended HE. Any attempt to collect these views into categories and to develop generic arguments involves regimentation. With these caveats in mind, it is useful to divide the arguments for HE into three categories: Cartesian, Wittgensteinian and behaviourist.

Cartesian views about animal language and thought have been influential on philosophers and linguists such as Vendler (1972) and Chomsky (1966). Although there is controversy about the exact nature of Descartes’ views about animals, the broad outlines are clear.

Chomsky credits Descartes with recognizing that language use is ‘creative’: it is both unbounded in scope and stimulus-free. Descartes wrote that while ‘magpies and parrots are able to utter words just like ourselves’ this is mechanical, ‘a movement of mere nature’ rather than a sign of thought (*Discourse on Method*, 1637: Part V). Having established to his satisfaction that animals do not have language, Descartes infers that they do not have thought, ‘for the word is the sole sign and the only certain mark of the presence of thought hidden and wrapped up in the body’ (letter to Henry More, 1649). Although Descartes is ambivalent about whether it can be proved that animals do not have thoughts, clearly he believes that they do not. Since animals do not have thoughts they do not have ‘real feeling’, for ‘real feeling’ involves propositional content and animals are incapable of propositional content because they do not have language.

Wittgensteinian accounts of animal language and thought have been given by Malcolm (1972-3) and Leahy (1991). Wittgenstein’s own views are characteristically difficult to unravel. In *Philosophical Investigations* (1953) he claims that animals ‘do not use language - if we except the most primitive form of language’, but appears to think that animals have sensations, emotions, intentions and first-order beliefs. However, Wittgenstein denies that animals have the power to simulate pain, to talk to themselves, or have attitudes about future events.

According to Malcolm, animals think but do not have thoughts. Having thoughts involves formulating and entertaining propositions, and he believes that animals are incapable of this. Although Malcolm (1972-3) does not identify thought with language, he claims that the relationship is ‘so close that it is really senseless to conjecture that people may *not* have thoughts, and also really senseless to conjecture that animals *may* have thoughts’. However, we know that animals think because ‘in real life we commonly employ the verb "think" in respect to animals’. Clearly Malcolm believes that animal thinking does not involve having thoughts, but says very little about how we are to understand it.

Behaviourists in both science and philosophy have denied animal language and thought (see Behaviourism, analytic; Behaviourism, methodological and scientific). Although some behaviourists deny thought to humans as well, most reserve their deepest scepticism for animals. Quine (1960) takes it as obvious that animals do not have language, and that ascribing thoughts to animals is an ‘essentially dramatic idiom’ - we imagine ourselves in the animal’s place and say what thoughts we would have were we the animal.

Davidson (1975) has produced the following argument for supposing that animals do not have thoughts: if an animal has a thought, then this thought must occur in a network of beliefs. This follows from Davidson’s holism which he takes from Quine: thoughts or beliefs come in ‘webs’, they cannot occur singly. In order to have a network of beliefs, an animal must have the concept of belief. This is because having a network of beliefs requires the ability to distinguish between someone holding a sentence to be true and the sentence actually being true. But having the concept of belief requires having language, for Davidson believes that this concept only arises in the context of linguistic interpretation. Since no animals have language they do not have the concept of belief. Since they do not have the concept of belief they do not have networks of beliefs. Hence, they do not have beliefs at all (see Davidson, D. §8).

What is striking about the arguments for HE, taken together, is that they turn on supposing a very close connection between language and thought. All of the arguments that we have reviewed suppose that having language is a necessary condition for having thoughts (although Malcolm grants that some kind of thinking may occur in the absence of language and thought). Moreover, some of these philosophers (for example, Malcolm) think that having language is sufficient for having thoughts.

It is against this background that the ape language experiments of the 1960s and 1970s caught the attention of philosophers. If it could be shown that apes could learn language then many philosophers would be convinced that apes are capable of having thoughts. Indeed, some philosophers would be convinced by nothing short of this.

3 The ape language controversy

Since the beginning of the twentieth century there have been at least half a dozen attempts to teach spoken language to an ape. In 1966 Beatrice and R. Allen Gardner began teaching American Sign Language to a chimpanzee named Washoe, and throughout the late 1960s and 1970s studies employing animals of different species using different communication systems were undertaken by David Premack, Duane Rumbaugh, Roger Fouts, Francine Patterson and others. In some cases extravagant claims were made about the linguistic abilities of these apes. These claims were deflated by Herbert Terrace in the late 1970s.

From 1973 to 1977 Terrace studied a chimpanzee named Nim Chimpsky. He concluded (1980) that there is no unequivocal evidence for supposing that apes can master syntactic, semantic or pragmatic aspects of language. Terrace’s results were devastating to the credibility and funding base of the ape language projects. The projects of Patterson and Fouts continue, funded primarily by private donations. The Rumbaugh project, now headed by Sue Savage-Rumbaugh, is one of the few that continues to produce significant scientific publications.

Beginning in 1981 Savage-Rumbaugh turned her attention to bonobos (so-called ‘pygmy chimpanzees’). Using a specially designed keyboard connected to a speech synthesizer one bonobo, Kanzi, has shown a surprising understanding of spoken English. He has a large vocabulary, is capable of communicating novel information and clearly follows simple syntactic rules. In a comparative study (1994) Savage-Rumbaugh concluded that the eight-year-old Kanzi’s linguistic skills were superior to those of a normal human two-year-old.

The ape language studies have been controversial from the beginning and continue to be so today. Terrace and other critics accuse most researchers of inaccurate observation and analysis, overinterpretation and cuing desired behaviour. Terrace has been faulted for failing to obtain results because of a lack of rapport with his subject and an impersonal training regime.

Animal behaviour, like human behaviour, is indeterminate, description-relative and open to interpretation. John Dupré (1991) has argued that the dispute over the ape language experiments primarily involves conflicts about the goals and methods of scientific research. The critics demand that claims about the linguistic abilities of apes be backed by compelling evidence that conforms to the most rigorous canons of scientific methodology. However, in the case of humans, language develops in a highly complex and emotive context and the application of principles of charity is an important part of the language-learning process (see Charity, principle of §4). It may be that there is an intrinsic conflict between teaching apes language in the most effective way possible and doing so in a way that will satisfy the scientific scruples of the sceptics.

It is clear that the ape language experiments have not convinced philosophers that at least some animals have language and therefore thoughts. However, one point is worth considering. Even sceptics may admit that some apes have demonstrated linguistic capacities to some degree or in some respect. But a similar view with respect to having thoughts is difficult to even understand. Having a thought appears to be an all or nothing matter. This may provide some evidence against the view that language and thought are as closely tied as some philosophers have claimed.

4 Animal communication

Even some scientists and philosophers friendly to the idea that animals have thoughts have been sceptical of the ape language experiments. For these experiments have focused on very few individuals from an even smaller number of species. They have been directed towards eliciting a variety of behaviours that these animals do not

manifest under natural conditions.

In addition to the ape language experiments there have been other interesting studies of the communication abilities of other captive animals (see Bekoff and Jamieson 1995; Ristau 1991). Louis Herman has claimed that bottle-nosed dolphins are capable of semantic and syntactic processing that ‘utilize a rich network of mental representations when responding to language-mediated tasks’, although they are less able to produce semantically and syntactically dense utterances. Irene Pepperberg has trained an African grey parrot to respond accurately to questions about the colour, shape, name and category of a variety of objects.

Under natural conditions many animals engage in highly sophisticated communicative activities. Cheney and Seyfarth have shown that vervet monkeys distinguish several different kinds of alarm call and behave appropriately with respect to each. Carolyn Ristau suggests that the broken wing displays in various species of plovers may be intended to lead intruders away from their nests. Deception has been claimed for a variety of animals including chimpanzees and elephants.

5 Thought without language

As we have come to know more about animals, the possibility that they may have thoughts without language has become increasingly plausible to many people. Even Joel Wallman (1992), one of the most thorough-going critics of the ape language experiments, holds that apes are ‘highly intelligent’, ‘reflexive’ and capable of ‘impressively abstract mentation’. Cognitive ethology and comparative psychology have emerged as the scientific fields that study the cognitive capacities of animals that may underlie some of their behaviour.

The dominant view in cognitive ethology, as well as among some philosophers who have written about cognitive ethology, is that an animal’s cognitive states must be inferred from its behaviour. Since an animal’s mental states are never observed but only inferred, the claim that an animal has a particular mental state or cognitive capacity can only be more or less probable. Often it is said even by cognitive ethologists that we can never really know that animals have thoughts.

In my view, this way of approaching the question of animal thought is fundamentally flawed. If by hypothesis the existence of animal thoughts can only be known by inference from behaviour, then it will always be an open question whether or not animals have thoughts. But if there are no convincing conceptual arguments for supposing that they do not, then we are justified in approaching the question of animal thoughts in the way that we approach the question of human thoughts. For there is no behaviour that an animal can engage in that will ineluctably drive us to the conclusion that it has thoughts. But if we were to view the claim that humans have thoughts as an inference from human behaviour, a sceptic could consistently deny the conclusion (see Other minds). As a matter of fact, however, we learn to view other humans as having thoughts in the process of psychological and social development. In many cases it is true to say that we see their thought in action. We can be wrong about the content of a particular thought, and indeed some versions of solipsism and scepticism may be logically possible or even defensible in a philosophy journal or seminar. But we do not confuse this with the practices of everyday life. Similarly, it seems that viewing some animals as sometimes having thoughts is part of our cultural practice. While we can take a sceptical stance and it may be useful to do so for certain methodological purposes, there is little reason to do so when it comes to fixing our beliefs. The idea that we are continuous with animals, unless it can be shown otherwise, appears to be a plausible naturalistic hypothesis.

Some philosophers may still be bothered by the fact that animals do not have full-blown human languages. But many animals may have powerful enough, nonlinguistic, representational systems to make their thoughts a significant possibility. For that matter, philosophers such as Ryle who deny that thinking necessarily involves manipulating representations may well be correct. At this stage in the investigation, the possibility that animals may have thoughts without language does not appear to be blocked by any convincing philosophical argument.

DALE JAMIESON

References and further reading

Bekoff, M. and Jamieson, D. (eds) (1995) *Readings in Animal Cognition*, Cambridge, MA: MIT Press.(This is a collection of articles on animal cognition by leading philosophers, psychologists and biologists, including Cheney and Seyfarth, Herman, Ristau and Savage-Rumbaugh.)

Chomsky, N. (1966) *Cartesian Linguistics*, New York: Harper & Row.(A defence of the Cartesian foundations of Chomsky's linguistic theory.)

Davidson, D. (1975) 'Thought and Talk', repr. in *Truth and Interpretation*, Oxford: Oxford University Press, 1984, 155-70.(This is the most influential recent article that denies that animals have thoughts.)

Dupré, J. (1991) 'Conversations with Apes: Reflections on the Scientific Study of Language', in J. Hyman (ed.) *Investigating Psychology: Sciences of the Mind After Wittgenstein*, London: Routledge, 95-116.(This is an especially interesting philosophical assessment of the ape language studies.)

Jamieson, D. (1998, forthcoming) 'Science, Knowledge, and Animal Minds', *Proceedings of the Aristotelian Society*.(Discussion of the 'other minds' problem as applied to animals.)

Leahy, M.P.T. (1991) *Against Liberation*, London: Routledge.(This attack on animal liberation argues that talk about animal thought and language is largely anthropomorphic.)

Malcolm, N. (1972-3) 'Thoughtless Brutes', repr. in D.M. Rosenthal (ed.) *The Nature of Mind*, New York: Oxford University Press, 1991, 454-61.(A Wittgensteinian defence of the view that animals think but do not have thoughts.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Foundational work in behaviourist philosophy of language.)

Ristau, C. (ed.) (1991) *Cognitive Ethology*, Hillsdale, MI: Lawrence Erlbaum Associates.(A collection of essays including the work of Ristau, Cheney and Seyfarth, and Pepperberg.)

Savage-Rumbaugh, E.S. and Lewin, R. (1994) *Kanzi: The Ape at the Brink of the Human Mind*, New York: Wiley.(The popular exposition of Savage-Rumbaugh's work.)

Savage-Rumbaugh, E.S. *et al.* (1993) 'Language Comprehension in Ape and Child', *Monographs of the Society for Research in Child Development* 233 (58).(The technical exposition of Savage-Rumbaugh's work. Co-authors are Jeanine Murphy, Rose Sevcik, Karen Brakke, Shelly Williams and Duane Rumbaugh, with commentary by Elizabeth Bates.)

Sorabji, R. (1993) *Animal Minds and Human Morals*, Ithaca, NY: Cornell University Press.(The best discussion of the ancient debate over animal minds.)

Terrace, H.S. (1980) *Nim*, London: Methuen.(A critique of the ape language experiments and the report of a failed attempt to teach language to a chimpanzee.)

Vendler, Z. (1972) *Res Cogitans*, Ithaca, NY: Cornell University Press.(A contemporary defence of a Cartesian theory of mind.)

Wallman, J. (1992) *Aping Language*, New York: Cambridge University Press.(A systematic, sceptical discussion of the ape language controversy.)

Wittgenstein, L. (1953) *Philosophical Investigations*, ed. G.E.M. Anscombe and R. Rhees, trans. G.E.M. Anscombe, Oxford: Blackwell.(Profoundly influential discussion of language and mind.)

Animals and ethics

Does morality require that we respect the lives and interests of nonhuman animals? The traditional doctrine was that animals were made for human use, and so we may dispose of them as we please. It has been argued, however, that this is a mere 'speciesist' prejudice and that animals should be given more or less the same moral consideration as humans. If this is right, we may be morally required to be vegetarians; and it may turn out that laboratory research using animals, and many other such practices, are more problematic than has been realized.

1 The traditional view

In some Eastern systems of thought, animals are accorded great respect. The Jains of India hold that all life is sacred, drawing no sharp distinction between human and nonhuman life. They are therefore vegetarians, as are Buddhists, whose sacred writings forbid all needless killing. In the West, however, it was traditionally believed that animals were made for human use. This idea, familiar from the Old Testament book of Genesis and elaborated by a long line of Jewish and Christian thinkers, also formed part of Aristotle's worldview. Aristotle taught that 'nature does everything for a purpose', and so, just as plants exist to provide food for animals, animals exist to provide food and other 'aids in life' for humans (see Regan and Singer (eds) 1989).

This was cosmology with a moral point. Aquinas, who emphasized that it was God himself who provided the animals for human use, made the point explicit: 'Therefore,' he said, 'it is not wrong for man to make use of them, by killing or in any other way whatever' (*Summa contra gentiles*). Are there, then, no limits on how animals may be treated? One might think we have a duty to be kind to them out of simple charity. But Aquinas insisted that this is not so. 'Charity,' he said, 'does not extend to irrational creatures.'

There was, however, one way in which animals could gain a degree of protection. They might be the incidental beneficiaries of obligations owed to humans. If someone has promised to look after your dog, he is obliged to care for it. But the obligation is owed to you, not to the dog. There might even be a general duty not to torment animals, because, as Kant put it, 'He who is cruel to animals becomes hard also in his dealings with men' (1780-81: 240). But once again, the point was to protect the men, not the animals. (This has sometimes been called the 'indirect duty view' - that we can have duties to animals, but only indirect ones.)

This view might seem extreme in its near total disregard for nonhumans. Nevertheless, the idea that animals are essentially resources for human use was accepted by almost every important thinker in the Western tradition - including such figures as St Francis, who is popularly but wrongly believed to have advocated a more charitable stance. For this view to be defensible, however, there must be some difference between humans and other animals that would explain why humans have a privileged moral status. Traditional thought cited two such differences. For Aristotle, the difference was that humans alone are rational. Religious figures added that man alone was made in the image of God. These explanations seemed sufficient until 1859, when Darwin's *On the Origin of Species* (1859) transformed our understanding of man's relation to the rest of nature (see Darwin, C.).

2 Challenges to the traditional view

Darwin demonstrated that humans are not 'set apart' from other animals, but are related to them by evolutionary descent (see Evolution, theory of). It is no accident that we bear such a startling resemblance to the apes. Our bones and muscles are but modified versions of the ape's bones and muscles - they are similar because we inherited them from the same ancestors. The same is true of our rational faculties. Man is not *the* rational animal, for other animals also possess a degree of rationality. How could it be otherwise, when our brains developed from a common source? Darwin went so far as to declare, 'There is no fundamental difference between man and the higher mammals in their mental faculties' (1859: 35). Such differences as do exist, he said, are matters of degree, not kind.

Today it is widely accepted that Darwin was right, at least in the main outlines of his view, and this poses an obvious ethical dilemma: if humans are similar in so many ways to other animals, and humans merit moral protection, then why should other animals not merit protection too? As Asa Gray, Darwin's friend in America, put it, 'Human beings may be more humane when they realize that, as their dependent associates live a life in which man has a share, so they have rights which man is bound to respect' (1880: 54). Darwin himself regarded cruelty

to animals, along with slavery, as one of the two great human moral failings.

Another nineteenth-century development also cast doubt on the traditional exclusion of animals from the range of moral concern. The utilitarians, led by Jeremy Bentham and John Stuart Mill, argued that morality is fundamentally a matter of seeking to promote happiness and prevent suffering (see Utilitarianism). But Bentham saw no reason to limit moral concern to human suffering. In fact, he suggested that disregard for animals was a form of discrimination analogous to racism:

> The day may come when the rest of the animal creation may acquire the rights which never could have been withholden from them but by the hand of tyranny. The French have already discovered that the blackness of the skin is no reason why a human being should be abandoned without redress to the caprice of a tormentor. It may one day come to be recognized that the number of the legs, the villosity of the skin, or the termination of the os *sacrum* are reasons equally insufficient for abandoning a sensitive being to the same fate…. The question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?
>
> (1789: 311; original emphasis)

It must be noted, however, that for most of Western history the moral status of animals did not seem to be much of an issue, and philosophers did not write very extensively about it (Bentham's discussion, for example, is confined to a footnote). The subject began to be widely discussed among philosophers only after the publication of Peter Singer's *Animal Liberation* in 1975.

3 The contemporary debate

One of the striking things about the debate concerning animals is that it is possible to reach radical ethical conclusions by invoking only the most common moral principles. The idea that it is wrong to cause suffering, unless there is a sufficient justification, is one of the most basic moral principles, shared by virtually everyone. Yet the consistent application of this principle seems to lead straight to vegetarianism or at least to the avoidance of factory-farmed meat. The argument is disarmingly simple. In modern factory farms, animals who are raised and slaughtered for food suffer considerable pain. Since we could easily nourish ourselves without eating them, our only reason for eating them seems to be our enjoyment of how they taste. So, unless one thinks our gustatory pleasure is a sufficient justification for causing torment, the obvious conclusion is that we are wrong to produce and consume such products.

Other arguments appeal to less commonplace notions. The word 'speciesism' was coined by Richard Ryder, a British psychologist who ceased experimenting on animals after becoming convinced it was immoral, and popularized by Singer in *Animal Liberation*. Speciesism is said to be analogous to racism. Just as racists unjustifiably give greater weight to the interests of the members of their own race, speciesists unjustifiably give greater weight to the interests of the members of their own species (see Discrimination §1).

Consider, for example, the very different standards we have for using humans and nonhumans in laboratory research. Why do we think it permissible to perform a painful and destructive experiment on, say, a rhesus monkey, when we would not perform the same experiment on a human? Someone might suggest that, say, humans are more intelligent than monkeys, or that their social relationships are more complex. But consider mentally retarded persons whose cognitive and social capacities are no greater than those of the animal. Would it be permissible to perform the same experiment on them? Many people think that, simply because they are human, it would not. This is speciesism laid bare: there is no difference between the human and the nonhuman in their abilities to think, feel or suffer, and yet the human's welfare is counted for more.

This line of thought suggests that animals may be treated differently from humans when, and only when, there are morally relevant differences between them. It may be permissible to admit humans, but not other animals, to universities, because humans can read and other animals cannot. But in cases where there are no relevant differences, they must be treated alike. This is the sense in which humans and nonhumans can be said to be morally 'equal': the bare fact that one is human never itself counts for anything, just as the bare fact that one has one skin colour or another never itself counts for anything. So we may not treat an animal in any way in which we would not be willing to treat a human with the same intellectual and emotional capacities.

Such arguments have, of course, provoked lively opposition. Many philosophers find it difficult to believe that

mere animals could have such powerful claims on us. Morality, they say, is fundamentally a human institution established to protect human rights and human interests (see Morality and ethics). Contractarianism, which has emerged in the latter half of the twentieth century as the principal rival to utilitarianism, makes this point most clearly. According to this view, morality rests on agreements of mutual benefit. Morality arises within a community when each person agrees to 'play the social game', respecting other people's rights and interests, provided others will do so as well. This agreement makes social living possible, and everyone benefits from it (see Contractarianism in ethics and political philosophy). But animals are unable to participate in such agreements, so they do not come within the sphere of moral protection.

In addition to initiating a philosophical debate, Peter Singer's book is perhaps the most conspicuous example of a philosophical work triggering a social movement. The animal rights movement, with its principled opposition to such practices as factory farming, the use of animals in commercial and scientific research, and the fur trade, has become a familiar part of contemporary life. Rarely, if ever, have philosophical thinking and social activism been linked so closely.

See also: Agricultural ethics §3; Environmental ethics; Evolution and ethics; Rights

JAMES RACHELS

References and further reading

Aquinas, T. (*c*.1259-65) *Summa contra gentiles (Synopsis [of Christian Doctrine] Directed Against Unbelievers)*, trans. English Dominican Fathers, Chicago, IL: Benziger Brothers, 1928, III.2, c.112.(Argues that animals are made by God for human use.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, New York: Hafner, 1948. (Argues on utilitarian grounds that the suffering of animals counts equally with the suffering of human beings.)

Darwin, C. (1859) *On the Origin of Species*, London: John Murray; repr. Cambridge, MA: Harvard University Press, 1975.(The classic source of the theory of evolution by natural selection.)

Finsen, L. and Finsen, S. (1994) *The Animal Rights Movement*, New York: Twayne Publishers.(An overall view of the movement, covering its history and the philosophical and religious arguments on which it is based.)

Frey, R.G. (1983) *Rights, Killing, and Suffering: Moral Vegetarianism and Applied Ethics*, Oxford: Blackwell. (Argues that the various pro-animal arguments do not succeed.)

Gray, A. (1880) *Natural Science and Religion: Two Lectures Delivered to the Theological School of Yale College*, New York: Charles Scribner's Sons.(Argues that we must acknowledge that animals have rights because we share common traits with them.)

Kant, I. (1780-81) *Lectures on Ethics*, trans. L. Infield, New York: Harper & Row, 1963.(Holds that humans have no duties to animals because animals are not self-conscious.)

Regan, T. (1983) *The Case for Animal Rights*, Berkeley, CA: University of California Press.(Defends the view that nonhuman animals have moral rights comparable to the rights of humans. One of the most important philosophical defences of animals.)

Regan, T. and Singer, P. (eds) (1989) *Animal Rights and Human Obligations*, Englewood Cliffs, NJ: Prentice Hall, 2nd edn.(A collection of readings on all sides of the issue. Includes a selection from Aristotle, which illustrates his view that because humans have a privileged moral status, on the grounds that they alone are rational, animals exist to provide food and other 'aids in life' for humans.)

Singer, P. (1975) *Animal Liberation*, New York: New York Review Books; 2nd edn, 1990. (A clear and well-written book, arguing that the same moral principles that govern how people should be treated also apply to animals.)

Anomalous monism

Anomalous monism, proposed by Donald Davidson in 1970, implies that all events are of one fundamental kind, namely physical. But it does not deny that there are mental events; rather, it implies that every mental event is some physical event or other. The idea is that someone's thinking at a certain time that the earth is round, for example, might be a certain pattern of neural firing in their brain at that time, an event which is both a thinking that the earth is round (a type of mental event) and a pattern of neural firing (a type of physical event). There is just one event, that can be characterized both in mental terms and in physical terms. If mental events are physical events, they can, like all physical events, be explained and predicted (at least in principle) on the basis of laws of nature cited in physical science. However, according to anomalous monism, events cannot be so explained or predicted as described in mental terms (such as 'thinking', 'desiring', 'itching' and so on), but only as described in physical terms. The distinctive feature of anomalous monism as a brand of physical monism is that it implies that mental events as such (that is, as described in mental terms) are anomalous - they cannot be explained or predicted on the basis of strict scientific laws.

1 Psychophysical identity and strict laws

Davidson proposed the following argument for a psychophysical identity thesis: let *m* be any arbitrary mental event that causally interacts with some physical event *p*. Events related as cause and effect fall under strict laws. But if *m* and *p* fall under a strict law, that strict law is a physical law as all strict laws are physical laws. If an event falls under a strict physical law, then it is a physical event. Hence, *m* is a physical event. As Davidson points out, if one makes the plausible assumption that every mental event causally interacts with some physical event, one can arrive by the same reasoning at the more general conclusion that every mental event is a physical event.

Although questions can be raised about various steps in this argument, its conclusion is fairly widely accepted (see Mind, identity theory of). The focus of attention has been on his claim that although mental events are physical events, there are no strict psychological laws (psychological anomalism) and there are no strict psychophysical laws (psychophysical anomalism) - the doctrines that make his version of monism count as a version of *anomalous* monism. Before discussing the anomalism theses, let us turn to the notion of a strict law itself.

According to Davidson, laws are true general statements that support counterfactuals and other subjunctive conditionals (see Counterfactuals), and which are confirmable by their positive instances. *Strict* laws are sometimes claimed to be laws that contain no escape clauses such as 'other things being equal', or 'typically', or 'for the most part', or the like. However, Davidson appears to have a stricter notion of law in mind than that. For the statement 'All sentient beings are mortal' contains no explicit escape clauses and expresses as exceptionless a law as one can hope to find, yet it appears that Davidson would not count it. His notion of a strict law is one that is as explicit, precise, and exceptionless as nature permits, or that can be refined into such a law by the addition of further explicit provisos and conditions stated in the same theoretical vocabulary as the original law statement. In contrast, nonstrict laws can be made exceptionless only by explicitly citing probabilities in the law itself, or, at the cost of explicitness and precision, by adding *ceteris paribus* clauses or hedges, or else by employing predicates (like 'sentient' and 'mortal') whose analyses require appeal to *ceteris paribus* clauses or hedges that cannot be fully spelled out in the same theoretical vocabulary as the original law statement. This difference between strict and nonstrict laws, according to Davidson, is due to the fact that strict laws, unlike nonstrict ones, are couched in the vocabulary of a closed, comprehensive theory. A theory *T* is closed if and only if events within the domain of *T* causally interact only with other events within the domain of *T*. A theory *T* is comprehensive if and only if every event within its domain satisfies a unique *T*-description under which it is subsumed by one of *T*'s laws. Davidson allows that while strict laws are, or can be refined into, laws as exceptionless as nature permits, a strict law, even fully refined, may prove to be probabilistic; but if it does, there will be no law covering exactly the same causal transactions it covers that is more explicit, precise, and has fewer exceptions. Nonstrict laws, Davidson says, can support causal claims concerning individual events. But they do so, he maintains, by providing evidence that there is a strict law at work that is free of either explicit or implicit escape clauses, and that precisely states all the causal factors at work in the particular causal transactions in question.

Davidson's notion of a strict law remains a subject of interpretation. The following is textually defensible. A strict

law is a law that is couched solely in a basic vocabulary of a closed, comprehensive theory, or in terms that can be defined by terms logically constructible from such a basic vocabulary, or be reduced by bridge laws to terms so constructible; a nonstrict law is a law not couched in such terms, or solely in such terms. A set of terms is a basic vocabulary of a closed comprehensive theory if and only if it is a set of terms sufficient for the formulation of a closed comprehensive theory, and no proper subset of it is.

2 Psychological anomalism

Davidson argues as follows for psychological anomalism. Psychology is not a closed theory; and, further, psychological terms cannot be incorporated by reductive bridge law or definition into the vocabulary of a closed, comprehensive theory. Therefore, there are no strict psychological laws.

Psychology is indeed not closed; there is overwhelming empirical reason to believe that there are nonmental events that causally interact with mental events. To be sure, biology is not a closed theory either; indeed, none of the special sciences is closed. Davidson tells us that most of science employs only nonstrict laws. He holds, however, that strict laws can be found in physics, either in current physics, or, if it falls short of what it promises to be (that is, closed and comprehensive), then in some (as yet unstated, perhaps never to be stated) improved version of current physics. All strict laws are either laws of physics or laws that can be refined to laws of physics by appeal to reductive laws and/or definitions, and (if necessary) adding physical provisos.

Some philosophers maintain, on the following grounds, that special science laws, including even chemistry, are typically not strict since they cannot be so refined: the event and state types cited in special science laws are, typically, widely and multiply realized by the event and state types cited in physics, and hence the former cannot be reduced to the latter. Davidson has, however, marshalled reasons for holding that certain mental predicates, in particular, cannot possibly be reduced to physical predicates, reasons which make no appeal to the notion of realization. The predicates in question are ones that contain essential occurrences of propositional attitude verbs such as 'believes', 'desires' and the like (see Propositional attitudes). Davidson maintains that propositional attitude concepts are such that predicates that express them could not possibly be reduced to physical predicates. His reasons are discussed in the following section. (Hereafter, 'mental event' will be used in a restricted sense to mean acquiring or losing a propositional attitude.)

3 Psychophysical anomalism

One way that psychophysical anomalism could fail is if there is a basic vocabulary for a closed, comprehensive theory that includes both mental and physical terms. A Cartesian interactionist (see Dualism) might argue that any basic vocabulary for a closed, comprehensive theory true of our world would have to contain both mental and physical terms; for the mental properties expressed by at least some mental terms are fundamental force-generating properties. However, this position is empirically implausible: mechanics seems to have no need of the hypothesis that there are mental events.

There is another way that psychophysical anomalism could fail, namely if mental terms can be defined by terms constructible from the basic vocabulary of a closed, comprehensive theory or be reducible to such terms by reductive bridge laws. Davidson offers reasons that appeal to the nature of mental concepts for holding that no mental predicate (more specifically, no propositional attitude predicate) could be reduced to any physical predicate, no matter how complex, in either of these two ways. If definitional psychophysical reduction were possible, then a physical predicate could express an unhedged sufficient condition of application for a mental predicate - one unqualified by any escape clauses. That, Davidson argues, is impossible: no physical predicate could express such an unhedged sufficient condition of application. Hence, definitional reduction is impossible. The same consideration arguably shows that reduction via bridge laws is impossible. For if a psychophysical bridge law were a reductive law, rather than an emergent law, the law would have to be a derivative law, rather than a fundamental law. The law would have to be implied by laws of the reducing theory, physical theory, together with physical conditions, and only analytical principles, and perhaps other necessary truths. Otherwise, the truths of physical theory itself would not imply the psychophysical truths in question. It is arguable that if the truths of physical theory itself imply such psychophysical truths, then unhedged analytical generalizations with physical (and logical) predicates in their antecedents and mental predicates in their consequents would be possible. But such generalizations are impossible if mental predicates could not have unhedged physical conditions of

application. Davidson does not consider the possibility, however, that there are metaphysically necessary truths that are not analytical truths, and that enable psychophysical reduction; of that consideration, more in §4.

There are, in Davidson's work, several lines of argument for the claim that mental predicates cannot have unhedged physical sufficient conditions of application; and other proponents of psychophysical anomalism have contributed to the development of some of these lines of argument. Much of the interest in anomalous monism has centred on whether a sound argument for the claim can be found.

Every line of argument appeals to what Davidson calls 'the holism of the mental' - (roughly) the view that propositional attitude types, identified by their intentional mode (for example belief) and their content (for example that the earth is round), are individuated holistically by their place in a network of other propositional attitude types. The place of a propositional attitude state type in such a network is partly determined by logical and semantical relationships between its content and the contents of a vast range of other state types in the network. For networks of propositional attitude types must, Davidson argues, exhibit a large degree of rational coherence. Irrationality is possible, obviously, but only against a background of largely rational coherence. To make coherent propositional attitude attributions, our attributions must conform to standards of rational coherence, standards concerning what it would make sense to believe given certain other beliefs, what it would make sense to value given certain beliefs and desires, and what courses of action would make best sense given a certain pattern of beliefs, desires, values, and so on.

One line of argument is that if unhedged physical conditions of application were possible, propositional attitude attributions could be made solely on the basis of such conditions, and thus without reliance on rational assessment. But that, it is claimed, would amount to changing the subject, since principles of rational assessment are constitutive of the very meaning of such mental predicates. This line of argument is, however, inconclusive. If there are constitutive principles of rational assessment that implicitly define mental predicates, then, indeed, unhedged sufficient physical conditions of application cannot permit attributions that contravene the principles without changing the subject. Nevertheless, so long as those principles would not be contravened by attributions made solely on the basis of unhedged physical conditions, there would be no change of subject. It thus remains to be seen why such attributions would have to contravene principles of rational coherence.

It has also been argued, however, that there is no complete set of impersonal, objective principles of rational coherence. This idea has been put by saying that there is no complete set of codifiable principles of rationality that can supply a common, objective standard in attributing mental states to others. We must rely on our own standards of assessment in understanding others; there is no single objective standard to which we can appeal. The claim that there are no codifiable objective standards of rationality is contentious. Decision theory and confirmation theory, for instance, attempt to formulate such standards (see Confirmation theory; Decision and game theory). Perhaps they cannot succeed, but if they cannot, it remains to be seen why they cannot. Suppose, however, that there is indeed no single set of codifiable objective standards of rationality, that there are at best rules of thumb, and that we must invariably rely to some extent on our subjective sense of what makes rational sense - what it makes sense to believe given other beliefs, and so on - in making mental attributions. The question arises, then: would this imply that unhedged physical conditions of application for propositional attitudes would invariably contravene our assessments of rational coherence?

If there is a set of codifiable principles of rational assessment for each attributor at a time, then such contravention might be avoidable. When two individuals' propositional attitude attributions are governed by somewhat different sets of principles, there would arguably be some difference in the meanings of the propositional attitude predicates they employ. Similarly, the meanings of propositional attitude predicates a given individual employs would arguably change somewhat over time as somewhat different sets govern the individual's attributions at different times. Propositional attitude predicates could be, in principle, rendered unambiguous by relativizing them to such sets of principles. Unhedged physical sufficient conditions of attribution for mental predicates understood as governed by a particular set of principles of rationality would, then, have to be such that they do not contravene the principles. However, if they did not, there would be no change of subject.

It might be responded that there will be no set of codifiable principles of rational assessment even for an attributor at a time. But a case for that remains to be made. In any event, for present purposes, suffice it to note that the reasoning in favour of psychophysical anomalism remains a topic of controversy.

4 Psychophysical anomalism and supervenience

While Davidson denies that mental characteristics reduce to physical characteristics, he maintains that they are supervenient on physical characteristics in this sense: there cannot be two events alike in every physical respect but differing in some mental respect. Various kinds of supervenience relations have been distinguished (see Supervenience; Supervenience of the mental). The kind Davidson appeals to appears to be a particularly weak kind. While he eschews talk of possible worlds, it can, nevertheless, be stated by quantifying over them: for any possible world, *w*, if any event *x* in *w* differs in any mental respect from event *y* in *w*, then *x* differs in some physical respect from *y*. This weak psychophysical supervenience thesis implies neither strict nor nonstrict psychophysical laws, and is thus compatible with psychophysical anomalism.

Many philosophers hold, however, that mental characteristics must supervene on physical characteristics in the following stronger sense, if mental characteristics indeed make a difference to causal relations: for any possible worlds *w* and *w**, and any event *x* in *w* and any event *y* in *w**, if *x* and *y* differ in some mental respect, then they differ in some physical respect. This strong supervenience thesis, in conjunction with the assumption that physical conditions are closed under complementation (negation will be part of the basic vocabulary of physical theory), implies that for any propositional attitude type, there is some unhedged, metaphysically sufficient physical condition for its occurrence. Strong psychophysical supervenience is thus incompatible with the thesis that mental predicates cannot have such physical conditions of application.

Strong psychophysical supervenience is, however, arguably compatible with psychophysical anomalism itself. For, laws, according to Davidson, must be confirmable by their positive instances, and that there are such metaphysically sufficient unhedged, physical conditions does not imply that there are true, unhedged psychophysical generalizations that are confirmable by their positive instances. Moreover, psychophysical anomalism remains a deeply interesting doctrine, even if it is compatible with the possibility of unhedged metaphysically sufficient physical conditions of application for propositional attitude predicates. For it is deeply interesting whether our propositional attitude concepts are such that generalizations purporting to state such conditions of application could be confirmable by their positive instances.

See also: Laws, natural; Reductionism in the philosophy of mind; Reduction, problems of

BRIAN P. McLAUGHLIN

References and further reading

Child, W. (1993) 'Anomalism, Uncodifiability, and Psychophysical Relations', *Philosophical Review* 102: 215-45.(Argues that Davidson's reasoning for anomalism can be reconstructed employing the idea that there is no complete set of codifiable principles of rationality.)

Davidson, D. (1970) 'Mental Events', in *Actions and Events*, Oxford: Clarendon Press, 1980.(The seminal paper on anomalous monism.)

Davidson, D. (1973) 'The Material Mind', in *Actions and Events*, Oxford: Clarendon Press, 1980.(Includes a detailed discussion of anomalous monism.)

Davidson, D. (1974) 'Psychology as Philosophy', in *Actions and Events*, Oxford: Clarendon Press, 1980.(Contains an elaboration and defence of anomalism monism along with some replies to objections to the doctrine.)

Davidson, D. (1980) *Actions and Events*, Oxford: Clarendon Press.(Collection of some of Davidson's papers on actions, events, causation, and anomalous monism.)

Davidson, D. (1993) 'Thinking Causes', in J. Heil and A. Mele (eds) *Mental Causation*, Oxford: Clarendon Press. (Davidson argues that anomalous monism and the principle of the nomological character of causality do not commit him to the view that mental types and properties make no difference to causal relations.)

Davidson, D. (1995) 'Donald Davidson', in S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*, Oxford: Blackwell.(Davidson's latest statement on anomalous monism.)

Heil, J. and Mele, A. (eds) (1993) *Mental Causation*, Oxford: Clarendon Press. (Contains essays discussing whether anomalous monism implies epiphenomenalism.)

Honderich, T. (1981) 'Psychophysical Lawlike Connections and their Problem', *Inquiry* 24: 277-303.(An attempt to show that anomalous monism is committed to property epiphenomenalism.)

Kim, J. (1985) 'Psychophysical Laws', in E. LePore and B.P. McLaughlin (eds) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.(Attempts to spell out Davidson's reasons for denying the existence of psychological laws and psychophysical laws.)

LePore, E. and McLaughlin, B.P. (1985) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.(Contains essays on anomalous monism.)

McLaughlin, B.P. (1985) 'Anomalous Monism and the Irreducibility of the Mental', in E. LePore and B.P. McLaughlin (eds) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.(Contains a discussion of Davidson's central arguments for psychophysical identity, and for psychological and psychophysical anomalism.)

Stanton, W.L. (1983) 'Supervenience and Psychological Law in Anomalous Monism', *Pacific Philosophical Quarterly* 64: 72-9.(Contains an examination of Davidson's argument for the identity thesis, a discussion of whether psychophysical supervenience is consistent with the anomalism of the mental, and offers the explication of a strict law presented in §1.)

Anscombe, Gertrude Elizabeth Margaret (1919-)

Elizabeth Anscombe has contributed to all principal areas of philosophy, most influentially to ethics and the philosophy of mind. She is the founder of contemporary action theory, and an important source of the revival of interest in virtue ethics. The chief influences on her thought are the work of her teacher, Ludwig Wittgenstein, much of which she has translated and of which she is an important interpreter, and the classical and medieval traditions, as found in Aristotle and Aquinas. She has also made a number of contributions to the defence of Roman Catholic religious belief.

1 *Intention*

Anscombe's most important work, *Intention* (1957a), is the founding document of contemporary philosophy of action. Much of her later thought may be said to move out from this essay in three main directions: towards ethics; general problems of explanation or 'causality'; and associated problems in the philosophy of mind.

The concept of intention is employed, she says, in three main connections: we speak of 'events in a man's history' as *intentional actions*, of the *intention with which* an action is performed, and of the *expression of intention*, or of the corresponding 'pure' intention for the future, which may exist though no action has yet been done with that intention. Taking the notion of intentional action as prior, she argues that an action is intentional when it is subject to a certain form of explanation: that is, when, as she puts it, 'a certain sense of the question "Why?" has application' to it. Her task is then to isolate this particular form of explanation. It is without application, she argues, if the agent is unaware of doing the thing; or if the agent knows that he or she is, but only by having noticed it - that is, 'by observation', or where, though the agent knows that the thing is going on without observation, he or she can give no account of it, or none without observation, conjecture and so forth. If a piece of behaviour passes these tests, it is an intentional action, unless perhaps it is a case of 'mental causality', like a gasp that is given at the hissing of a snake. These latter cases she excludes by a series of more complex tests (see Action §1; Intention §5).

Like Oedipus, I may *strike a person* intentionally, and *strike my father* unintentionally, though these are not two distinct events or two distinct actions of mine. Actions are thus intentional only 'under a description'. What is given in answer to the question 'Why?' is in fact often a further description of the same action. A series of such questions will thus reveal an order among many of the descriptions true of an action: 'Why are you moving your arm up and down?' -'Because I'm pumping water' - 'But why pump water?' - 'Because I'm replenishing the house water supply'. This chain of questions 'Why?' may often be pressed into the future, and thus beyond any description of what is now happening; the responses will then merely express the *intention with which* the action mentioned earlier is performed. In general, if the question 'Why?' has application to a first-person future-tense description of action, then the description is an *expression of intention* and not a mere prediction.

The idea that 'practical knowledge' (the knowledge one has of one's intentional action) does not spring from observation leads to some of the more striking claims of the work. She famously compares the relation that practical thought bears to action with the relation a shopping list bears to the contents of the shopper's basket. The corresponding model of non-practical or 'speculative' thought is given by the relationship between the same basket and the list of its contents constructed by the detective who follows its owner. The difference is in 'direction of fit', as it is now called: the detective amends a mismatch between list and basket by altering his list, the shopper by altering the contents of the basket. These matters are elucidated by an extended discussion of Aristotle's notion of a 'practical syllogism'. It is argued, among other things, that the order of descriptions of an action which we elicit with the question 'Why?' is the reverse of the order of descriptions articulated by the agent in reasoning from end to means (see Practical reason and ethics §3).

2 Other works

Much of Anscombe's ethical work also concerns the notion of intention and the relationship of thought to action. In 'Modern Moral Philosophy' (1958) she argues that contemporary thought on these matters is so wide of the mark that there is no point in practising moral philosophy at all. A naïve philosophy of mind and action has led, for example, to what she calls 'consequentialism' - the view that whether a person intends something or not is

irrelevant to the question whether he is responsible for it (see Consequentialism §1). (In 'Mr. Truman's Degree' (1957b) she claims that the familiar justifications for the bombing of Hiroshima and Nagasaki, which she argues was murderous, presuppose just such thinking.) She also maintains that the programme of attempting to elucidate the 'moral sense of "ought"' must be jettisoned: it presupposes a theory of ethics as founded in divine law. Only an Aristotelian ethical theory, she thinks, can be given an intelligible secular development. But this would require clarification of such concepts as 'virtue', 'human nature' and 'human flourishing', and thus renewed inquiry into philosophy of mind and action (see Virtue ethics §2).

In 'Causality and Determination' (1971), Anscombe attacks the notion that 'if an effect occurs in one case and a similar effect does not occur in an apparently similar case, there must be a relevant further difference' (1971: 133). Having attacked this view, she attempts to show how freedom of the will, which she argues is incompatible with it, is to be understood in the light of its rejection. An event not explicable in terms of prior determining physical causes may yet, she argues, be subject to other forms of explanation.

Much of Anscombe's work on the general philosophy of mind attempts to extend and clarify the teachings of Wittgenstein (§13). Her boldest claims are found in 'The First Person' (1974). There she argues that the word 'I' is not a referring term, that self-consciousness is not awareness of an object (a 'person' or 'self'), and that if either of these things *were* the case then the object of reference and awareness would have the character of a Cartesian ego.

See also: Causation §6; Free will §2; Reasons and causes§2

MICHAEL THOMPSON

List of works

Anscombe, G.E.M. (1981) *Collected Philosophical Papers*, Minneapolis, MN: University of Minnesota Press, 3 vols.(The definitive edition of Anscombe's works.)

Anscombe, G.E.M. (1957a) *Intention*, Oxford: Blackwell; 2nd edn, 1963.(On the concepts of action, intention and a reason for action.)

Anscombe, G.E.M. (1957b) 'Mr. Truman's Degree', privately published pamphlet; repr. in G.E.M. Anscombe (1981), vol. 3, 62-71.(A popular work attacking President Truman's use of the atomic bomb and defending an 'absolutist' account of the prohibition of murder.)

Anscombe, G.E.M. (1958) 'Modern Moral Philosophy', *Philosophy* 33; repr. in G.E.M. Anscombe (1981), vol. 3, 26-42.(Counselling a revival of ancient forms of ethical theory especially that of Aristotle; defines and attacks 'consequentialism'.)

Anscombe, G.E.M. (1959) *An Introduction to Wittgenstein's 'Tractatus'*, Philadelphia, PA: University of Pennsylvania Press; 2nd edn, 1971.(An introduction to some aspects of the teachings of Frege, Russell and the early Wittgenstein.)

Anscombe, G.E.M. and Geach, P. (1961) *Three Philosophers*, Ithaca, NY: Cornell University Press.(The first essay, an interpretation of Aristotle's doctrine of substance, is by Anscombe.)

Anscombe, G.E.M. (1963) 'The Two Kinds of Error in Action', *Journal of Philosophy* 60: 393-401; repr. in G.E.M. Anscombe (1981), vol. 3, 3-9.(On the role of knowledge and intention in morals and in criminal law.)

Anscombe, G.E.M. (1965a) 'The Intentionality of Sensation', in R.J. Butler (ed.) *Analytical Philosophy*, 2nd series, New York: Barnes & Noble, 143-58; repr. in G.E.M. Anscombe (1981), vol. 2, 3-20.(On the concept of 'an object of perception'.)

Anscombe, G.E.M. (1965b) 'Thought and Action in Aristotle', in R. Bambrough (ed.) *New Essays on Plato and Aristotle*, London: Routledge & Kegan Paul, 158-80; repr. in G.E.M. Anscombe (1981), vol. 1, 66-80.(An interpretation and critique of Aristotle's conception of deliberation.)

Anscombe, G.E.M. (1971) 'Causality and Determination', inaugural lecture, Cambridge University; repr. in G.E.M. Anscombe (1981), vol. 2, 133-47.(Defends a much looser account of the notion of a cause than is prevalent in contemporary philosophy; the essay ends with a very difficult discussion of freedom of the will.)

Anscombe, G.E.M. (1974) 'The First Person', in S. Guttenplan (ed.) *Mind and Language*, Oxford: Oxford University Press, 45-65; repr. in G.E.M. Anscombe (1981), vol. 2, 21-36.(On the first person and self-conscious thought.)

Anscombe, G.E.M. (1978) 'Rules, Rights and Promises', *Midwest Studies in Philosophy* 3: 318-23; repr. in G.E.M. Anscombe (1981), vol. 3, 97-106.(On the genesis of the concepts of a rule and of a right.)

Anscombe, G.E.M. (1979) 'Under a Description', *Nous* 13: 219-33; repr. in G.E.M. Anscombe (1981), vol. 2, 208-19.(A response to critics of *Intention*, which clarifies some of its teachings.)

Anscombe, G.E.M. (1982a) ‘Action, Intention and Double Effect’, *Proceedings of the American Catholic Philosophical Association* 56: 12-25.(A defence, from a modern point of view, of some aspects of the moral philosophy of Thomas Aquinas.)

Anscombe, G.E.M. (1982b) ‘Murder and the Morality of Euthanasia’, in *Euthanasia: A Clinical Perspective*, London: Linacre Centre, 23-36; repr. inL. Gormally (ed.) *Euthanasia, Clinical Practice and the Law*, Indianapolis, IN: Hackett, 1994, 37-50.(A semi-popular attempt at a definition of murder as a moral category.)

Anscombe, G.E.M. (1989) ‘Von Wright on Practical Reason’, in L. Hahn and P. Schilpp (eds) *The Philosophy of Georg Heinrik von Wright*, La Salle, IL: Open Court, 376-404; repr. in R. Hursthouse, G. Lawrence and W. Quinn (eds) *Virtues and Reasons*, Oxford: Oxford University Press, 1995, 1-34.(An important revision of the doctrine of *Intention* concerning practical reason.)

References and further reading

Bennett, J. (1966) ‘Whatever the Consequences’, *Analysis* 26: 83-102.(A spirited attack on the ‘absolutism’ of Anscombe 1957 and 1958.)

Davidson, D. (1963) ‘Actions, Reasons and Causes’, *Journal of Philosophy* 60: 685-700; repr. in D. Davidson (1980), 83-102.(An influential attack on the notion, apparently present in Anscombe’s *Intention*, that the reason why an action was performed is in no sense a cause of it.)

Davidson, D. (1978) ‘Intending’, in Y. Yovel (ed.) *Philosophy of History and Action*, Dordrecht: Reidel, 41-60; repr. in D. Davidson (1980), 83-102.(Attacking Anscombe’s methodological claim in *Intention* that the notion of *intentional action* should be treated in advance of that of *intention*.)

Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Oxford University Press.(The first six essays in this volume present the most influential alternative to Anscombe’s philosophy of action; despite the doctrinal divergence, they have contributed greatly to the currency of her central concepts and distinctions.)

Diamond, C. and Teichmann, J. (1979) *Intention and Intentionality*, Ithaca, NY: Cornell University Press.(A volume honouring Anscombe and containing many important criticisms of her work.)

Foot, P. (1967) ‘Abortion and the Doctrine of Double Effect’, *Oxford Review* 5: 5-15; repr. in P. Foot, *Virtues and Vices*, Berkeley, CA: University of California Press, 1978, 5-15.(An attack, friendlier than that of Bennett (1966), on Anscombe’s ‘absolutism’; Foot’s method of relating abstract principles to her picturesque examples was of great influence in later moral philosophy.)

Anselm of Canterbury (1033-1109)

Anselm of Canterbury, also known as Anselm of Aosta and Anselm of Bec or Saint Anselm, was first a student, then a monk, later prior and finally abbot of the monastery of Bec in Normandy, before being elected Archbishop of Canterbury in 1093. He remains one of the best-known and most readily engaging philosophers and theologians of medieval Europe. His literary corpus consists of eleven treatises or dialogues, the most important of which are the philosophical works Monologion and Proslogion and the magnificent theological work Cur deus homo (Why God Became a Man). He also left three meditations, nineteen prayers, 374 extant letters including Epistolae de Sacramentis (Letters on the Sacraments) and a collection of philosophical fragments, together with a compilation of his sayings (Dicta Anselmi) by Alexander, a monk of Canterbury, and a compilation of his reflections on virtue, De morum qualitate per exemplorum coaptationem (On Virtues and Vices as Illustrated by a Collage of Examples), possibly also by a monk at Canterbury.

At Bec Anselm wrote his first philosophical treatise, the Monologion, a title signifying a soliloquy. This work was followed by the Proslogion, the title meaning an address (of the soul to God). At Bec he also completed the philosophical dialogues De grammatico (On (an) Expert in Grammar), De veritate (On Truth), De libertate arbitrii (Freedom of Choice) and De casu diaboli (The Fall of the Devil). Near the end of his time at Bec, he turned his attention to themes more theological, drafting a first version of De incarnatione Verbi (The Incarnation of the Word) before September 1092 and completing the final revision around the beginning of 1094. During his time in office at Canterbury, which included two long exiles from England (1097-1100 and 1103-6), he wrote the Cur deus homo, followed by the concisely reasoned treatises De conceptu virginali et originali peccati (The Virgin Conception and Original Sin), De processione Spiritus Sancti (The Procession of the Holy Spirit) and De concordia praescientiae et praedestinationis et gratiae dei cum libero arbitrio (The Harmony of the Foreknowledge, the Predestination and the Grace of God with Free Choice).

Though his principal writings at Bec were more philosophical while his foremost writings as archbishop were more theological, still we must remember that Anselm himself made no express distinction between philosophy and theology, that at Bec he also wrote two meditations and sixteen prayers, and that his Cur deus homo and De concordia, in dealing with the weighty theological doctrines of atonement, predestination and grace, incorporate philosophical concepts such as necessitas praecedens (preceding necessity) and necessitas sequens (subsequent necessity).

Anselm's most famous philosophical work is certainly the Proslogion, while his most influential theological work is undoubtedly the Cur deus homo. The style of the Proslogion imitates that of Augustine in the Confessiones, where the soul invokes God as it prayerfully reflects and meditates. By contrast, the Cur deus homo is cast in dialogue form because, as Anselm states in I.1, 'issues which are examined by the method of question and answer are clearer, and so more acceptable, to many minds - especially to minds that are slower.' About his aims in the Proslogion there is no scholarly consensus. The traditional view holds that he is undertaking the twofold task of demonstrating the existence of God and demonstrating certain truths regarding God's attributes. In carrying out this task, he has recourse to a single consideration (unum argumentum), namely, that God is aliquid quo nihil maius cogitari potest (something than which nothing greater can be thought). This single consideration gives rise to a single argument form; the logical structure of the reasoning which purports to establish that quo nihil maius is actually existent is also the structure of the arguments which conclude that quo nihil maius is so existent that it cannot be thought not to exist, is alone existent per se, is omnipotent, merciful yet impassable, is supremely just and good, is greater than can be thought, and so on. According to this interpretation, the Proslogion seeks to establish most of the same conclusions that were reached in the earlier Monologion, but to establish them more directly, simply and tersely.

The central thrust of the Cur deus homo may be discerned from the title: namely, to explain why it was necessary for God, in the person of the Son, to become a man (that is, to become incarnate as a human being (homo)). Anselm uses the Latin word homo generically and not in the sense of male (vir). This fact is seen clearly in Cur deus homo II, 8: 'nil convenientius, quam ut de femina sine viro assumat [deus] illum hominem quem quaerimus' (nothing is more fitting than that God assume from a woman without a male that man [human being] about whom we are inquiring). Though the sense of homo varies in accordance with whether Anselm is speaking

about a human being or about a human nature, there is no doubt about the meaning of the title: the Son of God assumed a *human nature, thereby becoming* a *man; he did not assume another man (in other words, assume a human person together with a human nature) as the heretical Nestorians had taught, nor did he become man (in other words, become universal man, by assuming unindividuated human nature as such).*

Anselm's detailed theory of satisfaction for sin was in large measure a putative theoretical justification of the institutionalized practices of the confessional and the penitential system as found in the medieval Christian church, which understood every sin to constitute a punishable demerit and to require both the imploring of God's forgiveness and the making of amends for having dishonoured him. Throughout the intricate and sustained reasoning of the Cur deus homo, Anselm seeks to show one central truth: 'because only God can make this satisfaction and only a man ought to make it, it is necessary that a God-man make it' (Cur deus homo II, 6).

As in the Cur deus homo, so also in his other treatises Anselm proceeds insofar as he deems possible, sola ratione *(by recourse to rational considerations alone). Accordingly, he is rightly called the 'Father of Scholasticism'. He understands* ratio *in a broad sense, broad enough to encompass appeals to experience as well as to conceptual intelligibility. Although the main intellectual influence upon him was Augustine, he is less platonistic than the latter, and the influence of Aristotle's De interpretatione and Categories (from Boethius' Latin translations) is clearly discernible in his philosophical works.*

1 Biography

Major details of Anselm's life come down to us through his contemporary secretary and biographer, Eadmer, an English monk at Canterbury and author of the *Vita Anselmi (Life of Anselm)* and *Historia Novorum in Anglia (History of Recent Events in England)*. Eadmer knew little of, and reported meagrely on, Anselm's childhood and youth. It is known that Anselm was born of noble lineage in the Burgundian town of Aosta, near the border with the Kingdom of Lombardy (now in Italy). In 1056, however, several years after his mother's death and as a result of his father's hostility, Anselm left home. Some three years later, following intermittent studies, he arrived at the Benedictine Abbey of Bec in Normandy, having journeyed there to place himself under the tutelage of the Abbey's prior and schoolmaster, Lanfranc of Pavia. He was then twenty-six years old.

After his father's death (presumably in 1060), Anselm chose to enter the monastic order at Bec rather than return to the family estate. In 1063 he was elected prior of Bec, succeeding Lanfranc, who had been called to the abbey of St.-Etienne in Caen; in 1078 he was chosen abbot, in spite of his disinclination to assume the office. He showed even more reluctance and protestation when selected as archbishop of Canterbury in 1093, again in succession to Lanfranc. Eadmer tells of Anselm's vigorous and poignantly futile efforts to resist election, first to the abbacy at Bec and later to the archepiscopacy at Canterbury. When the monks of Bec besought Anselm to dispense with the customary protests and agree to become abbot, he threw himself prostrate on the floor, begging them not to weigh him down with so burdensome an office. A similar scene took place in England: when King William II, fearing that his sudden sickness was mortal, named Anselm archbishop and when the bishops carried him forcibly to William's bedside to receive from him the episcopal staff, Anselm kept his fist clenched, thus refusing the staff. Nonetheless, the staff was pressed against his hand by the bishops, the *Te deum* was chanted and he was proclaimed archbishop-elect on 6 March 1093. His consecration followed on 4 December.

Anselm's subsequent quarrels with William II (son of William the Conqueror) and with his brother and successor Henry I are well known. He contested William's exercise of jurisdiction over the church, in particular William's claim that he alone, as king, was entitled to convoke future reform synods and had the right to decide which rival for the papacy - Urban II or Clement III - to recognize. Relations became so tense that Anselm, acting on his own initiative, chose to leave England for three years (November 1097-September 1100) to consult with Urban II in Rome. Upon his return to England after William's death, he then quarrelled with Henry I over Urban's injunction against any bishop doing homage to a king or being invested with his bishop's office by a king or any other layman. Anselm, at Henry's behest, once again departed from England for three years (April 1103-August 1106), this time to consult with Pope Paschal II. The investiture controversy was settled only in 1107, twenty months before Anselm's death, when King Henry formally forswore the right of investiture at the Concordat of London.

Two conflicting portraits of Anselm's mature life and of his attitude towards his roles as theologian and prelate have emerged. The standard portrayal, that of R.W. Southern, who agrees in general with Eadmer, shows Anselm

as a devout monk who was committed to the ideals of monasticism and who aspired to a life of scholarly study and spiritual meditation. Dreading to trade the tranquillity of Bec for the combative milieu of Canterbury, he experienced his elevation to the archbishopric as 'his nearest approach to hell' (Southern 1988: 187). In his preface to the *Cur deus homo (Why God Became a Man)*, Anselm spoke of his tribulation of heart and his great suffering, presumably because of his clashes with King William over the extent of regal power. Anselm seemed most content when he could withdraw from the public arena, as he did in the Italian mountain village of Liberi while finishing the *Cur deus homo*, during his first exile.

A contrary portrait of Anselm is that of S.N. Vaughn (1987), who dismisses much of Eadmer's report as tendentious and too intent upon depicting Anselm as a saint. Anselm, says Vaughn, was a man politically astute and personally calculating, manipulative enough to enlist the support of his friends against his enemies and clever enough actively to fashion a positive historical image of himself by omitting from his collection of letters those that would have cast him in a less favourable light. Moreover, his very remonstrations and protestations when assuming high office were not only a formality - the expected and requisite display of humility - but also a test of the intensity of his electors' support. In fact, Vaughn claims, Anselm adroitly schemed to become selected as archbishop once he became convinced that God had destined him for this office. He did not desire the office and its burdens; but, despite this disinclination, he manoeuvred cleverly to obtain it in compliance with the perceived will of God. Once in the office he executed it with talent, protecting against regal confiscation the holdings and preserving from regal encroachment the prerogatives of the previously weakened English church.

Both these mutually incompatible pictures of the historical Anselm seem extreme. Neither is likely to be shown conclusively wrong by future scholarly debate, except in the unlikely event that new evidence is forthcoming. Scholarly caution must be exercised, however, in attempting radically to reshape Eadmer's testimony that Anselm neither sought nor desired to become archbishop - testimony corroborated by Anselm's own letters and not clearly contradicted by circumstantial evidence.

2 Philosophical method

Anselm did not write his first philosophical treatise, the *Monologion*, until he was forty-three. Even his earliest letters, prayers and meditations do not antedate 1070. Thus his entire corpus of writings arises from his mature years, after he had already been a monk at Bec for at least ten to sixteen years. Not surprisingly, his works show a consistency that cannot be found in a more prolific writer such as Augustine, who in later years felt obliged to write *retractationes* (reconsiderations) of his earlier work.

Stylistically, Anselm often uses streamlined Latin sentences that are set out in simple vocabulary. However, his concise style and his well-focused treatment of topics do not ensure that his ideas are easy to understand. On the contrary, understanding is frequently impeded by his failure to introduce more elaborate distinctions and to explore the metaphysical complexities that underlie his ostensibly straightforward assertions. More like Aristotle than like Augustine, he introduces into his argumentation appeals to common linguistic parlance, as when he explains that our saying 'Nothing caused it' ordinarily means that it is not the case that anything caused it. Similarly, when we say that God is unable to sin (an expression that seems to convey the idea of imperfection because of powerlessness), we usually mean that God has the ability always to keep from sinning (and to have this ability is to have a perfection).

Anselm is called the 'Father of Scholasticism' because he endeavours to show that revealed truths can be established on an independent rational basis. In the *Monologion*, where he professes to proceed *sola ratione* (by recourse to rational considerations alone), he regards himself as having demonstrated not only that God exists but also that God is triune and that the human soul is immortal. Similarly, in the concluding chapter of the *Cur deus homo*, he claims unhesitatingly that 'whatever is contained in the Old and in the New Testament has been proved, by the solution of the single problem which we have set forth.' Thus he goes further than his predecessor Augustine and his successor Aquinas, neither of whom supposed that either the doctrine of the Trinity or the doctrine of the inerrancy of Scripture is rationally demonstrable.

An example of Anselm's reasoning as a Scholastic may be drawn from *De casu diaboli (The Fall of the Devil)* 21, which asks the question of whether Satan was able to foreknow that he would fall. Anselm begins by stating that either Satan did know this or he did not. On the assumption that he foreknew, he either was willing that the fall

should occur, or he was unwilling. Were he willing, however, then he did not foreknow a future fall, since he was already fallen by virtue of that evil consent; and if he were unwilling, then he who because of his sinlessness deserved to be happy would because of his foreknowledge have been filled with grief, a theological inconsistency. Hence, concludes Anselm, Satan was not able to foreknow that he would fall. In reaching this conclusion, he extends theological truth beyond what is overtly taught in Scripture.

The scholastic method of *sola ratione* characterizes most of Anselm's philosophical and theological works, ranging from the *Proslogion* (whose original title was *Fides quaerens intellectum (Faith Seeking Understanding)*) to the *Cur deus homo*, which appears to be tacitly oriented toward Scripture in spite of its purporting to presuppose nothing about Christ. In *De incarnatione Verbi (The Incarnation of the Word)* 6, Anselm explicitly links the method of the *Proslogion* with that of the *Monologion*, when he states that both of these works show that 'what we hold by faith regarding the divine nature and its persons - excluding the topic of incarnation - can be proven by compelling reasons (*necessariis rationibus*) apart from appeal to the authority of Scripture.' In this context, to proceed *necessariis rationibus*, as do both the *Monologion* and the *Proslogion*, is to proceed *sola ratione*. Moreover, in the *Cur deus homo* he elaborates upon the notion of *rationes necessariae* but does so without this notion's being at odds with that of *remoto Christo* (arguing without recourse to any knowledge of Christ derived from Scripture or from history). Furthermore, in the terminal chapter of the *Cur deus homo*, the interlocutor Boso summarizes: 'You prove the necessity of God's becoming a man, and you do so in such a way that even if the few things you have introduced from our books are removed (for example, what you mentioned about the three persons of God and about Adam), you would satisfy not only the Jews but also the pagans *sola ratione*.' Finally, in *De incarnatione Verbi* 6, Anselm intimates that even the *Monologion* can be characterized as *fides quaerens intellectum*. The same is true of the *Cur deus homo*, as is attested by Boso's last speech in I, 25.

The distinction that Anselm makes in the Cur deus homo between *rationes necessariae* and *rationes convenientes* (fitting reasons) is the following: necessary reasons are reasons that by themselves are compelling, while fitting reasons are weaker reasons that, nonetheless, are compelling in the absence of more powerful contrary considerations. With regard to fitting reasons and their compelling power, Anselm's proposal to Boso in the *Cur deus homo* is crucial:

> I would like for us to agree to accept, in the case of God, nothing that is in even the least degree unfitting and to reject nothing that is in even the slightest degree reasonable unless something more reasonable opposes it. For in the case of God, just as an impossibility results from any unfittingness, however slight, so necessity accompanies any degree of reasonableness, however small, provided it is not overridden by some other more weighty reason.
>
> (*Cur deus homo* I, 10)

The fact that an overall consistency of method characterizes Anselm's treatises does not mean that he never relied upon Scripture to furnish a premise needed for a given argument. However, he sought to avoid doing so, and it is interesting to note that he did not write a single commentary on any book of Scripture.

All of the foregoing serve to distinguish Anselm from Augustine, to whom he is nevertheless deeply indebted. The *Proslogion*'s very description of God as *aliquid quo nihil maius cogitari potest* (something than which nothing greater can be thought) seems to be drawn from *Confessiones* 7.4, and its theme *fides quaerens intellectum* is taken from Augustine's *Sermon* 44.6.7 (and ultimately from Isaiah 7: 9), just as in the *Monologion* the main ideas regarding the Trinity are appropriated from Augustine's *De trinitate (On the Trinity)*. Likewise, Anselm's analysis of free choice in terms of the distinction between having an ability and exercising it, together with his reasoning about the impossibility of an upright human will's being overpowered by the force of temptation, harks back to Augustine. Yet, Anselm is no mere reiterator of Augustine's points. Just as his *Proslogion* argument advances well beyond any cognate systematic reasoning found in Augustine, so his theory of atonement lays a foundation different from Augustine's Devil-ransom theory, though a summary of that theory is incorporated into the *Cur deus homo*. Likewise, his definition of free choice has no parallel in Augustine's writings, even though his concept of 'free choice' is allied with that of Augustine.

3 Philosophical works: *Monologion*

The *Monologion* and the *Proslogion* should be studied together, since each in certain respects elucidates the other.

Anselm himself never thought of the *Proslogion* as replacing the *Monologion*. Although the more creditable (and traditional) view maintains that chapters 1-4 of the *Monologion* aim at proving the existence of God, this view has been challenged by F.S. Schmitt and others. Schmitt (1969) regards Anselm as striving to demonstrate something not about the existence of God but about God's essence: Anselm does not undertake to show that a Supreme Being exists as the cause of the universe, but endeavours instead to prove that the cause of the universe, a cause that is presupposed to exist, is a supreme being (*Analecta Anselmiana* vol. I: 45). Furthermore, Schmitt contends that in these chapters Anselm self-consciously deplatonizes Augustine, from whose *De trinitate* 8.3 he is borrowing. Anselm, he says, leaves aside three Neoplatonistic tenets found in Augustine's discussion: (1) that the concept of the good is impressed (by God) upon the human mind; (2) that the mind beholds both good things and the Good itself; and (3) that good things participate in the Good itself. Critics of Schmitt, however, question whether Anselm is actually borrowing from *De trinitate* 8.3 anything more than Augustine's topic; and if he is not borrowing, then he cannot be said to be deliberately leaving some particular doctrines aside (see Augustinianism).

There is a clear strand of Neoplatonism that runs throughout the *Monologion*. In *Monologion* 9 Anselm teaches the doctrine of exemplarism, although he goes on to maintain that there is only a single Form, or Exemplar, of creation: namely, the second member of the Trinity, or the Eternal Word of God (*Monologion* 30-3). He also accepts a doctrine of degrees of existing, whereby a plant exists more than does a stone, whereas a horse exists more than does a plant (*Monologion* 31). Moreover, he implies that existence is a perfection:

> No one doubts that created substances exist in themselves very differently from the way they exist in our knowledge. For in themselves they exist in virtue of their own being; but in our knowledge their likenesses exist, not their own being. It follows, then, that the more truly they exist anywhere by virtue of their own being than by virtue of their likenesses, the more truly they exist in themselves than in our knowledge.
>
> (*Monologion* 36)

Since Anselm correlates existing truly with existing greatly (*Monologion* 31; *Proslogion* 3), he regards something as existing more greatly (in other words, more perfectly) if it exists 'in its own being' in addition to existing in someone's knowledge. In contrast to the Neoplatonists, however, Anselm holds to a doctrine of analogy on the basis of which some of our predicative discourse about God bears some real but very remote likeness to the Divine Nature (*Monologion* 65; *Responsio* 8). He links our conceiving of God with our conceiving of ourselves:

> The mind, then, can very fittingly be called its own mirror, as it were, in which it beholds, so to speak, the image of this Being which it cannot see face to face. For if of all created things the mind alone can remember itself, understand, and love, then I do not see why we should deny that there is in it the true image of this Being, which exists as an ineffable trinity of self-remembrance, understanding and love.
>
> (*Monologion* 67)

Speaking more generally, Anselm states that 'every being in the degree to which it exists is in that degree similar to the Supreme Being' (*Monologion* 66). These resemblances, though very distant, serve as the foundation for analogical predications (see Existence; God, concepts of; Trinity).

4 Philosophical works: *Proslogion*

No portion of Anselm's literary corpus has evoked as much controversy as has his *Proslogion*. Interpretation of this 'pious breathing', as one translator calls it, is beset by a maze of criss-crossing textual and conceptual problems. Textually, a number of questions are raised. Does Anselm here aim to set forth a proof or proofs of God's existence, and if so, in what chapter(s) is the proof located? Does Anselm's use of the description *id quo maius cogitari nequit* and its variants differ from his use of the description *aliquid quo nihil maius cogitari potest* and its variants; are these really descriptions, or are they definitions? If they are descriptions, are they interchangeable with the description 'greatest conceivable being'? What is meant by *maius*, by *cogitari*, by *intellectus*? If *cogitari posse* means 'to be conceivable', can it be equivalently replaced by 'to be logically possible'? In the opening sentence of *Proslogion* 2, is Anselm asking to be granted to understand *quia es sicut credimus*, or *quia es, sicut credimus*? Does *Proslogion* 2 presuppose that existence is a perfection (an attribute or a property)? In *Proslogion* 4, what is the proper English rendering of *Qui ergo intelligit sic esse deum*? When Anselm writes in *Proslogion* 15,*es quiddam maius quam cogitari possit*, is he claiming that God is inconceivable?

Such textual problems cannot be resolved simply by re-examining the Latin passages, for the proper construal of the respective texts requires the elimination of ambiguities. Invariably, all such translations will be interpretative. Some translators, for example, construe the opening sentence of *Proslogion* 2 as Anselm's request to be granted to understand 'that You exist, even as we believe [You to exist]', so that *sicut credimus* is parenthetical. Others have taken *sicut* to mean 'in the manner that' and have dropped the editorial comma that precedes it, so that Anselm is seen as asking to understand 'that You exist in the manner that we believe [You to exist]'. According to this latter rendering, Anselm is not aiming to prove the existence of God; rather, he wants to show that God, whom he believes to exist, exists in a manner different from the mode of existence of all finite beings: that is, he exists so truly and eternally and immutably that he cannot even conceivably not exist. In comparison with this mode of existence, finite beings are as nothing. In this interpretation, first advanced by Stolz (1933) and subsequently repeated in modified form by many others, the *Proslogion* is primarily an exercise in mystical theology.

By contrast, Malcolm (1960) maintains that *Proslogion* 3 contains the elements of a second version of Anselm's argument, which Malcolm takes to be aiming at demonstration (as does also the first). This second version is supposed to be superior to the reasoning in Chapter 2, inasmuch as it does not require the dubious presupposition that mere existence is a perfection. Instead, it acceptably presupposes that necessary existence is more excellent than is contingent existence. Disagreeing with Malcolm and with almost everyone else, Anscombe (1985) views Anselm as proposing a decidedly different argument. She punctuates a key sentence in *Proslogion* 2 as *Si enim vel in solo intellectu est, potest cogitari esse et in re quod maius est*, and interprets this as: 'For if it [*quo nihil maius*] is only in the intellect, what is greater can be thought to be in reality as well.' Campbell, in turn, perceives Anselm as engaged in the analysis of a speech-act: Anselm's argument 'shows that if one speaks of something-than-which-a-greater-cannot-be-thought one is committed to asserting that such a thing exists' (Campbell 1976: 193-4), but nothing prevents an atheist from refusing to speak of something-than-which-a-greater-cannot-be-thought.

Even one of Anselm's earliest interpreters, Gaunilo, a monk from Marmoutier, had difficulty with Anselm's text. He misinterpreted Anselm to be saying that if *quo nihil maius* 'existed solely in the understanding, then whatever existed also in reality would be greater than it'. He also failed to see that his own counter-argument about an island that excelled in perfection above all other lands did not have a logical structure parallel to the structure of Anselm's argument. For he speaks of something than which no other actually existing land is more perfect; but Anselm speaks of something than which nothing conceivable is perfect. In other respects, Anselm himself also partly misapprehended Gaunilo's critique. He did not recognize that Gaunilo's phrase *maius omnibus* is shorthand for *illud maius omnibus quae cogitari possunt* and not for *illud maius omnibus quae sunt*. Moreover, he incorrectly accused Gaunilo of contradicting himself, an accusation resulting from a misreading of Gaunilo's words *non [posse] hoc aliter cogitare, nisi intelligendo, id est scientia comprehendo, re ipsa illud existere*. Anselm wrongly takes Gaunilo to be defining 'understanding *x*' as 'apprehending with certainty that *x* really exists', whereas Gaunilo is merely defining 'understanding' as 'apprehending with certainty'.

Going beyond the many textual issues, contemporary philosophers have also noted a variety of conceptual problems with the *Proslogion* 2 argument. First, does it confuse two different domains, *de dicto* and *de re*; that is, from something's inconceivable nonexistence, does it follow that that thing exists in reality? Second, does the argument beg the question by assuming that the existence of God is possible? Third, does it beg the question inasmuch as the proposition 'that than which a greater cannot be thought is not that than which a greater cannot be thought' is not self-contradictory unless the existence of that than which a greater cannot be thought is presupposed? This question involves a theory about the logic of definite descriptions: for example, 'the prime number between 3 and 5 is not between 3 and 5' is self-contradictory only if 'the prime number between 3 and 5 is between 3 and 5' is true; and it is true only if there is a prime number between 3 and 5.

Does the argument successfully establish the compossibility of the various perfections that it ascribes to *quo nihil maius*, and does it successfully demonstrate that *quo nihil maius* is indeed a unique being? Is the very concept of *quo nihil maius* intelligible, or is it like the concept of an integer than which none greater can be thought? Can a being that is maximally and absolutely perfect be intelligibly said to be comparatively greater than all existing things, things that are subject to degrees of perfection (greatness)? These and other questions have perpetuated continuing disagreement about just where and how Anselm's argument goes wrong. Among philosophers, the strong feeling that the argument is unsound has contributed to a tendency to regard the entire *Proslogion* as a

meditation on the majesty and greatness of God rather than as an attempted, and failed, set of proofs. However, the fact remains that Anselm, at the end of his reply to Gaunilo, spoke of himself as having engaged in proving:

> I have now showed, I believe, that in the aforementioned treatise I proved - not by inconclusive reasoning but by very compelling reasoning - that something than which a greater cannot be thought exists in reality…. For the signification of this utterance ['something than which a greater cannot be thought'] contains so much force that what is spoken of is, by the very fact that it is understood or thought, necessarily proved to exist in reality and to be whatever ought to be believed about the Divine Substance.
>
> (*Responsio* 10)

Anselm is not a mystic, as some have claimed him to be. Nor should his meditative reflections in the *Proslogion* misleadingly be called confessions, even though their style is reminiscent of Augustine's *Confessiones* (see God, arguments for the existence of).

5 Philosophical works: *De grammatico*

This dialogue reflects the emphasis at Bec on studying the trivium (dialectic, grammar and rhetoric, as opposed to the quadrivium: arithmetic, geometry, astronomy and music); moreover, it bears witness to the esteem in which Aristotle and his *Categories* were held within the School of Bec. The Latin word *grammatica* indicates not only the subject matter that we today call grammar but also the subject matter that we call literature. Anselm himself expresses a disaffection for teaching the former (*Epistola* 64). Contemporary interest in *De grammatico* is due largely to D.P. Henry (1964), who insightfully recognised it as a serious study in paronymy. Paronyms are words (such as *grammaticus*, meaning 'expert in grammar') that derive from other words from which they differ only in case ending (such as *grammatica*, meaning 'expertise in grammar') and that function as both adjectives and nouns (for example, we can say both that someone is expert in grammar and that someone is *an* expert in grammar). Accordingly, the title *De grammatico* is properly translated not as 'On the Grammarian' but as 'On (an) Expert in Grammar', where the parentheses around the indefinite article 'an' serve to signal that it is operative in English when 'expert in grammar' functions as a noun, but is dispensed with when 'expert in grammar' functions as an adjective. (Latin has neither a definite nor an indefinite article.)

In this light, the central question raised within the dialogue is *utrum grammaticus sit substantia an qualitas* (whether (an) expert-in-grammar is a substance or a quality)? Aristotle and medieval textbooks cited grammaticus as an example of quality because the word itself signifies a quality (namely, *grammatica*). However, in the work, Anselm's fictional interlocutor, the Student, points out that in customary parlance *grammaticus* is spoken of as a substance and not as either a quality or an accident. Anselm resolves the question by showing the Student that *grammaticus* signifies both a quality and a substance, for it signifies both expertise in grammar and man, in different ways. Principally and *per se* and substantially it signifies expertise in grammar; and because expertise in grammar is a quality, being expert in grammar is also a quality. Improperly and *per aliud* and accidentally it signifies man; and because man is a substance, being an expert in grammar is also a substance.

Anselm understands a word to have signification insofar as that word calls something to mind. 'Expert in grammar' calls to mind both expertise in grammar and man, though it does so in different ways. A word which signifies *per aliud* (that is, which does not in and of itself call something to mind but instead calls that thing to mind on the basis of something else and accidentally) is said by Anselm to be appellative of the thing that is thus signified *per aliud*. At this point Anselm tightens his terminology, reserving the word 'signifying' for properly signifying, that is, signifying *per se*: 'expert in grammar' signifies, but is not appellative of, expertise in grammar, while it is appellative of, but does not signify, man. Indeed, only man is called (an) expert in grammar because only man has the accident (the quality) of expertise in grammar. A term is appellative of a thing, says Anselm, 'if this thing is called by this name in the customary course of speaking'. Names signify things, not concepts (though there might be second-order names, which signify other names).

In interpreting *De grammatico*, we must beware of two traps: first, that of supposing that *significatio per se* has to do exclusively with meaning and that *significatio per aliud* (that is, *appellatio*) has to do exclusively with reference; and second, that of supposing that Anselm considers *grammaticus est grammatica* to state a logical truth (Henry 1974: 183). With regard to the first misconception, Anselm regards a speaker as at times able to refer to an object either by using the word that *per se* signifies that object, or by using a word that is an appellative of that

object. In a proffered example, he notes that someone might refer to a horse by using the word ‘horse’, which *per se* signifies a horse, that is, which in and of itself calls (the thought of) a horse to mind. Or he might refer to a horse by using the appellative *albus* (‘the white one’), in a situation where a white horse is standing next to a black ox. The hearer would know - not purely on the basis of the signification of the word ‘white’, but also and additionally on the basis of his knowledge that of the two animals, only the horse is white - that the horse was being referred to. Concerning the second misconception, neither Anselm nor the Student regards the phrase *grammaticus est grammatica* as stating a logical truth (while trespassing against *usus loquendi* (ordinary usage)). Rather, both agree that *grammaticus* signifies *grammatica* because it signifies *sciens grammaticam* (having expertise in grammar). Since *grammaticus* is an accident only of man, we say *grammaticus est sciens grammaticam*, rather than saying *grammaticus est grammatica*. This latter sentence is no more a logical truth than it is meaningful. By contrast, the former sentence (*grammaticus est sciens grammaticam*) has a double meaning which results from the fact that *grammaticus* is a paronym: ‘being (an) expert in grammar is having expertise in grammar’.

The real importance of *De grammatico* for Anselm lay in its pedagogical usage as an exercise in dialectic. Its contemporary importance lies in the fact that it attests to Anselm’s awareness of the need to be clear about the elusive relationship between signifier and signified; and it witnesses, further, to his desire to do so by recourse to *rationes necessariae* (*De grammatico* 1) (see Language, medieval philosophy of; Language, philosophy of).

6 Philosophical works: *De veritate, De libertate arbitrii, De casu diaboli*

These three dialogues centre around the notion of *rectitudo* (rectitude, rightness, uprightness). ‘Truth’, says Anselm, is definable as ‘rectitude perceptible to the mind alone’. Justice is rectitude kept for its own sake, and freedom is, also by definition, rectitude of will kept for its own sake. In eliciting his definition of ‘truth’, Anselm examines truth insofar as it is found in statements, thoughts, acts of will, actions, the senses and in the very being of things. A statement is true when it signifies what it ought to, that is, when it signifies correctly; and it signifies correctly when it either signifies to be the case that which actually is the case, or signifies not to be the case that which really is not the case. Thus, a statement’s truth is its rightness, correctness or rectitude.

However, a statement does not always do what it is designed to do (namely, signify correctly what is, or is not, the case), for the statement is also capable of being used to signify falsely (that is, to signify that what-is is not, or that what-is-not is). Even when the statement signifies falsely, it retains the capability to signify truly, because it remains a meaningful set of words. Indeed, such a statement never loses its capability to be used to signify truly and correctly. Insofar as the statement retains this capability, we may also say of it that it is as it ought to be, that it is correct and has correctness, or rightness, or truth.

Thus, reasons Anselm, a true statement has two truths: the truth that accords with its signifying what it is designed to signify, and the truth that accords with its being capable of signifying what it is designed to signify. This latter truth belongs to the statement naturally, inasmuch as it is retained even when the statement signifies otherwise than it is designed to (that is, even when it signifies that what-is is not or signifies that what-is-not is). The former truth belongs to the statement accidentally and only at such time as the statement signifies a correct state of affairs. Anselm follows Aristotle in regarding the statement ‘Socrates is sitting’ as true so long as Socrates is seated but as becoming false when Socrates stands up.

Anselm goes on to point out that the foregoing distinction between two truths applies to thoughts, acts of will, actions, the senses and the very being of things insofar as they all both do what they ought to do and are capable of doing what they ought to do, whether they actually succeed or not. He develops a sense in which thoughts, acts of will and so on may be said to signify, in a twofold way. For example, truth (which is a form of rightness) is said to be in the being of things when these things are as they ought to be (that is, when they conform to what God designed for things of their kind to be); in another sense they ‘signify’ by their very existence that what they are is what they ought to be, whether or not they are defective specimens. For example, a misbegotten animal is not what it ought to be, as measured by the perfection of its species; but it is what it ought to be insofar as it is an existent thing (because otherwise God would not have allowed it to be).

Since a natural rightness belongs to statements, thoughts, acts of will and so on (even when they do not do what they are designed to do, or are not what they are designed to be), this rightness through which they are right

continues to exist when the different statements, thoughts, beings and so on signify falsely. Indeed, this rightness accords with a Rightness that is one and the same in them all, and that continues to exist even when the things themselves and their natural rightness perish. Anselm (borrowing from his own *Monologion* 18) infers that this Rightness, which is perceptible by the mind alone, is eternal. The statement 'something was going to exist' was always true in the past, and the statement 'something has existed' will always remain true in the future; and since each of them can be true only if there is truth, truth is eternal. Like Augustine in *De libero arbitrio (On Free Choice of the Will)*, and like Scripture itself in John 14: 6, Anselm identifies Eternal Truth with God.

In *De libertate arbitrii (Freedom of Choice)*, Anselm searches for a definition of 'free choice' that will elucidate the essential freedom of God, unfallen angels, fallen angels, pre-fallen Adam, post-fallen Adam, Adam's earthly descendants and redeemed Adamic human beings in their sanctified, heavenly state. He concludes that 'freedom of choice' is unexceptionably definable as 'the ability to keep uprightness-of-will for its own sake', an ability which every rational being always possesses. This definition has both philosophical and theological aspects. Philosophically, the definition is meant to exclude the view that having freedom of choice consists essentially in having the power of alternative choice: the ability to choose either that which is morally right or that which is morally wrong. Neither God, the good angels, nor human beings in the heavenly state have this power; and yet Anselm considers them all to have free will. Even regarding cases where agents do have the power of alternative choice, their freedom does not consist therein. For example, when Satan and Adam first sinned (although they were under no necessity to do so), each sinned 'by his own choice, which was free'; but neither did so 'by means of that in virtue of which his choice was free', the ability to maintain a righteous will. Instead, each used another ability, namely, the ability to abandon uprightness of will. Anselm's distinction between having an ability and using that ability - a distinction appropriated from Augustine - allows him to assert, paradoxically, that evil acts of will and evil actions are done freely even though they are not done by means of that ability whose possession is essential to freedom (see Free will; Freedom, divine).

In further expounding his theory of freedom, Anselm encounters a second paradox. He distinguishes ability not only from use but also from motivation; in other words, he distinguishes being *able* to will from being *motivated* to will. However, in some contexts he proceeds to talk as if certain intense motivations were so disenabling that they deprived one's will of freedom:

> If the faithful were immediately transformed at baptism or at martyrdom into the state of incorruption, then merit would perish and men would be saved without merit…. For since men would see those who would be converted to Christ pass over immediately into the state of incorruptibility, there would be no one who would be able even to will to turn away from this very great happiness which he would behold.
>
> (*De concordia* III, 9)

This point of view is reinforced by Anselm's notion that in the future life, redeemed human beings will be like the good angels: they will be unable to sin because they will no longer see anything more to will than that which they shall already have (*De casu diaboli* 6). Anselm also closely links motivation and ability/inability when he states that if Satan had known that he would actually be punished if he sinned, rather than knowing simply that he could deservedly be punished, then 'he would not have been able freely to will what would have caused him to be unhappy' (*De casu diaboli* 23). In other words, he would not have been able to sin. However, if certain patterns of intense motivation are disenabling, how is it plausible that the will of a rational creature is really always free?

On the more centrally theological plane, a third paradox arises. Although Anselm teaches that each human being always retains the ability to keep uprightness of will, he also teaches that no human being has the ability, by his own efforts, to regain uprightness of will once it has been abandoned. Only God can restore this uprightness; but God does so only with respect to those who repent. This doctrine leaves Anselm in the awkward position of maintaining that if someone's will lacks uprightness, then he cannot will uprightly, even though he is still free by virtue of his having the ability to keep uprightness of will if uprightness of will were restored to him. In having an ability that cannot be used, such a man is said by Anselm to resemble someone who has the power of sight but who cannot use this power because he is located in a totally darkened room; still, he could actually see if light were restored to the room. There is something strange about Anselm's calling someone's will free when it is not free actually to will uprightly. Recognizing this point, Anselm refers to a man as both a servant and free. This paradox reaches its most concise expression in *De concordia* III, 13: 'without justice the will is never free, because

without justice the natural freedom of choice is futile'.

A fourth paradox emerges in conjunction with the third. If one's will is no longer upright by virtue of one's having willed evilly, and if God restores uprightness only to the will of those who repent, how can one be willing to repent unless one's will is to some extent already upright prior to the restoration of uprightness? Anselm expresses this paradox in terms of conversion (though it applies, as well, after conversion): 'Those who say "Convert us, O God" are already to some extent converted, because in willing to be converted they have an upright will' (*De concordia* III, 6). By prevenient grace, maintains Anselm, God guides toward repentance those whom he has foreordained.

A fifth paradox surfaces when Anselm contends that the force of temptation can never overpower an upright will. The force of temptation, if it were to succeed, would have to compel an individual to accede; but, philosophizes Anselm, again following Augustine, no one can be constrained to will. A soldier, for example, can be tortured or even killed against his will; however, he cannot be made to will against his will, for anyone who wills consents to will, that is, wills willingly. Anselm construes 'being constrained to will' as 'being made to will unwillingly'; and this latter expression he regards as incoherent, on the grounds that if one were unwilling to will, then there would be no actual willing. However, a critic could contend that Anselm's analysis of 'constrained to will' is tendentious and oversimplified.

A sixth paradox has to do with Anselm's view that Satan's will to persevere in justice both failed him and did not fail him. It failed him because, although he was created with a preponderant, supernatural inclination-for-justice, he nonetheless willed unjustly. On the other hand, it did not fail him because the supernatural inclination for justice, with which he was created, did not wane. In attempting to dissolve the paradox of how Satan, who was created with a strong inclination for justice and who was placed in an environment with reinforcing incentives, would have been motivated to will evilly, Anselm introduces a 'Scholastic-like' distinction: it is not the case that Satan's willingness to desert justice was caused by an antecedent waning of the willingness to keep justice but, rather, the willingness to desert caused the not-willingness to keep. Satan willed to desert justice in order to secure a desired benefit that it was unjust for him to have at that time. The deeper paradox now becomes one of why Satan, a rational creature, would irrationally and knowingly choose to have a benefit the choosing of which would risk the loss of all happiness? Anselm's answer, 'only because he willed to', transforms the paradox into a mystery.

A final paradox relates to Anselm's statement that:

> God causes all the things which are done by a just or an unjust will, viz., all good and evil deeds. Indeed, in the case of good deeds he causes what they are [essentially] and the fact that they are good; but in the case of evil deeds he causes what they are [essentially] but not the fact that they are evil.
>
> (*De concordia* I, 7)

These paradoxes Anselm himself regards as only apparent. Yet, whatever be one's own final appraisal of Anselm's theory of freedom, one must bear in mind that Anselm is a determinist and a compatibilist: only Satan's initial choice of evil was done simply because Satan willed thus to choose. All other reflective choices are such that they are 'caused' by factors that, in principle, are subject to investigation and description (and that do not deprive the agent of free choice):

> Even as every will wills something, so it also wills for the sake of something. And just as we must consider what it wills, so we must also notice why it wills. For a will ought to be upright in willing what it ought and, no less, in willing for the reason it ought. Therefore, every will has both a what and a why. Indeed, whatsoever we will, we will for a reason.
>
> (*De veritate* 12)

Having in *De veritate (On Truth)* defined 'justice' as 'uprightness of will kept for its own sake', and having indicated that he means 'kept for its own sake only', Anselm in *De casu diaboli (The Fall of the Devil)* interprets the evil of injustice to be the absence of justice from a will in which justice ought to be present. Accordingly, injustice is a privation; and a privation, he thinks, is a kind of not-being. However, if the evil that is injustice is a kind of not-being, how can the word 'evil' be significative, wonders the Student interlocutor. The Teacher,

Anselm, gives three replies. First, ‘evil’, ‘injustice’ and ‘nothing’ signify, respectively, a removal of good, justice and something. Since a removal can be signified only by also signifying that which is to be removed, expressions such as ‘not good’, ‘not just’ and ‘not something’ obliquely signify good, just and something, without signifying them by positing them. Second, ‘evil’ and ‘injustice’ do not signify anything *secundum rem* (according to fact) but they are significative *secundum formam* (according to linguistic form), for we use these terms to speak of evil and injustice as if these latter were something - as when we say ‘evil caused it’ or ‘injustice caused it’. Third, ‘evil’ is sometimes used to signify *incommodum* (detriment) rather than to signify injustice or an absence; in this respect it signifies something and not merely as-if-something (*quasi aliquid*).

In *De casu diaboli*, Anselm seeks to remind us that ‘we ought not to cling to the verbal impropriety concealing the truth as much as we ought to attend to the true propriety hidden beneath the many types of expression’. In the process of heeding his own injunction, he there appeals to many of the same distinctions that are also found in his *Philosophical Fragments*.

7 Theological works: *De incarnatione Verbi* and *De processione Spiritus Sancti*

De incarnatione Verbi (The Incarnation of the Word) contains Anselm’s defence of the doctrine of God’s triunity. This defence is mounted against Roscelin of Compiègne, a French cleric who maintained that if the three persons of God were only one thing and not three separate things, then the incarnation of the Son would, untowardly, have necessitated the incarnation of the Father and the Holy Spirit as well. Anselm examines the meaning of ‘thing’ (*res*) and explains that the three persons are three distinct (but not separate) things with respect to their distinguishing properties, but are one thing with respect to the nature that they have in common. Exhibiting masterly skill as a dialectician, he contends that Roscelin’s view leads either to tritheism or to Sabellianism, both of which doctrines are objectionable to Roscelin himself. He further contends that Roscelin’s own articulation of his own claim is such that it entails the very same untoward consequence from which Roscelin is attempting to escape: namely, the incarnation of the Father and the Holy Spirit along with that of the Son. After advancing an argument to show that there can be no more than one God, Anselm explains why this one God became incarnate in the person of the Son rather than in the person of the Father or of the Holy Spirit. To complete his dialectical strategy, he provides a symbolic illustration to lend credence to the very notion of triunity: the Nile, while being one body of water, is also a three-ness of spring, river and lake.

The divine trinity and oneness is explored more fully in *De processione Spiritus Sancti (The Procession of the Holy Spirit)*, where the Augustinian emphasis upon *relatio* looms larger. The three persons of God are three relations, three ways in which the one Divine Nature relates itself to itself. These relations are irreducible, inasmuch as the Father would not be a father unless he had a son, but the Father himself cannot be his own son; the Son would not be a son unless he had a father, but the Son himself cannot be his own father; and the Holy Spirit would not be a common spirit unless he proceeded from both the Father and the Son, but he himself cannot be either of the two from whom he proceeds. In line with the Nicene-Constantinople creed, Anselm defends the *filioque* doctrine against the Greek Orthodox teaching that the Holy Spirit proceeds from the Father alone. No other medieval treatises so concisely, clearly and vigorously display the rationale of orthodox trinitarianism as do these two Anselmian works (see Trinity).

8 Theological works: *Cur deus homo* and *De conceptu virginali*

Desirous to explain the reasons for the Incarnation, Anselm undertakes in the *Cur deus homo* an elaborate demonstration of his view that only by means of incarnation could God have made provision for human salvation. Adam’s sinful fall, as well as the ensuing sins of all his natural descendants, could not simply be overlooked by God, since doing so would do violence to God’s justice. Hence, every sinner is under obligation to make payment for his sins, sins which rob God of honour by detracting from the splendour of creatures and therefore from God’s own accomplishment as Creator. Payment must involve both restitution of what has been stolen and compensation for the injury done. Now, every rational creature, even if sinless, owes to the Creator perfect and voluntary obedience - indeed, owes itself and all that it is. A sinner, who has stolen honour from God by stealing from God the obedience due to him, is obliged to resume rendering the requisite obedience and, in addition to make compensation, or satisfaction, for the perpetrated wrong. In this twofold way the sinner would fully restore God’s honour.

However, once having sinned, the sinner's will is deprived of uprightness, which by its own efforts it cannot recover; and hence no sinner can attain perfect obedience. Moreover, the sinner has nothing with which to compensate God. Compensation must be equal to the gravity of the wrong that was done; but any disobedience to God's command is so grave, teaches Anselm, that it ought not to be done even if one could thereby save an infinite number of worlds from perishing. Since the sinner is obliged to offer as satisfaction, or compensation, something greater than is that for whose sake he is obliged not to dishonour God, and given that he is not to dishonour God even in order to save an infinite number of worlds from perishing, he must render to God something of greater value than are an infinite number of worlds. But only God can make such a payment, whereas only a human being of Adam's race ought to. Hence, if payment is to be made, it must be made by a God-man.

Yet, queries Anselm, what payment could a God-man make in addition to perfect and voluntary obedience? What is it that he would not already owe to God and that would count as satisfaction? The God-man, reasons Anselm, would make satisfaction by choosing righteousness in preference to life. Being sinless, and in this respect unlike all other human beings, he alone would not be under the penalty-of-death, which resulted from the Fall. He would yield up his life to the honour of God by allowing himself to be put to death for righteousness' sake - that is, to be put to death for refusing, when accused of blasphemy, to profess the untruth that he was not God. Since the evil of sinning against the person of God is the greatest conceivable evil, the goodness of the God-man's life must be the greatest conceivable goodness. Thus in choosing righteousness even in the face of death, the God-man restored honour to God by rendering a service more valuable than the value of the totality of things that are not God. Such a service deserves to be rewarded. And because the God-man needs nothing, he may admissibly transfer this reward to cancel the debt incurred by Adam and his natural descendants (see Atonement; Salvation; Sin).

In the course of his argument, Anselm explains once again why it was most fitting for the Son of God and not for either of the other two members of the Trinity to become incarnate. He hastens to point out that those who actually put to death the Jesus of history were guilty of only a venial sin, having acted in ignorance of his Messianic identity. In *De conceptu virginali et originali peccato (The Virgin Conception and Original Sin)*, he distinguishes between Adam's person and his nature. When Adam sinned personally, his sinful person contaminated his nature. As a result, his descendants are born with an individuated Adamic nature whose will lacks justice. This unjust nature contaminates their persons, predisposing them to sin personally when they reach the age of accountability. Anselm offers an explanation of how Jesus, who was of Adam's nature, was nonetheless able to remain free of original sin - that is, remain free of contracting the necessity-of-sinning upon attaining the age of reason. He insists that Adam's descendants are not condemned for Adam's sin but only for their own, even though Adam's sin is a cause of their own fallen nature. Anselm does not teach the doctrine of Mary's own immaculate conception (see Incarnation and Christology).

9 Theological works: *De concordia*

Oriented around the topics of grace, foreknowledge and predestination, *De concordia praescientiae et praedestinationis et gratiae dei cum libero arbitrio (The Harmony of the Foreknowledge, the Predestination and the Grace of God with Free Choice)* not only extends the views previously advanced in *De libertate arbitrii* and *De casu diaboli* but also places them squarely within a theological context. Predestination is a theological doctrine that relates primarily to the attendant doctrine of salvation. The New Testament seems to teach that some individuals are foreordained to salvation, whereas others are not. Anselm follows Augustine both in accepting this doctrine and in acknowledging that he has no explanation for why God singles out some and not others. As early as *Proslogion* 11 he prayed: 'But if we can somehow grasp why You can will to save those who are evil, surely we cannot at all comprehend why from among those who are similarly evil You save some and not others because of Your supreme goodness, and condemn some and not others because of Your supreme justice.' Yet predestination and enabling grace do not violate human freedom, believes Anselm, because inducing someone's assent and trust is not identical with compelling assent and trust.

Three considerations reassure Anselm that divine foreknowledge, like predestination, is not incompatible with human free will. First of all, God foreknows not only what a given individual will do but knows also *how* the individual will do it, whether freely or by necessity. Assume that the individual will act freely: then, if it is necessary that what-God-foreknows come to pass, it is necessary that the individual do freely what will be done. This necessity which guarantees one's freedom cannot at the same time be a necessity that deprives one of

freedom. Anselm believes that the expression ‘it is necessary that… ’ is misleading because it tends to suggest that the event in question happens by necessity, whereas in the present context it means only ‘it is certain that… ’. Second, something may be certain to occur either because God efficaciously wills that it occur, or because he foreknowingly wills to permit its occurrence. In the latter case, his foreknowledge does not efficiently cause the action or the event to happen. In fact, Anselm reminds us as early as *De casu diaboli* 21, God is improperly said to foreknow the future, since his eternal knowledge is knowledge of things as present. The present knowledge of something’s happening does not necessarily make it happen. Third, God’s will must be (conceived of as) free, because otherwise God could be thought to be more perfect - something which is impossible. By the same token, he must be regarded as knowing his own will. If his knowledge of his own will does not deprive it of freedom, then *prima facie* there is no reason to infer that his knowledge of the human will deprives it of freedom (see Eternity; Free will; Grace; Omniscience; Predestination; Salvation).

10 Conclusion

Anselm’s paramount publications, though only eleven in number, have had a significant (though never major) impact upon the history of Western philosophy. Peter Abelard, Thomas Aquinas, John Duns Scotus, William of Ockham, Nicholas of Cusa and others found it helpful to quote from them, sometimes approvingly, sometimes disapprovingly. Anselm’s learned command of Aristotle’s *logica vetus*, his insightful appeal to the Latin community’s *usus loquendi* and his bold utilization of the method of *sola ratione* bear ample witness to his gifted philosophical mind. Yet, like any good philosopher, he knew when to be tentative, as evidenced by his conclusion to *De conceptu virginali*:

> In accordance with the capacity of my understanding I have briefly made these statements about original sin - not so much by way of asserting them as by way of provisionally inferring them - until God shall somehow reveal to me something better. But if someone has a different view, I do not reject anyone’s opinion provided it can be proved to be true.
>
> (*De conceptu virginali* 29)

In spite of his emphasis upon *rationes necessariae*, he sensed a danger in trusting bloated, over-confident and self-confident reason. Consequently, he was not averse to declaring with respect to deep theological mysteries: ‘if someone thinks he knows something, he does not yet know it as he ought to know it’ (*De incarnatione Verbi* 1). Because he regarded such knowledge and understanding as a grace, he was willing to continue to believe - in the hope of one day better comprehending.

See also: Augustinianism; Free will; God, concepts of; Language, medieval theories of; Logic, medieval; Medieval philosophy; Omniscience; Platonism, medieval; Salvation; Trinity

JASPER HOPKINS

List of works

Anselm of Canterbury (1033-1109) *S. Anselmi Opera Omnia (Complete Works of Saint Anselm)*, ed. F.S. Schmitt, Stuttgart: Frommann.(This edition is a reprint, consisting of six volumes in two tomes, of Schmitt’s earlier edition of Anselm’s works, published by Thomas Nelson.)

Anselm of Canterbury (1033-1109) Philosophical Fragments, ed. F.S. Schmitt, *Ein neues unvollendetes Werk des hl. Anselm von Canterbury*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters 33. 3, 1936; ed. R.W. Southern and F.S. Schmitt, *Memorials of St. Anselm*, London: British Academy, 1969. (Southern and Schmitt includes the philosophical fragments as well as the compilations *Dicta Anselmi* and *De morum qualitate per exemplorum coaptationem*.)

Anselm of Canterbury (1033-1109) *Fragmenta Philosophica (Philosophical Fragments)*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 2, New York: Mellen, 1976.(A translation of philosophical fragments.)

Anselm of Canterbury (1070-1109) *Epistolae (Letters)*, trans. W. Fröhlich, *The Letters of Saint Anselm of Canterbury*, Kalamazoo, MI: Cistercian Publications, 1990-4, 3 vols.(Anselm’s collected letters in translation.)

Anselm of Canterbury (1072-1103) *Orationes sive meditationes (Prayers and Meditations)*, trans. B. Ward, *The Prayers and Meditations of Saint Anselm*, Baltimore, MD: Penguin, 1973.(Collected theological writings.)

Anselm of Canterbury (1076) *Monologion*, trans. J. Hopkins, *A New Interpretive Translation of St. Anselm’s*

Monologion and Proslogion, Minneapolis: Banning, 1986.(Anselm's first major philosophical work, showing the influence of Neoplatonism.)

Anselm of Canterbury (1077-8) *Proslogion*, trans. J. Hopkins, *A New Interpretive Translation of St. Anselm's Monologion and Proslogion*, Minneapolis: Banning, 1986.(Includes Gaunilo's *Pro insipiente (On Behalf of the Fool)* and Anselm's *Responsio (Reply to Gaunilo)*.)

Anselm of Canterbury (*c.*1080-5?) *De grammatico*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 2, New York: Mellen, 1976.(Grammatical work, notable for its emphasis on signification.)

Anselm of Canterbury (*c.*1080-5) *De veritate (On Truth)*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 2, New York: Mellen, 1976.(Contains Anselm's definition of truth, identifying eternal truth with God.)

Anselm of Canterbury (*c.*1080-5) *De libertate arbitrii (Freedom of Choice)*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 2, New York: Mellen, 1976.(Searches for a definition of 'free choice'.)

Anselm of Canterbury (*c.*1085-90) *De casu diaboli (The Fall of the Devil)*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 2, New York: Mellen, 1976.(Anselm's work on the relationship between sin and free will.)

Anselm of Canterbury (1092-4) *De incarnatione Verbi (The Incarnation of the Word)*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 3, New York: Mellen, 1976.(Anselm's defence of the doctrine of God's triunity.)

Anselm of Canterbury (completed 1098) *Cur deus homo (Why God Became a Man)*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 3, New York: Mellen, 1976.(Demonstrates Anselm's view that only through incarnation could God have made provision for human salvation.)

Anselm of Canterbury (1099-1100) *De conceptu virginali et originali peccato (The Virgin Conception and Original Sin)*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 3, New York: Mellen, 1976. (Discussion of the nature of sin, including original sin.)

Anselm of Canterbury (1102) *De processione Spiritus Sancti (The Procession of the Holy Spirit)*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 3, New York: Mellen, 1976.(A further exploration of Trinity and oneness, showing Augustinian influences.)

Anselm of Canterbury (1106-7) *Epistolae de sacramentis (Letters on the Sacraments)*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 3, New York: Mellen, 1976.(Anselm's views on the sacraments.)

Anselm of Canterbury (1107-8) *De concordia praescientiae et praedestinationis et gratiae dei cum libero arbitrio (The Harmony of the Foreknowledge, the Predestination and the Grace of God with Free Choice)*, trans. J. Hopkins and H. Richardson, *Anselm of Canterbury*, vol. 3, New York: Mellen, 1976.(Expands the views previously advanced in *De libertate arbitrii* and *De casu diaboli*.)

References and further reading

Adams, M. (1992) '*Fides Quaerens Intellectum*: St. Anselm's Method in Philosophical Theology', *Faith and Philosophy* 9: 409-35.(Relates Anselm's methodology to his station in life.)

Anscombe, G.E.M. (1985) 'Why Anselm's Proof in the *Proslogion* is not an Ontological Argument', *Thoreau Quarterly* 17: 32-40.(An extreme, but extremely intriguing, interpretation.)

Campbell, R. (1976) *From Belief to Understanding: A Study of Anselm's Proslogion Argument on the Existence of God*, Canberra: Australian National University.(More important for the major philosophical issues it insightfully develops than for the controversial conclusions about Anselm that it reaches.)

Eadmer (1093-1125) *Vita Anselmi (The Life of St Anselm, Archbishop of Canterbury)*, ed. and trans. R.W. Southern, London: Thomas Nelson, 1962.(Appears to have been written in two stages, roughly, between 1093 and 1100, and between 1109 and 1125.)

Eadmer (1109-14) *Historia Novorum in Anglia*, ed. M. Rule, London: Longmans, 1884.(Later edition trans. G. Bosanquet as part of *Eadmer's History of Recent Events in England*, London: Cresset, 1964. A record of Anselm's public life during his tenure at Canterbury, completed between 1109 and 1114.)

Evans, G. (ed.) (1984) *A Concordance To the Works of St. Anselm*, Millwood, NY: Kraus.(A valuable tool, in spite of containing too many trivial entries.)

Fröhlich, W. (1984) 'The Letters Omitted from Anselm's Collection of Letters', *Anglo-Norman Studies 6: Proceedings of the Battle Conference, 1983*, Woodbridge: Boydell, 58-71. (Assesses Anselm's motivation in not having certain letters copied for posterity and indicates the extent to which this 'suppression' occurred.)

Gombocz, W. (1980) 'Anselm von Canterbury. Ein Forschungsbericht über die Anselm-Renaissance seit 1960' (Anselm of Canterbury: A Research Report on the Anselm Renaissance since 1960), *Philosophisches Jahrbuch* 87: 109-34.(Contains a wealth of bibliography.)

Henry, D.P. (1964) *The De Grammatico of St. Anselm: The Theory of Paronymy*, Notre Dame, IN: Notre Dame University Press.(Provides a translation and an analysis.)
Henry, D.P. (1967) *The Logic of Saint Anselm*, Oxford: Clarendon.(Skillfully covers the principal aspects).
Henry, D.P. (1974) *Commentary on De Grammatico: The Historical-Logical Dimensions of a Dialogue of St. Anselm's*, Dordrecht: Reidel.(A painstakingly detailed further scrutiny of Anselm's views on paronymy.)
McIntyre, J. (1954) *St Anselm and his Critics: A Reinterpretation of the Cur Deus Homo*, London: Oliver & Boyd.(An older but penetrating examination of Anselm's theory of atonement.)
Malcolm, N. (1960) 'Anselm's Ontological Arguments', *Philosophical Review* 69: 41-62.(Claims that Anselm had two different ontological arguments in *Proslogion* 2 and 3, respectively.)
Rowe, W. (1976) 'The Ontological Argument and Question-Begging', *International Journal for Philosophy of Religion* 7: 425-32.(Makes a strong case for regarding the argument of *Proslogion* 2 as question-begging.)
Schmitt, F.S. (1968) 'Prolegomena seu Ratio editionis', *S. Anselmi Opera Omnia,* Stuttgart: Frommann, Tome 1, 1*-244*.(A cluster of critical considerations regarding the chronology, authenticity and manuscript tradition of Anselm's works.)
Schmitt, F.S. (1969) 'Anselm und der (Neu-)Platonismus' (Anselm and (Neo-) Platonism), *Analecta Anselmiana* I: 39-71.(Looks at Neoplatonic elements in Anselm's writings.)
Serene, E. (1981) 'Anselm's Modal Conceptions', in S. Knuuttila (ed.) *Reforging the Great Chain of Being: Studies in the History of Modal Theories*, Dordrecht: Reidel, 117-62. (Explores Anselm's *Philosophical Fragments*, which are not dealt with in this entry.)
Southern, R.W. (1988) 'Sally Vaughn's Anselm: An Examination of the Foundations', *Albion* 20: 181-204. (Sharpens the issues that separate the two conflicting interpretations of Anselm's life.)
Southern, R.W. (1990) *Saint Anselm: A Portrait in a Landscape*, Cambridge: Cambridge University Press. (Replaces his older work *Saint Anselm and his Biographer*, 1963.)
Stolz, A. (1933) 'Zur Theologie Anselms im Proslogion' (Anselm's Theology in the *Proslogion*), *Catholica* 2: 1-24.(Argues that the *Proslogion* is an exercise in mystical theology.)
Vaughn, S.N. (1987) *Anselm of Bec and Robert of Meulan: The Innocence of the Dove and the Wisdom of the Serpent*, Berkeley, CA: University of California Press. (Suggests that Anselm was a more gifted and scheming statesman than Eadmer makes him out to be.)
Vaughn, S.N. (1988) 'Anselm: Saint and Statesman', *Albion* 20: 205-20.(A response to Southern's critique of her historical portrait of Anselm.)
Vuillemin, J. (1971) *Le Dieu d'Anselme et les apparences de la raison (The God of Anselm and the Semblances of Reason)*, Paris: Montaigne.(A clear appraisal of Anselm's concept of God as it functions in the *Monologion* and the *Proslogion*.)

Anthropology, philosophy of

Anthropology, like philosophy, is multifaceted. It studies humans' physical, social, cultural and linguistic development, as well as their material culture, from prehistoric times up to the present, in all parts of the world. Some anthropological sub-fields have strong ties with the physical and biological sciences; others identify more closely with the social sciences or humanities. Within cultural and social anthropology differing theoretical approaches disagree about whether anthropology can be a science. The question of how it is possible to understand cultures different from one's own, and to transmit that knowledge to others is central to anthropology because its answer determines the nature of the discipline. Philosophy of anthropology examines the definitions of basic anthropological concepts, the objectivity of anthropological claims and the nature of anthropological confirmation and explanation. It also examines the problems in value theory that arise when anthropologists confront cultures that do not share their own society's standards.

1 Epistemological problems

Despite a shared commitment to fieldwork and to the role of anthropologists as participant-observers, different theoretical schools of anthropology disagree sharply about how it all works, and in particular about whether anthropology is or can be a science. According to one prominent school - cognitive anthropology - culture consists of a set of rules in the minds of its members (Frake 1969). To understand another culture, anthropologists must internalize those rules well enough to respond to new situations in a culturally appropriate way. In its approach to cultural behaviour as rule-governed, and in its broad interpretation of rules, cognitive anthropology accords with Winch's (1958) Wittgensteinian account of the nature of human social life (see Social norms). Winch, however, does not share cognitive anthropology's view that anthropology is a science. Linguistic rules - primary examples of cultural rules - offer an entry into the rest of the culture. To begin to learn a new set of cultural rules, cognitive anthropologists question informants about how they would classify various observable phenomena. The point is not merely to discover the names that others give to the objects and categories familiar to the anthropologist, but to grasp the other culture's own, possibly different, way of structuring the stream of experience into categories. To this end, cognitive anthropologists - who are also called 'ethnoscientists' - elicit indigenous classifications of features of the natural world (for example, colours, birds, fish, plants). They test their grasp of the rules by observing whether their own efforts at extending classifications gain the approval of their informants. When the anthropologists have mastered the new system, they compare and contrast it with their own (Western, scientific) system. The results of such investigations are usually put forth as evidence against radical forms of cognitive relativism (see Social relativism §3; Relativism).

Cognitive anthropology offers a clear answer to the first part of the central question raised above: one comes to understand another culture by becoming proficient in its rules through interaction with members of the culture. How that knowledge is transmitted to those not initiated into the culture remains problematic, however, for it is not clear that one system of rules can be translated into another. Although cognitive anthropologists view themselves as engaging in scientific studies of culture, and are concerned with the objectivity and predictive success of their claims, their account of the nature of cultural knowledge and how it is acquired works against its dispersal in the usual scientific channels, such as journal articles.

An opposing approach, symbolic anthropology, maintains that anthropology is ethnography. The business of anthropology is to write accounts of human culture that can be read, analysed, discussed and challenged by students and peers, now and in the future. Doing anthropology does not require one to internalize some set of rules, as cognitive anthropologists suppose, but instead to figure out what people are up to when they say and do certain things. Culture is not a set of rules located in anyone's mind, it is symbolic behaviour. Culture is thus a 'public document' that anthropologists are trained to 'read', just as literary critics read (interpret) poems and novels.

Comparing anthropology with literary interpretation highlights the issue of objectivity. Do standards exist for 'correct' interpretations of cultures or do anthropologists have the same latitude as literary critics to offer alternative, and perhaps incompatible, accounts? Geertz (1975), borrowing Ryle's terminology, notes that ethnography at its most basic level involves 'thick descriptions' rather than the reporting of 'raw' data. That is to say, any description of human behaviour (as opposed to mere bodily movement) imputes an intention to it, and

thus, at least partially, explains the behaviour. Collingwood in *The Idea of History* (1946) had already stressed the same point, using it to argue for a fundamental difference between interpretive explanations of human behaviour and causal explanations of 'mere' physical events (see Collingwood, R.G.). Nevertheless, he insisted that hypotheses about intentional descriptions could be subjected to rigorous examination, and accepted or rejected objectively, on the basis of evidence. Latter-day interpretivists, however, regard the impossibility of peeling away all layers of interpretation to get at 'the fact of the matter' as support for relativism (see Social science, contemporary philosophy of §3). One interpretation may be preferred to another because it is more coherent, richer, more subtle or relevant to contemporary concerns, but not because one is true and the other false. The latter terms are not applicable, at least in their usual sense, to interpretations. Because symbolic anthropologists emphasize the importance of interpretation over causal explanation and prediction, they see a significant gap between anthropology and science. While cognitive anthropology and symbolic anthropology are dominant theoretical approaches and loom large in philosophical discussions, many anthropologists, particularly those whose training was completed when earlier schools (for example, functionalism, historical particularism) flourished would not identify themselves with either school.

2 Ethical problems

Respect for the beliefs, practices and values of other cultures, no matter how different from one's own, is a hallmark of anthropological wisdom. Franz Boas (1940), whose name is inevitably linked with cultural relativism, rejected invidious comparisons between the 'high culture' of northern Europeans and the art forms, languages, myths and religious practices of indigenous Americans, just as he rejected 'progressive' evolutionary accounts, offered by physical anthropologists early in the twentieth century, that put the former at a more advanced stage of physical development than the latter. He insisted that the culture of each group should be studied in terms of its own historical development and appreciated in that context rather than judged by the standards of another culture.

Since moral beliefs and practices, like other beliefs and practices, depend on the cultural context, many anthropologists regard ethical relativism as an easy consequence of cultural relativism. Some moral philosophers have challenged this slide into relativism by claiming that while societies differ in their derivative moral judgments (for example, the propriety of cross-cousin marriage), they agree in their more fundamental moral judgments (for example, the immorality of incest). Whether universal agreement exists on any basic moral judgment is at least in part an empirical question - one to which Turnbull's studies of the Ik (1972) suggest a negative answer. A different attack on relativism maintains that establishing genuine relativism requires showing that people's basic ethical judgments would conflict even if they shared all the same factual beliefs and were fully enlightened as to the consequences of their views. On this view, the mere absence of universally accepted principles is not sufficient to prove relativism. By the same token, finding universally accepted moral principles would not disprove their dependence on particular cultures. The agreement could be accidental.

Despite such philosophical attacks on ethical relativism, most practising anthropologists continue to embrace it, for they equate ethical relativism with tolerance for the moral codes of others. Teachers of anthropology present relativism as the received view, usually also admonishing students to distinguish objective factual claims from subjective and relative value judgments. The student's ability to reserve judgment on the moral codes and practices of others is regarded as a prerequisite for anthropological fieldwork and for acceptance into the discipline.

Anthropologists' alleged commitment to relativism, however, is undermined by their own professional code of behaviour, as stated, for example, in the guidelines of the Society for Applied Anthropology (Bernard 1988). The dominant theme of these guidelines is Kantian: treat the people whom you study as ends not as means. Respect their right to self-determination, to arrange their lives as they see fit according to their standards. Anthropologists who are accused of violating this code must defend themselves satisfactorily or face professional sanctions.

Most field anthropologists believe that relativism permits or even requires them to defend 'their people' against interference from governments, missionaries, or other agents of so-called advanced civilizations. The common justification for such defences, however, is the society's right of self-determination. Anthropologists also take up the cause of oppressed minorities within the societies that they study. For example, feminist anthropologists work to improve the status of women in many cultures. Thus even when denying absolute moral values, anthropologists embrace moral guidelines that are supposed to hold across cultures. They unilaterally condemn behaviour that infringes on the rights of others.

See also: Universalism in ethics; Value judgments in social science

MERRILEE H. SALMON

References and further reading

Bernard, H. (1988) *Research Methods in Cultural Anthropology*, Newbury Park, CA: Sage Publications.(See Appendix A for 'Statement on professional and ethical responsibilities. Society for Applied Anthropology'.)

Boas, F. (1940) *Race, Language and Culture*, New York: Free Press.(Presents Boas' historical-particularist account of culture.)

Collingwood, R. (1946) *The Idea of History*, Oxford: Oxford University Press. (States the classic interpretivist position.)

Frake, C. (1969) 'The ethnographic study of cognitive systems', in S. Tyler (ed.) *Cognitive Anthropology*, New York: Holt, Rinehart and Winston, 28-41.(A brief introduction to cognitive anthropology.)

Geertz, C. (1975) *The Interpretation of Cultures*, London: Hutchinson.(See Introduction for modern interpretivist view by a prominent symbolic anthropologist.)

Hollis, M. and Lukes, S. (eds) (1982) *Rationality and Relativism*, Cambridge, MA: MIT Press.(Essays by contemporary philosophers and anthropologists.)

Salmon, M. (1992) 'Philosophy of the social sciences' in M. Salmon, J. Earman, *et al.*, *Introduction to the Philosophy of Science*, Englewood Cliffs, NJ: Prentice Hall, 404-25.(Discusses the senses in which social sciences can be scientific.)

Sperber, D. (1993) 'Remarques anthropologiques sur le relativisme moral' (Anthropological remarks on moral relativism), in J.-P. Changeux (ed.), *Fondements Naturels de l'Ethique (Natural Foundations of Ethics)*, Paris: Editions Odile Jacob, 319-34.(A critical account of ethical relativism.)

Turnbull, C. (1972) *The Mountain People*, New York: Simon & Schuster.(An account of the Ik.)

Winch, P. (1958) *The Idea of a Social Science and its Relation to Philosophy*, London: Routledge & Kegan Paul. (Wittgensteinian account of social science.)

Antiochus (*c.*130-68 BC)

For most of his career the Greek philosopher Antiochus of Ascalon, a pupil of Philo of Larissa, was an orthodox 'sceptical' Academic. He then changed his philosophy: some called him a Stoic, but he himself claimed to be returning to the Old Academy of Plato and his immediate successors. He took a generous view of his new home, urging that the Peripatetics and the Stoics were not new schools of thought but mere modifications of Platonism, and the philosophical position which he advocated was a 'syncretism' - an amalgam of ideas and doctrines and arguments taken from several sources. To philosophy itself he contributed little, but he was a figure of considerable importance in the larger world, where he presented Greek philosophy to an educated Roman public.

1 Life

Antiochus came from Ascalon in Syria. Circumstantial evidence places his birth *c.*130 BC. At some point he went to Athens, where he studied philosophy with Philo of Larissa and perhaps also with the Stoic Mnesarchus. The Academy won his heart (see Academy): he remained under Philo's wing for an unusually long period and wrote books in defence of Academic scepticism. After years of fidelity he defected - perhaps in the 90s BC. He abandoned scepticism, claiming to restore the true Academy of Plato. The autumn of 87 BC found him in Alexandria in the entourage of the Roman statesman Lucullus. There he was scandalized by Philo's 'Roman books' (see Philo of Larissa §2) and wrote his own version of the story. He may have stayed in Lucullus' company for some years. By 79 BC he was back in Athens, where Cicero (§1, 3) listened to his lectures for six months. A few years later he returned to the East, again in the company of Lucullus. He died, in Mesopotamia, in 68 BC.

Antiochus is referred to as head of a school, perhaps the Academy or perhaps a school of his own. Academic scholarch or not, he was a popular and an influential figure. He had a gentle and attractive personality, and a beguiling eloquence. He charmed Cicero, whom he did not convince, but who described him as 'the most accomplished and the most acute of all the philosophers I have known' (*Academics* II 113). He charmed - and convinced - the scholar Varro. He was the house-philosopher of Lucullus. He also had other friends among the eminent of Rome.

He wrote copiously: we know of certain sceptical books (written before 95 BC); of the *Sosus*, written against Philo in 86 BC; of a work on the agreement between the Stoics and the Peripatetics, dating from *c.*80 BC; of a late essay *On the Gods*; and of a work on epistemology called *Canonics*, the date of which is uncertain. All of these writings are lost; but the speech which Cicero attributes to Lucullus in *Academics* II is described by him as 'Antiochean', and the account of Academico-Peripatetic ethics in Cicero's *On Ends* is said to represent the views of Antiochus and of the Peripatetic Staseas. From these two Ciceronian works we can thus gain an idea of some of the general lines of Antiochus' thought. But Cicero is not translating, or even paraphrasing, Antiochus. Some scholars have supposed that Antiochus ghosted many or most of Cicero's philosophical works, and that he lies behind various other later philosophical texts. There is nothing to be said for these suppositions.

2 Thought

When, in 87 BC, Philo gave up scepticism and the Stoic definition of knowledge (see Philo of larissa §2; Stoicism §12), Antiochus was upset. Not because Philo had become a dogmatist, but because he had rejected the definition:

> when Philo weakens and destroys this, he does away with any criterion for what is known and unknown; hence it follows that nothing can be apprehended and he foolishly finds himself in the position he least wished.
>
> (Cicero, *Academics* II 18)

The criticism of Philo is weak, but Antiochus' own position emerges with clarity: he is a dogmatist, not a sceptic; and he believes that the Stoic definition of knowledge must be upheld.

Antiochus himself had converted from scepticism to dogmatism some ten years earlier. (Since Cicero did not know why he changed his opinion, we will never uncover his reasons.) In his own view, he removed from the New Academy to the Old. Others seemed to see matters differently:

> Antiochus led the Stoa into the Academy, so that it was actually said of him that he practised Stoic philosophy in the Academy for he tried to show that the Stoic doctrines are found in Plato.

(Sextus Empiricus, *Outlines of Pyrrhonism* I 235)

The accusation of Stoicism is not a Sextan absurdity: Cicero remarks that 'although he called himself an Academic, he was - if you changed a few items - a full-blooded Stoic' (*Academics* II 132). How so? And if so, why did not Antiochus simply convert to Stoicism (as Cicero himself pertinently asked)?

Cicero observes that the Old Academy and Zeno the Stoic 'disagree on one point only and agree wonderfully in everything else' (*Laws* I 53-4) - and he ascribes the observation to Antiochus. According to Antiochus, the Stoics were at one with the Old Academy; and in addition, 'the Stoics agree with Peripatetics in substance and dispute only over terminology' (Cicero, *On the Nature of the Gods* I 16). Thus Stoicism and Peripateticism are essentially forms of Platonism, so that a move to the Old Academy was by that very token an acceptance of Stoic and Peripatetic philosophy.

Granted this conglomerative or syncretistic account of Old Academic philosophy, Antiochus' 'Stoicism' ceases to amaze. But the syncretism itself may seem bizarre. None the less, it was not a crude fantasy: Antiochus knew something about the history of philosophy and based his syncretic claim on putative historical fact; nor did he pretend that Zeno and Aristotle swallowed Platonism whole and agreed perfectly with one another. Moreover, certain features which to us seem essential to Platonism - notably the theory of Forms - had no place in Antiochus' scheme of things; and other early Platonists had made a contribution to the philosophy of the Old Academy - Antiochus 'especially approved of Polemo' (*Cicero, Academics II 131*).

In any event, the syncretism is far from idiotic. One dogmatic school is omitted from it: Epicureanism. Antiochus realized that the cleft in dogmatic philosophy divided the Epicureans from everyone else. On the main questions, Old Academics, Peripatetics and Stoics were indeed in broad agreement: they were for knowledge and against scepticism, for virtue and against voluptuarism, for teleology and against mechanism, they were for continuity and against atomism.

Antiochus' syncretistic philosophy dealt with all three of the traditional 'parts' of the subject (logic, physics and ethics); but he urged that

> there are two things of the greatest importance in philosophy: the judgment of truth and the determination of the good. For no-one can be wise unless he is aware both of the beginning of knowledge and of the end of desire, unless he knows whence he is to begin and whither he is to journey.

(Cicero, *Academics* II 29)

For the beginnings of knowledge, Antiochus sketched a standard empiricist account of the development of science from sense-perception; he defended the Stoic definition of knowledge; and he argued that attacks by sceptics all failed. For 'the end of desire', Antiochus used the 'Carneadean division' (see Carneades §2) and urged that the 'end' or *telos* consists in the primary goods of nature together with virtue. He managed to marry this rudely Stoic thought to the Aristotelian idea that the *telos* is intellectual activity or *theōria*. He was also exercised by the question of whether 'external' goods such as riches and beauty are necessary for happiness (see Eudamonia). He answered with engaging good sense that you can be happy if you are poor and ugly but that to be *very* happy you need to be wealthy and handsome.

3 Influence

Antiochus had some professional pupils, but most of them deserted his doctrines. His syncretism was not welcomed by the Stoa. Indeed, there is no evidence that his new version of the Old Academy was taken seriously by any of his colleagues. (There is nothing in the old theory that it was Antiochus who inaugurated what we now call 'Middle Platonism' (see Platonism, Early and Middle).) But he was not an insignificant figure. He was writing primarily for a lay public - a public of educated and intelligent Romans. For such a public, technical treatises and professional pedantries were inappropriate, and the rough syncretism which Antiochus offered made sound sense. Antiochus succeeded in impressing Cicero, and Cicero was no slouch in philosophy.

JONATHAN BARNES

References and further reading

Barnes, J. (1989) ‘Antiochus of Ascalon’, in M. Griffin and J. Barnes (eds) *Philosophia Togata I*, Oxford: Clarendon Press.(Contains everything one needs to know, and more, about Antiochus.)
Cicero, M.T. (early 45 BC) *Academics*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933.(Varro in book I and Lucullus in book II represent Antiochus’ views.)
Cicero, M.T. (mid 45 BC) *On Ends*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1914.(Book V is largely Antiochean in inspiration.)
Glucker, J. (1978) *Antiochus and the Late Academy (Hypomnemata 56)* Göttingen: Vandenhoeck & Ruprecht. (Includes a detailed discussion of Antiochus and his relation to the Academy.)
Mette, H.-J. (1986/7) ‘Philon von Larisa und Antiochus von Askalon’ (Philo of Larissa and Antiochus of Ascalon), *Lustrum* 28/9: 9-63.(Collects the ancient testimonies on Antiochus.)
Sextus Empiricus (*c.* AD 200) *Outlines of Pyrrhonism*, trans. J. Annas and J. Barnes, *Outlines of Scepticism*, Cambridge: Cambridge University Press, 1994.(Fine translation with introduction and notes.)

Antiphon (late 5th century BC)

Antiphon was a Greek Sophist. His most famous work, On Truth, partially survives in two substantial papyrus fragments, plus a number of purported quotations. It sets up a bold antithesis between the claims of physis *(nature) and* nomos *(law/convention), arguing that it is more advantageous to follow nature when one can do so without detection. The antithesis suggests several important questions about the meaning of 'nature' and its role in ethics, the origin of social laws and their authority and the meaning and value of justice. It is disputed whether he is to be identified with the orator Antiphon of Rhamnus.*

1 Life and works

Nothing is known of Antiphon's life. Doubts over his identity with Antiphon the Orator rest on the apparent tension between the conservative views of the Orator's *Tetralogies*, in which obedience to law is highly praised, and the seemingly radical arguments of *On Truth*. For similar reasons some have doubted whether the same man could have written *On Truth* and the conventional gnomic utterances of *On Concord*. In defence of a single identity it has been urged that: (1) *On Truth* is not as radical as it appears, but simply a plea for legal reform; (2) its doctrines, although radical, are not endorsed by Antiphon himself; and (3) Antiphon changed his mind. Antiphon is also credited - although this too is disputed - with a *Politicus* (Statesman) and *On the Interpretation of Dreams*, but these attributions too are disputed.

2 *On Truth*

In *On Truth* Antiphon argues that 'justice' consists of not transgressing the laws and customs (the *nomoi*) of the state in which you are a citizen. Consequently, to treat justice in the way that is most advantageous to yourself you should respect the *nomoi* in the presence of witnesses, but follow the claims of nature when witnesses are absent. Disobedience to nature will always harm you; disobedience to the *nomoi* will harm you only if you are detected. Furthermore, most *nomoi* are hostile to nature and act as shackles, whereas those things laid down by nature as advantageous promote freedom and pleasure.

Nor can the law provide the protection it promises: it cannot prevent aggression, nor guarantee that aggressors will be found guilty. Laws and customs are also based on false distinctions of class and race: we all possess the same intrinsic needs and faculties. Finally, the *nomoi* that we call 'justice' are internally contradictory: they require an individual not to harm anyone except in retaliation, yet if called upon as witness a person is required to give potentially harmful evidence in court against someone who has never harmed them. Giving such testimony may also prove dangerous: the witness may have made an enemy for life.

3 The claims of nature

What precisely is meant by 'nature'? Internal evidence suggests Antiphon means 'human nature', but the opposition he draws between this and culture prompts the fundamental question whether such a thing as 'raw' unsocialized human nature exists, and how we could know about it even if it did.

Even if we grant that Antiphon's stark distinction between nature and culture is sustainable, why does he rank the claims of nature above those of society? His first argument appeals to the automaticity of nature's reprisals if flouted, as opposed to the uncertain outcome of a flouted law or convention. Antiphon is ascribing to nature the inevitability which Greek thought had traditionally ascribed to the anthropomorphic deities or fate; in this he may have been influenced by the mechanical physical systems of Leucippus and Democritus.

The inevitability of nature's workings would nevertheless be irrelevant without Antiphon's further claim of a systematic hostility between *nomos* and nature, with an individual's advantage tied to following the latter. In Greek 'advantageous' has medical connotations, and apparently Antiphon is primarily arguing that only by following the claims of nature can we safeguard our physical wellbeing, and indeed survival. His examples of social constraints, however, suggest that he is thinking not only in physiological terms but also of desire-satisfaction in general.

What are these constraints on our natural inclinations? Antiphon certainly holds (1) that the *nomoi* discourage us

from attacking a potential aggressor, and encourage us to honour our parents, even if they have maltreated us: in both instances our natural tendency towards self-protection is thwarted. He may, however, also be claiming (2) that society restrains us from aggressive behaviour in general, and from satisfying whatever desires we happen to have. If this is his position (the text is too fragmentary to allow certainty), then underlying his argument is a view of 'raw' human nature as comprised of fierce egoistic drives antithetical to social harmony.

This choice between readings substantially affects the tenability of Antiphon's position. On reading (1) it seems reasonable for him to argue that social conventions which hinder our ability to defend ourselves are injurious to the individual. On reading (2), however, one may question whether social restrictions on aggressive behaviour in general are in fact ultimately to the individual's disadvantage. Clearly such constraints aim to frustrate the immediate desire, but one may argue that in the long run all individuals are better served if certain aggressive and egoistic aspects of human nature are curbed wholesale: what Antiphon regards as fetters could also be regarded as the bonds which keep society together. One might have expected Antiphon's own words to prompt him in this direction. He does after all admit that death is natural yet disadvantageous; why then does he not concede that a considerable number of the *nomoi* are aimed at protecting the individual from premature death? It is true that he nowhere recommends discarding the *nomoi* altogether, but he still does not consider the general undermining effect on security that repeated disregard for the *nomoi* might have. Possibly his argument is intended only to apply to those who can take care of themselves in all eventualities.

Furthermore, if the *nomoi* are generally disadvantageous, how did they come to exist at all? Antiphon gestures towards a social contract theory: there are opaque references to the *nomoi* being 'agreed' and to 'those who made the agreement' (it is unclear whether this is intended to represent a historical event). This supposed agreement is used by Antiphon as a reason for disobeying the *nomoi* - 'agreed' is interpreted as 'artificial' - but it is equally possible to reach the opposite conclusion: the *nomoi* were agreed because they were perceived to be beneficial and this is a good reason for obeying them. Protagoras (§2) maintained precisely such a position, and it may be that *On Truth* is partly directed against him. Again, the question is raised whether Antiphon's argument is intended to apply to everyone or only to those who think they have sufficient resources to ignore the benefits of a social contract.

Antiphon's attack on the *nomoi* continues by claiming that they give unmerited preferential treatment to the high-born and Greeks (and given that most Greeks justified their practice of enslaving other races by appeal to the racial superiority which they claimed, he is probably criticizing the institution of slavery as well). Such criticisms suggest a reason why Antiphon called his work *On Truth*: he wishes to oppose the false ontological divisions of *nomos* to the true divisions of *physis*. In a separate fragment he also deplores the ambiguity of language; his criticisms in the papyrus fragments of the internal contradictions in the notion of 'justice' may be an example of this (Caizzi 1989).

4 Practical implications and influence

What, if anything, is Antiphon proposing that we actually do? On one view he is not recommending any course of action: he is not speaking *in propria persona* but simply setting up a debate. That would still leave the question whether the words in the text, even if not endorsed by Antiphon, are saying that we *ought* to follow nature when witnesses are absent; and if they are, whether this is an example of the 'naturalistic fallacy', the (allegedly) illegitimate inference of a prescription for action from a simple description of how things are (see Naturalism in ethics §3). The most plausible interpretation is that the text commends following nature as advantageous, but does not set it up as a moral imperative; it certainly does not propose an alternative description of 'justice'.

Another possibility is that Antiphon's real concern is with legal reform: he wishes to remodel the *nomoi* to harmonize with nature. The key issue here is again Antiphon's underlying view of humanity: if he thinks our natural inclinations tend towards unprovoked aggression rather than mere self-defence, then it is unclear how the *nomoi* could be remodelled accordingly yet still retain their social character. Nevertheless, his criticisms unquestionably force us to reconsider the origin and purpose of our own laws and customs and the nature and function of legal punishment.

Perhaps Antiphon's most important contribution to philosophy is his insistence that human physiology and psychology be included in ethical and political thought: our basic needs and desires cannot be ignored. The

tensions he perceived between these natural desires and the claims of society were taken up in Plato's *Republic*, he was thus the first to lay down what Plato saw as the fundamental ethical challenge.

See also: Callicles; Physis and nomos; Sophists

ANGELA HOBBS

References and further reading

Antiphon (5th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 2, 334-70.(The standard collection of the ancient sources; includes Greek texts with translations in German.)

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge & Kegan Paul. (Chapter 23 includes an English translation and a crisp appraisal.)

Caizzi, F.D. (1989) 'Antipho' in *Corpus dei papiri filosofici greci e latini (Corpus of Greek and Latin Philosophical Papyri)*, Part I , Florence: Olschki, 176-236.(Contains the most complete reconstruction in Greek of the papyrus fragments of *On Truth*, together with a sound and scholarly commentary in Italian and a very detailed bibliography.)

Furley, D. (1981) 'Antiphon's Case against Justice', in G.B. Kerferd (ed.) *The Sophists and their Legacy*, Wiesbaden: Steiner, 81-91; repr. in D. Furley, *Cosmic Problems*, Cambridge: Cambridge University Press, 1989.(Lucid and incisive; includes translation in English of some passages.)

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, vol. 3, Cambridge: Cambridge University Press; part of vol. 3 repr. as *The Sophists*, Cambridge: Cambridge University Press, 1971.(Interesting and scholarly discussions of Antiphon; see especially pages 107-13 and 285-94.)

Kerferd, G.B. (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press.(The best introductory handbook to the Sophists in general.)

Sprague, R.K. (ed.) (1992) *The Older Sophists*, Columbia, SC: University of South Carolina Press, 106-240.(Full English translation of the fragments and testimonia from Diels and Kranz (1952) plus the speeches of Antiphon of Rhamnus.)

Anti-positivist thought in Latin America

Anti-positivist philosophy arose in Latin America at the turn of the twentieth century in response to the dominance of closed positivistic systems of historical development in the climate of intellectual opinion. Argentina, Mexico and Uruguay were all centres of anti-positivist theorizing. Philosophers such as Mexicans Antonio Caso and José Vasconcelos, the Argentinian Alejandro Korn and the Uruguayan Carlos Vaz Ferreira attacked Auguste Comte's positivism, as well as deterministic forms of scientific Marxism and Herbert Spencer's social Darwinism for their denials of creative freedom and spiritual values. Latin American anti-positivism is characterized as a form of modernism although it incorporates elements of traditionalism. It is self-consciously critical of the limitations of modern progressivism and willing to supplement the modern paradigm with premodern discourses. Anti-positivist philosophy is also firmly committed to the modern embrace of process over fixed form.

Latin American anti-positivism is founded in a comprehensive interpretation of experience that embraces phenomena such as creative freedom, tentative and experimental thinking, imaginative coordination and charitable love. These aspects are excluded from the purview of what positivists allow to be objects of scientific knowledge. Anti-positivists interpret experience as a bi-polar struggle in which the free side of life battles to prevail over the forces of necessity, system, abstraction and egoism.

South American anti-positivists concentrated on issues of knowledge and the structure of thought and experience, whereas Mexican anti-positivists, who were products of a formal education modelled on Comte's prescriptions, undertook a more total revolt. This revolt had metaphysical, moral and political dimensions.

For the most part anti-positivists did not fully escape the doctrines they criticized. They took from these doctrines descriptions of unredeemed, degraded and mechanized life which they opposed to redemptive practices of struggle.

1 Motivation

At the turn of the twentieth century, positivism was the dominant tendency of thought, philosophy and ideology in Latin America. In the strongest and most specific sense, the Mexican dictatorship of Porfirio Díaz attempted to legitimate itself as the applier of Auguste Comte's positivistic philosophy to national development, which included basing the educational system on a Comtian curriculum (see Comte, A.). In a more general sense, evolutionary philosophies with naturalistic foundations, such as Comte's positivism, *fin-de-siècle* scientific Marxism and Herbert Spencer's social Darwinism, shaped the major component of educated opinion throughout Latin America, especially in Argentina and Uruguay (see Spencer, H.; Marx, K.). Anti-positivism is the intellectual response to the positivist milieu of the generation that came of age in the first two decades of the twentieth century.

Latin American anti-positivism is determined by a sense of intellectual suffocation from positivistic systems. Late nineteenth-century positivism was naturalistic, deterministic, evolutionary, rationalist and scientific. The generation that came of age in that climate of opinion rebelled against confinement in prisons of thought which justified and even glorified inadequate social conditions, reduced human beings to egoists and had already mapped out the future. Drawing upon the emerging vitalistic philosophies in Europe, such as Friedrich Nietzsche's life philosophy (see Nietzsche, F.) and Henri-Louis Bergson's vitalism (see Bergson, H.-L.; Vitalism) both of which opposed monistic nineteenth-century systems, the Latin American anti-positivists explored and advanced dimensions of human existence that had been neglected or silenced by positivism.

Latin American anti-positivism is a proclamation of the liberty to break out of the confines of nineteenth-century opinion and be open to a wealth of new experiences. As a variant of modernism, it is self-conscious of its own modernity and affirming of it, although, perhaps critical of it.

2 Major figures

The major Latin American anti-positivist philosophers include José Vasconcelos, Antonio Caso, and two generations later Leopoldo Zea in Mexico, Alejandro Korn in Argentina and Carlos Vaz Ferreira in Uruguay. The South Americans carried on their anti-positivist projects in the areas of epistemology and philosophy of experience, whereas the Mexicans took a more complete approach that included metaphysics, ethics, depth

psychology, sociology of knowledge and social philosophy, as well as reflections on knowledge and experience. Also, the South Americans based their modernism on an extension of the modern fixation on the category of development to an affirmation of the pure process of creative freedom. Whereas the Mexicans, while also insisting on creative freedom, brought to bear pre-modern thought (with its normative content) on their ethical and sociopolitical positions. The South Americans could be considered to be progressive modernists, kindred to William James in the USA, who reveal the intuition of continuous change which grounds and simultaneously deconstructs any evolutionary system of development. The Mexicans, in contrast, are complex figures who resemble Anglo-American traditional modernists, such as T.S. Eliot, Irving Babbitt and George Grant, in their return to pre-modern discourses, despite the utopian humanism rather than conservatism evident in their social philosophy.

As epistemology and philosophy of experience, Latin American anti-positivism has a basic structure which distinguishes it from other vitalist and existentialist tendencies (see Existentialism; Epistemology, History of). The anti-positivists share a commitment to an expanded empiricism which comprehends all aspects of human experience, including those that do not seem amenable to the methods of natural science (see Empiricism). They also subscribe to a dualistic philosophy of experience in which they set up irreducible polarities in human experience and then describe and prescribe a continuous struggle to overcome one side of those dualisms.

The substance of the dualisms developed by Latin American philosophers pits creative liberty, aesthetic synthesis, disinterested commitment to ideals, charity and the flux of fermentative thought, against mechanistic determinism, material entropy, utilitarianism and fixed and absolute systems (see Utilitarianism). Underlying this content is a deeper philosophical structure which involves uncompromising commitment to describe clearly the highest human ideals joined with the equally intense commitment to describe realistically the limitations, failings, frustrations and coercions of human existence. It is within the tension set up between uncompromising idealism and intense realism that human beings struggle to realize their provisional victories over physical and spiritual death.

Korn constructed his philosophy of creative liberty from a Kantian background, arguing that Kant's dualisms can be made empirical by interpreting the split between subject and object as the experienced opposition between liberty and necessity (see Kant, I). For Korn (1963), the field of experience is a dynamic process in which subject and object are reciprocal functions. Necessity is objective in that it comprehends the succession of facts linked by the principle of physical causality into mathematical laws and excludes personal will. Liberty is definitive of subjects, constituting them as a process of valuation presupposed by the plurality of ideals acknowledged by Korn, such as wellbeing, happiness, love, power, justice, sanctity, goodness, truth and beauty. According to Korn, epistemology and axiology are inseparable: as a process of valuation, the founding act of the subject is the valuation that there be a valuing process, that is, the affirmation of creative liberty raised against economic coercion, the more intimate coercion of the passions and the tyranny of necessitarian doctrines such as nineteenth-century positivism (see Axiology).

Vaz Ferreira, more strict an empiricist and more sceptical than Korn, based his dualistic philosophy of experience on what he called the greatest revolution in intellectual history which was taking place in his own time - the liberation of thought from words. Vaz Ferreira (1961) distinguished between thought which draws from its datum a system meant to be applied to every case, and thought which reserves and elaborates its datum so that it becomes available to enlighten deliberation on concrete problems. The latter, known as fermentative thinking, operates according to a 'living logic' that is personal and intransferable and reaches its fruition in 'hyper-logical good sense' (1961: 197). This is a process of balancing opposed arguments and facilitating the interplay of diverse ideas so as to prevent any single idea from dominating thought and protect against systematization. Like Korn, Vaz Ferreira prescribes a philosophy of struggle, a life of 'superquixotism' (1961: 274) in which the individual, on the one hand, renounces adherence to any single ideal or value theory as that would bind life to a system, and on the other, embraces an effort to appreciate a multitude of diverse and sometimes contradictory ideals. To Vaz Ferreira, struggle is defined with particular clarity because he challenges individuals to engage the mechanistic aspects of their own thought.

3 The Ateneo

Mexican anti-positivism was far more intricate, historically-based and comprehensive than its South American counterpart. The initiators of anti-positivist philosophy in Mexico were part of a larger group of rebels

representing different branches of the humanities. They were the privileged products of the positivistic curriculum which had been set up during the long rule of Díaz. In the first decade of the twentieth century, these young dissidents created a counter-educational institution of their own which they called the Ateneo de la Juventud (Athenaeum of Youth). The institution was partly devoted to fostering extensive study of the ancient Greek classics in order to liberate the mind from the confines of positivistic doctrine. Positivism in Mexico was a totalistic ideology that called for economic modernization, scientific education and administration of the country by technocrats, or *científicos*, who claimed they were above politics. Given the anti-metaphysical bias of positivism, the initial project of the Ateneists was the rehabilitation of metaphysics. However, they were not traditionalists or reactionaries, but modernists, who questioned the deprivation of modern values while maintaining the primacy of action and process over contemplation and form.

The aims of the Ateneists were not confined to educational reform, but touched on political morality. Caso, the only member of the Ateneo who supported Díaz, was still able to summarize his generation's criticism of the positivistic polity for making economic development the supreme goal of the society, believing that wealth was the sole basis of strong government and suppressing democracy in the name of national welfare. This is not to say that the members of the Ateneo were primarily concerned with political issues, or that they were direct forebears of the Mexican Revolution of 1910, but rather that their cultural politics, like that of the French philosophers of the eighteenth century, helped to undermine the legitimacy of established authority.

The two leading philosophers of the Ateneo, Caso and Vasconcelos, went on to write bodies of work that touched on most of the major branches of philosophy. The basic structure of their thought was the same as that of their South American counterparts. It was a comprehensive empiricism constituted by a polarized struggle. For Vasconcelos (1961), the struggle was between the ascending and erotic tendency of conscious experience as a process of coordinating heterogeneous contents into moving totalities, or aesthetic wholes, and the counter-tendency towards entropy, the dispersion of things into flat homogeneity. Like Vaz Ferreira, Vasconcelos advanced an informal logic to describe the coordinative process he identified. Vasconcelos (1961) opposed his 'organic logic', based on the categories of melody and counterpoint, to conceptual logic, which analyses things as components and then reconstitutes them through systems of abstraction. For Vasconcelos, life should be a ceaseless struggle to overcome entropy in favour of coordination.

Caso's (1943) dualism is moral as it puts into play opposing attitudes to existence. At one pole of experience is 'existence as economy', the principle of the positivistic polity which prescribes maximum advantage with minimum effort. At the other pole is 'existence as disinterest', or aesthetic contemplation and 'existence as charity', or the pure act of self-giving love, which transcend the economy of selfishness by prescribing, respectively, maximum effort with indifference to advantage and maximum effort with minimum advantage. For Caso (1943) life is a heroic struggle to overcome and reverse the normal selfish tendencies of the organism.

In their revolt against positivism both Vasconcelos and Caso retained its stamp: Vasconcelos by using the physical science of his time as a metaphor for the structure of experience and Caso by accepting as true the positivistic model of the individual as a selfish competitor. Their prime concern was not restrictive speculative systems parading as scientific, but forces in human nature that impede the fulfilment of spiritual value, which they understood as non-natural, not extra-experiential. In their later work, Vasconcelos and Caso, having witnessed the rise of Communist and fascist regimes in Europe, like many intellectuals throughout the Western world, turned to the Christian faith as a bulwark to protect the values they had sought to redeem from positivistic neglect and disparagement. Indeed, they recognized in twentieth-century totalitarianism the same phenomenon against which they had battled in their youth (see Totalitarianism).

4 Ramos, Zea

In the generation following the Ateneo, one major philosopher rose up as an organic critic of the anti-positivists. Samuel Ramos, a former student of Caso, broke with his master over the issue of basing philosophical reflection on intuition, which all the anti-positivists had done in some way, following Bergson. According to Ramos anti-positivism had erred from attention both by degrading reason and diverting thought to private states, to coming to terms with and improving objective conditions. Although he was a realist opposed to according primacy to imagination and introspection, Ramos's reaction did not represent a return to positivism. Rather, it represented an exploration of self-critical diagnosis of national character which culminated in his classic ***El perfil del hombre y***

la cultura en México (Profile of Man and Culture in Mexico) (1934). The book had many imitators in the decades following its publication. Later in his career, Ramos grew closer to anti-positivism, drawing on Max Scheler's anti-naturalistic axiology to ground a new humanism that would save civilization from totalitarian threats (see Scheler, M.F. §5).

The cycle of Mexican anti-positivism reached completion in the next generation in Leopoldo Zea's *El positivismo en México (Positivism in Mexico)* (1968), which applied Karl Mannheim's sociology of knowledge to positivism as an ideology. Zea argued that behind the facade of scientific administration were the particular interests of a small segment of the middle class who reaped their rewards by representing foreign economic interests. For Zea, the decline of positivism in Mexico was due to the increasing exploitation of the rest of the middle class by the positivistic elite and by the inability of positivism to preserve its claims to political neutrality against the attacks of liberals and Catholics. This led to the revolution of 1910. By the time the Ateneo arose, positivistic ideology and educational practice had become so diluted that the governing elite no longer had any clear legitimating principle.

From the viewpoint of sociology of knowledge, the anti-positivist generation in Mexico represented the middle class that had been marginalized by the positivistic elite. Their appeal to spiritual values sublimated the democratic demands that would be expressed in the revolution. Their subsequent fusion of utopian, often religious humanism and pre-modern discourses reflected the equivocal role of the post-revolutionary middle class as both essential to and alienated from the corporative structures that developed to integrate workers and peasants into the post-revolutionary regime.

See also: Positivist thought in Latin America; Tradition and traditionalism

MICHAEL A. WEINSTEIN

References and further reading

Caso, A. (1943) *La existencia como economía, como desinterés, y como caridad (Existence as Economy, as Disinterest and as Charity)*, Mexico: Ediciones de la Secretaría de Educación Pública.(A founding work of Mexican anti-positivism that counterposes a moral ontology of aesthetic contemplation and charity to the positivistic celebration of self-interest.)

Korn, A. (1963) *La libertad creadora (Creative Liberty)*, Buenos Aires: Editorial Claridad.(An anti-postivist defence of the creative liberty of the human subject against doctrines of metaphysical necessity.)

Ramos, S. (1934) *El perfil del hombre y la cultura en México*, Buenos Aires: Espasa-Calpe, 2nd edn, 1951; trans. P.G. Earle, *Profile of Man and Culture*, Austin, TX: University of Texas Press, 1962.(A critique of Mexican social character employing the perspective of Alfred Adler's psychology.)

Stabb, M.S. (1967) *Latin America in Quest of Identity*, Chapel Hill, NC: University of North Carolina Press.(A general background of the historical conditions for the development of anti-positivism in Latin America.)

Vasconcelos, J. (1961) *Obras completas (Complete Works)*, Mexico: Libreros Mexicanos Unidos. (Comprehensive empiricism constituted by a polarized struggle.)

Vaz Ferreira, C. (1961) *Estudios filosóficos (Philosophical Studies)*, Buenos Aires: Aguilar.(A description of the informal processes of thought that underlie formal logical systematizations.)

Weinstein, M.A. (1976) *The Polarity of Mexican Thought: Instrumentalism and Finalism*, University Park, PA: Pennsylvania State University Press.(An account of the dimensions of anti-positivist philosophy in Mexico through its successive generations.)

Zea, L. (1968) *El positivismo en México*, Mexico: Fondo de Cultura Económica; trans. J. Schulte, *Positivism in Mexico*, Austin, TX: University of Texas Press, 1974.(An account of the history of positivism in Mexico using the perspective of sociology of knowledge.)

Antirealism in the philosophy of mathematics

Realism in the philosophy of mathematics is the position that takes mathematics at face value. According to realists, mathematics is the science of mathematical objects (numbers, sets, lines and so on); mathematicians, to use the old metaphor, are discoverers, not inventors. Moreover, just as there may be truths about physical reality which we can never know, so too, realists say, there may be truths about mathematical reality which we can never know.

It is this claim in particular which antirealists find unacceptable. Equating what can be known in mathematics with what can be proved, they insist that only what can be proved is true. (Only what can *be proved: different accounts of what this 'can' means, facing different difficulties, generate different positions.) This leads antirealists to recoil not only from realism but also from the practice of mathematicians themselves. For the orthodox assumption that every mathematical statement is either true or false would be invalidated, on the antirealist view, by a statement that was neither provable nor disprovable. Not that antirealists themselves can see it in these terms. For if a statement were neither provable nor disprovable, that would itself be an unprovable truth about mathematical reality. Antirealists must learn how to be circumspect even in defence of their own circumspection.*

1 Realism and its rivals

Consider the following view. Any mathematical statement such as ' $7+5=12$' or ' $7+5=13$' or 'Every even number greater than 2 is the sum of two primes' is straightforwardly true or false. If mathematical reality is as the statement says it is, then the statement is true. Otherwise, it is false. How the statement connects with mathematical reality is indicated by its grammatical structure. To each singular term there corresponds an object, to each sortal noun a kind of object, to each quantifier a range of objects, and so on. Which grammatical category an expression belongs to depends solely on how it functions. Thus '7', which functions as a singular term, *is* a singular term: it stands for a particular mathematical object. 'Prime', which functions as a sortal noun, *is* a sortal noun: it picks out a particular kind of mathematical object. 'Every even number', which functions as a quantifier, *is* a quantifier: it ranges over a particular domain of mathematical objects. These objects, which exist outside space and time, are mind-independent. How they stand in relation to one another is not determined by how we think they stand in relation to one another. In some cases how they stand in relation to one another is completely unknown. It may even be unknowable.

This is, in its most extreme form, 'realism' (see Realism in the philosophy of mathematics). It is also known as 'Platonism' (see Plato §10). But if this is realism, then any number of things might reasonably be meant by 'antirealism'. Each part of this view can be challenged, and has been. Thus there are those who have claimed:

(1) That no mathematical statement is strictly true or false: mathematics consists in the manipulation of meaningless symbols in accordance with certain formal rules.
(2) That this is true of certain auxiliary parts of mathematics: not *all* mathematical statements are strictly true or false (see Hilbert's programme and formalism).
(3) That all mathematical statements are strictly true or false, but not straightforwardly so: ' $7+5=12$', for example, is like 'Hamlet was the prince of Denmark' - a fictional truth that is literally false.
(4) That the apparent grammatical structure of mathematical statements is misleading: '7' is no more a singular term than 'John's sake' in 'She did it for John's sake'.
(5) That the apparent grammatical structure is not misleading, but does not determine in the standard way how mathematical statements connect with mathematical reality: '7' does not stand for a particular object, but rather indicates part of a structure which any objects whatsoever could instantiate.
(6) That there are particular mathematical objects in just the way that the grammatical structure suggests, but they are mind-dependent creations of ours: how they stand in relation to one another is not totally independent of how we think they stand in relation to one another (see Intuitionism; Constructivism in mathematics). This allows for the possibility that there are certain mathematical statements, rather like 'Father Christmas is left-handed', whose truth or falsity has never been settled (see Cantor's theorem; Continuum hypothesis; Existence).
(7) Finally, that how things are in mathematical reality, though it is sometimes unknown, is never unknowable. This brings us full circle, for, as we shall see, this last challenge in turn becomes a challenge to the idea that every mathematical statement is true or false.

2 The antirealist challenge

The term ‘antirealism’ is most frequently reserved for the last of the above challenges, and that is what this entry will focus on. Not that only one view about the nature of mathematical truth is in question. On the contrary, as we shall see, there are radically opposed views which issue in this particular challenge: that is, there are radically opposed reasons for thinking that mathematical truth can never outstrip knowability.

Knowability in a mathematical context can be regarded as provability. The idea of a proof must then be suitably construed. For instance, a self-evident axiom must be allowed to count as a one-line proof of itself: otherwise it would be unprovable, and hence, *ex hypothesi*, unknowable. On the other hand, clearly not every mathematical statement can count as a one-line proof of itself. Let us assume that a satisfactory account of proof can be given. The question now is: can there be unprovable mathematical truths?

One reason for denying that there can be was mentioned in the previous section. If mathematical objects are mind-dependent creations of ours, then there is nothing to them beyond what is invested in the various principles of proof that constitute the creation. The best-known exponent of such a view was L.E.J. Brouwer (1912). He held further that the raw material for the creation is provided by something essentially private and incommunicable: our experience of time. This is the position known as ‘intuitionism’.

It is a consequence of intuitionism that the meaning of a mathematical statement is itself something essentially private and incommunicable. A diametrically opposed outlook is canvassed in the writings of Michael Dummett. According to this outlook, the meaning of a mathematical statement is something essentially public and communicable. So there can be nothing more to that meaning, ultimately, than how the statement and its constituents are used in actual mathematical practice: nothing else of potential relevance can be communicated. But this too leads to a denial of the possibility of unprovable mathematical truths. For if a mathematical statement were unprovably true, then its meaning would be something that went beyond how the statement and its constituents were used, or indeed could be used.

This second brand of antirealism, which I shall refer to as ‘Dummettian’ antirealism (though Dummett has never actually endorsed it; see Dummett 1973; Dummett, M.A.E. §3), depends on broadly Wittgensteinian considerations about the nature of meaning (see Wittgenstein, L. §12). Such considerations could just as well be applied in a non-mathematical context. Nevertheless, there are reasons for thinking that they are peculiarly suited to mathematics. For whether or not mathematical objects are mind-dependent creations of ours, there is surely something in the idea that they are nothing apart from the techniques and proof-procedures of mathematical practice. They lack the ‘solidity’ of physical objects. So even if there is some prospect that we can capture such ‘solidity’ in our use of language, and thereby express inaccessible truths about physical objects, there is no prospect that we can do anything similar in the case of mathematical objects.

A third brand of antirealism takes off from the observation that mathematics makes essential use of infinite domains. On one very attractive view, an infinite domain is unlike a finite domain in that its members cannot be given ‘all at once’. Somewhat less metaphorically, they cannot all have a separate involvement in any one state of affairs, or make a separate contribution to any one truth. For a statement about them all to be true, there has to be a single principle which determines its truth. That is, there has to be a proof.

All these brands of antirealism have a revisionary impact on standard mathematical practice. Consider a statement that has never been proved or disproved. At the time of writing, Goldbach’s conjecture - that every even number greater than 2 is the sum of two primes - is an example. It is part of standard mathematical practice to assume that the statement is either true or false. (This is an application of the law of the excluded middle.) Antirealists cannot share this assumption. To make the assumption is to register indifference about whether the statement can be proved or disproved. It is to allow that the statement may be true without proof, a kind of ‘infinite coincidence’. This is precisely what antirealists cannot allow.

3 What kind of possibility does antirealism involve?

The antirealist view is that only what can be proved is true. A much more radical view (not without adherents) is that only what *has been* proved is true. By holding back from this more radical view, antirealists inevitably draw

attention to the ‘can’. But what sense of ‘can’ is intended? Certainly, it is what can be proved *by us* that counts, and not, for example, by an infinite being existing outside space and time. But that does not answer the question.

Basically, antirealists need to decide whether they mean ‘can in principle’ or ‘can in practice’. Not that a simple one-or-other verdict will do. Both ‘can in principle’ and ‘can in practice’ require further elucidation. (How and where do scientific limitations figure? Or technological limitations? Or medical limitations?) But the broad distinction remains. And it is clear that if antirealists mean ‘can in practice’, then the revisionary impact on standard mathematics is going to be enormous. There are some numbers which are so big that we cannot, in practice, tell whether or not they are prime. So given such a number, we shall not be entitled to assume that the number is either prime or not. More generally, since mathematics cannot be extended to the unmanageably big, we shall not be entitled to assume that mathematical reality is infinite.

The resultant version of antirealism is known as ‘strict finitism’. Many antirealists think that strict finitism is incoherent. Call such antirealists ‘moderates’. For them, the ‘can’ must mean ‘can in principle’. But they need to explain why, given that their arguments would break down if the ‘can’ did not mean that, the arguments are not to be distrusted anyway. Dummettian moderates in particular are likely to experience some difficulty here. Others, whose very starting point is that mathematical reality is infinite, experience no difficulty at all.

A quite different problem for moderate antirealists is to explain why their own version of antirealism is not just a terminological variant of realism. For cannot any statement be proved or disproved *in principle*? Reconsider Goldbach’s conjecture. What is wrong with the following recipe for either proving it or disproving it? ‘Check successive even numbers greater than 2. In each case, ascertain whether or not it is the sum of two primes. Keep going until either every such number has been checked or a counterexample to the conjecture has been found.’ It is of no avail to reply simply that this process might never end. Something needs to be said to forestall the objection that if we spend half a minute checking 4, a quarter of a minute checking 6, an eighth of a minute checking 8, and so on, then the process will end in a minute - at most. Of course, we cannot in practice do this. But that is beside the point.

It is by no means clear that moderate antirealists have nothing to say here. But they may have nothing to say which is not question-begging. For instance, they may be forced to deploy a conception of what is possible in principle which itself incorporates an antirealist understanding of what it is to do infinitely many different things. If so, that is not an objection to moderate antirealism. But it does raise the spectre of realists and moderate antirealists operating with two quite different, incommensurable, individually coherent conceptions of what is possible in principle, together with two correspondingly different conceptions of how to do mathematics. Ultimately, perhaps, the realist is prepared to make certain assumptions in mathematics which the moderate antirealist is not prepared to make, and, to the extent that they are incapable of finding any independent leverage to settle their differences, they are talking past each other.

4 Implications for mathematical practice

The sheer fact that realists are willing to endorse standard mathematical practice might be thought to count in their favour. For, arguably, reversing Marx’s dictum, it is not the business of philosophers of mathematics to *change* mathematics, only to interpret it. (Wittgenstein famously wrote that ‘philosophy leaves mathematics as it is’.) Here, as before, there is a particular difficulty for Dummettian antirealists. For they are only too keen to cast actual mathematical practice in the role of a datum. Dummett’s own reaction to this difficulty is to argue that, even if actual mathematical practice is a datum, it is not an indissoluble datum: nothing in the view he canvasses precludes criticizing some parts of that practice for failing to harmonize with other parts.

Suppose there *is* room for a non-conservative philosophy of mathematics. Antirealists face a further difficulty. They have no satisfactory answer to the question ‘When exactly does standard mathematical practice lead us astray?’. They cannot answer, ‘Whenever a statement is neither provable nor disprovable’. For if a statement were neither provable nor disprovable, that would itself be an unprovable truth about mathematical reality. (To prove that a statement was not provable would already be to disprove it.) Likewise, if, say, there were no proof of Goldbach’s conjecture, nor, as it happened, a counterexample, that would itself be an ‘infinite coincidence’ of the very kind that antirealists disavow. Antirealists have no option, then, but to adopt a kind of stoic silence. Whenever others make assumptions in mathematics which they are not themselves prepared to make, they must

withhold assent, but they have no satisfactory way of saying what is holding them back. As far as anything they can say is concerned, their restraint may be a result, not of nonconformity, but of sheer reticence.

Antirealism remains of the utmost importance, despite this very real difficulty. Indeed, part of its importance may be in drawing the difficulty to our attention. For if antirealists are right, then it may be that their knowledge of correct mathematical practice is a prime example of a fundamental and neglected philosophical category: knowledge that is essentially ineffable.

See also: Logical and mathematical terms, glossary of

A.W. MOORE

References and further reading

Benacerraf, P. and Putnam, H. (eds) (1964) Introduction to *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 1-37.(Useful survey of the basic issues in the philosophy of mathematics, touching helpfully on both realism and antirealism.)

Bernays, P. (1935) 'Sur le Platonisme dans les mathématiques', *L'enseignement mathématique*, 1st series, 34: 52-69; trans. C.D. Parsons, 'On Platonism in Mathematics', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 258-71.(Classic survey of Platonism/realism and standard antirealist reactions.)

Brouwer, L.E.J. (1912) *Intuitionisme et Formalisme*, Groningen: Noordhoff; trans. A. Dresden, 'Intuitionism and Formalism', *Bulletin of the American Mathematical Society* 20: 81-96, 1913; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 77-89. (Classic statement of some of the fundamental tenets of intuitionism.)

Dummett, M.A.E. (1973) 'The Philosophical Basis of Intuitionist Logic', in H.E. Rose and J.C. Shepherdson (eds) *Proceedings of the Logic Colloquium, Bristol, July 1973*, Amsterdam: North Holland, 1975, 5- 40; repr. in *Truth and Other Enigmas*, London: Duckworth, 1978, 215-47; and in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 97-129.(Expounds the argument sketched in §2 for 'Dummettian' antirealism.)

Dummett, M.A.E. (1977) *Elements of Intuitionism*, Oxford: Clarendon Press, 1990.(A very thorough account of intuitionistic revisions to mathematics, together with a full discussion of antirealist reasons for adopting them.)

George, A. (1993) 'How Not to Refute Realism', *Journal of Philosophy* 90 (2): 53-72.(Identifies question begging in certain antirealist arguments, and lends support to the suggestion of §3 that realism and antirealism may be two equally acceptable but incommensurable alternatives.)

Lear, J. (1977) 'Sets and Semantics', *Journal of Philosophy* 74 (2): 86-102.(Argues that even realists have reasons for adopting intuitionistic revisions to mathematics.)

McDowell, J. (1989) 'Mathematical Platonism and Dummettian Anti-Realism', *Dialectica* 43 (1/2): 173-92. (Endorses the suggestion of §2 that the basic considerations in favour of 'Dummettian' antirealism apply only in a mathematical context, where the objects of investigation lack 'solidity'.)

Prawitz, D. (1977) 'Meaning and Proofs: On the Conflict Between Classical and Intuitionistic Logic', *Theoria* 43 (1): 2-40.(Good general account of the argument expounded in Dummett (1973) and sketched in §2).

Putnam, H. (1980) 'Models and Reality', *Journal of Symbolic Logic* 45 (3): 464-82; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 421-44.(Uses the Löwenheim-Skolem theorem to attack extreme realism.)

Rotman, B. (1993) *Ad Infinitum… The Ghost in Turing's Machine*, Stanford, CA: Stanford University Press. (Defends a version of strict finitism.)

Wittgenstein, L.J.J. (1956) *Remarks on the Foundations of Mathematics*, ed. G.H. von Wright, R. Rhees and G.E.M. Anscombe, trans. G.E.M. Anscombe, Oxford: Blackwell, 2nd edn, 1967; 3rd edn, 1978.(*Locus classicus* of Wittgenstein's later philosophy of mathematics, with its own distinctive brand of antirealism.)

Wright, C. (1980) *Wittgenstein on the Foundations of Mathematics*, London: Duckworth.(Very thorough discussion of antirealism in Wittgenstein's later philosophy of mathematics.)

Wright, C. (1982) 'Strict Finitism', *Synthese* 51: 203-82; repr. in *Realism, Meaning and Truth*, Oxford: Blackwell, 1986; and in A.W. Moore (ed.) *Infinity*, Aldershot: Dartmouth, 1993, 251-330.(Sympathetic discussion of strict finitism and its revisionary impact on mathematics.)

Anti-Semitism

Anti-Semitism is a form of racism which sees Jews as a dangerous and despicable group in society. It has solid philosophical sources in the work of German Idealism which emphasized the distinctiveness of Judaism and how it has been superseded by Christianity. Both Kant and Hegel made a sharp distinction between Judaism and what they regarded as more rational religions, and they questioned the capability of the Jewish people for playing an integral role in the state. Sartre used the notion of anti-Semitism to show how a sense of self-identity is created by the attitudes of others towards the individual and the group. That is, what makes Jews Jews is the fact that there is anti-Semitism, and there is nothing that Jews can do about anti-Semitism. Anti-Semitism is a problem for the anti-Semites themselves; anti-Semitism, by Sartre's account, is in fact an attempted solution to the difficulties of taking free and authentic decisions. Anti-Semitism has played an important role in Jews' self-definition, in attitudes to the State of Israel and to the religion of Judaism itself.

1 Kant

Kant presents generally critical views of Judaism. He argues that Judaism is primarily a religion of statutes, a system of positive law, and that it is also largely based on revelation. These characteristics lead him to deny its moral character, since the will of a supreme authority cannot be properly considered a source of morality. In Kantian philosophy, the essence of morality is the ability of subjects to place themselves under entirely general and rational laws, so a faith which is limited to a particular people and which is based on a particular form of revelation, does not fit his notion of morality. Kant even criticizes the laws themselves, regarding them as crude imperatives which demand a certain standard of conduct but do not insist on appropriate attitudes. This leads Kant even to deny that Judaism is a stage of religious development in the direction of Christianity, since Judaism has, by his account, so little genuinely religious content. Judaism makes worship of God primary and virtue subsidiary, and so is really just a form of belief limited to a particular group of people.

The emphasis Kant finds in Judaism on obedience to God leads it, he argues, to a form of legalism which he regards as far from a rational faith. Revelation and messianism are problematic from a rational point of view, since they relate to a particular relationship with a particular group of people, and can hardly be seen in universal terms (see Revelation). Also, the laws of the religion seem to be both pedantic and rigid, based more on the necessity to achieve isolation rather than on any quest to establish general forms of behaviour applicable to any human being. The empty legalism of Judaism plays no part in improving the human moral disposition, Kant argues, and so cannot be considered as a genuine expression of rational religion.

2 Hegel

Hegel's views on religions and their interrelationships are far more developed than those of Kant, but they are also quite negative with respect to Judaism (see Hegel, G.W.F.). The trouble with Judaism, says Hegel, is that it is a religion that emphasizes fear, a religion that stresses the object. Hegel has a view of the development of history as a movement of self-consciousness through different historical and religious epochs. In this movement, the ideal is to be approached through an ever-increasing perfection of that self-consciousness. Judaism has an important, but essentially limited, role on the route to perfection. Judaism regards religious law as an independent entity, embodying the self-alienation of human society. This means that Judaism obeys a legal system without the capacity to master society itself. This domination of life by law is total in Judaism, Hegel asserts, and the religion itself is a compulsory faith, with a notion of God that is represented as a command and not as a truth. There is no scope for the notion of individual rights, but a type of divine despotism that bases itself on the family as the ultimate social unit.

Judaism sees the important relationship with God as essentially external, where God is the subject and his creatures the object. This results in alienation from the universe, the effect of which on individual Jews is that they do not regard their lives as belonging properly to them. They instead belong to God; Jews, as a result, are entirely passive in their attitudes to the creator. This is hardly surprising, Hegel reasons, since the nature of Judaism is to deny the possibility of raising human existence to the realm of spirit; that is, Judaism makes such a huge distinction between the divine and the finite that it cannot conceive of transcending that gap. God is regarded as spirit, but to the extent that he entirely transcends the world, which makes it impossible for humanity to be

reconciled with God. Judaism is a constant struggle against idolatry. This struggle is represented by Abraham's leaving his family in Ur to become a wanderer. Jews cannot see the things in the world as embodying the divine. As a result, the ordinary world is literally emptied of the divine. Unlike both Christianity and the religion of classical Greece, Judaism does not offer a dialectical relationship with what is invisible and infinite. Judaism can only serve as an antithesis in the progress of the self-realization of the spirit. That is, it is a stage along the path of increasing self-consciousness and serves to reveal how unsatisfactory a particular way of understanding our links with the divine are. Judaism, concludes Hegel, is a way of thinking which is primitive and needs to be superseded.

3 Sartre

By far the most interesting and original use of anti-Semitism as a philosophical vehicle was created by the existentialist Jean-Paul Sartre. He reverses the traditional way of understanding racism, which was to look at the particular group selected as a target of hatred and to ask what it was about it which made that targeting feasible. Sartre argues, by contrast, that we need to wonder what it is that racism provides for the racist. Furthermore we also should examine the role of racism in creating the despised group itself. The racist, according to Sartre, chooses hatred as a way of avoiding confronting problems rationally. Racism is an attempted escape from the need to think for oneself and to make free decisions. It is essentially a way of trying to escape from the human condition. In his accounts of anti-Semitism, Sartre points out how natural it is for particular people to adopt this way of thinking, since it saves them from the necessity of distinguishing themselves from their social role.

Sartre sees anti-Semitism as a failure of rationality; anti-Semites fail to use rationality to understand that their lowly role in life may be due to their lack of talent, or to an unjust social system. The Jew is the person whom the anti-Semite thinks is a Jew, incorrectly or otherwise. This can result in Jews trying to deny their Jewishness in order to avoid persecution. Sartre criticizes this strategy by Jews, arguing that it is as inauthentic an attitude as the adoption of anti-Semitism itself. The important point is that racism is primarily a problem for the racists. There is nothing the despised group can do to change that attitude.

4 Anti-Semitism as a philosophical notion

It is wrong to see anti-Semitism as a specifically Christian movement, or as having anything particularly to do with Germany. From the very first times that Jews lived among Gentiles, in Greece and the Roman Empire, prejudice against them was often commonplace. The construction of philosophical theories to justify this prejudice did take place for the first time within Christianity. It stems from early Christian polemics against the rival faith and from claims as to the original responsibility of Jews for the crucifixion of Jesus. Embedded in a variety of doctrinal and institutional structures, anti-Semitism was intensified by those structures themselves; for example, by the mercantile the role that Jews were compelled to play in the medieval society.

The theories that bolstered Christian anti-Jewish and anti-Judaic attitudes were followed up in the nineteenth and twentieth centuries by theories that sought to place racism on a scientific footing, arguing that racial differences define human possibilities (see Race, theories of). There were also theories that tried to establish the existence of a Jewish conspiracy against the Gentile world. Such ideas, although largely bereft of evidentiary support, form the theoretical underpinnings of modern anti-Semitism.

It would be quite mistaken to draw a direct linkage between the views of Kant and Hegel on Judaism and the development of the systematic anti-Semitism which has played such an important role in Europe in the last two centuries. For one thing, both Kant and Hegel had many Jewish followers. Those followers saw the ideas of German Idealism as entirely compatible with the principles of Judaism (see German Idealism). Hermann Cohen, for example, was a fervent Kantian and Emil Fackenheim a Hegel enthusiast. On the other hand, the radical distinction which Kant and Hegel both made between Judaism and other systems of religion, which they took to be more rational, was echoed by many lesser thinkers to imply the radical inferiority of the Jewish people or its culture and religion. The idea that Judaism is a morally limited and outmoded religion fosters the notion that Jews are legalistic and are wedded to an archaic way of life and thought, incapable, for that reason of being proper constituents of civil society. Indeed, these modern versions of ancient canards are used, ultimately, in behalf of the claim that Jews are the enemies of society. Such ideas reached an extreme in the thinking that gave rise to the Holocaust (see Holocaust, the).

Some thinkers such as Karl Popper see Hegel as being vastly influential in the development of totalitarian forms of thought, and it is true that Hegel's views on Jews and Judaism do appear congruent with some aspects of later German anti-Semitism. But we need to make a distinction between theories which regard Jews as playing a distinctive or even outmoded role in society, as with Kant and Hegel, and theories which explicitly justify treating Jews in discriminatory or even murderous ways. It may be that there is a psychological, political or rhetorical link between the two types of theories, but there is need not be a logical link between the two types of theories, but there is certainly not a logical link between them.

5 Anti-Semitism and Jewish self-understanding

The significance of anti-Semitism for Jewish self-understanding has been considerable. It played an important part in the construction of theories of Zionism, since at least one rationale for a Jewish state is the difficulty that Jews face as a minority in states dominated by Gentiles (see Zionism). Had anti-Semitism not been so prevalent, the motivation for statehood would not have been so powerful. Also, there is a tendency for some Jews to see themselves as defined through anti-Semitism, as Sartre pointed out, so that they are either victims or defiant resisters of persecution. Some critics of this form of self-definition have argued that it perverts the nature of Judaism. Judaism, they argue, is a religion based on belief in God and the covenant (Goldberg 1995). Its message should not be deflected persistent (if necessary) attention to the need constantly to repel the hostility of actual and potential persecutors. Many Jews find the most potent source of their emotional relationship to Judaism to be their commitment to the State of Israel. But this attachment itself reflects in part the phenomenon of anti-Semitism. For the enthusiasm is directed to the survival of the State and of the Jewish people in the face of threats to the survival of both.

See also: Fascism; Holocaust, the; Jewish philosophy, contemporary; Race, theories of

OLIVER LEAMAN
CLIVE NYMAN

References and further reading

Goldberg, M. (1995) *Why Should Jews Survive? Looking Past the Holocaust Toward a Jewish Future*, New York: Oxford University Press.(A sustained critique of Jewish self-definition through the Holocaust and anti-Semitism.)

Hegel, G. (1807) *Phänomenologie des Geistes (The Phenomenology of Mind)*, trans. J. Baillie, London: Allen & Unwin, 1931.(Hegel's mature views on the development of religions as spirit tries to understand itself as spirit.)

Kant, I. (1783) *Die Religion innerhalb der Grenzen der blossen Vernunft (Religion Within the Limits of Reason Alone)*, trans. T. Greene and H. Hudson, New York: Harper & Row, 1960.(Kant's mature work in which he discusses religious institutions and compares faiths.)

Kaufmann, W. (1960) 'The Hegel Myth and its Method', in W. Kaufmann, *From Shakespeare to Existentialism*, New York: Anchor, 95-128.(A powerful critique of Popper's attack on Hegel as a source for Naziism.)

Popper, K. (1945) *The Open Society and its Enemies*, London: Routledge & Kegan Paul.(Analysis of thinkers whom Popper argues are the intellectual founders of racism and totalitarianism.)

Rose, P. (1990) *Revolutionary Anti-Semitism in Germany from Kant to Wagner*, Princeton, NJ: Princeton University Press.(Interesting argument that most German Idealists were anti-Semitic, with useful references to personal quotations about their attitudes to the Jews.)

Rotenstreich, N. (1963) *The Recurring Pattern: Studies in Anti-Judaism in Modern Thought*, London: Weidenfeld & Nicolson.(Interesting discussion of Kant, Hegel and Toynbee and their uniformly negative views on Judaism.)

Sartre, J-P. (1946) *Réflexions sur la question juive (Reflections on the Jewish Question)*, trans. E. de Mauny, *Portrait of an Anti-Semite*, London: Secker & Warburg, 1948.(The best formulation of Sartre's approach to the significance of anti-Semitism.)

Antisthenes (*c*.445-*c*.365 BC)

Antisthenes was one of the most devoted followers of Socrates. As a young man he was heavily influenced by the display speeches of Gorgias the rhetorician and the interpretation of Homer practised by the Sophists. He himself wrote much in the same vein, although almost all has been lost.

Antisthenes' influence can be recognized most in the writer Xenophon. Although it is likely that he succeeded in annoying Plato and Isocrates, his influence on Cynicism has been greatly exaggerated.

Little survives of his moral philosophy, but what there is is Socratic in conception, and indeed Socrates' own courage and tenacity are its avowed inspiration. Antisthenes focuses on virtue, conceived as inner strength, a fortress founded on wisdom and its unassailable reasonings. Virtue is acquired and maintained by 'exertions', a term deliberately recalling the labours of Heracles: these consist of the struggle to overcome the difficulties of, for example, poverty or unpopularity, by understanding how they can be viewed as good things - provided the riches of the soul are intact. Pleasure and sex are accordingly seen as threats to virtue's integrity. Antisthenes enjoins us to redraw our moral categories: the good and just are our true friends and kin.

In theory of language Antisthenes defended the paradox that contradiction is impossible, deriving his argument from the idea that there can be no successful reference to anything except by its own 'account', revealing what it is.

1 Life and work

Of all Socrates' followers none was perceived as closer to him than the Athenian Antisthenes (Xenophon, *Memorabilia* III 11.17). His origins were apparently humble, and his circumstances not affluent. Calculation of his dates depends on scanty and doubtful evidence. What is clear is that Antisthenes successively inhabited the world of the fifth-century Sophists familiar from Plato's early dialogues and the no-less-competitive intellectual scene of the early fourth century, where Socrates' philosophical heirs jostled among themselves as well as with Sophists and rhetoricians.

Xenophon's Socrates suggests that it was in fact Antisthenes who introduced Prodicus and Hippias to their patron Callias (*Symposium* 4.62). A report by Diogenes Laertius makes him a pupil initially of Gorgias the rhetorician. Much of Antisthenes' literary output reflects his immersion in this milieu : 'He brings the rhetorical style into his dialogues, especially *Truth* and *Exhortations*' (Diogenes Laertius, VI 1). Like the Sophists he wrote numerous works on Homeric interpretation. How early in life he became attracted to Socrates is not clear: Diogenes Laertius suggests middle age; Xenophon's *Symposium* (dramatic date 421 BC) may imply sooner. It is uncertain therefore whether Antisthenes' Sophistic writings belong to a period of his life separate from those recommending Socratic ethics. The Socratic dialogues he is credited with must be assumed to post-date Socrates' death; and his thesis that a person who has attained discretion would do better not to study literature is hard to square with an approach to truth through study of the poets.

The only items surviving intact from a huge body of work are a complementary pair of speeches, quite possibly authentic, entitled *Ajax* and *Odysseus*, in which the two heroes contend for the arms of Achilles. Also preserved is a fragment of a dialogue debating the meaning of Homer's description of Odysseus as a man of 'many ways' (fr. 51), in a manner reminiscent of Plato's *Hippias Minor*. A certain amount can be gleaned from Xenophon's *Symposium* and from the list of book titles found in Diogenes Laertius, (VI 15-18). For further information about Antisthenes' thought we are dependent on doxographical summaries (notably Diogenes Laertius, VI 10-11), occasional quotations or mentions of his views, and - for what they are worth - anecdotes: for example, his notorious anti-Platonic remark 'I see a horse, but I do not see horseness' (fr. 50). These ancient reports are dominated by preoccupation with his Socratic conception of ethics, but Aristotle in particular has some tantalizing references to logical doctrines (see §4).

Xenophon sketches a playful portrait of Antisthenes in his *Symposium*. In contrast with the urbane Socrates he is presented as an abrupt, sarcastic and pedantic interlocutor. But as the work unfolds his fierce attachment to Socrates becomes ever more apparent. This is not simply a matter of his practising a form of Socratic *elenchos* (cross-examination). Antisthenes develops in Socratic style the doctrine that true riches are found in the storehouse

of the soul. And before the end he shows and avows his love for Socrates.

2 Antisthenes and the Cynics

An ancient tradition makes Antisthenes the originator of the Cynic philosophy and teacher of Diogenes of Sinope. It shapes not only the account of Antisthenes but the whole presentation of Cynicism and Stoicism in Diogenes Laertius (see especially VI 2, 15, 21, 103-5), and it is assumed in writers of the imperial period like Epictetus and Dio Chrysostom. Many modern discussions of Antisthenes accept the tradition, or suppose him a principal influence on Diogenes. However, the verdict of the best critical scholarship (for example, Dudley) is that the story is a fabrication.

Although Diogenes' chronology is a good deal more obscure than that of Antisthenes', it is at least doubtful whether the two *could* have met. None of the fragments of early Cynic writers mentions Antisthenes. The distinctive Cynic garb of cloak doubled up, staff and wallet is ascribed to him only in late sources. If Diogenes had a precedent for adopting it, contemporary Pythagoreans like Diodorus of Aspendos seem likelier models. Although Diogenes earned himself the nickname 'dog' - and 'Cynic' means 'Dog-philosopher' - the attempt to derive the expression from the Cynosarges gymnasium where Antisthenes is alleged to have taught seems a desperate measure.

The idea that Diogenes was in some sense Antisthenes' 'successor' could build on Antisthenes' ethics of frugality and inner resilience. What explains the promotion of that idea in the ancient sources is the need of later philosophy, and more particularly Stoicism, to rewrite its own history. The Stoics could not easily deny the impact of Cynicism on their founder Zeno. Opponents represented this as disreputable. Construction of a pedigree for Stoicism through Cynicism via Antisthenes and back to Socrates was to prove an effective way of recovering the moral high ground for Stoics.

3 Ethics

The principal account of Antisthenes' ethical teaching in Diogenes Laertius reflects the preoccupations of Stoicism outlined in §2, certainly in its emphases (VI 10-13, 104-5). Fortunately enough specific items of information are recorded in his pages and elsewhere for a reader to get a sense of Antisthenes' key ideas. The figures of Heracles and Cyrus of Persia, each the subject of more than one of Antisthenes' writings, were evidently presented as exemplars of the proper acquisition of virtue. To achieve it 'nothing was needed but the strength of a Socrates' (VI 11). And Antisthenes famously taught that exertion or *ponos* (as in the 'labours' of Heracles) is a good (VI 2). Deeds, not words, are what matter (VI 11; fr. 86).

This was not conceived as an anti-intellectualist position: 'Get understanding - or a noose', said Antisthenes (fr. 67). The exertions of the good are pictured as a campaign waged from a stronghold of wisdom (*phronēsis*), which has to be constructed by unassailable reasonings. The weapon no one can rob the good of is virtue, and their allies are those who are just as well as brave. The most indispensable item of knowledge Cyrus of Persia acquires is to *un*learn what is evil (fr. 21). Few specific examples of appropriate exertions survive. A bad reputation is said to be good and 'equal to exertion' (Diogenes Laertius, VI 11).

The sayings ascribed to Antisthenes indicate a strong tendency (congenial to Stoicism) to redraw concepts like kinship and friendship in terms of moral notions such as justice and goodness, or to prefer the candour of an enemy to the blandishments of flatterers. Similarly, the law of virtue displaces the established laws of the city as the imperative the wise person will obey. Antisthenes represents wisdom as what enables people to make everything work for them, so that nothing is any longer 'alien'. This philosophy of life is interpreted in terms of the Stoic doctrine that virtue is self-sufficient in Diogenes Laertius' doxography (VI 11).

Antisthenes' views on pleasure attracted quotation by later authors. Most popular was: 'I would rather be mad than feel pleasure' (Diogenes Laertius, VI 3; fr. 108). Other texts reveal a more nuanced position: pleasure is a good if it does not require subsequent repentance (fr. 110); 'there is no pleasure in a symposium without concord or in riches without virtue' (fr. 93); 'we should prefer the pleasure after, not before, exertions' (fr. 113). The power of sex provoked the comment: '*I* would shoot down Aphrodite if I could apprehend her, because she has corrupted many of our fine and good women' (fr. 109). But his most notorious remark on the subject sounds thoroughly Cynic: one should sleep with 'those women who will be grateful' (VI 3), interpreted by Xenophon

(*Symposium* 4.38) as recommending as entirely sufficient immediate satisfaction of physical desire with women who have nobody else to want them. When the doxography in Diogenes Laertius reports that for the sake of having children the wise man will marry, 'sleeping with the women whose nature is best suited' (VI 11), the plural gives cause for suspicion that Antisthenes' amoral advice has been bowdlerized in the interests of moral propriety.

Scholars have harboured hopes of recovering from Xenophon more substantial tracts of Antisthenian moral argument than these brief maxims and summaries. A long speech on the sufficiency of a frugal life provided we possess riches in the soul is put in his mouth in the *Symposium* (4.34-44). This is very likely a free reworking of genuine Antisthenian ideas. But other attempts to identify specific Antisthenian material in the *Memorabilia*, even if it derives inspiration from his view of Socrates and of philosophy, have gone unproven.

4 Language and *logos*

Several texts credit Antisthenes with advancing the claim that contradiction is an impossibility (frs 47-9). One story makes controversy with Plato about the claim the occasion for Antisthenes' composition of his dialogue *Satho* (Diogenes Laertius, III 25, VI 16): the title was intended as a satirical play on the name 'Plato' (fr. 37) - it means 'prick'.

The argument that no one can contradict anyone else carries resonances of the Sophists. For Protagoras(§3) it was a consequence of the relativity of truth. Plato has the Sophist Dionysodorus derive it from the premise that the *logos* or account of each thing must say that thing as it is. So if I succeed in saying the *logos* of some particular thing, you cannot be referring to that same thing at all if you attempt to contradict. According to Aristotle Antisthenes followed Dionysodorus' line of reasoning: 'nothing could be said except by its own (Greek *oikeios*) *logos*, one to one - from which it followed that contradiction is impossible, and one might almost say falsehood in general' (*Metaphysics* 1024b32-4; fr. 49).

It is tempting to connect Antisthenes' premise about 'its own *logos*' with Diogenes Laertius' information that 'he was the first to define *logos*, saying: "*logos* is what reveals the *what it is or was*"' (VI 3). Scholars have disputed whether *logos* here should be taken as 'statement' or 'definition'. Statement has been thought to be too weak a notion to capture something that reveals the *what it is or was*. But if the connection just proposed is correct, any statement that succeeds in being about some particular thing will be a true account of that thing as it is and as nothing else is, and such a statement will therefore satisfy the terms of Antisthenes' definition (or statement). His position on *logos* resembles the view of names ascribed to Cratylus in Plato's *Cratylus*. Similarities with the 'dream' theory of names in Plato's *Theaetetus*, which actually speaks of *oikeios logos* (202a), have also been canvassed in the past, but not much illumination either of the dialogue or of Antisthenes' views has been achieved by the comparison.

Despite affinities between this Antisthenian material and positions argued by Sophists and others in Plato's dialogues, it may well be that Antisthenes conceived of himself as explicating the philosophy of Socrates in logic as he did for ethics. Epictetus attributes to Antisthenes the dictum: 'The starting-point for education is the examination of *onomata* (words or names)' (*Discourses*, I 17.12) and follows it with a very similar remark ascribed to Xenophon's Socrates. Since Socrates was constantly asking the question: 'What is *X*?', it would be natural for Antisthenes to want to clarify what kind of answer would in principle meet the inquiry. There is evidence that he specifically attacked Plato's assumption that what Socrates wanted was a *definition*. Aristotle, discussing the requirements of a good definition, says at one point:

> There is a certain timeliness in the difficulty raised by the Antisthenians and similarly uneducated persons, namely, that it is impossible to define what a thing is (because, it is claimed, a definition is a long *logos*), whereas it can be explained what sort of thing it is, for example it can be explained that silver is like tin, but not what silver is.
>
> (*Metaphysics* 1043b23-8)

(A 'long *logos*' refers to an evasive rigmarole told, for example, by a slave to conceal the truth.) It is obviously difficult to reconcile Antisthenes' doctrine of *logos* with the proposal that the most an attempt at saying what something is can achieve is explanation of what it is like. Perhaps Antisthenes was arguing *ad hominem*: given Plato's wrong assumption that one can reveal the nature of things in terms not special to such things but common

to other things, a comparison is the best one can do - brief and to the point, if not a proper *logos*.

Some have concluded that the only statement that could qualify as a *logos* under Antisthenian rules would be a tautology. It is hard to believe this was his intention, not least because of evidence of exploitation of the idea of *oikeios logos* in his rhetorical theory (fr. 51). But further progress in interpretation is crippled by the complete absence in our sources of any explicit examples of an Antisthenian *logos*, except for his *logos* of *logos* itself.

See also: Cynics; Socratic dialogues; Socratic schools

MALCOLM SCHOFIELD

References and further reading

Antisthenes (*c*.445-*c*.365 AD) Fragments, ed. F. Decleva Caizzi, *Antisthenis Fragmenta*, Varese and Milano: Istituto Editoriale Cisalpino, 1966.(A pocket-sized edition, with good brief notes and bibliography; no translations are supplied. Fragments are cited in this entry according to this edition.)

Aristotle (*c*. mid 4th century) *Metaphysics*, ed. H. Tredenrick, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933-5, 2 vols; trans. in J. Barnes (ed.), *The Complete Works of Aristotle*, Princeton, NJ: Princeton University Press, 1984.(Tredenrick contains Greek text and English translation. Passages in Books VI and VII constitute the principal evidence for Antisthenes' logic.)

Diogenes Laertius (early 3rd century AD) *Lives of the Eminent Philosophers*, trans. R.D. Hicks, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(Book VI, sections 1-19 contain his life of Antisthenes.)

Dudley, D.R. (1937) *A History of Cynicism*, London: Methuen.(Chapter 1 remains an excellent introduction to the problem of Antisthenes' relationship to Cynicism.)

Epictetus (late 1st- early 2nd century AD) *Discourses*, trans. W.A. Oldfather, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925-1928.(Greek text and English translation: mentions Antisthenes occasionally.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae (Surviving Fragments of Socrates and the Socratics)*, Naples: Bibliopolis, 4 vols. (Volume 2 section VA contains fragments and other evidence about Antisthenes, without translation; volume 3 has a full bibliography and indexes; volume 4 notes 21-40 (195-411) contains a magisterial scholarly review of all major and most minor questions of fact and interpretation relating to Antisthenes.)

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, vol. 3, Cambridge: Cambridge University Press.(Pages 209-16 and 304-11 contain the best account of Antisthenes in English, judicious and attractively written, with much useful comment on bibliography.)

Xenophon (*c*.370s BC and *c*.385 BC) *Symposium*, trans. O.J. Todd, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1923. (Contains a convincing portrait of Antisthenes.)

Xenophon (*c*.360s BC) *Memorabilia*, trans. E.C. Marchant, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1923.(Antisthenes appears in *Memorabilia*, whose presentation of Socrates may owe something to Antisthenes' understanding of Socratic ethics.)

Applied ethics

Applied ethics is marked out from ethics in general by its special focus on issues of practical concern. It therefore includes medical ethics, environmental ethics, and evaluation of the social implications of scientific and technological change, as well as matters of policy in such areas as health care, business or journalism. It is also concerned with professional codes and responsibilities in such areas.

Typical of the issues discussed are abortion, euthanasia, personal relationships, the treatment of nonhuman animals, and matters of race and gender. Although sometimes treated in isolation, these issues are best discussed in the context of some more general questions which have been perennial preoccupations of philosophers, such as: How should we see the world and our place in it? What is the good life for the individual? What is the good society? In relation to these questions, applied ethics involves discussion of fundamental ethical theory, including utilitarianism, liberal rights theory and virtue ethics.

'Applied ethics' and 'applied philosophy' are sometimes used as synonyms, but applied philosophy is in fact broader, covering also such fields as law, education and art, and theoretical issues in artificial intelligence. These areas include philosophical problems - metaphysical and epistemological - that are not strictly ethical. Applied ethics may therefore be understood as focusing more closely on ethical questions. Nevertheless, many of the issues it treats do in fact involve other aspects of philosophy, medical ethics, for example, including such metaphysical themes as the nature of 'personhood', or the definition of death.

1 Definitions

While the name 'applied ethics' is comparatively new, the idea is not. Philosophy has traditionally concerned itself with questions both of personal morality (what should I do?) and public morality (what is the good society?), but while these questions are fundamental to applied ethics, they could also be said to characterize ethics in general. Applied ethics is therefore distinguished commonly as that part of ethics that gives particular and direct attention to practical issues and controversies.

In the private sphere, ethical issues include, for example, matters relating to the family (see Family, ethics and the), or to close personal relationships (see Friendship), the care of the old or disabled, the raising of the young, particularly where matters of morality are concerned, or personal ethical problems arising for the individual in the work-place. In the public sphere, applied ethics may involve assessing policy in the light of the impact of advances in biomedical technology (see Life and death; Risk; Technology and ethics), or assessing international obligations and duties to future generations in the light of environmental problems (see Future generations, obligations to; Population and ethics). The public arena includes, too, a range of issues for the plural society, such as ethnicity or gender in relation to discrimination, cultural understanding and toleration; more widely still, it may extend to issues of interest also to political philosophy, such as terrorism and the ethics of war. In all these matters, the concern of applied ethics is not only to supply a personal ethical perspective, but also to provide guidelines for public policy.

Applied ethics includes, as well, the area of professional ethics; it examines the ethical dilemmas and challenges met with by workers in the health care field - doctors, nurses, counsellors, psychiatrists, dentists - and by a wide range of workers in other professions including lawyers, accountants, managers and administrators, people in business, police and law enforcement officers. Specific ethical issues such as confidentiality, truth-telling, or conflicts of interest may arise in all or any of these areas, and most professions seek to codify their approaches and provide guidance for their members.

2 Theory and practice

Underlying all such issues are questions about justice, rights, utility, virtue and community. The practice of distinguishing between theoretical and applied ethics must, therefore, be treated with some caution. Indeed, some have regarded the term 'applied' as redundant, on the grounds that there cannot be an 'ethics' which is not applied: on the one hand, they argue, theoretical concepts such as rights and justice should not be viewed as mere abstractions; and, on the other, applied ethics should not be detached from its roots in traditional morality. But while it is important to stress this continuity, there are certain characteristic features of applied ethics which mark

it out in practice from theoretical ethics. These are (a) its greater attention to context and detail and (b) its more holistic approach - its willingness to link ethical ideals to a conception of human nature and human needs (see Human nature; Needs and interests). Thus practitioners of applied ethics may be more willing than proponents of traditional academic moral philosophy to recognize that psychology and sociology, a knowledge of culture and history, the insights of good literature, and even an understanding of humans as biological entities, are all relevant to the determination of moral issues in personal and public life.

The demarcation line between applied and theoretical ethics which this suggests may be drawn at that point on the spectrum of ethics where ethical theory stops short of normative recommendations and confines itself to the analysis of moral concepts such as ‘right’, ‘good’, ‘responsibility’, ‘blame’ and ‘virtue’ and to discussion of what might be called the epistemology of ethics - such theories as ethical realism, subjectivism and relativism (see Analytic ethics; Moral knowledge; Moral realism; Logic of ethical discourse). This is the area sometimes described as ‘meta-ethics’. Drawing the line at this point may be useful so long as it is not allowed to obscure the truth that applied and theoretical ethics are not discrete but lie on a continuum from the particular to the general, the concrete to the abstract.

The ultimate focus of applied ethics may well be entirely particular: the individual case-study. And it is this that gives rise to a further characteristic feature of applied ethics: its concern with dilemmas - not necessarily in the hard logical sense of situations in which it is impossible to act rightly because each of two opposite courses of action is either judged to be mandatory or judged to be wrong; but in the looser sense of cases in which a choice between courses of action may be extremely difficult, the arguments on both sides being compelling, and the person who must act being strongly influenced in opposing directions (for example, to sanction drastic medical intervention to save a severely disabled baby which would otherwise die, or to allow nature to take its course). It should be said, though, that choosing between options which are not morally equal is not, strictly speaking, a dilemma, although it is admittedly likely to be emotionally traumatizing, while choosing between moral obligations that are indisputably of equal weight is not a *moral* problem. The question for applied ethics in such cases may well be whether or not the available options are indeed morally equal.

Because it focuses on individual dilemmas, applied ethics must confront the question of universalization, which may also be seen as a ‘free rider’ problem: many things are judged to be wrong as a result of asking the question, ‘What if everyone did that?’, even though, in a particular case, it might seem harmless and more convenient for an individual to ignore the rule, while benefiting from the fact that everyone else is following it (see Universalism in ethics). The applied ethicist, like the theoretical moral philosopher, must find a way to deal with this problem, but for the applied ethicist, the problem is bound up with the need to employ what is sometimes called moral casuistry. This ancient science is not necessarily to be despised, for while a secondary meaning of the term ‘casuist’ is indeed ‘sophist’ or ‘quibbler’, it was not originally a term of abuse, but simply meant accepting in a theological context people’s desire to work out the ‘right answer’ to a difficult issue of conscience in a particular set of circumstances (see Casuistry).

3 Method

One method of reasoning employed in applied ethics may be compared to that of a designer who starts with a blueprint, but has to adapt it to the materials to hand and to the situations in which it is required. There is some resemblance in this case to the Hegelian method of dialectical reasoning, as well as to the method of reflective equilibrium favoured by such contemporary writers as Rawls (1971), in which intuitions in response to particular cases are measured against principles, causing them to be revised and their implications for particular cases again reappraised (see Moral justification §2). According to this view of the subject, the method of applied ethics is neither purely deductive nor purely inductive. For others, however, the deductive model is more powerful, and the question to be answered in any particular case is simply *which* (inviolable) principle it falls under. Others again would favour the inductive model, according to which, by clearly seeing what is right in particular cases, it becomes possible to formulate a general principle encompassing these and other particular judgments (see Universalism in ethics §3).

In general, discussion of ethical theories in applied ethics aims to pursue, in the direction of the highest degree of generality and abstraction, the question of what humans should do. In practice, discussion of theories is often confined to their implications for the resolution of particular problems, since applied ethics characteristically seeks

to answer the broad question with a much greater degree of particularity.

4 Critics and opponents

In seeking answers to practical problems, applied ethics runs counter to much recent philosophy. For the view that prevailed during the dominance of empiricism and positivism (the greater part of the twentieth century) is that philosophy can have nothing to say about pressing practical problems.

This view is grounded in two important philosophical arguments: (a) Hume's objection to arguments that seek to derive an 'ought' from an 'is' (see Hume, D. §4; Logic of ethical discourse §2-4); and (b) Moore's argument that to identify moral characteristics with 'natural' or empirical ones is to commit a 'naturalistic fallacy' (see Moore, G.E. §1; Naturalism in ethics §3). Both of these arguments must be resisted if applied ethics is to succeed in closing the gap between factual descriptions of situations and moral judgments, and both may partially at least be answered by insisting that some facts 'speak for themselves' - torture, child-murder, genocide, for example.

The argument that facts and values are to be kept apart is, however, less of an obstacle to philosophers outside the English-speaking world; the notion of praxis, for example, is familiar from various continental traditions, including Marxism, the Frankfurt School, and the philosophy of Habermas (see Theory and practice §3); while the idea of the philosopher as *engagé* - as concerned with playing a part in the world - is an important part of French existentialist thought, made familiar in the works of Sartre. These sources have, however, produced a different kind of challenge to the notion of applied ethics as an impartial and essentially reason-based approach to ethical issues in society. Objections to the conception of universal moral norms and to foundationalist procedures in reasoning (the 'postmodernist' challenge) are associated with recent developments in Marxist theory, certain feminist approaches to ethics and epistemology, and the deconstructionist movement - schools of thought which may also adopt an analysis of power-structures in society incompatible with belief in individual freedom of action (see Feminism and psychoanalysis; Deconstruction). Supporters of these theoretical positions often make strong claims for the recognition of rights, but this is probably better seen as exploitation of the preconceptions of their opponents, rather than as recognition of universal ethical concepts and human freedom.

Other critiques of traditional ethics may, however, be more sympathetic to applied ethics. On the basis of research revealing the contextuality of many women's responses to ethical dilemmas, some feminist writers, most prominently Carol Gilligan (1982), have argued that women in general are likely to adopt an ethic of care and responsibility to particular others rather than an abstract morality of principles, rights or justice (see Feminist ethics §1). Such an approach may well seem better adapted to the resolution of 'hard cases' in, for example, health care or social work (see Nursing ethics).

Similarly, the approach known as 'virtue ethics', with its emphasis on seeking the good in particular situations, may seem well adapted to applied ethics, even if its proponents sometimes appear to view it in opposition, regarding their own stand as more objective, and wrongly equating applied ethics with subjectivism and relativism (see Virtue ethics).

Other stereotypes to be rejected are political: applied ethics has typically been associated with vegetarianism, pacifism, feminism and environmentalism. It should be noted, however, that it also includes criticism and evaluation of these positions: defences of meat-eating or animal experiments, scepticism about feminism, and resistance to new 'ecological ethics' are to be found alongside more orthodox publications on library shelves. There is nothing wrong with variety of opinion so long as this is within a broad ethical framework, for it is of the essence of applied philosophy in general to approach individual issues in their own right and not as part of an ideological package-deal.

Applied ethics, then, is part of a whole view of the human condition and takes a broad view of ethical decision-making. Essentially, this is ethical decision-making seen as practical policy that consciously recognizes the constraints of moral norms, rights and ethical principles capable of commanding universal respect. Where this is accepted, the object of applied ethics is plain: it is to gain clearer perceptions of right and wrong, with a view to embodying these insights in manners and institutions.

5 Historical context

The inception of applied philosophy could well be said to coincide with that of the Western philosophical tradition as a whole, for the first of the early Greek philosophers, Thales (*c*.585 BC), is recorded as having combined his speculative philosophical interests with economic acumen and an interest in legal and political reform. Later schools of philosophy in ancient times - Pythagoreans, Epicureans, Stoics - offered their followers principles for living and even indeed distinctive codes of practice.

For both Plato and Aristotle, ethical and political questions were posed in terms of such notions as the good for man, the ultimate good, or what is good in itself and for its own sake (see Plato §16; Aristotle §21). Their assumption was that this inquiry led both to a way of life for the individual, and to a conception of the good society. They disagreed about whether this would lead an individual necessarily to live according to the ethical insight thus gained, Aristotle, unlike Plato in his earlier writings, allowing for the intervention of weakness of will to divert the person who has recognized the good from pursuing it (see Akrasia).

Subsequent philosophers frequently applied their ethical assumptions to particular cases, and saw this, not as a way of fractionizing moral philosophy - making it the science of the particular - but as a route to formulating guiding principles. Aquinas treated a range of practical issues including marriage and the family in *Summa theologiae*, and this tradition was developed further by Suárez (1612-21) and Grotius (1625). Locke (1689) wrote on the issue of toleration, Kant (1785; 1797) on suicide and on the question of whether it is ever right to tell a lie from benevolent motives (see Locke, J. §7; Toleration §1; Suicide, ethics of §5). Bentham (1789) put forward a complex theory of punishment, even formulating plans for a new type of prison, to be called the ‘panopticon’. He also wrote on legal and political reform. Hegel’s philosophy included views on the family and on punishment.J.S. Mill’s writings on toleration, paternalism and feminism (1859) continue to be of interest in the present day, as the controversies involved in these areas remain subjects of disagreement and debate (see Feminism; Paternalism), and Dewey’s theories of education (1916) exercised enormous practical influence on education systems in the USA and Britain (see Education, history of philosophy of).

The tradition in moral philosophy unsympathetic to applied ethics is in fact of fairly recent origin. It was associated with the dominance of positivism and empiricism in the philosophy of science, and the vogue for linguistic analysis in epistemology. This is a twentieth-century phenomenon and, right up to the closing years of the nineteenth century, a more generous conception of ethics flourished. If a certain myopia on applied issues is recognized amongst philosophers in the English-speaking world, coinciding roughly with the first half of the twentieth century, various explanations may be offered for the gradual return of visual focus. For those with an interest in medical ethics, a research project in Tuskegee in the USA in which a control group with syphilis remained untreated for decades after safe treatment was known to be possible is often cited as a trigger generating widespread discussion of issues such as autonomy, beneficence and nonmaleficence, medical confidentiality, and the ethics of experiments on human subjects (see Medical ethics §§1-3). This case may have been, however, a symptom rather than a cause, for in general medicine moved during those decades from being a practice with little power to influence the natural course of disease, to being a powerful interventionist tool. Whatever the specific cause, then, from roughly this period medical ethics became an arena of critical and controversial discussion.

Again in the USA, the Vietnam War and the protests which it generated are cited as having promoted discussion of a different range of applied issues (civil disobedience, duty to conscience versus duty to society) and as having led in a fairly direct way to the setting up of the Society for Philosophy and Public Affairs and the journal *Philosophy and Public Affairs* (see Civil disobedience; Conscience).

Others, focusing on the applied philosopher’s interest in animal welfare, cite the publication of the volume *Animal Liberation* (1975) by Peter Singer as ushering in a new conception of ethics as a practical and possibly even campaigning area (see Animals and ethics §3). Already, too, Rachel Carson’s *Silent Spring* (1962) had alerted the general public to many environmental hazards and thus opened the way to an enlarged philosophical perspective in which developments in science and technology and the way in which these were applied by firms and governments to the environment were seen as matters of ethical concern. It was a decade or so later that the internal operations of businesses became matters for ethical scrutiny, prompted by scandals connected with sharp practices such as insider trading.

Finally, it must be said that philosophy itself no doubt provided a spur to the growth of applied ethics. The preoccupation of academic moral philosophy with entirely minor moral issues in a century which had witnessed

two world wars and many accompanying gross violations of human rights, was too remarkable to pass for long, particularly with wider access to higher education and hence to the hitherto elite and somewhat esoteric pursuit of philosophy.

This account of the rise of contemporary applied ethics raises the question of what kind of study applied ethics is. Is it merely another kind of academic study, or is it committed to the promotion of change in the world? Is it conservative or radical? Reactionary or revolutionary? The answer to this last question is that it can be either. Reflection may make one seek to promote change for the better, but it may also cause one to recoil from change and seek to preserve what is best from the past. The controversial nature of most of the issues involved is itself a spur to their philosophical study, for it is probably true to say that until recently, despite differences of religious or ideological background, a common moral approach could in general be assumed, and accepted norms of moral behaviour could be taken as a starting-point for ethical reasoning. Such moral consensus cannot now be presupposed, and, while absolutist approaches are by no means inconsistent with mainstream philosophical ethics, in practice the defence of an absolute conception of morality against relativistic, subjective and utilitarian approaches is often associated with a religious perspective.

Many writers on applied ethics, however, adopt a secular utilitarian stance. These include the Australian philosopher Peter Singer, and the Oxford philosopher Jonathan Glover, who has written especially in the area of medical ethics (see Utilitarianism). R.M. Hare, in *Moral Thinking* (1981), puts forward a prescriptivist theory which combines utilitarianism with Kantian universalizability (see Prescriptivism). Also influential is the ethic of care mentioned above, which is often linked to gender differences. Other views include those of the Australian philosopher John Passmore, who defends a liberal moral perspective, especially in relation to environmental ethics, and John Rawls, whose notion of reflective equilibrium combines intuitionism with contract theory (see Moral justification §2). Rawls' *A Theory of Justice* (1971) inaugurated a new more practical approach in ethics, which had implications for economics, law and political theory. Sissela Bok has written on the fine texture of issues in public life in *Lying: Moral Choices in Public and Private Life* (1978) and *Secrets* (1984) (see Truthfulness); Mary Midgley, in *Beast and Man* (1978) and elsewhere, has discussed the relations between humans and other species; and Onora O'Neill (1986) has brought a Kantian ethic to bear on the issues of famine and poverty (see Development ethics). The debate between communitarians and libertarians about the ethics of capitalism and the role of welfare can also be seen as a part of applied ethics (see Community and communitarianism; Market, ethics of the). The German philosopher, Jürgen Habermas, an influential figure both in continental Europe and the English-speaking world, has put forward a notion of consensus as the object of theory expressed in practice.

6 Professional ethics

Similar divisions may reveal themselves in professional ethics, although the idea that there should be special codes of ethics peculiar to particular professions has been current since ancient times, when the Hippocratic oath was required of those engaging in medical practice. Many modern groups, including engineers, nurses and lawyers, have adopted formal codes setting standards of ethical practice for their profession (see Professional ethics).

Ethics also plays an increasing role in the training of professionals. Often the preferred approach is through the use of case studies, sometimes fictional, sometimes using videos of actual cases. One problem with the case study approach is its possible negative effect. In stressing that there are at least two sides to many ethical problems, and in presenting ethical theories as giving conflicting outcomes, they may risk generating a facile moral or cultural relativism - the view that there are only opinions, not answers. The use of case studies and discussion based on situational ethics may also tacitly undermine principles (see Situation ethics). In contrast, some courses aim simply to increase the moral sensitivity of trainees, on the assumption that if this is successful they will go on to make good professional decisions (see Examples in ethics).

7 Are there ethical experts?

Applied ethics does not involve a claim of moral expertise, but often involves collaboration with specialists in practical areas in order to arrive at policy decisions that allow ethical considerations a determining role.

There is now wide acceptance of the principle of ensuring that a philosophical or ethical viewpoint is represented in certain kinds of forums, such as public enquiries, the reports of legislative committees or commissions of

inquiry, and hospital ethics committees. The USA has a President's Commission to report directly on bioethical issues to the US President, the UK has a National Bioethics Committee funded independently of government, while in France there is a French National Committee on Ethical Affairs in Public Debate. In 1985, the Council of Europe created a multidisciplinary body with experts appointed by each member country, now called the Comité Directeur de Bioéthique (CDBI). Canada set up a Royal Commission on New Reproductive Technologies, and the European Parliament commissions advice on scientific and technological policy options. Other countries are following a similar pattern. In addition, the Council of Europe in 1990 began working on a European Convention on biomedical ethics, which would be a legally binding instrument on all countries signing it, the object being ultimately to harmonize European legislation.

Individuals are also used as consultants on public policy issues. In Europe, Jonathan Glover, in collaboration with nationals of other European countries, produced a report on fertility and the family for the European Commission (1989), while Will Kymlicka has advised on this topic as a member of the Canadian Royal Commission and, in the USA, Arthur Caplan was a member of the President's Task Force on National Health Care Reform. In Britain, the philosopher Mary Warnock (1985) was responsible for official reports on the educational needs of children with disabilities and learning difficulties, and on new developments in reproductive medicine and embryology; Bernard Williams (1979) played a similar role in relation to pornography and censorship. The debate about euthanasia in the Netherlands has engaged philosophers, lawyers, and social theorists. Less happily, a visit by Peter Singer to Germany provoked widespread protest related to the debate on euthanasia and has led to the unpopularity of bioethics in some circles, and a general and unjustified rejection of applied ethics.

Some achievements in these areas may also be recorded; examination of the ethics of clinical trials, for example, particularly in relation to AIDS, led to a total reconceptualization of what clinical trials require, and to a multi-choice system being devised which is both scientifically acceptable and also allows a more acceptable level of choice to patients and physicians.

8 Research in applied ethics

In general, those who fund research regard the gathering of facts, often called the 'generation of new knowledge', as crucial; philosophy, in contrast, appears to involve reflection on facts, while normative philosophy generates proposals for action or policy. Applied ethics offers at its best an opportunity to combine these approaches: for facts to be made the fruitful object of analytic and morally sensitive reflection, and for philosophical inquiry to accept the discipline of the need to take account of the practical framework within which speculation is cast.

Research in applied ethics, then, ideally starts from a perceived problem and is motivated to find a solution to that problem. It is frequently interdisciplinary. A research programme is often inspired by technological progress, for it is this that has placed ethical considerations at the heart of many areas of public debate. Typical of these are the controversies already mentioned surrounding the new technologies of reproduction: embryo research, donation of gametes, surrogate motherhood, which raise questions about the status of the human embryo and the definition of parenthood (see Reproduction and ethics).

Other appropriate areas where ethics impinges on practical inquiry include, for example, the ethical implications of the Human Genome Project, the ethics of confidentiality, insurance in relation to AIDS or inherited disease, the care of the elderly, homelessness, and mental illness (see Genetics and ethics; Medical ethics §§4-5; Nursing ethics). One caveat to be noted here, however, is that simply gathering data about what people think is right is sociology, not ethics, applied or otherwise.

9 Institutions

Many research centres have been created in recent decades. Their function is usually to conduct research, to produce publications and to arrange lectures, seminars and conferences on practical issues of ethical concern.

North America has the best-established institutional network. First in the field was the Hastings Center, New York (1969), then the Center for Philosophy and Public Affairs, University of Maryland and the Center for the Study of Ethics in the Professions at the Illinois Institute of Technology (1976), the Center for the Study of Values, University of Delaware (1977), and the Social Philosophy and Policy Center, Bowling Green State University, Ohio (1981). There are now many other centres in universities in the USA and elsewhere, including, in the UK, the

Centre for Philosophy and Public Affairs at the University of St Andrews, the Centre for Medical Law and Ethics at King's College London, and the Social Values Research Centre at the University of Hull. The Netherlands has Bioethics Centres in Utrecht and Maastricht and work in applied ethics in the Scandinavian countries is increasing, with a strong interest in reproductive ethics in Aarhus, Denmark and in animal welfare issues in Copenhagen. The European Business Ethics Network (EBEN) began with an initiative from Switzerland, and business ethics is also well-established in Spain and Germany. Apart from university-based units, the Society for Applied Philosophy (1982) has general interests in most areas of applied ethics and has a broad membership not confined to professional philosophers.

Australia has been a pioneer in many fields of applied ethics: Peter Singer, together with Helga Kuhse, founded the Centre for Human Bioethics (1980) at Monash University, and there are now several other applied ethics centres in Australasia; it is worth noting the particular degree of interest there in environmental ethics, where the issues of species preservation, wilderness, and ecological threats such as damage to the ozone layer are of direct concern to residents.

The creation of a Chair of Environmental Ethics at Warsaw University represents the strong interest, partly political in origin, in environmental ethics in the former communist countries of Eastern Europe. Other countries where applied ethics is of growing interest are parts of Southeast Asia, including Thailand and Hong Kong, India, and several African countries.

See also: Agricultural ethics; Bioethics; Business ethics; Engineering and ethics; Environmental ethics; Journalism, ethics of; Information technology and ethics; Responsibilities of scientists and intellectuals; Sexuality, philosophy of

BRENDA ALMOND

References and further reading

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, ed. T. Gilby, London: Eyre & Spottiswoode, 1963-, 60 vols.(A theological work, which provides a place for the Aristotelian virtues, conscience and natural law, and also deals with practical ethical issues including marriage and the family.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, revised F. Rosen, Oxford: Clarendon Press, 1996.(Sets out the foundations of a utilitarian theory of ethics, law and punishment.)

Bok, S. (1978) *Lying: Moral Choices in Public and Private Life*, New York: Pantheon.(Practical discussions of ethical considerations raised by practice of lying in connection with many professional areas.)

Bok, S. (1984) *Secrets*, New York: Random House.(Practical discussion of issues of confidentiality and secrecy in public life.)

Carson, R. (1962) *Silent Spring*, London: Hamish Hamilton.(Although not itself a work in applied ethics, this was instrumental in drawing public attention to the threat to the environment poaed by the use of pesticides in agriculture.)

Clark, S.R.L. (1977) *The Moral Status of Animals*, Oxford: Oxford University Press. (A philosophical and ethical examination of the status of animals in their relations to human beings.)

Dewey, J. (1916) *Democracy and Education*, New York: The Free Press, 1966.(Sets out an approach to the education of young children, which includes the use of discovery methods and the arrangement of the classroom as wookroom for cooperative projects.)

Gilligan, C. (1982) *In a Different Voice: Psychological Theory and Women's Development*, Cambridge, MA: Harvard University Press; 2nd edn, 1993.(Account of research which first suggested that women might be articulating a different view of morality, contextual and personal rather than universal and abstract.)

Glover, J. (1977) *Causing Death and Saving Lives*, Harmondsworth: Penguin. (Comprehensive and wide-ranging discussion from a utilitarian standpoint of issues in medical ethics.)

Glover, J. *et al*. (1989) *Fertility and the Family: The Glover Report on Reproductive Technologies to the European Commission*, London: Fourth Estate Ltd.(A cooperative international report for the European Commission.)

Grotius (1625) *De iure belli ac pacis*, trans. F. Kelsey, *The Law of War and Peace*, Oxford: Oxford University Press, 1925.(Sets out the conditions for the conduct of a just war, and reasons which would justify going to war.)

Hare, R.M. (1981) *Moral Thinking*, Oxford: Oxford University Press.(Important statement of a distinctive

philosophical approach to practical ethics: universal prescriptivism.)
Hursthouse, R. (1987) *Beginning Lives*, Oxford: Blackwell.(A discussion of the ethical aspects of abortion.)
Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964. (Provides the fundamentals of Kant's moral theory according to which actions have moral worth if they are done from the motive of duty and fit with the principle of universalizability.)
Kant, I. (1797) 'On a Supposed Right to Lie from Altruistic Motives', in *The Philosophy of Immanuel Kant*, trans. L.W. Beck, Chicago, IL: University of Chicago Press, 346-9.(Argues, contrary to the usual opinion, that it is never right to tell a lie, even to achieve some good.)
Locke, J. (1689) *A Letter concerning Toleration*, ed. J.W. Gough and R. Kilbansky, Oxford: Oxford University Press, 1968.(Discusses political authority and its limits, and sets out the fundamental rights of individuals.)
Midgley, M. (1978) *Beast and Man*, Ithaca, NY: Cornell University Press. (Philosophical discussion of issues in sociobiology.)
Mill, J.S. (1859) *On Liberty*, Harmondsworth: Penguin, 1982.(Classic defence of liberty and toleration; includes practical examples of the way on which these principles are to be applied.)
O'Neill, O. (1986) *Faces of Hunger*, London: Allen & Unwin.(Kantian approach to the issues of famine and Third World poverty.)
Passmore, J. (1974) *Man's Responsibility for Nature: ecological problems and Western traditiions*, London: Duckworth; 2nd edn, 1980.(Comprehensive discussion by a leading Australian philosopher of ethical issues concerned with the environment.)
Rachels, J. (ed.) (1971) *Moral Problems*, New York: Harper & Row; 3rd edn, 1979. (Usefully representative collection of articles on applied ethics, including discussion of racism, discrimination, obligations to poorer countries.)
Rachels, J. (1986) *The End of Life: Euthanasia and Morality*, Oxford and New York: Oxford University Press. (Discusses ethical issues concerning the end of life and the issue of euthanasia.)
Rawls, J. (1971) *A Theory of Justice*, Oxford: Oxford University Press.(Seminal work which initiated a new era of practical philosophy; aims to establish universal principles for social living on rational grounds.)
Singer, P. (1975) *Animal Liberation*, New York: New York Review of Books Press; repr. London: Jonathan Cape, and New York: Random House, 1990.(Presents Singer's arguments for ethical treatment of animals as contrasted with current practise in food production and laboratory research.)
Singer, P. (1979) *Practical Ethics*, Cambridge: Cambridge University Press; 2nd edn, 1993.(Utilitarian perspective on applied ethics; includes reading guides and bibliographies on specific topics.)
Singer, P. (ed.) (1986) *Applied Ethics*, Oxford: Oxford University Press.(Classic articles, including Hume on suicide and Mill on the death penalty.)
Suárez, F. (1612-21) Selections from Three Works of Francisco Suárez, S.J., trans. G.L. Williams *et al*., in J.B. Scott (ed.) *Classics of International Law* 20, vol. 2, Oxford: Clarendon Press, 1944.(Discusses the nature of law, largely following Aquinas, and also the concept of obligation.)
Walzer, M. (1977) *Just and Unjust Wars*, New York: Basic Books.(An examination of arguments concerning war and morality.)
Warnock, M. (1985) *A Question of Life: The Warnock Report on Fertilisation and Embryology*, Oxford: Blackwell.(Based on a report on the ethics and law of reproductive medicine prepared for the British government, with recommendations for policy.)
Williams, B. *et al*. (1979) *Report of the Committee on Obscenity and Film Censorship ('The Williams Report') Cmnd 7772*, London: Her Majesty's Stationery Office; abridged version repr. as B. Williams (ed.) *Obscenity and Film Censorship*, Cambridge: Cambridge University Press, 1981.(A report prepared for the British government advising on the issue of censorship and the law.)
Ethics (1890-) (A journal which publishes articles in all areas of ethics.)
Philosophy (1925-) Cambridge: Cambridge University Press.(A journal which publishes articles in all areas of philosophy, usually of a non-technical area.)
Social Theory and Practice (1970-) (An interdisciplinary journal of social philosophy.)
Philosophy and Public Affairs (1971-) (A journal dedicated to the philosophical study of issues of public concern, including legal, social and political problems.)
Journal of Applied Philosophy (1984-) (A journal contributing to discussion of ethical and philosophical issues in many areas of practical debate.)
Public Affairs Quarterly (1987-) (A journal of philosophical studies of public policy issues.)

Apuleius (*c.* AD 125-180)

The Latin writer Apuleius of Madaura was a professional rhetorician, a novelist and an amateur Platonist. His handbook of Platonism and his essay on the guardian spirit of Socrates are valuable sources on Middle Platonism. The handbook is comparable to that of his probable contemporary Alcinous, but covers only physics (including metaphysics) and ethics.

Apuleius was born in Madaura in North Africa, of respectable provincial family. He received the best rhetorical education available in Carthage and then, around AD 150, set out for Athens to study philosophy. It is possible that, when there, he studied with the Platonist Calvenus Taurus, whose lectures Aulus Gellius also attended, but we cannot be certain. None the less, Apuleius acquired in Athens a good working knowledge of Platonism, which he puts to various uses.

After more time in North Africa, he returned to practise as a lawyer and rhetorician in Carthage, where he was celebrated in an inscription of *c.* AD 161 as a rhetorician, poet and Platonic philosopher. At some time after this he composed his chief contribution to philosophy, a basic handbook, *On Plato and his Doctrine*, of a similar nature to that of the more-or-less contemporary text of Alcinous. This is written in a flat, scholastic style, very unlike that of his rhetorical compositions, so much so that its authenticity has been questioned, but without adequate grounds. Another philosophical treatise, probably also from this period, is his essay *On the Daemon of Socrates*; in florid, rhetorical style, it does however contain useful Middle Platonist doctrine on daemons.

On Plato is in two books, the first dealing with 'physics' (including metaphysics), the second with ethics and politics. A third section, devoted to logic, is envisaged in the preface, but there is no sign of it; instead, we have the (possibly spurious) *De interpretatione*. *On Plato* begins with four chapters on the life of Plato, which constitute valuable evidence for the state of the Plato myth in Apuleius' time. Thereafter, Apuleius embarks on a systematic survey of Platonic philosophy, beginning with the first principles God, Matter and the Ideas (chaps 5-6), and continuing through the topics of the formation of the elements and of the world, the soul of the world, time, the heavenly bodies, classification of animate beings, gods and daemons, and fate and free will. This is followed by an anthropological section, on the parts of the soul and the body, the senses and bodily health and disease. All of this, as with the parallel treatise of Alcinous, is heavily dependent on Plato's *Timaeus* (see Plato §16). Most distinctive, perhaps, is his doctrine of fate and providence (ch. 12), postulating three levels of providence, in which he shows interesting affinities with the pseudo-Plutarchan treatise *On Fate*.

Book II of *On Plato* covers the various main topics of ethical theory, followed by a discussion of the nature of the perfect sage, and a short disquisition on politics, again not differing greatly from Alcinous, and identifying Apuleius as an adherent of the Peripateticizing wing of Middle Platonism. As for the work *De interpretatione*, it gives a summary account of Aristotelian logic, reflecting later developments in the Peripatetic school

Apuleius was a man of many parts. He is best known today for his novel *The Golden Ass*. As a philosopher, he would make no great claim to originality, but he provides useful evidence for the state of Platonism in the mid second century AD.

See also: Platonism, Early and Middle §§1,4, 7-9

JOHN DILLON

List of works

Apuleius (AD 158/9) *Apology*, ed. P. Vallette, Paris: Les Belles Lettres, 2nd edn, 1960. (With French translation.)

Apuleius (between AD 140 and 180) *On the Daemon of Socrates, On Plato and his Doctrine, On the World*, in J. Beaujeu (ed.) *Apulée, Opuscules philosophiques et fragments (Apuleius, Philosophical Minor Works and Fragments)*, Paris: Les Belles Lettres, 1973.(With French translation.)

Apuleius (between AD 140 and 180) *De interpretatione*, in D. Londrey and C. Johanson (eds) *The Logic of Apuleius*, Leiden: Brill, 1987.(In English translation with a useful logical commentary.)

Apuleius (between AD 140 and 180) *The Golden Ass* (or *Metamorphoses*), trans. J.A. Hanson, Cambridge, MA: Harvard University Press, 1989.(Of little interest philosophically, but notable as the only surviving novel in Latin.)

References and further reading

Dillon, J. (1977) *The Middle Platonists*, London: Duckworth.(Chapter 6 offers an introductory account of Apuleius.)

Hijmans, B.J. (1987) *Apuleius Philosophus Platonicus (Apuleius, Platonic Philosopher)*, in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin: de Gruyter, II 36: 1, 395-475.(A rather idiosyncratic account, concentrating more on Apuleius' style than on his content.)

Moreschini, C. (1978) *Apuleio e il Platonismo (Apuleius and Platonism)*, Florence: Olschki.(Essays in Italian on various aspects of Apuleius.)

Sullivan, M.W. (1967) *Apuleian Logic*, Amsterdam: Hakkert.(A good discussion of sources and influence.)

Aquinas, Thomas (1224/6-74)

Aquinas lived an active, demanding academic and ecclesiastical life that ended while he was still in his forties. He nonetheless produced many works, varying in length from a few pages to a few volumes. Because his writings grew out of his activities as a teacher in the Dominican order and a member of the theology faculty of the University of Paris, most are concerned with what he and his contemporaries thought of as theology. However, much of academic theology in the Middle Ages consisted in a rational investigation of the most fundamental aspects of reality in general and of human nature and behaviour in particular. That vast domain obviously includes much of what is now considered to be philosophy, and is reflected in the broad subject matter of Aquinas' theological writings.

The scope and philosophical character of medieval theology as practised by Aquinas can be easily seen in his two most important works, Summa contra gentiles (Synopsis [of Christian Doctrine] Directed Against Unbelievers) and Summa theologiae (Synopsis of Theology). However, many of the hundreds of topics covered in those two large works are also investigated in more detail in the smaller works resulting from Aquinas' numerous academic disputations (something like a cross between formal debates and twentieth-century graduate seminars), which he conducted in his various academic posts. Some of those topics are taken up differently again in his commentaries on works by Aristotle and other authors. Although Aquinas is remarkably consistent in his several discussions of the same topic, it is often helpful to examine parallel passages in his writings when fully assessing his views on any issue.

Aquinas' most obvious philosophical connection is with Aristotle. Besides producing commentaries on Aristotle's works, he often cites Aristotle in support of a thesis he is defending, even when commenting on Scripture. There are also in Aquinas' writings many implicit Aristotelian elements, which he had thoroughly absorbed into his own thought. As a convinced Aristotelian, he often adopts Aristotle's critical attitude toward theories associated with Plato, especially the account of ordinary substantial forms as separately existing entities. However, although Aquinas, like other medieval scholars of western Europe, had almost no access to Plato's works, he was influenced by the writings of Augustine and the pseudo-Dionysius. Through them he absorbed a good deal of Platonism as well, more than he was in a position to recognize as such.

On the other hand, Aquinas is the paradigmatic Christian philosopher-theologian, fully aware of his intellectual debt to religious doctrine. He was convinced, however, that Christian thinkers should be ready to dispute rationally on any topic, especially theological issues, not only among themselves but also with non-Christians of all sorts. Since in his view Jews accept the Old Testament and heretics the New Testament, he thought Christians could argue some issues with both groups on the basis of commonly accepted religious authority. However, because other non-Christians, 'for instance, Mohammedans and pagans - do not agree with us about the authority of any scripture on the basis of which they can be convinced... it is necessary to have recourse to natural reason, to which everyone is compelled to assent - although where theological issues are concerned it cannot do the whole job', since some of the data of theology are initially accessible only in Scripture (Summa contra gentiles I.2.11). Moreover, Aquinas differed from most of his thirteenth-century Christian colleagues in the breadth and depth of his respect for Islamic and Jewish philosopher-theologians, especially Avicenna and Maimonides. He saw them as valued co-workers in the vast project of philosophical theology, clarifying and supporting doctrine by philosophical analysis and argumentation. His own commitment to that project involved him in contributing to almost all the areas of philosophy recognized since antiquity, omitting only natural philosophy (the precursor of natural science).

A line of thought with such strong connections to powerful antecedents might have resulted in no more than a pious amalgam. However, Aquinas' philosophy avoids eclecticism because of his own innovative approach to organizing and reasoning about all the topics included under the overarching medieval conception of philosophical Christian theology, and because of his special talents for systematic synthesis and for identifying and skilfully defending, on almost every issue he considers, the most sensible available position.

1 Early years

Thomas Aquinas was born at Roccasecca, near Naples, the youngest son of a large Italian aristocratic family. As is

generally true of even prominent medieval people, it is hard to determine exactly when he was born; plausible arguments have been offered for 1224, 1225 and 1226. He began his schooling in the great Benedictine abbey at Monte Cassino (1231-9), and from 1239-44 he was a student at the University of Naples. In 1244 he joined the Dominican friars, a relatively new religious order devoted to study and preaching; by doing so he antagonized his family, who seem to have been counting on his becoming abbot of Monte Cassino. When the Dominicans ordered Aquinas to go to Paris for further study, his family had him abducted en route and brought home, where he was kept for almost two years. Near the end of that time his brothers hired a prostitute to try to seduce him, but Aquinas angrily chased her from his room. Having impressed his family with his high-minded determination, in 1245 Aquinas was allowed to return to the Dominicans, who again sent him to Paris, this time successfully.

At the University of Paris, Aquinas first encountered Albert the Great, who quickly became his most influential teacher and eventually his friend and supporter. When Albert moved on to the University of Cologne in 1248, Aquinas followed him there, having declined Pope Innocent IV's extraordinary offer to appoint him abbot of Monte Cassino while allowing him to remain a Dominican.

Aquinas seems to have been unusually large, and extremely modest and quiet. When during his four years at Cologne his special gifts began to be apparent, despite his reticence and humility, Albert assigned the still-reluctant Aquinas his first active part in an academic disputation. Having failed in his efforts to shake his best student's arguments on this occasion, Albert declared, 'We call him the dumb ox, but in his teaching he will one day produce such a bellowing that it will be heard throughout the world'.

In 1252 Aquinas returned to Paris for the course of study leading to the degree of master in theology, roughly the equivalent of a twentieth-century PhD. During the first academic year he studied and lectured on the Bible; the final three years were devoted to delivering in lecture form his commentary on Peter Lombard's *Sentences*, a standard requirement for the degree at that time (see Lombard, P.). Produced in 1253-6, Aquinas' massive commentary (often referred to as the *Scriptum super libros Sententiarum (Commentary on the Sentences)* is the first of his four theological syntheses. It contains much valuable material, but because it is superseded in many respects by his great *Summa contra gentiles (Synopsis [of Christian Doctrine] Directed Against Unbelievers)* and *Summa theologiae (Synopsis of Theology)* the *Scriptum* has not yet been studied as much as it should be.

During that same four-year period, Aquinas produced *De ente et essentia (On Being and Essence)*, a short philosophical treatise written for his fellow Dominicans at Paris. Although it owes something to Avicenna's *Metaphysics*, *De ente* is distinctively Aquinas' own, expounding many of the concepts and theses that remained fundamental to his thought throughout his career (see §9 below).

2 First Paris regency

In the spring of 1256, Aquinas was appointed regent master (professor) in theology at Paris, a position he held until the end of the academic year 1258-9. *Quaestiones disputatae de veritate (Disputed Questions on Truth)* is the first of his sets of disputed questions and the most important work he produced during those three years. It grew out of his professorship, which obliged him to conduct several formal public disputations each year. *Quaestiones disputatae de veritate* consists of twenty-nine widely ranging Questions, each devoted to some general topic such as conscience, God's knowledge, faith, goodness, free will, human emotions and truth (the first Question, from which the treatise gets its name). Each Question is divided into several Articles, and the 253 articles are the work's topically specific units: for example, q. 1, a. 9 is 'Is there truth in sense perception?'

The elaborate structure of each of those articles, like much of Aquinas' writing, reflects the 'scholastic method', which, like medieval disputations in the classroom, had its ultimate source in Aristotle's recommendations in his *Topics* regarding cooperative dialectical inquiry. Aquinas' philosophical discussions in that form typically begin with a yes/no question. Each article then develops as a kind of debate. It begins with arguments for the answer opposed to Aquinas' own position; these are commonly, if somewhat misleadingly, called 'objections'. Next come the arguments *sed contra* (but, on the other hand), which are in later works often reduced to a single citation of some generally accepted authority on Aquinas' side of the issue. The *sed contra* is followed by Aquinas' reasoned presentation and defence of his position. This is the master's 'determination' of the question, sometimes called the 'body' of the article (indicated by 'c' in references). An article normally concludes with Aquinas' rejoinders to each of the objections (indicated by 'ad 1', and so on, in references).

Conducting 'disputed questions' was one of the duties of a regent master in theology, but the theology faculty also provided regular opportunities for 'quodlibetal questions', occasions on which a master could, if he wished, undertake to provide replies to any and all questions proposed by members of the academic audience. These occasions were scheduled, for the master's own good, during the two penitential seasons of the church year. Aquinas seems to have accepted this challenge on at least five of the six such occasions occurring during his first regency at Paris, producing *Quaestiones quodlibetales (Quodlibetal Questions)* in which he offers his considered judgment on issues ranging from whether the soul is to be identified with its powers to whether the damned behold the saints in glory.

Aquinas' commentaries on Boethius' *De trinitate (On the Trinity)* and *De hebdomadibus* (sometimes referred to as 'How Substances are Good') are his other philosophically important writings from this period of his first regency. Although several philosophers had commented on those Boethian treatises in the twelfth century, the subsequent influx of Aristotelian works had left them almost universally disregarded by the time Aquinas wrote his commentaries (see Aristotelianism, medieval; Boethius, A.M.S.). No one knows why or for whom he wrote them, but he might well have undertaken these studies for his own edification on topics that were then becoming important to his thought. The *De trinitate* commentary (*Expositio super librum Boethii De trinitate*) presents Aquinas' views on the relationship of faith and reason and on the methods and interrelations of all the recognized bodies of organized knowledge, or 'sciences'. Boethius' *De hebdomadibus* is the *locus classicus* for the medieval consideration of the relation between being and goodness. Dealing with this topic in his commentary on that treatise, Aquinas also produced his first systematic account of metaphysical participation, one of the important Platonist elements in his thought. Participation, he claims, obtains when the metaphysical composition of something includes, as one of the thing's metaphysical components, *X*, which also belongs, to something else that is *X* in its own right in a way that is presupposed by the first thing's having *X*. In this way a running man participates in running, human being participates in animal, and an effect participates in its cause (see also §9 below) (see Platonism, medieval).

3 Naples and Orvieto: *Summa contra gentiles* and biblical commentary

Aquinas' activities between 1259 and 1265 are not well documented, but he seems definitely to have left his professorship at Paris at the end of the academic year 1258-9. He probably spent the next two years at a Dominican priory in Naples, working on the *Summa contra gentiles*, which he had begun in Paris and which he subsequently finished in Orvieto where, as lector, he was in charge of studies at the Dominican priory until 1265.

Summa contra gentiles is unlike Aquinas' three other theological syntheses in more than one respect. Stylistically, it is unlike the earlier *Scriptum* and the later *Summa theologiae* in not following the scholastic method; instead, it is written in ordinary prose divided into chapters, like his *Compendium theologiae (Compendium of Theology)* which he seems to have written immediately afterwards (1265-7). More importantly, the *Scriptum*, *Summa theologiae* and the *Compendium* are all contributions to revealed theology, which essentially includes the data of revelation among the starting points of its theorizing. In *Summa contra gentiles*, on the other hand, Aquinas postpones revealed theology to the last (fourth) book, in which he deals with the 'mysteries', the few doctrinal propositions that cannot be arrived at by natural reason alone and that have their sources in revelation only; and he takes these up with the aim of showing that even those propositions 'are not opposed to natural reason' (*Summa contra gentiles* IV.1.3348). He devotes the first three books to fully developing a natural theology, dependent on natural reason of course, but independent of revelation. As developed in Books I-III, this natural theology is able to accomplish a very large part of theology's job, from establishing the existence of God through working out details of human morality (see also §13 below).

Discussions important for understanding Aquinas' positions in many areas of philosophy are also scattered, not always predictably, among interpretations of the text in his biblical commentaries. During Aquinas' stay in Orvieto and around the time he was writing Book III of *Summa contra gentiles*, on providence and God's relations with human beings, he also produced his *Expositio super Iob ad litteram (Literal Commentary on Job)*, one of the most fully developed and philosophical of his biblical commentaries, rivalled in those respects only by his later commentary on Romans. The body of the Book of Job consists mainly of the speeches of Job and his 'comforters'. Aquinas sees those speeches as constituting a genuine debate, almost a medieval academic disputation (determined in the end by God himself), in which the thought develops subtly, advanced by arguments. His construal of the

argumentation is ingenious, the more so because twentieth-century readers have tended to devalue the speeches as tedious reiterations of misconceived accusations countered by Job’s slight variations on the theme of his innocence.

Aquinas’ interpretation of the book’s subject is also unlike the modern view, which supposes it to be the biblical presentation of the problem of evil, raised by a good God’s permitting horrible suffering to be inflicted on an innocent person. Aquinas seems scarcely to recognize that Job’s story raises doubts about God’s goodness. As he interprets it, the book explains the nature and operations of divine providence, which he understands as compatible with permitting bad things to happen to good people. As Aquinas sees it:

> If in this life people are rewarded by God for good deeds and punished for bad, as Eliphaz [one of the comforters] was trying to establish, it apparently follows that the ultimate goal for human beings is in this life. But Job means to rebut this opinion, and he wants to show that the present life of human beings does not contain the ultimate goal, but is related to it as motion is related to rest, and a road to its destination.
>
> (*Expositio super Iob ad litteram* 7: 1-4)

The things that happen to a person in this life can be explained in terms of divine providence only by reference to the possibility of that person’s achieving the ultimate goal of perfect happiness, the enjoyment of union with God in the afterlife.

In discussing Job’s lament that God doesn’t hear his prayers, Aquinas says that Job has that impression because God sometimes ‘attends not to a person’s pleas but rather to his advantage. A doctor does not attend to the pleas of the invalid who asks that the bitter medicine be taken away (supposing that the doctor doesn’t take it away because he knows that it contributes to health). Instead, he attends to the patient’s advantage; for by doing so he produces health, which the sick person wants most of all.’ In the same way, God sometimes permits a person to suffer despite prayers for deliverance, because he knows that those sufferings are helping that person achieve what he or she wants most of all (*Expositio super Iob ad litteram* 9:16).

4 Rome: disputed questions, Dionysius and the *Compendium*

In 1265 Aquinas went from Orvieto to Rome, having been appointed to establish a Dominican *studium* (something like a twentieth-century college) and to serve as regent master there. This Roman period of his career, which lasted until 1268, was particularly productive. Some of his major works dating from 1265-8 are just what would have been expected of a regent master in theology, in particular, three sets of disputed questions, *Quaestiones disputatae de potentia (Disputed Questions on [God’s] Power)*, *Quaestio disputata de anima (Disputed Question on the Soul)* and *Quaestio disputata de spiritualibus creaturis (Disputed Question on Spiritual Creatures)*. In the earliest of these, *De potentia*, there are eighty-three Articles grouped under ten Questions; the first six questions are on divine power, while the final four are on problems associated with combining the doctrine of Trinity with God’s absolute simplicity. The much shorter *De anima* is concerned mainly with metaphysical aspects of the soul, concluding with some special problems associated with the nature and capacities of souls separated from bodies (Articles 14-21). The eleven articles of *De spiritualibus creaturis* again address many of those same concerns but also go on to some consideration of angels as another order of spiritual creatures besides human beings, whose natures are only partly spiritual.

During this same period, or perhaps while he was still at Orvieto, Aquinas wrote a commentary on the pseudo-Dionysian treatise *De divinis nominibus (On the Divine Attributes)*, a deeply Neoplatonist account of Christian theology dating probably from the sixth century. Aquinas, like everyone else at the time, believed that it had been written in the apostolic period by the Dionysius who had been converted by St Paul. For that reason, and perhaps also because he had first studied the book under Albert at Cologne, it had a powerful influence on Aquinas’ thought. Very early in his career, while he was writing his *Scriptum*, he thought Dionysius was an Aristotelian (*Scriptum* II, d.14, q.1, a.2), but while writing the commentary on this text he realized that its author must have been a Platonist (*Expositio super librum Dionysii De divinis nominibus*, prooemium; *Quaestiones disputatae de malo* 16.1, ad 3). His commentary, which makes clear sense of a text that is often obscure, may, like his commentaries on Boethius, have been written for his own purposes rather than growing out of a course of lectures. In any case, his study of Dionysius is one of the most important routes by which Platonism became an essential ingredient in his own thought (see also Pseudo-Dionysius).

The *Compendium theologiae (Compendium of Theology)*, already mentioned in connection with *Summa contra gentiles*, was once thought to have been written much later and to have been left incomplete because of Aquinas' death. However, its similarity to *Summa contra gentiles* not only in style but also in content has lately led many scholars to assign it to 1265-7. Among Aquinas' four theological syntheses, the *Compendium theologiae* is unique in the brevity of its discussions and in having been organized around the 'theological virtues' of faith, hope and charity. Had it been completed, it might have provided a novel reorientation of the vast subject matter of medieval theology, but Aquinas wrote only ten short chapters of the second section, on Hope, and none at all of the third section, on Charity. He did complete the first section on Faith, but since most of the 246 chapters in the section simply provide much briefer treatments of almost all the theological topics Aquinas had already dealt with in *Summa contra gentiles*, the *Compendium* as he left it seems important mainly as a precis of material that is developed more fully in the other work (and in *Summa theologiae*).

5 Rome: Aristotelian commentary

While some of Aquinas' prodigious output in Rome from 1265-8 is, broadly speaking, similar to work he had already done, it also includes two important innovations, one of which is the first of his twelve commentaries on works of Aristotle. At the beginning of this commentary on *De anima* (*Sententia super De anima*), his approach is still a little tentative and (for Aquinas) unusually concerned with technical details. These features of the work once led scholars to describe the commentary on the first book of *De anima* as a *reportatio* (an unedited set of notes taken at his lectures), or even to ascribe this first third of Aquinas' commentary to another author. However, Gauthier (1984: *275-82) has argued persuasively that the difference between the commentary's treatments of Book I and of Books II and III of *De anima* is explained by differences between the books themselves, and that in fact none of Aquinas' commentaries on Aristotle resulted from lectures he gave on those books. Discrepancies within this work, the first of Aquinas' Aristotelian commentaries, are likely to be at least in part a consequence of the fact that he was finding his way into this new sort of enterprise, at which he quickly became very adept. In a recent volume of essays on Aristotle's *De anima*, Martha Nussbaum describes Aquinas' work as 'one of the very greatest commentaries on the work' and 'very insightful' (Nussbaum 1992: 3-4). T.H. Irwin, a leading interpreter of Aristotle, acknowledges that at one point in the *Sententia libri Ethicorum (Commentary on Aristotle's Nicomachean Ethics)* Aquinas 'actually explains Aristotle's intention more clearly than Aristotle explains it himself' (Irwin 1992: 467). Such judgments apply pretty generally to Aquinas' Aristotelian commentaries, all of which are marked by his extraordinary ability as a philosophical commentator to discern a logical structure in almost every passage he examines in every sort of text: not only Aristotle's but also those of others, from Boethius to St Paul.

Since commenting on Aristotle was a regular feature of life for a member of a medieval arts faculty but never part of the duties of an academic theologian, Aquinas' many Aristotelian commentaries were technically extra-curricular and therefore an especially impressive accomplishment for someone who was already extremely busy. Some scholars, admiring Aquinas' achievements in general but focusing on the fact that his professional career was entirely in the theology faculty, have insisted on classifying only the Aristotelian commentaries as philosophical works. Certainly these commentaries are philosophical, as purely philosophical as the Aristotelian works they elucidate. However, Aquinas wrote these commentaries not only to make good philosophical sense of Aristotle's very difficult texts but also, and more importantly, to enhance his own understanding of the topics Aristotle had dealt with. As he remarks in his commentary on *De caelo*, 'the study of philosophy has as its purpose to know not what people have thought, but rather the truth about the way things are' (*Sententia super libros De caelo et mundo* I.22.228), and he believed that the theologian's attempt to understand God and everything else in relation to God was the fundamental instance of the universal human drive to know the truth about the way things are. Moreover, his view of the best way of making intellectual progress in general looks very much like the age-old method of philosophy: ' But if any people want to write back against what I have said, I will be very gratified, because there is no better way of uncovering the truth and keeping falsity in check than by arguing with people who disagree with you' (*De perfectione spiritualis vitae* 26) (see Aristotelianism, medieval; Aristotle; Aristotle Commentators).

6 Rome: *Summa theologiae*

The other important innovation from Aquinas' three-year regency in Rome is *Summa theologiae*, his greatest and

most characteristic work, begun in Rome and continued through the rest of his life. *Summa theologiae*, left incomplete at his death, consists of three large Parts. The First Part (Ia) is concerned with the existence and nature of God (Questions 1-43), creation (44-9), angels (50-64), the six days of creation (65-74), human nature (75-102) and divine government (103-19). The Second Part deals with morality, and in such detail that it is itself divided into two parts. The first part of the Second Part (IaIIae) takes up human happiness (Questions 1-5), human action (6-17), the goodness and badness of human acts (18-21), passions (22-48) and the sources of human acts: intrinsic (49-89) and extrinsic (90-114). The second part of the Second Part (IIaIIae) begins with the three theological virtues and corresponding vices (Questions 1-46), goes on through the four 'cardinal virtues' and corresponding vices (47-170) and ends with special issues associated with the religious life (171-89). In the Third Part, Aquinas deals with the incarnation (Questions 1-59) and the sacraments (60-90), breaking off in the middle of his discussion of penance.

Aquinas thought of *Summa theologiae* as a new kind of textbook of theology, and its most important pedagogical innovation, as he sees it, is in its organization. He says he has noticed that students new to theology have been held back in their studies by several features of the standard teaching materials, especially 'because the things they have to know are not imparted in an order appropriate to a method of teaching': an order he proposes to introduce. It may well have been his enthusiasm for this new approach that led him to abandon work on his quite differently organized *Compendium theologiae*, and his natural preoccupation during this period with the writing of *Summa theologiae* Ia may also help to account for the fact that his other work of that time shows a special interest in the nature and operations of the human soul, the subject matter of Questions 75-89 of Ia (see §13 below).

7 Second Paris regency

In 1268 the Dominican Order again assigned Aquinas to the University of Paris, where he was regent master for a second time until, in the spring of 1272, all lectures at the university were canceled because of a dispute with the bishop of Paris. The Dominicans then ordered Aquinas to return to Italy.

Among the astounding number of works Aquinas produced in those four years is the huge Second Part of *Summa theologiae* (IaIIae and IIaIIae), nine Aristotelian commentaries, a commentary on the pseudo-Aristotelian *Liber de causis* (which, as Aquinas was the first to realize, is actually a compilation of Neoplatonic material drawn from Proclus), sixteen biblical commentaries and seven sets of disputed questions (including the set of sixteen *Quaestiones disputatae de malo (Disputed Questions On Evil)*, the sixth of which provides his fullest discussion of free choice). His literary productivity during this second regency is the more amazing because he was at the same time embroiled in various controversies.

Sending Aquinas back to Paris in 1268 seems to have been, at least in part, his order's response to the worrisome movement of 'Latin Averroism' or 'radical Aristotelianism', then gaining ground among members of the arts faculty who were attracted to interpretations of Aristotle found in the commentaries of Averroes (see Averroism). However, only two of his many writings from these years seem to have obvious connections with the Averroist controversy. One of these, his treatise *De unitate intellectus, contra Averroistas (On [the Theory of] the Unicity of Intellect, against the Averroists)* is an explicit critique and rejection of a doctrine distinctive of the movement; the theory, as Aquinas describes it, that the aspect of the human mind 'that Aristotle calls the possible intellect… is some sort of substance separate in its being from the body and not united to it in any way as its form; and, what is more, that this possible intellect is one for all human beings' (*De unitate intellectus*, prooemium). After briefly noting that this view's incompatibility with Christian doctrine is too obvious to warrant discussion at any length, Aquinas devotes the entire treatise to showing that 'this position is no less contrary to the principles of philosophy than it is to the teachings of the Faith', and that it is even 'entirely incompatible with the words and views' of Aristotle himself (*De unitate intellectus*, prooemium).

Besides the unicity of intellect, the other controversial theory most often associated with thirteenth-century Averroism is the beginninglessness of the universe. In many of his works Aquinas had already considered the possibility that the world had always existed, skilfully developing and defending the bold position that revelation alone provides the basis for believing that the world began to exist, that one cannot prove either that the universe must or that it could not have begun, and that a world both beginningless and created is possible (although, of course, not actual). The second of Aquinas' Parisian treatises that is plainly relevant to Averroism is *De aeternitate mundi, contra murmurantes (On the Eternity of the World, against Grumblers)*, a very short, uncharacteristically

indignant summary of his position. Aquinas could not complain that Aristotle had been misinterpreted regarding the eternity of the world; after initially supposing this to be the case, he had become convinced that Aristotle really did think he had proved that the world must have existed forever. Aquinas' position on this issue did not distance him enough from the Averroists in the view of their contemporary 'Augustinian' opponents, most notably the Franciscans Bonaventure and Pecham. In fact, the 'Grumblers' against whom Aquinas directed this treatise were probably not so much the Averroists in the arts faculty as those Franciscan theologians who maintained that they had demonstrated the impossibility of a beginningless world (see Augustinianism; Eternity of the world, medieval views of).

Aquinas' principled dissociation from some important Franciscans on this point must have helped to make his second Paris regency much more troubled than his first. In disputations conducted in Paris in 1266-7, the Franciscan master William of Baglione implicated Aquinas' views in the propositions he attacked, claiming that things Aquinas was saying encouraged the two heretical Averroist theses denounced by Bonaventure, namely the eternity of the world and the unicity of the intellect. 'The "blind leaders of the blind" decried by William evidently include Thomas as their chief' (Tugwell 1988: 226). It has also been persuasively argued that Aquinas' *De aeternitate mundi* was directed in particular against his Franciscan colleague in theology, John Pecham (Brady 1974). It seems, then, that Aquinas' development of a distinctly philosophical theology - which, like Albert's, was more Aristotelian than Augustinian - was dividing him from his colleagues in the Paris faculty of theology during these years. It may also have been bringing him closer to the philosophers in the arts faculty.

8 Last days

In June 1272 the Dominicans ordered Aquinas to leave Paris and go to Naples, where he was to establish another studium for the order and to serve as its regent master. Except for some interesting collections of sermons (originally preached in his native Italian dialect), the works dating from this period - two Aristotelian commentaries and the Third Part of *Summa theologiae* - were left unfinished. On or about 6 December 1273, while he was saying mass, something happened to Aquinas that left him weak and unable to go on writing or dictating. He himself saw the occasion as a special revelation. When Reginald of Piperno, his principal secretary and longtime friend, tried to persuade him to return to work on the Third Part of *Summa theologiae*, he said, 'Reginald, I can't.' And when Reginald persisted, Aquinas finally said, 'Everything I've written seems like straw by comparison with what I have seen and what has been revealed to me'. He believed that he had at last clearly seen what he had devoted his life to figuring out and, by comparison, all he had written seemed pale and dry. Now that he could no longer write, he told Reginald, he wanted to die. Soon afterwards he did die, on 7 March 1274 at Fossanuova, Italy, on his way to the Council of Lyons, which he had been ordered to attend.

9 Metaphysics

Every part of Aquinas' philosophy is imbued with metaphysical principles, many of which are recognizably Aristotelian. Consequently, concepts such as potentiality and actuality, matter and form, substance, essence, accident and the four causes - all of which are fundamental in Aquinas' metaphysics - should be considered in their original Aristotelian context (see Aristotle §11). He invokes such principles often, and he employs them implicitly even more often. Two of his earliest writings - *De principiis naturae (On the Principles of Nature)* and especially *De ente et essentia (On Being and Essence)* - outline much of his metaphysics, almost as if they had been designed to provide guidelines for the development of his philosophy. Perhaps the most important thesis argued in *De ente* is the one that became known as 'the real distinction', Aquinas' view that the essence of any created thing is really, not just conceptually, distinct from its existence. Metaphysically speaking, corporeal beings are composites of form and matter, but all creatures, even incorporeal ones, are composites of essence and existence. Only the first, uncreated cause, God, whose essence is existence, is absolutely simple.

Except for his commentary on Aristotle's *Metaphysics*, Aquinas devoted no mature treatise to metaphysics itself. However, since he considers metaphysics to be the science of being considered generally (*ens commune*), and since he argues that being itself is first of all God himself and that all being depends on God, his philosophy does begin with metaphysics insofar as the most systematic presentations of his thought (in *Summa contra gentiles* and *Summa theologiae*) start with the investigation of God-in-himself considered as the foundation of the nature and existence of everything (see for example, *Summa contra gentiles* III.25; *Expositio super librum Boethii De trinitate* V.4, VI.1; §14 below).

Being, Aquinas says, is intellect's most fundamental conception, 'inherently its most intelligible object and the one in which it finds the basis of all conceptions…. Consequently all of intellect's other conceptions must be arrived at by adding to being… insofar as they express a mode of being which is not expressed by the term "being" itself' (*Quaestiones disputatae de veritate* 1.1c). There are, he claims, just two legitimate ways of making such additions. The first results in the ten Aristotelian Categories, each of which is a 'specified [or specific] mode of being' - substance, quantity, quality and the rest. The results of 'adding to being' in the second way are less familiar. Aquinas takes them to be five modes of being that are entirely general, characterizing absolutely every being. That is, being, wherever and however instantiated, exhibits these five modes, which transcend the Categories because they are necessary modes of all specified being: thing (*res*), one, something (*aliquid*), good, true. These five, together with being itself, are the 'transcendentals', predicable correctly (if sometimes a little oddly) of absolutely anything that is. 'Good' and 'true' are the philosophically interesting cases, because some beings are obviously not good and because 'true' seems applicable only to propositions.

The claim that all beings are true depends on taking 'true' in the sense of 'genuine', as in 'true friend', a sense that had been explored in detail by Anselm of Canterbury. In Anselm's view, any being is true in this sense to the extent to which it agrees with the divine idea of such a thing (and is otherwise false, but only to some extent). Absolutely every thing that is agrees to some extent with the divine idea that is an ingredient in its causal explanation. Propositions are true if they correspond to the way things are in the world; things in the world are true if they correspond to what is in the mind, God's mind first, ours derivatively. So, Aquinas says, 'in the soul there is a cognitive and an appetitive power. The word "good", then, expresses the conformity of a being to appetite (as is said at the beginning of the *Ethics*: "The good is what all desire"). The word "true", however, expresses the conformity of a being to intellect' (*Quaestiones disputatae de veritate* 1.1c).

The central thesis of Aquinas' meta-ethics grows out of this theory of the transcendentals. The thesis is the metaphysical principle that the terms 'being' and 'good' are the same in reference, differing only in sense (*Summa theologiae* Ia.5.1). What all desire is what they take to be the good, and what is desired is at least perceived as desirable (see for example, *Summa contra gentiles* I.37; III.3). Desirability is thus an essential aspect of goodness. If a thing of a certain kind is genuinely desirable as a thing of that kind, it is desirable to the extent to which it is perfect of that kind: a complete specimen, free from relevant defect. But a thing is perfect of its kind to the extent to which it has actualized its specifying potentialities, the potentialities that differentiate its species from other species in the same genus. So, Aquinas says, a thing is desirable as a thing of its kind and hence good of that kind to the extent to which it is actualized and in being (*Summa theologiae* Ia.5.1). Generally, then, 'being' and 'goodness' have the same referent: the actualization of specifying potentialities. The actualization of a thing's specifying potentialities to at least some extent is on the one hand its existence as such a thing; it is in this sense that the thing is said to have being. However on the other hand, the actualization of a thing's specifying potentialities is, to the extent of the actualization, that thing's being whole, complete, free from defect: the state all things are naturally aimed at. It is in this sense that the thing is said to have goodness (see for example *Summa theologiae* IaIIae.1.5; 94.2; *Summa contra gentiles* III.3; *Quaestiones disputatae de veritate* 21.1-2.)

Aquinas' concept of analogy is important to his thought, though perhaps not so important as it has sometimes been made to seem. It is often presented, correctly, in terms of analogical predication. However, his concept of analogy can be explained at a more fundamental level in connection with causation. Setting aside 'accidental' causation - for example, a gardener's uncovering buried treasure - Aquinas thinks that efficient causation always involves an agent (*A*), a patient (*P*), and a form (*f*). In non-accidental efficient causation, *A* antecedently has *f*, somehow. *A*'s exercising causal power on *P* brings about *f* in *P*, *somehow*. Thus the efficient cause is *A*'s acting (or exercising a power it has), and the effect is *P*'s having *f*. The fact that *A* and *P* can have *f* in several different ways is what is brought out in '*somehow*'. The paradigm - straightforward efficient causation - is the kind Aquinas calls *univocal*: cases in which first *A* and then *P* have *f* in just the same way, and in which *f* can therefore be predicated truly of each in just the same sense. The metal hotplate and the metal kettle bottom resting on it are both called hot univocally: the form heat in these two causally related objects is the same specifically and differs only numerically.

However, Aquinas also recognizes two kinds of non-univocal efficient causation. The first, *equivocal* causation, characterizes cases in which there is no obvious respect in which to say that the *f* effected in *P* is found

antecedently in *A*, and yet there is a natural causal connection (as there standardly is an etymological explanation for equivocal predication). If *A* is solar power and its effect is the hardening (*f*) of some clay (*P*), then obviously the sun's power is not itself hard, as the clay is. To say what it is about solar power that hardens clay will not be as easy as explaining the heating of the kettle, and yet the hardening of the clay must, somehow, be brought about by that power. In such a case, *A* has *f* only in the sense that *A* has the power to bring about *f* in *P*.

Second, *analogical* causation occurs when, for instance, a blood sample (*P*) is correctly labelled 'anaemic', although of course the blood itself doesn't have anaemia and cannot literally be anaemic. The physiology of the sample's donor (*A*) brings about a condition (*f*) in the sample that is an unmistakeable sign of anaemia in *A*, thus justifying that (analogical) labeling of the sample. For theological purposes, Aquinas is interested not in natural analogical causation but rather in the artificial kind: the kind that involves ideas and volitions, the artisan's kind. 'In other agents [the form of what is to be brought about occurs antecedently] in keeping with *intelligible* being, as in those agents that act through intellect - the way a likeness of the house exists antecedently in the builder's mind' (*Summa theologiae* Ia.15.1c). Since the status of entirely univocal causation depends on there being a merely numerical difference between the *f* in *A* and the *f* in *P*, an intellective agent effecting its ideas is obviously not a univocal cause. But neither is this difference between the antecedent *f* and the consequent *f* so wide as to constitute equivocal causation. In fact, the kind of association between the idea and its external manifestation is closer than the kind found in natural analogical causation; and since, in Aquinas' view, 'the world was brought about not by chance but by God acting through intellect… it is necessary that there be a form in the divine mind, a form in the likeness of which the world was made' (*Summa theologiae* Ia.15.1c). God, then, is the non-univocal, non-equivocal, intellectively analogical efficient cause of the world (see Causation; God, concepts of).

10 Philosophy of mind

Aquinas' philosophy of mind is part of his more general theory of soul, which naturally makes use of his metaphysics. Obviously he is not a materialist - most obviously because God, the absolutely fundamental element of his metaphysics, is in no way material. Aquinas classifies every thing other than God as either corporeal or incorporeal (spiritual); he sometimes calls purely spiritual creatures - such as angels - 'separated substances' because of their essential detachment from body of any sort. However, this exhaustive division is not perfectly exclusive because human beings, simply by virtue of the human soul, must be classified not as simply corporeal but also as spiritual in a certain respect.

Merely having a soul of some sort is not enough to give a creature a spiritual component, however. Every animate creature has a soul (*anima*) - 'soul is what we call the first principle of life in things that live among us' (*Summa theologiae* Ia.75.1c) - but neither plants nor nonhuman animals are in any respect spiritual. Aquinas holds that even the merely nutritive soul of a plant, or the nutritive + sensory soul of a beast, is like the soul of a human being in being the *form* of a body. No soul, no first principle of life, can be *matter*. On the other hand, any vegetable or animal body has the life it has only in virtue of being a body whose special organization confers on it natural potentialities: that is, in virtue of the substantial form that makes it actually be such a body. Therefore, the first principle of life in a living non-human body, its soul, is no bodily part of that body but is rather its form, one of the two metaphysical components of the composite of matter and form that every body is. For plants and beasts, unlike humans, the form that is the soul goes out of existence when the composite dies, and it is in that sense that the souls of plants and beasts are not spiritual.

Only the soul of a human being is analysed as nutritive + sensory + *rational*. Aquinas thinks of this soul not as three nested, cooperating forms, but as the single substantial form that gives a human being its specifically human mode of existence. (In defending this thesis of 'the unicity of substantial form', Aquinas differed from most of his contemporaries.) He often designates this entire substantial form by its distinctively human aspect of rationality. He also thinks that the human soul, unlike the souls of plants and beasts, is subsistent: that is, it continues to exist after separating from the body in death. He says, for example: 'It is necessary to say that that which is the principle of intellective activity, what we call the soul of a human being, is an incorporeal, subsistent principle' (*Summa theologiae* Ia.75.2c). The human soul, just because it is distinctively mind (the principle of intellective activity), must therefore be described not only as incorporeal but also as subsistent.

It may seem impossible for Aquinas' account to accommodate the claim that souls persist and engage in mental acts after the death of the body. If the separated soul is a form, what is it a form of? Aquinas is not a universal

hylomorphist; unlike some of his contemporaries, he does not think that there is ‘spiritual matter’ that angels or disembodied souls have as one of their components, but rather that they are separated forms that configure no matter at all. Thus when he claims that the soul exists apart from the body, he seems to be holding the view that there can be a form with nothing of which it is the form. Moreover, Aquinas thinks that an angel or the soul separated from the body engages in mental activity. However, a form seems not to be the sort of thing that enages in acts of any sort, and so it appears that even if there were some way to explain the existence of the soul apart from the body, its acting could not be explained.

In this connection, it is helpful to examine Aquinas’ broader view of form. The world is ordered metaphysically in such a way that at the top of the universal hierarchy there are forms - God and angels - that are not forms *of* anything. Near the bottom of the hierarchy are forms that configure matter but cannot exist in their own right, apart from the corporeal composites they inform. The forms of inanimate things and of animate, non-rational things are of that sort. Those forms inform matter, but when the resultant composites cease to exist, those forms also cease to exist. In the middle - ‘on the borderline between corporeal and separated [that is, purely spiritual] substances’ - are human souls, the metaphysical amphibians (*Quaestio disputata de anima* 1c). Like angels, human souls are subsistent, able to exist on their own; but, like the forms of inanimate things, human souls configure matter.

Seeing the soul in this light helps to explain some of what is initially puzzling in Aquinas’ account. The human soul has a double character. On the one hand, unlike the forms of other material things, it is created by God as an individual entity in its own right, able to exist by itself as do purely immaterial angels. On the other hand, like the form of any corporeal thing, it exists in the composite it configures, and it comes into existence only with that composite, not before it (see Soul, nature and immortality of the).

11 Theory of knowledge

Nature, Aquinas thinks, must be arranged so as to enable human beings in general to satisfy their natural desire to know (*Sententia super Metaphysicam* I.1.3-4). His view of the arrangement actually provided seems at first too tight to be true, involving some sort of formal identity between the extra-mental object (*O*) and the cognizing faculty (*F*) in its actually cognizing *O*. However, Aquinas takes that (Aristotelian) identity-claim to mean only that the form of *O* is somehow *in F* (*Summa theologiae* Ia.85.2, ad 1). *O*’s form comes to be in *F* when *F* receives *species*, either sensory or intellective, of *O*. These species may be thought of as encodings of *O*’s form. If *O* is a particular corporeal object - an iron hoop, for instance - then in *O* itself *O*’s form informs matter to produce an iron hoop of just those dimensions at just that spatio-temporal location. (In Aquinas’ account of individuation, it is matter that is ‘designated’ or ‘determinate’ in this way that individuates *O*’s form: see for example *De ente et essentia* 2.) But when the appropriately encoded form is received in an external sense faculty *F* (which uses a bodily organ), then, even though it is received *materially* in *F*’s matter, it is nonetheless received differently from its reception in the matter of the hoop. The imposition of the form on the matter of the sense organ constitutes an ‘intentional’ or ‘spiritual’ reception of the form, contributing to a cognition of the hoop rather than metaphysically constituting a new, individuated matter-form composite.

Sensory species received in external senses are standardly transmitted to ‘internal senses’, the organs for which, Aquinas thought, must be located in the brain. Among the most important of these for purposes of cognition are ‘phantasia’ and ‘imagination’ (although Aquinas usually treats imagination as part of the power of phantasia). Phantasia and imagination produce and preserve ‘phantasms’, the sensory data that are necessary preconditions for intellective cognition. Imagination and phantasia are also indispensable to conscious sensory cognition. In Aquinas’ view, sensible species themselves are not the objects of cognition, and what he says about phantasia suggests that having sensible species isn’t sufficient for having sensory cognition. *O* itself, currently having a natural effect on the external senses, is consciously sensed because phantasia has processed *O*’s sensible species into phantasms.

The form presented in a phantasm has of course been stripped of its original, individuating matter, but a phantasm of *O* remains particularized as a phantasm in virtue of having been received in the *different* matter of phantasia’s organ, while remaining recognizably the form of *O* because of the details of *O* that are preserved in it. However, cognition of *O* as an iron hoop is conceptual, intellective cognition, for which phantasms are only the raw material.

In intellect itself, Aquinas distinguishes two Aristotelian ‘powers’. The first is *agent intellect*, the essentially active

or productive aspect of intellect, which acts on phantasms in a way that produces 'intelligible species'. These constitute the primary contents of intellect, stored in *possible intellect*, intellect's essentially receptive aspect. 'Through intellect it is natural for us to have cognition of natures. Of course, [as universals] natures do not have existence except in individuating matter. It is natural for us to have cognition of them, however, not as they are in individuating matter but as they are abstracted from it by intellect's consideration', the work of agent intellect, producing intelligible species (*Summa theologiae* Ia.12.4c). The intelligible species of *O* are unlike sensory species of it in that they are only universals, which occur as such only in possible intellect: for example, round, metallic, iron hoop. These 'universal natures' are not only received in the intellective faculty *F*, the possible intellect, but are also of course used regularly as the devices indispensable for intellective cognition of corporeal reality: 'Our intellect both abstracts intelligible species from phantasms, insofar as it considers the natures of things universally, and yet also has intellective cognition of them [the things] in the phantasms, since without attending to phantasms it cannot have intellective cognition of even those things whose [intelligible] species it abstracts' (*Summa theologiae* Ia.85.1, ad 5). It is in this way that 'in intellection we can have cognition of such [particular, corporeal, composite] things in universality, which is beyond the faculty of sense' (*Summa theologiae* Ia.12.4c).

Thus both sense and intellect have cognition of *O*, a particular corporeal thing. However, sense has cognition of *O* only in its particularity (*Sententia super Posteriora analytica* II.20.14). Further, an individual intellect that happened to have the concept 'iron hoop' would have cognition only of a universal nature that happened to be instantiated in *O*, and not also of any instantiation of that nature - unless that intellect were also attending to phantasms of *O*. It is as a result of this attending that intellect also cognizes *O* itself, but as exemplifying a universal, for example, as an iron hoop (*Summa theologiae* Ia.85.5c; *Sententia super De anima* II.12.377).

Although intellect regularly has cognition of a corporeal particular in the way described, its proper object, Aquinas says, is that particular's universal nature, or 'quiddity'. Intellect's 'first operation', then, is its cognition of a universal, its proper object (although as we have seen, agent intellect's abstracting of intelligible species is a necessary step on the way to the cognition of the quiddities of things). Aquinas sometimes calls this first operation 'understanding'. However, *scientia*, which is one of the last operations of intellect, a pinnacle of intellective cognition, also has the natures of things as its objects (see below). Universal natures, the proper objects of intellect's first operation and the objects of the culminating theoretical knowledge of nature, must then be thought of as proper objects of both the beginning and the culmination of intellective cognition. What is cognized in an unanalysed way in the first operation of the intellect - for example, *animal* - is in scientific cognition analysed into the essential parts of its nature - *sensitive animate corporeality* - which are themselves comprehended in terms of all their characters and capacities. In theory, in potentiality, the culminating cognitive state is all that could be hoped for: 'if the human intellect comprehends the substance of any thing - a rock, for example, or a triangle - *none* of the intelligible aspects of that thing exceeds the capacity of human reason' (*Summa contra gentiles* I.3.16).

Intellect's 'second operation' is the making of judgments, affirming by propositionally 'compounding' with one another concepts acquired in the first operation, or denying by 'dividing' them from one another. At every stage past initial acquisition, the cognition of quiddities will partially depend on this second operation, and on reasoning as well: 'the human intellect does not immediately, in its first apprehension, acquire a complete cognition of the thing. Instead, it first apprehends *something* about it - that is, its quiddity, which is a first and proper object of intellect; and *then* it acquires intellective cognition of the properties, accidents, and dispositions associated with the thing's essence. In doing so it has to compound one apprehended aspect with another, or divide one from another, and proceed from one composition or division to another, which is reasoning.' This is sometimes called intellect's third operation (*Summa theologiae* Ia.85.5c).

The framing of propositions and the construction of inferences involving them are necessary preconditions of the culminating intellective cognition Aquinas recognizes as *scientia*, which he discusses in greatest detail in his *Sententia super Posteriora analytica (Commentary on Aristotle's Posterior Analytics)*. The interpretation of his account of *scientia* is controversial, but one helpful way to view it is as follows. To cognize a proposition with *scientia* is, strictly speaking, to accept it as the conclusion of a 'demonstration'. Of course, many premises in demonstrations may themselves be conclusions of other demonstrations; some, however, must be accepted not on the basis of demonstration but *per se* (*Sententia super Posteriora analytica* I.7.5-8). Such propositions, knowable *per se* (although not always *per se* knowable by us) are Aquinas' first principles. Like Aristotle, he calls them immediate propositions; that is, they cannot themselves be the conclusions of demonstrations, and their truth is

evident to anyone who fully understands their terms, who not merely grasps their ordinary meaning but also comprehends the real nature of their referents. The predicate of an immediate proposition belongs to the *ratio* of the proposition's subject, and the *ratio* is the formulation of the subject's real nature (*Sententia super Posteriora analytica* I.10; 33). Thus for example, Aquinas considers 'God exists' to be self-evident, since according to the doctrine of simplicity God's nature is God's existence. 'God exists' is a good example of a proposition knowable *per se* but, as Aquinas insists in rejecting Anselm's ontological argument, not knowable *per se* by us. It is for that reason that he develops a number of *a posteriori* arguments for God's existence, among which the most famous are the 'Five Ways', found in *Summa theologiae* Ia.2.3c (see God, arguments for the existence of).

Anyone who has a developed concept of the subject's real nature is certain of the truth of such an immediate proposition, 'but there are some immediate propositions the terms of which not everyone knows. That is why although the predicate of such a proposition does belong to the ratio of its subject, the proposition need not be granted by everyone, just because its subject's [metaphysical] definition is not known to everyone' (*Sententia super Posteriora analytica* I.5.7). Because proper demonstrations are isomorphic with metaphysical reality, the facts expressed in their premises are regularly to be construed as causes of the facts in their conclusions (*Sententia super Posteriora analytica* I.2.9), although in some cases demonstrative reasoning goes the other way, from effects to causes. So, having *scientia* with respect to some proposition is the fullest possible human cognition, by which one situates the fact expressed by a conclusion in an explanatory theory that accurately maps metaphysical or physical reality.

According to Aquinas, then, what demonstration provides is not so much knowledge as it has been conceived of by foundationalists (for example, Descartes) as it is depth of understanding and explanatory insight. In general, Aquinas does not begin with self-evident principles and derive conclusions from them deductively; 'rather [he begins] with a statement to be justified (it will become the "conclusion" only in a formal restatement of the argument) and "reduce[s]" it back to its ultimate explanatory principles' (Durbin 1968: 82). When Aquinas himself describes his project generally, he says that there are two different processes in which human reason engages: *discovery* (or invention) and *judgment*. When we engage in discovery, we proceed from first principles, reasoning from them to other things; in judgment we reason to first principles on the basis of a kind of analysis. In his view, it is judgment's reasoning process, not that of discovery, that leads to *scientia*, and judgment is the subject of the *Posterior Analytics*: 'Judgment goes with the certitude of *scientia*. And it is because we cannot have certain judgment about effects except by analysis leading to first principles that this part of human reasoning is called "analytics"' (*Sententia super Posteriora analytica*, prooemium).

Sceptical worries seldom intrude on Aquinas' scattered development of his systematically unified theory of knowledge, largely because it is based on a metaphysics in which the first principle of existence is an omniscient, omnipotent, perfectly good God, whose rational creatures could not have been made so as to be standardly mistaken about the rest of creation (see God, concepts of; Knowledge, concept of).

12 Will and action

Philosophy of mind is obviously relevant to epistemology in its account of the mechanisms of cognition, especially of intellect. In its account of will it is just as obviously relevant to action theory and to ethics. Aquinas' concern with moral issues is even greater than his considerable interest in epistemological issues, and his ethics is so fully developed that he integrates his systematic treatment of acts of will into it rather than including such a treatment in his philosophy of mind.

As intellect is the cognitive faculty of the distinctively human rational soul, so will is its appetitive faculty. Will's metaphysical provenance is more primitive than intellect's; it is merely the most subtle terrestrial instantiation of an utterly universal aspect of creation. Not only every sort of soul but absolutely every form, Aquinas maintains, has some sort of inclination essentially associated with it; and so every hylomorphic thing, even if inanimate, has at least one natural inclination: 'on the basis of its form, fire, for instance, is inclined toward a higher place, and toward generating its like' (*Summa theologiae* Ia.80.1c). Inclination is the genus of appetite, and appetite is the genus of will. The human soul of course involves *natural appetites* - for example, for food - but its sensory and intellective modes of cognition bring with them *sensory appetites*, or passions - for example, for seafood - and *rational appetite*, or volition - for example, for food low in fat content.

In human beings, sensory appetite, or ‘sensuality’, is a cluster of inclinations (passions) to which we are subject (passive) by animal nature. Following an Aristotelian line, Aquinas thinks of sensuality as sorted into two complementary powers: the *concupiscible* - pursuit/avoidance instincts - and the *irascible* - competition/aggression/defense instincts. With the former are associated the emotions of joy and sadness, love and hate, desire and repugnance; with the latter, daring and fear, hope and despair, anger.

For philosophy of mind and for ethics, one important issue is the manner and extent of the rational faculties’ control of sensuality, a control without which the harmony of the human soul is threatened and morality is impossible - especially in Aquinas’ reason-centered ethics with its focus on virtues and vices. A human being who is not aberrantly behaving like a non-rational animal ‘is not immediately moved in accordance with the irascible and concupiscible appetite but waits for the command of will, which is the higher appetite’ (*Summa theologiae* Ia.81.3c). But the kind of control exercised by a cognitive rational faculty (standardly identified in this role as ‘practical reason’ rather than the broader ‘intellect’) is less obvious, and is particularly interesting in view of Aquinas’ account of intellective cognition. The rational faculties can direct the attention of the external senses and compensate to some extent for their malfunctioning, but they cannot directly control what the external senses initially perceive on any occasion. On the other hand, sensuality and the internal senses are not directly related to mind-independent external things, and so to some extent ‘they are subject to reason’s command’, although they too can fight against reason (*Summa theologiae* Ia.81.3, ad 3). Elaborating an Aristotelian theme (*Politics* I, 2), Aquinas observes that the soul’s rule over the body is ‘despotic’: in a normal body, any bodily part that can be moved by an act of will will be moved immediately when and as will commands. But the rational faculties rule sensuality ‘politically’, because the powers and passions that are the intended subjects of this rational governance are also moved by imagination and sense, and so are no slaves to reason. ‘That is why we experience the irascible or the concupiscible fighting against reason when we sense or imagine something pleasant that reason forbids, or something unpleasant that reason commands’ (*Summa theologiae* Ia.81.3, ad 2).

According to Aquinas, the volition for happiness in general is an ineluctable part of human nature (see §13 below). Nonetheless, ‘the movement of a creature’s will is not determined in particular to seeking happiness in this, or in that’ (*Quaestiones disputatae de veritate* 24.7, ad 6). This sort of freedom of will is freedom of specification or ‘freedom as regards the object’, freedom in the ‘determining’ aspect of volition. It is distinguished from freedom of exercise or ‘freedom as regards the act’, freedom associated with will’s ‘executive’ capacity, for either acting or not acting to achieve something apprehended as good.

The interpretation of Aquinas’ account of freedom of will is controversial. The very phrase ‘freedom of will’ is part of the difficulty, because it imports a concept from a later tradition. Aquinas conceives of freedom as *liberum arbitrium* (free decision or judgment), which cannot be attributed to will alone. It is a property that inheres in the system of intellect and will as a whole, that emerges from their interaction. However, it is perhaps safe to say that, since Aquinas emphatically denies that any volition caused by something extrinsic to the agent can be free, his account of freedom of will is not a version of compatibilism (see for example *Summa theologiae* IaIIae.6.4). The one apparent exception has to do with God’s acting on a human will. Aquinas holds that among extrinsic forces, God alone can act directly on some other person’s will without violating the will’s nature, that is, without undermining its freedom (see for example *Summa theologiae* IaIIae.9.6). On this basis, some interpreters characterize Aquinas as a theological compatibilist; however, the subtle complexities of his account of God’s action on human wills leads others to claim that a full appreciation of those complexities would show that Aquinas is not in any sense a compatibilist (see Determinism and indeterminism; Free will).

Aquinas’ analysis of human action, built on his account of will and intellect, is complicated and not readily summarized. Generally speaking, he finds elaborately ordered mental components in even simple acts. For instance, in a case of raising one’s hand to attract attention we are likely to suppose that the mental antecedents of the bodily movement are just the agent’s combined beliefs and desires, whether or not the agent is fully conscious of them. Aquinas would of course agree that the agent need not be completely aware of the overt action’s mental antecedents, but he sees them as having a complex, hierarchical structure.

On his analysis, the action begins when (I1) the agent’s intellect apprehends a certain end - attracting attention - as a good to be achieved in these particular circumstances. (I1) thus gives rise to a second component: (W1) the agent’s will forms a simple volition for that end. Then, (I2) the agent’s intellect considers whether the end can be

achieved at that time. If the result of (I2) is affirmative, then on that basis (W2) the agent's will forms an intention to achieve the end by some means or other. Next, (I3) the agent's intellect surveys the available means and settles on one or more that would be suitable to achieve the end and acceptable to the agent, and (W3) the agent's will accepts the means. If intellect has found more than one suitable and acceptable means, then (I4) intellect compares them and determines which is best in the circumstances, and (W4) will opts for that means. The process comes to its natural end when (W5) the agent's will exercises its control over the agent's arm, and the arm goes up. This ordered series looks deterministic, but as Aquinas views the interaction between intellect and will, the process could go otherwise at almost any point because will could direct intellect to reconsider, to direct attention in some other way, or even just to stop thinking about the issue (*Summa theologiae* IaIIae.6-17).

13 Ethics, law and politics

Aquinas' moral theory is developed most extensively and systematically in the Second Part of *Summa theologiae*. (Broadly speaking, the general theory is in IaIIae and the detailed consideration of particular issues is in IIaIIae.) Like almost all his predecessors, medieval and ancient, Aquinas sees ethics as having two principal topics: first, the ultimate goal of human existence, and second, how that goal is to be won, or lost. Of the 303 Questions making up *Summa theologiae*'s Second Part, 298 are concerned in one way or another with the second topic, and only the first 5 are concerned directly with the first (although in *Summa contra gentiles* III he devotes chapters 25-40 to a detailed examination of it).

Summa theologiae IaIIae.1-5, sometimes called the Treatise on Happiness, develops an argument to establish the existence and nature of a single ultimate end for all human action, or, more strictly, the kind of behaviour over which a person has 'control'. First, 'all actions that proceed from a power are caused by that power in accordance with the nature of its object. But the object of will is an end and a good', that is, an end perceived as good by the willer's intellect (*Summa theologiae* IaIIae.1.1c). From this starting point Aquinas develops an argument designed to show that a human being necessarily (though not always consciously) seeks everything it seeks for its own ultimate end, happiness.

Aquinas argues that the often unrecognized genuine ultimate end for which human beings exist (their 'object') is God, perfect goodness personified; and perfect happiness, the ultimate end with which they may exist (their 'use' of that object), is the enjoyment of the end for which they exist. That enjoyment is fully achieved only in the beatific vision, which Aquinas conceives of as an activity. Since the beatific vision involves the contemplation of the ultimate (first) cause of everything, it is, whatever else it may be, also the perfection of all knowledge and understanding (*Summa theologiae* IaIIae.1.8; 3.8).

Aquinas devotes just four questions of *Summa theologiae* IaIIae (18-21) to 'the goodness and badness of human acts in general'. Although considerations of rightness and wrongness occupy only a little more than ten per cent of the discussion in Questions 18-21, Aquinas nonetheless appears to think of rightness and wrongness as the practical, distinctively moral evaluations of actions. His emphasis on the broader notions of goodness and badness reveals the root of his moral evaluation of actions in his metaphysical identification of being and goodness (see §9 above).

What makes an action morally bad is its moving the agent not toward, but away from, the agent's ultimate goal. Such a deviation is patently irrational, and Aquinas' analysis of the moral badness of human action identifies it as fundamentally irrationality, since irrationality is an obstacle to the actualization of a human being's specifying potentialities, those that make *rational* the differentia of the human species. In this as in every other respect, Aquinas' ethics is reason-centred:

> In connection with human acts the words 'good' and 'bad' are applied on the basis of a comparison to reason, because… a human being's good is existing in accordance with reason, while what is bad for a human being is whatever is contrary to reason. For what is good for any thing is what goes together with it in keeping with its form, and what is bad for it is whatever is contrary to the order associated with its form.
>
> (*Summa theologiae* IaIIae.18.5c)

It would be a mistake, however, to suppose that Aquinas takes moral evil to consist in intellective error. Because of the very close relationship he sees between intellect and will, the irrationality of moral wrongdoing will be a

function of will as well, not just of intellect. In Aquinas' view, the moral evaluation of a human action attaches primarily to the 'internal act', the volition from which the external act derives. Since 'will is inclined toward reason's good [the good presented to will by intellect] by the very nature of the power of will', bad volition stems from defective deliberation (*Summa theologiae* IaIIae.50.5, ad 3). As intellect and will continually influence each other, so bad deliberation can also be an effect of bad volition. Moreover, practical intellect's mistakes in identifying the best available course of action may also have the passions of the sensory soul as sources.

Furthermore, 'because the good [presented by intellect] is varied in many ways, it is necessary that will be inclined through some habit toward some determinate good presented by reason so that [will's determining] activity may follow more promptly' (*Summa theologiae* IaIIae.50.5, ad 3). Habits of will are conditions necessary for our carrying out our volitions in particularly good or particularly bad ways, as regards both the 'executive' and the 'determining' aspects of volition; and the habits that play these crucial roles in Aquinas' moral theory are the virtues and the vices.

The four 'cardinal virtues' can be understood as habits of this sort. Reason's habit of good governance generally is *prudence*; reason's restraint of self-serving concupiscence is *temperance*; reason's persevering despite self-serving 'irascible' passions such as fear is *courage*; reason's governance of one's relations with others despite one's tendencies toward selfishness is *justice*. Aquinas' normative ethics is based not on rules but on virtues; it is concerned with dispositions first and only then with actions. In addition to the moral virtues in all their various manifestations, Aquinas also recognizes intellectual virtues that, like the moral virtues, can be acquired by human effort. On the other hand, the supreme theological virtues of faith, hope and charity cannot be acquired but must be directly 'infused' by God. Aquinas introduces these virtues and others in *Summa theologiae* IaIIae 49-88 and examines them in detail throughout IIaIIae (see Virtue ethics).

Passions, virtues and vices are all intrinsic principles, or sources, of human acts. However, there are extrinsic principles as well, among which is law in all its varieties. Consequently, Aquinas moves on in *Summa theologiae* IaIIae.90-108 to his Treatise on Law, a famous and original treatment of the subject. The best-known feature of the treatise is Aquinas' concept of natural law. Law in general is 'a kind of rational ordering for the common good, promulgated by the one who takes care of the community' (*Summa theologiae* IaIIae.90.4c), and 'the precepts of natural law are to practical reasoning what the first principles of demonstrations are to theoretical reasoning…. All things to be done or to be avoided pertain to the precepts of natural law, which practical reasoning apprehends naturally as being human goods' (IaIIae.94.2c). Human laws of all kinds derive, or should derive, from natural law, which might be construed as the naturally knowable rational principles underlying morality in general: ' From the precepts of natural law, as from general, indemonstrable principles, it is necessary that human reason proceed to making more particular arrangements… [which] are called human laws, *provided that* they pertain to the definition (*rationem*) of law already stated' (IaIIae.91.3c). As a consequence of this hierarchy of laws, Aquinas unhesitatingly rejects some kinds and some particular instances of human law, for example: 'A tyrannical law, since it is not in accord with reason, is not unconditionally a law but is, rather, a perversion of law' (IaIIae.92.1, ad 4). Even natural law rests on the more fundamental 'eternal law', which Aquinas identifies as divine providence, 'the very nature of the governance of things on the part of God as ruler of the universe' (IaIIae.91.1c) (see Natural law).

In *De regimine principum (The Governance of Rulers)*, his most important political work, Aquinas begins by sounding the familiar medieval theme: monarchy is the best form of government. However, he realizes that a single ruler is easily corrupted and that monarchy therefore has a tendency to turn into tyranny. He seems not to countenance revolution against a legitimate ruler who has become tyrannical (*De regimine principium* 6), but he maintains that radical means, including tyrannicide, may be justified against a usurper. Perhaps because he came to appreciate the dangers in monarchy, he gradually works republican elements into his theory of good government. His later commentary on Aristotle's *Politics* seems to erode the dominant monarchical model further in its treatment of the notions of the commonwealth (*res publica*) and of the citizen as one who rules and is ruled in turn (see Political philosophy, history of).

14 Theology: natural, revealed and philosophical

Because Aquinas developed most of his thought within the formal confines of thirteenth-century theology, and because this has in turn affected his place in the history of philosophy and the assessment of his work, some

attention must be paid to the ways in which much of what we recognize as philosophy was an essential component of what he thought of as theology.

Aquinas devotes the first three books of *Summa contra gentiles* to a systematic development of natural theology, which he saw as part of philosophy (*Summa theologiae* Ia.1.1, ad 2) (see Natural theology). As part of philosophy, natural theology must of course be based entirely on 'principles known by the natural light of intellect' (*Summa theologiae* Ia.1.2c), principles of the sort that underlie Aristotle's metaphysics, which Aristotle himself thought of as culminating in 'theology' (see Aquinas' interpretation of that thought in the prooemium to his *Sententia super Metaphysicam (Commentary on Aristotle's Metaphysics)*. In fact, the way Aquinas works in *Summa contra gentiles* I-III strongly suggests that he may have thought of natural theology as a science subordinate to metaphysics, somewhat as he would have understood optics to be subordinate to geometry.

However, there is something odd about that project of his. By Aquinas' day the churchmen governing universities had overcome most of their initial misgivings about the recently recovered works of the pagan Aristotle, and had acknowledged officially that the study of Aristotelian physics and metaphysics (with their integrated minor component of natural theology) was compatible with the then universally recognized availability of revealed truths about God. Medieval Christians had come to appreciate the ancient philosophers' attempts to uncover truths about God on the basis of observation and reasoning alone as having been justified, even commendable, given their total ignorance of revelation. However, no philosopher in Aquinas' circumstances could have justifiably undertaken a new project of natural theology heuristically.

Still, no opprobrium would attach to natural theology taken up expositionally. The aim of such an enterprise would be not to develop theology from scratch but rather to show, in the spirit of Romans 1: 20, the extent to which what had been supernaturally revealed could, in theory, have been naturally discovered. Such an enterprise is what *Summa contra gentiles* I-III seems to represent.

Evidence from a chronicle written about seventy years after Aquinas began *Summa contra gentiles* once led scholars to suppose that he had written it as a manual for the use of Dominican missionaries to Muslims and Jews. If that were so, then the work's presentation of natural instead of revealed theology in its first three books would have been dictated by the practical purpose of rationally deriving the truth about God, and about God's relation to everything else, for people who would not have acknowledged the revealed texts Aquinas would otherwise have cited as the source of that truth. But nobody, and certainly not Aquinas, could have supposed that Muslims or Jews needed to be argued into perfect-being monotheism of the sort developed in those first three books, which contain nothing that he would have taken to be contrary to Judaism or Islam. If Aquinas had intended *Summa contra gentiles* as a manual for missionaries to educated Muslims, Jews or Christian heretics, he would have wasted the enormous effort represented in the 366 copiously argued chapters of Books I-III (see Gauthier 1961, 1993, for a persuasive rejection of the earlier account).

What Aquinas himself says about his purpose in writing *Summa contra gentiles* suggests that what he wrote had at least its formal cause not in an attempt to aid missionary activities, but instead in his consideration of the interrelation of philosophy and Christianity. He begins by writing about the concerns of a wise person, one of those 'who give things an appropriate order and direction and govern them well' (*Summa contra gentiles* I.1.2). Obviously, such a person has to be concerned with goals and sources, and so the wisest person will be 'one whose attention is turned toward the universal goal, which is also the universal source', which Aquinas takes to be God (I.1.3). Because this natural theology is oriented as it is, 'it must be called the greatest wisdom itself, as considering the absolutely highest cause of all' (II.4.874). Therefore, the highest, most universal explanatory truth must be wisdom's concern. Anyone aspiring to wisdom will attend to metaphysics, since, Aquinas reports, Aristotle rightly identified metaphysics as 'the science of truth - not of just any truth, but of the truth that is the origin of all truth, the truth that pertains to the first principle of being for all things' (I.1.5). And, as he says in an observation that suits his own enterprise, 'sometimes divine wisdom proceeds from human philosophy's starting points' (II.4.875). However, since it is the business of one and the same science 'to pursue one of two contraries and to repel the other… the role of the wise person is to meditate on the truth, especially the truth regarding the first principle, and to discuss it with others, but also to fight against the falsity that is its contrary' (I.1.6). The truth regarding the first principle will be the truth about God, supposing natural theology can show that God exists; and so the explanatory truth associated here with metaphysics is the truth associated also with theology.

No one knows what title, if any, Aquinas himself gave to this work. In some of its medieval manuscripts, it is entitled *Liber de veritate catholicae fidei contra errores (A Book About the Truth of the Catholic Faith, Directed Against Mistakes)*, a title that comes closer to accurately representing the book's aim and contents than the more pugnacious, traditional *Summa contra gentiles (Synopsis [of Christian Doctrine] Directed Against Unbelievers)*. During the nineteenth century, when *Summa theologiae (Synopsis of Theology)* was instead normally referred to as *Summa theologica (Theological Synopsis)*, *Summa contra gentiles* was sometimes published under the deliberately contrasting title *Summa philosophica (Philosophical Synopsis)*. That contrast, although potentially misleading, has some truth in it, as may be seen in Aquinas' plan for *Summa contra gentiles* I-III: 'Since we intend to pursue by way of reason the things about God that human reason can investigate, the first consideration is of matters associated with God considered in himself [Book I]; second, of the emergence of created things from him [Book II]; third, of the ordering and directing of created things toward him as their goal [Book III]' (I.9.57).

In this pursuit by way of reason, Aquinas must and does shun 'authoritative arguments' of any sort, but he shows good sense in not restricting himself to 'demonstrative arguments' in developing natural theology. He does, of course, use demonstrative arguments when he thinks he has them, but, like almost all philosophers of any period, he recognizes philosophy's need for 'probable aguments' as well. A demonstrative argument takes as its premises propositions that explain the fact in the argument's conclusion by elucidating its causes (or, sometimes, its effects), and so it produces, or presents, scientific understanding. A probable argument - the sort that has always been most prevalent and most appropriate in philosophy - is one based on premises of any sort that are accepted widely or by experts in the relevant field, and so one group may be convinced by a probable argument that another group rejects. Of course, Aquinas has to make use of authoritative arguments in the fourth (and last) book, where he turns from natural to revealed theology, and his tolerance of them there is part of what distinguishes Book IV's argumentation from the sort that characterizes Books I-III.

In *Summa contra gentiles* IV, Aquinas engages in what has come to be called *philosophical theology*, the application of reason to revelation. Philosophical theology shares the methods of natural theology broadly conceived - in other words, analysis and argumentation of all the sorts accepted in philosophy - but it lifts natural theology's restriction on premises, accepting as assumptions revealed propositions. This includes those that are initially inaccessible to unaided reason, such as the 'mysteries' of Christian doctrine. In his many works of philosophical theology, Aquinas tests the coherence of doctrinal propositions (including the mysteries), attempts explanations of them, uncovers their logical connections with other doctrinal propositions and so on, in order to bear out his conviction that the doctrines themselves are eminently understandable and acceptable, and that the apparent incoherence of some of them is only a feature of our initial, superficial view of them.

Summa theologiae is the paradigm of philosophical theology. The very first Article of the very first Question makes it clear at once that it is not natural theology that *Summa theologiae* is a summa of, since it begins by asking whether we need any 'other teaching, besides philosophical studies'; which in Aquinas' usage means the studies that medieval beginners in theology would have just completed in the arts faculty. The question arises because philosophical studies are characterized not only as dealing with 'the things that are subject to reason', but also as encompassing 'all beings, including God', as a consequence of which 'part of philosophy is called theology'.

Although Aquinas accepts this characterization of philosophy's subject matter as universal and as including a part that is properly called theology, he offers several arguments to support his claim that revealed theology is nonetheless not superfluous. In one of those arguments, he claims that a thing's 'capacity for being cognized in various ways brings about a difference between sciences'. By this he means that different sciences can reason to some of the same conclusions on the basis of different premises or evidence. In his example, he points out that in order to support the proposition that the earth is round a naturalist uses empirical observations, while a cosmologist might support that same conclusion on a strictly formal basis. 'And for that reason', he concludes, 'nothing prevents the same things from being treated by philosophical studies insofar as they can be cognized by the light of natural reason, and also by another science insofar as they are cognized by the light of divine revelation. That is why the theology that pertains to *sacra doctrina* [in other words, revealed theology] differs in kind from the theology that is considered a part of philosophy' (ad 2).

In this argument, Aquinas might appear willing to concede that revealed and natural theology differ only in this methodological respect, that they simply constitute two radically different ways of approaching the very same

propositions about God and everything else. However, he would not actually concede this. There are propositions that belong uniquely to revealed theology's subject matter, simply because the different premises with which revealed theology begins can also lead to conclusions not available to unaided reason. And, of course, no doctrinal proposition that is initially available to human beings only in virtue of having been revealed by God can be part of natural theology's subject matter.

On the other hand, no propositions appropriate to natural theology are excluded from *Summa theologiae*'s subject matter. The propositions that belong to natural theology form a proper subset of those that belong to revealed theology:

> It was necessary that human beings be instructed by divine revelation even as regards the things about God that human reason can explore. For the truth about God investigated by a few on the basis of reason [without relying on revelation] would emerge for people [only] after a long time and tainted with many mistakes. And yet all human well-being, which has to do with God, depends on the cognition of that truth. Therefore, it was necessary for human beings to be instructed about divine matters through divine revelation so that [the nature of human] well-being might emerge for people more conveniently and with greater certainty.
>
> (*Summa theologiae* Ia.1.1c)

When he sums up his examination of *sacra doctrina*, or revealed theology, Aquinas says that its 'main aim… is to transmit a cognition of God, and not only as he is in himself, but also as he is the source of things, and their goal - especially of the rational creature' (*Summa theologiae* Ia.2, intro.). Thus the subject matter of *sacra doctrina*, the theology presented in this summa of theology, is the most basic truths about *everything*, with two provisos: first, it is about God and about things other than God as they relate to God as their source and goal; second, among things other than God it deals with, it is especially about human beings, whose study of theology should be motivated by the fact that their well-being depends specially on their grasp of certain theological truths. And, Aquinas insists, universal scope is just what one should expect in a rational investigation of the truth about God: 'All things are considered in *sacra doctrina* under the concept of God, either because they *are* God, or because they have an ordered relationship *to* God as to their source and goal. It follows from this that the subject of this science is really God', even though the intended explanatory scope of the science is universal (*Summa theologiae* Ia.1.7c).

In referring to *sacra doctrina* as a 'science', Aquinas means to characterize it as a systematic, reasoned presentation of an organized body of knowledge consisting of general truths about some reasonably unified subject matter. In that broadly Aristotelian sense, it is not obviously wrong to think of theology as a science (as it would be in the narrower, twentieth-century sense of 'science'). It is in that sense that the science of theology as Aquinas develops it in *Summa theologiae* would now be called philosophical theology, the enterprise of employing the techniques and devices of philosophy in clarifying, supporting and extending the propositions that are supposed to have been revealed for theology's starting points. Thus, some of the work of philosophical theology is an attempt to explain revealed propositions and systematically work out their implications.

Like natural theology, which is subordinate to metaphysics, philosophical theology is a subordinate science. However, because it begins its work on divinely revealed propositions, Aquinas identifies the 'science' to which it is subordinate as God's knowledge of himself and everything else, available to human beings directly only in the afterlife (*Summa theologiae* Ia.1.2c). As he says earlier, 'For us, the goal of faith is to arrive at an understanding of what we believe - [which is] as if a practitioner of a subordinate science were to acquire in addition the knowledge possessed by a practitioner of the higher science. In that case the things that were only believed before would come to be known, or understood' (*Expositio super librum Boethii De trinitate* 2.2, ad.7).

Not even the doctrinal mysteries are impervious to rational investigation, although unaided reason could never have discovered them. Regarding one central mystery, for example, Aquinas says: 'It is impossible to arrive at a cognition of the Trinity of the divine persons by means of natural reason' (*Summa theologiae* Ia.32.1c). However, he says this in the twenty-second of a series of seventy-seven articles of *Summa theologiae* devoted to analysing and arguing about the details of Trinity, in other words, in the midst of subjecting this mystery to philosophical theology. As he explains in the very Article in which he rules out the possibility of rationally discovering that there are three divine persons:

> There are two ways in which reason is employed regarding any matter… in one way to provide sufficient proof

of something fundamental… in the other way to show that consequent effects are suited to something fundamental that has already been posited…. It is in the first way, then, that reason can be employed to prove that God is one, and things of that sort. But it is in the second way that reason is employed in a clarification of Trinity. For once Trinity has been posited, reasonings of that sort are suitable, although not so as to provide a sufficient proof of the Trinity of persons by those reasonings.

(*Summa theologiae* Ia.32.1c)

Aquinas is also careful to point out that it isn't mere intellectual curiosity or even a defense of the faith that is served by a rational clarification of Trinity. In his view, this application of philosophical theology - confirming faith by reason, showing that Trinity is not after all irrational, exposing the intricate connections between these and other doctrinal propositions - aids one's understanding of creation and salvation (see Trinity).

See also: Albert the Great; Aristotelianism, medieval; Augustinianism; Averroism; Duns Scotus, J.; Durandus of St Pourçain; Giles of Rome; God, concepts of; Godfrey of Fontaines; Hervaeus Natalis; John of Paris; Kilwardby, R.; Knowledge, concept of; Logic, medieval; Medieval philosophy; Natural philosophy, medieval (§7); Natural theology; Pecham, J.; Peter of Auvergne; Richard of Middleton; Siger of Brabant; Thomism

NORMAN KRETZMANN
ELEONORE STUMP

List of works

Not all of Aquinas' works exist in critical editions, but the many volumes of the Leonine Edition ordinarily provide the best available Latin texts. Most volumes of the Marietti Editions reproduce the Leonine text in handier form with useful aids to research. There are many translations into English and other modern languages, but by no means all the works have been translated. For detailed lists of editions and translations of each work see Torrell (1993), Weisheipl (1983) or Gilson (1956). Ingardia (1993) is an indispensable bibliography. Deferrari and Barry (1948) is an indispensable lexicon. Busa (1974-80) provides an exhaustive but somewhat unwieldy resource for research in Aquinas. The following is an approximately chronological list of works, excluding letters and liturgies. Places of composition are given after each work.

Aquinas, Thomas (1248-73) *Opera omnia (Complete Works)*, ed. Leonine Commission, *S. Thomae Aquinatis Doctoris Angelici. Opera Omnia. Iussu Leonis XIII, P.M. edita*, Rome: Vatican Polyglot Press, 1882-.(Many of the editions in this series are repeated in the Marietti Editions.)

Aquinas, Thomas (1248-52, or 1252-6) *De principiis naturae, ad fratrem Sylvestrum (On the Principles of Nature, for Brother Sylvester)*.(Written either at Cologne, 1248-52, or Paris, 1252-6).

Aquinas, Thomas (1251/2) *Expositio super Isaiam ad litteram (Literal Commentary on Isaiah)*.(Written at Cologne.)

Aquinas, Thomas (1251/2) *Postilla super Ieremiam (Commentary on Jeremiah)*.(Written at Cologne.)

Aquinas, Thomas (1251/2) *Postilla super Threnos (Commentary on Lamentations)*.(Written at Cologne.)

Aquinas, Thomas (1252/3? or 1273?) *Postilla super Psalmos (Commentary on Psalms)*. (Written either at Paris, 1252/3, or Naples, 1273. Incomplete, covers Psalms 1-54.)

Aquinas, Thomas (1252-6) *De ente et essentia, ad fratres et socios suos (On Being and Essence, For His Brothers and Companions)*.(Written at Paris.)

Aquinas, Thomas (1253-6) *Scriptum super libros Sententiarum (Commentary on the Sentences)*.(Written at Paris.)

Aquinas, Thomas (1256) *Principia: 'Hic est liber mandatorum Dei' et 'Rigans montes de superioribus suis' (Inaugural Lectures: 'This Is the Book of God's Commandments' and 'Watering the Hills from His Places Above')*.(Written at Paris.)

Aquinas, Thomas (1256) *Contra impugnantes Dei cultum et religionem (Against Those Who Assail the Worship of God and Religion)*.(Written at Paris, a refutation of William of Saint-Amour's *De periculis novissimorum temporum*.)

Aquinas, Thomas (1256-9) *Quaestiones disputatae de veritate (Disputed Questions on Truth)*.(Written at Paris.)

Aquinas, Thomas (1256-9) *Quaestiones quodlibetales VII-XI (Quodlibetal Questions VII-XII)*.(Written at Paris.)

Aquinas, Thomas (1257/8) *Expositio super librum Boethii De trinitate (Commentary on Boethius' De trinitate)*. (Written at Paris; incomplete.)

Aquinas, Thomas (1259?) *Expositio super librum Boethii De hebdomadibus (Commentary on Boethius' De hebdomadibus)*.(Written at Paris; incomplete.)

Aquinas, Thomas (1259-65) *Summa contra gentiles (Synopsis [of Christian Doctrine] Directed Against Unbelievers)*.(Written at Paris, Naples and Orvieto.)
Aquinas, Thomas (1261-5) *Expositio super Iob ad litteram (Literal Commentary on Job)*.(Written at Orvieto.)
Aquinas, Thomas (1261-5, or 1265-8) *Expositio super librum Dionysii De divinis nominibus (Commentary on Dionysius' De divinis nominibus)*.(Written at Orvieto, 1261-5, or Rome, 1265-8.)
Aquinas, Thomas (1262-8) *Glossa continua super Evangelia (Catena aurea) (A Continuous Gloss on the Four Gospels (The Golden Chain))*.(Written at Orvieto and Rome.)
Aquinas, Thomas (1263/4) *Contra errores Graecorum, ad Urbanem IV Pontificem Maximum (Against Mistakes of the Greek [Fathers of the Church], for Pope Urban IV)*.(Written at Orvieto, on an anonymous treatise, *De fide sanctae trinitatis contra errores Graecorum*.)
Aquinas, Thomas (1264) *De rationibus fidei contra Saracenos, Graecos, et Armenos, ad cantorem Antiochiae (On Arguments for the Faith Directed against Mohammedans, Greek Orthodox Christians and Armenians, for the Cantor of Antioch)*.
Aquinas, Thomas (1265-6) *Quaestiones disputatae de potentia (Disputed Questions on Power)*.(Written at Rome.)
Aquinas, Thomas (1265-6) *Quaestio disputata de anima (Disputed Question on the Soul)*. (Twenty-one articles, written at Rome.)
Aquinas, Thomas (1265-7) *Responsio ad fr. Ioannem Vercellensem de articulis 108 sumptis ex opere Petri de Tarentasia (Reply to Brother John of Vercelli Regarding 108 Articles Drawn from the Work of Peter of Tarentaise [on the Sentences])*.(Written at Rome.)
Aquinas, Thomas (1265-7) *Compendium theologiae, ad fratrem Reginaldum socium suum (A Compendium of Theology, for Brother Reginald, his Companion)*. (Written at Rome, incomplete.)
Aquinas, Thomas (1266-72) *Quaestiones disputatae de malo (Disputed Questions on Evil)*.(Written at Rome and Paris.)
Aquinas, Thomas (1266-68) *Summa theologiae Ia (Synopsis of Theology, First Part)*.(Written at Rome. See below for IaIIae, IIaIIae and IIIa. The whole of the *Summa theologiae* can be found in the Leonine edition, vols 4-12.)
Aquinas, Thomas (1267) *De regno [or De regimine principum], ad regem Cypri (On Kingship [or On the Governance of Rulers], for the King of Cyprus)*. (Written at Rome. Authentic only through Book II, c. 4.)
Aquinas, Thomas (1267-8) *Sententia super De anima (Commentary on Aristotle's On the Soul)*.(Written at Rome.)
Aquinas, Thomas (1267-8) *Quaestio disputata de spiritualibus creaturis (Disputed Question on Spiritual Creatures [Angels])*.(Eleven articles, written at Rome.)
Aquinas, Thomas (1268-9) *Sententia super Physicam (Commentary on Aristotle's Physics)*.(Written at Paris.)
Aquinas, Thomas (1268-9) *Sententia super Meteora (Commentary on Aristotle's Meteora)*. (Written at Paris, incomplete.)
Aquinas, Thomas (1268-70) *Sententia super De sensu et sensato (Commentary on Aristotle's De sensu et sensato)*.(Written at Paris.)
Aquinas, Thomas (1268-71) *Summa theologiae IaIIae (Synopsis of Theology, First Part of the Second Part)*. (Written at Rome and Paris.)
Aquinas, Thomas (1268-72) *Quaestiones quodlibetales I-VI, XII (Quodlibetal Questions I-VI, XII)*.(Written at Paris.)
Aquinas, Thomas (1269) *De forma absolutionis sacramentalis, ad generalem magistrum Ordinis (On the Form of Sacramental Absolution, for the Master General of the Order [John Vercelli])*.(Written at Paris.)
Aquinas, Thomas (1269) *De secreto (On Secret Testimony)*.(Written at Paris. A committee report in which Aquinas is the lone dissenter, supporting the right of a religious superior to compel a subject to reveal a secret even under the seal of confession.)
Aquinas, Thomas (1269-70) *Lectura super Matthaeum (Lectures on the Gospel of Matthew)*.(Written at Paris.)
Aquinas, Thomas (1269-70) *De perfectione spiritualis vitae (On the Perfecting of the Spiritual Life)*.(Written at Paris, directed against Gérard d'Abbeville's *Contra adversarium perfectionis christianae*.)
Aquinas, Thomas (1269-72) *Sententia libri Politicorum (Commentary on Aristotle's Politics)*. (Written probably at Paris; incomplete.)
Aquinas, Thomas (1270?) *Sententia super De memoria et reminiscentia (Commentary on Aristotle's De memoria et reminiscentia)*.(Written probably at Paris.)
Aquinas, Thomas (1270) *Tabula libri Ethicorum (An Analytical Table of Aristotle's Ethics)*. (Written at Paris, incomplete.)

Aquinas, Thomas (1270) *De unitate intellectus, contra Averroistas (On the Unicity of Intellect, Against the Averroists)*.(Written at Paris.)
Aquinas, Thomas (1270-1) *Sententia super Peri hermenias (Commentary on Aristotle's De interpretatione)*.(Written at Paris, incomplete.)
Aquinas, Thomas (1270-2) *Lectura super Ioannem (Lectures on the Gospel of John)*.(Written at Paris.)
Aquinas, Thomas (1270-3) *Sententia super Metaphysicam (Commentary on Aristotle's Metaphysics)*.(Written at Paris and Naples.)
Aquinas, Thomas (1270-3) *Expositio et lectura super Epistolas Pauli Apostoli (Commentary and Lectures on the Epistles of Paul the Apostle)*.(Written at Paris and Naples.)
Aquinas, Thomas (1271) *De aeternitate mundi, contra murmurantes (On the Eternity of the World, Against Grumblers)*.(Written at Paris.)
Aquinas, Thomas (1271) *Responsio ad magistrum Ioannem de Vercellis de articulis 42 (Reply to Master John Vercelli Regarding Forty-Two Articles)*.(Written at Paris. Aquinas' answers to doctrinal questions which Vercelli submitted also to Albert the Great and Robert Kilwardby.)
Aquinas, Thomas (1271-2) *Summa theologiae IIaIIae (Synopsis of Theology, Second Part of the Second Part)*.(Written at Paris.)
Aquinas, Thomas (1271-2) *Contra doctrinam retrahentium a religione (Against the Teaching of Those Who Dissuade [Boys] from Entering the Religious Life)*.(Written at Paris, opposing the work of Gérard d'Abbeville.)
Aquinas, Thomas (1271-2) *Sententia libri Ethicorum (Commentary on Aristotle's Nicomachean Ethics)*.(Written at Paris.)
Aquinas, Thomas (1271-2) *Sententia super Posteriora analytica (Commentary on Aristotle's Posterior Analytics)*.(Written at Paris.)
Aquinas, Thomas (1271-2) *Quaestio disputata de virtutibus in communi (Disputed Question on the Virtues in General)*.(Thirteen articles, written at Paris.)
Aquinas, Thomas (1271-2) *Quaestio disputata de caritate (Disputed Question on Charity)*. (Thirteen articles, written at Paris.)
Aquinas, Thomas (1271-2) *Quaestio disputata de correctione fraterna (Disputed Question on Fraternal Correction)*.(Two articles, written at Paris.)
Aquinas, Thomas (1271-2) *Quaestio disputata de spe (Disputed Question on Hope)*.(Four articles, written at Paris.)
Aquinas, Thomas (1271-2) *Quaestio disputata de virtutibus cardinalibus (Disputed Question on the Cardinal Virtues)*.(Four articles, written at Paris.)
Aquinas, Thomas (1271-3) *De substantiis separatis, ad fratrem Reginaldum socium suum (On Separated Substances [Angels], for Brother Reginald, his Companion)*.(Written at Paris or Naples, incomplete.)
Aquinas, Thomas (1272) *Quaestio disputata de unione verbi incarnati (Disputed Question on the Unity of the Incarnate Word)*.(Five articles, written at Paris.)
Aquinas, Thomas (1272) *Expositio super librum De causis (Commentary on the Liber de causis)*.(Written at Paris.)
Aquinas, Thomas (1272-3) *Summa theologiae IIIa (Synopsis of Theology, Third Part)*.(Written at Paris and Naples, incomplete.)
Aquinas, Thomas (1272-3) *Sententia super libros De caelo et mundo (Commentary on Aristotle's On Heaven and Earth)*.(Written at Naples, incomplete.)
Aquinas, Thomas (1272-3) *Sententia super libros De generatione et corruptione (Commentary on Aristotle's On Generation and Corruption)*.(Written at Naples, incomplete.)
Aquinas, Thomas (1273, or 1261-8?) *Collationes in decem praecepta (Sermon Commentaries on the Ten Commandments)*.(Written at Naples, 1273, or possibly at Orvieto and Rome, 1261-8.)
Aquinas, Thomas (1273, or 1268-72?) *Collationes super Ave Maria (Sermon Commentaries on the Ave Maria)*.(Written at Naples, 1273, or possibly at Paris, 1268-72.)
Aquinas, Thomas (1273) *Collationes super Credo in Deum (Sermon Commentaries on the Apostles' Creed)*.(Written at Naples, 1273.)
Aquinas, Thomas (1273) *Collationes super Pater Noster (Sermon Commentaries on the Lord's Prayer)*.(Written at Naples, 1273.)

References and further reading

Aertsen, J.A. (1988) *Nature and Creature: Thomas Aquinas' Way of Thought*, Leiden: Brill.(Novel approaches to establishing the inner coherence and identifying the direction of Aquinas' philosophy and theology.)

Aertsen, J.A. (1993) ‘Aquinas’ Philosophy in its Historical Setting’, in N. Kretzmann and E. Stump (eds) *The Cambridge Companion to Aquinas*, Cambridge: Cambridge University Press, 12-37.(A concise presentation of the development of Aquinas’ thought.)

Ashworth, E.J. (1991) ‘Signification and Modes of Signifying in Thirteenth-Century Logic: A Preface to Aquinas on Analogy’, *Medieval Philosophy and Theology* 1: 39-67.(The best introduction to the topic and the recent literature.)

Bigongiari, D. (ed.) (1953) *The Political Ideas of St. Thomas Aquinas*, New York: Hafner.(A useful survey.)

Brady, I. (1974) ‘John Pecham and the Background of Aquinas’ *De Aeternitate Mundi*’, in A.A. Maurer (ed.) *St. Thomas Aquinas 1274-1974: Commemorative Studies*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, vol. II, 141-78.(A well-informed, persuasive presentation of the treatise’s Augustinian background.)

Burrell, D.B. (1986) *Knowing the Unknowable God: Ibn Sina, Maimonides, Aquinas*, Notre Dame, IN: University of Notre Dame Press.(A knowledgeable exploration of Aquinas’ intellectual relations with Islamic and Jewish thinkers.)

Burrell, D.B. (1993) ‘Aquinas and Islamic and Jewish Thinkers’, in N. Kretzmann and E. Stump (eds) *The Cambridge Companion to Aquinas*, Cambridge: Cambridge University Press, 60-84.(An updated synopsis of material in Burrell (1986).)

Busa, R. (1974-80) *Index Thomisticus*, Stuttgart-Bad Cannstatt: Fromann-Holzboog.(Incomparably the most complete, multi-faceted research tool devised for the works of any medieval philosopher, accompanied by a complete but not readily usable edition of Aquinas’ works.)

Chenu, M.D. (1964) *Toward Understanding St Thomas*, Chicago, IL: Regnery.(A classic interpretation.)

Copleston, F.C. (1955) *Aquinas*, Baltimore, MD: Penguin.(A philosophically and historically well-informed complete study; still useful.)

Davies, B. (1992) *The Thought of Thomas Aquinas*, Oxford: Clarendon Press.(An impressive attempt to present a complete picture of Aquinas’ thought, following the schema of *Summa theologiae*.)

Deferrari, R.J. and Barry, Sister M. Inviolata (eds) (1948) *A Lexicon of St. Thomas Aquinas*, Washington, DC: Catholic University of America Press.(Still useful, and handier than Busa.)

Doig, J. (1972) *Aquinas on Metaphysics: A Historico-Doctrinal Study of the Commentary on the Metaphysics*, The Hague: Martinus Nijhoff.(Historically useful.)

Donagan, A. (1977) *The Theory of Morality*, Chicago, IL: University of Chicago Press.(A philosophically sophisticated development of Aquinas’ moral theory.)

Donagan, A. (1982) ‘Thomas Aquinas on Human Action’, in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 642-54.(A good short analysis of Aquinas’ complex action theory.)

Durbin, P.T. (1968) *St Thomas Aquinas: Summa Theologiae*, Blackfriars edition and translation, vol. XII, New York: McGraw-Hill.(Useful notes and appendices on Aquinas’ theory of the nature and activity of intellect.)

Fabro, C. (1961) *Participation et causalité selon Saint Thomas d’Aquin (Participation and Causality in the Work of Saint Thomas Aquinas)*, Louvain: Publications Universitaires de Louvain.(An influential study bringing out Platonist elements in Aquinas’ metaphysics.)

Finnis, J. (1980) *Natural Law and Natural Rights*, New York: Oxford University Press.(A magisterial presentation of these notions and their place in Aquinas’ thought.)

Foster, K. (1959) *The Life of Saint Thomas Aquinas, Biographical Documents*, London: Longmans, Green.(A translation of the oldest biographies and other documents used in the canonization proceedings, 1317-23.)

Gauthier, R.-A. (1961) ‘Introduction’, in *Saint Thomas d’Aquin: Contra Gentiles, Livre premier*, Lyon: P. Lethielleux.(A ground-breaking study of the background and purposes of *Summa contra gentiles*, persuasively rejecting the ‘missionary’ account.)

Gauthier, R.-A. (ed.) (1984) *Sancti Thomae de Aquino: Opera Omnia iussu Leonis XIII P.M. edita, Tomus XLV, 1: Sentencia libri de anima*, Rome: Commissio Leonina.(A particularly impressive volume of the Leonine edition, with a very thorough scholarly introduction based on original research.)

Gauthier, R.-A. (1993) *Saint Thomas d’Aquin: Somme contre les gentils. Introduction*, Collection Philosophie Européene, dirigée par H. Hude, Paris: Éditions universitaires.(A magisterial survey of historical and textual evidence.)

Gilby, T. (1955) *The Political Thought of Thomas Aquinas*, Chicago, IL: University of Chicago Press.(A general introduction that is still useful.)

Gilson, E. (1956) *The Christian Philosophy of St. Thomas Aquinas*, trans. L.K. Shook, New York: Random House. (A classic; originally published in French in 1948, but still worth consulting.)

Henle, R.J. (1956) *Saint Thomas and Platonism*, The Hague: Martinus Nijhoff.(One of the best known studies of

neo-Platonist influences on Aquinas.)
Ingardia, R. (1993) *Thomas Aquinas: International Bibliography 1977-1990*, Bowling Green, KY: The Philosophy Documentation Center.(A well-organized, well-indexed list in which many entries are accompanied by synopses.)
Irwin, T. H. (1992) 'Who Discovered the Will?', in J.E. Tomberlin (ed.) *Philosophical Perspectives 6: Ethics*, Atascadero, CA: Ridgeview.(Uses Aquinas' interpretation in a convincing argument for attributing the concept of will to Aristotle.)
Jaffa, H.V. (1952) *Thomism and Aristotelianism*, Chicago, IL: University of Chicago Press.(A controversial study of Aquinas' Commentary on the *Nicomachean Ethics*.)
Jordan, M.D. (1986) *Ordering Wisdom: The Hierarchy of Philosophical Discourses in Aquinas*, Notre Dame, IN: University of Notre Dame Press.(A study of Aquinas' thought in terms of the 'discourses' of physics, metaphysics, and theology; good bibliography.)
Jordan, M.D. (1993) 'Theology and Philosophy', in N. Kretzmann and E. Stump (eds) *The Cambridge Companion to Aquinas*, Cambridge: Cambridge University Press, 232-51. (A well-argued presentation of the sort of interpretation that identifies Aquinas as a theologian rather than a philosopher.)
Kenny, A. (1980a) *Aquinas*, New York: Hill & Wang.(A philosophically sophisticated, critical introductory study, focusing on metaphysics and philosophy of mind.)
Kenny, A. (1980b) *The Five Ways: St Thomas Aquinas' Proofs of God's Existence*, Notre Dame, IN: University of Notre Dame Press.(An untraditional critical analysis.)
Kenny, A. (1993) *Aquinas on Mind*, London: Routledge.(Stimulating studies, occasionally too ready to reject parts of the account that need further investigation.)
Kenny, A. (ed.) (1969) *Aquinas: A Collection of Critical Essays*, Garden City, NY: Doubleday.(A somewhat dated but still stimulating set of papers on all aspects of Aquinas' philosophy.)
Kremer, K. (1971) *Die neuplatonische Seinsphilosophie und ihre Wirkung auf Thomas von Aquin (Neoplatonist Metaphysics and Their Effect on Thomas Aquinas)*, Leiden: Brill.(A study of Neoplatonist elements in Aquinas' metaphysics.)
Kretzmann, N. (1988) '*Lex iniusta non est lex*: Laws on Trial in Aquinas' Court of Conscience', *The American Journal of Jurisprudence* 33: 99-122.(The moral assessment of law in Aquinas' theory.)
Kretzmann, N. (1992) 'Infallibility, Error, and Ignorance', *Canadian Journal of Philosophy*, supp. vol. 17: 159-94.(An attempt to explain puzzling features of Aquinas' account of intellective cognition.)
Kretzmann, N. (1993) 'Philosophy of Mind', in N. Kretzmann and E. Stump (eds) *The Cambridge Companion to Aquinas*, Cambridge: Cambridge University Press, 128-59. (A concise critical account.)
Kretzmann, N. (1997) *The Metaphysics of Theism: Aquinas's Natural Theology in Summa Contra Gentiles I*, Oxford: Clarendon Press.(An examination of the philosophical account of 'God considered in himself'.)
Kretzmann, N. and Stump, E. (eds) (1993) *The Cambridge Companion to Aquinas*, Cambridge: Cambridge University Press.(Ten studies specifically designed to introduce all the important aspects of Aquinas' thought; includes bibliography.)
Lonergan, B. (1946) *Grace and Freedom: Operative Grace in the Thought of St. Thomas Aquinas*, ed. J.P. Burns, New York: Herder & Herder.(A sophisticated, learned account, still worthwhile.)
Lonergan, B. (1967) *Verbum: Word and Idea in Aquinas*, ed. D.B. Burrell, Notre Dame, IN: University of Notre Dame Press.(An original treatment of the metaphysical side of Aquinas' theory of cognition.)
MacDonald, S. (1984) 'The *Esse/Essentia* Argument in Aquinas' *De ente et essentia*', *Journal of the History of Philosophy* 22: 157-72.(A well-argued assessment and resolution of the crucial difficulty in *De ente*.)
MacDonald, S. (ed.) (1991) *Being and Goodness: The Concept of the Good in Metaphysics and Philosophical Theology*, Ithaca, NY: Cornell University Press.(Twelve essays, most of them not published elsewhere, on the most important relationship among the 'transcendentals'.)
MacDonald, S. (1991) 'Ultimate Ends in Practical Reasoning: Aquinas' Aristotelian Moral Psychology and Anscombe's Fallacy', *The Philosophical Review* 100: 31-66.(A careful analysis and defense of the central argument in Aquinas' 'Treatise on Happiness'.)
MacDonald, S. (1993) 'Theory of Knowledge', in N. Kretzmann and E. Stump (eds) *The Cambridge Companion to Aquinas*, Cambridge: Cambridge University Press, 160-95. (An examination of all aspects of the theory of cognition, but especially the notion of *scientia*.)
Maritain, J. (1951) *Man and the State*, Chicago, IL: University of Chicago Press.(A classic application of Aquinas' political theory.)
McInerny, R. (1990) *Boethius and Aquinas*, Washington, DC: Catholic University of America Press.(The best available study of the important influence of Boethius on Aquinas.)

McInerny, R. (1992) *Aquinas on Human Action*, Washington: Catholic University of America Press.(The fullest study of Aquinas' action theory.)

McInerny, R. (1993) 'Ethics', in N. Kretzmann and E. Stump (eds) *The Cambridge Companion to Aquinas*, Cambridge: Cambridge University Press, 196-216.(An engaging survey of the elements of Aquinas' moral theory.)

Nussbaum, M.C. (1992) 'The Text of Aristotle's *De anima*', in M.C. Nussbaum and A.O. Rorty (eds) *Essays on Aristotle's De anima*, Oxford: Clarendon Press, 1-6. (Appreciation of Aquinas as an Aristotelian interpreter is also found elsewhere in this volume of essays by various authors.)

Owens, J. (1980) *St. Thomas Aquinas on the Existence of God: Collected papers of Joseph Owens*, ed. J. Catan, Albany, NY: State University of New York Press.(Eleven important articles by a renowned student of Aquinas, the first five on topics other than the existence of God.)

Owens, J. (1986) 'Aquinas' Distinction at *De ente et essentia* 4.119-123', *Mediaeval Studies* 48: 264-87.(An important contribution to the debate represented in MacDonald (1984).)

Owens, J. (1993) 'Aristotle and Aquinas', in N. Kretzmann and E. Stump (eds) *The Cambridge Companion to Aquinas*, Cambridge: Cambridge University Press, 38-59. (A survey, distilling a lifetime's study of this intellectual relationship.)

Pegis, A.C. (1934) *St Thomas and the Problem of the Soul in the Thirteenth Century*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(A helpful introduction to Aquinas' philosophical psychology in its historical setting.)

Ross, J.F. (1985) 'Aquinas on Belief and Knowledge', in G. Etzkorn (ed.) *Essays Honoring Allan B. Wolter*, St Bonaventure, NY: Franciscan Institute.(A philosophically sophisticated, somewhat idiosyncratic interpretation.)

Sigmund, P.E. (1993) 'Law and Politics', in N. Kretzmann and E. Stump (eds) *The Cambridge Companion to Aquinas, Cambridge: Cambridge University Press*, 217-31. (A survey bringing out connections to twentieth-century institutions and events.)

Stump, E. (1990) 'Intellect, Will and the Principle of Alternate Possibilities', in M. Beaty (ed.) *Christian Theism and the Problems of Philosophy*, Notre Dame, IN: University of Notre Dame Press.(An attempt to show that Aquinas' libertarian account of human freedom does not need the principle.)

Stump, E. (1992) 'Aquinas on the Foundations of Knowledge', *Canadian Journal of Philosophy*, supp. vol. 17: 125-58.(Rejects the traditional account of Aquinas as a foundationalist.)

Stump, E. (1993) 'Aquinas on the Sufferings of Job', in E. Stump (ed.) *Reasoned Faith*, Ithaca, NY: Cornell University Press, 328-57.(Aquinas' account of providence and the problem of evil as shown in biblical commentary.)

Stump, E. (1993) 'Biblical Commentary and Philosophy', in N. Kretzmann and E. Stump (eds) *The Cambridge Companion to Aquinas*, Cambridge: Cambridge University Press, 252-68.(Shows the philosophical significance of Aquinas' biblical commentaries.)

Stump, E. and Kretzmann, N. (1988) 'Being and Goodness', in T.V. Morris (ed.) *Divine and Human Action*, Ithaca, NY: Cornell University Press, 281-312; reprinted in MacDonald (1991).(An analysis of the interrelation of these transcendentals in Aquinas, especially in his metaethics.)

Torrell, J.-P. (1993) *Initiation à Saint Thomas d'Aquin: Sa Personne et Son Oeuvre (Introduction to Saint Thomas Aquinas: His Character and His Work)*, Fribourg and Paris: Éditions Universitaires and Éditions du Cerf.(Supersedes Weisheipl and Tugwell on the life and works, offering persuasive evidence for new hypotheses.)

Tugwell, S. (1988) *Albert and Thomas: Selected Writings*, The Classics of Western Spirituality, Mahwah, NJ: Paulist Press.(In the scholarly but eminently readable introduction to the Aquinas translations, Tugwell supersedes Weisheipl on the life and works.)

Van Steenberghen, F. (1980) *Le problème de l'existence de Dieu dans les écrits de S. Thomas d'Aquin*, Louvain-la-Neuve: Editions de l'Institut Superieur de Philosophie. (The most complete critical study of Aquinas' many arguments for the existence of God.)

Van Steenberghen, F. (1980) *Thomas Aquinas and Radical Aristotelianism*, Washington, DC: Catholic University of America Press.(An analysis of the role played by Latin Averroism in Aquinas' second Paris regency.)

Vollert, C., Kendzierski, L. and Byrne, P. (eds) (1984) *On the Eternity of the World*, Milwaukee, WI: Marquette University Press.(A collection of medieval texts in translation, including Aquinas, with a helpful introductory study.)

Weisheipl, J. (1983) *Friar Thomas D'Aquino: His Life, Thought, and Works*, Washington, DC: Catholic

University of America Press.(A revised edition of the classic biography, updated in some respects; superseded by Tugwell (1988) and Torrell (1993).)

Westberg, D. (1994) *Right Practical Reason: Aristotle, Action, and Prudence in Aquinas*, Oxford: Clarendon Press.(A wide-ranging historical and theological discussion of practical reason in Aquinas' thought.)

Wippel, J.F. (1984) *Metaphysical Themes in Thomas Aquinas*, Washington, DC: Catholic University of America Press.(A valuable collection of the author's articles on Aquinas' metaphysics, with some attention to Henry of Ghent; excellent bibliography.)

Wippel, J.F. (1987) 'Thomas Aquinas and Participation', in J.F. Wippel (ed.) *Studies in Medieval Philosophy*, Washington, DC: Catholic University of America Press.(The best discussion of this important, difficult topic.)

Wippel, J.F. (1989) 'Truth in Thomas Aquinas (part I)', *Review of Metaphysics* 43 (2): 295-326.(The first of a pair of articles, which together constitute a very well-informed study of the topic.)

Wippel, J.F. (1990) 'Truth in Thomas Aquinas (part II)', *Review of Metaphysics* 43 (3): 543-567.(The second of the pair of articles referred to above.)

Wippel, J.F. (1993) 'Metaphysics', in N. Kretzmann and E. Stump (eds) *The Cambridge Companion to Aquinas*, Cambridge: Cambridge University Press, 85-127.(A magisterial survey, including a discussion of the close connection between philosophy and theology in Aquinas' thought.)

Zimmermann, A. (1965) *Ontologie oder Metaphysik? Die Diskussion über den Gegenstand der Metaphysik im 13. und 14. Jahrhundert (Ontology or Methaphysics? The Discussion of the Object of Metaphysics in the Thirteenth and Fourteenth Centuries)*, Leiden: Brill.(A useful historical introduction.)

Arama, Isaac ben Moses (*c*.1420-94)

Like many of his fifteenth-century Spanish contemporaries, Arama opposed the Aristotelianism of Maimonides. His philosophical sermons and biblical commentaries attack Jewish Aristotelians on charges of subordinating revelation to reasoning, upholding an eternal universe whose necessity limits God's power, and excluding miracles and individual providence. Yet while stressing the fallibility of human reason, Arama is no fideist. An eclectic, he values reason and philosophy as ways of deepening the understanding of Scripture through allegorical interpretation. He also develops striking philosophical theories of miracles, providence and the fundamentals of faith.

A leading rabbi of the generation that underwent expulsion from Spain in 1492, Isaac ben Moses Arama is best known for the philosophical sermons he composed to counter the conversionist Christian sermons which his congregants were compelled to attend. Preserved in the chapters of his popular and influential biblical commentary, *'Aqedat Yitzhaq (The Binding of Isaac)*, each of these discourses has two parts, one addressing a philosophical or theological problem, the other, like the scholastic *quaestiones* and *dubitationes*, raising issues in the exegesis of a biblical passage, which is then interpreted in light of the earlier philosophical account.

Arama is most influenced by Halevi and Crescas among Jewish thinkers; and, among the Muslims, by al-Ghazali. The essential beliefs of Judaism, he argues, transcend or contradict reason and must be accepted on faith. Even beliefs that lie within the scope of reason should be accepted on faith and only then verified by reason. Just as Abraham overcomes his reason when he determines to sacrifice his son, so the ideal religious person must seek out irresolvable paradoxes of reason, learning to subordinate intellect to faith, philosophy to revelation.

Arama harshly criticizes fourteenth-century Maimonists like Moses of Narbonne (Narboni) and Joseph ibn Caspi (see Averroism, Jewish §§3-4), charging that they favour philosophy over revelation. He is especially sensitive to their use of allegorical exegesis. He himself did much to revive and develop the allegorical method, which had fallen into decline after the 'Maimonidean controversy'; but he also defends the validity of the literal meaning, which is enriched by philosophical interpretation. The Maimonists, by contrast, are charged with seeking to replace the literal sense with the Greek philosophical ideas which they claim to discover as the true meaning of Scripture.

Arama's position in the fifteenth-century debate over the foundational beliefs of Judaism shows the subtlety of his approach to the relations of philosophy and revelation. His articles of faith are the principles essential not just to the idea of a divine law generically but to Judaism uniquely. Connecting these principles with the Mosaic commandments, he argues with Halevi and against Maimonides that one achieves immortality not through intellectual perfection but by performance of the commandments, the one feature that truly distinguishes Judaism from other religions and from the generic theism of a philosopher. The commandments are not themselves articles of faith. Rather, Arama identifies the true principles as those that are embodied in and derived from specific commandments, as belief in creation is embedded in the Sabbath laws.

Creation is the most important of Arama's principles. All the rest are derivable from it. Like Maimonides, he locates the problem of the origin of universe beyond rational demonstration. But creation for him is typical in this respect of all the essential truths of Judaism. Nonetheless, perhaps influenced by Nahmanides, he offers a quasi-Platonic account of the world's origin from pre-existent matter, itself created *ex nihilo* by God. Essential here is the commitment to God's unconstrained will: only if the world was created freely can there be providence, miracles, or divine omnipotence. As evidence against eternity, Arama appeals not only to the diverse and irregular motions of the spheres (as Maimonides did) but also to the occurrence of miracles as, attested in Scripture.

Despite the centrality of miracles and omnipotence in his scheme, Arama does not wish to reject the idea of natural laws. He therefore distinguishes between two coexisting natures: an unintelligent 'natural nature', whose laws apply without regard to human merit, and a 'supernatural nature' directed purposefully by an intelligence sensitive to human deserts. Miracles violate natural nature; but, as instances of supernatural nature, they are fully natural. Besides the observed violations of natural nature that we call miracles, Arama (recalling Nahmanides) argues that 'hidden' miracles occur constantly. Their miraculousness usually goes unnoticed, but their true explanation lies in

supernatural nature. All miracles therefore fall under *a* nature. Analogously, there is a ’natural’ governance that does not attend to the moral perfection of individuals but only to preservation of species, and a second, ‘providential’, governance that is sensitive to the moral perfection of individuals. In general the latter is limited to Israel.

Philosophers deny supernatural nature and providential governance, Arama reasons, only because they rank humanity beneath the celestial spheres and cannot understand why God would violate a higher order for a lower one. Following Saadiah’s premise that humankind is the final act of creation (see Saadiah Gaon), Arama argues that the spheres themselves were created for mankind’s sake and can therefore be manipulated in our behalf. Part of the original creation, moreover, gave man a special power over nature. When the soul rules, establishing the ’divine image’ in the microcosm of the human body, harmony results in the macrocosm as well; when sin disrupts the microcosmic harmony, the macrocosmic harmony is ruptured too, causing universal evils. Human beings can ensure cosmic harmony only by living in harmony with nature - following the natural law, which is most fully realized in the Mosaic Torah. Its commandments instill all moral and intellectual virtues, including those unknown to reason. Only a life according to this revealed law brings the highest, spiritual happiness, the ultimate goal of philosophy.

See also: Maimonides, M.

JOSEF STERN

List of works

Arama, I. (1522) *‘Aqedat Yitzhaq (The Binding of Isaac)*, ed. H. Pollack, Pressburg, 1849; English trans. E. Munk, *Aqaydat Yitzchaq*, Jerusalem: Feldheim, 1986, 2 vols. (Arama’s classic commentary on the Pentateuch; this standard edition in Hebrew also includes Arama’s commentaries on the Five Scrolls - *Song of Songs*, *Ruth*, *Lamentations*, *Ecclesiastes* and *Esther* - and his *Hazut Kashah (Grievous Vision)*, a polemic on the relations of philosophy and religion. The English translation is the only extended selection of Arama’s writings available in a language other than the Hebrew original; but it must be used with great caution.)

References and further reading

Bettan, I. (1939) *Studies in Jewish Preaching: Middle Ages*, Cincinnati, OH: Hebrew Union College Press, 130-91.(The classic study of Arama’s homiletics.)

Heller Wilensky, S.O. (1953) ‘Isaac Arama on the Creation and Structure of the World’, *Proceedings of the American Academy for Jewish Research* 22: 131-50.(An English translation of chapter 4 of Heller Wilensky (1956).)

Heller Wilensky, S.O. (1956) *R. Yitzhaq Arama Umishnato (Isaac Arama and his Philosophical System)*, Jerusalem and Tel Aviv: Bialik/Dvir Publishing House.(A systematic reconstruction of the philosophy scattered through Arama’s sermons.)

Kellner, M. (1986) *Dogma in Medieval Jewish Thought*, New York: Littman Library and Oxford: Oxford University Press, 159-61.(Contains a good but brief discussion of Arama’s views on the articles of faith.)

Pearl, C. (1971) *The Medieval Mind*, London: Vallentine, Mitchell.(An accessible, thorough exposition of Arama’s philosophy.)

Arcesilaus (*c*.316-*c*.240 BC)

Arcesilaus of Pitane came to Athens as a young man, and was seduced by Platonic philosophy. Around 265 he became head of the Academy. He turned the school in a sceptical direction, urging that Plato himself had been of a sceptical bent. He revived the Socratic practice of dialectical argument, in which he displayed remarkable logical skill and honeyed oratorical talent. His dialectical prowess led him to 'suspend judgment about everything'; but the main target of his arguments was Stoicism, and in particular Stoic epistemology, which he claimed to reduce to incoherence. Recognizing that a sceptic must live and act, he introduced the notion of 'the reasonable' as a criterion of sceptical action.

1 Life and thought

Arcesilaus was born in Pitane, in north-west Asia Minor. As a youth he was a pupil of the mathematician Autolycus, whom he followed to Sardis. He then travelled to Athens where he studied with Theophrastus. He was destined for a rhetorical career; but his head lay for philosophy. When he removed to the Academy and heard Polemo and Crantor and Crates, he deemed that they were 'either gods or else remnants of those men of old who were formed from the golden generation' (Philodemus, *History of the Academy* XV 5-10). The rest of his life he spent in the Academy. On the death of Crates in *c*.265 BC he became scholarch, a position he held until his death some twenty-five years later.

He was a celebrated figure, known for caustic wit and also for kindness, for oratorical skill and for the rigour of his argumentation. He wrote epigrams, he enjoyed dalliance and dinner parties - and he was regarded as one of the leading philosophers of the age. But he produced no philosophical writings, and what we learn of his views derives from hearsay: perhaps from first-hand hearsay, for we are told that one of his pupils, Pythodorus, took notes of his lectures.

The Stoic Ariston of Chios, his contemporary, parodied a Homeric verse in describing him as 'Plato in front, at the back Pyrrho, Diodorus between them'. Homer was describing the chimera, and Ariston insinuates that Arcesilaus was a philosophical monster, a three-fold hybrid.

Arcesilaus began his philosophical life as an orthodox Platonist (see Plato; Platonism, Early and Middle) - we are told that he acquired a copy of Plato's works as a boy, and that 'at first when he stated a thesis he argued in accordance with the tradition from Plato and Speusippus up to Polemo' (Philodemus, *History of the Academy* XVIII 7-12). He taught a dogmatic Platonism and then became a sceptic.

Later authors spoke of the foundation of a New Academy. But by his own lights Arcesilaus was no innovator - rather, he turned the Academy back to pure Platonism. Not (as some alleged) because he dissembled, secretly teaching unsceptical doctrine, but because (as he argued) Plato himself had been a sceptic. 'From several of Plato's books and Socratic dialogues he took the idea that nothing is certain' (Cicero, *The Orator* III 67). There are indeed sceptical touches in some of Plato's works - notably in the early dialogues, which generally end in puzzlement, and in the *Theaetetus*, which raises and conspicuously fails to answer the question 'What is knowledge?' Yet Arcesilaus could read the Platonic corpus and say that 'in his books nothing is asserted, many issues are argued on both sides, everything is a matter of investigation, nothing certain is said' (Cicero, *Academics* I 46).

Diodorus Cronus was celebrated for 'dialectic': his philosophical interests were absorbed by logical problems and puzzles. Arcesilaus did not himself engage in logical study - indeed, anecdote has him dismiss dialectic. He took after Diodorus in his practice, inasmuch as he too was renowned for his argumentative ingenuity. He excelled at arguing 'on both sides of the question': a proposition (no matter what) is put forward, and first you argue for it and then you argue against it, 'the arguments on each side being equally powerful' (Eusebius, *Preparation of the Gospel* XIV 4.15). The technique finds antecedents in Aristotle's dialectic. Arcesilaus also harked back to Socrates (§3) and the Socratic *elenchos* or method of refutation: a thesis (no matter what) is proposed, and you show by arguments which the proposer must accept that it is untenable.

Each of these two techniques demands logical versatility or sophistical sleight of hand. The ancient sources often link them, for if you can argue for and against any proposition then *a fortiori* you can argue against any thesis, and

if you can argue against any thesis whatsoever then you can argue for and against any proposition.

2 Scepticism

It is Arcesilaus' affinities with Pyrrho, the archetypal sceptic of the ancient world, which give him his philosophical bite. 'He created a new philosophy of non-philosophizing' (Lactantius, *Divine Institutions* IV 11), and introduced suspension of judgment and scepticism into the Academy. But even in antiquity, the nature of Arcesilaus' scepticism was a matter of dispute. According to Numenius (§1), Arcesilaus 'was a Pyrrhonist in all but the name' (Eusebius, *Preparation of the Gospel* XIV 6.6); and a half dozen texts, independent of one another and drawing on early sources, agree that Arcesilaus 'suspended judgment about everything': that is to say, he held no beliefs on any subject, and the end of his philosophizing was the eradication of all belief and the introduction of universal suspension of judgment.

Arcesilaus' logical techniques open a direct route to Pyrrhonism (see Pyrrhonism). If you can produce equally powerful arguments on each side of a proposition, then you will neither believe nor disbelieve that proposition, and if you can produce equally powerful arguments on each side of every proposition, then you will believe no proposition at all. And Arcesilaus had a second route to Pyrrhonism, for he championed *akatalēpsia* or 'inapprehensibility': 'he denied that there is anything which can be known - not even the one thing which Socrates allowed himself, that he knew that he knew nothing' (Cicero, *Academics* I 45). Hence if you have any beliefs, they will be mere opinions. But no one of any sense embraces what he takes to be a mere opinion. Hence no one of any sense will hold any beliefs.

Arcesilaus was a polemicist. He attacked all-comers and exploded any thesis anyone might propose, 'affirming nothing himself but merely refuting other positions' (Philodemus, *History of the Academy* XX 1-4). Such a polemical scepticism is the natural child of the Socratic *elenchos*, and some scholars have urged that Arcesilaus' philosophy was essentially a negative and a destructive thing. But polemical scepticism easily fades into Pyrrhonism. For if I refute a thesis which you propound, then I shall not uphold the thesis myself; if I can refute any thesis which is propounded to me, then I shall believe no thesis; and so I shall end up as a Pyrrhonist.

It is sometimes supposed that Arcesilaus always argued *ad hominem*: if you maintain a certain thesis, then Arcesilaus will show not that *the thesis* is untenable but that *you* have no good grounds for maintaining it. It is consistent with a successful *ad hominem* argument that *there are* good grounds for holding the thesis, and that Arcesilaus himself holds it on such grounds. Yet the *ad hominem* approach, universally applied, again fades into Pyrrhonism, for if Arcesilaus can argue *ad hominem* against all-comers, then in effect he can argue against any thesis.

Much of Arcesilaus' philosophical activity was directed against the Stoics. Thus he took the Stoic thesis that two stuffs may completely blend or interpenetrate and, relying wholly on ideas which the Stoics themselves maintained, reduced it to absurdity. (He chose a grotesque example: amputate your leg, grind it up, and blend it into the Aegean sea. By the thesis of interpenetration, the leg will blend with every part of the sea and the fleet of King Antigonus will sail through your leg.)

The Stoics held that if you know something, you must have 'apprehended' it; that apprehension is assent to an 'apprehensive appearance' (*phantasia katalēptikē*); and that an appearance of something is apprehensive if it is true of that thing, if it was caused by that thing, and if it could not have been produced by anything else (see Stoicism §12). Arcesilaus urged that the third of these conditions could never be met and hence that there are no apprehensive appearances. Hence 'everything being inapprehensible, it will follow that, for the Stoics too, the Sage suspends judgment' (Sextus Empiricus, *Against the Mathematicians* VII 155).

But if Arcesilaus was at his most vigorous against the Stoics, no ancient text suggests that anti-Stoicism exhausted his intellectual parts. The evidence suggests rather that the quarrel with the Stoics was one glamorous episode in his campaign to restore peace and scepticism to the land: the Stoics were the leading 'dogmatic' philosophers in town, and any sceptic would feel bound to have a confrontation with them. In any event, the arguments which Arcesilaus used against the Stoics could readily be adapted for use against others: a clever philosopher would not ignore and a combative philosopher would not decline these tempting possibilities. Anti-Stoicism, in other words, soon leaches into a more general *ad hominem* scepticism and hence, in the end, into Pyrrhonism.

It has often been wondered if Pyrrhonian scepticism is a serious or an interesting philosophical option. One argument which ancient dogmatists deployed against their sceptical colleagues was the following. Human actions are characteristically explained in terms of the beliefs (and the desires) of their agents. But sceptics have no beliefs. Hence sceptics cannot act. Hence sceptics cannot live - the only good sceptic is a dead sceptic.

Arcesilaus was aware of this argument, and had an answer. The argument, he claimed, supposes that three things are necessary for action: before you sink your teeth into the succulent flesh, you must have received an appearance (‘That apple looks ripe’), you must have given your ‘assent’ and formed a belief (‘That apple is ripe’), and you must have had an impulse or desire (you wanted to eat a ripe apple). Arcesilaus demurred:

> *Two* things are necessary for action, an appearance of something appropriate and an impulse towards the appropriate item which has appeared - and neither of these conflicts with suspension of judgement. It is belief, and not appearance or impulse, from which argument separates us.
>
> (Plutarch, *Against Colotes* 1122C-D)

The apple looks juicy and Arcesilaus wants an apple - so he eats it. No belief intervenes, and none is needed. It is enough to ‘follow the appearances’.

In addition Arcesilaus observed that ‘anyone who suspends judgment about everything will measure his choices and aversions, and in general his actions, by what is reasonable (*to eulogon*); and if he proceeds in accordance with this standard (*kritērion*) he will be successful’ (Sextus Empiricus, *Against the Mathematicians* VII 158). A sceptic will choose to perform those actions which, were they done, could be reasonably defended. Is Arcesilaus here telling us what *dogmatists* - what *Stoics* - will be obliged to do, once his arguments have reduced them to scepticism? Or is he reflecting more generally on the way in which any sceptics will make their way through the world? And in the latter case, are his reflections compatible with scepticism - can a sceptic consistently appeal to ‘the reasonable’? To such teasing questions our texts yield no safe answers.

See also: Carneades; Socrates §8

JONATHAN BARNES

References and further reading

Annas, J. (1992) ‘Plato the Sceptic’, in J. Klagge and N. Smith (eds) *Methods of Interpreting Plato and his Dialogues*, Oxford Studies in Ancient Philosophy supplementary vol., Oxford: Clarendon Press.(A discussion of the sceptical interpretation of Plato.)

Cicero, M.T. (45 BC) *Academics*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933.(A main source for Academic scepticism.)

Couissin, P. (1929) ‘The Stoicism of the New Academy’, in M.F. Burnyeat (ed.) *The Skeptical Tradition*, Berkeley, CA: University of California Press, 1983.(Argues that Arcesilaus’ scepticism was purely anti-Stoic.)

Long, A.A. (1986) ‘Diogenes Laertius, Life of Arcesilaus’, *Elenchos* 7: 429-49. (Primarily biographical.)

Maconi, H. (1988) ‘Nova non philosophandi philosophia’ (A new philosophy of not philosophizing), *Oxford Studies in Ancient Philosophy* 6: 231-53.(On the nature of Arcesilaus’ scepticism.)

Mette, H.-J. (1984) ‘Zwei Akademiker heute: Krantor von Soloi und Arkesilaos von Pitane’ (Two Academics: Crantor and Arcesilaus), *Lustrum* 26: 7-94.(Collects the ancient texts bearing on Arcesilaus.)

Philodemus (*c*.80-40 BC) *History of the Academy*, trans. T. Dorandi, *Filodemo. Storia dei filosofi. Platone e l’Academia*, Naples: Bibliopolis, 1991.(Greek text, with Italian translation and commentary; a biographical history.)

Striker, G. (1980) ‘Sceptical Strategies’, in M. Schofield, M.F. Burnyeat and J. Barnes (eds) *Doubt and Dogmatism*, Oxford: Clarendon Press.(An account of the aims and methods of Academic scepticism.)

Archaeology, philosophy of

Questions about the scientific status of archaeology have been central to field-defining debates since the late nineteenth century and have frequently involved appeals to philosophical sources. With the possible exception of Collingwood, however, there was little systematic exploration of the bearing of philosophical literature on these questions until the advent, in the 1960s and 1970s, of the New Archaeology, a self-consciously positivist research programme. The New Archaeology originated in North America but has been widely influential, especially in giving prominence to philosophical and theoretical issues. The New Archaeologists' advocacy of a positivist (Hempelian) conception of scientific goals and practice provoked intense debate which involved philosophers of science as well as archaeologists from the early 1970s. Although the positivist commitments of the programme were widely repudiated a decade later, philosophical exchange has continued and expanded to include consideration of a range of post-positivist models of scientific inference that emphasize the theory-ladeness of archaeological evidence, as well as hermeneutic and post-structuralist models of archaeological interpretation. The analysis of epistemological issues is also closely tied to foundational questions about how the cultural subject of archaeological inquiry should be conceptualized and has led, increasingly, to a consideration of normative questions about the values and interests that shape archaeological research and the ethical responsibilities of practitioners. In 1992 Embree argued that work in this area had achieved sufficient maturity to be recognized as a subfield which he designated 'meta-archaeology'.

1 The emergence of meta-archaeology

As advocates of an explicitly scientific research programme, the New Archaeologists rejected what they described as traditional 'inductive' practice; they insisted that law-governed, deductive-nomological explanation should be their primary goal, and that interpretive hypotheses should be rigorously tested following a hypothetico-deductive model of confirmation (see Explanation §2). When these commitments were made explicit in the early 1970s they generated debate within archaeology about philosophical theories of science which drew the attention of several philosophers. R.A. Watson consistently defended the positivist orientation of the New Archaeology against its critics. Most others objected that internal, philosophical critiques had substantially undermined Hempelian models by the 1970s (see Explanation §3), when they were taken up by archaeologists, and cautioned against treating philosophical models as authoritative recipes for scientific practice; these are properly 'theses to be argued', not 'established truths' (Nickles 1977: 164). Despite the contentious tone of some of these exchanges, a number of philosophers and archaeologists persisted in discussion across disciplinary boundaries, and have produced an increasingly sophisticated, diverse and resolutely interdisciplinary literature on philosophical issues raised by and about archaeological practice.

Contributors to this jointly archaeological and philosophical literature frequently address questions about how the project of meta-archaeology should be defined and situated. Some insist that there are irreducible differences between the interests of philosophers and archaeologists, even where similar questions seem to be at issue. Typically, however, the case is made for establishing meta-archaeology as an interdisciplinary venture that includes a range of science study disciplines. As such, philosophy of archaeology may exemplify a pattern of development within post-positivist philosophy of science which has led philosophers in many areas to establish increasingly close ties with the sciences they study.

2 Explanation

The New Archaeologists attempted to fit the explanatory goals of their discipline to a particularly narrow philosophical template. After an initial round of debate which focused on standard philosophical objections to covering-law models, attention turned to questions about explanation that arise from archaeological practice. At one end of a continuum of views are arguments that explanation should not be treated as the central objective of scientific inquiry, however conceptualized. Morgan (1973) took this position in debate with Watson *et al.* (1971). Kelley and Hanen (1988) have since argued that an antirealist, pragmatist view of explanation makes best sense of archaeological practice; on this account explanations are simply answers to 'why-questions' which deploy whatever scientifically credible information will satisfy a specific inquirer (see Scientific realism and antirealism). Although this suggests that explanation is a byproduct of scientific inquiry, not its goal, Kelley and Hanen also

treat explanatory power as a key consideration in the evaluation of competing hypotheses, so explanatory goals, reconceptualized, remain central to their account. A robustly realist alternative has been proposed by Gibbon (1989), and causalist models have been developed by Salmon (1982), who recommends a 'causally supplemented' statistical-relevance model of explanation, and by Wylie (1996a), who uses archaeological examples as the basis for a critique of unificationist accounts of explanation. As in these last cases, close consideration of archaeological practice frequently leads to the reassessment of philosophical models. For example, Nickles (1977) uses archaeological examples to establish a case for recognizing forms of 'singular causal' explanation which do not depend on laws, and Salmon (1982) argues that any adequate account of explanation must make sense of forms of functional explanation which are ubiquitous in archaeology (see Causation; Explanation; Functional explanation).

Analyses of explanation in archaeology are closely connected to foundational questions about the nature of the cultural subject matter. The New Archaeologists conceptualized their subject domain in terms that could support a search for laws; cultures were understood to be ecologically adaptive systems, and the New Archaeologists' ambition was to understand, not just the specific events and conditions that produced the archaeological record but, through them, larger scale processes shaping cultural systems in adaptation to their external environments. Archaeological critics insisted, on ethnographic and archaeological grounds, that this view is untenable; dynamics internal to the ethnographic life-world of cultural subjects can substantially shape the large scale, long term development of cultural systems. Likewise, most parties to the philosophical debate about explanation in archaeology insist that questions of content cannot be settled in advance; models of explanation must be flexible enough to accommodate reconstructions of beliefs, intentions, cultural conventions, social institutions, and various 'cognitive' dimensions of human life, even though some of these are unlikely to be law-governed (Nickles 1977) or accessible to material-causal analysis (Salmon 1982). Some urge recognition that, in practice, the 'archaeological record' is treated ambiguously, both as a fossil record which requires scientific modes of explanation and as a textual record of intentional, meaningful action, for which interpretive approaches are appropriate (Patrik 1985) (see Explanation in history and social science). The advocates of interpretivist approaches are influenced by Collingwood's argument that historians must grasp the 'insides of actions' (Hodder 1991); by symbolic and structuralist trends in anthropology (Hodder 1982); by hermeneutics, phenomenology and critical theory (Leone 1982; Preucel 1991) (see Structuralism in social science; Hermeneutics; Phenomenology, epistemetic issues in; Critical theory). Increasingly, those who endorse scientific modes of inquiry also insist on the importance of considering 'cognitive' dimensions of the cultural past (Gardin and Peebles 1992; Bell 1994).

As discussion of the goals of archaeology has expanded, it has become clear that archaeologists 'explain' in many different senses and at different levels. Frequently they explain archaeological data in quite localized terms, as the products of specific events, conditions of life, intentional actions, and the various 'formation processes' responsible for the surviving archaeological record. The more ambitious goal of understanding large scale cultural processes both depends upon and is continuous with these more modest explanations.

3 Inference

A second major theme in the philosophical literature on archaeology is the concern to explicate the forms of inference by which archaeological data are systematized, and hypotheses about the cultural past are generated and evaluated in light of these data. Reacting to the deductivism of the New Archaeology, philosophical analysts have explored a range of models which more adequately capture the inductive complexity of archaeological practice. Salmon (1982) proposes a modified Bayesian account as a framework for understanding the nuanced judgements archaeologists make about the significance of archaeological evidence; for example, where the confirming or disconfirming import of specific evidential outcomes is assessed in light of the prior plausibility of a test hypothesis and the likelihood that the evidence in question uniquely confirms this hypothesis (see Confirmation theory; Probability theory and epistemology). The comparative nature of hypothesis evaluation is emphasized by Kelley and Hanen (1988) who also argue that, in addition to considerations of empirical adequacy, explanatory power and consistency with established 'core beliefs' play a crucial role in hypothesis evaluation. They conclude that archaeological practice is best understood as a matter of 'inference to the best explanation' which proceeds through 'eliminative induction'; assessments of evidential support serve as grounds for eliminating implausible hypotheses as much as for accepting those that seem relatively well confirmed (see Inference to the best explanation). Gibbon (1989) holds a similar position but, as a realist, can argue that 'best explanations' are those

which afford the most comprehensive and plausible causal explanation of the available data. In 1994, Bell renewed the argument against ‘inductivism’ and formulated a set of guidelines for archaeological inquiry that translate, into practical terms, the central philosophical insights of Popperian refutationism (see Popper, K.R. §2). Bell urged archaeologists to treat hypothesis evaluation as a matter of endangering bold conjectures rather than of building evidential support for hypotheses.

A condition for evaluating claims about the cultural past, on any of these models, is the formulation of typologies and other tools of analysis which allow archaeologists to systematize their data. Since the 1930s, typological constructs have been the subject of ongoing debate within archaeology. The central questions are whether useful, perspicuous typologies are nothing more than problem-specific ‘tools’ for manipulating data or have greater significance, capturing structures inherent in archaeological assemblages and, perhaps, underlying cultural norms and categories (see Taxonomy). A sophisticated argument for ‘typological instrumentalism’ is developed by Adams and Adams (1991), and an account of the inferential processes underlying all forms of ‘archaeological constructs’ is proposed by Gardin (1980) using a ‘logicist’ approach. Both draw attention to the selective, interpretive dimensions of the process by which archaeologists describe and systematize their data.

4 Evidence

Since the mid-1980s the most pressing epistemological problem raised by archaeology is that of explaining how, and under what conditions, archaeological data may be taken as evidence of the cultural past. Early analysis focused on the role of analogical inference in establishing the evidential significance of archaeological data. Despite the New Archaeologists’ categorical rejection of analogical inference as a form of speculative induction - a projection of the present onto the past - several philosophical commentators argued that ampliative inference of this sort is inescapable, and that it can be closely controlled (Salmon 1982; Wylie 1985). Shelley (1996) has since developed a sophisticated account of abductive reasoning in archaeology that considers the role of visual mental imagery in generating hypotheses about the cultural significance of archaeological material.

By the early 1980s, both critics and proponents of archaeological positivism had accepted the point that archaeological data are necessarily ‘theory-laden’ (see Observation §§3-4). These data stand as evidence of the cultural past only under interpretation, and their interpretation depends on a wide array of auxiliary assumptions (‘middle-range theory’) which establish causal, functional, symbolic and other connections between elements of material culture and the conditions responsible for their production and survival in archaeological contexts. While many archaeologists have made it a priority to develop the necessary linking principles through experimental and ethnographic research, others insist that such principles can never stabilize ‘ascriptions of meaning’ to archaeological data: interpretively constituted evidence cannot provide an ‘independent’ test for explanatory or interpretive hypotheses. In a discussion of ‘bootstrapping’ models of confirmation, Wylie (1986) has argued that the reliance on auxiliaries need not entail vicious circularity: it is not necessarily the case that the linking principles required to establish evidential significance will ensure support for a favoured hypothesis even if both derive from a common theory. This strategy is further developed in analyses of various kinds and degrees of independence that may be realized between hypotheses and supporting or test evidence by Wylie (1996b; Pinsky and Wylie 1989), and by Kosso (1991, 1992, 1993) in an important series of articles on ‘observation of the past’.

5 Contextualism, ideals of objectivity and professional ethics

Challenges to ideals of objectivity, which have dominated archaeological debate since the early 1980s, arise not only from concerns about the instability of evidence but from reflection on the results of detailed empirical studies of the ‘sociopolitics’ of archaeology. These have been undertaken largely by archaeologists and document the influence on archaeology of its colonial, nationalist and imperialist entanglements (Trigger 1989); its relationship to intra-national and international elites; its class structure (Patterson 1986, 1995); its assimilation of racist and sexist presuppositions (Trigger 1980; Gero and Conkey 1991); and myriad features of its funding base, internal communication patterns, institutionalization, recruiting and training, and reward structures (Gero, Lacey and Blakey 1983; Kelley and Hanen 1988; Gibbon 1989). Although these analyses undermine objectivist ideals which make a primary virtue of neutrality and value freedom, most who urge attention to ‘contextual’ factors resist relativist conclusions. Their hope is that critical analysis will make archaeology more accountable for its presuppositions and will yield better informed judgments about the credibility and likely limitations of archaeological knowledge.

The pressure to engage normative issues has been mounting since the early 1970s when it became clear that the future of archaeology was threatened by accelerated destruction of archaeological resources and an unprecedented expansion of the international antiquities market. In addition, descendent populations and other national and public interest groups have challenged archaeologists' rights of access to archaeological sites and material, often on grounds that scientific investigation does not serve their interests in what they regard as their cultural heritage. Such changes in the contexts of archaeological practice raise ethical questions about whether archaeologists are ever justified in making professional use of looted or illegally traded material; whether they should consider the preservation of archaeological resources as important a goal as that of exploiting them for scientific purposes; what responsibilities they have to communities affected by their research; and how the goals of scientific investigation are to be weighed against heritage interests when these conflict (see, for example, Salmon 1997; Wylie 1996c).

See also: Postcolonial philosophy of science

ALISON WYLIE

References and further reading

Adams, W.Y. and Adams, E.W. (1991) *Archaeological Typology and Practical Reality: A Dialectical Approach to Artifact Classification and Sorting*, Cambridge: Cambridge University Press.(The product of collaboration between an archaeologist and a philosopher, this is an analysis of typological construction that is grounded in close consideration of archaeological practice.)

Bell, J.A. (1994) *Reconstructing Prehistory: Scientific Method in Archaeology*, Philadelphia, PA: Temple University Press.(Bell advocates Popperian refutationism as a model and source of guidelines for archaeological practice.)

Embree, L. (ed.) (1992) *Metaarchaeology: Reflections by Archaeologists and Philosophers*, Boston Studies in the Philosophy of Science, Boston, MA: Clair. (Embree provides a detailed survey of jointly philosophical and archaeological analyses which he designates 'meta-archaeology' and assembles representative essays by most of those he discusses in his introduction.)

Gardin, J.-C. (1980) *Archaeological Constructs*, Cambridge: Cambridge University Press.(A distinctively French, 'logicist' analysis of the conceptual categories and forms of reasoning that structure archaeological writing.)

Gardin, J.-C. and Peebles, C.S. (eds) (1992) *Representation in Archaeology*, Bloomington, IN: Indiana University Press.(The editors provide useful surveys of 'semiotic trends' in archaeology and of the prospects for scientific investigation of the cognitive aspects of human evolution.)

Gero, J.M. and Conkey, M.W. (eds) (1991) *Engendering Archaeology: Women and Prehistory*, Oxford: Blackwell.(The first collection of essays to appear in archaeology which made questions about women and gender their central concern.)

Gero, J.M., Lacy, D.M. and Blakey, M.L. (eds) (1983) *The Socio-Politics of Archaeology*, Research Report Number 23, Amherst, MA: Department of Anthropology, University of Massachusetts.(The first major collection of essays on the sociological conditions of archaeological practice.)

Gibbon, G. (1989) *Explanation in Archaeology*, New York: Blackwell.(An analysis of philosophical issues raised by the New Archaeology, undertaken by an archaeologist, which provides a survey of debates about the positivist commitments of the New Archaeology and an assessment of alternative philosophical models of science.)

Hodder, I. (ed.) (1982) *Symbolic and Structural Archaeology*, Cambridge: Cambridge University Press.(Hodder was one of the most influential critics of the New Archaeology and has been a primary exponent of symbolic and structuralist alternatives. In this collection of essays he assembled a range of theoretical and empirical studies that illustrate the potential of what came to be known as post-processualism.)

Hodder, I. (1991) 'Interpretive Archaeology and its Role', *American Antiquity* 56: 7-18. (Here Hodder argues the case for a hermeneutic approach to interpretation of the archaeological record, extending and reframing his earlier recommendations for post-processual archaeology.)

Kelley, J. and Hanen, M. (1988) *Archaeology and the Methodology of Science*, Albuquerque, NM: University of New Mexico Press.(A collaboration between an archaeologist and a philosopher which provides a review of the history of philosophical interest within archaeology and a model of archaeological practice based on the analysis of a wide range of case studies.)

Kosso, P. (1991) 'Method in Archaeology: Middle-range Theory as Hermeneutics', *American Antiquity* 56 (4): 621-7.(Kosso argues that, despite the antipathy between scientific, processual archaeologists and interpretive post-processualists, both make essentially the same use of linking principles - 'middle-range theory' - to

interpret archaeological data as evidence.)

Kosso, P. (1992) ‘Observation of the Past’, *History and Theory* 31 (1): 21-36.(A view recognizing the independence of archaeological evidence from the reconstructive hypotheses it is used to support, despite reliance on interpretive principles.)

Kosso, P. (1993) ’Middle-range Theory in Historical Archaeology’, *Studies in the History and Philosophy of Science* 24 (2): 163-84.(A sophisticated account, by a philosopher, of how archaeological data are constituted as evidence of the cultural past.)

Leone, M.P. (1982) ’Some Opinions about Recovering Mind’, *American Antiquity* 47 (4): 742-60.(An influential argument for attending both to intentional ideational dimensions of the cultural subject and to ideological-political factors that shape archaeological practice.)

Morgan, C.G. (1973) ’Archaeology and Explanation’, *World Archaeology* 4: 259- 76.(A sharply critical commentary, by a philosopher, on the New Archaeologists’ appeals to positivist models of science; Morgan focuses on Watson, Leblanc and Redman’s Explanation in Archaeology.)

Morgan, C.G. (1974) ’Explanation and Scientific Archaeology’, *World Archaeology* 6: 133-137.(A rebuttal to the response of Watson, Leblanc, and Redman in which Morgan expands elaborates his critique of the philosophical naïvete of the New Archaeology.)

Nickles, T. (1977) ’On the Independence of Singular Causal Explanation in Social Science: Archaeology’, *Philosophy of Social Science* 7: 163-87.(Examples of archaeological explanation serve as the basis for an argument against the Hempelian requirement that explanations presuppose covering laws.)

Patrik, L.E. (1985) ’Is there an Archaeological Record?’, in M.B. Schiffer (ed.) *Advances in Archaeology*, vol. 8, New York: Academic Press.(Patrik argues that archaeologists conceptualize the ’archaeological record’ in two fundamentally different senses, both as ’fossil’ record and as textual record.)

Patterson, T.C. (1986) ‘The Last Sixty Years: Toward a Social History of Americanist Archaeology in the United States’, *American Anthropologist* 88: 7-22. (Patterson identifies several distinct traditions of archaeological practice which he argues are aligned with the political economy of regional elites in the USA.)

Patterson, T.C. (1995) *Toward a Social History of Archaeology in the United States*, Orlando, FL: Harcourt Brace.(An examination of the history of professional archaeology in the USA, especially its class structure, which sets the development of the New Archaeology and associated philosophical debates in a larger socio-political context.)

Pinsky, V. and Wylie, A. (1989) *Critical Traditions in Contemporary Archaeology: Essays in the Philosophy, History and Socio-politics of Archaeology*, Cambridge: Cambridge University Press.(A collection of essays that provides an overview of philosophical debate in archaeology, and a selection of empirical studies of the historical formation of the discipline and its contemporary soco-political dynamics.)

Preucel, R.W. (ed.) (1991) *Processual and Postprocessual Archaeologies: Multiple Ways of Knowing the Past*, Carbondale, IL: Southern Illinois University Press, 17-29. (Preucel’s essay, ‘The Philosophy of Archaeology’, provides a useful survey of competing philosophical positions in archaeology through the 1980s.)

Salmon, M.H. (1982) *Philosophy and Archaeology*, New York: Academic Press. (Through the 1970s Salmon, a philosopher, published several essays for an archaeological audience on philosophical issues raised by the New Archaeology and then this monograph, the first sustained analysis of philosophical models for archaeological practice.)

Salmon, M.H. (1997) ’Ethical Considerations in Anthropology and Archaeology, or Relativism and Justice for All’, *Journal for Anthropological Research* 53: 47-63.(An analysis of ethical issues that arise for anthropologists generally; Salmon argues a commitment to the codes of ethics developed by a number of anthropological associations.)

Shelley, C. (1996) ’Visual Abductive Reasoning in Archaeology’, *Philosophy of Science* 63: 278-301.(A sophisticated account of the role of visual imagery in archaeology that is addressed to a philosophical audience.)

Trigger, B.G. (1980) ’Archaeology and the Image of the American Indian’, *American Antiquity* 45: 662-76.(An early critical analysis of racist presuppositions inherent in archaeological interpretation.)

Trigger, B.G. (1989) *A History of Archaeological Thought*, Cambridge: Cambridge University Press.(A comprehensive historical account that traces the emergence of distinct national traditions of archaeological practice.)

Watson, P.J., Leblanc, S.A. and Redman, C.L. (1971) *Explanation in Archaeology: An Explicitly Scientific Approach*, New York: Columbia University Press.(The most explicit and widely influential statement of the positivist commitments of the New Archaeology.)

Watson, P.J., Leblanc, S.A. and Redman, C.L. (1974) ’The Covering Law Model in Archaeology: Practical

Uses and Formal Interpretation', *World Archaeology* 6 (2): 125-31.(A rebuttal of Morgan 1974.)

Watson, P.J., Leblanc, S.A. and Redman, C.L. (1991) 'What the New Archaeology Has Accomplished', *Current Anthropology* 32: 275-91.(Watson here defends the New Archaeology, retrospectively, against its critics.)

Wylie, A. (1985) 'The Reaction Against Analogy', in M.B. Schiffer (ed.) *Advances in Archaeological Method and Theory*, New York: Academic Press, vol. 8, 63-111. (A model of analogical reasoning that counters the claim, associated with the New Archaeology, that archaeological uses of analogy are inevitably speculative.)

Wylie, A. (1986) 'Bootstrapping in Un-Natural Sciences: An Archaeological Case', in A. Fine and P. Machamer (eds) *PSA 1986*, East Lansing, MI: Philosophy of Science Association, vol. 1, 314-22.(Glymour's bootstrapping model provides the framework for showing how a limited independence of evidence from hypotheses can be realized even when both the test hypothesis and interpretive principles derive from common encompassing theory.)

Wylie, A. (1996a) 'Unification and Convergence in Archaeological Explanation: The Agricultural 'Wave of Advance' and the Origins of Indo-European Languages', in D.K. Henderson (ed.) *The Southern Journal of Philosophy*, Supplement Explanation in the Human Sciences 34: 1-30, 1996.(The adequacy of Kitcher's unificationist model of explanation is assessed in light of an archaeological debate over the credibility of explanatory appeals to the unifying power.)

Wylie, A. (1996b) 'The Constitution of Archaeological Evidence', in P. Galison and D.J. Stump(eds) *The Disunity of Science: Boundaries, Contexts, and Power*, Stanford, CA: Stanford University Press, 311-43.(A model is proposed for understanding how archaeologists stabilize evidential claims, exploiting local independence between lines of evidence, and between evidence and test hypotheses.)

Wylie, A. (1996c) 'Ethical Dilemmas in Archaeological Practice: Looting, Repatriation, Stewardship, and the (Trans)Formation of Disciplinary Identity', *Perspectives on Science* 4 (2): 154-94.(An overview of the history of debate about ethical issues within North American archaeology, and an assessment of epistemic and normative issues that underlie these debates.)

Archē

Archē*, or 'principle', is an ancient Greek philosophical term. Building on earlier uses, Aristotle established it as a technical term with a number of related meanings, including 'originating source', 'cause', 'principle of knowledge' and 'basic entity'. Accordingly, it acquired importance in metaphysics, epistemology and philosophy of science, and also in the particular sciences. According to Aristotle's doctrine of scientific principles, all sciences and all scientific knowledge are founded on principles (*archai*) of a limited number of determinate kinds.*

Archē is an ancient Greek term meaning 'rule' (in a political sense) and 'beginning'. The latter sense was developed in philosophical contexts to mean 'origin', 'starting point', and 'principle' or 'first principle'. It derives from the verb *archō*, 'to begin', 'to rule'. Both verbal meanings are found in Homer, although he does not use the noun '*archē*' to mean 'rule'. In historical and political writings *archē* means 'sovereignty', 'realm', 'political office', and, in the plural, 'authorities', 'magistrates'. The history of the noun as a philosophical term is controversial. Passages like Simplicius, *On Aristotle's Physics* 24.15 and 150.23 have led many to believe that Anaximander (§2) was the first to use *archē* in a philosophical sense, while according to others these passages simply mean that he was the first to refer to the first principle of his system (in Simplicius' vocabulary, the *archē*) as *apeiron*. Depending on the view taken, the history of *archē* as a philosophical term begins either in the sixth century BC or much later.

The latter view is probably correct. *Archē* is not used in a distinctively philosophical way in the authentic fragments of any Presocratic philosopher before Philolaus and although it carries philosophical weight in his writings (fragments 6, 8, 13) as well as in the Hippocratic medical writings, in Plato (especially *Phaedo* 101, *Phaedrus* 245 and *Republic* 510-11, 533), and possibly in early Greek mathematics, we must look to Aristotle for systematic discussion and use of the term (see Aristotle §6).

In *Metaphysics* V 1 Aristotle says the term *archē* is used in six ways, including the following: 'that from which (as an immanent part) a thing first arises, for example, the foundation of a house'; 'that from which (*not* as an immanent part) a thing first arises, and from which the movement or the change naturally first proceeds, as a child comes from the father and the mother'; 'that by whose choice that which is moved is moved and that which changes changes, for example, the magistracies in cities, and the arts'; and 'that from which a thing can first be known; for example, the hypotheses are the origins of demonstrations'. Aristotle points out that all causes (*aitia*) are *archai*, as are the nature (*physis*) of a thing, the elements, thought, choice, substance, and 'that for the sake of which'.

In accordance with some of these conceptions of *archē*, Aristotle and his Peripatetic successors refer to the basic principles of the Presocratics (whether they were conceived of as originative substances or as fundamental entities) as *archai*. In *Physics* I, Aristotle develops a more abstract account of physical *archai*, identifying them as two (matter and form) or three (matter, form and privation) (see Aristotle §8). The Stoics, who were material monists, believed in two *archai* - a passive principle, matter, and an active principle, god, which like Aristotle's physical *archai* are best understood as inseparable aspects of substance (see Stoicism §3).

It is unclear how, why or in what field the Greeks first conceived of basing proofs on unprovable principles (the fundamental step in axiomatics). By the late fifth century BC mathematics had made progress in this direction. In discussing mathematical method in the *Republic*, Plato emphasizes the way geometry establishes conclusions on the basis of unproved 'hypotheses' which are considered obvious, but which are not, strictly speaking, known. (Unfortunately, Plato does not make clear what these 'hypotheses' are.) By contrast, dialectic proceeds from hypotheses to an 'unhypothetical' *archē* which is the basis of certain knowledge.

Influenced to some extent by mathematics, Aristotle asserts in the *Posterior Analytics* that every science is based on three kinds of *archai*: definitions (*horismoi*), hypotheses, which seem to posit the existence of the basic entities which the science studies, and common principles (*koina*) or axioms (*axiōmata*) such as the Law of Non-Contradiction, used in more than one science. These kinds of principles correspond approximately, but not perfectly, to the principles on which the geometry of Euclid's *Elements* is based: definitions (*horoi*), construction postulates (*aitēmata*) and 'common notions' (*koinai ennoiai*).

The *Posterior Analytics* develops a general doctrine of scientific *archai*. Every science (*epistēmē*) consists of demonstrations of conclusions from *archai*. Aristotle argues that the *archai* cannot themselves be demonstrated. The *archai* must satisfy several strict conditions: they must be true, primary, immediate, and prior to the conclusions that follow from them. Moreover, they must be better known than and causes (that is, *aitia*: ‘grounds’, ‘explanations’) of these conclusions. ‘Primary’ and ‘immediate’ here mean ‘unprovable’ or ‘basic’. By ‘better known’ and ‘prior’ Aristotle means better known and prior ‘in nature’ to the conclusions; that is, more intelligible. By ‘causes’, Aristotle indicates that scientific *archai* are causally prior to the facts that follow from them; he is thus committed to believing in a real ordering of facts in which some are more causally basic than others, and to holding that scientific knowledge is a matter of knowing the basic facts, knowing that they are basic, and knowing how the derivative facts follow from them. It follows that scientific *archai* are basic both ontologically and epistemologically; an *archē* is an *archē* within the context of a single science. In general, an *archē* of one science has no status either as *archē* or as conclusion in another science. In one sense *archē* is a relative term: an *archē* is an *archē* of or for certain other facts or propositions (Aristotle did not always distinguish these from one another). In another sense an *archē* is absolute: a fact is basic whether or not anyone happens to recognize it as such.

RICHARD McKIRAHAN

References and further reading

Aristotle (*c.* mid 4th century BC) *Posterior Analytics*, trans. J. Barnes, Oxford: Clarendon Press, 2nd edn, 1994. (Includes notes.)

Aristotle (*c.* mid 4th century BC) *Metaphysics*, books IV, V, VI, trans. C. Kirwan, Oxford: Clarendon Press, 1971. (Includes notes.)

Fritz, K. von (1955) ‘Die *Archai* in der griechischen Mathematik’ (Archai in Greek Mathematics), *Archiv für Begriffsgeschichte* 1: 13-103; repr. in *Grundprobleme der Geschichte der antiken Wissenschaft (Fundamental Problems in the History of Classical Science)*, Berlin: de Gruyter, 1971, 335-429.(Discusses the kinds of principles Aristotle prescribes, and their relations with Greek mathematics.)

Huffman, C. (1993) *Philolaus of Croton, Pythagorean and Presocratic*, Cambridge: Cambridge University Press. (See pages 78-92 for the meaning of the term *archē* in Presocratic philosophy, the Hippocratic writings and early Greek mathematics.)

Lapidge, M. (1973) ‘*Archai* and *Stoicheia*: A Problem in Stoic Cosmology’, *Phronesis* 18: 240-78.(Important study of these concepts in Stoic physics.)

McKirahan, R.D. (1992) *Principles and Proofs. Aristotle’s Theory of Demonstrative Science*, Princeton, NJ: Princeton University Press.(Discusses the nature of science according to Aristotle, including the kinds of scientific principles and their roles in demonstrations.)

Plato (*c.*380-367 BC) *Republic*, trans. P. Shorey, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1930; repr. in *Plato, The Collected Dialogues including Letters*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1961.(Parallel Greek text and English translation; see books 6-7 for Plato’s notion of *archē*.)

Architecture, aesthetics of

The philosophy of architecture is a branch of philosophical aesthetics concerned with various issues arising from the theory and practice of building design. The oldest writings on architecture date from antiquity and link architectural principles to more general, metaphysical elements of form and order. This tradition persisted into and beyond the Renaissance, but in the eighteenth century it began to give way to new philosophies of mind and value, according to which the determining factors of aesthetic experience are the interests and attitudes of informed subjects. Thereby architecture came within the sphere of the theory of taste.

Nineteenth-century revivals of classical and Gothic styles produced renewed interest in the nature of architecture, its place within the scheme of arts and sciences, and its role in society. Following this, twentieth-century modernism offered various accounts of the rational basis of architectural form and combined these with utopian political philosophies. As it had been in antiquity and during the Renaissance, architecture was again viewed as central to and partly definitive of a culture. More recently, however, attention has returned to analytical questions such as 'What is the nature of the aesthetic experience of architecture?' and, relatedly, 'How is it possible for there to be reasoned, critical judgments about the meaning and value of buildings?'

In order to deal with such issues philosophers in different traditions have begun to develop accounts of the social aspects of architecture, recognizing that critical judgments presuppose the capacity to identify buildings as being of various types: public, domestic, formal, informal and so on. The nature of architecture is in part, therefore, a matter of social convention or more generally 'forms of life', and this limits the scope for abstract ahistorical theorizing. None the less, the resources of metaphysics, the theories of mind, action, meaning and value are all utilized in contemporary philosophy of architecture.

1 The subject and its central themes

Two works have influenced Western architectural theory throughout most of its history: Vitruvius' *De architectura libri X (Ten Books on Architecture)*; and Alberti's *De re aedificatoria (On Architecture)*. The first bears a dedication toAugustus Caesar, and the second, which derives from it, is a product of the Italian Renaissance. Ironically, although written 1,500 years apart, they were first published in printed form within a year of each other - Alberti in 1485 and Vitruvius in 1486 - and have stood side by side as foundational texts ever since.

Like most writers of treatises on architecture, Vitruvius and Alberti are largely concerned with practical questions of design and construction, but they also address aesthetic aspects of the subject. Commenting on the 'departments of architecture', Vitruvius writes

> all of these must be built with due reference to durability (*utilitas*) and beauty (*venustas*)… [and beauty will be assured] when the appearance of the work is pleasing and in good taste, and when its members are in due proportion according to correct principles of symmetry.
>
> (*De architectura libri X*, I, iii, 2)

Earlier he explains that symmetry consists in 'a proper agreement between the members of the work itself, and relation between the different parts and the whole general scheme, in accordance with a certain part selected as standard' (*De architectura libri X*, I, ii, 4).

The classical idea that architectural beauty rests upon compositional unity is taken up by Alberti and developed in a direction that raises a further philosophical issue, namely, that of the proper object of aesthetic assessment. At the outset of *De re aedificatoria* he writes

> It is the property and business of the design to appoint to the edifice and all its parts their proper places, determinate number, just proportion and beautiful order; so that the whole form of the structure be proportionable. Nor has this design anything that makes it in its nature inseparable from matter…. Which being granted, we shall call the design a firm and graceful pre-ordering of the lines and angles conceived in the mind, and contrived by an ingenious artist.
>
> (1452, I, i)

Subsequent authors have returned to these ideas of the nature and of the true bearer of architectural beauty; mostly to endorse them, but also to develop or reject them. In the seventeenth century, for example, the English theorist Sir Henry Wotton coined a much quoted formula that derives (via Palladio and Alberti) from Vitruvius, when he wrote that architecture must aim to provide 'commodity, firmness and delight'. In the following century the French architect and theoristÉtienne-Louis Boullée echoed Alberti's claim that architecture should be identified with abstract design rather than with material construction:

> What is architecture? Shall I join Vitruvius in defining it as the art of building? Indeed, no, for there is a flagrant error in this definition. Vitruvius mistakes the effect for the cause. In order to execute it is first necessary to conceive… it is this product of the mind, this process of creation, that constitutes architecture.
>
> (Boullée 1790)

Although none of these writers is a philosopher, the issues they raise are instances of central topics in aesthetic theory, namely, the basis of value and the objects of appreciation. Reflection on architecture gives rise to further questions concerning its status in respect of the fine arts, the character of our experience of designs and buildings, and the relation of the built environment to other aspects of human existence, for example, religion, morality and politics. The systematic exploration of these and similar issues constitutes the philosophy of architecture.

Conceived in these terms the subject is a modern one, for although the roots of architectural theorizing run deep into the foundations of Western culture, the writings so far cited contain, at best, incidental reflections on normative and ontological issues. It is only with the development of aesthetics as a distinct branch of academic philosophy, a process begun in the eighteenth century, that conceptual resources were fashioned for constructing comprehensive philosophical accounts of architecture. Indeed, it can even be argued that the existence of the subject dates from the publication of Roger Scruton's *The Aesthetics of Architecture* in 1979. For until then philosophical discussions were partial and generally ill-informed. What Scruton offers, by contrast, is a sustained investigation of the aesthetic experience of architecture which draws upon several parts of analytical philosophy, most prominently the philosophy of mind and action, and the theory of meaning.

One way of reading Scruton's work is as an attempt to answer the question of how there can be a critical experience of architecture. This follows the Kantian approach, according to which the general form of a philosophical question is 'How is such-and-such possible? Scruton is an avowed admirer of Kant and particularly values the account of aesthetic experience set out in *Critique of Judgment*, even though his own view departs significantly fromKant's. It is therefore unsurprising that the perspective developed in *The Aesthetics of Architecture* is broadly Kantian in nature. Before coming to this, however, and in order to ensure a proper appreciation of his work and its point of departure from a purely Kantian aesthetic, it is necessary to return to antiquity and to examine some of the central ideas that informed pre-Kantian thought.

2 Pre-Kantian traditions

In Book IX, Vitruvius discusses the wisdom of the ancients, illustrating this with accounts of the geometrical discoveries of Plato and Pythagoras. He cites the latter's famous triangle theorem and adds, 'When Pythagoras discovered this fact, he had no doubt that the Muses had guided him… and it is said that he very gratefully offered sacrifice' (1914: 253). AlthoughVitruvius is concerned with the practical applications of the theorem, mention of Pythagoras and the occult nature of his discovery expressed the common view that perceptible forms are underwritten by an abstract, numerically expressible, transcendental order.

Thus when, Vitruvius, Alberti, Palladio, Wotton and others write of the importance of proportion they are drawing upon a metaphysical theory of regularity. On this account beauty is obtained by designing compositions in which symmetry (*symmetria*) and due proportion (*eurythmia*) are realized, these being determined by relevant units or modules, and various operations ('modulations') performed upon them. The central Pythagorean idea (see Pythagoreanism §2), refashioned by Plato and subsequent Neoplatonists, is that empirical order results from the imposition or expression of abstract principles upon or through a medium, in this case matter. In some accounts the units and modes of combination are few and underlie all compositions; in others the modules differ according to the nature of the thing in question. Thus, one might hold that human anatomy expresses the same basic order as the relative positions and movements of the planets, or that each system is based upon its own units and modulations. Such differences, however, are less important than the extent and duration of the consensus that beauty attends

correct composition and that correct composition is a matter of cosmically legitimated proportion.

Although Vitruvius was not printed until the fifteenth century, many manuscript versions survive from the medieval period, and it is clear from this and other evidence that the Graeco-Roman metaphysics of architecture informed the theory and practice of design throughout antiquity and the Middle Ages. This raises the question of what philosophical difference, if any, underlay the development of European architecture from Greek to Romanesque and later to 'Gothic'.

At the philosophical level the difference was one of addition and interpretation rather than of replacement. For medieval builders, Platonism was maintained in a Christianized version that included the ideas of divinely ordained symmetry and proportion. Additionally these notions came to be associated with elements from scripture, and architecture came to be seen as an enduring medium for the symbolic representation of a transcendent reality. As before, the application of geometry to part and whole dominated the practice of design but a growing interest in natural forms and their variety led to an enrichment of architectural forms.

In *Gothic Architecture and Scholasticism* (1951), Erwin Panofsky proposed an interesting, though contestable, general parallelism between High Gothic cathedrals and High Scholastic philosophical and theological treatises (such as Aquinas' *Summa Theologiae*), arguing that each aspires to 'totality', 'articulation' and 'coherence'. As far as architecture is concerned, this involves an integration of theology, morality, nature and history in the plan, elevation and furnishings of the great cathedral churches; in consequence their interpretation and appreciation calls for more than (but not less than) an ability to discern and enjoy geometrical proportion.

Given their shared assumptions about the proper sources of architectural form it is unsurprising that people of the ancient and medieval worlds thought of its beauty as objective, and of aesthetic experience as an encounter with properties whose nature is independent of our experience of them. Aquinas' definition of beauty (*Summa Theologiae*, Ia.39.8.), involving integrity (*integritas*), proper proportion (*proportio sive consonantia*) and clarity (*claritas*) is an important post-Platonist statement of this idea, though its last condition introduces an element of relativity inasmuch as being 'clearly manifest' is a relational property requiring a possible knower.

Once introduced, this relational element was bound to give rise to a question of the degree to which the nature of the knower conditions the experience of beauty - and indeed of the extent to which the grounds of beauty are themselves relative. In the seventeenth century a famous dispute concerning just these matters took place between two French classical architects:Claude Perrault and François Blondel. Beginning with his edition of Vitruvius (published in 1673), Perrault (1674) contested the standard view that the object of aesthetic experience is harmonious unity established by true order and proportion. Instead he distinguished 'convincing' (*convaincantes*) and 'arbitrary' (*arbitraires*) types of beauty, the first being universally pleasing, the second depending on subjective factors such as convention, familiarity and contingent associations. On this basis he reasoned that proportion and its beauty are arbitrary, that is, not fixed by an independent reality but determined by intersubjective agreement. In reply, Blondel (1675) argued for the importance of architecture as a bridging art between painting and sculpture, and upheld the objectivity of the harmonious unity of proportionate orders. In this latter he was subsequently and emphatically supported by Boullée who insisted upon the certainty that proportion derives from natural symmetry: 'The basic rule and the one that governs the principles of architecture, originates in regularity' (Boullée 1790).

To some extent the debate was misconceived since, like Aquinas, Blondel acknowledged human relational elements in the analysis of beauty (as did Boullée and Perrault conceded the objectivity of certain kinds of aesthetic properties. None the less, it marked the beginning of a period in which philosophers and others turned towards nonobjectivist aesthetic theories. In his *Philosophical Inquiry into the Origin of our Ideas of the Sublime and the Beautiful* (1757), for example, Edmund Burke gives various psychological explanations of architectural features and of our approval of them, including the claim that Stonehenge is judged 'grand' because of the idea it induces of the difficulty of its creation (II, §12)

3 Post-Kantian perspectives

Like all such generalizations, the claim that philosophical aesthetics began with Kant is open to contention. It is undisputed, however, that his *Critique of Judgment* (1790), like his other major works, represents one of the points of definition of modern philosophy (see Kant, I. §12). In aesthetics, as in theoretical and moral thought, Kant's

principal innovation was to convert the relationship between subject and object, and to argue that sceptical doubts were answered by the consideration that since the structure of the human mind conditions the realm of its experience and understanding there is no general possibility that facts should elude the power of the mind to grasp them. The conditions of something being the case include its being a possible object of experience.

In the realm of aesthetics Kant's aim was to show how judgments of beauty could be subjective and yet assessable as correct or incorrect. When I say, 'This arched gateway is beautiful', I am not simply saying that I like it, but rather that my liking it arises from my judgment of its quality. For Kant an explanation of this involves the free play of the imagination engaged by something possessing form. Since form in this sense is a function of the mind's organizing tendency, and this tendency and the imagination are powers common to all rational subjects, if I regard the gateway as a formal object and view it apart from any practical or scientific interest then the experience I have and the pleasure that this involves will be similar for anyone else in an equivalent condition. In other words, aesthetic judgment admits the possibility of intersubjective validity.

From the perspective of the older metaphysical rationalism of Vitruvius, intersubjectivity is still subjectivity and thus falls short of what, on that account, a recognition of architectural beauty implies; but whether the Kantian view is incompatible with the kind of formal objectivism advanced by Aquinas and Blondel is another more subtle question. Concerning aesthetic experience in general and specifically that of architecture, even a supporter of Kant's general perspective may, however, take issue with his distinction between 'free' and 'dependent' beauty.

The experience of sensible forms such as a rainbow or a curling plume of rising smoke, attended to for their own sake and without any concern for their scientific nature or practical function, is the occasion of pure judgments of free beauty. Contrasted with these, and very much secondary to them, are applied judgments of dependent beauty. In the case of the latter, the experience and judgment is conditional upon a conception of the nature of the thing in question. Thus if in judging that a chapel is beautiful I take account of its religious function and relate its aesthetic qualities to this, judging it to be 'a beautiful chapel', then the beauty is dependent and the judgment is applied.

It is clear, however, that a theory of aesthetic experiences of architecture must accommodate the fact that buildings are functional objects. Someone whose judgments always abstracted from the fact that what they were looking at was an occupiable structure designed as such - a house, a church or an airport, for example - would rightly be held to be missing the whole point of these things. Architecture is not abstract sculpture and any theory of its nature and of our experience of it that seeks to give it this status is on the wrong track.

Scruton's appropriation of Kant takes the form of an acceptance of much of his general theoretical and practical philosophy, and the adoption of the structure of his aesthetic theory, with the major qualification that what was secondary and derivative in the *Critique of Judgment* becomes paradigmatic in *The Aesthetics of Architecture*. The experience of architecture is typically a judgment of feeling of something recognized to be a building and found to be pleasing as such. With this foundation in place, Scruton is able to build a theory that incorporates elements from idealist and other anti-empiricist sources. From Hegel and Wittgenstein, for example, he takes the idea that the conditions of individual subjectivity and hence of creative imagination include the pre-existence of a community within which relevant forms of practice are operative. The ability to design and to appreciate design, and the experience of occupying built designs, are constituents of given forms of life.

4 Postmodern prospects

While Scruton's book and subsequent articles comprise the most systematic and extensive conceptual study of the subject, the most active sources of writings now catalogued as philosophy of architecture lie within what is generally described as 'continental' philosophy, that is, those branches of speculative and political thought that derive from existential phenomenology and structuralism, and which include post- and neo-structuralism, deconstruction and postmodernism. Following the example of essays such as Heidegger's 'Building, Dwelling, Thinking' (1975), some writers have tried to construct a reflective phenomenology of the experience of place and of physical containment. Even when these are effective, however, they stand in need of some more general framework such as Scruton provides. It may therefore be useful to think of the two approaches as complementary, rather than as opposed to one another.

The question of how, if at all, buildings convey meaning is a recurrent theme of both analytical and continental writings, as is the issue of the connection between architecture and aspects of the wider culture. Writers of the left,

such as Habermas, and those of the right, principally Scruton and Watkin, both find reasons to followRuskin in relating aspects of architectural theory and practice to political ideas. A common target of much recent criticism is the utopian character of the Modern Movement as expressed in the writings and work of its leading figures, such as Gropius and Le Corbusier. To some extent the latter's theories of the nature of architecture and the basis of its aesthetic values recall the earlier, pre-modern, Neoplatonic traditions; but they also embody a notion of the architect as Messiah, bringing to a heedless world the redemptive truths of a revolutionary social message - a creed the advent of which has been determined by the logic of history. As Le Corbusier expressed it:

> A great epoch has begun. There exists a new spirit…. If we challenge the past, we shall learn that styles no longer exist for us, that a style belonging to our own period has come about; and that there has been a revolution.
>
> (1923)

Any plausibility such views might once have had was long ago undermined by the conspicuous failures of modernist architecture; but the philosophical attack upon them has been directed against their historicist and totalitarian assumptions. The collapse of Marxism-Leninism and the rise of radical relativism among thinkers of the left have produced less ambitious, more provisional and contextual ideas about the role of architecture as an element in social policy.

This last trend also contributes to one important strand of postmodernist thought. The term 'postmodernism', though since deployed very widely, had some of its earliest uses in the context of architectural criticism; and it was within this context that a distinction came to be drawn between two reactions to modernism. First there is that associated with Derrida andRorty, the defining characteristic of which has been the claim that rational legitimation is impossible. On this account, metaphysical theism and the foundationalist projects of Cartesian and Kantian rationalism have all failed and no other 'meta-narrative' is available. Appeals to reason are usually veiled exercises of power, and all that remains is the ironic affirmation of ideas and images that are without any means of validation.

The second postmodern reaction is characterized by Kenneth Frampton (1985) as 'critical regionalism'. Like its radical counterpart it rejects universal doctrines and policies, but not as part of a general attack on reason as such. Instead it favours local customs and practices and looks for vernacular solutions to contextually defined problems. This way of thinking about architecture has certain parallels with the style of moral and social philosophy advocated by Alasdair MacIntyre in a series of works beginning with *After Virtue* (1985).

Finally, thinking about the nature of the built environment cannot proceed for long without taking account of the natural one. The distinction between nature and artifice is indeed a philosophical question, but however this is defined it can hardly be denied that there are differences of kind or of degree between what has been built and the landscape within which it is set. As the continental and analytical traditions come into closer contact, and moral philosophers and aestheticians within each tradition learn about each other's concerns, it seems very likely that interest in the philosophy of architecture will grow and that it will become part of a larger philosophy of environment; at which point the metaphysics of symmetry can be expected to make a reappearance.

See also: Habermas, J.; Modernism

JOHN J. HALDANE

References and further reading

Alberti, L.B. (1452) *De re aedificatoria*; ed. J. Rykwert, trans. J. Leoni (1726), *Ten Books on Architecture*, London: Tirantif, 1955.(The first and most influential architectural treatise of the Renaissance, first printed in 1485.)

Blondel, F. (1675, 1698) *Cours d'architecture*, Paris; selections, with accompanying translations, appear in W. Tatarkiewicz, *History of Aesthetics*, vol. 3, *Modern Aesthetics*, ed. D. Petsch, The Hague: Mouton, 1974.(A defence, against Perrault, of the objectivity of correct proportion as the basis of architectural beauty.)

Boullée, É.-L. (1790) 'Architecture, essai sur l'art', trans. S. de Vallée (1953) in H. Rosenau, *Boullée and Visionary Architecture*, London: Academy Editions, 1976. (Unpublished until 1953, this is an engaging defence of the assumptions and values of neoclassicism.)

Burke, E. (1757) *A Philosophical Enquiry into the Origin of our Ideas of: the Sublime and the Beautiful*, ed. A.

Phillips, Oxford: Oxford University Press, 1990.(A youthful study of the connections between aesthetic qualities and psychological phenomena, in particular emotions.)

Frampton, K. (1985) ‘Towards a Critical Regionalism: Six Points for an Architecture of Resistance’, in H. Foster (ed.) *Postmodern Culture*, London: Pluto Press.(Written in a popular style, this argues against the universalism and rationalism of modernist architectural theory.)

Goodman, N. (1985) ‘How Buildings Mean’, *Critical Inquiry* 11: 642-53.(One of the major figures of analytical aesthetics examines the possibility of architecture having a symbolic function.)

Habermas, J. (1989) ‘Modern and Postmodern Architecture’, in S.W. Nicholson (trans. and ed.) *The New Conservatism*, Oxford: Polity Press.(An account of modernism and of possible alternatives to it by the leading advocate of post-Marxist critical theory.)

Heidegger, M. (1975) ‘Building, Dwelling, Thinking’, trans. and ed. A. Hofstader, in *Poetry, Language, Thought*, New York: Harper & Row.(A phenomenological-cum-existentialist exploration of spatial aspects of the experience of being-in-the-world.)

Horden, P. (1983) ‘The Function of Forms: Recent Architectural Aesthetics’, *Oxford Art Journal* 5: 39-45.(A useful discussion of psychological and philosophical approaches, focusing on Scruton as seen against the background of neo-Kantian idealism.)

Kant, I. (1790) *Critique of Judgment*, trans. W. Pluhar, Indianapolis, IN: Hackett Publishing Company, 1987.(Notwithstanding its obscurity and difficulty this is generally held to be the classic work of philosophical aesthetics.)

Le Corbusier (1923) *Vers une architecture*, trans. F. Etchells, *Towards a New Architecture*, London: Architectural Press, 1927.(A typically extravagant statement of the revolutionary aims of utopian modernism.)

MacIntyre, A. (1985) *After Virtue*, London: Duckworth, 2nd (corrected) edn.(Very influential presentation of the claim that contemporary liberal societies are in a state of moral chaos resulting from the fragmentation of traditions.)

Panofsky, E. (1951) *Gothic Architecture and Scholasticism*, London: Thames & Hudson, 1957. (A lively exploration of the idea that thirteenth-century architecture and philosophy exhibit significant parallels.)

Perrault, C. (1674) Abrégé des dix livres d’architecture de Vitruve, Paris; selections, with accompanying translations, appear in W. Tatarkiewicz, *History of Aesthetics*, vol. 3, *Modern Aesthetics*, ed. D. Petsch, The Hague: Mouton, 1974.(The work that opened the seventeenth-century debate about the objectivity of architectural beauty.)

Scott, G. (1914) *The Architecture of Humanism*, London: Methuen, 1961.(A defence of the values of classicism and of the autonomy of architectural aesthetics against several explanatory theories.)

Scruton, R. (1979) *The Aesthetics of Architecture*, London: Methuen.(The *locus classicus* of philosophical architectural aesthetics, written from a neo-Kantian and conservative perspective.)

(1994) *Classical Vernacular: Architectural Principles*, Manchester: Carcanet.(A further collection of essays in which Scruton develops his case against various aspects of modernism in architecture and defends the superiority of the classical understanding.)

Vitruvius (*c*.30 BC) *De architectura libri X*, trans. M. Morgan, *Ten Books on Architecture*, Cambridge, MA: Harvard University Press, 1914.(The primary text in the canon of architectural writings and the only such work to have survived from antiquity.)

Watkin, D. (1975) *Morality and Architecture*, Chicago, IL: University of Chicago Press.(A study of the historicist and utopian assumptions of modernist writings, particularly those of Pevsner.)

Archytas (early to mid 4th century BC)

Archytas of Tarentum (modern Taranto in southern Italy) was a contemporary and personal acquaintance of Plato, and the last of the famous Pythagoreans in antiquity. An ancient source (Proclus)chytas with those mathematicians 'who increased the number of theorems and progressed towards a more scientific arrangement of them' and ranks him among the predecessors of Euclid. His chief contribution in mathematics was to find a solution for the doubling of the cube. As a Pythagorean philosopher, Archytas gave mathematics universal scope: he viewed the four cardinal branches of Greek scientific knowledge - arithmetic, geometry, astronomy and music - as 'sister sciences' since they could be formulated mathematically. In both mathematics and music he emphasized the study of mean proportionals. He also conducted empirical investigations in acoustics and invented simple technical devices by which to illustrate the application of mathematical principles to mechanics. Archytas was able to combine his philosophical-scientific interests with an active political career; he was a leading statesman of Tarentum and served as a successful general.

1 Life

When Plato, on his third trip to Sicily, was forcibly detained in Syracuse by Dionysius II, he requested the help of Archytas and other friends in Tarentum, who sent a ship to rescue him. This took place in 361/360 BC, which locates Archytas' activity in the first half of the fourth century BC. (The label 'Presocratic', often used of him, describes Archytas' intellectual, rather than his strictly chronological, position.) Tarentum had continued as a stronghold of Pythagoreanism (§1) after the general dispersal of the Pythagoreans and their emigration from southern Italy. Archytas, who supposedly studied with Philolaus, devoted his interests chiefly to mathematics, musical theory and mechanics. His stature as a leading Pythagorean thinker was such that Aristotle wrote three separate treatises about him. In addition to his philosophical-scientific enterprises, Archytas achieved renown as a military commander of Tarentum, earning continual reappointments beyond the usual terms of office.

2 Mathematics and music

Archytas is famous for his geometrical solution to the problem of doubling the cube. This long-standing problem, which may originally have arisen among the Greeks from architectural consideration of how to double a solid body while retaining its shape, had been reduced by Hippocrates of Chios to that of finding the two mean proportionals. Building on Hippocrates' insight, Archytas solved the problem by means of moving, three-dimensional constructions (half-cylinders and cones), thus also introducing the concept of movement into geometry (previous Pythagoreans had not concerned themselves with motion). This has earned him a place in the annals of mathematics. But Archytas is of no less significance to historians of philosophy for at least two reasons: first, because he makes evident the high degree of sophistication that mathematics had attained by the fourth century BC, against which background the mathematical activities of Plato and his associates in the Academy must be understood; second, and more generally, because although his mathematics far surpasses in complexity the number speculation of early Pythagoreanism, Archytas none the less gives the clearest expression of the Pythagorean view that mathematics provides the philosophical key for the understanding of all of nature. The cosmic application of mathematics appears more pronounced in Archytas than in his fellow, non-Pythagorean Greek mathematicians (for example, Hippocrates of Chios, Eudoxus and, later, Euclid). The following fragment reveals the universal importance that Archytas assigned to mathematical insight:

> Mathematicians seem to me to have excellent discernment, and it is not at all strange that they should think correctly about the particulars that are; for inasmuch as they can discern excellently about the nature (*physis*) of the universe, they are also likely to have an excellent perspective on the particulars that are. Indeed, they have transmitted to us a keen discernment about the velocities of the stars and their risings and settings, and about geometry, numbers [arithmetic], sphericity [astronomy], and, not least of all, music. These seem to be sister sciences, for they concern themselves with the first two related forms of being [that is, number and magnitude].
> (fr. 1)

Mathematics is thus foundational for correct thinking about being. The view of number as an all-powerful explanatory concept for the orderly arrangement of the universe originated from the Pythagoreans' discovery of the numerical ratios governing musical harmony and seemed to them to be corroborated by the observation that

figures and shapes could be expressed arithmetically: for example, the line by two, the triangle by three, the pyramid by four points (see Pythagoreanism §2). That even three-dimensional bodies could in one way be accounted for by an aggregate of points led to a certain fusion of mathematical and physical properties in Pythagorean philosophy, allowing Aristotle to say of the Pythagoreans that 'they construct the whole universe out of numbers - only not numbers consisting of abstract units; they suppose the units to have spatial magnitude' (*Metaphysics* 1080b18-20). Archytas reports of another Pythagorean, Eurytus, that he arranged a number of pebbles to represent the figures of man or horse, and after counting the pebbles declared that such was the number of man and such of horse. Pythagorean number theory, even in the crude form it assumed in the pebble arithmetic of Eurytus, comes to a legitimate fruition with Archytas. He viewed geometry, arithmetic, astronomy and music - the classic quadrivium of medieval authors - collectively as 'sister sciences' (as later Plato would call astronomy and harmonics, expressly citing Pythagorean precedent (*Republic* 530d)), since they all have to do with number or numerical relations: geometry and arithmetic for obvious reasons, astronomy because it was treated mathematically, notably when the properties of the sphere were studied as a geometrical model to explain the movements of the celestial sphere, and music because it involved numerically expressed harmonic proportions. His high regard for these sciences was governed by the conviction - still held in modern science, whose theses often take the form of equations and formulas - that number supplies the precise, quantitative measures by which to comprehend the world.

Fragment 1, which concerns the universality of mathematical insight, forms the introduction to a work that is variously entitled *On Mathematics* or *On Harmonics*. Indeed, Archytas immediately continues to discuss the relation between pitch and frequency, citing a series of examples to show that swift movements produce high notes and slow movements low notes (for example, a stick moved at different speeds produces variations in the pitch of the sound, or the air emitted from the upper holes of a pipe yields higher notes than that from the lower holes). While Archytas' account also contains some inaccurate conclusions from his observations (the speed of the motion that produces a sound was confused with the speed of the sound itself), his acoustic theories are none the less informative about the type of rudimentary empirical investigation that characterized Presocratic science. Moreover, Archytas' acoustics provide an interesting footnote to Plato: discussing physical sense-experience in the *Timaeus* (67b, 80a), Plato appears to rely on Archytas to explain hearing and sound (where again the speed of the propagation of sound is confused with frequency: higher notes are said to reach the ear more quickly than lower notes).

In his treatment of harmonics proper, Archytas further developed Pythagorean musical theory (see Philolaus §5). He calculated the numerical ratios of the intervals in all three scales of the tetrachord (diatonic, chromatic and enharmonic) and specified a doctrine of 'means' to elucidate the connection of proportion and music. The three basic means are the arithmetic (in 6, 9, 12, the difference between the first and the second number equals that between the second and the third), the geometric (in 2, 4, 8, the ratio of the first to the second corresponds to that of the second to the third) and the subcontrary or the harmonic (in 6, 8, 12, the proportion of the first by which the second, the harmonic mean, exceeds the first ($\frac{1}{3}$) is the same as the proportion of the third ($\frac{1}{3}$) by which the third exceeds the second). The knowledge of the means is presupposed by Plato when in the *Timaeus* (35b-36b) he divides the world-soul into harmonic intervals.

However, Archytas' emphasis on means, as evidenced both in his drawing upon the mean proportionals to explain the doubling of the cube and in his determining the means of the musical proportions, goes beyond the confines of geometry and musicology. A single fragment that preserves Archytas' thoughts on political-ethical matters indicates that he regarded a just society as one ruled by equality in something like the mathematical sense of proportion:

> When calculation has been found, it checks political faction and increases concord, for there is no unfair advantage in its presence, and equality reigns. With calculation we smooth out differences in our dealings with each other. Through it the poor take from the powerful, and the rich give to the needy, both trusting in it to obtain an equal share…
>
> (fr. 3)

Archytas envisions here a society whose equilibrium depends upon right 'calculation' (*logismos*), a word that in Greek also means 'reasoning', but whose root meaning of 'counting' or 'calculating' was never wholly lost to

sight (the plural *logismoi* can in fact be translated as ‘arithmetic’) and such a form would especially be favoured by the mathematical mind of Archytas. Calculation, in its political application, leads (ideally) to figuring out the right proportion between rich and poor, establishing the means, as it were, between them, and thus ensuring political equality (*isotēs*, a term which also refers to mathematical proportion). That Archytas engaged in political reflections with mathematical overtones is not surprising, given that he was both a pre-eminent statesman and a mathematician. His stress on finding the right proportions in both domains make him what may be called ‘the philosopher of means’.

3 Miscellanea

The extent to which Archytas dedicated himself to specific problems of cosmology is not known, although fragment 1 allows the assumption that he brought mathematical insights to bear on physics and astronomy. It was most probably in the course of such pursuits that he took up the question of infinite space. Archytas posed to himself the question as to whether or not, standing at the heaven of the fixed stars (that is, the edge of the world) he could extend his hand or a staff (A24). The implication being that he could, it appears Archytas was trying to lend support to the Pythagorean idea of the ‘unlimited’ outside the heavens (see Philolaus §3).

Besides the theoretical concerns of Archytas, we are further informed of certain practical inventions. He is said to have constructed a wooden dove that could fly and a popular rattle for children to keep them amused and from breaking things in the house. These inventions, the technical details of which are not known, were probably devised in part to illustrate Archytas’ scientific theories, since he was noted for making mechanics systematic by applying mathematical principles.

In sum, although our knowledge of Archytas is comparatively meagre, it reveals one of the most versatile and creative thinkers of the fourth century BC - philosopher, statesman, mathematician and musical theorist combined in one person. Certainly, his indebtedness to previous Pythagoreans, Presocratics and contemporary Greek philosophers cannot be discounted, but despite this he evinces considerable originality. The original cast of his mind is well typified in the following fragment in which he presents, somewhat paradoxically, the philosophical craft as a process of independent discovery:

> To become knowledgeable about things one does not know one must either learn from others or find out for oneself. Now learning derives from someone else and is foreign, whereas finding out is of and by oneself. Finding out without seeking is difficult and rare, but with seeking it is manageable and easy, though someone who does not know how to seek cannot find.

(fr. 3)

HERMANN S. SCHIBLI

References and further reading

Archytas (early to mid 4th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidmann, 6th edn, 1952 vol. 1: 421-39.(The standard collection of the ancient sources, both fragments and testimonia, the latter designated by ’A’; includes Greek texts with German translations.)

Burkert, W. (1972) *Lore and Science in Ancient Pythagoreanism*, trans. E.L. Minar, Jr, Cambridge, MA: Harvard University Press.(Discusses *passim* the science and philosophy of Archytas; in particular, pages 379-83 consider Archytas’ acoustics in the context of Presocratic science, pages 384-9 examine his musical theory and pages 442-7 his number theory.)

Freeman, K. (1947) *Ancilla to the Pre-Socratic Philosophers*, Oxford: Blackwell, 6th edn, 1971.(Includes, at pages 78-80, an English translation of the major fragments of Archytas.)

Heath, T. (1921) *A History of Greek Mathematics*, Oxford: Oxford University Press; repr. New York: Dover, 1981.(Volume 1 pages 213-16 give a general introduction to the achievements of Archytas, followed by an account of his proof that there can be no number (the geometric mean) between two numbers of a superparticular proportion; pages 247-9 explain in detail Archytas’ duplication of the cube.)

Lloyd, G.E.R. (1990) ‘Plato and Archytas in the seventh letter’, *Phronesis* 35.2: 159-73.(Examines the evidence for the relationship of Plato and Archytas as given in Plato’s Seventh Letter.)

Plato (*c*.366-360 BC) *Timaeus*, trans. F.M. Cornford, *Plato’s Cosmology*, London: Routledge & Kegan Paul, 1937.

Portnoy, J. (1954) *The Philosopher and Music*, New York: Humanities Press.(A readable survey of the

relationship of philosophers to music in Western civilization; Chapter 1 provides a useful discussion of the significance of music for Greek philosophers and includes a short section on Archytas.)

Arendt, Hannah (1906-75)

Hannah Arendt was one of the leading political thinkers of the twentieth century. She observed Nazi totalitarianism at close quarters and devoted much of her life to making sense of it. In her view it mobilized the atomized masses around a simple-minded ideology, and devised a form of rule in which bureaucratically minded officials performed murderous deeds with a clear conscience. For Arendt the only way to avoid totalitarianism was to establish a well-ordered political community that encouraged public participation and institutionalized political freedom. She considered politics to be one of the highest human activities because it enabled citizens to reflect on their collective life, to give meaning to their personal lives and to develop a creative and cohesive community. She was deeply worried that the economically obsessed modern age discouraged political activity, and created morally superficial people susceptible to the appeal of mindless adventurism.

1 Life

Born into a Jewish family in Königsburg, Arendt went on to study philosophy, first under Heidegger and Bultmann at Marburg and then under Husserl and Jaspers respectively at Freiburg and Heidelberg. The Nazi rise to power put paid to her academic ambitions and she became active in Jewish politics. She was arrested and detained in 1933, but was soon able to escape to France. She was briefly interned in a French prison camp, from where she escaped to the USA in 1941. She held various academic positions there, the last at the New Centre for Social Research. She regularly wrote on public issues but was not otherwise politically active.

2 *The Origins of Totalitarianism*

Arendt made her name with *The Origins of Totalitarianism* (1951), a somewhat misleading title for a book that was concerned not only with totalitarianism but also with the rise of anti-Semitism, imperialism and racism, the ‘three elements of shame’ in the modern age as she called them. She argued that Nazi and Stalinist totalitarianism represented a wholly new form of rule based on ideology and its twin terror. Ideology, the *logos* of an idea, centred around a specific idea, be it race, class or nation. It teased out the logical implications of the idea, built a tightly knit system around them, and sought to reconstruct a polity on that basis. This necessarily involved terror, by which Arendt meant a systematic, institutionalized, carefully planned and legally unrestrained use of physical and psychological violence. In Arendt’s view totalitarianism found a fertile soil in an environment in which society had been dissolved into loose and rootless masses and the hollowed out state had been reduced to an unrestrained coercive apparatus. For her, totalitarianism, born out of the rejection of the ordinary world as we know it, had its paradigmatic expression in concentration camps - a mechanical, impersonal and unworldly ‘world’ utterly devoid of thinking, feeling, judgment, personal identity, privacy and all else that distinguishes human existence (see Ideology §1; Totalitarianism §2).

The Origins of Totalitarianism was an important work. It offered brilliant insights into the nature and role of political institutions, the inner contradictions of the modern nation-state, and the human need for rootedness. But it also had its weaknesses. Arendt treated totalitarianism as an independent and self-subsisting phenomenon relentlessly unfolding its inner logic and subject to no human constraints. She uncritically equated the Nazi and Stalinist ‘forms’ of totalitarianism, and her analysis was underpinned by theories of man and society that were nowhere clearly stated and defended. In her subsequent writings Arendt often returned to the large philosophical questions raised by the Nazi experience but inadequately explored in *The Origins of Totalitarianism*. She asked such questions as what it meant to be human, how we should live both individually and collectively, in what kind of world it was possible to lead a meaningful life, why human beings committed evil, how thinking was related to action, and what future modernity held in store for humankind.

3 *The Human Condition*

Arendt first dealt systematically with some of these questions in *The Human Condition* (1958). She distinguished two kinds of life, the *vita activa* and the *vita contemplativa*, and concentrated on the former. For her, humans are a part of nature and subject to its necessity, but are also able to transcend it and act in a truly free and gratuitous manner. She divided human activities into labour, work and action, each presupposing but going beyond and hence higher than the preceding one. Labour referred to the familiar daily round of activities centred around the

production and reproduction of life. Work referred to activities in which human beings controlled nature and interposed a durable and distinctively human world between themselves and nature. It included such things as building houses, crafts, writing books, painting pictures and composing music. Action was uniquely interpersonal and referred to activities in which human beings transcended nature, interacted with others, began something new, and made a distinct mark on the world. Action, the unique expression of the human capacities for freedom and transcendence, was a distinctively human achievement; it included such things as speaking, arguing, persuading, taking initiatives, standing up for a cause and protesting against an evil.

Although action occurred in all areas of life, Arendt argued that politics was its ideal home. The political life provided such necessary preconditions of action as a plurality of participants, publicity, public space, shared interests, the inspiring tradition of action and the possibility of immortal fame. By challenging people to ‘dare the extraordinary’ and leave behind an inspiring story that gave meaning to their existence and raised the level of communal life, the political community, a community living together in the mode of acting and speaking, realized the full potential of human existence. For Arendt this was why man was by nature a political ‘animal’.

For Arendt the classical world of Athens and Rome respected the hierarchy of the *vita activa* and nurtured a climate conducive to action, freedom and meaningfulness. The late middle and early modern ages gave the pride of place to work, and admired the craftsman. Modernity reversed the hierarchy and was centred around labour. It was distinguished by such features as an excessive preoccupation with life and its endless wants, a subjectivist morality, the loss of stable structures, the deterministic and process-like character of human existence and the reduction of government to administration. Since modernity did not nurture the necessary climate for action, people in the modern world lacked the opportunity to give meaning to their lives, and either led meaningless lives or sought a pseudo-meaning in following the allegedly objective laws of history.

4 Later works

After *The Human Condition* Arendt wrote several books, all containing stimulating insights but lacking in philosophical rigour and penetration and none matching its imaginative power. In 1961 she published *Between Past and Future*, a collection of six essays of varying quality. She explored the nature of political authority, freedom, political judgment and culture in greater detail than before, and showed her growing sympathy for Kant. She insisted that political thinking was public and representative in character and involved looking at a subject from a variety of different standpoints. The greater its range and imaginative sympathy, the more representative and valid was the resulting opinion or judgment.

In 1963 Arendt published *On Revolution*, bubbling with suggestive but sketchily explored ideas and showing signs of hurry. She argued that revolution, one of the highest forms of political action, was modern in origin and aimed to establish a secure framework of freedom. Since it was political in character, it succeeded in the USA where it did not overstep its limits, and failed in France where the revolution of 1789 was subverted by concern over the problem of poverty. In her view even the American Revolution was only a partial success. By failing to give a constitutional status to local assemblies and town hall meetings, it discouraged active political participation and the development of public spirit. Thanks to several factors, including the enormous influence of Karl Marx, the greatest theorist of revolution, it was not the American but the French Revolution that became the model of all subsequent revolutions, including the Russian revolution of 1917. For Arendt, they were all doomed from the start. She pleaded for a fuller appreciation of the ‘lost treasures’ of the revolutionary tradition, especially its concern to set up a participatory polity constructed from the bottom upwards (see Revolution §2).

In 1963 Arendt also published *Eichmann in Jerusalem* based on the Israeli trial of a prominent Nazi. Eichmann was a Nazi officer who dutifully carried out his superior’s orders and murdered several thousand Jews in one of the German concentration camps. Modifying some of her views in the *Origins of Totalitarianism*, she argued that Eichmann’s enormous crimes proceeded not so much from wickedness as from sheer thoughtlessness. He did not personally hate the Jews, nor was he a sadist or an evil man. All he did was mechanically to carry out his bureaucratic duties out of a blind sense of loyalty to the *Führer*, without once pausing to reflect on the enormity of what he was doing. His evil had neither moral depth nor deep roots in his psyche, and lacked the power to fascinate. He was a boring, banal and morally superficial man whose evil deeds, although not done inadvertently, had no deeper meaning for him and were incidental to his murderous job. Arendt’s book aroused heated controversy and made her a pariah in several Jewish and even non-Jewish circles. Her analysis was not wrong but

partial. It took no account of the passionate and fiercely moralized anti-Semitism of Nazism, did not explain why Eichmann did not think about what he was doing, and failed to explore the deeper nature and sources of evil (see Holocaust).

In the late 1960s, Arendt increasingly returned to an exploration of the *vita contemplativa* to which she had hitherto paid only cursory attention. The result was her posthumously published *The Life of the Mind* (1978). The work was intended to be in three volumes, devoted respectively to the analyses of the nature of and interrelations between the three basic human capacities to think, will and judge, but she managed to finish only the first two. For her the *vita contemplativa* took two forms, namely thinking and knowing, paradigmatically expressed in philosophy and science respectively. Thinking is concerned with a quest for meaning or significance, knowing with a search for truth. Science inquires into 'what is', and is motivated by a passionate love of truth, while philosophy inquires into 'what it means for it to be' and is motivated by an equally passionate love of wisdom. Science is analytical and investigative, philosophy reflective and meditative. Science offers hard conclusions, whereas philosophy is tentative and endlessly exploratory and, like Penelope's web, begins the day by unscrambling the certainties of the previous night. Science remains confined to the world, while philosophy transcends it and is a unique expression of human freedom.

Given her view of thinking, Arendt had some difficulty relating it to both willing and action. Willing was a worldly faculty involving other people, a wish to change the world, and a measure of moral certainty; by contrast, thinking was solitary, tentative and inconclusive. Arendt neither satisfactorily resolved their tension nor showed how the two were mediated. She faced a similar difficulty in relating thinking and action to philosophy and politics. She was puzzled as to why many philosophers displayed a great hostility to the politics of a free society and supported authoritarian rule. Sometimes she explained this in terms of the philosopher's inappropriate ambition to discover the truth about the world and to 'coerce' people into accepting it by the 'force' of logic. For the most part she stressed philosophy's solitary and unworldly nature which made it impatient of the plurality, unpredictability and the apparent chaos of political life. The latter view implied that the philosopher and the political community could never be at peace with each other, and that political philosophy was impossible. Arendt found the conclusion unpalatable, but could not see an alternative.

Arendt was a highly original thinker, who made substantial contributions to political philosophy. She offered a body of profound insights into the nature, structure and role of political life, developed a wholly new vocabulary, explored long-neglected dimensions of political experience, and highlighted some of the seductive but dangerous aspects of modernity. She helped sustain the tradition of political philosophy at a time when it was in a state of decline by asking new questions, offering new ways of answering old ones, and showing that a systematic political philosophy did not have to result in an elaborate system.

B. PAREKH

List of works

Arendt, H. (1951) *The Origins of Totalitarianism*, New York: Harcourt Brace; 2nd edn, enlarged, 1958.(A study of the development and operation of the Nazi and Communist systems of government.)

Arendt, H. (1958) *The Human Condition*, Chicago, IL: University of Chicago Press. (An examination of the nature of modernity in the light of Arendt's discussion of labour, work and action.)

Arendt, H. (1961) *Between Past and Future*, New York: Viking Press; 2nd edn, enlarged, 1968.(A collection of essays on such subjects as freedom, authority and culture.)

Arendt, H. (1963a) *On Revolution*, New York: Viking Press.(A critical examination of French and American revolutions.)

Arendt, H. (1963b) *Eichmann in Jerusalem*, New York: Viking Press.(A critical analysis of the Israeli trial of Eichmann.)

Arendt, H. (1972) *Crises of the Republic*, New York: Harcourt Brace.(A collection of essays on subjects of current political interest.)

Arendt, H. (1978) *The Life of the Mind*, New York: Harcourt Brace. (Investigations into the nature of thinking and willing.)

Arendt, H. (1982) *Lectures on Kant's Political Philosophy*, ed. R. Beiner, Chicago, IL: Chicago University Press. (An incomplete study of the nature and basis of judgment.)

References and further reading

Canovan, M. (1992) *Hannah Arendt: A Reinterpretation of her Political Thought*, Cambridge: Cambridge University Press.(A sympathetic and comprehensive study of Arendt's political thought, with an excellent bibliography. Stresses the centrality of the totalitarian experience to Arendt's thought.)

D'Entreves, M.P. (1994) *The Political Philosophy of Hannah Arendt*, London: Routledge.(A generally sympathetic study of Arendt's political thought against the background of her critique of modernity.)

Kateb, G. (1984) *Hannah Arendt: Politics, Conscience, Evil*, Oxford: Martin Robertson.(A critical study of Arendt's moral thought stressing the limitations of her account of evil.)

Parekh, B. (1981) *Hannah Arendt and the Search for a New Political Philosophy*, London: Macmillan.(A detailed examination of Arendt's critique of traditional political philosophy and her attempt to construct an alternative.)

Young-Bruehl, E. (1982) *Hannah Arendt: For the Love of the World*, New Haven, CT: Yale University Press.(A good biography of Arendt.)

Aretē

A pivotal term of ancient Greek ethics, aretē *is conventionally translated 'virtue', but is more properly 'goodness' - the quality of being a good human being. Philosophy came, largely through Plato, to recognize four cardinal* aretai*: wisdom (*phronēsis*), moderation (*sōphrosynē*), courage (*andreia*) and justice (*dikaiosynē*). Others, considered either coordinate with these or their sub-species, included piety, liberality and magnanimity. The term generated many controversies. For example, is* aretē *a state of intellect, character or both? Does it possess intrinsic or only instrumental value? Is it teachable, god-given or otherwise acquired? Is it one thing or many? If many, how are they differentiated, and can you have one without having all?*

In ordinary Greek, *aretē* functions as the abstract noun correlated with *agathos*, meaning 'good', and 'goodness' is in most contexts a correct translation. However, 'goodness', unlike *aretē*, lacks a plural, and so requires awkward periphrases such as 'kinds of goodness' or 'ways of being good'. Hence 'virtue(s)' (sometimes 'excellence(s)') is usually preferred. Similarly, 'vice(s)' is favoured for its opposite, *kakia*, more correctly rendered 'badness'. In early Greek, *aretē* has no narrowly moral use, but is contextualized to mean prowess in any field - athletic, military, political, and so on. Given the predominance of male values, it often approximates to 'valour'. This is reflected in its eventual Latinization as *virtus*, literally 'manliness', which has made 'virtue' the almost inevitable modern rendering.

A specifically moral use of *aretē* emerged gradually in the fifth and fourth centuries BC. The centrality of civic obligations in the Greek (especially the Athenian) value system gave cooperative virtues such as justice and courage a special standing, even before philosophers like Socrates and Plato began to scrutinize them. Many of the Sophists professed to teach *aretē* to the young (see Sophists). How if at all it could be acquired was a much-debated issue: would it be by teaching, practice, nature, divine favour or sheer luck?

Whereas modern virtue ethics tends to stress the culture-specific character of virtue implicit in such a background (see Virtue ethics), Greek thought rarely acknowledged this, and sometimes explicitly denied it. The dominant concern was to investigate goodness as a universal human property or ideal.

Aretē is above all functional goodness. In Plato, *Republic* I, Socrates investigates the 'goodness' of a soul by direct analogy with that of an eye: just as the *aretē* of an eye is what enables it successfully to perform its function, seeing, so the *aretē* of a soul is what enables it successfully to perform its own function, living. Aristotle's conception of *aretē* is founded on the analogous idea that there is a distinctively human function, which can be performed better or worse (see Aristotle §21). All this points to the intimate link between *aretē* and 'living well', a regular equivalent of 'happiness' (see Eudaimonia).

One recurring issue is the nature of that link. Socratic thought tended to make *aretē* and happiness extensionally equivalent. Aristotle's modification was to locate happiness in the active use of virtue, not its mere possession. A third tradition, sketched in Plato's *Protagoras* and later fully developed in Epicureanism, gives *aretē* purely instrumental value, as the prudential skill of maximizing the one intrinsic good, pleasure (see Epicureanism §10).

Socrates, on Plato's usual portrayal of him, has little sympathy with this instrumentalist account, but does favour one feature of it, the identification of *aretē* with some sort of knowledge or wisdom. Nothing is valuable unless used wisely; hence wisdom is the only underivatively valuable thing.

In Plato's *Protagoras*, Protagoras himself considers the political *aretai* - justice, courage and so on - to be innate human potentialities which can be realized by training. He sees them as separate capacities, so that someone might become, for example, brave but not wise. This separability assumption may well have been widespread, but philosophers in the wake of Socrates formed a united front against it. In the *Protagoras*, Socrates treats *aretē* as a single thing, never using the plural, and argues that justice, moderation and so on are not its 'parts' (perhaps meaning species?) but coreferential terms for it. This strong thesis of the Unit of Virtue was developed in the later Socratic tradition, especially by the Stoic, Ariston of Chios (§3), who maintained that 'justice', 'courage' and so on are all names for a single state of psychic health, differentiated purely by the contexts to which its possessor responds (apportionment, danger…). A version of the same thesis - in effect, that there is just one way of being a good person - is applied by Socrates in Plato's *Meno* to gender and class distinctions: the *aretē* of men and women,

free and slaves, is one and the same; they are all good in the same way.

Others considered the several *aretai* to be essentially distinct states of the person, but still inseparable. For mainstream Stoicism, this arose from the analysis of them as exact sciences, each with its own defining concerns, but made interdependent by their shared stock of theorems (see Stoicism §16). In the case of Plato and Aristotle, it was arrived at by concluding, *contra* Socrates, that the soul has emotive components as well as reason, and that moral *aretai* consist in a properly balanced relation between emotion and reason. Consequently the various *aretai* admit of complex analyses which enable them to be more clearly differentiated. For Plato in *Republic* IV, there are three psychic components - one rational, one spirited or competitive, one appetitive - and each of the cardinal *aretai* consists in a different relation between them. For Aristotle, each *aretē* is an educated disposition to make choices which strike a mean between excess and deficiency, especially in the relevant emotions; the rational component lies in the 'practical wisdom' (*phronēsis*) with which the choices must be informed (see Aristotle §§22-4).

The conception of *aretē* varied with the conception of human good. In his more other-worldly moods, especially in the *Phaedo*, Plato located true *aretē* in the soul's purification from bodily concerns and return to its natural discarnate state of purity and wisdom. This emphasis on intellectual (as distinct from moral) *aretē*, later taken up in earnest by Neoplatonism (see Neoplatonism) is also reflected in Aristotle. To *aretai* of character ('ethical' *aretai*) Aristotle adds intellectual *aretai*, including not just practical wisdom but also pure wisdom (*sophia*). The highest form of happiness, the life of contemplation, is achieved by concentration on the latter.

See also: Virtues and vices

DAVID SEDLEY

References and further reading

See also the entries on Socrates, Plato, Aristotle, Stoicism and Epicureanism for the further reading recommended on ethics.

Adkins, A.W.H. (1962) *Merit and Responsibility*, Oxford: Oxford University Press. (Useful guide to the early Greek background.)

Aristotle (*c.* mid 4th century BC) *Nicomachean Ethics*, trans. T.H. Irwin, Indianapolis, IN: Hackett, 1985.(Books II-V cover the ethical *aretai*, VI the intellectual ones.)

Plato (*c.*380s BC) *Protagoras*, trans. C.C.W. Taylor, Oxford: Clarendon Press, 2nd edn, 1991.(Debates the relation of *aretē* to knowledge.)

Plato (*c.*380s BC) *Meno*, trans. W.K.C. Guthrie, Harmondsworth: Penguin, 1956; repr. in *The Collected Dialogues of Plato including the Letters*, ed. E. Hamilton and H. Cairns, Oxford: Oxford University Press, 1975.(On the theme 'is *aretē* teachable?')

Plato (*c.*380s BC) *Phaedo*, trans. D. Gallop, Oxford: Clarendon Press, 1975, 68-9.(On the soul's detachment from the body.)

Plato (*c.*370s BC) *Republic*, trans. P. Shorey, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1930; repr. in *The Collected Dialogues of Plato including the Letters*, ed. E. Hamilton and H. Cairns, Oxford: Oxford University Press, 1975.(The former contains the Greek text too; Plato's most influential work on moral and political philosophy.)

Plotinus (*c.* AD 250-66) *Enneads* I 2, trans. A.H. Armstrong, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1966.(Parallel Greek text and English translation; a characteristic Neoplatonist account.)

Prior, W.J. (1991) *Virtue and Knowledge: An Introduction to Ancient Greek Ethics*, London: Routledge.(An elementary history from Homer to Stoicism.)

Argentina, philosophy in

Philosophy has been present throughout Argentine cultural life since the beginning of Spanish colonization. Despite institutional ups and downs, the teaching of philosophy was a practically constant component of higher and even secondary education. The principal currents that shaped that teaching for more than three centuries were Scholasticism, French ideology, eclectic spiritualism, positivism and in the twentieth century, all of the contemporary manifestations, such as, Husserlian phenomenology, existentialism, analytical philosophy and structuralism. A permanent characteristic, nevertheless, has been that the political vicissitudes of the country affected educational institutions.

In the nineteenth century, during the period of national independence and organization, public figures used philosophical ideas to analyse the problems of society and to make the political and institutional contributions that a country in formation required. Juan Bautista Alberdi and Domingo Sarmiento are, in this respect, two representative examples.

In the twentieth century, the figure of the professional philosopher, one who is interested in philosophical research for itself, emerged and expanded. However, thought that reflected direct interest in the problems of the community and in the ethical demands of praxis did not disappear during this era. This can be seen in such thinkers as José Ingenieros and Alejandro Korn and more recently in what has been called liberation philosophy.

Academic philosophy has made considerable progress. In the second half of the twentieth century, it has attained a high level of professional quality. In some cases, even original contributions have been made which go beyond assimilation or commentary about external philosophical influences.

In Argentina, as in the rest of Latin America, philosophy began as a pure transplant brought by those who conquered the continent. Upon creating centres of higher education (either as part of the religious orders or with the character of universities), the philosophical teaching being practised in the Spanish universities of Salamanca and Alcalá was reproduced in the Spanish colonies.

Argentine philosophy shares the same general characteristics and historical periods with the philosophies developed in other Latin American countries. In general terms, philosophy can be divided into three periods: the colonial period, the nineteenth century, or national period and the twentieth century.

1 The colonial period

Philosophical activity during the colonial period was exclusively didactic. The subject matter taught was Scholastic philosophy, which had predominated in Europe during the Middle Ages and continued to exist in Spain in a renewed form well beyond the Renaissance period. Those who transmitted this knowledge were members of the Church.

The teaching of philosophy began at the University of Córdoba in the seventeenth century. However, the first indications that allow us a better idea of its content come from the eighteenth century.

Following medieval tradition, the teaching of philosophy was imparted in the College of Arts where ‘grammar’ students who had studied Latin, the language of instruction, were admitted. The subjects taught were logic, physics, metaphysics and ethics. The study of the arts was necessary to enter into the course of study of the highest distinction, theology, law or medicine, where these were available.

Although scholastic philosophy in the River Plate lasted until the dawn of the nineteenth century, in the eighteenth century there was a perception of the advance of modern natural science and its consequences for the worldview of the time which was based on the science of Aristotle. Thus, there were attempts to accommodate traditional teaching to the new modalities. Some Jesuit professors at the University of Córdoba (before the expulsion of the order in 1767) were up-to-date on the ‘new physics’ without this altering the institution of scholastic metaphysics. In the Franciscan era (from the expulsion of the the Jesuits until 1808) at the same university, modern figures such as Descartes were discussed, but not always with the approval of their doctrines. Meanwhile, in Buenos Aires education at the San Carlos School was kept on a more traditional level.

2 The nineteenth century

Generally, in all Latin American countries after their declaration of independence, there was a long period of political instability often accompanied by bloody civil wars. Argentina was no exception. However, philosophy was not absent in the fifty years that followed the separation from Spain; that is, before the country was unified and began its march towards modernization and material progress, around 1860.

At the beginning of the nineteenth century there were calls to introduce the new philosophical and scientific currents into the curriculum. To begin with more practical studies linked to scientific progress were being requested. Such was the case of a new plan of studies prepared in 1813 by Gregorio Funes, rector of the University of Córdoba. Funes proposed limiting the duration of courses in logic and metaphysics to provide for an entire year dedicated to the study of mathematics, signalling a great innovation.

In 1819 at the School of the Union of the South (a continuation of the San Carlos School) Juan Crisóstomo Lafinur initiated a new era of teaching philosophy in Argentina. The content of that teaching was French ideology and classes began to be taught in Spanish. Juan Manuel Fernández de Agüero and Diego Alcorta continued the same trend until 1842 at the University of Buenos Aires. The philosophy taught was the sensualism of Etienne Bonnot de Condillac and the ideas of the French ideologues, especially Pierre-Jean Cabanis and Destutt de Tracy.

In the nineteenth century the first manifestation of a personal philosophical expression which was not derived from university teaching was the work of Juan Bautista Alberdi (1810-84), a public figure known for his contribution to the drafting of the Argentine constitution of 1853. Alberdi found ideology inadequate and was open to the influences of Cousin, Jouffroy, Lerminier and Leroux. He had a pragmatic view of philosophy in the sense of considering that philosophical studies in Latin American countries should be applicable to concrete problems of those countries: their art, laws, politics and industry. This provides the basis for a possible Spanish American philosophy in which a certain line of current Latin American thought finds an antecedent. The most important philosophical writing of Alberdi is his *Fragmento preliminar al estudio del derecho (Prologue to the Study of Law)* (1837). In this work he maintains that each society should address its own reality: the course of history has its own defined periods and each people should know the one in which they find themselves to apply authentic rather than imitative solutions to their problems. *Fragmento preliminar al estudio del derecho (Prologue to the Study of Law)* is also a philosophy of law that deals with natural and positive law and the theory of jurisprudence. The theoretical framework is borrowed primarily from Jouffroy and Lerminer.

In the nineteenth century other European currents were also present, like French eclecticism and Krausism. Eclectic spiritualism was represented in the secondary teaching of philosophy by French manuals from that school. Krausism, whose presence became notable around 1870, was adopted in the fields of pedagogy and law (see Krause, K.C.F.).

3 The twentieth century

The first Argentine philosophical movement which was widely disseminated and involved various major figures was positivism. In the Argentine case, this term implied that which was understood in Europe under that name (the doctrines of Comte and Spencer), but also the influence of Darwinian biological evolutionism and a high regard for the type of explanation provided by the natural sciences. The combination is more of a positivist or scientific naturalism than a positivism in the strict sense (see Positivist thought in Latin America).

There were positivist followers of Comte in Argentina, such as Alfredo Ferreira (1836-1938), who founded the Argentine Positivist Committee in 1924. Ferreira was principally an educator and despite adhering to the religion of humanity, his association with Comte was not dogmatic.

Darwinism was represented principally by the paleontologist Florentino Ameghino (1845-1911), who culminated vast scientific work in a brief text, 'Mi credo' (My Creed) (1915), in which he expressed his materialist and scientistic *Weltanschauung*.

The greatest individual figures of Argentine positivism are Carlos Octavio Bunge (1875-1918) and José Ingenieros (1877-1925), especially the latter. Bunge is not easily categorized in a particular positivist orientation. He applied his points of view to three principal themes: psychology, ethics and law. The ultimate tenets of his position in the three cases are biological. Ingenieros was a psychologist and psychiatrist of international renown

besides being a philosopher. He defended a moral idealism in his works, widely read throughout Latin America, such as *El hombre mediocre (Mediocre Man)* (1913a) and *Las fuerzas morales (Moral Forces)* (1925). His two principal philosophical works are *Principios de psicología (Principles of Psychology)* (1913b) and *Proposiciones relativas al porvenir de la filosofía (Propositions Relative to the Future of Philosophy)* (1918). The psychology of Ingenieros is of a naturalist nature and specifically biological, with clear application of Darwinian theory. In his work biology is explained in terms of energy, following Wilhelm Ostwald (1853-1932). This position did not prevent Ingenieros from considering that authentic philosophical problems were outside the field of science. In this sense, in his aforementioned *Proposiciones relativas al porvenir de la filosofía (Propositions Relative to the Future of Philosophy)*, he identifies philosophy with metaphysics, understanding philosophy as the realm of hypotheses that go beyond experience, but depart from it and cannot contradict it. This thesis represents one of the first expressions of the transition towards the next phase of the surpassing of positivism, which developed in the warm glow of new European philosophical trends, through forms of transition represented by Fouillé and Guyau, for example, and later through the enormous prestige of Bergson and Croce. In this transition two factors were pivotal: the founding of the School of Philosophy and Humanities of the University of Buenos Aires in 1896 and the first visit of José Ortega y Gasset to Argentina in 1916, with his prestige and the dissemination of the new European currents.

Following the example of Rodolfo Rivarola in the newly founded School of Philosophy and Humanities, who was interested in the thought of Kant, the two primary figures who were influential in the change of philosophical atmosphere were Coriolano Albernini (1886-1960) and Alejandro Korn (1860-1936). The first had a decisive influence on university life; the second has become a classic of Latin American philosophy. The criticism by Albernini of positivism was caustic and effective. Korn constructed his philosophy on the basis of the freedom of ethical conduct and rejected in that field the determinism of positivist naturalism (see Anti-positivist thought in Latin America). For Korn, freedom makes possible the moral voluntarism that facilitates human excellence. In addition to Albernini and Korn, Alberto Rougès (1880-1947), from the interior of the country, wrote *Las jeraquías del ser y la eternidad (Hierarchies of Being and Eternity)* (1943), one of the most important works of Argentine philosophy in the first half of the century.

Around 1940 when positivism had clearly been left behind, a new phase of Argentine philosophy began that has lasted despite its lack of uniformity over the course of its development. Its principal characteristics amount to a considerable increase in philosophical activity, especially in the number of people engaged in philosophy, the number of university professorships and the number of journals and publications; philosophical activity being carried out primarily in universities; an increase in the quality of professional training of philosophical practitioners and an openness to all the currents and problems of Western thought.

The first part of this phase witnessed the influence of German philosophy from the first decades of the century: Husserl and his direct disciples of the phenomenological school, along with Scheler, Hartmann and Heidegger. Such was the case of Francisco Romero (1891-1962), who developed his own philosophical position stemming from Scheler and Hartmann. In his day, Romero was considered one of the major Spanish-speaking philosophers. His position culminated in the publication of his principal work, *Teoría del hombre (Theory of Man)* (1952), which is a philosophical anthropology based on the concept of intentionality. He also outlined the fundamentals of axiology, a theory of culture and a metaphysics of transcendence. Various figures can be linked to Romero: Risieri Frondizi (1910-83), one of the few members of this group who was influenced by English-language philosophy and whose principal work is a theory of values; Aníbal Sánchez Reulet, who examined philosophy as a problem in *Raíz y destino de la filosofía (The Root and Destiny of Philosophy)* (1942) and Eugenio Pucciarelli (1907-95), who investigated the problems of reason, time and technique.

Carlos Astrada (1894-1970) and Luis Juan Guerrero (1896-1956) also moved into the realm of German philosophy. Astrada developed an extensive body of work influenced by Heidegger and concluding in Marxism. Guerrero developed a detailed aesthetics. For his part, Carlos Cossio (1903-87) developed his ‘egological theory of law’ in *La teoría egológica del derecho y el concepto jurídico de libertad (The Egological Theory of Law and the Juridical Concept of Freedom)* (1944) derived from Husserl, Heidegger and Kelsen.

Close to existentialism one finds Vicente Fatone (1903-62) and Miguel Angel Virasoro (1900-66). Fatone was one of the most outstanding expositors of existentialist philosophy, but also dealt in Hindu doctrines and philosophy of

religion. Oriental thought as well as philosophy of religion have also been studied by subsequent researchers. Virasoro combines the influence of Hegel and that of existentialism. In addition, Angel Vassallo developed themes that came from the great Western metaphysicians and from Blondel and Marcel (see Existentialist thought in Latin America).

The work of Juan Adolfo Vázquez is characterized by a high degree of interest in metaphysics and an affinity for the mystical. Following Fatone, Víctor Massuh continued the study of philosophy of religion, while Adolfo Carpio of Romero's group, authored a work of Heideggerian bent about the significance of the history of philosophy.

With respect to Catholic thought, the Thomist thinker of the most substantive work and the greatest renown is Octavio Derisi. An important Thomist figure in Córdoba was Nimio de Anquín, who wrote about ancient and medieval philosophy as well as contemporary thought. Two Jesuits who came later and who produced considerable bodies of work are Juan Sepich and Ismael Quiles. In the context of the great dissemination of existentialist philosophy, the latter developed his theory of 'in-sistentialism'. Similar to Nimio de Anquín, although a generation later, is Alberto Caturelli, who developed a metaphysics opposed to modern and contemporary immanentism. One should also note Diego Pró, who has worked with Aristotelian thought, and like Caturelli is the author of numerous studies about philosophy in Argentina.

Moving away from the predominance of French and German influences, a group of followers of English-speaking analytic philosophy was formed in the early 1970s. In 1972 the Argentine Society of Philosophical Analysis was founded, which brought together philosophers of varying thematic interests but with great *esprit de corps* and keen awareness of their difference from representatives of other movements. This group is in close contact with its counterpart in Mexico, which publishes the journal *Crítica*. This movement is characterized by a distrust of metaphysics, an interest in the accomplishments of science and in linguistic problems and an attempt to produce rigorous conceptual analyses. Similarly opposed to the majority of other currents in Argentina, but a thinker who is an independent figure of considerable international prestige, is the philosopher of science, Mario Bunge, who settled in Canada (see Analytical philosophy in Latin America).

At the beginning of the 1970s the philosophy of liberation movement began in Argentina and later spread throughout Latin America. It was considered by many to be the most authentic expression of Latin American thought. This style of philosophy sought to be distinguished from traditional academic forms, to hold that philosophy is justified only by immediate engagement with praxis and to maintain as its principal focus the question of the dependence of Latin America. This philosophy promotes the liberation of the region and that of its most overlooked sectors (the masses, the oppressed, the poor). It has natural affinities with theology of liberation. The first synthesizing overview of the movement was the work of Horacio Cerutti Guldberg, *Filosofía de la liberación latinoamericana (Latin American Philosophy of Liberation)* (1983). The criticism that the book contains about other members of the movement and the responses of the latter to the criticisms, illustrate the internal differences in the movement. However, aside from those differences, philosophy of liberation taken as a whole has a well-defined profile at the forefront of Argentine and Latin American thought (see Liberation philosophy; Liberation theology).

Dialectical materialism did not produce theoretical works of consequence in Argentina, but the founder of the Argentine Socialist Party and translator of Marx, Juan Justo (1865-1928) combined Marxism with biology and Darwinism in his work. Aníbal Ponce (1898-1938), who was a disciple of Ingenieros, moved from the biological orientation of his teacher to a Marxist position.

Recent philosophical activity is characterized by an intensification of the themes indicated as belonging to the post-1940 era. The international philosophical movement is experienced with first-hand immediacy and many philosophers are in a position to participate fully in international philosophical dialogue. In some cases, affinities of ideas and philosophical positions are reflected in institutional groupings or in publications that represent them. For example, the journal *Análisis Filosófico (Philosophical Analysis)* is for the analytic group, the Centre for Philosophical Studies is affiliated with the *Revista latinoamericana de filosofía (Latin American Journal of Philosophy)*, the National Academy of Sciences with *Escritos de filosofía (Philosophical Writings)*. This philosophical activity is not carried out solely in Buenos Aires, but also in various state and private universities in the rest of the country.

Part of the historical-philosophical task is devoted to Argentine and Latin American thought. The two classics of the history of ideas in Argentina are *La evolución de las ideas argentinas (Evolution of Argentine Ideas)* (1918-20) by José Ingenieros and *Influencias filosóficas en la evolución nacional (Philosophical Influences on National Evolution)* (1936) by Alejandro Korn. The most recent and complete panorama of the development of philosophy in the country is contained in the work of Francisco Leocata, *Las ideas filosóficas en Argentina (Philosophical Ideas in Argentina)* (1992, 1993). Among the outstanding authors who specialize in Argentine and Latin American thought are Gregorio Weinberg (1995) (history of education and science) and Arturo Roig (1981). The work of the latter, in addition to its historiographic importance, is considered a major contribution to the philosophical interpretation of Latin America. There are professorships of Argentine thought at the universities of Buenos Aires, La Plata, Córdoba, Tucumán and Mendoza where the journal *Cuyo: Anuario de historia de la filosofía argentina y americana (Cuyo: Journal of the History of Argentine and Latin American Philosophy)* has been published since 1965.

Since the beginning of the century, various foreign professors have contributed to the formation of contemporary Argentine philosophy. Perhaps the most notable is the Italian historian of philosophy Rodolfo Mondolfo.

JUAN CARLOS TORCHIA ESTRADA

References and further reading

Alberdi, J.B. (1837) *Fragmento preliminar al estudio del derecho (Prologue to the Study of Law)*, Buenos Aires: Facultad de Derecho y Ciencias Sociales, Universidad de Buenos Aires, 1942.(Alberdi maintains that each society should address its own reality.)

Ameghino, F. (1915) 'Mi Credo' (My Creed), in *Doctrinas y descubrimientos (Doctrines and Discoveries)*, Buenos Aires: La Cultura Argentina.(Written in 1906, but published posthumously.)

Biagini, H. (ed.) (1985) *El movimiento positivista argentino (The Positivist Movement of Argentina)*, Buenos Aires: Editorial de Belgrano.(Collective work which includes numerous aspects of the positivist movement.)

Cappelletti, A.J. (1995) *Filosofía argentina del siglo XX (Argentinian Philosophy of the Twentieth Century)*, Rosario: Facultad de Humanidades y Artes, Universidad Nacional de Rosario.(Provides a survey of twentieth-century philosophy and contains good essays on Alejandro Korn, Alberto Rougès, Alfredo Franceschi, Francisco Romero, Risieri Frondizi and others.)

Cerutti Guldberg, H. (1983) *Filosofía de la liberación latinoamericana (Latin American Philosophy of Liberation)*, Mexico: Fondo de Cultura.(The first work to synthesize the philosophy of liberation.)

Cossio, C. (1944) *La teoría egológica del derecho y el concepto jurídico de libertad (The Egological Theory of Law and the Juridical Concept of Freedom)*, Buenos Aires: Editorial Losada.(The egological theory of law, the author's own theory, is first propounded in this book.)

Gracia, J.J.E. (ed.) (1986) *Latin American Philosophy in the Twentieth Century*, Buffalo, NY: Prometheus Books. (An anthology including the works of Francisco Romero, Risieri Frondizi, Carlos Astrada, Alejandro Korn and Arturo Roig.)

Ingenieros, J. (1913a) *El hombre mediocre (Mediocre Man)*, Madrid: Editorial Renacimiento.(Moral idealism is defended in this work.)

Ingenieros, J. (1913b) *Principios de psicología (Principles of Psychology)*, Madrid: Jorro. (One of the author's main philosophical works.)

Ingenieros, J. (1918) *Proposiciones relativas al porvenir de la filosofía (Propositions Relative to the Future of Philosophy)*, Buenos Aires: Talleres L.J. Rosso.(Philosophy is identified with metaphysics.)

Ingenieros, J. (1918-20) *La evolución de las ideas argentinas (Evolution of Argentine Ideas)*, Buenos Aires: Losada.(One of the classic books on the history of ideas.)

Ingenieros, J. (1925) *Las fuerzas morales (Moral Forces)*, Buenos Aires: Talleres L.J. Rosso. (Widely read throughout Latin America.)

Korn, A. (1936) *Influencias filosóficas en la evolución nacional (Philosophical Influences on National Evolution)*, Buenos Aires: Editorial Claridad.(A classic in the area of the history of ideas.)

Leocata, F. (1992, 1993) *Las ideas filosóficas en Argentina (Philosophical Ideas in Argentina), Buenos Aires: Centro Salesiano de Estudios*, vols 1 and 2.(Covers the colonial period to the twentieth century. Although volumes of this work are still forthcoming, this is the most complete panorama to date.)

Roig, A. (1981) *Teoría y crítica del pensamiento latinoamericano (Theory and Critique of Latin American Thought)*, Mexico: Fondo de Cultura.(A major contribution to the philosophical interpretation of Latin America.)

Romero, F. (1952) *Teoría del hombre (Theory of Man)*, Buenos Aires: Losada.(The outline of metaphysics is developed.)

Rougès, A. (1943) *Las jeraquías del ser y la eternidad (Hierarchies of Being and Eternity)*, Tucumán: Facultad de Filosofía y Letras, Universidad Nacional de Tucumán. (One of the most important works of Argentine philosophy.)

Sánchez Reulet, A. (1942) *Raíz y destino de la filosofía (The Root and Destiny of Philosophy)*, Tucumán: Facultad de Filosofía y Letras, Universidad Nacional de Tucumán.(The problematics of philosophy are examined.)

Sánchez Reulet, A. (1954) *Contemporary Latin American Philosophy*, Albuquerque, NM: University of New Mexico Press.(Selections of texts with useful introductions to authors, such as, Korn, Ingenieros and Romero.)

Torchia Estrada, J.C. (1961) *La filosofía en la Argentina (Philosophy in Argentina)*, Washington, DC: Pan American Union.(The sections on ideology, Alberdi and Ingenieros and the anti-positivist revolt are still relevant.)

Weinberg, G. (1995) *Modelos educativos en la historia de América Latina (Educational Patterns in the History of Latin America)*, Buenos Aires: Editora AZ.(An historical study of the Latin American education system.)

Aristippus the Elder (*c*.435-*c*.355 BC)

Aristippus of Cyrene was a member of Socrates' entourage who after Socrates' death (399 BC) founded the Cyrenaic school. He was primarily interested in practical ethics. He focused on the concepts of pleasure and pain, and classed them as bodily motions of which we are conscious. He considered pleasure a major component of happiness, but also attributed intrinsic value to virtue and emphasized the importance of study and exercise as means to self-control.

Aristippus, son of Aritades, was born at Cyrene. He moved to Athens at a young age, where he became an associate of Socrates and also frequented Sophistic circles (see Sophists). After Socrates' death, he taught philosophy during his numerous travels to Syracuse, Asia Minor, Corinth, Megara and Aegina, and probably returned to Cyrene towards the end of his life. It may have been there that he founded the Cyrenaic school (see Cyrenaics) and that he trained his daughter, Arete, in philosophy. He must be distinguished from his grandson and second successor as head of the school, Aristippus the Younger (born *c*.380 BC), with whom he is occasionally confused.

Some sources maintain that he wrote nothing at all, perhaps in deliberate imitation of Socrates. But our most reliable informants report that he wrote several dialogues and a history of Libya in three books. The titles attest a vivid interest in practical ethics, but also a concern for rhetoric, history, literature, linguistic morphology and semantics. His argument establishing that no morally neutral object exists may be taken to indicate an additional interest in theoretical ethics. He disallowed the study of mathematics and of physics, but recognized that logic may be useful in so far as it sharpens our dialectical capacities.

Few tenets can be ascribed to Aristippus with confidence. It seems that he had no detailed epistemological views; the Cyrenaic theory of knowledge was probably formulated by his grandson. However, he focused on the notions of pleasure and pain which he classed as *pathē*, ways of being affected, and described them in terms of smooth and rough motions of the flesh which result in sensation. These *pathē* are short-lived due to their kinetic nature: the motions associated with them expire with time. To this extent, Aristippus foreshadows Cyrenaic subjectivism, which equated the *pathē* experienced in the present moment with bodily motions detected by 'internal touch', self-evident and incorrigibly known.

Regarding Aristippus' ethics, the evidence is divided. According to one tradition, he was a hedonist who defined bodily pleasure experienced in the present moment as the moral end, and thus formulated the core of the ethical doctrine held by his grandson and by the orthodox branch of the Cyrenaic school (see Hedonism). However, according to a second group of sources, he never defended a particular moral doctrine but only gave the impression of being a hedonist on account of his numerous lectures on pleasure and of his voluptuous life. There are good reasons for accepting this latter testimony: Aristippus attributed intrinsic value to virtue, and this is incompatible with a straighforward hedonism positing the short-lived pleasures of the present moment as the only intrinsic good; he emphasized the importance of study, exercise and self-control, all of which have non-pleasurable aspects; he recommended the enjoyment of bodily pleasures only if they do not endanger one's control over oneself; and he tended to look at his life as a single whole, not as a series of discrete pleasurable episodes, as is exemplified by his assertion that he aimed 'at the easiest and pleasantest life'. Such evidence indicates that Aristippus was a eudaemonist who considered pleasure the major component of happiness and spoke of happiness in terms of a pleasurable condition extending over one's lifetime (see Eudaimonia).

This conception of happiness influenced his social and political attitudes, which are marked by a strict individualism. He claimed for himself a freedom which entailed the rejection of the political condition of citizenship and abstention from duties and obligations to any particular city. In his view, part of the happiest life is 'to live as a foreigner in every land'.

VOULA TSOUNA

References and further reading

Classen, C.J. (1958) 'Aristippos', *Hermes* 86: 182-92.(Standard reference article in German.)
Doering, K. (1988) *Der Sokratesschüler Aristipp und die Kyrenaiker (Socrates' Pupil Aristippus and the*

Cyrenaics), Wiesbaden and Stuttgart: Akademie Verlag.(The most recent book-length study on Aristippus and the Cyrenaic school.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae (The Fragments of Socrates and the Socratics)*, Naples: Bibliopolis. (Volume 2 includes the fullest available collection of Greek and Latin testimonies on Aristippus.)

Xenophon (*c*.360s BC) *Memoirs of Socrates*, trans. H. Tredennick (1970), revised R. Waterfield, Harmondsworth: Penguin, 1990.(Aristippus appears as one of Socrates' interlocutors in II 1 and III 8.)

Ariston of Chios (early to mid 3rd century BC)

The Greek philosopher Ariston (alternatively Aristo), from the Aegean island of Chios, was an exceptionally independent-minded member of the early Stoic school. A pupil of the founder Zeno of Citium, he was among the most prominent philosophers working at Athens in the mid-third century BC. He concentrated on ethics, dismissing logic and physics as irrelevant.

Like many contemporary philosophers, including Zeno, Ariston undoubtedly saw his own views as the ones most authentically capturing those of Socrates. Virtue he considered a unitary intellectual state, its conventional fragmentation into kinds being misleading at best. He resisted Zeno's doctrine that nonmoral desiderata like health, although indifferent, were naturally 'preferable'. Total indifference to them, rather than rationally choosing between them, was the true goal of life. He rejected rules of conduct - much favoured by Zeno - as founded on the same mistake of treating indifferent things as if they could be ranked in terms of intrinsic values.

1 Life and work

Ariston of Chios was, during much of the third century BC, as important a figure in the Stoic school at Athens as Zeno of Citium, its official founder and his own teacher. But very little is known about Ariston's life. Probably he was Zeno's pupil in the early years of the third century, outliving him by a substantial period and perhaps well into the second half of the century.

Ariston's interests were almost entirely ethical, although there is some evidence of an interest in poetics, and possibly even in grammar. He stayed much closer to Cynic ethics (see Cynics) than Zeno himself did, and in the long run it was the differences between the two that were most noticed. The later Stoic tradition chose to revere Zeno but not Ariston, and, because history is written by the winners, Ariston has come to be seen with hindsight as a marginal and heretical figure. This was certainly not so in his own day, when his impact at Athens was enormous. For example, Arcesilaus, who had led the Academy into its sceptical phase (see Academy), appears to have engaged in debate with Ariston at least as much as with Zeno. Ariston's own pupils included a leading Stoic Apollophanes, and the celebrated scientist Eratosthenes. Ariston was an acute observer of contemporary philosophy, and his mock-Homeric line of verse about Arcesilaus as a philosophical chimera ('Plato in front, Pyrrho behind, Diodorus in the middle') became famous (see Arcesilaus §1).

What is harder to know is how far the later tradition may have exaggerated or even invented doctrinal differences between Ariston and Zeno. But there seems little doubt that some of these were real. During Zeno's lifetime, open disagreement with him was apparently acceptable in the school. After Zeno's death (262), however, his thought became canonized: Ariston's independence now began to look like heresy. It was probably at this stage that he set up his own school, said to have been in the Cynosarges gymnasium outside the city walls of Athens.

Ariston is reported to have written a number of works. The titles of fourteen are listed by Diogenes Laertius in a short biography of him. Some later Stoics disputed the authenticity of most of these, attributing them instead to the Peripatetic Ariston of Ceos, with whom Ariston of Chios certainly is sometimes confused in ancient sources. However, there is a good chance that all or most are genuine. They include *Chreiai* (a typically Cynic collection of moral anecdotes), *Dialogues*, *On Zeno's Doctrines*, *Against the Orators* and *Against the Dialecticians*. A further work, *Comparisons*, was a collection of Ariston's graphic philosophical similes (look out for some below).

Ariston objected to Zeno's tripartite philosophical curriculum, consisting of logic, physics and ethics. Echoing Socrates (Plato, *Apology* 19b-c), he dismissed physics as 'above us'. Dialectic, the formal study of argument, and hence a mainstay of the branch called 'logic', Ariston likened to cobwebs (technically complex, but useless), to quicksand, and to eating crabs (lots of bones, little nutriment). He judged these parts of philosophy - as also much of the traditional Greek educational curriculum - an irrelevance to what really mattered in life: moral knowledge. Consistently with this, his own important contributions were in ethics.

2 Indifference

Socrates had argued that most things conventionally judged good or bad, such as health and illness, are in themselves neither, since they only get their value from the way they are used. Health and sickness wisely used are

great goods, while unwisely used they are great evils. The consequence, that only virtue or wisdom is good, only vice or folly bad, while everything else is morally 'indifferent', was widely endorsed in the fourth-century BC Socratic tradition, including Cynicism. Zeno's innovation (see Stoicism §15) was to rank the indifferents on a scale of natural preferability, while continuing to call them indifferent. He also, in consequence, attached enormous importance to rules or 'precepts', as offering indispensable guidelines to a proper choice among the indifferents, and thus a start towards the goal of 'living in accordance with nature'.

Ariston fought a rearguard action against this dilution of Socratic/Cynic values: indifferents really are just that, indifferent. He denied rules any moral value. And in place of Zeno's formulation of the goal, he described it as 'living with a disposition of indifference towards what is intermediate between vice and virtue, not retaining any difference at all within that class of things, but being equally disposed towards them all'. He is said to have formed his view when, during Zeno's illness, he attended the lectures of Zeno's old Platonist teacher Polemo (see Platonism, Early and Middle §1). Paradoxically, Polemo himself was attacking Zeno from the other direction, for refusing to call such items as health and illness 'good' and 'bad'. No doubt Polemo argued for a straight choice: either such items are good and bad, or they cannot be valued relatively to each other at all. While Polemo took the former option, Ariston was persuaded by the latter.

Ariston's fundamental objection to rules of conduct seems to have been as follows. Typical precepts ('Don't get drunk', 'Look after your health', 'Avoid enslavement' and so on) misleadingly attach preferential value to items such as sobriety, health and freedom, which are in themselves totally indifferent and are worthy of choice purely according to circumstances. It was as if in spelling one were to favour some letters as intrinsically preferable to others. Rules of conduct, he concluded, have no place outside the kindergarten.

Ariston's rejection of rules has led some to call him an 'intuitionist'; and he is indeed reported as saying the sage will do 'whatever comes to mind' (Cicero, *On Ends* IV 43). But we also know that, as a Socratic, he held that virtue is an intellectual state - knowledge of good and bad (although also called 'health', that is, of the soul) - and that in place of rules he recommended reliance on 'doctrine'. He therefore must have thought correct moral decisions were arrived at by reasoning, not intuition. The serious problem that confronted him was, rather, how non-moral decisions should be made. If public image, comfort and so on are literally indifferent, how can we make such trivial decisions as which clothes to wear or which food to eat? Yet if we could not choose, we would wear and eat nothing. It was almost certainly to answer this that Ariston's appeal to intuition came in. If the objection was that the Aristonian sage will, like Buridan's ass, starve to death through inability to choose between indifferents, Ariston's answer was that in such situations it is rational simply to do the first thing that comes into your head.

3 Virtue

Since virtue is simply a matter of knowing good and bad, it becomes hard to see how it can have distinct parts, species or branches. In the wake of Socrates (see especially Plato's *Protagoras* and Socrates §5), every Socratic philosopher defended some version of his thesis of the unity of the virtues. To some (for example, Chrysippus: see Stoicism §16) this meant no more than their inseparability, but Ariston was one of those who took them to be literally one and the same thing. Why then do they have different names? His answer was that the very same state of mind was named differently according to the circumstances in which it was located. In situations of danger, for example, knowledge of good and bad was called 'courage'; in situations involving appetite, 'self-discipline' (*sōphrosynē*), and so on. It was, he said, as if one were to vary between calling the power of eyesight 'white-seeing' and 'black-seeing', according to the objects it happened to be confronted with. Chrysippus devoted a treatise to refuting Ariston on this issue.

The clear implication is that the different species names of virtue are superficial and accidental, misleadingly fragmenting what is in reality a unitary intellectual power. This disdain for conventional distinctions was characteristic of Ariston, and did much to give his pronouncements their strongly Cynic flavour. Consider his striking assertion of cosmopolitanism: 'A native land does not exist by nature, any more than does a house, a field, a smithy or a doctor's surgery. Each one of these comes to be so, or rather is so named and called, always in relation to the occupant and user.'

DAVID SEDLEY

References and further reading

Arnim, H. von (1903-5) *Stoicorum Veterum Fragmenta (Fragments of the Early Stoics)*, Leipzig: Teubner, with vol. 4, indexes, by M. Adler, 1924.(The standard collection of early Stoic fragments, in Greek and Latin, commonly abbreviated as *SVF*; Ariston is in volume 1, 75-90.)

Boys-Stones, G. (1996) 'The *epeleustikēdynamis* in Aristo's Psychology of Action', *Phronesis* 41: 75-94.(Defends a partly different view on Ariston's 'intuitionism' from that in §2 of the entry on him.)

Diogenes Laertius (early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols. (VII 160-4 in volume 2, is his life of Ariston.)

Ioppolo, A.M. (1980) *Aristone di Chio e lo stoicismo antico (Ariston of Chios and Early Stoicism)*, Naples: Bibliopolis.(The classic study, in Italian, of Ariston's work and his place in Stoicism.)

Long, A.A. (1988) 'Socrates in Hellenistic Philosophy', *Classical Quarterly* 38: 150- 71; repr. in A.A. Long, *Stoic Studies*, Cambridge: Cambridge University Press, 1996, 1-34. (Important on Zeno's and Ariston's rival interpretations of Socrates on moral indifferents.)

Schofield, M. (1984) 'Ariston of Chios and the Unity of Virtue', *Ancient Philosophy* 4: 83-96.(Subtle evaluation of Ariston's place in this debate.)

Seneca (*c.* AD early 60s) *Letters to Lucilius*, trans. R.M. Gummere, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1917- 25.(Latin text with English translation. Letters 94-5 place Ariston on one side of an extended debate about moral rules; but there is less genuinely Aristonian material here than is usually assumed.)

Aristotelianism in Islamic philosophy

In Arabic, Aristotle was referred to by name as Aristutalis or, more frequently, Aristu, although when quoted he was often referred to by a sobriquet such as 'the wise man'. Aristotle was also generally known as the First Teacher. Following the initial reception of Hellenistic texts into Islamic thought in al-Kindi's time, al-Farabi rediscovered a 'purer' version in the tenth century. In an allusion to his dependence on Aristotle, al-Farabi was called the Second Teacher. Ibn Rushd, known in the West as Averroes, was the last great Arabophone commentator on Aristotle, writing numerous treatises on his works. A careful examination of the Aristotelian works received by the Arabs indicates they were generally aware of the true Aristotle. Later, transmission of these works to Christian Europe allowed Aristotelianism to flourish in the scholastic period.

We should not take at face value the Islamic philosophers' claims that they were simply following Aristotle. The convention in Islamic philosophy is to state that one is repeating the wisdom of the past, thus covering over such originality as may exist. There was a tendency among Islamic philosophers to cite Aristotle as an authority in order to validate their own claims and ideas.

1 Early influence

Among the major differences between the Islamic philosophers and Aristotle are the questions of the eternity or creation of the world, the nature of Being and a real-world distinction between essence and existence. The ninth-century philosopher al-Kindi used Aristotle, in Arabic translations, as a base for his own philosophical works. Among other works, al-Kindi wrote one treatise specifically dealing with Aristotle, *Fi kammiya kutub Aristutalis wa ma yahtaj ilahi fi tahsil al-falsafa (The Quantity of the Books of Aristotle and What is Required for the Acquisition of Philosophy)*. The early part of this treatise gives an accurate summary of various logical works by Aristotle, such as *Categories* and *De Interpretatione*, before diverging into a decidedly non-Aristotelian précis on questions found in the physical treatises. Though he owed a large debt to Aristotelian thought, al-Kindi parted company with Aristotle in espousing the idea of creation from nothing by a Creator. Furthermore, in writing about creation al-Kindi does not ascribe the idea to Aristotle in the text of his treatises. Debate still continues over whether al-Kindi should be considered more 'Platonic' or 'Aristotelian'.

An inventory of those works attributed to Aristotle which were available to early Islamic philosophers appears in the *Fihrist* of the tenth-century bibliographer Ibn al-Nadim. The work known as the *Theology of Aristotle* appears in the *Fihrist*, although it is mentioned only in passing. Greater attention is paid to other correctly attributed works of philosophy and logic by Aristotle, including such detailed information as the translator, the number of sections and the work's Arabic commentators, suggesting the relative importance of these works to Ibn al-Nadim and his audience. Judging from the list available in the *Fihrist*, the Islamic philosophers would have been able to appreciate Aristotle's logic, physics and metaphysics. However, since the *Theology of Aristotle* was really a Neoplatonic work (see Neoplatonism in Islamic philosophy §1), based on the *Enneads* of Plotinus, accepting this attribution would have obscured understanding of Aristotle. Whatever the influence of Aristotle on Islamic philosophy, the Muslims were nonetheless obliged to work out for themselves certain underlying issues, such as conceptualizing ideas in their own language. In particular, they had to implement a philosophical terminology, as there was a lack of abstract nouns in Arabic.

2 Middle stage: Ibn Sina and al-Farabi

The Islamic philosophers picked and chose from Aristotle's texts, using him as an authority when it suited their purposes, and knowing that philosophy was a 'foreign science' in need of an external authority as it lacked an indigenous authority. While aspects of Avicennan philosophy continue the Aristotelian tradition in broad terms, Ibn Sina's ideas about the Necessary Existent and the Possible Existent do not have their antecedents in Aristotle's philosophy (see Ibn Sina). However, as Ibn Sina himself hailed from Khurasan, one cannot dismiss the possible influences of Buddhism, Zoroastrianism and Hinduism on his philosophy. The differences with Aristotle go back to the fact that the philosophers are writing in an Islamic milieu, and certain changes had to take place to correlate with the religious ideology. Seen in some lights, these changes may be considered peripheral; the philosophers continued to hold a solidly Aristotelian view of such basic ideas as the relationship of form and matter.

Among the scholars of the Middle Period - the fourth and fifth centuries AH (tenth and eleventh centuries AD) - al-Farabi is considered the foremost Aristotelian, and was indeed known as the Second Teacher (Aristotle himself being the First Teacher). Some scholars have divided his works into those which admit Aristotelian influence, such as *Kitab al-huruf (The Book of Letters)*, and more popular works, such as *Kitab fi mabadi' ara' ahl al-madina al-fadila (The Book of the Principles of the Opinions of the People of the Virtuous City)*, usually known simply as *al-Madina al-fadila (The Virtuous City)*, a utopian treatise which espouses Neoplatonic theories such as emanation, in which everything is said to flow from the One. His internalization of Aristotle is apparent in his treatment of the four causes in *Tahsil al-sa'ada (The Attainment of Happiness)*, echoing those found in *al-Tabi'a (The Physics)*. Here he shows a complete familiarity with the Aristotelian idea of the four causes, but is equally willing to propound his own interpretation, preferring the word *mabadi'* (literally, principles) rather than *asbab* (causes), which was the translator Ibn Ishaq's choice.

On another important point, however, al-Farabi is not conceptually Aristotelian. In *al-Madina al-fadila*, we do not find the long discourses on the inherent weakness of women, children and slaves, which are found in Aristotle; rather, he distinguishes the inhabitants of the virtuous city from those of the ignorant cities by their moral character.

Al-Farabi considered his *Kitab al-huruf*, which takes its title from the Greek letters which entitle Aristotle's chapters, to be a commentary on Aristotle's *Metaphysics*. While *al-Huruf* is inspired by Aristotle's concerns, and deals with many of the same subjects, it does not slavishly imitate or even follow the order of the *Metaphysics*. Al-Farabi also believed in the ultimate harmony of the opinions of Plato and Aristotle, a difficult notion for many philosophers today to accept.

One might ask why Ibn Sina (Avicenna) would be taken in by a false treatise, the *Theology of Aristotle*, when he had such a good command of Aristotelian concepts that he could quote accurately from memory. In his 'Letter to Kiya', Ibn Sina expresses doubt about the authorship of the *Theology of Aristotle*, remarking that the text is 'somewhat suspect' (Gutas 1988). The tone of his discussion indicates that while he included this work with other Aristotelian treatises, he has by no means concluded it is genuinely an Aristotelian text. On the other hand, in the *Danashnama-i 'ala'i (The Book of Knowledge for 'Ala')*, his account of metaphysics, Ibn Sina derives a quotation from Aristotle where he claims that Aristotle describes the First Being as having complete happiness in itself. It is uncertain to which part of the *Metaphysics* Ibn Sina is referring, as such a passage does not appear to exist.

Elsewhere, Ibn Sina claims to quote Aristotle from memory when discussing the theory of definition for his treatise on Definitions, when he suggests that in the *Topics*, Aristotle defines definition as 'a statement indicating the quiddity of a thing'. This is an exact quotation. It is remarkable that Ibn Sina appears to remember Aristotle's important ideas word for word after having, he says, read the books only once and thereafter being unable to refer to them. Given his life as a wanderer, this statement is credible.

3 Late period: the legacy of Aristotelianism

Unlike the Islamic east, where a Hellenistic tradition of philosophy flourished from the ninth century, philosophy reached al-Andalus later. Ibn Bajja, known as Avempace in Latin, was one of its first practitioners, active in the early part of the twelfth century. His heavily Aristotelian commentaries on the logical works of al-Farabi still survive. The socio-historian Ibn Khaldun ranked him with Ibn Rushd, and Ibn Bajja no doubt influenced this, the most famous, philosopher of Muslim Spain.

Ibn Rushd, better known in the West as Averroes, is considered not nearly as influential in the Islamic world as he was in medieval Europe. Here, either because he lived on the Western periphery of the Islamic world or because he wrote such extensive commentaries on Aristotle, he became renowned (see Averroism §1). Latin translations of Ibn Rushd's texts were available in Europe within a century of his death. Coming from a family of eminent jurists, Ibn Rushd had legal as well as philosophical training. He wrote commentaries on a wide range of Aristotle's works, including his *Physics*, *Metaphysics*, *Book of the Soul*, *On the Heavens* and *Posterior Analytics*, the last dating from 1170. In both long and intermediate commentaries as well as short paraphrases, Ibn Rushd tried to analyse the extent of his Islamic predecessors' deviation from Aristotle. He also exerted himself in reconciling religion and philosophy in his *Fasl al-maqal (Decisive Treatise On the Harmony of Religion and Philosophy)*. He discovers a duty to reflect with the intellect on existing beings and to seek knowledge in the Qur'anic injunction

found at Surah 49: 2: 'Consider, you who have vision.'

There is good reason to consider another of Ibn Rushd's works, the *Tahafut al-tahafut (The Incoherence of the Incoherence)* to be an attack on Neoplatonism and a defence of true Aristotelianism. On the question of the origin of the world, Ibn Rushd promulgated eternal creation but did not accept emanation. While he wrote the *Tahafut* primarily as a rebuttal of al-Ghazali's attack on the philosophers, he also disagreed with Ibn Sina's ideas about necessity.

Ibn Rushd was also to be the last in the line of Islamic Aristotelians. Throughout the classical period of Islamic thought, there were always some thinkers who distrusted rationalism and logic, certain that the study of philosophy results in a loss of faith. Al-Ghazali and Ibn Taymiyya are the two best known examples. Al-Ghazali studied philosophy to be able to rebut it; he suggested that knowledge is inferior to faith, as knowledge could not overcome doubts. His *Tahafut al-falasifa (The Incoherence of the Philosophers)* had a lasting influence. Here al-Ghazali attacked Aristotle and his followers, al-Farabi and Ibn Sina, particularly objecting to the Aristotelian notion of the eternity of the world, which he found irreconcilable with the Qur'anic description of God's creation of the world from nothing. Al-Ghazali also saw this as an idea which limited God in a totally unacceptable manner. Two centuries later, Ibn Taymiyya wrote *al-Radd 'ala al-mantiqiyyin (Against the Logicians)* as an attack on the method of definition and demonstration used by the philosophers who were influenced by Aristotle. He argued that logic is based on the faculty of human reason, which is necessarily inferior to divine revelation.

Despite the efforts of Ibn Rushd to rehabilitate philosophy, many scholars believe that Islamic philosophy never completely recovered from al-Ghazali's massive and brutal assault on it. In the Latin West, Islamic Aristotelianism was reincarnated as Averroism, that is, Aristotle's works as taught by Ibn Rushd and translated into Latin (see Aristotelianism, medieval §4; Averroism §1). His works also came to have great influence in Jewish philosophy, and for many years led to a strong strain of Aristotelianism among Jewish philosophers (see Averroism, Jewish). Aristotelianism continued to have an effect on Islamic philosophy through opposition to it from Illuminationist philosophy (see Illuminationist philosophy), and in particular thinkers such as al-Suhrawardi, al-Shahrazuri, Ibn Kammuna and others, often based in Persia. The latter sought to attack what they took to be the principles of Aristotelianism, especially its logical and ontological axioms, and produced critiques of Aristotelian essentialism which are sometimes quite similar to that of William of Ockham. It is accurate to say, however, that Aristotelianism as a school of philosophy in the Islamic world found no Muslim successors after the death of Ibn Rushd.

See also: Aristotelianism, medieval; Aristotle; Aristotle Commentators; al-Farabi; Greek philosophy: impact on Islamic philosophy; Ibn Rushd; Ibn Sina; Islam, concept of philosophy in; Logic in Islamic philosophy; Platonism in Islamic philosophy

KIKI KENNEDY-DAY

References and further reading

Abed, S. (1991) *Aristotelian Logic and the Arabic Language in Alfarabi*, Albany, NY: State University of New York Press.(A study on al-Farabi's use of Aristotle's works.)

Burrell, D.B. (1986) 'Essence and Existence: Avicenna and Greek Philosophy', *Mélanges de l'Institut dominicain d'études orientales du Caire* 17: 53-66.(An interesting discussion of Ibn Sina's distinction and his Greek forbears.)

El-Ehwany, A.F. (1963) 'Ibn Rushd', in M.M. Sharif (ed.) *History of Muslim Philosophy*, Wiesbaden: Harrasowitz, ch. 28: 540-64.(Summary of Ibn Rushd's philosophy.)

al-Farabi (*c*.870-950) *Kitab al-huruf (The Book of Letters)*, ed. M. Mahdi, Beirut: Dar al-Mashreq.(A new edition and translation is in progress, but no date has been set for its appearance.)

al-Farabi (*c*.870-950) *Alfarabi's Philosophy of Plato and Aristotle*, trans. M. Mahdi, New York: Free Press of Glencoe, 1962.(A reliable translation of three Farabian treatises: *Tahsil al-sa'ada (The Attainment of Happiness)*, *Falsafa Aflatun (Philosophy of Plato)* and *Falsafa Aristutalis (Philosophy of Aristotle)*.)

al-Farabi (*c*.870-950) *Sharh al-Farabi li-kitab Aristutalis fi al-'ibara*, ed. W. Kutch and S. Marrow, Beirut, 1960; trans F.W. Zimmerman, *Al-Farabi's Commentary and Short Treatise on Aristotle's De interpretatione*, Oxford: Oxford University Press, 1981.(Commentaries by al-Farabi on Aristotle.)

al-Ghazali (1058-1111) *Tahafut al-falasifa (The Incoherence of the Philosophers)*, ed. M. Bouyges and M. Fakhry, Beirut: Dar al-Mashreq, 4th edn, 1990; trans. S.A. Kamali, *Tahafut al-falasifah (Incoherence of the*

Philosophers), Lahore: Pakistan Philosophical Congress, 1963.(Al-Ghazali's attack on the philosophers.)

Gutas, D. (1988) *Avicenna and the Aristotelian Tradition*, Leiden: Brill.(A new interpretation of Ibn Sina's development as a philosopher.)

Ibn al-Nadim (before *c*.995) *Kitab al-fihrist*, ed. G. Flügel, Beirut: Khayyat, 1964; trans. B. Dodge, *The Fihrist of al-Nadim*, New York: Columbia University Press, 1970. (Early inventory of the works of Aristotle.)

Ibn Rushd (*c*.1170) *Averroes' Middle Commentaries on Aristotle's Categories and De Interpretatione*, trans. C.E. Butterworth, Princeton, NJ: Princeton University Press, 1983.(Contains the *Talkhis kitab al-maqulat (Middle Commentary on the Categories)* and *Talkhis kitab al-'ibara (Middle Commentary on De interpretatione)*. A detailed account of the approach by Ibn Rushd to these texts.)

Ibn Rushd (*c*.1179-80) *Kitab fasl al-maqal (Decisive Treatise on the Harmony of Religion and Philosophy)*, trans. G. Hourani, *On the Harmony of Religion and Philosophy*, London: Luzac, 1961.(Ibn Rushd's attempt to reconcile philosophy and religion.)

Ibn Rushd (*c*.1180) *Tahafut al-tahafut (The Incoherence of the Incoherence)*, ed. S. Dunya, Cairo: Dar al-Ma'aref, 1950, 2 vols; trans. S. Van Den Bergh, *Averroes' Tahafut al-tahafut (The Incoherence of the Incoherence)*, London: Oxford University Press, 1954, 2 vols.(Ibn Rushd's attack on al-Ghazali.)

Ibn Sina (980-1037) *Danishnama-i 'ala'i, ilahiyyat (Book of Knowledge for 'Ala')*, ed. M. Mu'in, Tehran, 1952; trans. P. Morewedge, *The Metaphysics of Avicenna (Ibn Sina)*, New York: Columbia University Press, 1973. (The metaphysical portion of the *Danashnama*.)

Ibn Taymiyya (1309-10) *al-Radd 'ala al-mantiqiyyin*, trans. W.B. Hallaq, *Ibn Taymiyya Against the Greek Logicians*, Oxford: Clarendon Press, 1993.(Translation with an introduction discussing in detail Ibn Taymiyya's objections to Aristotelian logic.)

al-Kindi (*c*.873) *Rasa'il al-Kindi al-falsafiyya (The Philosophical Treatises of al-Kindi)*, ed. M.A.H. Abu Ridah, Cairo: Dar al-fikr al-'arabi, 2 vols in 1, 1953.(The early edition is better than the recent reprint, but is hard to find. Many of the treatises have not been edited into European languages.)

Leaman, O. (1988) *Averroes and His Philosophy*, Oxford: Clarendon Press; repr. Richmond: Curzon, 1997.(A basic introduction to Ibn Rushd.)

Mahdi, M. (1990) 'Al-Farabi's Imperfect State', *Journal of the American Oriental Society* 110 (4): 691-726.(A review article that lays out many of the questions in the study of Islamic philosophy and its Greek origins.)

Peters, F. (1968) *Aristoteles Arabus*, Leiden: Brill.(Very much the standard work on Aristotle as understood in Arabic.)

Peters, F. (1996) 'The Greek and Syriac Background', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 3, 40-51.(Discussion of some of the important features of Greek and Syriac culture as sources of Islamic philosophy.)

Shayegan, Y. (1996) 'The Transmission of Greek Philosophy into the Islamic World', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 6, 98-104.(Detailed account of how the transmission took place, paying particular attention to the Persian background.)

Urvoy, D. (1996) 'Ibn Rushd', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 23, 330-45.(Detailed analysis of Ibn Rushd's thought and career.)

Walzer, R. (1962) *Greek into Arabic*, Cambridge, MA: Harvard University Press.(A traditional approach to Hellenistic influences on Islamic philosophy.)

Aristotelianism in the 17th century

Aristotelians in the seventeenth century comprised a group of mostly anonymous textbook writers whose chief claim to fame is that their philosophy was opposed by such as Descartes and Galileo. In line with the characterization of them by their opponents, their philosophy has generally been depicted as extremely conservative, monolithic and moribund. However, it is difficult to ratify such judgments. As Aristotelians, these philosophers do not seem particularly conservative; they appear to have assimilated many of the scientific developments of the seventeenth century, and the diversity and range of their views is quite broad. Some of the doctrines peculiar to them, or their particular developments of older views, can be seen as the background against which modern philosophy developed.

1 Some notable doctrines

What was taught in the schools during the seventeenth century was Aristotelian, but remains difficult to describe: probably not one of Aristotle's doctrines was held by all early modern scholastics. Some central Aristotelian theses were discarded. For example, Théophraste Bouju, in a work whose title page announces that all of it has the authority of Aristotle, rejected the Aristotelian four elements, discarding the sphere of fire and, as a consequence, argued against the radical heterogeneity of the sub-lunary and supra-lunary spheres (Bouju 1614). However, he safeguarded the *de facto* immutability of the heavens. Many of the theses that became canonical with later Aristotelians, such as the doctrine of substantial forms, also found early modern critics (Maignan 1653; Fabri 1686); there were even textbook writers who proclaimed the compatibility of peripatetic philosophy and atomism (Sennert 1618; Casimir of Toulouse 1674). Thus, it would be difficult to justify the epithet 'monolithic', although such pejorative labels have been applied to late scholasticism from the beginning. René Descartes wrote in a letter to Mersenne dated 11 November 1640 that there is nothing that seems as improbable to him than the philosophy of the schools: he does not think it difficult to refute, 'for one can easily upset all the foundations to which they agree and, once this is achieved, their particular disputes would appear inept'.

In fact, many Aristotelians of the seventeenth century were forward-looking, accepting the latest scientific developments including Galileo's celestial observations (Crassot 1618; du Chevreul 1623; du Moulin 1644, and others). Their philosophy of science can be characterized as probabilistic (Eustachius 1609 III.1: 1-3; and others), and some of their doctrines provide a background against which modern philosophy developed (see §§2, 5 of this entry). With the exception of the most noted (including Francisco Suárez and, possibly, Eustachius a Sancto Paulo, studied at least for his significance for Descartes) these thinkers are now generally neglected. In the Protestant world, such scholastic writers as Franco Burgersdijk at Leiden and Bartholomaeus Keckermann at Heidelberg were widely read, gaining fame primarily as logicians. In England, philosophers such as Thomas White and Kenelm Digby demonstrated flexibility in their attempts to graft the new philosophy on to Aristotelian roots.

2 Textbooks and notions of order

Eustachius a Sancto Paulo studied at the Sorbonne, receiving his doctorate in 1604. The following year he entered the Cistercian congregation of the Feuillants where he held various prominent positions, and became very influential in the French Catholic revival. He wrote two popular textbooks, a *Philosophy* (1609) and a *Theology* (1613-16), as well as two manuals of spiritual exercises.

Eustachius' *Philosophy* was considered by Descartes to be the best textbook in philosophy - Descartes even considered publishing it, together with his notes, other opinions, and his own philosophy. To understand the genre and its popularity, it must be contrasted with other such texts. Among the widely-read authors at the end of the sixteenth and the beginning of the seventeenth century were the Coimbrans and Franciscus Toletus. The Coimbrans (the Conimbricenses) were professors at the Jesuit Colègio das Artes, Coimbra (Portugal), who published a series of encyclopedic commentaries on Aristotle's works (see Collegium Conimbricense). The principal Coimbran was Petrus Fonseca, who separately published his own commentary on Aristotle's *Metaphysics*. Franciscus Toletus, a professor at the Jesuit Collegio Romano, similarly published commentaries on Aristotle's works, including an important *Logic* (1572), *Physics* (1573) and *On the Soul* (1574). (Other noted Jesuits who published textbooks for the collegiate curriculum included Rodericus Arriaga, Christopher Clavius and Antonius Rubius). In France, non-Jesuit philosophy texts from the same period included those by doctors

associated with the University of Paris, such as Eustachius and Charles d'Abra de Raconis (whose *Philosophy* was published in 1617), Jean Crassot (1618), Jean-Cécile Frey (1633) and François Le Rées. Judging from the number of editions, the texts of Eustachius and de Raconis were the most widely read Latin-language philosophies of the first half of the seventeenth century, Eustachius' *Philosophy* taking first place.

The seventeenth century also saw an enormous growth of philosophy textbooks in French, written by the tutors of the nobility (themselves often nobles). The movement began in the 1560s with the first French translations of Aristotle's works, but took off in the 1590s with the first French-language commentaries on Aristotle's *Physics*. Works in this genre include the 1614 textbook by Henry IV's almoner, Théophraste Bouju, and the 1643 volume by René de Ceriziers, a Jesuit who became a secular almoner of the Duc d'Orléans and later counsellor to the King. The most frequently reprinted work in the genre was the *Philosophy* (1627) by Scipion Dupleix, Cardinal Richelieu's favourite historian. This work alone seems to have exceeded Eustachius' in popularity.

The proliferation of textbooks in philosophy was a response to important changes taking place in pedagogy. The Jesuits, following the example of the University of Paris, had reorganized and standardized their curriculum. Textbooks, both Jesuit and non-Jesuit, were consequently modified. For example, the Coimbrans wrote volumes by committee, presenting the works of Aristotle that were taught in the curriculum; they followed the model of the great medieval commentaries, each volume treating a specific text (*Physics*, *On the Soul*, *On the Heavens* and so on), but with an elaborate (post-Renaissance) scholarly apparatus, giving both Aristotle's Greek text and its Latin translation, as well as Latin paraphrases (*explanationes*) and *quaestiones*, the analysis of standard problems relevant to the text being discussed. This pattern was generally followed by other textbook writers, although later editions of the Coimbran commentaries and textbooks such as those of Toletus omitted the Greek versions of Aristotle. Ultimately, Eustachius' *Philosophy* even omitted Aristotle's text itself. Eustachius simply arranged the *quaestiones* in the order in which the curriculum would have presented them, doing so for all the Aristotelian sciences within the frame of the whole philosophy curriculum - ethics and logic, physics and metaphysics - in a single volume. Dupleix followed the same pattern, as did de Raconis who also gave paraphrases along with the *quaestiones*. As their names generally indicated, the latter works were usually divided into four parts, following the collegiate curriculum. However, the *Philosophy* (1644) by the Protestant, Pierre du Moulin (whose logic text was also translated into English), was a three-part textbook (metaphysics having been omitted), and the *Philosophy* (1642) of Léonard Marandé added theology as a fifth part.

Underlying the format of these textbooks was a Renaissance concern with order or method. In one of the preliminary questions on the *Physics* Eustachius asked whether there is an order in the different parts of philosophy. He affirmed that there is one, appropriate both for the nature of things and for doctrine: namely, that which goes from the simplest to the more complex, from the principles to that of which they are constituted, and at the same time proceeding from the most universal things to the less universal, to the genera and species. Eustachius also asserted that Aristotle used such an 'order or method' in his writings on the various parts of philosophy. According to Eustachius, Aristotle in the *Physics* began with the principles, causes and general properties of natural things, then proceeded 'in part according to an analytic order and in part according to a synthetic order', from the most universal principles to the particular species of natural bodies. Eustachius consequently ordered his own presentation of natural philosophy into three parts: (1) natural bodies in general, from the principles of natural things to their causes and common properties, from matter and form, to causes, to place, infinity, void, time and motion; (2) inanimate natural bodies, from the world to the heavens and from elements to heterogeneous bodies; and (3) animate natural bodies, from soul in general to vegetative, sensible and rational soul. In fact, Eustachius reorganized Aristotle's topics and even reordered the *Physics* itself in keeping with his notion of order.

A general characterization of the doctrines of the first half of the seventeenth century is that the Jesuit textbooks (Coimbrans and Toletus, for example) usually propounded Thomist interpretations of Aristotle, while those associated with the University of Paris (Eustachius and de Raconis) did not, often preferring Scotist doctrines (see Aquinas; Duns Scotus, J.). The French-language authors varied in outlook: Bouju was a Thomist in many respects; Dupleix was vociferously anti-Thomist, resembling greatly the non-Jesuit Paris philosophers; and de Ceriziers seems to have supported a later, hybrid version of Thomism, answering many of the charges levelled against Thomism. By 1665 even the Jesuits, as evidenced by Pierre Galtruche's textbook (a work approved by the Order), seem to have rejected the Thomist positions in philosophy. However, the debate about Thomism and Scotism

continued into the seventeenth century. (To illustrate these generalizations, §§3 and 4 of this entry offer two examples from the foundations of natural philosophy.)

3 Matter and form

There was a debate in school texts about whether matter can exist without form. A positive resolution of this esoteric topic might lead one towards a dualistic, as opposed to a hylomorphic, conception of substance. Toletus (1589 I.13) discussed whether prime matter is a substance and detailed Aquinas' negative answer. According to Aquinas, prime matter is pure potency or has only potential being, so prime matter cannot be brought into being without form and cannot subsist without form. Toletus also explained Scotus' affirmative reply, that matter is a positive entity really different from the reality of form and can subsist in its own right distinct from form. Toletus shared Aquinas' view, his own doctrine being that prime matter is imperfect in itself (Toletus 1589 I.13: fol. 34 *verso*). Bouju also followed the Thomist line (Bouju 1614 I: 315-31). In contrast, Eustachius supported a variant of Scotus' doctrine: 'Though matter cannot be produced nor annihilated by any natural agent, God can create or annihilate it… God can strip naked all forms, substantial and accidental, from matter, or create it naked, without form, *ex nihilo*, and allow it to subsist by its own power in such a state' (1609 III.1.2.4: 16-17). Abra de Raconis agreed; quoting both Aquinas and Scotus, he said that matter is an incomplete substance, but maintained that God can create matter without substantial form ([1617] 1651 *Tractatus de Principiis* 4: 35-9). Scipion Dupleix threw into relief the disagreement between Thomists and Scotists:

> Thus matter deserves the name of substance because it subsists by itself and is not in any subject. This reply is based on the Philosopher's doctrine, but it does not satisfy everyone, particularly Saint Thomas Aquinas and his followers, who hold that such matter is not in nature, and cannot be in it, and even that this is so repugnant to nature that God himself cannot make it subsist thus stripped of all form. But this opinion is too bold, too mistaken, and it has been rejected by Scotus the Subtle [Doctor] and by several others.
>
> (Dupleix [1603] 1990: 131)

It is interesting to note that Dupleix argued against Aquinas' doctrine of prime matter by analogy to the sacrament of the Eucharist requiring real qualities and substantial forms, a difficulty that would haunt Cartesians and atomists later in the century. Some textbook writers got around the accusation by accepting the reality of matter as a miracle - for example, de Ceriziers argued that there can be no form without matter and no matter without form naturally, but added 'however, one must not deny that God can conserve matter without any form, since these are two beings that can be distinguished, that no more depend upon one another than accident from substance, the former being separated from the latter in the Eucharist' (1643 III: 51-2). The solution seems to have been unstable, so that by 1665 (II: 27) Gaultruche argued against the Thomists (*contra Thomistas*) about prime matter.

Not everyone gave up the Thomist doctrine of matter. Although Scotists such as Frassen seem to have had the best of the argument, and Thomists and Jesuits such as Barbay and Vincent needed to opt for middle ground, some Thomists resolutely maintained their position (Frassen 1686: 36-41; Barbay 1676b: 64-72; Vincent 1660: 2.74-7). For example, the Dominican Antoine Goudin wrote:

> it can be asked whether God by means of his omnipotence could create matter without it having a form. Scotus asserts it, as do some authors outside of Saint Thomas' school; Saint Thomas and all the Thomists deny it…. It seems that matter cannot exist without form even by means of God's absolute power. That is what Saint Thomas states (III quodlib. art 1). God himself cannot make it that something exist and not exist. He cannot make something that implies a contradiction and, consequently, he cannot make matter be without form.
>
> (Goudin [1668] 1864 II: 131)

4 The concept of place

Seventeenth-century school philosophy debated whether place itself is mobile or not - which is to say, whether there is a fixed reference for motion. Thomists distinguished between material place and formal place (where formal place is the real ground or *ratio* of place, in Aquinas' vocabulary). Place is then movable accidentally (as material place) and immovable *per se* (as formal place, defined as the place of a body with respect to the universe as a whole). Thus a ship is formally immobile with respect to the universe as a whole when the waters flow around it. Scotists rejected the distinction between material and formal place, arguing instead that place is a relation of the

containing body with respect to the contained body. Place is then a relative attribute of these bodies. (They also made use of the term *ubi*, sometimes referred to as inner place, to denote the symmetric relation of the contained body with respect to the containing body.) Since the relation changes with any change of either the contained body or the containing body, the place of a body does not remain the same when the matter around changes, even though the body in question might remain immobile. When a body is in a variable medium, the body is in one place at an instant and in another at an other instant; to capture what is meant by the immobility of place, Scotists said that the two places are distinct but *equivalent places* from the view of local motion.

Toletus took Aquinas' side against Scotus (Toletus 1589 IV.5: fol. 120 recto-121 recto). So did Bouju who also kept some Averroist elements. Bouju asserted that place is movable *per se* in what he called 'lieu de situation' and *per accidens* in what he called 'lieu environnant':

> The earth… is in a *lieu environnant* and can also be said to be in a *lieu de situation* with respect to the poles of the world. But it cannot change place with respect to its totality; thus it is immobile in that respect and mobile only with respect to some parts that can be separated from the totality and moved into others. The firmament is also in a *lieu de situation* with respect to the earth, but it cannot change except with respect to its parts and not in its totality, in the fashion of the earth.
>
> (Bouju 1614 I: 458-9; see also I: 460)

Eustachius, on the other hand, used Scotus' vocabulary: place and *ubi* being relations between the containing and contained bodies, and places being the same *by equivalence* (Eustachius 1609 III.3.2.1: 56-8). Abra de Raconis held a similar doctrine. De Raconis discussed two kinds of place, external and internal, external being the surface of the concave ambient body, and internal being the space occupied by the body. According to de Raconis, the ultimate heaven is in place internally, or occupies a space of three dimensions (Raconis [1617] 1651 *Physics* IV.2.1-2: 204-5). The distinction between external and internal place (or space) can also be found in Toletus and the Coimbrans; but they did not use the distinction in their resolution of issue of the mobility of place.

As was often the case, it was Scipion Dupleix who put the greatest contrast on the situation. He held that place was immobile in itself, while bodies change places. He took it that Aquinas had a different opinion, interpreting Aquinas' doctrine of formal place as the view that one can imagine a distance from each place to certain parts of the world with respect to which a given place, though changeable, may be said to be immobile. Scipion raged against the doctrine: 'But since all this consists only in useless imaginations, I am surprised that this opinion was received in several schools of philosophy; however, there are so many weak though opinionated brains who follow so closely the doctrine of certain persons that they would follow them right or wrong, and forget the golden sentence of the Philosopher: *I am a friend of Socrates, a friend of Plato, but rather more a friend of truth*' (Dupleix [1603] 1990: 149-50). On the subject of the place of the universe, Dupleix also rejected Aquinas' opinion, preferring a doctrine he attributed to Philoponus and Averroes, that when air is blowing around a house, one says that the place of the house changes accidentally. The house is in the same place *by equivalence* (that is, of the successive, distinct accidental places). Finally even Gaultruche rejected the Thomist doctrine of place, including the Thomist doctrine that the universe cannot move as a whole (Gaultruche 1655 II: 331). As with matter and form, the debate about the concept of place was not completely settled by the second half of the seventeenth century, overlapping similar disputes among the new philosophers (Frassen 1686: 357; Barbay 1676b: 261-72; Vincent 1660 II: 847-925; Goudin [1668] 1864 II: 504-6).

5 The origins of the modern concept of *idea*

Not all early modern scholastic debates took the form of Thomism versus Scotism. Some developments might have been equally at home in either camp. An interesting case, in view of the later philosophy of history, is a development of the Platonic and Neoplatonic theory of ideas within an Aristotelian context in the first few decades of the seventeenth century (see Neoplatonism). The discussion of ideas in Bouju's *Philosophy* is fairly standard. Enumerating the four Aristotelian causes, Bouju adds an account of 'exemplary causation'. Ideas are routinely identified with exemplars, either Platonic ideas or ideas in God's mind, and the question discussed is, whether in serving as models for creation, ideas as exemplars *cause* the things that imitate them in some fifth way. Further, the physician has an idea of health, the architect in building a house tries to make it like the one 'he has in his mind', and so on. Bouju is echoing a well-established Scholastic-Aristotelian tradition, in which ideas are either

the forms in God's mind according to which he makes things, or the exemplars in artificers' minds when they make their artefacts. Ideas as exemplars, however, are not strictly psychological. They are forms which are general, not particular, patterns to be followed in this or that case, rather than particular mental events (Bouju 1614 I: 297-8).

Possibly the first instance of a new, psychological usage of 'idea' in the philosophical literature seems to occur in the first part of the *Physics* of Eustachius. Again the question is whether exemplary causes constitute a fifth class in addition to the canonical four. Eustachius' answer is that in the case of natural causation exemplary cause may be taken to be a kind of efficient cause, but in the case of an artificer it is, rather, the formal cause:

> What the Greeks call Idea the Latins call Exemplar, which is nothing else but the explicit [*expressa*] image or species of the thing to be made in the mind of the artificer. Thus the idea or exemplar is in this case some image [*phantasma*] or work of imagination [*phantasiae*] in the artificer to which the external work conforms. And so in the artificer in so far as he is an artificer there are two internal principles of operation, namely the art in his mind or reason and the idea or exemplar in his imagination [*phantasia*]. Art is a certain disposition, but idea is a certain act or concept represented by the mind. So, the mind first represents a copy of the thing to be made through art, then it contemplates what it has represented, and directs the external work to its likeness.
>
> (Eustachius 1609 III.1.3: 36.)

Here the idea is an image and (particularly crucial for the Cartesian reading) it is 'an act' or 'concept expressed by the mind'. In this brief passage, then, we have the contemporary meaning that Descartes will exploit; idea is, as it were, an image, expressive of something, something which the mind contemplates. It is both something I do as well as something I 'see'.

One finds a similar account in Abra de Raconis ([1617] 1651: 94). Clearly, Descartes would not have to look far to find the term 'idea' used with the kind of ambiguity that he assigned to it in the Meditations.

6 Late seventeenth-century scholastics

By the end of the seventeenth century, Cartesianism and other versions of the new philosophy began to take hold and scholastics ranked themselves pro and con - mostly con. Many of them wrote critiques of Descartes (Huet 1689; La Grange 1692; Jean Duhamel 1692) or inserted critiques of Descartes into their textbooks (Frassen 1686; Barbay 1675-6; Vincent 1660, 1667; Goudin 1668; Jean Duhamel 1705); some even wrote satire of Cartesianism (Huet 1692; Daniel 1690, 1693) (see Huet, P.D.). And some attempted reconciliations between what they took to be Aristotelianism and Cartesianism (Jean-Baptiste Duhamel 1677; Le Bossu 1674). A few became Cartesians (Pourchot 1695). In England, Edward Stillingfleet and the 'Blackloist' John Sergeant were important critics of John Locke from an Aristotelian point of view.

See also: Aristotle; Aristotelianism, medieval; Aristotelianism, Renaissance; Suarez, F.

ROGER ARIEW

List of works

Arriaga, R. (1632) *Cursus philosophicus (Philosophy course)*.(A complete philosophy course by a Jesuit professor.)

Barbay, P. (1675) *Commentarius in Aristotelis logicam (Commentary on Aristotle's Logic)*, Paris.(The first portion of the philosophy course by a professor at the University of Paris. Barbay's texts were recommended by both Jesuits and Oratorians.)

Barbay, P. (1675b) *Commentarius in Aristotelis metaphysicam (Commentary on Aristotle's Metaphysics)*, Paris. (The third portion of Barbay's course.)

Barbay, P. (1676) *Commentarius in Aristotelis moralem (Commentary on Aristotle's Ethics)*, Paris.(The fourth portion of Barbay's course.)

Barbay, P. (1676b) *Commentarius in Aristotelis physicam (Commentary on Aristotle's Physics)*, 2nd edn, Paris. (The second portion of Barbay's course.)

Bossu, R. le (1674) *Parallele des principes de la physique d'Aristote, et de celle de René Des Cartes (Parallel between the Principles of Physics of Aristotle and Descartes)*, Paris.(An attempt to reconcile the philosophies of Aristotle and Descartes.)

Bouju, T. (1614) *Corps de toute la Philosophie (The whole of philosophy)*, Paris.(An early French-language

complete philosophy course.)
Casimir de Toulouse (1674) *Atomi peripateticae, sive tum veterum tum recentiorum atomistarum placita ad neotericae scholae methodum redacta, (Peripatetic atomism…)*, Toulouse.(An attempt to reconcile atomism and Aristotelianism.)
Ceriziers, R. de (1643) *Le philosophe français (The French philosopher)*, Paris.(A complete French-language philosophy course by a Jesuit.)
Chevreul, J. du (1623) *Sphaera (The Sphere)*, Paris.(A standard mathematics textbook teaching the rudiments of scholastic astronomy but taking Galileo's observations into account.)
Clavius, C. (1611-12) *Opera mathematica (Mathematical works)*, Rome.(The mathematical texts of the Collegio Romano Jesuit professor most responsible for the training of Jesuits who went on to teach in the colleges of the order.)
Conimbricenses (Jesuits of the University of Coimbra) (1592) *Commentarii in octo libros physicorum Aristotelis (Commentaries on the eight books of Aristotle's Physics)*, Coimbra.(The commentary of the Jesuits of the University of Coimbra dealing with bodies, motion, and their metaphysical foundations, that is, the beginning of the physics course.)
Conimbricenses (Jesuits of the University of Coimbra) (1598) *Commentarii in tres libros de anima (Commentaries on the three books on the soul)*, Coimbra.(The commentary of the Jesuits of the University of Coimbra dealing with animate creatures, usually the end of the physics course.)
Conimbricenses (Jesuits of the University of Coimbra) (1606) *Commentarii in universam dialecticam Aristotelis (Commentaries on the whole of Aristotle's logic)*, Coimbra.(The logic textbook of the Jesuits of the University of Coimbra.)
Crassot, J. (1618) *Physica (Physics)*, Paris.(A physics textbook from a professor at the University of Paris.)
Daniel, G. (1690) *Voyage du monde de Descartes*, Paris; trans. as *A Voyage to the World of Cartesius*, London, 1692.(A satire of Cartesian philosophy by a Jesuit.)
Daniel, G. (1693) *Nouvelles difficultés proposées par un péripatéticien à l'auteur du 'Voyage du monde de Descartes' (New difficulties proposed by a peripatetic to the author of 'Voyage to the World of Descartes')*, Paris. (The continuation of Daniel's satire of Cartesian philosophy.)
Duhamel, J. (1692) *Reflexions critiques sur le système cartesien de la philosophie de mr. Régis (Critical reflections on the Cartesian system of philosophy of Mr Régis)*, Paris.(A critical examination of the Cartesian textbook of Pierre-Sylvain Régis by a professor at the University of Paris.)
Duhamel, J. (1705) *Philosophia universalis sive commentarius in universam Aristotelis philosophiam ad usum scholarum comparatam (The whole of philosophy…)*, Paris.(The philosophy course of Duhamel, a professor at the University of Paris, containing many critical references to Descartes' philosophy, including a listing of condemnations by various authorities.)
Duhamel, J.-B. (1677) *Philosophia vetus et nova (Philosophy old and new)*.(An attempt to reconcile old (mostly Aristotelian) and new (mostly Cartesian and Gassendist) philosophies.)
Dupleix, S. (1603a) *La logique ou art de discourir et raisonner (Logic…)*, Paris: Fayard, 1984.(The first portion of Dupleix's course.)
Dupleix, S. (1603b) *La physique (Physics)*, Paris: Fayard, 1990.(The second portion of Dupleix's course.)
Dupleix, S. (1610a) *La métaphysique (Metaphysics)*, Paris: Fayard, 1992.(The third portion of Dupleix's course.)
Dupleix, S. (1610b) *L'ethique ou philosophie morale (Ethics)*, Paris: Fayard, 1994.(The fourth portion of Dupleix's course.)
Dupleix, S. (1627) *Corps de philosophie (Collection of philosophy)*, Geneva.(A popular complete French-language philosophy course, including the separately published *Logic*, *Physics*, *Metaphysics* and *Ethics*.)
Eustachius a Sancto Paulo (Asseline) (1609) *Summa philosophica quadripartita de rebus dialecticis, moralibus, physicis, et metaphysicis (Sum of philosophy in four parts…)*, Paris.(A popular Latin-language complete philosophy course, which Descartes wanted to publish together with his philosophy to enable readers to contrast the two.)
Fabri, H. (1666) *De plantis et de generatione animalium, de homine (On plants and on the generation of animals, on man)*, Lyons.(A physics text by a late seventeenth-century French Jesuit with atomist tendencies.)
Frassen, C. (1686) *Philosophia Academica (Academic philosophy)*, 3rd edn.(A philosophy textbook that defends Scotist school doctrines.)
Frey, J.-C. (1628) *Cribrum philosophorum qui Aristotelem superiore & hac aetate oppugnarunt (The sieve of the philosophers…)*, Paris.(An attack on anti-Aristotelians by a professor at the University of Paris.)
Frey, J.-C. (1633) *Universae philosophiae compendium (Compendium of the whole of philosophy)*, Paris.(The philosophy course by Frey, professor at the University of Paris.)

Gaultruche, P. (1665) *Philosophiae ac mathematicae totius clara, brevis, et accurata institutio (Instruction on all of philosophy and mathematics…)*, Caen, 5 vols.(The philosophy and mathematics courses by a Jesuit professor at Clermont, the main Jesuit college in Paris.)

Goclenius, R. (1613) *Lexicon Philosophicum (Philosophical dictionary)*, Frankfurt.(A comprehensive dictionary of late scholastic philosophical terms.)

Goudin, A. (1668) *Philosophie suivant les principes de Saint Thomas (Philosophy following the Principles of Saint Thomas)*, trans. T. Bourard, Paris, 1864.(A philosophy textbook that defends Thomist school doctrines.)

Grange, J.-B. de la (1682) *Les principes de la philosophie contre les nouveaux philosophes, Descartes, Rohault, Regius, Gassendi, le P. Maignan, etc. (The principles of philosophy against the new philosophers…)*.(A critique of Cartesian and other new philosophies by an Oratorian.)

Huet, P.D. (1689) *Censura philosophiae cartesianae (Judgment on Cartesian philosophy)*, Paris.(A critique of Cartesian philosophies by Huet, who was a Cartesian in his youth.)

[M.G. de L'A.] (1692) *Nouveaux mémoires pour servir à l'histoire du cartésianisme (New memoirs for the history of Cartesianism)*.(A satire of Cartesian philosophy, anonymously by Huet.)

Maignan, E. (1653) *Cursus philosophicus (Philosophical course)*, Toulouse.(A philosophy course by a Minim, which was thought too much a departure from Aristotelianism by some.)

Marandé, L. (1642) *Abrégé curieux de toute la philosophie (Curious summary of all of philosophy)*, Paris.(A summary of what is generally contained in philosophy courses, including a section on theology.)

Moulin, P. du (1644) *Philosophie mise en francois et divisee en trois parties, scavoir, elements de la logique, la physique ou science naturelle, l'ethyque ou science morale (Philosophy translated into French…)*.(A philosophy course, minus metaphysics, by a French Protestant.)

Pourchot, E. (1695) *Institutio philosophica (Philosophical instruction)*, Paris.(The first philosophy course by a professor of philosophy at the University of Paris who was sympathetic to Cartesianism.)

Raconis, C.F.d'A. de (1617) *Summa totius philosophiae (Sum of all philosophy)*, Paris, 1651.(A complete philosophy course by a professor at the University of Paris.)

Sennert, D. (1618) *Epitome naturalis scientiae*, Wittenberg; trans. as *Thirteen books of natural philosophy*, London, 1659.(A physics textbook by a physician who attempts to reconcile Aristotelianism and atomism.)

Toletus, F. (1572) *Commentaria una cum quaestionibus in universam Aristotelis logicam (Commentary on the totality of logic)*, Venice.(The logic text of a Jesuit *Collegio Romano* professor.)

Toletus, F. (1574) *Commentaria una cum quaestionibus in tres libros Aristotelis de amina (Commentary on the three books on the soul)*, Venice.(The commentary of a Jesuit Collegio Romano professor dealing with animate creatures, usually the end of the physics course.)

Toletus, F. (1589) *Commentaria una cum quaestionibus in octo libros de physica auscultatione (Commentary on the physics)*, Venice.(First edition published in 1573, this is the commentary of a Jesuit *Collegio Romano* professor dealing with the beginnings of the physics course.)

Vincent, J. (1660) *Cursus philosophicus (Philosophy course)*, Toulouse.(A complete Philosophy course by a Jesuit professor.)

Vincent, J. (1677) *Discussio peripatetica in qua philosophiae cartesianae principia (Peripatetic discussion regarding the principles of Cartesian philosophy)*, Toulouse.(A critical examination of Cartesian philosophy by a Jesuit.)

References and further reading

Blair, A. (1993) 'The Teaching of Natural Philosophy in Early Seventeenth-Century Paris: The Case of Jean-Cecile Frey', *History of Universities* 1993: 95-158.(Gives an account of the life and works of a seventeenth-century professor of philosophy.)

Bouillier, F. (1868) *Histoire de la Philosophie cartésienne (History of Cartesian philosophy)*, 2 vols.(Concerns the reception of Cartesianism, pro and con, including accounts of the 'persecution' of Cartesianism by scholastics and Jesuits.)

Brockliss, L.W.B. (1987) *French Higher Education in the Seventeenth and Eighteenth Centuries: A Cultural History*, Oxford: Oxford University Press.(Details the curriculum in French colleges and universities in the humanities, philosophy, theology, law and medicine.)

Dear, P. (1987) *Mersenne and the Learning of the Schools*, Ithaca, NY: Cornell University Press.(Includes an account of the scholastic, mostly Augustinian sources used by Descartes' mentor, Marin Mersenne, in his advocacy of mathematical physics.)

Des Chene, D. (1995) *Physiologia: Philosophy of Nature in Descartes and the Aristotelians*, Ithaca, NY: Cornell University Press.(Compares late scholastics and Descartes on such topics as motion, principles of change, the

essence of matter, mechanism and final cause.)

Gilson, E. (1976) *Études sur le rôle de la pensée médiévale dans la formation du système cartésien (Studies on the role of medieval thought in the formation of the Cartesian system)*, 2nd edn, Paris: Vrin.(Consists of a collection of essays which trace the roots of Cartesianism in late scholasticism.)

Lohr, C. (1988) *Latin Aristotle Commentaries, II: Renaissance Authors*, Florence: Olschki, 1995.(The most complete bibliography of Latin language late-scholastic texts, includes bibliographic and biographic references to Arriaga, Burgersdijk, Conimbricenses, Crassot, Eustachius, Frey, Goclenius, Keckermann, de Raconis, Sennert and Toletus, among others.)

Reif, P. (1962) *Natural Philosophy in Some Early Seventeenth Century Scholastic Textbooks*, unpublished dissertation, Saint Louis University.(Still the only comparative full-scale study of the contents of seventeenth century textbooks.)

Sortais, G. (1924) *Histoire de la Philosophie moderne depuis Bacon jusqu'à Leibniz (History of modern philosophy from Bacon to Leibniz)*, Paris: Beauchesne.(Imparts a history of seventeenth century philosophy that does not neglect minor figures, including scholastics.)

Verbeek, T. (1992) *Descartes and the Dutch*, Carbondale, IL: University of Southern Illinois Press.(Concerns the reception of Cartesianism at Utrecht and Leiden.)

Wallace, W. (1984) *Galileo and His Sources*, Princeton, NJ: Princeton University Press.(Includes an account of the natural philosophy, logic, and mathematics taught at the Collegio Romano, the Jesuits' main college in Rome.)

Aristotelianism, medieval

Although there are many possible definitions, 'medieval Aristotelianism' is here taken to mean explicit receptions of Aristotle's texts or teachings by Latin-speaking writers from about AD *500 to about* AD *1450. This roundabout, material definition avoids several common mistakes. First, it does not assert that there was a unified Aristotelian doctrine across the centuries. There was no such unity, and much of the engagement with Aristotle during the Middle Ages took the form of controversies over what was and was not Aristotelian. Second, the definition does not attempt to distinguish beforehand between philosophical and theological receptions of Aristotle. If it is important to pay attention to the varying and sometimes difficult relations of Aristotelian thought to Christian theology, it is just as important not to project an autonomous discipline of philosophy along contemporary lines back into medieval texts.*

The most important fact about the medieval reception of Aristotle is in many ways the most elementary: Aristotle wrote in Greek, a language unavailable to most educated Europeans from 500 to 1450. Aristotle's fate in medieval Europe was largely determined by his fate in Latin. Early on, Boethius undertook to translate Aristotle and to write Latin commentaries upon him in order to show the agreement of Aristotle with Plato, and also presumably to make Aristotle available to readers increasingly unable to construe Greek. He was able to finish translations only of the logical works, and to write commentaries on a few of them and some related treatises. Even this small selection from Aristotle was not received entire in the early Middle Ages. Of the surviving pieces, only the translations of the Categories and De interpretatione were widely studied before the twelfth century, though not in the same way or for the same purposes. Before the twelfth century, Aristotelian teaching meant what could be reconstructed or imagined from a slim selection of the Organon and paraphrases or mentions by other authors.

The cultural reinvigoration of the twelfth century was due in large part to new translations of Greek and Arabic works, including works of Aristotle. Some translators worked directly from the Greek, among whom the best known is James of Venice. Other translators based themselves on intermediary Arabic translations; the best known of these is Gerard of Cremona. Although the translations from Greek were often the more fluent, translations from the Arabic predominated because they were accompanied by expositions and applications of the Aristotelian texts. To have a Latin Aristotle was not enough; Latin readers also needed help in understanding him and in connecting him with other authors or bodies of knowledge. Hence they relied on explanations or uses of Aristotle in Islamic authors, chiefly Avicenna.

The thirteenth century witnesses some of the most important and energetic efforts at understanding Aristotle, together with reactions against him. The reactions begin early in the century and continue throughout it. The teaching of Aristotelian books was condemned or restricted at Paris in 1210, 1215 and 1231, and lists of propositions inspired by certain interpretations of Aristotle were condemned at Paris and Oxford in 1270 and 1277. However, interest in Aristotle continued to grow, fuelled first by the translation of Averroes' detailed commentaries, then by new translations from Greek. At the same time, some of the most powerful Christian theologians were engaged in large-scale efforts to appropriate Aristotle in ways that would be both intelligible and congenial to Christian readers. Albert the Great composed comprehensive paraphrases of the whole Aristotelian corpus, while his pupil Thomas Aquinas undertook to expound central Aristotelian texts so as to make them clear, coherent, and mostly concordant with Christianity.

Very different projects predominate in the fourteenth century. For John Duns Scotus and William of Ockham, the texts of Aristotle serve as distant ground against which to elaborate philosophical and theological teachings often radically anti-Peripatetic. If they are fully conversant with Aristotle, if they speak technical languages indebted to him, they are in no way constrained by what they take his teaching to be. Other fourteenth-century projects include the application of procedures of mathematical reasoning to problems outstanding in Aristotelian physics, the elaboration of Averroistic positions, and the rehabilitation of Albert's Peripateticism as both faithful and true to reality. By the end of the Middle Ages, then, there is anything but consensus about how Aristotle is to be interpreted or judged. There is instead the active rivalry of a number of schools, each dependent in some way on Aristotle and some claiming to be his unique interpreters.

1 Scope

Before it can be defined, the phrase ‘medieval Aristotelianism’ has to be stripped of a modern presupposition and restricted in its extension. ‘Aristotelianism’ and similar philosophical terms are neither ancient nor medieval, but modern. When these were popularized in the eighteenth century, they were not proposed neutrally. ‘Aristotelianism’ was coined with the meaning that the historical fate of Aristotle’s texts and teachings could be reduced to a pure position, an unhistorical set of propositions that could be analysed or criticized. By contrast, ancient historians of philosophy tended to think of Aristotle’s legacy as a school in some stronger sense, that is, as a succession of communities arising from his writings and attempting to practice the way of life they proposed as best (see Aristotle). Medieval Christians, who thought that their faith prevented them from claiming membership in pagan communities without qualification, often thought of philosophical teaching as an inheritance of human wisdom passed down through lines of authoritative texts and their recognized interpreters. On neither of these ancient or medieval views is the teaching of Aristotle understood or judged apart from one or another of its actual genealogies. This entry is aligned with the earlier views, and so understands ‘Aristotelianism’ as referring to historical engagements with the texts and teachings of Aristotle rather than to some abstract arrangement of ‘Aristotelian’ tenets.

Removing the modern presupposition makes the ambiguities of extension for ‘medieval Aristotelianism’ more difficult. There are, first, ambiguities of appropriation. Aristotelian teachings, texts and methods run through medieval learning from early to late, in almost every discipline and at every level. Many of these appearances are implicit or anonymous: Aristotle’s terminologies or procedures became common learning, no longer considered the property of an Aristotelian school. Clusters of terms such as ‘form/matter’, ‘act/potency’, ‘substance/accident’ or ‘formal/material/efficient/final’ appear in biblical commentaries, legal codifications, pharmaceutical handbooks and guides for composing poetry. Even explicit mentions of Aristotle come in half a dozen forms. Sometimes Aristotle figures merely as an ornament, in much the way that bits of Shakespeare or the Authorized Version of the Bible were once used by English speakers. At other times Aristotle is invoked merely to secure a general principle in no way specific to him. Explicit theoretical engagements with Aristotle themselves range from deployments of single phrases or sentences, through sustained amplification or criticism of arguments, to the detailed interpretation of whole texts. A single writer may show all of these relations to Aristotle over works in different genres or, indeed, in a single work.

Other ambiguities in extension are introduced by the interaction of Aristotle with Jewish, Christian and Islamic religious thinking. The Christian ambiguities are perhaps the most familiar. Almost all of the Christian Aristotelians in the Latin West were members of the clergy. Most spent their professional lives teaching and writing, not the liberal arts or philosophy, but Christian theology. It remains controversial whether or to what extent we can find an autonomous or even a textually distinguishable Aristotelian philosophy among them. Similar difficulties arise in trying to distinguish philosophy from other learned disciplines. Medieval students of Aristotle followed him through the host of topics that he broaches in his writings, from logic to poetics, from the physics of moving bodies through the species of fish to the motions of celestial spheres. We tend to divide these topics according to modern disciplinary divisions, but historically it is all ‘Aristotelianism’ and, in many ways important to medieval writers, all equally ‘philosophy’.

A final set of ambiguities lies around the term ‘medieval’. It is notorious that the ‘Middle Ages’ cannot be cut off neatly, either at their beginning or their end. Many learned medieval people considered themselves direct heirs and successors to pagan antiquity and the early centuries of the Christian churches. They admitted no chronological divide. At the other end, the ‘Renaissance’ is a slogan as much as a fact; there are as many continuities as discontinuities between philosophical thinkers of the thirteenth and fifteenth centuries (see Aristotelianism, Renaissance). Similar uncertainties affect geographical or cultural boundaries. Any span of time appropriately called medieval will include, in territories we now consider European, segments of both the Byzantine and Islamic traditions of Aristotle. Although these traditions figure prominently in any retelling of Latin receptions of Aristotle, they are separate and quite complex traditions that would require separate and substantial treatment (see Byzantine philosophy; Aristotelianism in Islamic philosophy).

Faced with these difficulties of matter, discipline, chronology and geography, the most prudent position is the most modest. Within this article, ‘medieval Aristotelianism’ will be restricted to explicit receptions and examinations of Aristotle’s texts or teachings by Latin-speaking writers from about AD 500 to about 1450. The emphasis throughout will be on explicit relations to Aristotle rather than on diffuse transformations or absorptions of him. No attempt

will be made to distinguish beforehand between philosophers and theologians, though much attention will be paid to the varying and sometimes difficult relations of Aristotelian thought to Christian theology. The arrangement will be according to traditional chronological divisions, though the chronological order is not mean to suggest that the topics or figures discussed can be subsumed under a single history. What follows are not incidents in a single narrative plot, but examples of the diversity of explicit receptions of Aristotle.

2 Boethius and the earlier Middle Ages

As Aristotle wrote in Greek, a language unknown to most educated Europeans from AD 500 to 1450, knowledge of his works in medieval Europe is largely determined by the extent to which they had been translated into Latin. The first important translator was Boethius in the sixth century. Before him, Aristotle seems not to have been translated into Latin in any systematic way. There certainly were Latin summaries of Aristotelian doctrine: one of these, a fourth-century outline of basic logic called *Categoriae decem* (Ten Categories), was widely studied in the earlier Middle Ages on the assumption that it had been written by Augustine. There were also paraphrases or criticisms of Aristotelian doctrine in Latin works, both pagan and Christian. However, these traces are slight, and for good reason. Well into the fifth century, the Roman empire conducted its philosophy in Greek. Citizens even of the westernmost provinces were taught Greek as the language not only of philosophy but also of medicine, natural science and, to some extent, of *belles lettres*.

Boethius may have suspected that this erudite bilingualism would not continue past his own lifetime. More importantly, he himself may have suffered from the increasing inaccessibility of Greek learning. His references to Aristotelian texts outside the Organon appear to derive from notations in the manuscripts he used or from Greek commentators on the logical works (see Aristotle Commentators). It may be that Boethius translated only the logical works of Aristotle because these were the only works he had at hand in Italy.

Whatever the problems of decline in the Latin West, Boethius' stated reasons for undertaking the translations would have been recognized by any Greek-speaking student of philosophy. Boethius translated Aristotle and wrote Latin commentaries on him in order to show the agreement of Aristotle with Plato. The project of reconciliation was already an old one, with multiple sources. Latin philosophy had been eclectic in its borrowing from Greek since Cicero. It was part of Cicero's philosophic identity to take what was best from every philosophic school. This tendency was reinforced and given grander theoretical justification in Neoplatonism from Porphyry on. If there was disagreement among Neoplatonists about the extent to which Aristotle had dissented from Plato, there was unanimity in thinking that Aristotle had accepted much from his teacher and that he had gone on to treat certain subjects with richer detail. Thus almost all of the Neoplatonists, beginning with Porphyry, wrote commentaries on Aristotle (see Neoplatonism). The appearance of these commentaries in Latin during the later Middle Ages influenced scholastic discussions at many points, but they also show that Boethius' project of translation and commentary had a long pedigree. The reconciliation of Aristotle and Plato, which meant most often the subordination of Aristotle to Plato, was a familiar programme by which Boethius could justify his project of translation and commentary.

In the end, Boethius translated the *Categories*, *De interpretatione*, *Prior* and *Posterior Analytics*, *Topics* and *Sophistical Refutations* - in short, all of the logical works of the Organon. He certainly wrote both rudimentary and advanced commentaries for *De interpretatione* and Porphyry's *Isagōgē(Introduction)* to the *Categories* and single commentaries on the *Categories* and *Topics,* as well as the *Topics* of Cicero. Even this small selection from Aristotle was not received entire in the early Middle Ages. The translation of the *Posterior Analytics* was lost early, as was the commentary on the *Topics*. Of the surviving pieces, only the translations of the *Categories* and *De interpretatione* were widely studied before the twelfth century, though not in the same way or for the same purposes. For example, Boethius' first commentary on *De interpretatione*, the more rudimentary one, predominated in the ninth and tenth centuries while the second commentary received much more attention in the eleventh century. With Boethius as in so many later cases, the rule is evident: translation is not the same as reception. Something can exist and even circulate in translation, but not yet be appropriated for speculative use. The reception can be limited for any number of reasons: because copies are scarce, because the work is too difficult, or because learned taste has turned to other topics.

The Aristotelian logic that Boethius had made available passed down to later readers in company with other works, among them the anonymous *Categoriae decem*. Together, these works constituted the curriculum not only in logic

but also in philosophy in general. They provided the instruments and the occasions for reflection on any number of topics, including theological ones. Indeed, the most striking early medieval appearances of Aristotle are in theological debate, for example in Carolingian debates about the Trinity (see Carolingian renaissance). There is also the solitary figure of Johannes Scottus Eriugena, whose *Periphyseon* enacts yet another appropriation of Aristotelian categorical logic into a Neoplatonic metaphysics. Some of Eriugena's readers later reversed the circle by copying bits of his teachings into texts of Aristotelian logical works as explanatory glosses. In the tenth and eleventh centuries, however, attention swung back to the *Categories* and *De interpretatione* with their Boethian commentaries. These were utilized regularly not only for the basic teaching of logic but also for theological debates (about the Eucharist, for example) and for scriptural exegesis (on passages such as Paul's doctrine of salvation in the letter to the Romans).

3 The twelfth century

The cultural reinvigoration of the twelfth century was due in large part to new translations of Greek and Arabic works in philosophy, the natural sciences and medicine. Prominent among these translations were works of Aristotle. It is important to distinguish the various routes by which the new Aristotle arrived in the West, especially as translations were often accompanied by strong interpretations - and not just in the sense that every translation is itself an interpretation.

The first stage in the recovery of Aristotle was neither translation nor interpretation, but the finding and using of manuscripts of Boethius' translations of works other than the *Categories* and *De interpretatione*. It was not that these manuscripts had been lost or hidden away; they simply were not read, taught, annotated or widely copied until a revived interest in logic made them pertinent. The turn back to these Boethian translations was already evident in the 1120s.

By the 1130s, translators had begun to work directly from Aristotle's Greek editions. These first translators worked in northern Italy, chiefly in cities such as Pisa and Venice that had active trading relations with Greek-speaking Byzantium (see Translators). The principal Aristotelian translator is James of Venice, who translated *Posterior Analytics* and *Sophistical Refutations* once again, as well as *Physics*, *On the Soul*, five of the seven so-called *Parva naturalia* (Smaller Natural Works) and *Metaphysics* 1-4.4. James also translated at least some parts of late Imperial Greek commentaries on Aristotle, but these translations did not circulate widely.

James wrote his own commentary, using Byzantine models, on the *Sophistical Refutations*, making him one of the earliest Latin writers to comment on the so-called 'new logic'. This 'new logic' contained segments of the Organon that were either unstudied or untranslated before the twelfth century. It provided not only any number of technical teachings, but also the doctrine of the *Posterior Analytics* on the nature of demonstratively organized bodies of knowledge. This doctrine made it possible to construct a comprehensive account of 'science' within which physical, mathematical, metaphysical and even theological doctrines could be ordered and criticized (see Logic, medieval). However, even with the recovery of Boethian versions and the new translations of James and his colleagues, something less than one-third of what we today consider the Aristotelian corpus was then available in Latin. Notably lacking were the foundational works in physical science.

Some of these gaps were soon filled by another species of translation, translations into Latin of Arabic versions of Greek works (see Islamic philosophy: transmission into Western Europe). These translations were made in areas of greatest contact with Arabic civilization, principally in Spain. Although these translators followed the translators in northern Italy by three decades, their versions of Aristotle taken from Arabic were to become more important than those made directly from the Greek. The reason is not to be found in the clarity of translation: the Greek versions are generally more fluent, not so say more accurate. Rather, what was needed was not so much texts as textbooks. Confronted with the bald text of Aristotle, many readers were at a loss. They needed tutors, and they had in their Arabic sources more powerful pedagogical aids than they had yet found in Greek. The Arabic texts of al-Farabi or Avicenna (see Ibn Sina) showed how Aristotle could be applied to questions of interest to Latin readers. They also offered more immediately attractive forms of specialized or technical knowledge. Some idea of the allure of this knowledge can be seen in the biographies of the translators: they came from all parts of northern Europe, including England and the Low Countries, to work in Toledo, Barcelona, Tarazona, Segovia, Leon, Pamplona and southern France (see Translators). The most prominent of the translators working in Spain was an Italian, Gerard of Cremona.

Gerard's work as a translator spans about forty years (*c.*1150-87), during which time he translated at least seventy works and perhaps as many as a dozen more. His translations of Aristotle include *Posterior Analytics* (with the commentary of Themistius), *Physics*, *On the Heavens*, *On Generation and Corruption* and *Meteorology* 1-3. Two of these works had, of course, already been translated by James of Venice from the Greek, but that was either not known or did not matter to Gerard. The separation of the Italian and Spanish translations can also be seen as Gerard began translating Aristotelian commentaries from the Arabic. Some of these commentaries were originally Greek (for example, Themistius on the *Physics*), though Gerard translated them from Arabic. Other commentaries and expository treatises had been written in Arabic as part of the Arabic culture's own appropriation of Greek learning (as in the case of al-Farabi's commentary on the *Physics*).

Two related facts provide important context for Gerard's Aristotelian translations. The first is that the Aristotelian translations were only a small part of his labour; he translated many more works on medical, astronomical and mathematical subjects. Gerard's Aristotle came into Latin surrounded by a small library of natural science. The second fact is that Gerard translated a number of works as part of the Aristotelian corpus that were not at all Aristotelian. The most important of these is the so-called *Liber de causis (Book of Causes)*, a compilation of material mostly from Proclus (see Liber de causis). This treatise on the cosmic participation of such transcendental features as goodness or unity was read for about a century after Gerard's translation as if it had been written by Aristotle, and so guided interpretations of the whole Aristotelian corpus. Other pseudonymous Aristotelian works circulating by the end of the twelfth century included treatises on the properties of elements, on how the soul causes basic bodily operations, on health and on alchemy. The presence of such heterogeneous works within the Aristotelian corpus would complicate the reading of Aristotle beyond the end of the Middle Ages.

The most dramatic effects of the Arabic and Neoplatonic Aristotle on Latin-speaking thought did not occur until well into the thirteenth century, but it is important to suggest how the new texts were already being received in the twelfth century. The most thorough reception was that of the logical works, which began to be appropriated in the schools of liberal arts as early as the 1130s. There are traces of the 'new logic' in the later works of Peter Abelard, but the impact is felt primarily in the next generation of writers, especially after the middle of the century, in the works of Adam of Balsham 'Parvipontanus', Alberic of Paris and Geoffrey of St Victor, with their colleagues and followers. The impact is not what one might expect. The first works of the 'new logic' to have an impact were the *Sophistical Refutations* and the *Topics*, which were read for what they could contribute to the analysis of procedure in disputation and to an understanding of signification and fallacies. The *Prior Analytics* was much slower in displacing Boethius' teaching on syllogism, and the *Posterior Analytics*, which might seem the most important work, was the last to make its teaching felt. Equally interesting is the fact that the new Aristotelian texts were not received through the writing of detailed, literal commentaries; these did not appear until the next century. The new logic was received rather as an aid or instigation to projects of logical research that had begun before these texts were known. The effects also vary by topic or tradition of inquiry. For example, Peter Helias' *Summa* on Priscian, written *circa* 1150, contains half as many references to Aristotle as to Boethius (see Language, medieval theories of).

The physical works had an even more patchy reception. Traces of the early translations from the Greek can be found in the writers associated with the cathedral school at Chartres (see Chartres, school of, and especially William of Conches). At the same time, if not earlier, bits of Aristotelian natural science began to figure in the teaching texts of the medical school at Salerno. By the end of the century, Urso of Salerno had written a treatise *On the Mixture of the Elements*, which was an attempt to resolve a vexed question in Aristotelian physics: what kind of virtual existence do the properties of individual elements have when those elements enter into compounds actualized by another substantial form? The whole of natural philosophy and medicine had hardly become Aristotelian, however. It could not, since so much of the scientific corpus remained untranslated. At century's end, the sum of translations available in Latin came to just half of the Aristotelian writings. There were, for example, no Latin versions of the bulky books on animals: together, these zoological treatises make up one-quarter of the whole corpus, and they contain the longest single Aristotelian book, *History of Animals* (see Natural philosophy, medieval). Nor had translators yet gone far into the ethical, political or literary parts of the corpus. There is an anonymous twelfth-century translation of *Nicomachean Ethics* II and III, but that is the only piece attested; there is nothing from *Politics*, *Rhetoric* or *Poetics*.

4 The thirteenth century

At the beginning of the thirteenth century, in scattered places and among diverse writers, there was a growing uneasiness about the Aristotelian and Islamic works entering Latin at such a rapid rate. This unease was evident particularly among theologians, who saw the possibility that the Aristotelian natural philosophy would contradict or displace views on the natural world considered essential for Christianity. One early official reaction came in 1210, at a provincial synod of the archdiocese of Sens, the ecclesiastical authority responsible for Paris and its schools. The archbishop was Peter of Corbeil, master of theology and of canon law at Paris from 1190-98. He knew and endorsed the view of his former colleagues that there were heresies in theology (linked to the works of Amaury of Bène and David of Dinant) and dangerous teaching in the faculty of liberal arts, sparked by the newly translated books. The heresies are difficult to reconstruct, in part because they were so effectively suppressed. The measures taken against the teaching of Aristotle and his interpreters, certainly Avicenna and possibly Alexander of Aphrodisias, were relatively milder. The assembled bishops ordered that no books of Aristotle on natural philosophy or their commentaries were to be lectured on at Paris, in public or in private, under penalty of excommunication. The prohibition seems to have been ineffective: a weaker prohibition is reiterated five years later, in the statutes of the papal legate, Robert of Courçon. His statues only prohibit public lecturing on ‘the books of Aristotle on metaphysics or on summaries of them’ within the faculty of arts. Even this prohibition was barely effective; it was reiterated in a papal letter of 1228, but softened to compromise three years later. Pope Gregory IX then ordered that the theologians should confine themselves to questions that could be settled from theological authorities, while the masters of liberal arts were to refrain from lecturing on the prohibited books until they had been examined and expurgated by an expert committee. A committee was indeed appointed, but seems never to have done its work. From 1235 on, Aristotelian works were studied with increasing diligence and attention in their original, unexpurgated versions. By the end of the 1240s, at the latest, their study was not only permitted but required in arts faculties and they were in constant use by the most eminent theologians, despite the fact that Gregory’s prohibitions were reiterated pro forma in 1245 and 1263.

It is difficult to say how much expert acquaintance informed the reaction against Aristotle. Aristotelian books were regularly confused with works translated alongside them or purporting to be drawn from them. Hence the decrees of 1210 or 1215 can be regarded as uninformed protests against a poorly known physicalism in which Aristotle was implicated. Even two or three decades later, theological authors at Paris who cited Aristotle frequently continued to judge harshly the materialism they thought latent in him, and which they found expressed by some of his commentators or disciples. William of Auvergne for example, rails against Alexander of Aphrodisias’ doctrine and heresies related to it, and against physicalist biases in Avicebron (see Ibn Gabirol) and Avicenna. The entangling of Aristotle with his interpreters was only made worse by the appearance in Latin of the works of Averroes (see Ibn Rushd).

Latin versions of many of the commentaries of Averroes began to circulate early in the second quarter of the thirteenth century. The most important group of these was produced by Michael Scotus in the 1220s and 1230s. Averroes had written three kinds of commentaries: epitomes, ‘middle’ commentaries, and ‘great’ or ‘large’ commentaries. These three genres represent increasing levels of attention to the letter of Aristotle’s texts, with the ‘great’ commentaries being by far the most detailed. Michael Scotus had a preference for these: he translated the ‘great’ commentaries on *Physics*, *On the Heavens*, *On Generation and Corruption*, *Meteorology*, *On the Soul* and *Metaphysics*. Other translators rendered Averroes’ middle commentaries on most of the Organon and on the *Nicomachean Ethics*. If these expositions helped Latin readers to interpret Aristotle and to avoid the difficulties or controversies that had trapped earlier readers, they also created many difficulties for Christians. These difficulties lay not so much in the differences between Christianity and Islam as in those between Christianity and Aristotle’s paganism, since Averroes was often painfully clear about Aristotelian teachings that contradicted the revelations of both the Bible and the Qur’an.

The story of condemnation and appropriation at Paris should not be taken as the only story about the reception of Aristotle. There are others elsewhere, including some at schools that had steady traffic with Paris. John Blund, for example, seems to have taught at both Paris and Oxford early in the thirteenth century. His *On the Soul*, which probably derives from a course of lectures given at Oxford in Arts, draws heavily from Avicenna in the interpretation of Aristotle. More important but more obscure is the career of Robert Grosseteste. Grosseteste may have studied and taught at both Oxford and Paris before 1214; by the 1220s he was a theologian of reputation at

Oxford. Before or around this time, he was engaged in writing commentaries on a range of Aristotelian works. After he was made bishop of Lincoln in 1235, Grosseteste organized a team of translators that produced a Latin version of the *Nicomachean Ethics* and Greek commentaries on it. Another example can be found in Richard Rufus Of Cornwall, who like many of his brethren was a mature scholar when he joined the Franciscans at Paris in 1238. He had already composed commentaries on both *Metaphysics* and *Physics*. However, the influence of these works would be felt most at Oxford after he began teaching theology there in 1250. Both commentaries show an acquaintance with Averroes and such parts of earlier commentary traditions as Averroes records. Rufus' work also shows his concern to resolve the contradictions between Aristotle and Christian revelation by holding Aristotle to the highest Aristotelian standards.

Aristotle was widely read by the middle of the century, but the work of receiving him was hardly finished. It was now up to Latin readers to construct coherent and defensible readings for an expanded and much controverted corpus. One response was to make Aristotle accessible by systematizing him with available bodies of knowledge. Around 1250, Albert the Great undertook to write paraphrases of the entire corpus in order to make Aristotle 'intelligible for the Latins'. The project stretched out over twenty years and required that Albert fill gaps in the corpus with what he thought Aristotle had taught. The concern throughout was to make whatever was true in Aristotle cohere with the rest of speculative knowledge, which in Albert's case meant to cohere both with a Neoplatonic metaphysics and with specialized sciences not discussed by Aristotle or not known to him.

Another interpretative project was more concerned with finding a coherent teaching across the Aristotelian texts. Around 1266, Thomas Aquinas inaugurated a series of line-by-line expositions of Aristotle. Their form depended partly on that of Averroes' 'great' commentaries, but more exactly on models that Thomas had learned in the arts faculty at Paris twenty years earlier. More importantly, Thomas was borrowing the modified Averroistic format to show that it was possible to read Aristotle against Averroes on disputed points such as the demonstrability of the eternity of the world (see Eternity of the world, medieval views of), the individuality of human intellects or the substantial union of human soul and body. Thomas undertook to expound only some of the most important Aristotelian books or parts of books, and his expositions limit themselves at most points to giving the briefest reading consistent with clarity and coherence. They do not rehearse many controversies or tease out many subtleties, and they rarely bring in alternate accounts of the things discussed. The simplicity of Thomas' commentaries made them influential even among readers not otherwise favourable to him.

Thomas was aided in his work of exposition by yet another effort of translation. A fellow Dominican, William of Moerbeke, had begun around 1260 to retranslate from the Greek most of the principal works of the Aristotelian corpus. More importantly, he translated for the first time some of the principal ancient commentaries on Aristotle, including that of Alexander of Aphrodisias on *Metaphysics*, Ammonius on *De interpretatione*, and both Philoponus and Themistius on *On the Soul*. These texts were acquired avidly by Thomas Aquinas and others as witnesses to the authentic meaning of Aristotle, a meaning different from that thought to have been imposed by Averroes. The Averroistic Aristotle was being confronted with an arguably Greek Aristotle, who was on many points made more congenial to Christian teaching.

The question of how far Albert and Thomas were Aristotelians in their own thinking has been debated for centuries without resolution. On the one hand, it is certainly the case that neither Albert nor Thomas called themselves Aristotelian. Albert several time reminds his readers that he is expounding Aristotle, not endorsing him. Thomas will not even use the word 'philosopher' when speaking of Christians, much less of theologians. On the other hand, many of their contemporaries in the faculties of theology thought that they were all too ready to accept Peripatetic doctrines uncritically. In 1273, for example, Bonaventure gave a series of lectures at Paris in which he excoriated those who based philosophy or theology on Aristotelian doctrines. Bonaventure identifies the principal Aristotelian errors in physics, ethics and metaphysics, thus showing his own technical mastery of the Aristotelian corpus. If much of Bonaventure's anger is directed at teachers of arts, some of it at least is reserved for theologians who embrace the same errors.

The controversies occasioned by the teaching of Aristotle in the arts faculty were not confined to the writing of alternate commentaries and counterattacks. They resulted in two sets of ecclesiastical condemnations at Paris, one in 1270 and a second, much larger and more important, in 1277. A prominent target of the first condemnations was Siger of Brabant. Around 1265, when Siger was a master of arts, the controversy over the radical reading of

Aristotle began to build between the faculty of theology and the faculty of arts at Paris. This has often been described, rather inaccurately, as a quarrel between 'Augustinians' and 'Latin Averroists' (see Augustinianism; Averroism). The 'Augustinian' theologians were in fact, if unknowingly, indebted to Plotinus and Avicenna as well as Augustine, and they often engaged in the constructive interpretation of Aristotle. The 'Latin Averroists' were so described not just because they took up Averroes' commentaries, but because they were thought to teach a number of Averroes' controversial doctrines, including the subordination of religion to philosophy (a doctrine misleadingly described as 'double truth'). Behind these labels, the quarrel was clearly a fight over claims for autonomy not only on behalf of philosophical speculation, but also of a philosophical way of life. The arts were repeatedly attacked for asserting not just that they could demonstrate conclusions while keeping out of view the articles of Christian faith, but that the study of philosophy or the practice of natural contemplation could lead to happiness.

Many of the accused denied the charges, and it is difficult to know how much the teachings of the 'artists' were distorted in polemic and in the ensuing legal proceedings. Siger himself was already in trouble with ecclesiastical authorities as early as 1266. A first set of general condemnations came in December, 1270. Etienne Tempier, Bishop of Paris, condemned thirteen propositions drawn from Siger's writings. Fifteen months later, masters in the arts faculty were prohibited from disputing theological matters or from determining philosophical questions in a way contrary to the faith. Finally, in 1276, Siger and two other masters from Brabant were summoned before the inquisitor of France to answer charges of heresy. It is not clear what happened to Siger afterwards. Finally, some 219 propositions were condemned by Tempier in Paris on 7 March 1277. The list was hastily composed by a mostly anonymous committee, and even its defenders would later admit that it contained inconsistencies and points which were not clear. A much shorter and better organized list was condemned at Oxford eleven days later. The Oxford list proscribed thirty propositions divided under the headings of logic, grammar and natural philosophy.

The 219 condemned propositions cannot be summarized: they are too imprecise and heterogeneous. Some of the propositions declared false evidently contradict the Christian faith. Others, however, seem to have been rejected because they relied on argumentative procedures or presuppositions that troubled the theological censors. Among these are some procedures that would figure prominently in speculative thought during the next two centuries: the distinction between God's power as unrestricted and as exercised (*potentia absoluta*, *potentia ordinata*), or the analysis of physical problems by conceiving alternate worlds or by transposing them to an imagined vacuum. In these ways, the list of propositions shows not just the topics of controversy between theologians and teachers of arts, but the budget of topics considered worthy of investigation by students of Aristotle.

There has been considerable contemporary dispute over the importance of the condemnations of 1277. Some have held that they were a local aberration forgotten after a few decades, others that they had a chilling effect on theological speculation for a century or more. In either case, it is worth remembering that the condemnations were not seen as a rejection of Aristotle, but rather of a certain sort of Aristotelianism. Henry of Ghent, for example, who authored some of the condemnations, placed himself as a theologian in the tradition of Augustine, but knew Aristotle well and used his knowledge frequently in exploring or defending theological points. Indeed, Aristotelian terms, principles, examples and arguments had so permeated theological learning by the end of the thirteenth century that no theologian, no matter how critical of certain points in Aristotle, could be anything less than fluent as an Aristotelian.

5 The fourteenth and fifteenth centuries

Two figures dominate the reception of Aristotle at the beginning of the fourteenth century: John Duns Scotus and William of Ockham. They would determine the reading of Aristotle not only by their own efforts, but by the prominence of their disciples over the next two centuries.

No brief description of the engagement of either Duns Scotus or Ockham with Aristotle can fail to be misleading. The thought of each is not only difficult, but continues to occasion strikingly different interpretations, let alone evaluations. Duns Scotus' writings are brutally original in manner of expression and teaching. Ockham writes more conventionally, but in the service of a critique of prevailing views no less sharp. Neither hesitated to reject central Aristotelian teachings: Duns Scotus denies Aristotle's accounts of intellection and substantial individuation, Ockham the accounts of linguistic foundations and ethical truth. Again, neither was to provide literal expositions of Aristotle's works or other rudimentary exegetical works, genres in which they might have

expressed general positions on Aristotle.

Both Duns Scotus and Ockham did write what are sometimes called commentaries on Aristotle. Duns Scotus wrote an early treatise in connection with *Metaphysics*, as well as multiple treatments of issues in *Categories*, *De interpretatione*, *Sophistical Refutations* and Porphyry's *Isagōgē*. These 'commentaries' engage prevailing disputes attached to the Aristotelian texts by only the frailest threads: Duns Scotus uses his writing to declare and defend positions on these disputes rather than to set out a simple reading of Aristotle. Ockham also wrote on the four logical works, with the aim of showing that none of them required a realist view of language or conception. He wrote in three different genres on *Physics*: in an exposition, a *summulae* (summary) and a set of disputed questions. The balance of exegesis and dispute varies in these three, but some indication of Ockham's authorial position may be given by the *summulae*; he begins that work with the assertion of a number of his characteristic positions, such as the non-existence of universals and the strictness of the criteria for scientific demonstration. Neither Duns Scotus nor Ockham was concerned to present a comprehensive interpretation of the Aristotelian corpus after the different styles of Averroes, Albert or Aquinas.

Still, it would be inaccurate and unjust to suggest that Duns Scotus and Ockham were incapable of close exegesis, just as it would be unjust to conclude that Scotistic or Ockhamistic views somehow precluded engagement with Aristotle. One of the most systematic commentators on Aristotle in the fourteenth century was John Buridan, a near contemporary of Ockham who shared many of the latter's concerns for linguistic and conceptual criticism. In the 1320s, Buridan began writing a mixture of expositions and disputed questions on Aristotle's books, frequently writing in both genres on a single book. Eventually he covered almost the entire corpus, with the exception of the *History of Animals* and the *Poetics*, doing so with careful attention to Aristotle's purposes and meanings. If Buridan's 'nominalism' was anti-Aristotelian, it certainly did not prevent him from being a tenacious expositor of Aristotle.

Scotistic or Ockhamistic views entered into combination with several other projects for explaining, extending and correcting Aristotle. One of these projects found an institutional home in Merton College, Oxford, with a group active from the 1320s onwards (see Oxford Calculators). The principal figures among these 'calculators' were Thomas Bradwardine, Richard Kilvington, William Heytesbury, Richard Swineshead and John Dumbleton. There are numerous differences among them, as there are numerous similarities linking them to their colleagues working in the established logical genre of sophisms. What the Mertonians share are procedures of mathematical reasoning as applied to the resolution of problems outstanding in Aristotelian physics. One central problem concerned the intensification and remissions of accidental forms, while another concerned the motion of projectiles. These problems in Aristotelian physics were not discovered in the fourteenth century; they had been known to late Roman and Islamic commentators, as well as to readers in the prior two centuries. What distinguishes the treatment of these problems among the calculators is not so much the statement of the problem or even its doctrinal resolution, but the sophistication of its mathematical development. In the case of projectile motion, for example, members of the groups posited an inner force or *impetus*, which they then proceeded to describe through accelerations and decelerations (though this doctrine too may derive from previous commentaries on Aristotle).

The methods of the Mertonians were to have considerable consequences for the reading of Aristotle's *Physics*. The theologian Nicole Oresme, a student of Buridan, having learned a basic reading of the principal texts of Aristotle and a nominalist critique of them, went on to take up the methods of mathematical analysis. He wrote sets of questions on a number of Aristotelian texts, as well as vernacular commentaries on the *Nicomachean Ethics* and the *Politics*. His masterwork is a dialectical response to Aristotle's *On the Heavens*. Here, Oresme first expounds the literal sense, next extracts the principal theses, then produces arguments for their negations and finally affirms either the thesis or its negation depending on his judgment of the matter at hand.

A second approach to the relation of Duns Scotus or Ockham to Aristotle was a frank rehabilitation of Averroes. The principal figures of this movement were John of Jandun, Thaddeus of Parma and Angel of Arezzo. John of Jandun was a master of arts at Paris in the first decade of the fourteenth century, and wrote disputed questions on the central works of Aristotelian physics and metaphysics. In these questions, John frequently rehearses the readings of his predecessors, including Thomas Aquinas (whom he calls 'old expositor'), but then asserts quite plainly the views of Averroes, which agree with those of Aristotle. John confronts contradictions between an Averroistic Aristotle and Christian doctrine with a simple and unexplained profession of faith.

A third aspect of the fourteenth-century reception of Aristotle, widespread in faculties of theology across northern Europe, took the form of a perceived recovery of essential Aristotelian truths, against all rivals. Jean Gerson, who became chancellor of the university of Paris in 1395, describes three contending schools of thought: ‘formalizers’, ‘nominalists’ and ‘Peripatetics’. Each of the three groups claimed medieval proponents: the formalizers claimed Duns Scotus, the nominalists claimed Ockham and Buridan, and the Peripatetics claimed Albert and Thomas. Each group was also held to reflect or descend from an ancient source: the formalizers from Plato, the nominalists from Epicurus and the Peripatetics, obviously, from Aristotle. The Aristotelian tradition invoked here is that of a specific line of thinkers. One writer traces it from Alexander of Aphrodisias and Themistius, through Boethius and John Philoponus, Avicenna and Averroes, to Albert the Great and Thomas Aquinas. The tradition is also characterized by a set of conclusions that emanate from the view (very roughly) that the human mind knows fixed, universal truths by abstracting universals from singulars, which it does without the need of any special divine assistance. This was taken to be the foundation of Aristotle’s teaching and, by its proponents, the foundation of sound philosophy or theology.

This school called itself Peripatetic and also, perhaps surprisingly, Albertist. The preference for Albert over Thomas is evident in a group of writers which included John of Maisonneuve and his student, Heimeric de Campo. John taught as master of arts in Paris from 1400 onwards, while Heimeric was master of theology at Cologne and Louvain from 1428 onwards. John wanted to defend against both nominalists and formalizers a cognitive realism grounded in appeals to the divine ideas. Heimeric extended this Albertist project into a sustained critique of Thomas Aquinas. If both John and Heimeric seem at times to attribute certain views to Albert unreflectively, they are equally capable of careful readings in which they argue for Albert’s Aristotle as the most accurate and the truest. Albert’s work is presented as the completion of Aristotelian inquiry.

There is obviously no final medieval consensus about the reception of Aristotle, as there is no end to the traditions of ‘medieval Aristotelianism’. The main medieval appropriations of Aristotle propounded around 1450, whether Albertist, Thomist, Scotist or Ockhamist, would continue to find advocates up into the twentieth century, especially in Roman Catholic schools and universities. Particular parts of the medieval Aristotle would also survive in general philosophical learning. The most obvious survival was of the logical doctrines, which attracted readers as astute as Kant, despite repeated attacks. Medieval versions of Aristotle on intellection would influence the development of modern epistemology at many points, saliently in the terminology of Descartes, Wolff and their various followers. Even medieval receptions of Aristotelian physics would have a long afterlife: Galileo’s critique of Aristotle owes much to late medieval debates over objections against Aristotelian texts in the ancient commentary tradition (see Galilei, Galileo).

See also: Albert the Great; Aquinas, T.; Aristotelianism in Islamic philosophy; Aristotelianism in the 17th century; Aristotelianism, Renaissance; Aristotle; Aristotle Commentators; Averroism; Duns Scotus, J.; Henry of Ghent; Islamic philosophy: transmission into Western Europe; Language, medieval theories of; Logic, medieval; Medieval philosophy; Natural philosophy, medieval; Oxford Calculators; Translators; William of Ockham

MARK D. JORDAN

References and further reading

Aristoteles Latinus (1957-), individual publication details as follows: >VII.2, *Physica: Translatio Vaticana*, ed. A. Mansion, Paris: Desclée de Brouwer, 1957; >I.1-5, *Categoriae vel Praedicamenta*, ed. L. Minio-Paluello, Paris: Desclée de Brouwer, 1961; >XXIX.1, *Politica: Translatio imperfecta*, ed. P. Michaud-Quantin, Paris: Desclée de Brouwer, 1961; >III.1-4, *Analytica priora*, ed. L. Minio-Paluello, Paris: Desclée de Brouwer, 1962; >II.1-2, *De interpretatione vel Periermenias*, ed. L. Minio-Paluello and G. Verbeke, Paris: Desclée de Brouwer, 1965; >XI.1-2, *De mundo*, ed. W.L. Lorimer, revised L. Minio-Paluello, Paris: Desclée de Brouwer, 1965; >I.6-7. *Categorium supplementa: Porphyrii Isagoge et Liber sex principorum*, ed. L. Minio-Paluello, Paris: Desclée de Brouwer, 1966; >XVII.2, *De generatione animalium*, ed. H.J. Drossaart-Lulofs, Paris: Desclée de Brouwer, 1966; >IV.1-4, *Analytica posteriora*, ed. L. Minio-Paluello and B.G. Dod, Paris: Desclée de Brouwer, 1968; >XXXIII.1-2, *De arte poetica*, ed. L. Minio-Paluello, Paris: Desclée de Brouwer, 1968; >V.1-3, *Topica*, ed. L. Minio-Paluello, Paris: Desclée de Brouwer, 1969; >XXV.1-1a, *Metaphysica: Translatio Iacobi (‘Vetustissima’) et Translatio composita (‘Vetus’)*, ed. G. Vuillemin-Diem, Paris: Desclée de Brouwer, 1970; >XXXVI.1-3, *Ethica Nicomachea*, ed. R.A. Gauthier, Leiden: Brill and Paris: Desclée de Brouwer, 1972-4; >VI.1-3, *De sophisticis elenchis*, ed. B.G. Dod, Leiden: Brill and Paris: Desclée de Brouwer, 1975;

>XXV.2, *Metaphysica: Translatio anonyma ('Media')*, ed. G. Vuillemin-Diem, Leiden: Brill, 1976; >IX.1, *De generatione et corruptione*, ed. J. Judycka, Leiden: Brill, 1986; >VII.1, *Physica*, ed F. Bossier and J. Brams, Leiden: Brill, 1990; >XXV.3, *Metaphysica*, ed. G. Vuillemin-Diem, Leiden: Brill, 1995.(Editions of surviving medieval translations of Aristotle into Latin, in progress. The texts and prefatory material are mostly in Latin. Three volumes of a survey of manuscripts have also been published; these are *Codices: pars prior*, ed. G. Lacombe *et al.*, La Libreria dell Stato, 1939; *Codices: pars posterior*, ed. G. Lacombe *et al.*, Cambridge: Cambridge University Press, 1955; and *Codices: supplementa altera*, ed. L. Minio-Paluello, Paris: Desclée de Brouwer, 1961.)

Blund, John (*c.*1200) *Tractatus de anima (Treatise on the Soul)*, ed. D.A. Callus and R.W. Hunt, Auctores Britannici Medii Aevi 2, London: British Academy, 1970.(An excellent, fully annotated edition.)

Corpus Latinum Commentariorum in Aristotelem Graecorum (1961-), Leiden: Brill. (Editions of medieval Latin translations of Greek commentaries on Aristotle, including Themistius and Philoponus on *On the Soul*, Ammonius on the *Peri hermeneias*, Alexander of Aphrodisias on the *Meteorology*, Simplicius on *Categories*, and various commentators on the *Nicomachean Ethics* and *Sophistical Refutations*. Texts are in Latin, prefatory material is in French or English.)

Dod, B.G. (1982) '*Aristoteles latinus*', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 45-79.(Includes a very helpful chart summarizing medieval Latin translations of Aristotle.)

Heimeric of Campo (1428) *Tractatus problematicus (Treatise of Problems)*, published as *Problemata inter Albertum Magnum et Sanctum Thomam ad utriusque opinionis intelligentiam*, Cologne, 1496.(This remains the only printing of Heimeric's treatise.)

Helias, Peter (1140s) *Summa super Priscianum (Summa on Priscian)*, ed. J.E. Tolson, *Cahiers de l'Institut du Moyen-âge Grec et Latin* 27-8, 1978: 1-210.(A careful edition, with a helpful preface by Margaret Gibson.)

John of Maisonneuve (1406-18) *De universali reali (On the Universal Considered as Real)*, in A.G. Weiler, 'Un traité de Jean de Novo Domo sur les universaux', *Vivarium* 6, 1968: 108-54.(The edition itself is on pages 126-52. Weiler provides a useful introduction to the edition.)

Lohr, C.H. (1967-74) 'Medieval Latin Aristotle Commentaries', *Traditio* 23-30 (in instalments).(A comprehensive listing of commentators, with details of the commentaries and a bibliography of pertinent scholarship.)

Lohr, C.H. (1988) *Commentateurs d'Aristote au moyen-âge latin: Bibliographie de la littérature secondaire récente (Medieval Latin Aristotle Commentators: A Bibliography of Recent Secondary Literature)*, Paris: Éditions du Cerf.(A supplement to Lohr 1967-74, bringing the scholarly bibliography up through 1987.)

Robert of Courçon (1215) Statutes, in H. Denifle and É. Chatelain (eds) *Chartularium Universitatis Parisiensis*, Paris: Delalain, 1889-97, vol. 1, no. 20.(Includes one of the prohibitions on certain teachings of Aristotle.)

Schmitt, C.B. (1983) *Aristotle and the Renaissance*, Martin Classical Lectures 27, Cambridge, MA: Harvard University Press. (Chapter 1 contains an interesting and partly retrospective discussion of Renaissance Aristotelians.)

Schmitt, C.B. (1986) 'Pseudo-Aristotle in the Latin Middle Ages', in J. Kraye, W.F. Ryan, and C.B. Schmitt (eds) *Pseudo-Aristotle in the Middle Ages: The Theology and Other Texts*, London: Warburg Institute/University of London.(A very useful survey of texts passing under the name of Aristotle.)

Sorabji, R. (ed.) (1987) *Philoponus and the Rejection of Aristotelian Science*, London: Duckworth, and Ithaca, NY: Cornell University Press.(A collection of essays that treats in part the recovery of Philoponus in the medieval West.)

Sorabji, R. (ed.) (1990) *Aristotle Transformed: The Ancient Commentators and Their Influence*, London: Duckworth, and Ithaca, NY: Cornell University Press.(Chapters 15-16 treat Boethius, and Chapter 19 considers the sources of medieval logic.)

Thijssen, J.M.M.H. (1991) 'Some Reflections on Continuity and Transformation of Aristotelianism in Medieval (and Renaissance) Natural Philosophy', *Documenti e studi sulla tradizione filosofica medievale* 2 (2): 503-8. (Discusses especially some varieties of Aristotelianism from the later Middle Ages.)

Urso of Salerno (*c.*1180?) *De commixtionibus elementorum (On the Mixture of the Elements)*, ed W. Stürner, Stuttgart: Klett, 1976.(A critical edition of an important text, it suffers from the lack of study of Urso's predecessors at Salerno.)

Van Steenberghen, F. (1955) *Aristotle in the West: The Origins of Latin Aristotelianism*, trans. L. Johnston, Louvain: E. Nauwelaerts.(A narrative survey of the reception of Aristotle in Paris and Oxford to 1277; somewhat dated.)

Van Steenberghen, F. (1980) *Thomas Aquinas and Radical Aristotelianism*, Washington, DC: Catholic

University of America Press.(A retelling of some of the central incidents treated in Van Steenberghen (1955).)

Aristotelianism, Renaissance

By the Renaissance here is meant the period of the fifteenth and sixteenth centuries during which there was a deliberate attempt, especially in Italy, to pattern cultural activities on models drawn from antiquity. However, Aristotelianism during that period was not cut off from medieval developments, since earlier interests and topics of discussion still held the attention of philosophers, theologians and non-academic intellectuals. Moreover, given that Aristotelianism was embedded in the university curriculum, the approach and activities of Renaissance Aristotelians often reflected earlier institutional developments. The educational reforms of the German Lutheran Philipp Melanchthon and of the newly-founded Society of Jesus (the Jesuits) ensured that Aristotle remained central to the curriculum. On the other hand, deliberate attempts to divorce themselves from earlier structures and approaches can be discerned in some Renaissance Aristotelians. Owing to the influence of humanism, professors of philosophy whose loyalty was to Aristotle came to study Greek and explicate Aristotle from the Greek text, to imitate the style of classical models, and to prefer the Greek commentators over the medieval Latins because their language was Greek.

Renaissance Aristotelianism did not constitute a uniform, coherent school of thought with a clearly defined body of doctrines shared by all adherents. A careful reading of the many commentaries, paraphrases, textbooks and treatises based on Aristotle's works reveals a surprisingly wide variation in interpretation and a strong tendency to modify or supplement the Stagirite's teachings with tenets derived from other philosophical or scientific sources or from contemporary interests and discoveries. It is perhaps wise to speak of a variety of Aristotelianisms rather than to perpetuate the long-standing caricature of 'modern' philosophy and science arising by throwing off the shackles of a monolithic Peripatetic orthodoxy. The various Aristotelianisms included Albertism, Thomism, Scotism and Averroism, but as a result of the new translations of the Greek commentators on Aristotle there were also Renaissance Aristotelians who approached Aristotle by way of Alexander of Aphrodisias, Themistius, Simplicius and John Philoponus. Another current is best described as 'eclectic Aristotelianism'. Some Aristotelians adopted a 'philological' approach, approaching Aristotle simply through analysis of the Greek text and not as a philosophical challenge. This approach made Aristotelianism irrelevant to the enterprise of philosophy, but fortunately did not predominate.

1 Aristotle, the Greek text and the Greek commentators

Although Aristotle was accorded the honorific title of 'The Philosopher' by most philosopher-theologians of the high Middle Ages (see Aristotelianism, medieval), there were also some who viewed him far less favourably. The fourteenth-century Franciscans Francis of Meyronnes, Antonius Andreas and John Canonicus carried Bonaventure's critique of Aristotle to its ultimate conclusion (see Bonaventure). The first called Aristotle 'the worst metaphysician', and both Andreas and Canonicus viewed him as a poor natural philosopher. Their critical views were known to Renaissance Aristotelians like Nicoletto Vernia, Agostino Nifo and Marcantonio Zimara (*c*.1475-1532). Nevertheless, Aristotle's influence on philosophy and science during the Renaissance was profound. Both his own works and those of his many late ancient and medieval interpreters attracted a wide audience. His writings on logic and natural philosophy formed the centrepiece of university studies in the arts and provided essential preparation for a career in medicine, law or theology, while his works on ethics, poetics and politics were widely read and discussed by a learned public increasingly educated in the schools of the humanists.

Despite the emphasis given to Renaissance Platonism in general histories of culture, a survey of the philosophical and scientific literature actually produced during the period from 1400 to 1600 clearly establishes the Renaissance as a golden age in the history of Aristotelianism. Since Renaissance philosophers trained in the arts faculties of major European universities had normally received a thorough grounding in Aristotle, those who became professors in their own right for the most part continued to base their instruction on the Aristotelian corpus. Their writings often reflected this, taking the form of commentaries on Aristotle, though works devoted to the systematic exposition of a particular topic - such as Pietro Pomponazzi's famous treatise *De immortalitate animae (On the Immortality of the Soul)* (1516) or Jacopo Zabarella's *De naturalis scientiae constitutione (On the Nature of Natural Science)* (published in 1586) - were not infrequent, particularly when aimed at a general learned audience rather than at other scholars (see Pomponazzi, P. §2; Zabarella, J. §6). Aristotle's dominance among university professors withstood the challenge mounted by the handful of appointments of chairs of Platonic philosophy in

Italy (see Platonism, Renaissance §5).

Much of the continued predominance of Aristotle was due to the work of humanists, their knowledge of Greek, and the new critical techniques they developed. The Greek text of Aristotle's works was published, notably in the edition of 1495-8 printed by Aldo Manuzio. New translations were done, and from the 1530s onward, bilingual editions were produced. From 1499 onward, but especially in the 1520s and 1530s, editions of the Greek commentators on Aristotle were printed, and these in turn were translated, many for the first time. Nor was attention paid only to the Greek commentators, for in 1483 Nicoletto Vernia edited an important Aristotle-Averroes edition, and 1550-2 saw the great Giunta Aristotle-Averroes edition containing many works not previously included. The Aristotle commentaries of such medieval Latin authors as Albert the Great and Thomas Aquinas were also printed. At the same time, attention was paid to the canon of Aristotle's work. Many dubious and spurious works had been excluded by 1600. On the other hand, the *Poetics*, which had been little read during the Middle Ages, came to have a great influence on literary criticism, and the pseudo-Aristotelian *Mechanics* also attracted a good deal of attention.

2 Logic and method

Renaissance Aristotelians developed an account of scientific methodology that built on insights put forward by Aristotle (§6) in his *Posterior Analytics* and *Physics*. In a well-known passage in the *Posterior Analytics* Aristotle distinguishes between two types of demonstration. The first he calls demonstration of the fact (*to hoti*) or demonstration *quia*. Its primary characteristic is that the middle terms of such demonstrations tell us only the fact that something is the case and not why this is the case. The second type of demonstration has a middle term that tells us the reason why (*to dioti*). Medieval thinkers inspired by this analysis connected Aristotle's remarks on demonstration with his remarks at the beginning of the *Physics* concerning the proper method for establishing the principles of nature. There Aristotle claims that one should start with those things that are more knowable to us and proceed to those things more knowable or intelligible by nature, though not to us. In addition, some later thinkers equated two methods set forth by Galen (§§3-4) with the two sorts of demonstration. In his *Art of Medicine*, Galen distinguishes between what he calls the method of resolution, in which an object is broken down into its component parts, and the method of composition, in which the components used in the resolution are put into their proper order. Late medieval Aristotelians, like Pietro d'Abano (1257-1315) in his *Conciliator differentiarum philosophorum et praecipue medicorum (Conciliator of the Differences between Philosophers and Especially Physicians)* (composed around 1300), conflated demonstration *quia* and *propter quid* with resolution and composition, and were thus able to offer a sophisticated account of both the method of scientific discovery and the proper way to order such knowledge.

Paul of Venice, a member of the Augustinian Order who studied at Oxford before returning to teach at Padua, sets forth the procedure of going from effect to cause and back to the effect, and defends it against the charge of circularity. He is also careful to note that demonstration in natural science does not yield the necessity and certitude of mathematics, since it deals with what happens for the most part. Natural science is thus a demonstrative science, but is only a science of what usually occurs. Two contemporaries of Paul, Hugo of Siena and Jacopo da Forli, also discuss the resolutive method. However, it is Nicoletto Vernia (§3) who examines in greater depth questions regarding demonstration and the resolutive method in his Paduan lectures on the *Posterior Analytics*. Vernia also takes up the question whether demonstration is circular. He points out that there must first be a movement from an effect by way of demonstration *quia* but thereafter there must be a return back to the effect if natural science is to be perfect. That is to say, after the resolution (*resolutio*) of an 'effect' - that is, what is experienced through the senses - into an initial universal knowledge, we still lack demonstration *propter quid*, which is knowledge based on the essential principles of the thing. Vernia therefore postulates a movement of the intellect (*negotiatio intellectus*) enabling it somehow to discern the true cause of the effect and the necessary causal connection of that cause to its effect. A genuine apodictic demonstration *propter quid* is the result.

The topic of resolutive or regressive method (*regressus*) was also treated by Agostino Nifo (§2), Vernia's former student. In his youth he too postulated a movement of the intellect (*negotiatio intellectus*) enabling it to discern the true cause of an effect so that a strict demonstration was possible. However, later in his career, after studying the Greek commentators, namely Alexander of Aphrodisias, Themistius, Simplicius and Philoponus, he denied the need for such a special movement of the intellect and held that the most that could be achieved in natural science

was a hypothetical syllogism and a hypothetical demonstration.

Discussions regarding these methodological issues are also to be found in such contemporaries as Marcantonio Zimara and Bernardinus Tomitanus (d. 1576). But it is Jacopo Zabarella (§5), a student of Tomitanus, who stands at the end of this long line of development. In his treatise *De regressu (On the Regress)*, Zabarella sets forth a systematic discussion of the nature of scientific investigation in which he assumes like the earlier theorists a basic continuity between Aristotle and Galen. He argues that demonstration *quia* provides us with a method of reasoning from effect to cause, but does not tell us the proper reason for the effect. Nevertheless, from the cause to which our initial knowledge of the effect leads, we can eventually achieve a full account of the proximate cause of the effect. That is, we can arrive at an argument that fully reveals the cause by giving a *propter quid* demonstration that states the reason why. What keeps this regressive method from being purely formal is the positing of the intermediate step of the movement (*negotiatio*) of the intellect, going from an initially confused and improper knowledge of the cause to a distinct and proper knowledge of it (see Galilei, Galileo).

3 Natural philosophy

One of the issues discussed by medieval Aristotelians that continued to concern Aristotelian philosophers during the Renaissance was the identification of the subject of natural philosophy. This issue was thought to be important because it concerned not only the relationship between metaphysics (the science of being *qua* being) and natural philosophy, but also the place of immaterial created beings, such as angels, within Aristotelian science. Thomas Aquinas (§9) had argued that mobile or changeable being (*ens mobile*), which seems to include all beings liable to change (that is, everything other than God), was the subject of natural philosophy, whereas his teacher Albert the Great (§4) took it to be mobile or changeable body (*corpus mobile*). Later, Paul of Venice states that the subject matter is natural body (*corpus naturale*), while his student, Cajetan of Thiene (Gaetano da Thiene) (1387-1485), maintained that it is sensible substance (*substantia sensibilis*). Subsequently the whole issue was approached in a systematic fashion by Cajetan’s student, Nicoletto Vernia (§3), who composed a separate question on whether mobile being is the subject of all natural philosophy. In it he reviews and compares a wide range of authors including Antonius Andreas and John Canonicus, whom he attacks for judging Aristotle to have erred in natural philosophy. Vernia accuses them of not speaking naturally but rather introducing theological considerations into the discussion, namely the question of the motion of angels. Vernia himself upholds and defends mobile being as the subject, which he took to be the position of Averroes. Thereafter Vernia’s student Agostino Nifo (§2) took up the question of the subject matter of natural philosophy in his commentary on the *Physics*. He rejects the respective positions of Albert and Aquinas but then attempts to conciliate them, arguing that mobile being can be considered the subject of natural science if by subject is meant the genus or most general predicate applying to the things considered in natural science, whereas mobile body can also be considered the subject if by that is meant the most general species to which all and only the things considered in natural science belong. He presents Averroes as holding that the total subject of natural science is a sensible thing (*res sensibilis*) in so far as it moves, that is, in so far as it contains the principle of motion.

Jacopo Zabarella (§6) who, like Vernia, wrote a systematic work on the question of the subject of natural philosophy, echoes the earlier discussions. He declares the common subject of all natural science to be body taken universally (embracing both the earthly bodies and also the bodies in the heavens), but in so far as body has within it a nature (that is, the principle of motion). He insists that this is the position of Aristotle (§10) in the *Physics* and *On the Heavens*. As to the subject of Aristotle’s *On the Soul*, he takes it to be animate body (*corpus animatum*). Further echoes of the discussion are still found in the seventeenth century (see John of St Thomas §3).

Another issue in the area of natural philosophy that merits mention is the challenge of Richard Swineshead and the fourteenth-century Oxford Calculators’ tradition of interest in physics (see Oxford Calculators). Central was their doctrine of measuring the intension and remission of forms, that is, physical properties, and stating this measurement in mathematical language. Paul of Venice was one of those who brought this manner of doing natural philosophy to Italy, and he promoted it in his teaching and writings. But it was also taught and fostered at Pavia in the teaching and writings of Giovanni Marliani (d. 1483). Nicoletto Vernia, who spent a year at Pavia, and studied Swineshead’s major work as well as other works in the tradition of the Oxford Calculators, emphatically rejected such an approach to natural philosophy as untrue to Aristotle and Averroes. Agostino Nifo also rejected the approach of the Calculators, insisting that their principles were those of mathematics and were therefore incapable

of providing the basis of a true natural science.

4 Psychology

The background for the discussion of psychology was provided by Aristotle's doctrine in his *On the Soul* of three types of soul, the vegetative, sensitive and intellective, and by his division of the intellect into an active element and passive or receptive element. The Aristotelian tradition would call these two elements the agent intellect and the potential or possible intellect respectively (see Aristotle §19). Some of the Latin medievals also saw Aristotle as maintaining a psychological power of the human soul, identified as the will (*voluntas*). These divisions raised such questions as whether the three souls were separate within the human being, whether the powers of intellect and will were distinct, and whether the agent and possible intellects were distinct either from each other or from the human being. Alexander of Aphrodisias had identified God as the agent intellect. Avicenna held that there was one agent intellect for all human beings, namely the lowest Intelligence, though each had an individual potential intellect (see Alexander of Aphrodisias §2; Ibn Sina §6). In the work on the soul known to the Latin medievals, Averroes held that there was just one separate intellect, but it was composed of both the agent and the potential intellects (see Ibn Rushd §3; Soul in Islamic philosophy, the).

Two particular problems about human cognition deserve mention. The first regards the nature of sensation as presented by the theory of the agent sense, proposed as an analogue of the agent intellect. Inspired by remarks of Averroes, John of Jandun had held that the object of sensation had to be spiritualized by an active power, an agent sense (*sensus agens*), if sensation was to take place. He thought that this agent sense was internal to and multiplied in each human being, though others like Giles of Rome (§3) held it to be a separate substance, that is, an Intelligence. The agent sense was discussed during the fourteenth century by such philosophers as John Buridan (§3), Nicole Oresme, Marsilius of Inghen (§1), and Taddeo of Parma (fl. early fourteenth century). It is also discussed by Blasius (see Blasius of Parma §3), Paul of Venice and Cajetan of Thiene. The latter proposes in a special question on the agent sense that a separate Intelligence provides the spirituality required for the act of sensation to occur. In like fashion, Agostino Nifo (§3) proposes in his own special treatise on the agent sense that God provides the spirituality required for sensation to take place. Cardinal Cajetan, Pietro Pomponazzi, Francesco Silvestri, and later the Jesuit scholastics also take up the question. In his own *Liber de sensu agente (Book on the Agent Sense)* (published 1590), Zabarella pays particular attention to Nifo's theory.

The second particular problem regards the need for intelligible species in the process of intellectual cognition. This need had already been proposed by medievals like Albert the Great, Thomas Aquinas and Duns Scotus (§13). What is involved is a universal representative image (*imago*) or likeness (*similitudo*) that is distinct from the act of intellection itself, yet needed to explain how humans can be united cognitively to external objects. In his explication of the cognitive psychology of Averroes, John of Jandun argued that Averroes had postulated intelligible species in the single possible or potential intellect. This interpretation of Averroes was heatedly debated during the Renaissance by philosophers like Antonio Trombetta (1436-1517) and Marcantonio Zimara, who agreed with Jandun, as well as by Alessandro Achillini (1463-1512) and Agostino Nifo (§3), who denied that Averroes had maintained intelligible species. In an early treatise on intelligible species, Pomponazzi maintained that Averroes had in fact held to such species; he accused Achillini and Nifo of having simply revived an old opinion, namely that of Walter Burley. Pomponazzi considers their opinion stupid and bestial. In fact, although Nifo maintained that neither Aristotle nor Averroes held to intelligible species, he himself takes as true that the human being does need intelligible species in order to think. This he maintains both in his early *De intellectu (On the Intellect)* (1503) and also in his later *De immortalitate animae (On the Immortality of the Soul)* (1518). The doctrine of intelligible species better enabled him to explain the knowledge enjoyed by the disembodied soul after death. In his own treatise on the topic, Zabarella (§§5-6) maintains that the intelligible species is the same thing as the act of thinking and does not remain after death. It is called an intellection if it is related to the intellect in which it exists, but it is called a species or image when it is related to the external thing known.

That intelligible species differ from the act of thinking and are images and likenesses necessary for human thinking was the common doctrine of Jesuit scholasticism. However, concepts as well as intelligible species were postulated, in line with Aquinas' distinction between the intelligible species and the inner word (see Aquinas §10). The need for such intelligible species and concepts was maintained by Franciscus Toletus (§5), the authors of the Coimbra commentaries (see Collegium Conimbricense), and Francisco Suárez (§3). But both Toletus and Suárez

also maintained, in clear opposition to Aquinas, that there are intelligible species and concepts of individuals as opposed to common natures.

The most important question about the intellective soul concerned its relation to the human body. The variously nuanced definitions of the soul that the Renaissance Aristotelians formulated took as their starting point Aristotle's statement in his *On the Soul* that the soul is the first actuality of a body that has life potentially in it (see Aristotle §17). What had been an issue in earlier discussions and continued to be an issue was whether this definition could be read in such a way as to allow a dualistic interpretation. The striking dualism to be found in the thought of Avicenna and the modified dualism that was to be found in the thought of Thomas Aquinas were of course not unknown. But what made the discussions somewhat more complicated was the availability by the end of the fifteenth century of translations of Alexander of Aphrodisias' own *On the Soul*, Themistius' paraphrases of Aristotle's *On the Soul* and the commentary on the same work that was traditionally attributed to Simplicius. Alexander's work presented a view of the soul as resulting from the harmony of bodily parts, while the commentary attributed to Simplicius presented a strikingly dualistic conception of the relation of the soul to the body (see Alexander of Aphrodisias §2).

While those who read Alexander and Simplicius would have found the human soul to be many, that is, one soul for each human being, the interpretation of Averroes was quite different. The theory of the human soul that emerged from Averroes' *Long Commentary on Aristotle's 'On the Soul'* was that each human being had an individual sensitive soul and a set of internal senses that were numerically distinct in each human being. On the other hand, there was numerically only one intellect for the entire human race, and that intellect served as the 'intellective soul' for each human being (see Ibn Rushd §3). Renaissance Aristotelians like Paul of Venice, Apollinaris Offredi (fl. fifteenth century), Johannes Argyropoulos (*c.*1415-87), Nicoletto Vernia, Agostino Nifo, Alessandro Achillini, Marcantonio Zimara, Pietro Trapolin (1451-1506), Marcantonio Genua (d. 1563) and Jacopo Zabarella, expended much of their philosophical and academic energies on explicating the text of Averroes and facing the serious problems that lurk in such a theory of the soul and intellect. In doing so they were in fact following in the footsteps of medievals like Thomas Aquinas, Siger of Brabant, Giles of Rome (§3), John of Jandun and John Baconthorpe among others. It bears noting that in his early thought Pomponazzi took Averroes to present the correct interpretation of Aristotle concerning the human soul. Only later did he attack Averroes, using Aquinas' arguments against Averroes' doctrine of the unity of the intellect.

5 The debate over immortality

The question of whether the personal immortality of the human soul could be demonstrated had already been answered in the negative by such medieval thinkers as Siger of Brabant (§2) (at least in his earlier works), John Duns Scotus, John of Jandun and John Buridan. Soon afterwards, Blasius of Parma (§2) denied the demonstrability of personal immortality. His position reflected the thought of Buridan and Alexander of Aphrodisias. Blasius, who had already stated in questions on Aristotle's *On the Soul* that human cognition has a material basis and is simply a grade of perfection of living matter, reasoned that inasmuch as the human soul has no operation independent from the body, there are no grounds for arguing that the human soul can survive the death of the body. He was reprimanded in 1396 by the Bishop of Pavia for remarks against the Catholic faith and the Church.

A little later, Paul of Venice in his commentary on Aristotle's *On the Soul* stresses, in opposition to Alexander of Aphrodisias, the intellect's ability to reflect upon itself and to know universals. The doctrine of the unity of the intellect that he ascribes to Aristotle he rejects as being counter to moral philosophy, since all rewards for those who are good and all punishments for those who are evil would thereby be destroyed. The view that virtue is its own reward and vice its own punishment he rejects as inadequate. And in his commentary on Aristotle's *On Generation and Corruption* he argues for personal immortality on the grounds that there is a natural desire in each human to enjoy a perpetual beatitude. Subsequently Agostino Nifo (§3) presented similar moral arguments for personal immortality in his early work *De intellectu (On the Intellect)* (1503).

Cajetan of Thiene, who succeeded Paul of Venice as the major professor at Padua, believed that Averroes' doctrine of the unity of the intellect could in fact be disproven by human reason and that probative arguments for personal immortality could be given. Johannes Argyropoulos, who had studied with Cajetan of Thiene, also offered arguments for personal immortality. In contrast, Nicoletto Vernia (§4), who succeeded to Cajetan of Thiene's chair at Padua, was a strict follower of Averroes regarding the unity of the intellect in an early treatise on the intellective

soul; in it, he denied that personal immortality could be demonstrated. In like fashion, Vernia's student, Agostino Nifo, stated in his earliest works that Averroes was the correct interpreter of Aristotle regarding the soul and intellect, and that personal immortality could not be demonstrated. However, a new set of factors dramatically changed the course of discussions regarding immortality, and both Vernia and Nifo came to hold that personal immortality could be demonstrated. These factors were the publication in 1481 of Ermolao Barbaro's Latin translation of Themistius' paraphrases on Aristotle's *On the Soul*, the publication in 1495 of Girolamo Donato's translation of Alexander of Aphrodisias' *On the Soul*, and the circulation of a now lost translation of the commentary on Aristotle's *On the Soul* attributed to Simplicius. These works, made available by the humanists, allowed philosophers to study the conflicting conceptions of the soul as subsistent and independent of the body and of the soul as the product of the harmony of the body. Vernia and Nifo fought over the correct interpretation of Alexander, since Vernia seemed unwilling to admit that Alexander denied personal immortality. Others like Giovanni Pico della Mirandola would also maintain the same mistaken interpretation of Alexander.

Both Vernia and Nifo later turned away from Averroes as the true interpreter of Aristotle regarding the soul and intellect and came to accept Themistius and Simplicius as truer guides. Pomponazzi too followed a similar path. He regarded Averroes as the true interpreter of Aristotle in his early questions on the soul and intellect, though he does also cite Alexander from Donato's translation. Eventually in his *De immortalitate animae* (1516) Pomponazzi (§2) uses arguments from Thomas Aquinas to discredit Averroes and his doctrine of the unity of the intellect, though he goes on to marshal arguments from Alexander to discredit Aquinas' attempt to demonstrate the intellective soul's independence of the body. He rejects moral arguments for immortality and retorts that virtue and vice are their own reward and punishment. It is noteworthy that there was a similar shift in the thought of the Dominican Cardinal Cajetan (§4) from acceptance of the demonstrability of personal immortality to a denial that such arguments have probative force. By this denial he was of course breaking with Thomas Aquinas, his intellectual master. Oddly enough, Antonio Trombetta, the major Franciscan theologian at Padua towards the end of the fifteenth century, did not follow his own master, Duns Scotus, on the question but held that immortality could be demonstrated.

Throughout the sixteenth century the possibility of demonstrating the immortality of the human soul continued to be debated. Simone Porzio (1496-1554), in his *De humana mente disputatio (On the Human Mind)* (first published in 1551), taught that the coming-to-be and passing-away of humans is not different from that of the animals. The intellect is inseparable from the body and can be considered to be material. Jacopo Zabarella clearly states in his commentary on Aristotle's *On the Soul* that according to true philosophy the human soul is a form giving existence to matter and yet separable and immortal, since it is not educed from the potency of matter but created by God. This view agrees with the Christian religion and with truth, but it is not the position of the philosophy of Aristotle. Indeed Zabarella notes in his own *Liber de mente humana (On the Human Mind)* (published in 1590) that it is doubtful that Aristotle knew of such a creation. Although he judges Pomponazzi to have come closest to the mind of Aristotle, he criticizes his views on what Aristotle meant by stating that the intellect comes from outside. On the other hand, Francesco Piccolomini (1520-1604) maintained that Aristotle considered the human soul to be immortal despite the contrary interpretation of others like Pomponazzi and Cardinal Cajetan. He takes Aristotle to have held that the human soul is not the form of the body but is related to the body as its actuality in the manner in which a sailor is the actuality of a ship. Finally it should be emphasized that Jesuit scholastics like Franciscus Toletus (§5) and Francisco Suárez (§3) offered rational arguments for the immortality of the soul (see Soul, nature and immortality of the).

6 Metaphysics

Various metaphysical topics debated during the Middle Ages continued to be of interest during the Renaissance. The nature of the distinction between essence and existence that divided Aquinas (§10) and Duns Scotus (§6) was studied by their Renaissance followers (see Báñez, D.; Cajetan §5; Fonseca, P. da §3; Nifo, A. §4; Suárez, F. §2), and whether being involved a univocal or an analogous concept was also debated (see Language, Renaissance philosophy of §4; Capreolus, J.; Cajetan §2; Fonseca, P. da §3; Silvestri, F.). The Dominican Paolo Barbo Soncinas (d. 1494) discussed analogy and the distinction of essence and existence. Trombetta and Maurice O'Fihely (1463-1514), both Franciscans, defended Scotus' views on the formal distinction and the univocal concept of being (see Duns Scotus §5). Cardinal Cajetan wrote a commentary on Aquinas' treatise *De ente et essentia (On Being and Essence)* and also claimed to set forth Aquinas' views in his own treatise *De Nominum*

Analogia (On the Analogy of Names) (1498). How close his thought is to Aquinas himself has been questioned. Achillini and Nifo both discussed the transcendentals, as did the Dominican Crisostomo Javelli (1470-1538) who defended Aquinas' views. Various Aristotelian thinkers like Elijah Delmedigo, Cajetan, Trombetta and Nifo all debated whether human reason could prove God's power to be infinite. Zimara, Delmedigo and Nifo also attempted to establish the mode of causality enjoyed by God and the Intelligences according to Aristotle and Averroes. Yet another topic of interest was whether according to reason and philosophy there could be a direct cognitive 'conjunction' of the human being with God and the Intelligences. This topic was inspired by Averroes and had already been discussed by medievals like Albert, Aquinas and John of Jandun. Nifo treated it seriously; Pomponazzi mocked it.

One of the most interesting developments to take place in Renaissance Aristotelianism concerned a conceptual scheme of metaphysical hierarchy first elaborated at the end of the thirteenth century. Albert the Great and Thomas Aquinas had spelled out this scheme, basing themselves on Proclus and Pseudo-Dionysius the Areopagite. God and nothingness (or matter) serve as two measures for the hierarchy of being. As a being 'approaches' God and 'recedes' from nonbeing, it holds an accordingly higher grade in the hierarchy. The scheme was also adopted with variations by Siger of Brabant (§4), Giles of Rome (§2), Henry of Ghent (§2), Godfrey of Fontaines (§3) and James of Viterbo. However, during the fourteenth century and thereafter it was challenged on the grounds that if God always remains at an 'infinite distance' from all creatures then no creature would be any closer to God than any other, and there would be no hierarchy of being.

One solution to this dilemma was to borrow from Richard Swineshead and the Oxford Calculators, who measured physical qualities like heat only from the lowest or zero grade (see Oxford Calculators). This route was pursued by Paul of Venice, who appears to be the target of Marsilio Ficino's attack on 'barbarians' (*barbari*) who do not take God as a measure of the hierarchy of being and settle on nonbeing alone (see Ficino, M.). A similar attack was mounted by Pomponazzi, who dismisses Swineshead's views and insists that the Aristotelian position is to measure things by the highest grade. On this topic Pomponazzi is anything but a 'radical philosopher'.

A major statement of the conceptual scheme of hierarchy as elaborated by Albert and Aquinas was given by Agostino Nifo, who attributed both the scheme and also a metaphysics of participation to Averroes. Marcantonio Zimara also attributed the scheme and a metaphysics of participation to Averroes. Other contemporaries who interested themselves in the scheme include Gabriele Zerbo (1435-1535), Trombetta, Cardinal Domenico Grimani (1461-1523), Achillini, the young Thomas de Vio (who was to become Cardinal Cajetan), Crisostomo Javelli and Gaspare Contarini (1483-1542). Later philosophers who examined and accepted the scheme include Francesco Buonamici (*c.*1540-1603) and Iacopo Mazzoni (1548-98). They were respectively the teacher and the colleague of Galileo Galilei at Pisa. Galileo himself presented arguments in his *juvenilia* against taking God as a measure of things (see Galilei, Galileo). The scheme was also rejected by Cesare Cremonini (1550-1631), Galileo's contemporary at Padua. One solution to the problem of how an infinite God could serve as a measure was to take God as finite. Achillini, Contarini and Buonamici accepted this solution.

7 Ethics and political philosophy

Aristotle's ethical writings interested and influenced a wide range of thinkers throughout the fifteenth and sixteenth centuries, from humanist authors such as Leonardo Bruni (1370-1444) to scholastic writers such as Francisco Suárez. Indeed the fifteenth century saw the diffusion of Aristotle's ethical thought to a new audience that was characterized by the concerns that cluster under the term 'humanism'. Bruni's approval of Aristotle marked the beginning of a new view, one that stressed Aristotle as a guide in rhetoric and moral philosophy and showed little interest in his views on logical, physical, psychological or metaphysical questions. Bruni thereby provided the inspiration for intellectuals who were not professional philosophers to use Aristotle as a guide to practical life and avoid more technical philosophical issues (see Humanism, Renaissance §§5, 7; Platonism, Renaissance §§1, 4).

At the beginning of the fifteenth century, the extant translations of Aristotle's *Nicomachean Ethics* were literal, technical and frequently relied on transliteration. To those taught according to humanist ideals, such translations were uninviting. Under the influence of Cicero and his own teacher Manuel Chrysoloras, Bruni developed a theory of translation unlike the typical medieval method. The sense of the Greek passage had to be understood and then translated according to the best models of Latin prose. Bruni claimed the translations used by scholastic

philosophers led them to misunderstand Aristotle even on minor points. He himself completed translations of the *Nicomachean Ethics* and the *Politics*, as well as the pseudo-Aristotelian *Economics*. His translation of the *Nicomachean Ethics* was hugely successful. He also wrote two works that reflect Aristotelian influence, namely an *Isagogicon moralis philosophiae (Introduction to Moral Philosophy)* (published between 1424 and 1426) and a *Vita Aristotelis (Life of Aristotle)* (published in 1428 or 1429). In the latter, Bruni takes Plato to task on several fronts, notably for his faulty methodology and for his views on such matters as the holding of wives and property in common.

The medieval Aristotelian tradition of political thought that was of Thomistic inspiration was passed on to Italy through two Florentine Dominicans who studied in Paris, namely Ptolemy of Lucca (*c.*1236-1327) and Remigio de'Girolami (1235-1319). Ptolemy had studied with Aquinas and wrote the continuation of Aquinas' *De regno (On Kingship)*, known also as *De regimine principum (On the Rule of Princes)*. Although Aquinas accorded primacy to monarchy, Ptolemy took Aristotle to prefer republican or elective rule regulated by law. Political rule is simply rule in accordance with written laws, while royal or despotic rule ignores law. Since the stars influence humans, some provinces are fit for liberty and others for servitude. The city-states of Northern Italy are suitable for political rule, since men there are self-confident and virile, like the ancient Romans. In contrast, the inhabitants of Sicily, Sardinia and Corsica are servile and therefore should have royal rule. Remigio attempted to conciliate Aristotle and Augustine, and also favoured Cicero as an authority and Ancient Rome as a model. In his *De bono communi (On the Common Good)* (written in 1302 or 1303), he discerned inordinate self-love as the cause of widespread discord; in his *De bono pacis (On the Good of Peace)* (written in 1304) he justified the ruler confiscating some individual's possessions for the greater good of the city.

Two Renaissance Dominicans need to be mentioned. Girolamo Savonarola (1452-98) wrote an early compendium of philosophy based on Aristotle and Aquinas that also reflects the thought of Ptolemy of Lucca (see Scepticism, Renaissance §3). Indeed his later treatise on governing Florence is a defence of republicanism dependent on Ptolemy's ideas in *De regimine principum*. He argues that since Florentines are vital and intelligent, their habit is to have civil rule. In the republicanism that he sets forth, he stresses the central role of a Great Council patterned on that of Venice. Another famed Dominican, Tommaso Campanella (§3), had a less favourable view of Aristotle. He noted that Aristotle denied that artisans, farmers and merchants could be citizens, since they lacked the virtues necessary to be either citizens or rulers. This would mean that only the nobles in Venice could be citizens: in fact they are tyrants, since other humans are thereby made their slaves and enemies. Consequently Aristotle's state is contrary to nature.

A more positive evaluation of Aristotle and the Venetian republic emerges from Francesco Piccolomini's *Universa philosophia de moribus (Universal Philosophy Concerning Ethical Matters)* (completed and published in 1583). He praises individual Venetians for their virtue and also the education offered young Venetian nobles at the University of Padua preparing them for service to Venice; nobles are more suited to be led to virtue. He also discusses at length how Venice is a mixed constitution and boldly claims that of all republics it approximates most closely to the republic that is ruled by God: the City of God.

Following Aristotle, Piccolomini takes ethics and political philosophy to be parts of one science. Whereas the subject of the *Nicomachean Ethics* involves human actions as they render humans good, the subject of the *Politics* is human actions as public good arises from them, namely the forming and the preserving of republics. The five grades of virtue are the natural, moral or civil, rational, heroic, and divine. Each represents a step towards becoming like God, the purpose of virtue. For Aristotle, human happiness is found primarily in contemplation and secondarily in moral goodness, while human virtue is found primarily in wisdom and only secondarily in moral virtue. Piccolomini rejects as against Aristotle the view of Averroes and Avicenna that the highest good in this life is a cognitive union with a separate intellect.

Piccolomini emphasizes the limits of Aristotle's ethics, noting that he knew neither another life nor union with God. Theologians know from divine revelation that the highest good of such a union is available only in the next life, but it is open to humans of every age and social class. In contrast, the highest good of Aristotle is available only in this life, and is not open to children, slaves, women and the poor. Aristotle's moral virtue forms one into an upright human being, citizen and ruler, but the theologian's infused or divine virtues of faith, hope and charity form one into a Christian and a citizen of heaven. And yet Piccolomini declares the Love Command of the Gospels

to be demanded by natural law.

See also: Aristotle; Aristotle Commentators; Aristotelianism in the 17th century; Aristotelianism, medieval; Averroism; Delmedigo, E.; Humanism, Renaissance; Language, Renaissance philosophy of; Logic, Renaissance; Melanchthon, P.; Nifo, A.; Renaissance philosophy; Vernia, N.

EDWARD P. MAHONEY
JAMES B. SOUTH

References and further reading

Copenhaver, B.P. and Schmitt, C.B. (1992) *Renaissance Philosophy*, Oxford and New York: Oxford University Press.(A readable general account that pays only limited attention to the scholastic tradition. Good bibliography.)

Garin, E. (1966) *Storia della filosofia italiana (History of Italian Philosophy)*, Turin: Piccola Biblioteca Einaudi, 499-580.(A detailed and insightful overview of Renaissance Aristotelianism by a master historian of philosophy.)

Kessler, E. (1988) 'The Intellective Soul', in C.B. Schmitt (ed.) *The Cambridge History of Renaissance Philosophy*, Cambridge: Cambridge University Press, 485-534.(An informative account of the cognitive theories of leading Renaissance Aristotelians.)

Kraye, J. (1993) 'The Philosophy of the Italian Renaissance', in G.H.R. Parkinson (ed.) *The Renaissance and Seventeenth-Century Rationalism*, London and New York: Routledge.(A brief introductory discussion with a topically-organized bibliography.)

Kristeller, P.O. (1974) 'Thomism and the Italian Thought of the Renaissance', ed. and trans. E.P. Mahoney, *Medieval Aspects of Renaissance Learning*, New York: Columbia University Press, 2nd edn; repr. 1992, 29-91. (A pioneering study of Aquinas' influence on Renaissance philosophy by a celebrated scholar.)

Kristeller, P.O. (1979) *Renaissance Thought and its Sources*, ed. M. Mooney, New York: Columbia University Press.(A model overview of Renaissance philosophy that discusses the different interpretations of Aristotle and the development of scholasticism.)

Lohr, C.H. (1988) *Latin Aristotle Commentaries: II, Renaissance Authors*, Florence: Olschki.(Alphabetically organized list of Renaissance Aristotelians, with biographies and bibliographies of primary and secondary sources.)

Mahoney, E.P. (1980) 'Albert the Great and the *Studio Patavino* in the Late Fifteenth and Early Sixteenth Centuries', in J.A. Weisheipl (ed.) *Albertus Magnus and the Sciences. Commemorative Essays 1980*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(Shows the important authority granted to Albert by major Renaissance Aristotelians.)

Mahoney, E.P. (1982a) 'Neoplatonism, the Greek Commentators, and Renaissance Aristotelianism', in D.J. O'Meara (ed.) *Neoplatonism and Christian Thought*, Albany, NY: State University of New York Press, 169-77, 264-82.(Demonstrates the impact of the Greek commentators and Ficino's translation of Plotinus on Renaissance Aristotelians.)

Mahoney, E.P. (1982b) 'Metaphysical Foundations of the Hierarchy of Being according to some Late Medieval and Renaissance Philosophers', in P. Morewedge (ed.) *Philosophies of Existence: Ancient and Medieval*, New York: Fordham University Press, 165-257.(A seminal article on the concept of the Great Chain of Being from Albert the Great to Cremonini and Galileo and early modern thought.)

Mahoney, E.P. (1988) 'Aristotle as "The Worst Natural Philosopher" (*pessimus naturalis*) and "The Worst Metaphysician" (*pessimus metaphysicus*): His Reputation among Some Franciscan Philosophers (Bonaventure, Francis of Meyronnes, Antonius Andreas, and Joannes Canonicus) and Later Reactions', in O. Pluta (ed.) *Die Philosophie im 14. und 15. Jahrhundert*, Amsterdam: B. Grüner, 261-73.(Some medieval attacks on Aristotle, and how they were received by Vernia and Nifo among others.)

Mahoney, E.P. (1995) 'From the Medievals to the Early Moderns: Themes and Problems in Renaissance Political Thought', in B. Carlos Bazán, E. Andújar and L.G. Sbrocchi (eds) *Les philosophies morales et politiques au Moyen Age/Moral and Political Philosophies in the Middle Ages*, Proceedings of the Ninth International Congress of Medieval Philosophy Ottawa (17-22 August 1992), Ottawa, Ont.: Legas, vol. 1, 193-225.(Pays particular attention to the influence of Plato and Aristotle as variously interpreted by Renaissance figures.)

Nardi, B. (1958) *Saggi sull'aristotelismo padovano dal secolo XIV al XVI (Essays on Paduan Aristotelianism from the 14th to the 16th Century)*, Università degli Studi di Padova: Studi sulla tradizione aristotelica nel Veneto 1, Florence: Sansoni.(Classic statement of Aristotelian tradition as found in Padua: old but still very useful.)

Olivieri, L. (ed.) (1983) *Aristotelismo veneto e scienza moderna. Atti del 25º anno accademico del Centro per la Storia della Tradizione Aristotelica nel Veneto (Venetian Aristotelianism and Modern Science)*, Padua: Antenore, 2 vols.(Rich collection of studies in several languages on a wide range of Aristotelian philosophers during the Renaissance.)

Poppi, A. (1966) *Causalità e infinità nella scuola padovana dal 1480 al 1513 (Causality and Infinity in the Paduan School from 1480 to 1513)*, Padua: Antenore.(Discusses issues such as God's intensive infinity and how this is related to God's efficient causality.)

Poppi, A. (ed.) (1983) *Scienza e filosofia all'Università di Padova nel quattrocento (Science and Philosophy at the University of Padua in the 15th Century)*, Contributi alla storia dell'Università di Padova 15, Padua: Edizioni Lint.(Studies on the philosophical and scientific ideas of fifteenth- and sixteenth-century thinkers including Copernicus and Pomponazzi.)

Schmitt, C.B. (1981) *Studies in Renaissance Philosophy and Science*, London: Variorum.(An important collection of informative studies by a distinguished Renaissance scholar.)

Schmitt, C.B. (1983) *Aristotle and the Renaissance*, Cambridge, MA, and London: Harvard University Press.(A very clear introduction to the subject that delineates the different types of Renaissance Aristotelianism.)

Skinner, Q. (1978) *The Foundations of Modern Political Thought*, Cambridge: Cambridge University Press, 2 vols.(A masterful account of late medieval, Renaissance and Reformation thought.)

Spruit, L. (1995) *Species Intelligibilis: From Perception to Knowledge. Vol. 2. Renaissance Controversies, Later Scholasticism, and the Elimination of the Intelligible Species in Modern Philosophy*, Brill's Studies in Intellectual History 49, Leiden: Brill.(Chronicles a large number of Italian and Spanish Aristotelians on the question of the representative nature of intellectual cognition.)

Wallace, W.A. (1972) *Causality and Scientific Explanation. Volume 1: Medieval and Early Classical Science*, Ann Arbor, MI: The University of Michigan Press.(Situates methodological doctrines of Renaissance Aristotelians in a tradition developing from Aristotle and his medieval commentators.)

Aristotle (384-322 BC)

Aristotle of Stagira is one of the two most important philosophers of the ancient world, and one of the four or five most important of any time or place. He was not an Athenian, but he spent most of his life as a student and teacher of philosophy in Athens. For twenty years he was a member of Plato's Academy; later he set up his own philosophical school, the Lyceum. During his lifetime he published philosophical dialogues, of which only fragments now survive. The 'Aristotelian corpus' (1462 pages of Greek text, including some spurious works) is probably derived from the lectures that he gave in the Lyceum.

Aristotle is the founder not only of philosophy as a discipline with distinct areas or branches, but, still more generally, of the conception of intellectual inquiry as falling into distinct disciplines. He insists, for instance, that the standards of proof and evidence for deductive logic and mathematics should not be applied to the study of nature, and that neither of these disciplines should be taken as a proper model for moral and political inquiry. He distinguishes philosophical reflection on a discipline from the practice of the discipline itself. The corpus contains contributions to many different disciplines, not only to philosophy.

Some areas of inquiry in which Aristotle makes a fundamental contribution are these:

(1) Logic. Aristotle's Prior Analytics *constitutes the first attempt to formulate a system of deductive formal logic, based on the theory of the 'syllogism'. The* Posterior Analytics *uses this system to formulate an account of rigorous scientific knowledge. 'Logic', as Aristotle conceives it, also includes the study of language, meaning and their relation to non-linguistic reality; hence it includes many topics that might now be assigned to philosophy of language or philosophical logic (*Categories*,* De Interpretatione*,* Topics*).*

(2) The study of nature. About a quarter of the corpus (see especially the History of Animals*,* Parts of Animals*, and* Generation of Animals*; also* Movement of Animals*,* Progression of Animals*) consists of works concerned with biology. Some of these contain collections of detailed observations. (The* Meteorology *contains a similar collection on inanimate nature.) Others try to explain these observations in the light of the explanatory scheme that Aristotle defends in his more theoretical reflections on the study of nature. These reflections (especially in the* Physics *and in* Generation and Corruption*) develop an account of nature, form, matter, cause and change that expresses Aristotle's views about the understanding and explanation of natural organisms and their behaviour. Natural philosophy and cosmology are combined in* On the Heavens*.*

(3) Metaphysics. In his reflections on the foundations and presuppositions of other disciplines, Aristotle describes a universal 'science of being qua *being', the concern of the* Metaphysics*. Part of this universal science examines the foundations of inquiry into nature. Aristotle formulates his doctrine of substance, which he explains through the connected contrasts between form and matter, and between potentiality and actuality. One of his aims is to describe the distinctive and irreducible character of living organisms. Another aim of the universal science is to use his examination of substance to give an account of divine substance, the ultimate principle of the cosmic order.*

(4) Philosophy of mind. The doctrine of form and matter is used to explain the relation of soul and body, and the different types of soul found in different types of living creatures. In Aristotle's view, the soul is the form of a living body. He examines the different aspects of this form in plants, non-rational animals and human beings, by describing nutrition, perception, thought and desire. His discussion (in On the Soul*, and also in the* Parva Naturalia*) ranges over topics in philosophy of mind, psychology, physiology, epistemology and theory of action.*

*(5) Ethics and politics (*Nicomachean Ethics*,* Eudemian Ethics*,* Magna Moralia*). In Aristotle's view, the understanding of the natural and essential aims of human agents is the right basis for a grasp of principles guiding moral and political practice. These principles are expressed in his account of human wellbeing, and of the different virtues that constitute a good person and promote wellbeing. The description of a society that embodies these virtues in individual and social life is a task for the* Politics*, which also examines the virtues and vices of actual states and societies, measuring them against the principles derived from ethical theory.*

*(6) Literary criticism and rhetorical theory (*Poetics*,* Rhetoric*). These works are closely connected both to Aristotle's logic and to his ethical and political theory.*

1 Life

Aristotle was born in 384 BC, in the Macedonian city of Stagira, now part of northern Greece. In his lifetime the kingdom of Macedon, first under Philip and then under Philip's son Alexander ('the Great'), conquered both the Greek cities of Europe and Asia and the Persian Empire. Although Aristotle spent much of his adult life in Athens, he was not an Athenian citizen. He was closely linked to the kings of Macedon, whom many Greeks regarded as foreign invaders; hence, he was affected by the volatile relations between Macedon and the Greek cities, especially Athens.

Aristotle was the son of Nicomachus, a doctor attached to the Macedonian court. In 367 BC Aristotle came to Athens. He belonged to Plato's Academy until the death of Plato in 347; during these years Plato wrote his important later dialogues (including the *Sophist*, *Timaeus*, *Philebus*, *Statesman*, and *Laws*), which reconsider many of the doctrines of his earlier dialogues and pursue new lines of thought. Since there was no dogmatic system of 'Platonism', Aristotle was neither a disciple of such a system nor a rebel against it. The exploratory and critical outlook of the Academy probably encouraged Aristotle's own philosophical growth.

In 347 BC Aristotle left Athens, for Assos in Asia Minor. Later he moved to Lesbos, in the eastern Aegean, and then to Macedon, where he was a tutor of Alexander. In 334 he returned to Athens and founded his own school, the Lyceum. In 323 Alexander died; in the resulting outbreak of anti-Macedonian feeling in Athens Aristotle left for Chalcis, on the island of Euboea, where he died in 322.

Aristotle married Pythias, a niece of Hermeias, the ruler of Assos. They had a daughter, also called Pythias. After the death of his wife, Aristotle formed an attachment to Herpyllis, and they had a son Nicomachus.

2 Order of Aristotle's works

By the end of Aristotle's life the Lyceum must have become a well-established school. It lasted after Aristotle's death; his successor as head of the school was his pupil Theophrastus. Many of the works in the Aristotelian corpus appear to be closely related to Aristotle's lectures in the Lyceum. The polished character of some passages suggests preparation for publication (for example, *Parts of Animals* I 5), but many passages contain incomplete sentences and compressed allusions, suggesting notes that a lecturer might expand (for example, *Metaphysics* VII 13). We cannot tell how many of his treatises Aristotle regarded as 'finished' (see §11 on the *Metaphysics* and §21 on the *Ethics*).

It may be wrong, therefore, to ask about the 'date' of a particular treatise. If Aristotle neither published nor intended to publish the treatises, a given treatise may easily contain contributions from different dates. For similar reasons, we cannot plausibly take cross-references from one work to another as evidence of the order of the works. External, biographical considerations are unhelpful, since we lack the evidence to support any detailed intellectual biography of Aristotle.

A few points, however, may suggest a partial chronology.

(1) Some of Aristotle's frequent critical discussions of Plato and other Academics may have been written (in some version) during Aristotle's years in the Academy. The *Topics* may reflect the character of dialectical debates in the Academy.

(2) It is easier to understand the relation of the doctrine of substance in the *Categories* and *Physics* I-II to the doctrine and argument of *Metaphysics* VII if we suppose that *Metaphysics* VII is later.

(3) The Organon (see §4) does not mention matter, perhaps because (a) Aristotle had not yet thought of it, or because (b) he regarded it as irrelevant to the topics considered in the Organon. If (a) is correct, the Organon precedes the works on natural philosophy.

(4) Some of the observations used in Aristotle's biological works probably came from the eastern Aegean. Hence, Aristotle probably pursued his biological research during his years away from Athens. We might trace his biological interests to the Academy (see Plato's *Timaeus*); he may also have acquired them from his father Nicomachus, who was a doctor. Probably, then, at least some of the biological works (or versions of them) are not the latest works in the corpus.

(5) The *Magna Moralia* (if it is genuine) and the *Eudemian Ethics* probably precede the *Nicomachean Ethics* (see §21).

The order in which Aristotle's works appear in the Greek manuscripts goes back to early editors and commentators (from the first century BC to the sixth century AD); it reflects their view not about the order in which the works were written, but about the order in which they should be studied. This entry generally follows the order of the corpus, except that it discusses *On the Soul* after the *Metaphysics* (see §17), not among the works on natural philosophy (where it appears in the manuscripts).

3 Appearances

The general aim of rational inquiry, according to Aristotle, is to advance from what is 'better known to us' to what is 'better known by nature' (see *Physics* I 1; *Posterior Analytics* 71b33; *Metaphysics* 1029b3). We achieve this aim if: (1) we replace propositions that we thought we knew with propositions that we really know because they are true and we understand them; (2) we find general principles that explain and justify the more specific truths that we began from; (3) we find those aspects of reality that explain the aspects that are more familiar to us.

The things better known to us in a particular area are the relevant 'appearances' (*phainomena*). Aristotle presents them through detailed collections of empirical data, reached as a result of 'inquiry' (*historia*; for example, *Parts of Animals* 646a8). Empirical inquiry proceeds from particular observations, by means of generalizations through induction (*epagōgē*) from these particular cases, until we reach experience (*empeiria*). Experience leads us to principles that are better known by nature (*Prior Analytics* 46a17); we also rely on it to test principles we have found (*Generation of Animals* 760b28).

Philosophical inquiry also relies on 'appearances'. However, the appearances that concern it are not empirical observations, but common beliefs, assumptions widely shared by 'the many and the wise'. The critical and constructive study of these common beliefs is 'dialectic'. Aristotle's method is basically Socratic. He raises puzzles in the common beliefs, looking for an account that will do them justice as a whole. Among common beliefs Aristotle considers the views of his predecessors (for example, *Metaphysics* I; *On the Soul* I; *Politics* II), because the puzzles raised by their views help us to find better solutions than they found.

Inquiry leads us to causes and to universals. Aristotle has a realist conception of inquiry and knowledge; beliefs and theories are true in so far as they grasp the reality that we inquire into (see Realism and antirealism §2). Universals and causes are 'prior by nature'; they are not created by, or dependent on, any theory, but a true theory must fit them.

If we attended only to Aristotle's remarks on what is better known to us and on the process of inquiry, we might regard his position as a form of empiricism (see Empiricism). But in his remarks on what is better known by nature, he insists on the reality of universals and on the importance of non-sensory forms of knowledge (see §15 on universals, §19 on thought).

4 Thought and language

One means of access to appearances, and especially to common beliefs, is the study of what words and sentences 'signify' (*sēmainein*). This is part of 'logic' (*logikē*, derived from *logos*, which may be translated 'word', 'speech', 'statement', 'argument' or 'reason': see Logos), which is discussed in the first section of Aristotle's works (*Categories*, *De Interpretatione*, *Prior Analytics*, *Posterior Analytics*, *Topics*). This section of the corpus came to be called the 'Organon' ('instrument'), because logic, as Aristotle conceives it, concerns statements and arguments in general, without restriction to any specific subject matter; it is therefore an instrument of philosophical inquiry in general, rather than a branch of philosophy coordinate with natural philosophy or ethics. The Organon includes some elements of philosophy of language, as well as formal logic (syllogistic; see §5) and epistemology (see §6).

According to Aristotle's account of signification (see especially *De Interpretatione* 1-4), as commonly understood, the word 'horse' signifies horse by signifying the thought of horse; in using the word, we communicate thoughts about horses. When the thoughts about horses we communicate are true, we communicate truths about the universal horse; even when our thoughts are not completely true, we may signify the same universal horse.

To understand the signification of a name '*F*', we look for the corresponding definition (*logos*, *horismos*) of *F*.

Aristotle distinguishes nominal definitions, stating the beliefs associated with the name, from real definitions, giving a true account of the universal that underlies the beliefs embodied in the nominal definition (see *Posterior Analytics* II 8-10. Aristotle himself does not use the labels 'nominal definition' and 'real definition'.).

Not every name corresponds to one nominal and one real definition. Some names correspond to no genuine universal; 'goatstag' signifies (in one way) animals that are both goats and stags, but it does not signify a genuine universal, since there is no natural kind of goatstag. Other names correspond to more than one universal, as 'chest' signifies both a container and a part of an animal. Chests are 'homonymous' (*homōnyma*) or 'multivocal' (*pollachōs legomena*; 'spoken of in many ways'); more than one definition is needed to capture the signification of the name. By contrast, since only one definition corresponds to the name 'horse', horses are 'synonymous' (*Categories* 1).

Other philosophers make serious errors, Aristotle believes, because they suppose they can give a single account of things or properties that are really multivocal. Once we see that different *F*s are *F* in different ways, we see that different, although (in many cases) connected, accounts of what it is to be *F* must be given. Some philosophically important cases of multivocity are cause (Aristotle's doctrine of the four causes; see §9), being (the doctrine of the categories; see §7) and good (the criticism of Plato's belief in a Form of the Good; *Nicomachean Ethics* I 6).

5 Deduction

Part of logic, as Aristotle conceives it, is the study of good and bad arguments. In the *Topics* Aristotle treats dialectical arguments in general. In the *Prior Analytics* he examines one type of argument, a 'deduction' (*syllogismos*; literally, 'reasoning', hence the standard term 'syllogism'). This is an argument in which, if propositions *p* and *q* are assumed, something else *r*, different from *p* and *q*, follows necessarily because of the truth of *p* and *q* (*Prior Analytics* 24b18-20, paraphrased). Aristotle insists that it is not possible for the premises of a deduction to be true and the conclusion false ('follows necessarily'); that a deduction must have more than one premise ('if *p* and *q* are assumed'); that the conclusion cannot be identical to any premise ('different from *p* and *q*'); and that no redundant premises are allowed ('because of the truth of *p* and *q*'). He takes deductions to express affirmative or negative relations between universals, taken either universally ('Animal belongs to every (no) man') or not universally ('Animal belongs (does not belong) to some man'). He takes the affirmative and negative claims to imply existence (so that 'Biped belongs to some dodo' follows from 'Biped belongs to every dodo'; the latter affirmation is not equivalent, therefore, to 'If anything is a dodo, it is biped').

These different features of an Aristotelian deduction differentiate Aristotle's account of a deduction from a more familiar account of deductively valid arguments. An argument may be valid even if it is redundant, or a premise is identical to the conclusion, or it has only one premise, or it is about particulars, or it contains neither 'some' nor 'every' nor 'belongs'; but no such argument is an Aristotelian deduction. Aristotle's theory of the different forms of deduction (often called 'the moods of the syllogism') examines the various forms of argument that necessarily preserve the truth of their premises. He begins from 'complete' (or 'perfect') deductions whose validity is evident, and classifies the different types of arguments that can be derived from (shown to be equivalent to) the complete deductions. He also explores the logical relations between propositions involving modalities ('Necessarily (possibly) animal belongs to every man' and so on). Since Aristotle accepts this relatively narrow account of a deduction, his exploration of the different forms of deduction is not a theory of valid arguments in general; the Stoics come much closer to offering such a theory (see Stoicism §11; Logic, ancient).

Aristotle's theory of deduction is developed for its own sake, but it also has two main philosophical applications. (1) Deduction is one type of argument appropriate to dialectic (and, with modifications, to rhetoric; see §29). Aristotle contrasts it with inductive argument (also used in dialectic), in which the conclusion does not follow necessarily from the premises, but is made plausible by them. (2) It is essential for demonstration (*apodeixis*), which Aristotle takes to be the appropriate form for exhibiting scientific knowledge.

6 Knowledge, science and demonstration

The progress from what is known to us to what is known by nature aims at *epistēmē*, the scientific knowledge whose structure is exhibited in the demonstrative pattern described in the *Posterior Analytics*. A demonstration is a deduction in which the premises are necessarily true, prior to and better known than the conclusions, and

explanatory of the conclusions derived from them. Aristotle assumes that if I know that *p*, then I can cite some justification *q*, to justify my belief that *p*, and I also know why *q* justifies *p* (*Posterior Analytics* I 2). The right sort of justification relies on things better known by nature - the general laws and principles that explain the truth of *p*. Since these are embodied in demonstrations, grasp of a demonstration of *p* expresses knowledge of *p*. Aristotle's theory of demonstration, then, is not intended to describe a procedure of scientific inquiry that begins from appearances; it is an account of the knowledge that is achieved by successful inquiry.

To show that a deduction is a demonstration, we must show that its premises are better known than the conclusion. Sometimes we can show this by demonstrating them from higher premises that are even better known. This process of justification, Aristotle claims, must be linear and finite. A circular 'justification' must eventually 'justify' a given belief by appeal to itself, and an infinite regress imposes on us a task that we can never complete. Since, therefore, neither a circle nor an infinite regress can really justify, a proper justification must ultimately appeal to primary principles of a science.

These primary principles are 'assumptions' (*hypotheseis*); we must see that they are better known and prior to other truths of a science, without being derived from any further principles. Since they are the basis of all demonstration, they cannot themselves be demonstrated; Aristotle claims that we have non-demonstrative understanding (*nous*: *Posterior Analytics* II 19) of the ultimate principles of each science (see Nous).

How are we entitled to claim understanding of an ultimate principle? Aristotle believes that the principles of a science are reached from appearances (perceptual or dialectical or both), which are the starting points known to us. He may believe that this relation of the principles to appearances justifies us in accepting them as first principles and in claiming to have understanding of them. This explanation, however, does not easily fit Aristotle's demand for linear and finite chains of justification. That demand suggests that the assumptions of a science must be self-evident (seen to be true without any inferential justification), so that his conception of knowledge expresses a foundationalist position (see Foundationalism §3). (On difficulties in foundationalism see Agrippa.)

Although Aristotle's aim of reaching a demonstrative science reveals some of his epistemological doctrines and assumptions, it does not evidently influence most of the structure or content of most of the surviving treatises. In his main philosophical works, the influence of dialectical methods and aims is more apparent.

7 Categories and beings

Part of the task of logic is to explain the nature of predication ('*A* is *B*', analysed by Aristotle as '*B* is predicated of *A*' or '*B* belongs to *A*', as in 'Animal belongs to every man'), which is presupposed by complex *logoi* (statements and arguments). In the *Categories* (*katēgoriai*; predications), Aristotle introduces ten 'categories' (usually called *schēmata tēs katēgorias*, 'figures (that is, types) of predication'). The categories correspond to different sorts of words (for example, count-nouns, adjectives, verbs) and to different grammatical functions (for example, subject, predicate), but they primarily classify the different non-linguistic items introduced in predications. The sentences 'Socrates is a man' and 'Socrates is a musician' are grammatically similar, but they introduce different sorts of things; the first predicates a second substance of a first substance, whereas the second predicates a non-substance of a first substance.

The first category is called *ousia* (literally, 'being'), which is translated into Latin as 'substantia', and hence usually called 'substance' (see Substance §1). The nine non-substance categories include quality, quantity and relative (the only ones that Aristotle refers to often; the categories are listed in *Categories* 4, *Topics* I 9). Each category contains both particulars and universals. The statement that this individual man is an animal predicates a second substance (that is, a universal in the category of substance) of a first substance (that is, a particular in the category of substance). 'White is a colour' predicates one universal quality of another.

The categories display the multivocity of beings (see §4). Whereas animals constitute an ordinary univocal genus with a single definition, beings do not constitute an ordinary genus; hence there is no single account of what it is for something to be a being. Aristotle believes Plato mistakenly pursued a single account of beings; the theory of categories is meant to avoid Platonic errors.

In marking categorial divisions, Aristotle is influenced by grammar and syntax, but also by his ontology - his classification of beings. This classification rests on his view of nature and change, which clarifies his analysis of

predication.

8 Change and substance

Aristotle's *Physics* discusses nature, *physis*. The nature of *x* is a principle (or 'source'; *archē*), internal to *x*, of change and stability in *x*; hence the inquiry into nature leads to a discussion of change in natural substances (the elements, plants and animals). Aristotle proceeds dialectically, raising and solving puzzles involved in the understanding of natural change. In solving the puzzles, he introduces the different types of beings that are presupposed by a coherent account of natural change.

In *Physics* I 7-8, Aristotle analyses a simple example of change - Socrates changing from being pale to being tanned. This change involves a subject (or 'underlying thing'; *hypokeimenon*), Socrates, who loses one contrary (his pale colour) and acquires another contrary (his tan). Neither of the contraries persists, but the subject persists (otherwise there would not be a change in Socrates). This particular subject that persists through change is what the *Categories* calls a first substance. First substances differ both from second substances and from non-substances by being capable of undergoing change; they persist while receiving opposites (as Socrates is first pale and then tanned). They cannot, however, remain in existence irrespective of any properties gained or lost; Socrates' ceasing to be a man is not a change in Socrates, but the perishing of Socrates.

The properties that a first substance cannot lose without perishing constitute (approximately) the essence of that first substance (see Essentialism). These essential properties define a kind to which the first substance belongs. A kind may be a species (*eidos*), for example, man or horse, or a genus (*genos*), for example, animal. In predicating a second substance of a first substance (as in 'Socrates is a man'), we place the first substance in the kind it belongs to. If we predicate one of the contraries that the first substance can lose without perishing, we introduce an item (Socrates' pale colour, his particular height, his ignorance, his being the husband of Xanthippe) in one of the non-substance categories (quality, quantity, relative, and so on). The kinds to which these non-substantial items belong are non-substantial universals.

Aristotle also examines the coming to be and perishing of a first substance. Here again, he distinguishes a persisting subject and two contraries. If we make a statue from bronze, the lump of bronze (the subject) acquires the shape of the statue, and loses the shapelessness it had, and so changes between contraries. But although the lump remains in existence, a new subject, the statue, has come into being. In this case, the subject of the change is the matter (*hylē*), and what it acquires is the form (*eidos*, also rendered 'species').

This analysis of change suggests an argument (*Physics* II 1) to show that the genuine subject, and hence the genuine substance, is the matter, whereas the apparent substance (for example, the statue) is simply matter with a certain shape. Socrates does not become another subject if he changes shape; hence (we may argue) the lump of bronze does not become another subject simply by acquiring the shape of a statue. Similarly, then, a natural organism might be understood as a piece of matter shaped in a certain way so as to embody Socrates. Natural organic 'substances', such as Socrates and this tree, turn out to be not genuine subjects, but mere configurations of the matter that is the real substance.

Aristotle does not endorse this eliminative attitude to natural organic substances. He uses the argument to raise a puzzle about whether matter or form is substance. He discusses this puzzle in *Metaphysics* VII (see §12-14). This discussion relies on his account of causation and explanation.

9 Causes

When we correctly answer questions such as 'Why does this event happen?' or 'Why is this object as it is?', we state the cause (or explanation; *aition*) of the event or object. Aristotle believes that causes are multivocal (see *Physics* II 3; *Metaphysics* I 3). Different accounts of a cause correspond to different answers to why-questions about (for example) a statue. (1) 'It is made of bronze' states the material cause. (2) 'It is a statue representing Pericles' states the formal cause, by stating the definition that says what the thing is. (3) 'A sculptor made it' states the 'source of change', by mentioning the source of the process that brought the statue into being; later writers call this the 'moving cause' or 'efficient cause'. (4) 'It is made to represent Pericles' states 'that for the sake of which', since it mentions the goal or end for the sake of which the statue was made; this is often called the 'final' (Latin *finis*; 'end') cause.

Each of the four causes answers a why-question. Sometimes (as in our example) a complete answer requires all four causes. Not all four, however, are always appropriate; the (universal) triangle, for example, has a formal cause, stating its definition, but no efficient cause, since it does not come into being, and no final cause, since it is not made to promote any goal or end.

Some have claimed that Aristotle's 'four causes' are not really causes at all, pointing out that he takes an *aition* to be available even in cases where the why-question (for example, 'Why do the interior angles of this figure add up to two right angles?') does not seek what we would call a cause (in Aristotle's division, an efficient cause). When explanations of changes are being sought, however, Aristotle seems to provide recognizably causal explanations. Even the *aitia* (material, formal, final) that do not initially seem to be causes turn out to play an important role in causal explanation; for this reason, the label 'four causes' gives a reasonably accurate impression of Aristotle's doctrine.

His comparison between artefacts and natural organisms clarifies his claims about formal and final causes. The definition of an artefact requires reference to the goal and the intended function. A hammer's form and essence is a capacity to hammer nails into wood. The hammer was designed to have this capacity for performing this function; and if this had not been its function, it would not have been made in the way it was, to have the properties it has. The form includes the final cause, by specifying the functions that explain why the hammer is made as it is.

Similarly, Aristotle claims, a natural organism has a formal cause specifying the function that is the final cause of the organism. The parts of an organism seem to perform functions that benefit the whole (the heart pumps blood, the senses convey useful information). Aristotle claims that organs have final causes; they exist in order to carry out the beneficial functions they actually carry out. The form of an organism is determined by the pattern of activity that contains the final causes of its different vital processes. Hence Aristotle believes that form as well as matter plays a causal role in natural organisms.

To claim that a heart is for pumping blood to benefit the organism is to claim that there is some causal connection between the benefit to the organism and the processes that constitute the heart's pumping blood. Aristotle makes this causal claim without saying why it is true. He does not say, for instance, either (1) that organisms are the products of intelligent design (as Plato and the Stoics believe), or (2) that they are the outcome of a process of evolution.

Aristotle's account of causation and explanation is expressed in the content and argument of many of his biological works (including those connected with psychology). In the *Parts of Animals* and *Generation of Animals* for instance, he examines the behaviour and structure of organisms and their parts both to find the final causes and to describe the material and efficient basis of the goal-direction that he finds in nature (*Parts of Animals* I 1). He often argues that different physiological processes in different animals have the same final cause.

Some ascribe to Aristotle an 'incompatibilist' view of the relation between final causes and the underlying material and efficient causes. Incompatibilists concede that every goal-directed process (state, event) requires some material process (as nutrition, for example, requires the various processes involved in digesting food), but they argue that the goal-directed process cannot be wholly constituted by any material process or processes; any process wholly constituted by material processes is (according to the incompatibilist) fully explicable in material-efficient terms, and therefore has no final cause.

Probably, however, Aristotle takes a 'compatibilist' view. He seems to believe that even if every goal-directed process were wholly constituted by material processes, each of which can be explained in material-efficient terms, the final-causal explanation would still be the only adequate explanation of the process as a whole. According to this view, final causes are irreducible to material-efficient causes, because the explanations given by final causes cannot be replaced by equally good explanations referring only to these other causes. This irreducibility, however, does not require the denial of material constitution.

10 Change

Aristotle studies nature as an internal principle of change and stability; and so he examines the different types of change (or 'motion'; *kinēsis*) that are found in the natural elements and in the natural organisms composed of them. In *Physics* III 1 he defines change as 'the actuality of the potential *qua* potential'. His definition marks the

importance of his views on potentiality (or ‘capacity’; *dynamis*) and actuality (or ‘realization’; *energeia* or *entelecheia*) (see *Metaphysics* IX 1-9).

The primary type of potentiality is a principle (*archē*) of change and stability. If *x* has the potentiality *F* for *G*, then (1) *G* is the actuality of *F*, and (2) *x* has *F* because *G* is the actuality of *F*. Marathon runners, for instance, have the potentiality to run 26 miles because they have been trained to run this distance; hearts have the capacity to pump blood because this is the function that explains the character of hearts. In these cases, potentialities correspond to final causes.

Potentiality and possibility do not, therefore, imply each other. (1) Not everything that is possible for *x* realizes a potentiality of *x*. Perhaps it is possible for us to speak words of Italian (because we recall them from an opera) without having a potentiality to speak Italian (if we have not learnt Italian). (2) Not everything that *x* is capable of is possible for *x*; some creatures would still have a potentiality to swim even if their environment lost all its water.

These points about potentiality help to clarify Aristotle’s definition of change. The building of a house is a change because it is the actuality of what is potentially built in so far as it is potentially built. ‘What is potentially built’ refers to the bricks (and so on). The completed house is their complete actuality, and when it is reached, their potentiality to be built is lost. The process of building is their actuality in so far as they are potentially built. ‘In so far… ’ picks out the incomplete actuality that is present only as long as the potentiality to be built (lost in the completed house) is still present. Aristotle’s definition picks out the kind of actuality that is to be identified with change, by appealing to some prior understanding of potentiality and actuality, which in turn rests on an understanding of final causation.

In the rest of the *Physics*, Aristotle explores different properties of change in relation to place and time. He discusses infinity and continuity at length, arguing that both change and time are infinitely divisible. He tries to show that the relevant type of infinity can be defined by reference to potentiality, so as to avoid self-contradiction, paradox or metaphysical extravagance. In his view, infinite divisibility requires a series that can always be continued, but does not require the actual existence of an infinitely long series. Once again, the reference to potentiality (in ‘can always… ’) has a crucial explanatory role.

11 Metaphysics

Some of the basic concepts of the *Categories* and *Physics* - including substance, particular, universal, form, matter, cause and potentiality - are discussed more fully in the *Metaphysics*. This is a collection of fourteen books, some of them loosely connected. Aristotle probably did not deliver a course of lectures in the order of the present treatise. Parts of book I are almost repeated in book XIII. Book V is a ‘philosophical dictionary’ that seems to interrupt the argument of books IV and VI. Book XI summarizes parts of book IV. Books II and XI were probably not written entirely by Aristotle.

Still, whatever their literary origins, all these books have a common subject matter, since they all contribute to the universal science that studies the common presuppositions of the other sciences. This universal science has four names. (1) ‘First philosophy’: it studies the ‘first principles’ and ‘highest causes’ (including the four causes of the *Physics*) presupposed by the other sciences. (2) ‘The science of being’: every science presupposes that it studies some sort of being, and the science of being examines and defends this presupposition. (3) ‘Theology’: first philosophy is not only first in so far it is most universal, but also in so far as it deals with the primary sort of being, the sort on which all other beings depend. The primary sort of being is substance, and the primary sort of substance is divine substance; hence the science of being must study divine substance. (4) ‘Metaphysics’ (*ta meta ta physika*; ‘the things after the natural things’): it is ‘after’ or ‘beyond’ the study of nature because (a) as theology, it studies entities outside the natural order, and (b) as first philosophy, it starts from the study of nature (which is prior and better known ‘to us’) and goes beyond it to its foundations and presuppositions (which are prior and better known ‘by nature’; see §3).

The first three of these names are used by Aristotle himself (*Metaphysics* IV 1-3, VI 1). The fourth was given to the treatise in antiquity (at an uncertain date); its use of ‘after’ captures Aristotle’s different claims about the relation of the universal science to other sciences.

The universal science is the science of being *qua* being - that is, being in so far as it is being - just as mathematics is the science of some beings *qua* mathematical objects (see §16) and physics is the science of some beings *qua* changeable. The science of being studies the beings that are also studied by other sciences, but it isolates the relevant properties of beings by a different level of abstraction; it does not rely on the fact that they have the properties of mathematical or natural objects, but simply on the fact that they are beings studied by a science (*Metaphysics* IV 1-2).

A special science assumes that it begins with a subject that has properties. The universal science is the science of being because it studies the sort of subject that is presupposed by the other sciences; and it is primarily the science of substance because substance is the primary sort of being. Aristotle's analysis of change in *Physics* I introduces substances as subjects; the *Metaphysics* asks what sorts of subjects and substances must be recognized by special sciences.

Aristotle argues that if we are to signify a subject, it is impossible for each of its properties both to belong and not to belong to it. This principle is often called the 'Principle of Non-Contradiction' (*Metaphysics* IV 3-4). To defend the principle, Aristotle considers an opponent who is willing to assert that a single subject, man, is both a bipedal animal and not a biped animal. If the opponent really says this about a single subject, then, when he uses 'man', he must signify one and the same subject, man. If he agrees that in using 'man' he signifies a biped animal, then he cannot also deny that man is a biped animal; for if he denies this, he can no longer say what 'man' signifies, and hence he cannot say what subject it is that he takes to be both a biped animal and not a biped animal. This property (which one cannot also deny of a subject) is an essential property. Hence, the attempt to reject subjects with essential properties is self-undermining.

Subjects of change must also, according to Aristotle, have objective properties (that is, properties that they have whether or not they appear to have them). An argument against Protagoras seeks to show that any attempt to reject objective properties undermines itself (*Metaphysics* IV 5). Protagoras denies that there are any objective properties, because he claims that how things appear to someone is how they are. If he is to maintain the infallibility of appearances against any possibility of correction, then, Aristotle argues, he must claim that it is possible for the same subject to change in every respect at every time (to match different appearances). This is possible, however, only if the same subject can remain in being, but change in all respects. Aristotle replies that if the same subject persists, it must keep the same essential property (the 'form'); hence it cannot change in every respect (IV 5).

12 From being to substance

In *Metaphysics* IV 2 and VII 1 Aristotle argues that, since substance is the primary type of being and other beings are in some way dependent on substances, the science of being must primarily be concerned with substance. The arguments of IV 4-5 describe some features of substances; they must be subjects with stable, objective, essential properties. Books VII-IX describe these subjects more fully, by re-examining the conception of substance that is presented in the *Categories* and *Physics* (see §§7-8).

Aristotle observes that we regard substance both as 'a this' and as 'essence' (or 'what it is'). We might assume that these two descriptions pick out two sorts of substances - a particular subject ('this') and a universal ('what it is'), corresponding to the first and second substances of the *Categories*. Aristotle, however, insists that his question 'What is substance?' will be satisfactorily answered only when we have found the one thing that best satisfies the conditions for being both a subject (a 'this') and an essence ('what it is'). Whatever best satisfies these conditions is primary substance.

The different candidates that Aristotle considers for this role are matter, form and the compound of the two. He argues against the first and third candidates, and defends the second. He regards matter and compound as types of substance, but argues that they are secondary to form because they do not meet the relevant conditions to the same degree. To show that form is primary substance, he argues that a form is both a subject and an essence of the right sort. In books VIII-IX he clarifies his answer by identifying form with the actuality for which the matter is the potentiality.

13 Why is form substance?

In claiming that form is substance, Aristotle relies on the connections between form, cause, essence and identity. He rejects the eliminative view (§8) that the so-called 'coming-to-be' or 'perishing' of an artefact or organism is simply an alteration of the matter. According to the eliminative view, this alteration does not involve the existence or non-existence of a distinct substance, any more than Socrates' coming to be musical involves the existence of a distinct substance, musical Socrates. Aristotle replies that the production of an artefact and the generation of an organism introduce a new subject, a substance that is neither identical to nor wholly dependent on the matter that constitutes it at a time (see Identity §2). Although this statue of Pericles has come into being from a particular piece of bronze, we may repair the statue by replacing damaged bits; we preserve the same statue but we cause a different bit of bronze to constitute it. Similarly, an organism remains in existence as long as it replaces its matter with new matter: it persists as long as its form persists (*Generation and Corruption* I 5).

When Aristotle speaks of the relation of form to matter, he may refer to either of two kinds of matter: (1) the proximate, organic matter (for example, the organs and limbs making up the organic body); and (2) the remote, non-organic matter (for example, blood, earth, water) of which the organic body is made. Remote matter can exist without the form of the organism, but the organism can persist without any particular piece of remote matter. Proximate matter cannot exist without the form (since it is the function of an arm or heart that makes it the limb or organ it is); the form is the actuality of which the proximate matter is the potentiality (*On the Soul* 412a10; *Metaphysics* 1038b6, 1042b10).

The role of the form in determining the persistence of an organism results from its role as the source of unity. The form, including the organism's vital functions, makes a heap of material constituents into a single organism (*Metaphysics* VII 16). A collection of flesh and bones constitutes a single living organism in so far as it has the form of a man or a horse; the vital functions of the single organism are the final cause of the movements of the different parts. The organism remains in being through changes of matter, as long as it retains its formal, functional properties. Since the structure, behaviour and persistence of the organism must be understood by reference to its form, the form is irreducible to matter (see §9); the organism, defined by its form, must be treated as a subject in its own right, not simply as a heap of matter.

These facts about organisms explain why Aristotle sees a close connection between primary substance and form. Organisms are substances primarily because of their formal properties, not because of their material composition; hence we cannot identify all the basic subjects there are unless we recognize the reality of formal properties and of subjects that are essentially formal.

14 What are substantial forms?

The conclusion that primary substance and form are closely connected, however, explains only why some substances are essentially formal; it does not explain why form itself is substance. To explain this further claim, we need to decide whether Aristotle regards a substantial form as (1) a species form (shared by all members of a given species, for example, the form of man or horse), normally taken to be a universal, or as (2) a particular form, proprietary to (for example) Socrates. (See *Metaphysics* VII 10-16, XII 5, XIII 10, *Generation of Animals* IV 3 for important evidence.)

Some points favouring the 'universal solution' are the following. (1) Aristotle often contrasts the form with the compound of form and matter, and describes particulars as compounds; hence he apparently does not regard particulars as forms. (2) Similarly, he says that a particular differs from a universal in having both form and matter; hence no particular seems to be simply a form. (3) He says the form is what is specified in a definition, but there is no definition of a particular; hence a particular apparently cannot be a form. (4) He says that substance is prior in knowledge to non-substance, but scientific knowledge of particulars is impossible; hence they apparently cannot be substances, and only a universal can be a substance.

In favour of the 'particular solution' it may be argued: (1) a substance must be a subject, whereas all universals are said of subjects; (2) a substance must be a 'this', as opposed to a 'such', and hence, apparently, some sort of particular; (3) Aristotle argues at length that no universal can be a substance.

We might be tempted to conclude that Aristotle's position is inconsistent. His conviction that substance as 'this' and substance as 'what is it' must be the same thing leads him to insist that the successful candidate for substance must satisfy the criteria for being both a this (a subject, and hence a particular) and an essence (a property, and

hence a universal). If one and the same thing cannot satisfy both criteria, then no one thing can satisfy all Aristotle's conditions for being a substance.

We need not draw this conclusion, however. We can maintain that Aristotle consistently favours the universal solution, if we can show: (1) a 'this' need not be a particular; (2) some universals are subjects; (3) a species form is not the sort of universal that cannot be a substance.

We can maintain that he consistently favours the particular solution, if we can show the following. (1) The contrast between form and matter does not imply that they are always mutually exclusive; some forms may be constituted by, or embodied in, particular bits of matter. Sometimes, indeed, Aristotle speaks as though a form is a subject that can persist and perish and can exchange its matter. (2) The sense in which particulars do not allow definition and scientific knowledge does not prevent them from also being, in an appropriate sense, prior in definition and knowledge to universals (*Metaphysics* XIII 10 may attribute the relevant priority to particular substances).

These two solutions are different ways of expressing Aristotle's belief that substances are basic. Both his metaphysics and his natural philosophy express and defend the conviction that natural organisms and their kinds are substances because they are fundamental; they are fundamental because they are irreducible to their constituent matter. It is more difficult to decide whether the individuals or their kinds are more fundamental. Perhaps, indeed, we ought not to decide; different things may be fundamental or irreducible in different ways.

15 Universals, Platonic Forms, mathematics

These disputes partly concern Aristotle's attitude to the reality of universals. One-sided concentration on some of his remarks may encourage a nominalist or conceptualist interpretation (see Nominalism §§1, 2). (1) He rejects Plato's belief (as he understands it) in separated universal Forms (see Plato §§10, 12-16), claiming that only particulars are separable. (2) In *Metaphysics* VII 13-16 he appears to argue that no universal can be a substance. (3) He claims that the universal as object of knowledge is - in a way - identical to the knowledge of it (*On the Soul* 417b23).

Other remarks, however, suggest realism about universals. (4) He claims they are better known by nature; this status seems to belong only to things that really exist. (5) He believes that if there is knowledge, then there must be universals to be objects of it; for our knowledge is about external nature, not about the contents of our own minds.

Aristotle's position is consistent if (1)-(3) are consistent with the realist tendency of (4)-(5). The denial of separation in (1) allows the reality of universals. Similarly, (2) may simply say that no universals are primary substances (which are his main concern in *Metaphysics* VII). And (3) may simply mean (depending on how we take 'in a way') that the mind's conception of the extra-mental universal has some of the features of the universal (as a map has some of the features of the area that it maps). While Aristotle denies that universals can exist without sensible particulars to embody them, he believes they are real properties of these sensible particulars.

He offers a rather similar defence of the reality, without separability, of mathematical objects (*Physics* II 2; *Metaphysics* XIII 3). While agreeing with the Platonist view that there are truths about, for example, numbers or triangles that do not describe the sensible properties of sensible objects, he denies that these truths have to be about independently-existing mathematical objects. He claims that they are truths about certain properties of sensible objects, which we can grasp when we 'take away' (or 'abstract') the irrelevant properties (for example, the fact that this triangular object is made of bronze). Even though there are no separate objects that have simply mathematical properties, there are real mathematical properties of sensible objects.

16 Metaphysics: God

When Aristotle claims that first philosophy is also theology (see §11), he implies that the general discussion of being and substance is the basis for the special discussion of divine substance. (Hence later writers distinguish 'special metaphysics', dealing with God, from 'general metaphysics', dealing with being in general.) The different features of substance explained in *Metaphysics* VII-IX are included in the divine substance of XII. (1) Primary substance is to be identified in some way with form rather than with matter or with the compound of form and matter; divine substance is pure form without matter. (2) Primary substance is in some way numerically one, a

'this' rather than a 'such'; divine substance is completely one and indivisible. (3) Primary substance is in some way actuality rather than potentiality; divine substance is pure actuality with no potentiality. (4) Primary substance is soul rather than body (see §17); divine substance is pure intellect without sense or body.

In each case the properties of primary substance are found in a sensible substance (an animal or a plant) only in so far as they belong to an object that also has other properties; hence primary substance in sensible reality is the form and actuality of an object (a horse, for example) that also has matter and potentiality. In divine substance, however, each feature is found in separation from these other properties; that is why a divine substance lacks matter, multiplicity, parts or potentiality. Aristotle argues that a substance with these pure substantial properties must exist if any sensible substances are to exist; for the existence of potentialities that can be actualized presupposes the existence of an actuality that does not itself include any potentiality (to avoid an infinite regress).

Since this primary type of substance is divine, it is what traditional belief in the Olympian gods was about, what the Presocratics were talking about when they spoke of 'the divine', and what Plato was talking about in speaking of a supreme god. Aristotle mentions the traditional Olympian gods without committing himself to acceptance of the traditional conception of them. He rejects anthropomorphic views of the gods, but he speaks of the divine nature as a kind of mind. He believes that there is something divine about the order and workings of nature, and still more divine in the heavenly substances (*Parts of Animals* I 5). Although he continues to speak of gods in the plural, he also speaks of one divine mind as the ultimate cause of the whole universe; these remarks help to justify the later interpreters who take him to speak of the one God who is the subject of (for example) Aquinas' 'Five Ways' (*Summa Theologiae* 1a q.2 a.3) (see Aquinas, T. §11).

Aristotle's God is the ultimate cause of the physical universe, but not its creator (as Plato's demiurge is), since Aristotle believes the universe is eternal. Nor does Aristotle suggest that God has providence or foreknowledge concerned with future contingent events. But he believes that the physical universe is dependent on God. In *Physics* VIII he argues that the explanation of motion requires recognition of a first cause of motion, and in *Metaphysics* XII this first cause is identified with divine, immaterial, substance. This first mover is itself unmoved; it initiates motion only as an object of love initiates motion by attraction. It is the ultimate final cause of the various movements in the universe.

In treating the divine substance as a god, and hence as a being with a soul and an intellect, Aristotle attributes some mental life to it. But since it would be imperfect if it thought of objects outside itself (because it would not be self-sufficient), it thinks only of its own thinking. This restriction, however, is not as severe as it may seem, since Aristotle believes that the various objects of thought are in some way identical to the mind that thinks them (see §15). In so far as God thinks of his own mind, he thereby also contemplates the order of the universe as a whole; this is the order that the different movements in the universe seek to embody.

Sometimes (as in *Physics* VIII) Aristotle argues for a single first mover. In *Metaphysics* XII, however, he argues that an unmoved mover must be postulated for each of the distinct movements of the heavenly bodies. This astronomical interpretation of his theological doctrine is difficult to reconcile with his belief, reaffirmed in *Metaphysics* XII 10, that in some way the universe is unified by a single first unmoved mover.

17 Soul and body

Aristotle's treatise *On the Soul* is placed among the works on natural philosophy, but should be read with *Metaphysics* VII-IX. In Aristotle's view, disputes about soul and body are simply a special case of the more general disputes about form and matter. He rejects both the Presocratic materialist assumption that the soul is simply non-organic matter, and the Platonic dualist claim that it must be something entirely non-bodily. He argues that soul is substance because it is the form of a natural body, and that the body is the matter informed by the soul. Although the soul is a substance distinct from the non-organic body (the collection of non-organic matter belonging to a living organism; see §13), it is not immaterial (if being immaterial excludes being composed of matter), nor is it independent of some non-organic body or other.

Aristotle assumes that the soul is the primary principle of life, and hence that it distinguishes the living from the non-living. A living organism is nourished, grows and diminishes, through itself - from a causal origin within itself rather than from the action of external agents. A living organism must, therefore, be teleologically ordered, since

(for Aristotle) nutrition and growth cannot be understood without appeal to final causation (see Teleology).

If life must be conceived teleologically, and the soul is the *primary* principle of life, then the soul is form rather than matter. For the primary principle is whatever explains our vital activities; since these are goal-directed activities, their explanation must refer to the goal-directed features of the subject, and so to the form rather than the matter. If the soul is what we live by primarily, it must be the final cause of the body, and so a formal, not a material, aspect of the subject. Soul must, therefore, be substance as form.

Aristotle attributes to the soul the features of substantial form (see §13). (1) It is a substance that is irreducible to a material non-organic body (remote matter); to that extent the soul is incorporeal, and not just some ordinary material stuff. (2) It is the source of unity that makes a heap of material constituents into a single organism. For a collection of flesh and bones constitutes a single living organism in so far as it is teleologically organized; the activities of the single organism are the final cause of the movements of the different parts. Since a single organism has a single final cause, it has a single soul and a single body. (3) The identity and persistence of the soul determine the identity and persistence of the creature that has it. If something has a soul in so far as it has life, then Socrates perishes if and only if his soul does. The truth of this Platonic claim (*Phaedo* 115c-e) does not imply Platonic dualism. (4) The definition of a soul must mention the proximate material subject (the organic body and its parts) whose capacities are actualized in the functions of the organism (*Metaphysics* 1036b28-30). A soul must be non-coincidentally connected to a specific sort of organic body (*On the Soul* 407b20-4).

Some of the puzzles in Aristotle’s doctrine of substantial form arise in his doctrine of soul and body. If, for instance, he recognizes particular substantial forms, then he also recognizes (as the previous paragraph assumes) the individual souls of Socrates and Callias; if, however, he recognizes only one substantial form for each species, then he recognizes only one soul for human beings, another for horses, and so on.

Since the soul is the form of the living body, an account of the different ‘parts’ or ‘capacities’ (or ‘faculties’; *dynameis*) of the soul does not describe the different physiological processes underlying the different activities of a living organism, but describes their formal and goal-directed aspects. Aristotle describes the capacities that distinguish the different types of souls: nutrition (characteristic of plants), perception and appearance (characteristic of animals) and rational thought (characteristic of rational animals) (see Psychē). He describes some of the physiological basis of these psychic capacities in the shorter treatises on natural philosophy, including the *Parva Naturalia*, the *Movement of Animals*, and the *Progression of Animals*.

18 Perception

To define perception, Aristotle returns to his contrast between form and matter. Perception happens in so far as (1) the perceiver becomes like the object (*On the Soul* 417a18); (2) the perceiver that was potentially *F* (for example, white) becomes actually *F* when it perceives the actually *F* object (418a3); (3) the perceiver acquires the form, but not the matter, of the object (424a18-24). These descriptions express a realist view of perception and its objects; Aristotle assumes in (2) that an object is actually white, square, and so on in its own right, before we perceive it.

He is sometimes taken to imply in (1) that perception requires physical similarity; but (3) counts against this interpretation. A sense receives the form without the matter in the way in which a house without matter is in the soul of the architect before the house is built. In the latter case, nothing that looks like a house is in the builder, but features of the house correspond to features of the builder’s design. Similarly, when we hear a tune, our ears do not necessarily sound like the tune, but a state of us systematically corresponds to the tune (as features of a map correspond to features of the area it maps).

A ‘common sense’ perceives common properties of sensible objects, such as size, shape and number, which are all perceived through the perception of motion (*On the Soul* 425a14-20). This is not a sixth sense independent of the other five, but the result of the cooperation of the five senses. Aristotle argues that we can explain our grasp of these common properties without supposing that they are objects of intellect rather than sense (contrast Plato, *Theaetetus* 184-6).

19 Appearance and thought

Appearance (or ‘imagination’; *phantasia*) links perception to goal-directed movement. A lion sees or smells a deer;

it takes pleasure in the prospect of eating the deer, and so wants to catch the deer. To connect perception with pleasure and desire, we need to say how the deer appears to the lion (as prey); this is what Aristotle calls the lion's appearance of the deer (*On the Soul* III 3, 7).

Aristotle denies that this appearance constitutes a belief (*doxa*). He argues that belief requires reason and inference, which non-human animals lack; in his view, they lack any grasp of a universal, and have only appearances and memory of particulars (*Nicomachean Ethics* 1147b4-5). The operations of sense, memory and experience are necessary, but not sufficient, for the grasp of a universal that is expressed in concepts and beliefs (*Posterior Analytics* II 19; *Metaphysics* I 1).

Concepts and beliefs require intellect (*nous*) actualized in 'understanding' or 'thinking' (*noein*; *On the Soul* III 4) (see Nous). Thought differs from perception in so far as it grasps universal essences - for example, what flesh is, as opposed to flesh. Perception does not include grasp of the universal as such; in grasping the universal, we recognize some feature of our experience as a ground for attributing the universal to a particular that we experience.

To explain how the mind is capable of grasping universals when we interact causally with particular perceptible objects, Aristotle distinguishes two aspects of intellect - passive and 'productive' (or 'active' or 'agent') - claiming that these two aspects must combine to produce thought of universals (*On the Soul* III 5). He does not say how productive intellect contributes to our grasp of universals. Later interpreters suggest that productive intellect abstracts the aspects relevant to the universal from the other features of particulars that are combined with them in perception (Aquinas, *Summa Theologiae* 1a q.79 a.3).

Aristotle takes the presence of this productive intellect to be necessary for any thinking at all. Moreover, he believes that productive intellect is capable of existing without a body. He still maintains his belief in the inseparability of soul from body; for since productive intellect is not a type of soul, its separate existence is not the separate existence of a soul.

20 Desire and voluntary action

Perception, appearance and thought are connected to goal-directed movement by means of desire. The appearance of something as desirable is the source of an animal's tendency to pursue one sort of thing rather than another. External objects, however, appear desirable to different agents in different ways. Aristotle distinguishes the appetite (*epithymia*) that animals have from the wish (rational desire; *boulēsis*) that only rational agents have; appetite is for the pleasant and wish is for the good (*On the Soul* 414b2-6, 432b5-7, *Politics* 1253a15-18).

A rational agent's wish differs from appetite in so far as it is guided by deliberation resting on one's conception of one's good. Such a conception extends beyond one's present inclinations both at a particular time and over time. Rational agents are aware of themselves as extending into past and future. Deliberation that is guided by reference to these broader aspects of one's aims and nature results in the rational choice that Aristotle calls 'decision' (*prohairesis*; *Nicomachean Ethics* III 3).

Agents who act on desire and appearance also act voluntarily (*hekousiōs*), in so far as they act on some internal principle (*archē*). While voluntary action is not confined to rational agents, their voluntary action has special significance, because it is an appropriate basis for praise and blame. Since it has an internal principle, it is in our control as rational agents, and therefore we are justly praised and blamed for it. We are held responsible for our actions in so far as they reflect our character and decisions (*Nicomachean Ethics* III 1-5).

Aristotle's defence of his belief that we are appropriately responsible agents does not confront the questions later raised by Epicurus' claim that responsibility is incompatible with the complete causal determination of our actions (see Epicureanism §12). An incompatibilist position is ascribed to Aristotle by Alexander in *On Fate* (see Alexander of Aphrodisias §4.) Aristotle neither explicitly presents an incompatibilist position nor explicitly endorses a compatibilist position of the sort later defended by the Stoics.

A discussion of time, truth and necessity (the 'Sea Battle'; *De Interpretatione* 9) has suggested to some interpreters that Aristotle is an indeterminist. His opponent is a fatalist, who assumes that (1) future-tensed statements about human actions (for example, 'There will be a sea battle tomorrow') were true in the past, and infers that (2) the future is necessarily determined, independently of what we choose. Aristotle certainly rejects (2). If he accepts the

validity of the fatalist's argument, and rejects (1), then he accepts indeterminism.

An alternative reply to the fatalist would be to accept (1) and to deny the validity of the argument. We might argue that the past truth of statements about my actions does not imply that my actions are determined independently of my choices. If on Friday Socrates decides to walk, and he acts on his decision on Friday, then it was true on Thursday that Socrates would walk on Friday, and also true that on Friday he would act on his decision to walk, but it was not true on Thursday that he would walk whether or not he decided to (see Stoicism §21). Probably Aristotle accepts this alternative reply to the fatalist, and hence does not endorse indeterminism.

21 The human good

Aristotle's account of rational agents, choice, deliberation and action is an appropriate starting point for his ethical theory. Ethics is concerned with the praiseworthy and blameworthy actions and states of character of rational agents; that is why it concerns virtues (praiseworthy states) and vices (blameworthy states) (see Aretē).

Aristotle's ethical theory is mostly contained in three treatises: the *Magna Moralia*, the *Eudemian Ethics* and the *Nicomachean Ethics*. The titles of the last two works may reflect a tradition that Eudemus (a member of the Lyceum) and Nicomachus (the son of Aristotle and Herpyllis) edited Aristotle's lectures. The *Magna Moralia* is widely agreed not to have been written by Aristotle; some believe, with good reason, that it contains a student's notes on an early course of lectures by Aristotle. The *Eudemian Ethics* is now widely agreed to be authentic, and generally (not universally) and reasonably taken to be earlier than the *Nicomachean Ethics*. Three books (*Nicomachean Ethics* V-VII = *Eudemian Ethics* IV-VI) are assigned by the manuscripts to both the *Eudemian Ethics* and the *Nicomachean Ethics*.

Aristotle conceives 'ethics' (*Magna Moralia* 1181a24) as a part of political science; he treats the *Nicomachean Ethics* and the *Politics* as parts of a single inquiry (*Nicomachean Ethics* X 9). Ethics seeks to discover the good for an individual and a community (*Nicomachean Ethics* I 2), and so it begins with an examination of happiness, (*eudaimonia*). ('Wellbeing' and 'welfare' are alternative renderings of *eudaimonia* that may avoid some of the misleading associations carried by 'happiness'; see Eudaimonia.) Happiness is the right starting point for an ethical theory because, in Aristotle's view, rational agents necessarily choose and deliberate with a view to their ultimate good, which is happiness; it is the end that we want for its own sake, and for the sake of which we want other things (so that it is the ultimate non-instrumental good). If it is to be an ultimate end, happiness must be complete (or 'final'; *teleion*) and self-sufficient (*Nicomachean Ethics* I 1-5, 7).

To find a more definite account of the nature of this ultimate and complete end, Aristotle argues from the human function (*ergon*), the characteristic activity that is essential to a human being in the same way that a purely nutritive life is essential to a plant and a life guided by sense perception and desire is essential to an animal (*Nicomachean Ethics* I 7). Since a human being is essentially a rational agent, the essential activity of a human being is a life guided by practical reason. The good life for a human being must be good for a being with the essential activity of a human being; hence it must be a good life guided by practical reason, and hence it must be a life in accordance with the virtue (*aretē*) that is needed for achieving one's good. The human good, therefore, is an actualization of the soul in accordance with complete virtue in a complete life. This 'complete virtue' appears to include the various virtues described in the following books of the *Nicomachean Ethics*; this appearance, however, may be challenged by *Nicomachean Ethics* X (see §26).

22 Virtue of character

From the general conception of happiness Aristotle infers the general features of a virtue of character (*ēthikēaretē*; *Nicomachean Ethics* I 13). He agrees with Plato in recognizing both rational and non-rational desires (see Plato §14). One's soul is in a virtuous condition in so far as the non-rational elements cooperate with reason; in this condition human beings fulfil their function well. The argument from the human function does not make it clear what states of a rational agent count as fulfilling the human function. Aristotle seeks to make this clearer, first through his general account of virtue of character, and then through his sketches of the individual virtues.

A virtue of character must be a 'mean' or 'intermediate' state, since it must achieve the appropriate cooperation between rational and non-rational desires; such a state is intermediate between complete indulgence of non-rational desires and complete suppression of them. (Aristotle is not recommending 'moderation' - for

example, a moderate degree of anger or pleasure - in all circumstances.) The demand for cooperation between desires implies that virtue is more than simply control over desires; mere control is ‘continence’ (*enkrateia*) rather than genuine virtue.

The task of moral education, therefore, is to harmonize non-rational desires with practical reason. Virtuous people allow reasonable satisfaction to their appetites; they do not suppress all their fears; they do not disregard all their feelings of pride or shame or resentment (*Nicomachean Ethics* 1126a3-8), or their desire for other people’s good opinion. Aristotle’s sketches of the different virtues show how different non-rational desires can cooperate with practical reason.

23 Virtue, practical reason and incontinence

A virtuous person makes a decision (*prohairesis*) to do the virtuous action for its own sake. The correct decision requires deliberation; the virtue of intellect that ensures good deliberation is prudence (or ‘wisdom’, *phronēsis*; *Nicomachean Ethics* VI 4-5); hence the mean in which a virtue lies must be determined by the sort of reason by which the prudent person would determine it (1107a1-2). Virtue of character is, therefore, inseparable from prudence. Each virtue is subject to the direction of prudence because each virtue aims at what is best, as identified by prudence.

In claiming that prudence involves deliberation, Aristotle also emphasizes the importance of its grasping the relevant features of a particular situation; we need to grasp the right particulars if deliberation is to result in a correct decision about what to do here and now. The right moral choice requires experience of particular situations, since general rules cannot be applied mechanically. Aristotle describes the relevant aspect of prudence as a sort of perception or intuitive understanding of the right aspects of particular situations (*Nicomachean Ethics* VI 8, 11).

These aspects of prudence distinguish the virtuous person from ‘continent’ and ‘incontinent’ people (*Nicomachean Ethics* VII 1-10). Aristotle accepts the reality of incontinent action (*akrasia*), rejecting Socrates’ view that only ignorance of what is better and worse underlies apparent incontinence (see Socrates §6; Akrasia §1). He argues that incontinents make the right decision, but act contrary to it. Their failure to stick to their decision is the result of strong non-rational desires, not simply of cognitive error. Still, Aristotle agrees with Socrates in believing that ignorance is an important component of a correct explanation of incontinence, because no one can act contrary to a correct decision fully accepted at the very moment of incontinent action.

The error of incontinents lies in their failure to harmonize the demands of their appetites with the requirements of virtue; their strong appetites cause them to lose part of the reasoning that formed their decision. When they act, they fail to see clearly how their general principles apply to their present situation. If their failure results from an error in deliberation, it is clear why Aristotle insists that incontinent people lack prudence.

24 Choice, virtue, and pleasure

It is initially puzzling that virtuous people decide to act virtuously for its own sake as a result of deliberation. If they decide on virtuous action for its own sake, then their deliberation causes them to choose it as an end in itself, not simply as a means. Decision and deliberation, however, are not about ends but about ‘the things promoting ends’ (*ta pros ta telē*, often rendered ‘means to ends’). Aristotle’s description of the virtuous person, then, seems to attribute to decision a role that is excluded by his explicit account of decision.

This puzzle is less severe once we recognize that Aristotle regards different sorts of things as ‘promoting’ an end. Sometimes he means (1) that the action is external and purely instrumental to the end; in this way buying food ‘promotes’ eating dinner. Sometimes, however, he means (2) that the action is a part or component of the end, or that performing the action partly constitutes the achieving of the end; in this way eating the main course ‘promotes’ eating dinner. Deliberation about this second sort of ‘promotion’ shows that an action is worth choosing for its own sake, in so far as it partly constitutes our end.

This role for deliberation explains how virtuous people can decide, as a result of deliberation, on virtuous action for its own sake; they choose it as a part of happiness, not as a merely instrumental means. Prudence finds the actions that promote happiness in so far as they are parts of the happy life. Such actions are to be chosen for their own sake, as being their own end; they are not simply instrumental means to some further end. The virtuous

person's decision results from deliberation about the composition of happiness; virtuous people decide on the actions that, by being non-instrumentally good, are components of happiness in their own right.

Aristotle's demand for the virtuous person to decide on the virtuous action for its own sake is connected with two further claims: (1) the virtuous person must take pleasure in virtuous action as such; (2) in doing so, the virtuous person has the pleasantest life. In these claims Aristotle relies on his views about the nature of pleasure and its role in happiness (*Nicomachean Ethics* VII 11-14, X 1-5).

He denies that pleasure is some uniform sensation to which different kinds of pleasant action are connected only causally (in the way that the reading of many boring books on different subjects might induce the same feeling of boredom). Instead he argues that the specific pleasure taken in *x* rather than *y* is internally related to doing *x* rather than *y*, and essentially depends on pursuing *x* for *x*'s own sake. Pleasure is a 'supervenient end' (1174b31-3) resulting from an activity that one pursues as an activity (*praxis* or *energeia*) rather than a mere process or production (*kinēsis* or *poiēsis*).

Aristotle insists, following Plato's *Philebus*, that the value of the pleasure depends on the value of the activity on which the pleasure supervenes (1176a3-29). The virtuous person has the pleasantest life, but the pleasantest life cannot aim exclusively at pleasure.

25 Virtue, friendship and the good of others

The virtuous person's deliberation, identifying the mean in relation to different desires and different situations, is articulated in the different virtues of character (described in *Nicomachean Ethics* III-V). The different virtues are concerned with the regulation of non-rational desires (for example, bravery, temperance, good temper), external goods (for example, magnificence, magnanimity) and social situations (for example, truthfulness, wit). Some concern the good of others to some degree (bravery, good temper, generosity).

Aristotle's Greek for virtue of character, *ēthikēaretē*, is rendered into Latin as 'virtus moralis'. The English rendering 'moral virtue' is defensible, since the virtues of character as a whole display the impartial concern for others that is often ascribed to morality. They are unified by the aim of the virtuous person, who decides on the virtuous action because it is 'fine' (*kalon*). Fine action systematically promotes the good of others; we must aim at it if we are to find the mean that is characteristic of a virtue (1122b6-7).

A second unifying element in the virtues, inseparable from concern for the fine, is their connection to justice (V 1-2). Aristotle takes justice to be multivocal (see §4), and distinguishes general justice from the specific virtue concerned with the prevention and rectification of certain specific types of injustices. General justice is the virtue of character that aims specifically at the common good of a community. Since it is not a different state of character from the other virtues, they must incorporate concern for the common good.

To explain why concern for the good of others, and for a common good, is part of the life that aims at one's own happiness, Aristotle examines friendship (*philia*; *Nicomachean Ethics* VIII-IX). All three of the main types of friendship (for pleasure, for advantage and for the good) seek the good of the other person. Only the best type - friendship for the good between virtuous people - includes *A*'s concern for *B*'s good for *B*'s own sake and because of *B*'s essential character (*Nicomachean Ethics* VIII 1-4).

In the best sort of friendship, the friend is 'another self'; *A* takes the sorts of attitudes to *B* that *A* also takes to *A*. Aristotle infers that friendship is part of a complete and self-sufficient life (IX 9-11). Friendship involves sharing the activities one counts as especially important in one's life, and especially the sharing of reasoning and thinking. Friends cooperate in deliberation, decision and action; and the thoughts and actions of each provide reasons for the future thoughts and actions of the other. The cooperative aspects of friendship more fully realize each person's own capacities as a rational agent, and so promote each person's happiness. Hence the full development of a human being requires concern for the good of others.

26 Two conceptions of happiness?

Although Aristotle emphasizes the other-regarding, social aspects of happiness, he also advocates pure intellectual activity (or 'study', *theōria*) - the contemplation of scientific and philosophical truths, apart from any attempt to apply them to practice (*Nicomachean Ethics* X 6-8). The connection between the human function and human

happiness (see §21) implies that contemplation is a supremely important element in happiness. For contemplation is the highest fulfilment of our nature as rational beings; it is the sort of rational activity that we share with the gods, who are rational beings with no need to apply reason to practice. Aristotle infers that contemplation is the happiest life available to us, in so far as we have the rational intellects we share with gods (see §16).

According to one interpretation, Aristotle actually identifies contemplation with happiness: contemplation is the only non-instrumental good that is part of happiness, and the moral virtues are to be valued - from the point of view of happiness - simply as means to contemplation. If this is Aristotle's view, it is difficult to see how the virtues of character are even the best instrumental means to happiness. Even if some virtuous actions are instrumental means to contemplation, it is difficult to see how the motives demanded of the virtuous person (see §§24-5) are always useful, rather than distracting, for those who aim at contemplation.

Probably, however, Aristotle means that contemplation is the best component of happiness. If we were pure intellects with no other desires and no bodies, contemplation would be the whole of our good. Since, however, we are not in fact merely intellects (*Nicomachean Ethics* 1178b3-7), Aristotle recognizes that the good must be the good of the whole human being. Contemplation is not the complete good for a human being.

If this is Aristotle's view, then contemplation fits the conception of happiness that is upheld in the rest of the *Nicomachean Ethics* and in the other ethical works. The virtues of character, and the actions expressing them, deserve to be chosen for their own sakes as components of happiness. In the virtuous person, they regulate one's choice of other goods, and so they also regulate one's choices about contemplation. The *Politics* may be taken to develop this conception of happiness, since (in book VII) it sets contemplation in the context of a social order regulated by the moral virtues.

27 Politics: ideal states

The *Politics* pursues three connected aims: (1) it completes the discussion of happiness, by showing what kind of political community achieves the human good (mainly books I, II and VII); (2) it sets out moral and political principles that allow us to understand and to criticize the different sorts of actual states and their constitutions (mainly books III and IV); (3) it offers some proposals for improving actual states (mainly books V and VI). The order of the books probably reflects Aristotle's aim of describing an ideal state after examining the strengths and weaknesses of actual states.

An individual's desire for happiness leads eventually to the city. A human being is a 'political animal', because essential human capacities and aims are completely fulfilled only in a political community; hence (given the connection between the human function and the human good) the individual's happiness must involve the good of fellow members of a community. The relevant sort of community is a *polis* ('city' or 'state') - a self-governing community whose proper function (not completely fulfilled by every actual political community) is to aim at the common good of its citizens, who (normally) share in ruling and in being ruled. The city is the all-inclusive community, of which the other communities are parts, since it aims at advantage not merely for some present concern but for the whole of life (*Nicomachean Ethics* 1160a9-30). Since happiness is complete and self-sufficient, the city is a complete and self-sufficient community (*Politics* 1252b28), aiming at a complete and self-sufficient life that includes all the goods needed for a happy life.

The connection between human nature, human good and the political community is most easily understood from Aristotle's account of friendship. Complete friendship, which requires living together and sharing rational discourse and thought, is restricted to individuals with virtuous characters, but this is not the only type of friendship that achieves self-realization in cooperation; a similar defence can be given for the friendship of citizens. Collective deliberation about questions of justice and benefit contributes to the virtuous person's self-realization because it extends the scope of one's practical reason and deliberation beyond one's own life and activities. Since the city is comprehensive, seeking to plan for everything that is needed for the complete good, a rational agent has good reason to want to share in its deliberations.

Since, then, Aristotle believes that political activity contributes in its own right to the human good, he argues against a 'social contract' theory that assigns a restricted instrumental function to the state (safety, or mutual protection, or the safeguarding of what justly belongs to each person; *Politics* III 9). Political life is to be valued for itself, apart from any instrumental benefit; the best city aims at the development of the moral virtues and at the

political participation of all who are capable of them.

In the light of these aims, Aristotle describes the best city. It has to assume favourable external conditions (geographical and economic) to allow the development of political life. Its criteria for citizenship are restricted, since they exclude everyone (including women and manual labourers) whom Aristotle regards as incapable of developing the virtues of character. Within the class of citizens, however, Aristotle is concerned to avoid gross inequality of wealth. and to ensure that everyone shares both in ruling and in being ruled. The institutions of the best state provide the political, social, economic and educational basis for the practice of the moral virtues and for contemplation.

28 Politics: imperfect states

Just as a correct conception of happiness is the basis of the ideal city, various incorrect conceptions of happiness define mistaken aims for different cities. These mistaken aims underlie the different conceptions of justice that are embodied in the constitutions of different cities. Partisans of oligarchy, for instance, take happiness to consist in wealth; they treat the city as a business partnership (*Politics* 1280a25-31). Partisans of democracy take happiness to consist simply in the satisfaction of desire; they assume that if people are equal in the one respect of being free rather than slaves, they are equal altogether, and should have an equal share in ruling (1280a24-5). Neither view is completely mistaken, since neither wealth nor freedom is irrelevant to questions of justice, but each is one-sided.

These one-sided views cause errors about the just distribution of political power or other goods. The proper basis for assigning worth in distribution will be whatever is relevant for the common good, since that is the aim of general justice. Since a correct conception of the common good requires a correct conception of happiness, a correct answer to the question about distribution must appeal to a true conception of happiness.

The criticism of existing constitutions seeks to show both how they fall short of the norms that are met by the ideal state, and how they can be improved. Aristotle wants to describe not only the ideal state, but also the best organization of each political system. In some circumstances, he believes, economic, social, and demographic facts may make (for example) democracy or oligarchy difficult to avoid. Still, an imperfect constitution can be improved, by attention to the aspects of justice, and hence the aspects of happiness, that this constitution tends to ignore. Even when Aristotle may appear to be engaged in empirical political sociology, or to be offering hints for the survival of a particular regime, he is guided by the moral and political principles that he defends in the more theoretical parts of the *Politics*.

29 Rhetoric and poetics

In Aristotle's classification, rhetoric and poetics (*poiētikē*; literally 'productive') count as 'productive' rather than 'practical' disciplines; they are concerned with 'production' (*poiēsis*) - purely instrumental action aiming at some external end - rather than with 'action' (*praxis*) - action that is also an end in itself. Rhetoric is a productive discipline in so far as it aims at persuasion in public speaking, and seeks the arguments, diction, language, metaphor, appeals to emotion and so on, that are most likely to persuade different types of audiences. Hence Aristotle's treatise on rhetoric contains sections on these different topics. Dialectic and logic are useful to a student of rhetoric, even though rhetoric does not aim at the truth; for true or plausible claims tend to be persuasive. *Rhetoric* II deals with another aspect of rhetorical persuasion, by describing the different emotions; the student of rhetoric must know how to arouse emotions in an audience.

Aristotle also takes his moral and political theory to be relevant to rhetoric, for two main reasons. (1) Rhetoric is concerned with the moral and political issues discussed in public assemblies or in courts, and the orator needs to be familiar with the convictions of a given audience. (2) Even more important, the orator should be guided by correct moral and political convictions (without necessarily grasping their philosophical basis). Aristotle does not endorse the conception of oratory as a technique of persuasion that is indifferent to the moral and political aims that it serves. This conception of oratory arouses Plato's criticism in the *Gorgias* (see Plato §7) Aristotle replies to such criticism by arguing that the orator should learn, and should be guided by, correct principles. He sets out some of these in the *Rhetoric*.

Moral and political principles are also relevant to Aristotle's treatment of literary criticism in the *Poetics*. The surviving part of this treatise deals mainly with tragedy. Some of it is similar to the *Rhetoric*, in so far as it

discusses matters of technique and psychology; Aristotle describes the various sorts of plots, characters, and dramatic devices that affect the audience in different ways. He is also concerned, however, about the moral aspects of tragedy; in this he may be responding to the criticisms of tragedy in book X of Plato's *Republic*. He argues that tragedy achieves its appropriate effect when it directs pity, fear, sympathy and revulsion at the appropriate sorts of people and situations; and he examines the plots and characters of various tragedies from this point of view (see Katharsis; Mimēsis).

30 Influence

Some aspects of Aristotle's philosophy have become so familiar that we do not even attribute them to him. When we say that an event was a mere 'coincidence', or that an ignorant person is 'ill-informed', or that someone's behaviour is forming good or bad 'habits', our vocabulary expresses Aristotelian assumptions, transmitted through Latin translations and interpretations.

The explicit influence of Aristotle's philosophical works and theories has been variable. In Hellenistic philosophy, he is not prominently cited or discussed (see Hellenistic philosophy); some have even doubted whether the major Stoics knew his works. From the first century BC, however, the study of Aristotle revived. This revival produced philosophers defending an Aristotelian position, often incorporating Stoic or Platonist elements, but sometimes sharpening contrasts between Aristotle and the Hellenistic schools (see Alexander of Aphrodisias; Peripatetics). These Aristotelians began a long series of Greek commentaries (lasting until the sixth century AD). Many of the later commentators were Neoplatonists; some of whom tried to reconcile Aristotelian with Platonic doctrines (see Aristotle Commentators; Platonism, Early and Middle §§8-9; Neoplatonism §1; Porphyry §2).

Between the sixth and the thirteenth centuries, most of Aristotle's works were unavailable in western Europe, although he was still studied in the Byzantine empire and the Islamic world (see Aristotelianism in Islamic philosophy). Two leading figures in the revival of Aristotelian studies and of Aristotelian philosophy in medieval Europe were the translator William of Moerbeke and Thomas Aquinas (see Aristotelianism, medieval). Aquinas' attempt to combine Aristotelian philosophy with orthodox Christian theology was at first rejected by ecclesiastical authority, but then came to be accepted (see Aquinas, T.).

The 'scholastic' philosophy of Aquinas and his successors is often opposed, but often presupposed, by Descartes, Locke, Hobbes and many of their successors, who often do not distinguish it from Aristotle's own philosophy. The reader who compares their representation of the scholastic position with Aristotle's own works (or with Aquinas) will often be surprised by the sharp differences between Aristotle's (and Aquinas') own positions and the positions that are attributed to him by the seventeenth-century philosophers who reject his authority (see Aristotelianism in the 17th century).

Modern historical study of Aristotle begins in the early nineteenth century. It has led to philosophical reassessment, and his works have once again become a source of philosophical insight and argument. Many of the themes of Aristotelian philosophy - the nature of substance, the relation of form to matter, the relation of mind to body, the nature of human action, the role of virtues and actions in morality - have reappeared as issues in philosophical debates, and Aristotle's contributions to these debates have influenced the course of philosophical discussion.

In some ways, Aristotle has suffered from his success. At different times he has been regarded as the indisputable authority in astronomy, biology, logic and ethics; hence he has represented the traditional position against which reformers have revolted. If he is regarded neither as the indisputable authority nor as a repository of antiquated and discarded doctrines, his permanent philosophical value can be more justly appreciated.

See also: Archē; Aristotelianism, Renaissance; Averroism; Being; Change; Dualism; Friendship; Metaphysics; Pneuma; Prudence; Strato; Teleological ethics; Tense and temporal logic; Virtue ethics; Virtues and vices

T.H. IRWIN

List of works

The works of Aristotle are usually cited by conventional Latin titles, or by English translations of these titles (often mere Anglicizations rather than proper translations). This list omits: works preserved in the Aristotelian corpus, but now generally agreed to be spurious; lost works; and the *Constitution of Athens* (probably not by Aristotle

himself; discovered after the standard arrangement of Aristotle's works was established).Neither the absolute nor the relative dates of individual treatises can be established (see §2). The list below follows the thematic order outlined in the entry.Recommended editions (Greek text with commentary) and translations of individual works are listed below. The standard text of most treatises appears in the Oxford Classical Texts (Oxford: Oxford University Press, various editors and dates), or, when these are lacking, in the Teubner texts (Leipzig: Teubner, various editors and dates). The Greek text, with facing English translation (not always reliable) appears in the Loeb Classical Library (Cambridge, MA: Harvard University Press and London: Heinemann, various editors and dates).

Aristotle (*c*. mid 4th century BC) *Aristotelis Opera*, ed. I. Bekker, Berlin: Reimer, 1831-70, 5 vols.(The first modern edition of the Greek text and the source of the page and line references normally used.)

Aristotle (*c*. mid 4th century BC) *The Works of Aristotle*, ed. W.D. Ross and J.A. Smith, Oxford: Oxford University Press, 1908-54, 12 vols.(The Oxford Translation.)

Aristotle (*c*. mid 4th century BC) *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984, 2 vols.(The standard English translation of the whole corpus; contains the revised Oxford Translation.)

Aristotle (*c*. mid 4th century BC) *The Basic Works of Aristotle*, ed. R. McKeon, New York: Random House, 1941. (A selection from the Oxford Translation.)

Aristotle (*c*. mid 4th century BC) *A New Aristotle Reader*, ed. J.L. Ackrill, Oxford: Oxford University Press, 1984. (Selections from existing translations.)

Aristotle (*c*. mid 4th century BC) *Aristotle: Selections*, trans. G. Fine and T. Irwin, Indianapolis, IN: Hackett, 1995. (New translations with notes.).

1 Logic

Aristotle (*c*. mid 4th century BC) *Categories (Categoriae)*, trans. J.L. Ackrill, Clarendon Aristotle Series, Oxford: Oxford University Press, 1963.(Outline of the theory of categories.)

Aristotle (*c*. mid 4th century BC) *On Interpretation (De Interpretatione)*, trans. J.L. Ackrill, Clarendon Aristotle Series, Oxford: Oxford University Press, 1963.(Thought, language and logic.)

Aristotle (*c*. mid 4th century BC) *Prior Analytics (Analytica Priora)*, trans. R. Smith, Indianapolis: Hackett, 1989. (Deductive logic: the theory of the syllogism.)

Aristotle (*c*. mid 4th century BC) *Posterior Analytics (Analytica Posteriora)*, trans. J. Barnes, Clarendon Aristotle Series, Oxford: Oxford University Press, 2nd edn, 1993.(The theory of demonstration: the structure of knowledge.)

Aristotle (*c*. mid 4th century BC) *Topics (Topica)* trans. A.W. Pickard-Cambridge in *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984, 2 vols, 167-314.(The theory, strategy and tactics of dialectical argument.)

2 Natural philosophy

Aristotle (*c*. mid 4th century BC) *Physics (Physica)*, ed. W.D. Ross, in *The Works of Aristotle*, Oxford: Oxford University Press, 1949; Books I-II, trans. W. Charlton, Clarendon Aristotle Series, Oxford: Oxford University Press, 1970; Books III-IV trans. E.L. Hussey, Clarendon Aristotle Series, Oxford: Oxford University Press, 1983.(General principles of natural philosophy - form, matter, explanation - and their application to motion, infinity, place and time.)

Aristotle (*c*. mid 4th century BC) *On the Heavens (De caelo)*, Books I-II trans. S. Leggett, Warminster: Aris & Phillips, 1995.(Cosmology: application of Aristotle's theory of motion to the four elements and their interaction in the universe.)

Aristotle (*c*. mid 4th century BC) *Generation and Corruption (De generatione et corruptione)*, ed. H.H. Joachim, Oxford: Oxford University Press, 1922; trans. C.J.F. Williams, Clarendon Aristotle Series, Oxford: Oxford University Press, 1982.(Account of change in natural substances; alteration, growth, coming-to-be and perishing; matter and the four elements.)

Aristotle (*c*. mid 4th century BC) *Meteorology (Meteorologica)*, trans. H.D.P. Lee, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1952.(Collection of observations on winds, tides and other aspects of inanimate nature.)

Aristotle (*c*. mid 4th century BC) *History of Animals (Historia animalium)* trans. A.L.Peck, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1965-91, 3 vols.(Collection of observations on different kinds of animals and their behaviour, providing a basis for Aristotle's biological

explanations.)
Aristotle (*c*. mid 4th century BC) *Parts of Animals (De partibus animalium)*, Book I, trans. D.M. Balme, Clarendon Aristotle Series, Oxford: Oxford University Press, 1972.(Introduction to the study of animals; explanation of physiology, organs and behaviour, in light of Aristotle's explanatory scheme.)
Aristotle (*c*. mid 4th century BC) *Generation of Animals (De generatione animalium)*, trans. A.L.Peck, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1942.(Application of Aristotle's explanatory scheme to reproduction and heredity.)
Aristotle (*c*. mid 4th century BC) *On the Movement of Animals (De motu animalium)*, ed. and trans. M.C. Nussbaum. Princeton, NJ: Princeton University Press, 1978.(Physiology and psychological explanation of animal movement.)
Aristotle (*c*. mid 4th century BC) *On the Progression of Animals (De incessu animalium)*, trans. E.S. Foster, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1937.(Physiology of animal movement.)

3 Metaphysics

Aristotle (*c*. mid 4th century BC) *Metaphysics (Metaphysica)*, ed. W.D. Ross, Oxford: Oxford University Press, 1924, 2 vols; Books IV-VI, trans. C.A. Kirwan, Clarendon Aristotle Series, Oxford: Oxford University Press, 1971; Books VII-VIII trans. D. Bostock, Clarendon Aristotle Series, Oxford: Oxford University Press, 1994; Books XIII-XIV trans. J. Annas, Clarendon Aristotle Series, Oxford: Oxford University Press, 1976. (Collection of treatises centred on the 'science of being', especially substance, form and matter, potentiality and actuality, culminating in theology.)

4 Psychology

Aristotle (*c*. mid 4th century BC) *On the Soul (De anima)*, ed. R.D. Hicks, Cambridge: Cambridge University Press, 1907; Books II-III trans. D.W. Hamlyn, Clarendon Aristotle Series, Oxford: Oxford University Press, 1968.(Application of theory of form and matter to questions about soul and body; perceptio, thought and action.)
Aristotle (*c*. mid 4th century BC) *Short Natural Treatises (Parva naturalia)*, ed. W.D. Ross, Oxford: Oxford University Press, 1955.(Short essays, physiological and psychological, on themes connected with *On the Soul*.)

5 Ethics

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics (Ethica Nicomachea)* commentary by J.A. Stewart, Oxford: Oxford University Press, 1892, 2 vols; trans. T.H. Irwin, Indianapolis, IN: Hackett, 1985.(Normally regarded as Aristotle's most important contribution to moral philosophy.)
Aristotle (*c*. mid 4th century BC) *Eudemian Ethics (Ethica Eudemia)*, Books I-II trans. M.J. Woods, Clarendon Aristotle Series, Oxford: Oxford University Press, 1982.(Similar in content to the *Nicomachean Ethics*, though different in important details. Normally regarded as an earlier version of Aristotle's ethical theory.)
Aristotle (*c*. mid 4th century BC) *Great Ethics (Magna moralia)*, trans. G. Stock, in *The Complete Works of Aristotle*, revised Oxford Translation, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984, 2 vols, 1868-1991.(The standard English translation of the whole corpus; contains the revised Oxford Translation.)

6 Politics

Aristotle (*c*. mid 4th century BC) *Politics (Politica)*, ed. W.L. Newman, Oxford: Oxford University Press, 1887-1902, 4 vols; Books I-II trans. T.J. Saunders, Clarendon Aristotle Series, Oxford: Oxford University Press, 1995.(Aristotle's major work in political theory, including ethics, history and sociology; examines the imperfections of actual states and proposes an ideal state.)

7 Rhetoric and Poetics

Aristotle (*c*. mid 4th century BC) *Rhetoric (Rhetorica)*, ed. E.M. Cope, Cambridge: Cambridge University Press, 1877, 3 vols; trans. G. Kennedy, *Aristotle on Rhetoric: A Theory of Civic Discourse*, New York: Oxford University Press.(The theory and practice of public speaking, based on Aristotle's logic, dialectic, psychology, ethics and political theory.)
Aristotle (*c*. mid 4th century BC) *Poetics (De arte poetica)*, ed. D.W. Lucas, Oxford: Oxford University Press, 1968; trans. R. Janko, Indianapolis, IN: Hackett, 1987.(Analysis of poetry (in the surviving part of the treatise,

tragic drama) from linguistic, stylistic, psychological, and moral points of view.)

References and further reading

Ackrill, J.L. (1981) *Aristotle the Philosopher*, Oxford: Oxford University Press.(This and Barnes (1982) are the best short introductions.)

Barnes, J. (1982) *Aristotle*, Oxford: Oxford University Press.(Along with Ackrill (1981), one of the best short introductions to Aristotle.)

Barnes, J. (ed.) (1994) *The Cambridge Companion to Aristotle*, Cambridge: Cambridge University Press.(Good bibliography.)

Barnes, J., Schofield, M. and Sorabji, R. (eds) (1975-9) *Articles on Aristotle*, London: Duckworth, 4 vols.(Good collections of essays, including translated selections from non-English works; full bibliographies.)

Bonitz, H. (1870) *Index Aristotelicus*, Berlin: Reimer.(Outstandingly useful guide to Aristotle's vocabulary.)

Broadie, S.W. (1991) *Ethics with Aristotle*, Oxford: Oxford University Press.(This and Hardie (1980) are the best general guides to the *Ethics*; Hardie is more accessible to a beginner.)

Charles, D. (1984) *Aristotle's Philosophy of Action*, London: Duckworth.(Detailed and sophisticated discussion, focusing on *Physics* and *Ethics*.)

Fine, G. (1993) *On Ideas*, Oxford: Oxford University Press.(On Aristotle's criticism of Plato.)

Furley, D.J. and Nehamas, A. (eds) (1994) *Philosophical Essays on Aristotle's Rhetoric*, Princeton, NJ: Princeton University Press.(Collection of essays; this and Rorty (1995) provide a survey of recent work on the *Rhetoric*.)

Furth, M. (1988) *Substance, Form, and Psyche*, Cambridge: Cambridge University Press.(Furth, Lewis (1991), Loux (1991) and Witt (1989) provide a guide to debates about the *Metaphysics*, especially Books VII-IX.)

Gotthelf, A. and Lennox, J. (eds) (1987) *Philosophical issues in Aristotle's Biology*, Cambridge: Cambridge University Press, 1987.(Collection of essays.)

Hardie, W.F.R. (1980) *Aristotle's Ethical Theory*, Oxford: Oxford University Press, 2nd edn.(Alongside Broadie (1991), one of the best general guides to the *Ethics*.)

Irwin, T.H. (1988) *Aristotle's First Principles*, Oxford: Oxford University Press.(This and Lear (1988) explore themes connecting several aspects of Aristotle's philosophy.)

Irwin, T.H. (ed.) (1995) *Classical Philosophy*, vols 5-7, New York: Garland.(Collection of essays on Aristotle.)

Jaeger, W. (1923) *Aristoteles: Grundlegung einer Geschichte seiner Entwicklung*, Berlin: Weidmann; trans. R. Robinson, *Aristotle: Fundamentals of the History of his Development*, Oxford: Oxford University Press, 2nd edn, 1948.(English translation; formerly influential account of Aristotle's intellectual development.)

Judson, L. (ed.) (1991) *Aristotle's Physics*, Oxford: Oxford University Press. (Collection of essays.)

Keyt, D. and Miller, F.D. (eds) (1991) *A Companion to Aristotle's Politics*. Oxford: Blackwell.(Collection of essays.)

Kraut, R. (1989) *Aristotle on the Human Good*, Princeton, NJ: Princeton University Press.(Clear and full discussion of happiness in *Nicomachean Ethics* I, X.)

Lear, J. (1988) *Aristotle: The Desire to Understand*, Cambridge: Cambridge University Press.(Explores themes connecting several aspects of Aristotle's philosophy; see also Irwin (1988).)

Lewis, F.A. (1991) *Substance and Predication in Aristotle*, Cambridge: Cambridge University Press.(Provides the most detailed guide to debates about the *Metaphysics*, especially books VII-IX; see also Furth (1988), Loux (1991) and Witt (1989).)

Loux, M.J. (1991) *Primary Ousia*, Ithaca, NY: Cornell University Press.(Provides a guide to debates about the *Metaphysics*, especially Books VII-IX; see also Furth (1988), Lewis (1991) and Witt (1989).)

Miller, F.D. (1995) *Nature, Justice, and Rights in Aristotle's Politics*, Oxford: Oxford University Press. (Discussion of major issues in Aristotle's political theory.)

Nussbaum, M.C. and Rorty, A.O. (eds) (1992) *Essays on Aristotle's De Anima*, Oxford: Oxford University Press.(Collection of essays, mostly recent.)

Organ, T.W. (1949) *An Index to Aristotle*, Princeton, NJ: Princeton University Press. (Moderately useful attempt to construct an index to the Oxford translation.)

Owen, G.E.L. (1986) *Logic, Science, and Dialectic*, Ithaca, NY: Cornell University Press.(Includes several seminal papers on Aristotle's logic and metaphysics.)

Rorty, A.O. (ed.) (1980) *Essays on Aristotle's Ethics*, Berkeley, CA: University of California Press.(Collection of essays, covering the main themes of the *Ethics* in order.)

Rorty, A.O. (ed.) (1992) *Essays on Aristotle's Poetics*, Princeton, NJ: Princeton University Press.(Collection of essays.)

Rorty, A.O. (ed.) (1995) *Essays on Aristotle's Rhetoric*, Berkeley, CA: University of California Press.(Alongside Furley and Nehamas (1994), provides a survey of recent work on the *Rhetoric*.)

Ross, W.D. (1923) *Aristotle*, London: Methuen.(Extremely useful summary of all of Aristotle's works.)

Sorabji, R. (1980) *Necessity, Cause, and Blame*, London: Duckworth.(Covers central areas in natural philosophy, metaphysics, and ethics in accessible style.)

Waterlow, S. (1982) *Nature, Change, and Agency*, Oxford: Oxford University Press.(Thought-provoking discussion of themes in *Physics*.)

Witt, C. (1989) *Substance and Essence in Aristotle*, Ithaca, NY: Cornell University Press.(Provides the most accessible guide to debates about the *Metaphysics*, especially Books VII-IX; see also Furth (1988), Lewis (1991) and Loux (1991).)

Aristotle Commentators

Aristotle's school treatises were given renewed prominence by Andronicus of Rhodes in the first century BC*, and from then on numerous commentaries were written on them. The main modern edition runs to 15,000 pages. They are not just commentaries, but represent the thought and classroom teaching on philosophy quite generally first of the Peripatetic (that is, Aristotelian) school, and then of the Neoplatonists between* AD *200 and 600, with further activity from the ninth century in the Islamic world and from the eleventh in the Byzantine.*

The commentary movement followed and is commonly thought to have been inspired by the work of Andronicus of Rhodes, who, perhaps around 60 BC, began a massive study of Aristotle's writings (see Aristotle). Of the extant commentaries, the earliest come from Peripatetics of the first half of the second century AD, Adrastus and Aspasius. The early commentaries culminate in the work of the greatest Peripatetic elaborator of Aristotle, Alexander of Aphrodisias (second to third century; see Peripatetics). But the later commentaries by Themistius (fourth century), though presented as mere paraphrases, and though informed by certain Neoplatonist developments, are similar in content to those of Alexander.

Outside the Peripatetic school, the chief interest in the first century BC and the first two centuries AD focused on Aristotle's *Categories* (see Categories §1; Aristotle §7). Under the influence of the Neoplatonist Porphyry (third century), who rejected the complaint of his teacher Plotinus Aristotle's *Categories* ignores the Platonic Forms, Aristotle's logic and a wide selection of his other texts became a standard prerequisite for studying Plato in the Neoplatonist schools, Neoplatonism being by then the dominant philosophy (see Neoplatonism).

Porphyry's insistence on the harmony of Plato and Aristotle inaugurated a Neoplatonist tradition. His pupil Iamblichus (third to fourth century) and Iamblichus' pupil Dexippus further defended the compatibility of Aristotle's *Categories* with Plato's theory of Forms, and Iamblichus is even said to have denied that Aristotle contradicted the theory of Forms at all. In fifth-century Alexandria, first Hierocles and then Ammonius, son of Hermeas represented Plato and Aristotle as agreeing that god was the artificer of a beginningless universe. This harmonization, though historically untenable, proved philosophically fruitful, producing a new amalgamation of the ideas of Plato and Aristotle.

After Iamblichus, the study of Aristotle as a propaedeutic to Plato acquired a new significance, because the study of Plato himself was seen as leading to the Neoplatonist ideal of ascent to god. In Iamblichus' Platonic curriculum, Plato's *Timaeus* and *Parmenides* were put last, and seen as theological treatises describing the supreme divinities. From Ammonius onwards formalized introductions were prefixed to commentaries on Aristotle's *Categories*, the first Aristotelian work in the curriculum. The introductions covered ten points which, we are told, had been made standard by Ammonius' teacher Proclus (fifth century), one of them being that the eventual aim of studying Aristotle is ascent to god through the theological works of Plato. Even before the *Categories*, the curriculum included Porphyry's introduction to the *Categories*, known as the *Isagōgē* or *Quinque Voces*. Commentaries on that work were prefaced by a description of the nature of philosophy.

The commentaries of the Alexandrian Neoplatonist school after Ammonius are sometimes divided into portions which would have taken an hour to deliver as lectures. Often there is a double discussion, a *protheōria* or treatment of Aristotle's doctrine in a passage, followed by an exegesis of the exact wording of the text (*exēgēsis tēs lexeōs*).

In Alexandria, the Christian Platonist Philoponus (490-570s) worked out a complete alternative to the Aristotelian physics which the Neoplatonists had accepted, and thus went on to influence the development of Islamic and Western physics. Many of his unorthodox ideas were included in his attack *Against Proclus*, published in 529, but he also began to introduce them into his commentaries on Aristotle. Meanwhile Simplicius, a pious devotee of Neoplatonist religion, had left Athens for Persia as one of the seven displaced Athenian philosophers after the Christian emperor Justinian closed the Athenian Neoplatonist school in 529. He wrote bitterly about Philoponus' Christianity and about the unorthodox interpolations in his commentaries. He emphasizes the harmony of virtually all Greek thought, partly in order to rebut Christian charges of contradictions. It has been suggested that Simplicius finished up in Ḥarrān, just within the borders of modern Turkey, and wrote his Aristotelian commentaries there after 532. It was from Ḥarrān that Thābit ibn Qurra later went, to found the Platonizing school in Baghdad. This

school became the driving force behind the translation into Arabic of Aristotle and his commentators in the ninth century, which was to inspire new philosophizing in the Islamic world.

Meanwhile commentaries were made available in Latin by Boethius (died *c.*525), but only on two logical works of Aristotle and on Porphyry's introduction (*Isagōgē*) to that logic. The Alexandrian School may have survived until the Arab capture of the city in 642, and a tradition continued in Constantinople. There in the twelfth century the princess Anna Comnena organized a circle that included the commentators Eustratius and Michael of Ephesus, whose commentaries were completed in 1138 or later. It may have been from Michael's workshops that James of Venice was able, around 1130, to collect Greek commentaries for translation into Latin. In the same century, Gerard of Cremona was translating Aristotelian commentaries into Latin from the Arabic versions. During the next century, the process of transmission from both languages to the Latin-speaking world turned from a trickle into a flood, and Thomas Aquinas was a beneficiary of this development. He was thus responding not just to Aristotle, but to Aristotle transformed by the ancient commentators.

The commentaries embed fragments from all 1,150 years of ancient philosophy from 550 BC to AD 600, notably those of the Presocratics. And many ideas which were previously thought to date from later times can actually be traced back to the commentators. In dynamics, the idea of an impetus, which in its medieval context has been hailed as a scientific revolution, can be seen to have travelled by an Arabic route from the sixth-century commentator Philoponus. Galileo in his early works mentions Philoponus more often than he mentions Plato. And Brentano in the nineteenth century got from the commentary tradition, and not from Aristotle himself, his idea that all activity of the mind is directed towards intentional objects.

See also: Aristotelianism in Islamic philosophy; Aristotelianism, medieval; Damascius; Neoplatonism in Islamic philosophy; Translators; Greek philosophy: impact on islamic philosophy

RICHARD SORABJI

References and further reading

Boethius (*c.* AD 510) Commentary on Aristotle's *Categories*, in J.-P. Migne (ed.) *Patrologia Latina*, vol. 64, Paris, 1860.(The primary surviving ancient Latin commentator.)

Boethius (*c.* AD 510s) Two Commentaries on Aristotle's *De interpretatione*, ed. R. Meiser, Leipzig: Teubner, 1877-80, 2 vols.

Diels, H. (ed.) (1882-1909) *Commentaria in Aristotelem Graeca*, Berlin: Reimer, 23 vols, with 3 supplementary vols.(Includes text of most of the ancient Greek commentaries. The commentators included are, in chronological order: Adrastus, Aspasius, Alexander of Aphrodisias, Porphyry, Dexippus, Themistius, Syrianus, Ammonius, Priscian of Lydia, Asclepius, Philoponus, Simplicius, Olympiodorus, Elias, David, Stephanus, Eustratius, Michael of Ephesus, Sophonias.)

Sorabji, R. (ed.) (1987-) *Ancient Commentators on Aristotle*, London: Duckworth and Ithaca, NY: Cornell University Press.(English translations, with introduction and notes, to be supplemented by a sourcebook.)

Sorabji, R. (ed.) (1990) *Aristotle Transformed: The Ancient Commentators and their Influence*, London: Duckworth and Ithaca, NY: Cornell University Press.(A comprehensive survey of the present state of knowledge, with extensive bibliography.)

Arithmetic, philosophical issues in

The philosophy of arithmetic gains its special character from issues arising out of the status of the principle of mathematical induction. Indeed, it is just at the point where proof by induction enters that arithmetic stops being trivial. The propositions of elementary arithmetic - quantifier-free sentences such as ' $7 + 5 = 12$*' - can be decided mechanically: once we know the rules for calculating, it is hard to see what mathematical interest can remain. As soon as we allow sentences with one universal quantifier, however - sentences of the form '* $(\forall x)f(x) = 0$*' - we have no decision procedure either in principle or in practice, and can state some of the most profound and difficult problems in mathematics. (Goldbach's conjecture that every even number greater than 2 is the sum of two primes, formulated in 1742 and still unsolved, is of this type.)*

It seems natural to regard as part of what we mean by natural numbers that they should obey the principle of induction. But this exhibits a form of circularity known as 'impredicativity': the statement of the principle involves quantification over properties of numbers, but to understand this quantification we must assume a prior grasp of the number concept, which it was our intention to define. It is nowadays a commonplace to draw a distinction between impredicative definitions, which are illegitimate, and impredicative specifications, which are not. The conclusion we should draw in this case is that the principle of induction on its own does not provide a non-circular route to an understanding of the natural number concept. We therefore need an independent argument. Four broad strategies have been attempted, which we shall consider in turn.

1 Formalism

Although the problem of accounting for both the objects apparently spoken of and the knowledge we appear to have about them in a way that explains both the apparent necessity and the apparent applicability of this knowledge is one which arises for all areas of mathematics, arithmetic has features which make some answers to the question seem especially plausible. In the language we use to express arithmetic every number has a canonical name (called a 'numeral'). This fact has made formalism a persistently attractive option in the case of arithmetic: since every number has a numeral to represent it, we can restrict our attention to the symbols and let numbers themselves drop out of the account entirely. It is not clear that this move helps very much with the epistemological part of the problem, since if the account is to license the whole of arithmetic, numerals will have to be abstract entities our knowledge about which demands explanation. But in any case formalism as just described is untenable, since it leaves the applicability of arithmetic wholly unexplained: we cannot rest content with the claim that arithmetic is just a game with symbols and not demand an account of why it should be *this* game that we play rather than any other.

If the fault in formalism is that it does not allow arithmetical statements to be meaningful, it might be thought an improvement to regard the terms occurring in our formal theory as being given their meaning by the role they play in the axioms of the theory. This variant of formalism has been advanced as an account not just of Peano's axiomatization of arithmetic but of the axiomatic method generally (most famously by Hilbert in his correspondence with Frege; see Frege 1980). It is far from clear, however, that the terms of a formal language can be invested with meaning this easily. In the 1920s and 1930s the members of the Vienna Circle (most notably their 'logician-in-chief', Rudolf Carnap) regarded arithmetic as true only 'by convention'. But if we were to adopt an inconsistent convention our language would be incapable of saying anything about the world. The general point, then, is that formalism cannot explain applicability without assuming consistency, and the most obvious sort of proof of the consistency of arithmetic appeals to our prior grasp of an infinite model, thus facing again the very difficulty formalism was intended to solve.

In his later work Hilbert had the idea of partitioning arithmetic into a real part, for which he gave a finitistic justification (see §3 below), and an ideal part, to be treated formally and justified instrumentally as a method of providing shorter or more comprehensible proofs of results in the real part. The consistency problem arises once more, of course, but Hilbert hoped to be able to solve it without supplying a model, by a combinatorial analysis of the syntax of the formal language. Such an analysis could, he thought, be carried out using the conceptual resources of the finitistically justified real part. Gödel's incompleteness theorems coupled with Turing's analysis of the notion of computability killed Hilbert's programme in the grand form originally envisaged. However,

mathematicians hardly ever exploit the full strength of the classical system and not everyone has given up hope of an instrumental justification on Hilbertian lines of the parts they actually use (see Gödel's theorems §6; Hilbert's programme and formalism §4; Turing, A.M.).

2 Empiricism

If objections to formalism lead us to believe that arithmetical statements really are *true*, we might try Mill's strategy, which treats numbers as empirical and arithmetical statements as empirical generalizations (see Mill, J.S. §3). This strategy has not been popular - our conviction that the hardness of the logical 'must' applies in arithmetic too seems difficult to dispel - but it is quite difficult to refute decisively. Frege objected to it on the grounds that if numbers were empirical, only what is empirical could be counted, whereas number, in Locke's phrase, 'applies itself to men, angels, actions, thoughts - everything that either doth exist or can be imagined' (1689: II.xvi.1). This is not, however, a decisive objection, since the empiricist may be prepared to accept that 'everything that doth exist or can be imagined' is ultimately explicable in empirical terms. A better objection is that empiricism seems ill-equipped to ground our knowledge about large numbers (for example, those larger than the number of atoms in the observable universe). Even though such numbers are practically useless (which is precisely why empirical evidence provides no confirmation for assertions about them), the revision to arithmetic which would be needed if we did not admit them is radical and the resulting system would be awkward to use. The most persuasive objection of all, however, is the difficulty the arithmetical empiricist has in justifying the principle of mathematical induction. It is very hard to see what would count as empirical evidence either for or against this principle.

3 Intuitionism

A third broad strategy is to appeal to intuition to explain how arithmetical knowledge is possible. This approach is due to Kant, whose account of arithmetic parallels quite closely his account of geometry. A geometric proof typically begins with a construction (of a triangle, for instance), on which various operations are then performed. Since I construct the triangle in intuition, Kant held that it will be subject to just the spatial structure which sensibility imposes on all my experience. As a consequence, the geometric theorem I prove by this means is applicable a priori to experience. Similarly, in order to prove that $7+5=12$ I must first construct the concept 'seven' in intuition and then perform on my construction various operations. What is less clear in the arithmetical case than in the geometric one is how the structure imposed by sensibility - what Kant calls the 'form' of intuition - constrains the construction. Kant admits that we do not depend on sensibility for the *concepts* of arithmetic, which are therefore purely intellectual in contrast to those of geometry, but he holds that constructing the concepts depends on counting, which is a process apparently subject to the forms of space and time. It was common in the nineteenth century to read Kant as advocating a neat parallelism - that arithmetic is the science of time just as geometry is that of space - but this is at best a simplification. It is in any case far from clear that counting depends on time. Notice, though, that if we abandon this dependence we thereby give up just what Kant saw as the principal advantage of his account, namely its explanation of the applicability of arithmetic. The moral is quite general: the appeal to intuition beyond the limits of logic, which is characteristic of Platonism, does nothing to explain the applicability of mathematics when it is removed from the Kantian framework (see Kant, I. §5).

If I assert an arithmetical proposition, I am, according to Kant, reporting the result of a mental construction which I have performed or could in principle perform. But as Frege said, 'another man's idea is another idea'. Kant's account makes '$7+5=12$' express a different statement on each person's lips. It is therefore unavoidably solipsistic.

A further difficulty is caused by Kant's insistence that sensibility is passive and that the intuitions it supplies are immediate representations: it is hard to believe that I can have in my mind an *immediate* representation of a very large number, such as 10^{100}. The notion (denied by Kant) that humans are capable not only of passive intuition through the faculty of sensibility but also of creative - Kant calls them 'intellectual' - intuitions is a theme of German romanticism in the nineteenth century. In application to arithmetic it was used by Brouwer to provide an account on which our experience of time unfolding gives us an intuition of succession - what Brouwer calls the bare two-oneness - which we can repeat for ourselves at will as creative subjects. On this view the law of the excluded middle is not justified: if numbers are mind-dependent and have no existence outside of our constructions, then, for example, there is no reason to believe that Goldbach's conjecture (that every even number greater than 2 is the sum of two primes) must be either true or false in advance of constructing either a proof or a

counterexample.

Hilbert's justification for the real part of his real/ideal partition mentioned in §1 above proceeds on similar lines to Brouwer's, although the intuitions on which numbers are based are for Hilbert those of finite arrangements of concrete objects. For this reason Hilbert's account seems less likely than Brouwer's to descend into solipsism. But the principal difference between their conceptions is that Brouwer regarded arithmetical statements and proofs as mental constructions available within arithmetic in the same way as numbers themselves. For Hilbert, on the other hand, any application of arithmetic to proofs is a matter for the metatheory. The result is that Hilbert's finitism can justify directly only a fragment of arithmetic with a limited induction principle (so-called primitive recursive arithmetic), whereas Brouwer's intuitionism has a much stronger induction principle available to it.

There is a persistent worry, however, as to whether either Hilbert or Brouwer overcame the problem Kant's account faced with large numbers. Both were ready to accept arguments on the basis of what humans can do in principle: these arguments seem to turn on a distinction between finite and infinite which is not available from their perspective without independent justification. If so, their position collapses into 'strict finitism', the view that we cannot have a general grasp of the notion of arbitrarily large finite numbers.

4 Logicism

The last strategy, the formulation of which is due to Frege, is to show that arithmetical truths are logical. One way of doing this is to postpone the attempt to explain the substantival use of number-words and to explain instead their adjectival use. This can be done by defining not the numbers themselves but the numerically definite quantifiers:

$$\exists^0 xFx \equiv_{\mathrm{Df}} {\sim}\exists xFx$$

$$\exists^{n+1} xFx \equiv_{\mathrm{Df}} \exists x(Fx \wedge \exists^n y(Fy \wedge y \neq x)).$$

From these definitions we can prove logically that, for example,

$$\exists^7 xFx \wedge \exists^5 xGx \wedge \exists^0 x(Fx \wedge Gx) \supset \exists^{12} x(Fx \vee Gx),$$

which we can interpret as being what is meant by '$7+5=12$'. By the same means we can generate the whole of the positive part of elementary arithmetic, that is, the true quantifier-free equalities. What we cannot generate is the negative part, that is, the true inequalities. For this we would need a guarantee, which logic alone cannot supply, that there is no finite bound to the number of objects that there are. (Wittgenstein in the *Tractatus* (1922) proposed a rather more general variant of the same strategy, but it seems to be susceptible, in the context of Wittgenstein's system, to just the same difficulty over proving inequalities.)

Frege considered the introduction of numbers via numerically definite quantifiers in his *Grundlagen* (1884). What led him to reject it was not the difficulty of deriving inequalities but a problem facing not just this but any strategy involving implicit definitions, namely that such definitions do not fully determine the identity conditions of the objects they attempt to introduce. This is nowadays known as the Julius Caesar problem because Frege posed it by asking how one could tell from the implicit definitions whether or not Julius Caesar is a number. In the present context this question creates a problem not for the project of explaining the adjectival use of number-words but for the strategy of explaining their substantival use in *terms* of their adjectival use. The problem bites, therefore, only when we attempt to move beyond elementary arithmetic and quantify over numbers.

Another logicist strategy considered and rejected by Frege, but the subject of renewed attention at the end of the twentieth century, is to derive arithmetic from the 'numerical equivalence', that is, the contextual principle (sometimes called 'Hume's principle' or '$\mathrm{N}=$') that the number of *F*s is the same as the number of *G*s if and only if the *F*s and the *G*s can be correlated one-to-one. As a technical programme this is perfectly feasible, as Frege was the first to realize. He sketched in the *Grundlagen* a construction of the natural numbers and a proof that they satisfy Peano's axioms, assuming only second-order logic and the numerical equivalence. He nevertheless rejected the strategy of basing arithmetic on this equivalence because the Julius Caesar problem applies to it as much as to the previous contextual strategy. Interest in the strategy has been revived by Crispin Wright, who argued (1983) that the Julius Caesar problem can be solved by appeal to an independently plausible 'sortal inclusion principle' to the effect that objects must be of different sorts if the content of their identity conditions is sufficiently different.

The proposal Frege settled on instead was to define the number of *F*s explicitly as the class of all concepts

equinumerous with *F*. He showed in *Grundgesetze der Arithmetik* (1893-1903) that on this definition of number arithmetic can be deduced from what he took to be logical principles. (One of his allegedly logical principles was inconsistent, however, and subsequent attempts - most notably by Whitehead and Russell - to repair Frege's system have had to appeal to principles even their advocates have baulked at calling logical.)

Frege's solution to the Julius Caesar problem (the underdetermination of the objects of arithmetic by the principles about them to which we are committed) is in any case under threat from the opposite problem: any non-arithmetical determination of the objects which solves the Julius Caesar problem will give numbers extra properties which, since they do not flow from the principles governing numbers, must be arbitrary and hence spurious. This problem was mentioned by Dedekind (in a letter to Weber dated 24 January 1888; see Ewald 1996), but it has come to prominence more recently through a much-cited paper by Benacerraf (1965). Two ways of dealing with it have been advanced under the generic label of 'structuralism'. Dedekind's way - that we are capable, once we have given a logical construction of one model of Peano's axioms, of abstracting away from its particular features to gain a conception of a model without those features - appeals to a mental process (abstraction) that many have found mysterious. Benacerraf's way - that arithmetic should be seen as the study not of one particular model of Peano's axioms but of the structure which all such models have in common - is in danger of relapsing into the axiomatic formalism we considered earlier.

What Russell's paradox has suggested to Michael Dummett is that the concept 'set' is what he calls 'indefinitely extensible' (see Dummett 1991; Dummett, M.A.E. §3). This means that any attempt to regard the objects falling under the concept as forming a definite totality leads inevitably to the realization that there are other objects not in the totality which we are nevertheless forced to admit as falling under the concept. Dummett holds that the presence of 'indefinitely extensible' concepts is a characteristic feature of mathematics which should lead us to espouse for it the anti-realism which his more general meaning-theoretic arguments make room for. He recommends that we abandon the law of the excluded middle and espouse the mathematics of intuitionism but not Brouwer's solipsistic conception of its objects. Dummett thinks that in this way we can retain Frege's logicist insight that numbers are abstract objects truths about which embody deductive subroutines whose application to the world is validated by logic alone.

Just as Brouwerian intuitionism has been accused of an instability which reduces it to strict finitism, not everyone is persuaded that Dummett's position does not collapse into the ultra-intuitionism of middle-period Wittgenstein, according to which the meaning of an arithmetical generalization is identical with its proof. On this view Goldbach's conjecture does not in the present state of knowledge have any meaning at all.

See also: Frege, G.; Husserl, E.; Logical and mathematical terms, glossary of

MICHAEL POTTER

References and further reading

Benacerraf, P. (1965) 'What Numbers Could Not Be', *Philosophical Review* 74: 47-73; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.(Referred to in §4. The impetus for much recent work on the philosophy of arithmetic.)

Benacerraf, P. and Putnam, H. (eds) (1964) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.(Includes work by Hilbert and Benacerraf referred to in this entry, as well as several other relevant articles.)

Carnap, R. (1934) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul, 1937.(The most developed expression of the Vienna Circle's view of mathematics as being true by convention.)

Dedekind, R. (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; trans. 'What are Numbers and What Should They Be?', in W.B. Ewald (ed.) *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Oxford: Clarendon Press, 1996.(A beautiful presentation of arithmetic as the study of 'simply infinite' systems, that is, models of what are now known as Peano's axioms.)

Dummett, M.A.E. (1991) *Frege: Philosophy of Mathematics*, London: Duckworth.(A lucid critique of Frege's logicism and a brief account of Dummett's own views on indefinitely extensible concepts.)

Ewald, W.B. (ed.) (1996) *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Oxford: Clarendon Press.(Includes the letter from Dedekind to Weber referred to in §4, as well as much relevant material by Kant, Hilbert, Brouwer and others.)

Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Jena: Pohle; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, 2nd edn, 1980.(Referred to in §4. An accessible, entertaining classic.)

Frege, G. (1893-1903) *Grundgesetze der Arithmetik*, Jena: Pohle, 2 vols; Part 1 of vol. 1 trans. M. Furth, *Basic Laws of Arithmetic*, Berkeley, CA: University of California Press, 1964.(Referred to in §4. Despite the contradiction at its core, includes much that is still of interest.)

Frege, G. (1980) *Philosophical and Mathematical Correspondence*, trans. H. Kaal, ed. G. Gabriel *et al*, Oxford: Blackwell.(Includes the correspondence between Frege and Hilbert referred to in §1.)

Hilbert, D. (1926) 'Über das Unendliche', *Mathematische Annalen* 95: 161-90; trans. 'On the Infinite', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.(A statement of Hilbert's programme.)

Locke, J. (1689) *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1979.(Referred to in §2. Includes an account of arithmetic that presages some aspects of Kant's.)

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1961.(Referred to in §4. The treatment of arithmetic is to be found in the 6.1s and 6.3s.)

Wittgenstein, L. (1929-30) *Philosophical Remarks*, ed. R. Rhees, trans. R. Hargreaves and R. White, Oxford: Blackwell, 1978; repr. Chicago, IL: University of Chicago Press, 1980.(The source for the ultra-intuitionistic conception of Wittgenstein's middle period.)

Wright, C. (1983) *Frege's Conception of Numbers as Objects*, Aberdeen: Aberdeen University Press. (Referred to in §4. A fascinating defence of the contextual strategy Frege rejected.)

Wright, C. (1986) 'Strict Finitism', in *Realism, Meaning and Truth*, Oxford: Blackwell.(A rare recent example of a philosopher taking strict finitism seriously.)

Armstrong, David Malet (1926-)

David Armstrong was born in Melbourne, and studied philosophy at the Universities of Sydney and Oxford. He returned to Australia to teach at the University of Melbourne and later moved to the chair at Sydney. He has made many major contributions to central topics in epistemology and metaphysics, including perception, laws, universals, the mind, belief and knowledge, and possibility. His overall programme has been the articulation of a naturalistic metaphysics, understood as the doctrine that nothing at all exists except the single world of space and time.

A notable feature of his work in these contentious areas has been its directness and clarity, and the central importance he attaches to squaring what the philosopher says with what science, especially physical science, teaches us. In both these respects he is like another important Australian philosopher, J.J.C. Smart, and together they have influenced the way a generation of philosophers in Australia do philosophy, as well as influencing the doctrines they espouse.

1 Biography

David Armstrong was born in 1926 in Melbourne, Australia. He studied philosophy at the University of Sydney under John Anderson, a major figure in the history of Australian philosophy. He went to Oxford and took the recently established B. Phil. degree in 1954, taught briefly at Birkbeck College, London, before returning to Australia to teach at the University of Melbourne. He succeeded J.L. Mackie in Anderson's chair at Sydney in 1964. Although he studied at Oxford during the heyday of linguistic philosophy (see Ordinary language philosophy, school of), his work has always been very much in the tradition of the systematic metaphysicians who saw their business as the articulation of a comprehensive picture of what there is, what it is like, and how we know about it.

2 Central state theory of mind

Armstrong is probably best known for his influential *A Materialist Theory of the Mind*. He was originally a Rylean behaviourist but became converted to the identity theory of mind, the theory that mental states are brain states, by John Jamieson Carswell Smart (see Identity theory of mind). The book is a defence of a central state version of the identity theory.

Armstrong argues that the concept of a mental state is the concept of a state that plays a distinctive, causally intermediate role between stimuli, other mental states and behavioural responses. Thus, to give the rough idea, pain is the state that is typically caused by bodily damage, and typically causes a desire that the state cease, a desire that in turn causes a behavioural response that is believed will tend to satisfy this desire and minimize the damage. This account makes good evolutionary sense of why we feel pain by making transparent the survival value of pain. This claim, and the corresponding claims for the other mental states, are put forward as conceptual analyses, and constitute the central state theory of mind. Most of the book is devoted to defending central state analyses of various mental concepts.

On the central state theory, the question of the identity of a given mental state M is the empirical question of what plays the distinctive, causally intermediate role assigned to M by the central state theory. Armstrong observed that in each case it will most likely turn out to be some state or other of the brain that plays the distinctive role, so deriving the identity theory from his central state theory.

Armstrong's central state theory gave a central place to causal connections between mental states as well as to those between stimuli, mental state and response. His theory was thus one of the first versions of functionalism (see Functionalism) though it was not marketed as such.

3 Perception, belief, knowledge

In the early- to mid-twentieth century, the view that perceptual experience is most directly an acquaintance with something mental, and the view that physical objects are some kind of logical construction out of perceptual experiences, were widely entertained (see Perception; Sense data). Armstrong's *Perception and the Physical*

World, published in 1961, defends direct realism - the view that we are directly acquainted with physical reality in perception, and that physical objects exist independently of our experiences. The most distinctive feature of his defence is his analysis of perception in terms of the acquisition of belief. Previous discussion of perception by philosophers had made perception implausibly distinct from what is after all its central function - the acquisition of belief.

Armstrong's treatment of belief follows a suggestion of F.P. Ramsey's that belief is like a map by which we steer (see Ramsey, F.P.). Inside our heads is a master map that moves us through the world in such a way that what we desire is achieved to the extent that the map is correct, and individual beliefs are thought of as sub-maps of the master map. This approach to belief is now the standard alternative to the internal sentence theory of belief (see Belief).

His treatment of knowledge is a version of reliabilism. It is widely accepted that knowledge necessarily involves true belief: if S knows that P, then S believes that P, and it is true that P. But not all true belief is knowledge - the truth of a belief may be a fluke, and flukiness is incompatible with knowledge. Armstrong's suggestion, roughly, is that S's true belief that P is knowledge if it is an empirically reliable sign that P.

4 Universals, laws and possibility

The role of a truth maker plays a central role in Armstrong's metaphysics. If some sentence or proposition is true, there must be something that makes it true; if some predicate applies to something, there must be something that makes it true that the predicate applies. You cannot say that the word 'red' applies to X, and that that is all there is to say. There must be something about X, maybe a relational something, that makes it true that 'red' applies to it.

Armstrong's answer to what makes it true that predicates apply to particulars is a species of realism about universals or properties and relations (see Universals). There are universals, and a predicate applies to X because of the universals that X instantiates. These universals have the following features. They exist independently of the classifications that we find natural; his theory is thus a version of realism, not of conceptualism. Second, they are not to be reduced to sets, or to resemblances between particulars; his theory is not a sort of nominalism. Third, there are no uninstantiated universals. Every universal is possessed by at least one thing; thus his theory is not a platonic realism. Fourth, there is not a distinct universal for each semantically distinct predicate. One and the same universal may, and typically does, make many different predicates true of a particular. His theory is thus a sparse theory of universals. Finally, which universals there are is an a posteriori matter to be settled by total science. Armstrong's theory is thus a version of scientific realism. This highly original theory has established itself as a major position in the debate over universals.

Armstrong's account of laws of nature and of possibility draws on his realism about universals. A long-standing challenge in philosophy has been to distinguish universal statements of the form 'Every F is G' that express laws of nature, are nomic universals, from those that express accidental regularities. 'Every massive body attracts every other massive body' is a law of nature. 'Everyone in this room speaks English' is an accidental regularity (in the sense of not being a law, not in the sense of being a fluke) (see Laws, natural). Armstrong's theory, roughly, is that the laws are the universal statements that correspond to relations of nomic necessitation between universals: 'Every F is G' is a law if being F necessitates being G.

Armstrong's account of possibility is a combinatorial one, drawing on his realism about universals. We can think of how things are as a huge arrangement of particulars and universals. The various possibilities can then be thought of as all the combinations and recombinations of these particulars and universals according to various rules for combining particulars and universals. Thus, to give the barest bones of the idea, suppose that there is in fact mass M at point p, and mass N at point q. What makes it possible that there be mass M at q, and mass N at p? The fact that putting being M with being at q, and putting being N with being at p, do not violate the rules of combination.

See also: Australia, philosophy in; Perception; Reliabilism

FRANK JACKSON

List of works

Armstrong, D.M. (1960) *Berkeley's Theory of Vision*, Melbourne: Melbourne University Press.(Critical discussion of George Berkeley's *A New Theory of Vision*.)

(1961) *Perception and the Physical World*, London: Routledge & Kegan Paul.(Discussed in §3.)
(1962) *Bodily Sensations*, London: Routledge & Kegan Paul.(An account of bodily sensations in terms of the perception of happenings inside the body.)
(1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.(Discussed in §3. A paperback edition with a new preface and bibliography appeared in 1993.)
(1973) *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.(Discussed in §3.)
(1978) *Universals and Scientific Realism*, Cambridge: Cambridge University Press, 2 vols.(Discussed in §4.)
(1983) *What is a Law of Nature?*, Cambridge: Cambridge University Press.(Discussed in §4.)
with N. Malcolm (1984) *Consciousness and Causality: a Debate on the Nature of Mind*, Oxford: Blackwell.(A debate between Armstrong and Norman Malcolm.)
(1989a) *Universals: an Opinionated Introduction*, Boulder, CO: Westview Press.(Introduction to the problem of universals from his general perspective on the problem.)
(1989b) *A Combinatorial Theory of Possibility*, Cambridge: Cambridge University Press.(Discussed in §4.)
(1997) *A World of States of Affairs*, Cambridge: Cambridge University Press.

References and further reading

Bacon, J., Campbell, K. and Reinhardt, L. (1993) *Ontology, Causality and Mind*, Cambridge: Cambridge University Press.(Good collection of papers on Armstrong's work with replies by Armstrong. Also contains a bibliography of his writings up to 1992.)
Bogdan, R. (ed.) (1984) *D.M. Armstrong*, vol. 4, *Profiles*, Dordrecht: Reidel.(Good collection of papers on Armstrong's work with replies by him and an intellectual autobiography. Also contains an annotated bibliography of his writings up to 1984.)

Arnauld, Antoine (1612-94)

Antoine Arnauld, a leading theologian and Cartesian philosopher, was one of the most important and interesting figures of the seventeenth century. As the most prominent spokesperson and defender of the Jansenist community based at Port-Royal, almost all Arnauld's efforts were devoted to theological matters. But early on, with his largely constructive objections to Descartes' Meditations in 1641, he established a reputation as an analytically rigorous and insightful philosophical thinker. He went on to become perhaps Descartes' most faithful and vociferous defender. He found Cartesian metaphysics, particularly mind-body dualism, to be of great value for the Christian religion. In a celebrated debate with Nicolas Malebranche, Arnauld advanced something like a direct realist account of perceptual acquaintance by arguing that the representative ideas that mediate human knowledge and perception are not immaterial objects distinct from the mind's perceptions, but are just those perceptions themselves. His criticisms of Leibniz gave rise to another important debate. He also co-authored the so-called 'Port-Royal Logic', the most famous and successful logic of the early modern period. The underlying motives in all Arnauld's philosophical writings were, however, theological, and his greatest concern was to safeguard God's omnipotence and to defend what he took to be the proper Catholic view on questions of grace and divine providence

1 Life and works

Antoine Arnauld was born in Paris on 6 February 1612, one of the many children of an established and well-connected family. He intended to become a lawyer, but the Abbé St Cyran who was spiritual director of Port-Royal (where Arnauld's sister was abbess) convinced him to follow the ecclesiastical life (see Port-Royal). He was ordained and received his doctorate in theology in 1641, and was admitted to the faculty of the Sorbonne in 1643. Most of Arnauld's work throughout his life was theological, devoted to, among other things, an explanation and defence of what he took to be the orthodox Augustinian doctrine of grace and a strict contritionism (see Augustine §5, 7). But he was also responsible for a significant and influential philosophical output, mostly polemical. In 1640, Arnauld was asked by Mersenne to comment upon Descartes' *Meditations*, and his objections were published, with Descartes' responses, as the fourth set in the first edition of the work (1641) (see Descartes, R. §1, 7). The most important of Arnauld's religious works, *De la fréquent communion*, appeared in 1643. This was a defence of the ethical principles of St Cyran and an indictment of what he saw as the indulgent morals of the Jesuits. In the early 1660s, Arnauld co-authored two important works on language and method: the *Grammaire générale et raisonnée* (1660, with Claude Lancelot) and *La Logique, ou l'art de penser* (1662, written with Pierre Nicole, adopting some ideas of Pascal). Better known as the 'Port-Royal Logic', the latter was a treatise on method and reasoning that drew heavily on Descartes' epistemological and methodological doctrines, particularly those found in the *Rules for the Direction of the Mind*.

Meanwhile, Arnauld, as the most prominent representative of the Jansenist movement centred at Port-Royal, continued to be persecuted for his religious views and, like all Jansenists, was suspected of Protestant persuasions and of harbouring politically subversive opinions. In 1656 he was excluded from the Sorbonne for his refusal to submit to the Church on the issue of whether or not Jansenius' *Augustinus* contained heretical propositions. After years of harassment and fearing for his safety, Arnauld left France for the Netherlands in 1679. From there he continued his theological and philosophical polemics. In 1683 he composed *Des vraies et des fausses ideés (On True and False Ideas)*, a philosophical attack upon Nicolas Malebranche's *De la recherche de la vérité (The Search After Truth)*. This was followed two years later by his *Réflexions philosophiques et théologiques sur le nouveau système de la nature et de la grace (Philosophical and theological reflection on the new system of nature and grace)*, in which he addressed Malebranche's theodicy and views on providence and grace. The debate with Malebranche, one of the great intellectual events of its day, continued until the end of Arnauld's life, often in harsh and highly personal terms. He also began a brief but philosophically rich correspondence with Leibniz in 1686 over Leibniz's metaphysical views. Arnauld died in exile in 1694.

2 Arnauld and Descartes

Arnauld's attraction to Descartes' philosophy began early. His objections to the *Meditations* are clearly offered in a constructive spirit by an ally who hopes to see the system move towards greater consistency. Descartes, in fact,

found Arnauld's comments to be the most reasonable and serious of all. Arnauld divided his objections into three parts: the first two dealing with 'philosophical' issues, the third concentrating on 'points which may cause difficulty for theologians'. In the first part, 'The Nature of the Human Mind', he questions Descartes' claim that, since it is possible to form a concept of oneself embodying nothing but the certain knowledge that one is a thinking thing, thought alone constitutes one's essence. The most that can be concluded with certainty from such a premise, Arnauld insists, is 'that I can obtain some knowledge of myself without knowledge of the body'; not, however, that there is a 'real distinction in existence between mind and body'. In the second part, 'Concerning God', Arnauld raises his famous objection to the circularity of Descartes' attempts to draw epistemic warrant from demonstrations of God's existence: 'I have one further worry, namely, how the author avoids reasoning in a circle when he says that we are sure that what we clearly and distinctly perceive is true only because God exists. But we can be sure that God exists only because we clearly and distinctly perceive this. Hence, before we can be sure that God exists, we ought to be able to be sure that whatever we perceive clearly and evidently is true' (Arnauld 1641: 32).

In the final part, Arnauld's most important remark concerns the consequences of Descartes' metaphysics for the Catholic doctrine of Eucharistic transubstantiation. Descartes has emptied the material world of sensible qualities (colour, taste, smell and so on), leaving behind only extension and its properties, modes which necessarily inhere in a substance. His ontology thus appears to Arnauld to be inconsistent with faith, which has traditionally been aligned with the view that the substance of the bread of the Eucharist is either converted into, or annihilated and replaced by, Christ's body, and only the accidents of the bread (colour, taste, smell) remain. Such a real existence of accidents, independent of any underlying substance, is ruled out on Cartesian principles. Descartes responded with one of his tentative reinterpretations of transubstantiation. Ironically, it would be on just this issue of the compatibility of Cartesian metaphysics with the Catholic dogma of the Eucharist that Arnauld would become Descartes' most loyal and vociferous defender over the next fifty years. He generally approved of Cartesianism not just because it seemed closer to the truth than any other system - especially the Aristotelian - but also because its doctrines were the most supportive of Christian piety. Arnauld believed that Descartes 'has demonstrated the existence of God better than anyone else', and that his mind-body dualism has laid the surest foundation for the immortality of the soul.

3 The Arnauld-Malebranche debate

In 1680, Arnauld came across the manuscript of Malebranche's *Traité de la nature et de la grace (Treatise on Nature and Grace)*, which was in the process of being printed. He was so astounded by what he read there that, unable to halt its publication, he decided to publicly refute Malebranche's entire system. His ultimate target was Malebranche's views on grace and on God's general *modus operandi*. But he chose to begin his attack by undermining what he took to be the philosophical foundations of Malebranche's theology. Thus, in 1683 he published *Des vraies et des fausses idées*, an attack on Malebranche's theory of ideas as presented in his most important philosophical work, *De la recherche de la vérité* (1674-5) (see Malebranche, N. §§2, 3, 6).

Malebranche had argued that ideas, the immaterial representations present to the mind in perception and knowledge, are not themselves modes of our thought, as sensations are, but are the very archetypes or essences of things in God's mind, to which we have access through a kind of divine illumination or union with God.

Like most seventeenth-century philosophers, Arnauld believed that representative ideas play an essential role in human cognition. His objection was to thinking of ideas as image-objects in their own right, independent of the mind and mediating its access to the external world. Malebranche's view, he alleged, is a result of the same confusions that gave rise to the Aristotelians' 'sensible species'. As children, we wrongly assume that the images or reflections through which we sometimes see things not actually before the eyes are themselves objects, and later come to suppose that it is through similar image-objects that the mind thinks of things in their absence. But philosophers have realized that even in ordinary sense-perception the material bodies before the eyes are not immediately present to the soul. They conclude that in sense-perception what is directly perceived are representative beings rather than bodies themselves. This line of reasoning, Arnauld argues, treats the soul as if it were material, assuming that the way the senses work and the way the mind works are analogous. More importantly, a theory that makes ideas into mind-independent entities mediating cognition of the world has the absurd consequence that we never know or perceive that world: all we ever perceive are ideas. The mind is surrounded by a 'palace of ideas' that keeps it from the world of things that God intended it to know. Malebranche

‘transports us to unknown lands… where a man sees, instead of the men toward whom he turns his eyes, only intelligible men; instead of the sun and the stars which God has created, only an intelligible sun and intelligible stars’ (Arnauld 1683: 227-8). Even Malebranche must reject such extreme Pyrrhonism (see Pyrrho; Pyrrhonism).

Arnauld goes on to argue that the ideas that function in human perception and knowledge are not ‘representative beings distinct from the mind’s perceptions’, but just *are* those perceptions: ‘I take the idea of an object and the perception of an object to be the same thing’ (1683: 198). To have an idea of a thing just is to perceive or think of that thing; it is not to have some proxy object standing before the mind. The idea is a mental act or operation which, through its ‘form’ (a term borrowed from Descartes, who defines an idea as the *forma cogitationis*), is directed at some object but which is not itself the object of perception. One can thus characterize a thought through its object (for example, as the idea of the sun) by attending to its form, or one can consider it simply as an act or mode of the mind:

> I have said that I take the perception and the idea to be the same thing. Nevertheless, it must be remarked that this thing, although single, stands in two relations: one to the soul which it modifies, the other to the thing perceived, in so far as it exists objectively in the soul. The word *perception* more directly indicates the first relation; the word *idea*, the latter.
>
> (Arnauld 1683: 198)

There is still a sense in which we perceive material objects only mediately or indirectly, since we perceive a thing through the form of the perceptual act (that is, through the idea of the thing). But it does not follow from this that we perceive things indirectly in the strong and unacceptable sense of ‘indirect’ entailed by Malebranche’s account.

In Arnauld’s eyes, then, his debate with Malebranche over the nature of ideas pitted something like a direct realist account of perceptual acquaintance with Malebranche’s representationalist or indirect realist account. But the debate is also a rich source for early modern theories of intentionality (see Intentionality). Arnauld claims that it is not his intention to do away with all representative beings, since he grants that the mind’s modifications are themselves representative of objects. This, in fact, is how his act-ideas achieve their relatedness to objects (the second relation in which every idea stands). Every perception is the perception *of* something because it has a representational content (what Arnauld, again following Descartes, calls its ‘objective reality’) and thus is representative of some object: ‘The perceptions that our soul has of objects are necessarily representative of these objects’ (Arnauld 1684: 381). This representative character is an intrinsic feature of the perceptual act and is what gives the act its intentionality, or directedness-towards-an-object. And for Arnauld this is true for every mental event - not just clear and distinct perceptions, but also sensations and passions. Malebranche, by contrast, claims that only intellections have intentionality, and their intentionality is explained by the real presence to the mind of some distinct object which the mind apprehends - that is, a divine idea - and not by some features intrinsic to the mental operation itself.

Arnauld also directed his considerable critical skills to the doctrine of the vision in God. Much of his concern was focused on Malebranche’s claim that our ideas of extended beings, or the idea of extension itself (what Malebranche calls the ‘infinite intelligible extension’) are in God. He suspected that this was tantamount to placing extension itself really or ‘formally’ in God and thus making God extended or material, and that Malebranche’s doctrine harboured a latent Spinozism or Gassendism (see Spinoza, B. de §4; Gassendi, P. §4). Arnauld accused Malebranche of distorting the thought of both Descartes and Augustine - whom Arnauld and Malebranche alike took as their mentors - and even of propounding anti-Cartesian, anti-Augustinian and anti-Christian views.

The clash over representative ideas - which continued until Arnauld’s death - was only supposed to be a preliminary, however, for the real issue: God’s manner of acting in the realms of nature and grace. Malebranche, in his theodicy, had argued that evil and sin occur because God acts only by what he calls ‘general volitions’ - volitions that carry out general and simple laws. God would like to forestall evil and to save every human being, but actually to do so would require a great number of *ad hoc*, particular volitions and would demand that God should violate the principles of his own nature, which determines him to carry out his plans by the wisest and most simple means. So God must *allow* imperfections in the world and the damnation of many (see Malebranche, N. §5).

Arnauld objected strongly to this model of God's activity. He accused Malebranche of undermining God's omnipotence and of treating God's agency no differently from human agency. He insists that, on Malebranche's account, God is like some distant king who only issues general edicts and has no concern over how his kingdom is run in its details. Such a picture threatens the true Catholic system of divine providence and removes God from direct governance of the world. Moreover, Arnauld rejects any attempt to limit God's absolute power, even if that limit comes from God's own nature. God's absolute freedom is the freedom of a will that determines itself and wills with a complete indifference.

Many of Arnauld's criticisms of Malebranche are rooted in his Jansenism. For it is clear that what really bothers Arnauld is the notion that God could will something (for example, that all humans should be saved or that the world should be without evil) and the object of his will not obtain, that is, that a divine volition might not be efficacious. For Arnauld, all God's volitions are necessarily efficacious, and if God had willed that all humans should be saved - and this claim is, by itself, unacceptable to Arnauld - then all humans would have been saved.

4 Correspondence with Leibniz

This same concern for safeguarding God's omnipotence and freedom is also apparent in Arnauld's critical remarks on Leibniz's metaphysics. In 1686, seeking to enlist Arnauld's support for his ecumenical project of bringing about a reconciliation of the Catholic and Protestant faiths, Leibniz sent a summary of his *Discours de la métaphysique* to Arnauld. Among the propositions upon which Arnauld focused was one related to Leibniz's notion of substance: that the individual concept of every person involves, once and for all, everything that will ever happen to them. Arnauld responded by saying that if that is so, then 'God was free to create or not to create Adam, but supposing he decided to create him, all that has since happened to the human race or which will ever happen to it has occurred and will occur by a necessity more than fatal' (Montgomery 1980: 73). God, having chosen to create Adam, therefore has no freedom or control over the course of events that constitutes the history of the world, since it all apparently follows necessarily from the concept of the first human. Arnauld suggested that Leibniz should cease such metaphysical speculations and think seriously of the condition of his soul and of entering the fold of the Catholic Church. Fortunately their correspondence continued, and Arnauld's objections and queries - concerning necessity, divine providence, causality, the being of possibles, the nature of individual essences and the distinction between essential and accidental properties - led Leibniz to think more deeply about many of his doctrines, and were certainly of great importance in the development of Leibniz's mature metaphysics (see Leibniz, G.W. §1, 4, 6-7).

See also: Dualism; Freedom, divine; Perception

STEVEN NADLER

List of works

Arnauld, A. (1775-83) *Oeuvres de Messire Antoine Arnauld, Docteur de la Maison et Societé de Sorbonne (Collected works)*, Paris: Sigismond D'Arnay, 43 vols.(This authoritative collection contains all works by Arnauld listed below.)

Arnauld, A. (1641) *Quartae objectiones*, in *Oeuvres*, vol. 38; trans. J. Cottingham, D. Murdoch and R. Stoothoff as 'Fourth Objections', *The Philosophical Writings of Descartes*, vol. 2, Cambridge: Cambridge University Press, 1984.(Arnauld's objections to Descartes' *Meditations*.)

Arnauld, A. (1643) *De la fréquent communion*, in *Oeuvres*, vol. 27.(A critique of Jesuit morality and a defence of Jansenist penitential discipline.)

Arnauld, A. (1660) *Grammaire générale et raisonnée*, in *Oeuvres*, vol. 41; trans. J. Rieux and B.E. Rollin as *General and Rational Grammar: the Port-Royal grammar*, The Hague: Mouton, 1975.(Treatise on language and grammar, co-authored with C. Lancelot. The translation contains a preface by A.C. Danto and a critical essay by N. Kretzmann.)

Arnauld, A. (1662) *La Logique, ou l'art de penser (Logic or the art of thinking)*, in *Oeuvres*, vol. 41; trans. J. Buroker, Manchester: Manchester University Press, forthcoming.(Treatise on language, reasoning, and method, co-authored with P. Nicole.)

Arnauld, A. (1683) *Des vraies et des fausses idées*, in *Oeuvres*, vol. 38; trans. E. Kremer as *On True and False Ideas*, Lewiston, NY: Edwin Mellon Press, 1990; also trans. S. Gaukroger, Manchester: Manchester University Press, 1990.(Arnauld's attack on Malebranche's theory of ideas and doctrine of the vision in God.)

Arnauld, A. (1684) *Défense de M. Arnauld contre la réponse au livre des vraies et des fausses idées (Monsieur*

Arnauld's defence against the response to the book on true and false ideas) in *Oeuvres*, vol. 38.(Arnauld's response to Malebranche's initial counterattack.)

Arnauld, A. (1685) *Réflexions philosophiques et théologiques sur le nouveau système de la nature et de la grace (Philosophical and theological reflection on the new system of nature and grace)* in *Oeuvres*, vol. 39. (Arnauld's critique of Malebranche's theodicy and account of God's causal agency.)

References and further reading

Descartes, R. (1620-*c*.28) *Rules for the Direction of the Mind* in *The Philosophical Writings of Descartes*, vol. 1, trans. J. Cottingham, D. Murdoch and R. Stoothoff, Cambridge: Cambridge University Press, 1984.(Descartes' early methodological treatise.)

Dominicy, M. (1985) *La naissance de la grammaire moderne (The birth of modern grammar)*, Brussels: Pierre Mardaga.(An examination of the emergence of linguistics as a specific discipline from the work of Arnauld, Nicole and Lancelot.)

Jansenius, C. (1640) *Augustinus*.(The Bishop of Ypres' monumental work of Augustinian theology.)

Kremer, E.J. (ed.) (1994) *The Great Arnauld and Some of his Philosophical Correspondents*, Toronto, Ont.:University of Toronto Press.(A collection of important essays by leading scholars on Arnauld's philosophy and philosophical theology.)

Kremer, E.J. (1996) *Interpreting Arnauld*, Toronto, Ont.: Univerity of Toronto Press.(Another collection of important essays by leading scholars on Arnauld's philosophy and philosophical theology.)

Laporte, J. (1923) *Exposition de la doctrine (d'après Arnauld)*, in vol. 2 of *La Doctrine de Port-Royal*, Paris: Presses Universitaires de France.(Part 1, 'Les vérités de la Grâce', provides a masterly study of Arnauld's views on grace, divine agency and freedom.)

Leibniz, G.W. *Discours de la méthode*, trans. G.R. Montgomery as *Discourse on Metaphysics/Correspondence with Arnauld*, La Salle IL: Open Court, 1980.(Leibniz's metaphysics; Arnauld's critique, 1686-7.)

Malebranche, N. (1674-5) *De la recherche de la vérité*, trans. T. Lennon and P.J. Olscamp as *The Search After Truth/Elucidations of the Search After Truth*, Columbus, OH: Ohio State University Press, 1980. (Malebranche's major philosophical work.)

Malebranche, N. (1680) *Traité de la nature et de la grace*, in *Oeuvres* vol. 5; trans. P. Riley as *Treatise on Nature and Grace*, Oxford: Oxford University Press, 1992. (Malebranche's solution to the theodicy problem with respect to natural evils and to the distribution of grace.)

Nadler, S. (1988) 'Arnauld, Descartes, and Transubstantiation', *Journal of the History of Ideas* 49 (2): 229-46. (Discusses Arnauld's initial objections regarding the Eucharist and his later defence of Descartes on this issue.)

Nadler, S. (1989) *Arnauld and the Cartesian Philosophy of Ideas*, Princeton, NJ: Princeton University Press.(An analysis of Arnauld's theory of ideas and his account of perception and intentionality.)

Ndiaye, A.R. (1991) *La philosophie d'Antoine Arnauld (The philosophy of Antoine Arnauld)*, Paris: J. Vrin.(A general analysis of Arnauld's philosophical opinions.)

Sleigh, R.C., Jr (1990) *Leibniz and Arnauld: A Commentary on Their Correspondence*, New Haven, CT: Yale University Press.(A rigorous and enlightening analysis of various philosophical themes from the correspondence.)

Art and morality

A complex set of questions is raised by an examination of the relationship between art and morality. First there is a set of empirical considerations about the effect that works of art have on us - one obviously contentious case is that of pornography. Many would argue that the artistic merits of a work are independent of any attitudes or actions it may lead us to adopt or perform. This claim does not survive scrutiny, however, though there is a distinction to be drawn between artistic value and the value of art as a whole. Though there are no coercive arguments to show that we have to take into account the moral qualities of works of art, it is in practice very difficult to ignore them, especially when the point of the work is insistently moral, or when the work is conspicuously depraved.

There is a long tradition, dating back toPlato, of regarding art with suspicion for its power over our emotions, and much of Western aesthetic theorizing has been a response to Plato's challenge. The longest-lasting defence justified art in terms of a combination of pleasure and instruction, though the two never hit it off as well as was hoped. In the early nineteenth century a new, more complex account of art was offered, notably byHegel, in the form of a historicized view in which art is one of the modes by which we come to self-awareness; the emphasis altered from truth to an independently existing reality to truthfulness to our own natures, as we explore them by creating art. Taken into the social sphere, this became a doctrine of the importance of art as an agent of political consciousness, operating in subtle ways to undermine the view of reality imposed on us by the ideologies that hold us captive.

1 The range of issues

The relationship between art and morality is the most fraught and complex in the philosophy of art. One reason for this is that, from Plato onwards, several key issues have not been adequately distinguished. For instance, one important question concerns the effect of works of art on the audience, which is an empirical matter. Does pornography, for example, inflame sexual drives or does it provide substitute gratification? Or does it have different effects on different people (the most plausible answer)? Philosophers, and many others, tend to argue about this question on the basis of nothing more than perfunctory introspection. A separate question is whether works of art should be designed to have moral effects, and if so, how best that can be achieved. The major tradition of Western criticism, from Aristotle to the end of the eighteenth century, has been that literary works should at least aim to instruct as well as delight, and a great deal of debate has been occupied with the relationship between the two goals. This discussion has been complicated, if not bedevilled, by Plato's postulation of the supreme trinity of forms: truth, beauty and goodness. If everything is to be allotted to one of these categories, it would seem that the object of art is beauty. But is it the case that all successful works of art are beautiful? It would seem that, unless one made it a matter of definition that they are, many clearly are not, including some of the greatest. This is especially the case with tragedies, where pain, evil and death predominate. Many artists have evidently felt it their duty to portray reality as vividly as possible, feeling an allegiance to truth rather than beauty. The characteristically philosophical hope that the three categories ultimately coincide seems to be more a case of wishful thinking than of rigorous argument. Nietzsche, in a late unpublished note, put the counter-position with typical incisiveness: 'For a philosopher to say, "the good and the beautiful are one", is infamy; if he goes on to add, ''also the true", one ought to thrash him. Truth is ugly.'

It is already clear that profound and very difficult questions proliferate in this area. It is tempting to hold them in check by postulating a kind of experience that is 'purely aesthetic', then to consider further matters, especially that of the relationship of art to morality, as separable issues. But examples of purely aesthetic experience are less common, it turns out, than might be hoped. And it is unclear what a purely aesthetic experience of, say, *King Lear* would be. Would it be a concentration on the form of the play without regard to its content? But what would that mean? It is unclear that in any of the cases which really bother us a distinction of that kind can be made. A related question asks whether there can be works of art that express strikingly uncouth worldviews. For example, is it a contingent matter that no great Nazi art was produced? And if it is not, how much light does that cast on the relationship between art and morality? And what of works that have an admirable moral content or message, but are of low aesthetic merit, such as Harriet Beecher Stowe's *Uncle Tom's Cabin*? Does that indicate something about the moral content itself, as some have held, or only about the author's lack of talent?

If art is not to be divorced from all other human activity - such a position is known as 'aestheticism', but has never been clearly stated - then the question of its place in the economy of human concerns is clearly crucial. After the effective demise of the classical tradition, the most striking series of attempts to relate art to our other endeavours has been that of the Idealist philosophers. These attempts begin with Hegel and continue at least as far as R.G. Collingwood (§3), and include the politically motivated theorists who stand in a complicated relationship to those earlier philosophers, and who have seen art as a mode of articulating our responses to a reality which we are at least in part responsible for creating. The stress has moved, consequently, from imitation to expression; from the registration of the way we find the world to the expression of our attitude towards it; and thus to sincerity, a term that connects truth and genuineness, and in assorted ways puts art at the centre of our moral experience, and vice versa.

2 The empirical issue

It is often plausibly said that totalitarian regimes take art more seriously than democratic ones. Whatever one's attitude to censorship, it shows at the least that art is felt to have an influence on people's behaviour. And in practice there are virtually no societies in which some form of censorship is not operated. Democracies pride themselves on not prohibiting art, but only various objectionable forms of entertainment or propaganda. It is not clear, however, that the definitions drawn here are more than terminological. 'Art' often has an honorific connotation, so that anything sufficiently degraded is allotted a separate category, such as 'pornography'; 'erotic art' is then claimed to be that which has sexual content but is not arousing (see Erotic art). This is all very suspicious. What it seems to show is that everyone agrees that there are artefacts with a *prima facie* claim to artistic status which can also be arousing, or can incite racial hatred, etc. If one simply decrees that in so far as they have effects on attitudes and behaviour they are not art, or that these effects are irrelevant to their artistic status or stature, this seems an easy way with a hard question. That some works of art have drastic practical effects seems undeniable: a notorious case was Goethe's *The Sorrows of Young Werther*, which led to a wave of romantically inspired suicides. That fact is not used to discredit the book now, though if a contemporary work were to have the same effect it undoubtedly would.

What remain perennial sources of dispute are those works which have a strong emotional effect, but which make it difficult for us to decide, on account of their complexity, how they influence our outlook and thus our conduct. This lies at the centre of the long-lasting debate about the effect of tragedy (see Tragedy). It seems strange that we should be unable to decide whether tragedies of great power, such as *Oedipus the King*, exalt, devastate, bewilder, affirm or question basic values, and so on. Yet all these positions are held and argued about with passion. That the issue remains so contentious is no doubt partly due to the fact that there is often a discrepancy between how they make us feel and what we think they ought to make us feel, and that it is difficult to keep in focus, after the immediate experience of such works, what we actually did feel. As any experience of great complexity recedes, we are apt to simplify its effect on us, especially if we are in the grip of a gratifying theory. The only thing that can be confidently asserted in this area is that there is no doubt that some works of art have moral (and of course political) effects.

3 Artistic value and the value of art

Although works of art do have moral effects, it has been argued that it is not part of their value that they do. It was largely in reaction to the oppressive moralism of Victorian criticism that Wilde remarked that there are no such things as moral or immoral works of art. And it is widely agreed (though not universally) that art which, in Keats' words, 'has palpable designs upon us' is inferior. But that does not mean that the moral effects of art are irrelevant to its value as art, only that it is less likely to achieve them if they are crudely manifest. Nor does the oft-cited fact that those who frequent art are not conspicuously better than those who do not have any cogency. It is very difficult to improve people and failure to do so may reflect more on people than on the attempt. Further, to claim that art in some cases derives its value from its moral qualities is not to claim that they are invariably relevant. It is not easy to envisage a plausible argument to show that much instrumental music or painting has moral properties or effects; and the same is true of a considerable amount of literature. The claim, made by Tolstoy, that art is only valuable in so far as it has a beneficial influence on people's behaviour, cannot be refuted, but equally cannot be established. It is an attitude one might choose to adopt, but it would certainly eliminate a great deal of what is normally called art. It might be claimed that any interesting aesthetic theory will do this, and that what matters is

how radical we are prepared to allow an aesthetic theory to be. Any theory that rules out much of what is widely valued, either on the ground that it is not art at all, or that it is bad art, risks appearing ridiculous.

There is a certain amount of art (not a great deal, but a significant amount) which takes the form of something approaching a parable. Tolstoy's great short story 'How much land does a man need?' is a good example. In such cases it is hard to say that someone could appreciate it without grasping its moral point. To confine oneself to admiration for its economy, tension, verve, but not to recognize the end those qualities serve would be to fail to recognize a realized intention. Similarly, in a rather more complex case, to see or listen to Beethoven's *Fidelio* without taking into account its concern with injustice, heroism and freedom (and that not surprisingly it favours the last two and opposes the first) would be so strange that we would normally say that anyone who claimed to be indifferent to its political and moral qualities was simply not responding to the work. If such a person said that they were only moved by the music, and regarded the drama as too crude to take any interest in, we would wonder how they could be moved appropriately by the music without acknowledging that it was articulating the dramatic action. Of course one could listen to the music in a purely abstract way, regarding the voices simply as instruments, but that would not be listening to *Fidelio,* but only to an aspect of it, somewhat as if one were to enjoy the sound of a poem in a language that one did not understand. Of course one need not agree with the morality that one takes a work to be propounding or embodying in order to value it. But once more it does not follow from this that one ignores its morality or regards it as irrelevant, nor does it follow that one thinks less well of the work. One may welcome art which puts a moral position that one can't share in a plausible light. The phenomenology of responses to art is itself a highly complex matter, but it seems to be the case that in imaginatively entering the world of a work one is enabled to test one's reactions to people who hold sets of moral views very different from one's own, and that this is one of the reasons that we value the experience of novels, dramas, etc. Though it is notoriously hard to say precisely what effect a powerful work of art has on us, there is wide agreement that we do sometimes feel changed by such works, which is a cause both for valuing them and also for feeling anxious about them. The latter tends to have dominated Western speculation on the subject, because many philosophers, Plato being the first and most influential, have taken it that art typically works on 'the passions', of which they have harboured deep suspicions, because they are thought of as hard to govern, irrational, selfish and destructive. And the greater the work of art in terms of its power to affect us, the more dangerous it may be.

This raises the issue of internal and external reasons for valuing art. Might it not be the case that the reasons we have for valuing a work as art are, at least sometimes, equally reasons for disapproving of it morally? An analogy with sport might make the point more clearly. Within such a sport as soccer one may behave in a way that is beneficial to the side one is playing on, in that it helps to score a goal; but the whole activity might be frowned on for fostering a spirit of competitiveness, if that is thought to be a bad thing. Victorian schoolmasters were very keen on competitive sports for a range of reasons which might need examining. But it would be possible to disapprove of the whole set of sporting activities that they promoted while being able to make judgments about how well someone was playing a specific game. Similarly, the reasons we give for putting a high value on a work of art might, in the context of our complete set of standards, be reasons for thinking that art itself is deplorable; or that most of what is highly esteemed is deplorable. In practice it turns out to be difficult to maintain this distinction, so that we often find people making what appear to be artistic judgments when they are really making judgments about the whole institution. Conversely, an overenthusiastic application of the distinction leads people to adopt a very narrow set of terms or concepts which they take to be appropriate for judging individual works, and many considerations which might be considered relevant are claimed to be external, that is, to be dealing with a different question - the nature and value of art itself.

4 Art and the self; art and society

Continuing this line of thought, we can deal briefly with two other lines of thought about art and its relation to other enterprises, both of which derive more or less directly from Hegel (§8). It was characteristic of him to take a historical perspective on art, valuing the art of different ages and cultures in terms of its connections with their other concerns. For us, living in a culture decisively formed in all its aspects by Christianity, art takes on a 'romantic' aspect, in his idiosyncratic use of that term. Hegel sees the tension in all kinds of art as the relationship between the medium, which is sensuous, and the content, which is spiritual. Operating along these lines, but without subscribing to Hegel's often bizarre views, Idealist aestheticians have seen artistic activity (which is by no means confined to artists) as the attempt to achieve a perfect congruence between the inner and the outer. Thus the

emphasis shifts to art as self-expression; it is seen as a demanding and ultimately moral activity, since one finds out who one is in the process of its creation.Nietzsche, hostile to Idealism in general, none the less took over this element and in his middle period advocated that we should make our lives into works of art, thus effecting a synthesis of the moral and the aesthetic.

Deviating much further fromHegel, but inconceivable without him, is the tradition of thought loosely called 'Marxist'. Marx himself contributed little on the subject of art, but many theorists working within his general framework have produced general theories of the nature and purpose of art and many specific judgments on works of art, both of which construe it as a crucial contribution to the class struggle. The most crass versions of such a line of thought were perpetrated in the Soviet Union and its satellites, where art was esteemed for bolstering the workers' spirits as they battled against the bourgeois and fascist enemies. The result, in terms of both the kind of art encouraged and endorsed and the fate of artists who failed to conform, is an eloquent testimony to the necessity for a large degree of freedom of expression. 'Socialist realism', which was the approved art of the Soviet bloc, like Fascist art, did not succeed even ephemerally. In the case of Marxism, however, there has been an immense and impressive body of theory in which it is clearly seen that if radical political goals are to be effectively furthered by art, they will not be achieved with slogans and propaganda, but involves the transformation of consciousness - a point that could have been taken from Hegel in the first place. The leading group of theorists has been the Frankfurt School, of which the salient figure was Theodor W. Adorno (§5), a man of immensely rich culture and a modernist in his artistic affiliations. His views on the function of art as critique of society, and the consequent particular value judgments he made on a vast range of works of art, especially music, represent what may be the most impressive single body of work in the field. He expounds the complex relations between the moral and social on the one hand and the artistic on the other with a detailed force that breaks through the obscurity with which they are expressed.

5 Conclusion

As is the case with many profound conceptual issues, it can confidently be predicted that the nature of the relationship between art and morality will never be settled. What does seem to have been established, though, is that the discussion must always take account of historical considerations. The appearance on the scene of new art-forms, such as the realistic novel and large-scale instrumental music, alone makes it unlikely that a theory produced at a given time will ever be adequate for later times. And the more art we have to ponder upon, the more intimate will be the relationship between our theorizing and what we are theorizing about. Hegel may have been exaggerating when he said that writing about art would finally take the place of art itself, but the element of truth in this claim means that the range of moral views available to us and their engagement with the vast body of art guarantees an indefinite future for meditation on the connections between them.

See also: Aristotle; Croce, B. §2; Emotion, in response to art; Johnson, S.; Kant, I. §12; Poetry §§1-3; Pornography; Schiller, J.C.F.; Tolstoy, L. §4

MICHAEL TANNER

References and further reading

Adorno, T.W. (1991) *The Culture Industry*, London: Routledge.(The most accessible introduction to Adorno's Marxist attack on mass culture, and a defence of the social role of art.)

Aristotle (*c*. mid 4th century BC) *Poetics*, trans. R. Janko, Indianapolis, IN: Hackett Publishing Company, 1987, chaps 6, 9, 11, 13, 14.(A defence of art, and of tragedy in p articular, in terms of its moral effects in purging our emotions.)

Collingwood, R.G. (1938) *The Principles of Art*, Oxford: Clarendon Press.(The introduction and first book provide a sophisticated exposition of art in terms of expression, and launch a vehement attack on all other theories).

Hegel, G.W.F. (1835) *Lectures on Aesthetics*, trans. T.M. Knox, Oxford: Clarendon Press, 1975.(The introduction gives Hegel's most succinct account of his general theory of the development of art, and its varying role in the progression of self-consciousness.)

Johnson, S. (1765) 'Preface to Shakespeare', in H.R. Wouhuysen (ed.) *Samuel Johnson on Shakespeare*, London: Penguin, 1989.(The culminating statement of the classical tradition, defending art when it shows us the good prospering and the bad coming to grief.)

Pater, W. (1858) *The Renaissance*, Berkeley, CA: University of California Press, 1980.(The conclusion offers the

classic statement of the irrelevance to art of all but 'purely aesthetic' properties - though Pater adds an important concession.)

Plato (*c*.380s-370s BC) *Republic*, trans. P. Shorey, Cambridge, MA: Loeb Classical Library, Harvard University Press, 1930, 602c-608b.(The first and most famous denunciation of art on moral grounds.)

Tolstoy, L. (1896) *What is Art?*, trans. A. Maude, New York: Bobbs-Merrill, 1962.(An attack on almost all art, except that which unites all people in fellow-feeling and that which brings us closer to God.)

Art and truth

Some things are true within the world of a literary work. It is true, in the world evoked by Madame Bovary, that Emma Roualt married Charles Bovary. In this entry, however, we are not concerned with truth in *fiction but rather with what it is for a work of art to be true of, or true to, the actual world. Representational works represent states of affairs, or objects portrayed in a certain way. The concept of truth naturally gets a grip here, because we can ask whether the represented state of affairs actually exists in the world, or whether a represented object exists and really is the way it is represented to be, or whether a representation of a kind of thing offers a genuinely representative example of that kind. If so, we could call the work true, or true in the given respect.*

A work will often get us to respond to what is portrayed in a way similar to what our response would have been to the real thing - we are moved to fear and pity by objects we know are merely fictions. But a work could also portray *characters responding in certain ways to the imaginary situations it conjures, often with the implication that the response is a likely human emotional or practical response to that situation, or a response to be expected of a character of the given type, and we could reasonably call the work true if we believed the portrayed reaction was a likely one.*

Arguably, if we judge a work to be in some respect true to life, we must already have known that life was like that in order to make the judgment. But, interestingly, works of art appear to be able to portray situations that we have not experienced, in which the portrayal seems to warrant our saying that the work has shown (that is, taught) us a likely or plausible unfolding of the portrayed situation, or shown us what it would have been like to experience the situation. It is also said, especially of narrative fiction, that, because of its power to show us what various alternative imaginary situations would be like, it can enlighten us about how we ought to live.

So we may consider how a work of art might be a vehicle of truths about the actual world. This gives rise to a further question - sometimes called the problem of belief - of whether the value of a work of art as *a piece of art is related to its truth. If a work implies or suggests that something is the case, ought I to value it more highly as art if I accept what it implies as the truth? Alternatively, should I take it as an aesthetic shortcoming if I do not?*

1 Literature as a vehicle of truth: the *Poetics*

Aristotle’s *Poetics* provides a good starting point for considering the claim that art informs us about or illuminates the actual world. Discussing tragedy (which in Aristotle’s terms is a form of poetry) he makes the following claims: (1) that plot is the most important aspect of a tragedy, which is a depiction of actions which form a unity, and (2) that tragedy (along with other forms of poetry) is more philosophical and more important than history, because it makes universal statements rather than statements about particular events.

Here is a suggestion that connects claims (1) and (2). Taking up the point in (1) that plot is a depiction of actions, we can assume that Aristotle means the actions of the characters in the story. A plot, we can now say, is unified to the extent that its depicted actions follow upon one another in a natural or plausible way. What we find plausible depends, of course, on our assumptions about human behaviour. Aristotle suggests that a good plot is able to bring an audience together in its responses to events and hence that some responses are practically universal, resting on truths about human conduct that are ‘necessary or probable’ - a phrase he uses repeatedly in the *Poetics*. This provides a connection with (2). By structuring itself around these ‘universal’ assumptions, a good tragedy can be said to be more like philosophy than history, in that it is biased away from the accidental, towards generality. History simply records events as they happen. It is a matter of indifference to the faithful recorder of history whether historical events are the product of accident or agency, or whether recorded actions are sane or mad, intelligible or whimsical. By contrast, a tragic plot, according to Aristotle, is spoilt if it is interspersed with accidents of nature or with actions that we cannot readily understand through our first-hand familiarity with ourselves and other people. Such a plot is ‘episodic’ in Aristotle’s terms - his word for a narrative sequence whose elements are not, as in a properly unified plot, linked by what is ‘necessary or probable’.

Consider *Madame Bovary* as our ‘tragedy’. Might it be said to depict actions as forming a unity? To begin with, Flaubert’s readers are carried along by the narrative because of the way the depicted actions are motivated and connected by familiar human impulses. His characters behave in ways that are accessible and plausible given their

different temperaments, histories and circumstances. The function of Flaubert's narrative is to draw us into the developing story in a way that depends on our tacit assent to the successive links in the plot, so that we become involved in the fates of its characters as if they were part of our own lives. This suggests the following way in which narratives can be informative. Out of the 'necessities and probabilities' that drive the plot, eliciting our implicit consent along the way, Flaubert develops a complex situation with a dramatic outcome that lies outside our experience and would normally beg an explanation. But his plot and characterization enable us to experience this world from within, so that we get to know what the fictional situation would have felt like if real. At the same time we come to understand how Emma feels, and hence why she does what she does. So we increase the range of our experience and advance our understanding, because a piece of human behaviour that may otherwise have seemed strange or impenetrable has been placed within the range of what is intelligibly human. Where a brief report of Emma's gruesome and tragic demise might have met with incomprehension or prompted a shallow and moralistic banality, the narrative elicits instead a response closer to Flaubert's: 'Madame Bovary, c'est moi!'

2 Literature and moral insight

Since literature seems able to provide an impression of what a possible situation would feel like to us if actual, it suggests itself as a useful tool for exploring the merits of various imagined alternatives to existing lifestyles or kinds of society. Hilary Putnam (1978) has claimed that since the choice between styles of living involves a full human response, including the capacity to feel, it is not an appropriate subject for science, which is essentially propositional in character. (Putnam surmises that we store certain kinds of information in the form of images rather than propositions.) The choice of lifestyles requires practical rather than theoretical reasoning, the former, as much as the latter, being a process subject to rational criticism. The search for better ways of living suggests a significant role for the imagination - and for imaginative literature. Putnam's idea is not that literature should present ideal lifestyles as 'solutions' to the question of how we ought to live, but that it should play a critical role. For example, by showing us what it would be like to live as a certain kind of person in a society organized in a specific way or having various imagined features, it can present us with 'perplexities' which enable us to refine our thinking about the desirability of the way of life represented.

The most substantial explication of the idea that literature can illuminate how we ought to live has come from Martha Nussbaum. Her detailed analyses of literary works - the novels of Henry James in particular - set out to show that literature provides a means of extending our moral awareness beyond the limits to which traditional moral philosophy can take us. 'Schematic philosophers' examples', Nussbaum says, 'almost always lack the particularity, the emotive appeal, the absorbing plottedness, the variety and indeterminacy, of good fiction; they lack, too, good fiction's way of making the reader a participant and a friend' (1990: 46). Only once its many aspects are adequately represented can a complex ethical question become the subject of clear reflection, and then it is easier to see the various and sometimes conflicting values that are at stake. For example, by taking us into the lives of its characters, Henry James' *The Golden Bowl* is able to show us how Maggie Verver's aspirations to a certain kind of moral perfection stand in the way of the full flourishing of her marriage; her love for the prince, fully acknowledged, calls for a more complex moral stance in which she must relinquish the hope of guiltless moral perfection. Thus, she must confront the fact that the full expression of her love may require 'a tragically necessary blindness' (1990: 144) - it may require her sometimes to turn away from or even wound others who are close to her.

3 Other arts

It is natural that representational art, especially literature, should offer the clearest examples of works that are in some way true of the world. But it may not offer the only examples. Turning to nonrepresentational art, Jerrold Levinson has argued (1990) that there are various ways in which music, a mainly nonrepresentational art form, can aptly be described as true. Suppose we accept (as seems reasonable) that anger is a destructive emotion. Passages in the fourth movement of Beethoven's *Pastoral* symphony express anger, Levinson suggests, and the anger is presented as destructive. In this case it seems natural to describe the passages as true. A different kind of example occurs where a transition in a piece of music from one emotional quality to another carries the implicit suggestion that that transition is psychologically plausible in human terms; and we judge the music true if we think it plausible in this way.

Nelson Goodman (§2) holds that all the arts, representational and nonrepresentational, serve a cognitive function

(1972). Thus it is a mistake to associate science alone with the cognitive and to limit the concerns of art to the evocation or expression of feeling. For him, inventiveness in both art and science consists in the development or modification of elements in a symbol system. Successful symbols illuminate the world through the aptness with which they fit their subject-matter, rewarding those who engage with them with enlightening new ways of seeing the world.

Goodman's ideas can fruitfully be read in conjunction with Ernst Gombrich's work on the visual arts (1951). Gombrich starts by attacking the 'myth of the innocent eye'. The recognition of what lies in our field of view when we perceive, he argues, requires the organization of visual input: what we see is structured by what we expect to see. Just as scientific discovery is preceded by the invention of successful hypotheses, the development of our awareness of the visual world requires inspired modifications to the 'schemas' we bring to bear upon it. The history of painting, Gombrich suggests, can be seen as an experimental process by which our actual visual capacities are gradually enhanced by painters' corrective modifications to the existing schemas available for representing the visible world. Rather than simply assuming that our visual awareness reflects the world in an unproblematic way, providing the touchstone by which the accuracy of a picture can be judged, Gombrich's account construes painting as a means by which that awareness is developed.

The view, shared by Goodman and Gombrich, that art is an instrument of perception, is not unproblematic. But it is an important development in twentieth-century aesthetics, and its exposition in the work of these philosophers provides some of the best reading the subject has to offer.

4 The artistic relevance of truth

If a work of art implies the accuracy or correctness of a certain claim, representation, perspective or attitude, can it or should it increase our estimation of the artistic worth of the work if we take the claim or representation to be true or the attitude or perspective to be accurate or correct? A negative answer is implied by I.A. Richards when he writes of our response to literature:

> the question of belief or disbelief, in the intellectual sense, never arises when we are reading well. If unfortunately it does arise… we have for the moment ceased to be reading poetry and have become astronomers, or theologians, or moralists, persons engaged in quite a different type of activity.
>
> (1929: 277)

Support for Richards's claim might come from the idea that art is a distinct category of human activity, with its own purposes and hence its own criteria of merit. On this view, just as we judge the merits of, say, the binding of a book by the standards of bookbinding, so we should judge poetry (remaining with the case of literature) by the special qualities that set poetry apart as an activity: the aptness of verbal choices, the displayed command of metre and rhythm, the elegance and unity of the poem as a whole, and so on. But it is unclear that poetry (or any other art form) comprises a self-contained activity with a fixed purpose that defines it as poetry, and dogmatic to suppose that its 'purpose' necessarily excludes the accurate representation of how things stand in the world. Furthermore, a view like Richards's is hard to reconcile with past and current critical practice. Henry James is widely admired for his moral subtlety, and critics generally assume (and it would seem bizarre to deny) that this quality adds to his stature *as* a serious novelist.

Malcolm Budd (1983) has produced an important argument that bears on this issue. He makes the point that the value of a work of art as art is always intrinsic to the experience it offers. It therefore does not include the beneficial effects the work may have upon our lives. For example, if a novel contains an insight that could illuminate our lives, its insightfulness may accrue to its value as art, but only in so far as the insight informs the reading experience itself, and not in virtue of the work's educative effects. This seems right, for someone could value and enjoy a work as art even though, due to complacency, forgetfulness or some other idiosyncracy, they failed to respond to it in such a way that it benefited or informed their life. But to say this is not to drive a wedge between artistic or aesthetic value, on the one hand, and moral or intellectual value, on the other. These values are not in opposition. It may be its intellectual or moral character that accounts for much of a work's impact as an object of experience; if it displays intellectual or moral shortcomings - philosophical immaturity, for example, or racist attitudes - then that is likely to diminish the value we place on the experience it offers. So, regarding the artistic relevance of truth, we may indeed value a work less as art if we fail to be persuaded about the truth of a

claim it implicitly makes about the world - provided its failure to persuade lessens its value as an object of experience.

See also: Art and morality; Art, value of; Fictional entities; Narrative

PAUL TAYLOR

References and further reading

All the articles listed are untechnical and should be accessible to someone with only a limited knowledge of philosophy.

Aristotle (*c*. mid 4th century BC) *Poetics*, trans. R. Janko, Indianapolis, IN: Hackett Publishing Company, 1987. (Aristotle's famous examination of the nature of poetry and tragedy.)

Budd, M. (1983) 'Belief and Sincerity in Poetry', in E. Schaper (ed.) *Pleasure, Preference and Value*, Cambridge: Cambridge University Press, 137-57.(A clear and careful discussion of the parallel issues of sincerity in literature and the problem of belief.)

Gombrich, E.H. (1951) 'Meditations on a Hobby Horse or the Roots of Artistic Form', repr. in *Meditations on a Hobby Horse and other essays on the theory of art*, London: Phaidon, 1-11, 1963.(A good entry point into Gombrich's philosophy of fine art.)

Goodman, N. (1972) 'Art and Inquiry', in *Problems and Projects*, New York: Bobbs-Merrill, 103-19.(A good introduction to Goodman's important and influential philosophy of art.)

Gulley, N. (1979) 'Aristotle on the Purposes of Literature', in J. Barnes, N. Schofield and R. Sorabji (eds) *Articles on Aristotle, 4: Psychology and Aesthetics*, London: Duckworth, 166-76.(A brief, authoritative and readable introduction.)

Levinson, J. (1990) 'Truth in Music', in *Music, Art and Metaphysics*, Ithaca, NY: Cornell University Press, 279-305.(This is a clear and fertile discussion of the topic of its title.)

Nussbaum, M.C. (1990) *Love's Knowledge: Essays on Philosophy and Literature*, New York: Oxford University Press, 3-53 and 125-47.(The fullest available treatment of the way literature illuminates morality.)

Putnam, H. (1978) *Meaning and the Moral Sciences, Part Two*, London: Routledge & Kegan Paul, 83-94.(A relatively lucid discussion of the capacity of literary works to give us knowledge; compares the cognitive function of literature with that of science.)

Richards, I.A. (1929) *Practical Criticism*, London: Routledge & Kegan Paul, 271-91.(Referred to in §4.)

Art criticism

To criticize a work of art is to make a judgment of its overall merit or demerit and to support that judgment by reference to features it possesses. This activity is of great antiquity; we find Aristotle, for example, relating the excellence of Sophocles' Oedipus Rex to the excellence of its plot construction. Criticism became a topic in philosophy because reflection on the kinds of things said by critics generated various perplexities and in some cases encouraged a general scepticism about the possibility of criticism. Two general and related problems in particular have taxed philosophers. The first is the question of whether criticism is a rational activity, that is to say, whether critics can give reasons for their judgments that would persuade potential dissenters of the rightness of those judgments. The second, a matter to which Kant andHume made notable contributions, is the problem of the objectivity of critical judgments, it being widely believed that critical appraisals are wholly subjective or just 'a matter of taste'. Arguments that use deductive or inductive reasoning to demonstrate the possibility of proofs of critical judgments are generally agreed to have failed. Another approach redescribes the critic altogether, not as someone who uses argument to prove their judgments to an audience, but as someone who aims to help the audience perceive features of the work of art and understand their role in the work. This entry will concentrate on the issues of the rationality and objectivity of art criticism.

1 A case for subjectivism

Disagreements in art criticism are widespread and frequently intractable. Subjectivists, wishing to demonstrate the impossibility of objective judgment in art, often begin with this apparently indubitable fact - though this tactic suffers from the fact that an observer is as likely to notice the widespread agreement in critical judgments over the pre-eminence of such figures asSophocles,Mozart, Tolstoy, Beethoven, Rembrandt and Shakespeare. Nevertheless, the existence of often intense disagreement cannot be denied, but to concede this is not as yet enough to establish a subjectivist case. For while there are vehement and intractable disagreements in, say, mathematical theory and in physics, these disagreements do not entail the subjectivity of physics or mathematics.

Apart from the fact that disagreements occur in it, there must then be some aspect of criticism that underlies the claim that it is subjective in a way that physics and mathematics are not. It is tempting to think that while mathematics and physics possess proof or decision procedures, agreed on by practitioners of those subjects, in terms of which enquiries in those subjects proceed and disputes in them are in principle resolvable, there are no such agreed procedures in criticism. There, in lieu of argument and proofs, we have only unsupportable opinion. The denial that criticism is a rational activity (one in which reasons can be given for judgments) becomes a principal ground for asserting that critical judgments are subjective.

2 The impossibility of induction and deduction

The above section sketches the claim that there are no proof procedures in art criticism, as there are in physics and mathematics. But what proof is offered for such a claim? One way to argue that proof procedures have no place in criticism would be to take the two most commonly accepted forms of proof - deduction and induction - and show that these cannot be invoked in support of critical judgments.

For induction, one might argue that since all pictures by Rembrandt that have been hitherto examined have been found to be great paintings, any hitherto unexaminedRembrandt is probably a great painting. This use of induction has two weaknesses. First, on what basis was it asserted that the first Rembrandt ever examined was a great painting? Since it was the first Rembrandt, the assertion cannot have been based on inductive proof. Instead, an appreciator probably looked at the picture and simply pronounced that it was a great painting; here the sceptic merely repeats the question of whether that judgment can be proved by reasons. Second, the inductive judgment I have sketched has a peculiar uselessness for the would-be appreciator. Induction might indeed lead the appreciator to conclude that an unexaminedRembrandt was probably a great one. But the appreciator wants not merely to *know indirectly* that the picture is a great one: they wish to see and experience its greatness directly.

Various attempts have been made to introduce inductive procedures into criticism in order to underpin its status as a rational proof procedure. One of the most famous uses the notion of the Golden Section. The claim is that a certain ratio is to be found in all pictures that are the subject of favourable judgments. Hence the presence of that

ratio in a picture constitutes a reason to believe that it is admirable. The claim that any picture exhibiting the Golden Section is admirable is not a self-evident truth, since it seems possible to imagine cases in which a picture that is admirable fails to exhibit the requisite ratio, or cases in which the ratio is exhibited in a picture that is a failure. Hence the claim must be an inductively based and probabilistic one: since admirable pictures have been found to exemplify the Golden Section, and since this picture exemplifies the Golden Section, we conclude that this picture is probably admirable. But, again, one wishes to know how the correlation between being an admirable picture and exemplifying the Golden Section was established in the first place. Presumably, someone asserted that a picture was admirable, then noticed that it exemplified a certain ratio and made a generalization from this. Then, however, the original judgment that the picture is admirable is not justified by induction, but rather underpins subsequent inductive arguments. Again, the sceptic can ask what, if anything, underpins the original judgment that the picture is admirable.

Deduction fares little better in justifying critical judgments. In deductive argument, one offers statements in support of one's judgment, and these statements, if accepted, absolutely force the interlocutor to accept the judgment. It is difficult to see how such a process could work with judgments of art criticism. If I say that a painting is superb and support this statement with the assertion that its composition is admirable, its drawing excellent and its colours radiant, one of two things may happen. First, you may agree that the composition, the drawing and the colour are as I say they are and yet not be forced to concede that the picture is superb. You might, for instance, claim that the composition, drawing and colour, though individually excellent, do not work together. Alternatively, you may deny that the colour, the drawing or the composition are as I say they are. How then am I to convince you that they are? I may go on to claim that what gives the picture its compositional quality is the presence of a patch of colour in a certain position. The problem is that, while you may agree that the patch of colour is in this position, you may not see the contribution that it makes to the composition, and thus not yet be forced to conclude that the location of the patch of colour entails the presence of the compositional quality to which I referred.

3 Generality

The conclusion that neither inductive nor deductive reasoning can be used to prove, and thus justify, critical judgments may be reinforced by another, related set of considerations. A reason has to have a generality. If your doing a certain action in certain circumstances is a reason for praising you, it is reason for praising anyone who does that action in those circumstances. Some thinkers (Stuart Hampshire (1954), for example) have argued that reasoning in criticism is impossible because of the impossibility of this sort of generality in that context. Thus, it is claimed, the fact that a painting has a patch of colour in a certain position may be the explanation for its admirable compositional features. But the existence of that patch in that location cannot be cited as a reason for concluding that the painting is admirably composed. For precisely that shade of colour in the same position in another picture may be the cause of that picture's bad composition. And if exactly the same feature can sometimes count for a conclusion and sometimes against it, it cannot be cited as a reason for believing that conclusion.

Care needs to be exercised here, however. Sibley has remarked that we can make a distinction between what he calls the 'neutral' features of a work of art and the 'merit' features. A neutral feature would be a feature such as the possession of iambic pentameter, an alliteration or a colour patch in a certain position. The feature is neutral with respect to merit conclusions because it is possible without any unnaturalness to say, for example, 'it is the alliteration that spoils this line', and, in the case of another poem, 'it is the (self-same) alliteration that makes this poem.' Statements about neutral features cannot, indeed, be used as reasons in support of critical judgments. However, asSibley observes, other terms do not have the neutrality of those just cited. If we take terms such as 'witty', 'radiantly coloured', 'elegantly composed', 'subtle in its harmonic variations', 'ham-fisted' or 'ponderously executed', then these terms seem to have a positive (or negative) merit force. Though there would be nothing unusual about saying 'What makes it so good is its wit', it would be odd to say, 'What makes it bad is the subtlety of its harmonic variations.' These terms do then seem to have a general positive or negative force and are generally (and so genuinely) reasons for thinking something good or bad. However, as Sibley also pointed out, this positive or negative force is at best *prima facie*. That is to say, although the possession of wit is a *prima facie* reason for saying that something is good, we cannot argue that because something possesses wit, it is for that reason good or has something good about it; for the wit might be out of place, as, for example, it is sometimes said to be in the Porter scene in *Macbeth*. For that reason we cannot deduce a work's value from the fact that it has wit

in it. Once again the critic's judgments seem not demonstrable by reason, a fact that, again, may appear to support subjectivism.

4 An alternative model

Hypnotized by the successes of the physical sciences and mathematics, many who thought about criticism - including, notably, the Russian Formalists - sought to remodel it along the lines of these activities and to look for inductive and deductive ways of proving critical judgments, the impossibility of such proofs being evidence of the unscientific subjectivity of criticism. In view of arguments already given no programme of this sort could succeed. Observable features, such as onomatopoeia, alliteration, patterns of plots, no less than sound patterns in music or colour areas in paintings are neutral features, as likely to count for merit as against it, and cannot support critical judgments in any deductive way.

In fact the model of the sciences and mathematics provides the wrong model for the procedures of critical judgment. What is required in criticism is not inductive and deductive argument but an ability to *see* the qualities of visual works of art, *hear* the qualities of music and *notice* the features of literature. The model that best fits the practices of criticism appears to be the model of getting someone to perceive something rather than arguing someone into something. This is not, as it is with the colours of traffic lights, simply a matter of pointing the gaze of a colour-sighted person in the right direction. Like wine-tasting, aesthetic perception may require practice and experience. The critic, in helping one to see, hear or notice, can use a variety of devices, ranging from simply pointing out the features believed to be present to the use of analogies, metaphors, comparisons and gestures, in the way in which a conductor may help a choir to sing a phrase in a certain way by hand movements.

If this kind of model is adopted - and, given that we use our eyes and ears in artistic appreciation, what more appropriate one suggests itself? - then the questions of rationality and objectivity assume a different aspect. First, the scope of reasoning in artistic judgment is immediately narrowed. What the critic wishes to do is to help the reader, viewer or listener to see or hear what is there to be seen and heard. And although critics can give reasons for looking and listening ('because the object will reward your contemplation'), and although they can give reasons, possibly of a deductive or inductive kind, for believing that something has merit or demerit ('most people think this is good, so try it'), they cannot give reasons that will make people see or hear something. The case is analogous to that of ordinary perception: I can give you reasons to look at the traffic light but not to see that it is red.

Although critical judgments are thus not objective in the sense that reasons can be given to prove them, this is not the only way in which objectivity is possible. We need to ask, then, what kind of objectivity is appropriate to critical judgments. Given that these are perceptual judgments, the kind of objectivity they will have, if any, will be the kind that can be possessed by perceptual judgments. We do have an inclination to believe that statements about the colours of traffic lights and the sounds of fog horns can be true and false, right or wrong. That possibility depends upon there being some kind of agreement in visual response among human beings in the presence of such things as grass and tomatoes. Some, notablySibley, have suggested that this kind of agreement is found in cases of art appreciation, and hence that this activity, too, has some claim to objectivity. That this objectivity is dependent upon human responses does not, as Hume argued (1757), prevent there being standards in terms of which we might adversely judge the adequacies of certain responses: for example, the response of someone who thought Barry Manilow superior toBach.

5 Final remarks

To assert thatBach is superior toBarry Manilow is not to rule out anyone's right to prefer Manilow to Bach. As Kant remarked in one of the most important treatises in aesthetics (1790), if all one wishes to say is that one likes a thing, then, at least in aesthetics, who is to deny one that right? But if one wishes to say that the thing is good, great or awful, one is making a claim that goes beyond any statement of one's personal preferences, a claim that, as I have suggested above, may invoke an appeal to a shared sentiment.

Next, it needs to be noted that discussions of subjectivity and objectivity are bedevilled by assertions that judgments must be either one or the other. Better perhaps to think of the subjective and the objective as poles of a spectrum; to think of the judgments we make, affected as they will almost certainly be by our life histories and our

distinctive human personalities, as lying somewhere along this spectrum; and to be characterized, at most, as tending towards one or other of its poles according to the perhaps excusable degree of idiosyncrasy they display.

Finally, we may sum up art criticism as the activity of detecting and of helping others to detect the perceptual value and devaluing features of works of art. In understanding that activity we need to distinguish two uses of the term ‘reason’: the justificatory sense, when, for instance, we say, ‘A reason for believing that it is good is that most competent critics say that it is so’, which is of doubtful use in art criticism; and the explanatory sense, when we say, ‘The reason it is balanced is the patch of red in the left-hand corner.’ Acute critics are often good at offering such explanations, and this is indeed one of the ways in which we might be helped to see the qualities to which our attention is being directed.

See also: Aesthetic concepts

COLIN LYAS

References and further reading

Beardsley, M. (1981) *Aesthetics*, 2nd edn, Indianapolis, IN, and Cambridge: Hackett Publishing Company.(Denies the relevance of critical references to sincerity. Chapter 10 contains a striking account of critical evaluation, vigorously debated in Sibley (1983). The postscript to the second edition contains important remarks on criticism.)

Bell, C. (1914) *Art*, London: Chatto & Windus.(Queries references to representation in criticism.)

Erlich, V. (1955) *Russian Formalism*, New Haven, CT: Yale University Press.(Informative and very accessible account of the Russian Formalist attempts, referred to in §4, to make criticism scientific.)

Hampshire, S. (1954) ‘Logic and Appreciation’, in W. Elton (ed.) *Aesthetics and Language*, Oxford: Blackwell. (Referred to in §3 as denying the possibility of generality of reasoning in critical appraisal.)

Hume, D. (1757) ‘Of The Standard of Taste’, in *Philosophical Works*, London: Longman Green, 1875.(Referred to in §4 as combining a commitment to subjectivism and a belief in standards of taste. A justly famous, elegant and enormously influential piece of writing.)

Kant, I. (1790) *The Critique of Judgment*, trans. J.C. Meredith, Oxford: Oxford University Press, 1928.(Part I, *The Critique of Aesthetic Judgment*, makes an important contribution to the analysis of aesthetic judgment in an extraordinarily demanding but essential work.)

Lyas, C. (1992) ‘The Evaluation of Art’, inO. Hanfling (ed.) *Philosophical Aesthetics*, Oxford: Blackwell.(A much expanded version of many of the arguments in this entry.)

Sibley, F.N. (1965) ‘Aesthetic and Non-Aesthetic’, *Philosophical Review* 74.(Clear and accessible statement of the claim, in §4, that aesthetic judgment is perceptual, and of the claim, in §5, that reason in judgment needs to distinguished from reason in explanation.)

Sibley, F.N. (1968) ‘Aesthetics and Objectivity’, in *Proceedings of the Aristotelian Society* Supplementary vol. 42. (Referred to in §4 as offering a clearly presented account of the analogy between colour and aesthetic judgments and of the case for treating the latter as having a degree of objectivity.)

Sibley, F.N. (1983) ‘General Criteria in Aesthetics’, in J. Fisher (ed.) *Essays in Aesthetics*, Philadelphia, PA: Temple.(Presents the distinction between merit and neutral terms referred to in §4 and argues the case for some area of generality in critical reasoning. Contains also a lucid and powerful account of the theory of critical judgment offered by Monroe Beardsley (1981).)

Strawson, P.F. (1974) ‘Aesthetic Appraisal’, in *Freedom and Resentment*, London: Methuen.(Relevant to the discussion of the generality of critical reasoning.)With the exception of the Kant essay, all of these are clearly and non-technically presented.

Art works, ontology of

In trying to decide what kinds of thing art works are, the most natural starting point is the hypothesis that they are physical objects. This is plausible only for certain works, such as paintings and sculptures; in such cases we say that the work is a certain marked canvas or piece of stone. Even for these apparently favourable cases, though, there is a metaphysical objection to this proposal: that works and the physical objects identified with them do not possess the same properties and so cannot be identical. There is also an aesthetic objection: that the plausibility of the thesis for painting and sculpture rests on the false view that the authentic object made by the artist possesses aesthetically relevant features which no copy could possibly exemplify. Once it is acknowledged that paintings and sculptures are, in principle, reproducible in the way that novels and musical scores are, the motivation for thinking of the authentic canvas or stone as the work itself collapses.

For literary and musical works, the standard view is that they are structures: structures of word-types in the literary case and of sound-types in the musical case. This structuralist view is opposed by contextualism, which asserts that the identity conditions for works must take into account historical features involving their origin and modes of production. Contextualists claim that works with the same structure might have different historical features and ought, therefore, to count as distinct works.

Nelson Goodman (1981) has proposed that we divide works into autographic and allographic kinds; for autographic works, such as paintings, genuineness is determined partly by history of production: for allographic works, such as novels, it is determined in some other way. Our examination of the hypothesis that certain works are physical objects and our discussion of the structuralist/contextualist controversy will indicate grounds for thinking that Goodman's distinction does not provide an acceptable categorization of works.

A wholly successful ontology of art works would tell us what things are art works and what things are not; failing that, it would give us identity conditions for them, enabling us to say under what conditions this work and that are the same work. Since the complexity of the issues to be discussed quickly ramifies, it will be appropriate after a certain point to consider only the question of identity conditions. For simplicity, this entry concentrates on works of art that exemplify written literature, scored music and the plastic and pictorial arts.

1 Works as physical objects

As in other parts of metaphysics, a theory of what art works are should seek to provide as simple and economical an account as possible, consistent with robust intuition about sameness and difference between works. We shall begin with maximum simplicity, introducing complications as required by consideration of particular cases, some of them real and some of them highly artificial constructs. By the end we shall have arrived at a position of some complexity, but one which has the virtue of treating all works of the kind we shall consider here in a uniform manner.

It seems relatively easy to say what a painting or a sculpture is: a canvas on a gallery wall, a piece of stone shaped in a certain way. In that case, these works are just physical objects. Three objections to this proposal have been made. The first is that art works have properties that a physical object could not have: aesthetic, expressive and representational properties, for example. This objection relies on the assumption that a physical object can have only physical properties. But a physical object might become, say, valued without ceasing to be physical, and being valued is not a physical property.

The second objection is this. If Michaelangelo's statue *David*, D, is identical with a certain block of stone, B, then D and B must have the same properties. But B existed prior to D, which came into existence only when Michaelangelo set to work. So 'existing prior to D' is a property B possesses and D lacks. One response to this objection is to say that physical objects are four-dimensional, extended in time as well as in space, and therefore possessed of temporal parts that are themselves physical objects. In that case we can say that D is identical with a physical object B*, which consists of that temporal part of B that began when *David* was fashioned from the stone and ends when the degraded condition of the stone no longer warrants our saying that *David* exists. But now the distinct-properties argument can be used again to show that D and B* are different. It is true of D that, if Michaelangelo had not lived, it would not have existed; the same is not true of B*. Substantial deformation would

destroy D but not B*.

An alternative and less ambitious proposal would be that art works (of the kinds we are currently considering) are not identical with physical objects; instead they are embodied in or constituted by physical objects. I shall not pursue this line of thought further here. Many of the arguments about identity and constitution apply equally to art works and to other kinds of things (for example, tools), and so belong to general metaphysics, as do the arguments we have considered so far. The only specifically aesthetic argument that has been developed in connection with this is an argument against both the idea that art works are physical objects and the idea that art works are constituted by physical objects. To this argument we now turn.

The claim that certain works are physical objects or are constituted by them depends on an intuitive contrast between singular and multiple works. Among works of the former kind are paintings and sculptures, where the object fashioned by the artist (the ‘authentic’ object) seems to have a unique status - a proper appreciation of the work requires that the viewer sees that object rather than any copy of it, however good. With novels, plays and poems, on the other hand, the original inscription of the work by the author (the autograph) has no special significance for appreciation of the work; any word-for-word copy of the autograph will do. But the significance placed on authenticity in painting and sculpture is due to the fact that aesthetically adequate copies of these works are very hard to achieve. Every visible feature of the work is potentially relevant to the proper appreciation of it, and so an aesthetically adequate copy of the work would have to look exactly like the original. With literature, mere sameness of spelling with the autograph is all that is required of a copy for it to allow us to appreciate the work fully. But this difference between painting and literature is a merely technical one and cannot be the basis for treating these two forms as fundamentally distinct. It is possible (though by no means easy) to produce copies of paintings and sculptures indistinguishable from their originals by the modes of perceptual access appropriate for those works. If this were frequently done, the aura of indispensability that surrounds originals would dissipate. They would be regarded as no more essential to the existence of the work than autographs of novels currently are. (Originals might continue to command high prices on grounds of their personal and historical interest, much as autographs of novels and poems do.) In that case, the claim that in painting and sculpture the work is the authentic object is only as plausible as the comparable claim that the novel is the autograph copy, and this is untenable, as we shall now see.

2 Works as types

Let us consider works of literary and musical art. There are physical objects significantly associated with both genres and these are analogous to authentic objects in painting and sculpture: autograph copies of the text or score. But neither Austen’s autograph nor any other copy of the text can be regarded as identical with the work *Emma*, since no particular copy need survive in order for the work to survive. (Word processing makes the ontological irrelevance of an autograph particularly obvious.) Nor can the work be identified with the class of copies of its text or score. Classes are so defined that a class could not have had members different from those it does have. But there could have been more or fewer or different copies of *Emma* than there are, without the identity of the novel being threatened. With music, drama and other ‘performance’ works, there are performances to consider as well as copies of the score. But a sonata cannot be identified with its original performance by the composer - there may be no such performance - nor with the class of its performances: a symphony might have had different performances from the ones it did, or might never have been performed, without being a different or non-existent work.

For these reasons we may choose to identify literary and musical works with ‘types’, of which copies of their text or score are ‘tokens’. It is common to make a distinction between word- (or letter-) types and tokens, since the same word (or letter) may be inscribed many times; these inscriptions are tokens of the type. The same distinction applies to items of musical notation; we say there are a number of D# semiquavers on the page. Sentence-types are sequences of word-types (which are sequences of letter-types), and texts are sequences of sentence-types. On the view we are now considering, literary works are texts. The closest parallel to this idea for musical works is that they are scores, which are similarly defined as sequences of note-types, the tokens of which are particular inscriptions of notes. Since texts and scores as defined here are abstract structures, we may call an approach of the kind just described a version of structuralism.

One objection to this proposal is that it does not allow us to say that works are created by their authors or composers, since types are abstract objects not capable of being affected by human action; instead, what we

normally think of as the act of composition would, according to the structuralist, be an act of discovery. This objection is not decisive. The structuralist need not deny any of the evident facts about composition: that hard work and talent are required for the composition of significant works; that without that talent and effort these works would not be available to us. We admire those who prove difficult and important mathematical theorems, and our admiration does not dissipate with the thought that these theorems are not created by the people that prove them. Whether the work of the artist is to be described as creation or as discovery can be counted as spoils to the victor in this debate.

Another, more serious objection, but one answerable within a generally structuralist outlook, is that the proposal does not make any provision for the obvious difference between ‘performance’ and ‘non-performance’ works, since it treats plays, symphonies and novels alike as having tokens which are particular inscriptions. We can rectify this by saying that musical works are sequences of sound-types rather than note-types. While note-types have as their tokens particular inscriptions, sound-types have as their tokens particular sounds. A sound-type might be identified by specifying a pitch, a duration and a degree of loudness. (To conform with our ordinary ways of individuating sounds we would need to preserve some vagueness in these specifications.) Thus defined, an instance of the work would consist in the actual production of tokens of the sound-types constitutive of the work; it would be a performance of the work. A comparable stipulation can be made for plays and other non-musical performance works.

There are facts in addition to mere performability which seem equally to demand reflection in our theory of the work. Musical works, at least typically, are intended to be performed *on certain instruments*, and the specification of the work in terms of sound-types alone fails to accommodate this: a sound-type can be produced in ways that would be inappropriate for many works. We could meet the performance-means objection by specifying that the work is a sequence of sound-types-as-performed-on-certain-instruments. But we must go further still. A person who produces the appropriate sequence of sounds on the appropriate instruments by accident, or while improvising, is not, strictly speaking, performing the work. We need to specify further that the work is a sequence of sound-types-as-performed-on-certain-instruments-as-a-result-of-intentionally-following-a-certain-score.

3 Works and their histories

What if our performer, Jones, was following the score of an existing work - the *Hammerklavier* sonata, for instance - but was doing so as a result of having hit upon that score himself by an act of composition undertaken in ignorance of Beethoven’s previous efforts? Would he be performingBeethoven’s *Hammerklavier* sonata? Structuralists say yes, assuming Jones was playing the instrument specified by Beethoven. Some writers have argued that the correct answer is that he is not playing Beethoven’s work, and that consequently the structuralist approach elaborated above is wrong; he is playing a distinct work composed by Jones himself. The argument for this is another version of the distinct properties argument of §1. For the well-informed listener, much of what is interesting and valuable in a musical work derives from the work’s art-historical features. Works are variously describable as stunningly original, fresh, deliberately anachronistic, shamelessly plundered from better composers. We notice and enjoy or deplore their quotation from and commentary on other works. If such features are features of the works themselves, we cannot say that Beethoven’s and Jones’s works are identical just because they are correctly performed on the same instruments in accordance with the same score. Jones’s twentieth-century work may reflect the influence of Brahms, express outrage at the practice of atonality and consciously submit itself to the discipline of an earlier age; Beethoven’s does none of these things. So these works have distinct properties and cannot be the same work. Structuralism’s identity conditions are ahistorical and fail to locate this vital divergence on historical properties, so they falsely identify distinct works.

Three responses to contextualism have been voiced. The first agrees that works do have the kinds of historically determined features just mentioned, but says that their having them is an objection to the structuralist account of work-identity only as long as we think of these historical features as essential features of works. This response is best clarified in terms of possible worlds. The contextualist is interpreted as saying that in the actual worldBeethoven’s work, B, has a certain historically determined feature, F, while in some merely possible world Jones composes an identically scored work, J, which lacks F. But it is agreed that objects have properties in one world that they lack in another; such properties are accidental rather than essential. If we think that historical properties of the kind just described are accidental, we can therefore continue to assert the identity of B and J. This

response misunderstands the contextualist's position. The world the contextualist imagines as a counterexample to structuralism is one in which Beethoven's composition of B is just as it is in the actual world *and* in which Jones also composes J. If it is agreed that in this (merely possible) world, B possesses F and J does not, then certainly B and J are distinct, since an object cannot possess and lack a property in a single world. Contextualism does not depend on any controversial assumption about what properties of works are essential.

The second response claims that historical features such as those described above are not properties of works at all. Instead they are properties of acts of composition. So there is no barrier to identifying B and J, and we may say that there is in this case one work and two acts of composition. The problem for someone who advocates this response is to find a principled way of distinguishing properties of the work from properties of the act of composition. An appeal to common speech or practice will not do, since we commonly speak of works themselves as original or conventional. We might say that, in the musical case, the properties of the work itself are just those that determine how it sounds; the historical properties just mentioned do not do this, since it is agreed that B and J sound the same (in the sense that one could not tell, by attending to sound alone, that something was a performance of B rather than of J). This would have the uncomfortable consequence of driving us back to the original structuralist position, according to which the correctness of a performance of a work depends only on how it sounds and not on the choice of instruments for performing it.

The third response to contextualism says that the historical features described above are features of the work, but insists that they are incompletely specified by expressions like 'is original'. Rather, they are all properties that need to be relativized to contexts of composition, much as velocity needs to be relativized to a frame of reference. On this interpretation, what looks like the paradox of saying 'B is F, while J, identical with B, is not' emerges as the consistent avowal that B is F-in-the-Beethoven-context while J is not-F-in-the-Jones-context. This position faces the same difficulty as the previous one: it must be supported by a principled distinction between relativized and unrelativized properties if it is not to be judged *ad hoc*. If such a distinction can be made, a version of structuralism may be defensible.

4 Autographic and allographic

In the previous section we focused on the question, 'Under what conditions is this work identical with that one?' We now turn to a different but related question: 'What makes an object an instance of this work rather than of that one?' The canvas painted byLeonardo is a genuine instance of *The Madonna of the Rocks*; it is also an authentic instance deriving from the hand of the artist, and those who think painting a singular art say that, for that reason, it is the only instance. The prints pulled from Rembrandt's plate by the artist himself (or authorized by him) are genuine instances of the etching *Tobit Blind*; they are also authentic instances. Are they the only instances? Those who think the artist's canvas is the only instance of a painting will presumably say yes, on the grounds that printmaking is like painting in that only authentic items are instances of the work. In line with this idea, Goodman classifies paintings, sculptures, prints and moulded figures as 'autographic'. Something is an instance of an autographic work only if it has a certain history of production; these arts are all autographic, then, because to be an instance of a work in any of these forms, the thing in question must issue from the hand of the artist, or by their instructions. Autographic arts can be singular, as with painting, or multiple, as with prints.

What, on the other hand, makes something an instance of the novel *Emma*? For Goodman, the criterion is that it should have the same spelling as Austen's autograph; in this and like cases we identify objects as instances of particular works by their structural features and not in terms of their histories of production. A typing monkey could produce something which is, in Goodman's sense, an instance of *Emma*, and a monkey with the luck to strike the right piano keys will produce a performance of the *Hammerklavier* sonata. Novels, poems, plays and musical works are all 'allographic' because of the irrelevance of history to the identification of work-instances. Allographic works are all multiple.

If we accept the argument of §1 according to which authentic instances of paintings, sculptures, prints and moulded figures have no privileged status, we shall deny Goodman's claim that, for works of these kinds, the only instances are the authentic ones. We shall say that indiscernible copies of authentic instances are also instances. Still, we would remain broadly in agreement with Goodman, for the condition of being a copy of an authentic instance is one that appeals to history of production, and so these arts remain autographic. A more radical disagreement with Goodman emerges when we consider the contextualist's argument of §3 which claims that

literary and musical works need to be identified historically. If that argument is right, and two distinct novels or sonatas could have the same structure, then, before we can know whether this performance is a performance of B or of J, we need to know whether the performers' knowledge of the score is traceable through a causal chain back to B's act of scoring or to J's. That way, being an instance of a literary or musical work is partly a matter of history of production, and so all works are to be categorized as autographic. Something which looks exactly like Leonardo's *Madonna of the Rocks* might not be an instance of that work because it is copied from another, independently produced, canvas of identical appearance. And this bundle of pages is a copy of *Emma* rather than of the same-spelled but independently produced *Schemma* because it derives causally from Austen's act of composition and not from that of her lesser-known rival Schmausten. We now have a uniform theory of identity conditions for art works of all the kinds considered here.

See also: Artistic forgery; Type/Token distinction

GREGORY CURRIE

References and further reading

Currie, G. (1989) *An Ontology of Art*, London: Macmillan.(Emphasizes the requirement that an ontology of art incorporate an historical dimension.)

Currie, G. (1991) 'Work and Text', *Mind* 100: 325-40.(Applies antistructuralist arguments to the case of literary works.)

Elgin, C.Z. and Goodman, N. (1988) 'Interpretations and Identity: Can the Work Survive the World?', in *Reconceptions in Philosophy*, London: Routledge.(Argues for the identification of the literary work with its text.)

Goodman, N. (1981) *Languages of Art*, 2nd edn, Brighton: Harvester Wheatsheaf.(A classic work containing, along with a great deal else, an account of the autographic/allographic distinction.)

Kivy, P. (1983) 'Platonism in Music: A Kind of Defence', *Grazer Philosophische Studien* 19: 109-29.(Defends the view that musical works are discovered.)

Levinson, J. (1990) *Music, Art, and Metaphysics*, Ithaca, NY, and London: Cornell University Press.(An important collection of essays giving arguments and examples which favour contextualism.)

Walton, K. (1970) 'Categories of Art', *Philosophical Review* 79: 334-67.(Argues that the aesthetic qualities of works depend on the genres to which they belong.)

Wollheim, R. (1978) 'Are the Criteria of Identity that Hold for a Work of Art in the Different Arts Aesthetically Relevant?', *Ratio* 20: 29-48.(Argues that questions about work identity must be settled by reference to the artist's theory of work identity.)

(1980) *Art and Its Objects*, 2nd edn, Cambridge: Cambridge University Press.(A clear statement of both the physical object and the structuralist hypotheses.)

Wolterstorff, N. (1980) *Worlds and Worlds of Art*, Oxford: Clarendon Press.(An influential approach somewhat different from those discussed here.)

Art, abstract

The use of the term 'abstract' as a category of visual art dates from the second decade of the twentieth century, when painters and sculptors had turned away from verisimilitude and launched such modes of abstraction as Cubism, Orphism, Futurism, Rayonism and Suprematism. Two subcategories may be distinguished: first, varieties of figurative representation that strongly schematize, and second, completely nonfigurative or nonobjective modes of design (in the widest sense of that term). Both stand opposed to classic representationalism (realism, naturalism, illusionism, mimeticism) understood as the commitment to a relatively full depiction of the subject matter and construed broadly enough to cover the traditional 'high art' canon through to Post-Impressionism. Analytic and Synthetic Cubism are model cases of the first subcategory while Mondrian's neoplasticism and Pollock's classic drip works are paradigms of the second. Though the effect was revolutionary, the positive motivations for this degree of abstraction in visual art were not wholly new. What was new was the elevation of previously subordinate aims to the front rank and the pursuit of certain principal aims in isolation from the full pictorial package. Thus abstract art variously celebrates structural and colour properties of objects, scenes and patterns; effects of motion, light and atmosphere; aspects of perceptual process, whether normal or expressively loaded; and forms expressing cosmic conceptions, visionary states or utopian ambitions. With a few exceptions (for example, the Futurists) the founders of abstract art were far from lucid or forthcoming about the significance of their work, and viewers have found successive waves of abstraction initially baffling and even offensive. But abstract art now forms a secure part of the 'high art' canon, though generally its appeal is less well understood than that of the classic modes of representation. Criticisms of abstract art have also become more lucid.

The chief philosophical issues affecting abstract art concern the definition of the term and the delineation of subordinate types; the relation between abstraction and other modes of avant-garde art that superficially resemble it; the magnitude of the artistic values so far achieved by the various forms; and finally the theoretical limits of significance attainable by abstraction as compared with the limits encountered in figurative art.

1 The history of the category

The nearest precedent of the category and its label is the use, from about 1870, of the term 'abstract music' for music without lyrics or programme. Until late in the nineteenth century, use of the vocabulary of abstraction in relation to the visual arts was rare and predominantly pejorative. For example, Gustave Courbet in 1861 claimed that abstraction, by which he meant undue emphasis on any partial aspect of art, puts the true end of art beyond reach. Purity was a more common metaphor in early writings about the new art, as it had been in relation to music. The positive implications of 'abstraction' were first exploited in a major way by Willhelm Worringer (1908) and Wassily Kandinsky (1911), though the two had quite different art in mind. They argued persuasively that the 'urge to abstraction' is a 'primal artistic impulse' (Worringer's phrases). The term gained additional currency from the proclamation by Apollinaire and other champions of the new trends in France of a new art of 'pure painting' which drew more on 'conceived reality' than on the data of everyday vision (and not coincidentally evaded rivalry with photography). The category and label remained problematic for decades. Some artists (Braque and Mirò, for example) objected to any of their works being called abstract although the present consensus favours taking many of them that way.

2 Basic distinctions

'Schematizing abstraction' covers here certain forms of depiction which severely curtail the extent to which the visible properties of their subjects are made manifest. The outcome might have been produced by beginning from a natural motif seen from a single point of view and then, for ends quite removed from verisimilitude, reducing its descriptive content as well as adding pattern that is either not descriptive or deviantly so, relative to traditional standards. In contrast, nonfigurative or nonobjective abstraction eliminates all literally descriptive references in order to free expressive or intellectual content from all encumbrance. The work might be produced by beginning from lines, forms, textures, and colours, which are then worked up into aesthetically self-sufficient totalities. Neither category has a sharp boundary, and claims for the superiority of one over the other are doubtful. But the distinction deserves respect since it denotes a significant difference of interest of both the artist and the appropriately responsive viewer.

3 Schematizing abstraction

The schematic rendering of figures, appearances and space of an appropriate sort, carried to the required degree, allows seemingly endless variations, for which no systematic classification has yet been devised. Analytic Cubism, a prime instance, geometricizes contours, evacuates or etherealizes solids, and expunges much other detail. On the constructive side, it repeats edges at different eye-levels, imports contours from quite different points of view, and fractures objects and regions of empty space into overlapping facets or shards, as if an eccentric crystalline structure were being revealed. Since many of the interpolations and displacements are depictively cryptic or altogether nonfunctional, the perceptible nature of the motif tends to be obscured or confused. In classic cases (for example, Picasso's *Portrait of Kahnweiler*, 1911) the central motif (a man dressed in a neat wool suit over a ribbed shirt, embellished by necktie and watch chain, with his hands folded, and a bottle on a table to his right, light coming in from his left, and so on) emerges against the odds from a farrago of translucent angular clutter. Carried to an extreme the process renders the motif unidentifiable. The appeal of this sort of abstraction has been explained in terms of its power to convey the intricacy, instability or illusoriness of perception or of the material world, or alternatively the inner vitality of objects - a power deemed beyond the range of any traditional mode of representation.

Synthetic Cubism revises the schematizing process in two directions. Aspects that are curtailed further include atmosphere and depth. Compositions are dominated by template-like forms, often closely stacked, blocking recession, and with it, atmosphere. Media motifs, printed materials and other inherently flat elements are deployed to similar effect. Suggestions of ambient lighting and perceptual process, typical of Analytic Cubism, are also generally reduced or entirely banished. All this, together with the evident arbitrariness of many of the forms, creates the impression of the composition having been built up (synthesized) from invented components rather than derived by analytic decomposition of a natural motif. Contrarily, descriptive content is in certain respects enhanced compared with the norm in Analytic Cubism. Some forms signify sizeable sections of recognizable objects (a guitar, a table, for instance).

Synthetic Cubist figures may be invested with an uncanny presence, as they are in Picasso's *Three Musicians*, 1921 (The Museum of Modern Art version). Having assumed some of the properties of the flattened planes of which they are composed, yet being represented in full view without any optical interference, the figures have become enigmatic new realities inhabiting an equally unnatural space. This contrasts with the effect in Analytic Cubism of normal objects seen through a radically fractured lens.

Another of the many currents of schematic abstraction is generally known as lyrical or expressionist and is represented by various works of Kandinsky, Franz Marc, Paul Klee and others. Objects and space are dematerialized by blurred and calligraphically naive transforms of the Cubist schemata just cited, and by a dispersal of emphasis over the entire picture plane. Colour is typically dramatic or evocative.

4 Nonfigurative or nonobjective abstraction

Preliminary steps towards nonfigurative abstraction are discernible in forms of schematizing abstraction, notably in the works of Kandinsky, Malevich and Mondrian, but it seems best to define the category independently of this lineage and of the implication of a higher aspiration that may cling to it. (For the same reason 'pure' is best avoided, as is 'impure' for schematizing abstraction.) The key concept is that of a design which conveys no implication of actual (physical or 'tactile') space, though it is optional how strictly one applies this criterion. In no case are figure-ground distinctions excluded, any more than they are in abstract decorative designs. It is enough that depth cues be too faint, too fragmentary, too flagrantly inconsistent, or too dispersed for the image as a whole to convey the sense of any conceivable three-dimensional space. 'Figures' in the weak sense are woven or soaked into the ground, or fused to its surface like letters on a page; or the ground may press so closely upon the figures or switch relations with them so extensively that the continuity, uniformity and ubiquity necessary to accommodate physical contents is implicitly denied. The depth relations suggested by the parts of the design taken in isolation thereby collapse into mere appearances, though they remain vital to the aesthetic effect. This accords with the description by Greenberg and Fried of the pictorial space in nonfigurative abstractions as not tactile, but optical, indeterminate, or virtual - in a sense doubly illusionistic.

Kendall Walton (1990) proposed an alternative explanation of the figurative-nonfigurative distinction. On this

view a nonfigurative Suprematist composition by Malevich represents the coloured patches that are part of the design itself fictionally standing in three-dimensional relations to each other. Were the design figurative, the patches would represent fictive three-dimensional forms, for example, ordinary physical objects, as standing in comparable relations.

Whatever the proper analysis, it seems plain that the appeal of most nonfigurative abstractions depends significantly on the viewer's susceptibility to spatial and more specifically pictorial ways of seeing. Under normal (if not always optimal) viewing conditions Mondrian's grids evoke lattices, windows, partitions or street patterns as if, or almost as if, 'seen' in the sense proper to representations. Similarly Helen Frankenthaler's works often evoke cloud and landscape forms. Even the Arthur Danto 'push-pull' tensions in works by Hans Hofmann and the titanic dimensions of Franz Kline's motifs capitalize on the involuntary exertion of the perceptual energies typically brought to bear on ordinary spatial presentations. The mystery felt to inhere in the more potent nonfigurative abstractions derives largely from the profusion of such subliminal underground connections, which trigger responses even when viewers are discouraged from making a literally representational reading. Possibly even the appeal of Ellsworth Kelly's hard-edged colour-blocks and Agnes Martin's misty stripes devolves in part from the challenge they pose to the space-obsessed visual system.

Closely related to the preceding is the dialectic of image and support stressed by formalist analysts such as Greenberg and Fried. For instance, recognizing that traditional art often seeks to make the support 'disappear' in favour of the illusory image (whereas in the ordinary, nonpictorial experience of a surface the percipient seeks to resolve perceptual flux and obtain a firm grasp of the objective reality) abstract artists have often sought to reverse priorities or play with tensions. Thus drips and slashing brushstrokes are intended to bring the surface to the fore, leaving only fleeting suggestions of images; and colour is poured directly on to unsized, unprimed canvas to bond image to fabric and evade the normal dichotomy of drawn edge and coloured area. Likewise solid 'op art' designs dissolve into fluctuating afterimages or visual squirm in the viewer's perceptual field, even when one attempts to see the surface merely as surface. Formalists regard the exploration of such paradoxes and inversions as a prime aim of nonfigurative abstraction. Others place greater weight upon the externally referential intellectual and expressive content which supposedly becomes available through the unsettling of normal perceptual expectations.

5 Intellectual and expressive values in abstract art

Writers such as Arthur Danto (1981) have demonstrated how much of a worldview may be teased out of seemingly inarticulate art. Thus the brushstroke paintings of De Kooning or their effigies in Lichtenstein are shown to have, in context, a wondrous depth of implication. Arguably the effect results from otherwise inexpressive elements acquiring magnified significance given an initial limitation to minimal means, a long tradition of thought and feeling conveyed by figurative means, and the conviction that authenticity demands zero excess, in the 'less is more' tradition. Where these conditions are met the sparest of patterns can express Zen simplicity, as in the late works of Ad Reinhardt; and Eva Hesse's crumpled cylinders, ragged sheets of plastic and wires wrapped in lumps can balance finely between a buoyant absurdity and pathos. Construed in this way, the game of viewing art can become one of sensitizing oneself to the merest or most idiosyncratic of signifiers.

From its inception the literature of abstraction has made much of abstract works of art conveying, reflecting, exploring, questioning or commenting upon scientific, semantical, aesthetic, or metaphysical concepts and theories. Abstract compositions are said to be creative responses to atomic or other physical theories, mathematical relationships, musical forms, laws of perception and other cognitive processes, unconscious psychological structures, conceptual or categorial truths, or to ideas and issues relating to art itself. Artists are said to engage in research in these domains. Difficulties arise, however, whenever works are presumed to do more than allude in an unspecific way to such referents and projects. To date, the specific content supposedly conveyed or the specific question posed has rarely been divulged. Mondrian, for example, repeatedly declares that art reveals and expresses 'laws of pure plastics' but he never enlightens us as to what these laws are or how we are to derive them from works of art. Nor do propounders of such interpretations ever raise the crucial question of whether a work misconceives or misrepresents its referent. Such reticence obviously casts a shadow over the credibility of these claims. Also, when explicit reference to intellectual content is implied by title, as it is in Georges Vantongerloo's paintings of the 1920s and 1930s (for example, *Composition 15 derived from the equation* $Y = ax + bx + 18$, 1930), doubts arise as to what aesthetic sustenance the viewer can derive from the connection

between design and content. Similarly, Josef Alber's *Homage to the Square* series is tied to his research into colour-interaction phenomena, but the paintings themselves would be ill-served by viewers ascribing artistic merit to them because they exemplify the principles governing such phenomena.

6 The limits of abstract art

The category of abstract art is defined in part by opposition to what is here called classic representationalism. But there is reason to think it should not embrace all nonrepresentational modes. A case in point is art which is 'concrete' in the sense of presenting material objects as such, including, for instance, Marcel Duchamp's unassisted ready-mades. Related examples occur here and there in the multifarious category of conceptual art: Joseph Kosuth's *Titled [Art as Idea as Idea], [water]*, 1966, which consists of an enlarged photocopy of a dictionary definition of 'water'. These works undeniably abstract from representation in that they disavow all fictive appearances, even purely 'optical' ones. But their obtrusive and calculated ordinariness seems to close off all connection to the accepted paradigms of abstract art.

This breach results in significant part from the fact that the ordinariness of the works, taken in context, rebuffs aesthetic contemplation or freely imaginative engagement. Instead of offering a feast for the visual system, the work purports to deliver a message concerning art, the art world or society at large, and one that typically excoriates commodity fetishism, elitism and reverence for art. The mode of signification, on which the work's standing as art depends, resembles that of a rebus or emblem, whose decipherment is most efficaciously accomplished by limiting one's attention to the properties that convey the meaning - convey it, that is, given a detailed context which may include statements or even lectures by the artist. The work cooperates by offering scant aesthetic distraction. Much the same syndrome is found in Robert Morris' quasi-minimalist sculpture. For example, the erect and fallen shafts comprising *Columns* (1961/73) forswear almost all of the modernist sources of sculptural appeal: refined finish, defiance of gravity, balance, precise placement, and so forth.

Perhaps from this analysis a criterion may be devised for distinguishing between minimalist or conceptual works that count as abstract art and those that are best placed elsewhere. Qualifying works would be those inviting fairly comprehensive and sustained aesthetic contemplation. A work that aims mainly at blocking or deflecting such engagement in favour of other ends would fall into one or another adjacent category. Naturally, delicate distinctions are required to apply the criterion. For example, Christo's wrappings may straddle the boundary, since they purport to offer variable but generally quite limited contemplation-worthy fare.

Another boundary question arises over pre-modern works that radically schematize: Cycladic figures, Sepik masks, and countless other 'primitive' images. Accepting such works into the category might seem justified inasmuch as they exemplify the 'urge to abstraction'. On the other hand, the works were certainly not inspired by a conscious rejection of classic representationalism. Further, nothing resembling the play with space fundamental to most twentieth-century abstraction can be plausibly ascribed to them. Thus including them would risk diluting the common ground on which the category was initially based. This issue is rarely discussed in the literature, but obviously deserves attention.

See also: Depiction; Expression, artistic

JOHN H. BROWN

References and further reading

Barr, A.H., Jr (1936) *Cubism and Abstract Art*, New York: The Museum of Modern Art, 1966. (A classic account of varieties of Cubism.)

Cheetham, M. (1991) *The Rhetoric of Purity: Essentialist Theory and the Advent of Abstract Painting*, Cambridge: Cambridge University Press.(Detailed analysis of philosophical theories influencing Gauguin, Kandinsky and Mondrian.)

Courbet, G. (1861) 'Letter in the *Courrier du dimanche*, December 25 1861', in L. Nochlin (ed.) *Realism and Tradition in Art, 1848-1900: Sources and Documents*, Englewood Cliffs, NJ: Prentice Hall, 1966, 34-6.(Gives Courbet's criticism of abstraction in visual art referred to in §1.)

Danto, A.C. (1981) *The Transfiguration of the Commonplace*, Cambridge, MA: Harvard University Press. (Contains ambitious interpretations of works by abstraction expressionists and their successors, among others; relevant to §5.)

Fried, M. (1965) *Three American Painters*, Boston: Garland Publishing Company. (Impressive formalist analysis of spatial effects in nonfigurative abstraction; relevant to §4.)

Greenberg, C. (1973) *Art and Culture*, London: Thames & Hudson.('Collage' and 'Abstract, Representational, and so forth' deal with the relation of image to surface discussed in §4.)

Kandinsky, W. (1911) *Concerning the Spiritual in Art*, trans. and with intro. by M.T.H. Sadler, New York: Dover, 1977.(An early, highly speculative attempt to provide a theoretical basis for abstraction by one of its principal creators.)

Museum of Modern Art (1984) *The Museum of Modern Art, New York: the History and the Collection*, New York: Harry N. Abrams and The Museum of Modern Art.(Especially useful compendium of illustrations of cited works.)

Osborne, H. (1979) *Abstraction and Artifice in Twentieth-Century Art*, Oxford: Clarendon Press.(The only comprehensive treatment of the subject by a philosopher available in English. Generally sound, provided that one abstracts from the information-theoretic concepts by substituting 'schematic' for 'semantic' and 'nonfigurative' for 'syntactic' [abstraction].)

Pohribny, A. (1979) *Abstract Art*, Oxford: Phaidon.(Rich in illustrations of cited works; also contains a useful historical summary of modes of abstraction through the mid-1970s, except for Cubism.)

Rosenberg, C.M. (1971) 'Cubist Object Treatment: a Perceptual Analysis', *Artforum* 9 (8): 30-36.(An insightful, though somewhat technical analysis of Cubist pictorial devices.)

Schapiro, M. (1978) *Modern Art: 19th and 20th Centuries, Selected Essays*, New York: George Braziller. ('Abstract Art' and 'Mondrian' give probing analyses of the relation of abstract art to its antecedents and social context.)

Walton, K.L. (1990) *Mimesis as Make-Believe: On the Foundations of the Representational Arts*, Cambridge, MA and London: Harvard University Press.(Major study of representation and fictional worlds.)

Whitford, F. (1984) *Understanding Abstract Art*, New York: E.P. Dutton.(A wide-ranging introduction to abstract painting.)

Worringer, W. (1908) *Abstraction and Empathy: A Contribution to the Psychology of Style*, trans. Michael Bullock, Cleveland, OH: World Publishing Company, 1967.(A seminal study of abstraction in traditional art, both fine and decorative.)

Art, definition of

Many of the earliest definitions of art were probably intended to emphasize salient or important features for an audience already familiar with the concept, rather than to analyse the essence possessed by all art works and only by them. Indeed, it has been argued that art could not be defined any more rigorously, since no immutable essence is observable in its instances. But, on the one hand, this view faces difficulties in explaining the unity of the concept - similarities between them, for example, are insufficient to distinguish works of art from other things. And, on the other, it overlooks the attractive possibility that art is to be defined in terms of a relation between the activities of artists, the products that result and the audiences that receive them.

Two types of definition have come to prominence since the 1970s: the functional and procedural. The former regards something as art only if it serves the function for which we have art, usually said to be that of providing aesthetic experience. The latter regards something as art only if it has been baptized as such through an agent's application of the appropriate procedures. In the version where the agent takes their authority from their location within an informal institution, the 'artworld', proceduralism is known as the institutional theory. These definitional strategies are opposed in practice, if not in theory, because the relevant procedures are sometimes used apart from, or to oppose, the alleged function of art; obviously these theories disagree then about whether the outcome is art.

To take account of art's historically changing character a definition might take a recursive form, holding that something is art if it stands in an appropriate relation to previous art works: it is the location of an item within accepted art-making traditions that makes it a work of art. Theories developed in the 1980s have often taken this form. They variously see the crucial relation between the piece and the corpus of accepted works as, for example, a matter of the manner in which it is intended to be regarded, or of a shared style, or of its being forged by a particular kind of narrative.

1 Definitions

If the purpose of a definition of art is to facilitate the unequivocal identification of items as art works, then it should characterize a property, or some combination of properties, displayed by each and every art work and belonging exclusively to art works, that is, a feature or set of features marking all art works and only them. Such a definition is called 'real' or 'essential' (see Definition); it specifies one or more necessary conditions which in the combination indicated are together sufficient for anything to be of the kind in question.

Definitions can serve goals other than that of unequivocal identification and they need be no more rigorous than is required by the chosen purpose. We might look to a definition simply for the sake of knowledge; for instance, in seeking a precise and systematic catalogue of things. We might aim, alternatively, to teach the meaning of a term and will use such definitional methods as are adequate to achieving that end - ostension, enumeration, dictionary meanings, reference to paradigms. We could wish to prescribe a new meaning or use for a term by an act of stipulation, either for a special purpose (in which case the definition is sometimes called 'operational') or in order to change its meaning altogether (as in revisionist definitions). We might be concerned, instead, to characterize a thing's typical features, or the properties that are significant in our use of that thing; these emphases might result in partial definitions drawing attention to non-essential features. As is apparent from this list, sometimes the task of definition is purely descriptive and at others it is regulative; sometimes it is concerned with the way the world is and at others with linguistic practices; sometimes it is controlled by our interests and at others is largely independent of them.

2 Early definitions of art

Many of the famous theories of art offered in the past - Plato's conception of art as *mimesis* (imitation or representation), Tolstoy's view of art as the communication of feeling,Clive Bell's account of art as significant form - fail very obviously when treated as real definitions. If the key notions are construed so broadly that all art works cannot help falling under them, these notions are also bound to cover many things that are not art works. If the central terms are read narrowly, then, they still seem certain to apply to some things that are not art, as well as not applying to some pieces that generally are agreed to be art. It is best to treat these views as recommending

fruitful approaches to art's interpretation, or indicating art's more salient or valuable features, rather than as real definitions. Indeed, this is the spirit in which most were offered. These theories are addressed to an audience already skilled in the identification of art, and take that common understanding for granted.

Is a more rigorous approach to a definition of art possible? Morris Weitz (1956) has famously and influentially argued that art has no fixed essence and, hence, that no real definition of art can be successful. He notes that when we look, we find no property common to all works of art. Art-making is creative and, hence, inevitably defeats the definer's attempt to congeal what is a fluid process. Weitz explains the unity of the concept of art with the idea of a network of 'family resemblances', a notion he adopts fromWittgenstein's *Philosophical Investigations* (see Wittgenstein, L. §§9-11). Works of art are appropriately grouped together in terms of similarities that link them, though there is no single feature or set of qualities shared by all.

Weitz's positive view faces serious difficulties. Similarity could not provide a basis for recognizing the first works of art, since these had no artistic forebears that they might resemble. Nor does the appeal to similarity explain the status of more recent art works. Some art works, such as ready-mades or representational works, more closely resemble things that are not art than they resemble art works; for instance, art films are more like TV 'soaps' and home videos than sculptures. A counter to this objection might insist that only relevant numbers, kinds, or degrees of similarity are significant in establishing the classification of things as art. To enumerate and clarify the types of resemblance that count towards something's being a work of art is to return to art's definition, however, for one would have to specify the set of similarities that are necessary and sufficient for something to count as art. Weitz's reliance on the notion of resemblance does not replace the need for definitions of the type he declared to be impossible.

On the face of it there is no significant property perceptible in all art works. If so, this counts against the kinds of theories Weitz was keen to attack, namely, those proposing that art might be defined in terms of shared aesthetic properties, these being conceived as qualities revealed directly to the senses. But it is not clear thatWeitz has demonstrated the impossibility of defining art, for the relevant properties might be imperceptible. (One cannot distinguish uncles from other males merely by examining their appearances, but this does not show that the idea of an uncle is indefinable.) It is plausible to expect that some complex, imperceptible relation between creators, the things they make and the audience that receives them will lie behind a definition of art. Hence, even if Weitz is correct in claiming that we do not see a property common to all art works, this does not show art to be indefinable.

What of Weitz's further claim - that a real definition of art will be refuted and repudiated by artists' creativity? Again, the claim appears plausible only when directed against definitions holding that art works must possess aesthetic qualities (given a limited set of these). A definition relating artists, their products and audiences might easily accommodate innovative kinds of art, because it emphasizes the context of creation and reception rather than the constitution of the piece involved in this transaction.

I have suggested that neither Weitz's arguments nor the fact that most adults have a secure grasp of the concept shows the irrelevance or impossibility of defining art. Moreover, there is an obvious need for such a definition, since the claim to art-status of many pieces created in the twentieth century is hotly debated. Some artists have deliberately produced works that challenge the border between art and non-art, provoking the question 'But is it art?' If we could define art we would have a means of resolving disputes about 'hard cases' of this sort. And even if the attempt to formulate a correct definition is likely to remain controversial, we might come to a deeper understanding of art and its context through the pursuit of such a definition. While Weitz's arguments have been influential and the impossibility of defining art is still asserted, the number of publications presenting new definitions indicates that reports of the death of the enterprise have been greatly exaggerated.

It might be said that it is not so much for us to *discover* whether things are works of art as a result of applying to them an independent standard captured in a definition but, rather, to *decide* whether they are art. I regard this response as misguided. As an aspect of culture, the nature of art is socially constructed and historically malleable, depending on human interests and judgments. If the nature of art is relative to, and affected by, human concerns and practices, this will be mentioned in an adequate definition; such a definition could play a role in settling the appropriateness of our deciding a particular hard case in one way or another. Even if it is for us to decide whether something is art, it does not follow that that decision can be entirely arbitrary, for there must be a difference between our coining an additional meaning for an old term and our resolving that some controversial case is to be

properly grouped with undisputed paradigms under the same conceptual umbrella.

3 Functionalism and proceduralism

Many definitions offered in recent decades can be classed as functional or procedural. Functional definitions give centrality to the necessary condition that works of art serve a purpose or purposes distinctive to art, whereas procedural definitions stress that they are created according to certain conventions and social practices. A composite definition mentioning both of these necessary conditions, as well as others, is possible. In practice, though, these two kinds of definition oppose each other, because the procedures by which the status of art is usually conferred have been used to create pieces that fail to serve functions traditionally met by art. Indeed, items may be presented as art, though they have as their point the goal of opposing the attempt to appreciate them in the orthodox fashion. Some functionalists offer their definitions with the goal of excluding such pieces from the realm of art, whereas proceduralists aim to include them. These approaches also differ concerning the connection between something counting as art and its having artistic value. Functionalists see the possession of a degree of aesthetic value, measured in terms of an item's success in fulfilling one or more of the functions of art, as essential to its qualifying as art, while proceduralists regard the artistic evaluation of a thing as separable from the determination of its status as art. The proceduralist's definition is purely descriptive, having little to say about the significance of art or about the reasons that might lead someone so authorized to confer art-status on one thing rather than another. By contrast, the functionalist's definition is normative.

Monroe C. Beardsley (1982), a functionalist, characterizes an art work as either an arrangement of conditions intended to be capable of affording an aesthetic experience valuable for its marked aesthetic character, or (incidentally) an arrangement belonging to a class or type of arrangement that is typically intended to have this capacity. A more recent version of functionalism is given by Robert Stecker (1994), according to which an item is a work of art at time *t* if and only if either (a) it is in one of the central art forms at *t* and is intended to fulfil a standard or correctly recognized function within the set of central art forms at *t* or (b) it is an artefact that achieves excellence in fulfilling a function belonging to the set of functions for central art forms (whether or not it is in a central art form and whether or not it was intended to fulfil such a function).

Among the tasks and difficulties faced by functionalist accounts are as follows. (1) Specifying the functions of art. ForBeardsley, the main purpose is that of providing an aesthetic experience. (2) Acknowledging both that the point of art might alter through time and that the art-historical context of creation affects the aesthetic character of the work and, thereby, its functionality. The historicism introduced by Stecker's time-indexing is designed to cover such considerations. (3) Explaining the dysfunctionality of very poor works of art. Both Beardsley and Stecker do so by allowing that something intended to serve the point or points of art might become an art work even if that intention is unsuccessful. (4) Resolving the status of the hard cases mentioned previously.Beardsley denies that Duchamp's pieces are works of art, whereasStecker argues that, within their art-historical setting, they serve accepted functions of art (reference to and rebellion against former artistic types and practices). He notes that they could not have served equivalent functions in earlier times.

The most detailed version of a procedural account is the institutional theory developed byGeorge Dickie. His most recent definition (1984) runs: (a) an artist is a person who participates with understanding in the making of an art work; (b) a work of art is an artefact of a kind created to be presented to an artworld public; (c) a public is a set of persons the members of which are prepared in some degree to understand an object which is presented to them; (d) the artworld is the totality of all artworld systems; (e) an artworld system is a framework for the presentation of a work of art by an artist to an artworld public. The 'artworld' is the historical and social setting constituted by the changing practices and conventions of art, the heritage of works, the intentions of artists, the writings of critics, and so forth.

Among the difficulties faced by the proceduralist are as follows. (1) Showing that the relevant procedures are established (and, in its institutional version, demonstrating that they mark an informal institution distinguishable from similar institutions with different goals). (2) Accounting for the art-status of works never presented to, or intended for, a public, including the products of isolated artists, of the earliest artists in history and of those working outside the officially recognized boundaries of the artworld, such as embroiderers.Dickie's definition requires not that the piece be presented, but that it be of a kind suitable for presentation; also, he could allow that some pieces are enfranchised as art from within the institution after their creation. (3) Avoiding a vicious

circularity in characterizing the procedures, or the institution in which they are applied, without assuming their products to be art works.Dickie claims that the circularity in his own account is benign. (4) Resolving the status of the hard cases mentioned previously. Dickie sees it as an advantage of his theory that it accommodatesDuchamp's ready-mades, but one might wonder if the procedural account is able to explain what makes such cases hard.

4 Recursive definitions

Weitz's suggestion that something is a work of art in virtue of its resemblance to other (prior) works of art indirectly acknowledges the historicist character of art-making. Artists frequently draw on, refer to or react against their predecessors. Moreover, what constitutes art and what can be done within art depends on what has been art and what has been done within art in the past; the art of the distant past of a culture might differ in many respects from the art of its present, despite the continuity of the process that links one to the other. The historicist character of art has received growing recognition within philosophical aesthetics since the 1950s; more recent attempts at a definition reflect this.

In crude outline, a historicist definition of art has two parts. The first explains how the first works in history came to be art - perhaps by stipulation, or because they served an appropriate function. The second, recursive part states that 'Something is an art work if it stands in an appropriate relation to art that predates it.' The 'appropriate relation' is characterized in various ways. A suitably historicized functionalist definition, for example, would construe the relation as holding between the (intended, central, significant) function of the present candidate and the (intended, central, significant) functions of past works. A suitably historicized proceduralist definition would construe the relation as holding between the procedures applied to the present candidate and the procedures used successfully in conferring art-status on prior works. (I have already noted the historicist aspect given to functionalism by Stecker. The institutional theory is ripe for and would be improved by a similar treatment.)

Some recent historicist definitions conceive the defining relation neither in terms of function nor procedure. Jerrold Levinson (1979) sees the defining relation in the intended treatment of the candidate - a work of art is a thing that has been 'seriously intended for regard-as-a-work-of-art'; that is, regard (meaning treatment, taking, engagement with or approach) in any way pre-existing works of art are or were correctly regarded. James Carney (1991) characterizes the defining relation as a shared style: an object is a work of art if and only if it can be linked by those suitably informed, along one or more various specific dimensions, to a past or present general style or styles exhibited by prior works of art. Noël Carroll (1993) takes the unifying relation to be that of narrative continuity, though he denies offering this as a definition. In his view, something is an art work if it can be linked to preceding art-making practices and contexts by a narrative committed to historical accuracy that reveals the piece as an intelligible outcome of recognizable modes of thinking and making of a sort already commonly adjudged to be artistic. If there is dispute about the artistic nature of the context from which the candidate work arose, then this is to be settled by appeal to a meta-narrative that links that context with acknowledged artworld practices, procedures and processes.

The detail of each of these theories might be examined critically. For instance, one might ask if Levinson can distinguish the art-making intention from other intentions that similarly invite a regard of something as if it were art without aiming, directly or indirectly, at making that thing art; and one might consider whether Carney could analyse the notion of artistic style, or Carroll could develop the relevant notion of continuity in narrative, without begging the definitional question. (Of course, a theorist might avoid such queries by further generalizing the recursive part of the definition - something is a work of art if and only if it stands in the appropriate art-creating relation to previous works. This approach meets these objections, though, only by emptying the definition of content.)

Instead of pursuing such matters here I will mention one concern about the general strategy. It seems that there is more than one tradition of art-making and appreciation; also, what is possible at a given time within one tradition might not be possible at the same time, or at any time, in others. Recursive definitions explain how something is art by relating it in the appropriate way to a given tradition. Such definitions will be at best incomplete, because so much of the explanatory burden is carried by the implicit, undefined notion of an artistic tradition. If something is a work of art within only one of many possible traditions, then the notion of art is not fully explicated until a basis is provided for distinguishing traditions of art from other historically continuous, cultural processes or practices and, also, for individuating one artistic tradition from another.

Two ways of attempting to dismiss this point fail, I think. First, it would be both false and offensive to confine art to a single cultural tradition, such as that arising from western Europe, and to dismiss other traditions merely as generating non-art that serves functions similar to those of art. And even if we allow for the many human artistic traditions, it might be implausible to reject the possibility of non-human, non-terrestrial art. Second, it would be an error to suggest that the proposed definition allows that something is a work of art if it relates appropriately to any pieces in *any* tradition of art-making, for the work then becomes decontextualized. This is unconvincing because it implies that, if something could become art within one tradition, it could become art in any; if Duchamp could make a work of art of a urinal in USA, a Chinese artist might have done the same in China. Rather than emphasizing that the art status of a piece depends on the piece's historico-cultural location, this approach treats the place of the piece in its given tradition as irrelevant to its status as art.

See also: Artist's intention; Collingwood, R.G. §3; Croce, B. §2; Tolstoy, L. §4

STEPHEN DAVIES

References and further reading

These readings are not introductory in style, though none is forbiddingly technical or formal.

Beardsley, M.C. (1966) *Aesthetics from Classical Greece to the Present*, New York: Macmillan.(Introductory history of aesthetics covering the main theories and definitions.)

Beardsley, M.C. (1982) 'Redefining Art', in M.J. Wreen and D.M. Callen (eds) *The Aesthetic Point of View*, Ithaca, NY, and London: Cornell University Press, 298-315.(A functionalist definition according to which art provides aesthetic experience.)

Beardsley, M.C. (1983) 'An Aesthetic Definition of Art', in H. Curtler (ed.) *What Is Art?*, New York: Haven, 15-29.

Bell, C. (1958) *Art*, New York: Capricorn Books.(Originally published 1914. Art is defined in terms of significant form.)

Carney, J.D. (1991) 'Style Theory of Art', *The Pacific Philosophical Quarterly* 72: 272-89. (A recursive definition identifying style as the crucial relation.)

Carroll, N. (1988) 'Art, Practice, and Narrative', *The Monist*, 71: 140-56.(Introduces a theory according to which the continuity of a narrative charts the course of art. The account is elaborated and defended in 1993 and 1994.)

Carroll, N. (1993) 'Historical Narratives and the Philosophy of Art', *Journal of Aesthetics and Art Criticism* 51: 313-26.(Elaborates and refines a theory first introduced in 1988 according to which the continuity of a narrative charts the course of art.)

Carroll, N. (1994) 'Identifying Art', in R.J. Yanal (ed.) *Institutions of Art*, University Park, PA: Penn State Press, 3-39.(Advances the theory of 1993 while criticizing alternative views.)

Davies, S. (1991) *Definitions of Art*, Ithaca, NY, and London: Cornell University Press. (Detailed review of philosophical treatments of the topic since the 1950s; contains a bibliography.)

Dickie, G. (1969) 'Defining Art', *The American Philosophical Quarterly*, 6: 253-6. (Dickie's first formulation of the institutional theory.)

Dickie, G. (1974) *Art and the Aesthetic: An Institutional Analysis*, Ithaca, NY, and London: Cornell University Press.(The first developed account of an institutional theory.)

Dickie, G. (1984) *The Art Circle: A Theory of Art*, New York: Haven.(The institutional theory revised and refined.)

Diffey, T.J. (1969) 'The Republic of Art', *British Journal of Aesthetics* 9: 145-56.(The most sophisticated of the first articulations of the institutional theory.)

Diffey, T.J. (1985) *Tolstoy's 'What Is Art?'*, London: Croom Helm.(An account of Tolstoy's theory of art as expression.)

Levinson, J. (1979) 'Defining Art Historically', *British Journal of Aesthetics* 19: 232-50. (First version of his recursive definition.)

Levinson, J. (1989) 'Refining Art Historically', *Journal of Aesthetics and Art Criticism* 47: 21-33. (Revises and defends Levinson (1979).)

Levinson, J. (1993) 'Extending Art Historically', *Journal of Aesthetics and Art Criticism* 51: 411-23. (Revises and defends Levinson (1979) and (1989).)

Schlesinger, G. (1979) 'Aesthetic Experience and the Definition of Art', *The British Journal of Aesthetics*, 19: 167-76.(Works of art defined functionally as artifacts that produce aesthetic experience.)

Sparshott, F. (1982) *The Theory of the Arts*, Princeton, NJ: Princeton University Press. (Detailed coverage of the

arts, including their definition.)
Stecker, R. (1994) 'Historical Functionalism or the Four Factor Theory', *British Journal of Aesthetics* 34: 255-65. (A historicist, sophisticated version of functionalism.)
Tolhurst, W. (1984) 'Towards an Aesthetic Account of the Nature of Art', *Journal of Aesthetics and Art Criticism,* 42: 261-9.(A historicist version of functionalism.)
Tolstoy, L. (1930) *What is Art? and Essays on Art*, trans. A. Maude, Oxford: Oxford University Press.(The title essay, which is philosophically naive, was originally published in 1898. Defines art as the communication and arousal of feeling.)
Weitz, M. (1956) 'The Role of Theory in Aesthetics', *Journal of Aesthetics and Art Criticism* 15: 27-35.(A famous attempt to show that art has no definable essence.)

Art, performing

Some works, such as plays and pieces of classical music, are created as instructions (either notated or implicit in an exemplar) for performers; performances of such pieces arise from the appropriate execution of those instructions. Because the instructions do not specify all features possessed by an accurate performance, performers inevitably contribute something to the performance; even ideally accurate performances differ in the interpretations they offer. Some such works serve primarily to highlight the performer's talents. Even where this is not so, some awareness of what is involved in rendering a piece is necessary to appreciate a work written for performance, since the skills and techniques of performance are the artistic media through which the work's contents are presented. Performances are evaluated for the life, integrity and interest of their interpretations, as well as for their accuracy. The desirability of one performance over another relates partly to the knowledge and experience of the intended audience.

Other works, such as films, involve performance in their creation rather than in their transmission. If these works are multiple, they are so because copies are cloned from a master. When completed, such pieces are not performed or interpreted; they are shown or displayed.

Free improvisation might stand as performance in its own right, being neither the creation of a work nor an instance of one. The criteria in terms of which improvisations are evaluated differ from those involved in the creation or transmission of works, taking into account the fact that the improviser's efforts involve the risks, as well as the delights, of spontaneity.

1 Arts involving performance

The performing arts include theatre, dance and other forms of movement, opera, film and television, as well as instrumental and vocal music. Some performing arts are now lost - sagas, narrative poetry and storytelling, for instance; others are marginal as performing arts - poetry, for example. Some activities that include performance are not counted among the arts, such as circus, striptease and televised advertisements.

2 Works for performance

In the case of works created for performance, artists (such as playwrights, composers, choreographers and authors) produce either instructions (in oral form, or in the form of a script or score) that performers execute or a model instance that they emulate in delivering instances of the works. A work for performance is complete when its score, script or model is complete. Such pieces are instantiated in their various performances. The works are distinct from their performances; they have been variously characterized as universals, types, kinds or classes (see Art works, ontology of §2). Pieces represented by a score or script might never be performed; that is, they might have no instances. Theorists sometimes distinguish performance for an audience from rehearsal, practising or private enactments. What should be acknowledged is that the activity takes its point from generating and publicly transmitting instances of given works.

In the case of works for performance, artists' instructions should be interpreted in light of the relevant performance practices; what can be presupposed might not be mentioned or notated, despite its being required, and not all that is mentioned or notated will be mandatory. To perform a work, performers typically produce an instance of the work by following the artist's instructions or model, or by copying other performances derived from those. (If a gust of wind by chance produces a sound acoustically indistinguishable from some performance of a Beethoven symphony, that sound-event would neither be a performance nor would it otherwise instantiate Beethoven's work.) A performance of a given work that contains some departures from the artist's instructions might still be regarded as a performance of that work, provided the work remains recognizable in the performance. To the extent that it diverges (whether this is intended by the performer or not) from the artist's determinative instructions, a performance is inaccurate or unauthentic as an account of the artist's work.

The relevant conventions and artists' notations underspecify or do not determine many of the features displayed in a performance of the work, and where performers emulate a model instance, the realization of many details is by convention left at their discretion. Where performers are left free, they are constrained only by the wider conventions of style or genre. The level of the performer's creative autonomy varies within and between the

performing arts; jazz dancers have more liberty in their actions than do ballerinas. But, even in those performing arts that provide highly detailed instructions, many crucial choices are left to performers; for instance, many aspects of speech delivery and gesture are not indicated by the playwright. It is by their treatment of those matters on which performers are free that equally faithful (but different) interpretations are distinguished. A single interpretation or production might receive more than one performance.

Though the work is distinct from the score, script or model instance produced by the artist, I hold that the identity of the work derives from the instructions notated or implicit in the model. These, in turn, are identified in relation to the artist and the period of creation. In that case, one cannot perform the work except by performing it as its creator's, and doing this requires faithfulness in the relevant respects (as determined by artistic conventions of the genre at the time) to the instructions produced by the artist. A performance can instantiate a given work only if it is faithful or authentic to the appropriate degree. I claim that we are interested in performances primarily as performances of artists' works and that this is how they are advertised and represented. (This is not to say, though, that we shun works from unknown provenances; a speculative account of a work consistent with the conventions of its time usually provides something worth considering, if not definitively its artist's work.)

The history of its performances and the constitution of the intended audience also are relevant in assessing the degree of faithfulness appropriate for a performance, as Jerrold Levinson has argued 1987). The first performance of a work should aim at a high degree of faithfulness, as should a performance directed at novices. Where works are well known, as are Shakespeare's plays in the West, much of the performer's duty of faithfulness has been discharged already; other desiderata, such as contemporary relevance, novelty and verve, become more prominent. Performances are evaluated in terms of the life, coherence, variety and interest they bring to the work, and these features are relative to the audience's prior experience of that work or similar ones, as well as to the materials furnished by the work's creator.

Stan Godlovitch (1993) specifies the following conditions for the integrity of live performances: only one work is performed at a time; its proper sequence is respected, as is the indicated rate of delivery; the performance is continuous, without unjustified breaks; performers comply with the appropriate roles (and do not, for example, swap parts midway through the work). Also, the audience should be in a position to receive the entire performance in every detail. Note that such conditions are required by the view presented earlier: the primary aim of performance is to deliver the artist's work, as specified, to an audience.

3 The centrality of the performer

That artists need the services of performers in instancing their work is not to say merely that performers are means for the work's transmission. The foundry workers who follow the sculptor's instructions and the film's projectionist help to create or transmit the work, but theirs is not the pivotal role of the performer. They might be replaced by technologically superior alternatives without thereby altering the artistic character of the statue or film, whereas the performer's task is ineliminably part of works created for performance.

Artists work with media; the appreciation of art requires the audience to be aware of the limits and possibilities of the media employed. In the performing arts, the requirements of performance are part of the medium in which artists operate. Artists do not create works that happen to be performed; rather, they write *for* performance, taking into account what will be involved for performers when they produce the outcome. (So it is that a new, though derivative, piece results when a musical work is transcribed for instruments different from those specified for the original.) Just as the viewers of a painting consider not only what it represents, but also its surface and the manner of its representation, so audiences in the performing arts consider the artist's use of the performers and their props or tools. They should be aware, for instance, that a dance depicts the death of a swan, that a given actor is playing the parts of several characters and that organists use their feet as well as their hands.

Many works for performance employ the performer's skills in order to achieve narrative, expressive, formal or other effects. These, rather than the performer's activity, are the proper focus of appreciative concern, even if that concern involves an awareness of the connection between the artist's instructions, the performer's efforts and the artistic result. In other pieces, though, the artistic point of the work is to highlight the expertise and techniques of the performer. This is the case with works providing virtuosic roles for one or two performers; some genres, such as the concerto and étude, are of this kind. As Thomas Carson Mark suggests (1980), such pieces are about the

talents required to perform them; the audience's fullest appreciation requires a recognition of the difficulties overcome by the performer in making the rendition seem effortless.

Combining the points just made with the earlier emphasis on the relation between the work's identity and its means of performance, it is possible to see why the tools, techniques and skills of the performer come to be valued and preserved in their own right, sometimes despite the availability of simpler alternatives. If a new ballet shoe were capable of doubling the height of a dancer's leap, it would not be appropriate for dancers to wear the new shoes for performances of nineteenth-century ballets; even if exaggerated elevation is a desirable feature in such works, the difficulty of achieving that elevation is also part of those works. Similarly, the programmed synthesizer is no substitute for the violin when it comes to performingBach's *Partitas*, even if it exactly reproduces the sound of a violin. It is not surprising, then, that musicians, dancers and actors of the past formed guilds, not only to train novices but also to keep secret the tricks of their crafts. Contemporary performers are expected to maintain the required standards.

The central place of artistic skill and creativity in performances helps to explain the distinction between, on the one hand, the disc jockey or ink-printer and, on the other, the performer. The former might be involved in delivering a work of art to an audience, but their roles lack the particular skills for which artists plan the shapes of their works. Tragedies are written for actors and concertos are written for pianists, but films are not made for projectionists and bronzes are not made for casters. Thus it is sometimes said that performers are artists' collaborators, not their servants.

4 Works made through performance

By contrast with pieces created for performance, some kinds of art involve performance not in instantiating the work but, rather, as an essential element in its creation. This is typically the case in cinema. Unlike a play, a movie is not completed as a work when it is scripted; movies must be made, and performers contribute to this process. Once finished, the cinematic work is screened, not performed. The cinematic art work is the master print, which is multiply instantiated by prints cloned from it. The same is also true for some kinds of music, such as electronic pieces that use tapes of the voice or instruments as their source material; the work is the finished master tape and the copies made from it.

An interesting puzzle is raised by popular music, in which discs derived from the master have a dominant status and 'performances' frequently involve 'lip- and hand-synching' to recordings. (Multi-tracking might preclude genuine live performance.) Is the importance of these recordings a sign that they are definitive of the work? If so, artists' rerecordings of their own compositions result in new works, not merely in new performances. Alternatively, is it that the recording is a model instance that may be performed live? In that case, the recording is the more important, not because it is that work's only instance, but because it sets the standard for the work's subsequent performances. (Or are these questions redundant, because the piece has become the music video, which has superseded the audio tape?)

5 Performance in the absence of works

So far I have concentrated on the connection between performances and art works, but performances might take place in the absence of works, as is sometimes the case with street theatre. Dancers might dance without making an instance of any work, or thespians might act without performing a play or making a film. Where there is no work to be followed, the content of the performance is improvised. While there is much to appreciate that is common to both improvised performances and performances of works, there also are differences to be recognized, as Philip Alperson has noted (1984). Obviously one cannot criticize free improvisers for lack of faithfulness to artists' specifications, since they follow none. (There can be mistakes in improvisation, however, when, for example, conventions of the adopted style are violated.) Spontaneity and inventiveness are valued in improvisation; meanwhile, some looseness of structure and lack of polish are less blameworthy in an improvised performance than in one in which such factors are produced by the artist or result after hours of rehearsal. And where there is common ground for appreciation the basis for evaluations can differ. Both improvisations and works for performance might be enjoyed for their narrative or formal structures, inner harmony or overall beauty, but it is relevant that, in the case of the former, responsibility for the achievement lies solely with the performers, who act freely and do so at the moment of performance.

Why not say that improvisation results in a work created by someone who acts both as performer and artist, even if the piece that is the outcome is not itself *for* performance? The difference between improvisation and the creation of a work through performance does not depend on the number of instances, because an improvisation might be taped and, thereby, duplicated, just as a film might have many prints. The results of improvisation are not more ephemeral in principle than are works created with the help of performers. The basis for the distinction, I suggest, is a matter of convention - we talk of films as works, yet we do not describe sessions of free improvisation by this term. This way of talking implicitly acknowledges differences in the goals of improvising and of creating permanent works through performance.

Why not allow that improvisation results in a model of a work for performance? The answer is as before: improvisation is not conventionally approached as providing a work recipe for others to realize through emulation. An artist might recall and notate a piece that was originally improvised, asJ.S. Bach is thought to have done with his *Musical Offering*, but this shows not that all acts of improvisation simultaneously involve the creation of works for performance but only that a work for performance might be composed through improvisation. It matters that the artist supplies a notation or specifies that the original improvisation be taken as a model performance, for without such indications there is no warrant for regarding the result as the creation or performance of a work.

See also: Dance, aesthetics of §§4, 6; Music, aesthetics of §3

STEPHEN DAVIES

References and further reading

Alperson, P. (1984) ‘On Musical Improvisation’, *Journal of Aesthetics and Art Criticism* 43: 17-30.(Argues, as in §5, that the goals of improvisation differ from those relevant to the performance of works but departs from the above in regarding improvisations as works.)

Carlson, M. (1984) *Theories of the Theatre: A Historical and Criticial Survey from the Greeks to the Present*, Ithaca, NY, and London: Cornell University Press.(A detailed survey of the methods, aims, functions and characteristics of Western theatre from the Greeks to the present.)

Carlson, M. (1984) *Theories of the Theatre: A Historical and Criticial Survey from the Greeks to the Present*, Ithaca, NY, and London: Cornell University Press.(A detailed survey of the methods, aims, functions and characteristics of Western theatre from the Greeks to the present.)

Davies, S. (1987) ‘Authenticity in Musical Performance’, *The British Journal of Aesthetics* 27: 39-50.(Outlines conditions for the authentic performance of works and emphasises their compatibility with performers’ creativity.)

Davies, S. (1991) ‘The Ontology of Musical Works and the Authenticity of their Performances’, *Nous* 25: 21-41. (Covers prevailing views on the ontology of musical works and on the relevance of authenticity in performance; argues for the connectedness of these issues; and includes a bibliography.)

Godlovitch, S. (1990) ‘Music Performance and the Tools of the Trade’, *IYYUN, The Jerusalem Philosophical Quarterly* 39: 321-38.(Argues, as in §3, that the difficulties of performing on musical instruments sometimes are essential, not incidental, features of the works that employ them.)

(1993) ‘The Integrity of Musical Performance’, *Journal of Aesthetics and Art Criticism* 51: 573-87.(Characterizes conditions for the integrity of musical performances as listed in §2.)

Kivy, P. (1988) ‘Live Performances and Dead Composers: On the Ethics of Musical Interpretation’, in J. Dancy, J.M.E. Moravcsik and C.C.W. Taylor (eds) *Human Agency: Language, Duty and Value*, Stanford, CA: Stanford University Press, 219-36.(Argues against alternative views that we have a strong, *prima facie* obligation to follow the composer’s specified intentions.)

Levinson, J. (1987) ‘Evaluating Musical Performance’, *The Journal of Aesthetic Education* 21 (1): 75-88. (Emphasizes, as in §2, that the features desirable in a performance are relative to, among other things, the audience’s context and knowledge.)

Levinson, J. (1990) ‘Authentic Performance and Performance Means’, in *Music, Art, and Metaphysics*, Ithaca, NY, and London: Cornell University Press, 393-408.(Argues, as in §2, that the means of performance are criterial for the identity of works for performance, because the use of the means generates some of the work’s artistically important properties.)

McFee, G. (1992) *Understanding Dance*, New York: Routledge.(Introductory text on the philosophy of dance.)

Mark, T.C. (1980) ‘On Works of Virtuosity’, *The Journal of Philosophy* 77: 28-45.(Argues, as in §3, that virtuosic works are about the virtuosity they display and that performances themselves can be works of art.)

Saltz, D. (1991) ‘How To Do Things On Stage’, *Journal of Aesthetics and Art Criticism* 49: 31-45.(Considers the

extent to which an actor's actions could be the actions of the character portrayed by the actor, with discussion of different models.)

Sparshott, F. (1988) *Off the Ground: First Steps to a Philosophical Consideration of the Dance*, Princeton, NJ: Princeton University Press.(Detailed discussion of the metatheory of dance.)

Thom, P. (1993) *For an Audience: A Philosophy of the Performing Arts*, Philadelphia, PA: Temple University Press.(Attempts to define the performing arts, distinguishing them from nonperformance arts and from nonartistic performance.)

Walton, K.L. (1988) 'The Presentation and Potrayal of Sound Patterns', in J. Dancy, J.M.E. Moravcsik, and C.C.W. Taylor (eds) *Human Agency: Language, Duty and Value*, Stanford, CA: Stanford University Press, 237-57.(Discusses how performances go beyond instantiating the pattern of the work as specified in its notation and how the notion of performance is both functional and normative.)

Art, understanding of

Art engages the understanding in many ways. Thus, confronted with an allegorical painting such asVan Eyk's The Marriage of Arnolfini*, one might want to understand the significance of the objects it depicts. Similarly, confronted with an obscure poem, such asEliot's The Waste Land, one might seek to understand what it means. Sometimes, too, we claim not to understand a work of art, a piece of music, say, when we are unable to derive enjoyment from it because we cannot see how it is organized or hangs together. Sometimes what challenges the understanding goes deeper, as when we ask why some things, including such notorious productions of the avant garde as the urinal exhibited by Marcel Duchamp, are called art at all. Some have also claimed that to understand a work of art we must understand its context. Sometimes the context referred to is that of the particular problems and aims of the individual artist in a certain tradition, as when the church of St Martin-in-the-Fields is understood as a contribution by its architect to the vexing problem of combining a tower with a classical façade. Sometimes the context is social, as when some Marxists argue that works of art can best be understood as reflections of the more or less inadequate economic organizations of the societies that gave rise to them. The understanding of art becomes a philosophical problem because, first, it is sometimes thought that one of the central tasks of interpretation is to understand the meaning of a work. However, recent writers, notably Derrida (1972), query the notion of the meaning of a work as something to be definitively deciphered, and offer the alternative view of interpretation as an unending play with the infinitely varied meanings of the text. Second, a controversial issue has been the extent to which the judgment of works of art can be divorced from an understanding of the circumstances, both individual and cultural, of their making. ThusClive Bell argued that to appreciate a work of art we need nothing more than a knowledge of its colours, shapes and spatial arrangements. Others, ranging from Wittgenstein to Marxists, have for a variety of different reasons argued that a work of art cannot be properly understood and appreciated without some understanding of its relation to the context of its creation, a view famously characterized byBeardsley and Wimsatt (1954) as the 'genetic fallacy'.*

1 Minimal understanding

Some, notably Clive Bell (1914), have argued that in order to appreciate a work of art we need a very minimal understanding of it. We do not need to understand what, say, an allegorical picture represents, or when, how and why it was painted. All we need is knowledge of its form as revealed in its colours, shapes and spatial arrangements. Apart from the unfortunate way in which this encourages appreciators to treat representations of suffering, grief and loss as exercises in aesthetic pleasure-seeking, and apart from the cavalier dismissal of the delight that we take in representation and expression, this theory fails even as an account of the appreciation of form. Sometimes it is only when we understand what is depicted that we can appreciate the formal composition of a work. In Stubbs's painting *The Duke of Richmond's Racehorses at Exercise*, it is only because we recognize a pointing hand leading our eyes in a certain direction that we understand the composition of the painting.

Wollheim (1980) has argued that the very possibility of representational painting entails a reference to the pictorial intentions of artists. Hence, granted that representation is a relevant concern to the appreciator of art, understanding is, on Wollheim's view, related to intention. For seeing a canvas as a representational painting differs from seeing, by an exercise of one's fancy, pictures in a moss-covered wall; the difference, Wollheim claims, being a matter of there being a standard of correctness for representation. We correctly understand a representation when there is a match between what we impute to the canvas and what the artist intended to represent. Both to understand that something is a representation and to understand what it represents we need a reference to intentional activity.

2 Categories

Some, including Kendall Walton (1970), have argued that evaluation and understanding are related, since in order to make a proper evaluation of a work of art it is sometimes necessary to understand to which category of art it belongs, a view also argued by Richard Wollheim, and vehemently denied byCroce in his attack on the notion of artistic genres. This view has some affinity with a discussion in moral philosophy initiated by Peter Geach in his distinction between predicative and attributive terms (1956). Consider the difference between 'grey' (predicative) and 'big' (attributive). 'This is a grey mouse' divides with no oddity into 'This is grey' and 'This is a mouse'. This

being so, we could know that a thing is grey without knowing to what category it belongs. But ‘This is a big mouse’ does not appear so easily to divide into ‘This is big’ and ‘This is a mouse’. The truth of assertions about bigness (unlike those for greyness) seems to be related to different standards of normal size for different categories of things. Similarly some have argued that judgments of merit in aesthetics are relative to categories. What is beautiful as the neck of a horse might not be beautiful as the neck of a Vice Chancellor, and the excellences of sonnets are not those of haiku. A clear case of a need to know the category of a work is the case in which we need to know that a work is ironic or a parody in order to appreciate it properly.

It is not, however, entirely clear that critical as opposed to classificatory judgments of works of art do require understanding of categories. It is legitimate to ask whether what is a good sonnet in the sense in which it meets the requirements of membership of that category is also a good literary work of art - a question that seems to invoke non-specific categories of appraisal in use across the arts.

3 Evaluation and explanation

Many have argued that it is one thing to evaluate a work of art and another to seek an understanding or explanation of its genesis. Beardsley and Wimsatt (1954), for example, claimed that understanding a poem as a prelude to critically appraising it requires only such knowledge of the public language as could be obtained from dictionaries or from any competent speaker. It does not require knowledge of the intentions of the poet to mean something, for the words of the poem belong not to the poet but to the language. This does not rule out historical studies. In the case of a work of some antiquity, as the critical apparatus of most Shakespeare plays will demonstrate, in order to secure understanding we might have to do considerable research to find out what the public meanings of the words were at the time of publication. According to Beardsley andWimsatt, this does not license biographical enquiries, however, since what we are interested in are the public words of a public text, not the private meanings of an author.

To investigate the genesis of the work would indeed be to understand more about it and how it came to be as it is, but this, Beardsley andWimsatt claim, would not be relevant or necessary to its assessment. A similar line was taken by Trotsky, who argued that a work of art is to be judged by its own laws (1924). That done, a Marxist could explain, as Marxists can for any human product, valuable or not, how that artistic product had ultimately arisen from a certain economic sub-structure of a society (see Marx, K.). The understanding of the work that this would yield would not, however, be relevant to judgments of artistic merit. In the next two sections we shall look more closely at such attempts to separate understanding and evaluation.

4 Understanding, truth and morality

Much of the debate about art and understanding is a debate about how much of what might rightly be called an understanding of art is relevant to questions of the evaluation of art. Bell, for example, would not have denied that it is a fact aboutFrith’s Victorian narrative painting *Paddington Station* that it is a representational painting. To understand that is to understand something about the painting. What Bell would have denied was that this understanding had anything to do with the appreciation or value of the work. Bell’s denial of the relevance of an understanding of the representational aspects of a painting, however, seems merely by fiat to eliminate aspects of paintings which people unhesitatingly enjoy and which, asWollheim and others have argued, are highly relevant to aesthetic effects.

We may take as a more promising example of the debates about the relevance of certain sorts of understanding the vigorously controversial issue of the relevance to evaluation of an understanding of the truth and morality of a work of art. Some, including Wilde, have denied that works of art can be true or moral at all. But among those who concede that a work of art might contain truths (as Kafka is said sometimes to have captured a truth about the human condition) and might articulate a moral stance (as Jane Austen is often said to do) there are those who deny that an understanding that a work truly has these aspects has any bearing on its evaluation. As to morality, there is a perfectly good sense in which anyone who missed the fact that a certain moral outlook pervades a novel by Jane Austen has not understood that novel. The question is whether that understanding is involved in the assessment of the work, a question to which F.R. Leavis categorically gave an affirmative answer and to which Croce gave an equally categorical negative reply (see Art and morality §3).

When we come to relevance of an understanding that a work of art articulates a view of life, including a view that can be characterized as morally correct, matters are initially muddier because of the complexities involved in assessing views of life. (How is pessimism to be weighed against optimism?Fielding against Kafka?Jane Austen againstSartre?)

One very important approach, adopted by certain Marxists, relates understanding the point of view of a work intrinsically to its evaluation. Suppose we allow that a work of art can articulate a view of life, and that to understand that work is, in part at least, to understand the view of life that it articulates. But, on one reading, Marxist theory claims to possess a privileged understanding of the objective laws of historical progress. In terms of those laws it is possible for a Marxist to say that such and such a state of society is a defective stage of human organization, to be surpassed in the forward march of history, and, further, to say that anyone endorsing that state of society shows a defective understanding of history. Then the way is open for a Marxist to say that a worldview articulated in a work of art can display a defective understanding of social relations (as some alleged was the case with Dickens's *Hard Times*). It would seem narrowly prescriptive to say, without further argument, that this judgment is irrelevant to an assessment of a work of art. For it attributes to a work a lack of understanding, perceptiveness, and possibly imagination. It certainly treats it as the expression of an inadequate state of social consciousness. But then, it seems that there is at least one account that links understanding a work of art with its evaluative judgment. For to understand the work of art is to understand it as the articulation of a worldview, and to understand that aspect is to open the possibility of assessing the work in terms of the adequacy of the worldview it articulates as well as the adequacy of its articulation of it.

Whether such an account can ever be made to work depends on the truth of the Marxist claim - vigorously contested - that they have a privileged access to the objective laws of history. Even if such a claim were false, it should not be forgotten that part of our understanding of a work may involve an understanding of the view of life articulated in it, and, further, that our reaction to a work is often very much bound up with our feelings, not merely about the quality of the way in which that view is expressed, but also about the view itself (see Art and truth §4). Hence Wittgenstein's frequent comment that he could not understand and engage with some works of art, for example the music ofMahler, as opposed to the works of Brahms, because he could not see the world from that viewpoint (1966).

5 Criticism as retrieval

Many have spoken as if the central task of all our dealings with art is evaluation, and other activities, such as the understanding of the whole context of a work, are irrelevant to criticism and appreciation. The element of truth in this is that we tend to embark upon enquiries into genesis and context after we have made the decision that the work is worth it. We study the origins of *The Waste Land* because, prior to any such study, we found *The Waste Land* rewarding. However, we cannot conclude from the fact that we are prompted to learn more about the circumstances of a work after it has impressed us favourably that information discovered about it subsequently is irrelevant to our assessment. Things that emerge on further enquiry might produce radical alterations in assessments (as when we discover that we were taken in by the excellence of a parody).

Second, on reading a work, we may find things in it that puzzle us. Beardsley and others are right that it is a fact about some works that they are puzzling. It is possible, however, not merely to settle for the knowledge *that* a work is puzzling. That leaves a gap in our understanding which we can plug by seeking reasons for why the work is as it is. The positioning of the figures in Picasso's *Les Demoiselles d'Avignon* will not change when we examine its very many drafts and sketches. But puzzlement as to why they are as they are will be replaced by a better understanding.

Finally, it is too easy to talk as if evaluation in some narrow sense were all there is to the appreciation of art. There is also such a thing as a love of a work of art, which, as is often the case with love, wishes to know all there is to know about the object of love and ultimately to understand it as fully as possible. To the lover of the work nothing about it is ultimately irrelevant. And this is related to the view that Richard Wollheim has defended of criticism as 'retrieval', where that involves:

> the reconstruction of the creative process, where the creative process must in turn be thought of as something not stopping short of, but terminating on, the work of art itself. The creative process reconstructed, or retrieval

complete, the work is then open to understanding.

(1980: 204)

Understanding and appreciation cannot be divorced, if for no other reason than that to understand a work of art may just be to hear, read or look at it with a certain kind of appreciative enjoyment. AsWittgenstein remarks, understanding that a Brahms rhythm has a certain queer quality is inseparable from experiencing that quality in it (1966: 20).

See also: Artist’s intention; Artistic interpretation

COLIN LYAS

References and further reading

Bell, C. (1914) *Art*, London: Chatto & Windus.(A classic statement of the case that only a minimal understanding of art is needed as a precondition for appreciation.)

Beardsley, M.C. and Wimsatt, W.K. (1954) ‘The Intentional Fallacy’, in Wimsatt, W.K. (ed.) *The Verbal Icon*, New York: Noonday.(Seminal statement of the view that the understanding of the work deriving from studies of its genesis is irrelevant to its evaluation.)

Croce, B. (1902) *Estetica come scienza dell’espressione e linguistica generale*, trans. C. Lyas, *The Aesthetic as the Science of Expression and of the Linguistic in General*, Cambridge: Cambridge University Press, 1992. (Referred to in §2 for its trenchant statement of the irrelevance of genre criticism and in §4 for its rejection of moralism in art.)

Derrida, J. (1972) *Margins*, trans. A. Bass, Brighton: Harvester Wheatsheaf, 1982.(Referred to in the introductory remarks as denying that the task of the interpreter is to understand a text by deciphering its meaning. Notoriously obscure writer, though the text cited is one of the more accessible, see esp. 316ff.)

Elliott, R.K. (1972) ‘The Critic and the Lover of Art’, in Mays, W. and Brown, S.C. (eds) *Linguistic Analysis and Phenomenology*, London: Macmillan.(Highly relevant and beautifully written contribution to the discussion of the love of art.)

Geach, P.T. (1956) ‘Good and Evil’, *Analysis* 17: 33-42.(The early discussion of the predicative/attributive distinction.)

Scruton, R. (1974) *Art and Imagination*, London: Methuen.(Argues that understanding might just be having the right sort of experience of a work. Ch. 12 discusses detailed examples.)

Trotsky, L. (1924) *Literature and Revolution*, trans. R. Stransky, London: Redwords, 1991. (Denies the relevance of Marxist understandings of art to its evaluation. Further details on the various Marxist positions mentioned in §4 can be found in the essay by S. Sim ‘Marxism and Aesthetics’, in O. Hanfling (ed.) *Philosophical Aesthetics*, Oxford: Blackwell, 1992.)

Walton, K. (1970) ‘Categories of Art’, *Philosophical Review* 79: 334-67.(Proposes the view that an understanding of the category of a work of art bears on its evaluation.)

Wittgenstein, L. (1966) *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*,ed. C. Barrett, Oxford: Blackwell.(Highly relevant to discussions of the relevance of understanding to appreciation, as are the remarks on various artists: for example, on Shakespeare and Mahler, in L. Wittgenstein (1977), *Culture and Value*, Oxford: Blackwell.)

Wollheim, R. (1980) ‘Criticism as Retrieval’ and ‘Seeing-as, Seeing-in, and Pictorial Representation’, in *Art and Its Objects*, 2nd edn, Cambridge: Cambridge University Press.(*Art and its Objects*, although sometimes difficult, bears importantly on questions of the place of understanding in art appreciation.)

Art, value of

Art has as many kinds of value as there are points of view from which it can be evaluated. Moreover, the benefits of art vary with the role of the participant, for there are benefits that are specific to the creation, the performance and the mere appreciation of art. But in the philosophy of art one value is basic, namely the distinctive value of a work of art, its value as *a work of art, which can be called its 'artistic value'. This value is intrinsic to a work in that it is determined by the intrinsic, rather than the instrumental, value of an informed experience of it, an experience of it in which it is understood. Artistic value is a matter of degree, but it is not a measurable quantity, and whether one work is better than another may be an indeterminate issue. A judgment about a work's artistic value claims validity, rightly or wrongly, not merely for the person who makes the judgment but for everyone. Both David Hume and Immanuel Kant tried to show how such a claim could be well-founded, but their attempts are usually considered failures, and there is no accepted solution to the problem they addressed. Many philosophers have been concerned with the relation between artistic value and other values. The most famous attack on art, founded on its supposed relation to other values, was made by Plato, who claimed that nearly all art has undesirable social consequences and so should be excluded from a decent society. Plato overlooked many possibilities, however, and the question of art's beneficial or harmful influence is a much more complex issue than he recognized.*

1 A multitude of values

There is no such thing as *the* value of art. For works of art can be evaluated from many different points of view and, corresponding to these points of view, they have many different kinds of value: moral, political, social, historical, religious, sentimental or therapeutic, for example. Moreover, for a particular kind of value, whether a work possesses that value, and the degree to which it does so, will often be a relative matter, depending on the kinds of people whose involvement with the work is in question. Furthermore, for any given person this value will be relative to the role they occupy in the creation, performance and appreciation of the work - the contribution of any one of these roles to the value in question being unlikely to coincide with the contribution of any of the others.

2 Artistic value

But although art has many kinds of value, within the theory of art one particular value is fundamental, namely the value that is distinctive of art, a work of art's distinctive value, the value of a work of art *as* a work of art, or, as I shall call it, its artistic value. Just as moral value is the kind of value that moral judgment is concerned with, so artistic value is the concern of artistic judgment. This value is fundamental in an account of the value of art; for whatever other kind of value a work of art may possess, the important question is how this other kind of value is related to its artistic value.

Perhaps it will be thought that there is no such thing as a work's artistic value, or that the identification of such a value must be the expression of a moral or political ideology, rather than the acknowledgement of what should be recognized from any position as the distinctive value of art. I believe that this view is mistaken. Although the identification of a certain value as the distinctive value of art must, given what has happened to the concept of art in this century, be to a certain extent stipulative, the identification should not and does not need to be the product of an ideology. It is a mistake to think that the identification of artistic value must have built into it a commitment to some other kind of value, of which the identification is a mere reflection, or a claim to the supremacy of artistic value over other values. Rather, artistic value can and should be identified in such a way that its own worth, as measured by some other standard, is left entirely open, as a matter for further investigation.

What is needed is the specification of a value that all great works of art possess to a high degree and that an artist, in their role as artist, attempts to endow their work with. Such a value exists and can be specified without prejudging any questions about how important it is or what its relations to other kinds of value may be. Clearly, a work's artistic value is simply how good a work of art it is. Accordingly, artistic value is a matter of degree: works are not just better or worse than other works, but they can be much better or worse, or perhaps only slightly better or worse. The question, therefore, is what kind of value is credited to a work of art when it is valued as a work of art, and what value is withheld if the work is thought not to be a good work of art?

A plausible way of conceiving of artistic value emerges from two considerations, the first concerning the importance of understanding a work for the viability of a verdict on its artistic value that rests on that understanding, the second concerning the way in which works of art matter to us when we value them as art.

The first consideration is that whether there is just one correct and complete understanding of a work or many, a verdict on a work's artistic value is insecure unless it is based on a correct understanding of the work. In other words, the fact that you do not properly understand a work undermines the authority of your evaluation. The second consideration is that when we value a work of art as a work of art we value it on account of what it provides us with *in* the experience of it, rather than for something it achieves *by means of* our interaction with it. Artistic value thereby contrasts with medicinal value, for example. The medicinal value of a drug is determined not by the nature of the experience of taking it but by the beneficial or harmful effects that it has on our health. But a work's artistic value is dependent on nothing other than the nature of the experience involved in the appreciation of it.

Putting these two thoughts together, the natural way to think of the artistic value of a work of art is as the intrinsic value of the kind of experience someone has when they experience the work with understanding. A work is valuable as art to the degree that the experience it offers is valuable, not in virtue of any beneficial effects it might bring about, but in itself. What directly determines the artistic value of a work is not the values that are realised in the various *effects* of the experience, but those that are realised in the experience. A fine work of art is its own reward - it rewards in the very experience of it. The suitability of a work, when experienced with understanding, to reward the reader, spectator or audience, and the nature of the rewards on offer, determine the artistic value of the work. If the work is such that it merits being found intrinsically rewarding to experience with understanding, then the experience it offers is intrinsically valuable and, accordingly, the work is valuable as art.

3 Incommensurability

It does not follow from the fact that artistic value is a matter of degree that each work of art possesses artistic value to a *precise* degree, enabling the construction of an order of rank in which each work is either a precise amount above or below, or at exactly the same level as, any other work. In fact, if one work is better than another, it is never better by a definite amount. Moreover, artistic value does not impose even a unique ordering on works of art, for issues of comparative artistic value are sometimes indeterminate: it is not true that, for any pair of works, either one must be better than the other or they must be exactly as good as each other. Sometimes, but not always, the most that can be said about the comparative ranking of two works, whether of the same art form or of different art forms, is that they are of roughly the same order of merit. For example,Vermeer's *Head of a Girl* is undoubtedly a better work than Murillo's *The Young Beggar*, andMozart's *Symphony in G Minor* (KV 550) is better than Schubert's *Symphony in C Minor*; but if it is conceded that neither the Vermeer nor the Mozart can be said to be superior to the other, the insistence that they must be precisely equal in value imposes an unreal precision on the concept of artistic value. This is not because the works of Vermeer andMozart belong to different arts, for the Mozart is better than the Murillo and the Vermeer better than the Schubert. Rather, the incommensurability of artistic value and the indeterminacy of many issues of the comparative rank of works of art arises from the fact that there are many different kinds of quality that can endow a work with artistic value; there is no common unit that would allow the contribution of different kinds of quality to a work's artistic value to be measured; and even if there were, a quality that in one context constitutes an artistic merit can in other contexts detract from a work's artistic value by being combined in an incongruous manner with other qualities of the work, so that the contribution of a quality to a work's artistic value is not an individual matter but holistic.

4 Hume and Kant

A difficult problem arises concerning the implicit claim of a judgment of artistic value to interpersonal validity. How, if at all, might such a claim be well-founded? Both David Hume and Immanuel Kant attempted to answer this question. Each sought to identify a point of view, the adoption of which is definitive of artistic value. The chosen point of view defines this not merely by imposing on any individual a criterion that determines how *they* should judge, but by introducing a condition that determines how *everyone* should judge. The condition secures this result because, it is claimed, the achievement of this condition is - in virtue of the identical operation in each relevant person of the human faculties involved in the appreciation of artistic value - open to all, or to all relevant

persons, by the adoption of the indicated point of view: from this point of view human emotional response to a particular work is uniform.

Hume thought to reconcile the fact that a work’s artistic value (‘beauty’) is not a mind-independent quality but the projection onto things of the pleasure they induce with the intersubjective character of judgments of artistic value, by the exploitation of a supposed parallel between artistic value and secondary qualities - colour, for instance. Just as, he believed, though colour lies in the eye of the beholder, there is such a thing as an object’s true colour, namely the colour it appears in daylight to a normal human being; so, though artistic value is based in the human mind, a work has a single, true artistic value, determined by the pleasurable or unpleasurable response of a human being of a certain kind (a ‘true judge’), someone of superior discriminatory powers, who interacts with the work in the right way and in the right conditions.

Kant’s fundamental thought was that interpersonal validity of a judgment of artistic value is warranted only if such a judgment is not based on anything that might be idiosyncratic or not common to all other persons. By defining such a judgment as the product of a disinterested pleasure in a work’s form; by construing this pleasure as the experience of the free and harmonious play of the faculties of imagination and understanding, the two faculties that yield perceptual knowledge; and by maintaining the essential uniformity of the operation of these faculties for all people; Kant believed that he had established the validity of the claim to intersubjective agreement demanded by a judgment of artistic value (see Kant, I. §12).

So whereas Hume bases his account on the idea of a person of exceptional powers of discrimination, Kant relies upon what he takes to be common to our powers of perception. But neither Hume’s nor Kant’s solution commands assent. Apart from any other considerations, the reasons that each provides for believing in the (supposedly definitive) uniformity of response postulated by their theories are not compelling; and their attempts to capture the idea of a correct judgment of artistic value in terms of the pleasure that would be experienced by someone who interacts with a work in a certain manner under certain conditions appear not to do justice to the normativeness that is integral to the judgment.

The issue of the intersubjective validity of a judgment of artistic value - whether such a judgment can be valid for all people, and, if so, what validates it - has still not been adequately resolved.

5 Non-artistic values

There are many interesting questions about the relations between artistic value and other kinds of value, especially cognitive value and ethical value. For example: Is there an inherent link between the kind and degree of artistic value of a work and values of other kinds? Do certain other kinds of value *determine* artistic value in the sense that they are essential conditions of it? Are works (or works of a certain kind) with a high degree of artistic value naturally suited to support or enhance certain other kinds of value? Some thinkers have attempted to establish a particularly close connection between artistic value and certain other kinds of value, especially ethical value. For example, Tolstoy defined the activity of art as the transmission of a feeling from artist to audience by means of the artist’s creation of a suitable public vehicle, and he drew the conclusion that the better the feeling transmitted - the most valuable feelings being moral-religious ones - the better the work of art that transmits it. His attempt to moralize artistic value, however, like other such attempts, proceeds by way of a tendentious characterization of the nature of art.

6 Plato’s critique of art

Plato proceeded in an entirely different manner. Rather than offering a definition of art from which a favoured criterion of artistic value can be extracted, he merely held up against various artistic practices and accepted paradigms of good or great art *other* values, by which standards these practices and works were, he claimed, found wanting. This was not an unreasonable procedure. For a justification of the importance of art in human life would be best founded on an account of the importance of works that are good *as art*, and a critique of the value of art must undermine the claims of good and great art to a valuable role in our lives.

But isPlato’s famous dismissal of most forms of art from his ideal state, the Republic, well-founded? His attack on art is directed at all three of the central artistic roles, the artist, the performer and the audience or spectator, and it is based on a number of grounds. I shall sketch just one or two of the principal claims. A central allegation is that

the works of such representational artists as the painter and the poet are twice removed from Platonic Reality, for they are mere representations of things that are only specimens of what is truly real - namely, the archetypes (Plato's Ideas or Forms) of what is represented, which are timeless entities more real than any specimens of them and the objects of the highest knowledge. Furthermore, a picture is only an imitation of the visual appearance of what it depicts - it is designed to present to the beholder an appearance similar to that of its subject - and, accordingly, the painter, as such, is not an expert about the reality whose appearance art imitates: the painter knows only how things look, not how they really are or whether they are well-suited to their natural or intended function. Likewise, poets lack the expertise to perform well in the occupations or roles of the characters they represent in their works. So artists can produce only more or less plausible images of the natural and human world: they possess only the art of imitation, not real knowledge; their works do not express or encourage knowledge of the highest Reality, nor do they display any other knowledge worth having. Hence works of art are cognitively worthless. Now this would not matter so much if works of art were harmless, both cognitively and otherwise. Plato's most serious charge is that they are not. On the contrary, they stimulate and foster the inferior part of the soul at the expense of the superior. Pictures, being mere imitations of visual appearances, can appeal only to the nonrational element of the soul, and an indulgence of an appetite for pictures weakens the superior, rational element by strengthening the inferior. Thus although pictures are delightful, they are not beneficial, but are indeed harmful. Likewise, the other main form of mimetic art, poetry, and especially the art of tragedy, encourages socially undesirable feelings and attitudes in people. It corrupts even the better kind of person by eliciting powerful emotional responses to characters who are not really present but only artistically represented as being present, and in unleashing these emotions threatens to usurp the governing power of reason in the life of the person outside art. Accordingly, the appeal of poetry should be resisted; the ideal state will proscribe it, allowing only models of human excellence, hymns to the gods and praises of good men.

It is clear that Plato's fundamental concern in considering whether art should be welcomed or even admitted into a just society is the social value of works of art, that is to say, their value in promoting or hindering the development of socially desirable characteristics in members of a society. This is underscored by his advocacy of the censorship of poetic misrepresentation of the nature of gods and heroes. For it is not so much falsehood that he objects to, but rather the power of poetry to engender beliefs and attitudes that are not beneficial to society. He is willing even to suppress the poetic expression of the truth if this is necessary to preserve a well-governed society.

Plato's specific criticisms of the social value of representational art are unconvincing. I believe that his misgivings about the effects of the powerful arousal of emotions in response to the thoughts, feelings, actions and fates of dramatic characters were effectively answered by Aristotle in his *Poetics*, and that his other charges have little force (see Aristotle §29). But whatever the truth of this, it is clear that Plato's critique suffers from the generality of his claims about the effects of works of art, which he was in no position to verify. We should in fact reject altogether the question, 'What is the social value of art?' For there are different social values - qualities that it is desirable for members of a society to possess and value in others - and works of art have different social values, not only by being variously beneficial or harmful with respect to the same social value, but by enhancing or weakening different social values, which themselves are of greater or lesser importance to society, to a type of society, or to a specific society. Moreover, whether a work produces a certain social effect - one that it produces in some people, say - and the degree to which it does so, depends on the nature of the individual who responds to it, understands or fails to understand it, is moved or unmoved by it, and how often and in what conditions they interact with it. And what holds for social value holds, *mutatis mutandis*, for other kinds of value in so far as these values are determined by the actual effects of works of art on people, in the short- or long-term. The great variety of works of art, the different ways in which they achieve artistic value, the different temperaments, personalities, histories and capacities of individuals, make any global connections between artistic value and any values not intrinsic to the experience of works unlikely. A more convincing account thanPlato gives of art's tendency or capacity to further or hinder non-artistic values must therefore be more exploratory, less partisan, more various and founded upon both a deeper understanding of human nature and a recognition of individual intellectual and emotional differences that affect the influence of works of art on people.

See also: Art and morality; Art, understanding of; Artistic taste

MALCOLM BUDD

References and further reading

Aristotle (*c* mid 4th century BC) *The Poetics of Aristotle*, trans. and commentary by S. Halliwell, Chapel Hill, NC: University of North Carolina Press, 1987.(Referred to in §6. Primarily an account of the nature of tragedy, and in part a response to Plato's hostile view of tragedy.)

Budd, M. (1995) *Values of Art*, London: Allen Lane, The Penguin Press.(Part 1 contains an account of artistic value and a consideration of the views of Hume and Kant referred to in §4.)

Goldman, A.H. (1995) *Aesthetic Value*, Oxford: Westview Press. (A wide-ranging account of judgments about artistic value advocating a nonrealist view of evaluative aesthetic properties.)

Hegel, G.W.F. (1835) *Aesthetics: Lectures on Fine Art*, trans. T.M. Knox, Oxford: Clarendon Press. (In parts extremely obscure, but an impressive and influential compilation that evaluates works of art by reference to what is claimed to be the highest function of art - the expression in a sensuous medium of the self-awareness of humanity's deepest interests.)

Hume, D. (1757) 'Of the Standard of Taste', in E.F. Miller (ed.) *David Hume, Essays: Moral, Political, and Literary*, Indianapolis, IN: Liberty Classics, 1987.(Referred to in §4. A classic attempt to explain the status of artistic value.)

Janaway, C. (1995) *Images of Excellence: Plato's Critique of the Arts*, Oxford: Clarendon Press.(A comprehensive and penetrating examination of Plato's views about the value of art.)

Kant, I. (1790) *Critique of Aesthetic Judgment*, trans. J.C. Meredith, Oxford: Clarendon Press, 1957.(Referred to in §4. Perhaps the most important attempt to elucidate the nature of a judgment of artistic value. Also tries to establish an intrinsic connection between artistic value and morality.)

Plato (*c*.375 BC) *Republic*, trans. P. Shorey, Cambridge, MA: Loeb Classical Library, Harvard University Press, 1930.(Books 2, 3 and 10 contain Plato's famous critique of the arts, referred to in §6.)

Schiller, J.C.F. (1795) *On the Aesthetic Education of Man: In a Series of Letters*, ed. and trans. E.M. Wilkinson and L.A. Willoughby, Oxford: Clarendon Press, 1982.(Schiller's celebrated argument that the ideal, harmonious human life, perfection of moral character, and a stable society that respects each person's inalienable freedom is achievable only through the aesthetic education offered by art.)

Tolstoy, L. (1898) 'What is Art?', trans. A. Maude, in *Leo Tolstoy, What is Art? and Essays on Art*, London: Oxford University Press, 1959.(Referred to in §5. A famous attempt to define the nature of art and to explain its importance in terms of the quality of the feelings it transmits.)

Artificial intelligence

Artificial intelligence (AI) tries to make computer systems (of various kinds) do what minds can do: interpreting a photograph as depicting a face; offering medical diagnoses; using and translating language; learning to do better next time.

AI has two main aims. One is technological: to build useful tools, which can help humans in activities of various kinds, or perform the activities for them. The other is psychological: to help us understand human (and animal) minds, or even intelligence in general.

Computational psychology uses AI concepts and AI methods in formulating and testing its theories. Mental structures and processes are described in computational terms. Usually, the theories are clarified, and their predictions tested, by running them on a computer program. Whether people perform the equivalent task in the same way is another question, which psychological experiments may help to answer. AI has shown that the human mind is more complex than psychologists had previously assumed, and that introspectively 'simple' achievements - many shared with animals - are even more difficult to mimic artificially than are 'higher' functions such as logic and mathematics.

There are deep theoretical disputes within AI about how best to model intelligence. Classical (symbolic) AI programs consist of formal rules for manipulating formal symbols; these are carried out sequentially, one after the other. Connectionist systems, also called neural networks, perform many simple processes in parallel (simultaneously); most work in a way described not by lists of rules, but by differential equations. Hybrid systems combine aspects of classical and connectionist AI. More recent approaches seek to construct adaptive autonomous agents, whose behaviour is self-directed rather than imposed from outside and which adjust to environmental conditions. Situated robotics builds robots that react directly to environmental cues, instead of following complex internal plans as classical robots do. The programs, neural networks and robots of evolutionary AI are produced not by detailed human design, but by automatic evolution (variation and selection). Artificial life studies the emergence of order and adaptive behaviour in general and is closely related to AI.

Philosophical problems central to AI include the following. Can classical or connectionist AI explain conceptualization and thinking? Can meaning be explained by AI? What sorts of mental representations are there (if any)? Can computers, or non-linguistic animals, have beliefs and desires? Could AI explain consciousness? Might intelligence be better explained by less intellectualistic approaches, based on the model of skills and know-how rather than explicit representation?

1 Historical beginnings

Artificial intelligence (AI) researchers make two assumptions. The first is that intelligent processes can be described by algorithms (sometimes called 'effective procedures'), which are rules where each step is so clear and simple that it can be done automatically, without intelligence. This is an empirical hypothesis, which some critics of AI accept (Searle, in Boden 1990: ch. 3) and others reject (Penrose 1989). The second is that all algorithms can be implemented on some general-purpose computer. This assumption is generally accepted. It is based on the Church-Turing thesis, which states that a universal Turing machine, to which general-purpose computers are approximations, can compute any algorithmically computable function (see Church's thesis; Turing machines).

The best-known types of AI - classical AI and connectionism - share these two assumptions. But they differ in other ways (see §§2-4). Classical AI involves serial, or one-by-one, processing of (sometimes complex) formal instructions, whereas connectionism involves parallel, or simultaneous, processing among many simple units. Classical computation uses programs made up of formal rules to generate, compare and alter explicit symbol structures (see Mind, computational theories of; Language of Thought). Connectionist computation typically uses numerical (statistical) rules to determine the activation within networks of locally interacting units and, in systems that can learn, to alter the firing thresholds of individual units and the (excitatory or inhibitory) 'weights' on their interconnections. The system's 'knowledge' is contained implicitly in the constellation of connection weights (see Connectionism). Some philosophers use 'computation' to apply only to the classical type, first defined by Turing in 1936. AI researchers themselves normally use the term to cover both kinds of information processing.

Despite their differences, both these types of AI started from the same source: a seminal article written in 1943 by McCulloch and Pitts (Boden 1990: ch. 1). This discussion of ‘A Logical Calculus of the Ideas Immanent in Nervous Activity’ integrated three powerful ideas of the early twentieth century: propositional logic, the neuron theory of Charles Sherrington, and Turing computability.

The authors showed that simple combinations of (highly idealized) neurons could act as ‘logic gates’. For instance, a McCulloch-Pitts neuron with two inputs could fire if and only if both inputs were firing (an ‘and’-gate), or if only one input were firing (an ‘or’-gate), or if some specific input were not firing (a ‘not’-gate). Since every truth-function can be expressed with ‘not’ and ‘or’ alone, McCulloch and Pitts were able to show that every function of the propositional calculus is realizable by some neural net; that every net computes a function that is computable by a Turing machine; and that every computable function can be computed by some net. Their work inspired early efforts in both classical and connectionist AI because they appealed to logic and Turing computability, but described the implementation of these notions as a network of abstractly defined ‘neurons’ passing messages to their neighbours.

The neural networks discussed in this entry were extremely simple. For example, any link always had the same amount of influence, whereas most modern connectionist systems allow for continuous changes in the weight of each connection. But the authors’ theoretical ambitions were vast. Perception, reasoning, learning, introspection, motivation, psychopathology and value judgments: all, said McCulloch and Pitts, could in principle be understood in their terms. The whole of psychology would in future consist of the definition of various kinds of nets capable of doing the things minds do - that is, capable of computing the sorts of things which minds compute. Neurophysiology and neuroanatomy would show how networks are implemented in the brain, but psychology would define their logical-computational properties. Their views on the relation between psychology and physiology (or mind and body) anticipated later developments in the philosophy of mind (see Functionalism).

McCulloch and Pitts’ 1943 paper made AI possible in three ways. It influenced von Neumann, in designing the digital computer, to use binary arithmetic and binary logic (see Neumann, J. von). It gave both psychologists and technologists the confidence to model propositional (symbolic) reasoning, as opposed to only arithmetical calculation, on logic-based computers. And it inspired people to start studying the computational properties of various types of neural network. Although classical and connectionist AI are often described as utterly distinct paradigms, research in both these approaches commenced because of this paper.

Early connectionist work was further encouraged by McCulloch and Pitts in a paper of 1947. They pointed out that the brain is a parallel-processing device, not a sequential one. Moreover, it can function acceptably even when some cells misfire or die, or when the input signal is ‘noisy’. The perfect input data assumed within their first paper are, in real life, neither necessary nor often available. The question arises, then, how we (and animals) manage without them. McCulloch and Pitts described a statistical technique, based on differential equations like those of thermodynamics, whereby a parallel-processing system could compute (learn to distinguish) various patterns despite slight variations in the input. These (statistical) ideas were less biologically unrealistic than their earlier (logic-based) discussion. Nevertheless, the 1947 paper was less influential over the next three decades than their earlier work. Only in the 1980s did statistical, parallel-processing models achieve prominence (see §3 below).

2 Classical AI

Classical AI is the best-known type of AI, and is sometimes called traditional AI. It uses sequential programming (do this, then do that), and employs internal representations of lists, semantic networks, arrays and other information-processing structures. These representational structures and their components are interpreted as symbolic representations of propositions and concepts (or beliefs and ideas). Accordingly, this approach is also called symbolic AI.

Most internal representations in classical AI are language-like, being constructed from components each of which has some distinct causal-semantic role (though just which role may vary according to context). Some philosophers, such as Jerry Fodor, explain human mental states, or propositional attitudes, in terms of a hypothetical ‘language of thought’ having logical properties (compositionality, productivity, systematicity) like those exploited in classical AI (see Fodor, J.A.; Language of thought). A ‘toy’ example of one simple type of classical AI program (a production system) might look something like this:

If thirsty
then set goal to drink.
If current goal is drink *and* weather is cool
then set goal to seek kettle.
If current goal is seek kettle *and* not in kitchen
then go to kitchen *and* locate kettle.
If kettle is empty
then fill kettle with water.
If kettle is full
then put kettle on hob *and* heat hob *and* locate teapot.
(and so on)

As this toy example suggests, every action, and every condition for action, has to be explicitly specified. Actions that undo previous actions (such as emptying the kettle you just filled) must be avoided. Some unintended consequences of actions have to be anticipated and tidied up (turn off the hob). Default steps must be specified in case any precondition is not satisfied (hot weather, not thirsty). Goal-subgoal structure must be recognized, and the program must be able to 'pop up' to the top goal-level when the lowest sub-goals have been achieved or abandoned. Moreover (what the toy program does not show), procedures must be provided for carrying out the tests (is it cool, and is the kettle full?) and for executing the lowest-level actions (going to the kitchen, locating and filling the kettle) (see Rationality, practical).

Classical AI modelling is widespread in computational research. It is used to study, for example, problem solving, planning, vision, robotics, learning, natural-language understanding, analogy and the perception and performance of music (Boden 1987, 1988, 1990; Rich and Knight 1991). It is applied also to phenomena often assumed to be intractable for a computational (or even scientific) explanation, such as motivation, emotion and creativity.

Among the advantages of classical AI are its ability to represent hierarchical structure and to provide relatively transparent models (whose workings can be well understood by inspecting the program). A further advantage is that it can define 'strong' (exceptionless) problem constraints. It is sometimes forgotten, especially by proponents of connectionism, that strong problem constraints are often needed. For instance, every sentence must have a noun phrase and a verb phrase; and waltz time in music demands that each bar have exactly three beats. Admittedly, a composer may produce some anomalous bars (for example, having only two beats in the upper voice along with three in the lower); but one cannot keep doing this, or break out into march time, without abandoning the goal of composing a waltz. Nor can one communicate intelligibly if one omits most noun phrases. Given that certain rules are mandatory, an AI system should respect them, not approximate them by blurring them with others.

Although it began no earlier than connectionism, classical AI achieved visible success before parallel-processing models did. The first major successes occurred in the 1950s. The logic theorist and general problem-solver of Newell and Simon introduced 'means-end analysis', wherein a program analyses the problem as a hierarchy of goals and sub-goals (on indefinitely many levels) and chooses the action most likely to reduce the difference between the current state and the desired state (the goal). This method was widely adopted in theorem proving, problem solving and planning. Another early landmark was Samuels' draughts (checkers) player, which played well and even learned to adapt to its opponent's individual style. And early language-using programs used stored English word strings and simple linguistic schemata to conduct 'conversations' in which the human interlocutors were occasionally (if briefly) persuaded that they were interacting with another person. (See Feigenbaum and Feldman 1963.)

By the early 1970s there had been considerable advance. For instance, natural-language processing could now be sensitive to highly complex syntactic structure or to the unspoken assumptions hidden in the semantics underlying the actual words - so that programs could 'answer' questions about things not explicitly mentioned. Machine learning was sometimes achieved through the program's having a model not only of the task domain but also of its own action strategies - which, with experience, it modified. Other advances followed. Various high-level AI programming languages were developed, such as LISP ('list-processing language') and PROLOG ('programming in logic'). And Newell and Simon developed 'production systems', a programming method based on if-then (condition-action) rules: if the condition is satisfied, then the action is taken. The condition may be a complex

conjunction or disjunction, including (sometimes) a statement of the system's current goal; similarly, the action may be complex and/or internal (see the toy AI program above). These developments affected both technological and psychological AI. Production systems, for instance, are the core of most 'expert systems', but were originally proposed as a model of human thinking. (An expert system is an AI program consisting of a set of 'If… then' rules, which can be used to aid human beings in solving specialist problems such as locating oil, planning a travel itinerary or diagnosing a disease.)

3 Classical AI and human thinking

Traditional AI began with the assumption that symbolic logic is a normative model for both human and automated reasoning (see Common-sense reasoning, theories of). This assumption sits well with some forms of reasoning, such as theorem-proving (although even there, human beings sometimes employ non-logical methods, such as imagery and analogy). But most human reasoning is approximate and qualitative. We can understand speech even when it is ungrammatical, heavily accented and partly obscured by noise; and we can recognize imperfect handwriting, shadowy scenes, and perceptual or linguistic analogies of many kinds. Also, we can use world knowledge in solving logical problems and in deciding whether strictly logical reasoning, as opposed to fallible heuristics based on experience, is appropriate in a given case. For example, even teachers of logic can solve a problem concerning rules of postage (reversing a minimum number of envelopes to inspect the postage stamps) more easily than a problem of identical logical form posed in abstract terms. In short, we have common sense (see Rationality of belief; Common-sense reasoning, theories of).

Work in AI has increasingly focused on common-sense reasoning, not least because classical AI systems tend to be 'brittle'. If data are missing or corrupted, a classical AI program may give an absurd answer, or none at all. For example, a story-writing program may allow a character to drown a few feet away from a potential rescuer on the river bank, because the programmer did not explicitly include the information that one is normally able to see whatever is going on in front of one's eyes. Similarly, an expert system might ask whether a particular three-year-old girl has any children, not knowing that pre-pubertal girls cannot conceive. People know many such facts about the world and often give something near the right answer (a good guess) even when they lack some relevant information. So AI research has studied probabilistic reasoning, non-monotonic logic (where not-*p* may turn out to be true even though *p* had been proved earlier), case-based or analogical explanation, deep (causal) reasoning, and common-sense semantic networks and belief systems (see Non-monotonic logic). As these new methods are incorporated into classical AI programs, the brittleness typical of first-generation AI models (for example, the rule-based 'expert systems' used by many commercial and public institutions) should be reduced.

To reduce brittleness, however, is not to avoid it entirely. Some critics believe that classical AI methods will never model everyday human thinking. A leading AI logicist has recently recanted (McDermott, in Boden 1990: ch. 9). Classical AI has been criticized also by philosophers who reject the Platonist assumption that all truths are analysable in terms of formalizable elements (Dreyfus 1979). Some philosophers see connectionism as immune to philosophical critiques of classical AI (see §3 below), while others regard it as acceptable only up to a point (Dreyfus and Dreyfus, in Boden 1990: ch. 13).

For AI to be deeply relevant to theoretical psychology and the philosophy of mind, it need not promise actual replication of all human behaviour. So someone (such as Fodor) who propounds a computational philosophy of mind may, without contradiction, doubt whether AI systems could ever in practice achieve more than a tiny fraction of human behaviour. Nor need AI researchers themselves believe in this possibility. They may allow that certain behaviour cannot, in principle or in practice, be replicated by AI (of any type). Moods, for instance, may be unreplicable in principle, and superb novel-writing may be unachievable in practice.

Even AI workers who do believe that full replication is possible need not accept the Turing test (see Turing, A.M. §3) as their criterion of intelligence. They all allow that a non-mental thing could sometimes fool us into thinking it intelligent; indeed, this happened so often with ELIZA, an early language-using program, that this user illusion is called the 'ELIZA effect'. And some, for example, argue that mere replication of behaviour, without evolutionary descent of the underlying causal mechanisms, would not suffice for true understanding.

One well-known philosophical critic of AI, especially (though not exclusively) in its classical form, is John Searle (see Boden 1990: ch. 3; Chinese room argument). Searle argues that even if an AI model passed the Turing test

(which he thinks is in principle possible), it would not really be thinking, or understanding, anything. Correlatively, a computational psychology could not explain how we can understand: at best, it could explain what we do with meanings once we have them. He argues that programs are defined purely syntactically (by formal rules defined over formal constituents), whereas intentionality - or meaning - involves more than mere syntax. Searle's argument has been challenged in many ways by philosophers and AI scientists, and remains controversial. Most opponents accept his assumption that computation is purely syntactic, disagreeing with him on other grounds. However, this assumption is also controversial (see Boden 1990: ch. 4). The nature of computation - and its connection with meaning - is less clear, and less universally agreed, than Searle supposes (see Language of thought).

More generally, many philosophical critiques - and defences - of AI turn on issues of semantics (not behavioural replication). But philosophical semantics is itself disputed. To speak of 'information processing' is problematic, because the nature of information is controversial (see Semantics, informational). Similar remarks apply to symbol manipulation. Other connected issues include whether meanings are 'in the head' (semantic internalism) or partly constituted by the things to which they refer (externalism) (see Content: wide and narrow), and the relevance (if any) of evolution in establishing meaning (see Semantics, teleological). Even within AI, researchers give differing accounts of semantics (and not all AI workers subscribe to 'strong AI' as Searle defines it). Some, for example, accept a procedural semantics wherein the execution of particular computations, or mini-programs, suffices for particular meanings (see Semantics, conceptual role). Others see causal interaction with the external world (and even evolutionary history) as necessary for meaning. In short, the semantics of AI are controversial both within and outside the field.

4 Connectionism and hybrid systems

Connectionist models are parallel-processing systems, involving mutually interactive computations grounded in local interactions between connected units (see Connectionism). Each individual computation is much simpler than a typical instruction in a classical AI program. Even so, connectionist units and computations (and learning rules, if any) vary significantly.

For instance, the semantic interpretation given to connectionist units by AI researchers differs. (This is true irrespective of philosophers' theories about semantics, and the relation of AI and meaning.) Some connectionist units are described as computing the truth-values of entire propositions; others are intended to code familiar concepts. The representational role of these units is thus comparable to the rules and rule components in the toy AI program in §2 (however, connectionist systems cannot specify action hierarchies, like that outlined above for quenching thirst). Yet other connectionist units stand for detailed subsymbolic micro-features, which are not always expressible in terms of everyday concepts or symbols. Thus an input unit might code for a tiny patch of a highly specific (unnamed) shade of purple.

The symbolic/subsymbolic distinction, so defined, is sometimes thought to distinguish (all) connectionist from (all) classical AI. This is a mistake. The distinction is vague: just which concepts count as everyday concepts? Moreover, not all connectionist systems employ subsymbolic processing, and many classical AI programs, especially in vision and natural-language processing, code for subsymbolic microfeatures. The distinction is better defined in terms of the presence or absence of causally efficacious and semantically evaluable logical constituents. Such (symbolic) constituents typify classical AI. Subsymbolic units (in this sense) have no fixed, context-free interpretation, since their effect on processing varies according to the simultaneous activation of the other units.

In the late 1950s and the 1960s, classical AI progressed faster than connectionism. None the less, some researchers persevered with early forms of connectionist modelling. But in 1969 Minsky - widely acknowledged as a father of classical AI, but also the first person to build a connectionist learning machine - proved surprising limitations on what very simple networks ('perceptrons') could compute (see Minsky and Papert 1969). Although he stated that more complex networks might be more powerful, there was a marked drop in interest in connectionist research.

A few individuals within AI, and some in psychology and physics, worked on connectionist systems during the 1970s and early 1980s. But not until 1986 did connectionism attract substantial attention. In the late 1980s it hit the headlines and attracted attention from philosophers, being widely hailed as a new, all-powerful computing methodology. More accurately, one specific type of connectionism attracted attention (others being largely

ignored): ‘parallel distributed processing’ (PDP). In a distributed connectionist system, a concept is not stored by a single unit (as in localist connectionism). Rather, it is represented by a global pattern of activation spread across the entire network, many different units making some contribution. It may be impossible to assign a specific, context-free interpretation to a given unit. Moreover, no individual unit is either necessary or sufficient for the whole network to represent some particular concept (recognize some particular pattern).

PDP models themselves vary in important ways. Not all employ subsymbolic processing, though most do. Some allow only for on-off variation in the activity of each unit, while others allow for continuous weights defining the unit’s influence. Not all PDP models can learn, though most can. Those which can learn employ various learning rules. The number of units can vary, with significant implications for the type of learning achievable. And although most PDP systems have only ‘feed-forward’ connections, passing from units nearer the input to units nearer the output, some have backward links also (called ‘back-propagation’). In general, PDP systems work by adjusting the simultaneous activity of the constituent units until some global equilibrium is reached (different concepts are represented by different equilibria within the same PDP system). This is achieved not by a sequence of symbolic rules, but by numerically described statistical processes like those of thermodynamics. The changing states of the system are described not as symbol structures (such as lists), but as numerical vectors.

Because of their statistical design, PDP systems are better able than classical models to perform well despite noisy input, and to retrieve an entire memory given only a fragment. Some classical AI programs can do this too, up to a point, if specifically pre-programmed to do so. PDP models, by contrast, ‘naturally’ achieve a plausible end-state given partially conflicting evidence, weighing both strong and weak constraints (and the extent of their mutual coherence). That is, they perform multiple constraint satisfaction, where the information may be partially conflicting and/or missing. And PDP systems with learning rules (which change the activation weights on the connections with experience) can learn -as people can - from being shown a range of examples.

PDP models are unlike real brains in many ways. For instance, the widely used back-propagation algorithm learns in a very un-biological fashion. Less neurophysiologically unrealistic forms of connectionism have been developed, but even these fall far short of neural networks in the brain.

Another drawback of first-generation PDP systems is that, unlike classical AI programs, they cannot model hierarchical structure or sequential processing. Some types of human thinking - many aspects of language and problem-solving, for example - require both these features. Most AI researchers use only classical, or only connectionist, models. This sociological fact has encouraged some philosophers to exaggerate the differences, and the supposed superiority of one approach over the other. However, these AI methods have complementary strengths and weaknesses. Accordingly, there is growing interest in hybrid models, which try to get the best of both worlds.

Various philosophers use connectionist ideas in addressing important philosophical problems. These include symbol grounding, the problem of how words and concepts acquire meaning (Cussins, in Boden 1990: ch. 15); the role of folk psychology in cognitive science (Clark 1989); family likenesses, paradigm cases, and prototypes of concepts (Clark 1993); eliminative materialism (Churchland, in Boden 1990: ch. 14); and scientific explanation (Churchland 1990). Many philosophers see affinities between connectionism and Wittgensteinian views of language, because connectionist representations are not cut and dried like those used in classical AI but allow for borderline cases and for varying degrees of similarity.

5 Situated robotics and anti-representationalism

Classical and connectionist AI share a commitment to internal representation as integral to intelligence. Both these approaches (and most cognitive scientists) posit identifiable data structures ‘in the head’ that are distinguishable from the system’s processing (which is done ‘on’ or ‘with’ them), and that stand for things in the world. This commitment has been (controversially) abandoned by AI work in situated robotics (Boden 1996; Maes 1991). Situated robotics is sometimes termed ‘*nouvelle* AI’ (in contrast to traditional AI), or ‘behaviour-based AI’ (in contrast to AI based on abstract task decomposition). It claims to be more biologically realistic than classical AI. It emphasizes ‘autonomous’ systems specifically adapted to their environment, not general-purpose computers controllable by many different programs. And it builds whole (sensory-motor) systems, rather than decomposing intelligence into distinct tasks (vision, planning, motor action) which then have to be integrated to provide a

functioning robot.

Situated robotics avoids using internal representations of the external (objective) world, although some systems use temporary representations of their own (subjective) place in or actions on their immediate environment. And it uses a bottom-up approach to generate complex behaviour. Because the environment is assumed to be noisy, dynamic and inconvenient, the detailed world-modelling and top-down planning typical of classical AI are rejected. No complex program is involved to decide on, monitor and control the creature's activities. Instead, the control of behaviour flows from the nature of the system itself, in the sense that the system is engineered (not programmed) to respond to environmental triggers in certain ways. By these means, researchers in situated robotics seek to avoid the notorious 'frame problem' (Boden 1990: chaps 7-9). This problem bedevils AI work on robot planning, language understanding and common-sense reasoning. It concerns foresight of the many intended and unintended consequences of action. For instance, if a box is moved across the floor then all its contents move also, whereas the chairs (and table and curtains) do not. A formal representation of action must explicitly allow for *all* the intended effects, or some may not happen. And the action's many unintended effects will be wholly irrelevant only if the agent is very lucky, or very thorough in explicitly anticipating potentially relevant outcomes.

Classical robots rely on planning, done within an internal world model. To avoid the frame problem, they must explicitly anticipate a host of intended consequences and unintended side-effects. Although this exhaustive listing of consequences is feasible in artificially impoverished environments, it is impractical in the real world (the thirst-quenching program in §2 would often lead to disappointment, because of unexpected facts about the house concerned). Moreover, because the real world cannot be relied on to remain unchanged, detailed anticipatory plans may fail on execution. Instead of manipulating complex internal representations of the world, situated robots deal directly with it. Situated roboticists eschew the abstract functional task-decomposition employed by classical AI. Instead, they analyse intelligence in terms of 'behaviours'. Their robots engage in simple, hardwired behaviour triggered by specific, ecologically relevant, environmental cues.

The anti-representationalist stance of *nouvelle* AI is controversial. Some situated robots use temporary representations that are not objective but 'deictic' (subject-centred), being closely bound to the robot's behaviour in *this* place on *this* occasion (see Content, indexical). Admittedly, some situated robots - including some of those which show apparently 'cooperative' behaviour, like that of some social insects - do not employ representations even in this sense. Instead, they are engineered to respond directly to certain conditions, without the central world-models favoured by classical AI. Nevertheless, whether human, or even much animal, intelligence can be modelled without internal representations of any kind is unclear.

Evolutionary AI uses genetic algorithms (GAs) to improve programs by artificial evolution. Evolution involves replication, random variation, and selection. GAs make random changes within a program, and (at each generation) identify the most efficient of the resulting rules. These rules are then assigned a high probability of 'breeding' the next generation. Eventually, a system evolves which is efficient, perhaps even optimal. This technique is widely used within classical AI, for inductive problem-solving and classification (Holland *et al*. 1986) and for evolutionary art - GA-directed computer graphics, or music - where the selection at each generation can be done by the human user (Todd and Latham 1992). GAs are used also for evolutionary robotics (in which the robot's design is automatically evolved), and for other work in artificial life, which studies self-organizing, self-replicating adaptive systems (Boden 1996).

6 The neo-Heideggerian challenge

Most AI researchers presuppose the Cartesian distinction between the mind as psychological subject and the real world as object of the mind's (representational) thinking (see Descartes, R.; Dualism). On this view, intelligence is located in the brain, for perceptual input from the world is followed by inner thinking, which in turn causes motor action. Organisms (and artefacts) are assumed to be located within, and adapted to, an environment whose properties are independent of any perceptions of it. Different creatures perceive and affect different world properties. But those properties are objectively there, independent of creatures' relations to them.

These neo-Cartesian assumptions have been questioned by philosophers drawing on Wittgenstein, Heidegger and phenomenology (see Dreyfus 1979; van Gelder 1995). Similar questions have been raised within AI and cognitive science. For instance, some work in situated robotics and artificial life offers an anti-Cartesian critique of

mainstream (classical and connectionist) AI. Such critics reject the subject-object distinction, theories positing internal representations, and the notion of intelligence as a causal series (perception, then thought, then action). Instead, they see organisms as dynamic systems closely coupled with their environment, which (so they argue) cannot be independently defined (see Port and van Gelder 1995; van Gelder 1995; Varela *et al*. 1991; Wheeler, in Boden 1996; Content: wide and narrow). 'Coupling' is a relation of simultaneous mutual causation whereby two systems constantly interact and interpenetrate, each causing changes not only in the other's superficial behaviour but also in its background parameters. If this interpenetration were wholly chaotic, we could not identify two systems. We can (approximately) distinguish systems, and subsystems, only when certain states are relatively stable through time.

Intelligence, on this view, is a function not of the brain but of the whole system constituted by nervous system, body and environment. Likewise, depth vision is a function not of a tiny region of cells, but of the whole visual cortex - indeed, of the entire brain. Computationalism, though not computer modelling, is rejected. It is replaced by dynamic systems theory drawn from physics, which employs differential equations rather than discrete algorithms, and focuses on intrinsic physical states (described by numerical vectors) rather than symbolic representations. (Strictly, since a dynamic system is one whose changes of state can be described by some rule, classical AI programs are a special case; however, real time and/or continuous variables do not feature in classical - or most connectionist - AI, whereas they do feature in the dynamic theories of physics.)

This anti-Cartesian challenge is highly controversial, both scientifically and philosophically. In scientific terms, dynamic systems theory (like first-generation connectionism) is largely a promissory note. Certainly, organisms are dynamic systems - but we need to know how this explains the specific phenomena of human and animal intelligence. Such explanations will require concepts at a higher level than dynamic systems in general.

MARGARET A. BODEN

References and further reading

Boden, M.A. (1987) *Artificial Intelligence and Natural Man*, London: MIT Press, 2nd edn, expanded.(A non-technical introduction to AI, including its philosophical, psychological and social implications; extensive bibliography.)

Boden, M.A. (1988) *Computer Models of Mind: Computational Approaches in Theoretical Psychology*, Cambridge: Cambridge University Press.(A textbook on computational psychology; extensive bibliography.)

Boden, M.A. (ed.) (1990) *The Philosophy of Artificial Intelligence*, Oxford: Oxford University Press.(Papers on the main philosophical-methodological disputes within AI, with a bibliography.)

Boden, M.A. (ed.) (1996) *The Philosophy of Artificial Life*, Oxford: Oxford University Press.(Papers on the philosophy of artificial life, with a bibliography.)

Churchland, P.M. (1990) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: MIT Press.(A defence of Churchland's approach to connectionism, with applications to various philosophical problems including the philosophy of science.)

Clark, A.J. (1989) *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*, Cambridge, MA: MIT Press.(An accessible introduction to connectionism with a discussion of its relation to classical AI and of the role of folk psychology in cognitive science.)

Clark, A.J. (1993) *Associative Engines: Connectionism, Concepts, and Representational Change*, Cambridge, MA: MIT Press.(A more advanced discussion of connectionism and its application to various problems in philosophy and psychology; many references.)

Cliff, D., Harvey, I. and Husbands, P. (1993) 'Explorations in Evolutionary Robotics', *Adaptive Behavior* 2: 73-110.(A survey of work in evolutionary robotics discussed in §5 above.)

Dreyfus, H.L. (1979) *What Computers Can't Do: The Limits of Artificial Intelligence*, New York: Harper & Row, 2nd edn.(A philosophical critique of the foundations of classical AI, in relation to Platonic, Cartesian and Continental philosophy.)

Feigenbaum, E.A. and Feldman, J. (eds) (1963) *Computers and Thought*, New York: McGraw-Hill.(A collection of classic papers in AI, with an extensive bibliography.)

Gelder, T. van (1995) 'What is Cognition, if not Computation?', *Journal of Philosophy* 91.(A defence of dynamical systems, as opposed to computation, as the basis of mental processing.)

Holland, J.H., Holyoak, K.J., Nisbet, R.E. and Thagard, P.R. (1986) *Induction: Processes of Inference, Learning, and Discovery*, Cambridge, MA: MIT Press.(Describes how genetic algorithms work, and how they

have been applied to various problems, including some of philosophical interest. Fairly difficult.)

Maes, P. (ed.) (1991) *Designing Autonomous Agents*, Cambridge, MA: MIT Press.(A collection of articles on situated robotics, with many references.)

Minsky, M.L. and Papert, S. (1969) *Perceptrons: An Introduction to Computational Geometry*, Cambridge, MA: MIT Press.(Limitations on what very simple networks ('perceptrons') can compute.)

Penrose, R. (1989) *The Emperor's New Mind*, Oxford: Oxford University Press.(An attack on the notion that all human thought can be described by algorithms.)

Port, R. and Gelder, T. van (eds) (1995) *Mind as Motion: Dynamics, Behavior and Cognition*, Cambridge, MA: MIT Press.(A collection of papers illustrating the dynamic-systems approach in cognitive science, and its critique of representation.)

Rich, E. and Knight, K. (1991) *Artificial Intelligence*, New York: McGraw-Hill, 2nd edn.(A comprehensive textbook of AI, including detailed descriptions of various AI methods; good bibliography.)

Todd, S. and Latham, W. (1992) *Evolutionary Art and Computers*, London: Academic Press.(A detailed description of the use of genetic algorithms by a professional artist to produce 'families' of three-dimensional computer sculptures.)

Varela, F.J., Thompson, E. and Rosch, E. (1991) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.(A defence of 'embodiment' and 'embeddedness' in anti-Cartesian cognitive science.)

Artistic expression

Many kinds of psychological state can be expressed in or by works of art. But it is the artistic expression of emotion that has figured most prominently in philosophical discussions of art. Emotion is expressed in pictorial, literary and other representational works of art by the characters who are depicted or in other ways presented in the works. We often identify the emotions of such characters in much the same way as we ordinarily identify the emotions of others, but we might also have special knowledge of a character's emotional state, through direct access to their thoughts, for instance.

A central case of the expression of emotion by works of art is the expression of emotion by a purely musical work. What is the source of the emotion expressed by a piece of music? While art engages its audience, often calling forth an emotional response, its expressiveness does not consist in this power. It is not because an art work tends to make us feel sad, for instance, that we call it sad; rather, we react as we do because sadness is present in it. And while artists usually contrive the expressiveness of their art works, sometimes expressing their own emotions in doing so, their success in the former activity does not depend on their doing the latter. Moreover, the expressiveness achieved has an immediacy and transparency, like that of genuine tears, apparently at odds with this sophisticated, controlled form of self-expression. It is because art presents emotion with simple directness that it can be a vehicle for self-expression, not vice versa. But if emotions are the experiences of sentient beings, to whom do those expressed in art belong if not to the artist or audience? Perhaps they are those of a fictional persona. We may imagine personae who undergo the emotions expressed in art, but it is not plain that we must do so to become aware of that expressiveness, for it is arguable that art works present appearances of emotions, as do masks, willow trees and the like, rather than outward signs of occurrent feelings. Expressiveness is valuable because it helps us to understand emotions in general while contributing to the formation of an aesthetically satisfying whole.

1 The expression of emotion

Thoughts and attitudes can be expressed. My concern here, though, is with artistic expressions of emotions, feelings and moods. I shall consider what and whose emotions are communicated in art, and the nature of expressiveness in art works.

Sometimes a person's expressions are distinguished from their dispassionate reports of their emotions (because the emotion is not directly present in the utterance). Also, expressions might be separated from uncontrolled ventings of emotions, these latter being regarded as symptoms, like the spots of measles, that betray or symptomize the condition without expressing it. My own approach is more liberal. I count as an expression any behaviour or display that communicates the agent's emotion, feeling or mood. Such instances of behaviour might be unintended and unthinking, or deliberate and self-conscious. (Indeed, their expressive character might depend on their being the one rather than the other. If my weeping is deliberate and controlled, this suggests pretence rather than expression; if my behaviour is unintended, then it cannot involve the use of social conventions for expression, even if it seems to match these.)

Typically, emotions depend on causal circumstances, take intentional objects and involve beliefs and desires (or make-beliefs and make-desires) concerning that object. For example, I hope for peace at a time of conflict because a treaty has been signed and because I believe treaties lead to the cessation of hostilities, which is what I desire. A person's emotion might be apparent to another who possesses knowledge of any suitable combination of these elements.

In some cases, a person's nonverbal actions alone will indicate that they feel an emotion. In fewer cases - those in which an emotion has an unambiguous mode of nonverbal expression - actions alone might indicate that a particular emotion is experienced. (Perhaps only the broadest categories for happiness and sadness have patterns of behavioural expression sufficiently distinctive for this to be the case; cognitively complex emotions, such as hope or jealousy, have many behavioural expressions none of which is distinctive.) More often, behaviour expresses the agent's emotions only where the wider context is known.

There are further possibilities for the communication of emotion: one can learn of a person's emotions from true

descriptions of them given by knowledgable third parties, or from their own sincere reports. If emotions can be individuated solely by their sensational character and the dynamic structure of their phenomenology, one’s knowledge of the detail of a person’s ‘internal’ experience could communicate their emotions. Finally, note that the expression of emotion has a social, arbitrarily conventional dimension. In some cultures, for instance, the wearing of black clothes and veils is an expression of grief or respect for the dead. The relevant conventions must be followed deliberately and sincerely if the resulting actions are to express an emotion the person feels.

2 Characters in works

If the work contains characters (for example, through depiction or description), then these characters might experience emotions to which their behaviour or circumstances give expression. Unless the audience is given reason in the work’s contents, its genre, or the context of presentation to make-believe otherwise, it is to assume that the beliefs, behaviour, bodily attitudes and causal circumstances of the work’s world correspond to those of the actual world. Accordingly, the audience can learn what emotions the work’s characters experience in the same manner as it recognizes the emotions communicated by others in the ordinary world, except that the audience’s relation to the world of the work depends on make-believe rather than belief.

Some differences are worth noting, however. In the case of narratives written in the first person, the audience might come to know ‘from the inside’ what a character experiences or believes, and hence what they feel, even if that feeling is not outwardly indicated. Second, the protagonists might be non-human or unreal concoctions, such as elephants or intelligent ants. In considering the emotions of such creations, information about their point of view will be relevant - their cognitive commitments and values, their vulnerabilities and aspirations, their intellect, physiology and the like. In addition, artists create expressive contexts that do not or could not arise in the actual world. For instance, the use of leitmotiv in opera to recall actions or words might reveal that a character’s passion is meant for X despite being directed at Y. Quotation and reference, both within and between works, might establish an expressive ambience one would not normally find or look for.

In addition to the emotions of their characters, art works seem to embody and express emotions of their own. This applies to all kinds of works but is perhaps most striking in abstract pieces, including pure music, where expressiveness is present in the absence of a narrated or depicted content. Whose emotions are expressed thereby and how are they expressed?

3 Arousal theory

One suggested answer to the above question is that we ascribe emotions to art works just because those emotions are awakened in us. This is the theory of emotivism or arousalism. Two cases need to be distinguished. In the first, the art work or some aspect of it is the emotional object of a response in the standard way. As a result of realizing that a character in a work is dying unloved one feels sadness for and pity towards that character. Or, believing that the dramatic potential of the last act was botched by the playwright, one feels disappointed by the play. Or one is delighted by the felicity of a turn in the melody. In the second case, one tends to respond with sadness to works called sad, or with happiness to works said to express happiness. Something in the work calls forth the reaction that mirrors the work’s expressive character. One does not then feel sad or happy *about* the work; indeed, the response seems to lack an emotional object, though the work is its perceptual object and cause.

Arousalism refers to this second kind of response in analysing artistic expressiveness: an art work expresses an emotion if and only if it has the power to arouse or tends to arouse that emotion without an object in an appropriate audience. In this view the sadness of the art work is like the greenness of grass; some property of the thing in question disposes it to affect the experience of those perceiving it. The sadness is attributable to the art work, not to the person in whom the feeling is aroused, because the work has the power to awaken the same feelings in a variety of suitably qualified perceivers. Similarly, we say it is the grass that is green, not the perceiving of it, just because its effect on perceivers is largely indifferent to their individuality and idiosyncrasies.

The arousal theory faces two main lines of objection. The first, pursued by Peter Kivy (1989), denies that there are any objectless responses of the kind described - sad music, for instance, never leads listeners to feel sad. A more plausible objection denies the match (postulated by arousalism) between artistic expressiveness and the audience’s tendency to respond. The audience might be unmoved, or might not feel what the work expresses,

despite their correctly recognizing its expressive character. In reply to this, the arousalist points out that a tendency to respond can be blocked or inhibited - for example, where one is distracted from or overexposed to the given piece, and so on. In some cases contemporary values and sentiments might permanently block the tendency that would have triggered a response at the time of the work’s creation. Though it might deal with some counterexamples, I am doubtful that arousalism successfully accounts for all the mismatches between artistic expressiveness and audience’s reactive tendencies, where these threaten its plausibility.

One can reject the arousalist account of artistic expressiveness while accepting that sad works sometimes evoke sad reactions. One might explain the echoing response as occasioned by the work’s expressiveness. Whereas arousalism holds that art works are sad because they make us feel sad, one might instead maintain that it is because they are sad that we respond as we do. We find the emotional moods of others contagious, even if we are not aware of having anything to be happy or sad about; perhaps we react to art works similarly. And perhaps we are open to this mode of response because we approach art as human communication.

4 Expression theory

In creating their works, do artists express feelings? Surely this is often so. In that case, are the emotions expressed in art works those of their artists? We approach many works, including abstract ones, as dealing not merely with the affective side of life but with personal feelings. One view, the expression theory, asserts that expressiveness can be attributed to art works only where there is this discharge of feeling, and because of it: art works are expressive because they stand in relation to artists’ occurrent emotions as do tears to sadness, as both arising from and revealing the feeling. Just as emotions are presented immediately and transparently in genuine tears, so that no inference from crying to sadness is required, we experience the expressiveness of art as residing in it. Also, we find the expressiveness of art works highly evocative of sharing or empathic reactions and this is how we respond to open, primary displays of emotion.

Despite its attractions, the expression theory seems to fail by entailing that when an art work expresses an emotion, the artist experienced that emotion. This generalization is patently false. The process in which art works have their genesis allows little scope for unthinking expression or for undergoing emotions powerful enough to produce the outcome as described. Moreover, some artists turn to creation to escape their traumatic circumstances and, in doing so, produce works that do not reveal the emotions dominating their lives at the time. The expressiveness of art works is usually achieved by their artists, but this happens typically by design. So structured and conventionalized is art, and so practical is the knowledge brought to its creation, that the making of art, even of an expressive variety, cannot continue long or far without reflection, including attention to technique, detail, the nature of the medium and overall structure. Besides, art works are not the kinds of things that arise causally or naturally as immediate, transparent expressions of occurrent emotions. A tendency to create art, unlike a tendency to tears, is not an essential part of sadness, so art works should not be the kind of thing from which sadness can be read directly. In a few cases an artistic action transfers its character to the product that results - violently produced brushstrokes often display the energy that went into their making. But in general, artists’ creative acts, even where these are impelled by emotions, are not such as to transfer that expressive character directly to the resulting piece.

The theory fares no better if performers (should the work have them) are substituted for artists, or where the approach is counterfactual in suggesting that a piece expresses such and such if it is the kind of art work that a person feeling such and such would create. The first alternative encounters objections like those confronting the original theory. The second presupposes art’s expressiveness, without analysing it; one could recognize the work’s aptness for expression only if it already independently displays the appropriate character.

Now, how could it be that art works display expressive directness while expressing the artist’s feeling if they do not relate to that feeling as tears relate to sadness? One way this could be achieved would be by the appropriation of something that itself possesses or simulates the immediate, primary presentation of feeling. For instance, a grieving person might employ professional mourners to weep on their behalf, or might show how they feel by deliberately putting on a sad face or by pointing to a mask of tragedy. Artists, in a similar fashion, might express their feeling through those of characters in the work or by matching the expressive tone of the work to their feelings. But in either case, the expressiveness present in the work has its character independently of the artist’s use of it, so expressions of artists’ feelings through the works they create presuppose, rather than explain, the expressiveness of those art works.

One version of the expression theory that has been influential in the recent history of aesthetics is that propounded by Benedetto Croce §2 and R.G. Collingwood §3. In outline, their account is this: the process of artistic creation is one in which, through the articulation of inchoate feelings and impulses, the artist comes to express a particular, unique emotion, thereby bringing it to their conscious awareness. The emotion is constituted through the act of expression, having no prior identity; that is, the emotion achieves its particular character through the manner of its expression. Collingwood regards art as expression at the level of imagination; for Croce, art is intuitive expression. Both tend to dismiss creation that does not satisfy this model as not truly artistic.

The impression that this view offers an account of the psychology of the creative process and not of the nature of the art work is misleading, for the work is regarded as primarily mental, as existing in the artist's mind and inseparable from the act of expression. The artist might work in a physical medium on a public object as they create the work, but this is not necessary, and a work can be created without being 'externalized'. Externalization is crucial, though, if the artist wishes to communicate the work to others. The object then created is not strictly the work, but the vehicle by which an awareness of the work can be transmitted. Communication is successful when the audience comes, through contemplation of the public object, to recognize or share the mental condition that is the work's existence. Collingwood denies, however, that it is the function of art to arouse emotion. Artists could pursue that goal only if they knew what emotion they wanted to arouse, and such knowledge comes only with successful expression and, hence, after the work's completion. Accordingly, art is primarily concerned with the self-awareness accomplished by the artist through their act of self-expression.

As an account of the creative process, which is much more variable than is specified, the theory is too narrow, and in so far as it applies to activities not normally regarded as issuing in art works, such as formulating and clarifying one's thoughts in language, it is also too broad. Second, in according primacy to the private, mental dimension of affective experience and in treating public expressions of emotion as dispensable ancillaries, the theory's account of expression is questionable. In addition, it leaves under-explained the process by which an audience comes to know the work and its expressive character. How do we map elements of the work's externalization to mental states and processes, thereby recreating those of the artist? Finally, in its account of the response appropriate to the appreciation of works of art, the theory seems at odds with the widespread view that most works admit a variety of interpretations and responses.

To the extent that it downplays the public context of presentation and appreciation, as well as the social determination and significance of artistic conventions, practices and genres, the account is committed to an uncomfortably idealist ontology for art works. It disregards the close identification we make in the case of singular pieces, such as oil paintings, between the work and the physical object in which it is realized. We speak of the properties and fate of such a work in terms of the qualities possessed and the vicissitudes undergone by that physical object. Art works are not deprived of their existence by the death of the artist, and their properties do not fade along with the artist's memories of its creation. Also, the theory cannot readily acknowledge the importance we accord to the artistic contribution of the performer in art forms such as drama, music and ballet. In general, the theory underplays the way in which the physical properties of a medium affect a work's artistic character by presenting distinctive possibilities and problems for the artist; or, to the extent that it admits such considerations, the account undermines the distinction it emphasizes between the (mental) act of creation and the (dispensable) activity of providing a public externalization of the work.

5 Fictional authors and personae

In the case of narrative or depictive works, it might be appropriate to regard the work as a communication from a fictional narrator or viewer, rather than from its artist. In line with critical practice, such an approach more readily tolerates the multiplicity of legitimate interpretations and the attribution of ideas and meanings to the work that were not held or intended by its artist. This fictional person, who stands outside the work and is distinct from its characters (including its internal narrator, should it have one), is constructed by the audience on the basis of the work's contents and conventions relating to what can be known and assumed in interpreting a work of the relevant time and place (see Narrative §2).

Once this strategy has been adopted, it will be natural to attribute to the fictional narrator or viewer attitudes and emotions expressed in the manner of the work's presentation, as well as beliefs and desires. For instance, we might

learn from the story that a character feels pride, but from the tone in which this is described also that the fictional narrator regards this pride as disappointing and contemptible. In that case it could be said that the fictional narrator is disappointed in the character's behaviour, even if there is not sufficient warrant for extending this reaction to the work's artist - indeed, even if there is reason to think the actual author might have had a more detached or ironic view.

The interpretive procedure outlined above might be extended to works that are neither narrations nor depictions, so that the expressiveness of such pieces is attributed to a fictional persona (whose thoughts and beliefs are largely absent from the expressive context). The work is approached, if not as a story, then as dealing piecemeal with the emotional experiences of a person about whom one knows little except that they undergo an emotion or sequence of feelings, the course of which is revealed in the work's contents and structure.

One might hear, say, a symphony as conveying the feelings of a fictional persona, but is this necessary for one to be aware of the work's expressiveness? If words for emotions always name the experiences of sentient beings, it might be thought that some such make-believing is involved, unconsciously if not consciously, where the work's expressiveness is not attributed to its artist, performer or audience, or to a character within it. Alternatively, if expressiveness can be present in the absence of occurrent emotions, this approach might be gratuitous.

6 Expressiveness as a property of art works

Several attempts to divorce the expressiveness of art works from expressions of occurrent feelings are unconvincing. We do sometimes say that art is expressive without indicating what it is expressive of. But, unlike Scruton (1983), I do not think that art deals with a kind of expression that is not the expression of an emotion or the like; rather, it reveals our concern (sometimes) to highlight the manner of a thing's being expressive, or the subtlety of its expressive nuance, above identifying or classifying the content of that expressiveness. Also unsatisfactory is the suggestion that art expresses emotions that are *sui generis* in being unfelt and non-cognitive, for this view, in divorcing artistic expressiveness from the world of human feeling, makes inexplicable art's power to move us as it does. And the supporting claim that art's expressiveness is ineffable is misplaced (in confusing description with duplication) and exaggerated (because it is only at a rather general level of expressiveness that there is sufficient interpersonal agreement to suggest that it is the work, rather than the responder, that is described). Nor does it help to call artistic expressiveness metaphorical. Metaphors take many forms and serve many purposes, so the suggestion is no substitute for an analysis of the phenomenon; analyses of literary metaphor seem not to be readily generalized to painterly or musical expressiveness; and many words for emotions seem no more lively as metaphors when predicated of art works than does talk of the necks of bottles.

Here is one argument for the view that art works are expressive without giving expression to occurrent emotions. We experience resemblances between art works and humanly expressive behaviour (voices, faces, deportment, actions). We do not notice similarities and infer a connection. (This is false to the phenomenology of the experience and, anyway, explains neither the direction of the inference nor our failure to connect art to many other things it might be seen as resembling.) Instead, as Kivy would have it, we are 'wired' to animate our experiences if we can. We see cars with faces, dolphins with smiles and willows with gloomy demeanours. The similarity resides, as I said, in the *experiences* we have of human and artistic expressiveness; basic resemblances that might underpin this experience, those between elements of human expressive behaviour and features of art works, are located, if at all, after the fact. In art, this potential for resemblance is frequently modified and structured by conventions, so it will be apparent only to those familiar with the appropriate artistic practices.

Why, though, should this experience of resemblance justify the attribution of expressiveness to art, since we know the analogy cannot be carried through because works of art cannot embody occurrent experiences? Sometimes attributions of expressiveness concern the character of an appearance (of a body, or face) without regard to feelings. We might identify a person's bearing as sad without meaning that they feel sad, or entertaining that thought, or regarding them as prone to the feeling, but merely as a description of their deportment. (Obviously this secondary use of terms of emotion follows on the primary one in which we are interested in expressive appearances for the feelings they display or convey. The behaviour that betrays a feeling in one context is likely, in the absence of feeling, to produce an appearance with a corresponding expressive characteristic.) It is arguable that when we say, for example, that a musical piece is sad, our use concerns not occurrent emotions but emotional characteristics presented in the sound of the music as we experience it. Such appearances are more compelling in

art than in nature, where they also occur, because we know them to have been deliberately employed by the artist.

Such a view meets several of the desiderata of an account of the expressiveness attributed to art works themselves. It locates the expressiveness in the work, explains the immediacy and directness with which this is experienced, and connects ascriptions of expressiveness to art with, if not the primary uses of those terms in connection with occurrent emotions, then with secondary uses of a familiar kind. Moreover, it fortunately does not describe the audience's experience as involving conscious inference or fanciful imaginings, since these seem to be absent from the reactions of many.

7 The value of artistic expressiveness

Expressiveness in art is usually thought to be value-conferring. Sometimes its value is described as instrumental - art is a source of knowledge about and mastery of the emotions, and is the more useful for presenting or arousing these in contexts that lack 'life implications', thereby permitting us to contemplate and savour their natures. The value of artistic expressiveness is also intrinsic in that it contributes to and is an aspect of narrative, depictive, formal and sensuous elements that together provide an integrated entity, the appreciation and understanding of which is pleasurable. Art's expressiveness is an invitation to engagement; a person's recognition of and response to its expressiveness can be as revealing of their understanding as would be their dispassionate description of the piece. In its expressiveness, art might best be regarded not as a mirror of life, a lesson on it or a preparation for it, but as a celebration of its affective side.

See also: Emotions, philosophy of; Emotion in response to art; Hanslick, E.

STEPHEN DAVIES

References and further reading

These works are not introductory in style, though none is forbiddingly technical or formal.

Bouwsma, O.K. (1950) 'The Expression Theory of Art', in M. Black (ed.) *Philosophical Analysis*, Englewood Cliffs, NJ: Prentice Hall, 71-96.(Widely reprinted. Classic discussion of a spread of relevant issues. See §3-6 of this entry.)

Budd, M. (1985) *Music and the Emotions: The Philosophical Theories*, London: Routledge & Kegan Paul.(Fine discussion of theories of music's expressiveness and ofHanslick's denial of such expressiveness; includes a chapter on the nature of the emotions.)

Budd, M (1989) 'Music and the Communication of Emotion', *The Journal of Aesthetics and Art Criticism*, 47: 129-38.(Considers the connection between expression and value in music and discusses several theories of musical expressiveness. See §5-7.)

Budd, M (1989) 'Music and the Expression of Emotions', *The Journal of Aesthetic Education*, 23 (3): 19-29. (Reviews various strategies for explaining music's expressiveness; opts for account making make-believe central. See §6.)

Budd, M (1995) *The Values of Art: Pictures, Poetry and Music*, London: Allen Lane, The Penguin Press.(Includes discussions of the nature and value of expressiveness in art with special attention to music.)

Davies, S. (1986) 'The Expression Theory Again', *Theoria* 52: 146-67.(Criticizes the expression theory and distinguishes modes of expressive behaviour.)

Davies, S. (1994) *Musical Meaning and Expression*, Ithaca, NY, and London: Cornell University Press.(Presents a detailed critique of theories of expressiveness in the arts, especially music; defends the account of artistic expressiveness outlined in §4; discusses the value of expressiveness; contains extensive bibliography.)

Elliott, R.K. (1966-7) 'Aesthetic Theory and the Experience of Art', *Proceedings of the Aristotelian Society* 67: 111-26.(Art is to be experienced imaginatively from within, and objectively from without.)

Elliott, R.K (1973) 'Imagination in the Experience of Art', *Royal Institute of Philosophy Lectures*, 6: 88-105. (Outlines the ways imagination enters in the experience of art and defends its place there.)

Gombrich, E.H. (1962) 'Art and the Language of the Emotions', *Proceedings of the Aristotelian Society* Supplement 36: 215-34.(Argues that, in art, the natural dimension of expressiveness is structured by conventions.)

Kivy, P. (1989) *Sound Sentiment*, Philadelphia, PA: Temple University Press.(Includes *The Corded Shell* (1980), in which he argues that music expresses emotion mainly by resembling expressive behaviour, and additional papers on the topic. See §§3, 6.)

Levinson, J. (1982) 'Music and Negative Emotion', *The Pacific Philosophical Quarterly*, 63: 327-346.(Discusses

the value of negative responses to art's expressiveness. See §7.)

Nolt, J. (1981) 'Expression and Emotion', *The British Journal of Aesthetics*, 21: 139-50.(Defends arousal theory. See §3.)

Osborne, H. (1982) 'Expressiveness in the Arts', *Journal of Aesthetics and Art Criticism*, 41: 19-26.(Outlines the difficulties of attributing expressive properties to art works and argues that music expresses moods, not emotions.)

Osborne, H. (1983) 'Expressiveness: Where is the Feeling Found?', *The British Journal of Aesthetics*, 23: 112-23. (Outlines the problems of attributing expressive properties to art works.)

Robinson, J. (1983) 'Art as Expression', in H. Curtler (ed.) *What Is Art?,* New York: Haven, 93-121.(Emphasizes that artistic expressiveness is treated as a communication from the artist, though not as a direct expression of occurrent feelings.)

Scruton, R. (1983) 'The Nature of Musical Expression', in *The Aesthetic Understanding*, London: Methuen, 49-61.(Regards musical expressiveness as gestural and metaphoric, and distinguishes 'transitive' from 'intransitive' expressiveness. See §6.)

Speck, S. (1988) '"Arousal Theory" Reconsidered', *British Journal of Aesthetics* 28: 40-7.(A defence of arousal theory.)

Stecker, R. (1984) 'Expression of Emotion in (Some of) the Arts', *Journal of Aesthetics and Art Criticism* 42: 409-18.(Holds that the arts differ in the manner of their expressiveness, depending on whether they possess a semantic, representational or other content.)

Tormey, A. (1971) *The Concept of Expression*, Princeton, NJ: Princeton University Press.(A detailed and sophisticated account of the notion of expressiveness in art; includes extended criticism of the expression theory.)

Vermazen, B. (1986) 'Expression as Expression', *The Pacific Philosophical Quarterly* 67: 196-224.(Holds that, in ascribing expressiveness to art works, including nonrepresentational ones, we are talking of the emotions of a fictional persona that we locate in the work.)

Wollheim, R. (1964) 'On Expression and Expressionism', *Revue Internationale de Philosophie*, 18: 270-289. (Discusses the nature of expressiveness.)

Wollheim, R. (1968) 'Expression', *Royal Institute of Philosophy Lectures* 1: 227-44. (Discusses expressiveness in painting, the artist's expression of emotion, and the conventions involved in this. See §4.)

Artistic forgery

Forgery in art occurs when something is presented as a work of art with a history it does not actually have. Typically this involves a false claim about the producer's identity. Forgeries are most usually works in the style of the artist whose work they falsely claim to be, while a forgery that is a copy of an existing work is a fake. Forgery is most common in the visual arts, but is also possible in other arts, such as literature and music.

The main aesthetic problem that forgery poses is that typically no deception is practised concerning what we might call the appearance of the forged object (generalizing from the pictorial case). Thus the forger does not deceive us about the disposition of colours on the canvas, the sequence of musical notes in the score, or the sequence of words in the text. If we adopt the widely held view that aesthetic value is a function of appearance alone, we shall conclude that something's being a forgery is irrelevant to its aesthetic worth; whatever false beliefs the viewer might be induced to have about the work, those beliefs could not affect an honest judgment of its aesthetic value. But in the art world it is universal practice to condemn forgery. If that practice is to be justified as anything other than artistic snobbery and the protection of prices in the art market, it must be shown that the aesthetic interest of a work is not exhausted by its appearance alone. In fact it can be shown that the aesthetic features of a work often depend on its historical features as well as on its appearance, and that these historical features are likely to be obscured by the deception that forgery involves.

1 Forgery and fakery

The most common type of artistic forgery is that of painting in the style of another, more valued artist; as, for example, when the failed Dutch artist van Meegeren painted and sold a series of 'Vermeers'. But forgery is possible in other arts, and poems, novels and musical works have had their origins intentionally misrepresented. Fakery also occurs, as when, for instance, a faked *Mona Lisa* was left in the Louvre to conceal the theft of the original. Fakery is less common than forgery, because the normally well-established facts of a painting's whereabouts make the deception involved in fakery hard to effect. The boundary between fakery and other kinds of forgery is hazy: Suppose A is a work, now lost, about whose appearance we have little information. I create B, without copying any work, and present it as A. Is B a fake of A, or merely an unfaked forgery? There are also works presented as the products of someone whose very identity is fabricated, such as the sixteen-poem oeuvre of the fictitious Ern Malley, composed as a trap for modernist enthusiasms by two poets in the Australian army.

Fakery and non-faked forgery, while different, pose questions for the aesthetician that can be given rather similar formulations. With a case of fakery (a case that is successful from the point of view of reproducing the given work's appearance) we ask, 'Is there an aesthetic difference between A, the original, and B, the fake, even though there is no difference in appearance between them?' In the case of non-faked forgery we ask, 'Is there a difference between A as it actually is, namely as something with a given appearance falsely purporting to be the work of X, and A as it might have been, namely as something with that same appearance but now genuinely the work of X?' In the case of fakery we are comparing objects within a single possible world; in the case of non-faked forgery we are comparing objects across possible worlds. In this brief survey we shall proceed informally, not making explicit this modal difference when arguing for the aesthetic relevance of forgery in general and fakery in particular. In this section we consider the varieties of forgery and fakery that are possible, and in the next section we shall tackle the aesthetic issue.

Forgery is possible in any art form, but Nelson Goodman (1981) has argued that fakery is not possible in music, literature and other 'allographic' arts, but only in 'autographic' arts such as painting and printmaking. (See Art works, ontology of §4 for critical discussion of this and other distinctions employed in this entry; here it is assumed that the autographic/allographic distinction divides the arts as Goodman claims, and that painting and printmaking are autographic while literature and music are allographic.) In autographic arts, a work's genuineness is a matter of its having the right history of production; in allographic arts it is not. Any sheaf of pages on which are inscribed the correct word sequence for *Northanger Abbey* is genuinely an instance of that work, irrespective of how, when or by whom it was produced. It is not possible to make a false claim that such a sheaf is genuine, because it is genuine. But it is possible to make a false claim that something which looks just like the authentic painting *Lucretia* byRembrandt is an instance (in fact the unique instance) of that work, or that something with the

appropriate appearance is an authentic instance of his etching *Tobit Blind*.

There is a difference between the painting case and the printmaking case. With painting, the unique instance of the work is identical to or constitutive of the work itself; with printmaking, where many authentic instances can be pulled from the same plate, no single instance, however authentic, is the work itself. So falsely presenting something as the painter's canvas amounts to falsely presenting something as the work itself; it is fakery of the work. Falsely presenting something as a print originating, in an authorized way, from the artist's block, amounts only to falsely presenting something as an instance of the work; it is not fakery of the work.

There is also a difference between autographic and allographic arts in the possibility of non-faked forgery: one can forge (but not fake) a work in the allographic arts without presenting anything that purports to be from the hand of the artist; this is not possible in the autographic arts. I might claim to be in possession of a poem by another hand, and produce as evidence for this a page of writing which I present as a mere copy of the alleged poet's autograph rather than as the autograph itself. This would be forgery. I can also falsely present something as a copy of a painting that is now lost. I could even present my copy as a perfect copy, and thus as a reliable guide to the appearance of that supposed work - how readily this would be believed is another matter. But it is doubtful whether this performance in the painting case constitutes forgery of any kind. Forgery is the activity of fraudulently claiming to present an instance of the work, and what counts as presenting an instance varies between the autographic and the allographic cases. In literature, an allographic art, any lexically correct copy of the original autograph is an instance of the work, and so the non-faking forger need claim only that what they present as an instance is lexically correct. For an autographic art form, the instances are just those with the requisite history, and the activity of presenting a copy not claimed to possess that history cannot count as presenting an instance of the work, and so cannot count as forgery.

2 The aesthetic significance of forgery

In this section we consider the potential impact that the discovery that an object is a forgery might have on our judgment of its merit as a work of art. It will not be argued that forgeries are automatically devoid of artistic worth, nor even that the fact of forgery is itself grounds for aesthetic reappraisal, though these claims have sometimes been made. It will be argued merely that the misrepresentation involved in a forgery is likely to have resulted in some other fact about the work being obscured, where the discovery of this other fact would justifiably lead to an aesthetic reappraisal.

The forger, in art and elsewhere, typically presents an object with a certain appearance, and hopes that we will make a plausible but erroneous inference from this appearance to certain other properties - its having been made by a certain person at a certain time and place, for example. If the aesthetic features of a work depend entirely on the appearance the forgery really has, and not on those inferred properties the forger hopes we will believe it has, forgery would have no aesthetic significance. Thus the visual appearance of van Meegeren's *Emmaus*, the word-sequence of the Malley poems, the ordering of musical notes in a Kreisler 'Mozart' piece, are evident features of these works. If they alone determine the aesthetic value of the work, the forger will not have practised any aesthetic deception. It seems we must say either that forgery is aesthetically irrelevant, or that the aesthetic qualities of the work depend on more than its evident features alone.

Someone unwilling to say either might appeal to an argument of Goodman's. Goodman claims that the best account we can give of the idea of two pictures looking exactly alike is one restricted to occasions of observation, that is, 'A and B are indistinguishable *for me now*'. So the fact that B is a forgery, though currently indistinguishable by me from A, may still be an aesthetically relevant fact for me now, because that fact will contribute to my future looking, and may result in my being able to see a difference between them. Even accepting Goodman's restriction of the notion of sameness of appearance, this argument is doubtful. To claim that the fact that one item is a forgery and the other is not constitutes an aesthetic difference because it *may* lead to the discovery of a difference of appearance seems to conflate possible and actual differences. If knowing that B is a forgery does lead me to notice a difference in appearance between A and B I may conclude that A and B differ aesthetically. If I do not find such a difference, then I have no reason to say that A and B differ aesthetically, whatever skills and knowledge I acquire in the process of examining them - so long, that is, as appearance is all that matters aesthetically.

If we are to establish the aesthetic significance of forgery we must abandon the idea that appearance alone determines a work's aesthetic value. A number of thought experiments have been suggested in which nonperceptible factors can be seen to make a difference to the value, *qua* works of art, of perceptually indistinguishable objects. To consider just the pictorial case: our judgment about this work's aesthetic success may crucially depend on whom we think it depicts. It may seem to capture the appearance and character of person X very successfully; when we learn that it is a picture of Y, our view may change. Its being a picture of Y rather than of X is not a feature of its appearance but of its history, and in particular of the history of its causal relations to X and to Y. Part of the impact of a work may consist in our recognition that it constitutes a strikingly original development of style or genre; a picture identical in appearance but produced much later would seem not original, but nostalgic or dull. Features of the work we would describe as elegant or delicate, which we thought had been produced by hand, would have little impact if we knew they were the product of a machine capable of producing lines of arbitrary thinness and complexity. There is little one can say about the appearance of an object that would be immune to these revisions based on a reassessment of the work's history. A description of the particular colours occurring at particular places on the surface might be the only claim immune to such revisions.

One response to the sorts of examples just cited is to say that they show simply that, when we judge a work according to representational, art-historical or technical criteria, we are not adopting a properly aesthetic standpoint. But it is unlikely that this narrowly aesthetic standpoint corresponds to the ways in which art works have traditionally been judged *qua* works of art, and its adoption would greatly impoverish our aesthetic discourse. The natural geometric partition of a picture, which we can describe in terms of abstract lines and planes, often depends on representational features, like the joined parts of a human body, that give salience to certain shapes. To ignore the representational features of the picture is to be no longer able to see what is natural about the geometric description. Apparently aesthetic predicates, such as 'is dynamic', when used for a painting, would not be aesthetic in this narrow sense; as Ernst Gombrich has emphasized (1977), a work like Mondrian's *Broadway Boogie-Woogie* can seem dynamic because works in the class to which it is naturally compared, namely Mondrian's other works, are rather static. Imagine *Broadway Boogie-Woogie* to have a different history, and it might be less appropriate to call it dynamic.

It is sometimes said that the upshot of our acquiring knowledge of the referential, art-historical and technical features of a work is that the object *looks different* to us from the way it had looked before we acquired that knowledge. And so, the objection goes, the examples described above are not really counterexamples to the thesis that there is no aesthetic difference without a difference of appearance. Rather, they are examples of the ways in which the work's appearance itself depends on extrinsic factors. If the objection were right, we should be able to answer the charge that forgery is aesthetically irrelevant immediately and without difficulty, saying simply that, since objects with different histories look different, there can be no objection to re-evaluating a work when we come to know that its history has been falsified, because that will change its appearance. However, the objection misses the point. To say that two objects are indistinguishable in appearance is not to say that how they look to the viewer is independent of the viewer's beliefs about their histories. A single object may look different to the viewer on two occasions of looking if they learn something about its history between the times of looking, but there is still a sense in which its appearance is unchanged. The sense in which objects can look the same despite differences between them in respect of facts about reference, history and technique is this: there could be a copy of the *Mona Lisa* so similar to the original that no one would be able to tell, on the basis of merely looking, whether the copy had been substituted for the original. The fact that, were the substitution pointed out, the copy would now look different to you from the way it had looked before (and different from the way the original had looked) is no objection to the claim that copy and original look exactly the same. It is in this sense of 'look the same' - being an indiscernible substitute for - that there can be an aesthetic difference between works that look the same.

We may conclude, then, that objects with the same appearance may yet have distinct aesthetic properties, and that therefore the aesthetic properties of the work are not fully determined by its appearance. In that case, the deception involved in forgery may result in the audience being misinformed about historical properties of the object potentially relevant to the determination of its aesthetic properties. Forgery is to that extent an aesthetically relevant fact.

See also: Artist's intention

GREGORY CURRIE

References and further reading

Currie, G. (1989) *An Ontology of Art*, London: Macmillan.(Discusses forgery within a framework that rejects the usefulness of the autographic/allographic distinction.)

Danto, A.C. (1981) *The Transfiguration of the Commonplace: A Philosophy of Art*, Cambridge, MA: Harvard University Press.(Argues that works identical in appearance may have distinct aesthetic properties.)

Dutton, D. (1983) *The Forger's Art*, Berkeley, CA, and Los Angeles, CA: University of California Press.(An anthology of work by art historians, critics and philosophers.)

Gombrich, E. (1977) *Art and Illusion: A Study in the Psychology of Pictorial Representation*, London: Phaidon, 5th edn.(An important historical work which, though it is not directly concerned with forgery, vividly illustrates some of the ways in which our perception of pictures depends on assumptions about their histories.)

Goodman, N. (1981) *Languages of Art*, 2nd edn, Brighton: Harvester Wheatsheaf.(A classic work that includes an account of the autographic/allographic distinction and its application. Sometimes intricate, but Part 3 on 'art and authenticity' is quite accessible.)

Sagoff, M. (1978) 'On Restoring and Reproducing Art', *Journal of Philosophy* 75: 453-70. (Argues for the aesthetic worthlessness of forgeries.)

Artistic interpretation

Interpretation aims to advance understanding by providing explanations of various kinds. In art, it should aim to maximize our understanding and appreciation of a work, and enable us to grasp its artistic values. When we interpret an art work we may explain why its elements are placed in their contexts, for example, to convey a certain meaning or express a certain feeling. In the case of literature we explain why words and passages are placed where they are, why characters and events are described as they are, and so on. When we interpret whole works, we explain how they fit into broader explanatory frameworks (for instance, Freudian or Marxist) or how they relate to various traditions so as to serve (or reject) the values emphasized in those traditions.

The distinction between description, or the presentation of the fundamental data constituting a work of art, and interpretation, which involves explaining why those elements exist in a work, what values they serve, may be used to justify claims that interpretations can never be known to be true, while descriptions are obviously true or false. While this is reinforced by the fact that a work may generate conflicting interpretations, the distinction does not imply that interpretations cannot be known to be correct.

In a related debate, many see the artist's intention in creating their work as the key to a valid interpretation. Since, however, many people find value in works in ways unintended by the artist, the onus is on the intentionalist to demonstrate the primacy of the value that the artist intended the work to have.

Ultimately, contending interpretations may not present as great a problem to a theory of interpretation as at first seems inevitable. Interpretations give priority in different ways to different artistic values; the choice of these values is simply a matter of artistic taste, not truth, and may not threaten the validity of any reasonably grounded interpretation.

1 Interpretation and description

Based on the fact that there is often more conflict among interpretations than among non-interpretive descriptions of works, many philosophers claim that interpretation is always on weaker epistemic ground than description. They claim that interpreters cannot know, or cannot know that they know, that their interpretations are correct. This epistemic weakness is supposed in itself to mark the distinction between interpretation and description of art works. But examples doom this attempt at distinguishing interpretation. Lady Macbeth can be known to be ambitious and manipulative, and this is an interpretation, albeit an obvious one, of her character. The distinction instead lies in the fact that interpretations are inferred, as explanations for what can be described without being interpreted. We non-interpretively describe what we directly perceive (as opposed to infer) in works; and such descriptions elicit universal agreement from audiences knowledgeable about the media in question, while interpretations need not elicit such agreement.

All artistic media present us with elements that can be described without being interpreted. Such elements constrain acceptable interpretations of the works that contain them, in that interpreters must explain why those elements exist as they do in the works. In painting and music the existence of such elements is obvious. Included among them are coloured shapes or notes and their formal relations. These make up the data or evidence for acceptable interpretations, which cannot be incompatible with their agreed descriptions. The counterparts in literature are not simply physical marks on paper, but words and sentences with their standard meanings. We do not perceive ink marks and infer that they are there to represent meaningful words. Instead, we directly perceive words and sentences, and, if we understand the language being used, agree on their standard meanings. Such elements make up texts, which should be defined in terms of standard lexical meanings at the times the texts are written. Even when we interpret such elements as metaphor or irony, we infer such interpretations as the best explanations for words with those standard or literal meanings being placed in those contexts. We make such inferences when doing so makes for more interesting readings or valuable literary experiences.

Some contemporary theorists of literary criticism deny that texts constrain acceptable interpretations. They see readers or critics as joining in the production of texts through their interpretive readings. To see pre-existing texts as constraints is, according to these theorists, to underestimate the degree of freedom that critics have in specifying their meanings and significance. Poetry is so notoriously open to radically different readings that it is empty to

claim that poetic texts, as opposed to critical communities, constrain what is acceptable as interpretation. But these theorists exaggerate. Even a poem such as William Blake's *The Tyger*, which is notorious for eliciting diverse interpretations, shows otherwise. For acceptable interpretations, no matter how divergent, must explain, for example, why Blake centrally used a term that ordinarily refers to tigers.

Others deny the distinction between description and interpretation by pointing out that what we take to be fictional facts in a work can depend on our interpretation of it. There are ghosts in *The Turn of the Screw* only if we interpret it as a ghost story. There is an egg in Piero della Francesca's Brera altarpiece only if we interpret a certain mysterious object in that way. But whenever interpretations are needed to determine what is (fictionally) true in a work, there remains a level of description of painted shapes or words used that cannot be placed in doubt. We may similarly grant that interpretations guide what we attend to or perceive in a work. The painted shapes to which we attend on a canvas will be those we take to contribute to the artistic value of the painting, and this will vary according to whether we interpret it as an allegory, a formal structure, and so on. But the shapes can nevertheless be described without being interpreted.

Finally, we may admit that the line between interpretation and description must sometimes be drawn finely, while maintaining that the distinction is crucial for understanding the constraints on acceptable interpretations. To point out, for example, that a musical passage leads back to the tonic key is to describe it non-interpretively, but to explain it simply as a bridge passage is to interpret it, however obviously.

2 Meaning

Most theorists of literary criticism agree that the interpretation of literature aims to disclose meanings in texts. They disagree about the ground of such meaning, whether it is simply the semantic conventions of the language, the intentions of authors, or the interpretive activities of readers or critics. But it should be clear that interpretation does not always consist in disclosing meanings (unless we think of meaning in the broad sense of significance, and of significance as the value of a work or the place of its parts in its broader structure); and disclosing meaning is not always interpreting. Trivially, in musical works and paintings there is not always semantic content to be revealed. In literature, if revelation of meaning were always interpretation, then every obvious paraphrase of every section of text would count as literary interpretation. But we need not interpret every simple sentence in a novel, and stating the ordinary dictionary meanings of the words in such sentences does not count as literary interpretation. As noted above, we simply grasp the meanings of such sentences as we read them, much as we see trees in a true-to-life landscape painting without having to interpret the painted shapes as trees.

One way of avoiding this objection is to view interpreting as determining the meaning of a whole text or a large segment of it from the lexical meanings of its words and sentences. But this view ignores the fact that phrases and even words in texts can be interpreted as well. Stating the meaning of some obscure phrase or line in a poem constitutes an interpretation when there is an implicit claim that the phrase is there to convey that meaning, and that its literary value lies at least partially in its doing so. Statements of meanings will rarely be complete interpretations of texts, since in literature the ways in which meanings are presented are relevant to the values of the works. Hence form and not merely content must be explained in complete interpretations. But certainly words often have value in a text largely because of the meanings they convey.

The idea that interpretation consists in determining the meaning of a whole text also implies that a simple paraphrase of a whole work is the paradigm interpretation of it. According to the explanatory account, it is most unlikely to be an interpretation at all. Simply to paraphrase a work is not normally to indicate the ways that its passages contribute to its overall artistic value. Conveying certain meanings is normally only part of that value, and paraphrase is needed to explain that part only when the text itself is ambiguous or obscure. Explanation of metaphorical, symbolic or ironic meaning is more often a part of genuine literary interpretation.

3 Intention

Much debate on this topic has centred on the relevance of artists' intentions. According to one side of this debate, if there is to be a standard of correct interpretation, it must lie in uncovering the intentions of artists regarding the meanings and expressive properties in their works, since otherwise these matters remain indeterminate and open to conflicting construals. According to this view, art is a form of communication, and the aim of the recipient, as in

all communication, is the discovery of the speaker's intentions. According to the other side, artists' intentions are inaccessible or irrelevant, since, if they are successfully realized, the results will be apparent in the works themselves; and if they are unsuccessful, they cannot determine how a work should be interpreted.

Viewing interpretation as value-enhancing explanation provides a different perspective on this issue. From this viewpoint there is certainly some value to be derived from seeing the world of an art work (and perhaps the real world as well) as its creator saw it and intended it to be seen. Seeing through another's eyes, or imagining through another's creative genius, so as to alter one's own imaginative vision, is a major benefit to be derived from the appreciation of art. This benefit requires fidelity to artists' intentions, where these are recoverable.

On the other side, to accept recoverable intentions as constraints on correct interpretations, to insist that there is always only one correct interpretation of any work of art and that this is the one intended by its artist, may be to rob contemporary audiences of valuable experiences they might have of the work. A Freudian reading of *Hamlet* may afford illuminating insights into its characters, whether or not such an interpretation was or could have been intended by Shakespeare. It is commonplace in the domain of music for conductors and performers not to be limited to conveying expressive properties specifically intended by composers. (It might be objected here that the performative interpretations of musicians do not fit the definition of interpretation given above, since they are not explanations. But they are informed by critical interpretations that are not constrained by composers' explicit intentions.)

The intentionalist can also be accused of confusing speakers' (artists') meanings, or what speakers intend to say, with utterance (text) meanings, or what their language conveys according to its semantic conventions. If an utterance or text is unclear or ambiguous, then a speaker's intentions cannot in themselves make it less so. In ordinary communicative contexts, we do not use speakers' intentions to clarify the meanings of their utterances; instead, we use the semantic conventions governing their utterances as guides to recovering their linguistic intentions.

Thus, the best argument for intentionalism - the claim that interpretation aims only to disclose artists' intentions - must be that art is a form of communication and, as such, shares this aim. This is partially correct in that, as indicated above, it points to one major value in the experience of art. But there are others. Since certain art works may be more interesting or expressive if not limited by their creators' specific intentions, the intentionalist must argue that the value indicated by the communicative model trumps all other values to be derived from the experience of art.

It should be noted finally here that we may ascribe intentions to artists on different levels of specificity. They may intend, for example, not only to convey specific meanings, but for their works to have certain dramatic or expressive effects, or, most broadly, for their works to be appreciated to the fullest extent possible. The realization of this broadest (and perhaps most common) intention may require creative acts of interpretation on the part of audiences. The problem with the orthodox intentionalist view is not (as some critics claim) that it forces us to search outside the work itself for its proper interpretation. We must in any case locate a work correctly in its broader artistic context and tradition to interpret and evaluate it correctly. The problem is rather that it focuses only on one sort of intention and one sort of value at the expense of others.

4 Incompatible interpretations

If acceptable interpretations guide perceptions or experiences of works towards enhanced appreciation of their values, and if art works have potential values that cannot be realized simultaneously in experiences of them, then incompatible interpretations might be equally acceptable. The second antecedent is true of many works and their parts. Iago's 'Credo' aria in Verdi's *Otello*, for example, can be interpreted as boisterously defiant or as broodingly sinister, and these different readings of the score and text will lead to different understandings of the character and his relations to other characters and to the dramatic events, and to different experiences of the work's expressive qualities. We cannot simply combine these interpretations by viewing the aria as ambiguous, since a reading or performance of it as ambiguous would differ from both these interpretations. It would be probably less satisfying, if more subtle, than either of the single interpretations, each of which could be equally satisfying or acceptable.

Interpretations are incompatible when they ascribe properties to a work (or its parts) that it cannot simultaneously

possess. They are acceptable when they produce understanding and experience of works that enhances the appreciation of central values in those works. Since great works of art are often multi-dimensional, it is rare that all their artistic values can be appreciated under single interpretations. When there are at least two incompatible but equally acceptable interpretations of a work, we cannot simply equate acceptability with truth. Inconsistent ascriptions cannot all be true, at least not absolutely. Talk of plausibility will not do here either, since plausibility must be defined in terms of probable truth. In such cases we can speak only of either acceptability or of truth relative to an acceptable interpretive scheme. When there is only one acceptable interpretation of some part of a work, we can continue to speak simply of truth.

That many works of art are intended to be susceptible to a variety of interpretations is indicated by their frequent use of metaphor and symbol, and by the incompleteness of their notations (especially in music). A theory of interpretation must explain why many apparent critical disagreements seem irresolvable. The theory presented above explains such disagreements as symptoms of incompatible but equally acceptable interpretations, and it explains the latter as resulting ultimately from differences in taste. If different interpretations enhance different artistic values by explaining elements within works as serving those values, then preference for one interpretation of a work over another will reflect a preference for particular kinds of experience that can be derived from the work. That is, it will reflect a particular taste in art.

Thus, interpretive disputes will mirror disagreements in evaluations of various works. Only realists or absolutists about the latter should be realists or absolutists about the former. Nevertheless, the acceptance of incompatible interpretations of the same works does not imply a lack of standards. Iago's aria may be interpreted in either of the irreconcilable ways mentioned above, but a reading or performance of it as light-heartedly humorous would be clearly inappropriate. An acceptable interpretation must adequately explain the work of art as it can be non-interpretively described.

See also: Art criticism; Art, understanding of; Artist's intention; Barthes, R.; Derrida, J.; Structuralism in literary theory §§2-3

ALAN H. GOLDMAN

References and further reading

Barnes, A. (1988) *On Interpretation*, Oxford: Blackwell.(Distinguishes interpretation on epistemic grounds.)

Beardsley, M. (1970) *The Possibility of Criticism*, Detroit, MI: Wayne State University Press. (Argues the anti-intentionalist position.)

Eco, U. (1990) *The Limits of Interpretation*, Bloomington, IN: Indiana University Press. (Reasserts limits to the freedom of interpreters.)

Fish, S. (1980) *Is There a Text in This Class?*, Cambridge, MA: Harvard University Press. (Argues that communities of critics produce texts, which do not exist without them.)

Gill, N.S. (ed.) (1982) *Structuralism and Literary Criticism*, Atlantic Highlands, NJ: Humanities Press.(A collection of French structuralist interpretation.)

Hirsch, E.D. (1967) *Validity in Interpretation*, New Haven, CT: Yale University Press. (Argues the intentionalist position.)

Young, R. (ed.) (1981) *Untying the Text*, Boston, MA: Routledge & Kegan Paul.(A collection of post-structuralist interpretation.)

Artistic style

Artistic style is a problematic notion in several ways. Sometimes the term refers to style in general, as it does in 'Good style requires good diction'. Sometimes it refers to style as a particular, as in 'Van Gogh's style' or 'the Baroque style'. In antiquity, style was a rhetorical concept referring to diction and syntax; consequently style is very often identified with the formal elements of a work of art as opposed to the content. However, the kind of subject matter an artist chooses may itself be a significant feature of style. One way of thinking about style is as a set of recurrent features of works of art that identify them as the product of a particular person, period or place. This may be adequate for some purposes, but it ignores the fact that a style has a unified 'physiognomy' or expressive character. The relation between style and expression is complex. A period style is often thought to express the cultural attitudes of the period, but it cannot do so in a very direct way. What a style expresses is a function of where it occurs in the history of style. Similarly, a work of art in an artist's individual style will be expressive only in the context of the possibilities of that style. According to the Romantic tradition, individual style is a genuine expression of the artist's self. But according to others, style is simply a construction by readers, viewers and listeners.

1 Historical background

The concept of style in the arts was first elaborated in Ancient Greek and Roman treatises on rhetoric - the art of public speaking. The ancients distinguished style from invention (subject matter) and organization (the arrangement of subject matter into parts). Style is diction, or word use, and composition, or the way in which words are combined into sentences. It is form rather than content: not what you say, but how you say it. Good style consists in correctness, clarity, ornamentation and decorum (appropriateness). However, since all good speech should be grammatically correct and clear, what chiefly makes the difference between one style and another is the use of ornamentation or rhetorical 'figures' and tropes. The principle of decorum stipulates that the style of a speech be appropriate to the total situation in which it is delivered, including who the speaker is, what they are talking about and what audience they are addressing. The ancients distinguished three kinds of style - plain, middle and grand - and each was appropriate to different occasions and purposes. It was very important to know how to adapt your style so as to secure the desired intellectual and emotional effect in your audience. Finally, style is sometimes thought of as an image of the speaker: Cicero's style to some extent reflects Cicero himself.

The rhetorical concept of style endured largely unaltered through the Enlightenment, and extended its area of application to music and the visual arts. The same 'subject', such as a portico or a representation of the Crucifixion, could be presented in different styles to achieve different effects. Different styles were individuated and organized into hierarchies. Decorum remained important: just as the epic poem demanded a more elevated style than the lyric, so in painting the grand style was suited to history painting as opposed to still life. Critics such as John Dryden took issue with Shakespeare because he mixed the grand style of tragedy with the low style of comedy in such works as *Hamlet* and *King Lear*, and the music theorist Johann Joseph Fux insisted that the styles of church music should not be confused with those of theatre and dance.

The advent of Romanticism and German Idealism radically altered the way in which style was conceived. The Romantics rejected the hierarchy of genres and the idea that a particular subject demanded a particular appropriate style. They argued that style and subject are not independent and that style cannot be defined in terms of a list of rhetorical ornaments. Coleridge, for example, celebrates the poem as a living, organic whole in which style and content are fused. To the Romantics a work of art was an emotional expression by an artist with a unique sensibility, whose emotional responses to a subject were embodied in both the style and content of their work. Individual style was the expression of all the peculiarities of the artist's qualities of mind and feeling.

On an altogether more grand scale, Hegel argued that different period styles are expressions of culture in general. For Hegel, art is a sensuous manifestation of Spirit, and the successive phases of art correspond to the inevitable movement of Spirit towards self-realization (see Hegel, G.W.F. §8). Each phase corresponds to a style - the Symbolic, the Classical and the Romantic (post-Classical or Christian) - and each is expressive of a different culture - the Ancient Egyptian, Ancient Greek and modern. Shorn of some of their metaphysical underpinnings, these ideas re-emerge in the theories of period style and style change developed by the great nineteenth-century

German art historians, Heinrich Wölfflin, Alois Riegl, Paul Frankl and others. For example, Wölfflin identified five contrasting style qualities - the most famous being ‘linear’ versus ‘painterly’ - through which he defined the contrast between the art of the High Renaissance and that of the Baroque period. He argued that there was a necessary evolution from the first set of qualities to the second, that this pattern of development recurred in most historical periods, and that it was due to principles internal to the history of art.

2 Style and form

The ancient rhetoricians discussed style in terms of rhetorical figures, both semantic (such as metaphor and personification) and syntactic (such as asyndeton and antithesis). Contemporary stylisticians and discourse analysts have used modern linguistic techniques to identify particular stylistic features of poems, plays and ordinary discourse. Style in this sense is identified as how something is said rather than what is said: with form rather than content. Very different things can be said in the same style, so style would appear to be independent of content. And the same content can be expressed in different styles: ‘The cat is on the mat’ is in plain style, in contrast to ‘The feline animal is situated upon the rug’, which is (inappropriately) in grand style, but they mean much the same thing. Style would therefore seem to involve choice of words and syntax - the ‘formal’ elements of the discourse rather than the content.

Are all formal elements part of style? Monroe Beardsley (Lang 1987) has tried to mark off stylistic from non-stylistic features of a discourse as those linguistic features which carry connotative or secondary meaning or which enable it to ‘reflect a subordinate illocutionary action’. So Caesar’s famous assertion, ‘Veni, vidi, vici’, primarily means that he came, he saw and he conquered, but in leaving out ‘and’ the utterance also implicitly asserts that Caesar operated quickly and decisively. The trouble is that any linguistic feature can have connotative or secondary meaning. Even when I say ‘The cat is on the mat’, I am implying that what I say is plain and straightforward. Indeed, any attempt to distinguish stylistic from non-stylistic linguistic features may well fail, since any word or grammatical construction in an appropriate context can contribute to style. The same is true of formal elements in the other arts, such as a particular sequence of chords or colours. The attempt to define a set of uniquely stylistic formal features seems to be hopeless.

3 Style and signature

In contemporary debates in aesthetics what is at issue is not normally style or stylistic features in general, but rather what is the nature of ‘a style’. Since a style is what picks out the work of a particular artist, period or place, perhaps a style can be thought of as the recurrent formal elements that identify a work as belonging to that artist, period or place.

The most important problem with the formalistic approach to style is that a style consists of more than just a set of formal elements. Styles have particular expressive qualities: they are plain, ornate, pompous, diffuse, sweet, euphonious, Miltonic, energetic, Latinate, abstract or flabby. Very often, subject matter is stylistically important: a penchant for subjects from Roman myth and legend together with fanciful Roman landscapes is arguably a feature of Poussin’s style, just as a tendency to domestic pastoral landscape is a feature of the Barbizon school style. A particular kind of iconography or conventional symbolism may also be important. Sometimes the use of certain materials - a preference for oil over watercolour or for bronze over marble - can be a feature of style, as can the use of certain techniques. In recognition of these multiple possibilities, Nelson Goodman has defined a style as ‘a complex characteristic that serves somewhat as an individual or group signature… in general stylistic properties help answer the questions: who? when? where?’ (1975).

The problem with this proposal is that not all identifying features are stylistic; the actual signature on a painting, for example, might not be part of the painter’s style. What, then, is distinctive about style regarded as signature? One answer is that it is only aesthetically salient qualities that count as stylistic. But on the one hand, some aspects of a work of art, such as its size and subject matter, are always aesthetically salient whether or not they are part of style; on the other hand, style features are not always particularly salient: often only a very careful study will unearth them.

One of the important facts about a style is that it comes across as an expressive unity. A set of recurrent features is not a style unless the features themselves combine to form a certain ‘physiognomy’: the style is pompous or

sentimental or Ciceronian. Hence style qualities do not just identify an artist, school or period; they also contribute to the expression of a particular ‘character’. One plausible suggestion, therefore, is to count as stylistic all those features of subject matter, form, expression, symbolism, materials and so on. which contribute towards the expression of the overall character of the individual or period in question.

4 Style and expression

The Hegelian idea that a period style expresses the collective spirit of an epoch or country and that the style of a particular work of art is a symptom of that spirit has been roundly criticized by twentieth-century art historians. Meyer Schapiro (1994) has urged that we should not extrapolate from a single painting to cultural attitudes in general: one-to-one correlations usually hold only between single aspects of a painting and the culture from which it originates. Erwin Panofsky (1955) has shown how we cannot tell what general attitudes are expressed by a painting unless we can place the picture in the history of style and the history of iconography. We can interpret what Tintoretto’s *Last Supper* expresses only if we know the history of renderings of the Last Supper. The picture has to be seen as a response to and a rejection of previous *Last Supper*s, such asGiotto’s and Leonardo’s. Ernst Gombrich (1960) has made a similar point with respect to works within a painter’s individual style: what a painting expresses is a function of its place within the artist’s ‘language’ or repertoire. Van Gogh’s painting of his room at Arles is relatively serene in the context of Van Gogh’s oeuvre, whereas if it were (*per impossibile*) by Cézanne it would express much greater turbulence and distress.

Deterministic theories of style change have also been criticized. All such theories treat style as inevitably moving towards some goal, but artists cannot be striving to achieve a perfection that has not yet come about and of which they cannot be cognizant. Style change is the result of individual artists responding to many influences, including the challenges posed by the art of the past, and should be thought of, says James Ackerman, not as ‘a succession of steps towards a solution to a given problem, but as a succession of steps away from one or more original statements of a problem’ (1963). Nevertheless, certain patterns of development do recur in the history of art, for example in the development of sculpture in Ancient Greece and the Renaissance. Ackerman suggests that this is due to a similarity in the way that the problems facing sculptors were conceived, as well as a similarity in the solutions they found. Thus both ancient and Renaissance sculptors were struggling with the same problem of how to create beautiful human forms out of blocks of stone, and there is the same development from a ‘blocky’ archaic style to the ideal Classical solution, and from that to the development of a freer, more dynamic style.

If style change and the expressiveness of style can only be understood as the result of individual artists responding to the art of the past and implicitly either accepting or rejecting specific alternatives, then it might seem as if style entails choice: in developing a style an artist chooses one particular form of words or configuration of lines in preference to others, depending upon what they want to express. Leonard Meyer, for example, defines style in music as a ‘replication of patterning… that results from a series of choices made within some set of constraints’ (1989). For example, given the constraints on the style of the classical sonata, Beethoven’s decision to begin his piano sonata, ‘Les Adieux’, with a deceptive cadence is an unexpected stylistic choice that permits the opening phrase to express an uncertain, plaintive quality. On the other hand, the word ‘choice’ may be misleading if it implies deliberation and conscious decision-making.

5 Style and personality

Most of the theories of style discussed so far have considered style from the historian’s point of view. It is, for example, a third-person viewpoint which treats Beethoven as working within certain constraints and as making certain choices. A different approach has been taken by Richard Wollheim (Lang 1987), who argues that there is an important theoretical distinction between the individual style of a particular artist and such general style categories as school style (the style of the school of Giotto), period style (Baroque concerto-grosso style, Augustan poetic style) and universal style (the geometric style, the heroic-epic style). General style categories are the invention of historians, who try to organize a body of knowledge according to their own interests and purposes. By contrast, an artist’s individual style has ‘psychological reality’ and can be captured only by a ‘generative’ conception of style that picks out and groups together elements of the artist’s work which are ‘dependent upon processes or operations’ characteristic of the artist’s acting as an artist. Wollheim restricts his theory to pictorial style, since he is thinking of style processes not only as psychologically dependent on the artist but also as physically embodied in motor habits and motor memory. However, the theory can be generalized to the other arts

if style is thought of as a way of doing or making something which is expressive of the artist's character, qualities of mind, attitudes and sensibility.

This way of thinking about style is reflected in Arthur Danto's maxim that style is 'what is done without art or knowledge' (1981). On this view, artists, in developing their individual styles, do not literally 'choose' among alternatives. Being of a certain character, the artist is able to paint or write only in accordance with that character. By the same token, a forgery of a painting, even one that cannot be distinguished from the original, cannot have style. It is a deliberate imitation of an already existent work or style, whereas the original is a genuine expression of the artist's self. For the same reason a forgery has no genuine aesthetic significance, for having a formed style is a precondition of expressiveness and hence of aesthetic interest (see Artistic forgery).

The theory that style is a way of doing something that expresses the artist's unique personality, character and ways of thinking and feeling explains many of the puzzles surrounding the concept of style. It explains why not all formal features of a work or oeuvre are stylistic, why subject matter can but need not be a feature of style, and why a feature can be a stylistic feature in one work or oeuvre and not in another: in every case it depends on whether the feature is expressive of the artist's character or personality. Similarly, the theory explains the unity of style in terms of the unity of the personality expressed and it explains the difference between style and signature: a recurrent feature of an artist's oeuvre is not stylistic unless it contributes to the expression of the artist's character. Thus a literal signature is not usually a feature of style.

However, the theory is unattractive to those who think of style as irremediably conventional, as the result of operations performed by readers. Some writers have argued that style is a function of the operations performed by the 'implied author' of the work, as constructed by the reader or viewer. Michel Foucault takes a more extreme position (1979). He argues that the notion of a unified style is simply one of the principles that identify the 'author-function' of a work, this function being characteristic of works which have a certain status and are designed to be received in a certain way. On this view style has no reality at all; it is just a social construction.

See also: Aesthetic concepts §4; Formalism in art

JENEFER M. ROBINSON

References and further reading

Abrams, M.H. (1953) *The Mirror and the Lamp*, Oxford: Oxford University Press.(Still perhaps the best introduction to the history of ideas in the Romantic movement in literature.)

Ackerman, J. (1963) 'Style', in J. Ackerman and R. Carpenter (eds), *Art and Archeology*, Englewood Cliffs, NJ: Prentice Hall, 174-86.(A theory of style in the visual arts by an art historian.)

Carter R. and Simpson, R. (1989) *Language, Discourse and Literature*, London: Unwin Hyman.(An introduction to stylistics and discourse analysis.)

Cooper, L. (1912) *Theories of Style*, New York: Macmillan.(Contains theories of literary style by many great writers, including Aristotle, Longinus and Coleridge.)

Danto, A.C. (1981) *The Transfiguration of the Commonplace*, Cambridge, MA: Harvard University Press, esp. ch. 7.(Application of Danto's general philosophy of art to style and expression.)

Foucault, M. (1979) 'What is an Author?', inJ.V. Harari (ed.), *Textual Strategies*, Ithaca, NY: Cornell University Press, 141-60.(Translation of a 1969 essay by the French philosopher.)

Gombrich, E.H. (1960) *Art and Illusion*, London: Phaidon, esp. ch. 11.(Fascinating discussion of the development of style in Western 'realistic' painting and the relation between style and expression.)

Gombrich, E.H. (1968-79) 'Style', in D. Sills (ed.), *International Encyclopedia of the Social Sciences*, New York: Macmillan, 18 vols, 15: 352-61.(A summary ofGombrich's views on style.)

Goodman, N. (1975) 'The Status of Style', *Critical Inquiry* 1: 799-811.(A theory of style by a philosopher, which complements the views developed in his *Languages of Art*.)

Kennedy, G. (1994) *A New History of Classical Rhetoric*, Princeton, NJ: Princeton University Press.(Gives a good account of the treatment of style in Ancient Greece and Rome.)

Lang, B. (1987) *The Concept of Style*, Ithaca, NY: Cornell University Press.(First edition 1979. Contains a number of important essays on style by philosophers, art historians and others, including several mentioned in this entry: Leonard Meyer, 'Toward a Theory of Style', Richard Wollheim, 'Pictorial Style: Two Views', and Monroe Beardsley, 'Verbal Style and Illocutionary Action'.)

Meyer, L. (1989) *Style and Music*, Philadelphia, PA: University of Pennsylvania Press.(An expansion of the essay

in Lang (1987) by a music theorist.)

Minor, V.H. (1994) *Art History's History*, Englewood Cliffs, NJ: Prentice-Hall.(Useful if brief introduction to the major figures in Western art history.)

Panofsky, E. (1955) *Meaning in the Visual Arts*, Garden City, NY: Doubleday, 26-54.(First published in 1939, this essay explains Panofsky's principles of art historical interpretation. Page reference marks 'Iconography and Iconology: An Introduction to the Study of Renaissance Art'.)

Podro, M. (1982) *The Critical Historians of Art*, London: Yale University Press.(The theories of post-Hegelian art historians.)

Robinson, J.M. (1985) 'Style and Personality in the Literary Work', *Philosophical Review*, 94: 227-47. (Development of some of the ideas in §5.)

Schapiro, M. (1994) *Selected Papers, Volume 4: Theory and Philosophy of Art: Style, Artist and Society*, New York: George Braziller, 51-102.(A seminal essay on style, originally published in 1962; contains a useful bibliography.)

Van Eck, C., McAllister, J. and Van de Vall, R. (1995) *The Question of Style in Philosophy and the Arts*, Cambridge: Cambridge University Press.(A collection of essays on style.)

Wölfflin, H. (1950) *Principles of Art History*, New York: Dover.(First published in 1915, this book explains the five principles which are said to distinguish Renaissance from Baroque style.)

Artistic taste

Taste has been variously understood as (1) the capacity to take pleasure in certain artistic and natural objects, (2) the capacity to identify the constituent elements in such objects, and (3) the capacity to discern certain special properties. Taste in sense (1) has been a topic since the early eighteenth century, culminating in the work of Hume and Kant. This conception of taste is annexed to the idea that 'beauty' or 'artistic excellence' is not itself an objective property of things, but that it is recorded in judgments of beauty as a report of a certain kind of pleasure felt by the judge in the presence of these things. Taste in sense (2), which is an analogue of the notion of taste as the ability to discriminate with the tongue and taste buds, has also been a topic since the eighteenth century, articulated perhaps most clearly by Hume. A connection between sense (1) and sense (2) is intended by eighteenth-century authors, but the connection has not been formulated clearly. Taste in sense (3) is a conception originating in the mid-twentieth century, notably in the work of Frank Sibley. It is primarily the idea that beauty, elegance, gracefulness and other properties - collectively called 'aesthetic properties' - require a special capacity for their discernment, although these are truly objective properties located in the objects being judged.

1 Hume

The idea of taste was given a central position in the philosophy of art in the eighteenth century, and though, afterKant, it ceased to be an especially lively topic, it has returned in the twentieth century and continues to receive considerable attention.

For more than two centuries taste has been thought of, roughly, as the human capacity to respond to beauty, and in early theories - David Hume's, in particular - the capacity to elicit certain responses of taste was taken to be the mark of beauty. Thus Hume understands the 'standard of taste and beauty' to be located in judges of particularly qualified taste.

In the eighteenth century, taste is characterized in two ways, whose relation is not obvious. (1) Taste is thought of as the capacity to respond with affirmative and negative feelings to items of beauty and ugliness. (2) Taste is thought of as the ability to discriminate the elements present in some item. The term 'discriminating', as in the expression 'a person of discriminating taste', tends to meld senses (1) and (2), sometimes without regard to an apparent logical difference between them. In one sense, a person of taste is said to be someone whose personal preferences attach to objects of merit: the person is someone who likes good things and dislikes poor things. In the other sense, a person of taste is said to be someone who can distinguish and identify the objects they deal with. People with taste in wine, for instance, are thought of both as people who truly enjoy better wines and are put off by poor wines, and as people who can identify and describe the qualities of any wine they happen to taste, and can, perhaps, identify the vintage. Both these senses of 'taste' are present in Hume's texts:

> There is a *delicacy* of *taste* observable in some men, which very much resembles this *delicacy* of *passion*, and produces the same sensibility to beauty and deformity of every kind, as that does to prosperity and adversity, obligations and injuries.
>
> (1741: 4)

> Where the organs are so fine, as to allow nothing to escape them; and at the same time so exact as to perceive every ingredient in the composition: This we call delicacy of taste, whether we employ these terms in the literal or metaphorical sense.
>
> (1757: 235)

When taste is thought of as a basis for normative judgments of beauty or artistic excellence, then the two senses of taste can be set in a kind of pragmatic relation. The determination of the aesthetic quality of something (its beauty, ugliness, artistic excellence, artistic failure, and so on.) is formulated in terms of how someone of adequate taste feels about the thing. The determination that this judge has adequate taste is in turn formulated in terms of the judge's ability to distinguish the components of the thing. This relation threatens to be circular unless either the merit of the thing or the qualifications of the judge can be independently established. Because those interested in the idea of taste have customarily hoped to define beauty in terms of the responses of judges of taste, their best recourse seems to be an attempt to show that the qualifications of the judge can be established independently of all

considerations of beauty. Thus Hume says:

> But if we consider the matter aright, these [questions of whether someone has adequate taste] are questions of fact, not of sentiment. Whether any particular person be endowed with good sense and a delicate imagination, free from prejudice, may often be the subject of dispute, and be liable to great discussion and enquiry…. Where these doubts occur, men can do no more than in other disputable questions, which are submitted to the understanding: They must produce the best arguments, that their invention suggests to them; they must acknowledge a true and decisive standard to exist somewhere, to wit, real existence and matter of fact…. It is sufficient for our present purpose, if we have proved, that the taste of all individuals is not upon an equal footing.
>
> (1757: 242)

This suggests a thesis that, in general, holds an object to be beautiful because it elicits pleasure in someone possessing the requisite capacity to respond, and it takes this requisite capacity to be identifiable in a person through the use of empirical, verifiable tests. Beauty is essentially a matter of human response, and is in this sense subjective, but since it is possible to authenticate the legitimacy of responses, the possession of taste is in this sense an objective fact; the attribution of beauty is thus a matter for empirical investigation (see Hume, D.).

2 Kant

Kant denies this possibility. He agrees that beauty is experienced essentially and solely in the exercise of taste, but denies that there is any empirical test that might determine whether taste is being properly exercised. ThusKant supposes that someone making what they take to be 'a pure judgment of taste' is justified in supposing that the object being judged is indeed beautiful, but that neither they nor anyone else can in principle *prove* that the judgment is correct. Despite his insistence that the judgment of taste is not an objective judgment, Kant argues that it is a unique and essential illustration of the general human capacity to make judgments of any kind, including those which are not 'aesthetical' and hence 'subjective', but are genuinely what Kant calls 'logical', 'objective' judgments.

One peculiarity of Kant's work is that he initially develops his theory of taste exclusively in terms of objects of nature. Later he turns to the consideration of the exercise of taste in judgments of works of art, developing an elaborate theory of such judgments, but declaring that judgments of beauty made about works of art are inevitably 'impure', because the judge is implicitly aware that the the object was *made*. Kant's argument for the impurity of such judgments is elliptical. It turns on his conviction that a pure judgment of taste does not involve the application of a predicate concept, while the recognition that an object was made involves a recognition that some concept was instrumental in the making of the object, this concept then inevitably figuring in one's taste-judgment of the object (see Kant, I. §12).

3 Sibley

In the mid-twentieth century the idea of taste was reconceived, notably in the work of Frank Sibley and the many commentators on his work. Whereas eighteenth-century theorists conceived taste either as the capacity to feel pleasure in the presence of certain objects or as the capacity to discern the properties of objects,Sibley conceives it as the ability to detect various special properties he calls 'aesthetic properties'. Although it is not clear whether these properties are distinguished by requiring taste for their perception, or whether taste itself is understood as the ability to perceive them,Sibley gives them an uncircular, working characterization by enumerating some of them: they are the properties named by 'graceful', 'elegant' and 'delicate', for instance, and he calls these words 'aesthetic terms'. He is content to identify aesthetic terms, in general, as terms similar to those that he enumerates. Taste thus becomes the ability to apply aesthetic terms, presumably by virtue of one's possession of the associated 'aesthetic concepts'. He also characterizes the exercise of taste independently of any considerations of feeling. Sibley's conception of taste is thus similar to the eighteenth-century idea of taste as an ability to detect properties, but unlike eighteenth-century authors, he regards the relevant properties as members of a special class, and again, unlike those authors, he does not treat the capacity for feeling as an essential aspect of the exercise of taste, while not denying that feelings might be attendant upon the exercise of taste. The success of Sibley's conception is a matter of current debate. Two questions deserve particular attention: (1) Can the ideas of taste and aesthetic properties be characterized independently of each other, or at least characterized in a manner that is not viciously

circular? (2) Is there an adequate criterion for distinguishing aesthetic properties from non-aesthetic properties?

4 Further issues

It may be thought that the only point of interest in the topic of taste arises from an epistemological conviction that judgments in aesthetics must be grounded in feelings, and, therefore, that the status of these feelings is implicated in any possibility of the objectivity of aesthetic judgments. This is not so. The question of one's personal taste retains philosophical interest even if there is an objective, independent standard of beauty, or if there can be no standard even in terms of taste.

In the first case, we imagine that there is some impersonal, objective measure of beauty. We will then think of a person of good taste as someone who has positive feelings about objects of merit (and, perhaps, negative feelings about meritless things). The question arises, should one like the objects of merit and dislike the inferior ones? What kind of obligation is this? Contemporary moral theorists have tended to separate questions of normative assessments of actions from questions of motivation. That is, the question of which action should be carried out is kept independent of the question of whether an agent is moved to do that action. One might think of a morally perfect person - a saint, for example - as someone who is always moved to do, and feels like doing the right thing. Analogously, one might think of an aesthetically perfect person as someone who always likes beautiful things, and only likes beautiful things. The relation of moral judgment to moral motivation is far from clear, but there is no doubt that a remark like this is sensible: 'I knew that it was the right thing to do, but I could not bring myself to do it.' Is there a parallel in the aesthetic case? In other words, is it sensible to say, 'I know that it is a beautiful thing, but I cannot bring myself to care for it'?

If a person making judgments about beauty distinguishes between judging something to be beautiful and merely having a positive feeling for the object, then in what way, if any, are such judgments and feelings related? Could one have positive feelings without being inclined to judge the object beautiful? Could one judge the object beautiful in the absence of any positive feelings of one's own? An explanation of the relation between feeling and judgment - in a single person - seems to be required, even if individual feeling were not a factor in the formulation of a standard of beauty.

In the second case, we imagine that there is no objective measure of beauty, that people making such judgments are only expressing their own feelings and preferences. It may still be wondered whether individual people, exercising their taste in expressions of liking, disliking and preferences, display self-consistency. If these people like something, say, a work of art, and believe that they like it on account of some property the work possesses, then what if they fail to like some other work that possesses the same property? Besides this question of an individual's aesthetic consistency, there is also a question of how to understand the totality of one's taste. It is, after all, the same person that likes one thing, dislikes another, prefers one thing to another, and so on. Since these exercises of taste exhibit and define a person's aesthetical self, it is necessary to ask how one would characterize the 'logic' of that self.

However taste is conceived, and however we are to understand its constitution and expression in a single person, there remains this question: is it better to have good taste? Why? Because it leads one to a life of greater pleasure? Surely that is but a hope, contingent on the objects one might encounter, and threatened by the fact that pleasures of taste are only some of the pleasures available to a person - they might be outweighed by other pleasures, the pursuit of which is blocked by the development and exercise of one's taste. Perhaps the possession of good or delicate taste makes for an improvement in one's well-being, and is necessary for enjoying a better life; it remains to explain why.

See also: Aesthetic concepts; Beauty

TED COHEN

References and further reading

Cohen, T. and Guyer, P. (eds) (1982) *Essays in Kant's Aesthetics*, Chicago, IL: University of Chicago.(Essays on various aspects of Kant's theory, with an editors' introduction which gives a brief description of the theory.)

Dickie, G. (1995) *The Century of Taste: Five Philosophers*, New York: Oxford University Press.(A critical history of the development of the theory of taste in Hutcheson,Gerard,Alison,Kant and Hume, which argues that

Hume's version of the theory is the most defensible.)
Dickie, G., Sclafani, R. and Roblin, R. (eds) (1989) *Aesthetics: A Critical Anthology*, 2nd edn, New York: St Martin's Press.(A useful reprinting ofSibley's original essay, along with some criticism and an excellent guide to the literature. It also contains remarks offered by Sibley for this edition of the anthology.)
Guyer, P. (1993) 'The Standard of Taste and the "Most Ardent Desire of Society"', in T. Cohen, P. Guyer and H. Putnam (eds) *Pursuits of Reason: Essays in Honor of Stanley Cavell*, Lubbock, TX: Texas Tech University Press.(An excellent analysis of Hume's theory, relating it to his other work.)
Guyer, P. (1979) *Kant and the Claims of Taste*, Cambridge, MA: Harvard University Press.(A difficult book, but the best one-volume discussion of Kant's theory of taste.)
Hume, D. (1741 and 1757) 'Of the Delicacy of Taste and Passion' and 'Of the Standard of Taste', both reprinted in Miller, E.F. (ed.) *David Hume, Essays: Moral, Political, and Literary*, Indianapolis, IN: Liberty Classics, revised edn 1987.(Also contains an excellent bibliographical account of the original publication of these essays in the eighteenth century.)
Kant, I. (1790) *Critik der Urtheilskraft*, Berlin: Lagarde & Friederich; trans. W.S. Pluhar, *Critique of Judgment*, Indianapolis, IN: Hackett Publishing Company, 1987.(The most reliable translation ofKant's text, with a somewhat contentious commentary.)
Kivy, P. (1976) *The Seventh Sense*, New York: Burt Franklin.(A very readable history ofHutcheson's aesthetics and its influence in Britain.)
Sibley, F. (1959) 'Aesthetic Concepts', *Philosophical Review* 68: 421-50.(The first exposition of Sibley's thesis.)
Sibley, F. (1963) 'Aesthetics Concepts: A Rejoinder', *Philosophical Review* 72: 79-83.(A clarification of Sibley's thesis by way of a response to early criticism.)
Sibley, F. (1965) 'Aesthetic and Nonaesthetic', *Philosophical Review* 74: 135-59.(An extension of the thesis.)

Artist’s intention

W.K. Wimsatt and Monroe C. Beardsley’s famous paper ‘The Intentional Fallacy’ *(1946) began one of the central debates in aesthetics and literary theory of the last half-century. By describing as a fallacy the belief that critics should take into account the author’s intentions when interpreting or evaluating a piece of literature, they were rejecting an entrenched assumption of traditional criticism - and a natural one, since we normally take it for granted that understanding actions, including acts of speech and writing, requires a grasp of the intentions of the agent. But they were expressing an idea that has been greatly influential; it was a central claim of the ‘new criticism’, while the marginalization of the author is also a marked feature of structuralist and poststructuralist literary theory. Most of the debate over the artist’s intentions - ‘artist’ here being used as a general word for writer, composer, painter, and so on - has centred on their relevance for* interpreting *art works. More particularly, the question has been whether* external *evidence about the artist’s intentions - evidence not presented by the work itself - is relevant to determining the work’s meaning.*

1 Some arguments against the relevance of intention

(1) Wimsatt and Beardsley (1946) argue that the intention of the author is not available to the interpreter of a literary work. More recently, Jacques Derrida has claimed that writing remains essentially readable even when its original context is forgotten, inviting the conclusion that writing is able to fulfil its function without the reader needing to consider the intentions of the writer. In partial support of these claims, it seems reasonable to say, at least, that for many art works (those of Homer, say) we know little about any of the day-to-day thoughts their authors had about them at the time of their composition. Since this seems to present no serious impediment to the process of interpretation and criticism, some theorists conclude that questions of intention are irrelevant to interpretation.

(2) It seems uncontentious that the critic’s basic task is to interpret and assess the work, for its own sake and not as a means of discovering something about the author. That the author had a certain intention is a fact about the author, not the work, and an intention with regard to the work is merely a seed in the author’s mind, until the intention is carried out. If it is successfully carried out, sayWimsatt and Beardsley, the feature is now visible in the work and it is not necessary to consider the author’s intentions, and if not, the feature has not been realized in the work, and so any concern with it is irrelevant to criticism.

(3) It is generally agreed that a work may have ‘personal’ features: it may be cheerful, sentimental, ironic, patriotic. However, Wimsatt and Beardsley say that the work is merely a dramatized response of an implied speaker to some situation. They rightly point out that there is no reason to identify the narrative voice with that of the author, but also suggest, more contentiously, that the tone of the work or any pervasive attitude to be found in it - cheerfulness, irony - are to be imputed to a purely literary or imaginary figure conjured up by the work and, if to the author at all, only by an act of biographical inference. Such an inference, by which we might try to pin down the author’s intentions, would take critics beyond the task of criticism; as critics, their concern is solely with the attitudes portrayed in the work.

(4) Few would dispute that, without generally observed rules of grammar and usage, no particular use of language could be understood. Certainly, the work of an author who uses a private code of ‘intended meanings’ would be entirely opaque. But literature is accessible. Beardsley and Wimsatt conclude that (with the exception of certain ‘private and semi-private’ meanings that may have become part of a word’s history) it is the public rules of syntax and semantics that determine a work’s meaning, and not the intentions of the author. A similar argument has been used by structuralists, for whom meaning depends on the system of ‘differences’ established by words in relation to one another, and on ‘codes’ governing their usage. It is held that a necessary and sufficient condition of interpretation is a knowledge of the system of inter-related words and the application of the codes, and that the author’s intentions are therefore irrelevant.

2 Privacy and the mind

Argument (1) needs to be assessed in the context of recent developments in the philosophy of mind, and especially of the way these have helped advance the debate on other minds. In at least some versions, the argument appears to

presuppose a dualistic conception of the mind by suggesting that a difficulty with intention-based criticism is that it could only work in those cases where authors have left detailed accounts of their intentions regarding their writing. But a widely preferred alternative to dualism - which we shall call the 'explanation' view - holds that we do not form our impressions of the mental states of others solely on the basis of their own reports, but according to the evidence of what they do more generally. We assume that their intentional activities are carried out in accordance with their beliefs and desires, and, starting from our available knowledge of their character and projects, construe those activities in the light of whatever motivational states best explain them. On this view any of their activities can reveal their intentions. Once it is accepted that, in general, someone's mental states show through in what they do, the question arises of why there should be any special difficulty about determining writers' intentions given the evidence of their writing itself. Those wishing to defend argument (1) have to show that the compositional activities of writers lack the evidential character of behaviour in general. They must establish this against the plausible claim that, among the various products of human effort, works of art are, if anything, unusually rich with the signs of their maker's cogitations and interests and so provide a wealth of evidence of the artist's intentions. (This reply to argument (1) can concede for the purpose of argument that no external evidence is available, though it is worth noting that even in the most challenging cases we are more than likely to have some general information about the historical and social circumstances of the artist, while more usually we also have a fair amount of biographical information.)

Opponents of intention-based criticism often point out that critics seldom feel bound by an artist's *avowals* of intention, and override them when they fail to make the best sense of what we find in the work. But this, while true, does not show that intentions are irrelevant to criticism. Proponents of the explanation view will want to point out that one kind of evidence for the presence of an intention can be overridden by another. Depending on the merits of the case, the artist's avowals (and other forms of external evidence) can sometimes be ignored in favour of indications of intention suggested by the configuration in the work itself.

3 Art and experience

Argument (2) rightly assumes that the critic's task is to understand the work itself, and that for a claim about the artist (for example, one concerning the artist's intentions) to be relevant to criticism, the claim must potentially make a difference to the way we experience the work. But this only tells us that the critic's interest in the artist should be for the sake of understanding the work and not vice versa. (An analogy: to inquire whether someone's hand gesture was intended as a greeting, not an insult, is in a sense to look past the gesture to the agent; but the answer would also help the inquirer to see the gesture itself more clearly.)

Argument (2) also seems to assume that a realized intention will automatically make its presence felt in the work, whereas an unrealized one can have no bearing on the experience of the work. These assumptions are called into question by a commonplace of the philosophy of perception, namely, that the way we experience an object is in part a product of our mental orientation towards it. In that case, even if an intention is realized we may sometimes not see that the work is the outcome of that intention unless alerted to the fact by external evidence. And if it is not realized, external evidence about the unrealized intention could conceivably modify and enrich the way the work is experienced. It is worth noting, finally, that the very idea of a straightforward distinction between internal and external evidence is called into question once it is conceded that what is visible in a work is modified by what is known about it.

Turning to argument (3), it must be conceded in its favour that in calling a piece of fiction or a poem cheerful we are not usually saying or implying that the writer felt cheerful at the time of writing it. But in the case of some personal qualities, the presence of the quality in the writing does seem to depend on whether the writer wrote the piece in an appropriate state of mind. Nor does it always seem possible to reconstrue the apparent viewpoint of the writer as a merely notional viewpoint within the work. Perhaps the underlying principle of argument (3) is that in aesthetic matters what counts is whether the work presents an *impression* of a certain attitude, while it is irrelevant whether that happens to coincide with what the artist felt. One difficulty with this is that the impression is likely to vary with what we take the underlying attitude of the artist to have been. A defender of the principle might now say that in that case the work is ambiguous, and that we should always apply the most charitable reading, again disregarding what the artist's actual attitude might have been. But this may be simplistic. Suppose that according to the charitable reading a poem expresses sympathy for the victims of a war, but that external evidence shows that

rather than (a) expressing an actually felt sympathy, the poet had (b) perfectly contrived the mere impression of sympathy in order cynically to exploit a general climate of concern for the victims. Someone tempted by the 'impression' principle should carefully consider whether it really is critically irrelevant that we have a case of (a) or (b). Another difficulty with the 'impression' principle is that some of the features we ascribe to writing cannot *but* be features of the writer. We cannot ascribe virtues like intelligence, perceptiveness or wit to a piece of writing without being prepared to ascribe these to the writer. And similarly, it is hard to avoid the implication that, when we attribute certain shortcomings to writing - pretentiousness, sentimentality, narcissism - we are implicitly criticizing the writer.

4 Meaning and use

Argument (4) turns on the claim that literary works are composed in a public medium. Certainly, poets and novelists do not choose the meanings they give to words but work within the constraints of general usage. But defenders of (4) need more than this to make their case. They need to establish that a knowledge of syntax, semantics and general usage is not merely necessary but also *sufficient* for interpretation. Counting against this is a commonplace of linguistic theory, namely, that when a sentence is encountered in isolation, the semantics and syntax of the language will enable native speakers to grasp only its propositional content, and that until the particular circumstances of its use have shown the user's specific intentions, it has been understood only in an attenuated sense. But perhaps a defender of (4) will argue that *literary* interpretation is not concerned with those aspects of meaning which are associated with intention. This, however, seems implausible, because it threatens to exclude tonal features of writing like irony and sarcasm, as well as the import of metaphors, allusions and the like, all of which depend on the intentions with which words are used. Alternatively, the claim might be that a work of art can be said to have whatever features of meaning most enhance the resulting aesthetic impression, provided that, given its syntactic and semantic character, the work can plausibly be seen or imagined *as* the outcome of intentions associated with that meaning. But this is also an extreme claim. Consider a painting that is a portrait of A. You could treat it as though it were intended as a portrait of B (the effect could be as satisfying as you like) but this would not make it a portrait of B. Returning to literature, a well-known problem case is Blake's use of the phrase 'dark satanic mills' in his poem *Jerusalem*. The phrase has often been interpreted as a reference to the factories of the Industrial Revolution, but it is now commonly accepted that this reading is historically impossible. Suppose, then, that Blake was not indeed referring to factories, but that the impossible reading is, none the less, more satisfying than more plausible ones. One reason for ruling it out, despite its appeal, would be that it is not, strictly speaking, an interpretation ofBlake's poem, since we know that Blake's poem does not refer to factories. Someone might attempt to legitimate the unintended but supposedly more attractive reading of the poem by saying that it is a valid critical activity to view a poem simply as a timeless verbal configuration, ignoring constitutive features it acquired through the context in which it was written. However, there appears to be only a difference of degree between a case like this and, say, the case of a free adaptation of *Macbeth* which transposes the play into a present day political setting with familiar contemporary politicians in leading roles - surely a case of a new play created out of the old. We may still call this an 'interpretation' of *Macbeth*, but then we are talking about a special activity, whose existence, however welcome, is no threat or rival to interpretation as exegesis.

See also: Art, understanding of; Artistic expression; Artistic interpretation §3; Artistic style §4; Barthes, R.; Intention; Structuralism in literary theory

PAUL TAYLOR

References and further reading

Armstrong, D.M. (1970) 'The Nature of Mind', in C.V. Borst (ed.) *The Mind/Brain Identity Theory*, London: Macmillan, 67-79.(A clear exposition and defence of what is referred to, in the discussion of argument (1), as the explanation view of the mind.)

Cioffi, F. (1963) 'Intention and Interpretation in Criticism', reprinted in D. Newton-De Molina (ed.) *On Literary Intention*, Edinburgh: Edinburgh University Press, 1974, 55-73.(A famous response to Beardsley and Wimsatt; particularly relevant to the issues raised by argument (1).)

Davies, S. (1982) 'The Aesthetic Relevance of Authors' and Painters' Intentions', *Journal of Aesthetics and Art Criticism* 41 (1): 65-76.(A good discussion of issues bearing on argument (4).)

Fish, S.E. (1982) 'With the Compliments of the Author: Reflections on Austin and Derrida', *Critical Inquiry* 8 (4): 693-721.(Discusses, with references,Derrida's version of argument (1) mentioned in §2, and offers a relatively

lucid interpretation and defence of a deconstructionist view of literary interpretation.)
Lyas, C. (1983) ‘The Relevance of the Author’s Sincerity’, in P. Lamarque (ed.) *Philosophy and Fiction*, Aberdeen: Aberdeen University Press, 17-37.(Especially helpful for coming to grips with arguments (3) and (4).)
Mele, A.R. and Livingston, P. (1992) ‘Intentions and Interpretations’, *Modern Language Notes* 107: 931-49. (Well informed on relevant work in the philosophy of mind and especially illuminating on argument (1).)
Sturrock, J. (1986) *Structuralism*, London: Paladin, ch. 4, 103-35.(A brief and clear introduction to structuralism as a critical movement. Shows the structuralist version of argument (4).)
Wimsatt, W.K. and Beardsley, M.C. (1946) ‘The Intentional Fallacy’, reprinted in D. Newton-De Molina (ed.) *On Literary Intention*, Edinburgh: Edinburgh University Press, 1974, 1-13.(The *locus classicus* of the ‘intentional fallacy’ debate; presents versions of all the arguments discussed in this entry.)
Wollheim, R. (1980) ‘Criticism as Retrieval’, in *Art and Its Objects*, 2nd edn, Cambridge: Cambridge University Press, 185-204.(Especially insightful on the issues raised by argument (2).)

Arya Samaj

*The Arya Samaj (*ārya-samāj*, 'The Association of Nobles') is a Hindu reform movement founded in 1875 by Swami Dayanand Saraswati (1824-83). Based on the supposition that the true religion of India was put forth in the ancient Vedas, rather than in later epics and cycles of myths, the principal aim of the Arya Samaj is to purge modern Hinduism of beliefs and practices associated with the devotional and mythic literature of India. Condemning the hereditary caste system and dismissing the practice of using icons and idols in worship, the society favoured a more rationalistic, humanistic and nationalistic form of Hinduism as India entered the modern era.*

The Arya Samaj, a prominent Hindu reform movement, was founded in the nineteenth century by Swami Dayanand Saraswati. Also known as Mul Shankar and nicknamed 'the Luther of India', Dayanand was born in 1824 into a Brahmanical family of Gujarat. As a boy he always doubted the divinity of idols of Hindu gods. The death of his sister turned him to pondering the problems of life and death. Consequently, he ran away from home and wandered for many years in search of a guru. At last in Mathura he found a blind teacher named Swami Brijanand, who taught him the philosophical interpretations of the Vedas, the most ancient collection of sacred writings in India.

Through his study of the Vedas, Swami Dayanand became convinced that selfish and ignorant priests had pervaded the Hindu religion during the post-Vedic period. Part of the reason for this corruption, he believed, were the Purāṇas, the eighteen collections of myths and legends of gods and heroes that form the basis of most devotional forms of Hinduism in the medieval and modern periods. These texts, which had informally acquired the status of sacred scripture and the contents of which had become much more familiar to most Hindus than the Vedas, were, Dayanand believed, full of false teachings. Linking priestly corruption with these (for him, apocryphal) Purāṇas, and seeing both priestly and textual corruption as the source of contemporary social problems, Dayanand began in 1863 to preach his doctrines publicly. Twelve years later, in 1875, he established the Arya Samaj (*ārya-samāj*, 'The Association of Nobles') in Bombay.

Dayanand regarded the Vedas as eternal and infallible and laid down his own interpretations of them in a book entitled *Satyārth Prakāś*. He considered the Vedas to be the inspired word of God and the fount of all knowledge. He rejected all later religious thought on the grounds of its conflict with the Vedas. Thus, unlike most traditional Hindu thinkers, Dayanand repudiated the authority of the post-Vedic texts, such as the Purāṇas, and regarded the epics, the *Rāmāyaṇa* and the *Mahābhārata*, as literary treasures and nothing more. His total dependence on the Vedas and their infallibility gives his teachings an orthodox colouring, for in orthodox Hindu thought, scriptural infallibility means that human reason is subordinate to sacred texts. Despite this appearance of orthodoxy, Dayanand's approach in fact has rationalistic and humanistic leanings, because the Vedas, though revealed, are to be interpreted by himself and other human beings. Thus, in the final analysis, Dayanand insisted that individual reason is the deciding factor in arriving at religious knowledge.

In keeping with his rejection of the mythology of the Purāṇas, Dayanand condemned the idolatry associated with their mythology. He also rejected the hereditary system of caste, giving it instead an ethical and occupational interpretation, somewhat as the Buddha had done. He questioned the authority and social superiority of Brahmans, and was against animal sacrifices and long pilgrimages. A confirmed nationalist, Dayanand believed that 'Aryans were the chosen people, the Vedas the chosen gospel and India the chosen land'. So the Arya Samaj looked upon the Vedas as India's 'rock of ages' and coined the slogan 'go back to the Vedas'.

The Arya Samaj, like the Brahmo Samaj, worked courageously to eliminate the social evils that it believed had crept into Hindu society. It was more a social reform movement than a strictly religious or philosophical school. In the area of social reform, the Arya Samaj struggled against child marriage and campaigned to fix the minimum age of marriage for boys and girls at 25 and 16 respectively. It advocated a status for women equal to that of men. Intercaste marriage was encouraged, as was the remarriage of widows. In addition to advocating progressive reforms, the Arya Samaj undertook charitable works during national disasters such as earthquakes, famines and floods, and it opened orphanages and homes for widows, thereby giving a new lease of life to the distressed.

The leaders of the Arya Samaj were aware of the supreme importance of education, but sharp differences arose within the movement over the question of the best system. One faction favoured the ancient system of Hindu education. One of its leaders, Swami Shardhanand, started the ‘Gurukul’ near Haridwar in 1902 to propagate the more traditional ideas. Another faction recognized the value of British education and established a network of ‘Dayanand Anglo-Vedic Schools and Colleges’ for both boys and girls throughout the country. Lala Hansraj played a leading part in this effort.

In order to counter Christian and Muslim missionary activities, the Arya Samaj started the *śuddhi* (purity) movement for the reconversion of those Hindus who had been unwillingly or forcibly converted to Islam or Christianity. They could now be readmitted to Hinduism after passing through a ceremony of purification. While the Arya Samaj’s work in social reforms tended to unite people, its religious work tended, though perhaps unconsciously, to work against the growing sense of national unity arising among the Hindus, Muslims, Parsis, Sikhs and Christians of India.

Though founded in Bombay, the Arya Samaj found its true home in Punjab, and it later spread far and wide over the whole of northern India. After Dayanand’s death in 1883, the work he had begun was continued by a band of followers. Even today, the spirit of the Samaj is active in the towns and villages of India, as well as in communities of Indians who have emigrated to other parts of the world.

See also: Brahmo Samaj

K.S. KUMAR

References and further reading

Dayanand Saraswati, Swami *Satyārth Prakāś*, trans. C. Bharadwaja, *Light of Truht* [sic], New Delhi: Saradeshik Arya Pratinidhi, 1975.(An introduction to the Vedas and their interpretation.)

Dayanand Saraswati, Swami *Ṛgvedādibhāṣyabhūmikā*, trans. Paramanand, *An Introduction to the Four Vedas*, New Delhi: Meharchand Lachhmandas Publications, 1981.(An introduction to the Vedas and their interpretation.)

Lajpat Rai, L. (1967) *A History of the Arya Samaj: An Account of its Origin, Doctrines and Activities, with a Biographical Sketch of the Founder*, Bombay: Orient Longman.(A very good book for the beginner, giving the historical development and basic principles of Arya Samaj, along with a biography of its founder, Swami Dayanand Saraswati.)

Sen, N.B. (1964) *Wit and Wisdom of Swami Dayanand*, New Delhi: New Book Society of India.(Sayings of Dayanand Saraswati classified under 250 subjects.)

Sharma, S.K. (1985) *Social Movements and Social Change: A Study of Arya Samaj and Untouchables in Punjab*, Delhi: B.R. Publishing Corporation.(A monograph describing the impact of the Arya Samaj movement on the untouchables in the northern Indian province of Punjab.)

Asceticism

The term 'asceticism' is derived from the Greek word, askēsis*, which referred originally to the sort of exercise, practice or training in which athletes engage. Asceticism may be characterized as a voluntary, sustained and systematic programme of self-discipline and self-denial in which immediate sensual gratifications are renounced in order to attain some valued spiritual or mental state. Ascetic practices are to be found in all the major religious traditions of the world, yet they have often been criticized by philosophers. Some argue that the religious doctrines that they presuppose are false or unreasonable. Others contend that they express a preference for pain that humans cannot consistently act upon.*

The chief ascetic practices are fasting, sexual continence, living in seclusion, living in voluntary poverty, and inflicting pain upon oneself. Such practices are elements in all the major religious traditions, and at least some of them are found in most of the religions of non-literate people that anthropologists have studied. Some philosophical movements, such as Stoicism, which is not ordinarily classified as a religion, also endorse certain ascetic practices (see Stoicism §19). Ascetic practices are relatively common in Jainism, Tibetan Buddhism, early Christianity and various branches of Hinduism (see Jaina philosophy §3; Buddhist philosophy, Indian §2; Sanctification §1; Hindu philosophy §5); they are relatively uncommon in Confucianism, Zoroastrianism and Judaism. All the major religions condemn extreme forms of asceticism, but pathological excesses have appeared in every tradition.

Some forms of asceticism involve an attitude of detachment from worldly things. This is expressed in the scriptural injunction to be in the world but not of it and in the *Bhagavad Gītā*'s recommendation of renunciation in action but not of action. Max Weber described the disciplined, methodical and controlled pursuit of one's vocation in the world, as service to God that is characteristic of certain kinds of Protestant Christianity, as 'inner-worldly asceticism', because worldly success is valued not for its own sake but only as a confirmation of one's salvation.

In theistic religions the ultimate aim of ascetic practices is to promote union with the deity. Many Christian ascetics take self-inflicted pain to further their identification with the suffering Jesus who is God the Son. In Christianity, ascetic practices also have other functions. One is penitential. Self-denial is considered an appropriate way to make reparation for past misdeeds that have offended God. Another is disciplinary. Unruly bodily appetites are, in the fallen human condition, the source of much wrongdoing, and asceticism can serve to check and subdue or even to extirpate them.

Non-theistic religions propose various ends for ascetic practices. In Hindu traditions, awareness is fostered of one's true self (*ātman*), which is identical with the ground of all being (*Brahman*). Theravāda Buddhism promotes the extinction of desire, which in turn leads to freedom from suffering and the illusions of the phenomenal world and to ultimate liberation (*nirvāṇa*) (see Nirvāṇa). In such traditions, ascetic practices, particularly the self-infliction of pain, also function as penance for past evil deeds in this life or in previous incarnations.

The rationale for a particular ascetic practice varies from one religious tradition to another. Christianity, for example, recommends fasting on the grounds that it produces discomfort which can function either penitentially or as a means to identification with the suffering deity, while Yoga recommends it on the ground that it alleviates discomfort by eventually rendering one oblivious to the body and thus better able to direct energy to meditation (see Buddhism, Yogācāra school of). There are also differences concerning who should engage in the more severe forms of ascetic discipline. Often they are reserved for people with special religious vocations, such as monks or nuns. Ordinary believers are frequently told that the extreme asceticism of saints such as Simeon Stylites, who lived atop a pillar for nearly forty years, is to be admired but not imitated.

It is not surprising that asceticism has not found favour with most of those who subscribe to the worldview of secular modernity. It has been criticized by some important modern philosophers. Bentham (1789) argued that humans cannot consistently practise asceticism because they are by nature motivated to seek pleasure and to avoid pain (see Bentham, J. §2). Typically, however, ascetics do not seek pain for its own sake but rather as a means to something they take to be a great good such as union with God. That humans seek painful medical treatment for the sake of their health suffices to show that they can practise asceticism, including the infliction of pain on

themselves, if they perceive doing so as a means to some good end. Nietzsche represented ascetics as weak people who, being unable to exercise power over the strong, express their will to power by turning it on themselves and exercising power over their own appetites (see Nietzsche, F. §6). He also portrayed asceticism as a means by which religious leaders exercise power over their followers. It seems clear that such things are true in some cases, but history also provides examples of strong ascetics who, like Ignatius Loyola, were capable of exercising power over others yet chose instead to devote themselves to ascetic practices at certain stages of their lives.

Some critics are prepared to concede that moderate asceticism makes sense within religious contexts, but argue that it should be rejected because the religious assumptions offered to justify it are false or unreasonable. For example, it is sometimes alleged that asceticism presupposes an untenable dualism of mind and body or of matter and spirit (see Dualism). There are, however, non-dualistic forms of asceticism, and dualism, though it faces philosophical difficulties, has not been shown to be untenable. It is also alleged that belief in the existence of God or of *Brahman* is false or, at least, irrational belief (see God, arguments for the existence of).

See also: Religion and Morality; Self-control; Sexuality, philosophy of; Suffering; Suffering, Buddhist views of origination of

PHILIP L. QUINN

References and further reading

Bhagavad Gītā (*c*.400-100 BC), trans. E. Deutsch, *The Bhagavad Gītā*, Lanham, MD: University Press of America, 1982.(Presents an ideal of asceticism that has influenced Hindu traditions.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, revised F. Rosen, Oxford: Clarendon Press, 1996.(Contains Bentham's argument against asceticism.)

Hardman, O. (1924) *The Ideals of Asceticism: An Essay in the Comparative Study of Religion*, New York: Macmillan.(A readable but somewhat dated overview, written from an explicitly Christian perspective.)

Kaelber, W.O. (1987) 'Asceticism', in M. Eliade (ed.) *The Encyclopedia of Religion*, New York: Macmillan.(A brief and up-to-date overview of religious asceticism.)

Nietzsche, F. (1887) *Zur Genealogie der Moral*, trans. W. Kaufmann and R.J. Hollingdale, *On the Genealogy of Morals*, New York: Random House, 1989.(The third essay presents Nietzsche's most fully developed critique of ascetic ideals.)

Oman, J.C. (1903) *The Mystics, Ascetics and Saints of India*; repr. New Delhi: Cosmo Publications, 1984.(A classic collection of information about Hindu and Muslim ascetics in India that is very readable if not always credible.)

Weber, M. (1904-05) *Die protestantischen Sekten und der Geist des Kapitalismus*, trans. T. Parsons, *The Protestant Ethic and the Spirit of Capitalism*, London: Routledge, 1985.(Discusses 'inner-worldly asceticism'.)

Wellman, C. (1967) 'Asceticism', in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York: Macmillan. (Contains a brief survey of philosophical arguments for and against asceticism.)

Ash‘ariyya and Mu‘tazila

The Mu‘tazila - literally ‘those who withdraw themselves’ - movement was founded by Wasil bin ‘Ata’ in the second century AH (eighth century AD). Its members were united in their conviction that it was necessary to give a rationally coherent account of Islamic beliefs. In addition to having an atomistic view of the universe, they generally held to five theological principles, of which the two most important were the unity of God and divine justice. The former led them to deny that the attributes of God were distinct entities or that the Qur’an was eternal, while the latter led them to assert the existence of free will.

Ash‘ariyya - named after its founding thinker, al-Ash‘ari - was the foremost theological school in Sunni Islam. It had its origin in the reaction against the excessive rationalism of the Mu‘tazila. Its members insisted that reason must be subordinate to revelation. They accepted the cosmology of the Mu‘tazilites but put forward a nuanced rejection of their theological principles.

1 Historical survey

The Mu‘tazila originated in Basra at the beginning of the second century AH (eighth century AD). In the following century it became, for a period of some thirty years, the official doctrine of the caliphate in Baghdad. This patronage ceased in AH 238/ AD 848 when al-Mutawakkil reversed the edict of al-Ma’mun, which had required officials to publicly profess that the Qur’an was the created word of God. By this time, however, Mu‘tazilites were well established in many other centres of Islamic learning, especially in Persia, and had split into two rival factions, the Basran School and the Baghdad School. Although their links with these two cities became increasingly tenuous, both schools flourished until the middle of the fifth century AH (eleventh century AD), and the Basran School only finally disappeared with the Mongol invasions at the beginning of the seventh century AH (thirteenth century AD). After the demise of the Mu‘tazila as a distinct movement, Mu‘tazilite doctrine - by now regarded as heretical by Sunnis - continued to be influential amongst the Shi‘ites in Persia and the Zaydis in the Yemen.

Al-Ash‘ari (d. AH 324/AD 935) was a pupil of Abu ‘Ali al-Jubba’i (d. AH 303/AD 915), the head of the Basran School. A few years before his master’s death, al-Ash‘ari announced dramatically that he repented of having been a Mu‘tazilite and pledged himself to oppose the Mu‘tazila. In taking this step he capitalized on popular discontent with the excessive rationalism of the Mu‘tazilites, which had been steadily gaining ground since their loss of official patronage half a century earlier. After his conversion, al-Ash‘ari continued to use the dialectic method in theology but insisted that reason must be subservient to revelation. It is not possible to discuss al-Ash‘ari’s successors in detail here, but it should be noted that from the second half of the sixth century AH (twelfth century AD) onwards, the movement adopted the language and concepts of the Islamic philosophers whose views they sought to refute. The most significant thinkers among these later Ash‘arites were al-Ghazali and Fakhr al-Din al-Razi.

2 Cosmology

Popular accounts of the teaching of the Mu‘tazilites usually concentrate on their distinctive theological doctrines. To the philosopher, however, their cosmology, which was accepted by the Ash‘ariyya and other theological schools, is a more appropriate starting point.

To the Mu‘tazila, the universe appears to consist of bodies with different qualities: some are living while others are inanimate, some are mobile while others are stationary, some are hot and some are cold, and so on. Moreover, one and the same body may take on different qualities at different times. For instance, a stone may be mobile when rolling down a hill but stationary when it reaches the bottom, or hot when left in the sun but cold after a long night. Yet there are some qualities which some bodies cannot acquire; for example, stones are invariably inanimate, never living. How are the differences between bodies, and between one and the same body at different times, to be explained?

The answer given by the Mu‘tazilites is that all bodies are composed of identical material substances (*jawahir*) or atoms (*ajza’*), on which God bestows various incorporeal accidents (*a‘rad*). This view was first propounded by Dirar ibn ‘Amr (d. *c.*AH 200/AD 815) and elaborated by Abu al-Hudhayl (d. AH 227/AD 841 or later), both of whom

were early members of the Basran School. Abu al-Hudhayl held that isolated atoms are invisible mathematical points. The only accidents which they can be given are those which affect their ability to combine with other atoms, such as composition or separation, motion or rest. Conglomerates of atoms, on the other hand, can be given many other accidents such as colours, tastes, odours, sounds, warmth and coldness, which is why we perceive them as different bodies. Some of these accidents are indispensable, hence the differences between bodies, whereas others can be bestowed or withdrawn, thus explaining the differences between one and the same body at different times.

This account of the world gained rapid acceptance amongst Islamic theologians, although to begin with it was rejected by two Mu'tazilites of the Basran School, al-Nazzam (d. AH 221/AD 836) and Abu Bakr al-Asamm (d. AH 201/AD 816?). The former, who was Abu al-Hudhayl's nephew, argued that atoms which were mere mathematical points would not be able to combine with one another and that, rather than being composed of atoms, bodies must therefore be infinitely divisible. Abu al-Hudhayl replied that God's bestowal of the accident of composition on an isolated atom made it three-dimensional and hence capable of combining (see Atomism, ancient). Al-Asamm, on the other hand, objected to the notion of accidents, arguing that since only bodies are visible their qualities cannot have an independent existence. Abu al-Hudhayl retorted that such a view was contrary to divine laws because the legal obligations and penalties for their infringement were not directed at the whole person but at one of his 'accidents', such as his prostration in prayer or his flagellation for adultery.

3 The five principles

According to the Muslim heresiographers, who are our main source of information about the Mu'tazila, members of the movement adhered to five principles, which were clearly enunciated for the first time by Abu al-Hudhayl. These were: (1) the unity of God; (2) divine justice; (3) the promise and the threat; (4) the intermediate position; and (5) the commanding of good and forbidding of evil.

The first and second principles are of major importance and will be discussed in detail below. The third principle is really only an adjunct of the second, and is here treated as such. The fourth principle is a relatively unimportant doctrine which probably only figures in the list because it was thought to have been the reason for the Mu'tazila's emergence as a distinct movement; it is said that when Hasan al-Basri was questioned about the position of the Muslim who committed a grave sin, his pupil Wasil bin 'Ata' said that such a person was neither a believer nor an unbeliever, but occupied an intermediate position. Hasan was displeased and remarked, 'He has withdrawn from us (*i'tazila 'anna*)', at which Wasil withdrew from his circle and began to propagate his own teaching. The historicity of this story has been questioned on the ground that there are several variants: according to one version the person who withdrew was Wasil's associate 'Amr ibn 'Ubayd (d. AH 141/AD 761), and according to another the decisive break came in the time of Hasan's successor Qatada. Moreover it is noteworthy that at least one influential member of the Basran school, Abu Bakr al-Asamm, rejected the notion of an intermediate position and argued that the grave sinner remained a believer because of his testimony of faith and his previous good deeds. This was also the view of the Ash'arites.

The fifth principle, which is derived from several passages in the Qur'an (for example, Surah 9: 71), and which the Mu'tazilites understood as an obligation incumbent on all Muslims to intervene in the affairs of state, was rarely put into practice. For the Ash'arites, the commanding of good and forbidding of evil was the prerogative of the head of state, who acted on behalf of the Muslim community.

4 The unity of God

The first half of the *shahada*, the Muslim declaration of faith, is the testimony that there is no god besides Allah. Thus the numerical unity of God is axiomatic for all Muslims. Nevertheless, although the Qur'an explicitly asserts that God is one, and equally explicitly rejects polytheism and the Christian doctrine of the Trinity, it speaks of God's 'hands' (Surah 38: 75), 'eyes' (Surah 54: 14) and 'face' (Surah 55: 27), and of his seating himself on his throne (Surah 20: 5), thus apparently implying that he has a body. Moreover, in describing the radiant faces of believers 'looking towards their Lord' on the Day of Resurrection (Surah 75: 23), it suggests the possibility of a beatific vision.

However, the Mu'tazilites emphatically rejected such notions, insisting that God is not merely numerically one but

also that he is a simple essence. This led them to deny that he has a body or any of the characteristics of bodies such as colour, form, movement and localization in space; hence he cannot be seen, in this world or the next. The Mu‘tazila therefore interpreted the Qur’anic anthropomorphisms as metaphors - God’s ‘hands’ are his blessing, God’s ‘eyes’ are his knowledge, his ‘face’ is his essence and his seating himself on his throne is his omnipotence - and argued that, since the Qur’an elsewhere asserts that ‘sight cannot reach Him’ (Surah 6: 103), the phrase *ila rabbiha nazira* means ‘waiting for their Lord’ rather than looking towards him.

Some of the later Ash‘arites accepted the Mu‘tazilite position on the Qur’anic anthropomorphisms. In al-Ash‘ari’s own view, however, they are neither to be dismissed in this way nor understood to imply that God has a body like human beings. They are ‘revealed attributes’, whose existence must be affirmed without seeking to understand how (*bi-la kayfa*). Furthermore, the possibility of beatific vision depends not on God’s embodiment, but on his existence. God can show us everything which exists. Since he exists, he can therefore show us himself. Hence the statement that ‘sight cannot reach Him’ must apply only to this world, where he impedes our vision.

Much more problematic than the Qur’an’s anthropomorphisms are the adjectives which it employs to describe God. He is said, for instance, to be ‘living’, ‘knowing’, ‘powerful’ and ‘eternal’. If we deny these qualities to God, we must then attribute to him their opposites, which are imperfections. But God is by definition free from imperfections; therefore God must always have had these qualities. But does this mean that he possesses the attributes of ‘life’, ‘knowledge’, ‘power’ and ‘eternity’ and that they are distinct from his essence? The Mu‘tazilites reasoned that this was impossible because it would imply plurality in the Godhead. When we speak of God as ‘living’, ‘knowing’, ‘powerful’ and ‘eternal’, we are, in their opinion, merely considering him from different points of view. God’s ‘attributes of essence’ (*sifat al-dhat*), as they are generally called, are a product of the limitations and the plurality of our own intellectual faculties; in reality, they are identical with God’s essence. Thus, according to al-Ash‘ari (*Maqalat*: 484), Abu al-Hudhayl maintained that ‘God is knowing by virtue of a knowledge which is His own essence’ and that he is likewise powerful, living and eternal by a power, a life and an eternity which are none other than his own essence. Al-Nazzam expressed this even more forcefully when he said, ‘If I say that God is knowing, I merely confirm the divine essence and deny in it all ignorance. If I say that God is powerful, living and so forth, I am only confirming the divine essence and denying in it all powerlessness, mortality and so forth’ (*Maqalat*: 484).

Al-Ash‘ari himself rejected this reductionist account of the ‘attributes of essence’ which made them artefacts of human reason, but his arguments for doing so are far from compelling. He alleged that since in the case of human beings knowing implies possessing knowledge as an entity distinct from the knower, the situation with God must be analogous. Hardly more cogent is the claim that if God knew by his essence, he would be knowledge. Finally, al-Ash‘ari’s assertion that the ‘attributes of essence’ are neither other than God nor identical with him is simply a retreat into paradox. However, al-Ash‘ari was not alone in wishing neither to affirm the independent existence of these attributes nor to deny it outright. Al-Jubba’i’s son Abu Hashim (d. AH 321/AD 933) attempted to resolve the problem by introducing the idea of ‘state’ (*hal*). A state is not something which exists or which does not exist; it is not a thing and it cannot be known in itself, only with an essence. Nevertheless it has an ontological reality. According to Abu Hashim, there are in God permanent states such as ‘his mode of being knowing’ (*kawnuhu ‘aliman*), ‘his mode of being powerful’ and so forth, which give rise to distinct qualicatives. This compromise was accepted by many of Abu Hashim’s fellow Mu‘tazilites of the Basran school, but was unanimously rejected by those of Baghdad.

In addition to the attributes of essence, the Qur’an employs a whole series of adjectives such as ‘providing’ and ‘forgiving’, which describe God in relation to his creatures. It is easy to imagine a time when God did not have these attributes. The Mu‘tazilites called these ‘attributes of action’ (*sifat al-fi‘l*) because they deemed them to come into being when God acts. In their reckoning, God’s ‘speech’ belongs to this category of attributes, for it does not make sense to think of his commandments as existing before the creation of the beings to whom they are addressed. Thus the Qur’an itself, although the Word of God, is temporal and not eternal. It was created initially in the ‘guarded tablet’ (Surah 85: 22) and subsequently recreated in the hearts of those who memorize it, on the tongues of those who recite it and on the written page. Although not denying the existence of attributes of action, al-Ash‘ari insisted that ‘speech’ - along with ‘hearing’ and ‘vision’ - was an attribute of essence. He argued that if God’s word were not eternal, it would have had to have been brought into being. Furthermore, since it is an attribute, it could not have been brought into being other than in an essence in which it resides. In which case

either God brought it into being in himself, or he brought it into being in another. But if he had brought it into being in himself, he would be the locus of things which come into being, which is impossible. If, on the contrary, he had brought it into being in another, it is the other, and not God, who would have spoken by the word.

5 Divine justice and human destiny

In addition to championing the unity of God, the Mu‘tazilites stressed his justice. They held that good and evil are objective and that the moral values of actions are intrinsic to them and can be discerned by human reason. Hence God’s justice obliges him to act in accordance with the moral law. For instance, he is thus bound to stand by his promise to reward the righteous with paradise and his threat to punish the wicked with hellfire. More importantly, the reward and punishment which he metes out must be merited by creatures endowed with free will (see Free will). Thus although the Qur’an says that God guides and leads astray those whom he wills (Surah 14: 4), it cannot mean that he predestines them. This and similar texts refer rather to what will happen after the judgement, when the righteous will be guided to paradise and the wicked will be caused to stray far from it. With regard to our acts in this world, God creates in us the power to perform an act but we are free to choose whether or not to perform it.

Many of the Mu‘tazilites held that the principle of justice made it requisite for God always to do for people what was to their greatest advantage. Al-Jubba’i went as far as to claim that God is bound to prolong the life of an unbeliever if he knows that the latter will eventually repent. In view of this, al-Ash‘ari is alleged to have asked him about the likely fate of three brothers: a believer, an unbeliever and one who died as a child. Al-Jubba’i answered that the first would be rewarded, the second punished and the third neither rewarded nor punished. To the objection that God should have allowed the third to live so that he might have gained paradise, al-Jubba’i replied that God knew that had the child lived he would have become an unbeliever. Al-Ash‘ari then silenced him by asking why in that case God did not make the second brother die as a child in order to save him from hellfire!

For al-Ash‘ari, divine justice is a matter of faith. We know the difference between good and evil solely because of God’s revelation, and not by the exercise of our own reason. God makes the rules and whatever he decrees is just, yet God himself is under no obligation: if he wished, he could punish the righteous and admit the wicked to paradise (see Voluntarism). Moreover, to suppose as the Mu‘tazilites did that human beings had free will would be to restrict the sovereign freedom of the creator. On the contrary, God creates in his creature both the power and the choice; then he creates in us the actions which correspond to these. Nevertheless, we are conscious of a difference between some actions, such as the rushing of the blood through our veins, which are involuntary, and others, such as standing up or sitting down, which are in accordance with our own will. Al-Ash‘ari argues that by approving of these latter actions, which God created in us, we ‘acquire’ them and are thus held responsible for them.

See also: Causality and necessity in Islamic thought; Free will; Islam, concept of philosophy in; Islamic theology; Karaism

NEAL ROBINSON

References and further reading

al-Ash‘ari (before 935) *Maqalat al-islamiyyin (Islamic Dogmas)*, ed. H. Ritter, Wiesbaden, 2nd edn, 1963. (Valuable source of information about earlier thinkers.)

Gimaret, D. (1990) *La doctrine d’al-Ash‘ari (The Doctrine of al-Ash‘ari)*, Paris: Éditions du Cerf. (A systematic and comprehensive treatment of the subject.)

Gimaret, D. (1992) ‘Mu‘tazila’, in *Encyclopaedia of Islam*, New Edition, Vol. VII, fasc. 127-8: 783-93.(Short survey of the topic.)

Hourani, G. (1985) *Reason and Tradition in Islamic Ethics*, Cambridge: Cambridge University Press.(Excellent defence of the Mu‘tazilite position.)

Leaman, O. (1985) ‘Are the Ethics of Religion Objective or Subjective?’, in *Introduction to Medieval Islamic Philosophy*, Cambridge: Cambridge University Press, ch. 4, 123-65.(Critique of the Mu‘tazilite interpretation.)

Nader, A.N. (1984) *Le système philosophique des Mu‘tazila (The Philosophical System of the Mu‘tazila)*, 2nd edn, Beirut: Dar el-Machreq.(Somewhat outdated; apparently a simple reprint of the first edition of 1956.)

Van Ess, J. (1984) *Une lecture à rebours de l’histoire du Mu‘tazilisme (A Controversial Reading of the History of Mu‘tazilism)*, Paris: Geuthner.(Brief and lively discussion of the early period.)

Asmus, Valentin Ferdinandovich (1894-1975)

One of the most accomplished thinkers in the Soviet Marxist tradition, Asmus wrote extensively in many areas of philosophy, and was widely regarded as the Soviet Union's principal Kant scholar. Early in his career, he became associated with the influential school of 'dialecticians' led by A.M. Deborin and produced a number of significant writings in the history of philosophy. When Deborin and his followers were condemned as 'Menshevizing idealists' in 1931, Asmus shifted the principal focus of his work to aesthetics and logic. His 1947 textbook of formal logic subsequently became the principal text for logic instruction in the USSR.

Throughout his long career, Asmus experienced a number of political difficulties. Nevertheless, he avoided imprisonment and published consistently, though he was never permitted to go abroad. His importance in Russian philosophy derives not so much from the significance of his theories, but from his role in preserving philosophical culture in Russia through the Stalin period. He aspired to high standards of scholarship and worked hard to foster the study of logic and the history of philosophy. The breadth of his interests and his excellence as a teacher made him an inspirational figure to the young scholars striving to revive Soviet philosophy in the 1960s.

1 Life

V.F. Asmus was born in Kiev. In 1919, he graduated from Kiev University, where he studied with V.V. Zenkovsky and A.N. Giliarov. Asmus taught philosophy and aesthetics in Kiev for a number of years, until he was appointed to Moscow's Institute of Red Professors in 1927. There his expert appreciation of the history of philosophy made him an ally of the school of 'dialecticians' led by Abram Deborin, which dominated the Soviet philosophical institutions at that time. Asmus participated in the dialecticians' controversies with Soviet positivists, or 'mechanists', and in the later conflicts with Party activists 'on the philosophical front' that led in 1931 to the dissolution of Deborin's school. Although condemned as a 'Menshevizing idealist' and temporarily denied the right to teach, Asmus was spared the fate that befell many Deborinites in the purges, probably because of his 'non-Party' status.

As Soviet philosophy increasingly became an instrument for the propagation of Party ideology, so Asmus shifted the focus of his work to aesthetics and the philosophy of literature, becoming a member of the Writer's Union in 1935 and completing his doctoral dissertation on 'Aesthetics in Classical Greece' in 1940, the first thesis to be defended at Moscow's Institute of Philosophy. Asmus could not, however, elude political controversy altogether. In 1938, he was endangered by his former association with Bukharin, though he fortuitously escaped arrest. And in 1944, Asmus was among the authors of the third volume of the *Istoriia filosofii (History of Philosophy)*, edited by G.F. Aleksandrov *et al*., which the Central Committee withdrew from publication because its sympathetic treatment of German idealism was deemed inappropriate in the climate of anti-German feeling provoked by the Great Patriotic War. It was fortunate for the contributors that they had earlier received the Stalin prize for an earlier volume. Thereafter Asmus, who had been a professor at Moscow University since 1939, sought refuge in its Department of Logic where he made an important contribution to the development of formal logic in Russia. He returned to the history of philosophy only after the Stalin period. A close friend of Boris Pasternak's, Asmus is reputed to have influenced the philosophical content of *Doctor Zhivago*. In 1960, he was once again embroiled in controversy when he gave a eulogy at Pasternak's graveside. Asmus continued to work actively in philosophy until his death.

2 Ideology, culture and explanation

Among the works which brought Asmus into the limelight was his debate with the mechanist A. (Sandor) Var'iash, published in *Pod znamenem marksizma (Under the Banner of Marxism)* (in 1926-7). In his *Istoriia novoi filosofii (History of Modern Philosophy)* (1926), Var'iash argues that if Marxism is to provide a consistently monist account of reality, it must explain how all forms of ideological activity, including literature, science and philosophy, arise from specific socioeconomic relations. By this, Var'iash means not just that philosophical or scientific theories cannot be understood without reference to the historical circumstances of their emergence, but that both the content and the logical structure of theories are ultimately causally determined by the forces of production. Var'iash's work examines how such causal relations can be traced. In reply, Asmus maintains that Var'iash succeeds in demonstrating only that, for any era, a correspondence exists between the social needs

defined by the relations and forces of production and the general themes of scientific research. Asmus argues that it is in principle impossible to derive the logical characteristics of a theory from considerations about the socio-historical conditions of its production, and that Var'iash's position must ultimately collapse into a disastrous relativism, for unless we distinguish between the 'genetic' analysis of a theory's origins and the logical analysis of its content, the concepts of proof and truth will be undermined. In addition, Asmus rejects Var'iash's view that the causal determination between 'economic base' and 'ideological superstructure' proceeds only in one direction.

The controversy is sometimes portrayed as a conflict between Var'iash's 'vulgar sociological' approach and Asmus' cultured historicism. This is misleading, for Var'iash was aware of his position's counter-intuitive aspects and developed his case with considerable ingenuity, while Asmus was sometimes overconfident in his arguments, no doubt because he knew they would find favour with the Deborinite orthodoxy. Nevertheless, the debate is testimony to the vigour and seriousness of Soviet philosophy in the 1920s, and illustrates Asmus defending the irreducibility of cultural phenomena against positivist or crude Marxist conceptions, a position which became a leitmotif of his work.

3 Philosophy and its past

One theme prominent in Asmus' dispute with Var'iash is the nature of philosophy's relation to its past. For Asmus, Marxism represents the outgrowth of a long historical development and incorporates insights from many philosophical positions. The history of philosophy is not simply a history of previous error. Accordingly, Asmus sought throughout his career to uphold standards of historical scholarship in philosophy, both in his own writings on ancient, early modern and nineteenth-century Russian philosophy, and in the many collaborative projects in which he was engaged. He was particularly well known for his studies of Kant, for whose thought he had a special affection.

Asmus' approach to the history of philosophy is exemplified by his *Ocherki istorii dialektiki v novoi filosofii (Essays on the History of Dialectics in Modern Philosophy)* (1930). Here he argues that dialectic represents a method of cognition, designed to capture how development occurs through the resolution of contradiction. Asmus traces the origins of Marx's conception of dialectic in the work of Descartes, Spinoza, Kant, Schelling, Fichte and Hegel. He argues that it was Kant who first grasped the significance of dialectical contradiction, though he mistakenly confined its influence to the realm of thought. Hegel, in contrast, correctly discerned that the development of being itself is dialectical, but his account is belied by his idealist metaphysics and teleological conception of history. It was Marx and Engels, Asmus argues, who identified the true empirical content of Hegel's system and turned dialectic into a method for the analysis of real human history.

Though Asmus' exposition conforms to the standard Soviet vision of philosophy's history, it contains many subtleties, both in the detailed accounts of particular philosophers and in his appreciation of the complexities of Marx's method. Though Asmus invokes familiar dialectical materialist formulas (such as the triad of thesis-antithesis-synthesis) to characterize the structure of dialectical development, he is adamant that all such schema represent crude simplifications. The central feature of Marx's dialectical method, he argues, is its sensitivity to the specific logic of development of particular objects of cognition, a sensitivity which cannot be captured in principles or schema. Thus no substantive content-neutral specification of dialectical method is possible. This view, undeveloped in the Stalin period, was later explored by Soviet thinkers in the 1960s.

For Asmus, the history of philosophy culminates in the final emancipation of humanity: Marxism renders transparent the laws of historical development and permits human beings to harness the forces which have hitherto shaped their lives. The relation of freedom and necessity is thus a pervasive theme in Asmus' historical writings. Like many of his Soviet colleagues, he follows Engels and Plekhanov in maintaining that freedom and necessity are not irreconcilable; we are free when we act consciously in full appreciation of the laws which govern our deeds. The proletariat, the first class to understand the circumstances of its own agency, thus reconciles freedom and necessity in its very practice.

As will be evident, Asmus presents a view of philosophy's development hardly less teleological than Hegel's, and in his more encyclopedic writings he is inclined to represent each pre-Marxist philosopher's contribution as a combination of 'progressive' and 'reactionary' elements, the former contributing to the eventual emergence of Marxism, the latter the result of constraints imposed by the socioeconomic climate in which the philosopher

worked. For example, his entry on Kant in the Soviet *Filosofskaia ėntsiklopediia (Philosophical Encyclopedia)* (1962) concludes by chiding Kant for his ‘agnostic’ denial that we can have knowledge of things-in-themselves, his ‘subjective idealist’ view of cognition as the imposition of a priori categories on the deliverances of experience, and his ‘formalistic’ approach to ethics and aesthetics. At the same time Asmus praises Kant’s insights about dialectical contradiction, his concern with the justification of scientific knowledge, and his view of aesthetics as transcending the conflict between theoretical and practical reason. Asmus’ scholarly writings, in contrast, provide a more nuanced (though no less linear) account of philosophy’s development and transcend the Manichean vision that dominates so much Soviet philosophical writing.

4 Logic and aesthetics

Asmus had an important influence on the study of logic in the USSR. In the 1930s and 1940s, it was widely held that formal logic’s allegiance to the law of non-contradiction made it ill-suited to represent dialectical thinking; formal logic must be supplanted by a higher ‘dialectical logic’. As a result, the study of formal logic was almost totally neglected. Alarmed by this, the Central Committee invited Asmus to write an elementary logic text. The work, published in 1947, was widely used and logic was gradually rehabilitated (though controversies about the relation of formal and dialectical logic continued into the 1970s). Though Asmus felt himself ill-equipped to pursue studies in mathematical logic, he did much to encourage its development at Moscow University. He also continued to write on broadly logical topics, particularly on the concept of proof (1954). In this work, Asmus denies that there is a sharp distinction between logical and empirical argument, and maintains that all axioms and definitions have empirical content and derive their justification ultimately from their role in human practice.

Asmus also made significant contributions to Soviet aesthetics. In addition to studies in classical aesthetics, he produced erudite and insightful pieces on Goethe, Lermontov, Tolstoi and Schiller.

5 Asmus’ legacy

Asmus will be remembered not so much for the novelty of his philosophical ideas as for his contribution to the survival of philosophical culture in Russia during the most oppressive periods of Soviet history. Though his work cannot be compared to that of Losev (who reputedly felt that Asmus had squandered his philosophical talents) and Bakhtin, there is no denying the extent of his influence on his peers and students, particularly the younger generation of philosophers who sought to reanimate Soviet philosophy in the late 1950s and early 1960s. For them, Asmus was an inspiring teacher who represented a living connection with the traditions of the pre-Stalin period and a writer whose breadth of interests and scholarly integrity were unparalleled among Soviet Marxists.

See also: Marxist philosophy, Russian and Soviet

DAVID BAKHURST

List of works

Asmus, V.F. (1924) *Dialekticheskii materializm i logika (Dialectical Materialism and Logic)*, Kiev.(Asmus’ first major work and one which brought him to the attention of Deborin and his followers in Moscow.)

Asmus, V.F. (1926) *‘Spornye voprosy istorii filosofii’ (Controversial Questions in the History of Philosophy)*, in *Pod znamenem marksizma (Under the Banner of Marxism)* 7-8: 206-25.(Asmus’ critique of Var’iash (1926).)

Asmus, V.F. (1927) *‘K spornym voprosam istorii filosofii’ (On Controversial Questions in the History of Philosophy)*, in *Pod znamenem marksizma (Under the Banner of Marxism)*, 1: 165-94.(Asmus’ rebuttal of Var’iash’s (1927) reply to Asmus (1926).)

Asmus, V.F. (1929) *Dialektika Kanta (Kant’s Dialectic)*, Moscow; repr. as *Immanuil Kant*, Moscow, 1973.(An assessment of Kant’s thought. Asmus later claimed that Kant’s philosophy was ’the foundation of his worldview’. The 1973 edition is reworked and expanded.)

Asmus, V.F. (1930) *Ocherki istorii dialektiki v novoi filosofii (Essays on the History of Dialectics in Modern Philosophy)*, Moscow and Leningrad.(A presentation of Marx’s ’synthetic’ dialectic in contrast to the views of Descartes, Spinoza, Kant, Fichte, Schelling and Hegel.)

Asmus, V.F. (1933) *Marks i burzhuaznyi istorizm (Marx and Bourgeois Historicism)*, Moscow and Leningrad.(Asmus traces the origins of Marx’s view of history in Bacon, Herder, Kant, Fichte, Schelling and Hegel, and contrasts it with the irrationalism of Schopenhauer and Spengler.)

Asmus, V.F. (1940s) ‘V.S. Solov’ëv: Opyt filosofskoi biografii’ (V.S. Solov’ëv: A Philosophical Biography), in

Voprosy filosofii (Questions of Philosophy), (1988) 6: 70-89; 'V.S. Solov'ëv: An Attempt at Philosophical Biography', in *Soviet Studies in Philosophy* 28 (1989), 2: 66-93.(The first chapter of an unfinished book on Solov'ëv, written in the 1940s. Asmus stresses the distinctiveness of Solov'ëv's thought, arguing that it represents a new stage in Russian idealism.)

Asmus, V.F. (1947) *Logika (Logic)*, Moscow.(Asmus' influential logic text.)

Asmus, V.F. (1954) *Uchenie logiki o dokazatel'stve i oproverzhenii (The Logic of Proof and Refutation)*, Moscow.(A broad discussion of issues in the philosophy of logic.)

Asmus, V.F. (1956) *Dekart (Descartes)*, Moscow.(An account of Descartes' philosophy.)

Asmus, V.F. (1960) *Demokrit (Democritus)*, Moscow.(A discussion of Democritus' atomism as a precursor of modern materialism.)

Asmus, V.F. (1962) 'Kant', in *Filosofskaia ėntsiklopediia (Philosophical Encyclopedia)*, vol. 2: 418-27.(A systematic presentation of Kant's philosophy, though significantly less sophisticated than Asmus' books on Kant.)

Asmus, V.F. (1963) *Nemetskaia ėstetika XVIII veka (German Aesthetics in the 18th Century)*, Moscow: Iskusstvo.(A scholarly discussion of the aesthetics of Baumgarten, Lessing, Kant and Schiller.)

Asmus, V.F. (1968a) *Voprosy teorii i istorii ėstetiki (Questions in the Theory and History of Aesthetics)*, Moscow: Istussvo.(A collection of essays on Aristotle, Schiller, Pushkin, Tolstoi, Stanislavskii and others; includes the influential 'Krug idei Lemontova' (Lermontov's Circle of Ideas).)

Asmus, V.F. (1968b) *Antichnaia filosofiia (Ancient Philosophy)*, Moscow, 2 vols; 2nd edn 1976.(A widely used textbook on classical philosophy.)

Asmus, V.F. (1969-71) *Izbrannye filosofskie trudy (Selected Philosophical Works)*, Moscow, 2 vols.(A representative selection of Asmus' works with an introduction by the author.)

Asmus, V.F. (1973) *Immanuil Kant (Immanuel Kant)*, Moscow.(A reworking and expansion of Asmus (1929).)

Asmus, V.F. (1975) *Platon (Plato)*, Moscow.(An account of Plato's thought.)

Asmus, V.F. (1988) 'V.S. Solov'ëv: Opyt filosofskoi biografii', in *Voprosy filosofii (Questions of Philosophy)* 6: 70-89; trans. 'V.S. Solov'ëv: An Attempt at a Philosophical Biography', in *Soviet Studies in Philosophy* 28 (1989) 2: 66-93.(The first chapter of an unfinished book on Solov'ëv, written in the 1940s. Asmus stressed the distinctiveness of Solov'ëv's thought, arguing that it represents a new stage in Russian Idealism.)

Asmus, V.F. (1995) 'Dialektika neobkhodimosti i svobody v filosofii istorii Gegelia' (The Dialectic of Necessity and Freedom in Hegel's Philosophy of History), in *Voprosy filosofii (Questions of Philosophy)* 1: 52-69.(A sympathetic discussion of Hegel's reconciliation of freedom and necessity; surveys conceptions of freedom in ancient, early modern and classical German philosophy.)

References and further reading

Aleksandrov, G.F., Bykhovskii, B.E., Mitin, M.B. and Iudin, P.F. (eds) (1941-3) *Istoriia filosofii (History of Philosophy)*, Moscow: Gospolitizdat, 3 vols.(A collective history of philosophy, to which Asmus was a significant contributor. Although the project had earlier won the Stalin prize, the third volume was denounced by the Central Committee for its sympathetic treatment of Western philosophy.)

Sadovskii, V.N. (1993) 'Filosofiia v Moskve v 50-e i 60-e gody' (Philosophy in Moscow in the 1950s and 1960s), in *Voprosy filosofii (Questions of Philosophy)* 7: 147-64.(A valuable account of the climate within Soviet philosophy after the Stalin period, including a discussion of Asmus' contribution.)

Var'iash, A. (1926) *Istoriia novoi filosofii (History of Modern Philosophy)*, Moscow and Leningrad.(The book which provoked the controversy between Asmus and Var'iash in 1926-7.)

Var'iash, A. (1927) 'Monisticheskii vzgliad na istoriiu filosofii i eë spornye voprosy' (The Monist View of the History of Philosophy and its Controversial Questions), in *Pod znamenem marksizma (Under the Banner of Marxism)* 1: 142-64.(Var'iash's reply to Asmus (1926).)

'V.F. Asmus - pedagog i myslitel' (V.F. Asmus - teacher and thinker) (1995), in *Voprosy filosofii (Questions of Philosophy)* 1: 31-51. (Insightful reminiscences of Asmus' life and work by former students and colleagues commemorating the hundredth anniversary of his birth.)

Astell, Mary (1666-1731)

Best known for her proposal to establish a women's college, Astell published on a variety of other topics: religious dissent, the social contract, the marriage contract, epistemic issues, mind-body dualism, immortality, proofs for God's existence, reason and revelation, and Locke's views on 'thinking matter'. Her correspondence with John Norris treated the pure love of God and occasionalism. On marriage she drew a shrewd contrast between the treatment of political tyranny by contractarians (such as Locke), and their failure to deal with domestic tyranny. Some of her reactions to the views of major philosophers anticipated later debates.

1 Philosophical influences on life and works

Mary Astell gained notoriety in her native England and beyond for *A Serious Proposal to the Ladies…* (1694), which advocated the founding of an Anglican academy for women. *Letters Concerning the Love of God* (1695) followed - her side of this correspondence with Cambridge Platonist John Norris was praised by Leibniz. Having failed to get support for her educational institution, Astell offered women a manual for improving their understanding: *A Serious Proposal to the Ladies, Part II* (1697) drew on Lockean views, Cartesian 'method', and the Port Royal *Logic* (see Port-Royal). After examining the marriage contract in *Some Reflections Upon Marriage* (1700), Astell became a pamphleteer on religious and political topics, criticizing Defoe, Swift, Shaftesbury and Locke. (Swift retaliated by satirizing her in the *Tatler*.) *The Christian Religion…* (1705) responded to Locke, Damaris Masham and Archbishop Tillotson. Her circle of friends included Lady Mary Wortley Montagu, for whose posthumously published letters she wrote a preface, and the Anglo-Saxon scholar Elizabeth Elstob.

2 Social and political thought

According to Springborg (1996), the real focus of Astell's *Reflections Upon Marriage* is the 'absurdity of contractarian voluntarism'. Here, and in her pamphlets, Astell wants to show that contractarians such as Locke are reluctant to accept arguments against domestic tyranny on the basis of consent, while they press these arguments against the state. Indeed, Springborg suggests that Astell may be the first systematic commentator on Locke's *Two Treatises of Government*.

In *Moderation Truly Stated…* (1704a) and *An Impartial Enquiry into the Causes of Rebellion and Civil War in This Kingdom* (1704c), Astell defended the royalist cause and the Established Church, while challenging the work of James Owen and Charles Davenant; *A Fair Way with the Dissenters…* (1704b) attacked Defoe; *Bart'lemy Fair or an Enquiry after Wit* (1709) criticized Shaftesbury's *Letter Concerning Enthusiasm* and charged Swift with irreligion.

3 The rejection of occasionalism

In *Letters Concerning the Love of God*, Norris had argued that God should be the sole object of our love and that Malebranche's occasionalism ensures this conclusion (see Occasionalism): all our perceptions and sensations, all our sources of pleasure, are caused solely by God. Astell had reservations about this view and, before allowing the correspondence to be published, appended a letter arguing specifically against occasionalism. Her criticisms are prefigured in medieval debates on secondary causation. The first is that 'it seems more agreeable to the Majesty of God, and that Order he has established in the World, to say that he produces our Sensations mediately by his Servant Nature, than to affirm that he does it immediately by his own Almighty Power' (1695: 281-2). Although Aquinas' *Summa Contra Gentiles* and Molina's *Concordia* had advanced similar arguments, this objection is not conclusive (see Aquinas §3; Molina, L. de). For occasionalists, to hold that God concurs with secondary causes is to hold that God's own causal role in natural change is not sufficient to determine effects in all their specificity. They can therefore claim that their doctrine is more consistent with God's majesty than that of concurrence.

Astell's other objection is more powerful: if bodies contain 'nothing in their own Nature to qualify them to be instrumental to the Production of… sensations' (1695: 279), and if God is sufficient to cause these natural changes, then it seems God has created bodies in vain - which would be contrary to God's perfection. Aquinas had stated a like objection in *Summa Contra Gentiles*.

In place of occasionalism, Astell maintains that sensation is directly caused by mind-body interaction, and

mediately caused by God. She goes beyond this Cartesian account to suggest that something like More's Neoplatonic 'plastic part of the soul' might explain the agreement between external objects and sensations (see More, H.).

4 Other metaphysical topics

The Christian Religion… attempts (1) to show how religious belief can and should be grounded in reason, (2) to determine the roles of reason and revelation, (3) to encourage women to examine Christian doctrines rationally, and (4) to examine the obligations that determine Christian practice. Towards these ends, Astell critically evaluates Archbishop Tillotson's sermons, Locke's *The Reasonableness of Christianity*, Masham's *A Discourse Concerning the Love of God*, and the anonymous *Ladies' Religion*. In place of the Socinianism and scepticism she finds there, she offers her own rational accounts of revelation and Christian practice, which frequently rely on substantive metaphysical arguments (see Socinianism). For example, in addition to an ontological argument for God's existence, Astell offers a cosmological argument that turns on a 'causal likeness principle' different from that of Descartes. Her argument is vulnerable, none the less, to the well-known criticisms of Descartes' version.

She also offers a two-part argument for the immortality of the human mind: first she argues for the immortality of immaterial things, and then offers a 'real distinction' argument to prove that the mind contains none of the properties of extended matter. Her Platonic first argument bears a striking resemblance to that of Leibniz for the 'natural indestructibility' of monads:

> A Being is Mortal and Corruptible, or ceases to Be, when those Parts of which it consists… are no longer thus or thus United…. Hence it follows, That a Being which is Uncompounded, which has no Parts, and which is therefore incapable of Division and Dissolution, is in its own Nature Incorruptible…. If then the Mind be Immaterial, it must in its own Nature be Immortal.
>
> (Astell 1705: 248-9, §§257-8)

In response to the possibility of divine annihilation, she adds: 'He does nothing in vain, and can't be suppos'd to Make a Creature with a design to Destroy or Unmake it'. In addition to depending upon speculation about the 'purposes of God', this response also does not address an issue which also remained for Leibniz: the possibility of instantaneous *natural* annihilation.

Astell's version of the real distinction argument is especially of interest in light of twentieth-century interpretations of Descartes' argument as an epistemic one (Wilson 1978). She reasons:

> When two Complete Ideas… have different Properties and Affections, and can be consider'd without any Relation to, or Dependance on each other, so that we can be sure of the Existence of the one, even at the same time we can suppose that the other does not Exist, as is indeed the case of a Thinking and of an Extended Being, or of Mind and Body; here these two Ideas, and consequently the things they represent, are truly Distinct and of Different Natures.
>
> (Astell 1705: 248-9, §§257-8)

Astell regarded the foregoing arguments and their theological consequences as threatened by the possibility of Locke's 'thinking matter'. She notes that according to Locke's *Essay*:

> it is impossible for *a Solid Substance to have Qualities, Perfections, and Powers, which have no Natural or Visible Connexion with Solidity and Extension*; and since there is no Visible Connexion between Matter and Thought, it is *impossible for Matter*, or *any Parcels of Matter to Think*.
>
> (Astell 1705: 259, §267)

This view parallels what Margaret Wilson has called Locke's 'official position': all properties and powers of an object 'stand in comprehensible or conceivable relations to its Boylean "primary qualities"' (1979: 144). Like Wilson, Astell argues that this position is inconsistent with Locke's view in his third letter to Stillingfleet: '*Some Parcels of Matter be so order'd by Omnipotence as to be endued with a Faculty of Thinking*' (Astell 1705: 259, §267). Astell concludes that God's superaddition of thought to matter could only create an 'Arbitrary Union' between matter and the faculty of thought, so that 'it is not Body that Thinks, but the mind that is United to it, Body being still as incapable of Thought as it ever was' (1705: 261, §269).

EILEEN O'NEILL

List of works

Astell, M. (1694) *A Serious Proposal to the Ladies, For the Advancement of their true and greatest Interest. By a lover of Her Sex*, London; repr. in *A Serious Proposal to the Ladies Parts I and II, by Mary Astell*, ed. P. Springborg, Brookfield, VT: Pickering & Chatto, 1996.(An important early English text concerning women's education, which advocates the founding of an Anglican academy for women.)

Astell, M. (1695) *Letters Concerning the Love of God, Between the Author of the Proposal to the Ladies and Mr. John Norris: Wherein his late Discourse, shewing That it ought to be intire and exclusive of all other Loves, is further cleared and justified. Published by J. Norris, M.A. Rector of Bemerton near Sarum*, London.(Astell's correspondence with the Cambridge Platonist who defends Malebranche's occasionalism; her appended letter makes clear her reasons for rejecting occasionalism.)

Astell, M. (1697) *A Serious Proposal to the Ladies, Part II. Wherein a Method is offer'd for the Improvement of their Minds*, London; repr. in *A Serious Proposal to the Ladies Parts I and II, by Mary Astell*, ed. P. Springborg, Brookfield, VT: Pickering & Chatto, 1996.(A manual for improving the understanding, designed for women; it draws on the views of Locke, Descartes and the Port-Royal logicians.)

Astell, M. (1700) S*ome Reflections Upon Marriage, Occasion'd by the Duke and Duchess of Mazarine's Case; which is also consider'd, London; Reflections Upon Marriage. The Third Edition. To which is Added A Preface, in Answer to some Objections*, London, 1706.(An examination of women's position in the 'estate of marriage'; it argues that contractarians use arguments from political tyranny against the state, but are reluctant to deal with domestic tyranny.)

Astell, M. (1704a) *Moderation truly Stated: Or, A Review Of A Late Pamphlet Entitul'd, Moderation a Vertue. With a Prefatory Discourse To Dr. D'Aveanant, Concerning His late Essays on Peace and War*, London.(An examination of the relation between church and state, by way of a detailed rebuttal of the defence of 'occasional conformity' to the Church of England; Astell argues for strict conformity.)

Astell, M. (1704b) *A Fair Way With The Dissenters and Their Patrons. Not Writ by Mr. L[esle]y, or any other Furious Jacobite whether Clergyman or Layman; but by a very Moderate Person and Dutiful Subject to the Queen*, London.(An attack against Defoe's satirical defence of religious toleration.)

Astell, M. (1704c) *An Impartial Enquiry Into The Causes of Rebellion and Civil War In This Kingdom: In an Examination of Dr. Kennett's Sermon, Jan. 31. 1703/4. And Vindication of the Royal Martyr*, London.(A royalist defence, arguing that Whigs and religious dissenters erode the authority of the state by placing supreme power in the people.)

Astell, M. (1705) *The Christian Religion, As Profess'd by a Daughter Of The Church of England*, London.(A response to Locke and Masham.)

Astell, M. (1709) *Bart'lemy Fair: Or An Enquiry after Wit in which due Respect is had to a Letter Concerning Enthusiasm, To my Lord * * *. By Mr. Wotton*, London.(An attack on Shaftesbury's argument that the application of good sense and pointed ridicule, rather than the elimination of liberty of conscience, are the best ways to deal with religious fanaticism.)

Astell, M. (1965) Preface to the Embassy Letters, in *The Complete Letters of Lady Mary Wortley Montagu*, ed. R. Halsband, Oxford: Clarendon Press, vol. 1, appendix 3.

Astell, M. (1996) *Astell: Political Writings*, ed. P. Springborg, Cambridge: Cambridge University Press.(This contains a useful introduction and bibliography, as well as the best modern editions of *Reflections upon Marriage*, *A Fair Way with the Dissenters and their Patrons*, and *An Impartial Enquiry into the Causes of Rebellion*.)

References and further reading

Atherton, M. (1993) 'Cartesian Reason and Gendered Reason', *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, ed. L. Antony and C. Witt, Boulder, CO: Westview Press.(Discusses Astell's and Damaris Masham's views about reason and its relation to gender.)

Ayers, M. (1991) *Locke: Epistemology and Ontology*, London: Routledge.(Contains important objections to Astell's charge of inconsistency in Locke's view of thinking matter.)

Ballard, G. (1752) *Memoirs of several Ladies of Great Britain…*, Oxford: W. Jackson; repr. ed. R. Perry, Detroit, MI: Wayne State University Press, 1985.(The only source of biographical data from someone roughly contemporary to Astell.)

Ferguson, M. (ed.) (1985) *First Feminists: British Women Writers 1578-1799*, Bloomington, IN: Indiana

University Press, 1985.(Places Astell’s social philosophy about women in its historical context; includes excerpts from *A Serious Proposal and Some Reflections Upon Marriage*.)

Kinnaird, J. (1979) ‘Mary Astell and the Conservative Contribution to English Feminism’, *The Journal of British Studies* 19 (1): 53-75.(An introduction to Astell’s social and political views.)

Perry, R. (1986) *The Celebrated Mary Astell: An Early English Feminist*, Chicago, IL: University of Chicago Press.(A detailed social and intellectual history of Astell and her times; appendices include some of Astell’s letters and poems.)

Smith, F. (1916) *Mary Astell*, New York: Columbia University Press.(The first major study of Astell; it provides a good survey of her thought.)

Smith, H. (1982) *Reason’s Disciples: Seventeenth-Century English Feminists*, Urbana, IL: University of Illinois Press.(An examination of seventeenth-century social and political philosophy, and the women writers of that period who, like Astell, foreshadowed the feminism of Wollstonecraft.)

Squadrito, K.M. (1987) ’Mary Astell’s Critique of Locke’s View of Thinking Matter’, *Journal of the History of Philosophy* 25 (3): 434-9.(A critical treatment of Astell’s objections to Locke’s position.)

Squadrito, K.M. (1991) ‘Mary Astell’, *A History of Women Philosophers*, vol. 3, ed. M.E. Waithe, Dordrecht: Kluwer Academic Publishers. (A survey of Astell’s philosophical contributions.)

Wilson, M.D. (1978) *Descartes*, London: Routledge & Kegan Paul.(Offers an epistemic reading of Descartes’ real distinction argument which parallels that of Astell.)

Wilson, M.D. (1979) ‘Superadded Properties: The Limits of Mechanism in Locke’, *American Philosophical Quarterly* 16 (2): 143-50.(As in *The Christian Religion*, this piece charges Locke with inconsistent views regarding thinking matter.)

Atheism

Atheism is the position that affirms the nonexistence of God. It proposes positive disbelief rather than mere suspension of belief. Since many different gods have been objects of belief, one might be an atheist with respect to one god while believing in the existence of some other god. In the religions of the west - Judaism, Christianity and Islam - the dominant idea of God is of a purely spiritual, supernatural being who is the perfectly good, all-powerful, all-knowing creator of everything other than himself. As used in this entry, in the narrow sense of the term an atheist is anyone who disbelieves in the existence of this being, while in the broader sense an atheist is someone who denies the existence of any sort of divine reality. The justification of atheism in the narrow sense requires showing that the traditional arguments for the existence of God are inadequate as well as providing some positive reasons for thinking that there is no such being. Atheists have criticized the traditional arguments for belief and have tried to justify positive disbelief by arguing that the properties ascribed to this being are incoherent, and that the amount and severity of evils in the world make it quite likely that there is no such all-powerful, perfectly good being in control.

1 The meaning of ‘atheism’

As commonly understood, atheism is the position that affirms the nonexistence of God. So an atheist is someone who disbelieves in God, whereas a theist is someone who believes in God. Another meaning of ‘atheism’ is simply nonbelief in the existence of God, rather than positive belief in the nonexistence of God. These two different meanings are sometimes characterized as positive atheism (belief in the nonexistence of God) and negative atheism (lack of belief in the existence of God). Barring inconsistent beliefs, a positive atheist is also a negative atheist, but a negative atheist need not be a positive atheist. One advantage of using ‘atheism’ in these two different senses is that negative atheism, but not positive atheism, characterizes the position of the logical positivists, who hold that statements purportedly about God, including the statement ‘God does not exist’, are cognitively meaningless. If one holds that the statements ‘God exists’ and ‘God does not exist’ are cognitively meaningless, and therefore neither true nor false, one cannot consistently believe that it is true that God does not exist or that it is true that God does exist. So the logical positivist cannot espouse positive atheism, but can be characterized as espousing negative atheism. Nevertheless, since the common use of ‘atheism’ to mean disbelief in God is so thoroughly entrenched, we will follow it. We may use the term ‘non-theist’ to characterize the position of the negative atheist. So instead of saying that the logical positivist is a negative but not a positive atheist, we shall say that the logical positivist is a non-theist but not an atheist.

Since human beings have worshipped many different gods, what god or gods is it whose existence, if denied, makes one an atheist? Generally, it is the dominant or official god of one’s country or culture that plays that role. (Early Christians were called atheists because they rejected belief in the official gods of the Roman state.) But given that there are a number of different conceptions of the divine in a given culture, it is best to distinguish a restricted or narrow sense of ‘atheism’ and ‘theism’ from a broader sense. In the major religions of the West - Judaism, Christianity and Islam - the traditional conception of God is of a purely spiritual, supernatural being who is the perfectly good, all-powerful, all-knowing creator of everything other than himself. As used in this entry, an atheist in the narrow sense of the term is anyone who disbelieves in the existence of this being, just as a theist in the narrow sense is anyone who believes in the existence of this being. In the broader sense, a theist is someone who believes in the existence of any divine being or divine reality, even if it is quite different from the idea of God just described. Similarly, an atheist, in the broader sense of the term, is someone who disbelieves in every form of deity, not just the God of traditional Western theology.

To avoid confusion, it is important to keep in mind both the narrow and the broader senses of these terms. In the narrow sense, the Protestant theologian Paul Tillich (§2) was an atheist, for he disbelieved in the existence of the God of traditional theism. But in the broader sense he was a theist, since he believed that there is a divine reality, being-itself (the God beyond the traditionalistic theistic God). The chief concern of this entry will be an investigation of the reasons supporting atheism in the narrow sense. While someone may readily undertake to give reasons for thinking that the God of traditional Western theology does not exist, it would be a vastly larger task to review all the ideas of the divine that human beings have generated over time and then undertake to justify belief in the nonexistence of each divine being or divine reality.

2 Historical sketch of Western atheism

Perhaps the best way to understand the struggle between atheism and theism is to note theism's insistence on an agent explanation of various natural phenomena, including the existence of the universe. We typically explain our actions and their results in terms of our purposes and our power as agents to make things happen. When the idea that the sun, moon and stars are themselves agents was abandoned, it seemed reasonable to explain their movements, and other natural happenings in the world, as the result of powerful agents (gods) acting upon inert material bodies. Thus the gods served to explain events in nature for which no other explanation was then available, particularly events directly affecting human welfare. And by worshipping and beseeching the gods, human beings undoubtedly hoped to influence the course of natural events in their favour.

The seeds of atheism in Western society were sown with the beginning of science. For the trend of science over the centuries has been to replace explanations of natural events by the activity of divine agents with explanations by means of other natural phenomena (see Religion and science §§1-3). As early as Epicurus, one finds an explicit attempt to rule out any explanation of natural phenomena by reference to the activity of supernatural agents (see Epicureanism §§8-9). But while it is one thing to observe the steady retreat of the gods from a significant place in explanations of phenomena within nature, it is quite another thing to discredit the view that the natural universe itself owes its existence to the creative activity of a supernatural deity. Moreover, the appearance of design in plants and animals made it difficult to find a plausible explanation in something other than the activity of an intelligent being. So, although the growth of the natural sciences did much to remove human dependence on the gods to explain events within nature, a foothold in ultimate explanations still remained; and certain phenomena in nature, particularly the apparent design in plants and animals, continued to suggest an intelligent being exercising a causal influence within nature. An attack on the need and possibility of ultimate explanations was left to philosophers, particularly to Hume (§6) and Kant (§8), although the argument from design received its most devastating blow from the work of Darwin and Wallace in the nineteenth century. It is not surprising, therefore, that atheism has flourished in the nineteenth and twentieth centuries, being advocated by such influential thinkers as Feuerbach, Nietzsche, Marx, Freud, Russell and Sartre.

3 Justification of atheism: insufficiency of proofs of the existence of God

As noted above, this entry will focus primarily on the reasons given in support of atheism in the narrow sense, the view that the God of traditional Western theology does not exist. If atheism in the narrow sense is to be established or shown to be probably true, two different sorts of reasons need to be given. First, one must give reasons that would be sufficient to justify disbelief in God if they constituted the only reasons we have that bear on the question of God's existence. Second, one must refute or show the insufficiency of reasons that have been given to justify belief in God. The necessity of the second undertaking is due to the possibility that reasons in support of a claim may outweigh reasons against the claim. The necessity of the first undertaking is due to the general principle that in the absence of good reasons in support of a claim the proper response is suspension of belief rather than disbelief. However, if we were to have sufficient reason to believe that the God of traditional Western theology (if he existed) would make available to us clear evidence of his existence, we would be entitled to adopt atheism once it had been shown that the reasons given in support of theism are inadequate. But it has been argued that if God made it altogether clear that he exists we would not be cognitively free in relation to him - that is, we would not be independent autonomous persons free to make our way in the world with or without God. Since it is difficult to know that this is not so, and difficult to know that God's purpose for us would not involve our being cognitively free in relation to him, we cannot be confident that the absence of sufficient reasons to believe in God warrants disbelief in God rather than suspension of belief. Thus the justification of atheism requires reasons in support of atheism as well as refutation of purported sufficient reasons in support of theism.

One popular argument for theism relies on the claim that belief in some sort of deity is universal among humans. What better explanation of this fact than that theism is true and God has implanted some recognition of himself, however poorly grasped, in the peoples of the world? Against this argument there are two decisive objections. First, if the argument supports anything, it more directly supports polytheism than theism, for the peoples of the world tend to believe in many distinct deities rather than the single supernatural, perfectly good, all-powerful, all-knowing being of traditional theism. Second, there are plausible naturalistic explanations of the near universal belief in some deity or other. For example, it has been argued that before our ancestors gained any significant

control over nature - or at least an understanding of its workings - they personified the forces of nature, made gods of them, and thus sought to control nature by praying to the gods. Such naturalistic explanations of the near universal belief in gods add to the implausibility of using this belief as support for the existence of the theistic god. In addition to such popular arguments, there are the more serious arguments advanced by philosophers and theologians. The most important of these are the argument from the idea of God, the argument from the existence of the world, the argument from the existence of a world supportive of life and consciousness, the argument from objective moral values and the argument from religious experience.

The argument from the idea of God as a being exhibiting every perfection in the highest possible degree (the ontological argument) has been found wanting by most philosophers. In its place, a modal version has been advanced that rests on the premise that it is logically possible that a being having all perfections in the highest degree exists in every possible world. But if it is logically possible that there is at least one possible world in which every being has some minor defect, the premise of the modal version is false. So although the argument is valid, there is no good reason to think it is sound.

The argument from the existence of the world (the cosmological argument) reasons that the world is contingent; it either might or might not have existed. Since whatever exists but might not have existed must owe its existence to some being that brought it into existence, the world owes its existence to a being that produced it. If the being who produced the world were itself contingent, then it would be part of the world and could not be the cause of the world. Therefore, the being who produced the world does not owe its existence to anything else; it exists necessarily. This argument faces two difficulties. First, if it is correct, it establishes only a necessary cause of the contingent world. It does not establish that the necessary cause of the world has the properties definitive of the theistic God. Attempts to carry out this further task have not been particularly successful. Second, the argument rests on a rather strong but unsubstantiated causal principle: every contingent thing (or collection of such things) has a cause of its existence.

The argument from the existence of a world supportive of life and consciousness is the successor to an earlier argument from apparent design in plants and animals (the design argument), an argument that has been seriously weakened, if not damaged beyond repair, by the naturalistic explanation based on Darwin’s theory of evolution through natural selection. According to the new argument, we know that a universe with initial conditions and laws permitting the emergence of life and consciousness is only one of vastly many ways the universe might have been. Had it been slightly different in any one of ever so many ways, life and consciousness would not have been possible. For example, had the rate of expansion after the Big Bang been slightly slower or slightly faster, life could not have occurred. Thus, it is extremely improbable that life should have emerged, given the vast number of alternative ways things could have gone. Of course, the initial conditions and order necessary for the emergence of life could have occurred randomly, but on the hypothesis of theism it is much more likely that these conditions and this order would occur. So, given that they have occurred, theism is more likely than any purely naturalistic account. This argument has merit provided we assume that our universe is a one-time affair. But if the present universe is only one of vastly many universes that have emerged and disappeared, it would not be so improbable that at least one such universe should contain the initial conditions and laws permitting the emergence of life and consciousness. And that it is our universe that contains these conditions and laws is, of course, not unlikely at all. For if it did not, we would not be here to ask why it does (see Religion and science §6).

The moral argument is generally presented in three steps. First, it is argued that there are objective moral truths. By this it is meant that independent of what we may feel or think, some acts just are morally wrong, others morally right (see Moral realism). Second, it is argued that these moral truths about right and wrong depend on an ultimate superhuman moral lawgiver who establishes these truths. Finally, it is argued that without some being who brings it about that human happiness is proportioned to our obedience to the demands of morality, we would lack sufficient incentive to be moral (see Religion, history of philosophy of §5). While this argument has always had some appeal, it has been severely criticized. It is argued that even the will of an omnipotent, perfectly good being cannot be the ultimate source of morality. For something is not morally right because an omnipotent, perfectly good being commands it; rather such a being would command it because it is morally right. Also, it is argued that there is no inconsistency in happiness being disproportionate to the moral life. For all we know, morality may be its own reward.

The argument from religious experience rests on the principle that if a person has an experience that seems to be of a particular object, they are justified in thinking that the object exists unless there is some reason for thinking otherwise. The application of this principle to experiences that seem to be of God gives grounds for thinking that God exists (see Swinburne 1979). Against this argument we should note three points. First, we know in advance roughly what an object such as a dog or a cat would look like were we to have a perception of it. Few people have any idea how the theistic God would appear if they were to perceive him. Second, in the case of physical objects we have some reasonably clear idea of how to discover reasons for thinking our perceptual experience was not a genuine perception of a physical object. Examples of such reasons might be finding that others who were in a position to see or hear it report seeing or hearing nothing, or determining that the supposed object is of a sort (a fire-breathing dragon, for instance) that we have good reason to think does not exist. Because we are not in such a position with regard to the supposed object of religious experience, the fact that it is difficult to discover reasons for thinking such experiences are not genuine perceptions provides little or no reason to think they are genuine. Third, religious experience seems incapable of establishing that the object of the experience has some of the properties (being creator of the world, being infinitely powerful, being eternal, and so on) essential to the God of traditional theism. Thus, many philosophers believe that the argument from religious experience provides little support for the view that the God of traditional theism exists (see Religious experience).

It is generally conceded that none of these arguments constitutes a proof of the existence of the God of traditional theism. Nor does any one of them appear to make the existence of this being so probable as to warrant belief. Some philosophers contend that considered together they provide sufficient grounds to warrant belief in theism. But critics insist that even this judgment is too strong. Some critics suggest that at best the arguments make it likely that some sort of divine intelligence had something to do with the ordering of the universe, but point out that this is a far cry from having grounds to believe that there is an omnipotent, omniscient, perfectly good creator. They claim that unless we have something else to go on we simply lack sufficient reason to believe in the God of traditional Western theology.

4 Justification of atheism: disproving the existence of God

What of the other side? Are there reasons to think that this being does not exist? Supporters of atheism typically focus on two issues: conceptual difficulties in the perfections ascribed to the God of traditional theism, and the problem of evil. With respect to the former, there are serious difficulties in formulating adequate accounts of omniscience and omnipotence. There are also formidable arguments to the effect that divine omniscience is inconsistent with immutability (an attribute long considered essential to the theistic God), and that God's essential moral perfection is incompatible with any significant divine freedom with respect to whether to create and what world to create (see Omniscience §2; Freedom, divine). Finally, there is the general problem of whether those divine perfections that vary in degrees (knowledge, power, goodness) have an upper limit. But while these difficulties raise genuine doubts as to whether the traditional theistic conception of God is coherent, many philosophers believe that they fall short of a proof of incoherence.

The fundamental issue in the problem of evil is this. Do we have good reason to think that evils occur in the world that an omnipotent, omniscient and perfectly good being would not be justified in permitting? In so far as we do have good reason to believe this, we have a good reason to believe that atheism is true. When we consider horrendous evils or the sheer magnitude of human and animal suffering, the idea that an omnipotent, omniscient, perfectly good being is in control of the world may strike us as absolutely astonishing, something almost beyond belief (see Rowe 1979). What could possibly justify his permitting such monstrous evils? Theists generally accept the view that there must be some good or goods whose realization justifies his permission of these monstrous evils. Some theists propose theodicies that purport to tell us what these goods may be. Other theists say that we do not understand what these goods might be and, furthermore, we should not expect to understand. After all, God's mind is infinite and the goods in question are likely to be very complex and thus beyond our comprehension. Critics respond that God could at least make his presence known and provide some direct assurance, for example, to the mother whose child has been raped, tortured and brutally murdered, assuring her that all this serves some good purpose. Would that not go some way towards comforting the mother in her suffering? Yes, some theists reply, but the mother's sense of the absence or silence of God is just one more evil that God permits in order to realize some justifying goods. But what are these goods and why must God permit such evils in order to realize them? After all, he is omnipotent. Theists reply that with our finite minds we should not expect to understand.

Although the proponent of the argument from evil may be able to show that it is quite likely, if not certain, that no goods we know of justify God in permitting all these horrendous evils, no one can provide a proof that there are no goods realizable in some future state of the world - an afterlife - that would be sufficient to justify the theistic God in permitting the monstrous evils that occur in this life. But the proponent of the argument from evil does argue that it is rational to believe that there are no such goods as these. Indeed, the atheist typically thinks that the argument from evil is of sufficient strength to justify belief in atheism (in the narrow sense) provided the case for theism is itself inadequate. But few theists share this judgment.

See also: Agnosticism; Evil, problem of; God, arguments for the existence of; God, concepts of

WILLIAM L. ROWE

References and further reading

Flew, A. (1976) *The Presumption of Atheism*, London: Pemberton Publishing Co. (Argues that the burden of proof is on the theist in the debate between theism and atheism.)

Gale, R. (1991) *On the Nature and Existence of God*, Cambridge: Cambridge University Press.(Sophisticated critique of the consistency of the divine attributes and of the major arguments for the existence of God. Includes a discussion of prudential arguments for belief in God.)

Mackie, J.L. (1982) *The Miracle of Theism*, Oxford: Oxford University Press.(A sophisticated study of the arguments for and against the existence of God. An important critique of theism.)

Martin, M. (1990) *Atheism: A Philosophical Justification*, Philadelphia, PA: Temple University Press.(A comprehensive critique of the major arguments for theism and a defence of various arguments in support of atheism.)

Rowe, W.L. (1979) 'The Problem of Evil and Some Varieties of Atheism', *American Philosophical Quarterly* 16 (3): 335-41.(Presents a version of the evidential argument from evil in support of atheism.)

Swinburne, R. (1979) *The Existence of God*, Oxford: Oxford University Press.(A careful, sympathetic presentation of most of the major arguments for the existence of God. Also works out a theodicy in answer to the problem of evil.)

Thrower, J. (1971) *A Short History of Western Atheism*, London: Pemberton Publishing Co.(A very readable summary of the development of Western atheism.)

Atomism, ancient

Ancient Greek atomism, starting with Leucippus and Democritus in the fifth century BC*, arose as a response to problems of the continuum raised by Eleatic philosophers. In time a distinction emerged, especially in Epicurean atomism (early third century* BC*), between physically indivisible particles called 'atoms' and absolutely indivisible or 'partless' magnitudes.*

The term 'atom' (*atomon*), literally 'uncuttable', was coined in the fifth century BC by the first atomists, Leucippus and Democritus (§2). As the name suggests, its primary sense is an unbreakable particle, and their theory was certainly a physical one about the ultimate constituents of phenomenal bodies. Later theorists, in the late fourth century BC and after, sometimes spoke of 'partless' or 'minimal' magnitudes or bodies, terms which focus more on the mathematical aspects of the entities in question. The Platonist Xenocrates (§2) and the Dialectician Diodorus Cronus (§2) developed such theories, although it is unclear whether, and if so how far, these were also applied to the problems of physics. In the early third century BC, Epicureanism (§§2-3) combined the physical and mathematical approaches, positing atomic physical particles which were themselves further analysable into absolute 'conceived as altogether', irreducibly small magnitudes.

Despite the above crude distinction between physical and mathematical indivisiblity, it is unlikely that the two concepts were originally distinct. The origins of atomism lay in the conceptual arguments of the Eleatic philosophers. Parmenides (§5) had argued that that-which-is is indivisible because it is 'all alike'. Some took this to mean that if a thing were divisible anywhere it would have to be divisible everywhere; hence it would consist of infinitely many parts of zero size, making the whole either infinitely large or sizeless. Zeno's puzzles about plurality and motion (Zeno of Elea §§4-6) were thought to raise similar problems. The atomist solution was that body is divisible in some places but not others: divisible in the void interstices between portions of body, but nowhere else. This is a metaphysical and mathematical thesis, but the resultant discrete portions of body easily became the basic particles of physics. Atoms, although themselves unchangeable, were held to come in varying shapes and sizes. By the motion of infinitely many of them in an infinite void, worlds and all their contents could be formed.

The emergence, especially with Diodorus and Epicurus, of a thesis of mathematical or absolute minima may reflect the feeling that atoms of varying shapes and sizes could not adequately answer the Eleatic puzzles. Such atoms must have parts, and, whether or not they could be physically fragmented into them, the same problems about measuring and counting the parts threatened to recur. Hence minima were conceived as altogether partless portions of body, and the same concept spread to the analysis of time and space. Thus all magnitudes came to be seen as granular in structure.

Opponents, whether sceptics (like Sextus Empiricus) or champions of the continuum (including Aristotle and the Stoics), could point out many conceptual difficulties, although whether these outnumbered the paradoxes of the continuum is debatable. Perhaps the toughest was: what would happen to two bodies approaching each other at equal speed across a distance consisting of an odd number of minima?

Ancient physical atomism won numerous adherents in the Renaissance and among early modern philosophers and scientists. Its most powerful exponent, Gassendi (§§2, 4), studied his Epicurean sources minutely. From him an unbroken line of influence runs to modern atomic physics.

See also: Matter, Indian conceptions of;Stoicism §4

DAVID SEDLEY

References and further reading

Aristotle (*c.* mid 4th century BC) *Generation and Corruption*, trans. C.J.F. Williams, Oxford: Oxford University Press, 1982.(Book I chapter 2 offers a particularly illuminating report and critique of Democritus' atomism; includes commentary.)

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge & Kegan Paul.(Includes an excellent chapter on Democritus' 'corpuscularian hypothesis'.)

Furley, D.J. (1967) *Two Studies in the Greek Atomists*, Princeton, NJ: Princeton University Press.(Seminal study

of atomism in Democritus and Epicurus.)
Kretzmann, N. (ed.) (1982) *Infinity and Continuity in Ancient and Medieval Thought*, Ithaca, NY: Cornell University Press.(Wide-ranging and important collection.)
Makin, S. (1989) 'The Indivisibility of the Atom', *Archiv für Geschichte der Philosophie* 71: 125-49.(Helpful analysis of the reasoning behind Democritean atomism.)
Sextus Empiricus (2nd century AD) *Against the Professors,* ed. and trans. R.G. Bury, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1936.(Greek text of Books IX-XI with English translation; Book X, which is also called *Against the Physicists* Book II, is an invaluable guide to ancient arguments on both sides of the debate about divisibility.)
Sorabji, R. (1983) *Time, Creation and the Continuum*, London: Duckworth.(Includes a magisterial study of the debate between atomicity and continuity, from the fifth century BC to the fourteenth century AD.)
White, M. J. (1992) *The Continuous and the Discrete: Ancient Physical Theories from a Contemporary Perspective,* Oxford: Oxford University Press.(Conceptually sophisticated reconstruction.)

Atonement

As a theological concept, atonement articulates the acts by which relations between God and creatures, disrupted by human offence, can be restored. Although other cultures show an awareness of the need for atonement, the Christian tradition understands it as provided by God's particular historical action in Jesus Christ. At its centre is the notion of reconciliation between God and his alienated creatures, which is achieved particularly by the death of Jesus. The distinctive philosophical and other problems of atonement theology derive from two features in particular: its claiming of universal significance for the historical life and death of Jesus of Nazareth (the problem of universality); and the moral difficulties, especially in the realm of human freedom and responsibility, which arise from the claim that he is the vehicle of atonement with God (the problem of human autonomy).

Although there were many theologies of atonement before Anselm of Canterbury's, his systematic treatment is the fountainhead of much modern discussion, both Roman Catholic and Protestant. Centring on the concept of satisfaction, it understands Christ as the God-man, satisfying both divine justice and human need by a free gift of his life. Criticisms of the formulation have centred on its understanding of sin and its tendency to understand atonement in external, transactional terms. Subsequent discussion of the concept has also raised questions about Christ's substitutionary and representative roles and about the relation between the justice and the love of God. A significant proportion of modern thinkers have rejected the need for any concept of atonement at all. They have preferred instead to understand Jesus as an example to be followed ('exemplarism') or to concentrate upon the effect his behaviour and example have on the believer ('subjectivism') - or to adopt a combination of both.

1 Atonement as moral necessity

The word 'atonement' is a Christian theological coinage, supposedly the only one made by an English theologian ('at-one-ment'). Along with many of the words found in its conceptual field - such as 'sacrifice' - its everyday uses are wide and general. Thus criminals are sometimes said to suffer punishment 'to atone' for their crime. This entry, however, will focus on the technical sense of the term, which applies primarily to relations between moral agents and the God to whom they are conceived to be responsible. In atonement theology, God is supposed to be in varying ways the active agent and passive recipient of atoning action. Most of the conceptual complications surrounding the topic derive from the various ways of construing this relation.

The concept of atonement, though developed in the Christian theological tradition in a number of related ways, is arguably of much broader provenance with respect to the moral and theological realities with which it purports to deal. Its underlying insight is that moral offences bring about an objective disruption of patterns of interpersonal life - indeed, in some accounts, of universal order also - of such a kind that some form of reparation or restoration is necessary if the resulting imbalance is to be corrected. To atone for one's offence is to act in such a way that the imbalance is corrected or reparation made to the offended. The specifically Christian understanding is that the impact of cumulative human offence is such that atonement must in some way be provided by God himself, the agents of moral offence having been rendered impotent by the weight of evil. Much ensuing discussion, particularly philosophical discussion, is centred on whether this can be conceived to take place without some violation of moral reality or human autonomy.

Classical Greek sources show that a need for atonement for wrong done is evidenced in cultures not directly affected by biblical categories. In Presocratic philosophy (Empedocles (§2), for instance) and Athenian drama there is a clear sense that moral agents are so bound up with both social and universal order that offence requires some form of atoning or purifying action. The Oedipus cycle makes clear the relation between moral impurity - particularly that deriving from the spilling of blood - and social disorder. The anthropologist René Girard (1977) has demonstrated the essentially rational way in which primitive religion understands and deals with violence as socially disruptive. Of particular importance for an understanding of substitutionary or vicarious aspects of atonement thought is his treatment of the fact that purificatory or avenging violence is often inflicted on someone or something other than the offender in order to break a cycle of revenge. Girard's is important evidence for the claim that matters of universal import underlie the particular forms of atonement theory in the Western Christian tradition.

In the Bible it is possible to discern a generally fourfold matrix within which atonement is understood. The moral

offence which must be atoned for takes place in a network of relations between God, the offender, society and the cosmic order. Despite a range of interpretations of the opening chapters of Genesis the general pattern is clear. The disruption of a due relation to God (Genesis 3) has consequences for both social order (for example, the story of Cain and Abel) and the human relation to nature (Genesis 3: 17-19). It is for the sake of a restored social order based on a renewed relation to God that the cultic and legal order of Israel, with its inextricably related atoning and juridical institutions, is described in succeeding books. It is in the light of such considerations that the New Testament and later theological discussion of the death of Jesus and its atoning significance should be understood.

2 The Bible and the Fathers

The New Testament, drawing as it does on a world of imagery deriving from the Old Testament as well as on non-biblical sources from the surrounding culture, contains the bases of the later, more systematically articulated theologies of atonement. Different writers centre their thought on the development of particular families of metaphor without drawing on one exclusively. St Paul developed metaphors derived from the law in expounding the atoning significance of Jesus of Nazareth. In the Letter to the Romans, his development of the theology of justification expounds the claim that the death and resurrection of Jesus are the way by which God is able to forgive and renew while remaining true to moral reality. It is noteworthy, however, that at a crucial stage of his argument he draws also on the imagery of the altar. The Authorized Version of the Bible controversially translates his word describing the atoning work of Jesus Christ as ‘propitiation’; later translations, fearing suggestions of substitutionary placation of an angry deity, tend to prefer ‘expiation’, implying more neutrally a means of taking away fault or pollution. All versions, however, draw on metaphors from the altar of sacrifice.

The author of the Letter to the Hebrews develops a theology of atonement drawing largely on such imagery. According to him, the eternal Son of God become man is at once human priest and victim, who perfects and replaces the Old Testament institutions of animal sacrifice. The author takes seriously the (atoning) purpose of those replaced institutions, and, while expounding the moral inadequacy of animal sacrifice (a view anticipated in the Old Testament prophets), theologizes the doctrine by centring atonement on the representative human achievement of the saviour. The effect of this metaphorical transformation of the language of the cult was recognized by later Christian writers who argued that Christ’s death as a universal sacrifice renders all animal sacrifice obsolete. The Johannine tradition likewise assimilates various aspects of the sacrificial tradition to Jesus (‘the lamb of God who takes away the sin of the world’ (John 1: 29)), but also fuses it with the military language of victory (see particularly Revelation 5).

For the purposes of this entry, the military imagery will be left to one side because the characteristic philosophical problems associated with atonement theology centre on legal and sacrificial notions, particularly those of representation, substitution and the forgiveness of sins. The notion of substitution encapsulates the claim made in some theologies that Jesus suffers *in place of* the human sinner in relation to God; that of representation the claim that Jesus acts *on behalf of* either the human race or (in some traditions) those elect for whom he is said to have died.

However these are to be understood, it can be argued that the New Testament concepts, which in different ways encourage both interpretations, come together in the expression often translated as ‘reconciliation’ (*katallage* and cognates). God in Christ is primarily the active subject of this verb (2 Corinthians 5: 18-20); believers are claimed (factually) to have been reconciled to God (Romans 5: 10-), but are also urged to ‘be reconciled to God’ (2 Corinthians 5: 20). The chief reference appears to be to a past, completed divine act which both determines the status of believers (and, possibly, everyone) and requires an appropriate religious and ethical response; in the Letter to the Ephesians, this is the reconciliation in one church of the once estranged Jews and Gentiles.

While the notion of substitution has little prominence in early theology, the notion of exchange (suggested by the primary meaning of *katallage*, which in turn is deeply indebted to the Old Testament - see Isaiah 52-3) prepares the way for later discussion. The anonymous (probably second-century) *Letter to Diognetus* celebrates the matter with no suggestion of the theological and moral difficulties that are later to arise: ‘O sweetest exchange! O unfathomable work of God!…. The sinfulness of many is hidden in the Righteous One, while the righteousness of the One justifies the many that are sinners’ (9.5). The foundations of the classical doctrine of the atonement were later laid by the patristic writers, for example Athanasius, who in *On the Incarnation* (*c*.316-18) employed forensic, sacrificial and military imagery in his account of the saving significance of Jesus. But, although this is

manifestly a theology of atonement, there was little systematic articulation of the problems in the early period.

Three generalizations are useful for an understanding of the later debates. First, the early writers, and the Eastern Orthodox tradition until today, tend to speak of redemption by Christ's self-sacrifice rather than atonement, and exposition is rather in terms of a wealth of images than of a systematic working out of a theology. Second, Orthodox theology tends to speak more of an ontological change brought about in the believer as a result of the self-giving sacrifice of Christ and the life of the church than of a moral or legal reconciliation concentrating on Christ's death. Third, theology in the West until the Middle Ages was very much shaped by Gregory of Nyssa's celebrated image of redemption: on the cross, Jesus cheats the devil of his human prey, deceiving him into thinking that here is but a man. The devil swallows his victim, only to be caught on the hook of his hidden divinity. Gregory's way of speaking was the catalyst for the critique of Christian theology which drew a definitive response from Anselm of Canterbury.

3 The doctrine of satisfaction

The doctrine of satisfaction has its origins in the rather juridical conception of the human relation to God developed in the West. In the first instance, it was used of the human response to God, with few overtones of atonement theology. For Tertullian, the human calling was conceived in terms of satisfying the just demands of God, the sovereign of the moral order. The concept was adopted and developed as a metaphor to express the atoning act of the Son of God by Anselm of Canterbury (§8) in his *Cur Deus homo (Why God Became a Man)* (completed in 1098). In response to criticisms of the irrationality of Christian atonement teaching made by opponents (probably Jewish and Muslim thinkers), he rejected the then reigning theology of atonement, which was heavily influenced by Gregory of Nyssa and according to which the death of Jesus on the cross had freed humankind from the legal power of the devil. Anselm's rejection was based on a conception of the sovereignty of God whereby God has no need to bargain with one who is his own creature. His own theory attempted to establish the necessity of the atonement understood in terms of satisfaction.

The assumption underlying Anselm's argument - and it is probably a necessary assumption for Christian theologies of atonement - is a belief in universal moral order and God's responsibility for upholding it. The human breach of this moral order has led, on the one hand, to human moral incapacity to atone, because of the infinite weight of accumulated offence; on the other, it has led to a situation in which God must either punish or provide some alternative (such as satisfaction) if his purposes in creation are not to be frustrated. Anything else (for example, the mere remission of sins) would involve an offence against universal order, and so be unjust, even (or especially) for God. Punishment would consist in annihilating the human race and so would involve an abandonment of God's purposes in creation; satisfaction requires a counterbalancing act of restitution which maintains that order.

Satisfaction is provided by the gift to God the Father of the God-man's life, offered by means of his free human obedience; as the life of God it outweighs even the infinity of human offence. This notion of satisfaction provides the basis of all claims for the universal significance of the particular historical act of Jesus. Because it is the act of God in the context of universal human fallenness - at one level, a gift by God to God; at another, an act of human reparation - it is of universal significance. Recent commentators (such as Steindl), in line with a long tradition of biblical theology of salvation, have emphasized that satisfaction, as an alternative to punishment, represents not the infliction of a penalty on a human Jesus in place of the actual offenders, but a creative act whereby God, by a new initiative, overcomes evil by good.

Anselm's aim was the limited one of establishing the rationality and necessity of the God-man, so there is a measure of injustice in insisting that he should have done things that he never set out to do. But subsequent theology has criticized him for omissions as well as commissions in his treatment of the atonement. Objectors have identified the excessively legal and mathematical terms of the conception, with its transactional rather than relational or personal emphasis, and the failure of the theology to give some account of the appropriation of salvation by the believer. Both criticisms observe a tendency to see the atonement as external to the human recipient of salvation and so inadequate as an account of reconciliation, which concerns relationships.

4 After Anselm

Important variations on Anselm’s theme were produced by Thomas Aquinas and John Calvin, both of whom strengthened it where it appeared weak, but remained broadly within his tradition. Aquinas broadened the terms in which the atonement was to be understood, balancing and expanding talk of a debt paid with that of a stain erased. Some of the external features are mitigated when the legal metaphor of satisfaction is enriched with sacrificial imagery which suggests cleansing rather than remission of penalty.

Calvin’s treatment of the atonement is the fountainhead of many recent disputes. The context is Luther’s reassertion and reshaping of the Pauline and Augustinian doctrine of justification by faith in terms of personal and relational categories (see Justification, religious §4; Luther, M.). Justification for Calvin is a metaphor of legality and therefore still within the Anselmian penumbra, but because his account centres on the subjective appropriation of the act of atonement by God in Christ, it helps to obviate the externalism of the Anselmian view. He buttressed his teaching of justification ‘by faith through grace’ with a theology of the objective atoning act of Christ. This act derives from the love of God (a note strangely absent from Anselm, who prefers to speak of God’s mercy). The wrath from which it saves us is, according to Calvin, the necessary means by which we are made aware of our need for salvation. Christ saves us by both a propitiatory sacrifice, which Calvin understands similarly to Anselm, and by a representative obedience. Another central aspect of Calvin’s teaching is the *participation* of the believer in the achievement of Christ, another blow against externalism (see Calvin, J. §4).

Differences over justification led to a parting of the ways between the two main streams of Western Christianity. Over against the Reformers, with their teaching of the complete forgiveness of sin, the sixteenth-century Council of Trent, building on Anselm and Aquinas in a very different way, stressed the merit earned by Christ on the cross and communicable to believers especially through the mass. The Roman Catholic church taught until recent times that sinners must continue to atone for their sins in this world and after death in purgatory (see Purgatory). In the teaching of the Second Vatican Council, juridical conceptions have given way to the view that by the death and resurrection of Jesus, God overcame evil and made humanity a new creation. However, because the chief intellectual battles since the Reformation have been between the orthodox and liberal wings of Protestantism, we shall concentrate on these developments.

Although Calvin did not say, any more than Anselm, that Jesus is punished instead of us, some of his successors were less careful in their expression of the doctrine. The neglect of his participatory language had two important consequences. First, the doctrine was again taken as teaching some external ‘transaction’, sometimes construed as a ‘penal substitution’ in which Christ was seen as suffering an equivalent penalty to that to which the guilty were liable. Second, and more important for the later history of the doctrine, it became liable to criticisms of varying degrees of severity.

On the one hand are the more radical critiques based on doctrines of moral autonomy. The source of this approach is the medieval philosopher Peter Abelard. While there is disagreement about the interpretation of his theology, it appears to make its chief appeal not to the reconciliation of those who are in some way at odds, but to a definitive act in which God shows his love to sinners. Those successors of Abelard who move in the direction of theories that are called subjective (because they deny an objective reconciling act) or exemplarist (because they see the work of Christ as essentially an example to be followed) can scarcely be said to produce theologies of atonement at all, because they have no place for a historic act of reconciliation of opposed or estranged parties. Their arguments are of interest here for their objections to forms of atonement theology proper.

The Socinians, the first influential opponents of a theology of atonement, taught the non-transferability of penalty and a strongly individualist conception of moral responsibility, and thus provide the strongest possible challenge to traditional conceptions (see Socinianism §§2-3). Their approach has its paradigmatic modern treatment in Kant’s *Religion within the Limits of Reason Alone* (1793) (see Kant, I. §11). Kant accepts two features of the orthodox teaching: a doctrine of the fallenness of the human race (‘radical evil’), and the requirement for some form of atonement. But the atonement cannot be anything other than a kind of individual philosophical and inner conversion. In describing this conversion, Kant uses all kinds of traditional atonement language, but transfers its reference from the historical cross to the individual’s soul.

On the other hand, the doctrine of atonement has been revised from within by a number of theologians who wished to remain true to its insight that Jesus Christ, and particularly his cross, is definitive for human salvation. Their criticism centred on the doctrine of a limited atonement taught by some Calvinists, that Christ died only for the

elect. Many have held this limitation on the universality of atonement to be inconsistent with the justice of God. Influential in the development of more universal theologies was John McLeod Campbell. Early in the twentieth century, the Scottish Congregationalist P.T. Forsyth, while accepting aspects of liberal criticism, yet felt it necessary to stress the holiness of God, in the light of which conception he argued for the need for the historic act of God in the substitutionary cross of Jesus. Similarly, Karl Barth (§§2-3) attempted to combine the elements of truth in substitutionary and representative theologies, while seeing the atonement as the historical outworking of God's universal electing and covenant love.

5 Theological and philosophical theses

Substitution and representation. There is a strong case for holding that Christian atonement theology requires some element of substitution if it is to give an adequate account of the teaching that human fallenness required the death of the saviour who on the cross took our place before God. Biblical teaching is, however, unanimous that there is no breach between the actions of God the Father and God the Son. The Son dies as a result of the love of God the Father, even though he also undergoes the effects of divine judgment on human sin. Here a distinction should be drawn between substitutionary atonement and substitutionary punishment. One cannot be punished for the offences of others, but one can make substitutionary atonement for them by taking their place in a transformative act of love. On this account, the Son of God, as representative of the human race, freely enters that relation to God which is the universal fate of those who reject God's love. His death is not so much a punishment as a sharing of the *relationlessness* which is the just fate of those who would deny their nature as existing in relation to God and others.

Justice and love. The classical Pauline view of the cross is that it expresses at once the love and justice of God. The cross and resurrection of Jesus show that he is at once just and justifies. That is to say, the reconstitution of human life by means of atonement establishes rather than undermines morality. Christian theology's claim is that God himself has provided the solution to the problem of moral evil by himself bearing its consequences, as a way of overcoming evil with good. This has encountered opposite objections from left and right: that God should be able to forgive simply by an exercise of power, and that morality is undermined by forgiveness. But the first objection fails to uphold the reality of the moral law, while the second ignores the transformative effects of atonement, the overcoming of evil by good.

Moral responsibility. A conception of atonement of this kind also requires a relational view of humanity. While I am responsible for my sins, what I am is also inextricable from my relation to God and to others. Modern objections, particularly those expressed in Kantian theories of autonomy, mistake the human condition in two important respects. They are individualist, and they assume a false or non-theological conception of human freedom. The participatory dimensions of atonement theology show that by participating in the (representative) achievement of Christ - particularly in those sacraments constitutive of the new community of forgiveness (baptism and the Lord's Supper) - the believer shares a form of community within which reconciled forms of life are realized.

The particular and the universal. As we have seen, Anselm's concept of universality derived from a particular conception of the divine act in Christ. Any attempt to show that a particular historical event or person is of universal moment depends on two features: its determination in divine intention to be what it is, and its answering to the needs of the human condition, particularly fallenness. If the human race does not have within itself the capacity for moral regeneration, there is a case for arguing that it is to be found in the reconstitutive acts of the life, death and resurrection of Jesus.

See also: Grace; Sin

COLIN GUNTON

References and further reading

Anon. (2nd or early 3rd century) *Letter to Diognetus*, in C.C. Richardson (ed.) *The Early Christian Fathers*, The Library of Christian Classics 1, London: SCM Press, 1953, 213-24.(An early summary of the Christian faith containing a notion of exchange.)

Anselm of Canterbury (completed 1098) *Cur Deus homo (Why God Became a Man)*, in E.R. Fairweather (ed.) *A Scholastic Miscellany: Anselm to Ockham*, The Library of Christian Classics 10, London: SCM Press, 1956,

100-83.(The medieval source of all modern discussion of atonement theology.)

Aquinas, T. (1266-73) *Summa theologiae* IIIa, qq.46-9, trans. with intro. by R.T.A. Murphy, London: Eyre & Spottiswoode, 1965, 2-115.(Important revisions of Anselm’s theology of atonement.)

Athanasius (c.316-18) *On the Incarnation*, in E.R. Hardy (ed.) *Christology of the Later Fathers*, The Library of Christian Classics 3, London: SCM Press, 1954, 55-110.(Classic early statement of a theology of redemption.)

Barth, K. (1956) *Church Dogmatics*, vol. 4/1: *The Doctrine of Reconciliation*, trans. G.W. Bromiley, Edinburgh: T. & T. Clark.(A major twentieth-century treatment of the doctrine.)

Calvin, J. (1559) *Institutes of the Christian Religion*, trans. J.T. McNeil and F.L. Battles, The Library of Christian Classics 20, London: SCM Press, 1960.(With Anselm, the fountainhead of modern debate about atonement.)

Campbell, J.M. (1878) *The Nature of the Atonement*, London: Macmillan.(A modern attempt to stress representative rather than substitutionary aspects of atonement.)

Cassirer, H. (1988) *Grace and Law: St Paul, Kant and the Hebrew Prophets*, Edinburgh: Handsel Press.(Argues that Paul’s account of the human condition is philosophically superior to Kant’s.)

Douglas, M. (1984) *Purity and Danger: An Analysis of the Concepts of Pollution and Taboo*, London: Ark Books.(A pointer to universal concerns underlying atonement theology.)

Feenstra, R. and Plantinga, C. (eds) (1989) *Trinity, Incarnation, and Atonement*, Notre Dame, IN: University of Notre Dame Press.(Contains essays discussing philosophical problems of atonement theology.)

Forsyth, P.T. (1965) *The Work of Christ*, London: Collins.(An important modern statement of a substitutionary theology of atonement.)

Girard, R. (1977) *Violence and the Sacred*, trans. P. Gregory, Baltimore, MD: Johns Hopkins University Press.(An anthropological study with discussion of the meaning of substitutionary sacrifice.)

Gregory of Nyssa (*c*.383) *An Address on Religious Instruction*, in E.R. Hardy (ed.) *Christology of the Later Fathers*, The Library of Christian Classics 3, London: SCM Press, 1954, 268-325.(An influential early theology.)

Gunton, C. (1988) *The Actuality of Atonement: A Study of Metaphor, Rationality and the Christian Tradition*, Edinburgh: T. & T. Clark.(An attempt to state a theology of atonement with the aid of recent philosophy of language.)

Kant, I. (1793) *Religion within the Limits of Reason Alone*, trans. T.M. Greene and H.H. Hudson, New York: Harper & Row, 1960.(The most important restatement of atonement as a philosophy of individual self-realization.)

McIntyre, J. (1954) *St Anselm and his Critics: A Reinterpretation of the Cur Deus Homo*, Edinburgh: Oliver & Boyd.(An important study of Anselm’s philosophy and theology.)

Steindl, H. (1989) *Genugtuung: Biblisches Versöhnungsdenken - eine Quelle für Anselms Satisfaktionstheorie?* (Satisfaction: Biblical Thought about Reconciliation - a Source for Anselm’s Theory of Satisfaction?), Freiburg: Universitätsverlag.(An important recent study of Anselm’s background and theology.)

Stump, E. (1988) ‘Atonement According to Aquinas’, in T.V. Morris (ed.) *Philosophy and Christian Faith*, Notre Dame, IN: University of Notre Dame Press, 61-91.(A discussion of Aquinas in the context of modern problems.)

Swinburne, R. (1989) *Responsibility and Atonement*, Oxford: Clarendon Press.(A modern philosopher’s Anselmian reflections on the Christian understanding of atonement.)

Taylor, V. (1941) *Forgiveness and Reconciliation: A Study in New Testament Theology*, London: Macmillan.(A summary of the main biblical teaching.)

Weingart, R.E. (1970) *The Logic of Divine Love: A Critical Analysis of the Soteriology of Peter Abailard*, Oxford: Clarendon Press.(A study of the source of modern exemplarist theories.)

White, V. (1991) *Atonement and Incarnation: An Essay in Universalism and Particularity*, Cambridge: Cambridge University Press.(A philosophically sophisticated discussion of claims of universality.)

Augustine (AD 354-430)

Augustine was the first of the great Christian philosophers. For well over eight centuries following his death, in fact until the ascendancy of Thomas Aquinas at the end of the thirteenth century, he was also the single most influential Christian philosopher. As a theologian and Church Father, Augustine was the person who did the most to define Christian heresy and so, by implication, to formulate Christian orthodoxy. Of the three most prominent heresies defined by Augustine - Donatism, Pelagianism and Manicheism - the latter two also have especially important philosophical implications. In rejecting Pelagianism and its thesis of human perfectibility, Augustine rejected one form of the principle, often associated with Kant, that 'ought' implies 'can', and in rejecting Manicheism, with its doctrine that good and evil are equally basic metaphysical realities, Augustine rejected one solution to the philosophical problem of evil.

The Categories may have been the only work of Aristotle that Augustine actually read. Plato he knew somewhat better. He seems to have been familiar with several Platonic dialogues and he clearly felt a special affinity for Plato and the Platonists, which is particularly evident in De civitate Dei (The City of God) and De vera religione (On True Religion). Although he could be said to have responded to classical Greek philosophy in consequential ways, it must be added that what he responded to had been filtered through Neoplatonism, Hellenistic scepticism and Stoicism. It was principally through the writings of Cicero that Augustine became schooled in the opinions of his philosophical predecessors, and it was through the works of the Neoplatonists that he developed his deep appreciation for Plato.

Augustine's philosophy thus draws significantly on the philosophy of late antiquity as well as on Christian revelation. Its originality lies partly in its synthesis of Greek and Christian thought, and partly in its development of a novel ego-centred approach to philosophy that anticipates modern thought, especially as exemplified in the philosophy of Descartes. In his De trinitate (The Trinity) and De civitate Dei, Augustine presents a line of thinking that foreshadows Descartes' famous cogito, ergo sum. *Through his Confessionum libri tredecim (*Confessions, *more usually known as Confessiones), the first significant autobiography in Western literature, and also through his Soliloquia (Soliloquies), which is a dialogue between himself and Reason, Augustine introduced a first-person perspective to Western philosophy.*

Early in his career, Augustine found himself attracted to philosophical scepticism. In his earliest extant work he offers his most extensive response to the main sceptical arguments of his day, including those that raise the possibility one might only be dreaming. His later responses to scepticism, though less extensive, are better focused; they concentrate on the self-knowledge he considers directly available to each knowing subject, including the knowledge that one exists. Taking the first-person perspective one can also develop, he tries to show, in his De trinitate, a convincing argument for mind-body dualism. But supposing, as he does, that each of us knows from our own case what a mind is raises, as Augustine is perhaps the first philosopher to realize, a problem about how one can ever know that there are minds in addition to one's own.

Augustine's account of language and meaning influenced the development of 'terminist' logic in the high middle ages. His thoughts on language acquisition in Confessiones provide a foil for Wittgenstein in the latter's Philosophical Investigations. Yet, some of Augustine's own reflections on ostensive definition in his dialogue De magistro (The Teacher) anticipate Wittgenstein's own views on language learning.

Augustine develops what is described as an 'active' theory of sense perception, according to which rays of vision touch objects whose consequent action on the body is 'noticed' by the mind or soul. Although his ideas on sense perception are interesting, his most influential epistemological conception is certainly his 'theory of illumination'. Instead of supposing that what we know can be abstracted from sensible particulars that instantiate such knowledge, he insists that our mind is so constituted as to see 'intelligible realities' directly by an inner illumination.

The modern concept of the will is often said to originate with Augustine. Certainly the idea of will is central to his philosophy of mind, as well as to his account of sin and the origin of evil. Strikingly, he uses psychological 'trinities', including the trinity of memory, understanding and will, to illuminate the doctrine of the Divine Trinity, where there is also a baffling unity in plurality. The theological warrant for this analogy Augustine finds in the

biblical idea that God created human beings, and specifically the human mind, in his own image.

Augustine's attempts to achieve a philosophical understanding of theology and religious belief set the framework for much later medieval and early modern philosophy. On the issue of how reason should bear on religious faith, Augustine develops the idea that reason should work out an understanding of what we must first accept on faith. Yet he also displays a keen sensitivity to those issues most likely to challenge one's religious faith. Prominent among his concerns is the philosophical problem of evil, to which he offers what has proved to be perhaps the most influential type of solution.

Particularly striking is Augustine's virtually lifelong preoccupation with human freedom and how the fact that human beings are free to make their own choices can be reconciled with the Christian doctrines of God's foreknowledge, predestination and grace. Almost every important medieval philosopher in the Christian West would later contribute to the continuing effort to achieve a satisfactory reconciliation of these issues. It is significant that Leibniz, who gave the problem of freedom, foreknowledge, predestination and grace one of its most sophisticated treatments, also gave much of his philosophical attention to the equally Augustinian problem of evil.

Although Augustine did present an argument for the existence of God, it is his understanding of the divine attributes, and especially his insistence on divine 'simplicity', that is, on the idea that God is not distinct from his attributes, that has been especially influential on later thinkers. Also influential are his various attempts to understand the created world. Augustine made several important efforts, perhaps most notably in the last books of his Confessiones and in his De genesi ad litteram (The Literal Meaning of Genesis) to give a philosophically sophisticated account of the creation story in the biblical book of Genesis. His contrast between God's eternity and human temporality set the stage for later medieval and modern discussions of these issues, and his discussion of the nature of time in Book XI of his Confessiones is sometimes taken to epitomize philosophy.

Augustine's descriptions of mystical experience are among the most eloquent in Western literature; they belong among the classic texts of mysticism. However, Augustine's attempts to understand ritual are perhaps more remarkable for the directness with which he identifies and confronts difficult issues than for the success of his efforts to solve them. Those efforts seem to be hobbled by his version of mind-body dualism.

Augustine is a thoroughgoing intentionalist in ethics. This feature of his thought, as well as his unflinching insistence that one can do what one knows one ought not to be doing, mark him off from ethicists of the classical Greek period. Yet Augustine also preserves in his own thinking important strands of ancient Greek thought. Thus, for example, his development of the doctrine of the Christian virtues includes an echo of Plato's idea of the unity of the virtues. His insistence that 'ought' does not, in any straightforward way, imply 'can', distinguishes him, not only from his contemporary Pelagius, whom he helped brand as a Christian heretic, but also from most modern ethicists as well.

The philosophy of history Augustine develops in De civitate Dei initiates a branch of philosophy that came into full flower in the nineteenth century. Also in that same work Augustine makes an influential contribution to what has come to be called 'just war theory', an applied ethical theory that has continued to develop even into the latter half of the twentieth century.

1 Life

Augustine was born Aurelius Augustinus, in the North African town of Tagaste (modern Souk Ahras in eastern Algeria), in the Roman Province of Numidia in the waning years of the Roman Empire. Except for a five-year stay in Italy, he spent his entire life in North Africa. Ordained a priest in AD 391 and made a bishop four years later, he lived out the remaining thirty-five years of his life, first as a coadjutor and then as the diocesan Bishop of Hippo (later Bône, now Annaba, Algeria), which was at the time the second most important port city in Africa.

Augustine's mother, Monica, was a devout Christian; his father, Patricius, a man of modest means, was not given a Christian baptism until he was on his death bed. Augustine received a classical education, first in the local grammar school, then in a higher school in nearby Madaura and finally, under the patronage of a local nobleman named Romanianus, at the university in Carthage. It was in Carthage as a student of rhetoric that he read Cicero's now lost dialogue, *Hortensius*, which, as he later wrote in *Confessionum libri tredecim* (Confessions, more usually

known as *Confessiones*) altered his sensibility and brought him under the spell of philosophy.

After a brief period as a teacher of rhetoric in his home town, Augustine returned to Carthage and then, in AD 383, sailed for Rome. The five years he spent in Italy included a period as professor of rhetoric at Milan. Also in Milan, Augustine joined a circle of Neoplatonists and turned away from the Manicheism he had embraced in Carthage. In Book VII of *Confessiones* he explains how profoundly the Neoplatonic works he read in that Milanese circle helped him think about the nature of God and the problem of evil.

It was also in Milan, after he had immersed himself in Neoplatonism, that Augustine finally became a Christian convert, under the tutelage of Ambrose, Bishop of Milan, and through the continuing influence of his mother, who had followed him to Milan. The year before his baptism in AD 387 Augustine withdrew with several philosophically-minded relatives and associates to a villa at Cassiciacum, perhaps near Lake Como, where he wrote four of his earliest works, including an extremely interesting dialogue critical of the scepticism of the New Academy, *Contra academicos (Against the Academicians)*, as well as the *Soliloquia (Soliloquies)*.

Shortly after Augustine's baptism, his mother died. Following a brief stay in Rome, Augustine returned to Carthage in AD 388 and never left North Africa again. When he became first a priest and then a bishop, he sought to combine his pastoral duties with extensive excursions into philosophy and theology. It was in response to the spread of Donatism, Pelagianism and Manicheism in North Africa that Augustine wrote great treatises to expose those trends as heretical (see Manicheism; Pelagianism). It was in further response to the decline and fall of the Roman Empire, blamed by critics of Christianity on the Christianization of Rome, that Augustine wrote *De civitate Dei*. Rome had already been sacked in AD 410; Hippo itself was under siege as Augustine died twenty years later.

Augustine's literary output, produced with the help of scribes, is enormous. Chadwick (1986) claims that what survives of his work is the largest body of writing left by any ancient author. In addition to approximately one hundred books and treatises there are some two hundred letters and over five hundred sermons. Three years before his death, Augustine went through all his works and listed and commented on them in a great compendium, the *Retractationes*, or reviews. In some cases he retracts claims made in the works he reviews, or expresses regret at having made them. However, most of what he had written he lets stand. On the basis of these reviews, we may conclude that at least 90 per cent of Augustine's writings have survived.

After Augustine became a priest in AD 391, he wrote no single work that could be said to be entirely philosophical. On the other hand, hardly anything he wrote at any time in his life was entirely devoid of philosophy. Philosophical reflections, analyses and explorations turn up, often quite unexpectedly, in his sermons and letters and, of course, in the great theological, exegetical and doctrinal treatises. By the time Augustine came to write *De civitate Dei*, his lingering admiration for the grand ambitions of speculative philosophy and natural theology had become tempered with an acute awareness of how the human mind on its own is too crippled by old vices to be able to 'enjoy and abide in the changeless light' (*De civitate Dei* XI.2). Be that as it may, the philosophical light in his own eye burned brightly until the very end of his long and singularly productive life.

2 Scepticism

In Rome at the beginning of his stay in Italy Augustine grew increasingly dissatisfied with Manicheism, to which he had provisionally given his allegiance in Carthage. He found himself attracted to the sceptical viewpoint of the Academics, the followers of Arcesilaus and the New Academy, who 'held that everything was a matter of doubt and asserted that we can know nothing for certain' (*Confessiones* V.10.19). Augustine seems to have learned of ancient scepticism, and of the debate between Arcesilaus and the stoic Zeno of Citium from Cicero's *Academica* (see Cicero).

Augustine's most extensive discussion of scepticism is to be found in his earliest surviving work, the dialogue *Contra academicos*, written at Cassiciacum just months before his baptism. The Academics, according to him, base their claim that nothing can be known on the application of a strict criterion for knowledge put forward by Zeno. Although Augustine formulates Zeno's criterion in several ways and it is difficult to be certain exactly how he wants it to be understood, the point seems to be that, according to this criterion, something can be known just in case it cannot even seem to be false. Indubitability would then be both necessary and sufficient for knowledge. Against accepting such a criterion, Augustine proposes a dilemma: either the criterion is known to be true, or it is not. If it is known to be true, then the sceptics are wrong, since something is known. If it is not known to be true,

then the sceptics have given us no adequate reason to become sceptics.

Augustine is not satisfied, however, with merely demonstrating the self-defeating character of the Academic position; he goes on to offer sample knowledge claims that he dares the sceptic to challenge. These sample claims fall naturally into three groups: *logical truths* (for example, 'There is one world or there is not'), *mathematical truths* (for example, 'Three times three is nine'), and *reports of immediate experiences* (for example, 'That tastes pleasant to me'). Of special interest is Augustine's response to the sceptical challenge, 'How do you know that this world exists if the senses are deceptive?' (*Contra academicos* III.11.24). Augustine answers: 'I call this entire thing, whatever it is, which surrounds and nourishes us, this object, I say, which appears before my eyes and which I perceive is made up of earth and sky, or what appears to be earth and sky, the world.' His idea is that, even if he is asleep and dreaming, he can know that 'the world' exists in the stipulated sense that there is for him at least a phenomenal world.

Although Augustine never goes on, in the fashion of Descartes (§9) to offer an argument for the existence of a world that is independent of him and his sense-impressions - what came later to be called 'the external world' - he does at least entertain the thought that there might not be such a world. As we know from later philosophy, that supposition is of great philosophical moment. Furthermore, his idea of an apparent world that fills our phenomenal space, whether or not there is an independent external world that resembles it, suggests a very Cartesian concept of mind.

Augustine never again devoted a whole treatise to the refutation of scepticism, but he did respond, again and again, to the challenge of Academic scepticism. Among the most interesting of these anti-sceptical passages are two in which he presents reasoning remarkably like Descartes' *cogito, ergo sum*:

> In respect of those truths I have no fear of the arguments of the Academics. They say, 'What if you are mistaken?' If I am mistaken, I am [*Si fallor, sum*]. Whoever does not exist cannot be mistaken; therefore I exist, if I am mistaken. Because, then, I exist if I am mistaken, how am I mistaken in thinking that I exist, when it is certain to me that I am if I am mistaken?
>
> (*De civitate Dei* XI.26)

Similar reasoning is to be found in *De trinitate (On The Trinity)*, except that there Augustine defends the 'inner' knowledge by which we each know, in our own case, that we are alive. (Presumably 'alive' must be understood here not in a specifically biological sense, but rather in the sense in which we may ask if there is life after death.)

Although Augustine's way of certifying his claim to know that he exists is indeed very similar to the reasoning of Descartes, the similarities should not obscure the equally important differences. Perhaps the crucial difference is that, unlike Descartes, Augustine does not use 'I exist' as a foundation stone in a rational reconstruction of knowledge. Indeed, although there are passages in Augustine in works such as *De libero arbitrio (On Free Choice of the Will)* that suggest a foundationalist approach to knowledge, he never really offers anything that could be called a systematic reconstruction of what one can be said to know. His defence of 'I know I exist', in particular, is meant simply to defeat universal scepticism, of the sort that occupied his attention in *Contra academicos*. His idea is that, if universal scepticism can be defeated by an unassailable knowledge claim, then the Academics are not after all 'wiser than the rest' and not everything is 'a matter of doubt', as he had earlier been tempted to think.

An important weapon in the attacks of the Academic sceptics on putative knowledge claims was always the question, 'What if you are dreaming?' In a passage from Book XV of *De trinitate* that parallels the *si fallor, sum* passage from *De civitate Dei* just discussed, Augustine considers the Academic's taunt: 'Perhaps you are sleeping, and you do not know, and you see in your dreams.' Augustine replies: 'He who is certain about the knowledge of his own life does not say in it: "I know that I am awake," but "I know that I live"…. He cannot be deceived in his knowledge of this even by dreams, because both to sleep and to see in dreams belong to one who lives' (*De trinitate* XV.12.21). In *Contra academicos*, Augustine rejects in a similar way the idea that not knowing whether one is dreaming might be a threat to one's claim to know logical and mathematical truths, or truths about one's immediate experiences.

However, Augustine never questions the assumption that we are sometimes awake and having veridical sense experiences, as well as sometimes dreaming. (He asserts that he sometimes knows, while he is dreaming, that he is dreaming, but he does not explain how he can know this.) Thus Augustine seems never to have considered the

ultimate Cartesian challenge that perhaps no thought that has ever entered one's mind is any more nearly true than the illusions of one's dream: in effect, all life might be simply one's dream.

3 Philosophy of language

In Book I of *Confessiones*, Augustine presents a classic view of language acquisition through ostension: 'When [my elders] named some object, and accordingly moved towards something, I saw this and I grasped that the thing was called by the sound they uttered when they meant to point it out' (*Confessiones* I.8.13). This passage, made famous by Wittgenstein in his *Philosophical Investigations*, may seem to most readers of the work to be little more than an aside. Yet, reflections on language appear throughout Augustine's writings, and language acquisition is a special concern of a relatively early work, the dialogue *De magistro (The Teacher)*, a dialogue between Augustine and his young son Adeodatus (who died at age 16, shortly after the dialogue was written).

In *De magistro*, Augustine clearly poses the problem of the ambiguity of ostension, and indeed does so in a most memorable way. First, the issue is raised as to how one can point to the colour of a body as distinct from the body itself. Then, there are more general worries:

> Augustine: Come now, tell me; if I, knowing absolutely nothing of the meaning of the word, should ask you while you are walking what walking is, how would you teach me?
> Adeodatus: I should walk somewhat more quickly…
> Augustine: Don't you know that walking is one thing and hurrying is something else?

(*De magistro* 3.6)

This last worry is surprisingly Wittgensteinian; as is Augustine's response. Although Augustine admits there is no way to eliminate all conceivable ambiguity in ostension, an intelligent learner, he thinks, will eventually catch on.

In *De magistro*, Augustine seems to want to account for meaning solely by appeal to what terms refer to. However, he does point out that for some pairs of words, each signifies 'as much' (*tantundem*) as the other; that is, the terms are co-referential, even though they do not signify 'the same' (*idem*) in that they do not have the same meaning or sense. Thus 'word' (*verbum*) and 'name' (*nomen*) are said to be co-referential, though not identical in meaning. Here Augustine relies on blurring the distinction, made in later medieval philosophy, between the personal and the material supposition of a term, or, we might say today, between *using* a term and *mentioning* it by quoting it (see Language, medieval theories of).

The discussion in later writings such as Book XV of *De trinitate* of the 'inner word' that is in no natural language seems to be, in part at least, an attempt to allow that a sense or concept mediates between a general term in a natural language and its extension:

> For the thought formed from that thing which we know is the word which we speak in our heart, and it is neither Greek, nor Latin, nor of any other [natural] language; but when we have to bring it to the knowledge of those to whom we are speaking, then some sign is adopted by which it may be made known. And generally this is a sound, but at times also a nod; the former is exhibited to the ears, the latter to the eyes, in order that that word which we bear in our mind may also become known by bodily signs to the senses of the body.

(*De trinitate* XV.10.19)

Although the idea that thinking might be inner talking is as old as Plato's *Theaetetus*, it is to Book XV of *De trinitate* that William of Ockham, for example, makes reference in the *Summa logicae* (I.1) when he wants backing for the idea of there being a mental language, as well as spoken and written languages.

Augustine begins *De magistro* with an inquiry into the purpose of using language. His son's first suggestion, that we speak either to teach or to learn, encourages us to accept the common assumption that we use language, either mainly or solely, to pass on information. Although Augustine and his son quickly come to include singing and praying among the language activities they discuss, the dialogue never gives proper consideration in the manner of, for example, Wittgenstein or J.L. Austin, to the full range of activities that make use of language or to the variety of roles individual utterances may play in our lives. On the contrary, Augustine concludes his dialogue

with the suggestion, backed by an intriguing but somewhat elusive line of reasoning, that language only reminds us of what we already know.

Other important discussions of language are to be found in Augustine's *De doctrina Christiana (On Christian Doctrine)*, especially at the beginning of Book II, as well as in the early treatise *De dialectica (On Dialectic)* (the authenticity of which, though contested, seems now to be quite generally accepted).

4 Epistemology

Augustine is said to have an 'active' theory of sense perception. The term 'active' in this context includes the idea that in vision the eyes emit rays that touch the object of vision. This is an idea that can be found in Plato's *Timaeus*, for example, as well as in other ancient sources (see Perception; Plato). More generally, it is Augustine's contention that, although bodily sense organs undergo change during perception, perception is not something undergone by the soul. 'Perception', he writes in a famous passage, 'is something undergone by the body *per se* that is not hidden from the soul [*non latens animam*]' (*De quantitate animae* 48). The soul takes note of what the body undergoes: 'For it is not the body that perceives, but the soul through the body, which messenger, as it were, the soul uses to form in itself the very thing which is announced from the outside' (*De genesi ad litteram* XII.24.51).

In Book XII of *De genesi ad litteram*, Augustine departs altogether from commenting on the biblical book of Genesis and offers instead a somewhat independent treatise on three kinds of vision: bodily vision, spiritual vision and intellectual vision. What Augustine calls *bodily vision* is in fact sense perception; *spiritual vision* is the entertaining of mental imagery, whether in memory or imagination; and *intellectual vision* is the non-imaginal perception of universal objects, structures and truths. This work includes Augustine's most serious attempt to account for error in sense perception. It also includes one of his most beautiful descriptions of mystical vision, and in fact this work took on great significance for the discussion of mysticism in the later Middle Ages (see Mysticism, history of).

Augustine's theory of sense perception seems not to be representational in the sense of making an image or sense-datum the direct object of perception. Thus in his discussion of vision in *De trinitate*, he claims that in seeing a body, we immediately form an image of it in our sense; yet we cannot discriminate between the form of the body seen and the form of the image we produce in our sense. He offers analogies and arguments to convince his readers that sight does have an image of the body seen, as soon and as long as it is seen, an image we may retain in memory after the perception. However, he seems to consider the formation of this inner image to be part of the process of seeing the *body*, not the production of something that could properly be said to be the direct object of perception.

Augustine's account of knowing, for example, what is virtue or what is a square, is not based on the idea of abstraction, as the accounts of Aquinas and sometimes even Aristotle are said to be. Rather, he espouses what is called 'the theory of illumination'. Augustine's talk of illumination is, in part, simply the deployment of an apt and traditional metaphor, that of light. He often uses this metaphor in discussions of cognition, as in the *Soliloquia*:

> So, whoever apprehends what is transmitted in the sciences, admits without any hesitation that this is absolutely true; and it must be believed that it could not be apprehended, if it were not illuminated by another sun, as it were, of its own.
>
> (*Soliloquia* I.5)

Platonic resonances in this passage are obvious. What exactly is meant by 'illumination' in this context is, however, less immediately obvious. What is the cash value of this light metaphor? Why does Augustine insist that whatever is apprehended through the sciences could not be apprehended if it were not appropriately 'illuminated'?

Perhaps the basic idea in Augustinian illumination is a generalization of the problem of learning by ostention. No group of instances of F-ness will display F-ness unambiguously as the single feature those items have in common; thus, if we ever come to understand what F-ness is, it will be only by an inner illumination that reveals something that cannot be unambiguously pointed to or displayed. In *De magistro*, Augustine concludes that no 'outward' teacher can teach us what anything is by asking or telling us something. At most, the 'outward' teacher can admonish or remind us to look 'within', where Christ the inner teacher dwells. Christ is identified as 'the

unchangeable excellence of God and the everlasting wisdom that every rational soul does indeed consult' (*De magistro* 11.38) (see Illumination).

More generally, Augustine, like Plato before him, insists that 'intelligible realities', presumably including what we might think of as a priori truths, cannot be learned or even confirmed in sense experience. However, Augustine explicitly rejects Plato's idea that the soul might have been introduced to the intelligible realm before birth; instead he espouses innatism. Referring to the Socratic interrogation of the slave boy in Plato's *Meno*, which is meant to demonstrate that there is latent knowledge of geometry even among the untutored, Augustine protests that not all would have been geometricians in their previous life, 'since there are so few of them in the human race that one can hardly be found.' He goes on:

> We ought rather to believe that the nature of the intellectual mind is so formed as to see those things which, according to the disposition of the Creator, are subjoined to intelligible things in the natural order, in a sort of incorporeal light of its own kind.
>
> (*De trinitate* XII.15.24)

5 Philosophy of mind: dualism and memory

Although arguments for soul-body, or mind-body, dualism are almost as old as Western philosophy, all such reasoning before Augustine seems to have started from an externalist or third-person perspective. Thus when Plato in the *Phaedo*, for example, has Socrates claim that we have knowledge we could not have acquired in this life and therefore our souls must have existed before they took on this bodily form, he is reasoning from a perfectly impersonal point of view.

By contrast, Augustine's approach is ego-centred, or first-personal. In Book VIII of *De trinitate*, he claims that we, from our own individual first-person perspective, 'know what a soul is, since we have a soul'. In the next book he adds, 'When the mind knows itself, it alone is the parent of its own knowledge' (*De trinitate* IX.12.18). In the following book he surveys philosophical theories about what the mind is (for example, that it is blood, that it is the brain, that it is a collection of atoms, that it is air, that it is fire, that it is some 'fifth body' or that it is a harmony). Reflection on this divergence of philosophical views suggests that the mind does not really know what it is. Yet if, as Augustine has maintained, the mind knows what a mind is from its own case, then it knows and is certain what a mind is. Therefore if the mind is uncertain whether it is blood, or the brain, or fire, or air, or indeed anything corporeal, it is really none of these things; if it is in fact uncertain as to whether it is any of these corporeal things, then it is none of them (*De trinitate* X.10.16).

Of course, the idea that the mind can know what a mind is only from its own case raises the problem of other minds. How do I know that there are minds to go with other human or animal bodies, if all I can ever observe are the motions of those bodies? Augustine's response is the earliest statement of the notorious argument from analogy. 'Just as we move our body in living', Augustine writes, 'so, we notice, those bodies are moved'. We come to think that there is something present in another body 'such as is present in us to move our mass in a similar way'(*De trinitate* VIII.6.9).

Recently, philosophers such as Norman Malcolm (1963) have argued that it is a mistake to suppose that one could come to know from one's own case what a mind is, or what thinking or feeling or having a pain is. They argue that we need a criterion to determine whether x is in pain or whether y even has a mind. Since one does not use a criterion in one's own case, they continue, either the argument from analogy could never get started (because we lack the needed criterion) or it is otiose (because we have such a criterion). Argumentation of this sort is often taken to be directed at Descartes' philosophy of mind; but in fact it is Augustine, rather than Descartes, who offers the argument from analogy and who maintains that each of us knows what a mind is from our own case, and can know what a mind is only from our own case (see Descartes, R. §§7-8).

Although Augustine's theory of sense perception seems not to be representational in the sense of making an image or sense-datum the object, or at least the direct object, of perception, his account of memory does seem to be representational. He is inclined to think of what is remembered, or perhaps what is remembered directly, as an image rather than as what the memory image portrays. Indeed, in *De magistro* he asserts flatly, 'When a question arises not about what we sense before us, but about what we have sensed in the past, then we do not speak of the

things themselves, but of images impressed from them on the mind and committed to memory' (*De magistro* 12.39). Augustine concludes, incorrectly, from the fact that we have no direct access to absent persons and things, or past events, that there is no way one can make claims about *them*, as distinct from their images, and hence there is no way we can, either falsely or truly, remember those past or absent things themselves.

Although Augustine sometimes uses *memoria* in ways that suggest that it means simply 'memory', he also uses the term much more broadly. Thus Book X of the *Confessiones* is a treatise on *memoria* as mind. In that book, Augustine tells us that the storehouse of his *memoria* includes not only images of objects and of past events but also feelings and experiences, certain facts, himself (!), and principles and laws of various sorts.

About feelings, Augustine asks how one could know what the terms for them mean unless on hearing, for example, the term 'pain', one had in one's *memoria* what that term means. However, if one had pain itself in one's *memoria*, he reasons, it would hurt just to think about the meaning of the word. Yet surely, he goes on, one can think about what 'pain' means, or what a pain is, without being in pain. But how? Augustine is not able to say. Although this apparently original criticism of imagist accounts of meaning seems quite devastating, it did not lead Augustine himself to turn away from his effort to appeal to images in explaining how words for feelings, or moods (such as 'happy' and 'sad') have the meaning they do.

6 Philosophy of mind: will

As Albrecht Dihle (1982) and others have argued, the concept of the will, important in much medieval and modern philosophy including the idea of a volition as an act of will, can be said to originate with Augustine. Plato's division of the self in his *Republic* into reason, spirit and appetite, by contrast, seems to make no room for the will as a distinct faculty or power, and Aristotle's subtle discussion of the voluntary in Book III of his *Nicomachean Ethics* seems not to presuppose the idea of any such force or power as the will.

In *De trinitate*, Augustine presents among other putative analogies to the divine Trinity the mental trinity of memory, understanding and will. His suggestion is that just as each of the three divine persons, Father, Son and Holy Spirit, is distinct from the other two - though each is also God and not just part of God - so memory, understanding and will are each distinct from each other, though each is also mind and not just part of mind. Yet Augustine also recognizes that the human mind, with its various trinities, is only an imperfect image of the divine Trinity. The mind-as-will may well operate in opposition to the mind-as-understanding; indeed, a will that is evil does just that. The idea of there being such a thing as what the Greeks called *akrasia* - doing what one knows one ought not to be doing - is thus not a conundrum for Augustine, in the way that it is for Socrates and Plato, or even for Aristotle in *Nicomachean Ethics* VII 3 (see Akrasia).

Sometimes Augustine suggests that what he terms 'the first cause of sin', namely the will, is itself uncaused. 'What cause of the will could there be', he asks rhetorically in *De libero arbitrio* (III.17), 'except the will itself?' This point is important to Augustine since, as he claims in this passage, there would be no such thing as sin if the acts of one's will were caused by something extraneous to the will. The idea of the will as a first, uncaused cause, should be coupled with Augustine's claim that the will is 'in our power'. Since it is in our power, he writes, it is free for us (*De libero arbitrio* III.3).

There is also, however, a competing suggestion also to be found in Augustine, to the effect that God, through his knowledge, is the cause of all that he foreknows, including free choices of the will (see §7). Since Augustine defines 'will' (*voluntas*) as 'a movement of the soul, under no compulsion, either toward getting or not losing something' (*De duabis animabus contra Manichaeos (On the Two Souls)* 10.14), it follows that the will is free from compulsion. In discussing this definition Augustine makes clear that ignorance, though it may excuse an action that would otherwise count as sinful, does not render the will unfree (*Retractationes* 14.3).

In his efforts to define and reject Pelagianism as a heresy, however, Augustine has to explain how the grace of God can work on the human will without destroying its freedom. 'Do we then by grace make free will void?' is how he puts the challenge to himself in *De spiritu et littera (On The Spirit and the Letter)* XXX.52. His answer is that, no, 'free will is not made void through grace, but is established, since grace cures the will whereby righteousness is freely loved.' This is the doctrine of 'prevenient' grace, which Augustine distinguishes from various other kinds of grace. The concept of the will is thus a point of intersection for several characteristic Augustinian doctrines. First, human wills are depraved as a result of Adam's original sin. Second, a human will is able to act rightly only

through the grace of God. Third, if our wills were not free, we could not be justly punished for choosing wrongly. Fourth, God foreknew who among human creatures would choose rightly and become saints, and he also predestined them, or foreordained that they would do so. It seems that Augustine's emphasis on the efficacy and importance of God's grace must inevitably qualify his defense of free will. Nevertheless, the anti-Pelagian writings Augustine composed in the last two decades of his life are eloquent testimony to the ingenuity with which he sought to reconcile the doctrine of divine grace with his insistence that human beings are free to make their own choices (see Grace).

Augustine's efforts to establish the compatibility of God's foreknowledge with human free will were almost as extensive as his attempts to show that God's grace is also compatible with free will. However, Augustine paid much less attention to the worry that predestination also poses a threat to the possibility of free will. Thus in his *De praedestinatione sanctorum (On the Predestination of Saints)* 37, he insists that God chose the elect 'before the foundation of the world', not simply because they were in fact going to be 'holy and immaculate' but that they might become so. Clearly he thinks of predestination as something additional to mere foreknowledge. So, even if we understand how it can be that God's foreknowledge does not render our choices unfree, we will still need to deal with the additional threat to our freedom that seems to be posed by predestination (see Predestination).

Knowing, as already noted, can be thought of as a sort of seeing; and we are not usually tempted to think of seeing as determining the nature of what is seen. However, predestination is a form of predetermination, and how predetermination can be compatible with human freedom is much more difficult to understand. Augustine, though he certainly does not ignore this problem, is less helpful in showing how it can be dealt with than he is with the problem of foreknowledge and free will.

7 Metaphysics: God and divine attributes

In *De libero arbitrio* (II.15.39), Augustine offers an argument for the existence of God. He first gets his interlocutor to admit that (1) *x* is God if and only if *x* is more excellent than our minds and nothing is more excellent than *x*. He then tries to establish that (2) truth exists and is more excellent than our minds. From these two premises he concludes that (3) something is God (in other words, God exists). His idea is that either nothing is more excellent than truth, so that, since truth is more excellent than our minds, truth itself is God; or else something is more excellent even than truth, in which case it (or, we could add, something even more excellent than it) is God.

In fact, Augustine has the idea that God is not only something more excellent than our minds, but is also something than which our minds can conceive nothing more excellent. Thus he writes that 'God is the supreme good, and that than which nothing can be nor can be conceived to be better.' This formula is remarkably close to the one Anselm uses in formulating the ontological argument: 'For we believe that you are something than which nothing greater can be conceived' (*Proslogion* 2) (see God, concepts of).

Augustine made important contributions toward the project of giving a philosophical analysis of the divine attributes, or divine 'names', as philosophers of the high Middle Ages sometimes called them. Thus he made clear that, although God is omnipotent, he is unable to do certain things, such as sin, die or make a mistake. According to Augustine, '*x* is omnipotent' means that '*x* does whatever *x* wills'. However, Augustine does not deal with seeming difficulties posed by this account, for example, whether 'if God wills to sin, God sins' can be true even if its antecedent is necessarily false, or what to say about an agent with only minimal power whose wants, because they are also minimal, never outrun the agent's power to satisfy them (see Omnipotence).

Augustine is perhaps the first in a long line of philosophers to maintain that God is perfectly simple, that is, that God is identical with his attributes (*De civitate Dei* XI.10; *De trinitate* VI.7.8). Thus God is his goodness, he is his wisdom, and so forth. Later philosophers who made this claim central to their philosophical theologies include Anselm and Aquinas (see Immutability; Simplicity, divine).

Augustine links divine simplicity to God's immutability. Sometimes he even explains simplicity in terms of immutability. 'The reason why a nature is called simple', he writes, 'is that it cannot lose any attribute it possesses, that there is no difference between what it is and what it has' (*De civitate Dei* XI.10). He goes on to add that even the incorruptible body we shall receive in the resurrection is neither simple nor unchangeable. That body will not be simple, he maintains, because the substance 'in virtue of which it is called a body is other than the

quality from which it derives the epithet incorruptible.' Even the human soul, he continues, is other than its wisdom. Only God *is* his wisdom.

In his effort to clarify the divine attributes, Augustine sometimes admits defeat. Thus, in Book I of *De doctrina Christiana*, he concedes that his efforts to clarify the doctrine of the Trinity have been unworthy of God and suggests that God is ineffable. He then notes that to say God is ineffable is to describe God and therefore to show, by saying it, that what one says is false.

8 Metaphysics: creation, time and eternity

Plato, in his *Timaeus*, pictures creation as the action of a divine craftsman who looks at eternal paradigms to form instances in an also pre-existent 'receptacle'. Augustine, interpreting Genesis 1: 1 to mean that God created the heavens and the earth out of nothing (*ex nihilo*), insists that God created unformed matter, as well as what he formed from it. His rejection of the Platonic alternative is explicit: 'You did not hold anything in your hand, of which you made this heaven and earth; for how could you come by what you had not made, to make something? For what is there that exists, except because you exist?' (*Confessiones* XI.5.7). Augustine's assumption is that nothing exists, except because God exists. Moreover, because everything changeable has a beginning, and the heavens and the earth are certainly changeable, God created them.

Not only did God create the world, and create it out of nothing, according to Augustine, he also conserves or sustains it, lest it disappear into nothing. If God's power 'ever ceased to govern creatures', he writes:

> their essences would pass away and all nature would perish. When a builder puts up a house and departs, his work remains in spite of the fact that he is no longer there. But the universe will pass away in the twinkling of an eye if God withdraws his ruling hand.
>
> (*De genesi ad litteram* IV.12.22)

Here is the precursor of Descartes' idea that 'the same power and action are needed to preserve anything at each individual moment of its duration as would be required to create that thing anew if it were not yet in existence' (*Meditation* III) (see Creation and conservation, religious doctrine of).

If the idea that God created matter is philosophically troublesome, even more puzzling is the contention, which Augustine affirms, that God created time. To make sense of that claim, one must study what Augustine says about time. In a much cited passage, Augustine asks, 'What then is time?' and responds: 'If nobody asks me, I know; but if I am asked what it is and want to explain, I don't know how to' (*Confessiones* XI.14). Wittgenstein (1958a: 26) thought this question epitomized philosophy as a misguided search for definitions (see Wittgenstein, L. §§15-17). By contrast, Bertrand Russell praised Augustine's theory of time as 'a very able theory, deserving to be seriously considered', and 'a great advance on anything to be found on the subject in Greek philosophy' (Russell 1945: 354).

In his discussion, Augustine appeals to the fact that we speak of 'a long time' and 'a short time'. He then develops a perplexity that Aristotle had raised in a highly condensed form in Book IV of his *Physics* but never resolved. 'How can anything that does not exist', Augustine asks, 'be either long or short?' (*Confessiones* XI.15.18). But the past is no more and the future is not yet, he continues; only the present exists. 'Can present time be long?' he asks. For any present period of time, he notes - the present century, the present year, the present day, or whatever - part of it is no more and part is not yet. All that is ever really present is the instantaneous divider between what is not yet and what is no more. As a simple divider between the past and the future, the present, the 'now', has no duration at all and so cannot be either long or short. Therefore, how can times be either long or short? Augustine concludes that it is in his own mind that he measures time. It is mental impressions that one measures, he says, and therefore time is the measure of something mental.

Early on in Book XI, Augustine had sought to deal with the sceptical question, 'What was God doing before he made heaven and earth?' (*Confessiones* XI.10.12). His answer is to say that 'there cannot possibly be time without a creature [*sine creatura*]' (*Confessiones* XI. 30), presumably because the mind of God is timeless, as time is a measure of created minds.

In Book XI of *De civitate Dei*, Augustine seems to revert to a more Platonic theory of time, according to which

time is the motion of bodies, perhaps especially the motion of the heavenly bodies. The explicit question up for discussion is why God made the world at one time rather than another. Augustine may have derived this sceptical question, and the earlier one in *Confessiones* introducing the discussion of time, from Cicero's *De natura deorum*, although Cicero was himself passing on discussions that have their roots in Presocratic philosophy.

It would have been odd for Augustine to conclude that time is the measure of bodily motion, since he explicitly rejects such a theory in *Confessiones* with the apt objection that time may equally be the measure of a body's rest. In Book XII of *De civitate Dei*, however, Augustine makes it clear that he considers the movements of the angels to be sufficient for there to be time. Angels have always existed, he claims, in the sense that there is no time at which they failed to exist. Nevertheless, they are created beings; indeed time was created with them.

In supposing that God creates time, Augustine does not of course suppose that this creation took place in time. 'There can be no doubt', he writes, 'that the world was not created *in* time but *with* time' (*De civitate Dei* XI.6). Augustine's idea is that God timelessly creates time; indeed, God's own being is immutable and timeless. In supposing that God timelessly creates time, Augustine avoids a paradox in the idea of time having been created. One might otherwise suppose that the idea of time having been created must somehow involve a contrast between a time when there was no time and the temporal period that began with time's creation. The absurdity of that idea seems to have led Aristotle to reason that time, like the universe, is eternal. But Aristotle did not consider the Augustinian idea that God timelessly creates time, perhaps even a first time (see Time).

Though Augustine often uses temporal words to talk about God and God's actions, those words are to be understood in a special sense. Sometimes Augustine warns his readers of this need. Thus he writes in *Confessiones* (XI.13.16), addressing God: 'It is not in time that you are before times; otherwise you would not be before all times; you precede all past times in the loftiness of your ever-present eternity.' He adds, 'Your years, because they abide, all abide at once.'

One might suppose that, if God is eternal in the sense of being 'before' or 'outside' all time, God could have no causal efficacy *in* time. With similar reasoning, Aristotle complained that Plato's Forms, being eternal and unchanging, could not be the cause of something's happening now rather than earlier or later. With Aristotle's criticism in mind, a Christian believer might complain that Augustine's depiction of God as being out of time means that God could not act in history. Augustine is, however, in a rather different position from Plato, since he supposes God to have created time. Though we may speak of God's actions *in* history, God in Augustine's view brings it about timelessly; there is time in which the entities of creation can act and be acted on by him. In the dialogue *De libero arbitrio* (II.3), the character Evodius maintains that 'God has decided once and for all how the order of the universe he created is to be carried out, and does not arrange anything by a new act of will' (see Eternity).

Following in the Platonic and Neoplatonic tradition, Augustine supposes there are different senses of 'exist', or ways in which something may be said to exist. He assigns the highest way, or the strictest sense, to the existence of God: 'Being is in the highest and truest sense of the term, proper to Him from whom being derives its name' (*De trinitate* V.2.3). Here Augustine links Platonic metaphysics with God's saying to Moses in Exodus 3:14, 'I am who am.' God truly is, Augustine writes in *De natura boni* 19, because he is immutable. 'For every change', Augustine goes on, 'makes what was, not to be; therefore he truly is who is unchangeable.'

9 Philosophy of religion: problem of evil

Although the author of the biblical book of Job questioned how it can be that unjust suffering is compatible with the nature of God, he seems not to have had a problem with the mere existence of evil. It may be that Epicurus was the first thinker to ask how the existence of evil could be compatible with the nature of God (Lactantius, *De ira Dei (The Wrath of God)* 13).

In *Confessiones* (VII.5.7), Augustine raises this compatibility problem in a particularly persistent way. Put in a somewhat more modern way, the problem is how the statements (1) God is all-good and (2) God is all-powerful, can be consistent with (3) there is evil. Although the conjunction of these three statements is not logically inconsistent, it is natural to assume that (4) if there existed some being both all-good and all-powerful, evil would not exist. The conjunction of (1), (2), (3) and (4) is indeed logically inconsistent (see Evil, problem of; Leibniz, G.W. §3).

When Augustine was a Manichean, he rejected (2) and, insofar as for Manicheans the principle of evil was coequal with the principle of good, there is a way in which he could also be said to have rejected (1). However, as Augustine moved toward Christianity, those options became closed. In *Confessiones* (VII.5.7), Augustine considers rejecting (3). ‘Can it be’, he asks, ‘that there simply is no evil?’ But if there really is no evil, he reasons, our fear of evil is unfounded; however, an unfounded fear of evil would itself be evil. If there is no evil, therefore, there is evil; so, there is evil. Augustine returns a few sections later in the same work to embrace the Neoplatonic idea that evil is not a reality but a mere privation and so, in a way, does not exist. Yet the fear of something nonexistent can itself be evil.

In *De civitate Dei* (XII.8), Augustine finds the root cause of evil in human free will. ‘A will could not become evil’, he writes, ‘if it were unwilling to become so; and therefore its shortcomings are justly punished, being not necessary but voluntary.’ In insisting that the existence of free but evil choices of human agents is compatible with the all-goodness and all-power of God, Augustine thus rejects (4) above. Since the problem of evil, conceived as a problem about the apparent inconsistency among (1), (2) and (3), is a purely *a priori* problem, all one would need to do to solve it would be to show that (4) is not a necessary or conceptual truth. Augustine, however, tries to do more; he tries to show that (4) is actually false.

When Augustine uses what has recently come to be called the ‘free-will defence’ to solve the problem of evil, he does not reject the logical possibility that human beings might have been created to have free will, yet never sin. For example, he supposes that the saints in heaven will have free will, yet they will sin no more. ‘Now the fact that they will be unable to take pleasure in sin’, he writes in *De civitate Dei* (XXII.30), ‘does not entail that they will have no free choice. In fact, the will will be the freer in that it is freed from taking pleasure in sin and immovably fixed in the pleasure of not sinning.’ Adam’s freedom, Augustine goes on in the same passage, was an ability not to sin, combined with the possibility of sinning. The freedom the saints will enjoy in heaven will bring with it, he says, the impossibility of sinning. Lest we protest, in the manner of philosophers who have discussed this issue in our own time, that the impossibility of sinning would destroy free will, Augustine adds that God cannot sin, yet surely God has free will.

Thus Augustine’s rejection of (4) does not rest on a rejection of the logical possibility that God might have created a sinless world in which agents exercise their free will. Rather, it rests on Augustine’s belief that the impossibility of sinning, when it finally comes to the saints in heaven, is a free gift of God’s grace, not anything God is constrained to bestow on creatures, not even by his all-goodness. Another way that Augustine tries to show (4) is false is to suggest that evil, or sin, is like a dark colour in a beautiful painting: in itself, ugly, but in context something that contributes to the beauty and goodness of the whole (*De civitate Dei* XI.23).

10 Philosophy of religion: divine foreknowledge and free will

The problem of how God’s foreknowledge of what an agent will do is compatible with that agent’s acting freely seems to be a specific version of the problem Aristotle discusses in *De interpretatione* about how the prior truth of a statement about a future event is compatible with that event’s being contingent. Aristotle’s discussion and the commentaries on it by Boethius, as well as the latter’s own discussion of the problem of foreknowledge in Book V of his *De consolatione philosophiae (The Consolation of Philosophy)*, were influential on later medieval discussions. Also influential, however, was Augustine’s discussion in Book III of his *De libero arbitrio*, and perhaps even more so in Book V of the *De civitate Dei*.

In *De civitate Dei*, Augustine takes Cicero as his target. As represented by Augustine, Cicero argues that (1) if God foreknows all events, then all events happen according to a fixed, causal order, and (2) if all events happen according to a fixed, causal order, then nothing depends on us and there is no such thing as free will. If we can then add (3) God foreknows all events, as Augustine insists we can, we can conclude that (4) nothing depends on us and there is no such thing as free will. Augustine attempts to defeat this argument by insisting that ‘our wills themselves are in the order of causes’ (*De civitate Dei* V.9), and adds it is necessary that, ‘when we will, we will by free choice’ (V.10). Moreover, among the things that God foreknows are the things that we will to do of our own free choice. Thus premise (2) is false and God’s foreknowledge, rather than being a threat to free will, is in a way its guarantor.

Talk of God’s foreknowledge, as Augustine himself realizes and points out, is somewhat misleading. God, as an

eternal being, is outside time. However, for any event that to us temporal creatures is still in the future, it is true to say of God that God knows, as we should say from our own point of view, 'already' what will happen. Augustine's explicit discussion of the problem of foreknowledge and free will proceeds as though God's knowledge perfectly reflects the causal order in which human wills operate freely, without itself causing that order. Thus he insists in *De civitate Dei* (V.10), for example, that 'it is not that one sins because God foreknew that one would sin'. But in fact, Augustine, like Aquinas after him, seems to hold the view that God's knowledge actually causes to happen what God knows will happen.

The context for Augustine's assertion that God causes what he knows is a discussion of how God can know the material world without himself having a body or any sense organs. Rejecting the suggestion that God might need messengers of any sort, whether sense organs or angelic witnesses, Augustine boldly denies that God's knowledge is in any way dependent on what he knows. God does not know creatures because they are, he writes in *De trinitate* (XV.13.22); rather, they are because God knows them. Indeed, Augustine adds in the same passage, God learns nothing about creatures from the creatures themselves. If then we are unable to surprise God with any of our free choices, and God's perfect and immutable knowledge not merely reflects but actually causes us to be the (to God) unsurprising creatures we are, how is there any metaphysical room for human free will? And how can it not be the case that we sin because God foreknows we will?

It seems to be Augustine's view that God's foreknowledge that, for example, Peter would deny Christ three times somehow causes Peter's denial without making it unfree and, therefore, without detracting from Peter's responsibility for it. Indeed, the view seems to be that God's foreknowledge of a free choice can cause there to be the very free choice that there is. Moreover, even though the choice is caused by God, it is the agent, not God, who makes the choice. Since the free choice is voluntary, the agent is responsible for it.

Of course, the problem mentioned earlier as to how predestination is consistent with free will is directly relevant here. If Augustine's account of God's grace can make clear how it is possible for God to foreordain and predestinate human choices without rendering those choices unfree, then perhaps it will also make clear how God can cause them without making them unfree. On the face of it, however, the claim that (1) God causes everything he foreknows, including the free choices human beings make, seems simply to contradict the assurance that (2) 'it is not that one sins because God foreknew that one would sin' (*De civitate Dei* V.10). If there should be no satisfactory way of reconciling (1) and (2), (1) should be qualified or rejected as a true reflection of Augustine's view, since it is far less prominent in Augustine's writings than the repeated insistence on (2) (see Free will; Omniscience; Predestination).

11 Philosophy of religion: faith, reason and mysticism

When discussing the relationship of faith and reason, Augustine characteristically insists that faith must precede understanding. 'For understanding is the reward of faith', he writes in his *In Ioannis evangelium tractatum (Homilies on the Gospel of John)*, adding: 'Therefore do not seek to understand in order to believe, but believe that you may understand' (29.6).

The comment of Augustine's interlocutor, Evodius, in the dialogue *De libero arbitrio* (II.2), 'But we want to know and understand what we believe', is apt, not only for much of Augustine's own writing but for much of medieval philosophy as well. The *Proslogion*, the treatise in which Anselm develops his famous ontological argument, carries the Augustinian subtitle, 'Faith in Search of Understanding'.

In Question 48 of *De diversis quaestionibus LXXXIII (Eighty-Three Different Questions)*, Augustine divides the classes of things to be believed into three. In the first class, ranging over the temporal dealings of human beings, are things that are 'always believed and never understood'. In the second group, which are 'understood as soon as they are believed', are 'human reasonings'. It is the third group, which concern divine dealings, that are first believed and afterward understood. However, even if 'I believe in order that I may understand' applies only to theological questions, there is still a problem. How can one believe something before one even understands it? As Augustine himself points out in his sermons, his hearers need to understand what he is saying before they are in position to believe what he says. How can one believe, for example, the biblical claim that the earth was once invisible and without form before one even understands what that claim means? In *Confessiones* (XII.23), Augustine makes a distinction between questions about truth and those about meaning. He seems to suppose one

can accept that, for example, the biblical account of creation in Genesis 1 is true without knowing what each of the statements that make it up means. Of course one would need to have some general idea of what those statements might mean in order for one's faith in the truth of the account to have any real content, but even a minimal understanding might be enough to give one's faith some purchase.

How, though, did Augustine think one is to choose among rival authorities in coming to accept, for example, some biblical or creedal statement, or some commentary on the Bible or the creed, as true? He seems to have allowed for both prior and subsequent contraints on what should be believed. As for prior constraints, he writes in *De moribus ecclesiae catholicae (Morals of the Catholic Church)* 7.11: 'We must have recourse to the teachings of those who were in all probability wise'. As for subsequent constraints, he looked for predictions that are not borne out as discrediting a putative authority. Thus he discredited astrologers in *Confessiones* (VII.6) by pointing out that it sometimes happens that a slave in misery has a twin brother who prospers as a freeman, though the twins were born under precisely the same astrological sign.

Later in *Confessiones* (VII.17), Augustine describes an early attempt at mystical ascent. Though 'in an instant of awe', his mind, he says, caught sight of God, he had not the strength to continue the vision; the memory remained with him as something he longed for. In the same work, he describes a more prolonged ascent that grew out of conversation with his mother:

> As the flame of love burned stronger in us and raised us higher towards the Selfsame, we passed through all corporeal things, and the heavens themselves, from which the sun and the moon and the stars shine down upon the earth…. At length we came to our own minds and passed beyond them to that place of everlasting plenty, where you feed Israel forever with the food of truth. There life is the Wisdom by which all things are made.
>
> (*Confessiones* IX.10)

Shortly thereafter Augustine's mother died.

Augustine wonders whether the moment when he and his mother 'reached out in thought and touched the eternal Wisdom' prefigured life in heaven. In perhaps his most eloquent description of the beatific vision, Augustine suggests that such an experience may help one live a more moral life on earth:

> There the virtues of the soul are not laborious and wearisome. For there desire is not bridled by the work of temperance, or adversities borne by the work of fortitude, or iniquities punished by the work of justice, or evils shunned by the work of prudence. There the one virtue and the whole of virtue is to love what you see and the greatest happiness is to have what you love. For there the heavenly life is drunk at its source, from which a little is splashed over onto this human life so that it is lived among the temptations of this world with temperance, with fortitude, with justice, and with prudence.
>
> (*De genesi ad litteram* XII.26.54)

In Chapter 7 of the treatise *De cura pro mortuis (On the Care of the Dead)*, Augustine raises a problem for the understanding of religious ritual. He wonders what can be the point of bending one's knees, holding out one's hands or prostrating oneself in supplication to God when the prayer we have in our hearts is already known to God. Augustine's problem is really two problems. First, why should we go through a prayer, whether inwardly or outwardly, when God already knows what we want to say? Second, why should we give outward expression to a prayer, since God, the intended recipient, knows already what we say inwardly in our hearts? He then gives an answer which is intended to apply to both problems: 'Although these motions of the body cannot come to be without a motion of the mind preceding them, when they have been made, visibly and externally, that invisible inner motion which caused them is itself strengthened' (*De cura pro mortuis* 5.7). The idea is that going through a prayer intensifies the thoughts and feelings that the prayer expresses, which answers the first problem. Moreover, doing this outwardly, with appropriate bodily motions, strengthens the inner feelings that such bodily motions express, thus answering the second problem. Augustine must have known as well as we do today that ritual sometimes actually gets in the way of our thinking the thoughts and having the feelings that the ritual is supposed to express. However, it would be enough for him if religious ritual regularly, or even just often, had the effect of nurturing and strengthening religious feelings and attitudes.

An odd thing about Augustine's response to his problem is that, according to his own metaphysical principles, the

response is inappropriate, or even false. The reason is that, in Augustine's view, the soul or mind is superior to the body and what is superior cannot be affected by what is inferior to it (*De genesi ad litteram* XII.16.33). So, neither the physical productions of sound in saying a prayer nor the physical movements of kneeling or stretching out one's hands can possibly have the effect of intensifying thoughts or feelings.

No doubt what Augustine should have concentrated on is not the bodily motions of tongue, hand or knee in prayer, but rather the importance to a relationship, whether it is a relationship with God or with another human being, of giving expression to one's feelings of gratitude, remorse or whatever else might be appropriate. Thus God's omniscience is really inessential to the issues Augustine is raising. Because some person, A, knows another person, B, very well, it might be the case that A knows that B is sorry for having wronged A, even though B has not yet apologized or asked for forgiveness. In that case, for A to apologize or ask for forgiveness would not have the purpose of passing on information to A; nevertheless, it might be important to their relationship. This might also be true of a relationship between a human being and God. This alternative response to the problems Augustine raises about ritual should remind us of Augustine's tendency, noted in the discussion of *De magistro* above, to try to reduce the role of language as well as that of gesture to one of passing on information, or even to merely reminding us of what we already know.

12 Ethics: sin, vice and virtue

For Socrates, Plato and Aristotle, there existed the problem of how we can ever do what we know we ought not to be doing. This is the problem of *akrasia* (see Akrasia). From this ancient perspective, perhaps the most striking thing about Augustinian ethics is its easy acceptance of *akrasia*. In *Confessiones* II, Augustine tells of stealing pears as a boy of sixteen. He spends two chapters ruminating on what might have motivated his theft. It was not the pears themselves, he says, for he had better ones at home. He concludes that it was the flavour of sinning that motivated him.

In *De libero arbitrio* (I. 2), Augustine admits that the question of why we do evil disturbed him greatly when he was young and moved him toward Manicheism. Once he accepted the idea of original sin, however, he found nothing paradoxical in saying of someone: 'He hates the thing itself because he knows that it is evil; and yet he does it because he is bent on doing it' (*De nuptiis et concupiscentia (On Marriage and Concupiscience)* I.28.31).

Augustine was an extreme intentionalist in ethics. In *De sermone Domini in monte (Commentary on the Lord's Sermon on the Mount)* I.12.34, he identifies three necessary and sufficient conditions for committing a sin: receiving an evil suggestion, taking pleasure in the thought of performing the act suggested and consenting to perform the act. Thus in Augustine's view, whether one commits a sin is in no way dependent on whether the contemplated action is actually carried out. Even when the action is carried out, it is the intention (understood as suggestion, pleasure and consent), rather than the action itself, or its consequences, that is sinful.

Augustine also devoted two treatises to the topic of lying. In the first of these, *De mendacio (On Lying)*, he first suggests that a person *S* lies in saying *p* if, and only if (1) *p* is false, (2) *S* believes that *p* is false and (3) *S* says *p* with the intention of deceiving someone. He then considers three cases: first, that of someone with a false belief who wants to deceive another by saying something that is, unknown to them, quite true; second, the case of someone who expects to be disbelieved and so knowingly says what is false in order to instill a true belief; and third, the case of someone who, also expecting to be disbelieved, knowingly speaks the truth in order to instill a falsehood. Augustine seems not to know what to do about these problem cases. He contents himself with insisting that the conditions (1)-(3) are jointly sufficient, without taking a stand on whether each is singly necessary (*De mendacio* 4.5).

Discussing virtue and vice, Augustine contrasts those things that are desirable in themselves with those that are desirable for the sake of something else. He says that things of the first sort are to be enjoyed (*frui*) whereas those of the second sort are to be used (*uti*). Vice, he adds, is wanting to use what is meant to be enjoyed or wanting to enjoy what is meant to be used (*De diversis quaestionibus LXXXIII* 30).

Ambrose had already added the Pauline virtues of faith, hope and love to the classical virtues of temperance, courage, wisdom and justice. Augustine follows Ambrose in this, and he follows St Paul in assigning first importance to love; in fact, he offers an interpretation of each of the seven virtues that makes it an expression of the love of God. Thus temperance is love 'keeping itself whole and incorrupt for God'; fortitude, or courage, is

love ‘bearing everything readily for the sake of God’, and so on (*De moribus ecclesiae catholicae (On the Morals of the Catholic Church)* 15.25). Virtue, he says, is nothing but the perfect love of God. In this way Augustine provides a Christian analogue to Plato’s idea of the unity of the virtues (see Virtues and vices).

13 Ethics: ‘ought’ and ‘can’

Augustine also attacked the Pelagians for their views on the avoidance of sin, focusing on the question of ‘ought’ and ‘can’. Two of his contemporaries, the British monk Pelagius and his disciple Coelestius, had made the principle that ‘ought’ implies ‘can’ a central tenet of their religious and ethical teaching. As already noted, Augustine was the person primarily responsible for defining their teaching, Pelagianism, as a Christian heresy (see Pelagianism). In his treatise *De perfectione justicia hominis (On Man’s Perfection in Righteousness)*, subtitled ‘In opposition to those who assert that it is possible for one to become righteous by one’s own strength alone’, Augustine describes the chief thesis of Coelestius as the contention that if something is unavoidable, then it is not a sin; there is simply no such thing as an unavoidable sin. Augustine responds to Pelagius and his disciple by rejecting the simple disjunction that either something is not a sin or it can be avoided. ‘Sin can be avoided’, he writes, ‘if our corrupted nature be healed by God’s grace.’ Thus in a way, Augustine agrees that ‘ought’ does imply ‘can’, but only with a crucial qualification. ‘Ought’ implies ‘can with the gratuitous assistance of God’, but it does not imply ‘can without any outside help’ (see Kant, I.).

Augustine’s response to dreaming as a possible threat to knowledge claims fits together with his intentionalism in ethics and his anti-Pelagianism to produce an interesting problem as to whether one is morally responsible for the acts of one’s dream self. He agonizes over this problem in *Confessiones* (X.30). Three ways of justifying a claim of no responsibility suggest themselves. I could say I am not responsible (1) because I am not my dream self, or (2) because what happens in a dream does not really happen, or (3) because I am powerless to avoid doing what my dream self does, and ‘ought’ implies ‘can’.

Augustine’s philosophical and theological commitments seem to undercut each of these three responses. Thus (1) is undercut, it seems, by his somewhat concessive response to scepticism. I can know that something tastes sweet to me, Augustine insists in *Contra academicos* (III.11.26), whether or not I am dreaming. It seems to be a consequence of this insistence that, if I am dreaming, I am my dream self. As for (2), it seems to be undercut by Augustine’s strong intentionalism in ethics. Thus when I commit adultery in my dreams, even if no ‘outward’ adultery takes place, still I entertain the evil suggestion, take pleasure in the evil suggested and give consent; so there is wrongdoing. As for (3), as noted above, Augustine rejects the Pelagian insistence that ‘ought’ implies ‘can’. or rather, he accepts it only with an important qualification. Although ‘I ought to refrain from consenting to fornicate’ does, in Augustine’s view, entail that I can so refrain with the help of God’s grace, it does not entail that I can refrain strictly on my own, that is, without any divine grace. Yet if I receive no grace and consent to fornicate, I sin, according to Augustine, and it is just for God to punish me.

14 Ethics: on killing

Although Augustine’s thoughts on suicide are not particularly original, they have been extremely influential. His position became Christian orthodoxy, which in turn influenced decisively the legal thinking in predominantly Christian countries. Augustine’s position is that, with certain specifiable exceptions (primarily, lawful executions and killings in battle by soldiers fighting just wars (see below), anyone who kills a human being, whether himself or anyone else, is guilty of murder (*De civitate Dei* I.21), and murder is prohibited by divine commandment (see Death; Suicide, ethics of).

Augustine did not invent the idea that certain requirements must be satisfied if a war is to count as just. The theory of just warfare - both the conditions that must be satisfied if a war is to be entered into justly (*jus ad bellum*) as well as the requirements of justice in the waging of war (*jus in bello*) - are already well developed by Cicero in his *On the Republic*. Nor was Augustine the first Christian thinker to develop a theory of just warfare; Ambrose had already done so (Christopher 1994). Nevertheless, Augustine is usually considered the father of the modern theory of the just war. Such deference is appropriate in that it is in Augustine, more than in Cicero or Ambrose or anyone else in the ancient world, that later theorists have found their earliest inspiration.

Although Augustine accepts the commandment, ‘Thou shalt not kill’, he interprets it in such a way that not

everyone who brings about the death of another can be properly said to kill. Thus, he writes in *De civitate Dei* (I.21), ‘One who owes a duty of obedience to the giver of the command does not himself kill; he is an instrument, a sword in its user’s hand.’ Thus an executioner may bring about the death of a convict without killing, and so may a soldier end another’s life without killing, especially when war is being waged ‘on the authority of God’.

In general, Augustine takes over the Roman principles of just war as set forth by Cicero and adds his own emphasis on the intention with which the acts of war are performed. This following passage is characteristic:

> What is the evil in war? Is it the death of some who will soon die in any case, that others may live in peaceful subjection? This is merely cowardly dislike, not any religious feeling. The real evils in war are love of violence, revengeful cruelty, fierce and implacable enmity, wild resistance, the lust of power, and such like; and it is generally to punish these things, when force is required to inflict the punishment, that, in obedience to God or some lawful authority, good men undertake wars.
>
> (*Contra Faustum manichaeum* 22.74)

Beyond such insistence that war should not be fought from love of violence, revengeful cruelty or lust for power, Augustine did not work out specific principles for the just conduct of war. Still, in making it plausible to many Christians that killing in war need not fall under the divine commandment not to kill, Augustine freed others to develop principles for what might be considered the just declaration of war, as well as the just conduct of war, once it has been justly entered into (see War and peace, philosophy of).

15 Philosophy of history

In the *Poetics*, Aristotle remarks that poetry is more philosophical than history. He seems to have thought of the historian as the mere chronicler of events, whereas the tragedian tells us what sort of person would, probably or necessarily, do or say what sorts of things. Augustine was the first important philosopher to treat the writing of history in a more philosophical way. In *De civitate Dei*, he rejects the idea of eternal recurrence, later associated with Nietzsche, according to which:

> just as Plato, for example, taught his disciples at Athens in the fourth century, in the school called the Academy, so in innumerable centuries of the past, separated by immensely wide and yet finite intervals, the same Plato, the same city, the same school, the same disciples have appeared time after time, and are to reappear time after time in innumerable centuries in the future.
>
> (*De civitate Dei* XII.14)

Although Augustine does not offer philosophical arguments for his own ‘linear’ view of history, he does offer a variety of criticisms of what he takes to be the main philosophical argumentation in support of the cyclical theories (*De civitate Dei* XII.18).

The account of history Augustine offers is meant to be a universal history, that is, a history of the world. It divides the history of the world into seven ages, analogous to the seven days of creation. The discussion proceeds, however, more as a gloss on scripture than by appeal to rational considerations or nontheological evidence. Although he does not try to establish the nature of ‘God’s universal providence’ by general considerations or evidence that is independent of Christian revelation, he does seek to establish that divine providence, like divine foreknowledge and divine predestination, is compatible with human free will. On a more general level he tries to establish that history can have a meaning, in fact the meaning God foreordains it to have, even though human agents are perfectly free; and indeed, are foreordained to be free (see History, philosophy of).

See also: Augustinianism; Boethius, A.M.S.; Encyclopedists, medieval §2; Ancient philosophy; Manicheism; Neoplatonism; Patristic philosophy; Pelagianism; Platonism, medieval; Religion, history of philosophy of; Scepticism

GARETH B. MATTHEWS

List of works

Augustine (386-429) Collected Works.(The seventeenth-century Maurist edition of Augustine’s works is to be found in J.P. Migne (ed.) *Patrologia Latina*, Paris, 1844-6, vols 32-47. Critical editions of many of his works are to be found scattered through the Corpus Scriptorum Ecclesiasticorum, Vienna: Tempsky, 1866-, as well as

in the Corpus Christianorum, Series Latina, Turnholt: Brepols, and The Hague: Nijhoff, 1953-. English translations of Augustine's major works are to be found in P. Schaff (ed.) *A Select Library of the Nicene and Post-Nicene Fathers of the Christian Church*, First Series, Edinburgh: T. & T. Clark, 1886-8, repr: Grand Rapids, MI: Eerdmans, 1971-74. A smaller selection is found in W. Oates (ed.) *Basic Writings of Saint Augustine*, New York: Random House, 1948, 2 vols. Individual works are also to be found scattered through the series Fathers of the Church, Washington, DC: Catholic University of America, 1947-, and in the series Ancient Christian Writers: The Works of the Fathers in Translation, New York: Newman, 1946-. The first English translation of Augustine's complete works is currently being published in 46 vols, edited by J.E. Rotelle, Hyde Park, NY: New City Press, in conjunction with the Augustinian Heritage Institute.)

Augustine (386) *Soliloquia (Soliloquies)*, trans. C.C. Starbuck, in W. Oates (ed.) *Basic Writings of Saint Augustine*, New York: Random House, 1948, vol. 1, 257-97.(A dialogue between Augustine and Reason, this work assumes the first-person persepctive that marks one of Augustine's special contributions to philosophy.)

Augustine (386) *Contra academicos (Against the Academicians)*, trans. P. King (along with *The Teacher*), Indianapolis, IN: Hackett, 1995.(Augustine's most sustained effort to refute the scepticism of the New Academy.)

Augustine (387) *De dialectica (On Dialectic)*, trans. B.D. Jackson, Dordrecht: Reidel, 1975.(The first book in a planned but later abandoned series on the liberal arts.)

Augustine (387/8) *De quantitate animae (The Magnitude of the Soul)*, trans. J.J. McMahon, Fathers of the Church vol. 4, Washington, DC: Catholic University of America, 1947, 49-149.(Augustine's early philosophy of mind.)

Augustine (387-9) *De moribus ecclesiae catholicae (Morals of the Catholic Church)*, trans. R. Stothert in P. Schaff (ed.) *A Select Library of the Nicene and Post-Nicene Fathers of the Christian Church*, First Series, 1886-8, vol. 4, 37-63.(Includes a Christian definition of the classical virtues.)

Augustine (388-95) *De diversis quaestionibus LXXXIII (Eighty-three Different Questions)*, trans D.L. Mosher, Fathers of the Church vol. 70, Washington, DC: Catholic University of America, 1977.(Short discussions of a variety of metaphysical, epistemological and theological topics.)

Augustine (388-95) *De libero arbitrio (On Free Choice of the Will)*, trans. T. Williams, Indianapolis, IN: Hackett, 1993.(The first philosophically important treatment of how to reconcile human free will with divine foreknowledge.)

Augustine (389) *De magistro (The Teacher)*, trans. J.M. Colleran, *The Teacher* (along with *The Greatness of the Soul*), New York: Newman, 1950; trans. P. King, *The Teacher* (along with *Against the Academicians*), Indianapolis, IN: Hackett, 1995.(An important attempt to consider whether the meanings of words can be learned by ostension.)

Augustine (389/91) *De vera religione (On True Religion)*, trans. J.H.S. Burleigh, *Augstine: Earlier Writings*, London: SCM Press, 1953, 218-83.(Augustine here finds truth even in doubting.)

Augustine (392/3) *De duabus animabus contra Manichaeos (On the Two Souls)*, trans. A.H. Newman in P. Schaff (ed.) *A Select Library of the Nicene and Post-Nicene Fathers of the Christian Church*, First Series, 1886-8, vol. 4, 91-124.(A spirited rejection of the Manichean idea that each person has both a good and an evil soul.)

Augustine (394) *De mendacio (On Lying)*, trans. M.S. Muldowney, Fathers of the Church vol. 14, Washington, DC: Catholic University of America, 1952, 45-120.(The earlier of two treatises devoted to an analysis of the sin of lying.)

Augustine (394) *De sermone Domini in monte (Commentary on the Lord's Sermon on the Mount)*, trans. J.J. Jepson, New York: Newman Press, 1948.(Perhaps the most important statement of Augustine's intentionalist ethics.)

Augustine (396-426) *De doctrina christiana (On Christian Instruction)*, trans. J.J. Gavigan, Washington, DC: Catholic University of America, 1947.(Augustine's handbook of biblical exegesis.)

Augustine (397-401) *Confessionum libri tredecim (Confessions)*, trans. F.J. Sheed, Indianapolis, IN: Hackett, 1993; trans. R.S. Pine-Coffin, Harmondsworth: Penguin, 1961; trans. H. Chadwick, Oxford: Oxford University Press, 1991.(This is no doubt the most widely and frequently translated of Augustine's works. The above are the English translations deserving of special mention.)

Augustine (399) *De natura boni contra manichaeos (The Nature of the Good)*, trans. A.H. Newman in W. Oates (ed.) *Basic Writings of Saint Augustine*, New York: Random House, 1948, vol. 1, 429-57.(An attempt to make plausible the idea that evil is a privation.)

Augustine (400) *Contra Faustum manichaeum (Reply to Faustus)*, trans. R. Stothert in P. Schaff (ed.) *A Select Library of the Nicene and Post-Nicene Fathers of the Christian Church*, First Series, 1886-8, vol. 4, 151-345. (A major polemic against the leading Manichee of Augustine's time.)

Augustine (400-16) *De trinitate (The Trinity)*, trans. S. McKenna, Washington, DC: Catholic University of America, 1962.(The last eight of this work's fifteen books constitute Augustine's most fully stated philosophy of mind.)

Augustine (401-14) *De genesi ad litteram (The Literal Meaning of Genesis)*, trans. J.H. Taylor, New York: Newman, 1982, 2 vols.(Augustine's commentary on Genesis includes important discussions of metaphysics and philosophical theology; the last treatise is a separate treatise on vision, including mystical vision.)

Augustine (412) *De spiritu et littera (On the Spirit and the Letter)*, trans. P. Holmes in W. Oates (ed.) *Basic Writings of Saint Augustine*, New York: Random House, 1948, vol. 1.(An attempt to argue that God's grace does not void free will, but rather grounds it.)

Augustine (413-27) *De civitate Dei (The City of God)*, trans. J. O'Meara, Harmondsworth: Penguin, 1972.(Several good English translations exist, of which the above is the most notable.)

Augustine (415-16) *De perfectione justicia hominis (On Man's Perfection in Righteousness)*, trans. P. Holmes and R.E. Wallis in P. Schaff (ed.) *A Select Library of the Nicene and Post-Nicene Fathers of the Christian Church*, First Series, 1886-8, vol. 5, 177-212.(A treatise on the accusations against Pelagius.)

Augustine (416) *In ioannis evangelium tractatus (Homilies on the Gospel of John)*, trans. J.W. Rettig, Fathers of the Church vols 78-9, 88, 92, Washington, DC: Catholic University of America, 1988-95.(Sermons on St John's Gospel arranged systematically so as to constitute a complete commentary on that gospel.)

Augustine (419-21) *De nuptiis et concupiscentia (On Marriage and Concupiscence)*, trans. P. Holmes and R.E. Wallis in P. Schaff (ed.) *A Select Library of the Nicene and Post-Nicene Fathers of the Christian Church*, First Series, 1886-8, vol. 5, 257-308.(A somewhat qualified defence of the good of marriage in response to the Pelagians.)

Augustine (421) *De cura pro mortuis (On the Care of the Dead)*, trans. J.A. Lacey, Fathers of the Church vol. 27, Washington, DC: Catholic University of America, 1955, 347-84.(A defence of burial services that raises interesting questions about ritual.)

Augustine (426/7) *Retractationes (Retractations)*, trans. M.I. Bogan, Fathers of the Church vol. 60, Washington, DC: Catholic University of America, 1968.(Augustine's efforts to review, in order of composition, all of his earlier works and, where he thinks it necessary, correct his earlier views.)

Augustine (429) *De praedestinatione sanctorum (On the Predestination of Saints)*, trans. R.E. Wallis in W. Oates (ed.) *Basic Writings of Saint Augustine*, New York: Random House, 1948, vol. 1, 775-817.(A final effort to insist that even the beginning of a believer's faith is God's predestined gift, though God does not act in human sin.)

References and further reading

Brown, P. (1967) *Augustine of Hippo*, Berkeley, CA: University of California.(The best general biography in English.)

Bubacz, B. (1981) *St. Augustine's Theory of Knowledge: A Contemporary Analysis*, New York: Mellen.(A good review of central passages on perception, memory and illumination.)

Chadwick, H. (1986) *Augustine*, New York: Oxford University Press.(A first introduction to Augustine.)

Christopher, P. (1994) *The Ethics of War and Peace*, Englewood Cliffs, NJ: Prentice Hall.(Chapter 3, 'Saint Augustine and the Tradition of Just War', describes Augustine's theory.)

Dihle, A. (1982) *The Theory of Will in Classical Antiquity*, Berkeley, CA: University of California.(One chapter is devoted to Augustine's theory.)

Gilson, E. (1960) *The Christian Philosophy of Saint Augustine*, New York: Random House.(A learned Thomist's best attempt to understand Augustine's philosophy.)

Holmes, R.L. (1989) *On War and Morality*, Princeton, NJ: Princeton University Press. (Judiciously places Augustine in the just war tradition.)

Kirwan, C. (1989) *Augustine*, London: Routledge.(The best treatment of Augustine's philosophy from a contemporary analytic point of view.)

Malcolm, N. (1963) 'Knowledge of Other Minds', in N. Malcolm, *Knowledge and Certainty*, Englewood Cliffs, NJ: Prentice Hall, 130-40.(A classic Wittgensteinian critique of the appropriateness and efficacy of the argument from analogy for the existence of other minds.)

Markus, R.A. (1967) 'Marius Victorinus and Augustine', in A.H. Armstrong (ed.) *The Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press, 331-419.(A learned by accessible summary of Augustine's philosophical views.)

Matthews, G.B. (1992) *Thought's Ego in Augustine and Descartes*, Ithaca, NY: Cornell University Press.(A discussion of the *cogito*, various dream problems, the problem of other minds and other issues that arise from

the attempt to pursue philosophy from a first-person point of view.)

O'Daly, G. (1987) *Augustine's Philosophy of Mind*, London: Duckworth.(A detailed treatment of Augustine's views on sense perception, imagination, memory and knowledge.)

Portalie, E. (1960) *A Guide to the Thought of Saint Augustine*, Chicago, IL: Regnery.(Still valuable as a guide to the interrelations between Augustine's various views in philosophical theology and theological ethics.)

Russell, B. (1945) *A History of Western Philosophy*, New York: Simon & Schuster.(A bold and always stimulating attempt to state succinctly and assess the contribution of each of the major Western philosophers.)

Sorabji, R. (1983) *Time, Creation and the Continuum*, London: Duckworth.(Puts Augustine's views on time, creation and mysticism into the context of ancient and early medieval thought.)

Wetzel, J. (1992) *Augustine and the Limits of Virtue*, Cambridge: Cambridge University Press. (Focuses on the moral relevance of Augustine's concept of will, and also on the idea that virtue is invulnerable to misfortune.)

Wittgenstein, L. (1958a) *The Blue and Brown Books*, Oxford: Blackwell.(Early formulations of views characteristic of Wittgenstein's later period.)

Wittgenstein, L. (1958b) *The Philosophical Investigations*, trans. G.E.M. Anscombe, Oxford: Blackwell.(The most important later work of Wittgenstein.)

Augustinianism

The influence of Augustine on Western philosophy is exceeded in duration, extent and variety only by that of Plato and Aristotle. Augustine was an authority not just for the early Middle Ages, when he was often the lone authority, but well into modern times. He was in many ways the principal author in contention during the Reformation and Counter-Reformation, and in France alone he was variously received by authors as diverse as Montaigne, Descartes, Malebranche, Arnauld and Pascal. The breadth of Augustine's influence makes it difficult to give precise sense to the term 'Augustinianism', even when considering only a single period.

Historians of medieval philosophy use the term 'Augustinianism' to describe three rather different relations to the thought of Augustine. The first relation is a comprehensive dependence on Augustine both for philosophical principles or arguments, and for instruction in the topics and procedures of ancient philosophy. Augustine serves as the trustworthy guide to philosophy as a whole. The second kind of relation is a defence of specific Augustinian teachings in the face of rival teachings, most especially those of Aristotle. These Augustinian teachings include the function of divine ideas in knowledge, the unity of the human soul's essential powers, and the unfolding of potential intelligibilities in material substances. The third relation is the reappropriation of Augustinian principles, especially those of his later writings, to address quandaries newly formulated with the tools of nominalist semantics and the mathematics of continuities. Among these quandaries are the contingency of future human actions and the certainty of human cognition.

These three relations to Augustine can be found in texts throughout the medieval period. They are not neatly correlated with particular centuries, but one or another does tend to be predominant at different times. Thus the first relation, of comprehensive dependence, is seen in the great majority of Latin writers on philosophic topics through the twelfth century. The second relation, of topical defence, appears prominently during the thirteenth-century contest between so-called 'Augustianians' and 'Aristotelians'. The third relation, of reappropriation in reaction to newly formulated quandaries, is found particularly in writings of the fourteenth century and beyond.

1 Scope

At face value, 'Augustinianism' would seem to refer generally to the teaching of Augustine or to one of the many adaptations of it. To use the term in this sense is to rob it of usefulness. The influence of Augustine on Western philosophy is exceeded in duration, extent and variety only by that of Plato and Aristotle. Augustine's writings served as authoritative works in the early Middle Ages, when he was often the sole philosophical authority referred to; the influence of Augustine is so pervasive in Latin Christian thought that any medieval Latin author on philosophic topics will show some borrowings from Augustine, and most of the best-known authors can be seen to engage him in sustained ways. Augustine was traditionally counted one of four 'Fathers' or 'Teachers' of the Latin-speaking Church, along with Jerome, Ambrose and Gregory the Great (see Patristic philosophy). His influence exceeded that of the other three in most theological areas and in all philosophical ones except for ethics, where Gregory's *Moral Readings on Job* provided more detailed if less systematic teaching about virtues. Further, the influence of Augustine can be seen well beyond the medieval period. During the Reformation and Counter-Reformation in the sixteenth century he was one of the principal authors referred to by both contending parties, and he was variously received by modern philosophic authors as diverse as Montaigne, Descartes, Malebranche, Arnauld and Pascal.

The term 'Augustinianism' can be made useful only by being applied to specific relations to Augustine. Three rather different relations have been used by historians of medieval philosophy. The first is a *comprehensive dependence* on Augustine, both for philosophical principles or arguments and for instruction in the topics and procedures of ancient philosophy. Second, there was a *defence* of specific Augustinian teachings in the face of rival teachings, in particular those of Aristotle (see Aristotelianism, medieval). Third, there was a *reappropriation* of Augustinian principles, especially those from Augustine's later works, in order to address quandaries newly formulated in the languages of nominalist semantics and the mathematics of continuities (see Nominalism; Oxford Calculators).

All three relations can seem to be progressively narrower or more restricted; indeed, it is tempting to understand

them as a sequence of comprehensive Augustianism, disputed Augustinianism and adapted or revivified Augustinianism. This conception of a sequence would be false in two important ways. First, it would ignore Augustine's permanent and unsurpassed authority in Christian theology, which provides so much of the context for medieval philosophy and, indeed, for modern European philosophy. Second, it would mislead one into thinking that philosophical interest in Augustine was at a low ebb at the end of the Middle Ages, whereas in fact Renaissance and early modern philosophy shows a striking resurgence of interest in Augustine. Any narrowing in the application of the term 'Augustinianism' should be imputed as much to the optical illusions of historiography as to the facts of reception. Contemporary historians of philosophy know more of philosophical writing before the thirteenth century than of what was written during or after that period, in large measure because there is so much more of the latter to know. Hence any narrative of medieval thought tends to become much more selective as it proceeds.

From the very beginning, Augustine's work was not easy to appropriate for philosophical uses. His body of writings is enormous and topical, not to mention contradictory in some of its details. Although there were early attempts to make his work both more accessible and more systematic, the interpretations of Augustine's writings depended very much on which works were being read. There was not always a choice in what to read. The reception of any author in the earlier Middle Ages depended on the hard facts of manuscript survival and distribution. In the cultural disarray that destroyed books and the means of producing them, Augustine was relatively fortunate; his works were copied steadily and widely. But to say that Augustine was frequently copied is not to say that he was uniformly copied. Not all parts of the Augustinian corpus were equally well known, and even in the thirteenth and fourteenth centuries certain Augustinian texts were much more rare than others. Thomas Aquinas revised a number of his opinions in mid-career, after coming into possession of texts of Augustine that he had not known before. The period 1325-35 in England also saw a sharp increase in commentaries on Augustine's knottiest writings, especially those of his later years.

Beyond problems of access there was also, and from early on, the problem of spurious works. Given Augustine's immense authority and his reputation for universal genius, it is not surprising that pious readers would attribute to him any number of works they admired. To extract the authentic works from this mass has been the labour of many centuries, a labour still unfinished with regard to some sorts of works, such as the homilies. These variations and uncertainties in the availability of Augustine's texts should be kept in mind through all that follows.

2 The reception of Augustine through the twelfth century

The first sense of 'Augustinianism' distinguished above refers to comprehensive influence exerted by Augustine on philosophic writers from his own time through the twelfth century. The most popular of his works not only set the terms and starting points for addressing many theological questions that were to vex early medieval authors, they also passed down precious information about ancient pagan philosophy and provided much-needed introductions to its teachings. Even after the conditions of learning had improved significantly in western Europe, Augustine continued to be followed as a sure guide through the confusing thicket of ancient philosophical debates.

Three examples can show the variety of relations to Augustine. The first is John Scotus Eriugena, who is best known for his technical command of late Greek Neoplatonism (see Neoplatonism). Some find in his work a turning away from the eclectic philosophy of Augustine, but it is just as possible to read Eriugena's *Periphyseon* as the incorporation of newer Neoplatonic elements within a framework that is deliberately and thoroughly Augustinian. Certainly Eriugena knew Augustine's work well, and just as certainly he engages the latter even when he disagrees with him.

The second, more obvious example is that of Anselm of Canterbury. In the prologue of his *Monologion*, Anselm explicitly submits his works to the test of conformity with Augustine. He advises readers to take Augustine's *De trinitate (On the Trinity)* as the measure of his own 'little work'. There is no irony or duplicity here. If Anselm goes on in his treatises to argue by reasoning alone, without frequent appeal to authorities, his arguments are still just so many variations on Augustinian themes. The famous definition of God used in the *Proslogion* to begin what came to be called the 'ontological argument' is Augustinian, and the argument's procedure follows closely several of Augustine's hierarchical arguments for the existence of God. Again, Anselm's analysis of truth as 'rightness' (*rectitudo*), however much it depends on more recent logical techniques such as the substitution of equipollent propositions, remains in its premises an Augustinian teaching about the communicative intention in language.

A final example of the comprehensive reception of Augustine in philosophy can be found in the work of Hugh of St Victor. Hugh is principally and formally a writer of theology, and he understands highest 'philosophy' to be the monk's prayerful contemplation of God. Hugh's works also rehearse a number of doctrines that are philosophic in a less exalted sense. These doctrines are drawn from many sources, but they are given shape after Augustinian patterns. In the survey of philosophical topics in the *Didascalicon,* Hugh quotes from a number of authors, including Boethius and his commentators, Cicero, Macrobius (see Encyclopedists §3), Calcidius and Augustine. The judgments on the right ordering of philosophy, as on particular philosophic doctrines, are taken for the most part from Augustine. If Augustine is invoked at times to supply some bit of learning about ancient philosophy, he is relied upon throughout in subordinating philosophy to Christian theology.

These two examples only begin to suggest the various relations between Augustine and later writers on philosophy. It should also be remembered that in the twelfth century, Augustine was the chief authority on theology. His works are given pride of place in Peter Lombard's influential theological codification, the *Sententiae in IV libris distinctae (Four Books of Sentences)*, and many of the topics under debate there are ones that Augustine would have recognized. Roughly the same is true in philosophy, despite the influx of newly translated Aristotelian works. A century later, however, Augustine's work was being increasingly challenged (see Lombard, P.).

3 The thirteenth century

The second sense of 'Augustinianism' distinguished above refers to the defence of particular Augustinian tenets against increasingly attractive Aristotelian alternatives. Indeed, some historians of medieval philosophy describe the principal philosophical developments from the middle of the thirteenth century onwards as a protracted contest between 'Augustinians' and 'Aristotelians'. They find evidence of the contest in a number of struggles within the universities of Paris and Oxford, and they see an important triumph for the Augustinian party in the condemnations pronounced at Paris and Oxford in 1270 and 1277. The 'Aristotelians', in this account, were not just the teachers of extreme positions in the Faculty of Arts but also, perhaps especially, Dominican theologians such as Albert the Great and Thomas Aquinas who approved some Aristotelian doctrines in philosophy and appropriated them integrally for theological use. By contrast, the 'Augustinians' were not just conservative masters of theology, but theologians who found the growing use of Aristotle damaging both to faith and to philosophy. Among the 'Augustinians' are usually accounted the Franciscans Alexander of Hales, Bonaventure and John Pecham. Because these controversies seem to have divided Dominicans from Franciscans, the quarrel of 'Aristotelians' and 'Augustinians' is often made into a quarrel between the two religious orders.

This way of describing the controversies makes caricatures of them. Augustine and Aristotle were in fact studied, quoted and defended by controversialists on both sides. Some of the best known 'Augustinians' spent considerable time explicating the text of Aristotle precisely in order to resist attempts to Christianize him, offering instead exacting interpretations of his original sense. More generally, Aristotelian terminology was used as the lingua franca by all members of the universities. On the other side, some of the leading 'Aristotelians' argued that the Augustine of their opponents was not the authentic Augustine at all, but a hyper-Platonist drawn largely from Islamic sources. As a result, great care should be taken when using the terms 'Augustianian' and 'Aristotelian' as designations for groups of medieval authors.

What is clear is that some of the Franciscan controversialists described themselves as defenders of Augustinian tenets in philosophy. They were quite willing to specify at least some of tenets they meant. In a letter of 1285, John Pecham wrote that he was not attacking the 'study of philosophers' so much as the 'profane novelties' of the 'Aristotelians'. These novelties are opposed to 'whatever Augustine teaches on the eternal rules and the unchangeable [intellectual] light, the powers of the soul, on the seed-like intelligibilities inserted into matter, and on innumerable other things like these' (Pecham, Letter of 1 January 1285). The three doctrines that Pecham mentions were indeed at the centre of the thirteenth-century disputes; they are known to historians of philosophy as doctrines of divine illumination, psychic essentialism and seminal reasons.

Augustine was understood to have taught that all knowledge worthy of the name was made possible only by the 'illumination', that is, by the intelligible presence of the 'divine rules' or 'divine reasons', God's own patterns for making creatures (see Augustine §10). On most readings, this did not mean that a human mind in the present life

could see the divine essence or could know something just by relying on divine illumination. Bonaventure, for example, is explicit that the presence of the divine ideas is ordinarily a necessary but not sufficient cause for cognition. By contrast, Thomas Aquinas argues that no ‘special’ illumination is required; the mind’s ‘participation’ in the divine light given at its creation is sufficient for ordinary cognition.

The second doctrine singled out by Pecham concerns the relation of a soul to its powers, such as reason and will. Some authors, claiming to follow Augustine, held that the soul’s powers were nothing more than its relations to various acts. Hence there could be no essential distinctions among the powers, which were all referred to the soul’s single essence. Other authors, following Aristotle, held that the soul’s powers inhered in the soul as accidents in a substance. In this view, the powers were essentially distinct. A third group attempted to split the difference; they held that the soul’s powers were only derivatively or analogically located in the category of substance. They should be considered as really different, not as essences are different, but as powers of a single essence can still be essentially different. Proponents of this third view, such as Bonaventure, thought that this was Augustine’s most careful teaching on the matter. Bonaventure also remarks that the debate is more curious than useful; none of the positions, he says, is opposed to faith or good morals.

The third disputed doctrine, that of seminal reasons, is adapted by Augustine from Stoic physics (see Stoicism §3). Knowing little of the genealogy, medieval readers held this doctrine as an affirmation of the completeness of divine creation and as an alternative to Aristotelian accounts of the pure potency of matter. They understood Augustine to teach that God had inserted into matter, at the moment of creation, intelligible patterns that could be actualized over time. So, for example, Bonaventure holds that the souls of non-rational animals and of plants were created not out of nothing nor simply out of some pre-existing matter, but rather ‘in the manner of a seed’. They were created, in other words, by actualizing an active potency in matter, which serves as a ‘seed bed’ of such potencies. After the moment of creation, animal souls are reproduced without divine intervention by the natural actualization of such ‘seed-like reasons’. These ‘reasons’ also contain the latent forms of certain lower-order substances which are produced by the corruption or commingling of other specific forms. In this way, every kind of form that would sometime attain actuality was put into the created world at the beginning.

Pecham’s letter also alludes to other doctrines disputed between the followers of Augustine and their adversaries. Principal among these were two, the plurality of forms and the universality of the form-matter distinction. The teaching on the plurality of forms, to which Pecham was particularly attached, held that multiple forms could exist in a single substance without having to be reduced to accidents of a single substantial form. This teaching was particularly important in analysing the relation of human souls and bodies. Pecham thought it nonsensical to say that the separable, immortal soul was the single substantial form of the body-soul composite. As to the other point, defenders of Augustine taught that something properly called a form-matter distinction could be found in all creatures, including the incorporeal substances known to theology as angels. If this ‘spiritual matter’ was not identical with ordinary matter, it was nonetheless a material principle.

Pecham’s version of Augustine can be found in other writers, if not so emphatically articulated. His Augustinian programme is in some ways more retrospective than prospective. Already by Pecham’s time, Augustine was being appropriated in more subtle and novel ways by a new group of authors, most notably Henry of Ghent. From Henry through John Duns Scotus and William of Ockham, one can trace profoundly original transformations of Augustinian principles and problems. Indeed, Scotus and Ockham have been counted, in their very different ways, as a second ‘school’ of Franciscan Augustinians. Here, however, the term ‘Augustinianism’ begins to become much too diffuse to be useful.

A more specific ‘Augustinianism’ is sometimes claimed for the writers of the religious order of Augustinian friars. The order was organized out of existing communities between 1244 and 1256, and its constitutions were ratified in 1290. Some historians have wanted to see in this order an institutional home for an ‘Augustinian’ school of philosophy and theology, but the matter is not so simple. Giles of Rome, first Parisian master from the order and later its Prior General, was a student of Thomas Aquinas. He often appears in narrative histories as an ardent if somewhat inexact Thomist. In fact, he differs from Thomas on a number of points, such as the nature of the essence/existence distinction and the plurality of substantial forms. For these reasons among others, Giles was claimed by later members of the Augustinian order as founder of a ‘school’ of thought (the *schola Aegidiana*) which those members understood themselves to be continuing. But Giles was hardly a programmatic Augustinian

in philosophy, and the most interesting philosophical adaptations of Augustine at the end of the thirteenth century were happening outside the Augustinian order.

4 The fourteenth century

The third relation to Augustine distinguished above is the most difficult to characterize. It combines motives of textual fidelity with motives of reappropriation. Textual fidelity led to an intensive re-reading of Augustine, culminating in Oxford during the second quarter of the fourteenth century. Reappropriation stemmed from the desire to respond to the sea-change in philosophy marked by Scotus and Ockham. The most prominent fourteenth-century Augustinians wanted to understand and to rehabilitate Augustine's central teachings in response to the quandaries produced by the 'modern way' in philosophy; some of these writers, notably Thomas Bradwardine, are distinguished by their willingness to restate difficult Augustinian principles against accepted conclusions of the latest semantics and the mathematical analysis of nature. Other writers are marked out by their fidelity to Augustine across a range of topics and in the face of the increasingly popular alternative ways of thinking; the most famous of these is Gregory of Rimini.

Bradwardine was prominent among the Oxford 'Calculators', a group of Masters of Arts centred in Merton College, Oxford, during the 1320s and 1330s (see Oxford Calculators). He wrote important works on logical quandaries and on proportions. While at Oxford, Bradwardine also wrote what is counted a centrepiece of fourteenth-century Augustinianism, *Summa de causa Dei contra Pelagium et de virtute causarum (On God's Cause against Pelagius and on the Power of Causes)*. The work is, as its title proclaims, a polemical one, written against certain 'new Pelagians' (see Pelagianism). Bradwardine seems to refer by this epithet to some of his Oxford contemporaries, perhaps especially Richard Fitzralph, Robert Holcot, Adam Wodeham and Thomas Buckingham. He accuses them of denying all of God's providence, cooperation, conservation and predestination with respect to creaturely activity. Denying these things, Bradwardine's Pelagians effectively deny the existence of God, because they deny that God is God. Against them, Bradwardine offers 'the grace of God, which precedes all good merits with a priority of both time and nature; I mean the gracious will of God, prior in both these ways, which wills to save the one who merits it, and naturally produces merit in him even before he does, just as He is the first mover in all motions.' This Augustinian doctrine came to Bradwardine in a moment of intellectual conversion, which he explicitly links to Augustine's narrative in the *Confessiones*.

Though Bradwardine confronts the Pelagians on a number of fronts, his sharpest and most technical attack comes in considering the problem of future contingents. This problem had been much debated in Oxford for several decades. Bradwardine considered it the crucial case for the Pelagians - and the one that revealed the depth of their error. According to Bradwardine, the new Pelagians held that while singular propositions about future contingents always possessed a determinate truth, the contingents were themselves not determined. A truth about them could become false. By contrast, the truth of propositions about present events was fixed, and the truth of propositions about past events was necessary. It might seem a long way from Bradwardine's Augustinian conversion to these logical issues, but the way is in fact quite short. Views on the truth of future contingents had immediate consequences for the question of divine foreknowledge and, indeed, on such specific topics as the truth of prophetic revelations about future events. For Bradwardine, the Pelagians' doctrines concerning the mutability of future contingents denied both the immutability of divine knowledge and the honesty of divine revelation. If the truth of propositions about the future is different from the truth of propositions about the past, then the character of God's knowledge must change with time; and if the truth of a specific proposition about a future event is contingent, then a prophecy of that event can be falsified. Bradwardine finds both conclusions monstrous. The first makes God into a creature; the second, into a deceiver. Bradwardine teaches, on the contrary, that all future events are necessitated in respect to higher causes, though some of them may be understood to be free in present time with respect to lower causes. This position is, on Bradwardine's reading, precisely that of Augustine.

With Gregory of Rimini, one may begin to speak of an Augustinian school within the order of Augustinian friars. He was elected Prior General of the order in 1357 after a distinguished teaching career in Paris and Italy. Gregory's appropriation of Augustine is wider and more nuanced than Bradwardine's. A sample can be had in Gregory's discrimination among the types of human knowledge. He agrees with Ockham in criticizing the pretensions of empirical knowledge, and he is equally firm in advancing the claims of introspection, which he describes by citation from Augustine's *De trinitate*. More strikingly, Gregory claims that intelligible knowledge

depends upon certain innate ideas that direct the soul to its cognitive and moral end. His explicit source here is Augustine's *De libero arbitrio (On Free Choice of the Will)*. In these and other instances, Gregory aligns problems or models from the most recent philosophy with principles retrieved from Augustine.

It is difficult to assess the immediate influence of Thomas Bradwardine or Gregory of Rimini without deciding how far they contributed to the heterodox reform movements that would eventually lead into the Reformation. A case in point is John Wyclif, who invokes Bradwardine as one of his intellectual masters. Indeed, the study of Augustinianism shows with particular clarity how arbitrary is the division between late medieval and early modern philosophy. The culmination of the medieval relations to Augustine's philosophical texts is not to be found in the Middle Ages; the conclusion comes in the sixteenth and seventeenth centuries, in the Reformation and in the genesis of several of the most important philosophical movements of early modernity, including Cartesianism and the philosophy of the Jansenists.

See also: Aristotelianism, medieval; Augustine; Bonaventure; Bradwardine, T.; Gregory of Rimini; Henry of Ghent; Medieval philosophy; Pecham, J.; Religion, history of philosophy of

MARK D. JORDAN

References and further reading

Courtenay, W.J. (1980) 'Augustinianism at Oxford in the Fourteenth Century', *Augustiniana* 30: 58-70.(A well-documented introduction to the reappropriation of Augustine.)

Crouse, R.D. (1987) 'Anselm of Canterbury and Medieval Augustinianisms', *Toronto Journal of Theology* 3 (1): 60-8.(A brief but very helpful introduction to the varieties of Augustinianism through the twelfth century; richly annotated.)

Genest, J.-F. (1992) *Prédetermination et liberé créé à Oxford au XIVe siècle: Buckingham contre Bradwardine (Predetermination and Free Will at Oxford in the Fourteenth Century: Buckingham versus Bradwardine)*, Paris: Vrin.(Detailed introduction to the debates over contingency; contains an edition of Buckingham's *Determinatio de contingentia futurorum (Determination Concerning the Contingency of Future Things)*.)

King, E.B., and Schaefer, J.T. (1988) *Saint Augustine and His Influence in the Middle Ages*, Sewanee, TN: Press of the University of the South.(Anthology of conference papers; the contributions by Chadwick and Evans are particularly pertinent.)

Marrou, H.I. (1959) *Saint Augustin et l'Augustinisme*, Paris: Éditions du Seuil.(There is also an English translation from an earlier edition by P. Hepburne Scott and E. Hill, *St. Augustine and His Influence through the Ages*, New York: Harper Torchbooks and London: Longmans, 1957. General and now somewhat dated.)

Oberman, H.A., and James III, F.A. (eds) (1991) *Via Augustini: Augustine in the Later Middle Ages, Renaissance, and Reformation: Essays in Honor of Damasus Trapp*, Leiden/New York: Brill.(Anthology of rather narrowly focused essays.)

O'Donnell, J.J. (1991) 'The Authority of Augustine', *Augustinian Studies* 22: 7-35.(Reflections on Augustine's influence and authority from the fifth century onwards.)

Pecham, John (1285) Letter of 1 January 1285, number 625 in C.T. Martin (ed.) *Registrum epistolarum Fratris Johannis Peckham, archiepiscopi Cantuariensis*, Rolls Series 77, London, 1885, vol. 3, 870-72.(Pecham's description of doctrines disputed between followers of Augustine and Aristotle.)

Wippel, J.F. (1977) 'The Condemnations of 1270 and 1277 at Paris', *Journal of Medieval and Renaissance Studies* 7: 169-201.(Richly annotated introduction to the motives and contexts of the two sets of condemnations.)

Aureol, Peter (*c.*1280-1322)

A master of theology at the University of Paris and a member of the Franciscan order, Peter Aureol helped shape the philosophical agenda of the fourteenth century. His original and provocative views were widely discussed during the later Middle Ages, but his influence was rather indirect since his views almost always met with hostility. Although Aureol wrote extensively on a wide range of philosophical and theological issues, his most-discussed contributions to philosophy, in epistemology and metaphysics, centre on his theory of esse apparens *(apparent existence).*

Aureol, known as the *Doctor facundus* (Fluent Doctor), was born near Gourdon in the south of France. He seems to have joined the Franciscan order in his teens, and to have studied at Paris, perhaps in 1304. There has been speculation that he was a student of John Duns Scotus at that time, but there is little evidence for the claim. Aureol's magnum opus, his *In quatuor libros Sententiarum (Commentary on Peter Lombard's Sentences)*, began taking shape in his lectures at Franciscan houses of study in Bologna (beginning in 1312) and Toulouse (in 1314). In 1316 he was sent to Paris to continue his lectures, and he remained there until 1320, becoming regent master in 1318, lecturing on the Bible, and completing a single quodlibetal question. He became an archbishop in 1321, and died shortly thereafter.

Aureol thoroughly edited the first book of his *Sentences* commentary, and this so-called *Scriptum*, together with the other parts of his commentary (the less polished *Reportationes*), constitute Aureol's most significant philosophical work. In it, he ranges widely over the theological and philosophical problems of his day, making interesting and original suggestions about the relationship between soul and body, the nature of grace and justification, and the status of future contingents and divine foreknowledge, among many other topics. However, the issues that have drawn most attention are those that surround his theory of *esse apparens* (apparent existence), a kind of non-real, merely apparent existence that is the key to Aureol's theories of universals, perception and conceptual thought.

Aureol first alludes to this *esse apparens* when he takes up Duns Scotus' distinction between intuitive and abstractive cognition (*Scriptum* pro. s.2). This distinction applies at both the intellectual and sensory levels, and is most easily described by examples: imagination (in the absence of the imagined object) counts as abstractive cognition, for instance, whereas ordinary sense perception is one kind of intuitive cognition. But although the extension of these categories was relatively uncontroversial, it proved much more difficult to give a general account of what distinguishes intuitive from abstractive cognition. Duns Scotus claimed that intuitive cognition takes as its object something that is present to the cognitive agent. Aureol objects to this claim by describing various cases of non-veridical perception, such as after-images and double-vision, where there is no object corresponding to what is being perceived. These should all be counted as instances of intuitive cognition, Aureol argues, and he goes on to claim that this sort of empirical evidence should be taken more seriously than any logical argument that might be put forward. In opposition to Scotus' proposal, Aureol suggests that intuitive cognition, at both the sensory and intellectual levels, be defined as cognition that is direct (not inferential), and that makes the object appear present, actual and existent.

These claims naturally lead Aureol to raise more general questions about perception, and in particular to ask how perception is able to make an object appear present, whether or not it is actually present. His answer is that the senses give the object apparent existence. (He also speaks of this as *intentional* existence.) When a stick is swung rapidly in a circle, something circular appears. There is nothing circular that really exists, because the appearance is there only for as long as someone is watching. However, Aureol insists that the appearance, this apparent entity, exists in the very place where it appears to exist. It is not in the eye or the brain, nor anywhere in the intervening air. (Aureol does incorporate sensible and intelligible *species* into his account, but he attributes real existence to them, not apparent or intentional existence, and he takes them to be pure intermediaries which are never themselves the objects of cognition.) Aureol uses these illusory cases to establish a general theory of perception: all perception involves the formation of *esse apparens*, but it is illusions that point to the true nature of perception, because ordinarily we can't distinguish *esse apparens* from *esse reale*: 'In the case of true vision they occur together' (*Scriptum* d.3 s.14 a.1).

Aureol gives a similar account at the intellectual level. Indeed, he explicitly defines cognition, sensory and intellectual, as 'having something present as an appearance'. Thus if a picture of Caesar on a wall were to make Caesar appear to that wall, then 'the wall would be said to be cognizing Caesar' (*Scriptum* d.35 a.1). At the intellectual level, an apparent entity exists within the intellect, and is in fact the focus of intellect's attention (*Scriptum* d.9 a.1; d.27 a.2). Aureol thinks that he can make this claim without epistemological difficulties, for he stresses in many places that it is the object itself that has apparent existence within intellect. One's concept of a rose is not distinct from the thing itself in the external world. It just is that rose, in *esse apparens*: thus 'the objects are cognized directly' (*Scriptum* d.3 s.14 a.3).

These considerations form the basis for Aureol's conceptualist theory of universals (*II Reportationes* d.3 q.2; d.9 q.2). Nothing outside mind is universal; mental concepts are universal only in the sense that some concepts have a less determinate content than others. Indeed for Aureol, 'everything, inasmuch as it is, is singular'. Individuation is a primitive feature of reality, and so 'to seek that through which a thing outside the intellect is singular, is to seek nothing' (*II Reportationes* d.9 q.2). However, Aureol's position seems readily distinguishable from nominalism because he rejects the view that mental concepts might merely *denote* objects in a certain indefinite and universal way (see Nominalism §§1-2). Mental concepts, as universals, are more than mere representations, because they *are* the objects, in *esse apparens*.

See also: Chatton, W.; Duns Scotus, J.; Perception

ROBERT PASNAU

List of works

There are no published translations of Aureol's work.

Aureol, Peter (1312-20) *In quatuor libros Sententiarum (Commentary on Peter Lombard's Sentences)*, Rome, 1596-1605.(This early edition is not widely available, but is currently the only printed source for most of the *Sentences* commentary, as well as for his *Quodlibet*.)

Aureol, Peter (1317) *Scriptum super primum Sententiarum*, ed. E.M. Buytaert, St Bonaventure, NY: Franciscan Institute, 1956, 2 vols.(Covers only the first eight of 48 distinctions.)

References and further reading

Boehner, P. (1948) '*Notitia intuitiva* of non-existents according to Peter Aureoli O.F.M. (1322)', *Franciscan Studies* 8: 388-416.(A good introduction to Aureol's disagreement with Scotus over intuitive cognition.)

Dreiling, R. (1913) *Der Konzeptualismus in der Universalienlehre des Franziskanererzbischofs Petrus Aureoli (Conceptualism in the Theory of Univerals of the Franciscan Archbiship Peter Aureol)*, Beiträge zur Geschichte der Philosophie des Mittelalters 11, 6, Münster: Aschendorff.(The best source for information on Aureol's theory of universals, although Dreiling judges Aureol quite harshly.)

Henninger, M. (1989) *Relations: Medieval Theories 1250-1325*, Oxford: Clarendon Press.(Devotes a chapter to Aureol's interesting claim that relations have no extramental existence.)

Normore, C. (1993) 'Petrus Aureoli and His Contemporaries on Future Contingents and Excluded Middle', *Synthese* 96: 83-92.(Discusses Aureol's claim that some future contingent propositions are neither true nor false.)

Perler, D. (1994a) 'What Am I Thinking About? John Duns Scotus and Peter Aureol on Intentional Objects', *Vivarium* 32: 72-89.(Compares the views of Aureol and Scotus.)

Perler, D. (1994b) 'Peter Aureol vs. Hervaeus Natalis on Intentionality', *Archives d'Histoire Doctrinale et Littéraire du Moyen Age* 61: 227-62.(Edits and supplies a useful introduction to part of *Scriptum* d.23, concerning so-called 'second intentions'.)

Tachau, K. (1988) *Vision and Certitude in the Age of Ockham: Optics, Epistemology and the Foundations of Semantics, 1250-1345*, Leiden: Brill.(Contains a detailed discussion of Aureol's theory of *esse apparens*.)

Teetaert, A. (1935) 'Pierre Auriol ou Oriol', *Dictionnaire de Théologie catholique* XII (2), Paris, 1810-81.(This book-length article provides the most complete treatment of Aureol's life and work.)

Vignaux, P. (1934) *Justification et prédestination au XIVe siècle: Duns Scot, Pierre d'Auriole, Guillaume d'Occam, Grégoire de Rimini (Justification and Predestination in the Fourteenth Century: Duns Scotus, Peter Aureol, William of Ockham, Gregory of Rimini)*, Paris: E. Leroux.(The most detailed study of Aureol's thought in these areas.)

Aurobindo Ghose (1872-1950)

Aurobindo Ghose was a leading Indian nationalist at the beginning of the twentieth century who became a yogin and spiritual leader as well as a prolific writer (in English) on mysticism, crafting a mystic philosophy of Brahman (the Absolute or God). Aurobindo fashioned an entire worldview, a system intended to reflect both science and religion and to integrate several concerns of philosophy - epistemology, ontology, psychology, ethics - into a single vision. Of particular importance to his cosmological thinking was evolutionary biology. But Aurobindo also understood the fundamental nature of matter to include - for metaphysical reasons - an 'evolutionary nisus' that ensures the emergence of individuals capable of mystical experience in which the supreme reality, Brahman, is revealed.

1 Life

Commonly referred to as Aurobindo, or, by followers, as Sri Aurobindo ('Sri' is honorific), Aurobindo Ghose was born in Calcutta. He spent fourteen years in England from the age of seven until graduating from King's College, Cambridge University. He died surrounded by disciples in the French colony of Pondicherry in South India.

Aurobindo's politics were not of the passivist variety later endorsed by Gandhi. He used the editorial columns of the nationalist newspaper *Bande Mataram (Hail to Mother India)* to call rather openly for open rebellion. Arrested on charges of sedition and then 'waging war', he spent a year in prison before being acquitted in 1909. Still harassed by the British authorities, he retreated to Pondicherry, retired from politics and wrote voluminously while practising meditation and yoga. In addition to his metaphysical writings, Aurobindo wrote an epic poem (*Savitri*, one of the longest in English) and works on psychology, political philosophy, ethics, culture and yoga.

2 Metaphysics

Aurobindo puts forth in *The Life Divine*, his principal work of philosophy, a thoroughly realist metaphysics: there is an essential God or Brahman (the two terms are with Aurobindo roughly interchangeable) and real physical objects - transcendent in the sense of existing independently of our experience - with much intermediate to these two 'poles of being'. He couples his realism with an empiricist epistemology that brooks no illusory element. Illusions and misperceptions are analysed as false combinations of presentations of realities. Objects stand in causal relations to perceivers, though causal media vary. Aurobindo defends an epistemic parallelism between the indications of sensory and mystical experiences, with the latter carrying most of the weight regarding the existence and nature of Brahman - although variations on teleological, cosmological and other arguments of rational theology are presented too. Aurobindo's realism is not unreflective; he considers and rejects several varieties of phenomenalism, including Buddhist and Vedāntic theories that use a phenomenalist understanding of sensory objects to defend the possibility of enlightenment experiences held to be a supreme personal good. Aurobindo endorses such possibilities of enlightenment (*nirvāṇa*, for example), but proposes that his realist views provide better explanations both of mystical experiences and of our everyday world. There is tension, in Aurobindo's metaphysical writings, between a mystical foundationalism, Aurobindo's sense of the veridicality of yogic experiences, and a speculative theology, a sense that Brahman as the Absolute cannot be perfectly known.

Throughout *The Life Divine*, Aurobindo weaves a view of a 'self-manifesting Brahman' with close attention to an overall coherence of theory. Brahman - in essence perfect Being, Consciousness-Will and Bliss or Value (so Aurobindo renders a traditional characterization of Brahman as *saccidānanda*) - involutes aspects of itself - that is, contracts - so that certain finite possibilities can emerge, a process that has an outer limit in the 'inconscient' energies of matter. But such apparent inconscience cannot remain because it is nothing but Brahman. Conceivable universes that are incompatible with Brahman's essential nature are strictly impossible. Thus conscious beings are destined to evolve. In other words, God (that is, Brahman) works within limits, and could not, for example, make 2 + 2 = 5. God cannot create an entirely insentient world since God is constrained by the metaphysical law *ex nihilo nihil fit* ('nothing from nothing') to create out of God's own nature of Consciousness and Bliss. Thus this world is destined to evolve sentient material beings and eventually a divine life conceived as a society where many have experience of Brahman.

3 Views on evil

Evil, which is rooted in the insentience of matter and the limitations it imposes on life (the converse of valuable possibilities it secures), is fated to diminish and even disappear as evolution proceeds - possibly past the human species. The future evolution that Aurobindo envisages is the working out of a divine intention in which human effort, however, has a crucial role. Humans have developed a sufficient measure of freedom and self-determination to further the evolutionary progression, as divine delegates as it were. In a second, revised and expanded, edition of *The Life Divine* (1943-4), Aurobindo elaborates his theory of individual progress, amplifying a mystical psychology and enlarging on a theory of rebirth. Brahman's self-manifestation includes other 'worlds', or 'planes of being', said to be accessible to us in a mystic trance and which the developing divine individual ('psychic being' or 'soul') is said to enter upon the death of the body. The material universe does not exhaust the manifestation of Brahman, but it is, Aurobindo claims, the only evolutionary world. The others are 'typal', with no evolutionary emergence. In this world, the soul, profiting from all its experience, develops an increasingly refined and finely etched personality in terms of body, life and mind. The value of this development discounts, Aurobindo reasons, the evil made possible and even necessary by the insentience of matter inasmuch as matter makes possible such an evolutionary world.

4 Relation to classical Vedānta

Aurobindo manages an extraordinary consistency; critics have tended to fault him not for tensions among his many claims but for wrongly weighing 'mystical data'. Though little known in the West, his system enjoys in India significant influence both among professional philosophers and a larger intellectual community.

The speculative originality evident in Aurobindo's intricate theodicy and theory of evolutionary progress sets his system apart from classical Vedānta, the traditional Hindu philosophy based on sacred texts (principally Upaniṣads and the *Bhagavad Gītā*). But Aurobindo also relies heavily on these works. He says he understands them as records of mystical experiences similar to his own, not as mainstays of a revealed tradition. Critics, particularly in India, have often misunderstood Aurobindo's approach to traditionally sacred texts, viewing his philosophy as Neo-Vedānta and fuelling worries that Aurobindo, like all mystics perhaps, is ethnocentric in his claims. But despite the distancing from traditional Hinduism, such worries about cultural shaping remain (see Mysticism, nature of §3).

Western influences on Aurobindo's thought have not been adequately studied. Platonic and Neoplatonic reverberations abound, and there are echoes of Hegel and Nietzsche. Aurobindo read widely in Western literature (he won prizes in Classics at Cambridge) as well as in Sanskrit, which he learned through English, in effect his mother tongue. A hero of the Indian nationalist movement could be expected to cast his thought in the rich philosophical terminology of the classical Sanskrit tradition (consider also the case of Gandhi). But Aurobindo's inspirations were arguably as much Western as Indian. There is also with Aurobindo, as with all great thinkers, an element of creativity that transcends cultural precedent.

See also: Brahman; Vedānta

STEPHEN H. PHILLIPS

List of works

Aurobindo Ghose (1973) *Sri Aurobindo Birth Centenary Library*, Pondicherry: Sri Aurobindo Ashram Trust, 30 vols.(Aurobindo's collected works.)

Aurobindo Ghose (1914-21) *The Life Divine*, Pondicherry: Sri Aurobindo Ashram Trust, revised edn, 1943-4, 2 vols.(Aurobindo's major work of philosophy, written in dozens of chapters, with much repetition of the principal themes.)

Aurobindo Ghose (1914-21) *The Synthesis of Yoga*, Pondicherry: Sri Aurobindo Ashram Trust, revised edn, 1959.(Aurobindo's teaching of yoga, that is, of 'self-discipline' and mystical technique.)

Aurobindo Ghose (1918-21) *The Foundations of Indian Culture*, Pondicherry: Sri Aurobindo Ashram Trust, 1973.(An early work of an incensed Aurobindo defending Indian culture against European belittlement.)

Aurobindo Ghose (1950-1) *Savitri*, Pondicherry: Sri Aurobindo Ashram Trust.(The *Mahābhārata* story of Sāvitrī, who by debate wins back into life the soul of her husband from Yama, Death, is used by Aurobindo to express mystical experiences and a worldview in a massive English epic.)

References and further reading

Bolle, K.W. (1965) *The Persistence of Religion*, Leiden: Brill.(Debunks religion sociologically with special reference to Aurobindo's mystic views.)

Heehs, P. (1989) *Sri Aurobindo: A Brief Biography*, Delhi: Oxford University Press. (Contains useful summaries of the various areas of Aurobindo's writing as well as a lively account of Aurobindo's political life.)

Maitra, S.K. (1941) *An Introduction to the Philosophy of Sri Aurobindo*, Calcutta: Culture, 2nd edn, 1965.(A sympathetic and highly readable overview.)

Phillips, S.H. (1986) *Aurobindo's Philosophy of Brahman*, Leiden: Brill.(An inquiry into Aurobindo's metaphysics and mystic epistemology.)

Singh, S.P. (1972) *Sri Aurobindo and Whitehead on the Nature of God*, Aligarh: Vigyan Prakashan.(Perhaps the best of many comparative studies that have appeared in India.)

Austin, John (1790-1859)

Although written in the early nineteenth century, Austin's is probably the most coherent and sustained account of the theory of legal positivism. The complex relationships between legal positivism and the concepts of morality and politics are explored by him but are often neglected or misunderstood in modern commentaries.

John Austin became in 1829 the first Professor of Jurisprudence at the newly established University of London, having previously spent a brief spell in the army, followed by a somewhat longer period of work in barristers' chambers as an equity draftsman. In preparation to take up his chair, he spent two years studying in Bonn, acquainting himself with classical Romano-German legal science. A neighbour of James Mill and Jeremy Bentham, he was much of their intellectual circle, and had John Stuart Mill as a student. His main work was a systematic explanation of some basic social concepts in terms which have been received as the classical statement of 'legal positivism', according to which law is dependent on the will of superior(s) over inferior(s) (see Legal positivism §2).

Austin wanted all lawyers to commence their studies with the cognitive sciences, logic and ethics. Politics and the study of law were seen by him as subcategories of ethics. More than half of his introductory series of lectures was given over to an exposition of the principle of utility - seen either as the index to the laws of God, or as the rational basis of ethics (see Utilitarianism).

He said that all laws (apart from scientific laws) were set to intelligent beings by intelligent beings, having power over them. There are two basic types of laws - those set by God for people and those set by people for others. Of the laws set by God, Austin said that he would normally call these 'natural' laws, but as this word is likely to be confused with aspects of sociology, he would instead call them 'divine' laws. He recognized that 'non-believers' would prefer the word 'ethics'. On either view, these provide the standard by which all other laws should be judged. Many modern positivist scholars think that positivism and natural law are opposed to each other - not so for Austin.

Austin said that of laws set by people there can also be two types - those set by them as political superiors (positive law) and those set by them in other capacities (positive morality). Having set out the basic categories of ethics, positive law and positive morality, Austin went on to explain their common and distinctive features. They have in common the concepts of 'commands', 'duties' and 'sanctions'. The intelligent being or person lays down some *general* rule for the guidance of human conduct. If it is not complied with, the maker of the rule can exact some *detriment* for non-compliance.

The distinctive features of 'positive law' (the law of the state) are that it comprises rules made by 'sovereigns' addressed to 'subjects' within an 'independent political society'. While these ideas are scorned by most contemporary legal theorists, they remain fundamental in international politics and foreign affairs policy. Austin made it clear that the study of politics is logically and socially prior to the study of state law. Without political stability there is no state, and without the state no law of the state. The most contentious aspects of Austin's views were his claims that 'sovereigns' cannot be *bound* by state laws, and that the 'laws' operating between states would be better viewed as 'international positive morality'. Both points relate to the role of coercion, assuming that the author of a coercive law cannot, logically, also be subject to it, and that state boundaries limit the enforceability of state laws, but not necessarily their range of application (see Sovereignty §§2, 3).

Some suggest that his separation of the concepts of positive (state) law, positive (existing or customary) morality, and ethics (the critical base), entails that the positive law can be understood or implemented without understanding its relationship to the other factors. This is plainly wrong.

There has been much debate about this 'separation of law and morality'. Most writers on the topic forget that Austin said - 'Positive law (or *jus*), positive morality (or *mos*) together with the principles which form the test of both, are the inseparably connected parts of a vast organic whole' (1861-3: 17). An unfulfilled project of his was to write a book on *The Principles and Relations of Jurisprudence and Ethics*. It follows from Austin's analysis that as we are dealing with three types of rules there are also three types of duties as well as three types of commands and sanctions. Austin said that if there should be any conflict between the duties of positive law and the duties of

ethics or divine law, then because the sanctions which God could impose would greatly outweigh those which could be imposed by the human makers of positive law, it would accord with the principles of utility to act in accordance with the natural or divine law and not the law of the state.

The law of the state, he said, had its basis in the positive morality of the wider community, and should be guided in its development by the positive morality of the legal and international communities.

Austin had no time for 'the childish fiction employed by our judges, that judiciary law or common law is not made by them, but is a miraculous something made by nobody, existing, I suppose from eternity, and merely declared from time to time by the judges' (1861-3: 634). He explains at some length how it is that judges *make* the law rather than just find it, and entitles chapters 35-9 of his *Lectures* 'Judicial Legislation'. In its conceptual aspects, Austin's approach influenced the work of people such as W.N. Hohfeld, and in its social aspects has much in common with the writing of the later legal realists (see Hohfeld, W.N.; Legal realism).

Austin's strongest supporter was his wife Sarah, who was accomplished and well known in the intellectual circles of London and Paris, and with a number of important translations to her credit. As a teacher, Austin was not conspicuously successful, and his periods of formal academic employment were quite brief, at the University of London (1829-33) and at the Inns of Court in London (1834). He also worked with the British Criminal Law Commission (1833) and for a short period as a Commissioner to the Royal Commission on Malta (1836-8).

See also: Law, philosophy of; Norms, legal §2

ROBERT N. MOLES

List of works

Austin, J. (1832) *The Province of Jurisprudence Determined*, London: John Murray; with intro. by H.L.A. Hart, London: Weidenfeld and Nicolson, 1954; with intro. by W.E. Rumble, Cambridge: Cambridge University Press, 1995.(The only book to be published by Austin in his lifetime, being a series of six lectures originally delivered as ten. It states his central tenets in clear terms, and has remained a focus of controversy, and of frequent attempts at refutation, more or less since its first publication. The 1995 edition includes a good bibliography.)

Austin, J. (1859) *A Plea for the Constitution*, London: John Murray.(This is the best known and most significant of Austin's later pamphlets; it is a forceful statement of opposition to radical reform of the franchise, based on the idea that habitual or customary acceptance of a constitutional order is the foundation of a stable polity and thus of law.)

Austin, J. (1861-3) *Lectures on Jurisprudence or the Philosophy of Positive Law*, ed. R. Campbell, London: John Murray, 5th edn, 1885.(This is the complete statement of Austin's views and of his legal-conceptual analyses, based on his London lectures. The two volumes were first published from his working papers as edited by Sarah Austin. First and subsequent editions contain the most moving piece about Austin's life, in the Preface written by Sarah.)

References and further reading

Anon. (1861) 'English Jurisprudence', *The Edinburgh Review* 114: 463.(This early review of Austin's *Province* exemplifies the nineteenth-century response thereto.)

Anon. (1863) 'Austin on *Jurisprudence*', *The Edinburgh Review* 118: 439.(A contemporary reviewer's response to the *Lectures*.)

Hamburger, L. and Hamburger, J. (1985) *Troubled Lives: John and Sarah Austin*, Toronto: University of Toronto Press.(A biographical account of a remarkable marriage, showing that behind a staid exterior storms may rage.)

Hart, H.L.A. (1961) *The Concept of Law*, Oxford: Clarendon Press, 2nd edn, 1994. (This exceptionally influential work contains a sustained critique of Austin's work, and an attempt to replace it with a different version of legal positivism.)

Hohfeld, W.N. (1919) *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, New Haven, CT: Yale University Press.(An influential account of such basic concepts as 'duty', 'right' and 'power' in a way that develops some of the ideas originally deployed in Austin's *Lectures*.)

Moles, R.N. (1987) *Definition and Rule in Legal Theory*, Oxford: Blackwell.(This presents a reasoned rejection of most conventionally accepted critiques of Austin's work, especially that of H.L.A. Hart, which it claims rests

on a serious misunderstanding of Austin.)
Morison, W.L. (1982) *John Austin*, London: Edward Arnold.(A readable, but in places polemical, account of Austin’s life, work, and jurisprudential doctrines, arguing for the substantial correctness of Austin’s views. Good bibliography.)
Rumble, W.E. (1985) *The Thought of John Austin*, London: Athlone Press.(This is a good scholarly account of Austin’s work; good bibliography, and comprehensive survey of the works.)

Austin, John Langshaw (1911-60)

J.L. Austin was a leading figure in analytic philosophy in the fifteen years following the Second World War. He developed a method of close examination of nonphilosophical language designed to illuminate the distinctions we make in ordinary life. Professional philosophers tended to obscure these important and subtle distinctions with undesirable jargon which was too far removed from everyday usage. Austin thought that a problem should therefore be tackled by an examination of the way in which its vocabulary is used in ordinary situations. Such an approach would then expose the misuses of language on which many philosophical claims were based.

In ‘Other Minds’ *([1946] 1961), Austin attacked the simplistic division of utterances into the ‘descriptive’ and ‘evaluative’ using his notion of a performatory, or performative utterances. His notion was that certain utterances, in the appropriate circumstances, are neither descriptive nor evaluative, but count as actions. Thus to say ‘I promise’ is to make a promise, not to talk about one. Later, he was to develop the concepts of locutionary force (what an utterance says or refers to), illocutionary force (what is intended by saying it) and perlocutionary force (what effects it has on others).*

1 Life and early period

John Langshaw Austin, a graduate of Balliol College, Oxford, was successively a Fellow of All Souls’ (1933) and Magdalen Colleges (1935) at Oxford until appointed White’s Professor of Moral Philosophy in 1952. He was a distinguished Intelligence Officer in the Second World War, latterly at Supreme Headquarters, Allied Expeditionary Force, being awarded the OBE, the Croix de Guerre and made an Officer of the Legion of Merit.

In the pre-war period Austin worked mainly on the history of philosophy, particularly on Aristotle and Leibniz, though publishing nothing on them. He was a caustic observer of the contemporary scene, as is clear from his only pre-war publication, the devastating, amusing and wholly negative paper ‘Are There A Priori Concepts?’ ([1939] 1961), but as yet he spoke with no clear voice of his own. He himself regarded his later work as quite different from anything he had done before the war.

2 Analysis of ordinary language

During his five years absence on military service Austin’s thought matured, and on returning to Oxford in 1945 he immediately became a leading force in philosophy there and, soon after, in analytic philosophy throughout the world. His work was based on three convictions.

(1) Ordinary language, by which is meant language other than that of philosophy, as the tool of communication, contains all the distinctions about the world that people have found it necessary to make - not, of course, all that can be made. As such, he thought, while it was not perfect, it was a much more powerful and subtle tool of thought than philosophers had traditionally recognized. What, for example, philosophers have hastily treated as synonyms, would usually be found on careful examination to mark distinctions that might be important. Where synonyms arose, the daily use of language would usually lead either to the abandonment of one or the development of a distinction in their employment.

(2) Philosophers consistently misused and abused ordinary language, blurring and perverting the distinctions it made. When they abandoned ordinary language in favour of a technical vocabulary of their own it was usually confused and imprecise, creating confusion and darkness rather than shedding light. Thus they neglected and damaged the powerful tool available to them in ordinary language in favour of a less efficient one; no wonder that little headway was made in answering the problems they tried to solve.

(3) While holding that much philosophical work was full of confusion, Austin did not, however, share the view that the sole task of the philosopher was to expose these confusions and to ‘show the fly the way out of the fly-bottle’ (see Wittgenstein, L. §17). He thought that progress could be made, and that philosophy could shed light as well as clear away fog. But this required slow and careful labour, especially including a thorough examination of the vocabulary available and used in the area where the problem arose, long before asking the huge and assault-defying major questions.

Some have wrongly thought that Austin believed that his way of working in accordance with these convictions was

the be-all and end-all of philosophy; this he denied, though he said that it might be the begin-all. But he once gave the title ‘One way of possibly doing one part of philosophy’ to an unpublished talk, and he ascribed his lack of use of more formal procedures to ignorance.

Most of Austin’s work based on these convictions was carried out in discussion and in lectures. He believed that his sort of work on language was best done by groups working together. He himself published only a handful of papers, never spontaneously, but almost always at the request of the Aristotelian Society and for symposia, the subjects of which were often not of his choosing. But from these, together with posthumous publications, it is possible to illustrate how his fundamental ideas were employed.

In his posthumously published lectures entitled *Sense and Sensibilia* (1962a) (Austin parodying Austen), he is mainly concerned to illustrate the second conviction above. He looks closely, very closely, at the invented philosophical terminology for discussing the problems of perception - terms such as ‘sensum’, ‘sensibile’ and ‘sense-datum’ - and also at the use (or, as he thought, abuse) by philosophers of such existing terms as ‘illusion’, ‘delusion’, ‘look’, seem’, ‘appear’ and ‘veridical’ (see Perception §1). Thus, for example, philosophers had tended to treat ‘look’, ‘seem’ and ‘appear’ as synonyms, which, he tries to show, they are not. The argument from illusion, he said, gained virtually all its force from misuse of words. That, for example, the fact that a penny looks elliptical from certain angles is part of our evidence that it is round, not an illusion that suggests that it is not round. He uses A.J. Ayer’s book *The Foundations of Empirical Knowledge* (1940) as his main illustrative text, attempting to show that Ayer’s problems are largely of his own making and result from abuse of language (see Ayer, A.J. §3). In his paper ‘Other Minds’ ([1946] 1961) he similarly seeks to show that much scepticism about the possibility of knowledge was based on continuing to use the word ‘know’, but with impossibly stringent criteria of application quite different from those of ordinary speech. If we cannot know according to criteria we never use, what then? In this paper Austin also insists that if, having made a claim to knowledge in circumstances which would normally be considered to provide complete justification for the claim, some freak circumstance arises (in Austin’s example, what the experts agree to be a goldfinch explodes or quotes Virginia Woolf), we should not naturally agree that the experts were wrong, but rather admit that words fail us. Language makes no provision for circumstances which do not in fact arise.

His more constructive work in this field is to be found in some of his papers. Thus there is the traditional problem of the nature of action. Austin was sure that to plunge in at the deep end by asking ‘What is an action?’ was mere folly. The way to start was by a close examination of the language ordinarily used for discussing action; a sub-area of this vocabulary is that which is concerned with the defects that actions may exhibit; they may, for example, be unintended, mistakes, inadvertences, involuntary or clumsy (see Action §5). As a deliberate illustration of his general ideas, he set out to examine some of these terms in a paper called ‘A Plea for Excuses’ ([1956a] 1961). His technique, as usual, was to take a real or imaginary situation and ask what we should say about it. Thus, take a situation in which you and I keep a donkey each in the same field and I, wishing to shoot mine, instead shoot yours. If I (a) fire accurately at your donkey, thinking it is mine, or (b) fire at mine, but the bullet ricochets and hits yours, which, if either, would you call a mistake or an accident? Or are the terms interchangeable? Austin had an unusual facility for bringing out differences which we implicitly observe in our daily use of language, but which we do not ordinarily make explicit and which we may fail to notice in philosophical discussion. ‘A Plea for Excuses’ also contains a short but illuminating account of his methods of work at a more general level.

Another important illustration of this aspect of Austin’s work is the paper ‘Ifs and Cans’ ([1956b] 1961). Here he typically approaches the problem of the freedom of the will by a careful examination of such expressions as ‘I could have…’ and, especially, ‘I could have if I had chosen…’, rather than by an immediate attempt on the big question (see Free will §§1-2). He brings out defects in the account given by G.E. Moore in his *Ethics* (1912), and by others, of the way such expressions are used in real life. Notably he brings out the inadequacy of the traditional accounts of hypotheticals. In the sentence, ‘There are biscuits on the sideboard if you want some’, for example, the truth of the conseqent should not be taken to follow from the truth of the antecedent. What is the force of ‘if I choose’? Does it give a necessary or sufficient condition of one’s being able to act, or does it function in some way more similar to that of ‘if you want some’ in the example above?

3 Philosophical linguistics

Austin thought that ordinary language did not provide us with ready-made distinctions for use in discussing the

nature of language, and that many of the technical terms already existing in philosophy were ill-suited to this purpose. Here was an area in which he thought that technical terms were necessary, not as a substitute for existing linguistic resources, but as a supplement to them. One of his most disliked targets was the distinction commonly made between the descriptive and evaluative uses of language, which he thought to blur a host of more effective distinctions that needed to be made. He started to remedy the situation by drawing attention to what he first called the performatory, later the performative, use of language. It was introduced in the paper ‘Other Minds’ with great brevity; a performative utterance is typically couched in an indicative sentence but is neither a description nor an evaluation, nor is it true or false; rather, one utters a formula or performs a ritual in appropriate circumstances which itself counts as the performance of an act which would not be naturally regarded as a linguistic act, provided that the utterer is in a position to make the utterance. Thus to say ‘I promise’ is to make a promise, not to talk about one, and to say ‘I name this ship "Generalissimo Stalin"’ is, if one is a duly authorized person, to give a ship a name. There are many types of action that can be best performed by such specified formulas, such as making contracts, making appointments and christening. Making a promise, for example, is more like shaking hands over a bargain in a recognized context, than describing anything whatsoever (see Performatives).

In his posthumously published William James Lectures, *How to Do Things with Words* (1962b), based on earlier lectures given at Oxford over a series of years with the title ‘Words and Deeds’, Austin set out to provide an ambitious classification of types of force of linguistic utterances on a far wider scale. Thus at a higher level he distinguished between different types of force, such as locutionary, illocutionary and perlocutionary force. The locutionary force is roughly the sense and reference of the utterance, the illocutionary force is what one is doing in making an utterance with a given locutionary force, and the perlocutionary force is the effect that is intended to be achieved by an utterance with a given illocutionary force. Thus the locutionary force of ‘The bull is charging’ is simply the predication of something of an animal; the illocutionary force may be that of a warning, a comment on the scenery or an exclamation; and if the illocutionary force is that of warning, the perlocutionary effect intended may well be to make somebody run for it. Not all these forces are always present; I may utter a sentence without even locutionary force merely to test a microphone or to give an example in a philosophical paper. So Austin further distinguishes the merely phonetic act of making a noise vocally and the phatic act of uttering an intelligible sentence (as when I say ‘The bull is charging’ merely to test a microphone) from each other and from the other forces (see Speech acts §1).

These and other distinctions have been widely used, sometimes with modifications, and also, needless to say, widely criticized (for example, Searle 1969). Their influence is apparent in theoretical linguistics as well as in philosophy. What is certain is that this work has had profound effects on philosophical linguistics. One notable feature is the way that Austin chose to present his classification of utterances as arising from a criticism and, eventually, the abandonment of the performative as a special type of utterance and its subsumption under the doctrine of illocutionary forces. Thus what had been counted as the action performed by the utterance of a performative now becomes the illocutionary force of that utterance, and the performative is no longer treated as importantly different from other speech acts having an illocutionary force. This may accurately reflect the historical development of Austin’s views, but it appears to some to involve a denaturing of the performative (Warnock 1973; Urmson 1977). In particular Austin now includes as performatives utterances which seem to lack the formulaic or ritual character early ascribed to performatives.

See also: Ordinary language philosophy, school of

J.O. URMSON

List of works

Austin, J.L. (1961) *Philosophical Papers*, Oxford: Clarendon Press; 3rd edn, enlarged, 1979.(Contains ‘Are There A Priori Concepts?’ (1939), ‘Other Minds’ (1946), ‘A Plea For Excuses’ (1956a), and ‘Ifs and Cans’ (1956b), and all Austin’s writings except two or three reviews and the books listed below.)

Austin, J.L. (1962a) *Sense and Sensibilia*, reconstructed from the manuscript notes by G.J. Warnock, Oxford: Clarendon Press.(Attacks all sense-datum theories of perception, but principally phenomenalism as developed by A.J. Ayer. It is urged that these theories stem from a perverted interpretation of the main terms which are used with reference to sense perception. Originally delivered as lectures at Oxford.)

Austin, J.L. (1962b) *How to Do Things with Words*, Oxford: Clarendon Press; revised edn by J.O. Urmson and M. Sbisà, 1975.(Develops a theory of speech acts based on a critique of Austin’s own theory of performative

sentences that we utter and the acts, such as stating, promising, condemning, that we perform in uttering them. The William James Lectures given at Harvard in 1955, based on lectures delivered at Oxford previously.)

References and further reading

Ayer, A.J. (1940) *The Foundations of Empirical Knowledge*, London: Macmillan. (A phenomenalist theory of perception.)

Ayer, A.J. (1973) 'Has Austin Refuted Sense-Data?', in *Essays on J.L. Austin*, Oxford: Clarendon Press.(Answers Austin's criticism of the above in *Sense and Sensibilia*.)

Berlin, I. *et al.* (1973) *Essays on J.L. Austin*, Oxford: Clarendon Press.(Essays discussing both the man and his work.)

Fann, K.T. (ed.) (1969) *Symposium on J.L. Austin*, London: Routledge & Kegan Paul. (Contains many relevant papers.)

Moore, G.E. (1912) *Ethics*, London: Williams & Norgate.(A utilitarian theory based on the idea of 'good' as an a priori concept.)

Searle, J. (1969) *Speech Acts*, Cambridge: Cambridge University Press.(A development and criticism of Austin's *How To Do Things With Words*.)

Urmson, J.O. (1977) 'Performative Utterances', in *Midwest Studies in Philosophy* 2. (Criticizes Austin's later treatment of performatives.)

Warnock, G.J. (1973) 'Some Types of Performative Utterance', in *Essays on J.L. Austin*, Oxford: Clarendon Press.(Criticizes Austin's later views on performatives.)

Warnock, G.J. (1980) *J.L. Austin*, London: Routledge.(A general critical account of Austin's philosophy.)

Australia, philosophy in

Australian academic philosophy has made an international impact disproportionate to the country's small population, though its beginnings contain little that might have suggested such influence. The first Philosophy Chair was established at the University of Melbourne in 1886 and its occupant Henry Laurie was more notable for extravagant shyness than public impact or academic achievement. Until the 1920s, the dominant philosophical outlook was idealism. After the arrival from Glasgow of the charismatic John Anderson to the Chair in Sydney in 1927, this outlook was challenged by his vigorous, distinctive, highly metaphysical and somewhat dogmatic version of realism. Anderson had little international recognition during his working life, but he had a powerful effect upon Australian cultural life and upon students who themselves achieved a significant international presence. Thinkers like David Armstrong, John Mackie and John Passmore diverged in many ways from Andersonianism but the indelible mark of the Sydney baptism remained with them even when they had accommodated to the international profile.

For twenty-five years or so, there was a strong contrast and rivalry between the style of philosophy done in Sydney and that done in Melbourne. Idealist influences persisted longer in Melbourne due to the two Boyce Gibsons (father and son) who occupied the Melbourne Chair successively, but the significant contrasts really began when Melbourne came under the sway of Wittgenstein's philosophy in the 1940s. This was due principally to the presence during the war years of G.A. Paul, one of Wittgenstein's pupils, and later Paul's friend Douglas Gasking, who had studied under Wittgenstein in Cambridge, and another pupil of Wittgenstein, A.C. ('Camo') Jackson. Where Anderson's orientation was systematic, metaphysical and provincial, the Melbournians were piecemeal, anti-metaphysical and (relatively) cosmopolitan. As a direct force in academic philosophy, Anderson's system died with him in 1962, as did the striking contrast in style between Melbourne and Sydney philosophy. With the expansion of universities and philosophy departments, the metaphysical emphasis of Sydney and the analytical professionalism of Melbourne merged in a technique that had no particular regional significance, even when some of its concerns were distinctive. Among these was the phenomenon known as Australian Materialism, associated principally with J.J.C. Smart and David Armstrong. This continued the metaphysical orientation of so much Australian philosophy, though deploying the analytical and argumentative skills by then common to English-speaking philosophy anywhere. Much of the passion surrounding the materialism debates of the 1960s and 1970s, involving the pros and cons of 'the scientific world view' and its reductionist enthusiasms, dissipated into broader metaphysical and psychological interests such as the discussion of universals and laws, realism versus antirealism, the ontology of space and time, and the status and pretensions of cognitive science.

There remains important work that is somewhat independent, even occasionally sceptical, of these metaphysical directions: work in epistemology, philosophical psychology, history of philosophy, and value theory. In pure value theory, there has been little home-grown work that is highly original though there have been many solid contributions by Australian philosophers to international debates, and Peter Singer is famous beyond philosophical circles for his theorizing of 'animal liberation' and opposition to 'sanctity of human life' outlooks in bioethics. The general tenor of Australian philosophy remains resolutely 'analytical' though there is a significant minority interest in 'continental' philosophy and some efforts to reach a modus vivendi *between the two. Until late in the twentieth century, women played no prominent role, but women philosophers and feminist philosophy have become increasingly significant and, although many find the 'continental' mode congenial to their approach, there is strong representation of the more 'analytic' tradition. Another prominent emphasis has been environmental philosophy which incorporates the traditional interest in metaphysics but with a less reductive touch than has been characteristic of the mainstream.*

1 Beginnings

Academic philosophy, the home for almost all contemporary philosophy, effectively began its Australian life in Melbourne where the first lecturer in logic was appointed at the University of Melbourne in 1881. The appointee was the Scottish-born journalist, Henry Laurie, who shortly afterwards (1886) became the country's first philosophy Professor. Prior to Laurie, some logic and political theory had been taught at Melbourne by the redoubtable W.E. Hearn, an extraordinary polymath with an international scholarly reputation, who also taught at

various times English, Ancient and Modern History, Political Economy and Classical Literature, before becoming the University's first Dean of Law. In Sydney, some logic had been taught along with classics since the University's foundation in 1850, but a lectureship in philosophy was not established until 1888. There had been a compendious Chair in English Language and Literature and Mental and Moral Philosophy established earlier at the University of Adelaide, but its first occupants had no philosophical claims. The professing of philosophy there really began in 1894 with the appointment of the gifted, if often obscure, William Mitchell who gave the prestigious Gifford Lectures in Aberdeen in 1924-6, later published as *The Place of Minds in the World*. Mitchell (later Sir William) was a great public figure in South Australia where he was Vice-Chancellor of the university for twenty-six years and Chancellor for a further six. He died in 1962 at the age of 101 and, in his nineties, was capable of vigorous exchanges with one of his young successors in the philosophy Chair, J.J.C. Smart.

Both Laurie and Mitchell were Scots and the Scottish influence on the formative years of Australian philosophy is even more remarkable than its notable influence upon Australian universities generally. In Sydney, the Scottish influence was if anything even more emphatic and enduring - it was Glasgow University that produced the first lecturer in philosophy (1888) and then the Challis Professor of Mental and Moral Philosophy (1890) in the (same) person of Francis Anderson. Like Mitchell, Francis Anderson was later knighted, but more for his services to education than philosophy. Some thirty-seven years after Anderson's appointment to the Chair, another (unrelated) Anderson came from Glasgow (via Cardiff and Edinburgh) as the third and most famous occupant of the Chair, though there was never any question of his being given, or accepting, Royal honours.

Philosophy in Australia in the late nineteenth century made little impact on the rest of the world and not much impact on the country's intellectual or social life. The US idealist philosopher, Josiah Royce, took leave from Harvard to sail to the Antipodes in 1887, seeking recovery from a bout of depression. The visit cured him, and much that he observed filled him with admiration for Australia and New Zealand. Upon his return to the USA, he wrote extensively of his impressions, but made no comment upon the fledgling state of academic philosophy or, indeed, of university life at all. In political philosophy, Royce was what would now be called a 'communitarian', and he was taken with what he thought was the more community-oriented nature of Australian attitudes compared to the individualism of the USA. If he made no contact with academic philosophers, he did become acquainted with some influential politicians and was greatly impressed by the philosophical outlook and statesmanship of the Victorian politician Alfred Deakin, who was to be one of the 'founding fathers' of the Federated Commonwealth, and eventually Australia's second Prime Minister. Royce described Deakin as 'a lover of metaphysics' and continued to correspond with him after returning to the US. Deakin seems to have drawn his own philosophical inspiration from overseas and, in any event, no Australian philosophers contributed significantly to theoretical debates about Federation or the form of the Australian Constitution as it was drawn up in 1900. Whether because of this, or because so few of its drafters had the philosophical and cultural leanings of a Deakin, the document is notably prosaic, legal, and uninspiring. This is in striking contrast to the highly philosophically oriented US constitution, but that was of course the product of a revolution of arms and of thought, whereas it was Australia's fate to achieve a peaceful evolution to independence.

2 Early influences

Well into the twentieth century, Australian philosophy was predominantly idealistic in orientation, eclectic in shape and inspired by foreign models. Its Scottish origins ensured the influence of Scots thought, but the great days of the Scottish School of Common Sense were past by the time philosophy began to be taught seriously in Australia, so it was Sir William Hamilton's rather windy, Neo-Kantian 'corrections' of Thomas Reid, and A.S. Pringle Pattison's idealism that prevailed, rather than Reid's own vigorous, realistic, somewhat Aristotelian elaboration of a common-sense philosophy (see Commonsensism). Earlier in the century, Reid had had a powerful influence upon the development of US philosophy, an influence which contributed to the growth of a distinctively national style of philosophy in the form of pragmatism. In Australia, however, the initial Scottish influences gave rise to no native school. Mitchell, it is true, made many shrewd philosophical observations and developed an original and, to some degree, internationally recognized version of idealism, but the opacity of his thought virtually prohibited its development by others. Laurie was a capable teacher, but he had no distinctive philosophy to impart and was, in addition, painfully shy. Francis Anderson, though more expansive, was notable for educational reforms rather than philosophical innovation.

Besides the Scottish influence, there were also sources from Continental Europe - Hegel, of course, but subsequently Bergson, Husserl and Eucken, these latter notably present in Laurie’s successor at Melbourne, W.R. Boyce Gibson. The older Gibson assumed the Chair in 1911 and his son Alexander followed him in 1935. The Gibson reign thus lasted fifty-four years, from 1911 to 1965. Gibson senior studied in Europe and, at Jena, came under the spell of Rudolf Eucken. Later, he attended Husserl’s seminars and translated one of his books. It is a curiosity of the history of thought that Eucken (now almost utterly forgotten) should have so influenced the older Boyce Gibson in Jena when there was in that same university at the time a philosopher at the height of his powers, destined to exert a massive influence upon modern analytical philosophy. Of course, Gottlob Frege was in the mathematics department and unacknowledged in his own time and his own university, but what a difference it might have made to the development of Australian philosophy had Gibson brought Frege rather than Eucken to Melbourne! Although Gibson had a mathematics background and was to co-author a logic text-book, he seems to have known nothing of Frege and would have found Frege’s realism nowhere near as sympathetic as Eucken’s personal idealism.

Idealism was thought to support certain forms of religion, indeed of Christianity, and there is much emphasis upon the significance of religious belief among the early Australian philosophers. This emphasis is both metaphysical and moral. It engages with the Victorian era’s anxieties about whether faith can survive the encounter with science and, even more urgently, whether morality can survive the demise of religion. The response is heavily, if sometimes inflatedly, metaphysical and the metaphysical interest will later survive the disappearance of the original question and the attempt to bolster religion. It is arguable that this metaphysical cast of mind is one of the most distinctive features of the Australian philosophical tradition, and it early received one of its most striking manifestations in the work of an expatriate Australian, Samuel Alexander.

Alexander was born in Sydney in 1859, but pursued undergraduate studies at the University of Melbourne before leaving Australia to study at Oxford. He was made Professor of Philosophy at Manchester in 1893, having earlier had the distinction of being the first Jew appointed to a Fellowship at either Oxford or Cambridge. During the First World War he gave the Gifford lectures in Glasgow, published as *Space, Time and Deity* in 1920. Alexander’s work is a fascinating blend of old and new. It has a traditional commitment to large-scale metaphysical vision, a commitment that would soon begin to appear outmoded, but it was determinedly realistic in a way that placed Alexander with the burgeoning revolt against idealism. He gave a primary emphasis to experience and science, but insisted on finding a role for religion and God. His moral philosophy was also markedly metaphysical and unusual. In spite of his birth and upbringing, Alexander can only be considered an Australian philosopher in a peripheral sense since his philosophical career was spent entirely in Britain, but his mode of thinking had a powerful influence in his birthplace through its impact upon John Anderson and, to a lesser degree, its very different effect on the second Boyce Gibson. Anderson took from him the opposition to idealism in any form, the resolute realism, the emphasis upon space-time and categories of reality, and the dedication to systematic thought. Gibson was impressed by the emphasis on the significance of religious experience. Interestingly, Alexander’s influence on Australian philosophy was not only abstractly intellectual - he was on the selection committees that appointed both of the Boyce Gibsons and seems also to have had a hand in Anderson’s appointment.

3 Anderson

When he was a lecturer at Glasgow, John Anderson had heard Alexander’s Gifford lectures, and when appointed to the Sydney Chair in 1927 Anderson had already begun to develop the highly individual outlook on philosophy that was to have such a powerful influence in his adopted country. It was an outlook that gave a distinctive cast to much of the philosophy done in Australia during his thirty-one years in the Chair.

The really remarkable features of Anderson’s work were less its realism and argumentativeness (aspects of a mood gathering force throughout much of the philosophical world during Anderson’s working life) than its commitment to systematic metaphysics and the eclecticism of its sources. These qualities set Anderson’s thought apart from the styles of philosophy that came increasingly to dominate Oxford and Cambridge from the 1920s to the 1950s, for these were mostly hostile to systematic, constructive metaphysics. To be educated by Anderson was to be inducted into a different world, though there was a similar commitment to close, often deflationary, examination and criticism of argument: the term ‘criticism’ had a mantra-like quality in Andersonian circles that sometimes consorted comically with the deferential use of the system.

Students found that there was not only an Andersonian position on everything from logic to love, from copula to community, from hypotheticals to Hegel, from reason to religion, but that these positions had an (at least apparent) interconnection that could be intoxicating. In addition, the system provided a comprehensive tool for the demolition of the conventional wisdoms of the day, whether religious, political, moral or institutional. Additional edge was provided by a special vocabulary which often employed familiar philosophical jargon in an unfamiliar sense. It was important to be ‘empiricist’ but not to believe in sense-data or ideas; ‘dualist’ errors abounded well beyond the philosophy of mind and were committed by believing in God or in any sort of ultimates. From within the system, the misguided were shot down in flames with monotonous regularity, though they were seldom persuaded since they rarely understood the terms of the debate. As the poet James McAuley put it: ‘Anderson had an answer to every conceivable question. It was "No"’. In Anderson’s case, as in many others, there was a sort of religious, even sectarian quality about his authoritarian presence, his campaigns and many of his followers, which may help explain the combination of liberation and enslavement that Anderson’s influence could produce.

David Stove, who absorbed Anderson’s influence, but rejected Andersonianism, and developed his own independent, highly polemical philosophical position, once remarked: ‘Anderson’s powers of dismissal were simply boundless…. Einstein scored just one two-line reference in print from Anderson: it said that his theory was "utterly illogical"’ (Grave 1984: 47). The conviction behind such sentiment owed nothing to the developments of modern mathematical logic which Anderson viewed with similar scorn. His logic was based on the traditional syllogistic, ingeniously adapted to try to deal with its apparent deficiencies, and had an almost mystically metaphysical solidity about it. The propositions it dealt with were not ideas or linguistic items or mathematical formulae, not anything representing or picturing other realities or states of affairs, but real states of affairs themselves. To suppose otherwise was to commit the cardinal sin against Anderson’s rigorous realism by proposing intermediaries between thinkers and the realities they were considering. For Anderson, the questions which greatly exercise many contemporary philosophers about the role of conceptual schemes in our understanding of reality would have been non-questions. Things, situations, processes existed in space-time and discussion should attend to determining the truth of such matters.

Anderson professed little concern for worldly success or academic advancement, and denounced the cosy certainties of consumerism and progressivism; indeed one of the few things he admired about religion was its potential for opposing such idols. Early on, he was involved with the Australian Communist Party, but became more and more critical of Marxism and moved to a strongly anti-communist position by the late 1930s. He wrote quite a lot but published little in major international journals, and some of his most important writings appeared in obscure places. Gilbert Ryle invited him to contribute to *Mind*, but he saw no point. He drew inspiration from the ancient Greeks, though selectively. He admired the objective attitude of the Presocratics which he contrasted unfavourably with the anthropocentric, subjective concerns of post-Cartesian philosophy, and he had a special admiration for Heraclitus. Hegel was respected, partly for his opposition to individualism and for the scope of his thought, and Moore, Russell and Alexander were considered significant enough to be critically adapted, but modern influences on his thought tended to come as much from outside the orthodox philosophical canon as from within: Marx and Freud, James Joyce and Sorel, Hilaire Belloc and James Burnham.

Some of his students pursued academic philosophy as rather imitative disciples, others became important contributors to philosophy on the Australian and the world scene. John Passmore, David Armstrong, John Mackie and Eugene Kamenka have all expressed a sense of indebtedness to Anderson and, although his system virtually died with him, its influence can be seen in some of their characteristic views. In other areas of the academy, and in the arts and the professions, people profoundly influenced by Anderson had a striking impact on Australian life, especially in New South Wales and Canberra. Two of Australia’s finest poets, James McAuley and Alec Hope, were students of Anderson and, as a young man, Hope engaged in print in respectful philosophical controversy about moral philosophy with his mentor. In mature life, neither could be regarded as Andersonians - indeed, McAuley converted to a very conservative version of Roman Catholicism - but typical Andersonian concerns remained with them, as with many journalists, senior lawyers, judges and public servants.

4 The Melbourne flavour

In Melbourne, Anderson’s influence was minimal during his heyday, for not only did idealism linger with the influence of Gibson in the 1930s, but Melbourne became an outpost of the new philosophical revolution associated

partly with logical positivism and more significantly with the later Wittgenstein. A primary influence in this development was the arrival of G.A. Paul in 1939, fresh from Wittgenstein's Cambridge. Stranded in Australia by the outbreak of the Second World War, Paul was another Scots-born philosopher to exert a great influence upon Australian philosophy. When he returned to England at the end of the war, he was replaced by his friend Douglas Gasking, whom he had earlier encouraged to migrate to a lectureship in Brisbane. Paul and Gasking had both studied under Wittgenstein as did A.C. ('Camo') Jackson, who went from Melbourne to Cambridge for Ph.D. studies in 1946 and returned to a lectureship in the Melbourne department in 1948. Paul not only brought the new philosophy to Australia, but had an immense influence upon the development of other disciplines at the University of Melbourne, most notably History. In one year, all the full-time members of the History department attended Paul's lectures on logic. Like Anderson's, Paul's impact could be partly explained by his being a big fish in a small pond (an especially small pond, in Paul's case, because of the drainage of intellectual talent caused by the war in the period of his greatest influence) but unlike Anderson, Paul was not an original thinker, and the power of his ideas came largely from afar even if he (like Jackson) clearly was a remarkable teacher. He published very little, and the best of that before he arrived in Australia; his paper 'Is there a Problem about Sense-Data?' was his most influential contribution.

The Wittgensteinian tradition thus established in Melbourne began with a somewhat positivist flavour, but later broadened under the impact of Oxford philosophy and the arrival of refugee intellectuals from continental Europe, many of whom had been absurdly deported from Great Britain and interned in Australia during much of the war. During the 1950s, the Melbourne department was host to a number of philosophers, both foreign-born and locally educated, who later left to pursue philosophy overseas. These included such well-known names as W.D. Falk, Kurt Baier (his New Zealand wife Annette also worked in Australia), Alan Donagan, Brian O'Shaughnessy, Paul Edwards, Michael Scriven and George Schlesinger. This established a trend for a later export industry - John Mackie, Jenny Teichman, Michael Devitt, Mark Johnston, to name a few - and the import-export aspect has continued with people like Michael Stocker and Michael Tooley coming to Australia from North America for long stays and then returning. Whereas Andersonian Sydney into the 1950s was confidently parochial and mostly contemptuous of recent developments in international philosophy, Melbourne stood for a more cosmopolitan and contemporary approach. It also hosted a diversity of outlooks and a professional, problem-oriented approach to the subject that Anderson dismissed as 'eclecticism'. The influence of Wittgenstein was strong, but not overwhelming, and indeed many of the best-known products of the department from that period, such as McCloskey, Baier, Donagan and Charlesworth were not in any sense Wittgensteinians. Douglas Gasking succeeded Boyce Gibson in the Melbourne Chair in 1966 and the Melbourne Department remained less metaphysically oriented than other departments in the country or indeed the new departments in the city of Melbourne. The rapid expansion of university education in Australia from the 1960s eliminated the dominance and the opposition of Melbourne and Sydney philosophy; what remains is mostly a Hegelian synthesis of the metaphysical interest derived from Sydney and the internationalist 'analytic' professionalism drawn from Melbourne.

5 Australian materialism

Since the late 1950s, the metaphysical stream in Australian philosophy has been in flood, first in the theory known as central state (or 'Australian') materialism, and then in a particularly forceful version of metaphysical realism. There has also been a persistent interest in the ontology of space and time. J.J.C. Smart, introduced to philosophy in Glasgow and continuing his studies in Oxford, was converted by U.T. Place from the influence of Ryle. In Adelaide he developed a version of materialism about the mind which was taken up and further developed by D.M. Armstrong, first in Melbourne and later when he moved to the Chair in Sydney in 1964. Armstrong's book, *A Materialist Theory of the Mind* (1968), became something of a bible for the materialist school. It propounded the view that 'the mind is the brain' and defended it against a variety of objections. There were differences of emphasis and formulation among the Australian materialists and there were similarities between their approaches and those of several physicalist and materialist philosophers in Britain and the USA.

The background to the movement was an increasing emphasis on the need to locate philosophy within the perspective of the physical sciences. The relation of philosophy to the developing modern sciences had posed a constant puzzle to philosophers since Descartes, but it became an abiding preoccupation once the 1960s saw the development of a dominance by the USA in academic philosophy. Much philosophy in the USA in the twentieth century has had a strong orientation to the physical (and sometimes social) sciences and, since Quine, the idea that

philosophy is at best a part of science has achieved a certain orthodoxy. Two aspects of the modern physical sciences particularly impressed the Australian materialists - comprehensiveness and reductionism. Armstrong writes with persistent admiration of 'the scientific world view' as something to which philosophy must not only be accommodated but subject, and the basic concept of the material is elucidated deferentially in terms of whatever may be the basic explanatory concepts of contemporary physics. The success of reductive strategies in chemistry and biology encouraged the idea that all manner of philosophically intractable realities would be more manageably understood in terms of scientific items. Much room remained for debate about the nature of the reduction. The identity theory that underpinned the approach generated worries about the respective merits of type identity and token identity, the precise role of functional explanations of mental concepts within the materialists' framework, and the relation of central-state materialism to eliminative materialism. Eliminativists treat the mental as rather like the magical, a category to be superseded by the march of science. Australian materialism, by and large, resisted eliminativism; its stand on mind-brain identities compared them with identities such as that of water with H_2O, and materialists thought that many of our 'folk' understandings of beliefs, pains, feelings and so on were, in their way, as valid and useful as much of our pre-theoretical thinking about water. Revision in the light of science was certainly possible, but elimination was a defiance of common sense.

The materialist mood was dominant for many years, but not all-conquering. Those more influenced by Wittgenstein or Oxford tended to think of the mental in linguistic or social terms; they were suspicious of ontology, especially a radically simplifying ontology, though some were attracted to the ambiguous physicalism of Donald Davidson. Epistemology, philosophy of language and 'moral psychology' were seen as more important than ontology, though the first two topics, at least, were also of concern to some of the materialists, as of course was the philosophy of science. (The Melbourne History and Philosophy of Science department was the second to be established anywhere, and produced such philosophers as Gerd Buchdahl, George Schlesinger, Hugh Lacey and Brian Ellis.) There were also critics of materialism, more sympathetic to the ontologizing mood, who held out for some residual dualism, often centred on the recalcitrance of 'qualia' to the materialist reduction (see Qualia). C.B. Martin in Adelaide was a significant influence here, and both Frank Jackson and Keith Campbell defended the ontological significance and irreducibility of the experiential, of what it is (and is 'like') to be aware of one's surroundings. Jackson even went so far as to support an unfashionable sense-data representational theory of perception in his book *Perception*. Campbell, a migrant from New Zealand, succeeded Armstrong in the Chair at Sydney, and Jackson (the son of A.C. Jackson) is a product of the University of Melbourne who was Professor at Monash University in Melbourne and then followed Jack Smart as Professor at the Australian National University. In a curious repetition of the Gibson dynastic tradition, Jackson succeeded his father in the Chair at Monash, and in 1995 became only the second Australian to give the prestigious John Locke Lectures at Oxford (his father was the first). The sympathies of Jackson and Campbell with the scientistic mood of their materialist colleagues, however, lead them to favour an epiphenomenalist account of 'qualia' so that the reality of the mental items they hope to reinstate is somewhat pallid and causally ineffectual (see Epiphenomenalism). The expatriate Australian, Brian O'Shaughnessy, developed a more positive and original, if sometimes more elusive, double aspect theory of the mind in *The Will* (1980).

More broadly, many critics and defenders of materialism in Australia tended to share a resolute realism in ontology that some believe to have climatic and geographical roots. One external influence has, however, been significant: regular visits since the 1970s of the eminent US philosopher David Lewis, himself an ultra-realist and something of an 'honorary Australian'. Some feel that Lewis' impact has been excessive; if so, it has merely reinforced a homegrown tendency. A sturdy realism is apparent even in some of the work on space and time; Graham Nerlich defends the view that space is not only absolute but a thing. What antirealism exists tends to be associated with the influence of Oxford's Michael Dummett and Harvard's Hilary Putnam (especially through the former's sometime pupil Barry Taylor) though there is a certain antirealist flavour about another interesting antipodean development, the work on alternative logics associated mostly with logicians at or connected with the Australian National University from the 1970s onwards. This was not indeed unique to Australia, but its development was vigorous and its interest in formalizing relevance and contextual relations and in allowing for 'good contradictions' was antiphonal to both the traditional Melbourne indifference to formal logic and the Andersonian scepticism about all its modern developments. It is perhaps significant that many of the practitioners of such arcana were migrants, notably Len Goddard and Graham Priest (from the UK), Bob Meyer (from the USA), and Richard Sylvan, formerly Routley (from New Zealand).

6 Besides metaphysics

Compared with achievements in philosophy of mind and metaphysics, Australians have produced little of originality in 'pure' value theory, though there have been some notable performances in applied ethics. Kurt Baier's *The Moral Point of View* (1958) and Alan Donagan's *The Theory of Morality* (1977) had genuine impact in moral theory, but both philosophers spent the major part of their careers in the USA. Similarly, J.L. Mackie's 'error theory' in his book *Ethics: Inventing Right and Wrong* has been much discussed, if seldom endorsed, but it was published long after he had migrated to the UK. D.H. Monro and H.J. McCloskey published books discussing fundamental moral theory, but they were basically intelligent refurbishings of, respectively, subjectivism and intuitionism. McCloskey's most influential work is found in his stern critiques of liberal political theory and his trenchant criticisms of utilitarianism. One notable feature of Australian moral philosophy in the latter part of the twentieth century has been the prevalence of utilitarianism. This parallels the dominance of materialism and the yearning after a 'scientific world view', although most of the materialists were indifferent to ethics, one of them famously sneering that 'ethics is for girls' - a delicious combination of sexism and philistinism. The materialist Jack Smart has, however, been a strong advocate of a rather traditional utilitarianism, and the theory fits the no-nonsense mood of much Australian metaphysics. In political philosophy, Philip Pettit and Robert Goodin also keep consequentialist and utilitarian flags flying (though C.L. Ten and Robert Young are notable non-utilitarians). In addition, Australia's best known moral philosopher, Peter Singer, has been an enthusiastic utilitarian and has deployed the theory in his various writings and activities in applied ethics. Singer's books, especially *Animal Liberation* and *Practical Ethics*, are among the few by professional philosophers in the latter part of the twentieth century to be widely read well beyond academic circles.

Singer has dedicated much of his work to an attack upon the pre-eminence traditionally given to human interests and values in moral thinking. Developing strands present in classical utilitarianism, especially in Bentham, and drawing on aspects of the thought of R.M. Hare, Singer has denounced any predominant ethical attachment to the value of human concerns as 'speciesist' on analogy with the sins of racism and sexism. Critics have found the analogy unpersuasive, and the Singer programme allows, in any event, that greater significance can be given the human where it can be shown that particular humans have endowments that deserve better treatment than particular animals or other living organisms. None the less, Singer's denunciation of speciesism has a radical and disturbing bite when it comes to the treatment of the very young, the very old and the seriously incapacitated, since many of these currently lack the endowments that differentiate most humans from animals. Such endowments (predominantly associated with the exercise of rationality) are those that Singer, along with many other philosophers, treat as definitive of a species-transcendent category of 'person'. Since only such persons are allowed a 'right to life' (though other beings should not be treated cruelly or callously without good reason) it follows that there is nothing even presumptively wrong with infanticide, and many other non-voluntary homicides. Consistent with this outlook, Singer has attacked any commitment to 'the sanctity of human life' as the outmoded vestige of religion and superstition.

If philosophy in Australia was marked until the 1920s, and to some degree beyond this, by a desire to reconcile religion with the rise of science and to show the continued relevance to ethics of a religious sense of life, there remains less sign of such concern at the end of the twentieth century. Not only is the mood of 'scientistic' metaphysics and the tenor of utilitarian thought generally hostile, or at best indifferent, to religious outlooks, but there has been little substantial positive work in the philosophy of religion. Several prominent Australian philosophers in the last quarter of the century have been committed Christians, and there is a burgeoning interest in certain Asian philosophies with strong religious components. But, apart from Max Charlesworth's founding of a journal in the philosophy of religion, *Sophia*, there has been little to indicate a revival of this project. Indeed, most of the contributions to *Sophia* come from overseas. Atheists have been more evident. C.B. Martin's *Religious Belief* (1959) accompanied a flurry of local interest in the 1950s in the problem of evil, mainly produced by atheists such as Mackie and McCloskey, but despite the high levels of religious attachment in the wider community (as attested by numerous surveys) academic philosophers in Australia have shown little recent interest in defending the religious outlook. There are some signs that this may be changing.

Something like the religious impulse does have a presence in philosophical engagement with problems of the environment, another area of applied philosophy which has had notable currency in Australia. A good deal of writing in this area invokes utilitarian or other more traditional considerations, but that which appeals to the

intrinsic value of the natural order often honours an ideal of respect and even reverence for nature that has distinct religious echoes. The metaphysical impulses of Australian philosophy are well to the fore in much writing on the environment, and appeals to scientific understanding are common enough, but the scientific models invoked are different from those beloved of the materialists. Freya Mathews' *The Ecological Self* (1991), for instance, presents a holistic metaphysics as the basis for appropriate environmental attitudes. There are points of connection between contemporary Australian environmental philosophy and attitudes to nature characteristic of the aboriginal inhabitants of the continent. Whether the aboriginal peoples had a philosophy of the environment is a moot point, but their religious cosmologies and ethical practices foreshadow some of the values and outlooks that many 'green' environmental philosophers seek to defend and systematize (see Ecological philosophy).

7 Feminism and other tendencies

Environmental and bioethical philosophers alike have an engaged dimension to their work, as do other Australian philosophers who work on questions that connect with different areas of public policy. But one of the most interesting developments of a style of philosophy allied to an activist programme has been the energetic expansion of feminist approaches to philosophy in the last quarter of the twentieth century. Genevieve Lloyd's *The Man of Reason* had a key role in this, and was also influential well beyond Australia. Lloyd, who became the first woman philosophy professor in Australia when appointed to the Chair at the University of New South Wales in 1987, has gone on to explore philosophical interests that are unusual in Australia. Like John Passmore, but in a different way, she works on the history of philosophical ideas and the connections between philosophy and literature. Partly because of the influence of feminism there are significant numbers of women in academic philosophy in Australia at the end of the twentieth century, where for most of its history there were very few. Virtually all women doing philosophy in Australia think of themselves as feminists, and most see their work in philosophy as having feminist aspects. There are complex questions here, not of course restricted to the Australian scene, about the sense in which there can be such a thing as a feminist philosophy (or a Catholic philosophy or a black philosophy) since there are constraints internal to any form of inquiry, whether it be mathematics, physics or philosophy, that rightly resist merely external 'doctrinal' imposition. This complexity, and debate about it, is complemented by another: the way in which feminist thinking is informed and often divided by different philosophical traditions, particularly those rather clumsily characterized as analytic and continental.

It is worth remarking in this connection, that most Australian philosophy departments approaching the twenty-first century are heavily 'analytic' in orientation, though there is usually some continental presence. The small Deakin University department and the General Philosophy department at Sydney University have long been much more continental in orientation, and the Faculties department at the Australian National University has been unique in forging an equal and amicable partnership between the traditions. The General department at Sydney resulted from an acrimonious division in the Sydney Department in 1973. The causes of the split were political and personal rather than philosophical: the Vietnam war and the upsurge of radicalism generated throughout the world in the late 1960s, plus a good deal of personal antagonism. The smaller half of the split, the Traditional and Modern department, included the Professor and Head, David Armstrong and his friend, David Stove, both tough-minded philosophers and dedicated anti-communists who supported the war and much else that the majority of the department opposed. The General Department initially had an assortment of radical feminists, Marxists and 'continentalists', but also contained several mainstream philosophers. Over the years, the General department became less political, and evolved into one specializing in continental philosophy, with a strongly historical approach to philosophy. But although continental philosophy had more presence in Australia at the end of the twentieth century, and there have been useful writings in a commentary style, there has been little distinguished creative philosophy produced. Of course, the terms 'analytic' and 'continental' can be distorting - a very interesting book in mixed mode is Max Deutscher's *Subjecting and Objecting* (1983) which acknowledges an equal debt to Gilbert Ryle and Jean Paul Sartre, and contains some pungent criticisms of materialism as a form of totalitarianism.

Some of the most distinctive features of Australian philosophy may have been partly a product of the country's isolation and its small academic population. Intellectual fashions were slow to arrive from overseas and, once transplanted, had plenty of time to develop their peculiar, local shapes. Schools like Andersonianism, Melbourne Wittgensteinianism and Australian Materialism flourished in a climate where you did not need to convert many people to have a dominant school. More recently, technology has considerably reduced the isolation, especially in

the world of ideas, and the philosophical community in Australia has greatly expanded from the small elite of the 1950s. Contemporary Australian work in philosophy remains very impressive, but it is much less distinctive than it once was; like philosophy elsewhere in the English-speaking world, it has acquired the polished professionalism that is part of a relatively homogenized international product. There remain some specifics of style that set it apart to some degree: on the positive side, a penchant for the direct and argumentative, and a widespread preference for a *modus vivendi* between the plain and the elaborately technical; more negatively, a certain tendency towards conformism, a sense that there are 'respectable' views behind which ranks should be closed to the exclusion of outsiders. This tendency was present in the old days and has curiously survived their passing. Maybe it reflects something in the complex national psyche. Australians romanticize rebels and outsiders, exalt egalitarianism, practice tolerance, and praise straight talking, but we have a strong need to belong and we find oddballs and deviants disturbing. A distinguished British philosopher visiting Australia in the late 1960s is said to have remarked of Australian philosophers (perhaps under the pressure of not being treated as deferentially as he had expected): 'They're either mad or boring.' If we read 'competent' for 'boring' and 'visionary' or 'unusual' for 'mad', then perhaps Australians should hope for more madness into the twenty-first century.

See also: Alexander, S.; Anderson, J.; Armstrong, D.M.; Passmore, J.; Smart, J.J.C.; Identity theory of mind; Materialism in the philosophy of mind; Scientific realism and antirealism

C.A.J. COADY

References and further reading

Alexander, S. (1920) *Space, Time and Deity*, London: Macmillan, 2 vols.(Gifford lectures by Australian expatriate philosopher with considerable impact on the development of Australian philosophy.)

Anderson, J. (1962) *Studies in Empirical Philosophy*, Sydney: Angus & Robertson.(A collection of Anderson's major articles and the primary source for understanding his thought.)

Armstrong, D. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul. (A crucial document for the development of 'Australian' materialism.)

Armstrong, D. (1984) 'Self-Profile', in R.J. Bogdan (ed.) *David Armstrong*, Dordrecht: Reidel. (An often amusing and elegantly presented short account of his own philosophical and intellectual development and the Australian and international influences upon it.)

Baier, K. (1958) *The Moral Point of View*, Ithaca, NY: Cornell University Press.(Vigorously defends an objective approach to ethics that treats moral rules as those that are for the good of everyone alike and would be apparent from the point of view of 'an independent, unbiased, impartial, objective, dispassionate, disinterested observer'.)

Baker, A.J. (1986) *Australian Realism: the Systematic Realism of John Anderson*, Cambridge: Cambridge University Press.(A thorough exposition and defence of Anderson's philosophy by a former student and colleague. It also contains an introduction by the Oxford philosopher, Anthony Quinton, placing Anderson in a broader context and speculating on the nature of his achievements.)

Brown, R. (1988) 'Recent Australian Work in Philosophy', *Canadian Journal of Philosophy* 18 (3): 545-78. (Challenges Sylvan's claims –- see below - about the existence and need for a regional philosophy, and gives a review of the contemporary scene in Australian philosophy.)

Brown, R. and Rollins, C.D. (1969) *Contemporary Philosophy in Australia*, London: Allen & Unwin.(Articles on sundry philosophical topics by Australians, with an introduction by Alan Donagan on the state of Australian philosophy and its background.)

Deutscher, M. (1983) *Subjecting and Objecting*, St Lucia: Queensland University Press. (Interesting and unusual employment of analytic and continental resources to discuss traditional philosophical problems.)

Donagan, A. (1977) *The Theory of Morality*, Chicago, IL: University of Chicago Press.(An ambitious attempt to give a classical rational account of morality drawing upon essential elements in the Judaeo-Christian tradition.)

Donagan, B. (1993) 'Alan Donagan: A Memoir', *Ethics* 104 (1): 148-53.(A historian wife's reflections on her distinguished husband's philosophical career. It includes perceptive comments on the Melbourne department's influences on Donagan and others.)

Eddy, W.C.H. (1961) *Orr*, Brisbane: Jacaranda Press.(A dedicated Orr supporter's account of the dismissal of Sydney Sparkes Orr as Professor of Philosophy at the University of Tasmania in 1956 and his subsequent tribulations. In what was probably the most lurid public scandal involving a philosopher anywhere, Orr was dismissed in controversial circumstances for sexual misconduct with a student. The Orr case divided Australian intellectuals while titillating the public, produced an eight-year staff association 'black ban' on filling the

Chair, and culminated in Orr accepting a settlement from the University on his deathbed in 1966.)

Grave, S.A. (1984) *A History of Philosophy in Australia*, St. Lucia: University of Queensland Press.(The definitive work on this subject by a former Professor of Philosophy at the University of Western Australia, Grave's finely researched book is comprehensive, thoughtful, and particularly valuable on the period up to 1970 or so.)

Jackson, F. (1977) *Perception: A Representative Theory*, Cambridge: Cambridge University Press.(Aims to defend a sophisticated 'sense-data' theory against standard objections.)

Lloyd, G. (1984) *The Man of Reason*, London: Methuen, 2nd edn, 1993.(Reviews the ways in which the idea of reason has developed in the philosophical tradition so as to make it appear the peculiar province of men.)

Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin. (Deflationary account of the status of ethics, claiming that moral language incorporates a pervasive mistake about the objectivity of moral discourse.)

Martin, C.B. (1959) *Religious Belief*, Ithaca, NY: Cornell University Press.(A wide-ranging and subtle critique of the philosophical underpinnings of religious belief. Historically important as one of the earliest comprehensive attempts to apply the new analytical techniques to philosophical theology.)

Mathews, F. (1991) *The Ecological Self*, London: Routledge.(Highly metaphysical treatment of environmental issues.)

Mitchell, W. (1933) *The Place of Minds in the World*, London: Macmillan.(Originally given as the Gifford Lectures at the Unversity of Aberdeen, 1924-6.)

O'Shaughnessy, B. (1980) *The Will: A Dual Aspect Theory*, Cambridge: Cambridge University Press, 2 vols. (Subtle and complex defence of a double-aspect theory of volition.)

Paul, G.A. (1936) 'Is there a Problem about Sense-Data? ', *Proceedings of the Aristotelian Society*, supplementary vol. 25: 61-77.(One of Paul's few publications, but a very influential piece which argues that central questions about sense-data should be construed linguistically rather than ontologically. The Wittgenstein influence is clear, but Paul's style foreshadows the Oxford philosophy of ordinary language.)

Pybus, C. (1993) *Gross Moral Turpitude: The Orr Case Reconsidered*, Port Melbourne: William Heinemann Australia.(An absorbing revisionist account of the Orr case, described in Eddy (1961), from a feminist - and somewhat patriotically Tasmanian - perspective that casts severe doubts upon the pro-Orr stance of most of the Australian philosophical community. Pybus has unearthed much new material - including information about how the undistinguished Orr was appointed in 1952 ahead of J.L. Mackie, Kurt Baier and H.P. Grice - and raises interesting questions about the professional ethic appropriate to sexual relations between university academic staff and students.)

Singer, P. (1976) *Animal Liberation: A New Ethics for our Treatment of Animals*, London: Cape, 2nd edn, 1990. (The bible of animal liberation, this book contains Singer's arguments against the outlook of 'speciesism', and more specifically attacks a variety of abuses of animals that should be repugnant even to many who do not accept his view that there is no moral basis for any preferential treatment of one's own species.)

Singer, P. (1979) *Practical Ethics*, *Cambridge: Cambridge University Press*, 2nd edn, 1993.(A more directly philosophical book, and something of a best-seller in those terms. Singer presents, with fluency and directness, an uncompromising case for a utilitarian solution to a range of issues in applied ethics.)

Sylvan, R. (1985) 'Prospects for Regional Philosophies in Australasia', *Australasian Journal of Philosophy* 63 (2): 188-204.(Sylvan argues that a distinctively regional philosophy is emerging in Australia and that this should be encouraged. He cites naturalism, especially its materialist version, and Australian environmentalism as especially distinctive.)

Toulmin, S. (1993) 'Alan Donagan and Melbourne Philosophy', *Ethics* 104 (1): 143-7. (Toulmin reflects, partly from personal experience, on the philosophical climate at the University of Melbourne in the 1940s, with special reference to the influence of Wittgenstein.)

Authority

The notion of authority has two main senses: expertise and the right to rule. To have authority in matters of belief (to be 'an authority') is to have theoretical authority*; to have authority over action (to be 'in authority') is to have* practical authority*. Both senses involve the subordination of an individual's judgment or will to that of another person in a way that is binding, independent of the particular content of what that person says or requires. If a person's authority is recognized then it is effective or* de facto *authority; if it is justified then it is* de jure *authority. The latter is the primary notion, for* de jure *authority is what* de facto *authorities claim and what they are believed to have. Authority thus differs from effective power, but also from justified power, which may involve no subordination of judgment. In many cases, however, practical authority is justified only if it is also effective.*

Political authority involves a claim to the obedience of its subjects. Attempts to justify it have always been at the core of political philosophy. These include both instrumental arguments appealing to the expertise of rulers or to their capacity to promote social cooperation, and non-instrumental arguments resting on ideas such as consent or communal feeling. Whether any of these succeed in justifying the comprehensive authority that modern states claim is greatly disputed..

1 The nature and forms of authority

What do theoretical and practical authority have in common? George Cornewall Lewis aptly defined the acceptance of theoretical authority as: 'the principle of adopting the belief of others, on a matter of opinion, without reference to the particular grounds on which that belief may rest' (1849). Authority over action shares this feature. Hobbes (§7) put it this way: 'command is where a man saith, *Doe this* or *Doe not this*, without expecting any other reason than the Will of him that sayes it' (1651). Authority thus offers what H.L.A. Hart (1982) calls 'content-independent' reasons for belief or action: the opinion of an expert or the directive of a ruler, parent, manager and so on is itself meant to be taken as a reason, irrespective of the grounds on which that opinion or directive is based.

A second feature of authority is that its requirements bind its alleged subjects: authority has an exigency that advice or requests lack. Some political realists offer a reductionist account of this, claiming that the exigency ultimately amounts to the credible threat of force. Others suggest it is the justification for force. Neither, however, is plausible, for force is necessary only as a backup when authority fails in its primary aim of directing behaviour. A better view is suggested in John Locke's remark: 'All private judgment of every particular member being excluded, the community comes to be umpire, by settled standing rules; indifferent and the same to all parties' (1690). The key here is neither force nor its justifications but rather the exclusion of private judgment (see Locke, J. §10). Joseph Raz (1986) has influentially argued that authoritative reasons pre-empt or exclude other reasons the subject might have. Authoritative reasons are thus 'exclusionary': they do not outweigh competing considerations but instead make them irrelevant. (It is, however, unclear how sharp a distinction between outweighing and excluding may be sustained.)

The normative character of authority as content-independent and binding has misled some into thinking that genuine authority must always be justified and that 'illegitimate authority' is a contradiction in terms, but this is wrong. 'Illegitimate authority' is on a par with 'invalid proof': while it purports to be or is accepted as valid, in fact it is not. While 'legitimate authority' is thus not superfluous as a term, it is still the sense generally implicit in the term 'authority': authorities are believed to be legitimate. This is fully compatible with a sociological analysis of the structure and origins of such beliefs, like Weber's three ideal types of rational-bureaucratic, traditional or customary, and charismatic legitimacy (the latter resting on the personal virtue or knowledge of an extraordinary leader) (Weber 1922).

2 Political authority

Three important features distinguish the authority of the state. First, the state claims and enforces compulsory jurisdiction over everyone within its territory. Some, such as the English jurist John Austin, consider that the authority of a sovereign state must also be legally absolute (Austin 1832). In fact, however, while the authority of most states is supreme - that is, binding on all other authorities in its territory - it is normally subject to

constitutional limits, although these may consist of the absence of a power to legislate in a certain manner or field rather than the presence of an enforceable duty not to do so.

Second, even when limited the authority of the state is always serious, and it regulates the most vital interests of everyone within its territory. Political authority, as Locke (1690) states, includes the power to regulate life, liberty and property, and to impose any penalties up to and including death.

Finally, political authority claims to impose obligations of obedience. Hobbes doubted this, maintaining that political authority consists only of a liberty to coerce. Friedrich Hayek goes further and states: ‘The ideal type of law… provides merely additional information to be taken into account in the decision of the actor’ (1960). However, neither of these views can explain the nature of political authority as it is seen by state officials and by many subjects. The state is not indifferent to its citizens either obeying its directives or disobeying while suffering the prescribed penalty; it prefers obedience. It is true that not all authoritative government action purports to create obligations - the state also grants permissions, makes declarations of status, and so on - but creating obligations is central to its other activities.

3 Justifications

Anarchists deny that political authority is justifiable at all (see Anarchism §§1-2). Some of their arguments hold that governments are inherently unjust. Others focus on the subordination that is at the heart of authority relations. William Godwin (§3) and, recently, Robert Paul Wolff argue that there is a deep conflict between personal autonomy and the authority of the state, for the subordination of judgment always violates our most urgent duty, namely, to do what we at that moment think right (Godwin 1793; Wolff 1976). This is, however, an odd view of autonomy, for we regularly subject ourselves to the will of others, for example by making plans and promises without which we would have even less control over our own lives. The real issue is thus not the subordination of judgment as such, but its subordination to the state.

Some religions claim that all political authority is ordained by God. Divine right theories, however, merely defer the problem: what justifies the authority of scripture or of divine commands? Humanistic theories, on the other hand, begin on the footing that practical authorities must in some way benefit people, especially their putative subjects. These benefits may be instrumental or non-instrumental.

Joseph Raz (1986) argues that if authorities are to serve their subjects then justifications for authority must normally show that the alleged subjects are likely to comply with the relevant ultimate reasons if they take the directives as binding and attempt to follow them. They must be more likely to do so than they would if they tried to follow the ultimate reasons directly by relying on their own judgment. On this view, all instrumental justifications for authority are indirect.

Arguments of this form underlie a variety of theories, including consequentialist, natural duty and hypothetical contract arguments, all of which purport to establish a need for authority (see Legitimacy §1; Obligation, political §1). Two main considerations are offered. The first appeals to the expertise of the rulers; they have knowledge or wisdom that their subjects lack. The second rests on the purported capacity of the state to sustain valuable forms of social cooperation.

The force of these arguments, however, is limited. It is doubtful whether there is relevant expertise in all moral matters (for example, about distributive justice); nor is there any prospect of agreement about who has it, and no a priori reason to suppose it always lies with the government. Moreover, considerations of equality and democracy argue against rule by platonic guardians.

The argument stemming from social cooperation is more plausible. Owing to short-sightedness, limited information and self-defeating motives, we often fail to coordinate action for the common good. By marking certain actions as obligatory and by providing incentives for compliance the state helps overcome this, solving coordination problems, providing public goods and so on. Objections to such arguments are twofold. First, authority systems are themselves the product of social cooperation, so necessarily there is a problem getting them started. Second, there is risk of overkill; even if authority is sometimes needed to secure cooperation, the comprehensive authority of a state may not be. This marks a general difficulty with all indirect, instrumental arguments. It is not sufficient to show that some set of binding rules would be better than individuals following

their own judgment; it needs to be shown that the actual set of state-sponsored rules is better than the feasible alternative rules. Moreover, as anarchists argue, to rely generally on the state may weaken the capacity to find better modes of cooperation.

Non-instrumental justifications attempt instead to show that authority relations have an inherent value or are constituent parts or logical results of other valued relations. This may hold if authority is consented to, or if the acceptance of authority expresses gratitude to the state or feelings of communal solidarity. However, actual consent is rarely given and does not always bind, and while obedience may be one way of expressing those attitudes and relations, it is not the only way nor always the best way. In a multicultural world of identity politics, it is not even clear that the state is an appropriate locus of these feelings. Perhaps Harold Laski (1919) and the pluralists were right to state that we are each the nodal point of many competing allegiances, none with inherent priority.

LESLIE GREEN

References and further reading

Austin, J. (1832) *The Province of Jurisprudence Determined*, ed. H.L.A. Hart, London: Weidenfeld & Nicolson, 1954.(See §2.)

Godwin, W. (1793) *Enquiry Concerning Political Justice*, ed. I. Kramnick, Harmondsworth: Penguin, 1976.(See §3.)

Green, L. (1988) *The Authority of the State*, Oxford: Oxford University Press.(An examination of the nature of and purported justifications for political authority.)

Hart, H.L.A. (1982) 'Commands and Authoritative Legal Reasons', in *Essays on Bentham*, Oxford: Oxford University Press.(See §1.)

Hayek, F.A. (1960) *The Constitution of Liberty*, Chicago, IL: University of Chicago Press, 148-61. (See §2.)

Hobbes, T. (1651) *Leviathan*, ed. C.B. Macpherson, Harmondsworth: Penguin, 1968.(See §§1-2.)

Laski, H.J. (1919) *Authority in the Modern State*, New Haven, CT: Yale University Press.(See §3.)

Lewis, G.C. (1849) *An Essay on the Influence of Authority in Matters of Opinion*, London: John W. Parker.(A rare volume but with a good treatment of theoretical authority and the connections between this and political authority; see §1.)

Locke, J. (1690) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 3rd edn, 1988.(See §1.)

Lukes, S. (1979) 'Power and Authority', in T. Bottomore and R. Nisbet (eds) *A History of Sociological Analysis*, London: Heinemann, 633-76.(A blend of historical and conceptual analysis.)

McMahon, C. (1994) *Authority and Democracy: A General Theory of Government and Management*, Princeton, NJ: Princeton University Press.(A careful examination of managerial authority in non-governmental organizations, situating it in broader democratic theory.)

Raz, J. (1986) *The Morality of Freedom*, Oxford: Oxford University Press, 21-105.(One of the most influential accounts of authority; see §§1, 3.)

Raz, J. (ed.) (1990) *Authority*, Oxford: Blackwell.(A collection of essays providing a good survey of the issues, with a helpful introduction.)

Simmons, A.J. (1979) *Moral Principles and Political Obligations*, Princeton, NJ: Princeton University Press.(A clear introduction to, and criticism of, theories of the duty to obey the law.)

Weber, M. (1922) *Economy and Society*, vol. 1, ed. G. Roth and C. Wittich, New York: Bedminster Press, 1968. (See §1.)

Wolff, R.P. (1976) *In Defense of Anarchism*, New York: Harper & Row, 2nd edn.(See §3.)

Autonomy, ethical

The core idea of autonomy is that of sovereignty over oneself, self-governance or self-determination: an agent or political entity is autonomous if it is self-governing or self-determining. The ancient Greeks applied the term to city-states. In the modern period, the concept was extended to persons, in particular by Kant, who gave autonomy a central place in philosophical discourse. Kant argued for the autonomy of rational agents by arguing that moral principles, which authoritatively limit how we may act, originate in the exercise of reason. They are thus laws that we give to ourselves, and Kant thought that rational agents are bound only to self-given laws. Much contemporary discussion has focused on the somewhat different topic of personal autonomy, and autonomy continues to be an important value in contemporary liberalism and in ethical theory.

It is important to distinguish different senses of autonomy because of variation in how the concept is used. Self-governance or self-determination appears to require some control over the desires and values that move one to action, and some such control is provided by the capacity to subject them to rational scrutiny. Thus, autonomy is often understood as the capacity to critically assess one's basic desires and values, and to act on those that one endorses on reflection. In other contexts, autonomy is understood as a right, for example as the right to act on one's own judgment about matters affecting one's life, without interference by others. The term is also sometimes used in connection with ethics itself, to refer to the thesis that ethical claims cannot be reduced to nonethical claims.

1 Autonomy in the early modern period

Central to autonomy is the notion of self-governance. The earliest use of the term was by ancient Greek writers [*autonomia* (n.), *autonomos* (adj.) from *autos* - self, and *nomos* - law], for whom it was a political concept. A city-state had autonomy if it had authority to enact its own laws and to manage its own affairs, independently of any foreign power. Moral and political philosophy of the early modern period allowed autonomy to be a basic feature of persons, even where the term was not used explicitly. Many seventeenth- and eighteenth-century rationalists held that our moral capacities create the capacity for self-governance. They believed that reason or conscience gives all individuals the ability to discover objective truths either about duty, or about the good, without external guidance through revelation, the church or political authority (see Moral knowledge §3). In addition, they held that agents could be motivated to act simply by their knowledge of moral norms, without the imposition of rewards or punishment (see Moral motivation). Since the content of such norms is desire-independent, the capacity to be motivated by one's moral knowledge gives one control over one's desires. On the assumption that reason or conscience occupies an authoritative role within the self, action guided by one's moral knowledge is self-determined.

Social contract theories of this period conceived of human beings as autonomous in a somewhat different sense, by attributing to them an 'original sovereignty' over themselves. The social contract theorists regarded individuals as by nature free, equal and independent, with authority to regulate their own conduct. It follows that one individual can become subject to the authority of another only through an act of consent or agreement, and that legitimate state powers are those it would be rational to agree to. In Locke's view (1690), individuals in the state of nature are bound by a law of nature requiring them both to preserve themselves, and others in so far as they are able. But all individuals in the state of nature have equal authority to judge and to punish violations of the law of nature. This is the basis of political authority. Locke argues that rational individuals in the state of nature would agree among themselves to entrust this authority to enforce the law of nature to a central power, for the limited purposes of preserving one's own life and liberty. Thus, his theory derives political authority from one's original sovereignty over oneself, which in addition limits the legitimate uses of state power (see Contractarianism in ethics and political philosophy; Locke, J. §10).

Rousseau takes the idea of self-governance a step further, claiming that sovereignty - the power to enact laws - resides in the collective body of all citizens, and that legitimate laws must be self-imposed: 'the people that is subject to the laws ought to be their author' (1762: 67) (see Rousseau, J.-J. §3). Rousseau argued that the freedom and independence of each citizen could be preserved only when individuals agree to submit to the 'general will' - the will of society as a whole concerning matters of common interest. The general will is expressed in laws which

preserve freedom and equality, and which are, accordingly, enacted by a political process in which all citizens participate (see General will). Thus his political ideal is a participatory democracy preserving the autonomy of each citizen, in which individuals are bound only to laws which they have a role in making, and which express each individual's will. Rousseau's famous remark that 'freedom is obedience to the law one has prescribed for oneself' (1762: 56) provided a model of autonomy for later theorists, in particular for Kant.

2 Kant's conception of the autonomy of rational agents

Two different senses of autonomy emerge in the above theories. One is the capacity to guide one's conduct by one's grasp of moral norms. The other is 'sovereignty over oneself' - a basic right to self-governance, which is the basis of further rights and standards of justification. These aspects of autonomy come together in Kant's moral theory. Kant takes the autonomy of rational agents to be the fact that they have the ability to legislate universally valid principles (the principles of morality) through their will, and are bound only by principles that originate in their own reason. This is a capacity for self-determination because action guided by such principles follows self-given laws. Finally, the autonomy of agents is the basis of specific duties, which are requirements of respect for autonomy (see Kantian ethics; Kant, I. §9).

Kant writes: 'Autonomy of the will is the property that the will has of being a law to itself (independently of any property of the objects of volition)' ([1785] 1903: 440). One implication of this claim is 'that man [humanity] is subject only to his own, yet universal, legislation and that he is bound only to act in accordance with his own will' ([1785] 1903: 432). Autonomy has both a negative and a positive aspect. Rational agents are not bound by any principles that do not originate in the exercise of reason, or by any sources of authority external to reason. That is, considerations such as desire, convention, the will of God, or uncritically accepted authorities are not inherently reason-giving; they provide reasons for action only if the agent takes them to. Positively, rational agency is the source of authoritative normative principles. A deliberative process expressing the basic features of rational agency generates authoritative principles of action, in particular, the basic principles of morality (see Practical reason §§2-3). Kant's conception of autonomy reflects the political origin of the concept in that he regards rational agents as a kind of sovereign authority who can give universal law through their willing. So understood, autonomy presupposes certain deliberative and motivational capacities: the capacities to assess critically any proposed reason for action and to be motivated by reasons that are independent of one's desires and presently held values.

Comparisons with earlier moral theories bring out the full dimensions of Kant's conception. In contrast to empiricists, Kant recognized reasons and principles that are not desire-based. Like rationalist theorists, Kant thought that morality consists of necessary desire-independent principles. But he did not think that such principles represent an objective order of values existing independently of rational volition. Rational deliberation is not the discovery of objective principles, but rather the process by which they are generated. Thus, Kant held that agents who act on moral principles act from self-given laws, since these are principles that originate in the use of one's reason. Kant argued that earlier attempts to ground obligation in, for example, the will of God, an objective order of values or obligations, or features of human psychology, all lead to 'heteronomy of the will', because such theories would subject rational agents to an authority external to reason.

Kant thought that the autonomy of rational agents is consistent with moral objectivity. The categorical imperative is the basic principle to which any agent with autonomy is committed, and generates universally valid principles of action. Agents express their autonomy by deliberating and acting from the categorical imperative because it is a rational procedure through which any agent can arrive at principles that other agents can acknowledge as authoritative; that is, it enables one to give law through one's will. Though Kant conceived of autonomy primarily as a feature of persons, individual actions guided by the categorical imperative are autonomous in a derivative sense, since they are guided by self-given principles. By contrast, actions determined by desires or uncritically accepted values are guided by principles external to reason.

Kant's conception of agents as having autonomy determines the content of his normative theory. Kant held that the capacity to give moral principles through one's will is the basis of human dignity, and that all agents are committed to valuing the exercise of their rational capacities. These assumptions underlie his general principle that rational nature should be treated as an end, and never as a means only (see Respect for persons §2). The application of this principle leads to requirements not to undermine the exercise of rational agency through, for example, coercion, deception, paternalistic interference, or contempt and ridicule, as well as positive duties to

support the exercise of rational agency through mutual aid and perfection of one's talents. Thus, the autonomy of rational agents is the basis of specific duties to respect the exercise of rational agency.

3 Contemporary accounts of the nature of personal autonomy

Some contemporary writers concerned with the nature of personal autonomy have distinguished between autonomy as a capacity for self-governance or self-determination, as the actual condition of self-governance, as a personal ideal, and as a right or a social value. (see Feinberg 1986; Hill 1991). Though these notions are defined and related in differing ways, the capacity for self-governance may be viewed as the basis of the other notions. An agent who exercises this capacity effectively is in the condition of autonomy. The personal ideal or virtue of autonomy would be the set of character traits associated with the complete development of the capacity, viewed as a component of the admirable or fulfilling life (see Eudaimonia; Virtues and vices §3). Autonomy viewed as a right is the general right to the unimpeded exercise of this capacity in matters concerning one's own life. The next two sections will focus on contemporary accounts of the capacity for autonomy, and on autonomy as a value.

Self-governance intuitively requires that one have control over the psychological states that determine one's actions, and that the desires and values that guide one's choices be 'truly one's own'. One might hold that one's desires and values are one's own when one identifies with or endorses them as a result of critical reflection. Thus, autonomy may be defined as the capacity to assess critically one's basic desires and values, to revise them if one judges that there is reason to, and to act on those that one identifies with or endorses upon critical reflection. This capacity enables one to take responsibility for one's basic desires and values, and to shape the direction of one's life.

This definition relies on a distinction between first-order desires (desires for certain objects or activities) and higher-order reflection about one's first-order desires, and makes the capacity for higher-order critical reflection central to autonomy. Higher-order values may be viewed as judgments about which of one's first-order desires and values one wants to be moved by; or alternatively, as evaluative judgments as to whether goals or activities towards which one is inclined are really worth pursuing, whether certain of one's values or character traits are good, and so on. Take, for example, a person who cares greatly about material goods. Agents with autonomy can ask whether they want to be moved by materialistic desires to the extent they currently are, or whether it is good to care about material goods to that extent. Moreover, such agents can modify their values and conduct if they see reason to. Note that autonomy should not be interpreted as a capacity to create one's desires and values *ex nihilo*. That would imply, implausibly, that no one possesses autonomy, since individuals are deeply influenced by social and cultural factors. Autonomy need not require that one be the ultimate source of one's desires and values, but only that one have the capacity to assess them critically.

The capacity for critical higher-order reflection is a normal feature of rational agency which may be developed to differing degrees by different agents, and which can be interfered with by such factors as psychological disorder, external manipulation of an agent's deliberative processes, and social conditioning. A full treatment of personal autonomy must spell out the kinds of influence that undermine autonomy. It must also provide some account of the process of critical reflection. Here several questions arise. What guides the formation of one's higher-order values? While autonomy would seem to require some control over one's higher-order values, even they are unavoidably shaped by upbringing, culture, and other factors beyond one's control. Which forms of social and cultural influence at this level are consistent with autonomy? Some of these problems are illustrated by the following example. Imagine a woman who has been socialized to believe that women ought to be subordinate to men and who has internalized this social role. On conscious reflection she judges that she is leading a good life, identifying with her subservient social role and endorsing the character traits of submissiveness that enable her to fulfil the expectations imposed upon women in her society. This sort of example suggests that a capacity to apply critical reflection to one's higher-order values - which appears lacking in this person - is necessary for autonomy. The question of when one's values are 'truly one's own' arises for higher-order values, just as it does for first-order desires and values. Should we then conclude that the higher-order values leading to endorsement of one's first-order desires and values should themselves be accepted or endorsed through critical reflection? If so, how do we avoid ascending levels of higher-order evaluation continuing *ad infinitum*? The worry is that either critical reflection must continue indefinitely, or the point at which it terminates is arbitrary.

Most theorists would agree that a capacity for ongoing critical reflection about one's higher-order values is

necessary for autonomy; that there are empirical limitations on the level to which individuals can carry higher-order reflection, and that contingency is inevitable in the higher-order values that determine one's first order identifications; and that higher-order reflection can be terminated in non-arbitrary ways. But there are different views about the resolution of higher-order reflection, which represent different views about the critical reflection needed for autonomy. One might hold that critical higher-order reflection, in principle, permits individuals to call into question any deeply held value. While never totally free of social and cultural influence, individuals can, either on their own or in dialogue with others, achieve sufficient distance from their basic values to view them critically. Eventually, one must decide which values to accept. But such a decision is not arbitrary when underwritten by the judgment that further reflection will confirm one's present decision.

Other theorists have argued that the process of critical reflection must principally satisfy 'conditions of procedural independence': it must be free from such obstructions as internal obstacles (such as psychological disturbance), manipulation, coercion, and unacceptable social conditioning (Dworkin 1988). The burden on such a theory is to spell out these conditions of procedural independence. This approach places no constraints on the substantive values that an autonomous agent could accept. Any values are consistent with autonomy, as long as one accepts them on one's own. A different approach holds that autonomy requires the capacity to modify one's values in the light of objective or fully reasonable values. This theorist would take the failure of an agent's reflective values to satisfy certain objective standards (for example, if one's values are clearly detrimental to individual fulfilment, or are morally flawed) as indicating a lack of autonomy.

4 Normative dimensions of autonomy

In much contemporary ethical and political theory, autonomy plays both a foundational and a normative role, as illustrated by the structure of Kant's ethical theory. The capacity for autonomy, as described in the previous section, is so central to agency that respect for persons is plausibly construed as respect for the exercise of this capacity. Thus in much contemporary ethical theory, a view of persons as having autonomy is the basis of a general right to act on one's own judgment of what one has most reason to do, from which one can derive more specific principles, such as duties not to interfere with a person's freedom, duties to refrain from coercion, manipulation, paternalistic interference, and so on, as well as positive duties to support autonomy. Many theorists hold that the wrongness of certain kinds of actions may be explained by noting how they interfere with, or deprive individuals of their autonomy. It is important to note that what autonomy-centred theories value is the opportunity to guide one's actions by one's exercise of the capacity for critical reflection. This is considerably more complex than simply acting on one's own desires.

Similarly, autonomy is a central value in modern liberalism and democratic theory (see Liberalism §3). A conception of persons as having autonomy figures in liberal conceptions of justice, the liberal principle that the state should not promote any particular conception of the good, and in arguments against state paternalism (see Justice §5; Paternalism §3). It is the basis of those rights and liberties that are the institutional means necessary for individuals to exercise autonomy - including liberty of conscience, rights of free expression, liberty to develop one's own plan of life (within the limits of justice), and rights of political participation. While autonomy is more prominent in the Kantian and social contract traditions, it can also play a role in utilitarian theories (see Utilitarianism). Mill, for example, argues in *On Liberty* (1859) that the exercise of judgment, choice and responsibility, and the development of individuality are essential to individual fulfilment. This conception of happiness permits him to argue on utilitarian grounds against state paternalism and for a set of civil liberties that allow individuals to exercise autonomy. His view, in short, is that institutionalized liberties protecting individual autonomy will promote general happiness (see Mill, J.S. §§11-12).

The importance of autonomy to contemporary liberalism is illustrated by features of Rawls' theory of justice (1993). One of the organizing ideas of Rawls' theory is a conception of persons as possessing two 'moral powers', which they have a fundamental interest in exercising. These are a capacity to develop, revise and pursue a conception of the good and a capacity for a sense of justice. Possession of these moral powers is the basis of moral equality, entitling individuals to equal consideration. The principles of justice which Rawls derives set out the political and social conditions that guarantee to each individual opportunity to exercise these moral powers effectively, with the principle of equal liberty listing basic individual liberties, and the principle of equal opportunity together with the difference principle (that inequalities should benefit the least advantaged) spelling

out further social and material conditions. In this way, a conception of persons as having autonomy plays a role in determining the content of a conception of justice. A further aspect of Rawls' theory directly parallels Kant's conception of autonomy. Kant held that rational agents are bound only by principles that originate in their own will. Similarly, in Rawls' theory, the final standards of justice are those that persons would autonomously choose for themselves. Rawls supports his conception of justice by arguing that the two principles would be chosen in the 'original position' - a conceptual device representing a fair choice by free and equal persons, with an interest in securing the exercise of their two moral powers. This particular aspect of autonomy is seen in the fact that the agents in the original position are not bound by any antecedently given moral principles. They are free to choose whatever conception of justice will best advance their basic interest in the exercise of their moral powers, and whatever would result from such a choice determines what is just (see Rawls, J.).

The value placed on autonomy is not unchallenged. Many theorists have argued that the emphasis on autonomy ignores or fails to leave room for other important values, such as the value of ties and attachments to others, loyalty to groups, respect for tradition, or the value of community. Versions of this objection hold that within autonomy-centred theories, personal commitments and attachments and obligations to others become purely voluntary, and thus cannot be definitive of the self. But surely, the objection continues, we can have commitments, attachments and obligations that are essential to our identity (see Morality and identity §4). In reply, it is sufficient to say that autonomy need not be conceived in a way that makes it inconsistent with such values, and that viewing autonomy as central does not entail viewing it as the sole value (see Moral pluralism). Nothing in the conceptions of autonomy surveyed above precludes agents from deciding as a result of critical reflection to take on binding obligations or to affirm attachments to others, or from concluding that certain commitments and ties are inescapable because constitutive of who they are. We may conclude that autonomy is deeply embedded in the modern notion of the person and is an important modern value.

5 The autonomy of ethics

The autonomy of ethics uses 'autonomy' in a sense not directly related to the autonomy of agents. It is the thesis that ethical principles or claims cannot be reduced to, or explained in terms of, statements containing no normative terms (see Logic of ethical discourse §§2-4). 'Normative terms', such as 'ought', 'good' and 'bad', 'right' and 'wrong', 'just' and 'unjust', are those which state the intrinsic value of, approve or recommend an action or state of affairs, or state that there is reason to perform certain actions or pursue certain ends. (They contrast with 'descriptive terms', which describe or report purely factual states of affairs.) The autonomy of ethics, for example, can allow claims about what is right to be explained in terms of the good (or vice versa), but holds that any such explanation of an ethical claim must use some normative language. Theorists who accept the autonomy of ethics can include intuitionists, for whom ethical statements are claims about mind-independent ethical properties, such as rightness or goodness (see Intuitionism in ethics); noncognitivists, who take ethical claims to express attitudes of approval or recommendation (see Analytic ethics); or practical reason theorists, who hold that ethical claims are based on principles of practical reason - principles which we accept in virtue of being rational (see Practical reason). Common to these theories is the view that the normative aspect of ethical discourse is not capturable in language devoid of normative terms.

More broadly, the autonomy of ethics includes the independence of ethical theory both from other areas of philosophy and from the natural and behavioural sciences. This does not mean that ethical theory can ignore these fields. Rather, it means that questions about the content and validity of fundamental moral principles do not depend on answers to questions of, for example, metaphysics, epistemology, or philosophy of mind, and that they are not dictated by the results of empirical inquiry. Ethical theory, as concerned with normative questions, has its own distinctive subject matter and tools of inquiry. To give one example, modern Kantians have thought that the moral principles that we accept should not be restricted by a picture of motivation derived solely from empirical inquiry, rejecting the idea that our motivational capacities can be described without any reference to normative principles. If we can be motivated by our acceptance of moral principles, then certain motivational states can be described only by reference to the reasoning that underlies them. Moral theory is then required for insight into certain of our motivational capacities (see Rawls 1975, 1993; Nagel 1970).

See also: Freedom and liberty; Free Will; Self-control

ANDREWS REATH

References and further reading

Christman, J. (ed.) (1989) *The Inner Citadel: Essays on Individual Autonomy*, Oxford: Oxford University Press. (A collection of contemporary essays. Good introduction to contemporary discussions of personal autonomy, with an extensive bibliography.)

Dworkin, G. (1988) *The Theory and Practice of Autonomy*, Cambridge: Cambridge University Press.(Accessible treatment of the nature and value of autonomy.)

Feinberg, J. (1986) 'Autonomy', in *Harm to Self*, Oxford and New York: Oxford University Press, ch. 18; repr. in J. Christman (ed.) *The Inner Citadel: Essays on Individual Autonomy*, Oxford: Oxford University Press, 1989. (A clear analysis of different meanings of autonomy.)

Frankfurt, H. (1988) *The Importance of What We Care About*, Cambridge: Cambridge University Press.(Contains influential essays on freedom of the will of importance to the nature of personal autonomy.)

Hill, T.E., Jr (1991) *Autonomy and Self-Respect*, Cambridge: Cambridge University Press, chaps 3, 4, 7. (Develops an accessible account of autonomy and of its normative significance.)

Hill, T.E., Jr (1992) *Dignity and Self-Respect*, Ithaca, NY: Cornell University Press, chaps 1, 5-7.(Develops an accessible account of Kant's views about autonomy.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten, in Kants gesammelte Schriften*, ed. Königlichen Preußischen Akademie der Wissenschaften, Berlin: Reimer, vol. 4, 1903; trans. J.W. Ellington, *Grounding for the Metaphysics of Morals* , Indianapolis, IN: Hackett Publishing Company, 1993.(References made to this work in the entry give the page number from the 1903 Berlin Akademie volume; these page numbers are included in the Ellington translation. A classic work of moral theory which bases morality on autonomy.)

Korsgaard, C. (1996) *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.(Essays on Kant's moral theory and on contemporary moral theory relevant to several aspects of autonomy.)

Locke, J. (1690) 'The Second Treatise of Government', in *Two Treatises on Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1988, 269-78, 323-33.(Classic social contract theory which bases the legitimacy of political authority on the consent of citizens.)

Mill, J.S. (1859) *On Liberty*, ed.E. Rappaport, Indianapolis, IN: Hackett Publishing Company, 1978, esp. 53-71. (Classic account of individual civil liberties developed within a utilitarian framework.)

Nagel, T. (1970) *The Possibility of Altruism*, Princeton, NJ: Princeton University Press.(Develops a contemporary Kantian view of motivation.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Seminal contemporary work arguing that principles of justice can be viewed as the result of an autonomous choice of free and equal persons.)

Rawls, J. (1975) 'The Independence of Moral Theory', *Proceedings and Addresses of the American Philosophical Association* 48: 5-22.(Argues for the independence of moral theory from other areas of philosophy.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press, paperback edition with new material, 1996.(Lecture II discusses the role of autonomy in the theory of justice.)

Reath, A. (1994) 'Legislating the Moral Law', *Nous* 28: 435-64.(Scholarly account of Kant's concept of autonomy.)

Rousseau, J.-J. (1762) *Du Contrat social*, trans. J.R. Masters, ed. R.D. Masters, *On the Social Contract*, New York: St Martin's Press, 1978.(Classic social contract theory that gives a central role to the notions of autonomy and self-legislation.)

Scanlon, T.M. (1972) 'A Theory of Freedom of Expression', *Philosophy and Public Affairs* 1: 204-06.(Develops a theory of free expression based on a conception of autonomy.)

Schneewind, J.B. (1986) 'The Use of Autonomy in Ethical Theory', in T.C. Heller, M. Sosna and D.E. Wellberry (eds) *Reconstructing Individualism*, Stanford, CA: Stanford University Press, 64-75.(Discussion of autonomy in Butler, Kant and Rawls.)

Schneewind, J.B. (1991) 'Natural Law, Skepticism and Methods of Ethics', *Journal of the History of Ideas* 52: 289-308.(Shows how eighteenth-century empiricist moral theories developed a morality of autonomy that dispenses with the need for moral authorities of various sorts.)

Schneewind, J.B. (1993) 'Modern Moral Philosophy: From Beginning to End?', in P. Cook (ed.) *Philosophical Imagination and Cultural Memory*, Durham, NC: Duke University Press.(Accessible overview of the history of early modern ethics, stressing the emergence of autonomy as a central notion.)

Avenarius, Richard (1843-96)

Richard Avenarius, a German philosopher, is known as a proponent of 'empiriocriticism' and the principle of economy of thinking. Empiriocriticism is a modern version of empiricism which attempts to restore the concept of the natural world and 'pure experience' through the elimination of 'introjection', understood as an insertion of redundant or distorted ideas and images into the objects of our knowledge. Avenarius traced back the origin of introjection to the cultural stages dominated by magic and mythology, yet his criticism applied also to traditional philosophy and science. His position is usually classified as a version of positivism, closely resembling the empiricist doctrine of Ernst Mach. Although his influence on some members of the Vienna Circle, especially Moritz Schlick, was considerable, the impact of his contribution has been hampered by his idiosyncratic use of language, especially in his masterpiece Kritik der reinen Erfahrung (The Critique of Pure Experience) (1888-90).

Avenarius was born in Paris, where his father owned a German publishing house. However, he spent his formative years in Leipzig and Berlin and received his Ph.D. in 1868 for a dissertation concerned with aspects of Spinoza's philosophy. Avenarius' *Habilitationsschrift*, presented in Leipzig under the title *Philosophie als Denken der Welt gemäss dem Prinzip des kleinsten Kraftmasses (Philosophy as Thinking of the World in Accordance with the Principle of the Least Amount of Energy Expenditure)* (1876), contributed to his fame and facilitated his appointment as Professor for Inductive Philosophy in Zurich, Switzerland. Avenarius held this position until his untimely death in 1896. During this period he published his significant works: *Kritik der reinen Erfahrung (The Critique of Pure Experience)* (1888-90) and *Der menschliche Weltbegriff (The Human Concept of the World)* (1891). In 1877 Avenarius founded an important philosophical journal, *Vierteljahrsschrift für wissenschaftliche Philosophie (A Quarterly for Scientific Philosophy)*, which he edited until his death. Perhaps the most famous of his disciples isJoseph Petzoldt, who edited posthumous editions of Avenarius' works.

Avenarius combines his version of empiriocriticism with the concept of the natural world which he wanted to restore. His method comes very close to Husserl's phenomenological technique of pure description (see Husserl, E.). 'Findings' (phenomena) are the point of departure for such pure description, and they encompass environmental components including the bodies of other persons. To distinguish human persons from other environmental components, Avenarius employs a linguistically founded hypothesis amounting to the interpretation of certain movements and sounds of fellow humans as statements. Accordingly, the endorsement of the natural concept of the world assumes that fellow human beings are not mere thoughtless and emotionless mechanisms. This principle offers then an opportunity to treat the empirical (verifiable) content of human knowledge through an appropriate analysis of meaningful statement-components - a very modern method indeed. Avenarius underscores another basic interpretative principle of empiriocriticism: the 'democratic' supposition of 'human equality'. This principle enables him to make short-cut inferences (by analogy) to similar experiences in other persons, as long as their statement-contents are meaningfully interpreted by the observer. It is through the contents of statements made by other persons that the so-called *E*-values are recognized; the *E*-values are then divided into 'elements' (for example, 'green', 'warm' and so on) and 'characters' (such as 'pleasant', 'ugly' and so on). It seems that Avenarius anticipates here the distinction between cognitive and noncognitive (emotive) constituents of statements, so dear to the later positivists. The *E*-values ultimately depend on the environmental components - on the so-called *R*-values (for example, trees, rivers, the sun). These are the things within the natural environment which a person (called *M*) finds as 'given' - the 'independent factors' of the first kind. Among the environmental components one must further distinguish the physical bodies of other persons, for whom Avenarius reserves the symbol *T*. His empiriocritical triad *MRT* symbolizes the presence and relationship between these three factors. The dependence of the *E*-values on the *R*-values is mediated by the brain and the entire central nervous system of the person in question. This empirical system *C* is then regarded as the 'independent factor' of the second kind. Impressed by the optimistic results of neurophysiology in his time, Avenarius strongly emphasizes this system *C*. He assumes that there is an obvious dependence relation between an *E*-value (say, of the person *M*) and some change in the person's system *C*, symbolized by ΔC. The complete account of these dependencies yields the following:

$$R \rightarrow M \rightarrow C_M \rightarrow E_M$$

We read this: a certain environmental component *R* is affecting the body of the person *M* and further, through the

proper nervous channels, the brain of the person *M* who then produces the respective *E*-value (expressed in statements). In *Kritik der reinen Erfahrung* Avenarius offers the following explanation: 'In every case, in which *E* is regarded as dependent upon *R*, *E* is taken as immediately dependent upon a *change* of *C*' (1888-90: 1, 80). Yet the indirect dependence relation of E_M upon *R* is treated by Avenarius as a purely logical functional relation, in accord with his positivistic rejection of causality. Avenarius undertook a grandiose programme of biological reductionism, supported by his principle of economy of energy expenditure: the limited amount of energy possessed by human organisms should not be wasted beyond necessity. Actively or passively, the human organism (governed by its system *C*) participates in a multitude of processes which are arranged into time-dependent 'vital series' of two basic types: independent and dependent. Avenarius characterizes the vital preservation maximum of an organism as the equality of its partial systematic factors:

$$\Sigma f(R) + \Sigma f(S) = 0$$

In this equation, the sum of environmental factors *R* equals the (negatively prefixed) sum of the bodily factors *S*. Variations in these interactions are responsible for a variety of 'vital differences' expressed in inequalities. The organism tends to reduce or eliminate such differences if they weaken its preservation value. Allegedly, the system *C* of the organism is setting up an idealized goal of perfect vital series and it affects variations of the actual vital series in this direction.

This biologically founded teleological doctrine is related to Avenarius' rejection of 'introjection' as an unwarranted insertion of something which distorts the purity of the original experience. Avenarius believed that the elimination of introjection will lead to restoration or 'restitution' of the natural concept of the world, with beneficial results for human knowledge and action. His ideals of 'pure experience' demanded the rejection of any philosophical dualism and a positivistic revision of traditional metaphysical beliefs. Thus he got close to the 'neutral monism' of Ernst Mach (§4), William James (§6) or the early Bertrand Russell (§13) (see Neutral monism). In the Preface to *Kritik der reinen Erfahrung*,Avenarius programmatically declared his intention to go directly to things as such ('*an die Sachen anzuknüpfen*'), thus foreshadowing the later goals of Husserl and Heidegger.

See also: Empiricism; Phenomenalism

AUGUSTIN RISKA

List of works

Avenarius, R. (1868) *Über die beiden ersten Phasen des Spinozischen Pantheismus und das Verhältniss des zweiten zur dritten Phase (On Both First Phases of Spinoza's Pantheism and the Relationship between the Second and the Third Phase)*, Leipzig: E. Avenarius; repr. 1980.(Avenarius' Ph.D. dissertation, concerned with the order and publication time of Spinoza's earlier writings.)

Avenarius, R. (1876) *Philosophie als Denken der Welt gemäss dem Prinzip des kleinsten Kraftmasses: Prolegomena zu einer Kritik der reinen Erfahrung (Philosophy as Thinking of the World in Accordance with the Principle of the Least Amount of Energy Expenditure)*, Leipzig: O.R. Reisland and Berlin: J. Guttentag, 1903, 1917.(A short book, yet widely read and cited.)

Avenarius, R. (1888-90) *Kritik der reinen Erfahrung (The Critique of Pure Experience)*, Leipzig: O.R. Reisland; 2nd and 3rd edn, ed. J. Petzoldt; repr. Ann Arbor, MI: University Microfilms International, 1980, 2 vols. (Avenarius' masterpiece, almost 800 pages long, very difficult to read. Volume 1 deals with the independent vital series, while volume 2 treats dependent vital series.)

Avenarius, R. (1891) *Der menschliche Weltbegriff (The Human Concept of the World)*, Leipzig: O.R. Reisland; 2nd edn, ed. J. Petzoldt, 1905; later editions, 1912, 1917. (English translation in progress by A. Riska; much more accessible presentation of Avenarius' 'metaphysics', as he himself called this relatively short work.)

Avenarius, R.(1913) *Zeitschrift für positivistische Philosophie (Journal of Positivist Philosophy)* 1 (4): 34-54. (Avenarius' own summary of his four major works. These summaries were originally published in *Vierteljahrsschrift für wissenschaftliche Philosophie (A Quarterly for Scientific Philosophy)* between 1888 and 1892.)

References and further reading

Numerous references to Avenarius' philosophy can be found in the literature relevant to Ernst Mach and logical positivism as well as to Malinowski's functional anthropology. A contemporary systematic treatment of

Avenarius' work is sorely lacking.

Arens, K. (1989) *Structures of Knowing: Psychologies of the Nineteenth Century*, Dordrecht: Kluwer.(Avenarius is here treated as one among the German conceptual psychologists of the nineteenth century.)

Bush, U.T. (1905) *Avenarius and the Standpoint of Pure Experience*, New York: Science Press.(Based on a Ph.D. thesis at Columbia University, New York; one of very few English publications on Avenarius' work.)

Carstanjen, F. (1896) 'Nachruf an Richard Avenarius' (In Memory of Richard Avenarius), *Vierteljahrsschrift für wissenschaftliche Philosophie* 20 (4): 361-91.(An obituary; a comprehensive biography and assessment ofAvenarius' philosophical contribution.)

Lenin, V.I. (1909) *Materialism and Empirio-Criticism*, Moscow: Zveno Publishing House; trans. Moscow: Foreign Language Publishing House, 1952.(While attacking his political opponents, Lenin criticized Mach's and Avenarius' empiriocriticism.)

Mach, E. (1900) *Die Analyse der Empfindungen und das Verhältniss des Physischen zum Psychischen*, 2nd expanded edn, Jena: Fischer; trans. *The Analysis of Sensations*, New York: Dover, 1959.(The section on pages 35-42, 'Mein Verhältniss zu R. Avenarius' (My Relationship to R. Avenarius), clarifiesMach's critical appreciation of Avenarius' positions.)

Raab, F. (1912) *Die Philosophie von R. Avenarius: systematische Darstellung und immanente Kritik (The Philosophy of R. Avenarius)*, Leipzig: F. Meiner.(A systematic exposition and an immanent criticism; a typical, dated attempt to provide a wider view of Avenarius' philosophy.)

Schlick, M. (1925) *Allgemeine Erkenntnislehre*, Berlin: Springer, 2nd edn; trans. A.E. Blumberg, *General Theory of Knowledge*, New York: Springer, 1974.(Many references to Avenarius show Schlick's indebtedness to Avenarius' ideas.)

Swoboda, W. and Hobek, F. (1975) 'Richard Avenarius: mit einer Bibliographie' (Richard Avenarius: with a bibliography), *Conceptus* 9 (26): 25-39.(A useful bibliographical source.)

Averroism

'Averroism', 'radical Aristotelianism' and 'heterodox Aristotelianism' are nineteenth- and twentieth-century labels for a late thirteenth-century movement among Parisian philosophers whose views were not easily reconcilable with Christian doctrine. The three most important points of difference were the individual immortality of human intellectual souls, the attainability of happiness in this life and the eternity of the world. An 'Averroist' or 'Radical Aristotelian' would hold that philosophy leads to the conclusions that there is only one intellect shared by all humans, that happiness is attainable in earthly life and that the world has no temporal beginning or end. Averroists have generally been credited with a 'theory of double truth', according to which there is an irreconcilable clash between truths of faith and truths arrived at by means of reason. Averroism has often been assigned the role of a dangerous line of thought, against which Thomas Aquinas opposed his synthesis of faith and reason. The term 'Averroism' is also used more broadly to characterize Western thought from the thirteenth through sixteenth centuries which was influenced by Averroes, and/or some philosophers' self-proclaimed allegiance to Averroes.

1 Rise of the terminology

Averroes, the twelfth-century Muslim commentator on Aristotle, exercised a strong influence on Latin scholastics from about 1230 onwards (see Ibn Rushd). Around 1270, the derogatory term *Averroistae* ([too ardent] followers of Averroes) began to be used, principally to characterize adherents of the view that there is only one shared human intellect. In 1277 the Bishop of Paris, Etienne Tempier, accused unnamed masters of arts of the University of Paris of paying more attention to heathen philosophers than to Christian revelation, and of behaving as if there were two truths, one of philosophy and another of faith. The theory of one shared intellect was among the 219 theses the bishop condemned. A generation later, Ramon Llull launched a series of attacks on university philosophers whom he saw as continuators of the lines of thought condemned in 1277, and used the term *Averroistae* to describe these philosophers.

Based on this medieval use of *Averroistae*, the term 'Averroism' was introduced in nineteenth-century historiography of philosophy. Averroism was conceived of as a movement of thirteenth-century thinkers faithful to Averroes, proclaiming that the same proposition could have different truth values in philosophy and theology, so that there was an unbridgeable inconsistency between philosophy and faith. Averroism was cast as a sinister force (a precursor of modern atheism), valiantly combatted by Albert the Great and Thomas Aquinas, notably in the latter's *De unitate intellectus (On the Unicity of Intellect)* and *De aeternitate mundi (On the Eternity of the World)*. Twentieth-century historiography came to identify three main currents in late thirteenth century philosophy: first, Augustinianism, mainly represented by Franciscan thinkers, which combatted the growing influence of Aristotelian philosophy (see Augustinianism); second, Averroism, which took a radically Aristotelian approach to philosophical problems, even though this must lead to conflict with Christian faith; and third, the current led by Albert the Great and Aquinas, who produced a synthesis of Aristotle and Christian faith (see Aristotelianism, medieval).

Early in the twentieth century, it was commonly assumed by historians that almost all the 219 theses condemned in 1277 were of Averroist provenance, and since there is medieval evidence that the main targets of the condemnation were Boethius of Dacia and Siger of Brabant, they were thought to have embraced most of the 'heterodox' opinions. Siger, who is known to have engaged in university politics, began to be seen as the leader of an Averroist party at the University of Paris in the 1260s and 1270s.

Subsequent research has undermined the foundations for the historiographical scheme in which a thirteenth-century 'Averroism' belongs. First, the majority of the theses condemned in 1277 were not inspired by Averroes. Moreover, scholars often apply the label 'Averroist' also to later philosophers who were influenced by Averroes or continued the views of writers such as Siger of Brabant: examples include the Parisian masters Ferrandus Hispanus in the late thirteenth century and John of Jandun in the early fourteenth century, and to a long list of Italian writers from Gentile da Cingoli in the 1290s and Angelo d'Arrezzo in the early fourteenth century to Agustino Nifo in the early sixteenth century. Some of these writers did indeed defend Averroes' views whenever possible, but such loyalty towards Averroes had not been a characteristic of the men who were condemned in

1277.

When these facts became apparent to historians of philosophy, they began to replace ‘Averroism’ with ‘radical Aristotelianism’ or ‘heterodox Aristotelianism’ as the name of this supposed thirteenth-century school of thought. However, so many historical misunderstandings and ideologically motivated judgments cling to all these labels that they are, in the 1990s, being abandoned. Yet, there are some interesting problems that these labels were meant to help explain, and which still have an important place in medieval philosophy.

2 ‘Averroist’ doctrine

In the later half of the thirteenth century, there was a common conviction that some philosophical tenets were inconsistent with Christian doctrine as standardly understood. ‘Philosophical’ in this connection means Aristotelian on the then standard interpretation of Aristotle, which leaned heavily on Arabic works including the writings of Avicenna and Averroes, and the *Liber de causis* (see Aristotelianism, medieval; Liber de causis).

Three issues stood in the foreground in the conflict between reason and faith: first, whether all humans share a common intellect (monopsychism); second, whether happiness is attainable in this life; and third, whether the world had a temporal beginning.

Monopsychism: It was generally accepted that the intellect (that is, the intellective soul) has both an active component, ‘the agent intellect’, which forms universal concepts on the basis of particular pieces of information provided by the senses, and a passive component, usually called the ‘possible intellect’, which is the initially blank wax tablet on which the active component leaves its imprints in the form of concepts and knowledge acquired. The question was, are the agent and possible intellects genuinely different, and if not, does each human being have its own intellect, or is there only one for all to share?

There was a tradition of considering only the agent intellect to be an extra-human separate substance, responsible for humans’ shared conceptual apparatus; in this view, the individuality of each person’s possible intellect explains why we do not share all our thoughts. One version of this view was held by Roger Bacon. By the 1260s, however, this radical separation of the agent and possible intellects had become rather old-fashioned. The main disputants of the time agreed that the two intellects are one substance, but they disagreed about whether that substance is extra-human. Averroes, as he was commonly understood after about 1250, taught that the intellect is a single impersonal substance with which individual souls enter into contact via their mental representations (phantasms) of extramental things; the intellect uses the phantasms as a basis for abstraction. Modern historiography has applied the term ‘monopsychism’ to this doctrine, which was attacked by Aquinas in his *De unitate intellectus*.

Monopsychism allows for the irrational part of a human soul to be destroyed on death without this affecting the intellect. Like the old assumption of a separate agent intellect, it also accounts for the ability of human beings to share knowledge; but it offers no convincing answer to the objection that if this is the case, then no thought belongs to one individual rather than another. During at least one phase of his career, Siger of Brabant accepted monopsychism, but believed that it was possible to save some private thought for the individual by making the operation of the intellect in a particular human depend on representations (*intentiones imaginatae*) with an origin in sensation unaided by intellect. In his somewhat obscure attempts to explain how the individual ‘plugs into’ (*continuatur*) the supra-individual intellect, Siger relied heavily on Averroes.

Contemporaries were alert to the Averroistic theory’s inability to explain how all humans can share an intellect without sharing all thoughts. However, to medieval thinkers the gravest objection against monopsychism was that it left no individual rational soul to carry responsibility for a deceased person’s acts. Nor was it easy to see how an immaterial intellect could fail to be eternal, which was contrary to Christian doctrine that God creates new souls every day and that they are in principle perishable (God could annihilate a soul if he wished). Nonetheless, for the next couple of centuries most philosophers seem to have held that monopsychism was one of the few rationally defensible views about the nature of the intellect, while standard Church doctrine continued to require the intellect to be both the form of the body and capable of separate, individual, existence. The issue was still very much alive in 1513 when the Fifth Lateran Council explicitly condemned the view that the intellective soul is either mortal or only one for all people, and explicitly asserted that it is the form of the human body, immortal, and as many in number as are the bodies into which it is infused.

Happiness in this life: Around 1260-70, masters from the Faculty of Arts at the University of Paris often expressed a great optimism about the attainability of happiness in this life. Their views strongly resemble those of ancient Neoplatonism, but the strongest impetus came from Arabic philosophy rather than directly from ancient sources (see Neoplatonism; Neoplatonism in Islamic philosophy). The way to happiness was thought to consist in an intellectual ascent to the contemplation of ever higher beings, culminating in contemplation of the First Cause and the (temporary) union of one’s possible intellect with the source of intellectual understanding, the agent intellect; In this tradition, the agent intellect was thought to be a separate substance and not identical with God. Such a state of intellectual bliss was held to be the fullest actualization of a person *qua* human, that is, a rational being.

This line of thought would seem to permit the construction of a naturalistic ethics with no need for either divine revelation or an individual life after death in order that human beings may reach their ultimate goal and happiness. Boethius of Dacia did indeed hold that a natural philosopher must deny the resurrection of the dead, and this was to be a common view for a long time. However, there is little evidence that anyone really wanted to abolish the belief in a second life. The philosophers’ point was simply that while it is known through revelation that supranaturally there will be such a life, a claim to that effect cannot be incorporated into a consistent theory of nature (see Natural philosophy, medieval).

Eternity of the world: Before the 1260s there had been some attempts to interpret Aristotle as if he accepted a temporal beginning of the world. Perhaps the first such attempt was made by William of Conches in the twelfth century. However, as Robert Grosseteste noted in the 1230s, such attempts had failed and the common assumption became that Aristotelian philosophy did in fact require the world to have existed for an unlimited time, partly because creation out of nothing could not be subsumed under any of the Aristotelian modes of change. Change implies the prior existence of something to be the subject of change, and so creation cannot be a species of change. By the 1270s, it was commonly recognized that the concept of creation out of nothing was consistent if not confused with change, but it also became a common conviction that this would wreak havoc on the natural sciences if incorporated among their concepts. The supposed Averroists were thought to have simply denied the temporal beginning of the world (see Eternity of the world, medieval views of).

Double truth: Averroists have been credited with a theory of double truth, occasioned by the fact that when medieval thinkers saw a conflict between philosophy (science) and the teaching of the church, they could not simply reject Church doctrine. Instead, they could hold that philosophers had misinterpreted some of the information obtained by natural means (as Aquinas held, for example), or they could hold that there was no way to detect any error in the derivation of the philosophical thesis, so that the only way out of the impasse consisted in rejecting the thesis on the authority of faith (as did Siger). Alternatively, they could try to explain how the assumption of a first cause makes it reasonable to expect that there are truths which no scientific theory can possibly account for; Boethius of Dacia, who distinguished the conditional truth of a scientific theorem from absolute truth, took that line. A fourth way, asserting that the same proposition can be absolutely true philosophically and also absolutely true theologically, had very few followers, if any at all, but has sometimes been imputed to the ‘Averroists’.

To understand how this misconception should arise, one should remember that most philosophers of the thirteenth to sixteenth centuries were masters of arts; it was their job to teach a non-Christian (Aristotelian) philosophy in a Christian society, and so they were caught in the contradiction between reason and faith. Guidelines on how to deal with this dilemma were given in a decision by the Faculty of Arts at Paris in 1272: henceforward, any master dealing with a problem that touched both philosophy and faith was bound by oath to solve it a way that was not contrary to faith. The result was a widespread use of the technique of first providing a philosophical solution and then adding one ‘according to the truth of faith’. For some twentieth-century scholars use of this technique has sufficed to stamp a philosopher as an adherent of a theory of double truth.

See also: Aristotelianism, medieval; Boethius of Dacia; Ibn Rushd; Islamic philosophy: transmission into Western Europe; Natural philosophy, medieval §9; Siger of Brabant

STEN EBBESEN

References and further reading

Boethius of Dacia [Boethus Dacus] (*c.*1275) *Opera (Works)* ed. H. Roos *et al.* in Corpus Philosophorum

Danicorum Medii Aevi IV-IX, Copenhagen: DSL/Gad, 1969-. (Critical edition of the extant works, in Latin. Volume IX, *Sophismata*, has not yet appeared.)

Boethius of Dacia [Boethus Dacus] (*c.*1275) *On the Supreme Good, On the Eternity of the World, On Dreams*, trans. J.F. Wippel, Medieval Sources in Translation 30, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1987.(Small philosophical treatises which forcefully present Boethius' views.)

Fioravanti, G. (1966) 'Boezio di Dacia e la storiografia sull'Averroismo', *Studi Medievali* 3a, serie 7: 283-322. (Excellent survey of the use of the term 'Averroism' in modern times, with pertinent criticism.)

Gauthier, R.A. (1983-4) 'Notes sur Siger de Brabant (I-II)' (Notes on Siger of Brabant), *Revue des sciences philosophiques et théologiques* 67: 201-32; 68: 3-49.(Rectifies a number of myths about Siger and Averroism; downplays the importance of Siger.)

Gorce, M.M. (1931) 'Averroïsme', in *Dictionnaire d'histoire et de géographie eccélsiastiques*, Paris: Letouzey et Ane, vol. 6: 1032-92.(Outdated, but representative of its time.)

Hisette, R. (1977) *Enquête sur les 219 articles condamnés à Paris le 7 mars 1277 (An Investigation of the 219 Articles Condemned at Paris, 7 March 1277)*, Philosophes Médiévaux 22, Louvain: Publications Universitaires de Louvain.(Contains the text of the condemned theses as well as a detailed study of the origin of each of them.)

Kuksiewicz, Z. (1965) *Averroisme bolonais au XIV^e siècle (Bolognese Averroism in the Fourteenth Century)*, Wroclaw: Ossolineum.(Study of Averroism in Italy.)

Kuksiewicz, Z. (1968) *De Siger de Brabant à Jacques de Plaisance. La théorie de l'intellect chez les averroistes latins des XIIIe et XIV siècles (From Siger of Brabant to James of Piacenza: The Theory of Intellect in the Work of the Latin Averroists in the Thirteenth and Fourteenth Centuries)*, Wroclaw: Ossolineum.(Valuable as a presentation of texts.)

Kuksiewicz, Z. (1982) 'Sense, Intellect, and Imagination in Albert, Thomas and Siger', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 602-22.(Survey of the debate about the intellect.)

Libera, A. de (1991) *Averroès et l'Averroïsme. Que sais-je? (Averroes and Averroism: Survey)*, Paris: Presses Universitaires de France.(Good overview of the subject.)

Ramon Llull [Raimundus Lullus] (1263-1316) *Disputatio Raimundi et Averroistae*, in H. Harada (ed.) *Opera latina VII*, Corpus Christianorum Continuatio Mediaevalis 32, Turnholt: Brepols, 1975.(One of Llull's anti-Averroist treatises.)

Mandonnet, P. (1899) *Siger de Brabant et l'Averroïsme au XIII^e siècle (Siger of Brabant and Thirteenth-Century Averroism)*, Fribourg: Librairie de l'Université.(Mandonnet's understanding of 'Averroism' was to dominate the historiography for a century.)

Renan, E. (1852) *Averroès et l'Averroïsme*, Paris: A. Durand.(Initiated the use of 'Averroism' in the historiography of philosophy.)

Siger of Brabant (*c.*1240-84) *Collected Works*, various editors, in *Philosophes Médiévaux* vols 12-14, 24, 25, Louvain: Publications Universitaires de Louvain, 1972-.(Works of Siger, who is often identified strongly with Averroism.)

Van Steenberghen, F. (1966a) *La philosophie au XIII^e siècle (Thirteenth-Century Philosophy)*, Philosophes Médiévaux 9, Louvain: Publications Universitaires de Louvain.(An influential defender of the 'radical/heterodox Aristotelianism' label.)

Van Steenberghen, F. (1966b) 'L'Averroïsme latin' (Latin Averroism), *Philosophia Conimbricensia* 1: 1-32. (Claims that full-fledged Averroism only appears with John of Jandun.)

Averroism, Jewish

Averroism was enthusiastically taken up by many Jewish philosophers and adapted in a number of ways that extended its scope beyond mere repetition of Averroes' own arguments. Jewish Averroists were particularly drawn by the potential they found in Averroism for resolving the delicate questions they faced about the relationship between philosophy and religion. The idea that both philosophy and religion are true even when they appear to produce different answers to the same question. Fascinated by the Averroistic idea that religious claims can be interpreted as popular expressions of philosophical truths, the Jewish Averroists followed up with vigour the programme of showing how to translate traditional religious statements into philosophical statements.

Many Jewish philosophers found themselves in a difficulty which they took great pains to resolve, namely, how to reconcile what they believed through faith with what they believed through reason. Averroism seems to be the solution to this problem, since it embodies a theory that explains how faith and reason are connected and makes it possiblc to be both religious and rational at the same time. It is not surprising, then, that many Jewish thinkers were attracted to this philosophical doctrine.

1 Averroes in the Jewish world

Averroes (see Ibn Rushd) came to have a significance in both Christian and Jewish settings that far exceeded his influence in the Islamic lands. There are many reasons for this. Jewish philosophers were impressed with the depth of Ibn Rushd's arguments and with his understanding of Aristotle, who was typically regarded as the paradigmatic philosophical thinker. They were undoubtedly also fascinated by the radical implications of his views. There were many translations of his works into Hebrew and transliterations of Arabic texts of his work into Hebrew script, and these extended far beyond Spain and the intellectual world of the Iberian Peninsula. Many Jewish philosophers wrote commentaries on his works and were influenced by him, in particular Crescas and Gersonides, and he set the agenda for much of Jewish philosophy. The positive references that Maimonides had made to him obviously helped his reception in the Jewish world, and his thought is often linked with that of Maimonides, since on some important topics they are clearly not far apart.

It would be wrong to think of all the Jewish philosophers who were influenced by Averroes as Averroists. Many often argued with his central ideas and criticized his conclusions, and there is no clear category of thinkers who are 'Averroists' in the sense that they accepted everything that Averroes argued. The philosophers discussed below are those who came closest to sharing Averroes' main principles and who tried to adapt them to Jewish intellectual life.

2 Isaac Albalag

Isaac Albalag came from the Pyrenees region during the second half of the thirteenth century. He was in no doubt at all concerning the merits of Averroes over his Islamic predecessors, or even over Maimonides. He translated al-Ghazali's *Maqasid al-falasifah (Intentions of the Philosophers)* into Hebrew, and argued that this book represented al-Ghazali's real views - a thesis that would have horrified the original author (see al-Ghazali §3). Albalag does accept al-Ghazali's argument that there are certain doctrines that must be accepted by religion. These are the existence of reward and punishment for our actions, the survival of the soul after the death of the body and the fact of providence whereby God watches over our actions. In his *Sefer Tikkun ha-De'ot (Setting Doctrines Right)*, he recognizes that philosophy in its Averroistic form does not think much of these religious ideas, and suggests that they are to be accepted by ordinary people who are not capable of philosophy. These ideas will enable them to achieve the highest level of well-being of which they are capable; but that must be contrasted with the sort of felicity that philosophers can achieve, since philosophers can understand far more about the nature of reality than can ordinary believers. The vital tool of understanding here is demonstrative argument, which Albalag, like Averroes, saw as the paradigmatic method of philosophy. Only philosophers are really capable of this sort of thought, and as a consequence, only philosophers can really be allowed to say that they know what is the case.

But what about all those religious texts which it is not possible to verify demonstratively, and which are based on prophecy? Albalag argues that these can be known but not necessarily demonstratively, and that this is no criticism of them. Here Albalag deviates from Averroes, since the latter criticized the claims of *kalam* (dialectical theology)

to understand religious texts, especially difficult religious texts (see Islamic theology §2). There is no point in thinking that theology can help us with such texts, Averroes argued, since it possesses no methodology that can derive a valid conclusion from premisses. Albalag tends to place the interpretations of those in the Jewish mystical tradition of the Kabbalah (see Kabbalah) in the same position as Averroes' *kalam* theologians. He holds that they are capable only of providing dubious and weak readings of Scripture. Indeed, he tends to separate philosophical and religious explanations more radically than Averroes allows. Such a separation is a common feature of Christian and Jewish Averroists (see Averroism). In Albalag's case, what we find is acceptance of both views, but the treatment of one is much less vigorous and intellectually satisfying than the other: when the literal sense of a text cannot be reconciled with a philosophical understanding of the topic, both the literal sense and the philosophical understanding have to be accepted, but in different ways. The literal sense is accepted as something which one would understand completely if one were in the position of the prophets who had originally transmitted the text. One must assume, even though one cannot see how, that this meaning is reconcilable with the philosophical meaning.

A crucial example in Jewish philosophy is the creation of the world. According to the philosophers, on Albalag's account, this doctrine must be understood to cohere with the eternity of the world. What results is a theory according to which God creates the world eternally (see Eternity of the world, medieval views of). Albalag criticizes Maimonides for claiming that Aristotle did not claim to know with certainty whether the world is eternal. In fact, Albalag believes, Aristotle had no doubts concerning the eternity of the world; Maimonides agreed with Aristotle, but did not wish to threaten the faith of the ordinary believers in Judaism. That is the reason for what Albalag sees as Maimonides' ambivalence about the demonstrability of the origins of the world. However, Maimonides should not have suggested that the eternity of the world cannot be established by reason. It can, and philosophers have no choice but to acknowledge that eternity, although they need not broadcast their views widely if they think hearing such things would upset the beliefs and practices of the naive. It is possible to accept the eternity of the world through reason and its createdness through faith, and there is no need to work out how to reconcile these two diverse positions. Albalag does not say why not; but he seems to go further than Averroes, who argues that there is one truth expressible in two different ways, whereas Albalag appears to argue that there are two truths. This takes him closer to the so-called doctrine of double truth often ascribed to the Christian Averroists in their more radical moments.

3 Joseph ibn Caspi

Joseph ibn Caspi was born in 1279 in Provence, and wrote a variety of theological and philosophical works. He was heavily influenced by Maimonides, Averroes and Abraham ibn Ezra, especially by the latter's construction of a philosophical grammar of the Hebrew language (see Ibn Ezra, A.). Caspi defended the literal sense of many passages in Scripture as accurate accounts of past events, and criticized as misleading many of Maimonides' attempts at explaining many of them as prophetic allegories. On the other hand, he shares Averroes' rather ambivalent attitude towards miracles, suggesting that there is a natural explanation for miracles, were we able to understand all aspects of the events in question. To understand a report of a miracle, it seems, we need to understand the point of view and knowledge of those actually present at the time. Prophecy also has to be interpreted in terms of the audience it is designed to impress, and if there are aspects of the event which we do not now entirely grasp, we should put this down to our distance from its occurrence and to our relative lack of knowledge of how the event was regarded at that time. Prophets are able to tell what is going to happen in the future because they are able to understand how the things they observe in the present link with what is to come. The role of the deity in this process is quite limited. Many religious statements are not capable of a truth value; their function is to move people to action and to teach them how to behave. Where philosophy and prophecy diverge, we should remember that they are different and it is hardly surprising that they do not always agree. If we really knew why prophets said what they say and why miracles take the forms they do, then we should understand how they might be reconciled; indeed, we might understand how prophecies are merely more popularly accessible expressions of philosophical truth. Since we are limited in our understanding of religious statements by our distance from their original formulation, we have to accept them as aspects of faith. We can remain confident nonetheless that such statements are in principle equivalent to philosophical truths. Unlike Maimonides and Averroes, Caspi had little sympathy with the idea that the secrets of interpretation should be restricted to the intellectual elite. His point was that they have to remain secrets, since there is no possibility now of finding out

precisely what the ancient statements originally meant, given the differences in audience, language and context (see Prophecy).

4 Moses Narboni

Moses Narboni was born in Perpignan around 1300 and died approximately sixty-two years later. He wrote many commentaries on theological and religious texts, together with some original works and several commentaries on the works of Averroes. He wrote extensively on Maimonides, and tended to criticize Maimonides using arguments drawn from Averroes, as he was one of the few philosophers of the time to recognize that Averroes was seeking to challenge the Neoplatonic metaphysics of Ibn Sina (Avicenna) which formed so important a part of Maimonides' thinking. Narboni also developed Averroes' theory of the 'active intellect' in such a way as to make it relevant to Jewish philosophy.

As human thinking becomes progressively better perfected it moves from being imaginative to becoming intellectual. Ultimately it fuses with the active intellect itself, which is the very principle of intellectual thought. As a result, the material part of us comes under the control of our thought. Using this theory, Narboni is in a position to explain miracles and prophecy as resulting from a certain sort of thinking that produces appropriate material effects, that optimally illustrate the ideas in the consciousness of the prophets but which adapt those ideas to the level of understanding of the audience the prophet has in mind. Here a kind of Neoplatonism creeps back in, for Narboni's language is clearly based on the idea of a hierarchy of intellects where each intellect is connected with an existent (see Neoplatonism in Islamic philosophy §2).

The parallelism between intellects and existents usefully echoes a similar parallelism between doctrines and acts. The idea of what is to be done results in the creation of that state of affairs, and similarly a religious doctrine has as its material aspect a particular form of practice. All of this accords nicely with the unified approach that Averroes takes to the relationship between such diverse phenomena as body and mind, the material and the spiritual, and the theoretical and the practical. The Torah, which is perfect, consists of doctrines which are true and accord with practices designed to bring about a desirable end in line with those true beliefs. Of course, it takes a remarkable individual to understand all aspects of the Torah, and only Moses fulfils that role, a role very similar to that which Averroes ascribes to Muhammad.

Narboni, like Averroes, upholds the principle of plenitude, according to which something is possible only if it is (at some time) actual; and he uses this principle to argue that in an eternal universe, if there can be a most perfect created being then (at some time) there will be. Moses fits the bill exactly. Lesser mortals will not be able to grasp perspicuously the reasons for all the doctrines in the Torah, and will have to accept these doctrines on the basis of faith alone. Narboni thinks it is bad policy to encourage ordinary believers to trouble themselves excessively with finding out the reasons for the commandments. Most people would not understand these reasons even if they were presented with them, and a fruitless search would only frustrate and undermine the faith of the seekers. The prophets are provided precisely for such people, since they are capable of representing philosophical truths in imaginative language which will impress the masses and keep them on the right path. Only those capable of philosophy will understand precisely what is actually intended, and only they should seek such understanding. We see here a reformulation of a genuine doctrine of Averroes, that there is one truth that is expressed in at least two different ways, one intellectually respectable and the other practical and effective.

5 Further consequences

Clearly, there are aspects of Jewish Averroism which are very different both from the philosophy of Averroes himself and from Christian Averroism. In the Jewish milieu, Averroes was commonly linked with Maimonides and, more surprisingly perhaps, with Abraham ibn Ezra. Some thinkers such as Narboni were interested in introducing Kabbalistic doctrines in their explication of the thought of Averroes. Narboni's discussion of providence in terms of astrological causation would have surprised Averroes and horrified Maimonides, but this is by no means a criticism of the approach of the Jewish Averroists to Averroes himself. They employed his thought, and what they considered could be derived from his thought, creatively and combined it with other relevant philosophical and theological theses to produce a novel account of the issues then of concern to the philosophical community.

There were a large number of other thinkers whose work is largely based upon Averroes but who have not been discussed here in detail. The work of Joseph ibn Waqar and Moses ibn Crispin, for example, provides evidence of considerable discussion on Averroistic themes within the Jewish community. It is only with the onset of the Renaissance and the last major Averroist thinker in the community, Elijah Delmedigo, that the passionate interest in Averroes started to decline. Averroes was long seen as the first commentator on Aristotle, and the relative decline in interest in Aristotle was matched with decreasing concern for the thought of Averroes.

One of the contributions of Jewish Averroism is its way of tackling the distinction between religious and philosophical truths. The argument that the pursuit of philosophy is not only permitted by religion but is even necessary for the intelligent adherent comes straight from Averroes himself in his *Fasl al-maqal (Decisive Treatise)*. The warning against trying to prove the truth of religion through philosophy was taken very seriously by the Jewish Averroists, for two reasons. It is a category mistake to try to explain through philosophy what is capable only of religious explanation. Also, the discovery that the truths of Judaism cannot be established philosophically might lead to disbelief or scepticism. This gives rise to an interesting question. If it is inappropriate to refer to the rational basis of a religion as the justification of that religion, what reason is there to prefer one religion to another? This was a lively issue during the Middle Ages, since there were strenuous efforts by Islam and Christianity to convert Jews, and equally strenuous efforts by Jewish thinkers to resist such pressure. The Jewish Averroists had no doubt of the superiority of their religion over its competitors, and they argued that Christianity in particular involves the acceptance of self-contradictory notions.

We must distinguish between those ideas which are in themselves possible and can be actualized through the miraculous intervention of the deity, and those ideas which even God could not bring about, since they are impossible and only an imperfect deity could wish to bring them about. This was the approach which Jewish philosophers took to the notion of God becoming man, the Incarnation, which they regarded as obviously an imperfection, along with a whole range of other crucial Christian doctrines such as the Trinity, transubstantiation and the Virgin Birth (see Incarnation and Christology; Trinity). A whole range of logical problems were discovered in these doctrines and contrasted with the bases of Judaism which, it was argued, is an acceptably rational faith. At least Judaism does not involve the acceptance of self-contradictory notions, although it is true that some of the stories of miracles offend against our understanding of natural necessity. As Aristotelians, the Jewish Averroists might have been expected to have been more wary of allowing for miraculous interventions in nature, since Averroes seems to have regarded such events as just as impossible as logical self contradictions. This is where the use of Maimonides and Abraham ibn Ezra in combination with Averroes proves so useful. Maimonides was used to suggest that the accounts of miracles need not be taken to be literally true (by philosophical readers), and Ibn Ezra was used to show that we are so distant from the time of the miraculous events that we do not really know precisely what the narratives about them are supposed to mean. We do not even know what the language used then really meant, nor how astrological forces brought about change in the sublunar world.

Jewish Averroism should not, then, be seen as a slavish adaptation of Averroes to issues of interest in Jewish philosophy. Some of Averroes' main theses were combined with the opinions of quite disparate philosophers and out of this heady mixture a rich variety of arguments emerges. These arguments are generally more than merely eclectic and they involve a sustained treatment of the logic of the relationship between religious and philosophical language.

See also: Aristotelianism, medieval; Aristotelianism in Islamic philosophy; Averroism; Ibn Rushd

OLIVER LEAMAN

References and further reading

Bland, K. (1981) *The Epistle on the Possibility of Conjunction with the Active Intellect by Ibn Rushd with the Commentary of Moses Narboni*, New York: The Jewish Theological Seminary of America.(An edition and translation of an important commentary.)

Golb, M. (1956-7) 'The Hebrew Translation of Averroes' Fasl al-Maqal', *Proceedings of the American Academy for Jewish Research* 25: 91-113; 26: 41-64.(Useful exploration of the popularity of the text in the Jewish world.)

Hayoun, M.-R. (1982) 'L'êpitre du libre-arbitre de Moise de Narbonne' (Moses Narboni's Letter on Free Will),

Revue des études juives 141: 139-67.(An edition and translation of Narboni's *Ma'amar ha-Behira (Treatise on Free Will)*.)

Hayoun, M.-R. (1986) *Moshe Narboni*, Tübingen: Mohr.(Edition and translation of Part 1 of Narboni's commentary on Maimonides' *Guide to the Perplexed*.)

Lasker, D. (1980) 'Averroistic trends in Jewish-Christian Polemics in the Late Middle Ages', *Speculum* 55: 294-304.(The relevance of Averroes for interreligious debate.)

Leaman, O. (1988) *Averroes and His Philosophy*, Oxford: Clarendon Press.(A general survey of Averroes and Averroism.)

Leaman, O. (1996) 'Jewish Averroism', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 769-80.(Good introductory article.)

Mesch, B. (1982) 'Principles of Judaism in Maimonides and Joseph ibn Caspi', in J. Reinharz and D. Schwetschinski (eds) *Mystics, Philosophers and Politicians*, Durham, NC: Duke University Press, 85-98. (Useful account of the relationship between these thinkers.)

Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Cambridge: Cambridge University Press. (Extensive bibliography of Jewish Averroists.)

Steinschneider, M. (1956) *Die hebräischen Übersetzungen des Mittelalters und die Juden als Dolmetscher (The Hebrew Translations of the Middle Ages and the Jews as Interpreters)*, Graz: Akademische Druck- under Verlagsanstalt.(Reprint of the celebrated 1893 bibliography.)

Vajda, G. (1952) 'A propos de l'Averroïsme juif' (On Jewish Averroism), *Sefarad* XII: 3- 29.(A useful general survey of the area with interesting remarks on, in particular, the thought of Moses ibn Crispin.)

Vajda, G. (1960) *Isaac Albalag - Averroiste juif, traducteur et commentateur d'Al-Ghazali (Isaac Albalag - Jewish Averroist, Translator and Commentator on al-Ghazali)*, Paris: Vrin.(Translation and study of the *Tikkun ha-De'ot*.)

Awakening of faith in Mahāyāna

*The Awakening of Faith in Mahāyāna (Dasheng qixinlun) is one of the most influential philosophical texts in East Asian Buddhism. It is most important for developing the Indian Buddhist doctrine of an inherent potentiality for Buddhahood (*tathāgatagarbha*) into a monistic ontology based on the mind as the ultimate ground of all experience. Its most significant contribution to East Asian Buddhist thought is its formulation of the idea of original enlightenment (*benjue*, or in Japanese,* hongaku*).*

Although attributed to Aśvaghoṣa and supposedly translated into Chinese by Paramārtha, there is general consensus among modern scholars that the *Awakening of Faith in Mahāhāyana (Dasheng qixunlun)* was most likely composed in China during the third quarter of the sixth century. Apocryphal texts such as the *Awakening of Faith* played a crucial role in justifying innovations in practice and belief that helped give shape to a uniquely East Asian form of Buddhism.

The *Awakening of Faith* was valued as a concise compendium of essential Mahāyāna doctrines. Its style is terse, often to the point of being cryptic. It thus allows for a range of different interpretations, so much so that some of the commentaries on the work sometimes seem to have been written on different texts.

It would be difficult to exaggerate the importance of this text on the development of East Asian Buddhist thought. Beginning with the influential commentary of Fazang (643-712), for example, it played an increasingly central role within the Huayan tradition, sometimes even eclipsing in importance the *Avataṃsaka (Huayan) Sūtra* on which the tradition based its authority. Its influence permeated early Chan texts in the Tang dynasty (618-907) as well. During the Song dynasty (960-1279), it was at the centre of the doctrinal debates that raged within the Tiantai tradition, as well as to a lesser extent those between the Tiantai and Chan (see Buddhist philosophy, Chinese). It also occupied a central place in the thought of Chinul (1158-1210), whose grand synthesis of Huayan (Hwaôm) and Chan (Sôn) was formative in defining the course of Korean Buddhism (see Buddhist philosophy, Korean). In Japan, it provided the cornerstone for the articulation of the so-called 'theory of original enlightenment' (*hongaku shisō*), which provided the doctrinal matrix out of which emerged the great medieval reform movements of Pure Land, Nichiren and Zen (see Buddhist philosophy, Japanese).

The *Awakening of Faith* developed the doctrine of *tathāgatagarbha* (embryo or womb of the Tathāgata) that had been introduced to China in such Indian Mahāyāna scriptures as the *Lion's Roar of Queen Śrīmālā* and the *Lankāvatāra Sūtra*. This doctrine had proclaimed the inherent potentiality of enlightenment in all sentient beings. Whereas this doctrine seems to have been of relatively minor importance in Indian Buddhism, it became the underlying premise on which the most distinctively East Asian Buddhist beliefs and practices were built. It was particularly valued because it guaranteed the universal accessibility of enlightenment and because it qualified the radical suspicion of language found in the Madhyamaka tradition by ascribing a series of positive predicates to the absolute (see Buddhism, Mādhyamika: India and Tibet).

The *Awakening of Faith* addressed a fundamental philosophical problem in the *tathāgatagarbha* tradition having to do with the relationship between enlightenment and ignorance. This problem was even more acute for *Awakening of Faith*'s ontologized understanding of mind. If the fundamental ground of all phenomena is the intrinsically enlightened mind, then how can ignorance be accounted for? The text tries to answer this question by linking the *tathāgatagarbha* with the *ālayavijñāna* (store consciousness), which explained the process of mental conditioning.

The *Awakening of Faith* begins by positing the one mind as the ultimate source of all phenomena. This mind has two aspects, which together totally comprehend all *dharmas*. The first is the mind as suchness (*xin zhenru men*), which the text defines as that which neither is born nor dies. The mind as suchness is the *tathāgatagarbha* in its true guise as the all-pervading, undifferentiated absolute reality (*dharmakāya*). Even though this mind is unborn and imperishable, differentiations come into being because of deluded thinking. Nevertheless, the true nature of the mind is beyond all predication and thought.

The second aspect of the one mind is referred to as the mind subject to birth-and-death (*xin shengmie men*). Founded on the *tathāgatagarbha*, this aspect of the one mind is identified with the *ālayavijñāna*, which the *Awakening of Faith* defines as 'the interfusion of that which is not subject to birth-and-death [that is, the mind as

suchness] and that which is subject to birth-and-death in such a way that they are neither one nor different.'

The *ālayavijñāna* also has two modes, which embrace and give rise to all *dharmas*. The first is enlightened (*jue*), and the second is unenlightened (*bujue*). 'Enlightened' is explained as meaning that the essence of the mind is free from thoughts. The characteristic of being free from thoughts is likened to the realm of empty space that pervades everywhere and is equal to the undifferentiated absolute body (*dharmakāya*) of the Buddha. Since the essence of mind is based on the *dharmakāya*, it is said to be intrinsically enlightened.

The *Awakening of Faith* goes on to distinguish intrinsic enlightenment (*benjue*) from experiential enlightenment (*shijue*). Experiential enlightenment, moreover, is contrasted with unenlightenment. In fact, the text states that experiential enlightenment can only be spoken of in the context of unenlightenment. Experiential enlightenment refers to the process by which one awakens to the ultimate source of the mind. Intrinsic enlightenment is at once the ontological ground that makes experiential enlightenment possible, and that which experiential enlightenment realizes.

At this point the *Awakening of Faith* introduces a famous metaphor to explain the relationship between ignorance and enlightenment: just as the originally tranquil ocean is stirred into waves by the wind, so the intrinsically enlightened mind is stirred into thoughts by ignorance. Under the heading of the unenlightened mode of the *ālayavijñāna*, the text then enumerates a series of nine stages that, like the classical twelve-link chain of interdependent origination (*pratītyasamutpāda*), explains how the process of karmic bondage develops. The process begins with the first subtle movement of thought. Based on the unenlightened mode of the *ālayavijñāna*, thought stirs the originally tranquil consciousness, which then manifests itself in terms of subject and object. This bifurcation of consciousness leads to the discrimation of likes and dislikes, and the awareness of pleasure and pain gives rise to associative patterns, which produce attachments, which lead to the generation of karma, which leads to becoming bound within the continuous process of birth and death (see Karma and rebirth, Indian conceptions of).

The theory of mind developed in the *Awakening of Faith* can be seen as an attempt to address the problem of ignorance. The relationship between the absolute and conditioned aspects of the one mind, or the *tathāgatagarbha* and the *ālayavijñāna*, traces back to a paradox at the core of the *tathāgatagarbha* doctrine: the *tathāgatagarbha* is at once intrinsically pure and identical with the *dharmakāya* and yet appears to be defiled. The two aspects of the one mind thus seem to be a matter of perspective, and their difference can be seen as corresponding to the point of view of ultimate and conventional truth. The *tathāgatagarbha* as seen from the enlightened perspective of a Buddha is perfectly pure and undefiled. It is only due to the deluded thinking of unenlightened beings that it appears to be otherwise.

The relationship between these two aspects of the one mind brings into focus the central philosophical problem for the *tathāgatagarbha* tradition: the origin of ignorance. If the mind is intrinsically enlightened, how can it become deluded? The problem arises because it is the consistent position of *tathāgatagarbha* theory that ignorance is only adventitious. The *Awakening of Faith*'s metaphor of the water and waves is thus an unsatisfactory resolution of the problem insofar as the comparison of the wind to ignorance posits a separate origin for ignorance. If ignorance had a separate origin, it would have its own autonomous ontological status placing it on an equal footing with enlightenment. The resulting dualistic ontology would undermine the axial premise of the *tathāgatagarbha* tradition that enlightenment is universally accessible to all beings.

In identifying the *ālayavijñāna* with the *tathāgatagarbha*, the *Awakening of Faith* grounds the process of conditioned origination on an intrinsically pure ontological foundation. This means that the defilements that appear to obscure the intrinsically enlightened mind of suchness are merely the manifestation of that mind as it adapts conditions and have no independent basis of their own.

See also: Buddhist philosophy, Chinese; Buddhist philosophy in India; Buddhism, Mādhyamika: India and Tibet; Mind, philosophy of; Ontology in Indian philosophy

PETER N. GREGORY

References and further reading

Gregory, P.N. (1986) 'The Problem of Theodicy in the *Awakening of Faith*', *Religious Studies* 22: 63-78.(This

article explores the problem of theodicy in the *Awakening of Faith*.)

Hakeda Yoshito (1967) *The Awakening of Faith*, New York: Columbia University Press.(Although not always satisfactory, this book offers the best translation of the *Awakening of Faith* into English.)

Awareness in Indian thought

Classical Indian schools all stake out positions on awareness, its intrinsic nature, its place in the causal processes crucial to human accomplishment, its relations to objects in the world, and the possibilities, according to certain religious or spiritual theories, of mystical transformation. In several prominent instances, stances taken on awareness may be said to constitute the most salient differentiation among schools, so central to a school's overall outlook is its view on the topic. Classical epistemological conceptions, for example, are in large part shaped by positions on awareness, and the spiritual philosophies for which Indian thought is best known present theories of awareness to guide meditation and mystical practice. Yogic, Vedāntic and Buddhist mysticism all came to be supported by views of the true nature of awareness or its native state.

In the professionalized debates that fill the immense proliferation of philosophical texts in the classical period (from approximately AD 100 to the eighteenth century and later), key issues are whether awarenesses have forms of their own or assume content only with reference to objects, and the precise nature of the relation, or relations, of awarenesses to objects in the world, including the role of awareness in human activity. Some important positions are shared across schools, and apart from the anti-theoretic polemics of Mādhyamaka Buddhists and others, a phenomenalist and idealist stance, a representationalism, and a direct or causal realism are the major theories concerning the content of awarenesses.

The world-oriented philosophies of Logic (Nyāya) and Exegesis (Mīmāṃsā) engage spiritual or mystical views (principally, Buddhist Yogācāra and Advaita Vedānta) on the issue of self-awareness or awareness of awareness. The exchange between upholders of Nyāya and Advaita Vedānta (Vedāntic Monism) on this score is, in particular, an admirable philosophical achievement.

1 Spiritual theories of awareness

According to preclassical texts called Upaniṣads ('secret doctrines'), human awareness offers a latent possibility of world-transcendence, of 'liberation' (*mukti*) of consciousness from embodiment and finite experience. The classical schools of Vedānta, Sāōkhya and Yoga develop views centred on such a mystical possibility, as do the socially revolutionary movements of Buddhism and Jainism (though the latter reject Upaniṣadic authority). In Advaita Vedānta, which of the Vedāntic (or Upaniṣadic) subschools is the most emphatic about the world-transcendent nature of awareness, the only veridical content of awareness is self-aware awareness itself. Awareness is by nature *kūṭastha*, 'stationed at the tip', at the far extreme, remote from the world; at least, it is so in its true nature or native state. In fact, awareness' worldly content is viewed as an illusion and a deformation, a falling away from what awareness is in itself. In this way, apparently, discoveries in meditation would be emphasized (see Monism, Indian §§1-2).

A dualistic view of awareness and the world can be teased out of the Advaita position: the world is contrary to the true nature of awareness, albeit the cosmic illusion depends on awareness, which is its substratum. In the schools of Sāōkhya and Yoga, in contrast, dualism is explicitly maintained: nature and awareness are separate realities, and it is problematic how they are related at all. As in Advaita, the only true content of awareness is awareness itself. With regard to the world, uninvolved witnessing - though, again, problematic (since the witness and the witnessed are unrelated in reality) - is upheld as the state the least estranged from what awareness is in itself (see Sāōkhya §§1-2, 6). (The similarities, by the way, among Advaita, Sāōkhya and Yoga on awareness hold despite the contrast between Advaita's claim that there is in reality only a single self, *ātman* - the appearance of multiple persons belongs to the cosmic illusion - and Yoga's and Sāōkhya's view that there are an infinite number of true persons, *puruṣa*, who are aware.)

The motivations for all three spiritual philosophies appear to be to show intellectually the possibility of a meditative state of self-absorption as well as to maintain its supreme value and desirability. By viewing awareness in its native state as without content other than self-aware awareness itself, the schools avoid having to explain the possibility of meditative self-absorption; rather, it is deviation from that state that is the mystery.

Taking a different angle on the three schools, we find that with respect to the everyday sort of awareness - perceptual awareness, in particular - intricate and carefully formulated theories emerge. According to Sāōkhya and

Yoga, nature ranges without break from luminous and malleable mentality to gross material elements, with the mind viewed as capable of presenting to the witnessing person or self representations of things that are continuous with the things represented both substantially and causally. In the *Yogasūtra*, which is the Yoga school's foundational text, this view is interwoven with a theory of personal transformation through yogic practice. That is, through psychophysical exercises centring on the ability to focus on and identify with any object whatsoever without distraction, there are, according to Yoga philosophy, practically unlimited possibilities of knowledge by extension of personal identity, a knowing of something ordinarily foreign to the self in the direct manner in which one knows oneself. According to Advaita Vedānta, modifications of a malleable 'internal organ' (*antaḥkaraṇa*) account for perceptions psychologically, although it is continually insisted that awareness itself belongs to another realm of reality. In Advaita's long history, a variety of cosmological theories are presented about how awareness comes to be associated with the internal organ, or mind.

Within Indian Buddhism, distinct views of awareness are constrained by the core teaching of an experiential *nirvāṇa* as the supreme good. Disidentification with the body, emotions, sensory presentations and thought is promoted by a 'no-self' doctrine (Sanskrit: *anātman*), a doctrine central to most Buddhist teachings about the way or path to *nirvāṇa*. Thus the real possibility of *nirvāṇa* experience, along with various prerequisites, is thought to demand a view of awareness as free-floating, unmoored by any enduring self or substratum, as well as a view of the apparent self as a bundle of awarenesses (see Buddhist philosophy, Indian §1; Nirvāṇa). Within this common frame, different schools of Buddhist philosophy uphold views of awareness that command attention in a host of classical texts that are less, or only peripherally, concerned with mystical possibilities. We shall take up the more refined Buddhist theories of awareness within that broader context.

2 Representationalism, idealism and direct realism

According to a view that appears in schools of early Buddhist philosophy (as well as in Sāōkhya and, arguably, Advaita Vedānta), awarenesses are of mental figures that represent objects external to awareness. External objects stand in causal relation to their representations, but are not what awarenesses are directly of. Such representationalism is attacked from two sides, from the perspective of phenomenalism and idealism (or at least scepticism about objects' independence of awareness) championed in the Yogācāra Buddhist school, and from the point of view of direct realism formulated in the sister schools of Nyāya (Logic) and Vaiśeṣika (Atomism), as well as in the traditionalist school of Brāhmanism known as Mīmāṃsā(Exegesis). The comparative advantages of representationalism are, first, that different types of awareness - perceptions, memories, imaginings, and so on - can be understood as all of the same type of object, namely, of mental objects, and, second, that the common intuition is affirmed of the independence from awareness of the pots and trees of the world as variously perceived, imagined, and so on, by different perceivers and by the same perceiver at different times. Direct realism has to assume that illusions have ontologically distinct objects from veridical awarenesses (the unreal and the real, respectively), although on any given occasion these cannot be differentiated from a first-person perspective: an illusion can seem to present a real object no less than a veridical awareness. And idealism embraces the counterintuitive position of putting the world in the head (or in the 'warehouse consciousness', *ālaya-vijñāna*; see below).

The Buddhist representationalists (sometimes known as Sautrāntikas) distinguish between the phenomenally existent (*saṃvṛti-sat*) and the substantially real (*dravya-sat*). Substantial existents are ultimate particulars (*svalakṣaṇa*), or atoms, which give rise to, for example, the phenomenon of a pot. Composites are not substantially real - the representation of a particular grouping of atoms, such as a pot, is mind-dependent - and particulars, that is, the substantially real atoms, while causally effective for appearances (though not for wholes *qua* wholes), do not themselves appear, at least not as atoms. The appearance of a pot represents a cluster of atoms, but the particular configuration is determined by human purposes and desires.

Dignāga (born *c*.480), a Yogācārin and an innovative logician (studied by philosophers of all schools), voices empiricist criticism of this position. The assumption that atoms cause the content of awareness although awareness does not represent the specific image of atoms means that a pot or a cup or anything might be perceived as anything else. The attributes of the atoms are not perceived, and everything that is perceived is considered, with respect to the atoms, to be like a double moon. 'It [the truly substantial objective support of awareness] is the object which exists internally to knowledge itself as knowable aspect and which appears to us as if it exists

externally' (*Ālambanaparīkṣā*, 1942: 48). Whether things really exist externally cannot be known.

According to Yogācāra, an awareness manifests a form internal to itself; it is *sākāra*, 'with form'. Yogācārins earlier than Dignāga maintained that a common 'warehouse consciousness' (*ālaya-vijñāna*) accounts for the intersubjectivity of objects among different subjects, and a beginningless psychological impulsion (*anādi-vāsanā*) for the ineluctable error of taking things to be external. Scholars have also supposed that the idealism here is motivated by a cosmological understanding of *nirvāṇa* or by a sense of a need to explain how *nirvāṇa* is an interactive possibility embracing everyone. Yogācāra is a Mahāyāna Buddhist school, and at the core of Mahāyāna is the teaching that we and everything are all interconnected, and that through the efforts of the Bodhisattvas (those of superlative wisdom and compassion) we shall eventually all arrive together at the supreme felicity (see Buddhism, Yogacārāschool of).

But whether Dignāga (or his equally illustrious follower, Dharmakīrti) himself embraced such an idealism has come to be disputed. What is clear is that the later Yogācāra philosophers work out a conceptual logic and a pragmatic epistemology that are supposed to be independent of such metaphysical questions as whether objects of awareness are merely internal to awareness or whether they also reflect external causes (see Epistemology, Indian schools of §1). Awareness guides action, and the touchstone of veridicality is success in action as shown by the coherence of the propositional content of one awareness with that of other awarenesses. The main criticism of representationalism is that it is otiose - as is also the direct realism of the Brāhmanical schools.

The direct realists, for their part, ably defend a contrary view of awareness, and voice criticisms of Yogācāra - including the pragmatism and scepticism of the late Yogācāra logicians and epistemologists - as well as of the earlier representationalism. We here focus on Nyāya (the 'Logic' school), although, especially with respect to attacks on Buddhist philosophy, Nyāya reasoners learned much from Mīmāṃsā(the 'Exegesis' school) (see Nyāya-Vaiśeṣika §6; MīmāṃsāČsect;2).

Awarenesses, according to Buddhists and Nyāya realists alike, are fleeting psychological episodes. So if they do not have extra-mental facts as content, then, since each awareness is itself a unique particular, linguistic practice (*vyavahāra*) could not be based on awarenesses. This is an unhappy ramification of the Buddhist theory. Communicable verbalization of awareness requires generalities or class characters as the content of awareness, because communication requires commonality in the experience of conversers (to talk to oneself would similarly require commonality among one's own experiential episodes). If each awareness has to itself content that is only an internal form, then each awareness, as itself a unique particular, will remain unverbalizable. According to Nyāya, awareness is *nirākāra*, 'without intrinsic form'; its content, including characters shared by things of the same type, is imparted to it by causal processes commencing with contact between sense organ and object.

The Buddhist representationalists, who, to their credit, saw the need for extra-mental causes of awareness content, still made two important mistakes, Nyāya advocates claim. First, generalities have to obtain in nature; neither awareness nor mentality can create the characters shared among things that we are aware of and that make communication possible (this thing as exhibiting the pothood that you from other experiences are acquainted with). Second, there is no need to postulate a third realm of mental figures, for to understand an awareness, itself a mental or internal event (properly, an episodic quality of the self or soul), as directly indicating facts, not mediated by mental figures, is the simpler hypothesis (*lāghavatva*). The occurrence of illusion does not establish such a third realm, because illusion can be explained on realist suppositions: due to an abnormality in a perceptual process, there is projection of a general character with which the subject is acquainted through prior experience and which is stored (associatively) in the subject's memory bank, projection on a particular that is being immediately experienced (see Error and illusion, Indian conceptions of §3).

The realist view is plainly vulnerable with respect to the role of memory it finds in illusion; the generality imparted to an awareness thereby is apparently better captured by a Yogācāra theory of a two-stage perceptual process, the first sensory and nonconceptual (involving only individuals or particulars), the second mental and conceptual (where salient groupings, or class characters, are imparted by the mind, in part based on human purposes). Philosophers of early Nyāya, in fact, struggled much with this, and tried, often confusedly, to concoct a similar two-stage realist theory. But the fourteenth-century 'New Logician' (*navya-naiyāyika*) Gaōgeśa (§2) managed to formulate a much clearer view of a pre-verbalizable stage of sensory awareness in which there is direct grasping of naturally obtaining generalities (and not only, as in the Yogācāra theory, particulars), followed by a verbalizable

stage in which these are attributed to particulars (see Sense perception, Indian views of §6).

3 Self-awareness

According to Nyāya, again in opposition to Yogācāra and also to Advaita Vedānta, self-awareness is not intrinsic to awareness, but is a matter of apperception (*anuvyavasāya*, literally 'after-cognition'), an awareness whose object is a previous awareness (that is, a 'scoped awareness', to use B.K. Matilal's happy phrase). Advaitins, such as the brilliant polemicist Śrīharṣa (*c.*1150), charge that the Nyāya view would make self-awareness impossible, in that as soon as one became aware of an awareness, A_1, the grasping awareness, A_2, not itself self-aware, would require an A_3 to be known, *ad infinitum*. The Advaita view, in contrast, is that every awareness is intrinsically self-aware (Yogcārins also defend this position). Moreover, awarenesses are infallibly known. Śrīharṣa argues, echoing the renowned Śaōkara (*c.*700, usually regarded as the greatest classical Advaitin), that if an awareness *qua* awareness could be doubted, then no resolution of any dispute would be possible. If anything is known, it must be awareness of awareness; awareness is self-luminous (see Śaōkara §2).

Gaōgeśa and other Nyāya philosophers respond by admitting the potential for a regress; it is simply that we normally have no call to make explicit in further apperception a lower-order apperception. We could do so, however, for as long as we wished to play that game. Normally we have no call to introspect, involved as we are in action and getting things done. Not every fact in the world is cognized by a human being. Similarly, not every awareness is apperceived. As the *Nyāyasūtra* indicates, however, some doubt does result from conflicts of theory, from the opposed positions of different schools, especially concerning the nature of awareness. Thus in philosophy we have special occasion to consult apperception to resolve higher-order doubts and disputes.

According to Gaōgeśa, in apperception, unlike in first-order awarenesses, the content of the scoped awareness is infallibly cognized as awareness content; that is to say, the scoped awareness is, *qua* awareness, known precisely as the phenomenon it is. Nyāya philosophers are generally fallibilists, taking the position that just about any bit of awareness content could prove to be erroneous. The exception is apperception, where the content of a previous awareness is infallibly known as awareness content. For example, the apperception verbalized as 'I took myself to be seeing a snake' will be veridical whether the scoped awareness' content (namely that something is a snake) obtains in fact. By taking this stance, Gaōgeśa is able to undercut the criticism of some of his opponents that Nyāya's realism about objects in the external world makes it blind to the infallible presentation of awareness to itself: Nyāya in fact holds that, *qua* subjective phenomenon, an awareness can be infallibly known. The view is also vital to Nyāya philosophers' being able to talk so discriminatingly about awarenesses, despite their direct realism.

See also: Knowledge, Indian views of; Self, Indian theories of

STEPHEN H. PHILLIPS

References and further reading

Bhattacharyya, S. (1993) *Gaōgeśa's Theory of Indeterminate Perception*, New Delhi: Indian Council of Philosophical Research.(Contains a translation of an important section from Gaōgeśa's philosophic masterpiece, and lucid explanation of the general Nyāya understanding of awareness.)

Dasgupta, S. (1922) *A History of Indian Philosophy*, vol. 1, Cambridge: Cambridge University Press, repr. 1969. (Remains probably the best overall introduction to classical Indian systems, although it is out of date concerning Buddhist views.)

Dignāga (*c.*500) *Ālambanaparīkṣa (Examination of the Object of Awareness)*, trans. N.A. Sastri, *Dignāga's Ālambanaparīkṣā*, Madras: The Adyar Library, 1942. (Sastri reconstructs the original Sanskrit, which is lost, from a Tibetan translation, as well as translating the text and much of a commentary by Vinītadeva into lucid English.)

Hayes, R.P. (1988) *Dignāga on the Interpretation of Signs*, Dordrecht: Kluwer. (Includes an annotated translation of two chapters of Dignāga's *Pramāṇasamuccaya (Collected Writings on the Acquisition of Knowledge)*; provides detailed contextualization.)

Matilal, B.K. (1986) *Perception: An Essay on Classical Indian Theories of Knowledge*, Oxford: Oxford University Press.(A modern classic by a premier scholar and philosopher intent on defending Nyāya.)

Nyāyasūtra (*c.* AD 150), trans. M. Gangopadhyay, *Nyāya-Sūtra with Vātsyāyana's Commentary*, Calcutta: Indian Studies, 1982.(A readable and accurate translation.)

Phillips, S.H. (1995) *Classical Indian Metaphysics: Refutations of Realism and the Emergence of 'New Logic'*, La Salle, IL: Open Court.(Contains elaboration of, in particular, what is discussed in §3.)

Axiology

Axiology is the branch of practical philosophy which studies the nature of value. Axiologists study value in general rather than moral values in particular and frequently emphasize the plurality and heterogeneity of values while at the same time adopting different forms of realism about values. Historically, three groups of philosophers can be described as axiologists: the original Austrian and German schools of value phenomenologists; American theorists of value who offered an account of value which reduces it to human interests; and an English school, influenced by Austro-German phenomenology, which included such diverse figures as G.E. Moore, Hastings Rashdall and W.D. Ross. Recent philosophy has seen a resurgence of interest in value realism in the broadly axiological tradition.

1 Introduction

'Axiology' is a translation of the German *Axiologie*, which simply means 'theory of value'. Axiology is the branch of practical philosophy which seeks to provide a theoretical account of the nature of values, whether moral, prudential or aesthetic (see Values). Values are understood in their tradition to be the proper objects of practical attitudes, analogous to the way truth is the proper object of our theoretical judgments. Historically, there have been three important groups of philosophers whose work can be described under the heading of 'axiology'. One of these is the American school of John Dewey and C.I. Lewis (see Pragmatism in ethics). The American school was subjectivist; values were viewed as created by subjective interests, in the form of desires. The Austro-German and British traditions, in contrast, were primarily realist or objectivist (see Moral realism §§1).

2 The Austro-German school of phenomenology

Historically the earliest group are the value theorists of nineteenth-century Austria and Germany, a group which includes Franz Brentano, Alexius Meinong, Max Scheler, and Nicolai Hartmann. Their work has been neglected, not least because of a perceived distinction between 'analytic' and 'continental' philosophy, which came to dominate post-war analytical philosophy to the detriment of both traditions. These philosophers based their approach to practical philosophy on a 'psychological' or 'phenomenological' method which involves giving a precise description of one's experience, freed as far as possible from all presuppositions and all theoretical interpretation, and paying special attention to the objects of experience precisely as they are experienced (see Phenomenological movement). They aimed to give an account of the contents and objects of evaluative acts which paralleled the move towards the investigation of the contents and objects of theoretical judgments on the part of Husserl, Frege and other logical realists at the turn of the century.

The work on value of Brentano (1889) and Meinong (1917) was grounded in Brentano's revision of the medieval concept of intentionality, which he and his successors took to mean the directedness of thought upon its objects (see Intentionality). Mental acts, for Brentano, are built up out of three elements: the presentation of an object, for example a sense datum; a dimension of positive or negative belief in the existence of the object of the belief; and finally a third dimension called a 'phenomenon of interest'. Just as judgments may be positive or negative, so also phenomena of interest are divided into positive interest phenomena, or 'love', and negative interest phenomena, or 'hate'. Just as some judgments are 'correct' or 'incorrect' according to whether they are marked by a special quality of clear and distinct evidence, so acts of love and hate, too, may be correct or incorrect. Brentano sees the origin of all ethical knowledge as lying in our experience of correct love and hate. At the same time, with his student Christian von Ehrenfels, he emphasized the role of organic unity in constituting valued objects.

Meinong's theory is a refinement of Brentano's doctrine. Meinong, however, objectified Brentano's phenomena of interest, conceiving of the corresponding acts as having special objects of their own in the realms of value and 'oughtness'.

Complementing the Austrian emphasis on the psychological complexities of valuing, the German contributions of Scheler (1913, 1916) and Hartmann (1926) focus on describing the structures of valuable objects. However, both theories also contain a doctrine of the 'emotional a priori' in which emotion serves as an a priori principle of practical thinking, a view influenced by Lotze and by Brentano's theory of correct and incorrect emotions (see Lotze, R.H.). The phenomenological investigations of Scheler and Hartmann focus on the classification of the

objects of practical judgments into classes such as the ‘higher’ and the ‘lower’, on the basis of such criteria as permanence, fundamentality, universality and so on. Hartmann’s version of the theory has the merit of emphasizing the possibility of tension and conflict amidst this experienced plurality of values (see Moral pluralism). Scheler sought to provide a ‘material’ ethics doing justice to the vast plethora of different types of things that different agents in different cultures value. He also emphasized the religious aspects of the value pantheon.

3 The British axiological tradition

In the analytical tradition axiological theses are to be found above all in Moore’s *Principia Ethica* (1903), and in the writings of Rashdall and Ross (see Moore, G.E.). Both Moore and Rashdall were ‘ideal consequentialists’, whose account of right action sees rightness as consisting in the production of goodness (see Consequentialism). Moore’s axiological theses in *Principia* reflect to some degree the influence of the Austrian school which Moore admired: Moore’s much criticized account of the faculty of moral intuition includes a reference to feeling and the will (see Intuitionism in ethics); his account of goodness and beauty is indebted to Brentano, as is his account of ‘organic unities’ in value; Moore even offers an account of the structure of the value realm, although it is perhaps rather too committed to the idiosyncratic values of the Bloomsbury group to be entirely convincing.

Hastings Rashdall’s now neglected *Theory of Good and Evil* (1912) offers an account similar to that of Moore, with the emphasis falling not on organic wholes, but on states of consciousness as the ultimately valuable objects of our judgment. Interestingly, Rashdall also believes that all values are ultimately commensurable (see Rashdall, H.).

W.D. Ross (1930) was the third great mid-century English axiologist. While his historical reputation rests largely on his neo-Aristotelian account of duty, axiological themes nevertheless run throughout his work, not least because he exploited the principle of Moore’s argument against the definability of ‘good’ to argue that the term ‘right’ was similarly undefinable. Ross saw the term ‘good’ as attaching to states of affairs, whereas ‘rightness’ is predicated of acts. Ross offers a three-fold classification of values, combined with a thesis of value incommensurability that marked his divergence from Rashdall’s otherwise similar position (see Ross, W.D.; Right and good).

4 The eclipse of value theory

Value theory has been eclipsed in recent Anglo-American analytical philosophy for several reasons: first, the rise of evolutionary psychology; second, the dominance of rational decision theory; third, the dominance of meta-ethical scepticism about the place of values in the world; fourth, the problem of motivational internalism; and finally certain key changes in the modern view of morality itself.

Evolutionary psychology seems to offer an account of the evolution of our ‘moral sense’ that dispenses with any reference to objective values. Its apparent elimination of objective values on the grounds of their explanatory dispensability has offered a new impetus to ethical scepticism, as illustrated above all in the work of J.L. Mackie and Michael Ruse, although this interpretation of evolutionary theory has been strongly contested, for example by Robert Nozick. Nozick (1981) argues that an evolutionary account of the moral sense can no more dispense with values than an evolutionary account of perception can dispense with perceptual objects objectively present in the world.

Decision theory seems to be evaluatively neutral, formulating simply the rules a rational agent must employ to determine the means to established ends; it therefore seems to suggest that values are theoretically eliminable, replaceable by subjective preferences (see Decision and game theory §2). Susan Hurley (1989) and Henry S. Richardson (1994), among others, have argued that the theory of practical reasoning cannot be an autonomous theory which can dispense with reference to values. A central line of argument is that decision theory rests on an insufficiently realistic model of the psychology of practical agents and that a more realistic theory must necessarily refer to values. Hurley argues that the theory of practical reasoning embodied in contemporary decision theory cannot be an empirical theory, applicable to real world agents, unless supplemented by assumptions about what agents value.

Moral scepticism questions how values could be part of the world as described by natural science (see Naturalism in ethics §1); the sceptic cannot see the justification for postulating values as real entities. On this point David

Wiggins (1987) and John McDowell concur with the naturalistic realists known as the 'Cornell school', such as Richard Boyd and David Brink (1989). Evaluative properties are postulated just like any other kind of properties: they are theoretical entities which must earn their keep, ontologically, by playing a role in explaining the responses of a rational agent. However, the Cornell school add the further insistence that such properties must ultimately be naturalistically reducible. This claim is resisted by Wiggins and McDowell; the insight of their much misunderstood secondary property model for evaluative properties is that values may be subjective, in the sense of tied to anthropocentric interests of ours, but nevertheless real in the only way any property is real - by pulling its weight in successful explanations (see Moral realism §4-5).

The problem of motivational internalism marks a major point of divergence between the British and American versions of moral realism. If one accepts the intuition expressed by internalism to the effect that moral judgments necessarily have an impact on action, how can a perception of objective properties explain such a connection? The Cornell realists simply dispense with the intuition and claim that the link between judgment and motivation is contingent - the thesis of externalism (see Moral motivation §1). Wiggins and McDowell argue that the connection is secured via a model of the ideal agent, the virtuous person, who is both equipped to perceive moral properties and disposed to assign these perceptions authority in their practical deliberation. Such a model, in their view, offers a defensible version of internalism.

The last reason for the decline in axiology turns on the fact that objective values seem dispensable to a modern conception of the self as autonomous, a spontaneous creator of value, which values freedom and autonomy above all else (see Autonomy, ethical). The value theorist can, however, argue that those features of this modern view of morality that seem to make reference to values dispensable in fact rest on suppressed commitments to certain specific evaluative frameworks - frameworks within which our commitment to such values as autonomy and freedom make sense. This is the central argument of Charles Taylor's *Sources of the Self* (1989), which seeks to overcome the philosophical basis of what he correctly calls an 'ethics of inarticulacy', which is based on the modern values of procedural rationality and freedom, but which is, by its own lights, unable to articulate the ways in which our moral commitments are ultimately based in the theory of value.

5 The revival of value theory in recent moral philosophy

Powerful forces, both outside and inside moral philosophy, have conspired to marginalize value theory in contemporary moral philosophy. Yet the foregoing demonstrates that value theory has undergone a considerable revival in recent moral philosophy.

A strong influence on recent developments was the work of the émigré moral phenomenologist Aurel Kolnai; other 'transmitters' of the Austro-German tradition to Anglo-American moral philosophy were J.N. Findlay (1963), R.M. Chisholm and Maurice Mandelbaum. Kolnai's work has in turn influenced the account of moral knowledge found in the work of Wiggins and McDowell.

Quite independent 'theorists of value' in contemporary ethics are Platonists such as Iris Murdoch and Neo-Kantian theorists such as John Rawls and Robert Nozick. Nozick in particular has looked back to the Austrian and German schools of axiology as inspiration for his work, which even includes a delineation of the valuable 'facets of being', including such categories as 'richness', 'completeness' and 'amplitude', in the manner of Scheler and Hartmann.

See also: Good, theories of the; Value, ontological status of; Values

BARRY SMITH
ALAN THOMAS

References and further reading

Brentano, F. (1889) *Vom Ursprung sittlicher Erkenntnis*, Leipzig: Duncker & Humblot; 4th edn, ed. O. Kraus, Hamburg: Meiner, 1955; trans. R.M. Chisholm and E.H. Schneewind, *The Origin of our Knowledge of Right and Wrong*, London: Routledge, 1969.(Classic presentation of Brentano's views, in English translation.)

Brink, D. (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press. (Representative of the naturalistic moral realism of the 'Cornell school'.)

Chisholm, R.M. (1986) *Brentano and Intrinsic Value*, Cambridge: Cambridge University Press.(Important and pioneering study of Brentano's theory of value.)

Findlay, J.N. (1963) *Meinong's Theory of Objects and Values*, Oxford: Clarendon Press.(General study of

Meinong's approach to the contents of thoughts.)
Grice, H.P. (1991) *The Conception of Value*, Oxford: Oxford University Press.(A recent work of value theory which looks to both Aristotle and Kant for an account of the 'construction' of value.)
Hartmann, N. (1926) *Ethik*, Berlin: de Gruyter; trans. S. Coit, *Ethics*, London: Macmillan, 1932, 3 vols. (Hartmann's comprehensive system in English translation.)
Hurley, S. (1989) *Natural Reasons*, Oxford: Oxford University Press.(Wide-ranging study, developing a theory of value in the context of the philosophy of mind more generally conceived. Contains an interesting critique of theories of practical reasoning that do not make assumptions about what agents value.)
Korsgaard, C. (1996) *The Sources of Normativity*, Cambridge: Cambridge University Press.(Presents an account of values from the point of view of the Kantian rationalist tradition.)
Meinong, A. (1917) *Über emotionale Präsentation*, Vienna: Holder; repr. 'Über emotionale Präsentation' *Sitzungsberichte der Akademie der Wissenschaften in Wien, philosophisch-historische Klasse* 183, 1924; trans. and ed. M.-L. Schubert Kalsi, *On Emotional Presentation*, Evanston, IL: Northwestern University Press, 1972. (Meinong's central work in value theory.)
Moore, G.E. (1903) *Principia Ethica*, ed. T. Baldwin, Cambridge: Cambridge University Press, revised edn, 1993. (One of the classic texts of modern moral philosophy and the origin of the modern British tradition in value theory. This revised edition has a valuable introduction by Thomas Baldwin and includes a reprint of 'The Conception of Intrinsic Value'.)
Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press, part V, chaps 1-6.(The work of contemporary value theory which stands closes to the Austro-German tradition; contains a comprehensive value theory of its own.)
Rashdall, H. (1912) *The Theory of Good and Evil*, London: Oxford University Press. (Rashdall's central work in value theory, much discussed by other mid-century British moral philosophers).
Richardson, H. (1994) *Practical Reasoning About Final Ends*, Cambridge: Cambridge University Press.(A recent work which argues that references to values are indispensable to a correct account of practical reasoning.)
Ross, W.D. (1930) *The Right and the Good*, Oxford: Clarendon Press.(Classic work, cited here for its contribution to value theory in addition to its more well-known account of *prima facie* duties.)
Scheler, M. (1913, 1916) *Der Formalismus in der Ethik und die materiale Wertethik*, Halle: Max Niemeyer; trans. M.S. Frings and R.L. Funk, *Formalism in Ethics and Non-Formal Ethics of Values*, Evanston, IL: Northwestern University Press, 1973. (Scheler's central work in translation.)
Smith, B. (1994) *Austrian Philosophy: The Legacy of Franz Brentano*, Chicago, IL: Open Court.(Historical study of Austrian philosophy which offers a comprehensive background account to the Austrian theory of value, especially Brentano's work.)
Taylor, C. (1989) *Sources of the Self*, Cambridge: Cambridge University Press. (Historically informed history of the modern view of morality, grounded in a value theoretic perspective.)
Wiggins, D. (1987) *Needs, Values, Truth*, Oxford: Blackwell; revised edn, 1991.(Collected papers of an important contemporary value theorist who defends a view of moral properties as anthropocentrically conditioned, but none the less indispensable to moral explanations and hence real.)

Axiom of choice

The axiom of choice is a mathematical postulate about sets: for each family of non-empty sets, there exists a function selecting one member from each set in the family. If those sets have no member in common, it postulates that there is a set having exactly one element in common with each set in the family. First formulated in 1904, the axiom of choice was highly controversial among mathematicians and philosophers, because it epitomized 'non-constructive' mathematics. Nevertheless, as time passed, it had an increasingly broad range of consequences in many branches of mathematics.

1 History of the axiom

Consider a whimsical example, due to Bertrand Russell, illustrating the need for an axiom of choice (stated above). Suppose a millionaire buys as many pairs of socks as pairs of boots and suppose that at last he has infinitely many pairs of each. Then (1) he has exactly as many boots as he does pairs of boots, and (2) he has exactly as many socks as he does pairs of socks. To show (1), we make use of the fact that there is a rule to pick a boot from each pair (for example, the left one). To establish (2) we must be able to pick a sock from each pair. Leaving aside artificial devices, such as the position of socks in space, there is no way to distinguish a sock in each pair and hence no way to prove (2). The axiom of choice implies the existence of such a choice from each pair, even when no rule is available to make the choice.

Thus the axiom involves infinitely many choices from sets. Despite misunderstandings by some philosophers, it has nothing to do with a single choice from a set or even with a finite number of choices, since these are already validated by first-order logic. Nor does the axiom concern the empirical possibility of actually doing the choosing, although this red herring was once taken seriously.

When first formulated by Ernst Zermelo in 1904, the axiom of choice was highly controversial. It led a number of British, French and German mathematicians to consider the philosophical side of mathematics, and induced several of them, especially in France, to adopt various constructivist philosophies of mathematics (see Constructivism in mathematics). At issue was the question of what is meant by 'existence' in mathematics.

The axiom of choice also motivated the axiomatization of set theory. Concerned primarily to defend this axiom against its numerous critics and secondarily to avoid the contradictions recently discovered (see Paradoxes of set and property), Zermelo axiomatized set theory in 1908. The axiom of choice figured prominently in his axiomatization. It turned out that many mathematical theorems required the axiom, and quite a few were actually equivalent to it (see §3 below).

The axiom is intimately involved with the emergence of important techniques for constructing models of set theory in first-order logic. The first such technique (1938) was Kurt Gödel's constructible sets (see Constructible universe). The second (1963) was Paul Cohen's method of forcing. These techniques established that the axiom of choice is consistent with the usual axioms of set theory, but also independent from them.

2 Philosophical questions surrounding the axiom

Today almost all mathematicians accept the axiom of choice, although many still remark when they use it. Earlier, Zermelo defended the axiom against its critics on two grounds: its usefulness and its self-evidence. He remarked that it had already been widely used, without being formulated explicitly. Such widespread use, he argued, could only be explained by the axiom's self-evidence.

Opponents of the axiom had various views. Some, such as Emile Borel, accepted the axiom when applied to a denumerable family of sets, but rejected it when applied to a family of larger cardinality, such as the set of all real numbers. Others, including René Baire, tentatively accepted the real numbers but argued against the existence of the set of all subsets of the real numbers, from which subsets Zermelo's choices were made. Many opponents rejected the axiom because it provided no rule for making the choices. Apparently thinking of axioms in the traditional way as self-evident truths, they often failed to understand that the axiom is needed precisely when no rule is available.

Opponents were encouraged by the axiom's counter-intuitive consequences. The most vivid of these was the

Banach-Tarski paradox, whereby a sphere *S* can be decomposed into a finite number of pieces and reassembled into two spheres each of the same size as *S*. The three-dimensionality is important here, since such a decomposition cannot be done with a two-dimensional figure such as a circle. Euclidean geometry also plays a role, since hyperbolic geometry (in which the parallel postulate is false) allows the various counter-intuitive decompositions without using the axiom.

The irony for the axiom's opponents was that if set theory is understood in a sufficiently constructivist way, then the axiom is true. In the extreme case, when all sets are finite, the axiom is true but trivial. If set theory is restricted to Gödel's constructible sets, then again the axiom is true but now distinctly non-trivial. It is only when set theory is taken in a non-constructive way that the axiom might be false. Even in intuitionism, various restricted forms of the axiom are true, but their meaning is changed since the underlying logic is not first-order logic (see Intuitionism).

3 Mathematical theorems that need the axiom

Zermelo introduced the axiom of choice to prove the well-ordering theorem: every set can be well-ordered (see Set theory). Previously Georg Cantor had assumed this well-ordering theorem as a 'law of thought', but mathematicians had been sceptical. Zermelo showed that the well-ordering theorem follows from an axiom that is conceptually much simpler: the axiom of choice.

The axiom of choice is vital to the arithmetic of infinite cardinal numbers. In particular, it is equivalent to the proposition that if two cardinals are unequal, then one is greater than the other. It is needed in order to define the addition or multiplication of infinitely many cardinals.

Various definitions of the finiteness of a set involve the axiom. The usual definition says that a set *M* is finite if it is empty or if there is some counting number *n* such that *M* can be put into one-one correspondence with $\{1, 2, \ldots, n\}$; otherwise *M* is infinite. Richard Dedekind formulated the first definition that made no use of numbers. A set *M* is said to be 'Dedekind-infinite' if there is a one-one function from *M* to a proper subset of itself; otherwise, *M* is 'Dedekind-finite'. The philosophical significance of this definition is not only that it dispensed with numbers but also that the infinite is defined positively, rather than as the absence of some property. Intuitively, a set should be finite if and only if it is Dedekind-finite, but this equivalence requires a weak form of the axiom of choice. Moreover, there are definitions of finiteness whose equivalence to the usual definition is itself equivalent to the axiom.

The axiom permeates most branches of mathematics, except geometry. Without the axiom, much of algebra, analysis and topology would not exist. A typical algebraic theorem equivalent to the axiom is that every vector space has a basis. In topology, an equivalent theorem is the proposition that the product of compact topological spaces is compact. In analysis, we rarely need the full strength of the axiom but can get along with a weaker assumption: the principle of dependent choices. This principle asserts the existence of a sequence of choices, each depending on the finitely many choices previously made.

In logic, the axiom of choice also plays an essential role in many important results. It is needed to prove the completeness theorem for first-order logic, as well as the compactness theorem and the completeness theorem for propositional logic, provided that the set of primitive symbols of the logic may have any infinite cardinality. With such a language, the downward Löwenheim-Skolem theorem is also equivalent to the axiom: every model of an infinite set *M* of sentences has a model whose cardinal is at most that of *M* (see Löwenheim-Skolem theorems and nonstandard models; Model theory). It is ironic, and of philosophical significance, that the axiom must be used to prove, via forcing, that there exists a model of set theory in which the axiom is false.

The axiom of choice is implied by various stronger propositions in set theory, such as the axiom of constructibility and the generalized continuum hypothesis (see Constructible universe; Continuum hypothesis). Another such proposition, which, unlike those two, is very intuitive, was discovered by John von Neumann: a class *M* is a proper class (that is, not a set) if and only if *M* can be put into one-one correspondence with the class of all sets. This beautiful proposition turns Burali-Forti's paradox (see Paradoxes of set and property) into an argument that the universe of all sets is well-ordered, and illustrates how paradoxical results, if viewed correctly, can often be transformed into useful theorems.

See also: Logical and mathematical terms, glossary of

GREGORY H. MOORE

References and further reading

Jech, T.J. (1973) *The Axiom of Choice*, Amsterdam: North Holland.(A careful study of the independence results surrounding the axiom.)

Moore, G.H. (1982) *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*, New York: Springer.(A detailed historical treatment of the axiom from its prehistory to the independence proofs. An appendix gives a translation of the most important document by the French opponents of the axiom.)

Rubin, H. and Rubin, J.E. (1985) *Equivalents to the Axiom of Choice II*, Amsterdam: North Holland. (Statements of many propositions equivalent to the axiom, with proofs.)

Russell, B.A.W. (1919) 'Selections and the Multiplicative Axiom', in *Introduction to Mathematical Philosophy*, London: Allen & Unwin; repr. London: Routledge, 1993.(An introductory discussion of the axiom of choice, which Russell calls the multiplicative axiom, by one of its most prominent philosophical opponents.)

Ayer, Alfred Jules (1910-89)

A.J. Ayer made his name as a philosopher with the publication of Language, Truth and Logic in 1936, a book which established him as the leading English representative of logical positivism, a doctrine put forward by a group of philosophers known as members of the Vienna Circle. The major thesis of logical positivism defended by Ayer was that all literally meaningful propositions were either analytic (true or false in virtue of the meaning of the proposition alone) or verifiable by experience. This, the verificationist theory of meaning, was used by Ayer to deny the literal significance of any metaphysical propositions, including those that affirmed or denied the existence of God. Statements about physical objects were said to be translatable into sentences about our sensory experiences (the doctrine known as phenomenalism). Ayer further claimed that the propositions of logic and mathematics were analytic truths and that there was no natural necessity, necessity being a purely logical notion. Finally the assertion of an ethical proposition, such as 'Stealing is wrong', was analysed as an expression of emotion or attitude to an action, in this case the expression of a negative attitude to the act of stealing.

During the rest of his philosophical career Ayer remained faithful to most of these theses, but came to reject his early phenomenalism in favour of a sophisticated realism about physical objects. This still gives priority to our experiences, now called percepts, but the existence of physical objects is postulated to explain the coherence and consistency of our percepts. Ayer continued to deny that there were any natural necessities, analysing causation as consisting in law-like regularities. He used this analysis to defend a compatibilist position about free action, claiming that a free action is to be contrasted with one done under constraint or compulsion. Causation involves mere regularity, and so neither constrains nor compels.

1 Life

A.J. Ayer was educated at Eton and Christ Church, Oxford. As an undergraduate he formed an interest in philosophy and was encouraged by Gilbert Ryle to spend some time in Vienna in 1932. It was at this time that he attended meetings of the Vienna Circle and was converted to logical positivism (see Logical positivism §4; Vienna Circle §2). During the Second World War he spent most of his time in military intelligence. After the war he became Grote Professor of Philosophy of Mind and Logic at University College, London, leaving London in 1959 to become the Wykeham Professor of Logic at New College, Oxford. During this time Ayer became a well-known public figure, appearing in radio and television programmes, and he was knighted in 1970. Ayer took a special interest in encouraging younger philosophers, to whom he was always known as 'Freddie'; and well after his retirement in 1978 he continued to enliven the annual British philosophical meetings (the 'Joint Sessions'). He married four times, including one remarriage, and wrote two autobiographies (1977; 1984).

Philosophically Ayer saw himself as being part of the tradition of British empiricism started by Locke and Hume and continued by Russell. He was a prolific writer, contributing especially to epistemology, but also producing many articles in the philosophy of science and philosophy of mind. In addition he wrote books on Hume, Russell and Moore, Wittgenstein, and the American pragmatists, Peirce and James.

2 Verificationism

The anti-metaphysical doctrines of logical positivism attracted Ayer at least partly because of his adverse reaction to the British idealists F.H. Bradley and McTaggart, who denied the reality of matter, space and time (see Bradley, F.H. §5; McTaggart, J.M.E. §3). Positivism provided Ayer with the weapon he needed to denounce such claims as metaphysical nonsense. That weapon was the verificationist theory of meaning, which divided meaningful sentences into those that were analytically true (or false) and so known a priori, and those that expressed propositions that were empirically verifiable. Sentences purporting to express the traditional metaphysical views were deemed to be either disguised analytic statements, reflecting a decision to use words in a certain way, or nonsensical. Ayer's first book, *Language, Truth and Logic* (1936), formulated the verificationist criterion of meaning and pursued its consequences for various philosophical controversies. The book became notorious partly because of its confident and aggressive style, and partly because of the distinctive theses endorsed concerning the nonsensical nature of assertions of God's existence and the analysis of ethical statements as expressive of our emotions, with no cognitive significance. Given the controversial nature of these claims, much rested on a satisfactory formulation of the positivist criterion of meaning, the verification principle.

In the first edition of *Language, Truth and Logic*, Ayer stipulated that a proposition is strongly verifiable if and only if its truth can be conclusively established by experience, and that a proposition is weakly verifiable if and only if it is made probable by experience. Ayer sharpened this definition of weak verifiability to the following: a proposition is weakly verifiable if and only if an observation statement is deducible from it in conjunction with other premises, and not deducible from those other premises alone. This definition proved too permissive: take *any* sentence '*S*' and observation statement '*O*'; the latter will follow from '*S*' and 'if *S* then *O*', without following from 'If *S* then *O*' alone. To remedy this defect Ayer reformulated the criterion in the introduction to the second edition of *Language, Truth and Logic* (1946). Here he calls a statement 'directly verifiable' if it is either itself an observation statement, or, in conjunction with other observation statements, it entails at least one observation statement not deducible from these other observation statements alone. And a statement is 'indirectly verifiable' if, in conjunction with other premises, it entails a directly verifiable statement which is not deducible from these other premises alone, and provided that these premises are themselves either analytic, or directly verifiable, or independently established as indirectly verifiable. A statement was meaningful if and only if it was either directly or indirectly verifiable in the above sense.

Unfortunately for Ayer this reformulated criterion of meaning was shown to be too permissive as well. Alonzo Church (1949) showed that one can again render *any* statement meaningful according to the reformulated criterion. Let O_1, O_2 and O_3 be logically independent observation statements and *S* be any statement.

(1) $(\sim O_1 \;\&\; O_2) \vee (O_3 \;\&\; \sim S)$

is then directly verifiable ((1) in conjunction with O_1 entails O_3). As long as O_2 does not follow from (1) alone, *S* is then indirectly verifiable, as O_2 follows from *S* and (1), and (1) is directly verifiable. Furthermore, if O_2 follows from (1) alone then O_2 follows from $O_3 \;\&\; \sim S$, which implies that $\sim S$ is directly verifiable, since O_2 and $O_3 \;\&\; \sim S$ are logically independent. Ayer acknowledged the force of this argument, but retained his belief that verifiability was crucial to meaning. He thought that once the theory of confirmation had been worked out in detail it would be possible to overcome the problems posed by Church.

Most of *Language, Truth and Logic* drew out the consequences of Ayer's verificationist theory of meaning. Material objects were construed as being constituted by actual and possible sense-data. Statements about other minds were meaningful on the basis of their verification by behaviour, whereas statements about one's own mind were not analysed as having a behaviourist basis. As a consequence mental predicates had different meanings, depending on whether they were used in judgments about others or oneself. Past-tense statements were originally thought to be rules for the prediction of the experiences which would verify them, a view which made them translatable into propositions about present and future experience. Ayer retracted this in the introduction to the second edition, where he asserted that the present unobservability of such events was accidental, so past-tense statements could be interpreted as analogous to statements about spatially distant events which we could not see. He proposed that they be analysed as being verifiable by those experiences we would have if we were to 'travel' to the historical event being described (see Meaning and verification §§2-4).

Two categories of expression were given fairly extended treatment in *Language, Truth and Logic*: those of logic and mathematics, and those of morality. The truths of logic and mathematics were too important to be declared meaningless, but Ayer resisted the thought that they were empirically verified. Instead such necessary truths were classified as analytic truths, and as such their truth depended solely on facts about how we used certain terms. For example, given that '$\rightarrow$' had been given a meaning in Russell and Whitehead's system of logic, the a priori proposition '$(q \rightarrow (p \rightarrow q))$' was analytically true. The merits of such an account were famously called into question by Quine (1936, 1951) (see Necessary truth and convention §1; Quine, W.V. §8). The category of moral statements was approached differently, and will be discussed later.

3 Perception

Verification of sentences proceeded by singling out a favoured class of sentences - 'observation statements' or 'basic sentences', which were the principal verifiers. There was a great deal of controversy among the positivists as to the nature of these sentences. Given the role they are assigned, they cannot themselves be verified by another set of sentences, so they have to be either self-verifiers or chosen independently of a verification criterion. The latter strategy - conventionalism - was favoured by Otto von Neurath, who persuaded Carnap of its virtues (see

Conventionalism §1). Ayer consistently opposed it, arguing that the use of any independent criterion, such as choosing those basic statements agreed upon by scientists, would require the exercise of some judgment as to what had been agreed (Ayer 1937). This judgment would itself require support; so the proposed independent criterion was accused of presupposing verification by experience. The self-verification option was the only one available.

Ayer's own thought about the nature of basic statements changed after his initial statement of it in *Language, Truth and Logic*, but he never gave up on the idea that the foundations of empirical knowledge were to be found in our sensory experience, in sense data or qualia. Sense-datum statements (or observation statements) formed the foundation because they were minimally committal: their function was simply to record what appeared to us without implying anything further about what may be true. Ayer vacillated as to whether such statements were infallible, but decided that the possibility of their fallibility did not make them unsuitable for their role as basic statements. What was essential was that they could not be false in one of the ways in which statements about physical objects could be false, through the person making the statement being misled by appearances. The beauty of sense-data was that their reality coincided with their appearance. Ayer's account of sense-data was criticized by J.L. Austin (1962), to which Ayer responded vigorously in 'Has Austin Refuted the Sense-Datum Theory?' (1967).

What of our knowledge of physical objects? In *Language, Truth and Logic* and *The Foundations of Empirical Knowledge* (1940) Ayer defended phenomenalism, the view that anything said about ordinary physical objects was translatable into a series of statements about sense-data. After the war he rejected phenomenalism on the grounds that no set of sense-datum statements could entail the truth of a statement about a physical object. In *The Central Questions of Philosophy* (1973) Ayer eventually settled for a form of realism which he called 'sophisticated realism' and which he differentiated from naïve realism, which holds that we directly perceive physical objects. For Ayer, our perceptual judgments about physical objects are always dependent upon an inference, usually implicit, from some more secure base. He claimed that it was possible to describe this base in ways which did not commit one to the existence of any physical objects, and named this epistemological basis the primary system. It consisted of neutral (that is, neither public nor private) qualia, such as loudness or greenness. These qualia form patterns in places and times, from which we postulate the more elaborate secondary system, a theory of 'visuo-tactual continuants'. Once this has been done the experiential basis is reinterpreted as consisting of states of the perceiver which are caused by the postulated physical objects. Critics have held that this view of perception is untrue to the facts of perceptual experience, which they claim is more likely to be veridically characterized as consisting in our awareness of the objects we are looking at, rather than these being postulated to exist on the basis of some more primitive sensory experience (see Perception §§1-2; Phenomenalism §1; Qualia §1; Sense-data §1).

4 Knowledge

In *The Problem of Knowledge* (1956) Ayer claimed that person *a*'s knowledge that *p* could be analysed into: (1) *p* being true; (2) *a* being sure that *p*; and (3) *a* having the right to be sure that *p*. Clause (1) is deemed necessary because we withdraw attributions of knowledge if we discover that the belief is false. Clause (2) is thought to be more or less obvious. Counterclaims trade on the fact that we often act on the basis of knowledge which we do not consciously admit. Clause (3) is inserted in order to avoid attributing knowledge to the lucky guesser - one who believes that *p*, and *p* happens to be true, but where the believer had no right to the belief. Ayer claimed that these three clauses were necessary and sufficient for knowledge.

Most of the subsequent discussion has concentrated on the sufficiency claim. Edmund Gettier (1963) produced examples which appeared to satisfy the three clauses but which did not seem to be a case of knowledge. One such example is that of a man, Smith, who has strong evidence for the belief (1) that Jones owns a Ford. On the sole basis of this evidence Smith forms the further belief (2) that Jones owns a Ford, or Brown is in Barcelona. Jones, as it happens, does not own a Ford, but Brown is in Barcelona. The evidence justifies Smith in believing (1), so justifies his belief in (2), and (2) is true. But, Gettier alleges, Smith does not *know* (2), so Ayer's analysis is wrong. Ayer thought that this type of example did not show that his analysis of knowledge was wrong, but simply that the notion of justification was trickier than initially thought. Once this had been sorted out, the counterexamples would be seen to involve unjustified beliefs (see Knowledge, concept of §4).

5 Ethics

One of the most controversial of the theses advocated by the early Ayer was that moral judgments lacked cognitive significance - they were not verifiable by any distinctive 'moral' facts or experience. They functioned, Ayer claimed, to express our emotions or attitudes. When we say 'Killing people is wrong', we are simply expressing a negative attitude to the actions of killing people. Analogously, when we say 'John is a good person' we are just evincing a positive attitude to John. This became known as the 'boo/hurrah' theory of ethics, as Ayer suggested that negative and positive moral judgments had the same cognitive status as saying 'boo!' or 'hurrah!'. Ayer continued to hold the emotivist theory, as it became known, with the slight emendation that the expression of attitude had a prescriptive element built into it. A positive moral judgment both expressed a 'pro-attitude' and enjoined others to have the same attitude. Ayer also made it clear that the attitude expressed was towards a *class* of acts, so when killing was condemned it was not a particular killing, but all killings, which were condemned.

Emotivism, and more generally the non-cognitivism about ethics which it espoused, has been widely discussed in the subsequent literature. One of the most difficult problems any such theory faces is to account for the status of moral expressions when they clearly do not have the force of expressing an emotion, as when they appear in conditionals. If somebody says, 'If stealing is wrong, then nobody ought to steal', there is no straightforward way of construing the meaning of the component sentence 'Stealing is wrong' as consisting just in the expression of emotion, since the person uttering the conditional may not have any negative attitude towards stealing. It is still a matter of some discussion as to whether this problem destroys the emotivist theory. An additional problem is whether the emotivist theory is accurate in its description of another facet of moral discourse, and that is moral argument. We seem to have moral disputes with one another about the correctness of our moral actions and judgments. On the emotivist theory these arguments are not about moral facts, so they were interpreted by Ayer as either being genuine arguments about nonmoral facts or not being genuine arguments (see Emotivism).

Ayer was puzzled as to whether we were morally responsible for our actions or not. It appears as though determinism, the thesis that all our actions are caused, destroys moral responsibility. We should not be held responsible for what we cannot help doing. On the other hand denying determinism did not seem to provide much support for responsibility either - making our actions independent of causal control made them independent of control by our minds, so again holding anybody morally responsible for what they cannot control seems misplaced. Ayer sought refuge in compatibilism, the thesis that determinism and freedom are reconcilable. He claimed that the mistake made by incompatibilists was to think that actions being caused entailed that they were compelled or constrained. For Ayer, causation was simply correlation of events, not involving any necessary connection between the events, and so an action could be caused without being compelled. Compulsion and free action were to be distinguished by the type of causation of involved. Ayer remained puzzled about whether this was sufficient for our ordinary concept of moral responsibility to have application (see Free will §1).

GRAHAM MacDONALD

List of works

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz; 2nd edn, 1946.(The first classic statement of Ayer's positivist philosophy. The second edition contains an important new introduction amending the verification principle.)

Ayer, A.J. (1937) 'Verification and Experience', *Proceedings of the Aristotelian Society* 46: 137-56.(An attack on conventionalist approaches to basic statements.)

Ayer, A.J. (1940) *The Foundations of Empirical Knowledge*, London: Macmillan.(Contains Ayer's most fully worked-out phenomenalism.)

Ayer, A.J. (1954) *Philosophical Essays*, London: Macmillan.(Essays on freedom, phenomenalism, basic propositions, utilitarianism, other minds, the past and ontology.)

Ayer, A.J. (1956) *The Problem of Knowledge*, London: Macmillan.(Formulates and defends Ayer's account of knowledge, with a discussion of scepticism.)

Ayer, A.J. (1963) *The Concept of a Person and Other Essays*, London: Macmillan.(Contains essays on truth, privacy and private languages, laws of nature, the concept of a person, and probability.)

Ayer, A.J. (1968) *The Origins of Pragmatism*, London: Macmillan.(On the American pragmatists Peirce, James and Dewey.)

Ayer, A.J. (1969) *Metaphysics and Common-Sense*, London: Macmillan.(Essays on knowledge, man as a subject for science, chance, philosophy and politics, existentialism, metaphysics, and Austin on the sense-datum theory.)

Ayer, A.J. (1971) *Russell and Moore: The Analytical Heritage*, London: Macmillan.(An extended essay on the origins of analytical philosophy.)
Ayer, A.J. (1972a) *Probability and Evidence*, London: Macmillan.(A concise presentation of Ayer's seminal work on the theory of probability and confirmation.)
Ayer, A.J. (1972b) *Bertrand Russell*, London: Fontana.(An introduction to Russell's philosophy.)
Ayer, A.J. (1973) *The Central Questions of Philosophy*, London: Weidenfeld.(An advanced introduction to major philosophical issues, such as perception, knowledge, freedom and morality.)
Ayer, A.J. (1977) *Part of my Life*, London: Collins.(An autobiographical account of Ayer's life up to his appointment to the Wykeham Chair of Logic at Oxford University.)
Ayer, A.J. (1980) *Hume*, Oxford: Oxford University Press.(A brief introduction to the philosophy of David Hume.)
Ayer, A.J. (1982) *Philosophy in the Twentieth Century*, London: Weidenfeld.(Ayer's views on the major developments in twentieth-century philosophy.)
Ayer, A.J. (1984) *More of My Life*, London: Collins.(The second and final part of Ayer's autobiography.)

References and further reading

Austin, J.L. (1962) *Sense and Sensibilia*, London: Oxford University Press.(Includes Austin's critique of Ayer's account of sense-data.)
Church, A. (1949) 'Review of *Language, Truth and Logic*', *Journal of Symbolic Logic*, 14: 52-3.(Includes Church's famous criticism of Ayer's criterion of meaningfulness.)
Foster, J. (1985) *A.J. Ayer*, London: Routledge & Kegan Paul.(An excellent book-length study of Ayer's philosophy, concentrating on verificationism and knowledge. Fairly rigorous.)
Gettier, E. (1963) 'Is Justified True Belief Knowledge?', *Analysis* 23: 121-3.(Classic criticism of Ayer's account of knowledge.)
Honderich, T. (1991) *Essays on A.J. Ayer*, Cambridge, Cambridge University Press.(A set of Royal Institute of Philosophy Lectures on aspects of the philosophy of Ayer.)
Lewis, D.K. (1988) 'Statements Partly About Observation', *Philosophical Papers* 17: 1-31.(A further attempt to resuscitate verificationism about meaning, including a response to Wright 1986.)
Macdonald, G. (ed.) (1979) *Perception and Identity*, London: Macmillan.(Essays presented to A.J. Ayer on his retirement from the Wykeham Chair of Logic, covering most of the topics discussed by Ayer. Includes Ayer's replies to the contributors.)
Macdonald, G. and Wright, C. (eds) (1986) *Fact, Science and Morality*, Oxford: Blackwell.(Essays commemorating the publication of *Language, Truth and Logic*.)
Quine, W.V. (1936) 'Truth by Convention', in O.H. Lee (ed.) *Philosophical Essays for A.N. Whitehead*, New York: Longmans.(An attack on the proposal that logical truth could be explained by reference to linguistic conventions.)
Quine, W.V. (1951) 'Two Dogmas of Empiricism', *Philosophical Review* 60: 20-43.(A challenge to the idea that distinguishing between analytic and synthetic truth could be philosophically rewarding.)
Wright, C. (1986) 'Scientific Realism, Observation and the Verification Principle', in G. Macdonald and C. Wright (eds) *Fact, Science and Morality*, Oxford: Blackwell, 1986, 247-74.(A sophisticated attempt to revive Ayer's verificationist criterion of meaningfulness.)
Wright, C. (1989) 'The Verification Principle: Another Puncture - Another Patch', *Mind* 98: 611-22.(A response to Lewis 1988.)

Bachelard, Gaston (1884-1962)

One indication of the originality of Bachelard's work is that he was famous for his writings both in the philosophy of science and on the poetic imagination. His work demonstrates his belief that the life of the masculine, work-day consciousness (animus), striving towards scientific objectivity through reasoning and the rectification of concepts, must be complemented by the life of a nocturnal, feminine consciousness (anima), seeking an expanded poetic subjectivity, as, in reverie, it creates the imaginary.

In common with other scientist-philosophers writing in the first half of the twentieth century, Bachelard reflected on the upheavals wrought by the introduction of relativity theory and quantum mechanics. The views at which he arrived were, however, unlike those of his contemporaries; he argued that the new science required a new, non-Cartesian epistemology, one which accommodated discontinuities (epistemological breaks) in the development of science. It was only after he had established himself as one of France's leading philosophers of science, by succeeding Abel Rey in the chair of history and philosophy of science at the Sorbonne, that Bachelard began to publish works on the poetic imagination. Here his trenchantly anti-theoretical stance was provocative. He rejected the role of literary critic and criticized literary criticism, focusing instead on reading images and on the creative imagination.

1 An unconventional philosopher

Gaston Bachelard did not enter philosophy through the standard French academic channels. Perhaps because of this, his work is unconventional, both in style and in range of subject matter. Bachelard was born and spent his childhood in Bar-sur-Aube, Champagne. From 1903 until the outbreak of the First World War he worked for the postal service and pursued scientific studies, receiving his *licence* in mathematics in 1912. He was drafted in 1914 and served in the army for the duration of the war. In 1919 he returned to Bar-sur-Aube to teach physics and chemistry at the local college. He had become interested in philosophy and received his *licence* in 1920, his *agrégation* in 1922 and his doctorate, written under the direction of Abel Rey and Léon Brunschvicg, in 1927. From 1930 until he moved to the Sorbonne in 1940 to succeed Rey in the chair of history and philosophy of science, he taught philosophy at the University of Dijon.

Bachelard's influence on philosophy in France has been much greater than accounts of contemporary French philosophy might suggest. He published twelve books on the philosophy of modern science, two on time and consciousness and nine on poetic imagination. These have been widely read, not least because they became an established part of the French university philosophy syllabus. When Bachelard retired from the Sorbonne in 1954, Georges Canghuilhem succeeded him in the chair of history and philosophy of science, and built on Bachelard's approach in his own work in the life sciences. Between them, Bachelard and Canghuilhem recast the disciplines of epistemology, history and philosophy of science in a way which, as Foucault has insisted, is essential to understanding not only his own work, but also that of Althusser and the followers of Lacan. In France, Bachelard has been criticized by postmodernists for remaining faithful to some of the methods of critical philosophy, and by Marxists for his humanism. His works on poetics inspired many of the French New Critics (see French philosophy of science §§1-2).

2 The new scientific spirit

To have taught science in the early part of the twentieth century is to have experienced dramatic changes in the character of physics and chemistry. Relativity theory and quantum mechanics did not represent simple additions to, or corrections of, existing science; their acceptance entailed the disruption of the whole framework of classical physics, requiring modifications in the concepts of space, time, causality and substance. These are the concepts which, according to Kant, are constitutive not merely of the framework of classical physics, but of our conception of the physical world. As such they are vital to the way in which the distinction is drawn between inner and outer, subject and object. Bachelard's epistemology is the result of absorbing the radical implications of these developments.

The fact that physics has reworked the very categories which Kant took to be a priori, grounded in the nature of the rational subject, invalidates any philosophy which starts from the presumption of a fixed rational framework,

whether found in formal logic, the structures of discourse or the rational subject. Similarly, Bachelard argues, science has reworked its empirical base. Sensory observation of given, natural objects has been replaced by the technologically mediated preparation of laboratory objects and phenomena. The objects of scientific investigation are made, not found; they are defined via their method of preparation or technical detection. We have passed from phenomenology to phenomeno-technique. Modern science thus requires us to put aside any epistemology erected on the foundation of the empirically given, whether sense-data or naturally occurring phenomena.

The epistemology of modern science cannot, therefore, be accommodated within a foundational epistemology, whether rationalist or empiricist; it is necessary to move to the open rationality of a non-Cartesian epistemology. As Bachelard reaches this conclusion he is led to introduce new concepts expressed in unfamiliar terminology: 'epistemological break' ('rupture'), 'epistemological value', 'epistemological obstacle', 'recurrent history' and the distinction between 'lapsed' and 'sanctioned' knowledge (see Incommensurability §1; Kuhn, T.S. §2-3).

Bachelard finds the clearest examples of what he calls 'dialectical reasoning' in mathematics. This is the reasoning characteristic of the open rationality of the new scientific spirit, which must reject the closed rationality of classical science as it moves beyond it and is aware of its own transitions. Bachelard's sense of dialectic is very much his own, and is not to be confused either with Hegelian or Marxist dialectic. A dialectical move, such as that from Euclidean to non-Euclidean geometry, takes one from a limited system, a framework which is closed in a certain respect, to one which is more general by being open in this respect. A dialectical development constitutes an 'epistemological break' because it introduces cognitive discontinuities, forcing far-reaching conceptual innovations on the one hand and reorganization and re-evaluation of cognitive fields on the other. In the case of geometry, the terms 'straight line' and 'space' continue to be used, but current concepts stand in no simple logical relation to their Euclidean/Newtonian predecessors. Although a straight line may still be the shortest distance between two points, the characteristics of lines with this property are now systematically related to the character of the space in which they occur. Moreover, a clear distinction is made between mathematical spaces and physical space, where previously there had been what, with hindsight, we judge to have been a complex but unclarified relation, presumed to be unproblematic only because it was taken for granted. In this sense the concept of space has been 'rectified' or 'corrected'.

The discontinuity involved in this epistemological break is not merely a disruption of logical relations; it requires a re-evaluation both of what was taken in the past to be geometrical knowledge, and of the relation between physics and mathematics. The 'epistemological value' attached to past proofs changes. The deductions are not discarded, but Euclid's theorems can no longer be treated as demonstrations of the properties of physical space. A dialectically advancing science will thus engage in periodic reassessments of the epistemological value to be attached to past science, or to what was in the past thought to be science, giving rise to what Bachelard calls 'recurrent history' of science. This history is evaluative; it evaluates the past of science from the standpoint of the present, distinguishing between those parts of past science which are 'sanctioned' (as having constituted progress) and those which are ignored because they are no longer valued ('lapsed science').

'Epistemological obstacles' are those (subjective) factors which contributed to the conviction that, for example, the parallel postulate was a necessary truth, something not open to questioning. They can only be identified with hindsight, after a break has been made and accepted as constituting a progressive move (a move towards greater objectivity). The factors which made the move difficult, which contributed to the closure of thought, then appear as factors which had to be overcome if progress was to be possible. These factors are necessarily hidden to consciousness while they are operative, since they result in the perception of something as self-evident, as beyond question. Epistemological obstacles thus have to be sought not in the cognitive foundations of a discipline, but among the various factors which play a role in shaping thought without our being aware of their operation. So epistemology has to concern itself with factors which condition and limit thought but which operate on a non-conscious, noncognitive level; it cannot restrict itself to considering only logical relations between concepts. Objective knowledge is progressively acquired by overcoming epistemological obstacles; this takes place on the borders of the conscious and the unconscious, between reflective theory and unreflective practice, between subject and object, and involves a continual redrawing of the line between them.

3 Dreamers and thinkers, images and concepts

To move beyond the boundaries set by current concepts, even the scientist must dream or engage in reveries. In

Bachelard's opinion, mathematics provides the space within which modern scientists dream. The two aspects of Bachelard's work come together in the figure of the creative thinker who must also be a dreamer and they do so in a way which gives a pivotal role to his highly original philosophy of mathematics, an aspect of his work which has received too little attention.

The axes of poetry and science are, Bachelard says, opposed in the sense that the very law of poetry is to go beyond thought, to take a voyage into the infinite, into an inner realm in which the unreal becomes real, in which images are forged and projected onto the external world. It is in this sense that mathematics is the poetry of science, and this is possible because, and to the extent that, here, as for Kant, there is a special relation between pure formal imagination and the construction of concepts. But even as scientists need to dream to go beyond, and ultimately extend, the boundaries of thought, so too their reveries may take them back to primordial themes, resistant to the teachings of scientific experiment, setting up resistances (epistemological obstacles) to psychological evolution. Drawing on elements of Jung's psychoanalytic theory, Bachelard wants to remind us of the power of elemental images and of the alchemist behind every engineer.

The life of concepts is essentially public. Within scientific communities conceptual boundaries are set by their logical interrelationships. These may be implicitly embedded in reasoning practices and rational standards, or explicitly codified in axioms, law, principles and definitions. The public role of concepts in the quest for objective knowledge is itself open to critical conceptual investigation. The life of images is, however, subjective and private (nocturnal). This is why Bachelard insists that the intellectual criticism of poetry can never lead to the centre where poetic images are formed, hence the image can only be studied through the image, by allowing images to gather in a state of reverie. Bachelard's own writing brilliantly evokes such gatherings of images. It is a form of writing which one cannot respect if one attempts to summarize its content. Bachelard concludes his *La poétique de la rêverie (Poetics of Reverie)* (1960) by saying, 'Since it was written in anima, we would wish that this simple book be read in anima'. And, as he explained at the beginning:

> It is *Anima* who dreams and sings. Dreaming and singing, that is the work of its solitude. Reverie - not the dream [*rêve*] - is the free expansion of all *anima*. It is doubtless with the reveries of his *anima* that the poet manages to give his *animus* ideas the structure of a song, the force of a song.
>
> From that point on, how can we read what the poet has written in an *anima* reverie without *anima* reverie? And that is how I justify not being able to read poets except by dreaming.

([1960] 1971: 67)

MARY TILES

List of works

Bachelard, G. (1928a) *Essai sur la connaissance approchée (Essay on Approximate Knowledge)*, Paris: Vrin. (Gives Bachelard's views on mathematics; these shaped his philosophy of science. He also indicates an approach to induction which differs significantly from anything developed in the analytic philosophy of science tradition.)

Bachelard, G. (1928b) *Étude sur l'évolution d'un problème de physique: la propagation thermique dans les solides (Study of the Evolution of a Problem in Physics: The Propogation of Heat in Solids)*, Paris: Vrin.(This gives a sense of Bachelard's early views as they emerge in the context of a study in the history of physics.)

Bachelard, G. (1929) *La valeur inductive de la relativité (The Indicative Value of Relativity)*, Paris: Vrin.(Here we find the first full exposition of the 'epistemological break (or rupture)'. Bachelard insists on the newness of the theory of relativity and on the way in which it breaks with past thought.)

Bachelard, G. (1932a) *Le pluralisme cohérent de la chimie moderne (The Coherent Pluralism of Modern Chemistry)*, Paris: Vrin.(An essay in historical epistemology.)

Bachelard, G. (1932b) *L'intuition de l'instant: étude sur la 'Siloë' de Gaston Roupnel (Intuition of the Insistent: A Study of Gaston Roupnel's 'Siloë')*, Paris: Gonthier.(Bachelard argues against Bergson's isolation of the subject from its world, rejecting his conception of pure consciousness and pure duration.)

Bachelard, G. (1933) *Les intuitions atomistiques: essai de classification (Atomisitic Intuitions: A Classification)*, Paris: Vrin.(Another example of the way in which Bachelard uses the history of science as a vehicle for exploring its epistemology.)

Bachelard, G. (1934) *Le nouvel esprit scientifique*, Paris: Presses Universitaires de France; trans. A. Goldhammer, *The New Scientific Spirit*, Boston, MA: Beacon Press, 1985.(A useful introduction to Bachelard's philosophy

of science. It is a forceful expression of the idea that scientific reality is dialectically related to scientific reason.)

Bachelard, G. (1936) *La dialectique de la durée (The Dialectic of Duration)*, Paris: Presses Universitaires de France.(In a further move in his polemic against Bergson, Bachelard turns to the adventures of consciousness. The self-proclaimed aim of the work is to present a propaedeutics for a philosophy of space.)

Bachelard, G. (1937) *L'expérience de l'espace dans la physique contemporaine (The Experience of Space in Contemporary Physics)*, Paris: Presses Universitaires de France.(Critically examines the role of space in substantiating realist views of science.)

Bachelard, G. (1938a) *La psychanalyse du feu*, Paris: Gallimard; trans. A.C.M. Ross, *The Psychoanalysis of Fire*, Boston, MA: Beacon Press, and London: Routledge & Kegan Paul, 1964.(Intended as a companion to 1938b. This marks a turning point in Bachelard's work; he begins to develop a theory of poetic imagination and at the same time discusses fire as an example of an epistemological obstacle.)

Bachelard, G. (1938b) *La formation de l'esprit scientifique: contribution à une psychanalyse de la connaissance objective (The Formation of the Scientific Spirit: A Contribution to a Psychoanalysis of Objective Knowledge)*, Paris: Vrin.(A pivotal work, dealing with the boundaries between subjective and objective in the construction of scientific knowledge. Here Bachelard introduces and develops the concept of an epistemological obstacle.)

Bachelard, G. (1939) *Lautréamont*, Paris: José Corti; trans. R. Duprée, *Lautréamont*, Dallas, TX: The Dallas Institute of Humanities and Culture Publications, 1984.(An exploration of analogies between the activities of scientists and poets. Mathematics and poetry are explicitly compared in the context of a discussion of Lautréamont's dynamic imagination.)

Bachelard, G. (1940) *La philosophie du non: essai d'une philosophie du nouvel esprit scientifique*, Paris: Presses Universitaires de France; trans. G.C. Waterston, *The Philosophy of No: A Philosophy of the New Scientific Mind*, New York: Orion Press, 1968.(Contains probably the clearest account of what Bachelard means by dialiectic and the dialectical development of modern science.)

Bachelard, G. (1942) *L'eau et les rêves: essai sur l'imagination de la matière*, Paris: José Corti; trans. E. Farrell, *Water and Dreams: An Essay on the Imagination of Matter*, Dallas, TX: The Dallas Institute of Humanities and Culture Publications, 1983.(A discussion of the relation between form and matter but not in the reasoned language of philosophical theory, rather in the subjective dream life which is shaped by the experience of handling liquids and materials. This leads Bachelard to the concept of the dynamic hand.)

Bachelard, G. (1943) *L'air et les songes: essai sur l'imagination du mouvement*, Paris: José Corti; trans. E. Farrell and F. Farrell, *Air and Dreams: An Essay on the Imagination of Movement*, Dallas, TX: The Dallas Institute of Humanities and Culture Publications, 1988.(Contains a discussion of Shelley's 'Prometheus Unbound' and of the metaphoric link between aspirations to greatness and the idea of taking wing, of soaring into the firmament.)

Bachelard, G. (1948a) *La terre et les rêveries du repos: essai sur les images de l'intimité (Earth and Reveries of Repose: An Essay on Images of Intimacy)*, Paris: José Corti.(In many ways this is a continuation and counterpoint to Bachelard 1942, but here the focus is on the matter which is imagined to underlie all form. This leads Bachelard to the notion of material repose.)

Bachelard, G. (1948b) *La terre et les rêveries de la volonté: essai sur l'imagination des forces (Earth and Reveries of Will: An Essay on the Imagination of Forces)*, Paris: José Corti.(An exploration focusing on the technically creative impulse and its origin in oneiric hypotheses.)

Bachelard, G. (1949) *Le rationalisme appliqué (Applied Rationalism)*, Paris: Presses Universitaires de France. (This work should be read in conjunction with Bachelard 1953. Together they give a sense of the power and originality and fruitfulness of Bachelard's mature philosophy of science.)

Bachelard, G. (1951) *L'activité rationaliste de la physique contemporaine (The Rationalist Activity of Contemporary Physics)*, Paris: Presses Universitaires de France.(An illustration of the concept of 'regional rationalism', first worked out in 1949.)

Bachelard, G. (1953) *Le matérialisme rationnel (Rational Materialism)*, Paris: Presses Universitaires de France. (Together with 1949 this work constitutes a sustained attack on idealism, Husserlian phenomenology and existentialism.)

Bachelard, G. (1957) *La poétique de l'espace*, Paris: Presses Universitaires de France; trans. M. Jolas, *The Poetics of Space*, New York: Orion Press, and Boston, MA: Beacon Press, 1969.(A counterpart to the earlier explorations of space in the scientific context. Follows up the insights of Bachelard 1929.)

Bachelard, G. (1960) *La poétique de la rêverie*, Paris: Press Universitaires de France; trans. D. Russell, *The Poetics of Reverie*, New York: Orion Press, and Boston, MA: Beacon Press, 1971. (Exploration of the reverie - something that played an increasing role in Bachelard's later works.)

Bachelard, G. (1961) *La flamme d'une chandelle*, Paris: Presses Universitaires de France; trans. J. Caldwell, *The Flame of a Candle*, Dallas, TX: The Dallas Institute of Humanities and Culture Publications, 1988.(A discussion of the relations between humans and things, subjects and objects which rests on uncompromising values concerning the nature of those relations that make us more fully human.)

References and further reading

Lecourt, D. (1975) *Marxism and Epistemology: Bachelard, Canguilhem and Foucault*, trans. B. Brewster, London: New Left Books.(Articulates Marxist criticisms of aspects of Bachelard's epistemology.)

McAllester, M. (ed.) (1989) *The Philosophy and Poetics of Gaston Bachelard*, Washington, DC: Center for Advanced Research in Phenomenology and University Press of America.(Useful collections of papers covering Bachelard's epistemology, his poetics and the relation between them.)

Roy, J.-P. (1977) *Bachelard ou le concept contre l'image*, Montreal, Que.: Les Presses de l'Université de Montréal.(Discusses the relation between Bachelard's epistemology and his poetics.)

Tiles, M. (1985) *Bachelard: Science and Objectivity*, Cambridge and New York: Cambridge University Press.(An introduction to Bachelard's philosophy of science, which compares his approach to that of analytic philosophy of science.)

Bacon, Francis (1561-1626)

Along with Descartes, Bacon was the most original and most profound of the intellectual reformers of the sixteenth and seventeenth centuries. He had little respect for the work of his predecessors, which he saw as having been vitiated by a misplaced reverence for authority, and a consequent neglect of experience. Bacon's dream was one of power over nature, based on experiment, embodied in appropriate institutions and used for the amelioration of human life; this could be achieved only if the rational speculations of philosophers were united with the craft-skills employed in the practical arts.

The route to success lay in a new method, one based not on deductive logic or mathematics, but on eliminative induction. This method would draw on data extracted from extensive and elaborately constructed natural histories. Unlike the old induction by simple enumeration of the logic textbooks, it would be able to make use of negative as well as positive instances, allowing conclusions to be established with certainty, and thus enabling a firm and lasting structure of knowledge to be built.

Bacon never completed his project, and even the account of the new method in theNovum Organum (1620) remained unfinished. His writings nevertheless had an immense influence on later seventeenth-century thinkers, above all in stimulating the belief that natural philosophy ought to be founded on a systematic programme of experiment. Perhaps his most enduring legacy, however, has been the modern concept of technology - the union of rational theory and empirical practice - and its application to human welfare.

1 Life

Francis Bacon was born into the political elite of Elizabethan England. His father, Nicholas, was Lord Keeper; his mother, Anne, sister-in-law to Lord Burghley, the Lord Treasurer. Much of Bacon's career and even some aspects of his philosophy can best be understood as resulting from an upbringing which made him familiar with the exercise of power, and the wealth that came with it. His perspective is always that of an insider, but of one who experienced considerable difficulty in establishing his own position as such.

In 1573 Bacon was admitted to Trinity College, Cambridge. In later recollection at least, he found little to admire in the Aristotelian philosophy to which he was introduced, and still less in the writings of such authors as Peter Ramus, who were becoming fashionable alternatives (see Aristotle; Aristotelianism, Renaissance; Ramus, P.). As was usual with undergraduates of his social rank, he did not take a degree. In 1576 he returned to London to train as a barrister at Gray's Inn, an institution with which he was to maintain a much more enduring connection. His father died in 1579, leaving him with only a modest inheritance. Throughout his life Bacon spent freely and lived beyond his income; quite apart from considerable personal ambition, much of his pursuit of office can be seen as an attempt to repair chronic indebtedness.

Though he was elected to successive parliaments from 1581 onwards, Bacon's career did not flourish under Queen Elizabeth, who recognized his abilities but seems to have found his personality unappealing. Burghley was more concerned to advance the career of his own son Robert, later Earl of Salisbury, and Bacon attached himself to Elizabeth's last favourite, the brilliant but insubstantial Earl of Essex. Essex's attempt in 1601 to restore his fortunes by staging an insurrection proved a complete fiasco, and made him liable to prosecution for treason. Bacon adroitly changed sides and prosecuted his former patron with a skill and vigour which provided ample confirmation both of his remarkable talents and of a fundamental coldness of character.

The accession of James I in 1603 presented the prospect - initially unfulfilled - of professional advancement. Bacon was knighted soon after the King's arrival in London, but he had to wait until 1607 before being given his first important office, that of Solicitor General. It was only after the death of Salisbury in 1612 that promotion became truly rapid: in 1613 he was appointed Attorney General, in 1617 Lord Keeper, and in 1618 Lord Chancellor. This last office brought admission to the peerage, first as Baron Verulam (1618) and then as Viscount St Albans (1621).

Bacon's fall was precipitous and catastrophic, though not entirely unpredictable. He had supplemented the income from his office by taking payments from those whose cases he heard, and though this was far from unprecedented it did make him vulnerable to attack. He was also important enough to be a substantial sacrifice to an angry House

of Commons, without being so close to James that he could not be dispensed with. At the beginning of May 1621 Bacon was deprived of office, imprisoned - albeit for only a few days - in the Tower of London, fined £40,000, barred from court and prevented from taking his place in the House of Lords.

Despite his best efforts, Bacon never returned to favour. He spent his last five years in retirement, writing incessantly - at first with the hope of regaining office, or at least influence, and then merely to leave a testament to posterity. He died on Easter Day 1626, according to John Aubrey (who had the story from Hobbes) from a cold contracted after an experiment of stuffing a chicken with snow. As has often been remarked, it was a fitting end for so fervent an advocate of experimental science.

2 Works

During the first two decades of his adult life Bacon wrote little, or at least little that survives; it was however in this period that his outlook and basic ideas were formed - certainly by the early 1590s, and probably earlier still; in 1625 he mentioned to a correspondent that forty years earlier he had advocated the reform of learning in a work (now lost) entitled *Temporis Partus Maximus (The Greatest Birth of Time)*. The direction of Bacon's interests is apparent in a letter of 1592, written to Lord Burghley, in which he (rather disingenuously) disclaimed any political ambition while simultaneously indicating the scope of his intellectual projects:

> I confess that I have as vast contemplative ends, as I have moderate civil ends: for I have taken all knowledge to be my province; and if I could purge it of two sorts of rovers, whereof the one with frivolous disputations, confutations, and verbosities, the other with blind experiments and auricular traditions and impostures, hath committed so many spoils, I hope I should bring in industrious observations, grounded conclusions, and profitable inventions and discoveries; the best state of that province.
>
> (1857-74 VIII: 109)

These themes, developed and articulated, were to preoccupy Bacon for the remainder of his life. No echo of them was however to appear in print for several years. Apart from some political tracts, the only one of Bacon's writings to be published during Elizabeth's reign was the first edition of the *Essays* (1597); the only portion of this volume of any philosophical significance is a short tract on 'The Colours of Good and Evil', which provides early evidence of Bacon's lifelong interest in fallacies and the pathology of the intellect.

The accession of James I stimulated a new burst of literary activity, of which the most visible result was *The Advancement of Learning* (1605), dedicated to the King and evidently written in the (unfulfilled) hope of munificent royal patronage. This was not the only project to have occupied Bacon's attention during the first years of the new reign. A large number of fragmentary treatises have survived, some in English, some in Latin. Several have strange, enigmatic titles: *Temporis Partus Masculus (The Masculine Birth of Time)*, *Valerius Terminus of the Interpretation of Nature with the Annotations of Hermes Stella*, *Filum Labyrinthi (The Thread of the Labyrinth)*. Others are more prosaic: *Redargutio Philosophiarum (The Refutation of Philosophies)*, *Cogitata et Visa de Interpretatione Naturae (Thoughts and Conclusions on the Interpretation of Nature)*. The diversity of the literary form displayed by these works is as striking as their unity of message: Bacon knew at least in outline what he wanted to say, but was undecided as to the most appropriate form in which to say it.

The last of these fragments probably dates from around 1608. For the next twelve years Bacon was increasingly busy with his official duties, and much of the time that remained was spent drafting and redrafting the *Novum Organum*. He did however find time to publish a second expanded edition of the *Essays* (1612) and one new work, *De Sapientia Veterum (On the Wisdom of the Ancients)* (1609), an interpretation of ancient myths as allegories of political and physical doctrine. The same pattern of thought can be found in the unpublished *De Principiis atque Originibus (On Principles and Origins)* (*c*.1610-12?), which also shows the considerable influence of Bernardino Telesio on Bacon's physical doctrines, as do two other works written around 1612, *Descriptio Globi Intellectualis (A Description of the Intellectual Globe)* and *Thema Coeli (Theory of the Heavens)*, both left unfinished and unpublished.

The first instalment of Bacon's chief work, the *Instauratio Magna*, was eventually published with appropriate magnificence in 1620, when Bacon was at the pinnacle of his success. The whole work was to contain six parts, but all that appeared at this stage were a general preface, an outline of the project as a whole (the *Distributio*

Operis), an incomplete section of the second part (the *Novum Organum*), and a short *Parasceve ad Historiam Naturalem et Experimentalem (Preparative towards a Natural and Experimental History)*.

In the years that followed, Bacon went some way towards filling the lacunae in his original plan. The missing first part was supplied in 1623 by *De Dignitate et Augmentis Scientarum*, a revised and greatly extended translation of *The Advancement of Learning*. Despite its evident incompleteness, nothing more was added to the *Novum Organum*; most of Bacon's efforts went into the natural histories intended to fill Part III, which he rather optimistically planned to produce at the rate of one per month. In the event only two were published before his death: on winds (*Historia Ventorum* 1622) and on life and death (*Historia Vitae et Mortis* 1623), although a work on the condensation and rarefaction of materials (*Historia Densi et Rari*) was also completed in 1623. Bacon's executors ignored this - it eventually appeared in 1658 - but did publish the *Sylva Sylvarum* (1627), a natural history in English filled with some very dubious material, which proved very popular during the remainder of the century, but which provided much material for Bacon's nineteenth-century detractors.

The final three parts of *Instauratio Magna* were never written apart from short prefaces to parts four and five (*Works* II: 687-92). The first of these, *Ladder of the Intellect*, was to contain actual examples of the new method in operation - something closer to perfection than the mere sketches provided in the *Novum Organum*. Part V, *Forerunners, or Anticipations of the Second Philosophy*, would, by contrast, exhibit discoveries made independently of the method, by the ordinary workings of the understanding. The content of the final part, the *Second Philosophy or Active Science*, can only be conjectured; one may suspect that Bacon himself had no very precise idea of what it would contain.

Perhaps the best picture of Bacon's final vision can be found in a work of a very different kind, published in the volume containing the *Sylva Sylvarum* but of uncertain date. *The New Atlantis* is an account of an imaginary voyage to an island in the Pacific Ocean, and of the scientific institution, Salomon's House, found there. Like most utopian narratives, this is deeply revealing of its author and provides the fullest picture we have of Bacon's vision of a reformed, active science, and of the kind of institution that he saw as necessary to its flourishing. It also had a profound influence both on the millennialist, visionary Baconianism of the 1640s and on the founders and early practice of the Royal Society.

3 The division of learning

The Advancement of Learning contains two books, the first on the dignity of learning and the reasons for the discredit with which it was often regarded, the second and much longer on the classification of its various branches; in the 1623 translation this latter was expanded further, and divided into eight books.

The primary division of the branches of learning reflects the faculties of the human mind: history corresponds to memory, poetry to imagination, and philosophy to reason. Philosophy itself has three subdivisions: divine philosophy or natural theology, natural philosophy, and human philosophy, this last including the doctrine of the soul, logic, rhetoric, ethics and politics). Metaphysics is a branch of natural philosophy, concerned with formal and final causes, in contrast with physics which studies the material and the efficient. Metaphysics is a more general and more abstract discipline than physics, and rests on it, just as physics in turn rests on a foundation of natural history. The image is that of a pyramid whose vertex is the summary law of nature, known to God but perhaps beyond the bounds of human enquiry.

Rather unusually, Bacon made a distinction between metaphysics and *philosophia prima* - primitive or summary philosophy. The three main subdivisions of philosophy are not like lines meeting at a point, but like branches of a tree that join in a common stem. Arboreal metaphors of this kind may appear to suggest the Cartesian picture of science, in which the trunk of physics grows out of and is sustained by the roots of an a priori metaphysical system, but the reality is quite different. Bacon's *philosophia prima* is a mere receptacle for such miscellaneous principles as have applications in several different disciplines - for example, that the force of an agent is increased by the reaction of a contrary, a rule with applications in both physics and politics.

Bacon's most important innovation was, however, the close linking of theoretical and practical disciplines. In the Aristotelian tradition these had been kept quite separate, but now (within natural philosophy at least) each speculative discipline was to have its operative counterpart: corresponding to physics there would be mechanics; corresponding to metaphysics, natural magic. Bacon had no illusions about the pervasive fraudulence of the

magical tradition, but - as in the parallel case of astrology - he sought reform, not abolition (see Alchemy).

This close association of theory and practice was of the utmost importance: Bacon saw the dismal record of earlier natural philosophy as stemming very largely from their divorce. The practitioners of the applied arts had made what progress they had in a purely empirical way, unaided by any method, while the philosophers - especially, although not exclusively, the schoolmen in the universities - had disdained experience and, like spiders, had spun metaphysical cobwebs out of their own insides. The only hope of progress lay in uniting the two approaches.

4 The new logic

The *Novum Organum* has had far fewer readers than either the *Essays* or *The Advancement of Learning*, partly because of its more difficult subject matter, and partly because it was written in Latin; it is, however, Bacon's most remarkable achievement, and the one which he himself regarded most highly. It cost him considerable trouble - William Rawley, his chaplain, described having seen no fewer than a dozen drafts revised year by year in the decade preceding publication. Bacon's chosen form is the aphorism: initially these are short and highly compressed, but as the work proceeds they grow longer. In the second book, clearly less thoroughly revised, Bacon's grip slackens and then loosens altogether, and the aphoristic form is abandoned except in appearance.

As its title makes plain, the *Novum Organum* was intended as an account of a new logic, designed to replace the Aristotelian syllogistic which Bacon saw as having hampered and indeed corrupted the investigation of nature. The full exposition of this is found in Book II; Book I contains a survey of the task and its difficulties.

The basic themes of the *Novum Organum* are set out in the first three aphorisms:

> Man, being the servant and interpreter of nature, can only do and understand so much… as he has observed in fact or in thought of the order of nature: beyond this he neither knows anything nor can do anything.
>
> Neither the naked hand nor the understanding left to itself can effect much. It is by instruments and helps that the work is done, which are as much wanted for the understanding as for the hand. And as the instruments of the hand either give motion or guide it, so the instruments of the mind supply either suggestions for the understanding or cautions.
>
> Human knowledge and human power meet in one, for where the cause is not known the effect cannot be produced. Nature to be commanded must be obeyed; and that which in contemplation is as the cause is in operation as the rule.

(Bacon 1620: i.1-3)

Natural philosophy needs to begin with observation. Though Bacon sharply separated himself from those whom he classed as 'empirics', his objection to them lay in their lack of method and consequent recourse to unsystematic experimentation, not in their reliance on experience itself. Method is absolutely essential: unmethodical experimentation is mere groping in the dark, and is no more likely to produce results than digging for buried treasure on a purely random basis.

It is an essential feature of the new method that it can be openly described, explained and taught. The new reformed science is seen as an essentially collective activity; though undoubtedly presupposing a certain minimum of intelligence in its operatives, such an enterprise does not require, and is therefore not dependent on, the appearance of individual genius:

> But the course I propose for the discovery of sciences is such as leaves but little to the acuteness and strength of wits, but places all wits and understandings nearly on a level. For as in the drawing of a straight line or perfect circle, much depends on the steadiness and practice of the hand, if it be done by aim of hand only, but with the aid of rule or compass, little or nothing; so it is exactly with my plan.

(Bacon 1620: i.61)

There is therefore nothing intuitionistic about Bacon's approach, nothing at all resembling the Cartesian reliance on clear and distinct ideas. Bacon distrusted any appeal to the supposedly self-evident at the outset of any enquiry. Validation could only be retrospective: it was the ability of a theory to endow its holders with power over nature that provided the best, and indeed the only genuinely satisfactory, evidence for its truth.

Previous attempts at discovery had failed because men had either complacently supposed the mind already to be adequately equipped for the task, or else had despaired altogether. Nature is comprehensible, but its subtlety far exceeds that of the human mind. In order for anything to be achieved, a new logic based not on the *anticipation* but on the *interpretation* of nature needs to be brought into use.

This contrast between anticipation and interpretation is central to Bacon's conception of his project. Anticipations are not hypotheses, but rather 'the voluntary collections that the mind maketh of knowledge; which is every man's reason' ([*c*.1603] 1857-74 III: 244). The root idea is one of superficiality: these are the notions of 'folk physics' - popular ordinary-language concepts such as arise in the ordinary conduct of life, sometimes refined and made more abstract by the labours of philosophers, but not fundamentally altered. 'There is no stronger or truer reason why the philosophy we have is barren of effects than this, that it has caught at the subtlety of common [*vulgarium*] words and notions, and has not attempted to pursue or investigate the subtlety of nature' ([*c*.1607] 1857-74 V: 421).

It was the all-pervasive unsoundness of the concepts used that made the old logic useless as a tool for the investigation of nature. Syllogisms incorporating confused and badly abstracted terms merely propagate error without supplying any means of correcting it; more generally, the teaching of deductive logic encourages the natural tendency of the mind to ascend hurriedly and without due examination to propositions of great generality, and then to regard these as securely established when investigating further. Bacon's method requires not the liberation but the regulation of the intellect, which *'must not… be supplied with wings, but rather hung with weights, to keep it from leaping and flying'* (1620: i.104).

Just as syllogisms are useless for any enquiry into nature, so too is the induction by simple enumeration described in logic textbooks. Bacon consistently regarded this with contempt - 'childish' was his favourite term of abuse. It operated on the surface of things, employing 'popular' notions, and was for that reason incapable of delivering certainty. Bacon was no fallibilist, prepared to settle for a natural philosophy of conjectures and merely provisional conclusions. Certainty was quite as important for him as it would be for Descartes, but what he was looking for was certainty of a very different kind - not immunity from sceptical doubt, but complete reliability. This could be furnished by induction, but it would have to be induction of a new and much more elaborate kind, one that could make use of negative as well as positive instances.

5 The idols of the mind

Before the new logic could be put to use, the weaknesses of the human mind which it was designed to correct or evade needed to be analysed. The central section of Book I is a counterpart to the analyses of sophistical reasoning provided in the logic textbooks. What emerged, however, was not merely a list of inductive fallacies, but rather one of the most memorable and original parts of Bacon's system.

Bacon distinguished four classes of idols. The 'Idols of the Tribe' arise from the limitations of human nature; they can be allowed for and guarded against, but not removed entirely. Bacon had in mind such weaknesses as the tendency to suppose more regularity than actually exists, to be over-influenced by the imagination, and even more by hopes and desires. A very different kind of limitation arises from the dullness of the senses. Bacon had no sympathy with radical sceptical doubts of the kind that were to preoccupy Descartes, but he was acutely aware of the weakness of the human senses, and of their complete incapacity to discern the secret workings of nature. The problem was not one to be abandoned to sceptical despair or solved by metaphysical validation. Some assistance could be gained from the use of instruments, but the real solution lay in experimental design. Hidden processes would be linked with observable consequences, and an experimental determination of the latter would reveal the nature of the former.

The 'Idols of the Cave' arise from the idiosyncrasies of individuals, either natural or implanted by education. Some minds are good at seeing distant resemblances, others at making fine distinctions; some are attracted to ancient wisdom, or what might pass for it, others only to novelty; almost everyone is influenced by those disciplines which they know well, and even more by those to which they have contributed.

The 'Idols of the Forum' (or 'Idols of the Market Place') arise from the deficiencies of human speech. Bacon had no respect for the categories of ordinary language, or the habitual thought-patterns of the uneducated; 'popular' is in his lexicon almost invariably a term of disparagement. Words devised for the ordinary purposes of life cannot

provide a satisfactory vocabulary for natural philosophy, and attempts to remedy the situation by making definitions achieve nothing: words are defined by other words, which themselves share the same defects.

These three classes of idols can be guarded against and to some extent allowed for, but never extirpated entirely. The fourth class is in this respect different. This consists of the 'Idols of the Theatre' - the point of the name was that rival philosophies were like stage-plays, with different casts and different plots, but all equally fictitious. The potential variety of such systems is clearly unlimited, but Bacon distinguished three main types. The natural philosophy of Aristotle and his followers was corrupted partly by logic, and partly by a reliance on common notions - popular conceptions quite unsuited to the task in hand. The empirical school (exemplified by the alchemists, but also including William Gilbert who investigated magnetism) was misled by too narrow a line of experimental enquiry: restricted ranges of data fill the imagination and lead to one-sided accounts of the world in chemical or magnetic terms. Platonism (Bacon had in mind not so much the doctrines of Plato himself - whom he generally treats with respect - as the Platonism of his own era) (see Platonism, Renaissance) was worst affected of all, being corrupted by theology and superstition. Bacon's own religious views are by no means easy to discern and have been very diversely interpreted, but one thing that is abundantly clear is that he was wholly opposed to the intrusion of religious doctrines, Christian or non-Christian, into natural philosophy; the result of allowing this to happen was a corruption of both, into a superstitious philosophy and a heretical religion.

6 Induction

Bacon's methodological proposals occupy Book II of the *Novum Organum*. The first stage in any investigation is the gathering together of a natural and experimental history. This might be quite broad in scope - for example, the history of heat in aphorisms 11-18 - but it could be much more narrowly focused: Bacon's own examples include histories of the rainbow, of honey and of wax. The idea of a natural history was an old one, going back through numerous Renaissance and medieval encyclopedias to Pliny, and ultimately to Aristotle's *Historia Animalium*. Bacon, however, made an innovation of crucial importance. His histories would record not only material gathered from the ordinary workings of nature, but also novel phenomena generated by human activity. In the Aristotelian tradition such artefacts would have been discounted as inappropriate material for investigation; Bacon, however, saw them not merely as legitimate subjects of enquiry, but as especially valuable: 'by the help and ministry of man a new face of bodies, another universe or theatre of things, comes into view' (1857-74 IV: 253). Nature was to be put to the question - a contemporary euphemism for torture.

Histories of this kind could not be assembled quickly, and the whole project would clearly absorb a very large amount of labour and money. Bacon was acutely aware of this, but could see no alternative. The human understanding needed to be purged and cleansed, and this had to be done not by any Platonic (or Cartesian) detachment from the data of the senses, but by an immersion into the world of experience in its full individuality and variety. Bacon was a good nominalist in the English tradition: for him, individuals alone are real and our most reliable cognitions are our direct sensory awareness of them. Withdrawal to a world of abstract objects supposedly accessible to reason leads merely to illusion and the enunciation of empty generalities; for Bacon the word 'abstract' - like 'popular' - almost invariably carries negative connotations.

We have to begin, therefore, with particulars; we have also to begin with as full a range of particulars as possible. Bacon did not require all this data to be correct, though manifestly false material ought to be kept out where possible, and dubious reports marked as such. Some falsehoods were bound to creep in, but these could be dealt with; what could not be dealt with were biases which affected the whole history. Initial attempts to impose criteria of relevance had therefore to be outlawed altogether.

Most histories would contain an immense quantity of data - far too much for any individual human mind to grasp as a whole - and an ordering of this material into some kind of structure was essential. Bacon proposed the use of three tables: first a 'Table of Essence and Presence', listing all the situations in which the nature under investigation is present; then a 'Table of Deviation or Absence in Proximity', describing all those situations which are as close as possible to those in the first table but where the nature under investigation is absent; and finally a 'Table of Degrees or Comparison', a list of those situations where the nature in question varies in intensity, together with details of the circumstances accompanying that variation.

When first drawn up, the second and third tables would both, in general, be incomplete in that they would contain

gaps corresponding to entries in the first. One of the chief functions of experiment was to remedy these defects: for example, given that the rays of the sun can be concentrated by a convex lens, a trial should be made to see whether such lenses can produce heat by focusing the rays of the moon, or any rays proceeding from heated stones or vessels containing boiling water.

When the tables have been drawn up it is possible to begin the inductive process itself:

> The first work therefore of true induction…is the rejection or exclusion of the several natures which are not found in some instance when the given nature is present, or are found in some instance where the given nature is absent, or are found to increase in some instance where the given nature decreases, or to decrease where the given nature increases.
>
> (1620: ii.16)

Only when this process of exclusion has been completed will it be possible to grasp the true essence (or form, to use Bacon's own term) of the nature in question.

This method clearly rests on several presuppositions, of which the most fundamental is a principle of limited variety. Though the world as we experience it appears unendingly varied, all this complexity arises from the combination of a finite, and indeed quite small, number of simple natures. There is an alphabet of nature, which cannot be guessed or discovered by speculation, but which will start to be revealed once the correct investigatory procedures are employed. The time needed is indeed not merely finite but quite short: once the natural histories are complete, the unearthing of all the secrets of nature will require no more than a few years.

Bacon also assumed there to be a direct one-to-one correlation between natures and the forms from which they arise. He was aware that critics might deny this and maintain (for example) that the heat of the heavenly bodies and of fire, or the red in a rose and in a rainbow, are only apparently similar, having quite different causes in reality. Bacon firmly denied this - however apparently heterogeneous, these things agree in the forms or laws which govern heat and redness; indeed even such diverse modes of death as by drowning, by hanging and by stabbing agree in the form or law which governs death.

This way of thinking reinforces a tendency already present in the alchemical tradition of considering bodies as collections of simple natures, each explicable (and therefore reproducible) in isolation. Bacon certainly did think in this way: gold is yellow, heavy, ductile, fixed (that is, unaffected chemically by fire) and so on. Whoever knows the forms of these natures can attempt to join them together in a single body, and thereby transform that body into gold. At other times, however, Bacon seems to have recognized that forms are seldom independent: 'since every body contains in itself many forms of natures united together in a concrete state, the result is that they severally crush, depress, break, and enthrall one another, and thus the individual forms are obscured' (1620: ii.24). They are not, however, hidden altogether: since expansion is part of the form of heat, all heated bodies must expand; but while the expansion of air is easily noticed, that of iron is less manifest to the senses.

The justification of the principles of limited variety and of the direct correlation of forms and natures could always be postponed; another problem, however, had to be faced at the outset. Exclusion involves the rejection of simple natures, 'and if we do not yet possess sound and true notions of simple natures, how can the process of Exclusion be made accurate?' (1620: ii.19). The old logic had proved inadequate because of this deficiency; what grounds are there for supposing that the replacement would fare any better?

Bacon was acutely aware of this problem and of the difficulty it posed for his project. His solution was to propose a series of supports of induction: the account of these occupies the last part of Book II of the *Novum Organum* and is (characteristically) lengthy, elaborate and unfinished; indeed all he managed to describe was the first of his nine kinds of support, the 'Prerogatives of Instances', of which he distinguished no fewer than twenty-seven different varieties. Bacon's account of these exhibits, perhaps more clearly than any other passage in his writings, the distinctive strengths and weaknesses of his mind. The discussion is often shrewd and sometimes much more - the *instantia crucis* has passed into modern science, under the name of a crucial experiment - but Bacon's addiction to elaborate systems of classification and portentous schemes of nomenclature is frequently apparent, above all in the nineteen species of motion described in aphorism 48. The immensity of his intellectual distance from such contemporaries as Galileo is nowhere more apparent than it is here.

7 Natural philosophy

Bacon's intellectual gifts were remarkable, but they were not those of a scientist. He was a lawyer, and it was here, as well as with human affairs in general, that his real area of expertise can be found. He was widely read in natural philosophy, but his approach remained that of an outsider, albeit a shrewd and exceedingly intelligent one. These limitations become particularly apparent when Bacon turned to astronomy, the most highly developed of all contemporary scientific disciplines. He rejected Copernicanism and, although he saw many of the weaknesses of the inherited astronomical tradition, unlike Kepler he had only vague and quite unhelpful ideas about how it might be reformed.

Bacon's own physics was fundamentally non-mechanistic. Bodies contain two types of matter - tangible and spiritual - and the operation of the latter, although never explained clearly, is certainly not conceived in mechanistic terms. Bacon did, however, employ several ideas that were to be taken over by the mechanical philosophers who followed him, in particular that the observable qualities of bodies are to be explained by the constitution of their internal parts. Glass can be made white by being crushed into tiny fragments, and water white by being beaten into foam; heat is not a scholastic real quality but a kind of motion. Later Baconians such as Boyle and Hooke were able to take over these ideas and express them in more unambiguously mechanistic terms.

8 Bacon's influence

Bacon's philosophical writings met with little appreciation in England during the 1620s and 1630s. Admirers of the older learning, from James I downwards, were for the most part uncomprehending, and the one major scientist then practising, William Harvey, was brutally dismissive. Bacon had more impact in France, where he was carefully read by Mersenne, Gassendi and Descartes, but even they only responded to selected parts of the system, notably the 'Idols' and the appeal to experiment.

The political turmoil in Britain in the 1640s stimulated a new interest in Bacon's thought, both among the advocates of universal reform like Samuel Hartlib, and among such natural philosophers as Robert Boyle and Robert Hooke. Baconianism indeed became the official philosophy of the Royal Society, celebrated in Thomas Sprat's semi-official *History* (1667). The hopes thus stimulated, however, proved difficult to satisfy. Newton paid little attention to Bacon, and the *Principia* was an achievement utterly unlike anything projected in the *Novum Organum*. Locke's debt was rather greater, especially in *The Conduct of the Understanding*, but by the early eighteenth century interest in Baconianism had started to decline.

Following the example of Voltaire, the French encyclopedists treated Bacon with great respect as an empirical, essentially secular thinker, to be contrasted favourably with Descartes who was now seen as scientifically discredited and too deferential to the Church. In Britain Bacon was ignored by Hume, but admired by Reid who helped create a widely influential methodological synthesis of Baconian and Newtonian ideas.

The Baconian revival reached its climax in the second quarter of the nineteenth century. Sir John Herschel's *Preliminary Discourse on the Study of Natural Philosophy* (1830) was a thorough attempt to recast Baconianism in a form compatible with contemporary science. John Stuart Mill and William Whewell, though disagreeing about almost everything, both acknowledged a deep debt to Bacon, and to the inductive method of science. The most accessible introduction to early-Victorian attitudes towards Bacon is however provided by Macaulay's essay 'Lord Bacon' (1837). Though respectful towards Bacon's thought, Macaulay took a less favourable view of his character, and it was in response to his account that James Spedding undertook the labours that led to his *Life and Letters* (1857-74: vols VIII-XIV), and to the critical edition of Bacon's works produced jointly with R.L. Ellis and D.D. Heath.

In the latter part of the nineteenth century, Bacon's reputation as a methodologist began to decline. The trend continued after 1900, Bacon's reputation reaching its nadir in the mid-century when Karl Popper proposed a method for science that eschewed induction altogether, and historians such as Alexandre Koyré offered accounts of the scientific revolution that made Bacon's contribution utterly marginal. Since then there has been a modest revival, but Bacon has still not recovered an assured place in the philosophical canon.

See also: Induction, epistemic issues in; Inductive inference; Technology, philosophy of

J.R. MILTON

List of Works

Bacon, F. (1857-74) *The Works of Francis Bacon*, ed. J. Spedding, R.L. Ellis and D.D. Heath, London: Longmans, 14 vols.(The standard edition, with prefaces that remain of great value. Volumes I-V contain the *Philosophical Works*, VI-VII the *Literary and Professional Works*, and VIII-XIV the *Life and Letters*.)

Bacon, F. (1903) *The Philosophical Works of Francis Bacon*, ed. J. M. Robertson, London: George Routledge & Sons.(A convenient one-volume abridgement of the Ellis and Spedding edition.)

Bacon, F. (1996-) *The Oxford Francis Bacon*, Oxford: Clarendon Press.(Of the twelve planned volumes, volume 6 has appeared - *Philosophical Studies*, *c.1611-1619*, ed. G. Rees. This contains Latin texts and English translations of a group of works on natural philosophy: *Phaenomena Universi*, *De Fluxu et Refluxu Maris*, *Descriptio Globi Intellectualis*, *Thema Coeli* and *De Viis Mortis*.)

Bacon, F. (1597) *Essayes or Counsell, Civill and Morall*, London; repr. Oxford: Clarendon Press, 1985.(Later considerably expanded editions were published in 1612 and 1625.)

Bacon, F. (1605) *The Advancement of Learning*; repr. London: Dent, 1973.(A greatly expanded Latin translation was published in 1623 as *De Dignate et Augmentis Scientarum*.)

Bacon, F. (1609) *De Sapientia Veterum (Wisdom of the Ancients)*, London; Kila, MT: Kessinger Publishing, 1992.(An exposition of ancient myths, taken as allegories of philosophical doctrines.)

Bacon, F. (*c*.1612) *De Principiis atque Originibus (On Principles and Origins)*; repr. in *Philosophical Studies, c.1611-1619*, ed. G. Rees, Oxford: Clarendon Press.(An exposition of Bacon's natural philosophy.)

Bacon, F. (1620) *Novum Organum*; repr. ed. T. Fowler as *Bacon's Novum Organum*, Oxford: Clarendon Press, 1888; trans. P. Urbach and J. Gibson, La Salle, IL: Open Court, 1994.(Bacon's most important philosophical work, containing a detailed though still important account of his work. Fowler reproduces the original Latin text with very useful notes; Urbach and Gibson offer a good modern translation.)

Bacon, F. (1627) *New Atlantis*, London; Kila, MT: Kessinger Publishing, 1992.(Incomplete description of a scientific utopia.)

References and further reading

Abbott, E.A. (1885) *Francis Bacon: An Account of his Life and Writings*, London: Macmillan.(Though now very old, perhaps still the most detailed and balanced one-volume biography.)

Coquillette, D.R. (1992) *Francis Bacon*, Edinburgh: Edinburgh University Press.(Deals primarily with Bacon's legal thought, but also his philosophy; good bibliography.)

Farrington, B. (1964) *The Philosophy of Francis Bacon*, Liverpool: Liverpool University Press.(Bacon from a Marxist perspective; contains English translations of *Temporis Partus Masculus (The Masculine Birth of Time)*, *Redargutio Philosophiarum (The Refutation of Philosophies)* and *Cogitata et Visa de Interpretatione Naturae (Thoughts and Conclusions on the Interpretation of Nature)*, left untranslated by Ellis and Spedding.)

Herschel, Sir J. (1830) *A Preliminary Discourse on the Study of Natural Philosophy*, London: Longmans. (Discussed in §8 above.)

Jardine, L. (1974) *Francis Bacon: Discovery and the Art of Discourse*, Cambridge: Cambridge University Press. (Links Bacon with debates among his predecessors and contemporaries about the use of rhetoric and dialectic.)

Leary, J.E., Jr (1994) *Francis Bacon and the Politics of Science*, Ames, IA: Iowa State University Press.(Argues for close connection between Bacon's scientific and legal/political projects, but against interpreters who see his work as politically 'progressive'.)

Locke, J. (1823) *The Works of John Locke*, London: T. Tegg *et al.*, 11th edn, 10 vols; repr. Aalen: Scientia, 1963. (*The Conduct of the Understanding*, a short manuscript posthumously published in 1706, can be found in volume 2.)

Macaulay, T. (1837) 'Lord Bacon', *Edinburgh Review* (July); repr. in *Critical and Historical Essays*, London: Longmans, 1993.(The most accessible introduction to early-Victorian attitudes towards Bacon.)

Pérez-Ramos, A. (1988) *Francis Bacon's Idea of Science and the Maker's Knowledge Tradition*, Oxford: Clarendon Press.(A detailed and fairly technical discussion of Bacon's theory of induction. Good on Bacon's influence.)

Pentonen, M. (ed.) (1996) *The Cambridge Companion to Bacon*, Cambridge: Cambridge University Press.(A collection of up-to-date surveys, with a full bibliography.)

Quinton, A. (1980) *Francis Bacon*, Oxford: Oxford University Press.(A useful short introduction.)

Rossi, P. (1968) *Francis Bacon: from Magic to Science*, London: Routledge & Kegan Paul.(An important work, stressing Bacon's relation to the magical tradition.)

Sprat, T. (1667) *The History of the Royal Society of London*, London: J. Martin; facsimile repr. ed. J.I. Cope and

H. Whitworth Jones, London: Routledge, 1959.(Mentioned in §8 above.)

Urbach, P. (1987) *Francis Bacon's Philosophy of Science*, Chicago, IL: Open Court. (A sophisticated though sometimes idiosyncratic account.)

Webster, C. (1975) *The Great Instauration: Science, Medicine, and Reform, 1626-1660*, London: Duckworth.(An immensely detailed account of Baconianism in England.)

Bacon, Roger (*c.*1214-92/4)

Associated with both the University of Paris and Oxford University, Roger Bacon was one of the first in the Latin West to lecture and comment on Aristotle's writings on subjects other than logic. After he came to know Robert Grosseteste's work in natural philosophy, he became the advocate of a curricular reform that emphasized scientific experiment and the study of languages. His views were often unpopular, and he constantly belittled all who disagreed with him.

Bacon's work in logic and semantic theory had some influence during his lifetime and immediately after his death. His work in science, however, had little impact. His renown in the history of science is due in part to his being viewed as a precursor of the Oxford Calculators, who in turn anticipated certain important developments in seventeenth-century science.

1 Life

Bacon was born into the minor nobility in Somerset, England. By the time of his death, he had earned the honorific title *Doctor mirabilis* (the Marvellous Doctor) for his prodigious learning in every area of philosophy. He had trained in the Latin classics before attending Oxford University and the University of Paris, and he was Master of Arts at Paris by 1237. At Paris between 1237 and 1247, Bacon wrote his *Sumule dialectices (Brief Summaries of Dialectic)* and *Summa grammatica (Synopsis of Grammar)*, and lectured on Aristotelian works, including the *Metaphysics* and *Physics* as well as the pseudo-Aristotelian *De plantis* and *Liber de causis* (see Liber de causis).

During this period, Bacon became intensely interested in another pseudo-Aristotelian text, *Secretum secretorum (Secret of Secrets)*, which was thought to have been written by Aristotle for his pupil Alexander the Great. In addition to instruction on kingship, the text contained occult lore ranging from astrology and numerology to the magical medicinal properties of plants, and an account of a unified science supposedly transmitted to Aristotle by God through the Hebrews and Egyptians. According to the treatise, this unified science could be recovered only by those intellectually and morally fit for it.

Bacon was so moved by the vision of knowledge revealed in the *Secretum secretorum* that in 1247 he left his teaching position at Paris, moved back to Oxford and spent the next ten years attempting to realize that vision. His efforts were concentrated on language, mathematics and experimentation because he believed that one could decipher Aristotelian texts only if one understood the nature and nuances of the language in which the texts had been written, or into which they had been translated. He also believed that the study of mathematics (including geometry and astronomy), along with experimentation, provided the key to the sciences. His efforts proved to be physically, emotionally and financially draining.

In 1257 Bacon entered the Franciscan order, hoping to find respect for his work and funds for his scientific instruments. Finding neither, he became embittered, and at every opportunity he railed against those of his contemporaries (such as Albert the Great and Thomas Aquinas) who were taking a more traditional approach to philosophical study. By the early 1260s Bacon had become unpopular among the Franciscans, and was apparently sent to their convent in Paris under an injunction to neither lecture nor circulate his writings outside the Order without approval.

Undaunted, Bacon set about communicating his controversial views to Cardinal Guy le Gos de Foulques. The latter, upon becoming Pope Clement IV in 1265, asked to see the writings in which these views were presented in detail. Although Bacon had not yet actually composed these writings, he managed to do so without his Order's knowledge and to have his work delivered to Clement in 1267. The result, the *Opus maius*, was an encyclopedia of the sciences and a proposal for educational reform. It was followed within the year by synopses, additions and corrections in the *Opus minus* and the *Opus tertium*.

Despite Clement's death during the time these additions were being written, Bacon continued for the next ten years to write on his favourite topics. He produced a Greek grammar, a Hebrew grammar and the *Compendium studii philosophiae (Compendium of the Study of Philosophy)* around 1272. The *Communia naturalium (General Principles of Natural Philosophy)* and *Communia mathematica (General Mathematical Principles)*, two surviving sections of a second encyclopedic work begun but never completed, are considered Bacon's most mature work.

Between 1277 and 1279, Bacon's teachings were condemned by the Minister General of the Franciscan Order, Jerome of Ascoli (the future Pope Nicholas IV), and Bacon himself was imprisoned. Why this happened is unclear. Although the official account is that his teachings contained 'suspect novelties' (perhaps some of the propositions condemned in 1277 by Etienne Tempier, the Bishop of Paris), it is possible that his imprisonment came as a response to his refusal to stop verbally abusing everyone whose views he did not share. The targets of this abuse included many of his Franciscan brothers. It is also possible that his imprisonment was the result of the controversial Joachimite views he had embraced (see Joachim of Fiore). Despite these serious problems, Bacon managed before his death some thirteen to fifteen years later to compose at least one more work, *the Compendium studii theologiae (Compendium of the Study of Theology)*, in which his criticisms of contemporary education were extended to cover the moral vices of Christendom as a whole.

2 Views on education

In the mid-1260s, Bacon began to insist that the current educational system needed to be completely revised, and that he himself knew exactly what was required. He shared his contemporaries' belief that God is man's ultimate end but, unlike most of them, believed that the pursuit of science should be instrumental in the study of God (and thus in one's pursuit of moral perfection). By understanding the nature of things, one would come to understand the nature of God who made them. Mathematics also figures in this scheme insofar as Bacon considered geometry to be the key to knowing the efficient and generating causes in nature. He believed that mathematics could be used to verify scriptural claims about creation, such as its chronology and geography.

Bacon was also unusual in believing that the study of texts in natural philosophy should be pursued in conjunction with the study of the languages in which the original texts were written, such as Hebrew, Greek and Arabic. He believed that one could also improve one's understanding of Latin with such study, as many Latin words could not be fully understood without a knowledge of Greek. In general, Bacon believed that all study was subordinate to theology and morality, and that the traditionally valued subjects of study such as metaphysics should be ranked below experimental science, mathematics and languages in the hierarchy of academic study. The ecclesiastical authorities in charge of education, he urged, ought to take account of these facts and revise the curriculum accordingly.

These ideas are expressed in detail in the *Opus maius*, along with a general account of why people are in their present predicament with respect to knowledge. Because of sin and humanity's fallen state, truth is difficult for us to discover. We cannot trust reason to provide us with demonstrative proof of anything; but we can use reason to help us formulate hypotheses which can then be confirmed by sense-perception, understood either as ordinary experience or as experience aided by instruments. Reason, according to Bacon, is part of an inner experience which is provided when God illuminates our mind (his debt to Augustine is evident in this view). However, sin affects our ability to reason in four specific ways. First, we submit to untrustworthy authority; that is, we mistakenly believe sources other than God (in the form of Scripture) or the Church Fathers (who were reliable interpreters of Scripture). Second, we are unduly influenced by long-standing custom, believing that what is traditional must therefore be right. Thus, we give more weight than we should to popular opinion. Fourth, we try to conceal our ignorance with a technical show of wisdom. In this connection Bacon condemned the current practice in the universities of using Peter Lombard's *Sentences* rather than the Bible as a basis for lecturing in theology.

That Bacon was altogether serious about the importance of educational reform is evident not only in his strident disapproval of great men of his day who did not embody the ideals he deemed appropriate - men such as Alexander of Hales, Bonaventure, Albert the Great and Thomas Aquinas - but also in his dedication to the study of optics, the making of experimental instruments and a firm understanding of languages (although he seemed to rely mostly on Latin translations of the treatises he used in his research).

3 Work in science

As already noted, Bacon emphasized the practical objectives of studying both philosophy in general and natural philosophy, or science, in particular. It was his view that science had to do with natural laws that could be expressed or elucidated mathematically, and that true science therefore had nothing to do with magic, which he characterized as a false mathematics based on superstition and the presumption that everything is causally necessitated. His view of mathematics as the key to the sciences seems to have had two sources. First, he accepted

Grosseteste's view that everything in the terrestrial and celestial realms is a manifestation of light, the diffusion of which takes place in accordance with laws of geometry (see Grosseteste, R.). Second, he believed that mathematical knowledge is certain knowledge insofar as all the other Aristotelian categories depend on a knowledge of quantity, the subject matter of mathematics, and insofar as things known to us and things in nature are absolutely the same only from a mathematical point of view. Given these reasons, it is not surprising that Bacon gave optics (*perspectiva*, or the theory of vision) pride of place among the sciences. He stood apart from his contemporaries in that instead of simply giving an account of optics, he tried to show how it fit into the whole of knowledge and to synthesize what was known about optics up to that time in the work of writers such as Euclid, Ptolemy and Alhazen.

Bacon also differed from his contemporaries in insisting that science, or a complete account of nature, must include an experimental as well as a theoretical component; without experience, he says, nothing can be known sufficiently. Aristotle's claim that 'a demonstration is a syllogism that makes us know' is to be understood not of the bare demonstration but of the case in which experience of the subject matter accompanies the demonstration. According to Bacon, by means of experiments one could (a) confirm hypotheses formulated on the basis of reason, (b) gain knowledge of that which is inaccessible to proof by deductive arguments, and (c) develop products to aid in everyday life. Examples of such experiments as carried out by Bacon include research into rainbows, the properties of magnets and gunpowder. In the traditional approach to education taken by Bacon's contemporaries, however, this aspect of knowledge was usually not even acknowledged, much less advocated.

Bacon himself did not make the kind of progress in empirical science that might have been expected, given the enthusiasm with which he advocated it, and given the accounts of his scientific prowess made popular during the seventeenth century, after his work had been rediscovered. He did make some advances; for example, he provided models for plane surfaces and for convex refracting surfaces, thereby advancing understanding of the nature of vision and the usefulness of lenses in aiding vision. On the whole, however, he appears not to have been a good scientist or mathematician. He produced no mathematical proofs or theorems, he held alchemy in high regard and he believed that by reading the stars one could come to know about the future. Because he had no disciples, his work was quickly forgotten after his death.

4 Linguistic work

As noted, Bacon's plan for curricular reform included a plea for the study of languages, which he represented as the vehicle of knowledge. He himself took the study of language seriously. Not only did he write grammars for both Greek and Hebrew, but he also devoted part of Part III of the *Opus maius* (called *De signis (On Signs)*) and part of the *Compendium studii theologiae* to discussions of signification in language, and wrote an analysis of the nature of ambiguity in his *Sumule dialectices*.

Bacon's comprehensive approach to the study of natural philosophy was replicated in his study of language. His interest in language was part of a wider interest in the nature of all signs, vocal and otherwise. 'Man', for instance, is a sign of an actually existing universal instantiated in all individual men, and smoke is a sign of fire. This approach was not typical in Bacon's day (see Language, medieval theories of).

Bacon's views on how words functions as signs were also unusual, and were developed in opposition to the views of people such as William of Sherwood, Lambert of Auxerre, Siger of Brabant and Boethius of Dacia. For example, it was Bacon's view that all words are equivocal. The basis of this view was his belief that (a) words signify conventionally (*ad placitum*) as a result of meaning being imposed on them, either expressly or tacitly, by an individual or group of individuals, and that (b) any semantic theory must explain how words can signify not only present but also past and future individuals as well as never-existent things. Bacon believed that he could provide such a theory if it were presupposed that first, each act of imposition makes a word signify exactly one thing (where 'thing' is broadly construed as a word or a non-word, particular or universal, existing or imaginary), and second, each imposition lasts only as long as does the signified thing. Thus, if the word 'man' is imposed by someone to signify an individual man, it would acquire another sense in addition to its standard sense (signifying the common nature or universal humanity). In this case, the word 'man' has two senses and is thus equivocal. When the man signified by 'man' in virtue of the special imposition dies, 'man' would not signify him any longer, although one could impose a third meaning on the word 'man' to use it as a sign of the physical remains of what once was a man. There is no determinable limit to the various senses that can be imposed on a word, either

consciously or unconsciously. With such a theory of signification, Bacon thought he could solve many of the linguistic puzzles of the day. For example, could Jesus be called a man during his three-day entombment? Can man be said to be an animal when no man exists? Is the sophisma-sentence 'This is a dead man, therefore this is a man' true? His questions, his grand schemes, his idiosyncratic pursuits, even his animosities and his mistakes, show him to have been very much ahead of his time.

See also: Grosseteste, R.; Language, medieval theories of; Natural philosophy, medieval; Oxford Calculators

GEORGETTE SINKLER

List of works

Bacon, Roger (*c.*1250) *Sumule dialectices (Brief Summaries of Dialectic)*, ed. R. Steele, Opera hactenus inedita Rogeri Bacon fasc.15, Oxford: Clarendon Press, 1940; Books I-II ed. A. de Libera, 'Les *Summulae dialectices* de Roger Bacon: I-II *De termino, De Enuntiatione*', *Archives d'Histoire Doctrinale et Littéraire du Moyen Âge* 61, 1986: 139-289; Book III ed. A. de Libera, 'Les *Summulae dialectices* de Roger Bacon: III *De argumentatione*', *Archives d'Histoire Doctrinale et Littéraire du Moyen Âge* 62, 1987: 171-278.(Bacon's writings on language, including an analysis of the nature of ambiguity.)

Bacon, Roger (1267) *Opus maius*, ed. J.H. Bridges, *The 'Opus maius' of Roger Bacon*, Oxford, 1879-90, 3 vols; repr. 1964; Part III, *De signis (On Signs)*, ed. K.M. Fredborg, L. Nielsen and J. Pinborg, 'An Unedited Part of Roger Bacon's "*Opus Maius: De Signis*"', *Traditio* 34: 75-136, 1978.(Bacon's encyclopedia of the sciences and proposal for educational reform.)

Bacon, Roger (1267-8) *Opus minus*, ed. J.S. Brewer in *Fr. Rogeri Bacon Opera quaedam hactenus inedita*, London: Longman, Green, Longman & Roberts, 1859; repr. 1965.(Contains synopses, additions and corrections to the *Opus maius*.)

Bacon, Roger (1267-8) *Opus tertium*, ed. J.S. Brewer in *Fr. Rogeri Bacon Opera quaedam hactenus inedita*, London: Longman, Green, Longman & Roberts, 1859; repr. 1965.(Contains further synopses, additions and corrections to the *Opus maius*.)

Bacon, Roger (after 1268) Greek and Hebrew Grammars, ed. E. Nolan and S.A. Hirsch, *The Greek Grammar of Roger Bacon and a Fragment of His Hebrew Grammar*, Cambridge: Cambridge University Press, 1902. (Grammars, produced after finishing the *Opus maius*.)

Bacon, Roger (after 1268) *Communia naturalium (General Principles of Natural Philosophy)*, ed. R. Steele, *Opera hactenus inedita Rogeri Bacon*, Oxford: Clarendon Press, fascs 2-4, 1905?-13.(Surviving section of a second, uncompleted encyclopedic work.)

Bacon, Roger (after 1268) *Communia mathematica (General Mathematical Principles)*, ed. R. Steele, *Opera hactenus inedita Rogeri Bacon*, Oxford: Clarendon Press, fasc. 16, 1940.(Surviving section of a second, uncompleted encyclopedic work.)

Bacon, Roger (*c.*1271-2) *Compendium studii philosophiae (Compendiusm of the Study of Philosophy)*, ed. J.S. Brewer, *Fr. Rogeri Bacon Opera quaedam hactenus inedita*, London, Longman, Green, Longman & Roberts, 1859; repr. 1965.(Intended as an introduction to Bacon's own philosophical writings.)

Bacon, Roger (*c.*1292) *Compendium studii theologiae (Compendium of the Study of Theology)*, ed. H. Rashdall, *Fratris Rogeri Bacon Compendium studii theologiae*, Aberdeen: University Press, 1911; repr. 1966; ed. and trans. T.S. Mahoney, *Compendium of the Study of Theology*, Leiden: Brill, 1988.(Bacon's final work, dealing with education and the moral vices of Christendom.)

References and further reading

Crombie, A.C. (1969) *Robert Grosseteste and the Origins of Experimental Science 1100-1700*, Oxford: Clarendon Press.(Discusses the ways Bacon is dependent on Grosseteste in optics; see especially pages 139-62.)

Crowley, T. (1950) *Roger Bacon: The Problem of the Soul in His Philosophical Commentaries*, Louvain: Institut Superieur de Philosophie.(Good study of Bacon's life.)

Hackett, J.M.G., and Mahoney, T.S. (1987) 'Roger Bacon Bibliography (1957-1985)', *The New Scholasticism* 61 (2): 184-207.(This bibliography is continued in *Roger Bacon Essays*, ed. J.M.D. Hackett, Leiden: Brill, forthcoming 1997. The volume contains studies of Bacon's views on science, philosophy of language and ethics.)

Pinborg, J. (1981) 'Roger Bacon on Signs: A Newly Recovered Part of the Opus Maius', *Miscellanea Mediaevalis* 13: 403-12.(Preliminary investigation into Bacon's importance as a philosopher of language.)

Bakhtin, Mikhail Mikhailovich (1895-1975)

Bakhtin is generally regarded as the most influential twentieth-century Russian literary theorist. His writings on literature, language, ethics, authorship, carnival, time and the theory of culture have shaped thinking in criticism and the social sciences. His name is identified with the concept of dialogue, which he applied to language and numerous other aspects of culture and the psyche.

Bakhtin viewed literary genres as implicit worldviews, concrete renditions of a sense of experience. Strongly objecting to the idea that novelists simply weave narratives around received philosophical ideas, he argued that very often significant discoveries are made first by writers and are then 'transcribed', often with considerable loss, into abstract philosophy. For example, he regarded the novelists of the eighteenth century as explorers of a modern concept of historicity long before philosophers took up the topic. He argued that considerable wisdom could be achieved by probing the form, as well as the explicit content, of literary works. In literature as in life, however, much wisdom is never fully formalizable, although we may approximate some of it and gesture towards more. Such partial recuperation was, in Bakhtin's view, the principal task of literary criticism.

Bakhtin's favourite genre was the realist novel. In his view, novels contain the richest sense of language, psychology, temporality and ethics in Western thought. He revolutionized the study of novels by arguing that traditional poetics, which employed categories suitable to poetry and to drama, had been unable to appreciate just what is novelistic and especially valuable about novels. Seeking the essence of 'prosaic intelligence', he therefore formulated an alternative to poetics, which critics have called 'prosaics'. This term also designates an important part of his worldview in approaching many other topics, especially language. Bakhtin stressed the prosaic, ordinary, unsystematic, events of the world as primary. In culture, order can never be presumed, but is always a 'task', the result of work that is never completed and always upset by everyday contingent events. Better than any other form of thought, great prose, especially realistic novels, captures this prosaic sense of life.

Believing in contingency and human freedom, Bakhtin described individual people, and cultural entities generally, as 'unfinalizable'. Human beings always manifest 'surprisingness' and can never be reduced to a fully comprehensible system. Paraphrasing the implications of Dostoevskii's novels, Bakhtin located humanness in the capacity of people 'to render untrue *any externalizing and finalizing definition of them. As long as a person is alive he lives by the fact that he is not yet finalized, that he has not uttered his ultimate word' ([1929] 1984: 59; original emphasis). Ethically, the worst act is to treat people as if some 'secondhand' truth about them were exhaustible. Psychologically,*

> *A man never coincides with himself. One cannot apply to him the formula of identity $A \equiv A$... the genuine life of the personality takes place at the point of non-coincidence between a man and himself... beyond the limits of all that he is as a material being ... that can be spied upon, defined, predicted apart from its own will, 'at second hand.'*
>
> *([1929] 1984: 59)*

Bakhtin therefore opposed all deterministic philosophies and all cultural theories that understate the messiness of things and the openness of time. He rigorously opposed Marxism and semiotics, although, strangely enough, in the West his work has been appropriated by both schools. Stating his own thought as a paraphrase of Dostoevskii, he wrote:

> *nothing conclusive has yet taken place in the world, the ultimate word of the world and about the world has not yet been spoken, the world is open and free, everything is still in the future and will always be in the future.*
>
> *([1929] 1984: 166)*

1 Life

The son of a bank manager, Bakhtin studied classics and philology at the University of Petrograd (1913-18). His meticulous knowledge of obscure ancient writers is reflected throughout his work. Most of his subsequent life can be seen as a series of escapes into obscurity. During the Russian Civil War, he moved to the small towns of Nevel and Vitebsk, where he worked as a schoolteacher, discussed philosophy and acquired his two best-known

disciples, Valentin Voloshinov and Pavel Medvedev. Interested in Neo-Kantianism, Bakhtin worked on a comprehensive treatise about ethics, authorship and the relation of the self to others (see Neo-Kantianism). In the 1920s, he also encountered the most influential non-Marxist school of criticism, Russian Formalism, learned from their ideas, and rejected their fundamental approach, which he saw as scientistic in its pursuit of laws at the expense of contingency and individual agency (see Russian literary Formalism). Their reduction of content to form ran counter to his view that literature and language are repositories of wisdom acquired by human experience over ‘great time’. In 1924, he moved to Leningrad, where he failed to find stable employment, perhaps because of a bone disease eventually leading to the amputation of his right leg in 1938. In 1929, he published his book on Dostoevskii, which was recognized as a classic.

Nevertheless, his bourgeois background, his interest in religion and his non-Marxist approach made him suspect in the Soviet Union, and in 1929 he was arrested. A sentence that would have meant his death in a harsh labour camp was commuted to six years of internal exile in Kazakhstan. During the 1930s he worked at odd jobs, including bookkeeper on a collective farm. During these years he wrote his classic essays on language and the novel. Out of step with the times, these essays were not published until 1975. He also worked on a study of Rabelais and the institution of carnival, which he described in terms of systematic parody by inversion of social norms. The Rabelais thesis was submitted for a doctorate, but only a lesser degree was awarded, and the book did not appear in print until 1965. Later, it became the first of his works to be widely known in the West.

In 1936 he became a professor at the remote Mordovia State Teachers College, but soon resigned so as to remain less visible during the years of mass arrests. He returned to his professorship at the end of the Second World War. During the 1950s, intellectuals in Moscow were again inspired by Bakhtin’s book on Dostoevskii and were surprised to discover that the author (unlike Voloshinov, Medvedev and so many others) had not perished during the years of purges and privation. Celebrated, sometimes worshipped as a secular saint, he inspired reminiscences of dubious accuracy. He continued to write essays on general problems of culture from the perspective of dialogue.

For a time, the rumour circulated that Bakhtin had written some of Voloshinov’s and Medvedev’s works, and publishers reissued their books under his name. But the weight of argument and evidence indicates that Bakhtin influenced but did not write these books.

2 Language as dialogic

In his book on Dostoevskii, in ‘Slovo v romane’ (Discourse in the Novel) (1934-5), in ‘Problemy rechevykh zhanrov’ (Problems of Speech Genres) (1952-3) and elsewhere, Bakhtin offers an alternative to the linguistic theories of Saussure and movements indebted to him, including formalism, structuralism and semiotics (see Semiotics; Structuralism in linguistics). His critique, and that of Voloshinov in *Marksizm i filosofiia iazyka (Marxism and the Philosophy of Language)* (1929), would also apply to the Chomskyite tradition (see Chomsky, N.).

For Bakhtin, the fundamental unit of language is not the sentence but the utterance. Whereas sentences are repeatable, utterances occur in a specific situation. Still more important, they are formulated as part of a dialogic exchange. Utterances are not only responded to, but are shaped in anticipation of a response. In a sense, they belong to both speaker and listener. Bakhtin finds simplistic the telegraphic model of communication, in which response (or decoding) simply follows the utterance; the same reasoning would apply to what literary critics think of as reader reception theory.

Bakhtin uses his key term ‘dialogue’ in three distinct senses: as a worldview or sense of truth, as a particular use of language and as a defining quality of language itself. When taken as a defining quality of language, the term indicates that all utterances are dialogic: they are constituted by their ‘addressivity’. Dialogic relations are (from the standpoint of Saussurean linguistics) metalinguistic. Metalinguistic relations are not reducible to logical or linguistic ones. Two sentences may contradict each other, but when they are ‘embodied’ (spoken by particular people on a particular occasion), new kinds of relations are established. Thus, if two people say, one after another, ‘Life is good’, from the standpoint of linguistics we have a repetition of the same sentence, from the standpoint of logic we have the relation of identity, but from a dialogic perspective, we have agreement. But agreement is never full, and the second utterance, necessarily embodied by a different person, carries with it a different experience that leads to different implications, which we all keenly sense. Tone, an ‘emotional-volitional’ stance, is crucial to

its meaning, and tone does not appear in the sentence but in the utterance. Indeed, some utterances (hm!) are spoken entirely as carriers of tone, something like verbal gestures. A mapping of dialogic relations (Bakhtin offers an incomplete typology) would necessarily involve complex relations outside the purview of either linguistics or logic. In life, we understand these relations, and if the study of language is to comprehend how language is used, it must describe them.

This approach differs from one focusing on the abstract system of language, which is in any case not as systematic as linguists suppose. Linguists typically avoid the issues on which Bakhtin focuses by 'smuggling' into a sentence features properly belonging to an utterance when convenient.

If one views utterances dialogically, other features not present in sentences come into view. For example, everything we speak about is 'already spoken about' and our utterances orient themselves dialogically not only to potential listeners but also to earlier utterances on the topic. No one is the 'biblical Adam' uttering the first word on a topic, and complexities of meaning, which are reflected in style, arise from the diversity of orientations to previous utterances. Traditional stylistics, which views choice of words simply as the expression of the author, misses the shaping role of both listener and previous speakers. Bakhtin compares 'the word' (utterance) to a ray of light entering a tension-filled, social 'atmosphere filled with alien words, value judgments and accents through which the ray passes on its way to the object' (1934-5: 277).

Thus, some elements of reported speech are present in every utterance. Bakhtin (and Voloshinov) offer detailed analyses of the various forms of reported speech in part because all speech is, obviously or surreptitiously, reported. The obvious forms of reported speech vary from culture to culture, depending on attitudes towards discourse, authority and the relation of self to other. The concept of the 'already-spoken-about' illustrates what Bakhtin means by saying that in calling language dialogic he is referring not only to 'compositionally expressed dialogue', as in play scripts, but also to 'the internal dialogism' of each utterance.

3 Heteroglossia

Because order is always a 'task', Bakhtin argues that 'a unitary language is not something given but is always in essence posited' (1934-5: 270). Language, like the rest of culture, is constantly subject to 'centrifugal' forces disrupting it in unsystematic ways; cultures often respond with 'centripetal' efforts to restore a minimum of unity by writing dictionaries and grammars. What we call *the* English language is in part the effect of nonlinguistic social pressures.

But other ways of speaking exist, and these cannot be adequately understood by referring to dialects. Instead, Bakhtin calls attention to the fact that groups speak differently because they conceive of experience differently. Each profession, generation, locale, ethnic group or any of countless other shifting identities, all have their own characteristic vocabularies, ways of addressing others, styles and intonations because they think and understand the world in their own way. Language is always languages; it manifests 'heteroglossia'. There is no single formal principle for identifying each of these 'languages' because what distinguishes them is conceptualizations of experience, which may be expressed in countless ways, each of which is 'the sclerotic deposit' of specific social activity.

All of us in fact know more than one language of heteroglossia because we function in diverse environments. We often approach one field of experience in one language and with the attitudes of another, and so languages of heteroglossia may enter into dialogue. This process not only changes our sense of experience but also, when repeated by many people, affects the development of language. Bakhtin preferred such messy and unsystematic accounts of language change to notions like those of the Formalists, who saw language as an abstract system developing by principles that were themselves abstract and systematic, so that, as Roman Jakobson and Iurii Tynianov observed, the history of a system is itself a system.

4 Double-voicing

All utterances are dialogic in one sense, but not in another. For utterances have different 'tasks'. Although all utterances can be shown to be cited or reported from other contexts, sometimes we do not want such allusiveness to be perceived; we just want to get our point across. At other times, for example when we engage in parody,

perception of ‘quotation marks’ is necessary to our purposes. Thus, some utterances are monologic, others dialogic; some ‘single-voiced’ and others ‘double-voiced’. In double-voiced utterances we have two speech acts and are meant to perceive both sets of intentions.

Double-voiced utterances may in turn be divided into passive and active. The distinction lies in the relation of the two utterances to each other. The utterance cited or alluded to may be subservient (passive) to the first. In parody, for example, the target utterance clearly enjoys lower semantic authority than the parodying utterance. The two disagree, and we know with whom we are expected to side. In another form of passive double-voiced word, stylization, we distinctly hear two utterances - say, one from another era and style and one currently spoken - that remain in qualified agreement. (One must not forget that agreement is also a dialogic relation.)

Sometimes the second utterance actively resists appropriation by the first. ‘In such discourse, the author’s thought no longer oppressively dominates the other’s thought, discourse loses its composure and confidence, becomes agitated, internally undecided and two-faced’ ([1929] 1984: 198). Sometimes, speech seems to take a ‘sideward glance’ at a possible hostile answer and responds to it in advance; the word has a ‘double orientation’, towards its topic and towards an expected response so active that it may utterly dominate the utterance. This sort of speech is rather common in everyday life, where it appears in various barbed words, overblown discourse repudiating itself in advance or the speech of deeply agitated people. It is especially well represented in literature. Dostoevskii’s *Zapiski iz podpol’ia (Notes From Underground)* contains actively double-voiced ‘words with a loophole’, in which speakers retain the right to reverse their position if they get the wrong response. One may sense a loophole in a tone of potential (but not definitive) self-mockery or in agitated exaggeration.

5 Psychology and ethics

Throughout his life, Bakhtin battled against what he initially called ‘theoretism’ and (after studying language) named ‘monologism’: the idea that the world is fundamentally orderly and that experience is in principle fully capturable in abstract theories. What theories that reason down to experience, instead of up from it, usually produce is bloodless ‘transcriptions’ of events. For example, the heritage of rationalism views events as utterly explicable in terms of underlying laws. Any such approach necessarily reads out of events their ‘eventness’, that which makes them and the present moment ‘momentous’ in the face of multiple future possibilities. When applied to language, theoretism leads to the Saussurean division between *langue*, the abstract system of language, and *parole*, the particular speech act, viewed as a mere instantiation of rules. Applied to people, it makes them entirely ‘coincident’ with themselves, a mere bundle or system of qualities without ‘unfinalizability’.

When ethics proceeds in this way, it grossly oversimplifies situations by fitting them into pregiven categories, whereas the most ethically relevant facts may turn out to be entirely unpredictable in advance. In stressing the importance of particulars, Bakhtin revives casuistry’s reasoning by cases (see Casuistry; Moral particularism). Also, the essence of ethical choice, its ‘oughtness’, depends on the ‘eventness’ of the event, which is lost in top-down reasoning. Morality involves taking into account a particular person, not a generalized agent, in a specific, unrepeatable situation: ‘What can be accomplished by me cannot be accomplished by anyone else, ever’ ([*c*. early 1920s] 1986: 129). For Bakhtin, the highest value is a human being, and the subordinated value is ‘the good’, and not the other way around. In the Soviet context, Bakhtin understandably opposed the sort of ideological thinking that displaces responsibility to an abstract authority or set of rules, which we are obliged to follow mechanically (the state of mind he calls ‘pretendership’). Rules and many other psychological mechanisms offer us an illegitimate ‘alibi’ whereas the essence of ethical life is our perpetual state of ‘non-alibi’. (This position is loosely comparable to existentialist analyses of authenticity and ‘bad faith’: see Existentialist ethics §3.)

Bakhtin’s early writings situate ethics within a larger discussion of the phenomenology of self. We see other people as they cannot see themselves. You cannot see the back of your head or the expression of your face when it is unselfconscious; your tears are framed for me by the blue sky behind and your own act of looking cannot appear in your field of vision. That is, I enjoy a ‘surplus of vision’ with respect to you; people exist in a state of ‘outsideness’ with respect to each other. For me you are part of the world, whereas your ‘I-for-myself’ experiences itself as entering into the world. Thus, we never see ourselves in our dreams as a finished presence, as we see others; but when we narrate our dreams later, we make ourselves into a character. In that case, the I-for-myself enters into the world by narrating.

Our selves consist in part of the ‘finalized’, almost artistically rendered images of ourselves that we receive from others. Bakhtin calls this sense of self received from others the ‘soul’, which is a ‘gift’ from outside. For the I-for-myself, the soul is something that we are always transcending, rendering partly untrue; the self is always ‘yet-to-be’. Ethically, outsideness and the surplus indicate that empathy, in the sense of merging, is the wrong way to respond to another’s suffering. For if merging were successful, no help could be given. Instead, what is needed is ‘living into’ another in which we put ourselves in the other’s position while also retaining our ‘surplus’, an outside position from which something new and helpful can be offered.

Bakhtin’s theories of language led him to the idea that, once a person acquires language, consciousness and selfhood are primarily linguistic in nature, so long as one understands language dialogically. We learn language by internalizing dialogic exchanges, and when we think we enact dialogues (often highly abbreviated) in our heads. Significant people in our lives become the repeated addressees of our inner speech; in effect, we are the voices that inhabit us. Thus, the full range of dialogic phenomena that appear in Bakhtin’s theories of language can be reinterpreted psychologically; Bakhtin also adds some new types of internal dialogic relations. In opposition to Freud, who viewed the self as something that must be socialized, Bakhtin saw the self as essentially social because it is dialogic.

6 Authorship and polyphony

Bakhtin’s early works contend that characters in a literary work cannot be truly free. After all, the author knows in advance their ultimate destiny and the significance of everything they do. Because they are part of a larger whole, they are themselves given ‘whole’, with no room for genuine surprisingness. Readers are always aware that the author is in full control, has planned everything in advance, and so no event is truly ‘eventful’. It would appear that narrative art excludes the possibility of representing anything but a pale image of human freedom.

In Bakhtin’s view, Dostoevskii discovered how to overcome the tendency of art to impose a destiny on characters. What they say and do is surprising to the author and truly eventful. Critics have plausibly discerned in the idea of polyphony a theological agenda, which could not be expressed directly in the Soviet Union. If we allow that the relation of God to people is analogous to that of an author to heroes, then if a polyphonic author could create characters capable of genuine surprise, God may have created people polyphonically as well. He may have deliberately created the world as a genuinely open process, where he cannot foresee what happens because human freedom genuinely exists.

To create the polyphonic novel, Dostoevskii surrendered an author’s ‘essential surplus’ with respect to his characters. This ‘essential surplus’ (which is stronger than the ordinary surplus each person enjoys with respect to others) consists in having a plan of the whole. Our knowledge that the author enjoys this surplus means that in a literary work the author must be the ‘ultimate semantic authority’: the meaning of any character’s utterances cannot be taken directly but must be mediated through the work as a whole. But in Dostoevskii’s novels, each major character enjoys semantic authority equivalent to the author’s. The author (not just the narrator) becomes simply another character. The result is that characters may surprise the author and the work becomes an open-ended, unfinalizable dialogue among distinct points of view.

To create such a work, Dostoevskii does not predetermine a plot, but, through a complicated process, concretely imagines fully realized voices, each expressing a whole sense of existence. He then creates a situation that will provoke the characters into dialogic confrontation; he records, in turn, what each one says, without knowing in advance where the conversation is leading. If a given character or narrator expresses Dostoevskii’s own opinions, that voice is just another in the dialogue, and, indeed, may lose the argument or wind up saying things the author has not imagined he could say. The results of each exchange are then allowed to lead to another, and the overall novel is simply the succession of thrilling dialogues. Plot is whatever happens to occur and ceases to be essential; it is no longer the ‘clamps’ holding the work together. The essence of the work lies in its ‘great dialogues’.

In addition to freedom, surprisingness and eventness, polyphony’s very form expresses another key idea: ‘monologism’, ‘theoretism’ and ‘rationalism’ are mistaken in recognizing only one kind of truth. Monologism’s truth is composed of abstract propositions that ‘gravitate’ to a system (an ‘ideology’), in which the particular voice speaking each proposition does not affect its truth value. But truth may also be dialogic, that is, may require many voices:

> It is quite possible to imagine and postulate a unified truth that requires a plurality of consciousnesses, one that in principle cannot be fitted within the bounds of a single consciousness, one that is, so to speak, by its very nature *full of event potential* [eventness] and is born of a contact among various consciousnesses.
>
> ([1929] 1984: 81)

Any attempt to 'transcribe' or 'monologize' this truth into a proposition will inevitably reduce it to a shadow of itself.

7 Theories of the novel

As we have seen, genres, like languages of heteroglossia, each express a view of experience. Because, according to Bakhtin, novels offer the richest view of life, his elucidation of their hidden wisdom in effect states his own views.

Bakhtin offers three distinct accounts of this genre's 'form-shaping idea'. The first follows from his theory of language. Keenly aware of the vast diversity of social beliefs implicitly carried by each 'language of heteroglossia', novels orchestrate an intense dialogue of those languages. Social values come into unexpected interaction as we see the 'image' of one language from the perspective of another. Novels may exaggerate actually existing dialogues of languages or create new dialogues among languages that have still not encountered each other in real life. The result is to deprive a given language and way of thinking of its naïveté: it can no longer present itself as the unchallenged best way to talk about its central topics. It is now 'contested, contestable, and contesting'. No longer able to regard itself as the unchallenged centre of things, a language ceases to move in a 'Ptolemaic' universe. The novel expresses a 'Galilean language consciousness'.

Poetics and stylistics, which rely on an examination of tropes, miss this essential activity of novels. They look for a style, whereas the novel manifests a peculiar style of styles, an 'interillumination of languages', each approaching others from various angles. Languages hybridize. Bakhtin offers a catalogue of how languages 'reaccentuate' each other. Because novels intensify hybridization always around us in daily life, they become a subtle sociological form. And because our inner speech also consists of such dialogues, novels are revealing about our mental processes in a way otherwise unavailable.

It is important to recognize that when Bakhtin speaks of such dialogues, he is not talking about the exchanges between characters. He has in mind the complex play of languages in the voice of narrators. The novel is distinguished from other genres precisely by its intense orchestration of voices and languages in this way. For example, a given character's particular way of speaking, which reflects a complex of social attitudes, infiltrates the narrator's speech and so may be felt even when the character is not present; the narrator's language is affected by this 'character zone'. Or the 'common language' of some group may be ironically reaccentuated in the narrator's speech, which weaves in and out of various languages. Bakhtin offers a number of examples as he develops a theoretical framework.

Bakhtin's second theory of the novel focuses on its temporality. He offers a catalogue and a history of literary genres, from antiquity to the nineteenth century, each of which embodies implicit assumptions about causality, agency, biography, the social environment's effect on action, a sense of historicity, fate or freedom and countless other topics concerned with our status as actors in a spatiotemporal, socio-historical world. These assumptions are rarely made explicit, but they may be detected in the characteristic ways a plot is constructed in each genre. For example, action that is plausible in one genre (last-minute escapes in an adventure story) may be completely implausible in another. Bakhtin calls a specific way of conceptualizing human action in time and space a 'chronotope'.

For Bakhtin, Goethe was the key figure in establishing the chronotope of the modern novel, and so he joins Rabelais and Dostoevskii as Bakhtin's favourite authors. The novel displays the most sophisticated chronotope for several reasons. It displays a sense of open time: 'Reality as we have it in the novel is only one of many possible realities… it bears within itself other possibilities'. Novels are also 'heterochronous': they display a social world in which many different chronotopes operate and interact. Novels understand how people change. Whereas in picaresque, the order of incidents could be changed because the hero or heroine remains essentially the same, novels show a gradual process of 'becoming'. Moreover, character change is not conceived as a mere 'unfolding' or revealing of inherent qualities, but as a complex process of interaction among social forces,

individual initiative and contingent events. Whereas in romances and other genres, the social world is mere background that does not significantly affect the characters, in novels the social world itself partially shapes their development. That is one reason why it has proven easy to transpose romances across centuries or national boundaries, whereas novelistic action is typically located precisely in time and place and would be almost impossible to relocate elsewhere. The novel recognizes both social constraints and individual choice and subtly shows their interaction. Moreover, social forces are themselves seen as in flux - a process of change that is often the key theme of novels. Above all, novels show their main characters as possessing what Bakhtin calls a ‘surplus of humanness’. In contrast to Soviet Marxism and other ideologies describing people as, in principle, fully knowable, the novel understands a person as transcending any conceivable ‘secondhand’ definitions: ‘There is no mere form that would be able to incarnate once and forever all of his human possibilities and needs, no form in which he could exhaust himself down to the last word’. For the novelistic hero ‘there always remains a need for the future, and a place for this future must be found’.

Deeply suspicious of Soviet and all other ‘theoretist’ dogmatisms, Bakhtin appropriated the novel as a counter-weapon. This appropriation is most evident in his third theory of the novel as a ‘carnivalized’ genre, that is, as a form embodying the spirit of various social rituals (especially the ancient Saturnalia and medieval carnival) that subject all received social norms to parody. Whereas theorists conventionally traced the novel to ancient epic, Bakhtin stressed its debt to menippean satire, a literary form beginning in antiquity that mixed philosophy and various kinds of parody in a highly idiosyncratic form. In the Renaissance, Rabelais combined menippean satire with the spirit of carnival to produce a distinctive amalgam leading, by way of Laurence Sterne and others, to the realist novel with its deep suspicion of all dogmatisms. Bakhtin recognized that it is possible to write long prose narratives without this menippean spirit, without the novelistic chronotope and without ‘dialogized heteroglossia’, but such works (which would include socialist realist fiction) are merely ‘unnovelistic novels’. Where ‘prosaic intelligence’, ‘unfinalizability’ and the spirit of open-ended dialogue join, as they do in the novel, we reach the most profound human vision - yet.

GARY SAUL MORSON

List of works

Bakhtin, H.M. (written 1919) ‘Iskusstvo i otvetstvennost’’, trans. V. Liapunov, ‘Art and Answerability’, in *Art and Answerability: Early Philosophical Essays by M.M. Bakhtin*, Austin, TX: University of Texas Press, 1990, 1-3.(Bakhtin’s first published work, a short essay on ethics and literature.)

Bakhtin, H.M. (written c. early 1920s) *K filosofii iskusstva (Towards a Philosophy of the Act)*, in *Filosofiia i sotsiologiia nauki i tekhniki 1984-5*, Moscow: Nauka, 1986, 80-160; Austin, TX: University of Texas Press, 1993.(This early work is Bakhtin’s key statement on ethics.)

Bakhtin, H.M. (written c.1920-3) ‘Author and Hero in Aesthetic Activity’, in M. Holoquist and V. Liapunov (eds) *Art and Answerability*, Austin, TX: University of Texas Press, 1990. (A meditation on the relation of selves to others, in life and literature.)

Bakhtin, H.M. (1924) ‘The Problem of Content, Material, and Form in Verbal Art’ in M. Holoquist and V. Liapunov (eds), *Art and Answerability*, Austin, TX: University of Texas Press, 1990.(Bakhtin’s critique of Russian Formalism.)

Bakhtin, H.M. (1929) *Problemy poetiki Dostoevskogo*; expanded 2nd edn 1963; ed. and trans. C. Emerson, *Problems of Dostoevsky’s Poetics*, Minneapolis, MN: University of Minnesota Press, 1984.(This is a translation of the expanded 1963 edition. Perhaps his single most important work, it discusses genres, ethics, the novel and the theory of language with Dostoevskii as constant illustration.)

Bakhtin, H.M. (written 1934-5) ‘Slovo v romane’, trans. C. Emerson and M. Holquist, ‘Discourse in the Novel’, in *The Dialogic Imagination*, Austin, TX: University of Texas Press, 1981, 259-422.(His theory of language and defence of a specifically prosaic approach to prose.)

Bakhtin, H.M. (written 1936-8) *Roman vospitaniia i ego znachenie v istorii realizma*, trans. V. McGee, ‘The Bildungsroman and its Significance in the History of Realism’, in *Speech Genres and Other Late Essays*, Austin, TX: University of Texas Press, 1986, 10-59.(These are surviving fragments from Bakhtin’s lost book on Goethe and the novel of education.)

Bakhtin, H.M. (written 1937-8) ‘Formy vremeni i khronotopa v romane’, trans. C. Emerson and M. Holoquist, ‘Forms of Time and of the Chronotope in the Novel’, in *The Dialogic Imagination*, Austin, TX: University of Texas Press, 1981, 84-258.(The key statement of his philosophy of time. Concluding remarks written 1973.)

Bakhtin, H.M. (written 1940) ‘Iz predystorii romannogo slova’, trans. C. Emerson and M. Holoquist, ‘From the

Prehistory of Novelistic Discourse', in *The Dialogic Imagination*, Austin, TX: University of Texas Press, 1981, 41-83.(Like 'Discourse in the Novel' this discusses the theory of novelistic prose.)

Bakhtin, H.M. (written 1941a) 'Epic and Novel', in *The Dialogic Imagination*, Austin, TX: University of Texas Press, 1981, 3-40.(A synthesis of Bakhtin's various theories on the novel.)

Bakhtin, H.M. (written 1941b) *Tvorchestvo Fransua Rable i narodnaia kul'tura*, trans. H. Iswolsky, *Rabelais and His World*, Cambridge, MA: MIT Press, 1968.(This thesis, unsuccessfully submitted for a doctorate, was not published in Russia until 1965.)

Bakhtin, H.M. (written 1952-3) 'Problemy rechevykh zhanrov', trans. V.W. McGee, 'The Problem of Speech Genres', in *Speech Genres and Other Late Essays*, Austin, TX: University of Texas Press, 1986, 60-102.(The importance of genres of daily speech in understanding dialogue.)

Bakhtin, H.M. (1970) 'Otvet na vopros redaktsii Novogo mira', trans. V.W. McGee, 'Reply to a Question from the Novy mir Editorial Staff', in *Speech Genres and Other Late Essays*, Austin, TX: University of Texas Press, 1986, 1-9.(A key statement on the theory of interpretation of texts from remote times and cultures.)

Bakhtin, H.M. (1970-1) 'Iz zapisei 1970-1', trans. V.W. McGee, 'From Notes Made in 1970-71', in *Speech Genres and Other Late Essays*, Austin, TX: University of Texas Press, 1986, 132-58.(A series of jottings on various topics.)

Bakhtin, H.M. (1975a) 'K metodologii gumanitarnykh nauk', trans. V.W. McGee, 'Toward a Methodology for the Human Sciences', in *Speech Genres and Other Late Essays*, Austin, TX: University of Texas Press, 1986, 159-72.(Stresses a dialogic approach.)

Bakhtin, H.M. (1975b) *Voprosy literatury i ėstetiki: Issledovaniia raznykh let (Questions of Literature and Aesthetics: Research of Various Years)*, Moscow: Khudozhestvennaia Literatura.(Inlcudes the first publication of some important articles.)

Bakhtin, H.M. (1979) *Ėstetika slovesnogo tvorchestva (The Aesthetics of Verbal Art)*, ed. S. Bocharov, Moscow: Iskusstvo.(Collection containing several key texts.)

Bakhtin, H.M. (1986) *Literaturno-kriticheskie stat'i (Literary-critical Articles)*, ed. S. Bocharov and V. Kozhinov, Moscow: Khudozhestvennaia literatura.(A comprehensive anthology.)

References and further reading

Bernstein, M.A. (1992) *Bitter Carnival: 'Ressentiment' and the Abject Hero*, Princeton, NJ: Princeton University Press.(Contains a subtle application and moral critique of Bakhtin's ideas on narrative, psychology and carnival.)

Clark, K. and Holquist, M. (1984) *Mikhail Bakhtin*, Cambridge, MA: Harvard University Press.(The first biography of Bakhtin.)

Mandelker, A. (ed.) (1995) *Bakhtin in Contexts: Across the Disciplines*, Evanston, IL: Northwestern University Press.(An anthology demonstrating how Bakhtin's ideas have been used in the humanities and social sciences.)

Morson, G.S. and Emerson, C. (1990) *Rethinking Bakhtin: Extensions and Challenges*, Evanston, IL: Northwestern University Press.(An anthology containing a range of views, including critiques; the introduction offers a detailed summary of *Towards a Philosophy of the Act*.)

Morson, G.S. and Emerson, C. (1990) *Mikhail Bakhtin: Creation of a Prosaics*, Stanford, CA: Stanford University Press.(Offers a detailed account of all Bakhtin's significant works.)

Bakunin, Mikhail Aleksandrovich (1814-76)

Bakunin was the leading proponent in the second half of the nineteenth century of a variety of anarchism rooted in a Romantic cult of primitive spontaneity, and one of the principal ideologists of Russian populism. But along with his public defence of the principle of 'absolute liberty' he attempted to set up networks of secret societies which were to direct the revolution and subsequently assume dictatorial powers. The contradiction between these two aspects of his activities has puzzled historians, many of whom have sought the answer in his personality, in which the urge to dominate was as strong as the urge to rebel.

1 Life

Born into a gentry family in the Russian province of Tver, Mikhail Aleksandrovich Bakunin became a leading figure in the Moscow intellectual circles of the 1830s as an exponent first of the idealism of Fichte, and subsequently of a conservative interpretation of Hegel. In Germany, where he went to study philosophy in 1840, he was converted to radicalism in philosophy and politics by the Young Hegelians (see Hegelianism; Hegelianism, Russian). In France he met Proudhon, who would have an important influence on his later thought. During the upheavals of 1848-9 he preached a revolutionary Pan-Slavism, calling for a democratic federation of all Slavic nations (see Pan-Slavism §3). Arrested in 1849 for revolutionary activities in Dresden, he was handed over first to the Austrian and then to the Russian government, which condemned him to prison followed by exile in Siberia. In 1861 he escaped to London, to collaborate with Aleksandr Herzen's émigré press (see Herzen, A.I.) on revolutionary propaganda directed at the Russian Empire. After the suppression of the Polish uprising of 1863 he turned his revolutionary hopes to Western Europe and began to expound an anarchist philosophy (see Anarchism). He founded his own international organization, the Alliance of Socialist Democracy, which he sought to have affiliated to the First International. He disbanded it as a tactical measure when his request was rejected, and joined the International with his supporters in order to challenge its domination by Marx's faction, which the Bakuninists accused of dictatorial centralism. Expelled from the International in 1872, he devoted his last years to revolutionary projects (including a rising in Lyon during the Franco-Prussian War) which all ended in failure.

2 Thought

Like many of his generation in Russia, Bakunin found in German Idealism a source of compensating fantasies which helped him sublimate his sense of alienation and achieve a foretaste of the wholeness which he craved. He was particularly attracted to Fichte's vision of absolute liberty as an earthly ideal to be attained through a protean feat of will (see Fichte, J.G. §5). Fantasies of self-affirmation alternated in the young Bakunin's thought with dreams of self-surrender through communion with the Absolute in religious or aesthetic contemplation, or through idealized platonic relationships with others of his circle (the cult of the 'beautiful soul').

Bakunin believed that his arrival in Europe had transformed him from a dreamer into a man of action; but his often-expressed contempt for 'theorizers' was based on the Romantic theory of the regenerative power of primitive spontaneity. He interpreted the unrest among the peoples of the Russian and Austro-Hungarian empires in the 1840s as the expression of elemental forces which were destined to sweep away all the artificial systems and institutions that sought to suppress them. This spirit of instinctive revolt, 'the sole creative force in history', was to be found at its purest in the most primitive, and thereby the least corrupt, of the common people: the peasantry. He placed high hopes on the Russian peasants who had a tradition of revolt, and whom he represents in his *Vozzvanie k Slavianam (Appeal to the Slavs)* of 1846 as a 'fiery ocean' which will engulf Moscow in blood and flame, and bury all the slavery in Europe beneath its own ruins. This faith was based on his version of the revolutionary dialectic of the Left Hegelians. In 'Die Reaktion in Deutschland' (Reaction in Germany) (printed in Arnold Ruge's *Deutsche Jahrbücher* in 1842 under the pseudonym 'Jules Elysard') he proclaims that the total destruction of the 'positive' (the existing order) by the forces of 'negation' will lead to 'a new heaven and a new earth… in which all the discords of our time will be resolved in harmonious unity'. This essay, which gave him his entrée into radical circles in Europe, ended with the famous line: *'Die Lust der Zerstörung ist auch eine schaffende Lust!'* ('The urge for destruction is also a creative urge!')

The anarchism which became his creed in the 1860s was a logical consequence of his cult of spontaneity. He now declared all states to be oppressive by their nature as institutionalizations of the rule of systems over life. The

infallible instincts of the masses were the only source of freedom and virtue. These instincts had created the peasant commune which still survived in Russia: a form of self-government which could become universal once the masses were freed from the tutelage of political and intellectual elites who could teach them nothing and who sought to prolong the tyranny of theory over life: as a prime offender in this respect he pointed to the scientific socialism of Marx, although Marx’s economic materialism was an important influence on his thought, along with Comtian positivism and Feuerbach’s critique of religion (see Comte, A.; Feuerbach, L.A.).

He considered himself a materialist, atheist and positivist, but the theoretical basis of his anarchism was flimsy and shot through with contradictions. In particular, his two principal theoretical works, *L’Empire Knouto-Germanique et la révolution sociale (The Knouto-Germanic Empire and the Social Revolution)* (1871) and *Gosudarstvennost’ i anarkhiia (Statism and Anarchy)* (1873) belie his self-image as the champion of ‘life’ against scientific and metaphysical abstraction. In these works he argues that all contemporary conflicts can be traced to the opposition between ’two polarities’: the state (equated with reaction), and revolution (identified with his brand of anarchism). He defines the primary characteristic of the masses as their spirit of rebellion: this forces him to exclude from the category of the ‘true’ people all those not inclined to rebel. Those who were so inclined are much idealized in Bakunin’s anarchist writings, along with the revolutionary secret societies for which he tirelessly recruited, but which existed more in his imagination than in fact.

His conspiratorial writings expound the notion of an ’invisible dictatorship’ which, by ensuring the unhindered expression of the people’s will, would bring about ’absolute liberty’, a state of being that Bakunin interprets collectivistically. He reverts to the mystical exaltation of his early idealism in passages in which he invites the revolutionary intellectual to ’drown’ in the purifying stream of popular revolt, thereby exchanging the ’appearance’ of power for its reality. His plans for dictatorship have been seen as attempts to act out idealist fantasies of spiritual self-fulfilment; they had disastrous consequences in his association with the notorious Jacobin Sergei Nechaev, whose *Catechism of a Revolutionary* justified all methods that furthered the cause of revolution, and who illustrated this theory in a murder that inspired Dostoevskii’s *Besy (The Possessed)*. Bakunin idealized Nechaev’s violent personality and methods as the embodiment of a purifying force of destruction, and in 1869-70 collaborated with him on the production of pamphlets designed to foment a peasant rising in Russia. Bakunin’s attempt to justify Nechaev’s crime to his fellow revolutionaries was a shameful episode in his career, exemplifying his reluctance to confront the contradiction between his goal of anarchist liberty and the authoritarian means to which, in his millenarian impatience, he was prepared to resort.

AILEEN KELLY

List of works

Bakunin, M.A. (1961-) *Archives Bakounine*, ed. A. Lehning, Leiden: Brill.(The first complete edition, still in progress, of Bakunin’s works, including previously unpublished manuscripts.)

Bakunin, M.A. (1873) *Gosudarstvennost’ i anarkhiia*, trans. and ed. M.S. Shatz, *Statism and Anarchy*, Cambridge: Cambridge University Press, 1990.(One of his most important theoretical works.)

Bakunin, M.A. (1973) *Selected Writings*, ed. and introduced by A. Lehning, trans. S. Cox and O. Stevens, New York: Grove Press. (Selections from letters and works from 1836-73, including ‘Reaction in Germany’.)

Bakunin, M.A. (1985) *From Out of the Dustbin: Bakunin’s Basic Writings 1869-1871*, trans. and ed. R.C. Cutler, Ann Arbor, MI: Ardis.(Propaganda writings as a member of the International.)

References and further reading

Carr, E.H. (1975) *Michael Bakunin*, London: Macmillan, 2nd edn.(Still the best biography of Bakunin.)

Kelly, A. (1982) *Mikhail Bakunin. A Study in the Psychology and Politics of Utopianism*, Oxford: Clarendon Press.(Focuses on the influence of idealist philosophy on his thought.)

Mendel, A.P. (1981) *Michael Bakunin. Roots of Apocalypse*, New York: Praeger Publishers.(A psychoanalytic approach to the contradictions between Bakunin’s authoritarianism and defence of freedom.)

Bar Hayya, Abraham (*c*.1016-*c*.1136)

Abraham bar Hayya (also called bar Hiyya) sought to reconcile Jewish tradition with contemporary philosophical thought, in his case that received from Arabic sources. Generally considered to be a Neoplatonist whose philosophy is enriched with Aristotelian accretions, he has also been called the first Jewish Aristotelian. He pioneered the writing of philosophy in Hebrew, and his work influenced later Jewish philosophers and the Kabbalah.

1 Life and works

Little is known of bar Hayya's life. Even the familiar form of his name, bar Hiyya, appears to be mistaken; there is clear evidence, including a rhyme in which he uses his name, that his patronymic was bar Hayya. He is known to have lived in Barcelona, but there is evidence that he visited France, probably Provence. He held high office, probably judicial, in the Jewish community and apparently also in the general community.

A polymath, bar Hayya wrote on astronomy and astrology, mathematics, geography, optics and music as well as philosophy. Some of his works, translated into Latin, became important sources in medieval European science. He himself was a translator from Arabic to Latin, collaborating on a number of scientific translations with the Christian scholar Plato of Tivoli, who transmitted the Ptolemaic system to the Latin world.

Bar Hayya's philosophy is found in two works, *Hegyon ha-Nefesh ha-Azuvah (The Meditation of the Sad Soul)*, an ethical work based largely on homiletical expositions of biblical texts, and *Megillat ha-Megalleh (The Scroll of the Revealer)*, written to fix an eschatological timetable according to which the Messiah will appear after 1136 and the dead will be resurrected in 1448, but also incorporating philosophical sections.

2 Metaphysics

Unlike many other medieval Jewish philosophers, bar Hayya omits proofs for God's existence and attributes, which he regards as self-evident. The structure of the universe clearly shows the unity of God and reveals him as its creator. Creation *ex nihilo* is assumed. The forms of all things and the archetypes of all species were first created in potentiality during the six days of creation, after which God brought them into actuality to endure throughout time. In potentiality things consist of form, matter and non-being. To bring them to actuality, God removed non-being and joined form to matter. Form and matter, the basic constituents of all that is, exist separately in God's pure thought, until conjoined by the divine will. Both are subdivided: matter into pure matter and sediment; form into closed and open forms. Creation results from the emanation of a light from the closed form (which itself is too pure to combine with matter and is identified with the light created on the first day in the biblical account) (see Creation and conservation, religious doctrine of). The upper world above the firmament (that is, the heavens) is divided into five worlds of light, the highest of which corresponds to the divine throne. The light emanating from within the closed form shines on the open form, enabling it to combine with matter. From this results the creation of the firmament and the four elements of which the world is composed. The human soul is a form temporarily attached to matter but ultimately destined to return to pure form. Expressed in traditional Jewish terms, the light can exist in this world and the world to come; the firmament and luminaries, in this world but not the next; and the lower creatures, in neither. Human beings are distinguished from other animals by their rational faculty, and those who act appropriately have their real existence in the next world, not this one.

3 Humility and eschatology

Humanity, the summit of creation, has three souls: vegetative, animal and rational (see Soul, nature and immortality of the). The goal of human life is to subdue the lower two souls to the rule of the rational. If the rational soul rules, one is righteous; if the others dominate, one is wicked. At death, one retains only the rational soul, and its state determines one's subsequent fate. The highest reward goes to the saint, aloof from affairs of the world, whose whole life is devoted to the hereafter, in fear and worship of God. Such an individual separates himself from this world and is not moved by its desires or lusts. The ascetic ideal bar Hayya expresses here is rare in Jewish contexts, but it is found in his contemporary Bahya Ibn Pakuda. The guiding virtue for bar Hayya is humility, and he despises worldly possessions; but only one in a thousand is perfect and merits the full rewards of the world to come, the world of pure form, in which the saint is completely emancipated from matter. Those who

have sinned have the freedom, while still alive, to repent, and the reward of sincere repentance is eternal bliss. The suffering of the righteous serves to test them in this world, so as to enhance their reward in the world to come. Evil, like good, emanates from God. It comes to the world not only to test the righteous but also to requite the wicked (see Evil, problem of). Free will ends with death; there is no subsequent repentance, and the actions of descendants cannot affect the fate of the dead (see Free will).

Things acquired in this world have no permanence. The righteous live in this world only to establish merits that will sustain them in the next. We gain life in the world to come through belief in God and his Torah and by acting only for the sake of God. The wicked have no future after death; their disembodied souls mingle, leaving none distinguishable from its fellow. Obliteration is the fate of all non-Jews; for, even if they repent, it is not for God's sake. The world to come is unalloyed bliss, 'life without death, being without non-being'.

Just as man is superior to other animals, so Israel is superior to the other nations. However, this superiority is not exclusive; it is open to all who are willing to accept the Torah. In the messianic era, all evil will end and enmity will be abolished. All the wicked and most of the nations on earth will perish. Only Israel and true proselytes to the Torah will survive. All the sufferings of Israel will be requited at the time of salvation by punishments exacted upon the other nations of the world. Salvation will be followed by the resurrection of the dead, but only of Israel, the souls of the righteous reunited with their bodies.

4 Significance

Bar Hayya is usually classed within the Neoplatonic tradition because of his emanationism and his doctrine of light (see Neoplatonism). However, his ideas of form and matter, potentiality and actuality show the impact of Aristotelian thinking. He imbibes ideas from both traditions. Stitskin (1960) classed him firmly as an Aristotelian, a pioneer of Jewish Aristotelianism and of the reconciliation of biblical with Peripatetic thought, but bar Hayya was well acquainted with the various schools of Greek thought known in his time in Arabic adaptations and abridgments, and he sometimes also reflects Arab mystical thought and the ideas of the Ikhwan al-Safa' (see Ikhwan al-Safa'). He was also familiar with Christian thought. His sources were eclectic, and his emphases often motivated by Jewish concerns. He cannot be tied exclusively to a single system.

Like other Jewish philosophers, bar Hayya sought to reconcile Jewish tradition with the rationalism of his day. All his philosophical work has a biblical or rabbinic basis, and it often uses exegetical methodology to link his philosophical reflections to scriptural tradition. Indeed two of the four chapters of the *Hegyon ha-Nefesh* take the form of homilies on the prophetic portions (Isaiah 57: 14-58: 14 and Jonah) read in the synagogue on the morning and afternoon of the Day of Atonement. His treatments of sin, repentance, and eschatology follow rabbinic tradition.

Along with his contemporary fellow-townsman Judah ben Barzillai, bar Hayya was the first Jewish philosopher to write in Hebrew, and as a result he had to coin many of the terms he used (Efros 1926-7). His predecessors had written in Arabic (in Hebrew letters) for an audience of Jews in Islamic countries. Bar Hayya's use of Hebrew in philosophic and scientific works indicates that he was writing for other Jews as well, perhaps in southern France. For them, philosophical thought would have been a novelty, and bar Hayya is at pains to stress that his ideas could be inferred from biblical sources. The influence of his writings can be traced in later Jewish philosophical thinkers, as well as in Kabbalistic thought, including the Book of Bahir and the German pietists.

See also: Bible, Hebrew; Eschatology; Neoplatonism

GEOFFREY WIGODER

List of works

Bar Hayya, Abraham (*c.*1016-*c.*1136) *Megillat ha-Megalleh (The Scroll of the Revealer)*, ed. A. Poznanski, Berlin: Mekitzei Nirdamim, 1924.(Account of bar Hayya's eschatological views, with their theoretical implications.)

Bar Hayya, Abraham (*c.*1016-*c.*1136) *Hegyon ha-Nefesh ha-Azuvah (The Meditation of the Sad Soul)*, ed. G. Wigoder, Jersualem: Mosad Bialik, 1971; trans. G. Wigoder, *The Meditation of the Sad Soul*, London: Routledge & Kegan Paul, 1969.(Very important text detailing bar Hayya's main philosophical principles.)

References and further reading

Efros, I. (1926-7) ‘Studies in pre-Tibbonian Philosophical Terminology’, *Jewish Quarterly Review* 17: 129-64, 323-68, and 20: 113-38. (Discussion of early Jewish philosophical terminology.)

Husik, I. (1942) *History of Medieval Jewish Philosophy*, Philadelphia, PA: Jewish Publications Society of America.(General history of Jewish philosophy.)

Stitskin, L. (1960) *Judaism as a Philosophy: The Philosophy of Abraham bar Hiyya*, New York: Bloch. (Full-length study of bar Hayya.)

Vajda, G. (1946) ‘Les idées théologiques et philosophiques d’Abraham bar Hiyya’ (The Theological and Philosophical Ideas of Abraham bar Hayya), *Archives d’Histoire doctrinal et littéraire du moyen âge* 15: 193-223.(Study of bar Hayya’s philosophy.)

Barth, Karl (1886-1968)

Karl Barth was the most prominent Protestant theologian of a generation shaken by the traumatic experience of the First World War and concerned with giving Christian theology a new grounding. He took a creative part in the struggle of the German Church against National Socialism, and, after the Second World War, exerted a worldwide influence that reached beyond the bounds of Protestantism. Although influenced at first by Christian socialism, Barth came to repudiate such 'hyphenated' versions of Christianity, which, he felt, underemphasize or ignore the otherness of God. There is an infinite qualitative distinction between the divine and the human; the Enlightenment attempt to historicize and secularize revelation was profoundly mistaken. This 'dialectical theology' attracted a number of leading theologians in the 1920s. Later, however, Barth felt compelled to close the gap with the divine, and developed a 'theology of the Word' to this end. Central to this approach is the concept of the knowledge conferred by faith, which makes theological understanding and rationality possible. It was on the basis of this that Barth constructed his massive Die Kirchliche Dogmatik (Church Dogmatics) (1932-70). In this, he emphasizes the self-expounding nature of Scripture (by contrast with nineteenth-century biblical scholarship, which stressed the need for a historical approach to the text) and the importance of Christ in the understanding of theology and human nature. He was a determined opponent of natural theology, and was critical of the idea that philosophy could complement theology.

1 Life

Karl Barth, a German-speaking Swiss theologian, was born in Basle, the son of Fritz Barth and Anna Sartorius. In 1889, his father, a minister and theologian of the liberal school, succeeded Adolf Schlatter at Bern University. Hence Barth was raised in Bern, where he attended the Free Gymnasium. Under his father's influence he read theology at the local divinity school. He completed his theological education in Germany, in Berlin, Tübingen and Marburg. In Berlin, he attended the lectures of the prominent liberal Church historian Adolf von Harnack. In Marburg, he fell under the influence of the Christocentric dogmatician Wilhelm Herrmann, Albrecht Ritschl's most prominent disciple. Barth's career began as assistant pastor in Geneva in 1909 and culminated as professor of theology at Basle University. Meanwhile he had played a crucial role in the German Church's struggle against National Socialism and had become, in Eberhard Jüngel's words, 'the most significant Protestant theologian since Schleiermacher'.

Barth served as a pastor in Safenwil in Aargau from 1911 to 1921. In 1913, he married Nelly Hoffmann, with whom he had five children. A liberal theologian by training, Barth was faced at Safenwil with the predicament of the working class for the first time in his life. To help the workers in his congregation he began to study law. He frequented the 'religious socialists', led by Hermann Kutter and Leonard Ragaz. In December 1911, he delivered a speech to Safenwil's Workers' Union, 'Jesus Christus und der soziale Bewegung' ('Jesus Christ and the Movement for Social Justice'), which identified Christianity with socialism. Barth claimed that 'Jesus is the social movement and that the social movement is Jesus today'. The start of the First World War broke the spell. His confidence in Culture-Protestantism was shattered. Liberal theologians had failed to prevent a resurgence of barbarism, in spite of their repeated emphasis on the religious, moral and social commitments of the Christian. The open support given by leading German theologians to the Kaiser shocked him deeply.

In September 1919, Barth delivered an epochal lecture, 'Der Christ in der Gesellschaft' ('The Christian in Society'), at a religious socialist conference in Thuringia. Won over to Christoph Blumhardt's millenarian eschatology, Barth deprecated altogether Adolf Stöcker's Christian-social movement, Friedrich Naumann's Protestant-social era, and Ragaz's religious-social programme. All these ventures, charged Barth, result in a 'hyphenated Christianity' that merges religion with socialism at the expense of revelation and the Beyond. Instead, he demanded that the otherworldliness of God be truly respected. Barth's speech marked a return to the theological paradigms of sixteenth- and seventeenth-century Protestantism. The same year, Barth published *Der Römerbrief* (*The Epistle to the Romans*), which heralded a new era in Protestant theology. A revised edition appeared in 1922 and an English translation in 1935. Emil Brunner, a Swiss colleague, praised Barth's commentary as the first victory over the platitudes of nineteenth-century Protestantism, while a dismayed Harnack saw it as part of the sickness of the age.

Barth's *The Epistle to the Romans*, which signalled the end of modernity for Protestantism, established his reputation, and he was called to the chair of Reformed theology at Göttingen in 1922. It also gave rise to the 'theology of crisis' school, the movement baptized by Bultmann as 'dialectical theology'. Eduard Thurneysen, Friedrich Gogarten, Emil Brunner (§1) and Rudolf Bultmann (§2) were among the school's most outstanding advocates. In 1923, the dialectical theologians started the journal *Between the Times*, under the editorship of George Merz, which lasted until the school's disruption in 1933. By the time of Hitler's accession to power, however, the dialectical theologians had gone separate ways. In 1925, Barth assumed the chair of systematic theology at Münster. He moved to the University of Bonn in 1930 and was dismissed in 1935 for refusing to take the oath of allegiance to Hitler. Meanwhile, he had taken part in the struggle against the German Christian movement. The 'Barmen Confession' (1934a), a rallying statement by a collection of evangelical ministers and teachers, was based on preparatory work by Barth. This confessional document spoke against the German Christians' nationalist, racist and anti-Semitic teachings, and affirmed Jesus Christ to be the one Word of God. It gave rise to the Confessing Church.

Barth returned to Switzerland to take a chair of systematic theology in Basel, a position he held until he retired at the age of 75. From there he continued his opposition to the Nazi regime, publishing polemical articles in *Between the Times* and in *Evangelical Theology*, as well as numerous pamphlets. He also pursued the writing of his major work, *Die Kirchliche Dogmatik* (*Church Dogmatics*), which he had begun in 1932. Through the years the work grew to thirteen volumes of over nine thousand pages. The *Church Dogmatics* is more than a systematic theology; it abounds with exegetical, historical, theological, practical and philosophical ideas, and is widely regarded as a classic.

During the Second World War, Barth championed the cause of the Allies. However, he was among the first to plead for reconciliation with defeated Germany. He went back to Bonn in the academic year 1946-7 and lectured at his former university. During the East-West conflict, he refused on moral grounds to view communism as on a level with National Socialism. He delivered the opening address at the first meeting of the World Council of Churches in Amsterdam in 1948, and lectured in many European countries and the USA. In 1966, he visited Rome at the invitation of Pope Paul VI and upon his return published *Ad limina apostolorum* (1967).

2 Dialectical theology

The 'theology of crisis' differs from the theological systems of neo-Protestantism (eighteenth- and nineteenth-century Protestantism) in its task and its method. What Barth proclaims in his *The Epistle to the Romans* is the otherness of God. Contrary to liberal theology, God is neither the sum of subjectivity and objectivity, nor a result of personal experience in either consciousness or history, but the Lord (*Gott der Herr*), the wholly other (*totaliter aliter*). Postulating the identity of God (*Gott ist Gott*) - a tautological axiom - Barth further predicates, under Kierkegaard's influence, an infinite qualitative distinction between time and eternity (see Kierkegaard, S.A. §§4-5). He postulates a radical separation between the divine and the human, and understands the relationship between transcendence and immanence as one of absolute apartness. In restoring the sixteenth-century Reformers' supernaturalism, he overturned the nineteenth-century account of the connection between revelation and history and repudiated the cultural objectives of later Protestantism. Barth accused the Enlightenment of having given rise to a way of thinking that psychologized, historicized, moralized and secularized revelation, the eternal and the Beyond. He thus rejected any form of relation between theology and philosophy, the former being the proclamation of God's sovereign grace. Barth put an end to the post-Kantian synthesis of neo-Protestantism, substituting a principle of separation for one of unity, causality and identity.

But if God is uncompromisingly transcendent, how can human beings speak about God? Wilhelm Herrmann's answer was that this is possible because God has revealed himself - we can speak of God because of the Incarnation. The way to God originates from God. Since for Barth human beings are finite, lacking the capacity for the infinite, all human ways are bound to end in failure. Thus faith is a crisis and grace a pure miracle. Barth thus stressed two of the Reformation's fundamental doctrines: the imputed justification by faith and the absent God. Justification comes from beyond us. The resurrection reveals the One who lives in the inaccessible light, the incognito of God.

Concerning the method of theology, Barth discarded the idea of a 'scientific' theology, much valued by the liberals, and instead proposed a theologizing process which starts from the object itself, the Beyond. He inverted

the traditional noetic process, which, since Descartes, had proceeded from the thinking subject. On the contrary, claimed Barth, the human knowledge of God originates from God's knowledge. This is a Hegelian legacy. Barth propounded the dialectical method in 1922 in 'Das Wort Gottes als Aufgabe der Theologie' (The Word of God as the Task of Theology). The divine Word is the source of theological knowledge, the foundation of Christian dogmatics and the guiding principle of every theological doctrine. What the dialectical method stresses is the paradoxical nature of the revelation. Barth later realized that the dialectical method hindered theological production and devised a new method.

3 Theology of the Word

The introduction of the Trinity in Barth's structuring of theology in 1924 marked a transition from 'dialectical theology' to the 'theology of the Word'. To bridge the divide between the divine and the human was essential to Barth. Besides Hegel, he followed Peter Lombard and Bonaventure (§4). In 1927 he published a dogmatics (*Die christliche Dogmatik in Entwurf (Christian Dogmatics in Outline)*), and in 1931 a study of Anselm (*Fides quaerens intellectum (Faith Seeking Understanding)*) in which he defended a fiducial interpretation of the ontological argument (see Anselm of Canterbury §§3-4). The Anselmian proof presupposes the fiducial knowledge of the God whose existence it claims to prove rationally. Understanding presupposes belief. The discovery of Anselm made possible what the dialectical method had rendered impossible. Owing to the centrality of the knowledge of faith in the noetic process, understanding and rationality - both indispensable in order to write a dogmatics - were possible. Fiducial knowledge is proper to theology; it becomes rational knowledge through analysis of the content of revelation, which communicates God's Word. And, again following Hegel (§4), revelation is a self-revelation. Faith and rationality are united in the understanding of faith. Fiducial knowledge is intrinsically rational because its essence, the Word of God, is rational.

From this perspective, Barth undertook the construction of the *Church Dogmatics*, a systematic theology which can best be characterized as biblical, Christocentric and ecclesiastical. The Scripture is the authoritative witness to God's revelation in Jesus Christ. It is the written Word, which points to the Word revealed and grounds the Word preached. The meaning of Scripture can only be obtained by allowing Scripture to be its own interpreter. This idea stems from the sixteenth-century Reformers. Calvin's doctrine of Scripture emphasizes two main ideas: clarity and sacred Scripture as its own interpreter. Like Calvin (§2), Barth asserts the self-expounding clarity of Scripture. Any prior understanding of God, a philosophical presupposition or anything else, is thus excluded from Barth's hermeneutics. God's self-revelation in Jesus Christ can only be known and confessed as it is attested and expounded by the direct witnesses. Against nineteenth-century biblical scholarship, Barth denied the ability of the historical-critical method to go behind the text. Hearing God's Word comprises three stages: explanation, reflection and appropriation.

Jesus Christ is the ground, content and object of Christian religion. Christian doctrines must therefore be interpreted in the light of the Incarnation. Barth acknowledged the Chalcedonian doctrine of two natures and asserted that the person of Jesus Christ is the union of true deity and true humanity (see Incarnation and Christology §1). The cross and the resurrection clear the way for human justification and sanctification. God's self-revelation, Jesus Christ, is at the same time the true human being. Jesus Christ discloses for us the essential truth about human nature; he is the source of knowledge for anthropology as well as for theology. Barth's Christocentric interpretation of the Bible involved a return to the traditional relationship between creation and redemption, as well as to the Lutheran dialectic of Law and Gospel. As creation must be seen in the light of redemption, so the Law must be understood in the light of the Gospel.

Barth's dogmatics is a systematic theology written within the Church for the Church. The ministry of the Church is threefold: it consists of the declaration, explication and application of God's Gospel. Such ministry may take the form of speech or action. Although salvation primarily concerns human existence, it is essentially eschatological. Hence the Church anticipates the *eschaton*, the fulfilment of redemption.

4 Philosophy

The relationship of Barthian theology to philosophy is ambiguous. Barth made ample use of philosophical structures and concepts, while at the same time claiming the independence of theology from philosophy. He ruled out any possible alliance, even one that would result in philosophy being construed as the handmaid of theology.

Barth discarded any philosophical preconditions for faith. The tautological axiom ('God is God') and the Christocentric principle preclude any knowledge of God from other sources. In Barth's thought, unlike Calvin's, there is no room for a general revelation of God, be it in human consciousness, in nature or in history. Thus he declared war against natural theology, a stance exemplified in the notorious dispute with Emil Brunner (§§2, 4) (Barth 1934b). For Barth, natural theology was closely associated with Roman Catholicism and particularly with the Thomist doctrine of the analogy of being. He objected on three major grounds: first, the thinking process starts from below, from finite human understanding; second, it sunders God's essence, which is naturally knowable, from God's action, which is known supernaturally by means of revelation; third, it jeopardizes the exclusiveness of God's revelation in Jesus Christ. Instead, Barth proposed an analogy of faith, or of relation. A faith relationship to God is the prerequisite for any discourse about God. And any true faith relationship exclusively depends on God's self-revelation in Jesus Christ. Furthermore, God's relationship to humanity depends on God's self-relationship, on the interrelatedness of the Trinity as revealer, revelation and revealed (*Offenbarer*, *Offenbarung* and *Offenbarsein*).

In 'Philosophie und Theologie' (1960), Barth appraises the possibility of philosophy and theology coexisting. He observes that philosophers may choose their starting-points whereas theologians are bound to start with the Incarnation. Starting from below, philosophizing is an ascending process. It moves upwards from the phenomenon to the idea, from reason to being. By contrast, theologizing is a retracing of God's condescending movement. Therefore the philosopher and the theologian are separated by a matter of order and priority. They may share a common objective, but their methods are incompatible.

JEAN-LOUP SEBAN

List of works

Barth, K. (1971-) *Gesamtausgabe (Complete Edition)*, Zurich: Theologischer Verlag. (Only a few volumes of this massive collection have been published so far.)

Barth, K. (1911) 'Jesus Christus und der soziale Bewegung', trans. G. Hunsinger, 'Jesus Christ and the Movement for Social Justice', in G. Hunsinger (ed.) *Karl Barth and Radical Politics*, Philadelphia, PA: Westminster Press, 1976.(In this lecture, delivered on 17 December 1911 in Safenwil, Barth sets out his early view of religious socialism as it relates to faith in Jesus Christ.)

Barth, K. (1919a) 'Der Christ in der Gesellschaft', in *Das Wort Gottes und die Theologie (The Word of God and Theology)*, Munich: Chr. Kaiser Verlag, 1924; trans. D. Horton, 'The Christian in Society', in *The Word of God and the Word of Man*, London: Hodder & Stoughton, 1928.(In this lecture, delivered on 25 September 1919, Barth distinguishes himself from various forms of religious socialism which do not do justice to God's otherness.)

Barth, K. (1919b) *Der Römerbrief*, Bern: Bäschlin; 2nd edn, much changed, Munich: Chr. Kaiser Verlag, 1922; 6th edn trans. E.C. Hoskyns, *The Epistle to the Romans*, Oxford: Oxford University Press, 1968.(A proclamation of the otherness and judgment of God, who nevertheless saves; this book had an incalculable impact on European theology.)

Barth, K. (1922) 'Das Wort Gottes als Aufgabe der Theologie' (The Word of God as the Task of Theology), in *Das Wort Gottes und die Theologie (The Word of God and Theology)*, Munich: Chr. Kaiser Verlag, 1924; trans. D. Horton, 'The Word of God and the Task of the Ministry', in *The Word of God and the Word of Man*, London: Hodder & Stoughton, 1928.(Barth expanded on his post-liberal, dialectical vision of theology in this lecture given on 3 October 1922 in Elgersberg.)

Barth, K. (1927) *Die christliche Dogmatik in Entwurf, 1: Die Lehre vom Worte Gottes (Christian Dogmatics in Outline, 1: The Doctrine of the Word of God)*, Munich: Chr. Kaiser Verlag.(An abandoned attempt at a full dogmatics.)

Barth, K. (1931) *Fides quaerens intellectum: Anselms Beweis der Existenz Gottes in Zusammenhang seines theologischen Programmes (Faith Seeking Understanding: Anselm's Proof of the Existence of God in Relation to his Theological Programme)*, Munich: Chr. Kaiser Verlag; trans. I.W. Robertson, *Anselm: Fides quaerens intellectum*, London: SCM Press, 1960.(Held by some to mark the transition from Barth's dialectical theology to a more expansive, substantive theology of the Word.)

Barth, K. (1932-70) *Die Kirchliche Dogmatik* 1 - 4, with index vol., Munich: Chr. Kaiser Verlag, and Zurich: Evangelischer Verlag; trans. G.W. Bromiley *et al.*, ed. G.W. Bromiley and T.F. Torrance, *Church Dogmatics*, Edinburgh: T. & T. Clark, 1956-77.(Barth's greatest, and longest, work: a sadly unfinished interpretation of all the major Christian doctrines.)

Barth, K. (1934a) 'Theologische Erklärung zur gegenwärtigen Lage in der Deutschen Evangelischen Kirche'

(Theological Declaration on the Present Situation in the German Evangelical Church), in *Junge Kirche* 2, Göttingen; trans. R. McAfee Brown, ‘The Barmen Confession’, in G. Cassalis, *Portrait of Karl Barth*, New York: Doubleday, 1963.(Barth was heavily involved in drafting this declaration directed against the capitulation of German Christians to Hitler.)

Barth, K. (1934b) ‘Nein! Antwort an Emil Brunner’, *Theological Existenz heute* 14; trans. P. Fraenkel, ‘No! Answer to Emil Brunner’, in K. Barth and E. Brunner, *Natural Theology*, London: Geoffrey Bles, 1946.(A famous and blunt rejection of natural theology.)

Barth, K. (1947a) *Dogmatik im Grundriss*, Zollikon-Zurich: Evangelischer Verlag; trans. G.T. Thomson, *Dogmatics in Outline*, London: SCM Press, 1949.(Not to be confused with Barth 1927 above, this is a beautifully clear introduction to Barth’s later work.)

Barth, K. (1947b) *Die Protestantische Theologie im 19. Jahrhundert*, Zollikon-Zurich: Evangelischer Verlag; trans. B. Cozens and J. Bowden, *Protestant Theology in the Nineteenth Century*, London: SCM Press, 1972. (Barth’s history of the century of theology after Schleiermacher, against which he reacted.)

Barth, K. (1954) *Against the Stream: Shorter Post-War Writings 1946-52*, trans. E.M. Delacour and S. Godman, London: SCM Press, 1954.(Political and theological talks and articles.)

Barth, K. (1960) ‘Philosophie und Theologie’, in G. Huber (ed.) *Philosophie und christliche Existenz: Festschrift für Heinrich Barth zum 70 Geburtstag am 3.2.1960*, Basle and Stuttgart: Helbing & Lichtenhahn.(Sets out Barth’s understanding of the relationship between these two human enterprises.)

Barth, K. (1961) *The Humanity of God*, trans. J. Newton Thomas, Richmond, VA: John Knox Press.(Several shorter pieces; in the title essay, Barth describes how and why his position has modified over the years.)

Barth, K. (1967) *Ad limina apostolorum*, Zurich-Zollikon: Evangelischer Verlag; trans. K.R. Crim, *Ad limina apostolorum: An Appraisal of Vatican II*, Richmond, VA: John Knox Press, 1968.(As this work shows, although Barth remained wary of Roman Catholicism, he had great sympathy for much that he found, particularly in Vatican II.)

References and further reading

Balthasar, H.U. von (1976) *Karl Barth: Darstellung und Deutung seiner Theologie*, Einsiedeln: Johannes Verlag, 4th edn; trans. E.T. Oakes, *The Theology of Karl Barth: Exposition and Interpretation*, San Fransisco, CA: Communio Books, 1992.(One of the most influential of commentaries on Barth, now a little dated but still important.)

Busch, E. (1976) *Karl Barth: His Life from Letters and Autobiographical Texts*, Philadelphia, PA: Fortress Press. (The standard biographical source.)

Ford, D. (1981) *Barth and God’s Story: Biblical Narrative and the Theological Method of Karl Barth in the ‘Church Dogmatics’*, Frankfurt am Main and Bern: Verlag Peter Lang.(The best introduction to Barth’s hermeneutics, drawing on P. Stern’s work on realism.)

Frei, H. (1992) *Types of Christian Theology*, ed. G. Hunsinger and W. Placher, New Haven, CT: Yale University Press.(Contains several sections on Barth, and, in particular, on the relationship of theology and philosophy in his work; argues that there is an *ad hoc* correlation between the two in Barth.)

Hunsinger, G. (1991) *How to Read Karl Barth: The Shape of his Theology*, Oxford: Oxford University Press.(An introduction to reading the *Church Dogmatics*, and a study of Barth’s treatment of truth and freedom.)

Jüngel. E. (1986) *Karl Barth: A Theological Legacy*, trans. G.E. Paul, Philadelphia, PA: Westminster Press. (Various essays by perhaps the greatest theological interpreter and follower of Barth.)

McCormack, B. (1995) *Karl Barth’s Critically Realistic Dialectical Theology: Its Genesis and Development 1909-1936*, Oxford: Clarendon Press.(The most detailed and lucid survey in English of Barth’s development, challenging many received understandings.)

Sykes, S. (ed.) (1989) *Karl Barth: Centenary Essays*, Cambridge: Cambridge University Press.(Several fine essays, including Ingolf Daflerth on Barth and Kantianism.)

Webster, J. (1995) *Barth’s Ethics of Reconciliation*, Cambridge: Cambridge University Press.(A clear, volume-by-volume survey of the *Dogmatics*, particularly ethical issues, and an investigation of ‘moral ontology’ in Barth’s work.)

Wildi, H.M. (1984-92) *Bibliographie Karl Barths*, Zurich: Theologischer Verlag, 3 vols.(A complete bibliography of many thousands of primary and secondary items, with comprehensive indices.)

Barthes, Roland (1915-80)

In the field of contemporary literary studies, the French essayist and cultural critic Roland Barthes cannot be easily classified. His early work on language and culture was strongly influenced by the intellectual currents of existentialism and Marxism that were dominant in French intellectual life in the mid-twentieth century. Gradually his work turned more to semiology (a general theory of signs), which had a close association with the structuralist tradition in literary criticism. In his later work, Barthes wrote more as a post-structuralist than as a structuralist in an attempt to define the nature and authority of a text. Throughout his writings Barthes rejected the 'naturalist' view of language, which takes the sign as a representation of reality. He maintained that language is a dynamic activity that dramatically affects literary and cultural practices.

1 Early writings

Roland Barthes was born in Cherbourg, France and studied French literature and Classics at the University of Paris. After a long illness that caused him to spend time in sanatoria during the war years, he taught French at the Universities of Bucharest and Alexandria, and began further studies in sociology. In 1947 he published a number of articles in *Combat* on literary criticism that formed the basis of his first book, *Le Degré zéro de l'écriture* (1953) (*Writing Degree Zero*, 1968). In 1960 he began teaching at the École Pratique des Hautes Études, and in 1976 he became Professor of Literary Semiology at the Collège de France. Barthes died after being struck by a van outside the Collège de France.

In *Writing Degree Zero* Barthes formulated a theory of writing on the basis of an account of language as a 'social object' and a 'field of action' in which the writer communicates. This account has much in common with the existentialist writer Jean-Paul Sartre, who, like Barthes, sought to break free of the orthodox view of literature as a-historical and considered it ideological in its unquestioned acceptance of the bourgeois universe. Against this view the writer is to be a liberator. Unlike Sartre, however, Barthes came to the conclusion that the writer can do little directly to effect social change and that the writer does not write exclusively for the sake of society. The writer also writes for the sake of writing. Barthes uses the special term 'writing' (*écriture*) to capture this double aspect of the writer. Writing is the middle ground between language, as that which resides in history (the material of writing), and style, as that which is indifferent to history and society (personal contribution to writing). Accordingly, writing is neither strictly historical nor strictly personal; it is an ambiguous reality arising as a confrontation of the writer with society and at the same time referring the writer back to the instruments of creation.

Increasingly Barthes came to adopt a more complex view of language, which enabled him further to develop his critique of the status quo. In general this view followed the linguistic structuralism of the Swiss linguist Ferdinand de Saussure, for whom language as a whole was composed of *langue*, the system of signs that allows for the construction of words in an oppositional phonological play, and *parole*, the actual experience of speaking. From the point of view of semiological analysis (analysis of signs) Barthes opposed the prevailing literary ideology that took the sign as a natural representation of reality rather than an arbitrary convention. This criticism was not limited to literature, which he regarded as only one among many signifying systems. Food, clothing, film, advertising and fashion were also viewed as signifying systems. In this broadened universe of signs Barthes directed his criticism at the bourgeois myths that naturalized objects and events so as to exempt them from political change. In *Mythologies* (1957), for example, Barthes showed how something as simple as a photograph appearing on the cover of the magazine *Paris-Match* can signify bourgeois myth. The photograph depicts a black soldier saluting the flag, signifying that France is a great empire without regard to race. Barthes, however, decoded this sign as an implicit defence of French colonialism. In both literary studies and cultural criticism Barthes' goal was to demythologize the sign system that fulfilled an ideological function (see: Structuralism; Structuralism in literary theory).

2 Post-structuralist criticism

Along with other French intellectuals in the late 1960s, Barthes began to distance himself from the possibility of a science of signs and expanded his critique of writing by adhering to a post-structuralist notion of the text (see Post-structuralism). In his critical analysis of a novella of Balzac, *S/Z* (1970), Barthes insisted that the literary text

must be explored in terms of the way the text outplays the literary codes which structuralism relies upon to make the text intelligible. The theoretical basis for this changed position is laid out in the brief essay 'The Death of the Author' (1968, in Barthes 1977a). Barthes pointed out that the author is a modern figure, the product of a society that has discovered the prestige of the individual. The author is regarded as the father and owner of the text, as the final signified, which preserves the unity of the text and to which all reading is directed. But Barthes insisted that what is written cannot be reduced to the authority of the author any more than language can be viewed as an unambiguous instrument of communication. Drawing from such literary writers as Joyce and Mallarmé, Barthes claimed that to write is to reach the point where it is not the subject who acts, but language. Writing destroys every point of origin so that the author is never more than the instance writing, 'just as I is nothing other than the instance saying *I*' (Barthes 1977a: 145).

With the death of the author the modern text is in the hands of the new 'scriptor' who practises a different kind of writing, one that no longer imposes a limit on the proliferation of signification. Rather than working from a given meaning, the modern scriptor works towards meaning by writing 'intransitively', infinitely deferring the signified and thereby producing a plurality of meaning. Such writing is viewed as a form of liberation, an anti-theological activity refusing God and his hypostases - reason, science, law. In this new mode of writing, traditional criticism, whether biographical, historical or formalist, is undermined. The role of the critic is not to decipher the meaning, looking for the secret of the text, but to disentangle the plurality of meaning, even transposing meaning from the text into different discourses.

Because of the work of the modern 'scriptor', the reader now assumes the important role of providing the unity of the text. This reader, though, unlike the author, is not personal since the reader is without history, biography or psychology. The reader is simply 'that someone who holds together in a single field all the traces by which the written text is constituted' (Barthes 1977a: 148) In holding the text together in this way, the reader no longer regards the text as an object of consumption, but as an activity of production. The reader becomes a collaborator in the very execution of the text, setting the text free towards meaning.

During the 1970s Barthes continued to develop ways to express his scepticism of normative notions in textual criticism. In *Le Plaisir du texte* (1973) (*Pleasure of the Text*, 1975), which theoretically draws loosely on the framework of psychoanalysis, he described reading as a form of pleasure or bliss (*jouissance*) that allows for a free play of meaning from the text. Reading is not unlike a sexual act with the erotic body of language; it is unpredictable, fleeting, and culminates, like orgasm, in its disappearance. The result of this approach to texts was the espousal of discourse as fragment. Stylistically Barthes' own writing became purposefully fragmentary. His autobiography *Roland Barthes par Roland Barthes* (1975) is a classic illustration of a text without unity. In place of an account of a subject with a coherent identity, Barthes writes, in the third person, on the experience of writing and on the distinctive themes that marked his own literary production.

See also: Deconstruction

JAMES RISSER

List of works

Barthes, R. (1953) *Le Degré zéro de l'écriture*, trans. A. Lavers and C. Smith, *Writing Degree Zero*, New York: Hill & Wang, 1968.

Barthes, R. (1957) *Mythologies*, trans. A. Lavers, *Mythologies*, London: Paladin, 1972.

Barthes, R. (1964) *Essais critiques*, trans. R. Howard, *Critical Essays*, Evanston, IL: Northwestern University Press, 1972.

Barthes, R. (1966) *Critique et vérité*, trans. K. Keuneman, *Criticism and Truth*, Minneapolis, MN: University of Minnesota Press, 1987.

Barthes, R. (1970) *S/Z*, trans. R. Miller, *S/Z*, New York: Hill & Wang, 1974.

Barthes, R. (1973) *Le Plaisir du texte*, trans. R. Miller, *The Pleasure of the Text*, New York: Hill & Wang, 1975.

Barthes, R. (1975) *Roland Barthes par Roland Barthes*, trans. R. Howard, *Roland Barthes by Roland Barthes*, New York: Hill & Wang, 1977.

Barthes, R. (1977a) *Image-Music-Text*, trans. S. Heath, New York: Hill & Wang.(A selection of Barthes' uncollected essays, including 'The Death of the Author' and 'From Work to Text'.)

Barthes, R. (1977b) *Fragments d'un discours amoureux*, trans. R. Howard, *A Lover's Discourse: fragments*, New York: Hill & Wang, 1978.

References and further reading

Culler, J. (1983) *Barthes*, New York: Oxford University Press.(A good introduction to Barthes' work.)
Lavers, A. (1982) *Roland Barthes: Structuralism and After*, Cambridge, MA: Harvard University Press.(A detailed account of Barthes' philosophy concentrating mainly on Barthes' earlier writing.)
Lombardo, P. (1989) *The Three Paradoxes of Roland Barthes*, Athens, GA: University of Georgia Press.(A highly nuanced reading of Barthes' writing, considered thematically.)

Bartolus of Sassoferrato (or Saxoferrato) (1313/14-57)

The Bartolist school of civil lawyers or 'commentators' dominated university law teaching from the fourteenth century. Challenged by the humanists in the sixteenth century, they remained influential in practice. Bartolus excelled among them in the ability to devise solutions to practical problems and provide clear and workable doctrines applying the civil law texts to legal and political problems.

Born in Sassoferrato, Bartolus studied civil law under Cinus of Pistoia at Perugia from 1328 but his baccalaureate in 1333 and doctorate in 1334 were at Bologna. He practised for a period and was a legal assessor at Todi, then probably at Cagli and finally at Pisa in 1339. In 1339 he became professor at Pisa. In 1342-3 he moved to Perugia, remaining there until his death in 1357. Baldus was his pupil and, from about 1351, his colleague there.

The older name for the 'commentators' is 'postglossators' because they worked from texts of the *Corpus iuris civilis* (body of the civil law) provided in the thirteenth century with a massive apparatus of glosses, the Accursian or great Gloss, which in effect sums up the work of the preceding school of civilians, the glossators. Their commentaries, a product of their teaching, form a major part of their literary output and give them their name. There and in their other works, using the Aristotelian 'new logic', they draw out from text and gloss ideas which they use to solve current problems. The original context of the texts used is scarcely relevant. From civil (and canon) law texts they thus created a living common law to which appeal could be made where local sources were inadequate.

Bartolus was pre-eminent in authority, noted for his practical sense and his ability to get to the heart of a problem and expound it clearly, for example, in conflict of laws, where more than one legal system is potentially applicable to a legal dispute, and in the issue of sovereignty. His views were constantly referred to and commonly deferred to. Like all medieval lawyers he drew on his predecessors, using especially the ultramontane (French) writers of the later thirteenth century who strongly influenced his first teacher, Cinus. Sometimes, therefore, he is given credit for originating ideas to which he rather gave currency, for example, the distinction between 'real' statutes, applying to things (*res*) and so territorial in operation, and 'personal' statutes, applying to those subject to them everywhere. Where he did adopt ideas he nevertheless gave them his own stamp, and by adopting them he ensured their lasting influence.

Bartolus left a huge body of writing, but both manuscripts and prints contain false attributions. His main works were his commentaries (or *lecturae*, literally, readings) on the various parts of the *Corpus iuris civilis*, along with special lectures (*repetitiones*) some of which are preserved separately, advisory opinions to parties or judges (*consilia*), academic discussions of problem cases (*quaestiones*) and treatises (*tractatus*) on individual topics, legal and political.

See also: Law, philosophy of; Roman law

WILLIAM M. GORDON

List of works

(The dates of publication given are mainly the dates of first printing. There are numerous later editions up to the seventeenth century.)

Bartolus (1333-57) *Quaestiones (Academic Disputations)*, Venice, 1471.(Eighteen disputations; four more in Venice, 1521 edition.)

Bartolus (1339-52) *Lecturae (Commentaria)*, 1st edns from 1470-5 at various locations; facsimile repr. of edn Venice, 1526-9, by and with notes of T. Diplovatatius, 9 vols, Rome: Il Cigno Galileo Galilei.(Commentaries on the texts of the *Corpus iuris civilis*.)

Bartolus (1473) *Consilia (Advisory Opinions)*, Rome.(Both manuscripts and prints contain opinions by others; 244 opinions, 117 more published Venice 1521, a further 38 in later editions.)

Bartolus (1477) *Lectura super Authenticis (Commentaries on the Authenticae [Latin versions of the Justinianic Novels])*, Milan.(Authenticity doubted by Diplovatatius.)

Bartolus (1481-2) *Opera* or *Opera omnia (Works or Complete Works)*, Lyon.(Collected editions appear from Lyon, 1481-2, through to the early seventeenth century. Some, like the first, contain only the commentaries.)

Bartolus (1504) *Lectura Institutionum (Commentaries on the Institutes)*, Pavia.(False attribution; true author may well be Jacques de Révigny.)

Bartolus (1983) *Tractatus de Guelphis et Gebellinis; de regimine civitatis; de tyranno*, ed. D. Quaglioni, Politica e diritto nel trecento italiano, Florence: Il pensiero politico, biblioteca, vol. 11. (Treatises on Guelphs and Ghibellines; the government of a city; the tyrant.)

References and further reading

Bellomo, M. (1995) *The Common Legal Past of Europe, 1000-1800*, trans. L.G. Cochrane, Studies in Medieval and Early Modern Canon Law, vol. 4, Washington, DC: Catholic University of America Press, esp. 190-95. (Places Bartolus and the commentators in context.)

Canning, J. (1987) *The Political Thought of Baldus de Ubaldis*, Cambridge Studies in Medieval Life and Thought, vol. 6, Cambridge: Cambridge University Press.(Refers back to the seminal work of Bartolus; useful short biographies of major medieval civilians and canonists.)

Clarence Smith, J.A. (1970) 'Bartolo on the Conflict of Laws', *American Journal of Legal History* 14: 157-74, 247-75.(A translation of the sections of Bartolus' commentaries on the *Digest* and *Codex* dealing with conflict of laws, with a brief introduction on Bartolus and the background to his work.)

Segoloni, D. (ed.) (1962) *Bartolo da Sassoferrato. Studi e documenti per il VI centenario*, Milan: Giuffrè.(An invaluable multilingual collection of articles on the work and influence of Bartolus.)

Ullmann, W. (1980) *Jurisprudence in the Middle Ages: Collected Studies*, London: Variorum Reprints.(Useful general background and includes a study of Bartolus' views on customary law.)

Woolf, C.N.S. (1913) *Bartolus of Sassoferrato: His Position in the History of Medieval Political Thought*, Cambridge: Cambridge University Press.(A standard work, with somewhat dated biography; numerous quotations, mainly in Latin, from Bartolus and other sources; gives the gist of quotations but no translation.)

Bataille, Georges (1897-1962)

Georges Bataille was born in Billom, France, raised in Reims, and spent much of his adult life in Paris. Never formally trained as a philosopher, he worked from 1922 to 1942 as a librarian at the Bibliothèque *Nationale. In addition to his philosophical works, Bataille also wrote on the history of art as well as a number of critical works and novels.*

Owing to his position outside academic philosophy, Bataille was able to treat diverse topics in ways which might have been unacceptable otherwise. His work addresses the importance of sacrifice, eroticism and death, as well as the kinds of 'expenditure' evidenced by what he called the general economy. *It draws on diverse sources (Hegel, Nietzsche, Marcel Mauss, anthropological research, and the history of religion, among others) and treats a wide range of topics: the role of art in human life, the practice of sacrifice in ancient and modern cultures, the role of death in our understanding of subjectivity, and the limits of knowledge.*

Bataille's major works include: *L'Expérience intérieure* (1943) (*Inner Experience*, 1988; composed 1941-2; materials from 1924 on), *La Part maudite* (1949) (*The Accursed Share*, vol. 1, 1949; vols 2-3, 1976; written 1946-9; planned as early as 1930), *L'Érotisme* (1957) (*Erotism, Death and Sensuality*, 1986; planned from 1930), *Sur Nietzsche* (1945) (*On Nietzsche*, 1992; written mid-1940s) and *Théorie de religion* (1974) (*Theory of Religion*, 1989; composed late 1940s). He is known largely for his connections with Surrealism, although his alliance with the movement was often strained. He is also considered an important forerunner of 'postmodernism', although he died before the word was coined. Much of his work manifests a resistance to systematicity and a desire to produce texts (which he called *heterologies*) which escape unitary interpretations. Such heterologies concern what is entirely other, and thus resist being reduced to the identities necessary for thought and language. The attempt to think the heterologous means to think that which is outside, and thus to transgress the bounds of thought (which remains, in an important way, an impossible undertaking). Bataille names that which transcends the bounds of science, the everyday, and time the *Sacred*. His work on general economy, however, owing to the influence of Mauss, shows certain systematic and structuralist tendencies.

Many of Bataille's most important texts comprise what he called *La Somme athéologique*, which analyses the Sacred at both individual and societal levels. At the most individual level (that of *inner experience*), Bataille investigates the possibility of transcending our everyday understanding of individuality without losing all notions of subjectivity. These investigations are oriented towards no particular goal or knowledge, but begin with a 'phenomenological' account of experience itself. Drawing on Nietzsche, Bataille insists that experiences and sensations take place before there is a subject to experience them; the subject is only established in and through experience. Moreover, if one chooses the right sorts of experience (those of intense suffering), a point is reached where pain ceases to be felt and one's own subjectivity is transcended. Since this experience is often followed by death, research into this kind of transcendence is both dangerous and difficult to replicate. It was these dangers, at least in part, which led Bataille to develop alternative accounts of the move towards the Sacred. What remains constant, however, is the important place he gave to the notion of death.

Later, Bataille developed the important systematic notion of *general economy* with its emphasis on *expenditure*. Focusing on surplus and expenditure as the primary notions of economics (rather than on scarcity, as standard economic models do) allows Bataille to link his work on political economy with his work on inner experience, eroticism and religion, all of which he characterizes as displays of excess or surplus. Here the Sacred is expressed in social practices rather than individual experiences. By focusing on social behaviour which exceeds the limits of (instrumental) rational explanation (such as Amerindian *potlatch* ceremonies), Bataille highlights the myriad ways in which human life and practice resists rational description.

The concept of a general economy of energy flows allows one to analyse not only economic phenomena but social, anthropological, biological and physical ones as well. The fundamental problem with which the general economy must deal is that of excess. The earth has a constant supply of new solar energy which must be either taken up or dispersed in some way. Societies, which draw on this energy, quickly reach a point where production exceeds necessity. The process by which this excess is dealt with is expenditure, which often takes place in a way which expresses the Sacred: through sacrifice or warfare. Such sacrifice may take a literal form (as in Aztec society) or a

more figural one as in the modern culture of conspicuous consumption.

Practices such as sacrifice and warfare serve the Sacred by elevating those who are destroyed, together with that in whose name the destruction occurs, above the realm of mere things. Even the victim is elevated; for the destruction that sacrifice is intended to bring about is not annihilation (Bataille 1974: 43).

The Sacred in general removes things from the realm of mere usefulness and thus elevates them above time and its laws of necessity and causality. It not only leaves the realm of reason and discourse behind (which is part of what makes it so difficult for Bataille to discuss it), but actually destroys them (at least temporarily) as well. The human move towards the Sacred is thus beset with a major difficulty. On the one hand, the Sacred allows humans to separate themselves off from the realm of necessity by moving towards transcendence. But if such transcendence leaves the realm of necessity entirely behind, it results in death. Bataille's proposed solution to this problem is to limit moments of sacrifice so that their transcendence *of* time is still caught up *in* time. The first example of this is the festival; the second is war (which still results in death for many of its participants).

Bataille also treats the problem of transcendence through investigations of eroticism. Eroticism is simultaneously the most potent form of embodied experience and one which, like a Dionysian festival, transcends (at least for a time) bodily and temporal limits. And, like the festival, such experiences risk absolute annihilation but usually end with a return to the everyday (although not the same everyday).

See also: Postmodernism; Structuralism

JONATHAN MASKIT

List of works

Bataille, G. (1970-88) *Oeuvres complètes*, Paris: Gallimard.(The standard French language edition of Bataille's complete works. Much of this material is unavailable in English.)

Bataille, G. (1927-39) *Visions of Excess: Selected Writings, 1927-1939*, ed.A. Stoekl, trans. A. Stoekl, C.R. Lovitt and D.M. Leslie, Jr, Minneapolis, MN: University of Minnesota Press, 1985.(A good selection of some of Bataille's early, shorter works on a variety of topics.)

Bataille, G. (1943) *L'Expérience intérieure*, Paris: Gallimard; trans. L.A. Boldt, *Inner Experience*, Albany, NY: State University of New York Press, 1988.(The most important work in *La Somme athéologique*. Treats the problem of transcendence from an individual perspective.)

Bataille, G. (1945) *Sur Nietzsche*, trans. B. Boone, *On Nietzsche*, New York: Paragon House, 1992.(Part of *La Somme athéologique*. Far more an investigation of Bataille's Nietzscheanism than of Nietzsche.)

Bataille, G. (1945-51) *The Absence of Myth: Writings on Surrealism*, ed. and trans. M. Richardson, London: Verso, 1994.(Together with *Visions of Excess*, a good selection of Bataille's shorter pieces. These treat not only Surrealism but also the notion of the Sacred.)

Bataille, G. (1949) *La Part maudite*, vol. 1, *La Consommation*, Paris: Éditions de Minuit; trans. R. Hurley, *The Accursed Share*, vol. 1, New York: Zone Books, 1991.(Bataille's definitive text on general economy. The most important work in the later part of *La Somme athéologique*. Volume 1 treats historical sources and the importance of sacrifice.)

Bataille, G. (1957) *L'Érotisme*, Paris: Éditions de Minuit; trans. M. Dalwood, *Erotism, Death and Sensuality*, San Francisco, CA: City Lights, 1986.(Very similar to what would be published in English as volume 2 of *The Accursed Share*. Treats eroticism in connection with death and transcendence.)

Bataille, G. (1962) *Les Larmes d'Éros*, Paris: Éditions Pauvert; trans. P. Connor, *The Tears of Eros*, San Francisco, CA: City Lights, 1989.(Discusses the importance of the Sacred in art.)

Bataille, G. (1974) *Théorie de religion*, Paris: Gallimard; trans. R. Hurley, *Theory of Religion*, New York: Zone Books, 1989.(More a theory of sacrifice and transcendence than a conventional theological text. A shorter treatment of many of the topics addressed in *The Accursed Share*.)

Bataille, G. (1976) *The Accursed Share*, vols 2 and 3, trans. R. Hurley, New York: Zone Books, 1993.(Compiled for the *Oeuvres complètes* from (mostly) unpublished materials. Volume 2 is a history of eroticism, volume 3 treats the notion of sovereignty and the development of the Soviet economy.)

References and further reading

Bergfleth, G. (1975) *Theorie der Verschwendung (Theory of Extravagance)*, Munich: Matthes & Seitz, 1985.(A very good treatment of the distinction between restricted and general economy, but also of the notions of

heterology and sovereignty. Also includes a chapter on the relevance of Bataille's thought for environmentalist philosophy.)

Boldt-Irons, L.A. (ed.) (1995) *On Bataille: Critical Essays*, Albany, NY: State University of New York Press.(A good collection of recent scholarship on Bataille, offering clear treatments of several important issues: heterology, inner experience, expenditure, and Bataille's relation to the history of philosophy.)

Cordier, S. (ed.) (1962) *Bataille*, special issue of *L'Arc* 44, Aix-en-Provence.(Published the year of Bataille's death. Contributions from his colleagues. An interesting collection.)

Derrida, J. (1972) 'From Restricted to General Economy: A Hegelianism without Reserve', trans. A. Bass, in *Writing and Difference*, Chicago, IL: University of Chicago Press, 1978.(Treats the importance of Hegel in Bataille's work on general economy, thus arguing against more Nietzschean interpretations.)

Fourny, J.-F. (1988) *Introduction à la lecture de Georges Bataille (An Introduction to Reading Georges Bataille)*, New York: Peter Lang.(A good introduction to Bataille's work as a whole.)

Habermas, J. (1985) 'Between Eroticism and General Economics: Georges Bataille', trans. F. Lawrence, in *The Philosophical Discourse of Modernity*, Cambridge, MA: MIT Press, 1987.(A very clear treatment of the concept of general economy, stressing Bataille's Marxist leanings.)

Hollier, D. (1974) *Against Architecture*, trans. B. Wing, Cambridge, MA: MIT Press, 1990.(Treats the importance of architectural metaphors in Bataille as a way of understanding his resistance to systematicity. Often obscures as much as it clarifies.)

Richardson, M. (1994) *Georges Bataille*, London: Routledge.(A clear but somewhat misleading introduction. Seeks to show that Bataille is not, *pace* many other commentators, a postmodernist.)

Richman, M. (1982) *Beyond the Gift: Reading Georges Bataille*, Baltimore, MD: Johns Hopkins University Press. (A very good introduction to Bataille focusing on the importance of expenditure and general economy. Probably the clearest book on Bataille in English.)

Wiechens, P. (1995) *Bataille: zur Einführung (Bataille: An Introduction)*, Hamburg: Junius Verlag. (A good general introduction to all of Bataille's work. Clearer than anything available in English.)

Baudrillard, Jean (1929-)

Jean Baudrillard taught for most of his career at the Nanterre campus of the University of Paris in the department of sociology. He began writing as a neo-Marxist in the tradition of Henri Lefebvre and Herbert Marcuse but very quickly developed his distinctive style of social and cultural criticism. He may be understood generally as a post-structuralist who focused on the importance of language in society and invented novel concepts and terms to understand the most advanced features of electronic communications. He has been hailed as the guru of postmodernity and berated as faddish trend follower. He has written polemical pieces like Oublier Foucault (1977)(Forget Foucault, 1987) and controversial ones such as La Guerre du Golfe n'a pas eu lieu (1991b) (The Gulf War Did Not Take Place, 1995).

Baudrillard's significance for social philosophy rests with his effort to theorize aspects of popular culture that were generally regarded with disdain (by the Frankfurt School and many others) or uncritically celebrated (by Marshall McLuhan in particular). Baudrillard brought an unprecedented seriousness of theoretical purpose to the questions of the media and consumer culture. It is important to recall this fact because recently he has been denigrated as an unserious intellectual 'star'. From his earliest works, *Le Système des objets* (1968) (*The System of Objects*, 1996) and *La Société de consommation* (1970), Baudrillard attempted to account for the increased prominence of the media in society and the rising significance of consumerism in daily life. In these works he deployed Marxist categories but added semiological and psychoanalytic dimensions to his analysis. Consumption, he argued, must be approached as a system of signs with its own internal quality of articulation, not explained away by categories geared to the process of production or by liberal political complaints. Already in these early writings he employed a post-structuralist method to make intelligible a newly emerging social/cultural sphere. This was so because Baudrillard aimed his theorizing of consumer culture against the liberal logic of *homo economicus* as well as the productivist emphasis in Marxism. In order to theorize the world of consumption, he argues, one must abandon the 'modernist' theoretical fiction of the autonomous rational individual, one who always calculates their utilities for happiness, and abandon as well the socialist view of the worker as a resisting agent upon whom capitalism imposes structures of domination. Baudrillard's early theory of consumer culture was post-structuralist in the double sense that it understood the subject as historically constituted and it recognized the salient role of language in that process of constitution. By bringing these interpretive strategies to the domain of everyday life Baudrillard revolutionized our understanding of them.

Baudrillard definitively broke with Marxism in his works of the mid-1970s: *Pour une critique de l'économie du signe* (1972) (*For a Critique of the Political Economy of the Sign*, 1981), *Le Miroir de la production* (1973) (*The Mirror of Production*, 1975) and *L'Échange symbolique et la mort* (1976) (*Symbolic Exchange and Death*, 1993). In these books Baudrillard specified how media culture and advertising in particular constituted a 'code' which had great influence on individuals. The 'code' was the essence of consumption. People consumed not so much objects but images, ideals, fantasies, styles - all of which were structured through advertising and presented in the electronic media, a strange new dimension of social life which altered forever the older 'bourgeois' culture of modern society. Individuals, Baudrillard argued, construct themselves, or better, their identities in and through their response to advertising and the media. It was pointless to bewail the quality of television shows or consumerism as many liberals and Marxists did because 'vulgarity' and 'exploitation' were irrelevant to the new consumer world. In the malls and shopping centres, in radio and TV ads, a culture was constructed that was capturing the attention and the imagination not only of the masses in the industrialized societies but in the Communist societies of Eastern Europe and in much of the 'Third World' as well. The problem for social theory was to understand how this new world of signification worked, not simply to condemn it in a futile gesture of snobbery. Baudrillard turned to semiology to analyse the structure of advertising and he demonstrated how ads restructured language in such a way that the word 'Coca Cola' would refer not so much to a brown bubbly liquid, its 'referent', but to the images of youth, sexuality and fun that were presented in ads for the product. In the 1970s Baudrillard went so far as to argue in *À l'ombre des majorités silencieuses* (1978) (*In the Shadow of the Silent Majorities*, 1993) that the masses' apparent apathy and lack of revolutionary fervour was in actuality a new form of resistance against the system.

In a series of works in the 1980s and 1990s Baudrillard developed his analysis of media advertising into a general

understanding of postmodern culture. From *Simulacres et simulations* (1981) (*Simulacra and Simulations*, 1994) to *L'Illusion de la fin* (1991a) Baudrillard presented an important argument that reality itself was changing as a consequence of the new consumer culture. As people spent more and more time with electronic communications (tuned into the radio, glued to TVs, jacked into computers, turned on to walkmen and ghetto blasters, conversing on telephones, sending faxes, receiving e-mail), more time exchanging symbols through the mediation of increasingly smart machines, the world of face-to-face was becoming the world of the 'interface'. Baudrillard called this emerging culture 'the hyperreal'. Hyperreality was built upon new cultural principles. Symbolic constructions were no longer rooted in an original reference such as a spoken conversation or a written letter. Now language was increasingly 'simulational' in the sense that the presentation is always both an original and a copy. The TV news does not really report about something in an 'external' world: it makes important what it states, creating news as it 'reports' about it. This difficult logic, 'Hyperreality', increasingly dominates the exchanges of words and images, gradually forming a new and very strange culture. If Baudrillard's importance rests on the attention he paid to new cultural formations, his limitations are an overly pessimistic assessment of them, a failure to recognize their limitations, and an inability to take into account new assemblages of humans and machines.

See also: Postmodernism; Post-structuralism

MARK POSTER

List of works

Baudrillard, J. (1968) *Le Système des objets*, Paris: Gallimard; trans J. Benedict, *The System of Objects*, New York: Verso, 1996.(In this ground-breaking study of consumer culture, Baudrillard throws some light on the vexed question of the language of commodities.)

Baudrillard, J. (1970) *La Société de consommation (The consumer society)*, Paris: Gallimard.(A systematic portrait of the twentieth-century consumer culture.)

(1972) *Pour une critique de l'économie du signe*, Paris: Gallimard; trans. C. Levin, *For a Critique of the Political Economy*, St Louis, MN: Telos Press, 1981.(Perhaps Baudrillard's theoretical masterpiece - a comparison of the mode of production to the mode of consumption.)

Baudrillard, J. (1973) *Le Miroir de la production*, Paris: Casterman; trans. M. Poster, *The Mirror of Production*, St Louis, MN: Telos Press, 1975.(Baudrillard's critique of, and farewell to, Marxism.)

Baudrillard, J. (1976) *L'Échange symbolique et la mort*, Paris: Gallimard; trans. I. Hamilton Grant, *Symbolic Exchange and Death*, London: Sage, 1993.(Baudrillard's theory of an alternative semiotic society, based on the model of the desert.)

Baudrillard, J. (1977) *Oublier Foucault*, Paris: Éditions Galilée; trans.S. Lotringer, *Forget Foucault*, New York: Semiotext(e), 1987.

Baudrillard, J. (1978) *À l'ombre des majorités silencieuses, ou la fin de la social*, Paris: Utopie; trans. P. Foss, P. Patton and J. Johnston, *In the Shadow of the Silent Majorities - or, The End of the Social and Other Essays*, New York: Semiotext(e), 1983.

Baudrillard, J. (1979) *De la séduction*, Paris: Éditions Galilée; trans. B. Singer, *Seduction*, Montreal: New World Perspectives, 1990.

Baudrillard, J. (1981) *Simulacres et simulations*, Paris: Éditions Galilée; trans. S. Faria Glaser, *Simulacra and Simulations*, Ann Arbor, MI: University of Michigan Press, 1994.(Baudrillard's most influential work; includes his theory of media effects as a new form of reality, the hyperreal.)

Baudrillard, J. (1983) *Les Stratégies fatales*, Paris: Bernard Grasset; trans. P. Beitchman and W.G.J. Niesluchowski, *Fatal Strategies*, New York: Semiotext(e), 1990.(A tragi-comedy of objects.)

Baudrillard, J. (1990) *La Transparence du mal: essai sur les phénomènes extrêmes*, Paris: Éditions Galilée; trans. J. Benedict, *The Transparency of Evil: Essays on Extreme Phenomena*, London: Verso, 1993.

Baudrillard, J. (1991a) *L'Illusion de la fin (The illusion of the end)*, Paris: Éditions Galilée.

Baudrillard, J. (1991b) *La Guerre du Golfe n'a pas eu lieu*, Paris: Éditions Galilée; trans. P. Patton, *The Gulf War Did Not Take Place*, Bloomington, IN: Indiana University Press, 1995.

Baudrillard, J. (1995) *Le Crime parfait*, Paris: Éditions Galilée; trans. C. Turner, *The Perfect Crime*, London: Verso, 1996.

References and further reading

Kellner, D. (1989) *Jean Baudrillard: From Marxism to Postmodernism and Beyond*, Stanford, CA: Stanford University Press.(A comprehensive, critical review of Baudrillard's ideas in relation to social theory.)

Kellner, D. (ed.) (1994) *Baudrillard: A Critical Reader*, New York: Blackwell. (Various views and evaluations of Baudrillard's work and standing.)

Poster, M. (ed.) (1988) *Jean Baudrillard: Selected Writings*, Stanford, CA: Stanford University Press. (A sample of Baudrillard's writings, with an introduction by the editor.)

Bauer, Bruno (1809-82)

The career of the Hegelian theologian Bruno Bauer is marked by his sudden turn from a reasoned defender of Christianity into one of its most extreme critics. His radical interpretation of Hegel's philosophy, which he first used to defend orthodox biblical hermeneutics, ultimately led him to become, as one of his admirers said, the 'Robespierre of theology'. As the leader of the so-called 'Young Hegelian' school, Bauer was one of Hegel's most gifted students. However, his condemnation of theology in general and his thesis that the New Testament was merely the fictional product of an unknown author contributed to the general distrust of Hegelianism among religious thinkers. Although his many theological and historical writings now remain largely unread, his 'Critical Philosophy' and his radical atheism exerted a strong influence upon Marx, who was his student and friend, and is still evident in such contemporaries as Jürgen Habermas.

1 The Young Hegelian

Bauer entered the University of Berlin in 1828, and gave himself completely over to Hegelianism. From the earliest period of his career until his death, Bauer was concerned with the 'reconciliation' of Hegelianism and orthodox religion - the theme of his doctoral dissertation and final works (see Hegel, G.W.F. §8; Hegelianism §2). However, as his thought developed, Bauer took 'reconciliation' to mean that the religious mind, which had created a history of God out of its own unconsciousness, could only cure itself by coming into a full and critical self-consciousness towards its own unconscious fantasies.

In his first years as a lecturer, Bauer published forty-three articles and reviews. In these first writings, it was his intention to elevate theological consciousness to a speculative level which would resolve the debates between faith and critical reason, a resolution which would occur in the higher synthesis of Hegelian speculation.

In 1835, David F. Strauss' *Das Leben Jesu (Life of Jesus)* was published (see Strauss, D.F. §1). For Strauss, the New Testament was fundamentally a literary creation generated out of the Messianic expectations of the Jewish people, having little or no historical foundation. The miracle stories were merely pre-Christian 'myths'. The orthodox Hegelians, whose careers were threatened by Strauss' claim that his work was inspired by Hegel, asked Bauer to refute Strauss. Bauer's reply to Strauss, which appeared in a series of articles, attempted to demonstrate that such gospel miracles as the Virgin Birth were the necessary consequences of the historical development of human self-consciousness. Bauer's refutation contained the seed of Bauer's own radical view of the Gospels: that they were merely the fictional creations unconsciously designed to satisfy the needs of the religious mind for some external salvation.

In 1838, Bauer first formulated his own view of the relationship between philosophy and religion in a two-volume work: *Die Religion des Alten Testaments in der geschichtlichen Entwicklung ihrer Principien dargestellt (The Religion of the Old Testament Presented in its Historical Development and Principles)*. This work continued Bauer's efforts to treat the Gospel stories as the unconscious expressions of the religious mind, and to trace their historical development. By 1840, he had written a number of multi-volume studies devoted to explaining that biblical history was fundamentally an imaginative exercise of the religious mind, with little or no actual basis in fact. This thesis was set forth in his *Kritik der evangelischen Geschichte des Johannes (Critique of the Gospel of John)*. From this point on, Bauer would identify his own position as that of 'critic' and would term his philosophy 'criticism'.

By the end of June 1840, Bauer began another major study of the gospels, the three-volume *Kritik der evangelischen Geschichte der Synoptiker (Critique of the Synoptic Gospels)*. In this work the problem of whether or not Jesus was in fact a historical figure was finally resolved: 'To the question of whether Jesus was an authentic historical figure we replied that everything relating to the historical Jesus, all that we know of him, relates to the world of fancy, to be more exact - to Christian fancies. This has no connection with any man who lived in the real world'. In the autumn of 1840, Bauer concluded the writing of *Die Posaune des jüngsten Gerichts über Hegel den Atheisten und Antichristen: Ein Ultimatum (The Trumpet of the Last Judgement Against Hegel the Atheist and Antichrist: An Ultimatum)*. This anonymous work was to rally all the various interpreters of Hegel into one camp from which they could enter into what he termed 'The Campaign of Pure Criticism'. It asserted that the prudent 'Old Hegelians' associated with German academic life had consciously concealed the total incompatibility of

Hegelian philosophy with traditional Christian belief and conservative political order.

In the early 1840s Bauer was the leader of Berlin's notorious club, 'The Free Ones'. This group was the focal point of the radical 'Young Hegelians' - a forum for the discussion of atheism and radical politics. Among the participants were his own brother Edgar, the young Friedrich Engels, and a new friend, the radical individualist Max Stirner. However, by 1844, Bauer's theoretical and theological concerns were rapidly losing their importance among the radical Hegelians, as they were becoming more concerned with applying their theory towards the practical overthrow of the reactionary forces controlling German political life. Concerned with revolutionary deeds, and preparing for the establishment of a new democratic order, the new generation of radicals rejected Bauer's speculative 'Terrorism of Pure Theory'. Within a few years, Bauer's influence upon such revolutionary Hegelians as Marx and Engels came to a end.

2 Later years

In 1843, Bauer wrote *Das Entdeckte Christentum (Christianity Exposed)*, a work intended to bring about a general appreciation of atheism through a rough dissection of Christian attitudes. His belief that the Enlightenment and its 'Age of Reason' embodied his Hegelian trust in the rationality of the real occasioned his next project, the four-volume *Geschichte der Politik, Cultur und Aufklärung des achtzehnten Jahrhunderts (History of the Politics, Culture and Enlightenment of the Eighteenth Century)*. In 1846, having concluded his final volume, he continued his historical studies with the issue of the two-volume *Geschichte Deutschlands und der Französischen Revolution (History of Germany and the French Revolution)*. These works constituted a complete cultural and political history of the Enlightenment. Given Bauer's scorn of pietistic Christianity and his admiration for the atheists of the Enlightenment, it is not surprising that he considered it to be mankind's greatest age.

In 1847, Bauer, with his usual energy, wrote a three-volume study looking back upon the political events of his own lifetime: *Vollständige Geschichte der Parteikämpfe in Deutschland 1842-1846 (The Complete History of Party Struggles in Germany During the Years 1842-1846)*. Contrary to its title, the third volume covered only the events up to 1844; a fourth volume was to extend the time frame to 1847. But the Revolution of 1848 intervened, and the final volume was never completed. The volumes are a chronicle of disappointments, from the first deceptively liberal appearance of King Friedrich Wilhelm IV, to the increasingly oppressive measures taken by both state and church throughout the 1840s to stifle democracy. They are bitter works, angry at the passivity of the people, and their temper was reflected in shorter articles of the same period, such as 'Die Gattung und die Masse' ('The Genus and the Crowd').

In 1852, Bauer returned to theological subjects, and composed *Die theologische Erklärung der Evangelien (The Theological Explanation of the Gospels)* - intended to be the fourth and final volume of his earlier series, *Kritik der evangelischen Geschichte der Synoptiker*. This particular volume brought to a close his criticism of Strauss' 'Tradition Hypothesis' and Bauer developed his theory of the *Urevangelist* - the actual writer who composed the original fictional life of Jesus.

In the next year, with the Crimean War imminent, Bauer wrote a series of studies dealing with the relationship of Russia to the rest of Europe. The first, in 1853, was *Rußland und das Germanentum (Russia and the Germanic World)*. It warned of the growing power of Russia, and was followed by other studies focused upon the same issue, such as his 1854 *Deutschland und das Russentum (Germany and the Russian World)*. This work was a model for later German nationalistic philosophies of history, in which Germany was seen as the destined, yet scorned, leader of the West.

In 1859, after a few years in various insignificant editorial posts, Bauer became the associate of Hermann Wagener, the editor of such ultra-conservative papers as the *Staats-und-Gesellschaftslexikon*, *Kreuzzeitung* and *Berliner Revue*. Wagener was not only scornful of the weak democratic forces in Prussia, but virulently anti-Semitic. However, Bauer's anti-Semitism, as his early essay on the 'Jewish Question' indicates, was not an expression of Prussian nationalism, but concerned the nature of the Jewish religion itself. He condemned Judaism as the source and support of Christianity - the prime obstacle to human progress. Bauer wrote no anti-Semitic articles for Wagener.

In his last works, which he produced sporadically until his death in 1882, Bauer fully developed the theory which he had first proposed in the early 1840s - that it was not Jesus or Paul, but Seneca and Philo who were the spiritual

creators of the basic gospel story. Today, Bauer's biblical studies are almost totally forgotten, and perhaps rightly. However, his reading of Hegelianism as a revolutionary and atheistic humanism not only set the general course of Young Hegelianism, but remains a viable perspective.

See also: Religion and epistemology; Hermeneutics, biblical

LAWRENCE S. STEPELEVICH

List of works

Bauer, B. (1838) *Kritik der Geschichte der Offenbarung. I, vol. 1-2: Die Religion des Alten Testaments in der geschichtlichen Entwicklung ihrer Prinzipien dargestellt (The Religion of the Old Testament Presented in its Historical Development and Principles)*, Berlin: F. Dümmler; repr. Aalen: Scientia, 1983.(Initiates Bauer's lifelong theory that the biblical narratives of both Jews and Christians were largely fictional.)

Bauer, B. (1839) *Herr Dr. Hengstenberg. Kritische Briefe über den Gegensatz des Gesetzes und des Evangeliums (Critical Letters on the Opposition of Law and Church)*, Berlin: F. Dümmler.

Bauer, B. (1840a) *Die evangelische Landeskirche Preussens und die Wissenschaft (Prussian Protestantism and Science)*, Leipzig: Otto Wiegand; repr. Aalen: Scientia, 1972.

Bauer, B. (1840b) *Kritik der evangelischen Geschichte des Johannes (Critique of the Gospel of John)*, Bremen: Carl Schünemann.(A reduction of St John's Gospel into a pure work of fiction.)

Bauer, B. (1841a) 'Der christliche Staat und unsere Zeit' (The Christian State and Our Times), *Hallische Jahrbücher* 135-40 (7 June-12 July): 537-58; repr. in *Feldzüge der Reinen Kritik*, ed. and with intro. by H.M. Sass, Frankfurt: Suhrkamp, 1968.

Bauer, B. (1841b) *Die Posaune des jüngsten Gerichts über Hegel den Atheisten und Antichristen: Ein Ultimatum*, Leipzig: Otto Wiegand; repr. Aalen: Scientia, 1969; trans. L.S. Stepelevich, *The Trumpet of the Last Judgement Against Hegel the Atheist and Antichrist: An Ultimatum*, Lewiston, MA: Edward Mellen Press, 1989. (This analysis of Hegel's views on religion and politics, treating them as radical in the extreme, expressed the basic premises of Young Hegelianism.)

Bauer, B. (1841c) 'Theologische Schamlosigkeiten' (Theological Shamelessness), *Deutsche Jahrbücher* 117-20 (15-18 November): 465-79; repr. *Feldzüge der Reinen Kritik (The Campaign of Pure Criticism)*, ed. and with intro. by H.M. Sass, Frankfurt: Suhrkamp, 1968.(A bitter criticism of pietistic theologians.)

Bauer, B. (1841, 1842) *Kritik der evangelischen Geschichte der Synoptiker (Critique of the Synoptic Gospels)*, Leipzig: Otto Wiegand, 3 vols; repr. Hildesheim: Georg Olms, 1974.(A further refinement of Bauer's thesis regarding the fictional nature of Jesus Christ. Volume 3 was published in 1842.)

Bauer, B. (1842a) *Hegel's Lehre von der Religion und Kunst von dem Standpuncte des Glaubens aus beurtheilt (Hegel's Teaching on Religion and Art seen through the Judgment of Faith)*, Leipzig: Otto Wiegand; repr. Aalen: Scientia, 1967.

Bauer, B. (1842b) *Die gute Sache der Freiheit und meine eigene Angelegenheit (Freedom and My Own Situation)*, Zürich: Verlag des literarischen Comptoirs; repr. Aalen: Scientia, 1972.

Bauer, B. (1843a) *Die Judenfrage*, Braunschweig: Friedrich Otto; trans. H. Lederer, *The Jewish Problem*, Cincinnati, OH: Hebrew Union College, 1958.(Karl Marx's review of this work marks his transformation from being mainly concerned with philosophic and religious questions into a radical critic of capitalism.)

Bauer, B. (1843b) *Das entdeckte Christentum. Eine Erinnerung an das achtzehnte Jahrhundert und ein Beitrag zur Krisis des neunzehnten (Christianity Exposed)*, Zürich: Verlag des literarischen Comptoirs; repr. Aalen: Scientia, 1989.(This is Bauer's first and clearest statement of his atheism.)

Bauer, B. (1843c) *Der 20. Juni und der 10. August 1792 oder der letzte Kampf des Königthums in Frankreich mit der Volksparthei (The Last Struggle of the French Monarchy with the People)*, Charlottenburg: Egbert Bauer.

Bauer, B. (1843, 1844, 1845) *Geschichte der Politik, Cultur und Aufklärung des achtzehnten Jahrhunderts (History of the Politics, Culture and Enlightenment of the Eighteenth Century)*, Charlottenburg: Egbert Bauer, 4 vols; repr. Aalen: Scientia, 1965.(A positive assessment of Enlightenment rationalism. Volume 1 was published in 1843, volume 2 in 1844 and the last two volumes in 1845.)

Bauer, B. (1844a) *Die Septembertage und die ersten Kämpfe der Partheien der Republik in Frankreich, vols 1-2 (The First Struggles of the French Republican Parties)*, Charlottenburg: Egbert Bauer.

Bauer, B. (1844b) *Die Process Ludwig XVI und der 21 Januar 1793 (The Trial of Louis XVI and 21 January 1793)*, Charlottenburg: Egbert Bauer.

Bauer, B. (1844c) *Briefwechsel zwischen Bruno Bauer und Edgar Bauer während der Jahre 1839-1842 aus Bonn und Berlin (Letters between Bruno Bauer and Edgar Bauer out of Bonn and Berlin from 1839 to 1842)*, Charlottenburg: Egbert Bauer; repr. Aalen: Scientia, 1969.(A primary biographical source, and an insight into

how Bauer’s reading of Hegel antagonized the conservative theological faculty at Bonn University.)
Bauer, B. (1844d) *Actenstücke zu den Verhandlungen über die Beschlagnahme der ‘Geschichte der Politik, Cultur und Aufklärung des achtzenten Jahrhunderts’ (File Extracts of the Proceedings in the Confiscation Trial of Geschichte der Politik, Cultur und Aufklärung des achtzenten Jahrhunderts)*, Christiania: C.C. Werner.
Bauer, B. (1844e) ‘Was ist jetzt der Gegenstand der Kritik?’ (What is Now the Object of Criticism?), *Allgemeine Literaturzeitung* 8: 18-26; repr. in *Feldzüge der Reinen Kritik (The Campaign of Pure Criticism)*, ed. and with intro. by H.M. Sass, Frankfurt: Suhrkamp, 1968.(Traces the influence of critical philosophy from the French Revolution to German reform movements.)
Bauer, B. (1844f) ‘Die Gattung und die Masse’, *Allgemeine Literaturzeitung* 10 (Sept): 42-8; trans. L.S. Stepelevich, ‘The Genus and the Crowd’, in *The Young Hegelians: An Anthology*, Cambridge: Cambridge University Press, 1983, 198-205.(A criticism of the public indifference to philosophic progress.)
Bauer, B. (1846) *Geschichte Deutschlands und der Französischen Revolution unter der Herrschaft Napoleons (History of Germany and the French Revolution During Napoleon’s Rule)*, Charlottenburg: Egbert Bauer, 2 vols.(A positive evaluation of the effects of Napoleonic reforms on German society. First volume reprinted Aalen: Scientia, 1979.)
Bauer, B. (1847a) *Sturtz des Königthums und die ersten Kämpfe der Republik (Fall of the Monarchy and the First Republican Struggles)*, Leipzig: Voigt und Fernau.
Bauer, B. (1847b) *Der neue eröffnete Edelmann, oder Auswahl aus Edelmanns Schriften (Newly Published Edelmann, or a Selection of Edelmann’s Writings)*, Bern: Jenni.
Bauer, B. (1847c) *Vollständige Geschichte der Parteikämpfe in Deutschland während der Jahre 1842-1846 (The Complete History of Party Struggles in Germany during the Years 1842-1846)*, Charlottenburg: Egbert Bauer; repr. Aalen: Scientia, 1964.(A survey of the largely frustrated efforts of German political reformers anticipating the 1848 revolution.)
Bauer, B. (1849a) *Die bürgerliche Revolution in Deutschland seit dem Anfang der deutsch-katholischen Bewegung bis zur Gegenwart (The Peasants’ Revolution in Germany since the Beginning of the German Catholic Movement to the Present Day)*, Berlin: Gustav Hempel; repr. Aalen: Scientia, 1979.
Bauer, B. (1849b) *Der Untergang des Frankfurter Parlaments. Geschichte der deutschen constituirenden Nationalversammlung (The Fall of the Frankfurt Parliament. The History of the German Constitutionalizing National Assembly)*, Berlin: Friedrich Gerhard.
Bauer, B. (1849c) *Verteidigungsrede Bruno Bauers vor den Wahlmännern des vierten Wahlbezirks am 22. Februar (Bruno Bauers Speech in his Own Defence…)*, Berlin: Gustav Hempel.
Bauer, B. (1850) *Der Fall und der Untergang der neuesten Revolutionen (On the Failure of the Newest Revolutions)*, Berlin: Gustav Hempel.
Bauer, B. (1851-2) *Kritik der Evangelien und Geschichte ihres Ursprungs (Critique of the Gospels and History of their Origin)*, Berlin: Gustav Hempel, 4 vols; repr. Aalen: Scientia, 1983.(The final and most coherent statement of Bauer’s reading of the Gospel narratives as an example of religious fiction.)
Bauer, B. (1852) *Rußland und das Germanentum (Russia and the Germanic World)*, Charlottenburg: Egbert Bauer; repr. Aalen: Scientia, 1972.(An exercise in prophetic history, stressing the future role of Russia in European history.)
Bauer, B. (1853) *Kritik der paulinischen Briefe (Critique of the Letters of St Paul)*, Berlin: Gustav Hempel.
Bauer, B. (1854a) *Die jetzige Stellung Rußlands (The Current Situation in Russia)*, Charlottenburg: Egbert Bauer.
Bauer, B. (1854b) *Rußland und England (Russia and England)*, Charlottenburg: Egbert Bauer; repr. Aalen: Scientia, 1972.
Bauer, B. (1854-5) *Deutschland und das Russentum (Germany and the Russian World)*, Charlottenburg: Egbert Bauer.(A continuation of the thesis that Russia represents a threat to Germany. An early nationalistic treatise.)
Bauer, B. (1863a) *Das Judentum in der Fremde (Judaism in Foreign Lands)*, Berlin: Heinicke.
Bauer, B. (1863b) *Freimaurer, Jesuiten und Illuminaten, in ihrem geschichtlichen Zusammenhange (Freemasons, Jesuits and Illuminati, and their Historical Relationship)*, Berlin: Heinicke.
Bauer, B. (1874) *Philo, Renan und das Urchristentum (Philo, Renan and Primitive Christianity)*, Berlin: Gustav Hempel; repr. Aalen: Scientia, 1972.(Bauer’s argument that the Gospel Narratives were, in principle, composed by Philo.)
Bauer, B. (1877) *Christus und die Cäsaren: Der Ursprung des Christentums aus dem römischen Griechentum (Christ and the Caesars: the Origin of Christianity from the Greco-Romans)*, Berlin: Eugen Grosser.(Continues the argument that either Philo or Seneca composed the Gospel narratives.)
Bauer, B. (1879) *Einfluß des englischen Quäkertums auf die deutsche Kultur und auf das englisch-russische Projekt einer Weltkirche (Influence of English Quakerism on German Culture and on the Anglo-Prussian*

Ecumenical Project), Berlin: Eugen Grosser; repr. Aalen: Scientia, 1972.

Bauer, B. (1880a) *Jesus und die Cäsaren (Jesus and the Caesars)*, Berlin: Eugen Grosser. (A further focusing on the fictional nature of the Christian gospels.)

Bauer, B. (1880b) *Das Urevangelium und die Gegner der Schrift: 'Christus und die Cäsaren' (Primitive Protestantism and the Opponents of Christus und die Cäsaren)*, Berlin: Eugen Grosser.

Bauer, B. (1882) *Disraelis romantischer und Bismarcks sozialistischer Imperialismus (Disraeli's Romantic Imperialism and Bismarck's Social Imperialism)*, Chemnitz: Schmeitzner; repr. Aalen: Scientia, 1979.

Schläger, E. (ed.) (1882) 'Bruno Bauer und seine Werke' (Bruno Bauer and his Work), *Internationale Monatsschrift. Zeitschrift für allgemeine und nationale Kultur und deren Literatur* 1: 377-400; A. Zanardo (ed.) 'Bruno Bauer hegeliano e giovanne hegeliano', *Revista Critica di Storia della Filosofia*, Florence, 1965. (These contain the most comprehensive, yet still incomplete, listing of Bauer's periodical publications.)

Barnikol, E. (1972) *Bruno Bauer: Studien und Materialien (Studies and Material)*, ed. P. Reimer and H.-M. Sass, Assen: Van Gorcum.(This work provides the most comprehensive listing of Bauer's unpublished works.)

References and further reading

Barnikol, E. (1972) *Bruno Bauer: Studien und Materialien*, ed. P. Reimer and H.-M. Sass, Assen: Van Gorcum. (A 572-page gathering of some of Ernst Barnikol's posthumous papers and research materials dealing with Bauer. An invaluable source, well-edited, but unavoidably uneven in both form and content.)

Brazill, W.J. (1970) 'Bauer, Stirner, and the Terrorism of Pure Theory', in *The Young Hegelians*, New Haven, CT: Yale University Press, 175-225.(A lucid survey of the main figures within the Young Hegelian movement.)

Mah, H. (1987) 'Bruno Bauer and the Crisis in Religious Theory', in *The End of Philosophy, the Origin of 'Ideology'*, Berkeley, CA: University of California Press, Part 2, 45-86.(A readable survey directed towards the political, mainly Marxian, outcome of Young Hegelianism.)

Moggach, D. (1996) 'Bruno Bauer's Political Critique, 1840-1841', *The Owl of Minerva* 27 (2): 137-54.(A detailed study of recent interpretations of Bauer's political thought during the period in which the Young Hegelian school turned its attention from philosophical speculation to political action.)

Rosen, Z. (1977) *Bruno Bauer and Karl Marx*, The Hague: Martinus Nijhoff.(The most detailed study in any language dealing with the six-year relationship between Bauer and Marx.)

Stepelevich, L.S. (1983) 'Bruno Bauer', in *The Young Hegelians: An Anthology*, Cambridge: Cambridge University Press, 173-205.(A brief introduction followed by translations of part of Bauer's *The Jewish Problem* and *The Trumpet of the Last Judgement* and 'The Genus and the Crowd'.)

(1989) 'Translator's Introduction', to B. Bauer, *Trumpet of the Last Judgement Against Hegel the Atheist and Anti-Christ: An Ultimatum*, Lewiston, MA: Edward Mellen Press, 1-56.(The most extensive biography of Bauer yet written.)

Strauss, D.F. (1835-6) *Das Leben Jesu*, Tübingen: Osiander; trans. M.A. Evans [George Eliot], *Life of Jesus*, Philadelphia, PA: Fortress, 1972.(Argues that the messianic Christ was merely a mythic figure.)

Towes, J.E. (1980) 'Bruno Bauer and the reduction of absolute spirit to human self-consciousness', in *Hegelianism: The Path Toward Dialectical Humanism, 1805-1841*, Cambridge: Cambridge University Press, 288-326.(A detailed study of the Young Hegelian movement drawn from the perspective of cultural history.)

Baumgardt, David (1890-1963)

Baumgardt's early works dealt with the problem of modalities in the philosophies of Kant, Husserl and Meinong and with German philosophical romanticism, especially in the mystic Franz von Baader. Although he never engaged in systematic inquiry into Judaism or Jewish philosophy, he was fascinated by the Jewish religious legacy, and his philosophical reflections on Jewish issues were integral to his philosophical work. A secular Jew, he associated himself with the liberal trends within Judaism. The Jewish philosophers he most highly prized were Maimonides, Spinoza and Mendelssohn. The chief goal of his Jewish studies was to promote those beliefs in Judaism that are of ethical significance and to draw out the import of the moral demands found scattered throughout the ancient Jewish Scriptures. Baumgardt's concern with ethics grew with his increasingly critical stance towards traditional religion. In that vein, he laid great stress on the distinction between knowledge and belief and on that between Jewish rituals and their underlying meaning.

Baumgardt's work did not receive the attention it deserved, because his career at the University of Berlin (1924-1935) came to an abrupt end with Hitler's rise to power. He was a visiting professor at the University of Birmingham, England from 1935 to 1938 and at Pendle Hill, Pennsylvania from 1939 to 1941. From 1941 to 1954 he served as a consultant on philosophy for the Library of Congress. He thus remained a kind of an outsider, publishing relatively little during his lifetime.

Refusing to draw any clear cut line between Jewish religious problems and general philosophical ones, Baumgardt was much attracted by ethical quandaries like that posed in the biblical story of the *Akedah* or binding of Isaac (Genesis 22), where the matter was not of choosing between two rights but between two wrongs: that of disobeying God's command and that of killing one's son. Finding no comfort in the many traditional responses to the ancient story, Baumgardt saw its outcome as a sheer aporia and turned increasingly to a form of ethical hedonism as the foundation of his moral philosophy. Some types of hedonism, he argued, must frame the 'supreme rational command'. But even such a philosophy, he argued, 'leaves to us the risk and responsibility of applying the command properly'.

Baumgardt's main contribution to philosophy was in the field of ethics. In 1933 he published *Der Kampf um den Lebenssinn unter den Vorlaeufern der modernen Ethik (The Struggle for the Meaning of Life among the Precursors of Modern Ethics)*. Because of the rise of the Nazis the work passed almost unnoticed. In it Baumgardt investigated and criticized the ethical ideas of Kant, whose formalism aroused his resentment, of Herder, Hemsterhuis and Jacobi, all of whom he presented as precursors of modern ethics. As a foil to the shortcomings of their ethical conceptions, he called attention to the ethical ideas to be found in the Bible and the Talmud. But these allusions to Jewish ideas, although they reflect Baumgardt's frame of mind and commitment to his Jewish legacy, did not play any important theoretical role in his ethical investigations.

In his unpublished *History of Modern Ethics* Baumgardt essayed a synoptic view of the ethical theories of the past two hundred and fifty years. He was especially drawn to the utilitarianism of Bentham, which he came to know well in England. His book *Bentham and the Ethics of Today* (1952) included hitherto unpublished writings of Bentham, and Bentham's utilitarianism provided the point of departure for his own ethical hedonism. Seeking to rehabilitate a philosophical tradition stigmatized since antiquity as advocating mere pleasure-seeking, Baumgardt drew inspiration from two sharply opposed philosophical orientations: Bentham's empiricist utilitarianism and Kant's metaphysical morals. What he admired in both was the attempt to lay rational and critical foundations for a secular and autonomous ethics. By creating a synthesis between the two, he hoped to deliver his 'critical hedonism' from the pragmatic bent of traditional utilitarianism. This project came to its fullest expression in his posthumous *Jenseits von Machtmoral und Masochismus (Beyond the Morality of Power and Masochism)* (1977), whose subtitle reads *Hedonistische Ethik als kritische Alternative (Hedonist Ethics as a Critical Alternative)*. Despite the work's many interesting ideas, it remains questionable whether it can forge the hoped-for alternative. It does not refer to any works on ethics later than Sartre's writings of the 1940s, and pays almost no attention to analytic philosophy or to Wittgenstein, whose impact on ethical theory cannot be ignored.

Kant and Bentham were contemporaries of one another, and for Baumgardt they remained our contemporaries as well. As regards Kant, Baumgardt was only one among many scholars who investigated his philosophy, but with

regard to Bentham he was one of the most insightful.

In chapters 5 and 6 of *Jenseits von Machtmoral und Masochismus*, Baumgardt asks whether the idea of pleasure can be made a basis for the maximization of happiness and the overcoming of meaningless suffering. This, he argues, requires a 'Copernican revolution' in ethics: while egoistic hedonism is concerned exclusively with the pleasure of the self and is indifferent to the suffering of others, a consistent hedonism will strive for the maximization of happiness. This much is little more than the undergirding of Bentham's move from egoism to the greatest happiness principle. But Baumgardt goes on to argue, in more Kantian vein, that appeals to happiness can be deemed morally valid only if they corroborate the principle of greatest happiness by creating a harmonious coherence among all concerned. The point, however, is not to praise or blame human motives but, rather, to evaluate the outcomes of human action.

Although Baumgardt, like most hedonists, grounded his ethics in naturalism, he did not offer any convincing proof of the chief argument of ethical hedonism, the claim that human pursuit of pleasure entails the ethical value of that pursuit. That is, he offered no convincing argument against the objection that in moving from is to ought one commits a naturalistic fallacy.

Although Baumgardt's ethical work, scattered in his articles and notebooks, and in part published posthumously, was not brought to a systematic conclusion, his rehabilitation of hedonism as a legitimate ethical option is a worthy achievement. The achievement is marred by his lack of a systematic meta-ethics and a weakness in the account of pleasure that he shares with Bentham: individual pleasure is subjective and context-dependent, while pleasure in his more general sense that is most relevant to the greatest happiness principle is linked to interests, ideals, aspirations and goals that can be shared by many. Baumgardt's critical hedonism, a typically reductionist theory, tries to reduce all these to a single measure, Bentham's pleasure/pain principle. And, like earlier hedonist theories, Baumgardt's account has not paid sufficient attention to the cognitive status of deontological assertions. It does, however, represent an important challenge, by a philosophically adept outsider, to modern ethical theory.

ZE'EV LEVY

List of works

Baumgardt, D. (1927) *Franz von Baader und die philosophische Romantik (Franz von Baader and Philosophical Romanticism)*, Halle/Saale: Max Niemeyer. (Unearthing and revitalizing Baader's thought and his critique of Kant.)

Baumgardt, D. (1933) *Der Kampf um den Lebenssinn unter den Vorlaeufern der modernen Ethik (The Struggle for the Meaning of Life among the Precursors of Modern Ethics)*, Leipzig: Felix Meiner.(A critical confrontation of Kant with his contemporaries, pointing out their continuous relevance to modern ethics.)

Baumgardt, D. (1952) *Bentham and the Ethics of Today*, Princeton, NJ: Princeton University Press.(A critical reexamination of Bentham's utilitarianism and first publication of various Bentham manuscripts.)

Baumgardt, D. (1961) *Great Western Mystics: Their Lasting Significance*, New York: Columbia University Press. (Has the same content as *Mystik und Wissenschaft*.)

Baumgardt, D. (before 1963) *History of Modern Ethics*, ed. A. Figen, unpublished manuscript, New York: Leo Baeck Institute.(Synopsis of ethical theories from the time of Kant and his contemporaries until Sartre. Due to the author's death, it is disorganized and chronologically inconsistent.)

Baumgardt, D. (1963) *Mystik und Wissenschaft. Ihr Ort im abendlaendischen Denken (Mysticism and Science: Their Place in Western Thought)*, Witten: Luther.(On the essence of mysticism, in particular mysticism of interiority, nature and history.)

Baumgardt, D. (1977) *Jenseits von Machtmoral und Masochismus. Hedonistische Ethik als kritische Alternative (Beyond Masochism and the Morality of Power: Hedonistic Ethics as a Critical Alternative)*, Meisenheim am Glan: Anton Hain.(Deals with the pluralism of moral, ethical, aesthetic and historio-philosophical positions and the ethics of critical hedonism, as well as a hedonistic critique of the criteria of ethical conduct.)

References and further reading

Frank, J., Minkowski, H. and Sternglass, E.J. (eds) (1963) *Horizons of a Philosopher: Essays in Honor of David Baumgardt*, Leiden: Brill.(Festschrift, including essays on Jewish and ethical issues.)

Levy, Z. (1989) *David Baumgardt and Ethical Hedonism*, Hoboken, NJ: Ktav Publishing House.(An analysis of Baumgardt's Jewish writings and of his ethical theory and a critical inquiry into ethical hedonism.)

Schlodder, H. (1980) *Der kritische Hedonismus David Baumgardts*, Magister-Hausarbeit, Mainz: Johannes-Gutenberg-Universität.(On hedonistic ethics as a means to overcome power morality.)

Baumgarten, Alexander Gottlieb (1714-62)

The German philosopher Baumgarten is known primarily for his introduction of the word 'aesthetics' to describe the affects of art and nature, which in the course of the seventeenth century replaced the older theory of beauty. Baumgarten derived the term from the Greek aisthanomai, *which he equated with the Latin* sentio *(1739: 79). He understood it to designate the outer, external or bodily sense, as opposed to the inner sense of consciousness. Thus aesthetics is the realm of the sensate, of sense perception and sensible objects. Baumgarten understood his usage to be consistent with classical sources, but he was aware also that he was extending logic and science into a new realm. Baumgarten's importance lay in adapting the rationalism of Leibniz for both the study of art and what came to be known after Kant as the aesthetic.*

1 The aesthetic as a 'confused' concept

Baumgarten accepted the fundamental rationalist, epistemological division between what is clearly and distinctly known, on the one hand, according to concepts and reason and what is known, on the other, by sense. The latter apprehension is 'confused' in a sense that can be traced back to medieval discussions of the problem of universals. According to medieval realism and conceptualism, only the universal deserves the definite article and provides clear, distinct apprehension. So 'the rose' refers not to some existing rose but to the universal type. To refer to a specific rose, one must individuate it either as 'this rose' when dealing with an existential application (cf. *Metaphysica*, para. 151) or as 'a rose' when dealing with an individual which may or may not actually exist. Both the individuation and its indefinite extension are dependent on sense in a way that the universal is not. Both can be said to 'confusedly' represent, therefore, because their referents are present only in limited perceptible and temporal forms, all of which lack some essential distinctness. In this sense, a confused presentation is not faulty or unclear; it is simply lacking in the kind of clarity which only the universal can provide. Its sensate forms limit it. Baumgarten adopted this distinction in describing aesthetic discourse. The aesthetic is sensate and confused or indistinct in this specialized sense.

It does not follow, however, that the aesthetic cannot be clear within its own cognitive sphere. Obscurity and clarity are matters of content. Something is clearer if a higher number of its sensate representations are available. So if one sees an object from all sides, for example, one sees it clearly. If one sees it only from one perspective at a distance, it is seen obscurely. The aesthetic permits an extensively clear form of representation which is determinate and individual. From the standpoint of promoting art, therefore, greater extensive clarity and determinateness are positive features, which enhance the aesthetic effect.

Baumgarten took discourse as the primary form of the aesthetic. Discourse can be construed broadly here, since Baumgarten acceded to Horace's equation of poetry and painting (1735: 52). Representations themselves include both images derived from sense and direct sense-perception. They are combined into thematic wholes on the basis of the connection of the images and resemblance between images. Such wholes are simple or complex, but since complexity is understood as having many themes, not parts, simplicity is an aesthetic virtue. Perceptions and images (which are secondary perceptions) are confused representations because they are not abstract, intelligible forms, which alone would be conceptually distinct. But while they lack intensional clarity, they may have great extensional clarity, and may form thematic wholes based on resemblance, the connection of the images, and ordering of the sensate elements. Baumgarten applied this scheme to poetry and poetic language in his *Meditationes philosophicae de nonnullis ad poema pertinentibus* (1735, translated as *Reflections on Poetry*), and he suggested ways in which it would also apply to painting, sculpture and music. His goal was a science of the perceptual realm that would parallel the more precise sciences of metaphysics and logic.

2 Aesthetic implications

A number of points in Baumgarten's aesthetics bear on later developments in aesthetics. He focused on the affective side of perception. Sensate representations are 'marked degrees of pleasure or pain' (1735: 47). Stronger impressions are more poetic because their impressions are extensively clearer (1735: 27). On the other hand, Baumgarten distinguished aesthetic effects quantitatively, not qualitatively. It is not that feelings of pleasure are intrinsically valuable, but that more effects contribute to a greater perfection of the discourse. The closest that Baumgarten came to a qualitative aesthetic is in his discussion of 'the wonderful', which he characterized as an

intuitive grasp of the inconceivable that is not present in perception (1735: 53). He was obviously trying to account for the sense of awe or wonder that is one of the classical marks of the sublime, yet he still explained it in essentially quantitative terms. The wonderful is an intuition of something that is not present. Hence it is an added element in the aesthetic. It depends on a mixture of the unfamiliar and the familiar, and since it cannot be traced to sense impressions, gives a scope to the imagination it would not otherwise have. However, it provides little justification for promoting feeling directly to prominence. Baumgarten is at the opposite extreme from Friedrich Schleiermacher in his regard for intuitive feeling.

Baumgarten also gave scope to the imagination in a way that reveals a contrast with later exaltations of imagination as the aesthetic faculty. The imagination is productive of fictions. But fictions are categorized in terms of possible worlds. True fictions are possible in this world. The imagination may go beyond what is actually perceived to what may very well be the case. When a painter poses figures to construct a tableau, his painting is a fiction, but a possible representation of an actual event. There could be such an event, and for all we know, this is how it may have looked. Fictions *per se*, however, are impossible in this world. They may be divided into those which are impossible in all worlds, called utopian fictions, and those which are impossible in this world but possible in some world, called heterocosmic fictions. There are no centaurs, and such combinations are biologically impossible. But they are nevertheless possible in some world, whereas a round square is not possible at all. Since utopian fictions cannot be represented, they cannot be aesthetically significant. But heterocosmic fictions can be represented, and in fact art depends on such fictions for its greater extensive clarity and determinateness. Baumgarten arrived at an implicit defence of fictions quite different from the earlier Aristotelian and subsequent Romantic defences, which are based on the positive emotional effects of the imagination. For Baumgarten, fictions and the imagination are needed because they extend the aesthetic beyond the reach of mere sense, without converting it into rational concepts.

Baumgarten thus remained firmly within the rationalist camp. The aesthetic is clearly a lower faculty: ‘Therefore *things known* are to be known by the superior faculty as the object of logic; *things perceived* [are to be known by the inferior faculty as the object] of the science of perception, or aesthetic’ (1735: 78). But Baumgarten expanded the scope of knowledge in interesting ways. He expected the aesthetic to be a science with its own logic, and in the final, fragmentary ‘Kollegium über die Ästhetik’, he acknowledged that sensate knowledge is the foundation of clarity and that the aesthetic must come to the aid of logic. The aesthetic is limited by its sensate representations and the imagination must be restricted in order to avoid licence, but within these limits the aesthetic is a legitimate source of a kind of knowledge. The standard remains rational distinctness, but while it is possible to mix distinct, conceptual elements with confused, perceptual elements in complex representations, the aesthetic is independent of the distinct conceptual forms. When conceptual forms dominate, one loses the aesthetic effects. The aesthetic may still be a lower form, but it is legitimate within its own realm, and Baumgarten came close to acknowledging that the aesthetic could be a clearer and more effective form of knowledge than a priori rational concepts for all but the most accomplished metaphysicians.

3 Tensions within Baumgarten’s system

The value that Baumgarten assigned to the aesthetic produced some tensions within his system. His ideal was knowledge, and the goal of knowledge is to build a science. So Baumgarten set out to create a science of the sensate. But he also recognized that the rational, conceptual distinctness which he took to characterize knowledge is antithetical to the confused, indistinct, sensible impressions and images on which the aesthetic depends. So the more scientific aesthetics becomes, the more it is in danger of ceasing to be aesthetic. The incompatibility of the rational and aesthetic realms (1735: 42) exposes the inconsistency between Baumgarten’s goal of pure knowledge and his defence of sensate discourse. The classical way to resolve this inconsistency was to leave the aesthetic behind as one progressed up the ladder of consciousness. Baumgarten continued to adhere to that classical model in which aesthetics is the product of a lower faculty and is limited to a helping role, but he also separated and gave independence to sensate representations, which suggests that there are simply two different forms of knowledge.

Two themes in Baumgarten’s work illustrate how the tension between reason and sensibility arose and how he tried, unsuccessfully, to resolve them. First, Baumgarten’s appeals to sensation are essentially quantitative. Consider an instance of sensate discourse. It will be more or less aesthetic depending on its extensive clarity, determinateness and specificity. All of these characteristics are consistent with, and in fact require, confused or

indistinct representations, since the discourse would cease to be sensate if it became distinct. The standard of judgment is aesthetic perfection. The aim of the aesthetic is the perfection of sensitive cognition as such; that is also what is meant by beauty (1750: 14). Baumgarten was led to this quantitative standard because it was consistent with his rational epistemology, which was based on completeness. (Quantity in this sense is not inconsistent with simplicity as long as the quantitative impressions are united by a single theme or object.) But while his aesthetics is based on sensibility, it is not based on feeling. So while Baumgarten recognized the close connection between sensate impressions and pleasure or pain, he had no way of incorporating those qualitative impressions into his epistemological scheme.

A second theme in Baumgarten's aesthetics is the requirement for thematic unity and order. The basis for order is the connection of impressions and images (1735: 62ff). The operative principle of connection is sufficient reason, and connections are themselves part of the observable order of indistinct representations. So despite his defence of the sensible realm, Baumgarten was still the disciple of Leibniz in holding that what gave order to sense data was a principle of reason which on Baumgarten's system was external to the data itself. He could forge a unity only by appealing to connections recognized by the mind but present between the sensate impressions. In spite of his desire to found a new science of aesthetics, Baumgarten could only appeal to a priori principles and try to find them reflected in his new aesthetic realm. He thus failed to resolve the tension between his rationalist epistemology and the new realm of science he had defined.

See also: Sublime, the

DABNEY TOWNSEND

List of works

Baumgarten, A.G. (1735) *Meditationes philosophicae de nonnullis ad poema pertinentibus*, trans. K. Aschenbrenner and W.B. Holther as *Reflections on Poetry*, Berkeley, CA: University of California Press, 1954. (English translation of Baumgarten's dissertation with a useful introduction.)

(1739, 1758) *Texte zur Grundlegung der Ästhetik*, trans. H.R. Schweizer, Hamburg: Felix Meiner Verlag, 1983. (Four key aesthetic texts, including a Latin text with parallel German translation of key sections of Baumgarten's *Metaphysica* of 1739.)

(1750, 1758) *Theoretische Ästhetik*, trans. H.R. Schweizer, Hamburg: Felix Meiner Verlag, 1983.(Latin text with parallel German translation of selected sections of Baumgarten's *Aesthetica*.)

References and further reading

Croce, B. (1909) *Aesthetic*, trans. D. Ainslie, New York: Farrar, Straus && Giroux, 1968, 212-19.(An essentially negative evaluation of Baumgarten's contribution to the history of aesthetics.)

Cassirer, E. (1951) *The Philosophy of the Enlightenment*, trans. F.C.A. Koelln and J.P. Pettegrove, Princeton, NJ: Princeton University Press.(In contrast to Croce, Cassirer credits Baumgarten with being the founder of modern aesthetics.)

Gregor, M.J. (1983) 'Baumgarten's *Aesthetica*', *Review of Metaphysics* 37 (2): 357- 85.(Argues that while there is a change in perspective, Baumgarten's definitions of the aesthetic remain consistently rationalist.)

Wessell, L.P., Jr (1972) 'Alexander Baumgarten's Contribution to the Development of Aesthetics', *Journal of Aesthetics and Art Criticism* 30 (3): 333-42.(Contends that Baumgarten did free aesthetics from rationalism, but only at the cost of consistency.)

Bayle, Pierre (1647-1706)

Bayle was one of the most profound sceptical thinkers of all time. He was also a champion of religious toleration, and an important moral philosopher. The fundamental aim of his scepticism was to curb the pretensions of reason in order to make room for faith. Human reason, he believed, suffers from two fundamental weaknesses: it has a limited capacity to motivate our actions, and it is more a negative than a positive faculty, better at uncovering the defects of various philosophical positions than at justifying any one of them. This conception of reason led Bayle to see, with an uncommon clarity, that the nature of the sceptic's arguments must be to proceed by internal demolition, showing how claims to knowledge undermine themselves in their own terms.

Bayle's moral thought is to be found essentially in his critique of attempts (such as that of Malebranche) to show how God, all-powerful and good, could have created a world in which there is evil. Such theodicies, he argued, rely on unacceptable models of moral rationality. Bayle's arguments reveal a view of moral reasoning that is of considerable interest in its own right. Like Malebranche (and contrary to Leibniz, who attacked Bayle's critique of theodicy), he believed that there are duties superior to that of bringing about the most good overall. But unlike Malebranche, Bayle saw these duties as lying not in what the rational agent owes himself but in what he owes to the inviolable individuality of others. This outlook had its psychological roots, no doubt, in Bayle's own experience as a Huguenot victim of religious persecution.

1 Scepticism

Pierre Bayle was one of the most important sceptical thinkers of all time, as well as a notable moral philosopher and advocate of religious toleration. The fundamental motivation of his scepticism was religious: his aim was to curb the pretensions of reason in order to make room for faith. Born into a Calvinist family in Carla in southern France, he became a professor of philosophy in the Protestant academy at Sedan. After its abolition (1681) and the revocation of the Edict of Nantes (1685), which ended Protestant toleration in France, Bayle fled to Holland and spent the rest of his life in Rotterdam. His Calvinist conviction that God is rationally inscrutable spurred his wide-ranging attack on the power of reason to shape our conduct and to make sense of the world. As he wrote in the article on 'Paulicians' in his most important work, the *Dictionnaire historique et critique (Historical and Critical Dictionary)* (1696: Paulicians, note E), 'The ways of God are not our ways… [We must accept] the elevation of faith and the abasement of reason'.

Bayle believed that human reason suffers from two fundamental weaknesses. The first is that reason is quite limited in its capacity to motivate our actions. Human beings act more often in virtue of their dominant passions than on the basis of their professed principles. In his first important work, *Pensées diverses sur la comète (Miscellaneous reflections on the comet)* (1682), Bayle made use of this observation to argue that, contrary to the accepted opinion of his time, atheists would be able to live together peacefully in society. Were they to follow through on all the consequences of God's non-existence, they would indeed plunge into a life of vice without remorse. For though convinced that we can know the principles of morality without relying upon belief in God, Bayle thought that without such a belief we would have no reason to subordinate our self-interest to them. Still, the atheist would be unlikely to take up the life of crime, Bayle insisted, because the rational calculation of advantage is a less powerful motive than the concern for honour. The vanity of wishing to be well regarded by others, combined with an interest in using moral principles to blame or approve their actions (if not one's own), would steer the atheist toward mutual cooperation with others.

For Bayle, this argument was not merely a philosophical curiosity. It formed part of his continuing campaign in favour of a society based on religious toleration. The expression of heterodox opinion need not by itself imperil social order. The motivational weakness of reason played a further role in the central argument of Bayle's principal treatise on religious toleration, *Commentaire philosophique sur ces paroles de Jésus-Christ, 'Contrains-les d'entrer' (Philosophical Commentary on the Following Words of Jesus Christ, 'Compel Them to Enter')* (1686). Just as our actions stem less from our reason than from our passions and feelings, so belief itself, Bayle argues here, is not directly under our rational control, but arises involuntarily from inner conviction. As a result, the use of force to impose religious belief must be a futile enterprise. Individual conscience ought to be respected, since sincere belief can have no other source.

The second fundamental weakness of human reason for Bayle was that it is more a negative than a positive faculty. It lends itself better to the refutation of opposing views than to the justification of one's own position. As he wrote in *Réponse aux questions d'un provincial (Reply to the Questions of a Provincial)* (1703-7: II.137), reason 'is better able to demolish than to build, it knows better what things are not than what they are'. When reason is instead put to the use of defending some particular position, it tends naturally to undermine itself. Reason is essentially destructive because it excels in uncovering the self-contradiction into which its positive employment inevitably falls.

This diagnosis underlies two other significant features of Bayle's thought. It allowed him, first of all, to see clearly how sceptical argumentation must proceed in general. The sceptic may not appeal to principles that are not admitted by the position under attack, since such an argument would be irrelevant to the partisans of that position and contrary to the sceptic's own professed lack of knowledge. Instead, Bayle observed, the sceptic must show how the position undermines itself, involving views which are mutually inconsistent or conflicting with opinions every reasonable person holds. The sceptic must work by internal demolition, attacking his adversaries 'on their very own dungheap', as he wrote in the preface to his *Pensées diverses* (see also the 'Second Clarification' in 'Spinoza', *Dictionnaire historique et critique*).

The idea that in its positive employment reason is ultimately self-destructive also shaped Bayle's attitude towards the fundamental conflicts between reason and faith which he doggedly uncovered (see §3). 'If reason were in agreement with itself', he wrote (1703-7: II.137), 'we should be more worried that it agrees so poorly with some of our articles of religion'. Many of Bayle's Enlightenment readers (such as Diderot, in the article 'Pyrrhonian or Sceptic' of the *Encyclopédie*) claimed him for one of their own. But they were wrong to suppose that his attack on rational theology was aimed at the rejection of religious dogma. On the contrary, Bayle's scepticism took the form of a fideism, intended to confirm his Calvinist belief in the radical disparity between God's ways and our own. Indeed, so far was Bayle from being a founding father of the Enlightenment that he was one of the first to express concern about what in the twentieth century has come to be called 'the dialectic of the Enlightenment'. If we follow reason alone, proportioning all our beliefs to the available evidence, we will end up by doing away, he once wrote, not only with superstition and barbarism, but eventually with every sort of conviction: 'Man's fate is so bad that the knowledge that delivers him from one evil throws him into another' (1696: 'Takiddin' Note A).

2 Epistemology and metaphysics

In arguing against the capacity of human reason to acquire knowledge of the world, Bayle focused chiefly on foundational matters. We cannot demonstrate the real existence of the external world, he maintained, nor grasp its fundamental principles. But he appears to have had no serious doubt that straightforwardly empirical questions admit of rational solutions. Indeed, Bayle frequently urged, against the authority of Descartes, that philosophy should recognize the validity and importance of historical knowledge, and his own scholarship in the history of philosophy and theology was significant. Bayle's scepticism was therefore restricted to speculative questions of principle. In his *Dictionnaire* article on Pyrrho, he described Pyrrhonism simply as the view that we have no knowledge of the underlying nature of things (see Pyrrhonism).

Such an outlook would seem to place Bayle within the important seventeenth-century current of thought generally called 'mitigated scepticism', the most important exemplars of which were Gassendi and Locke. Like them he believed that, unlike our beliefs about the observable features of the world, ultimate physical explanations can be at most probable, never certain (1696: 'Pyrrho' Note B), and that such explanations refer to what we cannot fully understand. For example, he endorsed Locke's view that we know too little about the nature of matter to rule out the possibility that God might have 'superadded' to it the power of thought (1696: 'Dicaearchus' Note M) (see Locke, J. §5).

None the less, Bayle's scepticism ran deeper than that of Gassendi and Locke, for he also believed that ultimate physical principles turn out, upon reflection, to be self-contradictory. This was, for instance, the verdict of his famous discussion of space in the *Dictionnaire* article on Zeno of Elea (1696: 'Zeno' Note G). The idea of space, he argued, is incoherent on any of the interpretations one might give of it. Space cannot consist ultimately of mathematical points since the addition of extensionless entities to one another cannot produce extension. Nor can it consist of extended but indivisible physical points, since anything extended is divisible. Nor can it be infinitely divisible, since this would preclude the immediate contiguity of its parts, or would permit the interpenetration of

any two contiguous bodies, and in any case would succumb to the well-known paradoxes of the infinite. Here we find a perfect specimen of Bayle's theme that reason tends inevitably to undermine itself. His discussion of space had an important influence on Hume's *Treatise of Human Nature* (1739-41).

3 Theodicy and ethics: Bayle versus Malebranche

Bayle's moral thought is best approached by way of his critique of philosophical *theodicy*. That critique is important, if for no other reason than the fact that Leibniz took it as his main target in his 1710 *Essais de théodicée (Theodicy)*. But it also embodies a view of morality that is of genuine interest in its own right.

A theodicy aims to justify the ways of God to humans, showing how God, all-powerful and good, could have created a world in which there is evil. It must therefore make use of a model of moral rationality that lays out the principles according to which one should choose among actions, where each action may involve bad consequences. In this light, theodicies retain a philosophical value even for those who today may find very foreign the idea of justifying the ways of God. Theodicies give vivid expression to models of moral rationality. Accordingly, in seeing why Bayle thought the two leading theodicies of his day were morally unacceptable, we bring into focus his own very interesting conception of moral rationality.

The first of these theodicies was presented by Malebranche in his 1680 *Traité de la nature et de la grâce (Treatise on Nature and Grace)*. Malebranche's guiding principle was that God, in creating the world, had to show his wisdom and goodness not only in the result he achieved but also in the means he used. Consequently, God could act to maximize the resulting good in the world only to the extent that at the same time he employed a 'simplicity' of means - that is, a system of universal and immutable laws. As he could foresee, these laws would sometimes result in evil that could have been avoided, had he set about single-mindedly to bring about the greatest good possible. God knew, for example, of the evil that men would do because of the freedom he had given them, when they made use of the laws he set up. This evil, though foreseeable, was not intended, however, since God aimed only to respect his own nature as wise and good. The model of moral rationality that Malebranche assigned to God is therefore one we would call 'deontological': right conduct is that which respects certain principles independently of the bad consequences that the action can foreseeably produce (see Malebranche, N. §6).

In his first writings, Bayle presented himself as an adherent of Malebranche's theodicy. Thus, in the *Pensées diverses*, he sought to show that sensational phenomena such as comets and monsters, and even sin, represent neither a failure of providence nor God's particular will, but are instead the unintended result of the simple laws by which God rules the world (1682: §§208, 230, 231, 234). It is important to note that, like Malebranche, Bayle did not draw this model of moral rationality from revelation or scripture. We must judge God's rationality, he wrote, by reference to our own (1682: §223). He was convinced that the first principles of morals are open to our reason even 'without knowledge of God' (1682: §178).

This 'autonomy of morality' with respect to religion forms one of the constants of Bayle's thought. It is the principal basis on which later (beginning with the *Dictionnaire* in 1696) he rejected Malebranche's theodicy. He had come to believe that Malebranche's underlying model of moral rationality is incompatible with what we can see to be the first principles of morality. It is in precisely these terms that, in one of his last writings, he explained his change of mind (1703-7: II.91, 155). Bayle's fideism was therefore not the result of any wholesale scepticism about the possibility of knowledge in general. Indeed, as noted in §2 above, Bayle regularly limited his scepticism to speculative questions - he did not question our ability to acquire knowledge of more straightforward matters. And so, in the case of theodicy, he denied that we can ever justify the ways of God because, continuing to consider Malebranche's theory the best conception possible, he found that it conflicted with moral truths we know already.

What exactly did Bayle find morally unacceptable in Malebranche's theodicy? To Bayle it was evident that 'those who permit an evil which it is easy for them to prevent are culpable' and that 'those who let perish a person that they could easily save are guilty of his death'. How, then, could God have let the whole human race fall into crime and misery? He could easily have prevented the disobedience of Adam and Eve or at least have stopped the consequences of their fall. Malebranche's own response, of course, was that God let these things happen because the pursuit of the good must yield to respect for certain general principles. Bayle was not opposed to the idea that there exist duties superior to maximizing the good. But he insisted that such duties consist in securing an urgent good for the agent or someone other, or in saving them from a disastrous evil. These are not, however,

considerations which could have moved God in the present case. No other parties but himself and humanity were involved and since, being perfect, he himself had need of nothing, nothing could therefore weigh more heavily than his benevolence toward humanity, his creation (1703-7: II.91, 150, 155).

It could be said that underlying this difference between Malebranche and Bayle were not only two models of moral rationality but also two opposing conceptions of love. On the one hand, there is the idea of love as merited, according to which if one is perfect, one must love oneself above all; on the other hand, there is the idea of love as a gift, which implies that the more perfect one is, the fewer obstacles there must be to the love one bears others. Thus, just as Malebranche wrote that ‘one cannot love anything except in proportion to how much one believes it lovable’ (Malebranche [1715] 1974 XVI: 96), so Bayle affirmed that ‘nothing fits better with true grandeur and supreme perfection than to put one’s power and knowledge in the service of others’ happiness’ ([1707] 1964-8 IV: 63).

In response to Bayle’s objection one could imagine Malebranche, like many deontologists, invoking the classic principle of ‘double effect’, according to which actions are to be judged morally by their intended effects, not by all their foreseeable effects. Is it not true that Adam and Eve were themselves free to prevent the evil, that the responsibility for it was their own, and that it was a paramount duty to respect their freedom of action, even at the price of not doing what one knew would produce the greatest good? Indeed, in a letter of 11 December 1706 to the Père André, Malebranche formally embraced the principle of double effect. But Bayle anticipated this reply, claiming that it is an evident principle that one ought not to give someone a good (such as freedom of action) that one knows the person will abuse (1703-7: II.81).

In the *Dictionnaire* Bayle presents a remarkable example to illustrate this point (1696: ‘Paulicians’ Remark E). A mother who sent her daughters to a ball knowing that they would succumb to temptation, but being content simply to encourage them to be virtuous and to threaten to disown them if they did not return home virgins, would not count, he observed, as loving either her daughters or chastity. Bayle did recognize duties of ‘strict obligation’, involving respect for the freedom of others which ordinarily take precedence over the maximization of the good. But he believed that such duties must be suspended in emergency cases, where a catastrophic evil is to be avoided (1696: ‘Paulicians’ Remark M). There, we must choose the lesser evil (See Malebranche, N. §§4-5).

4 Bayle and Leibniz

Bayle’s critique of Malebranche might suggest that he held a ‘consequentialist’ conception of moral rationality, according to which we should always seek the greatest net good overall, even at the price of using means that in themselves are bad. This conception underlies, in fact, the theodicy which Leibniz laid out in the *Essais de théodicée*. There he argued that God reasons according to the ‘règle du meilleur’ (‘the rule of the better’), so that what is evil in itself should be chosen if it forms the indispensable means to bringing about the greatest good overall (Leibniz 1710: §§22, 209). The general and immutable laws according to which God created the world constitute not limits (as Malebranche believed) but, rather, means to the maximization of the good.

Leibniz directed the *Essais* primarily against Bayle himself and his claim that theodicy is impossible. In his view, Bayle had sold theodicy short by failing to consider a version which makes rigorous use of ‘la règle du meilleur’. Leibniz’s position seems to look all the stronger if Bayle’s arguments against Malebranche are themselves consequentialist in spirit. However, the argumentative situation is more complex. Bayle’s conception of morality was not really consequentialist. And though Bayle died four years before the publication of Leibniz’s book, he was familiar with a theodicy of that form and rejected it as morally unacceptable (see Leibniz, G.W. §3).

The consequentialist theodicy with which Bayle was acquainted was one outlined by Isaac Jaquelot in his 1705 book, *La conformité de la raison et de la foi (The Conformity of Reason and Faith)*. Jaquelot himself rather closely followed the Malebranchian position. But he claimed that Bayle’s objections to Malebranche relied on the principle that one ought always to do that action which will bring about the most good overall. Suppose, he wrote (treating the question of theodicy once again through the story about the mother and her daughters), the mother was sending her daughters to the ball so that, in virtue of all the foreseeable consequences of that decision, a ‘great, noble, and comprehensive plan’ for the reformation of the kingdom would be realized. If Bayle required that the welfare of the daughters should count for more than the duty of respecting their freedom, he ought also to admit, so Jaquelot argued, that the still greater good of accomplishing that plan should outweigh their welfare.

They ought to be sent to the ball. God's decision not to prevent Adam and Eve's disobedience would thus be justified on the same basis - namely, by the greater good which his overall plan of eventual redemption would bring about.

Bayle took up Jaquelot's challenge in his *Réponse aux questions d'un provincial* (1703-7: II.153). Such a mother, he replied, would be even more wicked than the one first imagined. Agreeing that, taken absolutely, the good of the kingdom is greater than the good of the daughters, Bayle maintained that there are some things we should never do to another person, even at the cost of foregoing a tremendous good. This rejoinder to Jaquelot was, no doubt, the objection Bayle would have made to Leibniz had he lived to see Leibniz's *Essais de théodicée*. Once again we find him rejecting a theodicy on the basis of moral principles which he claims we know to be true.

An important question, however, is whether Bayle's moral thought, taken as a whole, is coherent. How can we harmonize his criticism of Malebranche, according to which the prevention of a significant evil must weigh more than the duties of strict obligation, with his objection to Jaquelot (and by anticipation to Leibniz) that the maximization of the good must yield to certain prohibitions on the treatment of others? How, we might say, can Bayle apparently reject both deontological and consequentialist conceptions of moral reasoning? In fact, Bayle was perfectly consistent. In his rejection of Malebranche, the operative principle was that we may suspend a strict duty with regard to an individual if we wish to avoid a very great evil to *this same individual*. Such a principle is clearly compatible with the principle Bayle invoked in his response to Jaquelot, namely that we must never cause a great evil to an individual in order to procure a great good *for others*.

Bayle was no doubt a deontological thinker, since he believed that there are duties superior to that of bringing about the most good overall. The fundamental coherence of Bayle's thought emerges once we recognize that a deontological ethic can take two different forms. Malebranche's version limits the maximization of the good by the demands of what a rational agent owes himself. Bayle's version limits it by the demands of what one owes to the inviolable individuality of others. The core conviction of Bayle's ethics was the refusal to sacrifice individuals to a greater whole. No doubt, it had its psychological roots in his own experience as a Huguenot refugee from religious persecution. Bayle's skill in giving it philosophical articulation shows him to be, despite his scepticism, one of the great moral thinkers.

See also: Spinoza, B. de

CHARLES LARMORE

List of works

Bayle, P. (1682) *Pensées diverses sur la comète (Miscellaneous Reflections on the Comet),* critical ed. A. Prat and P. Rétat, Paris: Nizet, 1984.(Best presentation of Bayle's moral psychology.)

Bayle, P. (1686) *Commentaire philosophique sur ces paroles de Jésus-Christ, 'Contrains-les d'entrer' (Philosophical Commentary on the following words of Jesus Christ, 'Compel them to enter'),* in *Oeuvres Diverses*, Hildesheim: Olms, 1964-8.(Bayle's defence of religious toleration.)

Bayle, P. (1696) *Dictionnaire historique et critique*; selective trans. R. Popkin and C. Brush, *Historical and Critical Dictionnary*, Indianapolis, IN: Bobbs-Merrill, 1965.(Bayle's most important philosophical work.)

Bayle, P. (1703-7) *Réponse aux questions d'un provincial (Reply to the Questions of a Provincial)*, in *Oeuvres Diverses*, Hildesheim: Olms, 1964-8.(Contains important clarifications of many of Bayle's views.)

Bayle, P. (1707) *Entretiens de Maxime et de Thémiste (Conversations between Maximus and Themistus)*, in *Oeuvres Diverses*, Hildesheim: Olms, 1964-8.(Devoted to the issue of theodicy.)

Bayle, P. (1727) *Oeuvres diverses*, The Hague, 4 vols; repr. Hildesheim: Olms, 1964-8.(Contains, with the exception of the *Dictionnaire*, most of Bayle's writings.)

References and further reading

Brush, C. (1966) *Montaigne and Bayle. Variations on the Theme of Scepticism*, The Hague: Martinus Nijhoff. (Very useful study of Bayle's scepticism.)

Delvolvé, J. (1906) *Essai sur Pierre Bayle. Religion critique et philosophie positive (Essay on Pierre Bayle. Critical religion and positive philosophy.)* Paris: Félix Alcan.(Comprehensive philosophical study which underestimates, however, the importance of Bayle's religious faith.)

Hume, D. (1739-41) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press, 1975.(Hume's classic. Bayle's influence on his treatment of space is evident in Book I, section 2.)

Jaquelot, I. (1705) *La conformité de la raison et de la foi (The Conformity of Reason and Faith)*, Amsterdam: Desbordes.(Bayle's example of a consequentialist theodicy.)

Labrousse, E. (1983) *Bayle*, Oxford: Oxford University Press.(The best short introduction by the leading contemporary Bayle scholar; includes bibliography.)

Larmore, C. (1993) *'Théodicée et rationalité morale' (Theodicy and Moral Rationality)*, in *Modernité et morale*, Paris: Presses Universitaires de France, 121-38.(Extended treatment of material in §3 and 4 above.)

Leibniz, G.W. (1710) *Essais de théodicée*; trans. E.M. Huggard as *Theodicy*, La Salle, IL: Open Court, 1985. (Referred to in §4; particularly in its introductory 'Discourse', it contains an important critique of Bayle.)

Malebranche, N. (1680) *Traité de la nature et de la grâce (Treatise on Nature and Grace)*, in *Oeuvres complètes* vol. 5, Paris: Vrin, 1976.(Referred to in §3; presents the version of theodicy which Bayle considered the strongest possible.)

Malebranche, N.(1715) *Réflexions sur la prémotion physique (Reflections on physical premotion)*, in *Oeuvres complètes*, vol. 16, Paris: Vrin, 1974.(Referred to in §3.)

Beattie, James (1735-1803)

James Beattie was famed as a moralist and poet in the late eighteenth century, and helped to popularize Scottish common-sense philosophy. At Marischal College, Aberdeen, Beattie cultivated a lecturing style which differed significantly from that of his Aberdonian predecessors. Because he believed that the form of abstract analysis characteristic of the science of the mind in his day often led students into the morass of Humean scepticism, Beattie endeavoured to inculcate sound moral and religious principles through the study of ancient and modern literature. Consequently his version of common-sense philosophy diverged from that developed by Thomas Reid. Beattie was more of a practical moralist than an anatomist of the mind, and his treatment of common-sense epistemology lacked the philosophical range and rigour of Reid's.

1 Life

Equally renowned in his day as a poet, the moralist James Beattie was born on 25 October 1735 in Laurencekirk, Scotland. He attended Marischal College, Aberdeen, from 1749 to 1753, when he took his M.A. Beattie shone as a student, and was a favourite of the principal of the college, the noted classicist Thomas Blackwell. After graduating, Beattie became a parish schoolmaster in the village of Fordoun near his home, and he also began to work towards a divinity degree. While at Fordoun, Beattie published occasional poems in the periodical press, and he later consolidated his literary reputation with *The Minstrel* (1771-4). Thanks to the influence of Blackwell and his colleagues, Beattie was appointed to teach at the Aberdeen Grammar School in 1758, and succeeded Alexander Gerard to the Chair of Moral Philosophy and Logic at Marischal in 1760.

Once installed as professor, Beattie became a prominent figure in the Aberdeen Philosophical Society (to which he was admitted on 10 February 1761) and the Aberdeen Musical Society. Because of his poetical works Beattie was also increasingly known outside of Aberdeen, and became the darling of the London literati in the early 1770s when he published *The Minstrel* along with his widely read attack on David Hume, *An Essay on the Nature and Immutability of Truth* (1770). For his part, Hume thought the *Essay* a splenetic performance, and his sentiments were echoed by Beattie's most combative critic, Joseph Priestley. However, the work was applauded by George III and, during a highly successful visit to London, Beattie was rewarded with a royal pension and given an honorary Doctor of Laws degree by Oxford University on 9 July 1773. Among the many friends he made on this trip was the painter Sir Joshua Reynolds, and while he was in the metropolis he sat for the allegorical portrait 'The Triumph of Truth', in which Beattie is juxtaposed with the figure of truth victorious over her opponents. On the Continent Beattie was similarly applauded as a champion of orthodoxy. In 1774 he was elected to the Zealand Society of Arts and Sciences in Holland, and the enthusiastic reception of the *Essay* prompted translations into Dutch, French, German and Italian.

The early 1770s can be seen as marking the high point of Beattie's career. Yet the promise of this period was never completely fulfilled, largely because his already fragile health declined markedly and his wife gradually became mentally unstable. Capitalizing on the popularity of the *Essay*, in 1773 he solicited subscriptions for a new edition which would include additional essays on other topics, and, after delays caused by illness, he saw this enlarged version through the press in 1776. Eventually his deterioriating condition and that of his wife forced him to abandon any hope of completing a systematic work on moral philosophy. He turned instead to revising segments of his lectures in the winter of 1780-1, and these eventually appeared in 1783 as his *Dissertations Moral and Critical*. The following year Beattie found some consolation in further public recognition, for he was appointed a fellow of the newly founded Royal Society of Edinburgh and elected an honorary member of both the Manchester Literary and Philosophical Society and the American Philosophical Society.

Beattie finally realized his long-standing ambition to publish an accessible defence of Christianity in 1786 when he completed his *Evidences of the Christian Religion*, modelled on an earlier apologetic work by Joseph Addison, whom he greatly admired. Like many of his previous writings the *Evidences* had its origins in the classroom, and he recycled his lectures one last time as the basis for *The Elements of Moral Science* (1790-3). The subsequent popularity of the *Elements* as a textbook attests to Beattie's standing as a pedagogue, but his ill health had somewhat compromised his teaching career. One of his major preoccupations during the 1780s was to find a suitable successor, and in 1787 his Marischal colleagues agreed to nominate his son, James Hay Beattie, as joint

professor. However, his son died tragically in 1790, and the issue of succession was not settled until 1796, when George Glennie's appointment as his assistant enabled him to retire from lecturing.

2 Professing moral philosophy

Prior to Beattie's election to the chair, the teaching of moral philosophy at Marischal College can be seen as a volatile mixture of moral exhortation with the rigorous study of the intellectual and active powers of the mind. George Turnbull was the first pedagogue at Marischal to argue that the lessons of practical morality must be rooted in the empirical investigation of human nature, and his call for a methodological reform in the science of morals was taken up at the college by David Fordyce and Alexander Gerard. Beattie, however, broke with this tradition. When he began lecturing in 1760, his course was virtually the same as that given by his teacher Gerard, but during the next two decades Beattie reoriented the moral philosophy curriculum. Whereas Gerard surveyed the various faculties of the mind in a comprehensive manner, over the years Beattie reduced the number of classes spent on pneumatology to make room for lectures on literary criticism, as well as greatly expanded treatments of language, rhetoric, composition and the writings of Cicero. Along with this literary turn went an increased interest in dreaming, taste and the workings of the imagination and memory; Beattie evidently valued his lectures on these topics in so far as he revised them for inclusion in his *Dissertations Moral and Critical* (1783). Beattie also enlarged the time spent on natural theology and, in the 1770s, his discussion of the rational grounds for Christianity took on a new polemical tone due to his growing obsession with the supposed threat posed by the fashionable scepticism of Hume.

Moreover, Beattie was more emphatic about the prerogatives of practical moralizing than Gerard had been. Whereas Gerard had espoused the view that the pursuit of virtue had to be based on a knowledge of the first principles of morality, Beattie virtually severed the connection between theory and practice by arguing that matters of conduct took priority over metaphysics (by which he meant the abstract study of the mind). Consequently, he eschewed any detailed discussion of the substantive issues debated in the science of the mind as being of little concrete application, and focused instead on the inculcation of sound moral, religious and political attitudes. Indeed, he explicitly warned his students against being seduced by purely speculative metaphysical problems, and urged them to attend to the fulfilment of their various practical duties. Beattie thus endeavoured to disentangle the role of the metaphysician from that of the practical moralist, and in so doing distanced himself from the style of philosophizing cultivated by most of his Scottish contemporaries.

Beattie's change of direction was largely prompted by his animus against David Hume. In Beattie's view, Scottish academic philosophy was far too preoccupied with the epistemological issues raised by Humean scepticism and had lost sight of the primacy of moral preaching. Fearing the dominance of Hume's supposed 'party' of sceptics in Edinburgh, he turned south of the border for inspiration, and cultivated both the teaching practices and the form of polite culture which he associated with the ancient English universities.

3 Beattie, Reid and common-sense philosophy

Joseph Priestley was the first to bracket together the works of Beattie and Thomas Reid in his critique of common-sense philosophy, and most commentators have subsequently followed his lead. Priestley's elision of their writings, however, conceals as much as it reveals. Beattie was introduced to the elements of common-sense epistemology by Gerard, and this largely accounts for the differences between his exposition of the notion of common-sense and that found in the works of Reid. Moreover, Beattie himself lamented Reid's generous public treatment of Hume, which for him showed that Reid was too much of a metaphysician and lacked warmth in the cause of virtue. Beattie's dislike of metaphysics also accounts for the fact that, unlike Reid, he did not engage in a systematic refutation of the theory of ideas as it had evolved in the works of Descartes, Locke, Berkeley and Hume. Given that Reid's painstaking analysis of the mechanisms of human perception was the cornerstone of his response to scepticism, Beattie's lack of interest in this topic serves to underline the contrasts between them. Beattie was much more of a moral preacher and polemicist than Reid, and their divergent styles suggest that Beattie's place in the development of Scottish common-sense philosophy requires reassessment (see Common Sense School).

See also: Oswald, J.

PAUL WOOD

List of works

Much of Beattie's unpublished correspondence, along with related manuscripts and student lecture notes, is held in Aberdeen University Library. Other important archival materials are to be found in the Scottish Record Office and Edinburgh University Library.

Beattie, J. (1996) *The Collected Works of James Beattie*, ed. R. Robinson, London: Routledge; Bristol: Thoemmes Press, 10 vols.(Facsimile reprints of all of Beattie's works, plus hitherto uncollected miscellaneous writings and manuscript materials, along with Forbes' biography (1806).)

(1760) *Original Poems and Translations*, London: Millar.(A collection of Beattie's earliest poems and translations of selections from Anacreon, Horace, Lucretius, and Virgil.)

(1770) *An Essay on the Nature and Immutability of Truth, in Opposition to Sophistry and Scepticism*, Edinburgh: Kincaid & Bell; London: Dilly; repr., intro. F.O. Wolf, Stuttgart-Bad Cannstatt: Frommann, 1973.(Established Beattie's reputation as a moralist, and popularized common-sense philosophy.)

(1771-4) *The Minstrel: or, the Progress of Genius*, London: Dilly; Edinburgh: Kincaid & Bell.(Rooted in debates over the nature of poetry, *The Minstrel* consolidated Beattie's standing as a writer.)

(1776) *Essays*, Edinburgh: Creech; repr. New York: Garland Press, 1971; Hildesheim: Olms, 1974.(An expanded text of the *Essay*, plus essays on our psychological response to poetry and music, humour, and the usefulness of classical learning.)

(1783) *Dissertations Moral and Critical*, London: Strahan & Cadell; repr. New York: Garland Press, 1971; Hildesheim: Olms, 1974.(Six lengthy essays on philosophical, psychological, linguistic, moral, aesthetic, and literary topics derived from his lectures.)

(1786) *Evidences of the Christian Religion, Briefly and Plainly Stated*, Edinburgh: Creech; London: Strahan & Cadell, 2 vols.(Popular exposition of the rational case for the truth of Christianity.

(1790-3) *The Elements of Moral Science*, London: Cadell; Edinburgh: Creech, 2 vols; repr. Hildesheim: Olms, 1974.(Influental textbook comprised of revised versions of Beattie's lectures on psychology, natural theology, ethics, economics, politics and logic.)

(1820) *The Letters of James Beattie, LL.D. Chronologically Arranged from Sir W. Forbes's Collection*, London: Sharpe, 2 vols.(The largest published collection of Beattie's letters.)

(1908) *James Beattie 'The Minstrel': Some Unpublished Letters*, ed. A. Mackie, Aberdeen: Aberdeen Daily Journal Office.(Twenty-five letters of primarily biographical interest.)

(1946) *James Beattie's London Diary 1773*, ed. R.S. Walker, Aberdeen: Aberdeen University Press.(Beattie's detailed chronicle of his extended trip to London in the summer of 1773.)

(1948) *James Beattie's Day Book 1773-1798*, ed. R.S. Walker, Aberdeen: The Third Spalding Club.(A selection from Beattie's personal accounts which sheds light on his dealings with his publishers and a number of other topics.)

References and further reading

Bevilacqua, V.M. (1967) 'James Beattie's Theory of Rhetoric', *Speech Monographs* 34: 109-24.(A thorough treatment of Beattie's approach to rhetoric.)

Bower, A. (1804) *An Account of the Life of James Beattie, LL.D. Professor of Moral Philosophy and Logic, Aberdeen*, London: Baldwin.(The earliest biography, which has interesting material on Beattie's Aberdeen context.)

Fabian, B. and Kloth, K. (1967-70) 'The Manuscript Background of James Beattie's *'Elements of Moral Science*'', *Bibliotheck* 5: 181-9.(Contains a useful list of sets of student notes from Beattie's lectures.)

Forbes, M. (1904) *James Beattie and His Friends*, London: Constable; repr. Bristol: Thoemmes, 1990.(A biography which contains some information and additional letters not found in Forbes (1806).)

Forbes, W. (1806) *An Account of the Life and Writings of James Beattie, LL.D.*, Edinburgh: Constable & Creech, 2 vols.(A detailed biography by one of Beattie's close associates which incorporates a large number of Beattie's letters.)

Grave, S.A. (1960) *The Scottish Philosophy of Common Sense*, Oxford: Clarendon Press.(The standard modern work on the philosophy of Beattie, Reid and James Oswald.)

Hewitt, D. (1987) 'James Beattie and the Languages of Scotland', in J.J. Carter and J.H. Pittock (eds) *Aberdeen and the Enlightenment*, Aberdeen: Aberdeen University Press, 251-60.(A suggestive essay on Beattie and the politics of language in eighteenth-century Scotland.)

King, E.H. (1977) *James Beattie*, Boston: Twayne.(An introduction to Beattie's career as a man of letters.)

Phillipson, N. (1978) 'James Beattie and the Defence of Common Sense', in B. Fabian (ed.) *Festschrift für Rainer*

Gruenter, Heidelberg: Winter, 145-54.(An insightful exploration of the psychological dynamics of Beattie's reaction to Humean scepticism.)

Wood, P.B. (1993) *The Aberdeen Enlightenment: The Arts Curriculum in the Eighteenth Century*, Aberdeen: Aberdeen University Press.(A detailed discussion of matters outlined in §2, which contains further references.)

Beauty

On the subject of beauty, theorists generally agree only on rudimentary points about the term: that it commends on aesthetic grounds, has absolute and comparative forms, and so forth. Beyond this, dispute prevails. Realists hold that judgments of beauty ascribe to their subjects either a nonrelational property inherent in things or a capacity of things to affect respondents in a way that preserves objectivity. In both cases acute problems arise in defining the property and in explaining how it can be known. Classical Platonism holds that beauty exists as an ideal supersensible Form, while eighteenth-century theorists view it as a quasi-sensory property. Kant's transcendental philosophy anchors the experience of beauty to the basic requirements of cognition, conferring on it 'subjective universality and necessity'. Sceptics complain that the alleged property is merely a reflection of aesthetic pleasure and hence lacks objective standing. Partly due to its preoccupation with weightier matters, the philosophic tradition has never developed any theory of beauty as fully and deeply as it has, say, theories in the domain of morality. Comparative neglect of the subject has been encouraged by the generally subjectivistic and relativistic bent of the social sciences and humanities, as well as by avant-gardism in the arts. However, several recent and ambitious studies have given new impetus to theorizing about beauty.

1 Areas of general agreement about beauty

Almost all theorists concerned with the concept of beauty accept the following propositions. The reflective use of terms of beauty standardly expresses not a mere effusion but a favourable aesthetic judgment of a thing as a whole or in part. The commendation is normally associated with the capacity of the thing to yield pleasure, and presupposes a basis of beauty-making properties, on which beauty is said to supervene. The commendation is relative to (a) a threshold which varies widely with speakers, below which lie grades of lesser aesthetic value without a determinable minimum and above which rise grades of eminence also without a fixed limit, and (b) a comparison-class, which in different contexts ranges from the highly restricted ('a beautiful daisy') to the virtually universal ('a beautiful thing'), or a grade of competition ('a beautiful drawing for a beginner'). It seems we need go only a small step beyond these truisms to conclude that all normative disagreements about beauty reduce to differences of comparative rankings (more, less, or equally beautiful or unbeautiful) and that differences regarding threshold or terminology have no theoretical importance. Beyond these elements of the 'logic' of beauty, there is scant agreement.

2 Specimen issues concerning beauty

Perhaps the most fundamental issue is whether beauty exists in any substantial sense. The common saying that beauty is in the eye of the beholder, taken straightforwardly, amounts to a denial that beauty is a property of anything or that anything is genuinely worthy of being aesthetically admired. This denial is a first principle of aesthetic subjectivism (or nihilism), which, if consistent, confines its theorizing to the analysis of aesthetic preference as a socio-psychological phenomenon and as a variably rational or nonrational part of life. Generally, subjectivists in aesthetics, like those in ethics, retain the usual value terms but reinterpret them as avowing, expressing or soliciting preferences. In contrast, aesthetic realists seek to identify the property or state of affairs that beauty consists in and to explain how it can be known.

Many issues are common to aesthetic realism and subjectivism. For instance: (1) What is the range of things to which terms of beauty can be meaningfully applied? Some take beauty to be a transcendental, in the medieval sense of being a category that is applicable to everything. Others deny that it applies to certain classes. Flavours, scents, bodily sensations, thoughts, theories, abstractions, virtues and even natural objects are excluded by one thinker or another. Some allege that the proper referent of terms of beauty is never a physical object but an appearance or 'semblance'. (2) To what extent can aesthetic value be subsumed under the beautiful? Are the sublime, the pretty, the cute, the witty and the tragic species or degree-ranges of beauty, or are they distinct values? (3) To what extent may things of different types be meaningfully compared in respect of beauty? Parrots of a given species may be judged beautiful relative to one other, but can they be ranked against horses or houses? If not, can the beautiful be a single category of appraisal? (4) How determinate can judgments of beauty be when many factors enter into the case? On the face of it we stand on firmer ground in judging that a musical work is beautifully tender or sprightly than when we pass a summative judgment on the total ensemble of its qualities. This

has an obvious impact on comparisons: Beethoven's Fifth and Sixth symphonies are replete with beautiful aspects and moments, but can we sum these so exactly as to say which work is more beautiful?

3 Beauty as an intrinsic property

The simplest form of realism about beauty takes it to be an intrinsic or nonrelational property with strong *de facto* and *de jure* ties to love. For Plato and Plotinus it is a supersensible abstract Form, better exemplified by abstractions than by concrete particulars, and supremely exemplified by itself. Acquaintance with beauty begins in commerce with particulars, but only pure thought, on the model of mathematical and moral intuition and demonstration, can elevate the opinions gained through acquaintance to the level of knowledge. Though their theoretical framework does not by itself entail particular normative principles, Platonically-minded thinkers usually favour Apollonian values of order, clarity, harmony and balance as opposed to Dionysian values of profusion, sensuality and vehemence.

A basic question left unanswered by theories of this type concerns the nature of the property of beauty. Neither Plato nor Plotinus offer to identify the property of beauty, and in their writings it tends to acquire a mystical air, due to the obscure nature of its purer exemplars (the Forms) and the extreme breadth of its range. The latter makes it difficult to imagine how any nonrelational property could account for all the indicated sorts of beauty, and the difficulty is compounded by suggestions of a single, universal rank-ordering. Answers to such questions are hampered by the vagueness of theorists' accounts of the relation between properties and their instances, which is especially acute if one takes literally Plato's claim that the property of beauty is supremely self-exemplifying. Platonic ontology aside, citations of primary values such as the medieval triad of clarity, splendour and proportionality invariably omit to supply a working criterion of any of the three, or a summation rule to decide between things which differ in more than one of them.

4 Response-related realist conceptions of beauty

A second sort of realism takes beauty to be essentially related to human response. The manner of relatedness varies. For example, G.E. Moore's definition of beauty as 'that of which the admiring contemplation is good in itself' avoids any reference to actual effects on human feeling. It speaks only of effects which, should they occur, would contribute to the intrinsic value of a state of mind. This allows things to be beautiful even if they are never experienced, or if experienced never admiringly contemplated. At the same time it ties beauty conceptually to human experience. At first sight such a theory has no difficulty explaining the diversity of beautiful things, for why should not endlessly diverse constellations of material and mental properties be intrinsically good to contemplate admiringly? But doubts arise when one asks how to determine which ones qualify. Obviously it will not do to answer, 'the beautiful ones'. And in the end the answer seems irretrievably buried in the nature of intrinsic goodness, since on Moore's view that property is not only simple and therefore unanalysable, but non-natural and consequently inaccessible to empirical observation. One must reach such results as one can by an intuition which, if rational, is inexplicably so.

British eighteenth-century theorists (Hutcheson and Hume, for example) advanced a naturalistic conception of beauty as a relational property, sufficiently similar to a sensory colour or sound to be known by a faculty called the sense of beauty. The sense is internal, in that it responds directly to mental representations, not to external stimuli. A considerably sanitized and amplified development of the central idea might go as follows. The surface colour of a red object is definable as a power (now more commonly called a disposition) to excite a red appearance in optimally colour-sighted percipients under optimal conditions; likewise the beauty of a sensible object is its power to produce, via impressions of the external senses, a beauty-datum (an 'idea' of beauty, in eighteenth-century parlance) in optimally beauty-receptive observers under optimal conditions. (Deviations are explained by non-optimality affecting some part of the sensory process.) The beauty-datum of choice is disinterested pleasure, defined (for instance, by Shaftesbury) as pleasure taken wholly in the object and not at all in the self, or more exactly not in the self's pleasure. Since disinterested pleasure is arguably also a mark of moral discernment (sometimes ascribed to a companion moral sense), an additional qualification is required if beauty is not to be identified with the moral good. (See the proposal by Kant, below.) To cover things of abstract beauty, the theory must assume that the sense also yields a beauty-datum in response to the mental representation of theorems, proofs and so forth, under appropriate conditions of optimality.

Discussions of erroneous judgments of beauty occur in the works of all the major British theorists, Hume's perhaps being most often cited. Pooling the suggestions therein, one might obtain an honestly empirical criterion of accuracy of the sense of beauty by requiring a consensus among maximal beauty-discriminators under optimal beauty-discriminating conditions. Failing that, beauty would not exist. To complete the analogy with sensory properties, the consensus would also have to correlate with a basis in the object comparable to wavelengths, reflectance and the like for sensory properties. Uniformity and variety 'in compound ratio' is Hutcheson's all-purpose candidate for this role. An appropriate brain structure for the sense must also be assumed. None of the theorists acknowledges the full set of necessary components or the theoretical and practical difficulties of establishing that all of them actually exist.

Immanuel Kant's theory of beauty introduces important new ideas, though in such a way as to provoke endless controversy over their precise content and validity (see Kant, I. §12). Aesthetic pleasure is distinguished from its moral counterpart by its freedom from concern with the actual existence or nonexistence of the object of pleasure. Kant circumvents the need to rely on actual agreement among qualified judges by an ingenious but cryptic account of the deep source of pleasure. This is, he says, the harmonious free play of the cognitive powers, the imagination and understanding - free in the sense of not being directed towards actual knowledge. In ordinary cognition the imagination supplements the received sensory data so as to fit them to a concept supplied by the understanding. The result is an objective judgment. In aesthetic contemplation the imagination is free to seek out relations of form without concern for cognitive relevance to the scene which contains them; and the understanding is free to accept the yield and set new pattern-finding tasks for the imagination. Since the routines of this play are indispensable staples of ordinary cognitive processing, they belong to the repertoire of all humans beings, and the pleasurable harmony of the faculties can be presumed universal and necessary - a pleasure we have a right to expect everyone to feel. Because the play is not controlled by definite concepts, the judgment of free beauty is irremediably singular and incapable of being generalized into a rule of beauty.

Kant also recognizes judgments of 'dependent' beauty, where a concept imposes a limitation. For example, to judge the beauty of a human body, one must take account of whether the properties of the body (for example, its proportions) are such as to facilitate the functions of a human being. Only such free beauty as fulfills this condition affects the overall assessment. Though less purely aesthetic than free beauty, dependent beauty is of far greater human significance. Artistic beauty falls under this head, since all art is to some extent constrained by a purpose.

5 Restoration of beauty as a theoretical subject

Twentieth-century philosophies of beauty are comparatively rare, due to a general shift of interest towards the aesthetic. An exception to the rule is a searching, broad-gauge theory by Guy Sircello, which construes virtually all aesthetic values as species of beauty and explains beauty in terms of non-defective and non-defective-seeming 'properties of qualitative degree'. As these properties are highly context-dependent both for their existence and for their status as non-defective and non-defective-seeming, it is hard to be sure of their ontological or, to an extent, their epistemological standing. Moreover, Sircello does not hazard a sufficient condition for the overall beauty of things as opposed to their beauty in respect of a property of the sort mentioned. Another effort was made to revive the theory of beauty by Mary Mothersill, after reflection on Kant's aesthetics, and especially on his dictum that there can be no laws of taste. On her view 'Any individual is beautiful if and only if it is such as to be a cause of pleasure in virtue of its aesthetic properties'. The latter are understood in a new way, namely as properties so sensitive to their context as to be incapable of being possessed by two individuals unless the latter are perceptually indistinguishable under standard conditions of observation. For example, the distinctive wavelike contour running throughEl Greco's *Burial of Count Orgaz*, when taken in the full context provided by the painting, is qualified by the totality of perceptible relationships between it and all other features of the design. In this way she supportsKant's idea of the judgment of beauty being logically singular, since such radically contextualized properties cannot figure in non-trivial laws or principles. Mothersill's theory, developed with skill and panache, raises daunting problems but, likeSircello's, opens up possibly profitable lines of research.

6 Subjectivism and relativism regarding beauty

Subjectivist theories of beauty, while relieved of the commitment to the identity and 'knowability' of beauty, raise important problems involving the rationality of aesthetic preference. If a judgment of beauty, taken as avowing or expressing an aesthetic pro-attitude, can be counted unjustified if based on error or ignorance regarding its object,

then aesthetic subjectivism can impute fault without implying that beauty is a property of anything. Is it amenable to principles of rationality beyond that point? Can it justify favouring consistency in aesthetic attitudes over time or the subsumption of particular attitudes under attitudes of wide scope? If people generally concur in an error-free aesthetic attitude regarding a thing, does that fact give one a reason for suspicion about one's own dissenting but equally error-free attitude?

Relativism holds that beauty varies with cultures or 'taste-publics' without collapsing into personal preference. A version close to the dispositional account in §4 could be obtained by relativizing the beauty-disposition to the maximal discrimination-capacities attainable in distinct cultures, especially if some neural explanation of the difference could be found - following the analogy of sensory colour, which is vulnerable to the same relativization if different cultures are found to have different but equally acute colour-sensibilities. Typically, however, relativists tend towards a notion of aesthetic taste so malleable as to leave no chance of a quasi-realist account - without, it must be observed, offering anything like compelling evidence. An alternative explanation of cultural difference as a specialization-phenomenon is at least as plausible. On this (realist) view, cultures develop special competences in distinct ranges of beauty, without there being any incompatibility among the values most reliably assessed by the respective cultures.

See also: Aesthetic concepts; Hume, D.; Plotinus §3; Artistic taste

JOHN H. BROWN

References and further reading

Brink, D.O. (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press, ch. 2. (Clearer and more detailed explanation of varieties of realism and antirealism with reference to values than can be found in the aesthetic literature.)

Hume, D. (1757) 'Of the Standard of Taste', in *Of the Standard of Taste and Other Essays*, ed. J. Lenz, New York: Macmillan, 1965.(Admirably clear discussion of criteria of accuracy of the supposed sense of beauty.)

Hutcheson, F. (1725) *An Inquiry concerning Beauty, Order, Harmony, Design*, ed., with intro. and notes by P. Kivy, The Hague: Martinus Nijhoff, 1993.(The fullest formulation of the eighteenth-century sense of beauty theory.)

Kant, I. (1790) *Critique of Judgment*, trans. J.H. Bernard, New York: Macmillan, 1951.(The first part, 'Critique of Aesthetic Judgment', contains Kant's theory of beauty and sublimity.)

Margolis, J. (1976) 'Robust Relativism', *Journal of Aesthetics and Art Criticism* 35: 37-46, reprinted in J. Margolis, *Philosophy Looks at the Arts*, Philadelphia, PA: Temple University Press, 1978, esp. 387-401.(Brief exposition and defence of aesthetic relativism regarding aesthetic judgments and critical interpretations.)

Moore, G.E. (1903, 1959) *Principia Ethica*, Cambridge: Cambridge University Press, ch. 4, 'The Ideal'.(A classic statement of the intuitionist position regarding moral and aesthetic value.)

Mothersill, M. (1984) *Beauty Restored*, Oxford: Oxford University Press.(Contains Mothersill's theory and discussions of historical positions regarding beauty; useful bibliography.)

Plato (*c*.380s BC) *Phaedrus*, trans. R. Hackford, in *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1963, 244-257.(Plato's ideas about love and beauty are presented mythopoeically in Socrates' speech to the god of love.)

(*c*.380s BC) *Symposium*, trans. M. Joyce, in *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1963, 209d-212a.(The classic account of the ascent of the soul to the vision of absolute beauty.)

Plotinus (*c*.AD260) *Enneads*, trans. S. McKenna, ed. J. Dillon, New York: Viking Penguin, 1991.(See Ennead I, 6th tractate, 'Beauty'; and Ennead V, 8th tractate, 'On the Intellectual Beauty'.)

Shaftesbury, A. (1711) *Characteristics of Men, Manners, Opinions, Times*, ed. J.M. Robertson, Indianapolis, IN: Bobbs-Merrill, 1964.(Shaftesbury's idea of disinterested pleasure is set forth in Treatise IV, Book 2, Part 2, section 1.)

Sircello, G. (1975) *A New Theory of Beauty*, Princeton, NJ: Princeton University Press.(A mostly non-technical exposition of the theory discussed in §5.)

(1989) *Love and Beauty*, Princeton, NJ: Princeton University Press.(Links the author's theory of beauty to an equally ambitious theory of love.)

Beauvoir, Simone de (1908-86)

Simone de Beauvoir, a French novelist and philosopher belonging to the existentialist-phenomenological tradition, elaborated an anthropology and ethics inspired by Kierkegaard, Husserl, Heidegger and Sartre in Pyrrhus et Cinéas (1944) and Pour une morale de l'ambiguïté (The Ethics of Ambiguity) (1947). In her comprehensive study of the situation of women, Le deuxième sexe (The Second Sex) (1949), this anthropology and ethics was developed and combined with a philosophy of history inspired by Hegel and Marx. The most prominent feature of Beauvoir's philosophy is its ethical orientation, together with an analysis of the subordination of women. Her concept of woman as the Other is central to twentieth-century feminist theory.

1 Life

Simone de Beauvoir was born on 9 January 1908 in Paris and lived there almost all her life. Her parents belonged to the bourgeoisie and provided her with a traditional Catholic education. After studies at the Sorbonne, Beauvoir took an *agrégation* in philosophy in 1929 (that is, a higher teaching exam) at the prestigious École Normale Supérieure. There she met Jean-Paul Sartre, with whom she entered a lifelong bond of intellectual companionship. They never married, nor had any children, but stayed together in a free liaison, allowing intimate relations with others.

In the 1930s, they both studied the phenomenology of Husserl and Heidegger, and the existential philosophy of Kierkegaard. During and after the war, Beauvoir developed an interest in Hegel and Marx, especially in the philosophy of the young Marx. In 1945 Beauvoir, Sartre and Maurice Merleau-Ponty founded *Les Temps Modernes*, a literary, philosophical and political journal. The same year, Sartre and Beauvoir became known as 'existentialists', a label they reluctantly accepted. Both became leading intellectuals of their generation, politically engaged but non-affiliated leftists. Beauvoir, for her part, also inspired and participated in the feminist movement of the 1970s and 1980s.

Generally, Beauvoir is better known as a novelist than as a philosopher. Still, during the 1940s she produced a number of philosophical essays, the most important being *Pyrrhus et Cinéas* (1944) and *Pour une morale de l'ambiguïté (The Ethics of Ambiguity)* (1947). This was followed by *Le deuxième sexe (The Second Sex)* (1949), labelled one of the most influential books of the twentieth century; its underlying philosophical structure reveals her thought in its mature phase. Novels like *L'Invitée (She Came to Stay)* (1943) and *Tous les hommes sont mortels (All Men are Mortal)* (1946), are also organized around philosophical questions.

2 *The Ethics of Ambiguity*

In *Pyrrhus et Cinéas*, Beauvoir, like Kierkegaard and Sartre, defines the human being as an existent that, lacking an inherent essence, has to form its life and give it meaning. This ethos is repeated in *The Ethics of Ambiguity*, where Beauvoir radically transforms Sartre's view of the human being as a 'useless passion'. According to Sartre, the human is characterized by both the lack of being and an unrealizable passion, the 'desire of being' (*désir d'être*), that is, to attain a fixed identity or essence. Beauvoir adds to this the Heideggerian notion of *Erschlossenheit* (disclosure; in French, *dévoilement*), therewith pointing also to the positive side of existence. According to her, the human being not only lacks and desires being, but also wants to 'disclose being'. Or, through the human's vain desire, the world is disclosed, that is, appears and is given meaning.

Affirmation of this positive side of existence is identified with authenticity: to 'deny the lack as lack', to go through a 'conversion', is to set the 'will to be "in parentheses"'; and further, a recognition of both oneself and the other as free - 'to will man free… is to will the disclosure of being in the joy of existence' ([1947] 1948: 135). Beauvoir thus affirms freedom as a fundamental human characteristic. At the same time she is careful to distinguish the freedom of human consiousness - a free spontaneity, a nothingness without given structures - from freedom as a social and political reality. Central to her ethics is the idea that freedom ought to be founded and defended by the individual. On the other hand, society has to facilitate conditions for the positive fulfilment of this freedom. There are thus objective differences between situations, a theme which is further developed in *The Second Sex* by the aid of such Hegelian-Marxist distinctions as that between 'abstract' and 'concrete' freedom and between 'positive' and 'negative' freedom.

Another central theme introduced in *Pyrrhus et Cinéas* and developed in *The Ethics of Ambiguity* concerns the interdependence of humans. In the latter work she states, 'The me-others relationship is as indissoluble as the subject-object relationship' ([1947] 1948: 72). In *The Second Sex* this interdependence is conceptualized in Heideggerian terms as *Mitsein* (Being-with). Human reality is a being-with-others, and the subject 'achieves freedom only through a continual reaching out towards other freedoms' ([1949] 1953: 27), a theme which points to the ethical orientation of her thought. Unlike Sartre in *Being and Nothingness*, Beauvoir does not see conflict as inevitable, nor intersubjectivity as impossible (see Existentialism; Existentialist ethics).

3 *The Second Sex*

If a prominent feature of Simone de Beauvoir's philosophy is its ethical orientation, another is the analysis of oppression. In *The Second Sex*, the existential anthropology and ethics is transformed under the influence of Hegel and Marx, to be combined with a philosophy of history. Consequently, consciousness is conceptualized as historically mediated and human beings as socially situated. Beauvoir declares, with Merleau-Ponty, that 'man… is a historical idea'. Following Hegel and Kojève, she understands the origins of humanity as characterized by a struggle for recognition between men, which led to the genesis of inequalities and oppression. This Hegelian master-slave dialectic is combined with a Marxist insistence on the importance of productive activity, or work, as key to the development of both the human being and society. Since women, owing to their reproductive function and their lesser physical strength, stood outside both the struggle for recognition and productive activity, and therefore outside the basic dialectic, they were defined by the males as the absolute Other (*l'Autre*). They were cast in the role of the object that never became the subject in relation to men, a situation that with the advent of private property and the state became institutionalized into patriarchal society.

The concept of woman as the Other is probably Beauvoir's most important contribution to philosophy and feminist theory. Accepting the existentialist credo 'essence does not precede existence', Beauvoir rejects any idea of an inherent femininity and asserts 'One is not born, but rather becomes, a woman' ([1949] 1953: 273). Gender is therefore conceptualized as simultaneously socially produced and self-created within the confines of the socio-historical situation. Beauvoir emphasizes that because women have historically been the subordinate sex, the Other from the dominant male point of view, they have also been defined as having a determinate nature, be it evil or good. Women have thus not been seen as subjects in their own right, something which Beauvoir criticizes as inauthentic.

The subordination of women is explained in *The Second Sex* not only as a social and historical phenomenon, but also from an existentialist perspective. If the 'desire of being', in Sartrean terms, is equated with an inauthentic flight from freedom and responsibility, it is equally for Beauvoir an explanation of oppression and submission. Man has sought to fulfil his desire by taking possession of a woman; yet the attempt is as ever vain, and ends only in the man's alienation of himself in the woman. Similarly, woman has tried to fulfil her desire by alienating herself in man as if he were an absolute subject and could take responsibility for her life. The existentialist 'desire of being' is combined in *The Second Sex* with a concept of alienation, influenced by that of the young Marx. This implies a distinction between alienation, which is synonymous with a search for being through the other, or through what one has, and authentic self-fulfilment through objectifying oneself in what one does, through conscious, freely chosen, object-creating activity. This authentic self-fulfilment is also related in *The Second Sex* to a 'conversion', which implies a renunciation of the oppressive 'possession' of others and a recognition of the other as a subject.

In *The Second Sex*, Beauvoir sought to combine an ethics with a theory of oppression and an existentialist phenomenology with a philosophy of history, the latter a combination not unusual in Paris in the 1940s, but none the less fraught with problems. The theory of authenticity, which is developed from Sartre's ahistorical, dualistic ontology, tends to collide with the dialectical and historical ontology inspired by Hegel and Marx. Even though Beauvoir is a forerunner for modern feminist theory, her philosophy has been criticized by feminists. The central concept of transcendence, which implies being a free, self-determining subject that can realize itself in self-chosen activities, is problematic since it tends not only to be conflated with authenticity, but also to be equated with men's traditional way of life and activities. For women to leave immanence, the confined animal-like existence that they have been assigned, they have to transcend the traditional female life-world, which is accordingly denigrated in Beauvoir's philosophy. Bearing and rearing children is not defined as transcendence. Simone de Beauvoir does not

question the apparent androcentricity in this, but uses the concept to criticize the exclusion of women from the public sphere and from the arts (see Feminism; Feminist ethics; Feminist political philosophy).

See also: Phenomenological movement §4

EVA LUNDGREN-GOTHLIN

List of works

Beauvoir, S. de (1943) *L'invitée*, Paris: Gallimard; trans. Y. Moyse and R. Senhouse, *She Came to Stay*, London: Secker & Warburg, L. Drummond, 1949.(Beauvoir's debut, a novel which is an account of a triangular relationship between one man and two women. Its philosophical theme is the conflictual relation between human beings, considering the fact that one exists for and is judged by the other.)

Beauvoir, S. de (1944) *Pyrrhus et Cinéas*, Paris: Gallimard.(Beauvoir's first philosophical essay, which analyses the human being as transcendence and human interdependence. It is also her first inquiry into existentialist ethics.)

Beauvoir, S. de (1946) *Tous les hommes sont mortels*, Paris: Gallimard; trans. L.M. Friedman, *All Men Are Mortal*, Cleveland, OH: World Publishing Company, 1955.(Through the story of an immortal man whose experiences reach over the centuries, Beauvoir investigates in this novel the importance of human mortality for the meaning of human life.)

Beauvoir, S. de (1947) *Pour une morale de l'ambiguïté*, Paris: Gallimard; trans. B. Frechtman, *The Ethics of Ambiguity*, New York: Philosophical Library, Citadel, 1948.(Simone de Beauvoir's existentialist ethics, centred around the concepts of ambiguity. Authenticity, freedom and the relation between ethics and politics are essential themes.)

Beauvoir, S. de (1948) *L'existentialisme et la sagesse des nations*, Paris: Nagel.(Contains four essays previously published in *Les temps modernes*, whose themes are the metaphysical novel, ethics and politics, revenge and justice, and whether existentialism is a philosophy of despair.)

Beauvoir, S. de (1949) *Le deuxième sexe, tome I, Les faits et les mythes, tome II, L'expérience vécue*, Paris: Gallimard; trans. and ed. H.M. Parshley, *The Second Sex*, London: Jonathan Cape, 1953.(A comprehensive study of the situation of women from prehistory to the 1940s, which shows Beauvoir's philosophy in its maturity. Introduces the important notion of woman as the Other. Note that the English translation is neither complete nor wholly correct philosophically.)

Beauvoir, S. de (1951-2) 'Faut-il brûler Sade', *Les temps modernes*, Dec. 1951, Jan. 1952; also in *Privilèges*, Paris: Gallimard, 1955; trans. A. Michelson, 'Must We Burn de Sade?', in *The Marquis de Sade: An Essay by Simone de Beauvoir, with Selections from his Writings*, New York: Grove Press, 1953.('Faut-il brûler Sade', an essay on Marquis de Sade's philosophy, also shows important aspects of Beauvoir's ethics. Republished in the essay collection *Privilèges* along with 'La pensée de droite aujourd'hui', which treats right-wing ideology, and 'Merleau-Ponty et le pseudosartrisme', a defence of Sartre's philosophy against Maurice Merleau-Ponty's critique in *Les aventures de la dialectique*.)

References and further reading

Butler, J. (1986) 'Sex and Gender in Simone de Beauvoir's Second Sex', in *Simone de Beauvoir: Witness to a Century, Yale French Studies* 72: 35-49, ed. H.V. Wenzel.(A critical account of *The Second Sex* from a feminist perspective.)

Heinämaa, S. (1997) 'What is a woman? Butler and Beauvoir on the Foundations of the Sexual Difference', *Hypatia* 12 (1).(Relevant critique of the usual interpretation of Beauvoir's *The Second Sex* as a theory of gender. Maintains instead that her book should be seen as a phenomenological description of the sexual difference.)

Kruks, S. (1990) *Situation and Human Existence: Freedom, Subjectivity and Society*, London: Unwin Hyman.(An important study of Beauvoir's concepts of freedom and situation and the philosophical relationship between Beauvoir, Sartre and Merleau-Ponty.)

Le Dœuff, M. (1989) *L'étude et le rouet: des femmes, de la philosophie, etc.*, Paris: Seuil; trans. T. Selous, *Hipparchias's Choice: An Essay Concerning Women, Philosophy, etc.*, Oxford: Blackwell, 1991.(Discusses the difficulties women face in gaining recognition as philosophers, with Beauvoir as a case study. Analyses the androcentric aspects of Sartre's philosophy and illustrates the differing approaches to existentialism taken by Sartre and Beauvoir.)

Lundgren-Gothlin, E. (1996) *Sex and Existence: Simone de Beauvoir's 'The Second Sex'*, London: Athlone, and New England: Wesleyan University Press.(An analysis of the philosophical foundations and structure of *The*

Second Sex. Expansion of the material of §§2-3 of this entry. Originally published in 1991 in Swedish.)

Moi, T. (1994) *Simone de Beauvoir, The Making of an Intellectual Woman*, Oxford: Blackwell.(A comprehensive study of her life and work from a combined socio-historical, psychoanalytic and feminist perspective.)

Seigfried, C.H. (1984) 'Gender-Specific Values', *Philosophical Forum* 15 (4): 425-42.(A feminist critique of Beauvoir's concepts of transcendence and immanence.)

Simons, M.A. (1983) 'The Silencing of Simone de Beauvoir: Guess What's Missing from The Second Sex', *Women's Studies International Forum* 6 (5): 559-64.(Gives an account of the omissions and of the mistranslations of philosophical terms in the English edition of *The Second Sex*. Essential for readers of the English edition.)

Simons, M.A. (1986) 'Beauvoir and Sartre: The Philosophical Relationship', in *Simone de Beauvoir: Witness to a Century, Yale French Studies* 72: 165-79, ed. H.V. Wenzel. (An account of the complex, two-way, philosophical relationship between Beauvoir and Sartre.)

Simons, M.A. (ed.) (1995) *Feminist Interpretations of Simone de Beauvoir*, Pennsylvania, PA: Pennsylvania State University Press.(A collection of essays treating various aspects of Beauvoir's philosophy, her ethics, her views on the body and sexuality, her concept of freedom, and the relationship between her philosophy and Sartre's. For the most part not difficult reading.)

Singer, L. (1985) 'Interpretation and Retrieval: Rereading Beauvoir', *Women's Studies International Forum* 8 (3): 231-8. (An analysis of Beauvoir's concept of freedom and its relation to her ethics.)

Vintges, K. (1996) *Philosophy as Passion: The Thinking of Simone de Beauvoir*, Bloomington and Indianapolis, IN: Indiana University Press.(An easily comprehensible summary of Beauvoir's philosophy and its relation to her life. Argues that Beauvoir's ethics is an 'art of living', which she formulated through her autobiography and fiction.)

Beck, Jacob Sigismund (1761-1840)

Beck played a brief but important role in the development of post-Kantian philosophy. A former student of Kant, he published at his teacher's instigation three volumes of 'Explanatory Abstracts' *of Kant's major writings. In the third volume Beck presented what he regarded as the 'Only Possible Standpoint' from which Critical Philosophy had to be judged if misunderstandings of Kant's work were to be avoided. His 'Doctrine of the Standpoint' involved a 'reversal' of the method of the Critique of Pure Reason and the elimination of the 'thing-in-itself' from Kant's theoretical philosophy.*

Beck was born in West Prussia in 1761. As a young man he went to Königsberg to study philosophy with Kant, and mathematics with C.J. Kraus and J. Schultz. In 1791 he habilitated at the University of Halle with a thesis in mathematics. In the same year, he accepted an offer from Kant to write *Erläuternde Auszüge (Explanatory Abstracts)* of Kant's major writings. Kant's hope was that Beck would make his philosophy more accessible to the general reader and answer some of the criticisms that had been levelled against it.

Over the next few years, a highly instructive correspondence ensued between them in whichBeck elicited often detailed comments from his teacher. For example, in response to an enquiry about the dynamical theory of matter, Kant admitted that the explanation of the difference in density in matter that he had given in his *Metaphysical Foundations of Natural Science* was circular. As a result, Kant later significantly revised his theory of matter in his *Opus postumum*.

In 1793 the first volume ofBeck's *Auszug* appeared, explicatingKant's *Critique of Pure Reason* and *Critique of Practical Reason*. The second volume, addressing Kant's *Metaphysical Foundations of Natural Science* and *Critique of Judgment*, came out in the following year. This volume also made available for the first time the original introduction to the *Critique of Judgment*, which Kant had discarded in favour of a slightly shorter one and later entrusted to Beck.

Most important, however, is the third volume of the *Auszug* (1796), which Beck called the 'Only Possible Standpoint from which Critical Philosophy Must Be Judged'. From the start Beck had planned to respond to critics of the Kantian philosophy in a separate volume. As his work progressed, however, he became convinced that the presentation of transcendental philosophy in the first *Critique* leads the reader astray on several accounts. If it is the function of the categories to make originally possible the reference of our representations to an object, he wrote to Kant (11 November 1791), one cannot define 'intuition' in the manner of the *Critique* as a representation which immediately relates to an object. Nor is it appropriate to define the categories simply as concepts: reference to objects is not the result of, but is presupposed by, the application of concepts. Transcendental philosophy must thus elucidate the fundamental act through which we represent to ourselves an object originally. For in all thought, Beck argued, we first of all fix for ourselves a point of reference to which we then attribute (*beylegen*) determinations. Concepts are the results of this operation, which Beck calls 'original representing'.

Beck's 'Standpoint' consequently 'reverses' the method of the first *Critique*. WhereasKant had led the reader gradually to the highest point of transcendental philosophy, or the synthetic unity of consciousness,Beck commences with it: with the postulate to represent originally. He then argues that the categories are but different modes of this original activity. Space emerges in original representing; it is the original synthesis of the homogeneous: 'Before this synthesis there is no space; we generate it, rather, in the synthesis'. When the synthesis is fixated in an 'original recognition', there first arises the notion of a determinate figure: 'Together with original recognition, original synthesis generates the objective unity of consciousness… that is, it generates the original concept of an object' (1793-6: 144). Beck thus undermines the sharp distinction, so crucial to Kant's thinking, between intuition and concept, between what is given and what is thought. Since all objective reference is the result of original representing, he regards as meaningless the notion of a 'bond' or connection between representation and object. Indeed, the concept of such a bond, Beck argues, 'is the source of all errors of speculative philosophy'. The concept of a thing-in-itself is an aberration of critical philosophy. As he later wrote toKant, 'My intention was to bar the concept of the thing-in-itself from theoretical philosophy' (Letter of 20 June 1797).

Beck's 'Standpoint' met with mixed reactions. The orthodox Kantians regarded it as heresy, but others were more positive: J.G. Fichte, for example, acknowledged its merits in both introductions to his *Science of Knowledge*. Beck himself, however, sought his teacher's approval; when his efforts to convinceKant of the importance of his 'Standpoint' seemed to fail, his disappointment grew proportionally. In 1797Kant named notBeck but his colleague Johann Schultz when challenged to state publicly who best understood his philosophy ('Declaration against Schlettwein'), and when, two years later, he included Beck in his notorious 'Open Letter on Fichte's *Wissenschaftslehre*', Beck's alienation from his former mentor was complete. However, unbeknown toBeck, Kant's own reflections at the time began to make him more appreciative of some of Beck's ideas. In correspondence with J.H. Tieftrunk, Kant admitted that Beck's method of beginning with the categories was possible, and in the *Opus postumum* he eventually experimented himself with what could be described as a reversal of the method in Beck's sense: 'that we have insight into nothing except what we can make ourselves. First, however, we must make ourselves.Beck's original representing' (1936-8: 114).

In 1799 Beck was appointed to a chair at the university of Rostock, where he soon became rector. Although he published several books over the next few decades, they were of little philosophical consequence.Beck died in Rostock in 1840.

See also: German idealism

ECKART FÖRSTER

List of works

Beck, J.S. (1793-6) *Erläuternder Auszug aus den critischen Schriften des Herrn Prof. Kant, auf Anrathen desselben (Explanatory Abstracts of Prof. Kant's Critical Writings, at the Latter's Instigation)*, Riga: Hartknoch, 3 vols; partly trans. in G. di Giovanni and H.S. Harris (eds) *Between Kant and Hegel*, Albany, NY: State University of New York Press, 1985.(This isBeck's principal text. Volume 3 is subtitled *Einzig-möglicher Standpunct, aus welchem die critische Philosophie beurtheilt werden muß (Only Possible Standpoint from which Critical Philosophy Must Be Judged)*.)

Beck, J.S. (1796) *Grundriß der critischen Philosophie*, Halle: Renger, trans. J. Richardson, *The Principles of Critical Philosophy*, London, 1797.(Another exposition of Kant's philosophy from Beck's standpoint.)

Beck, J.S. (1798) *Commentar über Kants Metaphysik der Sitten (Commentary onKant's Metaphysics of Morals)*, Halle: Renger.(Beck's moral philosophy.)

References and further reading

Fichte, J.G. (1797) 'Erste und zweite Einleitung in die *Wissenschaftslehre*' (First and Second Introductions to the *Science of Knowledge*), trans. and ed. P. Heath and J. Lachs, in *Science of Knowledge*, Cambridge: Cambridge University Press, 1982. (Fichte's subsequent 'introductions' to his main work.)

Kant, I. (1781) *Kritik der reinen Vernunft*, trans. N. Kemp Smith, *Critique of Pure Reason*, London: Macmillan, 1929.(Kant's principal work.)

Kant, I. (1936-8) *Opus postumum*, partly trans. E. Förster and M. Rosen in *Kant's Opus postumum*, ed. E. Förster, Cambridge and New York: Cambridge University Press, 1993.(Kant's last, unfinished *magnum opus*.)

Kant, I. (1900-2) *Kant's Briefwechsel*, partly trans. A. Zweig, in *Kant: Philosophical Correspondence 1759-99*, Chicago, IL: Chicago University Press, 1967.(A selection of letters to and from Kant.)

Meyer, T.L. (1991) *Das Problem des höchsten Grundsatzes der Philosophie bei Jacob Sigismund Beck (The Problem of a Highest Principle of Philosophy in Jacob Sigismund Beck)*, Amsterdam: Rudolphi.(A useful introduction with an extended bibliography.)

Schmucker-Hartmann, J. (1976) *Der Widerspruch von Vorstellung und Gegenstand. Zum Kantverständnis von Jakob Sigismund Beck (The Opposition of Representation and Object. On Jacob Sigismund Beck's Interpretation of Kant)*, Meisenheim am Glan: Anton Hain.(Thorough but more difficult.)

Wallner, I.M. (1984) 'A New Look at J.S. Beck's "Doctrine of the Standpoint"', *Kant-Studien* 75: 294-316.(A phenomenological interpretation of Beck's main text.)

Behaviourism in the social sciences

Classical behaviourism has had almost no direct reflection in the social sciences, in that there has never been a behaviourist social psychology or sociology. However, various features of the cluster of behaviourist doctrines have been widespread in the human sciences. Behaviourism as it developed from its roots in the proposals of Watson, and in its transformation by Skinner, had two influential aspects, one metaphysical and the other methodological. The metaphysics of behaviourism was positivistic. It was hostile to theory, favouring a psychology the subject matter of which was limited to stimuli and responses. It was hospitable to the conception of causation as regular concomitance of events, rejecting any generative or agent causal concepts. The methodology of behaviourism was hospitable to simple experimental techniques of inquiry, seeking statistical relations between independent and dependent variables. It was hostile to descriptions of human action that incorporated the intentions of the actor, favouring a laconic vocabulary of neologisms. Metaphysically and methodologically behaviourism favoured the individual as the locus of psychological phenomena. But, in practice, the use of statistical analyses of data abstracted psychological processes from real human beings leaving only simplified automata in their place.

1 The behaviourist legacy in social psychology

Despite the influence of aspects of behaviourism on mainstream psychology, in one important respect experimental (or 'old paradigm') social psychology was anti-behaviourist, in that it was in some ways anti-positivist. For example Tajfel's (1982) theory of intergroup relations proposed a cognitive mechanism of social comparison to explain observed correlations. Festinger's (1957) cognitive dissonance theory was based upon the alleged existence of cognitive discomfort in the presence of contradiction between expressed beliefs and actions. Yet the methodological influence of behaviourism is very clear in that both Tajfel and Festinger, and indeed many social psychologists of the post-war generation employed an experimental methodology that was a direct descendent of the methods of inquiry of behaviourism. Only the statistical trends which could be discerned in objectively described experimental manipulations of independent variables were to be admitted into an alleged 'truly' scientific social psychology as data. In this respect 'new paradigm' social psychology, particularly the role-rule model of Harré and Secord (1972), differed considerably in that a methodology of account analysis was recommended instead of simple experimentation. The history of social psychology can be seen in two stages: first the rejection of the metaphysics of behaviourism and only much later the abandonment of its methodology.

Behaviourism was both a learning theory according to which all human behaviour consisted in conditioned responses to types of environmental contingencies and a method of investigation in which the study of mature organisms simply consisted in the attempt to manipulate stimuli so as to discover which responses had indeed been installed in them. Even after the abandonment of behaviourism as general theory of human conduct (Danziger 1990) the use of the 'experimental' method persisted in social psychology. It was assumed that the complex situations in which human beings found themselves could be split up into simple states of the environment which could be treated as values of variables. Responses too were partitioned in a similar manner. Neither the role of the actor's interpretation of the stimulus conditions nor the actor's intention in responding were admitted as relevant, or as competing with the interpretations imposed by experiments (Milgram 1974). By an elementary application of Mill's Canons of Induction, that an effect which was always found to follow a certain phenomenon and which, in the absence of the phenomenon, did not appear, statistical trends in the relationship between stimuli and responses were offered as psychological laws. From the point of view of the social sciences the most important and most paradoxical applications of the experimental method were in social psychology.

The application of the methodological part of the behaviourist paradigm in social psychology led to the setting up of a number of experimental programmes centred around a variety of human social phenomena (Lindzey and Aronson 1968). For example the investigation of interpersonal liking (Zajonc 1968); of the fact that when they were with groups people tended to follow majority opinion (Asch 1956); of the conditions under which people behaved aggressively (Berkowitz 1962), all used the 'experimental' method. Paradoxically this approach to social behaviour in which the role of the individual human actor as agent undertaking various projects, alone or with others, was swept away and replaced by the idea of an automaton reacting with exactly those responses to which it had been conditioned, was dominant in that home of the ethos of the individual, the USA. This poses a fascinating

problem for the sociology of science.

2 The critique of behaviourism

It was the acceptance of a Cartesian metaphysics of mind that had led to scepticism about the possibility of public knowledge of mental states and processes from which the original impetus to a behaviourist treatment of problems in the human sciences had come. The most influential criticism of behaviourism was directed not at its Cartesian metaphysics, but at the limitations that it imposed on the possibility of obtaining scientific knowledge of other minds. This line of criticism developed into the 'first cognitive revolution' initiated by J.S. Bruner (1973) and G.A. Miller (Miller and Johnson-Laird 1976). Adopting the hypothetico-deductive conception of scientific method the new cognitivists saw the experiment as playing the role of a test for a hypothesis about publicly unobservable mental processes, rather than as a datum to be used in an induction to a correlative law of stimulus and response. The 'cognitive science' movement developed out of the insights of Bruner and Miller by the adoption of the computer analogy, as a way of systematically formulating hypotheses about mental processes. The computer is to the brain as the running of a programme is to the mental activity of that brain.

A more realistic approach to the study of social psychology, that developed in the 1970s, led to a growing emphasis on language as the main medium of human interaction (Harré 1977). In the behaviourist and immediately post-behaviourist eras, language appeared only in various so-called 'instruments', such as questionnaires, rating scales and so on. Many 'experiments' consisted in asking people to read stories and then to answer questions about their opinions of these stories. Why not, it was asked (for example by Shotter 1993), just abandon misleading talk of experiments, and turn to an analytical study of the many forms of human communication?

3 The behaviourist legacy in sociology

In sociology something akin to behaviourism developed under the influence of the three key factors identified above in the origins of behaviourism in psychology - a positivist philosophy of science, a search for statistical correlations between publicly observable phenomena and a 'regularity' conception of causality. The most influential proponent of positivistic sociology was Émile Durkheim. In his famous work on suicide (1908) he correlated local suicide rates with the values of broad social variables. Durkheim also offered to posterity another, non-statistical sociological method, in which publicly observable social phenomena, such as religious ceremonies, were subject to hermeneutical reinterpretation as covert symbolizations of social structure. The history of the debates over sociological method could be presented as a dispute between Durkheim and his *alter ego*.

Ethnomethodology was developed by Harold Garfinkel and his co-workers explicitly in opposition to Durkheimian social fact methodology. Garfinkel (1967) looked at the way social facts, such as suicide statistics, were produced. He found that they were the product of very complex conversational interactions, in the course of which people with various rights and standings took part in a complex and often quite long-running negotiation. This led Garfinkel to a very close analysis of the (conversational) procedures by which everyday reality was constituted as something normal and unattended. At this point we are a long way from positivism. However by a strange shift in focus the very method by which Garfinkel proposed to unveil the methods by which ordinary folk created their worlds turned into a kind of positivistic empiricism. In CA ('conversation analysis'), elementary conversational acts (as identified by the analyst) are sorted into statistical patterns, independent of all contextual features whatever. This development has been subjected to much the same kind of criticism as was classical behaviourism, namely that as a positivistic reduction of the phenomena that define a field of interest of a science it eliminates the thing that it should be attempting to discover by means of a context sensitive methodology, namely the cultural conventions by reference to which the micro-orderliness of social life is maintained.

4 Summary

The behaviourist paradigm has left its traces everywhere in the human sciences and particularly in social psychology. Despite paying lip-service to the idea that human beings in social interaction are active agents engaged in jointly realizing certain projects, plans and intentions and sometimes antagonistically attempting to force their projects on others, we have a view of human social actors as mindless automatons, mere spectators of processes over which they, as individuals, could have no control. This metaphysical position is still implicit in much of the work of US social psychologists (Wrightsman and Dew 1980) and to a much lesser extent in Europe

(Farr and Moscovici 1987). The metaphysical foundations of microsociology and social psychology have been strongly criticized. But to treat formalized discussions between psychologists and their subjects as experiments is, according to the critics, to do little more than offer a rhetorical re-description of something which is utterly unlike an experiment in the natural sciences. Instead it is argued that an acknowledgement of the fact that what are treated as experiments are really conversational interactions and should be analysed as such, is the way forward (Potter and Wetherell 1987). All that a psychological experiment can do, if seen in its true light as a kind of conversation, is to enable us to discover the local conventions of social discourse. Once this is achieved we can ask whether some such conventions can be identified in all human communities of which we have any knowledge.

See also: Behaviourism, analytic; Behaviourism, methodological and scientific; Positivism in the social sciences; Skinner, B.F.

ROM HARRÉ

References and further reading

Asch, S. (1956) ‘Studies of Independence and Conformity’, *Psychological monographs* 70: 416.(Study of the fact that when in groups people tend to follow the majority opinion.)

Berkowitz, L. (1962) *Aggression: A Social Psychological Analysis*, New York: McGraw-Hill.(Study of the conditions under which people behave aggressively.)

Bruner, J.S. (1973) *Beyond the Information Given: Studies in the Psychology of Knowing*, New York: W.W. Norton.(Initiated, with G.A. Miller, the ‘first cognitive revolution’.)

Danziger, K. (1990) *Constructing the Subject*, Cambridge: Cambridge University Press.(Selective history of modern psychology that sees psychology, not so much as a body of facts or theories, but as a special set of social activities intended to produce something that counts as psychological knowledge under certain historical conditions.)

Durkheim, É. (1908) *Suicide: A Study in Sociology*, trans. J.A. Spaulding and G. Simpson, London: Routledge & Kegan Paul.(Durkheim’s famous work in which he correlated local suicide rates with the values of broad social variables.)

Farr, R. and Moscovici, S. (1987) *Social Representations*, Cambridge: Cambridge University Press.(European work on the metaphysical foundations of social psychology.)

Festinger, L. (1957) *A Theory of Cognitive Dissonance*, Stanford, CA: Stanford University Press.(Cognitive dissonance theory based upon the alleged existence of cognitive discomfort in the presence of contradiction between expressed beliefs and actions.)

Garfinkel, H. (1967) *Studies in Ethnometholodogy*, Englewood Cliffs, NJ: Prentice Hall.(Develops ethnomethodology in opposition to Durkheimian social fact methodology.)

Harré, R. (1977) *Social Being*, Oxford: Blackwell, 2nd edn, 1993.(Emphasizes language as the main medium of human interaction.)

Harré, R. and Secord, P.F. (1972) *The Explanation of Social Behaviour*, Oxford: Blackwell.(Develops the ‘role-rule’ model which recommends a methodology of account analysis instead of simple experimentation.)

Lindzey, G. and Aronson, E. (1968) *Handbook of Social Psychology*, Reading, MA: Addison-Wesley.(Looks at the setting up of a number of experimental programmes centred around a variety of human social phenomena.)

Milgram, S. (1974) *Obedience to Authority*, New York: Harper & Row.(Study of experiments in social psychology and their interpretations.)

Miller, G.A. and Johnson-Laird, P.N. (1976) *Language and Perception*, Cambridge: Cambridge University Press.(Initiated, with J.S. Bruner, the ‘first cognitive revolution’.)

Potter, J. and Wetherell, M. (1987) *Discourse and Social Psychology*, London: Sage.(Argues that what are treated as experiments in social psychology are really conversational interactions and should be analysed as such.)

Shotter, J. (1993) *Conversational Realities: Constructing Life Through Language*, London: Sage.(Claims that naturally occurring psychological and sociological ‘realities’ are both socially constructed and sustained within everyday conversation.)

Tajfel, H. (1982) *Social Identity and Intergroup Relativism*, Cambridge: Cambridge University Press.(Theory of intergroup relations that proposes a cognitive mechanism of social comparison to explain observed correlations.)

Wrightsman, L.S. and Dew, K. (1980) *Social Psychology in the Eighties*, Monterey, CA: Brooks Cole. (Integrated introduction to social psychology.)

Zajonc, R.B. (1968) ‘Attitudinal Effects of Mere Exposure’, *Journal of Personality and Social Psychology*,

monograph supplement, 9 (2): 1-27.(Investigation of interpersonal liking.)

Behaviourism, analytic

Analytical behaviourism is the doctrine that talk about mental phenomena is really talk about behaviour, or tendencies to behave. For an analytical behaviourist, to say that Janet desires ice cream is to say that, all things being equal, she tends to seek it out. To say that Brad is now feeling jealous is to say no more than that he is now behaving in a way characteristic of jealousy, or perhaps that he would do so under appropriate provocation. Analytical behaviourism differs from methodological behaviourism in insisting that our ordinary use of mental language really is, in some sense, already about behaviour. The methodological version claims either that in doing psychology we should restrict ourselves to notions which can be defined behaviourally, or, sometimes, that our general psychological language, even if not already definable in this way, should be reformed in this general direction.

The most telling objection to this account of the mind is that it is inconsistent with the requirement that mental states are causes of behaviour. Ordinarily we might note that Brad has a tendency to display jealous behaviour with little provocation, and conjecture that this is caused by his feeling jealous (rather than, say, practising for his forthcoming part in a Jacobean tragedy). But according to analytical behaviourism his feeling jealous just is his tendency to the behaviour, and since nothing causes itself, his jealousy cannot be the cause of the pattern of behaviour.

1 Introduction

Analytical behaviourists hold that there is a conceptual connection between behaviour and mental states. Talk about mental states just is talk about behaviour and dispositions to behaviour. This can be discovered, the doctrine insists, by proper conceptual analysis of the mental language we use. In its strongest versions the idea is that there can be translations between mental language and behavioural language; for example they might have held that ‘Janet desires ice cream’ - apparently postulating something about her internal states - can be translated without loss of meaning as ‘Janet is likely to eat ice cream’. Sometimes analytical behaviourists admitted that the translations might be difficult or impossible to provide, but contended that there was, nevertheless, some unspecified sort of analytical entailment between mental talk and behavioural talk. Another issue, essentially terminological, is whether the analytical behaviourist claims to eliminate mental states. On one way of saying it, the behaviourist discovers that there are no mental states as such - but mental language is perfectly fine because it is about behaviour. The other way has it that mental states exist but are not internal, occult entities; rather they are to be identified with behaviour or behavioural dispositions. I will use the latter way throughout.

I will start with a few words about the doctrine’s roots, and why it seemed so appealing, and in addition explain the usual arguments for it. Then I will discuss the standard objections and assess the important contribution that the doctrine has made to the philosophical treatment of the mind.

2 Influences and arguments

A first, important influence on analytical behaviourism was the doctrine of verificationism (see Meaning and verification; Operationalism). According to one version of this view, the meanings of statements are given by the procedures used to find out if they are true. This was supposed to eliminate meanings as occult entities, and put the subject on a scientific footing. Anyone attracted by that view would have seen considerable merit in behaviourism. For we typically verify statements about other people’s mental states by looking at their current behaviour (including verbal behaviour), or their behaviour over time, drawing conclusions about the dispositional patterns that emerge.

A related influence was that of the success of methodological behaviourism (sometimes known as scientific behaviourism). This somewhat weaker doctrine stipulates that for the purposes of psychology we use terms for mental states which are defined in terms of behaviour (see Behaviourism, methodological and scientific). This was perhaps good advice in the early part of the century; in any case behavioural evidence was the only access then available to psychological information, and the advice yielded reasonable results. The dominance of this idea left philosophers in a difficult position. What scientific psychology talked about was defined in terms of behaviour. The denial of analytical behaviourism would therefore entail that ordinary discourse about the mind is simply not

about the same things as scientific discourse - bad news, indeed, for scientifically-minded philosophers. Rather than risk accepting that consequence, it seemed better to insist that everyone really was talking about the same thing, and thus that analytical behaviourism was true.

Another approach to this problem can be found among the methodological behaviourists. We might call it revisionary behaviourism - the view that, to the extent that ordinary talk is not about behaviour, it is defective and should be revised in line with scientific talk. This was never popular amongst philosophers, perhaps because of its consequence that all ordinary talk about psychology before behaviourism was seen as deeply defective. It is hard to believe that Aristotle was not saying something true when he attributed beliefs to Plato.

A crucial influence, especially in the British tradition, was Wittgenstein's scepticism about the possibility of private states contributing to the meaning of expressions in public language (see Private states and language). If you think that private states can have no role in fixing meaning, then behaviour is the only game in town.

The most attractive thing about analytical behaviourism is that it undoubtedly got some things right. While it might be strange to think that the meaning of mental language is exhausted by talk of behaviour, there surely is some kind of analytic link between behaviour and mental states. If you don't know that someone who desires something will, other things being equal, behave so as to seek it out, then something is wrong with your concept of a desire. If I believe that Janet truly does desire ice cream, and that her desire isn't overruled by another, then there is something deficient in my concept of desire if I am not remotely puzzled when she blithely ignores a convenient ice cream stall.

Behaviour can trump introspective evidence; we are familiar with people who sincerely claim to have desires which they make no attempt to fulfil, even when they claim to have no countervailing desires and believe that the means to their fulfilment are available to them. In such cases you might say that the person doesn't *really* have the desires they say they have, whatever they believe about their desires. If we are tempted by that idea, it might be because it follows from what we mean by desire: that someone who truly desires something behaves so as to fulfil that desire.

3 Objections

There are very few analytical behaviourists left. A range of objections has culled their number over the years, and has led to newer views in the philosophy of mind. I will discuss objections - again not entirely independent ones - in order of increasing severity.

A first objection is that some mental states could have no behavioural consequences. Simple examples of this may merely show that sometimes one mental state outranks another. Suppose that I love ice cream, but am so embarrassed by this that I resolve never to touch it, and succeed in never manifesting in any way whatsoever the difficulty this causes me. Have we not a counterexample to behaviourism?

The behaviourist has two options. One is to say that if this alleged desire has no impact at all - say in exciting my salivary glands when I see photographs of cool, creamy ice cream, and so on - then I indeed do not have any such desire. Another is to point out that there is a case of a resolved conflict of desires here: what I desire over all - to avoid embarrassment - is manifested in behaviour; and the analytic connection between all-things-considered desires and behaviour is maintained.

A stronger version of this objection involves mental states the expression of which is somehow blocked. Perhaps someone could be in pain but, because they are controlled in various ways, cannot express it. A disembodied brain might be an extreme case. Here the analytical behaviourist must resort to the disposition's being there, but being blocked. The dispositions are dispositions to behave in normal circumstances. Although one might worry about how to specify these normal circumstances, these sorts of objections do not seem too troubling.

Behaviourists have more trouble than most in accounting for consciousness. Whilst they can without conceptual difficulty accept that individuals know more about their own mental states than those of others - analysed as having behavioural dispositions which are themselves foreknowledge of their future behaviour - they are hard pressed to explain why this should be so, since the usual behavioural evidence is not available to them.

There is, moreover, no real account of the *nature* of consciousness. Perhaps no one has a satisfactory account, but

behaviourists are by their very own doctrine committed to refusing to countenance internal mental states. This leaves them with three options. They can accept conscious states but say they are not mental - but this is a terminological trick that leaves them with an ontology far from the spirit of behaviourism. They can deny that consciousness exists at all - which might be philosophically comforting for some but rather hard to do while keeping a straight face. Finally they can identify consciousness with behaviour or behavioural dispositions. The second, po-faced, option seems to have been most common; none is without difficulty.

The promise of behaviourism was that it would provide an analysis of the meaning of mental language in terms of behaviour. It rapidly became obvious that actually providing such an analysis was not easy: even a flawed one, and even for the simplest cases. The formulation that became popular was that mental states pick out a vague class of behaviour, and this is why the analysis cannot be done in practice. So mental language cannot be dispensed with, since the vagueness is useful to us in tying together the groups of behaviour-patterns, and is ineliminable. This mental language, nevertheless, has nothing but behaviour and behavioural dispositions to be about.

There was, however, a deeper reason for the difficulty than just vagueness. The problem is that there simply is no class of behaviour-patterns associated with individual mental states. Suppose I believe there is a dagger before me. Do I act as if nothing is there, thinking weapons in very poor taste and best ignored? Do I grab it, believing that the gods have put it there to assist in fulfilling my desire to take revenge on my enemies? Do I book in for analysis, thinking this is a hallucination brought on by feelings of remorse and guilt for what happened at a recent departmental committee meeting? Clearly it depends on my other beliefs and desires. This is very different from the sorts of worries mentioned above when we were considering ways in which the normal responses allegedly associated with mental states might be blocked. Here we see that there are, in fact, no normal responses associated with individual mental states.

So talk of 'vague' connections between individual mental states and behaviours misses the point. The point is that, to the extent that there are analytic connections, they are between complexes of mental states and complexes of behaviour.

The final objection, and perhaps the most telling, is that analytical behaviourists must deny that mental states ever cause behaviour. It is central to the idea of a mental state that they sometimes are the cause of our actions. Janet's desire for ice cream actually causes Janet's arm to go up and her legs to move her body towards the ice cream van. For the analytical behaviourist, however, desire for ice cream never *causes* ice cream-seeking behaviour, even when all the attendant beliefs and desires are in place. Since analytical behaviourists are typically physicalists, they admit that there are internal causes of behaviour, but say that these are not mental states. The mental states are the very behaviour caused by the internal states. Further, the behaviour cannot cause itself. The case of dispositions is even harder. The analytical behaviourist has to come up with something to say about the relationship between internal states and the behavioural dispositions. But the causes of the dispositions cannot be the dispositions themselves. So, to the extent that we think that behavioural tendencies like Brad's tendency to jealous behaviour are caused by some mental state rather than another, we cannot accept analytical behaviourism.

While analytical behaviourists were on to something in seeing that there were analytic connections between behaviour and the mental, they drew too strong a connection. What is analytic, they thought, cannot be contingent. But whatever causes Janet's legs to move is some physical system of neurones, nerves and muscles. And it is contingent that that system of neurones, nerves and muscles does the work. She could have had different neurones, used different muscles, or perhaps even had a prosthetic brain. The connection between the behaviour and the mental language, however, is necessary; it is impossible that she, Janet, could desire ice cream without having the appropriate dispositions. The connection is also graspable, in some sense, just through understanding the meaning of the mentalistic language involved.

What has gone wrong here is that the kind of analytic connection they are looking for was too direct. We are now more familiar with analysis proceeding indirectly via how some contingent matter actually is. Imagine a town that has had its water poisoned. The townsfolk coin the phrase 'death potion' to name whatever it is that is actually doing the killing. 'Death potion' is whatever is actually having a certain causal influence on the townsfolk; killing them, in this instance. As it happens, the poison is cyanide. Eventually the townsfolk find this out. They find out that 'death potion' is cyanide; and analysis played a key role in this, via the meaning of 'death potion'. Nevertheless it is quite contingent that the poisoners used cyanide - they could easily have chosen something else.

4 The legacy of behaviourism

The last objection to analytical behaviourism is the starting point for the families of views that have come to replace it. Once we see that analysis gives us indirect linguistic access to unknown internal things - such as whatever it is inside people's bodies that is killing them off - we can accept what is right about analytical behaviourism and yet embrace internal states.

In the example above, the analytic connection was between 'death potion' and whatever it was that (typically) played a certain causal role (the role of causing the sudden spate of deaths). So we sometimes use causal properties to identify states inside systems we do not know much about. It is a move like this that can meet the requirement that mental states cause behaviour while respecting the analytic insight. Janet's desire for ice-cream might be understood not as the ice cream seeking behaviour, but as whatever causes the behaviour. Instead of mental states being behaviour or clusters of behaviour, they might be analysed as states which are typically caused in various ways, and which typically cause the clusters of behaviour. Views of this kind are called functionalism (see Functionalism).

See also: Ryle, G.

DAVID BRADDON-MITCHELL

References and further reading

Armstrong, D.M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.(First book-length treatment which combines the grains of truth in the behaviourist analysis with an analysis of mental states in causal terms.)

Block, N. (1981) 'Psychologism and Behaviourism', *The Philosophical Review* 90: 5-43.(Advances a range of objections to tying mental states too closely to behaviour.)

Campbell, K. (1984) *Body and Mind*, Notre Dame, IN: University of Notre Dame Press: 2nd edn.(Chapter 4 of this book contains a very accessible discussion of behaviourism.)

Geach, P. (1957) *Mental Acts*, London: Routledge & Kegan Paul.(Contains an early articulation of something like the concerns expressed in §3 about the difficulties of analysing mental discourse in terms of behaviour.)

Putnam, H. (1963) 'Brains and Behaviour', in *Mind, Language and Reality: Philosophical Papers Volume 2*, Cambridge: Cambridge University Press.(An important paper, which advances arguments of the type discussed in §3, to the effect that some mental states have no behavioural consequences, and presaged the development of views of the type discussed in §4.)

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.(The central exposition of analytical behaviourism.)

Strawson, P.F. (1959) *Individuals*, London: Methuen.(An influential piece of analytical philosophy in the era of behaviourism. It includes attempts to make sense of the unsatisfactory 'vagueness' account of the difficulty of providing analyses of mental discourse in terms of behaviour.)

Skinner, B.F. (1974) *About Behaviourism*, London: Jonathan Cape.(A full-scale defence of behaviourism. Although he defends methodological behaviourism, it is clear that his view was of the reforming variety: to the extent to which analytic behaviourism is not true of ordinary discourse, ordinary discourse is deficient and should be revised.)

Watson, J.B. (1925) *Behaviourism*, London: Kegan Paul.(The *locus classicus* of a behaviourism of a methodological kind; but like Skinner after him, he believed that if ordinary talk was not about behaviour, it should be revised in line with scentific talk.)

Wittgenstein, L. (1953) *Philosophical Investigations*, trans. G.E.M. Anscombe, Oxford: Blackwell.(The source of Wittgenstein's influence on behaviourists.)

Behaviourism, methodological and scientific

Methodological *behaviourism is the doctrine that the data on which a psychological science must rest are behavioural data - or, at the very least, publicly observable data - not the private data provided to introspection by the contents of an observer's consciousness.* Scientific*, or, as it was sometimes called, 'radical', behaviourism contends that scientific psychology ought to be concerned only with the formulation of laws relating observables such as stimuli and responses; not with unobservable mental processes and mechanisms such as attention, intention, memory and motivation. Methodological behaviourism is all but universally embraced by contemporary experimental psychologists, whereas scientific behaviourism is widely viewed as a doctrine in decline. Both forms of behaviourism were articulated by J.B. Watson in 1913. B.F. Skinner was the most prominent radical behaviourist.*

In addition to its empiricist strictures against inferred mental mechanisms, radical behaviourism was also empiricist in its assumptions about learning, assuming that: (1) organisms have no innate principles that guide their learning; (2) learning is the result of a general-purpose process, not of a collection of mechanisms tailored to the demands of different kinds of problems; and (3) learning is a change in the relation between responses and the stimuli that control or elicit them. Many of these ideas continue to be influential, for example, in connectionism.

1 Methodological behaviourism and scientific behaviourism

Methodological behaviourism emerged as a reaction to introspectionism, which was the doctrine that psychology was the science of conscious states (see Introspection, psychology of §4). This latter doctrine placed the psychological scientist in a uniquely favourable epistemological position, because knowledge of the contents of consciousness seems direct, unmediated by sensory systems of questionable trustworthiness. However, it had two fatal weaknesses. (1) Because the data (the contents of consciousness) were not publicly observable, there was no way of settling disputes about whether the observer was reporting them accurately. (2) Because there was no way of securing reports from animals, psychology was confined to the study of human consciousness, a profoundly abiological state of affairs. These difficulties were obviated by the emergence of the doctrine that the data of psychology were behavioural. If you doubted the accuracy of someone's report about how their subjects behaved, you could come to their laboratory and see for yourself. And the behaviour of many nonhuman animals was as observable as human behaviour, so it was possible to approach psychology from an evolutionary perspective, dispensing with the pre-Darwinian notion that man was uniquely different from a scientific standpoint.

The doctrine that psychological theory must rest on behavioural (or, in any event, publicly observable) data in no way implies that psychological theory should eschew discussion of inferred mental processes and mechanisms, which was the essence of *scientific* behaviourism. If there were any such implication, then physics could not concern itself with such inferred entities as quarks; nor could classical genetics concern itself with genes, and so on. Cognitive psychology is thoroughly behaviourist in its methodology, but it is anti-behaviourist in its theorizing, which focuses on attention, perception and memory. Scientific or (as it was often called) 'radical' behaviourism is an eliminativist doctrine; it claims that traditional mental phenomena such as these have no place in a scientific psychology (see Eliminativism).

B.F. Skinner, the best-known radical behaviourist, argued that concepts such as memory, motivation, emotion, attention, intention and perception should not be part of scientific psychology. He argued that a scientific psychology should be concerned only with physical observables. This was the general position of logical positivism, which flourished for the first half of the twentieth century (see Logical positivism), but has vanished from contemporary physics, which routinely discusses unobservable entities such as quarks. Similarly, radical scientific behaviourism, which enjoyed a considerable vogue for several decades near the middle of the twentieth century, has diminished in significance with the rise of cognitive psychology.

2 The two schools of scientific behaviourism

At the peak of its popularity in the middle of the twentieth century, there were two principal schools of scientific behaviourism, one identified with J.B. Watson and B.F. Skinner, the other with Ivan Pavlov and Clark Hull. (The

latter are sometimes called neo-behaviourists.) Both schools were principally concerned with learning. Both took for granted the traditional empiricist assumption that the organism was initially a blank slate. That is, they assumed that learning was not constrained and directed by innate problem-specific information-acquisition mechanisms whose problem-specific structure made implicit a priori commitments to the structure of the information to be acquired. Such commitments, popular among Rationalists such as Descartes and Leibniz, involve innate principles of the mind (see Nativism), and behaviourists rejected the idea that innate principles guided learning. On the contrary, the learning process was thought to be a general-purpose process, applicable regardless of the substance of what was learned. The essence of the process was thought to be the bringing of a response the animal made in the presence of a stimulus under the control of that stimulus. This last assumption gave rise to the term ‘stimulus-response’ (or ‘S-R’) psychology.

The two schools had contrasting views on whether any kind of ‘theoretical construct’ was admissible. A theoretical construct was an entity (for example, a memory or an associative bond) or process (for example, attention) or state (for example, a motivational state) that played a role in explaining observed behaviour but was not directly observed (see Observation). The Skinnerian school thought that all such constructs should be rejected; even such traditional constructs as the associative bond. The Hull-Pavlov school, on the other hand, thought that a modest number of constructs were essential, but should be closely based on behavioural observation and should have plausible physiological embodiments. When a rat learned to press a lever to obtain food, Skinnerians explained this simply by an appeal to the ‘law of effect’, which says that when a response is emitted in the presence of a stimulus and followed by a reinforcer, the probability of that response in the presence of that stimulus increases. They thought this kind of Baconian explanation - subsuming observations under generalizations about lawful regularities - was all that psychology should aim for (see Bacon, F.). The Hullians, by contrast, aimed for a physiologically plausible conception of the unobserved mechanisms. They imagined that the food reduced a drive (defined as a state of high nervous arousal), which strengthened synaptic connections between sensory and motor neurons - the sensory neurons sensitive to the stimulus and the motor neurons that controlled the response in question. The drive and the change in synaptic connectivity (or associative strength) were the kinds of physiologically plausible theoretical constructs that Hullians thought should play a central role in psychological explanation. The proffered physiological interpretation of these constructs made it reasonable to suppose that one might someday actually observe them.

In short, the Skinnerians were opposed to inferred entities of any kind in psychological explanation, while Hullians wanted to restrict the inferred entities to those with a ready physiological interpretation. Both schools were sceptical of the explanatory value of the mental concepts and mechanisms that have traditionally played a central role in more cognitively oriented psychological theories - attention, perception, intention, memory, motivation (see Explanation).

Behaviourist ideas continue to be influential in contemporary psychology. The assumption of a general-purpose learning process, the assumption that the building blocks of psychological theories should be neurobiologically plausible, and the assumption that learning is fundamentally about changes in synaptic connectivity (associative strengths) are still points of departure for much contemporary work in psychology - both work on animal learning and work on human learning and perception done under the heading of connectionist or neural-net modelling (see Connectionism; Learning §1).

3 Critique of scientific behaviourism

The assumptions of scientific behaviourism may be challenged on a number of fronts. As a perfectly general proposition, the blank-slate assumption is demonstrably false. Fledglings of some species of song bird (for example, white-crowned sparrows) must hear an adult male song of their own species during a critical period if they are to develop normal song. It does not matter what other songs they hear. Even when those other songs are more numerous than that of their own species, the song-learning mechanism picks out the correct song, implying that it is innately tuned to aspects of the song to be learned that distinguish that song from those of other species. The mechanism that mediates song learning has been termed an instinct to learn. Like all instincts, it is a solution to one particular problem, obviously unsuited to many other learning problems. Some contemporary learning theorists argue that the problem-specific structure seen in this and many other examples from the zoological literature is characteristic of all learning mechanisms. Thus, it is open to question whether there is such a thing as a

general-purpose learning process (see Innate knowledge; Language, innateness of).

Also demonstrably false is the assumption that an animal must respond during learning. The juvenile white-crowned sparrow is incapable of singing during the critical learning period; all it does is listen. Similarly, nestling buntings, when they are too young to fly, learn the centre of rotation of the night sky by watching the stars rotate about it. Months later, they use this knowledge to keep oriented during the night-time portions of their migratory flights. These examples of purely passive observational learning cannot be conceptualized in S-R terms. Ironically, the phenomenon most studied by Skinnerians has proven to be passive in just this way. To get the pigeon to peck the key, it suffices to arrange for it to observe that soon after the key is illuminated the food hopper opens. If it observes this sequence several times, it begins to peck the key, even if its pecking has no effect on food delivery; indeed, even if its pecking forestalls food delivery! Most learning depends on the perception over time of a systematic stimulus relationship; whatever response the animal may or may not make during the period when the relationship is perceived is of little relevance to the learning that occurs.

There are also many examples of learning that cannot readily be conceptualized as the formation of an association. One example is the learning of one's position by 'dead reckoning', which has been found in animals ranging from insects to humans. Dead reckoning is the integration of one's velocity with respect to time to obtain net displacement: computing where you are from how fast you have moved in which direction for how long. This method of determining where you are relative to where you started is so unlike the traditional conception of learning by association that people when first hearing of it often reject it as an instance of learning. However, if the pre-theoretical definition of learning is the process or processes by which we acquire knowledge from experience, then dead reckoning is demonstrably a principal means by which mobile animals acquire knowledge of their current position in the world from their experience of the direction and rate at which they have moved through it. This example points towards learning as computationally derived knowledge of the world rather than as associated experiences (see Learning §2).

Finally, the desirability of constraining the theoretical constructs of psychology to be neurobiologically plausible is debatable. Constraining psychological theory in this way presupposes that neurobiologists have *already* identified those aspects of neural functioning that are relevant to understanding processes such as memory, attention, and thought-processes whose existence we infer from their behavioural manifestations. If the aspects of cellular and molecular neurobiology that are centrally relevant to our understanding of these processes remain to be discovered or identified, then attempts to ground psychological theorizing in well-established supposedly relevant neurobiological mechanisms are certain to lead us astray in that they make us build on what will turn out to be the wrong neurobiological foundations. For the first several decades of the twentieth century, classical geneticists did not attempt to ground their genetic theorizing in biochemistry. It is a good thing that they did not, because the discovery of the biochemical identity of the gene in 1954 revolutionized biochemistry. The biochemistry of the gene in the 1990s rests on concepts and mechanisms undreamed of in the biochemistry of the 1940s.

See also: Cognitive development

C.R. GALLISTEL

References and further reading

Blumberg, M.S. and Wasserman, E.A. (1995) 'Animal Mind and the Argument from Design', *American Psychologist* 50: 133-44.(A recent discussion of the continuing relevance of behaviourism in psychology.)

Boring, E.G. (1950) *A History of Experimental Psychology*, Englewood Cliffs, NJ: Prentice Hall.(Important early history of behaviourism.)

Catania, A.C. (1992) 'B.F. Skinner, Organism', *American Psychologist* 49: 1,521-30.(An appreciation of Skinner and his thought.)

Chomsky, N. (1959) 'Review of *Verbal Behavior* (by B.F. Skinner)', *Language* 35: 26-58.(An influential critique of behaviourism.)

Dennett, D. (1978) 'Skinner Skinned', in *Brainstorms*, Cambridge, MA: MIT Press.(An engaging discussion of some of the important philosophical issues surrounding behaviourism.)

Fodor, J. (1968) *Psychological Explanation*, New York: Random House.(A philosophical critique of behaviourism.)

Gallistel, C.R. (1990) *The Organization of Learning*, Cambridge, MA: Bradford Books/MIT Press.(A cognitive

analysis of learning, the antithesis of the behaviourist analysis.)
Kendler, H.H. (1987) *Historical Foundations of Modern Psychology*, Philadelphia, PA: Temple University Press. (A good historical discussion of behaviourism.)
O'Donnell, M. (1985) *The Origins of Behaviorism: American Psychology 1870-1920*, New York: New York University Press.(Discussions of the background and consequences of Watson's formulation.)
Schultz, D.P. and Schultz, S.E. (1992) *A History of Modern Psychology*, Fort Worth, TX: Harcourt Brace Jovanovich, 5th edn.(A good historical discussion of behaviourism.)
Skinner, B.F. (1990) 'Can Psychology be a Science of Mind?', *American Psychologist* 45: 1,206-10.(Skinner's view of cognitive psychology.)
Watson, J.B. (1913) 'Psychology as the Behaviorist Views It', *Psychological Review* 20: 158-77.(Historically the most important statement of both scientific and methodological behaviourism.)
Zuriff, G.E. (1985) *Behaviorism: A Conceptual Reconstruction*, New York: Columbia University Press.(A philosophically oriented discussion of both aspects of behaviourism.)

Being

Although 'being' has frequently been treated as a name for a property or special sort of entity, it is generally recognized that it is neither. Therefore, questions concerning being should not be understood as asking about the nature of some object or the character of some property. Rather, such questions raise a variety of problems concerning which sorts of entities there are, what one is saying when one says that some entity is, and the necessary conditions on thinking of an entity as something which is.

At least four distinct questions concerning being have emerged in the history of philosophy: (1) Which things are there? (2) What is it to be? (3) Is it ever appropriate to treat 'is' as a predicate, and, if not, how should it be understood? (alternatively, is existence a property?) (4) How is it possible to intend that something is? Twentieth-century discussions of being in the analytic tradition have focused on the first and third questions. Work in the German tradition, especially that of Martin Heidegger, has emphasized the fourth.

1 Which things are there?

Which sorts of things are there? This question asks for an inventory of the population of the world. It is natural to assume that there are middle-sized physical objects, such as organisms and artefacts, but reflection reveals a variety of puzzles. Given that middle-sized objects are composed of simpler material, should our inventory include only the objects, or only the material, or both? Given that objects have properties, and that different objects can have the same property, should our list include those properties, or merely the objects which have them? What about events, or numbers, or times?

It is not immediately clear what kinds of consideration are relevant to answering the population question. How does one decide just which entities there are? One possibility is to think of the question 'What things are there?' as analogous to 'What things are humans?' Ordinarily we think that we can answer the question regarding the population of humans if we know what conditions must be met for a thing to count as human. So, perhaps, we can know which things there are if we know what conditions must be met for a thing to count as something which is. But to know this is to know what it is we are saying when we say that something is. It is to have an answer to the question which Aristotle says 'was raised of old and is raised now and always, and is always the subject of doubt, viz. what being is'.

2 What is it to be?

The form of the question 'What is being?' initially suggests that we should think of 'being' either as a name for an entity or as picking out some property, but both views have proved problematic. When one asks a question of the form 'What is Patagonia?' one is requesting information regarding which entity is picked out by the word 'Patagonia'. So it is perhaps natural to think that the question 'What is being?' is similarly raising an issue regarding the object which is referred to by the word 'being'. But, unlike 'Patagonia', 'being' is not a noun: it is the present participle of the verb 'to be'. This suggests a different analogy. When one asks 'What is running?' (and is not asking which thing is running) one is asking about the character a thing must have in order to count as running. By analogy one might expect that the question 'What is being?' raises the issue of which character or property being is.

The notion that 'being' names an entity was the first possibility explored. Parmenides reaches the extraordinary conclusion that there is only one unchanging item in the world inventory - being. He argues in two ways. First, anything which is not being, or is different from being, is not, since it *is not being*, so at most one being is, and that being is being itself. Parmenides is here treating 'being' as a name for a special sort of entity, call it 'being itself', and suggesting that it is correct to say that A *is* just in case A is identical with being itself. Parmenides' second line of argument assumes that all words, including 'being', are meaningful just in case they stand for some thing which is. He suggests that it is impossible truly to say or think that something, A, is not, for if A is not then 'A' stands for nothing. This has suggested to some that words which apparently refer to non-existent entities really do refer to entities which are, even though they do not exist.

Ancient philosophers found Parmenides' conclusions troubling. In response, Plato and Aristotle clarified the distinction between properties and the entities which possess those properties, and recognized that an entity can

have some property without being identical to that property. So the fact that A is, does not imply that A is identical to is (or being), and the fact that 'is' is meaningful need not imply that 'is' is the name of an entity. In particular, Aristotle held that those entities which are in the primary sense are those self-subsistent individuals which have properties, the substances. And whatever else 'being' might signify, it does not stand for a self-subsistent individual substance, being-itself (see Aristotle §12).

These distinctions leave open the possibility that 'A is' should be understood as asserting that A has a property, being. And there are strains in Aristotle's thought which attracted the tradition he established to this possibility, as well as strains which undercut this option. On the one hand, Aristotle thinks that it is wrong to see 'being' as picking out a class of thing, as 'human' does. In the 'human' case we assume that we already have a list of the things which are; we ask merely which characteristics distinguish the humans on that list from the nonhumans. For 'human' to serve this function it must be possible for some items on the list to qualify as humans and possible for some to fail to qualify. In the case of 'being' it is precisely the contents of that initial list which are in question, and 'being' cannot differentiate among the items on the list, as all of those items *must* qualify as things which are.

On the other hand, Aristotle hoped to discover which entities are substances by examining those predications in which we specify *what* a thing is, as opposed to merely specifying that some individual has some property. But it is sometimes possible to specify what a thing is without committing oneself to the claim *that* it is: 'Hamlet is a man'. This suggested to later philosophers that a name, such as Hamlet, stands for a possible entity, and that those possible entities which exist have an additional property, actual being, or existence. In that case, the 'is' in 'A is' would serve to distinguish two classes of entities - the actual and the merely possible.

3 Is existence a property?

It is possible to understand what a thing is without asserting that it exists. Kant uses this fact to criticize the view that existence is a property (see Kant, I. §2). Consider a disagreement concerning whether or not some entity, say God, exists. Even though there may be disagreement concerning details, the two disputants agree on a general description of God as omnipotent, omniscient and so on. This description constitutes the concept of God. If there were not such agreement concerning the concept there would be no real dispute. The disagreement concerns only whether there is something which has those properties which, if someone had them, they would be God. Now, if existence were a genuine property there would not be agreement on the concept of God. For in that case the believer's concept of God would differ from the atheist's concept of God in so far as the first would include the property of existence and the second would not. So existence cannot be part of the concept of God, it is not part of a description of the entity which is God, or any other entity, and is no property.

If existence is not a property, there must be something misleading about the grammatical form of the sentence 'God is'. The surface grammar suggests that the function of 'is' is to assert the property of existence of God. Kant suggests we handle this problem by reading 'God is' as 'Something existing is God'. This has the effect of removing 'is' from the predicate position and thus removing the appearance that existence is a property. It also moves 'God' from the subject to the predicate of the sentence, which has further consequences for our understanding of 'being'. Parmenides had argued that words are meaningful just in case they stand for something which is. And many had argued that entities which are merely possible are, even if they are not actual. But if apparently referring expressions such as 'God' are really parts of the predicate, they can be treated as descriptions, rather than as names for things that are, possible or otherwise.

These suggestions have been systematically developed in the twentieth century by Russell and Quine. Quine's work in particular has important implications for reflections on being. Since antiquity it has been thought that the answer to the population question depends upon the answer to the question regarding what it is to be. Just as A being human involves A satisfying some description which specifies what it is to be human, for A to exist involves A satisfying some description. But if the existential use of 'to be' expresses no property, this model cannot be right. In the assertion 'A is', A can be seen as having descriptive content, while the existential assertion says that there is *something* which satisfies *this* description. Since 'something' serves as a pronoun, this amounts to saying that 'to be is to be in the range of reference of a pronoun', or, 'to be is to be the value of a variable'. The sorts of things we say there are are the sorts of things which can satisfy our descriptions. If we say that there are primes greater than five, we are saying that there is something such that it is a prime and it is greater than five. Since only numbers can be prime, this sentence commits us to the existence of numbers. And numbers *are* just in case some

sentences of this type *are true*. Our ordinary assertions commit us to the existence of various types of entities. To know which types of entities exist we need to know only which of those ordinary assertions are true.

Quine's work also has important implications for the relationship between possible and actual existence. Since names need carry no ontological commitment, asserting that some entity is not does not imply that there is some sense in which that entity, though not actual, is. Quine (1953) uses this to argue that there is no warrant for 'expanding our universe to include so-called possible entities' (see Ontological commitment).

Several philosophers have criticized aspects of Quine's views regarding being. David Lewis (1986) has argued that Quine's own criterion of ontological commitment commits us to the existence of possible entities, and Alvin Plantinga (1974) and Peter van Inwagen (1993) have argued that necessary existence, at least, is a property. These developments point to a reawakening of interest in the issues of what it is for something to be and how we should understand possible and actual being.

4 How is an understanding of being possible?

The twentieth-century philosopher Martin Heidegger attempted to renew interest in the question of being. For Heidegger 'being' is not a name for anything, and he carefully distinguishes among the various questions concerning being. But he follows a distinctive strategy in discussing these questions. He hopes to answer the population question by first answering the question of what it is to be, and to answer this question by answering the question of how it is possible to intend things that are as things that are. He thinks we can answer this question if we come to see how any understanding of being is possible.

Heidegger holds that the recognition that being is not a property leaves an unresolved problem: on what should we base decisions regarding which entities are candidates for descriptions? Quine thinks that this question must be answered pragmatically, holding that there are no general a priori conditions which restrict the range of reference of our pronouns. Heidegger disagrees, holding that there are such conditions which are grounded in necessary conditions for intending anything at all.

We can put Heidegger's strategy in context by returning to Kant. Kant thinks that the recognition that existence is not a property leaves a serious issue unresolved: What are the necessary conditions on the possibility of the objects of experience? What are the a priori constraints on the range of possible entities which we can come to know? He answers by appealing to the necessary conditions on possible experience. For Kant, our sentences can be true of a physical object only if that object is a possible object of experience, and it is possible to experience an object only if it is possible to perceive it. So the range of reference of our pronouns for physical objects is restricted by the condition of perceivability (and its extension in causality). ' The perception which supplies the content of the concept [of actuality] is the sole mark of actuality' (Kant 1781: A225/B273).

Heidegger thought that Kant's strategy had been developed by Edmund Husserl. Husserl asked how it is possible for a subject to intend something distinct from its own mental states. We intend a variety of such entities, including physical objects, numbers and norms, and it is puzzling how a subject can intend such independent entities. In order to solve this puzzle, Husserl generalized the notion of linguistic meaning. Each conscious act involves a meaning which is distinct from the act itself, and it is through this meaning that the act is related to its object. The meaning of an act depends upon its being placed within a temporal 'horizon' of past and possible future apprehensions. My current apprehension, for example, has the meaning of 'perception of a rock from this angle'. The act of rock perception has this meaning in virtue of its involving the retention of past apprehensions of the rock from different angles and the 'protention' that, were I to move, additional specific rock perceptions would occur. Intentions directed towards different classes of entities are characterized by different structures of relatedness among current apprehensions, retentions, and protentions, and what it is for an entity of some type to be varies as a function of the structure of the intentions directed towards entities of that type.

Heidegger has serious disagreements with Husserl regarding intentionality. In particular, he holds that intentionality is primarily a matter of action rather than consciousness. For Heidegger the primary way in which we intend a hammer as a tool, for example, is in an overt act of hammering, rather than in an internal conscious act, and the conditions on intending a hammer have more to do with the conditions under which an act can count as hammering than with how we think of a hammer (see Intentionality).

Nevertheless, Heidegger follows a strategy which is similar to Husserl's. The early Heidegger seizes upon the temporal character of Husserl's understanding of intentionality: ' intentionality … has the condition of its possibility in temporality and temporality's ecstatic horizontal character' (1975 (1982): 268). And, since one can understand what it is for a class of entities to be through determining the conditions necessary for intending that class of entities, Heidegger concludes that ' the central problematic of all ontology is rooted in the phenomenon of time' (1926 (1962): 40). Concretely, such an investigation involves fixing the range of reference for various classes of pronouns by determining the necessary temporal features of intentions directed at entities of that type. To intend a tool as a tool involves intending it as usable in various ways, and to be a tool involves serviceability. To intend a being as *Dasein*, or a being like us, on the other hand, is to intend it as itself capable of intentionality. Given the above, this implies that we are entities who are open to the dimensions of past and future, and Heidegger concludes that ' temporality [is] the meaning of the being which we call "Dasein"' (1926 (1962): 38).

Later, Heidegger came increasingly to concentrate on the contrast between the presence of an entity and the temporal horizon in which it becomes present. He concluded that philosophers have always answered the question regarding what it is to be with variations on the claim that being is constant presence, or being always available for interaction, but had failed to investigate the distinctive character of the horizon which makes an encounter with presence possible. Much of the work of the French philosopher Jacques Derrida, who has been heavily influenced by Heidegger, also focuses on the relationship between an object and the horizon in which it is encountered.

See also: Existence; Ontology

MARK OKRENT

References and further reading

Aristotle (*c*.mid 4th century BC) *Metaphysics*, trans. W.D. Ross in *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984.(Important discussions of being and substance appear in bks III, IV, and especially bks VII-IX.)

Heidegger, M. (1975) *Die Grundprobleme der Phänomenologie*, trans. A. Hofstadter as *The Basic Problems of Phenomenology*, Bloomington, IN: Indiana University Press, 1982.(Text of lecture course of 1927; the basic problems all concern being.)

Heidegger, M. (1926) *Sein und Zeit*, trans. J. Macquarrie and E. Robinson as *Being and Time*, New York: Harper & Row, 1962.(Heidegger's early masterpiece.)

Heidegger, M. (1969) *Zur Sache des Denkens*, trans. J. Stambaugh as *On Time and Being*, New York: Harper & Row, 1972.(Lectures which summarize Heidegger's later views on being, time and philosophy.)

Husserl, E. (1931) *Cartesian Meditations*, trans. D. Cairns, The Hague: Nijhoff, 1960.(Based on 1929 lectures; meditations 2 and 3 contain phenomenological discussions of being.)

Husserl, E. (1900) *Logische Untersuchungen*, trans. J.N. Findlay as *Logical Investigations*, New York: Humanities Press, 1970.(Meaning is discussed in Investigation I, being in Investigation VI.)

Inwagen, P. van (1993) *Metaphysics*, Boulder, CO: Westview Press.(Excellent introduction to metaphysics.)

Kant, I. (1781) *Critique of Pure Reason*, trans. N.K. Smith, London: Macmillan, 1933. (Crucial passages include A154-8, A219-26, A592-602.)

Lewis, D. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.(Forceful defence of modal realism.)

Parmenides (*c*.500 BC) 'Fragments', trans. and ed. G.S. Kirk, J.E. Raven, and M. Schofield in *The Presocratic Philosophers*, Cambridge, Cambridge University Press, 1983.(Among the first discussions of being.)

Plantinga, A. (1974) *God, Freedom, and Evil*, New York: Harper & Row.(Discusses necessary existence in the context of a defence of the ontological argument for the existence of God.)

Quine, W.V.O. (1953) 'On What There Is' in *From a Logical Point of View*, Cambridge, MA: Harvard University Press.(A succinct statement of Quine's views.)

Russell, B. (1918) *The Philosophy of Logical Atomism*, La Salle, IL: Open Court Publishing, 1985.(Sect. V treats existence and general propositions.)

Belief

We believe that there is coffee over there; we believe the special theory of relativity; we believe the Vice-Chancellor; and some of us believe in God. But plausibly what is fundamental is believing that *something is the case - believing a proposition, as it is usually put. To believe a theory is to believe the propositions that make up the theory, to believe a person is to believe some proposition advanced by them; and to believe in God is to believe the proposition that God exists. Thus belief is said to be a propositional attitude or intentional state: to believe is to take the attitude of belief to some proposition. It is about what its propositional object is about (God, coffee, or whatever). We can think of the propositional object of a belief as the way the belief represents things as being - its content, as it is often called.*

We state what we believe with indicative sentences in 'that'-clauses, as in 'Mary believes that the Democrats will win the next election*'. But belief in the absence of language is possible. A dog may believe that there is food in the bowl in front of it. Accordingly philosophers have sought accounts of belief that allow a central role to sentences - it cannot be an accident that finding the right sentence is* the *way to capture what someone believes - while allowing that creatures without a language can have beliefs. One way of doing this is to construe beliefs as relations to inner sentences somehow inscribed in the brain. On this view although dogs do not have a public language, to the extent that they have beliefs they have something sentence-like in their heads.*

An alternative tradition focuses on the way belief when combined with desire leads to behaviour, and analyses belief in terms of behavioural dispositions or more recently as the internal state that is, in combination with other mental states, responsible for the appropriate behavioural dispositions.

An earlier tradition associated with the British Empiricists views belief as a kind of pale imitation of perceptual experience. But recent work on belief largely takes for granted a sharp distinction between belief and the various mental images that may or may not accompany it.

1 Beliefs as sentences in mentalese: the language of thought

Fred's belief in the Devil cannot literally be a relation between Fred and the Devil. Otherwise he could not have the belief, unless the Devil existed. One response is to treat belief as a relation to sentences. To believe is to 'believe-true' a sentence: Fred believes-true 'The Devil exists'. But animals that lack a language have beliefs. My dog may believe that his master is home, or that it is time for a walk. Moreover, monolingual French speakers and monolingual English speakers may agree in what they believe, say, that it would be good if they knew more than one language, and yet they may not agree on which sentences they believe true. Finally, you might believe-true the sentence 'The Devil exists' and yet not believe that the Devil exists because you wrongly think that the word 'Devil' means 'God'. In this case what you believe is that God exists while wrongly thinking that the sentence 'The Devil exists,' is a good sentence to use to express this belief. For these reasons, and others, belief is usually thought of as a relation to a proposition. A proposition is what is expressed by a sentence; it is what is in common between sentences in French and English that mean the same; the proposition expressed is what is grasped when you understand a sentence. Monolingual speakers believe alike by believing the same propositions; dogs have beliefs by virtue of believing propositions despite not having a language to express them; someone who believes that the sentence 'The Devil exists' is true while thinking that 'Devil' means 'God' does not thereby believe that the Devil exists because they are wrong about what proposition 'The Devil exists' expresses. These remarks slide over a lively controversy concerning the ontological status of propositions (see Propositions, Sentences And Statements). Our immediate concern will be with a popular view that gives sentences a prominent role in the account of belief, but in a way which avoids the problems just rehearsed.

According to the language of thought hypothesis (LOTH), not only do certain sentences serve to provide the propositional objects of beliefs (and thoughts in general) but, in addition, the beliefs are themselves sentence-like. A sentence may be viewed as made up of significant parts put together according to certain rules. In the same general way, according to LOTH, beliefs have parts put together in certain ways (see Language Of Thought).

How does LOTH mesh with the idea that beliefs are relations to propositions? The idea is that a belief's propositional object is determined by how it is made up from parts which have representational or semantic

properties - that is, the parts stand for things, properties and relations much as the parts of a natural-language sentence do (see Semantics). In English 'biscuit' represents certain things, and 'crisp' represents a certain property, and when we combine them together to form the sentence 'Biscuits are crisp' we get a sentence that makes a claim that is true or false according to whether or not the things have the property. This is how the sentence expresses the proposition that biscuits are crisp (see Compositionality). In the same way, there are brain structures that represent things and properties, and when these brain structures are put together in the right way we get, says LOTH, a more complex structure, a sentence in mentalese, that represents the things as having the properties - as it might be, the sentence of mentalese that says that biscuits are crisp, that expresses that proposition, and that thereby provides us with a token of the belief that biscuits are crisp.

This theory can allow that dogs have beliefs. Dogs might have a language of thought even though they do not have a public language. It can explain how monolingual speakers of different language can agree in belief - their sentences of mentalese may express the same propositions. It also provides an explanation of a number of phenomena associated with belief. First, it explains how what a person believes can be causally relevant to what else they believe and what they do. If you believe that Mary is at the party and then learn that Mary is always accompanied at parties by Tom, you will typically come to believe that Tom is at the party. What you believe combines with what you learn to produce a new belief. LOTH explains these causal transactions as transactions between the structures that are the various beliefs. Much as a computer processes information by manipulating electronically coded structures so we arrive at new beliefs by virtue of our brains manipulating the symbols of mentalese. Similarly, what we believe contributes to explanations of what we do. My belief that there is coffee over there together with suitable desires may lead me to move over there by virtue of its being a belief that there is coffee over there. LOTH accounts for this fact in terms of the causal influence of the sentences of mentalese on the causal path to bodily movement.

Second, LOTH explains the fact that typically one who can believe that Jill loves Mary can believe that Mary loves Jill, and in general if you can believe that *aRb*, then you can believe that *bRa* (the phenomenon known as systematicity). The fact that if you can believe that *aRb*, then you can believe that *bRa* is explained by the fact that the state that encodes the former is a re-arrangement of the parts of the state that encodes the latter. And, of course, this explanation generalizes to explain more complex cases.

Finally, LOTH can explain our ability to form quite new beliefs (the property known as productivity). Just as we can form new sentences by novel combinations of the relevant words of a public language, so the brain can form new beliefs by means of novel combinations of the relevant words of mentalese.

LOTH thus explains a lot of what needs to be explained. Nevertheless, there are two serious problems for this view as applied to belief.

2 Two problems for LOTH

First, unless the claim that mentalese exists is trivialized - no matter what neuroscience reveals about how the brain processes information, what it reveals will count as the brain containing mentalese - LOTH involves risky speculation about how our brains work. The theory gives a hostage to fortune. Some are happy to accept this. If neuroscience reveals that there is no mentalese and that we do not process information in a sentential manner, we should say that we do not have beliefs and so embrace eliminativism about belief (see Eliminativism). This is, however, very much a minority view.

Second, LOTH leaves the intimate connection between belief and behaviour obscure. On the face of it predictions about the behaviour of highly complex organisms like ourselves should be enormously difficult. Trees bend in the wind whereas we put on jumpers, go inside houses, lean into the wind, cancel our games of tennis, or whatever. Unlike trees and simple machines, we respond to stimuli in enormously varied ways. Nevertheless we are quite good at predicting human behaviour. We all make many successful predictions of the following kind: someone who has uttered the word 'Yes' on hearing the sentence 'Would you like to come to dinner at 19.30 on the 21st?' will arrive around 19.30 at the house of the person the sentence came from. What we do, of course, is use hypotheses about what people believe and desire and predict in terms of the rule that subjects will tend to behave in such a way that they achieve what they desire if what they believe is true. Our subject's 'Yes' tells us what they desire, and what we predict - their turning up at the named time - is behaviour that will achieve their desire for

dinner.

Now we noted above how LOTH explains the way belief contributes to causing behaviour. In the same general way it explains how belief together with desire explains behaviour. For LOTH treats desires as like beliefs in being internal sentences of mentalese. The difference is that, as it is often put, the desires are stored in the ‘desire’ box, and the beliefs are stored in the ‘belief’ box. The metaphor of different locations marks the fact that beliefs and desires differ in how they relate to the world. Belief is a state that seeks to conform to how things are - the sight of coffee tends to extinguish my belief that there is no coffee near; whereas desire is a state that seeks to conform things to how it is - desire for coffee tends to bring one near coffee. The stored sentences that do the first job count as being in the belief box; the stored sentences of mentalese that do the second job count as being in the desire box. So the way belief and desire combine to produce behaviour is not a problem for the LOTH. The two ‘differently located’ stored sentences get together to produce the behaviour.

The problem, rather, arises from the fact that the connection between behaviour and what subjects believe and desire is most immediately one between behaviour and a rich system of belief and desire. Individual beliefs and desires grossly underdetermine behaviour. There is no behaviour that the belief that there is a mine near the tree, together with the desire to live, as such points to. It is, rather, a rich system of belief - to the effect, say, that there is a mine near the tree, that the mine is likely to be triggered by going near it, that moving one’s legs in such and such a way will not bring one near the tree, that there is not a bigger mine that can only be avoided by going close to the tree, that triggering mines tends to cause death, and so on and so forth, along with the desire to live being greater than the desire to test out the trigger system of the mines - that points to behaviour. When we give little illustrations of connections between subjects’ beliefs and desires and what they do, we take for granted a great deal about what they believe and desire. This is fine. It is by and large common knowledge. But the point remains that only rich systems of belief and desire have the intimate connection with behaviour. The same point could be made with the dinner invitation story. The prediction of our subject’s behaviour assumed a great deal by way of belief and desire. We assumed beliefs about what the words mean, about who uttered them, about which month was intended, …, and we assumed that there were no countervailing desires that outweighed the desire to go to dinner.

The problem for LOTH is that it takes as its starting point individual beliefs and desires. This leaves it seriously unclear what the theory has to say about the connection between a rich story about belief and desire, on the one hand, and behaviour, on the other. There is no behaviour that the individual belief that *p* and desire that *q* point to.*It is rich systems of both belief and desire that point in some reasonably determinate way to behaviour*. The challenge for LOTH is to find some kind of guarantee that the account of individual beliefs and desires it offers is such that if subjects have rich enough sets of these individual beliefs and desires, these rich enough sets of beliefs and desires will cause the reasonably determinate behaviour that tends to satisfy their desires if their beliefs are true.

3 Belief as a map by which we steer

One obvious fact about belief is the way we use sentences to state what we believe. An equally obvious fact is the connection between belief and behaviour via desire discussed above. F.P. Ramsey (1931) famously captured this idea by describing belief as a map by which we steer. The alternative to LOTH is an account of belief that sees belief as map like.

For LOTH, individual beliefs are fundamental; while on the map view systems of belief are fundamental. Inside us is a hugely complex structure that richly represents how things around us are in an essentially holistic way. When you believe the bank is bigger than the post office, there is no individual structure, no sentence of mentalese in your head, that represents your belief that the bank is bigger than the post office. Rather you believe that the bank is bigger than the post office by having a belief system according to which, among a great many other things, the bank is bigger than the post office. The key point can be made in terms of maps. A map of the Earth might represent the fact that the taller mountains are mostly near the deeper oceans, but there is no part of the map that says just that in the way that there may be a sentence that says just that - for instance the very sentence ‘The taller mountains are mostly near the deeper oceans’. Or consider holograms. Holograms are ’laser photographs’. When light from the laser is projected through the negative, the well known, three dimensional, coloured array is produced. The negative can be thought of as representing things as being the way the coloured array depicts them. However, no part of the negative has special responsibility for some part of the array. Each part contains

information about the whole array. In consequence, what happens if you damage part of the hologram is a loss of detail, a blurring, of the three dimensional array, not a loss in any particular part of it.

Many of the phenomena explained by LOTH can equally be explained by the map theory. We noted how LOTH can explain the evolution of belief over time in terms of the causal interactions of the internal sentences with each other, and how beliefs cause behaviour in terms of how the stored sentences figure in the causal path to behaviour. But internal maps guide rockets to their targets and evolve over time. The same goes for the maps we use every day - they guide our behaviour and evolve over time. We noted that LOTH can explain the fact that those with the capacity to believe that the bookshop is bigger than the post office are also able to believe that the post office is bigger than the bookshop. But maps (and holograms) that can represent that the bookshop is bigger than the post office can equally represent that the post office is bigger than the bookshop.

It has recently been argued that there is empirical evidence that our brains represent how things are around us in something like the way an internal map or hologram might. This has led to a renewed interest in the map theory of belief (see Connectionism).

The major question for the map theory concerns whether believing is closed under entailment. On the map theory to believe that *p* is to have a system of belief according to which *p*; that is, to have a system that could not be true unless *p*. But if *p* entails *q*, then a system that could not be true unless *p* must also be a system that could not be true unless *q*. This means that the map view must accept closure under entailment, the principle that if *p* entails *q*, anyone who believes *p* believes *q*. But is it not possible to believe that a triangle is equiangular without believing that it is equilateral - as many beginning geometry students know only too well? The usual reply by map theorists is to insist that one who believes that a triangle is equiangular does believe that it is equilateral; what they may lack is knowledge about the right words to capture what they believe. But this is a matter of lively debate.

See also: Behaviourism, analytic; Belief and knowledge; De Re/De Dicto; Functionalism; Intentionality; Probability theory and epistemology; Propositional attitudes

DAVID BRADDON-MITCHELL
FRANK JACKSON

References and further reading

Armstrong, D. M. (1973) *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.(Detailed development of the map theory that takes off from F. P. Ramsey's idea.)

Fodor, J. (1987) *Psychosemantics*, Cambridge, MA: MIT Press.(The appendix contains a very readable defence of LOTH by one of its major proponents.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford: Oxford University Press, 1888. (The Treatise is a classic widely available in many editions and is the standard source for the 'pale imitation of experience' view of belief, pretty much passed over in this entry.)

Lewis, D. (1994) 'Reduction of Mind', in S. Guttenplan (ed.) *A Companion to Philosophy of Mind*, Oxford: Blackwell.(Compares the map view to the language of thought view in detail.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Detailed, sympathetic discussion of whether belief should be thought of as 'believing-true' some linguistic item. Relates the discussion to the distinction between belief *de dicto* and belief *de re* not touched on in this entry.)

Ramsey, F.P. (1931) *The Foundations of Mathematics*, London: Kegan Paul.(Classic source of the view of belief as a map by which we steer, and of treatments of degree of belief in terms of betting behaviour.)

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.(Classic exposition of behaviourism about belief.)

Stalnaker, R.C. (1984) *Inquiry*, Cambridge, MA: MIT Press.(One of the most detailed discussions of whether belief is closed under entailment, along with two matters not touched on in this entry: treatments of what a subject believes in terms of where in the set of possibilities believers take themselves to be located, and the relation between belief and acceptance.)

Belief and knowledge

It is often said that for people to know that such and such is the case, they must have something like a belief that such and such is the case. Call this the 'entailment thesis'. It is usually added that the converse (call it the 'converse entailment thesis') is false: it is false that my belief-like attitude that such and such is the case always counts as knowledge. This standard view, combining the entailment thesis with the denial of the converse thesis, has been challenged in a number of ways.

The 'identity thesis' would retain the entailment thesis but would also endorse the converse entailment thesis. Knowledge and belief entail each other. (While no one has defended precisely this claim, Donald Davidson has come close.) The 'incompatibility thesis' rejects the entailment thesis as well as the converse entailment thesis, and says that knowledge and belief are mutually incompatible. Similarly, the 'separability thesis' also rejects the entailment thesis and the converse entailment thesis, but adds that knowledge and belief are mutually compatible. Those who defend the 'eliminativism thesis' hold that belief, like other elements of 'folk' or popular psychology, is an outmoded notion, and what is 'in our heads' when we know about the world is something other than beliefs.

1 Standard view

Perhaps the most common version of this view is that while some beliefs count as knowledge, not all do. And while epistemologists need to identify the conditions beliefs must meet if they are to acquire the status of knowledge, it is obvious that we cannot know that something is the case if we do not even believe it (see Knowledge, concept of). The common version of the standard view deals specifically with belief states. But some philosophers would claim that while knowledge entails belief-*like* states (but not vice versa), it is not precisely belief states that are entailed. Some of the substitutes that have been defended are psychological certainty, conviction, and acceptance.

A.J. Ayer is among those who say that knowledge entails psychological certainty, where I am psychologically certain that something is the case when I am not at all disposed to doubt it. Psychological certainty is not the same thing as infallibility, and it is only the former that Ayer links to knowledge (see Certainty). His view is based on an analysis of the meaning of the term 'know': according to him, 'to say of oneself that one knew that such and such a statement was true but that one was not altogether sure of it would be self-contradictory' (Ayer 1956: 16).

Keith Lehrer (1974) argued that, given common usage, knowledge need not entail conviction, but that philosophers ought none the less to restrict the use of 'know' so that knowledge does entail conviction. His grounds are that philosophers want to explain how it is possible for people to know what they *claim* to know: if I claim to know something, then I must be convinced that I do know it, but more importantly, I must be convinced that it is the case.

Later (1989) Lehrer retracted. He suggested that we need only to *accept* that something is the case in order to know that it is the case. There are two advantages to this. First, when we accept that something is the case, we do so for a purpose, and the relevant purpose for epistemic agents is the pursuit of truth. When we believe (or harbour a conviction) we need not have any such purpose. So the use of 'acceptance' rather than 'conviction' helps us keep in sight the goal-oriented nature of epistemic agents. Second, a belief (and conviction) can be the product of entirely irrational factors, such as wishful thinking, and one can believe things against one's better judgment. Acceptance, on the other hand, is governed by epistemic norms.

2 Identity thesis

While no one defends the claim that knowledge entails belief and vice versa, Donald Davidson (1983) comes close. He assumes the entailment thesis (that knowledge entails belief), but he also argues that belief is veridical, so that it is impossible for many of our beliefs to be false. Since belief is veridical, most beliefs constitute knowledge. Davidson bases his view that belief is veridical on theories of meaning and belief that appeal to a principle of charity and to a type of verificationism (see Charity, principle of §4; Davidson, D. §5).

First, Davidson assumes that when others interpret what we say, and attempt to figure out what we mean and what we believe, they must do their best to ensure that what we say is intelligible, and that requires that they avoid

ascribing to us very many false beliefs. So it is largely their own beliefs which they must attribute to us. Any interpreter must be *charitable*, and attribute to us beliefs that they consider to be largely true.

Next, Davidson adopts a form of verificationism. He claims that what interpreters of our discourse who are fully knowledgeable of our circumstances and speech behaviour would *say* we mean and believe *is* what we mean and believe. Combining his principle of charity with his verificationism, Davidson reasons as follows: The beliefs of ideally situated interpreters are by hypothesis true. Given their charitableness, when they interpret our discourse it will be largely these accurate views which they will attribute to us. But given verificationism, what they say goes: when they say that what we believe is largely true, they are right. 'Most of the sentences a speaker holds to be true… are true'; 'Belief is in its nature veridical' (Davidson 1983: 434, 432). If I wonder whether my beliefs are true, I have only to run through the above reasoning to see that 'beliefs are by nature generally true' (Davidson 1983: 437). This is not to say that it is impossible for me to believe a falsehood, however, for the ideal interpreter may be forced to attribute to me some false views.

There is plenty of room for doubt about Davidson's argument, however. One challenge made by Luper-Foy (1987) and others is that ensuring that a speaker's discourse is intelligible (which, according to Davidson, is the point of the principle of charity) need not involve attributing to the speaker what in the main we believe. What charitable interpreters, who are out to secure intelligibility, should attribute to speakers is what those interpreters would believe (and mean by what they say) if they were in the speaker's circumstances. Yet this policy might require attributing to speakers mostly false beliefs. For example, people who are in circumstances described by sceptics would be unintelligible to us if their views were not as mistaken as ours would be if *we* were there (see LePore 1986).

3 Incompatibility thesis

Arguments for the thesis that knowledge and belief are incompatible are not very strong. This thesis has been attributed to Plato, who suggested that belief (or opinion) was fallible while knowledge was not, but this suggestion does not imply the incompatibility thesis. People with a (fallible) belief state might also be related to the world in such a way that, given that relationship, their belief could not be false. For example, their (fallible) belief that *p* might be produced by a mechanism that infallibly detects the fact that *p*.

As developed by A. Duncan-Jones (1938), the incompatibility thesis is based on the claim that linguistic evidence precludes our saying that people know something once we have said that they believe it. People say things like 'I do not believe that Clinton is president, I *know* he is', which might suggest that only when we hold an opinion on weak or nonexistent grounds is the opinion a belief, so that belief is incompatible with knowledge. But the linguistic evidence need not suggest this. The evidence is consistent with the possibility that knowledge requires belief and something else, such as strong evidence. To say 'I believe that Clinton is president' when I have overwhelming evidence would be misleading to my listeners, who would assume that if I had overwhelming evidence I would say I *knew* that Clinton was president, not merely that I believed he was president (Grice 1989). Compare: 'I do not *grasp* the theory, I devised it' (see Implicature).

H.A. Prichard (1950) argued that knowledge entails psychological certainty, while mere belief always falls short of certainty. So the confident attitude we have when we know that something is the case rules out (merely) believing that it is the case. However, belief involves some degree of confidence, and surely I can increase my confidence indefinitely without ceasing to believe what I am so confident about.

4 Separability thesis

Perhaps the most prominent defence of the claim that knowledge and belief are separable is given by Colin Radford (1966). A.D. Woozley (1953) had argued that knowledge is compatible with the absence of confidence about what is known, for people who can pass a test prove themselves to know the material even if they are not confident about their responses. Radford uses a similar argument to suggest that knowledge is consistent with a complete absence of belief. Consider the diffident people just mentioned who pass the test. Suppose that they give the correct response 'Washington' to a particular question such as 'Who was the first president of the USA?' If they are sufficiently unconfident about their response, they might be prepared to *deny* that they believe that the first president was Washington, which is strong evidence that they do *not* believe it. Yet their accurate

performance on the test suggests that they knew the truth of their response. (Perhaps it would be inappropriate for them to *claim* that they knew, but that is not because they fail to know; rather, it is because they do not meet the conditions under which it is appropriate to claim knowledge, which would require that they believe that they have the knowledge they claim.)

Radford’s argument can be challenged on various grounds. First, we might point out that we are not infallible authorities about which beliefs we have. Perhaps we think we do not have a belief (say because we cannot detect it through introspection) when in fact we do (say because beliefs are behavioural dispositions such as the tendency to assent when asked the appropriate question) (see Introspection, epistemology of §§1-2). Second, we might deny that the people in Radford’s cases really do know what he claims that they know. After all, they have no evidence for the truth of their test responses; indeed, they have reason to think that their responses are incorrect. So, from their own subjective view, the truth of their responses would be an accident at best.

5 Eliminativism thesis

Patricia Churchland (1986), Paul Churchland (1979) and Stephen Stich (1983) offer a final, and somewhat oblique, challenge to the standard view. They suggest that the notion of belief is scientifically outmoded and ought to go the way of notions like phlogiston and ether. Moreover, theories of knowledge ought to focus on elements of cognition that are more primitive than any belief-like states could be.

Paul Churchland supports the latter point by attacking the common idea that the cognitive apparatus of epistemic agents functions essentially by manipulating sentences, and that rational agents are ones whose apparatus manipulates sentences in accordance with given logical and quasi-logical relations among those sentences. A problem with this common view is that it cannot possibly be true of infants. It is no more tempting to explain the behaviour of infants in terms of propositional attitudes than it is to explain the behaviour of protozoa or plants that way. Their development must be explained in terms of something more primitive than propositional attitudes. But if the rational development of adults is continuous with that of infants, then the rationality of adults must be explained in terms of the same primitives as that of infants.

Eliminativism is sketchy and hence hard to assess as a challenge to the standard view of knowledge. It is not clear what a less ‘folksy’, more scientific psychology would use in place of the notion of belief. It might still need a notion that is very much like the notion of belief, in which case a version of the standard view might survive (see Eliminativism; Folk psychology).

See also: Belief

STEVEN LUPER

References and further reading

All the following works are nontechnical, although demanding.

Ayer, A.J. (1956) *The Problem of Knowledge*, Harmondsworth: Penguin. (Referred to in §1. Analyses knowledge; discusses scepticism and certainty.)

Churchland, P.M. (1979) *Scientific Realism and the Plasticity of Mind*, Cambridge: Cambridge University Press. (Clear presentation of scientific realism which challenges the idea that knowledge should be understood in terms of belief. Referred to in §5.)

Churchland, P.S. (1986) *Neurophilosophy*, Cambridge, MA: Bradford Books/MIT Press.(Referred to in §5. Develops a view concerning the relationship between philosophy of mind and the scientific study of the central nervous system.)

Davidson, D. (1983) ‘A Coherence Theory of Truth and Knowledge’, in *Kant Oder Hegel*, Stuttgart: Klett-Cotta Buchhandlung, 423-38.(Referred to in §2. Attempts to refute scepticism by showing that belief is by its nature veridical.)

Duncan-Jones, A. (1938) ‘Further Questions about "Know" and "Think"’, *Analysis* 5 (5): 96-106; repr. in M. MacDonald (ed.) *Philosophy and Analysis*, Oxford: Blackwell, 1966.(Develops the incompatibility thesis. Referred to in §3.)

Grice, H.P. (1989) *Studies in the Ways of Words*, Harvard, MA: Harvard University Press.(Referred to in §3. Develops an account of the conversational implications of ‘believe’ and ‘know’.)

Lehrer, K. (1974) *Knowledge*, Oxford: Oxford University Press.(Referred to in §1. Provides an analysis of knowledge and justification.)

Lehrer, K. (1989) 'Knowledge Reconsidered', in M. Clay and K. Lehrer (eds) *Knowledge and Skepticism*, Boulder, CO: Westview Press.(Revises Lehrer's theory of knowledge. Referred to in §1.)

LePore, E. (ed.) (1986) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.(Collection of essays about the views ofDonald Davidson. Referred to in §2.)

Luper-Foy, S. (1987) 'Doxastic Skepticism', *The Southern Journal of Philosophy* 25 (4): 529-38.(Referred to in §2. Critically discusses Davidson's view that belief is by nature veridical.)

Prichard, H.A. (1950) *Knowing and Perception*, Oxford: Clarendon Press.(Analyses knowledge and perception. Referred to in §3.)

Radford, C. (1966) 'Knowledge - By Examples', *Analysis* 27 (1): 1-11.(Defends the separability thesis. Referred to in §4.)

Stich, S. (1983) *From Folk Psychology to Cognitive Science: The Case against Belief*, Cambridge, MA: Bradford Books/MIT Press.(Referred to in §5. Develops a philosophy of mind according to which common-sense psychological terms are not suited to cognitive science.)

Woozley, A.D. (1953) 'Knowing and Not Knowing', *Proceedings of the Aristotelian Society* 53 (1952-3): 151-72. (Discusses the relationship between knowledge and confidence. Referred to in §4.)

Belinskii, Vissarion Grigorievich (1811-48)

Belinskii was considered by his followers in the nineteenth century, and by the official ideology of the Soviet period, to be not only Russia's greatest literary critic, but also a leading Russian thinker. Soviet encyclopedias label him 'critic, publicist and philosopher'. His role in Russian cultural life has been given positive as well as negative assessments, but there can be no doubt as to his huge influence. He is largely responsible for the fact that Russian literature and art, for a century and a half now, have been considered an organ of society, a mirror of the Russian nation's destiny and a vehicle of its historical progress. It is largely his merit - or fault - that in Russia, art and literature have been accorded a lofty status of leadership and authority, and also that 'art for art's sake' never became respectable in Russia. The influence of Belinskii's philosophy of art extended through the entire political spectrum, far beyond his political legacy which was limited to the revolutionary left. The idea that art and literature are organic functions of society, nationhood and historical progress, which Belinskii took for granted, was passed on even to the Slavophile right and the liberal Westernizing centre. It was still an integral part of the doctrine of Socialist Realism.

1 Life and critical career

Belinskii was born in 1811 in Sveaborg, Finland, the son of a naval surgeon. He attended Moscow University from 1829 to 1832, but was dismissed 'for reasons of health and limited ability'. He was an active member of a discussion group led by Nikolai Stankevich and inspired by the philosophy of J.G. Fichte and F.W.J. Schelling. Among its members were Mikhail Bakunin, Timofei Granovskii and Konstantin Aksakov (see Schellingianism, Russian §1). In 1833 Nikolai Nadezhdin, Belinskii's professor, engaged him as a collaborator of *Teleskop (Telescope)*, a literary journal he had founded in 1831. Belinskii's first major essay, 'Literary Reveries', appeared there in 1834. It was the first of his many surveys of Russian literature and very much under the sway of Schellingian ideas. After the suppression of *Teleskop* in 1836, Belinskii worked, in succession, for *The Moscow Observer* (1838-9), *National Annals* (1839-46) and *The Contemporary* (1846-8). During his only trip abroad in 1847, Belinskii wrote his celebrated 'Letter to Gogol''. He died of consumption in 1848.

During his brief career, Belinskii published a series of annual surveys of Russian literature, scores of major essays and many hundreds of reviews of books on any conceivable subject. Many of his reviews were essays in their own right, for instance, when he would use a book for children as an occasion to expound his views on education. Belinskii correctly recognized every major poet and writer of his age (Pushkin, Gogol', Lermontov, Herzen, Turgenev, Dostoevskii) and dismissed minor lights, even if they were ephemerally successful. He established an approach to literature that saw it as an organic whole of which individual works were a living part.

Belinskii read French, but no other foreign language. However, his friends kept him well informed on recent developments in the West and all the major classics were by then available in Russian translation. Belinskii was reasonably well read in the major authors of the West and regularly responded to recent works of French, English and German literature as they were translated into Russian (Hugo, Janin, Sue, George Sand, Hoffmann, Heine, Walter Scott, Dickens, J.F. Cooper and many others).

2 Evolution of Belinskii's aesthetics

Belinskii's philosophy of art developed under the influence of a series of Western authors, yet always in direct connection with Russian life and the development of Russian literature. Specifically, he experienced, in succession, the stimulating influence of Romantic theories of art and poetry (Fichte, Schelling, the Schlegel brothers); Hegel (after 1837); and the Left Hegelians and French Utopian Socialists (in the 1840s) (see Romanticism, German §§1-3; Hegel, G.W.F; Hegelianism §§1-3).

Belinskii's conception of the organic nature of the work of art (its idea being to its form as soul to body); his notion of the symbolic nature of all true art (along with a dismissal of allegory and schematism as 'nonart'); and his belief in the cognitive power of art and the prophetic powers of genius were derived from Schelling's *System des transcendentalen Idealismus (System of Transcendental Idealism)* (1800). Given in 'The Idea of Art' (1841), Belinskii's famous definition of art as 'the *immediate* contemplation of truth, or thinking in *images*' (1952-9: 584; original emphases), squares with Schelling's: 'Art may be defined as the real representation of the forms of things

as they are in themselves - their proper, native forms, then'. Even more closely, it tallies with A.W. Schlegel's definition of poetry as 'expression of thought in sense images'.

In 1837, Belinskii, along with Stankevich and Bakunin (later his collaborator on *The Moscow Observer*), were converted to a Hegelian philosophy of art and history (see Hegelianism, Russian §§2-3). Specifically, Belinskii used a conspectus of Hegel's lectures on aesthetics, which M.N. Katkov had translated for him. The list of Belinskii's positions that may be derived from Hegel is a long one. As early as 1838, in a major essay on Shakespeare's *Hamlet*, Belinskii wrote of works of art as 'a manifestation of the Spirit, [as] a given stage of its consciousness'. Soon enough he developed a thoroughly historicist approach to literature, following the evolution of the Spirit in concrete historical developments. Concretely, Pushkin's verse epic 'Poltava' (1828) he considered anachronistic and hence a failure, because the age of epic poetry belongs to the past. Also quite concretely, Belinskii observed progress in literature, as Lermontov's Pechorin (in *A Hero of Our Time* (1841)) is perceived as an advance from Pushkin's Onegin (in *Evgenii Onegin* (1830)), while Gogol''s prose represents an advance from the poetry of his predecessors.

During his Schellingian period, Belinskii had been inclined towards Romantic ideas. Under Hegel's influence, his critical thought took a decisive turn to realism. (He called it 'poetry of reality'.) Belinskii's Hegelian historicism caused him to place Russia in the position of an emerging nation, which actually caused him to downgrade the achievement of his contemporaries Pushkin and Gogol', crediting the former with having created the formal tools for a literature of the future, the latter with being the first to deal with Russian life in a constructive way, and denying both international stature. He confidently predicted that Russia would make original contributions to world literature in the future. At the same time, Belinskii overestimated the importance of some Western writers, Walter Scott and J.F. Cooper in particular.

Up to 1840, Belinskii interpreted - or misinterpreted - Hegel's dictum, 'All that is rational is real, and all that is real is rational', to mean that Russians had to reconcile themselves to Russia's all too real backwardness, because it was a necessary stage in the nation's rational development. Later he rejected this notion, realizing instead that it was a Russian's moral duty to work for what was perceived as the rational advancement of Russian nationhood. Belinskii followed Hegel in assuming that the World Spirit realized itself in national spirits. In 1846 he polemicized with the critic Valerian Maikov, who had suggested that progress was proportional to an increase of independence from national ties. A Westernizer, Belinskii was also a Russian patriot.

Hegel's influence shows in some specific details, such as the meaning given to the word 'pathos' ('the rich, powerful individual quality in which the substantial movements of the Spirit are brought to life, achieving reality and expression', as defined in Hegel's *Aesthetics*), as it appears in Russian criticism to this day. Belinskii's conception of an ideal expressed through its negation is clearly Hegelian, for example, when he interprets Gogol''s comedy *The Inspector-General* as the presentation of a 'phantom reality' flowing from the irrational, chaotic principle of life and struggling against the emergence of a positive, rational reality ('Woe from Wit' 1840). Hegelian, too, is Belinskii's resolute rejection of both formalism and naturalism in art.

After 1842 Belinskii began to lean towards a Left-Hegelian or Utopian Socialist position, with an emphasis on *socialité* (*sotsial'nost'*), an overriding concern with social ills and their correction. He now developed the notion that true art is necessarily in step with social progress, in the same sense as he had always taken for granted that art is true to life. Induced by his social and political concerns, Belinskii departed from his previous, Hegelian position that art should be objective and dispassionate. He now suggested that in a 'critical epoch' (a Saint-Simonian term!), such as his own, works with a subjective tendency could play a useful role as catalysts of correct ideas, even if they were not artistically flawless. The practical consequence of this attitude was that Belinskii would now respond approvingly to works he admitted were 'unartistic', but were instrumental in leading the public in the right direction. Such was, for example, Aleksandr Herzen's novel *Whose Is To Blame?* (1847) (see Herzen, A.I. §1).

3 Art and socio-historical reality

Belinskii thus opened the door to didacticism, moralism and political propaganda in Russian literature (art was to follow soon enough, in the *Itinerant* movement). However, he was himself able to discriminate between the aesthetic merits of a given work and its moral and political virtues. In a way, Belinskii would - and did - eat his cake and have it, too, for he could demonstrate, at least to his own satisfaction, that the works which he loved and

respected were, by and large, morally and politically on the right track. When his beloved Gogol' turned reactionary, Belinskii could say, with some justification, that he was no longer the great artist he once was.

Belinskii considered a nation's world of letters (*pis'mennost'*) to be an organic whole, consisting of hierarchically layered, but independent parts: *poesy* (*poėziia*, German *Dichtung*), whose creations of genius reveal the truth of historical progress; *belles-lettres* (*belletristika*) which reduce the insights of poesy to terms readily understood by a broad reading public; and *journalism* (*zhurnalizm*) which applies these insights to the nation's day-to-day concerns.

Belinskii's aesthetic was, like that of his Western influences, an aesthetic of content (German *Gehaltsaesthetik*), meaning that the merits of a work of art were explored proceeding from an analysis of its social, moral, political and historical content, rather than from an analysis of its formal traits (composition, style, language, poetic devices). From the viewpoint of a Kantian aesthetic of form (German *Formaesthetik*), a Belinskiian approach ignores the very essence of art: the expression of aesthetic intuitions through appropriate artistic devices. In 'On the Poet's Calling' (1921), the poet Aleksandr Blok said that Belinskii had done more damage to poesy than Benckendorff, chief of gendarmes under Nicholas I, for the latter's censorship had only temporarily stopped the publication of some works, while Belinskii had induced Russian poets to betray their calling. A negative opinion of Belinskii's role in Russian letters was consistently expressed in the works of the Russian Formalist school of the 1920s (see Russian literary formalism).

The most telling criticism of Belinskii's method came from a radical critic who declared himself to be the master's faithful disciple: Dmitrii Pisarev (1840-68) correctly charged that Belinskii had projected his own humane and progressive ideas into works of 'pure art', such as Pushkin's *Evgenii Onegin*, because he wished to make his own progressive ideas agree with his love of Pushkin.

4 Legacy to Russian aesthetic thought

The generation of 'progressive' critics who claimed to be Belinskii's heirs, like Nikolai Chernyshevskii (1828-89), Nikolai Dobroliubov (1836-61) and Mikhail Saltykov (1826-89), developed the habit of using works of literature as texts for their political sermons (see Russian Materialism: 'the 1860s' §3). Belinskii was to some extent responsible for this practice. He interpreted Gogol''s *Dead Souls* or Dostoevskii's *Poor Folk* as indictments of the ills of Russian society and not as works of art, and therefore failed to see their whole depth, but he did base his observations on the text referred to. Later Chernyshevskii would assert that, inasmuch as all literature was inherently 'propaganda', the only thing that mattered was that it propagate the right ideas; Dobroliubov would often simply ignore the text under review; and Saltykov would merely dismiss as irrelevant works that did not clearly state their ideological position.

From a theoretical viewpoint, the right-wing followers of Belinskii's organic aesthetics, such as Apollon Grigor'ev (1822-64) and Fëdor Dostoevskii (1821-81), were closer to his position. They postulated, like Belinskii, that a true artist who would honestly follow the demands of his craft would *eo ipso* find the right content for his creations. But more unequivocally than Belinskii, they would insist that this agreement between art and the truth of life should be reached on art's terms - not those of ideology.

Belinskii was canonized by Chernyshevskii and Dobroliubov as the first of Russia's revolutionary democrats and as a precursor of the revolutionary movement of the 1860s. This happened largely on the strength of his 'Letter to Gogol'', his correspondence, and oral tradition, for his published writings were of course heavily censored. On balance, Belinskii was less radical than his disciples of the 1860s. He did claim, in his 'Letter to Gogol'', that the Russian people were at bottom atheist and anti-Church, but it is also quite clear from his correspondence and his essays on Russian folklore that he was less inclined to fall prey to a populist mystique than the radicals of the 1860s. He was of course passionately opposed to serfdom, but so were even some officials close to the Tsar. He considered the imperial bureaucracy to be a corporation of embezzlers and extortionists, but this was really an open secret: the Tsar himself had heartily guffawed at the first night of *The Inspector-General*. It appears that Belinskii's involvement with Utopian Socialism was a passing one and that his political vistas were rather down-to-earth. He mentions 'enforcement of existing laws' as a reasonable target in 'Letter to Gogol''.

VICTOR TERRAS

List of works

Belinskii, V.G. (1952-9) *Polnoe sobranie sochinenii (Collected Works)*, Moscow: AN SSSR, 13 vols.(Academy edition, with notes and indices.)

Belinskii, V.G. (1956) *Selected Philosophical Works*, Moscow: Progress.(Contains most of Belinskii's programmatic articles.)

References and further reading

Blok, A. (1921) 'O naznachenii poėta' (On the Poet's Calling), in *Stikhotvoreniya (Poems)*, Leningrad, 1936, 483-6.

Bowman, H. (1954) *Vissarion Belinski, 1811-1848: A Study in the Origins of Social Criticism in Russia*, Cambridge, MA: Harvard University Press.(A solid study of the social aspect of Belinskii's criticism.)

Fasting, S. (1972) *V.G. Belinskij: Die Entwicklung Seiner Literaturtheorie (V.G. Belinksii: The Evolution of his Theory of Literature)*, Bergen: Universitetsforlaget.(A meticulous analysis of the development of Belinskii's literary theory; combines attention to Belinskii's Western sources with careful attention to his critical practice.)

Terras, V. (1974) *Belinskij and Russian Literary Criticism: The Heritage of Organic Aesthetics*, Madison, WI: University of Wisconsin Press. (Concentrates on Belinskii's sources and the legacy of his criticism in the nineteenth and twentieth centuries.)

Bell's theorem

Bell's theorem is concerned with the outcomes of a special type of 'correlation experiment' in quantum mechanics. It shows that under certain conditions these outcomes would be restricted by a system of inequalities (the 'Bell inequalities') that contradict the predictions of quantum mechanics. Various experimental tests confirm the quantum predictions to a high degree and hence violate the Bell inequalities. Although these tests contain loopholes due to experimental inefficiencies, they do suggest that the assumptions behind the Bell inequalities are incompatible not only with quantum theory but also with nature.

A central assumption used to derive the Bell inequalities is a species of no-action-at-a-distance, called 'locality': roughly, that the outcomes in one wing of the experiment cannot immediately be affected by measurements performed in another wing (spatially distant from the first). For this reason the Bell theorem is sometimes cited as showing that locality is incompatible with the quantum theory, and the experimental tests as demonstrating that nature is nonlocal. These claims have been contested.

1 Survey

'Bell's theorem' is the generic name for a family of results that restrict the statistics for a certain type of quantum mechanical 'correlation experiment'. This is the sort of arrangement proposed by Einstein, Podolsky and Rosen ('EPR') (1935), where they argue that either quantum mechanics violates a condition of local causality or it is incomplete (see Einstein, A. §5; Bohr, N.). In the experiment a uniform source emits pairs of 'particles' in a special state. The particles then separate, each moving to a distant wing of the experimental apparatus. The separate wings contain devices that can perform any one of several different 'yes/no' measurements; that is, measurements with just two possible outcomes. For the most economical case suppose that in the 'A-wing' there are two possible measurements A or A' and similarly measurements B or B' in the 'B-wing'.

It is important for the analysis that in a given wing no two measurements can be performed simultaneously (technically, that the quantities being measured are incompatible - like position and linear momentum, or spin components in skew directions). A particular run of the experiment then corresponds to a choice of a single A-wing measurement and a single B-wing measurement, allowing four possible runs: AB, AB', $A'B$ and $A'B'$. From the frequencies of outcomes we can gather statistics that govern these runs. For an AB-run, for example, we can calculate the probability $\mathrm{P}(A)$ that the A-measurement turns up 'yes', the probability $\mathrm{P}(B)$ that the B-measurement turns up 'yes' and the probability $\mathrm{P}(AB)$ that both the A and the B measurements turn up 'yes'. (These probabilities suffice to fix the probabilities for 'no' outcomes as well.) In the same way, for other runs we can determine $\mathrm{P}(A')$, $\mathrm{P}(B')$, $\mathrm{P}(AB')$, $\mathrm{P}(A'B)$, and $\mathrm{P}(A'B')$. The experiment is so designed that $\mathrm{P}(A)$, from an AB-run, is the same as $\mathrm{P}(A)$ from an AB'-run, and similarly for the other single probabilities. By using sufficiently long runs, we should be able to approximate the probabilities experimentally to any desired degree of accuracy. We can also calculate them from the quantum mechanical state function that governs the pairs emitted from the source (see Quantum mechanics, interpretation of §§1-2).

Bell's theorem shows that under certain assumed conditions these probabilities satisfy a restrictive system of inequalities (the 'Bell inequalities') which, for special states of the pairs, are violated by the probabilities of quantum mechanics. The quantum probabilities have been verified to a high degree in various experimental tests and hence the Bell inequalities are violated. There are loopholes in the tests due to inefficiencies in the detection and pairing of results, but the experiments do suggest that the assumptions behind the Bell inequalities are not only incompatible with quantum theory but also with nature. One of the central conditions on which the derivation of the Bell inequalities depends is a species of no-action-at-a-distance, called 'locality': roughly, the condition that the outcomes in one wing are not immediately affected by measurements performed in the other wing. For this reason, the Bell theorem is sometimes cited as showing that locality is incompatible with the quantum theory, or that nature is nonlocal. These claims have been contested (Fine 1997).

2 The Bell inequalities

For the correlation experiment described above the strongest set of Bell inequalities are these:

$$\mathrm{P}(A)+\mathrm{P}(B)-1 \leq \mathrm{P}(AB)+\mathrm{P}(AB')+\mathrm{P}(A'B)-\mathrm{P}(A'B') \leq \mathrm{P}(A)+\mathrm{P}(B),$$

together with the three permutations of this obtained by interchanging *A* with *A*′, *B* with *B*′ and both together. This system has the following purely probabilistic significance. If we are given any four joint probabilities, $\mathrm{P}(AB)$, $\mathrm{P}(AB')$, $\mathrm{P}(A'B)$, $\mathrm{P}(A'B')$, and four 'compatible' singles, $\mathrm{P}(A)$, $\mathrm{P}(A')$, $\mathrm{P}(B)$ and $\mathrm{P}(B')$, (where 'compatibility' here means that, for instance, the probability for *A*-values derived from $\mathrm{P}(AB)$ is the same as the probability for *A*-values derived from $\mathrm{P}(A)$), then the system of Bell inequalities above are the necessary and sufficient conditions for all of the given probabilities to be derivable from one distribution, $\mathrm{P}(AA'BB')$, for all four variables. Put differently, the Bell inequalities are the conditions that the given probabilities can be represented on a classical probability space where all outcomes (even for incompatible measurements) are defined.

These inequalities fail maximally in the case most studied experimentally (for example, by the Aspect experiment that is often cited as decisive (Aspect *et al.* 1982)). Here quantum mechanics assigns 0.5 to all the single probabilities and assigns $\mathrm{P}(AB) = \mathrm{P}(AB') = \mathrm{P}(A'B) = 0.4268$ and $\mathrm{P}(A'B') = 0.0732$. The inequality becomes $0 \leq 1.2072 \leq 1$; and so it fails.

There are also cases where several correlations (represented by the joint probabilities) are strict (that is, 0 or 1); for example, where $\mathrm{P}(AB') = \mathrm{P}(A'B) = 0$, $\mathrm{P}(A'B') \neq 0$ and $\mathrm{P}(A \text{ or } B) = 1$. Although strict correlations are difficult to acheive in practice, Hardy (1993) has shown how to realize these probabilities, in principle, by means of a simple experimental design. Since $\mathrm{P}(A hbox or B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(AB)$, we have that $\mathrm{P}(AB) = \mathrm{P}(A) + \mathrm{P}(B) - 1$ and the left side of the Bell inequality reduces to $0 \leq -\mathrm{P}(A'B')$. Since probabilities are non-negative, this implies that $\mathrm{P}(\mathrm{A'B'}) = 0$, which contradicts the value assigned.

It is easy to construct a rationale for why a contradiction arises here. Because $\mathrm{P}(A'B') \neq 0$, there are runs where B′ yields 'yes'. In these runs if A were measured, then, because $\mathrm{P}(AB') = 0$, A would show 'no'. But if A were 'no' and B were measured, then B would be 'yes', since $\mathrm{P}(A \text{ or } B) = 1$. Hence where B′ yields 'yes', if B were measured instead, then B would yield 'yes'. But where B yields 'yes', A′would show 'no' because $\mathrm{P}(A'B) = 0$. So, where B′ yields 'yes', A′ yields 'no'; contradicting $\mathrm{P}(A'B') \neq 0$.

Clearly this rationale requires strong assumptions to make the counterfactuals along the sequence $B'A \rightarrow BA \rightarrow BA' \rightarrow B'A'$ mesh just right. Each arrow marks a use of locality, but in addition we need further assumptions that allow for transitivity so that we can transfer the *B*′ value from the far left to the far right. Assumptions of 'counterfactual definiteness' (that for each measurement there is an outcome such that if the measurement were made that outcome would result) and of pre-existing values (that each particle in a wing already possess the value - 'yes' or 'no' - that any measurement in that wing would reveal) have been considered. If locality forbids a measurement in one wing from altering the counterfactually definite or the pre-existing values in the other wing, then the stated rationale goes through. Speaking roughly, we can say the Bell theorem rules out the combination of locality plus determinism. If the result of a measurement is not predetermined, however, but is a matter of chance, corresponding to a random choice of 'yes' or 'no', then it is difficult to vindicate the claim that the assigned probabilities alone conflict with locality.

3 Behind the inequalities

In the general case, a standard rationale for the Bell inequalities involves two assumptions. One is an assumption that forbids a distant change in local probabilities. We can call it 'locality'. The other assumption is not related to locality so clearly; it is that there are no basic correlations between the measurement outcomes in the wings. This is 'factorizability' (other names are 'outcome independence' and 'completeness'). A simple framework for these is to consider an elementary sort of hidden variables ('noncontextual') model for a correlation experiment. In such a model, we introduce a set of parameters '*x*' (the hidden variables) that do not depend on any measurements or their results, and let $\mathrm{p}(A, x)$ be the probability at *x* that a measurement of A in the A-wing would produce a 'yes' (similarly for the other single measurements), and let $\mathrm{p}(AB, x)$ be the probability at *x* that both *A* and *B* would yield 'yes' (similarly, again, for the other products). The overall probabilities above (the $\mathrm{P}(A)$, $\mathrm{P}(AB)$, and so on) are to be realized by averaging over the parameters *x*.

Locality is automatically satisfied in this framework since the situation in one wing (that is, the single probabilities assigned at *x*) depend only on the measurement to be performed in that wing and on the particular parameter *x*, which is measurement-independent. (If *x* is measurement-independent violations of locality require that

probabilities at *x* in one wing depend on measurements made in the other wing, so give rise to a representation like $\mathrm{p}(S, x, T)$, where S represents the result of a measurement in one wing and T represents a measurement performed in the other wing. This would be a 'contextual' hidden variables' theory.) *Factorizability* is just the requirement that the joint probabilities factor; that is, that $\mathrm{p}(AB, x) = \mathrm{p}(A, x)\mathrm{p}(B, x)$, which is the condition for stochastic independence at *x* (similarly for the other products). This condition is equivalent to requiring for each parameter *x* that measurement outcomes in the two wings are uncorrelated.

It is straightforward to prove the Bell inequality of §2 from these assumptions provided a third condition is satisfied: namely, that each measurement produces, or is likely to produce, a result. Local, factorizable models can be constructed where this condition fails ('prism' models) and these models can accommodate the quantum probabilities and, so far, the statistics of all the experiments. (This is connected with the efficiency loophole mentioned in §1.) Finally, notice that these hidden variables' models are stochastic so that measurement outcomes are governed by probabilities and need not correspond to predetermined values (whether actual or counterfactual). Thus the Bell theorem here does not depend on determinism. On the other hand, it does not rule out locality alone but only the combination of locality and factorizability (given that the third condition holds).

Locality is an intuitively plausible constraint whose violation would in principle allow one to signal between the distant wings instantaneously, provided one could control the parameters *x*. If those parameters are not controllable, however, and instantaneous signalling is still ruled out, then locality may seem less compelling. (This is the situation in Bohmian mechanics, which is nonlocal but where instantaneous signalling does not occur. See Quantum mechanics, interpretations of §3). The rationale for factorizability is much less clear. It seems to represent a feeling that stable correlations need to be explained and that the only acceptable explanation traces correlations to statistically uncorrelated events. A contrasting point of view might accept some correlations as basic, not themselves in need of explanation, and try to account for others on that foundation.

Considerations about correlations are more pointed when the correlations are strict. Where the probabilities are 0 or 1 the stochastic case become deterministic. Locality then is not about affecting probabilities at-a-distance but about affecting values. So if $\mathrm{P}(AB') = 0$ (as in §2) then there are not just statistics to account for but, arguably, the fact that when a measurement in one wing produces a 'yes' the distant measurement always comes out 'no'. Factorizability becomes the requirement that the value at *x* assigned to the product of physical quantities just be the product of the values at *x* assigned to each quantity separately; using an obvious notation, that $\mathrm{val}(AB') = \mathrm{val}(A)\mathrm{val}(B)'$. This 'product rule' is studied in another theorem (the Bell-Kochen-Specker theorem) where it is shown to contradict quantum mechanics for certain systems, regardless of the state of the system, if the only values assigned are those allowed by the quantum theory. The product rule has been challenged on the grounds that the quantum theory only requires it in certain very special cases (technically, only in eigenstates of the quantities) and that these alone are not sufficient to produce a contradiction. Bohmian mechanics violates the product rule, as do a host of hidden variables' models, and each of these produce the very same statistical predictions as the quantum theory itself.

4 Without the inequalities

Although the Hardy example of §2 violates the Bell inequality, the rationale for why it gives rise to a contradiction does not involve all the machinery that lies behind deriving the Bell inequalities in general. There are other such 'Bell theorems without inequalities'. One set-up, due to Greenberger, Horne, Shimony and Zeilinger (GHSZ) (1990), involves three systems (labelled 1, 2 and 3) in a certain special state. Imagine these as moving away from a common centre along three straight lines in the same plane. For each system we are interested in yes/no quantities that relate to the vertical (*x*) and horizontal (*y*) directions (where *z* is the direction of motion of the system). So, for example, Q_{x_1} is the quantity relating to the *x*-direction for system 1, Q_{y_3} the quantity relating to the *y*-direction for system 3, and so on. Consider an assignment of values to the various quantities where $x_1 = \mathrm{val}(Q_{x1})$, $y_3 = \mathrm{val}(Q_{y3})\ldots$, then if 'yes' is assigned 1 and 'no' assigned 0, the following system of equations hold

$$x_1x_2x_3 = -1\, x_1y_2y_3 = 1 y_1x_2y_3 = 1 y_1y_2x_3 = 1$$

since if we add *Q* to the *x*s and *y*s these four equations give the values prescribed by quantum mechanics for the respective product operators. Since $(x_1)^2 = (y_1)^2 = 1$, if we multiply the first two equations we get $x_2x_3y_2y_3 = -1$. If we now multiply the second two we get a contradiction; namely, $x_2x_3y_2y_3 = 1$.

It is clear that the GHSZ example relies heavily on the product rule discussed in §3, extended here to the product of three quantities. Locality enters in the supposition that the same x_1 occurs in the first two equations; for example, that if x_1 were determined by distant measurements of Q_{x2} and Q_{x3}, and if instead we had measured Q_{y2} and Q_{y3}, then the value of Q_{x1} would not be altered. Similar uses of locality underlie the occurrence of the other terms shared between equations. As in the Hardy example, here too locality needs to be supported by further assumptions in order to justify the manipulations that lead to a contradiction.

In all the cases covered by the Bell theorem, simple-looking arguments that seem to overthrow locality are deceptive. Without question, however, the Bell theorem challenges certain features that one might desire, such as the picture of an observer-independent reality that measurement simply reveals (see Scientific realism and antirealism); it challenges them, that is, provided locality too is assumed.

See also: Probability, interpretations of; Quantum measurement problem

ARTHUR FINE

References and further reading

Aspect, A., Dalibard, J. and Roger, G. (1982) 'Experimental Test of Bell's Inequalities Using Time-varying Analyzers', *Physical Review Letters* 49: 1,804-7.(Frequently cited experimental test of the Bell inequalities, §2.)

Ballentine, L. and Jarrett, J. (1987) 'Bell's Theorem: Does Quantum Mechanics Contradict Relativity?', *American Journal of Physics* 55: 696-701.(Locality and no-signalling treated for contextual hidden variables; §3.)

Bell, J. (1987) *Speakable and Unspeakable in Quantum Mechanics*, Cambridge, Cambridge University Press. (Contains most of Bell's important papers, including his version of the Kochen-Specker theorem mentioned in §3.)

Branning, D. (1997) 'Does Nature Violate Realism?', *American Scientist* 85: 160-7. (Readable account of an optical experiment attempting to realize the Hardy example of §2.)

Cushing, J.T. and McMullin, E. (eds) (1989) *Philosophical Consequences of Quantum Theory*, South Bend, IN: University of Notre Dame Press.(Readable collection of articles on the Bell theorem representing many different philosophical points of view.)

Einstein, A., Podolsky, B. and Rosen, N. (1935) 'Can Quantum-Mechanical Description of Physical Reality be Considered Complete?', *Physical Review* 17: 777-80. (The famous EPR paper referred to in §1.)

Fine, A. (1982) 'Hidden Variables, Joint Probability and the Bell Inequalities', *Physical Review Letters* 48: 291-5. (Elementary calculations showing the probabilistic content of the Bell inequalities; §2.)

Fine, A. (1996) *The Shaky Game: Einstein, Realism and the Quantum Theory*, Chicago, IL: University of Chicago Press, 2nd edn.(Chapter 4 contrasts Bell with Einstein on locality, and discusses the prism models mentioned in §3. Chapter 9 considers the impact of the Bell theorem on quantum realism.)

Fine, A. (1997) 'Contextualism, Locality and the No-Go Theorems', in M. Ferrero and A. van der Merwe (eds) *New Developments on Fundamental Problems in Quantum Physics*, Dordrecht: Kluwer, ch. 16.(Contests the claim that nature is nonlocal, §1.)

Greenberger, D.M., Horne, M.A., Shimony, A. and Zeilinger, Z. (1990) 'Bell's Theorem without Inequalities', *American Journal of Physics* 58: 1,131-43.(Elementary exposition of the GHSZ theorem of §4.)

Hardy, L. (1993) 'Nonlocality for Two Particles Without Inequalities for Almost All Entangled States', *Physical Review Letters* 71: 1,665-8. (The source of the strict correlation example in §2.)

Kochen, S. and Specker, E.P. (1967) 'The Problem of Hidden Variables in Quantum Mechanics', *Journal of Mathematics and Mechanics* 17: 59-87.(The original of the Kochen-Specker theorem, §3.)

Mermin, N.D. (1990a) *Boojum's All the Way Through*, Cambridge, Cambridge University Press.(Chapter 12 is an informative elementary account of the Bell theorem.)

Mermin, N.D. (1990b) 'Quantum Mysteries Revisited', *American Journal of Physics* 58: 731-4. (Elementary account of the GHSZ theorem of §4.)

Mermin, N.D. (1993) 'Hidden Variables and the Two Theorems of John Bell', *Reviews of Modern Physics* 65: 803-15.(Comprehensive review of the Bell-Kochen-Specker theorem and recent simplifications, §3.)

Mermin, N.D. (1994) 'Quantum Mysteries Refined', *American Journal of Physics* 62: 880-7. (Elementary account of the Hardy example of §2.)

Redhead, M. (1987) *Incompleteness, Nonlocality and Realism*, Oxford: Clarendon Press.(Background and details on issues surrounding the Bell theorem.)

Benjamin, Walter (1892-1940)

Walter Benjamin was one of the most influential twentieth-century philosophers of culture. His work combines formal analysis of art works with social theory to generate an approach which is historical, but is far more subtle than either materialism or conventional Geistesgeschichte *(cultural and stylistic chronology). The ambiguous alignment of his work between Marxism and theology has made him a challenging and often controversial figure.*

1 Life and works

Benjamin was born into an affluent family of assimilated Berlin Jews. He wrote his doctorate on *Der Begriff der Kunstkritik in der deutschen Romantik (The Concept of Art Criticism in German Romanticism)*, and, in 1925 submitted *Ursprung des deutschen Trauerspiels* (*The Origin of German Tragic Drama*) for the degree of habilitated doctor at the University of Frankfurt. This book is now considered a classic. The application failed, however, and Benjamin abandoned his plans for a university career. After some years as a *feuilleton* journalist in Berlin, during which he met and worked with Bertolt Brecht, Theodor Adorno and other left-wing intellectuals, Benjamin was forced to flee to Paris in 1933. Under commission from the 'Institute for Social Research' (that is, the 'Frankfurt School' then in emigration in New York), he devoted himself to a major theoretical and historical project on nineteenth-century Paris (the 'Arcades Project'). After a financially and personally precarious decade, he was again forced to flee from the Nazis, and eventually took his own life after crossing the Pyrenees in a vain attempt to reach safety in Spain. Benjamin was little known during his lifetime; since 1955, however, under the stewardship of such erstwhile associates as Adorno, his work has been widely published and translated.

2 Art: Nietzsche and Marx

The agenda for modern aesthetic theory was set by Friedrich Nietzsche (§2), who believed that art expressed a realm more fundamental and constitutive than that accessible to the natural sciences. His many followers were happy to invoke such a superior legitimation for the human sciences. They were resisted by Marxists who, while assigning art a place in history, none the less insisted that this history was exclusively political, and that the nature of art was exhausted by determining the side it joined in the political struggle. Art, in other words, was not constitutive even of its own reality, but was merely a 'superstructural' reflection of the political 'base'.

Benjamin inclined by temperament and association towards the Marxists, not least because the most interesting art of the first third of the twentieth century was Marxist, at least by declaration (for example, Soviet Constructivism, Sergei Eisenstein, Brecht and the Bauhaus). It was clear, however, that 'base-superstructure' model of orthodox Marxism could scarcely account for such creativity.

Benjamin's project may be understood as an attempt to uncover the manner in which art engages in spheres describable in the terms of political economy, but autonomously and without adhering to simplistic criteria, such as 'progressive'. By abandoning the Marxist categorization of art as mere epiphenomenal superstructure, then, Benjamin accepts elements of Nietzsche's metaphysics.

3 Symbolism, melancholy and politics

Benjamin's thought runs through two phases. In his earlier work, which included the important essay 'Goethes Wahlverwandtschaften' (Goethe's Elective Affinities) (1922) and culminated in *The Origin of German Tragic Drama*, Benjamin is concerned to explore the manner in which art adopts pragmatic stances. His initial target is what he calls the 'symbolist' approach to art: the view, whether asserted by critics or implied by art works themselves, that art makes magical contact with essential structures of reality. It may issue (as it did with Goethe) in a superstitious fatalism, or (as in certain seventeenth-century dramas) in a naive faith in the capacity of art mimetically to capture God's creation.

Diametrically opposed to this stance is what Benjamin terms 'melancholy', a scepticism about the claims of science and empirical knowledge. The melancholic artist devises allegories and conceits to emphasize their despair at the inaccessibility of God's reality; the Baroque *Trauerspiel* is a typical example of this attitude. However, this is a rash response to the problems of mimetic realism, or 'symbolism', for there is a third possibility available to artists: an interventionist pragmatism. This depends on their ability to perceive their own activities within a wider,

political frame. In Benjamin's view, if they can do this, they will 'awaken under the open sky of history'; but the precise nature of interventionist art is something that, in the early work, still remains obscure.

4 Technology

The latter part of Benjamin's work - from the late 1920s onwards - was concerned to delineate with more precision how art assumes a political identity. This was a matter of describing how art manifested itself in the public arena at all, and how it assimilated itself, deliberately or not, to the conflicts dominating that arena. This discussion has two aspects: the theory of technology and the theory of history.

Benjamin's most important essay on art and technology is 'Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit' (The Work of Art in the Age of its Mechanical Reproducibility) (1935). He argues that there is a general tendency away from the 'auratic'. Under more primitive social conditions, art primarily performs a ritual function, for example, a symbolization of the divinity. It has a high 'cultic' value. This, however, goes hand in hand with low public availability; cultic or auratic works of art retain or increase their power by being confined to inaccessible ritual spaces. In the modern era, despite a decline of express ritual, aura is mimicked by high culture, which favours works that can be restricted to an elite (in museums, concert halls and opera houses).

Because this 'auratic' approach evades the issues of contemporary history, however, it scarcely deserves the designation of culture. Proper art has - indeed, has always had - its vehicles for engaging people in general. These are, nowadays, the instruments of mass dissemination. They have two aspects. In the first place, modern art dispenses with the notion of the unique object, invested with auratic magic. Modern art may be reproduced without losing its identity - as one sees in the case of film and photography and in all electronically storable works. In the second place, art loses its finality; it becomes part of a process of revision, testing and provisional application. Brecht's collaborative and plagiarizing treatment of art is one example; the collective contribution necessary for any movie is another. Because of its provisionality, art becomes subject to the interventions of many; and because of that, in turn, it becomes integrated into wider structures of social dynamics, such as politics.

5 History

The theory of history is the topic of Benjamin's last work, 'Über den Begriff der Geschichte' (On the Concept of History) (1940). Just as his aesthetics break with representationalism of orthodox Marxism, so his notion of history turns away from its characteristic faith in progress. The course of history becomes radically disrupted: that is to say, history is understood as being in a permanent state of emergency, where identities emerge only through isolated and contingent acts of struggle. The 'meaning' of history ceases to be theoretically recoverable, and yields only to redemptive recollection. From this perspective, the task of the historian is to cite those struggles for freedom which, by example or analogy, cast light on the conflicts of the present.

Benjamin represents this view as Messianism and thus as a form of theology. It should be noted, however, that the basic model applies also to common-law notions of legal precedent; this important work is not confined to religious realms, but, as with all of Benjamin's writings, has wide and continuing secular relevance.

See also: Frankfurt School §3

JULIAN ROBERTS

List of works

Benjamin, W. (1974-88) *Gesammelte Schriften*, ed. R. Tiedemann and H. Schweppenhäuser, Frankfurt am Main: Suhrkamp.

Benjamin, W. (1922) 'Goethes *Wahlverwandschaften*' (Goethe's Elective Affinities), in *Gesammelte Schriften*, vol. 1, 123-202.(A riposte to the conventional hagiographies of Goethe.)

Benjamin, W. (1935) 'Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit' (The Work of Art in the Age of its Mechanical Reproducibility), in *Gesammelte Schriften*, vol. 1, 431-508.(Benjamin's most famous work, and the source of his 'aura' theory. A translation is available in Benjamin (1968).)

Benjamin, W. (1940) 'Über den Begriff der Geschichte' (On the Concept of History), in *Gesammelte Schriften*, vol. 1, 691.(The 'Theses on History' which allegedly document Benjamin's turn from Marxism.)

Benjamin, W. (1968) *Illuminations*, ed. H. Arendt, New York: Fontana.(Contains works useful for a first encounter; the selection is tendentious, however, and the translation unreliable.)

Benjamin, W. (1973) *Understanding Brecht*, London: Verso.(An accessible exploration of Marxist cultural theory.)

Benjamin, W. (1977) *The Origin of German Tragic Drama*, London: Verso.(Benjamin's most challenging work.)

References and further reading

Buck-Morss, S. (1990) *The Dialectics of Seeing. Walter Benjamin and the Arcades Project*, Cambridge, MA: MIT Press.(A useful discussion of Benjamin's major project.)

Marcus, L. and Nead, L. (eds) (1993) 'The actuality of Walter Benjamin', special issue of *New Formations* Summer 20.

Bartram, G. (ed.) (1994) 'Walter Benjamin in the Postmodern', special issue of *New Comparison* Autumn 18.

Roberts, J. (1982) *Walter Benjamin*, London: Macmillan.(An intellectual and personal biography.)

Bentham, Jeremy (1748-1832)

Jeremy Bentham held that all human and political action could be analysed in terms of pleasure and pain, and so made comprehensible. One such analysis is how people actually do behave; according to Bentham, seeking pleasure and avoiding pain. Another such analysis is of how they ought to behave. For Bentham, this is that they should maximize utility, which for him is the same as producing the greatest happiness of the greatest number, which, again, is the same for him as maximizing pleasure and minimizing pain. His chief study was planning how there could be a good system of government and law; that is, how laws could be created so that people being as they actually are (seeking their own pleasure) might nevertheless do what they ought (seek the greatest pleasure of all). The instruments which government use in this task are punishment and reward, inducing action by threats and offers. For Bentham, punishment is done not for the sake of the offender, but to deter other people from doing the same kind of thing. Hence on his theory it is the apparent punishment which does all the good, the real punishment which does all the harm.

Bentham thought that the primary unit of significance was the sentence, not the word. He used this idea to produce profound analyses of the nature of law and legal terms, such as' right', 'duty' or 'property'. These are what he calls names of fictions - terms which do not directly correspond to real entities. However, this does not mean that they are meaningless. Instead, meaning can be given to them by translating sentences in which they occur into sentences in which they do not occur. Thus legal rights are understood in terms of legal duties, because sentences involving the former can be understood in terms of sentences involving the latter; these in turn can be analysed in terms of threats of punishment or, again, pleasure and pain. This gives sense to legal rights, but sense cannot be given in the same way to natural rights. For Bentham, we have no natural rights and the rights that we do have, such as property rights, are created by government, whose chief task is to protect them. Bentham also worked out how people could be protected from government itself, designing an elaborate system of constitutional law in which representative democracy was a central element.

Bentham invented the word 'international', and when he died he had an international legal and political influence. His chief influence in philosophy has been as the most important historical exponent of a pure form of utilitarianism.

1 Life and writing

Writing was the centre of Bentham's life. He shut himself away in remote cottages and even when in London described himself as a 'hermit'. Increasingly, the hermit merely produced large sheets of manuscript, and the task of selecting from these and turning them into books was left to others, such as the young John Stuart Mill, who produced five large volumes of Bentham's thought on evidence from a much larger mass of nearly illegible manuscript. When Bentham died he left 70,000 sheets of foolscap manuscript behind him - theoretical work, but also highly detailed designs for states, prisons, banknotes, and much else. His principal writings on language, ontology and the philosophy of law were only published posthumously.

Jeremy Bentham was born in London on 15 February 1748. He was the son and grandson of lawyers and was educated to follow them making money from the practice of law. However he soon became revolted at the current condition of the law and so, instead of making money from it, devoted the rest of his life to a study of how it could be improved. He started to design a perfect penal code; then diverted to write a criticism of the leading current legal thinker, William Blackstone; then diverted again from the main body of this criticism to write a lengthy refutation of one of Blackstone's digressions. This was published as *A Fragment on Government* (1776). He returned to working on the principles of penal legislation and printed the main part of his introduction to them in 1780. However, seeking to work out the identity conditions for a single law, he became entangled in a 'metaphysical maze' which meant that he had to lay the uncompleted book aside. He buried himself away and produced his main work on the philosophy of law, only published a century after his death. The work laid aside was finally published in 1789 as *An Introduction to the Principles of Morals and Legislation*. It had a new preface, but it was neither complete in itself nor accompanied by the worked-out penal code to which it was meant merely to be the introduction.

Since the *Introduction* and the *Fragment* are Bentham's two best known works, it is worth noting that they are

both parts of much larger uncompleted works. The *Fragment* is a fragment, the *Introduction* an introduction, and much of his most profound thought of the time, if it gets in at all, only makes the footnotes. They are also both relatively early works. However they do have the advantage, unlike most of what followed, that they were published by Bentham himself rather than by one of his disciples.

While the *Introduction* lay fallow, printed but not published, Bentham switched into writing in French with the hope of interesting Catherine the Great of Russia in his proposals for legal reform. He visited Russia, where his brother was living, and found a suitably remote cottage. He wrote on the principles of a civil code and of reward. Yet even when Catherine was near, he remained in his seclusion and failed to exert any influence. On his return to England he finally published the *Introduction*. However, it was the year of the French Revolution; public attention was elsewhere; and the work was half consumed by rats. Bentham bombarded the new French revolutionary government with proposals which had no effect beyond his being created an honorary citizen of their new republic. Then, in the chief diversion of his life from writing, he turned his attention to pushing the British government for a contract to build and manage a panopticon prison.

The idea of the panopticon, a circular building in which the unseen overseer in the centre would observe the inmates, derived from Bentham's brother in Russia. At first Bentham wrote about it relatively light-heartedly as a 'simple idea in architecture' which would solve all manners of different problems (such as allowing Turkish seraglios to be run with fewer eunuchs) (Bentham 1791). Editions of the work were printed in Dublin and Paris. A circular prison was built in Edinburgh. However, when Bentham tried to obtain a contract from the British government to manage the country's prisoners, matters became much more serious. He bought the iron and other materials, nearly ruined himself, and energetically lobbied the government for permission to build his 'mill for grinding rogues honest'. Several times in the 1790s it looked as if he would succeed; but the project was effectively over by 1803 (although not completely killed until 1813).

While Bentham was trying to get his panopticon contract, Etienne Dumont was working on his manuscripts. In 1802 Dumont produced the three volumes which first made Bentham's name, the *Traités de législation civile et pénale* (1802). Dumont moved on to editing other Bentham manuscripts, while Bentham returned to full-time production of the supply. In the 1800s Bentham wrote on economics, evidence and judicial organization. In the next decade he expanded into logic, language, ontology, and criticisms of the religious and legal establishment. His central concern became constitutional law. He offered to draft constitutional codes for all nations, and worked hard on the ideal code. This and pamphleteering for a more democratic government were his chief concerns in the 1820s. His reputation was now established. Newly created countries consulted him. Dumont's work was translated into English. A team of disciples produced other work from the manuscript. The *Fragment* and the *Introduction* were republished in second editions. A Benthamite journal was founded. His place in philosophy was secured, his influence being transmitted most of all by J.S. Mill, who was as a young man a great admirer and Bentham editor, and who arranged meetings of the younger utilitarians in Bentham's house. Bentham died in London on 6 June 1832.

2 The principle of utility

Both in the *Fragment* and the *Introduction* Bentham calls his central normative principle the 'principle of utility'. Utility provides the 'standard of right and wrong'. 'By the principle of utility', he says in the *Introduction* (I 2), 'is meant that principle which approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment or diminish the happiness of the party whose interest is in question'. Utility is therefore to be understood in terms of happiness; and the *Fragment* starts with the famous formula, declared as an 'axiom', that 'it is the greatest happiness of the greatest number that is the measure of right and wrong'. This formula reappears in his late work, and is stated in the *Constitutional Code* (1830), as the proper end of government. In this later work Bentham came to prefer 'the greatest happiness principle' as the best description of his central principle. However, in spite of Bentham's preference, the term 'utility' has stuck, and Bentham is normally thought of as a utilitarian. Indeed the prescient young Bentham dreamed once that he founded a 'sect of utilitarians'.

He founded a sect, but neither utility nor happiness originate with Bentham. They were plucked from the surrounding Enlightenment air. Bentham says that he used 'utility' because of Hume, and the famous 'greatest happiness' formula appears in the Italian legal theorist Cesare Bonesana Beccaria, whom Bentham much admired. Originality was not the point; indeed the more agreed, or hackneyed, the statement of the final goal was, the better.

Bentham's purpose was to work out in detail the means by which this goal could be achieved.

The greatest happiness is therefore the appropriate end of action. The next question is whose happiness. The famous formula says 'of the greatest number'; but this could render the formula indeterminate between recommending that the greatest happiness be achieved and recommending that the greatest number get happiness. In fact Bentham makes it clear that he always means the former: happiness is to be maximized whomever it may belong to. He therefore occasionally recommends omitting the 'of the greatest number' part of the formula, noting that otherwise the slight pleasures of a majority would count more than the severe pains of a minority.

'Pleasure' and 'pain' seem to introduce yet more alternative sources of value. However this promiscuous use of terms does not for Bentham indicate any diversity or conflict of values. For him these terms are all convertible. He says in the *Introduction* (I 3) that 'benefit, advantage, pleasure, good, or happiness' all 'come to the same thing' as also do their opposites, 'mischief, pain, evil, or unhappiness'. 'Interest' is also to be understood in these terms; something is in someone's interest when it tends to increase their pleasure. Bentham's is a monistic and consequentialist system of value. There is only one ultimate value, although it may have different descriptions, and actions should be done so as to bring about those states of affairs which have most of this value.

Bentham usually takes the proper aim of government to be concern with the greatest happiness of the people composing the country for which it is the government. This is less than all the people there are, which is the standard universalistic sense given to the ground principle of utilitarianism. This might be taken as a merely practical recommendation, so that Bentham is still supposed to have universal happiness as his end but is taken to be claiming that this is best achieved if every government restricts itself to maximizing the happiness of its own people. Or it could be taken, as David Lyons (1973) holds, to be explicitly intended by Bentham as a necessary feature of his brand of utilitarianism.

Another claim of Bentham's would, conversely, give his principle a wider scope than the standard utilitarianism. For, as he puts it in the *Introduction* (XVII 4.n), 'the question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?' So Bentham's utilitarianism goes beyond people and extends to all sentient creatures; animals also count.

In these statements of Bentham's ground principle there is constant use of the word 'tend'. Hence application of them inevitably has a generalizing effect. The precise utilities which follow on a single occasion of action are not as important as the general tendencies of that kind of action.

3 Duty and interest

At the start of chapter one of the *Introduction* (I 1), Bentham says that we are under the 'governance of two sovereign masters, *pain* and *pleasure*', and adds, 'it is for them alone to point out what we ought to do, as well as to determine what we shall do'. Two different things are involved here. On the one hand there is a standard of right and wrong determining what people ought to do. On the other hand there is a psychology of human action determining what they will actually do. Bentham is writing for legislators who have the task of bringing about the goal of maximizing happiness while working with people as they actually are. The legislator needs therefore to understand both the goal and also how people are; both value theory and psychology. These are both treated in the *Introduction*, but much more space is devoted to psychology than to value theory. Its central idea is that people seek their own happiness. Preventing pain and providing pleasure are the ways in which people, as they actually are, can be influenced in their actions.

'The state', says Bentham, 'has two great engines, *punishment* and *reward*' (*Introduction*: XVI 18). This does not mean that they have to be used. Follower of Adam Smith, Bentham realized that the desired end could often be achieved by leaving people alone to get on with their lives. So his chief economic prescription to the legislator is 'be quiet'. He even out-Smithed Smith in an early work, *The Defence of Usury* (1787), arguing against Smith's claim that interest rates should be controlled. However, interference is not ruled out in principle. It all depends upon what is needed to maximize happiness. Bentham was prepared, for example, to allow the control of corn prices in times of starvation; and he believed that relief of extreme poverty should be the concern of the state rather than being left to private charity.

People are moved in many ways. Bentham (1776, 1789) lists the 'sanctions', as he calls them, which operate on

them. The central one he is concerned with, the great engines of punishment and reward, he calls the 'political' (sometimes 'legal') sanction. There is also the 'moral or popular' sanction, which is Bentham's name for public opinion. Then there is the religious sanction, the force exerted by God's displeasure. These are the only ones mentioned in the *Fragment*, but Bentham brings it up to four in the *Introduction* by adding the 'physical' sanction: purely natural processes which influence behaviour (such as the pain of falling out of a window). Later he added the 'sympathetic' sanction, where people are moved merely by the pleasures and pains of others, but which he thinks is much weaker than the others.

If people are going to do the right thing under the operation of one of these other sanctions, the legislator does not need to interfere with them by use of the political sanction. The legislator may, however, seek to strengthen the other sanctions by what Bentham calls 'indirect legislation' - for example, by education. Sometimes, however, more direct action is required. In spite of the non-political sanctions, people seeking their own pleasure sometimes cause greater unhappiness to others and so diminish overall happiness. Then the legislator has to interfere, deploying the political sanction, and threatening punishment. By having a code of criminal law, which announces in advance the kinds of penalties that are attached to particular kinds of behaviour, the legislator changes the payoffs. Self-interested people who might otherwise have done these things are now deterred by the threat of punishment.

For Bentham all punishment is a pain. It is therefore, for him, an evil. Its justification is therefore indirect: it is a present harm done so that good may come (rather like drilling a tooth to prevent future toothache). In fact, as Bentham points out, it is the apparent pain which does all the good, the real pain which does all the harm. So if people could only appear to be punished, this would be even better. Punishment is not done for the sake of the offender, as in retributive theories, but for the future benefit of others. The amount of punishment is accordingly fixed as the minimum amount which is necessary to deter a sufficient number of similar actions by others; it does not depend directly on the seriousness of the offence.

The legislator can only calculate this amount with knowledge of psychology - that is, of how a typical individual calculates the value to themselves of a portion of pain or pleasure. Bentham thinks that the chief factors involved are, in his words, intensity, duration, certainty (or probability), and propinquity. That is, as well as obvious facts like the length or unpleasantness of punishment, the deterrent effect also depends on how imminent the threatened punishment is and how likely someone is to be caught.

Bentham developed this psychology in his *Civil Code* writings (1802), where he laid down what he calls 'axioms of mental pathology' (this is part of the material that was first published by Dumont in the *Traités*). Crucial among these is that equal increments of a good do not produce equal increments of happiness. In other words, the utility produced by goods diminishes at the margin. Therefore, other things being equal, happiness is maximized by an equal distribution of goods.

When Bentham moved on to design of prisons and poorhouses, and eventually to constitutional law, the root idea was similarly to distinguish between the *is* and the *ought* of human action, and then make them work together. The idea is, as he puts it in his work *Pauper Management Improved* 'to make it each man's interest to observe on every occasion that conduct which it is his duty to observe' (II iv. 2), which Bentham calls the 'duty and interest junction principle'. Similarly, the ground idea of the *Constitutional Code* is to construct a system of political offices, such that each office can be occupied by purely self-interested people who will nevertheless be led to behave as they ought to behave if the system as a whole is to deliver good government (that is, promote the greatest happiness).

4 Public reason

Both the prisoners in the panopticon and the guard in the centre will be led to behave well by the force of publicity. Publicity solves the old problem of who is to guard the guards. The same applies in the *Constitutional Code*, where a central role is given to what Bentham calls the 'public opinion tribunal'. Similarly, judges are to be forced to give public reasons for their decisions. This, he thinks, leads to better law.

So far, this is deployment of the 'popular' sanction. 'The eye of the public', as Bentham puts it, 'makes the statesman virtuous' (1843, vol. 10: 145). But it is also centrally connected with Bentham's espousal of his chief evaluative principle. For he thinks that only utility (and its cognates) are appropriate for use in such public justification. In the utility chapter of the *Introduction* (I 2) Bentham allows that 'that which is used to prove

everything else, cannot itself be proved'; but he does add some considerations designed to make people 'relish' the principle of utility. Chief among these is that there would not otherwise be any public standards of justification; anyone's opinions would be worth just as much or as little as anyone else's; and this would be 'despotical' if someone imposed them, 'anarchical' if not. In other words, unless all argument is to be 'at an end', utility must be taken as the standard. Only utility can be used in public reasoning.

Bentham puts this as a way of giving a 'meaning' to 'the words *ought*, and *right* and *wrong*, and others of that stamp' (I 10). His concern, here and elsewhere, is to give meaning; to clarify; to make things comprehensible. The chief justification of his foundational evaluative principle is that, by connecting *right* and *wrong* with pleasure and pain, it gives them public meaning.

5 Fiction and paraphrasis

Bentham carries his clarificatory mission into the centre of the law. His aim is that the central legal terms, such as 'right', 'duty', or 'property', should be understood. To make something clear for Bentham is to connect it with perception. These legal terms are names of what he calls fictional entities; and fictional entities are understood by connecting them with perceivable real entities like pleasure and pain.

For terms like 'duty' or 'obligation' Bentham invented the technique he called *paraphrasis*, most fully described in his *Essay on Logic*. For Bentham the primary unit of meaning is the sentence rather than the word, and he uses this insight to relate the tricky legal terms to perception. They do not themselves refer to things that can either be perceived or directly inferred from perception. However, Bentham's proposal is that if the difficult term is placed in a sentence, the whole sentence may then be given meaning by being translatable into another sentence the words of which can be more easily understood.

Armed with his technique of paraphrasis, Bentham accordingly analyses fundamental legal terms. Rights, for example, are analysed in terms of duties. That is, a sentence about rights, such as ‘John has a right to wear his coat', can be translated into a sentence about duties, such as ‘Everyone has a duty not to prevent John wearing his coat'. We are now still in the realm of fiction. But Bentham proceeds to the analysis of duty. Someone is said to be under a duty when they are threatened by punishment for non-performance. But the threat of punishment is the threat of pain. So, at last, we reach pain, an immediate object of experience; the law has been clarified and made comprehensible to all. As he puts it in the Preface to the *Fragment*, '*pain* and *pleasure*, at least, are words which a man has no need, we may hope, to go to a Lawyer to know the meaning of'.

For Bentham, rights are the benefits created by the imposition of duties. He analyses the varying kinds of rights according to the different kinds of duties, and in an embryonic deontic logic brings out the different connections between obligation and permission. Thus analysis of fictions (such as Bentham conducts in his *Of Laws in General* (1970)) gives clarification. The point of calling them 'fictions' is not to designate them as merely imaginary items which can be disregarded. Such things (ghosts; the pagan gods) Bentham calls, by contrast, 'fabulous' entities. Legal rights and obligations have what he calls a 'verbal' reality, and the paraphrastic analysis shows what this consists in: if I disobey a (verbally real) obligation, then I am liable to real pain.

Bentham holds that fictions are necessary for the use of language, yet he also sometimes uses 'fiction' in a merely pejorative way to stigmatize something as merely imaginary; so care has to be taken as to which use he intends on a particular occasion. He is being merely pejorative in the *Fragment* when he talks of the supposed original contract, which was meant to justify political obligation, as the 'sandy foundation of a fiction', or when he talks of 'pestilential breath of Fiction' poisoning the operations of law. In these cases he is talking about what he calls the lies of lawyers - of justification on the basis of purely imaginary happenings.

Bentham's account of natural rights illustrates both his techniques of analysis and this problem. 'From *real* laws come *real* rights', he says, 'from *imaginary* laws come imaginary ones' (*Anarchical Fallacies*, conclusion). Natural rights are just fictions. But, it might be objected, since all rights are fictions for Bentham, why can natural rights not also be given a meaning by paraphrastic analysis? The answer is as follows. Since (sentences about) rights can be analysed in terms of (sentences about) duties, the problem is not with natural rights, as such. They can be analysed in terms of natural duties. The problem is with the next step, where duties are analysed in terms of threats of pain (punishment). With the supposed natural duties there is, for Bentham, no such threat. There is no legislator. As he puts it in *Supply without Burthen* (1795), ‘a natural right is a son that never had a father' (1952,

vol.1: 334).

For Bentham, people talking of natural rights are really saying that they wished that there were (real, political) rights. But, as he says, 'want is not supply; hunger is not bread'. Even worse, for him, is when people suppose, as did the framers of the French *Declaration of the Rights of Man and of the Citizen*, that such rights are unalterable (impresciptible). As Bentham famously puts it, 'natural rights is simple nonsense, natural and impresciptible rights, rhetorical nonsense, nonsense upon stilts' (*Anarchical Fallacies* art. II).

6 Government

So, in Bentham, for real rights we need real laws. We need government. The argument for the states and for government is therefore straightforward: they confer the benefits we gain from the possession of rights. Of these benefits the chief, for Bentham, is security; and hence the principal task of government is to provide for the security of individuals. Possessing security, they can plan ahead in confidence, realize their plans, and increase their happiness. Bentham's main concern is to protect areas in which individuals may maximize their own utilities rather than having a government that constantly interferes to promote happiness. In this sense he is on the side of liberty - although this is a liberty which is only produced by government and does not predate it in the way that supposed natural liberty would.

Property is included by Bentham under security. Again, it is the creation of law and does not predate it. This means that Bentham avoids the problems about taxation or obedience to the state which philosophers run into when they start with natural rights to property. With John Locke we get an original contract argument for government, whereby people antecedently having property contract into government. Bentham mocks such arguments in the *Fragment*. For him justification of obedience is a matter of utilitarian calculation, of whether the 'probable mischiefs of obedience are less than the probable mischiefs of resistance' (I 43). Analogously, there are no problems about taxation. Government creates property and it can remove the property it creates.

As well as security, Bentham lists three other 'subordinate ends' of government in the *Civil Code*: subsistence, abundance, and equality. The utilitarian argument for equality was noted above; and both subsistence and abundance are naturally positively correlated with utility. They are therefore appropriate goals for the legislator, although it should be noted that this emphasis upon subordinate ends shows Bentham's indirect utilitarian aspect: the legislator protects security or provides subsistence rather than directly promoting utility. Security easily outranks the others, being thought by Bentham a necessary condition for their achievement.

When Bentham was writing the *Civil Code* he was happy to appeal to enlightened dictators to get his legal proposals put into effect. However he came to realize that the chief power against which people need security is the power of government itself. Hence his turn in his later writings to the construction of a constitutional code; and hence also his turn to democracy.

The argument for democracy follows simply from the central principles outlined above. Since, by the principle guiding human action, people tend to act in their own interest, so also do governors. Kings look after the interests of kings; oligarchs after the interests of the oligarchy; and so on. Yet the proper end of government is not just a sectional but a general interest. It is general happiness. The solution is to make the governors as far as possible the people themselves. The greatest happiness of the greatest number is safest in the hands of the greatest number. Merely by following their own interest, in accordance with the chief factual principle, they will also promote the general interest, in accordance with the chief evaluative principle.

The kind of democracy promoted is representative democracy. Bentham distinguishes between 'constitutive' and 'operative' powers. The people as a whole are to be the supreme constitutive power, electing governments and having final authority. But they elect, and can dismiss, the operatives. The operatives have to be controlled. Hence the *Constitutional Code*. As before, the guards are to be guarded. Politicians are to be forced by all the available sanctions so that, acting merely selfishly, they in fact promote general happiness.

See also: Democracy; Happiness; Law, philosophy of; Mill, J.S.; Utilitarianism

ROSS HARRISON

List of works

Bentham, J. (1843) *The Works of Jeremy Bentham*, ed. J. Bowring, Edinburgh, 10 vols.(The original collected edition, gradually being replaced, but still the source for *Essay on Logic* and *Essay on Language*, both in volume 8.)
(1776) *A Fragment on Government*, Cambridge: Cambridge University Press, 1988.
(1789) *An Introduction to the Principles of Morals and Legislation*, London: Athlone Press, 1970.
(1791) *The Panopticon Writings* London: Verso, 1995.(A selection which also contains *Fragment on Ontology*; the complete *Panopticon* writings are only in Bentham (1843).)
(1802) *Traités de législation civile et pénale*, Paris; greater part translated in J. Bentham, *The Theory of Legislation*, London: Kegan Paul, 1932.
(1817) *Chrestomathia*, Oxford: Oxford University Press, 1983.(Appendix 4 contains considerable material on language and fictions.)
(1830) *Constitutional Code*, *Vol. 1*, Oxford: Oxford University Press, 1983.
(1952) *Jeremy Bentham's Economic Writings*, ed. W. Stark, London: Allen & Unwin. (Contains *Defence of Usury* and *Supply without Burthen*, mentioned in the main text, as well as other economic writings.)
(1970) *Of Laws in General*, London: Athlone Press.
(1973) *Bentham's Political Thought*, ed. B. Parekh, New York: Barnes & Noble.(Wide selection of useful extracts edited from the original manuscript, including the *Anarchical Fallacies*.)

References and further reading

Harrison, R. (1983) *Bentham*, London: Routledge.(General introduction to all the philosophical aspects of Bentham.)
Hart, H.L.A. (1982) *Essays on Bentham*, Oxford: Oxford University Press.(Essays on Bentham's philosophy of law; moderately difficult.)
Kelly, P.J. (1990) *Utilitarianism and Distributive Justice*, Oxford: Oxford University Press.(Uses the *Civil Code* material to argue for an indirect utilitarianism account of Bentham.)
Long D. (1977) *Bentham on Liberty*, Toronto Ont.: University of Toronto Press.(Particularly full on the not yet published early manuscript material.)
Lyons, D. (1973) *In the Interest of the Governed*, Oxford: Oxford University Press, revised 1991.(Innovative reading of Bentham claiming that he held that people should seek particular rather than the universal interest.)
Postema, G. (1989) *Bentham and the Common Law Tradition*, Oxford: Oxford University Press.(A fuller account than the title suggests. A central problem considered is Bentham's thought about legal adjudication.)
Rosen, F. (1983) *Jeremy Bentham and Representative Democracy*, Oxford: Oxford University Press.(Accessible description of Bentham's later political thought, including the *Constitutional Code*.)

Bentley, Richard (1662-1742)

A towering figure in the history of textual criticism, Bentley's importance in the development of English philosophical thought rests on both public and private achievements. His great public contribution was made in the first Boyle lectures of 1692. In private, his correspondence with Newton sought the great man's blessing on the arguments and opinions advanced in these lectures, persistently questioning him on the possibility of a natural origin for the universe and on the role and nature of gravity in physics in general. Bentley's influence helped to establish the Newtonian consensus dominant in Europe until the end of the nineteenth century.

Richard Bentley was born into an English family of modest means in Oulton, Yorkshire. He entered St John's College, Cambridge in 1676. He later came under the patronage of Bishop Stillingfleet and served as tutor to his family. He took holy orders, his scholarly eminence resulting in a rapid rise to the post of Keeper of the Royal Libraries and election to the Royal Society in 1694. For an advocate of religion as a source of peace and concord, Bentley managed to provoke a great deal of acrimony. A famous dispute with Oxford scholars regarding the relative merits of the Ancients and the Moderns led to his great work of textual criticism - *Dissertation on the Epistles of Phalaris* - demolishing the claims to authenticity of the epistles of 'Phalaris'. In 1699 he became Master of Trinity College, Cambridge, a post he retained until his death. His despotic rule was not to the liking of the Fellows, whose unsuccessful efforts to remove him involved two trials before the Bishop of Ely.

The annual courses of lectures endowed by Robert Boyle became the main public forum for the exposition and defence of the Newtonian world picture as a basis for latitudinarian natural religion (see Latitudinarianism). Bentley gave the first series in 1692. The picture of a rationally ordered universe, in which the designing and maintaining hand of God could be discerned, was presented as a model for a rationally ordered social universe, providentially arranged for our delight and wellbeing. This happy state of affairs would be much facilitated, Bentley implied, if we followed the teachings of the Anglican 'broad church' party. According to Bentley (1692 vol. 3: 13), 'Religion itself gives us the greatest delights and advantages even in this life also, though there should prove in the event to be no resurrection to another'. It is significant that Bentley entitled his lectures *The folly and unreasonableness of atheism* (1692).

Bentley singled out gravity as peculiar evidence of the direct and arbitrary action of a wise creator. In an exchange of letters he pressed Newton on the role of gravity in the origin of the universe as we now observe it, and the possibility of a purely mechanical explanation of the origin of the material world. In his response of 10 December 1692 to Bentley's first letter Newton expresses a firm commitment to the principles of natural religion: 'I had an eye,' he says 'upon such Principles as might work with discerning men for the belief of a Deity' (Turnbull 1961). Bentley realized that the clumping of matter into discrete material bodies out of a primeval atomic chaos by any natural force, such as universal gravity, was implausible when the extremely low density of such a state was considered. Newton's first response was to distinguish between the general problem of the condensation of matter and particular difficulty of the separation of 'shining' matter into stars and 'opaque' matter into planets, without, as he says, the work of 'a voluntary Agent'. Bentley expressed his general point as follows: 'No Quantity of common Motion could ever cause those straggling Atoms to convene into great masses' (Letter of 18 February 1692); so there could be no natural explanation for the origin of the solar system. This provoked Newton's famous reply: 'Gravity must be Caused by an Agent acting constantly according to certain laws, but whether that agent be material or immaterial I leave to the consideration of my readers' (Letter of 25 February 1692). We can hardly doubt what he expected their opinion to be!

Of all Newton's followers it was above all Bentley and Samuel Clarke who made the Newtonian point of view available to an extensive lay audience, and, indeed created the Newtonian consensus which dominated English and later Continental thought for more than a century.

See also: Collins, A. §2

ROM HARRÉ

List of works

Bentley, R. (1838) *The works of Richard Bentley*, ed. A. Dyce, London.(These volumes are mainly concerned with

his scholarly controversies. *The Dissertation on the Epistles of Phalaris* is found in volume 1, pages 75-430 and volume 2, pages 1-181.)

Bentley, R. (1692) *The folly and unreasonableness of atheism*, London.(A comprehensive survey of the evidence for God's existence from human psychology and biology and from physics.)

References and further reading

Jacob, M.C. (1976) *The Newtonians and the English Revolution: 1689-1720*, Hassocks: Harvester Wheatsheaf.(A detailed account of the interactions between religious controversy, social theory and science in the seventeenth century.)

Turnbull, H.W. (1961) *The correspondence of Isaac Newton*, vol. 3, Cambridge: Cambridge University Press.(In this volume are collected letters to Locke, Huyghens, Leibniz as well Bentley and others, 1684-93.)

Berdiaev, Nikolai Aleksandrovich (1874-1948)

Nikolai Berdiaev, Russian religious idealist, was one of many non-Marxist thinkers expelled from Russia by communist authorities in 1922. Although attracted to Marxism in his youth, even then he tempered it with a Neo-Kantian ethical theory. Well before the Bolshevik Revolution, he became seriously disenchanted with Marxist philosophy (though not with the idea of socialism) and embarked on the career of elaborating a personalistic Christian philosophy that occupied him for the rest of his life.

Dubbed 'the philosopher of freedom', Berdiaev wrote prolifically on that subject and on related topics in metaphysics, philosophy of history, ethics, social philosophy and other fields (but not epistemology, which he rejected as a fruitless exercise in scepticism). Because his approach to philosophy was admittedly anthropocentric and subjective, he accepted the label 'existentialist' and acknowledged his kinship with Dostoevskii, Nietzsche and (to a lesser degree) Jaspers. Like them, he constructed no philosophical system, though he did expound views that were coherently interrelated in the main, if impressionistically and sometimes obscurely expressed. Among his more prominent ideas were his conception of freedom (for which he was indebted to the mystical philosophy of Jakob Boehme), his distinction between spirit and nature, his theory of 'objectification', his doctrine of creativity and his conception of time.

The most frequently translated of twentieth-century Russian thinkers, Berdiaev has been widely studied in the West since the 1930s, particularly in schools of religion and theology and by philosophers in the existentialist and personalist traditions. Although many Western readers considered him the voice of Russian Orthodox Christianity, his independent views drew fire from some Orthodox philosophers and theologians and also from strongly anti-Soviet Russian émigrés. His writings in emigration were eagerly embraced in his homeland once they could be published there, beginning in the late 1980s.

1 Life and works

Nikolai Aleksandrovich Berdiaev was born into an aristocratic family in 1874 near Kiev. A student of law at the University of Kiev, he was expelled in 1898 for his activity in radical student circles; this marked the end of his formal education, except for a semester of study with Wilhelm Windelband at Heidelberg in 1903. In 1900 Berdiaev, along with other members of the Social Democratic Party in Kiev, was banished to the northern province of Vologda. Philosophically, like many Russian thinkers of his generation, the young Berdiaev sought to complement Marxist socioeconomic views with Kantian transcendental idealism, and this effort is evident in his first philosophical book, *Sub''ektivizm i individualizm v obshchestvennoi filosofii (Subjectivism and Individualism in Social Philosophy)* (1901).

Allowed to return to Kiev in 1903, Berdiaev moved in the following year to St Petersburg and in 1908 to Moscow, where he became prominent in the lively cultural world of Russia's 'Silver Age'. The first decade of the century was a period of intense spiritual searching for Berdiaev as for many others; under the influence of a great range of thinkers, including Schopenhauer, Nietzsche, Vladimir Solov'ëv, Vasilii Rozanov, Fëdor Dostoevskii, Lev Tolstoi and Dmitrii Merezhkovskii (see Russian Religious-Philosophical Renaissance), he moved from Neo-Kantian Marxism (though without abandoning his socialist convictions, as we shall see in §5) to the religiously oriented, mystically coloured personalism that he would continue to elaborate throughout his life. His writings of the period 1907-11 may be considered transitional. It was not until 1916, in *Smysl tvorchestva (The Meaning of Creativity)* (his last philosophical book published in Russia), that Berdiaev provided a comprehensive account of the chief principles of his new worldview. This work, which Berdiaev in his autobiography called his most significant book, exhibits the impact of a major influence on his mature outlook - the mysticism of Jakob Boehme.

Berdiaev welcomed the Russian Revolution of February 1917, but he was opposed to the policies of the Bolsheviks, who seized power in October. Despite that opposition, he was able to take an active part in the cultural life of Moscow during the first years of communist rule. In 1919 he founded the Free Academy of Spiritual Culture, and in 1920 was named Professor of Philosophy at Moscow University. In 1922, however, Berdiaev and more than 100 other prominent non-Marxist intellectuals were abruptly stripped of their positions and required to leave the country, prohibited from returning on pain of death.

After a stay in Berlin, in 1924 Berdiaev with his family joined the growing Russian émigré community in Paris. He taught at the Russian Religious-Philosophical Academy there, an institution he had first organized in Berlin with the assistance of the Young Men's Christian Association (YMCA) of North America; he founded and edited (1925-40) the religious-philosophical journal *Put' (The Way)*; and he served as editor-in-chief (1924-48) of the YMCA-Press in Paris, the principal publishing outlet for émigré Russian religious philosophers.

Berdiaev won worldwide fame with his book *Novoe srednevekov'e (The New Middle Ages)* (1924), which was translated into a dozen languages. He continued to develop various aspects of his philosophical outlook in many subsequent books, also widely translated, until his death in 1948 at his home in Clamart, a suburb of Paris.

2 Metaphysics

In keeping with his existentialist orientation, Berdiaev defines 'metaphysics' not as the study of reality in general but as 'the philosophy of human existence', and he contends that this philosophy is not 'objective' but 'subjective'. In calling it subjective he means to indicate both that its focus is the subject and that it is not a sphere of demonstrable, impersonal truth. 'Truth and reality', he writes, 'are not identical with objectivity' ([1937b] 1952: 17). Yet, as this statement itself indicates, he does not reject truth or reality, and his worldview, however subjectively conceived, is in fact phrased as a general theory of reality for which truth is implicitly claimed. This theory was set forth most fully in his 1947 book, *Opyt ėskhatologicheskoi metafiziki (An Essay in Eschatological Metaphysics)*.

Berdiaev was willing to call his doctrine a form of dualism, but not a dualism of idea versus matter or divine versus human. Rather, he opposes *dukh* ('spirit', both divine and human) to *priroda* ('nature') or *byt'ë* ('being'), although he also frequently uses the latter word to signify reality in general. Spirit is the realm of freedom, personality and creative activity; nature is the realm of necessity, objects, routine and passivity. Berdiaev likens this dualism to the Kantian distinction between noumena and phenomena: the spiritual (noumenal) realm has primacy over the realm of nature (the phenomenal realm), in that it alone is truly 'real' and fully independent of the other (see Kant, I. §3). Unlike Kant, however, Berdiaev did not regard the world of spirit as unknowable (except for the ultimate mystery of the one perfect spirit, God); he considered it accessible through intuitive - essentially, spiritual and mystical - experience, which is the only true foundation for a philosophical system. Rational knowledge of the world of 'nature', on the other hand, is 'objective' knowledge, which has no philosophical value.

Berdiaev links God, human spirits and the objectified world of nature in a philosophical cosmology based on a controversial interpretation of the Christian idea of divine creation *ex nihilo*. Drawing on Boehme's mystical doctrine of the primitive *Ungrund* ('the groundless') that underlies all reality (see Boehme, J.), Berdiaev contends that the 'nothing' out of which God creates the world is not sheer emptiness but a positive potentiality, which he also calls 'meonic freedom' - a volitional, irrational, formless, creative potency or energy that is metaphysically prior not only to the world but to God the creator. God creates himself out of this irrational freedom, overcoming it in his own nature through his perfection. But in proceeding to create human beings out of the *Ungrund* as well, he produces creatures whose irrational freedom is neither generated nor controlled by him. Thus when humans misuse their freedom by choosing evil, God is not implicated in the choice; indeed he does not even have foreknowledge of it. Berdiaev valued this doctrine as providing an explanation of evil within a theistic philosophy; but for abandoning the metaphysical primacy and the omnipotence and omniscience of God, he was severely criticized by Russian philosophers who adhered more closely to traditional Orthodox theological doctrine.

The ability of the free human spirit to choose evil is also invoked by Berdiaev in his idealist account of the genesis of nature. The realm of objects is produced not by God but by the fallen human subject: 'the subject is the creation of God while the object is the creation of the subject' ([1947] 1952: 17). The proud individual subject, intoxicated by freedom, creates the world of exteriority by separating itself from God and other spirits. The subject engages in what Berdiaev calls 'objectification' (*ob''ektivatsiia*) - a process of self-alienation in which a spirit gives itself and others an objectified form as part of a system of external substances and relations. Characterized by division, disintegration, materiality and necessity, this system functions to enslave the very subject that created it. Although Berdiaev frequently speaks of this objectified world of nature as phenomenal - a realm of 'appearances' - he also calls it ontologically real, albeit of a lower order of reality. By this he apparently means that nature, once created by the objectification of spirit, is truly 'objective' as an independently subsisting force, not simply an imaginary

one. He regards the overcoming of objectification in the name of freedom as the daunting but imperative task of the human spirit.

3 Philosophy of history

Human history, as Berdiaev explains it (primarily in *Opyt ėskhatologicheskoi metafiziki* and in his 1923 *Smysl istorii (The Meaning of History)*), is a temporal process connected with the creation and conquest of objectification. History begins with the Fall. The dividedness or exteriority of the objective world that is generated by the sinful use of human freedom is manifested not only spatially but temporally, in the separation of past, present and future. The flow of time in the objective world may be viewed in two ways, according to Berdiaev: either cyclically, as an endless repetition of natural processes operating independently of human thought or action (he calls this *cosmic* time, symbolized by a circle), or linearly, as an unrepeatable succession of events involving human agency (this is *historical* time, symbolized by a line).

Berdiaev argues eloquently against interpreting the historical process as progressive, for ‘progress’ as he understands it signifies that existing generations of people are regarded simply as means for improving the lot of those to come. Berdiaev used this moralistic, Kantian condemnation of the idea of progress (anticipated by his Russian predecessor Nikolai Fëdorov) to great effect in his attack on Russian communism as a quintessential case of treating existing human beings as raw material for the production of a better future. At the same time, Berdiaev’s philosophy is strongly eschatological, and he argues that history would be meaningless if it had no termination and if objectification were not overcome. In his eyes the ultimate horror, adding moral crime to meaninglessness, would be a conception of history as *endless progress* - the perpetual instrumentalization of one generation after another with none ever becoming an end in itself.

The ‘end of history’, or conquest of objectification, postulated by Berdiaev takes place not in historical time but in time of still a third type, which he calls *existential*. This is the time of the free, creative act, and he symbolizes it by a point; it is durationless and integral. The creative act in existential time manifests itself as essentially an irruption of ‘meta-history’ into history, of the spiritual plane of eternity into the objectified world of historical time. One such irruption of great significance was the Incarnation, but in fact every creative act at any historical moment in the ‘divine-human’ process of redemption is a victory of spirit over nature and freedom over necessity. Berdiaev incorporates in his eschatology such traditional Christian concepts as the Kingdom of God, paradise, hell, universal resurrection (it must be universal, or some persons would not be valued as ends in themselves) and immortality, but in keeping with his notion of existential time he interprets these concepts as referring not to future events or conditions but to timeless and purely spiritual presences. Paradise and immortality are accessible to the creative spirit at any (historical) moment.

Berdiaev’s reasoning concerning the overcoming of objectification through creativity appears to be that since the succession of events in historical time is itself a product of objectification, the annulment of the latter cannot be considered an event in that succession but must occur in another temporal dimension - the extrahistorical dimension he calls ‘existential’. But this conception raises further questions that Berdiaev does not answer unequivocally. Commentators on his philosophy are divided as to whether the ‘overcoming’ of objectification requires that historical time literally come to an end or simply that it be rendered insignificant. In either case, if the ‘end of history’ is located outside historical time, that would appear once again to deprive history *per se* of meaning.

4 Ethics

The issue of the impact of creativity on the world is encountered once again in Berdiaev’s personalistic ethics, which ascribes supreme value to the individual spirit and identifies moral action with the creativity that combats and ultimately defeats the objectification produced by the Fall. His mature ethical views were first suggested in *Smysl tvorchestva* and were developed most fully in his widely read book *O naznachenii cheloveka (On the Destiny of Man)* (1931).

Berdiaev distinguishes three stages in the development of moral consciousness, each of which gives rise to a corresponding type of ethics. Earliest and lowest but still most widespread is the ethics of *law*, which subjects individuals to abstract norms of behaviour in the name of social order. Morality at this pre-Christian level is

negative and impersonal, and it is based on the distinction between good and evil - a distinction that is itself part of the objectification created by the Fall. Berdiaev concedes that the ethics of law is a social necessity in a fallen world, but argues that it is far from adequate to the developed moral consciousness.

A higher, Christian stage is marked by what Berdiaev calls the ethics of *redemption*, which speaks to the individual rather than the group and offers as its standard of value not an abstract law but the figure of a loving, all-forgiving saviour. This ethics of Christian love correctly acknowledges the Incarnation and redemption as elements in divine creation and in the transfiguration of the objectified world. But it, too, is flawed, according to Berdiaev: it can degenerate into 'transcendental egoism', an exclusive concern for one's own salvation.

The highest stage of the moral consciousness is reflected in the more advanced Christian ethics of *creativity*, in which human beings, as free spirits, transcend the good-evil distinction through unique and unrepeatable acts that bring novelty into the world. In responding freely in this way to God's command to overcome evil, the individual person becomes a co-creator of the universe. Human creators are 'beyond good and evil' in the sense that they, undetermined by social standards or any other form of objectification, define the moral sphere through personal decisions. At this highest level, morality and creativity are effectively coextensive for Berdiaev: 'the moral act is a creative act [and]… all creation has moral significance, even if it is the creation of cognitive or aesthetic values' ([1937b] 1952: 20-1). 'The creator is justified by his creative achievement' ([1931] 1935: 130).

To the critic who would charge Berdiaev with Nietzschean egoism here, he responds that it is precisely in creative acts that individuals are the most selfless, since all energies are focused on the acts and their products. Berdiaev denies, too, that the genuine creator would seek to wound or dominate other people: the free person recognizes that no one should be treated as a mere means. This continued appeal to the Kantian principle, however, suggests that Berdiaev is, after all, admitting a universal moral law to which individuals must conform - a suspicion furthered by his acceptance of 'conscience' as 'the organ of perception of the religious revelation, of goodness, righteousness and truth in its entirety' ([1931] 1935: 167). Berdiaev is also criticized for 'deifying' human beings; Sergei Levitskii (1975) and Piama Gaidenko (1994) argue that in glorifying human powers Berdiaev is forgetting the gulf that separates flawed and finite humanity from divine perfection.

The most widely debated feature of Berdiaev's philosophical outlook is the redemptive value he attributes to creativity in its dual capacity of giving meaning to history and expressing the highest morality. In both contexts, the value of creativity hinges on its effectiveness in overcoming objectification. But inasmuch as human creators are concerned with their 'products' and must operate in the external world (Berdiaev acknowledges that, unlike God, they require material other than 'meonic freedom'), they cannot escape the dead hand of objectification: 'every expression of creative action in the external [world] falls into the power of that [objectified] world' ([1939] 1943: 127). Thus, in the 'irruption' of existential time into historical time, creativity participates in the succession of historical events and is enmeshed in the very objectification it sets out to avoid. Berdiaev himself calls this situation 'tragic': 'It is the tragedy of creativeness that it wants eternity and the eternal, but produces the temporal and builds up culture which is in time and a part of history' ([1931] 1935: 136). V.V. Zenkovsky (1948-50) and some other critics contend that on Berdiaev's principles the creative act is not merely tragic but meaningless and futile; Fuad Nucho (1966) and others reply that creativity retains spiritual value despite its tragic character.

5 Social and political philosophy

Berdiaev's first book (1901) was devoted to social and political philosophy, and the subject figures prominently in virtually everything he wrote, including his last, posthumously published book, *Tsarstvo dukha i tsarstvo kesaria (The Realm of Spirit and the Realm of Caesar)* (1949b). In most of his mature works dedicated to the subject, such as *Filosofiia neravenstva (The Philosophy of Inequality)* (1923a), the critique of Russian communism is a principal concern. Two works - *The Origin of Russian Communism* (1937a) and *Russkaia ideia (The Russian Idea)* (1946) - include extensive critical discussion of the history of Russian social and political thought.

In using the term 'nature' for the world of objectification opposed to spirit, Berdiaev does not mean to exclude *social* factors from the objectifying forces that enslave personality; indeed, it is the objectifying power of society above all that he sees as a threat to freedom. He protests against the depersonalizing processes in which the self is moulded by objectified social relationships and comes to be considered a part of society rather than an independent bearer of value. The great error of Marxist communism and all forms of collectivism is that they view humanity as

simply a product of society, determined by social laws, and consequently have no regard for personal freedom. As early as 1906, Berdiaev predicted that the Bolshevik Party in Russia would produce a society united by force, ruled by an unlimited and ‘deified’ state (1906: 536).

At the same time, Berdiaev believed that individualism is as much a threat to personality as collectivism is, since setting the person *apart* from others is no less a form of objectification. In this regard Berdiaev retained what he called ‘a soft spot’ for Marxism; he shared Marx’s abhorrence of ‘bourgeois’, capitalist society as alienating and dehumanizing. A critic of the free market economy and of claims to an absolute right of private ownership, Berdiaev at times found points of agreement with Soviet ideology; in entertaining (especially during and immediately after the Second World War) the possibility that Soviet rule might become more benign, he incurred the wrath of much of the Russian émigré community.

In opposition to both collectivism and individualism, Berdiaev offered his own ‘personalistic socialism’ - a form of free sociality having much in common with the older Russian ideal of *sobornost’* (see Slavophilism §3). Personality, as a spiritual category, he argues, is not only free but universal, and as such it is inherently social in the sense of requiring others for its fulfilment. To achieve the fullness of personal being, the individual must be in communion with other individuals, where ‘communion’ signifies a complete but fully voluntary union based on love. It is in this essentially spiritual sense of ‘socialism’ that Berdiaev can affirm that a Christian not only may but must be a socialist.

Berdiaev’s hostility towards all established (‘objectified’) social arrangements - whether totalitarian or democratic - prevented him from spelling out ‘personalistic socialism’ beyond such broad descriptions of its aim as ‘a synthesis of the aristocratic, qualitative principle of personality with the democratic, socialist principle of justice and brotherly cooperation’ ([1937b] 1952: 20). The state as such appears to have no role in effecting such a synthesis, for Berdiaev views every political structure as an impediment to the realization of personality; he expresses no preference, for example, for a constitutional state based on the rule of law over a totalitarian state. Because he regards the state and law as essentially parts of the objectified world that must be vanquished, his political philosophy amounts to a form of anarchism.

See also: Existentialism; Existentialist theology; Neo-Kantianism, Russian §5; Nietzsche: impact on Russian thought §2

JAMES P. SCANLAN

List of works

Berdiaev’s principal works are listed here in order of their initial publication; since 1988 most of them have been reprinted in Russia, many with new editorial commentary and some in more than one edition.

Berdiaev, N. [Berdyaev] (1901) *Sub’’ektivizm i individualizm v obshchestvennoi filosofii. Kriticheskii ėtiud o N.K. Mikhailovskom (Subjectivism and Individualism in Social Philosophy: A Critical Study of N.K. Mikhailovskii)*, St Petersburg: Popova.(Berdiaev’s first book, in which he sought to complement Marxist socioeconomic views with Kantian transcendental idealism.)

Berdiaev, N. [Berdyaev] (1906) ‘Sotsializm, kak religiia’ (Socialism as Religion), in *Voprosy filosofii i psikhologii (Problems of Philosophy and Psychology)* 85: 508-45; trans. M. Schwartz, in B.G. Rosenthal and M. Bohachevsky-Chomiak (eds), *A Revolution of the Spirit: Crisis of Value in Russia, 1890-1924*, New York: Fordham University, 1990, 107-33.

Berdiaev, N. [Berdyaev] (1907) *Novoe religioznoe soznanie i obshchestvennost’ (The New Religious Consciousness and Society)*, St Petersburg: Pirozhkov.

Berdiaev, N. [Berdyaev] (1911) *Filosofiia svobody (Philosophy of Freedom)*, Moscow: Put’.

Berdiaev, N. [Berdyaev] (1916) *Smysl tvorchestva. Opyt opravdaniia cheloveka (The Meaning of Creativity: An Essay in the Justification of Man)*, Moscow: Leman & Sakharov; trans. D.A. Lowrie, *The Meaning of the Creative Act*, New York: Harper & Bros, 1955. (Berdiaev’s last philosophical book in Russia.)

Berdiaev, N. [Berdyaev] (1923a) *Filosofiia neravenstva. Pis’ma k nedrugam po sotsial’noi filosofii (The Philosophy of Inequality: Letters to Opponents on Social Philosophy)*, Berlin: Obelisk; trans. C. and A. Andronikof, *De l’inégalité*, Lausanne: Éditions l’Age d’Homme, 1976.(A critique of Russian communism.)

Berdiaev, N. [Berdyaev] (1923b) *Smysl istorii. Opyt filosofii chelovecheskoi sud’by (The Meaning of History: An Essay in the Philosophy of Human Destiny)*, Berlin: Obelisk; trans. G. Reavey, *The Meaning of History*, London: Geoffrey Bles, 1936.

Berdiaev, N. [Berdyaev] (1924) *Novoe srednevekov'e. Razmyshlenie o sud'be Rossii i Evropy (The New Middle Ages: Reflections on the Destinies of Russia and Europe)*, Berlin: Obelisk; trans. D. Attwater, *The End of Our Time*, New York: Sheed & Ward, 1933.(The book which won Berdiaev worldwide fame.)

Berdiaev, N. [Berdyaev] (1927-8) *Filosofiia svobodnogo dukha. Problematika i apologiia khristianstva (Philosophy of the Free Spirit: Christianity's Problems and Apologia)*, Paris: YMCA-Press, 2 vols; trans. O. Clarke, *Freedom and the Spirit*, London: Geoffrey Bles, 1935.

Berdiaev, N. [Berdyaev] (1931) *O naznachenii cheloveka. Opyt paradoksal'noi ėtiki (On the Destiny of Man: An Essay in Paradoxical Ethics)*, Paris: Sovremennye zapiski; trans. N. Duddington, *The Destiny of Man*, London: Geoffrey Bles, 1935.(Here Berdiaev's mature ethical views are developed most fully.)

Berdiaev, N. [Berdyaev] (1934) *Ia i mir ob''ektov. Opyt filosofii odinochestva i obshcheniia (The Self and the World of Objects: An Essay in the Philosophy of Solitude and Community)*, Paris: YMCA-Press; trans. G. Reavey, *Solitude and Society*, London: Geoffrey Bles, 1938.

Berdiaev, N. [Berdyaev] (1937a) *The Origin of Russian Communism*, trans. R.M. French, London: Geoffrey Bles. (Originally published in translation; the Russian text was first published in 1955 as *Istoki i smysl russkogo kommunizma (The Sources and Meaning of Russian Communism)*, Paris: YMCA-Press.)

Berdiaev, N. [Berdyaev] (1937b) 'Die Philosophische Weltanschauung N.A. Berdiaeffs (Selbstdarstellung)' (The Philosophical Worldview of N.A. Berdiaev (Self-Exposition)), in *Philosophen-Lexicon: Handwörterbuch der philosophie nach personen*, Berlin: Walter de Gruyter; repr. in *'Berdiaeff, Nicolei'* (1949) 1: 102-8, ed. W. Ziegenfuss; 1st Russian edn, 'Filosofskoe mirosozertsanie N.A. Berdiaeva (Avtoizlozhenie)' in *Vestnik Russkogo studencheskogo khristiankogo dvizheniia* (1952) 4/5: 15-21.(A text, first published as a German translation, in which Berdiaev summarized his own views in an unusually systematic fashion.)

Berdiaev, N. [Berdyaev] (1939) *O rabstve i svobode cheloveka. Opyt personalisticheskoi filosofii (On Slavery and the Freedom of Man: An Essay in Personalist Philosophy)*, Paris: YMCA-Press; trans. R.M. French, *Slavery and Freedom*, London: Geoffrey Bles, 1943.

Berdiaev, N. [Berdyaev] (1946) *Russkaia ideia. Osnovnye problemy russkoi mysli XIX veka i nachala XX veka (The Russian Idea: Basic Problems of Russian Thought of the Nineteenth and Early Twentieth Centuries)*, Paris: YMCA-Press; trans. R.M. French, *The Russian Idea*, London: Geoffrey Bles, 1947.(Includes extensive critical discussion of the history of Russian social and political thought.)

Berdiaev, N. [Berdyaev] (1947) *Opyt (skhatologicheskoi metafiziki. Tvorchestvo i ob''ektivatsiia (An Essay in Eschatological Metaphysics: Creativity and Objectification)*, Paris: YMCA-Press; trans. R.M. French, *The Beginning and the End*, London: Geoffrey Bles, 1952.(Sets forth a general theory of reality.)

Berdiaev, N. [Berdyaev] (1949a) *Samopoznanie. Opyt filosofskoi avtobiografii (Self-Knowledge: An Essay in Philosophical Autobiography)*, Paris: YMCA-Press; trans. K. Lampert, *Dream and Reality*, London: Geoffrey Bles, 1950.(Because of abridgment and other editorial changes, the English translation is untrustworthy; translations into many other languages exist.)

Berdiaev, N. [Berdyaev] (1949b) *Tsarstvo dukha i tsarstvo kesaria*, Paris: YMCA-Press; trans. D.A. Lowrie, *The Realm of Spirit and the Realm of Caesar*, London: Gollancz, 1952.(Social and political philosophy figures prominently in Berdiaev's last, posthumously published book.)

References and further reading

Ermichev, A.A. (1994) *N.A. Berdiaev: Pro et contra. Antologiia (N.A. Berdiaev: Pro and Contra: An Anthology)*, book 1, St Petersburg: Russkii khristianskii gumanitarnyi institut.(A collection of memoirs and critical studies by a long list of distinguished Russian contemporaries; includes reprints of Berdiaev [1937b] 1952 and Levitskii 1975; contains bibliography.)

Gaidenko, P.P. (1994) 'The Philosophy of Freedom of Nikolai Berdiaev', in J.P. Scanlan (ed.), *Russian Thought after Communism: The Recovery of a Philosophical Heritage*, Armonk, NY: M.E. Sharpe, 104-20.(Referred to in §4. A post-Soviet Russian critique of Berdiaev's philosophy with special reference to social and political questions.)

Klépinine, T. (1978) *Bibliographie des oeuvres de Nicolas Berdiaev*, Paris: YMCA-Press.(A well-organized bibliography of almost 500 works by Berdiaev; contains a detailed chronology of his life.)

Levitskii, S.A. (1975) *'Berdiaev: prorok ili eretik?' (Berdiaev: Prophet or Heretic?)*, in *Novyi zhurnal (New Journal)* 119: 230-53.(Referred to in §4. A critical assessment of Berdiaev's philosophy, field by field.)

Lowrie, D.A. (1960) *Rebellious Prophet: A Life of Nicolai Berdyaev*, New York: Harper & Bros.(A general biography by a long-time associate; contains bibliographies.)

Nucho, F. (1966) *Berdyaev's Philosophy: The Existential Paradox of Freedom and Necessity*, Garden City, NY: Doubleday.(Referred to in §4. A sympathetic analysis of Berdiaev's philosophy of freedom; contains

bibliographies, including an annotated bibliography of the philosopher's principal writings.)
Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, 2 vols; 2nd edn 1989; trans. G.L. Kline, A History of Russian Philosophy, vol. 2, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, 760-80.(Referred to in §4. Exposition and critique of the philosophy of Berdiaev by the authoritative historian of Russian philosophy.)

Bergson, Henri-Louis (1859-1941)

So far as he can be classified, Bergson would be called a 'process philosopher', emphasizing the primacy of process and change rather than of the conventional solid objects which undergo those changes. His central claim is that time, properly speaking and as we experience it (which he calls 'duration'), cannot be analysed as a set of moments, but is essentially unitary. The same applies to movement, which must be distinguished from the trajectory it covers. This distinction, he claims, solves Zeno of Elea's paradoxes of motion, and analogues of it apply elsewhere, for instance, in biology and ethics.

Bergson makes an important distinction between sensation and perception. He repudiates idealism, but claims that matter differs only in degree from our perceptions, which are always perfused by our memories. Perception free from all memory, or 'pure' perception, is an ideal limit and not really perception at all, but matter. Real perception is pragmatic: we perceive what is necessary for us to act, assisted by the brain which functions as a filter to ensure that we remember only what we need to remember. Humans differ from animals by developing intelligence rather than instinct, but our highest faculty is 'intuition', which fuses both. Bergson is not anti-intellectualist, though, for intuition (in one of its two senses) presupposes intelligence. He achieved popularity partly by developing a theory of evolution, using his élan vital, *which seemed to allow a role for religion. In ethics he contrasted a 'closed' with a (more desirable) 'open' morality, and similarly contrasted 'static' with 'dynamic' religion, which culminates in mysticism.*

1 Life

Bergson was born in Paris on 18 October 1859 with a musician as father and a mother from Yorkshire. He married Louise Neuberger, a relative of Proust, and had one daughter. After teaching in Angers, Clermont-Ferrand and Paris, he held a chair at the Collège de France from 1900 to 1921, where his lectures before the First World War attracted so many people that it was seriously proposed to move them to the Opéra. After the war, interest in his lectures declined and he turned from academic teaching (though only partly from writing) to promoting international understanding as a prophylactic against war. Fiercely patriotic, he died at France's darkest hour on 3 January or 4 January 1941, after seventeen years of crippling arthritis, and after supporting his fellow Jews by refusing an offer of exemption from anti-Semitic regulations; the same sympathy may have stopped him officially adopting the Catholic religion, to which in later life he became spiritually close (despite having his books placed on the Index in 1914).

Bergson was a man of wide intellectual attainments. At seventeen he won first prize in an open mathematical competition and also solved a problem left unsolved by Pascal. His subsidiary degree thesis (written in Latin) dealt with Aristotle on place, and he lectured on Lucretius. He devoted several years to a detailed study of the literature on aphasia, in connection with memory, and similarly used detailed scientific evidence to support his views on evolution. He was also a great stylist and his books can stand beside those of Berkeley, Russell and the early Plato as among the more readable works of philosophy.

2 Time and duration

The core of Bergson's philosophy, which, as he pointed out in a letter of 1915 (1972: 1148), every account of his philosophy must start from and constantly return to, on pain of distortion, is the 'intuition of duration'. Time, for Bergson, is of two fundamentally different kinds, or better, especially for his later philosophy, appears in two fundamentally different guises. For science, time is essentially particulate. It consists of an infinite, dense set of instants, and science uses the calculus to study the world as it is at these instants. Change is nothing over and above the world's being in different states at different instants, and the transition from one state to another is something science can take no account of except by using the calculus in this way. (This interpretation of the role of the calculus for Bergson has been disputed: see Milet 1974.) For experience, however, this transition is the very essence of time, now called duration (*durée*). We do not live from moment to moment, but in a continuous stream of experience (the similarity to William James' 'stream of consciousness' is unsurprising, given the close personal and professional friendship between Bergson and James, who reached their views independently).

One might wonder why change should not consist simply in being in different states at different instants, provided

the instants form a dense set, so that no two are adjacent (a feature Bergson unfortunately ignores in his favourite image of time as cinematographic). Bergson's reply, that this overlooks the phenomenology of experience, surely has merit, and helps to solve several problems. We experience the immediate past, and possibly the immediate future, along with the present, as actual, and we can perhaps avoid objections that have confronted James' independently developed 'specious present' if (with Bergson) we avoid treating the act of experiencing as itself separate and momentary. But be that as it may, Bergson can avoid Augustine's problem that time vanishes because only the present is actual and the present does not last long enough to be real at all. He also need not worry about how we acquire a concept of the past when experience only ever presents us with the present.

However, problems do arise. Duration is introduced as essentially linked to consciousness; but does duration exist in the outer world? Bergson's first major book *Essai sur les données immédiates de la conscience (Time and Free Will)* (1889) states unambiguously that it does not, but his next book *Matière et mémoire: Essai sur les relations du corps avec l'esprit (Matter and Memory)* (1896) does allow duration to the outer world, as do his later works. The change of view was well motivated, for how could a consciousness embedded in duration live in a world devoid of it? Science still treats the world as cinematographic, and so now falsifies it, but inevitably and harmlessly, so long as we do not expect from science more than it can give; it is for metaphysics, using 'intuition', to describe the world philosophically, but only science can give us our indispensable practical understanding of the world. Bergson, however, never seemed conscious of a real change of view, and in his much later *La pensée et le mouvant: essais et conférences (The Creative Mind)* (1934) talks simply of *Matter and Memory* as getting nearer to what he wanted to say. Nor did he ever satisfactorily explain the extent to which duration is bound up with consciousness.

Bergson's treatment of time and duration invites comparison with McTaggart's B-series and A-series respectively. McTaggart wrote in 1908, after Bergson's main treatments, but Bergson's later writings show no knowledge of him. In McTaggart's terms Bergson would be a thoroughgoing A-theorist, especially from *Matter and Memory* onwards (see McTaggart, J.M.E. §2).

Discrete plurality for Bergson is essentially spatial, and time with its multiplicity of moments is duration spatialized. This contrast between space and genuine time (duration) introduces an asymmetry between space and time which puts Bergson at odds with recent philosophy (which tends to treat them alike), and assimilates him in this respect to older philosophers such as the Greeks. One of his favourite examples for illustrating duration is a melody, which we can only hear as a melody if we hear it as a whole. Critics have pointed out that, similarly, we can only see a circle by seeing it as a whole (Boudot 1980: 349), and have claimed that in order to distinguish space and time Bergson uses a distinction between the psychological and the mathematical that applies within space and time equally (Berthelot 1913: 354-5). The critics are somewhat justified, though Bergson does in a lower key distinguish space from extensity, and could perhaps thereby deal with the circle. But the critics do scant justice to the real asymmetries between space and time in terms of directions and 'flow' which support Bergson's general approach (see Time §1).

3 Bergson and Zeno

The ideas so far outlined provide Bergson with a tool which he uses first to deal with Zeno's paradoxes of motion, but then goes on to apply in other spheres, such as biology and ethics. This tool is the distinction between a movement and its trajectory. The reason that Zeno's Achilles never overtakes his tortoise is that Zeno insists on applying to the movement, which occurs in time, the infinite process of division that really only applies to the trajectory, which is spatial (see Zeno of Elea §7). The movement is essentially unitary and indivisible. This gives the spirit of Bergson's views, though only as a rough approximation: evidently Achilles' movement does have parts - his steps - and it is these that have no parts. But Bergson never seems to succeed in giving adequate criteria for deciding just when a movement is unitary and so has no parts.

The use of this tool in other spheres begins with the treatment of free will in *Time and Free Will*, where it joins a sort of dialectical device that Bergson repeatedly employs: the insistence that two antagonistic approaches that together dominate a philosophical topic share a common error, though he often admits that his own view lies nearer to one pole than to the other. On free will, the poles are determinism and libertarianism, and the error, as so often, amounts to replacing a movement by its trajectory. Bergson's own view, that a free act will proceed from the self alone and 'express the whole of the self' ([1889] 1990: 165-6), is nearer to libertarianism, but the

libertarian, insisting that the agent ‘could have done otherwise’, shares with the determinist the view that the trajectory is already there before the action and that it makes sense to imagine a replay, stopping the action halfway through, as it were, and sending it off on a different course. His point seems to be that there is no ‘halfway through’ at which the action could be stopped; the process flowing from deliberation to completed action (the doing, as opposed to the things done) is unitary and indivisible (see Free will §1-2).

4 Process philosophy

Process philosophy is a philosophical tradition which goes back as far as Heraclitus, and if Bergson can be placed in any tradition, it is in this, despite his repudiation of allegiance to Heraclitus. Process philosophy stands in opposition to the tradition stemming from Aristotle’s scheme of categories, where the world consists of substances which have properties and undergo change. For process philosophy the world consists of processes, and Bergson often says things like, ‘There are changes, but there are underneath the change no things which change… movement does not imply a mobile’ ([1934] 1946: 173). Objects are like ‘snapshots’ of a flux, which is duration. This echo of the cinematographic approach of science illustrates another feature of Bergson: his pragmatism. He does not deny that language (itself a pragmatic device for dealing with the world) uses the Aristotelian apparatus of subject and predicate, but we see the world in terms of objects which change because that is the only way we can act in it, just as science gives us our only way of manipulating it.

5 Metaphysics and philosophy of mind

This pragmatism appears again in Bergson’s philosophy of mind, which, as we might expect from the way in which he links duration to consciousness, is itself closely linked to his metaphysics. He repudiates idealism, and begins the introduction to the 1911 edition of *Matter and Memory* by calling himself a dualist, ‘affirming the reality of spirit and the reality of matter’. But his dualism is not ‘vulgar’. It can be called a dualism of time and space (it is tempting to call it one of movement and trajectory), but from another point of view it could be called one of perception and memory, terms which he constantly contrasts as differing in kind, not in degree.

But though matter and spirit are both real, they differ only in degree, and here we reach a central part of Bergson’s metaphysics, and also his epistemology, for our knowledge of the world is essentially bound up with the nature of the world itself. Bergson is one of the few philosophers to distinguish clearly between sensation and perception. We cannot start with sensations, treated as unextended and inside ourselves, and somehow turn them into perceptions telling us of an extended outer world, just as we cannot get a concept of the past by starting from a momentary present, and treating memories simply as weaker (‘fainter’, as Hume would say) sensations or perceptions. Bergson’s target here is the associationism that underlies so much of eighteenth- and nineteenth-century empiricism, and his criticisms are of fundamental importance, whether or not his own view also faces difficulties.

Although officially dualist, Bergson’s view is somewhat akin to the ‘neutral monism’ of William James and others (see James, W. §6; Neutral monism). Though perception differs in kind from memory, it essentially involves it in varying degrees. Our perceptions are always affected by our experience, and if we had no memories we would have no real perceptions - another important criticism of Humean empiricism. Perception takes place not inside us but where its object is, and a perception unmediated by memory, and in that sense a ‘pure’ perception, is an ideal limit, and not really perception at all; it ‘is really part of matter’. In effect it is the object itself, or rather, since it now lacks duration, it is what we might now describe as a momentary time-slice of the object.

Bergson’s pragmatism reappears here: we perceive what we need to perceive in order to act (we might think this more obviously true in the case of animals), and the function of the brain is to filter memories so that only those enter consciousness which are of practical use, notably in perceiving; he used his study of aphasia to argue that the brain cannot be used as a storehouse for memories. Superficially, his treatment of memory involves an excessively crude dichotomy between picture-memory and habit-memory, but he was not concerned with many of the problems that interest later thinkers.

6 Humour

Le rire: essai sur la signification du comique (Laughter: An Essay on the Meaning of the Comic) (1900), probably Bergson’s most popular book, can be seen as an appendix to his philosophy of mind and body. For him, a human

being is a creature who is both body and spirit (to avoid too intellectual a term: see §7) and uses its body for practical purposes. But sometimes the body takes over and we act as though we were simply a body, either obeying only physical laws - when we slip on a banana-skin, for instance - or when our actions become wooden, mechanical, automatic or stereotyped. It is then that others laugh at us and we have the makings of comedy, low or high. Bergson also gives laughter a function, as a social corrective (his target has something in common with Sartre's 'bad faith'). 'A humorist is a moralist disguised as a scientist' (1900: 128).

7 Science and metaphysics: the *élan vital*

Bergson distinguishes three cognitive faculties: intelligence, instinct and intuition. As evolution has advanced, animals and humans have diverged and developed instinct and intelligence respectively as their tools for confronting the world. These are equally suited for their tasks, intelligence being extensible but hazardous, while intuition is limited but safe. Bergson uses detailed scientific evidence to illustrate the remarkable achievements of which instinct is capable. Intuition is a development of instinct, mediated by intelligence, which occurs only in humans but takes them to their highest level, and is the faculty used by metaphysics to say what reality is really like, while science uses intelligence to study reality in a manner inevitably distorted - but essential for practical living. But 'intuition' is ambiguous in Bergson. In one sense it turns quantity into quality, and, for example, enables us to see trillions of vibrations as the colour red, and experience duration; metaphysics uses it to study life and spirit. But in another sense it is insight, the getting of bright ideas, which both presupposes and is essential for the development of intelligence.

On evolution, Bergson again claims that two antagonistic theories, Darwinian mechanism and finalism or teleology, share a common presupposition, that the path or trajectory of evolution is somehow already laid out. His own view involves his famous *élan vital* ('vital impetus', usually left untranslated) which drives evolution on, though not towards any pre-ordained goal. It drives rather than draws, and to that extent resembles mechanism, but it also overcomes obstacles - a puzzling idea if it has no goals. Perhaps Bergson is here taking up a stance nearer to one extreme (teleology) than to the other, evolution having intermediate goals but no overall goal.

8 Morality and religion

Bergson turned to morality, and to an explicit discussion of religion, late in life in his last major work, *Les deux sources de la morale et de la religion (The Two Sources of Morality and Religion)* (1932). On both topics he uses a dualistic framework, but not, as in previous works, to show how two antagonistic approaches share a common premise. He contrasts closed morality with open morality and static religion with dynamic religion, and in each case his preference for the second term is unambiguous. The open morality is one of aspiration rather than impulsion and is universal in scope. Dynamic religion is somewhat similar, culminating in mysticism, of whose nature and development he gives an extended account. Obligation he sees as the pull of instinct against the waywardness introduced by intelligence, and he rightly emphasizes that we perform the great majority of our obligations as a matter of course and without any heroic Kantian struggle. The contrast between trajectory and movement is used twice here. Just as we can never build up a movement out of elements of its trajectory, but must treat it as something distinct and unitary, we can never construct a motive for moral action from individual intellectual considerations: the motive must already be there, given by instinct (there are echoes of Hume here). The second point is that we can never pass from ever-expanding group loyalties, which always require some out-group as a foil, to the universal love of mankind that open morality demands and that only the mystic can provide.

See also: Comedy §2; Humour §§2, 4

A.R. LACEY

List of works

Bergson, H.-L. (1889) *Essai sur les données immédiates de la conscience*, Paris: Alcan; trans. F.L. Pogson, *Time and Free Will*, New York: Swan Sonnenschein, and London: Allen & Unwin, 1990.(Introduces contrasts between time and duration and between movement and trajectory.)

Bergson, H.-L. (1896) *Matière et mémoire: Essai sur les relations du corps avec l'esprit*, Paris: Alcan; trans. N.M. Paul and W. Scott Palmer, *Matter and Memory*, New York: Swan Sonnenschein, and London: Allen & Unwin, 1911.(Translation contains a new introduction by Bergson.)

Bergson, H.-L. (1900) *Le Rire: Essai sur la signification du comique*, Paris: Alcan; trans. C. Brereton and F. Rothwell, *Laughter: An Essay on the Meaning of the Comic*, London and New York: Macmillan, 1911.(Applies his general ideas to the sphere of humour.)

Bergson, H.-L. (1903) 'Introduction à la métaphysique', *Revue de Métaphysique et de Morale* 29: 1-36; trans. T.E. Hulme, *Introduction to Metaphysics*, New York: Putnam, 1912; and in *The Creative Mind*, 1934.(More on epistemology than on metaphysics, this is where Bergson introduces his notion of 'intuition'.)

Bergson, H.-L. (1907) *L'evolution créatrice*, Paris: Alcan; trans. A. Mitchell, *Creative Evolution*, New York: Holt, 1911.(Important for Bergson's treatment not only of biology but of intuition, and also of different types of order and disorder and their relations, and of the concept of nothing.)

Bergson, H.-L. (1919) *L'energie spirituelle*, Paris: Alcan; trans. H. Wildon Carr, *Mind-Energy: Lectures and Essays*, New York: Holt, and London: Macmillan, 1920.(Collected essays, mainly on mind and body.)

Bergson, H.-L. (1922) *Durée et simultanéité: a propos de la théorie d'Einstein*, Paris: Alcan; trans. L. Jacobson, with introduction by H. Dingle, *Duration and Simultaneity*, Indianapolis, IN: Bobbs-Merrill, 1965.(The second edition in 1923 adds three appendices replying to criticisms. Bergson tries to defuse some paradoxical consequences of relativity theory, writing before these were empirically confirmed. It is now agreed he was wrong in the letter, though some say he was right in the spirit and anticipated later developments; see Čapek and Heidsieck.)

Bergson, H.-L. (1932) *Les deux sources de la morale et de la religion*, Paris: Alcan; trans. R.A. Audra, C. Brereton and W.H. Carter, *The Two Sources of Morality and Religion*, New York: Holt, 1935.(Distinguishes two levels, higher and lower, in each of these spheres.)

Bergson, H.-L. (1934) *La pensée et le mouvant: essais et conférences*, Paris: Alcan; trans. M.L. Andison, *The Creative Mind*, New York: Philosophical Library, 1946.(Essays, historical and on method, including 'Introduction à la métaphysique' and two new introductions which form perhaps the best entrance-point for newcomers to Bergson.)

Bergson, H.-L. (1959) *Oeuvres*, Paris: Presses Universitaires de France.(Centennial edition containing all the above except *Durée et simultaneité*, with an introduction by H. Gouhier and critical and historical notes by A. Robinet.)

Bergson, H.-L. (1972) *Mélanges*, Paris: Presses Universitaires de France.(Edited with notes by A. Robinet and with foreword by H. Gouhier, this contains virtually all Bergson's writings whose publication he allowed (except those in *Oeuvres* 1959), including his early thesis on place in Aristotle and *Durée et simultanéité*.)

References and further reading

Barlow, M. (1966) *Henri Bergson*, Paris: Presses Universitaires de France.(Brief French biography which succeeds in integrating Bergson's life and works.)

Barreau, H. (1973) 'Bergson et Einstein: A propos de *Durée et simultanéité*', *Les Études bersoniennes* 10: 73-134.(Extended but accessible discussion, critical of Bergson.)

Berthelot, R. (1913) *Un Romantisme utilitaire: étude sur le mouvement pragmatiste. Troisième partie. Un Pragmatisme psychologique; le pragmatisme partial de Bergson (A Utilitarian Romanticism: A Study of the Pragmatist Movement. Part 3: A Psychological Pragmatism; The Partial Pragmatism of Bergson)*, Paris: Alcan.(Full and scholarly, though rather unsympathetic, treatment of Bergson and influences on him.)

Boudot, M. (1980) 'L'Espace selon Bergson' (Space According to Bergson), *Revue de Métaphysique et de Morale* 85 (3): 332-56.(Mentioned in §2. Very hostile discussion, claiming, among other things, that the asymmetry Bergson sees between space and time is illusory.)

Čapek, M. (1971) *Bergson and Modern Physics*, Dordrecht: Reidel.(Full and scholarly treatment of Bergson, going well beyond its title, but accessible. Much more sympathetic to Bergson than Berthelot.)

Čapek, M (1980) 'Ce qui est vivant et ce qui est mort dans la critique bergsonienne de la rélativité' ('What is living and what is dead in Bergson's critique of relativity'), *Revue de synthèse* 101 (99-100): 313-44. (Sympathetic and accessible.)

Gale, R.M. (1973-4) 'Bergson's analysis of the concept of nothing', *The Modern Schoolman* 51: 269-300.(On a topic on which Bergson says some important things in chap. 4 of *Creative Evolution*, though not covered in this entry.)

Gunter, P.A.Y. (1974, 1986) *Henri Bergson: A Bibliography*, Bowling Green, OH: Philosophy Documentation Centre.(Massive work with over 6,000 entries by and on Bergson, many with summaries, extensive in the case of Bergson's main works.)

Heidsieck, F. (1957) *Henri Bergson et la notion d'espace (Henri Bergson and the Notion of Space)*, Paris: Le Cercle du Livre.(Influential in the rehabilitation of Bergson after the Second World War. See especially the

discussion, with some technicalities, of *Durée et simultanéité*, claiming that Bergson was right in spirit, though not in letter.)

Husson, L. (1947) *L'intellectualisme de Bergson*, Paris: Presses Universitaires de France.(Emphasizes that Bergson's use of intuition does not imply that he was anti-intellectualist.)

Kolakowski, L. (1985) *Bergson*, Oxford: Oxford University Press.(Brief elementary overview.)

Lacey, A.R. (1989) *Bergson*, London: Routledge.(The book on which this entry is based. Confined to philosophy, not the history of ideas.)

Milet, J. (1974) *Bergson et le calcul infinitésimal (Bergson and the Infinitesimal Calculus)*, Paris: Presses Universitaires de France.(Referred to in §2. Nontechnical discussion claiming Bergson saw the calculus as a means of dealing with duration, not as something limited to the 'cinematographic' method. See English résumé of his views in Papanicolaou and Gunter 1987.)

Moore, F.C.T. (1996) *Bergson: Thinking Backwards*, Cambridge: Cambridge University Press.(Good attempt to give brief and accessible expression to some of Bergson's more difficult doctrines, and to bring out their significance.)

Mullarkey, J. (ed.) (1998) *The New Bergson*, Manchester, Manchester University Press.(Essays aimed at bringing out the contemporary relevance of Bergson in a wide variety of spheres.)

Papanicolaou, A.C. and Gunter, P.A.Y. (eds) (1987) *Bergson and Modern Thought*, Chur, Switzerland: Harvard Academic Publishers.(Essays, sometimes technical, claiming Bergson anticipated elements of various modern scientific developments, which confirm many of his ideas.)

Berkeley, George (1685-1753)

George Berkeley, who was born in Ireland and who eventually became Bishop of Cloyne, is best known for three works that he published while still very young: An Essay towards a New Theory of Vision (1709), Three Dialogues between Hylas and Philonous (1713), and in particular for A Treatise concerning the Principles of Human Knowledge (1710). In thePrinciples he argues for the striking claim that there is no external, material world; that houses, trees and the like are simply collections of 'ideas'; and that it is God who produces 'ideas' or 'sensations' in our minds. The New Theory of Vision had gone some way towards preparing the ground for this claim (although that work has interest and value in its own right), and the Dialogues represent Berkeley's second attempt to defend it. Other works were to follow, including De Motu (1721), Alciphron (1732) and Siris (1744), but the three early works established Berkeley as one of the major figures in the history of modern philosophy.

The basic thesis was certainly striking, and from the start many were tempted to dismiss it outright as so outrageous that even Berkeley himself could not have taken it seriously. In fact, however, Berkeley was very serious, and certainly a very able philosopher. Writing at a time when rapid developments in science appeared to be offering the key to understanding the true nature of the material world and its operations, but when scepticism about the very existence of the material world was also on the philosophical agenda, Berkeley believed that 'immaterialism' offered the only hope of defeating scepticism and of understanding the status of scientific explanations. Nor would he accept that his denial of 'matter' was outrageous. Indeed, he held that, if properly understood, he would be seen as defending the views of 'the vulgar' or 'the Mob' against other philosophers, including Locke, whose views posed a threat to much that we would ordinarily take to be common sense. His metaphysics cannot be understood unless we see clearly how he could put this interpretation on it; and neither will we do it justice if we simply dismiss the role he gives to God as emerging from the piety of a future bishop. Religion was under threat; Berkeley can probably be judged prescient in seeing how attractive atheism could become, given the scientific revolution of which we are the heirs; and though it could hardly be claimed that his attempts to ward off the challenge were successful, they merit respectful attention. Whether, however, we see him as the proponent of a fascinating metaphysics about which we must make up our own minds, or as representing merely one stage in the philosophical debate that takes us from Descartes to Locke and then to Hume, Kant and beyond, we must recognize Berkeley as a powerful intellect who had an important contribution to make.

1 Life

George Berkeley was born in (or near) the town of Kilkenny, Ireland, and educated at Kilkenny College and at Trinity College, Dublin, where he took the degree of B.A. in 1704, and that of M.A. in 1707, becoming a Junior Fellow in the latter year. Before long he published the books for which he is now most renowned. However, mention must first be made of two notebooks, now known as the *Philosophical Commentaries*, which he filled during the years 1707-8. Since their first publication in 1871 (but more particularly since it was established that they had at some stage been bound together in the wrong order, thus giving a distorted picture of the development of Berkeley's thought) these have proved an invaluable resource for scholars seeking to understand the evolution of his thinking during this crucial period. The major fruits of that thinking were *An Essay towards a New Theory of Vision* (1709), *A Treatise concerning the Principles of Human Knowledge* (1710) - which was originally intended to be merely Part I of a three- or four-part work- and the *Three Dialogues between Hylas and Philonous* (1713), which Berkeley published after he had moved to London. In between the *Principles* and *Dialogues* he published a slighter work, *Passive Obedience* (1712), which gives the main insight into his thinking on ethics, and on the basis of which he has been described as a theological rule-utilitarian. Also dating from about this time there are essays published in Richard Steele's *Guardian* during the year 1713, which evidence his disdain for the antireligious sentiments of the 'free-thinkers'.

From this time onwards, Berkeley's life was active and interesting. He made two continental tours, the first (1713-14) as chaplain to Lord Peterborough, during which he apparently met Malebranche, and the second (1716-20) as tutor to George Ashe, son of the Bishop of Clogher. Towards the end of the second tour he wrote the Latin tract *De Motu* for submission to the Royal Academy of Sciences at Paris, which had offered a prize for an essay on the cause of motion. He published this in 1721, returned to Ireland in the same year, and was appointed Dean of Derry in 1724. Already, however, he had conceived a remarkable project that was to dominate his life for

ten years. During the spring of 1722 he resolved to found a college on the island of Bermuda, and before long he set about soliciting support for and gaining a charter for St Paul's College, which would, had it come into existence, have educated a number of young Native Americans, as well as the sons of English planters.

In fact he never reached Bermuda but, newly married, he set sail for Rhode Island in 1728, where he stayed for over two years awaiting a promised government grant, and where his house is preserved as a monument to him. The grant never materialized, so there was to be no college, either in Bermuda or, as he had come to think would be preferable, on the mainland. His time in Rhode Island was not, however, wasted. While there he wrote *Alciphron: or the Minute Philosopher*, an attack on atheism and deism in dialogue form, which was published in 1732, the year after his return to London. He also became a friend of Samuel Johnson, later the first president of King's College, New York. Johnson's *Elementa Philosophica* (1752) is dedicated to Berkeley, and two letters from Johnson written in 1729 and 1730 (published with Berkeley's replies in volume two of the standard edition of Berkeley's *Works*) reveal that he was basically sympathetic to, but also an acute critic of, Berkeley's main metaphysical doctrines.

Certainly the same could not be said of Andrew Baxter, who in 1733 included as part of his *Enquiry into the Nature of the Human Soul* what was, in fact, the first extended critique of Berkeley's *Principles*. Baxter's tone was hostile throughout. Berkeley chose not to respond, though in the same year he did answer an anonymous critic of the *New Theory of Vision* - a third edition of which had been annexed to *Alciphron* - by publishing *The Theory of Vision, Vindicated and Explained*. He also published a revised edition of the*Principles* and *Dialogues* in 1734. *The Analyst* (1734), which criticizes Newton's doctrine of fluxions, also relates to his earlier work in that Berkeley refers back to his observations on mathematics in the *Principles*, and it may be that remarks Baxter had made on his treatment of the mathematicians there played at least a minor role in encouraging him to publish it. Berkeley does not name the critic who, he says, had challenged him to 'make good' what he had said in the *Principles*, but if it was Baxter he treats him dismissively as someone who 'doth not appear to think maturely enough to understand either those metaphysics which he would refute, or mathematics which he would patronize' (*The Analyst* §50).

However, Berkeley also had to think about securing his and his family's future, and his efforts to gain preferment in the church were rewarded in 1734 when he was appointed Bishop of Cloyne in Ireland. There, he thoroughly earned the reputation he has had ever since as 'the good Bishop'. The tangible legacy includes *The Querist* (1735-7), which evidences his concern for the economic wellbeing of Ireland, and *Siris* (1744), so successful at the time that it went through six editions in the year of publication but which is now regarded as little more than a curiosity. However, this was to be his last original publication of any substance. He remained in Cloyne almost to the end of his life, moving to Oxford, where one of his sons was to study, in the summer of 1752. He died there in the following year.

2 Influences

The primary influence on Berkeley is unquestionably John Locke, whose *Essay concerning Human Understanding* Berkeley had studied as an undergraduate and continued to dwell on afterwards. The long introduction to Berkeley's *Principles* is for the most part a sustained attack on the view that we can frame abstract ideas, focusing on Locke's account of abstraction. Illegitimate abstraction is ultimately blamed for the supposedly untenable distinction between primary and secondary qualities, the belief in 'material substance', and the view that objects have an existence distinct from 'ideas', all of which are features of Locke's position (see Locke, J. §§2-5). Yet Berkeley also owed a great deal to Locke whom he likened in the notebooks to 'a Gyant' and who should be seen as his mentor as well as one of his philosophical targets. It is therefore understandable that Berkeley has most often been seen as the second of the three great British Empiricists, as successor to Locke and precursor of Hume, these three being placed in opposition to the three great Rationalist philosophers, Descartes, Spinoza and Leibniz. Certainly, it would be tempting to say that the importance of Locke's influence on Berkeley could hardly be overestimated, were it not for the fact that it sometimes has been.

If only as a corrective, then, it is important to stress that while it is evident that Locke was often in Berkeley's mind as he formulated his own position, and while there is no doubt that none of Berkeley's major works would have existed in their present form had Locke never published the *Essay*, Berkeley would have insisted that much more was at stake than whether Locke got things right. He targeted certain views and assumptions that were very widely held. Thus Locke is the only philosopher he actually identifies and quotes from in the attack on abstract

ideas, but even there he sees himself as opposing, not simply some quirky view of Locke's, but one which, as he put it in a letter, 'Mr. Locke held in common with the Schoolmen, and I think all other philosophers' (*Works*, vol. 2: 293). These certainly included Malebranche, for example, who, Berkeley elsewhere complained, 'builds on the most abstract general ideas' (*Works*, vol. 2: 214). Again, when he says that 'Some there are who make a distinction betwixt *primary* and *secondary* qualities' (*Principles* §9), he really does mean 'some', and not just Locke; and the same could be said of his opposition to the notion of 'material substance'. In short, Berkeley often had his eye on other thinkers too, and some of these must also count as influences. As is now widely recognized, these included writers in the Cartesian tradition, most notably Malebranche but also probably Pierre Bayle.

The relationship between Berkeley and Descartes is interesting - after all, it was Descartes who had introduced a radical dualism of 'matter' and 'mind', and although Berkeley rejected matter, he adhered to a broadly Cartesian view of the mind (see Dualism). However, Malebranche is particularly important in the story, both because Berkeley had studied his *De la recherche de la vérité* at an early stage, and because Berkeley's position struck many as remarkably close to that of Malebranche. In particular, Berkeley positively denies the existence of bodies 'without the mind', but Malebranche had already argued that it was impossible to prove their existence conclusively, thus paving the way for their dismissal. Again, Malebranche had insisted that there are no corporeal *causes*, and that, strictly speaking, God is the only cause, and Berkeley certainly holds that only spirits can act. Moreover, Malebranche held, and Berkeley at least suggested, that in perception, God's ideas are revealed to us. It is significant, then, that in his own day, despite his protestations, Berkeley was often seen as essentially a follower of Malebranche. We might note, finally, that while Malebranche had concluded that neither sense nor reason could conclusively establish the existence of bodies, he also held that faith in the Scriptures did require this belief. When in the *Principles* Berkeley considers a number of possible objections to his positive rejection of 'matter', this argument from the Scriptures is the last that he chooses to tackle. As he says, 'I do not think, that either what philosophers call *matter*, or the existence of objects without the mind, is any where mentioned in Scripture' (*Principles* §82).

There is evidence that Bayle too was an early influence, and when, as in the preface to the *Principles*, Berkeley refers to 'those who are tainted with scepticism', arguments he found in Bayle's *Dictionnaire historique et critique* were probably towards the front of his mind. Bayle had offered arguments against regarding extension and motion as any more objective than colour or smell (which the Cartesians recognized as mere 'sensations'), and for the view that the notion of real extension (for Cartesians the essence of matter) involved contradictions. Strict reasoning, Bayle argued, would thus lead us to deny the existence of bodies, in the face of our (fortunately) ineradicable beliefs. Berkeley could welcome and adapt these arguments to the extent that he was concerned to reject bodies 'without the mind', and while, unlike Bayle, Berkeley firmly denied that they lead to scepticism or to any conflict with common sense, it is hardly surprising if many of his contemporaries took a different view. As Andrew Baxter saw it, Berkeley was committed to the conclusion that 'he has neither *country* nor *parents*, nor any *material body* (but that all these things are mere *illusions*, and have no existence but in the fancy' (Baxter [1733] 1737, vol. 2: 260).

3 Berkeley's metaphysics

Berkeley is understandably best known for his (at first sight outrageous) claim that mind or spirit is the only substance, and that it is God who produces 'sensations' or 'ideas' in our minds. From the beginning, many regarded this view as sceptical at best or insane at worst, and Berkeley recognized that this might be the initial reaction. It is, then, an important feature of his position that, if rightly understood, his standpoint will be seen as common sense, and in accord with the views of the unsophisticated 'vulgar'. The purpose of the present section is to sketch in very general terms how Berkeley could see things in this way.

To begin with, we can hardly make sense of Berkeley's position unless we see him as starting from an assumption that he took both to be obviously true and to be shared by other philosophers, which was that each of us is aware only of the 'ideas', 'sensations' or 'perceptions' that are somehow or other produced in our minds. On the most common view - that taken by Descartes and Locke for example - these are produced in us by external objects, which objects we do not perceive 'immediately' because, as Locke put it (whatever precisely he meant by it), 'the Mind… perceives nothing but its own *Ideas*' (Locke, *Essay* IV 4: §3). Berkeley's first insight, and it is one that his reading of Malebranche and Bayle must have encouraged, was that if we set things up in this way - distinguishing

between the ‘ideas’ we perceive and the ‘real’ objects which lie hidden beyond them - scepticism becomes inevitable. At best we can hypothesize the existence of ‘real’ objects as the *most likely* causes of our ideas, but then we are vulnerable to the suggestion that there could be other causes, including, most plausibly, God. There are other difficulties too. Berkeley found it widely admitted that it is quite unclear how inert ‘matter’ could act on minds so as to produce ideas or perceptions in them (and Malebranche and other ‘occasionalists’ had denied that it in fact does) (see Occasionalism); moreover, Berkeley found only obscurities and incoherencies in the prevalent conceptions of ‘material substance’. Yet the most fundamental insight was to follow.

This insight was that when we - ordinary men and women - talk of houses, mountains, rivers and so on, we are talking about what we experience or are aware of, not of occult objects that we are not directly aware of at all. It follows, or at least it seemed to Berkeley to follow, that if when we refer to houses, mountains and rivers we are referring to things we are aware of, and if (as other philosophers agreed) we are aware only of ideas, houses, mountains and rivers must *be* ‘ideas’ or appearances or, better, ‘collections’ of such ideas. Certainly - and this was one thing that his readers found most difficult to handle, but which Berkeley himself was most insistent on - there is no need to deny that houses, mountains and rivers exist, but only to stress (common-sensibly) that they are the very things we perceive, which is to say that they are mind-dependent ideas. Their *esse* (being) is *percipi* (to be perceived); they exist only in the mind.

Berkeley’s major philosophical works, and in particular the *Principles* and *Dialogues*, are, in the main, a sustained defence of these insights and doctrines, together with a working out of their implications. For Berkeley, the implications, including those for religion and the sciences, are as important as the basic metaphysics. Yet the fundamental case for that metaphysics is supposed to be very simple indeed. Even by the end of section six of the *Principles* (under three pages in most editions) that case has supposedly been established.

4 The *New Theory of Vision*

Although Berkeley’s *An Essay towards a New Theory of Vision* (1709) was published just one year before the *Principles*, and Berkeley was already convinced that there was no such thing as ‘matter’, or bodies ‘without the mind’, this, his first major work, stopped short of making that claim. As he said in the *Principles*, although the earlier book had shown that ‘the proper objects *of sight* neither exist without the mind, nor are the images of external things (*Principles* §44; emphasis added), it had done nothing to disabuse readers of the view that tangible objects are external. At one level, then, the work can be seen as a sort of halfway house on the route to presenting his full case for immaterialism, but it is undoubtedly also true that he was fascinated by problems concerning vision in their own right. He was clearly very well read in optical theory, he had his own highly distinctive contribution to make, and for many years that contribution was esteemed by many who had little interest in, or were possibly quite blind to, any wider implications it may have had.

Ostensibly, then, the *New Theory of Vision* is merely an attempt to ‘shew the manner wherein we perceive by sight the distance, magnitude, and situation of objects’, though, still in the opening section, Berkeley also announces that he will be considering ‘the difference there is betwixt the ideas of sight and touch, and whether there be any idea common to both senses’ (*New Theory of Vision* §1). Broadly, the issue concerning ‘situation’, which others had recognized, is that of how we see things the ‘right’ way up (so to speak) when their images are inverted on the retina; that concerning ‘magnitude’ is how we judge objects at a distance to be small or large (one particular problem was why the moon on the horizon looks larger than the moon in the zenith, although they are virtually the same distance from us); and that concerning ‘outness’ or distance is that of how we come to see things as being at various distances, given that, as Berkeley observes, it was accepted that ‘distance being a line directed end-wise to the eye, it projects only one point in the fund of the eye, which point remains invariably the same, whether the distance be longer or shorter’ (*New Theory of Vision* §2). Berkeley’s solution is similar in each case. In the case of distance, for example, even when an object is relatively close, we do not, as others had supposed, make our judgments on the basis of what Descartes had described as a sort of ‘natural geometry’, and on facts such as that lines drawn from the two eyes to the object form a greater angle the closer the object is: the supposed lines and angles are only theoretical entities, and are not at any rate perceived. Rather, we *learn* to make these judgments solely on the strength of certain sensory cues including, for example, the sensations accompanying the turn of the eyes, and the increasingly confused appearance of an object as it comes closer to us. An explanation in terms of geometry is thus replaced by a psychology of vision in which, crucially, the connection between the cues

and the distance discoverable by touch turns out to be purely contingent. '[I]f it had been the ordinary course of Nature that the farther off an object were placed, the more confused it should appear, it is certain the very same perception that now makes us think an object approaches would then have made us to imagine it went farther off' (*New Theory of Vision* §26).

Though often regarded as controversial, Berkeley's work on the psychology of vision was also highly influential even though, and indeed partly because, Berkeley's ultimate metaphysical commitments are not apparent, and certainly not necessarily required for an acceptance, for example, that 'a man born blind, being made to see, would at first have no idea of distance by sight'. Admittedly Berkeley's account of our judgments is in terms of 'sensations', 'appearances' and 'ideas', as all we have to go on, and we are told, for example, not only that the man just cured of blindness would take the 'objects intromitted by sight' to be 'no other than a new set of thoughts or sensations, each whereof is as near to him as the perceptions of pain or pleasure, or the most inward passions of his soul', but that he would be right to do so (*New Theory of Vision* §41). Yet nothing is said to disabuse the reader of the thought that there is, for example, a distant moon, which is not at all dependent on the mind. There is a sense, therefore, in which the *New Theory of Vision* offers us some of the fruits of idealism without explicitly announcing the immaterialism, and one of those fruits is an indication of the existence of God (see Idealism). By the end of the work, Berkeley has concluded that there are no ideas common to sight and touch: the extension perceived by touch, for example, is quite distinct from, and has no likeness to, any visual idea. Here he considers a problem first raised by William Molyneux and discussed by Locke, agreeing with them that a man just cured of blindness who saw a cube and a globe for the first time would not know just by looking which was which, but seeing this answer as confirming his own view that visual ideas are merely 'signs'. These we learn to correlate with tangible ideas in much the same way as we learn a language. Berkeley takes this analogy very seriously. His conclusion in the first edition is thus that 'the proper objects of vision constitute the universal language of nature, whereby we are instructed how to regulate our actions', but by the third edition 'nature' has become 'the Author of nature', or God (*New Theory of Vision* §147).

5 The Introduction to the *Principles*

Berkeley prefaces *A Treatise concerning the Principles of Human Knowledge* (1710) with an important introduction which is for the most part devoted to an attack on abstract ideas, and in particular abstract general ideas. In it he quotes freely from Locke. Yet, as already stated, his target was wider, including philosophers generally and, ultimately, a variety of philosophical confusions. One needs to look outside the introduction to discover what these alleged confusions are. Sometimes this is fairly straightforward. Even in the *New Theory of Vision* the notion that there is an idea of extension common to both sight and touch is ascribed to the supposition that we can abstract it from all other visible and tangible qualities; while, in the *Principles*, the notion that the supposed 'primary' qualities exist in the outward object, although colours and the like are 'in the mind alone', is undermined by the observation that 'extension, figure, and motion, abstracted from all other qualities, are inconceivable'. Similarly, the idea of 'pure' or 'absolute' space is ruled out, it being 'a most abstract idea'. In one important case the connection is perhaps less obvious: Berkeley claims that holding that sensible objects can exist unperceived depends on illegitimate abstraction, but commentators have often found it difficult to see precisely how this is supposed to work. In yet other cases, the supposed connections have been less frequently explored in the literature, as for example when Berkeley has it that the Schoolmen were 'masters of abstraction' and, in the *Dialogues*, that Malebranche 'builds on the most abstract general ideas'. These matters can probably be sorted out. Malebranche had attacked the 'disordered abstractions' of the Schoolmen, who posited occult qualities and powers, and who supposed that matter is something distinct from its known attributes, and in particular from extension, and Berkeley had probably learned from that. Yet Malebranche himself fell foul of Berkeley's anti-abstractionism by talking of 'absolute' and 'intelligible' extension, by supposing that extension was the essence of matter, and by assuming an idea of 'being in general'. The connection between abstraction and the denial of the '*esse* is *percipi*' principle is trickier.

Berkeley's introduction attacks the view that, although the qualities of objects are always 'blended together' in them, we can frame a separate idea of each quality; that we can form, for example, an abstract idea of colour or extension in general; and that we can frame an idea corresponding to the word 'man' or 'triangle', as distinct from the ideas of particular men or particular triangles, as Locke had suggested. This in turn requires from Berkeley an alternative account of language to Locke's, which will not require that each general term stands for an idea. This

alternative account is not worked out very fully, but Berkeley does insist that 'a word becomes general by being made the sign, not of an abstract general idea but, of several particular ideas, any one of which it indifferently suggests to the mind' (*Principles*, Intro. §11). Moreover, suggestions towards the end of the introduction that words have other uses than to mark out ideas, including the production of appropriate emotions - 'May we not, for example, be affected with the promise of a *good thing*, though we have not an idea of what it is?' (*Principles*, Intro. §20) - have rightly been seen as significant, and further developments along these lines, in particular in the seventh dialogue of *Alciphron*, have even been seen as making him a precursor of Wittgenstein in this area.

6 The *Principles*

Berkeley's basic metaphysical position is usually known as 'idealism' or, because of what it denies, as 'immaterialism', and the classic defence of this position is offered in *A Treatise Concerning the Principles of Human Knowledge*. Like all Berkeley's works, this is well structured, with just 156 short sections: sections 1-33 argue the case for his idealism, sections 34-84 anticipate and answer possible objections, and the remaining sections take 'a view of our tenets in their consequences'.

As already indicated, Berkeley takes even his opponents to accept that, whatever else there may turn out to be in the world, we perceive only ideas. This assumption emerges in the opening section of the *Principles* (which is clearly modelled on the opening sections of the first chapter of Book II of Locke's *Essay*). Here Berkeley writes, or at least suggests, that 'the objects of human knowledge' are all 'ideas', adding that when certain ideas, for example a certain colour, smell and so on, are found going together they are 'reputed as one thing'. On the face of it, this blurs Locke's distinction between 'qualities' and 'ideas', and ignores Locke's supposition of a 'substratum' for the qualities. Yet Berkeley knows what he is doing, and clearly found encouragement in Locke's own preparedness not only to use 'idea' where he means 'quality', but also to assert that we have no other ideas of particular sorts of substances 'than that which is framed by a collection of those simple *ideas* which are to be found in them'. Certainly, we are supposed to start with 'ideas', although - as Berkeley points out in the second section -there are also the minds or spirits that perceive them. However, he soon insists that there can be no substance apart from mind. Given that sensible objects are ideas, and that ideas exist only when perceived, it becomes simply absurd to suppose that these objects could have any existence apart from perception; a fact that is confirmed, in Berkeley's view, simply by attending to 'what is meant by the term *exist* when applied to sensible things'. When I say that a table 'exists', I am referring to something that I perceive, or at least that I might perceive, and certainly not applying 'exists' to some object which, because it is not an idea, is not perceived at all.

This argument, like most of Berkeley's arguments, is tricky and needs careful handling. Ostensibly, it seems to have very little to do with the word 'exists' because, as Andrew Baxter observed, neither philosophers nor ordinary people seem to mean 'is perceived' by 'exists' in sentences such as 'the table exists'. That point is a fair one, and Berkeley's actual argument does seem to depend heavily on the underlying assumption that the only perceivable objects are mind-dependent items, which must consequently be actually perceived. The stress put on the word 'exists' remains puzzling, however, and one relevant fact seems to be that Locke had held that 'existence' was a simple idea 'suggested to the Understanding, by every Object without, and every *Idea* within' (Locke, *Essay*, II 7: §7). Berkeley had convinced himself both that the idea thus described was abstract (and hence impossible), and that this idea is involved when people suppose things to exist quite independently of perception. To perceive a table as existing and to simply perceive it are one and the same experience, and the existence cannot be separated from the perception so that we can attribute an 'absolute existence' to the thing.

That is at any rate what Berkeley concludes on the basis of the first few sections. But of course he expected resistance. His tactic now becomes, therefore, to seize on supposedly unsatisfactory features of his opponents' position and, by exposing them, to further his own case. If it is suggested, for example, that our ideas are merely the *likenesses* of external qualities, the counter is that an idea (or perceived thing) can be *like* nothing but an idea (or another perceivable thing). To those who argue that the supposed 'primary' qualities exist in outward objects but that colours and the like do not, his response is twofold: first, we cannot even conceive of an object having merely extension, figure and motion, but lacking any of the qualities these other philosophers recognize as mind-dependent; second, the basic argument deployed to prove that secondary qualities are mind-dependent (that is, that the appearance varies in varying circumstances) would prove the same of any quality whatsoever.

Furthermore, those who posit a material substratum as the *support* of qualities find that they can attach no clear meaning in this context even to the term ‘support’. There are other arguments, including a particularly tricky and much discussed one in which he proudly claims that it is impossible to conceive that there even *might* be a mind-independent object, for to conceive it would be to frame the idea of it, which would mean that it was an object of thought or perception after all. However, Berkeley is at his rumbustious best in sections 18-20, arguing that neither sense nor reason can establish that there are external bodies, and that they cannot even be posited as an hypothesis to account for our receiving the ideas we do. Even if we suppose, arbitrarily, that there are external bodies, the materialists ‘by their own confession are never the nearer knowing how our ideas are produced: since they own themselves unable to comprehend in what manner body can act upon spirit, or how it is possible it should imprint any idea in the mind’ (*Principles* §19). What emerges, predictably, is that the only possible cause of our ideas is another, superior spirit, who presents our ideas to us in orderly ways which in fact *constitute* the Laws of Nature, and which Berkeley also sees as constituting the *language* of God himself.

7 The *Principles* (cont.)

While the first thirty-three sections of the *Principles* are in an obvious sense basic, the sections in which Berkeley deals with possible objections to his thesis are important too. Here most readers new to Berkeley are likely to find that the first objections that spring to their minds have been anticipated, while the answers Berkeley gives help to clarify his basic thesis. The objections he envisages include, for example, that, given his idealism, everything becomes illusory or unreal; that we see things at a distance from us, so they are not ‘in the mind’; that, if the *esse* of sensible things is *percipi*, they will disappear when we stop perceiving them, which is absurd; and that, if objects are only ideas, or collections of ideas, there can be no causal interaction between them, so we will have to deny that fire heats and that water cools. Whether Berkeley’s answers to such objections satisfy us is another matter, but the objections are at least confronted, and the answers are always interesting. On the third objection mentioned above, for example, it is eventually suggested that for an object to exist it is necessary only for some mind to perceive it, with the implication that God’s perception may guarantee the continued existence of objects. The answer to the fourth objection above is that, just as we continue to say that the sun ‘rises’ despite scientific knowledge that it is the earth that moves, so this is another area where ‘we ought to *think with the learned, and speak with the vulgar*’ (*Principles* §51), recognizing that, strictly, the regularities in nature we describe as causal are ultimately down to God. In answering both these objections, Berkeley is typically quick to point out that his philosophical opponents are insecurely placed to make them. Even those who hold that there are external and material bodies are committed to the view that light and colours, or visible objects, are ‘mere sensations’, and thus to holding that these disappear when I shut my eyes; while, when it comes to causal relationships between objects, many other philosophers, both among the Schoolmen and modern philosophers, have held that God is the ‘immediate efficient cause of all things’.

In answering the second of the above objections, Berkeley predictably refers the reader back to the *New Theory of Vision*; but his answer to the first objection is more complex. There are, he stresses, decisive differences between the ‘faint, weak, and unsteady’ ideas of the imagination and those imprinted on the senses by God, and though he calls both ‘ideas’ to emphasize that they are equally in the mind, he would not object to simply calling the latter ‘things’. Nor does he deny even that there are corporeal substances, if ‘substance’ is taken ‘in the vulgar sense, for a combination of sensible qualities’. It is, he suggests, only other *philosophers* he opposes, for they take corporeal substance to be ‘the support of accidents or qualities without the mind’. We may well feel that this point glosses over the one big difference between Berkeley and the vulgar, which is that the vulgar do not recognize sensible qualities to be mind-dependent ideas, but it is one that Berkeley insists on. ‘The only thing whose existence we deny, is that which philosophers call matter or corporeal substance. And in doing of this, there is no damage done to the rest of mankind, who, I dare say, will never miss it’ (*Principles* §35).

8 The *Principles* (cont.)

The full title of the *Principles* describes it as a work ‘Wherein the chief causes of error and difficulty in the Sciences, with the grounds of Scepticism, Atheism, and Irreligion, are inquired into’. While Berkeley believes idealism to be true, he is as interested in the benefits that flow from accepting it. These include establishing the existence of God and attaining a proper understanding of God’s role in the world; the banishment of scepticism concerning the nature and the very existence of ‘real’ things, both of which result from distinguishing the ‘real’

from what we perceive; and the resolution of certain philosophical, scientific and mathematical perplexities. From section 85 onwards, therefore, Berkeley takes ‘a view of our tenets in their consequences’.

Some of the supposed advantages are obvious once stated, and they include the resolution of three issues Berkeley mentions at the outset: ‘Whether corporeal substance can think?’ (a possibility mooted by Locke, which threatened belief in the natural immortality of the soul); ‘Whether matter be infinitely divisible?’ (a long-standing issue, with Bayle in particular having exposed the paradoxes that arise whether we suppose that it is or it is not); and ‘how [matter] operates on spirit?’ (a problem that had exercised the Cartesians). None of these questions arises once it has been proved that there is no ‘matter’; that the soul is immaterial, or ‘one simple, undivided, active being’ which is therefore ‘indissoluble by the force of Nature’; and that, just as we can produce ideas in our own minds when exercising our fancies, so God (the superior spirit) can produce in our minds those ideas which constitute sensible things. In addition, however, Berkeley explores at some length the implications for natural philosophy and mathematics.

These, it must be stressed, were not simply casual interests for Berkeley. His very first publication - a compilation of two titles, *Arithmetica* and *Miscellanea Mathematica*, (1707) - evidences his early proficiency in mathematics, and the philosophically more significant manuscript *Of Infinites* was written at about the same time. The latter concentrates on the ‘disputes and scruples’ which infect modern analytical geometry, all arising from ‘the use that is made of quantitys infinitely small’. Moreover *De Motu* (1721) includes an examination of the role that such concepts as force, gravitation and attraction play in Newtonian mechanics. There would have been more on these topics in the additional parts of the *Principles* which Berkeley intended to write, as indeed there would have been on persons, perceivers or spirits. What he does say on the latter subject in the*Principles* as we have it is thin, and it is perhaps necessary only to note that Berkeley’s view is indeed broadly Cartesian, though the Berkeleian dualism is between ‘indivisible, incorporeal, unextended’ minds and ideas, not minds and ‘matter’; that he even convinces himself that the soul always thinks; and that the stress is on Berkeley’s claim that we do not know ourselves, or other spirits, by way of *idea*. This insistence underlies our earlier observation that in the opening section of the *Principles* Berkeley writes, or at least *suggests*, that all the objects of knowledge are ideas; for the truth is that, though Berkeley was prepared to give this impression at the outset (presumably so as not to raise an unnecessary complication early on), his own use of ‘idea’ for ‘any sensible or imaginable thing’, as he put it in the *Philosophical Commentaries*, rules out any ‘idea’ of spirit, or of the operations of the mind. Certainly, though, this is not supposed to be worrying, and Berkeley is not suggesting that the word ‘mind’ is insignificant. When he started penning the entries in the *Commentaries* he had indeed accepted the Lockian view that all significant words stand for ideas, but he had soon rejected that principle, partly as a result of deciding that the essentially active mind must be carefully distinguished from its passive objects or ‘ideas’.

9 *Three Dialogues between Hylas and Philonous*

The *Principles of Human Knowledge* is the most important book in the Berkeleian corpus and, had its reception not been so disappointing to Berkeley, the *Three Dialogues between Hylas and Philonous* (1713) would probably not have been written. People were readier to ridicule than to read a treatise that denied the existence of ‘matter’, while those who did read it usually misunderstood it. The *Dialogues*, therefore, were written, as Berkeley says in the preface,‘to treat more clearly and fully of certain principles laid down in the First [Part of the *Principles*], and to place them in a new light’, and the dialogue form proved an admirable way of allowing likely objections to be dealt with at each stage (as well as making the book still perhaps the most attractive introduction to Berkeley). The protagonists are Hylas (the name derives from the Greek word for ‘matter’) and Philonous (the ‘lover of mind’, representing Berkeley himself). At the outset Hylas assumes that the Berkeleian is the proponent of ‘the most extravagant opinion that ever entered into the mind of man’ (*Works*, vol. 2: 172), but, as the discussion progresses, Philonous is able to demonstrate that, although he accepts with other philosophers that ‘the things immediately perceived, are ideas which exist only in the mind’, his additional acceptance of the view of ordinary men and women that ‘those things they immediately perceive are the real things’ allies him with common sense (*Works*, vol. 2: 262).

Doctrinally there are no substantial innovations here, although Berkeley has Philonous take pains early on to convince Hylas that ‘sensible qualities’, or the things *immediately* perceived, are mind-dependent, making great play of how appearances vary for different perceivers, and for the same perceiver in different circumstances. Other

features include a striking passage, expanded in the third edition, which contains an anticipation of, and an attempt to answer, what is normally taken to be the Humean point that material and spiritual substance are on a par, so that if one is rejected, so too should the other. It is indeed a particularly attractive feature of the work that Hylas is allowed to be a quite pugnacious opponent who really does test the idealist's position. To give just one other instance, it is likely to occur to us that, if the things we perceive are identified with 'ideas' or 'sensations', surely each idea will be dependent on the particular mind that has or perceives it, with the apparently far from common-sense consequence that 'no two can see the same thing'. Berkeley's answer may or (more likely) may not satisfy us, but there is a deeper issue underlying Hylas' challenge which Berkeley himself may not have adequately explored. This concerns the relationship between particular ideas - whether described as 'sensations' or 'appearances' - and the 'collections' of ideas which, for Berkeley, constitute publicly observable objects. There are no more than hints that Berkeley may be prepared to countenance the notion that the permanently existent table is an archetypal idea in God's mind, and that we can be said to perceive it when we perceive any of the 'fleeting… and changeable ideas' which, to some degree, correspond to it.

10 *De Motu*

Berkeley intended to publish additional parts of the *Principles* and apparently made some progress on the second part, telling Samuel Johnson in 1729 that 'the manuscript was lost about fourteen years ago, during my travels in Italy, and I never had leisure since to do so disagreeable a thing as writing twice on the same subject', but that was as far as he got. Remarks in the *Philosophical Commentaries* suggest that one part would have been 'our Principles of Natural Philosophy', and we can assume that it would have included the sort of material covered in a work he did publish, *De Motu* (1721). This work reiterates and develops certain points already made in the *Principles* when Berkeley was taking 'a view of our tenets in their consequences', but although it is indeed assumed that minds are not corporeal, it would not have been apparent to the reader that Berkeley holds that the *esse* of sensible things is *percipi*. Rather, what is insisted on is that 'it is idle to adduce things which are neither evident to the senses, nor intelligible to reason' (*De Motu* §21), and that when we attribute gravity and force to bodies we are improperly positing occult qualities which take us beyond anything we can experience or conceive. 'Abstract terms (however useful they may be in argument) should be discarded in meditation, and the mind should be fixed on the particular and the concrete, that is, on the things themselves' (*De Motu* §4).

It is, therefore, idle to look to the qualities of bodies themselves in order to discover a cause of motion, for 'what we know in body is agreed not to be the principle of motion' (*De Motu* §24). Relying as we should on what we can conceive, we must look to mind for that principle, for we know from our ability to move our limbs that minds can act. On this basis we should conclude that 'all the bodies of this mundane system are moved by Almighty Mind according to certain and constant reason' (*De Motu* §32).

It is clear, then, that *De Motu* fits in with Berkeley's ultimate aim in all his philosophical writings, which is to bring out the dependence of the world upon God. Yet here, as in the case of everything he was to publish later, the elements of his metaphysics that had most perplexed the readers of the *Principles* and the *Dialogues* are either absent or in the background. Indeed, it is a feature of *De Motu* that Berkeley is anxious to present himself as representing a tradition going back to the ancient Greeks, but including the Schoolmen and the Cartesians, which recognizes the ultimate dependence of motion on God. Indeed, 'Newton everywhere frankly intimates that not only did motion originate from God, but that still the mundane system is moved by the same actus' (*De Motu* §32). It must be stressed, however, that it is not this supposed consensus that makes Berkeley's philosophy of science interesting, but his understanding of the proper role of the natural scientist as contrasted with that of the metaphysician. Terms such as 'gravity' and 'force', for example, have a legitimate use, in facilitating calculations on the basis of certain observable regularities in the behaviour of objects. We go wrong only if we confuse the discovery of regularities with genuine explanations of them. By contrast, absolute space and absolute motion, which were posited in Newtonian mechanics, are rejected outright, as indeed they were in the *Principles*. We should 'consider motion as something sensible, or at least imaginable,' and 'be content with relative measures' (*De Motu* §66). If there were but one body in the universe, it would make no sense to suppose that it moved (see Newton, I.).

11 *Alciphron* and *The Analyst*

Berkeley published *De Motu* in 1721 and nothing of any philosophical significance for over ten years thereafter.

Indeed none of his later writings matched in importance what had already appeared. Yet all were controversial, and some were taken very seriously at the time. These included *Alciphron* (1732) and *The Analyst* (1734) which represent, if in very different ways, Berkeley's commitment to defending religion against those seeking to undermine it.

Alciphron is composed of seven lively dialogues in which two Christian gentlemen, Euphranor and Crito, defend the religious and Christian standpoint against two 'free-thinkers', Alciphron and Lysicles. These are, of course, fictitious characters, but are allowed on occasion to present (or misrepresent, as many have claimed) the views of such actual, though unnamed, figures as the third Earl of Shaftesbury and Bernard Mandeville. Mandeville complained bitterly that his thesis that private vices are public benefits had been totally distorted in *Alciphron*; others have said the same of Berkeley's treatment of Shaftesbury's ethical theory. For all that, the book remains very readable. It contains, moreover, the only account of free-will published by Berkeley, and also the first explicit linking of the doctrine concerning the heterogeneity of the objects of sight and touch to a proof of the existence of God. Additionally there is a discussion in the Seventh Dialogue of particular interest in that it returns us to the topic of language.

The context is still the acceptability of religion, but at this point the objection from the free-thinker Alciphron is that the Christian religion is ultimately unacceptable, not because it can be shown to be false, but because it is straightforwardly unintelligible, involving, as it does, such meaningless notions as that of 'grace'. Here Alciphron appeals to the principle that 'words that suggest no ideas are insignificant'. Consequently this principle, which Berkeley had himself assumed in a demonstration of immaterialism nearly half way through the *Commentaries*, now becomes his explicit target. He reiterates his objection to abstract ideas, but also stresses the role of words in directing our practices, whether in mathematics and natural science, or in the religious sphere. It has been debated whether or not what we find here marks any decided shift from the line he had taken in the introduction to the *Principles*, and it is certainly true that Berkeley had long since moved towards the position he adopts here, but the discussion in *Alciphron* does reflect his mature consideration of the topic. It stresses the use of words as signs which, as he had put it to Samuel Johnson, 'as often terminate in the will as in the understanding, being employed rather to excite, influence, and direct action, than to produce clear and distinct ideas' (*Works*, vol. 2: 293).

By contrast with *Alciphron*, the *Analyst* is a technical work in the philosophy of mathematics, containing criticisms of Newton's calculus. The adequacy of these criticisms is still debated, but they were sufficiently acute to generate considerable controversy among the mathematicians. To this controversy Berkeley contributed two further works in 1735, *A Defence of Free-thinking in Mathematics*, and *Reasons for not replying to Mr. Walton's Full Answer*. Berkeley's theological preoccupations are again relevant in this area, for *The Analyst* was addressed to an unnamed 'infidel mathematician', who has generally been identified with Edmund Halley (of Halley's Comet fame). Halley had been reported as claiming that Christian doctrines were 'incomprehensible', and the religion an 'imposture'. Berkeley is able to take delight in answering that the objection comes ill from a mathematician. He targets what he saw as obscurities and contradictions in the calculus. Some of these result from assuming an increment of infinitesimal value which, without reaching zero, proceeds towards a limit of zero, allowing the analyst to predict the system's value at a conceptual point at which the increment becomes nothing. A consequence is that these 'ghosts of departed qualities' are both used and disregarded in one and the same proof. As already mentioned, Berkeley's interest in mathematics was of long standing, as was his opposition to infinitesimal quantities. He was able to show how these lead to absurdities in the calculus, and to argue against those who 'though they shrink at all other mysteries, make no difficulty of their own'. Moreover, he was able to do this without mentioning his own idealist view that, because *esse* is *percipi*, the smallest quantity must be what he had earlier called the *minimum sensibile*, which cannot be divided into parts.

12 *Siris*

Siris (1744), the last of Berkeley's writings of any substance, is also in many ways the strangest. His championship of tar-water as a useful remedy against many diseases (and as a possible panacea) is likely to strike us as foolish, though it was to some extent understandable given his apparently successful use of it in his diocese. Moreover, although it was practical experience that had led him to his belief in the virtues of tar-water, Berkeley does go deeply into the explanation of its effectiveness, relying on theories which gave prominence to the role of 'aether', or 'pure invisible fire', as the vital principle of the corporeal world. Here again we can now see that Berkeley was

wrong, although he was able to appeal to authorities, both ancient and modern. Indeed, this readiness to appeal to authorities, or to seek for maximum consensus, extends to the final sections in which his chain of philosophical reflections leads him to focus on God as 'the First Mover, invisible, incorporeal, unextended, intellectual source of life and being' (*Siris* §296). Here themes familiar from the early works re-emerge - including the view that 'all phenomena are, to speak truly, appearances in the soul or mind' (*Siris* §251) and that there are, strictly, no corporeal causes. Yet these are now tied in with what appear to be alien elements. There is a tendency to disparage the senses, and Berkeley's fascination with the philosophies of the ancients extends to a degree of sympathy for the Platonic Theory of Forms. That said, Berkeley's eclectic and somewhat hesitant approach in *Siris* is such that it would be wrong to look to it for evidence of a substantially new philosophical position. Though fascinating in its way, *Siris* now seems very dated indeed.

13 Concluding Remarks

Inevitably, Berkeley is famed for the metaphysics of the *Principles*, and *Dialogues*. It would be easy to multiply quotations from people who treated that metaphysics as absurd, but very wrong to suggest that all the reactions have been hostile, or that the more hostile responses have not frequently been based on misunderstandings. At the other extreme, John Stuart Mill was to refer to Plato, Locke and Kant among others when describing Berkeley as 'the one of greatest philosophic genius' (Mill [1871] 1978: 451), while A.A. Luce, the most prominent Berkeley scholar of the twentieth century, held Berkeley's views to be fundamentally correct, and to coincide with the common-sense view of the world. Even many who would be less effusive have at least seen Berkeley as playing an important role in the history of philosophy, if only as marking one important stage on the route from Locke to Hume, and then to Kant and modern idealism. Certainly, no serious commentators would judge that his views can be easily or simply dismissed, though they would often give very different accounts of what makes him important and interesting. Luce, for example, found the role God has to play in Berkeley's system attractive; Mill thought it an embarrassment. Phenomenalism, the theory of perception which Mill himself espoused, could indeed be described as 'Berkeley without God' (see Mill, J.S. §6; Phenomenalism).

The fact is that Berkeley was grappling with problems that are perennial in philosophy, including that of the relationship between appearance and reality, or between our experiences and what we take them to be experiences of. Their treatments of these issues have very often led philosophers to say things that would strike the 'vulgar' as strange, and if Locke's position, for example, seems initially more congenial (in that Locke never doubts the existence of a world corresponding to, but distinct from, our 'ideas', and treats scepticism in that area as absurd), Berkeley was neither the first nor the last to see him as, in effect, making knowledge of that world impossible. Berkeley did not invent the sceptical challenge that arises from insisting on a distinction between what we 'immediately' perceive and an external 'material' world; if his way of dealing with it is radical, one must recognize that 'idealism' in one form or another was to have quite a history - even now there are philosophers who are happy to use the label to describe their own philosophical positions.

Certainly Berkeley does sometimes exaggerate the extent to which he is at one with the 'vulgar', or with our ordinary views about the world. He may be quite right that he is at one with those who believe that '*those things they immediately perceive are the real things*'. Yet, as we saw in §9 above, it is only by combining this with the claim that '*the things immediately perceived, are ideas which exist only in the mind*', which he attributes to 'the philosophers', that he arrives at a theory concerning the nature of reality that is very much his own. Consequently, although he can chide his opponents for their commitment to such views as that 'the Wall is not white, the fire is not hot', remarking in the *Commentaries* (entry 392) that 'We Irish men cannot attain to these truths', many of his own claims, such as that 'Strictly speaking… we do not see the same object that we feel; neither is the same object perceived by the microscope, which was by the naked eye' (*Works*, vol. 2: 245), would strike the vulgar as equally odd. Berkeley's beliefs about what it is that we 'immediately' perceive may or may not be true, but clearly they are not vulgar views.

To be fair, Berkeley was not unaware that this was the position. For example, his comment that on the issue of causal relationships between objects we should '*think with the learned, and speak with the vulgar*' (*Principles* §51) suggests that the vulgar have not appreciated the truth of the matter; while claims that he opposes only other philosophers contrast with passages such as that in the *Principles* in which he actually refers to the 'mistake' of the vulgar who believe that the 'objects of perception [have] an existence independent of, and without the mind'

(*Principles* §56). To be sure, in the same discussion he suggests that they cannot *really* believe this, because the supposed belief involves a contradiction, and 'Strictly speaking, to believe that which involves a contradiction, or has no meaning in it, is impossible' (*Principles* §54), but the whole passage rests on the equation of the objects of perception with 'ideas', which is what *makes* the supposed belief contradictory.

The truth is, therefore, that for all his resolve in the *Commentaries* (entry 751) 'To be eternally banishing Metaphisics &c & recalling Men to Common Sense' (and what he seems to have in mind there is the arid metaphysics of the Schools), Berkeley does offer us what we would naturally describe as a metaphysics, and one that cannot be refuted simply on the ground that it might strike the average person as outrageous. His arguments must be examined on their merits, together with any underlying assumptions; attention has to be paid to the notion of 'immediate' perception which he works with; and account must be taken of possible problems generated by his metaphysical conclusions. These may include, as has often been claimed, an unrecognized tendency towards solipsism. Not that it is necessary to reject or accept his philosophy in total, for there may be insights alongside what we believe are mistakes. As with any philosopher of Berkeley's stature, doing justice to Berkeley's philosophy turns out to be a very complex, but also a rewarding exercise, which is why his philosophy still exercises the commentators today.

See also: Meaning and understanding; Vision

IAN TIPTON

List of works

Berkeley, G. (1948-57) *The Works of George Berkeley, Bishop of Cloyne*, ed. A.A. Luce and T.E. Jessop, Edinburgh: Thomas Nelson, 9 vols.(The standard edition, containing Berkeley's published and unpublished writings, both philosophical and non-philosophical. The philosophical correspondence between Berkeley and Samuel Johnson is in vol. 2.)

Berkeley, G. (1707) *Arithmetica* and *Miscellanea Mathematica*, in *Works*, vol. 4, Edinburgh: Thomas Nelson. (Published anonymously in one volume, these slight pieces are of relatively little interest today, even to mathematicians.)

Berkeley, G. (1707) *Of Infinites*, in *Works*, vol. 4, Edinburgh: Thomas Nelson.(Although Berkeley never published it, this essay is of some interest, both because of its anticipation of attacks on assumptions underlying the infinitesimal calculus that Berkeley developed in later works and for the use he makes of the Lockian principle - which he was to soon reject - that all words stand for ideas.)

Berkeley, G. (1707-8) *Philosophical Commentaries*, in *Works*, vol. 1, Edinburgh: Thomas Nelson.(Berkeley's notebooks, filled as he prepared to publish the *New Theory of Vision* and the *Principles*. They are of inestimable value to scholars. The dates given represent Luce's reasoned speculation; the title is that given to them by Luce.)

Berkeley, G. (1709) *An Essay towards a New Theory of Vision*, in *Works*, vol. 1, Edinburgh: Thomas Nelson.(The first of Berkeley's major works, and for long influential as a work on the psychology underlying our visual perception of the distance from us and the size and situation of objects. Although Berkeley's ultimate metaphysical commitments were not made manifest here, awaiting the publication of the *Principles* in the following year, this work prepared the ground for them.)

Berkeley, G. (1710) *A Treatise concerning the Principles of Human Knowledge*, in *Works*, vol. 2, Edinburgh: Thomas Nelson.(Undoubtedly Berkeley's most important work, in which he attacks 'material substance' and the notion that sensible objects exist 'without the mind'. Originally published as 'Part I', no further parts appeared.)

Berkeley, G. (1712) *Passive Obedience*, in *Works*, vol. 6, Edinburgh: Thomas Nelson.(An interesting little work in that it represents a rare excursion into political philosophy and ethics, and suggests a basically utilitarian position. Berkeley denies the legitimacy of resisting even a manifestly unjust ruler.)

Berkeley, G. (1713) *Three Dialogues between Hylas and Philonous*, in *Works*, vol. 2, Edinburgh: Thomas Nelson. (A lively exposition of the metaphysics argued for in the *Principles of Human Knowledge*, with an emphasis on its supposed conformity with common sense. Highly recommended to those reading Berkeley for the first time.)

Berkeley, G. (1713) Essays in the *Guardian*, in *Works*, vol. 7, Edinburgh: Thomas Nelson. (Short essays published in Richard Steele's short-lived periodical, evidencing Berkeley's opposition to any form of irreligion. Essays in the *Guardian* were published anonymously. Those included in *Works*, vol. 7, are the ones Luce judged were by Berkeley.)

Berkeley, G. (1721) *De Motu*, in *Works*, vol. 4, Edinburgh: Thomas Nelson.(Berkeley abandoned his plans to publish further parts of the *Principles*, including one on natural philosophy. This essay on the cause of motion covers some of the ground that might have been dealt with there.)

Berkeley, G. (1732) *Alciphron: or the Minute Philosopher*, in *Works*, vol. 3, Edinburgh: Thomas Nelson.(A defence of religion, and Christianity in particular, against 'free-thinkers'. Observations on the uses of language contained in the last of the seven dialogues into which the work is divided give insights into how Berkeley's thinking in this area developed.)

Berkeley, G. (1733) *The Theory of Vision, Vindicated and Explained*, in *Works*, vol. 1, Edinburgh: Thomas Nelson.(A third edition of the *New Theory of Vision* had been annexed to the first edition of *Alciphron*. In *The Theory of Vision, Vindicated and Explained*, Berkeley responded to an anonymous critic.)

Berkeley, G. (1734) *The Analyst*, in *Works*, vol. 4, Edinburgh: Thomas Nelson.(A technical work on mathematics, revealing alleged obscurities and contradictions lying at the root of the infinitesimal calculus. The work, which was addressed to 'an infidel mathematician', argued that mathematicians were poorly placed to object to obscurities in Christian doctrine.)

Berkeley, G. (1735) *A Defence of Free-thinking in Mathematics* and *Reasons for not replying to Mr. Walton's Full Answer*, in *Works*, vol. 4, Edinburgh: Thomas Nelson. (Two tracts, published in the same year, responding to two critics of *The Analyst*.)

Berkeley, G. (1735-7) *The Querist*, in *Works*, vol. 6, Edinburgh: Thomas Nelson.(A series of short rhetorical questions, motivated by Berkeley's deep concern with the economic state of Ireland and the poverty of the Irish people. By now little more than a curiosity, the work was deservedly popular at the time.)

Berkeley, G. (1744) *Siris: a Chain of Philosophical Reflexions and Inquiries* (etc.), in *Works*, vol. 5, Edinburgh: Thomas Nelson.(A work in which Berkeley moves from a consideration of the supposed virtues of tar-water as a medicine, through reflections on natural science, to observations on God and the Trinity. Although popular at the time, this work is of little interest now, except to specialist scholars.)

References and further reading

Atherton, M. (1990) *Berkeley's Revolution in Vision*, Ithaca, NY: Cornell University Press.(Argues that, far from being a mere halfway house on the route to the philosophy of the *Principles* and *Dialogues*, Berkeley's *New Theory of Vision* has value in its own right, and provides a key to a more adequate understanding of those works.)

Baxter, A. (1733), *An Enquiry into the Nature of the Human Soul*, London: G. Strahan; 2nd edn, London: A. Millar, 1737; repr. Bristol: Thoemmes Press, 1990. (The section in volume two entitled 'Dean Berkeley's scheme against the existence of matter, and a material world examined, and shewn inconclusive' contains the first extended critique of Berkeley's *Principles*.)

Berman, D. ed. (1989) *George Berkeley: Eighteenth-Century Responses*, 2 vols, New York: Garland Publishing. (Contains many of the early responses to various of Berkeley's works. These include some from important or influential figures such as James Beattie and Thomas Reid, the first reviews of the *Principles* and *Dialogues*, and materials not easily accessible elsewhere.)

Berman, D. (1994) *George Berkeley: Idealism and the Man*, Oxford: Clarendon Press.(A lively introduction to Berkeley's life and writings which is particularly interesting on works such as *Passive Obedience*, *Alciphron*, and *Siris*, and which throws light on the Irish context of Berkeley's thought.)

Brook, R.J. (1973) *Berkeley's Philosophy of Science*, The Hague: Martinus Nijhoff. (A thorough examination of Berkeley's writings on optics, physics and mathematics, against the background of his theory of meaning and signification.)

Grayling, A.C. (1986) *Berkeley: The Central Arguments*, London: Duckworth. (Argues that 'Berkeley's views are, in some important respects, more defensible than has in general been allowed', and - against Pitcher, Tipton and others - that Berkeley's 'theistic' realism can be plausibly presented as a defence of common sense.)

Johnston, G.A. (1923) *The Development of Berkeley's Philosophy*, London: Macmillan; repr. New York: Garland Publishing, 1988.(Originally published early in the twentieth century, Johnston's judicious and scholarly introduction to the Berkeleian corpus contains much that is still of value.)

Locke, J. (1689) *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1974. (Listed here because Locke has been mentioned frequently in the article, but also because Locke was undoubtedly the most important influence on Berkeley.)

Luce, A.A. (1934) *Berkeley and Malebranche*, Oxford: Oxford University Press, 1934; 2nd edn, Oxford: Clarendon Press, 1967.(While not underestimating the importance of Locke's influence on Berkeley, Luce's

important monograph showed that Berkeley's reading of Malebranche's *Search After Truth* made a very deep impression on him. Luce also argued for the likely influence of Bayle.)

Luce, A.A. (1949) *The Life of George Berkeley, Bishop of Cloyne*, Edinburgh: Thomas Nelson; repr. Bristol: Thoemmes Press, 1992. (The standard biography of Berkeley.)

Luce, A.A. (1963) *The Dialectic of Immaterialism*, London: Hodder & Stoughton.(A study of the *Philosophical Commentaries* published when Luce was in his eighties, this book rested on more than thirty years of devoted research.)

Mill, J.S. (1871) 'Berkeley's Life and Writings', *Fortnightly Review*, n.s. X (Nov.) repr. in *Collected Works of John Start Mill*, ed. J.M. Robson, Toronto: University of Toronto Press and London: Routledge & Kegan Paul, 1978, vol. 11.(Fascinating observations on Berkeley's philosophy from an important nineteenth-century philosopher.)

Muehlmann, R.G. (1992) *Berkeley's Ontology*, Indianapolis, IN: Hackett Publishing Company.(A fairly demanding but rewarding study, revealing an in-depth knowledge of the recent literature on Berkeley, as well as a sensitivity towards the intellectual background against which Berkeley wrote.)

Pitcher, G. (1977) *Berkeley*, London: Routledge & Kegan Paul.(Although he tends to write as if Berkeley was reacting solely against Locke, Pitcher's philosophical acuity and clarity of style make this a stimulating introduction to Berkeley.)

Tipton, I.C. (1974) *Berkeley: The Philosophy of Immaterialism*, London: Methuen; repr. Bristol: Thoemmes Press, 1994.(A careful examination of Berkeley's arguments in the *Principles* and *Dialogues*, this book aims to be of value to scholars and other philosophers as well as to those requiring a comprehensible introduction to Berkeley.)

Winkler, K.P. (1989) *Berkeley: An Interpretation*, Oxford: Clarendon Press.(This sympathetic exposition of Berkeley's metaphysics will appeal to the more advanced student. Concentrates on topics such as representation, abstraction, and cause and effect, in an attempt to deepen our understanding of Berkeley's central arguments.)

Berlin, Isaiah (1909-97)

Berlin said that he decided about 1945 to give up philosophy, in which he had worked up to that time, in favour of the history of ideas. Some of his best-known work certainly belongs to the history of ideas, but he continued in fact both to write philosophy and to pursue philosophical questions in his historical work.

His main philosophical contributions are to political philosophy and specifically to the theory of liberalism. He emphasizes a distinction between 'negative' and 'positive' concepts of liberty: the former is a Hobbesian idea of absence of constraint or obstacle, while the latter is identified with a notion of moral self-government, expressed for instance in Rousseau, which Berlin finds politically threatening. His anti-utopian approach to politics is expressed also in his view that values necessarily conflict; this irreducible 'value pluralism' may be his most original contribution to philosophy, though advances it through example and historical illustration rather than in semantic or epistemological terms. He also expresses himself against necessitarian interpretations of history, and in favour of an anti-determinist conception of free will.

1 Against reductionism

Isaiah Berlin was born in Riga, Latvia, of Jewish parents; the family emigrated in 1919 to Britain, where he was educated and spent his life. He worked first in general philosophy, and in the 1930s took part, with A.J. Ayer, J.L. Austin and others, in discussions of questions about knowledge and meaning which were raised by the logical positivist agenda of the time. (Papers related to these interests, mostly published rather later, are collected in Berlin 1978a.) Berlin was never tempted by positivism. Besides pressing realist objections against such proposals as the dispositional analysis of statements about the unobserved, he rejected its scientific paradigm of knowledge and its lack of interest in historical understanding. Paradoxically, the most positivist element to survive in Berlin's thought is his definition of philosophy, as concerned with questions that defy solution by the a priori and the empirical sciences. It may also have been a positivist influence that encouraged him to draw a sharp distinction, in his formulations though not in his practice, between his philosophical and his historical interests (see Logical positivism).

In his work in the history of ideas, Berlin paid particular and sympathetic attention to thinkers such as Vico and Herder who have emphasized cultural difference across time, have stressed the need to understand other ways of life 'from the inside', and have resisted the impulse of the more rationalistic strains of the Enlightenment to reduce the range of human concerns to some limited set of motivations which are met more or less efficiently by different social formations. He denied that there is a 'fixed and unalterable' human nature. At the same time, however, he firmly resisted relativism of any type.

At the level of interpreting other societies, he accepts that there are universal human potentialities, limitations, and indeed sentiments, and believes in fact that we would not be able to understand those societies except on this assumption. The denial of a fixed human nature comes to saying first, that there are many different and no canonically correct expressions of these potentialities; second, that what expressions these potentialities might receive cannot be recognized in advance of historical experience; third, that these two points hold good for the future, so that there can be no Hegelian (or - more particularly for Berlin's concerns - Marxist) total realization of human possibilities (see Hegel, G.W.F.; Marx, K.).

How far, in Berlin's view, we might go in forming a definite picture of human potentialities is unclear. On the one hand, Berlin, sharing with his favoured authors a keen sense of cultural particularity, is impressed by the unpredictable distinctiveness of different forms of life, as of artistic styles. On the other hand, we understand these forms of life, in part, through values that are expressed in them, and there is nothing in Berlin's work to rule out the idea that, although the cultural forms are manifold, the values expressed in them might be limited, and indeed quite few, in number. If so, there may be room for an account of human nature which would explain why, at a general level, only a certain range of values are candidates for expression in recognizably human cultures (see Human nature).

2 Value pluralism

These values, Berlin repeatedly urged, make conflicting claims and cannot be totally reconciled with one another

without loss. Political schemes, moral theories, and religious aspirations have repeatedly tried to deny this truth and to claim that, properly understood, values do not ultimately conflict: ‘true’ liberty, for instance, will not conflict with ‘true’ equality. Berlin rejects such outlooks, for several reasons. Politically, attempts to put them into practice have always been a disaster, in terms which only their most fanatical adherents can deny. Ethically, they are an evasion, and pretend that an intellectual construction can make life easier than it is. Philosophically, they are a mistake.

Not being concerned with meta-ethical analysis, Berlin does not try to counter the objection made by some critics that the status of this last claim is obscure, but concentrates on substantiating the first two claims. Berlin himself, because of his views on the nature of philosophy, is disposed to agree that the philosophical claim is distinct from political and ethical claims, but the best interpretation of his outlook may well be that the philosophical claim is to be understood through the others. Whatever theory of values we accept, our conclusions about their structure can be sensibly constrained, protected against being an arbitrary fantasy, only by serious reflection on political and ethical experience, and for Berlin this essentially depends on our best historical understanding.

Berlin’s claim of ultimate value pluralism encounters philosophical problems about its content, as well as about its status. In practice, policies do have to be adopted, some ways of life favoured over others. Berlin’s pluralism insists that such choices often involve value loss, denies that there is one currency in which the gains and losses can be calculated, but claims that the choices are not therefore irrational. Questions of how these views can be consistently interpreted are important to the theory of practical and evaluative reasoning: they are sometimes expressed as questions about the incommensurability of values.

Questions about the status and the content of value pluralism are important to Berlin’s defence of liberalism. He clearly regards liberalism as a political order which in some sense particularly respects value pluralism. It would be untrue to Berlin’s outlook to take this in a metaphysical sense, as implying that the liberal state best expresses the real structure of values. Rather, the arguments for liberalism use the same historical and political materials as the arguments for pluralism. To this extent, Berlin’s liberalism, with its emphasis on the human costs of non-liberal, ‘totalizing’, political visions, is close to what Judith Shklar (1989) called ‘the liberalism of fear’, an outlook which owes much to Montesquieu, Constant, De Tocqueville and a strong sense of modern historical experience.

Another line of interpretation suggests that, for Berlin, value pluralism is specially honoured by liberalism because liberal states emphasize the individual’s freedom to choose, in particular between different values and forms of life. This makes such freedom - freedom as autonomy - a privileged value in Berlin’s outlook, and associates him rather with von Humboldt and J.S. Mill. But it is doubtful whether autonomy is privileged by Berlin in this way. It would be barely consistent with his own treatment of liberty (see §3), and it is questionable whether value pluralism can be coherently used to promote one value over others. In any case, Berlin is notably ambivalent about the more defiantly individualist aspect of liberal societies. Above all because of his Zionist sympathies, he is at least as much interested in the individual’s ‘need to belong’ to a community, and in cultural rather than individual self-determination. To some extent, these sympathies stand in tension with his liberalism, but they also condition it, in such a way that radical autonomy and individual self-expression should not be seen as its most central values (see Moral pluralism).

3 Liberty

Berlin discussed more than one question related to liberty: he opposed ideas of historical inevitability, and resisted compatibilist theories of free will. He is best known, however, for his views on liberty as a distinctively political value, in particular for a distinction between ‘negative’ and ‘positive’ conceptions of liberty. Negative liberty is defined in traditionally Hobbesian or empiricist terms, as the absence of external constraint or interference. Positive liberty is understood, rather, as self-mastery or self-realization, two aspects which are themselves not very strongly distinguished from one another.

It is important that this is not only a distinction between two political conceptions, but a political distinction, which gets its content from its relation to two different styles of political thought. It is not, for instance, a distinction between ‘freedom from’ and ‘freedom to’. Though negative liberty centrally covers freedom from intentional interference, it does not cover the famous freedoms of post-war liberal politics, such as the freedoms from want,

disease and unemployment. Berlin is not opposed to the objectives that are expressed in this way, but in the general spirit of pluralism finds it more helpful to place them under values other than freedom, such as welfare or equality. Again, negative liberty does not cover freedom from internal compulsions. This freedom might be said in any case to represent an ethical or psychological aspiration, rather than a political conception. However, Berlin does associate it with positive liberty, and this precisely illustrates the point that on his view positive liberty is a political conception that aspires to have too much ethical content (it is, in John Rawls' terminology, 'perfectionist' (Rawls 1971: 25)). Negative liberty can extend to freedom from political manipulation, but seemingly only in gross forms which would be described even by everyday conceptions as restricting agents' freedom to act. It would not extend to freedom from ideological determination or false consciousness, which again would be associated with positive liberty.

The thread that holds together these various associations and exclusions seems to be that negative liberty is meant to be minimally normative in its definition. It is, of course, a value and brings with it a norm or positive evaluation, but in giving an account of its content, Berlin tries as far as possible to exclude other normative and evaluative ideas. This same point comes out in Berlin's treatment of the traditional difficulty that if liberty is measured by the ratio of actual desires to the capacity to fulfil them, it will follow that one can increase liberty just as much by reducing agents' desires as by increasing their capacities. Berlin answers this simply by appealing to possible as well as actual desires. This line involves the difficulty that if no limits are put on what an agent could possibly desire, everyone will be equally unfree; it seems that Berlin would rather face this problem than appeal to measures of liberty that bring in further normative elements, such as the idea of desires that agents might reasonably be expected to have, or of those that agents in society might reasonably expect to satisfy.

Berlin's treatment of liberty seems have three different sources. One is his pluralistic aim of keeping liberty as far as possible distinct from other political values. Another is that he is suspicious of accounts of political values which tie them too closely to non-political values, in particular to perfectionist ethical aspirations for the individual. Third, and relatedly, he is impressed by the fact that political programmes which advance such values have been historically associated with the actions of an elite trying to impose an interpretation of people's supposed real interests. Berlin's favoured negative concept of liberty, like other aspects of his thought, is not a purely analytical proposal, but is defined by conceptions of liberalism and of the political which are consciously shaped by responses to actual historical developments, above all the Russian Revolution (see Liberalism).

See also: Enlightenment, Continental; Freedom and liberty; Incommensurability; Historicism; Pluralism; Values

BERNARD WILLIAMS

List of works

Berlin, I. (1939) *Karl Marx: His Life and Environment*, London: Thornton Butterworth; repr. with a new introduction by A. Ryan, London: Fontana, 1955.(A vivid representation of Marx as well as of his ideas; some modifications in later editions.)

Berlin, I. (1969) *Four Essays on Liberty*, Oxford: Oxford University Press.(A central text. In his long introduction Berlin discusses the two central essays, 'Historical Inevitability' and 'Two Concepts of Liberty'.)

Berlin, I. (1976) *Vico and Herder*, London: Hogarth Press.(Basic to the work on the history of ideas.)

Berlin, I. (1978a) *Concepts and Categories*, ed. H. Hardy, London: Hogarth Press.(Contains some early contributions to general philosophy; and, among other essays, 'From Hope and Fear Set Free', which complements the essays on liberty, and 'Does Political Theory Still Exist?'.)

Berlin, I. (1978b) *Russian Thinkers*, ed. H. Hardy and A. Kelly, London: Hogarth Press.(Includes 'The Hedgehog and the Fox', 'On Tolstoy' and 'Fathers and Children', important for the understanding of liberalism.)

Berlin, I. (1979) *Against the Current*, ed. H. Hardy, London: Hogarth Press.(Essays in the history of ideas. Contains a complete bibliography of Berlin's writings, updated in a later edition 1991, Oxford: Clarendon Press, and further in 1997, London: Pimlico Press.)

Berlin, I. (1980) *Personal Impressions*, ed. H. Hardy, London: Hogarth Press.(Vivid and often moving biographical essays.)

Berlin, I. (1990) *The Crooked Timber of Humanity*, ed. H. Hardy, London: Murray.(Starts with Berlin's retrospective summary of his views, and includes a long essay on de Maistre.)

Berlin, I. (1993) *The Magus of the North: J.G. Hamann and the Origins of Modern Irrationalism*, ed. H. Hardy, London: Murray.(Short monograph on the obscure Prussian thinker regarded by Berlin as one of the fathers of

romanticism.)
Berlin, I. (1996) *The Sense of Reality: Studies in Ideas and their History*, ed. H. Hardy, London: Chatto & Windus.(Mostly historical essays written in the 1950s and 1960s, with one exception published for the first time.)
Berlin, I. (1997) *The Proper Study of Mankind*, ed. H. Hardy and R. Hausheer, London: Chatto & Windus.(An anthology of Berlin's best-known essays.)

References and further reading

Galipeau, C.J. (1994) *Isaiah Berlin's Liberalism*, Oxford: Oxford University Press.(Careful and well-documented discussion.)
Gray, J. (1995) *Isaiah Berlin*, London: HarperCollins.(Interpretation in terms of 'agonistic liberalism', emphasizing radical choice.)
Margalit, E. and Margalit, A. (eds) (1991) *Isaiah Berlin, A Celebration*, London: Hogarth Press.(Essays by various writers, about half of them on philosophical subjects.)
Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(A classic modern statement of liberal political philosophy.)
Ryan, A. (ed.) (1979) *The Idea of Freedom: Essays in Honour of Isaiah Berlin*, Oxford: Oxford University Press.(Essays by various writers, most on philosophical topics.)
Shklar, J. (1989) 'The Liberalism of Fear', in N.L. Rosenblum (ed.) *Liberalism and the Moral Life*, Cambridge, MA: Harvard University Press.(See §2.)

Bernard of Clairvaux (1090-1153)

Bernard was recognized by his contemporaries as the spiritual leader of western Europe. He was an indefatigable advocate of the monastic life and occasionally criticized the schools on moral grounds, but he was by no means an anti-intellectual. He encouraged a number of early scholastic philosopher-theologians in their work. Although he devoted the better part of his efforts to his wide-ranging pastoral duties, Bernard's own sermons and treatises make a significant contribution to twelfth-century theology and philosophy.

Bernard was born into a noble family in 1090 near Dijon in Burgundy, and studied grammar, rhetoric and logic at Châtillon. Abandoning plans for further study, he joined the Cistercian Order at Cîteaux in 1113. Only two years later, he became the first abbot of the new Cistercian foundation at Clairvaux, a position he held until his death in 1153. During his abbacy of almost forty years, Bernard exerted a profound influence over the intellectual and religious life of western Europe. The extent of his influence may be measured by the sixty-five daughter monasteries established by Clairvaux during Bernard's lifetime, his nearly single-handed resolution of the papal schism of 1130, and the fact that one of his own monks assumed the papacy as Eugenius III (1145-53).

Bernard was actively interested in the schools of his day (the precursors of the medieval universities) and was reasonably well-informed about academic discussions in theology. In his controversies, first with Peter Abelard in 1140 and then with Gilbert of Poitiers in 1148, Bernard wished to expose and correct what he saw as heretical opinions. Concerning atonement, for example, Bernard opposed Abelard's claim that Christ redeemed humanity simply by inspiring love for God and providing a perfect pattern of life. Regarding the trinity, Bernard maintained, also against Abelard, that the relation of the second person to the first was not a relation of species to genus, and that certain divine properties, such as power or wisdom, were not specific to just one or another person of the trinity (see Trinity). In the case of Gilbert, Bernard also disputed some of his assertions about the trinity, the incarnation and the divine essence.

In attacking certain of Abelard's and Gilbert's conclusions, Bernard also showed a distrust of their method, but he did not condemn outright their application of logic to doctrine. Bernard considered all knowledge good in itself, but he believed the pursuit of knowledge should be governed by practical concerns, especially salvation, and not by a desire to know simply for the sake of knowing.

Bernard's feuds with Abelard and Gilbert were more than counterbalanced by his support for other philosopher-theologians. He was, for example, a protector of Peter Lombard and a friend to John of Salisbury. He praised the English scholastic Robert Pullen for his 'sound doctrine' and endeavoured to keep him in Paris. Bernard responded to some inquiries of Hugh of St Victor with a long letter (*circa* 1127) on baptism, salvation before Christ's death and resurrection, and other disputed matters, which clearly influenced Hugh's massive work, *De sacramentis christianae fidei (On the Sacraments of the Christian Faith)*. More than once, Bernard admitted the practical value, even necessity, of scholastic theology in refuting heresy, clarifying obscure points of doctrine and guiding church leaders.

Bernard was an accomplished theologian in his own right, and his writings are characterized by strict adherence to scriptural and patristic sources rather than by philosophical development of doctrine. Most of his sermons and treatises are concerned with the exposition of Scripture and doctrine along moral and contemplative lines. *De gradibus humilitatis et superbiae (Steps of Humility and Pride)*, for example, describes the pursuit of truth through three kinds of knowledge: humility (self-knowledge), compassion (knowledge of one's neighbour) and contemplation (knowledge of God). *De diligendo Deo (On Loving God)* takes up the nature of disinterested love and describes the steps by which one attains it. His longest work, *Sermones super Cantica canticorum (Sermons on the Song of Songs)*, begun in 1135 and left unfinished at his death in 1153, combines scriptural commentary with an examination of the moral and contemplative aspects of religious life.

Bernard sometimes wrote works of theology that were more philosophical than scriptural. His treatise *De gratia et libero arbitrio (Grace and Free Choice)* is a substantial work of philosophical theology, markedly influenced by Augustine, in which he attempts to reconcile human freedom with God's grace. Bernard thinks that the will is necessarily free; in consenting to the good, the will cooperates with grace. His treatise is notable for its definition

of free choice as 'a spontaneous inclination of the will,' and for its innovative distinction among three types of human freedom: freedom from necessity, freedom from sin and freedom from sorrow (see Free will; Grace). This was the most influential of Bernard's works among scholastic thinkers, and elements of it appear in Peter Lombard's *Sentences* and, in the next century, in the works of Alexander of Hales and Albert the Great.

See also: Abelard, P.; Hugh of St Victor; Religion, history of philosophy of

SEAN EISEN MURPHY

List of works

Bernard of Clairvaux (1090-1153) *Sancti Bernardi Opera*, ed. J. Leclerq *et al.*, Rome: Editiones Cisterciensis, 1957-77, 8 vols; trans. as *The Works of Bernard of Clairvaux*, Cistercian Fathers Series, Kalamazoo, MI: Cistercian Publications, 1973-80.(The English translation series includes most of Bernard's treatises and many of his sermons.)

Bernard of Clairvaux (*c.*1125) *De gradibus humilitatis et superbiae (Steps of Humility and Pride)*, in J. Leclerq *et al.* (eds) *Sancti Bernardi Opera*, Rome: Editiones Cisterciensis, vol. 3, 1963; trans. M. Ambrose Conway, *The Steps of Humility and Pride*, Cistercian Fathers Series, Kalamazoo, MI: Cistercian Publications, 1973; repr. 1989.(A profound study of the nature of humility - as reflected in twelve manifestations of pride - and its importance in the attainment of truth; an important source for later medieval ethics.)

Bernard of Clairvaux (1126-41) *De diligendo Deo (On Loving God)*, in J. Leclerq *et al.* (eds) *Sancti Bernardi Opera*, Rome: Editiones Cisterciensis, vol. 3, 1963; trans. E. Stiegman, Cistercian Fathers Series, Kalamazoo, MI: Cistercian Publications, 1973; repr. 1985.(The best introduction to several overarching themes in Bernard's life and thought.)

Bernard of Clairvaux (*c.*1127) *De gratia et libero arbitrio (Grace and Free Choice)*, in J. Leclerq *et al.* (eds) *Sancti Bernardi Opera*, Rome: Editiones Cisterciensis, vol. 3, 1963; trans. D. O'Donovan, Cistercian Fathers Series, Kalamazoo, MI: Cistercian Publications, 1977; repr. 1988.(Takes a moderate, sometimes novel approach to one of the most complex and divisive issues in Christian thought.)

Bernard of Clairvaux (1135-53) *Sermones super Cantica canticorum (Sermons on the Song of Songs)*, in J. Leclerq *et al.* (eds) *Sancti Bernardi Opera*, Rome: Editiones Cisterciensis, vols 1-2, 1957-8; trans. K. Walsh and I.M. Edmonds, Cistercian Fathers Series, Kalamazoo, MI: Cistercian Publications, 1971-80, 4 vols.(A model of allegorical exegesis, this detailed commentary should be read more for its sustained analysis of Christian life and doctrine than for its nominal subject matter.)

References and further reading

Saint Bernard Théologien: Actes Du Congrès De Dijon, 15-19 Septembre 1953 (Saint Bernard the Theologian: Proceedings of the Dijon Conference, 15-19 September 1953) (1953), 2nd edn in Analecta Sacri Ordinis Cisterciensis IX, foreword and introduction by J. Leclercq, Rome: Editiones Cistercienses.(An unsurpassed collection of essays, all in French, on various aspects of Bernard's theology.)

Evans, G.R. (1983) *The Mind of St. Bernard of Clairvaux*, Oxford: Clarendon Press. (A very helpful introduction to Bernard's writings and his place in the intellectual life of the twelfth century, but sometimes weak in its exposition of his ideas.)

Gilson, E. (1940) *The Mystical Theology of St. Bernard*, trans. A.H.C. Downes, London and New York: Sheed & Ward; repr. Kalamazoo, MI: Cistercian Publications, 1990.(First published in 1934, this still indispensable classic reconstructs the theological system underlying Bernard's contemplative writings.)

Bernard of Tours (*fl.* 1147, d. before 1178)

Bernard of Tours, better known as Bernardus Silvestris, was closely acquainted with the major developments in science and theology which took place in the mid-twelfth century. His major work, the Cosmographia, an allegorical account of the creation of the universe and humankind, is dedicated to the philosopher-theologian Thierry of Chartres, who was probably also his teacher. However, Bernard himself was best known as a poet, and he seems to have made his living primarily as a teacher of grammar and rhetoric. His career perhaps reflects the fragmentation of the liberal arts curriculum in his day, including the segregation of literary studies from the increasingly specialized pursuit of the sciences.

Little is known of Bernard's career beyond his attachment to Tours. The city is recalled affectionately in the *Cosmographia*; Bernard is praised by his pupil Matthew of Vendôme, author of one of the earliest medieval treatises on the art of poetry, as 'the master of Tours', and we know that a nephew inherited his house in the city at some time prior to 1178. The *Cosmographia* also contains a fulsome compliment to Pope Eugenius III, and a manuscript gloss reports that the work was read before Eugenius, to the latter's great satisfaction. This was presumably in 1147, when Eugenius came to France to attend the trial of Gilbert of Poitiers.

No more is known of Bernard's life. Matthew notes his mastery of *dictamen*, the arts of composition, and later writers cite his poems almost exclusively as examples of literary style, so we may assume that this was the main focus of his teaching. Commentaries on Vergil and Martianus Capella, which use mythographic analysis to raise philosophical and religious questions and often closely echo the commentaries of William of Conches, have been attributed to Bernard, but despite suggestive thematic correspondences with the *Cosmographia*, they cannot be assigned to him with certainty.

The *Cosmographia* consists of alternating chapters of prose and verse, in emulation of Boethius' *De consolatione philosophiae* and the *De nuptiis Philologiae et Mercurii* of Martianus Capella, and its narrative of the cosmogony is based broadly on Plato's *Timaeus* (Wetherbee 1972: 152-86) (see Boethius, A.M.S.; Encyclopedists; Plato). Its theme is the process of creation, the realization of the 'seminal virtues' that come to fruition through the union of matter and form. The work begins with the noble and impassioned appeal of Nature to Noys, the agent of divine Wisdom, to order the formless existence of primal matter, which longs for 'the shaping influence of number and bonds of harmony'. The flowering of created life which results has important implications for man, the lesser universe, whose creation will be the culmination of the adaptation of Nature's vitality to the shaping influence of divine Wisdom. Man will find his fulfilment in recognition of the completeness with which his nature corresponds to the pattern and activity of the macrocosm. For man to comprehend this great affinity will be to simultaneously recognize his own lordship and destiny, a realization that mirrors the ordering activity of Noys, confirming that man's mind is created in the image of the Divine Wisdom.

Bernard's vision of the human condition is darkened by a strong sense of the precariousness of material existence and the uncertain power of reason to govern the aberrant tendencies of human nature. Like the visions vouchsafed to Adam in the last books of Milton's *Paradise Lost*, Bernard's foreshadowings of world history offer clear evidence of human guilt and folly. But they also celebrate the fruits of human genius in art, science and technology. The final vision of the *Cosmographia* is heroic, the stoic dignity with which man labours to resist his own instability and order and perpetuate his existence in the face of seemingly insuperable necessity. The work conveys the anxiety as well as the excitement of an age newly engaged with the study of nature and the tenuous linkage of cosmic order and physical law.

At times the pessimistic strain in the *Cosmographia* becomes a kind of determinism, and a similar note is struck in Bernard's other known works. The *Mathematicus (The Astrologer)*, a narrative in Latin elegiacs, tells of a young man confronted by a prophecy that he will slay his father, who resolves instead to take his own life, affirming that such a death will be a release from the prison of material existence (Dronke 1974: 131-41). Bernard's introduction to the *Experimentarius*, a Latin adaptation of an Arab manual of geomancy, affirms the power of God but acknowledges that the planets have a 'natural' power to affect 'the destinies of all mortal things'.

See also: Boethius, A.M.S.; Chartres, School of; Cosmology; Natural philosophy, medieval; Platonism, medieval

WINTHROP WETHERBEE

List of works

Bernard of Tours (*c.*1147) *Cosmographia*, ed. P. Dronke, Textus Minores 53, Leiden: Brill, 1978.(This work has been published in an edition translated by W. Wetherbee, Records of Civilization: Sources and Studies 89, New York: Columbia University Press, 1973.)

Bernard of Tours (*c.*1147) *Mathematicus (The Astrologer)*, ed. B. Hauréau, Paris: C. Klincksieck, 1895. (This work also appears in *Patrologia Latina* 171, cols. 1365-80, where it is attributed to Hildebert of Lavardin.)

Bernard of Tours (*c.*1147) *Experimentarius*, ed. M. Brina-Savorelli, *Rivista critica di storia della filosofia* 14, 1959: 283-342.(Latin adaptation of Arab work with introduction by Bernard.)

References and further reading

Dronke, P. (1974) *Fabula: Explorations into the Uses of Myth in Medieval Platonism*, Mittellateinische Studien und Texte 9, Leiden: Brill.(Places Bernard in a tradition of philosophical and religious interpretation of ancient myth.)

Edwards, R.R. (1993) 'Poetic Invention and the Medieval *Causae'*, *Mediaeval Studies* 55: 183-217.(On the *Mathematicus* and other Latin poems attributed to Bernard.)

Stock, B. (1972) *Myth and Science in the Twelfth Century: A Study of Bernard Silvester*, Princeton, NJ: Princeton University Press.(A detailed reading of the *Cosmographia* as a document in the history of ideas.)

Wetherbee, W. (1972) *Platonism and Poetry in the Twelfth Century*, Princeton, NJ: Princeton University Press. (Considers the *Cosmographia* as synthesizing the concerns of earlier twelfth-century philosophy.)

Bernier, François (1620-88)

Bernier was a minor figure who influenced the history of philosophy out of all proportion to his own strictly philosophical abilities. He was effective as a propagandist in the debates over the analysis of matter, and especially as a popularizer of the views of Pierre Gassendi, whose nominalism he sought to apply with greater consistency.

Bernier was and remains the best-known disciple of Gassendi, whose eyes he is supposed to have closed at his death. In the seventeenth and for at least the next two centuries, Bernier was rather less known as a philosopher than as a traveller, especially for his *Memoirs… of the Grand Mogul's Empire* (1670-1), a classic that detailed his ten-year visit to the Indian subcontinent. Dryden based his *Aureng-Zebe* on this work, and Locke, a physician like Bernier, was interested in it, particularly as a source for his investigation of religious psychology under the rubric of 'enthusiasm'.

In philosophy Bernier was a minor figure who none the less played an important role as a polemicist, as a popularizer and as a philosopher in his own right. As a polemicist, he wrote at greatest length in *Anatomia* (1651) and *Favilla* (1653) on Gassendi's behalf against the imprecations of the astrologist J.-B. Morin. More effectively, he turned his satirical pen in common cause with the Cartesians against the political machinations of the Jesuit Aristotelians in his *Requeste des maistres* (1671). Finally, he joined the debates over matter and transubstantiation and, as Pierre Bayle (who published his work) commented on *Eclaircissement* (1684), strongly opposed the Cartesians in order better to make his peace with the Jesuit La Ville.

By far the most influential of Bernier's works was his *Abrégé* of Gassendi's philosophy (1674-8). This abridgement, consisting of paraphrase and some translation, and written in straightforward, non-technical French, became for many their only access to the views expressed in the obscure, late scholastic Latin of Gassendi's lengthy tomes. Bayle's hope that it would remove the need to work through those tomes proved idle, however, for Bernier was an original thinker whose own views influenced the choice, ordering, emphasis and evaluation of the material.

Bernier's independence of Gassendi emerged explicitly in his *Doutes de M. Bernier sur quelques uns des principaux chapitres de son Abrégé de Gassendi* (1682). His eleven doubts raise questions about Gassendi's views on space, time, motion, collision and other important topics. However, in the preface to the second edition published with the second and last edition of the *Abrégé*, Bernier says that he 'doubts no longer, having despaired of being able ever to understand any of [these questions]'.

Beyond this negative or academic scepticism, however, the Doutes contains a more consistent application of Gassendi's own empiricism and nominalism than anything found in Gassendi himself (see Nominalism). A good example is the treatment of space, which for Gassendi was an uncreated, incorporeal tridimensionality, a third kind of being that was neither substance nor accident. In addition to the ubiquitous theological objection that such a space would be independent even of God, and thus pose a rival divinity, Bernier argued that it would not have parts and that it would be imperceptible. His deepest argument was that space is nothing at all: for there to be space between two things, there need not be a third thing between them, any more than for two things to be equal there need be a thing - equality - to make them so. Equality and distance are 'abstract terms, which like all others of this sort, lead us into error if we conceive something abstract or separated from the concrete' (1684 II: 387-8).

Explicitly exempted from Bernier's critique is Gassendi's atomism, for he does not think that one can reasonably philosophize on any system other than that of atoms and the void. Yet even here Bernier gives arguments that extend Gassendi's position beyond physical theory in an adumbration of the metaphysical atomism of Hume.

See also: Atomism, ancient

THOMAS M. LENNON

List of works

Bernier, F. (1651) *Anatomia, ridiculi muris, hoc est dissertatiunculae I.B. Morini astrologi adversus expositam a P. Gassendo Epicuri Philosophiam… (An analysis of an absurdity, that is, of the short dissertation of the*

astrologer I.B. Morin against the Epicurean philosophy of P. Gassendi), Paris.(A defence of Gassendi against the outré attacks of the astrologer Jean-Baptiste Morin.)

(1653) *Favilla ridiculi muris, hoc est dissertatiunculae ridicule defensae a Ioan. Bapt. Morino astrologo adversus expositam a Petro Gassendo Epicuri philosophiam (The remains of an absurdity, that is, of the little dissertation defended by the astrologer J.-B. Morin against Gassendi's exposition of the philosophy of Epicurus)*, Paris.(A continuation of the polemic with Morin.)

(1670-1) *Memoires du sieur Bernier sur l'empire du grand Mogul*, Paris; trans. T. Brock as *Travels in the Mogul Empire AD 1656-1668*, New Delhi: Schand, 1968. (Earning its author the nickname of "Mogul", this account of his ten-year stay in the Indian subcontinent was a standard reference work well into the nineteenth century.)

(1671) *Requeste des maistres ès-arts, professeurs et regents de l'Université de Paris, presènte à la Cour souveraine du Parnasse, ensemble L'arrest intervenue sur ladite requeste contre tous ceux qui pretendent faire enseigner ou croire de nouvelles découvertes qui ne soient pas dans Aristote (Request of the authorities at the university of Paris, presented to the sovereign court of Parnasse, to intervene against all those who claim to teach or believe the new discoveries which are actually not in Aristotle)*, in G. Gueret, *La guerre des auteurs anciens et modernes*, The Hague.(A propaganda tract directed against the Sorbonne's attempt to re-establish Aristotelianism in the schools.)

(1678) *Abrégé de la philosophie de Gassendi (Abridgement of the philosophy of Gassendi)*, Lyons, 7 vols, 2nd edn, 1684; repr. Paris: Fayard, 1992.(First published in proto-editions beginning 1674. Although it has been regarded as a principal source for the views of Gassendi, this translation, paraphrase, summary and interpolation of Gassendi's main work sometimes evidences Bernier's own, different views.)

(1682) *Doutes de M. Bernier sur quelques uns des principaux chapitres de son Abrégé de Gassendi (Doubts of M. Bernier about some of the main chapters of his Abrégé)*, Paris. (A set of sceptical doubts about Gassendi's views on such topics as space, time, motion and collision; published in a second edition in volume two of the second edition of the Abrégé.)

(1684) *Éclaircissement sur le livre de M. de la Ville intitulé: Sentiments de M. Descartes touchant l'essence et les propriétés des corps (Clarification of the work by M. de la Ville entitled 'Thoughts of Descartes regarding the essence and properties of bodies')*, in P. Bayle *Receuil de quelques pièces concernant la philosophie de M. Descartes*, Amsterdam.(A reply to the Jesuit Le Valois, who pseudonymously had attacked Bernier's views as incompatible with the doctrine of transubstantiation.)

References and further reading

Lennon, T.M. (1993) *The Battle of The Gods And Giants: The Legacies of Descartes And Gassendi, 1655-1715*, Princeton, NJ: Princeton University Press.(Sections 5-9 discuss Bernier as participant in the seventeenth-century contest between Cartesian Platonism and materialism.)

Murr, S. (1992) 'Bernier et les Gassendistes', *Corpus* special double issue 20-1. The most extensive treatment ever accorded Bernier, comprising thirteen articles on topics related to him, and ten brief, rare texts either about or by him, including the *Requeste des maistres*.)

Bernstein, Eduard (1850-1932)

Eduard Bernstein, an eminent German social democrat, is now noted as 'the father of revisionism'. He made a reputation as the radical editor of the German Social Democratic Party organ, Der Sozialdemokrat, and became a close associate of Friedrich Engels. However, after the death of Engels he abandoned revolutionary Marxism and argued that socialism could be achieved by legal means and piecemeal reform. In doing this, he raised fundamental questions concerning the validity of Marxism and the direction of socialist political strategy, thus provoking what is now known as the 'revisionist debate'.

Eduard Bernstein was born in Berlin. In 1872, a year after the fall of the Paris Commune, he joined the *Eisenach* wing of the German socialist movement and soon became known as an activist. He attended the Gotha Conference in 1875, at which the *Eisenachers* joined with the *Lassalleans* to form what was to become the German Social Democratic Party (SPD). However, in 1878 the Reichstag passed legislation effectively making the SPD illegal, and Bernstein fled to Switzerland. Here, upon reading Engels' recently published *Anti-Dühring*, he became a Marxist. He made contact with Marx and Engels and laid the basis for a long and fruitful collaboration (particularly with Engels). Around the same time, he met Karl Kautsky, with whom he collaborated for many years. In 1881 he became editor of the official party organ, *Der Sozialdemokrat*, in which capacity he established a reputation as a Marxist of impeccable orthodoxy and the guardian of the party's radical conscience.

In 1887 Bernstein moved to London where he continued editing *Der Sozialdemokrat* and consolidated his relationship with Engels. But by this time Bernstein had begun to nourish doubts about revolutionary Marxism as a basis for party policy. The terminal crisis of capitalism, predicted by Marx and Engels, had not occurred and, as far as Bernstein could see, it was not going to occur. There was no evidence that the means of production were being concentrated in fewer and fewer hands, or that cut-throat competition was wiping out large sections of the middle classes, or that the proletariat was being relentlessly reduced to abject poverty. Capitalism seemed to be in robust good health and likely to remain so for the foreseeable future. It was, therefore, idle for socialists to pin their hopes on an imminent collapse of bourgeois society. On the other hand, the advance of democracy in most industrialized countries had enabled working-class parties to enter the political arena, and there was a real prospect that significant reforms could be achieved by parliamentary means. Indeed, Bernstein argued, the victory of socialism might well be accomplished by the steady implementation of socialist principles by democratic and constitutional means. In his view, it was not a matter of pursuing utopian visions. It was a matter of hard political work guided by basic socialist principles. 'The final goal of socialism' was, as he put it, 'nothing to me, the movement is everything' (Tudor and Tudor 1988: 168-9).

From these observations Bernstein drew two general conclusions. First, that Marx's doctrines would have to be re-evaluated and, where necessary, 'revised'. Second, the SPD should abandon its revolutionary aspirations and acknowledge that it was now a democratic socialist party of reform.

Bernstein developed these views, in part in a series of articles published in *Die Neue Zeit* under the title 'Problems of Socialism' and in part in a polemical exchange with the English socialist Ernest Belfort Bax (Tudor and Tudor 1988: 168-9). The consequent uproar within the party culminated in the rejection of Bernstein's 'revisionism' at the party conference at Stuttgart in 1898. Early in the following year Bernstein published his *Die Voraussetzungen des Sozialismus und die Aufgaben der Sozialdemokratie* (*The Preconditions of Socialism*). In it, he commended Marx and Engels for their open-ended, scientific approach. Their own investigations had, he argued, quite rightly led them to revise and qualify the initial formulations of their theories. They had, for instance, been the first to recognize the abstract nature (and therefore limited usefulness) of the theory of surplus value. They had, in their later years, amplified the materialist conception of history to allow political and ideological factors greater autonomy in effecting historical change. And they had modified their analysis of capitalist development in ways that made it less deterministic.

In Bernstein's view, these scientific advances pointed towards an evolutionary interpretation of the transition from capitalism to socialism. However, it was undeniable that Marx and Engels had drawn a different conclusion. They had never abandoned their revolutionary expectations. The reason for this, according to Bernstein, was that, to the very end, their thinking had been confined within the straightjacket of Hegelian dialectics. It was Hegelian

dialectics, not their painstaking scientific work, that ultimately dictated their conclusions. Scientific socialists should, Bernstein argued, emulate Marx the scientist, not Marx the dialectician.

Bernstein's own philosophical predilections were ill-defined. His view of science and knowledge was vaguely positivist in character, and in ethics he was much influenced by the neo-Kantians, particularly Friedrich Albert Lange (see Neo-Kantianism). However, he was not a professional philosopher. Apart from the opening chapters of *Preconditions*, the nearest he came to stating his philosophical position was in his 1901 lecture, 'How is Scientific Socialism Possible?' (1976). Here he argued that science, by its nature, is disinterested; it is mere cognition and cannot move men to action. Socialism, however, is a movement with aims and objectives, and these embody, not the results of scientific investigation, but the interests of the working class. In short, socialism does move men to action, it is not disinterested, and it therefore cannot be scientific.

The controversy provoked by Bernstein's views lasted for many years. His opponents included luminaries such as Georgii Plekhanov, Rosa Luxemburg and, above all, his old friend Karl Kautsky (see Plekhanov, G. §2). The party conferences at Hanover (1899) and Dresden (1903) were devoted mainly to Bernstein and the 'revisionist' question. Bernstein himself returned to Germany in 1901 where he continued his literary activity and, for most of the rest of his life, served as Reichstag deputy. After various political vicissitudes, he died in 1932.

See also: Socialism; Marxism, Western

H. TUDOR

List of works

Bernstein, E. (1899) *Die Voraussetzungen des Sozialismus und die Aufgaben der Sozialdemokratie*, Stuttgart: Dietz; trans. and ed. H. Tudor, *The Preconditions of Socialism*, Cambridge: Cambridge University Press, 1993. (Bernstein's main theoretical work.)

Bernstein, E. (1901) *Zur Geschichte und Theorie des Sozialismus (Towards the History and Theory of Socialism)*, Berlin and Bern: J. Edelheim.(A collection of Bernstein's most significant articles.)

Bernstein, E. (1976) *Ein revisionistisches Sozialismusbild* (*A Revisionist Picture of Socialism*), trans. and ed. H. Hirsch, Berlin and Bonn: Dietz.(Three lectures, including 'How is Scientific Socialism Possible?' together with other materials from the revisionist debate.)

References and further reading

Gay, P. (1962) *The Dilemma of Democratic Socialism*, New York: Columbia University Press.(The only book-length study of Bernstein in English.)

Gustafsson, B. (1972) *Marxismus und Revisionismus (Marxism and Revisionism)*, Frankfurt: Europaische Verlagsanstalt.(A historical account of the development of Bernstein's ideas.)

Meyer, T. (1977) *Bernsteins konstruktiver Sozialismus (Bernstein's Constructive Socialism)*, Berlin and Bonn: Dietz .(On Bernstein's contribution to socialist theory.)

Tudor, H. and Tudor, J.M. (eds and trans) (1988) *Marxism and Social Democracy: The Revisionist Debate 1896-1898*, with intro. by H. Tudor, Cambridge: Cambridge University Press.(Contains all but two of the articles entitled 'The Problems of Socialism'.)

Beth’s theorem and Craig’s theorem

Beth’s theorem is a central result about definability of non-logical symbols in classical first-order theories. It states that a symbol P is implicitly defined by a theory T if and only if an explicit definition of P in terms of some other expressions of the theory T can be deduced from the theory T. Intuitively, the symbol P is implicitly defined by T if, given the extension of these other symbols, T fixes the extension of the symbol P uniquely. In a precise statement of Beth’s theorem this will be replaced by a condition on the models of T. An explicit definition of a predicate symbol states necessary and sufficient conditions: for example, if P is a one-place predicate symbol, an explicit definition is a sentence of the form $(x)(Px \equiv \phi(x))$*, where* $\phi(x)$ *is a formula with free variable x in which P does not occur. Thus, Beth’s theorem says something about the expressive power of first-order logic: there is a balance between the syntax (the deducibility of an explicit definition) and the semantics (across models of T the extension of P is uniquely determined by the extension of other symbols).*

Beth’s definability theorem follows immediately from Craig’s interpolation theorem. For first-order logic with identity, Craig’s theorem says that if ϕ *is deducible from* ψ*, there is an interpolant* θ*, a sentence whose non-logical symbols are common to* ϕ *and* ψ*, such that* θ *is deducible from* ψ*, while* ϕ *is deducible from* θ*. Craig’s theorem and Beth’s theorem also hold for a number of non-classical logics, such as intuitionistic first-order logic and classical second-order logic, but fail for other logics, such as logics with expressions of infinite length.*

1 The axiomatic method

Questions concerning the definability of concepts arose within the development of the formal axiomatic method by late nineteenth- and twentieth-century mathematicians such as Moritz Pasch, Giuseppe Peano, David Hilbert and Alfred Tarski. What sets apart the formal axiomatic method from the earlier axiomatic thinking of, say, Euclid’s geometry is that the primitive terms occurring in the axioms are uninterpreted and the axioms devoid of any meaning; and that the rules of reasoning have been made completely explicit and formal. The axioms state purely formal relationships between the terms. This suffices for the purpose of deducing theorems from the axioms by rigorous reasoning. The development of the formal axiomatic method is the culmination of the movement to rigorize mathematics started in the early nineteenth century by mathematicians such as Cauchy and Bolzano. But these earlier efforts were directed chiefly towards ontology, for example, cleansing the language of analysis of visual images and reference to movement - and a mathematical theory still had subject matter, albeit abstract. In the formal axiomatic method the subject matter is provided from the outside by an interpretation of the primitive terms of the theory. One and the same theory may be open to radically diverse interpretations.

A perspicuous system of axioms requires under the formal axiomatic method the independence of each axiom from the others: a dependent axiom can be dropped without loss of content. Peano developed a method to prove independence: give an interpretation of the primitive terms which makes the one axiom false and the others true. This now common method is the formalization of an earlier idea of Eugenio Beltrami who felt that the so-called non-Euclidean geometry of Lobachevskii lacked a ‘real foundation’, that is a foundation in actual physical space. In 1868 Beltrami offered an interpretation of this geometry in terms of the acceptable Euclidean geometry: Lobachevskian ‘geometry’ could be understood as being about a special kind of line (a geodesic) in a special kind of plane (a surface with constant negative curvature) in Euclidean space. But Beltrami’s project was one of meaning and he did not have the consistency of non-Euclidean geometry in mind nor was he concerned with the independence of Euclid’s parallel axiom. In fact, only two years later, the French mathematician Guillaume Hoüel pointed out that Beltrami’s construction showed the independence of the parallel axiom: while the other axioms are true for all planes, the parallel axiom holds only in planes of zero curvature (see Scanlan 1988).

Another idea from the pre-formal period was absorbed by the formal axiomatic method through the work of the Italian mathematician Alessandro Padoa, a close collaborator of Peano. Padoa was concerned with the definability of concepts. He pointed out that there exists a parallel between, on the one hand, the methodological concepts of being an axiom and being derivable as a theorem and, on the other hand, concepts from the theory of definition: the notion of primitive term corresponds to the notion of being an axiom, and the notion of definability corresponds to the notion of deducibility. A perspicuous system of axioms therefore also requires the

independence of the primitive concepts used in the axioms. Padoa was thus led to ask in 1900 whether there could be a method to prove independence of concepts, just as Peano’s method had shown independence of axioms.

Early in the nineteenth century the French mathematician Jose Diez Gergonne had suggested the contrasting terms ‘explicit’ and ‘implicit’ as regards definition (Gergonne 1918-19). Gergonne’s distinction was suggested by the difference, in algebra, between a set of solved equations, which gives as it were explicit definitions of the unknowns, and a set of unsolved equations which is strong enough to determine a unique solution for the unknowns. Gergonne characterizes implicit definitions as ‘phrases that make us understand one of the words that occur in it through the known meaning of the other words’. Analogously to the algebraic case, Gergonne requires that the number of unknown words should be equal to the number of phrases that together implicitly define them. Completely forgotten by the end of the nineteenth century it was Giovanni Vacca, an assistant of Peano, who around 1896 gave a short account of Gergonne’s paper on definitions. Padoa’s paper (1901), read at the First International Congress of Philosophy in Paris, is clearly motivated by Gergonne’s work. To prove the independence of a concept from the other concepts occurring in a theory Padoa proposed a new method: find a true interpretation of the theory, considered as an abstract system, that remains a true interpretation when solely the meaning of that concept is changed. Thus, though Padoa did not refer to Gergonne or use the term ‘implicit definition’, his so-called ‘two-model method’ establishes that a term is not implicitly defined by a theory in the sense of Gergonne.

Since Padoa did not indicate how to construct the two models it may be better to speak of the ‘Padoa criterion’ for undefinability. Padoa claimed without further proof that his criterion was both necessary and sufficient for the explicit undefinability of a given concept by means of the other concepts in a given theory. Sufficiency is clear: if an explicit definition were to be implied by the theory two such models could not exist since the truth of the explicit definition would force the uniqueness of the interpretation of the explicitly defined term given an interpretation of the other terms. But necessity is not obvious. Does the absence of two such models guarantee that an explicit definition exists and is derivable? With hindsight this question could not have been answered at that time for it requires a more careful specification of the underlying logical system than was available to Padoa.

Alfred Tarski (1935) answered the question affirmatively for a modification of the ramified theory of types of Whitehead and Russell’s *Principia Mathematica* (see Theory of types). His proof was a rather straightforward derivation within the system, since the meta-claim that Padoa’s two models do not exist can be expressed in the language of type theory.

2 Beth’s theorem and Craig’s interpolation theorem

In 1953 the Dutch philosopher and logician Evert Beth proved the necessity of Padoa’s criterion for first-order or elementary logic. Beth showed that if no explicit definition of a term can be deduced from a theory, two models of the theory exist that differ only in the interpretation of the term in question. Moreover, in his so-called semantic tableau method, Beth found the means to construct systematically, in the absence of definability, the two models required by Padoa’s criterion, albeit often through an infinite process, while a closed tableau makes it possible to find an explicit definition of the term in question (see Natural deduction, tableau and sequent systems §4). Beth thus took away some of the concerns of the American mathematician Oswald Veblen who had remarked in 1902 that what Padoa proposed ‘seems hardly adequate’ when the issue was to replace an axiomatic system by one with independent axioms and independent terms, since he gave no method to find the two models and thus prove independence, or to construct the explicit definition, in the case of dependency.

Let *L* be a first-order language and *P* an arbitrary non-logical constant not in *L*. Let $L(P)$ denote the language obtained by adding *P* to *L*. To simplify our notation we will assume that *P* is a one-place predicate symbol. If *T* is an arbitrary theory in the language *L*, then $T(P)$ will be a theory in the language $L(P)$. Deducibility in first-order logic will be denoted by ‘$\vdash$’. An interpretation or model *M* for *L* specifies extensions in a domain *D* for all the non-logical constants of *L* (see Model theory). If *M* is a model of *L*, a model of $L(P)$ will be denoted by (M, X), where *X* is a subset of the domain *D* of *M*. Thus *P* is here interpreted as the subset *X*.

We will now define the notions of explicit and implicit definability. $T(P)$ is said to *define P explicitly* if there is a formula $\phi(x)$ of *L* such that $T(P) \vdash (x)(Px \equiv \phi(x))$. If *P* is not a one-place predicate symbol, this definition can be modified in the obvious way. Furthermore, let *T* be the set of first-order consequences of $T(P)$ in the

language *L*, in which *P* does nor occur. Then $T(P)$ is said to *define P implicitly* when, for every model *M* of *T*, there is exactly one expansion (M, X) of *M* which is a model of $T(P)$.

‘Beth’s definability theorem’ for first-order logic states that a theory $T(P)$ defines a term *P* implicitly if and only if $T(P)$ defines *P* explicitly. Beth’s original proof of 1956 uses a modification of Gentzen’s ‘extended *Hauptsatz*’, which shows that, in first-order logic, every proof can be carried out without any detours. Nowadays Beth’s theorem is usually proved to be a direct implication of Craig’s interpolation theorem.

‘Craig’s interpolation theorem’ for first-order logic with identity says that if a sentence ψ of first-order logic entails a sentence θ there is an ‘interpolant’, a sentence ϕ in the vocabulary common to θ and ψ, that entails θ and is entailed by ψ. William Craig originally proved this theorem as a lemma to be used in obtaining a simpler proof of Beth’s theorem (Craig 1957). Since then, however, the result has come to stand on its own.

We will now sketch a proof that Craig’s theorem implies Beth’s theorem. Since first-order logic is complete, implicit definability, a model-theoretic condition, is equivalent to the following deducibility condition, which is in fact Beth’s original definition of (implicit) definability. Let P' be a one-place predicate not in *L* and distinct from *P*, and let $T(P')$ be the theory in $L(P')$ obtained by replacing *P* by P' in $T(P)$ wherever it occurs. Then *P* is implicitly defined by $T(P)$ if and only if $T(P) \cup T(P') \vdash (x)(Px \equiv P'x)$. Assume now that this condition holds, and that $T(P)$ is a finite set of sentences, or, rather, one big conjunction of axioms, and similarly for $T(P')$. So we can write $T(P)\ \&\ T(P') \vdash (x)(Px \equiv P'x)$. But then also $T(P)\ \&\ Pc \vdash (T(P') \rightarrow P'c)$, where *c* is a new individual constant not in *L*. By Craig’s theorem there is an interpolant $\phi(c)$ such that $T(P)\ \&\ Pc \vdash \phi(c)$ and $\phi(c) \vdash (T(P') \rightarrow P'c)$. Since P' does not occur in $\phi(c)$ it is also true that $\phi(c) \vdash (T(P) \rightarrow Pc)$. Thus $T(P) \vdash (Pc \equiv \phi(c))$. Since *c* does not occur in $T(P)$, we also have $T(P) \vdash (x)(Px \equiv \phi(x))$. This completes the proof that if $T(P)$ defines *P* implicitly, then $T(P)$ defines *P* explicitly. The other direction of Beth’s theorem follows independently of Craig’s theorem.

3 Further developments and applications

Again, let *T* be the set of first-order consequences of $T(P)$ in the language *L*. We said that $T(P)$ defines *P* implicitly when for every model *M* of *T* there is exactly one expansion (M, X) which is a model of $T(P)$. In Padoa’s method two models (M, X) and (M, X') of $T(P)$ are exhibited. Another way in which implicit definability can be violated is if there is a model of *T* that cannot in any way be expanded to a model of the full $T(P)$. Karel de Bouvère (1959) studied this so-called one-model method to show undefinability of addition and multiplication in number theory. In the philosophy of science literature this is called a failure of Ramsey eliminability of the term.

Not to be confused with the above concept of definability of a term in a theory is the concept of definability of a set in a model which, for Tarski, belongs to semantic definability rather than to the formal definability involved in Padoa’s question since now we have a fixed model for an interpreted language. Given a model *M* of *L*, a subset *X* of its domain *D* is *definable in the model M* if there is a formula of *L* with one free variable $\phi(x)$ such that $(x)(Px \equiv \phi(x))$ is true in (M, X), where *P* is interpreted as *X*. Obviously, if *P* is explicitly definable in $T(P)$ and if (M, X) is a model of $T(P)$, then *X* is definable in *M*. Moreover, the concept of definability in a model can be iterated, whereas definability of a term in a theory cannot since a set of terms is not itself a term of the language.

We say that the predicate *P* is definable in a model (M, X) of $T(P)$ if an explicit definition of *P* holds in (M, X). Different models of $T(P)$ may satisfy different, non-equivalent definitions. But a theorem proven by Lars Svenonius in 1959 shows that if *P* is definable in every model (M, X) of $T(P)$ then each model (M, X) of $T(P)$ satisfies one of a finite list of definitions. That is, $T(P)$ implies a (finite) disjunction of explicit definitions of *P*. This property is called ‘explicit definability up to disjunction’ or ‘piecewise definability’.

In model theory the concept of ‘a logic’ is defined and logics for which Craig’s interpolation theorem hold are said to have the Craig or interpolation property; similarly for the Beth property. Any usual logic with the Craig property has the Beth property, but the latter has been shown to be weaker.

See also: Definition; Geometry, philosophical issues in; Logical and mathematical terms, glossary of

ZENO SWIJTINK

References and further reading

Barwise, J. and Feferman, S. (eds) (1985) *Model-Theoretic Logics*, New York: Springer.(The papers by Ebbinghaus and by Makowsky give details about the relation between the Beth property and the Craig property in abstract model theory.)

Benthem, J.F. van (1978) ‘Ramsey Eliminability’, *Studia Logica* 37: 321-36.(A detailed exposition of Ramsey eliminability with a discussion of its relevance for the philosophy of the empirical sciences.)

Beth, E.W. (1953) ‘On Padoa’s Method in the Theory of Definition’, *Indagationes Mathematica* 15: 330-9.(Gives Beth’s result that for first-order logic implicit definability implies explicit definability in Beth’s original proof-theoretic formulation.)

Beth, E.W. (1956) *L’Existence en Mathématiques*, Paris: Gauthier-Villars.(Includes an example of the one-model method to show undefinability. This involves constructing a model of the consequences of the theory without the terms whose undefinability has to be demonstrated, a model which cannot be extended to a model of the full theory.)

Beth, E.W. (1962) *Formal Methods: An Introduction to Symbolic Logic and to the Study of Effective Operations in Arithmetic and Logic*, Dordrecht: Reidel.(Shows how Beth’s semantic tableau gives a method for constructing a Craig interpolant, and how this leads to the construction of an explicit definition in case Padoa’s method fails.)

Bouvère, K.L. de (1959) *A Method in Proofs of Undefinability, With Applications to Functions in the Arithmetic of Natural Numbers*, Amsterdam: North Holland.(De Bouvère’s dissertation with the one-model method, which shows that, under certain conditions, a term is essentially undefinable, that is, undefinable in any consistent extension.)

Chang, C.C. and Keisler, H.J. (1973) *Model Theory*, Amsterdam: North Holland.(Gives a complete account of interpolation and definition results for classical logic.)

Craig, W. (1957) ‘Three Uses of the Herbrand-Gentzen Theorem in Relating Model Theory and Proof Theory’, *Journal of Symbolic Logic* 22: 269-85.(Derives Craig’s theorem about the existence of an ‘intermediate’ formula or interpolant from a consequence of Gentzen’s extended *Hauptsatz*, and applies it in a proof of Beth’s definability theorem, formulating it as a result relating model theory and proof theory.)

Ebbinghaus, H.-D. (ed.) (1987) *Bibliography of Mathematical Logic*, vol. 3, *Model Theory*, Berlin: Springer. (§C40 gives an extensive bibliography on interpolation and definability.)

Gergonne, J.D. (1918-19) ‘Essai sur la théorie des definitions’, *Annales de Mathématiques Pures et Appliquées* 9: 1-35.(Referred to in §1. Defines the concept of implicit definition before the advent of the formal axiomatic method.)

Kleene, S.C. (1967) *Mathematical Logic*, New York: Wiley.(A good account of the formal axiomatic method.)

Padoa, A. (1901) ‘Essai d’une théorie algébrique des nombres entiers, précédé d’une introduction logique à une théorie déductive quelconque’, in *Premier Congrès International de Philosophie*, vol. 3, *Logique et Histoire des Sciences*, Paris: Armand Colin, 309-65; partial trans. ‘Logical Introduction to any Deductive Theory’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 118-23.(Referred to in §1. The ‘Logical Introduction on Deductive Theories in General’ has the first statement of Padoa’s criterion for the undefinability of a symbol on the basis of a theory.)

Scanlan, M.J. (1988) ‘Beltrami’s Model and the Independence of the Parallel Postulate’, *History and Philosophy of Logic* 9: 13-34.(Referred to in §1. Compares the work of Beltrami and Hoüel with the treatment of Euclidean and non-Euclidean geometry after the development of the formal axiomatic method by Pasch, Peano and others.)

Tarski, A. (1935) ‘Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe’, *Erkenntnis* 5: 80-100; trans. ‘Some Methodological Investigations on the Definability of Concepts’, in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press, 1956, 296-319.(Referred to in §1. Includes the proof that in the logical system of *Principia Mathematica* the two extensions required by Padoa’s method do exist if a term cannot be explicitly defined.)

Tarski, A. and Kuratowski, C. (1931) ‘Les opérations logiques et les ensembles projectifs’ (Logical Operations and Projective Sets), *Fundamenta Mathematicae* 17: 240-8.(The first occurrence of the expression ‘implicit definition’ in the sense of Gergonne in the modern literature. The authors point out that definitions by induction are a type of implicit definition.)

Tuomela, R. (1973) *Theoretical Concepts*, Vienna: Springer.(An accessible discussion of definability and Ramsey eliminability in the empirical sciences.)

Vacca, G. (1896-9) 'Sui precursori della logica mathematica. II. J.D. Gergonne' (On Precursors of Mathematical Logic. II. J.D. Gergonne), *Rivista di matematica* 6: 183-6.(Referred to in §1. Brings the work of Gergonne on definition to the attention of the new logic.)

Bhartṛhari (*c*. 5th century)

Bhartṛhari is the Indian philosopher of grammar par excellence. *Drawing on practically all the schools of thought of his time - religious, philosophical, linguistic and ritual - he uses elements from them to create a philosophy. This philosophy, while claiming to be grammatical, goes far beyond traditional grammar, constituting a new and remarkably original system of thought.*

1 Life, works and influence

Bhartṛhari is the author of the *Vākyapadīya* or *Trikāṇḍī* and probably the *Mahābhāṣyaṭīkā* or *Mahābhāṣyadīpikā*, perhaps the earliest commentary on the *Mahābhāṣya* of Patañjali, which survives only in part. The Indian tradition also ascribes to Bhartṛhari the *Vākyapadīyavṛtti*, the earliest surviving commentary on the first two books of the *Vākyapadīya*, but this is doubtful.

Bhartṛhari was long believed to have lived in the seventh century AD, according to the testimony of the Chinese pilgrim Yijing (eighth century AD). However, as his *Vākyapadīya* was known to the Buddhist philosopher Dignāga, this has pushed his date back to the fifth century AD. It is unlikely that Bhartṛhari was active any earlier than Kālidāsa, a famous Sanskrit poet and playwright, who is widely believed to allude to the Gupta emperor Candragupta II (375-413 AD).

The philosophy of Bhartṛhari is expressed in the *Vākyapadīya*, a difficult work whose serious scholarly study remains in its infancy. Later grammarians claim Bhartṛhari's philosophy to be inseparable from the grammatical tradition and cite from the *Vākyapadīya* selectively. However, other thinkers, such as Abhinavagupta and Dharmapāla found aspects of his text with which they concur. The Buddhist Dharmapāla is known to have written a commentary (now lost) on part of the *Vākyapadīya* and the Chinese pilgrim Yijing mistook Bhartṛhari for a Buddhist. Hindu Tantric thinkers (especially Abhinavagupta) took a deep interest in Bhartṛhari's work. His philosophy is also often looked upon as a form of Vedānta (see Vedānta). All these links and claims are justified by certain aspects of Bhartṛhari's thought, but they do not provide an integrated picture of his philosophy.

2 Philosophical outline

In order to understand Bhartṛhari on his own terms, he must be read in the context of his time. The *Vākyapadīya* must be understood against the background of the philosophical and religious currents with which Bhartṛhari was acquainted and to which he refers, although often implicitly. These references show that Bhartṛhari was a Vedic *brāhmaṇa* (most probably belonging to the Maitrāyaṇīya branch of the black Yajurveda), who was strongly influenced by the Vaiśeṣika philosophy and especially by Mādhyamika Buddhism (see Nyāya-Vaiśeṣika; Buddhism, Mādhyamika: India and Tibet). He was first and foremost a grammarian who claimed allegiance to Patañjali and tried to elevate grammatical studies to the rank of a philosophy with the further implication that its study would provide access to liberation. Bhartṛhari combined these and other ideas to produce a philosophy of his own for which he claimed no originality. However, his work stands apart from all that preceded and followed him, including the philosophy of the later grammarians who present themselves as his inheritors.

Modern scholarship has not yet reached agreement on the precise nature of Bhartṛhari's philosophy. His philosophy distinguishes between two levels of reality, a higher and a lower. Only the higher reality can be considered to be real; lower reality is not real in an absolute sense. Language cannot describe higher reality. Indeed, language plays a major role in bringing about the lower reality, which is the reality of everyday experience. The resemblance between these ideas and those current among the Buddhists of the time is striking. There are, however, a number of major differences. First, for Bhartṛhari language is not just any language: it is Sanskrit, the sacred language of the Brahmans, which took form in the Veda, the corpus of canonical texts believed to be eternal. The world of everyday experience is in this way created, or organized by the Veda. At this point, Bhartṛhari turns an essentially Buddhist argument into a confirmation of the pre-eminence of the Veda.

A second major difference concerns the nature of the higher, absolute reality. Variant schools of Buddhism had various ideas about this, depending on the school to which they belonged: consciousness, emptiness, the fundamental elements of existence (*dharma*). Bhartṛhari accepts none of these positions. For him the absolute is the totality of all there is, has been and will be. He used this idea, borrowed from earlier Brahmanical thinkers, as

an element in his own philosophy.

Lower reality is the result of a division of the absolute. The precise nature of this division is determined, among other things, by language. Sometimes Bhartṛhari also mentions analytical imagination (*vikalpa*), but there is reason to believe that he looked upon these different factors as amounting to the same thing. The ‘parts’ resulting from this division undergo the influence of a number of powers (*śakti*) of the absolute, foremost among them time and space. These operations bring about the objects of everyday experience, which are accordingly looked upon as consisting of ‘real’ and ‘unreal’ parts. Bhartṛhari is not categorical as to what constitutes the ‘real’ parts of familiar objects. Among the various possibilities he proposes are that the substance of, or the universal inhering in the object is its ‘real’ part. He does not choose between these alternatives. This is due to what J.E.M. Houben (1995) has called Bhartṛhari’s perspectivism: reality is different from different points of view. This perspectivism is a pervading characteristic of Bhartṛhari’s thought. It applies to lower reality, but not to the absolute.

Bhartṛhari’s vision of the absolute as the totality of all that exists, has existed and will exist, has repercussions in the realm of ordinary reality. For Bhartṛhari, any totality or whole is more real than its constituents. A vase, for example, is more real than its parts. Again we see how Bhartṛhari uses a Buddhist position for his own purposes, by turning it into its opposite. For the Buddhists wholes do not exist: only their ultimate parts (*dharma*) (see Buddhist philosophy, Indian). To Bhartṛhari, a whole is more real than its parts and the absolute whole is absolutely real.

3 Role of grammar

Bhartṛhari considers himself a grammarian. Generally, grammar deals with the analysis of language and produces words from stems and suffixes. Language also obeys the general rule that totalities are more real than their constituents. This is particularly clear in the case of words which convey meaning, whereas their constituents do not. Grammatical analysis, Bhartṛhari argues, is artificial as stems and suffixes are the inventions of grammarians. In this respect his position appears to be very close to that of Patañjali. Words, too, are merely the result of an artificial analysis of sentences which, in their turn, are parts of more encompassing and therefore more real, linguistic units.

Once again, Bhartṛhari uses ideas which he borrowed from Buddhism. The Sarvāstivādins postulated long before Bhartṛhari the existence of three entities (*dharma*), corresponding to individual phonemes, words and sentences. Bhartṛhari accepts these entities, but orders them in an hierarchical ontology in agreement with his overall vision of reality.

Grammar allows its practitioner to ‘ascend’ from the smallest elements isolated by grammatical analysis, such as phonemes, stems and suffixes, to ‘higher’ units of speech. This way he will learn about the world, which is largely determined by the linguistic analysis that is imposed upon it. He will also learn to appreciate the unreality of the everyday world. In the end, the realization that the highest reality is beyond language and concerns the totality of things can be attained. Insight into the all leads to liberation, as it does in a number of classical texts of Mahāyāna Buddhism. It is in this sense that Bhartṛhari states at the beginning of his *Vākyapadīya* that grammar is the door to liberation.

Bhartṛhari’s remarks concerning the nature of language should be understood in the light of the above. There has been much confusion among recent scholars about Bhartṛhari’s concept of the absolute, which is often depicted as being of the nature of speech. The *Vākyapadīya* does not support this point of view. Bhartṛhari does discuss the distinction between the real word, sometimes called *sphoṭa* and the sounds which manifest it. The real word, he believes, has no sequence. It is only the sounds that manifest it which are sequential. Among the manifesting sounds, he makes a distinction between primary (*prākṛta*) and secondary (*vaikṛta*) sounds. The former have the duration attributed to the real word, the latter are responsible for the differences of pronunciation between different speakers. Bhartṛhari’s perspectivism in the *Vākyapadīya* is pertinent here, implying that a different explanation of the same fact may be presented elsewhere in the same text without clear indication to that effect. As well as this explanation of the duration of a word in terms of primary sounds (*Vākyapadīya* 1.77) there are verses in the text (1.105-6), which speak of the *sphoṭa* as the first sound produced, whose duration is not affected by the sounds produced subsequently.

See also: Language, Indian theories of; Patañjali

JOHANNES BRONKHORST

List of works

Bhartṛhari (*c*.5th century AD) *Vākyapadīya (Bhartṛhari's Vākyapadīya)*, ed. W. Rau, Wiesbaden: Franz Steiner, 1977.(A difficult work whose serious scholarly study remains in its infancy.)

Bhartṛhari (*c*.5th century AD) *Mahābhāṣyaṭīkā(Mahābhāṣyadīpikā)*, ed. and trans. by V.B. Bhagavat, S. Bhate, J. Bronkhorst, G.V. Devasthali, V.P. Limaye, G.B. Palsule, Poona: Bhandarkar Oriental Research Institute, 1985-91, 8 vols.(Perhaps the earliest commentary on the *Mahābhāṣya* of Patañjali, which survives only in part.)

References and further reading

Bronkhorst, J. (1991) 'Studies on Bhartṛhari 3: Bhartṛhari on *sphoṭa* and universals', *Asiatische Studien/Études Asiatiques* 45 (1): 5-18.(A discussion of Bhartṛhari's notion of *sphoṭa* - the distinction between the real word and the sounds which manifest it.)

Bronkhorst, J. (1992) 'Études sur Bhartṛhari 4: l'Absolu dans le *Vākyapadīya* et son lien avec le Madhyamaka', *Asiatische Studien/Études Asiatiques* 46 (1): 56-80.(A discussion of the Absolute in Bhartṛhari's work and its link with the 'middle' school of Buddhism.)

Bronkhorst, J. (1996a) 'Sanskrit and reality: the Buddhist contribution', *Ideology and Status of Sanskrit: Contributions to the history of the Sanskrit language*, ed. J.E.M. Houben, Leiden: E.J. Brill, 109-35.(A discussion of the Sanskrit language.)

Bronkhorst, J. (1996b) 'Studies on Bhartṛhari 7: Grammar as the door to liberation', *Annals of the Bhandarkar Oriental Research Institute* 76: 97-106.(A discussion of Bhartṛhari's system of grammar.)

Houben, J.E.M. (1995) *The Saṃbandhasamuddeśa (Chapter on Relation) and Bhartṛhari's Philosophy of Language: a Study of Bhartṛhari Saṃbandhasamuddeśa in the context of the Vākyapadīya, with a translation of Helārāja's commentary Prakīrṇaprakāśa*, Groningen: Egbert Forsten.(Important study of Bhartṛhari's thought, taking as a point of departure a chapter of the third book of the *Vākyapadīya*.)

Iyer, K.A. Subramania (1969) *Bhartṛhari*, Poona: Deccan College.(The most comprehensive study of Bhartṛhari, his works and thought.)

Ramseier, Y. (1993) 'Bibliography on Bhartṛhari', *Asiatische Studien/Études Asiatiques* 47 (1): 235-267.(A complete bibliography of Bhartṛhari's work. Contains translations of the *Vākyapadīya* and the *Mahābhāṣyaṭīkā*.)

Bible, Hebrew

Although the Bible is not a work of systematic philosophy, it none the less contains a wide variety of philosophical and theological ideas which have served as the framework for rabbinic speculation through the centuries. Although these views about the nature and activity of God are not presented systematically, they do provide an overview of the ancient Israelite understanding of the Godhead, creation, divine providence and human destiny. Throughout rabbinic literature these notions served as the bedrock for theological speculation, and with the emergence of systematic Jewish philosophy in the Middle Ages, they came to preoccupy a variety of thinkers. Similarly, in the post-Enlightenment period until the present, scriptural teaching has served as the starting point for philosophical and theological reflection.

Foremost among scriptural beliefs is the conviction that one God has created the cosmos. As the transcendent creator of the universe, he reigns supreme throughout nature and is intimately involved in earthly life. God is both omnipotent and omniscient and exercises divine providence over all creatures - from on high he oversees all the inhabitants of the earth. In exercising his providential care, Scripture repeatedly asserts, God is a benevolent ruler who shows compassion and mercy to all. Furthermore, as lord of history, he has chosen Israel to be his special people and has revealed the Torah to them on Mount Sinai. The Jewish people are to be a light to the nations, and from their midst will come a Messianic redeemer who will inaugurate a period of divine deliverance and eventually usher in the world to come. Israel thus plays a central role in the unfolding of God's plan for all human beings.

1 l Divine unity

Pre-eminent among scriptural ideas is belief in the existence of one God. In the Hebrew Bible, the Israelites experienced God as the lord of history. The most uncompromising expression of his unity is the *Shema* prayer: ‘Hear, O Israel, the Lord, our God, is one Lord’ (Deuteronomy 6: 4). According to Scripture, the universe owes its existence to the one God, the creator of heaven and earth, and since all human beings are created in his image, all men and women are brothers and sisters. Hence the belief in one God implies that there is one humanity and one world (see God, concepts of).

Jewish biblical teaching therefore emphasizes that God alone is to be worshipped. As the prophet Isaiah declared, ‘I am the Lord, and there is no other, besides me there is no God;… I form light and create darkness, I make weal and create woe, I am the Lord, who do all these things’ (Isaiah 45: 5, 7).

2 Transcendence and immanence

For the biblical writers, God is conceived as the transcendent creator of the universe. Thus in Genesis 1: 1-2, he is depicted as forming heaven and earth: ‘In the beginning God created the heavens and the earth. The earth was without form and void, and darkness was upon the face of the deep; and the Spirit of God was moving over the face of the waters’. Throughout Scripture this theme of divine transcendence is repeatedly affirmed. Thus Isaiah proclaims:

> Have you not known? Have you not heard? Has it not been told you from the beginning? Have you not understood from the foundations of the earth? It is he who sits above the circle of the earth, and its inhabitants are like grasshoppers; who stretches out the heavens like a curtain and spreads them like a tent to dwell in.
>
> (Isaiah 40: 21-2)

Later in the same book Isaiah declares that God is beyond human comprehension: ‘For my thoughts are not your thoughts neither are your ways my ways, says the Lord. For as heavens are higher than the earth, so are my ways higher than your ways and my thoughts than your thoughts’ (Isaiah 55: 8-9).

In the book of Job the same idea is repeated - God's purposes transcend human understanding:

> Can you find out the deep things of God? Can you find out the limit of the Almighty? It is higher than heaven - what can you do? Deeper than Sheol - what can you know? Its measure is longer than the earth, and broader than the sea.
>
> (Job 11: 7-9)

According to the author of Ecclesiastes, God is in heaven whereas human beings are confined to the earth. Thus the wise should recognize the limitations of human knowledge: ‘Be not rash with your mouth, nor let your heart be hasty to utter a word before God, for God is in heaven, and you upon earth; therefore let your words be few’ (Ecclesiastes 5: 2). Despite this view of God’s remoteness from his creation, he is also viewed as actively involved in the cosmos. In the Bible, his omnipresence is stressed repeatedly. Thus the Psalmist rhetorically asks:

> Whither shall I go from thy Spirit? Or whither shall I flee from thy presence? If I ascend to Heaven, thou art there! If I take the wings of the morning and dwell in the uttermost parts of the sea, even there thy hand shall lead me.
>
> (Psalms 139: 7-12)

Throughout Scripture, God is also described as having neither beginning nor end. Thus the Psalmist stated, ‘Before the mountains were brought forth, or ever thou hadst formed the earth and the world, from everlasting to everlasting thou art God. (Psalms 90: 2).

In the Bible, the term *olam* is most frequently used to denote the concept of God’s eternity. In Genesis 21: 33 he is described as the Eternal God; he lives for ever (Deuteronomy 32: 40) and reigns forever (Exodus 15: 18; Psalms 10: 16). He is the living God and everlasting King (Jeremiah 10: 10); his counsel endures for ever (Psalms 33: 11) as does his mercy (Psalms 106: 1). For the biblical writers, God’s eternal existence is different from the rest of creation - he exists permanently without beginning or end (see God, concepts of).

3 Omnipotence and omniscience

Concerning God’s power, the belief in his omnipotence has been a central feature from biblical times. Thus in Genesis when Sarah expressed astonishment at the suggestion that she should have a child at the age of ninety, she was criticized: ‘The Lord said to Abraham, "Why did Sarah laugh, and say ‘Shall I indeed bear a child now that I am old?’ Is anything too hard for the Lord?"’ (Genesis 18: 13-14). Similarly, in the book of Jeremiah when the city of Jerusalem was threatened by the Chaldeans, God declared: ‘Behold, I am the Lord the God of all flesh: is anything too hard for me?’ (Jeremiah 32: 27). On such a view there is nothing God cannot do: what appears impossible is within his power (see Omnipotence).

Similarly, Scripture affirms that God is all-knowing (see Omniscience). As the Psalmist states, ‘The Lord looks down from heaven, he sees all the sons of men… he who fashions the hearts of them all, and observes all their deeds’ (Psalms 33: 13, 15).

4 Creation and providence

Turning from God’s attributes to his acts, Scripture declares that he created the universe:

> In the beginning God created the Heaven and the Earth. The Earth was without form and void, and darkness was upon the face of the deep; and the Spirit of God was brooding over the face of the waters. And God said, ‘Let there be light’: and there was light. And God saw that the light was good.
>
> (Genesis 1: 1-4)

That God controls and guides the universe is an essential belief. The Hebrew term for such divine action is *hashgahah*, derived from Psalm 33: 14: ‘From where He sits enthroned He looks forth (*hishgiah*) on all the inhabitants of the earth’. Such a view implies that the dispensation of a wise and benevolent providence is found everywhere - all events are ultimately foreordained by God. According to tradition, there are two types of providence: (1) general providence, God’s provision for the world in general, and (2) special providence, God’s care for each individual. In Scripture God’s general providence was manifest in his freeing the ancient Israelites from Egyptian bondage and guiding them to the Promised Land. The belief in the unfolding of his plan for salvation is a further illustration of such providential care for his creatures. Linked to this concern for all is God’s providential concern for every person. In the words of the prophet Jeremiah: ‘I know, O Lord, that the way of man is not in himself, that it is not in man who walks to direct his steps’ (Jeremiah 10: 23) (see Creation and conservation, religious doctrine of; Providence).

5 Goodness, revelation and sin

In exercising divine providential care for creation, Scripture repeatedly asserts that God is the all-good ruler of the universe. Thus in the Psalms, he is depicted as good and upright (25: 8); his name is good (52: 11; 54: 8); he is good and ready to forgive (86: 5); he is good and does good (118: 68); he is good to all (l45: 9). Throughout the biblical narrative God is viewed as the supremely beneficent creator who guides all things to their ultimate destiny. In the unfolding of his plan, he chose Israel as his messenger to all peoples. As creator and redeemer, he is the father to all. Such affirmations about God's goodness have given rise to intense speculation about the mystery of evil. In Scripture the authors of Job and Ecclesiastes explored the question why the righteous suffer, and this quest continued into the rabbinic period (see Goodness, perfect; Evil, problem of).

The Hebrew Bible also asserts that God revealed himself in history. According to tradition, God revealed 613 commandments to Moses on Mount Sinai: they are recorded in the Five Books of Moses (Genesis, Exodus, Leviticus, Numbers and Deuteronomy). These prescriptions, which are to be observed as part of God's covenant with Israel, are classified in two major categories: (l) statutes concerned with ritual performances characterized as obligations between human beings and God; and (2) judgments consisting of laws that would have been adopted by society even if they had not been decreed by God (such as laws regarding murder and theft). These 613 commandments consist of 365 negative prescriptions (prohibited) and 248 positive prescriptions (duties to be performed) (see Revelation).

For the biblical writers, sin is understood as a violation of these divine decrees. In Scripture the word *chet* (sin) means 'to miss' or 'to fail'. Here sin is understood as a failing, a lack of perfection in carrying out one's duty. The term *peshah* means a 'breach'; it indicates a broken relationship between human beings and God. The word *avon* expresses the idea of crookedness. Thus according to biblical terminology, sin is characterized by failure, waywardness and illicit action. A sinner is one who has not fulfilled his obligations to God.

6 The chosen people

A central feature of biblical Judaism is the belief that God chose the Jews as his special people. In the Bible the Hebrew root *b-h-r* (to choose) denotes the conviction that God selected the Jewish nation from all other peoples. As the book of Deuteronomy declares, 'For you are a people holy to the Lord your God: the Lord your God has chosen you to be a people for his own possession out of all the peoples that are on the face of the earth' (Deuteronomy 7: 6). According to Scripture, this act was motivated by divine love: 'It was not because you were more in number than any other people that the Lord set his love upon you and chose you, for you were the fewest of all peoples; but it is because the Lord loves you' (Deuteronomy 7: 7-8). Through its election Israel has been given an historic mission to convey divine truth to humanity. Thus, before God proclaimed the Ten Commandments on Mount Sinai, He admonished the people to carry out this appointed task:

> You have seen what I did to the Egyptians, and how I bore you on eagles' wings, and brought you to myself. Now therefore, if you will obey my voice, and keep my covenant, you shall be my own possession among all peoples; for all the earth is mine, and you shall be to me a kingdom of priests and a holy nation.
>
> (Exodus 19: 4-6)

God's choice of Israel thus carries with it numerous responsibilities. As God declares regarding Abraham in Genesis, 'For I have chosen him, that he may charge his children and his household after him to keep the way of the Lord by doing righteousness and justice' (Genesis 18: 19).

Divine choice demands reciprocal response. Israel is obliged to keep God's statutes and observe his laws. In doing so, the nation will be able to persuade the peoples of the world that there is only one universal God. Israel is to be a prophet to the nations, in that it will bring them to salvation. Yet despite this obligation, the Bible asserts that God will not abandon his chosen people even if they violate his covenant. The wayward nation will be punished, but God will not reject them: 'Yet for all that, when they are in the land of their enemies, I will not spurn them, neither will I abhor them so as to destroy them utterly and break My covenant with them: for I am the Lord their God' (Leviticus 26: 44).

As God's chosen people, Israel is to inherit a land of its own. Thus, in Genesis, God calls Abraham to travel to Canaan, where he promises to make him a great nation: 'Go from your country and your kindred and your

father's house to the land that will show you. And I will make of you a great nation' (Genesis 12: 1-2). This same declaration was repeated to Abraham's grandson Jacob who, after wrestling with God's messenger, was renamed Israel. After Jacob's son Joseph became vizier in Egypt, the Israelite clan settled in Egypt for several hundred years. Eventually Moses led them out of Egyptian bondage and the people settled in the land that was promised to their ancestors.

7 The Messiah

In the unfolding of God's providential plan for Israel and humanity, the Messiah is to play a pivotal role. The term 'Messiah' is an adaptation of the Hebrew *Ha-Mashiach* (the Anointed), a term used frequently in Scripture. Initially, in the Book of Samuel, the view was expressed that the Lord had chosen David and his descendants to reign over Israel to the end of time (2 Samuel 7; 23: 1, 3, 5). Eventually there arose the belief that the house of David would return in time to rule over the two divided kingdoms as well as neighbouring peoples. Such an expectation paved the way for the vision of a transformation of earthly life through a universal Messianic redemption.

See also: Enlightenment, Jewish; God, concepts of; Halakhah; Midrash; Maimonides, M.; Theology, Rabbinic

DAN COHN-SHERBOK

References and further reading

Anderson, G.W. (1972) *The History and Religion of Israel*, London: Oxford University Press.(Authoritative and well-written history of ancient Israel.)

Bright, J. (1996) *A History of Israel*, Philadelphia, PA: Westminster Press.(Excellent survey of biblical history.)

Cohn-Sherbok, D. (1996) *The Hebrew Bible*, London: Cassell.(Outline and brief description of the books of the Bible.)

Franfort, H., Frankfort H.A., Wilson, J.A. and Jacobsen, T. (1951) *Before Philosophy*, Baltimore, MD: Penguin.(Stimulating exploration of the relationship between the Bible and the ancient Near East.)

Jacobs, L. (1973) *A Jewish Theology*, London: Darton, Longman & Todd.(Definitive survey of the history of Jewish theology from biblical times.)

Biel, Gabriel (before 1425-95)

Biel was the last great systematizer of scholastic theology and philosophy. Not noted for originality, he sought to produce a synthesis of the work of his predecessors. His thought is pervasively religious; a profound sense of the freedom of God's will is basic to his perspective. He followed Ockham and Duns Scotus in emphasizing the sheer contingency of things. Nature, morality and salvation depend entirely on God's will, and God could have determined otherwise. Such a view places sharp limits on the ability of reason to discover the truth about the nature and will of God; Biel subordinates reason to faith (although he is a master in the use of reason to defend revealed truth). The radical freedom of God coexists with significant moral freedom in humanity, since it is decreed by God that humans should be free to play an active role in determining their own destiny. Implied in this view of the human situation is an activist, pragmatic tendency, an interest in concrete applications of theoretical insights rather than in abstract speculation for its own sake.

1 Life

Gabriel Biel was born sometime before 1425 at Speyer, and died on 7 December 1495, at Einsiedel in Schönbuch. He was the last and arguably the ablest champion of 'the modern way' in late medieval theology (that is, a nominalism based on the system of William of Ockham). He studied at Heidelberg (BA 1435; MA 1438), Erfurt (1442-3, 1451-3) and Cologne (from 1453). At Erfurt he was exposed to Ockham's thought, at Cologne to that of Albert the Great and Thomas Aquinas. Biel assumed the chair in theology at Tübingen in 1484; he was invested as rector in 1485 and again in 1489.

Pragmatic and spiritual concerns are interwoven into Biel's intellectual enterprise. His perspective is more pervasively pastoral and religious than is Ockham's. The deeply religious tone of Biel's thought reflects his role as a cathedral-preacher in Mainz and as a proponent of the 'modern devotion' (*devotio moderna*) throughout southern and central Germany. His concerns for spiritual guidance and moral growth find abundant expression in his *De communi vita clericorum (On the Common Life of Clerics)* (1468-77), *Canonis misse expositio (Exposition of the Canon of the Mass)* (1488) and *Collectorium circa quattuor libros Sententiarum (Commentary on the Four Books of Sentences)* (1501).

2 Practical concerns

The coherence of Biel's thought is derived from his commitment to Ockhamism. Although he consults other theologians on subjects that Ockham fails to treat in depth, Ockham's perspective provides the framework within which he appropriates their work. Yet Biel is no narrow partisan. He is eager to synthesize the best elements from the competing traditions of late medieval thought - the 'ancient way' (*via antiqua*) of Aquinas and Duns Scotus and the 'modern way' (*via moderna*) of Ockham and his disciples. Biel's thought is eclectic in the best sense: emphatic about his loyalty to Ockham's perspective, he does not hesitate to draw on the wisdom of Bonaventure, Aquinas and above all Duns Scotus. A wide-ranging dialogue with various philosophers and theologians makes Biel's work valuable as an encyclopedic overview of the diversity in Western thought at the end of the Middle Ages.

Biel seeks an integration not only of the ancient and modern 'ways' but also of knowledge and piety. His *Canonis misse expositio* (1488) is a remarkable example of the symbiosis between late medieval theology and spirituality.

The pragmatic tendency of Biel's thought can be seen in the attention he pays to social ethics, and the practical consequences of theological conclusions. This is seen most clearly in the fourth book of his *Collectorium circa quattuor libros sententiarum* (1501), where Biel explores the sources of legitimate political authority, the mutual obligations of rulers and citizens, the morality of usury and the place of money in a just economy. Biel's contributions to a modern economic perspective include his vindication of the role of money in economic transactions and of supply and demand (rather than abstract theoretical considerations) as determining a just price.

3 Human nature and justification

This practical emphasis is reflected in an activist impulse that is linked with Biel's optimism about the moral

potentialities latent in human nature: ‘To those who do what in them lies, God does not deny his grace’ (*Canonis misse expositio*, lect.59(D); II *Sent.* d.27 q.1 a.3 dub.3 (N)). Oberman (1963) regarded Biel’s doctrine of justification as ‘essentially Pelagian’, since Biel exalts the sinner’s ability to take certain crucial steps towards repairing the effects of sin by utilizing purely human moral resources, quite apart from any special enablement by God’s grace (see Pelagianism; Justification, religious §3). Others (Clark 1965; Ernst 1972; McGrath 1981) have contested that verdict. The themes of covenant and divine generosity make it clear that for Biel the ultimate basis of salvation is God’s mercy rather than human achievement. Furthermore, Biel teaches that the preparation for grace and the infusion of grace are simultaneous: even though the preparation of the soul for justification is accomplished without the assistance of grace, still there is never a moment when the required disposition for grace (sorrow for sin based on love for God above all things) exists in one who is not a recipient of grace (see Grace §2). Yet Biel’s doctrine seems at least quasi-Pelagian: a rhetoric of grace cloaks a spirituality of self-reliance. What Biel’s theology arouses is not a sense of dependence on the enablement of grace but rather an imperative to do what one can and must do for oneself. To be sure, Biel insists that sinners cannot save themselves or take the first step towards reconciliation with God; the process leading to justification begins with a covenant whose source is sovereign mercy, preceding every human act of goodness or merit. It is gracious that God established a covenant at all, gave the Law as a guide to its fulfilment, sent Jesus as Teacher/Example, created humanity with the moral endowment of conscience, reason, synderesis (an innate tendency of the will to prefer the good over the evil) and freedom, which make it possible (in some limited sense, at least) for sinners to merit the first grace. Once God has taken the initiative, however, Biel interprets the actual economy of salvation in such a way that the decisive issue is what the sinner’s free response will be. Technically, Biel’s doctrine may not be Pelagian, but for all practical purposes (psychologically and morally), it might as well be.

4 Nominalism, voluntarism and fideism

On the hotly-debated question of the ontological status of universals, Biel articulates a nominalist position, reflecting and reinforcing the inclination of his thought towards the concrete rather than the abstract (see Nominalism). Universals are merely names, applied arbitrarily to categories of entities; only individuals are actual existents. Other philosophical themes emerge from Biel’s doctrine of God, which follows Duns Scotus’ voluntarism in emphasizing God’s absolute freedom from any external necessity. The realm of what God could do, if God so willed, is limited only by the law of non-contradiction. Anything that is logically possible is a real possibility for omnipotence. From the infinite possibilities available, God freely chose a finite number to be actualized, but might have chosen differently (see Omnipotence §5). In the ultimate freedom of the divine will, God could have decided to create a vastly different physical, moral or salvific order; everything is utterly contingent on God’s absolutely free decision. This implies a radical contingency in the actual order of things. Right and wrong are defined by the arbitrary act of the divine will, for which no prior reason may be sought. ‘It is right because God wills it’ (*Canonis misse expositio*, lect.23 (E)), not vice versa. Biel is a master of dialectics in the service of revealed truth, but his voluntarism motivates a radical critique of the scope of reason’s competence. Reason can manipulate logical possibilities, but in the actual order of morality and piety, reason is powerless to discern the content of the divine will, which must be communicated through authority and accepted by a faith that is submissive to God’s revelation (in Scripture and tradition), interpreted according to the determination of the Church and, above all, the Pope. Biel seeks to be faithful, not original. His personal modesty combines with an emphasis on the primacy of faith over reason to foster an attitude of submission to the Church and a high view of papal supremacy.

5 Influence

Biel’s writings underwent numerous editions. His admirers included Johannes Altenstaig, Johann Eck and Wendelin Steinbach. Influential at the Council of Trent, Biel’s ideas continued over several centuries to be an important resource for a wide range of Roman Catholic thinkers.

The young Luther (§1) was a diligent student of the *Expositio* and *Collectorium*, but Biel’s relation to Luther is complex. Luther’s doctrine of justification is a massive assault on Biel’s synergism. Yet on several key issues, Luther’s theology betrays a basic continuity with Biel’s thought. Luther’s high view of the sacraments is faithful to much of what he learned from Biel. Biel’s voluntarism sets the stage for Luther’s radical view of divine sovereignty: God’s will is the sole basis for human salvation. When Luther stresses God’s hiddenness and calls

reason ‘the devil’s whore’, echoes of Biel’s fideism are unmistakable. Yet Luther’s dialectical skills were moulded by his early exposure to the clarity and rigour of Biel’s logic.

See also: Grace; Nominalism §§1-2; Omnipotence; Pelagianism

JOHN L. FARTHING

List of works

For works by Biel that are only available in Renaissance editions, see Oberman 1963 (in references and further reading).

Biel, G. (1468-77) *Tractatus magistri gabrielis Byell de communi vita clericorum (Treatise by Master Gabriel Biel on the Common Life of Clerics)*, ed. W. Landeen, *Research Studies* 28: 79-95.(An account of the spiritual disciplines of the *devotio moderna* as practised by the Brethren of the Common Life.)

Biel, G. (1488) *Canonis misse expositio (Exposition of the Canon of the Mass)*, Wiesbaden: Franz Steiner, 1963-76.(A commentary on the liturgy of the mass, outlining Biel’s position on moral and religious questions in dialogue with a variety of medieval intellectual traditions.)

Biel, G. (1501) *Epithoma pariter et collectorium circa quattuor libros Sententiarum (Synopsis and Commentary on the Four Books of Sentences)*, Mohr: Paul Siebeck, 1973-84.(Biel’s most thorough exposition of his position on a wide range of philosophical and theological issues.)

Biel, G. (1516) *Tractatus de potestate et utilitate monetarum*, trans. R.B. Burke, *Treatise on the Power and Utility of Moneys*, Philadelphia, PA: University of Pennsylvania Press, 1930.(An extract from the *Collectorium* (IV d.15 q.9), detailing a major aspect of Biel’s contribution to economic thought.)

References and further reading

Burkard, F. (1974) *Philosophische Lehrgehalte in Gabriel Biels Sentenzenkommentar unter besonderer Berücksichtigung seiner Erkenntnislehre (The Content of the Philosophical Teaching in Gabriel Biel’s Commentary on the Sentences, with Special Attention to his Theory of Knowledge)*, Meisenheim am Glan: Hain.(An exploration of the epistemology of Biel’s *Collectorium*.)

Clark, F. (1965) ‘A new appraisal of late-medieval theology’, *Gregorianum* 46: 733-65.(Referred to in §3. A defence of Biel’s teachings against the charge that his view of the human situation before God is Pelagian or at least semi-Pelagian.)

Ernst, W. (1972) *Gott und Mensch am Vorabend der Reformation (God and Man on the Eve of the Reformation)*, Leipzig: St Benno.(Referred to in §3. A defence of the orthodoxy of Biel’s teaching about justification.)

Farthing, J. (1988) *Thomas Aquinas and Gabriel Biel*, Durham, NC: Duke University Press.(An analysis of Biel’s citations and interpretations of materials drawn from Aquinas’ theology.)

Feckes, C. (1925) *Die Rechtfertigungslehre des Gabriel Biel und ihre Stellung innerhalb der nominalistischen Schule (Gabriel Biel’s Doctrine of Justification and its Place Within the Nominalist School)*, Münster: Aschendorff.(A critique of Biel’s doctrine of justification from a Thomistic point of view.)

Feckes, C. (1927) ‘Gabriel Biel, der erste grosse Dogmatiker der Universität Tübingen in seiner wissenschaftlichen Bedeutung’ (Gabriel Biel, the scholarly significance of the first great dogmatician of the University of Tübingen), *Theologische Quartalschrift* 108: 50-76.(An examination of Biel’s place in the early history of the University of Tübingen.)

McGrath, A. (1981) ‘The anti-Pelagian structure of "Nominalist" doctrines of justification’, *Ephemerides Theologicae Lovaniensis* 57 (1): 107-19.(Referred to in §3. A vindication of the orthodoxy of Biel’s doctrine of justification.)

Oberman, H. (1963) *The Harvest of Medieval Theology*, Cambridge, MA: Harvard University Press.(Referred to in §3. The most important comprehensive study of Biel’s thought, including invaluable notes, a helpful glossary and a very thorough bibliography, listing works by Biel available only in Renaissance editions.)

Ruch, C. (1932) ‘Biel, Gabriel’, *Dictionnaire de théologie catholique*, vol. 2, 814-25. (A thorough yet concise overview of religious themes in Biel’s writings.)

Bioethics

While bioethics, a part of applied ethics, is usually identified with medical ethics, in its broadest sense it is the study of the moral, social and political problems that arise out of biology and the life sciences generally and involve, either directly or indirectly, human wellbeing. Thus, environmental and animal ethics are sometimes included within it. In this regard, bioethics can be of broader concern than is either medical/biomedical ethics or the study of the moral problems that arise out of new developments in medical technology.

The interrelated issues of who or what has moral status, of what justifies a certain kind of treatment of one creature as opposed to another, and whether, if a creature has moral status, it can lose it, have proved especially important issues in this broadest sense of bioethics. The philosophical task of probing arguments for soundness appears essential to deciding these issues.

As a part of applied ethics, bioethics is exposed to the difficulty that (1) we do not agree in our moral convictions and principles about many of the cases that feature in bioethics, (2) we do not agree in the moral theories in which our moral principles find their home and by which we try to justify them, and (3) we do not agree in the test(s) of adequacy by which to resolve the disagreements at the level of moral theory. We seem left with no way of deciding between contending principles and theories.

1 Conceptions of bioethics

In contemporary discussions of applied ethics, the term ‘bioethics’ is used generally in three ways.

Narrowly, it refers to a number of moral problems that attend recent developments in biotechnology, especially developments in four overlapping areas. These areas concern life-saving technologies at the beginning and end of life, life-enhancing technologies to improve the quality of life, reproductive technologies, the technologies to do with genetic engineering and gene therapy, and the implications for humans of the recently announced technology for cloning sheep. It is morally difficult to decide when and how to use these various technologies. For example, advances in our ability to correlate diseases and illnesses with particular genes and advances in gene therapy/gene replacement obviously raise important moral issues about how we are to use and control this ability and the information about individuals that genetic testing increasingly yields us. Again, even rather small developments in reproductive technology raise the spectre in some countries, for example, of aborting foetuses of the ‘wrong’ sex (see Genetics and ethics).

More broadly, and much more commonly, ‘bioethics’ refers to medical ethics generally and to all the various problems - moral, social, political, economic - that arise therein, including those moral issues to do with the development of biotechnologies encompassed by the narrower view. Resource allocation, informed consent, abortion, euthanasia, physician-assisted suicide, surrogate motherhood, the doctor/patient relationship, genetic engineering/genetic enhancement: on this broader usage, all form part of bioethics (see Suicide, ethics of; Life and death; Reproduction and ethics). Since medical ethics has been perhaps the major growth area of applied ethics, this broader usage of the term ‘bioethics’ has become dominant in the Anglo-American world (see Medical ethics).

More broadly still, and in keeping with V.R. Potter’s introduction of the term (1971), ‘bioethics’ refers to the moral, social and political problems that arise from biology and the life sciences generally and that involve, directly or indirectly, human wellbeing. On this broadest usage, environmental ethics and animal ethics are parts of bioethics, though it should also be noted that Anglo-American medical ethicists increasingly include concerns to do with animals within their brief and so seek, at least in part, to incorporate these concerns into medical ethics (see Animals and ethics; Environmental ethics; Agricultural ethics).

Accordingly, there is no specific discipline that is bioethics. Rather, there are sets or series of moral problems that arise out of biotechnologies, medicine, and human interaction with animals and the environment and that, directly or indirectly, affect human wellbeing. Which set of problems is emphasized can vary among philosophers and ethicists generally, and some sets have come to be so widely discussed that they have taken on a life of their own. Thus, some issues in environmental ethics (such as whether and how an ecosystem has value) and a great many issues in animal ethics (including the moral status of animals and their use as experimental subjects) flourish in their own right, whatever their relationship to human medicine and biotechnology. Again, so prominent have the

problems of medical ethics become that medical ethicists, including those who speak of ‘biomedical ethics’, rarely engage directly with environmental issues. As indicated, the case of animal ethics is more complicated here, since the moral problems that surround medical experimentation upon animals for human benefit are increasingly regarded as important issues of medical ethics.

2 The role of the philosopher

Bioethics is a part of applied ethics (see Applied ethics). Since the mid-1960s, the application of ethical theory to moral, social and political issues has been one of the major growth areas of philosophy. To some extent, of course, philosophers had always been concerned with normative issues except, perhaps, in the very heyday of ordinary language philosophy. On the whole, however, they did little more than sketch some implications or draw some provisional conclusions based upon distant familiarity with a few empirical facts. All this has now changed, as philosophers have come to immerse themselves in the subject-matter of the areas and issues of normative concern to them, whether these come from medicine and biotechnology or from business, law and the affairs of social and political life (see Business ethics). Not only do they apply their theories in the midst of a much broader and deeper understanding of the empirical settings within which these issues arise but their works have also changed in other respects. Real-life examples abound; normative claims that are a part of ordinary life are legion; and attempts to justify such claims, as well as to unravel the concepts in terms of which they are expressed, are common. Expressed in more or less nontechnical language, these works in the main address the educated and not merely the philosophical public, and they have had the effect of taking philosophy to some extent into the public arena.

One must distinguish here between philosophers approaching an empirical discipline and members of such disciplines approaching philosophy, a distinction usually marked in the case of medical ethics as that between philosophical and clinical medical ethicists. The latter find that their practices raise certain moral issues that need to be resolved in order for them to act, or to advise others about how to act, and to act in what they think can be shown to be a justified manner; they turn to philosophy to help them achieve these ends. The task of the former, however, even when they are ethical consultants in hospitals, remains what the philosopher’s task has always been: to test arguments for soundness. To this end, they deploy the tools and canons of logic on behalf of accuracy in argument; they explore questions of meaning, implication, presupposition, derivation, relation, compatibility; they pry into and generate examples and counter-examples, both realistic and hypothetical; and so on. One important tool they use is the analysis of those concepts employed in the arguments being tested; analysis is not, however, an alternative view of what philosophers should be about, in some way competing with their assessment of arguments. Thus, while philosophers can immerse themselves in the factual material pertaining to the arguments they are assessing, increases in this material and in the knowledge pertaining thereto are not their concern. Applied ethics at its best, when done by philosophers, does not involve the philosopher as an agent of empirical discovery - therefore, complaints based upon lack of expertise or professional qualifications in the relevant empirical disciplines do not automatically succeed - but as the agent of sound and, therefore, probed or tested argument.

3 Moral status and justification of treatment

To consider only the very broad usage of ‘bioethics’, it is easy to see how this picture of what the philosopher is about inserts itself into the moral discussion. In the case of environmental ethics, for example, issues to do with the value, integrity and treatment of the environment, including whole ecosystems, are of fundamental concern, and the theoretical stances within which these notions are addressed become important. To focus upon value, traditional ethical theorists, such as utilitarians and Kantians, embrace value theories that are typically human-centred; environmental concerns figure in them through their effect upon and implications for human interests (see Kantian ethics; Utilitarianism). Nor will it matter that an essentially human-centred value theory is extended to some animals, since such a theory typically favours humans (and animals) on the strength of criteria (such as sentience and desire-satisfaction) that exclude inanimate nature. Many environmentalists claim to want a nature-centred, truly ecological ethic, in which the environment, viewed either holistically, as a well-functioning biotic community, or as separate parts of such a community, parts with a good of their own, counts for something quite apart from its service to human (and animal) interests, purposes and goals.

Traditional ethical theories incorporate a distinction between instrumental and intrinsic value (see Values). Nature is viewed as instrumentally valuable, and some facet of persons (or, more narrowly, agents) is viewed as intrinsically valuable, valuable in its own right. To the extent that some animals share in these facets, to that extent

they can be viewed as intrinsically valuable. It seems clear, however, that little of inanimate nature can be so viewed. Numerous environmentalists want to break out from these person- or agent-centred value theories and to extend the notion of intrinsic value into nature. Talk of 'intrinsic value', of course, is ambiguous, so we might characterize the matter this way: what some environmentalists appear to claim is that an ecosystem or nature is valuable independently of any reference to human (or animal) interests, states of mind or desires. Unless we can extend this concept of independent value to nature, we have no real case for regarding rocks, forests and dirt as part of an expanded moral community, in circumstances in which only those things within that community enjoy the moral protections that, for example, affect how they are treated. Of course, someone might think that we are under some duty not to pollute rivers, and so hold rivers as a result to enjoy moral protection, but this duty would not obviously make rivers count for anything in their own right. So the case for a wider view of the moral community seems in large measure to depend upon our agreeing that there are independent values in nature. The philosopher enters this discussion easily; for if we sever talk of value from any connection with humans (and their conative and cognitive states), so that there could in essence be values without valuers, we need to go on to provide an account of the generation of the whole notion of value. For this severance from humans has deprived us of what the traditional theories have seen as the origin of value. Thus, the generation and source(s) of value become important issues, ones that fall within the domain of the philosopher (see Value, ontological status of).

The issue of who or what is a part of the moral community has been an important one in bioethics with regard to foetuses, those in permanently vegetative states, members of future generations, and inanimate nature and animals. A related issue has been the matter of justification of treatment. If, for example, (some) animals are not members of the moral community, then the case for medical and scientific experimentation upon them seems relatively easy to make; if they are members of the moral community, then why do they not enjoy the protections accorded human members? Can we treat different members of that community differently? Why, if a heart transplant is required to save a child, should we use the heart of a healthy baboon instead of the heart of, say, an anencephalic infant?

A further issue is of obvious importance here: can, for example, a person or an agent, who has been a member of the moral community, cease to be a member? If membership is determined by some specific characteristic, then loss of that characteristic could remove one from the moral community, and the whole issue of how those outside the moral community can be treated arises. It seems important, then, to be able to retain membership. But how is this to be achieved? The integrative personality that we associate with human persons can come apart, and those in the advanced stages of Alzheimer's disease, unless they can be included by means of the interests of others in having them protected, can fall at risk. Yet, if they are included within the moral community through others, do they continue to count for anything in their own right? They remain alive, but they have lost, arguably, that which ensured their inclusion within the moral community; their inclusion now seems to depend upon the contingent fact of whether other people's interests encompass them.

Much of contemporary bioethics, then, is concerned with these interlocking questions of moral standing, justification of treatment, and loss of moral considerability in one's own right. The human cases of the brain dead, anencephalic infants, and those in a permanently vegetative state seem cases in point. May one permissibly use the organs of these individuals? If so, then the case for xenotransplantation, or the transplanting of animal organs into humans, is surely affected, even if not perhaps extinguished (see Moral standing).

4 The problem of disagreement

As a part of applied ethics, bioethics in its broad usages is subject to a difficulty that seems fundamental. The standard conception of applied ethics is one of the application of principles to practice, and the various subject areas of bioethics appear to share in this conception. Thus, much of recent medical ethics appears to be the working out in practice of the implications of those moral principles that one takes to be relevant to the cases at hand, and environmental and animal ethics are usually portrayed in the same light. Now this standard conception of applied ethics is thought by some to leave out such things as personal relations and emotions, as integral parts of how we view cases morally, but others think that it can be made to include these things (see Morality and identity §4). What seems quite clear, however, is that there is a deeper difficulty with this standard conception.

The fact is that we do not agree in our moral principles and do not agree, moreover, if several of them seem to apply to a particular case, on the measures or principles by which we weigh and balance the contending principles in order to decide which predominate in that case (see Moral judgment). This much is clear to all parties to the

debate, who stress different moral principles applicable to the situation or different weights to those principles. Nor can we be said to be moving towards a consensus in principles: in the cases of abortion, euthanasia, physician-assisted suicide, and homosexuality, for example, there is widespread, deep disagreement both over the permissibility of such acts and over the moral principles which are supposed to make them permissible or impermissible.

Moreover, we do not agree in the moral theories in which our moral principles find a home and by which we try to justify them. Utilitarianism, Kantianism, contractualism, virtue theory and more remain at odds with each other as philosophical accounts of morality. Those, for example, who use a broadly consequentialist theory continue to find themselves at odds with Kantians and others.

Yet the difficulty runs deeper still. For we do not even agree upon the test(s) of adequacy of the moral theories in which moral principles find their home. Indeed, this whole issue of the earmarks of adequacy in ethical theory is hotly disputed today, and resolution of it is hostage to our inability to agree on a number of things. One is how appropriate the model of the physical sciences and its determination of theoretical adequacy is for ethics. Another is the meanings and logical properties of the moral concepts and whether, say, the terms 'right' and 'good' are identical with and to be understood exhaustively in terms of natural properties, empirical properties occurring in the natural world (see Naturalism in ethics). Still another is the role of our moral intuitions in the assessment of theoretical adequacy in ethics (see Intuitionism in ethics). As for this last, any direct reliance upon intuitions is rejected by many, whereas indirect reliance through some form of 'reflective equilibrium', whether narrowly or broadly construed, has also proved contentious (see Moral justification §2). Of course, the overwhelming temptation may well be to think that moral intuitions have probative force, because the model of science can seem virtually to compel us to regard them as the data of ethics. But why should we think that that model is appropriate to ethics in the first place?

To many, then, bioethics suffers from a complaint that infects applied ethics generally, a complaint the resolution of which awaits further work in ethical theory rather than in applied ethics itself. The problem is not that people disagree over the moral principles (or the weight of such principles) by which they set out and discuss cases; it is that there is no obvious way at the moment of resolving this disagreement at the level of ethical theory wherein we seek to justify our principles. The result is that much of bioethics, and applied ethics generally, takes the form of setting out cases in terms of different moral principles and theories, without providing us with any way of deciding between them. One can pick and choose among those items on offer, but all the interesting philosophical work lies outside the scope of any such view of bioethics and applied ethics. This is one of the reasons why some challenge the view that applied ethics is, as it were, a separate branch of ethics.

See also: Bioethics, Jewish; Responsibilities of scientists and intellectuals; Technology and ethics

R.G. FREY

References and further reading

Beauchamp, T.L. and Childress, J.F. (1979) *Principles of Biomedical Ethics*, New York: Oxford University Press; 4th edn, 1994.(A major text in contemporary medical ethics.)

Brody, H. (1976) *Ethical Decisions in Medicine*, Boston, MA: Little, Brown; revised and extended 2nd edn, 1984. (A survey of a mix of issues in medical ethics from the standpoint of different theoretical positions.)

Culver, C.M. and Gert, B. (1982) *Philosophy in Medicine*, New York: Oxford University Press.(A general statement of theoretical positions in approaches to medicine and medical decision-making.)

Englehardt, H.T., Jr (1986) *The Foundations of Bioethics*, New York: Oxford University Press; revised and extended 2nd edn, 1995.(A defence of one important, philosophical position in medical ethics.)

Frey, R.G. (1996) 'Medicine, Animal Experimentation, and The Moral Problem of Unfortunate Humans', *Social Philosophy and Policy* 13: 181-211.(An attempt to show how views about animals affect certain cases involving humans in medical ethics.)

Gillon, R. (ed.) (1993) *Principles of Health Care Ethics*, Chichester: John Wiley & Sons.(A major survey of issues and debates in medical ethics and in bioethics, broadly considered.)

Harris, J. (1985) *The Value of Life: An Introduction to Medical Ethics* , London: Routledge & Kegan Paul.(An introduction to life and death decisions in medicine.)

Mepham, T.B., Tucker, G.A. and Wiseman, J. (eds) (1995) *Issues in Agricultural Bioethics*, Nottingham: University of Nottingham Press.(Essays in the area of bioethics, broadly understood.)

Pence, G. (1990) *Classic Cases in Medical Ethics*, New York, McGraw-Hill; revised and extended 2nd edn, 1995. (A survey of cases in medical ethics, including cases involving animal ethics.)
Potter, V.R. (1971) *Bioethics: Bridge to the Future*, Englewood Cliffs, NJ: Prentice Hall.(Considered to be the first text to introduce the term 'bioethics', where it is understood to include environmental concerns.)
Reich, W.T. (ed.) (1978) *Encyclopedia of Bioethics*, New York: Free Press, revised and extended 2nd edn, 1995. (Major compendium of discussions on issues in medical and animal ethics.)
Singer, P. (1994) *Rethinking Life and Death*, New York: St Martin's Griffin.(A challenge to the Judaeo-Christian ethic that has been traditionally understood to underlie Anglo-American medical animal ethics.)
VanDeVeer, D. and Regan, T. (eds) (1987) *Health Care Ethics*, Philadelphia, PA: Temple University Press. (Articles on important issues in medical and animal ethics.)
Veatch, R.M. (ed.) (1989) *Medical Ethics*, Boston, MA: Jones & Bartlett Publishers. (Major compendium on a number of important and lasting issues in medical ethics.)

Bioethics, Jewish

Jewish bioethics seeks to apply Jewish modes of normative discourse in bioethics. For some moral issues in medicine, explicit guidance may be found in the traditional sources of halakhah*; but many others require creative application of ancient or medieval precedents and norms. Much of the contemporary writing in this area takes the classical form of rabbinic Responsa to specific queries from adherents of the* halakhah*. But the field also includes contributions from thinkers who offer not rulings on religious law but the fruits of moral inspiration by the tradition.*

In the Judaic tradition the idea that each human being is created in God's image fosters a powerful commitment to saving and prolonging life. Murder and suicide are terrible sacrileges. Procreation is highly valued. Contraception is not easily countenanced, and still less is abortion. But abortion is clearly distinguished from homicide.

The symbolic preciousness of the divine image disallows disfigurement of human corpses. The implications of this prohibition for pathology and for the study of anatomy are hotly disputed. The prohibition is overridden, however, in cases of immediate life-saving.

Regarding triage and resource allocation, the universal egalitarianism implied by the idea of God's image fosters a powerful reluctance against bringing about any person's death, even to save a number of lives. Some tension arises between this egalitarianism and traditional structures of social and religious hierarchy.

1 Problems of method

Jewish bioethics, like other areas of applied Jewish ethics, faces foundational problems of method and disciplinary presuppositions. These problems arise from the conjunction of reasoned ethical discourse with a tradition whose core teachings are deemed to be divinely revealed.

Jewish religious law (see Halakhah) is a central aspect of the Jewish normative tradition, but when Jewish bioethics focuses on this aspect, it can become a discourse of authority, taking the form of commandments and injunctions addressed to those bound by that law. Such teachings, it has been forcefully argued, cannot validly claim relevance to non-Jews or even to secular Jews.

Certain writers, however, find in *halakhic* discourse a more universal aspect, whose norms do not draw their authority from the particular commitments of the Jewish covenant. Some locate this aspect of *halakhah* in the seven 'Noahide Laws', which are seen to represent a sort of natural law. Others find the universal aspects in specific broadly humanistic themes. But in either case, the challenge of imbuing Jewish bioethics with general relevance consists in identifying the appropriate strands within *halakhah*. Since *halakhah* is not only a positive legal system but an articulated code of values, the significance of its teachings need not depend on accepting the legal system's authority *en bloc*. The fact is that *halakhic* discourse often appeals to reasons that are intelligible independently of religious authority. It can thus be seen as constituting a language of moral discourse that is widely accessible and relevant and that bears its own distinctive emphases and valuations.

2 Basic attitudes towards medicine

Because Judaism attaches great value to human life, not only is homicide viewed as sacrilege, but there is a positive duty to rescue any person from mortal danger: 'Anyone who saves a single life - it is as though he had saved an entire world' (Mishnah Sanhedrin 4: 5). Yet according to one strand of Jewish tradition, this duty, particularly in the context of medical practice, may conflict with an ideal of utter acceptance of God's providence. Thus the rabbis found it necessary to deduce from Scripture an explicit divine dispensation for medical intervention: 'a physician is granted permission to heal' (Berakhot 60a).

Several medieval thinkers, most prominently Nahmanides (himself also a physician), retained some uneasiness about human self-reliance and depicted medicine as compromising perfect piety. Maimonides, by contrast, emphasized the divine origin of our capacities for healing and treated medicine as a most valuable art and profession that can secure the years human beings need to attain spiritual fulfilment. Mainstream *halakhic* policy follows Maimonides in accepting medical interventions and regards a physician's duty to heal as but one instance

of the general religious obligations to forestall dangers and to relieve human suffering. This duty is antecedent to any contractual relation between doctor and patient and, needless to say, holds even where the patient is unable to pay the costs of medical services.

In *halakhah* the obligation to forestall even a small risk to human life takes precedence over all other injunctions save three: the prohibitions against bloodshed, idolatry and incest. In practice, this confers great authority upon physicians. For it is they who can proclaim a state of emergency concerning the endangerment of human life. This authority does not extend, however, to the formulation of normative judgments. Thus ‘Medical Halakhah’ could not in principle be a set of special norms adopted by physicians. Rather, it takes the form of guidance for medical practice, derived from the *halakhic* tradition (see Medical ethics).

3 At the deathbed

In the classic sixteenth-century code of Jewish law and ritual, Joseph Karo’s *Shulkhan ‘‘Arukh*, the sections on caring for the sick are followed by a special section on ‘laws pertaining to *gosses* [that is, a dying person]’ (Yoreh De’ah 339). The sacredness of human life is conveyed by a Talmudic metaphor: the life of a *gosses* is like a flickering candle - ‘anyone who touches him is a shedder of blood’. When a person reaches this final phase, the soul’s departure must be neither hindered nor hastened, even where a quicker release is requested by the dying person. *Halakhic* writers place this rule in the context of the prohibition on suicide (see Genesis 9: 5-6 as interpreted at *Genesis Rabbah* 34: 13). The underlying principle, once again, is the exalted value of the life of the human being, created in God’s image.

Contemporary writers commonly ground the prohibition of suicide in affirmations of God’s sovereignty over life and death. Arguing against secular assertions of human autonomy, they condemn active euthanasia even at the patient’s explicit behest. ‘Passive’ euthanasia is cautiously accepted - extending, on some views, even to the removal of artificial life-support. This acceptance applies, however, only to a patient in the phase of *gosses* - a term traditionally limited to patients who could not be expected to live more than three days. In the context of modern medical capabilities, the proper scope of the term seems uncertain, but it is clearly narrower than that of ‘terminal’. In practice, determination of the point beyond which efforts to prolong life should be foregone is commonly left to the patient (and patient’s family), especially where further treatment itself involves some risk of death.

Against the main contemporary trend of opposition to active euthanasia and assisted suicide, stands the rabbinic endorsement of King Saul’s act in falling on his sword to avoid capture and torture. This biblical precedent, alongside the self-sacrifice of martyrs, is cited as an express qualification to the original prohibition on suicide. The precise circumstances justifying Saul’s suicide are variously defined. Some commentators, at least, find them simply in the expectation of inescapable death accompanied by great suffering.

It has also been argued that, under Jewish law, killing a terminal patient is less severe than murder, since capital punishment - the normative penalty for murder - does not apply to one who kills a terminally ill individual. It is not clear what moral significance the distinction carries, however, since it is but one among a host of instances in which the rabbis renounce capital punishment. In all these instances, the rabbis’ hesitation is arguably directed at the punishment, rather than at the culpability of the act.

4 Triage

Grounding the value of human life in the belief that human beings are created in God’s image implies a basic equality regarding life itself, transcending cultural boundaries and social hierarchy. No (innocent) person may be killed in order to save the life of another. But does the same standard point towards an absolute prohibition of killing, even to save a great many persons? An early rabbinic text reads:

> If a group is told by heathens: ‘Hand over one of your number and we shall kill him; otherwise, we shall kill all of you!’ they should all be killed, rather than surrender one individual of Israel.
>
> (Tosefta Terumot 7: 20)

The emphasis on ‘one individual of Israel’ has prompted an interpretation in terms of group solidarity in the face of ‘heathen’ persecution. But a prevalent alternative reading sees here a more general principle which would

clearly preclude, say, killing one person in order to provide transplant organs for the saving of several lives. Rabbinic debates are recorded regarding some exceptions to the rule, for instance, where the said individual is sure to perish anyway.

When the choice involves not killing but deciding whom to save, there is broad agreement that the numbers should count. This quantitative angle appears to extend to triage as well: the ‘one bed’ or ‘one kidney’ should go to the patient who is expected to gain a full lifespan, not to one whose death will only be somewhat postponed. According to some writers, there is room also for traditional considerations of personal status or merit: one learned in Torah might be given precedence over an ignoramus, one of priestly descent (a *kohen*) over an ordinary Jew, a man over a woman, or a Jew over a Gentile. Other writers reject any such distinctions; some strongly advocating ‘first come, first served’ (see Nursing ethics).

5 Procreation and abortion

Rabbinic traditions regarding abortion are far from consistent. Two oft-quoted passages (Mishnah Niddah 3: 7 and Babylonian Talmud Yevamot 69b) refer to pregnancy up to forty days as ‘mere fluid’. Beyond this point, some mandate abortion for the sake of a moderate interest of the woman, arguing that the embryo is simply ‘her body’ (‘Arakhin 7a). Others denounce it as ‘bloodshed’ (Sanhedrin 57b). But all draw a sharp boundary between a foetus, even at full term, which must be killed to save the woman, and a newborn, who may not be harmed (Mishnah Ohalot 7.6).

Contemporary views reflect this early diversity, differing greatly about the gravity of the prohibition of abortion at various stages of a pregnancy. The issue is not exhausted here, however, since the tradition emphasizes a positive duty of procreation. This is illustrated by the fact that infertility (of either spouse) has traditionally served as sufficient grounds for divorce. Contraception is in general not easily countenanced, and abortion is commonly viewed as slighting the value inherent in each new human life, created in God’s image.

Yet *halakhic* rulings on the morality of either contraception or abortion reflect also the force of the reasons weighing against pregnancy. These may include concerns about birth defects, the woman’s health, food supplies (in times of famine), or even the pursuits of one or both of the prospective parents. The weight assigned to such factors in modern rabbinic opinions can reflect in turn the extent to which patriarchal attitudes are replaced by more egalitarian ethical concerns.

The view of procreation as an important mitzvah (divinely commanded obligation) entails a generally positive attitude towards medical interventions against infertility. Artificial insemination and in-vitro fertilization are widely condoned, as long as they use no ‘third party’ gametes. But donor insemination is condemned by several *halakhists* as ‘artificial adultery’. The view that a woman may be fertilized only by her husband’s sperm is not accepted by all, however. Some argue that adultery means only sexual infidelity.

6 The human body as resource

The obligation to preserve one’s own life leads to some debate regarding the risks assumed in donating an organ such as a kidney. But most *halakhic* discussion regarding medical use of human bodies focuses on the bodies of the dead. First, it is necessary to determine when a person may be considered dead. Traditional *halakhic* teachings focus on breathing, and some contemporary scholars still adhere to this criterion, rejecting any notion of ‘brain death’. This rules out transplantation of a heart, liver or other vital organ. To remove the heart of one who has not ceased to breathe is viewed as outright murder. Others, including the Chief Rabbinate of Israel, have adopted innovative interpretations that permit such transplantations. According to one argument, the traditional criterion of ceasing to breathe is met upon complete death of the brain stem, since that precludes *autonomous* breathing.

The *halakhic* requirement of respect for the dead and the concomitant prohibition against disfiguring the dead raise serious problems regarding the use of corpses. Most interpreters understand the normative concern not in terms of the interests of the deceased but as a matter of reverence for God’s image. The prohibition is outweighed, however, by the duty of saving life. Thus transplants of kidneys, corneas and the like, which do not depend on endorsing brain death, are widely condoned.

But less immediate medical uses, such as dissections for the study and teaching of anatomy or surgery,

post-mortem examinations, and research using cadaveric organs or tissues, are hotly contended. Do such pursuits come under the overarching concept of *life-saving*, a duty that overrides virtually any prohibition? Or should that rubric be confined to immediate emergencies? The view adopted by many *halakhists* opposes post-mortem pathological practices, and poses a serious obstacle to the study of medicine. Those who follow such teachings yet avail themselves of modern medicine have consequently been criticized as free riders.

See also: Life and death; Halakhah; Medical ethics; Medicine, philosophy of

NOAM J. ZOHAR

References and further reading

Bleich, J.D. and Rosner, F. (1983) *Jewish Bioethics*, New York: Hebrew Publishing Company, 2nd edn. (Wide-ranging collection of essays offering a detailed introduction to the field from an Orthodox perspective.)

Feldman, D.M. (1974) *Marital Relations, Birth Control and Abortion in Jewish Law*, New York: Schocken. (Well-rounded presentation and analysis of the major sources and opinions.)

Green, R.M. (1985) 'Contemporary Jewish Bioethics: A Critical Assessment', in E.E. Shelp (ed.) *Theology and Bioethics*, Dordrecht: Reidel, 245-66.(An illuminating critique of what the author sees as a prevalent, unnecessarily rigid tendency in modern Jewish bioethics. Examples cover several important issues, with useful references.)

Jakobovits, I. (1975) *Jewish Medical Ethics*, New York: Bloch.(Revised text by the former Chief Rabbi of Great Britain, widely considered the founder of this academic subdiscipline.)

Newman, L.A. (1993) 'Talking Ethics with Strangers: A View from the Jewish Tradition', *Journal of Medicine and Philosophy* 18: 549-67.(Engaging the basic problem of particularism versus universalism in Jewish bioethics, including discussions of 'Noahide Law' and natural law.)

The Journal of Medicine and Philosophy (1983) 8 (3).(Issue devoted to Jewish bioethics.)

Steinberg, A. (1988-) *Encyclopedia of Medicine and Jewish Law*, Jerusalem: Schlesinger Institute, 5 vols.(In Hebrew; these are the first five volumes of the planned six. An English translation of the complete work by F. Rosner is in progress, and its publication is expected in 1998/9. The encyclopedia is a comprehensive and reliable reference work which presents virtually all issues addressed in *halakhic* bioethics. Its citations lead mainly to primary sources and are therefore most useful for those capable of reading rabbinic discourse in the original texts and languages.)

Zohar, N. (1997) *Alternatives in Jewish Bioethics*, Albany, NY: State University of New York Press.(A non-authoritarian presentation and analysis of various traditional voices in dialogue with contemporary philosophical bioethics.)

Blackstone, William (1723-80)

Blackstone produced the first systematic exposition of English law as a body of principles. His enterprise was founded upon the assumption that the detailed rules of English law embodied and enforced natural law. Blackstone's invocation of natural law has frequently been regarded as ornamental rather than substantial, but there is no good reason for taking this view. Blackstone is now remembered as much for Bentham's attacks upon him as for his own contribution.

Sir William Blackstone was the first Vinerian Professor of English Law in the University of Oxford; he subsequently became a judge of the King's Bench, and then of Common Pleas. He is most famous for his four-volume *Commentaries on the Laws of England*, which was published between 1765 and 1769.

Among philosophers, Blackstone is probably principally remembered for the attacks mounted upon his work by Jeremy Bentham (§1). This greatly underestimates his true significance, however, for the *Commentaries* constituted the first systematic treatment of English law as a body of orderly rights and principles. English law-books prior to Blackstone were unsystematic in arrangement: they were huge compendia of legal information arranged alphabetically, or collections of writs and pleadings with associated commentary. Blackstone introduced into English legal writing a tradition of exposition and analysis which related more specific rules to general principles, the latter forming an integrated structure of rights. This was not simply a matter of giving well-organized expression to traditional legal ideas: in order to express the common law in systematic form it was necessary to reconceptualize it fundamentally, and this Blackstone did by relating his account to what he conceived to be the law's foundations in natural law (see Natural law).

Blackstone begins his masterwork (1765-9: vol. 1, section 2) by embarking upon a general discussion of natural law. Anyone seeking to present Blackstone as a serious contributor to natural law theory is bound to be discouraged by this opening discussion: the points are loosely structured and ambiguous, and are asserted rather than argued for; the game of spotting contradictions has proved easy, if unrewarding. The insubstantial nature of the discussion has led many to conclude that Blackstone's invocation of natural law serves purely ornamental purposes. It is certainly true that such discussions were conventional among institutional writers of the period. Yet what does such a convention signify? We have not explained the presence of the discussion simply by pointing to its conventional nature; and an explanation of the convention itself requires an understanding of the deep assumptions underpinning the emergence of systematic legal writing.

Why should it be assumed that laws will form an orderly system? In so far as law is a product of authority, identified by its source, laws might well be expected to form a haphazard list, incapable of reduction to any limited set of general principles. For Blackstone, the relevant grounding perspective was provided by the assumption that English law was concerned to enforce natural rights. It was this perspective which enabled him to abandon the traditional framework of legal thought, within which the common law was thought of as a body of remedies, and to systematize the law in terms of substantive rights. Even though English law might not reflect natural law with perfect accuracy, it was capable of systematic study to the extent that it did so reflect natural law. Thus he endeavours to expound English law in terms of a system of natural rights, and he draws on the writings of natural lawyers throughout the *Commentaries* (and not simply in the opening section).

The tendency of Blackstone's work (in common with that of other treatise writers) to expound the extant legal rules in terms of the moral values which (in his view) served to justify them formed the focus for Bentham's attack upon him. Bentham argued that Blackstone conflated the law as it *is* with the law as it *ought to be*. Intellectual clarity required, in Bentham's view, a clearer separation between the roles of *expositor* and *censor*.

Those who deny the substantial nature of Blackstone's reliance upon a theory of natural law are unduly influenced by the *ad hoc* and broadly utilitarian character of many of his arguments. This does not seem to correspond to the conventional image of a natural law theory. In fact it is very much in line with what we might expect from a natural law writer in the mid eighteenth century. The theories of Grotius and Pufendorf, which were influential in the preceding century, had separated the mode in which we ascertain the *content* of natural law (in considerations of the requirements of social life) from the basis of its obligatory force (in the divine will). This meant that, behind

the deontological façade provided by the notion of the divine will, natural law developed a somewhat utilitarian and empirical character which is clearly reflected in Blackstone's work.

It is also suggested that Blackstone's defence of Parliamentary sovereignty involves a commitment to the legal positivist belief that legal validity is a matter of source alone, and not of content (see Legal positivism §1). The error here lies in the assumption that natural law theories must deny the source-based view of legal validity: the point of such theories, however, is to emphasize the extent to which the content of law reflects the requirements of natural law, rather than to offer any particular account of the criteria of legal validity.

Finally, Blackstone's extensive use of history is sometimes thought to be inconsistent with a commitment to timeless principles of natural law. There is indeed a tension between Blackstone's deployment of history and his invocations of natural law. He seeks to address this problem only on a rhetorical plane: he claims that the rights conferred by natural law are, in other countries, so debased that they may now fairly be called 'the rights of the people of England'. This enables him to use the language of one ideology, within which rights are an inheritance of the English, while not abandoning natural law theory, within which such rights are universal.

See also: Common law; Law, philosophy of; Legal hermeneutics

N.E. SIMMONDS

List of works

Blackstone, W. (1765-9) *Commentaries on the Laws of England*, 4 vols, Chicago, IL: Chicago University Press, 1979.(The 1979 edition contains valuable editorial introductions to each volume.)

References and further reading

Bentham, J. (1776) *A Comment on the Commentaries* and *A Fragment on Government*, ed. J.H. Burns and H.L.A. Hart, London: Athlone Press, 1977.(The two works, the former unpublished in the author's lifetime, which set forth Bentham's - frequently unfair - critique of Blackstone.)

Boorstin, D.J. (1996) *The Mysterious Science of the Law: An Essay on Blackstone's Commentaries*, 2nd edn, Chicago, IL: University of Chicago Press.(A valuable if idiosyncratic survey.)

Lobban, M. (1991) *The Common Law and English Jurisprudence 1760-1850*, Oxford: Clarendon Press.(Chapter 2 concentrates on Blackstone.)

Simmonds, N.E. (1988) 'Reason, History and Privilege: Blackstone's Debt to Natural Law', *Zeitschrift der Savigny-Stiftung fur Rechtsgeschichte: Germanistische Abteilung* 118: 200-13.(Examines Blackstone's invocation of natural law.)

Blair, Hugh (1718-1800)

Blair was the foremost literary critic and preacher of the Scottish Enlightenment. He participated in the thriving cultural life of eighteenth-century Edinburgh, and along with William Robertson, Adam Ferguson and other enlightened Moderate party clergymen was a close friend of that city's greatest philosopher, David Hume.

As minister of the High Kirk in St Giles Church, Edinburgh, from 1758 until his death, and Regius Professor of Rhetoric and Belles-Lettres at the University of Edinburgh from 1762 to 1784, Hugh Blair occupied Scotland's most prestigious platforms for literary and religious oratory. His phenomenally successful publications were direct outgrowths of his preaching and teaching. In 1807 the Critical Review declared Blair's *Sermons* (1777-1801) to be, excepting the Spectator, 'the most popular work in the English language'. His *Lectures on Rhetoric and Belles-Lettres* (1783), the first comprehensive guide to the rules of written and spoken English in the various branches of polite discourse, enjoyed comparable popularity in the literary world. One of his early academic lectures, revised and published in 1763 as *A Critical Dissertation on the Poems of Ossian*, did much to encourage the subsequent Ossianic vogue. These works by Blair were frequently reprinted and translated; along with his professorship and church living, they brought him international fame and considerable income.

Blair contributed to the philosophy of the Scottish Enlightenment (see Enlightenment, Scottish) in two major ways. First, his Lectures included influential discussions of the principles of literary taste and philosophical discourse. At their core is the idea that clarity, simplicity and perspicuity constitute the proper foundation of all forms of communication. In regard to philosophical writing specifically, Blair's thirty-seventh lecture states that 'beyond mere perspicuity, strict accuracy and precision are required in a Philosophical Writer', in addition to 'embellishment' attained through 'illustrations' and 'a polished, a neat, and elegant Style' that avoids 'too much ornament'. Plato and Cicero among the ancients, and John Locke among the moderns, are held up as the models of clear and elegant philosophical discourse, whereas Seneca and the Earl of Shaftesbury are censured for their excessively ornamental styles.

Second, Blair's *Sermons* popularized the ideal of Christian Stoicism, or the belief that happiness is to be found within - in resignation to the will of God - rather than in worldly matters. Blair repeatedly asserts 'that mind is superior to fortune; that what one feels within is of much greater importance than all that befalls him without'; that 'the happiness of every man depends more upon the state of his own mind, than … upon all external things put together', and that awareness of the nature of the human condition points out 'how submissive ought we to be to the disposal of Providence'. Such preaching represented a comforting, polite form of Presbyterianism, which complemented Stoic elements in the philosophy of other thinkers of the Scottish Enlightenment.

See also: Common Sense School; Ferguson, A.; Hume, D.; Stoicism

RICHARD SHER

List of works

Blair, H. (1763) *A Critical Dissertation on the Poems of Ossian.*(Laudatory introduction to James Macpherson's Ossianic poetry, revised in the second edition of 1765 to answer doubts raised by David Hume.)

(1777-1801) *Sermons*, 5 vols. (Most popular sermons of the late eighteenth and early nineteenth centuries.)

— (1783) *Lectures on Rhetoric and Belles-lettres*, ed. H.F. Harding, Carbondale, Il: Southern Illinois University Press, 1965, 2 vols.(Facsimile of first edition, with a useful introduction and bibliography.)

References and further reading

Bevilacqua, V. (1967) 'Philosophical Assumptions Underlying Hugh Blair's Lectures on Rhetoric and Belles-Lettres', *Western Speech* 31: 150-64.(Emphasizes Blair's ties with the Scottish common sense tradition.)

Daiches, D. (1990) 'Style Périodique and Style Coup: Hugh Blair and the Scottish Rhetoric of Independence', in R.B. Sher and J.R. Smitten (eds) *Scotland and America in the Age of the Enlightenment*, Edinburgh: Edinburgh University Press, 209-26. (A suggestive reading of the influence of Blair's rhetorical theory on the American Founding Fathers.)

Hill, J. (1807) *An Account of the Life and Writing of Hugh Blair*, Edinburgh.(The standard contemporary life.)

Miller, T.P. (1990) 'Witherspoon, Blair and the Rhetoric of Civic Humanism', in R.B. Sher and J.R. Smitten (eds)

Scotland and America in the Age of the Enlightenment, Edinburgh: Edinburgh University Press, 100-14. (Contrasts Blair's belletristic approach to rhetoric with Witherspoon's civic one.)

Schmitz, R.M. (1948) *Hugh Blair*, New York.(Still the standard modern biography, though dated and flawed.)

Sher, R.B. (1985) *Church and University in the Scottish Enlightenment: The Moderate Literati of Edinburgh*, Edinburgh: Edinburgh University Press.(A study of Blair's circle of Presbyterian clergymen of letters, with a lengthy bibliography.)

Warnick, B. (1993) *The Sixth Canon: Belletristic Rhetorical Theory and Its French Antecedents*, Columbia, SC: University of South Carolina Press.(Relates Blair's theory of rhetoric to earlier French theorists.)

Blanchot, Maurice (1907-)

Maurice Blanchot, has since the 1940s been a dominant voice in French philosophy and letters, initiating a postmodern discourse which has had a profound impact on Bataille, Levinas, Foucault and Derrida. His early writings, between 1930 and 1940, consisted of cultural and political criticism. The experience of the Second World War led him to disengage from politics and he became an essayist and novelist. His works have included novels, narratives, and criticism, notably. Since the 1970s he has produced a series of fragmentary writings in which the line between literature and philosophy is shattered and, since the 1980s, meditations on language, death, the 'disaster' and community.

Basing himself on Mallarmé, Blanchot distinguishes two kinds of language, or speech. Crude speech is utilitarian, descriptive, and representational; it puts us in touch with the object-world. Essential speech is poetic, what Blanchot designates as the 'pure language' through which, as he asserts in *L'Éspace littéraire* (1955) (*The Space of Literature*, 1982) 'it is being that tends to speak and speech that wants to be'. Essential language can reveal what is normally concealed, something which still is when everything has disappeared: being. 'The essence of being is to be there still where it lacks, to be inasmuch as it is hidden' ([1955] 1982: 41, 253). As with Heidegger, this poetic language has ontological priority for Blanchot. For him, the poem, literature, is an act of transgression; it reaches beyond all limits. As he says in *L'Entretien infini* (1969) (*The Infinite Conversation*, 1993), following Bataille, 'Transgression designates what is radically out of reach: assailment of the inaccessible, a surpassing of what cannot be surpassed' ([1969] 1993: 453).

Writing as transgression, literature as an expression of essential speech, occurs under the spectre of human finitude, and so as a confrontation with death. This link between writing and death is particularly clear in the novel *Le Très-Haut* (1948) (*The Most High*, 1996), in the last line of which, the hero, Henri Sorge, at the very moment of his death, cries out: 'Now, now I'm speaking' ([1948] 1996: 254). Death is the primordial experience for Blanchot, and it is through their encounter with it that humans make themselves mortal. Hence, as he says in *The Space of Literature*: 'Death, in the human perspective, is not a given, it must be achieved' ([1955] 1982: 96). Blanchot distinguishes between two kinds of death: death as the disappearing of consciousness, the biological extinction which humans flee, the dying which does not permit us to complete our death; and death as consciousness of disappearing, the death which the poet wants to shape, and through which humans seek to achieve their mortality. It is because death as consciouness of disappearing is so elusive, indeed even self-contradictory, that Blanchot is driven beyond the limits of discursive language in his effort to grapple with it.

For Blanchot, modernity's quest for community is haunted by totalitarianism, and by what he designates as the disaster. While his image of totalitarianism as an omnipresent feature of the political topography of the twentieth century has been shaped by the experiences of Nazism and Stalinism, it is not exhausted by them. For Blanchot, the danger of totalitarianism is that all negativity or independent action will be appropriated by the state, all opposition recuperated, all criticism absorbed. His fear, expressed in *L'Écriture du désastre* (1980) (*The Writing of the Disaster*, 1986), is that the totalization of the state in modernity will lead to a world in which 'the prisoners construct their prison themselves' ([1980] 1986: 45). The disaster, as Blanchot conceives it, is encapsulated by the totalitarian state, with its camps, in which what takes place is 'dying, as forgetfulness of death' ([1980] 1986:17). So powerful is the presence of the Holocaust, and its death camps, that Blanchot designates it 'the *absolute* event of history' ([1980] 1986: 47).

Yet in the face of the disaster, Blanchot sees humankind's search for community. Community appears in two forms in Blanchot's thinking. In the 1930s, in his earliest writings, his views were shaped by the extreme right-wing *Action Française*, to which he was then linked. At that time 'community' meant an organic community, based on rootedness in a common place, a native soil, and ethnicity. The experience of Nazi barbarism and occupation led him to reject any conception of a traditional community, opening the way to his thematization of *l'autrui*, the other, as a basis of his thinking. Here, we can see the profound influence of his friend Emmanuel Levinas. As the antithesis of totality, and identity, as the embodiment of *l'autrui*, the Jew becomes the point of departure for a different vision of community. This Blanchotian vision, articulated in *La Communauté inavouable* (1983b) (*The Unavowable Community*, 1988) is an elective community based on friendship and open to alterity, in which the

other is no longer despised, and execrated; a community in which *l'autrui* is an irreducible element. In contrast to the traditional communities imposed on humans by virtue of blood or race, such an elective community would gather its members around a choice, 'that gave permission to everyone, without distinction of class, age, sex or culture, to mix with the first comer as if with an already loved being, precisely because he was the unknown-familiar' ([1983b] 1988: 30).

Blanchot's daring linkage of the dangers of the totalizing state, and the disaster which confronts humankind, as well as the need for a community in which *l'autrui* is at home, to the experience of the Holocaust, and its death camps, make his a seminal voice in postmodern thinking.

See also: Holocaust, the; Totalitarianism; Alterity and identity, postmodern theories of

ALAN MILCHMAN
ALAN ROSENBERG

List of works

Blanchot, M. (1948) *Le Très-Haut*, Paris: Gallimard; trans. A. Stoekl, *The Most High*, Lincoln, NE: University of Nebraska Press, 1996.(A dystopian novel, in which the totalization of the state and the prospects for revolt are thematized.)

Blanchot, M (1955) *L'Éspace littéraire*, Paris: Gallimard; trans. A. Smock, *The Space of Literature*, Lincoln, NE: University of Nebraska Press, 1982. (Explorations of the relation of literature to artistic creation, time and death.)

Blanchot, M (1969) *L'Entretien infini*, Paris: Gallimard; trans. S. Hanson, *The Infinite Conversation*, Minneapolis, MN: University of Minnesota Press, 1993.(Blanchot's dialogues with Pascal, Kafka and Nietzsche, and his meditations on language, nihilism and what it is to be Jewish.)

Blanchot, M (1980) *L'Écriture du désastre*, Paris: Gallimard; trans. A. Smock, *The Writing of the Disaster*, Lincoln, NE: University of Nebraska Press, 1986. (An inquiry, in fragments, which seeks to write about the disasters of the twentieth century: total war, concentration camps and the Holocaust.)

Blanchot, M (1983a) *Après coup*, Paris: Éditions de Minuit; trans. P. Auster, *Vicious Circles: Two Fictions and 'After the Fact'*, Barrytown, NY: Station Hill Press, 1985.(Two pre-war narratives and a 1983 addendum which look forward to, and back from, the Holocaust, and raise the question of the difficulty of bearing witness.)

Blanchot, M (1983b) *La Communauté inavouable*, Paris: Éditions de Minuit; trans. P. Joris, *The Unavowable Community*, Barrytown, NY: Station Hill Press, 1988.(A meditation on the meaning and possibility of community, through reflections on the writings of Georges Bataille and Marguerite Duras.)

References and further reading

Foucault, M. (1986) *La Pensée du dehors*, Paris: Fata Morgana; trans. B. Massumi, 'Maurice Blanchot: The Thought From Outside', in *Foucault/Blanchot*, New York: Zone Books, 1987.(A particularly insightful treatment of Blanchot's understanding of writing and language as a thought that stands outside subjectivity.)

Gregg, J. (1994) *Maurice Blanchot and the Literature of Transgression*, Princeton, NJ: Princeton University Press.(A comprehensive overview of Blanchot, which includes an excellent bibliography of both primary and secondary sources.)

Levinas, E. (1975) *Sur Maurice Blanchot*, Paris: Fata Morgana.(An important contribution by a thinker both influenced by Blanchot, and an important influence on him.)

Mehlman, J. (1983) 'Blanchot at Combat: Of Literature and Terror', in *Legacies of Anti-Semitism in France*, Minneapolis, MN: University of Minnesota Press.(An analysis of Blanchot's early right-wing political journalism, and its intellectual sources.)

Ungar, S. (1995) *Scandal and Aftereffect: Blanchot and France since 1930*, Minneapolis, MN: University of Minnesota Press.(A significant contribution to elucidating Blanchot's origins as a theorist of the Young Right, and possible links to the themes of his mature writings; this work includes a comprehensive bibliography of the early writings.)

Blasius of Parma (d. 1416)

*Blasius of Parma was an important Italian philosopher, mathematician and astrologer who popularized the achievements of Oxford logic and Parisian physics in Italy. He questioned the Aristotelian foundations of medieval physical science, mechanics, astronomy and optics, thus helping to open the way to the mathematics, optics and statics of modern times. His teaching influenced the artists of the Florentine Renaissance in their rediscovery of linear perspective, and his discussion of proportions influenced the Paduan mathematicians up to the time of Galileo. He presented an atomist and quantitative account of physical reality, and a materialist account of the human intellect. His consequent denial of the immortality of the soul won him the title of 'diabolical doctor' (*doctor diabolicus*). His position on the human ability to avoid astrological determinism was equivocal. Though his work was scholastic in style, he enjoyed good relations with such Italian humanists as Vittorino da Feltre, whose request for lessons in mathematics he refused. In Florence, he took part in conversations between humanists and scholastics.*

1 Life

Blasius of Parma (Blasius de Pelacanis or Biagio Pelacani da Parma) was born on an uncertain date in Costamezana (Parma), but is known to have received his doctorate at Pavia in about 1374. He died in Parma and was buried there on 23 April 1416. He visited Paris during his youth, sometime before 1388, but spent most of his life in Italy. At various times he taught logic, natural and moral philosophy, medicine, mathematics, and astrology (which included natural philosophy or physics) at the Universities of Bologna, Padua, Pavia and Florence. He was forced to retire from the University of Padua in 1411 on the grounds that he had no students, and was no longer fit to teach. He wrote a large number of works, including commentaries on Aristotle's *On the Soul* (*Conclusiones de anima* in 1382; *Quaestiones de anima*, copying finished in 1385) and *Physics* (*Quaestiones physicorum*, two manuscript versions, 1382-8 and 1397), *Quaestiones de motuum proportionibus (Questions on the Proportions of Motions)* (two manuscript versions, 1389 and 1407) and *Quaestiones perspectivae (Questions on Perspective)* (two versions, before 1390 and 1403). Many of his works are available only in manuscript (for details, see Federici Vescovini 1979: 413-52). His writings show the strong influence of English and French authors, including William of Ockham, William Heytesbury, the Oxford Calculators, John Buridan and Albert of Saxony, as well as of Arabic science, astrology and optics.

2 Metaphysics

In the context of the fourteenth century, Blasius' metaphysics is strongly original, for he was a materialist. He worked out a doctrine of substance as prime matter, an indeterminate substratum common to all beings. This metaphysical conception of reality as prime matter resulted in its being conceived as ungenerable, incorruptible and eternal. The forms of particular beings originate from prime matter in so far as it is constituted internally by qualities which are dispositions to become this or that individual form, such as human being or horse. With the death of the individual, these forms disappear by returning to prime matter. Prime matter, however, is only a conjoint cause of the generation of forms. The other universal cause of generation is the universal movement of the heavens which, through the influence of their motion, activate the dispositions of prime matter, and bring individual forms to birth.

3 Human beings

A corollary of Blasius' metaphysical views and his desire to unify all observable phenomena is his inclusion of humans among those natural beings whose generation is accounted for in terms of prime matter, individual dispositions and heavenly influences. For Blasius, even the intellective soul of human beings is a naturally generated form, a material power which results from prime matter disposed to receive it by the movement of the stars. In this Blasius shows himself not a follower of Averroes (see Ibn Rushd §3), but rather of Alexander of Aphrodisias (§2). This materialist account was reinforced by Blasius' epistemology, which led him to argue that the existence of the intellect could be known only through inference from observation of the intellect's operations, and that such observable operations could not be independent of the body and matter in general. As a result Blasius denied the immortality of the human intellective soul. He was reprimanded by the Bishop of Pavia in 1396, but the reprimand was a mild one, without financial consequences or any effect on his teaching, which he pursued without

hindrance until his retirement. Nonetheless he showed greater caution in his later writings, and made reference to the priority of faith.

Blasius' materialist metaphysics raised obvious difficulties for human freedom, since the doctrine that humans can make free choices seems incompatible with astrological causality. Blasius resolved this problem of individual freedom by arguing that the human contains dispositions to act in one way or another, even when confronted with two objects which are equally attractive to reason, and that in providing these dispositions the stars incline but do not necessitate. As a result the individual is always confronted with freedom of choice (*libertas differentiae contradictionis*), though it is the impulse and not rationality that is impartial in the face of a decision. Blasius added in his later writings that the stars are only secondary instrumental causes. God is the true first cause, though his activity is only known through the activity of his instrument, the heavens.

4 Epistemology and science

Blasius adopted the English empiricism of William of Ockham (§§4-5) and his followers as filtered through the French teaching of John Buridan (§3) and Albert of Saxony. He thus held that we know only the singular, present individual (Peter, for example) whom we perceive in visual cognition (*intuitio*) by means of a repeated sensible experience. Through this sensible experience, a general image is formed in the memory, and this is the universal concept of some individual (Peter viewed as this man). Thus the difference between the intuitive cognition of the singular (Peter) and the abstractive cognition (the concept of the man Peter) does not depend on the thing cognized, which will always be a singular thing (*res*), but on the psychological and epistemological processes of our mind. Since Blasius is a materialist, he appeals to the doctrine of the agent sense (see John of Jandun and John Buridan) which plays a role analogous to the one the agent intellect was held to play in intellection. Blasius also assigned a role to species in perception, interpreting them as sensible impressions which are material signs of the external objects, while concepts are representational signs of the cause of the impressions. Blasius owed this doctrine of signs to Buridan and Albert of Saxony.

Given the centrality of visual experience in the elaboration of his empiricist epistemology, it is not surprising that Blasius devoted an important treatise, *Quaestiones perspectivae*, to the problems of optics, including the geometric laws of direct vision, reflection and refraction. The work consists of questions on the earlier work of John Pecham, and owes a great deal to the influence of the Arab Alhazen (see Ibn Hazm; Optics §1). He was critical of the perspectivists of the thirteenth century (the *antiqui*) in his discussion of the problem of the appearance of visual magnitudes which, he claimed, depended on distance and not only on the angle of vision. This view was important for the construction of the visual pyramid which, in connection with the geometrical approach to the perception of the distance of visual objects, was to generate the pictorial representation of perspective.

In accordance with his generally rationalist formulations, Blasius offered explanations of interesting superstitions of his time, as did Pietro Pomponazzi (§3) in his work *De incantationibus (On Incantations)* (1556). Blasius explained many phenomena held to be miraculous or diabolical as products of optical illusion, due to the laws of refraction. For instance he showed that some apparently miraculous apparitions in the sky of Busseto, near Milan, were phenomena brought about by the refraction of light on clouds during a thunderstorm.

His treatises on statics (*De ponderibus*) and on mathematical physics, *De motuum proportionibus (On the Proportions of Motions)* are also of great importance, for in them he established for the first time the concept of the quantitative measure of an extended physical substratum, understood as an atom (*quantum*) which is indivisible, limited and finite. These *quanta* could be both material and spiritual, for the human intellect can be regarded as an indivisible quantum. Blasius' atomism places him firmly in the anti-Aristotelian tradition. It is also important for the rise of mathematical physics through its emphasis on what is measurable and quantifiable as opposed to the Aristotelian doctrine of qualities. In this context he used the notion of latitude, or a range of degrees, which he took from Nicole Oresme (§2), to explain qualitative variations in a material subject, but he also followed the tradition of the Italian school of Simone di Castello, Giovanni da Casale and Jacopo da Napoli (see Clagett 1968: 66-107). He reduced the so-called intension of forms to extension and mathematical measure.

Blasius' logic is related to his science in that he applies terminist analysis, particularly with respect to supposition theory, to physical doctrines. He wrote a commentary on Peter of Spain which shows the influence of William Heytesbury and of Henry Hopton's treatise (formerly attributed to Heytesbury) on the significate of true

propositions, as well as of Albert of Saxony.

Through his scientific reflection on the world of nature and a philosophy freed of any subordination to theology, Blasius opened the way to a formulation of a mathematical method which would come to fruition in the modern science of Galileo Galilei (§2).

See also: Aristotelianism, Renaissance; Logic, medieval; Natural philosophy, medieval

Translated from the original Italian by E.J. Ashworth
GRAZIELLA FEDERICI VESCOVINI

List of works

Blasius of Parma wrote a large number of works, many of which are available only in manuscript. Details of these can be found in Federici Vescovini 1979: 413-52 (see references and further reading).

Blasius of Parma (1382-5) commentaries on Aristotle's *On the Soul*, in G. Federici Vescovini (ed.) *Le Quaestiones de anima di Biagio Pelacani da Parma*, Florence: Olschki, 1974; Italian trans. by V. Sorge, Naples: Morano, 1994.(Editions of the manuscripts *Conclusiones de anima* (1382) and *Quaestiones de anima* (copying finished in 1385).)

Blasius of Parma (1386) *In quodam iudicio anno currente [13]86 (Prognostication of the Year 1386)*, in G. Federici Vescovini (ed.) *Rinascimento* 22 (1971): 90-3.

Blasius of Parma (before 1390-1403) *Quaestiones perspectivae libri tres (Three Books of Questions on Perspective)*, in G. Federici Vescovini and V. Sorge (eds) *Quaderni di 'Physis'*, Florence: Olschki, forthcoming; G. Federici Vescovini (ed.) 'Le questioni di "Perspectiva" di Biagio Pelacani da Parma', *Rinascimento* 12: 163-243.(Devoted to problems of optics, Blasius' *Quaestiones perspectivae* appeared in two versions, the first before 1390 and the second in 1403.)

Blasius of Parma (before 1416) *Tractatus de ponderibus (Treatise on Statics)*, in E.A. Moody and M. Clagett (eds) *The Medieval Science of Weights*, Madison and Milwaukee, WI, and London: University of Wisconsin Press, 1952, 238-78.

Blasius of Parma (before 1416) *Le Quaestiones dialecticae di Biagio Pelacani da Parma (Dialectical Questions by Blasius of Parma)*, ed. G. Federici Vescovini, J. Biard and V. Sorge, Florence, forthcoming.

Blasius of Parma (before 1416) *La 'quaestio de intensione et remissione formarum' di Biagio Pelacani da Parma (The 'Question of the Intension and Remission of Forms' by Blasius of Parma)*, ed. G. Federici Vescovini, *Physis: Rivista internazionale di storia della scienza* 31 (1994): 433-535.(Federici Vescovini gives an introduction and notes.)

Blasius of Parma (before 1416) 'Questioni inediti di ottica' (Unpublished Questions on Optics), in F. Alessio (ed.) *Rivista critica di storia della filosofia* 16 (1961): 79-110, 188-221.

References and further reading

Barocelli, F. (1992) 'Per Biagio Pelacani, un Convegno e un "Centro Studi"' (A Convention and 'Core Study' for Blasius of Parma), in G. Federici Vescovini (ed.) *Filosofia, scienza e astrologia nel Trecento Europeo*, Padua: il Poligrafo, 21-38. (Bibliography of manuscripts and editions on pages 183-216; includes an interview with Raymond Klibansky.)

Clagett, M. (ed.) (1968) *Nicole Oresme and the Medieval Geometry of Qualities and Motions: A Treatise on the Uniformity and Difformity of Intensities known as Tractatus de configurationibus qualitatum et motuum*, Madison, WI: University of Wisconsin Press.(Referred to in §3. Contains Latin text and English translation of Oresme's principal work on his doctrine of configurations. Clagett also provides a chronological ordering of Oresme's writings.)

Federici Vescovini, G. (1979) *Astrologia e scienza: La crisi dell'aristotelismo sul cadere del Trecento e Biagio Pelacani da Parma (Astrology and Science: Blasius of Parma and the Crisis of Aristotelianism at the End of the 14th Century)*, Florence: Vallecchi.(Referred to in §1. Detailed analysis of Blasius' thought.)

Federici Vescovini, G. (1983) *'Arti' e filosofia nel secolo XIV ('Arts' and Philosophy in the 14th Century)*, Florence: Vallecchi.(Contains several papers on Blasius.)

Lindberg, D.C. (1975) *A Catalogue of Medieval and Renaissance Optical Manuscripts*, Toronto: Pontifical Institute of Mediaeval Studies.(Very useful on Blasius' manuscripts on optics.)

Maier, A. (1949) *Die Vorlaüfer Galileis im 14. Jahrhundert (Galileo's 14th-Century Predecessor)*, Rome: Storia e Letteratura.(Discusses the probable causes of the episcopal condemnation.)

Murdoch, J.E. (1976) 'Music and Natural Philosophy: Hitherto Unnoticed *Questiones* by Blasius of Parma',

Manuscripta 20: 119-36.(A well-founded attribution of work on music to Blasius.)

Bloch, Ernst Simon (1885-1977)

Bloch was one of the most innovative Marxist philosophers of the twentieth century. His metaphysical and ontological concerns, combined with a self-conscious utopianism, distanced him from much mainstream Marxist thought. He was sympathetic to the classical philosophical search for fundamental categories, but distinguished earlier static, fixed and closed systems from his own open system, in which he characterized the universe as a changing and unfinished process. Furthermore, his distinctive materialism entailed the rejection of a radical separation of the human and the natural, unlike much twentieth-century Western Marxism. His validation of utopianism was grounded in a distinctive epistemology centred on the processes whereby 'new' material emerges in consciousness. The resulting social theory was sensitive to the many and varied ways in which the utopian impulse emerges, as, for example, in its analysis of the utopian dimension in religion.

1 Life and works

Ernst Simon Bloch was born in Ludwigshafen, Germany. He studied at the Universities of Munich and Würzburg, and had associations with Simmel's private colloquium in Berlin and Weber's circle in Heidelberg. For much of his adult life he lived as a freelance writer and critic. Like many left-wing Jewish intellectuals he left Germany on Hitler's accession to power in 1933, and eventually moved to the USA. In 1948 he received his first university post, as Professor of Philosophy at Leipzig, East Germany. Although initially sympathetic to East German regime, he eventually came into conflict with it, and in 1961 did not return from a visit to West Germany. Subsequently he became Professor of Philosophy in Tübingen, where he died. His most notable works are *Geist der Utopie (Spirit of Utopia)* (1918) and *Das Prinzip Hoffnung (The Principle of Hope)* (1954-9).

2 Metaphysics and ontology

Throughout his working life, Bloch attempted to refurbish the ancient category of 'matter' to provide a dynamic materialist theory of the universe (see Materialism). Matter, the substance of the universe, is not inert, but rather, in nature and humanity, capable of movement and development. In this respect Bloch distances himself from much twentieth-century Western Marxism, which seeks to prioritize the social over the natural (see Marxism, Western §§2-3). He credits the 'Aristotelian Left' (notably Avicenna and Averroes) with the earliest formulation of this radical materialism (see Averroism). The universe is open-ended and rich with 'possibility', allowing development, novelty, intervention and alternative outcomes to occur. From a human perspective, this open-endedness is both an opportunity and a constraint, for the possibility of choice is beset by the hazards of uncertainty; human life is therefore experimental. Bloch attempts to resolve the Marxist conundrum of the relationship between freedom and necessity. He is at pains to avoid ultra-subjectivity or hyper-voluntarism: possibility can take the form of the nonsensical and the impractical, but rational movement will take the form of the 'objectively-real possible' - the possibility grounded in existing social tendencies. Aspiration is thus combined with a non-positivist empiricism, which he terms 'process empiricism'. It is thus to be distinguished from a determinist teleology, which, following the analogy of the acorn and the oak, envisages humanity in terms of a fixed, inevitable future growing out of iron historical trends.

A fundamental way in which Bloch conceptualizes the dynamic tension at work within matter is by building up composite concepts around the category of 'not'. 'Not' is meant to register the absent and unfinished dimensions of reality, as in his logical proposition that 'S is not yet P'. In the human world the 'not' is present in the form of need. Absence, initially of food, drives the individual on to more and more sophisticated forms of interaction with nature and society. Humans are thus 'not-yet'; the completion of their being lies in the future, they constantly hunger for themselves. 'Not' is the negative aspect of the historical process, 'Hope' is the positive. 'Subjective' hope is to be distinguished from 'objective'; the former involves the perpetual and ubiquitous representation of that which is deemed to be absent. Bloch's massive *Das Prinzip Hoffnung* (1954-9) contains an encyclopedic account of the many manifestations of hope in human history, from simple daydreams to complex visions of perfection. Objective hope is the concrete possibility present in each successive age, which enables subjective hope actively to develop the world (see Hope).

3 Epistemology

Bloch conceives of consciousness as a narrow field or band. Beyond the lower boundary lies the 'no-longer-conscious', the realm of the forgotten and the repressed, explored by Freud. The upper boundary delimits the 'not-yet-conscious', the place where new material enters consciousness, 'the psychological birthplace of the New'. It is the unexplored territory that Bloch sought to map, and which makes up his central epistemological category. The production of new material occurs through the stages of incubation, inspiration and explication. Incubation is the period of active fermentation of the new, much of it below the surface of consciousness, to a point where it bursts into the conscious world. This is the moment of inspiration, a sudden lucid moment of illumination. Bloch is keen to anchor the epistemological in the social. The newness, which he terms the 'Novum', emerges with the confluence of subjective and objective conditions. The historical timetable generates the material which is incubated in the individual, and the inspiration is as much historical as individual. A Marxist sense of history informs this interpretation; progressive classes are the fundamental fact in the emergence of the Novum. The initial entry, however, occurs through immensely gifted individuals, such as Marx. The realization of newness requires the third stage in the process - explication. This is the immensely difficult task of adequately representing the new, such that it re-enters the historical timetable as immanent potentiality. It also involves overcoming the resistances of the existing world to novelty. Bloch calls the site at which present and future meet the 'Front'. Like the military use of this word, it is meant to suggest an advancing, although not necessarily straight, line into as yet unconquered territory.

4 Utopianism

Nowhere is Bloch more untypically Marxist than in his self-conscious utopianism. He distinguishes 'abstract utopia' from 'concrete utopia'. The former manifests the utopian function in its weakest form; it is mere dreaming, unanchored in the real tendencies of the age. Concrete utopia, in contrast, is rooted in objective possibility; it is grounded in the ascending forces of the age, and is the most pregnant form of the utopian function (see Utopianism). Bloch's unorthodox Marxism is also apparent in his attempt to harness aspects of religion to his revolutionary project. In *Atheismus im Christentum* (*Atheism in Christianity*), he argues that religious consciousness has been a potent vehicle for utopian longings and asserts the paradoxical claim that 'only an atheist can be a good Christian; only a Christian can be a good atheist' ([1968] 1972: 9) (see Atheism). He is sensitive to the temporal aspects of the utopian impulse, and maintains that individuals may be contemporaries in a physical sense but not necessarily in terms of forms of consciousness - 'not all people exist in the same Now'. In the 1930s he deployed the term 'non-contemporaneity' to designate this experience, and cited the peasantry as an example. Marxism had been blind to the subversive and utopian elements in non-contemporaneity, and had left these strata to be co-opted and distorted by fascism.

Also out of joint with much twentieth-century Marxism is Bloch's radical materialist concept of a 'natural subject'. Given that matter is active and dynamic, and that developed consciousness has emerged out of it in the shape of humanity, is it not possible to conceive of further creative development in the realm of nature? Furthermore, since the human cannot be radically separated from the natural, Bloch conceives of a creative interaction of the two in the future. Nature too has its 'not-yet'. The complexity of such a universe, Bloch argues, requires the abandonment of classical conceptions of time and space. He is attracted to the non-Euclidian conception of space developed by Riemann, in which space is altered by local variables. He proposes, by analogy, 'a kind of "Riemannian" time' which is neither universal nor unilinear, but rather contains a plurality of differing time scales. Only in a future, interactive reconciliation between humanity and nature might convergences of time occur.

Bloch's work tends to be known about rather than known. His writing is forbidding: structural complexity and formal eclecticism are combined with a style studded with opaque metaphor, untranslatable puns, obscure neologisms and overblown rhetoric. Furthermore, the political cast of much of the work is a deeply unattractive, unreconstructed Marxism-Leninism. Apart from his youthful dialogue with the young Lukács, he has had little influence on other major figures in twentieth-century Marxism, and none at all on the pre-eminent philosophical currents of the century. There is, however, growing interest in his work. Radical theologians have long valued his insightful analyses of the utopian dimension in religion; literary theorists are increasingly using his work in their studies of utopias and science fiction; whilst social and political theorists have found his utopian socialism deeply stimulating.

VINCENT GEOGHEGAN

List of works

Bloch, E. (1918) *Geist der Utopie (Spirit of Utopia)*, Frankfurt: Suhrkamp Verlag, 1971.(Bloch's first important published work combines messianic utopianism with wide-ranging cultural criticism.)

Bloch, E. (1935) *Erbschaft dieser Zeit*, Frankfurt: Suhrkamp Verlag, 1962; trans. N. Plaice and S. Plaice, *Heritage of Our Times*, Oxford: Polity Press, 1971.(Contains a highly distinctive analysis of national socialism, focusing on the dreams and fantasies of Weimar Germany.)

Bloch, E. (1954-9) *Das Prinzip Hoffnung*, Frankfurt: Suhrkamp Verlag, 2 vols; trans. N. Plaice, S. Plaice and P. Knight, *The Principle of Hope*, Oxford: Blackwell, 3 vols, 1986.(Bloch's undoubted *magnum opus*, containing an encyclopedic presentation of the manifold forms of the utopian.)

Bloch, E. (1961) *Naturrecht und menschliche Würde*, Frankfurt: Suhrkamp Verlag; trans. D.J. Schmidt, *Natural Law and Human Dignity*, Cambridge, MA: MIT Press, 1986.(An analysis of the positive and negative dimensions of the European natural law tradition.)

Bloch, E. (1963-4) *Tübinger Einleitung in die Philosophie*, Frankfurt: Suhrkamp Verlag, 2 vols; vol. 1, trans. J. Cumming, *A Philosophy of the Future*, New York: Herder & Herder, 1970.(A succint presentation of many of the major components of Bloch's philosophy.)

Bloch, E. (1968) *Atheismus im Christentum*, Frankfurt: Suhrkamp Verlag; trans. J.T. Swann, *Atheism in Christianity*, New York: Herder & Herder, 1972.(The most sustained presentation of Bloch's attempt to combine atheism and Christianity.)

References and further reading

Geoghegan, V. (1996) *Ernst Bloch*, London: Routledge.(A basic introduction to the main ideas and themes.)

Harper, C. (1991) 'Ernst Bloch: A Bibliography of Primary Sources in English', *Bloch-Almanach*, vol. 12, 167-80.(A valuable guide to the English translations of Bloch.)

Hudson, W. (1982) *The Marxist Philosophy of Ernst Bloch*, London: Macmillan.(The best study to date of Bloch's thought.)

Jones, J.M. (1995) *Assembling (Post)modernism: The Utopian Philosophy of Ernst Bloch*, New York: Peter Lang. (Interesting attempt to relate Bloch's work to post-modern theory.)

Roberts, R.H. (1990) *Hope and its Hieroglyph: A Critical Decipherment of Ernst Bloch's Principle of Hope*, Atlanta, GA: Scholars Press.(Concentrated study of *The Principle of Hope*, focusing on the religious dimensions.)

West, T.H. (1991) *Ultimate Hope Without God: The Atheistic Eschatology of Ernst Bloch*, New York: Peter Lang. (General discussion of Bloch's work, ultimately concentrating on Bloch's 'religious' vision.)

Bobbio, Norberto (1909-)

The foremost legal and political theorist in Italy today, Norberto Bobbio founded in the 1940s Italian analytical legal positivism, trying to merge logical positivism and Kelsen's legal positivism. As a political thinker, he defends a synthesis of liberalism and socialism, focusing in particular on the defence of human and civil rights in democratic societies.

Norberto Bobbio taught Jurisprudence and Political Philosophy at the universities of Camerino, Siena, Padua and Turin for over forty years. After the Second World War, he founded the school of Italian analytical legal positivism. Bobbio took part in the Resistance. After the Liberation he withdrew from direct political involvement, still, however, taking an active part in Italian political debate. He defends a tolerant, secular point of view and a synthesis of liberalism and socialism. In 1984 he was made a Life Member of the Italian Senate.

Bobbio's original outlook was largely formed by Croce's historicism in philosophy, and Gobetti's radical liberalism in politics (see Croce, B.). Yet, while Bobbio's political thought has always been inspired by Gobetti's project of reconciling social justice with individual civil and political liberty, Crocean idealism he soon rejected. He examined the phenomenology of Husserl and Scheler as well as the existentialism of Heidegger and Jaspers. In the mid 1940s he finally abandoned these philosophies to embrace a very different approach, influenced by logical positivism, which he saw as simultaneously rigorous, rationalistic, empirical, and ethically and politically committed (see Vienna Circle §2) - and for these reasons suitable to his own epistemological preferences, which have always been against the trends of what he has called the 'Italian ideology', that is congenitally speculative and idealistic in bias.

These developments culminated in Bobbio's founding of Italian analytical legal positivism. It is worth emphasizing what kind of legal positivism it was, because Italian analytical legal positivism has been an eclectic but fruitful attempt to graft a philosophical outlook - logical positivism - on to the legal positivism of Hans Kelsen's *Pure Theory of Law* (see Kelsen, H.; Logical positivism §2).

The first phase of Bobbio's theory of law (1949-65) was marked by his acceptance of Kelsen's interpretation of legal theory as a scientific, value-free form of legal study not concerned with the moral or political evaluation of law, nor with the sociological description of legal phenomena, but with the analysis of fundamental legal concepts and with the structure and logical interrelation of the elements of a legal system. As a consequence, Bobbio's main concern was to clarify fundamental legal notions such as 'legal rule' and 'legal system' as well as to give an account of the structure of legal orders.

Bobbio defended a version of normativism (the opinion that law is made up of rules), but answered many of the criticisms raised against this approach by elaborating a theory of legal systems, which successfully addressed many problems that could not be solved analysing single legal rules. So, in his *Teoria dell'ordinamento giuridico (A Theory of Legal System)* (1958), Bobbio contended that the definition of law, as well as the distinction of law from other normative phenomena such as morals and customs, is possible only if the legal system is taken into consideration. There are no special features belonging to all legal rules and only to legal rules. In fact, it is the legal system as a whole that has identifying features such as effectiveness or coerciveness. A rule is legal because it belongs to a legal system; a system is legal because of the specific characteristics it has as a system (see Norms, legal).

This account of legal systems allowed Bobbio to acknowledge the existence of different types of legal rules. He refused Kelsen's reduction of all legal norms to duty- or sanction-imposing rules, taking into consideration the wide class of second-level rules (meta-rules), such as power-conferring rules, constitutive rules and so on. The classification of different kinds of legal rules and the description of their interrelationship within legal systems has been one of Bobbio's main contributions to legal theory.

Towards the mid 1960s, Bobbio came to a turning point in his interpretation of legal theory. It became clear that the two basic neo-positivistic philosophical assumptions at the root of Bobbio's outlook, namely the theory of discourse levels and the distinction between 'is' and 'ought' statements, were not consistent with Kelsen's interpretation of legal theory as a scientific endeavour. First, Bobbio distinguished between legal theory

(jurisprudence) and the discussion of the method of legal theory (meta-jurisprudence). Second, he criticized Kelsen's meta-jurisprudence as prescriptive - thus not scientific at all, as it did not aim at describing what jurists actually do but at prescribing what they should do. Third, Bobbio argued that Kelsen's doctrine of the basic norm, which gives unity and validity to a legal order, must rely on an ideological rather than a logical ground and so cannot be the basis for a value-free science of law.

In this second phase (from 1965 onwards) Bobbio thus acknowledged the prescriptive nature of the legal positivist's approach to law. Such an approach is not based on the desire to elaborate a scientific, value-free legal theory. Rather, the idea of a scientific description of the law is maintained by legal positivists because such an idea is logically required by the very notion of applying the law, which is central to the working of legal and political institutions based on the rule of law. Thus, the legal positivists' argument in favour of an objective description of the law simply shows their choice of the values of the rule of law (see Legal positivism §2).

Since the 1970s, Bobbio has on the one hand developed a sociological theory aimed at describing the social functions of law; on the other hand he has increasingly devoted his studies to political theory. In his functional analysis of law, Bobbio has focused on the 'promotional' function played by legal orders of developed countries, by stimulating desirable behaviour, mainly in economic and business activity, through positive sanctions such as subsidies, tax exemptions and so on. This function is one of the characteristic features of the welfare state, defended by Bobbio, as opposed to the liberal minimal state.

In fact, Bobbio's liberalism is basically a doctrine of constitutional guarantees for individual freedom and civil rights, not an economic theory of the free market. Bobbio sees no contradiction between liberalism and democracy. The basic rights - freedom of opinion, speech, association, and so on - on which the liberal state has been founded since its inception are the premises of the democratic state: if liberalism provides those liberties necessary for the proper exercise of democratic power, democracy guarantees the existence and persistence of fundamental liberties.

Bobbio's main concern in political theory has always been that of reconciling this understanding of liberal democracy with the demands of socialists for greater equality. He suggests that socialists must rethink their goals of social equality in ways compatible with the institutional framework of liberal democracies. Representative democracy has to be seen as a set of rules that cannot be given up if the risk of despotic regimes is to be avoided. Social rights, as the extension of civil and political rights, will be granted through the extension of representative democracy to the level of social life - to bureaucracies, to health and educational authorities, to the workplace and so on.

Bobbio has consistently played a part in active political debate in many fields: politics and culture, the defence of human and civil rights, the problem of peace in the nuclear age. A major contribution to political theory has been to show the strict link between human rights, peace and democracy.

See also: Democracy; Law and morality; Law, philosophy of; Liberalism

PATRIZIA BORSELLINO

List of works

Bobbio, N. (1938) *L'analogia nella logica del diritto (Analogy and Legal Logic)*, Turin: Istituto Giuridico della Reale Università.(The first important contribution by Bobbio to the analysis of legal reasoning.)

Bobbio, N. (1950) *Teoria della scienza giuridica (A Theory of Legal Science)*, Turin: Giappichelli.(States that legal theory is a value-free form of legal study, on the basis of a neo-positivistic account of science.)

Bobbio, N. (1955a) *Studi sulla teoria generale del diritto (Studies of General Theory of Law)*, Turin: Giappichelli.(Through a critical examination of the main positions in legal theory, Bobbio comes to accept Kelsen's interpretation of normativism.)

Bobbio, N. (1955b) *Politica e cultura (Politics and Culture)*, Turin: Einaudi.(Contains many important writings in defence of a synthesis of liberalism and socialism and about the political role of culture.)

Bobbio, N. (1958) *Teoria della norma giuridica (A Theory of Legal Norm)*, Turin: Giappichelli; *Teoria dell'ordinamento giuridico (A Theory of Legal System)*, Turin: Giappichelli, 1960; new edn of the two books together, *Teoria generale del diritto (General Theory of Law)*, Turin: Giappichelli, 1993.(A wide and uncommonly clear examination of the main problems of the philosophy of law, from the point of view of Kelsen's critically revised normativism.)

Bobbio, N. (1965) *Giusnaturalismo e positivismo giuridico (Natural Law Theory and Legal Positivism)*, Milan:

Comunità.(Develops and defends a normativist version of legal positivism and criticizes the natural law theory.)
Bobbio, N. (1976) *Quale socialismo?*, Turin: Einaudi; *Which Socialism?*, Oxford: Polity Press, 1987.(An examination of Marxist writings on the state and democracy, and the rethinking of socialism in liberal-democratic terms.)
Bobbio, N. (1977) *Dalla struttura alla funzione (From Structure to Function)*, Milan: Comunità.(The major contributions of Bobbio to a sociological theory aimed at describing the social functions of law.)
Bobbio, N. (1984) *Il futuro della democrazia*, Turin: Einaudi; *The Future of Democracy*, Oxford: Polity Press, 1987.(Defends the liberal idea of representative democracy, trying to extend it beyond parliamentary politics to many domains of social life.)

References and further reading

Borsellino, P. (1991) *Norberto Bobbio metateorico del diritto (Norberto Bobbio as a Meta-theorist of Law)*, Milan: Giuffrè.(A critical examination of Bobbio's contribution to methodology and general theory of law.)
Jori, M. (1987) *Il giuspositivismo analitico italiano prima e dopo la crisi (Italian Analytical Legal Positivism before and after its Crisis)*, Milan: Giuffrè. (A critical examination of Italian analytical legal positivism.)
Lanfranchi, E. (1989) *Un filosofo militante (A Committed Philosopher)*, Turin: Bollati Boringhieri.(An account of Bobbio's opinions about the relationship between politics and culture.)
Ruiz-Miguel, A. (1983) *Filosof'a y derecho en Norberto Bobbio (Philosophy and Law in Bobbio's Thinking)*, Madrid: Centro de estudios constitucionales.(A wide account of Bobbio's legal and political philosophy.)
Violi, C. and Maiorca, B. (1984) *Norberto Bobbio: 50 anni di studi (Norberto Bobbio: 50 Years of Studies)*, Milan: Angeli.(A general bibliography of writings by and about Bobbio.)

Bodily sensations

Bodily sensations are those feelings, or sensory experiences, most intimately associated with one's body: aches, tickles; feelings of pain and pleasure, of warmth, of fatigue. Many philosophers contrast bodily sensations with perceptions of the external world, claiming that sensations provide one with awareness of nothing independent of them. An alternative approach is to take sensations to be a form of awareness of one's body - on one view sensations are simply the perception of the state and properties of one's body. Bodily sensations have been seen as a major problem for any attempt to give an account of the mind that takes it to be part of the material world as investigated by the physical sciences.

1 Bodily sensations and the body

The most intimate causes of bodily sensation are changes or states within one's body: dental decay leads to toothache; general exhaustion to a feeling of fatigue. Bodily sensations are distinctive in acting as signs for the bodily changes that cause them. Pains are signs of damage to or disorder in one's body; an agent will react to feelings of warmth or cold in a way appropriate to the change in temperature of their body; feelings of hunger or satiation will control someone's feeding. Yet sensations are not invariably correlated with their typical bodily causes: someone can feel hunger when they are satiated; pains in parts of the body that are not damaged. This is most commonly the case with referred pain, where damage to one part of the body leads to feelings of pain in some other part.

One typically feels bodily sensations to have locations within one's body. One may feel a pain in one's ankle or a ticklish sensation behind one's knee. Sensations can also fill or suffuse body parts, such as a burning sensation throughout one's upper arm, or a feeling of being bloated in one's abdomen. Location in the body may be more or less determinate - a pin prick may be felt at a particular point on one's finger, while an erotic sensation may lack a precise location. Some sensations are associated with the body without having a location, such as feelings of fatigue or depression. Even sensations which normally have a location may sometimes lack one: certain patients who have lost general feeling in a limb can be brought to have 'deep pain' without knowledge of where it is.

With some sensations we can draw a distinction between the sensation and what it is a sensation of. One's body may actually be warm when one does not feel it to be so - while one's body may feel cold even when it is not, as when one has a fever. The same is true of hunger, satiation and exhaustion. For other sensations there is no such separation between a feeling and what is felt. It may seem obvious that there cannot be a pain without a feeling of pain, and that there is no such thing as false pain, the feeling of pain without actually being in pain. In the former kind of case, we talk of sensations of warmth, but not warm sensations; the term 'sensation' is used to pick out the state of feeling rather than any object of that feeling. In the latter kind of case, we use the term 'sensation' not only for the feeling, as in the phrase 'a sensation of pain', but also for the object of that feeling, as when we say that pain is a sensation.

However, some philosophers think that even in the latter case, a feeling and its object are distinct. It has been claimed that one can have pains without feelings of pain, since there can be unconscious pains, as when one only gradually comes to notice a pain in one's thumb. And conversely, it has been claimed that one can have feelings of pain without the accompanying pain: someone in a confused state might come to mistake a feeling of intense cold for a sharp pain. Neither example offers a conclusive counterexample. At best, they demonstrate the gap between having a feeling and making a judgment about it. In the former example, unless it is supposed that one can be conscious only of what one notices, all that has been described is someone who does not notice the pain in their thumb; this is not pain without consciousness, only pain without attention to it. The latter example can be interpreted as one in which someone mistakes a feeling of extreme cold for a feeling of pain. This may challenge a claim about the epistemology of sensations - that one has incorrigible knowledge of them - but not the link between feeling and its object.

2 Sensations as subjective

The close link between feeling and its object lends itself to one account of the nature of bodily sensation, which may be labelled 'subjectivist'. On this view, sensations are to be contrasted with sensory states such as

perceptions, or intentional states such as beliefs that can take external objects, as in beliefs about the sun. One can have a perception of an object that exists whether one perceives it or not - the table I see will still be there when I turn my back. I can also have a sensory experience when there is no appropriate object for me to perceive: I may hallucinate a table. By contrast, bodily sensations may be claimed to be purely subjective states of mind, which are directed on no object at all, or none that is external or independent of them. This view of sensation may, but need not, be associated with a more general account of sense perception, on which all perception of the world involves the subject's having a purely subjective state, a sensation (see Perception). On such a view of perception, bodily sensations will just be one special case of sensation in general. The view does not so obviously apply to bodily sensations such as feelings of warmth or fatigue where what is felt does have a nature independent of the feeling. But for these, subjectivists may appeal to a distinction between felt warmth or fatigue, which will be taken to be a subjective quality, and the objective qualities of heat or exhaustion associated with those felt qualities. In support of this, they may point out how it is imaginable that different objective qualities could have been typically responsible for feelings with those felt qualities - for example, that the feeling of warmth could have been typically brought about by tickling the skin.

Subjectivism can take different forms. One approach, often labelled 'the act - object' view of sensation, retains the idea that there is a genuine object of awareness in having bodily sensations, even if that object depends for its existence on the subject's awareness of it. Others repudiate the idea that there are any such objects of awareness and insist that to feel a pain is to be in pain, denying that there is any distinction to be made between the feeling of pain and the pain felt. For one may be suspicious of the idea that genuine objects could be dependent on our awareness of them, and doubtful that such things could have a location in the world. In place of a feeling of pain and a pain felt, this 'no-object' view talks of feeling in a painful manner, or being in a painful state. This is rather like insisting that there are not such things as individual dances: when we say that Mary danced a polka, we do not mean that there was a certain thing, a polka, which she danced, but rather that she danced in a certain manner. So, too, one may insist that to say that Mary felt a pain is not to talk of some thing of which she was aware, but to talk of the manner in which she felt, or to talk of the character of her feeling (see Mental states, adverbial theory of).

Much of our ordinary talk about sensations, in particular the way in which we can locate them in the body, is resistant to these ways of eliminating the objects of awareness. If states of feeling are to be located anywhere, it is plausible to locate them only in the brain and central nervous system, but one can feel a pain in one's ankle, or a crick in one's neck. Since it is not the feeling that is located in one's ankle, the 'no-object' view must deny that we speak literally when we say that the ankle hurts. What then does it mean to talk of a pain in one's ankle? One answer would be to suppose that there are certain non-spatial qualities of one's sensations which one associates with spatial vocabulary. This idea is not particularly plausible: for there seems to be no prior limit to the exact positions on or in one's body where one could feel a sensation to be located; and one can feel a sensation to be located in a place where one has never had a sensation before. In addition, the felt location of a sensation has a different status to the non-spatial qualities of a sensation. Take any feeling in your left hand that you like - an itch, a tingle or whatever - and now imagine a qualitatively identical sensation in your right hand. You need imagine no difference in the character of the sensation other than its location in one hand rather than the other. The felt location of sensation seems to play a similar role to spatiotemporal location for physical things: qualitatively identical but distinct objects can exist at the same time if they are in different locations; qualitatively identical but distinct sensations can be had at the same time if they are felt to be in different parts of one's body.

An alternative is to suppose that the attributed location of a sensation does not report any quality of the sensation at all. Perhaps talk of location reflects the subject's dispositions to act towards or talk about the location to which the sensation is referred. Someone who feels pain in an ankle may be inclined to rub that ankle, or say that that is the part of them which hurts - on this view, what it is for one to feel a pain to be located in one's ankle rather than anywhere else just is for one to be disposed to rub that ankle or say that that is where the pain is. But are there really any such definitional connections between felt location of pains and dispositions to behaviour? Someone who is paralysed will have no disposition to rub their ankle; and someone might be unbothered by pain and not inclined to answer any questions about whether they have a pain. Furthermore, we can explain why someone feels concerned about their ankle by reference to the fact that they feel pain and the pain they feel is in their ankle - both the quality and the location of sensation explain the subject's attitude. So it is implausible that we could simply explain away the location of sensations in terms of the non-linguistic and linguistic behaviour for which they give

reasons.

3 Awareness of one's body and sensations as perceptions

If we have to take seriously the felt location of sensations, then we also have to take seriously things being felt at those locations, and the elimination of an act - object approach to bodily sensations will consequently seem unattractive. This need not, however, mean a return to the idea of pains and itches as inner objects of mind. One may instead claim that in having a sensation one comes to be aware of one's body. When I feel a pain in my ankle, what I feel is that my ankle hurts in some way: the felt location of sensation is the body part that one is aware of in having the sensation. When we count pains, we count the number of parts of the body which hurt. One's ankles, toes or teeth are no less part of the objective world than are tables and chairs, so bodily sensations cannot be purely subjective states of mind which give one awareness of nothing independent of them. Rather, having sensations gives one an awareness of an item in the objective world: one's body.

What is it to say that sensations give one an awareness of one's body? Some philosophers have claimed that sensations are just a form of perception of one's body. This view is difficult to defend. For other kinds of perception, there is a contrast between the qualities perceived and the perceiving of them. A book may be square without looking square; and it may look square without actually being so. One can also have visual or tactual hallucinations, as when Macbeth 'sees' a dagger; or a disturbed individual may feel insects on their arm when nothing is there. It is more difficult to establish this separation in the case of sensations. Phantom-limb sensations are the best candidates for the example of bodily illusions or hallucinations, since the amputee feels hurt in a body part which no longer exists. But one cannot take feeling pain simply to be the perception or apparent perception of damage or disorder to one's body: a body part does not have to be damaged in order to hurt. Referred pain is not a form of illusion: we do not suppose that the pain is somehow illusory, and lose concern for the part of our anatomy which hurts; rather we distinguish between where we feel pain and the cause of that pain. Nor do we suppose that a damaged but anaesthetized limb is really hurting - hurt has to go with the feeling of hurt.

Second, even if some aspects of sensation can be treated as perceptual, there are other aspects of what our sensations are like which are not plausibly ascribed to any aspect of the state of our bodies: when I feel a nagging toothache, or a burning sensation, there does not seem to be any objective feature of my tooth or my arm of which I am aware in having that sensation. In both cases the characteristic seems to belong to my state of awareness and not my body as an object of awareness. Of course the fact that I have a burning sensation may tell a medic much about the state of my body, but that is not to say that I am aware of the change in state of my body through having the sensation.

The implausibility of a purely perceptual model of bodily sensation need not lead us to reject entirely the idea that sensation is a form of primitive awareness of one's body that gives one an intimate link to it and through which one knows it and has feelings about it in a way that one has for no other object in the world. Supposing that sensation is a form of awareness of one's body fits very well the physiological and psychological links between sensation and kinaesthesia and proprioception - the sense we all ordinarily have of the movement and position of our limbs. It also fits another feature of sensations: that, primarily, we feel the location of sensations to be locations within the body, and not within other parts of space. This is true even for persons having a phantom-limb pain - they do not feel the pain to be located out in empty space, but feel it to be in an apparent limb which extends beyond the point of amputation.

See also: Behaviourism, analytic; Mind, identity theory of; Intentionality; Private language argument; Qualia; Sense-data

M.G.F. MARTIN

References and further reading

Anscombe, G.E.M. (1957) *Intention*, Oxford: Blackwell, esp. 13-14, 49-51.

Anscombe, G.E.M. (1962) 'On sensations of position', *Analysis* 22: 55-8.(Anscombe rejects the view that we observe our sensations, and suggests a dispositional account of sensation location.)

Armstrong, D.M. (1962) *Bodily Sensations*, London: Routledge & Kegan Paul.(This is a detailed defence of the view that sensations are perceptions of one's body.)

Descartes, R. (1641) *Meditations on First Philosophy*, in *The Philosophical Writings of René Descartes*, J.

Cottingham, R. Stoothoff, D. Murdoch (eds), Cambridge: Cambridge University Press, 1984.(The sixth meditation contains an important discussion of the function of sensations, and the intimate awareness each of us has of our own body.)

Jackson, F. (1977) *Perception*, Cambridge: Cambridge University Press, ch. 3.(A defence of the view that there are inner objects of awareness.)

O'Shaughnessy, B. (1980) *The Will*, Cambridge: Cambridge University Press, vol. 1, chaps 5-7.(A detailed account of the need to attribute bodily locations to sensations, and of the special kind of awareness we have of our own bodies.)

Pitcher, G. (1970) 'Pain Perception', *Philosophical Review* 79: 368-93.(Another defence of the view that bodily sensations are perceptions.)

Reid, Thomas, (1764) *An Inquiry into the Human Mind on the Principles of Common Sense*, repr. Indianapolis, IN: Hackett Publishing Company, 1983, ch. 5.(An account of the role of bodily sensations in outer perception.)

Wittgenstein, L. (1952) *Philosophical Investigations*, Oxford: Blackwell, §242-315.(The famous private language argument, which some philosophers take to be an attack on the idea of sensations as inner or private objects.)

Wittgenstein, L. (1969) *Blue and Brown Books*, Oxford: Blackwell, 2nd edn, 49-55.(A discussion of the location and ownership of sensations.)

Bodin, Jean (1529/30-96)

Jean Bodin was one of the great universal scholars of the later Renaissance. Despite political distractions, he made major contributions to historiography and the philosophy of history, economic theory, public law and comparative public policy, the sociology of institutions, as well as to religious philosophy, comparative religion and natural philosophy. Among his most celebrated achievements are his theory of sovereignty, which introduced a new dimension to the study of public law, and his Neoplatonist religion, which opened new perspectives on universalism and religious toleration.

Many of these intellectual positions, moreover, were responses, at least in part, to great political issues of the time. Against doctrines of popular sovereignty and the right of resistance put forward in the course of the religious wars, Bodin sought to show that the king of France was absolute. Against the widespread corruption and laxity that weakened and undermined the monarchy, he argued for administrative reform. And against the party that pressed the king to impose religious uniformity, he cautiously supported religious toleration. In all these respects Bodin's thought helped to inform the policies of the early Bourbon dynasty esatblished by Henry IV.

1 Life

Bodin was born in Angers, France, into a modestly successful middle-class family. Obtaining an excellent humanist education in his youth, he was to become in the course of his career one of the most outstanding humanist scholars of his age. His erudition was formidable in scope as well as depth, and his many publications include important contributions to almost every field of learning pursued by his contemporaries.

Bodin's professional training was in law which he studied at the University of Toulouse during the 1550s. Unable to secure a regular faculty position, he embarked on a public career as a barrister in the *Parlement* of Paris. However, at no point in his career did he retreat from the encyclopedic programme of research and writing that he had projected in his early years. The success of his earlier publications on history (1566), public finance (1568), and public law and policy (1576) gave him access as adviser and confidant to high political circles in and around the royal court. In the 1570s he was often a dinner companion of Henry III, and he became counsellor to Henry's ambitious younger brother, the Duke d'Alençon. But in 1576, as a deputy to the Estates General of Blois, Bodin took a public-spirited if impolitic stand against requests for new taxation. He was thus at odds with Henry's policy and no longer enjoyed royal favour. The death of d'Alençon in 1584, marked the end of Bodin's involvement in high politics. He moved from Paris to Laon where he was a royal magistrate from 1587 until his death in 1596.

These last years were a time of trouble for Bodin. The French Religious Wars, which had gone on intermittently since 1562, were now entering a climactic phase. With the assassination of Henry III in 1589, a savage struggle had broken out over the claims to succession of the Protestant Henry of Navarre (King Henry IV). Large parts of France, including Laon, came under the control of the militantly revolutionary Catholic League, whose programme and doctrine contradicted Bodin's long-standing principles of legitimacy, non-resistance and religious tolerance. Yet Bodin, like many other royalist magistrates of the time, openly collaborated with the League. He sought to justify his course by mystical reflections on the preordained doom of the ruling dynasty. But he seems to have been driven by fears not only for his office and his property, but perhaps for his life as well; now, as in the past, he was under suspicion of heresy. He stood publicly for Navarre only in 1594 when the forces of the latter were victorious.

These troubles notwithstanding, Bodin never ceased to pursue his vast programme of scholarly and philosophic research. Between 1588 and his death he produced two short works on ethics, a major treatise on religion and a system of natural philosophy. His writings make it clear that Bodin's religion was a Judaizing Neoplatonism. But outwardly at least he remained within the church, and on his death he was buried as a Catholic in accordance with his will.

2 Public law - the theory of sovereignty

Bodin's most celebrated work is his *Les six livres de la république (The Six Books of a Commonwealth)* (1576), an encyclopedic treatise of public law and policy that appeared in 1576. The theory of sovereignty, which provides its framework, was a major event in the development of European political thought. Bodin's precise definition of

supreme authority, his determination of its scope and his analysis of the functions it logically entailed helped to turn public law into a scientific discipline. With Bodin and his followers (especially in Germany), the various jurisdictions of a state could be systematically ordered with respect to an ultimate centre of authority. And his elaboration of the implications of sovereignty through a vast synthesis of comparative public law helped to launch a whole new literary genre.

Bodin's doctrine of sovereignty, however, was seriously flawed by his erroneous views on the indivisibility of sovereignty. He believed that all the powers of the state had ultimately to be concentrated in a single individual or group. This was presented not only as a recommendation of political prudence but as the analytic condition of a coherent and coordinated legal system. Bodin could thus conclude that a mixed constitution, in which the prerogatives of sovereignty were shared or separated, was logically impossible. He therefore failed to see that shared or separated powers produced a compound sovereign, the components of which were coordinated by an underlying basic norm, or rule of recognition, accepted by the general community. Sovereignty, for Bodin, was always that of a ruler. What he needed, but could not imagine, was some notion of constituent authority distinct from the ordinary power of a government. (see Constitutionalism §1; Sovereignty §§1, 3)

3 Public law - the French monarchy and absolutism

Bodin's rejection of the mixed constitution would ultimately lead him to an absolutist interpretation of the French and other monarchies of Western Europe. (see Absolutism §§2-3). This was not his original intention, and in his *Methodus ad facilem historiarum cognitionem (Method for the Easy Comprehension of History)* (1566), he worked with a notion of limited supremacy. Ten years later, however, he had come to the conclusion that sovereign authority was absolute as well as indivisible. This seems to have resulted in part from further reflections on the logic of undivided power and in part from his deep fears of imminent anarchy arising from challenges to royal authority in the renewed religious wars. In the *République* Bodin developed an absolutist interpretation of the French kingship as well as of the Spanish and the English, which supplied a conservative reply to existing doctrines of resistance. Yet it must be emphasized that Bodin was no friend of arbitrary governance. Although a proper king was absolute juridically, prudence and decency required that he seek the advice of the Estates and respect the judgments of the *Parlements*. The king, moreover, was limited, morally at least, by the law of nature as well as by certain fundamental laws on the organization of the crown and its domain (see Natural Law). The law of nature, furthermore, was rigorously interpreted, and it even prohibited, at least in Bodin's reasoning, the imposition of new taxes without consent. The critical point for him politically was that a king's obligation to the law of nature was owed to God alone. Resistance by subjects was thus excluded, for according to contemporary notions, resistance to an authority that was absolute in the sense of not responsible to human agents was forbidden by the law of God.

4 Other contributions to social theory

Bodin's account of sovereignty and public law was only one of many pioneering contributions to social theory. His *Methodus* of 1566, which was a guide to the profitable study of universal history, included a critical method for evaluating historical statements, a system of universal chronology and a theory of progress in the arts and sciences, as well as an extensive preliminary exposition of his theory of public law. And both the *République* and the *Methodus* contain a long chapter, clearly anticipating Montesquieu, on how climate and geography shape the social and political temperament of nations. In *La response de Jean Bodin à M. de Malestroit (The Response of Jean Bodin to the Paradoxes of M. Malestroit)* (1568), Bodin explained the price revolution of the sixteenth century as the result of the sudden influx of precious metals from America. Although anticipated somewhat by Copernicus, he was the first to arrive at a clear explanation of the quantity theory of money. His thesis, furthermore, was illustrated and proven by a reconstruction of the historical movement of French prices, which was a model of sophisticated economic historiography. And his findings on the movements of bullion and goods across national boundaries led him to shrewd observations on the international division of labour.

5 Religious thought

Bodin's religious thought was also strikingly bold and highly influential. His main work on religion, the *Colloquium heptaplomeres de rerum sublimium arcanis abditis (Colloquium of the Seven about Secrets of the Sublime)*, was to seem scandalously freethinking to contemporaries. Probably written around 1593, Bodin left it in

manuscript form and wanted it burned upon his death. It survived in manuscript and obtained underground circulation among scholars until it was finally published in the nineteenth century.

Bodin seems to have resolved his lifelong search for religious truth with a theistic form of Neoplatonism (see Neoplatonism §5). Neoplatonic religiosity had strong appeal to Renaissance intellectuals. But Bodin's version is distinctive, however, in that it is Judaized and very strictly unitarian. Speculative reason teaches us that God exists, that he orders the cosmos by his angels and demons, and that he reveals his will for humankind and his purposes in history by inspiring his prophets. Thus disciplined, speculative reason conducts humans to their highest good, which consists of a kind of mystic unity with God. As Bodin describes it in the *Paradoxon (The Paradox of Jean Bodin)* (1596a), this unity is not, and cannot be, a participation in divinity, or even an active form of contemplation. It is rather a passive opening to God, by way of contemplation, that allows God's light to enter and illuminate the soul. Properly ordered, all aspects of life, both intellectual and moral, are subordinated to this goal. The science of nature, which Bodin treats in his *Universae naturae theatrum (Theatre of Nature in its Entirety)* (1596b), teaches the wonders and beauties of God's creation. Contemplative wisdom, however, leads beyond science in bringing us closer to God and culminates in illumination, the content of which, for those chosen to receive it, is the gift of prophetic powers.

Although this approach to God is possible in all religions, Judaism is held to be the oldest and truest, and the revelations of its prophets and sages are said to be the best. Christianity is portrayed as flawed, not only for its trinitarianism but also for its doctrine of original sin and the need for a saviour. Nevertheless, it is not the purpose of the *Colloquium heptaplomeres* to prove the claims of any one of the revealed religions. All the interlocutors - each representing a different theological position - agree that the differences among them cannot be resolved by argument, that sincere worship in any of the positive religions is pleasing to God, and that they will agree to disagree in the tolerant spirit of Venice, which is the imagined locale of the colloquium.

With respect to public policy, Bodin's recommendations, although still liberal for the time, are more cautious and politically aware. Where religious uniformity existed, it was to be preserved no matter what its form, since politico-religious factionalism was among the worst of evils. Since a state religion would not be unacceptable to God, the philosopher could observe it outwardly, while cultivating truth in private. But forced conversions are always to be shunned, and where a religious minority has become numerous, limited toleration is the prudent course.

Bodin's religious mysticism also had its darker side. He believed in astrology and numerology and attempted to apply both to political science. Darker still was his all too influential book, *De la démonomanie des sorciers (On The Demon-mania of Witches)* (1580), on the detection and punishment of witches. Nevertheless, these deviations into superstition were not uncommon in the Renaissance and ought not to obscure the fact that Bodin's religious thought was a significant moment in the development of universalism and religious toleration. This body of thought, together with his contributions to political and social theory, entitle him to be regarded as one of the foremost thinkers of his time.

See also: Toleration

JULIAN H. FRANKLIN

List of works

For an extended bibliography containing all known editions of each work and editions of Bodin's letters, see Denzer (1973: 492-500).

Bodin, J. (1566) *Methodus ad facilem historiarum cognitionem*; in P. Mesnard (ed.) *Œuvres philosophiques de Jean Bodin*, Paris: Presses Universitaires de France, 1951, 107-269; trans. B. Reynolds, *Method for the Easy Comprehension of History*; New York: Columbia University Press, 1945; repr. 1969.(A collection of essays on various topics designed to supply the reader with the information and critical tools needed to read historians with profit.)

Bodin, J. (1568) *La vie chère au XVIe siècle. La response de Jean Bodin à M. de Malestroit*; ed. H. Hauser, Paris: A. Colin, 1932; trans. of 1578 edn, G.A. Moore, *The Response of Jean Bodin to the Paradoxes of M Malestroit*, Washington, DC: Country Dollar Press, 1946.(An early statement of the quantity theory of money designed to show that the great price inflation of the time was caused by the importation of precious metals, mostly from America.)

Bodin, J. (1576) *Les six livres de la république*, Paris: Jacques du Puy, 1583; repr. Aalen: Scientia, 1961; trans. J. Bodin, *De republica libri sex*, Paris: Jacques du Puy, 1586; trans. R. Knolles, *The Six Books of a Commonwealth*, London: B. Bishop, 1606; repr. with apparatus by K.D. McRae, Cambridge, MA: Harvard University Press, 1962; ed. and partial trans. into modern English, J.H. Franklin, *Bodin: On Sovereignty*, Cambridge: Cambridge University Press, 1992.(Bodin's grand synthesis of comparative public law and policy centring around his investigation of the nature of sovereignty.)

Bodin, J. (1580) *De la démonomanie des sorciers*, Hildesheim and New York: G. Olms, 1988; trans. R.A. Scott, abridged J.L. Pearl, *On The Demonomania of Sorcerers*, Toronto, Ont: Centre for Reformation and Renaissance Studies, Victoria University in the University of Toronto, 1995.(A study of the satanic practices of witches and how to detect and convict them.)

Bodin, J. (*c*.1593) *Colloquium heptaplomeres de rerum sublimium arcanis abditis*, ed. L. Noack, Schwerin: F.G. Baereusprung, 1st complete edn, 1857; repr. Hildesheim and New York: G. Olms, 1970; trans. L. Daniels Kuntz, *Colloquium of the Seven about Secrets of the Sublime*, Princeton, NJ: Princeton University Press, 1975.(The text consists of a friendly debate among representatives of seven religions/religious positions as to which religion is the truest and best, ending in an agreement to disagree.)

Bodin, J. (1596a) *Paradoxon, quod nec virtus ulle in mediocritate, nec summum hominis bonum in virtutis actione consistere possit (The Paradox of Jean Bodin… that there is no virtue in mediocrity or in the mean between two vices)*, Paris; trans. J. Bodin, *Le Paradoxe de J. Bodin… qu'il n'y a pas une seule vertu en médiocrité, ny euy milieu de deux vices*, Paris, 1598; repr. P.L. Rose (ed.) *Jean Bodin, Selected Writings on Philosophy, Religion and Politics*, Geneva: Librairie Droz, 1980, 43-75.(Rejects the Aristotelian idea idea of the mean as the decisive criterion of virtue in favour of the pure love of divine truth.)

Bodin, J. (1596b) *Universae naturae theatrum (Theatre of Nature in its Entirety)*, Lyon and Paris: Jacob Roussoon; trans. Fr. de Fougerolles, *Le theatre de la nature universelle*, Lyon: Jean Pillehotte, 1597.(A systematic account, broadly Platonic in inspiration, of the natural order of the cosmos.)

References and further reading

Chauviré, R. (1914) *Jean Bodin, auteur de la République (Jean Bodin, Author of the République)*, La Flèche: E. Besnier; repr. Geneva: Slatkine, 1969.(Although now outdated in some aspects, still a useful treatment in one volume of Bodin's political thought as a whole.)

Denzer, H. (ed.) (1973) *Verhandlungen der internationalen Bodin Tagung in Munchen (Proceedings of the International Conference on Bodin held in Munich)*, Munich: C.H. Beck. (Contains a very extensive bibliography of secondary works as well as a number of important scholarly articles, some in English, on diverse aspects of Bodin's life and thought.)

Franklin, J.H. (1973) *Jean Bodin and the Rise of Absolutist theory*, Cambridge: Cambridge University Press.(A critical analysis of the genesis, structure and ideological impact of Bodin's theory of sovereignty.)

Goyard-Fabre, S. (1989) *Jean Bodin et le droit de la république (Jean Bodin's and the Law of the Commonwealth)*, Paris: Presses Universitaires de France.(A useful running commentary on the overall system of the *République*; includes an extensive bibliography of secondary works.)

Mesnard, P. (1929) 'La pensée religieuse de Bodin', *Revue du seizième siècle* 16: 77-121.(A judicious treatment of Bodin's religious thought by the most eminent of Bodin's commentators. There is no book-length study by Mesnard, but for listings of his numerous articles on many aspects of Bodin's thought and biography, see the bibliographies in Denzer 1973, and in Goyard-Fabre 1989.)

Rose, P.L. (1980) *Bodin and the Great God of Nature, The Moral and Religious Universe of a Judaizer*, Geneva: Librairie Droz.(An extensive study of the development and character of Bodin's religious thought, with heavy emphasis on its Judaizing element.)

Boehme, Jakob (1575-1624)

Boehme was a Lutheran mystic and pantheist. He held that God is the Abyss that is the ground of all things. The will of the Abyss to know itself generates a process that gives rise to nature, which is thus the image of God. Life is characterized by a dualistic struggle between good and evil; only by embracing Christ's love can unity be regained. Boehme was highly regarded by such diverse writers as Law, Newton, Goethe and Hegel.

A Lutheran theosopher, with a predilection for both mysticism and philosophy of nature, Jakob Boehme was a Silesian, a native of Alt Seidelberg near Görlitz. Situated between Catholic Poland and Lutheran Saxony, Silesia was a haven for heterodoxies in the late sixteenth century, although its tradition of hospitality ended abruptly with the Thirty Years War. Boehme spent most of his life in Görlitz, as a member of the Cobblers' Guild. He was an astute businessman, who had no formal training in the liberal arts but read voraciously and wrote inspiringly. His first mystical experience was in 1600, when he contemplated the 'Being of all beings, the Byss and the Abyss' in the sunlight reflected in a pewter dish.

Published in 1612, *Morgenröthe im Anfang (The Red Light at Dawn)* was Boehme's first attempt at solving the problem of theodicy. It immediately incurred the condemnation of Görlitz's Lutheran church. He was forbidden to write further, but his reputation was established. Boehme kept silent for seven years and then released the *Beschreibung der drey Principien Göttliches Wesen (Concerning the Three Principles of the Divine Essence)* in 1619, and *Hohe und tieffe Grund von dem drey fachen Leben des Menschen (The High and Deep Searching Out of the Threefold Life of Man)*, *De incarnatione verbi (On the Incarnation of the Word)*, *Sechs theosophischen Puncten (Six Theosophic Points)* and *Kurtze Erklärung von Sechs mystischen Puncten (Short Exposition of Six Mystical Points)* in 1620. A large commentary on Genesis, *Mysterium magnum (The Great Mystery)*, came out in 1623, followed in 1624 by a collection of small treatises, *Der Weg zu Christo (The Way to Christ)*. Written in 1622, *De signatura rerum (The Signature of All Things)* was posthumously published in 1635.

Like Valentin Weigel (1533-88), a subjective pantheist, Boehme began with the self, but he emphasized its will. The self, the source of all knowledge, is derived from the universal feeling for life (*Lebensgefühl*). Boehme saw himself as an agent of the Spirit, which, in his worldview, began the process of self-understanding that culminated in the inner vision of a universally present and active Christ. What is God? He is the Abyss (*Ungrund*), the ground of all things, the undifferentiated absolute, the eternal, natureless, unconscious Nothing that lies at the foundation of everything. At the core of the Abyss lies a will to self-intuition. This will initiates the process of self-knowledge, and its outgoing dynamic activity creates the inner world, which is the prototype of the outer world. In the self-noetic process, the will of the Nothing searches for something and discovers it within itself. Eternal nature finds its being in this process. With differentiation emerge evil, dualism and conflict. Boehme's voluntarism, which bears Luther's mark, is coupled with the doctrine of the Trinity and flavoured with the Manichean dualism of light and darkness, good and evil, love and hatred, grace and wrath.

For Boehme, nature is the image of God; he thus formulated the identity of God with nature half a century before Spinoza (§2). He also framed a theory of seven natural properties. In a letter dating from 11 November 1623, which furnishes a clear compendium of his metaphysics, Boehme defines these seven properties as desire, sensation, anxiety, fire, light, sound and being. The Trinity arises from the unfathomable will of the Father, which creates for all eternity the unfathomable will of the Son; from both emanates the Spirit, the 'moving life' that mirrors both the Father and the Son. History is where the struggle for life unfolds, which Boehme describes as a fight between good and evil, where the decision for or against God is made. Meaning is to be found in Christ. The purpose of life is to retrieve the lost unity by allowing the fire of love, Christ's heart, to embrace everything. Life should therefore be an imitation of Christ's suffering and triumph.

Boehme's theosophy, which can be characterized as a preparation for the mystical acknowledgement of Christ, shows the influence of Neoplatonism, Gnosticism, kabbalism, Paracelsian pansophism, Caspar Schwenckfeld's spiritualism, and Sebastian Franck's humanistic illuminism. Boehme's influence has been considerable. He was the most often translated German author of the seventeenth century. Descartes, Spinoza, the Cambridge Platonists and Newton read his works. His cosmic, metaphysical and ethical dualism enchanted the Romantics Novalis, Tieck and Goethe. Hegel celebrated him as the first true German philosopher, and Schelling owed to him his philosophy

of identity. Besides a crucial influence on the devotional writer William Law, the quietist Antoinette Bourignon and the poet William Blake, Boehme had an ecclesiastical following in the Low Countries (The Invisible Church of the Angel's Brothers) and in England (The Philadelphians).

JEAN-LOUP SEBAN

List of works

Boehme, J. (1955-61) *Sämmtliche Schriften*, ed. W.-E. Peuckert, Stuttgart: Frommanns, 11 vols.(The standard edition of Boehme's works.)

Boehme, J. (1612) *Morgenröthe im Anfang* (The Red Light at Dawn), trans. J. Sparrow, ed. C.J. Barker and D.S. Hehner, *The Aurora*, London: J.M. Watkins, 1960.(Written to preserve the memory of his mystical experience, and to probe God's relation to the world and its evils.)

Boehme, J. (1619) *Beschreibung der drey Principien Göttliches Wesen*, trans. J. Sparrow, ed. C.J. Barker, *Concerning the Three Principles of the Divine Essence*, London: J.M. Watkins, 1910.(A secretly composed and very opaque work in which Boehme dips into alchemy and astrology to 'clarify' his previous book.)

Boehme, J. (1620a) *Hohe und tieffe Grund von dem drey fachen Leben des Menschen*, trans. J. Sparrow, ed. C.J. Barker, *The High and Deep Searching Out of the Threefold Life of Man*, London: J.M. Watkins, 1909. (Boehme's system deepens: God is the absolute person who reveals himself by creating the world which is known by human thought.)

Boehme, J. (1620b) *De incarnatione verbi, Von der menschwerdung Jesu Christi*, trans. J.R. Earle and S.R. Webster, *De incarnatione verbi: Of the Incarnation of Jesus Christ*, London: Constable, 1934.(Sections on the Incarnation and Mary, the Christian's union with Christ's cross, and the way to being one in Spirit with God.)

Boehme, J. (1620c) *Sechs theosophischen Puncten*, trans. J.R. Earle, in *Six Theosophic Points and Other Writings*, Ann Arbor, MI: University of Michigan Press, 1958. (English edition printed with 1620d and other works. Further expansions and descriptions of Boehme's complete scheme.)

Boehme, J. (1620d) *Kurtze Erklärung von Sechs mystischen Puncten* (Short Exposition of Six Mystical Points), trans. J.R. Earle, in *Six Theosophic Points and Other Writings*, Ann Arbor, MI: University of Michigan Press, 1958.(See 1620c for further details.)

Boehme, J. (1620e) *Viertzig Fragen von der Seele*, trans. J. Sparrow, *Forty Questions on the Soul and The Clavis*, London: J.M. Watkins, 1911.(Further complexification of the system: Boehme treats God as the Ungrounded One who objectifies himself.)

Boehme, J. (1623) *Mysterium magnum* (The Great Mystery), trans. J. Sparrow, ed. C.J. Barker, London: J.M. Watkins, 2 vols, 1941.(A substantial commentary on the book of Genesis; the English translation leaves some material out.)

Boehme, J. (1624) *Der Weg zu Christo*, trans. P.C. Erb, *The Way to Christ*, New York: Paulist Press, 1978.(Nine mystical treatises: the best introduction to Boehme's spirituality.)

Boehme, J. (1635) *De signatura rerum*, trans. J. Ellistone, *The Signature of All Things, With Other Writings*, London and Cambridge: J.M. Clarke, 1969.(The original subtitle in English (1651) shows the ambition of Boehme's system: 'shewing the sign and signification of the severall forms and shapes in the creation, and what the beginning, ruin, and cure of every thing is'.)

Boehme, J. (1763-81) *The Works of Jacob Behmen, the Teutonic Theosopher*, trans. J. Sparrow and J. Ellistone, London: M. Richardson; microfilm, New Haven, CT: Research Publications, 1973.(This English edition contains a sketch of Boehme's life by William Law.)

Boehme, J. (1963-6) *Die Urschriften* (The Original Manuscripts), ed. W. Buddecke, Stuttgart: Frommanns, 2 vols.(A critical edition of previously unpublished materials.)

Boehme, J. (1989) *Jacob Boehme: Essential Readings*, ed. R. Waterfield, Wellingborough: Crucible.(Some obscure texts, fragments and letters which together provide an overview of Boehme's thought.)

References and further reading

Brown, R.F. (1977) *The Later Philosophy of Schelling: The Influence of Boehme on the Works of 1809-1815*, London and New York: Associated University Presses.(A long first section on Boehme treats several of his works in depth, and presents a systematic outline of his mature system.)

Buddecke, W. (1937-57) *Die Jakob Böhme-Ausgaben*, Göttingen: Ludwig Häntschel, 2 vols.(A complete bibliography of Boehme's works.)

Koyré, A. (1929) *La Philosophie de Jacob Boehme*, Paris: J. Vrin.(Remains the best book on Boehme.)

Peuckert, W.-E. (1924) *Das Leben Jakob Böhmes*, Jena: Eugen Diederichs.(The first critical biography, and still

the benchmark.)
Stoudt, J.J. (1957) *Sunrise to Eternity: A Study in Jacob Boehme's Life and Thought*, Philadelphia, PA: University of Pennsylvania Press.(A reliable and thorough biography.)
Walsh, D. (1983) *The Mysticism of Innerworldly Fulfillment: A Study of Jacob Boehme*, Gainesville, FL: University of Florida Press.(An exploration of Boehme's significance in the context of late-Renaissance magico-mystical movements.)
Weeks, A. (1991) *Boehme: An Intellectual Biography of the Seventeenth-Century Philosopher and Mystic*, Albany, NY: State University of New York Press.(Places Boehme firmly in his social and historical context.)

Boethius of Dacia (*fl. c.*1275)

Boethius developed an original theory of scientific knowledge designed to reconcile science with Christian doctrine without allowing one to determine the contents of the other. His main strategy was to consider each science as an independent system of axioms and theorems while also operating with a hierarchy of causes, the highest of which (God) is fundamentally unpredictable as to its operations. Boethius did, however, stress the powers of the human intellect and the possibility of reaching happiness through rational understanding; he vigorously objected to demands that natural science should adapt its axioms to the demands of Christian faith. This laid him open to suspicions of heresy. Boethius' work on grammar is the most complete application of his ideas of how to construct a science.

1 Life and works

All information about Boethius' life before 1277 is contained in the number and nature of his extant or attested writings, and in the title and the epithet accompanying his name in medieval sources, 'Magister Boethius de Dacia' or 'Dacus', which identifies him as a Dane. All his works, whether preserved or only attested, are such as a medieval master of arts would produce. Their number and quality suggest a rather long career, starting perhaps about 1265. If he was a master of arts by then, he cannot have been born much later than 1240.

Among his surviving works, the best known are the treatises *De summo bono (On the Highest Good)* and *De aeternitate mundi (On the Eternity of the World)*. However, there are a number of extant works in other fields, including logic (a sophisma on meaning, truth and knowledge, and *Quaestiones super librum Topicorum (Questions on Aristotle's Topics)*, natural philosophy (questions on Aristotle's *Physics* and *On Generation and Corruption* - the extant version of the questions is a reworking by someone else of Boethian material) and grammar (a sophisma, *Quaestiones super Priscianum maiorem (Questions on Priscian)*). His questions on Aristotle's *Metaphysics* has been lost, but passages borrowed from it seem to be preserved in a near-contemporary anonymous work. Boethius also wrote one of the very first Latin commentaries on Aristotle's *Rhetoric*, which appears to be irretrievably lost.

During the 1270s, Boethius taught at the University of Paris. In 1277, Etienne Tempier, Bishop of Paris, accused unnamed masters in the faculty of arts of propagating un-Christian doctrines and specifically prohibited the teaching of any of 219 theses (see Aristotelianism, medieval). Many of these theses were extracted from Boethius' works. It is the general, but unproven, assumption that he was still at Paris at the time of the condemnation of 1277, and that this event halted his university career. He may have sought a new life among the Dominican Order, for although he was almost certainly a secular during his regency (active professorship) in arts, his works occur in a medieval catalogue of books composed by Dominicans. Some of his books and ideas lived on, but Boethius himself was soon forgotten; only in the twentieth century has he re-emerged as an important figure in the history of philosophy.

2 Ethics, theory of knowledge and science

Boethius was an epistemological optimist. Real knowledge, even some knowledge about the First Cause, he thought, is within human grasp. The full realization of human nature, happiness, is the philosopher's life in which all lower powers are directed towards the supreme activity, the contemplation of truth, and of the First Truth in particular. By contrast, the layman's uncontemplative life is only quasi-human. As was common in his day, Boethius assumed a hierarchical structure of reality in which beings can be arranged on a scale according to their proximity to the First Cause. His philosopher first achieves an insight into the causal structures of the material (sublunary) world and thence proceeds to an understanding of higher forms of being, which are causes in relation to the lower ones. Eventually he arrives at the contemplation of the First Object of Knowledge, the First Cause.

Such intellectualism was not unusual for the time. Boethius' peculiar contribution was the way he connected it with a theory of science, based on Aristotle (in particular on the *Posterior Analytics*) but also inspired by the stylized disputations of thirteenth-century university life. To Boethius, the objects of knowledge are 'things with regard to their causes', not simply individual things nor hypostasized quiddities (or essences), for he would accept no uninstantiated quiddities. The proposition 'every human being is an animate being' expresses a piece of

knowledge about the (formal) cause of human beings, but since the existence of humans is contingent, Boethius claimed that the proposition is false unless taken to mean ‘supposing there exists at least one human being, it is true that every human being is an animate being’. Scientific propositions are only conditionally true, on the presupposition that their subject terms have referents, and in the natural sciences this presupposition could fail to be satisfied.

Each science has its own primitive terms and propositions, and each science incorporates insight into one causal network. However, one thing may enter into more than one network and one type of cause may annihilate the effect of another. To Boethius, scientific activity was similar to participation in a rule-governed university disputation in which a proposition should be granted or denied according to whether it cohered or not with certain other propositions, rather than on the basis of correspondence with external facts. He held that as long as one works within some particular science, *A*, one should grant any theorem properly derived from the axioms of *A* and deny any proposition inconsistent with them. The tacit presupposition is that there are such things as the objects of science *A*, and that only the causal mechanisms known through *A* are at work. In actual fact foreign causal mechanisms, known perhaps to science *B*, may prevent the predictions of *A* from corresponding to reality. Boethius may be said to have held a coherence theory of truth within a science, and a correspondence theory of truth simpliciter (see Truth, coherence theory of; Truth, correspondence theory of).

The hierarchy of entities and causes provides an explanation of why the work of lower causes can be frustrated by higher ones, and why the assumption of a divine free will makes the First Cause not totally transparent. Consequently, it is not possible to unify all sciences into one science of the First Cause (see Causation).

In Boethius’ view, it follows that the same person can consistently hold that it is a truth of natural science that every human being had a father and mother, and a revealed truth that there was a first couple of humans with no parents. The revealed truths are truths simpliciter, but truths knowable only thanks to revelation cannot be incorporated into scientific theories on pain of inconsistency. In this way Boethius tried to solve one of the burning questions of his day, how to reconcile faith and natural science, which seemed to clash over the question whether there can have been a beginning of the world as a whole and of natural species. Standard theology took the doctrine of creation to imply that the world is not infinitely old; standard science implied the contrary (see Creation and conservation, religious doctrine of; Eternity of the world, medieval views of; Natural philosophy, medieval).

3 Grammar

We see Boethius’ theory of science at work in his grammar. Grammar, he held, deals with the ways humans can express what their minds have grasped. Concepts have a core corresponding to the core, or ‘common nature’, of real entities, attended by ‘modes of understanding’ corresponding to ‘modes of being’ of the common nature, that is, ways in which it can manifest itself. The linguistic expression of a concept has a core, the significate (what is signified) attended by modes of signifying. For example, if someone’s back is aching, the pain is in one way something static like a substance and can carry accidental properties: thus it can be more or less acute. In another way, the pain is an accidental entity in the process of actualization. Pain ‘is’ in both ways, and is conceptualized in both ways; it is also signified in both ways, statically by the noun ‘a pain’, processively by the verb ‘to ache’.

It is none of the grammarian’s business to investigate what things there are or what properties they have; grammaticality is not measured by correspondence with the external world, so a grammatical explanation must never appeal to properties of the significate of a word. Nor can grammar establish the inventory of modes of signifying; they are a function of the modes of understanding which in turn are a function of the modal structure of extramental reality. To the grammarian, the modes of signifying are primitive terms. Neither can the grammarian predict what sounds will embody the significates and the modes of signifying. Any sound may conventionally represent any significate or mode of signifying, and languages may vary considerably; thus Greek (Boethius mistakenly thought) uses separate words (articles) to express what in Latin is expressed by the inflexion of nouns (case, number). What a Boethian grammarian can do is formulate the rules governing which combinations of modes of signifying yield well-formed sentences if instantiated. The rules of grammar are independent of the existence of any sentence of any language, but no actual sentence in any language can be fully intelligible if it breaks the rules (see Language, medieval theories of; Language, philosophy of). By strictly separating grammar from metaphysics, psychology and other sciences, Boethius tried to transform it into a formal discipline like

geometry or logic, unshakeable by any contingent fact but also unable to support deductions about the extralinguistic world.

See also: Aristotelianism, medieval; Averroism; Language, medieval theories of; Natural philosophy, medieval §7; Siger of Brabant

STEN EBBESEN

List of works

Boethius of Dacia [Boethius Dacus] (*c.*1275) *Opera*, ed. J. Pinborg, H. Roos, N.J. Green-Pedersen, S. Ebbesen and I. Rosier, Corpus Philosophorum Danicorum Medii Aevi vols IV-IX, Copenhagen: DSL/Gad, 1969-.(Vol. IX, *Sophismata*, has not yet appeared.)

Boethius of Dacia [Boethius Dacus] (*c.*1275) *De summo bono (On the Supreme Good)*, ed. N.J. Green-Pedersen in *Opera*, vol. VI, Copenhagen: DSL/Gad, 1976; trans. J.F. Wippel, *On the Supreme Good. On the Eternity of the World. On Dreams*, Mediaeval Sources in Translation 30, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1987.(Boethius' major work on the good.)

Boethius of Dacia [Boethius Dacus] (*c.*1275) *De aeternitate mundi (On the Eternity of the World)*, ed. N.J. Green-Pedersen in *Opera*, vol. VI, Copenhagen: DSL/Gad, 1976; trans. J.F. Wippel, *On the Supreme Good. On the Eternity of the world. On Dreams*, Mediaeval Sources in Translation 30, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1987.(Boethius' important work on eternity.)

Boethius of Dacia [Boethius Dacus] (*c.*1275) *Quaestiones super librum Topicorum (Questions on Aristotle's Topics)*, ed. N.J. Green-Pedersen and J. Pinborg in *Opera*, vol. VI, Copenhagen: DSL/Gad, 1976.(Boethius' major logical work.)

Boethius of Dacia [Boethius Dacus] (*c.*1275) *De somniis (On Dreams)*, ed. N.J. Green-Pedersen in *Opera*, vol. VI, Copenhagen: DSL/Gad, 1976; trans. J.F. Wippel, *On the Supreme Good. On the Eternity of the World. On Dreams*, Mediaeval Sources in Translation 30, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1987.(Boethius on natural philosophy.)

Boethius of Dacia [Boethius Dacus] (*c.*1275) *Quaestiones super libros Physicorum (Questions on Aristotle's Physics)*, ed. G. Sajó in *Opera*, vol. V.2, Copenhagen: DSL/Gad, 1974.(Boethius on natural philosophy.)

Boethius of Dacia [Boethius Dacus] (*c.*1275) *Quaestiones de generatione et corruptione (Questions on On Generation and Corruption)*, ed. G. Sajó in *Opera*, vol. V.1, Copenhagen: DSL/Gad, 1972.(Boethius on natural philosophy.)

Boethius of Dacia [Boethius Dacus] (*c.*1275) *Quaestiones super Priscianum maiorem (Questions on Priscian Major)*, ed. J. Pinborg and H. Roos in *Opera*, vol. IV, Copenhagen: DSL/Gad, 1969; trans. A.C. Senape McDermott, *Godfrey of Fontaine's Abridgement of Boethius of Dacia's Modi Significandi sive Quaestiones super Priscianum Majorem*, Amsterdam Studies in the Theory and History of Linguistic Science, Series III, Studies in the History of Linguistics vol. 22, Amsterdam: Benjamins, 1980.(Work on grammar which shows Boethius' philosophy of science at work.)

References and further reading

Kretzmann, N. and Stump, E. (1988) *Logic and the Philosophy of Language*, Cambridge Translations of Medieval Philosophical Texts 1, Cambridge: Cambridge University Press. (Contains translation of a part of *Sophisma* I.)

Marmo, C. (1994) *Semiotica e linguaggio nella scolastica: Parigi, Bologna, Erfurt 1270-1330 (Semiotics and Language in Scholasticism: Paris, Bologna, Erfurt 1270-1330)*, Rome: Istituto Storico Italiano per il Medio Evo.(A large-scale study of modism, drawing on both grammatical and logical sources.)

Pinborg, J. (1967) *Die Entwicklung der Sprachtheorie im Mittelalter (The Development of Theory of Language in the Middle Ages)*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters, Texte und Untersuchungen 42.2, Münster: Ashendorff, and Copenhagen: Frost-Hansen.(A pioneering and authoritative study of modistic grammar.)

Pinborg, J. (1972) *Logik und Semantik im Mittelalter. Ein Überblick (Logic and Semantics in the Middle Ages: A Survey)*, Stuttgart: Frommann-Holzboog.(Discusses Boethius' views on several matters.)

Pinborg, J. (1984) *Medieval Semantics, Selected Studies on Medieval Logic and Grammar*, ed. S. Ebbesen, London: Variorum.(Several relevant papers, including 'Zur Philosophie des Boethius de Dacia, Ein Überblick'.)

Boethius, Anicius Manlius Severinus (*c*.480-525/6)

Boethius was a principal transmitter of classical Greek logic from Aristotle, the Stoics and the Neoplatonists to the schoolmen of the medieval Latin West. His contemporaries were largely unimpressed by his learned activities, and his writings show him to have been a lonely, rather isolated figure in a world where the old Roman aristocrats were struggling to maintain high literary culture in an Italy controlled by barbarous and bibulous Goths, whose taste in music and hairgrease Boethius found painful.

Boethius himself was born into a patrician family in Rome, but was orphaned and raised instead by Q. Aurelius Memmius Symmachus, a rich Christian heir to a distinguished pagan line; Boethius later married the latter's daughter, Rusticiana. As well as Symmachus, Boethius had a small circle of educated friends, including the Roman deacon John (who probably became Pope John I, 523-6), who shared his enthusiasm for logical problems. The Gothic king of Italy at Ravenna, Theoderic, had met high culture during his education at Constantinople and made use of experienced Roman aristocrats as administrators. He employed Boethius to design a sundial for the Burgundian king and also a waterclock, specimens of advanced technology intended to impress a barbarian; he also sent a harpist to Clovis, the Frankish king, no doubt intended to soften the latter's bellicose spirit.

By 507 Boethius had gained the title 'patrician' and received letters addressed to 'your magnitude'. Symmachus was in a position to promote his public career. He was nominated consul for the year 510, a position without political power but of high standing and requiring large disbursements of private wealth; it also carried the perquisite that the consul's name stood on all dated documents for that year. In 522 his two sons were installed as consuls, a promotion that gave their father intense pride and pleasure, and he took up seriously the political post of Master of the Offices. In this capacity, his determination to eliminate corruption earned him numerous enemies among both Goths and his fellow Roman aristocrats. His relations with the courtiers at Ravenna became disastrous.

Boethius' fall came when he rashly defended a senator who had been delated to King Theoderic for conducting treasonable correspondence with persons high in the court of the emperor at Constantinople. There is no improbability in the notion that, along with other Roman aristocrats, Boethius would have preferred to be rid of the crude Goths and to see Theoderic replaced by a ruler congenial to the emperor. His great erudition had aroused fears that he was engaged in occult practices dangerous to the Ravenna dynasty. In 524 or early 525, Boethius was imprisoned at Pavia (Ticinum). Here, while awaiting the execution already decreed against him, he composed his masterpiece, De consolatione philosophiae (The Consolation of Philosophy).

De consolatione philosophiae, a bitterly hostile attack on Theoderic prefacing a philosophical discussion of innocent suffering and the problem of evil, must have been smuggled out of prison, no doubt with the aid of gold coins from Rusticiana or Symmachus. In the ninth century, the work captured the imagination of Alcuin at the court of Charlemagne, became a standard textbook in schools and was set on the way to being one of the greatest books of medieval culture, especially popular among laymen.

Boethius' earlier works have been the preserve of more specialized readers, especially concerned with the history of ancient philosophy. His stated original intention was to educate the West by translating all of Plato and Aristotle into Latin and to supply explanatory commentaries on many of their writings. That was too ambitious. He did not proceed beyond some of the logical works (Organon) of Aristotle, prefaced by a commentary on a Latin translation of Porphyry's Isagōgē(Introduction) made in the fourth century by Marius Victorinus, an African teaching in Rome, and then by a second commentary on a translation of the same text made by himself. This commentary underlay the medieval debates on universals. He also wrote a commentary on Aristotle's Categories and two commentaries on Aristotle's De interpretatione. In addition, Boethius adapted Nicomachus of Gerasa's Arithmetic for Latin readers, Nicomachus' introduction to music as a liberal art, a commentary on Cicero's Topics, a short treatise 'On Division', important treatises on categorical and hypothetical syllogisms and a further tract on different kinds of 'topic'.

Intricate theological debates between Rome and Constantinople convinced him that a trained logician could contribute clarification, and he composed four theological tractates on the doctines of the Trinity and the person of Christ, concentrating on logical problems. In addition, a fifth tract became a statement of orthodox belief

without much reference to logical implications. The five pieces, or Opuscula sacra, became hardly less influential than De consolatione philosophiae, especially from the twelfth century onwards. We hear of critics who thought contemporary theologians knew more about Boethius than about the Bible.

1 Life

Boethius was born in Rome into a wealthy Christian family of senatorial standing, in an age when barbarian soldiers ruled and the old aristocratic families had yielded power to them, yet remained indispensable to their Gothic masters for the good order of civil administration. Under the rule of the Ostrogoth king Theoderic, the old Roman families continued to assert their Roman-ness by the study and re-editing of classics of Latin literature - Livy, Cicero, Virgil, Seneca and so on - but also by retaining a politically hazardous contact with the eastern Roman emperor in Constantinople. Boethius' father died when Boethius himself was young, and he was taken in by Q. Aurelius Memmius Symmachus, whose daughter Rusticiana he later married.

The best-educated person of his time in the West, he could read Greek, even if not quite fluently, and his works are rich in literary allusions and reminiscences. He was well read in the Neoplatonic commentators on the logical and other writings of Aristotle, especially Porphyry and Proclus. He was also familiar with at least some of the major writings of Augustine of Hippo, and wrote five theological tractates (*Opuscula sacra*), four of which are devoted to clarifying logical problems in orthodox Catholic doctrine, especially in regard to the doctrines of the person of Christ and the divine Trinity. The quest for acceptable language to express Christian belief on these themes had a bearing on the break in communion between the papacy and the patriarchate of Constantinople from 484 to 518. Boethius and his circle of aristocratic friends in Italy were concerned to heal this breach, and this aspiration led him to attempt to use his dialectical skills to define the terms more closely than had been done previously. The set of five tracts on theology provided influential themes for exegesis by commentators in medieval times, especially during the twelfth and thirteenth centuries, including Thomas Aquinas. However, the contacts with Constantinople and the religious disagreements between the Arian Theoderic and the Orthodox Emperor Justin also played a part in bringing about Boethius' death. To Dante, he was a martyr and hero, and on 23 October 1883 his veneration as a saint was authorized at Pavia.

In his twenties, Boethius embarked on a programme of translation, commentary and adaptation to make available to the Latin West the logic of Aristotle and the standard Greek texts on the four mathematical 'arts', arithmetic, geometry, music and astronomy, for which he coined the term 'quadrivium', parallel to the 'trivium' of grammar (that is, literature), rhetoric and dialectic. Soon, King Theoderic was inviting him to revise the coinage system, to design a waterclock and sundial to send to the Burgundian king Gundobald and to review the system of weights and measures. He was asked to select a harpist to send to Clovis the Frankish king. His intellectual powers and the patronage of his powerful father-in-law Symmachus launched him on a meteoric career in the civil administration. He was nominated consul for 510; his young sons were consuls for 522, when he also delivered a panegyric on Theoderic (a text now lost but known to Cassiodorus). In the same year he became Master of the Offices, a post of considerable power, from which he tried to root out bribery and corruption in the bureaucracy, thereby ensuring that he made many enemies. He understood himself to be following Plato's principle that philosophers must accept government posts since, if they do not do so, wicked gangsters take over.

In 519 Theoderic's heir presumptive died, and speculation about his successor was rife. Boethius rashly defended another senator suspected of a correspondence with Constantinople, an act regarded by King Theoderic as treasonable. In the Eastern empire the new emperor Justin, guided by his nephew Justinian, had a programme of suppressing heresy to give social cohesion, and the sects being harassed included the Arians. However, Theoderic and his Gothic soldiers had received their faith from Arian missionaries. The eastern Roman imperial policy of intolerance therefore appeared to Theoderic at Ravenna as an ambition to eliminate both heresy and Goths in Italy. Tension became high, and Boethius was accused by other Roman aristocrats, keen perhaps to dissociate themselves from someone suspected of dangerously pro-Byzantine sentiment, and was imprisoned at Pavia (Ticinum) and sentenced to death. During his many months awaiting execution he wrote *De consolatione philosophiae (The Consolation of Philosophy)*. The scorn and hatred for Theoderic and other powerful Goths shown in the first book of this work, which was probably smuggled out of prison by his wife and other influential friends, proves that he was not writing an ingratiating work to submit to Theoderic in hope of a reprieve.

In effect, Boethius in his prison was a hostage under deteriorating conditions, being used by Theoderic in an

unsuccessful attempt to restrain Byzantine political and religious aspirations. Late in 525 or possibly early in 526 he was subjected to vile tortures and battered to death. Under Theoderic's anger his father-in-law Symmachus and his intimate friend Pope John I also lost their lives. A Ravenna chronicler, the so-called Anonymus Valesianus, thought that Arian heresy had addled the king's brain and deprived him of his wits. The Byzantine historian Procopius reports a story that Theoderic was haunted by guilt and died in bitter remorse (on 30 August 526). The king may well have come to perceive that the killing of Boethius and others had played into Constantinople's hands, enabling the Byzantine emperor to undermine the general admiration hitherto felt in Italy for Gothic government in cooperation with the old Roman families. Boethius' tomb is in the church of Ciel d'Oro at Pavia.

2 Works

De consolatione philosophiae (in late Latin *consolatio* came to mean help or support) owes its survival to being discovered in the ninth century, probably by Alcuin (see Carolingian renaissance), and was to become Boethius' most widely read work. Pervaded by passionate feeling, composed in a sophisticated Latin in alternating passages of verse and prose and in the last two books handling complex philosophical ideas such as, for example, divine providence, human freedom, eternity and time, the work became a major classic of western literature. The emergence of the *De consolatione philosophiae* and of the *Opuscula sacra* in the ninth century, followed soon after by the logical works, made Boethius' writings important in medieval education. Boethius taught good morality and fine Latinity in an age when both were hard to get. From this time onwards there survive a number of commentaries on the *De consolatione philosophiae*, as much concerned with style and metre as with the content of the argument. That initially the primary function of the work was educational is illustrated by manuscripts of the Latin text with marginal notes in vernacular languages. *De consolatione* also attracted vernacular translators; in England these included King Alfred in the ninth century and Geoffrey Chaucer in the fourteenth. The earliest of several German translations is by Notker of St Gallen *circa* 1010; Jean de Meun translated the work into medieval French in the late thirteenth century, and Jacob Vilt of Bruges translated it into Dutch in 1462. Queen Elizabeth I was among the later translators in England. In the Byzantine world, a Greek translation of *De consolatione philosophiae* was produced by Maximos Planudes (1260-1310), the humanist and monk who also produced a Greek version of Augustine on the Trinity. His translations reflected his support for the contemporary effort by Emperor Michael VIII to avert further military aggression from the West by ending the schism between Constantinople and Rome.

De consolatione philosophiae's first book protests that its author was innocent of the charges brought against him based on forged letters. Then, in contemplation of his hopeless position, the work moves on to a philosophical vindication of the goodness of providence, even in a world where the just are not rewarded with prosperity and the wicked are allowed power. *De consolatione* is a clearly religious work written by a Christian, and yet it is not a Christian work in the sense that it contains nothing about the forgiveness of sins or redemption. Nevertheless, while turns of phrase are much more Neoplatonic than Christian, there is one clear quotation from the Bible (III pr.12. 23-4). There Boethius expresses special pleasure at the fact that the lady Philosophy uses some words from the Wisdom of Solomon, admittedly to reinforce a truth of natural, not revealed, theology. As the title indicates, the work sets out to present the philosophical arguments for a Platonic theodicy, a subject to which all Platonists needed to give some attention and which in late antiquity was specially studied not only by Proclus but also by Hierocles of Alexandria (see Neoplatonism).

The *Opuscula sacra* regard faith and reason as independent but parallel and compatible ways of attaining to higher metaphysical truths, and the independent validity of logical reasoning is also an underlying presupposition throughout *De consolatione*. The apparently conscious reticence of *De consolatione* about Christianity led to conjectures, beginning in the eighteenth century, that perhaps the *Opuscula sacra* were by a different author. More recently it has been proposed that, under the stress of his misfortunes and perhaps in disappointment that the Pope and bishops were impotent to help him, Boethius abandoned Christianity and in his last work wrote as a pagan apostate. The manner of the citation from the Wisdom of Solomon makes the last conjecture inherently unlikely, and the old assumption that *De consolatione* presents Neoplatonic philosophy while the *Opuscula sacra* do not fails to withstand critical examination. In any event a fragment of Cassiodorus records subjects on which Boethius had written, and these include 'a book on the Holy Spirit and some dogmatic chapters and a book against Nestorius'. Cassiodorus, who succeeded Boethius immediately as Master of the Offices, clearly understood him to be a believer. The intention of the *De consolatione philosophiae* is to explore the philosophical arguments in

defence of belief in providence, where, if there is authority, it is located in Plato. Boethius had certainly read the three studies of providence written by Proclus, and in *De consolatione* I pr.4.30 cites loosely from Proclus' commentary on Plato's *Parmenides*: 'If there is a god, whence comes evil? Whence comes good, if there is not?' There are also themes from Proclus in the *Opuscula sacra*. In writing on the Trinity, Boethius uses the analysis of the relation between identity and difference discussed by Plato (§§15-16) in the *Sophist* and the *Parmenides* and thereafter by Proclus in his commentary on the *Parmenides*. The face that meets us in the *Consolatione* is identical with that of the *Opuscula sacra*.

Of the five *opuscula* on Christian theology, the fourth 'on Catholic Faith' stands apart and, as early as Reginbert of Reichenau in the ninth century, has been thought by some to belong to a different author. However, the manuscript tradition shows that it formed part of the corpus from a very early stage, and its diction is demonstrably Boethian. It states what believers accept on authority, succinctly summarizing Augustinian doctrine, and contains no allusion to the logical problems which mark the other four *opuscula*. The tractate 'Against Eutyches and Nestorius' is among Boethius' most original pieces, addressed to the confusions of thought and language in the debate about the unity of the person of Christ which, since the council of Chalcedon in 451, had split the Greek churches. Successive attempts at formulas of reconciliation angered the popes by implications that the controversy had not been fully settled by Pope Leo the Great. Boethius discerned confusion over the terms 'nature' and 'person', defining the latter as 'the individual substance of a rational nature'. In his second commentary on Aristotle's *De interpretatione*, 'person' is the incommunicable quality of a human individual. The third tractate answers questions put by John (probably the future Pope John I) arising out of another paper entitled *Hebdomads (Groups of Seven)*. A highly Neoplatonic piece, without theology, it inquires 'how substances are good in that they exist, yet are not substantial goods'. The influence of Proclus is pervasive. The second tractate on the Trinity illuminates the belief that God is one, yet three (in a way that Augustine thought hard to define in the traditional terms such as person or substance) by using the observation that to say two entities are 'the same' implies some distinction between them. Identity and difference are mutually related concepts.

De consolatione philosophiae reflects a mind soaked not only in classical Latin poetry but also in the arguments and philosophical methods of Aristotle and his Neoplatonic commentators. Boethius' declared programme was to translate all the works of Plato and Aristotle. He was persuaded by the contention of Porphyry (mirrored also in Augustine's *Contra academicos (Against the Academicians)*) that in the basic essentials Plato and Aristotle were not in disagreement. Dialectic had long shared with grammar (that is, the study of literature) and rhetoric an important place in the trivium of elementary education. However, dialectic could claim more, even that it was an indispensable instrument in all mental skills. The 'liberal arts' (the skills necessary for the education of a gentleman) included not only the trivium but also the four mathematical 'arts', arithmetic, music, geometry and (theoretical) astronomy. For these four, Boethius coined the term 'quadrivium'. In Platonic and Pythagorean education these were skills which trained the mind to cope with immaterial realities.

Boethius' intention to translate Plato came to nothing, but from 504 onwards he did serious work on Aristotelian logic. Porphyry's *Isagōgē(Introduction)* to this subject had been translated into Latin by Marius Victorinus, an African who taught in Rome in the middle years of the fourth century and who had also translated Aristotle's *Categories* and some other works of Porphyry and Plotinus. The *Isagōgē*, written in the late third century (see Porphyry), was the standard preface to the study of Aristotelian logic and was generally used in the Neoplatonic schools; Greek Neoplatonists wrote commentaries on it. Boethius used Victorinus' version as his basis for a commentary but, as he wrote, became increasingly and justifiably critical of it; so, for a second commentary, he produced his own version. He also produced his own translations of Aristotle's *Categories* and *De interpretatione*, and provided them with commentaries. In the case of *De interpretatione* there were two commentaries, one short and simple and the other full-length. He also translated the *Prior* and *Posterior Analytics*, *Topics* and *Sophistical Refutations*, and his notes on the *Prior Analytics* have been preserved. Porphyry's first commentary on the *Categories* largely served as his model for this; Boethius does not appear to have known Porphyry's second exposition, but he had access to the commentary by his own contemporary Simplicius. Much labour has been expended on the determination of Boethius' sources, and here it must suffice to say that numerous parallels of thought and language show his intellectual milieu to be identical with that of the eastern Neoplatonists and their commentaries on the logical works of Porphyry and Aristotle. A constant feature of his translations was the desire to be meticulously accurate.

Boethius' treatises on logic overlapped in some degree with his commentaries. He wrote *De divisione (On Division)* (that is, on the classification of ideas), *De syllogismo categorico (On Categorical Syllogisms)* expounding the *Prior Analytics*, *De hypotheticis syllogismis (On Hypothetical Syllogisms)* harmonizing Aristotle and the Stoics, *De topicis differentiis (On Topical Differences)* and a commentary on Cicero's *Topics*, transmitted incomplete. Boethius also mentions some scholia of his on Aristotle's *Physics*, which have not survived. In these he probably drew on Simplicius' huge commentary.

Boethius' commentaries amass the opinions of numerous previous exegetes of Aristotle, often borrowing such matter from Porphyry who, as far as the earlier authors are concerned, had done most of the work already. One cannot assume that Boethius had himself read every author he cites, but it is probable that he had access to manuscripts of others beside Porphyry and Proclus. J. Shiel (1958) ingeniously but controversially suggested that he may have worked from a single codex of Aristotle's Organon with wide margins filled with summaries of the opinions of commentators, in the manner of the contemporary biblical *catenae*. It is in any event non-controversial that as a logician, Boethius did not set out to be original or independent of the Greek authorities, whose work he was 'sweating' to make available to the Latin world. At the same time he was something more than a simple transcriber of others, and wrote as a man personally engaged by the logical problems on which he was writing. The historian of ancient logic has good reason to be grateful for all that he preserves of which otherwise there would be insufficient or even no record. The treatise *De hypotheticis syllogismis* is more informative than any other ancient source about this part of Aristotelian dialectic. His contemporaries did not manifest much gratitude for his educational labours, and he often had occasion to refer to unkind critics. In his commentary on the *Categories*, he remarked candidly that with the general neglect of the liberal arts, much of the knowledge acquired by past generations would soon be lost. Nevertheless, we hear also of a small circle of influential admirers.

The curriculum of Neoplatonic education influenced him to produce adaptations of the *Arithmetic* of the second-century Pythagorean Nicomachus of Geras (Iamblichus' exegesis of which is extant), and of Nicomachus' treatise on music, the latter with some additional dependence on the *Harmonics* of Ptolemy (extant also). In both cases, Boethius' Greek models survive. The Neoplatonists accepted the traditional view that the study of the mathematical arts prepares the mind for the contemplation of immaterial realities. They followed Plato's *Timaeus* in seeing harmonic ratios and exact proportionality located in the very structure of the cosmos: for them, the distance of the planets from the earth was determined by musical theory (see Plato §16). One does not go to Boethius' work on music to learn anything of the practice of making ordered sound. His latent theme is the providential harmony of the heavens and the seasons, the mathematical principles that operate in music and hold the diverse elements together in the grand consonance of the world-soul, and that are therefore a clue to the secret concord of God and nature in a world where the only source of discord is the evil in the human heart.

Cassiodorus records an introduction to geometry by Boethius, but the work has not survived. Medieval writers filled the gap in his name. The early sections of *De consolatione philosophiae* imply some serious study of astronomy, perhaps even that his expert knowledge in this field had been brought into the accusations against him, as if he had been using astrological almanacs to predict the future succession after Theoderic. A sentence in a letter drafted by Cassiodorus may mean that Boethius had made some adaptation of Ptolemy on astronomy, but there is no trace of the work surviving.

3 Logic

In presenting Aristotelian logic to the Latin world, Boethius' first task was to explicate the *Isagōgē* of Porphyry, who was interested in showing the compatibility of Aristotle's dialectic with Platonic metaphysics. The book Porphyry wrote on that agreement is lost, but its influence is apparent in Augustine and in the Greek Neoplatonic commentators (see Neoplatonism). He was concerned to rebut Christian polemic against the disagreements of the philosophical schools and to establish agreement among his own authorities. A subject on which Platonists and Aristotelians differed was that of universals. Are they prior to particulars, or is it the other way round? If nothing exists beyond particulars, must not the mind find knowledge impossible? At least the mind needs to have the capacity to see together things that are somehow linked, and to hold related particulars together under a common species or genus (see Particulars; Universals). The issue required some treatment, especially in Boethius' second commentary on Porphyry; but he declined to give any verdict between Plato and Aristotle, saying that this question was one for more advanced inquiries. At least he was able to follow the Peripatetic Alexander of

Aphrodisias, who was ready to grant that if the universal has no existence, there can be no particulars.

Boethius' commentary on the *Categories* has a preface to Book II which firmly dates the work in 510, Boethius' consular year. He expresses regret that his public social duties have interfered with his educational task. He takes Aristotle's intention to be the examination of verbal distinctions and the provision of an elementary introduction to lead on to the higher truths of his metaphysics. Here and throughout, dependence on the extant but incompletely transmitted commentary by Porphyry is evident. There is also, however, an explicit use of material taken directly or indirectly from the lost commentary by Iamblichus, whose work was also drawn upon in the surviving exegesis by Proclus' pupil Simplicius.

Aristotle's *De interpretatione* was regarded in antiquity as an exceptionally opaque work. From a mass of ancient commentaries, the two by Boethius and that of Ammonius of Alexandria are the principal survivors. The work of Porphyry is once again a dominant guide. A substantial part of Boethius' second commentary concerns the tracing of interconnections between an object (*res*), thinking about that object (*intellectus*), talking about it (*vox*) and then putting the spoken thought into writing (*litterae*). The second commentary also discusses a range of questions that had long presented difficulty. For instance, people can have the same general conceptions of justice and goodness; but if they then differ on what in particular is just and good, does that establish moral relativism or merely human fallibility? Aristotle's ninth chapter (the sea-fight tomorrow) provoked a large debate on future contingents and modal logic. In using the terms '*contingens*' and '*contingentia*', Boethius was anticipated by Marius Victorinus. The words were needed for matters which might be other than they are, especially where the chances are even either way.

Boethius shared the Peripetatic aversion to the Stoics' surrender to determinism (see Stoicism). Necessity controls the heavenly bodies; but in this world, human wills are capable of acting after rational deliberation. In Aristotle's vocabulary, 'necessity' is a term to use for what is invariably the case, and therefore not for individual historical events. The debate about determinism, partly paralleled in Alexander of Aphrodisias (in *On Fate*), has a theological bearing, partly because of belief in oracles and prophecy. Do inspired prophets make predictions of future events, or do they foresee trouble coming and, like a physician's prognosis or a weather forecast, utter sage warnings advising on actions to avoid? And if what they foresee is correctly discerned, is it their inspired knowing which in any sense makes the event fated to occur, or can it be said that the gods' knowledge of contingent things is itself contingent? Boethius' second commentary on *De interpretatione* therefore anticipates some of the material in the last book of the *De consolatione philosophiae*, where divine foreknowledge is a sign of future things but not in itself a cause, just as human beings can see what is going to happen without making it occur.

The treatise *De hypotheticis syllogismis* has attracted much attention, in part because of Boethius' awareness of the ambiguity surrounding conditional statements of the form 'If…, then… '. The characteristic account distinguishes hypothetical syllogisms which are affirmative (in the form, 'if A is, B is; if A is not, B is') from those which are negative (in the form, 'if A is, B is not; if A is not, B is not'). The consequent is the decider whether the hypothetical syllogism is affirmative or negative. Porphyry wrote on this subject, and perhaps Boethius made use of his work, which is not extant. The detail with which the subject is treated is significant of the strength of his commitment to the labour of convincing a society in cultural decline of the necessity of logic for clear statement: 'Those who reject logic are bound to make mistakes.' His Greek Neoplatonic masters had already taught him that in this field Aristotle was the supreme guide: '*Aristotelica auctoritas*', he writes in the second commentary on *De interpretatione* (218.26). This could readily be fitted into a Neoplatonic conception of the cosmos or even more readily into a Platonism from which Christian convictions had purged away some of the more mythological elements, such as reincarnation (see Logic, medieval).

4 *Opuscula sacra*

Debate in the Church from about AD 370 had revolved around the correct and adequate way to express (a) that Christ is both divine and human, and therefore combines 'two natures', and (b) that nevertheless Christ is one person and the union of divine and human in him is necessary to his work of redemption. How could the singleness of person be affirmed if a determination to safeguard the reality and spontaneity of Christ's humanity imposed an inextinguishable duality of natures? Boethius was not the first writer to try the use of Porphyry's logical terminology in the attempt to alleviate this formidable conundrum. Cyril of Alexandria, a prominent fifth-century critic of those who held an unmitigated duality of natures, had already employed Porphyrian terms. However, there

was still much defining and clarifying to be done. There remained thorny problems of inconsistency, for example between the meaning of terms expressing the doctrine of the Trinity and the meaning of the same terms of Christology. In both, the words ‘nature’ and ‘hypostasis’ were current, but not always in the same sense in the two contexts. Pietistic fideists who liked to pride themselves on the supernatural faith of Galilean fishermen could look down on the clever logicians who seemed to think Aristotle could better express their religious beliefs. In the West, the stress on the use of papal authority to determine controversies produced a comparable insistence both on submissive adherence to what had been canonically defined by the ecumenical council of Chalcedon (451) and confirmed (after some hesitations) by Pope Leo I, and on resistance to any further explanations or compromises made to placate the Monophysite critics of Chalcedon’s language. However, the logicians saw correctly that the methods and terms of the Neoplatonic commentators on Aristotle could do a little to clarify some obscurities and muddles (see Patristic philosophy).

Augustine (§3) had proclaimed that theologians could not safely neglect dialectic. The fifth tract in Boethius’ *Opuscula*, dedicated to John, a Roman deacon (probably the future Pope John I), was evoked by a critical moment in the debate at Rome with an immediate bearing on East-West relations in the Church and society. The then Pope, Symmachus (498-514), had execrable relations with the Byzantine emperor Anastasius, who found papal insults hard to bear, and with the patriarchs of Constantinople. Probably in 513 a Greek bishop, no doubt in the name of a group of bishops, had written to Symmachus affirming acceptance of Chalcedon’s language and approval of Pope Leo I, but in the qualified formula ‘in two natures *and of* two natures’, and rejecting the dissenting Monophysite formula ‘one nature after the union’. The Greek letter provoked consternation and a storm at Rome in a meeting attended by Boethius and other senators. Reflection on the incapacity of those present to define their terms led him to write the fifth tractate, entitled ‘Against Eutyches and Nestorius’ (Eutyches being for one nature, Nestorius for two, both in extreme and radical forms). The ‘in and of’ formula had been used by Pope Gelasius (492-6), but in 512-13 was unwelcome to the advisers of Pope Symmachus.

It is instructive that Boethius unreservedly welcomed the two prepositions. Close parallels to his language and method appear in Greek writers a few years after his time. Four definitions of ‘nature’ are given:

(a) ‘Nature consists of those things which, since in some way they exist, can be grasped by the mind’. (This language is found in Ammonius’ exegesis of Porphyry’s *Isagōgē*, 2.22.)
(b) ‘Nature is either that which can act or that which can be acted upon’. (Plato, *Phaedrus* 270d; *Sophist* 247d-e; Aristotle, *Topics* VI 10.148a18.)
(c) ‘Nature is the principle of motion *per se* and not as an accident’. (Aristotle, *Physics* II 1.192b20; VIII 4.255a32-; Proclus, *Commentary on Plato’s Timaeus* I 2.20.)
(d) ‘Nature is the specific difference imparting form to each individual thing’. (Aristotle, *Physics* II 1.193a28-31; Alexander of Aphrodisias, *Commentary on Aristotle’s Metaphysics* p.357. 25-.)

A definition of ‘person’ is more difficult. Person is subordinate to nature, as particular to general. ‘Person’ can be used only of a substance, and then only of a being endowed with mind and reason. Christian usage applies the term *persona* to God, human beings and angels. Therefore the definition of person is ‘the individual substance of rational nature’. This is the equivalent of the Greek term *hypostasis*. The Greek *ousia* (being), on the other hand, is used of the universal, with *hypostasis* used of the particular. The main thrust of Boethius’ argument is to show that ‘nature’ and ‘person’ are very distinct terms, and therefore that ‘two natures’ does not imply two persons and ‘one person’ does not imply one nature (in other words, denying the propositions associated with Nestorius and Eutyches, respectively). In connection with the Incarnation, it is impious absurdity to think either that the immutable God can be transformed into humanity or that the reverse is possible (see Incarnation and Christology §1). Change is possible only to entities which share a common substrate of matter (Aristotle, *On Generation and Corruption* 226a10). Nothing corporeal in genus can fall under an incorporeal species. Nevertheless, Boethius concedes a place to some of the language of devotion invoked by the critics of Chalcedon’s ‘two natures’, who wanted to proclaim the worship of the one person of Christ; so, he can say Christ is ‘one of the Trinity’, ‘God suffered’, and therefore it is acceptable to say ‘both in and of two natures’. The effect of Boethius’ argument, (which must have been unwelcome to Pope Symmachus and perhaps to his successor Hormisdas (514-523)), is to justify a middle path between the extremes, defending Chalcedon’s preposition *in*, yet also conceding the *of* designed to reconcile its more moderate critics. Thus, Boethius was advocating a position congenial to a powerful group at Constantinople and anticipating the policies of Justinian, who became Byzantine emperor a year or two

after Boethius' death but who was highly influential before his accession.

The first of the *Opuscula*, 'How the Trinity is one God not three', juxtaposes Catholic orthodoxy with Aristotle's map of ascending human knowledge (*Metaphysics* 1026a18), also set out early in Boethius' first commentary on Porphyry's *Isagōgē*. The study of nature deals with concrete physical entities. Mathematics is a half-way house towards the immaterial abstract realm of theology, where God is form without matter, one substance, indeed Being itself. In God there is no number, no multiplicity; and if 'substance' is a term applied to God, that must imply no substrate to accidents. Following Marius Victorinus, Boethius affirms God, Father, Son and Spirit to be *et ipse et idem* (the same, not identical). The sameness does not exclude differentiation in terms. The ten categories of Aristotle apply only to the realm of sense, not to that of divine forms. God transcends time and space; the everlastingness of the cosmos is other than the eternity of God. The one category which helps the speculative theologian is relation, since 'father' and 'son' are relational terms. The conventional term 'person' is in this context a source of confusion (as Augustine had already said). However, the language about sameness and differentiation within identity found in the *Parmenides* and in Proclus' commentary on it can offer more illumination. Sentences about the contrast between divine eternality and the everlasting temporal duration of the cosmos anticipate the section on the same subject in the *De consolatione* (V pr.6) and are closely related to the propositions in Plotinus III 7.

5 *De consolatione philosophiae*

By using the form of alternate verse and prose, Boethius was able to deploy all his intellectual resources, literary and philosophical. The work takes the form of a dialogue between Boethius, in a prison which is symbolic of his spiritual condition, and a visionary lady who is a personification of Philosophy and addresses her sick patient in discourse designed to give him fortitude in the indignity and injustice that have befallen him. The setting owes a conscious debt to the *Phaedo* of Plato, where Socrates' prison discourse (I pr.3) contrasts the immortality of the soul with the trivialities of life in a mortal material body and enables him to face the death sentence with total serenity. The dank cell and chains are symbolic of Boethius' all too earthbound mind. The lady Philosophy has a robe embroidered with a ladder connecting the initials Ξ and Π (for 'practical' and 'theoretical', or contemplative), and she will lead him upwards on a gradual ascent. The Π is also the letter on his prison dress standing for Thanatos: he is to die. Gradualism is essential to good medical treatment, and she cannot apply all her remedies at once. The verse passages ease the reader's path but are not presented as light entertainment; the Muses of amusement and love (in his youth Boethius wrote some erotic poems) are sharply dismissed. The poetic sections are in content integrated with the prose argument. In the first four books all the verse is uttered by Boethius; from V.3, Philosophy becomes the poetic speaker, adopting this form of address as a concession to his still weak and frail spiritual condition. However, the verses given to Boethius are also an ascent of the soul. At the central hinge-point of the whole work the famous poem *O qui perpetua* ('You who govern the world by perpetual reason… ') versifies the cosmology of Plato's *Timaeus* interpreted with the help of Proclus' commentary on that dialogue (III m.9).

Book I states Boethius' problem; an innocent person is suffering gross injustice. He complains of the way in which, as in the case of Seneca, a brilliantly successful career at the summit of the administration of government has been shattered by treacherous colleagues among his fellow Roman aristocrats and by the cruel tyranny of Theoderic. However, as the work proceeds Boethius himself fades to become a secondary and background figure, while the real exposition is that of Philosophy. She has to remind him of what he has forgotten, namely that power, wealth and honour are secondary matters at best, useful if they enable the possessor to do good to others but otherwise irrelevant to spiritual wisdom.

The second book uses mainly Stoic themes which had already been domesticated with the Neoplatonic scheme of things in writers such as Hierocles and Proclus. The pain of loss is made endurable by a psychological process of adjusting oneself to accept with resignation what cannot be changed. Those in exile must tell themselves that the wretched place to which they have been sent is home to the people who normally live there. To brood nostalgically for a lost past happiness is a peculiarly awful form of misery. Worldly secular honours are precarious and transient, dependent on others and wholly relative to the limited society in which they happen to be held. A holder of high office becomes a nobody beyond the imperial frontier. A philosopher when insulted should remain silent and not answer back. A wise person knows that in death all human beings are equal. Book 2 ends with a poem in

praise of the love which binds together the diverse elements in the cosmos and averts disintegration. Human beings would be happy if only the love by which the stars are ruled could reign in their minds. The poem is a bridge to the third book, which moves from Stoic to Platonic vindications of providence.

The third book, after *O qui perpetua*, seeks to establish the identity of the supreme Good with Good 'than whom nothing better can be thought', employing an Aristotelian argument from the imperfection of every individual good. The ladder of goods, which cannot have an infinite number of steps, has its end in the perfect Good which is also complete happiness. A good person participates in that goodness, which justifies the language of a deification synonymous with salvation. Individual goods give happiness not piecemeal but as a single totality. Boethius sees this principle as pointing to the truth that the supreme good is the One, and from this derives the universal experience of all living things seeking to avert disintegration and to survive by maintaining unity. In the cosmos the forces for disintegration are kept in check by providence; 'whatever holds everything together is what I mean by God' (III 12.25). God is Being itself (*esse*), and evil is a deficiency of being, a nothingness, a privation belonging to an inferiority in the hierarchy of being (see God, concepts of).

The fourth book depends on Plato's *Gorgias* (470-6): the wicked are the most miserable, and punishment benefits them by purification. Goodness is an essential element in true happiness. To the question as to why providence does not reward the virtuous with good and the wicked with evil, the classic Platonic answer is that virtue is the only really good fortune and is its own reward (see Good, theories of the §1). That presupposes some freedom of choice, and that in turn raises the question of providence and fate. From Plotinus onwards, the Neoplatonists held that providence controls the higher celestial order of things, while fate is the inexorable chain of cause and effect in this lower world. Like Proclus, Boethius thinks this fate may be controlled by 'angels' (a term as much as home in pagan Neoplatonists as in Christian writers) or by the world-soul or by the stars. The lower one's position in the great chain of being, the more tightly one is bound by fate. Providence is like the unmoved centre of a great wheel or sphere (a simile also in Plotinus and Proclus). Behind all change is the unmoved mover.

The fifth and last book takes Boethius to higher flights in the discussion of providence (see Providence). He begins from Aristotle's description of chance (*Physics* II 4-5) as an event with a traceable cause, yet an event falling outside the intention of human wills whose purposes are otherwise. As in the unexpected discovery of buried treasure, the action of digging had different intentions but another chain of causation intervened. For the Neoplatonists and for Boethius, such meetings of independent lines of causation are ultimately under the care of divine providence, not of a blind determinism. 'Luck' or 'fortune' is a way of talking about this hidden power of providence to surprise us. But if that were to mean that every event occurs within a closed system of causes, where is there room for determination by free choice and deliberation? Nature does nothing in vain, and human beings are not for nothing endowed with powers of deliberation enabling them to make choices between different options. This is inherent in being a rational being. Admittedly, Boethius does not think all members of the human race possess equal powers of exercising free choice. Moreover, a Platonist who had also read some Augustine, as Boethius demonstrably had, would not think freedom of choice neutral between truth and error, right and wrong; real freedom is to be liberated from the body's downward pull into error and mistaken judgements. So there are degrees of freedom, higher to those who are contemplating the divine mind, lower to those who slide down to the world of physical senses, reaching complete loss of freedom when vice has damaged the power of reason.

Free choice implies that not all events are wholly predictable. Some are contingent; they might turn out otherwise, and that is difficult to reconcile with perfect foreknowledge. At the initial stage of discussion it seems to Boethius more implausible than it appears to the lady Philosophy to hold that human wills do not make their choices because divine foreknowledge has foreseen them and cannot err, but rather that an omniscient divine foreknowledge foresees what choices human beings will make. After all, to know somebody's character intimately is to predict very accurately how they will react to a situation. *A fortiori* this must be true of the divine being who knows all hearts. Nevertheless, if there is contingency in the choices made, divine foreknowledge will, unthinkably, be in error if the choices are taken to be certain when they are inherently uncertain. To maintain that divine foreknowledge, which by definition cannot err, excludes indeterminacy seems to abolish distinctions of moral value between virtue and vice and undermines any sense in either hope or prayer for the avoidance of disaster. (The reference here to prayer does not step outside the conventions of Neoplatonic language, and does not have to be Christian.) The problem, long ago discussed by Cicero in his book on divination, is declared to be one on which Boethius has bestowed much study. The difficulty is alleviated by recalling that to foresee an event is not

causative of its occurrence, that the finite mind is out of its depth when the foreknowledge in question is divine and, crucially, that everything which is known is grasped not according to its inherent nature or power but relative to the capacity of the knowing mind (V 4.25). Foreknowledge of an event is not a sign that an event is absolutely necessary unless there are other causes that bring it about, and to say God knows contingent events to be contingent does not have to mean that because many outcomes are possible even God cannot really know which is will be (see Determinism and indeterminism; Omniscience; Free will).

On fate and destiny, Boethius had probably read the tract by Alexander of Aphrodisias where the correctness of divine foreknowledge is safeguarded by the proposition that the divine knows the contingent to be contingent and this knowing does not make the uncertain certain. Boethius, however, preferred the argument found in Iamblichus and Proclus that divine knowledge is qualitatively quite distinct from human. Events in the temporal process are indeed known to God, but divine knowledge transcends all the successiveness of past, present, and future. Divine knowing is eternal in its mode, and embraces all things and events in simultaneity without the transience inherent in this realm of sense and becoming. To speak of a divine foreknowledge suggests to our finite minds knowing beforehand, in advance. The temporal word ‘before’ is inapplicable in the case of a time-transcending divine perfection. What is before or after for limited minds is an absolute present (but not a frozen instant) for an eternal knower whose duration is paradoxically both infinite and atemporal. Both in his *De trinitate* and in the last section of the *De consolatione philosophiae*, Boethius stresses that eternality is more than mere perpetuity or everlastingness, but is ‘the simultaneous and complete possession of life without limits’ (*Interminabilis vitae tota simul et perfecta possessio*). The definition, as Boethius himself says, derives its force from contrast with temporal and finite life. It is not presented as an original reflection, and in fact closely echoes Plotinus and Augustine (see Eternity).

Aristotle, therefore, was not mistaken to think that the world had no beginning or end (*On the Heavens* I 12.283b 26), but that does not mean it is eternal. The cosmos can be dependent for its existence on the Creator without thereby being transcendent in relation to the temporal process. This created world is not coeternal with the Creator, as Plato observed in the *Timaeus* (37d). Accordingly, God’s knowledge is ‘in the simplicity of his present’ and imposes no necessity on choices and events in time. Moreover, as Aristotle pointed out in the *Physics* (II 9), ‘necessity’ is a term with more than one meaning. There is absolute necessity, for example, that all human beings are mortal, or that the sun rises, and there is also conditional necessity: it is necessary that if a man is walking, then he are walking. For Boethius, the kind of necessity attaching to events in the foreknowledge of God is conditional, not absolute. Certain future events which are to occur by the free choice of human wills and which might occur otherwise are, in divine foreknowledge, conditionally necessary. If you change your mind and intention, you do not empty providence; divine prescience knew that that was what you were going to do. (Augustine in *De civitate Dei (The City of God)* makes the same point: our human wills are included in the causes known to God (V 9)).

6 Influence

Boethius was born into an aristocratic senatorial society concerned to see that the past achievements of Roman culture would not be lost now that the West was controlled by barbarians. He was not the only author to speak the language of Neoplatonism in a way that the medieval world would absorb; that he shared with Calcidius, Martianus Capella, and Macrobius (see Encyclopedists, medieval). Among that company, however, he was distinctive in belonging to a Christian family and in writing influential tractates on controverted theological questions. His masterpiece, *De consolatione philosophiae*, presented the theodicy of Proclus as a flying buttress to a religious statement compatible with Christian belief. This work, taken together with his translations and commentaries, taught the medieval West its first steps in logic and the meaning of inference.

In the ninth century, the Carolingian educational programme gave special prominence to *De consolatione philosophiae* as a medium for the teaching of prose and poetry, and this drew attention to the *Opuscula sacra* as well. The surviving manuscripts of Boethius written in the Carolingian age are reasonably numerous (see Carolingian renaissance). Remigius of Auxerre was first to write a commentary on the tract on the Trinity, the first of the five *Opuscula*. However, it was in the twelfth century that these theological tractates rose to a rank of authority. They also came to stimulate controversy. This was largely because of Gilbert de la Porrée, who became bishop of Poitiers and caused alarm both by his catechetical teaching and his commentaries on four of the *Opuscula* (omitting *De fide catholica*) (see Gilbert of Poitiers). Conservatively minded critics were frightened

when a clever logician began applying Boethian methods of argument to such transcendent subjects as the Trinity and the person of Christ. It made them anxious to be told that without qualification the divine essence is not to be called God, or that it was muddled to say that the divine nature was made flesh and assumed our human nature.

In the same century, Thierry of Chartres more cautiously commented only on the first of the *Opuscula* (*De trinitate*) and on the highly Neoplatonic third which the medievals called *De hebdomadibus*. Clarembald of Arras joined the critics of Gilbert, also commenting on the two tractates chosen by Thierry. The consequence of these debates, in which Thomas Aquinas was to participate in the next century, was to insert into the consciousness of the Latin West a sharper awareness of some problems inherent in the logic of traditional Christian language about God, and at the same time to arouse among the less dialectical faithful a sense of disapprobation of theologians who knew more about Boethius than about the apostles.

See also: Encyclopedists, medieval §5; God, concepts of; Logic, medieval; Medieval philosophy; Natural philosophy, medieval §1; Neoplatonism; Patristic philosophy; Platonism, medieval; Plotinus; Porphyry

HENRY CHADWICK

List of works

Boethius, A.M.S. (*c*.480-525/6) translations of Aristotle, ed. L. Minio-Paluello, *Aristotles Latinus*, vols I-VI (vol. VI ed. B. Dod), Bruges: Desclée de Brouwer, 1966-75.(Collected editions of Boethius' translations of Aristotle's works.)

Boethius, A.M.S. (*c*.480-525/6) *In Isagogen Porphyrii commenta (Commentaries on Porphyry's Isagōgē)*, ed. S. Brandt, Corpus Scriptorum Ecclesiasticorum Latinorum 48, Vienna: Tempsky, 1906.(Includes the commentaries based on the translation by Marius Victorinus and that of Boethius himself.)

Boethius, A.M.S. (*c*.480-525/6) Commentary on Aristotle's *Categories*, in J.-P. Migne (ed.) *Patrologia Latina*, vol. 64, Paris, 1860.(Boethius' commentary uses the translation of Marius Victorinus as its base; he also produced his own translation.)

Boethius, A.M.S. (*c*.480-525/6) Two Commentaries on Aristotle's *De interpretatione*, ed. R. Meiser, *Anicii Manlii Severini Boetii in librum Aristotelis Peri hermeneias*, Leipzig: Teubner, 1877-80, 2 vols.(One commentary is full-length, the other is shorter and simpler.)

Boethius, A.M.S. (*c*.480-525/6) Notes on *Prior Analytics*, in L. Minio-Paluello (ed.) *Aristoteles Latinus*, vol. III, Bruges: Desclée de Brouwer, 1965.(Notes which accompanied Boethius' translation of the *Prior Analytics*.)

Boethius, A.M.S. (*c*.480-525/6) *In Ciceronis Topica (Commentary on Cicero's Topics)*, in J.C. Orelli and J.G. Baiter (eds) *Ciceronis Opera*, Vol. V, part 1, Zurich: Fuesslini, 1833; trans. E. Stump, *Boethius's In Ciceronis Topica*, Ithaca, NY: Cornell University Press, 1988.(This commentary has been transmitted incomplete.)

Boethius, A.M.S. (*c*.480-525/6) *De hypotheticis syllogismis (On Hypothetical Syllogisms)*, ed. L. Obertello, Brescia: Paideia, 1969.(A Greek paraphrase by Maximos Holobolos (*fl.* 1261-80) is edited by D.Z. Nikitas in *Eine byzantinische Übersetzung von Boethius' 'De hypotheticis syllogismis'*, Göttingen: Vandenhoeck & Rupprecht, 1982.)

Boethius, A.M.S. (*c*.480-525/6) *De divisione (On Division)*, in J.-P. Migne (ed.), *Patrologia Latina*, vol. 64, Paris, 1860; trans. E. Stump and N. Kretzmann in *The Cambridge Translations of Medieval Philosophical Texts* I, Cambridge: Cambridge University Press, 1988.(On the classification of ideas.)

Boethius, A.M.S. (*c*.480-525/6) *De syllogismo categorico (On Categorical Syllogism)*, in J.-P. Migne (ed.) *Patrologia Latina*, vol. 64, Paris, 1860.(Expounds on the *Prior Analytics*.)

Boethius, A.M.S. (*c*.480-525/6) *Introductio ad syllogismos categoricos (Introduction to Categorical Syllogisms)*, in J.-P. Migne (ed.) *Patrologia Latina*, vol. 64, Paris, 1860.(Shorter work on categorical syllogisms.)

Boethius, A.M.S. (*c*.480-525/6) *De topicis differentiis (On Topical Differences)*, in J.-P. Migne (ed.) *Patrologia Latina*, vol. 64, Paris, 1860; trans. E. Stump, *Boethius's De topicis differentiis*, Ithaca, NY: Cornell University Press, 1978.(Boethius' own writings on topical differences.)

Boethius, A.M.S. (*c*.480-525/6) *Opuscula sacra (Theological Tractates)*, ed. and trans. H.F. Stewart, E.K. Rand and S.J. Tester, *Boethius: Theological Tractates, Consolation of Philosophy*, Cambridge, MA: Harvard University Press, 1973; ed. and trans. M. Elsässer, *Die Theologischen Traktate*, Hamburg: Felix Meiner, 1988; trans. H. Merle, *Court traités de theologie*, Paris: Éditions du Cerf, 1991.(Elsässer's German translation has good notes. Merle's French translation is annotated, without the Latin text.)

Boethius, A.M.S. (*c*.480-525/6) *De consolatione philosophiae (The Consolation of Philosophy)*, ed. and trans. H.F. Stewart, E.K. Rand and S.J. Tester, *Boethius: Theological Tractates, Consolation of Philosophy*, Cambridge, MA: Harvard University Press, 1973; ed. L. Bieler, Corpus Christianorum, Series Latina 94,

Turnhout: Brepols, 1957; ed. J.J. O'Donnell, *Boethius Consolatio Philosophiae*, Bryn Mawr Latin Commentaries, Bryn Mawr, PA: Bryn Mawr College, 1984.(Part of Books IV and V are edited with English translation and good commentary by R.W. Sharples in *Cicero: On Fate (De Fato) & Boethius: The Consolation of Philosophy (Philosophiae consolationis) IV 5-7, V*, Warminster: Aris & Phillips, 1991. For an English translation only, see V.E. Watts, *The Consolation of Philosophy*, Baltimore, MD: Penguin, 1969.)

Boethius, A.M.S. (*c*.480-525/6) Introductions to *Arithmetic* and *Music*, ed. G. Friedlein, *De institutione arithmetica libri duo, De institutione musica libri quinque*, Leipzig: Teubner, 1867.(Boethius' works are adaptations of earlier works by Nicomachus of Geras; see §2.)

References and further reading

Chadwick, H. (1981) *Boethius, the Consolations of Music, Logic, Theology and Philosophy*, Oxford: Clarendon Press; Italian translation, Bologna: Mulino, 1986. (Describes Boethius' works in relation to the historical context of his life and to the intellectual background of his philosophical writings, with bibliography.)

Courcelle, P. (1967) *La Consolation du philosophie dans la tradition littéraire, antécédents et posterité de Boèce (The Consolation of Philosophy in Literary Tradition, Antecedents and Influence of Boethius)*, Paris: Études Augustiniennes. (Demonstrates the neoplatonic character of the *Consolatio*, and traces some of its later influence; a work of great erudition.)

Daley, B.E. (1984) 'Boethius' Theological Tractates and Early Byzantine Scholasticism', *Medieval Studies* 46: 158-91.(On the relation between the Tractates and the arguments of Byzantine schoolmen.)

Ebbesen, S. (1990) 'Boethius as an Aristotelian commentator', in R. Sorabji (ed.) *Aristotle Transformed*, London: Duckworth, 373-92.(Clear analysis by a logician.)

Fuhrmann, M. and Gruber, J. (eds) (1984) *Boethius*, Wege der Forschung 483, Darmstadt: Wissenschaftliche Buchgesellschaft.(Gathers twenty-three papers of major importance from journals.)

Gibson, M.T. (ed.) (1981) *Boethius, His Life, Thought and Influence*, Oxford: Blackwell.(A symposium of different authors, more on Boethius' influence than on his contemporary setting.)

Gibson, M.T. and Smith, L. (eds) (1995) *Codices Boethiani: A Conspectus of Manuscripts of the Work of Boethius*, London: University of London, The Warburg Institute.(Illuminating on Boethius' influence on medieval culture.)

Gruber, J. (1978) *Kommentar zu Boethius de Consolatione Philosophiae (Commentary on Boethius' Consolation of Philosophy)*, Berlin: de Gruyter.(Masterly on the literary side, indispensable on the detail of its structure.)

Kneale, W.C. and Kneale, M. (1981) *The Development of Logic*, Oxford: Clarendon Press, 1981.(Boethius' place in the history of logic.)

Kretzmann, N. (1987) 'Boethius and the Truth about Tomorrow's Sea Battle', in *Logos and Pragma, Essays on the Philosophy of Language in Honour of Gabriel Nuchelmans*, Nijmegen, *Aristarium* suppl. 3, 53-87.(On the exposition of Aristotle's *De interpretatione* IX.)

Lerer, S. (1985) *Boethius and Dialogue: Literary Method in the Consolation of Philosophy*, Princeton, NJ: Princeton University Press.(Applies modern techniques of literary criticism.)

Magee, J. (1989) *Boethius on Signification and Mind*, Philosophia Antiqua 52, Leiden: Brill.(Detailed examination of the commentaries on Aristotle's *De interpretatione*.)

Marenbon, J. (1983) *Early Medieval Philosophy (480-1150), An Introduction*, London: Routledge & Kegan Paul. (Clear and short.)

Masi, M. (ed.) (1981) *Boethius and the Liberal Arts*, Berne: Peter Lang.(A valuable symposium.)

Minnis, A.J. (ed.) (1987) *The Medieval Boethius*, Cambridge: D.S. Brewer.(Important studies of the versions of the *Consolatio*.)

Obertello, L. (1974) *Severino Boezio*, Genoa: Accademia Ligure di scienze e lettere, 2 vols.(Authoritative general study with full bibliography.)

O'Daly, G.P.D. (1991) *The Poetry of Boethius*, London: Duckworth.(A major study of the verse sections of *De consolatione philosophiae*.)

Patch, H.R. (1935) *The Tradition of Boethius*, New York: Oxford University Press.(On the influence of Boethius.)

Reiss, E. (1982) *Boethius*, Boston: Twayne.(Reliable biography.)

Rijk, L.M. de (1964) 'On the Chronology of Boethius' Works on Logic', *Vivarium* 2: 1-49.(Gives the chronology of Boethius' logical works.)

Schurr, V. (1935) *Die Trinitätslehre des Boethius im Lichte der Skythischen Kontroversen (The Trinitarian Doctrine of Boethius in the Light of the Scythian Controversies)*, Paderborn, 1935.(A pioneer exploration of the Aristotelian and Neoplatonic background of the *Opuscula sacra*.)

Shiel, J. (1958) 'Boethius' Commentaries on Aristotle', *Medieval and Renaissance Studies* 4: 217-44; repr. in M.

Fuhrmann and J. Gruber (eds) *Boethius*, Wege der Forschung 483, Darmstadt: Wissenschaftliche Buchgesellschaft, 1984; repr. in R. Sorabji, (ed.) *Aristotle Transformed*, London: Duckworth, 1990.(Acute conjectures about Boethius' philosophical works.)

Sorabji, R. (1980) *Necessity, Cause and Blame*, London: Duckworth.(Analysis of questions basic to Boethius' writings on *De interpretatione*.)

Sorabji, R. (ed.) (1990) *Aristotle Transformed*, London: Duckworth.(On the Neoplatonic commentaries.)

Stump, E. and Kretzmann, N. (1981) 'Eternity', *The Journal of Philosophy* 78: 429-57.(Careful philosophical analysis of the idea.)

Bogdanov, Aleksandr Aleksandrovich (1873-1928)

Aleksandr Aleksandrovich Bogdanov, né Malinovskii, was a Russian thinker who helped Lenin create the Bolshevik or Communist Party, broke with Lenin over a mixture of philosophical and political issues, yet would not quit the Revolution. His life and thought illuminate the interaction of philosophy and politics within the tumultuous context of a 'developing' country, which calls in question political philosophies that take for granted the conditions of 'developed' countries.

Bogdanov never won such widespread interest as those dissident communists - Georg Lukács, most notably - who turned Marxism away from claims of science towards theories of consciousness and wilful action. Bogdanov sought a positivist basis for his philosophy of action or practice. He offered 'empiriomonism' and 'organizational science' to creators of a 'free collectivism', but the creators of the Soviet system brushed him aside. He has been studied by scholars who wonder why the Russian Revolution - or twentieth-century revolutions in many developing countries - have failed to realize dreams of justice and freedom, and by a different cluster of scholars who conceive of a metascience that might unify the fragmented world of knowledge. Less known is Bogdanov's sense of tragic contradictions in revolutionary pragmatism, as we may call active belief in Marx's famous declaration that the point of philosophizing is not merely to interpret the world but to change it.

1 Life and thought

'Spiteful stupid administrators', who ran a provincial boarding school as a 'barracks, a jailhouse', taught young Bogdanov to 'fear and hate those who rule, to reject authorities', he reports in his 'Avtobiografiia' (Autobiography) (1926). Pragmatism - in the mundane sense of adapting to things as they are - must have tempered his fearful hatred, for he won a gold medal from those administrators, and then a degree in natural science at Moscow University, though revolutionary activity caused his student career to be interrupted by arrest and banishment from the capital. He also earned a medical degree, and intermittently worked as a physician for the rest of his life. Constant involvement in revolutionary propaganda brought repeated periods in jail and banishment, which gave him leisure to study and write. At the outset he was a populist of the 'People's Will' persuasion, believing in heroic action as the way to a just and prosperous Russia, but he soon turned to the Marxist or positivist belief that science reveals the way forward. He began a lifelong outpouring of publications with *Kratkii kurs ekonomicheskoi nauki (A Short Course of Economic Science)* (1897), which won high praise from the young Lenin and became a widely used textbook in Marxist circles.

Acting as a propagandist among provincial workers, Bogdanov discovered their need for a worldview, and responded with *Osnovnye elementy istoricheskogo vzgliada na prirodu (Basic Elements of the Historical View of Nature)* (1899). The most basic 'element' in this view was energy, Ostwald's substitute for the matter-and-force of mechanical or vulgar materialism. This book also won the admiration of Lenin, who failed to perceive the difference from the materialism, dialectical and historical, which he admired in the works of Plekhanov, 'the father of Russian Marxism'. By the time that Lenin learned the difference - talking with Plekhanov in Swiss exile - Bogdanov and he had become leaders of the emergent Bolshevik faction within the Russian Social Democratic Party. Though Plekhanov was a major figure in the rival Menshevik faction, Lenin still considered him the greatest teacher of orthodox Marxist philosophy. Bogdanov sneered at such orthodoxy as a fetish, not the philosophy of science that it claimed to be. That he found in a wide array of thinkers, including Nietzsche, and especially Mach, whose empiriocriticism he refashioned as empiriomonism, to unify all types of knowledge and action (see Russian empiriocriticism §2).

This disarray in political and philosophical commitments did not prevent Lenin and Bogdanov from joint action through the first Russian revolution in 1905-6. Their Bolshevik faction pressed the Russian Social Democratic Party towards more aggressive policies than the Mensheviks favoured, and thereby pushed towards a formal split into separate parties. When the Tsar's government had succeeded in putting down the first Russian Revolution, and Russian Marxists retreated from the politics of mass action to the quarrels of underground cells and exile circles, Lenin wrote *Materializm i ėmpiriokrititsizm (Materialism and Empiriocriticism)* (1909). It was a vituperative attack on Bogdanov and the other Marxists who had been drawing on Mach to remedy the movement's lack of proper philosophical grounding. Bogdanov responded with a counterattack on Lenin, *Padenie*

velikogo fetishizma; Vera i nauka (The Fall of a Great Fetishism; Faith and Science) (1910), as he had earlier responded to Menshevik attacks with *Prikliucheniia odnoi filosofskoi shkoly (The Adventures of a Philosophical School)* (1908a). Efforts to draw German notables into these disputes provoked illuminating refusals. In a private letter Mach declared his 'social democratic' sympathies to be separate from the philosophizing that he did in public, while Kautsky, the chief theorist of German Marxism, saw things the other way round: not politics but philosophy was a matter of individual taste - separate from the 'public sphere', he might have said, had he known a buzzword that German Marxism would generate later on.

One common way to sort out political from philosophical issues in the polemics of Russian Marxists is to blame Lenin's fanaticism for entangling two realms of discourse that would otherwise have been as civilly separated in Russia as in the West. This interpretation ignores the pattern of entanglement that preceded Lenin's outburst, and caused it. Empiriocriticism was linked with Bolshevik theorists, materialist orthodoxy with Mensheviks, and the correlation seemed logical. To ground Marxism in historical materialism, pointing to long-term social processes that determine human thought and action, seemed to support the go-slow politics of the Mensheviks, while emphasis on the active role of collective mentalities in ordering experience seemed to favour the urgent push to revolt that the Bolsheviks advocated. That scheme, presented to Western comrades by Bogdanov's article, 'Ernst Mach und die Revolution' (Ernst Mach and the Revolution) (1908b), in the major journal of German Social Democracy, was becoming conventional wisdom in Marxist circles. Lenin's outburst was designed to prove that his party was *not* revisionist but orthodox, as all true Marxists must be.

At a showdown meeting of Bolshevik leaders in 1909 Bogdanov and Lunacharskii mocked the push to establish a party line in philosophy. Lenin denied that he was doing any such thing, while pressing for expulsion of the unorthodox. Bogdanov tried to gain acknowledgement of the inconsistency, in vain. He was expelled from the Bolshevik organization, and intensified his warning against 'fetishism' - formulaic idolatry in place of creative theory - and against 'vampirism' - the desiccation of revolutionary thought by worship of formulas. His polemics against Lenin and Plekhanov declared them to be victims of 'the Vampire', which appeared in a novel of Bogdanov's as a hectoring monster within the revolutionary hero's mind, in a melodramatic imitation of the Devil appearing to Ivan Karamazov.

'Dilettantism' was Trotsky's diagnostic label for superficial pretensions of unifying knowledge and action, when knowledge is actually fragmented in professional disciplines apart from the Marxists' realm of political action. Trotsky himself was hardly immune. He wrote on almost as many topics as Bogdanov, from a supposedly Marxist viewpoint, and accepted an invitation to lecture at a school for revolutionaries that Bogdanov's group established on Capri, where the famous writer Gor'kii provided a haven. Bogdanov soon found there the same 'authoritarianism' that he opposed in the party high command. In a letter of resignation he noted his lifelong struggle 'against two enemies, authoritarianism and individualism, and the former is the more hateful to me'. By 1911 he announced his withdrawal from politics altogether, in favour of 'cultural' and 'scientific' work. He meant separation from any political party; his understanding of culture was thoroughly politicized in the broad sense of both terms.

2 Life and thought (cont.)

Brooding over the revolutionary passions of the lower classes, which promised imminent overthrow of the existing system, while their cultural backwardness portended their inability to use power effectively, Bogdanov began preaching the need to develop 'proletarian culture'. He was thinking by analogy with the flowering of bourgeois culture *before* the French Revolution brought the bourgeoisie to power. Lenin expressed a more wilful belief in practice as the criterion of truth. He asked why the Party might not mobilize lower-class anger to overthrow the rule of landlords and capitalists, and *then* have the 'cultural revolution' that would enable any cook to manage public affairs. Both men perceived the lower-class nature of the imminent revolution, and both took it for granted that a vanguard party must guide the masses, but Bogdanov was concerned to foster a 'free collectivism' at every level. Hence his paradoxical efforts to combine a sceptical epistemology with a totalizing worldview.

That dream was central both to his philosophy of 'empiriomonism' and to his 'organizational science' or 'tektology'. He emphasized the philosophy in publications of 1903-9, the science afterward, as he 'moved to a decisive recognition of "the decline of philosophy", of transition to construction of a purely scientific monism' ([1913] 1923: 328). The 'purely scientific monism', expounded in three volumes of *Tektologiia: Vseobshchaia*

organizatsionnaia nauka (Tektology: The Universal Organizational Science), (1913-25), drew its title from Haeckel, and its central thesis from Le Chatelier. Assuming 'equilibrium' to be inherent in all systems, mental and social as well as biological and mechanical, Bogdanov offered rules of 'organization' that could serve for scientific management in all spheres.

In 1917 mass revolt broke up the tsarist system and opened the way for Lenin's establishment of a Soviet state ruled by the Communist Party. The quarrels of émigré intellectuals shrank in significance and dissidents rejoined the Party to serve the victorious revolution, but Bogdanov declined to do so. He published gloomy criticisms of Lenin's policies in 1917-18, which he never retracted, though he stayed in Russia to help the cause. In a letter to Lunacharskii, declining the offer of a post in the Commissariat of Enlightenment, Bogdanov explained why he could neither join the new regime nor work against it. He found a repulsive distortion of socialism 'in the often absurd but almost always *compulsory* things that are done in your regime. I not only see, but think that *you* see the tragic quality [*tragizm*] of your situation' (1990: 352; original emphases). He summed up the tragedy with an aphorism - 'A situation is often stronger than logic' (1990: 355) - which implicitly challenged his lifelong faith in pragmatic reason fused with collective action, but he left the implication unexamined. He simply chose to stay in 'the wearisome isolation of a sighted person among the blind', doing 'cultural and scholarly' preparation for revival of a genuinely socialist party. One may call that a tragic version of revolutionary pragmatism, with only a trace of the usual reassurance that present defeat must surely turn into future triumph since scientific revolutionaries know what the future must bring. Bogdanov's philosophizing about knowing and acting usually emphasized some version of that reassurance - most melodramatically through the mouth of a novelistic heroine: 'In order to wage the struggle one must know the future', which Bogdanov's *Krasnaia zvezda (Red Star)* (1908) laid out in utopian detail. But he could also express dark suggestions that the struggle for justice and freedom may be a losing gamble.

For a time, 1918-20, he was a leader of 'Proletarian Culture' or Proletcult - a network of study circles, art studios and theatre companies, bringing traditional knowledge and skills to the masses while also developing their capacities to overcome the hegemony of bourgeois culture. Lenin was initially sympathetic, but he insisted that the movement must accept subordination to the new regime. Proletcult refused, and was shut down. Bogdanov thereafter confined his activities to the newly established Communist Academy, which published his works in philosophy and social science, and to experiments in blood transfusion at a new institute directed by himself. An official campaign against his philosophy during the drive to suppress Proletcult had put 'Bogdanovism' out of bounds for Soviet Marxists even in the relatively free 1920s. Logically, his views should have been central to the debate between neo-Hegelian and neo-positivist interpretations of Marxist philosophy, but the neo-positivists' affinity with an outright critic of Lenin had become a political smear, to be hurled or dodged without serious discussion of the critic's arguments. Thus his warnings against philosophy desiccated by idolatry fell against a wall of silence, erected by an ideological bureaucracy seeking formulas for a confessional Marxism.

The most poignant irony of this situation was the pragmatism that all these Marxists shared but could not openly acknowledge or freely debate. They all agreed that practice is the criterion of truth, that the way forward is through collective action based on 'the lessons of practice'. As the ideological establishment read the lessons, an obvious inference emerged: since Lenin was the supremely practical leader of the successful revolution, he must also be considered the supreme philosopher. The next step, the extension of that pragmatic logic to Stalin worship, would be taken in 1929-30, but Bogdanov did not live to see it. In 1928 he submitted his body to an exchange of all its blood for that of a man suffering from malaria and consumption, and died. No direct evidence of suicidal intent has come to light, though one notes that suicide was a debatable issue, not a taboo, in Bogdanov's philosophizing. The official eulogies described his experiment as heroic self-sacrifice for the advancement of science, but that did not rescue his philosophy from the limbo of enemy doctrines, recalled only to be reviled.

3 Reception and significance

Khrushchev's reforms of the 1950s and early 1960s revived a vision of socialism as a freely self-regulating system, but the most practical visionaries drew on mathematical economics rather than Bogdanov's legacy. Memory of Bogdanov's tektology, as a precursor of cybernetics and systems theory, cropped up in talk of improving the Soviet system by a 'scientific technological revolution' - fuzzy talk that served to evade serious engagement with the mounting problems of a command economy and an ideology that stifled thought. Bogdanov's

mockery of 'fetishism' and 'vampirism', and his insistence on free collectivism, were not recalled until the late 1980s, during Gorbachev's push for 'openness' in thought and 'restructuring' in social organization. But that push turned so quickly into systemic collapse, with total revulsion against Marxist thought in any form, that Bogdanov's period of respect in his native land was very brief. Whether Russian thinkers will ever have a sustained revival of interest in his thought depends still, as it has all along, on the stand they take between extremes of total revulsion and total passivity, between wild attempts to leap out of the system they are born into and fatalistic adaptation to it. If they attempt careful thought about ways to make effective use of their historic experience, the part that Bogdanov played will take on a new significance.

It seems unlikely that Bogdanov's thought ever will be considered of major significance apart from the life that gave it meaning within the Russian revolution. Few of his works have been translated into Western languages, though some historians and some believers in systems theory have called attention to him. Tektology may be 'a kind of metascience' (Gorelik 1980), comparable to Wiener's cybernetics or Bertalanffy's general system theory, but it has attracted less interest. The most searching studies of Bogdanov's thought have set it in an historical context that it does not transcend. Such studies emphasize either the logical inconsistencies of his efforts to combine epistemological scepticism with a monistic faith (see Kolakowski 1978), or the derivative quality of his popularized knowledge and overabundant theorizing (see Grille 1966), or the utopian faith that he struggled to keep creatively alive while the victorious revolutionaries were turning utopia into ritualized worship of a dream deferred (see Sochor 1988). Sympathy abounds, but it does not save the thought. Sochor dwells on Bogdanov's creative utopianism as evidence that the Russian Revolution might have developed less tyrannically than it actually did, and Grille pictures Bogdanov as a 'red Hamlet', whose balancing between the demands of action and of thought reflects honour on him. Both authors, and Kolakowski even more, avoid the hardest philosophical question posed by the history of political thought and action in Third World countries. They fail to ask whether practice as the criterion of truth in Russia - or China, or Angola, or Iran - is compatible with pragmatism as a philosophy of comfortable thinkers in advanced countries.

See also: Russian empiriocriticism; Marxist philosophy, Russian and Soviet

DAVID JORAVSKY

List of works

Bogdanov, A.A. (1897) *Kratkii kurs ekonomicheskoi nauki*; trans. *A Short Course of Economic Science*, London: Communist Party of Great Britain.(Bogdanov's first publication, which became a widely used textbook in Marxist circles.)

Bogdanov, A.A. (1899) *Osnovnye elementy istoricheskogo vzgliada na prirodu (Basic Elements of the Historical View of Nature)*, St Petersburg: Izdatel'.(An energeticist foundation for Marxism.)

Bogdanov, A.A. (1904) *Iz psikhologii obshchestva; stat'i 1901-04 (From the Psychology of Society; Articles 1901-4)*, St Petersburg: Delo; 2nd edn, enlarged, 1906.(Much on ideology, little on psychology, in the sense of an academic discipline.)

Bogdanov, A.A. (1905-7) *Ėmpiriomonizm; stat'i po filosofii (Empiriomonism: Articles on Philosophy)*, 3 vols, Moscow: Dorovatovskii & Charushkinov; 2nd and 3rd edn, 1906-8.(An assemblage of articles, rather than the unified treatise that this work is often assumed to be.)

Bogdanov, A.A. (1908a) *Prikliucheniia odnoi filosofskoi shkoly (The Adventures of a Philosophical School)*, St Petersburg: Znanie.(Response to criticisms by orthodox Mensheviks.)

Bogdanov, A.A. (1908b) 'Ernst Mach und die Revolution' (Ernst Mach and the Revolution), in Neue Zeit XXVI (1): 695-700.(Abridged translation of his preface to Russian translation of Mach's *Analyse der Empfindungen*.)

Bogdanov, A.A. (1908/1911) *Krasnaia zvezda/Inzhener Menni*, trans. C. Rougle, ed. L. Graham and R. Stites, *Red Star and Engineer Menni*, Bloomington, IN: Indiana University Press.(Translation of Bogdanov's two works of science fiction, with commentary.)

Bogdanov, A.A. (1910) *Padenie velikogo fetishizma (sovremennyi krizis ideologii); Vera i nauka (The Fall of a Great Fetishism; Faith and Science)*, Moscow: Dorovatovskii & Charushkinov.(The response to Lenin's *Materializm i ėmpiriokrititsizm*.)

Bogdanov, A.A. (1913) *Filosofiia zhivogo opyta: populiarnye ocherki; materializm, empiriokrititsizm, dialekticheskii materializm, ėmpiriomonizm, i nauka budushchego (A Philosophy of Living Experience: Popular Essays; Materialism, Empiriocriticism, Dialectical Materialism, Empiriomonism and the Science of the Future)*, St Petersburg: Semenov; 2nd and 3rd edn, enlarged 1920, 1923.(See analysis by Jensen.)

Bogdanov, A.A. (1913-25) *Tektologiia: Vseobshchaia organizatsionnaia nauka (Tektology: The Universal Organizational Science)*, St Petersburg, Moscow, Berlin: various publishers; trans. and ed. G. Gorelik, *Essays in Tektology*, Seaside, CA: Intersystems, 1980.(Three editions of his major work on systems theory; for bibliographic specifics see Susiluoto. Note also a German translation of 1926-8; and a two-volume Russian edition, Moscow, 1990.)

Bogdanov, A.A. (1924) *O proletarskoi kul'ture, 1904-24 (On Proletarian Culture, 1904-24)*, Moscow: Kniga. (Collection of articles with prefatory explanation.)

Bogdanov, A.A. (1926) 'Avtobiografiia' (Autobiography), in *Ėntsiklopedicheskii slovar' Granat* XLI (1): 29-33. (Includes Bogdanov's own list of his most important publications. German translation in Grille.)

Bogdanov, A.A. (1990) *Voprosy sotsializma: raboty raznykh let (Problems of Socialism: Works of Various Years)*, Moscow: Politizdat.(Includes criticisms of the Bolshevik Revolution originally published in 1917-18, the letter to Lunacharskii and so on.)

References and further reading

Bailes, K.E. (1967) 'Lenin and Bogdanov: The End of an Alliance', in *Columbia Essays in International Affairs* II: 107-33, New York: Columbia University Press.(Contains valuable archival material, including Trotsky's and Bogdanov's letters to the school on Capri.)

Grille, D. (1966) *Lenins Rivale: Bogdanov und sein Philosophie*, Köln: Wissenschaft und Politik.(The most detailed intellectual biography, emphasizing the derivative nature of his thought and his character as 'the red Hamlet'.)

Jensen, K. (1982) *Beyond Marx and Mach: Aleksandr Bogdanov's Philosophy of Living Experience*, Dordrecht: Reidel.(Analysis focused on Bogdanov's *Filosofiia zhivogo opyta*.)

Joravsky, D. (1961) *Soviet Marxism and Natural Science, 1917-1932*, New York: Columbia University Press. (Analyses the conflict between Bogdanov and Lenin before 1917, and sets Bogdanov in context of the 1920s.)

Kolakowski, L. (1978) *Main Currents of Marxism II: The Golden Age*, Oxford: Clarendon Press.(Bogdanov is analysed on 432-45.)

Rossiiskaia Akademiia Nauk, Institut Ekonomiki, Kommissiia po nauchnomu naslediiu A.A. Bogdanova (1992) *Trudy*.(Articles about Bogdanov and some items by him.)

Sochor, Z.A. (1988) *Revolution and Culture: The Bogdanov-Lenin Controversy*, Ithaca, NY: Cornell University Press.(A detailed analysis with emphasis on the role of utopian faith.)

Susiluoto, I. (1982) *The Origins and Development of Systems Thinking in the Soviet Union: Political and Philosophical Controversies from Bogdanov and Bukharin to Present-Day Reevaluations*, Helsinki.(A careful analysis of tektology and its reception.)

Vucinich, A. (1976) *Social Thought in Tsarist Russia: The Quest for a General Science of Society, 1861-1917*, Chicago: University of Chicago Press.(The last chapter credits Bogdanov with significant innovation.)

Bohr, Niels (1885-1962)

One of the most influential scientists of the twentieth century, the Danish physicist Niels Bohr founded atomic quantum theory and the Copenhagen interpretation of quantum physics. This radical interpretation renounced the possibility of a unified, observer-independent, deterministic description in the microdomain. Bohr's principle of complementarity - the heart of the Copenhagen philosophy - implies that quantum phenomena can only be described by pairs of partial, mutually exclusive, or 'complementary' perspectives. Though simultaneously inapplicable, both perspectives are necessary for the exhaustive description of phenomena. Bohr aspired to generalize complementarity into all fields of knowledge, maintaining that new epistemological insights are obtained by adjoining contrary, seemingly incompatible, viewpoints.

Bohr's elaborations of complementarity can be divided into two distinct periods marked by the challenge of the Einstein-Podolsky-Rosen (EPR) paper in 1935 (see Einstein, A. §5). The first period explores the compatibility of the quantum formalism with experiments through Heisenberg's uncertainty principle, which implies that in the quantum domain spacetime and energy-momentum (or causal) descriptions are not simultaneously applicable (see Heisenberg, W.). In Bohr's terminology, spacetime and causality are complementary. Similar complementarity relations hold between wave and particle attributes of atomic objects, which are exhibited in mutually exclusive experimental arrangements.

Bohr's initial formulations of complementarity relied on the idea of an 'uncontrollable' physical disturbance of the microscopic object by the measuring device, due to the 'indivisibility' of the quantum of action and necessitating the 'inseparability' of atomic phenomena from their means of observation. This disturbance is the source of the 'final renunciation' of causality and objective reality in the atomic domain. According to Bohr, realistic causal descriptions presuppose the possibility of the definition of a system's state, excluding in principle all disturbances.

The EPR paper undercut Bohr's notion of disturbance. Einstein and his co-workers devised a thought experiment where for two previously interacting particles one is free to predict either the position or the momentum of the first particle on the basis of observations made on the spatially separated second particle. Since the measurement on the second particle cannot immediately 'disturb' the state of the first one, they concluded that both position and momentum can be attributed to the first particle simultaneously, contrary to the uncertainty, complementarity and 'completeness' of the quantum theory.

Bohr countered this challenge by resorting to operational and relational definitions of concepts, replacing his idea of physical disturbance by a 'semantic' one. In the EPR experiment, measuring the position of the first particle precludes the prediction, or meaningful assignment, of momentum to the second one. Scientific concepts are to be defined only relative to the 'whole' experimental arrangement. After EPR, such positivistic statements are frequent in Bohr's writings.

The central claim of Bohr's philosophy after EPR is the indispensability of classical concepts, supplemented by ordinary language, for unambiguous communication of experimental results. Unambiguous description presupposes a strict separation between the observed object and the observing subject. This is possible only at the classical level, where the quantum of action can be neglected. Classical concepts, guided by complementarity in different experimental arrangements, are therefore the fundamental descriptive concepts in the quantum domain. There are no 'quantum concepts' and no 'quantum reality', only an abstract formalism for the calculation and prediction of measurement results. Progress in science consists in 'rational generalizations' of the old frameworks, not in their replacement by genuinely new conceptual schemes. This claim, placing a priori limits on scientific theorizing, was contested by Bohr's opponents and sympathizers alike.

Bohr's doctrine of the indispensability of classical concepts is rooted in his correspondence principle, which, by relying on classical analogies, guided the search for a consistent quantum theory before its creation in 1925. Yet the difference between correspondence and complementarity is striking. The former was a heuristic principle leading to many new discoveries, culminating in Heisenberg's quantum formalism. The latter had no new empirical import, being aimed at the philosophical legitimization of the quantum theory.

Bohr supports his view that the classical realm is privileged by arguments stressing direct accessibility of classical reality to immediate sense perception. These arguments, together with Bohr's preoccupation with the conditions and limits of the applicability of concepts, have strong Kantian roots. In this respect there is a substantial similarity between Bohr and the philosophy of his teacher, the Neo-Kantian philosopher Harald Høffding.

In his later years, Bohr aspired to extend the 'epistemological lesson' of complementarity to biology, psychology and anthropology, maintaining that in these fields, as in quantum physics, observational interaction can be neither neglected nor precisely determined. These generalizations are rarely developed, by Bohr or others, into more than suggestive analogies.

The value of Bohr's philosophy for the advancement of physics is controversial. His followers consider complementarity a profound insight into the nature of the quantum realm. Others consider complementarity an illuminating but superfluous addendum to quantum theory. More severe is the opinion that Bohr's philosophy is an obscure 'web of words' and mute on crucial foundational issues.

Opinions are also divided about the status of Bohr's philosophy after Bell's results (see Bell's theorem). Some scholars consider Bell's proof that rules out local deterministic realistic theories a vindication of Bohr's philosophy of indeterminism and inseparability. Others hold that Bohr's positivistic prohibitions are irrelevant to the contemporary quest for an adequate quantum ontology. Once a ruling 'dogma', Bohr's philosophy is no longer considered the final word about the nature of the quantum world.

See also: Logical positivism §4; Operationalism; Quantum mechanics, interpretation of

MARA BELLER

List of works

(The first three collections contain Bohr's main body of philosophical essays on quantum theory.)

Bohr, N. (1934) *Atomic Theory and the Description of Nature*, Cambridge: Cambridge University Press.

Bohr, N. (1963a) *Atomic Physics and Human Knowledge*, New York: Wiley.

Bohr, N. (1963b) *Essays 1958-1962 on Atomic Physics and Human Knowledge*, New York: Wiley.

Bohr, N. (1972-86) *Collected Works, Volumes 1-9*, Amsterdam: North Holland.(Contains Bohr's major scientific and philosophical papers on quantum theory until 1932, and those on nuclear physics until 1952.)

References and further reading

Beller, M. and Fine, A. (1994) 'Bohr's Response to EPR', inJ. Faye and H. Folse (eds) *Niels Bohr and Contemporary Philosophy*, Amsterdam: North Holland, 1-31. (A close reading of Bohr's response to the Einstein-Podolsky-Rosen challenge with an emphasis on a positivistic shift in Bohr's thought.)

Chevalley, C. (1991) 'Le dessin et la couleur' ('The drawing and the colour'), in *Introduction to Niels Bohr, Physique Atomique et Connaissance Humaine (Atomic Physics and Human Knowledge)*, Paris: Gallimard, 19-140, 307-641.(A presentation of Bohr's philosophy in the Neo-Kantian tradition.)

Darrigol, O. (1992) *From c-Numbers to q-Numbers. The Classical Analogy in the History of Quantum Theory*, Berkeley, CA: University of California Press.(Analysis of the origins and the applications of Bohr's correspondence principle.)

Faye, J. (1991) *Niels Bohr: His Heritage and Legacy. An Antirealist View of Quantum Mechanics*, Dordrecht: Kluwer.(Analysis of the impact of Høffding's philosophy on Bohr's interpretation of quantum physics.)

Faye, J. and Folse, H.J. (eds) (1994) *Niels Bohr and Contemporary Philosophy*, Amsterdam: North Holland.(A collection of essays on Bohr's philosophy, with an emphasis on realism and positivism.)

Folse, H.J. (1985) *The Philosophy of Niels Bohr: The Framework of Complementarity*, Amsterdam: North Holland.(Analysis of Bohr's extension of complementarity into domains beyond physics.)

Honner, J. (1987) *The Description of Nature: Niels Bohr and the Philosophy of Quantum Mechanics*, New York: Oxford University Press.(Argues that Bohr's philosophy is best understood as a transcendental inquiry into the very preconditions of the possibility of experience.)

Hooker, C.A. (1972) 'The Nature of Quantum Mechanical Reality: Einstein vs. Bohr', in R.G. Colodny (ed.) *Paradigms and Paradoxes*, Pittsburgh, PA: University of Pittsburgh Press.(Treats conceptual intricacies of the Bohr-Einstein debate.)

Murdoch, D. (1987) *Niels Bohr's Philosophy of Physics*, Cambridge: Cambridge University Press.(Analysis of different meanings of Bohr's complementarity, with an emphasis on the measurement problem.)

Pais, A. (1991) *Niels Bohr's Times: In Physics, Philosophy and Polity*, New York: Oxford University Press.
(Description of Bohr's life and thought by a physicist who adheres to Bohr's complementarity.)

Bold, Samuel (1649-1737)

Samuel Bold (or Bolde) was a Latitudinarian minister who defended John Locke's Reasonableness of Christianity and his Essay Concerning Human Understanding. Bold published a series of pamphlets and short books which argued a theological position substantially identical to that of Locke. He also mounted a philosophical defence of Locke's definition of knowledge and his supposition that it was possible that God could, if he so wished, superadd to matter the power of thought. In a book on the theological issue of the resurrection of the same body he defended Locke's account of personal identity.

Samuel Bold entered English public controversy in his Dorset ministries with a published sermon in defence of moderation for dissenters, followed by other controversial publications, including pamphlets in the late 1690s supporting Locke's *Reasonableness of Christianity* (Bold later surmised that Locke was the author of this anonymous work) (see Locke, J. §§1, 7). He was also much impressed with Locke's *Essay Concerning Human Understanding* and, when this was attacked by John Edwards, he defended it in a further pamphlet, *Some Considerations on the Principal Objections and Arguments Which have been Publish'd against Mr Lock's Essay of Humane Understanding* (1699).

Bold's defence of Locke was centred on two issues: Locke's definition of knowledge and his conjecture that God could add to matter a power of thinking, issues which were also central in Locke's controversy with Stillingfleet. On the first of these, Locke had defined knowledge in terms of seeing an agreement or disagreement between ideas. As Bold explained it, propositions are known to be true 'by perceiving that the Ideas, signified by the words of which the proposition doth consist, have such a connection or agreement, repugnancy, or disagreement, as the Proposition doth express' (1699: 6). He claimed that Locke was correct in holding that there was no way to truth except via ideas. He underlined that this was no threat to the Christian religion as long as we clearly distinguished knowledge (which is certain) from faith on evidence (which, although reasonable, is not).

With regard to the possibility of thinking matter Bold fully appreciated what Locke's detractors generally did not, that Locke's claim was an epistemic one: *for all we know* God might add the power of thinking to matter (because there was no inconsistency between the ideas of matter and thinking). It was not a disguised attempt to argue for the truth of the claim. As such, Bold held, it could not be faulted. The power of thinking could in principle be added by God to any substance, mental or physical, for we cannot understand in either case how that power actually produces its effect. All we know is that it does.

Bold shows himself to be an able, if unoriginal, philosopher. His next work, the book *A Discourse Concerning the Resurrection of the Same Body* (1705), is again a defence of Locke's philosophy, this time focusing on his account of personal identity and its implications for theology, especially as it relates to the necessary articles of the Christian faith. Although Bold's argument does not advance beyond Locke's position, it brings out clearly the close relationship between the epistemic issues which Locke raises and contemporary Christian theology. This is further confirmed by the subsequent section of the work, separately titled *A Discourse concerning the Immateriality of the Soul*, in which once again it is Locke's epistemology that is defended.

Bold is rarely more than a clear-headed commentator on Locke, but he is undoubtedly a superior philosopher to many of those who sought to attack the *Essay Concerning Human Understanding* for its supposedly unacceptable theological implications.

See also: Latitudinarianism; Socinianism

G.A.J. ROGERS

List of works

Bold, S. (1706) *A Collection of Tracts, Publish'd in Vindication of Mr Lock's Reasonableness of Chrsitianity, as deliver'd in the Scriptures; and of his Essay Concerning Humane Understanding*, London.(This contains many of Bold's writings, including those listed here.)

Bold, S. (1699) *Some Considerations on the Principal Objections and Arguments Which have been Publish'd against Mr. Lock's Essay of Humane Understanding*, London.(A strong defence of Locke's epistemology, including his definition of knowledge, especially against those critics who read it as a threat to established

religion.)
Bold, S. (1705) *A Discourse Concerning the Resurrection of the Same Body: with two Letters Concerning the Necessary Immateriality of Created Thinking Substance*, London.(Defends Lockean positions on identity of personhood and on the logical possibility of God superadding to matter the power of thought without claiming that God has actually done that in the case of human beings.)

References and further reading

Bayne, R. (1886) 'Samuel Bold' in *Dictionary of National Biography*, vol. 5, London, 317-8.(A brief account of Bold's life and works.)
Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, London and Oxford: Oxford University Press, 1975.(The work that Bold saw as expressing the philosophy to which he was committed.)
Locke, J. (1695) *The Reasonableness of Christianity, as Delivered in the Scriptures*, ed. J. Higgins-Biddle, Oxford: Oxford University Press, 1998. (Anonymous, but identified by Bold as by Locke's. It expressed Bold's own understanding of Christianity.)
Edwards, J. (1695) *Some Thoughts Concerning the Several Causes and Occasions of Atheism, Especially in the Present Age. With some Brief Reflections on Socinianism: And on a Late Book Entitled The Reasonableness of Christianity as deliver'd in the Scriptures*, London; repr. New York: Garland, 1984.(The work to which Bold's writings were, in a large part, a response.)

Bolzano, Bernard (1781-1848)

Bernard Bolzano was a lone forerunner both of analytical philosophy and phenomenology. Born in Prague in the year when Kant's first Critique appeared, he became one of the most acute critics both of Kant and of German Idealism. He died in Prague in the same year in whichFrege was born; Frege is philosophically closer to him than any other thinker of the nineteenth or twentieth century. Bolzano was the only outstanding proponent of utilitarianism among German-speaking philosophers, and was a creative mathematician whose name is duly remembered in the annals of this discipline. His Wissenschaftslehre (Theory of Science) of 1837 makes him the greatest logician in the period betweenLeibniz and Frege. The book was sadly neglected by Bolzano's contemporaries, but rediscovered by Brentano's pupils: Its ontology of propositions and ideas provided Husserl with much of his ammunition in his fight against psychologism and in support of phenomenology, and throughTwardowski it also had an impact on the development of logical semantics in the Lwów-Warsaw School.

1 Life and main works

Bolzano was son of a German mother and an Italian father. As a young man he became keenly interested in the foundations of mathematics (his first mathematical paper appeared in 1804), yet he decided to study theology. In his native city, Prague, he graduated with a Ph.D. and was ordained a priest in 1805. In the same year he was appointed professor of the philosophy of religion to a newly-founded chair at the Charles-Ferdinand University. Nevertheless his first book was *Beyträge zu einer begründeteren Darstellung der Mathematik (Contributions to a More Well-founded Presentation of Mathematics)* (1810) in which he attackedKant's views on geometry and arithmetic. Bolzano's university lectures are preserved in his monumental *Lehrbuch der Religionswissenschaft (Textbook of the Science of Religion)* (1834). His philosophy of religion is based on his ethics. He criticizes Kant's categorical imperative and his doctrine of postulates, and advocates a version of utilitarianism. Bolzano's logical notion of a manifold (*Inbegriff*) is the conceptual core of his version of a cosmological argument for the existence of God (see God, arguments for the existence of §1). Only the conclusion of this argument is the possible content of a religious belief. Bolzano maintains that a religious belief is, or strengthens the motivational power of, a moral belief. He takes beliefs to be indirectly under the control of our will and presents a lucid account of self-deception. Someone's religious belief may be due to self-deception, he argues, and yet the steps taken to acquire it may be justifiable by utilitarian principles. Taking divine revelation to be an attempt at communicating, he gives a very subtle analysis of this concept. He rejects the Humean definition of a miracle and tries to defend the possibility of miracles with the help of his theory of probability.

Owing to his public lectures, many of which are collected in his *Erbauungsreden (Sermons)* (1813, 1849-52), Bolzano became the protagonist of the 'Bohemian Enlightenment'. He pleaded for social reforms, attacked militarism and strongly criticized both the discrimination against Czech Bohemians by their German compatriots and the anti-Semitism on both sides. The Austrian Emperor Francis was infuriated by much of this and dismissedBolzano from his chair in 1820.

Bolzano withdrew to a small village in southern Bohemia where he composed some of his most important works. His metaphysical treatise *Athanasia (Immortality)* (1827) is inspired byLeibniz's *Monadology*. The four volumes of his *Theory of Science* contain a highly original (philosophy of) logic (vols 1-2), an empiricist epistemology and a theory of scientific discovery (vol. 3), and a methodology for writing useful textbooks (vol. 4). His *Größenlehre (Theory of Magnitudes)* comprising *Zahlenlehre (Theory of Numbers)*, *Functionenlehre (Theory of Functions)* and *Raumwissenschaft (Theory of Space)* remained unfinished (and will eventually be published in the complete edition of his works (1969-)). Bolzano spent the last seven years of his life back in Prague, where he frequently delivered papers (mainly on aesthetics and mathematics) to the Bohemian Academy of the Sciences. Among his most outstanding mathematical contributions are the following: he defined convergence and indicated convergence criteria (several years before Cauchy), and described a function continuous but not differentiable in an interval (several decades before Weierstrass). He realized that any infinite set contains a subset that stands in bi-univocal correspondence to it, and he realized that this is not a contradiction. This is a central topic in his *Paradoxien des Unendlichen (Paradoxes of the Infinite)* (1851). Cantor thought very highly of this book, and Peirce praised its author as conferring 'a singular benefit upon humanity' (1909).

2 Basic ontological categories

Bolzano's universe is a universe of objects (*Gegenstände*). Objects are either actual (*wirklich*, effectual) or non-actual. Something is actual if and only if it is an element of the causal order. (Bolzano outlines his ontology of actual objects at the beginning of his *Athanasia*.) Actual objects are either substances or adherences. An object is an 'adherence' if and only if it is actual and a feature of another actual object on whose existence and identity it depends. (Your last headache would be an example.) An object is a 'substance' if and only if it is actual and not an adherence. (Sometimes Bolzano takes substances to be non-composite as well.) Some adherences are mental: volitions, wishes, sensations and, most importantly for the *Theory of Science*, subjective ideas (*subjektive Vorstellungen*) and acts of judgment (*Urteile*). Properties as shareable by many objects are non-actual objects, and so are numbers. Seen against the background of earlier metaphysics *Sätze an sich* ('sentences-as-such': henceforth, 'propositions') and their constituents are the most remarkable non-actual objects in Bolzano's ontology as developed in his *Theory of Science*. In his use of '*Satz*' Bolzano can appeal to one prominent use of this term in German. A definite description such as '*der Satz des Pythagoras*' (Pythagoras' theorem) refers to something purportedly discovered by a Greek philosopher-mathematician which can be expressed by many sentences in many languages. In Bolzano's usage, the epithet '*an sich*' amounts to something like the gloss 'in the strict sense of this term'.

Propositions are not true or false in a language or in a context, but simply true or false, full-stop. They differ from true or false utterances or judgments in that they are non-actual. Thus something is a proposition if and only if it is both *simpliciter* true or false and non-actual. An utterance is true if and only if the proposition expressed by it is true. The proposition expressed by an utterance is the sense (*Sinn*) of this utterance. Bolzanian sense is not to be identified with the linguistic meaning of the sentence uttered. (Sentences with the same linguistic meaning might be used to express different propositions: your utterance of 'I have blood-type A' might express a true proposition, while my utterance of '*Ich habe Blutgruppe A*' expresses a false one.) Propositions also play a psychological role in being the potential contents (*Stoff* (matter)) of judgments. A judgment is true if and only if the proposition which is its content is true. Propositions are composite, structured entities. *P* is the same proposition as *Q* if and only if *P* is built in the same way from the same components as *Q*.

A component of a proposition which is not itself a proposition is a *Vorstellung an sich* ('idea-as-such': henceforth, 'idea'). The idea expressed by a non-sentential component of a sentential utterance is the sense of this component. (Not all words in an utterance express ideas: the phrase 'really sad' expresses the same idea as 'sad'. Not all ideas which are components of the proposition expressed by an utterance are expressed by components of the utterance: if you now say, 'It is snowing', the proposition expressed must contain an idea which represents a place, otherwise it would not be *simpliciter* true or false.) One and the same idea can be the content of many subjective ideas, that is, of many non-judgmental components of judgments. An idea is objectual (*gegenständlich*) if and only if there is at least one object to which it refers. An idea which refers to exactly one object and is not composed of other ideas is an 'intuition' (*Anschauung an sich*). An intuition is expressed by an indexical use of 'this'. An idea which neither is nor contains an intuition is a pure 'concept' (*reiner Begriff*). Equivalent ideas (*Wechselvorstellungen*) are objectual ideas which refer to the same objects. Bolzanian ideas are very fine-grained: 'trilateral' and 'triangular', for example, in spite of being analytically coextensive, do not express the same idea.

3 Truth and degrees of validity

A proposition *P* attributing property *y* to object *x* is true, Bolzano explains, if and only if *x* has *y*. Here '*P* attributes *y* to *x*' is meant to characterize not a linguistic or mental activity, but the internal structure of a proposition. The copula 'has' could be paraphrased by 'exemplifies' or 'instantiates'. Bolzano's conception of truth is thoroughly non-epistemic. Some truths may never be known (by any finite mind): 'The number of blossoms that were on a certain tree last spring is a statable, if unknown figure. Thus, the proposition which states this figure I call an objective truth, even if nobody knows it.' Nor is there any allusion in the definiens to what could be known under ideal conditions.

Bolzano's explication of truth presupposes that every proposition has the structure '*x* has *y*'. But we can say something true or false by sentences such as 'It is snowing' or 'Nobody is perfect', although *prima facie*, at least, they do not seem to ascribe a property to an object. In his philosophical grammar Bolzano offers reformulations with the appropriate structure: 'The idea of a snowfall occurring here and now has objectuality' and 'The idea of a

perfect person has lack of objectuality'. (Bolzano anticipates Frege's and Russell's conception of existence as a second-order notion.) But of course, the search for such a canonical paraphrase threatens to become rather tedious when we have to deal with sentences of the form 'If *P* then either *Q* or *R*', for example. Bolzano himself was by no means sure that his programme of paraphrase could always be carried out.

The basic procedure in Bolzano's logic is that of considering 'variants' of a given proposition with respect to one or more ideas contained in it. Let us take the idea expressed by the numeral '6' to be the variandum in the proposition P_1 that Beethoven's *Symphony No. 6* is in F major. (I shall refer to this idea as [6] and to variants of P_1 with respect to it as [6]-variants of P_1.) Let us only consider objectual [6]-variants of P_1, that is, only those variants the subject-idea of which is objectual. Some of these variants are true, some are false. If we want to determine the ratio between true variants and all variants we must forbid the replacement of [6] by any equivalent idea. Otherwise the floodgates would be opened for infinitely many ideas: [7-1], [8-2], [9-3], and so on. Using '$P[i/j]$' to denote the proposition which differs from *P* by containing the idea *j* at all places where *P* contains the idea *i*, we can introduce the following abbreviation: *Q* is a 'relevant' variant of *P* with respect to the idea *i* contained in *P* if and only if ($Q = P$) or ($Q = P[i/j]$, the ideas *i* and *j* are not equivalent, and *Q* is objectual).Bolzano defines the 'degree of validity' (*Grad der Gültigkeit*) of a proposition with respect to its component *i* as the ratio between the number of all its relevant *true i*-variants and the number of *all* its relevant *i*-variants. This degree of validity can be represented by a fraction the numerator of which is the first of those numbers and the denominator of which is the second. Thus the degree of validity of P_1 with respect to [6] is $\frac{2}{9}$. (Bolzano makes important use of this machinery in his theory of probability.) With respect to [6], the degree of validity of the proposition P_2 that Beethoven's *Symphony No. 6* contains more than one note is one, since all relevant variants of P_2 are true. Such propositions are universally valid (*allgemeingültig*) with respect to a component idea. With respect to [6] the degree of validity of the proposition P_3 that Beethoven's *Symphony No. 6* consists of a single note is zero, since all relevant variants of P_3 are false. Such propositions are universally invalid (*allgemein ungültig*) with respect to a component idea.

4 Analyticity, deducibility and consequentiality

A proposition is 'analytic' (in the broader sense), Bolzano explains, if and only if it contains at least one idea *i* such that it is either universally valid or universally invalid with respect to *i*. All other propositions are synthetic. Hence propositions P_2 and P_3 above are both analytic, whereas P_1 is synthetic. A proposition is 'logically analytic' if and only if it is either universally valid or universally invalid with respect to all the nonlogical ideas contained in it. Hence 'Whatever is square is square' expresses a logically analytic truth, since presumably the concept of a square is the only nonlogical idea in this proposition. Bolzano, like Tarski (1936), takes the borderline between logical and nonlogical ideas not to be sharp. Using a convenient terminology introduced by Quine we can say: a logically analytic truth contains only logical ideas essentially; any other idea occurring in it can be varied at will without engendering falsity. (Quine acknowledges the affinity of his view of logical truth to Bolzano's conception.) A truth which is analytic in the broader sense contains at least one idea nonessentially. (There is no counterpart to broad analyticity in Quine.)Bolzano explains analyticity as a property of propositions, not of sentences. This explanation is, like his account of truth, strictly non-epistemic, and it is vastly different from the Kantian one (see Kant, I. §4): for Bolzano, to illustrate just one of the many differences, your utterance of 'This cup which was taken from my cupboard is broken', expresses an analytic truth with respect to the intuition expressed by the demonstrative in your utterance if all the cups in your cupboard happen to be broken.

P and *Q* are 'compatible' (*verträglich*) with respect to idea *i* if and only if there is an idea *x* such that both $P[i/x]$ and $Q[i/x]$ are true. *Q* is 'deducible' (in the broader sense) from *P* with respect to idea *i* if and only if *P* and *Q* are compatible with respect to *i* and for all ideas *x*, if $P[i/x]$ is true then $Q[i/x]$ is true. Hence nothing is deducible (*ableitbar*) from a contradiction. As deducibility with respect to an idea is a relation between propositions, it is neither a syntactic nor a semantic relation in the modern sense. The proposition P_4 that in Prague it was warmer in August 1830 than it was in July 1830 is deducible from the proposition P_5 that in Prague the thermometer registered higher in August 1830 than it did in July 1830, and vice versa, with regard to the chronological and geographical ideas they contain, since every variation on any of these components that transforms P_4 into a true proposition also makes P_5 true, and vice versa. *Q* is 'logically deducible' from *P* if and only if *Q* is deducible from *P* with respect to all nonlogical ideas contained in them. This conception, or rather its generalization as found

in Bolzano, has often been described as a forerunner of Tarski's notion of logical consequence, and that is whatTarski himself took it to be.

Bolzano contrasts deducibility with 'consequentiality' (*Abfolge*) which only obtains between true propositions and which is never reciprocal. When a truth is deducible from and explained by another truth it is its consequence. Suppose propositions P_4 and P_5 above are true. Obviously P_4 is not a consequence of P_5: It is not the case that there was a rise in temperature *because* the thermometer registered higher. Rather, P_5 is a consequence of P_4. Bolzano also maintains that for all truths *P*, the proposition that it is true that *P* is a consequence of the proposition that *P*, while admitting that such propositions are always reciprocally deducible. Since nothing is its own consequence, Bolzano (as opposed to Frege and F.P. Ramsey (§4)) takes 'It is true that *P*' and '*P*' to express *different* propositions.

See also: Analytical philosophy

WOLFGANG KÜNNE

List of works

Bolzano, B. (1969-) *Gesamtausgabe (Collected Works)*,ed. J. Berg *et al.*, Stuttgart: Frommann/Holzboog.(The standard collected works. Volumes E 2/1 (1972), E 2/1, Supp. 1 (1982) and E 2/1, Supp. 2 (1988) contain the first instalments of an invaluable bibliography.)

Bolzano, B. (1810) *Beyträge zu einer begründeteren Darstellung der Mathematik (Contributions to a More Well-founded Presentation of Mathematics)*, Prague; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1974.(A sketch of an anti-Kantian philosophy of mathematics, insisting on logical rigour.)

Bolzano, B. (1813, 1849-52) *Erbauungsreden (Sermons)*, Prague, Vienna and Leipzig, 5 vols.(Public lectures on religious, moral and political topics.)

Bolzano, B. (1827) *Athanasia oder Gründe für die Unsterblichkeit der Seele (Grounds for the Immortality of the Soul)*, Sulzbach: Seidel, 2nd edn, 1838; repr. Frankfurt am Main: Minerva, 1970.(Metaphysics of simple substances and accidents.)

Bolzano, B. (1834) *Lehrbuch der Religionswissenschaft (Textbook of the Science of Religion)*, Sulzbach: Seidel, 4 vols.(Rationalist philosophy of religion and utilitarian ethics.)

Bolzano, B. (1837) *Wissenschaftslehre*, Sulzbach: Seidel, 4 vols; selections, trans. and ed. R. George, in *Theory of Science*, Oxford: Blackwell, 1972; selections, trans. B. Terrell, ed. J. Berg, in *Theory of Science*, Dordrecht: Reidel, 1973.(A highly original philosophy of logic followed by treatises on empiricist epistemology, heuristics and textbook methodology.)

Bolzano, B. (1851) *Paradoxien des Unendlichen*, Leipzig: Reclam; repr. Hamburg, Meiner, 1975; trans. and with intro. by D.A. Steele, *Paradoxes of the Infinite*, London and New Haven, CT: Routledge, 1950.(A theory of infinite manifolds and metaphysics of simple substances and forces.)

References and further reading

'Bolzano's *Wissenschaftslehre* 1837-1987, International Workshop', Biblioteca di Storia della Scienza 31, Florence, 1992.(Useful collection of articles on vols 1-2 of Bolzano (1837), all in English.)

Berg, J. (1962) *Bolzano's Logic*, Stockholm: Almqvist & Wiksell.(Pioneering study of all the topics covered in §§3-4.)

Berg, J. (1992) *Ontology Without Ultrafilters And Possible Worlds, An Examination of Bolzano's Ontology*, Salzburg: Academia Verlag.(Resolutely technical reconstruction, also covering Bolzano's mathematical work.)

Buhl, G. (1961) *Ableitbarkeit und Abfolge in der Wissenschaftstheorie Bolzanos (Deducibility and Consequentiality in Bolzano's Theory of Science)*, Köln: Universitätsverlag.(Lucid discussion of topics covered in §4.)

Cantor, G. (1883) 'Über unendliche, lineare Punktmannichfaltigkeiten' (On Infinite Linear Point-Manifolds), in *Gesammelte Abhandlungen*, Berlin, 1932, esp. 179-194, 212.(Important for assessing Bolzano's role in the history of set theory.)

Coffa, J.A. (1991) 'Bolzano and the Birth of Semantics', in *The Semantic Tradition from Kant to Carnap*, Cambridge: University Press, chap. 2, 22-40.(Useful overview with an interesting historical perspective.)

Dummett, M. (1993) *Origins of Analytical Philosophy*, London: Duckworth.(Places Bolzano in the prehistory of analytic philosophy and phenomenology.)

George, R. (1983) 'Bolzano's Consequence, Relevance and Enthymemes', *Journal of Philosophical Logic* 12: 299-325.(Points out significant differences between Bolzano's logic and that of the Frege-Russell tradition.)

Künne, W. (forthcoming) *Bernard Bolzano - Ein analytischer Philosoph im Schatten des deutschen Idealismus (Bernard Bolzano - An Analytic Philosopher in the Shadow of German Idealism)*.(Expands on all material in this entry.)

Morscher, E. (1973) *Das logische An-sich bei Bernard Bolzano (Logical Objects in Bolzano)*, Salzburg: Pustet. (Extremely well-documented and penetrating study of the topics covered in §2.)

Morscher, E. (ed.) (1987) 'Bolzano-Studien', *Philosophia Naturalis* 24 (4).(Useful collection of articles, some of them in English, on various parts of Bolzano's philosophy, including a biographical sketch.)

Peirce, C.S. (1909) 'Meaning (Pragmatism)', in *The New Elements of Mathematics*, The Hague and Paris: Humanities Press, 1976, vol. 4, esp. 117.(Interesting remarks onBolzano as mathematician and logician.)

Quine, W.V. (1954) 'Carnap and Logical Truth', *The Ways of Paradox*, Cambridge, MA: Harvard University Press, 1976, esp. §2, 109-.(Key text for comparing Bolzano's conception of logically analytic truth with modern notions of logical truth.)

Tarski, A. (1936) 'On the Concept of Logical Consequence', *Logic, Semantics, Metamathematics*, Oxford: Clarendon Press, 1956, 408-20.(Key text for comparing Bolzano's conception of deducibility with modern notions of logical consequence.)

Bonaventure (*c.*1217-74)

Bonaventure (John of Fidanza) developed a synthesis of philosophy and theology in which Neoplatonic doctrines are transformed by a Christian framework. Though often remembered for his denunciations of Aristotle, Bonaventure's thought includes some Aristotelian elements. His criticisms of Aristotle were motivated chiefly by his concern that various colleagues, more impressed by Aristotle's work than they had reason to be, were philosophizing with the blindness of pagans instead of the wisdom of Christians.

To Bonaventure, the ultimate goal of human life is happiness, and happiness comes from union with God in the afterlife. If one forgets this goal when philosophizing, the higher purpose of the discipline is frustrated. Philosophical studies can indeed help in attaining happiness, but only if pursued with humility and as part of a morally upright life. In the grander scheme of things, the ascent of the heart is more important than the ascent of the mind.

Bonaventure's later works consistently emphasize that all creation emanates from, reflects and returns to its source. Because the meaning of human life can be understood only from this wider perspective, the general aim is to show an integrated whole hierarchically ordered to God. The structure and symbolism favoured by Bonaventure reflect mystical elements as well. The world, no less than a book, reveals its creator: all visible things represent a higher reality. The theologian must use symbols to reveal this deeper meaning. He must teach especially of Christ, through whom God creates everything that exists and who is the sole medium by which we can return to our creator.

Bonaventure's theory of illumination aims to account for the certitude of human knowledge. He argues that there can be no certain knowledge unless the knower is infallible and what is known cannot change. Because the human mind cannot be entirely infallible through its own power, it needs the cooperation of God, even as it needs God as the source of immutable truths. Sense experience does not suffice, for it cannot reveal that what is true could not possibly be otherwise; so, in Bonaventure's view, the human mind attains certainty about the world only when it understands it in light of the 'eternal reasons' or divine ideas. This illumination from God, while necessary for certainty, ordinarily proceeds without a person's being conscious of it.

1 Life and works

The scholastic philosopher and theologian John of Fidanza, honoured as the 'Seraphic Doctor' but better known as Bonaventure, was born in Bagnoregio, a small town in Tuscany. After preparatory education in his home town, he completed a master's degree in arts at the University of Paris. There he joined the Franciscan order (*circa* 1243), taking the name Bonaventure probably to mark his entry into religious life. Bonaventure went on to study theology under the leading Franciscan masters at Paris during the years 1243-8, first with Alexander of Hales, a famous theologian whose influence is evident in Bonaventure's works, then with John of La Rochelle and Odo Rigaud. As an advanced theology student, he lectured on the Bible (1248-50) and the *Sentences* of Peter Lombard (1250-2). From 1253 to 1257, when he resigned his position to serve as minister general of the Franciscan Order, Bonaventure was regent master of the Franciscan school at Paris. Works composed during this period include the disputed questions *De scientia Christi (Concerning Christ's Knowledge)* and *De mysterio Trinitatis (On the Mystery of the Trinity)*. The *Breviloquium*, a highly condensed summary of theology for beginners, probably dates from around 1257.

The Franciscans were badly divided when Bonaventure became the Order's head. Many believed that the prophecies of Joachim of Fiore concerning a new, spiritual age of history were to be fulfilled by drastic institutional changes and commitment to the life of wandering mendicancy exemplified by St Francis. Though Joachim's teachings influenced Bonaventure's view of history, he himself thought that only a chosen few were suited to live as St Francis did: the time had not yet come for the transformation of the world. Bonaventure did his best to control the extremists, give the order a firmer institutional structure and articulate an understanding of Franciscan poverty compatible with university study and teaching.

Bonaventure's struggles as minister general often receive only passing mention in philosophical accounts of his work, and his conflicts with heterodox masters of arts at Paris naturally appear to have greater philosophical

import. However, the emphasis on the latter tends to conceal one of the most serious dilemmas Bonaventure faced. As he opposed the cult of Aristotle at Paris, so he opposed the cult of apocalyptic asceticism in his own order. The same thinker now sometimes regarded as an anti-intellectual theologian was regarded by some of his own confrères as a creature of the universities who had betrayed the spirit of St Francis. When viewed in context, Bonaventure's writings are remarkable less for the occasional polemics than for the balanced vision he consistently worked to communicate.

After two years of governing the Franciscan Order, Bonaventure retreated to Mount Alverno in Italy, where he wrote the *Itinerarium mentis in Deum (Journey of the Mind to God)*. The intellectual ascent described in this work suggests a solution to the problem of learning within the Franciscan Order (see Brown 1993). Bonaventure does not urge his readers to repudiate the world; he tells us how we may see there the power, wisdom and goodness of its creator. Beginning with contemplation of sensible things, we may rise by stages to contemplation of the soul as the image of God, God's presence within the soul and the attributes of God himself. Philosophical studies aid in the ascent, but only if pursued with humility and as part of a morally upright life: the external world has little use as a mirror 'unless the mirror of our soul has been cleansed and polished.' Bonaventure explains the relations between philosophy and other divisions of human knowledge in *De reductione artium ad theologiam (On Retracing the Arts to Theology)*, a highly condensed treatise of unknown date, possibly based on a sermon preached towards the end of his regency at Paris.

Bonaventure's 'collations' - *Collationes de decem praeceptis (On the Ten Commandments)*, *Collationes de septem donis Spiritus Sancti (On the Seven Gifts of the Holy Spirit)* and *Collationes in Hexaemeron (On the Six Days of Creation)* - represent three series of sermons given during Lent to Franciscans studying and teaching theology at the University of Paris. The third series was cut short in May of 1273, when Pope Gregory X appointed Bonaventure cardinal bishop of Albano. He left Paris to meet with the pope, to be consecrated as bishop (November 1273) and to help with preparations for the Second Council of Lyons. He died unexpectedly in July 1274, shortly before the Council ended, and was buried in the Franciscan church at Lyons.

The three series of Lenten sermons mentioned above bear witness to growing tensions at Paris. Some members of the arts faculty at the University of Paris had taken to defending views contrary to Christian doctrine - defending them, if not as true in the absolute sense, at least as positions that human reason, unaided by faith, seemed legitimately to reach. They tended to see Aristotle as the pinnacle of human reason, and Averroes, who advocated the separation of philosophy from theology, as Aristotle's foremost interpreter (see Aristotelianism, medieval; Averroism). Scandalized by the trend toward neo-pagan philosophizing, Bonaventure in his sermons cites various doctrines popular in the arts faculty as examples of the errors philosophers inevitably make when their reasoning is unillumined by faith. Some of the same views he attacked in 1267 and 1268 were formally condemned by the Bishop of Paris in 1270, but apparently with little effect. Bonaventure's sermons of 1273 accordingly display an even greater sense of urgency. The crisis culminated, three years after his death, in the most extensive doctrinal condemnation of the Middle Ages.

Bonaventure's works fall into different genres, related to the two stages of his career: the period when he was studying and teaching at Paris and the period beginning in 1257, when he resigned his university position to become minister general of the Franciscans. Most works from the first period are academic exercises with a conventional form. In his commentary on the *Sentences* and disputed questions, for example, Bonaventure cites authorities and marshals arguments on both sides of an issue before presenting his own resolution and answering objections to it. His later works differ strikingly from these academic compositions. The *collationes*, it should be remembered, are university sermons, and even the *Itinerarium* has some of the hallmarks of a sermon. The genre should be kept in mind when comparing Bonaventure's later works with academic productions by contemporaries, such as the *Summa theologiae* of Thomas Aquinas.

2 Neoplatonism and mysticism

In the second stage of his career Bonaventure's writing becomes rich in metaphor and heavily reliant on symbolic modes of expression. An elaborate hierarchical structure, itself of symbolic significance, replaces the pro-and-con method characteristic of scholastic compositions. Here we see a writer who, though eloquent, operates within a carefully controlled structure. Unfortunately, even the best translations cannot preserve the many layers of meaning in Bonaventure's language. Today's readers are likely to recognize that the metaphysics of participation

and exemplarism, like the themes of light and illumination, emanation and return, have older philosophical roots in Plato and the ancient Neoplatonists. At the same time, they may be tempted to dismiss all the complicated divisions and subdivisions in Bonaventure's writings as so much stylistic embellishment. The temptation should be resisted, for what looks like ornament turns out to be more substantial.

Works like the *Itinerarium* and the *collationes* draw on a Christian tradition well known to Bonaventure's contemporaries but now generally neglected. Perhaps the most basic idea is that all creation emanates from, mirrors and returns to its source - a fundamentally Plotinian doctrine prominent in the writings of Augustine (see Plotinus). Because the meaning of human life can be understood only from this broader perspective, the overarching aim is to show an interrelated totality hierarchically ordered to God. The structure of Bonaventure's later works reflects this ordering. As the structure is significant, so too is the symbolism, derived partly from mystical writings by the twelfth-century school of St Victor. Building on the thought of Augustine and the Pseudo-Dionysius, Hugh of St Victor taught that God communicates his plans everywhere, in a symbolic language we must learn to recognize and decipher. The world is like a book, revealing its author no less than Scripture does (see Zinn 1973). All creation reflects the triune character of God; all visible things represent an invisible, transcendent reality. The theologian must accordingly use symbols to show what cannot be expressed more directly. He must demonstrate the deeper meaning and order of the world, helping us to rise above the confusing multiplicity presented by our senses. To do so, he must teach especially of Christ, the Word through whom all things are created, the source of all true wisdom and the medium enabling us to return to God (see Mysticism, history of).

This sketch of the intellectual background may help to explain not only the hierarchical structure of Bonaventure's later works but also the significance he attached to numbers: threefold divisions symbolize the Trinity, sixfold divisions symbolize the six days of creation and so on (see Neoplatonism). Of course, Neoplatonic and mystical writings are hardly Bonaventure's only sources. Like all masters schooled at Paris in the mid-thirteenth century, he studied Aristotle closely and often appealed to Aristotle's teachings. Stoic doctrines, derived from Cicero and patristic writers, are equally present in his thought (see Patristic philosophy; Stoicism). The philosophical eclecticism displayed by the earliest Christian thinkers is even greater among the scholastics, so that scholarly debates about whether a given master is Aristotelian, Neoplatonic or something else - debates extremely common in the literature on Bonaventure - are fundamentally disagreements about the dominant strains in a mixed breed.

3 Happiness and the limits of philosophy

What Bonaventure considers to be the place and function of philosophy becomes easier to understand if one reflects upon happiness, the end of human life. By nature we strive for happiness, and we can never be fulfilled unless we attain it. This ultimate goal has both practical and speculative implications. As Bonaventure explains in Chapter 1 of the *Itinerarium*, 'Since happiness is nothing else than the enjoyment of the Supreme Good, and the Supreme Good is above us, no one can enjoy happiness unless he rise above himself - not, of course, by a bodily ascent, but by an ascent of the heart.'

The ascent of the heart involves the ascent of the mind, and yet the mind takes us only so far. The complete happiness experienced through union with God comes more from love than from knowledge, more from will than from intellect. Small wonder, then, that Bonaventure consistently emphasizes the higher purpose of speculation. Anyone who remains ignorant of this purpose, or who knows it but loses sight of it, is doomed not only to frustrated desire but also to intellectual confusion. A good philosopher must therefore understand the place of philosophy within the wider context of human life. As no one can perceive the beauty of a poem unless he sees the whole, Bonaventure argues, so no one can grasp the beauty of the order regulating reality unless he views it in its totality (see Faith).

Bonaventure's conviction that union with God constitutes the ultimate goal of all learning might well lead one to wonder whether philosophy has any genuine place in his thought. Why devote time to studying logic, much less the writings of pagan philosophers, when the example of St Francis proves such studies unnecessary for attaining the goal of eternal happiness? The title of one of Bonaventure's best-known treatises, *De reductione artium ad theologiam*, can reinforce the impression that he values the art of philosophy only insofar as philosophy can be 'reduced' to theology. This impression is misleading not only because the title, like the titles of many scholastic

works, was invented by later editors rather than the author himself, but also because the term *reductio* is properly translated as 'retracing' rather than 'reducing'.

For Bonaventure, *reductio* signifies both a process and a method of analysis (Bougerol 1964, 1988). As a process, it is the return to God, in whom all rational beings find their fulfilment. While individuals may or may not choose the path that leads them home, the process of humanity's return to its source continues to operate. God's grace, working ever to draw us back to himself, governs the historical process and gives it meaning. Hence, *reductio* can also serve as a method of analysis. Humans, as rational beings, find the path home more easily when someone shows them how the various areas of human knowledge are so many rungs of a ladder, all leading in the same direction. As a method of analysis, *reductio* accordingly seeks to demonstrate how the many lead to the one, the composite to the simple, even as the many and the composite are based on and derive a deeper meaning from the one and the simple. The aim is to articulate an intelligible order and direction by bringing out fundamental organizing principles; it is not to boil away everything else, as cooks do in 'reducing' a sauce. While philosophy should point in the right direction and understand its place in the wider context, it need not be distilled into theology to become valuable. Logic is well worth studying, even though it teaches us nothing about the meaning of life.

The culinary interpretation of Bonaventure's *reductio* of the arts to theology arises partly from his claim that all divisions of human knowledge are 'handmaids of theology'. They surely are 'handmaids' in the sense that they have a higher purpose. Anyone who pursues philosophy or any other art strictly as an end in itself, as if there were nothing more exalted or fulfilling, is making a disastrous mistake. On the other hand, to say that philosophy is not the pinnacle of human life is not to say that it has value only insofar as it can be distilled into theology. As Bonaventure makes philosophy a handmaid of theology, so he also makes agriculture and navigation handmaids of theology. Is he arguing that farmers are worthless unless they sow their crops in threes, or that sailors cannot chart a course to Rome without fixing on God? Not at all. To argue that an art has a higher purpose is not to insist that it has no legitimacy or value in its own domain.

Bonaventure's attitude towards Aristotle's philosophy should be understood within the context of his views on philosophy in general. In the *Collationes in Hexaemeron*, where he traces the most serious errors of philosophy to Aristotle's rejection of exemplar ideas and his affirmation that God knows only himself, Bonaventure's preference for a broadly Platonic metaphysics is evident. His diatribe, however, is much less against Aristotle than against members of the Paris arts faculty who defended as philosophically sound virtually every position they believed Aristotle had held. In Bonaventure's view, even the best philosophers of antiquity were ignorant of original sin and the need for God's grace. Knowing nothing of Christ, and having to rely exclusively on their own powers of reason, the ancients could not help but make grave mistakes. Bonaventure accordingly urges the arts masters to philosophize as Christians, with the illumination of faith to guide them (see Aristotelianism, medieval; Platonism, medieval).

4 Knowledge of God's existence

Though Bonaventure offered proofs for the existence of God in other works, the twenty-nine arguments in his disputed questions *De mysterio Trinitatis* represent his most developed treatment of the topic (see Doyle 1974). The arguments go on to prove three conclusions: that every truth impressed on all minds is an indubitable truth, that every truth which all creatures proclaim is an indubitable truth and that every truth that is in itself most certain and evident is an indubitable truth. Bonaventure's position combines these three conclusions into a single master conclusion: that the existence of God is an indubitable truth in its own right and can be doubted only because of some defect in the knower. A human being might doubt God's existence from failing to understand correctly what the term 'God' signifies. Doubt might likewise spring from failing to carry one's thinking far enough, or from having only a partial view of the evidence. In all cases, Bonaventure argues, doubt arises from failures of the human intellect, not from God's existence as a truth considered in itself.

In arguing that God's existence is self-evident, Bonaventure repeatedly appeals to Anselm's *Proslogion* (see Anselm of Canterbury). He argues, for example, that God is 'a being than which nothing greater can be thought'; but that which exists only in thought does not meet this description, for something existing in reality would be greater, and hence God cannot be thought not to exist in reality. Anselmian arguments are supplemented by Augustinian arguments for the existence of truth (see Augustinianism). To Bonaventure, each particular truth

implies the existence of an absolute truth that is its cause. To affirm any particular truth is thus to affirm, even though one might fail to recognize it, the existence of God (see God, arguments for the existence of).

Bonaventure's arguments reflect his belief that all human souls belong to a hierarchically ordered intelligible realm that is more real and more knowable than the world of the senses. Belonging to the same realm as God, the soul can know God's existence directly, through thought. This position stands in sharp contrast to the views of Bonaventure's contemporary, Aquinas. Sharing Aristotle's belief that we obtain knowledge from the initial data of sense experience, Aquinas taught that we can know God's existence by considering creatures and reasoning from effect to cause, but that God's existence is not self-evident to us.

5 The question of an eternal world

Bonaventure again differs from Aquinas on whether the world could have been created from eternity, a topic of heated debate in the thirteenth century. The problem, as he formulates it in his commentary on the *Sentences*, is 'Whether the world has been produced from eternity or in time.' The idea that the world was indeed produced is crucial to the question. To Bonaventure, there is nothing incoherent in believing that the world is eternal, only in believing that the world is both created and eternal (see Bonansea 1974).

Bonaventure's principal argument rests on the very notion of creation. Whatever is created in the truest sense - that is, produced from nothing rather than from pre-existing material - must come to have 'being after non-being': it must have a *beginning*. To say that something with a beginning is eternal, meaning that it has no beginning, is an obvious contradiction. Thus the thesis that the world was created from eternity is not merely false but unintelligible. In offering this argument, Bonaventure does not mean to imply that creation took place at some point within time, for he accepts the prevailing view that time began with creation. His challenge is to those who conceive of the world as both beginningless and created *ex nihilo*.

Another argument points to a problem with positing an actual, as opposed to merely potential, infinity. Bonaventure reasons that it is impossible to make the actually infinite greater, for this would be contrary to its nature, so that if the world existed from eternity, it would admit of no additional duration. In effect, extra days could not add to the number of days preceding the present, since the number preceding the present would already be infinite, and all actual infinites, Bonaventure assumes, must be equal. How could we even have reached the present day if infinitely many days had to elapse before today? One need not review all of Bonaventure's arguments to see how the notion of a world created from eternity raised issues, such as the problem of unequal infinites, that were of more than theological interest (see Eternity of the world, medieval views of).

6 Metaphysics

Bonaventure's metaphysics, by his own description, is a doctrine of emanation, exemplarism and consummation or return. To his mind, anyone who denies the existence of exemplar ideas might reasonably draw the kind of conclusions drawn by Aristotle: that the first cause knows only itself and 'moves' the world only as a final cause, as an object of desire rather than as an agent and efficient cause. Even though his teachings recall Plotinian doctrine on emanation from the One, Bonaventure's metaphysics owes much more to the distinctively Christian Neoplatonism of Augustine. In place of the One we find God; in place of a world emanating necessarily from the very nature of the One we find a world God created freely, out of nothing. The world comes from God as its efficient cause, mirrors God, the exemplar cause, and is destined to return to God as its end or final cause (see Creation and conservation, religious doctrine of).

As archetypes of all actual and possible creatures, exemplar ideas help to explain both how God made what exists and how he knows everything that could be. The plurality of ideas does, however, raise questions about the unity of God. Bonaventure explains that there is no real plurality of ideas in God. As they are not distinct from God himself, so they are not distinct from each other. Properly understood, the exemplar ideas are distinct only from the standpoint of reason. What 'idea' actually signifies is a creature's relation to God.

Bonaventure's 'universal hylomorphism' - the doctrine that all creatures are composed of matter and form - turns out to be less bizarre than it seems at first glance. For him, 'matter' is a principle of potentiality that may or may not be corporeal. Because all beings other than God are susceptible to change, and change is the actualization of potential, all beings other than God must have an element of potentiality. Thus the 'spiritual matter' of angels is

matter understood on an analogy with the corporeal matter of human beings, similar only in its status as a principle of potentiality and necessary complement to form.

7 The illumination of the intellect

Bonaventure's theory of knowledge roughly conforms to that of Aristotle as regards knowledge of the sensible world. Declaring the human mind a *tabula rasa* (blank slate) at birth, he teaches that we cannot acquire concepts of material objects, much less a knowledge of biology, without abstracting from sense experience. Aristotle's empiricism nonetheless has limitations: it fails to account for our idea of God, and fails even more miserably in accounting for certitude.

Bonaventure does not claim that the idea of God is innate in the sense that it is present at birth or will inevitably develop as the child matures. The idea might nonetheless be considered innate insofar as it is does not depend on abstraction from sense experience. To acquire the idea of God, the soul need only turn inward and reflect on its own nature, or on its natural desire for complete happiness, which God alone can provide. In arguing that the idea of God does not come from presentations of the senses, Bonaventure's purpose is mainly to emphasize that the human soul, made in God's image and belonging to the intelligible realm, need not have recourse to the material world to know its creator.

The theory of illumination, well presented in question 4 of Bonaventure's disputed questions *De scientia Christi*, aims above all to account for the certitude of human knowledge. In his view, there can be no certain knowledge without both infallibility on the part of the knower and immutability on the part of the object of knowledge. Because the mind of a creature cannot be entirely infallible through its own power, it needs the cooperation of God, even as it needs God as the source of immutable truths. Sense experience does not suffice, for it cannot reveal that what is true could not possibly be otherwise. The human mind attains certitude about the world only when it understands it in light of the 'eternal reasons' or divine ideas. In doing so, the mind will ordinarily be unconscious of divine illumination. Certitude, however, would be impossible if God did not provide an immutable object of knowledge and move the mind to assent (see Illumination).

In arguing for his theory of illumination, Bonaventure expressly denies that God is the sole source of human certainty. Such a view would fail to distinguish earthly knowledge from heavenly knowledge, knowledge of nature from knowledge of grace, and knowledge by reason from knowledge by revelation. On the other hand, Bonaventure thinks it insufficient for the knower to benefit from the eternal reasons without attaining to them. The human mind could not acquire certain knowledge if it did not in some way rise above the created order. According to Bonaventure, this is possible because, and insofar as, the soul is the image of God. Sense experience remains necessary as a source of our ideas about the world, but what certainty we attain comes from the cooperation of God.

8 Virtue

Bonaventure's doctrine of virtue, as presented in his *Collationes in Hexaemeron*, clearly reflects his metaphysics and theology (see Synan 1973). Since the end of virtue is happiness, which comes from the enjoyment of God in the afterlife, virtues must be ordered to that end. They must help us return to our creator. Return, however, is possible only through Christ; without Christ, human beings remain infected with original sin and doomed to remain forever separated from God. Indeed, at their fourth, highest level of reality, the virtues exist in Christ as exemplar ideas.

Virtues in human beings participate in these divine exemplars to various degrees. The cardinal virtues of prudence, justice, temperance and fortitude may accordingly be possessed at any of three levels. At the lowest level they are 'political' and belong to us insofar as we are social animals; at the next they are 'cleansing' and belong to us insofar as we are fit for God; and at the next they belong to those already completely cleansed. At all levels of the hierarchy human virtues depend for their reality on the exemplars. The cardinal virtues likewise also depend on the theological virtues of faith, hope and charity to attain their perfection and achieve their ends.

The four-level hierarchy of virtue comes from Macrobius' commentary on Cicero's *Dream of Scipio*, a work that explains the teachings of Plotinus. This Neoplatonic material is nevertheless transformed, not only by the addition of the theological virtues but also by Bonaventure's recasting of the cardinal virtues as products of grace and the

foundation of that ‘merit’ which makes us deserving in the eyes of God. In his commentary on the *Sentences*, he explains that merit is rooted in free decision (*liberum arbitrium*). Thus the cardinal virtues can belong only to the intellect and the will, those powers of the soul that share in free decision. Bonaventure even argues that *all* virtues, insofar as they are virtues, must belong to the rational part of the soul. Virtues are attributed to the lower, emotional part of the soul only because habituation makes it more submissive to reason; they cannot exist principally in the seat of the passions because they would then be beyond the scope of free decision. One would be attributing virtues to the part of the soul we have in common with animals instead of to the part we have in common with angels (see Kent 1995).

For Bonaventure, then, a virtue such as fortitude has less to do with our emotional responses to danger than with what we freely decide. The appropriate emotional responses are not essential to virtue; at most they are ‘annexed’ to it. While this view would seem to have distant origins in the Stoic restriction of virtue to what lies within the agent’s control, the emphasis on virtue as the basis of merit is of no small significance. It is one thing to see virtues as traits of character that make us happy in this life, another to see them as traits of character that make us deserving of happiness in the afterlife.

The influence of theology in Bonaventure’s thought appears to have two results that some of today’s moral philosophers - especially those sympathetic to Kant - would find appealing. First, virtue comes to depend on the freedom that all rational creatures have in common, even as the moral significance of emotions declines sharply. Second, it becomes quite reasonable to say that someone is both virtuous and unhappy, for virtue is now what makes someone deserving of happiness, even though internal emotional conflict, illness and misfortune may in fact cause the person great suffering (see Virtue ethics; Virtues and vices).

See also: Aquinas, T.; Aristotelianism, medieval; Augustinianism; Averroism; Mysticism, history of; Platonism, medieval

BONNIE KENT

List of works

Bonaventure (1250-73) *Opera omnia (Complete Works)*, Quaracchi: Collegium S. Bonaventurae, 1882-1902, 10 vols.(For English translations, see *The Works of Bonaventure*, 5 vols, trans. J. de Vinck, Paterson, NJ: St Anthony Guild, 1960-70.)

Bonaventure (1253-7) *De scientia Christi (Concerning Christ’s Knowledge)*, Q.4 published in E. Fairweather (ed.) *A Scholastic Miscellany*, Philadelphia, PA: Westminster, 1956.(Argues that divine illumination is necessary to account for the certitude of human knowledge.)

Bonaventure (1253-7) *De mysterio Trinitatis (On The Mystery of the Trinity)*, trans Z. Hayes, *Disputed Questions on the Mystery of the Trinity*, St Bonaventure, NY: Franciscan Institute, 1979.(Includes detailed analysis of arguments for God’s existence.)

Bonaventure (*c*.1257) *Breviloquium*, in *Opera omnia*, Quaracchi: Collegium S. Bonaventurae, 1882-1902, vol. 5; trans. J. de Vinck, *The Works of Bonaventure*, Paterson, NJ: St Anthony Guild, 1960-70, vol. 2.(A brief handbook of theology for beginners.)

Bonaventure (1259) *Itinerarium mentis in Deum (Journey of the Mind to God)*, trans. P. Boehner, *The Journey of the Mind to God*, ed. S. Brown, Indianapolis, IN: Hackett, 1993.(The mind’s ascent from the contemplation of the sensible world to the contemplation of God.)

Bonaventure (1267) *Collationes de decem praeceptis (On the Ten Commandments)*, in *Opera omnia*, Quaracchi: Collegium S. Bonaventurae, 1882-1902, vol. 5.(Christian teachings with the ethics of the pagan philosophers.)

Bonaventure (1268) *Collationes de septem donis Spiritus Sancti (On the Seven Gifts of the Holy Spirit)*, in *Opera omnia*, Quaracchi: Collegium S. Bonaventurae, 1882-1902, vol. 5.(Discusses piety, understanding, wisdom and other dispositions attributed by the theologian to God’s grace.)

Bonaventure (1273) *Collationes in Hexaemeron (On the Six Days of Creation)*, ed. F. Delorme, *Collationes in Hexaemeron*, Quaracchi: Collegium S. Bonaventurae, 1934.(This edition is preferred by some scholars to that published in the *Opera omnia*.)

Bonaventure (before 1274) *De reductione artium ad theologiam (Retracing the Arts to Theology)*, in *Opera omnia*, Quaracchi: Collegium S. Bonaventurae, 1882-1902, vol. 5; trans. J. de Vinck, *The Works of Bonaventure*, Paterson, NJ: St Anthony Guild, 1960-70, vol. 3.(The divisions, order and unity of human knowledge.)

Bonaventure (1250-2) *Commentarius in quattuor libros Sententiarum Petri Lombardi (Commentary on the*

Sentences of Peter Lombard), in *Opera omnia*, Quaracchi: Collegium S. Bonaventurae, 1882-1902, vols 1-4.(A selection, specifically *Sent*. II, d.1, p.1, a.1, q.2, is translated by P.M. Byrne in C.Vollert *et al.*, *On the Eternity of the World (Thomas Aquinas, Siger of Brabant, St Bonaventure)*, Milwaukee, WI: Marquette University Press, 1964.)

References and further reading

Bonansea, B. (1974) 'The Question of an Eternal World in the Teaching of St Bonaventure', *Franciscan Studies* 34: 7-33.(More complete and detailed presentation of Bonaventure on the eternity of the world.)

Bougerol, J.G. (1964) *Introduction to the Works of Bonaventure*, trans. J. de Vinck, Paterson, NJ: St Anthony Guild.(Expanded discussion of Bonaventure's life and works.)

Bougerol, J.G. (1988) *Introduction à Saint Bonaventure*, Paris: Vrin.(A revised edition of Bougerol (1964), in French.)

Brown, S. (1993) *The Journey of the Mind to God*, Indianapolis, IN: Hackett.(Brown's introduction and notes to this translation of *Itinerarium mentis in Deum*, written with elegant simplicity, provide valuable guidance not only to this particular treatise but to Bonaventure's thought as a whole.)

Doyle, J.P. (1974) 'Saint Bonaventure and the Ontological Argument', *Modern Schoolman* 52: 27-48.(A clear exposition of Bonaventure's arguments on the knowledge of God's existence and their Platonic background.)

Gilson, E. (1965) *The Philosophy of St Bonaventure*, trans. I. Trethowan and F. Sheed, Paterson, NJ: St Anthony Guild.(The standard exegesis of Bonaventure's philosophy, tendentious in places but probably still unrivalled.)

Kent, B. (1995) *Virtues of the Will: The Transformation of Ethics in the Late Thirteenth Century*, Washington, DC: Catholic University of America.(Discusses virtue and Bonaventure's *Sentences* commentary.)

Quinn, J.F. (1973) *The Historical Constitution of St Bonaventure's Philosophy*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(Provides a wealth of information about Bonaventure's thought while differing sharply from Gilson's interpretation.)

Synan, E. (1973) 'Cardinal Virtues in the Cosmos of Saint Bonaventure', in J.G. Bougerol *et al.* (eds) *S. Bonaventura, 1274-1974*, Grottaferrata: Collegium S. Bonaventurae, vol. 3: 21-38.(A brief but clear account of the moral doctrine of the *Hexaemeron* and its connection with exemplarism.)

Vollert, C. *et al.* (trans.) (1964) *On the Eternity of the World (Thomas Aquinas, Siger of Brabant, St Bonaventure)*, Milwaukee, WI: Marquette University Press.(Includes a translation of a section from Bonaventure's commentary on the Sentences.)

Zinn, G. (1973) 'Book and Word: The Victorine Background of Bonaventure's Use of Symbols', in J.G. Bougerol *et al.* (eds) *S. Bonaventura, 1274-1974*, Grottaferrata: Collegium S. Bonaventurae, vol. 2: 143-69.(Detailed look at mysticism in Bonaventure's philosophy.)

Bonhoeffer, Dietrich (1906-45)

Dietrich Bonhoeffer was a twentieth-century Lutheran theologian who associated Christian belief and political action in an exemplary fashion. His part in the struggle of the Confessing Church and of the German resistance against the National-Socialist dictatorship cost him his life. Christocentric and ecclesiocentric, he stressed personal and collective piety and revived the idea of the imitation of Christ; the concepts of obedience and of the suffering God are central to his view. His Ethik *(1949) was widely influential; in it, he argued that Christians should not retreat from the world, but have a duty to act within it. His answer to the secularization of the modern world was a 'religionless Christianity', a communocentric, pietistic, personal discipline.*

1 Life and theological background

Born in Breslau, Bonhoeffer was raised in Berlin in an intellectual milieu. Early in his life he felt a vocation to be a theologian and studied at Tübingen in 1923 and Berlin from 1924 to 1927. His dissertation was a theological reflection on the sociology of the Church, which critically examined Hegel, Max Weber and Ernst Troeltsch. Bonhoeffer had just discovered Karl Barth's dialectical theology and was anxious to relate Christology to ecclesiology (see Barth, K. §§1-2). In opposition to liberal and sociological views, he asserted that the Church is the body of Christ concretely. His other early writings show the influence of Kant, Heidegger, Barth and Luther.

Following the rise of Hitler, Bonhoeffer helped to organize the Pastor's Emergency League, and took a stand against the racist policies of National Socialism, publishing 'Die Kirche vor der Judenfrage' (The Church and the Jewish Question) (1933a) and Der Arierparagraph in der Kirche (The Aryan Paragraph in the Church) (1933b). His involvement with the Confessing Church, which Karl Barth had helped organize, led to his being banned from teaching and publishing. In 1938, he came into contact with the political resistance against National Socialism. He worked with the resistance until 1944, when, in the wake of the failed attempt on Hitler's life, he was arrested, imprisoned and later hanged.

Bonhoeffer's original contribution to Christian theology was prompted by historical circumstances, particularly the National Socialist dictatorship. However, a number of quite varied influences can be detected in his work, notably dialectical theology, which he discovered in the early 1930s, Moravian Pietism, Lutheranism, Roman Catholicism, and the neo-Protestantism of Wilhelm Herrmann, Adolf von Harnack, Ernst Troeltsch, Max Weber, Martin Kähler, Rudolf Otto, Karl Holl and Wilhelm Dilthey. Bonhoeffer's critical stance towards the phenomenon of religion was inspired by Barth's distinction between faith and religion, and his assertion that revelation abolished religion.

Bonhoeffer understood Christian faith as a combination of Old Testament legalism and New Testament Christocentrism. It seems that he experienced a conversion in 1931, while studying the Sermon on the Mount and Psalm 119, the famous love poem to the Law. To him, the central issue was that of obedience, and the concept of discipleship grew into a major theological theme. It culminated in *Nachfolge (Following After)* (1937), a study of the imitation of Christ in the Gospels, in which Bonhoeffer argues that 'only the believer is obedient, and only the obedient believes'. He believed that faith consists in following in Jesus Christ's footsteps by accepting God's messianic suffering. Two elements are constitutive of faith: the implementation of justice and the acceptance of divine suffering. Jesus' agony only ends with the end of the world. In Bonhoeffer's theology of the cross, God is not the triumphant and enthroned God of Isaiah 6, but a God of suffering, suffering in the world and for the world. Discipleship, then, is following Christ at Gethsemane.

2 Ethics

Begun in 1940, Bonhoeffer's *Ethik* was published as fragments in 1949 by Eberhard Bethge. It is Bonhoeffer's *magnum opus*, and must be understood against the background of his political involvement. In June 1939, he had refused to move to the USA, even though staying in Germany meant accepting a double life, simultaneously this-worldly and other-worldly. For Europe was in a such a predicament that, in his view, Christians were not morally allowed to retreat from the world. On the contrary, it behoved Christians to be involved in worldly affairs and to collaborate with non-Christians. Consequently, Christian theology had to address the question of the possibility of acknowledging the world from a Christocentric perspective. Bonhoeffer propounded five major ideas. First, though distinct, the Christian community and the world should not be kept apart. Christian action

always is action in the world. Faith is not an escape from the world, but is existentially this-worldly. Second, the world is the world accepted by Christ, who is the uniting principle of Western civilization. Third, a Christian's participation in the reality of Jesus Christ and in the reality of the world, even a Godless world, is possible because in Christ God enabled us to participate concomitantly in a dual reality, that of God and that of the world. Fourth, only in the reality of the world can one truly be a Christian. But there is no real worldliness except in the reality of Jesus Christ. Christ's sovereignty is the focal concept of Christian ethics. Finally, penitence clears the path for the imitation of Christ or discipleship.

3 Nonreligious interpretation of Christianity

In a letter from prison, dated 30 April 1944, Bonhoeffer introduces and discusses the concept of religionless Christianity. He often referred to this idea as the nonreligious interpretation of Christianity. As early as his dissertation, *Sanctorum communio* (1930), Bonhoeffer had argued for a nonreligious Christianity. His diagnosis of the 'churchliness of the modern bourgeoisie' led him to call the Church back to a more sincere and genuine community modelled on Christ. He was dismayed by the empty religiosity of the Church and further outraged by its passive enslavement to National Socialism. Under the joint influences of Dilthey, Kant and Barth, he developed a negative concept of religion. According to Bonhoeffer, Christian religion is 'a historically conditioned and transient form of human self expression'. Religion is a human answer to anxiety. In the human quest for security, God is the ultimate refuge. Towards the close of the Middle Ages, people ceased to believe in the authority of the Church. European thought turned away from other-worldliness towards this-worldliness, intellectual autonomy was established and personal emancipation encouraged. The Enlightenment and the subsequent nineteenth-century revolutions anticipated 'the world come of age' of the twentieth century. Progressive secularization transformed the Christian world into a world without God. The religious interpretation of life and history has ended, and religion is ending as well. This process is irreversible. The end of religion also implies the end of traditional Christianity. In a world beyond Christendom, a world come to adulthood and further progressing towards a religionless time, the only suitable apologetics consists in providing a nonreligious interpretation of Scripture. The Bible is not a religious book, but the story of a people progressing through suffering towards justice and God's kingdom.

Bonhoeffer's response to the demise of Christianity was both pietistic and ethical; it was a theological extrapolation of the piety of his home life. His call for a nonreligious Christianity was a call for a sincere Christocentric and communocentric Christianity. It was an attempt to bring Christian life into harmony with Christian discourse. In our secular world, Christianity can recover its identity only as a personal and secret discipline. In his most widely read book, *Gemeinsames Leben (Life Together)* (1939), he argued for the renewal of some forms of monastic life in order to serve the world. Bonhoeffer's courageous death at the hands of the National Socialists testified to his belief that 'before God and with God we live without God'.

JEAN-LOUP SEBAN

List of works

Bonhoeffer, D. (1986-96) *Dietrich Bonhoeffer Werke*, ed. E. Bethge *et al*., Munich: Chr. Kaiser Verlag, 16 vols; various translators, ed. W.W. Floyd, Jr, *Dietrich Bonhoeffer's Works*, Minneapolis, MN: Fortress Press, 1995-. (The new critical edition, with extensive introductions, notes and bibliographies; the English version is still appearing at the time of writing.)

Bonhoeffer, D. (1965-7) *Gesammelte Schriften (Collected Writings)*, ed. E. Bethge, Munich: Chr. Kaiser Verlag, 5 vols; selections trans. E.H. Robertson and J. Bowden, ed. E. Robertson, *No Rusty Swords: Letters, Lectures and Notes 1928-1936*; *The Way to Freedom: Letters, Lectures and Notes 1935-1939*; *True Patriotism: Letters, Lectures and Notes 1939-1945*, London: Collins, 1965-7.(The English volumes selected from this earlier complete edition contain much that is still not available elsewhere, as well as a detailed chronology of Bonhoeffer's life.)

Bonhoeffer, D. (1930) *Sanctorum communio: eine Untersuchung zur Sociologie der Kirche*, Berlin and Frankfurt-an-der-Oder: Trowitzsch & Sohn; trans. R. Gregor Smith, *Sanctorum Communio: A Dogmatic Inquiry into the Sociology of the Church* [in the USA: *The Communion of Saints:…*], London: Collins, 1963. (Bonhoeffer's 1927 Berlin doctoral dissertation, a somewhat technical sociological and theological discussion of personhood, community, sin and the Church.)

Bonhoeffer, D. (1931) *Akt und Sein: Transzendentalphilosophie und Ontologie in der systematischen Theologie*

(Act and Being: Transcendental Philosophy and Ontology in Systematic Theology), Gütersloh: C. Bertelsman; trans. B. Noble, Act and Being, London: Collins, 1964.(Bonhoeffer's *Habilitationsschrift*, a highly technical piece in which he claims that theology can mediate between an act-based philosophy associated with Kant - and Barth in theology - and a being-based philosophy associated with Heidegger, among others.)

Bonhoeffer, D. (1933a) 'Die Kirche vor der Judenfrage', in *Gesammelte Schriften*, ed. E. Bethge, vol. 2, Munich: Chr. Kaiser Verlag, 1965-7; trans. E.H. Robinson and J. Bowden, 'The Church and the Jewish Question', in *No Rusty Swords: Letters, Lectures and Notes 1928-1936*, ed. E.H. Robertson, London: Collins, 1965-7.(After the boycott of Jewish businesses was declared on 1 April 1933 and the German Church banned Jews from positions of authority on 7 April, Bonhoeffer delivered this lecture against anti-Semitism, trying to reclaim pro-Semitic emphases in Luther.)

Bonhoeffer, D. (1933b) 'Der Arierparagraph in der Kirche' (The Aryan Paragraph in the Church), in *Gesammelte Schriften*, ed. E. Bethge, vol. 2, Munich: Chr. Kaiser Verlag, 1965-7.(In this leaflet, Bonhoeffer presented a set of theses and antitheses, arguing against the anti-Semitic stance of the German Christians.)

Bonhoeffer, D. (1933c) *Schöpfung und Fall: theologische Auslegung von Genesis 1-3*, Munich: Chr. Kaiser Verlag; trans. J.C. Fletcher, *Creation and Fall: A Theological Interpretation of Genesis 1-3*, London: SCM, 1959.(In a very different mode, Bonhoeffer interprets Genesis in a way reminiscent of Karl Barth on Romans: treating it as a book of the Church, to be interpreted in and for the Church.)

Bonhoeffer, D. (1933d) 'Christologie', in *Gesammelte Schriften*, ed. E. Bethge, vol. 3, Munich: Chr. Kaiser Verlag, 1965-7; trans. J. Bowden, *Christology* [*Christ the Center* in the USA], London: Collins, 1961.(An unfinished lecture course, which treats of 'the present Christ' and 'the historical Christ', and presents Christ as the boundary and hidden centre of reality.)

Bonhoeffer, D. (1937) *Nachfolge (Following After)*, Munich: Chr. Kaiser Verlag; trans. R.H. Fuller, *The Cost of Discipleship*, London: SCM, enlarged edn, 1959.(A very influential nontechnical condemnation of 'cheap grace' - a reminder of the way of the Cross, and the harsh calls of Jesus' Sermon on the Mount.)

Bonhoeffer, D. (1939) *Gemeinsames Leben*, Munich: Chr. Kaiser Verlag; trans. J.W. Doberstein, *Life Together*, New York: Harper, 1954.(A description of Christian community, based, like *The Cost of Discipleship*, on Bonhoeffer's experience of common life in an irregular seminary in Finkenwalde, closed by the Gestapo in 1937.)

Bonhoeffer, D. (1949) *Ethik*, ed. E. Bethge, Munich: Chr. Kaiser Verlag; trans. N. Horton Smith, *Ethics*, London: SCM, 1955.(An incomplete and fragmentary work, *Ethik* was written secretly by Bonhoeffer in the years before his arrest, and deals very practically with Christian citizenship and conformity to Christ.)

Bonhoeffer, D. (1951) *Widerstand und Ergebung: Briefe und Aufzeichnungen aus der Haft (Resistance and Humility: Letters and Notes from Prison)*, Munich: Chr. Kaiser Verlag; trans. R. Fuller *et al.*, *Letters and Papers from Prison*, London: SCM, enlarged edn, 1971.(Now a religious classic, this collection casts much light on the directions Bonhoeffer's thought was taking when he died. Includes 'After Ten Years', Bonhoeffer's own assessment of the 1930s and early 1940s.)

Bonhoeffer, D. (1990) *Testament to Freedom: The Essential Writings of Dietrich Bonhoeffer*, ed. G.B. Kelly and F.B. Nelson, New York: Harper.(A useful introductory reader with a long introductory essay, and a selection of letters, poems and essays, each with a short introduction.)

References and further reading

Barth, K. (1919) *Der Römerbrief*, Bern: Bäschlin; 2nd edn, much changed, Munich: Chr. Kaiser Verlag, 1922; 6th edn trans. E.C. Hoskyns, *The Epistle to the Romans*, Oxford: Oxford University Press, 1968.(Barth was a strong influence on Bonhoeffer, whose *Creation and Fall* resembles this book.)

Bethge, E. (1967) *Dietrich Bonhoeffer*, Munich: Chr. Kaiser Verlag; trans. E. Mossbacker *et al.*, *Dietrich Bonhoeffer: Theologian, Christian, Contemporary*, London: Collins, 1970.(A massively detailed and absolutely essential biography from one of Bonhoeffer's closest friends.)

Dumas, A. (1968) *Une théologie de la réalité: Dietrich Bonhoeffer*, Geneva: Éditions Labor et Fides; trans. R. McAfee Brown, *Dietrich Bonhoeffer: Theologian of Reality*, London: SCM, 1971.(Something of a classic interpretation of Bonhoeffer's thought, stressing the continuities of the corpus, and Bonhoeffer's relationship to Hegel.)

Feil, E. (1971) *Die Theologie Dietrich Bonhoeffers*, Munich: Kaiser & Grumwald; trans. M. Rumscheidt, *The Theology of Dietrich Bonhoeffer*, Philadelphia, PA: Fortress Press, 1985.(A good, solid investigation of Bonhoeffer's theology.)

Floyd, W.W. and Marsh, C. (eds.) (1994) *Theology and the Practice of Responsibility: Essays on Dietrich Bonhoeffer*, Valley Forge, PA: Trinity Press International.(A collection of recent scholarship: Bonhoeffer and

modernity, and social analysis, and postmodernism, and so on.)

Kelly, G.B. (1984) *Liberating Faith: Bonhoeffer's Message for Today*, Minneapolis, MN: Augsburg.(An accessible, nontechnical introduction to Bonhoeffer's life and work.)

Marsh, C. (1994) *Reclaiming Dietrich Bonhoeffer: The Promise of His Theology*, Oxford: Oxford University Press.(Marsh reads Bonhoeffer as in pursuit of a Christologically redescribed philosophy within a roughly Barthian theology. Technical but clear; includes material on Bonhoeffer's relationship to Hegel and Heidegger.)

Ott, H. (1966) *Wirklichkeit und Glaube*, vol. 1: *Zum theologischen Erbe Dietrich Bonhoeffers*, Zurich: Vandenhoeck & Ruprecht; trans. A.A. Morrison, *Reality and Faith: The Theological Legacy of Dietrich Bonhoeffer*, London: Lutterworth, 1971. (A lengthy and detailed examination of Bonhoeffer's work in the context of German theology and philosophy.)

Peck, W.J. (ed.) (1987) *New Studies in Bonhoeffer's Ethics*, Lewiston and Queenstown: Edwin Mellen.(Useful, detailed essays on the text, context and content of Bonhoeffer's *Ethik*.)

Zimmerman, W.D. and Smith, R.G. (eds.) (1964) *Begegnungen mit Dietrich Bonhoeffer: Ein Almanach (Encounters with Dietrich Bonhoeffer)*, Munich: Chr. Kaiser Verlag; trans. K.G. Smith, *I Knew Dietrich Bonhoeffer*, London: Collins, 1966. (Chronologically arranged reminiscences, from childhood acquaintances to those who knew him in prison.)

Bonnet, Charles (1720-93)

In his youth, Bonnet made a meticulous and creative study of insects, which won him international fame for his discoveries, as well as his methods. He turned to psychology and offered a detailed, but speculative, account of the physiology of mental states. His empirical work was overtaken by speculative ambition. In later life, he developed (from elements already present in his early studies) a comprehensive view of the universe, of its history and its natural history, of theology and of moral philosophy. Christianity was proved, the great chain of being was mapped over time towards an ultimate perfection, and human morality, based on self-love, formed part of the Creator's scheme. The Creator, at the moment of creation, brought into being all the elements from which this vast unfolding would occur, without further intervention.

1 Life and works

Charles Bonnet, Swiss naturalist and philosopher, was born in Geneva and belonged to a family which had left France during the period of persecution of Protestants in the mid-sixteenth century. A youthful interest in natural history, and in insects in particular, led to his discovery, as a result of painstaking experimentation and observation, of the parthenogenesis of aphids. He was only 18 at the time of this important discovery, but received recognition from the Académie des Sciences in Paris to which his correspondent and mentor Réaumur presented his results in 1740. He then turned his attention to the regeneration of worms cut into parts, and to the respiratory mechanism of caterpillars. At the age of 23, he was elected fellow of the Royal Society in London, and his work *Traité d'insectologie (Treatise on Entomology)* was also published in 1745.

His vision deteriorated seriously, something which Bonnet attributed to overwork at the microscope. He did not at once abandon his empirical studies, but turned to the study of leaves. He investigated their movement and anatomy, maintaining that plants displayed a form of sensibility. Though lacking some of the essential theoretical concepts, he was in effect the first to investigate photosynthesis. He published his *Recherches sur l'usage des feuilles dans les plantes (Investigations into the Function of Leaves in Plants)* in 1754.

But his interests had already turned towards psychology and philosophy. His poor eyesight made it difficult for him to read or write, so that books were read to him. Leibniz made a special impression. His subsequent writings were dictated, including the *Essai de psychologie (Essay on Psychology)* (1754b), the *Essai analytique sur les facultés de l'âme (Analytical Essay on Mental Faculties)* (1760), the *Considérations sur les corps organisés (Considerations on Organic Bodies)* (1762-8), the *Contemplation de la nature (The Contemplation of Nature)* (1764), *La Palingénésie philosophique (Philosophical Palingenesis)* (1769), as well as a collected edition of his writings which he supervised, *Œuvres d'histoire naturelle et de philosophie* (1779-83).

2 Thought

Bonnet's thought has two organizing principles, potentially in conflict, both of which stemmed from his work as an entomologist. One is the importance of classification. The second is the idea that there are no gaps in nature. Between any two differently classified items, an intermediary item can be found.

In the *Essai de psychologie* and the *Essai analytique sur les facultés de l'âme*, Bonnet put forward a sensualist psychology, like, for instance, Hartley and Condillac. Sensations consisted in the vibration of nerve fibres in the brain. Once a 'virgin fibre' had first been caused to vibrate by an external stimulus, it could subsequently vibrate for internal reasons, giving rise to other types of mental phenomena. We pass smoothly from sensations to ideas. In developing this conception, Bonnet used the same model as Condillac: that of a statue successively presented with simple sensations of smell.

In spite of his speculative physiological account of sensation and other mental operations, Bonnet was not a materialist. On the contrary, he held that sensation required the activity of the soul, which, though linked to the body at a certain point in the nervous system, was an eternally existing 'germ'. (Bonnet did not have a fixed view about where this point was situated.)

Thus, Bonnet's psychology included an application of his general doctrine of 'preformation', already envisaged in his entomological work. He considered that the parthenogenesis of aphids demonstrated the existence of preformed

germs of organisms in the female germ cell. In his *Considérations sur les corps organisés*, he argued against epigenesis, the view that ontogenesis could be explained in some mechanical or developmental fashion, and maintained that from the time of creation, the universe contained a multitude of preformed ‘germs’, whose future development was already built into them. These ideas were brought together in an even more elaborate synthesis in the *La Palingénésie philosophique*. Here we see the ‘Great Chain of Being’ presented in dynamic form. We see an eschatological cosmology spread over time: the history of the earth consists of a series of cataclysms in which all organic life is destroyed, except the germs themselves. Over time, at each moment of rebirth, the germs produce a new and more perfect instantiation of what was already built into them at the moment of creation, until the final state is reached through the resurrection (or ‘palingenesis’) announced in the Christian gospel.

Bonnet also produced an apologetic work, *Recherches philosophiques sur les preuves du christianisme (Philosophical Investigations into the Proofs of Christianity)* (1770) in which, apart from reproducing some of the traditional proofs, he explained miracles, not as divine interventions in the natural order, but as apparently anomalous events which were already provided for at the moment of creation. His moral philosophy, developed in *Philalèthe* (before 1783), goes from the avoidance of pain and the search for pleasure, through self-love, but claims that this pre-utilitarian conception has its place in a divinely planned order. The ‘invisible hand’ of Adam Smith is the hand of God.

The grandiose system which Bonnet developed was widely influential in his day. But perhaps his more lasting philosophical impact arose from his detailed physiological theory of sensation, which influenced Cabanis and Maine de Biran.

See also: Evolution, theory of

F.C.T. MOORE

List of works

Bonnet, C. (1745) *Traité d’insectologie ou observations sur les pucerons (Treatise on Entomology; or Observations on Aphids)*, Paris: Durand.(Describes Bonnet’s famous discovery of the parthenogenesis of aphids.)

(1754a) *Recherches sur l’usage des feuilles dans les plantes (Investigations into the Functions of Leaves in Plants)*, Leiden: Luzac; repr. Hildesheim and New York: Olms, 1978.(Prefigures much later studies of photosynthesis.)

(1754b) *Essai de psychologie (Essay on Psychology)*, Leiden: Luzac; repr. Hildesheim and New York: Olms, 1978.(Attempts a materialist account of mental processes.)

(1760) *Essai analytique sur les facultés de l’âme (Analytical Essay on mental faculties)*, Copenhagen: C. & A. Philibert; repr. Hildesheim and New York: Olms, 1973.(Further development of associationist/materialist views of the mind, including the ’hypothesis of the statue’ which also occurs in Condillac.)

(1762) *Considérations sur les corps organisés: où l’on traite de leur origine, de leur développement, de leur reproduction etc., et où l’on a rassemblé en abrégé tout ce que l’histoire naturelle offre de plus certain et de plus intéressant sur ce sujet (Considerations on Organic Bodies; treating of their origin, development, reproduction etc., with a summary of the most certain and interesting findings of natural history on this subject)*, Amsterdam: Marc-Michel Rey.(The first of a series of works in which Bonnet argues that the epigenetic view of biological development, whch assumes that it can be explained mechanically, must be rejected in favour of ’preformation’.)

(1764) *Contemplation de la nature*, Amsterdam: Marc-Michel Rey; trans. anon. as *The Contemplation of Nature*, London: Longman, Becket, de Hondt, 1766.(Further development of ideas in Bonnet 1762.)

(1769) *La Palingénésie philosophique, ou Idées sur l’état passé et sur l’état futur des êtres vivants (Philosophical Palingensis; or Ideas on the Past and Future State of Living Beings)*, Amsterdam: Marc-Michel Rey.(A major, and somewhat mystical, synthesis of previous biological theories.)

(1770) *Recherches philosophiques sur les preuves du christianisme*, Geneva: C. Philibert & B. Chirol; trans. J.L. Boissier, *Philosophical and Critical Inquiries concerning Christianity*, London: Stockdale, 1787.(Reproduces some traditional theistic proofs, and adds the claim that miracles were prefigured.)

(1779-83) *Œuvres d’histoire naturelle et de philosophie (Works in Natural History and Philosophy)*, Neuchâtel: Fauche.(This edition additionally contains the following philosophical opuscula in volume 18: *Principes Philosophiques*; *Vue de leibnitzianisme*; *Hypothèse sur l’âme des bêtes et leur industrie*; *Philalèthe ou Essai d’une méthode pour établir quelques vérités de philosophie rationnelle*.)

(1788) *Précis de la doctrine de Kant sur l'entendement pur et remarques sur cette doctrine (Summary of Kant's Doctrine on Pure Understanding and Remarks on this Doctrine)* and *Remarques sur quelques endroits du livre de M. Kant intitulé 'Critique de l'entendement pur'*.(Comments on Kant.)

(1971) *Bonnet, Charles: Lettres à M. l'abbé Spallanzani*, ed. C. Castellani, Milan: Episteme.(Letters to Spallanzani.)

(1986) *Science against the Unbelievers: The Correspondence of Bonnet and Needham, 1760-1780*, ed. R. Mazzolini and S.A. Roe, Oxford: Voltaire Foundation at the Taylor Institution.(Correspondence.)

(1948) *Mémoires autobiographiques de Charles Bonnet de Genève (Autobiographical Memoirs of Charles Bonnet of Geneva)*, ed. R. Savioz, Paris: Vrin.

References and further reading

Anderson, L. (1982) *Charles Bonnet and the Order of the Known*, Dordrecht: Reidel.(A thorough and interesting survey of Bonnet's work, influenced by the approach of Foucault.)

Bonnet, G. (1929) *Charles Bonnet*, Paris: M. Lac.(A thorough study.)

Claparède, E. (1909) *La Psychologie animale de Charles Bonnet (The Animal Psychology of Charles Bonnet)*, Geneva: Georg.(A specialized study.)

Marx, J. (1976) 'Charles Bonnet contre les Lumières', *Studies on Voltaire and the Eighteenth Century*, Oxford: Voltaire Foundation at the Taylor Institute, 156-7.(A scholarly work which helps to place Bonnet in his intellectual environment, and which includes a thorough bibliography.)

Müller, G.H. and Pozzo, R. (eds) (1988) 'Bonnet, critico di Kant: due cahiers ginevrini del 1788', *Rivista di storia della filosofia*, 131-64.(Discussion of Kant's ideas.)

Offner, M. (1893) *Die Psychologie Charles Bonnets*, Leipzig.(Still a standard work.)

Rocci, G. (1975) *Charles Bonnet: filosofia e scienza*, Florence: Sansoni.(Should not be neglected.)

Roger, J. (1971) *Les Sciences de la vie dans la pensée française du XVIII siècle (Life Sciences in French Thought of the Eighteenth Century)*, Paris: A. Colin, 2nd edn.(A helpful historical study from a particular point of view.)

Rostand, J. (1945) *Esquisse d'une histoire de la biologie: Un préformationniste - Ch. Bonnet (Sketch of a History of Biology: A Prereformationist - C. Bonnet)*, Paris: Gallimard.(Important for those interested in the history and philosophy of biology, and Bonnet's relevance to it.)

Savioz, R. (1948) *La Philosophie de Charles Bonnet de Genève (The Philosophy of Charles Bonnet of Geneva)*, Paris: Vrin.(A helpful review.)

Boole, George (1815-64)

George Boole, a British mathematician, is credited with making a fundamental contribution to modern logic. If Leibniz's manuscript essays on logic, effectively unknown until the end of the nineteenth century, are excluded, thenBoole's algebra of logic (1847, 1854) was the first successful mathematical treatment of one part of logic. The treatment was mathematical in the broad sense of using a formal language expressed in symbols with definite rules. It was also mathematical in a narrow sense of being closely modelled after numerical algebra, from which it differed by an additional axiom, $x^2 = x$. *Letter symbols of this algebra were conceived as representing classes, 1 standing for a 'universe' of objects and 0 for the empty class. By identifying logical terms with their extensions, that is, with classes, inferences of a much more general character than those of the traditional syllogistic could be carried out. Boole also showed how this algebra could be used in propositional logic, presenting its earliest systematic general formulation.*

1 Beginnings

Boole, eldest child of a Lincolnshire shoemaker, had no formal secondary or higher education, though he was recognized early on as a gifted child. He taught himself Latin, Greek and the major European languages, reading widely in many subjects. At the age of 16, impelled by his family's financial hardship, he became assistant to a schoolmaster. At that time he also began the serious study of higher mathematics by reading works of the masters. At the age of 19 he opened his own school. While still teaching he managed to do mathematical research, attaining considerable prominence when he was awarded a Royal Society gold medal for a paper printed in the Society's *Philosophical Transactions*. In 1849, although without a university degree, he was appointed Professor of Mathematics at the newly opened Queen's College in Cork (now University College, Cork).

Boole's innovative ideas first appeared in a pamphlet, *The Mathematical Analysis of Logic* (1847). Boole was evidently a newcomer to the subject, as his exposition of the traditional syllogism paraphrased that contained in Richard Whately's popular treatise *Elements of Logic* (1844). Whately's conception of the subject as a formal, abstract discipline marked a sharp break with the view, prevalent through the eighteenth century, of logic as a language art, or as an epistemological tool in the search for truth. Moreover, Whately's explanations of logical principles were couched extensionally, logical terms being taken as representing classes of objects, not intensionally as representing combinations of attributes. To this extensional setting Boole adjoined a mathematical feature, adapting notions from operator algebra for this purpose. With each class (term) *X* he associated a selection operator, *x*, which selects from a universe of objects those which are *X*. In operator algebra, successive operation is indicated by 'multiplication' (of operators). Hence *yx* selects the *Y* from what *x* selects, that is, from the *X*. As selection in either order results in the same class, he has the 'law',

$$xy = yx$$

Similarly, since a repeated selection of the same class results in nothing new, he has

$$xx \text{ (or } x^2) = x$$

Propositions are represented by equations. The universal affirmative 'All *X* are *Y*', for example, is represented by '$x = xy$' (selection of the *X*s from the *Y*s yields just the *X*s if and only if all *X* are *Y*). The universal negative 'No *X* are *Y*' is represented by '$xy = 0$' (0 representing 'nothing'). Inferences are performed by the substitution and replacement of equals. Thus from '$x = xy$' and '$y = yz$', on replacing *y* in the first equation by *yz* and then the combination *xy* by *x*, one obtains

$$x = x(yz) = (xy)z = xz$$

justifying the validity of the AAA syllogism.

2 *The Laws of Thought*

These few items suffice to illustrate the nature of Boole's initial venture. Seven years later it was supplanted by his more carefully constructed and elaborated *The Laws of Thought* (1854). In a significant change from the earlier work, the symbols $x, y, z, \ldots$ are no longer operations selecting classes but stand directly for the classes themselves. The common understanding of the term class is extended so as to include 'universe' and 'nothing',

these being denoted by 1 and 0 respectively. Unlike '*xy*', defined for any two classes, the combinations '$x+y$' and '$x-y$' are only partially defined: '$x+y$' has meaning only if *x* and *y* have nothing in common, and '$x-y$' has meaning only if *y* is part of *x*. Citing examples from ordinary language for justification, he states a number of laws which hold for classes, such as the two mentioned above, and others such as '$z(x+y)=zx+zy$', for which he gives as an example 'European men and women' is the same as 'European men and European women'.

He contends that for purposes of logical deduction the only verb needed is 'is', or 'are', symbolically expressed by '$=$'. Logical deductions are accomplished by algebraic operations on equations, such as substituting equals for equals. What is new here is a more sophisticated justification for the use of algebra in doing logic, one based on the idea (at that time quite novel) of an algebra different from standard numerical algebra. Commenting on the analogy of his laws of class symbols with those of 'Number', all but '$x^2=x$' being true of numbers, and that also if *x* is restricted to being either 0 or 1, he says:

> Let us conceive, then, of an Algebra in which the symbols $x, y, z, \ldots$ admit indifferently of the values 0 and 1, and of these values alone. The laws, the axioms, and the processes, of such an Algebra will be identical in their whole extent with the laws, the axioms, and the processes of an Algebra of Logic. Difference of interpretation will alone divide them.
>
> (1854: 37-8)

With the adoption of this principle Boole vastly extended the range and applicability of traditional term logic. For - as in numerical algebra - terms of any size or complexity are constructible, resulting in new inference forms unlimited in number. Moreover, as in algebra, transformations on such complex terms can be carried out mechanically by rule, that is, 'computationally'.

But there were difficulties associated with the principle. According to it one may freely use algebraic expressions constructible by use of the binary operators $+$, $-$, $\times$, the (class) variables $x, y, z, \ldots$ and the constants 0 and 1. But then Boole had the problem of what to do with expressions such as '$1+1$' and '$0-1$' to which no meaning was attached. He argued that if the initial algebraic formulas were interpretable in terms of class notions, and if the end result was also so interpretable then the result was valid even if some of the intermediary steps were not. He likened this to the use of the imaginary $\sqrt{-1}$ in trigonometry. With considerable ingenuity Boole introduced procedures which enabled him to obtain correct results.

All this ingenuity became needless when Boole's partially defined addition was replaced by one based on non-exclusive 'or', (as suggested by W.S. Jevons (1864), C.S. Peirce (1867)), so that '$x+x=x$', resulting in what is now called 'Boolean Algebra'. An alternative simplification appeared in the twentieth century. This produced an algebra ('Boolean ring with unit') even closer to Boole's conceptions in having a meaningful subtraction. Here addition is taken to be 'symmetric difference' - what Boole would have written as '$x(1-y)+y(1-x)$' - resulting in the law '$x+x=0$' for this interpretation of addition.

The same algebra that Boole used for the logic of terms he also employed for propositional logic. Now, instead of $x, y, z, \ldots$ standing for classes $X, Y, Z, \ldots$ they stand for 'portions of time' for which propositions $X, Y, Z, \ldots$ are true. Then, for example, '$x=1$' stands for *X* being true (for all time), and '$y=0$' for *Y* not true and '$x(1-y)=0$' for 'If *X* is true, then *Y* is true' (since for no time is *X* true and *Y* not true). In this application, without explicitly realizing it, Boole was adjoining '$x=1$ or $x=0$', a property not necessarily holding for classes and, in effect, reducing his algebra to a special kind of algebra in which there are only two classes.

It will be no denigration of Boole's remarkable achievement in logic if its inadequacies are cited. As already noted, his choice of an addition, incompletely defined for classes, required excessively complicated methods to recover logical content. On the other hand, some of his devices (for example, expansion of a logical function into normal form) became valuable techniques. His treatment of the particular proposition (those involving 'some') was defective, as would be expected since his algebra had no formal means of expressing non-emptiness of a class. He thought his propositional logic had to be based on his logic of terms via the notion of 'portions of time'. Currently the logic of terms is treated as part of the predicate calculus, the development of which *presupposes* propositional logic. Finally, despite Boole's belief that his treatment was fully adequate for all of logic, it did not have the ability to express quantifiers.

See also: Boolean algebra; Logic in the 19th century §3

THEODORE HAILPERIN

List of works

Boole, G. (1847) *The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning*, Cambridge: Macmillan, Barclay & Macmillan; London: George Bell; repr. Oxford, Blackwell, 1948, and in Boole, 1952, 49-124.(Written in haste, Boole wished it to be replaced by his more maturely considered *Laws of Thought*.)

(1854) *An Investigation of the Laws of Thought, on Which are Founded the Mathematical Theories of Logic and Probabilities*, London: Walton & Maberly; repr. as *George Boole's Collected Logical Works*, Chicago, IL, and New York: Open Court, 1916, vol. 2; repr. New York: Dover, 1951.(This work is also notable for its contribution to the origins of modern algebra and, less so, for its extensive discussions, from an unconventional viewpoint, of probability matters.)

(1952) *Studies in Logic and Probability*, ed. R. Rhees, London: Watts & Co. (The most extensive commentary on Boole's ideas on logic and probability. Somewhat mathematical in parts, but much of it accessible to the general reader.)

References and further reading

Hailperin, T. (1976) *Boole's Logic and Probability: A Critical Exposition from the Standpoint of Contemporary Algebra, Logic and Probability Theory*, Amsterdam: North Holland; 2nd edn, 1986.(The second edition has been revised and enlarged, and contains an extensive bibliography.)

Jevons, W.S. (1864) *Pure Logic or the Logic of Quality apart from Quantity. With Remarks on Boole's System and on the Relation of Logic and Mathematics*, London: E. Stanford.(Of significance only for its introduction of the non-exclusive 'or'.)

MacHale, D. (1985) *George Boole: His Life and Work*, Dublin: Boole Press.(A highly recommended biography. Contains an extended bibliography.)

Peirce, C.S. (1867) 'On an Improvement in Boole's Calculus of Logic', *Proceedings of the American Academy of Arts and Sciences* 7: 250-61; repr. in *Collected Papers*, Cambridge, MA: Harvard University Press, vol. 3, 1933; also in *Writings of Charles S. Peirce*, Bloomington, IN: Indiana University Press, 1984, vol. 2.(Peirce's first paper on logic, the beginning of a long sequence of his distinguished contributions to the subject.)

Whately, R. (1844) *Elements of Logic: Comprising the Substance of the Article in the Encyclopedia Metropolitana: with additions, etc.*, 8th edn, revised, London: B. Fellowes.(This work, dating from 1827, appeared in many printings, editions, and by various publishers.)

Boolean algebra

Boolean algebra, or the algebra of logic, was devised by the English mathematician George Boole (1815-64) and embodies the first successful application of algebraic methods to logic.

Boole seems to have had several interpretations for his system in mind. In his earlier work he thinks of each of the basic symbols of his 'algebra' as standing for the mental operation of selecting just the objects possessing some given attribute or included in some given class; later he conceives of these symbols as standing for the attributes or classes themselves. In each of these interpretations the basic symbols are conceived as being capable of combination under certain operations: 'multiplication', corresponding to conjunction of attributes or intersection of classes; 'addition', corresponding to (exclusive) disjunction or (disjoint) union; and 'subtraction', corresponding to 'excepting' or difference. He also recognizes that the algebraic laws he proposes are satisfied if the basic symbols are interpreted as taking just the number values 0 and 1.

Boole's ideas have since undergone extensive development, and the resulting concept of Boolean algebra now plays a central role in mathematical logic, probability theory and computer design.

1 Basic facts

Boolean algebra, devised by the English mathematician George Boole, embodies the first successful application of algebraic methods to logic (see Boole 1847, 1854).

The algebraic structures implicit in Boole's analysis were first explicitly presented by Huntington (1904), and termed 'Boolean algebras' by Sheffer (1913). As Huntington recognized, there are various equivalent ways of characterizing Boolean algebras. One of the most convenient definitions is as follows. A 'Boolean algebra' is a structure $(B, +_B, *_B, -_B, 0_B, 1_B)$, with *B* a non-empty set, $+_B$ and $*_B$ binary operations on *B*, $-_B$ a unary operation on *B*, and 0_B, 1_B distinct elements of *B* such that the following are true for all x, y, z in *B*.

$$x + (y + z) = (x + y) + z \qquad x * (y * z) = (x * y) * z \qquad \text{(associativity)}$$

$$x + y = y + x \qquad x * y = y * x \qquad \text{(commutativity)}$$

$$x + (x * y) = x \qquad x * (x + y) = x \qquad \text{(absorption)}$$

$$x + (y * z) = (x + y) * (x + z) \qquad x * (y + z) = (x * y) + (x * z) \qquad \text{(distributivity)}$$

$$x + (-x) = 1 \qquad x * (-x) = 0 \qquad \text{(complementation)}$$

(We omit the subscript when confusion is unlikely.) The operations 'Čplus;' (Boolean 'addition' - here corresponding to *inclusive* disjunction), 'Čast;' (Boolean 'multiplication' - intersection) and '$-$' (taking the complement) are called Boolean operations. Note that the operation sending x, y to $x * (-y)$ corresponds to Boole's 'subtraction'. The elements 0 and 1 are called the 'zero element' and 'unit element' of *B*, respectively. We shall usually identify somewhat loosely a Boolean algebra by its underlying set; thus, for example, the Boolean algebra just introduced will be denoted simply by '*B*'.

The following basic identities can be readily shown to hold in any Boolean algebra.

$$-(x + y) = (-x) * (-y) \qquad \text{(De Morgan's laws)}$$
$$-(x * y) = (-x) + (-y)$$
$$-(-x) = x \qquad \text{(double complement law)}$$

We observe that in the axioms characterizing Boolean algebras the roles played by + and *, and by 0 and 1, are entirely symmetrical. This observation leads to the important 'principle of duality' for Boolean algebras which

may be formulated as follows. Let *P* be any statement about Boolean algebras which involves just the Boolean operations +, * and $-$ and the elements 0, 1. The 'dual' of *P* is obtained from *P* by interchanging + with * and 0 with 1. If *P* holds in all Boolean algebras, then so does the dual of *P*. Thus, for example, once we have established one of De Morgan's laws above, the principle of duality leads automatically to the other one.

We define a 'subalgebra' of a Boolean algebra *B* to be a non-empty subset of *B* which is closed under the Boolean operations in *B*.

2 Important types of Boolean algebra

Boole seems to have had several interpretations for his algebraic system in mind. In his original pamphlet (1847) he thinks of each of the basic symbols of his 'algebra' as standing for the mental operation of selecting just the objects possessing some given attribute or included in some given class. Later (1854) he conceives of these symbols as standing for the attributes or classes themselves.

Two standard types of Boolean algebra arise in set theory and logic, corresponding to these two interpretations. To obtain the first of these, consider any non-empty set *X* and let $\wp(X)$ be its power set, that is, the set of all its subsets. Then the structure $(\wp(X), \cup, \cap, -, \emptyset, X)$ is a Boolean algebra - the power set algebra of *X* - in which $\cup, \cap, -$ are the operations of set-theoretic union, intersection and complementation with respect to *X*, respectively.

A subalgebra of a power set algebra is called an algebra of sets. As an example, for any set *X*, let $Z(X)$ be the set of all finite and cofinite subsets of *X* (a 'cofinite' set being the complement of a finite set). Then $(Z(X), \cup, \cap, -, \emptyset, X)$ is an algebra of sets called the finite-cofinite algebra of *X*.

When *X* is a singleton $\{a\}$, $\wp(X)$ reduces to the set $\{0, 1\}$, with $0 = \emptyset$ and $1 = X$. This algebra is called the 'two-element Boolean algebra' and is denoted by '2'. Its operations are displayed in the table below.

x	y	$x+y$	$x*y$	$-x$
0	0	0	0	1
0	1	1	0	1
1	0	1	0	0
1	1	1	1	0

The standard types of Boolean algebra arising in logic are the so-called Lindenbaum-Tarski algebras. To obtain these, we start with a theory *T* in a propositional or first-order language *L*. Define the equivalence relation $\approx$ on the set of formulas of *L* by $\phi \approx \psi$ if $T \vdash \phi \leftrightarrow \psi$. For each formula ϕ let $[\phi]$ be its $\approx$-equivalence class. On the set $A(T)$ of such equivalence classes define the operations +, *, $-$ and the elements 1, 0 by:

$$\begin{aligned} [\phi] + [\psi] &= [\phi \vee \psi] \\ [\phi] * [\psi] &= [\phi \wedge \psi] \\ -[\phi] &= [\neg\phi] \\ 1 &= [\alpha_0 \vee \neg\alpha_0] \\ 0 &= [\alpha_0 \wedge \neg\alpha_0], \end{aligned}$$

where α_0 is a fixed but arbitrary formula. Then the structure $(A(T), +, *, -, 0, 1)$ is a Boolean algebra, called the Lindenbaum-Tarski algebra of *T*. The Boolean algebra $A(\emptyset)$ (which depends solely on the language *L*) is denoted by $Alg(L)$ and is called the Lindenbaum-Tarski algebra of *L*.

Algebras of sets and Lindenbaum-Tarski algebras are typical Boolean algebras in the following sense. Call two Boolean algebras $(B, +_B, *_B, -_B, 0_B, 1_B)$ and $(C, +_C, *_C, -_C, 0_C, 1_C)$ 'isomorphic' if there is a one-one, onto map (that is, an 'isomorphism') $f\colon\ B \to C$ such that $f(0_B) = 0_C$, $f(1_B) = 1_C$ and, for any x, y in *B*,

$$\begin{aligned} f(x +_B y) &= f(x) +_C f(y) \\ f(x *_B y) &= f(x) *_C f(y) \\ f(-_B x) &= -_C f(x). \end{aligned}$$

Then it can be shown that any Boolean algebra is isomorphic both to an algebra of sets and to the

Lindenbaum-Tarski algebra of some propositional theory. (Isomorphic Boolean algebras are structurally indistinguishable.) These facts (the first of which is the famous 'Stone representation theorem' of 1936) together show that Boolean algebras embody just the common structural features of set theory and logic.

Since, as is readily shown, there are Boolean algebras (for example, the finite-cofinite algebra of the set of natural numbers) which are *not* isomorphic to any power set algebra, nor to the Lindenbaum- Tarski algebra of a propositional language, it is of interest to try to determine the additional characteristic features of these two latter types of algebra.

For power set algebras, we need the following concepts. Given a Boolean algebra *B*, we define the relation $\leq$ on *B* by $x \leq y$ if and only if $x * y = x$. It is easily shown that $\leq$ is reflexive ($x \leq x$ for all *x*), transitive (if $x \leq y$ and $y \leq z$ then $x \leq z$) and antisymmetric (if $x \leq y$ and $y \leq x$ then $y = x$); in other words, $\leq$ is a 'partial ordering' on *B*. *B* is then said to be 'complete' if every subset *X* of *B* has both a greatest lower bound and a least upper bound with respect to $\leq$. (It is not hard to see that a *finite* subset $\{x_1, \ldots, x_n\}$ of any Boolean algebra has greatest lower bound $x_1 * \ldots * x_n$ and least upper bound $x_1 + \ldots + x_n$.) An element $a \neq 0$ of a Boolean algebra *B* is called an 'atom' if, for any *x* in *B*, if $x \leq a$, then $x = 0$ or $x = a$. *B* is said to be 'atomic' if for any $x \neq 0$ in *B* there is an atom *a* such that $a \leq x$. Then we have the fundamental result of Lindenbaum and Tarski (1935) characterizing power set algebras: a Boolean algebra is isomorphic to a power set algebra if and only if it is complete and atomic (see Halmos 1963: §16, theorem 5).

Turning to Lindenbaum-Tarski algebras of propositional languages, we define a subset *X* of a Boolean algebra *B* to be 'free' if for any finite subset $\{x_1, \ldots, x_n\}$ of *X* we have $y_1 * \ldots * y_n \neq 0$, where each y_i is either x_i or $-x_i$. *X* is said to 'generate' *B* if the only subalgebra of *B* containing *X* is *B* itself, that is, if every element *b* of *B* can be expressed in the form $b = y_1 + \ldots + y_n$, where each y_i is of the form $z_1 * \ldots * z_m$ with either z_i or $-z_i$ in *X*. Finally *X* is said to be 'freely generated' if it has a free set of generators. It can then be shown that, for any propositional language *L*, the algebra $Alg(L)$ is freely generated (with the equivalence classes of the proposition letters constituting a free set of generators) and, moreover, that any (infinite) freely generated Boolean algebra is isomorphic to some $Alg(L)$.

3 Boolean algebras as algebras of truth-values

The simplest Boolean algebra is the algebra $2 = \{0, 1\}$, which it is customary to think of as the algebra of two 'truth-values'; 0 standing for False and 1 for True. More generally, we may conceive of an arbitrary Boolean algebra as an algebra of 'generalized truth-values' - that is, containing elements corresponding to 'truth-values' different from True and False. In that case, it becomes natural to extend the usual idea of truth-valuations of (propositional) formulas (that is, in the algebra 2) to valuations in arbitrary Boolean algebras. Thus we define a 'valuation' of a propositional language *L* in a Boolean algebra *B* to be a map *v* from the set of formulas of *L* to *B* such that, for arbitrary formulas ϕ, ψ we have

$$v(\phi \vee \psi) = v(\phi) + v(\psi)$$
$$v(\phi \wedge \psi) = v(\phi) * v(\psi)$$
$$v(\neg\phi) = -v(\phi).$$

It can then be shown that, for any formula ϕ, the following assertions are equivalent:

(1) ϕ is a propositional theorem.
(2) $v(\phi) = 1$ for every valuation *v* in 2.
(3) $v(\phi) = 1$ for any valuation *v* in any Boolean algebra.

(The equivalence between (1) and (2) is the completeness theorem for propositional logic.) A similar result may be established for predicate logic (extending the relevant completeness theorem), this time using the idea of a valuation in a complete Boolean algebra to enable quantified formulas to receive 'truth-values'.

The idea of using Boolean algebras as algebras of truth-values scored what is regarded by logicians as its most spectacular success in the 1960s when it was shown by Robert Solovay and Dana Scott, and independently by Petr Vopenka, that Paul Cohen's celebrated proofs of independence in set theory could be formulated in terms of what has come to be known as the method of 'Boolean-valued models' (see Bell 1985). Briefly, the method runs as

follows. Let *L* be the language of Zermelo-Fraenkel set theory (ZF; see Set theory, different systems of), and suppose that σ is a sentence of *L* whose formal independence of ZF we wish to establish - the continuum hypothesis, for example (see Continuum hypothesis). A complete Boolean algebra *B* is chosen and a class $V^{(B)}$ of sets - the Boolean universe induced by *B* - constructed. Let $L^{(B)}$ be the language obtained by adding to *L* a name for each element of $V^{(B)}$. Now one can construct a 'valuation' *v* of the (sentences of) the language $L^{(B)}$ in *B*: for each sentence ϕ of $L^{(B)}$, $v(\phi)$ is the element of *B* representing the 'Boolean truth-value' of ϕ in $V^{(B)}$. This map *v* is defined so as to map all the theorems of ZF to the unit element 1 of *B*: accordingly the 'structure' $(V^{(B)}, v)$ is called a Boolean-valued model of set theory. If the Boolean algebra *B* is chosen with finesse, it is possible to arrange things so that $v(\sigma) = 0_B$. That is, for a suitable choice of *B*, $V^{(B)}$ is a Boolean-valued model of set theory in which the given statement σ is false. This amounts to a proof of the independence of σ from the axioms of set theory. It follows that the problem of establishing the independence of a particular set-theoretic statement boils down in principle to demonstrating the existence of a complete Boolean algebra with specified properties. This method has been exploited with great success in recent years.

See also: Boole, G.; Logical and mathematical terms, glossary of

J.L. BELL

References and further reading

Bell, J.L. (1985) *Boolean Valued Models and Independence Proofs in Set Theory*, Oxford Logic Guides 12, Oxford: Clarendon Press.(A detailed technical account of the subject.)

Boole, G. (1847) *The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning*, London: G. Bell & Sons; repr. in *Studies in Logic and Probability*, London: Watts, and LaSalle, IL: Open Court, 1952; also repr. in *An Investigation of the Laws of Thought*, New York: Dover, 1973.(The first appearance of Boole's system.)

Boole, G. (1854) *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities*, London: Walton & Maberley; repr. in *An Investigation of the Laws of Thought*, New York: Dover, 1973.(Further elaboration.)

Hailperin, T. (1976) *Boole's Logic and Probability*, Amsterdam: North Holland; 2nd edn, 1986.(An illuminating critical exposition of Boole's achievement from a modern standpoint.)

Halmos, P.R. (1963) *Lectures on Boolean Algebras*, Princeton, NJ: Van Nostrand.(An attractively written exposition of the basic theory of Boolean algebras.)

Huntington, E.V. (1904) 'Sets of Independent Postulates for the Algebra of Logic', *Transactions of the American Mathematical Society* 5: 288-309.(The first explicit postulate systems for Boolean algebra.)

Monk, J.D. (ed.) (1989) *Handbook of Boolean Algebras*, Amsterdam: Elsevier Science, 3 vols.(This set includes a series of articles on all mathematical aspects of Boolean algebras. Volume 1 is an excellent introduction to the subject.)

Rasiowa, H. and Sikorski, R. (1963) *The Mathematics of Metamathematics*, Warsaw: PWN.(A systematic technical account of the algebraic approach to logic.)

Sheffer, H.M. (1913) 'A Set of Five Independent Postulates for Boolean Algebras, With Application to Logical Constants', *Transactions of the American Mathematical Society* 14: 481-8.(Here it is shown that Boolean algebras can be axiomatized in terms of the single primitive operation corresponding to 'not both' - a fact apparently already known to Peirce.)

Bosanquet, Bernard (1848-1923)

One of the most prominent and prolific of the British Idealists of the late nineteenth and early twentieth centuries, Bosanquet ranged across most fields of philosophy, making his main contributions in epistemology, metaphysics, aesthetics and especially political philosophy. He was deeply influenced by Plato and byHegel. Bosanquet and F.H. Bradley were close on many matters, and each regarded the other as a co-worker; however, Bosanquet was always more Hegelian, less rigorous in argument than Bradley and lacking his sceptical approach. Bosanquet treats knowledge and reality as a single whole, working out the implications in the concrete 'modes of experience' of philosophy, science, morality, art, religion, and social and political life. He is at his best in explaining and developing the thoughts of others, particularly ofHegel, Bradley, Rousseau and T.H. Green.

Bernard Bosanquet was educated at Harrow and Balliol College, Oxford. He was much influenced by Green's philosophy, and by his example of active citizenship (see Green, T.H. §3). Elected a fellow of University College in 1870 (beating F.H. Bradley), Bosanquet taught philosophy and Greek history. He resigned in 1881 in order to have time to write and to engage in social work with the Charity Organisation Society in London; he was also a frequent adult education lecturer and organizer. He was professor of moral philosophy at St Andrews University in the years 1903-8.

Bosanquet propounds a version of absolute or objective idealism derived principally from Hegel. The key to Bosanquet's general metaphysics is 'the whole'. Our knowledge is always of particular parts of experience conceived in a particular way (scientific, aesthetic, religious, and so on); hence our judgments may conflict (for instance, the findings of science and religion). But what we experience as parts are in reality interrelated and constitute a whole, which reveals its different aspects in the different concrete forms of our experience. So all our knowledge is partial, inconsistent, incomplete and not fully true; we strive to unify it. Complete knowledge, or truth, would be comprehensive and systematic knowledge of the all-inclusive and self-consistent and coherent whole (or Absolute), and thus of every part in all its relations to every other part. The Absolute is not transcendent, in 'another' world, but immanent here in ours. Our knowledge rests on experience of appearances of reality.

Correspondingly, individual parts of the whole, for example, an animal, a work of art, or a person, are not completely real; only the whole of which they are part is real. Ultimately there is but one true, real, individual: namely, the whole. It follows that persons, for example, are 'finite individuals', incomplete and not as real as the communities to which they belong (which in turn are not fully real, relative to the whole). Reality and value are not found to the full in individual persons. Bosanquet drew the controversial conclusions, for example, that the traditional belief in an immortal soul peculiar to each person is groundless, a symptom of our mistaken religious individualism. He is not denying the separateness and uniqueness of persons: each is a particular 'modification' of the social whole. But each becomes more real by drawing on the ideas and institutions of their society. True individuality is judged not by the atomic and exclusive self, nor by what persons are, but by what they can become, and that is shown in their society and its cultural achievements.

Bosanquet's approach shows to advantage in his political philosophy. It arms him against the atomic individualism he finds in the 'theories of the first look' of Bentham, Spencer andJ.S. Mill. Instead, combining Rousseau andHegel, Bosanquet argues that the individual members of a state are linked together by their common ideas and through their participation in its institutions. Taken as a whole, they each have a 'general will', aiming at the common good, which is more real than their particular wills; and their freedom lies in acting according to the general will, as expressed in the law and practices of the state. Bosanquet also takes account of the various levels and kinds of groups (finite wholes) in a state: families, professions, trade unions, churches, local communities, for instance. He shows how these interact, how they express different and sometimes conflicting interests, and how the sovereign state harmonizes their demands. He is thus able, in later editions of *The Philosophical Theory of the State*, to incorporate the arguments of the Pluralists. He explains how change is effected in the state and in its component parts as ideas clash and are reconciled through a 'social logic'. This analysis works particularly well in the case of a parliamentary democracy. Bosanquet is no conservative. Because the Absolute is immanent, improvement is not postponable to another world: 'here or nowhere is your America', 'the Kingdom of God is on earth'. Not everything is as it should be; and the state should sometimes be used confidently to widen persons'

opportunities. The state is justified in restricting its members' narrow freedom from interference because its laws are in the interest of all, and bring greater freedom, the freedom to develop oneself. The state is more real than the person, the whole in which the person can enrich their life (but Bosanquet in no way considered the value of persons to be instrumental to the state). The state too is finite, in relation to other states. Bosanquet did not think the political conditions for international federations or world government existed yet, but there is space for it in his theory. The state is also finite relative to the deeper experience of religion, art and philosophy, which enable persons to enlarge their individuality.

Bosanquet suffered severely from Hobhouse's not altogether fair but very effective attacks on him as being politically authoritarian, confusing the state with society and idealizing actual states. Generally Bosanquet's style of philosophizing became unfashionable: Russell, for example, found it vague, loose, shallow, evasive, dogmatic and claiming more for philosophy than it could deliver. For most of the twentieth century Bosanquet has been ignored. It remains to be seen whether his ties to Hegel will benefit him as Hegel's reputation rises again.

See also: Absolute, the; Hegelianism §5; State, the

PETER P. NICHOLSON

List of works

Bosanquet, B. (1888) *Logic, or the Morphology of Knowledge*, Oxford: Clarendon Press, 2nd edn, 1911.(The systematic and fullest statement of his theory of truth.)

(1892) *A History of Aesthetic*, London: Swan Sonnenschein.(The aesthetic consciousness traced from the Greeks to the nineteenth century.)

(1899) *The Philosophical Theory of the State*, London & New York: Macmillan, 4th edn, 1923.(Bosanquet's major contribution to political philosophy.)

(1912) *The Principle of Individuality and Value; The Gifford Lectures for 1911*, London: Macmillan.(Outlines Bosanquet's metaphysics.)

(1913) *The Value and Destiny of the Individual; The Gifford Lectures for 1912*, London: Macmillan.(Further details of Bosanquet's metaphysics.)

(1924) 'Life and Philosophy', in J.H. Muirhead (ed.) *Contemporary British Philosophy, Personal Statements, First Series*, London: Allen & Unwin, 51-74. (Bosanquet's philosophical autobiography, stating the main principles of his philosophy in its typical language.)

Muirhead, J.H. and R.C. Bosanquet (eds) (1927) *Science and Philosophy and Other Essays by the Late Bernard Bosanquet*, London: Allen & Unwin.(Good selection of previously published papers in logic, metaphysics, ethics, politics and aesthetics.)

Muirhead, J.H. (ed.) (1935) *Bernard Bosanquet and His Friends. Letters Illustrating the Sources and the Development of his Philosophical Opinions*, London: Allen & Unwin.(Correspondence with other philosophers, particularly in his later years.)

References and further reading

Bosanquet, H. (1924) *Bernard Bosanquet. A Short Account of his Life*, London: Macmillan.(By his wife; the main biographical source.)

Gaus, G.F. (1994) 'T.H. Green, Bernard Bosanquet and the Philosophy of Coherence', in C.L. Ten (ed.) *The Routledge History of Philosophy, vol. 7: The Nineteenth Century*.(Treats complex issues of how Bosanquet's philosophy connects with his politics; he is helpfully compared with Green.)

Hobhouse, L.T. (1918) *The Metaphysical Theory of the State*, London: Allen & Unwin.(Famous attack on Bosanquet's political philosophy.)

Hoernlé, R.F.A. (1927) *Idealism as a Philosophy*, New York: George H. Doran, chaps 8-10.(Very clear and sympathetic account of the basics of Bosanquet's metaphysics and political philosophy.)

Lang, B. (1968) 'Bosanquet's Aesthetic: A History and Philosophy of the Symbol', *Journal of Aesthetics and Art Criticism* 26: 377-87.(Overview of Bosanquet's philosophy of art.)

Nicholson, P.P. (1978) 'A Bibliography of the Writings of Bernard Bosanquet (1848-1923)', *Idealistic Studies* 8: 261-80.(Full listing of the published and unpublished writings.)

(1990) *The Political Philosophy of the British Idealists: Selected Studies*, Cambridge: Cambridge University Press, Study 6.(Examines Bosanquet on the General Will. Extensive bibliography.)

Primoratz, I. (1994) 'The Word "Liberty" on the Chains of Galley-Slaves: Bosanquet's Theory of the General Will', in *History of Political Thought* 15: 249-67.(Argues that the theory fails to solve the problem of political

obligation.)

Russell, B. (1992) Reviews of Bosanquet's Gifford Lectures, in *Collected Papers, vol. 6: Logical and Philosophical Papers 1909-13*, ed. J.G. Slater, London & New York: Routledge, 368-70, 372-3. (ChallengesBosanquet's Hegelian conception of philosophy.)

Sweet, W. (1996) *Idealism and Rights: The Social Ontology of Rights in the Political Thought of Bernard Bosanquet*, Lanham, MD: University Press of America.(Clear, careful and detailed appraisal, basically sympathetic, of the main features of Bosanquet's political philosophy. Fullest bibliography.)

Bourdieu, Pierre (1930-)

Critically assessing both hermeneutic and structuralist approaches, Pierre Bourdieu's social theory aims at transcending the opposition between the individual and society. On the one hand, people exhibit practical skills which are adjusted to the constraints of the environment. On the other hand, society does not determine people's actions: the very same practical skills allow them to improvise and deal with an infinite number of situations. Although Bourdieu takes into account the individual, he does not succumb to the Cartesian notion of a self-sufficient subject. Also, his view is very much in opposition to rational choice theory. His theoretical framework has emerged out of his empirical research and vice versa. In his research Bourdieu applies his reflexive sociology: a critical reflection on the part of the social scientist towards their own practices.

Initially educated at the École normale supérieure and eventually Professor of Sociology at the Collège de France, Bourdieu is known for his contributions to the theory of society. The label 'social theorist' is, however, not entirely appropriate. First, his writings cover various fields in the social sciences, ranging from the abstract to the ethnographic. Second, rather than providing a theory that is independent of empirical research, Bourdieu aims at a method derived from and directed towards social research. Rather than embarking upon a grand theory, Bourdieu invites the reader to ask certain questions.

Pierre Bourdieu spent most of his academic life in France, apart from a short spell in Algeria. It is thus not surprising that many of his writings deal with French society. Neither is it far-fetched to read some of his writings as a dialogue with Sartre and Lévi-Strauss - the two leading French intellectuals at the time of Bourdieu's formative academic years (see Sartre, J.-P.; Lévi-Strauss, C.). Always sceptical of existentialist philosophy, Bourdieu was initially attracted to structuralism but soon also distanced himself from the latter (see Structuralism in social science §5). Where existentialism lacks objectivity, structuralism erroneously dismisses the relevance of hermeneutics (see Hermeneutics). Related to his critique of both philosophies, Bourdieu's project has been to transcend the opposition between the individual and society. However, Bourdieu's attention to the agent does not mean that he succumbs to a Cartesian notion of a self-sufficient subject; instead Bourdieu adopts some of the phenomenological and Wittgensteinian insights *vis-à-vis* shared, tacit knowledge (see Phenomenological movement; Phenomenology, epistemic issues in; Wittgenstein, L.). Likewise, his focus upon social structure does not entail the 'economism' of French Marxism. Indeed, Bourdieu pays much attention to power struggles in society, not merely at the economic level, but also in the symbolic sphere (see Marxist philosophy of science).

In spite of his reluctance to build a consistent theoretical system, Bourdieu's writings exhibit a particular philosophical outlook. He describes his own work as an attempt to transcend the alleged dualism between what he calls 'objectivism' and 'subjectivism'. Characteristic of structuralist perspectives, objectivism searches for underlying structures independent of people's knowledge or strategies. Central to humanist approaches, subjectivism attributes a pivotal role to people's experiences, strategies and improvisations. Focusing on one side only, each perspective inevitably distorts reality. Subjectivist accounts tend to conceive of social life as created *de novo* by individuals, failing to acknowledge the internalization of societal constraints within the individual. Objectivism tends to adopt a mechanistic model of human action, and fails to recognize that social life is a practical achievement of competent individuals. It mistakenly reduces the complexity of practices to a coherent and simplified cultural logic. Hence Bourdieu's argument for 'objectifying the act of objectification' implies that the researcher, while observing and objectifying, takes a similar critical distance towards the objectification itself. Bourdieu's proposal for 'participant objectivation' attempts to objectify the object of research, examine the validity and presuppositions of that objectification, and finally to take into account people's skilful achievements. Relatedly, the practices of the social scientist in general are embedded in structural conditions and power struggles themselves. Hence Bourdieu's 'reflexive sociology' aims at a critical distance regarding its own practice.

Central to Bourdieu's view is people's practical mastery of the logic of everyday life. Most of the time people know how to go on in their daily activities, without needing to formulate that knowledge discursively. This practical knowledge is thus not part of the realm of consciousness, nor is it, strictly speaking, unconscious. People's practical logic relies upon *doxa* or 'doxic experience', that is, the taken-for-granted nature of people's daily existence. Both practical logic and *doxa* are central to Bourdieu's concepts of *'habitus'*, 'strategy' and

‘field’. *Habitus* refers to an acquired generative scheme of dispositions. These dispositions are tacitly acquired in early childhood, and, once inculcated, they endure. Dispositions generate practices, perceptions or bodily *‘hexis’*, adjusted to the constraints of the social world in which the *habitus* has emerged. Hence different social backgrounds will produce a different *habitus*. The *habitus* provides a ‘feeling for the game’. It makes it possible for people to develop any number of strategies attuned to an infinite number of situations. The external social world consists of ‘fields’: that is, areas where, through strategies, struggles take place over goods or resources (capital). Fields are not restricted to struggles over economic capital: they might also deal with social capital (contacts and acquaintances), cultural capital (education, culture and related skills) or symbolic capital (distinction and prestige). Although Bourdieu’s use of ‘capital’ suggests that interests are at stake, these are not necessarily material, nor do individuals normally adopt a conscious calculative orientation towards them.

Bourdieu has been criticized mainly for his alleged lack of analytical precision and for his tendency to read too much into his empirical material. Although there is some truth in both criticisms, neither affects the core of Bourdieu’s frame of reference. In this context, his strengths reveal his weaknesses. Bourdieu’s incisiveness lies in accounting for the reproduction of society, as becomes apparent in his notion of skilful, unquestioned reproduction of structures or in the presupposition that the *habitus* tends to be adjusted to social constraints. One of the upshots of this is that Bourdieu pays less attention to the ability of individuals to distance themselves from the facticity of daily existence - the ability to turn tacit knowledge into theoretical knowledge. Bourdieu tends to do so only in so far as people’s distancing towards everyday life is a result of social scientific intervention. He thereby fails to acknowledge fully that people may also exhibit that ability for distancing in the absence of scientific interference. If people’s theoretical knowledge enters the public-collective realm, it might become an important source of change or deliberate maintenance of structures.

See also: Anthropology, philosophy of; Sociology of knowledge; Sociology, theories of; Social science, methodology of

PATRICK BAERT

List of works

Bourdieu, P. (1972) *Esquisse d’une théorie de la pratique: précédée de trois études d’ethnologie kabyle*, Geneva: Droz; trans. R. Nice *Outline of a Theory of Practice*, Cambridge: Cambridge University Press, 1977.(Bourdieu discusses the main outline of his social theory. Complex. The English version differs substantially from the original *Esquisse*.)

Bourdieu, P. (1980) *Le sens pratique*, Paris: Les Editions de Minuit; trans. R. Nice, *The Logic of Practice*, Cambridge: Polity Press, 1990.(Bourdieu explains very clearly some of the themes introduced in *Esquisse*. He uses his framework to analyse kinship and other social phenomena. More accessible than *Esquisse*.)

Bourdieu, P. and Wacquant, L. (1992) *An Invitation to Reflexive Sociology*, Cambridge: Polity Press.(Has an excellent introduction by Wacquant to Bourdieu’s sociology. Also contains a useful interview by Wacquant with Bourdieu.)

References and further reading

Calhoun, C., LiPuma, E. and Postone, M. (eds) (1993) *Bourdieu: Critical Perspectives*, Cambridge: Polity Press.(An inspiring collection of articles. Reaches a high level while remaining remarkably lucid.)

Jenkins, R. (1992) *Pierre Bourdieu*, London: Routledge. (An elementary introduction, but Jenkins’ criticisms are sometimes unclear.)

Harker, R., Mahar, C. and Wilkes, C. (eds) (1990) *An Introduction to the Work of Pierre Bourdieu: The Practice of Theory*, Basingstoke: Macmillan.(A lively reconstruction of the major themes in Bourdieu’s work, while taking into account the French academic context. One chapter is an interview with the man himself.)

Robbins, D. (1991) *The Work of Pierre Bourdieu: Recognising Society*, Milton Keynes: Open University Press.(A comprehensive and chronological account of Bourdieu’s writings.)

Boutroux, Émile (1845-1921)

The French philosopher Émile Boutroux wanted to reestablish metaphysics in the face of a growing tendency towards materialism, but without rejecting the natural sciences. He hoped to achieve this by showing that only an immaterial mind that is a free and final cause of everything that is determined can give an absolute foundation to the sciences and to nature. Scientific determinism, according to which all phenomena are governed by mathematical necessities, is not incompatible with freedom. Indeed, the contingency of things and of human reason, which one sees in scientific experience, shows that the mind is free; it is therefore only mind which can give a determined existence to things and necessity to scientific explanations. In trying to reconcile metaphysics and science through a philosophy of nature, Boutroux represents a major turning point in French spiritualism, foreshadowing not only Bergson but also Bachelard.

Boutroux defended his thesis *De la contingence des lois de la nature (On the contingency of the laws of nature)* in 1874. Having taught at the École Normale Supérieure, among other places, he moved to the Sorbonne, where he became Professor of the History of Modern Philosophy in 1888. As a member of the Academy of Moral and Political Sciences and of the Académie Française, he was one of the important thinkers of his time. He was a former pupil of Jules Lachelier and participated in a brilliant and original way in the growing renewal of spiritualist thought brought about by Ravaisson from the end of the Second Empire. As free from the declining eclecticism as he was hostile to scientific materialism, Boutroux upheld a rationalism influenced by Leibniz that tried to include elements of British empiricism and the positivism of Comte. He studied Aristotle, Descartes and Kant, but concentrated on German philosophy of the seventeenth century, and above all on Leibniz, whose *Monadology* and *New Essays on Human Understanding* (First book) he translated into French. His lecture courses on natural law, given in 1892-3, contain a critique of scientific rationalism through which he returned to the metaphysical position put forward in his thesis of 1874.

As a historian of philosophy,Boutroux denounced the Hegelian philosophy of history and its introduction into France by Victor Cousin (see Hegelianism §4). He did not want to judge works dogmatically in looking for a successive development of an immanent spirit in history. As history is not a system in which the historian is required to uncover a teleological regularity under the particularities, Boutroux practised a critical study of philosophies taken in themselves as individual. Without neglecting the historical context, he attempted to make clear the acts of free creativity that throughout history have led philosophers to return time and again to the great problems of metaphysics.

Boutroux's philosophy of nature gives meaning to this conception of history. In it he criticizes the modern idea of natural law and its regulating principle of determinism, according to which everything is governed by necessary mathematical relations. According to Boutroux, laws are contingent because the relations they establish are never entirely a priori, and because reality is incommensurable and irreducible to mathematical necessity. As mathematics is applied to experience, the laws become more determinate and particular, and thereby less necessitating. It is an error of scientism to confuse determinism with necessitarianism. Science, in its effort to reduce experience to the mathematical, loses its sense of the radical contingency at work in nature; it generalizes and takes things to extremes in transforming its useful regulative idea into a constitutive principle of nature.

This philosophy of science foreshadows certain ideas that are present both in Bergson's metaphysics andBachelard's epistemology (see Bachelard, G. §2; Bergson, H.-L. §5). For example, there is the empiricist idea, taken fromRavaisson and developed by Bergson, according to which scientific laws are habits of reproduction that assimilate reality to our minds so that we may act on it, but which only partially correspond to how things are. There is also the idea that scientific rationalism proceeds by using an artificial construction of experience without which it would never be able to apply itself; and that the specialization of the sciences results in an irreducible pluralism of local determinisms; ideas which Bachelard developed in more radical and unforeseen ways. Nevertheless, it is metaphysics that he wants to save, as much from being dissolved by science as from being divorced from it. Boutroux wanted to retain rationalism without falling either into scientism, or into the illusion denounced by Kant as the transcendental use of the laws of nature, or into the dangerous error of separating the deterministic natural world from the intelligible world of freedom, as Kant did (see Kant, I. §§8, 9). It is in his

reading ofAristotle, and still more of Leibniz, that Boutroux tries to reconcile determinism and freedom, in showing that mechanism presupposes finality. Contingency - which is not chance but determination through a *telos* (life), and, from the human point of view, determination by the idea of progress - is presupposed by determinism, because it is incapable of giving an account of it.

This movement that reaches above science to metaphysics is completed in a moral reflection on the relation between science and faith. Boutroux shows that, as human beings are conscious living things irreducible to mechanism, there is no contradiction between science and religion. Science presupposes not only a reality that extends beyond what it can assimilate, but also that the human mind goes beyond the intellectual faculties it employs. Both outside and inside us, it presupposes a creative vital force that, in human beings, implies faith, the ideal and love, the three principles of all religions. Religious works, dogmas and rites are neither adventitious for religion nor unjustifiable for reason, even if religion is a matter of feeling and if its truth, as it would be developed by Bergson, is a mystical sentiment. Thus it is once again a matter of demonstrating the possibility of a spiritualism that goes beyond the scientific outlook without opposing it.

Translated from the French by Robert Stern

DIDIER GIL

List of works

Boutroux, É. (1879) *De la contingence des lois de la nature*, Paris: Alcan; trans. F. Rothwell, *The Contingency of the Laws of Nature*, Chicago, IL and London: Open Court, 1916.(Contains Boutroux's first discussion of necessity and contingency.)

Boutroux, É. (1889) *La philosophie allemande au XVIIe siècle (German philosophy of the seventeenth century)*, Paris: Vrin.(Mainly contains material on Leibniz.)

Boutroux, É. (1895) *De l'idée de loi naturelle dans la science et la philosophie contemporaine*, Paris: Vrin; trans. F. Rothwell, *Natural Law in Science and Philosophy*, London: Macmillan, 1914. (Lectures of 1892-3, further developing Boutroux's views on necessity and contingency.)

Boutroux, É. (1908a) *Science et religion dans la philosophie contemporaine*, Paris: Flammarion; trans. as *Science and Religion in Contemporary Philosophy*, Associated Faculty Press Ltd, 1970.(Attempts a reconciliation between scientific rationalism and religious thought.)

Boutroux, É. (1908b) 'Rapport sur la philosophie en France depuis 1867' (Report on philosophy in France since 1867), in *Nouvelles études d'histoire de la philosophie*, Paris: Alcan, 1927.(Account of the state of contemporary French philosophy.)

Boutroux, É. (1926) *Nature et esprit (Nature and mind)*, Paris: Vrin.(Posthumously published work.)

References and further reading

Brunschvicg, L. (1922) 'La philosophie d'Émile Boutroux' (The philosophy of Émile Boutroux), *Revue de Métaphysique et de Morale* 1922.(A summary of Boutroux's position.)

Archambault, P. (1928) *Émile Boutroux*, Paris: Vald. Rasmussen.(General study.)

Gil, D. (1996) 'La philosophie de la nature d'É. Boutroux' (Boutroux's philosophy of nature), in *Les philosophies de la nature*, Paris: Presses Universitaires de France.(Discussion of Boutroux's philosophy of nature.)

Bowne, Borden Parker (1847-1910)

Bowne was one of the most influential thinkers and writers of the American personalist school of philosophy. His position is theistic and idealistic, and finds in human persons the key to meaning in the world. Knowledge comes only through personal experience, through which we understand ourselves to be enduring thinking entities with a certain degree of freedom. The uniformity of God's activity is such as to make nature intelligible to us, but our minds are nevertheless independent of God's.

Borden Parker Bowne was born in a New Jersey manse and educated at Pennington Seminary, New York University and in Europe (chiefly Paris, Halle and Göttingen). He taught at Boston University from 1876 to 1910, serving as the first dean of the graduate school. His views were strongly influenced by the ideas of Kant, Lotze (with whom he studied) and Bergson. He was ordained in the Methodist Episcopal Church as a local preacher in 1872 and an elder in 1882. During a long career of teaching and publishing he became the most influential exponent of American personalism, a philosophical school of which he was arguably the founder.

Bowne held that the basis for knowledge is personal experience: 'For each person, his own self is known in immediate experience and all others are known through their effects' (1902: 269). If the world is intelligible, there is continuity that underlies the flux of human experience; in self-consciousness, one is aware of various sensory and introspective phenomena that are unified in that they are objects of a single awareness. In memory, one recognizes the identity of one's current centre of awareness with prior centres. Thus one finds in oneself the continuity of a mental substance; one knows oneself as an enduring, thinking thing. One also experiences a measure of self-control, so that purposive action is possible, feelings can be controlled, and competing ends assessed; we find that we have a relative independence of our environment and of others. But we also experience a degree of dependence on our environment and on others that makes it clear that our independence is only relative. Bowne held that these facts are incontrovertible, and that no philosophy can be adequate that denies or ignores them.

One thing Bowne believed these facts to rule out is any view in which the thoughts and feelings that some or all human beings experience are really the thoughts and feelings that God has in or through them. Such a view is incompatible with both our relative independence (we are self-directing agents, not mere passive receivers or conduits) and our relative dependence (God cannot be dependent). These facts, even without further appeal to such things as our having false beliefs and our performing wrong actions, make it clear that no notion that our consciousness is really part of God's consciousness is defensible. Thus God is not to be thought of along the lines of 'the collection of all finite minds'; rather God possesses an independence and freedom on which our relative independence and freedom is modelled.

Bowne's view was that the sort of absolute idealism for which each human mind is somehow part of a cosmic mind, and the deterministic materialism for which mind reduces to matter, while at opposite ends of the metaphysical spectrum, are alike in certain of their most important consequences. Leaving no room for free human agency, both are incompatible with human responsibility. Both make evil action impossible: the former by making God the only agent, and the latter by allowing for no agents at all but only mechanistically produced effects.

Bowne had no difficulty in accepting evolutionary theory, and viewed the evolutionary process as directed by God so as to produce human beings. As a metaphysical idealist, he held that one's body at any particular time is simply a collection of one's sensory images, and over time is but a series of such collections. The bodies of persons are images that they do not create, and the laws of nature are descriptions of the uniformity of God's activity in causing our perceptual images. Since God's willing is rational, what we call 'nature' is intelligible. This makes science and planned practical activity possible. One consequence of this sort of view is that 'existing in this world' is a matter of having perceptual and introspective experiences of a familiar, if not easily characterized, form and type. Thus 'existing in another world' would be a matter of having perceptual and introspective images of a significantly different form and type. A complete world will be composed of various such 'worlds' and will exist within a single divine purpose, being unified in that each 'world' occurs within the framework of a single coherent divine plan that assures 'fit' between them.

See also: Personalism

KEITH E. YANDELL

List of works

Bowne, B.P. (1887) *Philosophy of Theism*, New York: Harper & Brothers.(Presents theism as the fundamental postulate of our total life.)

Bowne, B.P. (1897) *The Theory of Thought and Knowledge*, New York: Harper & Brothers. (Detailed account of Bowne's idealistic epistemology.)

Bowne, B.P. (1902) *Theism*, New York: American Book Company.(Revision and expansion of *Philosophy of Theism*.)

Bowne, B.P. (1908) *Personalism*, Boston, MA: Houghton Mifflin.(A more popular presentation of the personalist viewpoint, based on the N.W. Harris Lectures of 1907 at Northwestern University.)

Bowne, B.P. (1910) *Metaphysics*, New York: American Book Company.(Detailed argument for Bowne's personalist metaphysic.)

References and further reading

Cunningham, G.W. (1953) *The Idealistic Argument in Recent British and American Philosophy*, Westport, CT: Greenwood Press, 1969.(Discusses a century of Anglo-American idealist thought (1808-1910) in expository and critical terms; includes a chapter on Bowne.)

McConnell, F.J. (1929) *Borden Parker Bowne*, New York: Abington Press. (Biography of Bowne by close friend who was a Methodist bishop.)

Boyle, Robert (1627-91)

Boyle is often remembered for the contributions that he made to the sciences of chemistry and pneumatics. Like other natural philosophers in seventeenth-century England, however, he was a synthetic thinker who sought to advance knowledge in all areas of human concern. An early advocate of experimental methods, he argued that experimentation would not only reveal the hidden processes operative in the world but would also advance the cause of religion. Through the study of nature, experimentalists would come to understand that the intricacy of design manifest in the world must be the result of an omniscient and omnipotent creator.

Boyle's experimental investigations and theological beliefs led him to a conception of the world as a 'cosmic mechanism' comprised of a harmonious set of interrelated processes. He agreed with the leading mechanical philosophers of his day that the corpuscular hypothesis, which explains the causal powers of bodies by reference to the motions of the least parts (corpuscles) of matter, provided the best means for understanding nature. He insisted, however, that these motions and powers could not be known by reasoning alone, but would have to be discovered experimentally.

1 Life and works

Robert Boyle, born 1627 at Lismore Castle, Ireland, was the youngest son of the first Earl of Cork, one of the wealthiest members of the English aristocracy. Boyle's formal schooling was limited to four years at Eton (1635-9) after which he embarked upon a tour of Europe. He read the works of Galileo while visiting Florence, and studied the classics, languages, Calvinist theology and Epicurean philosophy during an extended stay in Geneva. Upon the death of his father in 1644 he returned to England and resided in London with his sister Katherine, Lady Ranelagh, through whom he met a number of leading educational theorists, mechanical philosophers and chemists such as Samuel Hartlib, John Milton, Kenelm Digby and Benjamin Worseley.

In 1646 Boyle retired to his inherited estate at Stalbridge where he wrote moral and theological essays, compiled catalogues of medicinal remedies, and began experimental studies in chemistry and anatomy. He moved to Oxford in 1655 in order to work with an informal group of experimentalists meeting there that included Robert Hooke, John Locke and Christopher Wren. While at Oxford he also associated with theologians and linguists such as Thomas Hyde, Samuel Clarke and Thomas Barlow, and he became involved in a number of business ventures including the Hudson's Bay Company and the East India Trading Company. In 1665 he received an honorary medical degree from Oxford. In 1668 he returned to London and lived there with his sister until both of them died, a week apart, in December 1691.

Boyle gained international recognition with the publication of his first major scientific work, *New Experiments Physico-Mechanical, Touching the Spring of the Air* (1660), in which he reported a series of experiments performed with an air-pump designed to support the notion of the weight of the air developed by Pascal and Torricelli and his own conception of the spring (elasticity) of the air. His work was seen by many as a significant improvement upon earlier Aristotelian speculations, yet it was also severely criticized, by Henry More and Thomas Hobbes among others. In 1662 he published responses to his critics in a second edition of *New Experiments* that also contained the first formulation of Boyle's law, which describes the inverse proportion between the pressure and volume of a gas.

Boyle's interests were broad and eclectic and his inherited wealth provided him with the leisure to pursue many areas of study. During his lifetime he published over forty works ranging from comprehensive experimental histories to defences of the Christian religion and the new corpuscular philosophy. In one of his last books, *The Christian Virtuoso* (1690), he explained that such eclecticism was warranted because ' true philosophy' is 'of greater extent, than the hypothesis of any one sect of philosophers, being indeed a comprehension of all the sciences, arts, disciplines, and other considerable parts of useful knowledge'.

2 Experimental philosophy

Boyle first became acquainted with the Baconian concept of useful knowledge through his association with the Hartlib group in London, and his own laboratory investigations, particularly in chemistry, increased his conviction that experimental practices would be required to improve upon what he perceived to be the 'barrenness' of the

natural philosophy taught in the schools. He agreed with Bacon that the Aristotelian distinction between natural and artificial processes was not tenable and that the experimental manipulation of natural bodies could yield more than immediate utilitarian benefits (see Bacon, F.). As Boyle explained in his *Usefulness of Experimental Philosophy* (1670), the production of effects under the controlled conditions of a laboratory may 'either hint to us the causes of them, or at least acquaint us with some of the properties or qualities of things concurring to the production of such effects'.

Although he designed his experimental method as a way by which to learn about the causal processes operative in nature, he opposed what he found to be the premature theoretical systems of many of his contemporaries. In *The Sceptical Chymist* (1661), for example, he produced experimental refutations of the prevalent Aristotelian and alchemical theories of elements, but he offered no theoretical account of his own with which to replace them (see Alchemy; Natural philosophy, medieval §1). In other works where he did speculate about the causal powers of bodies, his discussions were tentative and hesitant. Boyle's caution was in part a result of the many practical problems that he encountered in his laboratory. His reflection upon the epistemological significance of these 'contingencies of experiment' led him to write one of the earliest and most complete accounts of the systematic errors associated with experimental practice.

In his *Certain Physiological Essays* (1661), Boyle discussed how the use of impure ingredients, imperfect instruments or inappropriate techniques could lead to experimental failure. To mitigate the effects of systematic error, he designed a number of methodological strategies. Some were meant for specific areas of investigation, such as the indicator tests that he developed for the identification of acids and alkalies. Others, such as repetition and variation, were offered as general strategies to be used for all areas of investigation. In order to establish that an experimental result provides reliable evidence about a natural process, one has to ensure that it is not an artefact of the particular materials or instruments used in its production. Not only is it necessary to repeat experimental trials and vary their circumstances, it is also necessary to publish complete accounts of experiments so that others will be able to repeat the trials for themselves or at least judge the appropriateness of the experimental conditions and the soundness of any inferences drawn from them. Boyle's practical implementation of his programmatic methodological statements can be seen in the experimental histories that he wrote on *Colours* (1664), *Cold* (1665), *Blood* (1684) and *Air* (1692).

Unlike the stress upon individual reasoning found in the Scholastic and Cartesian philosophical traditions, Boyle insisted that collaboration and cooperation among naturalists were necessary for the advancement of knowledge. This belief was a product of his methodological dictates as well as a consequence of his view concerning the justification of theoretical speculations. According to Boyle, a theory is acceptable only after a 'concurrence of probabilities' has been produced whereby all of the relevant and available evidence supports a particular conclusion.

Experimentation should 'beget a confederacy, and an union between parts of learning', because often more than one area of investigation would be required to learn about a natural process. A full understanding of the circulation of the blood, for example, had required the efforts of workers whose areas of expertise were in physiology, anatomy, chemistry and mechanics. As he explained in his *Usefulness of Natural Philosophy* (1664), there is a 'dependency, continuation, and confederacy of causes' in nature that results from the 'secret correspondencies and alliances' between things. This ontological conception was supported by his experimental work as well as by his theological worldview.

3 Theology

In addition to being an experimentalist, Boyle was also a lay theologian who emphasized the importance of revealed religion. In *The Excellency of Theology* (1674) he argued for the intellectual superiority of scriptural studies over natural philosophy and early in his career he published a work on what we would today call biblical hermeneutics, *The Style of the Holy Scriptures* (1661), that is significant for understanding his views on interpretation. According to Boyle, the 'seeming contradictions' identified by critics of the Bible are the product not of its divine author but 'of our ignorance'. The Bible is a complex interrelated whole that must be 'coherent' because God would not require humans to believe inconsistencies. The proper method of interpretation, therefore, would require that readers attempt to eliminate apparent contradictions by a gradual process of reconciliation.

In *A Free Inquiry into the Received Notion of Nature* (1686), Boyle discussed how nature could also be described as a divine text because it is composed of a complex set of interrelated parts all of which were designed by God to produce one coherent whole. The relations that obtain in nature are the 'result of the universal matter, or corporeal substance of the universe, considered as it is contrived in the present structure and constitution of the world, whereby all bodies, that compose it, are enabled to act upon, and fitted to suffer from one another'. In anticipation of Leibniz, Boyle discussed the perfection of the world and maintained that there is an 'intercourse and harmony between truths'. Naturalists must attempt to discover the 'connection of physical truths, and the relations that material bodies have to one another' and thus the interpretation of nature would require the same type of reconciliation process that he had recommended for the Scriptures.

Boyle criticized the Cartesians and atomists who excluded considerations of final causality from their philosophical systems because he believed that the investigation of nature provided one of the best means for learning about its creator. In his *Disquisition about the Final Causes of Natural Things* (1688), he argued that God may 'declare truths to men, and instruct them, by his creatures and his actions, as well as by his words'. The study of nature could thus lead an experimentalist to the 'acknowledgement of the divine Architect's power, wisdom and beneficence'. He did agree with his contemporaries' criticisms of Aristotelian teleology, however, and he maintained that it was not legitimate to refer to God's purposes when explaining physical processes. Yet some appeal to God's initial act of creation could provide a valuable heuristic for natural investigations. Indeed, his conception of the coherence of the world guided his specific formulation of the corpuscular hypothesis and his belief in the omniscience of the creator led him to insist that corpuscular explanations could not be justified by a priori reasoning, but would require experimental proof.

4 Corpuscular hypothesis

In his 'Excellency and Grounds of the Mechanical or Corpuscular Hypothesis' (1674) Boyle argued that corpuscularianism was simpler and more intelligible than the alternative Aristotelian doctrine that appealed to the occult qualities and substantial forms of bodies. In his *Origin of Forms and Qualities* (1666) he also defended a mechanical account of nature by which physical phenomena were to be explained by considering 'only the size, shape, motion (or want of it), texture, and the resulting qualities and attributes of the small particles of matter'. These arguments were similar to those of other mechanical philosophers. Boyle's extensive experimental practice and his interest in the chemical composition of bodies led him to reject what he saw to be the purely quantitative and mathematical analysis of Descartes, however, and to advocate a more qualitative approach to the study of nature. According to Boyle, it is the set of qualities possessed by a body that defines its essence and gives it the causal power to produce changes in other bodies. He developed an account of the primary and secondary qualities of bodies, which would receive further elaboration and refinement in the philosophical writings of John Locke, in order to describe how corpuscular explanations would allow for a qualitative understanding of physical phenomena (see Locke, J. §4; Primary-secondary distinction §1).

Primary qualities are those 'affections', such as size and shape, that must belong to all parts of matter. Secondary qualities, on the other hand, are those properties of bodies such as colour and temperature that result from the internal configuration of the parts of matter of which they are composed as well as the external relations that they have to other bodies. Boyle maintained that these secondary qualities ' do as well seem to belong to natural bodies generally considered, as place, time, motion and those other things… treated of in the general part of natural philosophy'. Although the secondary qualities of bodies are causally responsible for human sensations, they are not 'like to the ideas they occasion in us'. Heat, for example, is a product of the rapid motion of the particles composing a body, such as the sun. The sun has the secondary quality of heat that gives it the power to produce the sensation of heat in a human, but even if all 'sensitive beings in the world were annihilated' the sun would retain its secondary quality and thus continue to have the power to affect other bodies such as ice and wax.

In *The Mechanical Origin or Production of Divers Particular Qualities* (1675), Boyle reported experiments that he had performed to show how secondary qualities could be added to or taken away from bodies simply by a mechanical alteration of their structure. Repeatedly striking a metallic body with a hammer will cause it to become hot, for example, whereas a piece of glass operated upon in the same way will lose its transparent quality and be reduced to a white powder. Although this work helped to establish the feasibility and intelligibility of mechanical explanations, Boyle was never able to specify the ultimate causal mechanisms responsible for the effects that he

produced in his laboratory. His lack of theoretical success may be attributed to his overly cautious attitude that resulted from his appreciation of the fallibility of experimental practice and the complexity of natural processes, as well as to his eclectic and comprehensive approach whereby he sought to combine researches in physics, chemistry and biology. His works are of lasting historical significance not so much because of the specific theories that he developed but because of his successful promotion of a new way of thinking about nature and a new method for investigating physical processes. A self-proclaimed 'underbuilder', he provided the materials and methods that later thinkers such as Newton and Lavoisier would use to great advantage (see Newton, I.; Chemistry, philosophical aspects of §3).

See also: Atomism, ancient; Experiment; Matter; Scientific method

ROSE-MARY SARGENT

List of works

Boyle, R. (1744) *The Works of the Honourable Robert Boyle*, ed. T. Birch, London, 5 vols. Available on microfilm in the *Landmarks of Science* series, eds Sir H. Hartley and D.H.D. Roller, New York: Readex Microprint, 1967-76.(Contains all of the works published by Boyle during his lifetime as well as some posthumous works, an autobiographical account of his youth, Birch's life of Boyle and correspondence between Boyle and a number of the leading figures of his day.)

Boyle, R. (1772) *The Works of the Honourable Robert Boyle*, ed. T. Birch, London, 6 vols; repr. Hildesheim: Georg Olms, 1965-6. (Contents the same as the 1744 edition. The reprint has an introduction by Douglas McKie.)

Boyle, R. (1991) *The Early Essays and Ethics of Robert Boyle*, ed. J. Harwood, Carbondale, IL: Illinois University Press.(Contains selections from Boyle's early ethical and theological manuscripts. Harwood's introduction provides helpful information concerning the historical context in which these papers were composed.)

Boyle, R. (1992) *The Letters and Papers of Robert Boyle*, Bethesda, MD: University Press of America.(A microfilm collection of all of Boyle's surviving manuscripts, notebooks and correspondence housed at the Royal Society of London. A separate introduction and guide to the manuscript material by Michael Hunter is also available.)

References and further reading

Alexander, P. (1985) *Ideas, Qualities and Corpuscles: Locke and Boyle on the External World*, Cambridge: Cambridge University Press.(An interesting and thorough examination of Boyle's views on the experimental investigation of the corpuscular constitution of matter and how his work influenced John Locke.)

Conant, J.B. (1970) 'Robert Boyle's Experiments in Pneumatics', in *Harvard Case Histories in Experimental Science*, vol. 1, Cambridge, MA: Harvard University Press. (Contains an edited version of Boyle's New Experiments Physico-Mechanical and a discussion of the methodological aspects of this work.)

Fulton, J. (1961) *A Bibliography of Robert Boyle*, London: Oxford University Press. (Contains the most complete bibliography of Boyle's works as well as secondary works published about Boyle up to 1940.)

Hall, M.B. (1958) *Robert Boyle and Seventeenth-Century Chemistry*, Cambridge: Cambridge University Press.(A detailed examination of Boyle's role in the chemical revolution of the seventeenth century.)

Hall, M.B. (1965) *Robert Boyle on Natural Philosophy: An Essay with Selections from His Writings*, Bloomington, IN: Indiana University Press.(Contains a brief account of Boyle's life followed by a discussion of his work. A detailed and well-documented study that contains substantial excerpts from Boyle's published works.)

Hunter, M. (ed.) (1994) *Robert Boyle Reconsidered*, Cambridge: Cambridge University Press.(Contains a helpful introductory essay by Hunter that discusses scholarship on Boyle prior to 1994. The individual essays cover Boyle's experimental, philosophical and theological works as well as his intellectual and social context. Contains a complete bibliography of all works published on Boyle between 1940 and 1993.)

Jacob, J.R. (1977) *Robert Boyle and the English Revolution*, New York: Franklin. (Speculates about the political motivations behind Boyle's scientific and religious views based upon early manuscript material and the known political ambitions of some of Boyle's associates.)

Maddison, R.E.W. (1969) *The Life of the Honourable Robert Boyle*, New York: Barnes & Noble.(The most definitive life of Boyle, corrects some errors contained in Birch's account as well as supplementing it with material from the private papers of Boyle's relations.)

Sargent, R.-M. (1995) *The Diffident Naturalist: Robert Boyle and the Philosophy of Experiment*, Chicago, IL: University of Chicago Press.(An analysis of Boyle's development and practice of experimental methodology from a philosophical perspective.)

Shapin, S. and Schaffer, S. (1985) *Leviathan and the Air-pump: Hobbes, Boyle, and the Experimental Life*, Princeton, NJ: Princeton University Press.(Discusses the controversy surrounding Boyle's pneumatic experiments primarily from the point of view of his critics, and focuses upon the political implications of the debates.)

Wojcik, J.W. (1997) *Robert Boyle and the Limits of Reason*, Cambridge: Cambridge University Press.(Examines Boyle's theological and philosophical works in the historical context of the religious debates in seventeenth-century England.)

Bradley, Francis Herbert (1846-1924)

Bradley was the most famous and philosophically the most influential of the British Idealists, who had a marked impact on British philosophy in the later nineteenth and earlier twentieth centuries. They looked for inspiration less to their British predecessors than to Kant and Hegel, though Bradley owed as much to lesser German philosophers such as R.H. Lotze, J.F. Herbart and C. Sigwart.

Bradley is most famous for his metaphysics. He argued that our ordinary conceptions of the world conceal contradictions. His radical alternative can be summarized as a combination of monism (that is, reality is one, there are no real separate things) and absolute idealism (that is, reality is idea, or consists of experience - but not the experience of any one individual, for this is forbidden by the monism). This metaphysics is said to have influenced the poetry of T.S. Eliot. But he also made notable contributions to philosophy of history, to ethics and to the philosophy of logic, especially of a critical kind. His critique of hedonism - the view that the goal of morality is the maximization of pleasure - is still one of the best available. Some of his views on logic, for instance, that the grammatical subject of a sentence may not be what the sentence is really about, became standard through their acceptance by Bertrand Russell, an acceptance which survived Russell's repudiation of idealist logic and metaphysics around the turn of the century. Russell's and G.E. Moore's subsequent disparaging attacks on Bradley's views signalled the return to dominance in England of pluralist (that is, non-monist) doctrines in the tradition of Hume and J.S. Mill, and, perhaps even more significantly, the replacement in philosophy of Bradley's richly metaphorical literary style and of his confidence in the metaphysician's right to adjudicate on the ultimate truth with something more like plain speaking and a renewed deference to science and mathematics.

Bradley's contemporary reputation was that of the greatest English philosopher of his generation. This status did not long survive his death, and the relative dearth of serious discussion of his work until more general interest revived in the 1970s has meant that the incidental textbook references to some of his most characteristic and significant views, for example, on relations and on truth, are often based on hostile and misleading caricatures.

1 Life and works

Bradley was born in Clapham, England, on 30 January 1846. The literary critic A.C. Bradley was a younger brother. Educated at fee-paying schools and University College, Oxford (where he studied classical languages and literature, and ancient history and philosophy), he began the study of German, and is recorded to have read some at least of Kant's *Critique*, while still at school. In 1870 he was elected to a fellowship requiring no teaching and terminable only on marriage, at Merton College, Oxford. Six months later he suffered a severe inflammation of the kidneys: he had little public life thereafter, for cold, fatigue or stress was apt to make him ill. He occupied his fellowship until his death from blood poisoning on 18 September 1924. His major works are *Ethical Studies* (1876), *The Principles of Logic* (1883), *Appearance and Reality* (1893), *Essays on Truth and Reality* (1914) and the posthumously published *Collected Essays* (1935). He was awarded honours both foreign and domestic, including the Order of Merit. Though a freethinker, he was said to be politically conservative. His writings reveal a character far from narrowly intellectual.

2 Philosophy of history

Bradley's first publication was the pamphlet 'The Presuppositions of Critical History' (1874). Though perhaps the earliest major theoretical study in English of the notion of historical fact, it had little impact at the time, but the kind of position it takes has been subsequently influential, especially in religious studies. Bradley's acknowledged sources were German historians of the origins of Christianity, but his position resembles that of Hume on miracles in its scepticism concerning astonishing historical reports. The question raised by Bradley's argument - by what criterion should the credibility of historical testimony be judged? - stimulated the reflections of such philosophers as R.G. Collingwood. This essay provides a good introductory sample of Bradley's writing: characteristic in its highly-charged style, frequent obscurity and disdain of example, it also anticipates some of his later holistic themes, for example, the fallibility of any individual judgment and the rejection of correspondence notions of truth.

3 Ethics

Greater recognition came with *Ethical Studies*, a work which more than any of his others reveals Hegel's influence

in both its ideas and its dialectical construction. This construction means that Bradley's prefatory remark that the essays 'must be read in the order in which they stand' should be taken seriously. Although the third essay is a *locus classicus* of arguments against hedonistic utilitarianism (see Hedonism; Pleasure; Utilitarianism) and the fifth presents with some passion a social conception of the moral life, the common idea that these two can be read in isolation as representing Bradley's own final views is mistaken. What *Ethical Studies* aimed at was a gradual working-out of an account of morality which, unlike the prevalent utilitarianism, did full justice to ordinary moral ideas and did not rely on a deficient notion of the self. (One of the book's governing notions is that ordinary moral thinking is not to be displaced by the fruits of moral philosophy.) This development originates in an examination of the 'vulgar' notion of moral responsibility, and a rejection of both determinism and indeterminism as one-sided views obtained by concentrating on different aspects of human action which coexist unproblematically but are made to appear as conflicting by abstraction from the whole (see Free will §§1-2, 4). It continues in the second essay by asking 'Why should I be moral?' His answer is that the moral end for each of us is self-realization, but as he holds all action to be self-realization whether the action be wicked or otherwise, he has to explain the kind of self-realization which is morality's goal: it is to realize oneself as an infinite whole (see Self-realization). One thing this may mean is that the fully moral self is not to be limited by any other self: that is, one aim of morality is the resolution of conflict between one's good and bad selves in favour of the former. Another is that self-realization can be accomplished only through the mutual dependence of self and society. But what it amounts to is meant to be revealed through consideration of representative philosophical theories, each of which, through its one-sidedness, is more or less unsatisfactory as it stands. One is hedonistic utilitarianism, from which, roughly speaking, Bradley drops the hedonism and individualism but retains the utilitarianism (being prepared to regard happiness as the goal of morality (see Happiness) provided it is not thought of as some independently identifiable state which could just as well be attained by some more convenient means, that is, as externally related to morality itself). Next comes a Kantian ethics of duty, from which he retains the idea that the performance of *duties* is essential to morality, while dropping the notion that duty should be done for duty's sake rather than because of the particular content of the individual duty (see Duty; Kantian ethics). Both these theories come to grief because of erroneous views about the nature of the self, which, Bradley thinks, is a concrete universal and essentially social. In the fifth essay he develops a Hegelian account of morality according to which the self is fully realized by playing its role in the social organism. While Bradley recognizes this account to be inadequate, because, for example, communities themselves can have moral imperfections, it nevertheless differs in status from the views he considered previously, for he regards it as merely requiring supplementation (undertaken in the next essay on ideal morality). But he does not make clear how self-realization as a part of the social organism can be an intelligible moral demand, for Bradley's argument against social contract theories - that they presuppose an impossible metaphysics of persons, in that the parties to the contract are only contingently social - requires him to hold that the self is already and necessarily social. Yet perhaps this does not matter, for the book closes by condemning morality itself as ultimately a self-contradictory enterprise, depending for its existence on the existence of the evil it seeks to overcome, and thus rendering impossible the ultimate realization of the ideal self, a realization obtainable only in religion.

Some of Bradley's metaphysical apparatus is deployed in his moral philosophy, along with anticipations of his idealism and hostility to external relations. One example is the concrete universal, a notion which arises from his rejection of the standard universal-particular distinction, on the grounds that this distinction abstracts unreal elements (that is, those incapable of independent existence) from actual things. Thus when we attribute greenness to a leaf, both the particular (the leaf without its greenness) to which the greenness is attributed and the universal (the greenness without the leaf) are figments of the intellect; any impression of their independent existence arises from the mechanisms of thought. Thought has to divide reality up like this in order to function at all, but thereby distorts its nature. The concrete universal and the concrete particular are both the one individual thing: a leaf is universal in collecting together diverse abstract particulars, such as its various stages over time, and particular in its being distinct from other leaves. Communities and individual persons are likewise concrete universals, the former retaining their identities over many generations, the latter through many actions, and Bradley thought moral philosophy had to recognize this. Although the expression 'concrete universal' rarely figures in Bradley's work subsequent to *The Principles of Logic*, the idea involved is fundamental to his thought in both logic and metaphysics in its encapsulation of the idea that abstraction is falsification (see Hegel, G.W.F. §6; Universals).

4 Logic

The Principles of Logic looks strange to anyone whose conception of logic has been shaped by the formalism of Frege and Russell. Devoid of mathematically inspired methods, of axioms and rules of inference, it proves no theorems and employs no calculi. Often the text seems contaminated with discussion of psychological matters. Familiar terms, such as 'identity of indiscernibles', are used in ways unfamiliar enough to make one unsure whether they have much to do with what other people mean by them, and the terminology is often bewilderingly loose. Nevertheless, and despite its idealist vocabulary and florid metaphor, its title is apposite, for the book is devoted to issues fundamental to logic. It brought the notion of meaning, for instance, to the centre of the philosophical stage; and the absence of calculi is at least partly the result of a principled opposition to formal logic and the mathematicization of reasoning on the grounds that inference was thereby detached from the practice of science and the acquisition of knowledge. Some of its doctrines have been greatly influential, often via their impact onRussell, such as the suggestion that the logical form of universal sentences is hypothetical (a prototypical instance of the distinction between logical and grammatical form). It is a transitional work, expressing in the vocabulary of ideas and judgments views which were to usher in the era of meanings, sentences and propositions.

Much of the book is polemical, and, as in *Ethical Studies*, Bradley develops his own views gradually through criticism of others. It is divided into three, the first dealing with judgment and the remainder with inference. He begins by arguing that atomists such as Hume, who thought of judgments in terms of ideas, failed to distinguish the sense of 'idea' in which ideas are important to logic: they are not datable occurrences like mental images but abstract universals. He is thus often portrayed as rejecting psychologism in logic, but this is an exaggeration, for he thinks of logic's subject matter as mental acts, not as sentences, or propositions in the sense of Russell or of Moore. He then rejects some standard accounts of judgment. He complains that a subject-predicate account cannot do justice to relational judgments, and that thinking of judgment as the coupling of two ideas makes it impossible to see how judgment can be about anything real, since all the ingredients of judgments are universal and belong to the realm of idea, whereas reality is stubbornly particular and actual. As Bradley also argued that there could be no unique designation of individuals, even grammatical names and demonstratives being disguised general terms, he may have planted the seed of Russell's elimination of grammatically proper names by application of the Theory of Definite Descriptions (see Russell, B.A.W.).Bradley's own account of judgment is 'the act which refers an ideal content … to a reality beyond the act'. (By 'ideal content' he means a universal abstracted from a mental image; later he realized that this overestimates the role of mental imagery in judgment.)

When Bradley turns to inference, his targets remain the same. He complains that the mathematical logics of his time cannot represent valid relational inferences. He rejects Hume's account of inference in terms of the association of ideas on the grounds thatHume's ideas, as datable particulars, are fleeting entities which cannot be revived by association (see Hume, D. §2). Association is possible between ideas only if they are universals. (He calls this process 'redintegration'.) He rejects both syllogistic and Mill's methods of induction for failing to recognize that reasoning can proceed only on the basis of the generality implicit in the universals essential to inference (see Inductive inference). His own account of inference is that it is 'ideal experiment': ideal in that it belongs to thought, experiment in that its results are not guaranteed in advance by a complete set of logical laws which infallibly determine their own application.

Underlying much of Bradley's criticism of previous accounts of judgment and inference is hostility to psychological atomism, whose particulars he regarded not as concrete universals, realities in their own right, but as abstractions from the continuous whole which is psychological life. But likewise he regarded judgments themselves as infected with abstraction, since their subject matter is necessarily selected from a background and accordingly falsifies reality. Thus the objections which destroy misleading accounts of logic start to threaten logic itself, and, consistently, the book ends by suggesting that no inference is ever really valid and no judgment ever really true. Here it spills over into the metaphysics it had ostensibly tried to avoid.Bradley's view is that logic presupposes a 'copy' (correspondence) theory of truth, but it is clear that he thinks this theory metaphysically inadequate, a view he develops in *Essays on Truth and Reality*, where he argues for 'the identity of truth knowledge and reality'. Thus the claim thatBradley held a coherence theory of truth (endlessly repeated in textbooks) is mistaken. (Because he held reality to be a unified whole, he thought the *test* of truth to be 'system', which includes what is commonly meant by coherence (see Truth, coherence theory of §1).) This identity theory of truth, that for a thought to be true is for it to be identical with the logical subject of which it is predicated - reality itself - has led to his being unfairly accused of confusing predication with identity. His relation to the

correspondence theory exemplifies a constant difficulty in trying to understand Bradley: he adopts various theories and suggestions temporarily only to abandon them as ultimately unsatisfactory, so that it can be hard to work out his commitments.

5 Metaphysics

Bradley's metaphysics gets its fullest exposition in *Appearance and Reality*, though this needs to be considered in the light of subsequent essays. It is divided into two books. The first, 'Appearance', is brief and destructive. It argues that 'the ideas by which we try to understand the universe' all involve us ultimately in contradiction. Some of these are philosophical, such as the suggestion that only primary qualities are real, while others belong to common sense, such as motion, space, time, relation, thing and self. The second, 'Reality', concerns the Absolute, the ultimate, unconditioned reality, undistorted by human conceptualization (see Absolute, the).

Many of the arguments in Book I are not unique to Bradley and make only part of his case: for example, primary qualities are inconceivable without secondary, motion involves paradoxes. But the arguments of the chapter 'Relation and Quality', which in generalized form allege that relations are unintelligible either with or without their terms, and terms unintelligible with or without their relations, are of a different order. Bradley himself said that a grasp of these arguments would lead the reader to condemn 'the great mass of phenomena', and it is clear that his views on relations are central to his thought. Thus it is unfortunate that the arguments are so sketchily and unconvincingly presented that even sympathetic commentators have found it hard to defend him. In part this is because the arguments are often read as designed to prove the doctrine of the internality of all relations (that is, their reducibility to qualities or their holding necessarily, depending on the sense of 'internal'). This is a misreading, but it is understandable, for Bradley flirted with the doctrine of internality in *Appearance and Reality*, only repudiating it in later works less frequently read, like the unfinished essay 'Relations'. Also he rejected the reality of external relations, and it is natural to interpret this as adherence to the doctrine of internality. But his considered view was that neither external nor internal relations are real. One of his main arguments for this conclusion was that if relations were another kind of real thing along with their terms, then a further relation would be required to relate them to the terms (and so on, *ad infinitum*). It is clear from this and from his own explanation that to be real is to be a substantial individual, that the denial of the reality of relations is the denial that they are independent existents (see Substance). But some have thought he meant that all relational judgments are false, for instance, that it is false that Galileo preceded Newton. AndBradley's theory of truth gives credence to such suggestions, for by that theory no ordinary judgment is perfectly true, so that to someone who assumes that truth is two-valued, his claim looks to be that such judgments are all false. But Bradley thought truth admitted of degrees, and that, provided we confine ourselves to everyday purposes and do not try to meet the exacting demands of metaphysics, it is true that Galileo preceded Newton. The imperfection of this truth is nothing to do with the judgment's being relational as opposed to predicative. A perfect truth would be one which did not abstract from reality at all. Such a complete description would have to be identical with reality itself and thus would no longer be a *judgment*. On Bradley's view, then, the final truth about reality is literally and in principle inexpressible.

An outline, though, is possible. When Bradley comes to ask what reality is, his answer is that it is experience, in a wide sense of that term: 'Feeling, thought and volition (any groups under which we class psychological phenomena) are all the material of existence, and there is no other material, actual or even possible' ([1897] 1930: 127) (see Idealism). His insouciantly brief argument for this challenges the reader to think otherwise without self-contradiction; he is more concerned to make clear that reality is not the experience of his individual mind, and that his doctrine is not solipsistic.

Bradley's criterion of reality is the absence of contradiction, and as he argues that the distinctions necessarily employed in judgment introduce contradiction, it follows that thought cannot capture reality. Nor can anything involving relatedness. Thus reality must be a non-relational unity, one which contains diversity because room must be found for appearances themselves. How? For illustration, Bradley appeals to a pre-conceptual state of immediate experience in which there are differences but no separations and from which the cognitive consciousness arises by imposing distinctions upon the differences. Reality, he thinks, is like this, except in transcending rather than falling short of thought, including everything in one comprehensive and harmonious whole (see Monism). Wollheim suggests that the best analogy for the relation of appearances to the Absolute is that of a painting (another might be an ecosystem): particular segments of the canvas would be falsified, even

made ugly, by abstraction from the whole to which they contribute, yet that character which makes for ugliness in isolation may nevertheless be itself beautiful in its surroundings and essential to the beauty of the whole. Bradley rejects the demand for detailed explanations of how phenomena like error are reconcilable with the Absolute, trying to shift the burden of proof to those who profess confidence in their incompatibility. His general answer is that anything that exists, even evil, is somehow real and thus belongs to the Absolute, which comprehends and transcends both good and evil (as well as religion): it is neither, but is further from the latter than the former.

Few philosophers have found Bradley's positive metaphysics persuasive. But it stands as a permanent challenge to the capacity of discursive thought to represent the world as it is in itself, a challenge posed by turning the mechanisms of thought upon themselves and demanding that they meet their own standards.

See also: Bosanquet, B.; Green, T.H.; Hegelianism §5; Herbart, J.F.; James, W.; Lotze, R.H.; Moore, G.E. §1-2

STEWART CANDLISH

List of works

Bradley's works are best consulted in the latest of the editions mentioned. These are the ones usually cited in recent discussion; they are also the most useful in that, while the earlier text is usually left intact, Bradley's later thoughts are added in the form of notes, appendices and essays. The library of Merton College, Oxford, holds Bradley's unpublished papers, notebooks and letters received. The Russell Archives at McMaster University contain letters from Bradley to Russell (some interesting extracts appear in *The Collected Papers of Bertrand Russell*, vol. 6, 349-53), and the John Rylands Library of the University of Manchester has letters from Bradley toSamuel Alexander.

Bradley, F.H. (1876, 1927) *Ethical Studies*, Oxford: Clarendon Press.(The second edition, published posthumously, contains unfinished notes by Bradley; the text is otherwise that of the first.)

Bradley, F.H. (1883, 1922) *The Principles of Logic*, London: Oxford University Press.(The second, revised edition adds extensive commentary and terminal essays to the text of the first. Out of print at the time of writing.)

Bradley, F.H. (1893, 1897) *Appearance and Reality*, London: Swan Sonnenschein.(The second edition contains a substantial and important Appendix. The most commonly encountered version now is the so-called 'ninth impression' of 1930, on which all subsequent printings have been based. This was published in Oxford at the Clarendon Press after the rights were secured from George Allen & Unwin, which as George Allen had taken over Swan Sonnenschein. This ninth impression is a textually corrected version of the second edition, but has a completely different pagination. Out of print at the time of writing.)

Bradley, F.H. (1914) *Essays on Truth and Reality*, Oxford: Clarendon Press.(Mostly published in *Mind* from 1899 onwards, these are major essays containing developments of Bradley's logic and metaphysics, incisive criticism of William James and Bertrand Russell, and reflections on religion. Out of print at the time of writing.)

Bradley, F.H. (1930) *Aphorisms*, Oxford: privately printed at the Clarendon press; Bristol: Thoemmes Press, 1993, facsimile edn.(The modern edition is bound together with 'The Presuppositions of Critical History' and contains an introduction to the latter by Guy Stock.)

Bradley, F.H. (1935) *Collected Essays*, Oxford: Clarendon Press.(Contains the two pamphlets 'The Presuppositions of Critical History' (1874) and 'Mr Sidgwick's Hedonism' (1877) as well as the valuable unfinished essay 'Relations' and a good bibliography of Bradley's published works. Out of print at the time of writing.)

Bradley, F.H. (1994) *Writings on Logic and Metaphysics*, ed. and with intros by J.W. Allard and G. Stock, Oxford: Oxford University Press.(Contains well-chosen extracts from the *Logic*, *Appearance and Reality* and *Essays on Truth and Reality*. The helpful introductions are both general and topic-specific; this is a very useful edition for undergraduates.)

References and further reading

Bradley Studies (1995-), Harris Manchester College, Oxford.(This journal 'aims to publish critical and scholarly articles on philosophical issues arising from Bradley's writings and from those of related authors' and 'to include each year an ongoing list of what has been published on Bradley and related themes'.)

Campbell, C.A. (1931) *Scepticism and Construction: Bradley's Sceptical Principle as the Basis of Constructive Philosophy*, London: George Allen & Unwin.(One of the clearest pictures of Bradley's metaphysics from

within his own tradition. Critical but sympathetic. Out of print.)

Candlish, S. (1978) 'Bradley on My Station and Its Duties', *Australasian Journal of Philosophy* 56 (2): 155-70.(A critical exposition of the place of Bradley's most famous essay in his moral thought, relating it also to his metaphysics. Suitable for undergraduates.)

Candlish, S. (1989) 'The Truth about F.H. Bradley', *Mind*, 98 (391): 331-48.(On analytic philosophy's distorted picture of Bradley, and his theory of truth; intelligible to advanced undergraduates.)

Horstmann, R.-P. (1984) *Ontologie und Relationen*, Königstein: Athenäum.(A thorough discussion, in German, of the treatment of relations by Bradley, Hegel and Russell, concluding in favour of Hegel. A work for scholars.)

Ingardia, R. (ed.) (1991) *Bradley: A Research Bibliography*, Bowling Green, OH: Philosophy Documentation Center.(Indexes primary sources, books, articles, dissertations, book reviews. Littered with trivial errors, but very comprehensive. Many of the articles attributed to Cresswell are by Crossley.)

Mander, W.J. (1994) *An Introduction to Bradley's Metaphysics*, Oxford: Clarendon Press.(Intended for serious beginners, but also of scholarly interest. Draws on such reassessments of Bradley as those in Manser and Stock and covers Bradley's best known ideas and arguments.)

Mander, W.J. (ed.) (1996) *Perspectives on the Logic and Metaphysics of F.H. Bradley*, Bristol: Thoemmes Press. (Papers from the 1993 F.H. Bradley Colloquium, accessible to advanced undergraduates.)

Manser, A. (1983) *Bradley's Logic*, Oxford: Blackwell.(A study of *The Principles of Logic* with particular attention to its historical background, which argues that the original text of 1883 is free of the later metaphysics and unfairly neglected. Not difficult, but presupposes knowledge of twentieth century analytic philosophy. Out of print.)

Manser, A. and Stock, G. (eds) (1984) *The Philosophy of F.H. Bradley*, Oxford: Clarendon Press.(A well-reviewed collection of essays, covering the full range of Bradley's work, with a useful scene-setting editorial introduction. Suitable for advanced undergraduates. Reprinted in paperback in 1986.)

Nicholson, P. (1990) *The Political Philosophy of the British Idealists: Selected Studies*, Cambridge: Cambridge University Press.(Study I is an accurate and sympathetic but somewhat uncritical account of *Ethical Studies*, readable by beginners. There is an extensive bibliography.)

Passmore, J. (1969) 'Russell and Bradley', in R. Brown and C.D. Rollins (eds), *Contemporary Philosophy in Australia*, London: George Allen & Unwin.(One of the most perceptive comparisons of the two philosophers, straightforwardly written and free of technicality. Out of print.)

Sprigge, T.L.S. (1993) *James and Bradley: American Truth and British Reality*, Chicago and La Salle, IL: Open Court.(A substantial critical comparison of the work of Bradley and William James, which pays more than usual attention to Bradley's views on religion, has a detailed summary of the philosophers' correspondence and gives chronologies listing their works by year of appearance. Clearly written but very long.)

Stock, G. (ed.) (1997) *Appearance versus Reality*, Oxford: Clarendon Press.(Originating in the F.H. Bradley Colloquium 1993, a collection of major articles on the metaphysics and logic, plus other material, accessible to advanced undergraduates.)

Taylor, A.E. (1924-5) 'Francis Herbert Bradley, 1846-1924', *Proceedings of the British Academy* 11 (2): 458-68. (The standard source for biographical information, some of which was obtained with family help.)

Wollheim, R. (1956) 'F.H. Bradley', in Ayer, A.J. *et al.*, *The Revolution in Philosophy*, London: Macmillan, 12-25.(Perhaps the best short depiction of central themes in Bradley's logic and metaphysics, written with great clarity and insight; accessible to complete beginners. Out of print.)

Wollheim, R. (1969) *F.H. Bradley*, Harmondsworth: Penguin, 2nd edn.(For many years the standard text on Bradley. Accessible to undergraduates. Admirably clear and free of jargon, but should not be regarded as reliable. Out of print.)

Bradwardine, Thomas (*c.*1300-49)

Thomas Bradwardine was a leading figure in fourteenth-century philosophy and theology from 1328, when he completed De proportionibus velocitatum in motibus (On the Ratios of Velocities in Motions), until his death in 1349, shortly after becoming Archbishop of Canterbury. His theory of ratios of velocities in motions was an important reinterpretation of Aristotle and was influential throughout Europe. The author of numerous mathematical and logical works, Bradwardine helped to initiate a style of natural philosophical analysis using a standard set of logical and mathematical tools. On the Continent, Nicole Oresme, Albert of Saxony and many others wrote works on the ratios of velocities in motions following Bradwardine's lead. In his De futura contingentibus (On Future Contingents) and De causa Dei Pelagium (On the Cause of God Against the Pelagians), Bradwardine staked positions emphasizing the symmetry of God's omniscience with respect to past, present and future.

1 Life and works

Born around 1300, Bradwardine is known for his association with Merton College, Oxford, and it was probably while he was at Merton that he composed the treatises *De insolubilibus (On Insolubles) De incipit et desinit (On 'It Begins' and 'It Ceases')*, *De proportionibus velocitatum in motibus (On the Ratios of Velocities in Motions)*, *De continuo (On the Continuum)*, *Arithmetica speculativa (Speculative Arithmetic)*, *Geometria speculativa (Speculative Geometry)* and possibly other logical works. In approximately 1332-3 Bradwardine, as a bachelor of theology, lectured on Peter Lombard's *Sentences*. Only fragments of these lectures remain, probably including what is known separately as *De futuribus contingentibus (On Future Contingents)*. In 1335 Bradwardine became a member of the group of scholars surrounding Richard de Bury, Bishop of Durham. He became a Master of Theology in the late 1330s, was named chancellor of Saint Paul's Cathedral in London in 1337, chaplain to Edward III in 1339 and Archbishop of Canterbury in June 1349. The *De causa Dei contra Pelagium et de virtute causarum ad suos Mertonenses (In Defence of God Against the Pelagians and On the Power of Causes)*, begun at Oxford, was completed in London probably in 1344 (see Augustinianism).

2 Theory of ratios in motions

Within the arts faculty at Oxford, Bradwardine's greatest impact undoubtedly came from his *De proportionibus velocitatum in motibus*. Aristotle, in the *Physics*, had supposed that the velocities of motions are related to their causes in such a way that when the force causing a motion is multiplied, the velocity will be similarly multiplied. Likewise, Aristotle had supposed that when resistance to motion is multiplied, the velocity is, inversely, reduced in a like ratio. Aristotle's discussion of the relationships of forces, resistances and velocities served his purposes in arguing that a vacuum is impossible. If there were a vacuum, Aristotle argued, then an elemental heavy body falling in the vacuum would have no resistance, and hence an infinite velocity would result. Since an infinite velocity is a self-contradiction, a vacuum is impossible.

In *De proportionibus*, Bradwardine first laid out the theory of ratios familiar from the theory of musical ratios as found in Boethius. He used this understanding of operations on ratios to reinterpret Aristotle's theory. Velocities vary, he said, as the ratio of force to resistance. When the ratio of force to resistance is 'doubled' the velocity is doubled, when the ratio is 'tripled' the velocity is tripled and so on, with the understanding that 'double' the ratio 3:1 is the ratio 9:1 (its square), and that 'triple' the ratio 2:1 is the ratio 8:1 (its cube). This new understanding of Aristotle's theory had the advantage of being expressed in similar words while avoiding a serious weakness of the theory itself, namely that it could not explain the mathematical relationships of forces and resistances in very slow motions. Everyone agreed that for motion to occur at all the force must be greater than the resistance. Suppose, then, that the ratio 2:1 produces a certain velocity. It follows on Aristotle's theory that the ratio of 1:1 should produce half that velocity, but in fact it produces no velocity at all, since the force is not greater than the resistance. Thus Aristotle's theory cannot account for any velocity smaller than half the velocity produced by the ratio 2:1. By contrast, Bradwardine's function provided values of the ratio of force to resistance greater than 1:1 for any velocity down to zero, since any root of a ratio greater than 1:1 is always a ratio greater than 1:1.

Bradwardine's proposed new function won immediate acceptance in universities all over Europe, beginning at Oxford where the so-called Oxford Calculators took it up. It also inspired people to quantify motions in every way

possible. Bradwardine's function was assumed to hold for rotations as well as rectilinear motions, and for alterations, augmentations and diminutions as well as local motions. In Paris, Nicole Oresme and Albert of Saxony wrote treatises on ratios, building on Bradwardine's rule. Oresme used the theory to argue that since most ratios are incommensurable with one another, when ratios of ratios are understood in the Bradwardinian sense, the velocities of motion of the planets caused by ratios of force to resistance will also likely be incommensurable, meaning that planets will never return to exact conjunctions in the same location in the sky. Astrological predictions based on such repetitions of positions will therefore be impossible.

Taking the ratio of force to resistance as the measure of motion 'as if with respect to cause' (*tanquam penes causam*), philosophers then asked about the measure 'as if with respect to effect' (*tanquam penes effectum*). For instance, should a rotation be measured by the average velocity of the points of the rotating body or by the distance traversed by the fastest moved point, as Bradwardine had argued in the last book of *De proportionibus*? Learning how to operate with ratios as Bradwardine did became a standard part of the arts curriculum in many later medieval universities, and subsequently became part of the theology curriculum as well.

3 Other works

In contrast to the novelty of Bradwardine's *De proportionibus*, his *De continuo* argued for what were at the time well-established conclusions. As Aristotle had argued, no continuum is composed of indivisibles. Moreover, indivisibles such as points, lines, planes and instants, despite their use by mathematicians, do not exist in the real world (a position also taken by Ockham, Buridan and other fourteenth-century nominalists). However, although most scholastics believed that the composition of continua from indivisibles was inconsistent with both Euclidian geometry and Aristotelian physics, in the period just before Bradwardine no less a scholar than Henry of Harclay, Chancellor of Oxford, had been persuaded by theoretical considerations of God's knowledge that the points in a line must be immediate to each other. Harclay's defence of immediate indivisibles derived from the view proposed by Robert Grosseteste that God measures a line by intuiting the infinitely many mediate points contained within it. In Harclay's opinion, Grosseteste's view implied his own: if God intuits all the points in a line, Harclay argued, they must be immediate to each other.

Perhaps the most original aspect of Bradwardine's *De continuo* was his strategy for combatting indivisibilists (or 'atomists') of every kind. Instead of arguing in turn against each of the theories that posited the composition of continua from indivisibles, Bradwardine began in an axiomatic format and then argued against key theses combined in different ways by the different theories. Thus he argued first against immediately conjoined indivisibles, a thesis common to the Pythagoreans (see Pythagoreanism), Harclay and Walter Chatton. Then he argued against the composition of a continuum from finitely many indivisibles, a thesis maintained by the Pythagoreans and Chatton but not by Grosseteste or Harclay, proving his point by running through every known discipline or science. Turning next to the theory that a continuum is composed of infinitely many indivisibles, a view common to Harclay and Grosseteste, Bradwardine used both natural philosophical and mathematical arguments. A key conclusion in the axiomatic part of *De continuo* was that if one continuum has (infinitely or finitely many) immediate atoms, any continuum has them. This enabled Bradwardine to argue that if a line has immediate atoms, so do bodies, motion and time, and vice versa.

Despite the supposed reciprocity between conclusions in the different disciplines, natural philosophical arguments were fundamental to Bradwardine's project. What would happen to indivisible surfaces if a single body of water were divided in two or if two bodies of water were joined together? Would the division create new surfaces and the joining destroy them and, if so, what would cause such creation or destruction of indivisibles? It was in light of such physical considerations that Bradwardine in Part VIII of *De continuo* finally denied the existence of indivisibles altogether. Thus despite the prominence of mathematics in his work, Bradwardine remained an Aristotelian, implicitly assuming that natural philosophy gives a truer description of reality than mathematics.

Bradwardine's other mathematical and logical works do not seem to have been particularly notable. In *De insolubilibus*, Bradwardine may have been the first to suppose that insoluble propositions - for example, the liar paradox - imply not only their own falsity but also their own truth. In *De incipit et desinit*, Bradwardine not only restated standard views concerning first and last instants of permanent and successive things, but also raised questions about the relations of successive continua such as motion and time to future contingency. If, for instance, no motion has a last instant of being, then if a body is in motion at an instant, it will necessarily be true that it must

continue in motion in the future, since the instant at which it is in motion cannot be the last instant. However, the continuation of its motion is, presumably, contingent. How then can the continuation of the motion after the present instant of motion be both necessary and contingent?

Bradwardine's *De futuris contingentibus* and *De causa Dei contra Pelagium* were directed against contemporaries who had argued that there are real future contingents despite God's omniscience and who had adopted a Pelagian position with regard to human action, admitting the possibility that human merit and willpower could lead to salvation without God's freely given grace (see Pelagianism). William of Ockham was the most prominent target of Bradwardine's arguments in *De futuris contingentibus*, and his major opponent in *De causa Dei contra Pelagium* was most likely Thomas Buckingham, who modified his own position as a result of Bradwardine's views.

Like *De continuo*, *De causa Dei contra Pelagium*, a work of 876 printed folio pages, has an unusual format, combining strings of conclusions with prolific quotations from authoritative authors. Admitting that earlier in his life the Pelagian position had seemed more reasonable to him, Bradwardine emphasized instead the necessity of grace. The revival of a Pelagian position among Bradwardine's contemporaries seems to have arisen in part from considerations of God's freedom. Could God choose to allow some individuals to lead meritorious lives without divine grace? To say no would seem to limit God's liberty.

In response to such a line of thought, Bradwardine reaffirmed God's power, goodness and immutability. God is the first and most immediate cause of everything that happens, including the choices humans make by their free wills. Being the cause of everything that happens, God causes evil as well as good, but this is only relative evil and contributes to the overall good (see Evil, problem of). Absolutely central to Bradwardine's argument was the assertion that God is eternal and therefore immutable: God is not in time, and any implication that he might 'change his mind' or be affected by human choices after they occur is mistaken (see God, concepts of).

In *De consolatione philosophiae*, written in the early sixth century, Boethius had attempted to address questions of God's foreknowledge and human free will by arguing that God's timeless knowledge of the future does not determine the future any more than human observation determines what is seen. Using the ideas of logical possibility introduced into the discussion of free will by Duns Scotus - who had argued that for an act to be freely chosen it is not necessary that the individual making the choice ever actually make a different choice, only that it not be logically contradictory to have made a different choice - Bradwardine argued that God's relations to past, present and future are identical. God freely chooses how he designs the cosmos, though he does so only once. It is not logically contradictory that God could make a different choice, but he will not in fact do so since he is not in time. In a way beyond human comprehension, human free will and divine determination are said to be not inconsistent, but rather compatible with each other (see Eternity; Free will; Omnipotence).

See also: Augustinianism; Natural philosophy, medieval; Oxford Calculators

EDITH DUDLEY SYLLA

List of works

Bradwardine, Thomas (*c*.1322-5?) *De insolubilibus (On Insolubles)*, ed. M.L. Roure in 'La problématique des propositions insolubles du XIII[e] siècle et du début du XIV[e], suivie de l'édition des traités de William Shyreswood, Walter Burleigh et Thomas Bradwardine', *Archives d'histoire doctrinale et littéraire du moyen âge* 37, 1970: 205-326.(Covers solutions to the problem of propositions that seem to falsify themselves, such as 'I am telling a lie.')

Bradwardine, Thomas (*c*.1323-5) *De incipit et desinit (On 'It Begins' and 'It Ceases')*, ed. L.O. Nielsen, *Cahiers de l'Institut du moyen âge grec et latin* 42, 1982: 1-83.(Bradwardine's contribution to the common fourteenth-century problem of assigning intrinsic or extrinsic temporal limits to entities and/or processes.)

Bradwardine, Thomas (*c*.1323-5) *Geometria speculativa (Speculative Geometry)*, ed. and trans. A.G. Molland, *Thomas Bradwardine, Geometria Speculativa*, Wiesbaden: Franz Steiner Verlag, 1989.(For an analysis of the significance of this work, see A.G. Molland (1978) 'An Examination of Bradwardine's Geometry', *Archive for History of Exact Sciences* 19: 113-175.)

Bradwardine, Thomas (*c*.1323-5) *Arithmetica speculativa (Speculative Arithmetic)*, Paris, 1495. (Following this edition the work was reprinted many times.)

Bradwardine, Thomas (1328) *De proportionibus velocitatum in motibus (On the Ratios of Velocities in Motions)*,

ed. and trans. H.L. Crosby, Jr in *Thomas of Bradwardine: His Tractatus de Proportionibus*, Madison, WI: University of Wisconsin Press, 1955.(Bradwardine’s extremely influential exposition of one traditional interpretation of the Euclidean theory of operations on ratios, along with its application to the Aristotelian problem of the relations of forces, resistances and velocities in motions.)

(*c*.1328-35) *De continuo (On the Continuum)*, ed. J.E. Murdoch in ‘Geometry and the Continuum in the Fourteenth Century: A Philosophical Analysis of Thomas Bradwardine’s *Tractatus de continuo*’, Ph.D. thesis, University of Wisconsin, 1957. (For a description of this work, with Latin text of the conclusions, see J.E. Murdoch, ‘Thomas Bradwardine: Mathematics and Continuity in the Fourteenth Century’, in E. Grant and J.E. Murdoch (eds) *Mathematics and its Applications to Science and Natural Philosophy in the Middle Ages*, Cambridge: Cambridge University Press, 1984, 103-37.)

Bradwardine, Thomas (*c*.1332-3) *De futuris contingentibus (On Future Contingents)*, ed. J.-F. Genest, *Recherches augustiniennes* 14, 1979: 249-336.(See Genest (1992) for a description of what the issues were, with regard to future contingents, for Bradwardine and other fourteenth-century thinkers.)

Bradwardine, Thomas (*c*.1332-3) Commentary on the *Sentences* of Peter Lombard.(Previously thought to be lost, but some questions in MS Paris Bibl. Nat. lat. 15805, f. 40-49v, identified as coming from Bradwardine, are discussed in J.-F. Genest and K. Tachau, ‘La lecture de Thomas Bradwardine sur les Sentences’, *Archives d’histoire doctrinale et littéraire du moyen âge* 57, 1990: 301-6. The *De futuris contingentibus* may have been derived from this commentary.)

Bradwardine, Thomas (1344) *De causa Dei contra Pelagium et de virtute causarum ad suos Mertonenses, libri tres (In Defense of God Against the Pelagians and On the Power of Causes)*, ed. H. Savile, London: ex officina Nortoniana apud Ioannem Billium, 1618; repr. Frankfurt: Minerva, 1964.

References and further reading

Dolnikowski, E.W. (1995) *Thomas Bradwardine: A View of Time and a Vision of Eternity in Fourteenth-Century Thought*, Leiden: Brill.(Links Bradwardine’s mathematics, logic, natural philosophy and theology through his concept of time.)

Genest, J.-F. (1992) *Prédétermination et Liberté Crée à Oxford au XIVe Siècle. Buckingham contra Bradwardine (Predetermination and Created Liberty at Oxford in the Fourteenth Century: Buckingham against Bradwardine)*, Paris: Vrin.(Unravels existing confusions concerning Bradwardine’s position on the controverted issues of God’s predetermination and free will and edits questions of his main opponent, Thomas Buckingham.)

Knuuttila, S. (1982) ‘Modal Logic’, in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 342-57.(For understanding concepts of necessity, contingency and possibility presupposed by Bradwardine’s *De causa Dei*.)

Molland, A.G. (1996) ‘Addressing Ancient Authority: Thomas Bradwardine and Prisca Sapientia’, *Annals of Science* 53: 213-33.(Considers Bradwardine’s attitudes about authority and the relations of science and religion on the basis of *De causa Dei*.)

Murdoch, J.E. (1970) ‘Bradwardine, Thomas’, in *Dictionary of Scientific Biography*, vol. 2, New York: Charles Scribner’s Sons, 390-7.(A description of Bradwardine’s mathematical and scientific work, with references to the standard earlier books and articles on Bradwardine.)

Normore, C. (1982) ‘Future Contingents’, in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 358-81.(Essential contextual background for understanding Bradwardine on future contingents and on human free will.)

Oberman, H. and Weisheipl, J.A. (1958) ‘The ‘Sermo Epinicius’ ascribed to Thomas Bradwardine (1346)’, *Archives d’histoire doctrinale et littéraire du moyen âge* 26: 295-329.(Contains Weisheipl’s reconstruction of the time line of Bradwardine’s works, as well as the text of Bradwardine’s sermon.)

Sbrozi, M. (1990) ‘Metodo matematico e pensiero teologico nel "De causa Dei" di Thomas Bradwardine’ (Mathematical Method and Theological Thinking in *De causa Dei* of Thomas Bradwardine), *Studi Medievali*, serie terza, XXXI: 143-91.(Argues that the axiomatic form (*more geometrico*) of *De causa Dei* reflects Bradwardine’s view that God’s knowledge is atemporal, or in other words that God’s knowledge of future contingents does not differ in kind from God’s knowledge of past and present.)

Brahman

The Sanskrit word brahman *(neuter) emerged in late Vedic literature and Upaniṣads (900-300* BC*) as the name (never pluralized) of the divine reality pervading the universe, knowledge or experience of which is a person's supreme good. The word's earliest usage (often pluralized) is to refer to the verses of the oldest work in Sanskrit (and in any Indo-European language), the Ṛg Veda (*c.*1200* BC*), which is a compilation of poems and hymns to Indo-European gods. The individual verses of the poems are mantras (*brahmāṇi*), whose proper enunciation in the course of ritual and sacrifice was thought to secure various aims. Thematically, the Ṛg Veda and other early Indian literature presents a sense of pervasive divinity. Apparently through an assimilation of the idea of the magic of mantras to the divine immanence theme the word* brahman *assumed its later meaning. In any case, Brahman - the Absolute, the supremely real - became the focus of Indian spirituality and the centre of much metaphysics for almost three thousand years, down to the present day. In the Upaniṣads, which are mystic treatises containing speculation about Brahman's nature and relation to ourselves and the world, the central positions of Vedānta schools emerge, all of which are philosophies of Brahman. But not even in the narrow set of the earliest and most universally accepted Upaniṣads (numbering twelve or thirteen) is there expressed a consistent worldview. Important themes about Brahman may be identified, but there is no overall unity of conception, despite what later exegetes claim. The unity of the early Upaniṣads concerns the premier importance of mystical knowledge or awareness of Brahman (*brahma-vidyā*), not precisely what it is that is to be mystically known. The classical Indian philosophical schools of Vedānta systematized the thought of early Upaniṣads.*

1 The Vedas and Vedic literature

The poems of the *Ṛg Veda*, most but not all of which address gods of an Indo-European pantheon, were transmitted until modern times through the care of a priestly caste. A Vedic (later, Hindu) priest is known as a *brāhmaṇa* or even a *brahmán* (masculine, as opposed to *bráhman*, neuter). The apparent etymological connection between the word for the Absolute and that for the caste has provoked no little consternation and speculation in a sociological vein among modern scholars. The solution to the mystery appears to lie in the shamanic role of the poet-cum-priest in early Indo-European society. According to Vedic scholars, the poems of the *Ṛg Veda* seem the results of competitions in eloquence among seers (*kavis*) or shamans over generations, each poem conceived as inspired by a shaman's sense of 'occult correspondences between the sacred and profane' (Renou 1953: 10). As Vedic society became more settled in its practices, rituals and sacrifices became fixed, with shamans assuming the offices of priests and keepers (that is, memorizers) of the sacred hymns. Original composition ceased, and the *Ṛg* and three other Vedas assumed their canonical forms. Individual verses of the poems, referred to in the Vedas themselves as *bráhman* (neuter), became central to ritual performances, and the importance of the overall content - stories of the gods, identifications of correspondences among spheres of life, and so on - waned. The verses, known as 'mantras', were considered invested with magical power.

Vedic themes were not forgotten, however. Throughout the *Ṛg Veda*, a sense of pervasive divinity is expressed. Gods correspond to, or inhere in, natural forces, and there is a divine structure and rhythm to the universe as a whole. Moreover, although the gods hymned are plural, each when addressed is said to be supreme. At places, a kind of monotheism is evident: 'They have styled Him Indra [the Chief of the gods], Mitra [the Friend], Varuṇa [the Venerable], Agni [Fire], also the celestial, great-winged Garutmā; for although one, poets speak of Him diversely; they say Agni, Yama [Death]), and Mātarīśvan [Lord of breath]' (*Ṛg Veda* 1.164.46). This verse has been interpreted as an early expression of Indian henotheism: one God, that is, Brahman, takes various divine forms, and is worshipped in various forms by people according to personal proclivities. Thematically, the Vedas express a multifarious divinity and unitive spirituality important to most subsequent Indian religion.

In later Vedic literature - prose appendages called *brāhmaṇa* and *āraṇyaka* - the supreme force moving the gods and operative throughout the universe is designated for the first time *brahman* (neuter). The term conveys a sense of mystery, and distinct proposals are made about just what Brahman is (for example, wind, breath, the sun). Brahman is said to be *svayam-bhū*, 'self-existent', and is identified with Prajāpati, 'Father of Creatures'. The *Śatapatha Brāhmaṇa* (*c.*900 BC) declares at 10.2.3: 'In the beginning this universe was just Brahman; Brahman created these gods.'

2 Upaniṣads

Early Upaniṣads (from 800 to 300 BC) represent a break with the ritualism of later Vedic literature, and speculation about Brahman becomes decidedly more pronounced. Debates on metaphysical topics called *brahmôdya*, 'discourses on Brahman', held in the courts of kings and princes, are recorded. (A *brahmôdya* was apparently only a riddle contest in the earlier period.) In particular, the reasonings of the 'Brahman-knower' Yājñavalkya against various opponents in the court of King Janaka, as reported in the Bṛhadāraṇyaka Upaniṣad (*c*.800 BC), may be taken to usher in a new era of Indian thought, marked by abstraction and self-conscious argument.

However, the early Upaniṣads present at best variations on central views. No single coherent worldview is expressed, but rather arguments and proposals centred on Brahman and Brahman's relation to ourselves and the universe. Despite what later exegetes claim, there is no unity of theory. Later proponents of Upaniṣadic philosophies (called Vedāntins) look for an overall unity because they view the Upaniṣads as revealed texts.

Many Upaniṣadic passages are exploratory, playfully spinning stories with rather abstruse morals, or etymologies, usually false, of words with psychological meanings. Brahman is usually approached psychologically; the early Upaniṣads are predominantly mystical texts. The self as known in meditation is the medium for knowing the Absolute, the self's mysterious ground, which turns out to be the ground of the entire universe. As though announcing a discovery, some Upaniṣads proclaim the identity of the self and Brahman. Several rich psychological conceptions are worked out and asserted with a tone of confidence that contrasts with the tentativeness of statements about Brahman.

Nevertheless, there are at least ten themes about Brahman that reverberate throughout both early Upaniṣads themselves and later Vedāntic philosophy: (1) Brahman is self (*ātman*) and consciousness; (2) Brahman is world ground; (3) Brahman is transcendent of 'names and forms' (*nāma-rūpa*), that is, transcendent of finite individuality; (4) Brahman is unitary, the coincidence of opposites, and omnipresent; (5) Brahman has 'nondual' (*advaita*) self-awareness; (6) Brahman is the essence or finest part of everything; (7) Brahman is the locus of value, with awareness of Brahman as the 'supreme personal good' (*parama-puruṣārtha*) and 'liberation' (*mukti*) from fear and evil; (8) Brahman is mystically discoverable; (9) Brahman is beyond the power of thought; (10) Brahman is the creator and inner controller of all things.

It would be difficult, if not strictly impossible, to read the Upaniṣads uninfluenced by the centuries of later commentary and interpretation. The great debate among classical Vedāntins concerns the question of the theism of early Upaniṣads, or, more broadly, how Brahman relates to the world. The theistic interpretation - of Brahman as God creating a world of real particulars - is eschewed by the Advaita (Non-Dualist) school. Theistic Vedāntins cite Upaniṣadic passages stating that Brahman determines individual names and forms, *nāma-rūpa*. In their view, Brahman is a primordial Will and Controller - that is, 'God', *īśvara*. The dispute is irresolvable in the Upaniṣads themselves. Theme (10) is as pronounced as any, textually speaking. But Advaitin exegetes find a way of subordinating it to other themes. Conversely, Vedāntic theists interpret Upaniṣadic monism as underscoring their understanding of creation as emanation. Brahman as God looses forth (or manifests, *sṛjate*) the world out of God's own substance, as a spider's web is spun out of its own body.

The Advaita interpretation emphasizes themes (4) and (5), the unity and self-awareness of Brahman. At Bṛhadāraṇyaka 4.1.7, Yājñavalkya says, 'With only an awareness of *ātman*, self, should one meditate, for here [in the self] all these things become one.' The prominence of monism here - a spiritual monism in accordance with the idea (theme (1)) that Brahman is self or consciousness - provides grounds for interpreting individuality ('names and forms') as *mere* names and forms, as Advaita would do. The logic of the reasoning is not complex: if there is just one thing, how can there be many?

Moreover, in other Upaniṣadic passages, Brahman is explained through an analogy with dreaming, an analogy that fuels an Advaita assimilation of the theistic doctrine of creation to the idea of the One. Bṛhadāraṇyaka 4.3.9-10: '[Sleeping,] one takes along the stuff of this all-embracing universe [and] tears it apart himself [and] shapes it himself.… He is the all-maker, for he is the maker of everything.' Advaitins understand the 'He' as Brahman identical with *ātman*, who enjoys various states of himself - specifically, waking, dreaming and a state transcending both where the soul is 'aware only of its own light' (see also the Māṇḍūkya Upaniṣad). In the dream state, emanationism seems valid, and waking, says the Upaniṣad, is like dreaming. But states involving awareness

of objects other than the self are expressly declared to be less valuable than self-illumination. Thus a subordination of theist emanationist cosmology to an illusionism about diversity seems to be called for, too: this is the Advaita reading.

Vedāntic theists, for their part, propose a stratified view of reality, of Brahman, ranging from an essential Divine to material things. Passages such as Taittirīya Upaniṣad 2.1-, with its theory of five sheaths progressively manifesting an essential soul, and Chāndogya Upaniṣad 6.8.7-, with its proclamations that Brahman is an indwelling essence (theme (6)) can be read as supporting the stratified reading: there is an essential Brahman, both transcendent and dwelling in the heart of everything, who progressively manifests in this world. Physical things may be furthest from what God is in God's own nature, but, as Taittirīya 3.1.2 says, 'matter [too] is Brahman'. Perhaps the most significant theistic passage comes from the Bṛhadāraṇyaka (3.7.1-23): 'He who, dwelling in the earth, is other than the earth, whom the earth does not know, whose body the earth is, who controls the earth from within - is the Self, the Inner Controller, the Immortal' (this last refrain - verse 3 - is then repeated with substitutions). Finally, the Sanskrit word for God, *īśvara* and cognates (such as *īś*), appear in early Upaniṣads dozens of times as apparent synonyms of *brahman*.

3 Advaita Vedānta

The classical Indian philosophical schools of Vedānta systematized the thought of early Upaniṣads. These schools divide broadly into the psychological monism of Advaita Vedānta and theistic views of Brahman championed by non-Advaitins. According to Advaita, the central Upaniṣadic teaching is that Brahman is the self. The world, and God, are illusions of false consciousness. According to Indian theism, in contrast, Brahman is God, the real creator of a real universe. Despite this opposition, concessions regarding Brahman can be discerned in each camp.

Developments in thought about Brahman continue long past the early Upaniṣads to such modern philosophers as Aurobindo Ghose and Sarvepalli Radhakrishnan. Most contributions occur under the banner of the Advaita Vedānta school or a confederation of Indian theists. There is also speculation about Brahman as Speech (*śabda-brahman*) in the writings of Bhartṛhari. Both Advaita and Indian theism are intellectually long and complex movements, and we shall confine ourselves exclusively to the the most significant developments in the thought about Brahman.

Advaitins devote special attention to Brahman's nature and how it is possible for us to conceive of that nature on the one hand, and to the relation, or nonrelation, of Brahman to the world on the other. In the first case, the concern is principally to shore up a sense of the value of liberation. In the second, it is to defend a view of Brahman as an absolute unity admitting no differentiation whatsoever.

Advaitin thought about Brahman's nature appears to be soteriologically motivated; that is to say, a view of experience or realization of oneself as Brahman, considered the supreme goal of life, dictates the direction of its elaborations. Classical Advaita also asserts that nothing positive can be predicated of Brahman (themes (3) and (9)), since positive characterization is confined to differentiating finite things. But inasmuch as this view stands in tension with elaborations of Brahman's nature, it is usually conveniently ignored or muted through speculation, not very successful, about the power of metaphor and indirect indication (*upalakṣaṇa*). The soteriological need is overriding.

A stock characterization of Brahman as *saccidānanda* emerges at the centre of this project: Brahman as existent or existence (*sat*), as awareness (*cit*) and as bliss (*ānanda*: the compound *saccidānanda* is formed by euphonic combination). Brahman is in fact said to be the sole existent, the single reality; the sole consciousness, a single self of everyone; and a supreme bliss. Who would not want to realize this?

There is controversy in particular concerning how Brahman is bliss. Some insist that this is said only to indicate that Brahman is not subject to the hedonic content of our normal experience; others that this underscores the supreme value of personal realization of Brahman. In either case, a soteriological context is key. The spirit of the elaboration is not to spell out how Brahman underpins our everyday experience, how Brahman as existent underlies the existence of everything, how Brahman as conscious sustains consciousness, and the like. Advaitins when confronted with such interpretation typically retreat to their stance about metaphor and indirect indication, though some do say that the bliss of Brahman is (distantly) reflected in our finest moments of worldly pleasure and happiness.

Exegetically, conflict with theistic Vedāntins seems to have inspired the notion of *nirguṇa brahman* contrasting with *saguṇa brahman*, Brahman as 'without' and 'with qualities', though the distinction also dovetails with the Advaita type of negative theology. Brahman-without-qualities is supremely real; Brahman-with-qualities is talked about in scripture as a concession to obtuse minds. Scripture is like a patient teacher (*guru*), and it is difficult to appreciate that Brahman as supremely real has no qualities. Scripture talks about God - Brahman-with-qualities - as a preparation for the austere truth.

Advaitins face obvious difficulties in upholding the view that Brahman is an absolute unity admitting no differentiation. Since Brahman is the only reality, diversity has to be illusory. But how can what is illusory even appear? And how did our spiritual ignorance - or awareness of diversity (and not of Brahman) - originate? How could it possibly originate, given that Brahman, as understood by Advaita, is the sole reality? Much reflection is devoted throughout the long history of Advaita to everyday perceptual illusion, since this is the analogy used for Brahman's relation, or nonrelation, to the world. One Advaita camp also maintains that the world of appearance has structure, that it can be studied and explained in its own terms, though not in relation to Brahman, which is a reality of an entirely different order. Another, dialectical, camp eschews such a two-tiered view, arguing that the reality of Brahman means that no sense can be made of the appearance of diversity, even in its own terms. This camp presents a barrage of arguments against all known pluralist ontologies, particularly that of Nyāya-Vaiśeṣika.

4 Post-Upaniṣadic Indian theism

The understanding of Brahman as God - the Creator and Sustainer of the universe as both its stuff and a will that fundamentally shapes things - is developed in devotional texts of popular religion, such as the *Bhagavad Gītā(Song of God*, *c.*200 BC), as well as in polemical treatises by theistic philosophers. The philosophers are much occupied with explaining precisely how the individual soul is distinct from Brahman, and the sense, if any, in which the two may be said to be identical. Rāmānuja (eleventh century) works an analogy between a substance as quality-bearer and the qualities it bears: God bears souls as accidental qualities, and is the necessary support of their appearance. Other theists use other metaphors; that of the ocean (God) and waves (souls) is common. The question is shaped by the controversy with Advaita and the Advaita insistence that the soul and Brahman are one.

Theistic philosophers also tend to stress God's love for the soul, teaching, in accord with the *Bhagavad Gītā*, that the best way to mystical knowledge of Brahman is not meditation or asceticism (associated with Advaita, Yoga and Buddhism), but rather a corresponding *bhakti*, or love of God. The whole world is God's play (*līlā*), say theistic Vedāntins; and through love of God, and worship and devotion, we are eventually to realize this and find in every experience the embrace of the Divine (see Madhva). Indian theists from the classical age into the modern conceive of a supreme personal good as a spiritual act of love-making.

The *Bhagavad Gītā* uses a Vedic motif, developed by later theists, to explain the process of emanation: sacrifice. Brahman sacrifices its infinity in becoming finite, and thereby creates the world. Several theists trace a process of contraction through gods and goddesses and earthly incarnations (*avatāra*) of God, through humans and animals and down to rocks and dust. Through sacrifice, Brahman emanates the world as its body. The *Bhagavad Gītā* teaches that through a reverse sacrifice of offering of the finite, a soul finds Brahman transcendent, the supreme good.

See also: Cosmology and cosmogony, Indian theories of §1; God, Indian conceptions of; Monism, Indian; Pantheism §§1-3; Śaōkara; Self, Indian theories of §§1-2, 7; Vedānta

STEPHEN H. PHILLIPS

References and further reading

Bhagavad Gītā (200 BC-AD 200, disputed), trans. F. Edgerton, Harvard Oriental Series 38-9, Cambridge, MA: Harvard University Press, 1944, paperback repr. 1972. (Though Edgerton's interpretive essay - orginally volume 39 in the HOS - too much reflects outworn assumptions of nineteenth-century indologists, his translation is excellent, faithful and elegant. There are, however, dozens of acceptable translations into English and other modern languages.)

Dasgupta, S. (1922-55) *A History of Indian Philosophy*, Cambridge: Cambridge University Press, 5 vols.(The latter volumes trace the history of Indian theism; volume 1 presents the core of Advaita as well as views of the early Upaniṣads and the Vedas; volume 2 presents Advaita in greater detail and much of its history.)

Phillips, S.H. (1995) *Classical Indian Metaphysics: Refutations of Realism and the Emergence of 'New Logic'*, La Salle, IL: Open Court.(Contains a section on the early Upaniṣads that is the model for the discussion of Upaniṣadic views of Brahman here.)

Potter, K.H. (ed.) (1981) *Encyclopedia of Indian Philosophies*, vol. 3, *Advaita Vedānta*, Princeton, NJ: Princeton University Press.(Contains an excellent introduction to Advaita philosophy, as well as summaries of works in the early Advaita school.)

Renou, L. (1953) *Religions of Ancient India*, New York: Schocken.(A study of Vedic religion and society written by the great Vedic scholar for nonspecialists.)

Ṛg Veda (*c.*1200 BC), trans. K.F. Geldner, *Der Rig-Veda*, Harvard Oriental Series 33-6, Cambridge, MA: Harvard University Press, 1951.(A translation commonly cited by scholars.)

Thieme, P. (1952) 'Brahman', *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 102: 91-129.(A watershed study of the semantic development of the word *brahman*, a topic with which much late nineteenth- and early twentieth-century scholarship was preoccupied.)

Upaniṣads (800-300 BC), trans. P. Olivelle, *Upaniṣads*, Oxford: Oxford University Press, 1996.(A new translation that may well be the best; readable and accurate. The translations in the present entry are by Stephen Phillips.)

Warrier, A.G.K. (1977) *God in Advaita*, Simla: Indian Institute of Advanced Study. (Shows that Śaōkara, the great Advaitin, takes seriously the Upaniṣadic teaching of *saguṇa brahman* - that is, God.)

Brahmo Samaj

The Brahmo (or Brahma) Samaj (‘Society of Brahma’) is the name of a theistic society founded by Raja Rammohun Roy in Calcutta in 1828. It advocated reform, and eventually abolition, of the traditional caste system, as well as legislation aimed at improving the social status of women and greater protection of children. Also dedicated to Hindu religious reform, the Brahmo Samaj stressed a monotheistic doctrine with a policy of tolerance and respect for all major religions of the world. The society split into two factions in 1866, largely over the issue of the speed of reform. Another split occurred in 1878 over whether the society’s constitution was to be fully democratic. The democratic wing, called the Sadhāraṇ Brahmo Samaj (‘Universal Brahma Society’), is still active in India.

1 Rammohun Roy

The early history of the Brahmo (or Brahma) Samaj is closely tied to the life of its founder, Rammohun Roy. Born in the village of Radhanagar in the Hoogly district of Bengal on May 22, 1772 into an orthodox Brahmanical family, Rammohun Roy (Rāmamohan Raya) worked for the East India Company from 1803 to 1814. He received in his boyhood the traditional education of the country and soon attained remarkable proficiency in Arabic, Persian and Sanskrit. Later in life he learned Greek, Latin and Hebrew. The study of Islamic theology shook his faith in the popular polytheistic and idolatrous forms of Hindu worship, and made him a lifelong admirer of the uncompromising monotheism of Islam. His subsequent acquaintance with the Upaniṣads, the *Brahmasūtra* and the *Bhagavad Gītā* convinced him that the concept of the unity of the Godhead constitutes the essence of Hinduism. He also came to have profound respect for the moral precepts of Jesus Christ.

Like many of his contemporaries, Rammohun Roy saw Indian society in the nineteenth century as caught in a vicious web created by religious superstition and social obscurantism. Hinduism, as Max Weber observed in his celebrated sociological study of Indian religion (1916), had become a compound of magic, animism and superstition. One of the features of society that Roy found most distressing was the socially inferior position of women. The birth of a girl was usually unwelcome; her marriage became a financial burden for her parents, who customarily had to pay the husband’s family a dowry; and, if her husband died before her, her widowhood was seen as inauspicious. Because of these attitudes, attempts to kill girl infants at birth were not unusual. At the other end of life, it was also not unusual for widows to burn themselves alive on the funeral pyres of their deceased husbands, a practice known as *satī*, which literally means ‘good woman’. Roy described this as ‘murder according to every religious text’.

Another debilitating factor in Hindu society was the system of hereditary caste, which sought to maintain hierarchical social segregation on the basis of ritual status. Rammohun Roy came to be convinced that the rules and regulations of caste hampered social mobility, fostered social divisions and sapped individual initiative. Above all other injustices of the caste system was the humiliation of untouchability, which militated against human dignity. There were innumerable other practices in the Hindu society of his day that Roy saw as marked by arbitrary constraints, credulity, uncritical acceptance of authority, bigotry and blind fatalism. Rejecting them as features of a decadent society, Roy’s Brahmo Samaj (‘Society of Brahma’), established in Calcutta in 1828, sought to create a social climate for modernization.

Rammohun Roy was fully alive to the challenge that had come to India in the form of Western civilization and felt strongly the need for a new philosophy which would, without sacrificing the genuine spiritual heritage of India, absorb and assimilate the modernism imported from the West. He warmly advocated the introduction of Western science and technology in the educational curriculum and became a pioneer of English education and socially progressive journalism in India.

Rammohun Roy and his Brahmo Samaj advocated the emancipation of women. They opposed polygamy, indentured servitude and the custom of *satī*. It was due to Roy’s efforts that William Bentinck, then Governor General of India, passed the famous regulation number XVII declaring *satī* a criminal offence. The Brahmo Samaj also opposed child marriage and the rigidity of the caste system, and it supported the remarriage of widows, equal rights for men and women, and the right of daughters to inherit property.

The creed of the Brahmo Samaj was declared to be universal monotheism. Its worship could be joined by anyone irrespective of their religious affiliations. Hence, according to Brahmo Samaj doctrine, there is no divine incarnation and no priestly mediation. Nature, earth and heaven were all said to have been created by a single God. Since there was no need for the mediation of priests, there was to be no hereditary priestly class and no performance of sacrifices. The Brahmo Samaj laid an emphasis on universal love for all human beings, irrespective of race or creed. Despite the universalistic features of the Brahmo Samaj, Roy gave a decidedly Hindu character to the form of worship within it. Worship at this early stage consisted of readings from the Vedas and the Upaniṣads, a sermon and devotional music.

2 Debendranath Tagore and Keshub Chunder Sen

Next to give organizational machinery to the Brahmo Samaj was Debendranath Tagore (Devendranatha Thakura, 1817-1905). He established the Tattvabodhini Sabha in 1839 with the objective of propagating Brahmo *dharma*. Until 1866 he remained the accredited leader of the Calcutta Brahmo Samaj and carried forward the best tradition of the days of Rammohun Roy. A doctrinal change that occurred in Brahmoism during this time was the abandonment of the belief in the infallibility of Hindu scriptures. Tagore laid more pronounced emphasis on *bhakti*, or devotion, in his exposition of religious texts. Under his inspiring leadership the Brahmo Samaj played a distinguished role in sponsoring social reforms such as the remarriage of widows, spreading education and vigorously opposing the efforts of Christian missionaries to gain converts from the ranks of Hindus.

The next phase of the Brahmo movement is dominated by Keshub Chunder Sen (Keśavacandra Sena, 1838-84), who joined the Samaj in 1857 and became Tagore's right-hand man. Having imbibed more of Western culture and Christian influence, Sen advocated a much more aggressive programme than Tagore, the latter favouring a slow and cautious approach to social reform. The two radically different temperaments led to a parting of the ways and paved the way for the first schism in the society. In 1866 Sen established the Brahmo Samaj of India (Akhila-Bhārata Brahma-samāja, 'Pan-Indian Brahma Society'), after withdrawing along with his supporters from the parent body, which henceforth came to be known as Ādi Brahmo Samaj ('The Original Brahmo Samaj').

Sen's Brahmo Samaj of India adopted a much more radical and comprehensive scheme of social reform than before, including programmes for female education and emancipation, and the complete abolition of caste distinction. Its activities led to the formation of the Indian Reform Association in 1870 and the enactment of the Native Marriage Act of 1872, which fixed the marriageable age of girls and boys at 14 and 18 respectively.

Probably owing to Christian influence, the doctrine of the Brahmo Samaj of India placed much more emphasis on the sense of sin, the spirit of repentance and the efficacy of prayer. The scriptures of Hinduism, Islam, Christianity and Buddhism were studied with great respect. An infusion of devotional fervour into Brahmoism made it a practical religious culture. Sen had a sympathetic and respectful attitude towards all religions and proclaimed a comprehensive synthesis of religions under the title of the Nava Vidhāna ('New Dispensation') in 1880. The New Dispensation called for faith in a living God, and explained that the several religions of the world are but varying interpretations of this one God. Though diverse and fragmentary, the religions of the world were seen as mutually complementary rather than exclusive.

The second schism of the Brahmo Samaj occurred in 1878, when a band of Sen's followers demanded the introduction of a democratic constitution for the Samaj, which was not conceded. To make matters more complicated, Sen's daughter had been married in violation of the provisions of the Native Marriage Act of 1872. The dissenters formed the Sadhāraṇ Brahmo Samaj ('Universal Brahma Society'), which took shape under the leadership of Anand Mohan Bose (Ānandamohan Basu) and framed a democratic constitution based on a universal adult franchise. It declared in 1882 that it was about to establish a worldwide republic. This new body has proved up until now the most powerful and active branch of the Brahmo Samaj of India.

See also: Arya Samaj; Vedānta §4

K.S. KUMAR

References and further reading

Chakrabarti, S.K. (1991) *Role of Brahmo Samaj in the History of Bengal*, Calcutta: Barnali.(Describes the historical background of the origin of the Brahmo Samaj movement in Bengal.)

Dasgupta, B.N. (1980) *The Life and Times of Rajah Rammohun Roy*, New Delhi: Ambika.(A good biography of Raja Rammohun Roy, the founder of the Brahmo Samaj.)

Kopf, D. (1979) *The Brahmo Samaj and the Shaping of the Modern Indian Mind*, Princeton, NJ: Princeton University Press.(Illustrates the impact of the Brahmo Samaj on the people of India after independence.)

Roy, Raja Rammohun (1906) *The English Works of Raja Rammohun Roy*, ed. J.C. Ghose, Allahabad: Panini Office, 4 vols.(A collection of some of Roy's administrative, political, religious and philosophical writing.)

Roy, Raja Rammohun (1977) *Selected works of Raja Rammohun Roy*, New Delhi: Government of India Ministry of Information and Broadcasting.(A collection of articles by Roy.)

Sen, K.C. (1980) *Keshub Chunder Sen in England: Diaries, Sermons, Addresses and Epistles*, Calcutta: Writers Workshop.(A monograph containing the diary kept by the founder of the Brahmo Samaj of India during a visit to England, along with some of his public addresses.)

Weber, M. (1916) *The Religion of India: The Sociology of Hinduism and Buddhism*, trans. and ed. H.H. Gerth and D. Martindale, Glencoe, IL: Free Press, 1958.(One of a series of works dealing with the economic ethics of the world's religions.)

Brazil, philosophy in

It is possible to distinguish between European philosophy in Brazil and Brazilian philosophy. The former refers to Brazilians who participate in discussions of issues occurring in the European philosophic tradition without any reference to Brazilian reality and its problems; the latter to those Brazilian intellectuals who respond to the problems growing out of situations which have confronted the nation historically whether their philosophical orientations have originated in Europe or elsewhere. This entry focuses on the latter and generally follows a historical progression. This progression spans from the precabralian Tupi-Guarani speaking societies of eastern South America to the healthy development of Brazilian philosophy since 1950 after the founding of the Institute of Brazilian Philosophy.

1 Precabralian Tupi-Guarani worldview

Pedro Álvarez Cabral was a Portuguese navigator who accidentally discovered the Brazilian coast in 1500. A nascent precabralian civilization extended along eastern South America which still exists in so far as a common set of assumptions, attitudes and values is operative among extant Tupi-Guarani speaking societies. When the Portuguese arrived they found a lingua franca uniting these people. A comparative study of myths of creation and destruction of the earth reveals a creative spirit, in the very act of creativity, emerging out of fundamental chaos in which irrational forces of annihilation struggle against any emergent deity with the capability to dispel darkness and nihility - 'he-who-creates-his-own-body-out-of-primeval-darkness' (Gillette Sturm 1991) - sustaining himself in the process, while simultaneously creating the heavens and the earth through creative wisdom. The creative process is intentionally dialectic, a dynamic struggle between opposites, involving feminine-masculine collaboration. The view of temporality is neither cyclical nor linear, but helical. Creation of language, love and sacred song lays the foundation for humanity, but the temporal process is marked by a tendency toward deterioration: tiring, aging, dying. An annual ceremony at planting time is performed to drive away malevolent forces which threaten the germination of seeds and growth of plants with death and destruction. From the earliest records there are reports of groups who have followed leaders who have predicted catastrophes and the end of the world, marching towards the Atlantic and serving as models for other messianic movements in the Brazilian interior.

2 Philosophy in sixteenth-century Portugal

The sixteenth century witnessed the most exciting period in Portuguese intellectual history marking a golden age of Portuguese philosophy. It was the time of the voyages of discovery when the tiny nation of Portugal established commercial outposts on the shores and islands of four continents and briefly became a world power in trade and commerce. Portuguese philosophers and intellectuals reflected the latest developments in European thought and also contributed actively to the formation of the mind of modern Europe.

Early in the century the Dutch humanist Desiderius Erasmus attracted a group of Portuguese disciples. In 1547 the king founded the Royal College of Humanities at the University of Coimbra and entrusted it to humanists, although within eight years its control was transferred to the Jesuits (see Collegium Conimbricense). None the less, northern European and Italian humanism made their impact on sixteenth-century Portugal, even in formulations of scholastic resurgence later in the century at Coimbra. Among leading Portuguese humanists during the sixteenth century were João de Barros, Damião de Gois, Francisco de Holanda, André de Resende and Francisco de Sá de Miranda.

The famous School of Sagres in southern Portugal, which provided training in navigational skills and the use of new navigational instruments, cannot be considered a research centre for experimental science. However, the publications of several Portuguese involved in navigation, or who had the opportunity for travel to foreign lands, point to a decided change in orientation concerning the study of natural science, as well as to the beginnings of experimental methodology in the effort to understand natural phenomena. Among the experientialists were Duarte Pacheco Pereira and Garcia de Orta. The experimentalists include Pedro Nunes and José de Castro. Worthy of mention also is the sceptic and forerunner of Cartesian methodology Francisco Sanches.

The instruments of the Counter-Reformation, including Inquisition and Index, became operative during the

sixteenth century and the Society of Jesus was given responsibility for the direction of Portuguese educational curricula in 1555. However, the resurgence of Scholasticism, which had one of its points of origin at the University of Coimbra, was not a retrenchment to earlier orthodoxy, but an advance in the great tradition. Those who moulded the Second Scholasticism at Coimbra hailed from both Portugal and Spain. These were Pedro da Fonseca, the ‘Portuguese Aristotle’, Luis de Molina and Francisco Suárez. They produced a new philosophy curriculum referred to as *Cursus Conimbricensis*.

3 Colonial Brazil: philosophical problems

Portugal never established a university in Brazil. Students were expected to matriculate at the Universities of Coimbra or Évora. Several monastic orders founded colleges and seminaries where prospective university students could pursue an undergraduate curriculum based on the approved model at Coimbra.

The presence of indigenous people posed several unique problems. Questions were posed regarding whether these people were human, what kind of humans they were and what rights they had to land and liberty. In 1548 when Portugal sent the first governor-general to consolidate its territories in Brazil the crown made clear that priority was to be given to Christian missions among the indigenous. Eleven years earlier a papal bull had declared the Indians to be fully human and proscribed deprivations of their freedom or possessions. In the light of royal regulation and the papal bull the question of the humanity of the Indians and their rights under natural law should not have been raised. Yet by 1557 Manuel da Nóbrega, director of Jesuit missions, published his *Dialogue concerning the Conversion of the Gentiles*, in which the central question was ’Do these have a soul like us?’ It also dealt with issues of slavery, land sovereignty, nomadism, cannibalism, differences in ‘nations’ of Indians, social institutions and ethics. Both sides of the debate are represented equally with standard arguments taken from scripture and classic Greek and Roman philosophy. In the end the equality of all persons is declared. The nature/nurture distinction is emphasized: Indians do not differ essentially from other humans; the differences lie in the social milieu. Other Jesuits who concerned themselves with new issues were José de Anchieta and Antônio Vieira.

4 Adaptations of African ideas and values

In their defence of the Brazilian Indians the Jesuits recommended the use of African slaves. This led to the introduction of cultural traditions differing from both the indigenous and the European. Traditional societies in subSaharan Africa are characterized by a strong sense of integrated community extended to elements in the nonhuman environment. The soil upon which the society was established by the ancestors becomes a focal point uniting the generations as both burial-place and source of fertility which sustains the living and guarantees continuity of community. The slave trade forced Africans into permanent exile from their homeland and alienated them from their communities. This threatened human survival through a rootlessness which made existence impossible unless a process of transplantation to new soil and re-establishment of community occurred which again made possible contact with one’s ancestral homeland and historical social heritage. The achievement of this end included large-scale slave revolts and the founding of *quilombos*, autonomous communities of runaway slaves in the interior, coronation ceremonies for African chieftains in which historic traditions were incorporated with reaffirmation of allegiance to a sovereign representing the continuity of community; the invocation of nature and ancestral spirits in possession ceremonies.

After 1950 a body of literature emerged giving intellectual articulation to the worldview implicit in the practices of the African-Brazilian community. An important issue is the extent to which the African past remains essential to the community. José Ribeiro de Souza speaks for the ‘re-Africanization’ movement and works through the Institute for Afro-Brazilian Studies for the preservation in Brazil of west African languages. Alfred Costa Moura calls for total de-Africanization. W.W. da Matta-Silva takes a middle position in which African roots constitute an essential base for a universal intellectual and religious position.

5 Enlightenment thought in eighteenth-century Brazil

Academies, arcadias and literary societies based on Graeco-Roman models and encouraging revival of classical Arcadian literary style appeared in southern Europe during the Renaissance. Early in the eighteenth century they were transformed from centres of debate on literary questions articulated in the baroque style of Gongorism into

conduits for reception, diffusion and discussion of Enlightenment ideas. They were private associations selecting their own members and determining their own objectives and procedures. In 1724 the Brazilian Academy of the ‘forgotten ones’ was founded to produce an interdisciplinary history of Brazil based on observation and primary documentation. The Scientific Academy of Rio de Janeiro discussed questions of physics, chemistry, botany, medicine, surgery and agriculture, in addition to maintaining a botanical garden and carrying on active correspondence with the Swedish Royal Academy of Sciences. Many similar societies appeared throughout Brazil which gradually added sociopolitical topics to the agenda.

Agitation for educational reform began in Portugal early in the century, leading to the decision to change the Coimbra curriculum drastically, giving responsibility for Portuguese education to the Order of the Oratory instead of the Society of Jesus. Philosophy was no longer handmaid to theology; physics was liberated from Aristotelian philosophy; natural science was divided into autonomous scientific disciplines stressing laboratory research. Students returning to Brazil from studies in Europe brought these new ideas and procedures with them, leading to radical changes in secondary and collegiate education.

Brazilian authors addressing Enlightenment issues included Matias Aires Ramos da Silva (1700-63), whose book *Problems of Civil Architecture* (*c*.1740) was an exposition of the latest developments in the natural sciences of physics and chemistry; Tomaz Antônio Gonzaga (1744-1810), whose *Treatise on Natural Law* (*c*.1790) was a critical exposition of new directions in natural law theory; Nino Marques Pereira (1652-1735), whose novel depicting a journey from Salvador to São Paulo ridicules the professors of the Second Scholasticism; José Joaquim de Cunha de Azeredo Coutinho (1742-1821), economic theorist who argued for physiocrat principles and was founder of a model college focusing on scientific studies and experimentation emphasizing practical application to the Brazilian situation and Francisco Luís Leal (1740-1820), author of the first history of philosophy in Portuguese which distinguished between the ‘ancients’ and the ‘illustrious moderns’.

6 Philosophical ambiance at Independence

The first articulate stirrings towards independence from Portugal began with the Minas conspiracy of 1789, the 1798 revolt of the tailors in Bahia and the Pernambuco Insurrection of 1817. Ideological roots are found in the literary and scientific academies, the curricular reform at Coimbra and the lodges of freemasonry. Five intellectuals, alive at the time of the declaration of independence by Pedro I, are representative of the philosophical spectrum.

José da Silva Lisboa, Viscount of Cairu (1756-1835) was a political conservative, out of sympathy with the French Revolution, who favoured retaining the United Kingdom of Portugal and Brazil, which had been instituted in 1815. He was a disciple of Adam Smith and advocated the lifting of all restrictions imposed by policies of feudalism and mercantilism in favour of the *laissez-faire* principle of British economic liberalism.

José Bonifacio de Andrada e Silva (1763-1835), known as the Patriarch of Independence, returned to Brazil in 1819 as a member of the leading scientific academies of the European continent after spending thirty-six years there as a geologist and metallurgist. The first grand master of the Grand Orient of Brazilian Freemasons, he believed in the rational structure of the universe and the guarantee of rights of all humans by natural law. However, his insistence that Brazil was not yet ready for complete social transformation did not endear him to the republicans.

Silvestre Pinheiro Ferreira (1769-1846) supported the United Kingdom of Portugal and Brazil, which had been established in 1815 and the moderative power of the monarch. Fundamentally a Lockean, he was disturbed none the less by the failure of the empiricists to provide a complete system to support scientific and political theory and attempted to derive such a system through a reinterpretation of Aristotelian categories (see Aristotle §7;Locke, J.).

Frei Joaquim do Amor Divino e Caneca (1774-1825) was a Mason imbued with the spirit of French encyclopedism (see Diderot, D. §1.). He spoke for the nascent social revolutionary movement and was concerned especially with guaranteeing rights for the lower classes of Brazilian society.

Diogo Antônio Feijó (1784-1843) was a priest-philosopher and politician. He was also the leader of a small group of Kantians in São Paulo who broke with the stale ideas of French encyclopedism and the sensualism of Condillac, aware that Kant provided a radically new direction in philosophy (see Condillac, E.B. De; Kant, I.).

7 Philosophy during the Empire: eclecticism and positivism

Diverse motives had impelled the leaders of movements for independence. Some favoured a new economic order, but with a continuation of political ties with Portugal. Others urged political independence without changing the monarchical form of government. Voices were heard favouring republicanism. Some called for transformation of the socioeconomic order. This diversity of viewpoints, combined with Brazil's geographical immensity and a partial slave economy, caused some to predict that Brazil would become fragmented into several autonomous states. Leadership concerned with maintaining unity called for conciliation in the face of the divisiveness existing during the first decades of the empire.

Francisco José de Carvalho Mont'Alverne (1784-1858) is credited with directing Brazilian thought during the Regency and the period of 'conciliation' along the lines of Victor Cousin's eclectic spiritualism (see Cousin, V.). As H.-A. Taine put it, eclecticism proposed a peace treaty to all systems, just as representative government provided a forum to satisfy all elements of society. Its popularity in Brazil introduced intellectuals to the European philosophic scene in a panoramic, if vague, view. The political realm was also well served, because eclecticism provided an ideological basis for moderates. Eclecticism appealed especially to medical scientists concerned with issues of physiological psychology, such as Eduardo Ferreira Franca (1809-57), professor of medical chemistry in Bahia and José Maria de Morais Valle (1824-86), professor of medicine in Rio de Janeiro. In that vein are the works of Domingos José Gonçalves de Magalhães (1811-82), *Facts of the Human Spirit* (1858) and *Soul and Brain* (1876).

By the mid-nineteenth century Brazilian students of mathematics, physics and medicine were encountering Auguste Comte in France and Comtean positivism began to replace eclecticism in scientific circles. A series of doctoral dissertations began to appear in the 1850s written from a Comtean perspective. In 1876 Benjamin Constant (1833-91), professor of mathematics at the Military School and Luís Pereira Barreto (1840-1923), author of *As Três Filosofias* (1874) founded the Brazilian Positivist Association. The association was committed to realizing Comtean sociopolitical values. Two years later when Miguel Lemos (1854-1917) and Raimundo Teixeira Mendes (1855-1927) returned from France, the positivist Apostolate of Brazil was established. Comtean positivism had profound influence on the Brazilian intelligentsia, but was fragmented into scientific, political and religious factions (see Positivist thought in Latin America §5).

8 Reactions to positivism: School of Recife, spiritualism and modernism

The first reaction against Comtean positivism took the form of a polemic between Comteans in Rio and adherents of Spencerian evolutionary naturalism in Recife (see Spencer, H.). Silvio Romero (1851-1914), historian of Brazilian philosophy and literature, recounted the polemic in *Doctrine Against Doctrine: Evolutionism and Positivism in the Republic* (1894). Along with Clovis Bevilaqua (1859-1944), Fausto Cardoso (1864-1906), Tito Lívio de Castro (1864-90) and Artur Orlando (1858-1916), he is identified with the School of Recife, centred in the Law School of Recife and headed by Tobias Barreto de Meneses (1849-89). Tobias Barreto had abandoned eclecticism because of its vagueness and Comtean positivism because of its rejection of metaphysics. Spencerianism was followed by an enthusiastic espousal of 'Germanism' and the development of a naturalistic monism inspired by Noiré.

Two of Tobias's students broke with him as they moved in different directions, but both strongly influenced the birth of Brazilian modernism. José Pereira da Graça Aranha (1868-1931) articulated a unique metaphysical position of monism based on aesthetics and mysticism. Raimundo de Farias Brito (1862-1917), impelled by a sense of bankruptcy in European culture, spent much time critically examining the development of modern philosophic and scientific thought, and formulating a 'philosophy of spirit' (see Anti-positivist thought in Latin America).

9 Brazilian philosophy since 1950

Philosophic activity has been especially productive since the 1950s. Much credit goes to the founding in 1950 of the Institute of Brazilian Philosophy, a truly inclusive national association, and the *Revista Brasileira de Filosofia (Review of Brazilian Philosophy)*, which began publication in 1951, founded by Miguel Reale. Through national and regional congresses and an active programme of adult education the Institute of Brazilian Philosophy has

fostered much philosophical dialogue. Reale has been a controversial figure and a firm believer in intellectual and political pluralism. He is known especially for his tridimensional theory of law and his philosophical position of ’ontognosiology’. He is also one of the founders of the Interamerican Society of Philosophy and the International Association for Philosophy of Law and Social Philosophy.

A concern for study of Brazilian philosophy is seen in the work of João Cruz Costa (1962) and Antônio Paim (1967; 1983) and several series of publications of the classic texts of the tradition. Official interest in the utility of philosophy in determining national policy was evident in President Juscelino Kubitshek’s establishment of the Advanced Institute of Brazilian Studies (ISEB) as part of his administration. Under the direction of the philosopher Roland Cavalcanti de Albuquerque Corbisier, one of the divisions was concerned with philosophical issues and involved collaboration with Álvaro Vieira Pinto and Hélio Jaguaribe.

Brazilian philosophers who have established an international reputation in their fields include Leônidas Hegenberg and Newton Carneiro Alfonso da Costa who have been associated with the Centre for Logic, Epistemology and History of Science at the Federal University of Campinas, Romano Galeffi, founder of the Brazilian Centre for Studies in Aesthetics, Henrique C. de Lima Vaz, Leonardo and Clovis Boff, Hugo Assman, Dom Helder Câmara, Rubem Alves, theologians of liberation (see Liberation theology), Paulo Freire, educational theorist, Leôncio Basbaum, Caio Prado Junior, J. Chasin, Marxists, Stanislavs Ladusãns, Catholic philosopher and Gerd Bornheim, phenomenologist.

Antônio Paim (1983) has pointed to a group of Brazilian philosophers whose concerns are similar enough to constitute a movement with roots in Brazil’s intellectual past and a commitment to a pluralistic, dialogic approach to philosophical issues with a concern for cultural context. Paim calls this the Culturalist School and mentions as prime movers, *inter alia*, Miguel Reale, Djacir Menezes and Luís Washington Vita.

See also: Latin America, colonial thought in; Cultural identity

FRED GILLETTE STURM

References and further reading

Crippa, A. (ed.) (1978) *As idéias filosóficas no Brasil (Philosophical Ideas in Brazil)*, São Paulo: Convívio, 3 vols. (A survey of philosophical history in Brazil from the perspective of a leading Catholic thinker. Vol. 1 covers the eighteenth and nineteenth centuries; vols 2 and 3 deal with the twentieth century, the former covering Marxists and Catholics, Farias Brito and Vicente Ferreira da Silva and the Culturalist movement, the latter dealing with natural law theorists, philosophy of education, aesthetics, logic and philosophy of science.)

Cruz Costa, J. (1962) *Panorama of the History of Philosophy in Brazil*, Washington, DC: Pan American Union. (A condensed version of the *History of Ideas in Brazil* (1964) commissioned as part of a series of brief introductions to philosophy in the Americas.)

Cruz Costa, J. (1964) A *History of Ideas in Brazil: the Development of Philosophy in Brazil and the Evolution of National History*, Berkeley, CA: University of California Press.(The leading historian of Brazilian philosophy relates national and societal development, covering colonial formation and its basis in Portuguese philosophy and culture.)

Gonçalves de Magalhães, D.S. (1858) *Facts of the Human Spirit*, Paris: Garnier.(Deals with eclecticism and concerns itself with issues of physiological psychology.)

Gonçalves de Magalhães, D.S. (1876) *Soul and Brain*, Rio de Janeiro: Garnier.(This book also hinges on eclecticism and provides an ideological basis for moderates.)

Gonzaga, T.A. (*c*.1790) *Treatise on Natural Law*, Rio de Janeiro: National Library of Lisbon, 1953.(An exposition of changing directions in natural law.)

Nóbrega, da M. (1557) *Dialogue concerning the Conversion of the Gentiles*, Lisbon: Comissão do IV Centenário da Fundação de São Paulo, 1954.(Important issues, such as slavery, land sovereignty and differences between ‘nations’ of Indians are dealt with.)

Paim, A. (1967) *História das idéias filosóficas no Brasil (History of Philosophical Ideas in Brazil)*, São Paulo: Editora Convívium, 3rd edn, 1984.(Paim addresses the relation between the history of philosophy in general and national philosophic traditions and considers the unique problematic confronting Brazilian philosophers.)

Paim, A. (1983) *Bibliografia filosófica brasileira, 1808-1985 (Bibliography of Brazilian Philosophy, 1808-1985)*, Salvador: Centro de documentação do pensamento brasileiro, 3 vols.(An account of all philosophical books published in Brazil from 1808 to 1985, including translations into Portuguese.)

Pereira Barreto, L. (1874) *As Três Filosofias (The Three Philosophies)*, part 1, *Theological Philosophy*, Rio de Janeiro: Laemmert; part 2, *Metaphysical Philosophy*, Jacarel: Comercial, 1876.(Published in two parts.)
Revista Brasileira de Filosofia (Review of Brazilian Philosophy) (1951-), São Paulo: Instituto Brasileira de Filosofia.(A quarterly journal.)
Romero, S. (1894) *Doctrine Against Doctrine: Evolutionism and Positivism in the Republic*, Rio de Janeiro: Tipografia de J.B. Nunes.(The polemic between Comteans in Rio and adherents of Spencerian evolutionary naturalism in Recife is recounted in this book.)
Sturm Gillette, F. (1991) 'Ontological Categories Implicit in the Mbyá-Guaraní Creation Myth', in M.H. Preuss (ed.) *Past, Present and Future: Selected Papers on Latin American Indian Literatures*, Culver City, CA: Labyrinthos, 117-22.(León Cadigen also transcribed the myth in *Guaraní en Ayvu Rapyta: míticos de los Mbyá-Guaraní del Guaira*, 1959.)

Brentano, Franz Clemens (1838-1917)

Brentano was a philosopher and psychologist who taught at the Universities of Würzburg and Vienna. He made significant contributions to almost every branch of philosophy, notably psychology and philosophy of mind, ontology, ethics and the philosophy of language. He also published several books on the history of philosophy, especially Aristotle, and contended that philosophy proceeds in cycles of advance and decline. He is best known for reintroducing the scholastic concept of intentionality into philosophy and proclaiming it as the characteristic mark of the mental. His teachings, especially those on what he called descriptive psychology, influenced the phenomenological movement in the twentieth century, but because of his concern for precise statement and his sensitivity to the dangers of the undisciplined use of philosophical language, his work also bears affinities to analytic philosophy. His anti-speculative conception of philosophy as a rigorous discipline was furthered by his many brilliant students. Late in life Brentano's philosophy radically changed: he advocated a sparse ontology of physical and mental things (reism), coupled with a linguistic fictionalism stating that all language purportedly referring to non-things can be replaced by language referring only to things.

1 Life and intellectual development

Franz Brentano came from a talented family. He mastered ancient and scholastic philosophy as well as the works of Comte and the British empiricists. He always prized Aristotle above other philosophers, and regarded German Idealism as the nadir of philosophy. Brentano's vision of philosophy as an exact discipline sharing its true method with natural science inspired his many famous students, who included such noted philosophers as Anton Marty,Carl Stumpf, Alexius Meinong, Christian von Ehrenfels, Edmund Husserl, and Kazimierz Twardowski, as well as later political leaders German Chancellor Georg Hertling and the first President of Czechoslovakia, Thomas G. Masaryk.

Brentano's life was one of controversy and disappointment. In 1864 he became a Roman Catholic priest and played an important role in the discussion of the proposed doctrine of papal infallibility. In a position paper commissioned by the Bishops of Germany he recommended that they reject the doctrine. When the doctrine was officially proclaimed, Brentano felt justified in giving in to other doubts that had been tormenting him, concluding, for example, that the doctrine of the Trinity was contradictory. In 1873 he resigned from the priesthood, the church, and his position in Würzburg. His career as a Professor in Vienna was short-lived. A legal controversy surrounding the question whether ex-priests could marry led him to resign his Professorship in 1880, a position to which, in a *cause célèbre*, he was never reinstated, remaining an instructor (*Privatdozent*) until he left Austria for Italy in 1895. In later years he became blind and was estranged from several of his older students. Brentano did not relish publication: apart from the uncompleted *Psychology* he published mainly short papers, lectures, and monographs on the history of philosophy. His views underwent continuous revision, but a major change occurred during the first decade of the twentieth century, when his ontological views went through a 'Copernican revolution' whose results even his closest followers found difficulty in accepting. Large quantities of letters, lecture notes and dictated pieces remained unpublished at his death. Many of these were edited from Prague between the wars, with support fromMasaryk.

2 Psychology

Brentano's interest in psychology dated from his early occupation with the work of Aristotle and the British empiricists. *Psychologie vom empirischen Standpunkt* (*Psychology from an Empirical Standpoint*) (1874) helped establish psychology as an independent discipline. Following Comte in deliberately eschewing metaphysical controversy,Brentano, though believing in the soul, determined psychology as the science not of the soul but of mental *phenomena*. He writes: 'All the data of our consciousness are divided into two great classes: the class of physical and the class of mental phenomena.' Brentano initially restricted scientific investigation to phenomena or appearances and regarded the assumption that there are things in themselves as very uncertain. Physical phenomena are those sense-objects (for example, colours, sounds, odours) that we experience whenever we have a sensation or an imagined or dreamed counterpart of a sensation.

Comte had held inner observation to be impossible, since it would require us to split ourselves mentally in two. Brentano countered that inner perception is possible, because every mental act is accompanied by a secondary

awareness of itself. Inner perception and memory form the solid experiential basis of psychology. Brentano was concerned in the *Psychology* to establish a proper taxonomy of mental acts. Following Descartes, he divides them into three classes: ideas (*Vorstellungen*), judgments and a third class comprising emotions, feelings, desires and acts of will, variously called interests or phenomena of love and hate. Ideas merely present something, judgments accept as existent or reject as nonexistent something presented, while interests take a pro- or con-attitude to something judged. Thus interests presuppose judgments and these in their turn presuppose ideas. So all mental phenomena are either ideas or are founded on ideas. Brentano's classification has not been widely accepted.

From the late 1880sBrentano divided psychology into descriptive psychology, which he also sometimes called 'phenomenology' (see Phenomenological movement), and genetic psychology. The former is an a priori, philosophical discipline concerned with the basic elements of consciousness and their modes of structural combination, resting on the certain evidence of inner perception. The latter is a posteriori, empirical and probabilistic, concerned with the causal laws governing how mental phenomena arise and perish and the connections between the mental and the physiological.

3 Intentionality

In the search for a positive criterion marking off mental from physical phenomenaBrentano revived the Aristotelian-Thomistic conception of intentional inexistence in the *Psychology*. Every mental phenomenon is intentional, contains an object in itself, or no physical phenomenon does. In seeing something is seen, in judging something is accepted or rejected, in hoping something is hoped for and so on. Hence psychology cannot be reduced to or replaced by a physical science. Brentano maintained this demarcation principle but modified his conception of intentionality in a more epistemologically realist direction. A middle phase finds him rejecting the immanent conception of objects, accepting that the object of intentionality may be outside consciousness, and may be real, or an *ens rationis*, like a state of affairs, or even nonexistent. This middle view was most influential among his students.

Brentano's student Twardowski went on to distinguish the mental acts themselves from their subjects, from their objects and their contents. Your seeing Jones would be, according to Twardowski's account, a mental act that has you as its subject, Jones as its object and a visual image as its content. Every mental act then has an object, but not all such objects exist or are real. This view influenced Meinong's theory of objects (see Meinong, A. §§2-4; Twardowski, K. §3).

However, Brentano himself later rejected all non-real and nonexistent objects (see §4 below). Since someone who judges (correctly) that unicorns do not exist seems to stand in a rejecting relation to something nonexistent, Brentano came to regard intentionality as not a relation but as merely relation-like, the similarity being that we think of relations in a similar way. In thinking of John touching Mary, we think directly of John and indirectly of Mary, but she too exists. In thinking of Mary rejecting unicorns we think directly of the unicorn-rejecting subject Mary and only indirectly of what she rejects. There are no non-real objects; they are fictions engendered by a careless use of language.

Brentano emphasizes that, in knowing that one is in a mental state, one knows directly and immediately that there is a certain individual thing - namely, the one who is in that state. And *you*, of course, are the one who is in your mental state. We may single out three different phases of this situation: (1) I can know that I hope for rain; (2) as a rational being, I can conceive what it is to hope for rain; and (3) that the only type of entity that can have the property of hoping for rain is an individual thing or substance.

Following Leibniz, Brentano distinguishes two types of certainty: the certainty we can have with respect to the existence of our conscious states, and the a priori certainty that may be directed upon axioms and other necessary truths. These two types may be combined in a significant way. At a given moment, I may be certain, on the basis of inner perception, that there is believing, desiring, hoping and fearing; and I may also be certain a priori that there cannot be believing, desiring, hoping and fearing unless there is a substance that believes, desires, hopes and fears. In such a case, it will be certain for me (Brentano says that I 'perceive') that there is a substance that believes, desires, hopes and fears. It is also axiomatic,Brentano says, that if one is certain that a given substance exists, then one is identical with that substance.

4 Types of being: substance, accident and boundary

InBrentano's earlier writings he assumed that in addition to concrete individual things or *entia realia*, there are also 'non-things' or *entia irrealia,* falling into several different categories. They include: (1) 'intentionally inexistent' objects, such as that devil that is supposed to 'exist in mind' when one thinks about a devil; (2) those objects, often called 'propositions', thought to be designated by 'that'-clauses or sentential gerundives (as in 'that there are living beings on Venus', or 'there being living things on Venus'); (3) properties or universals, such as redness and being-both-round-and-square; and (4) states and events.

Brentano's rejection of such 'non-things' was based upon two general principles. First, the only objects that we can think about, strictly speaking, are concrete individual things. We do think of concrete things; 'think of' is univocal; therefore whatever we think of is a concrete thing. 'Concrete' does not here imply materiality. God and the soul, according to Brentano, are concrete but non-material. Second, an adequate understanding of language will show us that any plausible statement that ostensibly refers to such non-things may be construed as pertaining only to concrete things. Brentano defends the second principle with considerable ingenuity.

The theory of 'concrete predication' tells us that all of our thoughts may be adequately expressed by using concrete terms in place of predicates. We may best understand this theory if we imagine that all predicates in our language have been replaced by terms, and consider a view of the world which might naturally suggest itself to us if our language were in fact of such a sort, as it very well could have been. 'Concrete predication' may recall whatAristotle had said of simple judgments: 'An "affirmation" is a statement affirming something of something; a "negation" is a statement denying something of something'. In simple affirmative judgments, Aristotle said, we combine things; and in simple negative judgments we separate them. Since such statements as 'Mary is thinking' are for us a natural way of describing people, one might suppose that, when a person is thinking, they stand in a certain relation to the property of thinking. But if we say 'Some persons are *thinkers*' (Brentano uses '*ein Denkendes*', which one might translate more literally as 'a thinking-thing') then it may be more natural to suppose that we are describing not a relation between an individual thing and a property, but a relation between those individual things that are persons and those individual things that are thinkers. What, then, is the relevant relation between persons and thinkers?

Brentano uses Aristotle's term and says that the thinker is an 'accident' of its subject.Aristotle had said, however, that accidents are 'in' their subjects.Brentano saw that there was a reason for reversing this, claiming that accidents *contain* their subjects. The relation of substance to accident is similar in fundamental respects to that of part to whole. Brentano tells us what may seem surprising at first, that a substance is a part of its accidents. An accidental determination of a substance is a larger whole which is not necessary to its substance. If I happen to be thinking at the moment, I am a substance which, for now, is a thinker. The thinker is, at the moment, an accident of that substance which is identical with me. An accident is something containing its substance but not necessary to the substance.

A substance is *necessary to* its accidents. The accidents could not exist unless the subject exists - just as, according to the principle of 'mereological essentialism', every whole is ontologically dependent upon its parts. But no accident is necessary to its subject; hence the use of the word 'accident'. The expression 'X is an accident of Y' does not then introduce any unfamiliar concept into ontology, but refers only to a more generalized version of the part-whole concept.

In addition to distinguishing those individuals that are accidents from those that are substances, Brentano points out that there are individuals of still another sort. These include the points, lines and surfaces that constitute the inner and outer 'boundaries' of spatial objects. Boundaries exhibit a unique type of ontological dependence. Every boundary is necessarily such that it is a boundary of some constituent of the thing of which it *is* a boundary; but this is not to say that, for any thing of which it is a boundary, the boundary is necessarily such that it is a constituent of it. The point is of fundamental importance to ontology and the theory of categories, and may be put more precisely. Every boundary *x* is necessarily such that it is a boundary of (and therefore a constituent of) some constituent *y* of the entity *z* of which it happens to be a boundary. But from this fact, it does not follow that there is a constituent *w* of *z* which is such that *x* is necessarily a constituent of *w*. For example, a spatial point P is necessarily such that it is a constituent of a constituent C of the line L of which P happens to be a constituent. But, for any constituent C of L that you might mention, P need not be a constituent of C; some smaller constituent of C

would do just as well. Analogously, this holds for the relation of lines to surfaces and for the relation of surfaces to three-dimensional bodies.

By contemplating the concept of a boundary in this way, Brentano claimed, we acquire the concept of infinite divisibility by means of which we can then understand what it is for an entity to be ‘continuous’. The general problem of the continuum occupied Brentano throughout his philosophical career.

5 Truth

Like Aristotle, Brentano holds that the primary sense of ‘true’ is its application to beliefs or judgments; a secondary sense is its application to sentences. This theory of truth is properly called a ‘doxastic theory’. Brentano originally upheld a version of the correspondence theory of truth, but since his mature ontology has no place for such entities as facts or propositions, his final theory is not a correspondence theory as it is usually understood (see Truth, correspondence theory of §§1-2).

By reflecting upon judgments that are certain, according to Brentano, we obtain the concept of the ‘correctness’ of judgment. Once we have this concept, we are able to extend it beyond the sphere of the certain and thereby derive the broader concept of truth. We may speak, then, of a strict sense and of an extended sense of ‘correct’. In its strict sense, ‘correct’ will mean the same as ‘certain’. (Actually, Brentano uses ‘evident’ in place of ‘certain’; but he uses it in the way in which most epistemologists now use ‘certain’.) If, however, we use ‘correct’ in its extended sense, then we may also say that a judgment is correct if it agrees with a judgment that is certain - that is, if it accepts what a judge who has certainty would accept or rejects what such a judge would reject. In *Wahrheit und Evidenz (The True and the Evident)*, Brentano defines truth thus:

> Truth pertains to the judgment of one who judges correctly - one who judges about a thing in the way in which a person who judged with evidence would judge about it; it pertains to the judgment of one who asserts what the person who judges with evidence would assert.
>
> (1930, 139; 1966, 122)

The concept of correctness also plays an essential role in Brentano’s moral philosophy and general theory of value.

6 Ethics and the theory of value

Brentano’s moral philosophy is based upon his theory of value. His theory of value, in turn, is a theory of ‘intrinsic value’, about what is good in itself, or good as an end, about what is bad in itself, or bad as an end, and more generally about the relation of intrinsic preferability.Brentano’s theory of intrinsic value is based upon the analogy he believes to hold between intellectual and emotive phenomena. One may take an intellectual stand towards an object of thought, thereby having a positive belief or a negative belief about that object; one may ‘affirm’ the object or ‘deny’ it. One may also take an emotive stand with respect to the object, in which case one ‘loves’ the object or one ‘hates’ it. Brentano uses ‘love’ and ‘hate’ somewhat broadly to cover what we might call, respectively, ‘pro-feelings’ and ‘anti-feelings’.

What does it mean to say that a thing is intrinsically good or intrinsically bad? Brentano appeals again to the concept of correctness which is central to his theory of knowledge and truth. To say that a thing is intrinsically good, according to Brentano, is to say that it is ‘correct to love’ that thing as an end; and to say that a thing is intrinsically bad is to say that it is ‘correct to hate’ it as an end. Brentano believed that we can be immediately aware of the correctness of certain of our emotive attitudes, just as we can be immediately aware of the correctness (that is, the truth) of certain of our intellectual attitudes. In each case, the correctness consists in a relation of appropriateness or fittingness between the attitude and its object.

Brentano constructs a hierarchy of values in which pleasure plays a subordinate role. The principal bearers of intrinsic value, he claims, are conscious states. Some but not all conscious states are intrinsically good. And every conscious state *as* a conscious state contains some part that is intrinsically good. Every correct judgment is intrinsically good; so too is every correct emotion and every enrichment of our intellectual life. Since every intrinsic evil is itself a conscious state, there can be no intrinsic evil that does not include some intrinsic good. Among the things that are ‘predominantly bad’ are error, pain, every unjustified act of hate (especially the hatred of that which is good), and every unjustified act of love (especially of that which is bad). Sensory pleasure is, as

such, intrinsically good.

A complete account ofBrentano’s value theory would take into consideration a detailed analysis of aesthetic value, and also Brentano’s views on instrumental or practical value, which are broadly utilitarian.

7 Philosophical theology

Brentano’s philosophical theology depends upon a combination of the traditional arguments from motion and from contingency (see God, arguments for the existence of §1). He attempts to show that there can be no uncaused events and that the hypothesis according to which there is just one necessary substance, upon which all contingent things depend, has a probability approaching certainty. He combines this proof with an appeal to the evidence for design that we find when we contemplate ourselves and other living beings. Any sound epistemology, Brentano believes, will concede that there is such evidence. He then argues that the necessary substance is personal in that it is an intelligent being having both intellectual and emotive consciousness. It is also an immaterial being and is not a subject of accidents.

What distinguishes Brentano’s conception of God from that of most theologians is the thesis that God, like everything else, is a temporal being. Brentano describes an instance of God’s temporal consciousness this way:

> He now knows, for example, that I am writing down these thoughts. Yesterday, however, he did not know this, but rather that I will write them down later. And similarly he will know tomorrow that I have written them down.

(1976: 105; 1988: 87)

8 Historiography of philosophy

Brentano published five monographs on Aristotle, and several of his works on the history of philosophy have been published posthumously. His most characteristic view is that philosophy goes through cycles of advance and decline in four phases: a constructive theoretical phase, then in successively greater decline a practical phase, a sceptical phase and finally a mystical phase. The cycle then repeats itself. Brentano discerned three such cycles from the pre-Socratics to the German idealists, one each in ancient, medieval and modern philosophy. He clearly saw himself as ushering in a fourth constructive phase, a view to which one may give assent. Despite its obvious simplifications, the model is better than most others in accommodating philosophy’s uncertain progress.

See also: Intentionality; Psychology, theories of §§1-2

RODERICK M. CHISHOLM
PETER SIMONS

List of works

Brentano, F. (1862) *Von der mannigfachen Bedeutung des Seienden nach Aristoteles*, Freiburg im Breisgau: Herder; repr. Darmstadt, Wissenschaftliche Buchgesellschaft, 1960; trans. and ed. R. George, *On the Several Senses of Being in Aristotle*, Berkeley, CA: University of California Press, 1975.(Brentano’s Berlin doctoral dissertation: a catalogue of the ways Aristotle says something may be said to be; includes a classification and justification of the categories.)

Brentano, F. (1874) *Psychologie vom empirischen Standpunkte*, Leipzig: Duncker && Humblot; 2nd edn, ed. O. Kraus, Leipzig: Meiner, 1924; repr. Hamburg: Meiner, 1973; trans. A.C. Rancurello, D.B. Terrell and L.L. McAlister and ed. L.L. McAlister, *Psychology from an Empirical Standpoint*, London, Routledge, 1969; 2nd edn, 1995. (Brentano’s best-known and most influential work, though incomplete. Contains a defence of psychology as an autonomous discipline, the famous delimitation of its subject matter, mental phenomena, via intentional existence, and a threefold classification of mental phenomena into ideas, judgments, and phenomena of love and hate (interests), which include both feeling and will.)

Brentano, F. (1889) *Vom Ursprung sittlicher Erkenntnis*, Leipzig: Duncker & Humblot; 4th edn, ed. O. Kraus, Hamburg: Meiner, 1955; trans. R.M. Chisholm and E.H. Schneewind, *The Origin of our Knowledge of Right and Wrong*, London: Routledge, 1969.(Brentano’s ethics in a small compass, based on a theory of intrinsic value defined in terms of correct interests and preferences, supplemented by a utilitarian account of instrumental value.)

Brentano, F. (1926) *Die vier Phasen der Philosophie und ihr augenblicklicher Stand (The Four Phases of*

Philosophy and Its Present Status), Leipzig: Meiner; repr. Hamburg: Meiner, 1968.(Brentano's theory of the cyclical rise and decline of philosophy.)

Brentano, F. (1929) *Vom Dasein Gottes*, ed. A. Kastil, Leipzig: Meiner; repr. Hamburg: Meiner, 1968; trans. S. Krantz, *On the Existence of God*, Dordrecht, Nijhoff, 1987. (Posthumously published treatise setting out Brentano's probabilistic proof for the existence of God.)

Brentano, F. (1930) *Wahrheit und Evidenz*, ed. O. Kraus, Leipzig: Meiner; repr. Hamburg: Meiner, 1974; trans. R.M. Chisholm, I. Politzer and K.R. Fischer, and ed. R.M. Chisholm, *The True and the Evident*, London: Routledge, 1966.(A chronologically arranged series of writings on truth, beginning with Brentano's somewhat half-hearted espousal of a correspondence theory and ending with its rejection in favour of the view that a true (correct) judgment is one which someone judging with evidence would accept.)

Brentano, F. (1933) *Kategorienlehre*, ed. A. Kastil, Leipzig: Meiner; repr. Hamburg: Meiner, 1974; trans. R.M. Chisholm and N. Guterman, *The Theory of Categories*, Dordrecht: Nijhoff, 1981.(A miscellaneous collection of late writings on categories and ontology, setting out Brentano's opposition to Aristotle and his economical ontology of reism. Discusses meanings of 'be', substance and accident, relations and linguistic fictions.)

Brentano, F. (1952) *Grundlegung und Aufbau der Ethik*, Hamburg: Meiner; ed. F. Mayer-Hillebrand, trans. E.H. Schneewind, *The Foundation and Construction of Ethics*, London: Routledge & Kegan Paul, 1973.(Compiled from the notesBrentano made for his enormously popular lectures on practical philosophy in Vienna, 1876-94.)

Brentano, F. (1976) *Philosophische Untersuchungen zu Raum, Zeit und Kontinuum*, selected from the unpublished papers by A. Kastil, ed. S. Körner and R.M. Chisholm, Hamburg: Meiner; trans. B. Smith, *Philosophical Investigations on Space, Time, and the Continuum*, London: Croom Helm, 1988.(Subject matter as the title describes: contains Brentano's account of time-consciousness, his anti-Cantorian theory of the continuum and his unorthodox views on boundaries.)

Brentano, F. (1982) *Deskriptive Psychologie*, ed. R.M. Chisholm and W. Baumgartner, Hamburg: Meiner; trans. B. Müller, *Descriptive Psychology*, London: Routledge, 1995.(Edited from Vienna lecture notes: psychology fromBrentano's influential middle period, when he had distinguished descriptive psychology (phenomenology) from genetic psychology.)

References and further reading

Baumgartner, W., Burkard, F.-P. and Wiedman, F. (eds) (1988-) *Brentano Studien*, Dettelbach: Röll.(An international yearbook devoted toBrentano. Essays in English and German.)

Chisholm, R.M. (1982) *Brentano and Meinong Studies*, Amsterdam: Rodopi.(Contains essays on several aspects ofBrentano's philosophy.)

(1986) *Brentano and Intrinsic Value*, Cambridge: Cambridge University Press.(Introductory account of Brentano's views, with discussion of hierarchy of values, organic unities and evil.)

Chisholm, R.M. and Haller, R. (eds) (1978) *Die Philosophie Franz Brentanos (The Philosophy of Franz Brentano)*, Amsterdam: Rodopi; also in *Grazer Philosophische Studien* 5.(Nineteen essays on various aspects of Brentano's philosophy.)

(eds) (1987) *The Descriptive Psychology of the Brentano School*, special issue of *Topoi* 6 (1).(Seven essays on Brentano's philosophy.)

Kastil, A. (1951) *Die Philosophie Franz Brentanos, Eine Einführung in seine Lehre (The Philosophy of Franz Brentano, An Introduction to his Teaching)*, Munich: Lehnen.(A paraphrase of Brentano's opinions by the person who probably knew them best, still a useful all-round introduction, but lacking scholarly apparatus.)

McAlister, L.L. (ed.) (1976) *The Philosophy of Brentano*, London: Duckworth.(An informative collection including personal reminiscences by Stumpf and Husserl.)

Bridgman, Percy William (1882-1961)

Bridgman founded high-pressure experimental physics and was committed to a classical empiricist view of science - a view challenged by twentieth-century developments in relativistic and quantum mechanics. He argued that developments in special relativity showed the experimental operations scientists performed were suitable substitutes for basic constituents of matter, thus founding operationalism, a methodological position which influenced logical positivism and, transformed beyond his recognition, was expropriated by the behaviourist school in the social sciences. As Bridgman grappled with the challenges of general relativity and quantum mechanics, he increasingly parted company with his positivistic and behaviourist followers by moving more towards subjectivist views of science and knowledge. These later views led him to see and explore intimate connections between foundations of scientific knowledge and human freedom.

Bridgman developed a leak-proof seal for pressure-pistons which enabled him to obtain far higher pressures than his predecessors. He discovered allotropic forms such as ‘hot ice’ - solid water at $80°$ - and abrupt phase transitions wherein materials went from solid to liquid without an intermediary plastic stage. Existing theory neither predicted nor explained such phenomena, and so experimental work proceeded in a theoretical vacuum. This reinforced Bridgman’s classical empiricist belief that observation of, and experimentation on, reality was the basis of all science, not theorizing.

Bridgman construed Einstein’s 1905 special relativity theory (see Einstein, A. §2; Relativity theory, philosophical significance of; Spacetime), as showing that Newtonian space and time had no objective standing, thereby abolishing the absolute dichotomy between matter and motion and calling into question the ontological status of the quantities experimental science measured. He wrestled with the relativistic challenge in the 1914-26 ‘dimensional analysis’ disputes over the ontological status of measured quantities in physics, coming to the following conclusions: the quantities of physics are the products measurement produces according to a particular set of operations. Primary qualities are merely those regarded as fundamental. The dimensions of a primary quantity have no absolute significance, being merely reflections of the rules of operation in measurement. Since equations express equality of measured numbers, they too have no absolute physical significance.

Bridgman’s operational philosophy, introduced in 1927, essentially generalizes his dimensional analysis position and is strongly conditioned by his highly successful experimental discoveries done in an area where existing theory provided little guidance. His philosophical concern was making sense of experimentation and its place in an increasingly theoretical physics if absolute qualities did not exist. He concluded that special relativity also undercut such absolute principles as the uniformity of nature. The focal issue became how to make sense of truth without lapsing into irrationalism if an absolute or transcendental reality is rejected.

Bridgman interpreted Einstein’s special relativity theory as showing that objectivity is to be found in what the physicist *does*. What is real *is* what is meaningful and ‘we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations’. Each operation thus defines a separate concept. For Bridgman the measuring operations were the *only* permanent physical entities. Bridgman saw operationalism as a method for analysis - for sorting out what was and was not real among scientific assertions - and not a criterion for regimenting or demarcating the scientific from the unscientific as others interpreted it (see Operationalism).

Bridgman later moved from the idea that meanings of scientific claims were the operations the scientist performed to the view that knowledge is the product of human activity and rests on acts of the understanding - an essential private knowledge component that is not to be compromised by externally imposed demands for consistency. Critics accused Bridgman of flirting with solipsism, a charge he steadfastly denied.

Bridgman interpreted general relativity theory as reintroducing absolutes to science. Since he had come to reject all absolutes, Bridgman rejected general relativity theory. His attempts to reconcile quantum theory and thermodynamics focused on what they showed about the limits of human subjectivity, reinforcing his rejection of all absolutes as unknowable. Metaphysical and other absolutes including absolute ethical principles thus were illegitimate vehicles for imposing conformity of intellect on others.

In subjectivity of knowledge Bridgman found the source of cognitive freedom: to be scientific is to rely on the intellect, accepting only what is operationally knowable, and thus is the paradigm of intellectual integrity. Intellectual integrity is the essence of freedom. In social, moral and political realms, the path to freedom is to be scientific. What was real socially and what was morally binding were what was true. To assess truth one relied on operational analysis. In *The Intelligent Individual and Society* (1938) he subjected social and moral concepts to operational analysis, arguing that 'duty', 'responsibility', the idea of a 'right', and 'justice' all fail to have operational meaning and thus have only coercive utility. So, too, the state fails to exist as an entity in its own right. Bridgman rejected the idea that 'the best and deepest interests of the individual are coincident with those of society'. Rather than embracing solipsism he found that the key to individual freedom from authority is one's isolation and subjectivity. Freedom is the only ultimate.

See also: General relativity, philosophical responses to; Logical positivism; Measurement, theory of

FREDERICK SUPPE

List of works

Bridgman, P.W. (1922) *Dimensional Analysis*, New Haven, CT: Yale University Press.(His position here lays the basis for his operationalist philosophy.)

Bridgman, P.W. (1927) *The Logic of Modern Physics*, New York: Macmillan; repr. Salem, NH: Ayer Co., 1993. (Original statement of his operationalism.)

Bridgman, P.W. (1931) *The Physics of High Pressure*, London: G. Bell & Sons.(Based on his high pressure scientific experiments.)

Bridgman, P.W. (1936) *The Nature of Physical Theory*, Princeton, NJ: Princeton University Press; repr. in *Philosophical Writings of Percy Williams Bridgman*, New York: Arno Press, 1980.(In this work Bridgman pursues the subjectivist tendencies in his operationalism.)

Bridgman, P.W. (1938) *The Intelligent Individual and Society*, New York: Macmillan.(Basic articulation of his extension of operational method to social and moral philosophy with an emphasis on issues of individual freedom.)

Bridgman, P.W. (1941) *The Nature of Thermodynamics*, Cambridge, MA: Harvard University Press.(This summarizes Bridgman's attempts to deal with the epistemological implications of nature having indeterministic stochastic properties.)

Bridgman, P.W. (1952) *The Nature of Some of Our Physical Concepts*, New York: Philosophical Library; repr. in *Philosophical Writings of Percy Williams Bridgman*, New York: Arno Press, 1980.(Later restatement of his operationalist philosophy similar to his 1936 work.)

Bridgman, P.W. (1955) *Reflections of a Physicist*, New York: Philosophical Library, 2nd edn; repr. New York: Arno Press, 1980.(This contains reprints of nearly all his philosophical and nontechnical papers.)

Bridgman, P.W. (1959) *The Way Things Are*, Cambridge, MA: Harvard University Press.(A reiteration of his 1938 views on society and freedom.)

Bridgman, P.W. (1962) *A Sophisticate's Primer of Relativity*, Middletown, CT: Wesleyan University Press; 2nd edn, 1983.(First edition contains a critical commentary by A. Grünbaum, the second an introduction by A. Miller, unpublished notes of Bridgman's from the 1920s, and drafts.)

Bridgman, P.W. (1964) *Collected Experimental Papers*, Cambridge, MA: Harvard University Press, 7 vols. (Contains a complete bibliography of his scientific writings.)

References and further reading

Moyer, A.E. (1991) 'P.W. Bridgman's Operational Perspective on Physics. Part I: Origins and Development' and 'Part II: Refinements, Publications, and Receptions', *Studies in the History and Philosophy of Science* 22: 237-58, 373-97.(Historical account of development of his ideas.)

Walter, M.L. (1990) *Science and Cultural Crisis: An Intellectual Bibliography of Percy Williams Bridgman (1882-1961)*, Stanford, CA: Stanford University Press. (An invaluable, detailed biography based on archival and published sources. Does an excellent job of relating how Bridgman's scientific work informs his philosophical views. Contains comprehensive bibliography.)

Brinkley, Richard (*fl.* 1350-73)

Richard Brinkley was a Franciscan theologian at the University of Oxford in the latter half of the fourteenth century. Probably at the request of his superiors, he undertook an attack on nominalism and conceptualism, resulting in his best-known work, Summa logicae (Synopsis of Logic). Other works include a commentary on Peter Lombard's Sentences, which survives only fragmentarily and in a student's shortened version. Brinkley had a significant influence on several generations of Oxford logicians and Parisian theologians.

Brinkley was active at Oxford University sometime between 1350 and 1373. His successors called him *Doctor Bonus* or *Doctor Valens* (the 'Good Doctor' or 'Capable Doctor'). His principal surviving philosophical work is the lengthy *Summa logicae (Synopsis of Logic)* comprising over one hundred manuscript folia (roughly equivalent to 500 ordinary printed pages). Brinkley's extant theological works include fragments of his *Commentum super Sententias*, a commentary on the *Sentences* of Peter Lombard, which also survives in a shortened version (*abbreviatio*) by Stephen Gaudet, who delivered his own commentary on the *Sentences* in 1361-2 (see Kaluza 1989: 181-88). The same manuscript in which Gaudet's *abbreviatio* is found also contains *Quaestiones magnae* and *Quaestiones breves (Long Questions and Short Questions)* on various philosophical issues in theological contexts, probably also by Brinkley. Lost works by Brinkley include a *Distinctiones scholasticae (Scholastic Distinctions)* listed by John Bale in the sixteenth century, and a *Quaestiones theologicales Biligam et Brinkel (Theological Questions of Brinkley and [Richard] Billingham)* (probably a student's compilation), mentioned in an old Prague University library catalogue. The *Lectura super Sententias (Lecture on the Sentences)* mentioned by Bale is very likely identical with the *Commentum*. Bale also mentions a *Determinationes (Determinations)*, but this work is probably by William of Foville, a Franciscan of Cambridge University.

Brinkley's *Summa logicae* seems to date from between 1360 and 1373 (see Gál and Wood 1980: 77-8), perhaps after his theological works. It was unusual but not unheard of for a medieval thinker to produce purely philosophical works after entering the faculty of theology. In Brinkley's case, this unusual sequence might have resulted from his having been ordered by his superiors to write a logical work against the predominant nominalism and conceptualism of his time. The *Summa logicae* advertises itself as a manual for young students, but primarily it is a polemical defence of the realist position regarding universals. It consists of seven parts: 'On Terms', 'On Universals', 'On the Categories', 'On the Supposition of Terms', 'On Propositions', 'On Insolubles' and 'On Obligations'.

Brinkley advances his position with sophisticated theories, some of which are of contemporary interest. For instance, in Part V of the *Summa logicae*, continuing a debate between Adam Wodeham and Walter Chatton, Brinkley argues against Richard Billingham (Regent Master at Oxford in 1349), William of Bermingham (a Doctor of Theology by 1362) and Richard Feribrigge (Rector of Shelton in 1361), who hold, respectively, that what a proposition signifies is a mode of a thing, a mental act or simply the significate of the subject term. Brinkley rejects the need for any intensional semantic entity, and holds the extentionalist position that the primary significate of any proposition is the coordinated things in the world signified by the terms of the proposition. Brinkley's views may have been known to Henry Hopton, Albert of Saxony and John Wyclif.

Brinkley's theory of supposition is thoroughly informed by his realist views. For instance, for Brinkley simple supposition occurs whenever a term designates an independent, extra-mental universal, as does the term 'man' in the proposition 'Man is a species'. Even compared to other realists, Brinkley adopted some unusual positions, such as his rejection of the usual interpretation of material supposition, associated with a term's standing for itself (as does the term 'man' in the proposition 'Man is monosyllabic'). All supposition, Brinkley says, requires a relation, and nothing can be related to itself.

Brinkley's theological writings are dated between 1350 and 1360. Like most ostensibly theological works of his day, his writings are an amalgam of physics, moral philosophy, epistemology and metaphysics, as well as pure theological speculation. It is difficult to assess the full range of Brinkley's thought definitively, given the fragmentary state of his literary remains. Much of what we know of him depends on allusions to his writings in the works of other authors. While Brinkley never achieved the status of an authority, he prompted both assent and dissent among several generations of Parisian theologians. In the fourteenth century alone, authors whom he

influenced include Angelus of Dobelin, Dionysius of Montina, Galerand of Pendref, Henry Totting of Oyta, Jacob of Eltville, John Bramarth, John Hiltalingen of Basel (and his associate Paul de Fonte), Peter of Candia, Stephen Gaudet and William Centueri of Cremona.

See also: Language, medieval theories of; Nominalism; Realism and antirealism

ROBERT ANDREWS

List of works

Brinkley, R. (*c.*1350-60) *Commentum super Sententias (Commentary on the Sentences)*, selection edited in Z. Kaluza, 'L'oeuvre théologique de Richard Brinkley, OFM', *Archives d'histoire doctrinale et littéraire du Moyen-âge* 56, 1989: 169-273.(Only fragments of this work survive. For a manuscript, see Bibl. Nazionale (Florence), conv. soppr. B 6 910 (*sine fol.*).)

Brinkley, R. (*c.*1350-60) *Quaestiones magnae* and *Quaestiones breves* (Long Questions and Short Questions), selection edited in Z. Kaluza, 'L'oeuvre théologique de Richard Brinkley, OFM', *Archives d'histoire doctrinale et littéraire du Moyen-âge* 56, 1989: 169-273.(Original manuscript in Paris (Bibl. Nat., lat. 16535, ff. 128r-129r) also contains Stephen Gaudet's *abbreviatio*.)

(*c.*1360-73) *Summa logicae (Synopsis of Logic)*, Part V ed. in M.J. Fitzgerald, *Richard Brinkley's Theory of Sentential Reference: 'De significato propositionis' from Part V of His Summa nova de logica*, Studien und Texte zur Geistesgeschichte des Mittelalters XVIII, Leiden: Brill, 1987; Part VI ed. in P.V. Spade, *An Anonymous Fourteenth Century Treatise on 'Insolubles': Text and Study*, Toronto, Ont.: The Pontifical Institute of Mediaeval Studies, 1969; Part VII ed. P.V. Spade and G. Wilson, *Richard Brinkley's Obligationes: A Late Fourteenth Century Treatise on the Logic of Disputation*, Münster: Aschendorff, 1995; five various chapters ed. in G. Gál and R. Wood, 'Richard Brinkley and His *Summa logicae'*, *Franciscan Studies* 40, 1980: 59-101.

References and further reading

Fitzgerald, M.J. (ed.) (1987) *Richard Brinkley's Theory of Sentential Reference: 'De significato propositionis' from Part V of His Summa nova de logica*, Studien und Texte zur Geistesgeschichte des Mittelalters XVIII, Leiden: Brill.(Edition and English translation (by M.J. Fitzgerald) of Part V of the *Summa logicae*, with a discussion of Brinkley's semantic theory and its influence.)

Gál, G. and Wood, R. (1980) 'Richard Brinkley and His *Summa logicae*', *Franciscan Studies* 40: 59-101.(An introduction to the *Summa logicae*, with an edition of five scattered chapters from Parts I-IV.)

Kaluza, Z. (1989) 'L'oeuvre théologique de Richard Brinkley, OFM' (The Theological Work of Richard Brinkely, OFM), *Archives d'histoire doctrinale et littéraire du Moyen-âge*: 169-273.(The most thorough discussion of Brinkley's theological work, with a complete edition of Stephen Gaudet's *Abbreviatio* of the *Commentum super Sententias*, a surviving fragment of the *Commentum*, question 3, and a selection from the *Quaestiones magnae*.)

Spade, P.V. (1969) *An Anonymous Fourteenth Century Treatise on 'Insolubles': Text and Study*, Toronto, Ont.: The Pontifical Institute of Mediaeval Studies.(A transcription of Part VI of the *Summa logicae*.)

Spade, P.V. (1991) 'Richard Brinkley's *De insolubilibus*: A Preliminary Assessment', *Rivista de storia della filosofia* 46: 245-56.(A critical analysis of Brinkley's theory of *obligationes*.)

Spade, P.V. (1993) 'Opposing and Responding: A New Look at "Positio"', *Medioevo* 19: 233-70.(Edition of selected passages of Part VI of the *Summa logicae*, the *Insolubilia*.)

Spade, P.V. and Wilson, G. (eds) (1995) *Richard Brinkley's Obligationes: A Late Fourteenth Century Treatise on the Logic of Disputation*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters, NF 43, Münster: Aschendorff.(Edition, with an introduction and some comments, of Part VII of the *Summa logicae*.)

Brito, Radulphus (*c*.1270-*c*.1320)

Radulphus Brito was a prominent master of arts at the University of Paris around 1300. In order to secure the foundation of concepts in extramental reality, he devised a system of four types of 'intentions', first and second, abstract and concrete. As a philosopher of language, similar concerns made him claim a formal identity between the modes of signifying (of words) and the modes of being signified (of things).

Probably a native of Brittany, Radulphus Brito (Ralph the Breton) must have been born about 1270, as he became a master of arts in Paris no later than 1296 and doctor of theology, also in Paris, about 1311-2. He died in 1320 or later.

The extant body of Radulphus' work is mainly unedited. Most of it relates to his activity as a master of arts, and comprises questions on the Aristotelian Organon, including Porphyry's *Isagōgē*, the anonymous *Six Principles* and Boethius' *De differentiis topicis (On Topical Differences)*, and on Aristotle's *De anima*, *Physics*, *Meteorology* and *Metaphysics*, as well as on some mathematical and astronomical textbooks. There are also *sophismata* (investigations of particular logical problems), and a work on grammatical theory, *Quaestiones super Priscianum Minorem (Questions on Priscian Minor)*, also called *Modi significandi (Modes of Signifying)*. From Radulphus' career in theology we have questions on Peter Lombard's *Sentences*, a *Quodlibet* and *Quaestiones in Vesperis (Evening Questions)*.

Radulphus became quickly famous, and younger contemporaries such as Bartholomew of Bruges attempted to dispute his views. The logical works continued to be read until some time in the fifteenth century, especially in Italy, and in the 1430s some parts were translated into Greek by Georgios (Gennadios) Scholarios (see Byzantine philosophy).

Radulphus' fame rests on his epistemology, and in particular on his distinction between concrete and abstract first and second intentions, the conceptual counterparts of such series of nouns as 'man, humanity; universal, universality'. Traditionally, first intentions such as 'horse' were distinguished from second intentions such as *species*, which presuppose the first ones. Radulphus divided both first and second intentions into abstract and concrete ones. The abstract first intention is a formal concept, such as 'humanity'. It is based on the features or manifestations (modes of being, *apparentia*) of some essence (or 'nature'). Reasoning, for example, is a manifestation of human nature. This abstract first intention is a thought (*cognitio*) whose object is man, but it does not include its object. The concrete first intention is the object of the abstract one, but it is not purely extramental. It is the thing (for example, man) considered simply as thought of by means of the formal concept (humanity). Alternatively, it is described as an aggregate of the thing and the thought by which we grasp it. This ontological duplicity was often criticised in later times, but what Brito wanted was to secure the lifeline between concepts and their objects without moving the objects into the mind.

Likewise for the second intentions, the modes of being form the basis of concept formation, but this time the modes in question are not proper to some nature but common to natures. The abstract second intention, (for example, universality), is a concept derived from the common feature of being capable of occurring in several individuals or types. This feature is shared by, for example, 'man' and 'donkey', both of which can occur in several individuals, and also by 'animal', which can occur in several species. Our intellect need perform no comparison to construct a Porphyrian tree; sensory acquaintance with a single individual suffices to recognize, for instance, that sensing (which characterizes animals) is a trait apt to be shared by more beings than reason (which is reserved for humans). The corresponding concrete second intention (universal) is the thing (man) considered simply as conceived of by means of the formal concept of universality.

Earlier thinkers had held that second intentions arise from inspection and comparison of things already grasped by the mind. Brito wanted to generate them through direct inspection of the entities that gave rise to the first intentions. His aim was clearly to strengthen the link of the second intentions to reality by making them thoughts of things rather than of thoughts.

The linguistic theory (modism) prevalent in Radulphus' days took grammatical categories to be features (modes of signifying) of significative words but derived from features (modes of being) of things (see Language, medieval

theories of). Some decades of discussion had revealed weaknesses of the theory, and Brito fought an uphill battle to save it. Thus, to avoid difficulties arising from the existence of equivocal words, modes of signifying had been split into active modes, residing in words, and passive ones (modes of being signified), residing in things, such that one active mode could correspond to several passive ones. Seeing that this compromised the basic modistic idea of deriving features of thought and of language from features of reality, Radulphus claimed that the active modes are formally, but not materially, identical with passive modes, which in turn are materially identical with passive modes of being and of understanding. This view was taken over by the influential grammarian Thomas of Erfurt.

See also Language, medieval theories of; Logic, medieval

STEN EBBESEN

List of works

Brito, Radulphus (*c*.1270-*c*.1320) *Quaestiones super librum De anima (Commentary on Aristotle's On the Soul)*, ed. W. Fauser, *Der Kommentar des Radulphus Brito zu Buch III De anima*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters NF 12, Münster: Aschendorff, 1974.

Brito, Radulphus (*c*.1270-*c*.1320) *Quaestiones super libra Topicarum Boethii (Questions on Boethius' Topics)*, ed. N.J. Green-Pedersen and J. Pinborg, 'Radulphus Brito: Commentary on Boethius' *De differentiis topicis* and the Sophism "Omnis homo est omnis homo"', *Cahiers de l'Institut du Moyen-Age Grec et Latin* 26, 1978.

Brito, Radulphus (*c*.1270-*c*.1320) *Sophisma 'Omnis homo est omnis homo' (Sophism: 'Every Man is Every Man')*, ed. N.J. Green-Pedersen and J. Pinborg, 'Radulphus Brito: Commentary on Boethius' *De differentiis topicis* and the Sophism "Omnis homo est omnis homo"', *Cahiers de l'Institut du Moyen-Age Grec et Latin* 26, 1978.

Brito, Radulphus (*c*.1270-*c*.1320) *Quaestiones super Priscianum Minorem (Questions on Priscian Minor)*, ed. H.W. Enders and J. Pinborg, *Grammatica Speculativa* 3.1-2, Stuttgart-Bad Cannstatt: Fromman-Holzboog, 1980.(On pages 15-19 there is a complete list of extant works by Radulphus of Brito, and of published extracts. Supplements to this list can be found in Ebbesen and Pinborg (1981-2): 318, n.3-5.)

Brito, Radulphus (*c*.1270-*c*.1320) *Quaestiones super Artem Veterem (Questions on the Old Logic)*, ed. S. Ebbesen and J. Pinborg, 'Gennadios and Western Scholasticism. Radulphus Brito's *Ars Vetus* in Greek Translation', *Classica et Mediaevalia* 33, 1981-2: 263-319.(This is an edition of the prologue of the question.)

Brito, Radulphus (*c*.1270-*c*.1320) *Quaestiones super librum Elenchorum (Questions on the Sophistical Refutations)*, ed. S. Ebbesen and J. Pinborg, 'Gennadios and Western Scholasticism. Radulphus Brito's *Ars Vetus* in Greek Translation', *Classica et Mediaevalia* 33, 1981-2: 263-319.(This is an edition of the prologue of the question.)

Brito, Radulphus (*c*.1270-*c*.1320) *Sophisma 'Aliquis homo est species' (Sophism: 'Some Man is a Species')*, ed. J. Pinborg, 'Radulphus Brito's Sophism on Second Intentions', *Vivarium* 13, 1975: 119-52. (Edition of the sophism.)

Brito, Radulphus (*c*.1270-*c*.1320) *Quaestiones super librum Porphyrii (Questions on Porphyry's Isagōgē)*, ed. J. Pinborg, 'Radulphus Brito on Universals', *Cahiers de l'Institut du Moyen-Age Grec et Latin* 35, 1980: 56-142. (Edits long section from the questions on Porphyry. An extract from this work also appears in H. Roos, 'Zur Begriffsgeschichte des Terminus "apparens" in den logischen Schriften des ausgehenden 13. Jahrhunderts', in J. Moller and H. Kohlenberger (eds) *Virtus politica. Festgabe zum 75. Geburtstag von Alfons Hufnagel*, Stuttgart-Bad Cannstatt: Frommann-Holzboog, 1974, 323-34.)

References and further reading

Ebbesen, S. (1980) 'Is 'canis currit' ungrammatical? Grammar in Elenchi commentaries', *Historiographia Linguistica* 7 (1/2): 53-68.(Inter alia, on the problem of active and passive modes of signifying.)

Marmo, C. (1994) *Semiotica e linguaggio nella scolastica: Parigi, Bologna, Erfurt 1270- 1330 (Semiotics and Language in Scholasticism: Paris, Bologna, Erfurt, 1270-1330)*, Rome: Istituto Storico Italiano per il Medio Evo.(A large-scale study of modism, based on both grammatical and logical texts.)

Pinborg, J. (1967) *Die Entwicklung der Sprachtheorie im Mittelalter (The Development of the Theory of Language in the Middle Ages)*, Beiträge dur Geschichte der Philosophie und Theologie des Mittelalters, Texte und Untersuchungen 42.2, Münster: Ashendorff, and Copenhagen: Frost-Hansen.(The foundation of all later work on modism.)

Pinborg, J.(1972) *Logik und Semantik im Mittelalter. Ein Überblick (Logic and Semantics in the Middle Ages: A Survey)*, Stuttgart-Bad Cannstatt: Frommann-Holzboog.(Pays special attention to Brito.)

Pinborg, J. (1984) *Medieval Semantics, Selected Studies on Medieval Logic and Grammar*, ed. S. Ebbesen,

London: Variorum.(Several relevant papers, including one on the reception of Brito’s theory of intentions.)

Broad, Charlie Dunbar (1887-1971)

A Cambridge contemporary of Russell, Moore and Wittgenstein, C.D. Broad wrote on an exceptional range of topics, including causation, perception, the philosophy of space and time, probability and induction, mind and body, ethics and the history of philosophy. He typically set out a number of received positions on a topic, explored their consequences with great clarity, and then came to a cautious estimate of where the truth probably lay. However, Broad made some notable contributions of his own, especially on perception (he defended a representative theory), induction (he argued that our inductive practices require the existence of natural kinds), and time (he argued that tensed facts cannot be analysed away). Although his talents lay in very careful analysis, Broad insisted that there was a proper place in philosophy for metaphysical speculation; he particularly admired McTaggart, and his monumental Examination of McTaggart's Philosophy (1933, 1938) contains some of Broad's best work.

1 Conception of philosophy

Broad was born near London and educated at Dulwich College and Trinity College, Cambridge, initially as a natural scientist. He then switched to philosophy, winning a prize fellowship at Trinity in 1911: his revised dissertation was published as *Perception, Physics and Reality* (1914). After teaching at St Andrews, Broad became professor of philosophy at Bristol University in 1920, then succeeded McTaggart as Fellow in Moral Science at Trinity College in 1923. Broad's return to Cambridge coincided with the publication of *Scientific Thought* (1923). *The Mind and Its Place in Nature* (1925) and *Five Types of Ethical Theory* (1930) quickly followed, and the first volume of his *Examination of McTaggart's Philosophy* appeared in 1933. The same year, Broad became Knightbridge Professor of Moral Philosophy in Cambridge. For most of the tenure of the professorship he was, at least by his own account, an 'unbelieving Pope' who 'no longer believed in the importance of philosophy' - though he continued to publish a steady flow of articles, and after retiring in 1953 he wrote a vigorous 'Reply to My Critics'. He also wrote extensively on psychical research, in which he had a lifelong interest.

Broad's later disaffection lay, more precisely, with the style of philosophy practised in a Cambridge that was increasingly under the spell of Ludwig Wittgenstein. In *Perception, Physics And Reality*, Broad had written that the only way of assessing a fundamental philosophical thesis is to see 'what would follow if it were true and then try to judge whether these results make that which implies them probable or improbable'. And he continued to regard philosophy as similar in this way to science: in his 'Reply', Broad still maintained that ' the philosophy of the physical world and of our perception of it [is] analogous in certain respects (though profoundly dissimilar in certain others) to the making and testing of a far-reaching scientific hypothesis' (Schilpp 1959). This conception of philosophy (now widely shared by many philosophers influenced, for example, by Quine) put him profoundly at odds with the 'Wittgersnappers', who saw thinkers such as Broad as conceptual delinquents in the grip of pseudo-problems, and recommended close attention to our ordinary language as appropriate therapy. But for Broad, 'in philosophy it is equally silly to be a slave to common speech or to neglect it'. And he always insisted that there is a proper place not just for Critical Philosophy - 'the analysis and definition of our fundamental concepts, and the clear statement and resolute criticism of our fundamental beliefs' - but also for Speculative Philosophy which hopes 'to reach some general conclusions as to the nature of the Universe, and as to our position and prospects in it'.

Much of his writing directly reflects Broad's conception of philosophy: like a judicious scientist, he weighs up the evidence for and against various theories, and then comes to a guarded verdict about their relative acceptability. The final chapter of *Mind and Its Place in Nature* is an extreme example, where he defines seventeen possible accounts of the relation between the mental and the physical, gradually whittles these down to a short list of three, finally judging one of them - 'emergent materialism' - as 'on all the evidence which is available to me… the most likely view'. In other hands this approach could make for tedium, but Broad writes with model clarity, and often with admirable elegance enlivened by flashes of mordant wit.

2 Perception

Broad discussed perception more often than any other topic. In particular, *Scientific Thought* contains a classic development of the sense-datum theory. When perceiving a stick half immersed in water, it looks bent; and ' the

most obvious analysis of the facts is that, when we judge that a straight stick *looks* bent, we are aware of an object [a 'sensum', to use Broad's preferred term] which really *is* bent, and which is related in a peculiarly intimate way to the physically straight stick'. Broad stresses that there are promising alternatives to this analysis of perceptual experience in terms of the apprehension of some special kind of transitory object. Still, in *Scientific Thought* he assumes that the sensum analysis is basically correct, and then explores how it should be developed in detail.

This approach has fallen into considerable disfavour since Austin's classic assault on sense-data theories, for example (see Austin, J.L. §2); but it would be wrong to conclude that there can now be nothing of interest in Broad's investigations. As he remarks in *Mind and Its Place in Nature*, 'a perceptual situation [such as seeing a cat] is "intuitive", while a thought-situation with the same kind of epistemological object [such as thinking of a cat] is "discursive"'. Judging *my cat is ginger* involves thinking of something as being a cat; but one can see a ginger cat without having the concept of a cat at all - the perceptual experience has, to use a later idiom, a non-conceptual content. And much of Broad's discussion of sensa can be read as an examination of such non-conceptual contents in a framework which need not presuppose that sensa are genuine 'objects'. For example, he revealingly investigates how our visual experiences interlock with our experiences of bodily movement and touch as we explore the world so as simultaneously to determine the rather different spatial characters of our visual and tactual perceptual contents.

Broad defends a refined version of Locke's account of our knowledge of the physical world (our experiences are caused by a world which, in certain respects, they resemble). But in *Scientific Thought*, Broad suggests that there is no conclusive proof to be had here. Rather, we can only indicate those facts about our sense experience 'which would give a high final probability to the belief in a physical world, provided it had a finite [non-zero] antecedent probability'. This probabilistic realism sharply distinguishes Broad from Russell, who at the time was aiming to 'construct' the external world from sense-data, rather than riskily infer it - and also from G.E. Moore, who was a realist but not a probabilist (see Perception §2; Sense-data §§1-3).

3 Induction

Another pregnant suggestion involving probability appears in Broad's early account of causality. In *Perception, Physics and Reality*, he examines some alleged difficulties about causation raised by idealist philosophers such as Bradley, and suggests that the way out is to hold that the proper form of a causal law is 'the occurrence of *X* at any moment increases the probability of *Y*'s occurrence at a moment τ later over what it would have been if *X* had not happened'.

But this novel proposal is not further explored (although in *Examination of McTaggart's Philosophy* Broad continues to make room for indeterministic causation). What Broad does concentrate on, in his many discussions related to probability, is the question of the character and justification of inductive reasoning. His long paper, 'The Relation Between Induction and Probability' ([1918] 1968), is noteworthy. Broad argues that inferences such as 'All observed *S*'s have been *P*: hence the next *S* will be *P*' are irreducibly probabilistic. However, he also argues that we cannot immediately apply to scientific inductions the results from probability theory that govern artificial situations like predicting the colour of a ball drawn from a bag. On certain assumptions we can show that if a random sample of *m* balls has proved to be all red, then the probability that the next ball out of the bag is red is $(m+1)/(m+2)$, so almost 1 when *m* is large. But we cannot carry over such a result in order to guarantee a high probability to the prediction that the next swan will be white, having observed many white swans, because the assumptions that apply in the artificial case do not apply in nature.

If an appeal to probability theory is not enough to warrant our induction here, what more is needed? Our most confident inductions are just those which deal with things that 'are believed to belong to what Mill would call a Natural Kind' (such as 'swan'): taking his cue from this, Broad argues that it is the real existence of natural kinds with characteristic causal powers which is required to underpin our inductive practices - and equally, is bound up with the very idea of a world containing 'continuants', that is, particular things that persist through time.

4 Time

Broad's *Examination of McTaggart's Philosophy* is a very detailed commentary on McTaggart's *Nature of Existenc*e (it concludes that 'none of McTaggart's more characteristic conclusions have been established by his

arguments', a persuasive verdict that still leaves Broad rating McTaggart alongside Leibniz as a speculative metaphysician). But it also contains a series of interwoven discussions in which Broad offers his own independent treatment of various issues. Perhaps the most impressive of these is 'The "Nature" of a Continuant', where Broad extends his work on kinds, arguing, for example, that the only way of making sense of counterfactual claims about what a given thing would have done if its circumstances had been different 'presupposes constancy of nature in individuals and the existence of natural kinds'.

However, the best-known part of *Examination of McTaggart's Philosophy* deals with time. We think of time as essentially involving change, and change as involving events that were still in the future becoming present, and then receding into the past - yet McTaggart famously argues that this conception leads to contradiction. Broad's analysis of McTaggart's reasoning here is definitive. 'He assumes that what is meant by a sentence with a *temporal copula* [such as 'is now', 'was', 'will be'] must be completely… expressible by a sentence or combination of sentences in which there is no temporal copula, but only temporal predicates and non-temporal copulas.' Thus, the truth that event *E will* occur consists in *E*'s having (timelessly!) the property of futurity. But eventually, the truth will be that *E is now* happening, so - on the same assumption - that different truth consists in *E*'s timelessly having the incompatible property of presentness. Contradiction ensues: hence, Broad concludes, we must reject McTaggart's tacit assumption that temporal copulas can be replaced by temporal predicates.

Broad also criticizes Russell's very different attempt to analyse away temporal copulas by treating, for instance, '*E* will occur' as meaning '*E* is later than this utterance' - where the 'is' is timeless. Broad argues that this theory requires us to conceive of past, present and future events as timelessly 'coexisting' in a way that it is impossible for us to render intelligible, or at least not compatibly with 'the fact that at any moment [certain events] are marked out from all the rest by the quality of presentness'. He concludes that tensed facts (as expressed using temporal copulas) are fundamental; or, in other words, 'Absolute Becoming' is an irreducible feature of the world. Broad's reasoning here is debatable - but his discussion remains a paradigm presentation of the 'tensed' view of time (see Tense and temporal logic).

5 Ethics

At the end of his widely read *Five Types of Ethical Theory*, Broad favours a mild form of ethical intuitionism; but later articles are non-committal. Indeed, in his 'Reply', when asked to commit himself on meta-ethical questions such as 'Do moral claims express *judgments* which are true or false?' or 'Are moral characteristics non-natural?', Broad confesses that he has 'no decided opinion'. On first-order questions, such as the merits of utilitarianism, Broad's usual philosophical caution is compounded by the belief that no single criterion of right action is likely to do justice to the complexity of our ethical lives. But lacking an axe to grind, his ethical discussions (if not especially original) are exceptionally careful and judicious, and remain valuable - where more recent writers have overlooked his writings, such as those on moral-sense theories, it has often been at the cost of labouring to rediscover distinctions already elegantly made by Broad.

PETER SMITH

List of works

Broad, C.D. (1914) *Perception, Physics and Reality*, Cambridge: Cambridge University Press.(Defends a Lockean causal theory of perception.)

Broad, C.D. (1923) *Scientific Thought*, London: Kegan Paul, Trench, Trubner & Co.(Part I - on space, time and motion - includes an excellent early treatment of the significance of the theory of relativity. Part II further develops the sensum analysis of perception.)

Broad, C.D. (1925) *The Mind and Its Place in Nature*, London: Kegan Paul, Trench, Trubner & Co.(Extended treatments of the unconscious and 'The Alleged Evidence for Human Survival of Bodily Death' are sandwiched between illuminating and influential discussions of the mind-body problem.)

Broad, C.D. (1930) *Five Types of Ethical Theory*, London: Kegan Paul, Trench, Trubner & Co.(Discusses Spinoza, Butler, Hume, Kant and Sidgwick: a textbook from before the era of textbooks, deservedly very widely used for many years.)

Broad, C.D. (1933, 1938) *Examination of McTaggart's Philosophy*, Cambridge: Cambridge University Press. (Over 1,250 pages, the 'mausoleum' in which Broad inters his admired teacher's philosophy contains some of his own most impressive work.)

Broad, C.D. (1952) *Ethics and the History of Philosophy*, London: Routledge & Kegan Paul.(Mostly reprinted historical papers; but includes his influential inaugural lecture 'Determinism, Indeterminism and Libertarianism'.)

Broad, C.D. (1953) *Religion, Philosophy and Psychical Research*, London: Routledge & Kegan Paul.(Further reprinted papers; 'The Relevance of Psychical Research to Philosophy' usefully sums up Broad's stance on the topic - he is far from credulous, but insists that it behoves philosophers to examine the evidence carefully, and not merely dismiss the possibility of psychic phenomena on shaky a priori grounds.)

Broad, C.D. (1962) *Lectures on Psychical Research*, London: Routledge & Kegan Paul.

Broad, C.D. (1968) *Induction, Probability and Causation*, ed. J. Hintikka, Dordrecht: Reidel.(Reprints Broad's important papers on these topics, including 'The Relation Between Induction and Probability' from *Mind* 1918, 1920.)

Broad, C.D. (1971) *Broad's Critical Essays in Moral Philosophy*, ed. D. Cheney, London: Allen & Unwin.(A comprehensive collection - includes the fine 'Some Reflections on Moral-Sense Theories in Ethics'.)

Broad, C.D. (1975) *Leibniz, An Introduction*, ed. C. Lewy, Cambridge: Cambridge University Press.(Edited from undergraduate lectures given in 1949-50: a useful student text, with a particularly good treatment of Leibniz's theory of corporeal substances.)

Broad, C.D. (1978) *Kant, An Introduction*, ed. C. Lewy, Cambridge: Cambridge University Press.(Edited by Lewy from lectures given in 1951-2: again, a very accessible and reliable student text.)

Broad, C.D. (1985) *Ethics*, ed. C. Lewy, Dordrecht: Nijhoff.(Broad lectured on ethics almost every year from 1933-4 to 1952-3, the evolving course providing the setting from which a number of his papers were excerpted: this book is edited from his final set of lectures.)

References and further reading

Britton, K. (1978) 'Charlie Dunbar Broad, 1887-1971', *Proceedings of the British Academy* 64: 289-310.(An accessible review of the man and his work.)

McTaggart, J.M.E. (1921, 1927) *The Nature of Existence*, Cambridge: Cambridge University Press, 2 vols. (Broad edited the second, posthumously published volume of McTaggart's *magnum opus*.)

Schilpp, P.A. (ed.) (1959) *The Philosophy of C.D. Broad*, New York: Tudor.(Contains Broad's 'Autobiography', twenty-one critical essays - still overall the most useful discussions of his philosophy - and Broad's 'Reply to My Critics'. It also contains a detailed bibliography of Broad's publications up to 1959.)

Brown, Thomas (1778-1820)

Thomas Brown was the last prominent figure in the Scottish philosophical tradition deriving from David Hume and Thomas Reid. Like Reid, he took the mind's knowledge about itself to be a datum it is pointless to challenge or try to justify, since no other grounds can be more certain for us. But he defended Hume's account of causation as nothing more than invariable succession. The mind, therefore, is a simple substance, whose successive states are affected by and affect the states of physical objects: the laws according to which these changes take place are no harder to grasp than the effects of gravitation. Brown's lectures, published as delivered daily to Edinburgh students, seek to classify the laws of the mind so that we can conveniently understand ourselves, and direct our lives accordingly; the last quarter of his course draws conclusions for ethics and natural religion.

1 Life

Son of the minister of Kilmabreck, Thomas Brown was a precocious child who entered Edinburgh University in 1793 at 15, and after hearing Dugald Stewart's lectures in the following year submitted searching criticisms of them in writing (see Stewart, D.). He soon made a name in Edinburgh literary circles, and after abortive legal studies qualified as a medical doctor in 1803, submitting a thesis in Latin (which he was said to speak as fluently as English; his knowledge of modern European languages was also considerable). He entered medical practice, but kept up his philosophical interests, and when Stewart's health failed in 1809 Brown was made first his assistant and then his 'colleague' in the Edinburgh Chair of Moral Philosophy, in fact discharging all its duties until he himself fell ill in 1819. He died in the following year. His lectures, mostly written during the day before their first delivery, were highly thought of. They were published unrevised after his death, and throw an interesting light on his youthful student audience, being written in a powerful and complex Ciceronian style and peppered with quotations from a wide range of authors in Latin, French and Italian: Juvenal, Horace, Cicero and particularly Seneca are assumed to be familiar and are often not translated. He also quotes extensively from his favourite English poets, Alexander Pope and Mark Akenside. He published several volumes of his own ponderous poetry, in a similar style, which made little impact, but his lectures continued to be in demand (21st edition, 1851) until William Hamilton and idealism eclipsed the Scottish school; he has scarcely been remarked since.

2 Writings

Brown was one of the last notable exponents of the Scottish Common Sense School of philosophy, owing much to David Hume and Thomas Reid, but more to the former. His mature position is a substance dualism, in which Hume's 'theatre of the mind' has become a simple substance, the successive states of which - a series of 'feelings' - constitute its subject's mental life. His first publication, *Observations on the Zoonomia of Erasmus Darwin* (1798), an attack on the Zoonomia, was of little philosophical interest, but in his *Short Inquiry into the Nature of the Relation of Cause and Effect* (1803, much extended by the 3rd edition, 1818) he clarifies and purges Hume's view of causation. He defends Hume's account of causation as nothing but invariable succession, mounting a powerful attack on the notion of power, particularly that of unexercised power, as wielded by Reid. However, he also uses tools furnished by Reid: the world consists of substances; general terms are essentially abbreviations of lists of substances, suiting our own practical purposes, and descriptions are abstractions from states of those substances; we have by virtue of our nature some beliefs about the world that are not optional and are in no need of justification, and indeed could not be justified by anything more certain. Brown maintains, against Hume, that our belief that like antecedents are always followed by like consequences is just one such fundamental and pervasive belief; thus Hume's explanation of the effects of custom is out of place (experiments explore what the true causes of an effect are; we already believe there is some cause or other). He also mounts a vigorous demolition of Hume's whole impression/idea distinction, so as to detach belief from 'vivacity'; as a consequence Brown is able to combine bare Humean causal laws with complete absence of scepticism about both matter and God.

3 Philosophical outlook

Brown's philosophical acumen is better displayed in criticism of rival theories, through careful analysis and telling use of concrete examples, than in supporting his own. He was inclined (perhaps as a result of his scientific training, which he puts to good use as a source of examples) to apply Ockham's razor where possible, though not

at the expense of convenience - for example, he does not recommend that we should abandon speaking of powers, merely that we should be careful not to be taken in by the language which reifies them. Although he shares Reid's reliance on consciousness as our only evidence for the mind's operations, as well as Reid's use of first principles, he scoffs at multiplying 'powers' of the mind. Rather, the philosophical ambition should be to seek out and systematize the kinds of changes it invariably displays in various kinds of situation. His lectures, published under the title *Lectures on the Philosophy of the Human Mind* (1820), also promise practical benefits, since understanding the mind and its operations (and its limitations) will give us a better picture of our divinely appointed nature and destiny. The extensive range of rhetorical aids which so recommended him to his audiences is consciously deployed to this exhortatory purpose, particularly in the later lectures on ethics and natural theology.

For Brown, then, the mind is a simple substance, changes in which (throughout called 'feelings') follow changes in the bodily organs of sense, and vice versa. Its simplicity, and hence distinctness, enables him to fend off materialism: the bodily organs are mere aggregates of particles which our senses are too weak to distinguish and the physiological unity of which is a mere matter of descriptive convenience; but there cannot be half a thought at the northwest corner of the mind; therefore the mind is not the physical lump. Its distinctness also affords him an argument for immortality. The particles of the body are all conserved at death, being merely redistributed: why therefore, in the face of the obvious presumptions, should God see fit to single out the separate mental substance for destruction? Problems about how we can understand a causal relation between the physiological and the mental are put firmly aside: there is no a priori grasp of any causal relation (nothing about the look of gunpowder can tell you what will happen if you put a match to it), and so we must just observe what correlations actually obtain, as in any other science (gravitation's influence at a distance is at first sight equally problematic). It may well seem that this picture of the mind is open to Reid's criticisms of the theory of ideas (see Reid, T.§1); but Brown characteristically meets this doubt head on, arguing that Reid misguidedly took as literal what for most theorists since the collapse of Aristotelian real universals was merely convenient metaphor, and that, for such as Hume, 'ideas' in the proper sense were not the *object* of mental attention but its vehicle.

However, like Reid, Brown claims that our knowledge of things is through their properties, that is, how they behave. So the mind can be considered as effect and cause of contemporary states of other substances and as itself a succession of states. Although at any given moment its state is simple, this can be analysed, like a chemical substance, into the result of simpler inputs. 'Consciousness' does not name a separate power, but is a generic name for every kind of feeling - that is, mental awareness; nor is perception distinct from sensation. Abandoning previous ways of classifying mental operations, Brown deals first with changes in the mind which we discover to depend on outside causes (in our organs) and then with those which the mind itself produces. It emerges that though belief in a 'permanent external system of things' is for adults not optional, it may be explainable in terms of early experience. He speculates at length about the way in which kinaesthetic sensations may provide the associations with vision which can account for how the infant solves George Berkeley's problem of deriving spatial information from colours (see Berkeley, G.).

The affections of the mind caused by its own internal states Brown describes in terms of ingeniously various laws of 'suggestion': the extensive and detailed discussions of particular emotions are full of interest. Like Reid, Brown takes care to avoid regressive accounts: the mind cannot aim at producing a volition without circularity, and so desires and 'volitions' differ only as long-term and short-term states, causing effects in muscular action. The resulting ethics is a moral-sense account reminiscent of Francis Hutcheson's, classifying moral feelings towards human agents separately from the other emotions (see Hutcheson, F. §2). This requires, and delivers, criticism of Adam Smith's rival account (see Smith, A.), but Brown is really more interested in classifying our duties and reminding us at full oratorical stretch of the powerful motives in human nature for fulfilling them.

CHRISTOPHER BRYANT

List of works

Brown, T. (1798) *Observations on the Zoonomia of Erasmus Darwin*, Edinburgh: Mundell.(Academic infighting, of no enduring interest.)

(1803) *A Short Inquiry into the Nature of the Relation of Cause and Effect, with Observations on the Nature and Tendency of the Doctrine of Mr Hume*, Edinburgh: Walker & Greig; enlarged 3rd edn, Edinburgh: Constable, 1818.(Study of Hume's constant-conjunction theory, detached from Hume's metaphysics and defended against

Reid’s criticisms. Clear statement of problems of evidence and experimental research in science.)
(1820) *Lectures on the Philosophy of the Human Mind*, Edinburgh: Tait, 4 vols.(Brown’s lecture course as delivered to Edinburgh undergraduates day by day. A broadly associationist psychology which seeks to classify and analyse mental operations of all kinds - perception, judgment, thought, desire, value.)

References and further reading

Davie, G. (1961) *The Democratic Intellect*, Edinburgh: Edinburgh University Press.(General work on Scottish culture in Edinburgh at the Enlightenment period.)
Dictionary of National Biography (1886) vol. 7: 31, article on Thomas Brown, London: Smith Elder.(Full biography, but not very helpful about philosophical views.)
Laurie, H. (1902) *Scottish Philosophy in its National Development*, Glasgow: Maclehose.(Now somewhat dated in approach.)
McCosh, J. (1875) *The Scottish Philosophy: Biographical, Expository, Critical*, New York: Carter; repr. Bristol: Thoemmes, 1990.(Probably still the most useful general work on the Scottish Common Sense School. Several pages devoted to each individual figure, including biography and summary of philosophical positions.)
Robinson, D.S. (1961) *The Story of Scottish Philosophy*, New York: Exposition Press.

Browne, Peter (1666-1735)

Peter Browne, an Irish bishop, was a critic of Locke's theory of ideas. His chief philosophical concern was to explain how human beings can conceive of God. He proposed that God's existence and attributes can be understood analogically, by their real - though inevitably partial - resemblance to human things. He distinguished between analogy, which turns on a 'real resemblance', and metaphor, which turns on a merely imagined one.

Browne entered Trinity College Dublin in 1682 and became a fellow in 1692. He served as provost of Trinity from 1699 until 1710, when he became Bishop of Cork and Ross in the south of Ireland. George Berkeley was a student and fellow at the college during Browne's tenure as provost. Years later, in the Fourth Dialogue of *Alciphron*, Berkeley, without naming Browne, attacked his analogical theology (see Berkeley, G. §11). He took Browne to be denying that we have any notion at all of God's attributes. Browne's angry reply occupies the long final chapter of his *Things Divine and Supernatural* (1733).

Browne's first publication of philosophical interest was his *Letter* (1697) on John Toland's *Christianity Not Mysterious* (1696) (see Toland, J.). Relying on a simple version of the doctrine of analogy elaborated in his later works, Browne argues that although religious mysteries can be understood only in part, they *can* be understood and can therefore be believed. He also argues that they can be believed with good reason. It is perfectly legitimate to trust informants we know to be truthful and able, even if we cannot fully comprehend what they relate. And if their authority is confirmed by miracles our 'evidence or degree of knowledge' will mount even higher.

Browne's main philosophical work, *The Procedure, Extent, and Limits of Human Understanding* (1728), is in part a logic textbook meant to correct 'the false and Pernicious principles in some of our modern Writers of Logic and Metaphysics' (1728: 34). John Locke is the foremost example (see Locke, J. §§3-6). Like John Sergeant, Browne rejects Locke's definition of knowledge as perception of the agreement or disagreement of ideas. 'The Word *Idea*,' he writes, 'should be limited and confined to our simple Sensations only', and to the 'various Alterations and Combinations of them' which he calls 'compounded ideas' (1728: 63). Compounded ideas differ from 'complex notions' or 'conceptions', in which ideas are combined with an immediate consciousness of the mind's operations and affections. Locke's over-broad use of the word 'idea' is, Browne suggests, one sign of his tendency to overestimate the mind's passivity. Another is Locke's denial, in his assault on the syllogism, of 'all true *Illation*, or the *Actual* infering one thing from another' (1728: 422). Instead Locke gives us 'a mere naked *juxta-Position* of Ideas' (1728: 422).

Browne's mature doctrine of analogy is presented in the *Procedure*, but it is spelled out more fully in *Things Divine and Supernatural* (1733). The doctrine rests on a distinction between analogy and metaphor. In each case we use one idea or conception to stand for another. But in analogy there is a 'real resemblance' or 'true correspondence' between the terms; in metaphor the resemblance is merely imagined. In what Browne calls 'human' analogy, each term is known - or knowable - directly; it follows that 'human' analogy can, at least in principle, be paraphrased away. In 'divine' analogy, the second, divine term cannot be known directly. Divine analogy is therefore 'necessary' or inescapable, if things divine are to be understood at all. 'Thus we contemplate things *Supernatural*', as he had written in the *Procedure*, 'not by looking directly *Upward* for any *Immediate View* of them; but as we behold the heavenly Bodies, by casting our Eyes *Downward* to the *Water*' (1728: 425).

KENNETH P. WINKLER

List of works

Browne, P. (1697) *A letter in answer to a book entitled Christianity Not Mysterious; as also to all those who set up for reason and evidence in opposition to revelation & mysteries*, Dublin.(A reply to John Toland's *Christianity Not Mysterious*.)

Browne, P. (1728) *The procedure, extent, and limits of human understanding*, London; repr. New York: Garland, 1976.(Browne's main work, develops an anti-Lockean theory of knowledge.)

Browne, P. (1733) *Things divine and supernatural conceived by analogy with things natural and human*, London; repr. New York: Garland, 1976.(The fullest statement of Browne's doctrine of analogy and metaphor.)

References and further reading

Berkeley, G. (1732) *Alciphron*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, vol. 3, London: Thomas Nelson, 1950.(Berkeley criticizes Browne's doctrine in the Fourth Dialogue, §§17-18, and presents his own account of analogical predication in §21.)

Berman, D. (1982) 'Enlightenment and Counter-Enlightenment in Irish Philosophy', *Archiv für Geschichte der Philosophie* 64: 148-65.(Assesses Browne's place in the history of Irish philosophy in the early eighteenth century.)

Berman, D. (ed.) (1993) *Alciphron in Focus*, London: Routledge.(A convenient edition of selections from Berkeley's *Alciphron* with critical commentary. Includes the Fourth Dialogue and a brief excerpt from Browne's reply.)

Brunner, Emil (1889-1966)

Emil Brunner was one of the most influential Protestant theologians of the twentieth century. He was a minister of the Swiss Reformed Church, a professor at the University of Zurich, and held distinguished lectureships in England, the USA and Japan. He joined the 'dialectical school' early in his career, but tried to rehabilitate natural theology, which led to a rift with Barth. His works were widely read and often served as basic texts in Reformed and Presbyterian seminaries. He rejected the historicist reduction of Christ to a wise teacher figure that was characteristic of neo-Protestantism. He was also critical of modern philosophical anthropologies - as propounded by Marx or Nietzsche, for example - because he felt that they reduced human essence to a single dimension. Only theological anthropology can fully interpret human essence; and of central importance here is the 'I-Thou encounter', whereby the fulfilment of the human 'I' is achieved through a relationship with the divine 'Thou'. Brunner also unfolded an original view on the relation of theology to philosophy. Reason, he argued, is essential for the elucidation and communication of faith. Philosophy, in so far as it indicates the limitations of reason, can serve to prepare us for the revelation of the Absolute.

1 Theology of crisis

Emil Brunner was born in Winterthur in Switzerland. His religious and theological education was marked by three main influences: a socio-eschatological, a historico-critical and a phenomenological. The first of these was mediated through his father, who was acquainted with religious socialism and especially enthralled by Christoph Blumhardt's eschatological message. He also read Hermann Kutter and Leonhard Ragaz and studied Herknew's and Sombart's socio-economic works. At the University of Zurich, Brunner received a solid training in the historico-critical method. His teachers were liberals, and one of them, Walter Köhler, was a disciple of Ernst Troeltsch. In 1911, he spent a semester in Berlin, where he attended Adolf von Harnack's lectures. The phenomenological influence manifested itself when Brunner, searching for a rigorous formulation of Christian belief, read first Kant, and then discovered Edmund Husserl.

After a year in England (1913-14), where he frequented the Christian Labour Movement and the Brotherhood Movement, he received his doctorate in theology in Zurich. His dissertation examined the use of symbol in religious knowledge. Brunner took his theme from Kant and Bergson and his method from Husserl (§§2, 6). As knowledge of a higher world, religious knowledge is inherent in the consciousness of ethical norms. In the wake of Kant (§11), Brunner's theologizing process proceeded from ethics to faith.

The First World War destroyed Brunner's hopes for a better world. Blumhardt's belief in the imminent coming of the kingdom of God, to which Brunner had held, was shattered dramatically. Like Karl Barth, Brunner turned back to the foundations of Christian doctrine and action. He spent a year of research at Union Theological Seminary in New York (1919-20), then, in 1921, defended his *Habilitationsschrift* at the University of Zurich, which led to his appointment there the following year; in 1924, he took the chair of systematic and practical theology. It was at this time that Brunner discovered Barth's *Der Römerbrief (The Epistle to the Romans)* (1919) and joined the 'theologians of crisis' contributing regularly to the journal *Zwischen den Zeiten (Between the Times)* (see Barth, K. §§1-2). He subsequently became a major exponent of 'the theology of crisis'. Like Barth, Brunner found inspiration in Kierkegaard's warning against the accommodation of Christian faith to human culture (see Kierkegaard, S.A. §§4-5). His main concern was to restore the sovereignty of God. Thus he centred theology on the gospel of sin and grace while emphasizing the infinite qualitative distinction between God - the Wholly-Other - and human beings. Brunner denied human beings any natural knowledge of God, whether by reason, feeling or will; the knowledge of God is exclusively supernatural. Faith in revelation is the sole medium for our knowledge of God. Brunner insisted on the primacy of biblical revelation.

Having returned to the supernaturalism of the sixteenth-century Protestant Reformation, Brunner launched a spectacular offensive against neo-Protestantism, or liberalism. *Die Mystik und das Wort (Mysticism and the Word)* (1924) distinguishes Christian faith from the modern understanding of religion, as exemplified by Friedrich Schleiermacher (§7); *Der Mittler (The Mediator)* (1927) claims allegiance to Nicene and Chalcedonian Christology. Brunner defended the divinity of Christ the Mediator against the lives of Jesus presented by the historicizing and psychologizing schools of the nineteenth century. What he most disapproved of in

neo-Protestantism was the reduction of Christ to a wise teacher or a heroic figure, the identification of progress with divine providence, and uncritical confidence in human nature.

2 Eristic and natural theology

In 1929, Brunner published a programmatic article, ‘Die andere Aufgabe der Theologie’ (The Other Task of Theology), which presaged the coming break with Barth. Responding to an epistemological, a missionary and an ethical motive, Brunner’s theology was shifting from the content of revelation (the ‘what’) to the appropriation of revelation (the ‘how’). Whereas expounding God’s revelation in Christ through Scripture defines the first theological task, the second one consists in clearing the path for its appropriation by human beings. Brunner called ‘eristic’ the type of apologetics that stresses the delusion of reason; he claimed to have borrowed this theology from Pascal, Hamann and Kierkegaard. An article entitled ‘Die Frage nach den "Anknüfungspunkt" als Problem der Theologie’ (The Question of the ‘Point of Contact’ as a Theological Problem) (1932) argues in favour of an innate human disposition to the divine. In it, Brunner contends that the Word of God does not encounter a passive, unresponsive being, but a thinking subject, a responsive and responsible person. Brunner called the ‘point of contact’ that which corresponds in epistemology to the human ability to hear and respond to God’s revelation, in anthropology to the infralapsarian image of God, and in ethics to evil conscience.

The rehabilitation of natural theology, though of a different kind, came in 1934. Brunner’s essay ‘Natur und Gnade’ (Nature and Grace) prompted Barth’s uncompromising censure (1935). The dispute culminated in a rupture between the two, though Brunner did not advocate a Roman Catholic type of natural theology. For Brunner, natural theology serves a threefold purpose: to ground personal and social ethics; to provide a philosophical substructure to dogmatics; and to enable Christian preaching and education. It lays the foundation for expositions of revelation. Brunner drew a distinction between subjective natural theology (Roman Catholic) and objective natural theology (Scripture and Calvin). Whereas the former implies the natural knowability of God, the latter consists only in a gnoseologically inoperable general revelation. Sin has disconnected the knowledge of God from general revelation. Hence Brunner argued that God’s special revelation in Christ constitutes the source of knowledge for God’s general revelation in nature, history and human consciousness. Contrary to Roman Catholic natural theology, Brunner’s so-called ‘Christian natural theology’ claims to preserve a Christocentric gnoseology. Notwithstanding this, Barth felt that Brunner’s call for a rehabilitation of natural theology in Christian theology disrupted the post-Reformation emphasis on faith.

3 Theology of encounter

Brunner made ample use of the psychological-anthropological ‘I-Thou’ relationship in all his major works, including *Das Gebot und die Ordnungen (The Divine Imperative)* (1932), *Der Menschen im Widerspruch (Man in Revolt)* (1937), *Wahrheit als Begegnung (Truth as Encounter)* (1938), *Offenbarung und Vernunft (Revelation and Reason)* (1941) and the three-volume *Dogmatik* (1946-60), which unfolds a liberal, modernized version of Reformed theology. The I-Thou relationship was borrowed from the personalist philosophy of existence that Martin Buber (§4) and Ferdinand Ebner professed. In *Man in Revolt*, Brunner argues that all philosophical anthropologies, whether realist (Aristotle), idealist (Plato), pantheist (Stoics) or naturalist (Darwin), provide a distorted image of the human being because of their false interpretation of human essence. The leading philosophical anthropologies of our time, namely those of Darwin, Nietzsche, Marx and Freud, reduce human essence to a single principle and thus fail to account for human reality as a whole. Only theological anthropology, which is biblical, Christocentric and based on faith, correctly explains human essence. Theological anthropology deals with human responsibility. The Word of God is the source of theological anthropology and also makes it knowable by us. Thus a human being is not only a sensitive and thinking subject, but also an ‘I’ whose fulfilment depends on a relationship to the human-divine ‘Thou’. For Brunner, this scheme offers an escape from the traditional alternatives of objective or subjective anthropology. For the truth of faith is neither an objective nor a subjective truth but ‘truth as encounter’. Likewise, revelation is not the disclosure of supernatural facts or truths, but God’s self-revelation. The truth of revelation is thus also ‘truth as encounter’. The knowledge of faith is understood as a synthetic theory that removes the contradictions of human thought and life. Life without Christ is life in contradiction.

4 Faith and reason

From the vantage point of Christian theology, the relationship between faith and reason concerns three things: the use of philosophical terminology, the auxiliary function of philosophical analysis, and the legitimacy of the philosophical endeavour (see Faith §§4-5). Like Barth, Brunner occasionally used philosophical concepts. Divine revelation cannot avoid human reason nor is it fundamentally opposed to reason. For reason is the quintessence of a human being's natural ability to know: it is the capacity to conceive and to formulate ideas. Thus faith needs reason for its rational elucidation; it is a matter of understanding and communication. Consequently, the employment of reason is essential to Christian education. Brunner viewed reason (*Vernunft*) not merely as understanding (*Verstand*) but also as part of the image of God. Whereas understanding is limited to finite and objective knowledge, reason is the power to transcend, to relate to the beyond. The use of philosophical terminology is necessary and legitimate, and may go as far as a partial or global appropriation of a philosophical scheme; an example is the 'I-Thou' relationship, which enables theological discourse to avoid dualism without falling into monism.

The issue of whether philosophy should be regarded as the handmaid of theology was much debated among the 'theologians of crisis'. Rudolf Bultmann (§§2-3) favoured this type of relationship, whereas Barth disapproved of it. In *Erlebnis, Erkenntnis und Glaube (Experience, Insight and Belief)* (1921), his *Habilitationsschrift* for Zurich, Brunner advocated a constructive relationship between theology and a type of philosophy that he labelled 'philosophy of origin' (*Philosophie des Ursprungs*). Critical and deconstructive, the 'philosophy of origin', as developed by Plato, Descartes and Kant, emphasizes and respects the finitude of human nature. It is an inverted science, which postulates the idea of a world beyond, an original and absolute other world, but acknowledges the limitations of this idea. Thus, the 'philosophy of origin' can be a valuable auxiliary to Christian faith. By stressing the inadequate outcome of all rational endeavour, the 'philosophy of origin' prepares human beings to receive the Absolute as ultimate principle and meaning, and thereby clears the path for the 'miracle' of God's Word, namely for revelation and faith. As a preparation for faith, the 'philosophy of origin' functions as a handmaid to theology.

Regarding the legitimacy of philosophical endeavour, Brunner justified and restricted it. In *Revelation and Reason*, he enunciated a law of contiguity which allows for a Christian philosophy. The choice is no longer between theology and philosophy but between existential faith and philosophical and theological reflection. In faith as encounter, faith restores reason's original purpose. Whereas reason provides access to an objective, impersonal, universally valuable truth, faith, as encounter with the divine 'Thou', is given a valid and life-transforming truth. In faith as encounter, the believing/knowing subject does not acquire the knowledge of an objective and objectifiable truth, but becomes God's possession. Finally, faith and reason together build a dialectical relationship similar to the dialectic of Law and Gospel.

JEAN-LOUP SEBAN

List of works

Brunner, E. (1979-) *Emil Brunner: Werke (Collected Works)*, ed. R. Wehrli, Zurich: Theologischer Verlag. (Multiple volumes, still being produced.)

Brunner, E. (1921) *Erlebnis, Erkenntnis und Glaube (Experience, Insight and Belief)*, Tübingen: Mohr. (Brunner's *Habilitationsschrift*, in which he advocates the 'Philosophy of Origin', which can do justice to human finitude and thus fit with theology.)

Brunner, E. (1924) *Die Mystik und das Wort: der Gegensatz zwischen moderner Religionsauffassung und christlichen Glauben, dargestellt an der Theologie Schleiermachers (Mysticism and the Word: The Conflict between Modern Religious Belief and Christian Faith, with Reference to the Theology of Schleiermacher)*, Tübingen: Mohr.(Like Karl Barth, Brunner defined himself to some extent over against Schleiermacher, particularly in this work.)

Brunner, E. (1927) *Der Mittler: Zur Besinnung über den Christusglauben*, Tübingen: Mohr; trans. O. Wyon, *The Mediator: A Study of the Central Doctrine of the Christian Faith*, London: Lutterworth, 1934.(A lengthy presentation of Brunner's broadly orthodox Christology.)

Brunner, E. (1929) 'Die andere Aufgabe der Theologie' (The Other Task of Theology), *Zwischen den Zeiten* 7. (Theology must, *pace* Barth, ask how revelation is humanly appropriated - this is its other task.)

Brunner, E. (1932a) 'Die Frage nach den "Anknüfungspunkt" als Problem der Theologie' (The Question of the 'Point of Contact' as a Theological Problem), *Zwischen den Zeiten* 7.(The question of the human reception of revelation led Brunner to a break with Barth through positing a fundamental human receptivity: the *Anknüfungspunkt* of the title.)

Brunner, E. (1932b) *Das Gebot und die Ordnungen: Entwurf einer protestantisch-theologischen Ethik (The Law and the Rules: Outline of a Protestant Theological Ethics)*, Zurich: Zwingli-Verlag, 4th edn, 1939; 2nd edn trans. O. Wyon, *The Divine Imperative: A Study in Christian Ethics*, London: Lutterworth, 1937.(An ethics consonant with Brunner's focus on the encounter between God and fallen humanity, and Protestant ambiguity about 'law'.)

Brunner, E. (1934) 'Natur und Gnade: zum Gespräch mit Karl Barth', *Theological Existenz heute* 14; trans. P. Fraenkel, 'Nature and Grace', in K. Barth and E. Brunner, *Natural Theology*, London: Geoffrey Bles, 1946. (Brunner suggests a rehabilitation of natural theology as preparatory work for theology. The article prompted, and in English is printed with, Barth's famous published reply, 'Nein!')

Brunner, E. (1935) *Unser Glaube: Eine Christliche Unterweisung*, Bern and Leipzig: Gotthelf Verlag; trans. J.W. Rilling, *Our Faith*, London: SCM, 1949.(A brief and simple guide to Brunner's understanding of credal Christian faith.)

Brunner, E. (1937) *Der Menschen im Widerspruch: die christliche Lehre von wahrem und von wirklichen Menschen (Man in Revolt: The Christian Doctrine of True Man and of Man as He Appears in the World)*, Berlin: Furche-Verlag; trans. O. Wyon, *Man in Revolt: A Christian Anthropology*, London: Lutterworth, 1939. (Despite his strong natural theology, Brunner retained a deeply Protestant view of fallen humanity, best seen in this book.)

Brunner, E. (1938) *Wahrheit als Begegnung*, Zurich: Zwingli-Verlag; trans A.W. Loos, *Truth as Encounter*, London: SCM, 1964.(An earlier, incomplete translation was published as *The Divine-Human Encounter* (London: SCM, 1944), but the original German title captures the 'paradoxical' subject-matter best.)

Brunner, E. (1941) *Offenbarung und Vernunft: die Lehre von der christlichen Glaubenserkenntnis*, Zurich: Zwingli-Verlag; trans. O. Wyon, *Revelation and Reason: The Christian Doctrine of Faith and Knowledge*, London: SCM, 1947.(Deals with broadly epistemological issues, and the foundation of faith in revelation.)

Brunner, E. (1943) *Gerechtigkeit: eine Lehre von den Grundgesetzen der Gesellschaftsordnung (Justice: A Study of the Foundations of the Social Order)*, Zurich: Zwingli-Verlag; trans. M. Hattiger, *Justice and the Social Order*, London: Lutterworth, 1945.(Brunner's constant social-ethical emphasis is most explicit in this work on the nature of society and the justice which must lie at its heart.)

Brunner, E. (1946-60) *Dogmatik*, Zurich: Zwingli-Verlag, 3 vols; trans. O. Wyon (vols 1 and 2) and D. Cairns (vol. 3), London: Lutterworth, 1949-62.(Volume 1: the doctrine of God; volume 2: the doctrines of creation and redemption; volume 3: the doctrines of Church, faith and 'the consummation'.)

Brunner, E. (1947-8) *Christianity and Civilization*, London: Nisbet, 2 vols.(Brunner's Gifford Lectures, originally published in English.)

References and further reading

Barth, K. (1919) *Der Römerbrief*, Bern: Bäschlin; 2nd edn, much changed, Munich: Chr. Kaiser Verlag, 1922; 6th edn trans. E.C. Hoskyns, *The Epistle to the Romans*, Oxford: Oxford University Press, 1968.(Brunner was one of those theologians inspired by Barth's seminal work.)

Barth, K. (1934) 'Nein! Antwort an Emil Brunner', *Theological Existenz heute* 14; trans. P. Fraenkel, 'No! Answer to Emil Brunner', in K. Barth and E. Brunner, *Natural Theology*, London: Geoffrey Bles, 1946.(A famous and blunt rejection of Brunner's natural theology.)

Brunner, H.H. (1986) *Mein Vater und sein Altester: Emil Brunner in seiner und meiner Zeit (My Father and his Oldest Son: Emil Brunner in His and My Time)*, Zurich: Theologischer Verlag.(One of the only biographical works available.)

Buber, M. (1958) *I and Thou*, trans. R. Gregor Smith, New York: Scribner.(Perhaps the central philosophical influence on Brunner, this book proposes the centrality of I-Thou rather than I-It relationships.)

Ebner, F. (1921) *Das Wort und die geistigen Realitäten*.(A personalist influence on Brunner's early development, although less important than Buber.)

Kegley, C.W. (ed.) (1962) *The Theology of Emil Brunner*, New York: Macmillan. (Contains essays by prominent theologians, autobiographical material and a reponse from Brunner, and a bibliography complete to 1962.)

Kutter, H. (1904) *Sie Müssen! Ein Offenes Wort an die christliche Gesellschaft (You Must! An Open Word to Christian Society)*, Berlin: Hermann Walther Verlagsbuchhandlung.(This initiated the religious socialist movement which so influenced Brunner.)

Hesselink, J. (1989) 'Emil Brunner: A Centennial Perspective', *The Christian Century*, Dec. 13.(Discusses the state of Brunner scholarship and the possibilities for a Brunner revival.)

Bruno, Giordano (1548-1600)

Giordano Bruno was an Italian philosopher of nature and proponent of artificial memory systems who abandoned the Dominican Order and, after a turbulent career in many parts of Europe, was burned to death as a heretic in 1600. Because of his unhappy end, his support for the Copernican heliocentric hypothesis, and his pronounced anti-Aristotelianism, Bruno has often been hailed as the proponent of a scientific worldview against supposed medieval obscurantism. In fact, he is better interpreted in terms of Neoplatonism and, to a lesser extent, Hermeticism (also called Hermetism). Several of Bruno's later works were devoted to magic; and magic may play some role in his many books on the art of memory. His best-known works are the Italian dialogues he wrote while in England. In these Bruno describes the universe as an animate and infinitely extended unity containing innumerable worlds, each like a great animal with a life of its own. His support of Copernicus in La Cena de le ceneri (The Ash Wednesday Supper) was related to his belief that a living earth must move, and he specifically rejected any appeal to mere mathematics to prove cosmological hypotheses. His view that the physical world was a union of two substances, Matter and Form, had the consequence that apparent individuals were merely collections of accidents. He identified Form with the World-Soul, but although he saw the universe as permeated by divinity, he also believed in a transcendent God, inaccessible to the human mind. Despite some obvious parallels with both Spinoza and Leibniz, Bruno seems not to have had much direct influence on seventeenth-century thinkers.

1 Life

Giordano Bruno was born in Nola, near Naples, and frequently called himself 'the Nolan'. His baptismal name was Filippo, and he adopted the name 'Giordano' on entering the Dominican Order in 1565. He fled from the Dominicans in 1576 on learning that he was to be accused of heterodox views, and wandered in Italy for a while, supporting himself by private teaching and tutoring. By 1579 he was in Geneva, where the municipal records show that he petitioned against a denial of participation in the Calvinist sacraments. This denial was apparently the result of Bruno's attack on a leading philosopher, which also led to brief imprisonment. Next he went to France, finally settling in Toulouse, where he lectured on Aristotle's *On the Soul*. He was in Paris 1581-3, where his lectures attracted the attention of Henri III. From 1583 to 1585 he was in London, staying at the house of the French ambassador, Michel de Castelnau. Unknown to Castelnau, Bruno was probably a spy, reporting to Sir Francis Walsingham, Elizabeth I's secretary, about the plots of the English Catholics (see Bossy 1991). He had an unfortunate experience in Oxford in 1583, being forced to abandon a series of lectures when a member of the audience pointed out that he was plagiarizing the works of Ficino. When Michel de Castelnau was recalled to Paris at the end of 1585, Bruno went with him. After some problems there, including opposition roused by a disputation in which his view of the universe was defended by a pupil, he turned to the German Protestant universities. In 1586 he went to Marburg, and matriculated at the university there but (as the Rector has recorded) angrily withdrew his name when denied the right to teach philosophy publicly. He then went to Lutheran Wittenberg, where he was allowed to lecture. In 1588 he visited the court of Rudolph II in Prague, and then went to Helmstadt, where he seems to have been excommunicated by the Protestants. In 1590 he was resident in the Carmelite monastery in Frankfurt. At this point he was invited by a Venetian nobleman, Zuan (a form of 'Giovanni') Mocenigo, to go to Venice and teach him the art of memory, and he took up the invitation in the autumn of 1591. In May 1592 Mocenigo, perhaps disappointed at what he had learned, handed Bruno over to the Venetian Inquisition. The following year they handed Bruno over to the Roman Inquisition. A long sequence of interrogations followed. In the end he refused to recant, though the list of final charges is not known. Despite the fact that heretics were normally strangled in prison and burned only in effigy, Bruno was burned to death on February 17, 1600, in the Campo dei Fiori in Rome.

2 Works

Bruno was an extremely prolific writer who wrote more than fifty works and opuscula, of which nearly forty, all from the period 1582-91, are now available in print. He wrote one comedy, *Il Candelaio (The Torch-Bearer)* (1582a), which influenced Molière, and several didactic and critical works. His works on magic, probably written in Helmstadt, were all unpublished during his lifetime. His most significant works include those on the art of memory and Ramon Llull's combinatory method. His first surviving work on these (or any other) issues is *De*

umbris idearum (On the Shadows of Ideas) (1582b), published with *Ars memoriae (On the Art of Memory)*. In 1583 there followed four works in one volume: *Ars reminiscendi (On the Art of Remembering)*, *Explicatio triginta sigillorum (The Explanation of the Thirty Seals)*, *Sigillus sigillorum (The Seal of Seals)*, and a letter to Oxford setting out his claims to be heard. Another work in the same series is *Lampas triginta statuarum (The Lamp of Thirty Statues)* (*c*.1587).

Much of Bruno's fame rests on his six Italian dialogues. The cosmological dialogues are: *La Cena de le ceneri (The Ash Wednesday Supper)* (1584a); *De la causa, principio et uno (Cause, Principle and Unity)* (1584b); *De l'infinito universo et mondi (On the Infinite Universe and Worlds)* (1584c). The moral dialogues are: *Spaccio della bestia trionfante (The Expulsion of the Triumphant Beast)* (1584d), modelled on the satirist Lucian; *Cabala del cavallo Pegaseo con l'aggiunta de l'Asino cillenico (The Kabbalah of the Pegasean Horse with an Appendix on the Cillenican Ass)* (1585a); *De gl'Heroici Furori (The Heroic Frenzies)* (1585b), a sonnet sequence with prose commentaries.

The final version of Bruno's philosophical and cosmological speculations is found in three Latin poems with prose accompaniment, published in 1591: *De triplici minimo et mensura (On the Threefold Minimum and Measure)*; *De monade, numero et figura (On the Monad, Number and Figure)*; *De innumerabilibus, immenso et infigurabili (On the Innumerables, the Immense and the Infigurable)*.

Bruno was never a professional philosopher with a permanent university post, a fact which helps explain both his chaotic style of writing and his exploration of various literary forms. He did not write organized textbooks, and there is little close argumentation or precise definition in his works. On the other hand, particularly in his dialogues, he is witty, bawdy, disputatious and hostile. Women are 'a chaos of irrationality' ([1584b] 1964: 118); his contemporary Patrizi is that 'excrement of pedantry, who has soiled many quires with his Peripatetic Discussions' ([1584b] 1964: 99); Oxford doctors, who know beer better than Greek, are characterized by 'discourteous impoliteness and brazen ignorance' ([1584a] 1975: 142). His invective is balanced by elaborate praise of 'the Nolan', who has 'set free the human spirit and cognition' ([1584a] 1975: 60).

One feature of Bruno's style that seems particularly strange to modern readers, is his elaborate, extravagant, use of verbal descriptions of images, seals and emblems (symbolic devices representing abstract qualities, astral powers and so on). These devices are often arranged in groups of thirty, though *The Expulsion of the Triumphant Beast* is organized around forty-eight constellations, each linked to a vice which is to be expelled and replaced by the contrary virtue.

The influences on Bruno's work are manifold. As a Dominican, he had studied Thomas Aquinas, whom he greatly admired, and Thomistic distinctions and categories still pervade his works. Obviously he knew Aristotle well, though he is generally opposed to his views. He frequently refers to the Presocratics (see Presocratic philosophy), and also to Lucretius, but his general framework is Neoplatonic. Plato and Plotinus are of particular importance, along with the *Corpus Hermeticum*, Ibn Gabirol and Nicholas of Cusa. Bruno also made much use of Ramon Llull, Marsilio Ficino and Agrippa von Nettesheim (§5), especially Agrippa's *De occulta philosophia* (1533).

3 Memory and magic

The art of memory, which grew out of rhetoric, had a long history (see Memory §1). It developed rules for the cultivation of so-called artificial memory. These rules focused on places and images. First, one should imagine a place, such as a house with several rooms; then one should place images in each room. Later, by wandering in thought through the rooms, one would be enabled to remember the topics or things associated with each image. Bruno was trained in this tradition, but when he went to Paris, he became acquainted with Ramon Llull's combinatory art, which involved using letters and figures to represent basic concepts, and writing them on moving wheels in order to make new combinations (see Llull, R. §2). Bruno adapted Llull's technique by employing images and emblems in place of mere letters. On turning the concentric wheels, one could relate the images in various ways and multiply the triggers of memory, as well as the complexity of what was represented.

Bruno's use of images has been linked to magic and Hermeticism (or Hermetism) as well as to the privileging of the imagination in epistemology. On the traditional view, the imagination was taken to be one of the inner senses, and had two functions, the forming of images which help us to retain sense impressions in memory, and the manipulation of images to form images of things never experienced, such as a golden mountain. Given a realist

epistemology, images both reflected the outer world and formed a basis for the operations of the intellect. Bruno seems not to have departed from this account, and while he naturally emphasized the imagination when speaking of artificial memory, his other accounts of the human search for knowledge always subordinate the imagination to the intellect. The noetic process is triggered by sense perception, and involves an ascent beyond the imagination to as much understanding as the human mind can achieve. In *The Heroic Frenzies* Bruno emphasizes that love is the stimulus, and that the final ascent to union with God through understanding is unattainable in this life. There are difficulties with Bruno’s epistemology, in that it is not always clear how much room is assigned to illumination from above, or how it is that human notions (the ‘shadows of ideas’) reflect the divine ideas that structure the world without being innate, but it is not the case that his account of cognition is reducible to his account of the imagination.

The link between magic and the art of memory can be explained by appeal to Bruno’s realist framework, together with the premises that symbols and reality are linked and that knowledge as such has operative force. If external symbols and internal images represent not just the world but the natural forces which pervade it, the magician can achieve mastery of these natural forces by increasing his ability to represent them. Not only that, he can even influence others by projecting images into their minds. While Bruno’s works on magic were all written at the end of his literary activity, he does speak of natural magic (as opposed to black magic) approvingly in *The Expulsion of the Triumphant Beast*. He describes the cult of the ancient Egyptians as ‘magic and divine’ ([1584d] 1964: 239), and says that their ‘magic and divine rites’ provide a way of ascent ‘to the Divinity by means of the same ladder of Nature by which Divinity descends even to the lowest things in order to communicate herself’ ([1584d] 1964: 236).

Frances Yates links both these references to magic and Bruno’s praise of the ancient Egyptian religion to the *Corpus Hermeticum*, a group of writings dating from AD 100 to AD 300, and attributed to the mythical Egyptian sage Hermes Trismegistus. She argues that Bruno’s main aim in all his writings was to preach Hermeticism, and to establish it as the one true religion. Although her thesis is overstated, it does provide a way of reconciling Bruno’s works on memory and magic with his more directly philosophical works (see Hermetism).

4 Cosmology

Bruno is best known for his championship of Copernicus in *The Ash Wednesday Supper*. In *De Revolutionibus Orbium Coelestium (On the Revolutions of the Heavenly Spheres)* (1543), Copernicus had argued both that the earth had its own daily rotation, and that it rotated around the sun (see Copernicus, N.). These theses challenged Aristotelian cosmology, but the force of the challenge was recognized only gradually. It was not until 1616 that *De Revolutionibus* was put on the Index of books forbidden by the Roman Catholic Church, and ironically, it may have been in part Bruno’s defence of Copernicus that led to this result, for he pointed out that Copernicus’ theory was inconsistent with the standard ways of interpreting the Bible at that time. Bruno took Copernican cosmology more seriously and less metaphorically than Yates suggests when she writes ‘The sun-centred universe was the symbol of Bruno’s vision of universal magical religion, inspired by the works of "Hermes Trismegistus"’ (Yates 1982: 219). Nonetheless, it is true that Bruno showed little interest in the mathematical basis of Copernicus’ work. He criticized Copernicus for ‘being more intent on the study of mathematics than of nature’ ([1584a] 1975: 57), a point which ties in with his other attacks on mathematics, and his emphasis on numerology in such writings as *De monade*. Moreover, Bruno got some of the technical details wrong, perhaps because he was drawing on the writings of a French bishop, Pontus de Tyard, who was favourable to Copernicus, but muddled.

In both *The Ash Wednesday Supper* and *On the Infinite Universe and Worlds* Bruno makes a series of cosmological claims that owe much to Lucretius and Nicholas of Cusa. First, the universe is infinite, which means that it can have no centre, though there are many world-systems each of which may have its own centre. Second, these worlds may be inhabited. Third, the stars can be regarded as suns, that is, as self-luminous bodies, and they should not be seen as fixed on spheres. Fourth, the earth is made of the same stuff as the other worlds: there is no difference of kind between the sublunar realm and the heavenly realms, as Aristotelians argued. Finally, the celestial bodies that constitute the universe are ‘intelligent animals’ ([1584a] 1975: 46). Indeed, it is because the earth is animate that it must rotate. It has an ‘innate animal instinct’ ([1584c] 1950: 266), and there is no need to postulate extrinsic movers.

5 Metaphysics

Bruno's main argument for the infinity of the world relied on his view of God as infinite in power and goodness. Goodness diffuses itself, and the only appropriate product of infinite power must itself be infinite. He went on to give an account of the created world in largely Neoplatonic terms. In *De umbris idearum* he claims that all things are ordered and connected, and he presents a fairly standard account of the 'ladder of nature' (*scala naturae*, or the Great Chain of Being), with pure act or infinite unity at the top and matter or infinite number at the bottom. The intelligences or pure spirits come below God, followed by the corporeal world, with soul as an intermediary.

The theme of the ladder of nature goes together with two other themes, the division into three worlds (found in the *Corpus Hermeticum*), and the Neoplatonic scheme of *exitus* and *reditus*, or going out and coming back, ascending and descending. The three worlds are described in *Sigillus sigillorum*. The first world is the *fons idearum* (the fount of ideas) which is God himself as containing the divine Ideas by which the created world is structured; the second is the *mundus ideatus*, the created world; and the third is the human mind which reflects the created world. Just as God's power flows down the ladder of nature, creating the third world through the second, so human contemplation ascends the ladder, knowing the first world through the second.

This Neoplatonic schema was considerably complicated by Bruno's account of matter in *Cause, Principle and Unity*. While he retained Aristotle's account of physical reality in terms of form and matter, he attacked Aristotle's view of prime matter as purely receptive and devoid of form. Rather, matter contains all things within itself, and 'unfolds what it holds folded-up' ([1584b] 1964: 131). Forms both desire and need matter. In language very reminiscent of Ibn Gabirol (§2), Bruno developed his theory by claiming that there are just two substances (apart from the transcendent God, who is the only genuine substance), namely Matter and Form, and these are united. Universal Form is identified with the World-Soul, which through its faculty of intellect serves as efficient cause of the world, and through its status as form, animates and informs the world. As a result, it is both a part of the universe (as form) and not a part (as efficient cause). It was this part of Bruno's theory that led to mistaken accusations of pantheism.

His view of matter has two consequences. First, he rejected the Neoplatonic hierarchy in relation to the created world. One can no longer think of matter as at the bottom of the ladder, for it is indivisible from form and indeed from life: 'there is not the least corpuscle that doesn't contain internally some portion that may become alive' ([1584b] 1964: 87). Second, he abandoned the Aristotelian account of individuals as substances belonging to a species. Neither *humanitas* ('humanity') nor *Socrateitas* ('Socrateity') are substantial forms, that is forms which by union with matter constitute a substance capable of receiving accidents. Instead, Socrates is himself an accident of the one material substance, and so what we take to be an individual is a collection of fleeting accidents.

Bruno also believed in atomism (see Atomism), but his account in *De triplici minimo et mensura* seems to retain the dualism of form and matter. Atoms are the smallest parts of things, and are indivisible, but there are different types of atom. Some atoms are the constituents of material things, other atoms are soul, while a third simple element, the monad, is found rationally in numbers and essentially in all things ([1582-1600] 1962 vol. I, part 3: 140). Indeed, Bruno even calls God the monad of monads (*monadum monas*) ([1582-1600] 1962 vol. I, part 3: 146) because he is one in the fullest sense.

6 Influence

At first, Bruno had little influence, partly because his works were placed on the Index in 1603, but also because some of them were in Italian, while Latin remained the language of science and scholarship well into the seventeenth century. Kepler cites him a number of times, and there seem to be hints of his thought in Spinoza and in Leibniz, but it is difficult to tell whether there was any direct influence. He became popular in Germany in the late eighteenth and early nineteenth centuries, when such thinkers as Schelling and Jacobi praised him for his supposed pantheism. In the later nineteenth century he became an Italian national hero, and was regarded as a martyr for freedom of thought and modern science.

See also: Copernicus, N.; Hermetism; Ibn Gabirol; Neoplatonism; Nicholas of Cusa

E.J. ASHWORTH

List of works

Bruno, G. (1582-1600) *Opera latine conscripta (Collected Works in Latin)*, ed. F. Fiorentino, F. Tocco, H. Vitelli, V. Imbriani and C.M. Tallarigo, Naples: Dom. Morano, Florence: Le Monnier, 1879-91, 3 vols in 8 parts; repr. Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1962.(This edition has only a few textual notes and no indexes.)

Bruno, G. (1582-1600) *Opere italiane (Works in Italian)*, ed. G. Gentile and B. Spampanato, Bari: Laterza, 1907-9, 3 vols; 2nd revised edn, 1925-7; 3rd edn of first 2 vols, *Dialoghi italiani (Dialogues in Italian)*, ed. G. Aquilecchia, Florence: Sansoni, 1958; repr. 1985.

Bruno, G. (1582a) *Il Candelaio (The Torch-Bearer)*, ed. G. Barberi Squarotti, Turin: Giulio Einaudi, 1964.

Bruno, G. (1582b) *Le opere latine. Edizione storico-critica vol. I: De umbris idearum (Works in Latin: Historical-Critical Edition. Vol. I: On the Shadows of Ideas)*, ed. R. Sturlese, Florence: Olschki, 1991.(The first of a proposed full critical edition of Bruno's Latin works.)

Bruno, G. (1582c) *Ars memoriae (On the Art of Memory)*, in F. Fiorentino, F. Tocco, H. Vitelli, V. Imbriani and C.M. Tallarigo (eds) *Opera latine conscripta*, vol. 2, part 1, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1962.

Bruno, G. (1583a) *Ars reminiscendi (On the Art of Remembering)*, in F. Fiorentino, F. Tocco, H. Vitelli, V. Imbriani and C.M. Tallarigo (eds) *Opera latine conscripta*, vol. 2, part 2, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1962.

Bruno, G. (1583b) *Explicatio triginta sigillorum (The Explanation of the Thirty Seals)*, in F. Fiorentino, F. Tocco, H. Vitelli, V. Imbriani and C.M. Tallarigo (eds) *Opera latine conscripta*, vol. 2, part 2, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1962.

Bruno, G. (1583c) *Sigillus sigillorum (The Seal of Seals)*, in F. Fiorentino, F. Tocco, H. Vitelli, V. Imbriani and C.M. Tallarigo (eds) *Opera latine conscripta*, vol. 2, part 2, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1962.

Bruno, G. (1584a) *La Cena de le ceneri*, ed. G. Aquilecchia, Turin: Giulio Einaudi, 1955; trans. S.L. Jaki, *The Ash Wednesday Supper. La Cena de le Ceneri*, The Hague and Paris: Mouton, 1975; trans. E. Gosselin and L. Lerner, *The Ash Wednesday Supper: La cena de le ceneri*, Hamden, CT: Archon Books, 1977.

Bruno, G. (1584b) *De la causa, principio et uno*, ed. G. Aquilecchia, Turin: Giulio Einaudi, 1973; trans. J. Lindsay, *Five Dialogues by Giordano Bruno: Cause, Principle and Unity*, New York: International Publishers, 1964; trans. S.T. Greenburg, *The Infinite in Giordano Bruno with a Translation of his Dialogue 'Concerning the Cause, Principle, and One'*, New York: Octagon Books, 1978.

Bruno, G. (1584c) 'De l'infinito universo et mondi', trans. D.W. Singer, 'On the Infinite Universe and Worlds', in *Giordano Bruno: His Life and Thought*, New York: Henry Schuman, 1950.(Singer's work also contains valuable secondary material: see Singer 1950 for details.)

Bruno, G. (1584d) *Spaccio della bestia trionfante*, ed. M. Ciliberto, Milan: Biblioteca Universale Rizzoli, 1985; English trans. A.D. Imerti, *The Expulsion of the Triumphant Beast*, New Brunswick, NJ: Rutgers University Press, 1964; repr. London and Lincoln, NB: University of Nebraska Press, 1992.

Bruno, G. (1585a) *Cabala del cavallo Pegaseo con l'aggiunta de l'Asino cillenico (The Kabbalah of the Pegasean Horse with an Appendix on the Cillenican Ass)*, in *Opere italiane*, ed. G. Gentile and B. Spampanato, Bari: Laterza, 2nd edn, 1925-7.

Bruno, G. (1585b) *Des fureurs héroïques (De gl'Heroici Furori)*, ed. and trans. P.-H. Michel, Paris: Les Belles Lettres, 1954; English trans. P.E. Memmo, *Giordano Bruno's 'The Heroic Frenzies'*, Chapel Hill, NC: University of North Carolina Press, 1964.

Bruno, G. (*c*.1587) *Lampas triginta statuarum (The Lamp of Thirty Statues)*, in F. Fiorentino, F. Tocco, H. Vitelli, V. Imbriani and C.M. Tallarigo (eds) *Opera latine conscripta*, vol. 3, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1962.

Bruno, G. (1591a) *De triplici minimo et mensura (On the Threefold Minimum and Measure)*, in F. Fiorentino, F. Tocco, H. Vitelli, V. Imbriani and C.M. Tallarigo (eds) *Opera latine conscripta*, vol. 1, part 3, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1962.

Bruno, G. (1591b) *De monade, numero et figura (On the Monad, Number and Figure)*, in F. Fiorentino, F. Tocco, H. Vitelli, V. Imbriani and C.M. Tallarigo (eds) *Opera latine conscripta*, vol. 1, part 2, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1962.

Bruno, G. (1591c) *De innumerabilibus, immenso et infigurabili (On the Innumerables, the Immense and the Infigurable)*, in F. Fiorentino, F. Tocco, H. Vitelli, V. Imbriani and C.M. Tallarigo (eds) *Opera latine conscripta*, vol. 1, parts 1-2, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1962.

Bruno, G. (1957) *Due dialoghi sconosciuti e due dialoghi noti (Two Unknown and Two Famous Dialogues)*, ed. G. Aquilecchia, Rome: Edizioni di Storia e Letteratura.

Bruno, G. (1964) *'Praelectiones geometricae' e 'Ars deformationum' (Lectures on Geometry and the Art of Deformations)*, ed. G. Aquilecchia, Rome: Edizioni di Storia e Letteratura.

References and further reading

Bossy, J. (1991) *Giordano Bruno and the Embassy Affair*, New Haven, CT, and London: Yale University Press.(Referred to in §1. Lively discussion of Bruno's supposed career as a spy, his life and personality.)

Ciliberto, M. (1990) *Giordano Bruno*, Bari: Laterza.(General discussion of Bruno's philosophy, most useful for its chronology of life and works and its full, topically organized bibliography. This work is written in Italian.)

Feingold, M. (1984) 'The Occult Tradition in the English Universities of the Renaissance: A Reassessment', in B. Vickers (ed.) *Occult and Scientific Mentalities in the Renaissance*, Cambridge: Cambridge University Press, 73-94.(A re-evaluation of Bruno at Oxford.)

Gosselin, E.A. (1988) 'Bruno's "French Connection": A Historiographical Debate', in I. Merkel and A.G. Debus (eds) *Hermeticism and the Renaissance: Intellectual History and the Occult in Early Modern Europe*, Washington, DC: The Folger Shakespeare Library, London and Toronto, Ont.: Associated University Presses, 166-81.(Llullism and Copernicanism in France and Bruno's intellectual development.)

Kristeller, P.O. (1964) *Eight Philosophers of the Italian Renaissance*, Stanford, CA: Stanford University Press.(Contains a readable introduction to Bruno's thought.)

Michel, P.-H. (1973) *The Cosmology of Giordano Bruno*, trans. R.E.W. Maddison, London: Methuen.(A reliable account which complements Védrine 1967.)

Singer, D.W. (1950) *Giordano Bruno: His Life and Thought With Annotated Translation of His Work 'On the Infinite Universe and Worlds'*, New York: Henry Schuman.(Good biography, with a useful chronological bibliography of Bruno's writings.)

Spruit, L. (1988) *Il problema della conoscenza in Giordano Bruno (The Problem of Knowledge in Giordano Bruno)*, Naples: Bibliopolis.(A helpful study of Bruno's epistemology in relation to his other doctrines.)

Védrine, H. (1967) *La conception de la nature chez Giordano Bruno (Giordano Bruno's Understanding of Nature)*, Paris: Vrin.(A reliable account which complements Michel 1973.)

Yates, F. (1964) *Giordano Bruno and the Hermetic Tradition*, Chicago, IL, and London: University of Chicago Press.(Referred to in §3. The classic account of Bruno's Hermeticism.)

Yates, F. (1966) *The Art of Memory*, Chicago, IL: University of Chicago Press, London: Routledge & Kegan Paul.(A rich discussion of the art of memory in the medieval and Renaissance periods, with much material on Bruno and on his relation to Ramon Llull.)

Yates, F. (1982) *Lull and Bruno: Collected Essays, Volume I*, London: Routledge & Kegan Paul.(Referred to in §4. A useful spectrum of Yates' views.)

Brunschvicg, Léon (1869-1944)

Brunschvicg occupied a central place in French philosophy during the first part of the twentieth century. In 1909 he became a professor at the Sorbonne, teaching there and at the École Normale Supérieure for the next thirty years. His indefatigable activity, wide curiosity and erudition made him a leading figure of French philosophy. His influence is manifest in the work of Bachelard, Piaget, Guéroult, Nabert, Koyré and Sartre. His most important work lay in the field of the philosophy of mathematics, where (among other things) he introduced French philosophers to the work of Frege and Russell.

1 Life

Léon Brunschvicg was born in Paris in 1869, and educated alongside Louis Couturat at the Lycée Condorcet and the École Normale Supérieure. In 1893 he helped to found the *Revue de Métaphysique et de Morale*, which remains one of the most distinguished French journals. In 1897 he received his doctorate for his dissertation, *La modalité du jugement (The Modality of Judgment)*. He wrote here that ‘ the proof of a single theorem is sufficient to give us the joy and security of certainty, since it implies truth’ (1897: 237), which is the starting-point for his philosophy of mathematics. His subsequent teaching career was spent in Paris; but he died at Aix-les-Bains, where he had taken refuge during the war.

2 *Les étapes de la philosophie mathématique* (The Stages in the Philosophy of Mathematics) (1912)

In his most important work, *Les étapes de la philosophie mathématique (The Stages in the Philosophy of Mathematics)* (1912), Brunschvicg sets out the principal stages of philosophical reflection on mathematics. What fundamentally concerns him here is the problem of truth: he writes in the preface (unfortunately omitted from the 1972 edition) that he aims to resolve this problem by ‘ a meditation on the discipline which has employed the greatest scrupulousness and subtlety in its search for the truth’ (1912: xi). Brunschvicg’s concern to elucidate the underlying commitment within mathematics to a conception of truth can be contrasted with the conventionalism of his older contemporary Henri Poincaré who looked primarily to the practical achievements of mathematics (see Conventionalism §1).

Brunschvicg undertakes a massive survey of the evolution of mathematical thought, from the first attempts of primitive peoples to develop a number system to the most advanced results of the modern age. Among the most important discussions are those devoted to classical antiquity (which deal with the number theories of Pythagoras and Plato and with Aristotle’s logic, and that of the work of Descartes (concerning his analytic geometry and his conception of a universal mathematics). Brunschvicg also sets out the philosophical account of mathematics to be found at the start of the modern period in the works of Kant and Comte, who both believed that mathematics was a completed science. This account was overturned during the nineteenth century, however, when profound changes in mathematical techniques were introduced, and new lines for research opened up in all directions. In discussing this, Brunschvicg devotes a long chapter to the development of modern logical theory - to Boole, Peano, Frege and, above all, Russell.

The final chapters are devoted to drawing the moral of this long historical inquiry. Brunschvicg here clarifies his own point of view, and presents a qualified form of ‘intellectualism’ in his final chapter on ‘The reaction against "mathematism": the sense of intellectualism in mathematics’. Up to now, he claims, the philosophy of mathematics has lacked a proper appreciation of the nature of truth in mathematics. In accepting that the logical order of exposition reverses the direction of the psychological order of discovery, philosophers of mathematics admit implicitly that the concern for rigour that is characteristic of reasoning is foreign to, indeed opposed to, the discovery of mathematical truth. Thus their conclusion should be that the logical formalization of mathematics makes no difference to its truth. So even in mathematics one should bear in mind the two distinctive marks of the intellect - ‘an indefinite capacity for progress’ joined with ‘a perpetual disquiet about verification’. As such, mathematics represents one of the most powerful and lasting achievements of the human genius; it reveals to us the capacities of the human intellect, and should be as much a foundation for our knowledge of the mind as it is for the natural sciences. ‘ The activity of the mind has been free and productive only since the epoch when mathematics brought to mankind the true standard of truth’ ([1912] 1972: 577).

3 Other works

Brunschvicg's numerous writings are diverse. One area of interest was the history of philosophy, especially seventeenth-century French philosophy. He produced the standard edition of Pascal's *Pensées et opuscules* (1897); wrote *Spinoza et ses contemporains (Spinoza and His Contemporaries)* (1923); and in 1942 completed *Descartes et Pascal: lecteurs de Montaigne (Descartes and Pascal: Readers of Montaigne)* (1944). In an earlier work on a much grander historical scale, entitled *Le Progrés de la conscience dans la philosophie occidentale (The Progress of Consciousness in Western Philosophy)* (1927), he had opened up wider and bolder lines of thought.

Brunschvicg also wrote on the philosophy of science. He took the view that the development of science was not autonomous: philosophy is not to be separated from science and should not pretend to a superior or distinctive kind of truth. This line of thought, which continues that of *Les étapes de la philosophie mathématique*, is presented in *L'expérience humaine et la causalité physique (Human Experience and Physical Causality)* (1922) and *La physique du vingtième siècle et la philosophie (Philosophy and Twentieth-Century Physics)* (1936). For Brunschvicg, the development of science reveals the work of the human intellect and shows reason at work; as Plato held, the philosopher or the scientist should be able to concern himself with truth in a completely disinterested spirit, following to the limit the movement of commitment which, from Copernicus to Einstein, has manifested the character of the human intellect. In the same spirit, the philosopher and scientist should seek to bring together the pursuit of truth with that of beauty and the good. Grasped in its essence, science should aim to be what it was for Plato, Descartes and Kant: our way of gaining access to the truths of the human spirit.

See also: French philosophy of science

MAURICE LOI

List of works

None of these works have been translated into English.

Brunschvicg, L. (1897) *La modalité du jugement (The Modality of Judgment)*, Paris: Alcan.(Thesis defended at the Sorbonne on 29 March 1897. Outlines the programmes of subsequent books.)

Brunschvicg, L. (1897) *Pensées et opuscules de Blaise Pascal*, intro. L. Brunschvig, Paris: Hachette; repr. 1945. (Highly commended by the Académie française.)

Brunschvicg, L. (1912) *Les étapes de la philosophie mathématique (The Stages in the Philosophy of Mathematics)*, Paris: Alcan; revised edn, Paris: Blanchard, 1972.(A description of the different stages in the mathematical philosophical process.)

Brunschvicg, L. (1922) *L'expérience humaine et la causalité physique (Human Experience and Physical Causality)*, Paris: Alcan.(A description of contemporary science. Book XV deals with theories of relativity.)

Brunschvicg, L. (1923) *Spinoza et ses contemporains (Spinoza and His Contemporaries)*, Paris: Alcan. (Biographical and comparative.)

Brunschvicg, L. (1924) *Le génie de Pascal (The Genius of Pascal)*, Paris: Hachette.(Deals more thoroughly with certain aspects of Pascal's work already discussed in earlier works.)

Brunschvicg, L. (1927) *Le progrés de la conscience dans la philosophie occidentale (The Progress of Consciousness in Western Philosophy)*, Paris: Alcan.(Gives an account of the development of consciousness.)

Brunschvicg, L. (1931) *De la connaissance de soi (On Self-Conciousness)*, Paris: Alcan.(A description of the necessity for self-knowlede.)

Brunschvicg, L. (1936) *La physique du vingtième siècle et la philosophie (Philosophy and Twentieth Century Physics)*, Paris: Hermann.(A complement to *L'expérience humaine et la causalité physique*.)

Brunschvicg, L. (1944) *Descartes et Pascal, lecteurs de Montaigne (Descartes and Pascal: Readers of Montaigne)*, New York: Brentano's.(Places an emphasis on the literary aspects of Descartes and Pascal.)

References and further reading

Boirel, R. (1964) *Brunschvicg*, Paris: Presses Universitaires de France.(An account of Brunschvicg's life and works, with an exposition of his philosophy.)

Bryce, James (1838-1922)

James Bryce, British statesman and writer, combined a distinguished public life with scholarship in history, politics and law. As a jurist his interest lay in historical jurisprudence, but he is best remembered for his comparative politics. He contributed significantly to democratic political theory and to a liberal-historicist approach in philosophy of law.

Bryce was born in Belfast in 1838, of Ulster Scottish descent. He was educated at Glasgow University and Oxford University. Although called to the Bar in 1867, he preferred scholarship, travel and public service. As Regius Professor of Civil Law at Oxford (1870-93), he reinvigorated the study of Roman law in Britain and helped to reform legal education. He entered Parliament as a Liberal MP in 1880 and held several ministerial offices between 1886 and 1906. He was Ambassador to the USA (1907-13), was elevated to the peerage as Viscount Bryce of Dechmont in 1914, and remained active in public life until his death.

Bryce diffused his energies widely and was a prolific writer and scholar as well as a man of affairs. In *The Holy Roman Empire* (1864) he propounded a thesis of historical unity stretching from ancient Rome to nineteenth-century Germany. In other works he portrayed England's legal system as emulating, in the modern world, Roman law in the ancient, and emphasized an Anglo-American affinity. His *The American Commonwealth* (1888) was a monumental treatise designed to supersede Tocqueville's more impressionistic account, and incidentally testified to the success of democracy and of federalism. It was an instant classic. Bryce wrote influentially on the classification and characteristics of constitutions and on comparative politics, for example in *Modern Democracies* (1921). His studies of political systems were determinedly empiricist and his research was invariably supplemented by contacts, conversation and travel: one modern critic suggested that his 'genius largely consisted of an infinite capacity for taking trains'.

Studies in History and Jurisprudence (1901) was perhaps his most representative work. As a typical nineteenth-century liberal, he espoused *laissez-faire* economics, favoured religious and political liberty, distrusted state intervention, but was not free from apprehensions concerning majoritarian democracy. As a historian, he implied biological laws of development, and cannot be acquitted of what Popper was to term historicism (see Popper, K.R. §4). In his philosophy of law, he dismissed natural law theories as a priori and speculative, but faulted aspects of the analytical jurisprudence of Austin and Bentham as being not only historically indefensible but also inapplicable to contemporary societies (see Austin, J.; Bentham, J.). In juristic interests and endeavour he was closer to the comparative method and historical jurisprudence of Maine.

However, he lacked Maine's inclination to theorize. Harold Laski's observation on Bryce's 'insatiable appetite for facts and grotesque inability to weigh them' was unkind but not entirely unfair. In a changing world which he struggled to comprehend, the public servant was garlanded with honours, but the scholar restricted himself to fact-gathering and classification or lofty pronouncements on public affairs.

See also: Common law; Jurisprudence, historical; Law, philosophy of; Rule of law

COLIN MUNRO

List of works

Bryce, J. (1864) *The Holy Roman Empire*, London: Macmillan, 4th edn, 1875.(A history of the Romano-Germanic Empire as the centrepiece of European civilization.)

Bryce, J. (1888) *The American Commonwealth*, London and New York: Macmillan, 4th edn, 1914.(An account of, and reflections on, the US constitutional and political system as it had developed by the mid-nineteenth century.)

Bryce, J. (1901) *Studies in History and Jurisprudence*, Oxford: Oxford University Press.(Essays on history, Roman law, English law, constitutional theory, ancient and modern political systems, and jurisprudential concepts.)

Bryce, J. (1921) *Modern Democracies*, London: Macmillan. (Comparative treatise on democratic theory and institutions, with studies of some systems.)

References and further reading

Fisher, H.A.L. (1927) *James Bryce*, London: Macmillan.(This is the standard biography, readily accessible to the general reader.)

Tulloch, H. (1988) *James Bryce's 'American Commonwealth'*, Woodbridge: The Royal Historical Society/The Boydell Press.(Gives an account of Bryce's political thought in the context of his intellectual biography.)

Buber, Martin (1878-1965)

Martin Buber covered a range of fields in his writings, from Jewish folklore and fiction, to biblical scholarship and translation, to philosophical anthropology and theology. Above all, however, Buber was a philosopher, in the lay-person's sense of the term sense: someone who devoted his intellectual energies to contemplating the meaning of life.

Buber's passionate interest in mysticism was reflected in his early philosophical work. However, he later rejected the view that mystical union is the ultimate goal of relation, and developed a philosophy of relation. In the short but enormously influential work, Ich und Du (I and Thou). Buber argued that the I emerges only through encountering others, and that the very nature of the I depends on the quality of the relationship with the Other. He described two fundamentally different ways of relating to others: the common mode of 'I-It', in which people and things are experienced as objects, or, in Kantian terms, as 'means to an end'; and the 'I-Thou' mode, in which I do not 'experience' the Other, rather, the Other and I enter into a mutually affirming relation, which is simultaneously a relation with another and a relation with God, the 'eternal Thou'.

Buber acknowledged that necessity of I-It, even in the interpersonal sphere, but lamented its predominance in modern life. Through his scholarly work in philosophy, theology and biblical exegesis, as well as his translation of Scripture and adaptations of Hasidic tales, he sought to reawaken our capacity for I-Thou relations.

1 Life

Martin Mordechai Buber was born in Vienna. Following the divorce of his parents when he was three years old, he spent most of his childhood in Lemberg, Galicia, with his grandparents Solomon and Adele Buber. This separation from his parents, particularly his mother, had a profound effect on the subsequent development of Buber's philosophy of relation. His abiding childhood memory was not of his mother leaving him - he expected her to come for him any day - but of being told at the age of four by a baby-sitter that his mother would never return. 'I suspect,' he later wrote, 'that everything I experienced of genuine encounter during the course of my life had its origin in that moment on the balcony' ([1960b] 1972: 18-19).

At the age of fourteen, Buber returned to his father's house, where he lived with his father and step-mother. He rejected the traditional Judaism of his father and grandparents, and became increasingly interested in philosophy, particularly the works of Kant and Nietzsche. In 1896 Buber enrolled as a student of philosophy at the University of Vienna. Two years later he moved to the University of Leipzig and joined the Zionist movement (see Zionism). Buber was still not interested in the religious tradition he had rejected as an adolescent, but he devoted his energies to the renewal of Jewish culture.

As a delegate to the Third Zionist Congress in 1899, Buber spoke on behalf of the Propaganda Committee, but he used the platform to emphasize the importance of education rather than propaganda. In 1901 he was appointed editor of the weekly Zionist publication *Die Welt*, in which he again stressed the need for a Jewish cultural renaissance. Later in the same year he became a member of the Zionist Democratic Fraction which was opposed to the programme of Theodor Herzl and, at the Fifth Zionist Congress, he resigned as editor of *Die Welt*. Shortly after the Congress, Buber withdrew from political affairs. From about 1903 onwards, he immersed himself in the study of Hasidism. Although initially attracted to the literary qualities of the Hasidic tales, he gradually developed an appreciation of their spiritual content and took it upon himself to communicate the message of Hasidism to the assimilated Jews of western Europe, and to humanity at large.

During the years 1906 to 1911, Buber attended lectures at the University of Berlin, especially those given by Wilhelm Dilthey. Dilthey's hermeneutic theory was an important influence on Buber's approach to interpretation, as manifested both in his Hasidic writings and in his later work of biblical translation and commentary (see Hermeneutics). Buber resumed public life in 1909. He exerted a profound influence on Jewish youth in central Europe through his lectures and publications. In 1916 he founded *Der Jude*, a monthly publication which promoted the Jewish cultural renaissance. Buber's Zionism continued to diverge from the mainstream. He sought complete equality and cooperation between Jews and Arabs and believed that Palestine could become the shared homeland of two autonomous peoples. His political philosophy also diverged from the dominant conception of socialism. He

believed that a socialist society could never come about through the mechanism of the state; rather, it depended on a renewal of relationships among individuals.

Buber's reflections on the relations between people developed into his most famous work, *Ich und Du*, which was published in 1923. Although he continued to write for over forty years, *Ich und Du* is unquestionably his masterpiece. 'It is the vessel into which he pours the learning and wisdom acquired over the years' and 'everything that he wrote afterwards can be traced back to it' (Vermes 1994: 27). In 1925, Buber took on a project that would occupy him for decades: a new German translation of the Hebrew Bible. Begun in collaboration with Franz Rosenzweig, the work was continued by Buber on his own after Rosenzweig's premature death in 1929. The final volume of *Die Schrift* was published in 1961. Buber did not actively work on the translation between 1932 and 1949, but he continued to reflect on the Hebrew Bible, publishing works of biblical criticism and theology, including *Königtum Gottes (Kingship of God)* (1932), 'Die Frage an den Einzelnen' ('Question to the Single One') (1936), *Torat ha-Neviim (The Prophetic Faith)* (1940), and *Moshe (Moses)* (1946).

In 1930, Buber was appointed professor of religion at the University of Frankfurt. He retained this position until 1933, when, following the Nazi rise to power, he was forced to leave the university. In the same year he became director of the Central Office for Jewish Education and head of the Jüdisches Lehrhaus in Frankfurt. He travelled throughout Germany, lecturing and teaching, until 1935, when he was prohibited from speaking at Jewish gatherings. Buber moved to Jerusalem in 1938, when he was appointed to the newly created chair of social philosophy at the Hebrew University. He taught there until his retirement in 1951. After retiring Buber continued to write, and he lectured extensively abroad. In his later works he wrestled with the themes of good and evil and the suffering of the innocent, publishing *Bilder von Gut und Böse (Good and Evil)* (1952) and *Gottsfinsternis (Eclipse of God)* (1953).

2 Early writings

Buber's period of intense research into Hasidism resulted in two books of Hasidic legends: *Die Geschichten des Rabbi Nachman (The Tales of Rabbi Nachman)* (1906) and *Die Legende des Baalschem (The Legends of the Baal-Shem)* (1908). Influenced by both Nietzsche and Dilthey, Buber did not simply translate the Hasidic tales, but retold them, attempting to renew the old stories for his generation. Although determined to retain the original spirit of the tales, he believed, following Dilthey's hermeneutics, that his empathy with the Hasidic masters gave him licence to edit and embellish their stories. Buber's interpretation of Hasidism was criticized by scholars, notably Gershom Scholem.

Buber's early enthusiasm for mysticism is also evident in *Daniel: Gespräche von der Verwirklichung (Daniel: Dialogues on Realization)* (1913), the main theme of which is the quest for meaning in human life. Although many of the observations in *Daniel* prefigure the more mature *Ich und Du*, the earlier work specifies a form of mystical union as the ultimate aim of relation. The eponymous Daniel urges his companion to come to know a tree not by observing it, but by identifying with it, to the extent that she feels its bark to be her own skin.

3 *Ich und Du*

In *Ich und Du*, the mystical ideal of union is replaced by a dialogical notion of relation. The relation between a person and a tree is again presented as an example, but this time the aim is not to identify with the tree, but to encounter it, entering into dialogue with it. Dialogue with a tree? Many readers find this notion mysterious, if not mystical. Yet it is essential to note that, according to Buber, dialogical relation is not limited to the interpersonal sphere. The thesis of *Ich und Du* is that there are two fundamental modes of relating to others. Buber names these two fundamental attitudes 'I-Thou' and 'I-It':

> To man the world is twofold, in accordance with his twofold attitude. The attitude of man is twofold, in accordance with the twofold nature of the primary words which he speaks. The primary words are not isolated words, but combined words. The one primary word is the combination *I-Thou*. The other primary word is the combination *I-It*; wherein, without a change in the primary word, one of the words *He* and *She* can replace *It*. Hence the *I* of man is also twofold. For the *I* of the primary word *I-Thou* is a different *I* from that of the primary word *I-It*.

(1923: 3)

Buber explores the dual nature of human existence through the pronouns that a person speaks. He insists that the 'I' does not exist independently of the others it addresses; 'there is no *I* taken in itself, but only the *I* of the primary word *I-Thou* and the *I* of the primary word *I-It*' (1923: 4). The world of It is the world of discrete objects; the corresponding I is the self who experiences something, and this 'something' can as easily be a person as a thing. In a clear departure from his earlier mysticism, Buber also insists that there is no fundamental difference between 'inner' and 'outer' experiences - both belong to the world of It.

The I-Thou, by contrast, is a reciprocal relation in which there is no experience of the Other but, rather, a mutual encounter, in which each party affects and affirms the Other. It can only take place if 'the primary word I-Thou' is spoken 'with the whole being'. The relation to the Thou is direct, not mediated by the I's concepts, preconceptions, or intentions. The encounter is lived as a discovery of the uniqueness and concrete presence of the Other *qua* Other and, in a sense, as revelation, for Buber insists that 'in each *Thou* we address the eternal *Thou*' (1923: 6).

Buber discusses three spheres in which the I-Thou relation is possible: life with nature, life with people, and life with 'spiritual beings' (*geistige Wesenheiten* or, as Wood translates it, 'forms of spirit' (1969: 43)). Buber is not concerned here with mystical beings, but with aesthetic inspiration. He asserts that the true artist does not express himself, but responds to an encounter with a 'form which desires to be made through him into a work' (1923: 9).

Of the three spheres, only that of inter-human relations permits a literal dialogue of I with Thou. However, Buber employs the linguistic metaphor of 'saying Thou' to discuss relations in all three spheres. He notes that when we address creatures as Thou, 'our words cling to the threshold of speech' and that, in addressing forms of spirit as Thou, we speak with our being, not with our lips (1923: 6). Linguistic terms in *Ich und Du*, including 'word', 'language', 'dialogue' and 'speech', are primarily used metaphorically to convey the qualities of presence, dynamism and reciprocity which characterize the I-Thou relation (Kepnes 1992: 31); indeed, according to Kohanski's reading of Buber, 'the very act of turning to another in relation is an act of speaking, even when not a word is uttered between them' (1982: 268).

I-Thou is clearly the privileged mode of existence. For Buber, 'all real living is meeting [the *Thou*]' (1923: 11). But he also recognizes that I-It, the ordinary mode of existence, is indispensable. In *Ich und Du* he does not call for the abandonment of I-It, which would be impossible, but aims rather to re-awaken the reader's potential for true relation.

4 Relationship with Jewish philosophy

Ich und Du is not a work of Jewish philosophy in the sense of 'philosophy of Judaism'; it deals with universal philosophical themes and is intended for a broad audience. Nevertheless, the philosophy expounded in its pages shares many key features with those of other twentieth-century Jewish philosophers, notably Hermann Cohen, Franz Rosenzweig and Emmanuel Levinas. The conception of the 'I-Other' relationship within these Jewish thinkers' writings differs markedly from the dominant view in modern western philosophy - epitomized by Sartre's exclamation 'Hell is the Other!' - that this relationship is fundamentally antagonistic. Buber, Cohen, Rosenzweig and Levinas all reject this view of the I-Other relationship. They see a positive, welcoming response to the Other as the *sine qua non* of authentic selfhood.

According to these thinkers, the 'I' emerges from addressing the 'you' (or 'thou'). For each of them, in different ways, the interpersonal relationship is also the site of transcendence. In Buber's terms, 'in each *Thou* we address the eternal *Thou*'; in Levinas', the Other's face 'is the manifestation of the height in which God is revealed' (1961: 79).

Nevertheless, despite the resemblances between these thinkers, both Rosenzweig and Levinas criticized *Ich und Du*. Levinas took issue with Buber's characterization of the I-Thou. One of his major criticisms was that Buber 'understood the Thou primarily as partner and friend and thus gave primacy to a relationship of reciprocity' (see Bernasconi 1988: 105), in contrast with Levinas's own emphasis on the asymmetry of the 'Je-Vous', or 'face-to-face' relation, through which I become aware of my inescapable responsibility for the Other, without regard to the Other's responsibility for me.

Commenting upon the galley proofs of *Ich und Du*, Rosenzweig wrote to Buber: 'In your setting up the I-It, you

give the I-Thou a cripple for an opponent. True, this cripple rules the modern world: however, this does not change the fact that it is a cripple' (see Horwitz 1988: 227).

5 Theology and hermeneutics

Although Buber did not alter *Ich und Du* in response to Rosenzweig's criticism, Rosenzweig's 'New Thinking' was a major influence on the development of Buber's subsequent work, particularly in the areas of hermeneutics and theology. Steven Kepnes (1992) has traced the development of Buber's hermeneutics and theology from his early pantheistic, romantic period, through the dialogical thought of *Ich und Du*, to Buber's later version of narrative theology, which both transcends and incorporates the insights of *Ich und Du*. Buber's later work on the Bible continued to emphasize moments of encounter, both with God and between human beings, but his equally strong interest in the Bible as a source for the common memory of the Jewish people reveals the influence of Rosenzweig's notions of creation, revelation and redemption. In Buber's later thought, a Jew reads the Bible not simply to facilitate an I-Thou encounter with the text, but to discover in the biblical narratives resources for addressing the reader's own situation.

From the early 1940s, Buber's reflections on the Bible centred, unsurprisingly, on the theme of innocent suffering. He drew upon Job, Deutero-Isaiah and the Psalms in his attempt to express the suffering and fragile hope of his generation, describing the Holocaust era as a time of the 'eclipse of God'. Buber attempted no theodicy. He sought rather to express the Job-like faith of those who await the end of the eclipse:

> Do we stand overcome before the hidden face of God like the tragic hero of the Greeks before faceless fate? No, rather even now we contend… with God …. In such a state we await His voice, whether it comes out of the storm or out of a stillness that follows it. Though his coming appearance resemble no earlier one, we shall recognize again our cruel and merciful Lord.
>
> (1952b: 225)

See also: Holocaust, the; Jewish philosophy, contemporary; Hasidism

TAMRA WRIGHT

List of works

Buber, M. (1962-4) *Werke*, vol. 1, Munich and Heidelberg: Kosel-Verlag and Verlag Lambert Schneider. (Contains all of Buber's major philosophical works.)

Buber, M. (1906) *Die Geschichten des Rabbi Nachman; Ihm nacherzählt von Martin Buber*, Frankfurt: Rütten und Loening; trans. M. Friedman, *The Tales of Rabbi Nachman*, Atlantic Highlands, NJ: Humanities Press International, 1988.(The first fruit of a five-year period of research into Hasidism, this book is not a simple anthology of tales and legends which Rabbi Nachman of Bratslav taught to his disciples, but Buber's own retelling of the stories. It opens with 'Jewish Mysticism', an essay on Hasidism.)

Buber, M. (1908) *Die Legende des Baalschem*, Frankfurt: Rütten und Loening; trans. M. Friedman, *The Legends of the Baal-Shem*, New York: Schocken, 1969.(This work established Buber's reputation as a first-class German writer. It is a free retelling of the legendary life of Rabbi Israel ben Eliezer (1700-60), the founder of the Hasidic movement, with an introductory essay on 'The Life of the Hasidim', in which Buber outlines his interpretation of Hasidic thought.)

Buber, M. (1913) *Daniel: Gespräche von der Verwirklichung*, Leipzig: Insel-Verlag; trans. M. Friedman, *Daniel: Dialogues on Realization*, New York: Holt, 1964.(This collection of five dialogues on the themes of direction, reality, meaning, polarity and unity is Buber's first work of general philosophy. Since Buber later rejected some of its teachings, the work is primarily of importance for readers interested in the development of his thought.)

Buber, M. (1923) *Ich und Du*, Leipzig: Insel Verlag; trans. R.G. Smith, *I and Thou*, New York: Charles Scribner's Sons, 2nd edn, 1958; trans. W. Kaufman, *I and Thou*, New York: Charles Scribner's Sons, 1970. (The page numbers used in references to the work in this entry are those from the Smith translation; Kaufman's translation substitutes 'you' for 'thou' everywhere except the title. *Ich und Du* is Buber's most famous and influential work, in which he describes two modes of relating to others: 'I-Thou' and 'I-It'. The first is a reciprocal, mutually affirming encounter between self and Other; it is the route to authentic selfhood. The second designates the commonplace attitude of using or experiencing an object. While recognizing that the I-It mode of relation is indispensable, even in the realm of interpersonal relations, Buber argues that a renewal of

society can take place only through a reawakening of our potential for I-Thou encounters.)

Buber, M. (1932) *Königtum Gottes*, Berlin: Schocken; trans. R. Scheimann, *The Kingship of God*, Atlantic Highlands, NJ: Humanities Press International, 1988.(The first of Buber's purely scholarly studies of the Hebrew Bible, which are mainly concerned with the theme of messianism. Buber interprets the covenant between God and Israel as the means by which the kingship of God is to be actualized in communal life.)

Buber, M. (1936) 'Die Frage an den Einzelnen', trans. R.G. Smith, 'The Question to the Single One', in M. Friedman (ed.) *Between Man and Man*, New York: Macmillan, 1965.(Here, Buber responds sensitively yet critically to Kierkegaard's attempt to prepare himself for a true relation with God by renouncing both marriage and the body politic. Buber himself rejects the path of renunciation, and holds that 'the man who loves God and his companion in one… receives God for his companion'.)

Buber, M. (1942) *Torat ha-Neviim*, Tel Aviv: Bialik; trans. C. Witten-Davis, *The Prophetic Faith*, New York: Harper & Row, 1960.(Considered by many scholars to be Buber's finest work of biblical exegesis, this study examines the role of God's 'servants' - Job, Abraham, Moses, David and Isaiah - as partners in the revelation of the God of history. Buber rejects traditional interpretations which identify the suffering servant of Deutero-Isaiah with either corporate Israel or Christ. Instead, he sees the servant as a perfected human being whose task is to bring Israel back to the service of God, thus heralding the beginning of God's kingdom on earth.)

Buber, M. (1946) *Moshe,* Jerusalem: Schocken; trans. *Moses: The Revelation and the Covenant*, New York: Harper & Row, 1958; repr. Atlantic Highlands, NJ: Humanities Press International, 1988.(In this work Buber continues the exposition of Hebrew messianism begun in *Königtum Gottes*, and develops his approach to the Bible as a historical account of a particular people's response to moments of I-Thou encounter with God.)

Buber, M. (1946, 1947) *Or ha-Ganuz - Sippurey Hasidim*, Jerusalem and Tel Aviv: Schocken, 2 vols; trans. O. Marx, *Tales of the Hasidim: The Early Masters and Tales of the Hasidim: The Later Masters*, New York: Schocken, 1995.(These two collections of legendary anecdotes about the great Hasidic masters are of importance for tracing the development of Buber's hermeneutics. In the introduction to the first volume, Buber criticizes his earlier collections of Hasidic legends for being too free. The anecdotes in these two volumes are related without embellishment or imaginative additions.)

Buber, M. (1947) *Netivot be-Utopia*, Tel Aviv: Am Oved; German translation, *Pfade in Utopia*, Heidelberg: L. Schneider, 1950; English trans. R.F.C. Hull, *Paths in Utopia*, New York: Macmillan, 1988.(An important contribution to both social philosophy and the history of socialism, this work includes chapters on Proudhon, Marx, Lenin and other utopian thinkers. Buber also discusses a variety of experiments in collective living, including the Israeli *kibbutz*, which he describes as 'an experiment that did not fail'.)

Buber, M. (1950) *Zwei Glaubenweisen*, Zurich: Manesse; trans. N. Goldhawk, *Two Types of Faith*, New York: Macmillan, 1986.(In this controversial study of Jesus and Paul, Buber presents Jesus' conception of faith as *emunah*, a Hebrew term which, Buber notes, implies both 'fidelity' and 'trust'. For Paul, by contrast, faith is cognitive and entails belief in a proposition.)

Buber, M. (1951) 'Distance and Relation', *Hibbert Journal* 49; trans. R.G. Smith, in M. Friedman (ed.) *The Knowledge of Man: A Philosophy of the Interhuman*, Atlantic Highlands, NJ: Humanities Press International, 1988.(This essay is the most important exposition of Buber's mature social anthropology. Buber analyses the movements of 'setting at a distance' and 'entering into relation' which underlie both the 'I-It' and 'I-Thou' relations.)

Buber, M. (1952a) *Bilder von Gut und Böse*, Cologne: Hegner; trans. R.G. Smith, *Good and Evil*, New York: Charles Scribner's Sons, 1952.(Buber develops an anthropological analysis of evil based on both biblical texts and Zoroastrian myths.)

Buber, M. (1952b) *An Der Wende*, Cologne: Hegner; English translation ed. N. Glatzer, *On Judaism*, New York: Schocken, 1972.(A collection of Buber's speeches on Jewish themes, delivered in Germany and central Europe from 1909 to 1918, and the four lectures he gave in America in 1951-2. It includes 'Der Dialog zwischen Himmel und Erde', translated in the Glatzer edition by E. Jospe as 'The Dialogue Between Heaven and Earth', which is of particular importance for readers interested in Buber's response to the Holocaust.)

Buber, M. (1953) *Gottsfinsternis*, trans. M. Friedman, E. Kamenka, N. Guterman and I.M. Lask, *Eclipse of God: Studies in the Relation Between Religion and Philosophy*, New York: Harper & Row; repr. with introduction by R. Seltzer, Atlantic Highlands, NJ: Humanities Press International, 1988.(In this collection of nine essays on religion and philosophy, Buber criticizes contemporary thinkers, including Nietzsche, Bergson, Heidegger, Sartre and Jung, who have denied transcendence and thus contributed to the 'eclipse of God'.)

Buber, M. (1954) 'Elemente des Zwischenmenschlichen', in *Die Schriften über das dialogische Prinzip*, Heidelberg: L. Schneider; trans. R.G. Smith, 'Elements of the Interhuman', in M. Friedman (ed.) *The*

Knowledge of Man: A Philosophy of the Interhuman, Atlantic Highlands, NJ: Humanities Press International, 1988.(Buber argues for distinguishing between the realm of the 'interhuman' and that of the 'social'. The former refers to the sphere of direct dialogical relation, the latter describes impersonal group relations.)

Buber, M. (1955) 'Der Mensch und sein Gebild', Heidelberg: Lambert Schneider; trans. M. Friedman, 'Man and His Image-Work', in M. Friedman (ed.) *The Knowledge of Man: A Philosophy of the Interhuman*, Atlantic Highlands, NJ: Humanities Press International, 1988.(An important statement of Buber's aesthetics, based on his epistemological theory that 'I-It' knowledge is derived from the 'I-Thou' relation.)

Buber, M. (1956) 'Dem Gemeinschaftlichen folgen', *Die Neue Rundschau* 67 (4): 582-600; trans. M. Friedman, 'What is Common to All', in M. Friedman (ed.) *The Knowledge of Man: A Philosophy of the Interhuman*, Atlantic Highlands, NJ: Humanities Press International, 1988.(Buber draws upon Heraclitus to analyse the emergence of a common world, and thus a 'We', through meaningful human speech.)

Buber, M. (1957) 'Guilt and Guilt Feelings', *Psychiatry* 20 (2); trans. M. Friedman, in M. Friedman (ed.) *The Knowledge of Man: A Philosophy of the Interhuman*, Atlantic Highlands, NJ: Humanities Press International, 1988.(Contra Freud and other thinkers, Buber argues that guilt is not always reducible to neurosis or social convention. The possiblity of real guilt is implicit in a philosophy of personal responsibility.)

Buber, M. (1960a) 'Das Wort, das gesprochen wird', in R. Oldenbourg (ed.) *Wort und Wirklichkeit*, Munich: Bayerische Akademie der Schönen Künste; trans. M. Friedman, 'The Word that is Spoken in Word and Reality', in M. Friedman (ed.) *The Knowledge of Man: A Philosophy of the Interhuman*, Atlantic Highlands, NJ: Humanities Press International, 1988.(Buber analyses the importance of lived speech for human existence, arguing that it is ambiguity that constitutes the tension and fruitfulness of living language.)

Buber, M. (1960b) *Begegnung: Autobiographische Fragmente*, Stuttgart: W. Kohlhammer; trans. and ed. M. Friedman, *Meetings*, La Salle, IL: Open Court, 1972.(A collection of twenty autobiographical anecdotes, some of which appear in Buber's earlier work, which Buber assembled in place of a conventional autobiography for *The Philosophy of Martin Buber*. Each fragment relates an event, meeting or 'mismeeting' which had an important influence on the development of his thought.)

Buber, M. and Rogers, C.R. (1960) 'Dialogue between Martin Buber and Carl R. Rogers', *Psychologia - An International Journal of Psychology in the Orient* 3 (4); repr. in M. Friedman (ed.) *The Knowledge of Man: A Philosophy of the Interhuman*, Atlantic Highlands, NJ: Humanities Press International, 1988.(Transcript of a 1957 discussion between Buber and the American psychologist Carl Rogers. The dialogue brings out differences between their approaches to psychology, and elaborates the significance of Buber's thought for psychotherapy.)

Buber, M. (1965) *Between Man and Man*, ed. M. Friedman, New York: Macmillan. (Collection of essays including 'Die Frage an den Einzelnen', 'Dialogue', 'Education', 'Education and Character' and 'What is Man?'.)

Buber, M. (1967) *The Philosophy of Martin Buber*, trans. M. Friedman *et al.*, ed. M. Friedman and P.A. Schilpp, Library of Living Philosophers, Cambridge: Cambridge University Press.(Includes Buber's 'Autobiographical Fragments' and 'Replies to Critics', as well as essays on Buber's thought by Gabriel Marcel, Emmanuel Levinas, Emil Fackenheim and Ernst Simon, among others.)

References and further reading

Bernasconi, R. (1988) '"Failure of Communication" as a Surplus: Dialogue and Lack of Dialogue Between Buber and Levinas', in R. Bernasconi and D. Wood (eds) *The Provocation of Levinas: Rethinking the Other*, London: Routledge.(Bernasconi analyses Levinas' various criticisms of Buber and Buber's responses to them.)

Friedman, M. (1991) *Encounter on the Narrow Ridge: A Life of Martin Buber*, New York: Paragon House.(An accessible intellectual biography of Buber which includes an annotated bibliography of Buber's most important works in English translation.)

Gibbs, R. (1992) *Correlations in Rosenzweig and Levinas*, Princeton, NJ: Princeton University Press, ch. 7. (Comprises a brief comparative study of the I-Thou relation in Cohen, Buber, Rosenzweig and Levinas, which is useful for students and scholars alike.)

Horwitz, R. (1988) *Buber's Way to 'I and Thou': The Development of Martin Buber's Thought and His 'Religion as Presence' Lectures*, New York: The Jewish Publication Society.(An accessible account of the development of Buber's thought up to the publication of *Ich und Du* which includes Horwitz's translation of Buber's 1922 lectures on 'Religion as Presence' and a selection of letters from the Buber-Rosenzweig correspondence.)

Kepnes, S. (1992) *The Text as Thou: Martin Buber's Dialogical Hermeneutics and Narrative Theology*, Bloomington, IN: Indiana University Press.(A scholarly but accessible analysis of the development of Buber's

hermeneutics and theology.)
Kohanski, A. (1982) *Martin Buber's Philosophy of Interhuman Relation*, London: Associated University Presses. (An introduction to Buber's philosophy which includes a helpful 'Index of Terminology'.)
Levinas, E. (1961) *Totalité et infini*, The Hague: Martinus Nijhoff; trans. A. Lingis, *Totality and Infinity: An Essay on Exteriority*, Pittsburgh, PA: Duquesne University Press.(Develops a quasi-phenomenological account of the ethical relationship between self and Other which is in many ways similar to Buber's I-Thou. Levinas, however, insists on the asymmetry of the 'face-to-face' relation, in which I become aware of my inescapable responsibility for the Other, without thinking about the Other's responsibility for me.)
Theunissen, M. (1984) *The Other: Studies in the Social Ontology of Husserl, Heidegger, Sartre and Buber*, trans. C. Macann, London: MIT Press.(Scholarly study of the phenomenology of the I-Other relation.)
Vermes, P. (1994) *Buber on God and the Perfect Man*, London: Littman Library of Jewish Civilization.(An introduction to Buber's life and thought which includes a précis of *Ich und Du*.)
Wood, R. (1969) *Martin Buber's Ontology*, Evanston, IL: Northwestern University Press.(A study of Buber's thought, grounded in a systematic analysis of the form and content of *Ich und Du*.)

Buddha (6th-5th century BC)

The title of Buddha is usually given to the historical founder of the Buddhist religion, Siddhārtha Gautama, although it has been applied to other historical figures, Buddhist and non-Buddhist, and to many who may be mythological. The religion which he founded was enormously successful and for a long period was probably the most widespread world religion. It is sometimes argued that it is not so much a religion as a kind of philosophy. Indeed, Buddhism bears close comparison with some of the philosophical schools of the Hellenistic world in this respect. The Buddha himself does not seem to have known the concept of a transcendent God and most schools of Buddhism have repudiated it on the grounds, among others, that it undermines personal responsibility for action. Buddhism could be considered as a kind of 'metareligion', open to many religious practices and tolerating others, but not identifiable with religious activity as such - more a kind of philosophical structuring of religion together with a methodology for self-development. Associated with this latter is an elaborate and sophisticated account of mental states and the functioning of consciousness. Characteristic of earlier Buddhist thought is a positive emphasis upon balanced states and a strong rejection of any form of underlying substance and most types of changelessness.

The date when the Buddha, or Siddhārtha Gautama, lived is not certain. For some time most scholars thought that the main period of his activity was in the late sixth century BC, but many believe it more likely that he died close to the end of the fifth century BC, making him a near contemporary of Socrates. Varied information about his life has been preserved, although it is difficult to differentiate between accurate biography and later legend. Scholars are divided on this issue.

It is known that he was born in a noble family among a people called the Sākyas who dwelt near the present-day borders of India and Nepal. After marriage and the birth of a son he experienced some kind of existential crisis. Extant literature presents him as perceptive to the sickness, old age and mortality inherent in life and perceiving the possibility of a solution to such problems. So he went forth ultimately 'for the welfare and benefit of the manyfolk' (as many early Buddhist texts put it) and took up the wandering lifestyle of a homeless religious mendicant. He studied under various spiritual teachers and then, seeking to go beyond what he had learned from them, he adopted, together with a group of companions, the common Indian practice of severe asceticism for around six years. He rejected this method as ineffectual and returned to more moderate practices at which point he achieved the spiritual breakthrough that made him a *buddha*. The remainder of his life was spent teaching the way he had found, or rather rediscovered, to his disciples.

This way is often characterized as the 'Middle Way' between the extremes of pursuing sensual pleasures and suffering torment as a means of self-purification and applies both to spiritual praxis and intellectual understanding. The Middle Way also seeks to avoid the extremes of 'eternalism' and 'annihilationalism'. These are terms for commitment as a result of a psychological bias - either towards the belief that there is an unchanging immortal soul that survives death or that death is the final end and there is no continuity after the destruction of the body. Similarly any suggestion that the soul or life principle (*jīva*) is identical to, or completely distinct from the body is avoided. Preferences with regard to the universe were also considered to be the result of psychological bias, for example, whether or not it is eternal in time, and whether it is finite or infinite in space.

With regard to practice, the Buddha especially advocated the development of the meditation states known as *jhānas* (*dhyāna*). Since these were characterized by a pleasant state of mind, they contrasted with the painful methods of practice widespread in the Indian religion of the day. These states also placed a stress on clarity and conscious awareness, as distinct from the apparent valuation of unconscious trance-like states in pre-Buddhist Indian religion. This emphasis led to the development of methods of study and practice which strongly asserted the importance of insight and understanding (*prajñā*). For most later forms of Buddhism and probably for the Buddha himself, the actual goal of Buddhist meditation represented an awakening which permanently integrated both the meditative states of the *jhānas* and a high development of understanding. It is this which is referred to in the early texts as acquiring the 'vision of truth' or '*dhamma* eye' and is seen as the basis for the awakening or *bodhi*, taught by the Buddha, or Awakened One.

According to early Buddhists, the Buddha presented his teaching on two levels. At an introductory stage he put

forward a simple model of the good life, advocating both the practice of generosity and moral restraint. This was probably linked to a picture of the world as one where living beings were reborn in a series of after-death destinies, including rebirth as a human being or animal (see Karma and rebirth, Indian conceptions of). Such a view appears to have been already widely known, even by the Greeks of that time. Living such a life was seen as creating conditions for a better mental state, capable of understanding more profound truths. For the Buddha these more advanced truths were his higher teaching and were referred to as the Four Noble Truths.

The Buddha probably rejected monism. Early on his followers developed a type of process philosophy which emphasized the universality of change in ordinary experience and rejected the idea that there was any kind of fixed essence behind things. This applied both to the world at large and to the individual. Thus, notions of a world-soul, or ground of being and a permanent, unchanging individual nature were repudiated. The picture they adopted saw life in terms of harmonious interaction of multiple processes. Fundamental to this was the notion of *dhamma* (Sanskrit *dharma*) - the lawful and harmonious nature of things. This *dhamma* was rediscovered by the Buddha and his teaching was understood as the presentation of a universal law which exists in some sense, whether or not it is known about.

Our knowledge of the Buddha's teachings and of early Buddhism in general derives primarily from the collection of texts known as the *tipiṭaka*, or 'three baskets': the canonical writings of early Buddhism. The earliest extant version of these is that preserved in the Pāli language by the southern Buddhists. At first it was preserved within the oral tradition, but was written down in Sri Lanka, formerly Ceylon in the first century BC. Translations of similar texts of diverse origins and dates are preserved in the Chinese Buddhist Canon. Undoubtedly some of this material is as old as the corresponding (and often similar) sources in the Pāli Canon, but it is mixed with material of later origin. Some texts in Sanskrit were discovered in libraries (mostly in Nepal) in the nineteenth century and more recently in the monastic libraries of Tibet, in a ruined library at Gilgit in Kashmir and (often fragmentary) in various desert locations in central Asia. However, most of these appear to be slightly later than the equivalent texts in Pāli. Most recent of all is the acquisition by the British Museum in 1995 of a number of fragments of a version of the Canon in the Gāndhārīlanguage, apparently dating from the first century AD or earlier.

See also: Buddhist philosophy, Indian; Buddhist philosophy, Chinese; Buddhist philosophy, Japanese; Buddhist philosophy, Korean

L.S. COUSINS

References and further reading

Carrithers, M. (1983) *The Buddha*, Oxford and New York: Oxford University Press. (A succinct account.)

Gombrich, R. (1988) *Theravāda Buddhism: a Social History from Ancient Benares to Modern Columbo*, London and New York: Routledge & Kegan Paul.(Chapters 3-5 include a valuable account of the ancient Indian background to early Buddhism.)

Harvey, P. (1990) *An Introduction to Buddhism: Teachings, History and Practices*, Cambridge and New York: Cambridge University Press.(A good introduction to Buddhism.)

Lamotte, É. (1958) *History of Indian Buddhism: from the Origins to the Śaka Era*, Louvain: Peeters Press, 1988. (The standard history. Magisterial, although a little dated since its first publication in French.)

Ñāṇamoli, B. (1972) *The Life of the Buddha as it Appears in the Pali Canon: the Oldest Authentic Record*, Kandy, Sri Lanka: Buddhist Publication Society.(A useful collection of many of the original texts in translation.)

Pye, M. (1979) *The Buddha*, London: Duckworth.(An interesting attempt to distinguish myth, legend and history.)

Thomas, E.J. (1927) *The Life of Buddha*, London: Routledge & Kegan Paul, 1975.(A classic, but still useful account.)

Buddhism, Ābhidharmika schools of

During the first centuries after the Buddha, with the development of a settled life of scholarly study and religious practice, distinct schools began to emerge within the Buddhist community. In their efforts to organize and understand the Buddha's traditional teachings, these schools developed a new genre of text, called 'Abhidharma', to express their doctrinal interpretations. More importantly, the term 'Abhidharma' was also used to refer to the discriminating insight that was not only requisite for the elucidation of doctrine but also indispensable for religious practice: only insight allows one to isolate and remove the causes of suffering.

Abhidharma analysis is innovative in both form and content. While earlier Buddhist discourses were colloquial, using simile and anecdotes, Abhidharma texts were in a highly regimented style, using technical language, intricate definitions and complex classifications. The Abhidharma genre also promoted a method of textual exegesis combining scriptural citation and reasoned arguments.

In content, the hallmark of Abhidharma is its exhaustive classification of all factors that were thought to constitute experience. Different schools proposed different classifications; for example, one school proposed a system of seventy-five distinct factors classified into five groups, including material form, the mind, mental factors, factors dissociated from material form and mind, and unconditioned factors. These differences led to heated doctrinal debates, the most serious of which concerned the manner of existence of the individual factors and the modes of their conditioning interaction. For example, do the factors actually exist as real entities or do they exist merely as provisional designations? Is conditioning interaction always successive or can cause and effect be simultaneous in the same moment? Other major topics of debate included differing models for mental processes, especially perception.

1 Abhidharma

By about the first century BC, about 200 to 300 years after the death of the Buddha, the community of his followers had changed significantly. From a loose grouping of itinerant mendicants, practising largely as individuals, there emerged communities with a settled life of scholarly study and religious discipline. In this more organized setting, Buddhist monks began to re-examine the received teaching, to extend it to new areas, and to elaborate and refine techniques designed to ensure its preservation. Eventually, this new scholastic enterprise became not only the medium for transmitting and understanding the traditional teaching but also tradition in its own right and, for the various schools that were soon to emerge, the source of multiple, even rival traditions.

The Buddhist community developed into separate schools for many reasons: geographic dispersion and isolation; selective patronage; contact with non-Buddhists; disagreements over disciplinary codes; the absence of a single institutional authority; the divergent views of influential teachers; specialization in segments of Buddhist scripture; and, finally, doctrinal disagreements. Among these, disputes about doctrine, or rather the elaboration and analysis of doctrine in the context of disputation with rivals, became a defining activity of the emergent schools. This analysis was called 'Abhidharma'. Although its original meaning is not certain, most interpretations agree that it refers to its purpose of supplementing (*abhi*-) or clarifying the teaching (*dharma*). Abhidharma is not only a genre of texts but also, more importantly, the analytic method and the insight through which one correctly comprehends the Buddhist teaching. For Abhidharma is not an end in itself; the scholastic enterprise, like the original teaching, serves a soteriological goal. Abhidharma enables the practitioner to discriminate those aspects of experience that are defiling and so lead to suffering from those that are virtuous; through this discriminating insight, one can remove the defilements, cultivate virtue, and thereby emulate the Buddha and attain enlightenment (see Suffering, Buddhist views of origination of).

According to the self-understanding of the later Buddhist schools, the Abhidharma simply organizes and explicates the often implicit and embryonic teaching preserved in the collections of discourses on doctrine (sūtra) that are attributed to the Buddha. But while the tradition did not consider Abhidharma as innovative - indeed, most schools attributed the content, if not the very words, of the Abhidharma to the Buddha himself - it is none the less clear that the Abhidharma treatises record doctrinal debate that stimulated new doctrines and new methods of exegesis.

Though evidence suggests that each of the early Buddhist schools preserved and transmitted its own complete set

of texts, which would have included an Abhidharma collection, only two are extant: that of the Theravāda school, which became predominant in Sri Lanka and spread to southeast Asia; and that of the Sarvāstivāda school, which became predominant in north, especially northwest, India and spread to central Asia. Although at opposite ends of the geographical and doctrinal range of schools, these two have Abhidharma collections that are similar in number (seven base texts with commentaries, despite no direct correspondence between the texts of the two collections) and, more significantly, in style and doctrinal subjects. Also, each collection fortunately has a text (the *Kathāvatthu* and its commentary among the Theravādin texts, and the *Mahāvibhāṣā* among the Sarvāstivādin) that, by extensively citing the views of other schools lacking extant Abhidharma works, completes our picture of early Buddhist sectarian views.

2 Philosophical significance: methods of exegesis and argument

With the Abhidharma emphasis on discriminating analysis as the proper mode of religious praxis, Buddhist thought becomes ‘philosophical’. However, the philosophical contribution of Abhidharma cannot be understood apart from its specific sectarian and doctrinal context. Only later, with the appearance of Buddhist logicians (see Dignāga; Dharmakīrti), do familiar, abstract philosophical problems become central, and pragmatic soteriological concerns recede somewhat into the background.

In structure and style of presentation, even more than in subject matter, Abhidharma is radically innovative. The colloquial, earlier discourses that persuaded with simile, metaphor and anecdote are replaced by a precise, regimented style and technical terminology that demonstrates by definition, exegesis, catechesis and taxonomy. Despite its technical character, Abhidharma was not a sterile tangent in the development of Buddhist thought, but was the representative core of the Buddhist monastic worldview.

As is to be expected in a strongly traditional culture, the innovative character of Abhidharma did not spring up in an abrupt or discontinuous way; instead, it grew over centuries, with roots in two tendencies that can be traced to the earlier Buddhist discourses. The first of these tendencies was a penchant for analysis pursued with the goal of encompassing in an intricate classificatory system all the factors (*dharma*) thought to constitute experience. This included topical lists of factors distinguished by numeric and qualitative criteria as well as complex combinations of sets of categories yielding matrices (*mātṛkā*). In this exhaustive listing, all factors of experience would be strictly defined and their relation to all other factors would be clearly indicated by their placement within the various categories; guided by this exhaustive enumeration, the religious practitioner could then distinguish those factors to be cultivated from those to be removed. The second tendency was more expository than analytic and evolved from dialogues about doctrine conducted in the early Buddhist community. Such discussions would often begin with a quotation from earlier discourses or a doctrinal point to be defined and would then proceed using a pedagogical technique of question and answer.

Abhidharma texts also document the concurrent development of organized methods of exegesis and argument. While similar due to shared beginnings, the canons of Buddhist schools became widely divergent; contradictions occur even within the texts of a single school. In order to judge the authenticity and authority of different passages and texts, to interpret them accurately and to mediate conflicting positions, the schools began to elaborate a systematic hermeneutics. In general, the interpretative principles applied were inclusive and harmonizing: any statement deemed in conformity with the teaching of the Buddha or with his enlightenment experience was accepted as genuine; hierarchies were created that incorporated divergent passages by valuing them differently; contradictory passages in the discourses or within Abhidharma texts were said to represent the variant perspectives from which the Buddhist teaching could be presented. Notable for its parallel with later Buddhist ontology and epistemology was the hermeneutic technique whereby certain passages or texts were judged to express absolute truth or reality explicitly, while others, apparently expressing mere conventional truth, nevertheless were judged to have implicit meaning. For the Ābhidharmika schools, accordingly, the Abhidharma texts were considered explicit, whereas the collections of discourses were merely implicit and thus in need of further interpretation.

More formal methods of argumentation begin to appear in Abhidharma texts of the middle period (first centuries AD), which record doctrinal disagreements and debates among rival schools. These texts exhibit a stylized pattern using both supporting scriptural references and reasoned investigations, both of which were considered equally effective in argument. In the earliest examples of such arguments, the reasoned investigations did not yet possess the power of independent proof and were considered valid only in conjunction with supportive scriptural citations.

However, in texts of the later period, supporting scriptural references became decontextualized commonplaces - cited simply to validate the occurrence of terms - and reasoned investigations began to be appraised by independent nonscriptural criteria such as internal consistency and the absence of fallacious causal justification and other logical faults.

3 Philosophical significance: nature of reality and conditioning relations

The exhaustive Abhidharma analysis of experience occasioned doctrinal controversies that could be termed ontological: namely, the manner of existence of the different factors constituting experience; the nature of the reality that they represent; and the dynamics of their interaction, or conditioning. Simple enumerations of factors representing various analytical perspectives are found even in the earlier discourses. Abhidharma texts preserve these earlier analyses, but elaborate more comprehensive and complex intersecting classifications that attempt to clarify the unique identity of each factor as well as all possible modes of interaction among factors. Virtually every early Buddhist school proposed some method of classification, but the lists of specific factors varied greatly. For example, the Sarvāstivādins enumerated seventy-five factors (*dharma*) categorized in five groups: material form (eleven); mind (one); mental factors (forty-six); factors dissociated from material form and mind (fourteen); and unconditioned factors (three). The forty-six mental factors were then subcategorized into groups reflecting their moral quality, a determination essential for correct practice.

The variant lists proposed by other schools often reflect differences of opinion on major issues. Implicit in the Sarvāstivādin categorization of factors is a model of mental processes whereby each moment is characterized by the occurrence of a single instance of mind (*citta*), which serves as a focal point necessarily arising together with ten mental factors that perform specific psychological functions: for example, such factors as feelings, volition, concepts, insight and mindfulness. In addition to these omnipresent and necessary mental factors, mind may also arise with other factors of either a virtuous or unvirtuous moral quality. Other schools, among them the Dārṣṭ āntikas, rejected this model and claimed that more than one mental activity could not occur simultaneously in the same moment, but rather could arise only sequentially in successive moments. Indeed, some sources suggest that certain Dārṣṭāntika masters recognized the existence of only three such mental factors - feelings, concepts and volition - which were thought to arise in successive moments.

Perhaps the most significant controversy debated in the context of these lists concerns modes of existence, specifically, the manner of existence of the individual factors and their conditioning interaction. Within the earlier discourses, the enumerating of factors was motivated not simply by a need to organize the teaching to meet the demands of oral transmission, but rather had a doctrinal function. The fundamental Buddhist teaching of nonself (see Buddhist philosophy, Indian §1) was thought to be validated by demonstrating that no perduring, unchanging, independent ‘self’ (*ātman*) could be found in an entity as a whole, in any of the constituents that the entity comprises, or outside the entity. All entities were thus proven to be impermanent, that is, arising and passing away without any constant essence. In Abhidharma treatises, the increasingly detailed enumerations of factors were intended to demonstrate this fundamental truth of existence as essenceless and impermanent; nevertheless, they led to markedly different ontological models.

A distinctive ontology was constructed by the Sarvāstivādins (‘those who profess that everything exists’), who offered a radical interpretation of impermanence as momentariness, where all factors constituting experience exist separately, arising and passing away within the span of a single moment. In a reality that is an array of radically momentary factors, continuity as commonly experienced and indeed any interaction among the discrete factors would seem to be a logical impossibility. A factor arising in one moment could not act as a condition for an as yet nonexistent future factor, and that subsequent factor could not be said to be conditioned by a past and nonexistent factor of the previous moment. To safeguard ordinary experience, the Sarvāstivādins suggested a novel reinterpretation of existence: each factor, they claimed, is characterized by both an intrinsic nature, which exists unchanged in the past, present and future, and an activity, which, arising and passing away due to the influence of conditions, exists only in the present moment. Only factors that are thus both defined by intrinsic nature and characterized by activity can be considered to exist as real entities; composite objects, consisting of factors and constituting ordinary experience, exist only provisionally. The Sarvāstivādins saw this as the only model that would preserve the Buddhist doctrine of impermanence, since each factor’s activity is conditioned. Conditioning interaction in the past, present and future would still be possible given an existing intrinsic nature. Later

Sarvāstivādins, in an effort to clarify this causal functioning and maintain temporal distinctions, claimed that a factor's present functioning should be distinguished from its functioning in the past or future. Present functioning, or 'activity', refers to homogeneous causation through which a factor conditions the arising of a similar factor in the subsequent moment. Past and future functioning, or 'capability', would include various causal functions, such as serving as an object of cognition in the case of memory, or acting as an immediately prior cause in the arising of a subsequent effect.

The Sarvāstivādin ontological model became the subject of heated debate and was rejected by other Ābhidharmika schools such as the Dārṣṭāntikas, who claimed that factors exist only in the present and not in the past and future. Factors do not exist as isolated units of intrinsic nature that manifest a particular activity through the influence of other isolated conditions. Instead, their activity is to be equated with their very existence. The process of causal activity, the Dārṣṭāntikas argued, is experience itself; the fragmentation of this process into discrete factors possessed of individual existence and unique efficacy is nothing but a mental fabrication.

These ontological investigations generated complex theories of conditioning and intricate typologies of causes and conditions. There is evidence of several rival classifications: the Theravādins proposed a set of twenty-four conditions; the Sarvāstivādins, separate sets either of four conditions or of six causes; and the school represented by the **Śāriputrābhidharmaśāstra*, ten conditions. Each of these individual causes and conditions accounts for a specific mode of conditioning interaction: for example, one cause explains the association among mental factors; another, the maturation of efficacious action; still another, the reciprocal influence of simultaneous factors; and yet another, the arising of a similar subsequent factor. Besides establishing different typologies of causes and conditions, the schools also disagreed on the causal modality exercised by these specific types. The Sarvāstivādins acknowledged that certain of these causes and conditions arise prior to their effects, while others, which exert a supportive conditioning efficacy, arise simultaneously with them. The Dārṣṭāntikas, however, allowed only successive causation; a cause must always precede its effect. In these debates about causality, the nature of animate or personal causation - that is, efficacious action, or *karman* - and the theory of dependent origination intended to account for its activity were, naturally, central issues because of their fundamental role in all Buddhist teaching (see Causation, Indian theories of §§1, 6).

4 Philosophical significance: perception and insight

Abhidharma investigations also extended into the field of epistemology, not for its own sake, but rather for an understanding of how the liberating insight that annuls ignorance and the suffering it causes can arise. To this end, Abhidharma texts carried out a thorough analysis of mental processes, which were all interpreted as varieties of perception. In general, every instance of consciousness, which includes five externally directed varieties of consciousness, and mental consciousness as a sixth, was understood as intentional, occurring only in dependence upon three conditions: a sense organ; a suitable sense object; and the previous moment of consciousness. However, schools disagreed about the specific roles of the sense object, sense organ and consciousness, and about the distinctive character of mental consciousness.

For the Sarvāstivādins, the sense organ grasps the object and consciousness apprehends the general character of the object as grasped by the sense organ: that is, its character as visual form, sound, odour, and so on. The various mental factors simultaneous with that moment of consciousness grasp the particular characteristics of the object. When apprehended in the present moment by consciousness and its associated mental factors, the object becomes internalized as an 'object-support'. In this model of consciousness, erroneous cognition is possible, but such cognition is grounded in an existent object, despite being erroneously grasped by the sense organ or erroneously apprehended by consciousness. Error thus lies in the internal object-support, not in the external object.

In the case of the five externally directed sense organs, the object exists as a real entity and is apprehended only when it is present; indeed, its function as an object verifies its existence as a real entity. The object and sense organ condition the arising of a simultaneous apprehending consciousness. Thus, perception assumes a model of simultaneous conditioning. Mental consciousness also functions according to the model of external perception. The previous moment of consciousness, whatever the type, serves as the mental sense organ for the arising of the subsequent moment of mental consciousness. Any object, whether past, present or future, can be apprehended by mental consciousness. This includes really existent objects - for example, mental factors such as feelings and concepts or any object of the other five varieties of consciousness - or the composite entities of ordinary

experience, which exist only provisionally. Both the organ and object then condition the arising of a present moment of mental consciousness.

The Dārṣṭāntikas consider this Sarvāstivādin model a mere provisional description of perception; no experience, they maintain, including perception, can be analysed into components that actually exist as discrete factors possessing individual activities. The sense organ, sense object, and consciousness are simply provisionally discriminated in the midst of an experienced causal stream. Further, they claim that one cannot establish a relationship of cause and effect within the same moment; as a result, one cannot look to an object in one moment producing consciousness in that very moment. If that were the case, perception would become a successive causal process whereby the sense organ and object in one moment condition the arising of consciousness in a subsequent moment; in effect, perception would always involve apprehension of a past and - for the Dārṣṭāntikas, who reject the existence of past factors - nonexistent object.

Building upon the Abhidharma texts, later Buddhist thought would continue to refine both its dialectical method and its examination of core philosophical issues in the context of debates largely conducted with a variety of extra-Buddhist traditions. But with the passing of the Ābhidharmika schools and the historical conditions that stimulated their particular intersectarian mode of disputation, with its interweaving of hermeneutical and logical argument, there passed also a unique creative phase in the history of Indian thought. Only some of their unique legacy of subtle analyses remains, never fully subsumed into the later tradition and relatively unstudied to this day.

See also: Buddhist philosophy, Japanese; Sense perception, Indian views of

COLLETT COX

References and further reading

Bareau, A. (1947-50) ‘Les Sectes bouddhiques du Petit Véhicule et leurs Abhidharmapiṭaka’ (Lesser-Vehicle Buddhist Sects and their Abhidharmapiṭaka), *Bulletin de l’École Française d’Extrême-Orient* 44, fasc. 1: 1-11. (Review of Abhidharma collections of early Indian Buddhist schools.)

Bareau, A. (1955) *Les Sectes bouddhiques du Petit Véhicule (Lesser-Vehicle Buddhist Sects)*, Saigon: École Française d’Extrême-Orient.(Extensive collection of the doctrinal positions of early Indian Buddhist schools.)

Buddhaghosa (5th century AD) *Visuddhimagga*, trans. Bhikkhu Ñāṇamoli, *The Path of Purification*, Colombo: A. Senage, 1956.(Later compendium of virtually the entirety of Theravādin Abhidhamma analysis.)

Cox, C. (1995) *Disputed Dharmas: Early Buddhist Theories on Existence. An Annotated Translation of the Section on Factors Dissociated from Thought from Saōghabhadra’s Nyāyānusāra*, Studia Philologica Buddhica Monograph Series, Tokyo: The International Institute for Buddhist Studies.(Study of later Sarvāstivādin *dharma* classification and ontology.)

Dhammasaōgaṇi (3rd-1st centuries BC), trans. C.A.F. Rhys Davids, *A Buddhist Manual of Psychological Ethics*, London: Pali Text Society, 1974, reprint of 1900 edn.(Theravādin Abhidhamma work employing elaborate matrices to specify types of mind (*citta*).)

Frauwallner, E. (1995) *Studies in Abhidharma Literature and the Origins of Buddhist Philosophical Systems*, trans. S.F. Kidd, Albany, NY: State University of New York Press.(Studies of northern Indian Abhidharma texts.)

Hirakawa, A. (1990) *A History of Indian Buddhism*, trans. P. Groner, Honolulu, HI: University of Hawaii Press. (History of early Indian Buddhism, focusing on doctrine.)

Kathāvatthu (3rd-1st centuries BC) trans. S.Z. Aung and C.A.F. Rhys Davids, London: Pali Text Society, 1969, reprint of 1915 edn.(Theravādin Abhidhamma work detailing contending positions on key doctrinal points.)

Lamotte, É. (1988) *History of Indian Buddhism*, trans. S. Webb-Boin, Louvain and Paris: Peeters Press.(Thorough history of all aspects of early Indian Buddhism.)

La Vallée Poussin, L. de (ed.) (1925) ‘La Controverse du temps et du *pudgala* dans le *Vijñānakāya*’ (The Controversy over Time and pudgala in the Vijñānakāya), *Études asiatiques* 1: 343-76.(Sections from the Sarvāstivādin Abhidharma text, the *Vijñānakāya* (3rd-1st centuries BC).)

La Vallée Poussin, L. de (ed.) (1930) ‘Documents d’Abhidharma: Textes relatifs au *nirvāṇa* et aux *asaṃskṛtas* en général I - II’ (Abhidharma Documents: Texts Concerning nirvāṇa and asaṃskṛtas in General, I - II), *Bulletin de l’École Française d’Extrême-Orient* 30: 1- 28, 247-98.(Sections from Sarvāstivādin Abhidharma texts of the 2nd-5th centuries AD.)

La Vallée Poussin, L. de (ed.) (1936-1937a) ‘Documents d’Abhidharma: la controverse du temps’ (Abhidharma

Documents: The Controversy over Time), *Mélanges chinois et bouddhiques* 5: 7-158.(Sections from Sarvāstivādin Abhidharma texts of the 2nd-5th centuries AD.)

La Vallée Poussin, L. de (ed.) (1936-1937b) ‘Documents d’Abhidharma: les deux, les quatre, les trois vérités’ (Abhidharma Document: The Two Truths, the Four Truths and the Three Truths), *Mélanges chinois et bouddhiques* 5: 159-87.(Sections from Sarvāstivādin Abhidharma texts of the 2nd-5th centuries AD.)

Stcherbatsky, T. (1923) *The Central Conception of Buddhism and the Meaning of the Word ‘Dharma’*, Delhi: Motilal Banarsidass, 1970.(Brief study of Abhidharma classifications.)

Vasubandhu (5th century AD) *Abhidharmakośa*, trans. L. de la Vallée Poussin, *L’Abhidharmakośa de Vasubandhu*, Paris: Librairie Orientaliste Paul Geuthner, 1923-31, 6 vols; trans. (of La Vallée Poussin edn) L. Pruden, Berkeley, CA: Asian Humanities Press, 1988-90, 4 vols. (Later digest of northern Indian Abhidharma teachings, including also polemical arguments on doctrinal points.)

Williams, P. (1981) ‘On the Abhidharma Ontology’, *Journal of Indian Philosophy* 9: 227-57.(Study of Sarvāstivādin ontology.)

Buddhism, Mādhyamika: India and Tibet

Madhyamaka (‘the Middle Doctrine’) Buddhism was one of two Mahāyāna Buddhist schools, the other being Yogācāra, that developed in India between the first and fourth centuries AD*. The Mādhyamikas derived the name of their school from the Middle Path* (madhyamapratipad*) doctrine expounded by the historical Siddhārtha, prince of the Śākya clan, when he gained the status of a* buddha*, enlightenment. The Madhyamaka, developed by the second-century philosopher Nāgārjuna on the basis of a class of* sūtras *known as the* Prajñāpāramitā *(‘Perfection of Wisdom’), can be seen in his foundational Mūlamadhyamakakārikā(Fundamental Central Way Verses). Therein he expounds the central Buddhist doctrines of the Middle Path in terms of interdependent origination (*pratītyasamutpāda*), conventional language (*prajñapti*), no-self nature (*niḥsvabhāva*) and voidness (*śūnyatā*). He grants that the* dharma *taught by the enlightened ones is dependent upon two realities (*dve satye samupāśritya*) - the conventional reality of the world (*lokasaṃvṛtisatyam*) and reality as the ultimate (*satyam paramārthataḥ*). Although voidness is central to Madhyamaka, we are warned against converting* śūnyatā *into yet another ‘ism’.*

Historically, Madhyamaka in India comprises three periods - the early period (second to fifth century), represented by the activities of Nāgārjuna, Āryadeva and Rāhulabhadra; the middle period (fifth to seventh century) exemplified by Buddhapālita and Bhāvaviveka (founders respectively of the Čast;Prāsaōgika and Čast;Svātantrika schools of Madhyamaka), and Candrakīrti; and the later period (eighth to eleventh century), which includes Śāntarakṣita and Kamalaśīla, who fused the ideas found in the Madhyamaka and Yogācāra systems. Many of the Indian Madhyamaka scholars of the later period contributed to Madhyamaka developments in Tibet.

1 The early period (second to fifth century)

Nāgārjuna (*c.* AD 150-200). Nāgārjuna’s life story can be seen in such writings as Xuanzang’s *Travels in the Western World* and Kumārajīva’s *History of the Bodhisattva Nāgārjuna* (both in Chinese), and the histories of Buddhism by Bu-ston and Tāranātha (in Tibetan). Most of the accounts found in these works, however, are fictions and there are discrepancies between the Chinese and Tibetan traditions. But since Kumārajīva (AD 350-409) lived relatively close to the period attributed to Nāgārjuna and was a great scholar of *Prajñāpāramitā* (‘Perfection of Wisdom’) literature, certain information given in his account can be trusted from certain perspectives.

Both archeological records and textual evidences point to Nāgārjuna as a historical person. For example, in southern India, near the river Kushnara, there remains a cave named Nāgārjunakoṇḍa, and Nāgārjuna’s association with such kings as the King of Satavana is recorded in the *Suhṛllekha (A Letter to a Friend)* and the *Ratnāvali (Garland of Jewels)*.

Many works are attributed to Nāgārjuna, among which the following are considered by scholars to be genuine: *Mūlamadhyamakakārikā(Fundamental Central Way Verses)*, *Yuktiṣaṣṭikā(Sixty Verses on Reasoning)*, *Śūnyatāsaptati (Seventy-Verse Commentary on Emptiness)*, *Vigrahavyāvartani (Refutation of Objections)*, *Vaidalyasūtra (Teaching of Vaidalya)*, *Ratnāvali*, *Suhṛllekha* and *Pratītyasamutpāda-hṛdayakārikā(Verses on the Heart of Interdependent Origination)*.

Nāgārjuna developed his philosophy of *śūnyatā* (voidness) as a critical response to views held by Indian Realists such as Sāōkhya and Vaiśeṣika, the Indian schools of logic such as the Nyaiyāyika, and such Hīnayāna Buddhists as the Sarvāstivādas and other Abhidharma philosophers. Although the Abhidharma ideas that Nāgārjuna criticized did not coincide totally with those found in Sarvāstivāda, the idea of existence developed by the Realists did (see Buddhism, Ābhidharmika schools of). Nāgārjuna also aimed to refute non-Buddhist schools such as Sāōkhya and Vaiśeṣika. It was the idea that the objects of perception possess substantive qualities in the external world - the common thread among these Realist schools - that Nāgārjuna attacked.

The Indian system of logic was both the mother and opponent of Nāgārjuna’s system of logic. Without doubt, tenets such as *chala* (quibble), *jāti* (analogue), *nigrahasthana* (defeat), and so on, found in the *Nyāyasūtra (Teaching on Logical Arguments)*, were forerunners of Nāgārjuna’s system of critique. Also his use of *prasaōga* (*reductio ad absurdum*) is an application of the *tarka* (reasoning) method found in the *Nyāyasūtra* (see

Nyāya-Vaiśeṣika §6).

Āryadeva (*c.* AD 170-270). It is said that Āryadeva was born in southern India or perhaps Sri Lanka and was a disciple of Nāgārjuna. His life story can be seen in such writings as Xuanzang's *Travels in the Western World*, Kumārajīva's *History of the Bodhisattva Āryadeva* and Candrakīrti's commentary, *Catuḥśatakavṛtti*.

Three works are attributed to Āryadeva: *[Bodhisattvayogācāra-]Catuḥśataka (Text of Four Hundred Verses)* (Sanskrit, Chinese translation and Tibetan), *Śataśāstra (A Hundred Verses)* (Chinese translation only) and *Akṣaraśataka* (Tibetan and Chinese translation). Kumārajīva's Chinese translation of the *Śataśāstra* is a popular text with the Sanlun school of Chinese and Japanese (Sanron) Buddhism. With regard to the *[Bodhisattvayogācāra-]Catuḥśataka*, there remain Candrakīrti's commentary in Tibetan translation and Dharmapāla's (530-61) commentary in Chinese translation; however, in each case, only the last nine chapters of the text appear. Only about one-third of the Sanskrit original remains. Although the whole work can be found in Tibetan translation, the difficulties of the first half of the Tibetan text mean that very little work has been done on this part.

Rāhulabhadra (*c.* AD 200-300). Little is known of Rāhulabhadra other than that he was Āryadeva's disciple. Of his compositions, two Sanskrit works remain: *Prajñāpāramitāstotra (Stotra on the Perfection of Wisdom)*, attributed in Tibetan Tripiṭaka to Nāgārjuna, and *Saddharmapuṇḍarīkastava (Stava on the Lotus of the Wonderful Law)*. Other works seem to have been attributed to him, as some of his verses are quoted by Asaōga and Paramārtha in their works.

2 The middle period (fifth to seventh century)

In the first chapter of Avalokitavrata's (seventh century?) commentary on the *Prajñāpradīpa (Lamp of Wisdom)* and in the postscript found in Nāgārjuna's own work, the *Akutobhayā (Nāgārjuna's Auto Commentary)*, it is said there were eight commentaries on the *Mūlamadhyamakakārikā*; however, only those by Buddhapālita, Bhāvaviveka, and Candrakīrti are renowned, as the remainder are either not very useful, or are not extant at present. Along with Avalokitavrata, Śantideva (*c.*650-700) and Prajñākaramati (*c.*950-1030) were active during this period.

Avalokitavrata's *Prajñāpradīpaṭīkā (Lamp of Wisdom Commentary)*, extant in Tibetan translation, is a huge text and a very erudite commentary on the various Indian and Buddhist schools; consequently, it serves as a very important source book. Avalokitavrata knew about Candrakīrti and Dharmakīrti (*c.*600-60), but because he did not utilize their logic, we can surmise that he was their contemporary.

In the histories of Buddhism by Bu-ston and Tāranātha, Śantideva is credited as the author of the *Śikṣāsamuccaya (A Compendium of Buddhist Doctrine)*, the *Sūtrasamuccaya (A Compendium of Buddhist Sutras)* and the *Bodhicaryāvatāra (Entering the Path of Enlightenment)*. The *Sūtrasamuccaya*, extant in the Tibetan canon, purports to be by Nāgārjuna, but because some of the sixty or so *sūtras* quoted in it were composed after Nāgārjuna's death, this attribution is difficult as the text stands.

Although commentaries on Śantideva's *Bodhicaryāvatāra* abound, it is Prajñākaramati's *Bodhicaryāvatārapañjika (Entering the Path of Enlightenment Commentary)* which is the most famous. It comments on the first nine chapters, but not the tenth. Prajñākaramati probably began it as a commentary on the ninth chapter, and then added commentary on the first eight chapters later. Judging from his criticism of Sautrāntika and Yogācāra, we can assume that Prajñākaramati was a Mādhyamika, but it is difficult to determine to which school he belonged. Although in terms of his dates Prajñākaramati falls into the later period, he is discussed here as a commentator on Śantideva's work.

According to later Tibetan doxographies, the Middle Period is also characterized by Madhyamaka splitting into the Čast;Prāsaōgika and Čast;Svātantrika schools.

Buddhapālita (*c.*470-540). Buddhapālita's *Mūlamadhyamakavṛtti (Fundamental Central Way Verses Commentary)*, a commentary in twenty-seven chapters, exists only in Tibetan translation. Although from the seventh *kārikā* of chapter twenty-three the content of the text is identical with Nāgārjuna's *Akutobhayā*, its importance can be inferred from Bhāvaviveka's vehement criticism of it and Candrakīrti's concerted defence. The text's distinctive character lies in Buddhapālita's presentation of Nāgārjuna's use of dilemma and tetralemma from

the standpoint of plural *prasaōgas* - that is, he divides the tetralemma into two or four different *prasaōga* (*reductio ad absurdum*) arguments. It is probably for this reason that Buddhapālita is seen as the founder of the Čast;Prāsaōgika school and was criticized by Bhāvaviveka.

Bhāvaviveka (*c.*500-70). Bhāvaviveka, or Bhavya, is known by many different names, but because Candrakīrti, in his *Prasannapāda (Lucid Exposition)*, calls him Bhāvaviveka, he is known generally by that name. He is said to have written four works: *Prajñāpradīpa[-mūlamadhyamakavṛtti]*, *Madhyamakahṛdayakārikā*, *Tarkajvālā (Blaze of Argument)* (*Madhyamakahṛdayavṛtti*) and *Dashengzhangzhenlun* (extant in Chinese only). In the Tibetan canon there is another work, the *Nikāyabhedavibhaōga-vyākhyāna*, attributed to Bhāvaviveka, but its content corresponds to the discussion of the Hīnayāna found in fourth chapter of the *Tarkajvālā*. The authenticity of two other works ascribed to him, the *Madhyamakaratnapradīpa (Jewel of the Central Way Lamp)* and the *Madhyamakārtha-saṃgraha*, is doubtful.

The *Prajñāpradīpa[-mūlamadhyamakavṛtti]* is the first full-scale commentary on the *Mūlamadhyamakakārikā* and is important in two respects. First, Bhāvaviveka's interpretation of the various arguments presented by Nāgārjuna was strongly influenced by Dignāga's logic. Bhāvaviveka established many categorical syllogisms, the so-called *svatantra-anumāna* (independent inference), and hence his system is known as the Čast;Svātantrika school. Second, he criticized Buddhapālita's version of *prasaōga* as being merely *prasaōga* without the minor and major premises.

The *Madhyamakahṛdayakārikā (Verses on the Heart of the Central Way)* consists of eleven chapters, but it is only the first three that discuss Madhyamaka philosophy. In the rest, Bhāvaviveka discusses the doctrines of Hīnayāna Buddhism, Yogācāra, Sāōkhya, Vaiśeṣika, Vedānta and other schools. In this manner, he establishes the Madhyamaka on the one hand, and on the other he criticizes other schools and thus develops his work around the system of doctrinal classification that was further developed by scholars such as Śāntarakṣita. The Sanskrit manuscript of this text was discovered circa 1938.

The *Tarkajvālā*, extant only in Tibetan translation, is Bhāvaviveka's auto-commentary on the *Prajñāpradīpa*. Because within this text, reference to Bhāvaviveka is made by the title *ācarya*, there is some doubt as to whether Bhāvaviveka is the author. The *Dashengzhangzhenlun* has been translated into French by Louis de la Vallée Poussin, but because the text is difficult and its contents differ from the *Prajñāpradīpa* and the *Madhyamakahṛdayakārikā*, research on it has been slow and laborious.

Candrakīrti (*c.*600-50). Candrakīrti's *Prasannapadā* is not only the sole extant Sanskrit commentary on Nāgārjuna's *Mūlamadhyamakakārikā* but is also the one from which the Sanskrit *kārikās* are derived. His *Madhyamakāvatāra* discusses the essentials of the Madhyamaka in accordance with the stages of the ten perfections found in the *Daśabhūmikasūtra (Teaching on the Ten Stages)*. Other works attributed to Candrakīrti, in Tibetan translations, are the *Pañcaskandhaprakaraṇa (Five Aggregates Commentary)*, and commentaries on the *Śūnyatāsaptati*, the *Yuktiṣaṣṭika* and the *Catuḥśataka*.

Čast;Svātantrika and Čast;Prāsaōgika. The fact that Bhāvaviveka criticized and Candrakīrti defended Buddhapālita indicates a two-way split in the Madhyamaka school, although the terms 'Prāsaōgika' and 'Svātantrika' are not to be found in any of the extant Sanskrit texts. They are probably derived from translating into Sanskrit the Tibetan terms *thal 'gyur ba* and *rang rgyud pa* respectively. Although these terms are found in Tibetan Buddhist literature prior to Tsong kha pa, it is he who can be credited with giving them prominence, after which they began to be utilized in the Tibetan *Grub-mtha'* texts. To understand the manner in which the Čast;Prāsaōgikas and the Čast;Svātantrikas develop their respective theses requires a comprehensive background in the development of logic in India. The controversy between the two was not settled by Candrakīrti; its development is closely aligned with the logical proofs for *reductio ad absurdum* arguments produced after the eighth century and with the idea of *antarvyāpti* (intrinsic determination of universal concomitance) introduced by eleventh-century Buddhist logicians. Thus, in discussing the controversy between the two, one must take into consideration the developments in logic that occured after Candrakīrti.

3 The later period (eighth to eleventh century)

The special characteristics of the later period, which began with Śāntarakṣita, were: (1) the influence of Dharmakīrti's logico-epistemological school (although there was still dependence on Nāgārjuna's thought); (2) the

fact that most philosophers belonged to the Svātantrika school (an exception is Prajñākaramati); and (3) the incorporation of tenets of the Yogācāra tradition. Madhyamaka of this period can be seen as a rendezvous between Madhyamaka and Yogācāra (see Dharmakīrti; Buddhism, Yogācāra school of). As well as the thinkers described below, there were others, such as Jitari, Bodhibhadra and Advayavajra (eleventh century), who wrote expositions of the doctrines of the various Buddhist schools, while others still - for example, Kambala (*c.*700), Ratnākaraśanti (*c.*eleventh century) and Atīśa (982-1055) - were actively engaged in the synthesis of Madhyamaka and Yogācāra in India and in the development of Madhyamaka Buddhism in Tibet.

Jñānagarbha (eighth century). Very little is known about Jñānagarbha, except that he was Śāntarakṣita's teacher and the author of the *Satyadvayavibhaōga (Analysis of the Two Realities)*, the *Satyadvayavibhaōgavṛtti (Analysis of the Two Realities Commentary)* and the *Yogabhāvanāmārga (Path to the Practice of Yoga)*, all extant in Tibetan only. He translated texts on Madhyamaka and epistemology into Tibetan and consequently may have entered Tibet before Śāntarakṣita.

Śāntarakṣita (*c.*725-84). Śāntarakṣita was a scholar-monk in Nālandāwho composed the *Tattvasaṃgraha*, the *Madhyamakālaṃkāra*, and his own commentary on it, the *Madhyamakālaṃkāravṛtti*, and a commentary on Jñānagarbha's *Satyadvayavibhaōga*. With the exception of the first, which can also be found in Sanskrit, all are extant in Tibetan only. In criticizing the Indian philosophical traditions, both Buddhist and non-Buddhist, Ś āntarakṣita gives a wealth of information regarding the philosophical status of the time. The reason for his critical analysis was of course to establish Madhyamaka as the highest doctrine of all. Although he considered *śūnyatā* as the highest truth, he claimed that from the worldly level one should accept the Yogācāra claim of *vijñaptimātra* (information-only), and thus he can be seen as an adherent of the Yogācāra-Madhyamaka. In so far as he respected the system of inference, he can be seen as an adherent of the Čast;Svātantrika position. In 763, he entered Tibet and then returned to India. He returned again to Tibet in 771 and worked together with Padmasaṃbhava to establish the bSam-yas temple. He also initiated the first six monks of Tibet and devoted the rest of his life to the development of Buddhism there.

Kamalaśīla (*c.*740-97). A student of Śāntarakṣita, Kamalaśīla wrote brilliant commentaries on his teacher's works, the *Tattvasaṃgrhapañjika (Compendium of Principles Commentary)* and the *Madhyamakālaṃkārapañjika (Central Way Ornament Commentary)*. Besides these, he wrote the *Madhyamakāloka (Light of Central Way)*, the *Tattvāloka (Light of Reality)* and the *Sarvadharmaniḥsvabhāvasiddhi (Establishment of the No-Self-Nature of all Dharmas)*. Invited by King Khri srong lde brtsan, Kamalaśīla entered Tibet in 794. It is said that he defeated the Chinese monk, Hva-shang Ma-he-yan, in a debate held in bSam-yas (Samyay) monastery. The Madhyamaka tradition is supposed to have become firmly established in Tibet as a result. For the Tibetan Buddhists he composed a trilogy known as the First, Second and Third *Bhāvanākrama (Steps of Religious Meditative Practices)*, in which he denounced the 'sudden enlightenment' view held by Chinese Chan Buddhists (see Buddhist philosophy, Chinese §9).

Vimuktisena (eighth century). Vimuktisena, teacher of Haribhadra, was a well-known figure in the Tibetan tradition, but his biography is not clear. His *Abhisamayālaṃkāravṛtti (Ornament of Realizations Commentary)* is a commentary on the *Abhisamayālaṃkāra (Ornament of Realizations)* attributed to the teacher Maitreyanātha, who expounded the foundational doctrine of Yogācāra. As not much is known about this text, it is difficult to determine the extent of Vimuktisena's dedication to Madhyamaka thought. According to the Tibetan tradition, he belongs to the same lineage as Śāntarakṣita and Kamalaśīla.

Haribhadra (*c.*800). Haribhadra composed a huge treatise called the *Aṣṭasāhasrikāprajñāpāramitāvyākhyā Abhisamayālaṃkārāloka*, a commentary on the *Aṣṭasāhasrikāprajñāpāramitāsūtra (Teaching of the Perfection of Wisdom in 8,000 Verses)* in accordance with the divisions of the *Abhisamayālaṃkāra* attributed to Maitreyanātha. In this treatise, he criticized the Sarvāstivāda, the Sautrāntika, the Alīkākāravāda Yogācāra (who hold the theory of the unreal image of cognition) and the Satyākāravāda Yogācāra (who hold the theory of the real image of cognition) by following the method of critique in Śāntarakṣita's *Madhyamālaṃkāra*. He also explained the basic Madhyamaka teaching of no self-nature (*niḥsvabhāva*) for all entities and having classified the causal principle into four types (many causes, one effect; many causes, many effects; one cause, many effects; one cause, one effect) he disproved all of them.

4 Madhyamaka in Tibet

Buddhism was first introduced into Tibet during the reign of King Khri srong lde brtsan (741-97). At that time, Ś āntarakṣita was invited to inaugurate the bSam-yas monastery and initiated the first six monks (*Sad Mi Drug*) of Tibetan origin. Later, Kamalaśīla was invited. Buddhist texts were translated and Buddhist studies advanced, but because both Śāntarakṣita and Kamalaśīla were adherents of Madhyamaka thought (*dBu-ma-pa*), Madhyamaka had a great influence on Tibetan Buddhism.

However, influence was not exerted from India alone, for Chinese Chan was also making its presence felt in Tibet. This interaction led to the famous debate between the Chinese Chan master Hva-shang (also Ho-shang) Ma-he-yan and Kamalaśīla in the latter part of the eighth century. The outcome established the superiority of Kamalaśīla and the Indian tradition, and resulted in Khri srong lde brtsan requesting Kamalaśīla to explain what the nonsubstantive view of phenomena meant in the context of the three learnings of listening, thinking and practice; as a response, Kamalaśīla proceeded to compose the First *Bhāvanākrama (Stages of Practice)* (*sGom-rim dang-po*). When asked how to put into practice such a teaching, he composed the Second *Bhāvanākrama* (*sGom-rim bar-pa*), and finally, when asked what the result would be, he composed the Third and last *Bhāvanākrama* (*sGom-rim tha-ma*). In this text, he criticized the mistake made by Hva-shang Mahāyāna. He presented this to the king, who was overjoyed with it. Consequently, Kamalaśīla composed a commentary (*Don-'grel*), the *Madhyamakāloka* (*dBu-ma snang-ba*), in which he synthesized the correct teaching and the correct principle.

Although Śāntarakṣita and Kamalaśīla are seen as Yogācāra-Mādhyamikans, their activities are more representative of an Indian Madhyamaka tradition in Tibet rather than an indigenous Tibetan Madhyamaka. Unlike the case of the Indian subcontinent, Madhyamaka ideas in Tibet are found within the various doctrines of Tibetan Buddhist orders rather than constituting an independent Madhyamaka Buddhist school. The Tibetan *Grub-mtha'* texts discuss Madhyamaka as a single system, but what these texts explain are the tenets of Madhyamaka with respect to a general classification of doctrine rather than with respect to the Madhyamaka school in particular. Consequently, a unique feature of Tibetan Buddhism is that the indigenous schools incorporate tenets of several Indian Buddhist traditions, such as Sarvāstivāda, Sautrāntika, Yogācāra and Madhyamaka, and therefore it is difficult to discuss Madhyamaka as an indigenous school in Tibet. This means that to investigate Madhyamaka in Tibet is to examine how the Tibetan Buddhist schools, such as the dGe-lugs-pa (Gelukba), the Sa-skya-pa (Sagyaba), the rNying-ma-pa (Nyingmaba), and the bKa'-brgyud-pa (Gagyuba), incorporate the foundational Indian Madhyamaka tenets of *śūnyatā*, *saṃvṛtisatya*, *paramārthasatya*, and so on, within their own systems.

Among the various schools, it is with respect to the dGe-lugs-pa school that we find the most comprehensive discussion of the Madhyamaka system. The dGe-lugs-pa was founded by Tsong kha pa Blo bzang grags pa (Dzongkaba Losang dragba, 1357-1419), who at the age of nineteen heard a lecture on the *Prajñāpāramitā* from Kun-dga'-dpal, who belonged to the Sa-skya-pa school, and also heard lectures on the *Madhyamakāvatāra* and *Abhidharmakośa (Treasury of Higher Dharma)* from Red mda' ba (Rendawa, 1349-1412). He wrote the *gSer-phreng*, a commentary on the *Abhisamayālaṃkāra*, when he was thirty-one years old, but we cannot see Tsong kha pa's own thinking appearing yet. At thirty-three he began to study under dBu-ma-pa (Umaba), to whom he addressed questions regarding the Čast;Svatāntrika (Rang-rgyud-pa, ranggyüba) and the Čast;Prāsaōgika (Thal-'gyur-ba, talgyurwa). When he was forty-six he wrote the *Lam-rim-chen-mo (The Great Stages of the Path)*, a text which was modelled on Atīśa's *Bodhipathapradīpa* and finally established Tsong kha pa's original thinking regarding Madhyamaka. He severely criticized the normative interpretation in Tibet that *śūnyatā* referred to the view of neither existence nor nonexistence held by the mTha'-bral dbu-ma'i lugs ('school of Madhyamaka free of extremes'). Instead, he stressed that *śūnyatā* was absolutely beyond any claim that, for example, the absolute is freed from the two extreme views and also from conventional language, which resembles the view held by Hva-shang Mahāyāna that Buddhahood is attained when one is freed from all discriminations.

Tsong kha pa's understanding of Madhyamaka is expressed again and again in his works, for example, the *Legs-bshad snying-po (Essence of Good Explanations)*, the *Rigs-pa'i rgya-mtsho (Ocean of Reasoning)* and the *gDongs-pa rab-gsal (Exposition of Intention)*. He was severely criticized by Śā-kya mchog-ldan (1428-1507) in his *dBu ma rnam nges (Determination of the Middle)* and by such mTha'-bral dbu-ma'i lugs scholars as Śā-skya sTag-tshang-pa (1405-?) in his *Grub-mtha' kun-shes (All-knowing Tenets)* and Go ram pa (Goramba, 1429-89) in his *lTa ba'i shan 'byed*. In response to these criticisms, Se-ra rJe-tsun-pa Chos-kyi rgyal-mtshan (1469-1546) and his disciple bDe legs nyi ma attacked Śā-skya sTag-tshang-pa and Go ram pa in their *lTa-ngan-mun-sel (Light Within Views)* and the great scholar Jam dbyangs bzhad pa (Jamyang shayba, 1648-1722) attacked Śā-kya

mchog-ldan in his *Grub-mtha' chen-mo (Great Tenets)*.

See also: Buddhist concept of emptiness; Buddhist philosophy, Indian; Buddhist philosophy, Japanese; Seng Zhao; Tibetan philosophy

LESLIE S. KAWAMURA

References and further reading

Āryadeva (*c.* AD 170-270) *Catuḥśataka (Text of Four Hundred Verses)*, trans. K. Lang, *Āryadeva's Catuḥśataka, On the Bodhisattva's Cultivation of Merit and Knowledge*, Indiske Studier 7, Copenhagen: Akademisk Forlag, 1986.(A useful English translation with which to gain insight into Āryadeva's thoughts.)

Harada, S. (1982) 'Chibetto Bukkyō no Chūkan Shisō' (Madhyamaka Thought of Tibetan Buddhism), in A. Hirakawa, Y. Kajiyama and J. Takazaki (eds) *Kōza Daijō Bukkyō (Lectures on Mahāyāna Buddhism)*, Tokyo: Shunjusha, vol. 7: 284-314.(An important source for the present entry.)

Kajiyama, Y. (1982) 'Chūkan Shisō no Rekishi to Bunken' (The History and Literature of Madhyamaka Thought), in A. Hirakawa, Y. Kajiyama and J. Takazaki (eds) *Kōza Daijō Bukkyō*, Tokyo: Shunjusha, vol. 7: 1-83.(An important source for the present entry.)

Kajiyama, Y. (1989) *Studies in Buddhist Philosophy*, ed. K. Mimaki, Kyoto: Rinsen Book Co. (An important source for the present entry; see especially the article on Madhyamaka, pages 21-7.)

Kalupahana, D.J. (trans.) (1986) *Nāgārjuna: The Philosophy of the Middle Way*, Albany, NY: State University of New York Press.(An English translation of the *Mūlamadhyamakakārikā*.)

Lindtner, C. (trans.) (1982) *Nagarjuniana: Studies in the Writing and Philosophy of Nāgārjuna*, Copenhagen: Akademisk Forlag.(A translation, with explanatory text, of Nāgārjuna's major works. A must for Nāgārjunian study.)

Matsumoto, S. (1989) 'Tsonkapa to gelukupa' (Tsong kha pa and dGe-lugs-pa) in G. Nagao *et al*. (eds) *Chibeto Bukkyō (Tibetan Buddhism)*, Tokyo: Iwanami Shoten, 224- 62.(An important source for the present entry.)

Mimaki, K. (ed.) (1977) 'Le *Grub mtha' rnam bźag rin chen phreō ba* de dKon mchog 'jigs med dbaō po (1718-91): Texte tibétain édité, avec une introduction', in *Zinbun* 14: 55-112, Memoirs of the Research Institute for Humanistic Studies, Kyoto University.(Useful reference for Tibetan Buddhist Schools.)

Mimaki, K. (ed. and trans.) (1982) *Blo gsal grub mtha'. Chapitres IX (Vaibhāṣika) et XI (Yogācāra) édités et Chapitre XII (Mādhyamika) édité et traduit*, Kyoto: Zinbun Kagaku Kenkyūsho.(An informative exposition on Madhyamaka in India and Tibet.)

Nāgārjuna (*c.* AD 150-200) *Mūlamadhyamakakārikā (Fundamental Central Way Verses)*, trans. K.K. Inaba, *Nāgārjuna, A Translation of his Mūlamadhyamakakārikā with an Introductory Essay*, Tokyo: The Hokuseido Press, 1970.(English translation of the verses with Sanskrit text.)

Nāgārjuna (*c.* AD 150-200) *Ratnāvalī (Garland of Jewels)*, ed. M. Hahn, Bonn: Indica et Tibetica Verlag, 1982. (Sanskrit, Tibetan and Chinese text.)

Ramanan, K.V. (trans.) (1971) *Nāgārjuna's Philosophy as Presented in the Mahā-Prajñāpāramitā-Śāstra*, Varanasi: Bharatiya Vidya Prakashan.(An English translation with a commentary by the author on a Chinese Madhyamaka text.)

Ruegg, D.S. (1981) *The Literature of the Madhyamaka School of Philosophy in India*, Wiesbaden: Otto Harrasowitz.(A must for further research in the field.)

Ruegg, D.S. and Schmithausen, L. (eds) (1990) *Earliest Buddhism and Madhyamaka*, Leiden: Brill.(Five outstanding articles on Madhyamaka by leading writers in the field.)

Sprung, M. (trans.) (1979) *Lucid Exposition of the Middle Way*, Boulder, CO: Prajñā Press. (A useful translation of the *Mūlamadhyamakakārikā*.)

Streng, F.J. (trans.) (1967) *Emptiness: A Study in Religious Meaning*, New York: Abington Press.(A readable English translation of the *Mūlamadhyamakakārikā* interpreted from a philosophical view.)

Buddhism, Yogācāra school of

Yogācāra is one of the two schools of Indian Mahāyāna Buddhism. Its founding is ascribed to two brothers, Asaōga and Vasubandhu, but its basic tenets and doctrines were already in circulation for at least a century before the brothers lived. In order to overcome the ignorance that prevented one from attaining liberation from the karmic rounds of birth and death, Yogācāra focused on the processes involved in cognition. Their sustained attention to issues such as cognition, consciousness, perception and epistemology, coupled with claims such as 'external objects do not exist' has led some to misinterpret Yogācāra as a form of metaphysical idealism. They did not focus on consciousness to assert it as ultimately real (Yogācāra claims consciousness is only conventionally real), but rather because it is the cause of the karmic problem they are seeking to eliminate.

Yogācāra introduced several important new doctrines to Buddhism, including vijñaptimātra, *three self-natures, three turnings of the* dharma-*wheel and a system of eight consciousnesses. Their close scrutiny of cognition spawned two important developments: an elaborate psychological therapeutic system mapping out the problems in cognition with antidotes to correct them and an earnest epistemological endeavour that led to some of the most sophisticated work on perception and logic ever engaged in by Buddhists or Indians.*

Although the founding of Yogācāra is traditionally ascribed to two half-brothers, Asaōga and Vasubandhu (fourth-fifth century BC*), most of its fundamental doctrines had already appeared in a number of scriptures a century or more earlier, most notably the Saṃdhinirmocanasūtra (Elucidating the Hidden Connections) (third-fourth century* BC*). Among the key Yogācāra concepts introduced in the Saṃdhinirmocanasūtra are the notions of 'only-cognition' (*vijñaptimātra*), three self-natures (*trisvabhāva*), warehouse consciousness (*ālayavijñāna*), overturning the basis (*āśrayaparāvṛtti*) and the theory of eight consciousnesses.*

The Saṃdhinirmocanasūtra proclaimed its teachings to be the third turning of the wheel of dharma*. Buddha lived around sixth-fifth century* BC*, but Mahāyāna Sūtra did not begin to appear probably until five hundred years later. New Mahāyāna Sūtra continued to be composed for many centuries. Indian Mahāyānists treated these Sūtras as documents which recorded actual discourses of the Buddha. By the third or fourth century a wide and sometimes incommensurate range of Buddhist doctrines had emerged, but whichever doctrines appeared in Sūtras could be ascribed to the authority of Buddha himself. According to the earliest Pāli Sutta, when Buddha became enlightened he turned the wheel of* dharma*, that is, began to teach the path to enlightenment. While Buddhists had always maintained that Buddha had geared specific teachings to the specific capacities of specific audiences, the Saṃdhinirmocanasūtra established the idea that Buddha had taught significantly* different *doctrines to different audiences according to their levels of understanding; and that these different doctrines led from provisional antidotes (*pratipakṣa*) for certain wrong views up to a comprehensive teaching that finally made explicit what was only implicit in the earlier teachings. In its view, the first two turnings of the wheel - the teachings of the Four Noble Truths in Nikāya and Abhidharma Buddhism and the teachings of the Madhyamaka school, respectively - had expressed the* dharma *through incomplete formulations that required further elucidation (*neyārtha*) to be properly understood and thus effective. The first turning, by emphasizing entities (such as* dharmas *and aggregates) while 'hiding' emptiness, might lead one to hold a substantialistic view; the second turning, by emphasizing negation while 'hiding' the positive qualities of the* dharma*, might be misconstrued as nihilism. The third turning was a middle way between these extremes that finally made everything explicit and definitive (*nītartha*). In order to leave nothing hidden, the Yogācārins embarked on a massive, systematic synthesis of all the Buddhist teachings that had preceded them, scrutinizing and evaluating them down to the most trivial details in an attempt to formulate the definitive Buddhist teaching. Stated another way, to be effective all of Buddhism required a Yogācārin reinterpretation. Innovations in* abhidharma *analysis, logic, cosmology, meditation methods, psychology, philosophy and ethics are among their most important contributions. Asaōga's magnum opus, the Yogācārabhūmiśāstra (Treatise on the Stages of Yoga Practice), is a comprehensive encyclopedia of Buddhist terms and models, mapped out according to his Yogācārin view of how one progresses along the stages of the path to enlightenment.*

1 Historical survey

Asaōga and Vasubandhu became the first identifiable Yogācārins, each having initially been devoted to other

schools of Buddhism. Both were prolific authors, although Asaōga attributed a portion of his writings to Maitreya, the future Buddha living in Tuṣita heaven. Some modern scholars, such as H. Ui (1929) have argued that this Maitreya was an actual human teacher, not the future Buddha, but the tradition is fairly clear. After Asaōga has spent twelve years of fruitless meditation alone in a cave, in a moment of utter despair Maitreya appears to him and transports him to Tuṣita heaven where he instructs him in previously unknown Yogācārin works. These works Asaōga then introduces to his fellow Buddhists. Precisely which texts these are is less clear since the Chinese and Tibetan traditions assign different works to Maitreya.

According to tradition, Vasubandhu first studied Vaibhāṣika Buddhist teachings, writing an encyclopedic summary of those teachings, the *Abhidharmakośa (Treasury of Abhidharma)*, that has become a standard work throughout the Buddhist world. As he grew critical of Vaibhāṣika teachings, he wrote a commentary on that work refuting many of its tenets. Intellectually restless for a while, Vasubandhu composed a variety of works that chart his journey to Yogācāra, the best known of these being the *Karmasiddhiprakaraṇa (Investigation Establishing Karma)* and *Pañcaskandhakaprakaraṇa (Investigation into the Five Aggregates)*. These works show a deep familiarity with the Abhidharmic categories discussed in the *Abhidharmakośa* and attempts to rethink them; the philosophical and scholastic disputes of the day are also explored and the new positions Vasubandhu formulates in these texts bring him closer to Yogācārin conclusions. On the basis of some conflicting accounts in old biographies of Vasubandhu, a few modern scholars have argued that these texts along with the *Abhidharmakośa* were the work of another author. However, as the progression and development of his thought is so strikingly evident in these works and the similarity of vocabulary and style of argument so apparent across the texts, the theory of the existence of two Vasubandhus has little merit.

The writings of Asaōga (and/or Maitreya) and Vasubandhu ranged from vast encyclopedic compendia of Buddhist doctrine (for example, *Yogācārabhūmiśāstra*, *Mahāyānasaṃgraha*, *Abhidharmasamuccaya*), to terse versified encapsulations of Yogācāra praxis (for example, *Triṃśikā*, *Trisvabhāvanirdeśa*), to focused systematic treatises on Yogācāra themes (for example, *Viṃśatika*, *Madhyāntavibhāga*), to commentaries on well-known Mahāyānic scriptures and treatises, such as the *Saddharmapuṇḍarīka (Lotus)* and the *Vajracchedika (Diamond-cutter) Sūtras*.

As the *Saṃdhinirmocanasūtra* offers highly sophisticated doctrines, it is reasonable to assume that these ideas had been under development for some time, possibly centuries, before this scripture emerged. Since Asaōga and Vasubandhu lived a century or more after the *Saṃdhinirmocana* appeared, it is also reasonable to assume that these ideas had been further refined by others in the interim. Thus the traditional claim that the two brothers are the founders of Yogācāra at best could be considered to be a half-truth. According to tradition Asaōga converted Vasubandhu to Yogācāra after having been taught by Maitreya; he is not known to have had any other notable disciples. Tradition does assign two major disciples to Vasubandhu: Dignāga, the great logician and epistemologist, and Sthiramati, an important early Yogācāra commentator. It is unclear whether either ever actually met Vasubandhu (current scholarship deems it unlikely). They may have been disciples of his thought, acquired exclusively from his writings or through some forgotten intermediary teachers. These two disciples exemplify the two major directions into which Vasubandhu's teachings split.

After Vasubandhu Yogācāra developed into two distinct directions or branches: a logico-epistemic tradition exemplified by such thinkers as Dignāga, Dharmakīrti, Śāntarakṣita and Ratnakīrti; an Abhidharmic psychology, exemplified by such thinkers as Sthiramati, Dharmapāla, Xuanzang (Hsüan-tsang (*c*.659 bc)) and (again) Ratnakīrti (see Buddhism, Ābhidharmika schools of). While the first branch focused on questions of epistemology and logic, the other wing refined and elaborated the Abhidharma analysis developed by Asaōga and Vasubandhu. These branches were not entirely separate, and many Buddhists wrote works that contributed to both. Dignāga, for instance, besides works on epistemology and logic also wrote a commentary on Vasubandhu's *Abhidharmakośa*. What united both branches was a deep concern with the process of cognition, that is, analyses of how we perceive and think. The former wing approached that epistemologically, while the latter wing approached it psychologically and therapeutically. Both identified the root of all human problems as cognitive errors that needed correction.

Several Yogācāra notions basic to the Abhidharma wing came under severe attack by other Buddhists, especially the notion of *ālayavijñāna*, which was denounced as something akin to the Hindu notions of *ātman* (permanent, invariant self) and *prakṛti* (primordial substrative nature from which all mental, emotional and physical things evolve) (see Sāōkhya). Eventually the critiques became so entrenched that the Abhidharma wing atrophied. By the

end of the eighth century it was eclipsed by the logico-epistemic tradition and by a hybrid school that combined basic Yogācāra doctrines with Tathāgatagarbha thought. The logico-epistemological wing sidestepped much of the critique by using the term *cittasantāna*, or mindstream, instead of *ālayavijñāna*, for what amounted to roughly the same idea. It was easier to deny that a stream represented a reified self. The Tathāgatagarbha hybrid school was no stranger to the charge of smuggling notions of selfhood into its doctrines, as for example, it explicitly defined *tathāgatagarbha* as 'permanent, pleasant, *self* and pure'. Many Tathāgatagarbha texts argue for the acceptance of selfhood (*ātman*) as a sign of higher accomplishment. The hybrid school attempted to conflate *tathāgatagarbha* with the *ālayavijñāna*. Key works of the hybrid school include the *Laōkāvatārasūtra*, *Ratnagotravibhāga (Uttaratantra)* and in China the *Awakening of Faith*.

In China during the sixth and seventh centuries, Buddhism was dominated by several competing forms of Yogācāra (see Buddhist philosophy, Chinese). A major schism between orthodox versions of Yogācāra and Tathāgatagarbha hybrid versions was finally settled in the eighth century in favour of a hybrid version, which became definitive for all subsequent forms of East Asian Buddhism. Yogācāra ideas were also studied and classified in Tibet. The Nyingma and Dzog Chen schools settled on a hybrid version similar to the Chinese Tathāgatagarbha hybrid; the Gelugpas subdivided Yogācāra into a number of different types and considered them preparatory teachings for studying Prāsaōgika Madhyamaka, which Gelugpas ranked as the highest Buddhist teaching. The Tibetans, however, tended to view the logico-epistemological tradition as distinct from Yogācāra proper, frequently labelling them Sautrāntika instead.

2 Yogācāra contrasted with metaphysical idealism

The school was called Yogācāra (Yoga practice) because it provided a comprehensive, therapeutic framework for engaging in the practices that lead to the goal of the *bodhisattva* path, namely enlightened cognition. Meditation served as the laboratory in which one could study how the mind operated. Yogācāra focused on the question of consciousness from a variety of approaches, including meditation, psychological analysis, epistemology (how we know what we know, how perception operates, what validates knowledge), scholastic categorization and karmic analysis.

Yogācāra doctrine is summarized in the term *vijñaptimātra*, 'nothing-but-cognition' (often rendered as 'consciousness-only' or 'mind-only'), which has sometimes been interpreted as indicating a type of metaphysical idealism, that is, the claim that the mind alone is real and everything else is created by the mind. However, the Yogācārin writings themselves argue something very different. Consciousness (*vijñāna*) is not the ultimate reality or solution, but rather the root problem. This problem emerges in ordinary mental operations and it can only be solved by bringing those operations to an end.

Yogācāra tends to be misinterpreted as a form of metaphysical idealism primarily because its teachings are taken for ontological propositions rather than as epistemological warnings about karmic problems. The Yogācāra focus on cognition and consciousness grew out of its analysis of karma and not for the sake of metaphysical speculation. Two things should be clarified in order to explain why Yogācāra is not metaphysical idealism: the meaning of the term 'idealism' and the important difference between the way Indian and Western philosophers regard philosophy.

The term 'idealism' came into vogue roughly during the time of I. Kant (although it was used earlier by others, such as G.W. Leibniz) for one of two trends that had emerged in reaction to Cartesian philosophy. R. Descartes had argued that there were two basic yet separate substances in the universe: extension (the material world of things in space) and thought (the world of mind and ideas) (see German idealism). Subsequently, opposing camps took one or the other substance as their metaphysical foundation, treating it as the primary substance, while reducing the remaining substance to a derivative status. Materialists argued that only matter was ultimately real so that thought and consciousness derived from physical entities (such as chemistry, brain states). Idealists countered that the mind and its ideas were ultimately real and that the physical world derived from the mind (for example, the mind of God, G. Berkeley's *esse est percipi* (to be is to be perceived), or from ideal prototypes) (see Berkeley, G.). Materialists gravitated toward mechanical, physical explanations for why and how things existed, while Idealists tended to look for purposes - moral as well as rational - to explain existence. Idealism meant 'idea-ism', frequently in the sense of Plato's notion of 'ideas' (*eidos*), denoting ideal types that transcended the physical, sensory world and provided the form (*eidos*) that gave matter meaning and purpose. As materialism buttressed by advances in materialistic science gained wider acceptance, those inclined towards spiritual and theological aims turned

increasingly toward idealism as a countermeasure. Before long there were many types of materialism and idealism.

Idealism in its broadest sense came to encompass everything that was not materialism, which included so many different types of positions that the term lost any hope of univocality. Most forms of theistic and theological thought were, by this definition, types of idealism, even if they accepted matter as real, since they also asserted something as more real than matter, either as the creator of matter (in monotheism) or as the reality behind matter (in pantheism) (see Pantheism). Extreme empiricists who only accepted their own experience and sensations as real were also idealists. Thus, 'idealism' united monotheists, pantheists and atheists. At one extreme were various forms of metaphysical idealism which posited the mind as the only ultimate reality. The physical world was either an unreal illusion, or not as real as the mind that created it. To avoid solipsism (which is a subjectivized version of metaphysical idealism), metaphysical idealists posited an overarching mind that envisages and creates the universe.

A more limited type of idealism is epistemological, which argues that since knowledge of the world only exists in the mental realm we cannot know actual physical objects as they truly are, but only as they appear in our mental representations of them. Epistemological idealists could be ontological materialists, accepting that matter exists substantially; they could even accept that mental states derived at least in part from material processes. What they denied was that matter could be known in itself directly without the mediation of mental representations. Although unknowable in itself, the existence of matter and its properties could be known through inference based on certain consistencies in the way material things are represented in perception.

Transcendental idealism contends that not only matter, but also the self remains transcendental in an act of cognition. Kant and Husserl, who were both transcendental idealists, defined 'transcendental' as that which constitutes experience but is not itself given in experience. A mundane example would be the eye, which is the condition for seeing even though the eye does not see itself. By applying vision and drawing inferences from it, one can come to know the role eyes play in seeing, although one never sees one's own eyes. Similarly, things in themselves and the transcendental self could be known if the proper methods were applied for uncovering the conditions that constitute experience, although such conditions do not themselves appear in experience. Even when epistemological issues are at the forefront, it is ontological concerns, such as the status of self and objects, that are really at stake. Western philosophy rarely escapes this ontological slant. Those who accepted that both the self and its objects were unknowable except through reason and that such reason was their cause and purpose for existing - thus epistemologically and ontologically grounding everything in the mind and its ideas - were called absolute Idealists, such as F.W.J. von Schelling, G.W.F. Hegel and F.H. Bradley, since only such ideas are absolute while all else is relative to them (see Absolute, the).

With the exception of some epistemological idealists, what unites all the positions described thus far, including the materialists, is that they are ontological. They are concerned with the ontological status of the objects of sense and thought, as well as the ontological nature of the self who knows. Mainstream Western philosophy since Plato and Aristotle has treated ontology and metaphysics as the ultimate philosophic pursuit, with epistemology's role being little more than to provide access and justification for one's ontological pursuits and commitments. Since many of what are decried as philosophy's excesses - such as scepticism, solipsism and sophistry - have been accused of deriving from overactive epistemological questioning, epistemology has often been held to be suspect and in some theological formulations, considered entirely dispensable in favour of faith. Ontology is primary and epistemology is either secondary or expendable.

3 The role of epistemology

In Indian philosophy one finds the reverse of the positions described. Epistemology (*pramāṇavāda*) is primary both in the sense that it must be engaged in prior to attempting any other philosophical endeavour and that the limits of one's metaphysical claims are always inviolably set by the parameters established by one's epistemology. Before one can make claims, one must establish the basis on which such claims can be proven and justified. The Indians went so far as to concede that if one wishes to debate with an opponent, one must first find a common epistemological ground upon which to argue. Failing that, no meaningful debate can take place.

Since one's ontology (*prameya*) depends on what one's epistemology makes allowable, many Indian schools tried to include things in their list of valid means of knowledge (*pramāṇa*) that would facilitate their claims. Hindus, for

instance, considered their Scriptures to be valid means of knowledge, but other Indians, such as Buddhists and Jains, rejected the authority of the Hindu Scriptures. Therefore, if a Hindu debated with a Buddhist or Jain, he or she could not appeal to the authority of Hindu Scriptures, but had to find common epistemological ground. In the case of Buddhism that would be perception and inference; in the case of Jainism, it would only be inference. All schools except Jains accepted perception as a valid means of knowledge, meaning that sensory knowledge is valid (if qualified as non-erroneous or nonhallucinatory). What is not presently observed but is in principle observable can be known by inference. Without actually seeing the fire, one knows it must exist on a hill when one sees smoke in that location because both fire and smoke are in principle observable entities and an observed necessary relation (*vyāpti*) exists between smoke and fire (think of the expression, there is no smoke without fire). If one were close to the fire on the hill, the fire could be observed. One cannot make valid inferences about things impossible to perceive, such as unicorns, since no necessary observable relation is obtainable so one cannot infer that a unicorn is on the hill. Perceptibility therefore is an indispensable component of both perception and inference, and thus for Buddhists, of all valid knowledge. In order to be considered 'real' (*dravya*) by the standards of Buddhist logic, a thing must produce an observable effect. Buddhists argued among themselves whether something was real only while it was producing this observable effect (the Sautrāntika position), or whether something could be considered real if it produced an observable effect at some moment during its existence (the Sarvāstivāda position). However, it was generally agreed that a thing must have observable causal efficacy (*kāraṇa*) in order to be considered real. This helps explain the centrality of perception and consciousness for Yogācāra theory.

The logico-epistemological branch of Yogācāra drew a sharp distinction between perception and inference. Perception involves sensory cognitions of unique, momentary, discrete particulars. Inference involves linguistic, conceptual universals, since words are meaningful and communicative only to the extent they designate and participate in universal classes commonly shared and understood by users of the language. Inferences are true or false depending on how accurately they approximate sensory particulars, but even when linguistically true, they are only still true relative (*saṃvṛti*) to the sensations they approximate. Conversely, sensation (and only sensation) is beyond language. Sensory cognition devoid of linguistic overlay or theoretical assertions (*samāropa*) is correct cognition and precisely, not approximately, true (*paramārtha*). While this seems to involve metaphysical claims about categories such as particulars and universals, sensation and language, in fact it is a request that we should cognize things as they are without imposing any metaphysical assertions or conceptual framework whatsoever. The cognitive and epistemic, not the metaphysical, is at stake. What is the case is beyond description, not because it is something ineffable residing outside or behind human experience, but because it is the very sensory stuff of human experience whose momentary unique actuality cannot be reduced to universalistic, eternalistic language or concepts. To interpret this position itself as a metaphysics of particularity is to miss its point.

4 Primacy of epistemology

Epistemological concerns pervade Indian philosophy. This is especially true of Buddhist philosophy. Many Buddhist texts assert that higher understanding has nothing to do with ontology, that focusing on the existence or nonexistence of something (*astināsti*, *bhāvābhāva*) is a misleading category error. They typically remove important items - such as emptiness and *nirvāṇa* - from ontological consideration by explicitly declaring that these have nothing to do with existence or nonexistence, or being and nonbeing. They further warn that this is not a licence to imagine a higher sense of existence or being into which such items are then subsumed or sublated. The Buddhist goal is not the construction of a more perfect ontology. Instead its primary target is always the removal of ignorance. Hence, while Buddhists frequently suspend ontological and metaphysical speculation (*tarka*), denouncing it as useless or dangerous, correct cognition (*samyagjñāna*) is invariably lauded. Even Madhyamakas, who question the feasibility of much of Buddhist epistemology, insist that we should understand where the errors lie and correct the way we cognize accordingly. Stated bluntly, Buddhism is concerned with seeing, not being, that is, epistemology rather than ontology.

Tellingly, no Indian Yogācāra text ever claims that the world is created by the mind. What they do claim is that we mistake our projected interpretations of the world for the world itself, that is, we take our own mental constructions to be the world. Their vocabulary for this is as rich as their analysis: *kalpanā* (projective conceptual construction), *parikalpa* and *parikalpita* (ubiquitous imaginary constructions), *abhūtaparikalpa* (imagining something in a locus in which it does not exist), *prapañca* (proliferation of conceptual constructions), to mention a

few. Correct cognition is defined as the removal of those obstacles which prevent us from seeing dependent causal conditions in the manner they actually become (*yathābhūtam*). For Yogācāra these causal conditions are cognitive, not metaphysical; they are the mental and perceptual conditions by which sensations and thoughts occur, not the metaphysical machinations of a Creator or an imperceptible domain. What is known through correct cognition is euphemistically called *tathatā*, or 'suchness', which the texts are quick to point out is not an actual thing, but only a word (*prajñaptimātra*).

What is crucial in the foregoing for understanding Yogācāra is that its attention to perceptual and cognitive issues is in line with basic Buddhist thinking and that this attention is epistemological rather than metaphysical. When Yogācārins discuss 'objects' they are talking about cognitive objects, not metaphysical entities. Rather than offer one more ontology, they attempt to uncover and eliminate the predilections and proclivities (*āśrava*, *anuśaya*) that compel people to generate and cling to such theoretical constructions. According to Yogācāra, all ontologies are epistemological constructions. Therefore, to understand how cognition operates is to understand how and why people construct the ontologies to which they cling. Ontological attachment is a symptom of cognitive projection (*pratibimba*, *parikalpita*). Careful examination of Yogācāra texts reveals that they make no ontological claims, except to question the validity of making those ontological claims. Such ontological silence derives from the fact that, were they to offer a metaphysical description, that description would be appropriated by its interpreters whose proclivities would project onto it what they wish reality to be, thereby reducing the description to their own presupposed theory of reality. Such projective reductionism is the problem and is the definition of *vijñaptimātra*: to mistake one's projections for that onto which one is projecting. Vasubandhu's *Triṃśikā(Thirty Verses)* states that if one clings to one's projection of the idea of *vijñaptimātra*, then one fails to dwell truly in the understanding of *vijñaptimātra* (verse 27 [1984]). Enlightened cognition free of all cognitive errors is defined as *nirvikalpajñāna*, or 'cognition without imaginative construction', that is, without conceptual overlay. Ironically, Yogācāra's interpreters and opponents could not resist reductively projecting metaphysical theories onto what Yogācārins did say, at once proving Yogācāra was right and making actual Yogācāra teachings that much harder to understand. Interpreting their epistemological analyses as metaphysical pronouncements fundamentally misconstrues their project.

The arguments Yogācāra deploys frequently resemble those made by epistemological idealists. Recognizing those affinities, early twentieth-century Western scholars compared Yogācāra to Kant. More recently scholars have begun to think that Husserl's phenomenology comes even closer. Indeed, there are intriguing similarities, for instance, between Husserl's description of noesis (consciousness projecting its cognitive field) and noema (the constructed cognitive object) on the one hand, and Yogācāra's analysis of the (cognitive) grasper and the grasped (*grāhaka* and *grāhya*) on the other hand. However, there are also important differences between the ideas of those Western philosophers and Yogācāra, most markedly that Kant and Husserl play down notions of causality, while Yogācāra developed complex systematic causal theories it deemed to be of the greatest importance; there is no counterpart to either karma or enlightenment in the Western theories, while these are the very *raison d'être* for all of Yogācāra theory and practice; finally, Western philosophies are produced to afford the best possible access to an ontological realm (at least sufficient to acknowledge its existence), while Yogācāra is critical of that motive in all its manifestations. To the extent that epistemological idealists can also be critical realists, Yogācāra may be deemed a type of epistemological idealism with the proviso that the purpose of its arguments was not to engender an improved ontological theory or commitment, but rather an insistence that we shift our attention to the epistemological and psychological conditions that compel us to construct and attach to ontological theories (seeEpistemology, Indian schools of; Ontology in Indian philosophy).

5 Karma

The key to Yogācāra theory lies in the Buddhist notions of karma which they inherited and rigorously reinterpreted (see Karma and rebirth, Indian conceptions of). As earlier Buddhist texts explained, karma is responsible for suffering and ignorance and consists of any intentional activity of body, language, or mind. Since the crucial factor is intent and intent is a cognitive condition, whatever is noncognitive must necessarily lack intent and be nonkarmic. Hence, by definition, whatever is noncognitive can have no karmic influence or consequences. Since Buddhism aims at overcoming ignorance and suffering through the elimination of karmic conditioning, Buddhism, they reasoned, is only concerned with the analysis and correction of whatever falls within the domain of cognitive conditions. Hence questions about the ultimate reality of that which is noncognitive are irrelevant for solving the

problem of karma. Further, Yogācārins emphasize that categories such as materiality (*rūpa*) are cognitive. Materiality is an abstract classification of the colours, textures and sounds experienced in acts of perception and it is only to the extent that they are experienced, perceived and ideologically grasped, thereby becoming objects of attachment, that they have karmic significance. Intentional acts also have moral motives and consequences. Since effects are shaped by their causes, an act with a wholesome intent would tend to yield wholesome fruits, while unwholesome intentions produce unwholesome effects.

In contrast to the cognitive karmic dimension, Buddhism considered material elements (*rūpa*) to be karmically neutral. The problem with material things is not their materiality, but the psychology of appropriation (*upādāna*) - desiring, grasping, clinging, attachment - that infests our ideas and perceptions of such things. It is not the materiality of gold that leads to problems, but rather our *ideas* about the value of gold and the attitudes and actions we engage in as a result of those ideas. Those ideas were acquired through previous experiences. By repeated exposure to certain ideas and cognitive conditions, one is conditioned to respond habitually in a similar manner to similar circumstances. Eventually these habits are embodied, becoming reflexive and presuppositional. For Buddhists this process of embodied conditioning (*saṃskāra*) is not confined to a single lifetime, but accrues over many. *Saṃsāra* (the continuous cycle of birth and death) is the karmic enactment of this repetition, the reoccurrence of cognitive embodied habits in new life situations and life forms.

For all Buddhists this follows a simple sensory calculus by which we are conditioned: pleasurable feelings we wish to retain, or repeat; painful feelings we wish to cut off, or avoid; pleasure and pain, reward and punishment, approval and disapproval. Our karmic habits (*vāsanā*) are constructed this way. Since nothing is permanent, pleasurable feelings cannot be maintained or repeated permanently; painful things (such as sickness and death) cannot be avoided permanently. However, we do imagine all sort of permanent things, from God to souls to essences, in an effort to avoid facing the fact that none of us has a permanent self. We believe that if we can prove something is permanent then we too have a chance for permanence. The greater the dissonance between our actual impermanent experience and our expectations for permanent desired ends, the more we suffer and the greater tendency (*anuśaya*) towards projecting our desires onto the world. The anxiety about our lack of self and all the cognitive and karmic mischief it generates is called several things by Yogācāra, including *jñeyāvaraṇa* (obstruction of the knowable, that is, our self-obsessions which prevent us from seeing things as they are) and *abhūtaparikalpa* (imagining something, such as permanence or a self, to exist in a locus in which it is absent).

6 The storage and ripening of karma

Previous Buddhists, especially in the Abhidharma schools, had developed a sophisticated metaphoric vocabulary to describe and analyse the causes and conditions of karma in terms of seeds (*bījā*). Just as a plant develops from its roots unseen underground, so do previous karmic experiences fester unseen in the mind; just as a plant sprouts from the ground when nourished by proper conditions, so karmic habits under the right causes and conditions reassert themselves as new experiences; just as plants reach fruition by producing new seeds that re-enter the ground to take root and begin regrowing a similar plant of the same kind, so karmic actions produce wholesome or unwholesome fruit that become latent seeds for a later, similar type of action or cognition. Just as plants reproduce only their own kind, so do wholesome or unwholesome karmic acts produce effects after their own kind. This cycle served as a metaphor for the process of cognitive conditioning, as well as the recurrent cycle of birth and death (*saṃsāra*). Since Yogācāra accepts the Buddhist doctrine of momentariness, seeds are said to perdure for only a moment during which time they become the cause of a similar seed that succeeds them. Momentary seeds are causally linked in sequential chains, each momentary seed a link in a chain of karmic causes and effects (see Momentariness, Buddhist doctrine of).

Seeds are basically divided into two types: wholesome and unwholesome. Unwholesome seeds are the acquired cognitive habits preventing one from reaching enlightenment. Wholesome, or pure, unpolluted seeds give rise to more pure seeds, which bring one closer to enlightenment. In general Yogācāra differentiates inner seeds (personal conditioning) from external seeds (being conditioned by others). One’s own seeds can be modified or affected by exposure to external conditions (external seeds), which can be beneficial or detrimental. Exposure to polluting conditions intensifies one’s unwholesome seeds, while contact with pure conditions, such as hearing the correct teaching (*Saddharma*), can stimulate one’s wholesome seeds to increase, thereby diminishing and ultimately eradicating one’s unwholesome seeds.

Another metaphor for karmic conditioning that accompanies the seed metaphor is ‘perfuming’ (*vāsanā*). A cloth exposed to the smell of perfume acquires its scent. Similarly one is mentally and behaviourally conditioned by what one experiences. This conditioning produces karmic habits, but just as the odour can be removed from the cloth so can one’s conditioning be purified of perfumed habits. Typically three types of perfuming are discussed: linguistic and conceptual habits; habits of self-interest and ‘grasping self’ (*ātmagrāha*), that is, the belief in self and what belongs to self; and habits leading to subsequent life situations (*bhāvāōgavāsanā*), that is, the long-term karmic consequences of specific karmic activities.

Yogācāra literature debates the relation between seeds and perfuming. Some claim that seeds and perfuming are really two terms for the same thing, that is, acquired karmic habits. Others claim that seeds are simply the effects of perfuming, so that all conditioning is acquired through experience. Still others contend that ‘seed’ refers to the chains of conditioned habits one already has (whether acquired in this life, in some previous life, or even ‘beginninglessly’), while ‘perfuming’ denotes the experiences one has that modify or affect the development of one’s seeds. ‘Beginningless’ might be understood as a corollary to Husserl’s term ‘transcendental’, that is, a causal sequence constituting a present experience whose own original cause remains undisclosed in this experience. Some claimed that one’s possibilities for enlightenment depended entirely on the sort of seeds one already possessed; perfuming merely acted as a catalyst, but could not provide wholesome seeds if one did not already possess them. Beings utterly devoid of wholesome seeds were called *icchantikas* (incorrigibles); such beings could never reach enlightenment. Some other Mahāyāna Buddhists, feeling that this violated the Mahāyāna dictum of universal salvation, attacked the incorrigibility doctrine.

The karmic cause of the fundamental disease (*duḥkha*) is desire expressed through body, speech, or mind. Therefore Yogācāra focused exclusively on cognitive and mental activities in relation to their intentions, that is, the operations of consciousness, since the problem was located there. Buddhism had always identified ignorance and desire as the primary causes of suffering and rebirth. Yogācārins mapped these mental functions in order to dismantle them. Because maps of this sort were also creations of the mind they too would ultimately have to be abandoned in the course of the dismantling, but their therapeutic value would have been served in bringing about enlightenment. This view of the provisional expediency of Buddhism can be traced back to Buddha himself. Yogācārins describe enlightenment as resulting from overturning the cognitive basis (*āśrayaparavṛtti*), that is, overturning the conceptual projections and imaginings which act as the base of our cognitive actions. This overturning transforms the basic mode of cognition from consciousness (*vi-jñāna*, dis-cernment) into *jñāna* (direct knowing). The *vi-* prefix is equivalent to *dis-* in English (dis-criminate, dis-tinguish) meaning to bifurcate or separate from. Direct knowing was defined as nonconceptual (*nirvikalpa*), that is, devoid of interpretive overlay.

7 Matter and cognitive appropriation

The case of material elements is important for understanding one reason why Yogācāra is not metaphysical idealism. No Yogācāra text denies materiality (*rūpa*) as a valid Buddhist category. On the contrary, Yogācārins include materiality in their analysis. Their approach to materiality is well rooted in Buddhist precedents. Frequently Buddhist texts substitute the term ’sensory contact’ (Pāli: *phassa*, Sanskrit: *sparśa*) for the term ’materiality’. This substitution is a reminder that physical forms are sensory, that they are known to be what they are through sensation. Even the earliest Buddhist texts explain that the four primary material elements are the sensory qualities solidity, fluidity, temperature and mobility; their characterization as earth, water, fire, and air, respectively, is declared an abstraction. Instead of concentrating on the fact of material existence, one observes how a physical thing is sensed, felt, perceived. Yogācāra never denies that there are sense-objects (*viśaya*, *artha*, *ālambana*, etc.), but it denies that it makes any sense to speak of cognitive objects occurring outside an act of cognition. Imagining such an occurrence is itself a cognitive act. Yogācāra is interested in why we feel compelled to so imagine.

Everything we know, conceive, imagine, or are aware of, we know through cognition, including the notion that entities might exist independent of our cognition. The mind does not create the physical world, but it produces the interpretive categories through which we know and classify the physical world. It does this so seamlessly that we mistake our interpretations for the world itself. Those interpretations which are projections of our desires and anxieties become obstructions (*āvaraṇa*) preventing us from seeing what is actually the case. In simple terms we are blinded by our own self-interests, prejudices (which means what is already prejudged) and desires.

Unenlightened cognition is an appropriative act in which cognitive objects are apprehended (*upalabdhi*). Yogācāra does not speak about subjects and objects; instead it analyses perception in terms of graspers (*grāhaka*) and what is grasped (*grāhya*).

Yogācāra at times resembles epistemological idealism, which does not claim that this or any world is constructed by mind, but rather that we are usually incapable of distinguishing our mental constructions and interpretations of the world from the world itself. This narcissism of consciousness Yogācāra calls *vijñaptimātra*, 'nothing but conscious construction'. A deceptive trick is built into the way consciousness operates at every moment. Consciousness projects and constructs a cognitive object in such a way that it disowns its own creation, pretending the object is 'out there' in order to render that object capable of being appropriated. Even while what we cognize is occurring within our act of cognition, we cognize it *as if* it were external to our consciousness. Realization of *vijñaptimātra* exposes this trick intrinsic to consciousness's workings, thereby eliminating it. When that deception is removed one's mode of cognition is no longer termed *vijñāna* (consciousness); it has become direct cognition, *jñāna*. Consciousness engages in this deceptive game of projection, dissociation and appropriation because there is no 'self'. According to Buddhism, the deepest, most pernicious erroneous view held by sentient beings is that a permanent, eternal, immutable, independent self exists. There is no such self and deep down we know that. This makes us anxious as it indicates that no self or identity endures forever. To assuage that anxiety, we attempt to construct a self, to fill the anxious void, to do something enduring. The projection of cognitive objects for appropriation is consciousness's main tool for this construction. If I own things (ideas, theories, identities, material objects), then 'I am'. If there are eternal objects that I can possess, then I too must be eternal. To undermine this desperate and erroneous appropriative grasping, Yogācāra texts such as the *Madhyāntavibhāga* say: 'Negate the object, and the self is also negated' (1: 4, 8).

Yogācārins deny the existence of external objects in two senses. In terms of conventional experience they do not deny the existence of objects such as chairs, colours and trees, but rather reject the claim that such things appear anywhere other than in consciousness. It is externality, not objects *per se*, that they challenge. They believe moreover that while such objects are admissible as conventions, in more precise terms there are no chairs, or trees. These are merely words and concepts by which we gather and interpret discrete sensations that arise moment by moment in a causal flux. These words and concepts are mental projections. The point is not to elevate consciousness, but to warn us not to be fooled by our own cognitive narcissism. Enlightened cognition is likened to a great mirror that impartially and fully reflects everything before it without attachment to what has passed nor in expectation of what might arrive.Yogācārins refuse to provide answers to the question of what sorts of objects enlightened ones cognize, apart from saying it is purified of karmic pollution (*anāśrava*), since whatever description they might offer would only be appropriated and reduced to the habitual cognitive categories that are already preventing us from seeing properly.

8 Eight types of consciousnesses

The most famous innovation of the Yogācāra school was the doctrine of eight types of consciousness. Standard Buddhism described six types of consciousness, each produced by the contact between its specific sense organ and a corresponding sense object. When a functioning eye comes into contact with a colour or shape, visual consciousness is produced. When a functioning ear comes into contact with a sound, auditory consciousness is produced. Consciousness does not create the sensory sphere, but is an effect of the interaction of a sense organ and its proper object. If an eye does not function but an object is present, visual consciousness does not arise. The same is true if a functional eye fails to encounter a visual object. Consciousness depends on sensation. There are a total of six sense organs (eye, ear, nose, mouth, body and mind) which interact with their respective sensory object domains (visual, auditory, olfactory, gustatory, tactile and mental spheres). These domains in turn amount to six distinct types of consciousness. It is worthy of note that the mind is considered another sense as it functions like the other senses with the activity of a sense organ (*manas*), its domain (*manodhātu*) and the consciousness (*manovijñāna*) resulting from the contact of organ and object. Each domain is discrete so that vision, audition and each of the remaining spheres function apart from each other. Hence the deaf can see and the blind can hear. Objects, too, are entirely specific to their domain and the same is true of consciousnesses. Visual consciousness is entirely distinct from auditory consciousssness, and so on. These eighteen components of experience, that is, six sense organs, six sense object domains and six resulting consciousnesses, were called the eighteen *dhātus*. According to standard Buddhist doctrine these eighteen exhaust the full extent of everything in the universe, or

more accurately, the sensorium.

Early Buddhist Abhidhamma, focusing on the mental and cognitive aspects of karma, expanded the three components of the mental level - mind (*manas*), mental objects (*manodhātu*) and mental consciousness (*manovijñāna*) - into a complex system of categories. The apperceptive vector in any cognitive moment was called *citta*. The objects, textures, emotional, moral and psychological tones of *citta*'s cognitions were called *caittas* (literally, 'associated with *citta*'). *Caittas* were subdivided into numerous categories that varied in different Buddhist schools. Some *caittas* are universal, that is, they are components of every cognition (for example, sensory contact, hedonic tone, attention); some are specialized, that is, they only occur in some, not all, cognitions (for example, resolve, mindfulness, meditative clarity). Some *caittas* are wholesome (for example, faith, tranquillity, lack of greed, hatred, or misconception), some unwholesome; some are mental disturbances (*kleśa*) (appropriational intent, aversion, arrogance), or secondary mental disturbances (anger, envy, guile, shamelessness) and some are karmically indeterminate (torpor, remorse).

As Abhidharma grew more complex, disputes intensified between different Buddhist schools across a range of issues. For Yogācāra the most important problems revolved around questions of causality and consciousness. To avoid the idea of a permanent self, Buddhists claimed *citta* was momentary. Since a new *citta* apperceived a new cognitive field each moment, the apparent continuity of mental states was explained by claiming each *citta*, in the moment it ceased, also caused its successor. This was fine for continuous perceptions and thought processes, but difficulties arose since Buddhists identified a number of situations in which no *citta* at all was present or operative, including certain meditative conditions explicitly defined as devoid of *citta*. If a preceding *citta* had to be temporally contiguous with its successor, how could one explain the sudden restarting of *citta* after a period of time had elapsed since the prior *citta* had ceased? Where had *citta* or its causes been residing in the interim? Analogous questions arose, such as, from where does consciousness re-emerge after deep sleep? How does consciousness begin in a new life? The various Buddhist attempts to answer these questions led to more difficulties and disputes.

Yogācārins responded by rearranging the tripartite structure of the mental level of the eighteen *dhātus* into three novel types of consciousnesses. *Manovijñāna* (empirical consciousness) became the sixth consciousness (and operated as the sixth sense organ, which previously had been the role of *manas*), surveying the cognitive content of the five senses as well as mental objects. *Manas* became the seventh consciousness, redefined as primarily obsessed with various aspects and notions of self and thus called 'defiled *manas*' (*kliṣṭamanas*). The eighth consciousness, *ālayavijñāna* (warehouse consciousness) was novel. Warehouse consciousness was defined in several ways. It is the receptacle of all seeds, storing experiences as they 'enter' until they are sent back out as new experiences, like a warehouse handles goods. It was also called *vipāka* consciousness, meaning the maturing of karmic seeds when the consequences of previous karmic actions have matured sufficiently to actualize in present experience. Seeds gradually matured in the repository consciousness until karmically ripe, when they reassert themselves as karmic consequences. *Ālayavijñāna* was also called the basic consciousness (*mūlavijñāna*) as it retains and deploys the karmic seeds that both influence and are influenced by the other seven modes of consciousness. When, for instance, the sixth consciousness is dormant (while one sleeps, or is unconscious), its seeds reside in the eighth consciousness and they restart when the conditions for their arising are present. The eighth consciousness is largely a mechanism for storing and deploying seeds of which by and large it remains unaware. *Cittas* occur as a stream in *ālayavijñāna*, but they mostly cognize the activities of the other consciousnesses, not their own seeds. In those states devoid of *citta*, the flow of *cittas* are repressed, but their seeds continue to regenerate without being noticed until they reassert a new stream of *cittas*. Warehouse consciousness acts as the pivotal karmic mechanism, but is itself karmically neutral. Each individual has its own warehouse consciousness which perdures from moment to moment and life to life. Although it is nothing more than a collection of everchanging 'seeds', it is continually changing and therefore not a permanent self. There is no universal collective mind in Yogācāra.

Enlightenment consists in bringing the eight consciousness to an end, replacing them with enlightened cognitive abilities (*jñāna*). Overturning the basis transforms the five sense consciousnesses into immediate cognitions that accomplish what needs to be done (*kṛtyānuṣṭhānajñāna*). The sixth consciousness becomes immediate cognitive mastery (*pratyavekṣaṇajñāna*), in which the general and particular characteristics of things are discerned just as they are. This discernment is considered nonconceptual (*nirvikalpajñāna*). *Manas* becomes the immediate

cognition of equality (*samatājñāna*), equalizing self and other. When the warehouse consciousness finally ceases it is replaced by the great mirror cognition (*Mahādarśajñān*) that sees and reflects things just as they are, impartially, without exclusion, prejudice, anticipation, attachment, or distortion. The grasper-grasped relation has ceased. It should be noted that all of these 'purified' cognitions engage the world in immediate and effective ways by removing the self-bias, prejudice and obstructions that previously had prevented one from perceiving beyond one's own narcissistic consciousness. When consciousness ends, true knowledge begins. One more Yogācāra innovation was the notion that a special type of cognition emerged and developed *after* enlightenment. This post-enlightenment cognition was called *pṛṣṭhalabdhajñāna*. Since enlightened cognition is nonconceptual its objects cannot be described.

9 Three self-natures

The three self-nature theory (*trisvabhāva*), which is explained in many Yogācāra texts including an independent treatise by Vasubandhu, maintains that there are three 'natures' or cognitive realms at work: the conceptually constructed realm (*parikalpitasvabhāva*) ubiquitously imputes unreal conceptions, especially permanent 'selves', into whatever it experiences, including oneself; the realm of causal dependency (*paratantrasvabhāva*), when mixed with the constructed realm, leads one to mistake impermanent occurrences in the flux of causes and conditions for fixed, permanent entities. It can be purified of these delusions by the third element, the perfectional realm (*pariniṣpannasvabhāva*), which, like the Madhyamaka notion of emptiness on which it is based, acts as an antidote (*pratipakṣa*) that 'purifies' all delusional constructions out of the causal realm. When the causally dependent realm is cleansed of all defilements it becomes 'enlightened'. These self-natures are also called the three non-self-natures, as they lack fixed, independent, true, permanent identities and therefore should not be hypostatized. The first is unreal by definition; the third is intrinsically 'empty' of self-nature. The second (which is the only 'real' one) is of an unfixed nature as it can be 'mixed' with either of the other two. Understanding the purified second nature is equivalent to understanding dependent origination (*pratītyasamutpāda*), which all schools of Buddhism accept as Buddhism's core doctrine and which tradition claims Buddha came to realize under the Bodhi tree on the night of his enlightenment (see Self, Indian theories of).

10 Five stages

Yogācāra literature is so vast that one should not be surprised to find that many of the attempts to provide detailed systems run into conflict with each other. Since it was a self-critical scholastic tradition, it was not uncommon for Yogācāra texts to discuss and criticize the positions of other Yogācāra works, as well as their more obvious opponents. Yogācāra positions on the stages of the path are diverse. The *Daśabhūmikasūtraśāstra*, a commentary on the *Ten Stages Scripture* attributed to Vasubandhu, describes the progress of the Bodhisattva path to Mahāyānic liberation in ten stages, comparable to the ten stages implicit in the Mahāyānic formulation of the ten perfections of wisdom. Asaōga's *Yogācārabhūmiśāstra* describes a series of seventeen stages. There are other formulations, such as the five-stage path that offers a useful survey of the other formulations.

The first stage is called 'provisioning' (*sambhārāvasthā*) since this is the stage at which one collects and stocks up on provisions for the journey. These provisions primarily consist of orienting oneself toward the pursuit of the path and developing the proper character, attitude and resolve to accomplish it. It begins the moment the aspiration for enlightenment arises (*bodhicitta*). The next stage is the 'experimental' stage (*prayogāvasthā*), in which one begins to experiment with correct Buddhist theories and practices, learning which of them work and which are true. One begins to suppress the grasper-grasped relation and begins to study carefully the relation between things, language and cognition. After honing one's discipline, one eventually enters the third stage, 'deepening understanding' (*prativedhāvasthā*). Some texts refer to this as the path of corrective vision (*darśanamārga*). This stage ends once one has acquired some insight into nonconceptual cognition.

Nonconceptual cognition deepens in the next stage, the path of cultivation (*bhāvanāmārga*). The grasper-grasped relation is utterly eliminated as are all cognitive obstructions. This path culminates in the overturning of the basis, or enlightenment. In the final stage (*niṣṭhāvasthā*), one abides in unexcelled complete enlightenment and engages the world through the five immediate cognitions. All one's activities and cognitions at this stage are 'postrealization'. As a Mahāyānist, from the first stage one has been devoting oneself not only to one's own attainment of enlightenment, but to the attainment of enlightenment by all sentient beings. In this stage such a pursuit becomes one's sole concern.

See also: Buddhist philosophy, Indian; Buddhist philosophy, Japanese; Wônch'ūk

DAN LUSTHAUS

References and further reading

Anacker, S. (1984) *Seven Works of Vasubandhu*, Delhi: Motilal Banarsidass.(Important translation and discussion of key works by Vasubandhu, including some of his pre-Yogācāra treatises and the *Triṃśikā(Thirty Verses)*, 183-90.)

Buswell, R.E. (1989) *The Formation of Ch'an Ideology in China and Korea: The Vajrasamādhisūtra, a Buddhist Apocryphon*, Princeton, NJ: Princeton University Press.(The introductory essay contains the most thorough published account in English of the debate between Chinese Yogācāra and Tathāgatagarbha schools. However, its presentation represents the latter's viewpoint which substantially distorts actual Yogācāra positions.)

Griffiths, P. (1986) *On Being Mindless: Buddhist Meditation and the Mind-Body Problem*, La Salle, IL: Open Court.(Analytic philosophical discussion of the 'mindless' cessation meditations, translating and examining some relevant sections of Theravāda, Vaibhāṣika and Yogācāra texts. Helpful for understanding how Yogācārin positions relate to Buddhism at large.)

Griffiths, P., Hakamaya, N., Keenan, J. and Swanson, P. (1989) *The Realm of Awakening: Chapter Ten of Asaōga's Mahāyānasaṃgraha*, New York and Oxford: Oxford University Press.(Collective effort from a class taught by Hakamaya at the University of Wisconsin. Compares and translates Chinese and Tibetan versions of the root text as well as major commentaries. A good presentation of the scholastically dense style of some Yogācāra texts.)

Hayes, R.P. (1988) *Dignāga on the Interpretation of Signs, Studies of Classical India*, vol. 9, Dordrecht: Kluwer Academic Publishers.(A superb and challenging examination of Buddhism's most famous logician.)

Kochumuttom, T. (1982) *A Buddhist Doctrine of Experience*, Delhi: Motilal Banarsidass.(Translation and critical analysis of Vasubandhu's major Yogācāra texts, including the *Triṃśikā(Thirty Verses)*, 127-63. It argues that Vasubandhu's texts should be interpreted as critical realism rather than idealism.)

Laōkāvatārasūtra (*c*.5th-6th century), trans. D.T. Suzuki, Boulder, CO: Prajña Press, 2nd edn, 1978.(Claims to be based on the Sanskrit, but more closely follows the Chinese editions.)

Lusthaus, D. (forthcoming) *Buddhist Phenomenology: The Cheng Weishilun in Philosophic Context*, Albany, NY: State University of New York Press.(An investigation of the transmission of Yogācāra philosophy from India to China.)

Nagao, G. (1991) *Mādhyamika and Yogācāra*, trans. L. Kawamura, Albany, NY: State University of New York Press.(Fine collection of essays by one of Japan's leading Yogācāra scholars.)

Powers, J. (1991) *The Yogācāra School of Buddhism: A Bibliography*, Metuchen, NJ: Scarecrow Press.(A fairly comprehensive bibliography, listing virtually all known Yogācāra works in Sanskrit and Tibetan and most standard Western language works. Its coverage of East Asian Yogācāra is less complete.)

Powers, J. (1995) *Wisdom of the Buddha: The Saṃdhinirmocana Mahāyāna Sūtra*, Berkeley, CA: Dharma Publishing.(Based on one of the Tibetan versions, this is the only complete English translation. It draws on Lamotte's work, but is not as extensive.)

Rahula, W. (trans.) (1971) *Le Compendium de la Super-Doctrine d'Asaōga (Abhidharmasamuccaya)*, Paris: Publications de l'École Française d'Extrême Orient.(The only Western language translation of this important Asaōga text.)

Sparham, G. (trans.) (1993) *Ocean of Eloquence: Tsong-kha-pa's Commentary on the Yogācāra Doctrine of Mind*, Albany, NY: State University of New York Press.(Translation of an interesting work by the great Tibetan Gelugpa reformer from his early days.)

Takasaki, J. (1966) *A Study of the Ratnagotravibhāga*, Rome: Serie Orientale Roma 33, Instituto Italiano per il Medio ed Estremo Oreinte.(Contains translations of the text and a critical study of its history and contents.)

Tatz, M. (1986) *Asaōga's Chapter on Ethics with the commentary of Tsong-kha-pa*, Lewiston, NY: Edwin Mellen Press.(Translation and discussion of the section on ethics from the Bodhisattvabhūmi chapter of the *Yogācārabhūmi*.)

Ui, H. (1929) 'Maitreya as an Historical Personage', in *Indian Studies in Honor of Charles Rockwell Lanman*, Cambridge, MA: Harvard University Press.(A discussion of Maitreya as an actual human teacher.)

Vasubandhu (4th or 5th century AD) *Abhidharmakośa*, ed. and trans. L. de la Vallée Poussin, *L'abhidharmakośa de Vasubandhu*, Paris/Louvain: Paul Geuthner, 1923-31, 6 vols; repr. Brussels: Institut belge des hautes études chinoises, 1971, 6 vols; trans. L.M. Pruden, Berkeley, CA: Asian Humanities Press, 1988.(La Vallée Poussin's is still the best translation available of this key work from Vasubandhu's phase as an Abhidharma scholar;

Pruden's translation, from La Vallée Poussin's French, is to be used with great caution, but is the only one available in English.)

Vasubandhu (4th or 5th century AD) *Karmasiddhiprakaraṇa*, trans. S. Anacker, 'Discussion for the Demonstration of Action', in Anacker 1984; trans. E. Lamotte, 'Le Traité de l'acte de Vasubandhu. *Karmasiddhiprakaraṇa*', *Mélanges chinois et bouddiques* 4: 151-288, 1936.(A work in which Vasubandhu criticizes Sarvāstivādin views of the mechanics of ethically charged actions and their fruitions.)

Vasubandhu (4th or 5th century AD) *Madhyāntavibhāgaśāstra*, trans. S. Anacker, 'Commentary on the Separation of the Middle from Extremes', in Anacker 1984; trans. T. Kochumuttom, 'Discrimination Between Middle and Extremes', in Kochumuttom 1982.(An important Yogācāra work.)

Vasubandhu (4th or 5th century AD) *Pañcaskandhakaprakaraṇa*, trans. S. Anacker, 'A Discussion of the Five Aggregates', in Anacker 1984.(An analysis of the classes of characteristics that constitute the physical body and the mentality of human beings.)

Vasubandhu (4th or 5th century AD) *Saṃdhinirmocanasūtra*, trans. É. Lamotte, Louvain and Paris: Université de Louvain and Adrian Maisonneuve, 1935.(Richly annotated French translation drawing on Tibetan and Chinese versions).

Vasubandhu (4th or 5th century AD) *Trisvabhāvanirdeśa*, trans. S. Anacker, 'The Teaching of the Three Own-Beings', in Anacker 1984; trans. T. Kochumuttom, 'A Treatise on the Three Natures', in Kochumuttom 1982; trans. F. Tola and C. Cragonetti, 'The *Trisvabhāvakārikā* of Vasubandhu', *Journal of Indian Philosophy* 11: 225-66, 1983.(An important work in early Yogācāra theory.)

Vasubandhu (4th or 5th century AD) *Triṃśikākārikāvṛtti*, trans. L. de la Vallée Poussin, along with the *Viṃśatikākārikāvṛtti* and a Chinese commentary by Xuanzang as *La Siddhi de Hsuan-tsang*, Paris, 1928-48, 3 vols; trans. S. Anacker, 'The Thirty Verses', in Anacker 1984; trans. T. Kochumuttom, 'A Treatise in Thirty Stanzas', in Kochumuttom 1982.(An important Yogācāra treatise.)

Xuanzang (Hsüan-tsang) (*c*.659 BC) *Cheng weishilun (Treatise Establishing Vijñaptimātra)*, TaishōShinshūDaizokyō, 1585, 1-59; 1st edn trans. L. de La Vallée Poussin, *Vijñaptimātrāsiddhi*, Paris: Librairie Orientaliste Paul Geuthner, 2 vols, 1928; 3rd edn trans. W. Tat, *Ch'eng Wei-Shih Lun: The Doctrine of Mere Consciousness*, Hong Kong, 1973; partial English translation by S. Ganguly, *Treatise in Thirty Verses on Mere-Consciousness*, Delhi: Motilal Banarsidass.(Xuanzang's seventh-century Chinese work, a commentary on Vasubandhu's *Triṃśikā* that drew on Sanskrit commentaries, became one of the standard expositions of Yogācāra doctrine in East Asia. La Vallée Poussin's rendition is loose, drawing on old Japanese scholarship and Chinese commentaries. He interprets the text idealistically. Tat's version is an English rendition of La Vallée Poussin's French text, minus the latter's extensive annotations. Ganguly's abridgement is convenient but frequently mistaken.)

Buddhist concept of emptiness

*'Emptiness' or 'voidness' is an expression used in Buddhist thought primarily to mark a distinction between the way things appear to be and the way they actually are, together with attendant attitudes which are held to be spiritually beneficial. It indicates a distinction between appearance and reality, where the paradigm for that distinction is 'x is empty (*śūnya*) of y', and emptiness (*śūnyatā*) is either the fact of x's being empty of y or the actual absence itself as a quality of x. It thus becomes an expression for the ultimate truth, the final way of things.* Śūnya *is also a term which can be used in the nontechnical contexts of, for example, 'The pot is empty of water'. These terms, however, are not univocal in Buddhist thought. If* x *is empty of* y, *what this means will depend upon what is substituted for* 'x' *and* 'y'. *In particular, any simplistic understanding of 'emptiness' as the Buddhist term for the Absolute, approached through a sort of* via negativa, *would be quite misleading. We should distinguish here perhaps four main uses of 'empty' and 'emptiness': (1) all sentient beings are empty of a Self or anything pertaining to a Self; (2) all things, no matter what, are empty of their own inherent or intrinsic existence because they are all relative to causes and conditions, a view particularly associated with Nāgārjuna and the Mādhyamika school of Buddhism; (3) the flow of nondual consciousness is empty of hypostasized subject-object duality, the Yogācāra view; (4) the Buddha-nature which is within all sentient beings is intrinsically and primevally empty of all defilements, a notion much debated in Tibetan Buddhism.*

1 Empty of Self or anything pertaining to Self

Conceptually and probably chronologically, one of the earliest uses of 'emptiness' (*śūnyatā*, also translated as 'voidness') in Buddhist thought is to refer to an absence specifically of an immutable and permanent 'Self' (*ātman*), a constant referent for the term 'I', the Self postulated by non-Buddhist teachers and presupposed at least implicitly even by 'the person in the street'. Grasping after 'me and mine' is held to generate an impetus towards continued rebirth and therefore suffering. Discovery of emptiness, the fact that really there is no such thing as a Self and therefore no coherent real basis for the destructive patterns of 'me and mine', leads to a letting-go of grasping which finally issues in *nirvāṇa*, the cessation of rebirth and suffering, a state sometimes itself spoken of as 'the empty'. This *nirvāṇa* is said to be 'seeing things the way they really are', that is, seeing emptiness, the absence of Self (see Nirvāṇa). Already, therefore, we can detect the use of 'emptiness' to designate the ultimate truth, that is, what is ultimately the case, how things really are. This analysis was clarified by the claim that all sentient beings are empty (*śūnya*) of a Self or anything pertaining to a Self, and really consist of ever-changing patterns of physical form, sensations, conceptions, other mental factors such as volitions, and the flow of consciousness. This list of five ever-changing 'aggregates' (*skandhas*) which are what is really there, contrasted with a Self which is really not there (that is, the aggregates are empty of a Self), was elaborated even further in the development of Abhidharma thought and eventually issued in a clear distinction between conventional truth (or reality, *satya*) and ultimate truth.

2 Emptiness in Vaibhāṣika Abhidharma

In the classical Vaibhāṣika (Sarvāstivāda) Abhidharma, all things (not just sentient beings) can be divided into primary existents (*dravyasat*) and secondary existents, conceptual constructs (*prajñaptisat*) (see Buddhism, Ābhidharmika schools of). Conceptual constructs - such as a forest, a table, a pot, the mind or a sentient being - are presented linguistically and cognitively, but are thought to be conventionalities and not ultimate realities since they are constructs out of primary existents (also known as *dharmas*), those data (such as sense-data) which are immediately presented and could not be further analysed away for fear of infinite regress and/or complete destruction. Most of these primary existents are the results of causes and are radically impermanent, indeed almost instantaneous; they form a series, a flow. They are the ultimate truths/realities. Thus from this Abhidharma perspective, although tables and sentient beings appear to be realities in their own right, really they are conceptual constructs out of more basic and ineliminable data into which they can be analysed. On the level of what is really there - 'really' in the sense of primary, ineliminable, unconstructed and therefore 'simple' reality, in other words, on the final, ultimate level, the 'ultimate truth' - there are no tables and sentient beings, for they are conceptual constructs and therefore only 'conventional truths'. Reality is empty of conceptually constructed and reified entities. The soteriological correlate of this is renunciation (an emptying), the letting-go of attachment to conceptual constructs, which as constructs are sure to perish when their causes cease to operate. Attachment to

things which by nature are perishable leads to suffering and continued rebirth. These conceptual constructs are empty of any ultimate significance, are ultimately valueless and are not worthy of attachment. In this letting-go lies liberation, final immutable freedom from suffering.

In this Vaibhāṣika distinction between primary and secondary existence there is an essential ontological aspect with axiological implications which is thought to be fundamental to enlightenment. It is arguable that this aspect, while implicit, is less central in some other Abhidharma traditions. The ontological distinction is marked by the concept of *svabhāva* ('own' or 'inherent' existence). While most primary existents are radically impermanent, the results of causes, they still exist in a different sort of way from conceptual constructs. Primary existents have 'own-existence', in contrast to secondary existents, which are simply constructs. Secondary existents are *niḥsvabhāva* - they lack (or are empty of) own-existence. If *x* is the result of the particular type of causation associated with conceptual construction then *x* is a secondary existent (*prajñaptisat*) and is empty of own-existence. If, on the other hand, *x* exists but is not the result of this particular sort of causation then it has own-existence; secondary existents are constructed out of these primary existents (*dharmas*) - primary existents inasmuch as they are not empty of own-existence.

We have seen already a move from the idea that all things are empty of a Self - that is, no matter where one looks, one cannot find anything which could be called one's Self (*ātman*), and therefore there is also nothing which is truly 'mine' - to an idea by no means identical that certain things are empty of own-existence. By an understandable but stretched sense of the word 'Self' to mean something like 'essence' ('self-identity') one could say that conceptual constructs are also empty of a Self, not just because I cannot find my Self in them but also because *qua* conceptual constructs they lack an essence. The problem here is that this could lead to an ambiguity in the term 'Self'. From the earliest Buddhist sources we have the statement that all things are empty of a Self. From a Vaibhāṣika point of view, however, while this is true it could not mean that all things are empty of own-existence, that is, that all things are secondary existents, conceptual constructs (*niḥsvabhāva*), for that would be to deny the distinction between primary and secondary existents and produce the absurdity that all things are constructs with nothing left for them to be constructed out of. That could only mean that nothing exists at all.

3 Madhyamaka critique and development

It is precisely in the claim that *all* things absolutely and without exception, no matter what, are empty that the concept of emptiness really comes to the fore in Buddhist thought. This claim is most frequently associated with the Perfection of Wisdom (*Prajñāpāramitā*) literature and its philosophical clarification and development in the Madhyamaka school, particularly the work of Nāgārjuna (*c.*150-200) and his commentators, such as Candrakīrti (*c.*600-50) (see Buddhism, Mādhyamika: India and Tibet). Here emptiness indeed means that for all *x*, *x* is empty of own-existence (*y*). The key to understanding this lies in the relationship between emptiness and causation. If *x* is the result of causes then *x* is not nonexistent, but also *x* does not have own-existence, that is, the existence of *x* is given to it by its causes, its existence is contingent and not intrinsic to it. *X* as contingent is arguably ontologically simply the intersection point of a set of causal forces - pure relativity. *X* does not have inherent existence, and the concept of *svabhāva* slides from that of own-existence in contrast to the sort of existence possessed by conceptual constructs such as forests and tables to that of inherent, noncontingent and therefore uncaused existence. Nothing which is the result of causes can have its existence inherent in itself. If something were inherently existent it would be permanently (that is, necessarily) existent. There is therefore a contradiction between causation and inherent existence. Thus in these terms even the Vaibhāṣika primary existents which are part of a causal flow become empty of inherent existence. While there may be a relative distinction between primary and secondary existents (clearly tables and suchlike can still be analysed into parts), this distinction, the follower of Madhyamaka wants to argue, ceases to have any ontological significance. In fact in the last analysis all things, no matter what, are secondary existents - they are all conceptual constructs.

For Madhyamaka, emptiness, that very absence - a complete and total absence of something which simply does not exist anywhere, a simple nonexistence (*abhāvamātra*) - of inherent existence in any *x* (because *x* is always in some sense the result of causal conditioning) is the ultimate truth in that if one analyses in order to find the fundamental simples (primary existents) out of which all things are constructed, nothing will ever be found. Analysis can dissolve away *everything* into secondary existence. Thus if we ask what exists ultimately, that is, what has the type of existence called ultimate or primary, it is found (it can be shown through analysis, commonly a sort of *reductio*

ad absurdum, Nāgārjuna would say) that nothing has that sort of existence. On the level of ultimate existence there is nothing. That absence of ultimate existence (that is, that emptiness) applied to everything (including emptiness itself) is the ultimate truth, what is found if one searches for any ultimate status for *x*. Therefore it follows not only that if *x* is causally conditioned it is empty (of inherent existence), but also that emptiness must be the very absence of inherent existence in something or other (an *x*). In terms of actuality, emptiness (the ultimate truth) requires causally conditioned existence (the conventional truth). In fact, emptiness is precisely what makes conditioned conventional reality what it is - conventional (and not ultimate) truth. If emptiness is the absence of *y* in *x* then it can be argued that there has to be an *x* - an *x* which turns out to lack *y* - in order for it to be empty of inherent existence. *X* therefore indeed exists, but as a noninherently existing, conventional entity, a conceptual construct. In order for causally conditioned existence to be causally conditioned existence it must be empty of inherent existence, that is, empty. Thus Nāgārjuna and Candrakīrti want to argue that holding all things to be empty (all things as secondary existents) does not imply the existence of nothing at all (that is, nihilism, *pace* the follower of Vaibhāṣika Abhidharma, who is unpersuaded and still finds something very strange in all of this). Rather, since emptiness is the result of causal conditioning emptiness must imply the existence of things in the causally conditioned way they actually are and must be (a point made strongly - for his opponents rather too strongly - in fifteenth-century Tibet by Tsong kha pa Blo bzang grags pa).

Emptiness, Nāgārjuna says, far from being nihilism and destroying all things (Buddhism included), is the very condition for their being. It is his opponent, who holds to a static world of inherent existences, who destroys all things. Any understanding of emptiness which does not see or show how emptiness implies existence as it actually is, Nāgārjuna would want to say, is not the correct understanding of emptiness. This includes any understanding of emptiness as itself an inherently existent Ultimate Reality. If emptiness is the pure absence of inherent existence in *x*, then without *x* there could be no emptiness. Thus just as *x* is contingent, causally conditioned, so in some sense must be its emptiness. Therefore emptiness too is empty of inherent existence. Madhyamaka here is emphatically not a form of *via negativa* leading to a pure, nonconceptual Absolute. Emptiness is empty, the Buddha is empty, enlightenment is empty - not in the sense that they are Realities beyond all mundane conceptualizing, but inasmuch as they result from some form of causal conditioning. Nevertheless, it does not follow that enlightenment, for example, ceases to be positively important (is empty in *that* sense). For Nāgārjuna, as a follower of Mahāyāna Buddhism, the welfare of sentient beings is supremely (one might say absolutely) important, but sentient beings and their welfare are still empty of inherent existence. It is because they and their welfare are empty that they can be helped. One could not change an inherently existent misery.

4 Emptiness of subject-object duality in Yogācāra

The claim that all things without exception are conceptual constructs, secondary existents (*prajñaptisat*, that is, *niḥsvabhāva*, empty of inherent existence) is a characteristically Madhyamaka claim. For the follower of Vaibhāṣika it is a patent contradiction which must make Madhyamaka an ontological nihilism, notwithstanding any protest to the contrary. To make the counterclaim that something must exist as a foundation for conceptual construction in order to avoid nihilism is to posit that at least one thing must have primary existence (*dravyasat*). To make this claim is to differ from Madhyamaka in ontologically the strongest possible way. It seems that we have this radical disagreement not only in Vaibhāṣika thought but also in a different way in Yogācāra. Yogācāra texts too ridicule as nihilism the claim that all things are conceptual constructs. Thus Yogācāra texts *must* postulate at least one thing as a primary existent, not a conceptual construct. In fact in Yogācāra thought that thing is the flow of nondual consciousness. It is a flow of consciousness which serves as the actual substratum for an erroneous bifurcation into the polarized (and linguistically reified) subjects and objects of unenlightened duality, and also as the necessary substratum for realizing that actually the flow is empty of hypostasized, reified and polarized subjects and objects. Thus from this perspective the ultimate, enlightening truth is also said to be emptiness, but here the ultimate truth is that something which really exists in the strongest possible sense - the flow of consciousness - is empty of projected subjects and objects, which do not really exist at all. As in Vaibhāṣika, we have here at least an implicit denial of an equation between causal conditioning and absence of own-existence. Since for Yogācāra too it would appear that all things including the flow of consciousness, are in some sense causally conditioned, the alternative to the denial of this equation is to hold (with Madhyamaka) the absurdity that all things are secondary existents, that is, putative nihilism (see Buddhism, Yogācāra school of §§1-4).

5 Intrinsic purity of the innate Buddha-nature

There is also in Buddhist thought a teaching of the Buddha-nature (*tathāgatagarbha*), which asserts the presence within each sentient being of something which enables that being to become a *buddha* ('awakened one'). Commonly that thing is spoken of as intrinsically pure, primevally present, and sometimes as primevally enlightened. The terms 'Mind' and even 'Self' (*ātman*) are sometimes used of it, but the Buddha-nature is also said to be empty. Often this is taken as meaning that the Buddha-nature is intrinsically and primevally empty of all the defilements of unenlightenment, which never really taint its nature. They are adventitious or contingent, which is why they can be removed. One can become enlightened because there is a sense in and a level at which there is no unenlightenment. Yet this Buddha-nature is also spoken of as not-empty, for it is not empty of all the attributes and qualities of enlightenment which are intrinsic (that is, necessary) to it and therefore never cease. The great question much debated in Tibet was whether in accepting the Buddha-nature, it should nevertheless be taken as empty of its own inherent existence (on the Madhyamaka model) or not. In one common Tibetan view (widespread also in East Asian Buddhism, but strongly opposed by others, particularly in Tibet) the Buddha-nature really does inherently and immutably exist in all sentient beings, primevally pure and enlightened, an emptiness (Mind) empty of all defilements, including the defilements of conceptuality. As nonconceptual, this emptiness cannot even be said to be 'emptiness', certainly not in the sense of '*x* is empty of *y*', which is just a relative, mundane emptiness. It can be approached only through nonconceptual gnosis (see Mi bskyod rdo rje). It is just possible that here we reach something akin to emptiness seen as an Absolute approached through a *via negativa*.

It is perhaps necessary to distinguish the foregoing philosophical discussions of emptiness in Indo-Tibetan thought from the use of 'emptiness' within the practical context of descriptions of meditation practice (for example, as an 'empty mind'). In Tibetan meditation it is common for the practitioner to reduce everything to emptiness and then generate out of that emptiness the forms of *buddhas*, or oneself in the form of a *buddha*, and so on. It would be interesting to explore further the relationship between the preceding conceptual discussions of Buddhist philosophy and purported 'nonconceptual' mental states, 'blank-minds', and so on.

See also: Buddhist philosophy, Indian; Buddhist philosophy, Chinese §12; Causation, Indian theories of §6; Error and illusion, Indian conceptions of §1; Fazang; Hasidism §2; Linji; Nominalism, Buddhist doctrine of

PAUL WILLIAMS

References and further reading

Buddhadāsa Bhikkhu (1994) *Heartwood of the Bodhi Tree: The Buddha's Teaching on Voidness*, trans. Dhammavicayo, ed. Santikaro Bhikkhu, Boston, MA: Wisdom Publications.(Very clear oral teachings by an eminent Thai Theravāda meditation master on 'empty of Self and pertaining to Self'.)

Cabezón, J.I. (1994) *Buddhism and Language: A Study of Indo-Tibetan Scholasticism*, Albany, NY: State University of New York Press.(An excellent and not too technical introduction to Madhyamaka and language within the context of Tibetan scholasticism.)

Hookham, S.K. (1990) *The Buddha Within: Tathāgatagarbha Doctrine According to the Shen-tong Interpretation of the Ratnagotravibhāga*, Albany, NY: State University of New York Press.(A detailed, sympathetic exposition of the teachings of an inherently existent Buddha-nature and their relationship to other Buddhist traditions.)

Nagao, G. (1991) *Mādhyamika and Yogācāra: A Study of Mahāyāna Philosophies*, ed. and trans. L.S. Kawamura, Albany, NY: State University of New York Press, ch. 5. (An important paper on the Yogācāra approach to emptiness in relation to that of Madhyamaka.)

Napper, E. (1989) *Dependent Arising and Emptiness*, Boston, MA: Wisdom Publications.(Probably the best introduction to Tsong kha pa's influential approach to emptiness. Some material, particularly the translations, is technical and difficult, but the earlier chapters, which include a critique of contemporary approaches in the light of Tsong kha pa are very accessible and contain all the important points.)

Ruegg, D.S. (1981) *The Literature of the Madhyamaka School of Philosophy in India*, vol. 7, fasc. 1 of J. Gonda (ed.) *A History of Indian Literature*, Wiesbaden: Otto Harrassowitz.(Rather like a detailed annotated bibliography at times, but a standard work on Indian Madhyamaka by its leading authority. Highly recommended, with many insights.)

Streng, F. (1967) *Emptiness: A Study in Religious Meaning*, Nashville, TN and New York: Abingdon Press. (Dated but extremely readable introduction to Madhyamaka, with some translated texts. Places Madhyamaka within the context of Religious Studies and the history of Buddhism. An excellent bibliography up to its date of

publication.)
Williams, P. (1981) 'On the Abhidharma Ontology', *Journal of Indian Philosophy* 9: 227-57.(Very technical study of Vaibhāṣika Abhidharma, particularly the *svabhāva* and primary/secondary distinction of existents.)
Williams, P. (1989) *Mahāyāna Buddhism: The Doctrinal Foundations*, London: Routledge, chaps 2-5.(Accessible introductions to Perfection of Wisdom, Madhyamaka, Yogācāra and Buddha-nature teachings within the context of Mahāyāna Buddhism.)
Williams, P. (1992) 'Non-Conceptuality, Critical Reasoning and Religious Experience: Some Tibetan Buddhist discussions', in M. McGhee (ed.) *Philosophy, Religion and the Spiritual Life*, Cambridge: Cambridge University Press.(A paper aimed at philosophers with no background in Buddhist Studies, introducing Tibetan Madhyamaka and the role of reasoning within it, with a discussion of some problems concerning nonconceptual states and 'empty minds'.)

Buddhist philosophy, Chinese

When Buddhism first entered China from India and Central Asia two thousand years ago, Chinese favourably disposed towards it tended to view it as a part or companion school of the native Chinese Huang-Lao Daoist tradition, a form of Daoism rooted in texts and practices attributed to Huangdi (the Yellow Emperor) and Laozi. Others, less accepting of this 'foreign' incursion from the 'barbarous' Western Countries, viewed Buddhism as an exotic and dangerous challenge to the social and ethical Chinese civil order. For several centuries, these two attitudes formed the crucible within which the Chinese understanding of Buddhism was fashioned, even as more and more missionaries arrived (predominantly from Central Asia) bringing additional texts, concepts, rituals, meditative disciplines and other practices. Buddhists and Daoists borrowed ideas, terminology, disciplines, cosmologies, institutional structures, literary genres and soteric models from each other, sometimes so profusely that today it can be difficult if not impossible at times to determine who was first to introduce a certain idea. Simultaneously, polemical and political attacks from hostile Chinese quarters forced Buddhists to respond with apologia and ultimately reshape Buddhism into something the Chinese would find not only inoffensive, but attractive.

In the fifth century AD, Buddhism began to extricate itself from its quasi-Daoist pigeonhole by clarifying definitive differences between Buddhist and Daoist thought, shedding Daoist vocabulary and literary styles while developing new distinctively Buddhist terminology and genres. Curiously, despite the fact that Mahāyāna *Buddhism had few adherents in Central Asia and was outnumbered by other Buddhist schools in India as well, in China* Mahāyāna *became the dominant form of Buddhism, so much so that few pejoratives were as stinging to a fellow Buddhist as labelling him '*Hīnayāna*' (literally 'Little Vehicle,' a polemical term for non-*Mahāyānic *forms of Buddhism). By the sixth century, the Chinese had been introduced to a vast array of Buddhist theories and practices representing a wide range of Indian Buddhist schools. As the Chinese struggled to master these doctrines it became evident that, despite the fact that these schools were all supposed to express the One Dharma (Buddha's Teaching), their teachings were not homogenous, and were frequently incommensurate.*

By the end of the sixth century, the most pressing issue facing Chinese Buddhists was how to harmonize the disparities between the various teachings. Responses to this issue produced the Sinitic Mahāyāna *schools, that is, Buddhist schools that originated in China rather than India. The four Sinitic schools are Tiantai, Huayan, Chan and Pure Land (Jingtu). Issues these schools share in common include Buddha-nature, mind, emptiness,* tathāgatagarbha*, expedient means (*upāya*), overcoming birth and death (*saṃsāra*), and enlightenment.*

1 Historical overview

The development of Chinese Buddhist philosophy can be divided roughly into four periods: (1) the early introduction of Indian and Central Asian Buddhism (first-fourth centuries AD); (2) the formative development of Chinese versions of Indian and Central Asian Schools (fifth-seventh centuries); (3) the emergence of distinctively Sinitic Buddhist schools (seventh-twelfth centuries); and (4) the continuance of Chinese Buddhism into the present day (thirteenth century onwards).

From the fourth through the seventh centuries, Buddhists periodically realized that the positions being engendered in China were at variance with their Indian antecedents, and attempted to correct the problem, either through the introduction of additional translations or by clarifying differences between Buddhist and native Chinese ideas. By the eighth century, the Chinese had apparently become satisfied with the types of Buddhism they had developed, since from then on they lost interest in Indian commentaries and treatises and instead turned their attention toward Chinese commentaries on the Buddhist scriptures - such as the *Lotus Sutra* and *Huayan Sutra* - that had assumed importance for Chinese Buddhist traditions. Moreover, even though missionaries continued to arrive in China and new translations continued to be produced through the thirteenth century, none of the significant developments in Indian Buddhism (such as Buddhist syllogistic logic) from the seventh century onwards had any lasting impact on Chinese Buddhism, and many important texts and thinkers (for example, Dharmakīrti, Candrakīrti, Śāntarakṣita) remained virtually unknown in East Asia until modern times (see Buddhist philosophy, Indian).

2 Earliest developments

The first undisputed reference to Buddhism in China is an edict by Emperor Ming to Liu Ying, king of Qu, in the year AD 65, which mentions sacrifices performed by the king to Buddha as well as favourable treatment for Buddhist monks and laymen; the edict also identifies King Liu Ying as a follower of Huang-Lao Daoism. For the next few centuries, the Chinese continued to view Buddhist texts and practices as a part of or supplement to Daoism. Buddhism seemed to share important issues with the types of Daoism practised in this period, including the metaphysical primacy of emptiness, meditation techniques, dietary and behavioural disciplines, afterlife theories connected to moral and behavioural discipline, expansive pantheons and cosmologies, striving for the soteriological transformation of the ordinary human condition, rigorous and subtle intellectual traditions, and magical and yogic powers. These affinities, however, were more apparent than real, since the Buddhist approaches to these issues usually differed sharply from their Daoist counterparts, though Buddhists did not assert their distinctiveness until the fifth century.

Buddhist monks and businessmen representing a variety of Buddhist schools and disciplines continued to arrive in China, establishing Buddhist communities in Loyang and elsewhere. It took several centuries for the Chinese to notice how disparate the various forms of Buddhism really were. Initially the Chinese were most interested in Buddhist meditation techniques, including chanting and visualizations, which they adopted as supplements to Daoistic techniques. Daoism held out the promise that one could become a sage or perfected person, or even an immortal, but the exact details of how to accomplish this transformation remained elusive and vague. In comparison to many of the Daoist texts which were esoteric, hard to find, and frequently obscure in presentation, Buddhist texts seemed systematic and detailed, providing step-by-step procedures for practitioners.

Along with meditation manuals, the earliest Buddhist texts to become popular in China were Āvadana materials (legends of the Buddha and Buddhist heroes) and the Perfection of Wisdom Scriptures (*Prajñāpāramitā Sutras*). About half a dozen schools formed around varying interpretations of the Perfection of Wisdom Scriptures, mixing ideas found in these and other Buddhist texts with concepts prominent among Chinese intelligentsia. One Prajñāschool, called the Original Nothingness school (*Benwu*), adopted a neo-Daoist cosmology: everything has emerged from a primordial, original emptiness, and everything returns to that void. This was a thorough misconstrual of Buddhist emptiness (see Buddhist concept of emptiness). Another school, called the Mind Empty school (*Xinwu*), equated the primordial Nothing with the nature of mind (see Zhi Dun). Each of the Prajñāschools managed either to promote a metaphysical substantialized emptiness which they opposed to form, or smuggle an eternal self or spirit into their formulations, despite Buddhism's emphatic rejection of the notion of permanent selfhood.

Dao'an (AD 312-85) criticized the Prajñāschools, challenging their faithfulness to authentic Buddhist positions as well as the translation methodologies behind the texts they and other Chinese Buddhists had come to rely on. In particular, he criticized the practice of 'matching the meanings' (*geyi*), by which translators seeking Chinese equivalents for Indian Buddhist technical terms and concepts borrowed heavily from Daoist literature. This 'matching of meanings' was a mixed blessing. Packaging Buddhist ideas in familiar terms made them amenable and understandable, but the 'matches' were often less than perfect, distorting or misrepresenting Buddhism. For instance, early translators chose a well-known Daoist and Confucian term, *wuwei* (nondeliberative activity), to translate *nirvāṇa*. Arguably, *wuwei* and *nirvāṇa* represent the teloi of Daoism and Buddhism, respectively, but it is not obvious that they denote the same telos (see Daoist philosophy §6; Nirvāṇa). Later, to emphasize the uniqueness of Buddhist *nirvāṇa*, translators dropped *wuwei* in favour of a transliteration, *niepan*. *Wuwei* was retained to render another important Buddhist notion, *asaṃskṛta* (unconditioned). The semantic connotations of Daoist *wuwei* and Buddhist *wuwei*, while possibly overlapping in some senses, were nonetheless quite distinct: for Daoists it meant a mode of interacting effortlessly and naturally with the world, while for Buddhists it denoted something unaffected by causes and conditions, that neither arose nor ceased. Chinese readers inevitably came to conflate the semantic ranges of such terms, which over the centuries led to some distinctively Chinese Buddhist concepts. After Dao'an, Chinese Buddhism asserted its distinctiveness from native Chinese traditions and Buddhists adopted increasingly critical hermeneutic approaches to translation.

3 Indian transplants: Madhyamaka and *icchantikas*

In the critical environment that followed Dao'an, two sets of events moved Chinese Buddhism in new directions. First, Kumārajīva, a Mahāyāna Buddhist from Kucha in Central Asia, was brought to Changan, the Chinese

capital, in 401. Under the auspices of the ruler, he began translating numerous important works with the help of hundreds of assistants, including some of the brightest minds of his day. Some works, such as the *Lotus Sutra*, *Vimalakīrti Sutra* and *Diamond Sutra*, quickly became popular classics. He also introduced the emptiness philosophy of Nāgārjuna's Madhyamaka thought (see Buddhism, Mādhyamika: India and Tibet; Nāgārjuna), which in China came to be called the Three Treatise School (Sanlun) after the three Madhyamaka texts he translated: the *Madhyamaka-kārikās*, the *Twelve Gate Treatise* and Āryadeva's *One Hundred Verse Treatise*. In a series of famous letters exchanged with a disciple of Dao'an, Huiyuan (344-416), who had mastered most of the Buddhist theory and practice known in China up to that time, Kumārajīva attacked the shortcomings of the current Chinese Buddhist theories and argued persuasively for the preeminence of Madhyamaka in matters of both theory and practice. His leading disciple, Seng Zhao (384-414), further popularized Madhyamaka thought by packaging it in an exquisite adoption of the literary style of Laozi (see Daodejing) and Zhuangzi, both of whom were extremely popular amongst literati at that time. Sanlun thought continued to spread through the fifth through seventh centuries, greatly influencing other Buddhist schools. After Jizang (549-623), who attempted to synthesize Madhyamakan emptiness with the Buddha-nature and *tathāgatagarbha* thought gaining prominence at his time, the Sanlun school declined, its most important ideas absorbed by other schools.

Second, in 418 Faxian (the first Chinese monk successfully to return to China with scriptures from pilgrimage to India) and Buddhabhadra produced a partial translation of the Mahāyāna *Nirvāṇa Sutra*. One of the topics it discusses is the *icchantika*, incorrigible beings lacking the requisites for achieving enlightenment. Daosheng (*c*.360-434), a disciple of Huiyuan, convinced that all beings, including *icchantikas*, must possess Buddha-nature and hence are capable of enlightenment, insisted that the *Nirvāṇa Sutra* be understood in that light. Since that violated the obvious meaning of the text, Daosheng was unanimously rebuked, whereupon he left the capital in disgrace. In AD 421, a new translation by Dharmakṣema of the *Nirvāṇa Sutra* based on a Central Asian original appeared containing sections absent from the previous version. The twenty-third chapter of Dharmakṣema's version contained passages declaring that Buddha-nature was indeed universal, and that even *icchantikas* possessed it and could thus reach the goal. Daosheng's detractors in the capital were humbled, suddenly impressed at his prescience. The lesson was never forgotten, so that two centuries later, when Xuan Zang (600-64) translated Indian texts that once again declared that *icchantikas* lacked the requisite qualities to attain enlightenment, his school was attacked from all quarters as promoting a less than 'Mahāyānic' doctrine. However, it should be noted that there is no clear precedent or term in Indian Buddhism for 'Buddha-nature'; the notion probably either arose in China through a certain degree of license taken by translators when rendering terms like *buddhatva* ('Buddhahood', an accomplishment, not a primordial ontological ground), or it developed from nascent forms of the theory possibly constructed in Central Asia. However, from this moment on, Buddha-nature become one of the foundational tenets of virtually all forms of East Asian Buddhism.

4 Indian transplants: *tathāgatagarbha* and Yogācāra

A dispute at the start of the sixth century presaged a conflict that would take the Chinese Buddhists more than two centuries to settle. Two Indian monks collaborated on a translation of Vasubandhu's *Daśabhūmikasūutra śāstra (Treatise on the Ten Stages Sutra; in Chinese, Shidijing lun, or Dilun for short)*. The *Dilun* described the ten stages through which a *bodhisattva* proceeded on the way to *nirvāṇa*, and Vasubandhu's exposition of it highlighted aspects most in accord with the tenets of the Yogācāra school (see Buddhism, Yogācāra school of; Vasubandhu). While translating, an irreconcilable difference of interpretation broke out between the two translators, Bodhiruci and Ratnamati. Bodhiruci's reading followed a relatively orthodox Yogācāra line, while Ratnamati's interpretation leaned heavily toward a Buddhist ideology only beginning to receive attention in China, *tathāgatagarbha* thought. Bodhiruci went on to translate roughly forty additional texts, and was later embraced by both the Huayan and Pure Land traditions as one of their early influences (see §§8, 10). Ratnamati later collaborated with several other translators on a number of other texts. Both sides attempted to ground their positions on interpretations of key texts, especially the *Dilun*. The Yogācāra versus Yogācāra-*tathāgatagarbha* conflict became one of the critical debates amongst sixth and seventh century Chinese Buddhists.

Yogācāra focused on the mind and distinguished eight types of consciousness: five sensory consciousnesses; an empirical organizer of sensory data (*mano-vijñāna*); a self-absorbed, appropriative consciousness (*manas*); and the eighth, a warehouse consciousness (*ālaya-vijñāna*) that retained the karmic impressions of past experiences and coloured new experiences on the basis of that previous conditioning. The eighth consciousness was also the

fundamental consciousness. Each individual is constituted by the karmic stream of one's own *ālaya-vijñāna*, that is, one's karmic conditioning. Since, like a stream, the *ālaya-vijñāna* is reconfigured each moment in response to constantly changing conditions, it is not a permanent self, although, being nothing more than a sequential chain of causes and effects, it provides sufficient stability for an individual to maintain a sense of continuity. According to classical Yogācāra texts, the mind (that is, *ālaya-vijñāna* and the mental events associated with it) is the problem, and enlightenment results from bringing this consciousness to an end, replacing it with the Great Mirror Cognition (*ādarśa-jñāna*); instead of discriminating consciousness, one has direct immediate cognition of things just as they are, as impartially and comprehensively as a mirror. This type of enlightenment occurs during the eighth stage according to the *Dilun* and other texts.

The term *tathāgatagarbha* (in Chinese, *rulaizang*) derives from two words: *tathāgata* (Chinese, *rulai*) is an epithet of the Buddha, meaning either 'thus come' or 'thus gone'; *garbha* means embryo, womb or matrix, and was translated into Chinese as *zang*, meaning 'repository'. In its earliest appearances in Buddhist texts, *tathāgatagarbha* (repository of buddhahood) signified the inherent capacity of humans (and sometimes other sentient beings) to achieve buddhahood. Over time the concept expanded and came to signify the original pristine pure ontological Buddha-ness intrinsic in all things, a pure nature that is obscured or covered over by defilements (Sanskrit, *kleśa*; Chinese *fannao*), that is, mental, cognitive, psychological, moral and emotional obstructions. It was treated as a synonym for Buddha-nature, though Buddha-nature dynamically understood as engaged in a struggle against defilements and impurities. In Chinese Buddhism especially, the soteriological goal consisted in a return to or recovering of that original nature by overcoming or eliminating the defilements. The battle between the pure and impure, light and dark, enlightenment and ignorance, good and evil and so on, took on such epic proportions in Chinese Buddhist literature that some scholars have compared it to Zoroastrian or Manichean themes, though evidence for the influence of those religions on Buddhist thought has been more suggestive than definitive (see Manicheism; Zoroastrianism).

In their classical formulations the *ālaya-vijñāna* and *tathāgatagarbha* were distinct items differing from each other in important ways - for instance, enlightenment entailed bringing the *ālaya-vijñāna* to an end, while it meant actualizing the *tathāgatagarbha*; the *ālaya-vijñāna* functioned as the karmic mechanism *par excellence*, while *tathāgatagarbha* was considered the antipode to all karmic defilements. Nonetheless some Buddhist texts, such as the *Laōkāvatāra Sūtra*, conflated the two. Those identifying the two argued that the *ālaya-vijñāna*, like *tathāgatagarbha*, was pure and its purity became permanently established after enlightenment. Those opposing the conflation countered that the *ālaya-vijñāna* was itself defiled and needed to be eliminated in order to reach enlightenment. For the conflators, *tathāgatagarbha* was identified with Buddha-nature and with mind (*xin*) (see Xin). Mind was considered pure, eternal, and the ontological ground of reality (*Dharma-dhātu*), while defiled thought-instants (*nian*) that engaged in delusionary false discriminations had to be eliminated. Once *nian* were eliminated, the true, pure nature of the mind would brilliantly shine forth, like the sun coming out from behind the clouds.

A third view was added when Paramārtha, another Indian translator with his own unique interpretation of Yogācāra, arrived in the middle of the sixth century. For his followers the most important of his translations was the *She dasheng lun* (Sanskrit title, *Mahāyānasaṃgraha*), or *Shelun*, a quasi-systematic exposition of Yogācāra theory by one its founders, Asaōga. In some of his translations he added a ninth consciousness beyond the usual eight, a 'pure consciousness' that would pervade unhindered once the defiled *ālaya-vijñāna* was destroyed. His translations, which sometimes took liberties with the Sanskrit originals, offered a more sophisticated version of the conflation theory.

5 The *Awakening of Faith in Mahāyāna*

These debates and their ramifications dominated Chinese Buddhist thought in the sixth century. On one side was a substantialistic nondual metaphysic whose eternalistic ground was variously called Buddha-nature, mind, *tathāgatagarbha*, *Dharma-dhātu* and suchness (*tathatā*; in Chinese, *rulai*). On the other side was an anti-substantialist critique that eschewed any form of metaphysical reification, emphasizing emptiness as the absence of permanent selfhood or independent essence in anything. To the anti-substantialists the *tathāgatagarbha* position sounded dangerously close to the notion of eternalistic, reified selfhood that Buddha had rejected. Mahāyāna texts had declared that there were four conceptual perversions or reversals behind human

delusion: (1) seeing a self in what lacks self; (2) seeing permanence in the impermanent; (3) seeing happiness in what is suffering; and (4) seeing purity in the impure. Yet starting with the earliest *tathāgatagarbha* texts - such as *The Lion's Roar of Queen Śrīmāla* - *tathāgatagarbha* was brazenly defined as 'self, eternal, happiness and pure'. In the face of these and other disparities, the Chinese asked how, if there is only one *dharma* (teaching), there can be such incommensurate variety.

The *Awakening of Faith in Mahāyāna*, a Chinese composition purporting to be a translation by Paramārtha of an Indian text, became an instant classic by offering a masterly synthesis of Buddhist teachings that seemed to resolve many of the disparities (see Awakening of Faith in Mahāyāna). Its central tenet is that there is one Mind that has two aspects. One aspect is suchness and the other is *saṃsāra*, the cycle of birth and death, arising and ceasing. Suchness also has two aspects, emptiness and non-empty. Emptiness in this text means suchness is beyond predication, neither one nor many, neither the same nor different. Non-empty means it is endowed with all the marvellous qualities and merits of a Buddha, 'as numerous as the sands along the banks of the Ganges'. The link between suchness and the realm of arising and ceasing is *tathāgatagarbha* in association with the *ālaya-vijñāna*. Ignorance, enlightenment and pursuit of the Path are all on the arising and ceasing side.

In a pivotal passage that would become foundational for most forms of Chinese, Korean and Japanese Buddhism, the *Awakening of Faith in Mahāyāna* states that on the basis of Original Enlightenment there is non-enlightenment; on the basis of non-enlightenment there is initial enlightenment; and on the basis of initial enlightenment there is final enlightenment, which is the full realization of original enlightenment. Beyond the problem of theodicy that it raises (the text does not offer a clear explanation for why or how non-enlightenment arises), several issues emerge. First, suddenly there is no longer simply one enlightenment that is achieved at the culmination of a spiritual path, but instead several enlightenments, one of which (original enlightenment) precedes even entering the path. What the text calls initial enlightenment had been termed *bodhicitta* or *cittotpāda* (arousing the aspiration for enlightenment) in previous Buddhist literature. Arousing this aspiration is what the title *Awakening of Faith in Mahāyāna* signifies. Now, rather than marking a singular, ultimate achievement, the term 'enlightenment' referred to several things: an atemporal originary ground upon which everything else plays out, including non-enlightenment; one's initial resolve or insight that leads one to begin pursuing the path; the final achievement at the end of the path, an achievement that is not only anticlimactic, but is little more than an unravelling of the intersection of original and initial enlightenment. This reinforced the conviction of Chinese Buddhists that the conflationist approach, with its emphasis on Buddha-nature or mind as ground, was the correct view.

One of the first to recognize the importance of the *Awakening of Faith in Mahāyāna* was a Korean monk named Wônhyo. He wrote a commentary on the text that reached China, where it influenced Fazang, a foundational thinker of the Huayan school, who used its ideas as a major cornerstone for his thinking. Since the *Dilun*, originally an independent text, had eventually been incorporated into the *Huayan Sutra* as one of its chapters, and that scripture became the basic text of the Huayan school, many of the issues that had emerged from the debates on the *Dilun* were absorbed and reconfigured by the Huayan thinkers. In a sense, it was ultimately Ratnamati's interpretation that prevailed after two centuries of debate. The *Awakening of Faith in Mahāyāna* became pivotal for Chinese Buddhism, is still one of the foundations of Korean Buddhism and, though it has been eclipsed in Japan by other texts such as the *Lotus Sutra*, many of its ideas, such as the idea of original enlightenment (*hongaku*), still exert a profound influence. This text set the stage for the development of distinctively East Asian forms of Buddhism.

6 The Chinese Buddhist Schools

Although the ideas and literature of many different forms of Buddhism reached China - including Sarvāstivāda, Mahīśāsika, Saṃmitīya, Dharmaguptaka, Sautrāntika and others - only Madhyamaka and Yogācāra developed Chinese schools and lineages. Madhyamaka disappeared as an independent school after Jizang, but its influence and the preeminence of Nāgārjuna never abated. The sixth century was basically a battleground of competing Yogācāric theories.

In the seventh century the famous Chinese pilgrim Xuanzang (600-64) spent sixteen years travelling in Central Asia and India. He returned to China in 645 and translated seventy-four works. Due in part to his accomplishments as a traveller and translator, and in part to the eminent favour bestowed on him by the Chinese emperor upon his

return, Xuanzang became the most prominent East Asian Buddhist of his generation. He promoted an orthodox form of Yogācāra as it was then being practised in India, and students flocked to him from Japan and Korea as well as China. Not everyone was enamoured of the Buddhist ideology he had brought back. Zhiyan (602-68), who would later be considered one of the patriarchs of the Huayan school, was openly critical of Xuanzang's teachings, and Fazang had joined Xuanzang's translation committee late in Xuanzang's life, only to quit in disgust at Xuanzang's 'distorted' views. While in India, Xuanzang had discovered how far Chinese Buddhism had deviated from its Indian source, and his translations and teachings were deliberate attempts to bring Chinese Buddhism back in line with Indian teachings. The ideas he opposed (primarily but not exclusively those that had been promoted by Paramārtha's school) were already deeply entrenched in Chinese Buddhist thinking. While he was alive his pre-eminence made him unassailable, but once he died his detractors attacked his successor, Kuiji (632-82), and successfully returned Chinese and East Asian Buddhism to the trajectory established by the conflationists. (Wônch'ūk, a Korean student of Xuanzang, was a rival of Kuiji who fared better with the revivalists since he attempted to harmonize the teachings of Paramārtha and Xuanzang.) The underlying ideology of this resurgence, which reached its intellectual apex over the course of the Tang and Song Dynasties (sixth-twelfth centuries), was neatly summarized by the label '*dharma*-nature' (*faxing*), that is, the metaphysical ground of Buddha-nature *qua dharma*-*dhātu qua* mind-nature *qua tathāgatagarbha*. Fazang argued that orthodox Yogācāra only understood *dharma* characteristics (*faxiang*), that is, phenomenal appearances, but not the deeper underlying metaphysical reality, '*dharma*-nature'. After Fazang, all the Sinitic Buddhist schools considered themselves *dharma*-nature schools; Yogācāra and sometimes Sanlun were considered merely *dharma* characteristics schools.

Four *dharma*-nature schools emerged. Each school eventually compiled a list of its patriarchs through whom its teachings were believed to have been transmitted. Modern scholarship in Japan and the West has shown that these lineages were usually forged long after the fact, and frequently were erroneous or distorted the actual historical events. For instance, while Huiyuan was an active promoter of the Sarvāstivādin teachings introduced during his time by Sanghadeva and Buddhabhadra, the later Pure Land schools dubbed him their initial Chinese patriarch on the basis of his alleged participation in Amitābha rituals, allegations that were probably first concocted during the Tang Dynasty. Similarly, the lineage of six Chan patriarchs from Bodhidharma to Huineng is unlikely; the Huayan lineage (Du Shun to Zhiyan to Fazang to Chengguan) was largely an invention of the 'fourth' patriarch, Chengguan: his predecessors were unaware that they were starting a new lineage and rather thought that they were reviving the true old-time religion of Paramārtha. It was also during the Tang dynasty that Tantra briefly passed through China, from whence it was brought to Japan and became firmly established as the Shingon school.

7 The Chinese Buddhist Schools: Tiantai

Though considered its third patriarch, the intellectual founder of the Tiantai school was Zhiyi (538-97). Responding to the proliferation of different Buddhist theories and practices, he proposed a masterly, detailed synthesis that definitively set Chinese Buddhism in its own direction. To the question of why there was an abundance of incommensurate teachings despite the fact that there could only be one *dharma*, Zhiyi replied that all the different vehicles of Buddhism were ultimately one vehicle (*eka*-*yāna*), an idea championed by the *Lotus Sutra*. More specifically, he offered a *panjiao*, or classificatory scheme of teachings, to explain the discrepancies. His *panjiao* was complex and brilliant (and further refined much later by Chegwan, a Korean Tiantai monk in China), but in simple form it can be summarized as follows.

Buddha offered different teachings to different audiences based on the differing capacities of audiences to comprehend what he preached. According to the basic narrative, which became the way all Chinese Buddhists thought of Buddha's teaching career, upon reaching enlightenment under the Bodhi Tree, Buddha, enraptured by his new vision, began to describe that vision in immediate and exuberant terms. This became the *Huayan Sutra*. (In reality, this 'sutra' is a collection of disparate texts - none probably composed earlier than the third century AD - that were gradually compiled together over several more centuries.) When he finished (it took two or three weeks) he realized that no one had understood the sublime meaning of his words, and immediately began to teach a simplified, preparatory teaching which became the Hīnayāna teachings. After twenty years of preparatory teachings, he introduced the next level, beginners' Mahāyāna (basically Yogācāra and Madhyamaka). In the next period he introduced advanced Mahāyāna (the *Vaipūlya Sutras*), and finally in his last days, having now trained many advanced students, he preached the *Lotus Sutra* and the *Nirvāṇa Sutra*. In effect this *panjiao* asserts that the two highest sutras offered by Buddha were the *Huayan* and *Lotus*; but whereas the *Huayan* was too sublime to be

understood by any save the most advanced or enlightened students, the *Lotus* represented Buddha's most comprehensive, cumulative, mature and accessible teaching, every bit as sublime as the *Huayan*, but now presented in a pedagogically effective manner. For that reason, Zhiyi made the *Lotus Sutra* the foundational text of Tiantai. As for the remaining teachings, as the *Lotus* itself explains, different 'truths' can be superseded once they have served their task of raising one to a higher level where a different 'truth' holds sway. Buddhism, according to the *Lotus* and Tiantai, is a system of expedient means (*upāya*) leading one with partial truths to ever greater, more comprehensive truths. Tiantai teachings are 'Round Teachings', meaning that they encircle or encompass everything and, lacking sharp edges, are therefore Perfect. Other forms of Buddhism are not 'wrong', but are only partial visions of the One Vehicle that Tiantai most perfectly and completely embodies.

Zhiyi, based on an exhaustive exposition of a verse from Nāgārjuna's *Madhyamaka-kārikās* (24: 18), devised a theory of three truths: provisional, empty and middle. The first two are mirror images of each other, two ways of speaking about causes and conditions. A table can provisionally be called a table, since its perceptible form has arisen through causes and conditions, and it only exists provisionally on the basis of those temporary conditions. The table is empty because, being the product of causes and conditions, it lacks its own intrinsic, independent nature. It is 'middle' because neither the provisional nor the empty truth about the table fully captures its reality. It is both provisional and empty, and simultaneously neither provisional nor empty. As Zhiyi put it, 'wondrous being is identical to true emptiness'. Zhiyi sought many ways to express the nondual middle truth. For instance, rejecting the obvious dualism of the distinction most of his contemporaries made between pure mind (*xin*) and deluded thought-instants (*nian*), Zhiyi declared that every deluded thought-instant was identical to three thousand chilicosms. The details of the formulas he used to arrive at the number three thousand is less important than fact that it is meant to encompass the full extent of Buddhist cosmological metaphysics. The whole universe in all its dimensions is entailed in every moment of thought. Rather than attempt to eliminate deluded thinking to reach a purified mind, Zhiyi claimed each moment of deluded thinking was already identical to enlightenment. One merely has to see the mind and its operations as they are. This idea was later taken over by the Chan (Zen) school, which expressed it in sayings such as 'Zen mind is everyday mind'.

The middle approach is also evident in the Tiantai notion of three gates, or three methods of access to enlightened vision: the Buddha-gate, the gate of sentient beings and the mind-gate. The Buddha-gate was considered too difficult, too abstruse, too remote; one had to be a Buddha already to fully comprehend it. The sentient-being gate (the various methods taught and practised by any sort of being) was also too difficult because there are too many different types of sentient beings all with their own types of delusions, so that this gate is a confusing cacophony of disparate methods, some which may not be appropriate for some beings. The easiest and hence preferable gate was the mind-gate. It is no more remote than this very moment of cognition, its diversity can be observed in every thought-instant, and nothing could ever be more appropriately suited for an individual than to observe one's own mind. Tiantai cultivated many types of meditation for that purpose.

8 The Chinese Buddhist Schools: Huayan

Drawing on a *panjiao* similar to that of Zhiyi, the Huayan school chose the *Huayan Sutra* (Sanskrit title *Avataṃsaka Sutra*, Chinese *Huayan jing*) for its foundational scripture. What immediately differentiates Huayan from typically Indian approaches is that instead of concentrating on a diagnosis of the human problem, and exhorting and prescribing solutions for it, Huayan immediately begins from the point of view of enlightenment. In other words, its discourse represents a nirvanic perspective rather than a samsaric perspective. Instead of detailing the steps that would lead one from ignorance to enlightenment, Huayan immediately endeavours to describe how everything looks through enlightened eyes.

Like Tiantai, Huayan offers a totalistic, encompassing 'round' view. A lived world as constituted through a form of life experience is called a *dharma-dhātu*. Chengguan, the 'fourth' Huayan patriarch, described four types of *dharma-dhātus*, each successively encompassing its predecessors. The first is *shi*, which means 'event', 'affair' or 'thing'. This is the realm where things are experienced as discrete individual items. The second is called *li* (principle), which in Chinese usage usually implies the principal metaphysical order that subtends events as well as the rational principles that explicate that order. Often *li* is used by Buddhists as a synonym for emptiness. The first sustained analysis based on the relation of *li* and *shi* was undertaken by the Korean monk Wônhyo in his commentary on the *Awakening of Faith in Mahāyāna*, which influenced early Huayan thinkers like Fazang. The

li-shi model went on to become an important analytic tool for all sorts of East Asian philosophers, not just Buddhists. In the realm of *li*, one clearly sees the principles that relate *shi* to each other, but the principles are more important than the individual events. In the third realm, one sees the mutual interpenetration or 'non-obstruction' of *li* and *shi* (*lishi wu'ai*). Rather than seeing events while being oblivious to principle, or concentrating on principle while ignoring events, in this realm events are seen as instantiations of principle, and principle is nothing more than the order by which events relate to each other.

In the fourth and culminating *dharma-dhātu*, one sees the mutual interpenetration and non-obstruction of all events (*shishi wu'ai*). In this realm, everything is causally related to everything else. Huayan illustrates this with the image of Indra's net, a vast net that encompasses the universe. A special jewel is found at the intersection of every horizontal and vertical weave in the net, special because each jewel reflects every other jewel in the net, so that looking into any one jewel, one sees them all. Every event or thing can disclose the whole universe because all mutually interpenetrate each other without barriers or obstruction.

This form of nondualism is not monistic because *shishi wu'ai* does not obliterate the distinctions between things, but rather insists that everything is connected to everything else *without* losing distinctiveness. Identity and difference, in this view, are merely two sides of the same coin, which, though a single coin, still has two distinct sides that should not be confused for each other. Mutual interpenetration is temporal as well as spatial; past, present and future mutually interpenetrate. Hence according to Huayan, to enter the path towards final enlightenment is, in an important sense, to have already arrived at that destination.

9 The Chinese Buddhist Schools: Chan

Better known in the West by its Japanese pronunciation, Zen, Chan emerged as a reaction against the increasing scholastic complexities of the Tiantai and Huayan schools and their voluminous, hairsplitting literature, which, some Chan practitioners believed, could be more of an obstacle than an aid to enlightenment. The Pāli term for meditative absorption, *jhāna* (Sanskrit, *dhyāna*), was transliterated into Chinese as *Channa*, and then shortened to *Chan*. Until the early Tang Dynasty, *chanshi* (Chan master) meant a monk adept at meditation, though it did not specify what sorts of meditation he was practising. Some monks were called *dharma* masters (*fashi*), some were called scriptural masters (*zangshi*), some were called disciplinary masters (*lushi*) and some were meditation masters. These titles could be applied to a monk (or nun) of any school, since they denoted one's methodological focus rather than one's ideological leanings.

Chan begins to denote a specific doctrinal and meditative ideology around the time of Huineng (638-713). Although Chan tradition describes a transmission by five patriarchs culminating in Huineng as the sixth patriarch, as noted above, that transmission is more fiction than fact. Huineng's followers established the Southern School of Chan, which unleashed a polemical tirade against the Northern School. Since the Northern School disappeared about a thousand years ago, our only source of information on these schools was the prejudiced accounts of the Southern School, until the discovery at Dunhuang early this century of Northern School documents. We now know that many different versions of lineage histories were circulated, and, more importantly, that the positions attributed to the Northerners by their Southern rivals were grossly inaccurate and unfair. In fact, the Northern School had initially been the more successful of the two, but its success led to its ultimate ruin, since its growing dependence on Imperial patronage made it a vulnerable target during times of Imperial persecution of Buddhism. The Southern School, because it had taken root in remote areas less affected by actions of the Central government, survived the persecutions relatively intact.

Huineng is depicted in the *Platform Sutra* (authored by his leading follower and promoter, Shenhui) as an illiterate seller of firewood who experiences sudden enlightenment while overhearing someone reciting the *Diamond Sutra*. He joins a monastery where, without any official training in scriptures or meditation, he demonstrates that his enlightenment is more profound than all the monks who had been practising for years. Hence sudden enlightenment is one of the main tenets of the *Platform Sutra* (and subsequently for all forms of Chan). Another is 'direct pointing at mind', which, similar to the Tiantai approach, means that what is important is to observe one's own mind, to recognize that the nature of one's mind is Buddha-nature itself (see Platform Sutra).

While some Buddhists had argued that the goal was wisdom, and meditation was merely a means to that goal, Huineng argued for the inseparability of meditation and wisdom. Using an analytic device probably introduced by

the so-called neo-Daoist Wang Bi (226-49), the *tiyong* model (see Ti and yong), Huineng claimed that meditation is the essence (*ti*) of wisdom, and wisdom is the function (*yong*) of meditation. Wisdom does not produce meditation, nor does meditation produce wisdom; nor are meditation and wisdom different from each other. He drew an analogy to a lamp: the lamp is the *ti*, while its light is the *yong*. Wherever there is a (lit) lamp, there is light; wherever there is lamplight, there is a lamp. Lamp and light are different in name but identical in substance (*ti*), hence nondual.

Huineng's style of Chan was still sober, calm, rational, and rooted in commonly accepted Buddhist tenets. New and more radical elements were soon incorporated into Chan, some iconoclastically renouncing meditation and practice as well as scholasticism, and others trying earnestly to work out a rational system by which Chan could be syncretized with the other schools. Zongmi (780-841) considered a patriarch of both the Chan and Huayan schools, attempted just such a synthesis, but his sober approach was soon overshadowed in China by more abrupt, startling forms of Chan.

Of the 'Five Houses of Chan', only the Linji school survives today in China, Taiwan and Korea. Based on the teachings of Linji (d. 867), this school possibly provided Buddhism with its most 'Chinese' voice. Chan literature of the Linji and related schools were among the first texts ever written in vernacular as opposed to classical Chinese. Daoist elements also began to appear prominently. Zhuangzi's 'true man' becomes Linji's 'true man of no rank' who is going in and out of each person's face this very moment, and is always right here before one. The anecdotal humour associated with Zhuangzi's stories and the irreverent exploits of the Bamboo Sages of the Six Dynasties period clearly infused the style of Chan anecdotes. Rather than indulge in elaborate, complicated theoretical abstractions, Chan focused on experience as lived, in terms familiar to anyone immersed in Chinese culture (though often exotic to Western students, which has led to the common misconception that Chan is nonsensical or obscurantist).

Teaching techniques began to overshadow doctrinal content. At the heart of Chan training are the exchanges between teacher and student. Records, called *gongan*, were compiled of classic encounters, and even these eventually became part of the teaching techniques, as they were presented to students as riddles to concentrate on during meditation. To disrupt the sort of idle or pernicious speculation that could prove a hindrance to enlightenment, abrupt and shocking techniques were employed, from radical statements such as, 'If you meet Buddha on the road, kill him!', to exchanges punctuated by blows and shouts (all the more startling in the subdued monastic atmosphere in which they would unexpectedly occur). Linji's methods were designed to make students confront and overcome their mental and emotional habits and crutches, so as to become truly free and independent. Even dependency on Buddhism could be such a crutch. Linji summarized his teaching with the phrase: 'Don't be deceived.'

10 The Chinese Buddhist Schools: Pure Land

All forms of Chinese Buddhism, including Chan, contain devotional elements and rituals, but for Pure Land Buddhists devotionalism is the essence. The origins of Pure Land Buddhism are somewhat unclear. While undoubtedly devotional practices were imported to China by monks and laity (and these were blended with native Chinese forms of devotionalism), there does not seem to be a distinct school in India devoted to rebirth in Amitābha's Pure Land. As noted earlier, the traditional lineages are not very helpful for reconstructing the school's history. According to tradition, early contributors to Pure Land thought and practice include Tanluan (476-542), Dao Chuo (562-645) and his student Shandao (613-81). The term 'Pure Land' (*jingtu*) may itself be largely a product of certain license taken by translators. The term *jingtu* appears in Kumārajīva's translation of the *Vimalakīrti Sutra* where the Tibetan version simply has 'Buddha lands' (the Sanskrit version is no longer extant). Apparently, Xuanzang was the first explicitly to associate Sukhāvati (Amitābha's Paradise) with the term 'pure land'. The main scriptures for Pure Land practice were the Larger and Smaller *Sukhāvati-vyuha Sutras* and the *Guan Wuliangshuofo jing*.

At the beginning of the Tang Dynasty, several forms of Buddhist devotionalism were popular, including cults devoted to Mañjuśrī(Bodhisattva of Wisdom), Guanyin (Bodhisattva of Compassion, at that time particularly popular as a patron saint and protector of travellers), Maitreya (the future Buddha) and Amita (a conflation of Amitābha and Amitāyus whose names mean 'Infinite Light' and 'Infinite Life' respectively, and are possibly deities of Central Asian origin). Arguably the most popular form of devotionalism was the Maitreya cult. The

Empress Wu (r. 683-705), a great patron of Buddhism but generally reviled in Chinese history as an unscrupulous usurper, considered herself an incarnation of Maitreya. Due to her unpopularity once dethroned, people wanted to distance themselves from her and anything associated with her. Unfortunately this effectively extinguished Maitreya worship in China. Worshippers of Amita filled the void.

Pure Land theology maintained that people were living in the age of degenerate *dharma*, when study and personal effort were insufficient for making progress on the path to liberation. Relying on one's own efforts was in fact deemed a form of self-theory, or the selfishness and arrogance that comes from erroneous views of self. Rather than indulge in egoistic fantasies, one ought to rely on the power and grace of Amita. Amita was a *buddha* (whether he was an earlier incarnation of the historical Buddha or another person altogether is answered differently by different Pure Land sources) who, while still a *bodhisattva*, vowed to help sentient beings once he became a *buddha*. He has the power to transfer to anyone he deems worthy sufficient merit to enable them to be born in his Pure Land, the Western Paradise. In the earliest forms of Pure Land devotionalism a variety of practices were cultivated, but these were eventually pared down to chanting the *nianfo*, literally 'remembrance of Buddhas' (in Sanskrit, *buddhānusmṛti*), which in Chinese is 'Na-mu A-mi-to Fo' (Hail Amita Buddha).

11 Sinicizing Buddhist concepts

Since the time of Mencius, the ultimate ontological issue in China was the question of human nature and mind (which Mencius and most Chinese thinkers treated as synonyms). Pre-Han Chinese philosophers had debated whether human nature was originally good, bad or neutral. The written Chinese character for 'nature', *xing*, consists of two parts: the left side means 'mind' and the right side means 'birth', which led Chinese thinkers to debate whether human nature was determined by what one is born with, namely appetites and desires, or whether it reflects the nature of one's mind, which in Chinese thought invariably carried an onto-ethical rather than strictly cognitive connotation. The word *xin* literally means 'heart', indicating that - unlike Western conceptions that draw a sharp line between the head and the heart - for the Chinese, thinking and feeling originated in the same bodily locus. Feeling empathy or compassion as well as rationally abstracting principles and formulating ethical codes were all activities of *xin*, heart-and-mind (see Xin; Xing).

Indian Buddhism had little to say about human nature, with many forms of Buddhism rejecting the very concept of essential nature. Some of the early polemics against Buddhism in China explicitly attacked it for neglecting to address the question of human nature. The notion of Buddha-nature was developed, in part, to redress that failing. Since Indian Buddhists were deeply interested in the mind in terms of cognitive processes such as perception, thinking, attention and so on (see Mind, Indian philosophy of; Sense perception, Indian views of), it was a natural step for the Chinese to read these initially in the light of the Chinese discourse on mind, and to further develop interpretations of this material in line with Chinese concerns. Hence passages that in Sanskrit dealt primarily with epistemology or cognitive conditions often became, in their Chinese renderings, psycho-moral descriptions. The Sanskrit term *ekacitta*, a mind with singular focus (but literally meaning 'one mind') becomes the metaphysical one mind of the *Awakening of Faith in Mahāyāna*. Similarly, Indian and Chinese philosophers had developed very different types of causal theories. Indian Buddhists accepted only efficient causes as real, while Chinese Buddhists tended to interpret Buddhist causal theories as examples of formal causes.

12 Sinicizing Buddhist concepts: emptiness

Before Buddhism entered China Daoists had already embraced a notion of emptiness which it took Buddhists several centuries to realize was significantly different from their own (see Daoist philosophy). Laozi had contrasted the empty or open (*xu*) with the solid. What made a wheel functional was its empty hub; what made a vessel or room functional was its open space. Hence emptiness (or openness) is not worthless but rather the key to functionality and usefulness (see Daodejing). Later Daoists contrasted existents (*you*) with nonexistence (*wu*), and claimed that all existence emerges from nonexistence and ultimately returns to nonexistence (see You-wu). Some Chinese metaphysicians, such as Wang Bi, wrote about primordial nonexistence (*yuan wu*, *benwu*) as the metaphysical source, destination and substratum for all existent things. Thus form and emptiness were opposed, contrasting poles, and emptiness had primacy.

Some early Chinese Buddhists interpreted Buddhist emptiness in the same fashion, especially in the Prajñāschools. Eventually Buddhists realized, as the *Heart Sutra* says, that form and emptiness are not opposed to each other, but

that ‘form itself *is* emptiness, emptiness itself *is* form, form is not different from emptiness, emptiness is not different from form.’ In other words, Buddhist ‘emptiness’ did not mean ‘open’ or ‘nonexistence’. Emptiness (*śūnyatā*) signified the absence of an eternal, independent, self-causing, invariant, essential self-nature (*svabhāva*) or selfhood (*ātman*) in any thing or person. Whatever existed did so by virtue of a perpetually changing web of causes and conditions that themselves were products of other causes and conditions. Stated simplistically, emptiness does not mean that a table is unreal or nonexistent, or that its solid texture or colour are unreal; it does mean that the concept of tableness is unreal, and that the abstractions ‘solidity’ and ‘colour’ are unreal apart from the discrete and particular sensations one has at specific moments due to specific causes and conditions. Buddhist emptiness is not a primal void, but the absence of self-essence (see Buddhist concept of emptiness). To avoid being confused with Daoist concepts of emptiness, the Buddhists eventually chose a new term, *kong*, to render their ‘emptiness’.

Emptiness is neither the origin nor terminus for forms; forms themselves at any moment are emptiness. Since everything is causally connected with everything else, and there are no independent identities beyond or behind such causes and conditions, everything, according to Huayan, mutually interpenetrates and conditions everything else. Every thing defines and is defined by every other thing.

13 Sinicizing Buddhist concepts: suffering and ignorance

For all forms of Indian Buddhism, the fundamental fact with which Buddhism begins, and the problem it attempts to solve, is the problem of suffering (*duḥkha*). The first of the Four Noble Truths is: ‘All is suffering.’ Suffering does not mean simply pain. Buddhism does not deny joy, pleasure, delight and so on; but it claims that all is impermanent, so that whatever the source of a particular pleasure, that pleasure can never be permanent. The more pleasure one feels, the greater becomes one’s attachment to the presumed source. The greater the attachment, the greater the pain at the loss of that pleasure. Since everything is impermanent, such loss is inevitable. So, ironically, pleasure itself is ‘suffering’. Suffering is the affective reaction to impermanence. According to general Buddhist causal analysis, the causes of suffering are desire and ignorance. We desire permanent pleasures because we are ignorant of the fact that all is impermanent, empty of eternal selfhood. As the Four Noble Truths state, these causes of suffering can be eliminated, and Buddhism is the method or path for eliminating those causes. The purpose of Buddhism, then, is the elimination of suffering.

Chinese Buddhist texts do occasionally mention suffering, but usually in passing. Instead the root problem became ignorance. Discussions of the dialectical conflict between ignorance and enlightenment grew so pervasive that suffering was all but forgotten. This shift helped reinforce the emphasis on mind and mind-nature. Enlightenment was no longer defined as awakening to the causes of suffering, but instead denoted seeing the nature of the mind itself (see Suffering, Buddhist views of origination of).

14 Sinicizing Buddhist concepts: is Buddha-nature good or evil?

The pre-Han debate about whether human nature is good, evil or neutral was echoed in debates between Chinese Buddhists about Buddha-nature. Huayan contended that Buddha-nature and *tathāgatagarbha* were pristinely pure and good, filled with infinite good merits and qualities. In the fully realized perfection of Buddha-nature all evil, impurities and delusions have been eradicated.

This position was opposed by Tiantai, which argued that some evil (that is, some ignorance) remains in Buddha-nature. Following the Daoist sense of nonduality, in which good and evil or pure and impure are complimentary opposites as impossible to separate from each other as East from West, Tiantai accused Huayan of dualistic extremism. From the Tiantai perspective, Huayan’s ‘obsession’ with purity and goodness was one-sided and dualistic. Moreover, Tiantai insisted that it is necessary for Buddha-nature to retain some traces of evil and delusion in order to understand and empathize with the plight of ordinary sentient beings. If one becomes too rarefied, too transcendent, one loses touch with the everyday reality in which people wander deludedly, and thus one becomes incapable of effectively saving such people. Buddhahood, for Tiantai, was not simply a matter of correctly seeing or understanding in a ‘pure’ way, but was at its core salvific; Buddhahood is the active liberation of sentient beings from ignorance.

The debate on Buddha-nature heated up during the Song Dynasty. Heterodox forms of Tiantai tinged with

Huayan's 'purity obsession' appeared, and these were challenged sharply by the orthodox Tiantai thinkers from their headquarters on Tiantai mountain (from which the school took its name). The heterodox schools were labeled the Off-Mountain groups, while the orthodoxy styled itself the On-Mountain group. Zhili (959-1028), one of the On-Mountain leaders, had a keen intellect alert to the subtlest hints of Huayan-like thinking lurking in the rhetoric of Off-Mountain thinkers; his writings systematically ferret out and refute those implications with a logical sophistication rarely equalled amongst Chinese Buddhist philosophers.

These debates gain additional importance when viewed in the larger context of Chinese intellectual history. In the pre-Han period, Mencius' contention that human nature is originally good did not prove persuasive. Others argued that human nature was essentially neutral and subject to the influence of external conditions. Another early Confucian, Xunzi, had argued that human nature was basically selfish and greedy, which is why human society needs sages such as Confucius to guide them beyond the baseness of their own nature (see Xing). Han Confucians sided with Xunzi rather Mencius. The Tiantai position, by insisting that some evil and ignorance exists even in Buddha-nature, was close to some readings of Xunzi's position, while the idealistic optimism of the Huayan view clearly showed parallels with Mencius. Between the Han and Song Dynasties (third through tenth centuries), Confucianism was by and large intellectually stagnant. It found new vitality in the Song in part by reabsorbing back into itself the elements it had 'lent' to the Buddhists (and to some extent Daoists as well). The elements they took back had been modified and expanded by the Buddhists, and given metaphysical foundations that the neo-Confucians retained and continued to rework. Neo-Confucian thinkers, especially after Zhu Xi (1130-1200), rediscovered Mencius and unanimously embraced his view of the original goodness of human nature. Looked at another way, neo-Confucianism adopted Huayan's metaphysics of nature. Zhu Xi's famous dialectic of principle (*li*) and 'material-energy' (*qi*) owed more than a little to Huayan's *li* and *shi* metaphysics (see Li; Qi).

See also: Awakening of Faith in Mahāyāna; Buddhist philosophy, Indian; Buddhist philosophy, Japanese; Buddhist philosophy, Korean; Chinese philosophy; Daoist philosophy; Fazang; Linji; Neo-Confucian philosophy; Platform Sutra; Seng Zhao; Ûisang; Wônch'ūk; Wônhyo; Zhi Dun; Zhiyi; Zongmi

DAN LUSTHAUS

References and further reading

Buswell, R. (ed.) (1990) *Chinese Buddhist Apocrypha*, Honolulu, HI: University of Hawaii Press.(An important collection of essays detailing the impact of apocryphal texts - Chinese creations purporting to represent Indian originals - on the development of Chinese Buddhism.)

Ch'en, K. (1964) *Buddhism in China*, Princeton, NJ: Princeton University Press.(Though somewhat dated, still a classic overview of the early history of Buddhism in China.)

Ch'en, K. (1973) *The Chinese Transformation of Buddhism*, Princeton, NJ: Princeton University Press.(Another dated classic.)

Cook, F. (1977) *Hua-yen Buddhism: The Jewel Net of Indra*, University Park, PA: Pennsylvania State University Press.(A good introduction to basic Huayan doctrine.)

Gimello, R. (1976) 'Chih-yen (602-668) and the Foundations of Hua-yen Buddhism', Ph.D. dissertation, Columbia University.(Though unpublished, this remains the best historical analysis available in English of Chinese Buddhism in the sixth and seventh centuries.)

Gregory, P. (1995) *Inquiry Into the Origin of Humanity: An Annotated Translation of Tsung-mi's Yüan-jen lun with a Modern Commentary*, Honolulu, HI: Kuroda Institute and University of Hawaii Press.(A translation of an important Chinese text, richly annotated by Gregory, who intends this book as a survey and primer of Chinese Buddhist thought for intermediate students.)

Hurvitz, L. (1960-2) *Chih-I (538-597): An Introduction to the Life and Ideas of a Chinese Buddhist Monk*, Mélanges chinois et bouddhiques vol. 12.(A thorough overview of the foundational Tiantai thinker, Zhiyi.)

Liu Ming-wood (1994) *Madhyamaka Thought in China*, Leiden: Brill.(Useful discussion of the key developments in Chinese Madhyamaka, especially Liu's treatment of Jizang.)

Robinson, R. (1967) *Early Mādhyamika in India and China*, Madison, WI: University of Wisconsin Press.(Classic work, detailing the efforts of Kumārajīva and his contemporaries.)

Swanson, P. (1989) *The Philosophy of T'ien-t'ai*, Berkeley, CA: Asian Humanities Press.(Important historical overview of the texts and issues that lead to Zhiyi's distinctive philosophy, with a partial translation of one of Zhiyi's discussions of the *Lotus Sutra* that serves as a good example of the dense textual style of this form of Buddhism.)

Wright, A. (1990) *Studies in Chinese Buddhism*, ed. R. Somers, New Haven, CT: Yale University Press.(Essays dealing with important people and issues during the formative years of Buddhism in China.)
Yampolsky, P. (1967) *The Platform Sutra of the Sixth Patriarch*, New York: Columbia University Press. (Yampolsky's translation of this Chan sutra is somewhat problematic, but his introductory essays demonstrate the importance of critical historical methods for approaching this material.)
Zürcher, E. (1959) *The Buddhist Conquest of China*, Leiden: Brill, 2 vols.(Another classic, dealing with the formative periods of Chinese Buddhism.)

Buddhist philosophy, Indian

Buddhism was an important ingredient in the philosophical melange of the Indian subcontinent for over a millennium. From an inconspicuous beginning a few centuries before Christ, Buddhist scholasticism gained in strength until it reached a peak of influence and originality in the latter half of the first millennium. Beginning in the eleventh century, Buddhism gradually declined and eventually disappeared from northern India. Although different individual thinkers placed emphasis on different issues, the tendency was for most writers to offer an integrated philosophical system that incorporated ethics, epistemology and metaphysics. Most of the issues addressed by Buddhist philosophers in India stem directly from the teachings attributed to Siddhārtha Gautama, known better through his honorific title, the Buddha.

The central concern of the Buddha was the elimination of unnecessary discontent. His principal insight into this problem was that all dissatisfaction arises because people (and other forms of life as well) foster desires and aversions, which are in turn the consequence of certain misunderstandings about their identity. Discontent can be understood as frustration, or a failure to achieve what one wishes; if one's wishes are generally unrealistic and therefore unattainable, then one will naturally be generally dissatisfied. Since the Buddha saw human frustration as an effect of misunderstandings concerning human nature, it was natural for Buddhist philosophers to attend to questions concerning the true nature of a human being. Since the Buddha himself was held as the paradigm of moral excellence, it was also left to later philosophers to determine what kind of being the Buddha had been. A typical question was whether his example was one that ordinary people could hope to follow, or whether his role was in some way more than that of a teacher who showed other people how to improve themselves.

The Buddha offered criticisms of many views on human nature and virtue and duty held by the teachers of his age. Several of the views that he opposed were based, at least indirectly, on notions incorporated in the Veda, a body of liturgical literature used by the Brahmans in the performance of rituals. Later generations of Buddhists spent much energy in criticizing Brahmanical claims of the supremacy of the Veda; at the same time, Buddhists tended to place their confidence in a combination of experience and reason. The interest in arriving at correct understanding through correct methods of reasoning led to a preoccupation with questions of logic and epistemology, which tended to overshadow all other philosophical concerns during the last five centuries during which Buddhism was an important factor in Indian philosophy.

Since the Buddha saw human frustration as an effect that could be eliminated if its cause were eliminated, it was natural for Buddhist philosophers to focus their attention on a variety of questions concerning causality. How many kinds of cause are there? Can a multiplicity of effects have a single cause? Can a single thing have a multiplicity of causes? How is a potentiality triggered into an actuality? Questions concerning simplicity and complexity, or unity and plurality, figured prominently in Buddhist discussions of what kinds of things in the world are ultimately real. In a tradition that emphasized the principle that all unnecessary human pain and conflict can ultimately be traced to a failure to understand what things in the world are real, it was natural to seek criteria by which one discerns real things from fictions.

1 Human nature

A key tenet of Buddhist doctrine is that discontent is an outcome of desires grounded in false beliefs. The most important of these false beliefs are that (1) one's own individual existence is more important than those of other individuals, and that (2) fulfilment can be achieved by acquiring and owning property. If these misunderstandings can be replaced by an accurate view of human nature, suggested the Buddha, then unrealistic craving and ambition will cease, and so will frustration. Happiness, in other words, can be achieved by learning to recognize that (1) no one is more important than anyone else, since all beings ultimately have the same nature, and that (2) the very idea of ownership is at the root of all conflicts among living beings. The methods by which one achieves contentment, according to the Buddha, are both intellectual and practical. One can gradually become free of the kinds of beliefs that cause unnecessary pain to oneself and others by carefully observing one's own feelings and thoughts, and how one's own words and actions affect others. To counter the view that one's own individual existence is more important than the existence of other beings, Buddhist philosophers adopted the radical strategy of trying to show that in fact human beings do not have selves or individual identities. That is, an attempt was made to show that

there is nothing about a person that remains fixed throughout a lifetime, and also that there is nothing over which one ultimately has real control. Failure to accept the instability, fragmentation and uncontrollability of one's body and mind is seen as a key cause of frustration of the sort that one could avoid by accepting things as they really are. On the other hand, realizing that all beings of all kinds are liable to change and ultimately to die enables one to see that all beings have the same fundamental destiny. This, combined with the recognition that all living beings strive for happiness and wellbeing, is an important stage on the way to realizing that no individual's needs, including one's own, are more worthy of consideration than any other's.

The notion that one does not have an enduring self has two aspects, one personal and the other social. At the personal level, the person is portrayed in Buddhist philosophy as a complex of many dozens of physical and mental events, rather than as a single feature of some kind that remains constant while all peripheral features undergo change. Since these constituent events are incessantly undergoing change, it follows that the whole that is made up of these constituents is always taking on at least some difference in nature. Whereas people might tend to see themselves as having fixed personalities and characters, the Buddha argued it is always possible for people either to improve their character through mindful striving, or to let it worsen through negligence and obliviousness. Looking at the social aspects of personal identity, the Buddha maintained, in contrast to other views prevalent in his day, that a person's station in human society need not be determined by birth. According to the view prevalent in ancient and classical Indian society, a person's duties, responsibilities and social rank were determined by levels of ritual purity; these were in turn influenced by pedigree and gender and various other factors that remained constant throughout a person's lifetime. In criticizing this view, Buddhist philosophers redefined the notions of purity and nobility, replacing the concept of purity by birth with that of purity by action (karma) (see Karma and rebirth, Indian conceptions of §5). Thus the truly noble person, according to Buddhist standards, was not one who had a pure and revered ancestry, but rather one who habitually performed pure and benevolent actions.

Given these basic ideas of human nature as a starting point, later generations of Buddhist thinkers were left with the task of explaining the mechanisms by which all the components of a person work together; this also involved trying to explain how human beings can gradually change their character. While there was general agreement on the principle that the intentions behind one's actions led eventually to resultant mental states, that benevolent actions resulted in a sense of wellbeing, while malevolent actions resulted in uneasiness and vexation, the precise details of how karmic causality took place were a matter of much dispute. Especially difficult was the question of how actions committed in one lifetime could influence the character of a person in a different lifetime, for Buddhists accepted the notion of rebirth that was common in Indian systems of thought. Discussions of how people could improve their character presuppose that the people in question have not become irreversibly depraved. One controversy that arose among Buddhist thinkers was whether there are beings who become so habitually perverse that they can no longer even aspire to improve their character; if so, then such beings would apparently be heir to an unending cycle of rebirths.

The view of the person as a set of interconnected modules, the precise contents of which were always changing, was characterized by Buddhists as avoiding the untenability of two other hypotheses that one might form about human nature. One hypothesis is that a person has some essential core that remains unchanged through all circumstances. This core survives the death of the physical body and goes on to acquire a new body through a process of reincarnation. According to this view, the unchanging essential part of a person is eternal. The second hypothesis is that a person takes on an identity at birth and carries it through life but loses the identity altogether at death. The Buddhist view, characterized as a middle way between these two extremes, is that a person's character is always in flux, and that the factors that determine the particular changes in a person's mentality continue to operate even after the body housing that mentality dies. So Buddhists tended to claim that what goes from one living body to another is not an unchanging essence, but rather a set of tendencies to behave in certain ways.

2 Ethics

The strongest motivation for accepting the doctrine of rebirth was to support the notion that people are accountable for their actions to the very end of their lives; the doctrine thus plays a central role in Buddhist ethical theory. It was noted in the preceding section that the Buddhist view of the person was described as a middle path between two equally untenable extremes. In the realm of conduct also, the Buddhists described theirs as a middle path or a moderate position that avoided extreme views of human conduct. In order to understand the various positions

against which the Buddhists defined their views on appropriate conduct, it should be borne in mind that the central question being asked by the Buddha and his contemporaries was how to achieve contentment. The strategies recommended by different thinkers were closely related to their views of life after death. Those who held that a person has only one life tended to argue that one's life should be spent in the pursuit of as much pleasure as is possible without bringing pain and injury to oneself. Restraint in the pursuit of pleasure was seen as necessary only to the extent that excessive indulgence might shorten one's life and decrease one's opportunities for future pleasure seeking.

Philosophers who accepted the doctrine of rebirth, on the other hand, tended to argue that the only kind of happiness worth pursuing was lasting freedom from the pains and turmoil of life; this could be won only by bringing rebirth to an end. After death, they said, all living beings are eventually reborn in a form of life determined by the accumulated effects of deeds done in previous lives. Although some forms of life might be very pleasant and offer a temporary reward for previous good actions, every form of life involves some amount of pain and suffering, even if it is only an anxiety that one's present peace and happiness will eventually come to an end and be replaced by more direct forms of physical and mental pain. Therefore, the only hope of any lasting freedom from the pains of existence is to remove oneself from the cycle of birth and death altogether. Exactly how this was to be achieved was a matter of much controversy, but some drastic methods involved undergoing extreme forms of austerity and even self-inflicted pain. The Buddhist middle path, therefore, was one that avoided two extremes: one extreme was the self-indulgence of those who denied life after death altogether, and the other was the self-torture recommended by some ascetics as the only way to gain freedom.

Buddhist philosophers tended to agree that a person's mentality at any given moment is either virtuous, vicious or neutral. This means that all of one's mental characteristics in a given moment have the same orientation, which is either towards a state of happiness, the natural consequence of virtue, or towards a state of discontent, the natural consequence of vice. The principal virtues that were said to cooperate in a healthy mentality were correct understanding, which manifested itself as a sense of shame, and a sense of decency, usually interpreted as respect for oneself and respect for others. Thus if one has the virtue of having a sense of shame, then while that sense of shame is functioning, one will also have the virtues of being generous, free of malice and open-minded; having these virtues makes one likely to behave in ways that conduce to the health of oneself and others. If, on the other hand, one has the vice of being shameless, then one will also have the vices of being deluded and agitated and therefore prone to behave in ways likely to bring harm to oneself and others. While virtuous and vicious mental qualities cannot be present in the same mentality at the same moment, it could very well be that a person vacillates between virtuous and vicious frames of mind. Indeed, this is said to be the condition of the vast majority of living beings (see Duty and virtue, Indian conceptions of §3; Virtues and vices).

Despite a tendency to agree on these basic matters, Buddhist philosophers disagreed with one another over several other questions. There was, for example, controversy over whether people could arrive at stages of attainment from which they could never backslide. Some argued that once people gained certain insights into reality, then they could no longer be deluded in the ways that result in acting on self-centred motivations. Others argued that, even if backsliding might be unlikely for some people, it is in principle always possible, and therefore a person can never afford to be complacent. Another controversy arose over whether a vicious person could be fully aware of a virtuous person's virtues, some Buddhists holding the view that only a virtuous person can recognize that another person is also virtuous. Yet another matter of controversy had to do with whether the merit of being virtuous could be transferred to others. Some argued that each person is strictly accountable for their own actions and that no one can escape the ill effects of their intentionally harmful actions. Others claimed that merit can be transferred to others, enabling them to experience levels of happiness that they could never have deserved on the merit of their own actions. Closely tied to this controversy was the question mentioned above, concerning whether some beings fall into such states of depravity that they can no longer even aspire to be good. Those philosophers who accepted that beings could become depraved to this extent but denied that merit can be transferred had to conclude that some beings would never attain *nirvāṇa* (see Nirvāṇa). Other philosophers, for whom the prospect of eternal suffering in the cycle of birth seemed unjust, favoured the doctrine that merit could be transferred, thus enabling these thoroughly depraved beings to undergo the change of mentality necessary to begin leading a life of virtue.

3 Buddha-nature

Even in the earliest strata of Buddhist literature that has survived to the present, the Buddha is portrayed in a variety of ways. Some passages depict him as a man who skilfully answers questions that have been put to him, either by answering the questions or by showing why the question as asked cannot be answered. The passages were clearly designed to portray the Buddha as a paragon of wisdom, whose careful and analytic thinking could be used as a model for those seeking to arrive at correct understanding. The Buddha is also portrayed as a model of virtue, a man who has mastered the art of living in the world without bringing harm to other living beings and whose concern for the welfare of all living things around him is unsurpassed. Interspersed with these passages that focus on the Buddha as a remarkable man, there are other passages that portray the Buddha as a superhuman miracle worker whose mastery of yogic technique has given him the power to travel hundreds of kilometres in the blink of an eye, transport himself and his followers through the air, know the precise thoughts of other people, see into the past and future, heal serious wounds merely by looking at them, and soothe wild and dangerous animals by merely speaking gently to them. Some texts show him inviting his followers to question everything he says and to accept nothing on his authority; in other passages, he is portrayed as a man to whom even the wisest and most knowledgeable gods come so that the profound mysteries of the universe can be explained to them in clear words. Given the diversity of things said about the Buddha in the texts that Buddhists regarded as authoritative, it is no wonder that among the points about which there was considerable controversy was the nature of the Buddha himself. Discussions about the nature of the Buddha were as important to some Buddhist philosophers as discussions about the nature of God were to the theologians of theistic traditions (see Buddha).

The earliest extant record of controversies concerning the nature of the Buddha is a work known as *Kathāvatthu (Points of Controversy)*, supposed to have been written around 246 BC by an elder monk known as Tissa Moggalīputta. This treatise mentions over two hundred topics over which there was controversy among Buddhists, of which several pertain to the nature of the Buddha. According to this text, some Buddhists held to the view that the Buddha pervades all regions of space at all times and has the power to suspend all the laws of nature at will; others argued that the Buddha exists only where his human body is located and that he is bound by all the natural laws by which other living beings are bound. Those who accepted the Buddha as a ubiquitous and eternal entity tended to claim that the human Buddha was merely a manifestation in human form that appeared for the sake of guiding human beings. This apparition, they claimed, had no real need for food and shelter or other material requirements of life, but it accepted such gifts from devotees so that they might learn the benefits of generosity. Moreover, this apparition was said to be wholly lacking any of the unpleasant physical or mental traits of a human being and never had any thoughts that were not directed at teaching people how to cultivate virtue and attain *nirvāṇa*. Other Buddhists rejected this view of the Buddha altogether and argued that he was a mortal just like all other mortals, except that among the limited range of topics about which he had knowledge was the important matter of how to achieve lasting peace and happiness. This issue was controverted for over a millennium in India, with Dharmakīrti and some of his followers taking up the view of the Buddha as an ordinary mortal, while some members of the Yogācāra movement took up the position that the Buddha was more of a cosmic principle; the eleventh-century Buddhist Ratnakīrti eventually argued that all particular acts of individual awareness are merely parts of a single, universal consciousness, which he identified as the mind of the Buddha (see Buddhism, Yogācāra school of).

The position taken by Ratnakīrti may be the logical conclusion of an idea first mentioned in the *Kathāvatthu*, namely, that one becomes an Awakened One (*buddha*) by acquiring a quality known as awakening (*bodhi*). Tissa Moggalīputta himself rejected this idea, arguing that if awakening were something that one could attain, it would also be something that one could lose, in which case a *buddha* could cease to be a *buddha*; his view, therefore, was that awakening is not a positive trait but merely the absence of delusions. Dissenting from this view, other Buddhists (and especially members of the Yogācāra school) argued that *buddhas* become *buddhas* as a result of realizing an innate potential to become awakened. This innate potential, called the embryo of the knower of truth (*tathāgata-garbha*), was said by some to exist in all living beings, thereby making all living beings *buddhas*, or at least *buddhas* in the making. From this view that every sentient being has the essential quality of a *buddha*, even if this essence is somehow obscured from view by others, it was a short step to the view that all sentient beings are identical in their essence and therefore not really different from one another. Closely related to this controversy over the nature of the Buddha's essence was the issue of whether there are degrees of buddhahood or different ranks of *buddhas*. Tissa Moggalīputta had argued that being a *buddha* is a matter of being free of delusion, and either one is free or one is not; there can be no degrees of freedom. Other thinkers took the view that although all

beings are essentially *buddhas* in their nature, they manifest their essences to a different extent, and therefore one may speak of degrees of buddhahood.

4 Epistemology

Siddhārtha Gautama the Buddha is portrayed in Buddhist literature as ridiculing the sacrificial rituals of the Brahmans and accusing the priests of fabricating them for no better reason than to make money from the wealthy and to manipulate the powerful. Attacking the sacrificial practices of the Brahman priests in this way eventually led to challenging the authority of the Vedic literature that the priests considered sacred. An early Buddhist philosopher who challenged the authority of sacred texts was Nāgārjuna, whose arguments called into question the very possibility of justified belief. In a text called *Vigrahavyāvartanī(Averting Disputes)*, Nāgārjuna argued that all opinions are warranted by an appeal to experience, or to various forms of reasoning, or to the authority of tradition. Now among the opinions that one may hold, said Nāgārjuna, is the opinion that all opinions are warranted in one of those ways. Nothing, however, seems to warrant that opinion. If one should claim that that opinion is self-warranting, then why not grant that all other opinions are also self-warranting? On the other hand, if that opinion requires substantiation, the result will be an infinite regress. Therefore, concluded Nāgārjuna, no opinion can be grounded. Realizing that one can never arrive at certainty thus becomes for Nāgārjuna the most reliable way of freeing oneself from the various delusions that cause unhappiness in the world. Dispelling delusions is therefore not a matter of discovering truth, but a matter of realizing that all opinions that pass as knowledge are not really knowledge at all.

Although Nāgārjuna's scepticism managed to capture the spirit of some passages of Buddhist literature that depict the Buddha as questioning the authoritarianism of other teachers, it did not leave adequate room for distinguishing truth from error. Most Buddhist philosophers who came after Nāgārjuna, therefore, placed an emphasis on both eliminating error and securing positive knowledge. Dignāga, modifying theories of knowledge that Brahmanical thinkers had developed, argued that there are just two types of knowledge, each having a distinct subject matter unavailable to the other: through the senses one gains knowledge of particulars that are physically present, while the intellect enables one to form concepts that take past and future experiences into consideration.

In an important work called *Ālambanaparīkṣa (Examining the Support of Awareness)* Dignāga developed an argument that his predecessor Vasubandhu had made. Here Dignāga argued that a cognition is accurate only if the subject matter of the cognition is identical to that which causes the awareness to arise. So, for example, if one sees a dead tree in the dark and takes it to be a man, then the subject matter of the cognition is a man, which is not identical to the dead tree that is causing the cognition; the cognition is therefore inaccurate. Given this principle, said Dignāga, it follows that none of our sensory cognitions is accurate, because each of them is really caused by atoms massed together; nevertheless, we are never aware of anything as a mass of atoms. Instead, what we are aware of is such things as human beings, elephants and trees. These notions of things as human beings or elephants, however, are purely conceptual in nature and are not in accord with the realities that exist in the external world. Therefore, he concluded, the only objects of our awareness are concepts; we are never directly aware of realities as they occur outside the mind.

Dignāga's essays rekindled an interest in epistemological questions among Buddhist philosophers that lasted for several centuries. As influential as his theories in logic and the sources of knowledge were, there was little in them that explicitly referred to previous Buddhist doctrine. It was left to the systematic philosopher Dharmakīrti to draw out the implications that questions of logic and epistemology had for people interested in the traditional Buddhist preoccupation with eliminating delusions concerning the nature of the self in order to win freedom from discontent.

Dharmakīrti's system of epistemology was centred around his criticism of the Brahmanical doctrine of the special authority of the Veda, which the Brahmans supposed had been revealed to humanity by God. He combined this criticism with a defence of the doctrine that the Buddha was a source of knowledge. The Brahmanical claim of authority for their scriptures is based upon the notion that God is omniscient and compassionate; this idea, said Dharmakīrti, is extravagant and laughable. The Buddhist claim for the authority of the Buddha, on the other hand, is based on the modest claim that the Buddha was an ordinary man who could see the root cause of discontent, knew how to eliminate the cause, and took time to teach other human beings what he had discovered. Moreover, argued Dharmakīrti, the Buddha taught nothing but principles that every human being could confirm. Full

confirmation of the Buddha’s teachings was said to be impossible for a person whose vision was still clouded by delusions. On the other hand, to a person who had learned to listen to wise counsel, reflect on it and then put it into practice, all the teachings of the Buddha on the question of winning *nirvāṇa* would be confirmed.

The form of reflection that Dharmakīrti recommended was based upon a systematic study of the principles of legitimate inference; most delusions, he contended, stem from forming hasty generalizations from limited experience. Most doctrinal matters about which philosophers dispute, he said, cannot be decided with certainty. In this respect, Dharmakīrti adopted the cautious attitude of the Buddha and Nāgārjuna towards unwarranted opinions and assumptions. At the same time, he tried to show that the extreme scepticism that had characterized the work of such thinkers as Nāgārjuna was also unwarranted. This epistemological ’middle path’, consisting of modest claims about the extent of the Buddha’s knowledge and yet insisting that the Buddha’s doctrines were true and distinguishable from falsehoods, set the tone for most of the Buddhist philosophy that evolved in India until the time that Buddhism stopped being an important factor in the philosophical milieu of the Indian subcontinent. The focus on epistemological issues enabled Buddhists to set aside the sometimes bitter disputes over the question of which Buddhist scriptures were authentic (see Epistemology, Indian schools of §1).

5 Metaphysics

Given the emphasis in Buddhist teachings on the role of erroneous belief as a cause of unhappiness, it was natural that Buddhist philosophers should focus on questions of ontology and the theory of causation. Ontology was important, since a kind of intellectual error that was supposed to lead to unhappiness was being mistaken about what exists. The theory of causation was important, since the eradication of the cause of unhappiness was supposed to result in the removal of unhappiness itself.

The earliest attempts to systematize the teachings of Buddhism were in the genre of literature known as ‘Abhidharma’, in which all the factors of human experience were classified according to a variety of schemata (see Buddhism, Ābhidharmika schools of). The study of the relationships among these classes of factors eventually evolved into a detailed theory of causality, in which several types of causal relationship were enumerated. There were many schools of Abhidharma, and each had its own set of schemata for the classification and enumeration of the factors of experience. Indeed, each had its own interpretation of what the very word ‘Abhidharma’ means; among the possible interpretations of the word, a common one is that it means a higher or more advanced doctrine, or a doctrine that leads to a higher form of wisdom. The variety of approaches taken in Abhidharma literature makes it difficult to discuss this literature in any but the most general way. Among most schools of Abhidharma, there was a commitment to the idea that the best strategy for coming to an understanding of any complex being is to analyse that being into its ultimate parts. An ultimate part is that which cannot be analysed into anything more simple. Most Buddhist systematists held to the principle that the ultimately simple building blocks out of which things are made are ultimately real, while complex things that are made up of more simple parts are not ultimately real; they are held to be real only through the consensus of a community. As was seen above in the section on human nature (§1), for example, there was a strong tendency for Buddhists to accept that a person’s character is the product of many components; these components were held to be real, but the person was held to be ultimately unreal. The idea of a person may be a fiction, but it is one that makes the running of society more manageable, and therefore it can be regarded as a consensual reality, in contrast to an ultimate reality.

The philosopher Nāgārjuna questioned the whole attempt to make a distinction between consensual and ultimate truths. One interpretation of his philosophical writings is that he was trying to show that every attempt to understand the world can only be an approximation on which there may be some degree of consensus; there is, however, no understanding that can claim to have arrived at an adequate description of things as they really are. Along with this radical criticism of the very enterprise of trying to discern ultimate from consensual realities, Nāgārjuna criticized the doctrine that the simple constituents that serve as causes of more complex beings are more real than the complex beings themselves. This principle had rested on the assumption that the more simple a being is, the closer it is to being independent. In fact, he argued, the apparently simple constituents are no less dependent than their apparently complex effects. To this fact of being dependent upon other things, Nāgārjuna gave the name emptiness; since all beings are dependent for their existence on other beings, he said, all beings are empty.

Later Buddhist philosophers, beginning especially with Dharmakīrti, devoted their energy increasingly to refuting the claims, advanced by some Brahmanical thinkers, that the whole universe can be traced back to a single cause.

Dharmakīrti argued that if all things in the history of the world had a single cause, such as God or some type of primordial matter, then there would be no way to account for all the formal variety in the world at any given time, nor would it be possible to account for the fact that events unfold in sequences. If all the formal and temporal diversity are already inherent in the cause, he argued, then the cause is not a single thing after all. One might argue that the diversity exists in the single cause only as a potential of some kind; this, however, only raises the problem of explaining how that potential is actualized. If the potential is activated by something outside the cause that possesses it, then the outside agency must be counted among the causal factors along with the primary cause, in which case there is no longer a single cause. Besides general arguments directed against the view that all things could have any single cause, Dharmakīrti also gave arguments against the existence of a creator God in particular. The universe, he observed, shows no signs of having been designed by anyone intelligent. Even if it were conceded for the sake of argument that the world might have been made by some intelligent being, there is still no sign that this being had any concern for any living beings. Later generations of Buddhist philosophers expanded upon Dharmakīrti's arguments against the existence of a single intelligent creator, but most of these expansions took the form of replying to the objections of opponents rather than formulating new arguments (see Causation, Indian theories of §6).

During the last five hundred years that Buddhism was an important factor in Indian philosophy (600-1100), criticism of Buddhist doctrines by Brahmanical and Jaina religious philosophers, as well as from anti-religious materialists, forced Buddhist thinkers to refine some of their arguments and even to abandon some of their doctrinal positions. Arguments among Buddhists became much less a feature of Buddhist philosophy than arguments against non-Buddhist opponents. Within Buddhism itself, there was a tendency to try to reconcile differences that in earlier centuries had divided Buddhists against each other, at least doctrinally. This new spirit of overcoming sectarianism resulted in several ingenious attempts to fuse the old Abhidharma schools, the Mādhyamika schools, the Yogācāra schools and the Buddhist epistemologists. Few new issues were raised in this last five-hundred-year period, and not many new arguments were discovered to defend old positions. Careful scholarship tended to replace philosophical innovation as the principal preoccupation of later Buddhist intellectuals, such as Śāntarakṣita, Kamalaśīla, Jñānaśrīmitra and Ratnakīrti. It was during this period of the decline of Buddhism in India that Buddhist philosophy was introduced into Tibet. Once established there, it received a new impetus from a range of Tibetan intellectuals who were able to study many of these doctrines with a fresh perspective.

See also: Buddhism, Mādhyamika: India and Tibet; Buddhist concept of emptiness; Buddhist philosophy, Chinese; Buddhist philosophy, Japanese; Buddhist philosophy, Korean; Hindu philosophy; Jaina philosophy; Momentariness, Buddhist doctrine of; Nominalism, Buddhist doctrine of; Suffering, Buddhist views of origination of; Tibetan philosophy

RICHARD P. HAYES

References and further reading

Conze, E. (1967) *Buddhist Thought in India*, Ann Arbor, MI: University of Michigan Press.(A good survey of the issues and the schools of Indian Buddhism.)

Frauwallner, E. (trans.) (1956) *Die Philosophie des Buddhismus*, ed. W. Ruben, Berlin: Akademie-Verlag, 3rd revised edn, 1969.(A collection of important Indian Buddhist philosophical texts translated into German.)

Griffiths, P.J. (1994) *On Being Buddha*, Albany, NY: State University of New York Press.(The first two chapters present a cogent general account of the place of doctrine and philosophy within Indian Buddhism.)

Kalupahana, D.J. (1976) *Buddhist Philosophy: A Historical Analysis*, Honolulu, HI: University of Hawaii Press. (Informative, though biased to a view of Buddhism as 'empirical' and 'anti-metaphysical'.)

Mookerjee, S. (1980) *The Buddhist Philosophy of Universal Flux*, Delhi: Motilal Banarsidass.(Excellent account of the doctrines held and arguments advanced in the later scholastic period.)

Nāgārjuna (*c.* AD 150-200) *Vigrahavyāvartanī(Averting Disputes)*, trans. K. Bhattacharya, 'The Dialectical Method of Nāgārjuna', *Journal of Indian Philosophy* 1: 217-61, 1971.(A superb study, including a translation from the original Sanskrit with introduction and notes; see §4.)

Tissa Moggalīputta (*c.*246 BC) *Kathāvatthu (Points of Controversy)*, trans. Shwe Zan Aung and C.A.F. Rhys Davids, London: The Pali Text Society, 1979.(The standard English translation of the earliest collection of points disputed by Buddhists during the first two centuries after the Buddha's career, part of the *abhidhamma* section of the Pāli canon; see §3.)

Warder, A.K. (1970) *Indian Buddhism*, Delhi: Motilal Banarsidass.(This thorough study of the history of Buddhism in India contains several chapters on the history and principal schools of its philosophy.)

Williams, P. (1989) *Mahāyāna Buddhism: The Doctrinal Foundations*, London: Routledge.(The first six chapters of this excellent and readable study deal mostly with Buddhist philosophy in India. Also contains discussions of Tibetan and Chinese developments.)

Buddhist philosophy, Japanese

Buddhism transformed Japanese culture and in turn was transformed in Japan. Mahāyāna *Buddhist thought entered Japan from the East Asian continent as part of a cultural complex that included written language, political institutions, formal iconography and Confucian literature. From its introduction in the sixth century through to the sixteenth century, Japanese Buddhism developed largely by incorporating Chinese Buddhism, accommodating indigenous beliefs and reconciling intersectarian disputes. During the isolationist Tokugawa Period (1600-1868), neo-Confucian philosophy and Dutch science challenged the virtual hegemony of Buddhist ways of thinking, but served more often as alternative and sometimes complementary models than as incompatible paradigms. Only since the reopening of Japan in 1868 has Japanese Buddhist thought seriously attempted to come to terms with early Indian Buddhism, Western thought and Christianity.*

Through the centuries, Buddhism gave the Japanese people a way to make sense of life and death, to explain the world and to seek liberation from suffering. When it engaged in theorizing, it did so in pursuit of religious fulfilment rather than of knowledge for its own sake. As an extension of its practical bent, Japanese Buddhist thought often tended to collapse differences between Buddhism and other forms of Japanese religiosity, between this phenomenal world and any absolute realm, and between the means and end of enlightenment. These tendencies are not Japanese in origin, but they extended further in Japan than in other Buddhist countries and partially define the character of Japanese Buddhist philosophy.

In fact, the identity of 'Japanese Buddhist *philosophy' blends with almost everything with which we would contrast it. As a development and modification of Chinese traditions, there is no one thing that is uniquely Japanese about it; as a Buddhist tradition, it is characteristically syncretistic, often assimilating* Shintō*and* Confucian philosophy *in both its doctrines and practices. Rituals, social practices, political institutions and artistic or literary expressions are as essential as philosophical ideas to Japanese Buddhism.*

Disputes about ideas often arose but were seldom settled by force of logical argument. One reason for this is that language was used not predominately in the service of logic but for the direct expression and actualization of reality. Disputants appealed to the authority of Buddhist sūtras *because these scriptures were thought to manifest a direct understanding of reality. Further, as reality was thought to be all-inclusive, the better position in the dispute would be that which was more comprehensive rather than that which was more consistent but exclusive. Politics and practical consequences did play a role in the settling of disputes, but the ideal of harmony or conformity often prevailed.*

The development of Japanese Buddhist philosophy can thus be seen as the unfolding of major themes rather than a series of philosophical positions in dispute. These themes include the role of language in expressing truth; the non-dual nature of absolute and relative, universal and particular; the actualization of liberation in this world, life or body; the equality of beings; and the transcendent non-duality of good and evil.

1 Language

The practical, soteriological nature of Buddhism makes it wary of what Wittgenstein called the bewitchment of the mind by language. Yet the view that Buddhism disregards language is an oversimplification, especially in Japan. Japanese Buddhism has a positive appreciation of language, particularly of its non-referential usage. Buddhist thought first came to Japan in the form of *sūtras* and commentaries written in Chinese characters. Written language was still so novel, and categories of thought so limited, that the content of these texts could be read as manifestations of new realities rather than as ideas referring to Buddhism. Some of the earliest literature composed in Japan, such as the poetry collection *Man'yōshū*, mentions *kotodama*, the spiritual power of words that makes things present. Although scholars articulated this theory of language only much later and aligned it with Shintō(see Motoori Norinaga), it may apply to the early understanding of Chinese Buddhist as well as indigenous words. Words had the power not only to manifest things in the world but also to change them; before Buddhist texts represented a doctrinal and ethical system, they provided incantations to heal illness and bring prosperity to the land.

The ritualistic use of language continued through the centuries and was often central to the expression of doctrine.

Practitioners often used single words or phrases as the condensed form of a doctrine or lengthy *sūtra*. Kūkai, for example, recited the Sanskrit formulas for the Womb and Diamond Mandalas to better envision these pictorial representations of the cosmic order. He proposed that intoning mantras could make the basic sounds of the cosmos audible. He used *dhāraṇī* or magical formulae not only as a means of purifying body and mind and allowing him to understand the point of every Buddhist scripture, but also as the means by which his patron bodhisattva could fulfill all wishes. Even when later Pure Land School teachers such as Shinran suspected discursive language and discouraged belief in worldly benefit through magical transformation, they taught that the sincere invocation of Amida Buddha's name has the power to actualize the salvation of all sentient beings. Buddhists of the Nichiren schools chanted the name of the *Wondrous Lotus Sutra*, *namu myōhōrenge kyō*, as a condensation and realization of all doctrines contained in it. Rinzai Zen teachers advocated the practice of *kanna* or 'contemplating the [crucial] phrase' of a dialogue, or of compressing the already condensed *Heart Sutra* into a single word. As long as the words are 'live' and not dead repetitions, Zen teachers conceived such practices as shortcuts to enlightenment that immediately put the practitioner in a frame of mind to realize the point of Buddhist doctrines.

The transformative as well as expressive nature of language is evident in perhaps the most important collection of premodern Japanese philosophical literature, the *Shōbōgenzō(Treasury of the True Dharma Eye)*, by the thirteenth-century Zen master Dōgen. One chapter begins: 'As for the Buddha Way, not to voice it is impossible.' This statement does not command one to proclaim the teachings of Buddhism, but rather connects one's attainment of truth with its expression in the world. Such expression includes but is not limited to language. Another chapter of the *Shōbōgenzō* states, 'All buddhas and patriarchs are able to voice the Way', that is, to express truth in all their words and actions. By virtue of their realization of non-duality or no ultimate opposition, they can directly express the Way and not merely refer to it. They can also experience as the words of the Buddha 'the preaching of non-sentient beings', the 'sounds of the valley streams and the forms of the mountains'. This idea, explained further below (see §5), seems to collapse any ultimate distinction between sign and signified. Words express themselves, as do valley streams and the forms of mountains. This 'expressing' or 'voicing of the Way' does not stand for or represent something other than itself, and language used representationally is not privileged to express reality. The limitation placed on the power of representational language here is accompanied by an expanded meaning of 'expression' or 'voicing of the Way' that includes both linguistic and nonlinguistic forms.

2 Doctrine and truth

When early readers began to understand Buddhist texts as expressions of doctrines, and implicitly to grasp written language as signs signifying ideas, they came to understand the world in terms of formerly unseen relations. The Buddhist categories they applied were Chinese translations of Sanskrit terms, and this usage remained predominant even after modern Japanese scholars turned to Indian sources. The categories included names for the basic factors or elements of existence and the relations of cause and effect, conditioned and unconditioned, part and whole, relative and absolute.

The first systematic study of these categories and relations occurred in the so-called six schools of Nara, during the historical period of the same name (710-81). These 'schools' were more bodies of doctrine than independent sects with their own adherents. Often the doctrines were studied together in the same locale, such as the great temple Tōdaiji. The Japanese version of the Indian Abhidharmakośa school stressed the analysis of some seventy five conditioned and unconditioned constituents (*dharmas*) of existence and some ten types of direct and indirect causes that give rise to the conditioned factors.

As abstract as it appears, this analysis was intended as a kind of psychological examination to remove obstacles on the path to liberation. The Jōjitsu (Establishment of Truth) school proceeded to show the emptiness, or lack of substantial reality, of all constituents. The Ritsu (Precept) school conveyed very specific ethical and administrative rules to regulate the ordination and life of monks and nuns, and taught that the observance of precepts was one purification necessary for enlightenment. As discussed below, however, it did not provide an ethical theory or general principles for determining good and evil (see §6).

The Sanron (Three Treatises) school derived from Indian Mādhyamika philosophy (see Buddhism, Mādhyamika: India and Tibet). It practiced the refutation of all claims of the existence or non-existence, truth or falsity, of separate elements and doctrines. The discursive language used to refute all views ultimately pointed to its own limit, beyond which lay absolute truth. The Hossōschool formed the scholastic counterpart to Indian Yogācāra or

'consciousness-only' philosophy (see Buddhism, Yogācāra school of). It criticized the Mādhyamika tendency to absolutize emptiness, explained the role of consciousness in the appearance of self and world, and taught that enlightenment requires a conversion of the deep unconsciousness, called the 'storehouse' consciousness. It rejected the teaching that enlightenment is equally attainable by all. The sixth school, the Kegon (Flower Garland), was historically no more influential than these last two, but introduced a principle crucial to an understanding of the treatment of doctrines and truth in subsequent Japanese Buddhism.

The Nara schools took over doctrines and analyses from their Chinese counterparts virtually without change. These doctrines were partially incompatible with each other, but the discrepancies seemed little cause for dispute. Various doctrines and categories show up in later Japanese Buddhist philosophy, often as a foil for more innovative ideas. An example is the way the Abhidharma breakdown of reality and time forms the springboard of one chapter of Dōgen's *Shōbōgenzō*, 'Ocean Reflecting Samādhi'. This work first accepts the Abhidharma analysis of the causal arising of all elements but then disrupts it to proclaim the totality and sufficiency of each moment. In general, the failure of scholastic doctrines to challenge later generations of Japanese Buddhist thinkers is a sign that the latter have been more at home with practical reasoning and reconciliation than with formal argumentation and adjudication of claims. Indeed, until its introduction from the West there was in Japan no formal logic, where the rules of validity are separable from the content of the argument (see Logic in Japan). This lack does not necessarily point to a deficiency in the ability to reason, however, but may have to do with a way of relating absolute and relative, universal and particular. The philosophy of the Kegon school best epitomizes that way.

The Kegon school accounted for the variety of doctrines by placing them in a scheme of increasing difficulty and comprehensiveness. The level of the teachings matches the hearer's capacity to understand. Partial teachings, in other words, are 'skillful means' designed to lead one to more difficult and comprehensive teachings. The Kegon school typifies the inclusion characteristic of much of Japanese and Chinese Buddhist thought. One Buddhist school would incorporate rather than exclude the doctrines of rival schools, but would regard them as incomplete and place them on a lower level. To be sure, the hermeneutic of inclusion was often politically motivated, intended to increase a power base by incorporating and not alienating other groups; but it also followed the dialectical principle that rival or alternative beliefs were partial truths that fit into a larger scheme.

The Kegon scriptural hermeneutic privileged its own source *sūtras* and was rivalled by a similar scheme offered in the tremendously influential *Lotus Sutra*, but it developed a philosophy of totality and interrelation crucial for understanding most Japanese Buddhist thinking. According to this philosophy, each part or constituent of reality perfectly reflects the whole, and the whole depends upon each and every constituent part. All individual parts are therefore equivalent as conditions of the whole, and at the same time they are distinct and interdependent. The implications of this philosophy of nature for a theory of truth are the key to approaching Japanese Buddhist thought. There is no whole, universal or absolute, without its manifestation in concrete, distinct and relative particulars. In the twentieth century, Nishida Kitarō reformulated this principle paradoxically: the more relative a truth is - that is, the more deeply embedded or embodied in particulars - the more absolute it is. The absolute must encompass the relative, not stand in opposition to it. In general, Japanese Buddhist philosophy developed through a kind of synecdochic argumentation that appealed not to *a priori* reasons or empirical evidence, nor simply to scriptural sources of authority, but to this mutual accommodation of relative and absolute.

3 Universal and particular

Three interrelated themes display the mutual accommodation of relative and absolute, universal and particular: the locus of liberation, its actuality and the equality of beings.

Indian Mahāyāna *sūtras* introduced the idea that liberation does not lead to a transcendent realm different from this world, but they often spoke of the 'three immeasurable eons' required for its attainment. In contrast, many seminal Buddhist thinkers in Japan developed one Chinese view of the locus of liberation: the idea of becoming a buddha in this very body (*sokushin jōbutsu*). For Saichō(767-822), the founder of the Tendai school, it meant only a partial realization of buddhahood; but by the time of Tendai scholar Annen (841-89?) it referred to buddhahood attained in this very life. Kūkai (774-835), founder of the rival Shingon school, taught that yogic practices involving body (*mudra*s or symbolic hand gestures), speech (*mantras*) and mind (mental concentration and visualizations) could unify the practitioner with the Buddha Mahāvairocana. Such practices made it possible to attain buddhahood in this very body. This doctrine, popularized by Kūkai, provided a theoretical basis for the bodily aspects of practice

and the enactment of rituals that were important especially in the esoteric traditions of Tendai and Shingon Buddhism.

In summary, the idea of becoming a buddha in this very body or life involved a temporal and a metaphysical reduction. First, it collapsed the various stages of the path to enlightenment that were so important in various Buddhist traditions, and reduced the ‘three immeasurable eons’ to one lifetime. Second, it collapsed any remnant of a difference between the physical reality of life on earth and a transcendent realm, between the transient human body and the transcendent *dharma*-body. The body we are born with is the site of realization and symbolically embodies the entire universe, the ‘truth-body’ of the Buddha. Dōgen modified the idea further and proclaimed that ‘this very mind is buddha’ (*sokushin zebutsu*) when the discursive mind was disengaged. Although Dōgen speaks of ‘mind’ and not body, no psychophysical dualism is implied here.

Other Buddhist reformers of the Kamakura period (1192-1333) qualified the idea that liberation was possible in this world or life. Hōnen and Shinran placed this possibility in the Pure Land, which they usually considered as immediately accessible if intermediate between this world and the realm of final enlightenment. Even where it was imagined as a different world, this Pure Land was not nearly so remote a goal as Indian views of *nirvāṇa*, and was sometimes in effect made immanent in this world. Ippen explicitly stated that the Pure Land was indeed identical with this world. Nichiren too emphasized this world as the locus of salvation.

The affirmation of this world in Japanese thought is often called phenomenalism (see Phenomenalism), but this description neglects the principle or absolute said to interpenetrate all phenomena. One tenet of Kegon philosophy was that ‘principles are not impeded by things’ (*rijimuge*). Tendai Buddhism stressed the Mahāyāna doctrine of the ‘identity of the phenomenal and the real’ (*genshōzoku jissō*), and proposed that the phenomena of the world are aspects of Buddha. Dōgen wrote that the ‘true reality’ (*jissō*) is in fact all things, and that birth and death (impermanence) (see Mujō) is Buddha or *nirvāṇa*. To understand these developments in Japanese Buddhism, however, one needs to consider the question of the actuality of liberation.

4 Universal and particular (cont.)

Many Indian Mahāyāna *sūtras* teach that all sentient beings possess the nature of Buddha. The *Lotus Sutra* proclaims further that the historical Buddha is not a man who attained awakening but rather is a manifestation of the universal buddhahood that is available to all. It rejected schemes that differentiated levels of potential in beings and that excluded the lowest level from eventual enlightenment. Saichōadvocated universal buddhahood, based on the language of the *Lotus Sutra*. Universal buddhahood and ‘becoming Buddha in this very body’ are ideas based on the notion of *hongaku*, the Chinese doctrine that all beings are ‘originally or inherently awakened’. Inherent enlightenment, in contrast to ‘acquired enlightenment’, is timeless and independent of spiritual development. For some contemporary scholars such as Tamura Yoshirō, the *hongaku* idea is definitive not only of Buddhism in medieval Japan but of Japanese culture in general, since it underlies the ideals of equality, harmony and conformity (Tamura 1987).

The ideas of inherent and acquired enlightenment do not form logical contraries, since both depend upon the buddha-nature in all beings, but rather pose a practical question: how is this innate enlightenment related to practice? Legend has it that Dōgen encountered this question in his boyhood Tendai training and made it an existential problem: why undergo rigorous practice when we are endowed with ‘*dharma* nature’ by birth? Dōgen’s answer is thought to be a criticism of the *hongaku* idea, but can also be seen as its limit. He proposed ‘the unity of practice and realization’: ‘The *dharma* [truth] is not manifested unless one practices; it is not attained unless there is realization’ (*Shōbōgenzōbendōwa*). The practice of concentrated mind-body, epitomized in *zazen* or seated meditation, is a way to actualize our inherent nature. In §1 above, Dōgen’s words were quoted: ‘As for the Buddha Way, not to voice it is impossible… ’ and this sentence concludes: ‘not to study [practice] it is remote.’ Dōgen thus collapses the distinction between practice as means and enlightenment as end. Instead, *zazen*, the practice of awakening, becomes the manifestation of enlightenment.

Like Dōgen, other Kamakura Period Buddhist reformers reacted against the complacency encouraged by the idea of original enlightenment, but modified rather than rejected it. For Hōnen and Shinran the only ‘practice’ that counts is the *nembutsu* or invocation of the Buddha Amida. In his form as the Bodhisattva Dharmakara, Amida enunciated a primal vow that made his enlightenment contingent on that of all sentient beings. Those with sincere

faith in Amida’s vow were assured of rebirth in the Pure Land, where final enlightenment will be possible. In the original story, countless lifetimes elapse before Dharmakara becomes Amida Buddha, but the logic of the conditions may suggest a non-dual, timeless relation: since Amida Buddha is indeed here to save us, our liberation need only be actualized in this moment. Some writings of Shinran explicitly suggest that to recollect Amida Buddha with a believing mind (*shinjin*) is already to actualize buddhahood, without waiting to be reborn in the Pure Land. Moreover, this entrusting invocation is strictly speaking not a practice at all; that is, it is not accomplished through one own efforts or ‘self-power’ but rather is a gift of Amida, a result of Amida’s ‘other-power’. Ippen stated that ‘the *nembutsu* is what recites the *nembutsu*’. Although faith in personal effort is considered futile, as in some interpretations of inherent enlightenment, true reality must be activated by personal faith. This again is a variation on the theme of the actualization of reality (see Faith; Salvation).

5 Universal and particular (cont.)

The idea of universal buddhahood supported both syncretistic tendencies in Japanese Buddhism and the equality of all beings. The first undermined any ultimate difference between Buddhist and Shintōobjects of veneration, and the second any difference between the human and natural worlds, or what we today call culture and nature.

Syncretistic tendencies were formalized in the theory of *honji suijaku*. This theory regarded indigenous Shintōdeities (*kami*) as trace manifestations (*suijaku*) of buddhas and *bodhisattva*s, who were their original ground or true nature (*honji*). The theory appears to establish a hierarchical relationship, with the *bodhisattva*s as more basic and important than the *kami*, and thus Buddhism as more fundamental than Shintō. This interpretation implies that the Shintōgods cannot exist without the Buddhist ideal beings. It not only reverses the historical order of their appearance in Japan; it misplaces the emphasis. In fact, the *honj-suijaku* theory often worked to set the two realms and their objects of veneration on a par with one another. If a *kami* functioned as a trace or attribute of a bodhisattva, the particular *bodhisattva* was best manifested in the *kami*. The reality of the one depended upon that of the other. This interdependence was explicit in a Tendai version of *honji suijaku* theory according to which the true nature could only be perceived in its manifestations. A fourteenth-century Shintōschool simply reversed the order of *bodhisattva* and *kami*. Moreover, specific Buddhist and Shintōideal beings are sometimes identified with one another and then contrasted with other pairs. In the twentieth century, a ‘new religion’ called Gedatsukai in effect identifies the Buddha Mahāvairocana and the Shintōdeity Tenjinchigi as *honji*, and gives other Buddhist-Shintōdeities the status of *suijaku*. These examples illustrate that the accessibility of ideal beings was more important than any rank order, which was historically variable (see Shintō).

The theory of original ground and trace manifestation was developed in the late Heian Period (794-1185), contemporaneous with the idea of inherent enlightenment. Yet the underlying association of Shintōand Buddhism is evident much earlier and lasted until much later. At the very beginning of Japanese Buddhism, the *Nihon shoki (Chronicles of Japan)* report that the Emperor Yōmei (d. 587) ‘believed in the Law of the Buddha and reverenced the Way of the Gods’. Yet the name ‘Way of the Gods’ (Shintō) was a Chinese, quasi-Buddhist concept. In effect, Buddhism prevailed and was established as the state religion, at least temporarily, through the efforts of Yōmei’s sister, Empress Suiko, and especially of his son, Prince Shōtoku, who wrote commentaries on *sūtras* and built temples (see Shōtoku Constitution). In the Nara Period (710-81), Shintōshrines were incorporated into Buddhist temples in order to protect land interests, and Buddhist altars built near shrines ensured the protection of Shintō. At one locale the god of war, Hachiman, was named the Bodhisattva Daijizaiten in 783; at others he was associated with different *bodhisattva*s. In the ninth century, Kūkai proposed a ‘dual aspect’ theory that, for example, identified the sun goddess Amaterasu with Mahāvairocana, the Buddha of Great Illumination, whom Kūkai further identified with the *dharmakāya* or truth-body of Buddha. Although Tokugawa Period (1600-1868) nativists such as Motoori Norinaga reasserted the superiority of indigenous Shintō, the Buddhist inclusion of Shintōwas not effectively challenged on a national scale until the early Meiji government’s policies (in the 1870s) of ‘separating gods and buddhas’, persecuting Buddhism and establishing State Shintō. Religion and state have had separate legal status since the Second World War, but Buddhist and Shintōteachings and customs still coexist in Japanese family life.

Japanese Buddhist thought extended the idea of universal buddhahood to the equality of all beings. In particular, the idea of inherent enlightenment lent support to the doctrine of the Buddha-nature of non-sentient beings that is often alluded to in medieval Japanese literature. It was both Tendai and Nichiren doctrine that the grass and trees -

in other words, beings not possessed of a sensitive mind - can become buddhas: ‘Grasses, trees, mountains and rivers all attain buddhahood’ became a frequent saying in medieval Buddhist texts. A chapter of Dōgen’s *Shōbōgenzō* quotes the enlightenment verse of Chinese poet Su Dongpo:

> The sounds of the valley streams are His long, broad tongue;
> The forms of the mountains are His pure body.
> At night I heard the myriad *sūtra*-verses uttered
> How can I relate to others what they say?
>
> (Cook 1989: vii)

For Dōgen, this verse provides one answer to the questions, ‘where is buddha?’, ‘what is buddha?’ and ‘how does he preach?’ The preaching of nonsentient beings, a frequent theme in Dōgen, is echoed earlier in the poetry of Saigyō. It expands Dōgen’s view of ‘expression’ (discussed in §1) to the world of nature. Whether the traditional Japanese love of nature preceded or succeeded the absolute significance accorded to the phenomenal world cannot be determined.

6 Theodicy and ethics

Japanese Buddhist treatments of theodicy and ethics also reveal ways in which Buddhism was transformed in Japan. Theodicy is meant here in Max Weber’s general sense of an explanation that make the injustices of life intelligible and accounts for evil in the world (see Weber, M.). In classical Indian thought, the notion of *karma* answers the problem of both moral and natural evil. Karmic retribution means that every deed or action has consequences and evil deeds bring about suffering. Suffering is not punishment by a God who is free to punish or not, but rather is an inevitable consequence of evil actions. There are accordingly practical reasons to avoid evil and do good. In early Indian thought the notion of *karma* entailed that of transmigration. *Karma* is, as it were, the momentum of our actions that propels us through *saṃsāra*, the continuous cycles of birth and death. This transmigration occurs through various lifetimes and life forms, mythologized as the six realms of hell, hungry ghosts, animals, fighting demons, humans and gods. The Buddha taught a way of liberation that is a release from transmigration and its cause, *karma*, whether good or evil (see Karma and rebirth, Indian conceptions of). In later Japanese Buddhist philosophy, attachment to self and the concomitant ignorance of the reality of no-self were stressed as the root cause of *saṃsāra*.

After Buddhism was introduced to Japan, people easily accepted the notion of karmic retribution but not the literal belief of rebirth as animals or lower life forms. Transmigration through the six realms of *saṃsāra* played a larger role in Japanese literature and theatre than in philosophical discussions, which focused more on the practical path to liberation. Good deeds were not the means to liberation; they did not, for example, ensure rebirth in the Pure Land, whether taken as the physical abode or spiritual state wherein final liberation is possible. Hōnen, and later Shinran in his *Tannishō(Lamenting the Deviations)* went so far as to state that: ‘If a good person attains birth in the Pure Land, how much more so the evil person’.

The apparent incompatibility between any notion of rebirth and the Indian Buddhist doctrine of no-self or no-soul was not an issue in traditional Japanese Buddhist thought. Moreover, even after the Japanese understood this doctrine they, like the Chinese, continued to believe in the spirits of ancestors who remain somehow present among the living. That folk belief has played a stronger role in everyday life and religious practices than has Buddhist doctrine. Logically, the disassociation of good deeds and liberation would seem to compromise the theodicy offered by the notion of karmic retribution, and ancestor veneration to compromise the doctrine of no substantial self which is the basis of compassion with all living beings. In effect, however, the compromises have meant little more than finding a middle course, or superimposing the two metaphysically incompatible beliefs. Belief in karmic retribution has encouraged the avoidance of deeds conventionally deemed evil, and thus supported conventional ethics; but karmic retribution does not entail any final resolution to misery. Only religious practice, whether conceived as self-power or other-power, holds out the prospect of final liberation, or at least the actualization of one’s inherent enlightenment. In the meantime, one is to do good and avoid evil; more specifically, one is to adhere to conventional Buddhist precepts such as not killing, stealing, lying and so on. Similarly, veneration of ancestral spirits, including Buddhist patriarchs, has encouraged a respect for tradition and social order, while non-ego has been extolled as the basis of right action in both Japanese Buddhism and Confucianism.

If we seek in Japanese Buddhist philosophy a coherent theory of ethics as a separate branch of philosophy, we find instead a variety of practical reasons; for a moral theory, we would have to refer to Confucianism (see Confucian philosophy, Japanese).

This general description receives closer definition in the views of Shinran and Dōgen. Dōgen's ethical views reflect both the penchant to practical reasoning and the transcendence of ethics in the realization of enlightenment. Dōgen's sermons to monks training under him admonish them to keep the precepts while recognizing that their content is relative to the situation. Ultimately, the practice of keeping precepts is subsumed into *zazen* or the practice of realization. Dōgen's more philosophical writings transform practical admonitions such as 'do good' and 'do no evil' into proclamations of the realized state: 'the non-production of evil, the performance of good'. Ultimately, both Dōgen and Shinran would have us transcend the duality of good and evil to manifest, through uncontrived actions, absolute non-dual reality or 'suchness', which is perfect as it is. There is no cause-effect relation between morality and enlightenment, and so the issue is not the situational relativity of morality but, once again, the problem that the absolute is not complete without its manifestation in the phenomenal. Here ethical issues would give way to the problem of inherent enlightenment. The contemporary interpreter D.T. Suzuki advocated Zen enlightenment as the solution and simply proclaimed it to be 'beyond good and evil' (Suzuki 1964). On the other hand, the idea of inherent enlightenment has come under attack in the 1990s as perpetuating the status quo of society and blinding Japanese Buddhism to the need for reform guided by a social ethics.

See also: Aesthetics, Japanese; Buddhist philosophy, Chinese; Buddhist philosophy, Indian; Confucian philosophy, Japanese; Dōgen; Kūkai; Kyoto school; Mujō; Nichiren; Shinran; Shintō

JOHN C. MARALDO

References and further reading

Cook, F. (1989) *Sounds of Valley Streams*, Albany, NY: State University of New York Press.(An introduction to enlightenment in Dōgen's Zen, with a translation of nine essays from the *Shōbōgenzō*.)

Dobbins. J. (1989) *Jōdo Shinshū: Shin Buddhism in Medieval Japan*, Bloomington, IN: Indiana University Press. (A systematic treatment of the doctrines and major figures, such as Hōnen and Shinran, of the Pure Land school.)

Groner, P. (1984) *Saichō: The Establishment of the Japanese Tendai School*, Berkeley, CA: Berkeley Buddhist Studies Series.(A systematic study of the founder and teachings of the Tendai school, including the doctrine of attaining Buddhahood in this very existence.)

Hubbard, J. and Swanson, P. (eds) (1997) *Pruning the Bodhi Tree: The Storm Over Critical Buddhism*, Honolulu, HI: University of Hawaii Press.(Translations of articles by two major critics of the legacy of inherent enlightenment, N. Hakayama and S. Matsumoto, with responses from differing perspectives.)

Hakeda, Y. (1972) *Kūkai: Major Works*, New York: Columbia University Press.(An introduction to and translations of the seminal writings of Kūkai, the founder of the Shingon or 'Truth-word' school of esoteric Buddhism.)

Kitagawa, J. (1987) *On Understanding Japanese Religion*, Princeton, NJ: Princeton University Press.(A collection of essays by a historian of religions that treats both specific topics and the general character of Japanese Buddhism in a manner accessible to the general reader.)

Kiyota, M. (ed.) (1987) *Japanese Buddhism: Its Tradition, New Religions and Interaction with Christianity*, Los Angeles, CA and Tokyo: Buddhist Books International.(Essays in the first section elaborate the doctrines of the universal and the particular, and of rebirth and transmigration.)

LaFleur, W. (1983) *The Karma of Words: Buddhism and the Literary Arts in Medieval Japan*, Berkeley, CA: University of California Press.(Presents the Buddhist philosophical background to Japanese literary works, often in contrast to Western, especially Platonic thinking.)

Matsunaga, A. (1969) *The Buddhist Philosophy of Assimilation*, Tokyo: Sophia University and Charles Tuttle Co. (Presents the *honji suijaku* theory.)

Matsunaga, D. and Matsunaga, A. (1974) *Foundations of Japanese Buddhism*, Los Angeles, CA and Tokyo: Buddhist Books International, 2 vols.(A detailed historical survey from the beginnings through the Muromachi Period, 1338-1573.)

Nakamura, H. (1964) *Ways of Thinking of Eastern Peoples*, Honolulu, HI: University of Hawaii Press.(Expounds the acceptance of the phenomenal world as absolute, the social matrix of values and the emphasis of symbolic expression over logical thinking.)

Suzuki, D.T. (1964) *An Introduction to Zen Buddhism*, New York: Grove Press.(Essays by the most influential

advocate of the distinctiveness of Zen Buddhism.)

Tamura Yoshirō (1987) 'Japanese Culture and the Tendai Concept of Original Enlightenment', *Japanese Journal of Religious Studies* 14: 203-10.(A brief argument for the positive cultural significance of the *hongaku* idea of inherent enlightenment.)

Unno, T. (trans.) (1984) *Tannisho: A Shin Buddhist Classic*, Honolulu, HI: Buddhist Study Center Press.(A translation of Shinran's most influential writing, 'Lamenting the Deviations [from the true teaching]', and an essay on the central doctrines of the True Pure Land school.)

Watanabe, S. (1970) *Japanese Buddhism: A Critical Appraisal*, Tokyo: Kokusai Bunka Shinkokai.(A realistic and detailed but non-technical treatment of the characteristics of Japanese Buddhism.)

Buddhist philosophy, Korean

Buddhism was transmitted to the Korean peninsula from China in the middle of the fourth century AD. Korea at this time was divided into three kingdoms: Kokuryô, Paekche and Silla. Both Kokuryô and Paekche accepted Buddhism as a state religion immediately after it was introduced, to Kokuryô in 372 AD and to Paekche in 384 AD. However, it was not until two centuries later that Silla accepted Buddhism as a state religion. This was because Silla was the last of the three kingdoms to become established as a centralized power under the authority of one king.

It is not coincidental that Buddhism was accepted by these three states at the very same time that a strong kingship, independent of the aristocracy, was created. These newly established kingships needed a new ideology with which to rule, separate from the age-old shamanistic tradition which had been honored among the previous loose confederations of tribes. Buddhism fulfilled this need. It became a highly valued tool which kings used shrewdly, not only to provide their societies with a political ideology but to give them a foundation from which to build a viable system of ethics and philosophical thinking. Given this historical legacy, Korean Buddhism came to possess a feature which set it apart from the other East Asian traditions: it became 'state-protection' Buddhism. Although this was not a particularly sophisticated phenomenon on a philosophical level, this feature had a lasting influence on all aspects of Buddhist thought in Korea. In general, Korean Buddhism has followed a course of development more or less parallel to that of the greater East Asian context, although with notably closer ties to China than to Japan. There is no historical evidence which indicates any direct intellectual transmission from India, Buddhism's birthplace; rather, most of the philosophical development of Buddhism in Korea occurred as Korean monks travelled to China to study and obtain Buddhist texts which had either been written in or translated into Chinese. Despite such close ties to China, however, Korean Buddhism has developed its own identity, distinct from that of its progenitor.

Compared to Indian and Central Asian Buddhism, which developed along clear historical lines, the development of Buddhism in China was largely dependent on the personalities of individual monks, and was thus affected by such factors as their region of origination and the particular texts which they emphasized. Thus, in the process of assimilating Indian Buddhism, the Chinese created and developed a number of widely varying schools of Buddhist thought. In Korea, however, such a diverse number of philosophical traditions was never established. Rather, one of the distinct features of Korean Buddhism has been its preference for incorporating many different perspectives into a single, cohesive body of thought.

1 Three Kingdoms period (372-668 AD)

When Buddhism arrived on the Korean peninsula, Korea was in a period of transition, changing from loose tribal societies to states ruled by centralized kingships. Buddhism was strongly patronized by the kings, as it provided them with a valuable political ideology, but nevertheless it had to face the challenge of countering the indigenous belief system based firmly on shamanistic ritualism. In response to this challenge, certain Buddhist doctrines which tended to undermine animistic beliefs received special emphasis in Korea. An example of these is the doctrine of *karma*, which asserts the importance of ethical action by humans and denies the power of gods and spirits in shaping human destiny. However, Buddhism also resorted to the tactic of assimilation. The heavenly being Hanunim, a shamanistic deity, was absorbed into the Buddhist conception of heaven. Similarly, the indigenous shamanistic worship of mountains, earth deities and dragons all became integrated into the Buddhist religious system.

Only fragmentary bits of information are available concerning Buddhist philosophy during its early existence in Korea. As previously discussed, the development of Buddhist studies in Korea ran almost parallel to that in China. Both Mahāyāna schools, such as Tientai, Huayan and Dilun, as well as Hīnayāna schools, including Abhidharma and the Sarvāstivāda and Vinaya traditions, were introduced. However, it seems that Mādhyamika philosophy formed the mainstream of Buddhist doctrinal thought during this period. Names of Korean monks from the Kokuryô and Paekche kingdoms are often mentioned, particularly in Japanese records, as masters of Mādhyamika theory or as founders of the Sanron (in Chinese, Sanlun) school in Japan (the Sanlun school was based on the sinicized interpretation of the Indian Mādhyamika).

We can learn much about the Korean interpretation of Mādhyamika theory by investigating the contributions of Sûngnang, an eminent Korean monk. Although we have no exact record of his date of birth, according to the Biography of Fadu in the *Liang Gaosengzhuan (The Liang Biographies of the Eminent Monks)*, Sûngnang succeeded to the lineage after the Master Fadu passed away in 500 AD. Sûngnang's unique interpretations represented a turning point not only for Sanlun studies but for Chinese Buddhist studies as a whole. Through the decisive contributions of Sûngnang, Sanlun was established by Jizang (549-623 AD) as one of the major schools in Chinese Buddhist history. Sûngnang's own works have not been transmitted, but his philosophical ideas are cited throughout Jizang's writings.

During Sûngnang's time, Chinese Buddhist studies were divided into Northern/Abhidharma and Southern/Satyasiddhi. Sanlun was considered part of the Satyasiddhi school, whose doctrine was based on Abhidharma theory and centered around the Hīnayāna concept of the Four Noble Truths. Sûngnang, however, made a distinction between Satyasiddhi and Sanlun, and he demarcated a boundary between the Old and New Sanlun. He is best known for his two-truth theory, in which he distinguished between conventional and ultimate truth. In his view, these are not ontological truths, but represent rather a convenient pedagogical device for teachers. In discussions concerning truth, some say that it is existence, some say that it is nonexistence, others say that it is either existence or nonexistence, while still others say that it is neither existence nor nonexistence. Sûngnang claimed, however, that none of these statements represent the complete truth, but merely indicate one aspect of it. Thus, in his view, the two truths represent different expressions of the excellent teaching of the Middle Path.

For Sûngnang, the Middle Path was not a state situated between existence and nonexistence, or duality and non-duality. Rather, he saw it as consisting of three levels: on the first level, the existence of all beings is considered as conventional truth, and the emptiness of all things is regarded as ultimate truth. Here, emptiness is called ultimate truth in response to the followers of Abhidharma who advocated the existence of *dharmas*. On the second level, the existence and emptiness of the first level are both viewed as conventional truth while the denial of both is seen as ultimate truth. This is Sûngnang's rebuttal to the Satyasiddhi school, which admits the dualism of existence and emptiness, positing emptiness as ultimate truth. Finally, on the third and last level, it is recognized that both truths as well as all three levels, are merely teaching devices. Thus, the three gates or three levels are expounded only in order to realize the truth of not three. This non-acquisition of, and non-abiding in, one level or one truth is initially named the ultimate.

The structure of Sûngnang's theory indicates a process of continuing dialectics to a level of infinity. As such, it negates any fixed concepts which regard the ultimate as a state arrived at through progression. For Sûngnang and his followers, a highest level does not exist; rather, this dialectic method of denial continues endlessly. The ultimate truth, then, cannot be the ultimate; it is not an absolute reality. This is why Sûngnang asserted that the two-truth theory was not a principle but only a pedagogical tool. He showed clearly that it is not possible to verbalize or conceptualize about ultimate truth. Any attempt to do so will only bring one back to the level of conventional truth. This unique philosophy marked the beginning of the New Sanlun school and was elaborated on first by Zhouyong and then by Jizang, who finalized its ideas in the latter half of the sixth century.

2 Unified Silla period (668-935 AD)

This period was one of religious maturity and innovation; as such, it represented the culmination of Buddhist influence in Korea, both culturally and doctrinally. Buddhism fully blossomed and flourished at this time, reaching heights that it would never see again. On the one hand, influences from China were still felt, as new schools of thought were introduced by Silla monks returning from the Tang dynasty. On the other hand, Korean Buddhist scholarship by such eminent monks as Wônhyo (617-86) played an important role in the development of Chinese Buddhist thought as well. However, unlike the Tang tendency towards sectarianism, Unified Silla leaned towards a synthesis of various aspects of Buddhist thought into a type of interdenominational philosophy, the so-called *t'ong pulgyo*, or Buddhism of total interpenetration. The syncretic approach of Korean Buddhist philosophy can be best understood under the following headings: the reconciliation of doctrinal disputes, the Silla Hwaom School, the Silla Vijñānavāda School, and the introduction of Sôn (Chan) Buddhism from China.

3 Unified Silla period: reconciliation of doctrinal disputes

One of the most important proponents of this syncretic approach was the noted monk and scholar Wônhyo. In his view, the various disputes which arose among Buddhist scholars were similar to the story of the blind men's description of an elephant: each man was only defining one aspect of the animal. In one sense, they were all wrong, because each failed to fully describe the elephant. In another sense, however, they were not entirely wrong, as what they were describing was not something other than the elephant.

Wônhyo was not interested in merely partial or incomplete understandings, however; rather, he focused on integrating a variety of doctrines into one complete view, as seen from a higher perspective. This view has been termed *hwajaeng*, or the 'reconciliation of doctrinal controversy'. The hermeneutical device used in this doctrinal reconciliation is well represented in his unique genre of Buddhist commentarial work, *chong-yo*, as well as in his concepts of *kae* (unfolding or opening) and *hap* (folding or sealing) the truth. Wônhyo is believed to have composed approximately seventeen of these *chong-yo*, of which only five are extant. *Chong* translates as doctrine or theme, and *yo* means essence. Thus, for example, one of his works, the *Yolban'gyong chongyo*, translates as the 'Doctrinal Essentials of the Nirvāṇa Sutra'. Here, 'doctrine' refers to the various ideas which are discussed, and 'essentials' integrates them into one primary framework of thought. Similarly, by his usage of the word 'unfolding' (*kae*), Wônhyo indicates a variety of ideas and analyses, while through 'folding' (*hap*) he denotes how these themes may be synthesized into one cohesive unit of understanding. Doctrine, then, reflects the unfolding of the one into the many, whereas essence refers to the folding of the many back into the one.

Another device which Wônhyo used in order to reconcile opposites and thus denote their non-duality was his theory of essence-function (Korean: *ch'e-yong*) (see Ti and yong). In his *Taesung kisillon so (Treatise on Awakening Mahāyāna Faith)* he asserts that Mahāyāna is a synonym for One-Mind. This One-Mind represents essence, or *ch'e*, while the variety of mental and physical states, or the arising of *dharmas*, represents function, or *yong*. In terms of its internal logic, essence-function is no different from the *kae-hap* (folding and unfolding) formulation. According to Wônhyo, when the essence of Mahāyāna, or One-Mind, is unfolded, immeasurable and limitless meaning can be found in its doctrine. Conversely, when the doctrine is folded, it returns to its essence, which consists of two aspects: the absolute and the phenomenal. The limitless meanings of One-Mind are identical to this essence and are completely amalgamated within it. Therefore, One-Mind unfolds and folds freely, and establishes and refutes without restrictions.

In this way, the myriads of sensory impressions are brought to a permanent end and return to the source of One-Mind. Thus, there is nothing that is not advocated and nothing that is not refuted. Wônhyo calls the advocacy concession or acceptance (*hwanho*), and he labels the refutation variously as deprivation (*t'al*), negation (*kyon*), return (*wang*) or prohibition (*purho*).

Wônhyo was particularly concerned with using his hermeneutic tools to reconcile the Mādhyamika and Yogācāra doctrines. In his view, Nāgārjuna's *Mūlamadhyamakaśāstra* and *Dvādaśanikāyaśāstra* offer a thorough dialectic critique of all views which a person might advocate, to the extent that they abolish both the views refuted and the act of refutation (see Nāgārjuna). However, because Mādhyamika dialectics do not affirmatively acknowledge both the subject and object of this refutation, the approach is purely negative, and thus nothing is established. On the other hand, the Yogācāra school accomplishes a thorough analysis of all mental states and accordingly establishes both the shallow and profound teachings. However, because the Yogācārins do not continue on to refute what they have established, theirs is a purely positive approach, and thus, nothing is refuted. Doctrinal controversy arose between the two schools due to the fact that the Mādhyamikas attached to refutation and deprivation, whereas the Yogācārins attached to establishment and acceptance. Each recognized only a part of the truth; although neither was entirely wrong, at the same time neither was altogether correct.

Wônhyo was able, based on his unique usage of unfolding and folding, to reconcile these differences by including them into a single Buddhist teaching, that of One-Mind, which allows for both affirmation and negation. According to Wônhyo, the *Taesung kisillon*, a noted Mahāyāna text upon which he based his well-known commentary, is a syncretic work which embraces the positions of both Mādhyamika and Yogācāra (see Awakening of faith in Mahāyāna). He asserts that while it refutes all relative views (as represented by the two above-mentioned schools), it also embraces them comprehensively within the all-encompassing One-Mind.

4 Unified Silla period: Silla Hwaôm school

Another vital contribution to Korean Buddhist thought was made by Ûisang (625-702), who was a contemporary of Wônhyo and is known as the founder of the Korean Hwaôm school (Huayan in Chinese) based on the *Avatāmsaka Sūtra*. Unlike Wônhyo, who had never been to China, Ûisang spent almost ten years there and studied under Zhiyan (602-68), the second patriarch of the Chinese Huayan school. The *Hwaôm ilsung popke to (Diagram of the Dharmadhātu of the One Vehicle of Hwaôm)* is Ûisang's only major extant work, and is comprised of his *Haein sammae to (Diagram of Ocean Seal Samādhi)* and an autocommentary.

The *Ocean Seal*, as it is commonly known, written while Ûisang was in China, is a poem consisting of 210 characters. It opens with the word *pop* (*dharma*) and closes with *pul* (Buddha), thus signifying the cause-and-effect relationship between the two. By placing both of these characters in the centre of a diagram in the form of a maze, Ûisang articulated the basic theme of Hwaom philosophy which asserts that beginning and end, or cause and effect, occupy the same position while still retaining their own distinctive characteristics. In this way, he depicted Hwaôm philosophy as a mystery of simultaneous mutual penetration. In addition, he pointed to the Hwaôm soteriological position that within the fifty-two stages of a *bodhisattva*'s career, the first stage of initial faith itself embodies all fifty-two stages, including the final stage of marvellous enlightenment. Of this position, a well-known Hwaôm aphorism states, 'The moment one arouses an enlightenment thought (*bodhicitta*), instantly perfect enlightenment is attained.'

Although Ûisang's expression of Hwaôm theory in the *Ocean Seal* was not his own creation, it succeeded in ably representing the main ideas of the school, both symbolically and graphically. The *Ocean Seal* is traditionally presented to monks upon completion of their course of study, a mark of how highly esteemed it is in the Korean Buddhist tradition. It is also chanted in Korean Buddhist ceremonies as a *dhāranī*, having been endowed with a special mantric power. Along with Wônhyo's philosophy of the reconciliation of opposites, Ûisang's contributions to Hwaôm theory helped to create the Korean tradition of *t'ong pulgyo* and provided the foundation for the synthesis of the doctrinal or scriptural schools (Kyo) and the meditational schools (Sôn) by Chinul in the following Koryô dynasty.

5 Unified Silla period: Silla Vijñānavāda school

There were actually three different systems of Yogācāra (or Mind-Only) philosophy which were transmitted into Korea from China during the Unified Silla period. The first can actually be traced back to the Three Kingdoms period, when Won'gwang (d. 630) returned to Korea after studying the *Mahāyānasanùgraha* (in Chinese, *Shelun*) in China. However, it was not until Wônch'ūk (612-96) arrived on the scene that any significant study of Yogācāra was undertaken in Korea. Wônch'ūk is considered the major figure among Silla monks studying Yogācāra doctrine. His system is classified as the Old Yogācāra, founded by the Indian monk Paramārtha (499-569). This is in order to distinguish it from the New Yogācāra, which was founded by the Chinese monk Xuanzang (602-64). Xuanzang translated Dharmapāla's *Vijñāptimātratāsiddhi* into Chinese and his new school, which was established on the basis of this text, became known as the Faxiang (Dharma Characteristics) school.

The two Yogācāra schools, Old and New, differed in their theoretical approaches and there were conflicts between them. Wônch'ūk and his disciples eventually established their own school, called the Ximing school. This was followed by the Silla Vijñānavāda school, which included such noted monks as Tojung and T'aehyon. Vijñānavāda theory in Silla developed its own unique characteristics quite distinct from those of Xuanzang's school in China. The Silla school not only effected a synthesis of the various Vijñānavāda schools in China, but it also adopted theories from other Chinese schools, such as Tientai and Huayan. It achieved this synthesis through its understanding of the term 'One Vehicle', an approach made possible by the work of Wônhyo. Like Wônhyo, the Silla Vijñānavāda school attempted to reconcile the teaching of Mādhyamika with that of Yogācāra, but by using the concept of one sound as its reference point. Here, Wônch'ūk was referring to the Buddha's teaching, which has one meaning but is given different explanations depending on one's level of understanding. Wônch'ūk's concept of 'one sound' is, of course, comparable to the concept of the One Vehicle, as both are nothing but teaching devices used to comprehend various understandings from a higher vantage point.

6 Unified Silla period: introduction of Sôn

A new epoch occurred in Korean Buddhism upon the transmission of Sôn (Chan) Buddhism from China. This new movement, with its strong emphasis on meditational practice, entered Korea not long after it was introduced into

China. According to traditional accounts, it was Pômnang (*fl.* 632-46) who first introduced Chan Buddhism to Korea, following his study under Daoxin (580-651), the fourth patriarch of the Chinese Chan school. His lineage was passed down through Sinhaeng, Chunbôm and Hyeûn until it reached Chisôn Tohon (824-82), who founded the Mount Huiyang school, the oldest Sôn lineage in Korea, in 879. Within a hundred-year period, from the eighth to the ninth centuries, eight other mountain schools were founded. From this time onwards, the term Nine Mountains has come to represent Korean Sôn schools in general, and this number remained unchanged during the early Koryô period.

In contrast to the Chinese Chan schools, which were categorized as the Five Schools and Seven Orders and which maintained sharp sectarian distinctions, the Korean Nine Mountain schools were classified according to the identities of their founders. Only a monk who had originally studied in China was entitled to be the founder of a new school. Furthermore, in the Korean tradition, the relation between teacher and disciple was considered more important than the philosophical doctrine espoused. This differed sharply from the Chinese Chan tradition, in which lineage affiliation was based directly on doctrine. The Korean approach was to regard what they called the Five Schools and Seven Orders not as different schools, but rather as different families of the same school. For this reason, the Nine Mountain schools, regardless of their founders, were all considered to belong to the Chogye order. The name Chogye (in Chinese, Zaoxi) was taken from the mountain associated with Huineng, who was regard as the progenitor of Chan before its division into five schools and seven orders.

Eight of Korea’s nine mountain schools (the exception was the Mount Sumi school) were derived from or related to the Chinese Hongzhou school. As noted by Zongmi (780-841), the noted Chinese Buddhist commentator and philosopher, the Hongzhou school had a close doctrinal affiliation with Huayan. Most of the Korean monks who came to China to study Buddhism were already familiar with Hwaôm philosophy, as it was the predominant school in Korea at that time. Therefore it is not surprising that they were attracted to the Hongzhou school. However, the fundamental difference between Huayan and Chan must not be overlooked: while the former emphasizes scriptural study, the latter rejects it, choosing rather to espouse ‘a separate transmission outside the scriptures’. This crucial difference may also have attracted the Silla monks, who were perhaps frustrated with the philosophically elaborate and priest-centred Hwaôm school.

Immediately upon Sôn’s introduction to Unified Silla, an antagonistic tension was created between the new Sôn and the old doctrinal schools, particularly Hwaôm. Sôn masters perceived the scripture-based Hwaôm school as antithetical to their practice-oriented mission. Later, however, they altered their approach in an effort to synthesize the messages of the two schools.

7 Koryô period (918-1392): parallel cultivation of doctrinal study and contemplation

Upon the collapse of the Silla dynasty, the Korean peninsula reverted back to three kingdoms, until King T’aejo unified the country and subsequently founded the Koryô dynasty. Under the Koryô, Buddhism was fully supported not only by the royal family but by the entire apparatus of the Korean government. The price, however, was tight government control.

Despite these controls, discord between the radical Sôn schools and the more conservative doctrinal proponents continued, and it was not till Ûich’ôn (1055-1101) that a serious attempt was made to reconcile these conflicts. Although his background was in Hwaôm, Ûich’ôn studied various schools of Buddhist thought throughout his career. Most notably, he travelled to China in 1085 and studied with renowned teachers of the Huayan, Tientai, Pure Land, Vinaya and Chan sects there. Upon returning to Korea, he attempted to incorporate Sôn into the system of Tientai (in Korean, Ch’ônt’ae). He felt that meditation was essential not only for followers of Sôn but for members of all Buddhist schools, and he criticized the polarization of the two approaches, doctrinal and meditational. From his perspective, which favoured scriptural study yet also emphasized meditational practice, although *dharma* is devoid of words and appearances, it is not separate from them. Similarly, doctrinal study and meditation do not oppose each other but should be seen as complements; each is a requisite tool for the attainment of awakening.

Yet Ûich’ôn was unable to accept the radical doctrine of the Sôn of his time. He acknowledged the validity of Sôn only on the condition that it remained based on scriptural teachings. For him, the Sôn dictum of ‘transmission outside the scriptures’ was not acceptable as a valid Buddhist teaching. He believed that the practitioners of his

time, by abandoning the scriptures, were veering dangerously towards heresy. He felt that his own *kyogwan kyomsu*, or 'parallel cultivation of doctrinal study and contemplation', was the correct approach. Viewing the system of Ch'ônt'ae as one which harmonized Sôn and scriptural study, Ûich'ôn attempted to consolidate the Sôn school into the Ch'ônt'ae order. However, he was unable to accomplish this, thanks not only to his early death, but also to of his narrow understanding of Sôn.

It was Chinul (1158-1210) who effectively assimilated Sôn into the mainstream Korean Buddhist tradition, if not in its entirety, at least in the area of doctrine. Chinul lived during one of the most difficult times in the history of Koryô Buddhism. Having become dependent on the support of the government, Koryô Buddhism had lost most, if not all, of its earlier spiritual vitality; corruption was at its peak and discipline in the monasteries was increasingly lax. Chinul perceived the Buddhism of his day as subject to 'ten kinds of disease', and his endeavours to effect a rapprochement between the Sôn and doctrinal schools were founded on his deep concern about this decline.

If Wônhyo's philosophy of harmonization was aimed at ending the doctrinal disputes among the various schools, Chinul's intent was to provide a theoretical scheme for harmonizing Sôn with the scriptural schools, and to offer a radical approach to the issue of final enlightenment based on the doctrinal teachings of Mahāyāna. He attempted ultimately to provide a systematic, theoretical framework which would incorporate both theory and experience. Being a Sôn monk, he was deeply aware that only through one's own experience can one attain final, ultimate liberation. The primary value of his theory, then, was that it clearly displayed an understanding of the necessity of practice.

Unlike Ûich'ôn, whose viewpoint was rooted in the doctrinal teachings, Chinul's foundation was in Sôn. His syncretic vision was therefore to establish a new Sôn school which would incorporate not only various teachings of the traditional Chan schools in China, but various doctrinal teachings as well. His primary intent was to systematize a comprehensive soteriological scheme for Sôn which would be pertinent to practitioners of various levels of capacity. He was inspired in this task by the great Chinese Buddhist thinker Zongmi, a patriarch of both the Huayan and the Heze school of Chan. Chinul, deeply frustrated with the deterioration of Sôn in his time, was especially attracted by the balanced approach of the Heze school, which encompassed both the intellectual understanding of the scriptural teachings and the radical, anti-intellectual spirit of Chan. Chinul's soteriological scheme of sudden enlightenment followed by gradual self-cultivation reflects the Heze approach.

In an inscription written by Kunsu Kim (*fl.* 1216-20), it is said that Chinul established three main approaches to Sôn practice, directly reflecting his own enlightenment experiences: the balanced practice of meditation and wisdom, inspired by the *Platform Sutra* (see Platform Sutra); faith and understanding according to the complete and sudden teachings, inspired by the Li Tongxuan's *Exposition of the Huayanjing* (*Avatamsaka Sūtra*); and finally, the shortcut approach of *hwadu* (in Chinese, *huatou*) investigation. We can elucidate Chinul's philosophical system by examining these three approaches.

8 Koryô period: balanced cultivation of meditation and wisdom

In his *Susim kyol (Secrets on Cultivating the Mind)*, Chinul elucidates a method for cultivating the mind after a sudden, initial enlightenment experience. According to him, sudden enlightenment represents a realization of the nature of true mind as having two aspects: voidness and calmness in its essence, and numinous awareness in its function. These in turn correspond to the two aspects of *dharma*, immutability and adaptability. Mind is in its essence originally void and calm, and yet at the same time it adapts freely in infinite ways, depending on one's level of consciousness. Thus if one's consciousness is deluded, Mind appears ignorant, while if it is awakened, Mind manifests a numinous awareness.

In Chinul's system, gradual practice should follow after the sudden, initial awakening to the nature of True Mind. Chinul correlates the essence and function of Mind with meditation (*samādhi*) and wisdom (*prajñā*), respectively: in the task of cultivation, meditation is the essence and wisdom is the function of Mind; they are inseparable, nondual aspects of the same thing. This dynamic concept of the interrelationship of meditation and wisdom further entails that in cultivating Mind one should be alert in the void and calm mental state, and calm in the state of numinous awareness: calm and alertness represent meditation and wisdom, which are inseparable and nondual, and thus must be practiced simultaneously in cultivation.

Even if one attains a sudden awakening to the nature of Mind and thus realizes the identity of meditation and

wisdom, the balanced cultivation of meditation and wisdom is still important: due to the residue of defilements and the karmic force deriving from long-standing habits, one's mind-nature does not always have its original harmony. The two nondual, inseparable aspects of mind, essence and function, can become bifurcated, as if they were two different, separate entities. An excess of essence can produce too much calmness, which in turn leads to dullness, while an excess of function can produce over-alertness, thereby easily causing distraction.

Since Chinul follows the sudden approach to enlightenment, in which enlightenment precedes cultivation, his interpretation of meditation and *samādhi* as essence and function differs from that of the gradual school of Sôn. In the latter, where practices are centred on the gradual removal of defilements in order to achieve perfect Buddhahood, meditation is used to counter the defilements in the conditioned realm, and wisdom is used thereafter to counter discriminative thinking. These two different methods are intended to ultimately achieve a penetrating insight into the nature of reality. However, although Chinul does not exclude the possibility that the gradual approach to meditation and wisdom may be used as an expedient means for counteracting defilements, he felt it should be allowed only after the experience of sudden awakening. In fact, Chinul urges that at all levels of practice, as well as at all stages of a practitioner's development, both meditation and wisdom should be combined, just as the principle of balance, that is, of calmness and alertness, must be maintained in order to access the enlightenment experience.

9 Koryô period: faith and understanding according to the complete and sudden teaching

More often than not, the doctrinal teachings based on scriptures were considered by Sôn masters as a grave fallacy, an impediment to attaining final enlightenment. From their perspective, this is especially true because the conceptual or discriminatory nature of doctrinal teachings cannot invoke the Absolute, which is ineffable. On the other hand, Hwaôm, as one of the doctrinal schools, considered Sôn as merely representing the sudden teaching, which they considered a lesser teaching than the final, complete doctrine of Hwaôm. These two main branches of Mahāyāna Buddhism thus seemed to contradict each other, not only in terms of doctrinal framework, but also in their soteriological schemes.

Chinul believed that Sôn and Hwaôm were not necessarily contradictory, but could serve to complement each other. He felt that the conceptual framework of Hwaôm could help to enhance practitioners' understanding, especially those with a lesser capacity, of the mind-to-mind transmission of the esoteric teachings of Sôn. He was especially concerned about the Sôn practitioners of his time, whom he felt had strayed from correct practice due to their lack of a proper understanding of scriptural teachings. He feared that they were meditating in a state of ignorance and thus in vain. Chinul felt that proper guidance in the scriptural teachings was necessary for them to achieve final enlightenment

However, it was not an easy task to synthesize the two, thanks to their different orientations. Hwaôm philosophy systemized the path to enlightenment into fifty-two stages, beginning with the ten levels of faith and ending with the stage of Buddhahood. Often a temporal scheme was assumed, such as a period of three incalculable aeons, or more than several lives, from the initial stage of faith to the final attainment of Buddhahood. Within this progressive scheme, Hwaôm soteriology postulated a process of learning, practice and realization, in which each stage proceeds sequentially in a causal relationship. Chinul thus incorporated the soteriology of Hwaôm, which assumes a progressive development within a temporal scheme, into the sudden teaching approach of Sôn. It was in Li Tongxuan's *Exposition of the Avatamsaka Sutra* that he found a means of incorporating the two systems into a viable soteriological scheme.

Inspired by Li Tongxuan's unique interpretation of the Huayan soteriological structure, and recapitulating the latter's essential message, Chinul was able to propose, as another possible approach to Sôn practice, the method of faith and understanding as detailed in his *Wondon Songbullon (Complete and Sudden Attainment of Buddhahood)*. Chinul noted that unmoving wisdom, or the wisdom of universal brightness, is essentially identical with the deluded mind. Through a sudden awakening to this essential identity, one can enter directly the initial abiding stage, the stage of the arising of *bodhicitta*, in which one may directly experience the fact that one is a Buddha. From this stage onward, the subsequent stages of the *bodhisattva* path can be instantly achieved because of the functioning of inherent Buddhahood. Such a soteriological structure is possible due to the fact that unmoving wisdom is not merely a fruit to be attained at the final stage, but also a cause for the final attainment of buddhahood. Moreover, from the initial stage of faith to the final attainment of buddhahood, it is the unmoving

wisdom of Buddha which operates unceasingly.

10 Koryô period: shortcut approach of *hwadu* investigation

Although Chinul incorporated the theoretical framework of Hwaôm into Sôn practice, he did not forget to indicate that theories and conceptual frameworks need to be abandoned at a certain stage. In their place, Chinul proposed his method of *hwadu* investigation, which is like a finger pointing to the moon.

Chinul never lost sight of the fact that the goal of Sôn practice is to transcend conceptual understanding, thus leaving behind all discriminatory thoughts. As a means of achieving this goal Chinul found the investigation of *hwadu* most effective, a method, which in his words, cuts through iron and split nails.

Hwadu is often considered a synonym for *kongan* (in Chinese, *gongan*). More precisely, however, the *hwadu* is the essential point of the *kongan*, which is used as a topic of meditation in the Sôn school. The purpose of *hwadu* investigation is to help practitioners break through views based on their conceptual understanding of the *dharma* and to ultimately return them to the source of all discriminative thought. Thus, Chinul referred to the function of the *hwadu* as one of cleansing knowledge and understanding.

The *hwadu* method is also called the shortcut approach, as it involves no conceptual descriptions but rather points one directly to the truth right from the inception of practice. In his *Kanhwa Kyorui Ron (Resolving Doubts About Observing the Hwadu)*, Chinul discusses the three mystery gates, or three different methods, differentiated according to one’s capacity for attaining the ultimate teaching of the Sôn school: (1) the mystery of essence, which involves the conceptual understanding of the Buddha-*dharma*, such as the doctrine of the unimpeded interpenetration of all phenomena of the Hwaôm school; (2) the mystery of the word, which helps one eliminate the defects of conceptual understanding and ultimately destroy it; and (3) the mystery of mystery, in which practitioners abandon the interpretive analytical approach altogether and investigate the word itself. This method transcends any trace of a sign which a word may signify and finally results in the awareness of the noumenal state, that is, the void and calm state of mind.

In making *hwadu* the cardinal approach of Sôn practice, Chinul was influenced by Dahui Zonggao (1089-1163), a disciple of Yuanwu Gejin of the Linji school in China. Inspired by the *Dahui yulu (The Records of Dahui)*, Chinul was the first Korean Sôn master to introduce the *hwadu* into Korean Sôn practice. *Hwadu* became a hallmark of Korean Sôn and, combined with the concurrent cultivation of meditation and wisdom, and the doctrine of sudden awakening and gradual practice, represents the essence of Korean Sôn.

In summary, Chinul systematized a Korean approach to Sôn practice which stood apart from all other schemes. Its uniqueness lay in the fact that it synthesized the doctrinal teachings into Sôn practice, provided a theoretical framework for Sôn practice, and maintained the fundamental spirit of Sôn teaching by postulating *hwadu* investigation as a shortcut to the final attainment of enlightenment. With only slight variations through the ages, this structure has been maintained by Korean Sôn practitioners ever since.

After Chinul’s time, Koryô Buddhism began to gradually decline. This was primarily due to the political instability of the Koryô court after the Mongol invasion, but it was caused in part by the corruption of the Buddhist monasteries and monks. However, there were still contributions made by such monks as Kyônghan Paegun (1290-1374), T’aego Pou (1301-82) and Naong Hyegûn (1320-76), who tried to restore the strength of Buddhism in the latter Koryô. All three had returned from Yuan China after studying under the Linji (in Korean, Imje) school, which places a strong emphasis on *hwadu* practice. The most radical and confrontational methods of the Linji school became very popular in Korea and had a great impact on Korean Sôn practice. However, due to the general social instability of the times, there was insufficient support for a sustained reinvigoration of Buddhism.

In this period of decline the Korean literati, whose official education was grounded in neo-Confucian philosophy, succeeded in supplanting Buddhism as a state-sponsored ideology (see Confucian philosophy, Korean). Chinese philosophies, such as Daoism and neo-Confucianism, became increasingly influential, and monks like Kyônghan and Naong began to study these schools of thought. This represents the beginning of another unique trend in Korean Buddhism, as monks began to emphasize the study of non-Buddhist philosophies. Korea became more and more Confucianized, first in a political sense and later throughout society as a whole. This move toward Confucianism began gradually in the latter period of Koryô and accelerated with the beginning of the Chosôn

dynasty, which was founded in 1392.

11 Chosôn period (1392-1910)

The Chosôn dynasty was founded on the basis of neo-Confucianism, which combined a strong political ideology with a practical ethics emphasizing the importance of the family. Neo-Confucianism was strongly antagonistic towards Buddhism, as Buddhist monks left their families and maintained strict celibacy. During the Chosôn dynasty Buddhism was not only suppressed by those in political power but also largely ignored by intellectuals. In terms of social stratification, monks were ranked in the same class as the servants and were not allowed to enter the capital city. Amidst such harsh circumstances, Buddhism became completely marginalized.

The Buddhists of this period used their adverse circumstances as an opportunity to systematize a harmonized perspective which reconciled the philosophical conflicts among neo-Confucianism, Buddhism and Daoism (the three systems were often called *samgyo* or the Three Religions.) This synthesis in turn served once again to set Korean Buddhism apart as a unique system of thought.

Kihwa (1376-1433), otherwise known as Hamhô Tûkt'ong, continued the syncretic view of Korean Sôn systematized by Chinul, and was also the first Buddhist monk to advocate Buddhism against the neo-Confucian attack. He was also the first Buddhist monk to assert the intrinsic unity of the three religions, Buddhism, Confucianism and Daoism. Kihwa's life and times are reflected in his philosophical system. He was a prolific writer of Buddhist exegeses, in which he freely used Confucian concepts as well as citations from Confucian texts in order to clarify Buddhist ideas: he saw Buddhism and Confucianism as supplementary and not contradictory systems. He objected not only to the exclusivist attitude of Sôn practitioners toward the doctrinal schools, but also of Confucianism toward Buddhism. Before he studied Buddhism he was educated, like most sons from upperclass families, in the Confucian classics. He often used poetry, both from his own pen and from the classics, as a means of elucidating Buddhist texts (see Chinese Classics).

His syncretic approach is well demonstrated in his *Hyônjông non (Treatise on Manifesting Righteousness)*. In this apologetic essay on the unity of Buddhism and Confucianism, he not only defends Buddhism using the language of Confucians, but he goes one step further and advocates Buddhism using scriptural evidence from the Confucian classics. In this treatise, Kihwa identifies the various mental functions, which he termed usually as the arising of *dharma*, with emotions, often described as the obstruction of the true nature of mind in neo-Confucianism. In so doing, he is pointing out the similarities between Buddhism and Confucianism in the sense that both attempt to eliminate the arising of *dharma* or obstructions of the mind so that the true nature will be manifested. Buddhists call this the 'enlightened state', while Confucians call it 'being a saint'. Kihwa stressed that in terms of the final goal there is no discrepancy between them, and in fact there exists an intrinsic unity between the two. Further, as the title *Treatise on Manifesting Righteousness* implies, Kihwa saw that both ultimately reveal the truth. He also demonstrated that Buddhist cultivation does not differ from the Confucian principle of cultivating the mind, step by step, beginning with cultivation of the personal life, continuing with regulation of the family and national order, and finally ending with world peace (see Daxue). Moreover, he claimed that the Buddhist teaching allowed people to pursue different levels of practice according to their capacity, from the *bodhisattva* path to the merit-making of ordinary people. With this, Kihwa intended to demonstrate the superiority of the Buddhist teaching, which in his view embraces a universality that includes all kinds of people, in contrast to the soteriology of Confucianism, which is often limited to the highly educated.

Regarding the social aspect of the two systems, Confucianism was often considered to be superior to Buddhism. Kihwa believed otherwise. He felt that the Confucian system, by using tactics of reward and punishment in governing the people, enforced a mere superficial obedience. In contrast, as Buddhism based its teachings on the law of cause and effect, and as it further propagated the value of silence through meditation, it promoted a more spontaneous response.

Kihwa also developed the issue of social ethics, including filial piety, the most important aspect of Confucian ethics, into the greater context of truth itself, which he termed *to* (in Chinese, *dao*), or the Way. He defined the Buddhist perspective of the truth, the Way, as consisting of two aspects, the unchangeable principle and temporary expediency. According to him, only by both maintaining principle and adapting to change through the use of expedients can one achieve completion of the Way. However, the formalistic and family-centred ethics of

Confucianism cannot achieve both of these goals. Similary, he felt that one who is still subject to passions is not able to be loyal to the state and, at same time, to be filial at home. Only when one becomes free from one's own passions and conceptual boundaries can one extend oneself to universal altruism, as represented in the Confucian concept of humanity, *in* (in Chinese, *ren*), or the great compassion of Buddhism. Residing in the mountains away from home is neither unfilial nor disloyal, but is a greater form of filial piety and loyalty to the state, unimpeded by any particular phenonema. The Buddhist aim of achieving an unimpeded state between principle and phenomena, as well as among phenomena, was focused not on leaving society or the world, but rather on realizing the perfection of society.

Unlike the fourth and fifth century Chinese apologetic defenses of Buddhism, Kihwa defended Buddhism more positively and often advocated its superiority. However, before he became a Buddhist he was well known as a promising young Confucian scholar. One of his biographical accounts tells us that his interest in Buddhism was motivated by the sudden death of a friend, and also by a monk who questioned him about the Confucian practice of encouraging people to serve their parents meat, yet espousing universal affection toward all living beings. As this account illustrates, his spiritual and intellectual quest was more concerned with ontological and soteriological issues. Although he seems to have found some insight through Sôn meditation, his quest did not stop there, but further impelled him to discern an intrinsic unity among the various philosophical and religious teachings of his time. His pseudonym was Tûkt'ong, which means attainment of interpenetration; it is an apt title, as he diligently pursued this goal not merely in regard to Sôn and the Buddhist doctrinal teachings, but also in connection to all the existing teachings of his time. Dismissing the question as to which of the three teachings, Confucianism, Buddhism or Daoism, was superior, he instead felt that they were intrinsically identical. For Kihwa, the Daoist teachings of non-action and the Confucian teachings of being responsive but always calm were no different from the teaching of Sôn Buddhism, which espouses being: 'Calm but always illuminating; illuminating but always calm.'

Wônhyo's synthesis of the various doctrinal disputes was succeeded and developed by Chinul's syncretic view attesting to the unity of Sôn and scriptural teaching; this in turn was followed by Kihwa's harmonization of the three religious teachings. This tradition of the syncretic perspective in Korean Buddhism was further developed by Hyujông (1520-1604), known as Sôsan Taesa (Great Master of the Western Mountain) (see Sôsan Hyujông).

Hyujông was not only the central Buddhist figure during the Chosôn period, but also one of the most influential figures in the entire history of Korean Buddhism. He not only continued to propagate the syncretic perspective of the unity of Sôn and doctrinal teachings as formulated by Chinul, but he also attempted to reconcile the three teachings in terms of their ultimate message of truth.

By Hyujông's time, the various doctrinal schools had been reconfigured into one unified school, called Kyo, or scriptural teaching, while the Sôn schools were categorized under the single, unified name of Sôn. The conflict between Sôn and the doctrinal teachings still existed as an active point of controversy among Korean monks. As the titles of his major works, such as *Sôn'ga kwigam (The Mirror of the Sôn school), Sôn'gyo sôk (An Interpretation of Sôn and Kyo)* and *Sôn'gyo kyôl (The Secret on Sôn and Kyo)*, illustrate, Hyujông's major concern was to reconcile Sôn and the scriptural teachings by incorporating the doctrinal schools into Sôn. While generally agreeing with Chinul's syncretic teachings, Hyujông contrasted Sôn practice more sharply with Kyo and ultimately espoused Sôn's supremacy. He felt that since the intrinsic unity of Sôn and Kyo had already been clarified by the efforts of Chinul, his mission was not to reassert this point, but rather to establish a correct relationship between the two in terms of the means of attaining ultimate enlightenment.

According to Hyujông, the source of both Sôn and Kyo are the Buddha, but one refers to his mind and the other to his words. Regarding the relation between Sôn and Kyo, he asserted that Kyo is what reaches wordlessness from the word, while Sôn is what reaches wordlessness from wordlessness itself. Wordlessness means here the true nature of mind or the enlightenment state which cannot be reached through any verbal, logical description. By saying that both Kyo and Sôn reach wordlessness, however, Hyujông did not mean that one can reach the ultimate state through Kyo. Rather, he meant that Kyo, beginning with the word (that is, scriptural teaching), should be applied and then continued ultimately for the purpose of attaining the state of wordlessness. In other words, one may begin practice by studying Kyo, but in the process of attaining enlightenment, Kyo must be discarded and Sôn must be practised. This relation between Sôn and Kyo is well elaborated in his theory of *sagyo ipson*, the principle of abandoning Kyo and entering Sôn, which is still followed in present day Korean Buddhism. In this way,

Hyujông acknowledged the efficacy of Kyo as a part of Buddhist practice, yet he also asserted the supremacy of Sôn practice by postulating the limitations of intellectual understanding. Thus, he refused to recognize the possibility of attaining enlightenment through the use of Kyo alone. Of course, not every practitioner needs to follow both Kyo and Sôn practices successively. According to Hyujông, the person of high capacity may not find it necessary to follow both practices, but the person of middle or low capacity cannot omit doctrinal study, Kyo.

In his other work, *Samga Kwigam (Mirror of the Three Teachings)*, Hyujông's purpose was to demonstrate that all of the three East Asian intellectual traditions, Confucianism, Buddhism and Daoism, ultimately lead to the truth, *to*, or the Way. The intrinsic unity of the three teachings was first expostulated by his predecessor, Kihwa. However, while Kihwa's approach can be said to incorporate Confucianism and Daoism into Buddhism, especially Sôn Buddhism, and thus ultimately to advocate the superior position of the Buddhist teaching, Hyujông's approach was different. He shows us the core or the essentials of the three teachings, and by doing so he reveals the similarities among the three, although he does not ignore their differences. He felt that to espouse the superiority of one teaching over another was to be guilty of sectarianism. In interpreting each teaching of the three religions, he used a higher perspective to reconcile their minor differences into the greater context of their intrinsic unity. His approach, the synthesis of discrepancies into a higher perspective, immediately reminds one of Wônhyo's method of *kae* (unfolding) and *hap* (folding): by unfolding discrepancies, the various functions of the truth are displayed, and by folding them, the discrepancies are reconciled into the essence of the truth itself. In the epilogue of the *Samga Kwigam*, he mentions that his intention is to provide a communication channel while overcoming the limits of sectarian views. He called the three teachings 'three gates', indicating his perception of them as different means of attaining the ultimate goal. This goal, which is the realization of the ultimate truth, has often been termed *to*, or the Way, in the East Asian tradition, and carries no sectarian connotation. To realize this *to*, one's mind-eye, or wisdom, is the key. Hyujông's all-inclusive spirit, which perpetuates the long syncretic tradition of Korean religious thought, continues to have a great impact on contemporary Korean Buddhism.

12 Conclusion

Since Buddhism was introduced into the Korean peninsula during the latter half of the fourth century, Korean Buddhism has undergone three major paradigm shifts. The first occurred as a response to Korea's indigenous belief system, which may be characterized as shamanistic animism; this period extends from Buddhism's incipient stage to the Three Kingdoms period. The second shift represents the Korean effort to understand various doctrines and corresponds to the Unified Silla and Koryô periods. The third shift involved the defense of Buddhism from the criticisms of neo-Confucianism during the Chosôn period. Each paradigm shift represented a Buddhist response to challenges from either within or outside itself; in this way, Korean Buddhist philosophy developed continually. The new understandings which occurred at each shift did not disappear with the next shift, but rather remained as integral aspects of the Korean Buddhist tradition.

Buddhism's first challenge came from the indigenous shamanistic beliefs which had helped to consolidate the various tribes of the peninsula into unified kingdoms. The Korean Buddhist response was *hoguk pulgyo*, or state-protection Buddhism. Lectures on Buddhist *sūtras* as well as ceremonies to ensure national security were held regularly during the Koryô period. Even during the Chosôn period, when Buddhism was suppressed, the tradition of *hoguk pulgyo* continued in modified form as armed monks organized into armies to resist the Japanese invasion. Although the concept of *hoguk pulgyo* was never elaborated into a sophisticated philosophy, it is nevertheless a pervasive element in the development of Korean Buddhism.

In the second shift characterizing the Koreanization of Buddhism, both Wônhyo and Chinul played pivotal roles. By Wônhyo's time, the major Buddhist texts, as well as the doctrines of the numerous Chinese Buddhist schools, had been introduced to Korea. Wônhyo, using his own hermeneutics of *kae* (unfolding) and *hap* (folding), and his *hwajaeng* (reconciliation of disputes) theory, attempted to reconcile the various doctrinal disputes. In doing so, he neither accepted nor denied any one sectarian perspective. While refusing to accept any one particular doctrine as the whole truth, he recognized all doctrines as the unfolding of one mind, or the Buddha-nature. On the other hand, by 'folding' the various disputes, he synthesized them into a higher perspective of the truth. This approach reflected the fundamental spirit of Korean Buddhism, which preferred to emphasize the similarities rather than the discrepancies among Buddhist schools and among other religions. Thus the primary characteristic of Korean Buddhism, *t'ong pulgyo*, the Buddhism of total interpenetration, began with Wônhyo. It was Chinul, during the

Koryô period, who attempted to reconcile the disputes between Sôn and Hwaôm, the latter representing the scriptural tradition in Korean Buddhism. To establish a new approach to Sôn, he assimilated the theoretical framework of Hwaôm, long considered to be the anithesis of the Sôn approach.

The third shift came from the challenge of neo-Confuciansm, which became the new state ideology of the Chosôn dynasty. Kihwa, and later Hyujông, attempted to maintain the tradition of Chinul's syncretic approach to Sôn while at the same time creating the theoretical framework for the intrinsic unity of the three major intellectual traditions of Korea and all East Asia: Confucianism, Buddhism and Daoism. Continuing the approach, begun by Wônhyo and Chinul, of reconciling disputes by way of a higher perspective, Hyujông suggested that each of the three teachings is a gate leading to the same goal, that is, the *to*, or Way.

In the early twentieth century Korean Buddhists, together with the rest of their countrymen, suffered from colonization by Japan. During this time, from 1910 to 1945, many Korean Buddhist monks such as Master Yongsông (1864-1940), Master Hanyông (1870-1948) and, most notably, Master Manhae (1879-1944) became engaged in a variety of political activities. These worldly involvements, which pulled them away from their practice in the mountains, were undertaken in the *bodhisattva* spirit, that is, as a means of helping to directly alleviate the suffering of others. Due to their participation in these activities, philosophical study was also undermined. Those monks who wished to study Buddhism on a scholarly level travelled to Japan and enrolled in universities there. Japanese Buddhists in turn left their own country for Europe and absorbed the Western perspective from such countries as England, Germany and France. They gradually incorporated this new knowledge into their previous systems of understanding, and in the process Japanese Buddhist philosophy was significantly altered. The Korean monks who visited Japan were exposed to this new perspective and dutifully introduced it to their fellow Buddhists upon returning to Korea. From this time also, Buddhism began to be taught in the universities. This period marks the beginning of a bifurcation between monks and scholars, the significance of which became increasingly reflected in Buddhist scholarship.

In the 1970s and 1980s the philosophical study of Buddhism began to re-appear with renewed interest and energy. This was due to both external and internal factors: externally, Korean Buddhism faced new challenges from the West, primarily in the forms of Christianity and Western philosophy, while internally, the division between monks and scholars had become even more pronounced. In the 1980s a debate arose among monks and scholars over the issue of 'sudden' versus 'gradual'. Although this debate centred around the age-old conflict between direct experience and intellectual understanding, and thus between Sôn and doctrinal schools, due to the above-mentioned influences the conflict has become a new focus for modern Korean Buddhist thought. The debate itself has become well known not only throughout Korea but internationally as well. Yet to this day, no satisfactory resolution has been reached: some emphasize the importance of direct experience while others stress the necessity of intellectual understanding. Perhaps the conflict cannot be resolved, and this may be a reflection of our modern times. Yet what seems to be called for is a new methodology with which to consider the problem. The *t'ong pulgyo*'s perspective is unique to Korea and has served well in the past. Can Korean Buddhists find another way to embrace the understanding which it signifies, or do they need to discover an altogether different means of confronting the issue, one that stands outside of *t'ong pulgyo*? The times are unique; certainly any approach towards resolving the difficulties must be unique as well.

See also: Awakening of faith in Mahāyāna; Buddhist philosophy, Chinese; Chinul; Platform Sutra; Sôsan Hyujông; Ûisang; Wônch'ūk; Wônhyo

SUNGTAEK CHO

References and further reading

Buswell, R. (1983) *The Korean Approach to Zen: The Collected Works of Chinul*, Honolulu, HI: University of Hawaii Press.(This work includes a brief outline of Korean Buddhism before Sôn, but mainly deals with Chinul's life and works. It includes an English translation of most of the major works of Chinul, and a bibliography of works on Korean Buddhism written in Asian languages.)

Ha Tae-hung and Mintz, G. (1972) *Samguk Yusa: Legend and History of the Three Kingdoms of Ancient Korea*, Seoul: Yonsei University Press.(This is a translation of the *Samguk Yusa*, one of the most important historical sources for studying Korean Buddhist history during the Three Kingdoms period and the Unified Silla period. The original work in classical Chinese was written by monk Iryon in the thirteenth century.)

Han'guk Pulgyo Chônsô (Complete Collection of Korean Buddhist Works) (1979-89), Seoul: Dongguk University Press.(This complete collection of Korean Buddhist works is the most recent critical edition of Korean Buddhist works from the Silla period to the Chosôn period.)

Kamstra, J.H. (1967) *Encounter or Syncretism: The Early Growth of Japanese Buddhism*, Leiden: Brill.(Part 3 includes a very useful survey of Korean Buddhism from its beginnings to the period of its introduction into Japan in the sixth century.)

Keel Hee-sung (1977) 'Chinul: The Founder of Korean Son (Zen)', Ph.D. dissertation, Harvard University. (Discusses Chinul's contribution to Korean Sôn and the uniqueness of the Korean Sôn tradition.)

Korean Buddhist Research Institute (ed.) (1993) *The History and Culture of Buddhism in Korea*, Seoul: Dongguk University Press.(This edited translation of works on Korean Buddhism covers the Three Kingdoms period to modern times.)

Lancaster, L. and Yu Chai-shin (eds) (1989) *Introduction of Buddhism to Korea*, Studies in Korean Religions and Culture vol. 3, Berkeley, CA: Asian Humanities Press.(This edited English translation of articles on Korean Buddhism, originally written in Korean or Japanese, is a very useful guide for studying Korean Buddhism during the Three Kingdoms period and has a comprehensive bibliography.)

Lancaster, L. and Yu Chai-shin (eds) (1991) *Assimilation of Buddhism in Korea: Religious Maturity and Innovation in the Silla Dynasty*, Studies in Korean Religions and Culture vol. 4, Berkeley, CA: Asian Humanities Press.(This volume covers Korean Buddhism during the Unified Silla period.)

Lancaster, L. and Sung Bae-park (1979) *The Korean Buddhist Canon: A Descriptive Catalogue*, Berkeley, CA: University of California Press.(This compiles all basic bibliographical information concerning Buddhist classical texts in Chinese.)

Lee, P. (ed.) (1993) *Sourcebook of Korean Civilization: From Early Times to the Sixteenth Century*, New York: Columbia University Press.(This includes various excerpts from classical works, originally in classical Chinese. It includes an English translation of important Buddhist philosophical exegeses.)

Lee, P. (1969) *Lives of Eminent Korean Monks: The Haedong Kosung Chon*, Cambridge, MA: Harvard University Press.(This is a translation of the biographies of eminent monks of the Three Kingdoms period. The original work in classical Chinese was written by Kakhun during the Koryô period.)

Odin, S. (1982) *Process Metaphysics and Hua-Yen Buddhism: A Critical Study of Cumulative Penetration Vs. Interpenetration*, Albany, NY: State University of New York. (This comparative study of Huayan Buddhism and Whitehead's process philosophy includes in its appendix a translation of Ûisang's autocommentary on the *Ocean Seal Samādhi*.)

Park Sung-bae (1983) *Buddhist Faith and Sudden Enlightenment*, Albany, NY: State University of New York. (This is a most provocative discussion of Korean Sôn practice, presenting the spirit of Buddhist meditation in the East Asian tradition and also ably depicting the uniqueness of Korean Sôn practice.)

Shim Jae-ryong (1979) 'The Philosophical Foundation of Korean Zen Buddhism: The Integration of Son and Kyo by Chinul', Ph.D. dissertation, University of Hawaii.(Discusses Chinul's contribution to Korean Sôn and the uniqueness of the Korean Sôn tradition.)

Büchner, Friedrich Karl Christian Ludwig (Louis) (1824-99)

Ludwig Büchner wrote one of the most popular and polemical books of the strong materialist movement in later nineteenth century Germany, his Kraft und Stoff (Force and Matter) (1855). He tried to develop a comprehensive worldview, which was based solely on the findings of empirical science and did not take refuge in religion or any other transcendent categories in explaining nature and its development, including human beings. When Büchner tried to expose the backwardness of traditional philosophical and religious views in scientific matters, his arguments had some force, but the positive part of his programme was not free of superficiality and naivety. Büchner's writings helped to strengthen progressive and rational traditions inside and outside philosophy, but they can also serve as the prime example of the uncritical nineteenth-century belief in science's capacity to redeem humankind from all evil.

Büchner is commonly grouped together with Karl Vogt and Jacob Moleschott as one of the classical triumvirate of materialist philosophers in nineteenth-century Germany. Their special variant of materialism is often distinguished from other brands as 'mechanical', 'bourgeois', 'scientific' or 'vulgar' (see Materialism §3). Büchner's *Kraft und Stoff (Force and Matter)* became the most popular and enduring work of the materialist movement. Up to the First World War, it went through twenty-one editions and was translated into seventeen foreign languages, where it often saw numerous editions of its own.

Ludwig Büchner was born in Darmstadt, Germany. Like his brothers and sisters, he had a predilection for literary and journalistic activity and was engaged in republican and revolutionary opposition against the oppression of the *Vormärz* era. Büchner took up medicine at the university of Gießen in 1843 and became active in the revolutionary democratic movement of the time. In 1848 he entered his father's medical practice. In 1852 he became lecturer in medicine at the university of Tübingen, but soon lost his post when *Kraft und Stoff* appeared. He returned to Darmstadt where he stayed for the rest of his life, practising medicine, popularizing science and writing on materialist philosophy, political and social issues.

In *Kraft und Stoff*, Büchner set out to formulate the philosophical consequences which he saw in the science of his day. He propagated a thorough empiricism, rejected metaphysics and speculative philosophy and every aprioristic tendency. No supranaturalism or idealism was to be allowed in the explanation of natural events and processes.

The arguments of the book are more or less all derived from its central claim, taken from Moleschott, that there is 'no matter without force and no force without matter'. Force is inseparably bound to a material substrate and thus cannot be regarded as a kind of supernatural, transcendent entity. All changes and events follow mechanical laws. These laws are not imposed on nature from outside, but are built into matter itself. Matter is eternal and has no bounds, neither on a microscopic nor on a macroscopic scale. It is ruled by rigid and universal laws which do not allow for miracles. Soul or mind, as well as organic life, are the product of specially combined materials that are endowed with special mechanical forces. There is neither an immaterial spiritual substance, nor a vital force, nor an externally set purpose of nature.

Neither Büchner's claims nor his supportive arguments can be said to be original. *Kraft und Stoff* was less an elaborated philosophical essay than a racily written summary of the materialistic trends of his time which did not mince its words and was intelligible to the layperson. At the same time, it carried a strong political significance. The flourishing of research in the natural sciences and technology was seen as a new and effective form of opposition against the reactionary political powers which had defeated all liberal-republican and national aspirations by suppressing the revolution of 1848. The growing labour movement took Büchner's materialism as a refutation and exposure of the ideology of the ruling classes and devoured his book. This sort of writing also quenched the thirst for knowledge of several generations of students at the *Gymnasium*. By appealing to the impartiality and the common sense of the autonomous reader and leaving the intricacies and obscurities of German speculative philosophy behind, Büchner's work gave rise to a new genre of popularizing literature in which a rational and empiricist *Weltanschauung* is developed on the basis of natural science. Ernst Haeckel, Wilhelm Ostwald, Wilhelm Bölsche, the early Vienna Circle and to a certain extent even Ludwig Wittgenstein in his *Tractatus* all continued this tradition. As Büchner put it: 'Philosophical elaborations not accessible to every educated person are not worth the printer's ink used for them. What is thought out clearly can be said clearly and

plainly' (1855).

In the course of time Büchner enlarged the topics of his writings mainly in two respects. After the rise of Darwinism he emphasized its intimate relation to or even virtual identity with materialism and tried to show its positive implications for a general theory of progress. The other attempt at expanding the realm of his philosophy was to erect a humanistic and rational sociopolitical theory on his materialism which still owed very much to his liberal ideas of 1848. After the foundation of the new *Reich* he did not succumb to the temptations of nationalism, although his social and political ideals started to become obsolete and did not find a large number of supporters. His considerations included a far-sighted condemnation of the anti-Semitism of his time.

Büchner's work met with harsh opposition from many different quarters. From a philosophical point of view, the most important and momentous criticism was expressed by Friedrich Albert Lange (§2) in his *Geschichte des Materialismus und Kritik seiner Bedeutung in der Gegenwart (History of Materialism and Criticism of its Present Importance)* (1866). While standing up for Büchner against the accusation that materialism leads to immorality, and conceding a materialist methodology as even a necessity for the working scientist, Lange gave a penetrating analysis of the internal difficulties, weaknesses and inconsistencies of the philosophy of Büchner and other materialists.

This kind of criticism formed a crucial motive for rising Neo-Kantianism and led to a general recovery of philosophy's standing in Germany. Even among those who admitted the centrality of science for modern society and saw much truth in materialism, the conviction arose that philosophy had more to say than is contained in *Kraft und Stoff* and that it had not lost the right of autonomy *vis-à-vis* science.

MICHAEL HEIDELBERGER

List of works

Büchner, L. (1848) *Beiträge zur Hall'schen Lehre von einem excito-motorischen Nerven-System (Contributions to Hall's theory of an excito-motorous nervous system)*, Gießen: Schild.(Büchner's medical dissertation from 9 September 1848.)

(1855) *Kraft und Stoff. Empirisch-naturphilosophische Studien. In allgemein-verständlicher Darstellung*, Frankfurt am Main: Meidinger; trans. J.F. Collingwood, *Force and Matter, or Principles of the Natural Order of the Universe*, London: Trübner, 1864.(Subtitle is changed in later editions. 5th German edn, 1858, 21st German edn, Leipzig: Thomas 1904. The 4th English edn of 1884 was much reprinted.)

(1857) *Natur und Geist. Gespräche zweier Freunde über den Materialismus und über die real-philosophischen Fragen der Gegenwart. In allgemeinverständlicher Form (Nature and mind. Conversations of two friends on materialism and the real philosophical questions of the present)*, Frankfurt am Main: Meidinger.

(1861, 1875) *Physiologische Bilder (Physiological images)*, Leipzig: Thomas, 2 vols.(On physiological topics.)

(1862, 1884) *Aus Natur und Wissenschaft. Studien, Kritiken und Abhandlungen (About nature and science. Studies, critiques and treatises)*, Leipzig: Thomas, 2 vols. (A collection of essays on different subjects.)

(1863) *Herr Lasalle und die Arbeiter. Bericht und Vortrag über das Lasalle'sche Arbeiterprogramm, erstattet auf dem Arbeitertag in Rödelheim am 19. April 1863, im Auftrag des Central-Comités der Arbeiter des Maingaus (Herr Lasalle and the workers. Report and address on Lasalle's Worker-programme)*, Frankfurt am Main: Baist.(A discussion and criticism of the views of Ferdinand Lasalle, an early proponent of the social-democratic movement in Germany.)

(1868) *Sechs Vorlesungen über die Darwin'sche Theorie von der Verwandlung der Arten und die erste Entstehung der Organismenwelt (Six Lectures on Darwin's theory of the transformation of species and the origin of the organic world)*, Leipzig: Thomas.(Lectures on Darwin's theory of evolution.)

(1869) *Die Stellung des Menschen in der Natur, in Vergangenheit, Gegenwart und Zukunft. Oder: Woher kommen wir? Wer sind wir? Wohin gehen wir?*, Leipzig: Thomas; trans. W.S. Dallas, *Man in the Past, Present and Future*, London: Asher, 1872. (A popular account of the results of recent scientific research as regards the origin and prospects of the human race.)

(1874) *Der Gottes-Begriff und dessen Bedeutung in der Gegenwart. Ein allgemein-verständlicher Vortrag (The concept of god and its meaning for the present)*, Leipzig: Thomas.

(1876) *Aus dem Geistesleben der Tiere oder Staaten und Thaten der Kleinen*, Berlin: Hofmann; trans. A. Besant, *Mind in Animals*, London: Bonner, 1903.

(1879) *Liebe und Liebesleben in der Thierwelt (Love and love-life in the animal kingdom)*, Berlin: Hofmann.

(1882) *Die Macht der Vererbung und ihr Einfluß auf den moralischen und geistigen Fortschritt der Menschheit*

(The power of inheritance and its influence on the moral and spiritual progress of mankind), Leipzig: Günther.
(1882) *Licht und Leben. Drei allgemeinverständliche naturwissenschaftliche Vorträge als Beiträge zur Theorie der nathrlichen Weltordnung (Light and life. Three popular scientific addresses as contributions to the theory of the natural world-order),* Leipzig: Thomas.
(1884) *Der Fortschritt in Natur und Geschichte im Lichte der Darwin'schen Theorie (Progress in nature and history in the light of Darwin's theory),* Stuttgart: Schweizerbart.
(1885) *Der neue Hamlet. Poesie und Prosa aus den Papieren eines verstorbenen Pessimisten (The new Hamlet. Poetry and prose from the papers of a deceased pessimist),* Zürich: Verlags-Magazin, J. Schabelitz.(This is a collection of early literary work by Büchner which includes an interesting autobiographical fragment on 189-196. It appeared under the pseudonym Carl Ludwig. A second edition appeared under the author's real name in Gießen, E. Roth, 1901.)
(1887a) *Thatsachen und Theorien aus dem naturwissenschaftlichen Leben der Gegenwart (Facts and theories from the scientific life of the present)*, Berlin: Hofmann.
(1887b) *Über religiöse und wissenschaftliche Weltanschauung. Ein historisch-kritischer Versuch (On religious and scientific world-views. A historical and critical essay),* Leipzig: Thomas.(On the difference between religious and scientific world-views.)
(1889) *Das künftige Leben und die moderne Wissenschaft. Zehn Briefe an eine Freundin (Life beyond death and modern science. Letters to a female friend)*, Leipzig: Spohr.
(1890a) *Fremdes und Eignes aus dem geistigen Leben der Gegenwart (Notes from myself and others on the spiritual life of the present)*, Leipzig: Spohr.(A collection of essays on different subjects.)
(1890b) *Zwei gekrönte Freidenker. Ein Bild aus der Vergangenheit als Spiegel für die Gegenwart. Dem deutschen Volke gewidmet (Two crowned free-thinkers. A portrait from the past as a mirror for the presence. Dedicated to the German people),* Leipzig: Thomas.
(1891) *Das goldene Zeitalter, oder das Leben vor der Geschichte. Nebst einem Anhang: Das Kulturmetall der Zukunft (The golden age, or life before history. With an appendix: The cultural metal of the future)*, Berlin: Allgemeiner Verein für deutsche Literatur.
(1892) *Das Buch vom langen Leben oder die Lehre von der Dauer und Erhaltung des Lebens (Makrobiotik) (The book of longevity or the doctrine of the endurance and preservation of life (macrobiotics))*, Leipzig: Spohr.
(1894a) *Darwinismus und Sozialismus, oder: Der Kampf ums Dasein und die moderne Gesellschaft (Darwinism and Socialism, or: the struggle for existence and modern society),* Leipzig: Günther.
(1894b) *Meine Begegnung mit Ferdinand Lasalle. Ein Beitrag zur Geschichte der sozialdemokratischen Bewegung in Deutschland. Nebst fünf Briefen Lasalles (My encounter with Ferdinand Lasalle. A contribution to the history of the Social-Democratic movement in Germany. With five letters by Lasalle)*, Berlin: Hertz & Süßenguth.(Recollection of his encounter with Lasalle, the founder of the German social-democratic party.)
(1898) *Am Sterbelager des Jahrhunderts. Blicke eines freien Denkers aus der Zeit in die Zeit,* Gießen: Roth; partly trans. J. MacCabe, *Last Words on Materialism,* London: Watts, 1901.(A collection of essays on different subjects; chapter 12 is on the Jewish question.)
(1900) *Im Dienste der Wahrheit. Ausgewählte Aufsätze aus Natur und Wissenschaft (In the service of truth. Selected papers on nature and science)*, Gießen: Roth.(Again a collection of essays. Contains a biography of Ludwig Büchner by his younger brother Alex.)
(1901) *Kaleidoskop. Skizzen und Aufsätze aus Natur und Menschenleben (Kaleidoscope. Sketches and essays from nature and human life)*, Gießen: Roth.(This collection contains an appreciation of Büchner's work by Wilhelm Bölsche: 'Zur Geschichte der volkstümlichen Naturforschung' (On the history of popular research of nature).)
Wittich, D. (ed.) (1971) *Vogt, Moleschott, Büchner: Schriften zum kleinbürgerlichen Materialismus in Deutschland (Vogt, Moleschott, Büchner: writings of vulgar materialism)*, Berlin: Akademie, 2 vols.(This reprint of the important writings of materialists includes also a reprint of the 1855 edition of *Kraft und Stoff*. The editor gives a useful 82-page introduction.)

References and further reading

Berglar, P. (1976) 'Der neue Hamlet - Ludwig Büchner in seiner Zeit' (The new Hamlet - Ludwig Büchner in his time), *Archiv für Kulturgeschichte* 58 (1): 204-26. (An account of Büchner's development.)
Büchner, A. (1963) *Die Familie Büchner. Georg Büchners Vorfahren, Eltern und Geschwister (The Büchner family: Georg Büchner's ancestors, parents, brothers and sister)*, Hessische Beiträge zur deutschen Literatur, Darmstadt: Roether.(On the background of Büchner's family. Pages 64-81 give an account of Büchner's life by his grand-nephew and a reprint of Büchner's autobiographical fragment.)
Frauenstädt, J. (1856) *Der Materialismus. Seine Wahrheit und sein Irrtum. Eine Erwiderung auf Dr. Louis*

Büchner's 'Kraft und Stoff', Leipzig: Brockhaus.(A contemporary critique of Büchner's work by an adherent of Schopenhauer. Translated as *The Materialism of the Present Day: A Criticism of Dr Büchner's System*, London, 1867.)

Gregory, F. (1977) *Scientific Materialism in Nineteenth-Century Germany*, Dordrecht: Reidel.(Standard account, comprehensive and very readable.)

(1977) 'Scientific versus Dialectical Materialism: a Clash of Ideologies in Nineteenth-Century German Radicalism', *Isis* 68: 206-23.(Deals with the relation of Büchner's brand of materialism with that of Marx and Engels.)

Janet, P. (1864) *Le Matérialisme contemporain en Allemagne, examen du système du docteur Büchner*, Paris: Baillière; trans. G. Masson, *The Materialism of the Present Day. Critique on Dr. Büchner's System*, London, 1865.(Criticism of Büchner's work by a French philosopher which was also influential in its German translation.)

Lange, F.A. (1866) *Geschichte des Materialismus und Kritik seiner Bedeutung in der Gegenwart*, Iserlohn: Baedeker; augmented and altered 2nd edn, 2 vols, 1873-75; trans. E.C. Thomas, *The History of Materialism and Criticism of its Present Importance*, London: The English & Foreign Philosophical Library, 1877-9, 3 vols; repr. with intro. by B. Russell, London: Kegan Paul, 1925, 1 vol.(The second edition was particularly influential in its criticism of mechanical materialism. Lange's account of the relation of science and philosophy after Kant is still extremely valuable today.)

Lehmann, S., ed. (1987) *Georg Büchner 1813 - 1837: Revolutionär, Dichter, Wissenschaftler (Katalog der Ausstellung Mathildenhöhe, Darmstadt 1987)*, Basel and Frankfurt am Main: Stroemfeld/ Roter Stern, 1987.(Catalogue of a grandiose exhibition on life, time and work of Büchner's brother Georg; with informative and original articles, also on the other members of the Büchner family, including on Ludwig Büchner by G. Böhme, on pages 376-93.)

Lübbe, H. (1963) *Politische Philosophie in Deutschland. Studien zu ihrer Geschichte (Political philosophy in Germany: studies of its history)*, Basel: Schwabe; repr. München: Deutscher Taschenbuch Verlag, 1974.(The 3rd part deals with the intimate relation of scientific materialism with politics/political philosophy and its transformation into the movement of 'scientific monism' in Germany in later years.)

Moleschott, J. (1852) *Der Kreislauf des Lebens. Physiologische Antworten auf Liebig's Chemische Briefe (The circulation of life: physiological answers to Liebig's chemical letters)*, Mainz: Zabern.(Response to a book by the chemist J. von Liebig. This book of a fellow materialist was of particular influence on Büchner.)

Negri, A. (1981) *Trittico materialistico: Georg Büchner, Jakob Moleschott, Ludwig Büchner (A materialist tryptych: Georg Büchner, Jakob Moleschott, Ludwig Büchner)*, Roma: Cadmo.(Valuable information on Marx's and Engel's comments on Büchner and on the concept of death in nineteenth-century materialism.)

Schnädelbach, H. (1983) *Philosophie in Deutschland 1831-1933*, Frankfurt: Suhrkamp; trans. E. Matthews, *Philosophy in Germany 1831-1933*, Cambridge: Cambridge University Press, 1984.(A useful introductory text which puts the materialistic movement in nineteenth-century German philosophy in a wider context. See 66-108 on science.)

Schreiner, K. (1977) 'Der Fall Büchner. Studien zur Geschichte der akademischen Lehrfreiheit an der Universität Tübingen im 19. Jahrhundert' (Büchner's case: studies in the history of academic freedom at the University of Tübingen in the nineteenth century), *Beiträge zur Geschichte der Universität Tübingen 1477-1977*, ed. H. Decker-Hauff, Tübingen: Attempto, 307-46.(An account of Büchner's time at Tübingen, his dismissal from his academic position and its influence on the general political development. Based on extensive archival sources that give a vivid account of the circumstances.)

Ziegler, T. (1899) *Die geistigen und sozialen Strömungen des neunzehnten Jahrhunderts (The intellectual and social movements of the nineteenth century)*, Berlin: Bondi.(Chapter 11, 'Naturwissenschaft und Philosophie um die Mitte des Jahrhunderts', gives a vivid, still readable and valid account.)

Buffier, Claude (1661-1737)

A French Jesuit who flourished in the early eighteenth century, Buffier developed an outlook that he referred to as common-sense philosophy. While deeply influenced by the philosophies of Descartes and Locke, he saw their reliance on the testimony of inner experience to be conducive to scepticism concerning the external world. In reaction to this, he sought to establish the irrevocable claims of various 'first truths', which pointed towards external reality and qualified it in various respects. His work anticipates certain themes that surfaced later in the common-sense philosophy of Thomas Reid.

Buffier was born in Poland to French parents. He pursued his studies in the Jesuit college at Rouen, and entered there as a novitiate at the age of 18. His career was spent almost entirely at the college of Louis-le-Grand in Paris, first as a teacher then, after 1699, as 'scriptor', a position that enabled him to devote all his energies to writing. He wrote extensively on philosophy, geography, history, and grammar (often in verse, to facilitate memorization) and made important contributions to the *Dictionnaire de Trévoux* and the *Mémoires de Trévoux*.

Buffier's philosophical writings provide little indication of his Jesuit commitments - perhaps it was this that led Voltaire to refer to him as 'the only Jesuit to have put forward a reasonable philosophy in his works'. It is Descartes and Locke who emerge as principal sources of Buffier's thinking, though he was no slavish disciple of either. To both of these thinkers he owed the belief that inner experience (*le sentiment intime*) provides the strongest possible certification for truth, and that any logically correct inferences based on such experience are equally incontrovertible. Yet he realized that if we take such evidence to be the only philosophically legitimate evidence, solipsism and scepticism follow immediately. And these positions he regarded as absurd.

Rather than call into question the Cartesian premises, as some later thinkers were to do, Buffier sought to establish that various propositions pointing towards the external world and indicating something of its nature are every bit as acceptable, philosophically, as the claims of immediate experience. These propositions he termed the first truths of common sense (*le sens commun*), and he held that they reflect certain dispositions that nature has placed in us all in order to bring unity to what would otherwise be a heterogeneous mass of idiosyncratic judgments springing from the immediate experience of each of us.

First truths are not established by any straightforwardly deductive process, for they are only to be reasoned *from*. They reveal themselves instead through their clarity and universality: no propositions offered to support or supplant them could be formulated more clearly than they are; and they show themselves either to be explicitly acceptable to all sane persons who have reached the age of reason, or to be embedded in the conduct of all, even of those who might indulge in their denial. Some examples of these first truths are:

Bodies exist;
There are other beings and other people than myself in the world;
I possess something I call intelligence and something independent of this intelligence that I call my body;
The presence of design, such as we find in a watch, requires an intelligent designer;
The self or soul is absolutely one;
I possess a free will.

While Buffier does recognize certain differences between the claims of outer and inner experience, he refuses simply to place them on the same scale and declare the former to be weaker than the latter. Instead, he will say that both are equally 'certain', although through inner experience we grasp truths with greater 'vividness'; or that to deny inner experience would take one out of 'oneself', while to deny our outer experience would place one outside 'reason' (by leaving one with no ultimate premises to reason from); or that denial of the former might lead to contradiction, but denial of the latter would indicate some form of madness (taking one outside the domain of reasonability). The exigencies of life, in short, require us to accept and live in accord with the dictates of common sense.

Buffier accepts calling such truths innate ideas, if we mean by this that they constitute certain dispositions by which the mind grasps reality. However, he rejects any belief in innate ideas (see Innate knowledge) if these are construed as particular items standing permanently before the mind for contemplation. In fact Buffier rejects the

very conception of ideas as ‘things’ that stand before the mind, regarding them instead purely as modifications of our soul in so far as it is thinking, which suggests that the mind, when it receives an idea, is no more distinguishable from that idea than a ball set in motion is distinguishable from the motion it receives. This point, touched upon but lightly by Buffier, was to become the very focus of Reid’s critique of the ‘ideal system’ (see Reid, T.).

Noticeably absent from Buffier’s list of first truths is any that makes mention of God’s existence. He believed, rather, that this particular truth followed reasonably from the first truth by which we infer the existence of an intelligent designer from the observation of design in an object. Quite generally, where philosophical and scientific questions are concerned, Buffier maintains an empiricist’s reserve, deploring such speculative excesses as one finds in Descartes’ physics or Malebranche’s metaphysics. And while his principal philosophical priority involves overcoming the hyperbolic scepticism associated with solipsism, he freely affirms our ignorance of the ultimate causal forces at work in the universe. The essences of things that we can know are but nominal essences, while real essences stand beyond our reach - real essences being understood by Buffier not as those that ground the qualities of a given substance, but as those that render an individual truly unique. Out of this ignorance, however, emerges an emphasis on the important position occupied by probability (*vraisemblance*) in forming sound opinions and making judicious decisions.

See also: Common Sense School

JAMES W. MANNS

List of works

Buffier, C. (1843) *Oeuvres philosophiques du Père Buffier (Philosophical Works of Father Buffier)*, ed. Bouillier, F., Paris: Charpentier.(Contains the *Traité*, the *Eléments*, the *Examen*; preceded by a substantial and insightful introduction.)

(1704) *Examen des préjugez vulgaires, pour disposer l’esprit à juger sainement de tout (Examination of common prejudices)*, Paris: Chez Mariette.(A dialogue aiming to resolve twelve commonly disputed issues; showing, for example, that two people can disagree yet be both in the right and that women can be adept in all the sciences.)

(1714) *Les principes du raisonnement exposez en deux logiques nouvèles (Principles of Reasoning Expounded in Two New Logics)*, Paris: Chez Pierre Witte.(An examination of the many different senses in which the notion of truth is employed, leading to the recommendation that we should broaden our criteria of this concept.)

(1724) *Traité des premières véritez et de la source de nos jugements, où l’on examine le sentiment des philosophes sur les premières notions des choses*, Paris: Veuve Mongé; trans. anon. *First Truths, and the Origin of Our Opinions Explained, with an inquiry into the sentiments of moral philosophers relative to our primary notions of things to which is prefixed a detection of the Plagarism, Concealment and Ingratitude of Doctors Reid, Beattie, Oswald*, London: Johnson, 1780.(The work in which the central tenets of Buffier’s doctrine of common sense are most clearly and systematically exposed.)

(1725) *Eléments de métaphisique à la portée de tout le monde (Elements of Metaphysics, Accessible to Everyone)*, Paris: Pierre François Giffart.(An examination, presented in dialogue form, of the nature, significance, and utility of metaphysical reflection.)

References and further reading

Aguilar, J.A.V. (1957) *El Sentido Commún en las Obras Filosóficas del P. Claude Buffier, S.I. (Common Sense in the Philosophical Works of Father Claude Buffier, S.J.)*, Barcelona: Seminario Conciliar.

Marcil-Lacoste, L. (1982) *Claude Buffier and Thomas Reid: Two Common-Sense Philosophers*, Kingston and Montreal: McGill-Queen’s University Press.(Good comparison between Buffier and Reid, requiring some philosophical sophistication; good bibliography of secondary literature.)

Montgomery, F.K. (1930) *La vie et l’oeuvre du Père Buffier (The Life and Work of Father Buffier)*, Paris: Association du Doctorat.(Brief biography and general look at the whole corpus of Buffier’s work, not just the philosophical items. Contains a complete enumeration of Buffier’s writings.)

Sortais, G. (1928) ‘Le Cartésianisme chez les Jésuites français aux dix-septième et dix-huitième siècles’ (Cartesianism among the French Jesuits of the Seventeenth and Eighteenth Centuries), *Archives de Philosophie* 6 (3): 1-93.

Wilkins, K.S. (1969) ‘A Study of the Works of Claude Buffier’, *Studies on Voltaire and the Eighteenth Century* 66, Geneva: Bestermann.(Good biography, including detailed treatment of Buffier as an educator. Views him generally from the philosophical perspective. Good secondary bibliography.)

Buffon, Georges Louis Leclerc, Comte de (1707-88)

Both as a scientist and as a writer, Buffon was one of the most highly esteemed figures of the European Enlightenment. In depicting the perpetual flux of the dynamic forces of Nature, he portrayed the varieties of animal and vegetable species as subject to continual change, in contrast with Linnaeus, whose system of classification based on physical descriptions alone appeared timeless. But Buffon's definition of a species in terms of procreative power excluded the evolutionary hypothesis that any species could become transformed into another. Hybrids, as imperfect copies of their prototypes, were in his scheme ultimately destined to become sterile rather than to generate fresh species. By virtue of the same definition, he judged that the different races of mankind formed family members of a single species, since the mating of humans of all varieties was equally fertile.

1 Career

In his own lifetime among the most highly esteemed of all the *philosophes* of the age of Enlightenment, Buffon achieved eminence within the intellectual and political establishments of his day rather than, like either Diderot or Voltaire, as an independent writer outspokenly critical of prevalent orthodoxies. Born in Montbard, he first gained distinction and early admission to both the Académie des Sciences and the Royal Society for his work on mathematics. By 1739 his writings on organic nature had won him sufficient renown to secure his appointment as keeper of the Jardin du Roi in Paris, which formed both a botanical garden and a zoological museum. From that base until his death, he became one of the most celebrated men of science and letters in France. On account of its being published by the Imprimerie Royale, his massive *Histoire naturelle (Natural History)* (1749-88) largely escaped the clutches of censorship and official criticism which befell so many of his prominent contemporaries. In 1753 he was elected a member of the Académie Française, and he was to become a corresponding member of more scientific and learned societies than perhaps any other figure of the eighteenth century. Hume remarked of him that he seemed less like a writer than a marshal of France.

2 His *Histoire naturelle*

Buffon conceived his *Histoire naturelle* principally as an account of the generative and degenerative properties of Nature, along lines which contrasted with the descriptive taxonomy of Linnaeus, and it was set within an even wider cosmology opposed to the mechanistic philosophy of Newton. As distinct from the Linnaean system of classification based upon observable similarities of organic structure, Buffon's causal and developmental scheme of Nature addressed its procreative powers and variable manifestations through time and space. Through his reading of such works as Madame Du Châtelet's *Institutions de physique (Instruction in Physics)* of 1740, he was persuaded that Christian Wolff's conception of Nature as a set of dynamic forces of impulsion which engendered the life-forms of matter was more plausible than Newton's notion of inert Nature framed by absolute time and space (see Wolff, C.; Newton, I.).

Supposing that the planets had been formed from the impact of comets with the sun, he argued that the topography of the earth was attributable to the effects of erosion. Just as civil history records the epochs of the revolutions of human affairs and the occurrence of moral events, so does natural history record the physical development and transformations of the world, he claimed. In his creation of heaven and earth, God had made only the matter of the universe. The earth's consolidation, with the introduction of water, the birth of volcanoes, the separation of continents, the generation of animals and the emergence of the human race, constituted the history of all organic substance, which Buffon was to term *les époques de la nature* in the most celebrated volume of his work, bearing that title, published in 1778. From his manuscripts it appears that he supposed the earth might be 3 million years old, but in print he allowed only a total of around 75,000 years.

3 His notion of a species

Buffon's early interest in mathematics had included research in infinitesimal calculus, a subject which accorded well with the ancient doctrine to which he subscribed, of a natural chain of being, or *scala naturae*, linking all vegetables and animals by a series of gradations of their physical form. In part because he accepted that Nature's ladder consisted of unbroken steps, each in its place from the start, Buffon doubted the central premise of what

would become evolutionary or Darwinian biology, according to which one species generates another. Along lines corresponding to Plato's theory of Forms, he even maintained that, because species are perpetual and permanent, they comprise the only true beings in Nature. Each individual organism, or what is now termed a 'phenotype', he thought, was patterned by the *molécule organique* or *moule intérieur* of its prototype - which today would be termed the 'genotype' of its species as a whole.

But Buffon's conception of each species' form was more dynamic than that of any philosopher of nature before him. The most striking difference between the world's animal and vegetable matter, he claimed, is the power of movement possessed by animals alone, allowing for the fact that some animals, such as oysters, lack that power as well. Buffon noted that individual members of species may be altered or improved by climate or nourishment, and he therefore stressed that the phenotypes of organisms were marked by their differences or variability. Permanence or fixity, however, was generically inherent in every species, he insisted. Contending that Nature is in continual flux, he perceived the diverse instantiations of species as either improving or degenerating according to their circumstances, a point which he illustrated most strikingly with respect to the flora and fauna of the New World as contrasted with the Old, in so far as he deemed the North American puma to be a diminutive lion and the South American llama just a small camel.

The crucial factor which determined the nature of a species, according to Buffon, was not physical resemblance of bodily parts, as in Linnaeus' taxonomy, but procreative ability - a species being defined as a constant succession of similar individuals that can reproduce together. The sterility of hybrids - such as a mule interbred from the horse and ass - made it difficult for Buffon to lend credence to any ideas of descent with modification or the transformation of one species into another, as had been mooted by his contemporary, Maupertuis (see Species).

4 His anthropology

Buffon's definition of a 'species' in terms of its power of procreation led him to view the whole of humanity as forming only one species, whose populations were differentially tinged by the sun in higher and lower latitudes. There could be other determinants of race, he allowed, such as bile or blood, but in general he supposed that skin colour was an acquired characteristic shaped essentially by climate, thus following Montesquieu's conjectures on the climatological determinants of men's moral traits (see Montesquieu). Believing that the progenitors of the human race must have been white, Buffon held that the skin colour of blacks was merely a deformation of the archetype of humanity. This proposition excludes the evolution of mankind from the ape, since a white archetype would scarcely have arisen from a black animal without a soul, he thought. But it did permit a proposition that others would pursue to a racist conclusion - that apes might be degenerate blacks.

Buffon himself categorically rejected that hypothesis, both because he thought the human faculty of reason established an unbridgeable gulf between mankind and the rest of Nature, and because there was no evidence of viable offspring from matings between apes and humans such as were produced from the racial interbreeding of blacks and whites. Stressing that animation - or life - is a property of matter, while thought is a property of spirit alone, he was convinced that the entire human race sprang from the same family and formed only one species. A disciple, Cornelius De Pauw, nevertheless contended in his *Recherches philosophiques sur les américains* that the native inhabitants of America had degenerated to the level of orang-utans, a proposition rejected by Henry Home, Lord Kames, in the polygenist theory of multiple races which he set out in his *Sketches of the History of Man* of 1774, largely as a refutation of the monogenetic anthropology of Buffon.

See also: Darwin, C.; Enlightenment, Continental; Evolution, Theory of; Species

ROBERT WOKLER

List of works

Buffon, G.L.L. (1749-88) *Histoire naturelle, générale et particulière (General and particular natural history)*, Paris: Imprimerie Royale.

References and further reading

Blanckaert, C. (1993) 'Buffon and the Natural History of Man: Writing History and the ''Foundational Myth'' of Anthropology', *History of Human Sciences* 6: 13-50.(Addresses diverse interpretations of Buffon's pioneering contribution to the history of anthropology, with a substantial bibliography.)

Duchet, M. (1971) *De l'homme (On Man)*, Paris: François Maspero.(Includes an introductory essay on Buffon's anthropology.)
Farber, P.L. (1972) 'Buffon and the Concept of Species', *Journal of the History of Biology* 5: 259-4.(Traces the changes and development of Buffon's philosophy of nature.)
Fellows, O. and Milliken, S.F. (1972) *Buffon*, in *World Authors Series*, vol. 243, New York: Twayne.(A synoptic treatment of Buffon's life, the nature of his theory of evolution and his style.)
Gerbi, A. (1973) *The Dispute of the New World: The History of a Polemic, 1750-1900*.(A comprehensive account of the influence of Buffon's denigration of the natural species of America, originally published in Italian in 1955.)
Lovejoy, A.O. (1959) 'Buffon and the Problem of Species', in B. Glass, O. Temkin and W. Strauss, Jr. (eds) *Forerunners of Darwin, 1745-1859*, Baltimore, MD: Johns Hopkins University Press, 84-113.(Originally published in 1911 in *Popular Science Monthly*, a wide-ranging assessment of Buffon's steps towards evolutionism.)
Roger, J. (1993) *Les sciences de la vie dans la pensée française du XVIII^{e} siècle*, 3rd edn, Paris: Michel.(The third edition of a text first published in 1963, which comprehensively locates Buffon's natural history within its scientific and philosophical context.)
Sloan, P. (1995) 'The Gaze of Natural History', in C. Fox, R. Porter and R. Wokler (eds) *Inventing Human Science: Eighteenth-Century Domains*, Berkeley, CA and Los Angeles: University of California Press, 112-51. (Addresses Buffon's contributions to eighteenth-century natural history and anthropology, by contrast with those of Linnaeus).
Wilkie, J.S. (1956-7) 'The Idea of Evolution in the Writings of Buffon', *Annals of Science* 12: 48-62, 212-27 and 255-66.(Offers various explanations of Buffon's failure to produce a fully generalized theory of evolution.).
Wokler, R. (1976) 'Tyson and Buffon on the Orang-utan', *Studies on Voltaire and the Eighteenth Century* 155: 1-19.(Assesses Buffon's account of the animal limits of humanity in terms of a scheme of comparative anatomy.)

Bulgakov, Sergei Nikolaevich (1871-1944)

*A luminary of the Russian Religious-Philosophical Renaissance, Bulgakov moved from Marxism, to idealism, to Christianity in the early twentieth century. He rejected historical determinism, class struggle and all theories of progress that accept the suffering of one generation as a bridge to the happiness of another. He regarded the abolition of poverty as a moral imperative, insisted that Christianity mandates political and social reform, and wanted to create a new culture in which Orthodox Christianity would permeate every area of Russian life. His most important philosophical works, Filosofiia khoziaistva, chast' pervaia (The Philosophy of the Economy, Part I) (1912) and Svet nevechernyi (Unfading Light) (1917), reflect his turn to a Solov'ëvian mysticism which apotheosized transfiguration, Sophia and Godmanhood (*Bogochelovechestvo*). Bulgakov saw the cosmos as an organic whole, animated and structured by a World Soul, an entelechy that he called Sophia, Divine Wisdom. Sophia mediates between God and his creation, working mysteriously through human beings. In emigration, Bulgakov developed new interpretations of Orthodox dogmatics and participated in the ecumenical movement. His lifelong concerns were the Church in the world and the interconnection of religion and life. His writings on contemporary political, social and cultural issues helped inspire the Russian Religious-Philosophical Renaissance.*

1 Life and early writings

Bulgakov was born in Livny, Orel province, a descendant of six generations of priests. He started out to be one, but left the seminary in 1888, studied in a secular gymnasium and then entered Moscow University's juridical faculty, graduating in 1894. He had advanced training in economics and statistics and taught these subjects at the Moscow Technical Institute. A 'legal Marxist' in the mid-1990s, Bulgakov contributed to left-leaning journals and was a progenitor of the 'back to Kant' movement, which tried to supplement Marxism with an autonomous ethic. His book, *O rynkakh pri kapitalisticheskom proizvodstve (On Markets in Capitalist Production)* (1897), which argued that Russia could achieve capitalism on the basis of its domestic market alone, brought him a national reputation. Between 1898 and 1900 Bulgakov did research for his doctoral dissertation in Western Europe, mostly in Germany. His chance discovery of Raphael's Sistine Madonna at an art gallery in Dresden made him realize, though not immediately, that he needed 'not a philosophical idea but a living faith in God, Christ, and the Church'. In 1898 he married Elena Tokmatova. They had two sons and a daughter; a mystical experience at the funeral of a third son, who died in 1909 at the age of 3, deepened Bulgakov's religiousness.

After returning to Russia, Bulgakov taught political economy at the Kiev Polytechnical Institute and Kiev University. In *Kapitalizm i zemledelie (Capitalism and Agriculture)* (1900) he argued that small farms are more productive and stable than large ones, contradicting the Marxist view. His essay, 'Basic Problems of the Theory of Progress' appeared in the symposium *Problemy idealizma (Problems of Idealism)* (1902) and in his own collection, *Ot Marksizma k idealizmu (From Marxism to Idealism)* (1903). One essay asked the question 'What does the philosophy of Solov'ëv give to the contemporary consciousness?' 'Positive all unity' and fusion of Christian theory and Christian practice (see Solov'ëv, V.S.) was Bulgakov's answer. Already critical of abstract German Idealism, he believed that philosophy must encompass 'living experience'. In 1904 he became a co-editor of Merezhkovskii's journal *Novyi Put' (New Path)* and, in 1905, of its successor *Voprosy zhizni (Problems of Life)*. During the Revolution of 1905 Bulgakov advocated a Union of Christian Politics, and helped found the Christian Brotherhood of Struggle and the Moscow branch of the Religious Philosophical Society. In 1906 he left Kiev to become Professor of Political Economy at the Moscow Commercial Institute and to teach at Moscow University. The same year, he was elected to the second Duma as a non-Party Christian Socialist on the Kadet slate. Bulgakov regarded the popularity of socialism as a punishment for the sins of historical Christianity and a call to repent for ignoring poverty and injustice. In 1908 he returned to the Russian Orthodox Church. Around the same time Bulgakov began a second intensive reading of Solov'ëv and became close friends with the philosopher-priest Pavel Florenskii. In 1909, Bulgakov attacked the revolutionary intelligentsia as egoistic and destructive in 'Geroizm i podvizhnichestvo' (Heroism and Selfless Devotion), his contribution to *Signposts* (see Signposts movement). The same year, he called for an Orthodox work ethic as part of a larger attempt to create a distinctive Orthodox culture that would permeate every area of Russian life from the most exalted to the most prosaic. In 1910, Bulgakov helped found the Orthodox publishing house Put'. In 1911, he resigned from Moscow University in protest against

the suspension of university autonomy, but resumed teaching there in 1917. As a delegate to the All Russian Council of the Orthodox Church (1917-18), he worked to restore the Patriarchate.

In June 1918 Bulgakov was ordained as a priest, losing his professorship as a result. He taught at the University of Simferopol, in the Crimea, until 1921. After being expelled from Russia on 30 December 1922, he lived in Prague for two years and then settled in Paris. He helped found the Russian Student Christian movement and the Orthodox Theological Institute, serving as its Dean and as Professor of Dogmatic Theology. In 1939, Bulgakov lost his voice almost completely as a result of surgery for throat cancer, but he continued to celebrate the liturgy and to conduct his classes.

2 *The Philosophy of the Economy* and *Unfading Light*

These works represent Bulgakov's attempt to create an Orthodox philosophy. In *Filosofiia khoziaistva, chast' pervaia (The Philosophy of the Economy, Part I)*, he discussed the meaning and significance of the economy, including its epistemology, ontology and phenomenology, and promised to discuss ethics and eschatology in Part II. His purpose was to reveal an inherent cosmic meaning in the most prosaic human actions - production and consumption - as well as the intrinsic kinship of man and the cosmos and their basis in God and in Sophia, the mediator between God and the world. Interpreting 'Marx in the spirit of Boehme and Boehme in the spirit of Marx', Bulgakov stressed the centrality of labour, the need to increase production, and the Hegelian and Marxist concepts of freedom and necessity, mingling these ideas with Boehme's concept of the World Soul (see Boehme, J.). Bulgakov taught that human labour, physical and mental, has metaphysical and cosmic significance because through it, Sophia permeates and transfigures the world. All human activity contains elements of freedom and necessity, of Sophia and anti-Sophia respectively. Explaining economic activity (activity conducted out of necessity) in terms of the Fall, Bulgakov maintained that before it, labour was disinterested and loving, but then 'nature, damned by God, became a hostile force, armed with the powers of hunger and death'. Economic activity became a struggle for survival, concealing the Sophianic significance of labour and engendering the materialistic ideology of economism, the view of life as only or primarily an economic process. Science arose as part of man's attempt to subjugate 'Tsar Nature', and therefore also contains Sophianic and anti-Sophianic elements, as does art - the most Sophian human activity. Creative activity is of divine origin, by way of Sophia, for only God can create. But God allows man to participate in his salvation and to be a co-creator. Labour is humankind's means of redemption.

Objecting to the Kantian bifurcation of subject and object, Bulgakov maintained that man is the subject *and* the object of the economy, and stressed the former, the active subject, the proprietor (*khoziain*) of a farm or a business. He described economic creativity as a primarily psychological phenomenon, 'a phenomenon of [internal or] spiritual life', and associated creativity with individuality and freedom. An Orthodox economy would be characterized by *sobornost'*, unity in love and freedom, not impersonal collectivism or egoistic individualism. Egoism was a result of the Fall, which shattered the wholeness of the cosmos, and separated people from one another.

Bulgakov looked forward to a new era when the tyranny of nature would be overcome, when man would rule the economy instead of being ruled by it. Labour would again be voluntary and joyful, and would merge with artistic creativity. This 'leap from necessity to freedom' was not foreordained, he insisted, because the principle of Sophia is freedom. But elsewhere in the book, Bulgakov explained the apparent discord and chaos of economic life as the workings of a superpersonal force, a mystical version of Hegel's 'cunning of reason' and suspiciously akin to 'necessity'.

Bulgakov never wrote Part II. He considered *Svet nevechernyi (Unfading Light)* the continuation of Part I, but it was more theological in nature. Bulgakov argued that God created the world out of nothing as an emanation of his own nature, not as something alien or external to him, and then pronounced his creation 'very good' (Genesis 1.31). This means that Christians must not reject the world, or nature, or 'the flesh'. Bulgakov also argued that God created the world as dynamic becoming, not as eternal being. Therefore, the world is not evil but incomplete. In the section titled 'The Sophianicity of the Creature', Bulgakov described Sophia as the living link between God and the creature, as a kind of fourth hypostasis outside and different from the Trinity, and as the object of God's love. He was ambiguous on whether Sophia is a person or a force, but he did say that Sophia is revealed to the world as beauty. In later works, Bulgakov distinguished between the 'divine Sophia' and the 'created' or

'creaturely Sophia', but he continued to emphasize their ultimate metaphysical identity and hence the consubstantiality of God and the cosmos. In *Svet nevechernyi*, Bulgakov attacked Boehme's mysticism as pantheism or 'immanentism' and declared that this 'heresy' became the basis of subsequent German thought. But Bulgakov also argued that pantheism contains a grain of religious truth and must be absorbed into an Orthodox cosmology which does not negate the world. In later works he distinguished panentheism from pantheism. Apropos of the promised treatment of ethics and eschatology, Bulgakov attributed evil to the nothingness or non-being that is the substratum of the cosmos and that erupts in human wilfulness, but he provided no guidelines for conduct except to condemn egoism and individualism and praise humility and love. Bulgakov's eschatology featured the transfiguration of the world through art, which he called 'Sofiiurgy', not 'theurgy', as the Symbolists did, and the divinization of all humankind, that is, Godmanhood. Resurrection would be a universal phenomenon, for man is 'one Adam' and the cosmos is 'one corporeality and one body'.

3 From Orthodox philosopher to dogmatic theologian

Svet nevechernyi, *Die Tragödie der Philosophie (The Tragedy of Philosophy)* and *Filosifiia imeni (The Philosophy of the Name)*, the latter two written between 1919 and 1921, mark Bulgakov's transition from Orthodox philosophy to dogmatic theology, and his rejection of philosophic idealism. In *Die Tragödie der Philosophie*, Bulgakov declared that philosophy must be the handmaiden of religion, not just of theology; that is, philosophy must encompass religious experience, ritual, and personal revelation. He criticized Kant, Fichte, Hegel and Spinoza extensively, partly because they dichotomized consciousness into 'I-not I'. Bulgakov posited a dialogic consciousness - 'I am thou, he, you, and we' - relating this conception to a triune theory of human relations which he in turn related to the Holy Trinity. *Filosofiia imeni* was Bulgakov's response to a dispute in the Orthodox Monastery of Mount Athos about whether the name of God was itself divine and also to Symbolist writers' debates on the nature of the Word. Bulgakov argued that the name of God was inseparable from the divinity but not identifiable with it, and that the very act of naming connects God and man. Words are living symbols that connect the empirical and the invisible worlds.

In emigration Bulgakov turned to patristics as part of a heightened emphasis on Orthodox tradition and on the Church as an institution. He described the Church as a living body, the Body of Christ, and maintained that tradition, including dogma, encompasses 'the new in the old and the old in the new'. Bulgakov interpreted Orthodox dogmatics in this spirit. Sophiology, treated somewhat differently, remained his central theological concept. The émigré Church in Karlovtsi, Serbia, and the Moscow Patriarchate alike accused him of heresy. Underlying their attack was hostility to any reinterpretation of dogma at all. Bulgakov's dogmatic works are not available in English. To date, they have not been reprinted in Russia, but his earlier writings are available and are attracting a great deal of attention.

See also: Russian religious-philosophical renaissance

BERNICE GLATZER ROSENTHAL

List of works

Bulgakov, S. (1897) *O rynkakh pri kapitalisticheskom proizvodstve (On Markets in Capitalist Production)*, Moscow: A.G. Kolchugin.(Argues that Russia could achieve capitalism on the basis of its domestic market alone.)

Bulgakov, S. (1900) *Kapitalizm i zemledelie (Capitalism and Agriculture)*, St Petersburg: V.A. Tikhonov.(Argues that small farms are more productive and stable.)

Bulgakov, S. (1903) *Ot Marksizma k idealizmu (From Marxism to Idealism)*, St Petersburg: Obshchestvennaia Pol'za.(Collection including 'Basic Problems of the Theory of Progress'.)

Bulgakov, S. (1909, 1911) *Dva grada (Two Cities)*, Moscow: Put', 2 vols.(Articles written between 1904 and 1909.)

Bulgakov, S. (1912) *Filosofiia khoziaistva chast' pervaia (The Philosophy of the Economy, Part I)*, Moscow: Put'. (Discusses the meaning and significance of the economy, including its epistemology, ontology and phenomenology.)

Bulgakov, S. (1917) *Svet nevechernyi (Unfading Light)*, Moscow: Put'.(Bulgakov considered this a continuation of *Filosofiia khoziaistva chast' pervaia*, though more theological in nature.)

Bulgakov, S. (1918) *Tikhie dumy (Quiet Thoughts)*, Moscow: Leman & Sakharov.

Bulgakov, S. (1926) *Sv. Pëtr i Ioann (Saints Peter and John)*, Paris: YMCA-Press.

Bulgakov, S. (1927a) *Die Tragödie der Philosophie (The Tragedy of Philosophy)*, Darmstadt: Otto Reichl.(Marks Bulgakov's transition from Orthodox to dogmatic theology.)

Bulgakov, S. (1927b) *Kupina neopalimaia (The Burning Bush)*, Paris: YMCA-Press.(This volume about the Virgin Mary, and the following two, constitute Bulgakov's 'small trilogy'.)

Bulgakov, S. (1927c) *Drug zhenikha (Friend of the Bridegroom)*, Paris: YMCA-Press.(St John the Baptist is the subject of this book.)

Bulgakov, S. (1929) *Lestnitsa Iakovlia: Ob angelakh (Jacob's Ladder: About Angels)*, Paris: YMCA-Press.(This volume is about angels and higher states of being.)

Bulgakov, S. (1933) *O bogochelovechestve I, agnets Bozhii (On Divine Humanity I, The Lamb of God)*, Paris: YMCA-Press.(This volume about Jesus, and the following two, constitute Bulgakov's 'great trilogy'.)

Bulgakov, S. (1936) *O bogochelovechestve II, uteshitel (On Divine Humanity II, The Comforter)*, Paris: YMCA-Press.(This volume is about the Holy Spirit.)

Bulgakov, S. (1937) *Sophia, The Wisdom of God*, London; revised edn, Hudson, NY: Lindesfarne, 1993.(An outline of Sophiology, from an unpublished Russian manuscript.)

Bulgakov, S. (1945) *O bogochelovechestve III, nevesta agnetsa (On Divine Humanity III, The Bride of the Lamb)*, Paris: YMCA-Press.(This work concerns the Church.)

Bulgakov, S. (1946) *Avtobiograficheskie zametki (Autobiographical Notes)*, ed. L.A. Zander, Paris: YMCA-Press.

Bulgakov, S. (1948) *Apokalipsis Ioanna (The Apocalypse of John)*, Paris: YMCA-Press.

Bulgakov, S. (1953) *Filosofiia imeni (The Philosophy of the Name)*, Paris: YMCA-Press. (Concerns the name of God, and the nature of the Word.)

Bulgakov, S. (date unknown) *Pravoslavie*, Paris: YMCA-Press; repr., Moscow: Terra, 1991; abridged trans. *The Russian Orthodox Church*, London, 1935; repr. Crestwood, NY: St Vladimir's Seminary, 1988.

Bulgakov, S. (1993) *S.N. Bulgakov, Sochineniia v dvukh tomakh (S.N. Bulgakov, Selected Works in Two Volumes)*, ed. S. Khoruzhi (vol. 1) and I. Rodianskaia (vol. 2), Moscow: Nauka.(Volume 1 includes Khoruzhi's introductory article 'Vekhi filosofskogo tvorchestva o. Sergeia Bulgakova' (Landmarks of the Philosophical Work of Father Sergei Bulgakov), Bulgakov's *Tragediia filosofii*, and selected articles. Volume 2, contains Rodianskaia's introductory essay, 'S.N. Bulgakov - Publitsist i obshchestvennyi deiatel'' (S.N. Bulgakov - Writer and Public Figure) and selected articles, mostly from *Dva grada*.)

Bulgakov, S. (1976) *A Bulgakov Anthology*, eds N. Zernov and J. Pain, Philadelphia, PA: Westminster.(Brief excerpts from Bulgakov's works.)For additional bibliography see Evtuhov (1996) and Zander (1948).

References and further reading

Evtuhov, C. (1996) *The Cross and the Sickle: Sergei Bulgakov and the Fate of Russian Religious Philosophy*, Ithaca, NY: Cornell University Press(Bulgakov's thought in the context of the Russian Silver Age.)

Kornblatt J.D. and Gustafon, R. (eds) (1996) *Russian Religious Thought, Contexts and New Perspectives*, Madison, WI: University of Wisconsin Press.(Contains three articles on Bulgakov: B.G. Rosenthal, 'The Nature and Function of Sophia in Bulgakov's Prerevolutionary Thought'; M.A. Meerson, 'Sergei Bulgakov's Philosophy of Personality'; and P. Valliere, 'Sophiology as the Dialogue of Orthodoxy with Modern Civilization'.)

Rosenthal, B.G. (1991) 'The Search for a Russian Orthodox Work Ethic', in E. Clowes, S. Kassow and J.L. West (eds) *Between Tsar and People*, Princeton, NJ: Princeton University Press, 57-74. (Bulgakov's search for an ethic based on labour that is creative and loving.)

Zander, L. (1948) *Bog i mir (God and the World)*, Paris: YMCA-Press, 2 vols.(A respectful study of Bulgakov's worldview.)

Bultmann, Rudolf (1884-1976)

Rudolf Bultmann was one of the most influential Protestant theologians of the period that immediately followed the Second World War. A founding member of the school of dialectical theology in the 1920s, he was a major New Testament scholar, who refined the method of form criticism. He argued that the Synoptic Gospels reveal not the historical Jesus, but the Christ of faith, the Christ-myth developed by the early church. The existentialist philosophy of Martin Heidegger was a major influence, and he adapted it to the needs of Christian theology, devising an existential access to faith. He contrasted Historie - *objective, factual accounts of historical events* - *with* Geschichte - *the meaning that people choose to give to those events. One must demythologize the New Testament* - *strip it of its prescientific imagery* - *before one can interpret its significance for oneself. Bultmann defined biblical hermeneutics as an inquiry into the reality of human existence and proposed a new understanding of the person and teaching of Christ. Central to this is the concept of the* kerygma, *the proclamation of the salvation-event focused on Christ. It is in response to the* kerygma *that a human being can actively opt for faith. Bultmann reinterpreted the Lutheran doctrine of justification and the theology of the cross in the light of this.*

1 Life

Rudolf Bultmann was the son of a minister of the German Evangelical Lutheran Church. At Oldenburg he attended the same gymnasium as Karl Jaspers. He read theology at Tübingen, Berlin and Marburg. Karl Müller (Church history), Adolf von Harnack (history of dogma), Hermann Gunkel (Old Testament), Adolf Jülicher and Johannes Weiss (New Testament), and Wilhelm Herrmann (systematic theology) were among his teachers. He graduated in New Testament studies in 1912. His teaching activity began at the University of Marburg, where he spent a lot of time in the company of Martin Rade, whose theological journal *Die christliche Welt* had a wide audience. In 1916 he was invited to teach at Breslau and received a call to a chair at Giessen in 1920. A year later, he was offered a chair at the university of Marburg. Meanwhile he had released *Die Geschichte der synoptischen Tradition (The History of the Synoptic Tradition)* (1921).

At Marburg, Bultmann taught New Testament for thirty years. There he became acquainted with Rudolf Otto and Martin Heidegger. During the National Socialist dictatorship, he courageously joined the Confessing Church and spoke openly against the ‘Aryan Paragraph’. In 1941, he delivered a highly influential lecture to a circle of ministers: ‘New Testament and Mythology: The Problem of Demythologizing the New Testament Revelation’. Bultmann’s demythologizing programme aroused ecclesiastical indignation which resulted in its official rejection at the General Synod of the German Evangelical Church in 1952. His years at Marburg were among his most creative: *Jesus* (1926), *Das Evangelium des Johannes (The Gospel of John)* (1941), *Theologie des Neuen Testaments (Theology of the New Testament)* (1948-51), *Das Urchristentum in Rahmen der antiken Religionem (Primitive Christianity in its Contemporary Setting)* (1949) and *History and Eschatology* (1957) were among the works he wrote. Most of his articles were collectively published in *Glauben und Verstehen (Faith and Understanding)* (1933-65). Bultmann retired in 1951.

2 History and myth

As a member of the school of dialectical theology that Karl Barth (§§1-2), Eduard Thurneysen and Emil Brunner (§1) had started in the early 1920s, Bultmann rejected the liberal way of treating Christian faith as a phenomenon of the history of religion. However, unlike Barth, he carried forward the tradition of biblical criticism. In spite of the acknowledgement of the paradoxical nature of revelation and faith in Barth’s *Der Römerbrief (The Epistle to the Romans)* (1919), Bultmann did not surrender the method of history of religion that he had successfully applied in his doctoral dissertation. He was interested in understanding early Christianity within the wider religious and cultural context of its time, and endeavoured to retrace the influence of pre-Christian Gnosticism, Jewish apocalyptic and Manicheism on the New Testament writings. This orientation led to the later theory of demythologization. Like David Friedrich Strauss (§1) and the School of Marburg, Bultmann had a modern worldview and wanted to cleanse the New Testament of its naïve prescientific one. His exegetical research primarily concerned the pre-literary traditions of the New Testament. He investigated the internal development of the text, its ‘setting in life’ (*Sitz im Leben*) and its literary genres. In this way, he made significant contributions to form criticism.

What are the results of Bultmann's historical investigations? *The History of the Synoptic Tradition* (1921) provides an analysis of the origin and the redactorial process of the traditions underlying the narrative of Jesus' life. In the Synoptic Gospels, the original proclaimer, Jesus, has become the Christ, the one proclaimed as God's decisive act for human beings. The Synoptic Gospels offer an extended picture of the faith of the early Church in Jesus the Christ. They do not proclaim the historical Jesus but the Christ of faith (*der Christus des Glaubens*), the *kerygma*, the Christ-myth. Bultmann unveiled a process in Christology whereby Jesus the historical being developed into the mythical idea of a celestial being, a God-elected king of the eschaton, the pre-existing *logos* and co-begetter of the universe. In a famous monograph (1926), Bultmann characterized this process as a transformation of the Gospel of Christ into the Gospel about Christ. He claimed that the supernatural events in Jesus' life, including the miracles attributed to him, were legendary, mythical accounts. Jesus probably was a disciple of John the Baptist, an eschatological prophet. He created his own community of followers which rivalled that of the Baptist. But unlike the Baptist, Jesus was both a rabbi who taught the Law of God and a prophet who proclaimed the imminent end of the world. Jesus died as a messianic prophet.

Throughout his work, Bultmann thoroughly respected the principles of positivist historiography. In his view, there is only one historical method, common to both the believer and the unbeliever. Against Nietzsche, who claimed that there are only interpretations, he maintained that historical events are knowable objectively. He was aware of the historian's subjectivity, but subjectivity is not subjectivism. In this respect, the most subjective history is the most objective. Bultmann, however, did not confine himself to the quest for the historical Jesus. The meaning of the Christ-event for the believer, its significance for the individual, was of crucial importance to him. Owing to the adoption of Heidegger's existential analysis, Bultmann was able to draw an all-important distinction between *Historie* and *Geschichte*. *Historie* concerns objective, factual accounts of past events, whereas *Geschichte* points to the intentionality or the meaning of events for the people who encounter them. Historical facts are neutral, but *geschichtliche* events concern people's existence and prompt a for-or-against decision on their part. In order to discover the response of Christian faith to the Christ-event, one must divest the New Testament of the prescientific imagery in which it has been cast; one must 'demythologize' its message. This was the purpose of Bultmann's demythologizing programme.

3 Existential interpretation

Once demythologized, theological doctrines such as virgin birth, pre-existence, resurrection, ascension, original sin, miracles, angels, demons, and so on must be interpreted in order to attain their significance for the individual. In accordance with the early Martin Heidegger (§§2-3), Bultmann put forth an existential interpretation. Heidegger's *Sein und Zeit (Being and Time)* (1927) phenomenologically describes the existential structure of the human being. In this analysis of the ontological constitution of *Dasein* (being-in-the-world), existence is understood as a special mode of being independent of natural determination. Fundamental ontology is related to, but not identical with, anthropology. As Bultmann retained from Barth a trans-subjective understanding of revelation and faith, so he borrowed an existentialist ontology from Heidegger. Bultmann's concern for existence derived from a need to secure a realm of freedom in a world of causality and necessity. Contrary to the rational being, who is confined to objectivity, the free and volitional being exists in and through their decisions. For, by its essence, will is intentionality, project and decision; freedom is openness to the future. According to Bultmann, a human being's existence is realized in life acts, in the intentionality of decisions which are made ever anew and involve the whole being. A human being may exist either authentically or inauthentically. Inauthentic existence is the search for satisfaction and security in tangible reality, whereas authentic existence is life based on intangible realities, in free and responsible openness to the future, in faith.

Bultmann repeatedly claimed that Heidegger's fundamental ontology was neutral. It therefore can profitably be used to interpret any biblical text, and does not imply a subordination of the Bible to a philosophical framework. For the historian cannot investigate *for me* what that text means about God and his action, nor answer the question of the meaning of a given biblical text *for me*. For any understanding presupposes a prior understanding of the subject. A prior understanding or prior knowledge of the possibilities of one's life is the presupposition for the understanding of any biblical text. Bultmann also adopted Heidegger's idea of a hermeneutical circle. But he further drew a distinction between the existentialist understanding of philosophical analysis, which he called *das Existentiale*, and the pre-philosophical, concrete and ontic self-understanding that he labelled *das Existentielle*. Only the latter clears the way for the understanding and the appropriation of the revelatory action of God.

Consequently, biblical interpretation involves more than linguistic, grammatical and historical research. It implies both an ontology and an anthropology. The connection of text and interpreter, which is founded on the relationship of the interpreter to life, is a further presupposition for the understanding of any biblical text. I can recognize God's action only if I already have an existential (*existentiell*) knowledge of God. I can understand the *kerygma* when it speaks to me of God because I have an existential self-understanding. Thus biblical hermeneutics is understood as an inquiry into the reality of human existence. However, Bultmann did not substitute the existential-eschatological for the historical-objective, but established an authentic relationship instead of the perversion of objectivization or mythologization.

4 *Kerygma* and eschatology

In his early writings, Bultmann established a relationship between theology and anthropology. Theology is a discourse based on divine revelation. However, such a discourse is only possible if it concomitantly speaks about human beings. Thus theology is dialectic on two accounts: the dialectic of the hidden god and the dialectic of human existence. According to Bultmann, theology is the conceptual description of the existence of human beings as being determined by God. God should not be objectified or mythologized; God is the *eschaton*. Revelation is not the disclosure of a set of objective or objectifiable truths, but is God's judgment on the world; it concerns human existence at its deepest. As anthropology is existential (*existentiell*), so theology is eschatological.

The *kerygma*, or proclamation of the salvation-event, is the focal concept of Christian theology. According to the New Testament, Jesus Christ, the Crucified and Risen One, is the eschatological event, the action by which God has set an end to the old world (see Eschatology §1). The *kerygma* is message, testimony and proclamation; preaching is part of the essence of the Word. As eschatological event, the proclaimed *kerygma* is not only the occasion for the decision as to faith, but also the condition for faith. Jesus Christ is present in the *kerygma* and encounters human beings through the *kerygma*. Preaching is in itself an eschatological event prompting an eschatological decision on the part of the hearer.

Bultmann was a Lutheran. The doctrine of justification and the theology of the cross receive a special emphasis in his eschatological theology. The God that human beings encounter through justification by faith is no object of knowledge and thought; it is the acting, transcendent and hidden God. Thus, to have faith is to believe in spite of experience. In faith, a human being surrenders all human securities and overcomes the predicament of human existence, the despair in which human existence is thrown, by choosing freely in favour of authentic existence. Under God's grace, a human being acknowledges a new self-understanding, embraces a new life and accepts a new responsible mode of action born of love. Through the human response to the *kerygma* the 'old man', torn by sin and anxiety, is transformed into a 'new man', freed from this world and rooted in the world beyond. The demythologizing programme that Bultmann launched in the 1950s is fundamentally related to the Lutheran doctrine of justification (see Justification, religious §4). The programme was a radical implementation of the doctrine of justification by faith in the field of knowledge and thought.

The cross of Christ is the salvation-event proclaimed by the New Testament. By the cross, Bultmann does not mean the historical-objective (*historisch*) fact of the crucifixion, but the historical-personal (*geschichtlich*) faith-event that the early Church experienced in relation to the crucifixion. The cross is God's liberating judgment on the world. Because Christ was crucified *for us*, the cross creates a new situation between God and human beings. The cross is not a mere event in the past which can be objectively contemplated. The cross is an occurrence of faith. To the believer, it is an ever-present and repeatable reality, in the sacraments as well as in daily life. Regarding the resurrection, which is inseparable from the cross, Bultmann taught that it was a further proclamation of the eschatological meaning of the cross. The resurrection is no historical fact, but an occurrence of faith. The Easter faith of the early Church was a faith in the glorified Christ. The resurrection is the main eschatological event, the salvation-event that signifies the believer's newness of life. The preached word renders the cross and the resurrection, or the *kerygma*, repeatedly present for all to hear.

In Bultmann's theology, ecclesiology is not less important than Christology. Because God and the Word are eschatological, God's revelation in Christ is not merely a past occurrence to be historically traced. The Word is an ever-living reality which is encountered where it is preached, and among those in which it is alive. Since justification by faith is not only freedom from sin but also freedom for obedience, Christians are called to an

exemplary ethical life and to responsible action in political life. The living Christ reveals himself to the world through the work of salvation that is constantly operating in and through the Church. The end of the world and the coming of the kingdom of God are proclaimed to the world only through the Church. From there the Word calls today's humanity to experience crucifixion and resurrection; the eschatological event is continued in the Church. The Church is the eschatological community, and, like Jesus Christ, the incognito of God.

See also: Existentialist theology; Hermeneutics, biblical §4

JEAN-LOUP SEBAN

List of works

Bultmann, R. (1921) *Die Geschichte der synoptischen Tradition*, Göttingen: Vandenhoeck & Ruprecht; trans. J. Marsh, *The History of the Synoptic Tradition*, New York: Harper & Row, 1963.(A classic work on the process of redaction which led to the Synoptic Gospels as we know them.)

Bultmann, R. (1926) *Jesus*, Berlin: Deutsche Bibliothek; trans. L.P. Smith and E. Huntress Lantero, *Jesus and the Word*, New York: Scribner, 1934.(Bultmann's Christology, based on the Jesus preached by the early Church.)

Bultmann, R. (1933-65) *Glauben und Verstehen*, Tübingen: Mohr, 4 vols; vol. 1 trans. L.P. Smith, *Faith and Understanding*, Philadelphia, PA: Fortress Press, 1987; vol. 2 trans. J.C.G. Greig, *Essays Philosophical and Theological*, London: SCM, 1955. (Bultmann's collected essays, on any and every topic.)

Bultmann, R. (1941) *Das Evangelium des Johannes*, Göttingen: Vandenhoeck & Ruprecht; trans. G.R. Beasley-Murray, R.W.N. Hoare and J.K. Riches, *The Gospel of John: A Commentary*, Oxford: Blackwell, 1971.(A very detailed and innovative form-critical commentary.)

Bultmann, R. (1948-51) *Theologie des Neuen Testaments*, Tübingen: Mohr, 2 vols; trans. K. Grobel, *Theology of the New Testament*, New York: Scribner, 1951-5.(Material on Johannine theology, the rise of the early Church, and the development of doctrine.)

Bultmann, R. *et al.* (1948-55) *Kerygma und Mythos: ein theologisches Gesprach*, ed. H.W Bartsch, Hamburg: Reich & Heidrich, 5 vols; trans. R.H. Fuller, *Kerygma and Myth: A Theological Debate*, London: SCM Press, 2 vols, 1953, 1962.(Selections from the first two German volumes are in the first English volume, and selections from the last three are in the second. This is the definitive guide to the theological debate on demythologization.)

Bultmann, R. (1949) *Das Urchristentum in Rahmen der antiken Religionem*, Zurich: Artemis-Verlag; trans. R.H. Fuller, *Primitive Christianity in its Contemporary Setting*, New York: Meridian, 1957.(Sections on the Old Testament, Jewish, Greek and Hellenistic backgrounds to primitive Christianity. It is particularly dated in its treatment of Judaism.)

Bultmann, R. (1957) *History and Eschatology*, Edinburgh: Edinburgh University Press. (Bultmann's Gifford Lectures, demythologizing eschatological claims.)

Bultmann, R. (1958) *Jesus Christ and Mythology*, New York: Scribner.(Bultmann's 1951 Yale Shaffer Lectures, once again stating the demythologizing programme.)

Bultmann, R. (1961) *Existence and Faith: Shorter Writings of Rudolf Bultmann*, ed. S. Ogden, London: Hodder & Stoughton.(Some overlap with the *Glauben und Verstehen* volumes mentioned, but a good collection with a strong introduction.)

Bultmann, R. and Barth, K. (1971) *Karl Barth - Rudolf Bultmann Briefwechsel 1922-1966*, ed. B. Jaspert, Zurich: Theologischer Verlag; trans. and ed. G.W. Bromiley, *Karl Barth - Rudolf Bultmann Letters 1922 - 1966*, Edinburgh: T. & T. Clark, 1982. (Useful both for the controversy between Barth and Bultmann, and for their autobiographical content.)

Bultmann, R. (1985) *New Testament and Mythology*, ed. S. Ogden, London: SCM.(Some overlap with the *Glauben und Verstehen* volumes. Contains the important 1941 article on demythologizing, 'New Testament and Theology'.)

References and further reading

Barth, K. (1919) *Der Römerbrief*, Bern: Bäschlin; 2nd edn, much changed, Munich: Chr. Kaiser Verlag, 1922; 6th edn trans. E.C. Hoskyns, *The Epistle to the Romans*, Oxford: Oxford University Press, 1968.(Bultmann was deeply influenced by this initator of 'dialectical theology'.)

Braaten, C.E. and Harrisville, R. (eds) (1962) *Kerygma and History: A Symposium on the Theology of Rudolf Bultmann*, New York: Abingdon, 1962.(An important collection of essays by prominent theologians.)

Evang, M. (1988) *Rudolf Bultmann in seiner Frühzeit (Rudolf Bultmman's Early Years)*, Tübingen.(A meticulous survey of Bultmann's development and the influences that shaped him.)

Fergusson, D. (1992) *Bultmann*, London: Chapman.(A short but helpful volume in the Outstanding Christian Thinkers series.)

Heidegger, M. (1927) *Sein und Zeit*, Tübingen: Neomarius Verlag, 10th edn, 1963; trans. J. Macquarrie and E. Robinson, *Being and Time*, London: SCM, 1962. (The main philosophical influence on Bultmann.)

Jaspert, B. (ed.) (1989) *Rudolf Bultmanns Werk und Wirkung*, Darmstadt: Wissenschaftliche Buchgesellschaft. (Thirty essays from prominent scholars, covering the state of the field in Bultmann studies. Five are in English.)

Johnson, R.A. (1987) 'Introduction', in R.A. Johnson (ed.) *Rudolf Bultmann: Interpreting Faith for the Modern Era*, London: Collins.(A good short introduction to the main issues.)

Jones, G. (1991) *Bultmann: Towards a Critical Theology*, Cambridge: Polity Press. (Good on the philosophical background and political implications of Bultmann's work.)

Kegley, C.W. (ed) (1967) *The Theology of Rudolf Bultmann*, New York: Harper & Row.(Biographical material and a response by Bultmann, several good essays, and a bibliography complete to 1967.)

Morgan, R. (1989) 'Rudolf Bultmann', in D. Ford (ed.) *The Modern Theologians*, Oxford: Blackwell.(Helpful, particularly on Bultmann's relationship to his theological forebears.)

Painter, J. (1987) *Theology as Hermeneutics: Rudolf Bultmann's Interpretation of the History of Jesus*, Sheffield: Almond Press.(Concentrates on the extent to which Bultmann's whole position is intended as an interpretation of the New Testament.)

Buridan, John (*c.*1300-after 1358)

Unlike most other important philosophers of the scholastic period, John Buridan never entered the theology faculty but spent his entire career as an arts master at the University of Paris. There he distinguished himself primarily as a logician who made numerous additions and refinements to the Parisian tradition of propositional logic. These included the development of a genuinely nominalist semantics, as well as techniques for analyzing propositions containing intentional verbs and paradoxes of self-reference. Even in his writings on metaphysics and natural philosophy, logic is Buridan's preferred vehicle for his nominalistic and naturalistic vision.

Buridan's nominalism is concerned not merely with denying the existence of real universals, but with a commitment to economize on entities, of which real universals are but one superfluous type. Likewise, his representationalist epistemology accounts for the difference between universal and singular cognition by focusing on how the intellect cognizes its object, rather than by looking for some difference in the objects themselves. He differs from other nominalists of the period, however, in his willingness to embrace realism about modes of things to explain certain kinds of physical change.

Underlying Buridan's natural philosophy is his confidence that the world is knowable by us (although not with absolute certainty). His approach to natural science is empirical in the sense that it emphasizes the evidentness of appearances, the reliability of a posteriori modes of reasoning and the application of certain naturalistic models of explanation to a wide range of phenomena. In similar fashion, he locates the will's freedom in our evident ability to defer choice in the face of alternatives whose goodness appears dubious or uncertain.

1 Life and work

It is testimony to Buridan's stature and influence that many of the events traditionally associated with his life are the stuff of legend. The stories that the King of France ordered that he be thrown into the Seine in a sack because of a scandalous affair with the Queen, or that, after being driven from Paris for his nominalist teachings, he went on to found the University of Vienna, cannot be verified. What we do know of his life is derived from the facts of his academic career (Faral 1950; Michael 1985). From these it can be inferred that Buridan was born shortly before 1300, probably in Béthune in northwestern France. He served as rector of the University of Paris in 1328, and again in 1340. University documents indicate that he received several benefices to support his work, and that he was much esteemed by his colleagues. He is last mentioned in a document of 1358, and probably died a short time after.

Buridan's academic career was unusual in two respects. First, he spent his entire career as a Master of Arts, without ever seeking an advanced degree in theology. He accordingly produced no commentary on the *Sentences* of Peter Lombard, one of the main genres of philosophical writing in the fourteenth century. Almost all of his written work is based on the Parisian arts curriculum, and reflects his pedagogical concerns as a member of the arts faculty. Second, he remained a secular cleric rather than joining an order such as the Dominicans or Franciscans. This freed him from the doctrinal disputes which often arose between religious orders, something which can be seen in the eclectic character of his work.

Buridan's logical writings are in the form of handbooks and commentaries intended for use by students of logic. Most of his non-logical works appear as short commentaries (*expositiones*) or longer critical studies (*quaestiones*), usually in several versions, based on his lectures on the works of Aristotle. In addition to the entire Organon, these included lectures on the *Metaphysics*, *Physics*, *On the Heavens*, *On the Soul*, *Parva naturalia*, *Nicomachean Ethics* and *Politics*. Only a few of these works are available in modern editions, and none can be dated precisely. However, copies and early printed versions were distributed throughout European universities, where they often served as primary texts for courses in logic and Aristotelian philosophy. As a result, Buridan's teachings continued to shape European thought well into the Renaissance.

2 Logic

Buridan belongs to the 'terminist' tradition of medieval logic, so called because its practitioners regarded terms as the primary unit of logical analysis. He wrote his most comprehensive logical work, the *Summulae de dialectica (Summary of Dialectic)*, as a commentary on the *Tractatus* or *Summulae logicales (Summary of Logic)*, a terminist

textbook composed by Peter of Spain almost a century earlier. However, because Buridan charged himself with ‘analyzing and supplementing’ Peter’s rather elementary remarks ‘sometimes differently’ than Peter himself had expressed them, his comments served as a forum for his own innovative ideas in logic (see Pinborg 1975).

One section of Peter’s text rewritten entirely by Buridan is the treatise on supposition, which deals with the referential function of significative terms in propositional contexts. In this section, Buridan presents a revised account of the divisions of supposition that better expresses his nominalist ontology. He rejects in particular the notion of ‘simple supposition’, by which some logicians held that common terms such as ‘man’ could refer to ‘universal natures… distinct from singulars outside the soul’ (*Summulae* IV, 3). In Buridan’s view, terms refer only to particulars, although they can do so in different ways depending upon their semantic function. Buridan also rejects as superfluous the idea of his Parisian contemporary Gregory of Rimini that what propositions signify, or ‘make known’, is an abstract object that is ‘complexly signifiable’. Further expression of Buridan’s nominalism can be found in the *Tractatus de consequentiis (Treatise on Consequences)*, an advanced study of the theory of inference, in which he argues that one need posit no additional cause for the falsity of a sentence beyond what makes its corresponding negation true.

Buridan moved the Parisian tradition in new directions by developing and then systematically applying semantic theory to solve philosophical problems (Klima 1991, 1992). As well as characterizing his revision of the doctrine of supposition, this emerges in his analysis of the meaning of propositions containing intentional verbs (for example, ‘you know the one approaching’), in which he appeals to the counterpart doctrine of appellation to show how we can ‘understand the same thing in many different ways’ (*Sophismata* IV, seventh remark) without altering the particular denotation of the concept or term used to signify it. Buridan has also achieved some notoriety among modern logicians for his solution to the liar paradox, in part because it preserves bivalence (see Semantic paradoxes and theories of truth). This turns on his suggestion that a proposition *P* is true if and only if (1) it signifies things as they are, and (2) it entails, in conjunction with a proposition asserting that *P* actually exists, another proposition asserting that *P* is true. Buridan then argues that self-referentially paradoxical propositions (for example, ‘every proposition is false’) are false because they cannot consistently satisfy both conditions. Even if we assume that they signify things as they are, the propositions they entail are false because the subject and predicate of the entailed propositions do not ‘stand for the same’ (if *P* is ‘every proposition is false’, then *P* itself must be false, not true), thus failing the second condition (*Sophismata* VIII, seventh sophism) (for discussion see Spade 1978; Hughes 1982) (see Language, medieval theories of; Logic, medieval).

3 Metaphysics and epistemology

Like most medieval nominalists, Buridan argues that commonality is not a real feature of the world. ‘Universals,’ he says, ‘do not exist outside the soul, but are only concepts of the soul by which it conceives of many things indifferently’ (*Summulae* IV, 3). Medieval nominalism, however, was a complex and evolving doctrine whose adherents were not merely interested in the question of universals. The distinctiveness of Buridan’s nominalism consists in the way he implements the more general nominalist aim to economize on entities, of which real universals are but one superfluous type (see Nominalism; Universals).

Buridan’s contribution to the nominalist theory of universals was to develop the semantics for an intentionalist account of common concepts. In a discussion influenced by his Parisian predecessor Peter Abelard, Buridan argues that concepts such as ‘humanity’ and ‘rationality’ are best understood as ways of conceiving individual human beings, rather than as mirroring abstract entities (de Rijk 1992). Although Buridan, like William of Ockham, accepts only substances and qualities in his basic ontology, his explanation of certain kinds of physical change reflects this tendency to multiply modes or ‘ways’ of being, rather than things. Thus he regards changes in the shape or position of a physical substance as changes in the mode(s) of its qualities, that is, in *how* it is, whereas the unadorned substance-quality view would be forced to posit the generation of new qualities. Medieval nominalists also disagreed about where (and how) to use the razor, or principle of parsimony, a fact which distinguishes Buridan’s nominalism from other versions. Buridan differs from his older English contemporary William of Ockham, for example, in his willingness to embrace quantitative forms to explain the impenetrability of bodies.

In epistemology, Buridan is a representationalist who rejects the idea that human knowers possess any intuitive modes of cognition, that is, those which (following William of Ockham and John Duns Scotus) involve the direct and unmediated awareness of some present object (Zupko 1993a) (see Perception, epistemic issues in). Buridan

maintains that on the contrary, all cognition, including singular cognition, occurs when the intellect apprehends an object by means of a particular species or concept representing it. How we use the concept in our act of apprehension determines whether we cognize singularly or universally: singularly, if its object appears to us 'in the manner of something existing in the presence of the person cognizing it'; universally, if we focus on certain features of the concept to the exclusion of others, for example, if we use Socrates's humanity to understand 'all human beings, indifferently' (*Quaestiones super De anima*, third redaction, III, q.8).

Buridan touches on the issue of epistemic justification in the course of rebutting certain sceptical objections to the possibility of scientific knowledge (some of which are associated with his Parisian contemporary, Nicholas of Autrecourt) with naturalistic arguments based on the reliability of human cognitive processes. Thus, he contends that induction and sense perception are justified on the a posteriori grounds that both tend to produce beliefs that are free from doubt and error (although not infallible), and have verifiable results (Zupko 1993a). In general, however, Buridan is much more interested in explaining how we come to have knowledge than he is in exploring the grounds for knowledge claims, an orientation his work shares with most pre-Cartesian epistemology.

4 Natural philosophy

Historians of science have long recognized Buridan's contributions to the field of mechanics, most notably his development of the theory of impetus, or impressed force, to explain projectile motion (Grant 1977). Rejecting the discredited Aristotelian notion of antiperistasis, according to which the tendency of projectiles to continue moving is due to a proximate but external moving cause (such as the air), Buridan thought that only an internal motive force, transmitted from the original mover to the projectile, could explain its continued motion. His account of impressed force differs from others in that he did not see it as self-dissipating: '[this] *impetus*,' he says, 'would be of infinite duration, were it not diminished and corrupted by a contrary which resists it, or by something inclining it to a contrary [motion]' (*In Metaphysicam Aristotelis quaestiones* XII, q.9). He also views impetus as a variable quality whose strength is determined by the speed of and the quantity of matter in its moving subject, so that the acceleration of a falling body is understood in terms of its accumulation of successive increments of impetus. Buridan did not himself use the concept of impetus to revolutionize mechanics, however, since he remained an Aristotelian in other respects, for example in his assumptions that motion and rest are contrary states of bodies, and that the world is finite in extent.

Buridan's explanation of motion is a product of his approach to natural science, which is empirical in the sense that it emphasizes the evidentness of appearances, the reliability of a posteriori modes of reasoning and the application of certain naturalistic models of explanation to a wide range of phenomena. Against the notion that scientific knowledge must be demonstrative and its truths absolutely certain, Buridan argues that natural science, at least, concerns only what happens assuming the common course of nature. This stems from his theological assumption that it is always possible for an omnipotent deity to deceive us in ways we could never detect, tempered by his confidence (for which he often cites empirical evidence) that our powers of perceptual judgment and inductive inference are reliable enough to make 'the comprehension of truth with certitude possible for us' (*In Metaphysicam Aristotelis quaestiones* II, q.1). Likewise, Buridan is confident that certain models of explanation can be used to grasp a whole range of natural phenomena. Thus, the concept of impetus re-emerges in his psychology to explain the difference between occurrent and dispositional modes of intellectual cognition, as well as in his ethics where it is used to account for the relative ease with which virtuous people are able to perform good actions.

Buridan also made an important contribution to the fourteenth-century debate on infinity and continuity, arguing for the divisibilist view that continuous magnitudes are composed not of indivisible atoms but of parts divisible without end, and that the parts of a continuum constitute a potentially infinite set. Again, what is distinctive about Buridan's divisibilism is the way he defends it, by using semantic theory to explain the truth conditions of propositions containing (apparently non-referring) terms such as 'point', the meaning of the term 'infinite' in categorematic and syncategorematic contexts, and the use of the term 'possible' as a modal operator to express the notion of potential infinity (Thijssen 1985; Knuuttila 1992; Zupko 1993b). The mathematical and physical considerations prominent in other medieval discussions of continua occur secondarily for Buridan, almost as an afterthought (see Natural philosophy, medieval).

5 Ethics

Among the most influential of Buridan's commentaries on Aristotle was his *Quaestiones super decem libros ethicorum Aristotelis ad Nicomachum (Questions on the Ten Books of the Nicomachean Ethics)*, a work containing sophisticated accounts of the structure of the will, the nature of human freedom and the phenomenon of *akrasia*, or weakness of the will (see Akrasia). In moral psychology, Buridan appears to effect a compromise between the 'intellectualist' or 'naturalist' tradition associated with Aristotle (§§21-5) and Thomas Aquinas (§13), and the 'voluntarist' tradition of Augustine (§§12-14) and Franciscan thinkers such as Duns Scotus (§14) and William of Ockham (§10). Thus he contends (with the former) that human happiness ultimately consists in an intellectual act, 'the perfect comprehension of God', rather than in a volitional act such as willing or loving, although (with the latter) he emphasizes the role of the will as a self-determining power in achieving that end. The compromise appears to turn on Buridan's innovative concept of free choice, which builds upon Albert the Great's notion that moral certainty admits of degrees (Saarinen 1993) (see Albert the Great §5). Buridan's idea is that even if the will cannot choose evil as such, it still has the power to defer its choice whenever the goodness of the alternatives presented to it remains doubtful or uncertain: a move which, given our poor epistemic position in this life, seems to guarantee the will a certain amount of freedom in practice. The compromise is only apparent, however, because Buridan also insists that deferment is possible only if 'the intellect would judge it to be good to consider the matter further' (*Quaestiones super decem libros ethicorum* III, q.5), a claim which, together with his assumption that the will can choose non-optimally only through ignorance or impediment (q.3-5; 9), places deferment squarely in the province of the intellect, which weighs the relative goodness and badness of possible courses of action, including deferment. The voluntarist language Buridan appropriates to describe the will's freedom is thus used in an intellectualist sense, perhaps to dispel the cloud of heterodoxy that had surrounded several of the main principles of intellectualist moral psychology since the condemnations of 1277 (Zupko 1995).

Buridan's peculiar conception of freedom also provides the most plausible explanation of the use of his name in the example traditionally known as 'Buridan's ass', in which a donkey starves to death because it is unable to choose between two equidistant and equally tempting piles of hay. This particular example is unknown in Buridan's writings, although there are similar cases. The example probably originated as a parody of Buridan's theory introduced by later authors, who perhaps found absurd the idea that the will's freedom could consist in inaction, that is, in its ability to defer the act of either accepting or rejecting some proposed course of action (see Free will).

See also: Aristotelianism, medieval; Burley, W.; Gerard of Odo; Language, medieval theories of; Marsilius of Inghen; Nominalism; Universals; William of Ockham

JACK ZUPKO

List of works

Buridan, John (*c.*1300-58) *Summulae de dialectica (Summary of Dialectic), Perutile compendium totius logicae Ioannis Buridani cum praeclarissima solertissimi viri Ioannis Dorp expositione*, Venice, 1499; repr. Frankfurt: Minerva, 1965. (Incunabular edition of the *Summulae de dialectica*, with a commentary by John Dorp of Leiden.)

Buridan, John (c.1300-58) *Quaestiones super octo physicorum libros Aristotelis (Questions on Aristotle's Physics), Ioannis Buridani subtilissime quaestiones super octo physicorum libros Aristotelis*, Paris, 1509; repr. as *Kommentar zur Aristotelischen Physik*, Frankfurt: Minerva, 1964.(Incunabular edition of the questions on Aristotle's *Physics*.)

Buridan, John (*c.*1300-58) *Quaestiones super decem libros ethicorum Aristotelis ad Nicomachum (Question's on Aristotle's Nicomachean Ethics), Quaestiones Ioannis Buridani super decem libros ethicorum Aristotelis ad Nicomachum*, Paris, 1513; repr. as *Super decem libros ethicorum*, Frankfurt: Minerva, 1968.(Incunabular edition of the questions on Aristotle's *Nicomachean Ethics*.)

Buridan, John (*c.*1300-58) *Quaestiones super octo libros politicorum Aristotelis (Questions on Aristotle's Politics)*, Paris, 1513; repr. Frankfurt: Minerva, 1969.(Incunabular edition of the questions on Aristotle's *Politics*.)

Buridan, John (*c.* 1300-58) *In Metaphysicam Aristotelis quaestiones (Questions on Aristotle's Metaphysics), In Metaphysicen Aristotelis quaestiones argutissimae magistri Joannis Buridani*, Paris, 1588 (actually 1518); repr. as *Kommentar zur Aristotelischen Metaphysik*, Frankfurt: Minerva, 1964.(Incunabular edition of the questions on Aristotle's *Metaphysics*.)

Buridan, John (*c*.1300-58) *Quaestiones super libris quattuor De caelo et mundo (Questions on Aristotle's On the Heavens)*, ed. E.A. Moody, *Iohannis Buridani Quaestiones super libris quattuor De caelo et mundo*, Cambridge, MA: Medieval Academy of America, 1942.(Edition of the questions on Aristotle's *On the Heavens*.)

Buridan, John (*c*.1300-58) *Tractactus de consequentiis (Treatise on Consequences)*, ed. H. Hubien, *Iohannis Buridani Tractatus de consequentiis*, Louvain: Publications Universitaires de Louvain, 1976; trans. P. King, *Jean Buridan's Logic: The Treatise on Supposition; the Treatise on Consequences*, Boston: Reidel, 1985. (Edition and English translation of Buridan's study on inference.)

Buridan, John (*c*.1300-58) *Sophismata (Sophisms)*, ed. T.K. Scott, Jr, *Iohannis Buridani Sophismata*, Stuttgart: Frommann, 1977.(Translations include that of T.K. Scott, Jr., *John Buridan: Sophisms on Meaning and Truth*, New York: Appleton-Century-Crofts, 1966, and that in G.E. Hughes, *John Buridan on Self-Reference*, Cambridge: Cambridge University Press, 1982, Chapter 8.)

Buridan, John (*c.* 1300-58) *Quaestiones in Praedicamenta (Questions on Aristotle's Categories)*, ed. J. Schneider, *Iohannes Buridanus: Quaestiones in Praedicamenta*, Munich: Beck, 1983.(Edition of the questions on Aristotle's *Categories*.)

Buridan, John (*c.* 1300-58) *Quaestiones longe super librum Perihermeneias (Questions on Aristotle's De interpretatione)*, ed R. van der Lecq, *Johannes Buridanus: Questiones longe super librum Perihermeneias*, Nijmegen: Ingenium, 1983.(Edition of the questions on Aristotle's *De interpretatione*.)

Buridan, John (*c.* 1300-58) *Quaestiones in Aristotelis De anima (Questions on Aristotle's On the Soul)*, first redaction, ed. B. Patar, *Le Traité de l'âme de Jean Buridan [de prima lectura]*, Louvain-la-Neuve: Éditions de l'Institut Supérieur de Philosophie, 1991.(Edition of the first version of Buridan's questions on Aristotle's *On the Soul*. The reference in the text of this entry is to the as yet unpublished third redaction.)

Buridan, John (*c*.1300-58) *Tractatus de infinito (Treatise on Infinity)*, ed. J. M. Thijssen, *John Buridan's Tractatus de infinito*, Nijmegen: Ingenium, 1991.(Edition of this quasi-independent treatise, contained in Book III of the final version of *Quaestiones super octo physicorum libros Aristotelis*.)

Buridan, John (*c*.1300-58) *Tractatus de suppositionibus (Treatise on Supposition)*, ed. M.E. Reina, 'Giovanni Buridano: *Tractatus de suppositionibus*', *Rivista critica di storia della filosofia* 12, 1957: 175-208; 323-52; trans. P. King, *Jean Buridan's Logic: The Treatise on Supposition; the Treatise on Consequences*, Boston: Reidel, 1985.(Reina is an edition of Treatise IV of *Summulae de dialectica*; King is an English translation.)

Buridan, John (*c*.1300-58) *Quaestio de puncto (Question on Points)*, ed. V. Zoubov, 'Jean Buridan et les concepts du point au quatorzième siècle', *Mediaeval and Renaissance Studies* 5, 1961: 43-95.(Edition of Buridan's independent treatise on points.)

References and further reading

de Rijk, L.M. (1992) 'John Buridan on Universals', *Revue de Métaphysique et de Morale* 97 (1): 35-59.(Explains Buridan's semantic approach to the question of universals, as well as its similarities to that of Peter Abelard.)

Faral, E. (1950) *Jean Buridan: Maître dès Arts de l'Université de Paris (Jean Burdian: Master of Arts of the University of Paris)*, Extrait de l'Histoire littéraire de la France, Tome XXVIII, 2e partie, Paris: Imprimerie Nationale. (Still the best account of the many sources on Buridan's life and times.)

Grant, E. (1977) *Physical Science in the Middle Ages*, Cambridge: Cambridge University Press, 50-5.(Clear explanation of the application of Buridan's theory of impetus to the case of projectile motion.)

Hughes, G.E. (1982) *John Buridan on Self-Reference: Chapter Eight of Buridan's Sophismata with a Translation, an Introduction, and a Philosophical Commentary*, New York: Cambridge University Press.(Discusses Buridan's treatment of self-referential paradoxes.)

Klima, G. (1991) 'Latin as a Formal Language: Outlines of a Buridanian Semantics', *Cahiers de l'Institut du Moyen-Âge Grec et Latin* 61: 78-106.(Offers a model-theoretic construction of key elements in Buridan's semantic theory.)

Klima, G. (1992) '"*Debeo tibi equum*": A Reconstruction of the Theoretical Framework of Buridan's Treatment of the Sophisma', *S - European Journal for Semiotic Studies* 4: 141-59.(Formalizes the role of the doctrine of appellation in Buridan's analysis of intentional verbs.)

Knuuttila, S. (1989) 'Natural Necessity in John Buridan', in S. Caroti (ed.) *Studies in Medieval Natural Philosophy*, Florence: Olschki.(Investigates Buridan's application of the statistical interpretation of modality to natural possibilities and nomic natural necessities.)

Michael, B. (1985) 'Johannes Buridan: Studien zu seinem Leben, seinen Werken und zur Rezeption seiner Theorien im Europa des späten Mittelalters' (John Buridan: Studies on His Life, His Work and the Reception

of His Theories in the Later Middle Ages), Ph.D. dissertation, University of Berlin.(Authoritative inventory and overview of the complicated textual tradition surrounding Buridan's writings.)

Normore, C. (1985) 'Buridan's Ontology', in J. Bogen and J.E. McGuire (eds) *How Things Are: Studies in Predication and the History and Philosophy of Science*, Dordrecht: Reidel, 189-203.(Discusses the motivation for Buridan's realism about modes.)

Pinborg, J. (ed.) (1975) *The Logic of John Buridan*, Acts of the Third European Symposium on Medieval Logic and Semantics, Copenhagen: Museum Tusculanum.(Contains ten studies on different aspects of Buridan's logic.)

Saarinen, R. (1993) 'John Buridan and Donald Davidson on *Akrasia*', *Synthese* 96 (1): 133-54.(Discusses Buridan's explanation of the phenomenon of weakness of will, including his indebtedness to Albert the Great.)

Spade, P.V. (1978) 'John Buridan on the Liar: A Study and Reconstruction', *Notre Dame Journal of Formal Logic* 19 (4): 579-90.(Formalizes Buridan's solution to the liar paradox in a way that is sensitive to interpretive difficulties in the relevant texts.)

Thijssen, J.M. (1985) 'Buridan on Mathematics', *Vivarium* 23 (1): 55-77.(Explains and illustrates Buridan's logico-semantic approach to mathematics.)

Walsh, J.J. (1980) 'Teleology in the Ethics of Buridan', *Journal of the History of Philosophy* 18 (3): 265-86. (Examines the intellectualist and voluntarist aspects of Buridan's moral teleology.)

Zupko, J. (1993a) 'Buridan and Skepticism', *Journal of the History of Philosophy* 31 (2): 191-221.(Explains Buridan's response to scepticism as a consequence of his naturalistic epistemology.)

Zupko, J. (1993b) 'Nominalism Meets Indivisibilism', *Medieval Philosophy and Theology* 3: 158-85.(Compares Buridan's divisibilist account of continuous magnitudes with those of Ockham and Adam Wodeham in the context of their replies to the indivisibilist argument that a sphere coming into contact with a plane surface must touch it at an indivisible point.)

Zupko, J. (1995) 'Freedom of Choice in Buridan's Moral Psychology', *Mediaeval Studies* 57: 75-99.(Argues that the novel Buridanian act of deferment does not represent a voluntarist departure from what would otherwise be a straightforwardly intellectualist theory of volition, because as Buridan presents it, deferment does not make the will any more autonomous or free.)

Burke, Edmund (1729-97)

Edmund Burke's philosophical importance lies in two fields, aesthetics and political theory. His early work on aesthetics, the Philosophical Enquiry into the Origin of our Ideas of the Sublime and Beautiful (1757), explored the experiential sources of these two, as he claimed, fundamental responses, relating them respectively to terror at the fear of death and to the love of society.

Active in politics from 1759, and Member of Parliament from 1765, he wrote and delivered a number of famous political pamphlets and speeches, on party in politics - Thoughts on the Causes of the Present Discontents (1770), on the crisis with the American colonies - On Conciliation with America (1775), on financial reform and on the reform of British India - Speech on Mr Fox's East India Bill (1783). While clearly informed by a reflective political mind, these are, however, pièces d'occasion, *not political philosophy, and their party political provenance has rendered them suspect to many commentators.*

His most powerful and philosophically influential works were written in opposition to the ideas of the French Revolution, in particular Reflections on the Revolution in France (1790), which has come to be seen as a definitive articulation of anglophone political conservatism. Here Burke considered the sources and desirability of social continuity, locating these in a suspicion of abstract reason, a disposition to follow custom, and certain institutions - hereditary monarchy, inheritance of property, and social corporations such as an established Church. His Appeal from the New to the Old Whigs (1791) insisted on the distinction between the French and Britain's revolution of 1688; while his final works, Letters on a Regicide Peace (1795), urged an uncompromising crusade on behalf of European Christian civilization against its atheist, Jacobin antithesis.

1 Early work

Edmund Burke was born on 12 January 1729, of Anglo-Catholic parents, in Dublin. As a young man at Trinity College his early aspirations were literary. He produced a periodical - *The Reformer* - devoted, in the manner of Addison and Steele's *Spectator*, to the reformation of taste as part of a self-conscious moral and political programme, and he retained a concern with the role of manners and taste in sustaining a viable and stable political society. Leaving Dublin to read law at the Temple Inn, he settled in London. There he published a philosophical satire, *A Vindication of Natural Society* (1756), followed a year later by *A Philosophical Enquiry into the Origin of our Ideas of the Sublime and Beautiful*. This achieved prominence for him both in London society, where he moved in the literary and artistic coterie of Samuel Johnson, David Garrick and Joshua Reynolds, and abroad. He was contracted to edit the *Annual Register* from 1758, for which he also wrote an annual political survey and the book reviews.

Burke's conservatism was founded on a deep and often explicit mistrust of metaphysics and abstraction, which for some commentators has rendered doubtful his status as a true political theorist, and even more so as a philosopher. The description of his writings as philosophical is further called into question by their occasional, polemical and often party-political character. However, these early mature writings - and a surviving notebook - reveal a pervasive sceptical epistemological position which is entirely consistent with, and arguably underpins his later political theory. In *A Vindication of Natural Society* Burke satirized the confident rationalism which the First Viscount Bolingbroke, in his posthumously published *Philosophical Works* (1755), had applied critically to religion. The *Vindication* was an ironic 'defence' of the state of nature, exercised through a hyperbolically rationalist critique of existing society, in a kind of *reductio ad absurdum* of Deist arguments (see Deism). A preface to the second edition clarified his concern to warn against the terrible consequences of applying human reason to areas beyond its very limited competence.

A year later his *Philosophical Enquiry* was published; seventeen English editions were produced in Burke's lifetime and the work was an important influence on European aesthetics. It was translated and discussed by Gotthold Ephraim Lessing in his *Bemerkungen über Burkes philosophische Untersuchungen (Remarks on Burke's philosophical enquiries)* (1758), absorbed by Denis Diderot, and was a significant influence on Immanuel Kant (see Beauty; Sublime, The). French and German translations were published in 1765 and 1773 respectively. The work was a resolutely empiricist enquiry into the experiential sources of our responses to the sublime and the beautiful rooted in Lockean sensationist psychology (see Locke, J.). Burke rejected the Third Earl of

Shaftesbury's claim to be able to derive moral or aesthetic standards from empirical observation, and implicitly criticized Francis Hutcheson in rejecting the ascription of a new 'sense' for each dimension of human response (see Shaftesbury, Third Earl of; Hutcheson, F.). He allied himself methodologically to Isaac Newton in seeking only to describe the phenomena actually observed or experience gained through introspection, and not going beyond that to hazard underlying general principles or forces at work.

Burke saw the sources of the beautiful in the social and pleasurable, those of the sublime in a vicarious appreciation of that which threatens self-preservation, and pain. It was thus, in an important sense, as Burke claimed, a 'theory of the passions' rather than a narrowly aesthetic work, and he stressed its relevance for all persons concerned with 'affecting the passions'. Burke was critical of a number of contemporary theories of beauty, including those neoclassical, utilitarian or Platonic theories which related it to 'proportion', 'fitness' or 'goodness'. Beauty was evoked by the soft, the smooth, and the delicate, and its affective power was by far the weaker of the two responses. Unsurprisingly, as has been pointed out by critics since Dugald Stewart, 'the idea of female beauty was uppermost in Mr Burke's mind when he wrote', and a particular conception of female beauty too: Burke's aesthetic is heavily gendered, a source of comment both for the modern psychobiographer and for feminist critics.

Burke's greater originality lay in his account of the much stronger principle of the sublime. Here, controverting established neoclassicism, Burke stresses the greater power of ideas that are obscure, indistinct and infinite. He anticipates romanticism in stressing the more intense emotional effects produced by imprecise images - and therefore by words, rather than painting, which suggest without delineating. He explains this power through the sublime's capacity to 'over-take our reason', which in turn derives from the inability of reason, on a Lockean account conceived of as the comparison of ideas, to operate where ideas are not (as the sublime never is) 'clear and distinct'. The power of the sublime is linked to religion and politics. Burke follows Hutcheson in drawing attention to the sublime effects produced by the obscurity of 'heathen temples' and the social distancing of political leaders, but goes beyond him in developing a distinction between the social virtues, which are associated with the beautiful response, and the political virtues, linked to austerity, power and the sublime.

2 The party-political pamphleteer

Burke's commitment to a political career began when he entered the service of Lord Rockingham, the opposition Whig leader, in 1765 and took a seat in Parliament later that same year. His most famous political pamphlets and speeches from that period include *Thoughts on the Causes of the Present Discontents* (1770) on party in politics; *On Conciliation with America* (1775) on the crisis with the American colonies; and *Speech on Mr Fox's East India Bill* (1783) on financial reform and on the reform of British India.

The political pamphlets are extremely issue-oriented, and invoke, rather than articulate, a conception of politics. Joining a century-long debate, Burke defends embryonic party association - 'when bad men combine, the good must associate' - on the basis of principled opposition, against the ministerial view that opposition parties were at the least 'factious', and possibly disloyal. His defence of party grouping *within* Parliament contrasts with his deprecation of it outside. In a famous speech to his Bristol constituents he rejected their right to organize so as to 'instruct' MPs, articulating a famous distinction between a representative and a delegate. Here as elsewhere Burke is looked back on as a founder or contributor to a Whiggish developmental account of constitutional practice, which it is not clear the historical record will sustain. In his speeches and writings on America he pursued a policy of reconciliation that stressed the futility of making stands on issues of principle which might never emerge, and could in any case never be enforced. His Economical Reform Bill of 1780, on the other hand, did pursue a principled distinction - between the King's private household finances and those of the Exchequer - seeking, for reasons widely urged, to extinguish the public financial sources of the Crown's influence and patronage threatening parliamentary independence.

These works fit broadly within an ideological position which derives in part from moderate oppositionist thought and Court Whig sentiments. Burke accepted the institutions of political and economic modernization, the importance of safeguarding credit and the moral and political role of a state Church - more especially in times of crisis. However, like opposition Whigs he wanted to limit the Crown's control of patronage through places and supply beyond parliamentary control, and he was anxious about the moral and political effects of Empire - especially where, as in the case of India, through enabling the creation of large fortunes which could be used to

buy political influence, and providing the experience of unconstitutionally restrained power, it threatened liberty at home. Another source of political destabilization in this political orthodoxy was insecurity of property. Despite Burke’s acceptance of modernizing Whig institutions - the Bank of England, a national debt and a market in stock and bonds - he was acutely aware of the potentially destabilizing effects of property speculation. The image of gambling as destructive of both individual rationality and social coherence haunted the eighteenth century since at least the South Sea Bubble, and Burke draws on all these resources when he comes to denounce in *Reflections on the Revolution in France* (1790) the Revolutionaries’ political economy and for their having reduced France to ‘one great play table’ and produced a people whose ‘impulses, passions, and superstitions [are] of those who live on chance’.

Despite their congruities with well-established patterns of argument, neither these, nor the works of his later, French revolutionary period, are in any sense works of systematic philosophy, as the *Philosophical Enquiry* is. Burke’s wider political position emerges episodically in parenthetical observations, epigrams and observations, which engage and resonate with a number of important political languages and issues of his time, and which only contentiously, and with some effort at reconstruction, constitute a political philosophy. One consequence of this is a considerable diversity of interpretation about Burke’s precise - or even general - theoretical location. Within the last hundred years Burke has been characterized as a utilitarian, a radical Lockean, a Court Whig, a natural law theorist, a common law theorist, and a proto-romantic.

3 Philosophical orientation

Even in the more theoretically-oriented political works is a series of issues of philosophical significance and in the treatment of which a broad continuity can be seen with Burke’s earlier works. Central among these are his fear of atheism; his conception of the political order as resting on an ‘opinion’ which is potentially unstable and a product of cultural and practical experiences; and a defence of tradition, both in substance, and on grounds sceptical of other sources of knowledge. In each of these cases it is an abstract ‘reason’ which Burke sees as the threat. Pride in the power of abstract reason fuelled the deist attacks on Christianity both in his youth and in Jacobin France. Philosophical speculation, especially when taken up by the politically inexperienced, detached people from valuable socialized dispositions and created political expectations and opinions which were not only unrealistic, in the sense of being cut adrift from the empirically validated institutions, practices and customs of the society, but positively harmful. For such abstract principles seemed to justify not merely the adjustment of institutions and practices to circumstance - a necessary condition, he recognized, for a tradition to incorporate and accumulate experience - but their reconstruction *de novo*. Such rationalist appraisal of tradition undermined the presumption in its favour necessary for its preservation, for although tradition provided real benefits in terms of social continuity and stable expectations, Burke believed that such a defence could be articulated with only the greatest of difficulty to the enthusiastic rationalist mind.

Burke’s response to the French Revolution (1790) was directed in particular at the language of radical natural rights, and the fear of its being imported to England. But the occasion of that fear was the Unitarian Richard Price’s sermon ’On the Love of our Country’, in which Price compared the deficiencies of Britain’s Glorious Revolution of 1688 to what the French were succeeding in achieving. Price’s religious rationalism was of the stamp Burke had attacked in his early *Vindication of Natural Society*. Radical deists (which Burke judged Unitarians to be) were now applying in earnest to politics those very arguments of religious rationalists which Burke had attacked in his youth when he had applied them ironically to politics. Natural right, far from providing a continuing standard of probity in political society, was inconsistent, Burke urged in Hobbesian fashion (see Hobbes, T.), with its very existence - the inconsistency being amply demonstrated by the fact that it is only in virtue of surrendering our natural rights that we can enjoy our civil rights. Conversely, in reasserting the former we effectively dissolve the social bonds. While this account seems to entail some version of social contract theory, Burke’s conception of the emergence of society is so historicized that such a contract cannot be the act of any one set of natural agents, but is rather a partnership across time and between generations. The corollaries drawn are that no one generation has the right to abrogate the contract, and exclude future generations from benefiting thereby, and moreover, that if such an abrogation were to occur, the re-establishment of society would be not an act, but a long and hazardous historical process.

Burke’s response to Price’s invitation to imitate France comprised, as well as an attack on radical natural rights

doctrine, a more general attack on abstract theory in politics, and an argument about English historiography and the status of 1688 - continued at some length a year later in his *Appeal from the New to the Old Whigs* (1791). To natural right, Burke opposed tradition; to rationalism, sentiment; and to the interpretation of 1688 as a failed revolution, the view of it as a desperate bid to sustain continuity - ‘a revolution not made but prevented’.

Burke had a complex conception of tradition owing much to Edward Coke, Hale and the English common law tradition. Custom - and even ‘prejudice’ - comprising the accumulated adjustment to circumstance of many generations, represented a kind of collective rationality which individuals cannot create. There was also, for Burke a sense in which, rational or not, the sustenance of such collective practices over time - or even the belief that they have been so sustained - constitutes the identity of the society.

Burke’s more general arguments against attempts to use rational criteria for social criticism focus on a number of deficiencies he had long identified in rationalism. A major feature was the indeterminacy of reason when not embedded in social practice. This insistence on social embeddedness has been a feature of much recent political debate, not only about values and principles but about personality itself. Burke here presages Hegel, as well as modern communitarians such as Ernest Mandel and Alasdair MacIntyre (see Community and communitarianism). For Burke, assessing the consequences of any social institution apart from an actual practice in its historical context is extremely hazardous, since its consequences and those of its fit with other institutions are unpredictable. The presumption in favour of natural right over positively existing institutions is a recipe for anarchy, for it represents an attempt to dissociate individuals from those shared beliefs and conventions which alone make social life possible. Even the assumption of majority decision-making, Burke points out, so far from being natural is a ‘violent fiction of positive law’. The indeterminacy of rationalism was psychological as well as logical: at an individual level moral rationalism left the individual perplexed as to how to apply a rule, whereas ‘prejudice renders a man’s virtue his habit’ and so provides an unhesitating and culturally appropriate response.

A second deficiency in reason - already famously stressed by Hume - was its ineffectual properties as a motivator. To replace traditional institutions and beliefs with rationalist schemes, however good, risked the failure of a people to adhere to them. The prejudicial partiality for one’s immediate family and locality, for known and accepted patterns of life, was the basis of any moral or political reality, both psychologically and in extent. The love of family and the ‘little platoon’ was the affective basis of ever-widening social circles through which we ultimately conceived affection for our country. By contrast, Price’s conditional patriotism, and the internationalism of the revolutionaries, Burke characterized as cold and ultimately ineffectual, leading to a love of humankind but not of any individuals. Moreover, the political consequences of abandoning custom and the motivating sentimental attachments to it would, Burke thought, be the increasing need to resort to coercion.

Burke identified British political culture as peculiarly suited to avoiding the intrusion of a-prioristic reasoning; it is praised as comprising a proper and continuing predisposition to conserve. The events of 1688 thus represented the successful avoidance of constitutional discontinuity at the price of some minor deviation in the line of descent. The principle of inheritance - in the monarchy and the Lords, as well as in the private property of family estates - links an important natural process - that is, reproduction - and natural sentiments - that is, family affection - to social practices conducive to continuity, and so to the epistemological collective benefits outlined above. Religion plays an important role in underpinning respect for these institutions, which might not otherwise be forthcoming, or might otherwise be undermined by a variety of processes at work in society, including moral or religious scepticism.

The French Revolution thus not only promulgated a political doctrine which Burke believed to be fundamentally flawed, but its actions - in undermining the Church and destabilizing property and even the currency - were destroying those institutions participation in which sustained shared patterns of belief and knowledge that comprised the community. Burke successfully predicted that a dictatorship would emerge from the atomized, institutionally dislocated or uprooted society.

Crucial to Burke’s pessimism is an Augustinian sense of humankind’s fallen and imperfect nature, from which we are preserved only by the painfully acquired veneer of civilization, which can be very quickly lost. Rejecting the Rousseauean art-nature polarity (see Rousseau, J.-J.), Burke affirms that ‘art is man’s nature’. It is nevertheless a second nature that is only acquired with care over a long time, and failure to acquire or sustain it reveals a savage and violent nature of a Hobbesian kind, which, Burke insists, will result from the reassertion of natural right.

Throughout the mid twentieth century Burke's thought was regularly appealed to by practising conservative politicians in Britain, and by American Cold-War ideologues. While much recent conservative thought has been so organized around the role of the market, as, in Burke's terms, to be the application of 'abstract' principle, a thinker such as Michael Oakeshott can be seen as operating in a recognisably Burkean intellectual environment.

See also: Conservatism; Contractarianism; Revolution; Rights; Tradition and traditionalism

IAIN HAMPSHER-MONK

List of works

Burke, E. (1886) *The Works of the Right Honorable Edmund Burke*, London: Bell.

(1981-) *The Writings and Speeches of Edmund Burke*, ed. P. Langford, Oxford: Oxford University Press.(In preparation.)

(1757) *A Philosophical Enquiry into the Origin our Ideas of the Sublime and Beautiful*, ed. J.T. Boulton, London and New York: Routledge & Kegan Paul; Columbia University Press, 1958.(This contains an excellent historical and critical introduction.)

References and further reading

Furniss, T. (1993) *Edmund Burke's Aesthetic Ideology*. Cambridge: Cambridge University Press.

Kramnick, I., (1977) *The Rage of Edmund Burke*, New York: Basic Books.(Explores the gender dimension of Burke's personality and political theorising.)

Lessing, G.E. (1758) *Bemerkungen über Burkes philosophische Untersuchungen (Remarks on Burke's philosophical enquiries)*(Discussion of Burke's *Philosophical Enquiry*.)

O'Brien, C.C. (1992) *The Great Melody*, London: Sinclair-Stevenson.(A very full and reflective biography.)

Pocock, J.G.A. (1960) 'Burke and the Ancient Constitution', *Historical Journal*, 3 (2).

Pocock, J.G.A. (1982) 'The Political Economy of Burke's Analysis of the French Revolution', *Historical Journal*, 25 (2); repr. in J.G.A. Pocock, *Virtue Commerce and History*, Cambridge: Cambridge University Press, 1985.

Wilkins, B.T. (1967) *The Problem of Burke's Political Philosophy*, Oxford: Oxford University Press.(The most cautious of the 'natural law' interpretations of Burke.)

Burley, Walter (*c*.1275-*c*.1345)

Active in the first half of the fourteenth century, Burley received his arts degree from Oxford before 1301 and his doctorate in theology from Paris before 1324. At one time a fellow of Merton College, he - along with Thomas Bradwardine, Richard Kilvington and others - became a member of the household of Richard de Bury, Bishop of Durham and served several times as envoy of the King of England to the papal court. Despite his extra-university activities, Burley continued to compose Aristotelian commentaries and to engage in disputations to the end of his life. A clear and prolific writer, Burley has been labelled an 'Averroist' and a 'realist' because of his arguments against Ockham, but it would perhaps be more accurate to see him as a middle-of-the-road Aristotelian whose intellectual activity coincided with the transition between the approaches of Thomas Aquinas and Duns Scotus on the one hand and those of William of Ockham and the Oxford Calculators on the other.

Born around 1275, Walter Burley became an Oxford Master of Arts sometime before 1301. He was a fellow of Merton College, Oxford in 1305, but by 1310 he was a student of theology at Paris. A colophon dated 1324 cites Burley as a doctor of theology. In 1327 Burley was appointed King Edward's envoy to the papal court, where he was again in 1343. In between, he was a member of the household of Richard Bury, Bishop of Durham, and conducted a disputation in Bologna in 1341. He probably died soon after 1344, when he is last mentioned.

Although Burley became a doctor of theology, he continued to write philosophical works throughout his life, commenting on all of Aristotle's logical works and on many of his other works as well, sometimes more than once. He also wrote independent treatises on philosophical subjects. Of his commentary on Peter Lombard's *Sentences* there is no known copy, nor are other explicitly theological works ascribed to him (although the *Tractatus primus* probably resulted from Burley's 'Principium' on the *Sentences*). His *De puritate artis logicae (On the Purity of the Art of Logic)* exists in a shorter version written about 1324 and a longer version written in 1325-8, responding to Ockham's *Summa logicae* of 1324. His commentary on Aristotle's *Physics* survives in manuscript in a version written before 1316; revised versions were printed many times, no doubt because of the commentary's completeness and clarity. Accurate, extensive quotations from earlier commentators are included.

In natural philosophy, Burley is best known for having held the 'succession of forms' theory of intension and remission, which treats qualitative forms, such as whiteness or heat, as similar to substantial forms, such as humanity. Each form is thought of as corresponding to one degree of quality. As a body is heated, at each instant it loses the degree of heat it had and acquires a greater degree. There are infinitely many indivisible degrees of heat, and between any two degrees there are other degrees. Godfrey of Fontaines had held a similar theory in explaining augmentation and diminution. In the later fourteenth century, most natural philosophers followed not Godfrey and Burley, but Duns Scotus, who had developed the so-called 'addition theory' of intension and remission: when a quality is intensified new parts are added, which produce a higher degree. Thus the previous degree of quality is not lost, as in the succession theory, but becomes part of the new degree. However, as long as the issue of how to explain the intension and remission of forms continued to be discussed, the succession theory was nearly always taken seriously.

Burley's treatise on first and last instants represents a standard view. For Burley, as for most other fourteenth-century authors, 'permanent' entities - that is, those that are wholly existent at once, unlike motion - have a first instant of being but no last instant. When water is heated and becomes air, there is a first instant at which the air exists but no last instant of the water. On the other hand, for 'successive' entities such as motions, there is neither a first instant at which the motion exists, nor a last instant of motion before the body comes to rest. Combining his ideas of first and last instants and his ideas of intension and remission, Burley argued in the *Tractatus primus* that all degrees of heat must be of the same species, rejecting such separate, contrary species as hotness, temperateness and coldness.

William of Ockham and other fourteenth-century nominalists such as John Buridan developed physical theories that minimized the types of entities in the world, maintaining that there are only substances and qualities unless there are special reasons for relaxing this rule. Burley, on the other hand, seems not to have been motivated by any preference for ontological parsimony. Thus, for Ockham, motion cannot exist separately from the moving body. Alteration as a kind of motion is simply the body altered and the degrees of quality (*forma fluens*) that it

successively takes on. Buridan adhered to Ockham's view for alteration, but relaxed it for local motion. If local motion were not something more than the body and the places it successively takes on, Buridan argued, then if God rotated the cosmos as a whole and there was nothing at rest to act as a reference point, it would be impossible to distinguish the cosmos in rotation from the cosmos at rest. Finding this conclusion unacceptable, Buridan concluded that in local motion there must be a flux (*fluxus formae*) as well as the places occupied. Burley, when confronted with similar questions, did not even try to develop a theory of motion that did not involve a *fluxus formae* as well as a *forma fluens*, but instead easily accepted the assumption that there is a *fluxus formae*. Likewise, while the nominalists emphasized that distances can exist and be measured only on bodies, so that when bodies move, the distances measured on them must likewise change, Burley took it as unproblematic that when two bodies are at rest, the distance between them remains the same, whether or not there is an unchanging body on which to measure the distance.

Because Burley belonged to no religious order, he has had no modern confrères interested in perpetuating a knowledge of his views. In the current state of scholarship, Burley appears more often as someone holding views contrasting with those of the nominalists than as someone whose theories are of interest for their own sake. Yet Burley's works were widely used, even after the advent of printing, probably more because of their thoroughness and clarity than because of the unique positions that Burley upheld. Burley's work would repay further study as a source of information about the views that were typical of early fourteenth-century scholasticism.

See also: Buridan, J.; Language, medieval theories of; Natural philosophy, medieval; William of Ockham

EDITH DUDLEY SYLLA

List of works

For a more complete list, see Weisheipl (1969) or Wood (1988b). Wood lists 50 surviving authentic works plus 17 lost, doubtful or spurious works. Included here are longer works and works that have received modern editions. For many more details on manuscripts and early printed editions of Burley's Aristotelian commentaries, including many works not listed here, see C. Lohr (1988) 'Medieval Latin Aristotle Commentaries', *Traditio* 24: 171-87.

Burley, W. (1301) *Quaestiones in Librum Perihermeneias (Questions on the Perihermeneias)*, ed. S. Brown, 'Walter Burley's Quaestiones in Librum Perihermeneias', *Franciscan Studies* 34, 1974: 200-95.(Twenty years later, Burley composed his so-called Middle Commentary on the *Perihermeneias* (ed. S. Brown, 'Walter Burley's Middle Commentary on Aristotle's Perihermeneias', *Franciscan Studies* 33, 1973: 42-134).)

Burley, W. (*c*.1302) *Obligationes (Obligations)*, part trans. in N. Kretzmann and E. Stump, *Logic and Philosophy of Language*, The Cambridge Translations of Medieval Philosophical Texts vol. I, Cambridge: Cambridge University Press, 1988, 370-412.(Burley's treatment of the common scholastic disputational exercise called 'obligations'.)

Burley, W. (*c*.1302) *Insolubilia (Insolubles)*, ed. M.L. Roure, 'La Problématique des propositions insolubiles au XIIIe siècle et au début de XIVe, suivie de l'édition des traités de W. Shyreswood, W. Burley and Th. Bradwardine', *Archives d'histoire doctrinale et littéraire du moyen âge* 37, 1971: 262-84.(Related to the previous work and perhaps composed at about the same time, this work treats those propositions that seem to falsify themselves, such as 'I am telling a lie.')

Burley, W. (early version pre-1316; later versions 1324-7 and 1334-7) *In Physicas Aristotelis. Expositio et Questiones (On Aristotle's Physics: Exposition and Questions (various revisions and titles))*, Pavia, 1488; Venice, 1482, 1491, 1501 (repr. Hildesheim: Olms, 1972), 1508; Bologna, 1589.(Existing in several versions, Burley's exposition of and questions on Aristotle's *Physics* combined a detailed passage-by-passage explanation of Aristotle's text, and often of Averroes' commentary on it, with occasional digressions to answer questions or doubts.)

Burley, W. (before *c*.1316?) *Tractatus expositorius super libros posteriorum Aristotelis (Expository Treatise on the Books of Aristotle's Posterior Analytics)*, Venice, 1497, 1504; Oxford, 1517; repr. of edition with Robert Grosseteste's *In Aristotelis Posteriorum analyticorum libros*, Frankfurt: Minerva, 1966.(Burley's commentary on the *Posterior Analytics*.)

Burley, W. (before 1320?) *Tractatus de suppositionibus (Treatise on Supposition)*, ed. S. Brown, 'Walter Burleigh's Treatise *De Suppositionibus* and its influence on William of Ockham', *Franciscan Studies* 32, 1972: 15-64.(Burley's work on a central concept of the new logic.)

Burley, W. (1320-3) *Tractatus de formis (Treatise on Forms)*, ed. F.J.D. Scott, Veröffentlichungen der Kommission für die Herausgabe ungedruckter Texte aux der mittelalterlichen Geisteswelt, Bd. 4, Munich: Verlag der Bayerischen Akademie der Wissenschaften, 1970.(One of Burley's shorter works that has received

a modern edition, this provides a general overview of Burley's theory of forms. Various other short, mostly logical works have been edited by H. Shapiro and various collaborators; see Murdoch and Sylla (1970) for details.)

Burley, W. (1320-7) *Tractatus primus* or *Tractatus de activitate, unitate et augmento formarum activarum habaentium contraria et suscipientium magis et minus (First Treatise, or Treatise on the Activity, Unity and Augmentation of Active Forms Having Contraries and Undergoing More or Less)*, MSS Bruges, Stadsbibl. 501, 70r-105r; Vat. lat. 817, ff. 203-223r.(Exists only in manuscript form. Contains Burley's theory of the succession of forms in its formative stages.)

Burley, W. (after 1327?) *Tractatus secundus or Tractatus de intensione et remissione formarum (Second Treatise, or Treatise on the Intension and Remission of Forms)*, Venice, 1496.(The mature statement of Burley's succession of forms theory and of its justifications. For a discussion, see Maier (1968).)

Burley, W. (1320s?) *Utrum contradicto sit maxima oppositio (Whether Contradiction is the Maximum Opposition)*, ed. R. Palacz, 'Gualterii Burleii Quaestio: Utrum contradicto sit maxima oppositio (MS. Vat. Ottob. 318, f. 141va-145vb)', *Mediaevalia Philosophica Polonorum* XI, 1968, 128-39, 152-6; ed. L.M. de Rijk, 'Burley's So-called *Tractatus Primus*, with an Edition of the Additional Quaestio "Utrum contradicto sit maxima oppositio"', *Vivarium* 34, 1996: 161-91.(Brief logical work.)

Burley, W. (1320s?) *De exclusivis (On Exclusions)*, ed. L.M. de Rijk, 'Walter Burley's Tract *De exclusivis*: An Edition', *Vivarium* 23, 1984: 23-54.(Brief logical work.)

Burley, W. (1320s?) *De exceptivis (On Exceptions)*, ed. L.M. de Rijk, 'Walter Burley's *De exceptivis*, an Edition', *Vivarium* 24, 1986: 22-49.(Brief logical work.)

Burley, W. (shorter version before 1324; longer version 1325-8) *De puritate artis logicae (On the Purity of the Art of Logic)*, ed. P. Boehner, St. Bonaventure, NY: Franciscan Institute, 1955; section on consequences trans. in N. Kretzmann and E. Stump, *Logic and the Philosophy of Language*, The Cambridge Translations of Medieval Philosophical Texts vol. I, Cambridge: Cambridge University Press, 284-311.(A typical fourteenth-century textbook of new logic, which may be compared to Ockham's logical works to see the points on which fourteenth-century logicians tended to agree or disagree with each other.)

Burley, W. (before 1327) *De primo et ultimo instanti (On the First and the Last Instant)*, ed. H. Shapiro and C. Shapiro, *Archiv für Geschichte der Philosophie* 47, 1965: 157-73.(This is a working rather than critical edition of Burley's ideas concerning first and last instants.)

Burley, W. (*c*.1333-41) *Exposition super libros Ethicorum Aristotelis (Exposition of the Books of Aristotle's Ethics)*, Venice, 1481, 1500.(Towards the end of his life, Burley continued to compose commentaries on the major works of the Aristotelian corpus.)

Burley, W. (final version 1337) *Expositio super artem veterem Porphyrii et Aristotelis (Exposition of the Old Art of Porphyry and Aristotle)*, Venice, 1481, 1497 (repr. Frankfurt: Minerva, 1967), 1519 and many others.(A commentary on the old logic.)

Burley, W. (1340-3) *Expositio super librum Politicorum (Exposition on the book of Politics)*.(Exists in many manuscripts, though relatively neglected by recent scholars.)

pseudo-Burley (?) *De vita et moribus philosophorum (The Lives and Customs of the Philosophers)*.(Previously thought to be by Burley; see J. Prelog (1990) 'De Pictagora phylosopho. Die Biographie des Pythagoras in dem Walter Burley zugeschriebenen Liber de vita et moribus philosophorum,' *Medioevo. Rivista di storia della filosofia medievale* 16: 191-251.)

References and further reading

Jung-Palczewska, E. (1986) 'Le problème d'averroïsme de Walter Burley dans son Commentaire sur la "Physique"' (The Problem of the Averroism of Walter Burley in his Commentary on the *Physics*), *Studia Mediewistyczne* 24: 101-9.(Argues that Burley was not an Averroist, but rather tried to expound Aristotle's meaning, Averroes' meaning where relevant, and then in some cases to give his own opinion, which was likely to follow Catholic views when these diverged from Aristotle (for instance, on the eternity of the world or on free will).)

Kitchel, M.J. (1971) 'The *De Potentiis Animae* of Walter Burley', *Medieval Studies* 33: 85-113.(See for Burley's views on the human intellect and soul.)

Maier, A. (1955) 'Ein unbeachteter "Averroist" des XIV Jahrhunderts: Walter Burley', (An Unnoticed 'Averroist' of the Fourteenth Century: Walter Burley), *Medioevo e Rinascimento: Studi in onore di Bruno Nardi*, Florence; reprinted in A. Maier, *Ausgehendes Mittelalter: Gesammelte Aufsätze zur Geistesgeschichte des 14. Jahrhunderts*, 2 vols, Rome: Edizioni di Storia e Letteratura, 1964-7.(Puts Burley's work in its fourteenth-century context.)

Maier, A. (1968) *Zwei Grundprobleme der scholastischen Naturphilosophie (Two Fundamental Problems of Scholastic Natural Philosophy)*, Rome: Edizioni di Storia e Letteratura, 3rd edn.(See the chapter 'Nachträge zu Walter Burleys Traktat De intensione et remissione formarum'. As always for fourteenth-century natural philosophy, Maier's work is fundamental.)

Martin, C. (1964) 'Walter Burley', in *Oxford Studies Presented to Daniel Callus*, Oxford: Clarendon Press, 194-230.(Comprehensive collection of the biographical information available about Burley.)

Murdoch, J.E. and Sylla, E.D. (1970) 'Burley, Walter', in *Dictionary of Scientific Biography*, New York: Charles Scribner's Sons, vol. 2, 608-12.(Provides more detail about Burley's work insofar as it is relevant to the history of science.)

Sylla, E.D. (1988) 'Walter Burley's *Tractatus Primus*: Evidence Concerning the Relations of Disputations and Written Works', *Franciscan Studies* 44: 257-74.(Deals with Burley's *Tractatus primus* and the various opponents whose opinions are considered in that work.)

Una Juarez, A. (1978) *La Filosofia del Siglo XIV. Contexto Cultural de Walter Burley (The Philosophy of the Fourteenth Century. The Cultural Context of Walter Burley)*, Madrid: Biblioteca 'La Cuidad de Dios' Real Monasterio de el Escorial.(A survey of fourteenth-century philosophy as relevant to the work of Burley.)

Weisheipl, J.A. (1969) 'Repertorium Mertonense' (Mertonian Inventory), *Mediaeval Studies* 31: 174-224. (Includes a list of Burley's known works together with manuscript locations.)

Wood, R. (1988a) 'Walter Burley's *Physics* Commentaries', *Franciscan Studies* 43: 275-327.(Distinguishes several editions, and shows that the commentaries ascribed to Burley in Erfurt manuscripts are not in fact authentic.)

Wood, R. (1988b) 'Studies on Walter Burley 1968-1988', *Bulletin de philosophie médiévale* 30: 233-50.(Lists facsimile editions, modern editions and articles on Burley.)

Wood, R. (1989-90) 'Walter Burley on Motion in A Vacuum', *Traditio* 45: 191-217.(As Wood explains, Burley suggest there could be a quantitative extension without a substance present (as occurs in the Eucharist), and if so, motion might occur in it.)

Burthogge, Richard (*c.*1638-*c.*1704)

Richard Burthogge, perhaps the first but certainly not the least interesting modern idealist, was a minor philosopher who responded to a variety of English and Dutch influences. His epistemology, constructed as an undogmatic framework within which to debate theological and metaphysical issues, contains remarkable resemblances to later, even recent idealism. He argued that, since our faculties help to shape their objects, we never know things as they are in themselves: all the immediate objects of thought are appearances. 'Metaphysical truth' is therefore beyond us, but we approach 'logical truth' in so far as our notions harmonize or cohere with one another and with experience. In this spirit, Burthogge advocated a tolerant reasonableness in religion, while in metaphysics he postulated a universal mind, united with matter, of which individual minds are local manifestations.

1 Life

Richard Burthogge was born in Plymouth, England, and educated at Exeter Grammar School, at Oxford (B.A. 1658) and at Leiden in Holland (M.D. 1662). He returned to his native Devon and became a prosperous medical practitioner. His first theological works, *Divine goodness* (1672) and *Causa Dei* (1675), wrestled with the problems of evil, grace and divine justice. *Organum Vetus et Novum* (1678) was Burthogge's epistemological response to the 'Proud *Ignorance*, Ignorant *Zeal*, and Impertinent Reasoning' of his critics. It rested on an anti-dogmatic probabilism characteristic of such advocates of toleration as William Chillingworth and the Cambridge Platonists (see Cambridge Platonism), together with a thesis which Burthogge could have heard expounded in Leiden by Geulincx, that the objects of logic and knowledge are shaped by the forms of our understanding. Between 1684 and 1691 further works called for 'a general toleration', and attempted to settle religious disputes by 'harmonious' interpretation of scripture and historical evidence. In 1694 *An Essay upon Reason, and the Nature of Spirits* set out Burthogge's theory of knowledge as prolegomenon to a theory of mind which, although dualistic, owed something to Spinoza. This work was also influenced by Locke, to whom it was dedicated and with whom a correspondence ensued. *Of the Soul of the World, and of Particular Souls* (1699), again addressed to Locke, gave further arguments for Burthogge's hypothesis. The Calvinist doctrine of predestined salvation was attacked once more in *Christianity a Revealed Mystery* (1702), an unorthodox exegesis of *Romans* VIII.28-30. In 1703 this argument was pursued in letters to Locke, who then offered the same interpretation in his *Paraphrase* of St Paul. It seems not to be known when Burthogge died.

2 Theory of knowledge

Burthogge's fundamental epistemological principle, expounded in both *Organum* and *An Essay*, is that things are known by us only as they are in our faculties (that is sense, imagination and understanding), and that 'Every Faculty hath a hand, though not the sole hand, in making its immediate Object'. Philosophers recognize that 'Colours are [not] without the Eye… although they seem so to Sense' (1678: 13-14). Yet the understanding equally 'doth *Pinn* its Notions upon Objects' (1694: 58): substance, accidents, powers, similitude, whole, part, cause, effect and so on 'own no other kind of Existence than… an Objective [that is, intentional] one'. Yet it does not follow that nothing is real, since intentional or 'cogitable' beings are characteristically grounded in realities: 'Notions of the Minde are *bottomed* on *Sentiments* of Sense; so that as Realities are Grounds to Sentiments, so Sentiments are Grounds to Notions' (1678: 15-16). The hinge here is our awareness in sensation that external realities are acting on us - just that awareness which masks the contribution of the faculty itself. Voluntary fictions obviously lack such grounding, while, although sensory illusions and dreams are both 'real' in having causes outside the mind, in the former case the causes are unusual, and in the latter they are 'Causes only, and not Objects as well as Causes' (1694: 78).

Understanding, it is explained in *An Essay*, arises from sense by means of language. Words are images which aid thought either by calling other images or notions to mind or by standing in place of them. Indeed, ideas of the understanding are nothing but definitions, both 'of words' and 'made by words'. This allows for notions of things 'perfectly intelligible and mental', but such notions are necessarily shaped by language and even further from reality than the images or sensations on which they are based (1694: 27-34).

Burthogge recurs to the traditional point that to apprehend something through an attribute is to apprehend it

imperfectly under a notion or aspect or *modus concipiendi*, and yet we can only apprehend things through their attributes. In *An Essay* this thought is developed on Lockean lines: our experience is only of qualities, but we conceive 'that other things are under them, which do uphold and support them, and consequently, that are *subjects*, or substances' (1694: 98) (see Locke, J. §§4-5). In *Organum*, Burthogge had emphasized a comparison with our imperfect knowledge of God. As to know God through his attributes is to know him only by analogy, so in general 'to us men *things* are nothing but as they stand in our *Analogie*' (1678: 12). There is, however, a difference between the Light of Faith and the Light of Reason. Revelation supplies analogies for spiritual, rather than natural things. They should be employed as such, and reason should endeavour to harmonize them, but none should be insisted on 'as if it were adequate, or just'. Still less should they be used in opposition to natural reason (1678: 26-34).

Organum also contains what is perhaps the clearest and most forthright statement of a coherence theory of truth in the seventeenth century (see Truth, coherence theory of). Because we cannot get behind our notions, 'metaphysical truth' is of no concern to us. 'Logical truth', however, when 'rightly shewn', is naturally distinguished by the understanding. The criterion of logical truth cannot be the Stoic or Cartesian clear and distinct perception, since any criterion must lie in the object, not in the way in which we perceive it. Nor does it lie in innate tendencies or principles. Truth, as the ground of assent, is the objective harmony, congruity and coherence of things 'in the Frame and Scheme of them in our Mindes'. Natural reasoning assumes a harmonious world as the object of a unified science: in 'science as it is in Arch-work, the Parts uphold one another'. To prove something is to fit it into 'a Scheme and Frame of Notions bottomed on things'. Any hypothesis is in principle vulnerable to a more coherent hypothesis, but to doubt without reason is 'unreasonable and contradictious' (1678: 47-60).

3 Theory of mind

In contrast to Locke, Burthogge argues in *An Essay* that, despite our ignorance of their real natures, we cannot suppose that mind is a power of matter; nor can we, with Spinoza, suppose them identical (see Locke, J. §5; Spinoza, B. de §§5-6). Since the attributes through which we distinguish them, extension and thought (and, with thought, agency), are mutually irreducible by us, we simply have to abide by our 'Refracted, Inadequate, *Real-Notional* way of conceiving' (1694: 107). Pure matter appears as passive; pure mind, which is God, as active. Mind as we know it in ourselves and in animals, however, appears inextricably 'concreted' with matter. Experience shows that reason itself requires a properly working organ. The best hypothesis is that everything in nature is a combination of matter with a universal spirit which activates and controls it. The species of living, sensitive and rational things are produced according to the 'Texture and Quality' of their material structure. The organs of perception and thought 'catch and retain' specific forms of cogitation as a mirror catches an image from motions in the medium, or a glass catches the vibrations of a lute-string. Without matter so in harmony with mind, there would be no individuation of thoughts or, indeed, of souls, which are 'portions' of 'the Soul of the World'. This panpsychist hypothesis allows that human souls are generated in the same general way as animal souls, a piece of common sense doubted only by the Cartesians.

MICHAEL AYERS

List of works

Burthogge, R. (1921) *The Philosophical Writings of Richard Burthogge*, ed. M.W. Landes, Chicago, IL, and London: Open Court.(Contains the *Organum Vetus et Novum*, *Of the Soul of the World*, with selections from *An Essay upon Reason*. Introduction presents Burthogge as precursor of Kant.)

Burthogge, R. (1672) *ΤΑΓΑΘΟΝ, or, Divine Goodness Explicated and Vindicated from the Exceptions of the Atheist*, London.(Attempts to reconcile the existence of evil, and the doctrines of grace and of salvation by faith, with divine goodness and justice.)

Burthogge, R. (1675) *Causa Dei, or an Apology for God*, London.(As above, allowing means of salvation to heathens, since God cannot require impossibilities.)

Burthogge, R. (1678) *Organum Vetus et Novum: or, A Discourse of Reason and Truth*, London; repr. in *The Philosophical Writings of Richard Burthogge*, ed. M.W. Landes, Chicago, IL, and London: Open Court, 1921. (Bases an anti-dogmatic epistemology and coherence theory of truth on the principle that the objects of knowledge are shaped by our faculties and notions.)

Burthogge, R. (1683) *An argument for infants baptisme, deduced from the analogy of faith, and harmony of the scriptures*, London; 2nd edn, 1684.(Typical of Burthogge's theological arguments, this one finds analogical

and biblical grounds in favour of infant baptism.)

Burthogge, R. (1687) *Prudential reasons for repealing the penal laws against all recusants, and for a general toleration*, London.(An appeal for religious toleration, published anonymously.)

Burthogge, R. (1690) *The Nature of Church-government, freely discussed and set out*. (Anonymously published argument that the form of church government is a practical matter, not determined by scriptural authority.)

Burthogge, R. (1694) *An Essay upon Reason, and the Nature of Spirits*, London; repr. in *British philosophers and theologians of the 17th and 18th centuries*, 10, New York and London: Garland, 1976.(Restates the theory of *Organum* as a methodological framework for a speculative, panpsychic theory of the relation between mind and body.)

Burthogge, R. (1699) *Of the Soul of the World; and of Particular Souls: in a Letter to Mr. Lock*, London; repr. in *The Philosophical Writings of Richard Burthogge*, ed. M.W. Landes, Chicago, IL, and London: Open Court, 1921.(Further argument, from the coherence of scripture, various authorities and purported facts, for panpsychism.)

Burthogge, R. (1702) *Christianity a Revealed Mystery: or, the gracious purpose of God toward the gentiles*, London.(An attack on the Calvinist doctrine of the elect which apparently influenced Locke.)

Burthogge, R. (1979-82) *The Correspondence of John Locke*, vols 5, 6 and 7, ed. E. S. de Beer, Oxford: Oxford University Press.(Contains letters from Burthogge, some on predestination.)

References and further reading

Gruenbaum, J. (1939) *Die Philosophie Richard Burthogges*, Bern: J. Kleiner.(This doctoral dissertation is the only modern monograph.)

Lennon, T.M. (1993) *The Battle of the Gods and Giants*, Princeton, NJ: Princeton University Press.(Chapter 3, §11, discusses Burthogge's 'tendency' to idealism in the context of debates about 'enthusiasm', toleration and analogy.)

Nuchelmans, G. (1983) *Judgement and Proposition from Descartes to Kant*, Amsterdam: North Holland.(Chapter 6 contains a discussion of Burthogge's relation to Geulincx.)

Yolton, J.W. (1968) *John Locke and the Way of Ideas*, Oxford: Oxford University Press.(Contains discussions of Burthogge's relation to Locke.)

Bushi philosophy

Bushi *is one of several terms for the warrior of premodern Japan;* samurai *is another. The 'way of the warrior' - that is, the beliefs, attitudes and patterns of behaviour of the premodern Japanese warrior - is commonly called* bushidō *(literally, the 'way of the* bushi*'). However,* bushidō *is actually a phrase of rather late derivation, and in premodern times was never exclusively used to describe the warrior way.*

Two of the earliest and most enduring phrases for the way of the warriors who rose in the provinces of Japan in the late ninth and tenth centuries were the 'way of the bow and arrow' and the 'way of the bow and horse'. These phrases, however, referred to little more than prowess in the military arts, the most important of which, as the second phrase clearly specifies, were horse riding and archery. For many centuries no one in Japan undertook to define systematically what the way of the warrior in a larger sense was or should be. Warrior beliefs, ideals and aspirations - including loyalty, courage, the yearning for battlefield fame, fear of shame and an acute sense of honour and 'face' - were widely recognized, but neither warriors nor others apparently felt the need to codify them in writing.

*Not until the establishment of the Tokugawa military government (shogunate) in 1600, which brought two and a half centuries of nearly uninterrupted peace to Japan, did philosophers begin to study and write about the warrior way (*bushidō*). Concerned about the meaning and proper role of a ruling warrior class during an age of peace, philosophers posited that warriors should not only maintain military preparedness to deal with fighting that might occur, but should also develop themselves, through education based primarily on Confucianism, to serve as models and moral exemplars for all classes of Japanese society.*

1 Early beliefs

A warrior class emerged in the provinces of Japan in the late ninth and tenth centuries because the court government in Kyoto, headed by an emperor but administered mainly by regents of the Fujiwara family, left local governance largely to local families. In the absence of firm control from the court, men of locally powerful families, holding offices in provincial and district governments, took up arms to maintain the peace and deal with disputes over land and other issues.

However, with few exceptions, these warriors did not seek to assert their independence from the Kyoto court. On the contrary, they often turned to the court for support in conflicts with other warriors or officials. They also greatly respected and admired court society and culture. In ensuing centuries, warriors repeatedly sought to obtain court ranks and offices, even though these might be purely ornamental, and to acquire knowledge of court culture. A distinctive feature of the warrior way as it evolved over the ages was the steady acquisition by warriors of courtly ways. As the records put it, warriors combined the *bun* (the courtly and cultural) with the *bu* (the military).

As warrior bands formed and contended for local and regional power in the middle and late Heian period (794-1185), they evolved feudal institutions very similar to the feudalism of western Europe in its medieval age. Chief among these institutions was the lord-vassal relationship. Chronicles known as war tales, records of the lives and battles of warriors, depict the intimate personal attachments that bound lords and vassals, telling of vassals prepared to give their lives for their lords at a moment's notice and of lords who bestowed upon vassals the kind of love normally given by fathers to their sons. Of course, not all warriors behaved like the idealized characters of the war tales; but the tales presented models for ages to come of how warriors ought to conduct themselves, especially in their roles as lords and as vassals.

As noted above, warriors fought primarily on horseback with the bow as their principal weapon. In battles between armies, warriors usually paired off to contend one-against-one, using their swords when they entered into close combat. Since it was difficult to kill an enemy with a sword from horseback, the warrior tried to unseat his foe, then leap down and stab him to death. Although in later centuries the sword came to be looked upon as the soul of the warrior, in this early age the warrior's great pride lay in his skill with the bow, wielded while riding a horse.

From earliest times the warrior displayed great pride in *na* (name), a term that connoted not only a sense of pride in family but also a determination to uphold personal honour and to achieve fame on the battlefield. Concern about *na*, however, had the potential to conflict with the warrior's obligation of loyal service to his lord. The warrior

anticipated reward for victory in battle; and indeed usually received from his lord a portion of the lands and other wealth seized from defeated enemies. However, if the warrior pursued his own interests too aggressively in a battle - not only in the quest for material reward but also to achieve personal fame and glory - he could be difficult to command and could even undermine his lord's battle strategy.

The converse of the warrior's pride in name and thirst for fame was his fear of shame. The warrior was of course shamed if he was perceived to be timid or cowardly, but shame could also be experienced simply by defeat in battle, no matter how well a warrior had fought. The war tales often speak of warriors driven by the single-minded desire to expiate shame, suggesting that shame was a major factor in the frequent undertakings of revenge that we find in warrior history. In later centuries, the pursuit of revenge became institutionalized in the vendetta.

2 The warrior way in the age of the country at war

One of the most important periods in the shaping of the warrior way was the *sengoku jidai*, the 'age of the country at war' (1478-1573). The Ōnin War, fought largely in Kyoto during the decade 1467-77, destroyed Kyoto and plunged Japan into a state of disunion and conflict that lasted nearly a century. Beginning about 1500, territorial domains were established by warrior chieftains whom historians call *daimyō*, and through much of the early and middle sixteenth century these *daimyō* fought fiercely among themselves. Finally, one of the *daimyō*, Oda Nobunaga, emerged as a unifier. Entering Kyoto in 1568, Nobunaga began a process of national unification that was completed in 1590 by his vassal and successor, Toyotomi Hideyoshi.

So intense and widespread was the fighting in the *sengoku* age that there were real fears it would tear apart the social fabric of Japan. Warrior values were sorely tested; indeed, some were openly abandoned. If the lord-vassal relationship, based on the vassal's loyalty to his lord and the lord's loving care of his vassal, was the essence of Japanese feudalism, then feudalism fared poorly in the *sengoku* age. This was a time of frequent, outright treachery, which witnessed *daimyō* violating treaties with impunity and vassals betraying their lords or simply abandoning them in favour of those who offered richer rewards.

It is instructive to note that, in the legal codes they compiled for their domains, the *daimyō* of the *sengoku* age said little about loyalty but a great deal about disloyalty. Hereditary (*fudai*) vassals - that is, vassals whose families had served a *daimyō* house for generations - were expected to be loyal. Non-hereditary (*tozama*) vassals, on the other hand, were not only not held to high standards of commitment and loyalty, they were scarcely criticized if they shifted from one *daimyō* to another, seeking a better (more powerful and more generous) lord.

The *daimyō* especially feared fighting between vassals, which could cause serious ruptures in both their armies and their governments. To prevent such fighting, they included in their domainal codes what can be called 'no questions asked' laws. These laws specified that, if two vassals began to fight, both would be punished without any inquiry into the cause of the clash. The more powerful *daimyō* decreed that the punishment for fighting between vassals was death. The vassal of one prominent *daimyō*, however, charged that the 'no questions asked' law went against the warrior's 'manly way' (*otoko no michi*), since it encouraged the warrior to restrain himself even if insulted or otherwise provoked. The vassal said that warriors worthy of being called warriors had to be aggressive, ready to fight whenever challenged, and predicted that the 'no questions asked' law would nurture meek, if not cowardly, vassals.

A major event during the *sengoku* age was the appearance of Europeans in Japan for the first time. The first Europeans to set foot on Japanese soil were Portuguese traders who arrived in 1543 in a Chinese junk. Soliciting trade, the Portuguese introduced the Japanese to Christianity and guns (muzzle-loading muskets); during the ensuing century, trade, Christianity and guns became the main staples of intercourse with Europeans. Japanese warriors had been exposed to the use of gunpowder during the two unsuccessful invasions of Japan by Mongol troops in 1274 and 1281. However, this exposure had not led to any significant experimentation with gunpowder by the Japanese themselves during the next three centuries. Hence, we may suppose that the Japanese were utterly astonished when they first observed the firing of guns by the Portuguese. Immediately recognizing the extraordinary value of guns for warfare, the *daimyō* of the *sengoku* age eagerly sought to acquire them. However, guns were not readily obtainable; the Portuguese could only import so many, and establishing gun foundries in Japan was expensive. Some important battles were won with guns in the late sixteenth century, but by the time guns became widely available around 1600, the fighting had ceased. Thus guns did not revolutionize warfare in

Japan during this period. Had they done so, they would surely also have undermined if not destroyed the special status of the warrior or *samurai* class. As it was, the 'way of the warrior' remained the way of elite fighters specially trained in the complex military arts of horsemanship, archery and use of the sword.

3 The Tokugawa period and the concept of *bushidō*

With the establishment of the Tokugawa shogunate in 1600, Japan's ruling warrior class lost its principal *raison d'être*, which was fighting. No one could have known at the time that peace would prevail, with few exceptions, for more than two and one-half centuries. Hence, throughout the seventeenth century at least, the rulers of the shogunate were absorbed first and foremost with security, taking measures to prevent a reversion to the bloodshed and chaos of the *sengoku* age.

One of the most important steps that had been taken to end the warfare, unify the country and make unification lasting was the separation of peasants and warriors. In the late sixteenth century, those warriors who had not already done so were obliged to move from the countryside into the castle towns of their *daimyō*; simultaneously the peasants, who were ordered to remain in their villages, were disarmed in a national 'sword hunt' that deprived them of their swords and other weapons.

Removed from the land and transformed into urban residents, warriors became stipendiaries, the recipients of regular stipends based on the revenue derived from agricultural fiefs held by them or their ancestors during the *sengoku* age. High-ranking warriors received, for the most part, generous stipends and were employed by their *daimyō* in the domainal governments. Those on the lower end of the warrior scale were usually not so employed and were obliged to get along on meagre stipends.

The existence of a large number of idle stipendiaries among the warriors troubled philosophers. Yamaga Sokō(1622-85), a Confucian scholar whose name is closely associated with the formulation of *bushidō* during the Tokugawa period (although he himself used the abbreviated term *shidō*), observed: '… the *samurai* eats food without growing it, uses utensils without manufacturing them, and profits without buying or selling. What is the justification for this?' (Tsunoda *et al.* 1958: 398). Sokō's answer was that the function of the *samurai* in society was to serve his lord loyally and to stand as an exemplar of high moral standards to the lower classes, peasants, artisans and merchants. The acquisition of high moral standards - that is, Confucian moral standards - was premised on the acquisition of education (see Confucian philosophy, Japanese).

Strongly influenced by the thought of Yamaga Sokō, *bushidō* came to comprise two parts. The first part was military preparedness and a romantic looking back to the great fighting traditions of the past. Countless editions of the war tales and other records of ancient and medieval battles, biographies of the great warrior heroes of the past and the like were published during the Tokugawa period. In addition, many schools in military tactics and the martial arts were founded, including one by Yamaga Sokō.

The second part of *bushidō* concerned the ethical training urged by Sokō. This not only resulted in the articulation of an ethical code for warriors of a kind that we do not find prior to the Tokugawa period, but also greatly stimulated education in general among warriors. Hence, *bushidō* contributed importantly to the evolution of the warrior class as an intellectual as well as ruling elite.

If the *bushidō* just described was based primarily on Confucian rationalism, there was another *bushidō* nurtured by a different, powerfully 'irrational' school of thought. This school derived from the belief, adumbrated above in the discussion of the feudal lord-vassal relationship, that vassals ought to serve their lords in a spirit of absolute self-sacrifice: vassals should, indeed, be prepared to relinquish their lives in a moment's notice for their lords. The war tales, for example, speak of vassals who considered their lives as no more important than a 'speck of dust' or a 'goose feather' in the service of their lords.

The bible of this irrational school of *bushidō* was the *Hagakure*, an early eighteenth-century compilation of the views of Yamamoto Tsunetomo, vassal to the *daimyō* of a western domain. In one of the most famous proclamations in all of Japanese literature, Tsunetomo asserts in the *Hagakure* that: 'The way of the warrior (*bushidō*) is found in death' (trans. Wilson 1979: 17). But, as Tsunetomo explains, it is not sufficient for the warrior to be resigned to the fact that he may be killed. He must be prepared at all times to face death at a moment's notice, and when confronted with a situation in which he must decide between living or dying, he should

unhesitatingly choose to die. In such a situation there can be no calculating or 'rationalizing'. The warrior must fight like a madman to his death, even though aware that others may later judge his actions to have been wrong or foolish. Utterly unconcerned about 'achieving' anything, the true warrior devotes himself single-mindedly to acting in a death-defying - or more precisely, a death-attaining - way.

The *Hagakure*, with its extreme views about proper warrior behaviour, was written after the passage of a century of Tokugawa peace and during a time (the Genroku cultural epoch) that celebrated the lives of merchants and artisans, whose chief goals in life were to earn money and enjoy the luxuries it could buy. Genroku culture was totally antithetical to traditional warrior values, at least as those values were idealized in writings like the war tales. A *cri de coeur* against what Yamamoto Tsunetomo viewed as the moral deterioration of the times, the *Hagakure* was an anachronistic attempt to rally the spirit of the warrior class, most of whose members could only have been bewildered at the prospect of rushing to their deaths in an age when warfare had been suspended for so long that no one had any experience of it.

But in fact, an incident had occurred just a few years before the publication of the *Hagakure* that seemed to suggest that the traditional warrior values centred on the selfless service of vassals to their lords, which Yamamoto Tsunetomo so admired and so yearned to revive, were still alive. One of the most famous events in all of Japanese history, this incident was the sensational revenge of the forty-seven *rōnin* (masterless *samurai*). In 1701, a *daimyō* in attendance at the shogun's castle in Edo (modern Tokyo) assaulted and wounded one of the shogun's officials. For having drawn his sword in the castle the *daimyō* was condemned to death, and that same day he disembowelled himself - that is, he committed *seppuku*, the warrior's way of suicide. Two years later, a group of the *daimyō*'s vassals, who had become *rōnin* because of their lord's death, fulfilled a secret pact by attacking the residence of and killing the official the *daimyō* had assaulted (the exact reason for the assault is not known). Seven weeks later the *rōnin,* condemned to death for violating shogunate law, themselves committed *seppuku*.

Although this act of revenge - this carrying out of a vendetta (*katakiuchi*) - has stirred the spirits of many Japanese over the years, it was in fact an isolated incident and not in any sense illustrative of the way warriors behaved or were expected to behave. Yamamoto Tsunetomo in fact criticized the *rōnin* because they did not spring into action immediately against the enemy of their lord but instead calculated and schemed for two years. More fundamentally, this vendetta was unusual because nearly all the other vendettas by warriors during the Tokugawa period were based on avenging kin rather than feudal lords.

Nevertheless, the revenge of the forty-seven *rōnin* was important in the history of *bushidō* because it drew attention to the special character of the lord-vassal relationship among warriors, reaffirming the ideal of vassals acting selflessly, even to the point of forfeiting their lives, in the service of their lord. In the modern age, after the dissolution of the *samurai* class and the establishment of a conscript army, this spirit was still held to set Japanese soldiers apart from those of other countries. Even as late as the Second World War the spirit of the Japanese soldier, prepared to die if necessary for his emperor, was exalted and accorded special consideration in the strategic planning of many Japanese commanders.

See also: Confucian philosophy, Japanese; Honour; Japanese philosophy; Political philosophy, history of; War and peace, philosophy of

PAUL VARLEY

References and further reading

Mills, D.E. (1976) '*Kataki-uchi*: The Practice of Blood-Revenge in Pre-Modern Japan', *Modern Asian Studies* 10 (4): 525-42.(Article on the practice of vendetta.)

Tsunoda Ryusaku *et al.* (eds) (1958) *Sources of Japanese Tradition*, New York: Columbia University Press. (Includes the work by Yamaga quoted in this entry.)

Varley, P. (1994) *Warriors of Japan, As Portrayed in the War Tales*, Honolulu, HI: University of Hawaii Press. (Description of the *samurai* class.)

Yamamoto Tsunetomo (1716) *Hagakure (In the Shadow of the Leaves)*, trans. W.S. Wilson, Tokyo: Kodansha, 1979.(One of the most prominent works of *bushi* philosophy.)

Business ethics

Business ethics is the application of theories of right and wrong to activity within and between commercial enterprises, and between commercial enterprises and their broader environment. It is a wide range of activity, and no brief list can be made of the issues it raises. The safety of working practices; the fairness of recruitment; the transparency of financial accounting; the promptness of payments to suppliers; the degree of permissible aggression between competitors: all come within the range of the subject. So do relations between businesses and consumers, local communities, national governments and ecosystems. Many, but not all, of these issues can be understood to bear on distinct, recognized groups with their own stakes in a business: employees, shareholders, consumers, and so on. The literature of business ethics tends to concentrate on 'stakeholders' - anyone who occupies a role within the business or who belongs to a recognized group outside the business that is affected by its activity - but not in every sort of business. Corporations are often discussed to the exclusion of medium-sized and small enterprises.

Theories of right and wrong in business ethics come from a number of sources. Academic moral philosophy has contributed utilitarianism, Kantianism and Aristotelianism, as well as egoism and social contract theory. There are also theories that originate in organized religion, in the manifestos of political activists, in the thoughts of certain tycoons with an interest in social engineering, and in the writings of management 'gurus'. Recently, business ethics has been affected by the ending of the Cold War, and the breakdown of what were once command economies. These developments have encouraged enthusiasts for the market economy to advocate moral and political ideas consistent with capitalism, and the handing over to private companies of activity in certain countries that has long been reserved for the state.

1 Stakeholders and the range of business ethics

The concept of the stakeholder - someone who occupies a role within the business or who belongs to a recognized group outside the business that is affected by its activity - lies at the centre of business ethics. The literature tends to concentrate on stakeholders in bigger companies on the American model: companies quoted on stock exchanges, with impressive turnovers, large numbers of employees, and several layers of management. Such companies are involved with a wide range of stakeholders, both internal and external to the organization. The 'internal' stakeholders include directors, managers, ordinary shop-floor or 'blue-collar' employees, and shareholders. Outside the company there are customers, suppliers, competitors, local communities, and one or more levels of government. Business ethics as a whole may be regarded as a unified understanding of the obligations of those who run a business - directors or top managers - to the rest of a business' stakeholders. Alternatively it may be seen as a unified understanding of the typical ethical risks or problems facing the internal stakeholders in their dealings with one another and outsiders. Both conceptions are well represented in the literature, and both accommodate the central issues of business ethics.

2 Internal stakeholders: the case of employees

Perhaps the leading 'internal' stakeholders of a firm are its employees, whose fortunes depend on those of the firm in the most direct way. The firm may be the only source of income for each of its employees, and the income may be used to support many others. Employees may depend on their job for a considerable proportion of their social life and their most rewarding activity. Having a particular job, or having a job at all, may be an important source of self-esteem (see Work, philosophy of). With so many connections to the good things in life, the role of employee cannot fail to matter morally. Equal opportunity and employment protection legislation reflect the importance of having and keeping a job. In many industrialized countries discrimination in recruitment on the grounds of race or sex or religion is against the law. In western Europe dismissal from a full-time job can be legally challenged, and redundancy can be accompanied by significant financial compensation. Different types of employee have different sorts of stake in different businesses, and businesses have different stakes in different employees. A worker on the factory floor and the manager above him are both employees, but their roles in the firm are likely to make different moral demands on them and their employers. It may fall to the manager to sack factory workers for some offence or to make them redundant during a recession. It may fall to a worker to make known a costly production defect or to complain of harassment by someone higher up in the organization. The manager may strongly disagree with the

decision to sack the workers, and yet be responsible for carrying it out. The worker may wonder whether making known the defect will indirectly cost them their job: putting right the defect may be expensive, and redundancies may be the firm's way of covering its losses. On the other hand, if the defect is not made known, perhaps the safety of a very large number of consumers will be threatened. Serious consequences may also attend the report of harassment. The complainant risks not being believed, and probably the embarrassing details will be made the subject of gossip. Perhaps the person complained about will retaliate in some way. And the firm may be forced to decide between two employees, neither of whose talents it can afford to lose.

Not every moral theory is suited to weighing what is at stake in such cases, and several different theories may be required to do justice to all stakeholders in business ethics. Utilitarianism and Aristotelianism have resources for capturing the value of employment in general, since each works with a concept of wellbeing or welfare which embraces many different goods, including those associated with having a job (see Utilitarianism; Aristotle §21). Kantianism is able to give weight to the way in which consumers are put at risk by not publicizing the production defect, and the way in which the wellbeing of employees can be sacrificed for a short-term reduction in the costs of a business (see Kantian ethics). Kantianism is also able to make sense of the wrong of harassment, and therefore to justify a decision to press a complaint against harassment. In all of these cases, albeit in different ways, persons are treated as means, and Kantianism highlights such treatment as a kind of immorality. Aristotelianism is sometimes thought to be able to coordinate the moral demands that fall on people in their role as manager or worker with the moral demands that fall upon them in the rest of their lives. The manager who is uncomfortable with sacking people may be someone who tries to pursue a policy of benevolence or charity whenever possible. The worker who balks at overlooking the production defect in the factory may be in the habit of being honest and open away from work.

Aristotelianism prescribes the cultivation of patterns of behaviour that serve one in one's private as well as in one's public life, and that cannot be abandoned without a morally significant loss to the agent. Utilitarianism and Kantianism, on the other hand, sometimes diagnose the reluctance to depart from characteristic patterns of behaviour as a kind of squeamishness or self-indulgence. Sacking the workers may be unpleasant, according to these theories, but it may be best on balance for everyone, or necessary if one is to honour promises to shareholders or creditors.

3 External stakeholders: social responsibility

It is not only Kantianism, utilitarianism and Aristotelianism which are invoked to deal with the obligations of a firm to the wider world. Sometimes the firm is said to be party to a social contract which confers a duty to transfer profits or services to the communities in which it operates, in return for the good will or custom or labour of members of those communities (see Contractarianism). Sometimes it is said to be in the firm's self-interest to benefit communities. These things are said, at any rate, by those who think that businesses have obligations to external stakeholders at all. There are well-known advocates of the view that company philanthropy within a host community violates a prior and overriding obligation of a firm to its shareholders. The more profit or employee time is diverted to noncommercial and non-profit-making activity, so it is argued, the less there is to redistribute to shareholders, who after all own the company and invest in it usually on the understanding that the company will maximize returns to them.

The view that firms have no business helping the community is now widely contradicted by the behaviour of large commercial enterprises themselves. The directors of such enterprises accept that philanthropy is often expected by local communities and shareholders alike, and sometimes endorsed by consumers in the form of greater approval for, and loyalty to, firms that go in for it. In the UK retail sector, for example, market research has confirmed the commercial benefits of a caring public image.

If it is in the interest of firms to give money or time to helping the community, and if they give the help because of the commercial benefits, is the help morally creditable? Much depends on whether the help is given only for the commercial benefits. If it is only for the commercial benefits, then probably only egoism - the theory that something is right if and only if it serves one's own interest - makes the act moral, and egoism is strongly counterintuitive as a moral theory (see Egoism and altruism). On the other hand, although business people see the value of self-interest, it is uncharacteristic for them to say or think that acts of corporate charity serve only self-interest. Perhaps they would not go in for donations of money or time if they were not going to pay the firm in

some way; this does not mean that nothing else matters. The motivation can be mixed, as it is when directors of enterprises themselves have strong philanthropic impulses, but can also see that the philanthropy will be good for public relations. In the case where the motivation is mixed and the benefits to the community are real, then several theories assign moral credit to corporate giving, including utilitarianism, different versions of theories that say that businesses are parties to a social contract, and Aristotelianism. Even a Kantian theory can assign considerable positive value to company philanthropy with a mixed motivation, so long as it is the right sort of mix. The action that is in accordance with duty but motivated by extra-moral forces can still be honourable, beautiful, and so on according to Kant.

What are the appropriate forms of corporate philanthropy or ‘social responsibility’? The question has an edge in places where the private sector is increasingly expected by government to pay directly for, or sponsor programmes in, schools and hospitals, and when businesses are partners with the state in big transportation and housing projects. Does a firm go beyond what can reasonably be demanded of it when it takes over responsibility in one place for a service or institution that the state is responsible for elsewhere? Should it help only with public services or institutions, refraining from taking a leading role? And if it gives help, should the help have some relation to its business activity? Should a construction firm donate construction advice or labour for a public housing scheme; or should it lend the scheme one of its accountants? Should a supermarket take charge of recycling packaging on the goods it sells rather than lending its property-buying expertise to the public sector? Here a version of social contract theory, according to which business has an obligation to fulfil the expectations it creates in society, provides a clue to the answer. The closer a firm’s philanthropy to its public business activity, the less the risk of creating new, perhaps over-demanding, expectations among the public and confusion among internal stakeholders.

Many questions about the obligations of businesses to the wider world are less tractable than those we have been pursuing. For example, what are the obligations of transnational firms in societies with very different levels of prosperity? In what special circumstances are the assets of firms open to expropriation in the name of the public interest? To what extent is the law rather than corporate policy the best medium for stipulating the limits of corporate obligation?

4 Conflicts between obligations to stakeholders

Fulfilling obligations to one set of stakeholders can involve a firm in conflicts with another set. A policy of reducing costs may be pursued to maximize returns to shareholders, and yet it may involve the dismissal of productive and loyal employees. Recalling faulty goods for the sake of consumer safety may similarly conflict with the duty to maximize returns. Keeping suppliers waiting for their money may be justified by a duty to manage the liquid assets of a firm as profitably as possible, but it may drive the supplier into bankruptcy; and so on. The existence of these conflicts does not show, as might be thought, that it is unprofitable to gear obligations in business ethics to stakeholders. But it does raise the question of whether there is a pecking order among stakeholders, so that obligations to one group typically outweigh obligations to another. The literature sometimes suggests that one or another group of internal stakeholders is pre-eminent: sometimes it is employees and sometimes it is shareholders. Neither group is uncontroversially the pre-eminent one, and much depends on facts about the firm with the conflicting obligations. It may be that for some firms at some times meeting obligations to employees makes the difference between staying in business and bankruptcy; for other firms it may be shareholders. Either way, considerations about the requirements of staying in business at a given time seem relevant to the pecking order; and when the survival of the business is not in question, it may be considerations about which group is the more important to the growth of the business.

5 Further questions

Not all questions in business ethics can be resolved into questions of what a business owes to its stakeholders. Questions can be asked about the morality of the market economy or of pursuing profit in general (see Market, ethics of the). Capitalism as an economic system, or democratic capitalism as a politico-economic system, can be subjected to moral appraisal. Occasionally the large question of the morality of capitalism is pursued in the literature without an enthusiastic audience among business people. Yet the relevance of the question has increased with the collapse of the command economy and the communist state. A market economy may be a condition for membership in the community of nations for countries once within the orbit of the Soviet Union, just as it has become the unquestioned condition for aid to the developing countries of Africa and Asia. On the other hand, it

may be that a bias in some places towards a ‘social market’ and in others towards a purer form of the market will simulate the old rivalry between capitalism and socialism.

The collapse of the Soviet bloc has put one sort of gloss on the question of the desirability of capitalism; the Green movement is responsible for another (see Green political philosophy; Environmental ethics). An economic system predicated on the possibility of indefinite growth, and the satisfaction, no questions asked, of a very wide variety of consumer demands, seems to many to be immoral in its profligacy and its relentless draining of natural resources. The Green complaint against capitalism goes far deeper than a protest against pollution and the loss of wilderness. It is a protest against the picture of technological and economic progress that both market and command economies have in common. More generally, it is a protest against the idea that nature is an economic resource, and that the human ability to produce by transforming nature is a good thing. Perhaps more fundamentally still, it is a protest against the identification of the good with the satisfaction of human desire. Business ethics has yet to confront these protests.

See also: Applied ethics; Development ethics; Economics and ethics; Professional ethics

TOM SORELL

References and further reading

Beauchamp,T. and Bowie, N. (eds) (1988) *Ethical Theory and Business*, Englewood Cliffs, NJ: Prentice Hall; 5th edn, 1997.(Collection of essays and cases prefaced by a discussion of the bearing on business ethics of utilitarianism and deontological theories.)

Hoffman, W.M. (1989) ‘The Cost of a Corporate Conscience’, *Business and Society Review* Spring 1989. (Acknowledges that social responsibility can be a burden to business.)

Schmidheiny, S. and Business Council for Sustainable Development (1992) *Changing Course: A Global Business Perspective on Business and the Environment*, Boston, MA: MIT Press.(Articles on ‘Green’ business initiatives throughout the world.)

Solomon, R. (1992) *Ethics and excellence: Co-operation and Integrity in Business*, New York: Oxford University Press.(Adapts the Aristotelian concept of virtue to the demands of business ethics; works out virtues for managers.)

Sorell, T. and Hendry, J. (1994) *Business Ethics*. Oxford: Heinemann.(A review of the subject from a British and western European perspective.)

Butler, Joseph (1692-1752)

Joseph Butler the moral philosopher is in that long line of eighteenth-century thinkers who sought to answer Thomas Hobbes on human nature and moral motivation. Following the Third Earl of Shaftesbury, he rejects any purely egoistic conception of these. Instead, he analyses human nature into parts, of which he notices in detail appetites, affections, and passions on the one hand and the principles of self-love, benevolence and conscience on the other. His ethics consists in the main in showing the relation of these parts to each other. They form a hierarchy, ordered in terms of their natural authority, and while such authority can be usurped, as when the particular passions overwhelm self-love and conscience, the system that they constitute, or human nature, is rightly proportioned when each part occupies its rightful place in the ordered hierarchy. Virtue consists in acting in accordance with that ordered, rightly proportioned nature.

As a philosopher of religion, Butler addresses himself critically to the eighteenth-century flowering of deism in Britain. On the whole, the deists allowed that God the Creator existed but rejected the doctrines of natural and, especially, revealed religion. Butler's central tactic against them is to argue, first, that the central theses associated with natural religion, such as a future life, are probable; and second, that the central theses associated with revealed religion, such as miracles, are as probable as those of natural religion. Much turns, therefore, on the success of Butler's case in appealing to what is present in this world as evidence for a future life.

1 Life

For more than a century following the death of Thomas Hobbes in 1679, philosophers, theologians, and thinking men generally felt obliged to address his alleged reduction of morality to self-interest and his alleged banishment of God from the universe. Both views represent misreadings of Hobbes, but both quickly settled into received interpretations of him. The first view provoked an outpouring of moral philosophy that today, still, we associate with the work of the Cambridge Platonists, the Third Earl of Shaftesbury, Bernard Mandeville, Francis Hutcheson and, eventually, David Hume and Adam Smith. The second became allied with the claims of the British deists and so with the writings of men such as Herbert of Cherbury, John Toland, Anthony Collins, Matthew Tindal, Thomas Chubb, and, later, Hume, Edward Gibbon and Thomas Paine. The two views were connected, since religious scepticism was widely thought to be destructive of morality.

No figure is more central to these intellectual currents of the period than Joseph Butler. With *The Analogy of Religion, Natural and Revealed, to the Constitution and Course of Nature* (1736), he dealt Deism a serious blow. Though one of the chief monuments of English philosophical theology, the *Analogy* was a topical work, directed against Tindal and the deists, and today, since the disputes to which it is a response have died out, it is little read. With Butler's other major work, matters are quite different. His *Fifteen Sermons Preached at the Rolls Chapel* (1726) is one of the enduring classics of British moral philosophy, and parts of it, such as the refutation of psychological egoism, are widely celebrated. To the *Analogy*, Butler appended two indices - one on ethics, the other on personal identity, which mounts a lasting criticism of John Locke on the subject.

Butler was born in May 1692, in Wantage, Berkshire. The son of a dissenter, he was sent to Samuel Jones' academy in Tewkesbury, from where he corresponded with Samuel Clarke about several matters arising out of Clarke's Boyle Lectures (1704-5) on the being and attributes of God (see Clarke, S.). (Today, this exchange is little read.) Butler impressed Clarke and subsequently enjoyed his support. At Tewkesbury, Butler converted to the established Church and so was able to enter Oriel College, Oxford, in 1714, where he again made important and powerful friends. Upon his taking his degree in 1718, Butler's supporters secured his appointment as Preacher at the Rolls Chapel in London, and his rise in the Church began. In time, he would become Bishop of Bristol, Dean of St. Paul's, and Bishop of Durham. That he enjoyed royal favour was obvious from his appointment as Clerk of the Closet, first to Queen Caroline in 1736, then to George II himself in 1747. Shortly after his assumption of the See of Durham in 1750, Butler fell ill and eventually took himself away to Bristol and then Bath, where he died in June 1752.

2 Moral philosophy

In the *Sermons* (preface: 8), Butler's aim is to explain 'what is meant by the nature of man, when it is said that

virtue consists in following, and vice in deviating from it' and to show that this claim about virtue or morality 'is true' (1765: 8). Everything turns, therefore, upon what this nature is taken to be. Over this matter, however, commentators do not agree.

According to Butler, human nature is a system of parts - these are the particular passions, benevolence, self-love and conscience - and of the relations of these parts. In this last regard, Butler makes several points. Some parts of human nature are superior to others in terms of natural authority, which in turn orders the system; natural authority is to be distinguished from actual strength; and all cases of the usurpation of authority by strength are violations of human nature. The usual interpretation of Butler can now be stated. Virtue consists in acting in accordance with the nature of man, when the parts of that nature and system are in right proportion. These parts are in right proportion when our inward principles (whereby he means general principles of action or motives) exhibit their ordered authority, and these principles exhibit their ordered authority when the particular passions are controlled and regulated by benevolence and self-love and when conscience or the principle of reflection controls and regulates benevolence and self-love, as well as the particular passions, and so reigns supreme over the system. Human nature is thus a three-level affair: the particular passions; the principles of benevolence and self-love; and conscience.

A more radical interpretation is possible. This takes seriously passages in which Butler marks off the authority of self-love from that of benevolence, as when he remarks that 'reasonable self-love and conscience are the chief or superior principles in the nature of man' (*Sermons*, III: 13) as well as the fact that his examples of benevolence almost invariably involve particular passions. It also takes seriously Butler's claim about self-love and conscience that 'an action may be suitable to this nature, though all other principles are violated; but becomes unsuitable if either of those are'(*Sermons*, III: 13). Accordingly, human nature can appear to be a two-level affair, with the particular passions and benevolence on one level and self-love and conscience, now reigning jointly over the system, on another. Plainly, on this interpretation, the role of self-love in Butler's ethics is much more enhanced than on the usual interpretation.

When the inward principles in man exhibit their ordered authority, the system of human nature is in harmony, balance, and right proportion. Acts may be proportionate or disproportionate to man's nature; when they are proportionate, they are natural, when disproportionate, unnatural. An act is unnatural or disproportionate, not because it fails to be in accordance with the strongest principle in one at the time, but because it involves the usurpation of authority by strength. If hatred overwhelms self-love, then even though hatred is as much a part of our nature as self-love, the resultant action is unnatural. Thus, virtue consists in acting in accordance with the constitution of man, when the parts of this constitution are in right proportion; the act in question, natural and proportionate to human nature, is right. Vice consists in allowing a lower part of our constitution in terms of authority to predominate over a higher one and acting accordingly; the act, unnatural and disproportionate, is wrong. Thus, rightness and wrongness are matters of proportion, and the disproportion that a wrong or unnatural act involves is determined, not by its consequences, but by comparison of it with the nature of the agent, with, that is, whether it is suitable to the ordered system that is our nature. Thus, Butler is no consequentialist.

Particular passions. We have all manner of appetites, affections, and passions, from hunger, thirst, and bodily needs generally to compassion, love, and hate. Butler normally lumps all these together in order to contrast them with self-love and conscience, so that he usually writes as if the passions formed a single principle of action or motive. With regard to the particular passions, Butler emphasizes three points. First, we have all kinds of such passions or desires, and every one of them is 'a real part of our nature' (*Sermons*, III: 1). Second, some of our passions seem 'to respect self, or tend to private good'; others seem 'to respect others, or tend to public good' (*Sermons*, I: 6). That we have other-regarding passions, such as love, pity and compassion, Butler never doubts, and he castigates Hobbes for holding otherwise. These other-regarding passions are as much a part of our nature as our self-regarding ones. Third, our self-regarding and other-regarding passions are not to be identified with self-love and benevolence, respectively. For example, we cannot equate satisfied desire with self-love; for the gratification of a particular desire is often at the expense of self-interest.

Self-love. Self-love is a general desire for one's happiness, and it is 'inseparable' from any creature 'who can reflect upon themselves and their own interest or happiness, so as to have that interest an object to their minds' (*Sermons*, XI: 3). It is thus, unlike passion, a reflective or rational principle. As such, it provides us a measure of

control and regulation of the passions, in which task it is assisted by conscience, and it is a calculative principle, ranging over the likely consequences of acts to determine which are likely to be in our (long-term) interest. Butler's emphasis is always on restraining the passions, else they usurp the authority of self-love and conscience and so motivate us at the expense of our interest and judgment. Significantly, too, whereas the passions are occurrent - particular desires for this or that thing - self-love is an abiding, general desire for our own happiness.

Benevolence. If benevolence were a general principle of action on a par with self-love, we should expect Butler to maintain that it is an abiding desire for the happiness of others and so is an independent and reflective principle that aims at their happiness and restrains and regulates the passions. But he nowhere asserts that we have any such abiding desire, especially a general desire aimed at others indiscriminately or at humanity in general. His examples of benevolence are more limited - typically, confined to family and friends - and thus appear to be particular passions. He does insist that benevolence is not contrary to self-love, but this only makes the point, dubious if taken too far, that being benevolent can make us happy and serve private ends.

Happiness. Though self-love aims at happiness, it is not the same thing as happiness. Happiness consists in 'the enjoyment of those objects, which are by nature suited to our passions'; self-love simply 'helps us to gain or make use of' those objects (*Sermons*, XI: 6). We do not make things suitable to our passions; nature does this. We simply find that we take delight in certain things as a result of their engaging our passions, and self-love helps put us in the way of securing them. Without particular passions, then, happiness could not exist, since it 'consists in the gratification of particular passions' (*Sermons*, XI: 6). Thus, to aim at happiness makes no sense; one needs to aim at securing the objects of some of one's particular passions.

3 Moral philosophy (cont.)

Butler now adds an important point. Particular passions are particular desires for particular objects, and we individuate these desires by means of their objects. One is a desire for food, another for revenge, and what satisfies these desires is obtaining their particular objects. This feature of desires is quite independent of whether the desire, or its satisfaction, is to our interest or happiness. A desire for heroin is aimed at having the drug, and an injection of heroin satisfies this desire; this is true even if such a desire, or its satisfaction, is not to our interest.

Psychological egoism. The refutation of psychological egoism applies the content of the preceding discussion. For example, psychological hedonism is the view that each of us pursues exclusively our own pleasure. Were this view true, all of our desires would have the same object, pleasure, and this is plainly false. As a doctor who desires to cure my patient, I do not desire pleasure; I desire that my patient be made better. In other words, as a doctor, not all my particular desires have as their object some facet of myself; my desire for the well-being of my patient does not aim at alteration in myself but in another. My desire is *other*-regarding; its object is *external* to myself.

Of course, pleasure may arise from my satisfied desire in such cases, though equally it may not; but my desire is not aimed at my own pleasure. The same is true of happiness or interest: my satisfied desire may make me happy or further my interest, but these are not the objects of my desire. Here, Butler simply notices that desires have possessors - those whose desires they are - and if satisfied desires produce happiness, their possessors experience it. The object of a desire can thus be distinguished from the possessor of the desire: if, as a doctor, my desire is satisfied, I may be made happy as a result; but neither happiness nor any other state of myself is the object of my desire. That object is other-regarding, my patient's well-being. Without some more sophisticated account, psychological egoism is false.

Conscience. Conscience is a principle of reflection that judges acts right or wrong and characters and motives virtuous or vicious. It also, however, approves or disapproves of what it finds. Thus, in the dissertation *Of the Nature of Virtue*, an appendix to the *Analogy*, conscience is described both as a 'sentiment of the understanding' and 'a perception of the heart', and in the *Sermons*, Butler maintains that it is in virtue of conscience that man 'is a moral agent'. Thus, he gives every evidence of rejecting the view that morals are either a matter of reason or a matter of sentiment, an option upon which Hume was subsequently to insist.

As a reflective principle, conscience reflects in the case of acts upon whether they are natural and proportionate to our nature, and this brings in the authority of parts. To say of conscience that it possesses 'supreme' authority is to say that no act may violate or usurp it and remain proportionate to the system that is our nature. But this Butler says of self-love as well, and the implication seems to be that self-love and conscience are the 'chief or superior

principles' in man. Exactly in what the supremacy of principles consists is a matter of controversy, particularly as regards conscience; but naturalness and unnaturalness are crucial to its understanding.

The pronouncements by conscience of naturalness or unnaturalness are not calculative, for example, based on consequences, but immediate and, presumably, infallible. We have 'the rule of right within', and we have only to attend to it, to appreciate its authority, if not its strength, to see our way, morally (*Sermons*, III: 3). Much is unclear here, both because Butler does not give particular examples of how determinations of rightness are arrived at by appeals to naturalness and because the kind of intuitionism that seems implied by his remarks is simply assumed in its results to be uniform throughout all 'plain' and 'honest' men (*Sermons*, III: 4).

4 Moral philosophy (cont.)

If benevolence does not conflict with self-love, neither does conscience; indeed, conscience and self-love are 'perfectly coincident', if not here, then in the hereafter. They always, Butler says, 'lead us in the same way' (*Sermons*, III: 13). Apart from the invocation of God and the hereafter, there is no warrant for this claim in Butler's account of human nature, which poses a problem. If duty and interest pull us sometimes in different directions, and if we cannot violate either and remain true to our ordered nature, what do we do? A virtual guarantee of the coincidence of duty and interest seems required to avoid this impasse, and an appeal to the goodness of God is supposed to provide it.

Revelation. Curiously, this invocation of the goodness of God strikes an odd note. For Butler stresses that, exclusive of any belief in revealed religion, man is 'a law to himself' (*Sermons*, III: 3) and that one's obligation to obey this law flows from 'its being the law of your nature' (*Sermons*, III: 6). This obligation remains in force, even if 'the prospect of a future life were ever so uncertain'. But if we assume that a future life is uncertain, how can we use the hereafter to guarantee that we will not have to face a choice between duty and happiness? And if we do face that choice, Butler says that, when we 'sit down in a cool hour', we cannot justify the pursuit of virtue to ourselves 'till we are convinced that it will be for our happiness or at least not contrary to it' (*Sermons*, XI: 21). What if we are not convinced? Even treating the 'cool hour' passage as aberrant, the problem of a choice between duty and interest remains.

Strength and motivation. One of the most pronounced features of the *Sermons*, namely, Butler's repeated insistence that benevolence, conscience, and virtue are not opposed to self-love, can now be understood. Human nature may be hierarchical, based on the authority of parts, but the strength of self-love, prudence, and the self-regarding passions is great. Even if we have other-regarding passions and benevolent affections, will they be strong enough to motivate us at the expense of interest? If it could be shown that they were not at the expense of interest but actually served it, then one could hitch, for example, self-love to the cause of benevolence, and so perhaps increase the latter's chances of motivating us. Otherwise, we have to contend with the facts that (1) our other-regarding passions are rarely as strong as our self-regarding ones; (2) we have no general desire for the happiness of others; (3) our particular passions for the happiness of certain others typically do not carry us beyond the persons whose happiness is the object of those passions; and (4) prudence and self-love are powerful driving forces of our nature. What reason have we to think that we will choose benevolence over self-love?

Nor does conscience come off any better. For even if we allow it a motivational element and so regard it as more than judgment or reflection, that is, even if we allow that it can generate a desire in opposition to the desires in which the self-regarding particular passions, prudence, and self-love flow, Butler laments that its 'strength' and 'power' may be inappropriate to its authority. In fact, then, we may not be sufficiently motivated either to act benevolently or to pursue virtue. What Butler needs is a policy of insurance in these regards. That policy in both cases is to bring happiness and self-love to the rescue; hence, the repeated insistence that benevolence, conscience, and virtue are not 'contrary' to happiness or interest.

In the case of conscience and virtue, then, the motivational problem can be acute. Butler never says of actual individuals that they will come to love to be virtuous for its own sake; accordingly, it matters to his position that duty and interest or happiness coincide. But if they imperfectly coincide in this world, and if I am powerfully driven to seek my own happiness, then exactly how is the knowledge that God will rectify matters in the hereafter supposed to motivate me sufficiently powerfully to forego my happiness in this world? And even if we assume this knowledge about God gives rise to or is accompanied by a desire to forego my happiness in this world, why is this

desire assumed to be more powerful than my desire for my happiness, unless I do in fact love virtue for its own sake? An answer is required, if conscience is to be the 'guide' of our moral lives.

5 Philosophy of religion

By the time the *Analogy* was published, deism was rampant. It consisted not so much in a positive body of doctrine as in doubts about the Christian stories and the Church, and the sowing of such doubts, even when belief in God remained, was worrisome to the orthodox. For mere belief in God's existence, obviously, would not validate the specific claims of the Christian revelation. To validate these, Butler's effort is in two parts. First, he tries to show that, if the deists accept the existence of an intelligent Creator of the natural world, then when they examine God's handiwork, in order to find out what it contains and how it operates, they will find probable: (1) that there is a future life, in which 'our capacity of happiness and misery' persists; (2) that 'our happiness and misery Hereafter' depend 'upon our actions Here', in terms of God's rewards and punishments; and (3) that 'our present life is a state of probation for a future one', in that our present life exposes us to temptations that we must resist, if we have a concern for our future happiness. This material occupies Part I of the *Analogy*. Second, Butler turns in Part II to parts of the Christian revelation, and tries to show that when we look at what we find in this life, it is probable not only that there would be such a revelation, parts of which are 'imperfectly comprehended' by us and so about which we are doomed to 'ignorance', but also that many of the distinctive doctrines of Christianity, such as the coming of the Messiah and miracles by a divine personage, pass muster. His aim in this latter regard is to show that 'there is no presumption against a revelation as miraculous' and that 'the general scheme of Christianity, and the principal parts of it, are conformable to the experienced constitution of things, and the whole perfectly credible'.

While, doubtless, there are individuals who regard Part II of the *Analogy* as an integral part of their defence of Christianity, most modern readers will find Part I the more intriguing. It is there that Butler displays his argument form - analogical or inductive reasoning from what we find here in this life to what is probable in the life to come - and shows us what he is going to regard as evidence here for the probability of a claim about the hereafter. On both counts, the modern temper is likely to register deep scepticism, on the one hand about the usefulness of inductive reasoning from here to the hereafter, on the other about exactly what we will take as evidence for the probability of claims about the hereafter, for example, that God will reward and punish us for what we do here. Of course, to some extent, this scepticism and talk of evidence for claims about the hereafter get wrapped up in all the discussion since Hume's *Dialogues Concerning Natural Religion* (1779) about design and purpose in the natural world, and there is no doubt that, as far as the Christian revelation is concerned, Butler must achieve rather impressive feats with the evidence to make probable that God has 'by external revelation, given us an account of himself and his moral government over the world' (*Analogy*, II, vii: 31) (see Hume, D. §6). Then, too, general worries about induction and what it can establish are never far from the surface here, as Butler meticulously marshals his evidence in favour of the truth of natural and revealed religion.

Butler certainly believes that the evidence for natural and revealed religion is considerable and their probable truth high. But what if the deist does not go along with this, as, indeed, he might not, given that at the time what passed as evidence was being increasingly called into question? Every fresh attack on or alternative explanation of it lowered the probable truth of natural and revealed religion. Here, in a way, is the test Butler must meet. He tries to meet it with two claims. First, even if the probability of Christianity is low or slight, it does not follow that it is false. Second, even if its probability is low or slight, if we act, then we must act 'upon that presumption or low probability, though it be so low as to leave the mind in very great doubt which is the truth'. In life, probability must be our guide, and this is true even when the probability in question is slight. Put differently, if we have any regard for our own happiness, we must act on the basis of what is more probable, though this probability be extraordinarily small. This is true in the case of Christianity: if its truth be ever so slightly more probable than its falsity, then we should act upon it. In *A Charge to the Clergy of the Diocese of Durham* (1751), Butler turns this into something akin to Pascal's wager when he remarks of Christianity that, even if it were only slightly probable, 'it ought in all reason, considering its infinite importance, to have nearly the same influence upon practice, as if it were thoroughly believed' (see Pascal, B.). In the *Analogy*, Butler puts this point about conduct very emphatically: whether Christianity is true or may be true, whether it is certain or doubtful 'in the highest supposable degree', makes no practical difference. Doubting, he says, 'necessarily implies some degree of evidence for that, of which we doubt', and some, even slight evidence for Christianity 'does as really lay men under obligations, as a full

conviction that it is true'.

This argument for Christianity can appear a modest one, far removed from the ringing endorsements to which the orthodox were accustomed; in fact, this very modesty - few assumptions, few grand claims, no a priori reasoning, reliance upon evidence - would appear to be part of Butler's appeal. Yet, the argument itself seems to point the way to a beginning of criticism of it. If slight evidence in favour of Christianity should induce us (to believe it or) to act in accordance with it, should slight evidence against it induce us (to disbelieve it or) to act contrarily? If probability is the guide of life, what if there is slight probability against an afterlife, miracles, and so on? If Butler responds that there can never be such evidence or probability, this seems implausible; after all, Hume adduces at least *some* evidence for rejecting miracles. If Butler responds that such evidence or probability is never at a level to be compelling, then, even if this is true, why are his opponents forced to a level of compelling evidence, in order to have evidence enough to act upon? And if Butler responds that we have to weigh and balance the evidence in order to decide where the balance of evidence falls, then we need to be told what the principles are by which we weigh and balance. No such principles are set out, and weighing and balancing evidence is not a process on display in the *Analogy*. In short, at least one reading of Butler is that he holds that there is nothing in the natural world (and the claims of natural and revealed religion) that could give one even the slightest evidence against belief in an afterlife or miracles, so that such things can be held by him to be probable because those who oppose them are bereft, not of compelling evidence, but of evidence altogether. As it were, religion wins because there is in effect nothing to oppose it. The deists would not have accepted this, nor today, perhaps, would anyone not already religious, either in the natural or, more realistically, the revealed sense.

6 Personal identity

In the dissertation *Of Personal Identity* Butler turns to an issue of central importance to the *Analogy* - an afterlife, and whether we shall be the same person then as we are now. This leads him directly into a discussion of in what sameness of person consists, and on this score he makes (at least) four points, which in the main are with respect to Locke's discussion of personal identity in his *Essay Concerning Human Understanding* (1689).

First, 'upon comparing the consciousnesses of one's self, or one's own existence, in any two moments', there 'immediately arises to the mind the idea of personal identity'. Such a comparison shows that one is the same self from moment to moment.

Second, though this consciousness now of what is past shows our identity to ourselves, it does not constitute our identity. As Butler puts it, 'consciousness of personal identity presupposes, and therefore cannot constitute, personal identity; any more than knowledge, in any other case, can constitute truth, which it presupposes'. Trying to analyse personal identity in terms of memory, therefore, as Locke at one point is taken to have done, will not work; memory presupposes personal identity.

Third, in response to Locke's question of whether the same self is the same substance, a question that arises because consciousness of one's existence at two different times are two successive consciousnesses, Butler observes that such successive consciousnesses, while numerically different, are 'consciousnesses of one and the same thing or object; of the same person, self, or living agent'. Consciousness that one is the same person or self is consciousness that one is the same thing, which, if this thing is called substance, as in Locke, is consciousness that one is the same substance. Butler notices that Locke defines person as 'a thinking intelligent being' and personal identity as 'the sameness of rational being', and he maintains, accordingly, that the same rational being is the same substance, since being and substance here 'stand for the same idea'.

Fourth, we may distinguish between loose and strict senses of 'sameness'. In the loose sense, a tree remains the same over time even though it loses or gains parts or properties; indeed, over time, all of its parts or properties may change and it remain the same tree. In the strict sense, however, this is impossible; that is, it is a contradiction to say of the tree then and now that it is the same 'when no part of their substance, and no one of their properties is the same'. Sameness of persons refers to this second sense, so that sameness in this regard 'cannot subsist with diversity of substance'.

All four of these points have figured prominently in discussions of personal identity, and theorists even today often develop elaborate accounts of sameness of person either around or with them. *Of Personal Identity* remains very much alive (see Personal identity).

See also: Conscience; Shaftesbury, Third Earl of

R.G. FREY

List of works

Butler, J. (1896) *Collected Works*, ed. W.E. Gladstone, Oxford: Clarendon Press, 2 vols.(Volume 1 contains *The Analogy*, the dissertations *Of Personal Identity*, *Of the Nature of Virtue* and the *Correspondence with Samuel Clarke*. Volume 2 contains Butler's various sermons and the *Charge to the Durham Clergy*.)
Butler, J. (1804) *The Works of Joseph Butler*, Edinburgh: Constable.(This includes a life of Butler by Dr. Kippis.)
Butler, J. (1900) *Collected Works*, ed. J.H. Bernard, London: Macmillan.(This edition of Butler's collected works is more difficult to find and employs a different division of the texts from Gladstone's division.)
Butler, J. (1716) *Several Letters to the Reverend Dr. Clarke, from a Gentleman in Gloucestershire*, London.(A discussion of some points in Clarke on the being and attributes of God.)
Butler, J. (1726) *Fifteen Sermons Preached at the Rolls Chapel*, London.(One of the classics of British moral philosophy.)
Butler, J. (1736) *The Analogy of Religion Natural and Revealed, to the Constitution and Course of Nature*, Dublin: Ewing.(One of the great achievements of British theology.)
Butler, J. (1751) *A Charge to the Clergy of the Diocese of Durham*, London.(An exhortation to the Clergy of Durham to look to the well-being of their flock.)
Butler, J. (1765) *Fifteen Sermons to which are added Six Sermons Preached on Public[k] Occasions*, London: Horsfield.(Six sermons on topics of interest in the period, including charity schools and liberty.)

References and further reading

Broad, C.D. (1923) 'Butler as a Theologian', *The Hibbert Journal* 21: 637-56.(Analysis of the main argument lines of the *Analogy*.)
Broad, C.D. (1930) *Five Types of Ethical Theory*, London, Routledge & Kegan Paul.(Survey of Butler's ethics.)
Cunliffe, C. (ed.) (1992) *Joseph Butler's Moral and Religious Thought*, Oxford: Clarendon Press.(Collection of original articles on Butler's ethics and philosophy of religion.)
Duncan-Jones, A. (1952) *Butler's Moral Philosophy* Harmondsworth: Penguin.(Overview of Butler's ethics, from an ordinary language perspective.)
Frey, R.G. (forthcoming) *Joseph Butler*, Oxford: Oxford University Press.(Overview of Butler's ethics primarily, with chapters on his philosophy of religion and his views on personal identity.)
Hume, D. (1779) *Dialogues Concerning Natural Religion*, ed. N Kemp Smith, London: Collier Macmillan Publishers and New York: Macmillan Publishing Company, 1947.
Jeffner, A. (1966) *Butler and Hume on Religion*, Stockholm: Diakonistyrelsens Bokforlag.(Compares and contrasts the two thinkers.)
Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Oxford University Press, 1975.(A classic work in British empiricism and epistemology.)
Mossner, E. C. (1990) *Bishop Butler and the Age of Reason*, Bristol: Thoemmes.(Primarily sketches Butler's place in the deist controversies of the eighteenth century.)
Pascal, B. (1670) *Pensées and Other Writings*, New York: Oxford University Press, 1995, 152.(Jottings and aphorisms on God, religion and religious scepticism.)
Penelhum, T. (1959) 'Personal Identity, Memory and Survival', *Journal of Philosophy* 56: 82-903.(Discusses and makes use of Butler on personal identity.)
(1985) *Butler*, London, Routledge & Kegan Paul.(Analyses the main arguments of Butler's ethics and philosophy of religion.)
Perry, J. (ed.) (1975) *Personal Identity*, Berkeley, CA: University of California Press.(Collection on personal identity, including remarks on Butler.)
Shoemaker, S. (1959) 'Personal Identity and Memory', *Journal of Philosophy* 56: 868-82.(Refers to and builds upon Butler's discussion of personal identity.)
Stephen, L. (1962) *A History of English Thought in the Eighteenth Century*, London: Rupert Hart-Davis, 2 vols.(Discusses both Butler's ethics and philosophy of religion and places him in both respects amongst the thinkers of his age.)
Sturgeon, N. (1976) 'Nature and Convention in Butler's Ethics', *Philosophical Review* 85: 316-56.(Discussion of what it means 'to follow' in Butler's ethics.)
White, A.R. (1952) 'Conscience and Self-love in Butler's Sermons', *Philosophy* 27: 329-44.(Assesses the relation

between conscience and self-love in Butler's ethics.)

Byzantine philosophy

In Byzantium from the ninth century through to the fifteenth century, philosophy as a discipline remained the science of fundamental truths concerning human beings and the world. Philosophy, the 'wisdom from without', was invariably contrasted with the 'philosophy from within', namely theology. The view that philosophy is 'the handmaiden of theology', which the Greek Church Fathers derived from Philo and the Alexandrian school of theology, was not the dominant position in Byzantium as it was in the West; philosophy, and logic in particular, was never treated as a mere background to, or tool of, theology. By the same token, theology in Byzantium never developed into a systematic method of dialectical inquiry into Christian truths, or a science. Thus the initial distinction between philosophy and theology remained intact.

In terms of institutional practice, theological schools and studies did not exist in Byzantium and the main purpose of higher studies was to train state functionaries. This instruction, based on philosophy and the quadrivium, was mainly private, but it received support from the emperor and the church and we do hear of occasional interference by the secular or ecclesiastical authorities, perhaps because of professional or personal rivalries among the philosophy teachers. Furthermore, Byzantium had no independent universities or centers of study instituted by monastic orders as there were in the West, where social and political conditions were different.

Philosophy in Byzantium also steered clear of involvement in the theological controversies that arose from time to time. The prevalent model of the thinker in Byzantium was a sort of encyclopedic teacher of philosophy, an erudite scholar who kept in touch with the sciences of the quadrivium (arithmetic, geometry, astronomy and music) and other disciplines and set the philosophical tone of the scientific curricula. The development of philosophy in Byzantium was thus very different from that of Western scholasticism.

1 Historical outline

Although early Christian writers on the ascetic theory of life had adopted the term *philosophia*, the earliest manifestations of autonomous philosophical thought in Byzantium appeared in the ninth and tenth centuries with the 'Christian humanists' such as Photios, Patriarch of Constantinople, Arethas of Patras, Bishop of Caesarea, and Leo the Mathematician (or Philosopher). Photios elaborated the doctrine of the Trinity in the dispute over the procession of the Holy Spirit (the *filioque* dispute) using the armoury of Aristotle's theory of substances (the distinction between 'first substance' and 'second substance'). He was keenly interested in Aristotelian logic, rejecting Plato's self-existent 'ideas', and he collected works by many ancient writers. Arethas copied and commented on works by Plato and Aristotle and wrote critical notes on logic, ontology and psychology.

In the late eleventh and twelfth centuries, the growing study of philosophy reflects the great boost given to higher education and learning by the foundation in 1045 of the 'University' of Constantinople. Among the teachers known as *hypatoi tōn philosophōn* (first among philosophers) were Michael Psellos, undoubtedly the most important and most prolific of the Byzantine polymaths, Ioannes Italos, Theodoros of Smyrna, Eustratios of Nicaea and Michael of Ephesos. The last two are better known as commentators on Aristotle. The general outlook of the pre-eminent philosophers of this period, and the particular tendencies in their work, display the basic characteristics of Byzantine philosophy but with some distinctive features, such as an even stronger leaning towards the classical models of Greek philosophy and attempts to pursue a more autonomous line of inquiry into problems of knowledge, the natural world and human nature.

The temporary conquest of the Byzantine Empire by the Latin crusaders in 1204 shifted the centre of Byzantine intellectual life away from Constantinople. The flowering of literature and learning in at Nicaea, in Asia Minor, and the presence there of excellent teachers and writers of philosophy such as Nikephoros Blemmydes and Theodoros II Laskaris led to the emergence of generations of scholars well versed in philosophy and science. These men produced an impressive body of original work, especially in astronomy, during the politically troubled but culturally brilliant Paleologan period (1261-1453), the final two centuries of the Byzantine Empire. Outstanding among this group were Theodoros Metochites, Nikephoros Gregoras, Theodoros Choumnos, Georgios Pachymeres, Maximos Planoudes, Gennadios-Scholarios and Bessarion.

This splendid renaissance which coincided with the end of the empire also spread from the capital to other centres

such as Thessalonika and Mystras. Thessalonika was associated with the fourteenth-century Hesychast movement of Gregorios Palamas and his followers, a movement that had a considerable impact on Byzantine philosophy and also, more importantly, on the survival of Orthodoxy as a source and driving force of spirituality in the ensuing centuries of Turkish supremacy throughout the Balkans. Mystras, in the Peloponnese, was the home of the last great philosopher and perhaps the most original thinker of Byzantium, Georgios Gemistos-Plethon.

The awareness of Greek national identity had been cultivated, to a greater or lesser extent in earlier centuries but especially in the Palaeologan period, along with the development of Byzantine humanism. This movement had many features in common with the Italian humanism of the Renaissance. In particular, they shared a belief in the value and utility of the ancient Greek civilization with all its achievements, in the sciences no less than in other fields (see Humanism, Renaissance). There was a great surge of interest in the sciences, particularly mathematics and astronomy, and a number of major writers on these subjects emerged. Many works were also written on natural phenomena and cosmology.

2 The basic tenets of philosophical thought in Byzantium

The basic tenets that consistently characterized Byzantine philosophical thought throughout its history are first, the personal hypostasis of God as the principle not only of substance but also of being; second, the creation of the world by God and the temporal finitude of the universe; third, the continuous process of creation and the purpose behind it; and fourth, the character of the perceptible world as ‘the realization in time of that which is perceptible to the mind’, having its eternal hypostasis in the divine intellect (*nous*), the *Logos* (see Creation and conservation, religious doctrine of).

Phenomena, in Byzantine philosophy, are real hypostases to which the creator has given material existence by uniting them with the matter already created. Matter has no eternal existence: the ancient doctrine of uncreated, incorruptible matter is rejected utterly. Ideas have no self-existent hypostasis but are conceived by God and are instruments of God’s creative will and omnipotence. The general characteristics of sensible things (*universalia*) can be apprehended by the process of abstractive recognition; here the conceptual realism of the Alexandrian commentators on Aristotle is adopted and nominalism is rejected (see Aristotle Commentators; Nominalism).

Both the world and the human race are subject to divine providence, which is personal in character. Human self-determination does not conflict with the manifestation of divine providence, and Byzantine thinking on the subject presents a full affirmation of human free will. The soul as a spiritual substance, being by its very nature immortal, always belongs to one particular body. This personal soul is directly connected with the human being’s intellectual powers, which help him or her to achieve happiness by means of the freedom of decision. The relationship between God and human beings is based on love, which explains the central place of the human race in creation. Comparing Byzantine with ancient ontology, one can say in general that where the latter has ‘being’ the former has the doctrine of ‘existence’. The doctrines of the personal relationship between soul and body and the immortality of the personal soul are elaborated much more lucidly than in ancient philosophy (see Soul, nature and immortality of the).

3 General characteristics of Byzantine philosophy

The most important general characteristic of Byzantine philosophy, apparent in every period, is the absence of philosophical systems with real originality or independence. Instead, we find individual philosophers with an excellent classical education, extremely well versed in the writings of the ancient philosophers and with a discerning eye for subtleties of meaning. As far as Byzantine philosophers were concerned, the world view was fixed and crystalized and the study of philosophy was always regarded as a preparatory means to the ultimate end, which transcends the bounds of nature and cognition and consists in closer communion with God. Thus inquiry into such problems as the relation of faith to opinion and to knowledge - issues to which Christianity gave a new dimension with its concept of the individual - are the main areas in which philosophy is brought face to face with theology, and also Christianity with Hellenism.

On the one hand, therefore, Byzantine philosophers were quite familiar with the doctrines of the ancient Greek philosophers and had no difficulty in assimilating their ideas. It helped, of course, that they had a better knowledge of the language in which those texts were written than did those living outside the Greek-speaking world. A close

attunement to the Greek spirit led Byzantine philosophers to preserve the ancient philosophical texts, and to produce a fine set of commentaries on the most important of them (especially the Aristotelian and Neoplatonic writings) in the early post-classical period. In time this also generated an authentic (if unsystematic) philosophical tradition in the Byzantine world and produced some fine philosophers. Their main sources of inspiration were twofold: patristic theology (though with many of the ideas and methods of ancient Greek philosophy grafted on to it, after suitable modification) and first-hand study of Greek philosophical writings. The main purpose of studying these texts was to provide training in philosophical methods without regard to the 'rightness' or 'wrongness' of the views expressed in them, with the additional aim of resolving difficulties or rebutting theses that conflicted with the new (Christian) world view.

On the other hand, ancient Greek and Christian thought remained basically opposed to one another, despite the ever-increasing use of the ideas and methods of ancient philosophy which had helped to shape Christian doctrine in its final form. This antithesis was not all-embracing: some principles of pagan teaching in specific areas, chiefly ontology and cosmology, were rejected while others, especially in the realm of logical reasoning (for example, the apodictic syllogism, but not the dialectic method) and the natural sciences, were accepted. Moreover, the study of the ancient philosophers was not limited to commentary and interpretation but sometimes prompted critical discussion or even original elaboration of problems and the formulation of solutions. Thus for example, theological difficulties led to the clarification of philosophical concepts relating to the distinction between the essence (*ousia*) and activity (*energeia*) of God in the Hesychast controversy, where the debate was conducted using the armoury of classical metaphysics. The Byzantines always remained genuinely interested in grafting the teaching of the Scriptures and the Church Fathers onto ancient Greek wisdom, and in 'capturing concepts in Christ' and the eventual synthesis of 'philosophy' and 'knowledge of God'. This explains the ever-present influence of the great philosophers of antiquity throughout the Christian Middle Ages.

As regards the more specific and controversial question of the Byzantines' Platonism, Aristotelianism or Neoplatonism, the trends outlined above explain the frequent transformations undergone by many of the Platonic, Aristotelian and even Neoplatonic elements adopted by Byzantine thinkers, especially on such great issues as the concept of God, the relationship of the perceptible world to the absolute and the generation and essential nature of the soul. This has to be taken into account for an understanding of the so-called Platonism or Aristotelianism of the Byzantines and of their affinity with the thinking of the Neoplatonists. The affinity with any of these schools is largely external, inasmuch as the elements borrowed from them are chiefly methodological and logical, and the same language and terminology are used. This is particularly true of Neoplatonism, the phase of ancient Greek philosophy closest chronologically to the Byzantine period (see Neoplatonism). As regards the relationship of the most typical Byzantine philosophers with Plato and Aristotle, we find a balanced synthesis of logical and intuitive tendencies, with Aristotle on the side of logic and Plato on the side of intuition (see Aristotelianism, medieval; Platonism, medieval).

Leaving aside certain specific cases of accurate interpretation of the classical philosophers by Byzantine philosophers of the first rank, this attitude to both Plato and Aristotle gives Byzantine thought a typical manner of interpretation which leads away from the historical (that is, authentic) Plato and Aristotle. Furthermore, the preoccupation with the two great ancients ranges from pure study and interpretation for educational purposes to the fierce wrangling between rival schools that characterized the final centuries of the Byzantine Empire. In the former category we have the long line of Byzantine commentators on Aristotle, whose work established an authoritative and textually reliable Byzantine hermeneutic tradition which followed in the footsteps of the Alexandrian commentators, whose work they knew well and made good use of (see Aristotle Commentators). Thus they were instrumental in rehabilitating Aristotle's philosophy in the West after the Renaissance and ending the dominance of scholasticism and Averroism (see Averroism; Medieval philosophy; Renaissance philosophy).

The 'Platonism' of the Byzantines was the Platonic system which, having undergone many transformations, had been handed down to them from the Early Christian period and was often invoked for the defence of Orthodoxy against heresies. In the strife between the Hesychast movement of Gregorios Palamas and its opponents, many of the points at issue were Platonic concepts which had often exercised the minds of theologians and Christian philosophers in the past. In the fifteenth century, shortly before and immediately after the fall of Constantinople, individual interpretations of Plato and Aristotle led to the fierce philosophical dispute over the superiority of one or the other among the Greek scholars who had fled to Italy and were working as teachers or translators.

The interest taken in Neoplatonism by philosophers from Psellos to Plethon was manifested in a similar way, but it provoked a strong reaction, not only in terms of opposition to individual Neoplatonists such as Iamblichus and especially Proclus, but also in more general opposition on the part of the 'Aristotelians' to Orthodox theology and Western scholasticism in the dispute of the fourteenth and fifteenth centuries. The most obvious case in point is mystical theology, which is at once an ally and an adversary of Neoplatonism, especially in its attitudes to both Byzantine scholarship and Latin rationalism. During these two centuries mysticism was again the arena in which many philosophical concepts continued to exist not only externally but in personal practice. Here in mysticism, as also in Byzantine art, are to be found genuine metaphysical riches and experiential philosophical participation, which owes much to its contact with Neoplatonism. This explains the opposition of the church (or at least the official church) to this particular form of Byzantine spirituality, and it also accounts for Plethon's conflict with the church. Plethon had turned philosophy into a stirring experience and was fighting to resuscitate the Greeks' awareness of their national identity through philosophy.

The relationship between the Greek Orthodox East and the Catholic West (leaving aside the keen interest of both in Aristotle's philosophy) is of interest in connection with the conflict between the philosophical approaches of the two cultures. Despite attempts to improve mutual understanding in the Palaeologan period, when pro-Latin theologians and philosophers were active in the East, and despite the existence of numerous translations of Latin works (mostly by Thomas Aquinas) in the East, Byzantium remained closed to Western scholasticism right to the end. On the other hand, it is now known that Byzantium exerted a fertile influence on the West even before the forced migration of Greek scholars following the collapse of the Byzantine Empire in 1453 (see Aristotelianism, Renaissance; George of Trebizond; Platonism, Renaissance).

4 The study of Byzantine philosophy

Byzantine philosophy began attracting keen interest among scholars after the Second World War, long after most other periods and regions in the history of philosophy. A great deal of historical material is now available concerning the writings and teachings of major Byzantine philosophers and, more generally, concerning the cultivation of ideas and the trends of philosophical thought during those six hundred years.

Scholars now generally accept such fundamental premises as the unbroken continuity of Greek philosophy throughout the Greek-speaking world from classical times through late antiquity to Neoplatonism, recognizable in many tenets of the Church Fathers and in Christian Byzantium. In particular, the importance of linguistic continuity, which meant that ancient writings were widely used and understood in Byzantium, should not be underrated. Also, the accepted view of the relationship between theology and philosophy in Byzantium has been revised, with a more correct assessment of the dependence of Byzantine philosophy on patristic theology. Lastly, new facts have emerged and received opinions have been revised with regard to the relations and interaction between the Greek East and the Latin West in the realm of philosophy.

In conclusion, some methodological observations are called for concerning the background of Byzantine philosophy and the possibility of its systematic presentation. The philosophical output of this period, of which we have only fragmentary knowledge, is extremely varied and is spread over a wide variety of writings. Moreover, many important works are still unpublished or are available only in old and imperfect editions. Although many useful specialized books and papers have been written in recent years, the Byzantines' indebtedness to ancient philosophy has not been sufficiently studied or definitively assessed. Consequently any survey of the period could only be more or less schematic, though it can and must show the vigour and vitality of ancient Greek philosophy throughout the long era of Christian Byzantium.

See also: Aristotle Commentators; Greek philosophy: impact on Islamic philosophy; Islamic philosophy; Medieval philosophy; Neoplatonism; Patristic philosophy; Platonism, early and middle; Renaissance Philosophy

PHIL LINOS BENAKIS

References and further reading

Anastos, M. and Benakis, L.G. (1979-80) 'Philosophy in Byzantium', in *History of the Greek Nation*, Athens: Ekdotike Athinon, vol. 8, 266-73 (Anastos); vol. 9, 348-71 (Benakis).(In Greek.)

Benakis, L.G. (1971) 'Byzantine Philosophy. Forschungsbericht 1949-1971' (Byzantine Philosophy: Report on Research 1949-71), *Philosophia* 1: 390-433.(In Greek with German summary.)

Benakis, L.G. (1982) 'The Problem of General Concepts in Neoplatonism and Byzantine Thought', in D.J. O'Meara (ed.) *Neoplatonism and Christian Thought*, Norfolk: International Society for Neoplatonic Studies, 75-86.(Includes discussion of Neoplatonist elements in Byzantine philosophy.)

Benakis, L.G. (1986) 'Die Stellung des Menschen im Kosmos in der byzantinischen Philosophie' (The Position of Man in the Cosmos in Byzantine Philosophy), in *L'homme et son univers au Moyen Age, Actes du VIIème Congrès Internationale de philosophie médiévale (Man and His Universe in the Middle Ages, Transactions of the VIIth International Congress on Medieval Philosophy),* Louvain-la-Neuve, vol. I, 56-75.(Cosmology in Byzantine philosophy.)

Benakis, L.G. (1987) 'Grundbibliographie zum Aristoteles-Studium in Byzanz' (Basic Bibliography of the Study of Aristotle in Byzantium), *Aristoteles. Werk und Wirkung, Paul Moraux gewidmet*, Berlin: de Gruyter, 352-79. (Bibliography of Aristotelian works from Byzantium.)

Benakis, L.G. (1988) 'Commentaries and Commentators on the Logical Works of Aristotle in Byzantium', *Gedankenzeichen. Festschrift für Klaus Oehler*, Tübingen: Stauffenburg, 3-12.(Byzantine commentators on the Organon.)

Benakis, L.G. (1991a) 'Commentaries and Commentators on the Works of Aristotle (except the Logical Ones) in Byzantium', *Historia Philosophiae Medii Aevi. Festschrift für Kurt Flasch*, Amsterdam: Grüner, 45-54. (Byzantine commentators on Aristotle's other works.)

Benakis, L.G. (1991b) 'Bibliographie internationale sur la Philosophie Byzantine (1949-1990)' (International Bibliography of Byzantine Philosophy (1949-90)), *Bibliographie Byzantine publié à l'occasion du XVIIIe Congrès Internationale d'Études Byzantines (Byzantine Bibliography Published on the Occasion of the XVIIIth International Congress on Byzantine Studies)*, Athens: Comité Hellénique des Études Byzantines, 319-77. (Includes more than 400 titles in systematic and chronological order, plus bibliography on the Church Fathers and philosophy.)

CPhMA - Philosophi Byzantini (1984-92), Athens: The Academy of Athens.(The first critical editions of works by Nicholas of Methone, Nikephoros Blemmydes, Georgios Gemistos-Plethon, Georgios Pachymeres and others. Each work presents the full Greek text, accompanied by a critical apparatus, an introduction, a translation into English (Pachymeres), French (Plethon) or German (Blemmydes) and full indexes. Further volumes are in preparation. Works of philosophical interest by Michael Psellos and Plethon have also been published in recent years in editions by L.G. Benakis, U. Criscuolo, J. Duffy, P. Gautier, D. O'Meara and G. Weiss; see Benakis (1991b).)

Hunger, H. (1978-91) 'Philosophie', in *Die hochsprachliche profane Literatur der Byzantiner*, vol. I, Munich: Beck, 3-62; Greek translation with new bibliography by L.G. Benakis, Athens: Cultural Foundation of the National Bank of Greece, 1991, 39-122.(Summary of Byzantine philosophical writings.)

Hunger, H. (1993) 'Philosophie. B. Byzanz', *Lexikon des Mittelalters (Dictionary of the Middle Ages),* vol. VI, Munich: Artemis & Winkler, col. 2092-100.(The most recent survey, taking into consideration many new editions and relevant studies.)

Lemerle, P. (1971) *Le premier humanisme byzantin. L'enseignement et la culture à Byzance des origines au Xe siècle (The First Byzantine: Education and Culture in Byzantium from the Beginning to the Tenth Century)*, Paris: Presses Universitaires de France, Bibliotheque Byzantine.(Discusses humanism in Byzantium during the so-called Macedonian Renaissance.)

Meyendorff, J. (1973) *Défence des saints hésychastes. Grégoire Palamas (Gregory Palamas and the Defence of the Hesychast Saints),* Leuven: Spicilegium Sacrum Louvaniense.(On the fourteenth-century mystical movement, the Hesychasts, and their most prominent member, Gregory Palamas.)

Oehler, K. (1969) *Antike Philosophie und byzantinisches Mittelalter (Ancient Philosophy and Medieval Byzantium),* Munich: Beck.(On the inheritance of ancient philosophy in Byzantium.)

Oehler, K. (1990) 'Die Byzantinische Philosophie' (Byzantine Philosophy), in *Contemporary Philosophy: A New Survey*, vol. 6/2, Philosophy and Science in the Middle Ages, Dordrecht: Kluwer, 639-49.

Podskalsky, G. (1977) *Theologie und Philosophie in Byzanz. Der Streit um die theologische Methodik in der spätbyzantinischen Geistesgeschichte (14/15. Jh.) (Theology and Philosophy: The Debate About Historical Method in Late Byzantium (14th-15th Centuries))*, Munich: Beck.(Relations between theology and philosophy at the end of the Byzantine period.)

Tatakis, B.N. (1949) *La Philosophie byzantine (Byzantine Philosophy),* Paris: Presses Universitaires de France. (The first systematic survey of the subject, now somewhat out of date.)

Tatakis, B.N. (1969) 'La philosophie grecque patristique et byzantine' (Greek Patristic and Byzantine Philosophy), in *Histoire de la Philosophie*, vol. I, *Orient-Antiquité-Moyen âge*, Encyclopedie de la Pléiade, Paris: Gallimard, 936-1005.(Survey of late Greek, patristic and Byzantine philosophy.)

Woodhouse, C.M. (1986) *Gemistos Plethon: The Last of the Hellenes*, Oxford: Clarendon Press.(The life of Gemistos-Plethon, perhaps the most original thinker in Byzantium.)

Báñez, Domingo (1528-1604)

Domingo Báñez, once spiritual advisor to St Teresa of Avila, was a prominent Spanish theologian. In his commentaries on the Summa theologiae of Thomas Aquinas, he challenged an essentialist reading of Aquinas, and insisted that esse *(being) was an act. He is best known for his opposition to Molina's attempt to reconcile human free choice with divine foreknowledge, providence and grace. He also wrote on logic, and commented on Aristotle's On Generation and Corruption.*

Báñez studied at Salamanca, where he entered the Dominican Order in 1546. He held a series of teaching and administrative posts in Avila, Alcalá, Valladolid and Toro. In 1577 he returned to the University of Salamanca where he taught theology until he retired in 1600.

In logic, his *Summulae* (1599) stands out for its organization and its moderate use of nominalism (see Nominalism §1). Some of Báñez's points were original. For instance, in his section on signs, he rejected formal signs, reducing them to concepts and images, and taking away their character of mental language. As a result, he only accepted instrumental signs. He maintained the traditional division into natural and conventional signs, but explained that linguistic signs (especially onomatopoeia) were not so fully conventional that they had no relation with what is natural (see Language, Renaissance philosophy of §1). He restricted the scope of the section on consequences or inferences, and subordinated this material to syllogistic logic. His *Summulae* was very influential in Spain.

In metaphysics, he expounded an existential Thomism in his commentary on the first part of the *Summa theologiae* of Thomas Aquinas (§§6, 9). He argued that *esse* (being) cannot be seen as a categorial accident. Nor does it enter reductively into the categories, as Cajetan (§5) and Capreolus held, except in so far as what is participated in is that which realizes *esse*. Nor is *esse* an essence or something arising from essential principles. Rather, it is an act or actuality, and not just the last act, as Cajetan held, but both the first and the last. That is, *esse* not only completes a being, but makes it possible for there to be a being at all. All actuality comes from *esse*, for by it an essence is actualized. This is so in the ideal order as well as the real, for we can only think of something as existent, yet there is still no implication that being should be viewed as quiddity or essence. Being is limited by potentiality and essence, but it also gives content to things, not essential content, but that of existence or actuality. Although *esse* must suffer some imperfection in order to be received in created beings, it perfects essence without being perfected by it, and as a result *esse* is more perfect than essence (see Renaissance philosophy §4; Aristotelianism, Renaissance §2).

Báñez is best known for his work on human free choice. He replied to Luis de Molina's *Concordia* (first edition 1588) in a short work *De vera et legitima concordia liberi arbitrii (On the True and Legitimate Reconciliation of Free Choice)* (1600) (see Molina, L. de §2). Following Aquinas, Báñez argued that the primacy in any action belongs to God's efficacious grace, and that no reconciliation can remove this primacy. There is a physical premotion on God's part which does not affect human freedom, for God moves all things in this way, and humans are able to choose. Moreover, God sees humans as free without his omniscience being affected. In fact, Molina had sought to defend human freedom, but he thought it was enough that there should be a general or simultaneous concurrence of God, as primary cause, and human beings, as secondary causes, in the production of a determinate act. For Báñez this was not enough: a stronger concurrence, including physical premotion on the part of God, was also required. Molina also appealed to God's solicited grace which aided humans to do supernatural works without thereby necessitating their will, and to middle knowledge (*scientia media*) through which God knows what may be brought about with or without secondary causes. God can thus see future contingents (*futuribilia*) in his essence without having to establish predetermined decrees concerning them. Báñez, on the other hand, argued that God's grace must be efficacious, and he believed that God does not need any middle knowledge in order to preserve the purity of human free action. God sees everything at once, and so he sees what is determined as determined, and what is free as free. Báñez showed that Molina's doctrines could not be reconciled with those of Aquinas.

See also: Language, Renaissance philosophy of; Molina, L. de §2; Molinism

Translated from the Spanish by E.J. Ashworth
MAURICIO BEUCHOT

List of works

Báñez, D. (1571-1600) *Commentarios inéditos a la Tercera Parte de Santo Tomás (Unpublished Commentaries on the Third Part of Aquinas' Summa theologiae)*, ed. V. Beltrán de Heredia, Madrid: Instituto Francisco Suárez, 1951-3, 2 vols.

Báñez, D. (1583-1600) *Commentarios inéditos a la Prima Secundae de Santo Tomás (Unpublished Commentaries on the First Part of the Second Part of Aquinas' Summa theologiae)*, ed. V. Beltrán de Heredia, Madrid: Instituto Francisco Suárez, 1942-8, 3 vols.

Báñez, D. (1584) *The Primacy of Existence in Thomas Aquinas*, trans. B.S. Llamzon, Chicago, IL: Henry Regnery Company, 1966.(Translation of Báñez's commentary on *Summa theologiae* Ia.3.4.)

Báñez, D. (1584-8) *Scholastica commentaria in primam partem Summae Theologiae S. Thomae Aquinatis (Scholastic Commentary on the First Part of Aquinas' Summa theologiae)*, ed. L. Urbano, Madrid and Valencia: Editorial F.E.D.A., 1934; repr. Dubuque, IA: W.E. Brown Reprint Library, [1964?].(The Brown reprint only includes up to question 26.)

References and further reading

Gutiérrez Vega, L. (1954) 'Domingo Báñez, filósofo existencial' (Domingo Báñez: An Existential Philosopher), *Estudios Filosóficos* 3: 83-114.(On the emphasis that Báñez placed on *esse* within Thomistic metaphysics.)

Llamzon, B.S. (1964) 'The Specification of *Esse*: A Study in Bañez', *The Modern Schoolman* 41: 123-43.(A useful study of being and essence.)

Llamzon, B.S. (1965) 'Suppositional and Accidental *Esse*: A Study in Báñez', *New Scholasticism* 39: 170-88.(On the ways in which the being of substances and accidents differ.)

Muñoz Delgado, V. (1965) 'Domingo Báñez y las súmulas en Salamanca a fines del siglo XVI' (Domingo Báñez and the Summulae in Salamanca at the end of the 16th century), *Estudios* 21: 3-20.(On Báñez's reforms in logic and the great influence of his work.)

Cabanis, Pierre-Jean (1757-1808)

Cabanis believed in the possibility of a 'science of man', having its basis in medicine. He tried to show how a materialist conception of the human organism can throw light on our mental and moral life. The properties of living matter were derived from physical laws, but had their own peculiarities. In particular, the property of sensibility (being able to have sensations) and the property of motility (involving the experience of effort and of resistance to it) were the keys to understanding human nature.

Though the thrust of Cabanis' thought is materialistic, his emphasis on medical science distinguishes him both from the mechanistic tradition as represented by La Mettrie, and from the intellectualist tradition represented by Condillac, in which sensations are taken as given mental items, from which the rest of our mental life is constructed by operations of reasoning or association.

1 Life and works

Cabanis was born in Cosnac in the Limousin. He was registered as a medical doctor in Reims in 1784, after seven years' study in Paris (during which he had already become a protégé of Mme d'Helvétius, encountering Condillac, Condorcet, Benjamin Franklin, Mirabeau and Theodore Roosevelt in her circle). His radical ideas about the reform of medical practice and education would perhaps have made it difficult for him to be accepted by the medical establishment in Paris at the time. However, he did not make his profession as a doctor (though he treated Mirabeau, and published an account of Mirabeau's illness and death in 1791). Instead, he put his medical knowledge to political and philosophical use. In 1790, he wrote his *Observations sur les hôpitaux (Observations on Hospitals)* and this led to public office, including membership of the Commission on Hospitals, under the revolutionary régime. He also took an active interest in educational reform.

Like many members of the salon of Mme d'Helvétius at Auteuil, he withdrew from the public scene during the Reign of Terror, for fear of Robespierre. When he re-emerged, it was to stand for the ideals of reason, the perfectibility of the human species, and freedom. He was made Professor of Hygiene, and Professor of Clinical Medicine at the École de Médecine in Paris, and was elected to the Conseil des Cinq-Cents. On the creation of the Institut de France, he became a member of the short-lived 'Class of Moral and Political Sciences', where he delivered the series of memoirs which formed his major work, published in 1802: the *Rapports du physique et du moral de l'homme (Relations between the Physical and the Mental in Man)*. He became a senator in 1797, having supported Bonaparte's *coup d'état* of 18 Brumaire. However, he overtly opposed Bonaparte's growing authoritarianism. He belonged to a group of thinkers devoted to *idéologie* (the 'science of ideas'). The word was coined by Destutt de Tracy, a leader of that group, which included figures like Condorcet, Laplace and Lavoisier. But Napoleon soon adopted a repressive approach, and the appellation *idéologue* came to connote intellectual, social and political subversiveness. In about 1807, Cabanis seems to have composed a letter to Fauriel on first causes, in which he made concessions to the religious revival. The letter was published only in 1824, but we may surmise that Cabanis felt the need to seek some accommodation with the Imperial authority, as did other thinkers in the first decade of nineteenth-century Paris.

2 Thought

Cabanis was among those figures of the Enlightenment and post-Enlightenment period who believed passionately in the possibility and importance of a 'science of man'. The systematic understanding of brute matter which the mechanical philosophy had made possible should be matched in our understanding of the human species. This was not, of course, a new ideal, but what was distinctive in Cabanis' pursuit of it was his view that this understanding must be rooted in medical science, broadly conceived.

The memoirs which make up his *Relations between the Physical and the Mental in Man*, though materialist in cast, are concerned primarily with the special properties of living matter, especially in the human species. However, this was not necessarily a vitalist position. Indeed, Cabanis places his study of these properties in the context of a much more general principle, that of *attraction*, illustrated by gravity in physics, and by 'affinities' in chemistry and biology. Nevertheless, it is clear that he gives a certain autonomy to our understanding of living matter. Cabanis began this work by insisting that the moral and medical sciences must deal with human beings as whole creatures.

Like Descartes, he insisted on the union of mind and body, but unlike Descartes he was no dualist, adopting a broadly materialist approach, which gave a central role to sensation. The property of being able to have sensations, though peculiar to living matter, was derived from more general physical laws, and could give rise to unconscious as well as conscious psychological phenomena. The physiological underlayers of sensation could be seen at work in differences between the sexes, and other mental differences between people arising from their inherited physical constitution in interaction with environmental factors. Cabanis had no account of how life arose, and thought that we should not speculate about causes whose existence could not (as he thought) be experimentally verified, but he believed that one source of the development of life was the inheritance of acquired characteristics. This led him to argue for selective breeding of humans.

This summary of a diverse and wide-ranging set of memoirs is very incomplete, but it shows how Cabanis held a position which already had a substantial background, for instance, in the work of Boerhaave, Stahl, Bonnet, Haller, La Mettrie and others, authors whose importance he fully acknowledged. But these thinkers were divided. Could our understanding of living matter be a simple derivation from our understanding of matter in general? Or were there special types of explanation at play in the case of living matter? Some maintained that a single principle was at work in living matter, others, like Haller, made a radical distinction between a physical property of 'irritability' responsible for lower and unconscious functions not involving sensation, and a sensitive property which was responsible for sensation and other mental functions which were derived from it.

Cabanis, as we have already indicated, preferred a monistic position. Indeed, he was particularly known for his claim that just as the stomach digests food, so the brain is a device for digesting sense impressions: he said that it carried out 'the secretion of thought'. Some of his contemporaries, such as Maine de Biran, strongly rejected this claim. Nevertheless, Cabanis' systematic treatment of the physiological underlayer of our mental existence, with its exploration of internal factors such as sex, health, bodily chemistry, inherited dispositions, habituation, as well as external ones such as nutrition and climate, marked a distinct break with the intellectualist analyses of eighteenth-century sensualism, typified by Condillac. Like Bonnet, whom he admired, Cabanis could be seen as attempting to 'naturalize' sensualism. *Relations* is a wide-ranging work which also gives an important role to language, envisages 'transformism' (a precursor of evolutionary theory) and emphasizes the importance of motility: Cabanis claims that the consciousness of self requires the experience of effort, and of resistance to it (this view influenced Maine de Biran, who made it a main element in his own philosophy).

There is significant disagreement about how to read Cabanis' position. Some, including some modern commentators, have regarded him simply as a materialist. Others have viewed his work as endorsing some version of vitalism. The situation is further complicated by the posthumously published letter to Fauriel on first causes (*Lettre à Fauriel sur les causes premières*, 1824). Although called a 'letter', this is in fact a book-length opus. Some commentators have gone so far as to ignore it; others have viewed it as a recantation of Cabanis' former robust materialism; others again consider that it raises questions of method which were Cabanis' concern all along. On the latter interpretation, we should point out how, all along, Cabanis abjured the search for 'essences', 'occult powers' or 'first causes', a commonplace at the time. We should be empiricists, neither postulating nor speculating about the unobservable, but confining ourselves to what can be established by observation. The letter to Fauriel, however, distinguishes between what can be 'demonstrated' in this way, and what can be argued for only by calculating probabilities. The latter may, according to this 'letter', re-admit a religious world-view as something with a certain probability, even though it cannot be demonstrated by scientific methods. In the background were Châteaubriand's *Le Génie du Christianisme* (1802), and the influence of the Illuminati, and the many other factors which were bringing the end of the Enlightenment (see Illuminati).

Whether or not we regard this posthumously published 'letter' as a recantation, Cabanis' principal work had a substantial and fruitful impact on the early development of psychology and its connections with physiology, being reprinted several times up to the middle of the nineteenth century.

See also: Human nature, eighteenth-century science of; La Mettrie, J.O. de

F.C.T. MOORE

List of works

Cabanis, P.-J.G. (1956) *Œeuvres philosophiques (Philosophical works)*, ed. J. Cazeneuve and C. Lehec, Paris:

Presses Universitaires de France.(Cabanis' collected philosophical works.)
Cabanis, P.-J.G. (1790) *Observations sur les hôpitaux (Observations on Hospitals)*.(Included in Cabanis (1798).)
Cabanis, P.-J.G. (1791) *Journal de la maladie et de la mort d'Honoré-Gabriel-Victor Riquetti de Mirabeau (Journal of the Illness and Death of Mirabeau)*.(Included in Cabanis (1798).)
Cabanis, P.-J.G. (1798) *Du degré de certitude de la médecine (On the Degree of Certainty of Medicine)*, Paris: Firmin Didot; 2nd edn, 1803.(Second edition included *Observations on Hospitals* and *Journal of the Illness and Death of Mirabeau*.)
Cabanis, P.-J.G. (1798) *Rapport fait au Conseil des Cinq-Cents sur l'organisation des Écoles de Médecine (Report to the Council of Five Hundred on the Organization of Medical Schools)*; photographic repr. Paris: Éditions de la Cité des sciences et de l'industrie, 1989.(Works of Cabanis about the administration and practice of medicine.)
Cabanis, P.-J.G. (1802) *Rapports du physique et du moral de l'homme (Relations between the Physical and the Mental in Man)*, Paris; 2nd edn, Paris: Crapelet, 1805; photographic repr., Paris: Slatkine, 1980.(Distinctive views on the mind-body question. The 1980 reprint is of an 1844 edition.)
Cabanis, P.-J.G. (1804) *Coup d'œil sur les révolutions et sur la réforme de la médecine*, Paris: Crapart; trans. A. Henderson, *Sketch of the Revolutions of Medical Science, and Views relating to its Reform*, London: Johnson, 1806.(Proposals for medicine.)
Cabanis, P.-J.G. (1807a) *Observations sur les affections catarrhales en général: et particulièrement sur celles connues sous les noms de rhumes de cerveau et rhumes de poitrine (Observations on Catarrhal Infections Generally: In Particular on Those Known as Head Colds and Chest Colds)*, Paris: Crapart.(On catarrhal infections.)
Cabanis, P.-J.G. (1807b) *Lettre à Fauriel sur les causes premières (Letter to Fauriel on First Causes)*, ed. F. Bérard, Paris: Gabon, 1824.(On first causes; published posthumously.)

References and further reading

Azouvi, F. (1992) *L'Institution de la raison: la révolution culturelle des idéologues (The Institution of Reason: the Cultural Revolution of the 'Idéologues')*, Paris: Éditions de l'École des hautes études en sciences sociales, Vrin.(A collection of articles on the ideals and activities of Cabanis' intellectual circle.)
Cailliet, É. (1943) *La Tradition littéraire des idéologues (The Literary Tradition of the Idéologues)*, Philadephia, PA: American Philosophical Society.(A thorough study of the immediate intellectual milieu to which Cabanis belonged.)
Colonna d'Istria, F. (1911-7) Five articles in *Revue de métaphysique et de morale*, vols. 19-21, 24-5.(These articles constitute an effective overview of some main aspects of Cabanis' work.)
Moravia, S. (1974) *Il pensiero degli 'idéologues': scienza e filosofia in Francia 1780-1815 (The Thought of the Idéologues: Science and Philosophy in France 1780-1815)*, Florence: La Nuova Italia.(Another authoritative account of 'ideology'.)
Poyer, G. (ed.) (1910) *Cabanis: choix de textes et introduction (Cabanis: Selected texts and introduction)*, Paris: Louis Michaud.(Contains a brief but useful biography.)
Staum, M.S. (1980) *Cabanis: enlightenment and medical philosophy in the French revolution*, Princeton, NJ: Princeton University Press.(A very useful study.)
Tissot, C. J. (1843) *Anthropologie spéculative générale, (Speculative General Anthropology)*, Paris.(A valuable document showing the impact of the work of Cabanis and Maine de Biran on the early stages of empirical psychology.)

Cabral, Amílcar (1924-73)

Amílcar Cabral was founder and leader of the African Party for the Independence of Guinea and Cape Verde (PAIGC), which led a war of liberation in the Portuguese colonies of Guinea Bissau and Cape Verde that ended with the recognition of their joint independence by the Portuguese government in October 1974. Cabral was assassinated in 1973, the victim of an attempted coup aimed at taking over the PAIGC leadership. Thus he did not live to see the independence for which he had struggled. Cabral's importance for African political philosophy lies in his having developed an undogmatic left-wing analysis of the situation of the Guinean peasantry. While familiar with Marxist analysis, Cabral was always willing to adapt it to the empirical realities of the Guinean situation. His writings on the role of culture in the nationalist struggle, which have important affinities with Gramsci, combine theoretical ingenuity with detailed local knowledge.

1 Life

Amílcar Lopes Cabral was born on the Guinean mainland in Bafatá. Later he moved to the island of São Tiago in the Cape Verde islands. He began primary education only at the age of twelve, but by twenty he had completed in eight years an education that normally took eleven. During this period, Cabral participated in the literary current that grew up around the journal *Claridade* (1936-60). He wrote poetry that was nationalist but not explicitly anti-colonial, like much of the work of this group.

In 1945 he won a prestigious scholarship to the Higher Institute of Agronomy in Lisbon. He gained his degree in March 1952. While in Portugal he met other students from Portuguese Africa. He came to know a number of those who later became leaders in the independence struggles in Portugal's African colonies, among them Agostinho Neto, Angola's first president and Mario de Andrade, later head of the Popular Movement for the Liberation of Angola (MPLA); but unlike many of them, Cabral never joined the Portuguese Communist Party.

During this period Cabral was also engaged with what he called the 're-Africanization of the spirit'. He argued that this was a necessary step for 'assimilated' Africans like himself if they were to play a positive role in the nationalist struggle. Cabral's work at that time was strongly influenced by his reading of the *négritude* poets and of *Présence Africaine*, the major journal of that movement (see African philosophy, Francophone §4). Cabral returned to Guinea in 1952 to direct the research of a government experimental station near Bissau, the colony's capital. In 1953 he was asked by the authorities to conduct an agricultural survey of the colony. This experience played a crucial role in the development of his ideas. It provided him with first-hand knowledge of peasant life throughout the country and allowed him to 're-Africanize' himself by cultural immersion. In 1955 he returned to Portugal.

On returning to Guinea for a short visit in 1956 he helped found the African Party for the Independence of Guinea and Cape Verde (PAIGC). For the next four years Cabral worked as an agronomist in Portugal and Angola, where he was involved with the foundation of Angola's Marxist liberation movement (MPLA) through contacts from his student days.

In 1959 Cabral returned to Guinea to take part in a momentous meeting of the PAIGC at which some crucial decisions were taken. These included shifting the focus of the struggle from urban to rural areas, moving their political headquarters out of the country and promoting the idea of armed struggle against colonial rule. Moving their headquarters was possible because Guinée, the neighbouring French colony, had recently become independent. For the remainder of his life Cabral had a home in Guinée and the PAIGC conducted its programme of decolonization from a base there. From 1960 Cabral primary focus was on the political education of the party's core membership. Over the next decade the political mobilization of the Guinean peasantry played a central role in the work of the PAIGC.

From 1962 the PAIGC developed a guerrilla army trained to work with the peasantry. The group also built support from village chiefs and elders. Cabral was involved in the organization of both the war and the party's diplomacy as it sought to gain international support. He also organized the elections in 1972 which saw the creation of the National Assembly that declared independence eight months after his assassination in January 1973. The overthrow of the fascist regime in Portugal in the spring of 1974 led to Portugal's recognition of the independence

of Guinea and Cape Verde in October of that year.

2 Political theory

In ‘The Weapon of Theory’ (1966), probably his most widely-cited work, Cabral makes a number of quite specific criticisms and modifications of the Marxist tradition, as he understood it. For example, it cannot be the case that history begins only after the development of classes, as many European Marxists in the post-war period had argued. He claimed that there were no classes in the Marxist sense in much of Africa, Asia and Latin America before colonialism, or in early European history. His analysis of Guinean (and other) societies taught him there was not always class struggle throughout history, therefore, something else must have been the ‘motive force of history’ ([1966] 1980: 123) before the existence of class systems. This force is the mode of production, which means ‘the level of the productive forces and the system of ownership’ ([1966] 1980: 123). Class struggle did not exist in earlier forms of society in which there was a low level of development of the productive forces, such as technology and no private ownership of land. Cabral maintained that there was still historical development and offered an analysis of three stages of development of the mode of production, communal, agrarian and then socialist.

Within this analysis Cabral thought that European colonialism in Africa and elsewhere interfered with the historical development of African societies. He saw national liberation as ‘the regaining of the historical personality’ ([1966] 1980: 130) of a people. Foreign domination of the productive forces of an agrarian colony by a capitalist metropolis drives the history of a people according to a logic inappropriate to the level of productive forces in their society. Cabral went on to analyse the forms of stratification and the relations of production that imperialism imposed upon colonies.

Cabral’s political theory was developed in the context of his experience as a colonial subject and as the leader of a war of national liberation. Like all successful twentieth-century revolutionaries he succeeded in a context where there was almost no proletariat. He had to develop an account of his struggle which faced the reality that the PAIGC consisted of a petty-bourgeois leadership mobilizing a peasantry.

The key to national liberation, he thought, lay in the sphere of culture, but the role of culture depended on which class you were addressing. This was because of the ‘class-character’ of culture in the struggle for liberation. He resolutely opposed the notion implicit in *négritude* that the ‘re-Africanization of the spirit’, or ‘return to the source’ ([1966] 1980: 63) he had experienced as a member of the petty bourgeoisie was relevant to the situation of the rural peasantry. He felt this was because imperialism had had little effect on peasant culture. He also rejected *négritude*’s notion that there was a single African culture.

Cabral argued instead that it was necessary to nurture the popular culture of the indigenous peoples as the basis for a developing nationalist consciousness. With a critical approach to this aim a humanism that transcended nationality could be achieved. He also believed, perhaps because of his training in agronomy, that it was important to develop a scientific culture that would support the technology required for economic development.

The greatest challenge for national liberation was that the petty bourgeoisie would be tempted to become more bourgeois and ally itself with international capital. This is what happened to those societies that became neocolonial states after formal independence. (In Cabral’s view this would have included most of the states of Francophone Africa, except for Guinea.) The result was that the task of the revolutionary party was to shape the consciousness of the petty bourgeoisie before independence in such a way that they would become ‘capable of committing suicide as a class’ because they were ‘completely identified with the deepest aspirations of the people’ ([1966] 1980: 136). Cabral placed a great deal of emphasis on the cultural and ideological work of shaping the consciousness of the petty bourgeoisie, just as he had insisted on shaping the national consciousness of the peasantry. It is this aspect that bears a resemblance to Gramsci. This view is substantially at odds with much Marxist thinking which has denied the possibility that there can be a revolutionary petty bourgeoisie. Subsequent African history suggests that it is Cabral who is mistaken on this issue.

Cabral’s detailed analysis of the role of the popular culture of the peasantry as a basis for the development of national consciousness has had a great influence on cultural thought among intellectuals in Africa since the 1980s.

See also: Fanon, F.; Gramsci, A.; Marx. K.

K. ANTHONY APPIAH

List of works

Cabral, A. (1969) *Revolution in Guinea*, London: Stage One.(Speeches and writings in which Cabral discusses the conduct of the war of liberation in Guinea.)

Cabral, A. (1973) *Return to the Source*, New York: Monthly Review Press.(Selected speeches, including ’The Weapon of Theory’ and an important essay on ‘Identity and Dignity in the Context of National Liberation’, which discusses the idea of a ‘return to the source’ quite extensively.)

Cabral, A. (1966) ‘Weapon of Theory’, in *Unity and Struggle*, London: Heinemann, 1980.(Translates most of the major essays including ‘The Weapon of Theory’ and an important essay on ‘National Liberation and Culture’.)

References and further reading

Chabal, P. (1983) *Amílcar Cabral: Revolutionary Leadership and People’s War*, Cambridge: Cambridge University Press.(The first scholarly biography with a substantial chapter devoted to Cabral’s social and political thought. Also contains a thorough bibliography of Cabral’s writings.)

Chilcote, R.L. (1991) *Amílcar Cabral’s Revolutionary Theory and Practice: A Critical Guide*, Boulder, CO: Lynne Rienner.(Contains an annotated bibliography of work on and by Cabral.)

Claridade (1936-60), repr. in *Claridade: revista de arte e letras*, Linda-a-Velha, Portugal: Africa, Literatura, Arte e Cultura, 1986.

Présence Africaine (African Presence) (1947-), Paris.(This quarterly is the major journal of the *négritude* movement. It is published by the publishing house of the same name that has also produced many influential journals.)

Cajetan (Thomas de Vio) (1468-1534)

Thomas de Vio, better known as Cajetan, has long been considered to be the outstanding commentator on the philosophical thought of Thomas Aquinas. He has had a great influence not only on discussions about Aquinas' theory of analogical predication regarding God and creatures but also on discussions about Aquinas' fundamental notions of essence and existence. On both counts his interpretations are at variance with Aquinas himself. He also set himself in opposition to Aquinas when he denied in his later writings that the immortality of the human soul could be demonstrated, arguing that it is a doctrine that must be accepted simply on faith, like the doctrines of the Trinity and the Incarnation. His explication of Aquinas' cognitive psychology is an interesting development that goes beyond Aquinas.

1 Life

Thomas de Vio was born at Gaeta. Entering the Dominican Order there in 1484, he first studied at Naples. In 1491 he was sent to Padua where he studied theology and metaphysics. Valentino da Camerino, a Dominican, had been appointed to the university's chair of Thomistic metaphysics in 1489. After Valentino was elected in 1494 to be provincial of the Roman province, he resigned his posts as professor of Thomistic metaphysics and as regent master in the Dominican *studium generale*. This was done before the 1494-5 academic year began. Despite his youth, Cajetan was called upon to fill Valentino's two posts temporarily. At the end of that academic year, he left Padua. Subsequently he taught theology at Pavia from 1497 to 1499. By 1501 he was in Rome as the procurator of the Dominican Order. He was named vicar general in 1507 and the following year he was elected master general - the head of the Dominican Order. In 1517 he was named a cardinal by Pope Leo X. The following year, while on a diplomatic mission for Leo in Germany, he had discussions with Martin Luther. He later wrote treatises defending the papal position. During the 1520s Cajetan turned his attention to biblical exegesis and he published commentaries on almost all the books of the Bible.

2 Logic and analogy

Although Cajetan wrote commentaries on Aristotle's *Categories* and *Posterior Analytics* and on Porphyry's *Isagōgē*, he is better known for his views on analogy, which supposedly explicate the position of Thomas Aquinas (§9). He begins his *De Nominum Analogia (The Analogy of Names)*, which was finished at Pavia in 1498, by lamenting that contemporary explanations of analogy are badly mistaken and lead astray those who study metaphysics. He is thereby attacking the views of fellow Dominicans, some of whom are his contemporaries. Cajetan takes 'analogy' to mean proportion or proportionality, and he claims to derive this account of the word from the Greeks, that is, from Aristotle (§7). According to Cajetan, there are three types of analogy to which all other analogies can be reduced. They are (1) the analogy of inequality, (2) the analogy of attribution or proportion, and (3) the analogy of proportionality. According to the first, the analogy of inequality, there is a common term or word for the analogous things and the notion connected to the common term is wholly the same. However, the analogous things share unequally in the nature involved, for example, 'body' as found in corruptible bodies on the surface of the earth and 'body' as found in the heavenly bodies, which Cajetan (following Aristotelian tradition) takes to be incorruptible. Since the notion of body is wholly the same, namely a substance subject to three dimensions, the logician considers the analogous things to be univocal. On the other hand, the philosopher studies the actual natures of the analogous things and therefore takes the analogous things to be equivocal (different in kind). The second type of analogy, analogy according to attribution, involves things of which there is a common name and also a notion that is the same. The distinguishing feature of this type of analogy is that there are different notions involved, according to the different relationships between the analogous things and that common name. An example is the term 'healthy', when it is applied to an animal, a medicine and urine. Only the first of the analogous things actually has the perfection, namely health as pertaining to the animal, whereas the perfection is said of medicine and other things by an extrinsic denomination, that is, by a term denoting an external relationship such as 'cause of' or 'sign of'.

Finally there is analogy of proportionality, which Cajetan considers to be analogy properly so-called. The name or term and the notion are the same only proportionally. What is involved in proportionality is a comparison of proportions; for example, eight is to four as six is to three. Cajetan explains that philosophers have extended the

term ‘proportionality’ to any similarity of proportions, so that one can say substance is to its existence as accident is to its existence. He believes that only this type of analogy enables us to know the intrinsic being, goodness and truth of things. Accordingly, he ranks analogy of proportionality above the other two types of analogy. Indeed, he judges that those who do not know and accept it are incapable of successfully studying metaphysics. Cajetan’s position in fact stands in opposition to that of Thomas Aquinas, who held to analogy of proportionality only in his *Quaestiones disputatae de veritate (Disputed Questions on Truth)*. Otherwise Aquinas only maintained the analogy of proportion based on attribution or reference to one First Being, namely God, and on all things other than God participating in an existence given them directly by God (see Language, Renaissance philosophy of).

3 Natural philosophy

In an early question entitled ‘Whether Mobile Being is the Subject in Natural Philosophy’, finished at Milan in 1499, Cajetan states that the limits of natural philosophy extend to and include the human soul, which is partly separated from and partly joined to matter. Neither the Intelligences that cause the motion of the heavenly bodies nor the angels are studied in natural philosophy. However, philosophers disagree as to whether their subject-matter is mobile being, mobile body, natural body or sensible substance. Cajetan takes the followers of Thomas Aquinas to hold that mobile being (which includes more than mobile body) is the true subject of natural philosophy. Accordingly, things are studied in natural philosophy from the formal aspect of their mobility (see Vernia, N. §3).

Another key topic treated by Cajetan concerns the principle of individuation, the basis for there being a multiplicity of material individuals in the same species or kind. In his early commentary *In De Ente et Essentia divi Thomae Aquinatis (Commentary on ‘Being and Essence’)*, which was finished at Padua in 1495, Cajetan took prime matter and its capacity for a particular quantity - this quantity rather than that quantity - to be the principle of individuation. He calls it ‘designated matter’ (*materia signata*) and attributes the notion to Aquinas. He later rejected this position on the grounds that the individual unity of Socrates would then be based on something accidental to him, namely quantity. Cajetan thus holds in his *Commentaria in Summa Theologiae divi Thomae Aquinatis (Commentary on Aquinas’ Summa theologiae)* that prime matter, which is the root and cause of quantity, is prior to quantity, and of itself numerically distinct. Matter somehow pre-contains quantity.

4 Psychology

Cajetan’s position on the immortality of the human soul underwent an evolution. In a sermon on immortality preached before Pope Julius II in 1503, Cajetan gave three arguments for the immortality of the human soul: that the soul knows all bodies and stands above them in judgment; that the soul has a natural desire to live forever; and that human beings enjoy a freedom of action that radically distinguishes them from the animals and can be explained only by immortality. He again maintained that human immortality was demonstrable in his *Commentaria in Primam Partem Summae Theologiae (Commentary on the First Part of Aquinas’ Summa theologiae)*, finished in 1507. He accepts Aquinas’ argument that the soul must be immaterial and subsistent (that is, able to exist on its own), since it knows all material objects and therefore has an operation which it does not share with the body. He also accepts Aquinas’ idea that even though the intellect depends on the phantasm for the object of cognition it is independent of the body in its very act of cognition (see Aquinas, T. §10). Subsequently in his *Commentaria in De anima Aristotelis (Commentary on Aristotle’s On the Soul)*, finished in 1509, he points out that learned men have given opposing interpretations of Aristotle’s words. He himself takes Aristotle to deny human immortality, emphasizing that for Aristotle all human cognition requires the presence of a phantasm. On the other hand, he emphasizes that he rejects the false position that, according to the principles of philosophy, the possible intellect (as opposed to the agent intellect) is corruptible. Indeed, he offers arguments for immortality on his own account. Nonetheless Cajetan was attacked for his interpretation of Aristotle by such fellow Dominicans as his former teacher, Valentino da Camerino, as well as by Francesco Silvestri, Bartolomeo de Spina and Crisostomo Javelli.

Some years later Cajetan came to reject any rational demonstration of immortality. In his commentary on the Epistle to the Romans, finished at Gaeta in 1528, he lists immortality with the Trinity and the Incarnation as mysteries that he accepted on faith. And in his commentary on Ecclesiastes, finished at Rome in 1534, he takes Chapter 3 to place immortality in doubt, since humans and animals are made from dust and return to dust. He then states explicitly that no philosopher has ever demonstrated immortality, that there is no demonstrative argument for immortality, and that immortality is believed on faith or is accepted on the basis of arguments that are only

probable (see Soul, nature and immortality of the).

In *Commentaria in De anima Aristotelis*, Cajetan observes that the human soul is treated by the natural philosopher as if it is joined to the body, but by the metaphysician as if it can be considered in abstraction from the body. Of special interest is Cajetan's interpretation of Aquinas' cognitive psychology found in his *Commentaria in Summa Theologiae*. Following Aquinas, he states that knowledge is gained by way of assimilation through a similitude (*similitudo*) of the thing known, namely an intelligible species that precedes the act of cognition. He rejects a view that had been defended by Agostino Nifo (§3) and other contemporaries, namely that an intelligible species is unnecessary since the act of cognition itself serves as the similitude. Cajetan explains that the light of the agent intellect enables the nature or quiddity that is present in the phantasm to shine forth without there also shining forth the individuality (*singularitas*) of the thing. The intelligible species is thus jointly produced in the possible intellect by a phantasm so-illuminated and by the agent intellect. Considered formally (*formaliter*), the intelligible in act is the intelligible species. But considered objectively (*obiective*), that is, from the point of view of the object known, the intelligible in act is the quiddity that the light of the agent intellect causes to shine forth in the phantasm.

5 Metaphysics

Some of Cajetan's basic metaphysical ideas differ notably from those of Aquinas, his supposed master. In his early commentary on the *Sentences* of Peter Lombard (which was never published), Cajetan closely follows Johannes Capreolus, an earlier Thomist, when he speaks of essence and existence as really distinct. He adopts from Capreolus the terminology of 'the being of essence' (*esse essentiae*) and 'the being of actual existence' (*esse existentiae actualis*). The same dependence on Capreolus is found in Cajetan's *Commentary on 'Being and Essence'*, where he again characterizes the 'being of essence' and the 'being of actual existence' as really distinct. They are 'things' (*res*) from which each creature is composed. Cajetan makes the surprising and questionable claim that this is the position of Aquinas (§9). It should be noted that in fact while Aquinas maintains as a basic doctrine that being (*esse*) is the first actuality of all and the source of all other actualities in an individual thing, Cajetan holds that being is the last or ultimate actuality (*ultima actualitas*) in an individual thing. Indeed, it is added to essence, which appears to possess in some way a being of its own for Cajetan.

See also: Aquinas, T. §§9, 10; Aristotelianism, Renaissance; Thomism

EDWARD P. MAHONEY

List of works

Cajetan [Vio, T. de] (1495) *In De Ente et Essentia divi Thomae Aquinatis*, ed. M.-H. Laurent, Turin: Marietti, 1934; trans. L.H. Kendzierski and F.C. Wade, *Cajetan: Commentary on 'Being and Essence'*, Mediaeval Philosophical Texts in Translation 14, Milwaukee, WI: Marquette University Press, 1964.(*In De Ente* was finished at Padua in 1495 and first published at Venice in 1496. While commenting on Aquinas' text, Cajetan develops metaphysical ideas of his own.)

Cajetan [Vio, T. de] (1496) *Aristotle: On Interpretation. Commentary by Saint Thomas and Cajetan (Peri Hermeneias)*, trans. J.T. Oesterle, Mediaeval Philosophical Texts in Translation 11, Milwaukee, WI: Marquette University Press, 1962.(Cajetan here completes Aquinas' commentary on the *De interpretatione*, which stopped after Book II, Lesson 2. The text of Cajetan's addition to the commentary was first published at Venice in 1496 in Aquinas' *Expositio in Libros Posteriorum et in De Interpretatione*.)

Cajetan [Vio, T. de] (1496) *In libros Posteriorum analyticorum Aristotelis additamenta (Additions on Aristotle's Posterior Analytics)*, Venice: Simon de Luere, 1505.(This is Cajetan's own commentary on the *Posterior Analytics*, completed at Brescia in 1496.)

Cajetan [Vio, T. de] (1497) *Commentaria in Porphyrii Isagogen ad Praedicamenta Aristotelis (Commentary on Porphyry's Isagōgē regarding Aristotle's Praedicamenta)*, ed. I.M. Marega, Rome: Angelicum, 1934. (Completed around 1497; first published in 1505. Contains references to Albert the Great, Aquinas and Averroes.)

Cajetan [Vio, T. de] (1498) *Scripta philosophica: Commentaria in Praedicamenta Aristotelis (Commentary on Aristotle's Praedicamenta)*, ed. M.-H. Laurent, Rome: Angelicum, 1934.(Completed at Pavia in 1498 and first published at Venice in 1505. Cajetan thought that this work should hold first place among Aristotle's logical writings and therefore be studied even before Porphyry's *Isagōgē*.)

Cajetan [Vio, T. de] (1498) *De Nominum Analogia*, in P.H. Hering (ed.) *Scripta philosophica: De nominum*

analogia. De conceptu entis, Rome: Angelicum, 1952; trans. A. Bushinski and H.J. Koren, *The Analogy of Names and the Concept of Being*, Pittsburgh, PA: Duquesne University Press, 1953.(The *De Nominum Analogia* was completed in Pavia in 1498 and published at Venice in 1506. In it Cajetan presents his views on predications about God.)

Cajetan [Vio, T. de] (1499) *Quaestio de subiecto naturalis philosophiae (Question on the Subject of Natural Philosophy)*, in *Opuscula omnia*, Lyons: Juntas, 1587; repr. Hildesheim: Olms, 1995, 207-11.(Finished at Milan in 1499; first published at Venice in 1506. Cajetan defends as the position of Thomas Aquinas and his followers, the Thomists, that mobile being is the subject of natural philosophy.)

Cajetan [Vio, T. de] (1499) *Quaestio de infinitate Dei (Question on the Infinity of God)*, in *Opuscula omnia*, Lyons: Juntas, 1587; repr. Hildesheim: Olms, 1995, 192-206.(Finished at Pavia in 1499 and first published at Venice in 1506. Drawing on Aquinas, Cajetan argues that God's creative power is intensively infinite since it requires no passive potency on which to operate.)

Cajetan [Vio, T. de] (1503) *Oratio habita coram Julio secundo pontifice de immortalitate animorum (Oration before Pope Julius II on the Immortality of Souls)*, in *Opuscula omnia*, Lyons: Juntas, 1587; repr. Hildesheim: Olms, 1995, 186-8; trans. J.K. Sheridan, *On the Immortality of Minds*, in L.A. Kennedy (ed.) *Renaissance Philosophy: New Translations*, The Hague: Mouton, 1973, 46-54.(In this sermon delivered in 1503, Cajetan maintains that immortality can be proven and offers various arguments for it.)

Cajetan [Vio, T. de] (1507) *Commentaria in Primam Partem Summae Theologiae (Commentary on the First Part of Summa theologiae)*, in T. Aquinas, *Opera omnia*, vol. 5, Rome: Typographia Polyglotta, 1889.(This commentary was finished at Rome in 1507 and first published in 1508.)

Cajetan [Vio, T. de] (1509) *Commentaria in De anima Aristotelis (Commentary on Aristotle's On the Soul)*, ed. P.I. Coquelle, Rome: Angelicum, 1938-9, 2 vols; G. Picard and G. Pelland (eds) *Commentaria in libros Aristotelis De anima liber III*, Bruges: Desclée de Brouwer, 1965.(Cajetan's original *In De anima Aristotelis* was completed at Rome in February 1509 and published that same year at Florence; the first two books of the commentary appear in Coquelle, the third in Picard and Pelland.)

Cajetan [Vio, T. de] (1509) *De conceptu entis (On the Concept of Being)*, in P.H. Hering (ed.) *Scripta philosophica: De nominum analogia. De conceptu entis*, Rome: Angelicum, 1952, 97-102; trans. A. Bushinski and H.J. Koren, *The Analogy of Names and the Concept of Being*, Pittsburgh, PA: Duquesne University Press, 1953, 79-83.(Finished at Rome in 1509 and published in 1511.)

Cajetan [Vio, T. de] (1978) *Cajetan Responds: A Reader in Reformation Controversy*, trans. J. Wicks, Washington, DC: The Catholic University of America Press.(Translation of various short works written against the Lutherans: introductory essays on Cajetan's biography and on Cajetan and the Reformation.)

References and further reading

Ashworth, E.J. (1992) 'Analogical Concepts: The Fourteenth-Century Background to Cajetan', *Dialogue* 21: 399-413.(Identifies the sources of three views attacked by Cajetan.)

Bobik, J. (1956) 'The *Materia Signata* of Cajetan', *The New Scholasticism* 30: 127-53.(On Cajetan's discussion of the principle of individuation.)

Gilson, E. (1953) 'Cajetan et l'existence', *Tijdschrift voor Philosophie* 15: 267-87.(Gilson poses the question: what did Cajetan think of the distinction between, or rather the composition of, essence and existence or act of being?)

Gilson, E. (1955) 'Cajetan et l'humanisme théologique', *Archives d'histoire doctrinale et littéraire du moyen âge* 22: 113-36.(On Cajetan's response to Aristotle and Aquinas on the immortality of the soul.)

Gunten, A.F. von (1997) 'Cajetan et Capreolus', in G. Bedouelle, R. Cessario and K. White (eds) *Jean Capreolus et son Temps 1380-1444*, special issue of *Mémoire Dominicaine* 1: 213-38.(Shows both that Cajetan depended on Capreolus for information regarding critics of Aquinas and also that he frequently studied those critics' writings. The author also compares Cajetan's and Capreolus' positions on individuation and other topics.)

Hegyi, J. (1959) *Die Bedeutung des Seins bei den klassischen Kommentatoren des heiligen Thomas von Aquin: Capreolus, Silvester von Ferrara, Cajetan (The Meaning of Being in the Classic Commentators on Aquinas: Capreolus, Silvestri and Cajetan)*, Pullacher Philosophische Forschungen 4, Pullach bei München: Berchmanskolleg, 109-48.(Full examination of Cajetan's teaching about being, beings and analogy.)

Klubertanz, G. (1960) *St Thomas Aquinas on Analogy: A Textual Analysis and Systematic Synthesis*, Chicago, IL: Loyola University Press, 5-17, 118-23, 136-7.(Some discussion of Cajetan in relation to the texts of Aquinas.)

Marc, A. (1933) 'L'idée de l'être chez Saint Thomas et dans la scolastique postérieure' (The Idea of Being in Aquinas and the Later Scholastics), in *Archives de philosophie*, Paris: Beauchesne, vol. 10, cahier 1.(Pages 50-66 of this monograph discuss Cajetan on the idea of being, including abstraction and proportional unity.)

Maurer, A. (1966) 'Cajetan's Notion of Being in his Commentary on the *Sentences*', *Mediaeval Studies* 28: 268-78.(Cajetan's position on essence and existence, as found in the unedited manuscript of his commentary on the *Sentences* of Peter Lombard. Strong dependence on Johannes Capreolus noted.)

Montagnes, B. (1963) *La doctrine de l'analogie de l'être d'après Saint Thomas d'Aquin (The Doctrine of the Analogy of Being according to Aquinas)*, Philosophes médiévaux 6, Louvain: Publications Universitaires and Paris: Béatrice-Nauwelaerts, 126-58.(Discussion of Cajetan on analogy. Montagnes makes it clear that Cajetan is not a trustworthy guide to Aquinas.)

Pinchard, B. and Ricci, S. (eds) (1993) *Rationalisme analogique et humanisme théologique. La culture de Thomas de Vio 'Il Gaetano' (Analogical Rationalism and Theological Humanism. The Culture of Cajetan)*, Naples: Vivarium.(A collection of fifteen essays on Cajetan's philosophy and theology.)

Poppi, A. (1966) *Causalità e infinità nella scuola padovana dal 1480 al 1513 (Causality and Infinity in the Paduan School from 1480 to 1513)*, Padua: Antenore, 170-85.(Discusses Cajetan on God's intensive infinity, and how this relates to God's efficient causality.)

Schwartz, H. (1954) 'Analogy in St Thomas and Cajetan', *The New Scholasticism* 28: 127-44.(Argues that Cajetan misunderstood Aquinas.)

Stöve, E. (1991) 'De Vio, Tommaso (Tommaso, Gaetano, Caetano)', *Dizionario biografico degli italiani* 39: 567-78.(Full biographical details and full bibliography of both primary and secondary sources.)

Tavuzzi, M. (1993) 'Some Renaissance Thomist Divisions of Analogy', *Angelicum* 70: 93-121.(Discussion of Cajetan's contemporaries on analogy.)

Tavuzzi, M. (1994) 'Valentino Da Camerino, O.P. (1438-1515): Teacher and Critic of Cajetan', *Traditio* 49: 287-386.(Discusses Valentino's possible influence on Cajetan; gives useful details of Cajetan's early career.)

Wells, N.J. (1968) 'On Last Looking into Cajetan's Metaphysics: A Rejoinder', *The New Scholasticism* 42: 112-7. (Relates Cajetan's position on the real distinction between essence and existence to that of Albert the Great.)

Calcidius (*c.* 4th century AD)

*The Platonist Calcidius (sometimes less correctly spelt Chalcidius) was the author of a Latin work containing a partial translation of, and partial commentary on, Plato's Timaeus. Although of uncertain date, the doctrinal content of his commentary reflects the thought of the Middle Platonist era (*c.*50 BC-AD 200).*

Calcidius' date and place of operation are uncertain, the only clue residing in his dedication of his work to one Osius, who has been taken, following identifications in a number of manuscripts, to be the Bishop of Corduba (AD 256-357) and spiritual advisor of Constantine. This identification has been challenged, mainly on the grounds that Isidore of Seville makes no mention of Calcidius in his enumeration of all the Spanish writers that he knows, but this is not conclusive. It is also true that his language is more consistent with a fifth-century than a fourth-century date, but even this is not conclusive when weighed against considerations of content. The main issue is whether Calcidius is to be regarded as exclusively dependent on Middle Platonist sources or as influenced also by the *Timaeus* commentary of Porphyry. In fact, there are sufficient indications that, unless Calcidius was being very selective (which does not seem to be his method), he knows nothing of Porphyry. His work seems primarily to be a translation of a number of Middle Platonist or Peripatetic Greek sources (or even of a single such source). Identifiable sources include the second century Peripatetic Adrastus of Aphrodisias, who wrote a commentary on the *Timaeus* concentrating on the mathematical and 'scientific' aspects, and the Neo-Pythagorean Numenius.

What we have is a translation of the dialogue as far as 53c, followed by a detailed commentary, beginning with *Timaeus* 31c (Calcidius ignores the introductory portion) and breaking off also at 53c (although a translation of, and commentary on, the whole seems to have been envisaged). It is hardly suitable in the case of Calcidius to speak of a distinct philosophical stance, since it is doubtful that he is putting much of himself into the commentary, but he reflects various interesting Middle Platonic positions. In chapter 176, for example, he presents a sequence of supreme god, intellect and world-soul which accords well with the systems of Numenius and Alcinous. The supreme god is identified with the Good of Plato's *Republic*, but there is no need to take it as a Neoplatonic One. At the other end of the scale, his doctrine that matter is neither corporeal nor incorporeal, but potentially both (chaps 319-20), agrees with that of Alcinous, Apuleius and Hippolytus.

See also: Neoplatonism; Neo-Pythagoreanism; Peripatetics; Platonism, Early and Middle

JOHN DILLON

List of works

Calcidius (*c.* 4th century AD) *Timaeus a Calcidio translatus*, ed. J.H. Waszink, London: Warburg Institute, 1962; trans. J.C.M. van Winden, *Calcidius on Matter: His Doctrine and Sources*, Leiden: Brill, 1965, chaps 268-354; trans. J. den Boeft, *Calcidius on Fate: His Doctrine and Sources*, Leiden: Brill, 1970, chaps 142-90; trans. J. den Boeft, *Calcidius on Daemons*, Leiden: Brill, 1977, chaps 127-36.(Waszink is an excellent edition, with introduction, but the text is in Latin. Van Winden and den Boeft provide translations of selected chapters, with commentary.)

References and further reading

Dillon, J. (1977) *The Middle Platonists*, London: Duckworth, 401-8.(Introductory account of Calcidius.)

Gersh, S. (1986) *Middle Platonism and Neoplatonism, The Latin Tradition*, Notre Dame, IN: University of Notre Dame Press, vol. 2, 421-92.

Waszink, J.H. (1964) *Studien zum Timaioskommentar des Calcidius (Studies on Calcidius' Commentary on Timaeus)*, vol. 1, Leiden: Brill.(Essays on various topics arising in the first half of the commentary; the second volume was never published.)

Callicles (late 5th century BC)

*Callicles, although known only as a character in Plato's Gorgias (the dramatic date of which is somewhere between 430 and 405 BC), was probably an actual historical person. Employing a distinction between nature (*physis*) and convention (*nomos*), he argues eloquently that the naturally superior should seize both political power and a greater share of material goods: it is only a convention of the weak majority which labels such behaviour unjust. In private life the superior should indulge their desires freely: excess and licence are true virtue and happiness.*

Callicles is a wealthy and aristocratic young Athenian with ambitions to be a democratic leader, and the detail that Plato bestows on his portrait strongly suggests that he did in fact exist. Plato may have deliberately selected someone whose youthful promise was known to have come to nothing, as a warning against moral and intellectual indiscipline (there is a hint that Callicles was later arrested). Despite his proclaimed contempt for Sophists, his views appear indebted to those of Gorgias and may consequently be distorted by Plato's anti-Sophistic bias (see Plato; Sophists).

Callicles attacks Socrates' claim that doing wrong is more shameful than being wronged (see Socrates §4). This, he argues, is simply the conventional notion of justice, which is nothing but a craven invention of the weak majority to restrain the stronger few: the weak proclaim that equality of both power and wealth is 'just' because equality is the most they can expect. In nature, however, might is right, as is clear from both the animal kingdom and interstate relations. This is not simply brute fact, but a prescription: it is true justice, the 'law of nature'. It is thus naturally more shameful to suffer wrong than to do it, as the former shows that one is too weak to protect oneself and one's own. Natural justice decrees that those who are superior in courage, manliness and practical competence should appropriate both political control and the greater share of goods, and it is similarly a natural imperative that the superman should freely indulge his physical desires; the convention that praises self-restraint again results from a conspiracy of the weak majority who lack the natural force to satisfy their own urges. Breaking free from convention is the mark of nature's lions, and such 'real men' are in sharp contrast to the feeble, impractical philosopher.

By appealing to an alternative 'natural justice', Callicles is more radical than contemporary supporters of *physis* over *nomos*, such as Antiphon and Thrasymachus, who employ 'justice' solely in its conventional sense. He implies an antithesis between reason and desire, the latter being the true self. Any restraint on the desires, whether from society or from one's reason, is thus an unacceptable infringement of personal liberty. He is adamant that it is he, not Socrates, who truly understands human nature.

Callicles' defence of natural force is open to challenge, particularly for its appeal to the animal world. His notion of 'natural justice', being based on nothing more than description of the existing order, would be accused by some philosophers of committing the 'naturalistic fallacy' (see Naturalism in ethics §3). Furthermore, it is debatable whether his depiction of animal predators is correct. Do they really pursue individual interests to the total exclusion of group conventions? And if such apparent 'conventions' are simply a manifestation of their 'nature', does this not dissolve the antithesis on which Callicles relies?

It is also unclear whether human aggression is as natural as Callicles claims (he implies that everyone would imitate the superman if they could). Some would argue that, although we all possess egoistic drives, they only develop into ruthlessly competitive behaviour when encouraged by society. The only 'natural' aspect of the aggressive superman might be that it is natural to absorb social conventions, for good or ill. This, however, would only demonstrate again that Callicles' required antithesis between nature and culture is false.

Nor does he appear to have thought through what a society run by supermen would be like. How would they behave towards each other? Or would a single superman emerge from each community? And presumably the laws made by such supermen would have to command Callicles' respect. In disparaging man-made *nomoi*, he simply assumes a democratic framework. (The relation between Callicles' contemptuous dismissal of the majority and his democratic aspirations is an intriguing question in itself.)

Despite these and other inconsistencies and philosophic shortcomings, Callicles is an important figure. There is

evidence that he was a strong influence on Nietzsche and it is not difficult to see why. He challenges us to reconsider the origins and nature of social 'morality' and makes a vigorous bid for the possibility of self-directed virtues. He raises pertinent questions about the nature of 'manliness' and the relative merits of the active and the contemplative life. Above all, he forces us to reflect on the issues of personal identity and freedom, on which all his arguments depend.

See also: Physis and nomos

ANGELA HOBBS

References and further reading

Berman, S. (1991) 'Socrates and Callicles on Pleasure', *Phronesis* 36: 117-40.(A lucid and lively discussion.)

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, vol. 3, Cambridge: Cambridge University Press; part of vol. 3 repr. as *The Sophists*, Cambridge: Cambridge University Press, 1971.(Helpful discussions of Callicles, especially pages 101-7.)

Kerferd, G.B. (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press.(The best introductory handbook to the Sophists in general.)

Plato (*c*.395-387 BC) *Gorgias*, ed. E.R. Dodds, Oxford: Clarendon Press, 1959; trans. T. Irwin, Oxford: Clarendon Press, 1979.(The earlier edition is scholarly and sensitive; it includes Greek text, with introduction and commentary and contains a useful appendix on Callicles and Nietzsche; the translation, with introduction and notes, contains a crisp and rigorous appraisal of Callicles' philosophical position.)

Calvin, John (1509-64)

John Calvin, French Protestant reformer and theologian, was a minister among Reformed Christians in Geneva and Strasbourg. His Institutes of the Christian Religion (first edition 1536) - which follows the broad outline of the Apostles' Creed and is shaped by biblical and patristic thought - is the cornerstone of Reformed theology.

Calvin's religious epistemology links self-knowledge and knowledge of God. He identifies in humans an innate awareness of God, which is supported by the general revelation of God in creation and providence. Because sin has corrupted this innate awareness, Scripture - confirmed by the Holy Spirit - is needed for genuine knowledge of God. Scripture teaches that God created the world out of nothing and sustains every part of it. Humanity, which was created good and with free will, has defaced itself and lost significant freedom due to its fall into sin. Calvin sees Christ the mediator as the fulfilment of the Old Testament offices of prophet, priest and king.

Calvin insists that God justifies sinners on the basis of grace and not works, forgiving their sins and imputing Christ's righteousness to them. Such justification, received by faith, glorifies God and relieves believers' anxiety about their status before God. On the basis of his will alone, God predestines some individuals to eternal life and others to eternal damnation.

Calvin dignifies even ordinary occupations by seeing them as service to God. He recognizes the distinction between civil government and the Church, although he says that government should protect true worship of God and Christians should obey and support their government. Calvin's thought was dominant in non-Lutheran Protestant churches until the eighteenth century and has enjoyed a resurgence since the mid nineteenth century.

1 Life

John Calvin was born in Noyon, France. He spent five years studying Latin grammar and the arts at the University of Paris, where he became acquainted with both scholastic and humanist thought. After his father withdrew him from preparation for an ecclesiastical career, Calvin studied civil law at the University of Orléans and the University of Bourges. He received the licentiate in law in 1532 and, in the same year, published a commentary on the Stoic philosopher Seneca's *De clementia*, thereby establishing his credentials as a humanist familiar with ancient thought.

By 1534 Calvin publicly advocated ecclesiastical reform in France, which meant that he had broken with the papacy; he spent the rest of his life as a refugee. The first edition of the *Institutes of the Christian Religion*, written while he was in exile in Basle, was published in 1536. Several major revisions and expansions of the work appeared in both Latin and French, including the definitive Latin edition of 1559. Addressed to King Francis I of France, the preface to the *Institutes* argues that, since the reform movement is the legitimate heir to the Christianity of the Scriptures and the early Church and is not seditious, it should be protected from persecution.

In 1536, during a stop in Geneva on his way to Strasbourg, Calvin was persuaded to stay and assist the ecclesiastical reformers. Tensions between the reformers and the city council led to Calvin's expulsion from Geneva in 1538. After ministering in Strasbourg, he returned to Geneva in 1541 and remained there until his death. His pastoral work in Geneva and Strasbourg focused on ministry to French Protestant refugees.

Calvin has been criticized for his involvement in the death of the notorious heretic Michael Servetus in 1555. Having already escaped execution elsewhere for his heterodox Trinitarian views, Servetus appeared in Geneva, was arrested at Calvin's insistence, and was condemned to death by the city council. He was burned at the stake despite Calvin's request that he be beheaded instead.

In addition to the *Institutes*, Calvin wrote many polemical works as well as commentaries on most of the books of the Bible, and engaged in preaching and pastoral work. He founded the Genevan Academy in 1559.

2 Religious epistemology

'Nearly the whole of sacred doctrine consists in these two parts: knowledge of God and of ourselves.' In this opening line of the 1536 edition of the *Institutes* (and every edition thereafter), Calvin announces his fundamental project of describing how we acquire this twofold knowledge and summarizing its content. Self-knowledge and

knowledge of God are interrelated: self-knowledge leads us to be displeased with ourselves, thereby arousing us to seek God, the source of all good; and knowledge of God is required for a clear awareness of ourselves - especially of our folly and corruption.

Knowledge of God is not ‘empty’ or ‘cold’ speculation that ‘merely flits in the brain’, but rather knowledge that ‘takes root in the heart’ and includes the honouring of God (*Inst.*, 1.5.9; 1.12.1). Such reverence for and love of God, which Calvin calls ‘piety’, results from recognizing that we owe everything to God, are nourished by his fatherly care and should seek nothing beyond him.

Calvin insists that all humans have, by natural instinct, an awareness of divinity (*sensus divinitatis*). The universal need for religion - even idolatrous religion - shows that God has implanted in all people a seed of religion, a natural awareness ‘that there is a God and that he is their Maker’ (*Inst.*, 1.3.1). Religion was not ‘invented by the subtlety and craft of a few to hold the simple folk in thrall’ (*Inst.*, 1.3.2).

Lest anyone be excluded from access to happiness by lack of knowledge of God, says Calvin, God has also ‘engraved unmistakable marks of his glory’ upon the universe (*Inst.*, 1.5.1). The ‘skillful ordering of the universe is for us a sort of mirror’ reflecting the invisible God (*Inst.*, 1.5.1). And nursing infants ‘have tongues so eloquent to preach [God’s] glory that there is no need at all of other orators’ (*Inst.*, 1.5.3). Although we have within ourselves ‘a workshop graced with God’s unnumbered works’ that should lead us to break forth in praise of God, instead we are ‘puffed up and swollen with all the more pride’ because we see nature rather than God as the source of these good gifts (*Inst.*, 1.5.4).

God’s glory is also displayed outside the ordinary course of nature, in his declaring clemency to the godly and showing severity to the wicked, protecting the innocent, and caring for the poor in their desperate straits. Yet most people, ‘immersed in their own errors, are struck blind in such a dazzling theater’, and do not recognize the glory of God (*Inst.*, 1.5.8).

Although God represents himself in the mirror of his works with great clarity, we grow dull towards such clear testimonies and presumptuously fashion a deity of our own imagining. Our minds are like springs pouring forth gods we have invented for ourselves. Even philosophers adore a shameful diversity of gods (*Inst.*, 1.5.11-12).

Since there is no pure religion founded upon common understanding alone, we need the witness of God himself, ‘illumined by the inner revelation of God through faith’ (*Inst.*, 1.5.13, 14). ‘Just as old or bleary-eyed men and those with weak vision’ can scarcely see until they put on spectacles, ‘so Scripture, gathering up the otherwise confused knowledge of God in our minds, having dispersed our dullness, clearly shows us the true God’ (*Inst.*, 1.6.1). Calvin recognizes, however, that Scripture must be confirmed as the living word of God. Neither the consent of the Church nor disputation in defence of Scripture will imprint on people’s hearts the certainty and assurance that piety requires. The Spirit who spoke through the prophets must penetrate into our hearts to persuade us that they faithfully proclaimed what God spoke to them; the inner testimony of the Holy Spirit ‘is more excellent than all reason’ (*Inst.*, 1.7.4). In response to religious fanatics or enthusiasts who abandon Scripture and claim a new revelation from the Spirit, Calvin argues that God has forged a mutual bond between Word and Spirit, with the Spirit, as the author of Scripture, attesting the Word of God in Scripture and the Word leading us to recognize the true Spirit of God.

For Calvin, faith is ‘a firm and certain knowledge of God’s benevolence toward us, founded upon the truth of the freely given promise in Christ, both revealed to our minds and sealed upon our hearts through the Holy Spirit’ (*Inst.*, 3.2.7). Knowledge of the unseen God does not involve comprehension, as with objects of sense perception, but rather consists in assurance of God’s goodness towards us, which depends on our ‘truly feeling its sweetness and experiencing it in ourselves’ (*Inst.*, 3.2.15). Then, despite temptations and doubts, believers will ‘either rise up out of the very gulf of temptations, or stand fast upon their watch’, confident of God’s mercy to them (*Inst.*, 3.2.21, 37).

3 God, creation and providence

Calvin emphasizes God’s sovereignty and glory. He introduces the notion of accommodation in order to account for Scripture’s speaking of God’s mouth, ears, eyes, hands and feet (*Inst.*, 1.13.1); of God’s repenting of having created the world or of having made Saul king of Israel (*Inst.*, 1.17.12); and even of God’s wrath against sin (*Inst.*,

1.17.13; 2.16.2, 4). Such images 'do not so much express clearly what God is like as accommodate the knowledge of him to our slight capacity' (*Inst.*, 1.13.1).

Calvin affirms that God created the world by his own free will and follows Augustine in rejecting speculation about why God created the world when he did. He shows little interest in the duration or sequence of the events of creation, focusing instead on the spiritual value of observing that God created, out of nothing, an abundance and variety of creatures, each with its own nature and assigned functions, that God formed humanity as 'the most excellent example of his works', and that God has provided for the preservation of the entire creation (*Inst.*, 1.14.20). The goodness of creation should lead us to take 'pious delight in the works of God open and manifest in this most beautiful theater' (*Inst.*, 1.14.20).

Proper self-knowledge, says Calvin, is twofold: knowing what we were like when first created and knowing our condition after the Fall of Adam. As created, humans had rational minds to know their duty, and wills to direct the appetites under the guidance of reason. Only if humanity had not fallen would those philosophers be correct who hold that moral accountability depends on having free choice between good and evil. By sinning, Adam lost God's good gifts for himself and for his descendants, with the result that fallen human beings lack the ability to carry out their duty. Still, says Calvin, in being subject to the depravity and corruption of our nature, or original sin, we are not held liable for the transgression of another; the contagion imparted by Adam resides in us (see Sin §2).

Calvin insists that we see God's power as much in the continuing state of the universe as in its creation. God not only drives the 'celestial frame', but also 'sustains, nourishes, and cares for everything he has made, even to the least sparrow' (*Inst.*, 1.16.1). Since all events proceed from God's plan, 'nothing takes place by chance' - not even if a branch falls and kills a passing traveller (*Inst.*, 1.16.4, 6). Calvin argues that, unlike the Stoic dogma of fate, the doctrine of providence sees God as the ruler of all things who 'in accordance with his wisdom' has from eternity decreed what he was going to do and now carries it out (*Inst.*, 1.16.8).

If God wills all things, does he also will what is evil? Calvin argues that it is absurd to substitute 'bare permission' in the place of God's providence 'as if God sat in a watchtower awaiting chance events, and his judgments thus depended upon human will' (*Inst.*, 1.16.8; 1.18.1). Humans can accomplish nothing except by God's secret command. In response to the charge that God would then have two contrary wills - a revealed will that forbids sin and a secret will that decrees evil actions - Calvin says that God's will is one and simple in him but appears manifold to us because we cannot grasp how God both wills and does not will something to take place. Many Reformed theologians, while insisting with Calvin that all things are governed by God's providence, have departed from him by distinguishing between the active and the permissive will of God.

Calvin's affirmation of God's goodness in relation to evil events hinges on Augustine's insight that there is a great difference between what it is fitting for a human being to will and what is fitting for God. In opposition to those who 'wish nothing to be lawful for God beyond what their own reason prescribes for themselves', Calvin says that nothing but right flows from providence, although the reasons have been hidden from us (*Inst.*, 1.17.2). Conceding that his doctrine of providence might seem harsh, Calvin says we should not refuse to accept a teaching attested by clear Scriptural proofs just because it exceeds our understanding. In addition, believers can have gratitude for favourable outcomes, patience in adversity, and freedom from worry about the future because they know that all things are ruled by their heavenly father.

4 Christ and salvation

Even sinless humanity, says Calvin, would have needed a mediator in order to reach God. Yet the need for this mediator to be truly human as well as truly divine was due to our sin. The purpose of the incarnation was for Christ the mediator to become the redeemer, restoring the fallen world and assisting lost humanity in its distress.

Calvin's theology of the atonement emphasizes both Christ's victory over the powers of evil and his satisfaction of God's righteous judgment. Death held humanity captive, says Calvin, but Christ, in our stead, submitted to the power of death, grappled hand-to-hand with the armies of hell, and arose from the dead victorious over them. Alternatively, Calvin says that humanity's reconciliation with God requires the countering of disobedience with obedience, which Christ does in our place, paying what we could not pay and thereby acquiring righteousness for us (see Atonement §4).

Like Martin Luther, Calvin says that the doctrine of justification is the main hinge on which true religion turns. To be *justified* in God's sight means that one is counted righteous by God; to be justified *by faith* means that one who is in fact a sinner 'grasps the righteousness of Christ through faith, and clothed in it, appears in God's sight' not as sinful but as righteous (*Inst.*, 3.11.2). God justifies sinners by remitting their sins and imputing Christ's righteousness to them. God first freely embraces the sinner who is utterly void of good works and then conveys a sense of divine goodness that leads the sinner to despair of works and to rely wholly on God's mercy. So the faith through which sinners are justified is not itself a work that qualifies them for salvation, but rather God's gracious gift through which they receive God's mercy (see Justification, religious §4).

In his doctrine of predestination, Calvin attempts to avoid the twin dangers of either probing into matters that God wills to remain hidden or neglecting what Scripture says. Scripture presents two types of election: the general election of a nation or people, such as ancient Israel or the Church; and the election or predestination of individuals. Although God does not grant regeneration and perseverance to every member of an elect people, he works in the heart of each elect individual, granting perseverance so that this election is assured. Calvin observes that election involves the adoption 'in Christ into the eternal inheritance' of those who in themselves are unworthy of this adoption (*Inst.*, 3.22.1) (see Predestination; Providence §§3-4; Reprobation).

Predestination, then, is God's eternal decree regarding each individual, foreordaining eternal life for some and eternal damnation for others. This foreordination is not based on God's having foreseen the holiness of the elect or the sinfulness of the reprobate, but only in God's will, which to us is inscrutable. Mirroring election, reprobation is based in God's will, not in foreknowledge of the sinfulness of the reprobate. Calvin's doctrine of reprobation, which is rooted in his rejection of the idea of God's permissive will (see §3 above), differs from that of many Reformed theologians, who hold that God wills to elect some, thereby leaving the others in the ruin they incur by their sins.

5 Social and political thought

Calvin recognizes the dignity of each occupation - not just that of the clergy - as service to God. God has given each individual a calling or vocation 'as a sort of sentry post', to prevent heedless wandering about throughout life. Each person will fulfil his or her calling willingly, bearing its weariness and anxieties, if it is seen as a duty to God. 'No task', says Calvin, 'will be so sordid and base, provided you obey your calling in it, that it will not shine and be reckoned very precious in God's sight' (*Inst.*, 3.10.6).

Calvin argues that, since 'Christ's spiritual Kingdom and the civil jurisdiction are things completely distinct', legal or social distinctions of rank or privilege should not apply in Christ's realm (*Inst.*, 4.20.1). Nevertheless, civil government and the Church are to be mutually supportive. Accordingly, civil government should 'protect the outward worship of God' and even defend sound doctrine (*Inst.*, 4.20.2-3). Civil magistrates, who are given authority by God and serve as God's deputies, should embody divine providence, goodness, benevolence and justice. Governments may levy taxes, but should regard the revenues as treasuries of the people that must not be squandered.

Christians, for their part, must regard all magistrates as God's representatives, not as a necessary evil. Subjects should obey their government, pay taxes and accept service in public office or for military defence. Even wicked and unjust magistrates have been ordained by God and should be held by their subjects in the same esteem as they would hold the best of monarchs. Still, Calvin recognizes means for holding unjust rulers in check. Sometimes God raises up avengers to punish the wicked government, as when a just nation defeats a nation ruled by an evil leader; sometimes God appoints lesser magistrates to restrain unjust kings. Even common subjects must not allow their obedience to rulers to lead them to follow commands contrary to God's revealed will (*Inst.*, 4.20.32; *Comm.* on Daniel 6: 22) (see Religion and political philosophy).

6 Influence

Calvin's thought has had a wide influence in the Church, society and the academic world. His theology sparked movements for ecclesiastical reform that led to Reformed churches in continental Europe, Presbyterian churches in the United Kingdom, and Puritan groups that sought refuge in the American colonies. Owing to the effects of the Enlightenment and liberal Protestant thought, Calvinist theology declined in influence during the eighteenth and

nineteenth centuries. Since the middle of the nineteenth century, however, Calvin's thought has enjoyed a resurgence, having provided impetus to a Calvinist revival in the Netherlands led by Abraham Kuyper and to the neo-orthodoxy that flourished in the middle of the twentieth century under the leadership of Karl Barth and Emil Brunner. Calvin's religious epistemology has also influenced the development of 'Reformed epistemology' by Alvin Plantinga and Nicholas Wolterstorff (see Religion, history of philosophy of §8; Religion and epistemology).

RONALD J. FEENSTRA

List of works

Calvin, J. (1528-64) *Ioannis Calvini opera quae supersunt omnia*, ed. J.W. Baum, A.E. Cunitz and E. Reuss, Brunswick and Berlin: C.A. Schwetschke, 1863-1900, 59 vols.(The standard scholarly edition of Calvin's works, published as part of the Corpus Reformatorum.)

Calvin, J. (1528-64) *Selected Works of John Calvin: Tracts and Letters*, ed. H. Beveridge and J. Bonnet, 1844-58; repr. Grand Rapids, MI: Baker, 1983, 7 vols.(Reprint of a nineteenth-century translation of Calvin's tracts and letters.)

Calvin, J. (1536) *Institutes of the Christian Religion*, trans. F.L. Battles, Grand Rapids, MI: Eerdmans, rev. edn, 1986.(A valuable translation of the first edition of the *Institutes*.)

Calvin, J. (1540-64) *Calvin's Commentaries*, trans. Calvin Translation Society, 1843-55; reprint, Grand Rapids, MI: Baker, 1979, 22 vols.(Reprint of a good nineteenth-century translation of Calvin's commentaries, including the best English translation of his New Testament commentaries.)

Calvin, J. (1551-64) *Calvin's Old Testament Commentaries*, Rutherford House translation, ed. D.F. Wright, Grand Rapids, MI: Eerdmans and Carlisle: Paternoster, 1993-.(The best English translation of Calvin's Old Testament commentaries.)

Calvin, J. (1552) *Concerning the Eternal Predestination of God*, trans. J.K.S. Reid, Cambridge: James Clarke, 1961.(An important treatise by Calvin on predestination.)

Calvin, J. (1559) *Institutes of the Christian Religion*, ed. J.T. McNeill, trans. F.L. Battles, The Library of Christian Classics, Philadelphia, PA: Westminster, 1960, 2 vols.(The standard translation of the definitive 1559 edition.)

References and further reading

Dowey, E.A., Jr (1952) *The Knowledge of God in Calvin's Theology*, New York: Columbia University Press; expanded edn, Grand Rapids, MI: Eerdmans, 1994.(A classic study of Calvin on the knowledge of God.)

Duke, A., Lewis, G. and Pettegree, A. (eds) (1992) *Calvinism in Europe, 1540-1610: A Collection of Documents*, Manchester and New York: Manchester University Press. (A selection and translation of important documents related to the development and spread of Calvinism.)

Ganoczy, A. (1987) *The Young Calvin*, trans. D. Foxgrover and W. Provo, Philadelphia, PA: Westminster.(An important study of Calvin's background and early theological development.)

Graham, W.F. (1971) *The Constructive Revolutionary: John Calvin and His Socio-Economic Impact*, Richmond, VA: John Knox.(A standard work on Calvin's social and political thought.)

Greef, W. de (1994) *The Writings of John Calvin: An Introductory Guide*, trans. L.D. Bierma, Grand Rapids, MI: Baker.(A solid introductory survey of Calvin's life and writings, with good primary and secondary bibliographies.)

McKim, D.K. (ed.) (1984) *Readings in Calvin's Theology*, Grand Rapids, MI: Baker. (Includes several important essays on Calvin's thought.)

McNeill, J.T. (1954) *The History and Character of Calvinism*, Oxford: Oxford University Press.(A classic study of Calvin, his background and his influence.)

Muller, R.A. (1986) *Christ and the Decree: Christology and Predestination in Reformed Theology from Calvin to Perkins*, Durham, NC: Labyrinth Press.(An important study of the doctrine of predestination in sixteenth-century Reformed theology.)

Niesel, W. (1956) *The Theology of Calvin*, trans. H. Knight, London: Lutterworth.(A classic study of Calvin's theological system.)

Parker, T.H.L. (1975) *John Calvin: A Biography*, Philadelphia, PA: Westminster.(A very good biography of Calvin.)

Pettegree, A, Duke, A. and Lewis, G. (eds) (1994) *Calvinism in Europe, 1540-1620*, Cambridge: Cambridge University Press.(A collection of original essays on the early development and spread of Calvinism.)

Plantinga, A. and Wolterstorff, N. (eds) (1983) *Faith and Rationality: Reason and Belief in God*, Notre Dame, IN and London: University of Notre Dame Press. (Includes seminal essays on Reformed epistemology.)

Prestwich, M. (ed.) (1985) *International Calvinism, 1541-1715*, Oxford: Clarendon Press.(Valuable essays on the development of Calvinism.)
Schreiner, S.E. (1991) *The Theater of His Glory: Nature and the Natural Order in the Thought of John Calvin*, Durham, NC: Labyrinth Press.(An important study of nature, providence and natural law in Calvin.)
Steinmetz, D. (1995) *Calvin in Context*, New York and Oxford: Oxford University Press.(A study of Calvin's thought in its historical context and in relation to his contemporaries.)
Wendel, F. (1963) *Calvin: Origins and Development of His Religious Thought*, trans. P. Mairet, New York: Harper & Row.(An excellent introduction to Calvin's life and to the *Institutes*.)

Cambridge Platonism

Cambridge Platonism was an intellectual movement broadly inspired by the Platonic tradition, centred in Cambridge from the 1630s to the 1680s. Its hallmark was a devotion to reason in metaphysics, religion and ethics. The Cambridge Platonists made reason rather than tradition and inspiration their ultimate criterion of knowledge. Their central aim was to reconcile the realms of reason and faith, the new natural philosophy and Christian revelation. Although loyal to the methods and naturalism of the new sciences, they opposed its mechanical model of explanation because it seemed to leave no room for spirit, God and life.

In epistemology the Cambridge Platonists were critics of empiricism and stressed the role of reason in knowledge; they also criticized conventionalism and held that there are essential or natural distinctions between things. In metaphysics they attempted to establish the existence of spirit, God and life in a manner consistent with the naturalism and method of the new sciences. And in ethics the Cambridge Platonists defended moral realism and freedom of the will against the voluntarism and determinism of Hobbes and Calvin. Cambridge Platonism was profoundly influential in the seventeenth and eighteenth centuries. It was the inspiration behind latitudinarianism and ethical rationalism, and many of its ideas were developed by Samuel Clarke, Isaac Newton and the Third Earl of Shaftesbury.

1 Origins and context

The term ‘Cambridge Platonism’ has been used to refer to a group of thinkers active in Cambridge from the late 1630s to the 1680s, who were in one form or another inspired by the Platonic tradition of philosophy. Most of them were either fellows or students of two colleges, Emmanuel and Christ’s. Their inner circle consisted of Henry More (1614-87), Ralph Cudworth (1617-88), John Smith (1618-52) and Benjamin Whichcote (1609-83), who was its leading figure. Their outer circle of associates within Cambridge comprised John Sherman (d. 1666), John Worthington (1618-80), Peter Sterry (1613-72), George Rust (1626-70) and Nathaniel Culverwell (1618-51). There were also thinkers outside Cambridge who were closely connected with, and often shared the views of, the Cambridge Platonists: John Norris (1657-1711), Joseph Glanvill (1636-80) and Richard Burthogge (*c*.1638-*c*.1704). Among the disciples of the Cambridge school were some important latitudinarian divines: Simon Patrick (1626-1707), Edward Fowler (1632-1714), John Tillotson (1630-94), John Moore (1646-1714), Gilbert Burnet (1643-1715), Edward Stillingfleet (1635-99) and Thomas Tenison (1636-1715).

Cambridge Platonism represents the most concerted and systematic attempt in seventeenth-century English thought to reconcile the claims of reason and faith, philosophy and religion. Cudworth and More were the first thinkers in the English Protestant tradition to develop a sophisticated natural theology. Some of their most substantive works - More’s *Antidote to Atheism* (1653) and *Immortality of the Soul* (1659), Cudworth’s *A Treatise concerning Eternal and Immutable Morality* (1731) and *True Intellectual System of the Universe* (1678) - were devoted to precise and detailed demonstrations of the existence of God, providence and immortality. True to the tradition of Florentine Platonism (see Platonism, Renaissance §4), and contrary to Bacon and orthodox Protestantism, the Cambridge men sometimes completely identified the realms of reason and faith, as if nature and revelation were only different hypostases or degrees of manifestation of the divine goodness. They held that the truths of revelation are only a more spiritual form of reason, and the truths of reason only a more natural form of revelation. But they did not always adhere to this equation and sometimes abandoned it in the face of the more recalcitrant Christian mysteries, such as the Trinity and Incarnation, which were conceded to be above, although never contrary to, reason.

Cambridge Platonism began as a reaction to the more severe tenets of Calvinist theology, especially its doctrine of predestination. In their early years Whichcote, Cudworth and More were repelled by Calvin’s *deus absconditus*, which by some ‘dark and inscrutable decrees’ predestined people, regardless of their merits, to either salvation or damnation (see Calvin, J. §4). This undermined not only certainty regarding salvation, but also moral responsibility. The great attraction of Platonism was that it made the ways of God accountable to man, so that he could be sure of his eternal election. Platonism meant that there are eternal forms of good and evil, which God never violates, and which are discernible by reason.

In the seventeenth-century dispute concerning ‘the rule of faith’ or the ultimate criterion of religious knowledge, the Cambridge Platonists stood firmly in the camp of reason. They opposed enthusiasm, which appealed to

inspiration, no less than Roman Catholic ‘dogmatism’, which referred to apostolic tradition. It is important to recognize, however, that the Cambridge men defended reason only because they had a Platonic conception of it. They saw reason as a mystical faculty, a power of vision, which is guided by divine grace. They opposed the formal concept of reason prevalent in the nominalist tradition, according to which it is merely a power of inference. They also rejected Socinianism because of its critical attitude toward traditional Christian doctrine (see Socinianism).

During the seventeenth-century controversy regarding ecclesiastical polity, the Cambridge Platonists attempted to walk a middle path between Laudianism and Puritanism, High-Church Anglicanism and Presbyterianism. They denied a common premise to these extreme positions: that church government and discipline is *jure divino*, prescribed by divine decree and determinable by Scripture. They argued that all such matters are indifferent, and should be determined by the civil sovereign alone. Anticipating the latitudinarians, they stressed the need for a comprehensive church whose creed was broad enough to accommodate all Christians (see Latitudinarianism).

2 Epistemology

The epistemology of the Cambridge school was chiefly a defence of essentialism and rationalism against the conventionalism and empiricism of the nominalist tradition (see Conventionalism; Empiricism; Essentialism; Rationalism). In his *A Treatise concerning Eternal and Immutable Morality*, the main epistemological work of the Cambridge school, Cudworth argued that things are what they are in virtue of the necessity of their nature or essence alone, and not in virtue of some divine command or power. God cannot make two plus two equal five, or the nature of a triangle such that its three angles are not equal to two right angles. It is an eternal and necessary truth that, given the concept of two and the concept of addition, two plus two equals four, and that, given the concept of a triangle, its three angles are equal to 180 degrees. Although the will of God is indeed the supreme efficient cause of all things, it is not their formal cause.

The Cambridge Platonists criticized empiricism as much as conventionalism. Anticipating Leibniz and Kant, Cudworth and More pointed out how the conditions of knowledge cannot be satisfied by the senses: knowledge requires judgment, but the senses do not judge. Knowledge also involves universality and necessity, whereas the senses show us only particularity and contingency. Furthermore, the senses are entirely passive, though knowledge demands activity, making sensations conform to innate laws. Finally, the object of the senses is given, distinct from the knower, but knowledge requires the identity of the knower and known, reason’s contemplation of its own creations. Contrary to empiricism, Cudworth advocated a rationalist paradigm of knowledge, according to which true knowledge derives from deduction, consisting in ‘a descending comprehension of a thing from the universal ideas of the mind, and not an ascending perception of them from individuals by sense’.

The Cambridge Platonists became notorious for their doctrine of innate ideas, which was the apparent target of Locke’s famous polemic (see Locke, J. §2). But Cudworth, More and Culverwell explicitly rejected the view that the mind has pre-formed ideas prior to experience. They understood innateness in terms of the inherent *activities* and *faculties* of the mind, and stressed that these had to be stimulated by experience.

Cudworth and More had a clear response to Descartes’ radical scepticism, which posed a serious challenge to the rationalism of the Cambridge school. They remarked that Descartes could not escape his radical doubt by his proof of the existence of God, since such a demonstration presupposes the truth of our faculties. They also maintained that there is a false premise behind Cartesian doubt: that knowledge requires the correspondence between concepts and an external reality. Since reason creates its object in the act of knowing it, we need not seek truth outside ourselves. Finally, they contended that we have no general reason to doubt our faculties if they normally supply clear and distinct ideas on specific occasions (see Descartes, R. §4).

3 Metaphysics

The Cambridge Platonists had an ambivalent attitude towards the new natural philosophy which blossomed during their era. They admired its methods of observation and experiment; they shared its distaste for the verbiage of the old scholasticism; and they accepted its naturalism, its belief that everything conforms to law. But they feared the consequences of the new mechanical model of explanation of Descartes and Boyle. If this paradigm were generalized, then it would leave no room for spirit, life or God. The aim of the Cambridge Platonists was therefore

to secure a place for spirit, life and God while upholding the naturalism and methods of the new philosophy.

Cudworth and More argued that mechanism alone cannot account for such phenomena as gravity and cohesion in the material world, or generation and growth in the organic world. To explain such phenomena, they postulated the existence of a ‘plastic power’ or ‘spirit of nature’, a concept deriving from the *logoi spermatikoi* or *rationes seminales* of the Stoic and Neoplatonic tradition (see Stoicism §5). This concept posits a living force within matter, a self-generating and self-organizing power. Although it is subconscious, this power is also purposive. It is the instrument of God himself, the means by which he achieves his ends in the material world. Cudworth and More denied that their spirit amounts to an occult force, and maintained that it alone avoids the extremes of supernaturalism and mechanism. If supernaturalism admits constant miracles in nature, mechanism derives design from chance. Only the plastic power explains design according to the requirements of naturalism.

To Henry More, the dangers of the new mechanism were most apparent in Descartes’ distinction between the mind and body as thinking and extended substance. If *only* matter is extended, then spirit cannot exist, because what is unextended does not exist anywhere or in any place. If something exists, then it must be somewhere, and therefore extended. Boldly, More concluded that spirit too is extended. He maintained that its essential nature consists in a fourth dimension, ‘spissitude’, the power to expand or contract the space it occupies. Far from being the exclusive attribute of matter, infinite space has many attributes in common with God himself, and it amounts to the divine presence in nature itself. This conception of space later inspired Newton to call space ‘the sensorium of God’ (see Descartes, R. §8; Newton, I. §3).

Although the Cambridge Platonists criticized Cartesian dualism, they introduced one of their own. They attributed opposing characteristics to spirit and matter: spirit is indivisible, active and penetrable, while matter is divisible, passive and impenetrable. Although they postulated their plastic power partly to uphold the continuity of nature, they also understood it to be a spiritual substance distinct in kind from matter. Unlike the vital materialists of the eighteenth century, they did not employ this doctrine to show that mind and body differ only in degree, as if they were only more or less organized forms of living force (see Dualism).

4 Ethics

The hallmarks of the ethics of Cambridge Platonism are realism and rationalism. Its realism grew from its reaction against Hobbes’ and Calvin’s voluntarism. What makes something good or evil, the Platonists contended, is not simply the will of some sovereign power, whether that is God or the civil ruler; rather, it is the thing’s eternal essence or nature, which exists independent of contract, convention or will.

The rationalism of their ethics is apparent from the insistence that we know the nature of good and evil through reason alone. They stress that reason determines not only the means, but also the ends of action. Although they were rationalists, they did not advocate an ethics of duty in the modern Kantian sense. Unlike Kant, they did not distinguish between adherence to principle and conformity to nature, the promotion of human perfection.

The Cambridge men also opposed Hobbes’ and Calvin’s determinism no less than their voluntarism. They defended freedom of the will in the strong sense as a *liberum arbitrium*: the power to choose and act otherwise, and not simply to act on my desires without constraint. They contended that moral responsibility is incompatible with the necessity prevalent in nature. They were especially repelled by the bleak view of humanity in Hobbes and Calvin, and denied that human nature had been so ruined by the Fall that it could not do good by its own efforts. Although they did not dispute the reality of sin, they regarded it as a *perversion* or *disorder* of an inherently good human nature. They also insisted that people are not naturally selfish and competitive. They stressed the social nature of man, and maintained that a person is only happy and fulfilled through a moral life.

In general, ethics played a central role in Cambridge Platonism and its doctrines became especially influential. It maintains that action, not contemplation, is the end of life, and that the highest knowledge is only the result of good conduct. Its emphasis upon reason, the social nature of man and the goodness of human nature constantly reappear in the late seventeenth and early eighteenth centuries. The Cambridge school was particularly influential on Samuel Butler, Shaftesbury and Richard Cumberland, and it was the inspiration behind the ethical rationalism of Samuel Clarke, John Balguy and Richard Price.

See also: Plato; Neoplatonism

FREDERICK BEISER

References and further reading

Anderson, P. (1933) *Science in Defence of Liberal Religion: A Study of Henry More's Attempt to Link Seventeenth Century Religion with Science*, New York: Putnam. (Somewhat out of date, but clear and incisive.)

Cassirer, E. (1953) *The Platonic Renaissance in England*, London: Nelson.(Still the best study.)

Cragg, G. (ed.) (1968) *The Cambridge Platonists*, Oxford: Oxford University Press. (A useful anthology.)

Cudworth, R. (1731) *A Treatise concerning Eternal and Immutable Morality*, ed. E. Chandler, London: Printed for James Knapton.(The fundamental ethical work of the Cambridge school.)

Cudworth, R. (1678) *The True Intellectual System of the Universe*, London; repr. ed. J. Harrison, London, 1845; repr. Stuttgart-Bad Cannstatt, 1964.(The only major work of philosophy by Cudworth published in his lifetime.)

De Pauley, W.C. (1937) *The Candle of the Lord: Studies in the Cambridge Platonists*, London: Macmillan. (Sometimes helpful.)

Lichtenstein, A. (1962) *Henry More: The Rational Theology of a Cambridge Platonist*, Cambridge MA: Harvard University Press.(Incisive treatment of More.)

More, H. (1662) *A Collection of Several Philosophical Writings*, London: William Morden.(An indispensible and important anthology.)

Passmore, J. (1951) *Ralph Cudworth: An Interpretation*, Cambridge: Cambridge University Press.(Helpful analysis).

Patrides, C.A. (ed.) (1968) *The Cambridge Platonists*, Cambridge: Cambridge University Press.(A good anthology.)

Smith, J. (1660) *Select Discourses*, London: Morden.(The only edition of Smith's writings.)

Stewart, J.A. (1910) 'The Cambridge Platonists', in J. Hastings (ed.) *Encyclopedia of Religion and Ethics*, Edinburgh: T&T Clark, vol. 3, 167-73.(A useful introduction to religious aspects).

Tulloch, J. (1874) *Rational Theology and Christian Philosophy in England in the Seventeenth Century*, vol. 2, Edinburgh: Blackwood.(A useful survey).

Whichcote, B. (1751) *The Works of the Learned Benjamin Whichcote*, Aberdeen: Chalmers, 4 vols.(Although indispensible, this is not a complete edition. It also lacks dating of texts.)

Campanella, Tommaso (1568-1639)

Tommaso Campanella was a Counter-Reformation theologian, a Renaissance magus, a prophet, a poet and an astrologer, as well as a philosopher whose speculations assumed encyclopedic proportions. As a late Renaissance philosopher of nature, Campanella is notable for his early, and continuous, opposition to Aristotle. He rejected the fundamental Aristotelian principle of hylomorphism, namely the understanding of all physical substance in terms of form and matter. In its place he appropriated Telesio's understanding of reality in terms of the dialectical principles of heat and cold; and he adopted a form of empiricism found in Telesio's work that included pansensism, the doctrine that all things in nature are endowed with sense. Especially after 1602, Campanella's exposure to Renaissance Platonism also involved him in panpsychism, the view that all reality has a mental aspect. Thus his empiricism came to show a distinctly metaphysical and spiritualistic dimension that transformed his philosophy. At the same time his epistemology embraced a universal doubt and an emphasis on individual self-consciousness that are suggestive of Descartes' views.

Campanella's career as a religious dissident, radical reformer and leader of an apocalyptic movement presents a political radicalism that was oddly associated with more traditional notions of universal monarchy and the need for theocracy. The only one of his numerous writings that receives attention today, La Città del Sole (The City of the Sun) (composed 1602, but not published until 1623), has come to occupy a prominent place in the literature of utopias though Campanella himself seems to have expected some form of astronomical/apocalyptic realization.

Campanella's naturalism, especially its pansensism and panpsychism, enjoyed some currency in Germany and France during the 1620s, but in the last five years of his life it was emphatically rejected by the intellectual communities headed by Mersenne and Descartes, as well as by Galileo.

1 Life

Son of an illiterate Calabrian cobbler, Giovanni Domenico Campanella adopted the name 'Tommaso' when he joined the Dominican Order in 1582. He was attracted to Naples as an intellectual centre, where he encountered the primitive experimentalism and magic of Giambattista Della Porta and his circle. There too in 1591 he published his first work, *Philosophia sensibus demonstrata (Philosophy Demonstrated by the Senses)*, which registered both his emphatic anti-Aristotelianism and his reception of Telesio's naturalism. As a result of his increasingly radical thought, he was hounded by the Inquisition. He suffered a series of trials, tortures and imprisonments, culminating with his twenty-seven year incarceration (1599-1626) by the Spanish viceregal regime for leading an abortive uprising. No matter how severe the conditions of his confinement, he managed to produce a vast number of works, some of which have never been recovered. It was through his writing that he managed to win his way out of prison to freedom in Rome (1626-34) where his astrological talents won him high favour from Pope Urban VIII. However, he was once more threatened by Spain, and fled to Paris where he was granted asylum by Louis XIII and Richelieu. During the last years of his life he prepared his works for publication while under royal protection.

2 Epistemology

It was in Campanella's *De sensu rerum et magia (On the Sense and Feeling in All Things and On Magic)*, first composed in 1590 and published in 1620, that he initially presented his emerging empiricism and his view that the perceived world is alive and sentient in all its parts. This work became the most widely read and influential of all his philosophical works, and remains central to his combination of naturalism and a degree of magic that enables us to manipulate natural processes. However, the most comprehensive and formal presentation of his philosophy is found in his *Metafisica* (1602-24; first published in 1638), which is representative of his mature thought. He divides the work into three parts: *principia sciendi* (epistemology), *principia essendi* (metaphysics) and *principia operandi* (moral and political philosophy); and it is noteworthy that he begins with the theory of knowledge. Thus, before he constructs a metaphysics, Campanella finds it necessary to establish the reality of knowing and of individual self-consciousness. He seems to have been the first philosopher to feel the need of explicitly stating the problem of knowledge as an introduction to his philosophy.

Although he had previously been satisfied through his long exposure to Telesio that the nature of all things

including human beings was basically sensory, by 1604 he had become convinced that the mind that God has infused into human beings not only has sensation and animal memory but something higher and more divine. This higher level of the intellect's functioning established individual self-consciousness. Here Campanella was more inspired by Augustine (§§5-6) than by Aquinas (§10) (his ostensible guide in most matters), especially in his pursuit of a universal theoretical doubt. He entertained the possibility of his own self-deception, and he made explicit reference, paraphrasing it at length, to the famous passage in *De civitate Dei (The City of God)* where Augustine counters the Academic sceptics with his 'If I am mistaken, I am' (XI.26). It is this direct intuition of oneself, this direct knowledge of our being, our knowing and our willing that is now grafted onto the general sensory nature in which the self, like all other things, participates. Although between 1617 and 1623 Descartes (§§3, 7-8) read more of Campanella than he ever wished to admit, the striking similarities between Descartes' views and Campanella's universal doubt, his emphasis on self-consciousness, and his principle *cognoscere est esse* (knowing is being) are more apparent than real. While Descartes creates an emphatic distinction between soul and body, Campanella, instead of separating the two substances, makes a Neoplatonic identification of being with thought, in which all things now share. The difference between them is that between a philosophy of instinct, sensation and *élan vital*, and a philosophy of mathematical certitude, clear and distinct ideas, and a universal, impersonal reason.

By intellect, Campanella understands two distinct faculties. The first is sensation (the *intellectus sensualis*), which humans have in common with animals, and which is capable of grasping particulars but is unable to go beyond the senses. The second is mind (the *intellectus mentalis*) which aspires to invisible and eternal realities. This two-fold view of the intellect is quite similar to that offered by Telesio (§5) in the later editions of his *De rerum natura iuxta propria principia (On the Nature of Things According to Their Own Principles)* (1586). Sense exhibits its own inherent certainty in its apprehension of singulars. Experience shows that to sense fire one need not apprehend the form of fire; rather it suffices simply to be slightly burned. On the basis of this initial passive perception, sense 'can infer the rest of the power of the object acting upon itself'. For Campanella, however, knowledge does not remain at this sensory, Telesian, level, but comes to involve a process of assimilation. Proceeding on the basis of his principle that knowing is being, he claims that things are known inasmuch as the knower becomes similar to them. However, he means by this something quite different from the assimilation of form to which Aristotle and Aquinas had referred. Each thing is sensed and known in so far as it is itself a sentient part of nature, and the knowing essence of the thing becomes the object to be known.

In practical terms, Campanella's empiricism manifests a clear respect for *testimonia* (the evidence of direct witness or experience) in preference to *opinio* (traditional belief based upon some distant authority). The best and most repeated example of this distinction is the witness of Christopher Columbus against the authority of Augustine and Lactantius on the existence of antipodes. Sense, here reading from Nature, God's book, attests to things as they are; opinion to what we have read from others.

Especially in his *Apologia pro Galileo* (written in 1616, published in 1622), Campanella accepts and pays considerable attention to the popular notion of the two Books of Scripture and of Nature. In the *Apologia*, he presents an important statement about freedom of thought and the relationship between God's two Books, though without in any way committing himself to the Copernican system that Galileo had defended (see Copernicus, N.; Galilei, Galileo §§1, 4). In what seems to be a continuous adjustment of the evidence from the two Books in order to achieve their conformity or agreement, Campanella grants both precedence and pre-eminence to the first Book, that of Nature. Yet while the new discoveries provided by the first Book would seem to dictate a reinterpretation of the second that would bring Scripture into conformity with Nature, the entire intellectual enterprise is predicated on the underlying assumption and confidence that there can be no disagreement between what is offered in the Book of Scripture and the findings of the Book of Nature.

3 Metaphysics

Campanella's philosophy of nature is inspired by Telesio's effort to explain nature through its own principles (*iuxta propria principia*) (see Telesio §2). From his first published work, he announced his opposition to Aristotle's hylomorphism. He rejects the idea of matter as pure potency actualized by form, and attributes to it a measure of reality all its own. Nor does he accept form in the scholastic sense, seeing it as the mode or quality of an object. Scorning the Aristotelian structure of matter as potency and form (whether substantial or accidental) as

act, Campanella hoped to build a new edifice on the Telesian foundation of heat and cold as active principles contending for sole mastery of matter as passive substratum. Unfortunately, inconsistency in terminology obscures the clarity of his distinctions and he was by no means completely successful in disentangling himself from the conceptual world and terminology of the Aristotelian tradition.

Campanella confidently asserted that the senses perceived body directly: body is body in its own right without needing an abstract form to make it so. Having renounced substantial forms, he substituted the two active principles of heat and cold. Form is replaced by *temperamentum*, the arrangement and blending of internal parts, as the structure of matter. A more positive evaluation of Campanella's rejection of hylomorphism has detected in this departure from the traditional Aristotelian metaphysical supports a desire to reduce occult phenomena to mechanical contact, to some sort of material force, or to an inhering structural resonance in all things.

Also in reaction to Aristotle, Campanella presented a Neoplatonizing metaphysics that seeks to be in accord with Telesian physics. The subtitle to his *De sensu* (1620) reads: 'the world is shown to be a living and truly conscious image of God, and all its parts and details to be endowed with sense perception, some more clearly, some more obscurely to an extent sufficient for their preservation and that of the entirety in which they share sensation…'. Sense makes the entire world a gigantic, feeling creature.

Campanella goes on to explicate the nature and structure of Being by means of his original doctrine of the primalities (*primalitates*), Power, Wisdom and Love. His reading of the universe as consisting of power, sense or knowledge, and love seems to be a natural outgrowth both of his pansensism and his Augustinian base. The three primalities, his 'Monotriad', provided him with the philosophical counterpart to the central Christian theological doctrine of the Trinity. As metaphysical principles, the primalities are so inherent in the effects they produce that Campanella can speak of their 'essentiating' a being. Lesser beings do not merely participate in the primalities; they are totally pervaded or 'coessentiated' by them.

While the three principles of this 'coessentiation' are the same by virtue of their function, they differ in terms of their origin. Love derives from wisdom and power, and wisdom derives from power alone. Power is thus the source of the other two, though the operation and functioning of all three primalities presupposes their transcendent unity. These relationships are seen in all created beings, for being itself is a transcendental composite of power, knowledge and love. Wisdom, the second primality, is particularly important, for its effects help explain Campanella's doctrine of universal sensation. Primal wisdom endows all things, whether rational, animal, or material, with some measure of sense perception, and it is in this way that all created beings are essentially related to one another and participate in the essence of the infinite being. Love as the third primality is the manifestation of wisdom, for love and knowledge, appetite and perception are conjoined. Love involves a process of becoming something else, the subject becoming imperfectly the object loved. Thus the doctrine that knowledge involves assimilation has a metaphysical basis. Here, as with Campanella's understanding of magic and of natural religion, a general indebtedness to Ficino (§§2, 4) is perceptible.

From this cosmology, the human being emerges as a microcosm or little universe, though Campanella preferred the term *epilogo* (epilogue) for this familiar Renaissance idea. The human microcosm conjoins five worlds or orders: (1) the archetypal world, which is infinite, eternal, and contains the archetypes of all possible worlds; (2) the mental or metaphysical world, that of angelic and human minds, composed of the primalities; (3) the mathematical world or universal space, which constitutes the basis or substratum of all bodies in an unchanging space; (4) the material world whose regular alterations place it above time, so that it participates in the perpetuity of the mathematical world; (5) the fifth, localized world, characterized by time, which is found within the fourth world and is the product of the two active principles. The five orders of being are so interrelated that whatever is in a superior world is by participation in the lower worlds, and whatever is in an inferior world is in an eminent way in those above. The human being as *epilogo* is understood as the ultimate epitome of all five worlds. This account of the microcosm is the corollary of the doctrine that the universe itself, the macrocosm, is a most perfect animal with its own body, spirit and soul (see Cosmology §1).

4 Moral and political philosophy

Campanella distinguishes three aspects of moral or practical philosophy: ethics, economics and politics, the last of which far outweighed the other two in importance. In his ethics he rejects the idea of the supreme good as

something extrinsic to humans, and he also rejects Aristotle's belief that the contemplation of truth is the supreme end for humans. Instead, he affirms the ideal of happiness as consisting in self-conservation, and the direction of all one's efforts towards God, who alone can guarantee this conservation. Beatitude then becomes a union effected with God through power, wisdom and love rather than through any intellectual contemplation. Campanella's understanding of virtue and the virtues follows from his understanding of the supreme good, for virtue has its ground in the three primalities. Virtue is seen as a power proceeding from the very nature of being, and manifesting itself in the operations of power, wisdom and love, which it helps perfect. In his deployment of a number of virtues, Campanella departed from the traditional framework of the four cardinal virtues, and made considerable use of St John Chrysostom, among the Church Fathers.

The focus and realization of the basically social nature of human beings occurs in the individual household, which includes the immediate family, its servants and even the domestic animals. In considering the economy of the family, which differs only in degree rather than in kind from the political state, Campanella asserts the centrality of marriage as a contract of natural law, sanctified by Christ as a sacrament. If some of his recipes for sexual relations, child-rearing and the education of the young suggest the calculations and concerns of the Solarians in his imaginary City of the Sun (see below), there is no hint of the Solarian community of women to be found in his treatment of marriage. Campanella is thus generally orthodox in his understanding of marriage within the present temporal dispensation, although the wife exercises a responsible role within the conservative structure headed by the *paterfamilias*.

Politics early became the pre-eminent consuming interest of the radical Dominican reformer - an interest, which when driven by the heady combination of astrology, magic and messianism led to a popular uprising, and to Campanella's long imprisonment. His famous work *La Città del Sole (The City of the Sun)* (first published in 1623), traditionally understood as being simply the account of an ideal society, and thus relegated to utopian literature (see Utopianism §1), is more accurately understood as representing part of Campanella's millenarian expectation. It seems hardly accidental that he composed the work in 1602 only a year before the expected astronomical conjunction of planets figuring the sun's descent towards the earth, set for December 1603. Despite the apparent incommensurability of the then-current Christian economy with the rationally organized, communistic millenarian economy which Campanella saw as impending, there is a complementarity between the City of the Sun and the theocratic papal monarchy advanced by his other political writings: namely, the amalgamation of secular and spiritual into a comprehensive universal monarchy ruled by a priest-king.

From the very beginning two themes pervade Campanella's political philosophy: (1) the necessity of religion's informing politics; (2) the necessity for universal monarchy. These two threads, running through his entire thinking, intersect in his affirmation of the papal office as the effective expression and realization of power on earth. Monarchy is not simply the noblest form of government, but in Campanella's hands it came to be understood as something universal, all-inclusive and above the present clutter of kingdoms and empires. With Christ seen as the founder of universal monarchy, the designation of monarch properly pertains to the Pope, who when endowed with both temporal and spiritual arms, can serve as moderator among princes, and can prevent any Christian prince from arriving at the monarchy of Christendom. Campanella was scandalized by Dante's dualism of church and empire (see Alighieri, Dante §3), and wished to fuse secular and spiritual power in the monarchy of the priest-king. Thus one can only understand the monarchy of Spain (the ruler of Naples, and together with the Roman Inquisition, author of Campanella's misfortunes) as the Arm of God when it is in close association with, dependence on, and service to the papacy; the territorial aggrandizement of Spain can have the papal blessing only in so far as it serves the universal propagation of the true religion. After 1634, Campanella was to put France forward as the Pope's secular agent.

Campanella is a Machiavellian of sorts. Although apart from Aristotle himself no other so provoked the hostility of Campanella as did Machiavelli and his reason of state, none the less Campanella could accept much of his predecessor's appreciation of power, its defence and increase through deception and harsh, sometimes immoral measures. On the other hand, he could not admit the divorce of politics from religion or the reduction of religion to some form of political manipulation and a mere means of ruling (see Political philosophy, history of §5). Affirming religion's political utility, Campanella seeks to incorporate the state into his church, empowering Christianity under the aegis of the armed priest-king. He wants to have his religion both ways: as justified on utilitarian grounds, and as the comprehensive, informing, living truth.

Campanella's entire thought reflects an emotional predisposition to innovation and a commitment to mint a new philosophy, eclectic and Platonizing, that is ostensibly more supportive of Christianity than the philosophy of Aristotle.

See also: Aristotelianism, Renaissance; Platonism, Renaissance; Telesio, B.

JOHN M. HEADLEY

List of works

Most of Campanella's works went through numerous stages of complex redaction before publication; for details see Firpo (1940) in references and further reading.

Campanella, T. (1591) *Philosophia sensibus demonstrata (Philosophy Demonstrated by the Senses)*, Naples: Orazio Salviano; ed. L. de Franco, Naples: Vivarium, 1992. (Campanella's first publication and first formal venture into philosophy, composed 1589 but not published until 1591. Registers his emphatic anti-Aristotelianism, the reception of Telesio's philosophy and its transfer into a metaphysical key; precedes his magical, astrological and Gnostic speculations and his later interest in celestial, political and religious palingenesia.)

Campanella, T. (1620) *De sensu rerum et magia (Of the Sense and Feeling in All Things and on Magic)*, Frankfurt: Gottfried Tampach; repr. in L. Firpo (ed.) *Tommaso Campanella, Opera latina, Francofurti impressa annis 1617-1630*, Turin: Bottega d'Erasmo, 1975; English trans. of Book I and some chapters of Book IV (incorrectly designated Book V) in A.B. Fellico and H. Shapiro (eds) *Renaissance Philosophy; the Italian Philosophers*, New York: Modern Library, 1967, 338-79; partial English trans. of Book II by B. Dooley, *Baroque Italy: Selected Readings*, New York: Garland Press, 1995.(Composed first in Latin in 1590, then recomposed in Italian in 1604; a critical edition of the more definitive Italian version is available as *Del senso delle cose e della magia*, ed. Antonio Bruers, Bari: Gius. Laterza & Figli, 1925. After its publication in 1620, *De sensu* became the most widely read and influential of Campanella's philosophical works during the seventeenth century and remains central to his total naturalistic, magical position.)

Campanella, T. (1622) *Apologia pro Galileo*, Frankfurt: Gottfried Tampach; repr. in L. Firpo (ed.) *Tommaso Campanella, Opera latina, Francofurti impressa annis 1617-1630*, Turin: Bottega d'Erasmo, 1975; trans. R.J. Blackwell, *A Defense of Galileo, the Mathematician from Florence*, Notre Dame, IN, and London: University of Notre Dame Press, 1994.(Composed in 1616, presents an important statement of *libertas philosophandi* and the relationship between God's two Books without in any way committing the author to the Copernican system; Blackwell's is the best English translation.)

Campanella, T. (1623) *La Città del Sole: Dialogo Poetico; The City of the Sun: A Poetical Dialogue*, trans. and ed. D.J. Donno, Berkeley and Los Angeles, CA, and London: University of California Press, 1981.(Although composed in 1602, *La Città del Sole* was not published until 1623; most famous as a work of utopian literature.)

Campanella, T. (1636) *De gentilismo non retinendo (On that Paganism that must not be Maintained)*, Paris: Toussaint Dubray; Italian trans. R. Amerio, *Della necessità di una filosophia cristiana*, Turin: Società editrice internazionale, 1953.(Composed originally 1609-10; more than usually pungent effort on Campanella's part to mint a new Christian philosophy, eclectic but exclusive of Aristotle.)

Campanella, T. (1638) *Metafisica: universalis philosophiae seu metaphysicarum rerum propria dogmata (Metaphysics: The Essential Doctrines of Universal Philosophy or Metaphysical Things)*, ed. P. Ponzio, Bari: Levante, 1994; repr. L. Firpo, Turin: Bottega d'Erasmo, 1961; Latin-Italian abridged version, G. di Napoli (ed.) *Metafisica*, Bologna: Zanichelli, 1967, 3 vols.(First published in Paris in 1638, following four redactions 1602-24. Constitutes the most comprehensive, cumulative and formal presentation of Campanella's philosophy. One of the great dinosaurs of seventeenth-century philosophical production, di Napoli's Latin-Italian abridged version is more manageable.)

References and further reading

Amerio, R. (1972) *Il sistema teologico di Tommaso Campanella (The Theological System of Tommaso Campanella)*, Milan and Naples: R. Riccardi.(The most formidable interpretation of Campanella as a Christian thinker, by the editor of Campanella's theological works.)

Augustine (413-27) *De civitate Dei (The City of God)*, trans. J. O'Meara, Harmondsworth: Penguin, 1972.(Quoted in §2. Several good English translations exist, of which this is the most notable.)

Badaloni, N. (1965) *Tommaso Campanella*, Milan: Feltrinelli.(A thematic, perceptive study, emphasizing the magical and the influence of Origen in Campanella's thought.)

Blanchet, L. (1920) *Campanella*, Paris; repr. New York: Burt Franklin, 1964. (Although somewhat dated, a brilliant, comprehensive treatment of Campanella as a radical philosopher of nature.)

Bock, G. (1974) *Thomas Campanella: Politisches Interesse und Philosophische Spekulation (Thomas Campanella: Political Interest and Philosophical Speculation)*, Tübingen: Max Niemeyer.(A thorough, balanced Marxist interpretation of Campanella's political thought.)

Bolzoni, L. (1989) 'Tommaso Campanella e le donne: fascino e negazione della differenza' (Thomas Campanella and Women: Charm and the Denial of Difference), *Annali d'Italianistica* 7: 193-216.(Sensitive, perceptive analysis of Campanella's ambivalent view of women.)

Bonansea, B.M. (1969) *Tommaso Campanella: Renaissance Pioneer of Modern Thought*, Washington, DC: Catholic University of America Press.(Fundamental, though excessively scholastic and impervious to Renaissance intellectual currents.)

Copenhaver, B.P. and Schmitt, C.B. (1992) *A History of Western Philosophy: Vol. 3. Renaissance Philosophy*, Oxford and New York: Oxford University Press.(Contains an important, positive appreciation of Campanella as a critic of Aristotle who significantly undermined the current allegiance to occult qualities through his rejection of hylomorphism.)

Ernst, G. (1991) *Religone, ragione e natura: Ricerche su Tommaso Campanella e il tardo Rinascimento (Religion, Reason and Nature: Research on Tommaso Campanella and the Late Renaissance)*, Milan: Franco Angeli.(An insightful thematic study of aspects of Campanella's thought.)

Firpo, L. (1940) *Bibliografia degli scritti di Tommaso Campanella (Bibliography of Tommaso Campanella's Writings)*, Turin: Vincenzo Bona.(Fundamental for understanding the complicated history of Campanella's writings, though in need of revision.)

Firpo, L. (1974) 'Campanella, Tommaso', in *Dizionario biografico degli Italiani* 17: 372-401, Rome: Società Grafica Romana.(Basic biographical and bibliographical information by the leading Campanella scholar.)

Headley, J.M. (1997) *Tommaso Campanella and the Transformation of the World*, Princeton, NJ: Princeton University Press, 1997.(Seeks to contextualize the thought of Campanella according to major issues in the Europe of the seventeenth century.)

Lerner, M.-P. (1987) 'Le "livre vivant" de Dieu: la cosmologie évolutive de Tommaso Campanella' (God's 'Living Book': the Developing Cosmology of Tommaso Campanella), in *Le discours scientifique du Baroque: Actes de la Xe session internationale d'étude du Baroque*, Montauban, 1983.(Thorough discussion of Campanella's interpretation of nature.)

Mönnich, M. (1990) *Tommaso Campanella: Sein Beitrag zur Medizin und Pharmazie der Renaissance (Tommaso Campanella: His Contribution to the Medicine and Pharmacy of the Renaissance)*, Stuttgart: Wissenschaftliche Verlagsgesellschaft.(Excellent presentation of the scientific and medical features of Campanella's thought in their larger philosophical context; notable analyses of *sensus* and *spiritus*; good bibliography.)

Campbell, George (1719-96)

George Campbell, Scottish minister, professor and religious thinker, is now remembered primarily for The Philosophy of Rhetoric (1776). Here he employed the Scottish Enlightenment's developing science of human nature to explain the effectiveness of the classical rules of rhetoric. He did this by relating the various ends of persuasive discourse to the natural faculties and propensities of the human mind. In his own time Campbell was better known as a religious apologist, using an enlightened theory of evidence in A Dissertation on Miracles (1762) to defend the believability of Christian miracles against the sceptical attack of David Hume.

George Campbell was born in Aberdeen and educated at the town's Marischal College. He was ordained in the established Church of Scotland in 1748 and was minister of a country parish for nearly a decade. Upon returning to Aberdeen, he helped found the influential Aberdeen Philosophical Society, whose members included Thomas Reid, Alexander Gerard and James Beattie (see Aberdeen Philosophical Society). He was appointed to the posts of Principal and Professor of Divinity in Marischal College in 1759 and 1771 respectively, which he held jointly until the year before his death.

Campbell's best known work today is *The Philosophy of Rhetoric* (1776), which earned its title by combining classical rhetorical theory with the Scottish Enlightenment's attempt to develop a comprehensive science of human nature (see Enlightenment, Scottish; Human nature, 18th-century science of). Campbell argued that the rhetorical standards established by the ancients were definitive, and could be used to explore the principles of human nature that made those standards universal. He linked the four ends of speaking, which were 'to enlighten the understanding, to please the imagination, to move the passions, or to influence the will', with the qualities of style, perspicuity and vivacity, that most effectively brought about persuasion (Campbell 1776: 1). He also employed the Scottish Enlightenment's principle of sympathy to explain how an orator communicated passions as well as believability to his audience.

Perhaps the most important philosophical aspect of *The Philosophy of Rhetoric* is its theory of evidence. Campbell divided evidence into two types, intuitive and deductive. Intuitive evidence is immediately convincing to the mind and provides a necessary foundation for rational and empirical argument. It includes evidence from pure intellection (self-evident mathematical axioms and definitions), consciousness (one's own being and sensations), and common sense (indemonstrable but indispensable premises of moral reasoning). Deductive evidence, by contrast, produces conviction mediately, by either logical or factual demonstration, denominated demonstrative and moral evidence. Demonstrative evidence concerns invariable mathematical and geometrical relations and is capable of absolute proof. Moral evidence concerns contingent matters of fact rather than necessary relations, and its proofs are only probable at best. It is, nevertheless, the most important kind of evidence, for it includes all human experience and testimony, and constitutes the proper province of rhetoric.

Campbell's philosophy was influenced by that of David Hume, though Campbell employed much effort in correcting Hume. Campbell followed Hume in equating experience with the habitual associations of ideas in the mind, but thought that Hume's extreme emphasis on personal experience seriously undervalued the importance of memory (the only voucher for the past evidence of the senses) and of testimony to human knowledge as a whole. He therefore employed Common Sense philosophy to defend the necessary reliance of human beings on their own memories and on the testimony of others (see Common Sense School). His theory of evidence was a significant and original contribution to Common Sense philosophy, particularly in its explication of intuitive knowledge and its emphasis on testimony, though his use of that philosophy was otherwise minimal and not nearly as extensive as the version employed by Thomas Reid.

The methods and concerns of *The Philosophy of Rhetoric* were those of the Scottish Enlightenment, but its rhetorical strategy was aimed primarily at religious persuasion. Campbell's theory of evidence provided a philosophical foundation for his Christian apologetics, as exemplified by *A Dissertation on Miracles* (1762). This work was written against Hume's essay 'Of Miracles' (1748), which had claimed that a miracle could not be believed because the testimonial evidence in its favour necessarily contradicted our uniform experience of the laws of nature. Campbell argued that our knowledge of the laws of nature depended more upon the testimony of others than upon our personal experience. We could therefore no more reject the testimony of reliable witnesses

concerning particular facts than we could the general testimony that had established the empirical laws of nature in the first place. He then attempted to demonstrate that only early Christian claims of miracles were sufficiently reliable to overturn the general experience of human history.

Campbell was a relentless advocate for religious toleration and freedom of expression. His *Address to the People of Scotland* (1779; see Campbell 1762) defended the civil rights of Roman Catholics, though his outspokenness earned him the wrath of many of his Protestant countrymen. He argued that persecution was unchristian, counterproductive and philosophically indefensible. In a later manuscript (now in Aberdeen University Library) he contended that belief in abstract doctrines is the involuntary consequence of the passive mind’s perception of evidence. Only argument and persuasion are compatible with the realities of the human mind and with the spirit of the gospel of Christ.

Campbell’s reputation in his own lifetime rivalled that of almost any participant in the Scottish Enlightenment, and rested primarily on his religious writings, which included his posthumously-published divinity lectures. His contemporaries believed his masterpiece to be *The Four Gospels* (1789), which contained dissertations on the proper method of translating ancient documents, and which advocated enlightened principles of biblical criticism. Since his death, *The Philosophy of Rhetoric* has become a standard text on the subject as well as the object of considerable scholarly interest. It has eclipsed all his other writings.

See also: Blair, H.; Miracles; Rhetoric

JEFFREY M. SUDERMAN

List of works

Campbell, G. (1762) *A Dissertation on Miracles*, Edinburgh: Kincaid & Bell; 3rd edn, Edinburgh: Bell & Bradfute, 1797, 2 vols; repr. New York: Garland, 1983.(The third edition contains most of Campbell’s sermons, including *An Address to the People of Scotland*.)

(1776) *The Philosophy of Rhetoric*, ed. L. Bitzer, Carbondale, IL and Edwardsville, IL: Southern Illinois University Press, 1988.(The now standard edition of the *Rhetoric*; Bitzer’s introduction is the best short scholarly introduction to Campbell’s life and thought.)

References and further reading

Bevilacqua, V. (1965) ‘Philosophical Origins of George Campbell’s *Philosophy of Rhetoric*’, *Speech Monographs* 32: 1-12.(A reasonably accessible introduction to Campbell’s rhetorical thought.)

Bitzer, L. (1969) ‘Hume’s Philosophy in George Campbell’s *Philosophy of Rhetoric*’, *Philosophy and Rhetoric* 3: 139-66.(A compelling case for Campbell’s debt to Hume.)

Burns, R. (1981) *The Great Debate on Miracles*, Lewisburg, PA: Bucknell University Press.(A treatment of the whole eighteenth-century miracles debate that puts Campbell’s arguments in context.)

Howell, W. (1971) *Eighteenth-century British Logic and Rhetoric*, Princeton, NJ: Princeton University Press.(The standard account of the subject, which gives considerable space to Campbell.)

Hume, D. (1748) ‘Of Miracles’, in *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 2nd edn, 1972.(The most famous eighteenth-century attack on miracles, although coming rather late in the debate, and against which Campbell and many others wrote.)

Campbell, Norman Robert (1880-1949)

Campbell made important contributions to philosophy of science in the 1920s, influenced by Poincaré, Russell and his own work in physics. He produced pioneering analyses of the nature of physical theories and of measurement, but is mainly remembered for requiring a theory, for example, the kinetic theory of gases, to have an 'analogy', that is, an independent interpretation, for example, as laws of motion of a swarm of microscopic particles.

The British philosopher of science Norman Robert Campbell, who became a Fellow of Trinity College Cambridge in 1904, was also an experimental physicist, and worked on the research staff of the British General Electric Company from 1919 to 1944. His main contribution to philosophy, published in 1920, is his account of how physical theories explain laws. It maintains an absolute distinction between laws relating observable properties of objects, on which agreement can be achieved, and theories used to explain them. It could allow a weaker distinction, letting accepted theories come to state laws needing further explanation. But only an implausible view of the significance of the distinction can save its claim that theories need analogies.

Campbell's account of theories credits them with three components, illustrated by a simplified version of the kinetic theory of gases. First there is a theory's 'hypothesis', its mathematical propositions, empirically uninterpreted. Then there is a 'dictionary', linking terms of the hypothesis to observable terms used to state the laws the theory explains. Thus in his example the dictionary identifies the volume V, mass M, pressure P and absolute temperature T of a gas with combinations of constants and variables postulated by the hypothesis: for example, $V = l^3$, where l is a constant, $M = nm$, where m is a constant and $3n$ the number of variables dependent on the independent variable t (time). This hypothesis and dictionary entail the perfect gas law, $PV \propto T$.

But arbitrarily many formal systems, suitably interpreted, would also entail this law. To *explain* it, ' the propositions of the hypothesis must be analogous to some known laws': here 'the laws which would describe the motion of … [n] infinitely small and highly elastic particles [of mass m] … in a cubical box [of size l]' (1920: 128-9). Thus for Campbell an analogy is an essential part of a theory, not a dispensable aid to its formation: ' to regard analogy as an aid to the invention of theories is as absurd as to regard melody as an aid to the composition of sonatas' (1920: 130).

The importance of Campbell's main distinction, between propositions linking a theory's terms to each other and those linking them to others in the laws it explains, is now widely accepted if differently expressed. (For example, Nagel (1961) uses 'calculus' for Campbell's 'hypothesis' and 'correspondence rules' for his 'dictionary'.) But few now accept that a theory needs an analogy (which Nagel called a 'model'), a thesis that falls between two stools. First, a theory needs no analogy on an instrumentalist reading of it as a formal device linking laws statable in other terms: for then, since its hypothesis states nothing either true or false, its terms need no interpretation. But nor does it need an analogy on a realist reading of its hypothesis as consisting of true or false propositions. For then what these propositions state if true are not merely *analogous* to laws governing (for example) the motion of particles: they *are* laws governing the motion of the particles which the kinetic theory says compose a gas. Only if a hypothesis comprises propositions that are true or false but different in kind from those that state laws is Campbell's case for analogy tenable; but his - basically epistemological - argument for this was always weak and has been weakened further by later work on the meaning of theoretical terms in science (see Theories, scientific §§2, 7).

Although most physical laws relate measurable quantities, measurement remains an underrated topic in the philosophy of science. Campbell was one of the first philosophers to recognise its importance, observing that physics 'might almost be described as the science of measurement'. He saw that measurement itself depends on laws, like those giving physical sense to functions (addition, subtraction, multiplication) of the numbers used to measure quantities like length. His systematic account of this dependence, of the difference between fundamental and derived magnitudes, of the significance of units and dimensions, and of errors of measurement, set the agenda for later theories, which remain indebted to it.

See also: Duhem, P.M.M.; Measurement, theory of; Models; Scientific realism and antirealism

D.H. MELLOR

List of works

Campbell, N.R. (1920) *Physics: the Elements*, Cambridge: Cambridge University Press.(Republished in 1957 as *Foundations of Science*, New York: Dover Publications. Part I contains Campbell's account of theories, Part II his account of measurement.)

Campbell, N.R. (1921) *What is Science?*, Cambridge: Cambridge University Press. (Republished in 1952, New York: Dover Publications. An introductory book on the scope of science, laws, measurement, mathematics in science, and its applications.)

Campbell, N.R. (1928) *An Account of the Principles of Measurement and Calculation*, London: Longmans.(The final version of Campbell's account of measurement.)

References and further reading

Kyburg, H.E., Jr (1984) *Theory and Measurement*, Cambridge: Cambridge University Press.(A recent account of measurement, drawing on Campbell's work.)

Nagel, E. (1961) *The Structure of Science*, New York: Harcourt Brace.(Chapters 5 (II-III) and 6 (I) of this classic work discuss Campbell's account of theories.)

Camus, Albert (1913-60)

Albert Camus was awarded the Nobel Prize in 1957 for having 'illuminated the problems of the human conscience in our times'. By mythologizing the experiences of a secular age struggling with an increasingly contested religious tradition, he dramatized the human effort to 'live and create without the aid of eternal values which, temporarily perhaps, are absent or distorted in contemporary Europe'(1943). Thus the challenge posed by 'the absurd' with which he is so universally identified.

Camus' most celebrated work is *L'Étranger* (*The Stranger*) (1942). Depicting the 'metaphysical' awakening of an ordinary Algerian worker, Camus concretizes the Pindarian injunction, provided as life's answer to 'the absurd' in an epigram to *Le Mythe de Sisyphe* (*The Myth of Sisyphus*) (1943): 'Oh my soul do not aspire to immortal heights but exhaust the field of the possible.'

But if the 'absurd' defines our world, it was never treated by Camus as a conclusion, only 'a point of departure'.

> What else have I done except reason about an idea I discovered in the streets of my time? That I have nourished this idea (and a part of me nourishes it still) along with my whole generation goes without saying. I simply set it far enough away so that I could deal with it and decide on its logic.
>
> (1954)

How and what morality is still possible, then, in view of the experience of 'the death of God' which has given birth to the experience of absurdity? While the absurd leaves humans without justification and direction, rebellion bears witness to the refusal of human beings to accept this incipient despair.

By demanding an end to oppression, rebellion seeks to transform - by revolution if necessary - the conditions that gave rise to it. Rebellion thus testifies to the human being's incessant demand for dignity. But it is often a vain yearning without a revolutionary transformation of the institutional structures of exploitation and oppression. Yet that transformation only promises further and even greater humiliation if it is not continually guided by the spirit and concerns of rebellion. Appalled by the totalitarian direction of many modern revolutionary movements, Camus thought he detected a messianic nostalgia lurking at the core of Western rebellions. He saw them driven by an often unexpressed need to replace the failed vertical transcendence of Judaeo-Christianity with a new horizontal transcendence.

L'Homme revolté (*The Rebel*) (1951) his major work in political theory, thus seeks to diagnose a way of thinking that has led rebellious thought from its initial generous impulses down the path of destruction, thus undermining one of the few sources of hope in our post-Christian world.

Rebellion and revolution have often, however, been falsely seen as polarized for Camus. Their opposition, rather than being necessary and celebrated, as has been claimed by both right- and left-wing detractors, is the death of both of them. The ground of any meaningful development of rebellion must be the implicit community of humans for whom value is demanded. Here is the crux of Camus' critique of 'legitimate murder', including capital punishment, which has led him, mistakenly, to be defined as a pacifist. Self-defence is justified, both individually and collectively, but pre-meditated or logical murder in the service of any cause whatsoever is not. It undermines the one undeniable community of humans confronting the universe, destroying the grounds of the possibility of coherent social values.

Thus Camus rejected any theory that argued that the ends justify the means. There are no transcendent ends. All ends are visions of transformed futures which themselves will simply be means for further action. If the ends justify the means, what then can justify the ends? His answer is the means - because they are simply more proximate ends in the service of a transformed quality of life that must always be lived concretely in the temporally unfolding present, ever confronted with injustice and exploitation, envisaging a transformed future that action may aspire to bring into being.

Throughout his life his thought was deeply marked by his experience. A French-Algerian from a poor working-class district of Algiers, Camus confronted Western civilization from the margins. Separated by ethnicity and culture from the Muslim majority, and by class from the ruling French *colons*, his attitude toward Europe

remained ambivalent: drawn by the brilliance of its cultural expressions yet repelled by the scope of its brutalizing inhumanity.

To an emerging class sensitivity was added an anguished awareness of human finitude, poignantly brought home by contracting tuberculosis and facing death at the age of 17. A very personal urgency thus vitalized his metaphysical reflections. His mature work is marked by the conflicted awareness of the sensuality of the body in direct contact with a vibrant nature confronting the inevitability of ageing and death in a stark landscape without illusions.

Almost alone among traditional intellectuals, Camus remained tied by sensitivity and vision to his working-class origins. His values drew more from the communal experiences of college soccer than from the theoretical speculations of our greatest writers. The egalitarian collective effort of the Workers' Theatre that he led in the mid-1930s, where stars did not take bows and everyone pitched in with the staging, suggested to him the outlines of a truly just society.

An advocate of 'relative utopias', the renaissance, to the creation of which Camus always remained passionately committed, meant the qualitative transformation of daily life, the creation of dialogic communities at work and at home that gave voice and sustenance to the struggles for dignity of ordinary people. He continued to believe that only when the dignity of the worker and respect for intelligence are accorded their rightful place can human existence hope to realize its highest ideals, and our life find the collective meaning and purpose that alone can truly sustain us in the face of an infinite and indifferent universe.

See also: Existentialism; Existentialist ethics

DAVID A. SPRINTZEN

List of works

Camus, A. (1938) *Noces*, Algiers: Charlot; trans. E. Conroy Kennedy, 'Nuptials', in *Lyrical and Critical Essays*, New York: Alfred A. Knopf, 1968.

Camus, A. (1942) *L'Étranger*, Paris: Gallimard; trans. M. Ward, *The Stranger*, New York: Alfred A. Knopf, 1988.

Camus, A. (1943) *Le Mythe de Sisyphe*, Paris: Gallimard; trans. J. O'Brien, *The Myth of Sisyphus and Other Essays*, New York: Alfred A. Knopf, 1955.

Camus, A. (1944) *Caligula*, Paris: Gallimard; trans. J. O'Brien, *Caligula and Three Other Plays*, New York: Alfred A. Knopf, 1958.

Camus, A. (1947) *La Peste*, Paris: Gallimard; trans S. Gilbert, *The Plague*, New York: Alfred A. Knopf, 1948.

Camus, A. (1950a) *Les Justes*, Paris: Gallimard; trans. J. O'Brien, The Just Assassins, in *Caligula and Three Other Plays*, New York: Alfred A. Knopf, 1958.

Camus, A. (1950b, 1956, 1958) *Actuelles*, vol. 1, *Chroniques 1944-1948*, vol. 2, *Chroniques 1948-53*, vol. 3, *Chroniques algériennes 1939-58*, Paris: Gallimard; trans. J. O'Brien, *Actualities*, and partially reprinted in *Resistance, Rebellion, and Death*, New York: Alfred A. Knopf, 1960.

Camus, A. (1951) *L'Homme revolté*, Paris: Gallimard; trans. A . Bower, *The Rebel*, New York: Alfred A. Knopf, 1954.

Camus, A. (1954) *L'Été*, Paris: Gallimard; trans. E. Conroy Kennedy, *Summer*, in *Lyrical and Critical Essays*, New York: Alfred A. Knopf, 1968.

Camus, A. (1956) *La Chute*, Paris: Gallimard; trans. J. O'Brien, *The Fall*, New York: Alfred A. Knopf, 1957.

Camus, A. (1957) *L'Éxile et le royaume*, Paris: Gallimard; trans. J. O'Brien, *Exile and the Kingdom*, New York: Alfred A. Knopf, 1957.

Camus, A. and Koestler, A. (1957) *Réflections sur le peine capitale: introduction et étude de Jean Bloch-Michel*, Paris: Calmann-Levy; trans. J. O'Brien, *Reflections on the Guillotine*, and partially reprinted in *Resistance, Rebellion, and Death*, New York: Alfred A. Knopf, 1960.

Camus, A. (1958) *'Discours de Suède' et 'L'Artiste et son temps'*, Paris: Gallimard; trans. J. O'Brien, 'Camus in Stockholm', *Atlantic Monthly* May 1958; 'The Artist and His Time', in *Resistance, Rebellion, and Death*, New York: Alfred A. Knopf, 1960.

Camus, A. (1994) *Le Premier Homme*, Paris: Gallimard; trans. D. Habgood, *The First Man*, New York: Alfred A. Knopf, 1995.

References and further reading

Bree, G. (1964) *Camus*, New York: Harcourt Brace & World.(Popular and sympathetic presentation from a literary perspective.)

Hanna, T. (1958) *The Thought and Art of Albert Camus*, Chicago, IL: Regnery.(A popular philosophical overview from a liberal perspective.)

Isaac, J.C. (1992) *Arendt, Camus, and Modern Rebellion*, New Haven, CT: Yale University Press.(A thoughtful, detailed and sympathetic critical comparison of the two thinkers.)

Lazere, D. (1973) *The Unique Creation of Albert Camus*, New Haven, CT: Yale University Press.(A serious study with a psychoanalytic bent.)

Lottman, H. (1979) *Albert Camus: A Biography*, Garden City, NJ: Doubleday & Co.(A very detailed factual account of Camus' life.)

Luppé, R. (1966) *Albert Camus*, trans. J. Cumming and J. Hargreaves. London: Merlin.(A succinct and original French interpretation.)

McCarthy, P. (1982) *Camus*, New York: Random House.(An intelligent critical review of Camus' philosophy that downplays the importance of *The Rebel*.)

O'Brien, C.C. (1970) *Albert Camus of Europe and North Africa*, New York: Viking Press.(An original analysis that critically evaluates the Cold War dimensions of Camus' life and writings.)

Parker, E. (1965) *Albert Camus: The Artist in the Arena*, Madison, WI: University of Wisconsin Press.(A presentation of Camus' journalistic writings in the immediate post-war period.)

Sprintzen, D.A. (1988) *Camus: A Critical Examination*, Philadelphia, PA: Temple University Press.(A comprehensive critical analysis of the philosophical structure and development of Camus' thought.)

Thody, P. (1961) *Albert Camus: 1913-1960*, London: Hamish Hamilton.(A popular literary overview of Camus' writings.)

Cantor, Georg (1845-1918)

Georg Cantor and set theory belong forever together. Although Dedekind had already introduced the concept of a set and naïve set theory in 1872, it was Cantor who single-handedly created transfinite set theory as a new branch of mathematics. In a series of papers written between 1874 and 1885, he developed the fundamental concepts of abstract set theory and proved the most important of its theorems. Although today set theory is accepted by the majority of scientists as an autonomous branch of mathematics, and perhaps the most fundamental, this was not always the case. Indeed, when Cantor set out to develop his conception of sets and to argue for its acceptance, he initiated an inquiry into the infinite which raised questions that have still not been completely resolved today.

1 Cantor's development of set theory

Cantor was born to a wealthy merchant family in St Petersburg and first went to Germany (Frankfurt) at the age of eleven. He studied in Zurich, Göttingen and (mainly) in Berlin, under Kummer, who directed his dissertation (completed in 1867), Weierstrass and Kronecker. In 1869, after a brief intermezzo as a teacher, Cantor went to Halle as a *Privatdozent*. Despite persistent attempts to secure a better position at a more prestigious university, he spent his entire career there, dying in 1918 after several bouts of manic depression. Thanks to the efforts of Hilbert, Zermelo, Fraenkel and many others he did see his ideas accepted by the majority of mathematicians.

The beginnings of Cantor's work on set theory emerged towards the end of 1873 in an exchange of letters with his colleague Richard Dedekind (see Cantor and Dedekind 1937). In a letter dated 29 November Cantor asks whether there is a one-one correspondence between the real numbers and the natural numbers. He already knew, from Dedekind, that it was possible to put the algebraic numbers into one-one correspondence with the natural numbers. He doubted that the same was true of the real numbers because the reals form a continuum while the natural numbers are discrete. As far as we know, Dedekind did not answer the question, believing it to be of no practical importance (although he saw its theoretical relevance).

In his next letter (7 December) Cantor answered his own question: he informed Dedekind that he had a proof of the impossibility of a one-one correspondence between the real numbers in the open unit interval (that is, those between, but not including, 0 and 1) and the natural numbers. This was the first proof that the real numbers are 'uncountable'. A few months later (1874), Cantor published his proof under the rather misleading title 'Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen' (On a Property of the Set of all Real Algebraic Numbers). The main result, not mentioned in the title, had as a consequence the proof of the existence of 'transcendental' - that is, non-algebraic real - numbers.

This little-noticed paper showed for the first time that different 'levels' of infinity must be distinguished and this result, perhaps more than any other, triggered the further development of set theory. Dedekind was probably the only mathematician at the time who understood the real significance of this work, correcting his previous opinion of the practical unimportance of Cantor's question as 'conclusively rejected by Cantor's proof'. But what is its significance? First, the proof shows that the distinction between different levels of infinity is not empty - by no means a trivial result because until Cantor almost all mathematicians believed there to be essentially only one 'size' of infinity. Second, it challenges the Aristotelian dogma that the infinite exists merely potentially, but not actually.

In his next paper, 'Ein Beitrag zur Mannigfaltigkeitslehre' (1878), Cantor considered the 'power' of sets: two sets are of equal power (cardinality) if there is a one-one correspondence between their elements. He tried to show that there exist sets of higher cardinality than the continuum by showing that a continuum of two dimensions cannot be mapped one-one onto a continuum of one dimension. To his own surprise, however, the result was exactly the opposite: all continua of any countable dimension have the same cardinality as the linear continuum (the real numbers).

$$|\mathbb{R}\times\mathbb{R}\times\mathbb{R}\times\ldots| = \ldots = |\mathbb{R}\times\mathbb{R}| = |\mathbb{R}|$$

This unexpected result led Cantor to state for the first time his famous 'continuum hypothesis': that in the linear continuum of real numbers exactly two classes of infinities are represented - countable (the cardinality of the natural numbers) and uncountable (the cardinality of the reals) - with none in between (see Continuum hypothesis;

Set theory §3).

The following six years were the most fruitful of Cantor's life. Not only did he publish (in six parts) the paper 'Über unendliche lineare Punktmannigfaltigkeiten' (On Infinite Linear Manifolds of Points), but he also developed his 'arithmetical' theory of the infinite. The first step in the development was the introduction of 'derived sets' represented by the symbols '∞', '$\infty + 1$', and so on. Later he took these symbols as representing transfinite ordinal numbers. He defined a generalized 'successor' operation, and a 'supremum' operation which gives the limit of an infinite sequence of ordinal numbers. Starting from 0, the ordinal of the empty set, the 'successor' operation generates the sequence of natural numbers and the 'supremum' operation secures the existence of the first transfinite number - the ordinal of the set of natural numbers - called ω_0 or ω. The 'successor' operation in turn generates the numbers $\omega + 1, \omega + 2, \ldots$ and the 'supremum' operation then guarantees the existence of $\omega + \omega$, and so on. (See Set theory §2.)

Cantor divided the ordinals into number classes. The first number class is the class of natural numbers, which has cardinality $\aleph_0$ ('aleph-nought' or 'aleph-null'). It has to be exhausted before ω, the first number of the second number class, is created. The second number class is the set $\{\omega, \omega + 1, \omega + 2, \ldots, \}$, each element of which has cardinality $\aleph_0$. The cardinality of the second number class itself is $\aleph_1$. Cantor proved that $\aleph_1$ is the next largest cardinal after $\aleph_0$. In this terminology the continuum hypothesis becomes the conjecture that the cardinality of the continuum is $\aleph_1$ (see Continuum hypothesis §2).

2 The philosophical implications

Cantor's work on the infinite did not end with his arithmetical analysis of it. In his 1883 essay 'Grundlagen einer allgemeinen Mannigfaltigkeitslehre', he also tried to answer philosophical objections to the idea of actually existing, completed infinite classes.

One objection rested on the supposition that infinite numbers must obey exactly the same laws as finite numbers. But, as Cantor had shown, transfinite numbers have their own peculiar arithmetical laws - for example, the commutative law of addition $(a + b = b + a)$ does not hold for ordinals. This means that transfinite numbers are not finite numbers but not that they are not numbers.

Another objection related to the impossibility of completing a limiting process such as that of approaching the transcendental number π. This was more difficult to deal with. Cantor developed his counter-argument in two steps: first, he argued that the infinite as it occurs in a limit-process is, properly speaking, only a variable approaching a certain limit but still itself always remaining finite. In the second step, he pointed to the logical aspect of quantification; in particular, to the fact that the use of a variable presupposes the specification of a closed domain over which it varies. Hence, a limit-process - like a general statement concerning all numbers - presupposes logically that the infinite domain of numbers is actually and not only potentially given.

Neither argument is, however, completely convincing. With respect to the first, it turned out that the question of consistency is more difficult to resolve than Cantor had imagined. Indeed, set theory, as it was conceived by Cantor, entailed also 'inconsistent sets' and, hence, inconsistencies as Cantor himself noted in a letter to Hilbert in 1897 (see Paradoxes of set and property §1). But even when the inconsistency had been removed by Zermelo, the problem of proving the consistency of the improved set theory remained.

The second argument blurs the distinction between mathematics and logic. This results from the fact that mathematical induction - as it is used in the limit-processes in analysis - is logically analysed as a universal claim about an infinite domain, which is supposed to be completely finished. 'In the system of set theory mathematics has no specific content, it is nothing but logic, which has come to maturity' as Weyl said (1925). Indeed Cantor's argument for the existence of transfinite numbers, or, as he says, for their 'immanent reality', rests on the assumption that we can 'run through' an infinite sequence of numbers and exhaust it. But this is - epistemically speaking - impossible. It presupposes, as Hilbert has shown (1926: 174-8), certain ideal assumptions regarding the logical notions of 'all' and 'existence'.

See also: Cantor's theorem; Infinity §§6-7; Logical and mathematical terms, glossary of; Paradoxes of set and property; Set theory

ULRICH MAJER

List of works

Cantor, G. (1932) *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts (Collected Papers)*, ed. E. Zermelo, Berlin: Springer, 1980.(With a biography by A. Fraenkel.)

Cantor, G. (1874) 'Über eine Eigenschaft des Inbegriffs aller reellen algebraischen Zahlen' (On a Property of the Set of all Real Algebraic Numbers), *Crelle Journal* 77: 258-62.(The title of this paper is misleading because it describes only the first half of the content - the countability of the algebraic numbers. It refers neither to the main result - the uncountability of the open unit interval - nor to its main consequence - a new proof of the existence of 'transcendental' (that is, non-algebraic real) numbers.)

Cantor, G. (1895, 1897) 'Beiträge zur Begründung der transfiniten Mengenlehre', parts 1 and 2, *Mathematische Annalen* 46: 481-512, 49: 207-48; trans. P.E.B. Jourdain (1915), *Contributions to the Founding of the Theory of Transfinite Numbers*, New York: Dover, 1955.(Two of Cantor's most important memoirs on transfinite numbers.)

Cantor, G. (1991) *Briefe (Letters)*, ed. H. Meschkowski and W. Nilson, Berlin: Springer.(Includes Cantor's 1899 letters to Dedekind and 1897-1900 letters to Hilbert relating to the paradoxes.)

Cantor, G. and Dedekind, R. (1937) *Briefwechsel Cantor-Dedekind (Cantor-Dedekind Correspondence)*, ed. E. Noether and J. Cavaillès, Paris: Hermann; French trans. in J. Cavaillès, *Philosophie mathématique*, Paris: Hermann, 1962.(Includes most of the correspondence between Dedekind and Cantor.)

References and further reading

Dauben, J.W. (1979) *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Cambridge, MA: Harvard University Press; repr. Princeton, NJ: Princeton University Press.(An excellent intellectual biography which traces the development of Cantor's mathematical ideas, putting them in the context of his philosophy and professional life.)

Ferreiros, J. (1993) 'On the Relation between Georg Cantor and Richard Dedekind', *Historia Mathematica* 20: 343-63.(An interesting re-examination of the mutual relation between Cantor and Dedekind.)

Hallett, M. (1984) *Cantorian Set Theory and Limitation of Size*, Oxford: Clarendon Press.(A detailed study of the historical development of set theory.)

Hilbert, D. (1926) 'Über das Unendliche', *Mathematische Annalen* 95: 161-90; trans. 'On the Infinite', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.(An abridged version of an entire lecture held in the winter term 1924-5.)

Majer, U. (1993) 'Das Unendliche - eine bloße Idee?', *Revue internationale de Philosophie* 47: 319- 41.(A logico-philosophical investigation of the infinite.)

Purkert, W. and Ilgauds, H.J. (1987) *Georg Cantor 1845-1918*, Basle: Birkhäuser.(An extensive biography of Cantor's life and work.)

Weyl, H. (1925) *Die heutige Erkenntnislage in der Mathematik (The Present State of Knowledge in Mathematics)*, Erlangen: Weltkreis; repr. in *Gesammelte Abhandlungen*, vol. 2, ed. K. Chandrasekharan, Berlin: Springer, 1968.(A very condensed exposition of Weyl's point of view regarding the relation between, on the one hand, logic and set theory and, on the other, arithmetic including analysis and the continuum.)

Cantor’s theorem

Cantor’s theorem states that the cardinal number (‘size’) of the set of subsets of any set is greater than the cardinal number of the set itself. So once the existence of one infinite set has been proved, sets of ever increasing infinite cardinality can be generated. The philosophical interest of this result lies (1) in the foundational role it played in Cantor’s work, prior to the axiomatization of set theory, (2) in the similarity between its proof and arguments which lead to the set-theoretic paradoxes, and (3) in controversy between intuitionist and classical mathematicians concerning what exactly its proof proves.

Cantor’s theorem states that the cardinal number (‘size’; see Set theory §3) of the power set $\wp(A)$ of any set *A* (that is, the set of all subsets of *A*) is greater than the cardinal number of *A* itself.

One of the seemingly paradoxical properties of infinite sets is that they can be put into one-one correspondence with proper parts of themselves. For example, the function $f(n) = 2n$ maps each natural number to exactly one even number. If one-one correspondence is taken to indicate sameness of size for infinite sets, as it does for finite ones, then this means that the set of natural numbers must be said to be the same size as the set of even numbers, even though intuitively one would think that there should be twice as many natural numbers as there are even numbers. If any infinite set could be put into one-one correspondence with any other, there would be little point in trying to use ‘numbers’ as means of ‘size’ comparison for infinite sets. This is why Cantor’s proof that there are infinite sets which cannot be put into one-one correspondence with each other was crucial to the development of his theory of transfinite ordinal and cardinal numbers.

Cantor was seeking a characterization of the structure of point continua and his original result (1874) was that no denumerable sequence of elements in an interval $[a, b]$ of the real line can contain all the elements of $[a, b]$. From this he concluded that a continuum of real numbers cannot be uniquely correlated with the natural numbers. To transform this result into what has become known as Cantor’s theorem required making sense of the idea that every set, finite or infinite, has a cardinal number (sets between which there exists a one-one correspondence being deemed to have the same cardinal number); developing an arithmetic of infinite cardinal numbers; and establishing the existence of a one-one correspondence between the real numbers in any continuous interval and the set $\wp(\mathbb{N})$ of all subsets of the set $\mathbb{N}$ of natural numbers. The non-denumerability of the real numbers then becomes a special case of the more general set-theoretic result.

Russell was able to produce a contradiction in Frege’s system of logic by considering the class of all classes that do not belong to themselves (see Paradoxes of set and property §4). A similar strategy is used to prove that there would be a contradiction if there were a one-one function *f* mapping a set *A* onto its power set $\wp(A)$. Consider $d = \{x \in A\colon\ x \notin f(x)\}$. This is a subset of *A* and hence $d \in \wp(A)$. Since *f* correlates every member of $\wp(A)$ with a unique member of *A*, there is a unique element $a_d \in A$ such that $d = f(a_d)$. But both $a_d \in d$ and $a_d \notin d$ entail contradictions. Since the existence of *A* and of $\wp(A)$ are not in question, there can be no such *f* which is both one-one and onto.

In the case where *A* is finite or denumerable, *d* coincides with the diagonal set, which is defined by interchanging 0’s and 1’s down the diagonal of a square array such as the following (in which 0 indicates that a_i belongs to $f(a_i)$ and 1 that it does not).

	$f(a_1)$	$f(a_2)$	$f(a_3)$	$f(a_4)$	$\dots$
a_1	0	1	1	0	$\dots$
a_2	0	0	1	1	$\dots$
a_3	1	1	1	0	$\dots$
a_4	1	0	0	1	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

While there is general agreement that such an argument can be used to establish, for example, that for any given enumeration of the real numbers it is possible to specify a real number which is not included in that enumeration, the step to concluding from this that there exists a set whose cardinal number is greater than that of the natural

numbers is disputed by those who, like the intuitionists, do not accept the existence of actually infinite totalities (see Constructivism in mathematics; Intuitionism). Instead they may interpret the result as reflecting the difference between natural numbers and real numbers. The latter must be defined via potentially infinite sequences of rational numbers or potentially infinite decimal expansions, and so are never fully determinate objects. A function from a domain of such entities to a collection of determinate objects can never be guaranteed to be one-one. It is thus concluded that the real numbers form a potentially infinite domain of a different kind from that of the natural numbers, but not one which could be said to have a higher cardinality.

See also: Continuum hypothesis; Logical and mathematical terms, glossary of

MARY TILES

References and further reading

Cantor, G. (1895, 1897) ‘Beiträge zur Begründung der transfiniten Mengenlehre’, parts 1 and 2, *Mathematische Annalen* 46: 481-512, 49: 207-46; trans. P.E.B. Jourdain (1915), *Contributions to the Founding of the Theory of Transfinite Numbers*, New York: Dover, 1955.(Two of Cantor’s most important memoirs on transfinite numbers.)

Dauben, J.W. (1979) *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Cambridge, MA: Harvard University Press.(An excellent intellectual biography which traces the development of Cantor’s mathematical ideas, while putting them in the context of his philosophy of the infinite and his professional life.)

Dummett, M.A.E. (1978) *Truth and Other Enigmas*, London: Duckworth.(See the essays ‘The Philosophical Significance of Gödel’s Theorem’, ‘Platonism’ and ‘The Philosophical Basis of Intuitionistic Logic’.)

Wilder, R.L. (1952) *Introduction to the Foundations of Mathematics*, Krieger; 2nd edn, New York, London and Sydney: Wiley & Sons, 1965, 91-101. (Discussion of diagonal arguments.)

Wittgenstein, L.J.J. (1956) *Remarks on the Foundations of Mathematics*, ed. G.H. von Wright, R. Rhees and G.E.M. Anscombe, trans. G.E.M. Anscombe, Oxford: Blackwell, 2nd edn, 1967, part 1, appendix 2, 54-63; 3rd edn, 1978.(Includes a discussion of what exactly the proof of Cantor’s theorem proves.)

Capreolus, Johannes (*c*.1380-1444)

Thomist philosopher and theologian, Capreolus composed a lengthy commentary on Aquinas' work on Peter Lombard's Sentences, *known as Defensiones theologiae divi Thomae Aquinatis (Defences of the Theology of Thomas Aquinas) (first printed in 1483-4). He sought to refute the criticisms of Thomism by competing scholastic traditions during the fourteenth century. The Thomistic school was so impressed with Capreolus' achievement that it came to refer to him as* Princeps Thomistarum *(leader of the Thomists). Twentieth-century Thomists have, generally, considered him more faithful to the teachings of Aquinas than later commentators such as Cajetan. His philosophical opinions which have received most attention concern analogy, the formal ontological constituent of the person and the individuation of material substances.*

Capreolus was born in Rodez, France, around 1380 and entered the Dominican Order there. In 1407 he was assigned to the University of Paris as bachelor of the *Sentences*. He was licensed to graduate as master of theology in 1411. He subsequently taught in Dominican convents in Toulouse and Rodez. He died in Rodez on 6 April 1444.

Capreolus' reflections on analogy appear in his attack on Duns Scotus' affirmation that the concept of being is predicated univocally of God and creatures (see Duns Scotus, J. §5). He does not present a systematic theory of analogy but solely a tentative clarification of the analogy of being (see Language, Renaissance philosophy of §4).

He accepts as paradigmatic Aquinas' threefold division of analogy (see Aquinas §9). It is the third part of the division that deals with the analogy between God and creatures: the analogical concept is predicated of things which have neither a proper notion nor their manner of being in common. To explain it, he uses the distinction between the formal and the objective concept thereafter adopted almost unanimously by the Thomistic school. The formal concept is the subjective, mental representation of a common nature produced by the possible intellect after it has been actualized by an impressed intelligible species abstracted by the agent intellect from a phantasm. The objective concept is the extramental common nature considered precisely in so far as it is the object of an act of understanding.

Duns Scotus would be correct, Capreolus claims, if he held that it is the formal concept of being that is predicated univocally of God and creatures. But Duns Scotus errs by asserting this of the objective concept of being which, in fact, is predicated analogically. An objective concept is analogical when it is predicated of several things 'by attribution'. Predication by attribution is the essential characteristic of analogical predication. By this Capreolus means a predication which reflects an order of priority and posteriority founded on real relations of causality, dependence or correlativity.

Capreolus proposes a theory of the formal, ontological constituent of the person. The person is a complete being in the intellectual order. Being is predicated primarily of first or individual substances. The person is, then, an individual substance in the intellectual order. An individual substance does not derive the characteristics of its completeness (autonomy and incommunicability) so much from its essence as from its act of existence. The act of existence, which is the formal constituent not only of the essence but of the substance as such, must also be the formal constituent of the person. Accordingly, ontological personhood is to be distinguished from the instantiated intellectual nature in the same manner as the act of existence is distinguished from essence. Capreolus' theory has no place for the explanation of ontological personhood in terms of an added modality, as would be proposed by later commentators such as Cajetan, and is founded directly on Aquinas' real distinction between essence and the act of existence.

The problem of the individuation of material substances concerns the possibility of their solely numerical distinction and multiplication within a single species. The classic Thomistic answer is that such distinction and multiplication presupposes division, which in turn presupposes divisibility. Divisibility itself presupposes quantification, which in turn presupposes materiality. Individuation is to be explained, therefore, in terms of matter as marked (*signata*) by quantity. Divergences arise over the precise meaning of this marking (*signatio*). Capreolus' opinion is that it means the actual quantification of corporeal matter and not merely a capacity for, or some other kind of relation to, quantification on the part of prime matter. Later commentators, such as Cajetan,

preferred one of these alternative interpretations and accused Capreolus of absurdly reducing the principle of individuation to an accident of material substances. But, for Capreolus, the principle of individuation is not this actual quantification but corporeal matter in so far as it is actually quantified.

Capreolus' work was very popular during the century which followed its first printing in 1483-4. This was due to the wide range of authors against whom he argued: Duns Scotus, William of Ockham, Peter Aureol, Gregory of Rimini, Durandus of St Pourçain, Hervaeus Natalis and many others. He seems to have relied, though, on the expositions of their views which he found in Aureol's commentary on the *Sentences*. Particularly important for the diffusion of his doctrines were the many digests, concordances and indices made at this time by such authors as Paul Soncinas, Sylvester Prierias and Matthias Aquarius. Since then his thought has been generally neglected with the exception of the few, especially controversial, issues mentioned above.

See also: Aquinas, T.; Cajetan; Thomism

MICHAEL TAVUZZI

List of works

Capreolus, J. (1483-4) *Super libros IV Sententiarum (Defensiones theologiae divi Thomae Aquinatis) (Defences of the Theology of Thomas Aquinas)*, Venice; ed. C. Paban and T. Pègues, Tours: Cattier, 1900-8, 7 vols; repr. Frankfurt am Main: Minerva, 1967.(The only modern edition of Capreolus' work.)

References and further reading

Bedouelle, G., Cessario, R. and White, K. (eds) (1997) *Jean Capreolus et son temps 1380-1444 (Capreolus and His Times)*, special issue of *Mémoire Dominicaine* 1.(Includes articles on Capreolus' background, doctrine and influence.)

Bellerate, B.M. (1960) 'L'analogia tomista nei grandi commentatori di S. Tommaso' (Thomistic Analogy in the Great Commentators on Aquinas), Rome: Salesian University, 28-44.(Abstract of a Ph.D. thesis that gives the most detailed account of Capreolus' theory of analogy.)

Degl'Innocenti, U. (1940) 'Il Capreolo e la questione sulla personalità' (Capreolus and the Question on Personality), *Divus Thomas* 43: 27-40.(Account of Capreolus' theory of ontological personhood.)

Degl'Innocenti, U. (1971) *Il principio di individuazione nella Scuola Tomistica (The Principle of Individuation in the Thomistic School)*, Rome: Lateran University, 72-93.(Account of Capreolus' theory of individuation.)

Grabmann, M. (1956) 'Johannes Capreolus O.P., der *Princeps Thomistarum* (†1444), und seine Stellung in der Geschichte der Thomistenschule' (Capreolus, the Prince of the Thomists, and his place in the history of the Thomistic school), in *Mittelalterliches Geistesleben: Abhandlungen zur Geschichte der Scholastik und Mystik*, Munich: Hueber, vol. 3, 370-410.(Capreolus' influence on the Thomistic School.)

Hegyi, J. (1959) *Die Bedeutung des Seins bei den klassischen Kommentatoren des heiligen Thomas von Aquin: Capreolus, Silvester von Ferrara, Cajetan (The Meaning of Being in the Classic Commentators on Aquinas: Capreolus, Silvestri and Cajetan)*, Pullacher Philosophische Forschungen 4, Pullach bei München: Berchmanskolleg, 9-51.(Account of Capreolus' metaphysics.)

Kaeppeli, T. (1975) *Scriptores Ordinis Praedicatorum Medii Aevi (Medieval Dominican Writers)*, Rome: Sabina, vol. 2, 395-6.(Essential biographical and bibliographical data.)

Cardano, Girolamo (1501-76)

The Renaissance Italian Girolamo Cardano is famous for his colourful personality, as well as for his work in medicine and mathematics, and indeed in almost all the arts and sciences. He was an eclectic philosopher, and one of the founders of the so-called new philosophy of nature developed in the sixteenth century. He used both the Aristotelian and the Neoplatonic traditions as starting points, and following the medical paradigm of organic being, he transformed the traditional Aristotelian universe into an animated universe in which, thanks to their organic functional order, all individual parts strive towards the conservation both of themselves and of the whole universe. As a result, they can be subjected to a functional analysis. In his more casual writings on moral philosophy, Cardano showed his orientation to be basically Stoic.

1 Life

Girolamo Cardano (also known as Gerolamo Cardano or Hieronymus Cardanus) was the illegitimate son of a Milanese lawyer, and grew up in very difficult conditions. He studied medicine and mathematics at Pavia (1518-23) and Padua (1524-6). He received his MA and his MD from Padua, and practised as a physician, first at Saccolongo near Padua (1526-32) and then, after some controversy with his local colleagues, in Milan (1539-43). In 1552 he successfully cured John Hamilton, Bishop of Edinburgh, of asthma, and for the rest of his life enjoyed his reputation as a famous physician. Cardano started teaching mathematics at Milan in 1534, and medicine in Pavia from 1543. After the execution of his son for murder Cardano left Pavia, and he taught at Bologna from 1562. He was arrested by the Inquisition in 1570 for reasons which remain unknown, and was released on condition that he abandon teaching and publishing. Cardano then went to Rome where he received an annuity from the Pope, and continued to practise as a physician. He also worked on his frank and self-revelatory autobiography up to the time of his death on 21 September 1576, a date that he had himself foretold from his horoscope.

Cardano was a prolific writer. In his will of 1571 he mentions 103 printed works and forty-three in manuscript form. In his autobiography *De vita propria liber (The Book of My Life)* he counts fifty-five works in print and forty-five in manuscript. His *Opera omnia* contains seventy-one works published during his lifetime and forty other works. His style is inelegant, his Latin faulty, his argumentation often incomplete and disorganized, and as a result his writings are often unclear and difficult to understand. This does not, however, mean that he had nothing of value to say.

2 Non-philosophical achievements

Cardano was extremely creative, and a torchbearer for scientific progress. He took pride in being called *vir inventionum* (the man of inventions) and opened up new perspectives on every field he entered. In medicine, Cardano claimed to have made 40,000 important and 200,000 less important inventions. In mathematics, he taught a method for solving cubic equations, though the mathematician Tartaglia accused him of plagiarism, and he made the first moves towards the theory of probability. In mechanics 'Cardano's suspension' is still regarded as a success. He wrote on music and on dreams, believed in demons, and was regarded as a reliable astrologer, who even cast the horoscope of Jesus Christ.

3 Metaphysics

Cardano claimed that Plotinus and Aristotle were his main inspiration in philosophy, and though he argued explicitly against Aristotle, his theory is implicitly based on the Aristotelian tradition. While he was not a systematic thinker, his work rests on some basic assumptions which allow for a systematic understanding of his main philosophical teaching, and which are themselves of major philosophic interest.

His notion of unity is fundamental. It was developed in a little treatise *De uno (On the One)*, written around 1560 and published in Basle in 1562. Cardano himself recommended it as an introduction to his natural philosophy. In it he argued that everything that exists is one, so that the structure of unity is identical to the structure of reality. Analysis of a given real being, a human being for example, shows that its unity consists both in the single principle behind its various operations and in the organic structure of its corporeal basis, which guarantees that its various parts cooperate in the effort to achieve self-conservation. Thus every real being can properly be regarded as a system constituted by a certain number of organic parts that cooperate in function. Every organ, in so far as it is a

real being and therefore one, has necessarily to be such a system; and every real being, as part of a greater, more comprehensive, whole, has to function as an organ of this larger unit. As a result, reality is reduced to a system of functions, in which every individual being is at the same time an organ or subsystem of its supersystem, and the superior unit of its subsystems. This kind of ordering is also to be found in the sphere of history, which is governed by fate.

4 Physics

In Cardano's treatise *De natura (On Nature)*, written at the same time as *De uno* (though not published in his lifetime) and like *De uno* recommended as introductory reading, Cardano developed his concept of nature in accordance with his basic ontological principle of unity. The single principle which coordinates the various operations of natural beings and causes the organic order of their unity is called the soul. As a result the whole of the material world is animated, and Aristotle's distinction between organic and non-organic nature is abandoned. Different souls are systematically ordered in the same way as the different natural beings which they animate. Their principal task is to care for the self-preservation of the bodies which correspond to them, so they are endowed with whatever means of cognition are necessary for that task. They act in accordance with the principles of sympathy and antipathy. The intellectual soul, being supernatural, is one and the same for all human beings, as Averroes had held (see Ibn Rushd §3).

Bodies themselves are constituted by the heat of the heavens. This heat is the active principle that serves the soul, and operates on the passive principle's density or extension. In *De subtilitate (On Subtlety)* Cardano identifies this passive principle with prime matter (*prima materia*) which, according to Averroes, is endowed with indefinite quantity; in *De natura* the passive principle is understood as moisture. Thus Cardano reduces and transforms Aristotle's doctrine of the four primary qualities (hot, cold, wet and dry) that constitute the four elements.

5 Epistemology

Natural beings are perceived as organic units through the senses, and it is the task of science, through a functional analysis, to transform this confused knowledge of the whole into a distinct knowledge of its parts and their ordering in a system. Thanks to the infinity of the universe's various parts, human beings will never be able to analyse and thereby know the whole of the universe. Moreover, thanks to the infinite subtlety of the smallest parts, which escape the senses, these too will necessarily remain unknown. As a result, as Cardano points out in *De arcanis aeternitatis (On the Secrets of Eternity)*, it is the greatest and indeed most disastrous error to claim that humans can achieve absolute knowledge. Nonetheless, there are organic units which are not disproportionate to the human intellect in terms of the variety of their parts and the subtlety of their structure, and these can be the subject of scientific analysis. Moreover, they enjoy ontological autonomy as far as their place in the system's order is concerned, and hence can be known, if not according to their essence, at least according to their functional structure. It is this knowledge of the functional structure of natural beings that, if soundly based, can be used in technical or mechanical inventions.

6 Main works in natural philosophy

It is in the context of the principles discussed in §§3-5 that Cardano's main works in natural philosophy, the twenty-one books of *De subtilitate* from 1550, and their supplement, the seventeen books of *De rerum varietate (On the Variety of Beings)* written in 1557, should be read. While their titles refer to the aspects of reality that constitute the limits of human knowledge, the books themselves offer a comprehensive description of the whole universe. Following the order of the traditional Aristotelian doctrine of nature, they start with the principles of physics and descend to the elements, the mixed bodies, metals and stones, plants, animals, and human beings; then they ascend again from the psychological faculties of the senses, the will and the intellect, through their effects, the arts and sciences, to the demons, first substances, and God himself. This apparently Aristotelian order led Julius Caesar Scaliger to criticize Cardano from the perspective of orthodox Aristotelianism; the fact that, for no apparent reason, Cardano mixed the theoretical with the practical and the general with the particular does make it difficult to follow his arguments. However, contrary to Scaliger's belief, Cardano did not intend to teach Aristotelian natural philosophy. Nor did he, contrary to modern expectations, intend to display a comprehensive and consistent theory of nature. His way of writing and arguing must be understood in the light of his denial of the possibility of absolute truth and his doubts about the fruitfulness of merely theoretical argumentation. He was content to offer a

functional analysis of individual natural beings through sense perception, and to validate his analysis through technical application.

See also: Aristotelianism, Renaissance

ECKHARD KESSLER

List of works

Cardano, G. (*c*.1540-76) *Opera omnia*, ed. C. Spohn, Lyons, 1663, 10 vols; repr. New York and London: Johnson, 1967.(Only comprehensive edition of Cardano's works.)

Cardano, G. (*c*.1540) *De arcanis aeternitatis (On the Secrets of Eternity)*, in C. Spohn (ed.) *Opera omnia*, New York and London: Johnson, 1967, vol. 10, 1-49.(Incomplete; demonstrates the limits of human knowledge.)

Cardano, G. (*c*.1541) *Liber de ludo aleae*, in C. Spohn (ed.) *Opera omnia*, New York and London: Johnson, 1967, vol. 1, 262-82; trans. S.H. Gould, 'The Book on Games of Chance', in O. Ore, *Cardano the Gambling Scholar*, Princeton, NJ: Princeton University Press, 1953; repr. New York: Dover, 1965.(Cardano's theory of probability.)

Cardano, G. (1545) *Artis magnae sive de regulis algebraicis liber unus*, Nuremberg: Ioannes Petreius; ed. and trans. T.R. Witner, *The Great Art, or the Rules of Algebra*, Cambridge, MA: MIT Press, 1968.(Cardano's most important and influential algebra manual.)

Cardano, G. (1550) *De subtilitate (On Subtlety)*, Nuremberg: Ioannes Petreius; in C. Spohn (ed.) *Opera omnia*, New York and London: Johnson, 1967, vol. 3, 357-672; M.M. Cass (ed. and trans.) *The first book of Jerome Cardan's De subtilitate, Translated from the Original Latin with Text, Introduction and Commentary*, Williamsport, PA: Bayard Press, 1934.(Cardano's major work in natural philosophy.)

Cardano, G. (1557) *De rerum varietate (On the Variety of Beings)*, Basle: Henricus Petri; in C. Spohn (ed.) *Opera omnia*, New York and London: Johnson, 1967, vol. 2, 1-356.(Supplement to Cardano's *De subtilitate*; written in 1557.)

Cardano, G. (*c*.1560) *De natura (On Nature)*, in C. Spohn (ed.) *Opera omnia*, New York and London: Johnson, 1967, vol. 2, 283-98.(Written around 1560, *De natura* was only published posthumously.)

Cardano, G. (1562) *De uno (On the One)*, Basle; in C. Spohn (ed.) *Opera omnia*, New York and London: Johnson, 1967, vol. 1, 277-83.(Written around 1560, but not published until 1562; develops Cardano's concept of what constitutes the 'oneness' of natural beings.)

Cardano, G. (before 1576) *De vita propria liber*, in C. Spohn (ed.) *Opera omnia*, New York and London: Johnson, 1967, vol. 1, 1-54; trans. J. Stoner, *The Book of My Life*, New York: Dover, 1962.(Cardano's autobiography; he was working on this work up to his death in 1576.)

References and further reading

Eckman, J. (1946) 'Jerome Cardano', *Bulletin of the History of Medicine*, supplementary vol. 7, Baltimore, MD: Johns Hopkins University Press.(Discussion of Cardano's life and fortunes, with extensive bibliography in the notes.)

Fierz, M. (1977) *Girolamo Cardano (1501-1576)*, Basle and Stuttgart: Birkhäuser.(Discussion of Cardano's achievements as philosopher and scientist.)

Ingegno, A. (1980) *Saggio sulla filosofia di Cardano (Essay on the Philosophy of Cardano)*, Florence: La Nuova Italia.(Describes the different aspects of Cardano's philosophy; bibliography on pages 21-31.)

Kessler, E. (ed.) (1994) *Girolamo Cardano. Philosoph, Arzt, Mathematiker (Philosopher, Physician, Mathematician)*, Wiesbaden: Harrassowitz.(Contributions to the international Cardano Colloquium, held in Wolfenbüttel in 1989; covers the different aspects of Cardano's thought.)

Margolin, J.-C. (1976) 'Cardan, interprète d'Aristote' (Cardano as Interpreter of Aristotle), in *Platon et Aristote à la Renaissance. XVIe Colloque international de Tours*, Paris: Vrin, 307-33.(Places special emphasis on moral philosophy and the immortality of the soul.)

Carlyle, Thomas (1795-1881)

Although widely influential as a historian, moralist and social critic, Carlyle has no real claim to be considered a philosopher. He does have some importance as one of the transmitters of the ideas of the German Idealists, such as Kant and Fichte, to Britain, and as one of the chief British spokesmen for the Romantic exaltation of the imagination above the understanding; but his grasp of philosophical issues is vague. His later writings are dominated by the idea, derived from his childhood Calvinism, of a divine justice working in history through the medium of great men ('heroes') who are its conscious or unconscious instruments.

Carlyle was born in Ecclefechan, Scotland, the eldest son of James Carlyle, a farmer and stonemason, and his wife Margaret. Educated at Annan Academy and Edinburgh University, he eked out a meagre living by private tutoring and journalism until leaving Scotland and moving to London in 1834. His first major work, *Sartor Resartus*, was published in 1833: the most significant of his later writings were *The French Revolution* (1837), *Past and Present* (1843), *Letters and Speeches of Oliver Cromwell* (1845) and *The History of Frederick the Great* (1858-65).

Carlyle is important as a literary figure, a moralist, a historian and a social critic, but has little serious claim to be considered a philosopher, though he was sometimes regarded as such by his contemporaries. His chief interest in this regard is as one of the transmitters of German Idealism to Britain in the 1820s. The tradition of Scottish Calvinism in which he had been brought up had been shaken by his encounter with the ideas of the Enlightenment at Edinburgh University, and he found in the German Idealist tradition a resolution of the problems this caused him. His enthusiasm for this tradition was expressed in a series of review articles from the 1820s which first brought his name before the public - notably on the 'State of German Literature' (1827) and 'Novalis' (1829) (see also 'Voltaire' (1829) and 'Diderot' (1833) for his repudiation of the rationalism of the Enlightenment).

In these articles he refers enthusiastically to the work of Fichte and Kant, insisting particularly on the characteristic German Idealist distinction between *Vernunft* ('Reason') and *Verstand* ('Understanding'), and exalting 'Reason', understood as comprehending feeling and imagination, at the expense of the empirical and logical 'Understanding'. However, Carlyle was much more interested in the literature of the German renaissance than in its philosophy (it is significant that he never wrote specifically on Kant or Fichte) - in particular, he idolized Goethe - and his understanding of the strictly philosophical issues at stake was limited. He is accordingly less important as a transmitter of German Idealist philosophy than Coleridge - whom, in his later years, Carlyle knew, but whom he revealingly rejected, lacking patience for the fineness of Coleridge's distinctions and what he saw as his intellectual indecisiveness.

The fullest expression of Carlyle's philosophical position at this stage of his life is found in *Sartor Resartus*, the only one of his books which makes any claim to philosophical status. Its form - an examination of the life and writings of an imaginary 'Professor Teufelsdrockh' - is maverick, but the influence of German Idealism is manifest. The title ('The Tailor Re-Tailored') refers to Teufelsdrockh's supposed 'Philosophy of Clothes', in which clothes become the metaphor for the material world which simultaneously expresses and conceals the transcendental world within. This material world is conditioned by the categories of time and space, but these are themselves illusions, concealing the realities of eternity and infinity. These realities cannot be apprehended rationally, but only intuitively by the 'deep infinite faculties' of 'Fantasy and Heart'. In this plea for the imaginative faculties (anticipated in two important earlier essays, 'Signs of the Times' (1829) and 'Characteristics' (1831)) can be discerned one of the major manifestos of Romanticism, then at its height as a literary movement. The inspired artist - the ideal is Goethe - is seen by Carlyle as the hero who alone can discern and make clear to his readers the spiritual realities which underlie their material garments (the origin of the idea of 'hero worship', which becomes increasingly dominant in Carlyle's later work).

However, after *Sartor* Carlyle's concerns turned increasingly away from philosophy towards history and social criticism (the categories to which all his later writings listed above belong). His friends included both John Stuart Mill and Ralph Waldo Emerson; but he passionately rejected both Mill's utilitarianism (which he characterized as 'Pig Philosophy') and his political liberalism, and showed little more enthusiasm for Emerson's transcendentalism. In his later writings, of which *The French Revolution* is the most remarkable, the influence of his earlier Calvinist heritage becomes increasingly apparent, in his insistence that the events of history can be properly understood only

when they are seen as the workings of a vaguely-conceived divine providence which ensures the final triumph of justice in all the processes of history. This providence is purely moral - Carlyle does not see history as progressive (and seems not to have read Hegel). This sense of a providence working through history forms an approximate parallel to his earlier insistence on a transcendental order which can only be discerned through the material, but it is certainly not identical with it: it is moreover weakened by Carlyle's inability to provide any firm independent criterion of justice other than success, so that the argument becomes circular, and his biographies of Cromwell and Frederick the Great, especially the latter, are hard to distinguish from encomiums of *Realpolitik*.

A lack of close argument and clear definition pervades all Carlyle's work, and must undermine any serious claim for him as a philosopher. This very incoherence, however, enabled him to give forceful and sometimes memorable expression to a wide range of antirationalist and anti-empiricist attitudes characteristic of the Romantic reaction against the Enlightenment - a range that also includes a recurrent insistence on moral authenticity that sometimes anticipates some existentialist positions. His real significance, however, lies in other fields.

See also: German idealism

A.L. LE QUESNE

List of Works

Carlyle, T. (1833-4) *Sartor Resartus*, ed. C.F. Harrold, New York: Doubleday, 1937. (Carlyle's only major work of philosophical interest.)

Carlyle, T. (1896) *The French Revolution*, London: Chapman and Hall, Centenary edition. (Generally recognized as the greatest of Carlyle's writings, this is interesting for its underlying philosophy of history, strongly influenced by German Idealism.)

Carlyle, T. (1896-9) *Critical and Miscellaneous Essays*, London: Chapman and Hall.(This Centenary edition contains 'Voltaire' and 'Diderot', marking Carlyle's repudiation of the Enlightenment; 'State of German Literature' and 'Novalis', summarizing his understanding of the German Idealist tradition; and 'Signs of the Times' and 'Characteristics', rejecting mechanical and utilitarian modes of thought in favour of imagination and faith.)

Carlyle, T. (1898) *Latter-Day Pamphlets*, London: Chapman and Hall, Centenary edition. (Interesting for the satirical attack on utilitarianism in the last chapter.)

References and further reading

Harris, K.M. (1986) 'Reason and Understanding Reconsidered: Coleridge, Carlyle, and Emerson', in *Essays in Literature* 13: 263-81, Macomb, IL: Western Illinois University.(Discusses Carlyle's understanding of Kant's distinction.)

Harrold, C.F. (1934) *Carlyle and German Thought*, Yale, NH: Yale University Press. (The only substantial treatment of the topic.)

Kaplan, F. (1983) *Thomas Carlyle*, Cambridge: Cambridge University Press.(The most recent authoritative biography.)

Tennyson, G.B. (1965) *Sartor Called Resartus*, Princeton, NJ: Princeton University Press.(The fullest study of *Sartor Resartus*.)

Carmichael, Gershom (1672-1729)

Gershom Carmichael was a teacher and writer of pivotal importance for the Scottish Enlightenment of the eighteenth century. He was the first Professor of Moral Philosophy at the University of Glasgow, predecessor of Francis Hutcheson, Adam Smith and Thomas Reid. Carmichael introduced the natural law tradition of Grotius, Pufendorf and Locke to the moral philosophy courses he taught at the University of Glasgow (1694-1729). His commentaries on Samuel Pufendorf's work on the duty of man and citizen (1718 and 1724) made his teaching available to a wider readership in Great Britain and in Europe. He also composed an introduction to logic, Breviuscula Introductio ad Logicam, (1720 and 1722) and a brief system of natural theology, Synopsis Theologiae Naturalis (1729).

Gershom Carmichael began his teaching career at St Andrews University in 1693; he was appointed to the University of Glasgow in 1694, and became the first Professor of Moral Philosophy at that university in 1727. His moral philosophy is remarkable particularly for the manner in which he justified the natural rights of individuals. Those rights included self-defence, property and the natural right to services contracted for with others. He also argued that slavery is incompatible with the rights of men and citizens, and that subjects have a right to resist rulers who exceed the limits of their powers. His arguments for natural rights derived from his understanding of the divine or natural law that lasting happiness or beatitude may be found only in reverence for God. One may signify reverence or veneration directly; but reverence for God may also be signified indirectly, by reverence or respect for his creatures, and specifically by respect for mankind. The most evident manner in which the latter may be signified is by acknowledging the natural rights of all human beings. By such acknowledgement one signifies observance of the divine or natural law, and one finds, in such observance, beatitude or lasting happiness.

Carmichael's theory that rights derive from a universal longing for beatitude was consistent with the Reformed scholastic theology that was taught in Scottish universities in the late seventeenth and early eighteenth centuries. His theory of moral motivation presented a problem, however, for his successors. Was it possible for an action or a character to be moral or right or good if it was inspired by some motive other than veneration of God? Hutcheson, Smith and Reid found in benevolence, in sympathy with others and in conscience, moral inspiration which they judged to be more consistent with experience than Carmichael's moral psychology. But, in their disagreements with Carmichael and with one another, they generated those fruitful speculations concerning the moral life of mankind that we have come to call the Scottish Enlightenment (see Enlightenment, Scottish). Carmichael's importance as a moral philosopher turns, then, not only upon his distinctive contribution to the natural rights tradition of Grotius, Pufendorf and Locke; nor upon his redirection of Reformed or Presbyterian scholastic ethics to speculation on the divine inspiration for human rights and obligations; it was, above all, perhaps the manner in which he helped to shape the agenda of moral philosophy in eighteenth-century Scotland that makes him a figure worthy of attention in the history of philosophy.

See also: Natural law; Rights

JAMES MOORE
MICHAEL SILVERTHORNE

List of works

Carmichael, G. (1699) *Theses Philosophicae ...Sub Praesidio Gerschomi Carmichael, P.P., Glasguae MDCXIX (Philosophical Theses [defended] under the presidency of Gershom Carmichael, Professor of Philosophy)*, Glasgow.(An early statement of Carmichael's ideas of thinking and acting which reflects the influence of the Port-Royal logic, Malebranche and Locke.)

(1707) *Theses Philosophicae ...Sub Praesidio Gerschomi Carmichael, P.P., Glasguae MDCCVII (Philosophical Theses [defended] under the presidency of Gershom Carmichael, Professor of Philosophy)*, Glasgow.(An early version of Carmichael's theory of natural law).

(1718) *Supplementa et Observationes ad C[larissimi] V[iri] Sam[uelis] Pufendorfii. Libros Duos De Officio Hominis et Civis, Glasguae MDCCXVIII (Supplements and Observations upon the distinguished Samuel Pufendorf's two books On the Duty of Man and Citizen)*, Glasgow.(The first edition.)

(1720) *Breviuscula Introductio ad Logicam, Glasguae MDCCXX (A Short Introduction to Logic)*, Glasgow; 2nd edn, (Edinburgi MDCCXXII) Edinburgh, 1722.(An adaptation of the Port-Royal logic for students in British

universities.)

(1724) *S[amuelis] Pufendorfii De Officio Hominis et Civis, Juxta Legem Naturalem, Libri Duo. Supplementis et Observationibus in Academicae Juventutis auxit et illustravit Gerschomus Carmichael, Philosophiae in Academia Glasguensi Professor. Editio Secunda priore Auctior et Emendatior. Edinburgi MDCCXXIV (Supplements and Observations upon Samuel Pufendorf's On the Duty of Man and Citizen according to the Law of Nature composed for the use of Students in the Universities by Gershom Carmichael, Professor of Philosophy in the University of Glasgow)*, Edinburgh.(This is a second edition of the 1718 work with additions and amendments: it includes the text of Pufendorf's work and provides many observations not included in the edition of 1718.)

(1729) *Synopsis Theologiae Naturalis, sive Notitiae, De Existentia, Attributis et Operationibus, Summi Numinis, ex ipsa rerum Natura, Studiosae Juventutis usibus accomodata. Edinburgi MDCCXXIX (Synopsis of Natural Theology, or Considerations upon the Existence, Attributes and Operations of the Divine Will, based upon the nature of things themselves, adapted for the use of young men in their studies)*, Edinburgh(A brief system of natural theology in the Reformed scholastic tradition.)

References and further reading

Hutcheson, F. (1747) *A Short Introduction to Moral Philosophy in three books, containing the Elements of Ethicks and the Law of Nature*, Glasgow.(Acknowledges his indebtedness to Carmichael's work in the introduction and in notes to the text.)

Moore, J. and Silverthorne, M. (1983) 'Gershom Carmichael and the natural jurisprudence tradition in eighteenth-century Scotland', in I. Hont and M. Ignatieff (eds) *Wealth and Virtue: the Shaping of Political Economy in the Scottish Enlightenment*, Cambridge: Cambridge University Press, 73-87.(Describes how John Locke's ideas of property and the original contract were communicated to Scottish students of moral philosophy.)

(1984) 'Natural Sociability and Natural Rights in the Moral Philosophy of Gershom Carmichael', in V. Hope (ed.) *Philosophers of the Scottish Enlightenment*, Edinburgh: Edinburgh University Press, 1-12.(Contrasts Carmichael's understanding of moral ideas with the ideas of Pufendorf.)

Reid, T. (1990) *Practical Ethics: Being Lectures and Papers on Natural Religion, Self-Government, Natural Jurisprudence and the Law of Nations*, ed. K. Haakonssen, Princeton, NJ: Princeton University Press, 1990. (Edited from the manuscripts with an introduction and commentary. The commentary explores the relevance of Carmichael's natural rights theories and other juridical theories for Thomas Reid's lectures on jurisprudence.)

Stein, P. (1982) 'From Pufendorf to Adam Smith: the natural law tradition in Scotland', in N. Horn (ed.) *Europäisches Rechtsdenken in Geschichte und Gegenwart (European Legal Thought in History and at the Present Time)*, Munich: C.H. Beck'sche Verlagsbuchhandlung, 667-79.(Examines continuities and differences in Scottish moral and legal philosophy from Carmichael to Adam Smith.)

Veitch, J. (1837) 'Philosophy in Scottish Universities', *Mind* 2: 74-91 and 207-34.(Remains a valuable account of the teaching of philosophy in early modern Scottish universities.)

Carnap, Rudolf (1891-1970)

Carnap was one of the most significant philosophers of the twentieth century, and made important contributions to logic, philosophy of science, semantics, modal theory and probability. Viewed as an enfant terrible *when he achieved fame in the Vienna Circle in the 1930s, Carnap is more accurately seen as one who held together its widely varying viewpoints as a coherent movement. In the 1930s he developed a daring pragmatic conventionalism according to which many traditional philosophical disputes are viewed as the expression of different linguistic frameworks, not genuine disagreements. This distinction between a language (framework) and what can be said within it was central to Carnap's philosophy, reconciling the apparently a priori domains such as logic and mathematics with a thoroughgoing empiricism: basic logical and mathematical commitments partially constitute the choice of language. There is no uniquely correct choice among alternative logics or foundations for mathematics; it is a question of practical expedience, not truth. Thereafter, the logic and mathematics may be taken as true in virtue of that language. The remaining substantive questions, those not settled by the language alone, should be addressed only by empirical means. There is no other source of news. Beyond pure logic and mathematics, Carnap's approach recognized within the sciences commitments aptly called a priori - those not tested straightforwardly by observable evidence, but, rather, presupposed in the gathering and manipulation of evidence. This a priori, too, is relativized to a framework and thus comports well with empiricism. The appropriate attitude towards alternative frameworks would be tolerance, and the appropriate mode of philosophizing the patient task of explicating and working out in detail the consequences of adopting this or that framework. While Carnap worked at this tirelessly and remained tolerant of alternative frameworks, his tolerance was not much imitated nor were his principles well understood and adopted. By the time of his death, philosophers were widely rejecting what they saw as logical empiricism, though often both their arguments and the views offered as improvements had been pioneered by Carnap and his associates. By his centenary, however, there emerged a new and fuller understanding of his ideas and of their importance for twentieth-century philosophy.*

1 *Aufbau*

The last of twelve children, Carnap was born in Ronsdorf, a small village now incorporated into Wuppertal, Germany. His father, who died when Carnap was seven, was born poor but became a prosperous factory owner. Carnap's mother (née Dörpfeld) was from a more academic family. Her father was a well-known educational reformer, and her oldest brother, Wilhelm, was a renowned archaeologist who, along with Heinrich Schliemann, discovered the remains of Troy. Carnap's undergraduate studies at Freiburg and Jena were in physics, mathematics and philosophy, and Gottlob Frege was among his teachers at Jena. While serving in the First World War he became familiar with relativity theory, and in 1919 with Whitehead and Russell's *Principia Mathematica* (1910-13) (see Russell, B.; Whitehead, A.N.). His doctoral dissertation in Jena, *Der Raum: Ein Beitrag zur Wissenschaftslehre (Space: A Contribution to the Theory of Science)* (1922), combined his undergraduate interests (as his career was to do). Anticipating his later principle of tolerance, this work argues that apparent disagreements among physicists, geometers and philosophers arise from the fact that they articulate wholly different concepts: physical, mathematical and visual space. The section on the last has a character reminiscent of Kant which he later repudiated, though some Kantian features continued to shape his work (see Kant, I. §5). This was especially true in the 1920s, during which he published papers on space, time and concept formation, as well as *Der logische Aufbau der Welt (The Logical Structure of the World)* (1928) and a textbook in mathematical logic. In 1923 he became friends with Hans Reichenbach, who was then also a Neo-Kantian. In 1926, at the invitation of Moritz Schlick, Carnap became a *Privatdozent* at the University of Vienna, where an early version of the *Aufbau* was his *Habilitationsschrift*.

The *Aufbau* established his international reputation. Taken at face value, it is a system of definitions. He presents the first stages, constituting the objects of one's own experience, with great formal rigour, and the later stages carefully but in outline form. Finally he discusses the philosophic consequences of such a system. The *Aufbau* assumes a single undefined concept, partial, remembered similarity, ranging over time-points in the stream of experience. On that basis Carnap sets out to define (or constitute) every other concept and object in the whole system of science. The resulting system is constitution theory. Similarity is used to define properties, for every object having a given property must be similar to every other object having the same property. The partialness of

the basic relation is exploited to order properties into families, and the 'rememberedness' of the basic relation is used to achieve a temporal ordering of experiences. Later, physical objects were to be defined, then other minds, and finally cultural objects.

Though plainly inspired by Mach and Russell, just what Carnap hoped to achieve by the *Aufbau* is unclear. One interpretation holds that the enterprise is straightforward ontological reduction to sensory phenomena, but Carnap seems not to have meant 'reduction' as understood by English-speaking philosophers. He shows complete indifference to such ontological issues and draws from his constitution theory a moral of metaphysical neutrality. Moreover, he says that a different basic relation or even a physical basis would have served his purposes equally well, and he calls his definitions 'epistemic analyses', not ontological ones. Carnap intends not to reproduce the actual processes of concept formation, but rather to give a rational reconstruction of them. This, and the fact that he starts, like Descartes, with one's own experiences and tries to range the various concepts and claims in a (rational) epistemic order, has suggested to some that Carnap is trying to justify the claims of science by providing them with a philosophical foundation. Against this, Carnap seems to take the results of contemporary science completely for granted; he claims no synonymy between defined and defining concepts, but only extensional equivalence in ways that presuppose rather than defend the science in question. This, surely, would undermine such a foundational programme. He denies that philosophers can make the sciences more certain. One intriguing interpretation developed by Michael Friedman (1992) among others, is that Carnap was undertaking the task of showing that intersubjectively applicable concepts can be developed, even though we each begin with our own subjective experience (see Foundationalism).

Whatever the interpretation, are even those portions of the system that are fully worked out satisfactory? Notably, Nelson Goodman (1951) has argued that though each member of a 'property class' is indeed similar to each other, the class may fail to pick out a property because its members are similar in different respects. The seriousness of this problem will depend on the remedy and on how it serves Carnap's purposes.

2 Vienna

Moritz Schlick, named professor of inductive sciences in 1922, drew Carnap, Otto Neurath, Hans Hahn and others as members of a new Vienna Circle, active in discussing scientific methodology (see Vienna Circle). With such figures as Wittgenstein, Popper and Gödel in the neighbourhood, Vienna in the late 1920s and early 1930s was lively indeed. The Circle combined serious interests in empirical science, its methods, and their logical analysis. Schlick had a doctorate in physics (under Max Planck); Carnap's speciality was logic, foundations of mathematics, and physics; Neurath was a sociologist, Hahn a mathematician, and others included economists and psychologists. The interest in scientific methodology was inspired largely by the writings of scientists: Mach, Duhem, Poincaré, Hilbert and Einstein. Members were uninterested or openly hostile to Cartesian methodological discussions, and they knew little of the British empirical philosophers, such as Berkeley, Hume and Mill. Russell was no ordinary British empiricist, and Viennese interest in him was very selective.

The nineteenth and early twentieth centuries witnessed major scientific and mathematical developments with implications for Kantian philosophy. First came non-Euclidean geometries which encouraged a distinction between physical geometry, which was synthetic, and mathematical geometry, which was a priori (see Geometry, philosophical issues in). Hilbert dealt with the latter by deploying highly formal analyses and a theory of implicit definition. With the advent of Einstein's relativity theory, non-Euclidean structures invaded even physical geometry (see Relativity theory, philosophical significance of). And Schlick was among the first philosophers to discuss and embrace the new physics. These developments did not refute the Kantians, but threw them into disarray. Rather than giving no answer, they had too many. Other nineteenth-century mathematical developments diminished the role for Kantian intuition. Finally, Frege and Russell developed modern symbolic logic and succeeded in deriving all of mathematics from it (and set theory) (see Logicism). In this they were inspired by the fear of both psychologism and empiricism in mathematics. Technically, their achievement only transferred mathematics from one Kantian classification (synthetic) to another (analytic) without saying much about what the analytic is. Russell himself seemed to prefer some sort of Platonistic account, and Frege preferred silence. Nevertheless, these achievements revealed how meagre and inadequate were Kant's discussions of the analysis of concepts. Both the geometric and arithmetic developments undermined the foundations of Kant's account of the synthetic a priori and showed the need for much more careful and systematic formal analyses. On neither front did

the nineteenth-century empiricist tradition seem to have much to say that was useful.

Members of the Circle were empiricists, but of what sort? Schlick and Carnap were initially Neo-Kantian empiricists, and Neurath seems to have come from an Aristotelian empiricist tradition (see Aristotle §9). Clearly they were not trying to reprise phenomenalism nor to substitute philosophic arguments and speculation for scientific evidence and methods. Perhaps what held them together was their conviction that science is not subservient to philosophy. Philosophers have no special source of news, and when they try to go beyond science or neglect it altogether they simply talk rubbish. There was a long tradition of calling this habit metaphysics, and deploring it (even, and perhaps especially, within the Kantian tradition). The result was the famous verification principle: that meaning is the mode of verification (or better, of confirmation) and that no claim is meaningful unless it can be confirmed or disconfirmed, either by logical (analytic) or empirical means. The verification principle itself is not an empirical claim, sociological or otherwise, about what claims people find meaningful. It is rather a proposal about how we should structure our language. If such a proposal were adopted, the principle would be an analytic claim in the metalanguage. In neither case would the verification principle be self-defeating (see Meaning and verification).

The Vienna Circle embraced a variety of evolving views. Among the most important cleavages was that between the ‘right wing’, led by Schlick and Frederick Waissmann and influenced strongly by Wittgenstein, and the ‘left wing’, led by Carnap and Neurath. One plank in the platform of the left was to liberalize the criterion of meaningfulness; a second was antifoundationalism. They argued that a physicalist language was sufficient for and conducive to the purposes of science, including its most basic observational evidence. They also embraced a confirmation holism and a thoroughgoing fallibilism about both observation reports and theories, and insisted that logic and mathematics were revisable (see Fallibilism). The left wing also emphasized the practical, especially the use of science as part of a programme of social reform. This practical emphasis eventuated in Carnap’s intriguing suggestion that empiricism itself was a proposal to be adopted on pragmatic grounds. Carnap’s version of the unity of science was that the language of science should be unified in having all of its concepts appropriately tied to a common, publicly testable, observational basis. Regarding the unity of the laws of science, he urged that while philosophers might have something to say about what the reduction of some laws to others might mean, we should avoid speculation and wait on experience to see whether that sort of unity really obtained. Such unity is neither foundationalist nor phenomenalist (see Logical positivism).

3 Syntax

Carnap’s most impressive undertaking of the 1930s was the development of a metalanguage that was adequate for discussing logic and the foundations of mathematics. In the process he worked out a new pragmatic and conventionalist epistemology with consequences for the character of observation and the interpretation of scientific theories. In the background were disputes among logicists, formalists and intuitionists over the foundations of mathematics, and between classicists and intuitionists over the character of logic. While Wittgenstein seemed to be saying that one could not talk about logical form, only show it, Carnap tried to develop a way of talking about logic in which its structure or form could be expressed. If the original language were suitably restricted, its structure could be described, even within that language itself, thanks to Gödel’s device of arithmetization. Other object languages could be described in this metalanguage, and still more powerful object languages could be described in distinct metalanguages. This required a choice which could be settled neither by empirical means, nor, on pain of circularity, by logical ones. The issue resembled that which had faced geometers at the end of the nineteenth century, and Carnap met the new problem as Hilbert had met the old, with a version of a theory of implicit definition. The alternative logical structures are conventions defining the terms they contain, and we should tolerate alternatives. Hence, Carnap’s famous principle of tolerance: ‘In logic there are no morals’, meaning roughly that there is no uniquely correct logic. The alternatives can be described, though not always in the language itself, and their consequences investigated and compared. Some may be more useful (more convenient, simpler, more powerful), and so a choice might be made on these grounds, but not because one of them is the right one. This is the central doctrine of *Logische Syntax der Sprache (The Logical Syntax of Language)* (1934a). The English translation included several sections containing a virtual semantics originally omitted for lack of space.

The doctrine was illustrated by presenting two languages. Language I is a coordinate language (using symbols of position rather than names of objects) and is highly restrictive, allowing, either in its formulas or in the definitions

of terms, only quantifiers that can be replaced with finite conjunctions or finite disjunctions. Carnap thought this the language preferred by intuitionists, though they never entirely agreed. Language II is classical in using unlimited quantifiers and is sufficient for expressing classical mathematics. Gödel had shown that no language this rich could be both consistent and complete in the sense that every sentence or its negation would be derivable from the axioms. Carnap's striking innovation in Language II is to develop a consequence relation much more powerful than the traditional idea of derivation, and to show that mathematics could be complete in this new sense.

Most logical discussions, including those of Languages I and II, begin with lists of grammatical categories and of terms therein, and go on to specify what shall count as sentences, derivations, proofs and so on. But this provides no general terms which are useful for the discussion of other languages. Part IV of *Logical Syntax* provides definitions of grammatical categories and everything else to discuss and compare alternative languages, including translation. This is done with just two primitive terms: 'is a sentence' and 'is a direct consequence of'. The definitions do not always work, but, even so, are an amazing achievement. He tried to distinguish the logical from the descriptive vocabulary and specifically allowed rules of inference which were not wholly logical in character. He also developed a general notion of analyticity, which he later modified. This discussion was essentially of a sort later called semantical, and in limited ways surpassed even Tarski's concurrent discussions (see Semantics; Tarski, A.). Carnap denied only that a syntactical discussion of truth was possible and, hence, that a truth theory was part of logic. Unsurprisingly, therefore, he soon changed his mind and embraced semantics. He and Tarski then viewed the new truth theory as lying within what Carnap meant by 'syntax'.

Characteristically, the last part of the book is devoted to the philosophical significance of the rest. Having discussed standards of translation, Carnap notes that many claims which seem to be about extra-linguistic objects can be translated into sentences about only words, sentences or other linguistic items. Before such translation, these sentences are said to be in material mode; after, in formal mode. Speaking in the material mode is fine, but it can disguise the implicit relativization of such claims to specific languages. This fosters the illusion that there is a uniquely correct answer to the unrelativized questions, and hence that there are genuine disagreements about it. Where we might be misled, we would do better to speak in the formal mode.

Carnap generalizes this observation. *All* philosophic questions, properly understood, are really about language. The traditional practice of philosophy is a mixture of empirical (psychological and sociological) questions with ones about the logical structure of language. This is not bad, but the concoction allowed metaphysicians, who certainly did not think their enterprise an empirical one, to think they were talking about extra-linguistic objects to which they somehow had a transcendent non-empirical access. To this Carnap says: Humbug! When the empirical component is assigned to science, where it belongs, only the logic of the language of science remains. Properly speaking, this is what philosophy is and all that it is. Such logic of science Carnap calls 'syntax', but his sense is far broader than that of later generations. Specifically, Carnap supposed that the grammatical form of observation sentences could be determined or at least marked syntactically and also that the question of which sentences implied which others could likewise be answered syntactically. Such an account of the structure of observation and inference would be the non-psychological core of a proper philosophic epistemology; it would be the logical structure of science.

This syntax project had two noteworthy dividends for philosophy of science. The first was 'Über Protokollsätze' ('On Protocol Sentences') (1932a), Carnap's first presentation of the new conventionalist/pragmatist view. The Circle was debating the character of observational evidence. Is it certain? In cases of conflict with theory might the observational claims give way? Can or must it be about physical objects as opposed to private sensory states? Carnap says that the alternatives are not claims in conflict but rather the expression of different languages among which we must pragmatically choose. His own recommendation was a physicalist observational language (one far less regimented than Neurath's) in which the individual protocols could be revised in the light of other evidence and theory.

The second major dividend was the paper 'Testability and Meaning' (1936-7). Having decided that observation and implication relations were at the heart of logical analysis, Carnap saw that such analysis need not require that all empirically significant terms be explicitly definable in the observation language. Partial definitions would do if they reveal the implication relations between a given term and other terms antecedently established as empirically meaningful. These partial definitions Carnap calls reduction sentences, and they provide more adequately for

disposition terms even if observational conditions which are jointly necessary and sufficient cannot be given. The reduction does not allow the elimination of the partially defined term. That this fact does not bother Carnap shows that the point of reduction in his sense is not economy, either in ontology or in conceptual resources. More likely, the point is a logical one, or epistemic in the structural sense just noted. This analysis of scientific terms also resulted in Carnap's denial in 'Wahrheit und Bewährung' ('Truth and Confirmation') (1936) that strict translations of claims from one language into another were always possible. Elsewhere in 'Testability and Meaning' he denies a sharp observational-theoretical distinction and suggests that it matters little where or how it is made more exact. Finally, he says that the varieties of empiricism themselves are conventions, though ones he strongly endorses, and he prefers the most liberal of those considered (see Observation; Reduction, problems of).

4 America

While Carnap is strongly associated with Vienna, he resided there for just five years. In 1931 he accepted a professorship in natural philosophy at the German University in Prague, though he still frequently attended Circle meetings. With the rise of Nazism, he moved in 1935 to the University of Chicago. This move coincided with his turn to explicitly semantical concerns, and the significance of his syntax project was largely lost on his new American audience. In Chicago an encyclopedia for unified science finally found a home, with the help of Charles Morris.

A preliminary result of the semantical work was *Foundations of Logic and Mathematics* (1939) with fuller accounts in *Introduction to Semantics* (1942) and *Formalization of Logic* (1943). These works broaden the previous syntax programme, which now clarifies meaning, rather than forbidding talk of it. They deal only with extensional languages, but in *Meaning and Necessity* (1947) Carnap provided a new semantical method of intension and extension in place of the older method of the name relation. This allowed the treatment of intensional languages (such as for modality) involving even quantification. Modality was taken as an essentially linguistic affair, rather than in metaphysically more profligate ways. He added a distinction between two senses of synonymy: a weak sense amounting to mutual entailment, and a stronger sense, called intensional isomorphism, which holds whenever expressions are built up in the same way out of weakly synonymous parts. Hence, any pair of arithmetic truths will be (weakly) synonymous but will not in general be intensionally isomorphic (see Intensional logics).

5 Probability

A project to develop a theory of logical probability or degree of confirmation dominated the last three decades of Carnap's life. This notion of probability is essentially a generalization of deductive or logical implication, and, as such, a relation between sentences or propositions. Elementary statements of such probabilities are analytic or contradictory (just as claims of logical entailment would be). There is another concept of probability, the statistical frequency notion developed by Carnap's friends Richard von Mises and Hans Reichenbach, according to which the probabilities are factual ratios of classes of events. For Carnap, these two probability notions are not rivals, not even rival explications of the same ordinary concept. They are explications of two entirely different notions, both useful and both exhibiting the mathematical structure required to merit the term 'probability' (see Probability, interpretation of).

Carnap developed this notion most extensively in *Logical Foundations of Probability* (1950a), which discusses what it is to explicate an ordinary concept and proceeds to provide the required analysis. Some results were initially implausible, such as that the degree of confirmation for any scientific law on any evidence is zero. This is mitigated, however, by the result that the instances of such a law need not have zero confirmation values. Kemeny and Bar-Hillel showed independently that Carnap's formulations exclude primitive predicates (such as disposition terms) having meaning relations to other primitive terms. Kemeny proposed a solution which Carnap adopted in 'Meaning Postulates' (1952), but still the account applied to a restricted range of languages, and Carnap's later years were devoted to broadening the theory's range and adequacy. Much of this new work appeared posthumously (see Confirmation theory).

Throughout the 1940s Carnap happily spoke of abstracta, such as properties and propositions. This sparked fears of incompatibility with empiricism, and so, in 'Empiricism, Semantics, and Ontology' (1950b), Carnap argued persuasively that this fear was unfounded. The paper reaffirmed his longstanding linguistic conventionalism, now

formulated in terms of a distinction between external and internal questions of existence. In the former, what is at issue is the adoption of a conceptual framework or language. There is no question of truth or falsity here, only of the utility of a way of speech. Once a framework has been adopted, genuine questions can be asked, but the answers are either factual (settled by empirical means) or logical. In neither case is there any deep metaphysical realism or any threat to empiricism.

While the issue was raised privately much earlier, in 1951 W.V. Quine (§8) launched a public attack on the intelligibility of Carnap's notion of analyticity (or truth in virtue of meaning). Quine's view came to be very influential, though not universal. Whatever the merits of the case, Quine came to accept a notion of analyticity that applied to what had been the central examples under debate, namely, elementary logic and the so-called truths of essential predication (for example, 'All bachelors are unmarried'). This did not entirely resolve all of the issues, but it was unclear just how much any unresolvable disagreements should affect the evaluation of Carnap's philosophy (see Analyticity).

From the early 1950s until its appearance in 1964 Carnap worked on the massive *The Philosophy of Rudolf Carnap*. It included an intellectual autobiography, a then-complete bibliography, and twenty-six critical essays with Carnap's extensive replies covering every aspect of his work.

In 1956 Carnap moved to UCLA (after two years at the Institute for Advanced Study in Princeton) to fill the post left vacant by Reichenbach's death. He continued to work on probability and confirmation, and also resumed work on topics which were more explicitly part of general philosophy of science. In 'The Methodological Character of Theoretical Concepts' (1956) he ventured a sophisticated criterion of empirical significance, and in other papers he advanced an account of analyticity for theoretical concepts. The last book published in his lifetime was *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science* (1966), which was essentially a reworked transcript of a course, and reissued under its less formidable subtitle a few years later. It contains wonderfully accessible treatments of quantitative concepts and measurement, non-Euclidean geometry, relativity and a variety of other issues, such as determinism and scientific realism.

Carnap left a legacy of clarity of thought, philosophic achievement and personal kindness that has rarely been equalled. After a period of eclipse, his work has been partially 'rediscovered', and it seems likely to inform and inspire succeeding generations of philosophers much as it had done throughout the middle third of the twentieth century.

See also: Analytical philosophy

RICHARD CREATH

List of works

Carnap, R. (1922) *Der Raum: Ein Beitrag zur Wissenschaftslehre (Space: A Contribution to the Theory of Science)*, Berlin: Reuther & Reichard.(Dissertation and first use of distinctions among senses in order to reconcile conflicting philosophic views.)

Carnap, R. (1926) *Physikalische Begriffsbildung (Physical Concept Formation)*, Karlsruhe: G. Braun.(Early discussion of qualitative and quantitative concepts.)

Carnap, R. (1928) *Der logische Aufbau der Welt*, Berlin-Schlachtensee: Weltkreis-Verlag; trans. R. George, *The Logical Structure of the World*, Berkeley, CA: University of California, 1967.(Dramatic demonstration of constructional techniques.)

Carnap, R. (1929a) *Abriß der Logistik (Survey of Symbolic Logic)*, Vienna: Springer.(One of the first textbooks on symbolic logic, with applications, based on *Principia Mathematica*.)

Carnap, R., Hahn, H. and Neurath, O. (1929b) *Wissenschaftliche Weltauffassung (The Scientific Conception of the World)*, Vienna: Wolf.(The so-called manifesto of the Viennese logical empiricists.)

Carnap, R. (1932a) 'Über Protokollsätze', *Erkenntnis* 3: 215-28; trans. R. Creath and R. Nollan, 'On Protocol Sentences', *Nous* (1987) 87: 457-70.(Carnap presents a conventionalist account of the form of observation reports.)

Carnap, R (1932b) 'Überwindung der Metaphysik durch logische Analyse der Sprache', *Erkenntnis* 2: 219-41; trans. A. Pap, 'The Elimination of Metaphysics Through the Logical Analysis of Language', in *Logical Positivism*, ed. A.J. Ayer, Glencoe, IL: Free Press, 1959, 60-81.(Carnap's most famous rejection of metaphysics.)

Carnap, R (1932c) 'Die physikalische Sprache as Universalsprache der Wissenschaft' (Physical Language as the Universal Language of Science), *Erkenntnis* 2: 432-65; trans. M. Black as *The Unity of Science*, London: Kegan Paul, Trench, Trubner & Co., 1934.(Along with 1932d, Carnap's classic statement of his physicalism.)

Carnap, R (1932d) 'Psychologie in physikalischer Sprache', *Erkenntnis* 3: 107-42; trans. G. Schick, 'Psychology in Physical Language' *Erkenntnis.* 165-98. (With 1932c, Carnap's classic statement of his physicalism.)

Carnap, R (1934a) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul, Trench, Trubner & Co., 1937. (Carnap's most important presentation of his conventionalism.)

Carnap, R (1934b) 'On the Character of Philosophic Problems', trans. W.M. Malisoff, *Philosophy of Science* 1: 5-19.(Carnap distinguished between assertions and proposals and identifies philosophy as the logical syntax of the language of science.)

Carnap, R (1935) *Philosophy and Logical Syntax*, London: Kegan Paul, Trench, Trubner & Co.(A less formal presentation of the basic ideas of the preceding item.)

Carnap, R (1936) 'Wahrheit und Bewährung', *Actes du Congrès international de philosophie scientifique, Sorbonne, Paris* 1935 4: 18-23; trans. H. Feigl, 'Truth and Confirmation', in *Readings in Philosophical Analysis*, ed. H. Feigl and W. Sellars, New York: Appleton-Century-Crofts, 1949.(Embraces semantics, a distinction between truth and confirmation, and a claim that meanings can change as theories do, so as to make exact translation impossible.)

Carnap, R (1936-7) 'Testability and Meaning', *Philosophy of Science* 3: 419-71; 4: 1-40. (Embraces partial definitions to treat disposition terms, makes further remarks about the theory-observation distinction, and treats empiricism as a proposal.)

Carnap, R (1939) *Foundations of Logic and Mathematics*, Chicago, IL: University of Chicago Press.(Treats a variety of philosophic issues, including the structure of scientific theories, from his emerging semantical point of view.)

Carnap, R (1942) *Introduction to Semantics*, Cambridge, MA: Harvard University Press. (Textbook-like presentation of Carnap's semantical views.)

Carnap, R (1943) *Formalization of Logic*, Cambridge, MA: Harvard University Press. (Extensive treatment of non-normal interpretations in the formalization of logic.)

Carnap, R (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Presents Carnap's distinction between intension and extension, a sophisticated account of synonymy, a surprising ambiguity about names, and a treatment of modalities and quantification.)

Carnap, R (1950a) *Logical Foundations of Probability*, Chicago, IL: University of Chicago Press.(Presents Carnap's first systematic treatment of his distinctive views on probability; though the view is not widely accepted the book is a gold mine of epistemological insight and argument.)

Carnap, R (1950b) 'Empiricism, Semantics, and Ontology', *Revue internationale de philosophie* 4: 20-40. (Presents Carnap's important distinction between external and internal questions.)

Carnap, R (1952a) *The Continuum of Inductive Methods*, Chicago, IL: University of Chicago Press.(Broadens the range of available treatments of probability.)

Carnap, R (1952b) 'Meaning Postulates', *Philosophical Studies* 3: 65-73.(Repairs a defect in Carnap's account of meaning generated by his work on probability.)

Carnap, R (1955) 'Meaning and Synonymy in Natural Languages', *Philosophical Studies* 6: 33-47.(Carnap's attempt to provide a pragmatics for natural languages, that is, to satisfy Quine's demand for behavioural criteria for meaning.)

Carnap, R (1956) 'The Methodological Character of Theoretical Concepts', *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, ed. H. Feigl and M. Scriven, Minneapolis, MN: University of Minnesota Press.(Carnap's most sophisticated criterion of meaningfulness.)

Carnap, R (1963) 'Intellectual Autobiography' and 'Replies and Systematic Expositions', *The Philosophy of Rudolf Carnap*, ed. P.A. Schilpp, La Salle, IL: Open Court, 3-84, 859-1013.(The volume, which includes twenty-six critical essays, is an impressive exploration of many facets of Carnap's philosophy.)

Carnap, R (1966) *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science*, ed. M. Gardner, New York: Basic Books.(A clear and accessible treatment of Carnap's views on science; excellent discussions of measurement, non-Euclidean space and relativity.)

Carnap, R (1971, 1980) 'A Basic System of Inductive Logic', *Studies in Inductive Logic and Probability*, vol. 1, ed. R. Carnap and R. Jeffrey, vol. 2, ed. R. Jeffrey, Berkeley, CA: University of California Press, 33-165, 7-155.(A substantial revision and enrichment of the account given in *Logical Foundations of Probability*.)

Carnap, R. and Quine, W.V. (1990) *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work*, ed. R. Creath, Berkeley, CA: University of California Press.(Besides the correspondence, the volume contains a previously unpublished reply to Quine's 'Two Dogmas of Empiricism'.)

References and further reading

Coffa, J.A. (1991) *The Semantic Tradition from Kant to Carnap: To the Vienna Station*, ed. L. Wessels, Cambridge: Cambridge University Press.(Pioneering re-evaluation of Carnap and the tradition out of which he comes.)

Creath, R. (1991) 'Every Dogma Has Its Day' *Erkenntnis* 35: 347-89.(Detailed review of the arguments of Quine's 'Two Dogmas of Empiricism' and surrounding literature.)

Creath, R. (1994) 'Functionalist Theories of Meaning and the Defense of Analyticity', in Salmon and Wolters (eds) 1994, 287-304.(An attempt to meet Quine's demands for behavioural criteria for natural languages.)

Friedman, M. (1992) 'Epistemology in the *Aufbau*', *Synthèse* 93: 15-57.(A significant reinterpretation of the *Aufbau*.)

Friedman, M. (1994) 'Geometry, Convention, and the Relativized A Priori', in Salmon and Wolters (eds) 1994, 21-34.(An examination of a view of the a priori according to which it is both relativized and revisable.)

Goldfarb, W. and Ricketts, T. (1992) 'Carnap and the Philosophy of Mathematics', in *Science and Subjectivity*, ed. D. Bell and W. Vossenkuhl, Berlin: Akademie Verlag, 61-78.(An important modern discussion of Carnap's philosophy of mathematics.)

Goodman, N. (1951) *The Structure of Appearance*, Cambridge, MA: Harvard University Press.(Detailed technical discussion of Carnap's *Aufbau*.)

Goodman, N. (1963) 'The Significance of the *Aufbau*', in Schilpp, 1963, 545-58.(A spirited defence of the *Aufbau* project by one of its major critics.)

Hintikka, J. (ed.) (1975) *Rudolf Carnap, Logical Empiricist*, Dordrecht: Reidel.(An important collection of essays on Carnap's work.)

Martin, R.M. (1952) 'On "Analytic"', *Philosophical Studies* 3: 42-7(An important reply to Quine's 'Two Dogmas of Empiricism'.)

Oberdan, T. (1992) *Protocols, Truth, and Convention*, Amsterdam: Rodopi. (Extensive discussion of Carnap's views in the 1930s.)

Quine, W.V. (1951) 'Two Dogmas of Empiricism', *Philosophical Review* 60: 20-43.(Quine's most famous attack on Carnap's notion of analyticity.)

Quine, W.V. (1963) 'Carnap and Logical Truth', in Schilpp, 1963, 385-406.(Quine's most extensive attack on analyticity.)

Richardson, A. (1997) *Carnap's Construction of the World: The Aufbau and the Emergence of Logical Empiricism*, Cambridge: Cambridge University Press.(Important re-examination of Carnap's *Aufbau* and the Neo-Kantian tradition.)

Russell, B. and Whitehead, A.N. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1927.(Enormously influential, this book carries out the reduction of mathematics to logic, including set theory.)

Salmon, W. and G. Wolters (eds) (1994) *Logic, Language, and the Structure of Scientific Theories*, Pittsburgh, PA: University of Pittsburgh Press, and Konstanz: Universitätsverlag Konstanz.(Significant collection of essays on Carnap and Reichenbach.)

Sarkar, S. (ed.) (1992) 'Carnap: A Centenary Reappraisal', *Synthèse* 93.(A centennial collection of essays on Carnap.)

Schilpp, P.A. (ed.) (1963) *The Philosophy of Rudolf Carnap*, La Salle, IL: Open Court.(Its 1,100 pages contain twenty-six critical essays on Carnap's work, Carnap's autobiography, replies, and a bibliography.)

Spohn, W. (ed.) (1991) *Hans Reichenbach, Rudolf Carnap: A Centenary*, Dordrecht: Kluwer.(A centennial collection of essays on Reichenbach and Carnap.)

Übel, T. (1992) *Overcoming Logical Positivism from Within*, Amsterdam: Rodopi. (Extensive examination of the protocol sentences debate from Neurath's point of view.)

Carneades (214-129 BC)

The Greek philosopher Carneades was head of the Academy from 167 to 137 BC. Born in North Africa he migrated to Athens, where he studied logic with the Stoic Diogenes of Babylon; but he was soon seduced by the Academy, to which his allegiance was thereafter lifelong. He was a celebrated figure; and in 155 BC he was sent by Athens to Rome as a political ambassador, where he astounded the youth by his rhetorical powers and outraged their elders by his arguments against justice.

*Under Carneades' direction the Academy remained sceptical. But he enlarged the sceptical armoury - in particular, he deployed sorites arguments against various dogmatic positions. He also broadened the target of sceptical attack: thus he showed an especial interest in ethics, where his 'division' of possible ethical theories served later as a standard framework for thought on the subject. But his major innovation concerned the notion of 'the plausible' (*to pithanon*). Even if we cannot determine which appearances are true and which false, we are able to distinguish the plausible from the implausible - and further to distinguish among several grades of plausibility. It is disputed - and it was disputed among his immediate followers - how, if at all, Carneades' remarks on the plausible are to be reconciled with his scepticism.*

1 Life

Carneades' life is ill documented. He was born in 214 BC at Cyrene in North Africa. He moved to Athens where he studied logic with the Stoic Diogenes of Babylon and no doubt also listened to the other leading philosophers of the day. He gravitated to the Academy (see Academy), which was still following the sceptical path laid down by Arcesilaus, and in 167 BC he was chosen as scholarch. The school flourished under his leadership: later authors deemed that he had opened a new phase in its history. Blindness and miserable health obliged him to resign his position in 137; but he continued to philosophize vigorously until his death in 129. And when he died there was an eclipse of the moon.

He was a loud man, and admired even by his opponents for his powerful and persuasive style of argument: he had the irresistible force of a mighty river, the subtlety and resilience of a hunted beast. He lost no argument and left no hearer unmoved. When in 155 BC the Athenians petitioned the Roman Senate for the remission of a fine, they sent three philosophers as ambassadors - Diogenes the Stoic, Critolaus the Peripatetic, and Carneades. At Rome the three men displayed their professional talents to an enthusiastic public. Carneades revealed a particular power to charm and excite - the more so when, on successive days and with equal power, he urged the claims first of justice and then of injustice. Stern Cato hurried business through the Senate and bundled the Greeks off home before they could corrupt the Roman youth.

Like Arcesilaus - and Socrates - before him, Carneades wrote nothing. Several of his pupils took notes on his lectures and wrote memorials of his arguments; and our meagre knowledge of Carneades depends ultimately on these works - all of which are lost. Carneades was persuasive, but he was also elusive; and his pupils disagreed on how to interpret him: his successor Clitomachus, who wrote voluminously about his master, confessed that 'he could never understand what Carneades believed' (Cicero, *Academics* II 139); and Metrodorus claimed that 'everyone has misunderstood Carneades' (Philodemus, *History of the Academy* XXVI 8-10).

Also like Arcesilaus - and Socrates - before him, Carneades argued against all-comers, showing himself 'the most acute and fertile' proponent of the *elenchos* or method of refutation (Cicero, *Tusculan Disputations* V 11). He also argued 'on both sides' - notoriously during his ambassadorial visit to Rome. And he was familiar with at least the rudiments of Stoic logic, so that his arguments against the Stoics were often and deliberately expressed as instances of Stoic inference patterns (see Stoicism §11).

In addition, he had a weakness for 'soritical' arguments - arguments which proceed 'little by little' from apparently true premises to an apparently false conclusion (see Vagueness §2). Such arguments were not invented by Carneades. Moreover, they had already been used against the Stoics; for Chrysippus had tried to answer them: before he came to the difficult cases he would 'fall silent' (*hēsychazein*). Carneades thought little of this reply:

> As for falling silent, as far as I'm concerned you may snore. What good does it do you? Someone has only to wake you up and pose the same question again.

(Cicero, *Academics* II 93)

The sorites cannot be slept away (see further Stoicism §11).

Like Arcesilaus, Carneades was especially energetic in his attacks on the Stoa: he read Stoic texts assiduously, and confessed that ‘ had Chrysippus not existed, I would not have done either’ (Diogenes Laertius IV 62). But his aim extended more widely - we happen to know that he discussed some aspects of Epicureanism; and Sextus Empiricus affirms that ‘with regard to the criterion of truth, Carneades attacked not only the Stoics but all his predecessors’ (*Against the Mathematicians* VII 159). Moreover, he maintained that, on some issues at least, ‘the dispute between Stoics and Peripatetics concerned not substance but terminology’ (Cicero, *On Ends* III 41), so that an argument against the former was thereby an argument against the latter (see Antiochus §2). It should not be thought (as some have thought) that his philosophy was purely polemical and *ad hominem*.

Carneades had a particular interest in ethics, where he not only discussed philosophical matters but also composed an idiosyncratic work on consolation. He also reflected on philosophical theology, and on such related issues as divination and the doctrine of fate. But it was primarily for his epistemological attitudes that he was notorious: he argued for some version of scepticism - and at the same time he advocated a ‘criterion’, namely *to pithanon* or ‘the plausible’. This last item was - or at any rate seemed to some of his disciples to be - a profound innovation; and it was presumably ‘the plausible’ which encouraged the suggestion that Carneades had not merely continued the tradition of Arcesilaus but had instituted a New Academy.

2 Ethics

Dogmatic philosophers disagreed over the *telos* or (Latin) *summum bonum*, the goal of life or the highest good. Carneades catalogued the differing doctrines and incorporated them into a schedule - the ‘division of Carneades’ - which set out ‘not only the opinions on the highest good which have been championed by philosophers up to now but all opinions which can possibly be maintained’ (Cicero, *On Ends* V 16). The division was celebrated, but no coherent version has survived. Carneades himself urged that the *telos* is ‘the enjoyment of the goods to which nature first inclines us’ (Cicero, *Academics* II 131): it is not known how he advocated the view - which in any event he advanced only as part of an *ad hominem* argument against the Stoics.

On the two Roman speeches on justice we have a little more information. The first speech was an anthology: Carneades ‘put together everything which had been said in favour of justice’ (Lactantius, *Divine Institutions* V 14.5). On the following day he first distinguished between ‘civil’ justice and ‘natural’ justice. He then urged that civil justice makes good sense: that is to say, it is in the interest of the legislators to enforce the laws which constitute civil justice and which they have conventionally established. But civil ‘justice’ is not just at all. Natural justice, on the other hand, is just - but it is either a phantom or a folly. If the Romans, with their wealth and their conquests, ‘wish to be just, that is, to give back what is not theirs, then they must return to their cottages and live in poverty and misery’. For the community as a whole, natural justice is absurd. And for the individual too. Suppose that you are selling a house which you alone know to be riddled with dry rot: do you show the rot to a prospective purchaser? If you do not, you are prudent - and unjust. If you do, you are just - and foolish. Suppose that you have been shipwrecked and find a fellow sailor clinging weakly to a floating plank: do you dislodge him and grab the plank yourself? If you do, you are prudent - and unjust. If you do not, you are just - and dead.

The second speech, no less than the first, was a cento of familiar arguments and examples. Carneades’ originality lay not in invention but in presentation. What he himself thought of the matter, and what he intended in arguing ‘on both sides’, we cannot tell. But he produced, as Socrates had produced, perplexity and disquiet and hostility. Cato was enraged. (But the Christian Lactantius was pleased; for Carneades had unwittingly shown that pagan philosophy, ignorant of divine truth, had no understanding of the nature and the glory of justice.)

3 Theology

Carneades collected arguments against the existence of gods. Some of these mockingly followed Stoic patterns of reasoning. Thus: the gods, being living creatures, must be capable of sense perception; anything which perceives is thereby pleased or displeased; anything which may be displeased may change for the worse; anything which may change for the worse may be destroyed: therefore if there are any gods they can be destroyed. But gods cannot be destroyed. Hence there are no gods. There was also a batch of soritical arguments. Thus:

> If Zeus is a god, then Poseidon is a god; if Poseidon is a god, then Achelous is a god; if Achelous is a god, then the Nile is a god;…then streams are gods. But streams are not gods. Therefore Zeus is not a god. But if there were any gods, then Zeus would be one. Hence there are no gods.
>
> (Sextus Empiricus, *Against the Mathematicians* IX 183)

According to Cicero, ‘Carneades said these things not in order to do away with the gods - what could be less appropriate in a philosopher? - but in order to show that the Stoics explain nothing about the gods’ (*On the Nature of the Gods* III 44). Perhaps so - in any event, it would be wrong to conclude that Carneades was a convinced atheist; for if even the arguments are not *ad hominem* - *ad Stoicos* - they will surely have formed the one half of a set of arguments ‘on both sides’.

Carneades also argued against divine providence, and in particular he disputed the Stoic view that beasts exist for the benefit of mankind - what about bats and scorpions and crocodiles? And he rejected the science of divination (which the Stoics had generally applauded), urging that when a seer gets things right, the success is mere luck. Now divination, according to Carneades, presupposes causal determinism; for ‘even Apollo cannot foretell an event unless there are causes in nature which so constrain it that it is necessary for it to occur’ (Cicero, *On Fate* 32).

The Stoics, who defended divination, held that everything does depend on such naturally necessitating causes - such was their doctrine of fate (see Stoicism §20). The doctrine had been variously attacked. Carneades allowed that some of the arguments used against it were fallacious: they falsely conflated the thought that an event is causally determined with the thought that someone may have truly said in advance that it will occur. (To say truly that something will happen is not the same as to divine or foretell it.) And he produced a better argument:

> If everything comes about by antecedent causes… then necessity determines everything; if so, then nothing is in our power. Now some things are in our power. But if everything comes about by fate, then everything comes about by antecedent causes. Therefore not everything which comes about comes about by fate.
>
> (Cicero, *On Fate* 31)

The Epicureans also rejected fatalism. In doing so they introduced their celebrated theory of ‘atomic swerves’ (see Epicureanism §12): atoms may suddenly, and for no cause, change trajectory; and these swerves ensure that there may be ‘voluntary motions of the mind’ which leave some actions in our power. But Carneades criticised the Epicureans no less than the Stoics, maintaining (truly enough) that voluntary motion can be better defended than by the introduction of uncaused swerves.

4 Scepticism

> Carneades and Clitomachus and all their friends trample on the doctrines of everyone else and themselves expressly affirm that everything is inapprehensible and that a false appearance always neighbours a true one.
>
> (Hermias, *Mockery of the Pagan Philosophers* 15)

According to Carneades, we can apprehend nothing - not even (as Cicero adds) that we can apprehend nothing; for next to and indistinguishable from any true appearance there will always be found a false appearance.

Carneades fired broadsides against every philosophical attempt to provide a criterion of truth. He peppered them with bent oars and round towers and moving coastlines - with all the ‘illusions’ which have diverted later sceptics. In particular, he reflected on the reigning Stoic account of knowledge, which turned on the notion of an ‘apprehensive appearance’ (*phantasia katalēptikē*) (see Stoicism §12).

> An appearance (*phantasia*), being the origin of cognition in animals, ought, like light, both to reveal itself and to indicate the evident item which produces it.

But *phantasiai*, like unreliable messengers, do not always report correctly the information with which they are charged. What is more, even if a *phantasia* does manage to give a correct report,

> none is true in such a way that it could not be false; rather, for any *phantasia* which is thought to be true, a false *phantasia* indistinguishable from it (*aparallaktos*) can be found.

(Sextus Empiricus, *Against the Mathematicians* VII 163-4)

He offered numerous proofs and illustrations of this indistinguishability. Consider the twins, Castor and Pollux: it looks to me for all the world as though Castor is sitting opposite me at the dining table - and indeed he is sitting there. But is this appearance apprehensive, does it give me *knowledge* that Castor is there? No - for an indistinguishable appearance would have been caused by his twin, Pollux.

In general, if you know something, you must have apprehended it; apprehension is assent to an apprehensive appearance; and an appearance of something is apprehensive if it is true of that thing, if it was caused by that thing, and if it could not have been produced by anything else (see Stoicism §12). Like Arcesilaus, Carneades fastened on to the third condition:

> They say that they reject this one element, that something true can appear in such a way that what is false could not appear in the same way.

(Cicero, *Academics* II 33)

If an appearance is apprehensive, then were it not the case that so-and-so it could not seem to me that so-and-so. But this is never so: it *could* seem to me that Castor was sitting opposite me even if (say) not he but Pollux were there. No appearances are apprehensive. Nothing is known.

Thus far, destruction. But there is also an apparently positive side to Carneades' scepticism. Alongside the distinction between true and false appearances there is a second and independent distinction between 'plausible' (*pithanos*) and implausible opinions. (The word *pithanos* was often Latinized as *probabilis* and hence rendered in English as 'probable'; but 'probable', in its modern sense, is a false translation of *pithanos*.) Truth and falsity are determined by the relation of an appearance to the items *of which* it is an appearance. Plausibility and implausibility are determined by the relation of an appearance to the item *to whom* it appears. An appearance is plausible if it tends to persuade or to evoke assent: it looks to me as if there is a coil of rope in the potting-shed - and the appearance is plausible in so far as I find myself inclined to accept that there is a coil of rope there.

What is plausible for me may not be plausible for you. And there are degrees of plausibility, some appearances being more persuasive than others. Moreover, we may *scrutinize* our appearances. Perhaps it is really a snake, not a coil of rope? I prod about in the potting-shed with a stick: nothing budges - and I now have a 'scrutinized' plausible appearance of a coil of rope. Again, appearances do not come to us as isolated atoms: they arrive in molecules, so that a plausible appearance may be controverted or *uncontroverted* by its associates. There seem to be no sloughed skins in the shed, I see no notice urging me to Beware of Serpents in the Grass, and so on. None of the other appearances associated with the appearance of the coil of rope controverts that appearance: I have an 'uncontroverted and scrutinized' plausible appearance of a coil of rope.

Plausible appearances, however well scrutinized and however uncontroverted, may always be false. None the less, we shall prefer the plausible to the implausible, the more plausible to the less plausible, and the uncontroverted and scrutinized plausible to anything else.

Such is Carneades' account of plausibility: herein appeared to lie his chief innovation - on account of which his 'New' Academy was later distinguished from the Middle Academy of Arcesilaus. What was the philosophical role and function of the account? The problem is this: on the one hand, the plausible, of which Carneades seems to approve, is that which evokes *assent*; on the other hand, Carneades apparently argues for a scepticism which *repudiates* any act of assent. Carneades' own pupils were puzzled, and interpreters have found no solution which commands general agreement.

One of Carneades' pupils probably construed the theory of plausibility as a theory of knowledge. Carneades' destructive arguments proved that the standard account of knowledge was flawed. Hence he had to devise another account - and the theory of the plausible lay at its heart. Carneades was not a sceptic - on the contrary, the point of the *pithanon* pricks the bubble of scepticism (see Philo of Larissa §2). Later, St Augustine (on what authority is uncertain) affirmed that Carneades had never intended to attack 'ordinary' knowledge: his sceptical assault was limited to philosophical knowledge; and the theory of plausibility offered an account of our ordinary knowledge of ordinary things.

Or perhaps plausibility offers not knowledge but a knowledge substitute. We cannot strictly *know* anything - but that is no reason to suspend judgement, nor does it imply that all appearances are on an equal footing. Plausibility provides a criterion for rational belief - and rational belief is all we have on earth and all we need to have.

Sextus Empiricus presents Carneades' 'plausible' as a criterion of action: it is concerned with 'the discrimination of what is good and what is bad' (*Outlines of Pyrrhonism* I 226). Carneades suggests not that it is rational to *believe* what is plausible but that it is rational to *act on* what is plausible. (And rationality will recommend satisfaction with different degrees and types of plausibility in different practical circumstances.)

Or does the plausible play a purely descriptive and explanatory role? The question to which the plausible offers an answer is perhaps not 'How *should* a sceptic act?' but rather 'How *can* a sceptic act?' Dogmatists argued that without beliefs a sceptic could never stay alive (see Arcesilaus §2). But according to Carneades, a sceptic's actions may be determined by the appearances; for not all appearances are equally plausible - and a sceptic will be moved by the most plausible appearance. Not, of course, that a sceptic will *believe* the appearance to be true: the appearance evokes 'assent' merely insofar as it is acted upon.

These brief suggestions do not exhaust the gamut of interpretations; and the surviving texts scarcely allow a rational choice among them: only in the Elysian fields - if anywhere - shall we learn the truth about Carneades.

JONATHAN BARNES

References and further reading

Bett, R. (1989) 'Carneades' *Pithanon*: A Reappraisal of its Role and Status', *Oxford Studies in Ancient Philosophy* 7: 59-94.(Discusses the nature and the consistency of Carneades' account of plausible appearances.)

Cicero (45 BC) *Academics*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933.(Our most important single source on the sceptical Academy.)

Cicero (mid 44 BC) *On Fate (De fato)*, trans. R. Sharples, Warminster: Aris & Phillips, 1991.(Dialogue on causation and determinism; incompletely preserved.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius: Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols. (Book IV includes a brief biography of Carneades.)

Mette, H.-J. (1985) 'Weitere Akademiker heute: von Lakydes bis zu Kleitomachos' (More Academics: Lacydes to Clitomachus), *Lustrum* 27: 39-148. (Collects the ancient testimonies on Carneades.)

Philodemus (*c*.80-40 BC) *History of the Academy*, trans. T. Dorandi, *Filodemo. Storia dei filosofi. Platone e l'Academia*, Naples: Bibliopolis, 1991.(Greek text, with Italian translation and commentary; a biographical history.)

Sextus Empiricus (*c*. AD 200) *Against the Mathematicians*, trans. R.G. Bury, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1935.(Book VII includes a long presentation of some of Carneades' arguments.)

Striker, G. (1980) 'Sceptical Strategies', in M. Schofield, M.F. Burnyeat and J. Barnes (eds) *Doubt and Dogmatism*, Oxford: Clarendon Press.(An account of the aims and methods of Academic scepticism.)

Carolingian renaissance

The 'Carolingian renaissance' is the name given to the cultural revival in northern Europe during the late eighth and ninth centuries, instigated by Charlemagne and his court scholars. Carolingian intellectual life centred around the recovery of classical Latin texts and learning, though in a strictly Christian setting. The only celebrated philosopher of the time is Johannes Scottus Eriugena, but the daring Neoplatonic speculations of his masterpiece, the Periphyseon (On the Division of Nature) are not at all characteristic of the time and are based on Greek sources (Pseudo-Dionysius, Gregory of Nyssa, Maximus the Confessor) generally unknown to his contemporaries. The mainstream of Carolingian thought is important for the history of philosophy in three particular ways. First, it was at this time that logic first started to take the fundamental role it would have throughout the Middle Ages. Second, scholars began to consider how ideas they found in late antique Latin Neoplatonic texts could be interpreted in a way compatible with Christianity. Third (as would so often again be the case in the Middle Ages), controversies over Christian doctrine led thinkers to analyse some of the concepts they involved: for instance, the dispute in the mid-ninth century over predestination led to discussion about free will and punishment.

1 General characteristics

It was Charlemagne himself who was in large part responsible for instigating the northern European cultural and intellectual renaissance of the late eighth and ninth centuries. From the mid-770s onwards, leading scholars gathered at his court, first men from southern Europe (where the tradition of classical learning still flourished) such as the Lombards Peter of Pisa, Paul the Deacon and Paulinus of Aquileia and the Spanish Goth Theodulf, who became bishop of Orleans, and then later the Anglo-Saxon Alcuin and his circle. During the reign of Louis the Pious (814-40) the intellectual movement, no longer encouraged at Court, was continued in the great monasteries such as Tours (where Alcuin had ended his life as abbot) and Fulda. By contrast, Charlemagne's grandson Charles the Bald (d. 877) was an enthusiastic promoter of learning and acted as patron and protector to the outstanding thinker of the period, Johannes Scottus Eriugena.

The Carolingian renaissance was one aspect of a wider renewal instituted by Charlemagne, which included reform of corruption and laxness in the clergy, restoration of correct latinity and the regularization of biblical and liturgical texts. The small elite of intellectuals at court, as well as writing history and engaging in theological dispute, composed classicizing poetry (quite often in panegyric of Charlemagne) in a self-conscious return to ancient models, parallel to the political renewal of antiquity which culminated in Charlemagne's imperial coronation in 800. Alcuin adopted the nickname 'Flaccus' (Horace) and addressed poems in classical metres to his group of intimate pupils. A library was built up at the court which contained copies of classical authors such as Lucan, Terence, Horace, Juvenal, Martial and Cicero; and after Charlemagne's death, Einhard wrote his life in a form imitated from Suetonius. In the mid-ninth century Lupus, Abbot of Ferrières, wrote elegant Latin letters which, in their concern for discovering and collating ancient texts, anticipate the interests of early modern humanists. Even Greek was studied in this period, especially by scholars of Irish origin such as Martin of Laon, Eriugena and Sedulius Scottus (also a fine Latin poet).

Historians often present the role of the Carolingian renaissance in the history of philosophy in terms simply of this rediscovery of ancient texts. Apart from Eriugena (who is envisaged as an isolated figure, inhabiting the thought world of Greek Neoplatonism or even nineteenth-century German idealism), Carolingian thinkers are described as passive assimilators of the ancient writings they began to study. Certainly, the forms and techniques favoured by thinkers of the time - excerption, paraphrase, compilation and glossing - give an impression of servility. Yet, through their particular interests and emphases, Carolingian thinkers set the direction for the subsequent course of medieval philosophy, a direction quite different from what late antique precedent would have suggested. With the exception of Eriugena, Carolingian thinkers did not take up the Neoplatonic systematizing of the Greek tradition after Plotinus, nor did they reflect the strong emphasis on rhetoric characteristic of classical Latin culture. Their thought is typified rather by the important place of logic and grammar; by the attempt to reach a Christian interpretation of pagan motifs and philosophical ideas; and by the importance of controversies over Christian doctrine in stimulating argument and analysis.

2 The revival of logic

It is no surprise that grammar should have been important in Carolingian education, since Latin was now (even in romance areas) not a mother tongue but a language which needed to be learned. But it is striking how, even in this period, grammarians went beyond straightforward linguistic instruction to explore some of the semantic problems raised in Priscian's *Institutiones grammaticae (Principles of Grammar)*. There was no such obvious didactic reason for the study of logic, yet this subject was central to Carolingian education. Already in the *Libri Carolini*, (written very probably by Theodulf of Orleans, *circa* 789-92), ostentatious use is made of formal logical methods of argument. In the 790s, Alcuin prefaced a late antique paraphrase of the *Categories*, which he attributed to Augustine, with a dedicatory verse epistle to Charlemagne: the *Categoriae decem (Ten Categories)*, as it was called, would be the most widely studied logical text in the next century. Alcuin also composed a treatise on logic; this was a patchwork of borrowed passages, especially from two late antique Christian encyclopaedists, Cassiodorus and Isidore of Seville (see Encyclopedists, medieval §§6-7), and from the *Categoriae decem*. Whereas Isidore had treated logic as a verbal art, and both he and Cassiodorus had placed great emphasis on syllogistic argument, Alcuin saw the ten categories as central to the subject, and he used them in understanding the Trinity and in order to distinguish God from his creation. Alcuin's pupils (especially a fellow Anglo-Saxon, Candidus) continued and extended these interests. A set of passages (the 'Munich passages', *circa* 800) connected with Candidus contains logical extracts and exercises, discussions of the Trinity and also a short dialogue which makes an original adaptation of Augustine's arguments for the existence of God. At much the same time another of Alcuin's pupils, Fredegisus, composed a little treatise *De substantia nihili et de tenebris (On the Substance of Nothing and on Darkness)*, which attempts, rather unsuccessfully, to elucidate the reference of 'nothing'.

From the 850s onwards, interest in logic intensified. For example, in his *Liber de anima (Book on the Soul)* in the early 860s, Ratramnus, a monk at Corbie, turned a question raised by Augustine about the relationship between individual souls and the hypostasis Soul into a logical discussion about the relations between individual souls and the universal soul. The earliest glossed manuscripts of the *Categoriae decem* date from about ten years later. Some use the logical doctrines here as opportunities to develop the characteristic ideas of Eriugena (often without much real connection with the text). Others concentrate on more down-to-earth explanation of fundamental logical terms and concepts. They indicate the beginning of a process of careful study of ancient logical texts, which would ensure that, by the eleventh century, a high level of logical proficiency was common among well-trained thinkers (see Language, medieval theories of; Logic, medieval).

3 Platonism and Christianity

The sources for ancient philosophy which first began to be used in the Carolingian period would remain the principal ones until the late twelfth century. On the one hand, there were writings by Church Fathers which were strongly marked by Neoplatonism - for example, Augustine's *De trinitate (On the Trinity)*, heavily influenced by Plotinus, which was available earlier but not exploited by medieval writers until the time of Charlemagne (in, for instance, Alcuin's *De fide sanctae trinitatis (On Faith in the Holy Trinity)* (see Augustine; Patristic philosophy). There were also the pseudo-Dionysian writings, redolent of Proclus, which reached the West in 827 and were translated into Latin first by Hilduin, Abbot of St Denis and then, more successfully, by Eriugena in the late 850s (see Pseudo-Dionysius). On the other hand, there was a handful of more direct witnesses to ancient philosophy. Plato's *Timaeus*, in the partial translation of Calcidius, was in Charlemagne's library. Alcuin is the first medieval author to use Boethius' *De consolatione philosophiae (On the Consolation of Philosophy)*. The allegorical prosimetrum by Martianus Capella, *De nuptiis Philologiae et Mercurii (On the Marriage of Philology and Mercury)* began to be read in the early ninth century (and was being glossed intensively a few decades later), and Macrobius' commentary on Cicero's *Somnium Scipionis (Scipio's Dream)* came into use at much the same time (see Encyclopedists, medieval §§3-4).

These direct witnesses posed Carolingian scholars with the problem of how to treat explicitly pagan material. This arose most directly in connection with the *De consolatione philosophiae*. Boethius was a Christian, but he omits any openly Christian doctrine from the *De consolatione* and, in sections such as the metrum (III, m.9) *O qui perpetua*, his epitome of the *Timaeus*, he mentions pagan ideas (reincarnation, the World Soul) which are rejected by Christianity (see Boethius, A.M.S. §5). Two commentaries from the end of the ninth century illustrate diametrically opposed ways of tackling the problem. Remigius of Auxerre, the most prolific of late Carolingian

commentators on secular texts, chooses to interpret everything in an explicitly Christian sense and either treats references to pagan doctrines as mere analogies or ignores them entirely. By contrast, Bovo, a monk of Corvey, begins his commentary (solely on *O qui perpetua*) by expressing surprise that Boethius, though a Christian, 'did not discuss anything of Church doctrine, but wished to reveal the teachings of the philosophers, especially the Platonists, to his readers'; and he then goes on, with obvious relish, to give a surprisingly accurate exegesis of the pagan ideas in the poem. But Bovo's bold position would have few imitators, whereas Remigius' Christianizing interpretation of pagan texts set a model for the next three centuries (see Platonism, medieval).

4 Doctrinal dispute and philosophical analysis

Early in the Carolingian renaissance, the need to answer and convincingly reject the Greek position on image-worship proposed at the second Council of Nicaea provoked the most ambitious theological treatise of the period, the *Libri Carolini*. Throughout the period, doctrinal controversies (on Christology, the Trinity and predestination) continued to provide an impetus towards systematic thought. Not only did they force scholars to collect and order dossiers of authoritative material from the Church Fathers to support their case (as with Alcuin's voluminous works against the Adoptionist heresy), they also encouraged them to analyse and explore difficult and important concepts.

The mid-ninth-century controversy over predestination was particularly important in stimulating philosophical analysis. The dispute was begun by Gottschalk of Orbais, who, so his opponents claimed, held that divine predestination was dual: of the blessed to salvation and of the wicked to damnation. Eriugena's contribution to the debate in his *De praedestinatione (On Predestination)* is well known for its use of logic and its adoption of a number of Neoplatonic themes in establishing its radical conclusions. However, the sophistication of other participants in the controversy is less often noted. For example, both Hraban Maur (Hrabanus Maurus), Abbot of Fulda, and Hincmar, Archbishop of Rheims, made the distinction in their attacks on Gottschalk between predestining and foreknowing, and both recognized (though they did not fully overcome) the danger that, even if God is said to predestine only to salvation, he will seem arbitrary in his choice of whom so to predestine. Those who attacked Eriugena's treatise also showed greater acumen than is commonly recognized. For instance, Prudentius of Troyes challenges Eriugena's conceptual analysis of free will - especially his distinction between the will itself and its strength and power, and Florus of Lyons enters into a discussion of how different types of words signify in order to show the weakness of Eriugena's theory of sin (see Predestination).

See also: Byzantine philosophy; Chartres, School of; Encyclopedists, medieval; Eriugena, J.S.; Logic, medieval; Platonism, medieval

JOHN MARENBON

References and further reading

Alcuin (795-802) *De dialectica (On Dialectic)*, in J.-P. Migne (ed.) *Patrologia Latina*, Paris, 1851, vol. 101, col. 951-76.(The earliest medieval Latin logic textbook.)

Alcuin (795-802) *De fide sanctae Trinitatis(On Faith in the Holy Trinity)*, in J.-P. Migne (ed.) *Patrologia Latina*, Paris, 1851, vol. 101, col. 11-58.(A discussion of the Trinity, based on Augustine's *De trinitate*.)

Anonymous (after *c*.870) Glosses to *Categoriae Decem*.(A selection of these can be found in J. Marenbon, *From the Circle of Alcuin to the School of Auxerre*, Cambridge: Cambridge University Press, 1981.)

Bischoff, B. (ed.) (1965) *Karl der Grosse: Lebenswerk und Nachleben (Charlemagne: His Life's Work and its Influence)*, vol. II *Das geistige Leben (Cultural Life)*, Düsseldorf: Schwann.(Authoritative essays on all aspects of the early Carolingian renaissance.)

Bovo of Corvey (*c*.880-90) Commentary on *O qui perpetua*, in R.B.C. Huygens, 'Mittelalterliche Kommentare zum *O qui perpetua*', *Sacris Erudiri* 6, 1954: 373-427.(Comments on what Bovo perceives as pagan ideas in Boethius' poem.)

Bullough, D. (1991) *Carolingian Renewal: Sources and Heritage*, Manchester: Manchester University Press. (Wide-ranging studies in Carolingian intellectual history and its political background.)

Candidus Wizo (*c*.800) Philosophical and theological passages, in J. Marenbon, *From the Circle of Alcuin to the School of Auxerre*, Cambridge: Cambridge University Press, 1981.(Discusses logic, trinitarian theology and quotes from Timaeus.)

Florus of Lyons (early 850s) *Adversus Ioannis Scoti Erigenae erroneas definitiones*, in J.-P. Migne (ed.) *Patrologia Latina*, Paris, 1852, vol. 119, col. 101-250.(Criticism of Eriugena's *De praedestinatione*.)

Fredegisus of Tours (*c.*800?) *De substantia nihili et de tenebris (On the Substance of Nothing and on Darkness)*, in Monumenta Germaniae Historica, Epistolae IV, *Epistolae Karolini Aevi II*, ed. E. Euemmler, Berlin: Weidmann, 1895, 552-5.(Considers what the referent of ‘nothing’ is.)

Gottschalk of Orbais (late 830s-840s) Works on Predestination, in C. Lambot (ed.) *Oeuvres théologiques et grammaticales de Godescalc d’Orbais*, Louvain: Spicilegium Sacrum Lovaniense, 1945.(Gottschalk sets out his theory of dual predestination, basing himself on Augustine.)

Hincmar of Rheims (849) *Ad reclusos et simplices (Hincmar’s letter against Gottschalk’s theory of predestination)*, ed. W. Gundlach, ‘Zwei Schriften des Erzbischofs Hinkmar von Reims - 2’, *Zeitschrift für Kirchengeschichte* 10, 1889: 258-309.(Hincmar’s attack on Gottschalk.)

Hrabanus Maurus (*c.*840) Letter to Bishop Noting (against Gottschalk on predestination), in J.-P. Migne (ed.) *Patrologia Latina*, Paris, 1852, vol. 112, cols. 1530-53.(Hrabanus attacks the idea of dual predestination.)

McKitterick, R. (ed.) (1994) *Carolingian Culture: Emulation and Innovation*, Cambridge: Cambridge University Press.(Essays by various specialists on Carolingian thought, music, literature, art and so on.)

Marenbon, J. (1981) *From the Circle of Alcuin to the School of Auxerre*, Cambridge Studies in Medieval Life and Thought, 3rd series, vol. 15, Cambridge: Cambridge University Press.(Logic, philosophy and theology from *circa* 790-890.)

Prudentius of Troyes (early 850s) *De praedestinatione (On Predestination)*, in J.-P. Migne (ed.) *Patrologia Latina*, Paris, 1852, vol. 119, col. 101-250.(Attack on Eriugena’s *De praedestinatione*.)

Theodulf of Orleans (*c.*788-92) *Libri Carolini (Caroline Books)*, ed. H. Bastgen, Monumenta Germaniae Historica Legum: III Concilia II, suppl., Hannover/Leipzig: Hahn, 1924.(The Western response, issued in Charlemagne’s name, to the Greek position on image worship.)

Ratramnus of Corbie (*c.*860-3) *Liber de anima (Book on the Soul)*, ed. D.C. Lambot, Namur: Godenne; Lille: Giard, 1952.(Most important for its rejection of realist views about universals.)

Remigius of Auxerre (*c.*900) Commentary on Boethius’ *De consolatione philosophiae*, extracts in E.T. Silk (ed.) *Saeculi noni auctoris in Boetii consolationem Philosophiae commentarius*, appendix, Rome: American Academy in Rome, 1935.(Discovers hidden Christian meanings in Boethius’ text.)

Cassirer, Ernst (1874-1945)

Cassirer is one of the major figures in the development of philosophical idealism in the first half of the twentieth century. He is known for his philosophy of culture based on his conception of 'symbolic form', for his historical studies of the problem of knowledge in the rise of modern philosophy and science and for his works on the Renaissance and the Enlightenment. Cassirer expanded Kant's critique of reason to a critique of culture by regarding the symbol as the common denominator of all forms of human thought, imagination and experience. He delineates symbolic forms of myth, religion, language, art, history and science and defines the human being as the 'symbolizing animal'. All human experience occurs through systems of symbols. Language is only one such system; the images of myth, religion and art and the mathematical structures of science are others.

Being of Jewish faith, Cassirer left Germany in 1933 with the rise of Nazism, going first to Oxford, then to university positions in Sweden and the USA. In the last period of his career he applied his philosophy of culture generally and his conception of myth specifically to a critique of political myths and to the study of irrational forces in the state.

1 Life

Ernst Cassirer was born in the German city of Breslau (now Wrocław, Poland) on 28 July 1874; he died suddenly, of a heart attack, on the Columbia University campus in New York on 13 April 1945. His life was a personal and intellectual 'odyssey' that took him from Europe to the USA, and led him from the Marburg Neo-Kantianism of his teacher, Hermann Cohen, to his own broad vision of human culture and a critique of the modern state. Cassirer lectured as *Privatdozent* at the University of Berlin from 1906 until 1919, when he accepted a professorship at the newly founded University of Hamburg; he served as its rector in 1929-30.

After Hitler's assumption of the chancellorship of Germany in January 1933, Cassirer left Germany. He taught from 1933 to 1935 at All Souls College, Oxford and then accepted a professorship at the University of Göteborg, Sweden. In 1941 he moved to Yale University in the USA, and then went to Columbia University for the academic year 1944-5. Cassirer published nearly 125 books, essays and reviews and left a number of unpublished papers.

2 Philosophy of symbolic forms

Cassirer published the major work of his philosophy, *Philosophie der symbolischen Formen (The Philosophy of Symbolic Forms: vol. 1, Language; vol. 2, Mythical Thought; vol. 3, The Phenomenology of Knowledge)*, from 1923 to 1929, but his conception of 'symbolic form' goes back to the philosophy of science he formed a decade earlier. Modern scientific thought, Cassirer holds, is based on the 'functional concept'. As opposed to the Aristotelian theory of concept formation, in which a common substantial element is sought through a comparison of the similarities and differences of a class of particulars, the functional concept is formed by articulating a principle by which a set of particulars can be ordered as a series. This principle of serial arrangement of a group of particulars, unlike a substance, has no reality or meaning independent of the elements it orders, and these elements have meaning only in terms of the positions they each occupy in the series. Cassirer formulated this indissoluble bond between universal and particular of the functional concept as F (a, b, c, …). It suggested to him a model for how the mind forms experience in all spheres of human activity, cognitive and noncognitive.

The historical source for this insight is Kant's idea of the 'schema', a conception of sensuous-intellectual form that is presupposed by all acts of human knowledge. What Kant delineates abstractly as one of the principles of his first *Critique*, Cassirer finds as a phenomenon within human experience: the symbol. The critique of reason becomes the critique of culture. Each area of human culture has its own way of bringing sensed particulars together in symbolic orders. Each area of culture has its own 'inner form' - its own formation of the object, its own causality, its own apprehensions of space, time and number. These various symbolic forms of culture differ from each other in their individual 'tonality', and human culture as a whole is ideally a harmony of these forms.

The symbolic forms are frequently thought of as a list, following the chapter titles of Cassirer's *An Essay on Man* (1944): myth and religion, language, art, history and science. Cassirer also suggests the possibility of additional symbolic forms, such as economics, morality and technology. In *The Phenomenology of Knowledge*, the third volume of *The Philosophy of Symbolic Forms* (1923-9), Cassirer presents three symbolic forms as corresponding

to the fundamental functions of the development of consciousness. He makes clear that he is using the term ‘phenomenology’ not in Husserl’s sense but in Hegel’s, that is, as developmental, not descriptive phenomenology. All knowledge and culture originates in the ‘phenomenon of expression’, the *Ausdrucksfunktion* of consciousness. At the level of ‘expression’, (*Ausdruck*) the object is ‘felt’ in its immediacy. Consciousness at this level takes the form of myth. Symbol and symbolized occupy the same level of reality. The dancer who dons the mask of the god *is* the god. The mythic image in its felt immediacy gives way to the logical powers inherent in language; this produces the ‘representational function’ (*Darstellungsfunktion*). This function builds a world of common-sense objects, of thing-attribute relationships and classes. Symbol and symbolized now are different orders of reality. Symbols refer to things. Beyond this is the purely ‘significative function’ (*Bedeutungsfunktion*) of scientific and theoretical thought. At this level the power of the symbol to generate ‘symbolic systems’ occurs. Here symbols can refer in fully determinate ways to other orders of symbols. The purest examples of this are mathematics and mathematical logic.

In a fourth volume of *The Philosophy of Symbolic Forms*, left incomplete in manuscript at his death, Cassirer considered ‘the metaphysics of symbolic forms’. He examined how the expressive function of consciousness is the most fundamental manifestation of spirit (*Geist*) and how spirit is a transformation of life (*Leben*). Cassirer discusses a number of conceptions of life in modern philosophy and is led to his own doctrine of ‘Basis-phenomena’ (*Basisphänomene*), the foremost of which, he claims, is life. Life is the ongoing flow of existence that is first formed by the human power of expression, out of which, as described above, arise all forms of human culture.

3 Historical studies

Cassirer did not approve of ‘hurling one’s ideas into empty space’ without showing their relation to the historical development of philosophy. Not only does he ground his original ideas in their historical sources and in the fields he discusses, he is the author of a large corpus of work in intellectual history. He wrote books and essays on Leibniz, Descartes, Kant and Rousseau, and edited a three-volume edition of Leibniz’s philosophical works as well as one of the standard editions of Kant’s works. Cassirer published essays on figures in humanist thought such as Hölderlin, Kleist, Humboldt, Schiller, Shaftesbury, Pico della Mirandola, Thomas Mann and Schweitzer. In the philosophy and history of science he wrote on Galileo, Newton, Einstein’s relativity and Bohr’s indeterminacy principle. Cassirer’s systematic interpretation of the history of philosophy is centred on two series of works: his four-volume history, *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit (The Problem of Knowledge in Philosophy and Science in the Modern Age)* (1906, 1907, 1920) and his trilogy, *Individuum und Kosmos in der Philosophie der Renaissance (The Individual and the Cosmos in Renaissance Philosophy)* (1927), *Die Platonische Renaissance in England und die Schule von Cambridge (The Platonic Renaissance in England)* (1932a) and *Die Philosophie der Aufklärung (The Philosophy of the Enlightenment)* (1932b). Cassirer began his study of the problem of knowledge intending to show how this problem develops in the simultaneous rise of modern philosophy and modern science, beginning with Nicholas of Cusa and culminating in Kant. Later he continued the theme in the post-Kantian systems through in Hegel, and much later, when in Sweden, he considered the shape of the contemporary sciences.

His trilogy of studies goes over some of the same ground as his earlier work, but in a more agile way and to a different purpose. He begins with Nicholas of Cusa, but his aim is to show how the Renaissance can be understood as a whole in terms of the problem of the individual and the cosmos. He then shows how the ideas of the Renaissance were transmitted via the Cambridge Platonists to culminate in the Enlightenment. Cassirer wishes to present a ‘phenomenology of philosophic spirit’. This is not a progression of problems of pure thought, but the generation of a philosophical point of view on the individual, the world and society, in which the problems of knowledge are tied to the whole of human activity (see Platonism, Renaissance; Renaissance philosophy).

It is not possible to understand the basis of Cassirer’s philosophy without an awareness of his debt to Goethe (see Goethe, J.W. von §3). Goethe’s understanding of organic form is important for Cassirer’s conception of the symbol, and his cosmopolitanism influences Cassirer’s conception of culture. Goethe is the source of Cassirer’s grasp of human creation as a process of self-liberation, a sentiment Cassirer extends to the whole of culture. He wrote on various aspects of Goethe’s thought, but he was influenced more by Goethe’s spirit and sense of life than by his interest in particular questions of interpretation.

4 Myth and the state

Cassirer does not have a political philosophy in the traditional sense. He has a critique of the modern state and a definite view of the relation of philosophy to politics. In his inaugural lecture at Göteborg in 1935, Cassirer recalled Kant's distinction between a 'scholastic' conception of philosophy and a conception of 'philosophy as related to the world' (*conceptus cosmicus*). Cassirer said that he as well as others have been guilty of the former and he aligns himself with the latter. He quotes Schweitzer, his ethical hero, who calls philosophy a 'watchman' who slept in the hour of peril, and did not keep watch over us during the rise of totalitarianism. Cassirer holds that philosophy does not and cannot cause the events of political life, nor can it resolve them, but it has a duty to act as our conscience, to inform us of them by the use of its powers of reflection. In *The Myth of the State* (1946) and other writings of this period, Cassirer attacks Heidegger's conception of 'thrownness' (*Geworfenheit*) as a conception of the human condition that puts philosophy in a position where it can no longer 'do its duty'. This goes back to Cassirer's debate with Heidegger in 1929 at Davos, Switzerland. Quoting Goethe, Cassirer sees human freedom as tied to the human project of spirit (*Geist*), of the creation of culture. Heidegger sees freedom as requiring a 'breakthrough' (*Einbruch*); freedom is not part of the human condition itself, but is contingent (*zufällig*) (see Heidegger, M. §5).

Cassirer's philosophy of mythology is the most original part of his epistemology and phenomenology of knowledge. He shows that myth is not a collection of errors or a world of unchecked emotions but a total way of thinking and symbolizing, which exists at the beginning of human culture and is present as a phase in the development of any subsequent symbolic form. Myth is always present as the expressive moment in any act of cognition. In *An Essay on Man*, Cassirer claims that his philosophy of culture is an extension of the ancient ideal of self-knowledge and that human culture as a whole is the process of humanity's 'self-liberation'. The key to self-knowledge and culture is freedom from the immediacy of the object.

In *The Myth of the State*, Cassirer is able to apply the force of his entire philosophy of culture towards understanding the logic of modern political myths. They are a revival of the logic of the primordial forms of expressive consciousness. Modern political myths are not natural; they are manufactured products joined to the technology of mass communication. Such myths shape the life of the state and become the substitute for its rational principles. Cassirer sees this not only as true of Nazism, but also as a danger in the modern state itself. He claims that myth is impervious to argument, but that philosophy can warn us of it and allow us to understand it.

See also: State, the §1

DONALD PHILLIP VERENE

List of works

Cassirer, E. (1906, 1907, 1920) *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit (The Problem of Knowledge in Philosophy and Science in the Modern Age)*, Berlin: Bruno Cassirer, 3 vols. (History of the problem of knowledge from Nicholas of Cusa through in Hegel and Neo-Kantian systems of philosophy.)

Cassirer, E. (1910) *Substanzbegriff und Funktionsbegriff. Untersuchungen über die Grundfragen der Erkenntniskritik*, Berlin: Bruno Cassirer; trans. W.M. Swabey and M.C. Swabey, in *Substance and Function and Einstein's Theory of Relativity*, Chicago, IL: Open Court, 1923.(Cassirer's philosophy of science, showing how the functional concept reflects the thought-form of modern science.)

Cassirer, E. (1923-9) *Philosophie der symbolischen Formen*, Berlin: Bruno Cassirer, 3 vols; trans. R. Manheim, *The Philosophy of Symbolic Forms*, New Haven, CT: Yale University Press, 3 vols, 1955-7: vol. 1, *Language*; vol. 2, *Mythical Thought*; vol. 3, *The Phenomenology of Knowledge*.(The major statement of Cassirer's systematic philosophy.)

Cassirer, E. (1925) *Sprache und Mythos. Ein Beitrag zum Problem der Götternamen*, vol. 6 of *Studien der Bibliothek Warburg*, Leipzig: Teubner; trans. S.K. Langer, *Language and Myth*, New York: Harper & Bros, 1946.(Study of the interconnections of the symbolic forms.)

Cassirer, E. (1927) *Individuum und Kosmos in der Philosophie der Renaissance*, vol. 10 of *Studien der Bibliothek Warburg*, Leipzig: Teubner; trans. M. Domandi, *The Individual and the Cosmos in Renaissance Philosophy*, New York: Harper & Row, 1963.(The major themes that unify Renaissance philosophy, with an emphasis on Nicholas of Cusa.)

Cassirer, E. (1932a) *Die Platonische Renaissance in England und die Schule von Cambridge*, vol. 24 of *Studien*

der Bibliothek Warburg, Leipzig: Teubner; trans. J.P. Pettegrove, *The Platonic Renaissance in England*, Austin, TX: University of Texas Press, and Edinburgh: Nelson, 1953.(Emphasis on the Cambridge Platonists and Shaftesbury.)

Cassirer, E. (1932b) *Die Philosophie der Aufklärung*, Tübingen: Mohr; trans. F.C.A. Koelln and J.P. Pettegrove, *The Philosophy of the Enlightenment* , Princeton, NJ: Princeton University Press, 1951.(The unity of the Enlightenment's views of nature, religion, aesthetics and the state.)

Cassirer, E. (1942) *Zur Logik der Kulturwissenschaften. Fünf Studien*, vol. 47 of *Götesborgs Högskolas Arsskrift*, Göteborg: Wettergren & Kerbers; trans. C.S. Howe, *The Logic of the Humanities*, New Haven, CT: Yale University Press, 1961.(Five essays on the nature of the humanities, especially the differences between concept-formation in the sciences and the humanities.)

Cassirer, E. (1944) *An Essay on Man: An Introduction to a Philosophy of Human Culture*, New Haven, CT: Yale University Press.(Cassirer's summary of his philosophy of symbolic forms and the conception of human nature presupposed by it.)

Cassirer, E. (1946) *The Myth of the State*, New Haven, CT: Yale University Press.(Historical and systematic analysis of the modern state employing Cassirer's conception of mythical thought.)

Cassirer, E. (1950) *The Problem of Knowledge: Philosophy, Science and History since Hegel*, trans. from the German manuscript of 1940 by W.H. Woglom and C.W. Handel, New Haven, CT: Yale University Press; German edn, *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit. Von Hegels Tod bis zur Gegenwart (1832-1932).* Stuttgart: Kohlhammer, 1957.(The final volume of *Das Erkenntnisproblem*, see above.)

Cassirer, E. (1979) *Symbol, Myth, and Culture: Essays and Lecturers of Ernst Cassirer 1935-1945*, ed. D.P. Verene, New Haven, CT: Yale University Press.(Essays and lectures from the last decade of Cassirer's life left in manuscript at the time of his death.)

Cassirer, E. (1995) *Zur Metaphysik der symbolischen Formen*, ed. J.M. Krois, vol. 1 of *Nachgelassene Manuskripte und Texte*, ed. J.M. Krois and O. Schwemmer, Hamburg: Meiner; trans. and ed. J.M. Krois and D.P. Verene as *The Metaphysics of Symbolic Logic*, vol. 4. of *The Philosophy of Symbolic Forms*, New Haven, CT: Yale University Press, 1996.(The fourth volume of Cassirer's major work (written in 1928), with further text on the conception of 'basis-phenomena' (written *c*.1940) left in manuscript at his death.)

References and further reading

Braun, H.-J. *et al*. (eds) (1988) *Über Ernst Cassirer's Philosophie der Symbolischen Formen (On Ernst Cassirer's Philosophy of Symbolic Forms)*, Frankfurt: Suhrkamp.(Essays by European scholars on major themes in Cassirer's philosophy. From an international conference held in 1986 in Zurich.)

Eggers, W. and Mayer, S. (1988) *Ernst Cassirer: An Annotated Bibliography*, New York: Garland. (Comprehensive bibliography of writings by and about Cassirer; annotations are uneven.)

Hamburg, C.H. (1964) 'A Cassirer-Heidegger Seminar', *Philosophy and Phenomenological Research* 25: 208-22. (Account of the exchange between Cassirer and Heidegger in 1929 at Davos, Switzerland.)

Heidegger, M. (1928) Review of *Das mythische Denken* by E. Cassirer, *Deutsche Literaturzeitung* 21: 1000-12; trans. J.G. Hart and J.C. Maraldo, 'Book Review of Ernst Cassirer's *Mythical Thought*', in *The Piety of Thinking: Essays by Martin Heidegger*, Bloomington, IN: Indiana University Press, 1976.(Heidegger's review of volume 2 of The Philosophy of Symbolic Forms.)

Krois, J.M. (1987) *Cassirer: Symbolic Forms and History*, New Haven, CT: Yale University Press.(Definitive study of the various aspects of Cassirer's philosophy.)

Langer, S.K. (1974) 'De Profondis', *Revue Internationale de Philosophie* 28: 449-55.(Comparison of Cassirer's and Freud's views of the symbol, by a philosopher who was greatly influenced by Cassirer.)

Schilpp, P.A. (1949) *The Philosophy of Ernst Cassirer*, Evanston, IL: The Library of Living Philosophers, Inc. (Twenty-three essays on aspects of Cassirer's thought, a biographical essay and a bibliography of his writings to 1946.)

Verene, D.P. (1969) 'Kant, Hegel and Cassirer: The Origins of the Philosophy of Symbolic Forms', *Journal of the History of Ideas* 30: 33-46.(Assesses the extent to which both Kant and Hegel are sources for Cassirer's philosophy of symbolic forms.)

Casuistry

Casuistry, from the Latin casus *(cases), has been understood in three separate yet related senses. In its first sense casuistry is defined as a style of ethical reasoning associated closely with the tradition of practical philosophy influenced by Aristotle and Aquinas. In its second sense it is reasoning about 'cases of conscience' (*casus conscientiae*). The third sense, moral laxism, arose out of Pascal's famous critique of casuistry, which did much to diminish its influence. In recent years, however, a renewed interest in the first and second senses of casuistry has been witnessed in the areas of practical reasoning and applied philosophy.*

In its widest sense, casuistry can be described as a method of ethical reasoning which, drawing on the tradition of practical philosophy of Aristotle and Aquinas, aims to construct a 'dialectic' between the facts of particular cases and the antecedent assumptions, evaluations and convictions which individual agents bring to bear in their consideration of such cases. The purpose of the dialectic is to enable agents to arrive at informed decisions as to what is morally possible and impossible for them to do in particular cases. In a narrower sense, the term casuistry has been employed to characterize different systems of moral theology within the Christian, Jewish and Islamic traditions, in which all-inclusive norms are derived from judgments in particular cases, instead of being laid down in advance by absolute moral codes. In its narrowest sense, casuistry refers to the use of subtle definitional distinctions in the handling of the problems of moral theology, with the aim of drawing fine dividing lines between what is and is not permissible at the level of action. The technique has at times been used to excuse crimes and sins, thereby exculpating the immoral, and such is the extent of the modern association of casuistry with all varieties of obfuscation, quibbling and laxism, that a pejorative connotation of the word itself is now established in most European languages.

The identification of casuistry with its narrowest definition is due to Blaise Pascal's *Lettres écrites à un provincial (The Provincial Letters)* (1656-7) (see Pascal, B.). This famous satire delivered to casuistry a near fatal blow, from which, in the popular mind, it has never recovered. It was Pascal's urbane yet brutal vilification of the Jesuit confessors of the University of Paris, individuals who had attempted to base an account of practical conduct on the analysis of cases (*casus*) and circumstances (*circumstantiae*), that stereotyped casuistry as the doctrine that sought 'to excuse the inexcusable'.

Pascal's attack on casuistry can be considered unjust since it was based on a partial understanding of the theories he was aiming to expose, and of the individual casuists he was attempting to ridicule. His aim in writing this work was to enlist the reader's support for a version of moral absolutism that was peculiar to Jansenism, a rigoristic Roman Catholic sect to which Pascal was aligned. Pascal was forever trying to illustrate the laxism inherent within Jesuit casuistry in order to establish an unfavourable contrast with the moral rigorism of the Jansenists.

The method embodied in the second definition of casuistry as case-reasoning was developed in the penitential handbooks of the Middle Ages and reached its fullest expression in Roman Catholic textbooks of moral theology of the Counter-Reformation. The method was also embraced by a number of Anglican divines and members of the Reformed tradition. The impetus behind the case-method lay in the desire among theologians and philosophers to discover the moral norms embodied in divine law in the circumstances of human life, rather than finding them in antecedent absolute norms which one could then apply to a set of cases. Thus, the case-reasoning of the theologians was able to respond to new situations and novel problems with a commendable degree of good sense, originality and doctrinal flexibility.

In recent times the case-method of the theologians has been revived in the field of bioethics (see Bioethics). For bioethicists the value of casuistry resides in its emphasis of practical problem-solving by means of a nuanced interpretation of individual cases. This is thought to have greater consonance with the actual conditions of medical practice and decision-making. Problems arise, however, when one attempts to achieve the transposition of the case-method of the theologians to the modern secular context of bioethics. The confessors of Roman Catholic casuistry, in particular, possessed both juridical authority and the authority of being experts within an institutional setting. Yet the new casuists of bioethics must rely on professional knowledge alone to determine morally adequate judgments in particular cases. The problem here is one of analogy. The older casuistical method relied upon interpretative authorities such as confessors and canonists which helped to bolster the idea that the judgments

made in particular cases would reflect the assumptions of a common morality. Given the context of moral pluralism in which bioethical judgments are made, it is difficult to see how bioethics can draw upon the same type of paradigmatic examples of 'right' and 'wrong' moral judgment on which the earlier model of casuistry was based (see Moral pluralism; Moral expertise).

It is in its widest sense, in the context of moral philosophy, that casuistry holds its greatest promise. As a dialectical method of moral reasoning inspired by Aristotle and Aquinas, casuistry may well prove relevant here in virtue of its opposition to inflexible and literal interpretations of moral principles, and its resistance to any ethical attitude which first absolutizes a universal moral norm and then insists on its all-round and unyielding application, while denying any abatement or adjustment to changing contingencies.

See also: Logic of ethical discourse; Moral Judgment; Situation ethics

MARTIN STONE

References and further reading

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, ed. T. Gilby, London: Eyre & Spottiswoode, 1966, vol. 28, IaIIae.90-97. (Questions 94 and following cover Aquinas' remarks on natural law. Many casuists conjoined Aquinas' remarks on natural law to Aristotle's account of the nature of ethics.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985, books I 3, V. (Provides discussion of the nature of ethics, and considers the importance of equity in law, both of which exerted a significant influence on the development of casuistry.)

Jonsen, A.J. and Toulmin, S. (1988) *The Abuse of Casuistry*, Berkeley, CA: University of California Press. (An influential defence of the casuistical tradition, particularly influential in applied philosophy, especially recent bioethics.)

Keenan, J.F. and Shannon, T.A. (eds) (1995) *The Context of Casuistry*, Washington, DC: Georgetown University Press. (A rich collection of articles which discuss the roots, practice and legacy of casuistry in moral philosophy and theology.)

Leites, E. (ed.) (1988) *Conscience and Casuistry in Early Modern Europe*, Cambridge: Cambridge University Press. (A collection of articles that illuminate the historical context in which casuistry was practiced.)

Pascal, B. (1656-7) *Lettres écrites à un provincial*, trans. A.J. Krailsheimer, *The Provincial Letters*, Harmondsworth: Penguin, 1967, letters VI, VII. (A strong attack on Jesuit doctrines.)

Stone, M.W.F. (1998) *The Subtle Art of Casuistry: An Essay in the History of Moral Philosophy*, Oxford: Oxford University Press. (A defence of casuistry from the standpoint of recent moral philosophy. Emphasizes the Aristotelian-Thomistic roots of the casuistical method in contrast to other authors, most notably Jonsen and Toulmin.)

Categories

Categories are hard to describe, and even harder to define. This is in part a consequence of their complicated history, and in part because category theory must grapple with vexed questions concerning the relation between linguistic or conceptual categories on the one hand, and objective reality on the other. In the mid-fourth century BC,Aristotle initiates discussion of categories as a central enterprise of philosophy. In the Categories he presents an 'ontological' scheme which classifies all being into ten ultimate types, but in the Topics introduces the categories as different kinds of predication, that is, of items such as 'goodness' or 'length of a tennis court' or 'red', which can be 'predicated of' subjects. He nowhere attempts either to justify what he includes in his list of categories or to establish its completeness, and relies throughout on the unargued conviction that language faithfully represents the most basic features of reality. In the twentieth century, a test for category membership was recommended by Ryle, that of absurdity: concepts or expressions differ in logical type when their combination produces sentences which are palpable nonsense.Kant, working in the eighteenth century, derives his categories from a consideration of aspects of judgments, hoping in this manner to ensure that his scheme will consist exclusively of a priori concepts which might constitute an objective world. The Sinologist Graham argues that the categories familiar in the West mirror Indo-European linguistic structure, and that an experimental Chinese scheme exhibits suggestively different properties, but his relativism is highly contentious.

1 Categories in Aristotle

Despite the historical importance of category theory in Western philosophy, it is remarkably difficult to grasp what a category is and how a category theory might achieve legitimacy. To a degree, such elusiveness is simply the product of the lengthy and elaborate historical development these theories have sustained, but the profundity of the issues they broach also contributes to the puzzle - and hence to the abiding interest - of categories.

Orthodoxy has it that Aristotle in the treatise appropriately entitled *Categories* introduces categories to the Western tradition: ' of things said without any combination, each signifies either substance or quantity or qualification or a relative or where or when or being-in-a-position or having or doing or being-affected' (1b 25-7). Despite what his phrasing might easily be (and often has been) taken to imply, Aristotle plainly intends to refer, not to linguistic items, but rather to their referents. Thus his sentence might be paraphrased 'of things to which we can refer in uncombined (namely syntactically simple) language, each is.….', and it enumerates all the types of being there are: substances, such as human beings; quantities, such as three feet long; qualities, such as red; and so forth. But although his manner of expression is in one sense misleading, it should not be deplored or dismissed as merely something like a trivial confusion of material and formal modes of speech. Rather, here, as so often elsewhere, Aristotle is moving from what we would regard as features of language to features of the world without so much as signalling any recognition that there is a transition to be made. Nor does he explain how he arrived at his list, or why one should feel confident that it is comprehensive (see Aristotle §7).

To complicate matters further, there are compelling grounds for the belief that, despite the tradition, the genera of being with which the *Categories* deals are not (Aristotelian) categories. The word which we transliterate 'category' is employed by Aristotle to mean 'predicate' or '(type of) predication', so that a 'category theory' ought to be a theory of predication, not of being. And in a relatively unfamiliar text, *Topics* (I 9), this is just what we do find. After saying that the types of predication must be distinguished, Aristotle provides a list identical to that in the *Categories* - except that the first item is not 'substance', but 'what it is' (103b 21-3). In the *Categories*, individuals such as Socrates are substances pre-eminently, and species such as human being are substances secondarily: in the *Topics*, in contrast, colour is the 'what it is' of red no less than human being is the 'what it is' of Socrates.

What the *Topics* delivers, then, is a theory of predication according to which the predicates which can characterize any subject whatsoever will fall into ten ultimate kinds. Again, defence of this list in particular (or of the possibility of constructing *any* non-arbitrary, correct list) is lacking; but reflection on why the *Topics* was written might make the absence of this defence considerably less shocking, though at the cost, ironically enough, of undermining the status which the categories of the *Categories* have historically enjoyed. The *Topics* as a whole comprises prescriptions for the classification and analysis of a vast array of the sorts of argument then current in

Greek philosophy. The intention is combative as well as constructive: Aristotle also teaches how to detect and dismember what he considers sorts of fallacious reasoning, and prime among these is ignorant or malicious exploitation of confusion in manner of predication. For example, what is good about food? That it does something - it produces pleasure. What is good about human beings? That they are 'of a certain quality, such as temperate or courageous or just' (*Topics* I 15.107A). So, if viewed as a digest of types of predication already familiar from the practices of Greek dialectic, the categories of the *Topics* require no defence beyond reasonable fidelity to the range of predicates actually found in the typical dialectical repertoire. In particular, there need be no presumption that predicational categories rooted in philosophical practice will correspond to a significant, let alone universal, ontological classification.

This is not to say that predicational and ontological categories are unrelated. Because Aristotle moves with disconcerting freedom from language to the world, he takes it for granted that predicate expressions do usually, without the possibility of radical misrepresentation, refer to real entities. So a scheme of predicational categories might indeed suggest at least the outline of a corresponding ontological scheme, one justified in detail and scope precisely to the extent that the original classification of predicates, and its extension, are well-founded. If this is the correct account of how these theories developed in Aristotle, what prompted the substitution of ontological 'substance' for predicational 'what it is'? Perhaps the governing idea of the *Categories* is that substance alone can serve as the subject in which items from other categories inhere: for instance, Socrates, but not his height, is pale-skinned and can properly be said to be pale-skinned. Therefore the design of the treatise is to convince us that, at its deepest level, ontology is marked by a crucial asymmetry between substantial and non-substantial being, quite possibly in a spirit hostile to the elevation of Platonic Forms which, far from being most real, would figure as no more than non-substantial, parasitic things. If so, a further irony is that the *Categories* was intended to establish a metaphysical thesis essentially independent of the grand classificatory ambitions with which it came to be historically associated. In the event, the philosophical tradition which was dominated by the followers of Aristotle since late antiquity, elevated the ontological categories into a system uniquely capable of displaying the lineaments of what is (see Being §2; Substance).

2 Categories in Ryle

The most noteworthy modern contribution to the theory of categories is that of Gilbert Ryle, in that he tries to throw light on the difficulty which Aristotle left entirely obscure: how to identify and discriminate between categories (see Ryle, G. §2). Ryle contends that one determines the logical category or type of an expression by ascertaining the field of sentence-forms into which it can enter without resultant absurdity. For example, 'four' can complete the frame '.….is a prime number' to yield a meaningful (if false) sentence, while 'Socrates', or so Ryle would urge, cannot. Such nonsense exemplifies what Ryle famously dubbed the 'category-mistake', the production of absurdity as the upshot not of lexical or grammatical irregularity but of the vain effort to combine the logically uncombinable. He conceived of category theory as a diagnostic tool for the exposure and resolution of chronic philosophical disputes. Ryle proposed that disputants on either side of, say, the mind-body problem are really at cross-purposes: they have put forward propositions which only seem to conflict, since their difference in type removes the possibility of any logical relations, whether of implication or of incompatibility, holding between them. At one point Ryle was willing to go so far as to proclaim that ' philosophy is the replacement of category-habits by category-disciplines' (1949: 10).

Although Ryle regarded Aristotle's theory as the ancestor of his own, an obvious dissimilarity is that Rylean categories, defined as they are by multiple logical relations, are potentially unlimited in number: Ryle was keen to explore what he termed 'the logical geography of concepts', not to engage in pigeonholing. The great awkwardness looming over Ryle's theory is that all ventures to make out a principled difference between mere falsehood and the nonsense allegedly distinctive of category-mistakes have come to grief. But if Ryle did not finally achieve a strict criterion for the identification of categories, he nevertheless did successfully sharpen the powerful and once widely-shared intuition that certain propositions - some of philosophical importance - must be rejected not as false, but as (veiled) nonsense.

3 Categories in Kant

The basic problem of specifying what a category theory is, is unavoidably aggravated by Kant, the second major historical influence to shape the tradition (see Kant, I. §4). Kant arrives at his alternative categories by considering

the conditions which allow there to be a logic of judgments, a procedure which he explicitly attributes to Aristotle, while emphasizing the divergence of their conclusions (Kant 1781/7: A80, B105). He asks: Which judgement forms does logic treat of? His answer is that all judgements have a 'quantity' (universal, particular or singular) and a 'quality' (affirmative, negative or infinite). They must all instantiate one of the 'relations' (categorical, hypothetical or disjunctive), and one of the 'modalities' (problematic, assertoric or apodeictic) (Kant 1781/7: A70, B95). Kant proposes to derive his categories from these forms or aspects of judgment, obtaining the corresponding list of unity, plurality and totality under 'quantity'; reality, negation and limitation under 'quality'; inherence and subsistence, causality and dependence, and community under 'relation'; and possibility-impossibility, existence-non-existence and necessity-contingency under 'modality' (Kant 1781/7: A80, B106). He pretends that their derivation from the types of judgement makes his categories superior to Aristotle's, because this procedure supposedly guarantees that all the concepts of pure understanding, and no others, are systematically discovered.

Every element of empirical reality must fall into place within this scheme, since the categories exhaust the understanding's a priori concepts. Kant's eventual aim in the Transcendental Deduction is to prove that, since every experience involves judgement, every possible experience must involve use of the categories. Specifically, his claim is that one must conceive of experience as experience of a world organized by the categories.

The conviction that Aristotelian genera of being and Kantian concepts of understanding cannot, *pace* Kant, be in competition, is hard to avoid since the philosophical projects which gave a point to these various theories are so diverse. But we should not hastily conclude that these schemes have nothing of philosophical importance in common. Aristotle's transitions from what we are likely to classify as linguistic or conceptual data to robustly realist theses seem surprisingly bold, even unmotivated; but that is precisely because we are heirs to the Kantian legacy which insists on our right to empirical realism but concedes the obligation to earn it by espousing transcendental idealism.

Aristotle blithely assumes that the fabric of the world is accessible to us, disclosed by the range of predications people, at any rate those of a philosophical disposition, have been disposed to make; he mounts no argument for the reliable accuracy of predication. True, it is not as if later Aristotelians ignored the issue of what the categories are categories of. Because the *Categories* was incorporated into the *Organon* - which came to be the traditional logic course, culminating in the analysis of arguments - some interpreters were indeed inclined to the position that the *Categories*, which began the logic course, ought to deal with the simples from which propositions making up arguments are constructed, and so favoured the identification of categories with words. Other strands in the tradition preferred a realistic interpretation, while nevertheless asserting a thorough-going isomorphism between the structures of the world, our concepts and the words in which we express ourselves (this 'isomorphic' interpretation is vividly and compendiously presented in, for example, the Renaissance *Conimbricenses* commentary on the *Organon*). Still, even Aristotelians sensitive to the question 'what are categories categories of?', neglected the gulf which might separate our conceptual and linguistic resources from what really is: such isomorphism is viable only if grounded in a pre-modern assurance that an impersonal teleology or a divine providence secures and protects the fit between us and the world.

Kant's transcendentalism attempts to redeem such hostages to scepticism. We might accordingly conceive of both theories of categories as intimately concerned with the relation of language or our conceptual equipment to the world. To the Kantian, Aristotle seems primitive because he naïvely fails to perceive the paramount challenge confronting philosophy; to the modern follower of Aristotle, the Kantian seems the decadent victim of unjustifiable doubts.

4 Are categories universal?

The impression that Aristotelian and Kantian theories have a vital disagreement in common is strengthened if we move beyond the Western tradition in which they are such towering presences. Angus Graham (1989) endorses the contention, most frequently linked with the name of the famous linguist Benveniste, that the Aristotelian categories represent nothing more than yet another bogus reification of Indo-European linguistic forms. For example, 'quantity' in Aristotle's list is the translation of a Greek word which functions as both an interrogative ('how much?') and an indefinite adjective ('so much'). Thus it is argued that peculiarities of Indo-European grammar have encouraged the construction and success of a categorical scheme which, viewed from a broader linguistic perspective, is nothing more than the deluded inflation of quite parochial traits of familiar languages into general

features of reality itself.

Graham goes on to try to substantiate this charge of Indo-European parochialism and to show that moderate versions of 'linguistic relativism' are acceptable - that is, that the narrowly linguistic properties of a language (its syntax, its degree of inflection and so on) encourage or discourage certain thought-patterns in its users. He sketches a scheme of Chinese categories derived from Chinese interrogative patterns, on the assumption that Aristotle had unreflectingly extrapolated his from Greek ones. For example, if the Aristotelian category of substance is the natural linguistic correlate of the question 'what is it?', the analogous Chinese interrogative *ho* solicits an answer in terms of *lei*, kind or type of thing (for example, human) rather than essence or basic nature (for example, rational biped). His global hypothesis is that the Chinese preference for 'which?' over 'what?' questions betrays a holistic cast of mind, felt throughout Eastern philosophy, originating in the inevitable congruence of Chinese thought with Chinese language.

Graham's version of linguistic relativism is arresting but very hard to vindicate. Quite apart from the obstacle that in the *Topics* the category 'what it is' groups non-substantial subjects as well as substances, Graham's relativistic interpretation presupposes that Aristotle proceeded in blinkered ignorance of the (alleged) fact that his native language variously stimulated and inhibited his tendency to propound ontological doctrines. But inspection of his works amply demonstrates that Aristotle is acutely aware of the linguistic traps into which philosophers are prone to fall. After all, in large measure the goal of the *Topics* is to alert us to such dangerous possibilities, and Aristotle's exercises in linguistic analysis often reach a pitch of extreme sophistication. This is not to say that gauging the impact of language on category theories is fruitless: only that they must not be reduced to a paltry imposition of fortuitous linguistic structure onto a world actually untouched by such strictly local projections. Relativistic objections do not foreclose the debate between Aristotelian confidence and Kantian caution.

See also: Ontology in Indian philosophy; Universals

ROBERT WARDY

References and further reading

Aristotle (4th century BC) *Categories*, trans. and notes J. Ackrill, Oxford: Clarendon Press, 1979.(Discussed in §1 above.)

Aristotle (4th century BC) *Topics*, Oxford: Clarendon Press, 1958.(Discussed in §1 above.)

da Fonseca, P. (1606) *In universam dialecticam*, Coimbra.(Renaissance commentary on the *Organon*, written by several authors of whom de Fonseca is the sole one known.)

Frede, M. (1987) 'The Title, Unity, and Authenticity of the Aristotelian *Categories*', in *Essays in Ancient Philosophy*, Oxford: Clarendon Press.(Magisterial exposition of the account relating the *Categories* and *Topics*, adopted in §1 above.)

Graham, A.C. (1989) *Disputers of the Tao*, La Salle, IL: Open Court.(Appendix 2 is the source for the relativistic arguments relayed in §4.)

Kant, I. (1781/7) *Critique of Pure Reason*, trans. N.K. Smith, London: Macmillan, 1929. (Discussed in §3.)

Ryle, G. (1937-8) 'Categories', *Proceedings of the Aristotelian Society*, reprinted in A.G.N. Flew (ed.) *Logic and Language*, 2nd series, Oxford: Blackwell, 1959.(Pithily brilliant classic introduction of the views outlined in §2.)

Ryle, G. (1949) *The Concept of Mind*, Harmondsworth: Penguin, 1978.(The most sustained and influential attempt to do philosophy by uncovering category-mistakes.)

Reding, J.-P. (1986) 'Greek and Chinese Categories', *Philosophy East and West* 36: 349-74. (Contains a thoughtful critique of Graham's relativism.)

Sommers, F. (1963) 'Types and Ontology', *Philosophical Review* 72: 327-63.(One of a series of works in which Sommers attempts to formalize and improve on criteria for types put forward by Ryle and others.)

Category theory, applications to the foundations of mathematics

Since the 1960s Lawvere has distinguished two senses of the foundations of mathematics. Logical foundations use formal axioms to organize the subject. The other sense aims to survey 'what is universal in mathematics'. The ontology of mathematics is a third, related issue.

Moderately categorical foundations use sets as axiomatized by the elementary theory of the category of sets (ETCS) rather than Zermelo-Fraenkel set theory (ZF). This claims to make set theory conceptually more like the rest of mathematics than ZF is. And it suggests that sets are not 'made of' anything determinate; they only have determinate functional relations to one another. The ZF and ETCS axioms both support classical mathematics.

Other categories have also been offered as logical foundations. The 'category of categories' takes categories and functors as fundamental. The 'free topos' (see Lambek and Couture 1991) stresses provability. These and others are certainly formally adequate. The question is how far they illuminate the most universal aspects of current mathematics.

Radically categorical foundations say mathematics has no one starting point; each mathematical structure exists in its own right and can be described intrinsically. The most flexible way to do this to date is categorically. From this point of view various structures have their own logic. Sets have classical logic, or rather the topos Set has classical logic. But differential manifolds, for instance, fit neatly into a topos Spaces with nonclassical logic. This view urges a broader practice of mathematics than classical.

This article assumes knowledge of category theory on the level of Category theory, introduction to §1.

1 Categorical thinking

The notion of 'isomorphism' aptly illustrates categorical thinking. In algebra two groups with the same group structure are 'isomorphic'. Two spaces with the same topological structure are 'homeomorphic'.

Before category theory, group isomorphism was defined group-theoretically, homeomorphism topologically, and so on. The definitions are unified in the categorical one (see Category theory, introduction to for an introduction to the concepts and notation). In any category, arrows $f : A \to B$ and $g : B \to A$ are called 'inverse' to one another if $gf = 1_A$ and $fg = 1_B$. Objects *A* and *B* are isomorphic if there are inverse arrows between them.

Category theory generally defines objects only up to isomorphism. That is, a categorical property will not have a unique instance in a category A but all its instances in A will be isomorphic and any object isomorphic to an instance will be an instance. For example, a 'terminal object' in a category A is an object *T* such that each object of A has exactly one arrow to *T*. A category A may have any number of terminal objects or none. Suppose *T* and T' are two such. Then there are arrows $T \to T'$ and $T' \to T$, and they are inverse since 1_T is the only arrow from *T* to *T* and similarly for T'. Hence *T* and T' are isomorphic. Conversely if *T* is terminal and there is an isomorphism $T \to T'$ then every object *B* of A has a unique arrow $B \to T \to T'$ to T', so T' is terminal.

2 The elementary theory of the category of sets (ETCS)

The 'elementary theory of the category of sets' (ETCS) axiomatizes the category Set (see Category theory, introduction to §1). We give the axioms in the next section.

In ETCS a 'singleton' is any set terminal in Set. We pick one and call it 1. We define an 'element' *x* of a set *A* to be a function from 1 to *A*, $x : 1 \to A$. Composing with any $f : A \to B$ gives an element of *B*, $fx : 1 \to B$. Trivially, each singleton has exactly one element.

We define a 'product' of sets *B* and *C* to be a set $B \times C$ and functions $p_1 : B \times C \to B$ and $p_2 : B \times C \to C$ such that for any set *A* and functions $f : A \to B$ and $g : A \to C$ there is a unique $u : A \to B \times C$ with $p_1 u = f$ and $p_2 u = g$. We can use $\langle f, g \rangle$ to name *u*, so $p_1 \langle f, g \rangle = f$ and $p_2 \langle f, g \rangle = g$. Thus elements $\langle x, y \rangle$ of $B \times C$ correspond to elements *x* of *A* and *y* of *B*.

Now 1 is only defined only up to isomorphism, as is the product $B \times C$, p_1, p_2. This is no laxity on our part. If a set *S* has any property expressible in ETCS without using names then so does every set isomorphic to *S*. Thus 1 is

distinguished from other singletons only by its name. We can pick a product of *B* and *C* to call $B \times C,\ p_1,\ p_2$ but it is distinguished from other products of *B* and *C* only by its name.

The elements of an ETCS set are definite in number but without distinguishing properties. Each is merely an element of *S* and is itself and no other. Like Cantor's cardinals ETCS sets are distinguished only by cardinality and their elements are 'mere units' (see Lawvere 1994).

This is a point of controversy. In practice even in Zermelo-Fraenkel set theory (ZF) where sets can be uniquely defined we rarely bother to do so. We say 'Let *S* be a singleton' without saying which one. We speak of 'the product $B \times C$' without specifying its elements (that is, without choosing among the many ways to code ordered pairs $\langle b, c\rangle$ as sets in ZF). But some people claim this will not do in foundations (see Mayberry 1988).

Others favour such indeterminacy. Since Dedekind raised the issue for modern mathematics, some people claim mathematical objects should have only relevant properties. Natural numbers, say, should have only arithmetic properties. Every ZF set has non-arithmetic properties given by its particular elements. So on this view numbers cannot be ZF sets. But ETCS models of arithmetic have no irrelevant properties. So this view could say the numbers form an ETCS set (see Benacerraf 1965).

Specifically, 'recursion data' on a set *S* means an element *x* of *S* and a function $f : S \to S$. We define a 'natural number object' to be a set *N* with recursion data 0 and $s : N \to N$, called zero and successor, such that: for any recursion data *x*, *f* on any *S* there is a unique function $u : N \to S$ with

$$u(0) = x$$
$$u(sn) = f(u(n)) \text{ for all } n \in N$$

Provably in ETCS there are infinitely many different natural number objects but all are isomorphic. None has any properties but the shared ones which follow from the definition (see McLarty 1994).

A 'subset' of an ETCS set *A* is any one-to-one function $i : S \to A$, that is, a function such that for all elements *x* and *y* of *S*, $ix = iy$ implies $x = y$. An element *x* of *A* is a 'member' of *i* if there is some element *z* of *S* with $iz = x$. We often write $S \rightarrowtail A$ for a subset, leaving the function *i* implicit. A 'relation' between sets *A* and *B* is a subset $R \rightarrowtail A \times B$ of the product. If $\langle x, y\rangle : 1 \to A \times B$ is a member of the subset we say *x* and *y* stand in the relation *R* or we write *xRy*.

A 'power set' of a set *B* is a set PB and an internal membership relation $\in^B \rightarrowtail B \times PB$ which ties subsets of *B* to elements of PB as follows: for each set *A* and relation $H \rightarrowtail B \times A$ there is exactly one function $h : A \to PB$, and vice versa, such that for all *x* in *A*, an element *y* of *B* has yHx if and only if $y \in^B hx$. Through *H*, each *x* in *A* defines a subset of *B* containing all *y* with *yHx*. Then *hx* is the element of PB tied to that subset. Power sets are defined up to isomorphism.

Now take any natural number object *N* and choose a power set $PN,\ \in^N$ for it, and then a power set for PN, and so on for as many levels as we want to work with. From these we can construct sets of rational, real and complex numbers in the usual way of arithmetized analysis. From here on ETCS works very much like ZF.

Mathematical practice, or even the typical set theory chapter in a maths text, relies on too few details and too much abuse of notation to say whether it is using ETCS or ZF. It can be cleaned up in either system. Yet ETCS claims to begin closer to practice. The basic ZF devices of transfinite accumulation, and coding pairs and numbers as sets, serve foundations only and are forgotten in practice. The basic devices of ETCS (the category axioms, finite limits, universal constructions - see Mac Lane 1986) are used throughout mathematics.

3 ETCS axioms and toposes

The topos axioms are the category axioms (see Category theory, introduction to §1) plus these two:

(1) There is a terminal object 1, and any objects *A* and *B* have a product $A \times B,\ p_1,\ p_2$.
(2) Each object *B* has a power set $PB,\ \in^B$.

Products and power objects are defined as in §2, reading 'object/arrow' for 'set/function'. We do the same for natural number objects, and many toposes satisfy the further 'axiom of infinity':

(3) There is a natural number object *N*, 0, *s*.

The characteristic feature of Set is that each function is uniquely determined by its effect on elements:

(4) Given any *f* and *g* both $A \to B$, if for all $x : 1 \to A$ we have $fx = gx$, then $f = g$.

The ETCS axioms are the category axioms plus (1)-(4).

In general topos theory any arrow $f : B \to A$ may be called a 'generalized element' of *A*. We call an arrow $i : S \to A$ 'monic' if it is one-to-one for all generalized elements of *S*: for any object *T* and arrows $h : T \to S$ and $k : T \to S$, $ih = ik$ implies $h = k$. We call any monic $i : S \to A$ a 'sub-object' of *A*. The definition of subsets in §2 agrees with this in Set but not in general.

We say *i* is 'included in' another sub-object $j : P \to A$ if there is some $f : S \to P$ with $i = jf$. Trivially, $1_A : A \to A$ is monic and includes all sub-objects of *A*. Call it all of *A*. Every topos has an initial object ϕ, and for each *A* the unique $\phi \to A$ is monic and included in every sub-object of *A*. Call it empty. Any two sub-objects of *A* have an intersection, defined as the largest sub-object included in both. They have a union defined as the smallest including both. Each sub-object *i* has a negation (or pseudocomplement), defined as the largest sub-object whose intersection with *i* is empty.

The operations on sub-objects give each topos a natural logic. For details see Johnstone (1977: ch.5) or other introductions to topos theory. The law of noncontradiction holds in every topos since by definition the intersection of any sub-object with its negation is empty. In terms of logic, every formula $P \mathbin{\&} {\sim}P$ is false. But in many toposes the union of a sub-object of *A* with its negation need not be all of *A*, and so a formula $P \vee {\sim}P$ need not be true. The law of the excluded middle (LEM) fails in some toposes.

For example, in the topos Spaces every object has a spatial structure and every arrow has a derivative. There is an object *R* which we think of as the line. The arrows $R \to R$ are the classical differentiable functions on the real numbers. The line has a point 0, that is, an element $0 : 1 \to R$. So there is a sub-object $\{0\} {\to} R$ and its negation $\{x \neq 0\} {\to} R$. But the union $\{x | x = 0 \vee x \neq 0\}$ is not all of *R*. It is 'all of *R* except the cohesion at 0'. An arrow from this union to *R* may lack a derivative at 0 and even be discontinuous at 0 simply because it can be defined separately on the pieces $\{0\}$ and $\{x \neq 0\}$. So LEM fails in Spaces for the formula $x = 0$, among many others.

Heyting's first-order intuitionist logic is sound in Spaces as it is in every topos (see Intuitionism §7). It is not complete since in Spaces LEM holds for closed formulas. Every disjunction $P \vee {\sim}P$ with no free variables is true in Spaces. In Set the union of any subset and its negation is the whole set so LEM holds. Of course classical first-order logic is sound and complete in Set. Heyting's logic is complete in some toposes, and so for logic in toposes collectively.

The axiom of choice is false in many toposes. It is easy to see it must fail in Spaces. Consider the function $f : R \to R$ defined by $fx = x^3 - x$. For each *y* in *R* there is an *x* with $fx = y$ so if the axiom of choice held there would be a choice function $g : R \to R$ such that for each *y*, $f(gy) = y$. But the graph of *f* shows no such function *g* can be continuous everywhere (let alone differentiable), so no such function can exist in Spaces. In fact the axiom of choice fails in any topos where LEM fails, since it implies LEM (see Johnstone 1977 or other topos texts).

In another respect topos logic is not even formally intuitionistic. In the logic of any topos, for every object *A*, every formula with just one free variable of sort *A* defines a sub-object of *A*. Brouwer's philosophical misgivings about set formation find no reflection here.

When Grothendieck created toposes for applications in topology, algebraic geometry and number theory (see Category theory, introduction to §3) he paid no attention to intuitionistic mathematics, let alone to Brouwer's epistemology. And the theory today does not say reasoning must all be intuitionistic or constructive. Reasoning should be just as classical or constructive as will suit its context.

4 Radically categorical foundations

Lawvere offered ETCS as a logical foundation and to support an ontological claim. Since ETCS sets have no

determinate contents, ‘we seem to have partially demonstrated that even in foundations, not Substance but invariant Form is the carrier of the relevant mathematical information’. But ETCS is not his only logical foundation. For other purposes ‘a more satisfactory foundation will involve a theory of the category of categories’ (1964: 1,506, 1,510).

Lawvere (1966) axiomatized a ‘category of categories’. We call these axioms CCAF. They give general theorems of category theory without implying the existence of many specific categories. Lawvere gave various stronger axioms. Positing that Set exists is one. The axioms then give all the categories constructed from Set.

Another would be to posit a ‘category of all categories’ as one object in the CCAF universe which represents the whole universe. The set theory NF (see Set theory, different systems of §7) includes a set of all sets. It also has a category of all categories in the most naïve sense. But this set theory works so awkwardly with functions that the category of all categories lacks crucial properties (it is not Cartesian closed). Axiomatizing this does not seem promising. A more promising approach not yet much explored would use ‘fibered’ categories in the sense of Bénabou (1985).

The radically categorical position, though, does not lean on any one logical foundation. It says mathematical objects exist each in its own right. Given any type of structure one goal should be to describe it as intrinsically as possible, and the most intrinsic descriptions of structures to date are categorical. For example, axioms for the topos Spaces describe differential manifolds including the line *R* in their own terms. These axioms do not presume set theory or anything else. They are on a par with ETCS or ZF in giving a self-contained logical foundation for their subject. But axioms only organize our knowledge. Spaces, sets and more exist whether we axiomatize them or not.

If spaces are as real as sets in this sense, then the logic of Spaces is as genuine a logic as that of Set. This is the point of controversy over topos logic. The axioms for Spaces can be modelled in classical set theory. (Moerdijk and Reyes (1991) give many variants.) And Set can be modelled in Spaces (McLarty 1988). You cannot have one without the other. But you could claim the logic of sets is genuine logic while Spaces and other toposes give merely formal interpretations. The radically categorical position says otherwise: each topos (and other categories we do not go into here) has its own genuine logic.

Logical foundations can also describe relations between various branches of mathematics. It is useful in geometry to know how spaces relate to sets. Plus, modelling Spaces in Set and vice versa gives relative consistency results - the Spaces axioms are consistent if ETCS is and vice versa. Johnstone (1977: ch.9) gives relative consistency results for variants of ZF and ETCS. This can increase our confidence in all the axiom systems.

What no logical foundations can do, on this view, is give a starting point to justify the rest of mathematics. No such powerful axiom system is transparently consistent and there is no prospect of proving any one consistent (see Gödel’s theorems). The justification of mathematics rests on its whole place in our knowledge. All consistency proofs are relative consistency proofs.

See also: Logical and mathematical terms, glossary of

COLIN McLARTY

References and further reading

Bénabou, J. (1985) ‘Fibered Categories and the Foundations of Naïve Category Theory’, *Journal of Symbolic Logic* 50: 10-37.(Requires basic category theory. Cited in §4.)

Benacerraf, P. (1965) ‘What Numbers Could Not Be’, *Philosophical Review* 74: 47-73.(Discusses the classical problem for set-theoretic foundations that arithmetic has no unique set-theoretic representation. Requires basic logic.)

Corry, L. (1992) ‘Nicolas Bourbaki and the Concept of Mathematical Structure’, *Synthese* 92: 315-48. (Includes Bourbaki’s internal debates over category theory.)

Johnstone, P. (1977) *Topos Theory*, London: Academic Press.(The standard reference. Chapters 1-4 are quite demanding. Others are readable with basic topos theory. Cited in §§3-4.)

Lambek, J. and Couture, J. (1991) ‘Philosophical Reflections on the Foundations of Mathematics’, *Erkenntnis* 34: 187-209.(Claims moderate logicism, formalism, Platonism and intuitionism are all realized in the ‘free topos’. Requires basic formal logic.)

Lawvere, F.W. (1964) ‘An Elementary Theory of the Category of Sets’, *Proceedings of the National Academy of*

Science 52: 1,506-11.(Requires basic category theory.)

Lawvere, F.W. (1966) ‘The Category of Categories as a Foundation for Mathematics’, in S. Eilenberg (ed.) *Proceedings of a Conference on Categorical Algebra in La Jolla*, Berlin: Springer, 1-21.(Cited in §4. Requires basic category theory.)

Lawvere, F.W. (1969) ‘Adjointness in Foundations’, *Dialectica* 23: 281-96.(Includes a rapid introduction to the terms.)

Lawvere, F.W. (1994) ‘Cohesive Toposes and Cantor’s ‘‘*lauter Einsen*’’’, *Philosophia Mathematica* 2 (3): 5-15. (Describes aspects of Cantor’s thought, such as his cardinals, lost in ZF set theory. Cited in §2.)

Mac Lane, S. (1986) *Mathematics Form and Function*, New York: Springer, 386-402.(Easy account of the categorical terms in §2.)

Marquis, J.-P. (1995) ‘Category Theory and the Foundations of Mathematics: Philosophical Excavations’, *Synthese* 103: 421-47.(Refers to much mathematics but readable without relevant background.)

Mayberry, J. (1988) ‘What are Numbers?’, *Philosophical Studies* 54: 317-54.(Argues that working methods in mathematics since Newton have disguised the role of concrete foundations, as in §2.)

McLarty, C. (1988) ‘Defining Sets as Sets of Points of Spaces’, *Journal of Philosophical Logic* 17: 75-90. (Compares the nineteenth-century creation of set theory from analysis to the relation between Set and Spaces. Requires basic logic. Cited in §4.)

McLarty, C. (1994) ‘Numbers Can Be Just What They Have To’, *Noûs* 27: 487-98.(Shows the structuralism of ETCS as in §2. No technical knowledge is required.)

Moerdijk, I. and Reyes, G. (1991) *Models for Smooth Infinitesimal Analysis*, New York: Springer. (Many variants on the Spaces of §3-4. Requires some differential geometry plus basic topos theory or skill in algebra.)

Category theory, introduction to

A 'category', in the mathematical sense, is a universe of structures and transformations. Category theory treats such a universe simply in terms of the network of transformations. For example, categorical set theory deals with the universe of sets and functions without saying what is in any set, or what any function 'does to' anything in its domain; it only talks about the patterns of functions that occur between sets.

This stress on patterns of functions originally served to clarify certain working techniques in topology. Grothendieck extended those techniques to number theory, in part by defining a kind of category which could itself represent a space. He called such a category a 'topos'. It turned out that a topos could also be seen as a category rich enough to do all the usual constructions of set-theoretic mathematics, but that may get very different results from standard set theory.

1 Examples and the category axioms

Consider sets and functions. Write $f : A \to B$ to say *f* is a function defined on the set *A* and all its values are in the set *B*. Each set *A* has an identity function $1_A : A \to A$ with $1_A(x) = x$ for every *x* in *A*. Given any $f : A \to B$ and $g : B \to C$ there is a composite $gf : A \to C$ with $gf(x) = g(f(x))$ for all *x* in *A*.

Now consider a formal language with deductive rules. Let $f : A \to B$ say *f* is a proof whose sole assumption is the formula *A* and conclusion is the formula *B*. Each formula *A* gives a one-line proof $1_A : A \to A$ with *A* as assumption and conclusion. Proofs $f : A \to B$ and $g : B \to C$ concatenate to give $gf : A \to C$. We view a proof as transforming its assumption into its conclusion.

The features which are common to these two examples are widespread in mathematics. Category theory gives an axiomatic treatment of them. The category axioms speak abstractly of 'objects' and 'arrows': each arrow has an object as domain and another as codomain. We write $f : A \to B$ to say *f* is an arrow with domain *A* and codomain *B*. If the codomain of one arrow is the domain of another, as in the case of $f : A \to B$ and $g : B \to C$, then the arrows have a 'composite' $gf : A \to C$. Each object *A* has an 'identity arrow' $1_A : A \to A$. The axioms require that for any arrows $f : A \to B$, $g : B \to C$ and $h : C \to D$ we have $h(gf) = (hg)f$, $f1_A = f$ and $1_B f = f$.

Thus we have the 'category' Set, whose objects are sets and whose arrows are functions, and the category Proof, whose objects are formulas and whose arrows are proofs (see Lambek 1994). The axioms clearly hold for Set and for Proof. Other useful categories include: Top, with topological spaces for objects and continuous functions for arrows; and Group, with groups (as in abstract algebra) for objects and group homomorphisms for arrows.

Picture a category as a network of arrows between points. So Set is an infinite network with a point for each set and an arrow for each function. Each path of arrows connected head-to-tail has a composite going from the first point on the path to the last.

2 Working with categories

We can look at some simple categorical structures. An object *T* of a category A is a 'terminal object' in A if each object of A has exactly one arrow to *T*. An object *W* of A is a 'weak terminal object' in A if each object of A has at least one arrow to *W*.

Terminal objects of Set are singleton sets. Weak terminal objects of Set are non-empty sets. In Proof the weak terminal objects are the provable formulas. If *W* is provable with no assumptions then it is provable from any assumption. Nothing is provable in only one way, so Proof has no terminal object.

Reversing arrows in the definition of a terminal object defines an 'initial object': an object *I* of A such that each object of A has exactly one arrow from *I*. Similarly for a 'weak initial object'. A set is initial in Set if and only if it is empty, and is weakly initial on the same condition. No formula is initial in Proof. Inconsistent formulas are weakly initial.

These definitions deal only with patterns of arrows. Yet in Set and Proof they coincide with specific concepts such as non-emptiness or provability. Other definitions give richer structures such as Cartesian products or power sets

in Set, the several connectives in Proof, and topological or group structures in Top or Group. So category theory has a unifying role where one argument proves different theorems in different contexts. More importantly, it focuses on the relevant structural relations between objects, be they sets or spaces or whatever. To see this at work we ascend to transformations of categories, called ‘functors’.

We write $\mathbf{F} : \mathbf{A} \to \mathbf{B}$ to say F is a functor from the category A to the category B. This means that for each object *A* of A there is an object $\mathbf{F}A$ of B, and for each arrow $f : A \to A'$ of A there is an arrow $\mathbf{F}f : \mathbf{F}A \to \mathbf{F}A'$ of B, and F preserves identity arrows and composites. That is, for each *A* of A we have $\mathbf{F}(1_A) = 1_{\mathbf{F}A}$, and for any composite gf we have $\mathbf{F}(gf) = (\mathbf{F}g)(\mathbf{F}f)$. So F draws a picture of the network A in the network B.

Each category A has an identity functor $1_\mathbf{A} : \mathbf{A} \to \mathbf{A}$ with $1_\mathbf{A}(A) = A$ and $1_\mathbf{A}(f) = f$ for all objects *A* and arrows *f* of A. Functors compose in the obvious way. So categories and functors themselves are the objects and arrows of a category of categories (see Category theory, applications to the foundations of mathematics §4).

Functors were invented to focus on relevant structure. Since about 1900 homology theory has been central to topology. To simplify somewhat, by 1940 homology theory was a way of taking each topological space *S* and constructing from it a family of groups $H_n(S)$, one for each natural number *n*. By a further construction each continuous function $f : S \to S'$ between topological spaces gives homomorphisms $H_n(f) : H_n(S) \to H_n(S')$. So group theory could be used to prove theorems of topology. But there were many different ways to construct the groups. The subject was thriving and chaotic.

Eilenberg and Mac Lane looked away from constructing the groups $H_n(S)$ and saw homology instead as drawing the network of continuous functions in the network of homomorphisms. They saw each H_n as a functor $H_n : \mathbf{Top} \to \mathbf{Group}$. To do this they created category theory and wrote their classic paper, ‘The General Theory of Natural Equivalences’ (1945). Category theory is now the standard language for homology. It cleared away old complexities and made room for vast new ones.

The new homology (and related cohomology) spread to algebra and number theory taking category theory with it. Then category theory took other new directions. (Applications before 1970 are surveyed in passing in Mac Lane (1971).)

3 Toposes

Around 1960 Grothendieck attacked some famous problems in geometry and number theory. His project achieved its initial aims in 1973 with Deligne’s proof of the Weil conjectures, certain theorems concerning integer roots of polynomials (see McLarty 1990). For this project Grothendieck created ‘toposes’ (or, ‘topoi’). These were categories which he viewed as spaces. Every ordinary topological space can be represented by a topos but toposes are considerably more general than that. The topological method of cohomology can be applied to all toposes and thus Grothendieck defined the ‘étale’ cohomology that Deligne used. This cohomology is now central to number theory but not all its applications require toposes.

About 1970 Lawvere and Tierney gave first-order axioms which apply to all of Grothendieck’s toposes plus some other categories. Today we use ‘topos’ or ‘elementary topos’ to mean any model of the axioms, and ‘Grothendieck topos’ to mean one got by Grothendieck’s method. To a logician a topos is analogous to the universe of sets. In any topos the constructions of ordinary mathematics exist, but not always in classical form. Classical logic is not valid in all toposes.

The category Set is a topos. Classical logic is valid in it and it gives classical mathematics. But there is also a topos Eff in which all functions from the natural numbers to themselves are recursive. Classical logic would prove there are non-recursive functions; but classical logic is not valid in Eff. There is a topos Spaces in which each object has a spatial structure and each function is differentiable. This topos includes all the usual differentiable functions of calculus. Classical logic proves there are non-differentiable functions but it is not valid in Spaces. In any topos the valid logic is somewhere between classical and intuitionistic. This is uncontroversial as a fact in model theory. For controversy over its philosophical interpretation see Category theory, applications to the foundations of mathematics §4.

See also: Logical and mathematical terms, glossary of

COLIN McLARTY

References and further reading

Bell, J.L. (1988) *Toposes and Local Set Theories*, Oxford: Oxford University Press.(A logician's account of toposes using their analogy with the classical universe of sets. Requires facility with sets and formal logic. Includes a bibliography.)

Eilenberg, S. and Mac Lane, S. (1945) 'The General Theory of Natural Equivalences', *Transactions of the American Mathematical Society* 58: 231-94.(The original paper. Requires basic knowledge of vector spaces and topology.)

Lambek, J. (1994) 'What is a Deductive System?', in D.M. Gabbay (ed.) *What is a Logical System?*, Oxford: Oxford University Press, 145-64.(Surveys styles of proof theory including the categorical one in §1.)

Mac Lane, S. (1971) *Categories for the Working Mathematician*, Berlin: Springer.(The standard. The main text requires some sophistication; the examples call for much background. Includes a thorough bibliography.)

Mac Lane, S. (1986) *Mathematics Form and Function*, New York: Springer, 386-402.(Basic category theory, including Set, is described with no substantial prerequisites.)

Mac Lane, S. and Moerdijk, I. (1992) *Sheaves in Geometry and Logic*, Berlin: Springer.(The general theory of Grothendieck toposes, with many examples, including Cohen's forcing as a topos construction. Requires some category theory.)

McLarty, C. (1990) 'The Uses and Abuses of the History of Topos Theory', *British Journal for the Philosophy of Science* 41: 351-75.(Debates some views of the history of toposes.)

McLarty, C. (1992) *Elementary Categories, Elementary Toposes*, Oxford: Oxford University Press.(A categorical account, with chapters on Set, Spaces and Eff. Requires facility with algebraic thinking and formal logic. Includes a bibliography.)

Cattaneo, Carlo (1801-69)

The figurehead of the Italian democratic movement prior to the unification of Italy, Carlo Cattaneo developed a theory of federalism as a practice of self-government, envisaging a United States of Italy. He identified the bourgeoisie as the most dynamic force in contemporary history and regarded scientific culture as the engine of progress. Often dubbed the first Italian positivist, he perceived empirical philosophy as a kind of synthesis of all the sciences, but also stressed its anthropological and psychological dimensions and above all its character as a methodology of knowledge; his objective was to study human thought. The great themes of Cattaneo's philosophy are nature, the individual and society; particularly the last.

Cattaneo represents the most progressive trends in Italian thought in the first half of the nineteenth century. He aimed to encourage the propagation of technical and scientific knowledge in his native Lombardy in order to equip the entrepreneurial class with the expertise necessary for economic development; and to this end he founded the journal, *Il Politecnico* (1839-44; 1851-63). Until 1848, Cattaneo had formed part of the legal opposition to the regime, but when Milan rose against the Austrians on 18 March of that year, Cattaneo took an active, leading role in the Five Days' Uprising. He was critical of the provisional government, however, and when the Austrians returned, left for Paris, where he published *L'Insurrection de Milan en 1848 (The Milan Uprising of 1848)*. Shortly afterwards, he left Paris for Switzerland, where he settled permanently in Castagnola, near Lugano. From his exile in Switzerland, Cattaneo voiced his often fierce opposition to the principal political currents in Italy, and forged a role for himself as the great 'prompter' and moving spirit of the Italian democrats, and the leading proponent of federalism against unitary solutions.

Partly as a result of the influence of the philosopher G.D.R. Romagnosi, from whom he inherited the notion of sociability as the foundation of civil society, Cattaneo insisted on the need to intervene practically to shape the reality of one's society. Associated with this basic principle were his profound mistrust of the notion of the solitary consciousness and stress on human action in all its aspects. Rejecting the 'four opposing exaggerations' of idealism, materialism, scepticism and mysticism,Cattaneo saw the human environment as containing a balance of material and ideal forces. He hoped that the kind of renewal that had taken place in the natural sciences could also be brought about in the human sciences.

Cattaneo regarded history as the history of peoples and civilizations. It might even be studied through the examination of languages: a method which Cattaneo employed in a new and often anti-Romantic manner. The perspective adopted in his great historical canvasses is that of liberal, bourgeois culture, in which social changes and transformations are possible and necessary, and social mobility contains subversive forces. At least until 1848, Cattaneo regarded England as the apogee of modern economic development and triumphant capitalism (*Della conquista d'Inghilterra pei Normanni (On the Norman Conquest)* (1839); *Alcuni Stati moderni (Some Modern States)* (1842)), even though he did not ignore the effects of colonialism and its abuses of power (*Dell'India antica e moderna (India, Past and Present)* (1846)). Progress, for Cattaneo, is made up of cultural graftings, of conflicts and exchanges of ideas and principles between different peoples and cultures. Social conditions are the source of good or evil. Not for nothing have some critics seen Cattaneo's famous *Interdizioni israelitiche (Israeli Interdicts)* of 1835 as containing the germs of historical materialism, in view of the importance he attributes there to economics as a factor in cultural development; while in his *Notizie naturali e civili sulla Lombardia (Notes on the Natural and Social History of Lombardy)* (1844) - perhaps his masterpiece - the portrait of a region becomes a model of historical method in its presentation of the dynamic interaction of the forces of economic interest and human intelligence.

Cattaneo's idealization of the Lombard bourgeoisie is just one aspect of his celebration of the bourgeoisie in general as the driving force of capitalism, within the context of which science and technology constitute the instruments of development and progress. Addressing himself to the problem of the origins of the Italian bourgeoisie, Cattaneo located them in the rise of the city, which interested him as the historically decisive force in the struggle against feudalism and as the catalyst for the whole process of socioeconomic development in Italy (*Della formazione e del progresso del Terzo Stato (On the Genesis and Development of the Third Estate)* (1854); *La città come principio ideale della istoria italiana (The City as the Ideal Origin of Italian History)* (1854). For

Cattaneo, urban life was the typical and distinctive embodiment of Italian culture; and the city was also the birthplace of scientific culture and of modern empirical thought. It was no coincidence, for him, that the Italian revolutionary movement of 1848 had its roots in the cities.

It has been remarked of Cattaneo that his thought oscillates between a conception of philosophy as a synthesis of the empirical sciences and, contrastingly, as an analysis of the human spirit. Philosophy needed to draw on science, but at the same time it needed to be a psychological and anthropological exploration of collective life. In his *Collective Psychology* (five lectures delivered at the Lombard Institute of Sciences between 1859 and 1865), he takes as his object of study that most social of human acts: thought. And social ideology, the subject of his Lugano lectures, implies a study of human society from a perspective which foreshadows that of modern anthropology. Cattaneo employs the notion of a ‘system’. A system may be closed, which is to say reactionary, or open and progressive; and it is the succession of one system by another that constitutes progress. In this way, Cattaneo develops the thought of Vico (§6) and integrates it with the notion of progress, the great common theme of eighteenth- and nineteenth-century writers.

During his exile, as he watched Italian society embark on a process of development very different from that for which he had hoped,Cattaneo abandoned political activism for the sphere of ‘opinion’ and ideas. He attached great importance to opinion-forming, which he regarded as essential as a means of combating the ignorance of the masses and preparing them for struggle on a national scale.Cattaneo’s thought can thus be distinguished from socialist positions in that he envisaged the problems of the rural and industrial proletariat being dealt with by the bourgeoisie itself.

When attempts have been made to trace the ideological antecedents of twentieth-century Italian culture, Cattaneo has often been identified as the ‘founding father’ of the anti-moderate tradition in recent Italian history, the great protagonist of Italian ‘anti-history’ and the most authentic example the nineteenth century provides of a dissenting intellectual.

See also: Italy, philosophy in §2

Translated from the Italian by Virginia Cox
DELIA FRIGESSI

List of works

Cattaneo’s writings were first published in his journal *Il Politecnico*.

Cattaneo, C. (1845) *Alcuni scritti (A Selection of Writings)*, Milan: Borroni & Scotti. (Contains *Della conquista d’Inghilterra pei Normanni (On the Norman Conquest)* (1839), *Alcuni Stati moderni (Some Modern States)* (1842) and *Dell’India antica e moderna (India, Past and Present)* (1846).)

Cattaneo, C. (1848) *L’Insurrection de Milan en 1848 (The Milan Uprising of 1848)*, Paris: Amyot.(Cattaneo’s personal account of the Milan Uprising, highly critical of the provisional government.)

Cattaneo, C. (1860) *Memorie di Economica pubblica dal 1833 al 1860 (Records of Public Economy from 1833 to 1860)*, Milan: Libreria F. Sanvito.(Contains *Ricerche economiche sulle interdizioni imposte dalla legge civile agli Israeliti (Economic research on the bans imposed by civil law on the Israelites)*.)

Cattaneo, C. (1948) *Scritti letterari, artistici, linguistici e vari (Writings on Literary, Artistic, Linguistic and Other Subjects)*, ed. A. Bertani, Florence: Le Monnier, 2 vols.(Essays and articles by Cattaneo on aesthetics, linguistics and other subjects.)

Cattaneo, C. (1949-56) *Epistolario (Collected Letters)*, ed. R. Caddeo, Florence: Barbera, 4 vols.(Cattaneo’s correspondence on a range of subjects.)

Cattaneo, C. (1950) *Scritti filosofici (Philosophical Writings)*, ed. N. Bobbio, Florence: Le Monnier, 3 vols.(This volume contains his lectures *Collective Psychology* (1859-65).)

Cattaneo, C. (1955) *Scritti economici (Writings on Economics)*, ed. A. Bertolino, Florence: Le Monnier, 3 vols. (Cattaneo’s writings on industry, capitalism and economic issues.)

(1957) *Scritti storici e geografici (Historical and Geographical Writings)*, ed. G. Salvemini and E. Sestan, Florence: Le Monnier, 4 vols.(The second volume contains *Notizie naturali e civili sulla Lombardia (Notes on the Natural and Social History of Lombardy)* (1844) and *Della formazione e del progresso del Terzo Stato (On the Genesis and Development of the Third Estate)* (1854).)

Cattaneo, C. (1964-5) *Scritti politici (Political Writings)*, ed. M. Boneschi, Florence: Le Monnier.(Cattaneo’s political writings.)

Cattaneo, C. (1968) *Scritti scientifici e tecnici, 1823-1848 (Scientific and Technical Writings, 1823-1848)*, ed. C.S. Lacaita, Florence: Giunti and Barbera.(Cattaneo’s technical and scientific writings.)

Cattaneo, C. (1972) *Opere scelte (Selected Works)*, ed. D. Castelnuovo Frigessi, Turin: Einaudi, 4 vols.(A selection of Cattaneo’s major writings in a range of subjects.)

References and further reading

Ambrosoli, L. (ed.) (1989) *‘Il Politecnico’, 1839-1844*, Turin: Bolleti Boringhieri, 2 vols.(A very full selection of articles from the famous Lombard journal, edited by Cattaneo, including extracts documenting its role in the diffusion of cultural and scientific knowledge.)

Ambrosoli, L. *et al.* (1975) *L’opera e l’eredità di Carlo Cattaneo. vol. 1, L’opera; vol. 2, L’eredità (The Works of Carlo Cattaneo, their Fortunes and Influence: vol. 1, Cattaneo’s Works; vol. 2, Fortunes and Influence)*, Bologna: Il Mulino.(A collection of essays that gives a good overall picture of the different tendencies in Cattaneo studies and indicates the reasons for the continuing interest in his works.)

Bobbio, N. (1971) *Una filosofia militante. Studi su Carlo Cattaneo (A Militant Philosophy. Studies on Carlo Cattaneo)*, Turin: Einaudi.(Cattaneo is presented as a paradigm of rationalist thought in these studies, which cover his federalism, his philosophy and philosophical teaching and the critical fortunes of his works.)

Castelnuovo Frigessi, D. ‘Lo specchio di Cattaneo’ (Cattaneo’s Mirror), *Studi storici* 11 (2): 254-78. (Interpretations of Cattaneo and their connection with the development of his thought.)

Cattaneo, C. (1943) *Lombardia antica e moderna (Lombardy, Ancient and Modern)*, ed. M. Fubini under the pseudonym M. Fusi, Florence: Sansoni.(An edition of Cattaneo’s famous *Notizie naturali e civili sulla Lombardia (Notes on the Natural and Social History of Lombardy)* brought out during the Fascist period by a distinguished Italian literary scholar, in which Cattaneo is presented first and foremost as a great historian.)

Puccio, U. (1977) *Introduzione a Cattaneo (Introduction to Cattaneo)*, Turin: Einaudi. (Marxist study, stressing the unity ofCattaneo’s work as a historian and ideologue of capitalism and underlining the ideological element in his thought.)

Causality and necessity in Islamic thought

Discussions of causality and necessity in Islamic thought were the result of attempts to incorporate the wisdom of the Greeks into the legacy of the QurČamp;lsquo;'an, and specifically to find a philosophical way of expressing faith in the free creation of the universe by one God. Moreover, that article of faith was itself a result of the revelation of God's ways in the free bestowal of the QurČamp;rsquo;'an on a humanity otherwise locked in ignorance, which a purportedly Aristotelian account of the necessary connection of cause and effect might be taken to rule out. Thus free creation of the universe and free gift of the Qur'an formed a logical unit. The challenge, therefore, was to compose an account of metaphysical and ethical matters which permits rational discourse about them, without obscuring their ultimate source or precluding divine action in the course of world events and human actions.

*The scheme of emanation elaborated by al-Farabi sought to give 'the First' the place of pre-eminence which the Qur'an demanded for the Creator, but did so by modelling creation on a logical system whereby all things emanated necessarily from this One. It was this necessity, further articulated by Ibn Sina, which al-Ghazalii took to jeopardize the freedom of God as Creator and as giver of the Qur'an. al-Ghazali's objections were honed by a previous debate among Muslim theologians (*mutakallimun*), who had elaborated diverse views on human freedom in an effort to reconcile the obvious demand for free acceptance of the Qur'an with its claims regarding God's utter sovereignty as Creator over all that is. Natural philosophy was also affected by these debates, specifically with regard to the ultimate constitution of bodies as well as accounts that could be given of their interaction. However, the primary focus was on human actions in the face of a free Creator.*

1 Metaphysical issues

The understanding of causality that prevailed among the classical Islamic philosophers was decidedly Neoplatonic in character (see Neoplatonism §3). Intellectual coherence was assured by a scheme of emanation itself modelled on the necessity inherent in a logical system. Thus the connections between events shared in the connections between propositions that followed logically from one another. In this fashion, al-Farabi's emanation scheme offered a cosmic pattern for all causality as well as a master metaphor for causal interaction (see al-Farabi §2). All that is and all that happens was conceived as depending, for its being and its intelligibility, ultimately on the first cause (or 'the One') by way of intermediary influences which are cosmic in character yet linked together by an intellectual inherence assuring necessary linkage between cause and effect. Aristotle's ideal of 'scientific explanation' in the *Posterior Analytics* could only have been projected, after all, if the world itself was so constructed that things and events were properly connected with one another so as to form a coherent whole (see Aristotle §3).

That is the sense to be made of Ibn Sina's division of 'being' into 'necessary being' and 'possible being', with 'necessary being' restricted to the One from which all the rest emanates while the remainder is characterized as 'possible in itself yet necessary by virtue of another' (see Ibn Sina §5). In this way the order of the natural world is assured, since it derives from the one principle of being in a way that is modelled on logical derivation. In this way also, the necessity of causal interaction becomes virtually identical with that of logical entailment, thereby linking the entire universe in a necessary order with the first cause. Furthermore, the pattern of logical entailment extends to the action of that first cause as well: the universe comes forth from it necessarily, as premises from a principle.

Such a model for causal activity cannot be easily imported into a world believed to be freely created by one God. The order described by the emanation scheme threatened the hegemony of the God revealed in the Qur'an, by removing the freedom of that God to reveal as well as to create. Accommodation with Neoplatonic thought required too much by way of concession from believers in the Qur'an, and it was only a matter of time before this effort to harmonize creation with emanation was challenged. That challenge came notably from al-Ghazali, whose frontal attack, entitled *Tahafut al-falasifa (The Incoherence of the Philosophers)* was in turn countered by Ibn Rushd in his *Tahafut al-tahafut (Incoherence of the Incoherence)*. But while Ibn Rushd's defence of a repristinated Aristotle would continue to influence Western thought, al-Ghazali's spirited attack succeeded in virtually marginalizing such philosophical reflection on metaphysical matters in the Islamic world to the activities of an elite who would come to be known as 'the philosophers', and whose adherence to Qur'anic faith would often be

suspect.

2 Human action and divine action

If such quasi-logical necessity attributed to causality in the universe ran counter to the freedom of the divine agent, it also threatened by implication the freedom of human beings to respond to divine revelation; and such freedom is clearly presupposed by the very structure of the Qur'an, which calls constantly for a response to its warnings and means to elicit wholehearted response to the guidance it offers (see Islamic theology §3). Yet the controversy here was not carried out on the terrain staked out by the philosophers but rather among expressly religious thinkers (called *mutakallimun* because of their desire to articulate the faith by way of argument). The earliest among these, the Mu‘tazilites, could only see their way to securing human freedom by considering free human actions as utterly autonomous, and so as the creations of human agents, not of God. These early Islamic thinkers fashioned their position on such matters without benefit of the later philosophical reflections noted above, and appear to have conflated notions of origination, causation and creation in an effort to assure that humans bear complete responsibility for the actions for which they will be rewarded or punished. Their concern was for justice, as it can be applied to human beings and to the God of the Qur'an.

Again, however, to withdraw a sector of creation from the purview of the creator of heaven and earth, and to insist that human actions be our creation and not God's, could hardly be sustained in Islam. The challenge this time came from one of the theologians' own number, al-Ash‘ari, who shared their conceptual conflations; he developed a purportedly intermediate position, whereby human actions are *created* by God yet *performed* by us. The key notion introduced was *kasb* (or *iktisab*), which attempts to distinguish responsibility for one's actions from their sheer origination. Originating in the marketplace to describe transactions, it alludes to the fact that actions which God creates are ‘acquired', or perhaps better, ‘performed' by human beings as created agents (see Ash‘ariyya and Mu‘tazila §§1, 5).

While this position obviates the removal of human actions from the domain of God's creation, it seems to complicate unduly the issue of human agency and responsibility. The conceptual question turns on the extent to which the notion of ‘agent' can be used of both creator and creature. Can there be an authentic agent other than the creator?

This is a question, of course, which touches every religious tradition which avers the free creation of the universe by one God: how does the first cause relate to other causes? Indeed, can there even be other causes in the face of a sovereign ‘First'? The response of *kalam* thinkers to this question was complicated by their commitment to an atomistic metaphysics, which seemed designed to remove all causality other than the divine from the realm of nature: this is the celebrated ‘Islamic occasionalism'. Later, however, the question was debated by al-Ghazali without reference to any such metaphysical theses. Rather, the terms were those introduced by ‘the philosophers', turning on the necessity of the connections between events in nature, specifically between those which we recognize as causes and their effects. Clearly these connections could not be akin to logical necessity, or there would be no room whatsoever for miracles like the ‘descent' of the Qur'an; yet if the universe is the result of God's free action and not of necessity, the creator will continue to be free to act within creation. Thus the account given of causal agency in general, and of personal agency in particular, will have to allow for just that: a kind of agency proper to creatures, yet always subordinate to the influence of a free creator. al-Ghazali responds to this challenge by comprehending created causes under a patterned regularity of the *sunna Allah*, action willed by God: creatures do indeed contain such powers, yet always subject to the will of the One who so created them. In this way, a key Islamic religious thinker such as al-Ghazali can simultaneously insist that God alone is the only agent and yet, by God's power, others are agents as well. Thus causality can be attributed to creatures, but not causal connections of the quasi-logical sort demanded by the emanation scheme.

3 Natural philosophy

Early *kalam* thinkers, as noted above, presumed an atomistic conception of nature by which the universe was divinely sustained by being freely created at ‘each moment' by God. This conception is clearly an attempt to affirm the omnipresence of divine causal action in the universe God creates, and just as clearly evacuates created causal efficacy. It also runs foul of Aristotle's refutations of atomism in terms of the irreducibility of the continuum of time, space and matter to discrete moments (see Atomism, ancient). It is these arguments which

persuaded al-Ghazali that the Ash‘arite presumption of the atomistic constitution of nature was gratuitous and unnecessary as a ploy to assure the omnipresence of the creator’s action in nature. Thus Islamic thought has not been caught between the two extremes of ‘occasionalism’ (whereby all action is effectively God’s) (see Occasionalism) and the pervasive necessity associated with the Neoplatonic emanation scheme of its notable philosophers; there is a middle course, intimated by al-Ghazali and explicitly developed in the early twentieth century in Egypt in the celebrated Qur’an commentary *al-Manar*, whereby the *sunna Allah* is evoked to explain the consistency of a created world of nature.

In summary, Islamic thought must always reconcile the assertion that the entire universe is the free creation of God, who neither requires anything in order to create it nor stands to gain anything by creating it, with the fact that this created world has a consistency associated with causes and effects as we observe them and as we use them to explain natural phenomena. If that causal consistency is articulated in terms which presume a necessity inimical to the free action of a creator, either in the beginning or at any moment of the universe’s duration, then Islamic thinkers will feel constrained, as some indeed have, to deny created causality in favour of divine sovereignty. However, as might be expected, developments internal to Islamic thought on this matter have found a way of affirming the free creation of the universe together with its causal consistency, and in doing so have suggested a pattern for causal connections in nature which distances them from the quasi-logical while respecting their reality, specifically by invoking an analogy with the patterns which the God who reveals the Qur’an sets up between human actions and their recompense: the *sunna Allah*. In this way the world of nature can be seen to have a consistency proper to it, yet at the same time is affirmed to be the result of an intentional agent whose order it reflects in the operations proper to it (see Intention §4). The pattern to be found by scientific investigation will be a reflection of that with which the natural world has been endowed by its creator: the celebrated *ayat*, or ‘signs’ of divine wisdom and ordering available to human reason.

See also: Aristotelianism in Islamic philosophy; Causation; Causation, Indian theories of; Creation and conservation, religious doctrine of; al-Ghazali; Islamic fundamentalism; Islamic theology; Natural philosophy, medieval; Neoplatonism in Islamic philosophy; Occasionalism; Omnipotence

DAVID BURRELL

References and further reading

Alon, I. (1980) ‘Al-Ghazali on Causality’, *Journal of the American Oriental Society* 100 (4): 397-405.(A clear expositions of al-Ghazali’s options.)

al-Ash‘ari (873-936) *Kitab al-luma’ (The Book of Radiance)*, ed. and trans. R. McCarthy, *The Theology of al-Ash‘ari*, Beirut: Imprimérie Catholique, 1953.(Edition of al-Ash‘ari’s major work.)

Fakhry, M. (1958) *Islamic Occasionalism*, London: Allen & Unwin.(Origin of the cliché.)

al-Farabi (*c*.870-950) *al-Madina al-fadila (The Virtuous City)*, trans. R. Walzer, *Al-Farabi on the Perfect State*, Oxford: Clarendon, 1985.(Clear presentation of the emanation scheme.)

Frank, R. (1983) ‘Moral Obligation in Classical Muslim Theology’, *Journal of Religious Ethics* 11: 210, 228 n.19.(Argues that *kasb* should be translated as ‘performance’.)

Frank, R. (1992) *Creation and the Cosmic System: Al-Ghazali and Avicenna*, Heidelberg: Carl Winter.(A careful study of al-Ghazali’s relation to Ibn Sina.)

Frank, R. (1994) *Al-Ghazali and the Ash‘arite School*, Durham, NC: Duke University Press.(Very innovative argument that al-Ghazali is in many ways closer to Ibn Sina than to the Ash‘arites on a number of topics, including causality.)

Gardet, L. (1950a) ‘*‘illa*’ (Cause), in *Encyclopedia of Islam* 3: 1127-32.(A classic compilation of sources on this topic.)

Gardet, L. (1950b) ‘*Kasb*’ (Acquisition), in *Encyclopedia of Islam* 4: 690-4.(A classic compilation of sources on this topic.)

al-Ghazali (1095) *Tahafut al-falasifa (The Incoherence of the Philosophers)*, ed. M. Bouyges, Beirut: Imprimerie Catholique, 1927; trans. S.A. Kamali, *Al-Ghazali’s Tahafut al-Falasifah*, Lahore: Pakistan Philosophical Congress Publications, 1963.(Al-Ghazali’s refutation of Islamic philosophy.)

Gimaret, D. (1980) *Théories de l’acte humain en théologie musulmane (Theories of Human Action in Islamic Theology)*, Paris: Vrin.(A comprehensive study of human actions, with attention to *kasb*.)

Gimaret, D. (1990) *La doctrine d’al-Ash‘ari*, Paris: Éditions du Cerf.(Places the work of al-Ash‘ari in the perspective of *mutakallimun*.)

Ibn Rushd (1180) *Tahafut al-tahafut (The Incoherence of Incoherence)*, ed. and trans. S. Van den Bergh, London: Luzac, 1978.(Also contains the complete text of al-Ghazali's *Tahafut al-falasifa*, which it criticizes.)

Goodman, L. (1978) 'Did al-Ghazali deny Causality?', *Studia Islamica* 47: 83-120.(Examines a number of texts to find a coherent thread in his assertions.)

Schwartz, M. (1972) '"Acquisition" (*kasb*) in Early Kalam', in S.M. Stern and A. Hourani (eds) *Islamic Philosophy and the Classical Tradition*, Columbia, SC: University of South Carolina Press.(A careful tracing of Qur'anic and other early sources of *kasb*.)

Causation

Two opposed viewpoints raise complementary problems about causation. The first is from Hume: watch the child kick the ball. You see the foot touch the ball and the ball move off. But do you see the foot cause *the ball to move? And if you do not see it, how do you know that that is what happened? Indeed if all our experience is like this, and all of our ideas come from experience, where could we get the idea of causation in the first place?*

The second is from Kant. We can have no ideas at all with which to experience nature - we cannot experience the child as a child nor the motion as a motion - unless we have organized the experience into a causal order in which one thing necessarily gives rise to another. The problem for the Kantian viewpoint is to explain how, in advance of experiencing nature in various specific ways, we are able to provide such a complex organization for our experience.

For the Kantian the objectivity of causality is a presupposition of our experience of events external to ourselves. The Humean viewpoint must find something in our experience that provides sufficient ground for causal claims. Regular associations between putative causes and effects are the proposed solution. This attention to regular associations connects the Humean tradition with modern statistical techniques used in the social sciences to establish causal laws.

Modern discussions focus on three levels of causal discourse. The first is about singular causation: about individual 'causings' that occur at specific times and places, for example, 'the cat lapped up the milk'. The second is about causal laws: laws about what features reliably cause or prevent other features, as in, 'rising inflation prevents unemployment'. The third is about causal powers. These are supposed to determine what kinds of singular causings a feature can produce or what kinds of causal laws can be true of it - 'aspirins have the power to relieve headaches' for example.

Contemporary anglophone work on causality has centred on two questions. First, 'what are the relations among these levels?' The second is from reductive empiricisms of various kinds that try to bar causality from the world, or at least from any aspects of the world that we can find intelligible: 'what is the relation between causality (on any one of the levels) and those features of the world that are supposed to be less problematic?' These latter are taken by different authors to include different things. Sensible or measurable properties like 'redness' or 'electric voltage' have been attributed a legitimacy not available to causal relations like 'lapping-up' or 'pushing over': sometimes it is 'the basic properties studied by physics'. So-called 'occurrent' properties have also been privileged over dispositional properties (like water-solubility) and powers. At the middle level where laws of nature are concerned, laws about regular associations between admissible features - whether these associations are deterministic or probabilistic - have been taken as superior to laws about what kinds of effects given features produce.

1 Hume's legacy

The chief historical locus for the attack on causation is the work of David Hume, which is based on the associationist theory of concept formation that he shared with Locke and Berkeley. Ideas for Hume are either copies of sensory impressions or built out of other ideas, themselves copies of sensory impressions, by simple association. Sensory impressions are of the way things look, taste, feel and sound (see Hume, D. §2). We can have sensory impressions of redness and of hardness, of sourness and shrillness, but not of necessity or causality. Whence come these ideas then? In our experience events of one kind regularly follow those of another. By association we are led to expect the second when we experience the first. It is our impression of our own state of expectation that provides the idea of necessity or of causation. So the idea of causation, according to Hume, is a copy of an impression not of a feature of events in the outside world but of events that take place within us.

Hume therefore steered agnostically away from the direct question 'what is causality?', holding that the right answer (if there is one) must be unintelligible to us. Instead he asked what makes us believe in causal connections and what state of mind is such a belief. Much twentieth-century 'Humeanism' about causation is not so Hume-like as is often implied.

Hume's positive account has been developed in two different directions. The first attempts to analyse causal claims

as claims about regular associations between sensible qualities (or others deemed acceptable) in the external world. The second bases causality in our particular ways of representing the world. A popular version of this looks not to how we represent the world but to how we intervene in it (Von Wright 1993): roughly, we regard as causes those kinds of features that we can manipulate. This idea needs careful formulation if it is to account for our usual causal discourse; for example, it is standardly assumed that craters on the moon cause their shadows but not the converse, though we typically manipulate neither.

As we have seen, Hume’s theory of the one-to-one association of impressions and ideas lies at the heart of his answer to the second central question about causality, the question about the relation between causings and other ‘less problematic’ features of the world. This theory is also central to his answer to the first question about singular causation. There is no impression of one feature causing another, hence no corresponding idea about the world. What then about the world makes a singular causal claim true? The most that can be said is that we believe causal claims at the first level - that of singular causation - when they are instances of second-level claims - about causal laws - which in turn we believe when we meet with regular associations between sensible properties. The only facts that matter about the individual case are that the cause precedes the effect and is spatiotemporally contiguous with it. If in addition there is something about the world that makes effect follow cause we can have no comprehension of it and *a fortiori* no knowledge. Sometimes later accounts try to avoid the assumption that causes precede their effects in order to leave open the question of backward causation; other even aim to reduce claims about temporal order to ones about causal order (Reichenbach 1956) (see Time §4; Relativity theory, philosophical significance of §3). But this has proven difficult to do.

The associationist theory of concept formation central to Hume’s account has been entirely abandoned in twentieth-century philosophy. But his conclusions have not. Modern day defenders of Hume’s theses face three challenges:

(1) Explain why singular causal facts are less legitimate than noncausal facts like ‘the cat was on the mat’.
(2) Show that no causal facts are required to make sense of those features counted as unproblematic and explain what makes these features superior.
(3) Produce a regularity account of causal laws that either accords with or corrects our usual judgments and the uses to which we put them.

The answers to questions (1) and (2) tend to be based on epistemological or methodological considerations: causal facts are so much harder to find out about than the privileged alternative that they can have no place in the known or intelligible world. This answer had a special grip in the heyday of sense-data when facts about our ‘internal’ experiences - which purportedly do not include causal facts - were supposed to be incorrigible. The construction of a reasonable sorting criterion is a more delicate job nowadays when it is widely assumed that nothing is incorrigible and the defence of any single claim will generally require a wealth of knowledge including some laws of nature. Similarly judgments about intelligibility are more difficult at a time when no theory of meaning seems entirely acceptable.

2 Kant’s reply

In *Critique of Pure Reason* (1781/1787), Kant argued that our minds impose the category of causality on the otherwise unorganized manifold of sensory experience, and by this imposition constitute the external world of objects causally interacting within a uniform space and time. Therefore we have a priori assurance that all changes in the phenomenal world proceed in accordance with causal laws which we can discover by observation and experiment (see Kant, I. §§4-7).

Kant rejected Hume’s suggestion that the mind is always passive in forming its expectations and that all our empirical beliefs arise from the association of ideas. The principle of determinism - that every event in the natural order is the effect of some cause - is a necessary truth because it is the presupposition of any experience of an objective world of things. Without this constraining principle we could not distinguish, as we in fact do, between the subjective order of our perceptual experience and the objective order of events in nature. Hume insisted that the proposition that nature is throughout regulated by causal laws could not itself be derived from experience and he was right to do so, according to Kant. This proposition is rather the necessary presupposition of all experience of objects existing in an objective temporal order. We cannot but think of nature as the domain of universal laws

valid for all observers.

3 Mill's empiricism

John Stuart Mill was an empiricist. He believed that all knowledge - even knowledge of mathematics - derives from experience. But he did not share Hume's concern about the idea of causation; nor did he embrace a regularity account of causal law. What occurs in nature is the consequence of a vast number of jointly acting causes. Seldom does any particular arrangement of causes stay in place long enough for regular behaviour to appear. Even if it were to, the regularity would not match with the causal laws we most need to know - laws about what each cause contributes separately.

In many domains there will be laws of this kind. In mechanics, for example, we can find the law governing the force due to gravity and that governing the force due to electricity. Mill called these 'tendency laws' because they tell what the cause contributes, or 'tends to do', but do not describe the effect that actually occurs (here, the 'total' force). What the cause tends to do and what it actually does would coincide exactly were the cause in a position to operate alone. The idea of a tendency law makes sense in mechanics because there is a well-defined rule for the composition of causes: when causes act together, the actual force produced is the vector sum of the forces ascribed to each individually by the basic tendency laws. Chemistry is different: when a cause acts jointly with another it may have no systematic relationship to what it does when the other is absent. In Mill's opinion political economy is like mechanics, not like chemistry.

Mill was concerned not with what causes are but with the proper empiricist question of how we find out about them. He was keen on the experimental method but realized that in many domains experiment is impossible. For these he articulated 'Mill's methods' which involve looking for factors that are present and absent in tandem with the effect and that co-vary in magnitude with it. These methods are a qualitative precursor to modern statistical techniques for confirming causal hypotheses from naturally occurring data (see Statistics and social science §3). In political economy, where he supposed the fundamental driving factor to be behaviour resulting from 'the economic motive', Mill argued that basic causal laws can be found by looking inside ourselves at our motives and actions (Economics, philosophy of §3).

4 Logical positivism

Like various reductive empiricisms inspired by Hume, the logical positivism of the Vienna Circle also eschewed causes (see Logical positivism §4; Vienna Circle §3). Causal talk, according to Rudolf Carnap, imperspicuously casts in the material mode claims that properly belong in the formal mode: it looks as if we are talking about things in the world when instead we are talking about our representations of it. Causality for logical positivism became derivability: 'C causes E' if E can be derived from C in our representation of the world; that is, if we accept that all cases of C are cases of E. The account, however, was beset by problems about accidental regularities. 'All coins in my pocket are silver' does not ground a causal connection. So some notion of necessity or lawlikeness has generally been admitted: 'all Cs are *by law* Es'. Still whatever is involved in separating lawlike from accidental regularities has been supposed to be far weaker than a full-fledged concept of causation (see Laws, natural §2).

This account has obvious counter-intuitive results, pinpointed in an example by Sylvan Bromberger (1966). Given the angle of the sun and the laws of geometrical optics, from the height of the flagpole the length of its shadow can be derived - and the converse. Yet surely the height of the pole causes the length of the shadow and not the reverse. For staunch positivists this argument illustrates what is wrong with causal talk - it postulates differences in the world where in fact there are none. For others it constitutes a *reductio ad absurdum* of the reductivist/positivist position.

If strict derivability under universal laws is too weak as a stand-in for causal laws, it is also too strong, as C.G. Hempel admitted in his *inductive-statistical* model of explanation: E can be explained by (or accepted on the basis of) C if we admit as law that the probability of E given C is high, even though the association is not 100 per cent (see Explanation §2). (Many would find the expression 'as law' redundant here for they take probability, as opposed to finite frequency, to be already a modal notion.) The move to less-than-universal associations is motivated not only by the irreducibly probabilistic nature of quantum associations but also because we could admit few explanations indeed in either the natural or the social sciences if universal (deterministic) association is

demanded.

But high probability does not work either, as Michael Scriven argued with the example of paresis, which arises only in the last stages of syphilis. The probability of paresis given syphilis is low. Nevertheless when paresis occurs, syphilis is clearly the cause. This suggests that increase in probability is a better requirement: C causes E just in case the probability of E is higher than C than without.

Over and above these problems it is well-known that ‘correlation is not causation’. Two kinds of cases illustrate. First, joint effects of a common cause are generally correlated with one another. For example, there is an anti-correlation between eating candy and getting divorced. But low candy consumption in a given period does not cause divorce during that period; rather, ageing is responsible for both. Second is a case due to Wesley Salmon *et al.* (1971). A box contains either of two radioactive elements, L and H, but never both. Each element produces W particles, L at a lower rate than H. L causes the presence of Ws in the box: if this situation persists L will predictably (or necessarily) produce Ws. Yet L may reduce the probability of W since the presence of L is correlated with the absence of even more effective cause H. Indeed, with a little care in the choice of the levels of intensity of L and H, the presence of L will raise the probability of W being present (see Determinism and indeterminism; Explanation §4).

The example suggests that there may be no universal rule for connecting probabilistic associations with causal laws although there may well be correct case-by-case connections. In the radioactivity case, the fact that Ls predictably cause Ws, plus facts about the situation, explain why the probability of W given L is exactly as it is. If one adopts some form of inference to the best explanation, the probabilities - whatever they are - will be evidence for the causal law (see Inference to the best explanation). Moreover, given sufficient collateral information it is possible to deduce from the probabilistic facts that Ls produce Ws in this set up. (Although uncommon in the social sciences, this kind of ‘bootstrapping inference’ that allows the deduction of the hypothesis from data, given general background assumptions and other information specific to the situation, is familiar in experiments in physics.) Causality may after all be an umbrella notion with different criteria in different contexts.

5 Causal laws

By this point, it seems, the logical positivist account that locates the truth of causal-law claims in our representations of the world has merged with the Hume-inspired account that locates their truth in laws about regular associations. The difference between them depends on whether one talks about the regular associations we accept in our scientific theories - laws of science - or about laws of nature. Nevertheless they share the insistence that there must be a universal rule that connects causal laws and the laws of association, since neither tradition is willing to admit any concept of causation that cannot be reduced to facts they find less problematic. It remains to produce such a rule. The most promising attempts, originally formulated by Patrick Suppes (1970, 1973), try to restore increase in probability as the true sign of causality.

In both the candy-eating and radioactivity examples, the putative cause and the putative effect are each correlated with a third factor that is itself a cause of the putative effect. Candy consumption and divorce are correlated with age, which causes increase in divorce levels; L is correlated with H, itself a cause of W. Stratification is often used to control for features like age or the presence of other materials that may confound the correlations we expect between a genuine cause and its effects. The proposal then is that ‘candy consumption prevents divorce’ is true just in case candy consumption lowers the probability of divorce in subgroups where everyone is the same age. But what features should be used for stratification? Mathematically, correlations between C and E may appear and disappear if we stratify along any third variable correlated with both. The most natural choice is to use all and only other factors - like age - that are themselves causes of E.

Besides the obvious question of how exactly this suggestion should be formulated, the proposal faces three problems. First, in cases where a cause may produce two effects in tandem - cases of the production of products and by-products for example - joint effects may still be correlated with each other even though we look only at stratified subpopulations in which the cause is present to the same degree.

Second, the proposal abandons the programme of replacing causal laws with law-like regularities. Probabilistic associations may still help with causal inference, but only in so far as one has a great deal of causal information to begin with: no causes in, no causes out.

Third is the problem of validating the proposal - this one or any other. All available validations (Simon 1971; Spirtes, Glymour and Scheines 1993) put strong constraints on what kinds of causal relationships are possible among a set of variables as well as on the general relationship between causal laws and probabilities. Thus the task for those who wish to maintain some single situation-independent connection between causal laws and laws of associations is either to generalize these kinds of validations to cover laws or to produce convincing reasons why other kinds of cases are impossible.

The task for those who wish to reduce causal laws to probabilistic associations (Papineau 1993) is even more difficult, as can be seen by considering results of groups working with Clark Glymour and Peter Sprites and with Judea Pearl (1988). These provide the most systematic treatment available of causal inference from probabilistic information. In this work, for a given set of variables it is assumed that both a causal structure (a set of causal laws satisfying certain conditions like asymmetry) and a probability measure over the variables exist and some plausible constraints are assumed to hold between the probabilities and the causal structure. It is then possible to generate the set of all causal structures consistent with any particular set of observed probabilities given the constraints. For most cases, however, the causal structure is not unique. The probabilistic associations do not seem enough to fix the causal laws. This is the problem of the flagpole and the shadow reasserting itself again after a long detour.

Alternative conclusions can be defended. The results are sympathetic to the view that causal laws do not reduce to laws of association but are rather about what kinds of singular causal events happen predictably or by law. On the other hand the results fit with a programme of locating the choice among different causal structures consistent with the same probabilities in facts about our perspective or our interests, or even with the claim that causal talk is hopelessly underdetermined and nothing fixes what we should say. And of course there is the project of articulating further constraints so that the probabilistic associations fix the causal laws uniquely. What definitely can be concluded is that no reduction of causal-law claims to claims about law-like or necessary associations has yet been achieved, let alone the stronger Hume-inspired view that causal-law talk can be reduced to claims about associations that employ no concepts of law or necessity at all.

A final outstanding problem for any account is to link the analysis of what causal claims say with our usual methods for confirming them. The regularity accounts described previously relate readily to standard techniques based on Mill’s methods for the use of population data. The connection to results of controlled experiments is less clear (Cartwright 1995). Yet these are supposed to be our surest methods for causal inference; indeed so much so that some have taken causal-law claims simply to be claims about the results that would be obtained from controlled experiments (Holland 1986) (see Experiment; Scientific method).

6 Singular causal claims

Singular causal claims may be true because of singular causal facts. Two much discussed problems seem to presuppose this. First is ‘overdetermination’: two factors, each by law sufficient for an effect, occur and the effect ensues. Which is actually responsible? The second is ‘causal pre-emption’: a cause occurs sufficient by law to bring about an effect but a second cause intervenes and produces the effect first (Menzies 1996). Both problems suggest that there is more to causation in the single case than reductive empiricists and logical positivists allow. Yet some resist, feeling that this claim is unintelligible. The challenge they pose is to ground this claim in some plausible theory of meaning that will support their intuition.

Others reject singular causal facts because they find no grounds for judgments about them. To this writers like G.E.M. Anscombe (1971) reply that we can see the cat lapping up the milk as well as we can see anything! This defence is supported by theories of perception in which the ability to judge correctly that something has any feature at all by seeing it depends on having a good deal of collateral knowledge, including knowledge of both singular causal facts and causal laws. There are in addition, especially for causal processes in physics, accounts of singular causation in terms of other features conventionally admissible, such as energy interchange (Salmon 1984; Dowe 1992). It is also argued that the operation of a cause often leaves traces that can be used to determine its occurrence.

The question of whether there is significantly more to singular causation than lawful association is different from the question: for every singular causal fact is there a law that it can be brought under? Donald Davidson (1980) replies ‘yes’ to this latter question, and the same answer is presupposed by J.L. Mackie (1980), who takes causes

to be INUS conditions - insufficient but non-redundant parts of unnecessary but sufficient conditions. Mackie's account supposes that there are general laws of the form

$$E \equiv (C_1^1 C_1^2 \ldots C_1^n) \vee (C_2^1 C_2^2 \ldots C_2^m) \vee \ldots \vee (C_p^1 C_p^2 \ldots C_p^k).$$

This allows us to focus on one of the contributing components, C_i^j, as the cause of interest in a given context, even though this component alone is not sufficient. (However it does not capture Anscombe 's denial that causes must be sufficient for their effects since she has in mind purely probabilistic or chance causation: could we not have a cause that is enough for the effect and yet the effect fail?) Some considerations of how to characterize properties enter here: if this C that is supposed to have caused an E is not doing what it does regularly, what makes it a C? But the usual arguments are epistemological: how could we know that a C caused E if this fact neither falls under a causal law nor is the expression of a known causal power? For instance, how could we count a factor as a trace if there were no law ensuring that the operation of the cause generally produces this factor as a side-effect?

Midway between the view that there are singular causal facts and reductivist/positivist doctrines is the counterfactual account (Lewis 1986): roughly, C caused E just in case if C had not occurred E would not have. One task is to formulate this correctly, especially to avoid problems of overdetermination and pre-emption and to ensure that one effect of a common cause does not count as cause of a joint effect. The second task to explain is counterfactuals (see Counterfactual conditionals). One common attempt offers a possible-world analysis. The counterfactual is true if in the nearest possible world where E fails, C fails too. Possible worlds are generally taken to be worlds which have the same laws as our own, except for small miracles to allow C to fail. On one reading possible worlds are mere calculational devices: one function maps from laws and non-causal facts to a set of structured descriptions ('the set of possible worlds'), and a second takes the set of descriptions plus the counterfactual in question into either 'true' or 'false'. Singular causal claims would then turn out to be indirect ways of talking about laws and non-causal facts, and the logical positivist programme would be in part borne out. If, however, the ontology of possible worlds plays a genuine role, the challenge is to defend the advantage of possible worlds over singular causal facts. A similar challenge must be met by probabilistic accounts of singular causation that use special notions of probability like single-case propensities (Eells 1991): why are these better than the cat's lapping up the milk?

7 Causal powers

The idea of a causal power or a disposition (Harré and Madden 1975) is closely bound to notions of potentiality and possibility related to the notions of necessity that reductive empiricists and logical positivists find unintelligible. They understand only claims about what occurs, or perhaps what occurs by natural law, not about what can occur by natural law. Proponents of causal powers often counter that there is no distinction between powers or dispositions and occurrent properties (Mellor 1971, 1991, 1995). Objects have properties and by virtue of having these properties they have the power to do a variety of things; otherwise the notion of property makes no sense. Perhaps properties are just nexus of causal powers (Shoemaker 1984). This raises the question of whether an object could have a particular power all on its own without thereby having any other usually related features (see Primary-secondary distinction; Property theory).

For the most part, however, questions centre on the relation between causal powers on the one hand and causal laws and singular causation on the other. Some powers are, in Gilbert Ryle's terminology, highly specific or determinate (1949). They tend to be back-named from some episodic verb that is taken to describe their natural expression (as in 'soluble' and 'dissolve'). This suggests a one-to-one correspondence between powers and laws. Aspirins have the power to relieve headaches just in case, by law, aspirins properly administered relieve headaches. Talk of powers then seems otiose.

The matter appears differently when more generic powers are considered. 'Electrons repel other electrons' describes a power that can be harnessed to produce an indefinite variety of different law-like behaviours in different situations; situations can even be rigged in which electrons reliably, by repelling other electrons, cause them to move closer. One can restore the one-to-one correspondence by inventing or choosing an appropriate episodic verb, though the attempt can seem strained. (As Ryle urges, grocers sell tea, wrap butter and weigh bacon. But do we want to say that grocers groce?) Nor does this unproblematically serve the reductivist/positivist purpose of rendering talk about powers as talk about laws, for is not 'repelling electrons' in the situation where they move

closer a specific case of 'exercising a power'? The relationship between powers and the indefinite and un-inventoriable class of laws that express them thus remains a central philosophical concern.

See also: Causality and necessity in Islamic thought; Causation, Indian theories of; Paranormal phenomena

NANCY CARTWRIGHT

References and further reading

Anscombe, G.E.M. (1971) *Causality and Determination*, Cambridge: Cambridge University Press.(Sophisticated but accessible to beginners.)

Bromberger, S. (1966) 'Why-Questions', in R.G. Colodny (ed.) *Mind and Cosmos: Essays in Contemporary Science and Philosophy*, Pittsburgh, PA: University of Pittsburgh Press.(Accessible to beginners.)

Carnap, R. (1934) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul, Trench, Trubner & Co., 1937.(Advanced study in logical positivism.)

Cartwright, N. (1989) *Nature's Capacities and Their Measurement*, Oxford: Oxford University Press.(See chapter 3. Advanced.)

Cartwright, N. (1995) 'Probabilities and Experiments', *Journal of Econometrics* 67: 47-59. (Advanced.)

Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Clarendon Press. (Intermediate-level study of singular causation.)

Dowe, P. (1992) 'Wesley Salmon's Process Theory of Causality and Conserved Quantity Theory', *Philosophy of Science* 59: 195-216.(Intermediate.)

Eells, E. (1991) *Probabilistic Causality*, Cambridge: Cambridge University Press. (Advanced study of singular causation.)

Harré, R. and Madden, T.E. (1975) *Causal Powers*, Oxford: Blackwell.

Hempel, C. (1966) *Philosophy of Natural Science*, Englewood Cliffs, NJ: Prentice Hall.(Very good for beginners.)

Holland, P. (1986) 'Statistics and Causal Inference', *Journal of the American Statistical Association* 81: 945-60. (Advanced.)

Hume, D. (1739-40) *A Treatise of Human Nature*, Oxford: Clarendon Press, 1888.(Essential.)

Hume, D. (1902) *An Inquiry Concerning Human Understanding*, Oxford: Clarendon Press. (Essential.)

Kant, I. (1781/1787) *The Critique of Pure Reason*; trans. N.K. Smith, London: Macmillan, 1929.(Essential; difficult.)

Lewis, D. (1986) *Philosophical Papers*, Oxford: Oxford University Press, vol. 2, 159-84.(Advanced.)

Mackie, J.L. (1980) *The Cement of the Universe: A Study of Causation*, Oxford: Clarendon Press. (Intermediate-advanced; a good survey.)

Mellor, D.H. (1971) *The Matter of Chance*, Cambridge: Cambridge University Press.(See chapters 4 and 8. Intermediate.)

Mellor, D.H. (1991) *Matters of Metaphysics*, Cambridge: Cambridge University Press. (Intermediate.)

Mellor, D.H. (1995) *The Facts of Causation*, London: Routledge.(Intermediate.)

Menzies, P. (1996) 'Probabilistic Causation and the Pre-emption Problem', *Mind* 105: 85-117.(Intermediate.)

Papineau, D. (1993) 'Can We Reduce Causal Directions to Probabilities?', in A. Fine and M. Forbes (eds) *PSA 1992*, East Lansing, MI: Philosophy of Science Association, vol. 2, 238-52.(Intermediate.)

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA: Morgan Kaufmann Publishers.(Advanced.)

Reichenbach, H. (1956) *The Direction of Time*, Berkeley, CA: University of California Press.(Intermediate, but parts are good for beginners.)

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.(Good for beginners.)

Salmon, W., Jeffrey, R.C. and Greeno, J.G. (1971) *Statistical Explanation and Statistical Relevance*, Princeton, NJ: Princeton University Press.(Intermediate, but parts are good for beginners.)

Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*, Princeton, NJ: Princeton University Press.(Intermediate, but parts are good for beginners.)

Shoemaker, S. (1984) *Identity, Cause and Mind: Philosophy Essays*, Cambridge: Cambridge University Press. (Advanced.)

Simon, H. (1971) 'Spurious Correlation: A Causal Interpretation', in H.M. Blalock (ed.) *Causal Models in the Social Sciences*, Chicago, IL: Atherton.(Advanced.)

Spirtes, P., Glymour, C. and Scheines, R. (1993) *Causation, Prediction and Search*, Lecture notes on statistics 81, New York: Springer.(Advanced.)

Suppes, P. (1970) *A Probabilistic Theory of Causality*, Amsterdam: North Holland. (Advanced.)

Suppes, P. (1973) ‘Logic, Methodology and the Philosophy of Science’, in E. Sosa and M. Tooley (eds) *Causation*, Oxford: Oxford University Press, 1993.(Intermediate-advanced.)
Wright, G.H. Von (1993) ‘On the Logic and Epistemology of Causal Relations’, in E. Sosa and M. Tooley (eds) *Causation*, Oxford: Oxford University Press, 1993. (Intermediate-advanced.)

Causation, Indian theories of

*Causation was acknowledged as one of the central problems in Indian philosophy. The classical Indian philosophers' concern with the problem basically arose from two sources: first, the cosmogonic speculations of the Vedas and the Upaniṣads, with their search for some simple unitary cause for the origin of this complex universe; and second, the Vedic concern with ritual action (*karman*) and the causal mechanisms by which such actions bring about their unseen, but purportedly cosmic, effects. Once the goal of liberation (*mokṣa*) came to be accepted as the highest value, these two strands of thought entwined to generate intense interest in the notion of causation. The systematic philosophers of the classical and medieval periods criticized and defended competing theories of causation. These theories were motivated partly by a desire to guarantee the efficacy of action and hence the possibility of attaining liberation, partly by a desire to understand the nature of the world and hence how to negotiate our way in it so as to attain liberation.*

Indian philosophers extensively discussed a number of issues relating to causation, including the nature of the causal relation, the definitions of cause and effect, and classifications of kinds of causes. Typically they stressed the importance of the material cause, rather than (as in Western philosophy) the efficient cause. In India only the Cārvāka materialists denied causation or took it to be subjective. This is unsurprising given that a concern with demonstrating the possibility of liberation motivated the theories of causation, for only the Cārvākas denied this possibility. The orthodox Hindu philosophers and the heterodox Buddhists and Jainas all accepted both the possibility of liberation and the reality of causation, though they differed sharply (and polemically) about the details.

*The Indian theories of causation are traditionally classified by reference to the question of whether the effect is a mode of the cause. According to this taxonomy there are two principal theories of causation. One is the identity theory (*satkāryavāda*), which holds that the effect is identical with the cause, a manifestation of what is potential in the cause. This is the Sāōkhya-Yoga view, though that school's particular version of it is sometimes called transformation theory (*pariṇāmavāda*). Advaita Vedānta holds an appearance theory (*vivartavāda*), which is often considered a variant of the identity theory. According to the appearance theory effects are mere appearances of the underlying reality, Brahman. Since only Brahman truly exists, this theory is also sometimes called* satkāraṇavāda *(the theory that the cause is real but the effect is not).*

*The other principal theory of causation is the nonidentity theory (*asatkāryavāda*), which denies that the effect pre-exists in its cause and claims instead that the effect is an altogether new entity. Both adherents of Nyāya-Vaiśeṣika and the Buddhists are usually classified as nonidentity theorists, but they differ on many important details. One of these is whether the cause continues to exist after the appearance of the effect: Nyāya-Vaiśeṣika claims it does, the Buddhists mostly claim it does not.*

Finally, some philosophers try to take the middle ground and claim that an effect is both identical and nonidentical with its cause. This is the position of the Jainas and of some theistic schools of Vedānta.

1 The context of the Indian theories

The commitment to the ideal of liberation (*mokṣa*) provides the context for understanding the classical Indian philosophers' concern with causation. Typically the theoretical problem of Indian philosophy is to provide an account of the world which allows for the possibility of our successfully entering into it as agents set on liberation from suffering (*duḥkha*). In order to guarantee the feasibility of liberation we need to be assured that there are reliable causal connections between events and actions such that it is possible for a person to enter into the course of events as a conscious agent whose actions have predictable consequences. To do this the Indian philosophers sought to identify those causal chains relevant to liberation, and to analyse the nature of the causal relation as exhibited among the members of those chains.

There are a number of such causal chains proposed. One of the oldest and best known is the Buddhist chain of dependent origination (*pratītyasamutpāda*). This twelvefold chain runs: ignorance; dispositions; consciousness; body and mind; the six sense fields; sense contact; sensation; desire; clinging; becoming; birth; old age and death. Each of these factors is both conditioning and conditioned. This is expressed in a traditional Buddhist formula

characterizing the relation between the links: ‘When *A* is, *B* is; *A* arising, *B* arises; when *A* is not, *B* is not; *A* ceasing, *B* ceases.’ The possibility of liberation (called *nirvāṇa* in Buddhism) requires that we be able to break the chain at certain points where the links are only necessary conditions for what follows. In Buddhism the favoured weak links are ignorance (*avidyā*) and desire (*tṛṣṇa*), which can be eliminated with (respectively) knowledge and nonattachment (see Suffering, Buddhist views of origination of).

The alternative causal chains of the Jainas, of Nyāya-Vaiśeṣika and of Sāōkhya-Yoga differ in their details. However, all include ignorance as a link and hence guarantee the possibility of liberation through the possibility of the elimination of ignorance by right knowledge. It is essential for the possibility of liberation that the causal chains which bind us to suffering are such that at least some of the members are necessary but not sufficient conditions for other members, otherwise we could not enter the chain and reverse our condition of bondage. On the other hand, the chain must not have gaps, for if there is no necessary connection between the links none of our actions can be relied upon to bring about the goal of liberation. Thus once we have identified the causal chain that leads to bondage, we also need to say something more about the nature of the causal relation itself.

Two basic models dominate Indian thinking about the nature of the causal relation. One is the model favoured by the identity theory of Sāōkhya-Yoga. A standard illustration is the case of milk and curds: the milk is the cause and the curds are the effect. But the milk is the same stuff as the curds; one is merely transformed into the other. Generalizing from this kind of example leads to the view that the effect pre-exists in its cause. In fact a cause and its effect are not two separate, discrete entities but instead two states of the same enduring substance. The major difficulty with this model (in both its Sāōkhya and Advaitin versions) is that the causal relation threatens to be too strong. If nothing can become other than what it is already, how can we unenlightened beings ever achieve our own liberation?

The competing model of the causal relation is the one favoured by the nonidentity theory of Nyāya-Vaiśeṣika and (some) Buddhists. A standard Naiyāyika illustration is the pot and its two halves: the two halves of the pot are its causes and the pot is the effect. (Indian potters make a pot by first making two halves and then joining them.) According to this model the effect is not pre-existent in the cause. Effects are instead conceived of as wholes inhering in pre-existing parts which are the causes.

The Buddhist versions of the nonidentity theory are rather different because Buddhism generally denies that there are any persisting substances at all; everything is momentary. A standard illustration is the way in which whirling a flaming torch creates in watchers the impression of a persistent object: a circle of fire. Similarly, the momentary occurrences of things in certain patterns create the impression of persisting objects existing in the world.

Both versions of the nonidentity theory claim to avoid the error of making the causal relation too strong, for effects are not identical with their pre-existing causes. They also both claim to avoid making the causal relation too weak, for both in their different fashions try to guarantee the regularity of causal relations. However, to vindicate the latter claim each school has to develop a rather elaborate ontology and epistemology, the details of which come in for criticism from other schools.

A third position tries to occupy the middle ground between the identity theory and the nonidentity theory. This is the stance taken by the Jaina theory of nonabsolutism. According to this theory there is a variety of aspects any entity can be viewed from. Thus from one viewpoint the effect is pre-existent in the cause and from another viewpoint it is not. The theory attempts to provide a compromise account of the causal relation that is neither too strong (as the identity theory threatens to be), nor too weak (as the nonidentity theory threatens to be).

Common to all of these causal theories, however, is the assumption that causation is real and not merely subjective. This is natural enough, since all these theories accept the original problematic: how to analyse the nature of the causal chain so as to guarantee the feasibility of attaining liberation. The one exception to this in Indian philosophy is to be found in the views of the Cārvāka materialists. These philosophers were sceptics about both causation and the possibility of attaining liberation. They espoused an antireligious materialism and a subjectivist account of causation as being merely observed conjunctions of events.

We can arrange, then, the Indian theories of causation in the following sequence. First, we have Cārvāka scepticism about causation, with its attendant scepticism about the possibility of liberation. All the other theories can be represented as various defensive responses to this scepticism. Thus the identity theory seeks to guarantee

the possibility of liberation with an account of causation that makes it a very strong relation. Two distinct versions of the theory are developed: the transformation theory of Sāõkhya-Yoga and the appearance theory of Advaita Vedānta. The nonidentity theory, on the other hand, tries to preserve the possibility of liberation while making causation a rather weaker relation. Again, two versions of the theory are developed: the Nyāya-Vaiśeṣika theory and the Buddhist theory of conditioned origination. The Jaina theory is an attempted synthesis of the identity and nonidentity theories.

2 Cārvāka scepticism

As antireligious materialists, the Cārvākas had no interest in vindicating the possibility of liberation (see Materialism, Indian school of §4). And as strict empiricists in their epistemology, they refused to admit anything but perception as a valid means of knowledge. Accordingly they refused to admit causation as an invariable and unconditional relation. All we can know is what we perceive and all we perceive are conjunctions of events, not a dependence relation between events. These conjunctions may be regarded as purely accidental: hence their views are sometimes known as accidentalism (*yadṛcchāvāda*). Rather than supposing some things are effects dependent on other things which are causes, the Cārvākas held that it is more reasonable to suppose things occur because of their own natures (*svabhāva*): hence their views are sometimes called naturalism (*svabhāvavāda*).

This sceptical position was criticized by all the other schools. The most common criticism was of the overly restrictive epistemology that led to this sceptical result. Indeed, the Cārvāka position was often viewed as a *reductio ad absurdum* of its strict empiricist premises. Accordingly the response of the other schools was usually to try to develop from a less restrictive epistemology a more generous ontology which included causal relations.

Another common complaint, however, was that Cārvāka scepticism was self-refuting. One Naiyāyika version of this charge claims that the scepticism about causation is refuted by the sceptics' own behaviour: they purport to doubt the causal relation between fire and smoke, but light a fire when they want to produce smoke. Of course, the Cārvāka might just reply here that we are so constituted psychologically that we expect a uniform regularity between instances of fire and smoke. Notwithstanding this, there is no real justification for the expectation; it is just a habit of expecting what has previously occurred in certain circumstances to reoccur in similar circumstances. But for the liberation-oriented philosophers, the Cārvāka's philosophical anthropology here is far too pessimistic. The agent trapped in the patterns of habit is paradigmatically the unliberated being; liberation is freedom from such bondage and understanding the causal chains which lead to bondage also allows us to discover the route to freedom.

3 Sāõkhya-Yoga and the transformation theory

Sāõkhya-Yoga espouses the identity theory (*satkāryavāda*), so called because it holds that an effect (*kārya*) is already existent (*sat*) in its cause in a potential form. The Sāõkhya version of the theory is also called transformation theory (*pariṇāmavāda*) because it holds that the cause undergoes a real transformation into its effect through the causal process.

The standard Sāõkhya arguments for the identity theory presented in Īśvarakṛṣṇa's *Sāõkhyakārikā (Verses on Distinctionism)* (verse 9) are basically that something cannot emerge out of nothing, that the effect must be of the same material as the cause and that specific causes can only produce specific effects. In order to guarantee all of this the effect must pre-exist in the cause; it is a modification of what was already present. Thus according to the metaphysics of Sāõkhya the manifest world must have an existing cause that the effect pre-exists in. This is nature (*prakṛti*), conceived of as a unitary principle underlying observable phenomena, which are transformations of this substance. The self (*puruṣa*) is merely the passive witness of all this.

The key Sāõkhya examples of causation all involve material causation: as when a seed grows into a plant, or milk is transformed into curd, or oil seeds into oil. In these cases a cause and its effect are plausibly just two states of a single continuing substance. The Naiyāyikas objected that the Sāõkhya theory abolishes the distinction between material and efficient causes. But this is not quite true, for Sāõkhya-Yoga also admits another type of cause: the efficient or instrumental cause (*nimittakāraṇa*). However, this is not supposed to act upon the material cause and transform it into an effect. Rather it simply removes the barriers which check the material cause (*prakṛti*) from transforming from a relatively unmanifested state to a more manifested state. The *Yogabhāṣya* (4.3) compares it to

how a farmer allows water to flow from a filled bed to another just by removing the obstacles.

The transformation theory insists that there is a necessary relation between cause and effect. To this extent it responds to the Cārvāka sceptical challenge to the possibility of liberation: there is indeed an invariable concomitance between cause and effect, since the latter is just a manifested state of the former. But the theory threatens to make the causal relation too strong. Transformation theory holds causation to involve a real transformation of a common stuff. But then nothing can become other than it is already and the presently unenlightened can never attain liberation. Sāōkhya-Yoga responds to this objection by developing a radical dualism between nature (*prakṛti*) and self (*puruṣa*). The self is essentially unaffected by the causal transformations of *prakṛti*. As the *Sāōkhyakārikā* (62) startlingly puts it: 'No one, therefore, is bound; no one released, likewise no one transmigrates. Only *prakṛti* in its various forms transmigrates, is bound and is released'. Because the *puruṣa* is just pure, contentless consciousness it cannot be bound or liberated. Realizing the absolute separation of *prakṛti* and *puruṣa*, ceasing to misidentify ourselves with our bodies, we come to appreciate our true natures as pure consciousnesses. This realization leads to liberation (called *kaivalya* in Sāōkhya-Yoga), a condition apart from all suffering. The difference between bondage and liberation, then, is not an ontological one, but an epistemological one. The removal of the epistemological condition of ignorance is sufficient for liberation (see Sāōkhya).

4 Advaita Vedānta and the appearance theory

The Sāōkhya epistemological model of the route to freedom has much in common with the Advaitin approach to the problem. However, the Advaitin view came to be called appearance theory (*vivartavāda*), for it differs importantly from the Sāōkhya-Yoga version of the identity theory. In particular, the transformation theory views an effect such as a pot as a genuine transformation of the clay which constitutes it; both cause and effect are real. By contrast, the appearance theory views the effect (the pot) as not real, but only an appearance (*vivarta*). This is because Advaita espouses a radical monism: only the Absolute (Brahman) is real and the Self (*ātman*) is identical with Brahman (see Brahman; Monism, Indian). Accordingly the Advaitin theory of causation is a version of the identity theory in that the effect (the illusory world around us) is in a sense not different from its cause (Brahman/*ātman*). However, the effect is ultimately unreal, though the cause is real (thus the theory is sometimes called *satkāraṇavāda*, 'existent-cause-theory').

It is only on the level of phenomenal reality, then, that the Advaitins are willing to defend the identity theory. The Advaitin philosopher Śaōkara, for instance, endorses on this level the familiar Sāōkhya arguments that otherwise anything might come of anything and that clearly nothing comes of nothing. He also adds others of his own, including the suggestion that since the perceptibility of cause and effect are not independent it is reasonable to suppose they are identical. But ultimately, when we consider the relation between the world and Brahman, the effect is merely an *apparent* effect. This is because the Advaitins accept both that Brahman is an eternal being and that an eternal cause must have eternal effects. But since worldly phenomena are clearly not eternal, they conclude that they cannot be genuine effects of Brahman, merely illusory or apparent effects. In so far as Brahman underlies these appearances, however, it can be viewed as the material cause of the world. This is in accord with the Advaitin theory of perceptual error, according to which there must be something real that underlies a false appearance.

On the appearance theory, then, causality is an apparent relation between a (comparatively) unreal effect and a (comparatively) real cause, between a thought construction and that which grounds such a construction. The focus of the theory of causation thus shifts away from a concern with external relations between objects to a concern with the epistemic or awareness relation involved in such constructions. Accordingly liberation is conceived of epistemically as the realization of what one already essentially is: Brahman.

Clearly this theory of causation is only as plausible as the concept of Brahman as pure being upon which it rests. But this latter notion was vigorously rejected by many other Indian philosophers and the appearance theory thus requires a controversial monism to support it. Moreover, the identity theory's attempt to guarantee the causal relation, begun by Sāōkhya and continued by Advaita, seems to end up with too strong an account of causation. The only way out of this difficulty is to insist that, in some sense, we are already liberated but do not know it. Then liberation becomes an epistemological matter, not an ontological one. But with bondage no longer conceived in material terms, we have a corresponding drift both away from epistemological realism and towards metaphysical dualism or monism.

5 Nyāya-Vaiśeṣika and the nonidentity theory

Nyāya-Vaiśeṣika represents a robust commitment to both epistemological realism and metaphysical pluralism (see Nyāya-Vaiśeṣika). The Naiyāyikas define a cause as an invariable and independently necessary antecedent of an effect. That is, the causal relation is a uniform temporal relation that is necessary in the sense that there can be no counterinstances (though the relation is not a logical one in the Western sense). Moreover, the constant conjunction involved is a relation between properties, rather than between particular events.

Nyāya recognizes three kinds of causal factor: inherent cause (*samavāyikāraṇa*); noninherent cause (*asamavāyikāraṇa*); and instrumental cause (*nimittakāraṇa*). The inherent cause is that substance in which the effect abides by the relation of inherence. Thus the pot-halves or the threads are the inherent causes of the pot or the cloth because the latter effects inhere in the former causes. Note that for Nyāya the halves or the threads are not that out of which the pot or the cloth are composed. Rather the effects inhere in the cause so that, for instance, the cloth is not produced *out of* the threads, but subsists *in* the threads.

The noninherent cause is a cause which (directly or indirectly) inheres in an inherent cause. For example, in the production of a pot, the pot-halves are the inherent cause of the pot and the contact between the pot-halves, which inheres in the pot-halves, is the noninherent cause of the pot.

The first two kinds of causal factor are together necessary but not sufficient to produce an effect. The category of instrumental (or efficient) cause lumps together all the remaining causal factors. These include the agents of actions and other supporting factors.

The Nyāya theory takes the effect to be an absolutely new thing. The Sāōkhya argument that a nonexistent effect cannot be brought into existence is dismissed by the Naiyāyikas as confusing an absolute nonentity (like the hare’s horn, which is nonexistent for all time) with what is merely nonexistent before a particular time (like the pot before it is produced by the potter).

Nyāya also rejects the Sāōkhya argument that since not just anything can produce anything, there must be a necessary relation between cause and effect requiring that they coexist contemporaneously. The Naiyāyikas claim the relevant necessity is supplied by the fact that the relation is between universals, not particulars.

Essential to the Nyāya theory is the notion of inherence (*samavāya*). Inherence is the relation that connects wholes and parts (like pots and pot-halves, threads and cloth); it also connects substances and their qualities. Inherence is defined as the relation between two inseparable things related as located to locus. Inherence explains the relation of the pot to the pot-halves which are its material cause without falling into the identity theory’s mistake of identifying the effect with its cause (and hence being unable to explain why the effect does not come into existence as soon as the cause does). Moreover, inherence relates the self (*ātman*) to its qualities, including wrong notions. This allows for the wrong notions to be destroyed without thereby destroying the self, thus guaranteeing the possibility of liberation.

The most popular objection to the Nyāya theory is the infinite regress argument against inherence: if two entities *A* and *B* are to be related by the inherence relation *R*, which is itself a distinct entity, then it is also necessary that *A* and *R* be related by a different inherence relation *R**, itself a distinct entity. But then, of course, *A* and *R** have to be related by a yet different inherence relation *R***, and so on *ad infinitum*.

The Naiyāyikas reply that there is no regress because there is no other relation to connect inherence to its relatum. Clearly they cannot mean that the relatum and its relation are identical, for then, by the transitivity of identity, *A* would not only be identical with *R*, but also with *B*! Instead Navya-Nyāya (‘New Nyāya’, the later school of Nyāya) appeals to the notion of a self-linking connector (*svarūpasambandha*). The idea here is that while *A* requires the inherence relation *R* to connect it to *B*, *A* can be its own connector to *R*.

6 Buddhist theories

The Buddhists reject the Nyāya version of nonidentity theory because it seems to them that inherence is too strong a relation for causation. This is because the Buddhist theory of momentariness (*kṣaṇikavāda*) implies that there can be no persisting relation between any two entities, nor any persisting entities (see Momentariness, Buddhist

doctrine of). Instead they espouse an ontology of momentary events, each of which is causally efficacious, grouped into various patterns. Moreover, the theory of dependent origination is understood to imply that an effect is not the result of a single cause, but of many causes working together. The Buddhist schools attempted various classifications of this totality of causes and conditions.

The Abhidharma analyses reality into elements (*dharmas*) (see Buddhism, Ābhidharmika schools of). A distinction is also admitted between a *dharma* and its characteristics (*lakṣaṇas*). But this distinction quickly leads to a quasi-substantialism in the Sarvāstivāda school, as the concept of a *dharma*'s enduring essence or 'own-nature' (*svabhāva*) is introduced as the bearer of a *dharma*'s 'own-characteristics' (*svalakṣaṇa*). The concept of *svabhāva* is utilized by the Sarvāstivādins to explain the continuity of phenomena, which are analysed into momentary existences: one aspect of a *dharma* changes while another (the *svabhāva*) remains unchanged. This idea is used to explain the connection between cause and effect: a mango seed gives rise only to a mango tree because of the unchanging essence of 'mango-ness' that is in the seed and which connects the seed and the tree. Thus *svabhāva* is a kind of underlying substratum of change, a quasi-substance.

The Sautrāntikas rejected this theory as incompatible with the Buddha's doctrine of 'no-self', for to say a thing arises from its 'own-nature' is just to say it arises from the self. Instead the Sautrāntikas held existence to be but a series of successive moments. A seed is but a series of such point-instants and the seed-series gives rise to the tree-series in the sense that the latter succeeds the former. Causality, then, is just contiguity or immediate succession. But what of the origin of the series themselves? The Sautrāntikas maintain that the seed-series, at one time nonexistent, comes into existence: that is, the effect does not pre-exist.

Hence the Sarvāstivādins, with their appeal to a quasi-substantial essence, end up with a causal theory that threatens to become a Buddhist version of the Sāōkhya identity theory. The Sautrāntikas, on the other hand, espouse a Buddhist version of the nonidentity theory which fails to provide for any kind of necessity in the causal relation. Either way the possibility of liberation is not guaranteed: the first account is too strong, the second too weak.

This situation provides the context for the Mahāyāna developments. The Yogācārin idealists give up the reality of the external object and join the drift away from epistemological realism (see Buddhism, Yogācāra school of §§1-4). But the Mādhyamikas take a different line, exemplified in the celebrated critique of causation by the second-century Buddhist philosopher Nāgārjuna in his *Mūlamadhyamakakārikā (Fundamental Verses on the Middle Way)* (see Buddhism, Mādhyamika: India and Tibet). Nāgārjuna refers to and rejects four types of causal theory: (i) self-causation; (ii) external causation; (iii) both (that is, self- and external causation); and (iv) noncausation. The first type of theory includes the Sāōkhya identity theory; it also includes the Sarvāstivādin theory. The second type includes the Nyāya nonidentity theory; it also includes the Sautrāntika theory. The third is the Jaina theory. The fourth is the Cārvāka theory.

In a virtuoso dialectical display Nāgārjuna argues that the first option is absurd since it supposes the production of what already exists. The second option is absurd because the cause cannot be totally extraneous to its effect, or anything might cause anything. The third option is also untenable, since it just combines the first two options. The fourth option is unacceptable because it implies randomness and the inefficacy of action.

The last claim makes it clear that Nāgārjuna does not deny causation *per se*. Rather causality is interdependence: that is, all things are on a par, dependent on one another. Accordingly everything is empty (*śūnya*) of an independent essence. But all the causal theories criticized understand causation as an asymmetrical dependence relation with one relatum self-existent and hence more real. Instead, the Buddha's teaching of dependent origination is that everything is interdependent, and this is equivalent to the truth of emptiness (*śūnyatā*), that nothing has any self-existence or essence. Liberation is the realization of this emptiness (see Buddhist concept of emptiness).

7 Jaina nonabsolutism

The Jainas also agree that everything is interdependent. However, they insist too that it is still possible to distinguish the more real from the less real. Jaina nonabsolutism (*anekāntavadā*) is the theory that everything in the world has various aspects which permit everything to be seen from various viewpoints (see Manifoldness, Jaina theory of). With respect to causation this means that cause and effect are partly identical and partly

nonidentical. A cause has a power (*śakti*) to produce an effect and from this viewpoint the effect is pre-existent in the cause. But the effect is a new substance *qua* its form and from this viewpoint the effect is not pre-existent in its cause. This explains both why a particular effect can only be produced from a particular cause and why an extra effort is necessary to bring about that effect. Thus a pot is pre-existent in the clay in so far as its matter is concerned, but not in so far as its shape is concerned. The potter's effort is required to shape the clay into a pot.

The Jaina view seeks, then, to combine the merits of both the identity and nonidentity theories, while avoiding the difficulties of each. The theory is also very close to the identity-in-difference (*bhedābheda*) theories of certain theistic Vedāntins. The major difficulty with the theory from its opponents' point of view is that it just doubles the trouble by trying to have things both ways. To the extent that the Jaina theory of causation is a version of the identity theory, the causal relation is too strong to guarantee the possibility of liberation; to the extent that it is a version of the nonidentity theory, the causal relation is too weak.

See also: Causation

ROY W. PERRETT

References and further reading

Āraṇya, Svāmī Hariharānanda (1963) *Yoga Philosophy of Patañjali*, Albany, NY: State University of New York Press, 1983.(Sanskrit texts and translations of the Yogasūtra and the Yogabhāsya.)

Bhaduri, S. (1947) *Studies in Nyāya-Vaiśeṣika Metaphysics*, Poona: Bhandarkar Oriental Research Institute, 2nd edn, 1975.(Chapter 12 reviews the Naiyāyika theory of causality.)

Bhartiya, M.C. (1973) *Causation in Indian Philosophy*, Ghaziabad: Vimal Prakashan. (Good comparative review with special reference to Nyāya-Vaiśeṣika.)

Garfield, J.L. (1995) *The Fundamental Wisdom of the Middle Way*, New York; Oxford University Press. (Translation of the Tibetan version of Nāgārjuna's *Mūlamadhyamakakārikā*, plus an extensive and valuable philosophical commentary thereon.)

Īśvarakṛṣṇa (*c.* AD 350-450) *Sāōkhyakarika (Verses on Distinctionism)*, in G.J. Larson, *Classical Sāṃkhya*, Delhi: Motilal Banarsidass, 1969; 2nd rev. edn, 1979. (Sanskrit text and a translation, in a good study of Sāōkhya.)

Kalupahana, D.J. (1975) *Causality: The Central Philosophy of Buddhism*, Honolulu, HI: University Press of Hawaii.(Good study of dependent origination with special reference to early Buddhism.)

Nāgārjuna (*c.* AD 150-200) *Mūlamadhyamakakārikā (Fundamental Verses on the Middle Way)*, trans. K. Inada, *Nāgārjuna: A Translation of his Mūlamadhyamakakārikā with an Introductory Essay*, Tokyo: Hokuseido Press, 1970. (Includes the Sanskrit text.)

Potter, K.H. (1963) *Presuppositions of India's Philosophies*, Englewood Cliffs, NJ: Prentice-Hall.(Chapters 6-9 present an excellent and original philosophical treatment of the Indian problem of freedom and causation.)

Potter, K.H. (ed.) (1970) *Encyclopedia of Indian Philosophies*, vol. 1, *Bibliography*, Delhi: Motilal Banarsidass, 2nd rev. edn, 1983.(Standard bibliographical source; includes details of Sanskrit texts, translations and secondary literature.)

Radhakrishnan, S. and Moore, C.A. (eds) (1957) *A Source Book in Indian Philosophy*, Princeton, NJ: Princeton University Press.(A very useful collection of translated primary sources; see the index under 'causality'.)

Smart, N. (1964) *Doctrine and Argument in Indian Philosophy*, London: Allen & Unwin.(Chapter 14 offers a useful synoptic review of some of the Indian argumentation about causation.)

Cavell, Stanley (1926-)

Born in Atlanta, Georgia, Stanley Cavell has held the Walter M. Cabot Chair in Aesthetics and the General Theory of Value at Harvard University since 1963. The range, diversity and distinctiveness of his writings are unparalleled in twentieth-century Anglo-American philosophy. As well as publishing essays on modernist painting and music, he has created a substantial body of work in film studies, literary theory and literary criticism; he has introduced new and fruitful ways of thinking about psychoanalysis and its relationship with philosophy; and his work on Heidegger and Derrida, taken together with his attempts to revitalize the tradition of Emersonian Transcendentalism, have defined new possibilities for a distinctively American contribution to philosophical culture. This complex oeuvre is unified by a set of thematic concerns - relating to scepticism and moral perfectionism - which are rooted in Cavell's commitment to the tradition of ordinary language philosophy, as represented in the work of J.L. Austin and Wittgenstein.

1 Criteria and scepticism

Cavell's Ph.D. dissertation (later published, much revised, as *The Claim of Reason* (1979)) elaborates an unorthodox interpretation of the notion of a criterion - pivotal for both J.L. Austin and Wittgenstein as that which is elicited by their method of responding to philosophical confusions with reminders of 'what we say when', of how we apply words in the circumstances of everyday life. Cavell's position resembles orthodoxy in conceiving of criteria as fixing a word's grammar by linking it to other words in a schematism that aligns language with the world and users of those words with one another. However, he takes the precise extent of this linguistic community as something philosophical exploration helps to determine, not something it can take for granted; and he takes the capacity exercised in such explorations to be a species of self-knowledge rather than knowledge of a predetermined body of rules, since a word's complex grammatical schematism sets real but inherently flexible limits to its projectibility into new contexts of use, and so requires an ineliminably personal evaluation of the mutuality of words and world in specific contexts.

Cavell further argues that Wittgensteinian criteria go deeper than Austinian ones, by showing that grammatical reminders cannot be used to refute scepticism in the manner favoured by Austin and many Wittgensteinians (see Criteria; Scepticism). Wittgensteinian criteria are criteria of identity rather than of existence; they specify what it is for something to count as a certain kind of thing, thus determining the applicability of a concept and making possible the judgments which deploy it, but they do not and could not determine the correctness of any such judgments. In this sense, sceptical denials that we can be certain of the existence of the external world or of other minds embody a truth; but they misrepresent this truth, since if criteria are not a species of knowledge-claim, they can no more be doubtful than they can be indubitable. Nevertheless, criteria so understood cannot be grounded in evidence or argument, but solely in our agreeing in their continued use; so the possibility of a sceptical refusal of that agreement is ineliminable. Since, however, our capacity to use words presupposes some such agreement, its sceptical refusal must result in emptiness - in the sceptics' saying something other than they take themselves to mean, or in saying nothing whatever. Accordingly, scepticism must be combatted not by claiming that the repudiation of criteria is impossible or irrational, but by demonstrating its true cost.

That cost turns out to be high. In so far as criteria distinguish phenomena from one another, their repudiation amounts to transforming the world into an undifferentiated plenum. And in so far as they give expression to human interests in phenomena, marking out the distinctions which matter to us and so our shared responses to the world, their refusal amounts to a denial that the world matters, that its phenomena are of interest. But why, then, might such costs seem worth bearing? How can otherwise competent speakers come to relinquish the mastery of grammar which governs their everyday intercourse when under the pressure to philosophize? Any answer partly depends on which criteria are subject to repudiation; the needs and interests served by affirming a fantasy of the essential privacy of language are not those served by an emotivist fantasy of morality. But at its most general, the sceptical impulse refuses criteria as such - the conditions of human knowledge and meaning; and Cavell's claim is that the sceptic in us all is motivated by the sense that such conditions are constraints, that the limits of sense (that which makes thought and knowledge possible) are in fact limitations (perspectival blinkers imposed on an otherwise unmediated knowledge of the world). In other words, criteria are rejected in the name of a fantasized perspective on reality which transcends all limits, a view from nowhere; scepticism is a refusal of the conditioned

nature of human knowledge, a denial of human finitude in the name of the unconditioned, the inhuman. But of course, nothing is more human than the desire to deny one's own humanity.

2 Literature, cinema and psychoanalysis

For Cavell, then, Wittgenstein's later philosophy aims at overcoming a variety of manifestations of scepticism in philosophy - scepticism understood not simply as an intellectual doctrine, but as the modern inflection of a perennial human impulse that finds expression in domains regarded as distinct from that of philosophy. By interpreting the repudiation of criteria as bringing about the death of the world and our interest in it, Cavell links Ordinary language philosophy with the concerns of Romanticism, in both its English and German forms - particularly the writings of Coleridge, Wordsworth and the Schlegels. He also interprets Shakespeare's plays and certain genres of Hollywood movie as responding to sceptical struggles, by following out the fates of human beings caught up in relationships in which scepticism about other minds is lived out and presented as exemplary of scepticism in general (see Other minds). The human capacity to revive self and world by recovering an interest in the latter's autonomous yet environing life is there figured by the capacity of these couples to acknowledge one another as separate and yet related; the vicissitudes of their weddedness to one another symbolize the vicissitudes of human weddedness to the world.

For Cavell, these readings are not literary illustrations of independently derived philosophical theses. Their focus on sceptical themes is rather understood to unsettle any received wisdom about divisions between philosophy and what lies outside it; and they are often Cavell's primary motivation for revising his more obviously philosophical investigations of these issues. For example, he takes his reading of *The Winter's Tale* to demonstrate that scepticism is inflected by gender; and the same reciprocity can be seen in his conception of the relationship between philosophy and psychoanalysis. Interpreting scepticism as a tragic repudiation of the humanity in oneself and others leads him to see Freud's attempts to recover human meaning from aberrant behaviour as a mode of overcoming scepticism, and to view the general psychoanalytic commitment to the reality of the unconscious mind as itself responding to Descartes' failure to refute scepticism by staking the mind's reality upon its existence as consciousness (see Freud; Descartes). At the same time, he utilizes Freudian theory to diagnose the motivational sources of confusions in philosophy; and he uses Freud's fundamental but politically fraught alignment of masculinity with activity and femininity with passivity (understood as two dimensions of, or orientations towards, human experience) to argue that the essential passivity invoked in the Wittgensteinian practice of recalling the true significance of one's words (as well as in Heidegger's emphasis upon *Gelassenheit*) embodies a specifically feminine mode of philosophical thought and expression.

3 Perfectionism and modernism

Cavell's vision of the human mind as torn between its active and passive sides, between scepticism and its overcoming, is further underpinned by reference to Emerson and Thoreau - and specifically by their perfectionist conception of the self as ineluctably split or doubled, as always capable of moving beyond the state in which it finds itself. When that capacity is active, the self embarks on an endless process of self-development, with each attained state neighbouring an attainable state that forms its possible future; when it is in eclipse, perhaps because of the attractions of one's attained state or by personal and social distractions from the draw of one's unattained state, the self's capacity to grow is also eclipsed - and losing one's capacity to change oneself in the name of a better state of self and society means losing an essential aspect of the self's autonomy: the capacity to revise one's conception of the good. As this formulation suggests, Cavell concludes that Emersonian perfectionism should form an essential dimension of Rawlsian (and more generally, of Kantian) liberal democracy, since a non-autonomous self cannot internalize the moral law which should govern relations with others in a liberal society (see Liberalism).

Cavell's work is a species of philosophical modernism. His writings relate themselves to a number of intellectual and cultural traditions by regarding their continuation as an undismissible problem. He can neither accept their prevailing paradigms as they stand, nor reject them as no longer philosophically and humanly meaningful; so he aims to inherit them by subjecting them to a radical but internal critique that can ultimately be grounded only upon his own (and his readers') willingness to acknowledge his words as worthy continuations of those traditions. This means that his writings are bound to appear idiosyncratic and self-regarding; but it also means that that appearance can no more justify their dismissal than it can guarantee their value.

STEPHEN MULHALL

List of works

Cavell, S. (1969) *Must We Mean What We Say? A Book of Essays*, New York: Charles Scribner's Sons.(A set of interrelated essays on philosophical method and the arts, including a long and hugely influential essay on *King Lear*.)

Cavell, S. (1971) *The World Viewed: Reflections on the Ontology of Film*, New York: Viking Press.(This analysis draws on Cavell's earlier work on theatre and painting; an expanded edition, published in 1979, includes replies to critics.)

Cavell, S. (1972) *The Senses of Walden*, New York: Viking Press.(A highly compressed, detailed reading of Thoreau's famous text.)

Cavell, S. (1979) *The Claim of Reason: Wittgenstein, Skepticism, Morality and Tragedy*, Oxford: Clarendon Press. (Cavell's much-revised dissertation, this contains his most detailed elaboration of his interpretation of criteria, and sets the scene for all his future writing.)

Cavell, S. (1981) *Pursuits of Happiness: The Hollywood Comedy of Remarriage*, Cambridge, MA: Harvard University Press.(These six essays identify a genre of screwball comedies from Hollywood's golden age, in which issues of interpersonal acknowledgement are worked out in relation to Romanticism and scepticism.)

Cavell, S. (1984) *Themes Out of School: Effects and Causes*, San Francisco, CA: North Point Press.(A disparate collection, including an important early essay on existentialism and analytical philosophy.)

Cavell, S. (1987) *Disowning Knowledge: In Six Plays of Shakespeare*, Cambridge: Cambridge University Press. (This collects a number of already-published pieces, as well as two new ones.)

Cavell, S. (1988) *In Quest of the Ordinary: Lines of Skepticism and Romanticism*, Chicago, IL: University of Chicago Press.(The core of this collection is a sequence of lectures on American and English Romanticism in literature and philosophy.)

Cavell, S. (1989) *This New Yet Unapproachable America: Lectures after Emerson after Wittgenstein*, Albequerque, NM: Living Batch Press, and Chicago, IL: University of Chicago Press.(Contains two lectures, including a long piece on Wittgenstein as a philosopher of culture.)

Cavell, S. (1990) *Conditions Handsome and Unhandsome: The Constitution of Emersonian Perfectionism*, Chicago, IL: University of Chicago Press.(Cavell here relates his perfectionist concerns to recent work byRawls and Kripke.)

Cavell, S. (1994) *A Pitch Of Philosophy: Autobiographical Exercises*, Cambridge, MA: Harvard University Press. (As well as explicitly autobiographical reflections, this collection contains important essays on Derrida and J.L. Austin, and on opera.)

Cavell, S. (1995) *Philosophical Passages: Wittgenstein, Emerson, Austin, Derrida*, Oxford: Blackwell.(Includes a much-revised version of Cavell's lectures on the opening of the*Philosophical Investigations*.)

Cavell, S. (1996) *Contesting Tears: The Hollywood Melodrama of the Unknown Woman*, Chicago, IL, and London: University of Chicago Press.(This collection identifies a genre of film adjacent to that of the remarriage comedies studied in *Pursuits of Happiness*, and systematically relates cinema to psychoanalysis.)

References and further reading

Cohen, T., Guyer, P. and Putnam, H. (eds) (1993) *Pursuits of Reason: Essays in Honor of Stanley Cavell*, Lubbock, TX: Texas Tech University Press.(*Festschrift*, with some useful critical essays.)

Fischer, M. (1989) *Stanley Cavell and Literary Skepticism*, Chicago, IL: University of Chicago Press.(Applies Cavell's interpretation of scepticism to post-structuralist literary theory.)

Fleming, R. and Payne, M. (eds) (1989) *The Senses of Stanley Cavell*, Lewisburg, PA: Bucknell University Press. (Useful collection of critical essays, with two interviews with Cavell.)

Mulhall, S. (1994) *Stanley Cavell: Philosophy's Recounting of the Ordinary*, Oxford: Oxford University Press. (The first full-length general philosophical study of Cavell's work.)

Mulhall, S. (ed.) (1996) *The Cavell Reader*, Oxford: Blackwell.(Contains selections from every phase of Cavell's career, with a detailed bibliography, and a general introductory survey by the editor.)

Cavendish, Margaret Lucas (1623-73)

The only seventeenth-century woman to publish numerous books on natural philosophy, Cavendish presented her materialism in a wide range of literary forms. She abandoned her early commitment to Epicurean atomism and, rejecting the mechanical model of natural change, embraced an organicist materialism. She also addressed the relations that hold among philosophy, gender and literary genre.

1 Philosophical influences

Margaret Cavendish, Duchess of Newcastle, made contact with Hobbes, tutor to the Cavendish family, during the English Civil War. She became a member of the 'Newcastle Circle', which included Hobbes, Charleton and Digby, and which was influenced by interaction with Mersenne and Gassendi. While exiled in Paris, Rotterdam and Antwerp, she met Descartes and Roberval. In 1667, she became the first woman to attend a session of the Royal Society of London. She corresponded with Christian Huygens about 'Rupert's exploding drops', and with Glanvill about witchcraft and Neoplatonic notions such as 'plastic faculties' and the 'soul of the world'. She was one of the first Englishwomen to gain recognition for her publications. In additional to writing treatises such as *Philosophical Fancies* (1653) and *Philosophical and Physical Opinions* (1655, 1663) she experimented with a wide range of genres to express her views: poetry, orations, plays, autobiobraphy (*Nature's Pictures…*, 1656), biography (*The Life of ... William Cavindishe…*, 1667), allegories (*The World's Olio*, 1655), epistolary narrative (*CCXI Sociable Letters*, 1664a) and fiction (*The Description of a New World, called the Blazing-World*, 1668). Introduced to twentieth-century readers by Henry Ten Eyck Perry (1918) and Virginia Woolf (1925), her philosophy only came in for serious, sustained evaluation beginning in the 1980s.

2 Organicist materialism

Cavendish's earliest works, such as *Poems, and Fancies* (1653), combine a version of Epicurean atomism with Lucretius' strategy of using poetry to achieve the highest form of the good: intellectual pleasure (see Epicureanism §§3-5). But it is her *Philosophical Letters* (1664b), *Observations upon Experimental Philosophy* (1666), and *Grounds of Natural Philosophy* (1668) that present her mature natural philosophy. She attacks the Platonists' and hermeticists' postulation of incorporeals in their theory of natural change. She agrees with Hobbes that an incorporeal substance is inconceivable for 'natural reason', but concedes that, on the basis of faith, one might accept 'supernatural spiritual beings'.

Like Hobbes, she holds that all natural change is change of motion. But instead of mechanism, the model for change in her mature philosophy is one of vital agreement or sympathetic influence of parts, as within a single organism. For, she rejects Hobbes' principle: every motion in a body is produced by a contiguous body through impact. Rather, motion and vitality cannot be transferred from an external source, but are inherent in corporeal body. Cavendish rejects the transfer model of causation, which she thinks must underlie mechanical philosophy, for two main reasons. First, if motion is a mode of body, then it cannot be transferred outside of the substance in which it inheres. This would give to motion the unacceptable status of a 'real quality': a mere modification of substance that is none the less treated as if a 'complete thing'. Second, since motion is naturally inseparable from material body, if motion could be transferred, then a portion of body would be transferred too. But, since all change reduces to changes in motion, it follows that corporeal body would quickly be diminished. Cavendish here adapts a standard argument against atomism, cited by Lucretius.

Corporeal individuals, that is to say parts of body, can act on others at a distance because of their sympathy and vital agreement, just as a seriously wounded appendage can affect change in the organism as a whole. Cavendish is even committed to a version of panpsychism, whereby all corporeal individuals contain some degree of sense and reason (see Panpsychism). (In *De Corpore*, Hobbes admitted that he could not refute those who would ascribe sense to inanimate bodies.) She holds that corporeal nature is endowed with something analogous to understanding: Bodies know how to 'pattern out' the figure of a distant object in perception; they know how to duplicate themselves in generation. Her views - (1) that nature is a single, unified corporeal body, intrinsically possessing self-motion or vital force, (2) that all parts of corporeal nature have some degree of sense and intellect, (3) that causation is understood through the vital affinity one part of nature has for another, rather than via the mechanical model - are all adaptations from Stoicism. Cavendish appears to have been familiar with some version

of Chrysippus' views, for she makes reference to some of his analogies and arguments (see Stoicism §3; Chrysippus).

3 Theory of sense perception

Cavendish characterizes her anti-mechanist model of change in the following way:

> A Watch-maker doth not give the watch its motion, but he is onely the occasion, that the watch moves after that manner, for the motion of the watch is the watches own motion, inherent in those parts ever since that matter was…. Wherefore one body may occasion another body to move so and so, but not give it any motion, but everybody (though occasioned by another, to move in such a way) moves by its own natural motion.
>
> (Cavendish 1664b: 100)

She uses this vital concomitance model in her theory of sense perception: the 'corporeal motions' of external objects are the 'occasion' for the 'sensitive and rational motions' in creatures to imitate or 'pattern out' the motions and figures of the external objects. An 'occasion' is any circumstance which has no intrinsic connection to or direct influence on the effect, is not necessary for the production of the effect, but has an indirect influence on the production of the effect by inducing the primary cause through command or example. For Cavendish, a 'primary cause' is that which is necessary and sufficient for the production of the effect. An external object cannot be *necessary* for perception since 'the sensitive organs can make such like figurative actions were there no object present' (1668b: 56). But neither is it *sufficient*; when the sense organs operate 'irregularly', misperception of objects takes place. So, the sensitive body is the principal cause of perceptions, and external objects are the exemplar causes that induce the body to pattern out one sense perception rather than another.

Cavendish's main argument against Descartes and Hobbes on perception is that their talk of translation of motion, or of imprinting an image, can only be interpreted via a transfer model of causation. But, as has been shown, such a model is inconsistent with their substance/mode ontology. Cavendish also has specific criticisms of Hobbes' theory of memory that derive from Stoic arguments. If sense is produced by imprinting and if, as Hobbes suggests, the bodily organs retain these imprints, then mechanism entails an eventual jumble of corporeal images. This criticism, derived from Chrysippus, was also proffered by Digby and Glanvill. Hobbes would have charged his critics with interpreting his metaphors too simplistically, since the impressions he refers to are infinitesimal motions. But without the mathematical model later provided by the calculus, the only models in terms of which to depict the impact of infinitesimally small motions were crude mechanical ones. So, however quaint these criticisms may seem to us, they were contemporary demands for an intelligible mechanical model of perception.

4 Philosophy and gender

In 'Female Orations', a section of *Orations of Divers Sorts* (1662), Cavendish explores the following questions: Is woman's social subordination to man a fact or no? If the former, is it due to oppression by men, to a natural inferiority in women, or to unfair cultural practices such as inferior education and constricted possibilities for acquiring moral, physical and intellectual experience? Finally, is woman's subordination to man inevitable? Rather than arguing for a position, Cavendish is content to lay out the logical space of the gender politics of her day. The prefaces and introductions to her books also contain defences of her desire to publish and her ability, as a woman, to do philosophy. Her experimentation with varied philosophical genres is, in part, an attempt to find a form which her readers would find acceptable: one in which a seventeenth-century woman, explicitly seeking fame as a natural philosopher, might give voice to her views.

See also: Atomism, ancient; Materialism

EILEEN O'NEILL

List of works

Cavendish, M. (1653) *Poems, and Fancies*, London; facsimile repr. Menston: Scolar Press, 1972; *Poems, and Phancies*, The Second Impression, much Altered and Corrected, London, 1664; *Poems, or Several Fancies in Verse: with the Animal Parliament, in Prose*, The Third Edition, London, 1668.(Atomism expressed in verse form, together with poems on moral topics such as cruelty to animals, and whimsical verses about fairies and other incorporeals.)

Cavendish, M. (1653) *Philosophicall Fancies*, London.(A prose exposition of her materialist natural philosophy.)

Cavendish, M. (1655) *Philosophical and Physical Opinions*, London; 2nd edn, London, 1663.(This revised version of *Philosophicall Fancies* is a transitional work, in which Cavendish begins to reject mechanistic atomism in favour of an organicist materialism; the substanially revised second edition, reissued as *Grounds of Natural Philosophy* (1668), expounds her mature system of natural philosophy.)

Cavendish, M. (1655) *The World's Olio*, London; The Second Edition, London, 1671.(A collection of mostly prose works, it also contains some poems on atomism first published as Part I of *Poems, and Fancies*.)

Cavendish, M. (1655/6) *Nature's Pictures Drawn by Fancie's Pencil to the Life*, London; 2nd edn, *Natures Picture Drawn by Fancies Pencil to the Life*, London, 1671.(The first edition of this collection of largely fictional sketches includes her autobiography, 'A True Relation of my Birth, Breeding and Life', which is not found in the second edition.)

Cavendish, M. (1662) *Playes*, London.(Some plays include discussions about gender politics, such as *The Female Academy* and *Bell in Campo*; excerpts from the latter play appear in *Women Writers of the Seventeenth Century*, ed. K. Wilson and F. Warnke, Athens, GA: University of Georgia Press, 1989.)

Cavendish, M. (1662/3) *Orations of Divers Sorts, Accomodated to Divers Places*, London; 2nd edn, London, 1668.(Includes 'The Female Orations' reprinted in *The Norton Anthology of Literature by Women*, ed. S. Gilbert and S. Gubar, New York and London: W.W. Norton & Co., 1985.)

Cavendish, M. (1664a) *CCXI Sociable Letters*, London; facsimile repr. Menston: Scholar Press, 1969.(An epistolary narrative praised by Charles Lamb, which has been compared to Samuel Richardson's works, it provides details of the historical and intellectual setting in which Cavendish was writing.)

Cavendish, M. (1664b) *Philosophical Letters: or, Modest Reflections Upon some Opinions in Natural Philosophy, maintained By several Famous and Learned Authors of this Age, Expressed by way of Letters: By the thrice Noble, Illustrious, and Excellent Princess, the Lady Marchioness of Newcastle*, London.(Letters to a fictious noblewoman, in which she criticizes the views of Descartes, Hobbes, More and van Helmont in light of her own system of natural philosophy; letters concerning Descartes are reprinted in *Women Philosophers of the Early Modern Period*, ed. M. Atherton, Indiannapolis, IN: Hackett Publishing Company, 1994.)

Cavendish, M. (1666) *Observations upon Experimental Philosophy, To which is added, The Description of a New Blazing World*, London; the additional text was reissued as *The Description of a New World, called the Blazing-World*, London, 1668; repr. in *The Blazing World and Other Writings*, ed. K. Lilley, London and New York: Penguin, 1992.(A treatise of her natural philosophy, published with a utopian fiction that intimates some of her philosophical positions and parodies views about incorporeal spirits.)

Cavendish, M. (1667) *The Life of the thrice Noble, High and Puissant Prince William Cavendishe, Duke, Marquess, and Earl of Newcastle…*, London; 2nd edn, 1675; Latin trans. W. Charleton as *De vita et Rebus Gestis…Guilielmi Ducis Novo-Castrensis…*, London, 1668; numerous reprints.(This early English biography has gained Cavendish a place in literary history; she has been known primarily for this work.)

Cavendish, M. (1668a) *Plays, never before Printed*, London.(Some plays include discussions about gender politics, such as *The Convent of Pleasure*, excerpts from which appear in *First Feminists: British Women Writers 1578-1799*, ed. M. Ferguson, Bloomington, IN: Indiana University Press, 1985.)

Cavendish, M. (1668b) *Grounds of Natural Philosophy… The Second Edition, much altered from the First which went under the Name of Philosophical and Physical Opinions*, London.(A reissue of the 1663 work; together with *Philosophical Letters* and *Observations upon Experimental Philosophy*, this work constitutes the expressions of her mature natural philosophy.)

References and further reading

A Collection of Letters and Poems: Written by several Persons of Honour and Learning, Upon divers Important Subjects, to the Late Duke and Duchess of Newcastle, (1678), London.(Documents important for reconstructing her intellectual circle, including letters to her by Hobbes, Glanvill, Charleton and Digby.)

Alic, M. (1986) *Hypatia's Heritage: A History of Women in Science from Antiquity through the Nineteenth Century*, Boston: Beacon Press.(Historical information about her natural philosophy.)

Ballard, G. (1752) *Memoirs of several Ladies of Great Britain…*, Oxford; repr. ed. R. Perry, Detroit: Wayne State University Press, 1985.(Biographical information.)

Bowerbank, S. (1984) 'The Spider's Delight: Margaret Cavendish and the "Female Imagination"', *English Literary Renaissance* 14; repr. in *Women in the Renaissance: Selections from English Literary Renaissance*, ed. K. Farrell, E.H. Hageman and A.F. Kinney, Amherst, MA: University of Massachusetts Press, 1991.(A feminist examination of Cavendish's natural philosophy.)

Cavendish, W. (ed.) (1676) *Letters and Poems in Honour of the Incomparable Princess Margaret, Dutchess of Newcastle*, London.(Documents important for reconstructing her intellectual circle.)

Clucas, S. (1994) ‘The Atomism of the Cavendish Circle: A Reappraisal’, *The Seventeenth Century* 9 (2): 247-73. (Chronicles Cavendish’s changing views in natural philosophy, arguing that she never entirely rejected atomism, broadly construed.)

Dumée, J. (*c*.17th century) *Entretien sur l’opinion de Copernic touchant la mobilité de la terre (Discussion of the Opinion of Copernicus Concerning the Mobility of the Earth)*, Paris: La Veuve Bertrand, ms. Bibliothèque National Fonds français 1941.(One of the only other works on natural philosophy written by a seventeenth century woman.)

Grant, D. (1957) *Margaret the First: A Biography of Margaret Cavendish, Duchess of Newcastle, 1623-1673*, London: R. Hart-Davis.(A standard source for biographical information.)

Jones, K. (1988) *A Glorious Fame: The Life of Margaret Cavendish, Duchess of Newcastle, 1623-1673*, London: Bloomsbury.(A biography which places Cavendish within the context of other women intellectuals of the period.)

Kargon, R.H. (1966) *Atomism in England from Hariot to Newton*, Oxford: Clarendon Press.(One of the few histories to treat the ‘Newcastle Circle’; only Cavendish’s earliest publications are discussed.)

Merchant, C. (1980) *The Death of Nature*, San Francisco, CA: Harper & Row.(A feminist/environmentalist revisionist history of science, which treats her rejection of mechanism.)

Meyer, G.D. (1955) *The Scientific Lady in England 1650-1760*, Berkeley, CA: University of California Press. (Focuses on her *Observations upon Experimental Philosophy* and places it in the context of other early modern scientific works by Englishwomen.)

Rogers, J. (1996) *The Matter of Revolution: Science, Poetry, and Politics in the Age of Milton*, Ithaca, NY and London: Cornell University Press.(Comparable to earlier treatments of Hobbes, this work explores the connection between Cavendish’s natural philosophy and political views.)

Sarasohn, L.T. (1984) ‘A Science Turned Upside Down: Feminism and the Natural Philosophy of Margaret Cavendish’, *Huntington Library Quarterly* 47 (4): 299-307.(A treatment of Cavendish’s natural philosophy that draws conclusions about the feminist implications of her work.)

Schiebinger, L. (1991) ‘Margaret Cavendish, Duchess of Newcastle’, in *A History of Women Philosophers*, vol. 3, ed. M.E. Waithe, Dordrecht: Kluwer.(A survey of Cavendish’s natural philosophy.)

Smith, H. (1982) *Reason’s Disciples: Seventeenth-Century English Feminists*, Urbana: University of Illinois Press.(Focuses on the feminism in her non-scientific works.)

Ten Eyck Perry, H. (1918) *The First Duchess of Newcastle and Her Husband As Figures in Literary History*, Boston, MA and London: Ginn & Company.(An examination of Cavendish’s corpus and a historically useful account of its critical reception.)

Woolf, V. (1925) ‘The Duchess of Newcastle’, *The Common Reader*, London: Published by Leonard & Virginia Woolf at the Hogarth Press.(An introduction of Cavendish to twentieth-century readers.)

Celsus (late 2nd century AD)

The Greek philosopher Celsus of Alexandria was a Middle Platonist, known only for his anti-Christian work The True Account. The work is lost, but we have Origen's reply to it, Against Celsus. In it Celsus defends a version of Platonist theology.

Celsus is known only as the author of a polemical work against the Christians entitled *Alēthēs logos*, which may be translated *The True Account*, although other connotations of *logos* are also present in the title (see Logos). We know of this work only through the reply (*Against Celsus*) composed to it in AD 248 by the Church Father Origen, who does not in fact know who Celsus is (see Origen §1). We too know nothing about Celsus, but can date his work fairly closely from references in it to a persecution of Christians under a joint rulership which must be that Marcus Aurelius and Commodus (AD 177-80).

The title of Celsus' book may also refer to a passage of Plato's *Meno* (81a), where Socrates speaks of the ancient doctrine he has heard concerning the immortality of the soul as a 'true account' (see Plato §11). One of Celsus' polemical points certainly is that Platonic philosophy is in accord with the wisdom of the most ancient authorities, such as Orpheus and Homer - a common enough view among the Neoplatonists. As regards his own philosophical position, it is hard to pin him to any particular tendency within contemporary Platonism. He shows the expected contempt for a notion like the resurrection of the body (Origen, Against Celsus V 14), or the idea that god made man in his own image (VI 63), there being nothing that could resemble god. At VII 42, he gives a basic account of the Platonist view of the supreme god, which is incompatible with the notion of his involving himself too closely with matter. He alludes, indeed, to the notion of the supreme god being 'beyond being' (Plato, *Republic* 509b), and the title of his work may indicate his adoption of the *logos* as a secondary god, but we cannot be sure. He mentions three ways of attaining a conception of god - synthesis, analysis and analogy, which correspond approximately to the three ways distinguished by Alcinous in his *Didaskalikos*, chapter 10, although there is no indication that Celsus knows that work. Contact between man and god is effected, of course, through the agency of daemons, whom the Christians are criticized for disdaining (VIII 28, 33, 35).

All this is very non-specific. *The True Account* is an interesting document, none the less, as a sign that Christians are coming to be of some weight in society, to the extent of meriting the sort of polemical refutation which hitherto the philosophical schools had directed against each other. Celsus' work is of the same genre as his contemporary Atticus' attack on the Peripatetics, Plutarch's attacks on the Epicureans and Stoics (see Plutarch of Chaeronea §1) and, later, Plotinus' treatise *Against the Gnostics* (Enneads II 9).

See also: Neoplatonism; Platonism, Early and Middle

JOHN DILLON

List of works

Celsus (*c.* 180 AD) *The True Account*, ed. O. Glöckner, *Celsus' Alēthēs logos*, Bonn: A. Marcus und E. Webers Verlag, 1924; trans. R.J. Hoffman, *Celsus, On the True Doctrine*, Oxford: Oxford University Press, 1995. (Celsus' lost text as reconstructed from extracts of Origen's work *Against Celsus*.)

References and further reading

Andresen, C. (1955) *Logos und Nomos: Die Polemik des Kelsos wider das Urchristentum (Logos and Nomos: Celsus' Polemic against Early Christianity)*, Berlin: de Gruyter. (Authoritative study of Celsus from a religious perspective.)

Origen (AD 248) *Against Celsus*, trans. H. Chadwick, *Contra Celsum*, Cambridge: Cambridge University Press, 1965.(Our sole source on Celsus.)

Certainty

'Certainty' is not a univocal term. It is predicated of people, and it is predicated of propositions. When certainty is predicated of a person, as in 'Sally is certain that she parked her car in lot 359', we are ascribing an attitude to Sally. We can say that a person, S, is psychologically certain *of a proposition, p, just in case S believes p without any doubts. In general, psychological certainty has not been a topic which philosophers have found problematic.*

On the other hand, certainty as a property of propositions, as in 'The proposition that Sally parked her car in lot 359 is certain for Sally', has been discussed widely by philosophers. Roughly, we can say that a proposition, p, is propositionally certain *for a person, S, just in case S is fully warranted in believing that p and there are no legitimate grounds whatsoever to doubt that p. The philosophical issue, of course, is whether there are any such propositions and, if so, what makes them certain.*

1 Types of certainty

Before directly considering certainty, it would be useful to distinguish it from another notion - that of infallibility. Although neither concept has been employed in a universally accepted way by philosophers, generally it has been held that a person, *S*, is infallible with regard to a proposition, *p*, just in case it is not possible for *S* falsely to believe that *p*. For example, it might not be possible for persons falsely to believe of themselves that they have beliefs. But the range of such contingent propositions, if any, is very small indeed; and philosophers have been interested in whether certainty extends to a much wider range of contingent propositions. In addition, a person could be infallible with regard to some propositions, even if those propositions are not at all warranted. For example, consider any necessary truth. It would not be possible for *S* mistakenly to believe it, since it is not possible for that proposition to be false. But it may have little or no warrant for *S*. Since propositional certainty is inextricably connected with warrant, infallibility and certainty should not be conflated (see Fallibilism).

So let us turn to a direct examination of certainty. Psychological certainty is an attitude that persons can have towards a proposition, and propositional certainty is a measure of the epistemic warrant for a proposition. Thus, it is clear that psychological certainty and propositional certainty are logically independent. For *S* could be certain that *p* on inadequate grounds or, in some extreme cases, on no grounds whatsoever. On the other hand, a person could have the best grounds for believing *p* and no basis for doubt, but none the less fail to believe *p* with a high degree of resolve because of timidity.

Although the two types of certainty are logically independent, there is a connection between them. Presumably a person would want the degree of belief in a proposition to parallel the degree of epistemic warrant for it. Descartes' *Meditations* (1641), one of the starting points of contemporary epistemology, can be seen as his attempt to determine which, if any, propositions are certain, in order to adjust his degree of belief accordingly (see Descartes, R. §4). If he is correct, very few propositions are certain (for example, 'I exist' and others revealed by the 'light of nature'). In particular, he argued that those based upon experience (for example, 'there is a table before me') are never certain because there is always some legitimate basis for doubt. Other philosophers, for example, G.E. Moore, argued that many propositions based upon experience can be certain (see Moore, G.E. §3; Commonsensism). The remainder of this entry discusses some of the more influential accounts of propositional certainty.

A contextualist account has been developed by both Moore and Wittgenstein (see Contextualism, epistemological). Although the specifics of the views vary, a contextualist account of propositional certainty claims, roughly, that the range of propositions that are fully warranted and beyond doubt or challenge is determined by the presuppositions of the context of discussion. As Aristotle remarked:

> [Some] people demand that a reason shall be given for everything, for they seek a starting point, and they seek to get this by demonstration, while it is obvious from their actions that they have no conviction. But their mistake is what we have stated it to be; they seek a reason for things for which no reason can be given; for the starting point of demonstration is not demonstration.
>
> (Aristotle *Metaphysics*: 1010a 8-14)

Wittgenstein (1949-51) and Moore (1959) have argued that the proposition, 'here's one hand', for example, is

certain in most contexts. It might not be certain if, for example, we knew that there are several fake hands resting on a table and we also know that there is a hole in the table through which people sometimes put their hands just to fool us. But in the normal circumstances, it is certain. (It should be noted that Wittgenstein and Moore differ about whether knowledge entails certainty. Wittgenstein held that 'knowledge' and 'certainty' belong to different categories because knowing requires justification whereas a proposition is certain only if it does not require a justification.)

Contextualism can be challenged because it blurs the distinction between psychological and propositional certainty, for the latter is characterized in a way that makes it roughly equivalent to group psychological certainty, thus eschewing the notion of epistemic warrant. For example, in Salem, Massachusetts in the 1600s, it would have been propositionally certain (given a contextualist account) that there were witches. But, it could be argued that although such a proposition was taken for granted, it was never fully warranted and was always subject to legitimate doubt. More generally, it could be claimed that propositions that are certain might not be, and typically are not, equivalent to the ones that are taken for granted.

A modification of the contextualist view, suggested by David Lewis (1979), can be employed to mitigate this objection. Let us grant the basic contextualist point that in order for genuine doubt to occur, there must be some propositions that 'stand fast' (to use Wittgenstein's expression) which can be employed to remove or substantiate the doubt. None the less, what stands fast at the beginning of a discussion (even if held with oneself) could become doubtful as the discussion progresses. Grounds for doubt can increase as knowledge expands or simply because new possibilities are envisaged. Suppose, for example, that one of the participants asks 'Are you really sure that is a hand? After all, does not the certainty of that proposition depend upon the belief that your senses are generally reliable? Can you be certain of that?' Now, of course, the other participants can refuse to rise to the philosophical bait. But, by introducing the general question of the reliability of the senses, one of the participants has exposed one of the contextual presuppositions, thus making it eligible for doubt. Lewis' modification of the contextualist view provides a partial response to the objection because the worthiness of belief in a proposition that once stood fast can be addressed by participants by altering the context.

Nevertheless, the fundamental challenge to contextualism remains unanswered: Why should the contextual presuppositions, even the shifting ones as described by Lewis, determine which propositions are actually worthy of belief? Indeed, are any propositions worthy of belief, even those which play the quasi-foundational role ascribed to contextual presuppositions? That is the Cartesian question.

2 Propositional certainty

Some philosophers (for example, Hegel and Dewey) reject the Cartesian question, holding that it incorrectly presupposes an 'independently existing' reality whose nature is the object of inquiry (see Idealism; Pragmatism). Dewey wrote:

> Knowing is, for philosophical theory, a case of specially directed activity instead of something isolated from practice. The quest for certainty by means of exact possession in mind of immutable reality is exchanged for the search for security by means of active control of the changing course of events. Intelligence in operation, another name for method, becomes the thing most worth winning.
>
> (Dewey [1929] 1960: 204)

Other philosophers would argue that although *some* properties of *some* objects are affected by our interactions with them during the process of inquiry, it does not follow that *all* the properties of the object are affected or even that *all* objects have some properties that are affected by inquiry. To use an example cited by Dewey, suppose that the location or velocity of an atomic particle is altered when it is observed. None the less, its property of being within 100 miles of the earth's surface would not be affected. In addition, presumably, some very large objects can be observed without affecting them at all (stars that are millions of light years away, for example). Indeed, many properties (for example, being a conifer) or objects (for example, trees) appear to be mind-independent. But more importantly, even if all objects and properties are somehow mind-dependent, the Cartesian question appears to remain: Is any claim about objects and properties, whatever they are, fully warranted and not subject to any legitimate doubt?

In order to answer this question, it is important to note as did Peter Unger (1975) that in the relevant sense, ‘certainty’ (like, for example, ‘perfect’) is an absolute term. It does not come in degrees. Although we do say that some propositions are more certain than others, it has been pointed out that such expressions are equivalent to the claim that some propositions are more nearly certain than others. Certainty, like perfection, does not come in degrees although we can approach it by degrees.

Roderick Chisholm’s (1977) account of certainty captures this absolute sense. He suggests that a proposition is certain just in case it is warranted and there is no proposition more warranted than it. None the less, although that account incorporates an important necessary condition of propositional certainty, it ignores two crucial features - namely that a proposition is certain only if it is fully warranted and it is beyond legitimate doubt. A proposition could satisfy the proposed condition and still be only partially warranted and subject to doubt, as long as no propositions are more warranted or subject to less doubt.

So, let us focus on what makes a proposition fully warranted and not the least bit doubtful. The first condition is rather easily defined. Given any account of warrant (for example, either an internalist or reliabilist account), we can say that a proposition is fully warranted just in case it is warranted and there is no incompatible proposition that is at all warranted (see Justification, epistemic §3).

Immunity to legitimate doubt is more difficult to characterize. None the less, two senses of doubt seem relevant here. We can say that a proposition, *p*, is *subjectively doubtful* for *S* if and only if there is some proposition which *S* is not warranted in denying which is such that it lowers the warrant of *p* (even the slightest degree). This is roughly equivalent to Descartes’ account of certainty employed in the *Meditations* and is subjective because whether a proposition is immune to this type of doubt depends upon *S*’s particular prior beliefs. They could be false. Indeed, a proposition which is subjectively certain could be false. So, subjective immunity from doubt cannot be the entire story.

A second sense of propositional certainty is relevant: objective certainty. We can say that a proposition is *objectively doubtful* for *S* if and only if there is some *true* proposition which if added to *S*’s beliefs lowers the warrant of *p* (even to the slightest degree). This is objective because *S*’s warrant for *p* will be diminished if some of that warrant depends upon a false belief of *S* or if there is some true proposition that counts against *p* for which *S* has no explanation (based only on true propositions within the corpus of warranted beliefs). Finally, if *p* were false, its negation added to *S*’s beliefs would undercut the warrant for *p*. Hence, propositions immune to objective doubt are true.

It is now possible to suggest an account of propositional certainty. We can say that a proposition, *p*, is certain for *S* if and only if it is fully warranted and is neither subjectively nor objectively doubtful. Philosophers do not agree about either the range of such propositions or about whether knowledge entails certainty. Rationalists, typically, hold that certainty is a necessary condition of knowledge and that (only) a priori propositions are certain. Some - but only some - empiricists will hold that certainty is not a necessary condition of knowledge. Finally, many sceptics have held that knowledge entails certainty, and since no proposition is beyond legitimate doubt, no proposition is known.

See also: A posteriori; A priori; Doubt; Empiricism; Internalism and externalism in epistemology; Knowledge concept of; Rational Beliefs; Rationalism; Scepticism

PETER D. KLEIN

References and further reading

Aristotle (*c*. mid 3rd century BC) *Metaphysics*, trans. W.D. Ross, in R. McKeon (ed.) *The Basic Works of Aristotle*, New York: Random House, 1941.(The account of foundational propositions is also developed in *Posterior Analytics*, translated by G.R.G. Mure in the McKeon collection.)

Chisholm, R. (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice-Hall, 2nd edn, 1977.(General introduction to contemporary epistemology. Later editions are equally good introductions.)

Descartes, R. (1641) *Meditations on First Philosophy*, in E.S. Haldane and G.R.T. Ross (eds) *The Philosophical Works of Descartes*, vol. 1, Mineola, NY: Dover Publications, 1955.(Standard English edition of one of the starting points of contemporary epistemology.)

Dewey, J. (1929) *The Quest for Certainty*, New York: Putnam, 1960.(Contains a defence of a pragmatic

epistemology.)
Firth, R. (1976) ‘The anatomy of certainty’, *Philosophical Review* 76: 3-27.(Treatment of various accounts of what is termed ‘objective’ certainty in this entry.)
Klein, P. (1981) *Certainty*, Minneapolis, MN: University of Minnesota Press. (Development of the nature of objective propositional certainty discussed in this entry.)
Klein, P. (1986) ‘Immune belief systems’, *Philosophical Topics* 14 (1): 259-80. (Development of the nature of subjective certainty discussed in this entry.)
Lewis, D. (1979) ‘Scorekeeping in a language game’, *Journal of Philosophical Logic* 8: 339-59.(Development of the modified contextualist account discussed in this entry.)
Markie, P. (1992) ‘The cogito and its importance’, in J. Cottingham (ed.) *The Cambridge Companion to Descartes*, Cambridge: Cambridge University Press.(Account of Descartes’ discussion of the conditions of what was termed ‘subjective certainty’ in this entry.)
Moore, G.E. (1959) *Philosophical Papers*, New York: Macmillan.(Contains several papers developing Moore’s account of certainty and his defence of common sense.)
Unger, P. (1975) *Ignorance: A Case for Scepticism*, Oxford: Oxford University Press.(Defence of scepticism based upon an analysis of psychological certainty, not propositional certainty, as is the more typical basis for scepticism.)
Wittgenstein, L. (1949-51) *On Certainty*, Oxford: Blackwell, 1969.(The development of the contextualist account of certainty discussed in this entry.)

Certeau, Michel de (1925-86)

*Michel de Certeau, a French philosopher trained in history and ethnography, was a peripatetic teacher in Europe, South America and North America. His thought has inflected four areas of philosophy. He studied how mysticism informs late-medieval epistemology and social practice. With the advent of the Scientific Revolution, the affinities the mystic shares with nature and the cosmos become, like religion itself, repressed or concealed. An adjunct discipline, heterology, thus constitutes an anthropology of alterity, studying the 'other' and the destiny of religion since the sixteenth century. De Certeau opens the hidden agendas that make representations of the past a function of social pressures, so that sometime histories are rearticulated in mirrored or subversive forms. This subversion makes accessible a general philosophy of invention that works within and against the strategic policies of official institutions. De Certeau's writings also belong to activism, the history of ideological structures, psychoanalysis, and post-1968 theories of writing (*écriture*) as defined by Deleuze, Derrida, Foucault and Lyotard.*

For de Certeau, mystics are those strange beings who claim to have left the material and sensory world and encountered an unnameable being, or an unearthly experience, but have returned with visible marks - scars or mutilations - that attest to the truth of what they say. They eagerly tell of their visions to others. They are creatures of passage who invent places in which they disappear, and whose words convey to their listeners the supraterrestrial qualities of the world and language. They liberate those whom they meet, inspiring them to cognizance of areas that cannot be controlled by rational or symbolic means. Mystics are frequently female, often rootless, sometimes grimy and occasionally without formal education. The mystic belongs to all religions but appears on the European horizon at the time major edifices of belief slowly crumble under the pressure of science, oceanic travel, schematic reason (Ramus, Descartes) and pre-capitalist economy.

When the substantive *mysticism* is coined in the early seventeenth century, the behaviour and philosophy become the subject of scientific inquiry, thus relegating a formerly accepted, but almost unnameable condition to the margins of civilization. Mystical activity returns, however, in subjective areas of everyday life, has roots in politics (for example, ecology and feminism), and is a principal concern in Freudian and Lacanian psychoanalysis. Included in de Certeau's studies are close readings of Jean-Joseph Surin, Nicolas of Cusa, Hieronymus Bosch and the wily, nomadic Freud of *Moses and Monotheism*.

De Certeau's contributions to the philosophy of history importantly concern the silencing of mystical language since the French classical age. Since the advent of print culture official historians have laboured to produce monuments reflecting the structures of power that sustain them. For de Certeau, the 'historiographical operation' constitutes the laws governing the creation of these pantheons of 'truth'. But the conceptual/social order of this historiography is also contested - by those (for example, Machiavelli) who tamper with its codes or who make visible its ideology.

Heterology, the second of de Certeau's philosophical inflections, is his name for the 'science' of otherness or alterity. Its topic is whatever resists being named, classified, or organized in a body of knowledge, often what generates or inspires it. This resistant other takes the form of those nocturnal musings, beings, impressions, dreams or epiphanic flashes that fascinate but cause consternation to the diurnal being. Frequently, the *other* who irrupts into the familiar world of sameness is the Indian (such as the Tupinamba in Jean de Léry's *Voyage to Brazil* of 1578), the wanderer (the footprint of Friday on Crusoe's island, or Panurge who disrupts the princely realm of Rabelais's *Pantagruel*), or the savage, the peasant isolated from urban development (those whom the Abbé Grégoire encounters in rural France in his 1791 study of the provinces). They can wear the garb of a literary double (Jekyll's Hyde) but may also resemble the *intercessor* (as Gilles Deleuze and Félix Guattari define the term in *What is Philosophy?*). The welcome extended to and cultivation of the other are vital, argues de Certeau, for the grounding displacement and deracination that mobilize inquiry, belief, affect and knowledge.

Invention, from its Latin sense (*invenire*, to happen upon, to choose) describes a vital area of de Certeau's philosophy of everyday life. Since capitalism has mapped out the world and, through mass media, colonized the subjective imagination, every individual becomes responsible for creating practices that do not serve the economy, but that give definition to autonomous and unofficial ways of living. Invention entails conceiving of *other* spaces in which the individual subject cannot be plotted, located or, as have been most third world natives, bartered, bled

or tortured. Based in part on Lévi-Strauss's concept of *bricolage*, which is the indigenous 'science of concrete activity' that fashions new meanings from fragments of myth and quotidian objects, the art of invention includes practices of walking (a performative 'space act' analogous to J.L. Austin's ' speech act'; see Speech acts), of making different things from a gamut of inherited forms (cooking, playing), and of opening different mental spaces (via thinking or reading) in areas that otherwise lie under strict ideological control. In this way the anonymous subject, no matter what their origins, can philosophize and act so as to live within but also to change the tenor of inherited, often limiting, social conditions.

De Certeau's non-philosophical writings are no less varied than his philosophy. They range from dense and erudite historical studies (including critical editions) to collective and polemical tracts aimed at changing educational philosophy and the tradition of subjectivity. He ranks among the most important figures who pragmatized critical theory and the human sciences in France after the intellectual revolution of May 1968.

See also: Alterity and identity, postmodern theories of; Deleuze, G.; Derrida, J.; Foucault, M.; Lacan, J.; Lyotard, J.F.; Mysticism, history of; Mysticism, nature of

TOM CONLEY

List of works

Certeau, M. de with Julia, D. and Revel, J. (1975) *Une politique de la langue: la révolution française et les patois*, Paris: Gallimard. (Groundbreaking study that blends history, linguistics and ethnography of the 1789 era.)

Certeau, M. de (1980) *La Possession de Loudun*, Paris: Gallimard; trans. M. Smith, *The Possession of Loudun*, Chicago, IL: University of Chicago Press, 1997. (Studies how possessed women in witchcraft trials invented discursive space through language-play.)

Certeau, M. de with Giard, L. and Mayol, P. (1980) *L'Invention du quotidien*, vol. 2, *Habiter, cuisiner*, Paris: Union Générale d'Éditions; trans. T. Tomasik *The Practice of Everyday Life*, vol. 2, *Living, Cooking*, Minneapolis, MN: University of Minnesota Press, 1998. (Mobilizes the concept of everyday practice by analyzing space as it is lived in urban Paris and elsewhere.)

Certeau, M. de (1982a) *L'Écriture de l'histoire*, Paris: Gallimard; trans. with introduction T. Conley, *The Writing of History*, New York: Columbia University Press, 1988 and 1992. (Most synthetic of de Certeau's work in general; includes chapter on possessed women of Loudun.)

Certeau, M. de (1982b) *La Fable mystique*, Paris: Gallimard; trans .M. Smith *The Mystic Fable*, Chicago, IL: University of Chicago Press, 1993. (Comprehensive study of mysticism and heterology.)

Certeau, M. de (1987a) *Histoire et psychanalyse entre science et fiction*, ed. L. Giard, Paris: Gallimard; trans B. Massumi, introduction W. Godzich, *Heterologies*, Minneapolis, MN: University of Minnesota Press, 1986. (English translation contains an essay on South American Indian political federations.)

Certeau, M. de (1987b) *La Faiblesse de croire*, ed. L. Giard, Paris: Éditions du Seuil. (Argues that unstressed belief has political and ecological potential in an age of multinational capitalism.)

Certeau, M. de (1990) *L'Invention du quotidien*, vol. 1, *Arts de faire*, ed. L. Giard, Paris: Gallimard; trans. S. Rendall, *The Practice of Everyday Life*, Berkeley, CA: University of California Press, 1984. (First introduction to the philosophy in English; the translation is of the first edition, published in 1980.)

Certeau, M. de (1993 and 1994) *La Culture au pluriel*, and *La Prise de parole*, ed. L. Giard, Paris: Éditions du Seuil; trans. with introduction T. Conley, *Political Writings*, Minneapolis, MN: University of Minnesota Press, 1996. (Mobilizes the philosophy in terms of issues surrounding politics and cultural reform after 1968.)

References and further reading

Ahearne, J. (1995) *Michel de Certeau*, London: Polity Press. (Terse and informative study for anglophone readers.)

Conley, T. and Terdiman, R. (eds) (1992) 'Michel de Certeau', *Diacritics* 22 (1). (Includes translation of 'Mysticism' by de Certeau and articles by French and American scholars.)

Giard, L. (ed.) (1987) *Michel de Certeau*, Paris: Centre Georges Pompidou. (Incisive essays by Derrida, Rabinow, Terdiman, *et al*.)

Giard, L. (ed.) (1988) *Le Voyage mystique: Michel de Certeau*, Paris: Éditions du Cerf. (Contains complete bibliography of writings compiled by L. Giard and essays by students and followers.)

Giard, L. (ed.) (1991) *Histoire, mystique, et politique: Michel de Certeau*, Grenoble: Jérôme Millon. (Offers perspective on work by French philosophers and historians.)

Chaadaev, Pëtr Iakovlevich (1794-1856)

Pëtr Chaadaev was the first Russian thinker for whom his own country became a philosophical problem. His works initiated the powerful Russian tradition of reflecting on Russia's whence and whither: that is to say, the meaning of Russian history, the character of Russian national identity, and the possible, or necessary, paths of Russian historical development in the future. However, Chaadaev's answer to these questions was mostly negative: he defined Russia not by what it was, but by what it was not.

A paradoxical feature of Chaadaev's s position was that his general philosophical views did not apply to his native country. He was a convinced Westernizer, identifying Western development with universal human history, but Russia was in his view the opposite of the West, an exception to the general rules. His general social philosophy, deeply influenced by the French theocratic traditionalists, was inherently conservative, stressing the importance of supra-individual unity and of continuous historical traditions; in contrast with this, his philosophy of Russian history defined Russia as a country without unity and without history, thus lacking the basic conditions for a genuine conservatism. This view provoked a strong reaction among Russian Romantic conservatives: they accepted some aspects of Chaadaev's conservative critique of atomistic individualism but tried to refute his pessimistic view of Russia, by arguing that, in fact, not Russia but the West represented atomistic disintegration and incapacity for organic development.

1 Metaphysics and philosophy of man

In his youth Chaadaev, a promising Russian officer and a friend of Pushkin, was a Freemason (from 1814) and a sympathizer of the Decembrist movement, named after the ill-fated military uprising which took place in December 1825. During his travels in the West in 1822-6 he intensely studied the counter-Enlightenment trends in European thought and, as a result, underwent a spiritual crisis which confirmed him in his leanings towards Roman Catholicism. The failure of the Decembrist uprising he saw as an additional proof of the bankruptcy of the rationalistic belief in progress through political change. On his return to Russia in 1826 he withdrew into almost complete seclusion and devoted himself entirely to the task of formulating his philosophical view of the world. He fulfilled this task in the eight 'Lettres Philosophiques' (Philosophical Letters) (written in French between 1828 and 1831), of which only the first, devoted to Russia, was published during his lifetime, in the journal *Teleskop (Telescope)* in 1836.

Following the French traditionalists - de Maistre, de Bonald and the early Lamennais - Chaadaev attacked in his Letters the moral and intellectual autonomy of the individual. The natural order of things, he argued, is based upon dependence and submissiveness. Moral law, like truth, is not something autonomous, as Kant claimed, but a force outside us; individual reason separates man from the universe and makes true understanding impossible; hence it should be subordinated to the universal reason of humankind. Being is hierarchically stratified: at the summit is the transcendent God; his emanation is the 'world consciousness', that is, supra-individual social consciousness, living in tradition and developing with it; below is the empirical consciousness of isolated individuals; on the lowest rung is pre-human nature. In this way Chaadaev combined the traditional theistic conception of a transcendent God with pantheistic emphasis on God's immanent presence in the world. This was in tune with Christian Neoplatonism (which reached Chaadaev through the esoteric tradition in Freemasonry, as well as through Schelling, whom he personally met in 1825) and with the panentheistic religious ideas of the German Romantics.

Chaadaev's conception of the 'social sphere' and supra-individual social consciousness provides the key to his philosophy of history. Knowledge, he argued, is a form of collective consciousness. Without society, that is, the supra-individual sphere which allows traditions to be handed down, human beings would never have emerged from the animal state. Also in religious experience the social sphere is of decisive importance: through it alone can the individual come to know God and to become a vessel for the divine truth. Therefore the surest way to God leads not through individualistic self-perfection or solitary ascetism but through strict observance of traditional norms and active participation in social life. The highest aim of man is 'the annihilation of his personal being and the substitution for it of a perfectly social or impersonal being'. The foundation of the inner unity of society is religion, whose necessary guardian is the institutionalized Church. Ecclesiastical mediation between man and God is indispensable for salvation, since efforts to achieve unmediated, individualized contact with God weaken the

discipline of the soul and bring about social disintegration.

In his philosophy of history Chaadaev attempted to reconcile the notion of a transcendent Providence with an immanentist approach, looking for the inner patterns that govern events and transform history into a meaningful process. The instruments of history are great, chosen individuals and historical nations, that is to say, nations which constitute supra-individual ‘moral personalities’. The mission of historical nations is to rise towards universality; hence they cannot lock themselves up in nationalistic particularisms and superstitions. Since the time of Christ the substance of history and the focal point of ‘world consciousness’ is Christianity, and the purest, most ‘historical’ and ‘social’, manifestation of Christianity is Catholicism. The Renaissance and Reformation had destroyed the splendid unity of medieval Christendom and pushed humanity towards increasing social atomization and neo-paganism. Now, however, this great spiritual crisis was drawing to a close. The process of corruption had reached its lowest point but Christianity had not collapsed; on the contrary, there were signs of its imminent regeneration and progressive transformation.

In defining this transformation Chaadaev followed the favourite idea of the post-Revolutionary French thinkers, both traditionalist (like de Maistre), and utopian socialist (like the Saint-Simonians): the idea of a renewed, socialized Christianity which was to bring about the ‘socialization of societies’, the extension of Christianity from private life to public sphere, that is the effective ‘ethicization’ of social and political relations. Christianity, he felt, was becoming social, and humankind was entering the last phase of the establishment of the Kingdom of God on earth.

2 Philosophy of Russian history

In his views on Russian history Chaadaev drew a sharp contrast between Russia and the West, as well as between Russia and ‘historical nations’ in general.

According to Chaadaev, Russia had no history. He meant by this that Russian history was merely a collection of meaningless events, lacking the continuity of tradition and the unity of purpose. This was so because of Russia’s separation from the universal Church, the Church of Rome. True, the Russians were Christians, but so were the Ethiopians; in both cases religious affiliation did not ensure membership of the Christian civilization. The European nations were marching through the centuries hand-in-hand, had fought together to free the Holy Sepulchre and prayed to God in one and the same language. The history of the West was sacred history, the Kingdom of God had to some extent already materialized in Europe. Russia, however, was a country forgotten by Providence, having no supraindividual consciousness and moral personality of its own, not covered by the ‘universal education of humankind’. Therefore the Russians had no guidelines and support in their past, lived without fixed customs and rules for anything, knew nothing about the ideas of duty, justice, rights and order. The ‘Western syllogism’ - logic and methodical thinking - was completely alien to their undisciplined minds. Small wonder that they grew but did not mature, that ‘not a single useful thought had sprouted in the sterile soil’ of their country.

The impression evoked by the publication of the ‘Philosophical Letter’ was enormous and almost completely negative. The most important exception was Aleksandr Herzen who appreciated the ‘Letter’ as a powerful protest against the hopelessness of Russian life under autocracy. The Tsar Nicholas I decided that Chaadaev should be treated as a madman and subjected to compulsory supervision by physicians and the police. This explains the title of Chaadaev’s text ‘Apologie d’un fou’ (The Apology of a Madman) (1837), in which he tried to make his views more palatable for Russian patriots. He admitted that his interpretation of Russian history was too severe and drew from his main thesis a different conclusion. The ‘lack of history’, he now argued, might also be a kind of privilege: without the burden of the past Russia would meet no obstacles in learning from the experience of Europe and in grounding its future on purely rational foundations.

In this way the Russian admirer of tradition formulated the view that being a blank sheet of paper made Russia a country of the future, destined to solve the grave social problems which had arisen in the West. This view, thanks to Herzen, was later to serve the purposes of Russian revolutionaries.

It is hard to say to what extent Chaadaev identified with his new view on Russia. It is likely that he really did admit the possibility of a great future for Russia, but was far from holding this view unreservedly. He could hardly be enthusiastic about the image of a civilization deprived of a religious basis and founded on purely scientific and

anti-historical principles. Hence he tried to combine his new ideas with his old dream about Russia's inclusion into the universal history. This implied the removal of the main cause of Russia's 'lack of history' - its separation from the Church of Rome.

3 Ecumenical ideas

Chaadaev's manuscripts, especially his 'Fragments et pensées diverses 1828-1830' (Fragments and Thoughts 1828-1830), show that his overall view on the Orthodox Church was in fact far from the one-sidedly negative picture presented in 'Philosophical Letters'.

The Western Church, Chaadaev argued, was called to make history. It was ambitious, intolerant and did not despise the goods of this world, but this was the reverse side of its sublime mission in universal history. The mission of the Eastern Church was quite different, unhistorical, as it were. It developed the contemplative and ascetic aspect of Christianity, and hence all forms of secular ambition were forbidden to it as incompatible with its inner essence. In this way Chaadaev arrived at the conception of the Roman Catholic and the Orthodox Church as the two poles of the absolute truth. On the one side there was 'activism' and 'sociality', on the other, monastic concentration of spirit, contemplation and the evangelical purity of Christ's teaching. But there can be no doubt that Chaadaev wanted the Kingdom of God to be realized in history and saw the Church as its historical incarnation. It followed from this that the Catholic Church was historically much more important, although the Orthodox Church (unlike the Protestant churches) was also a legitimate and valuable part of the Christian heritage. Hence the need of reconciliation and union between the two Churches: not a simple conversion of the Orthodox Christians to Catholicism but a union on equal rights, with the due recognition of the separate traditions of both Churches.

Very similar, if not identical, views were developed later (partially under Chaadaev's influence) by Vladimir Solov'ëv. It is justified to see Chaadaev as an important precursor of Solov'ëv's philosophical ecumenism.

In the history of Russian thought Chaadaev's name was firmly associated with pro-Catholic sympathies. Even Herzen saw him as representing 'neo-Catholicism', or 'revolutionary Catholicism'. The first edition of Chaadaev's works was published by the Russian Jesuit, Father Ivan Gagarin, whose conversion to Catholicism was caused by Chaadaev's 'Philosophical Letter' on Russia. Small wonder that Chaadaev's heritage could not be fully appropriated by Russian religious thinkers; nor, of course, by the programmatically secularist ideologies of the 'progressive intelligentsia'. In Soviet Russia attempts were made to present Chaadaev as a 'progressive' thinker, but publishing his collected works, let alone honestly discussing his ideas, was impossible. In post-communist Russia Chaadaev's works have been published in full and, for the first time, widely distributed. None the less, his place in the rich tradition of Russian religious philosophy and in the history of Russia's national self-identification has not become sufficiently clarified.

See also: Slavophilism

ANDRZEJ WALICKI

List of works

Chaadaev, P.Ia. (1914) *Sochineniia i pis'ma (Works and Letters)*, ed. M. Gershenzon, 2 vols, Moscow: A.I. Mamontov.

Chaadaev, P.Ia. (1969) *The Major Works of Peter Chaadaev*, ed. and trans. R. McNally, Notre Dame, IN: University of Notre Dame Press.(Includes commentary and bibliographical notes.)

Chaadaev, P.Ia. (1991) *Polnoe sobranie sochinenii i izbrannye pis'ma (Collected Works and Selected Letters)*, ed. Z.A. Kamenskii, 2 vols, Moscow: Nauka.(In Russian and French, with bibliographical references.)

Chaadaev, P.Ia. (1828-30) 'Fragments et pensées diverses 1828-1830' (Fragments and Thoughts 1828-1830), in [P. Tchaadaev], *Oeuvres inédités ou rares (Unpublished or Rare Works)*, ed. R. McNally, F. Rouleau and R. Tempest, Meudon: Bibliothéque Slave, 1990.(This collection includes bibliographical references and indexes.)

Chaadaev, P.Ia. (1828-31) 'Lettres Philosophiques' (Philosophical Letters), trans. M.-B. Zeldin, in *Philosophical Letters and Apology of a Madman*, Knoxville, TN: University of Tennessee Press, 1969.(Only the first of these letters was published in Chaadaev's lifetime, in *Teleskop*.)

Chaadaev, P.Ia. (1837) 'Apologie d'un fou' (Apology of a Madman), trans. M.-B. Zeldin in *Philosophical Letters and Apology of a Madman*, Knoxville, TN: University of Tennessee Press, 1969.(Chaadaev's presentation of

his views in a form more palatable to his contemporaries.)
[P. Tschaadaïeff] (1862) *Oeuvres choisies (Selected Works)*, Paris-Leipzig: A Franck.(The first edition of Chaadaev's works, published by Father I. Gagarin.)

References and further reading

Gershenzon, M. (1908) *P.Ia. Chaadaev. Zhizn' i myshlenie (P.Ia. Chaadaev, Life and Thought)*, St Petersburg: M.M. Stasiulevich.
Kamenskii, Z.A. (1991) *'Paradoksy Chaadaeva'*, introduction to Chaadaev, *Polnoe sobranie sochinenii* (1991) (see above).
Koyré, A. (1950) *Études sur l'histoire de la pensée philosophique en Russie (Studies in the History of Russian Philosophical Thought)*, Paris: J. Vrin.
McNally, R.T. (1971) *Chaadaev and His Friends*, Tallahassee, FL: Diplomatic Press.
Quénet, Ch. (1931) *Tchaadaev et les Lettres philosophiques (Chaadaev and the Philosophical Letters)*, Paris: H. Champion.
Walicki, A. (1975) 'The Paradox of Chaadaev', in *The Slavophile Controversy*, Oxford: Clarendon Press; 2nd edn, Notre Dame, IN: University of Notre Dame Press, 1989.

Chaldaean Oracles

The Chaldaean Oracles were a collection of revelatory verses purportedly compiled in the second century AD. Along with the Orphic texts, Neoplatonists regarded them as divine words. When the Oracles appear in philosophical works, they lend support to select cosmological, metaphysical or psychological propositions which have already been formulated.

According to Neoplatonists, the *Chaldaika logia*, or Chaldaean Oracles, originated with a certain Julianus, a late second-century AD ex-soldier in the eastern Roman army, and with his son, also named Julianus, who was their author. The father, surnamed 'Chaldaean' and a 'philosopher', may have collected handed-down material which he passed on to his son, a 'theurgist', who was 'divinely inspired' to write new oracles. This origin, some seventy years before Porphyry, by whom they are first mentioned, is by no means certain. However, it may be corroborated by similar material in Numenius, an important Neo-Pythagorean. Numenius flourished at that time in Apamea (Syria), where later Plotinus' theurgic-minded follower Amelius went to teach, as did Iamblichus, who was a major proponent of theurgy and the Oracles. Both Porphyry and Iamblichus had Syrian parentage. Moreover, in a unique fragment preserved by Proclus (*Commentary on Plato's Parmenides*, 594-5), 'the true theologians' used Syriac terms. Yet the question remains, how much was handed down and how much made up in a culture where Greek and Oriental had mixed since the third century BC and religions had absorbed Platonic ideas.

'Chaldaean' was often an honorific title for long-established wisdom, owing to the ancient astronomy of the Babylonians. In classical, and especially biblical, Greek, *logia theou* are God's own sayings (for example, the 'Hebrew *logia*' given to Moses, and those revealed in the New Testament). Neoplatonists, such as the emperor Julian, Proclus and Damascius, referred to the Chaldaean sayings as 'of the gods' or 'God-given' (*theoparadota logia*; an expression later used by John Damascene for revelation in the Christian Church). For comparison, Iamblichus also spoke of 'Pythagoras' *logia*. Thus the *logia* are statements directly attributed to a supreme authority.

Of the various fragments, some cite distinctive entities particularly, the 'wry necked' (*iynges*) birds, as divine powers. Others reflect general Babylonian and Mithraic themes (fire as creative and divine). Many rely on Platonism: 'cyclo-spiral' time (from Plato's *Timaeus*, where the planets have compound orbits and are cosmic clocks); divinity is 'once' (*hapax*) and 'twice transcendent' (*dis epekeina*) (based on the Good in Plato's *Republic*). However, two cosmological themes seem to be Chaldaean and important. First, there is a hierarchy of divinity, of which the highest, called 'Father', is an 'abyss'. Neoplatonists tried to associate this highest divinity with their transcendent One, or with Being or essence, which is opaque to ordinary conception but intuitively intelligible. Second, the universe is divided into 'material', 'aetherial' and 'empyrean' 'zones', of which the 'aetherial' indicated the semi-materiality of vital soul, and the 'empyrean' the transcendent world. However, the Neoplatonists associated the 'material' with earth and physical matter; the 'aetherial' with the heavens populated by stars and planets; and the 'empyrean' (fiery) with the incorporeal domain. Later, when amalgamated with Aristotelian and Christian themes, the whole scheme flourished in graphic representations of the universe through the Middle Ages to the seventeenth century.

The Chaldaean Oracles are referred to by all the later Neoplatonists down to Simplicius (sixth century), and by some Christians. The Oracles also emerge in Byzantine sources; for example, Psellus (eleventh century) and Plethon (fifteenth century), who collected them in a manuscript titled 'Sayings of the Zoroastrian Magi'. Finally, they were transmitted to Renaissance Italy, where they were used mainly by Patrizi da Cherso, Ficino and Pico della Mirandola for lending support to their own anti-Aristotelain arguments.

See also: Iamblichus; Neoplatonism §1; Porphyry §2; Proclus §5

LUCAS SIORVANES

References and further reading

Chaldean Oracles (2nd century AD?) *Oracles Chaldaïques*, ed. E. des Places, Paris: Les Belles Lettres, 1971; trans. R. Majercik, Leiden: Brill, 1989.(The former is parallel Greek text and French translation with notes, and includes the treatises by Proclus and Psellus; the latter is parallel Greek with English translation of fragments

and commentary, but without the tracts from Proclus and Psellus.)

Lewy, H. (1978) *Chaldaean Oracles and Theurgy*, ed. M. Tardieu, Paris: Études Augustiniennes, 2nd edn.(The seminal work on the Oracles, with updated evaluations, notes and articles.)

Proclus (5th century AD) *On the Chaldaean Philosophy*, ed. E. des Places, Paris: Les Belles Lettres, 1971; trans. T. M. Johnson 1907; repr. in S. Neuville (ed.) *Iamblichus: Exhortation to Philosophy*, Grand Rapids: Phanes Press, 1988.

Proclus (5th century AD) *Commentary on Plato's Parmenides*, trans. J.M. Dillon and G.R. Morrow, Princeton, NJ: Princeton University Press, 1987; repr. 1992.(English translation, including some Oracles and Orphic associations; see especially 593-5 for the fragment at 60 Klibansky, which is not cited in the current edition of the Chaldaean Oracles.)

Saffrey, H.D. (1981) 'Les néoplatoniciens et les Oracles Chaldaïques' (The Neoplatonists and the Chaldaean Oracles), *Revue des Études Augustiniennes* 26: 209- 25.(Important study on Neoplatonism, theurgy and the origin of the Oracles.)

Suda (10th century AD) *Suidae Lexicon*, ed. E. Adler, Leibzig: Teubner, 1931; repr. Stuttgart, 1967.(Greek text of the dictionary, where many ancient reports survive; the Chaldaean Julians can be found at Iota numbers 433 and 434.)

Change

Change in general may be defined as the variation of properties (whether of things or of regions of space) over time. But this definition is incomplete in a number of respects. The reference to properties and time raises two important questions. The first concerns whether we need to specify further the kinds of properties which are involved in change. If we define change in an object as temporal variation of its properties we are faced with the problem that some properties of an object may alter without there being a consequent change in the object itself. The second question concerns the passage of time: does temporal variation constitute change only in virtue of some feature of time itself, namely the fact (or putative fact) that time passes? Some philosophers have wished to reject the notion of time's passage. Are they thereby committed to a picture of the world as unchanging?

1 Cambridge change

In *The Principles of Mathematics* (1903), Russell defines change as follows:

> Change is the difference, in respect of truth or falsehood, between a proposition concerning an entity and the time T, and a proposition concerning the same entity and the time T', provided that these propositions differ only by the fact that T occurs in the one where T' occurs in the other.
>
> (Russell 1903: §442)

Geach (1969) has dubbed this kind of change 'Cambridge change', a term now familiar in the metaphysical vocabulary. Note that Russell defines change *simpliciter*, not specifically change in an entity, even though the definition is restricted to propositions about entities. But if we took it as a criterion for change in an entity, then we would have to allow that, for example, a cup of tea changes, not only when it cools down but when someone forms the intention to drink it. This second change passes Russell's test, since 'At T someone intends to drink the tea' is false, and 'At T' someone intends to drink the tea' is true. Yet, intuitively, someone's merely forming the intention to drink the tea is not a genuine change in the tea itself. We need something stronger than Russell's 'Cambridge criterion', as we might call it, if we wish to analyse genuine change in objects. Two possibilities will be briefly sketched here.

An object 'Cambridge-changes' if merely its relations to other objects change, as when for example it becomes smaller than another object simply because the latter has grown in size. If we can distinguish an object's *intrinsic* properties from its *extrinsic* properties, then we can define genuine change in an object as the variation over time of its intrinsic properties. We might attempt to characterize 'intrinsicness' as follows: *F* is an intrinsic property of *x* if and only if *x*'s being *F* does not logically depend upon the existence or properties of any other object. On this account, being an aunt, being ten yards from the top of Mount Etna and being read about do not count as intrinsic properties, whereas being ten yards in diameter and made of lead do count as intrinsic. However, this characterization is not without its difficulties. Change of shape is, surely, a genuine change in an object, but the shape of a thing does not count as intrinsic on the above criterion since having shape is, arguably, not logically independent of the existence of the containing space.

An alternative means of strengthening the Cambridge criterion employs the notion of causality. A genuine change in an object must, on this account, involve causal consequences contiguous to the object. This is not true of, for example, becoming an aunt or becoming famous - the effects of such 'changes' need not be near the object in question. Instead of defining real change in terms of intrinsic properties, we could define the intrinsic properties of an object as those whose alteration *must* involve contiguous effects. On this account, shape turns out, as we would expect, to be an intrinsic property of an object. But we cannot simply define real change in *x* as alteration which *must* have effects contiguous to *x*. The 'must' in italics is surely weaker than a logical must: there is no contradiction in supposing a change in mass, for example, to have no contiguous effects. But then if 'must' means instead 'must according to physical laws' then some mere Cambridge changes would pass the causal criterion. If *x* undergoes the Cambridge change of becoming less massive than some *y* as a result of *y*'s increase in mass, then this must (in the physical sense) have gravitational consequences, however small, in *x*'s vicinity. However, the chain of events leading to these effects in *x*'s vicinity must have started in *y*'s vicinity, so we could modify the account as follows: a real change in an object is such that, if it has effects, these must (in the physical sense) be mediated by effects contiguous to that object.

2 Change and the passage of time

In developing his famous argument for the unreality of time, McTaggart (1927) quotes, and proceeds to criticize, Russell's definition of change (see McTaggart, J. §2). McTaggart's first criticism is, in effect, that real change must involve, not difference in truth-value between *different* propositions, but alteration in truth-value of the *same* proposition. Suppose, to use McTaggart's example, a poker is hot on Monday and cold thereafter. Now it is true at all times that the poker is hot on that particular Monday and cold thereafter, so these facts about the poker do not change. The only real change consists of the fact that the poker's being hot is first a present state and then a past state. In other words, real change implies the passage of time.

Why should we accept this objection of McTaggart's? Let us describe the event of the poker's cooling down, that is its being hot at T and cold at T', as a first-order change. Then McTaggart seems to require a second-order change: the change in the event as it shifts from future to present to past. So it remains to be seen why someone who denies the existence of second-order change, but accepts first-order change, should be thought to be denying change altogether.

McTaggart, however, has another objection. We can construct a spatial analogue of Russell's Cambridge change: there are two spatial points, S and S', such that the proposition 'At S [for example, London] the Greenwich meridian is within the UK' is true, while the proposition 'At S' [for example, Paris] the Greenwich meridian is within the UK' is false. Now this purely spatial variation is clearly not change, but why not? It will not do to say that Russell's definition of change involves times and so excludes purely spatial variation, for this leaves unanswered the deeper question of why time, not space, is the dimension of change. A number of writers who deny that time passes have addressed themselves to this question. One account has it that time is the dimension of causality (since causes must precede their effects) and real change involves causally connected states.

A third argument, not considered by McTaggart, may be advanced for the view that real change in objects must involve the passage of time. It is a commonplace that, for an object to be said to undergo change, it must be the same object which first has one property and then loses that property (see Continuants §1). However, if we deny that time passes, then we must also deny that things can 'move through' time in order to have different properties at different times. Or, if this kind of talk is too metaphorical, we may put the point in the following terms. In order to have a property (or at least an intrinsic property) at a time, an object must exist at that time. Since change takes place over time, an object cannot change unless it exists at more than one time. But now suppose that there is no passage of time and that every proposition about an object is, if true, true for all times. Then, if *x* is F at T and Not-F at T', then it is true for all times both that *x* exists at T and that it exists at T'. Is this not a contradiction? An influential solution to this problem is to say that objects are extended in time in the same way as they are extended in space: that is, they have different (temporal) parts at different times, just as your feet are located in different regions of space. But this introduces another problem, for neither the temporal parts of a thing, nor the collection of temporal parts which constitutes the temporally extended object, can be said to change in the sense we have defined it. A temporal part is confined to an unextended instant and so cannot have different (intrinsic) properties at different times; and the properties that can be attributed to the collection as a whole, such as occupying threescore years and ten, do not vary over time (see Continuants §2).

An adequate account of change, therefore, needs to address these questions: is there such a thing as McTaggart's second-order change in the temporal properties of events? If there is not, must we think of things as having temporal parts? If we must, can we reconcile this with the apparently obvious fact that things change?

See also: Events; Processes

ROBIN LE POIDEVIN

References and further reading

Geach, P.(1969) *God and the Soul*, London: Routledge & Kegan Paul.(Referred to in §1 above.)

Helm, P. (1975) 'Are "Cambridge" changes non-events?', *Analysis* 35 (4): 140-4.(Argues that mere Cambridge changes are not genuine events, and that therefore they cannot be the causes of other events, and are not examples of uncaused events.)

Le Poidevin, R. (1991) *Change, Cause and Contradiction*, London: Macmillan.(Argues against the passage of time, defends temporal part theory and attempts to show how this can be reconciled with the existence of

change. Quite difficult.)

McTaggart, J.M.E. (1927) *The Nature of Existence*, vol. 2, ed. C.D. Broad, Cambridge: Cambridge University Press.(See pages 9-22 both for the material discussed in §2 above and for the famous proof of the unreality of time. Not an easy read, but does not presuppose a great deal of philosophical knowledge.)

Mellor, D.H. (1981) *Real Time*, Cambridge: Cambridge University Press.(Provides a causal criterion of genuine change, and argues that the denial of time's flow need not commit one to temporal parts and so can avoid the difficulties such parts raise for change. Lucid but quite demanding.)

Russell, B. (1903) *The Principles of Mathematics*, London: Allen & Unwin.(Quoted in §1 above.)

Chaos theory

Chaos theory is the name given to the scientific investigation of mathematically simple systems that exhibit complex and unpredictable behaviour. Since the 1970s these systems have been used to model experimental situations ranging from the early stages of fluid turbulence to the fluctuations of brain wave activity. This complex behaviour does not arise as a result of the interaction of numerous sub-systems or from intrinsically probabilistic equations. Instead, chaotic behaviour involves the rapid growth of any inaccuracy. The slightest vagueness in specifying the initial state of such a system makes long-term predictions impossible, yielding behaviour that is effectively random. The existence of such behaviour raises questions about the extent to which predictability and determinism apply in the physical world. Chaos theory addresses the questions of how such behaviour arises and how it changes as the system is modified. Its new analytical techniques invite a reconsideration of scientific methodology.

1 Defining chaos theory

Although there is controversy over the proper scientific definition of chaotic behaviour, chaos theory may be defined as the qualitative study of unstable aperiodic behaviour in deterministic nonlinear dynamical systems.

A dynamical system is a mathematical entity that usually results from the attempt to model a situation in the natural world (see Models). It has two components: a procedure for producing a mathematical description of the instantaneous state of a physical system, and a rule (the evolution equations) for transforming the current state description into a description for some other time. For some dynamical systems, the evolution equations yield a simple formula that allows the direct calculation of future states of the system.

The dynamical systems of interest for chaos theory contain nonlinear terms in the equations. These terms, such as x^2 or $\sin(x)$ or $5xy$, usually render an exact solution impossible. Researchers into chaotic phenomena instead seek a qualitative account of the system's behaviour, investigating the general character of its long-term behaviour, rather than seeking to arrive at numerical predictions of its exact future state (this special sense in which chaos theory is 'qualitative' must not be understood as contrasting with 'mathematical'). For example, a qualitative study of a planetary system would seek to discover what circumstances will lead to elliptical orbits as opposed to hyperbolic ones. This approach is called 'dynamical systems theory' (or sometimes simply 'dynamics') and began with the work of Henri Poincaré. Chaos theory is a specialized application of dynamical systems theory that typically treats such questions as: what characteristics will all solutions of this system ultimately exhibit? and, how does this system change from exhibiting one kind of behaviour to another kind?

Dynamical systems are usually studied in terms of trajectories in state space (sometimes called 'phase space'), a mathematically constructed abstract space where each dimension corresponds to one variable of the system. Every point in state space represents a full description of the system, and the evolution of the system manifests itself as the tracing out of a path, or trajectory, in state space. It is often possible to study the geometric features of these trajectories, even in actual experimental systems, without explicit knowledge of the solutions they represent. This allows one to characterize trajectories according to their topological features, and the investigation of how those features change as the parameters of the dynamical system are altered.

While qualitative questions can be asked about almost any dynamical system, chaos theory focuses on certain forms of behaviour - behaviour which is unstable and aperiodic. The form of instability known as sensitive dependence on initial conditions is a distinguishing characteristic of chaotic behaviour. A dynamical system that exhibits sensitive dependence on initial conditions will produce markedly different trajectories for two initial states that are initially very close together. In fact, given any specification of initial conditions, there is another set of initial conditions arbitrarily close to it that will diverge from it by some finite distance, given enough time. In most chaotic behaviour, sensitive dependence on initial conditions results from the exponential growth of any small initial difference. It is this instability that makes chaotic behaviour unpredictable (see §2).

Aperiodic behaviour occurs when no variable describing a property of the system undergoes a regular repetition of values. Furthermore, chaos is an appropriate label only when such behaviour occurs in a bounded system. An explosion thus does not qualify as chaotic behaviour. And systems studied by chaos theory bear the label

‘deterministic’ because their equations make no explicit reference to chance mechanisms (see §3).

For a dissipative dynamical system, characterized by the gradual loss of energy, trajectories in state space will asymptotically approach a shape known as an ‘attractor’ when transient effects have died away. Until the advent of chaos theory, only three types of attractor were generally recognized: the fixed point (corresponding to eventual equilibrium), the limit cycle (corresponding to periodic behaviour), and the torus (corresponding to behaviour with multiple periods in rationally incommensurable ratios). Chaotic behaviour in dissipative systems requires the introduction of a fourth variety: the so-called ‘strange attractor’.

A crucial geometric feature of strange attractors is their combination of stretching and folding. The action of such a system takes nearby points and ‘stretches’ them apart in a certain direction, thus creating the exponential divergence responsible for unpredictability. But the system also acts to ‘fold together’ points that are at some distance, bringing about convergence of trajectories in a different direction, and hence asymptotic attraction.

The stretching and folding of chaotic systems gives strange attractors the distinguishing characteristic of a nonintegral, or fractal, dimension. These attractors often appear as stacks of two-dimensional sheets packed in a self-similar structure that seems to intrude into three-dimensional space. The dimension of such an object - more than two but less than three - describes its scaling properties, giving a quantitative indication of the stretching and folding at work in the dynamical system. Another quantitative characterization of chaotic systems is given by the Lyapunov exponents, which measure the degree of sensitivity to initial conditions and thus the degree of unpredictability.

It should be noted that strange attractors appear only in dissipative systems. For systems where energy is conserved, there is no convergence onto an attractor. Instead, trajectories are confined to a surface of constant energy in state space. Chaotic behaviour can occur in such systems, but the trajectories will fill the allowed energy surface (which may itself display an extremely complicated structure).

The examination of the transition to chaotic behaviour has yielded important results for understanding the way complex behaviour arises. For example, period-doubling behaviour is found in a wide variety of nonlinear systems, and this type of transition has certain universal features. This universality allows us to make exact predictions about a system’s period-doubling behaviour as soon as we have verified that certain qualitative conditions hold.

Sensitive dependence on initial conditions, strange attractors and other elements of chaos theory have been reported in experimental systems in the physical and biological sciences. Some of these examples of chaotic behaviour have been convincingly documented in laboratory settings, while some of the examples of chaos outside the laboratory are the subject of lively debate.

2 Predictability

Moderately accurate initial conditions will always allow us to predict the chaotic behaviour of a system for some short time into the future, but since the vagueness in our initial conditions grows exponentially with time, we can see that for some chaotic systems, a useful prediction would require initial conditions to be specified with more accuracy than is possible. Inaccuracy can be due to equipment error, ineliminable perturbations from outside influences, round-off error in the representation of real numbers, or even the inherent indeterminacy of the quantity being measured.

Some may contend that so long as the prediction task for a chaotic system requires only some finite degree of accuracy, it is incorrect to say that such accuracy is impossible to achieve.

Since chaos theory, unlike special relativity or quantum mechanics, introduces no new fundamental postulates about the physical world, it provides no new physical theory that rules out the possibility of such predictions. Yet these same systems may nevertheless be unpredictable *in principle* because our inability to make such calculations in the longer term is not due to some constraint we could overcome. The very finitude of human resources implies that it is impossible to obtain the accuracy necessary for predictions.

The limitation posed by chaotic systems presents a challenge to the methodological assumption that a small amount of vagueness in measurements will lead to only a small amount of vagueness in predictions. The standard

technique of perturbation analysis treats a difficult case as a version of a predictable system plus some small 'perturbation' that complicates the situation but keeps it mathematically tractable. Sensitive dependence guarantees that such approximative techniques will be worthless, thus calling into question some of the metaphysical beliefs that underwrite their general applicability. One such belief is the conviction that systems governed by simple rules must behave simply.

3 Determinism

Scientists and philosophers use the term 'determinism' in several senses (see Determinism and indeterminism §1). By one criterion, a system is deterministic if the way it changes in time can be specified by a set of equations that make no reference to chance. That is, there are no probabilities built into the state description due to an averaging over many indistinguishable substates, and the evolution equations do not involve probabilistic branching. Thus, the systems treated by chaos theory are deterministic by definition.

Another criterion consists in the uniqueness of a system's evolution: if two systems are identical in physical properties at one time, they must remain identical at all other times. That is, the complete instantaneous description of a deterministic system 'fixes' the past and future with no alternatives.

A third notion of determinism is that a system is predictable in principle by an all-powerful intelligence or computational scheme, given complete information of instantaneous conditions and the complete set of physical laws. Since measurements and calculations cannot be made with unlimited accuracy, a chaotic system will be unpredictable even though it satisfies other requirements for determinism. In other words, we may not be able to tell *which* unique trajectory a system is following, but that does not mean such trajectories do not exist.

On one interpretation, the randomness and apparent indeterminism of chaotic behaviour are merely artifacts of our epistemic limitations. Another interpretation would treat unique trajectories as inadmissible idealizations, arguing against uniqueness of evolution in chaotic behaviour. This instrumentalist approach argues that there is no point in insisting that the future is fixed, given the present, if that future cannot be predicted.

A third approach would draw on the implications of quantum mechanics to show that unique evolution encounters difficulty due to the workings of chaotic dynamics on intrinsically indeterminate quantities. Such an approach would require an account of how to combine the Newtonian formulations of nonlinear dynamics with the mathematical formalism of quantum mechanics - the difficult task of making sense of 'quantum chaos'.

4 Chaos and scientific understanding

One of the apparent paradoxes of chaos theory is that it represents a scientific study of unpredictable systems that actually has significant predictive power. Chaos theory provides an understanding of how predictable large-scale or long-term patterns appear in behaviour which is nonetheless unpredictable in detail. It may seem odd to assert that chaos theory has 'great predictive power', but such claims refer not to an ability to predict the exact value of some property of a system, but to an ability to foresee and understand changes in the overall behaviour of that system. The crucial point here is the distinction between specific *quantitative* predictions, which are typically impossible for chaotic systems, and *qualitative* predictions, which are at the heart of dynamical systems theory.

The fruitfulness of chaos theory militates against the methodological assumption that it is always appropriate to seek to understand the behaviour of a system by trying to determine the equations governing the interactions of its parts (see Reduction, problems of). For most nonlinear dynamical systems, such a reductionist strategy does not help us understand how the chaotic behaviour sets in, or what kind of attractor characterizes it, or how the system will respond to changes in parameters. Understanding this behaviour requires the use of the new mathematical techniques of nonlinear dynamics. While chaos theory argues against the universal applicability of the method of reductionism, it provides no examples of 'holistic' properties which could serve as counter examples to metaphysical doctrines of reductionism, such as physicalism.

Chaos theory shows us the mechanisms responsible for unpredictable behaviour, but these mechanisms are best characterized as geometrical rather than causal. Systems with very different underlying causal substrata are grouped together in order to study their qualitative behaviour. Chaos theory does not need to reveal the causal process at work, and it would be impossible in principle to trace out the workings of the actual causal mechanism

in a chaotic system. To do this would require calculations using the evolution equations of the system.

According to some accounts, science furnishes understanding by widening the scope of events that happen out of necessity and not contingency - by revealing laws of nature and subsuming phenomena under them (see Explanation §2). While chaos theory establishes powerful generalizations, its practitioners do not portray their work as discovering new laws of nature. For instance, there is no set of necessary and sufficient conditions that would allow us to foresee which type of transition to chaotic behaviour a system will follow. Chaos theory emphasizes finding patterns and connections, while jettisoning the requirement that the patterns must yield necessity in a detailed and deterministic sense. Some may contend that this search for patterns actually constitutes striving to discover new laws governing qualitative features of systems.

Perhaps chaos theory does not provide us with any genuine understanding at all. After all, it does not introduce any fundamental revisions in our laws of nature. Some might argue that it merely introduces stop-gap measures for data analysis in systems with intractable equations. Perhaps it is only a source of new observations or empirical generalizations. But nonlinear dynamics does require a wholesale revision of notions about the applicability of classical models, and provides powerful new analytic techniques. It should not be disparaged on the grounds that fundamental theoretical structures are left unchanged.

5 Applications outside the natural sciences

Chaos theory has attracted attention from outside the natural sciences. In the social sciences, the techniques developed for analysis of nonlinear systems have been applied to economic data, for instance. Such techniques can determine the extent to which random fluctuations in a chosen variable can be modelled with simple deterministic mechanisms.

Sensitive dependence on initial conditions has also served as a metaphor for the precarious swings of human history. Metaphorical interpretations of chaos theory have also appeared in natural theology and literary theory. Other areas where chaos theory may have philosophical implications are: the foundations of statistical mechanics, the problem of freedom of the will, and the role of the cultural context of science.

See also: Randomness

STEPHEN H. KELLERT

References and further reading

Batterman, R.W. (1991) ‘Chaos, Quantization, and the Correspondence Principle’, *Synthèse* 89: 189-227.(A useful discussion of some of the issues surrounding ‘quantum chaos’.)

Bergé, P., Pomeau, Y. and Vidal, C. (1984) *Order Within Chaos*, trans. L. Tuckerman, Paris: Wiley & Sons. (Useful guide to chaos in the physical sciences.)

Devaney, R. (1986) *An Introduction to Chaotic Dynamical Systems*, Menlo Park, CA: Benjamin-Cummins Publishing.(A standard presentation of the mathematics of chaos at the introductory level.)

Dyke, C. (1990) ‘Strange Attraction, Curious Liason: Clio Meets Chaos’, *Philosophical Forum* 21: 369-92. (Examines the implications of chaos theory for history and the social sciences.)

Earman, J. (1986) *A Primer on Determinism*, Dordrecht: Reidel.(A comprehensive examination of philosophical issues about scientific determinism. Includes a technical discussion of chaos.)

Ford, J. (1989) ‘What is Chaos, That We Should be Mindful of It? ’, in P. Davies (ed.) *The New Physics*, Cambridge: Cambridge University Press, 348-72.(Presents a number of arguments and speculations regarding predictability and determinism as they relate to chaos theory.)

Hao, B.-L. (ed.) (1984) *Chaos*, Singapore: World Publishing Company.(A useful collection of historically important scientific papers on chaos.)

Hobbs, J. (1991) ‘Chaos and Indeterminism’, *Canadian Journal of Philosophy* 21: 141-64.(Presents an argument on the relevance of chaos theory to issues of determinism.)

Kellert, S.H. (1993) *In the Wake of Chaos*, Chicago, IL: University of Chicago Press. (Discusses the philosophical implications of chaos theory, especially regarding predictability, determinism, scientific understanding and social context.)

Smith, P. (1991) ‘The Butterfly Effect’, *Proceedings of the Aristotelian Society* 91: 247-67.(Presents an argument on the implications of chaos theory for scientific prediction.)

Charity

Within at least some branches of Christianity, the term 'charity' has been used to mean the love mandated by Jesus. In recent theological writings, however, there has been a tendency to replace it with the Greek word agapē*. There has been some disagreement in the twentieth century concerning the precise nature and functioning of Christian love, a major catalyst for debate having been Anders Nygren's book Agapē and Eros (1930-6). Numerous scholars have complained that charity does not have a high profile nowadays and have noted that, in common parlance, the word usually has the meaning of benevolence or beneficence. Some attempts have been made to place greater emphasis on Christian love and relationships within Christian ethics. Of some interest in this regard is the notion of an ethic of care, which is not confined to Christian circles but has been the subject of some debate in recent times.*

1 Christian love

In much Christian literature, perhaps most especially in Roman Catholic theological writings, the word 'charity' has been used to denote the love for God and neighbour commanded by Jesus (John 15: 12) and extolled by St Paul (1 Corinthians 13). In the writings of Aquinas and others, charity (in Latin *caritas*) is also described as the greatest of the theological virtues (see Theological virtues). The more generic term 'love', which is commonly used in the context of erotic relationships and friendship, has been said by many authors to be too imprecise to convey the distinctiveness of Christian love (see Love §2). 'Charity' too, however, is not devoid of ambiguity. In fact, it is most commonly used nowadays to indicate benevolence, beneficence, philanthropy or an organization dedicated to philanthropic pursuits (see Help and beneficence). Some writers have made use of the expression 'charity-love', but, as Leandro Rossi has pointed out (1976), this term, in addition to being too complex and divorced from what is normal, might serve only to combine the ambiguities. The most suitable word, he concludes, is *agapē*. This Greek word is, of course, widely used in the New Testament. Moreover, it would seem to be the preferred term nowadays in theological circles - including Roman Catholic ones. Indeed, Gene Outka goes so far as to say that, in current usage, *agapē* 'is almost uniformly the referent for any alleged distinctiveness in Christian love' (1978: 7).

A mere change of word, however, does not suffice to eliminate all disputes about Christian love. Writing in the 1930s, Anders Nygren claimed that the classical Catholic idea of *caritas*, which stemmed largely from the writings of Augustine, has little to do with specifically Christian love. This is so, he maintains, because, in the Catholic notion, the primitive Christian *agapē* motif is combined with the *eros* motif. *Eros* here means the human desire for union with God. *Agapē*, on the other hand, is the distinguishing mark of Christianity. Christian fellowship with God depends entirely upon God's love (*agapē*). On the human side, there is no way that leads to God. There is only a way for God to come to humans, and that is *agapē*. In other words, there is simply no connection between *agapē* and *eros*. They are totally incompatible. He goes on to say that Christians have nothing of their own. The love they show to others is the love that God has infused into them (Romans 5: 5).

The renowned Protestant scholar Karl Barth (1967) also saw the need to distinguish *eros* and *agapē*. He went on to say, however, that the kind of relationship in which only God was at work and humans were mere channels of divine action could not be described as a covenant relationship, and yet, according to Scripture, that is the true relationship between them (see Barth, K.). The initiative in this relationship is wholly and exclusively on the part of God, but that initiative aims at a correspondingly free human act, not the reaction of a puppet. The Roman Catholic moral theologian Bernard Häring (1979), on the other hand, takes the view that all human love can be redeemed. When, in response to God's *agapē*, believers commit themselves to God, he says, all their prior dispositions, including *eros* and friendship, are gradually transformed, purified and raised to a higher level.

There has been some complaint in recent times that Christian love - whatever we wish to call it and whatever its relationship to *eros* may be - does not figure highly enough in Christian ethics. Referring specifically to Roman Catholic works, and writing in the 1950s, Gérard Gilleman noted that manuals of moral theology had law rather than love as their dominant theme. He therefore sought to establish 'a method of exposition in which charity will play the role of a vital principle, just as it does in the message of Christ and in Christian life' (1959: xxxvi). More recently, some Roman Catholic moral theologians have laid more emphasis than had previously been apparent in

their discipline upon the distinction between personal goodness and the rightness of acts. Personal goodness exists when someone is truly loving (charitable). Mere rightness of acts does not necessarily indicate personal goodness. One could perform an ethically right act for a variety of motives, not all of them loving.

Within Protestantism, attempts to place greater emphasis on Christian love in Christian ethics can be found among some situation ethicists (see Situation ethics). One of the best known is Joseph Fletcher, who holds that principles and rules can be enlightening, but never the deciding factor. That status can only be given to love. In a given situation, one has to decide what is the loving thing to do. Numerous writers have lamented the fact, already noted, that the word ‘charity’ is now used chiefly to denote benevolence rather than Christian love. It could therefore seem surprising that Fletcher should describe Christian love as precisely benevolence. Having noted, however, that such words as ‘benevolence’ and ‘goodwill’ have taken on a tepid meaning, Fletcher goes on to say that ‘*Agapē* goes out to our neighbors, not for our own sakes nor for theirs, really, but for God’s’ (1966: 105).

Whatever one makes of this contention or of Fletcher’s situation ethics in general, the fact remains that, in recent times, the word charity has been used most often to refer to benevolence in a sense that is not specifically theological, or to beneficence. One can hardly claim, moreover, that the reason is purely linguistic - other senses of ‘charity’ having been taken over, so to speak, by *agapē*. Outside the confines of theological circles, the Greek word is seldom encountered.

2 An ethic of care

Moreover, it is claimed by many that, beyond the confines of theological circles, there has been very little emphasis in recent times upon the demands of charity in ethics. Even discussions of philanthropic endeavours such as those aimed at the relief of poverty, it is pointed out, are discussed habitually in terms of justice. Given this state of affairs, recent discussion about an ethic of care has been particularly interesting (see Feminist ethics §1). The debate began with Carol Gilligan’s book *In a Different Voice* (1982). Gilligan had noted that developmental psychology had been based largely on a study of men’s lives. Lawrence Kohlberg’s theory of moral development, moreover, was built on research that involved no females at all. His six stages of development in moral judgment through which he says one should pass in order to achieve maturity in moral reasoning are founded empirically on a study of eighty-four boys whose development he followed for over twenty years. Gilligan noted that groups not included in the original sample rarely reached the higher stages. This was particularly true of women, whose judgments, on the whole, seemed to exemplify Kohlberg’s third stage. At that stage goodness is seen in terms of helping and pleasing other people, and morality is conceived in interpersonal terms. At the ‘higher’ stage four, ‘relationships are’, in Gilligan’s words, ‘subordinated to rules’, while rules are, in turn, subordinated to universal principles of justice at stages five and six (1982: 18). Kohlberg’s notion of maturity derives from a study of men’s lives and, says Gilligan, ‘reflects the importance of individuation in their development’. If, however, developmental constructs are derived from a study of women’s lives, a conception of morality as concerned with care emerges. Moral development is then seen as centred around responsibility and relationships, whereas, when morality is seen in terms of fairness, moral development is tied to the understanding of rights and rules. Gilligan has noted, however, that the care perspective is not confined to women (see Moral development).

An ethic of care does not necessarily involve the specifically Christian concept of charity or *agapē*, but, no doubt, many would be of the opinion that it could, if it were adopted in Christian circles. Whether such charity is involved or not, however, it would seem that both philosophical ethics and specifically Christian ethics might benefit from further debate about Gilligan’s basic idea. Interestingly, as John R. Donahue noted (1977), the biblical notion of justice could be described in terms of faithfulness to the demands of relationships.

See also: Faith; Hope; Theological virtues

BERNARD HOOSE

References and further reading

Barth, K. (1967) *Church Dogmatics*, vol. 4, *The Doctrine of Reconciliation*, Edinburgh: T. & T. Clark, part II, 727-840. (A detailed exposition of Barth’s ideas on Christian Love.)

Donahue, J.R. (1977) ‘Biblical Perspectives on Justice’, in J.C. Haughey (ed.) *The Faith that Does Justice: Examining the Christian Sources for Social Change*, New York: Paulist Press, 68-112. (An analysis of the concept of justice in the Bible.)

Fletcher, J. (1966) *Situation Ethics: The New Morality*, London: SCM Press. (A basic work on situation ethics seeing only love as always good.)
Gilleman, G. (1959) *The Primacy of Charity in Moral Theology*, Westminster, MD: The Newman Press. (An attempt to centre morality in charity.)
Gilligan, C. (1982) *In a Different Voice: Psychological Theory and Women's Development*, Cambridge, MA: Harvard University Press. (The book which is at the heart of the debate regarding an ethic of care.)
Häring, B. (1979) *Free and Faithful in Christ*, vol. 2, *The Truth Will Set You Free*, Slough: St Paul Publications, 419-91. (A Roman Catholic view of Christian love.)
Keenan, J.F. (1992) *Goodness and Rightness in Thomas Aquinas's Summa theologiae*, Washington, DC: Georgetown University Press. (A scholarly investigation of the goodness/rightness distinction, including reference to the role of charity in the distinction.)
Larrabee, M.J. (ed.) (1993) *An Ethic of Care: Feminist and Interdisciplinary Perspectives*, London: Routledge. (A collection of contributions to the debate on an ethic of care by various authors, with particular attention being given to matters concerning moral philosophy, psychological aspects and cultural aspects.)
Nygren, A. (1930-6) *Agapē and Eros*, London: Society for Promoting Christian Knowledge. (A long and very detailed study of the meaning of Christian love. The author also sets out to show the major changes in meaning that have taken place in the course of history.)
Outka, G. (1978) *Agapē : An Ethical Analysis*, New Haven, CT and London: Yale University Press. (An examination of modern writings on *agapē*, including those by Anders Nygren, M.C. D'Arcy, Gérard Gilleman, Reinhold Niebuhr, Paul Ramsey and Karl Barth, special attention being given to the work of the last-mentioned.)
Rossi, L. (1976) 'Carità', in L. Rossi and A. Valsechi (eds) *Dizionario enciclopedico di teologia morale*, Rome: Edizioni Paoline, 89. (An analysis of Christian love.)

Charity, principle of

The principle of charity governs the interpretation of the beliefs and utterances of others. It urges charitable interpretation, meaning interpretation that maximizes the truth or rationality of what others think and say. Some formulations of the principle concern primarily rationality, recommending attributions of rational belief or assertion. Others concern primarily truth, recommending attributions of true belief or assertion. Versions of the principle differ in strength. The weakest urge charity as one consideration among many. The strongest hold that interpretation is impossible without the assumption of rationality or truth.

The principle has been put to various philosophical uses. Students are typically instructed to follow the principle when interpreting passages and formulating the arguments they contain. The principle also plays a role in philosophy of mind and language and in epistemology. Philosophers have argued that the principle of charity plays an essential role in characterizing the nature of belief and intentionality, with some philosophers contending that beliefs must be mostly true. A version of the principle has even served as a key premise in a widely discussed argument against epistemological scepticism.

1 Versions of the principle of charity

The principle of charity is typically taken to be a principle about how to interpret the words and thoughts of others, favouring interpretations that maximize the accuracy and rationality of their utterances and beliefs. There are a variety of more precise formulations of the principle, emphasizing different aspects. Principles of charity also differ with regard to the strength and source of the commitment to charitable interpretation.

Truth and rationality are two dimensions along which one can urge charity. It can be that a person's situation is such that the most reasonable thing for them to believe happens to be false. If there is a question about how to interpret the person, where the choice is between attributing a reasonable but false belief and a true but unreasonable belief, the various versions of the principle of charity will choose differently.

Principles of charity can also differ in strength. The weakest merely recommend consideration of charitable interpretations, leaving open the possibility that in some circumstances one can properly interpret others in ways that make them irrational or mistaken. Moderate positions demand interpreting others as rational or right except in those rare cases in which there are compelling reasons to do otherwise. Stronger versions hold that it is never appropriate to interpret others as irrational or systematically mistaken. It is possible to regard the commitment to charity as merely a contingent but universal truth about the interpretation of humans by imperfect human interpreters or as an a priori requirement about all interpretation. Some versions of the principle are not primarily directives about how to interpret others but are, instead, principles to the effect that intentional states are by their nature rational or accurate so that interpreting others as irrational or mistaken is necessarily incorrect.

2 Charity and critical thinking

In courses on critical thinking (or argument analysis or informal logic), students learn a method for interpreting and evaluating the arguments and assertions that they read and hear. The principle of charity often plays a central role in the methodology taught in such courses. The version of the principle advanced in this context is typically one of the weaker ones mentioned above, recommending that charitable interpretations be favoured, but acknowledging other considerations as well. Advocates of charity differ over the basis for the principle. Some see it as an ethical matter. On this view, being charitable is part of being fair to those with whom one interacts. An alternative view emphasizes a more epistemological and pragmatic basis for being charitable: being charitable best contributes to one's own pursuit of the truth. Interpreting others in ways that make their words less plausible may make it easy to refute them. But interpreting and reflecting on the most plausible interpretation of their words is apt to increase the interpreter's own understanding of the issues addressed. When there is a more plausible and more insightful interpretation available, there is little point in considering a less plausible interpretation of the assertions or arguments of others.

3 Charity and philosophy of mind

Something resembling the principle of charity plays a key role in Daniel Dennett's influential philosophy of mind.

According to Dennett, when we describe a thing as an intentional system (a thing with propositional attitudes), we must take it to be rational. Dennett's claim is not endorsing any of the weaker theses described above. Rather, his claim is about the possibility of explaining or characterizing irrationality in intentional terms. He says of a purported case of irrationality, 'we must descend from the level of beliefs and desires to some other level of theory to describe' the case, since such cases 'defy description in ordinary terms of belief and desire' (Dennett 1982: 66-7). The claim here is not that we are always rational, but just that we cannot give intentional explanations of seemingly irrational behaviour. On Dennett's view, charity is not just a pragmatic, moral or epistemic principle to help guide interpretation. It is an essential part of intentional interpretation. Dennett argues for this partly by appeal to general considerations about belief and other intentional states, partly by appeal to the difficulty he finds in filling out the details of intentional explanations in cases of alleged irrationality.

4 Charity, philosophy of language and epistemology

The most widely explored use of the principle of charity is in discussions of radical translation. W.V. Quine and Donald Davidson are the most influential defenders of the principle in this setting. Quine discusses cases of radical translation - translation in cases in which the words to be translated are those of people in a foreign and unfamiliar culture. He contends that if we come up with a translation scheme that renders their beliefs and utterances regularly silly or mistaken, we will eventually be driven to question our translation scheme and replace it by one that makes more sense of them. Thus, if a person utters a certain sentence when and only when a rabbit passes by, one is to interpret that sentence to mean something like the true 'That's a rabbit', rather than the false 'That's a duck'. Of course, other things the person says might make the latter interpretation best, but then these other things are interpreted as truths, thereby preserving the truth of most of what the person says (see Quine, W.V. §9; Radical translation and radical interpretation).

Donald Davidson defends an account similar to Quine's, but he puts it to a surprising use, developing an influential argument against epistemological scepticism (see Davidson, D. §5). Sceptics deny that we have knowledge, either globally or in some particular domain. Davidson appeals to the principle of charity in an argument against global scepticism. According to Davidson, what the principle of charity really requires is that the interpreter construe the subject's words so that the subject's beliefs mostly match the interpreter's. There's no guarantee of truth here; at best there is a guarantee of agreement. This would hardly worry a sceptic, as Davidson is willing to concede. But his argument has a second stage designed to overcome scepticism. He asks that we imagine an omniscient interpreter who believes all and only truths about the world of some subject. The omniscient interpreter is also bound by the principle of charity, and thus must interpret the subject's beliefs so that they largely match the interpreter's own. Since the corresponding beliefs of the omniscient interpreter are all true, the beliefs of the subject must be largely true. Thus, the nature of interpretation rules out the possibility that any interpretable being has beliefs that fail to be largely true. Hence, any scepticism based on the possibility that one's beliefs are mostly false is ruled out.

Quite a few philosophers have found Davidson's argument questionable, though they have differed over its fundamental flaw. Among the criticisms are these:

(1) The appeal to the omniscient interpreter raises a cluster of complex questions about possibility. Davidson surely does not want to rest his argument upon the assumption that there actually is an omniscient interpreter. Rather, his claim is merely that there could be one. Suppose that assumption is granted. Suppose further that in some possible world the omniscient interpreter interprets someone's beliefs. It may follow that, in that world, anyone interpreted must have mostly true beliefs. But nothing follows about the accuracy of anyone's beliefs in the actual world. Hence, nothing contrary to scepticism about actual humans follows.

There are various ways for Davidson to formulate his argument to try to get around this objection, perhaps the most promising being the following. There may be an omniscient interpreter in some other world who has true beliefs about the actual world and who interprets the beliefs of any individual, *S*, in the actual world. That interpreter is bound by the same rules of interpretation, and thus *S*'s beliefs must largely coincide with the true beliefs of the interpreter. Hence, *S* has beliefs that are mostly true. Critics of this response may doubt the justifiability of assuming that there is a possible interpreter who is (a) omniscient, (b) capable of interpreting *S*, and (c) required to make *S*'s beliefs match the interpreter's own beliefs.

(2) Even if we grant Davidson's conclusion that it is impossible that most of *S*'s beliefs are false, it does not follow that *S* knows much. From the fact that most of *S*'s beliefs are true, and even from the fact that *S* knows this, it does not follow that with respect to almost any individual belief that *S* considers that *S* knows that it is true. Consider the analogy of a lottery in which most of the tickets will win small prizes. One can know that most tickets are winners without knowing of any particular ticket that it is a winner. Similarly, even if *S* knows that most of *S*'s beliefs are true, it may be that there are none (other than that one and its consequences) that *S* knows to be true. Thus, even if sound, Davidson's argument leaves unscathed an extreme sceptical view. Davidson might reply that *S*'s knowledge that *S* is mostly right provides justifying evidence for each of *S*'s beliefs, thereby making each true belief justified, and thus knowledge. This, however, seems to rely on a remarkably low standard for knowledge-level justification.

(3) Much of the interest in Davidson's argument stems from the fact he attempts to draw such an epistemologically significant conclusion from a fact about the nature of radical interpretation. Thus, much of the criticism has focused on the connection between the principle of charity and scepticism. However, the principle itself is not without critics. Indeed, one may well wonder why Davidson thinks charity is essential for interpretation. Surely, in general interpreters do not always have to make their subject's beliefs largely agree with their own: interpreters will have many beliefs about their own home environment, beliefs remote subjects need not share. What may be more plausibly required is that interpreters attribute to subjects beliefs that the interpreters would hold given the subject's evidence. In other words, perhaps in interpretation one must assume that one's subjects respond to evidence in much the way one does oneself. Even this will seem dubious to many. Why, it could be asked, can one not take a subject to be responding erroneously to evidence? But even if this sort of principle of charity is granted, it is doubtful that the desired anti-sceptical conclusion follows. This is because one can grant that another person is responding properly to evidence but is nevertheless massively deceived. Thus, a brain in a vat, or a victim of the Evil Demon, might respond properly to evidence, but still have consistently false beliefs because that evidence is systematically misleading. Hence, a revised, and perhaps more plausible, version of the principle of charity requires only that interpreted subjects respond properly to evidence. This leaves open the possibility of massive deception and thus does not yield the anti-sceptical result.

(4) Facts about how best to interpret others are in general logically independent of how those others actually must be. It would therefore be surprising if a requirement that we interpret others so as to make them not globally mistaken could yield the conclusion that they cannot actually be globally mistaken. It is, therefore, noteworthy that in some descriptions of his argument against scepticism, Davidson presents it in a way in which the principle of charity is not fundamentally a principle about how to interpret others but rather a principle about what intentional states must be like. He writes, 'What stands in the way of global skepticism is, in my view, the fact that we must, in the plainest and methodologically most basic cases, take the objects of a belief to be the causes of that belief. And what we, as interpreters, must take them to be is what they in fact are' (Davidson [1983] 1986: 317-18). The idea here is that the content of a belief is determined by the external states that cause it. Thus, for example, the belief that a subject has when undergoing a certain sort of sensory stimulation has as its content the external state regularly causing that sort of stimulation. This is a causal theory of meaning or content. On this view, it is the nature of language and belief that requires that our beliefs and utterances be largely true. This makes charitable interpretation appropriate, but only because independent facts about belief and language imply that uncharitable interpretations are always inaccurate.

Whatever the nature and role of the principle of charity in the version of Davidson's argument just described, the argument is not without its critics. Some dispute the causal theory of meaning upon which it depends. Others contend that even if it is sound, Davidson's argument does not really refute scepticism. They argue that even if it shows that that we know that our beliefs are true, we can still fail to know the content of our beliefs. This is because we cannot tell which external states are systematically causally connected to our beliefs. For example, it is possible that a particular sort of belief is systematically connected either to the presence in one's visual field of a certain kind of object or to a certain sort of computer state in the computer to which one's brain is connected. Davidson's view apparently implies that the content of the belief depends upon which sort of cause it has, and, the critics contend, Davidson has not shown how one can know which sort of cause one's belief has. This seems to leave intact a radical form of scepticism. Whether this objection succeeds depends upon whether there is room within Davidson's view for some sort of privileged access to the contents of one's beliefs.

See also: Rational beliefs; Scepticism

RICHARD FELDMAN

References and further reading

Brueckner, A.L. (1986) 'Charity and Skepticism', *Pacific Philosophical Quarterly* 67 (4): 264-8.(Develops the last objection to Davidson's argument mentioned in §4 above.)

Brueckner, A.L. (1991) 'The Omniscient Interpreter Rides Again', *Analysis* 51 (4): 199-205. (Defends Davidson's argument from objection 1 in §4 above.)

Davidson, D. (1982) 'Empirical Content', *Grazer Philosophische Studien* 16/17: 471-89; repr. in E. LePore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell, 1986, 320-32.(Includes a version of the anti-sceptical argument discussed in §4 above.)

Davidson, D. (1983) 'Coherence Theory of Truth and Knowledge', in D. Henrich (ed.) *Kant oder Hegel?*, Stuttgart: Klett-Cotta; repr. in E. LePore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell, 1986, 308-19.(Develops the argument discussed above in §4.)

Dennett, D. (1982) 'Making Sense of Ourselves', in J.I. Biro and R.W. Shahan (eds) *Mind, Brain, and Function: Essays in the Philosophy of Mind*, Norman, OK: University of Oklahoma Press, 63-81.(Discusses the difficulty of attributing irrational beliefs to people.)

Foley, R. and Fumerton, R. (1985) 'Davidson's Theism?', *Philosophical Studies* 48 (1): 83-90.(Develops objection 1 from §4 above.)

Govier, T. (1987) *Problems in Argument Analysis and Evaluation*, Dordrecht: Foris Publications.(Discusses the role of charitable interpretation in critical thinking and argument analysis.)

Klein, P. (1986) 'Radical Interpretation and Global Skepticism', in E. LePore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell, 1986, 369-86.(Clarifies Davidson's anti-sceptical argument and develops objection 2 in §4 above.)

McGinn, C. (1986) 'Radical Interpretation and Epistemology', in E. LePore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell, 1986, 356-68.(Develops a version of objection 3 in §4 above.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Most widely discussed statement of Quine's views about radical translation.)

Charleton, Walter (1620-1707)

The physician Walter Charleton was the first to introduce Epicurean atomism into England in the form advocated in France by Gassendi. Charleton's version of atomism, although largely derivative, was nevertheless influential. Together with his advocacy of a Christian hedonism, it helped to make both atomism in natural philosophy (with its associated mechanistic account of nature) and utilitarian theories in ethics acceptable to such thinkers as Robert Boyle, Isaac Newton, John Locke and others associated with the foundation of the Royal Society, of which Charleton was himself an active early member.

Walter Charleton was a physician who served both Charles I and Charles II, but his interests were always broader than medicine. He entered practice after studying at Magdalen Hall, Oxford, under John Wilkins, also one of the founders of the Royal Society. Charleton's quick rise to fame made him enemies, which in part explains his sad decline to poverty in later years after serving as President of the College of Physicians from 1689-91.

Charleton's major philosophical writings were two books on the philosophy of Epicurus and two on central religious themes (the absurdity of atheism and the immortality of the soul). But it is important to appreciate that although he clearly saw himself as an exponent of a modified Christian version of the philosophy of Epicurus, in which he substantially followed in the footsteps of Gassendi, he was also almost equally influenced by the philosophy of Descartes. In each of them he found a commitment to mechanism as the basic explanatory concept for an understanding of nature, together with a theism which rejected a materialist account of the mind.

In his *Physiologia Epicuro-Gassendo-Charletoniana: or a Fabrick of Science Natural upon the Hypothesis of Atoms* (1654), Charleton openly follows the radical natural philosophy espoused by Gassendi in considerable detail, even though its full exposition was not available until 1658, after Gassendi's death. The *Physiologia* offers a detailed and comprehensive account of matter and its properties in terms of atoms and motion. Although close to Epicurus, it is much fuller than any of the extant writings. Contrary to the traditional teaching of the Schools, the motion of atoms in the void is argued to be the basic explanation for all the phenomena of nature and of our awareness of those phenomena through sensation. In so doing, Charleton argues for the distinction between primary and secondary qualities, made famous by Boyle and Locke later in the century, and presents accounts of light, colours, sounds and other phenomena within the atomic model. Charleton makes no claim to absolute certainty - it is the *hypothesis* of atoms that he expounds. But he makes a very good case for the mechanical account of nature from within the new learning associated with Bacon, Descartes, Gassendi and even his friend Thomas Hobbes, although he never accepted Hobbes' materialism.

Charleton had made clear his hostility to a full blown-materialism in *The Darknes of Atheism dispelled by the Light of Nature* (1652). A similar position was also presented in the later *Immortality of the Human Soul demonstrated by the Light of Nature* (1657). In these works Charleton follows Descartes rather than Gassendi, not only in his proof of God where he gives (with full acknowledgement) the causal argument offered by Descartes in the third Meditation, but also in a sharp dualism between mind and matter. The soul is a substance perfectly distinct from that of body, he claims, and it is endowed with immortality by the character of its essence. He follows Descartes, too, in holding that the idea of God is innate.

In drawing on Descartes' philosophy Charleton makes it clear that he wishes to distance himself from all those aspects of Epicurus that might be supposed to support atheism. Like Gassendi and Descartes he was keen to make his account of the new philosophy not just compatible with, but actually supportive of a Christian theology. At the same time he was quite sure that standard criticism of Epicurus was far too severe, as he made clear in the Introduction to his edition of Epicurus' moral philosophy, *Epicurus's Morals* (1655). Rather than being, as he was so often depicted, an advocate of impiety, gluttony and drunkenness, Epicurus was, Charleton urged, a great master of temperance, sobriety, continence and fortitude. Charleton is quite prepared not only to admit but to argue that Epicurus was wrong on several fundamentals. His three cardinal errors were his rejection of both the immortality of the soul and the worship of God, and his acceptance of suicide. There were, nevertheless, extenuating circumstances which went some way towards excusing Epicurus these undoubted errors, principally his living in a pagan country before the advent of Christianity. Granted the errors, Charleton, makes clear that he believes Epicurus has much to teach us about happiness and how it may be best obtained.

In all his writings Charleton is careful not to appear dogmatic. Nor was he ever an uncritical supporter of those who influenced him. This is especially true of his later writings. In his *Natural History of the Passions* (1674), for example, Charleton takes Descartes to task both for his bad anatomy (Charleton was an undoubted expert on anatomy) and for his inadequate explanation of the connection between mind and body.

Charleton died in poverty in Nantwich. His major influence was through the *Physiologia*, which introduced many English speakers to atomism, including the young Isaac Newton. His other works contributed to that assault on traditional teaching and the adoption of the mechanistic and hedonistic philosophies that came to dominate the eighteenth century.

See also: Atomism; Hedonism

G.A.J. ROGERS

List of works

Charleton, W. (1652) *The Darknes of Atheism dispelled by the Light of Nature*, London.(A defence of theism against Hobbism and other supposedly atheistic philosophies.)

Charleton, W. (1654) *Physiologia Epicuro-Gassendo-Charletoniana: or a Fabrick of Science Natural upon the Hypothesis of Atoms. Founded by Epicurus, Repaired by Petrus Gassendus, Augmented by Walter Charleton*, London.(Charleton's most important and influential work which argues for an Epicurean atomism, as compatible with traditional Christian doctrine, following Gassendi's interpretation.)

Charleton, W. (1656) *Epicurus's Morals collected partly out of his owne Greek text in Diogenes Laertius, and partly out of the Rhapsodies of Marcus Antoninus, Plutarch, Cicero, and Seneca*, London.(Argues for the rehabilitation of Epicurus as a moral thinker.)

Charleton, W. (1657) *The Immortality of the Human Soul demonstrated by the Light of Nature in two dialogues* London.(Rejects Epicurus' materialism but not his atomism.)

Charleton, W. (1659) *Natural History of Nutrition, Life and Voluntary Motion*, London.(An important attempt to give a mechanistic and mathematically based physiology.)

References and further reading

Kargon, R.H. (1964) 'Walter Charleton, Robert Boyle and the Acceptance of Epicurean Atomism in England', *Isis* 5: 184-92.(Examines Charleton's impact on English matter theory.)

Webster, C. (1967) 'The College of Physicians. ''Solomon's House'' in Commonwealth England', *Bulletin of the History of Medicine* 61: 393-412.(Highlights scientific innovation in the College of Physicians and Charleton's role in it.)

Rattner Gelbart, N. (1971) 'The Intellectual Development of Walter Charleton', *Ambix* 18: 149-68.(Important biographical paper.)

Sharp, L. (1973) 'Walter Charleton's Early Life and Relationship to Natural Philosophy in mid-Seventeenth-Century England', *Annals of Science* 30: 311-43. (Another important paper on Charleton's intellectual biography.)

Osler, M.J. (1979) 'Descartes and Charleton on Nature and God', *Journal of the History of Ideas* 40: 445-56. (Contrasts Descartes' rationalism with Gassendi's and Charleton's empiricism and its implications for their differing understandings of knowledge and of God.)

Charron, Pierre (1541-1603)

Pierre Charron was a French Catholic priest of the late sixteenth century who used Montaigne's sceptical thought, which he presented in didactic form, in order to refute Calvinists, non-Christians, and atheists. He advanced a fideistic defence of religious thought which was based on accepting complete scepticism while appealing to faith alone as the source of religious knowledge. His De la Sagesse (On Wisdom) (1601) is one of the first significant philosophical works to be written in a modern language. It is also one of the first modern works to set forth a naturalistic moral theory independent of religious considerations, and based primarily on Stoic ideas. Charron's views were extremely popular in the seventeenth century, and they influenced many sceptically inclined thinkers in France and England. His sceptical 'defence' of religion was regarded as insincere by some of the orthodox theologians, but other important religious thinkers defended him.

1 Life and works

Pierre Charron, philosopher and theologian, was born in Paris, one of twenty-five children. He studied Greek, Latin and philosophy at the Sorbonne, and then, after possibly studying law at Orleans and Bourges, went to the University of Montpellier from which he received the degree of Doctor of Law in 1571. After practising law unsuccessfully for a few years in Paris, he turned to theology and soon became renowned as a preacher and theologian. He was chosen by Queen Marguerite to be her regular preacher, and Henri IV, even before his conversion to Catholicism, often attended his sermons. He then became lecturer in divinity in various cities in the south of France, and canon and official teacher of the church of Bordeaux. Despite his great success he desired to give up worldly activities and to join a cloister, but he was turned away by two orders because of his age (forty-eight at the time). The most important event in Charron's life was his friendship with Michel de Montaigne (though doubt has been cast on its extent). After Montaigne's death in 1592, Charron demonstrated his immense debt to Montaigne by presenting the union of scepticism and Catholicism in his writings.

Charron's first work, *Les Trois Veritez (The Three Truths)* (1594), was an apologetical work of theology which attacked Calvinists in particular. His major philosophical work, *De la Sagesse (On Wisdom)*, appeared in 1601, as did his Discours chrestiens (Christian Discourses). His final work, *Le Petit Traicté de la sagesse (Little Treatise on Wisdom)*, was written in 1603 and published posthumously in 1606. All of his works were very popular, and were often republished during the seventeenth century. *De la Sagesse* was most influential in disseminating sceptical ideas and arguments in the philosophical and theological discussions of the time.

There were several serious efforts to refute, reject and even suppress Charron's sceptical views. He was attacked by the Jesuit, François Garasse, who accused him of providing a breviary for libertines, and by the Protestant medical doctor, Pierre Chanet, among others. He was defended by the *libertins érudits* Guy Patin, Gabriel Naudé, François de La Mothe Le Vayer, and Pierre Gassendi, as well as by some French leaders of the Counter-Reformation (see Libertins). Pierre Bayle considered Charron a fine exemplar of Christian fideistic thought.

2 Theological thought

In his first work, *Les Trois Veritez*, Charron sought to undermine Calvinism by sceptical means. The three truths which he considered, that God exists, that Christianity presents the correct view of God, and that Catholicism is the true presentation of Christianity, are each supported by attacking the opponents in a negative manner rather than by presenting positive reasons for accepting the truths. Although most of the enormous work is devoted to the third truth, it begins with a brief discourse about knowledge of God, in which a scepticism is developed about the possibility of human beings having knowledge in this area. Human rational capacities are so restricted and unreliable that it is doubtful that humans can know anything in either the natural or supernatural realms. Charron then joined his sceptical claims about the inadequacy and unreliability of human knowledge to the contention of the negative theologians that God is unknowable because his nature is infinite, and thus surpasses all human attempts to define or delimit it. Hence we cannot rationally know what he is, and as a result, the greatest philosophers and theologians know just as much or just as little about God as do the humblest persons. What we know is only what God is not. So, Charron said, 'the true knowledge of God is perfect ignorance about him'.

Charron's combination of negative theology and scepticism was then used to attack the claim of the atheist who denies that God exists. Such a denial, Charron argued, is presumably the consequence of a definition of God from which absurd or contradictory conclusions are drawn, but any such definition can only be the result of a human attempt to measure divinity by human means, an attempt which has no value, since atheists do not and cannot know what they are talking about.

Finally Charron argued negatively that it is unreasonable not to believe that God exists, and not to believe that Christianity is the correct statement of God and his role, and not to believe that Catholicism is the right form of Christian belief. He claimed that his opponents, especially the Calvinists of the day, argue on the basis of views derived by means of weak and miserable human capacities, and then use these unreliable results to measure divine truths.

3 Sceptical philosophy

Charron's sceptical defence of faith was elaborated in *De la Sagesse* and *Le Petit Traicté de la sagesse*. He claimed that since human beings can only discover truths by means of revelation, morality should be based on following nature except when it is guided by Divine Light. This view was first developed by setting forth Montaigne's scepticism in an organized didactic form. We must begin by knowing ourselves, which involves knowing the limits of what we can in fact know. Charron presents the traditional sceptical doubts about sense knowledge. Do we have the requisite sensory capacities to gain genuine knowledge? Are we able to distinguish illusions and dreams from veridical experiences? And can we ascertain which of our many varied sense experiences actually correspond to some objective state of affairs?

Charron then turned to questions that can be raised about our rational capabilities. He contended that we possess no certain or adequate criterion for distinguishing truth and falsehood. We in fact believe what we do as the result of our passions and of social pressures on us, rather than on the basis of reasons and evidence. We actually function as beasts do, rather than as rational beings. Hence we should accept Montaigne's claim that we possess no genuine principles unless God has revealed them to us. Everything else is just dreams, smoke, or illusion.

In the second book of *De la Sagesse* Charron set forth a discourse on method as a way of avoiding error and discovering truth in the light of the human situation. The Charronian method anticipates that of Descartes. First, one should examine each question freely and dispassionately. Next, one should keep prejudice and emotion out of all decisions. After that, one should develop a universality of mind. Finally, one should reject any decisions which are doubtful in even the slightest degree. This sceptical method or method of doubt was, Charron declared, of greater service to religion than any other possible method. If we followed it, we would reject all doubtful opinions until our minds became 'blank, naked and ready' to receive divine revelation on the basis of faith alone. So, Charron insisted, the complete sceptic will never be a heretic, for the sceptic, having no opinions, cannot have the wrong ones. On the other hand, if it pleases God to give the sceptic any information, he will possess true knowledge.

Until the sceptic receives any revelation, he ought to live according to a provisional morality. This is described in the third and final book of *De la Sagesse*, which sets forth Charron's theory of natural morality based on ancient sceptical and Stoic views. Without divine guidance the best one can do is live naturally as a Stoic sage or a noble savage. Here Charron is one of the first modern philosophers to set forth a moral theory independent of religious considerations.

At the time, some readers saw Charron's work as a basic didactic presentation of Pyrrhonian scepticism that challenged traditional philosophical claims to knowledge as well as religious views. By so doing, these readers believed, Charron was preparing the basis for a completely naturalistic view of human nature and conduct. However, Charron averred that *De la Sagesse* contained only part of his theory, namely that which deals with the human situation seen completely apart from divine guidance. When one considers the views in his different works taken as a whole, together with his ecclesiastical career and the religious sentiments that appear in his Discours chrestiens, one can find evidence for the case that Charron was a sincere fideist who propounded scepticism as a way of destroying those whom he saw as the enemies of the true faith, while preparing others for salvation by God's actions. Whatever his true views, Charron greatly influenced the so-called *libertins*, and his ideas, as interpreted by them, provided some of the basis for irreligious naturalism in the Enlightenment.

See also: Scepticism, Renaissance; Stoicism

RICHARD H. POPKIN

List of works

Charron, P. (1594-1603) *Oeuvres (Works)*, Paris, 1635; repr. Geneva: Slatkine, 1970.(The only edition of Charron's works.)

Charron, P. (1594) *Les Trois Veritez (The Three Truths)*, in *Oeuvres*, Geneva: Slatkine, 1970.(Uses sceptical arguments to attack Calvinist beliefs.)

Charron, P. (1601) *De la Sagesse (On Wisdom)*, ed. A. Duval, Paris: Chassériau, 1824, 3 vols; repr. Geneva: Slatkine, 1968.(An important source of sceptical ideas for Renaissance theologians and philosophers. For a list of other editions and translations, see Charron 1961.)

Charron, P. (1606) *Le Petit Traicté de la sagesse (Little Treatise on Wisdom)*, in *Oeuvres*, Geneva: Slatkine, 1970.(Presents a sceptical defence of faith; written in 1603, the treatise was first published posthumously in 1606.)

References and further reading

Busson, H. (1933) *La pensée religieuse française de Charron à Pascal (French Religious Thought from Charron to Pascal)*, Paris: Vrin.(Discusses the irreligious consequences of Charron's thought in seventeenth-century France.)

Charron, J.D. (1961) *The 'Wisdom' of Pierre Charron, an Original and Orthodox Code of Morality*, University of North Carolina Studies in the Romance Languages and Literatures 34, Chapel Hill, NC: University of North Carolina.(Includes a list of editions and translations of *De la Sagesse* on pages 147-51.)

Horowitz, M.C. (1971) 'Pierre Charron's View of the Source of Wisdom', *Journal of the History of Philosophy* 9: 443-57.(Emphasizes the Stoic side of Charron's thought.)

Popkin, R.H. (1954) 'Charron and Descartes: the Fruits of Systematic Doubt', *Journal of Philosophy* 51: 831-37. (Shows the similarities and differences between Charron's and Descartes' methods of doubt.)

Popkin, R.H. (1979) *The History of Scepticism from Erasmus to Spinoza*, Berkeley, CA: University of California Press.(Places Charron in the development of Renaissance scepticism.)

Rice, E.F. (1958) *The Renaissance Idea of Wisdom*, Cambridge, MA: Harvard University Press.(An examination of how the notion of wisdom changed during the Renaissance, culminating in Charron's naturalistic, Stoic view.)

Sabrié, J.B. (1913) *De l'humanisme au rationalisme: Pierre Charron (1541-1603), l'homme, l'oeuvre, l'influence (From Humanism to Rationalism: Pierre Charron, the Man, his Work and its Influence)*, Paris: Alcan.(Emphasizes the irreligious results of Charron's scepticism.)

Soman, A. (1970) 'Pierre Charron: A Revaluation', *Bibliothèque d'Humanisme et Renaissance* 32: 57-79. (Presents evidence that Charron was not a genuine friend and disciple of Montaigne.)

Chartres, School of

In the first half of the twelfth century, the most advanced work in teaching and discussion of logic, philosophy and theology took place in the schools attached to the great cathedrals. Chartres was undoubtedly one of the more important of these schools, and Gilbert of Poitiers and Thierry of Chartres were certainly connected with it. To some historians, Chartres was the *great intellectual centre of the period, and the greatest achievement of early twelfth-century thought was a brand of Platonism distinctive of this school. However, this view has been challenged by scholars who stress the pre-eminence of Paris, where the schools emphasized logic.*

1 Who taught at Chartres?

When Clerval (1895) first claimed the particular importance of early twelfth-century Chartres as an intellectual centre, he believed that he could point to four outstanding figures who had been schoolmasters there. First, there was Bernard, described as teaching at Chartres in John of Salisbury's *Metalogicon* (see John of Salisbury); Clerval identified him with Bernard Silvestris (Bernard of Tours), author of the *Cosmographia*, a prosimetrum recounting the creation of the world and of man in terms closely reminiscent of Plato's *Timaeus*. Second was the polymath Thierry of Chartres and third was Gilbert of Poitiers, one of the leading theologians of the mid-twelfth century. Finally, most scholars of Clerval's time believed that William of Conches, the grammarian and commentator on Platonic texts, was also a teacher at Chartres, although more recent research has cast doubt on various aspects of this claim.

The Bernard described in the *Metalogicon* - who was probably born *circa* 1060, and is mentioned in charters as master of the school at Chartres between 1114 and 1119 and as Chancellor between 1119 and 1124 - is no longer identified with Bernard Silvestris, who taught at Tours in the mid-twelfth century and whose only link with Chartres was his friendship with Thierry. The Chartrian Bernard was a grammarian, interested in Priscian and the theory of grammar as well as in expounding classical Latin tests. Although a commentary on Plato's *Timaeus* has been attributed to him with some plausibility, he was not, like the author of the *Cosmographia*, one of the great writers of the time.

The distinction between Bernard of Chartres and Bernard Silvestris is now accepted by all historians. The relation of other leading teachers to Chartres remains controversial: Sir Richard Southern (1970) has been the leading sceptic, while Edouard Jeauneau (1973), Peter Dronke (1969) and Nikolaus Häring (1974) have argued for the importance of Chartres. No one queries that Thierry ('the Breton' or 'of Chartres') was linked with Chartres, where he became Chancellor in about 1142. However, it is questionable how much teaching he did there; he is also known to have taught John of Salisbury on the Mont Sainte Geneviève in Paris in 1136-8. Where was his teaching predominantly based? A similar question applies to Gilbert of Poitiers, who was a canon of Chartres in 1124 and chancellor there from 1126 to 1137. Gilbert certainly taught in Paris, and one student mentions attending his lectures both at Paris among an audience of three hundred, and at Chartres in an audience of only four. Does this mean that Chartres was a scholastic backwater, or rather (as Dronke has argued) that Gilbert reserved his most advanced teaching for Chartres?

The connection of William of Conches with Chartres is even more problematic. He very probably studied at Chartres under Bernard and is regarded by John of Salisbury as a continuer of Bernard's approach. Jeauneau has found a few remarks in his work which might hint that he was himself teaching at Chartres, and an unclear and much discussed passage in John of Salisbury's *Metalogicon* (II.10) can be argued to support this surmise. However, this hardly constitutes solid evidence for placing him at Chartres, and there is much to recommend Southern's final view, that William taught neither at Paris nor at Chartres. In sum, Chartres may well have been a more important scholastic centre in the first half of the twelfth century than Southern allows, but there is no denying the pre-eminence of Paris in almost all branches of learning from about 1110 onwards.

2 'Chartrian Platonism'

The dispute about the school of Chartres has gone beyond the merely factual questions of who taught where, to become a debate about the very nature of twelfth-century thought. Many modern supporters of the school of Chartres hold that, in Dronke's words, it stands for 'what is freshest in thought, richest and most adventurous in

learning, in northern Europe, in the earlier twelfth century' (Dronke 1969: 117), and some have suggested that 'Chartres' should be understood less as the name of a place where certain masters taught and more as the label for a distinctive brand of Platonism which was followed by masters such as Thierry, William of Conches, Gilbert of Poitiers and Bernard Silvestris.

Southern disputes this. In his first attack on the idea of a school of Chartres, he wrote of Bernard (of Chartres), Thierry and William of Conches that 'all their thoughts were old thoughts' (Southern 1970: 83). More recently, he has refined his criticisms. The Platonism of the masters who are described as 'Chartrian' was, he suggests, a very scholastic Platonism. It did not involve adopting, knowingly or instinctively, Plato's underlying positions or attitudes, nor some special 'poetic Platonism', but centred rather on the close scrutiny of the one work of Plato's they know, the *Timaeus* in Calcidius' partial translation. The 'Chartrian' scholars, he says (Southern 1979: 10), regarded Plato as an authority on 'the primitive organization of the elements of the universe' and used him not to reach a 'philosophical Platonism' but as 'a contribution to the vast jig-saw puzzle of universal knowledge about the origin of the world'. Such an authority was important as a doorway, but was quickly left behind when more scientific knowledge became available, especially from Arab sources. The 'Chartrian' scholars are better understood, he urges, 'by being freed from the school of Chartres and placed in the wider setting of a common scholastic enterprise' (Southern 1979: 40).

Powerful though it is, Southern's argument has some serious weaknesses. It overlooks the close link, in the work of Bernard of Chartres, William of Conches and Bernard Silvestris, between the study of grammar (which included both the interpretation of classical texts and the theory of grammar, based on Priscian) and what we would now call science. The breaking of this link in the course of the twelfth century should not be seen straightforwardly as scientific progress, but as a fundamental shift in assumptions and methods. Moreover, Southern fails to bring out an important peculiarity shared by at least three of these masters. Unlike most of the leading teachers of their time, Bernard of Chartres, William of Conches and Bernard Silvestris apparently remained arts masters and did not venture into sacred doctrine at all. Further, none of them seems to have been very interested in logic, the area to which (apart from theology) most twelfth-century masters gave the greatest attention (see Logic, medieval). Thierry fits this pattern only partly, since he did write about the Trinity and the creation and seems to have been interested in logic. Gilbert of Poitiers fits it not at all, conforming rather to the more usual twelfth-century mould of the logician who goes on to write theology. Perhaps what the enthusiasts for the school of Chartres have hit upon, half unconsciously, is a group of twelfth-century masters (working in various centres) who rejected the pattern and priorities of intellectual life which were becoming the norm and who, in their lifelong dedication to the arts as opposed to theology, might be seen as forerunners of late thirteenth-century arts masters such as Siger of Brabant and Boethius of Dacia.

See also: Bernard of Tours; Clarembald of Arras; Gilbert of Poitiers; Platonism, medieval; John of Salisbury; Thierry of Chartres; William of Conches

JOHN MARENBON

References and further reading

Bernard Silvestris (early 1150s) *Cosmographia (Cosmography)*, ed. P. Dronke, Leiden: Brill, 1978; trans. W. Wetherbee, Ithaca, NY: Cornell University Press, 1973.(Describes the origins of the universe and man, strongly influenced by Plato's *Timaeus*.)

Clerval, A. (1895) *Les écoles de Chartres au moyen-âge (The Schools of Chartres in the Middle Ages)*, Paris: Picard.(Fundamental, though in some respects dated, study of the historical evidence for teaching at Chartres.)

Dronke, P. (1969) 'New Approaches to the School of Chartres', *Anuario de estudios medievales* 6: 117-40; repr. with new bibliographical postscript in P. Dronke, *Intellectuals and Poets in Medieval Europe*, Rome: edizioni di storia e letteratura (Raccolta di studi e testi 183), 15-40.(Spirited defence of the School of Chartres against Southern.)

Dutton, P.E. (1984) 'The Uncovering of the *Glosae super Platonem* of Bernard of Chartres', *Mediaeval Studies* 46: 192-221.(Dutton argues that a set of glosses to the *Timaeus* - which he has now edited in full - are the work of Bernard of Chartres).

Häring, N. (1974) 'Chartres and Paris Revisited', in J.R. O'Donnell (ed.) *Essays in Honour of Anton Charles Pegis*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 268-329.(Arguments against Southern.)

Jeauneau, E. (1973) *Lectio philosophorum*, Amsterdam: Hakkert.(Collects Jeauneau's detailed and important

essays on masters connected with Chartres.)
John of Salisbury (1159) *Metalogicon*, ed.J.B. Hall, Corpus Christianorum, Continuatio Mediaeualis 98, Turnhout: Brepols, 1991; trans. D.D. McGarry, Berkeley, CA: University of California Press, 1955.(Defence of the study of ancient classics; important source for evidence about twelfth-century education.)
Southern, R.W. (1970) 'Humanism and the School of Chartres', in R.W. Southern, *Medieval Humanism and Other Studies*, Oxford: Blackwell, 61-85.(Southern's original attack on the notion of a School of Chartres.)
Southern, R.W. (1979) *Platonism, Scholastic Method and the School of Chartres*, Reading: University of Reading.(Describes Chartrian Platonism as scholastic rather than distinctive.)
Southern, R.W. (1982) 'The Schools of Paris and the School of Chartres', in R.L. Benson and G. Constable (eds) *Renaissance and Renewal in the Twelfth Century*, Oxford: Oxford University Press, 113-37.(Historical study, stressing the greater importance of Paris over Chartres.)
Wetherbee, W. (1972) *Platonism and Poetry in the Twelfth Century: The Literary Influence of the School of Chartres*, Princeton, NJ: Princeton University Press.(Urges an understanding of 'school of Chartres' as the label for a type of thought.)

Chatton, Walter (*c.*1290-1343)

Chatton was an English philosopher and theologian who developed a detailed critique of the work of William of Ockham, causing the latter to revise some of his earlier writings. Chatton was also at times an opponent of Peter Aureol and Richard of Campsall; he generally, though not always, followed John Duns Scotus and responded to his critics. He is known also for his writings on physics, where he held views in line with those of Pythagoras and Plato, and on the Trinity, where he was strongly attacked by Adam Wodeham.

The English Franciscan philosopher and theologian Walter Chatton was born in the village of Catton, near Durham, around 1290. He was a contemporary of William of Ockham and Adam Wodeham at the Franciscan custodial school in London from 1321 to 1323. There he delivered his *Reportatio* lectures on all four Books of the *Sentences* of Peter Lombard, in preparation for his later *Lectura* on the *Sentences* at Oxford sometime between 1324 and 1330 (most likely in 1328-30). He was one of the examiners of the works of Durandus of Saint-Pourçain and Thomas Waleys at the papal court in Avignon, and is believed to have died there in 1343.

Chatton was such a detailed critic of Ockham that it is impossible to follow many of the arguments in the latter's *Quodlibets* without having Chatton's *Reportatio* at hand. Ockham's famous razor was sharpened partly in response to Chatton's critique. Its early formulation, 'Beings should not be multiplied without necessity', was challenged by Chatton, who counters with an anti-razor that he calls 'my proposition': 'When a proposition is made true by things, if two things are not sufficient for its truth, then it is necessary to posit a third, and so on.' In response, Ockham in his later works reformulates his razor to say: 'When a proposition is made true by things, if two things are sufficient for its truth, then it is superfluous to posit a third, and so on.'

Chatton also attacked Ockham's view of concepts, at least the view Ockham held in the first redaction of his *Sentence* commentary, where he defends the *fictum* theory, which holds that the concept does not exist in the mind as in a subject but only has the reality of an object created by an act of understanding. Chatton himself held the *intellectio* theory, which contends that the concept is nothing else but the very act of knowing. Such an act of knowing is a true quality existing in the soul as in a subject, and is also a natural sign of the object that is immediately understood by means of it. Ockham not only did not despise Chatton's critique of the *fictum* theory, he incorporated his opponent's *intellectio* theory into his later treatments of concepts, at first reducing the *fictum* theory to a 'less probable' opinion and finally abandoning it.

The Ockham-Chatton exchange, however, is not limited to the period when both were in London. In discussing the dependence of a second cause on the first cause in essentially ordered causes, Chatton in his *Lectura* defends Duns Scotus' position against Ockham's challenges, quoting verbatim from Ockham's last philosophical work, the *Quaestiones in libros Physicorum Aristotelis(Questions on Aristotle's Physics)*, probably written at Oxford about 1324.

In dealing with his own questions concerning Aristotle's *Physics*, Chatton is known especially for joining the early fourteenth-century minority including Henry of Harclay, Gerard of Odo and Nicholas Bonet, thinkers who opposed Aristotle's claim that continua cannot be composed of indivisibles. Although Chatton had contemporary allies among the atomists, he seems to be alone in holding that continua are composed of finite numbers of indivisibles. Thomas Bradwardine, in his *Tractatus de continuo (Treatise on the Nature of a Continuum)*, makes Chatton a follower of Pythagoras and Plato, who held the same position (see Aristotelianism, medieval; Platonism, medieval).

Besides Ockham, Chatton had a number of other debating partners. When Peter Aureol attacked Duns Scotus' theory of the univocity of being, Chatton came to its defence. Aureol attacked Duns Scotus for claiming that we can have a univocal concept of being, a concept that is predicable in the same sense both of God and of creatures. Scotus achieved this univocal concept at a price, since his concept of being leaves outside its ambit the modes 'infinite' and 'finite' which, if they were included, would impede 'being' from being predicable both of God and of creatures. Chatton grants this objection, but considers it irrelevant. If Aureol, he argues, wants to include modes and differences in his concept of being, then 'being' becomes a most general concept of all that is opposed to nothing, and this is merely a logical and not a metaphysical concept. It is the latter, according to Chatton, that

Scotus had in mind (see Duns Scotus, J.).

Richard Campsall became Chatton's opponent when he argued that intuitive and abstractive cognition are not really distinct, 'since numerically the same knowledge is intuitive when the object is present and abstractive when it is absent, because plurality should not be admitted without necessity.' Against this position, Chatton raised twelve difficulties and then refuted Campsall by appealing to his anti-razor, arguing that it is not impossible that God conserve in existence the intellect with its abstractive cognition and make the object present without the intellect grasping it as present. Thus for the proposition 'He sees that object' to be true, it is not enough to have the intellect, its abstractive cognition and the object present. A distinct thing has to be added: intuitive cognition.

Chatton clashed with Campsall also over the logic involved in statements of non-identity related to the Christian teaching on the Trinity. Chatton's own treatment of the Trinity had its logical and metaphysical problems, turning the divine essence into a collection of persons. He was severely ridiculed and criticized for such Trinitarian views by Adam Wodeham, who quite likely was the student who wrote down Chatton's *Reportatio*. In the margin of Chatton's text, Wodeham wrote: 'In all this discussion the report is not in accord with the mind of the speaker. Nor is there any wonder, since when the author said these things he was not quite sane. Later on, he thought things out better and had another go at it. And then the reporter naturally expressed things in a better way' (*Lectura secunda* I, 13*, n.28). Wodeham was Chatton's chief critic, often accusing him of misunderstanding or misrepresenting Ockham, or of accepting Ockham's views but pretending, by petty quibbles, that he was differing. Despite Wodeham's frequent attacks on him, however, the influence Chatton had on both Ockham and Wodeham shows his philosophical importance.

See also: Bradwardine, T.; Henry of Harclay; William of Ockham; Wodeham, A.

STEPHEN F. BROWN

List of works

Chatton, Walter (*c.*1321-30) *Reportatio et Lectura super Sententias: Collatio ad librum primum et prologus (Commentary on Peter Lombard's Sentences)*, ed. J.C. Wey, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1989.(This volume contains the introductory sermon and the prologue, which in effect is the prologue for both the *Reportatio* (1321-1323) and the *Lectura* (1324-1330). This work replaces earlier editions of separate prologue questions, but some earlier editions also contained excellent analyses and valuable studies; these are noted in the list of further readings, below. Other parts of the *Reportatio* and *Lectura* that have been edited can be found listed in Wey's introduction, 3-4, n. 10. Not included in Wey's list is *Lectura in I Sent.*, d. 4, q.1, aa.1-2, ed. S.F. Brown, *Medieval Philosophy and Theology* 3, 1993, 135-57, which discusses supposition theory.)

Chatton, Walter (*c.*1321-30) *Quaestio utrum quantum et continuum componantur ex indivisibilibus sicut ex partibus integrantibus*, ed. J.E. Murdoch and E.A. Synan, *Franciscan Studies* 26, 1966, 234-66. (Contains Chatton's writings on the continuum; attributed to Chatton with some reservations (see Murdoch and Synan 1966).)

References and further reading

Adams, M.M. (1982) 'Universals in the Early Fourteenth Century', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 411-39. (Includes an outline of Chatton's criticisms of William of Ockham, and the latter's response.)

Brown, S.F. (1985) 'Walter Chatton's *Lectura* and William of Ockham's *Quaestiones in libros Physicorum Aristotelis*', in W.A. Frank and G.J. Etzkorn (eds) *Essays Honoring Allan B. Wolter*, St Bonaventure, NY: Franciscan Institute, 81-93.(Shows the interrelationship of Chatton's *Lectura* and Ockham's *Quaestiones*.)

Campsall, Richard (*c.*1350-60), cited in Chatton, *Reportatio et Lectura super Sententias: Collatio ad librum primum et prologus (Commentary on Peter Lombard's Sentences)*, ed. J.C. Wey, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1989, prol., q.2, a.1, 77, n.3.(Campsall's argument, refuted by Chatton.)

Courtenay, W.J. (1978) *Adam Wodeham: An Introduction to his Life and Writings*, Leiden: Brill.(Provides Chatton's *curriculum vitae*.)

Cova, L. (1973) *Walter Catton, Commento alle Sentenze: Prologo, Questione Terza*, Rome: Edizione dell'Ateneo. (Analysis of the question as to whether we can have many concepts proper to God).

Etzkorn, G.J. (1987) 'A Heretofore Unknown *Quodlibet* of Walter Chatton', *Bulletin de philosophie médiévale* 29: 230.(Cod. Paris. Nat. lat. 15, 805 contains a Chatton *Quodlibet* that ends in the middle of Question 29.)

Fitzpatrick, N.A. (1971) 'Walter Chatton on the Univocity of Being: A Reaction to Peter Aureoli and William of Ockham', *Franciscan Studies* 31: 88-177.(Defends Scotus' teaching on univocity against Aureol and Ockham.)

Knudsen, C. (1982) 'Intentions and Impositions', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 479-95.(Includes Chatton's criticisms of Ockham's work on intentions and impositions.)

Knudsen, C. (1985) 'Chatton contra Ockham über Gegenstand und Einheit von Wissenschaft und Theologie' (Chatton's Attack on Ockham Concerning the Object and Unity of Science and Theology), *Cahiers de l'Institut du moyen-âge grec et latin* 50: 3-112.(On the unity of the science of theology).

Maurer, A. (1984) 'Ockham's Razor and Chatton's Anti-Razor', *Mediaeval Studies* 46: 463-75.(Presents Ockham's reformulation of his famous razor in response to the criticism of Chatton.)

Murdoch, J.E. (1982) 'Infinity and Continuity', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 564-94.(Contains a brief account of Chatton's views on indivisibles.)

Murdoch, J.E. and Synan, E.A. (1966) 'Two Questions of the Continuum: Walter Chatton (?) and Adam Wodeham, OFM', *Franciscan Studies* 26: 212-88.(Edition of the text concerning the nature of the continuum attributed to Chatton.)

Wodeham, Adam (1990) *Lectura Secunda*, ed. R Wood and G Gál, St Bonaventure, NY: St Bonaventure University, vol. I, 12*-18*.(Contains Wodeham's evaluation of Chatton.)

Chemistry, philosophical aspects of

Chemistry, like all theoretical sciences, is deeply rooted in philosophical inquiry. Early Greek atomism was a response to Parmenides' argument that the very concept of change is unintelligible. Aristotle in turn argued that a vacuum is impossible and proposed that qualitative change could be better understood in terms of four elements and an underlying prime matter.

During the Arabic and Latin Middle Ages, philosophical commentaries on the nature of materials were brought into juxtaposition with the practical arts of the alchemist, miner and pharmacist. As chemical speculations became more closely connected to observations during the time of the Scientific Revolution, natural philosophers became more and more interested in the methodological aspects of chemistry. Galileo and Locke tried to clarify the relationship between primary and secondary qualities. Boyle struggled to understand how the selective affinities so characteristic of chemical reagents could be explained within the framework of Descartes' mechanical philosophy. Lavoisier's textbook was organized around principles drawn from philosophes *such as Condillac.*

As chemistry became an autonomous science, chemists turned less often to philosophy as a source of theoretical inspiration. However, they frequently appealed to philosophies of science in order to defend their own theories or criticize those of their opponents. The so-called 'atomic debates' amongst chemists in the British Association during the 1860s were primarily disputes about the epistemological legitimacy of appeals to unobservable entities. Many of the same issues were taken up at the end of the century by Ostwald, Mach and Duhem.

Logical empiricists in the twentieth century have often turned to chemistry as an example of the successful reduction of a secondary science to physics. Thus, it is claimed that the laws of classical thermodynamics, expressed using concepts such as enthalpy and entropy, can be derived from the laws of statistical mechanics describing the motions of particles. The details of this purported derivation have been questioned. Even more controversial is the attempt to derive accounts of the chemical bond and the shape of molecules from the first principles provided by quantum mechanics.

1 Early concepts of atoms and elements

To a modern student of chemistry there seems to be no tension between the concepts of atom and element. The periodic table of the elements illustrates how the chemical properties of a simple substance such as sulphur can be related to the physical properties of the atoms of which it is composed, especially its atomic number. Historically, however, there has often been conflict between atomic approaches and research programmes that talk about elements. Examples of the opposition already appear in pre-Socratic cosmological theories about the origins and composition of the world. The typical features of classical atomic theories are present in the fragmentary accounts we have of Leucippus and Democritus (fifth century BC): the world is composed of atoms moving in a void (see Atomism, ancient). The atoms themselves are unchanging, indivisible bits of matter which differ from each other only in size or shape. The processes we observe around us are the result of changing groupings or arrangements of atoms; the qualitative aspects of experience are the results of their collisions with the atoms of our soul or mind. However, Democritus is very explicit that we cannot expect to have knowledge of qualities.

> By convention are sweet and bitter, hot and cold, by convention is colour; in truth are atoms and the void. …There are two forms of knowledge, one genuine, one obscure. To the obscure belong all of the following: sight, hearing, smell, taste, touch.
>
> (Kirk and Raven 1962: 422)

The atomistic tradition was continued by Plato who, in the *Timaeus*, gave an elaborate geometrical account of atoms in terms of the five regular polyhedra (see Plato §16). In *The Nature of Things*, Lucretius provided speculative corpuscular explanations of everything from evaporation to sexual attraction, but in part because of its historical connections to Epicurean moral philosophy, atomistic theorizing became theologically suspect and was not further developed until the rise of modern science (see Epicureanism §§3, 12).

The dominant theory of matter was Aristotle's sophisticated account of the four elements - earth, air, fire and water (an account earlier developed by Empedocles). Aristotle posited an underlying substratum of prime matter that could neither be created nor destroyed. On it were impressed various combinations of the four fundamental

qualities, the hot or the cold, the wet or the dry. The four elements could be transmuted into each other. For example, by adding more warmth to water (formed from the cold and the wet) one could change it into air (characterized as the hot and the wet) (see Matter §1).

In trying to explain the multiplicity of properties which we actually observe, Aristotle associated additional qualities with each element. Thus softness and ductility and a bland taste are characteristic of water. Brittle materials are earthy, while sour or spicy foods contain fire. Not only do ordinary materials contain varying proportions of the elements, they also differ in how intimately the parts are combined.

Although there are places where Aristotle seems to presuppose a sort of conservation law for his qualities, there are also passages in which he says that the more potent properties can completely transform the weaker. Thus if a great quantity of water is added to wine, in the end the water will actually convert the wine into water. In a similar way, Aristotle says, living things can transform food into flesh and bone (*De generatione et corruptione* I 5). The Aristotelian system thus provided the philosophical foundation for alchemy (see Alchemy).

2 Corpuscles, forces and conservation

Greek atomism had postulated that all processes could be understood as the result of collisions between, and regroupings of, invisible particles, but it had very little to say about how or why the atoms moved or what caused them to cohere into relatively stable clusters. Both of these deficiencies were addressed by the theories of motion developed in the seventeenth century. Although Descartes rejected the possibility of a vacuum, his programme of explaining all natural phenomena in terms of corpuscles moving in accordance with laws of inertia and collision was otherwise a continuation of the classical approach. Like Epicurus and Lucretius, Descartes wrestled with phenomena - colour, patterns of iron filings round a bar magnet, and physiology - which seemed particularly recalcitrant to mechanical explanations (see Descartes, R. §11).

Isaac Newton not only corrected the Cartesian laws of motion, he also revolutionized its ontology. The universal law of gravitation plus his three laws of motion gave a complete account of the movements of both heavenly bodies and middle-sized terrestrial objects such as projectiles. Newton was convinced that additional forces, each with a corresponding force law, would be needed to explain behaviour at the atomic level. Thus, he postulated a repulsive force varying inversely with the distance between nearest neighbours to explain the ’spring’ in air which makes it resist compression. And in the famous thirty-first query of the *Opticks* Newton suggests that a variety of cohesive forces acting at short range may be needed to account for the ’sociability’ of chemicals for each other. John Locke tried to give a philosophical account of how the world of sense experience with its rich diversity could be reduced to the ’solid, massy, hard, impenetrable, moveable particles’ proposed by Newton which varied only in size, figure and velocity (see Locke, J. §4; Primary-secondary distinction §1).

Even if the relationship between primary and secondary qualities could be clarified to the satisfaction of philosophers, there still remained the difficulty of how working chemists, who dealt with what British schoolchildren came to know as the science of ‘bangs and stinks’, could relate atomic models to the practical problems of describing the properties and reactions of materials in the laboratory. And as chemists made progress in classifying materials according to their reactivities into, for example, acids, bases and salts or metals, calces and reducing agents, they found it more useful to think in terms of underlying elements or principles characterized in terms of their qualitative chemical natures, instead of the geometrical properties of corpuscles. The chemists working in this tradition, while not denying the possibility of underlying atoms, nevertheless ignored them when theorizing about chemical reactions.

Aristotle had accepted the Parmenidean axiom that a persisting underlying reality was a prerequisite for understanding change, and in his system prime matter is conserved. However, most of the things chemists were interested in, such as the elements themselves, were not conserved on Aristotle’s account. By the eighteenth century there was a concerted effort to find a small fundamental number of chemical elements or principles which were conserved.

For the atomists, of course, conservation was built into the system - atoms are by definition invariant - and they prided themselves on the clarity and parsimony of their list of primary properties, size, figure, and so on. But atomists also had no way to limit or predict how many different types of atom there should be.

Those who postulated elements defined in terms of qualities permitted themselves a richer set of basic properties, but placed much stricter limits on the number of different kinds of element. Stahl, for example, retained a version of the four traditional elements but added to them the three ’earths’ proposed by Becher: the saline earth (the principle of acidity), the sulphurous earth or phlogiston (the principle of inflammability) and the mercurial earth. Although originally called ’earths’, these chemical principles were characterized by their great activity and, as became evident in later variants of the phlogiston theory when phlogiston was assigned a zero or even negative weight, their material nature was sometimes in doubt. But Stahl was adamant that the principles were conserved.

In his treatises on both the sulphurous and saline principles, Stahl extolled what he calls the joint method of analysis and synthesis. The standard chemical practice of determining the composition of a complex material by analysing it into its simpler constituents is fallible because either some of the analysands may be lost or something may be inadvertently added. Stahl recommends that chemists should ’complete the proof’ by resynthesizing the material in question. Thus he believed that he had proved that sulphur contained phlogiston because he could prepare it using charcoal (a major source of phlogiston). It is an irony that Stahl’s system was eventually rejected on the basis of Lavoisier’s famous experiments in which the red calx of mercury was first analysed, thereby forming oxygen gas, and then resynthesized from the gas and metallic mercury.

Lavoisier’s textbook of 1789 presents a list of dozens of elements, most of which are familiar to us today, and emphasized that not only mass but also the quantity of each individual element is conserved in all reactions. Lavoisier also gives the concept of element an operational definition - we are to take as elements those simple substances which cannot be further broken down in the laboratory. And, unlike Stahl, Lavoisier stressed the importance of tracking what happens to elements in reactions by weighing products and reactants. However, Lavoisier was not entirely consistent in his approach: two imponderable fluids which cannot be isolated, light and caloric, are the first elements on his list.

3 Positivism and nineteenth-century atomism

In 1808 Dalton provided a link between the atomistic research programme and Lavoisier’s descriptive chemistry. Each material element on Lavoisier’s list consisted of atoms of a characteristic weight surrounded by caloric. The relative weights of atoms could be calculated from empirically determined combining weights. At last there was a direct way to specify one of the primary properties attributed to atoms. Over the next century a great deal of experimental work went into determinations of atomic weights and molecular formulae.

However, Dalton’s atomic theory provided no conceptual resources for explaining what held atoms together when they formed compounds. Unlike earlier ’hooks and eyes’ speculations, Dalton pictured all atoms as spherical. And their effective volumes seemed to reflect how much caloric surrounded them rather than their intrinsic sizes. Caloric itself caused atoms to move apart, thus only heightening the mystery of the nature of chemical bonding.

Boscovich (1763) had developed Newtonian matter theory by drawing speculative graphs which showed alternating attractive and repulsive forces as a function of distance from the nucleus (see Mechanics, classical), but no one had any idea of how to relate these diagrams to chemical reactivities. Most puzzling was the fact that the reactivity of a material depends on its ‘sociability’ or compatibility with other reactants. Thus, as the alchemists were very aware, although gold is in general less reactive than silver, aqua regia will dissolve gold much more readily than it will silver. Aqua fortis, on the other hand attacks silver but not gold. Furthermore, the extent to which chemicals ’elected’ to combine was sometimes a function of temperature or concentrations. It was difficult to fathom how Boscovichean forces could depend on such variables. This was one major reason why Kant concluded that chemistry would never be a theoretical science founded on universal mathematical laws. But after the discovery of electrochemical phenomena, philosophically minded chemists hoped that chemical affinities could be explained in terms of electrical attractions. After all, in voltaic cells electricity was produced from chemical reactions while in electrolytic cells electricity could be used to decompose complex bodies back into their elements. Berzelius’ ’dualistic theory’ proposed that all bonds were formed between electropositive and electronegative units. But this theory could not cope at all with the fact that the common atmospheric gases such as nitrogen and oxygen were diatomic and it was of little help in understanding the ever-growing number of organic compounds.

It is tempting to view nineteenth-century chemistry as the progressive articulation of atomic theory. To the laws of

definite and multiple proportions were added Avogadro's hypothesis that equal volumes of gases contain equal numbers of particles and Kekulé's discovery that organic compounds could be represented as chains or rings of tetravalent carbon atoms. It even became possible to assign definite geometric structures to molecules.

Yet there was on-going scientific and philosophical resistance to the atomistic research programme. Some of it was a result of general methodological caution. Chemists were still very aware of the phlogiston debacle which John Herschel described as ' a lamentable epoch in which…as if to prove the perversity of the human mind, of two possible roads, the wrong was chosen' (1830: 300). The so-called 'atomic debates' at the British Chemical Society in 1869 illustrate the concern of chemists to avoid a naïve belief in hypothetical entities that were admittedly useful but not directly tied to experiment.

Other objections stemmed from specific conceptual lacunae or flaws within the contemporary theory. For example, if one took seriously the rigid geometrical structure attributed to carbon, how could one account for the so-called 'Walden inversion' in which the molecule apparently turned wrong-side out? Why should atoms have discrete valences or combining powers in the first place, such that nitrogen hooked up with three hydrogens and oxygen combined with two, while chlorine was satisfied with only one? And how could the family relationships and trends among the elements, so nicely summarized in Mendeleev's periodic table, ever be explained in terms of periodicities or trends within the numerical values of atomic weights?

By the end of the century the most vigorous opponents of atomism, Ostwald, Mach and Duhem, could offer an alternative research programme, that of classical thermodynamics. This approach appeared to be much more closely tied to measurable quantities such as temperature and heats of reaction and its treatment of chemical processes required no assumptions about either mechanisms or the microstructure of the reactants. Mach and Duhem both had strong philosophical reasons for favouring phenomenological theories and Ostwald developed an entire metaphysical system based on 'energetics'.

Thermodynamics did indeed give valuable insights into reaction equilibria and the stability of chemical compounds, but it did not displace atomism. Instead, Boltzmann, Planck and others argued that it could be reduced to statistical mechanics, a theory of moving atomic particles (see Thermodynamics). And through experiments with X-ray diffraction and radioactivity it became possible to measure quite directly exactly how many such particles there were in a gas sample and the distances between them in a crystal lattice.

4 Chemistry with electrons

In the early twentieth century chemistry was revolutionized by the discovery that the atom has parts, and by 1932, when the neutron was discovered, the basic features of our present picture of the chemical atom were firmly in place. The defining characteristic of a chemical element now became the number of protons in the nucleus - atomic weights were no longer universal constants but mere averages of isotopic distribution in the earth's crust. Valence, a useful empirical concept that had previously been devoid of theoretical interpretation, could now be related to the configurations of the outer electrons surrounding the chemical nucleus.

Since there was little opposition to the new chemistry, explicit recognition of just how radically concepts had been changed was slow in coming. Not only were chemical atoms divisible (through ionization as well as through nuclear fission), but even more fundamentally there were now no units that remained invariant during chemical processes. The closest candidate perhaps would be the kernel of the chemical atom - the nucleus plus the filled inner electron shells - but then, paradoxically it is the electrons outside of the kernel that are essential to chemical bonds.

The new concept of chemical element also deviated considerably from the traditional one. Not only was the law of conservation of matter revised by relativity theory, nuclear reactions violated Lavoisier's law of the conservation of elements. In a sense Lavoisier's operational definition of element as the last product of analysis could still be relied on to reveal the simplest homogeneous 'stuffs' out of which everything was constructed, but the arché of the universe now turned out to have no resemblance to the elements of the chemist, but to be instead the fundamental particles of the physicist. Even if one could restrict the methods of analysis to so-called chemical ones in a non-question-begging way, it was now clearer than ever that the relationship between the properties of the simple substance (for example, elemental carbon) and the properties of the stuff which was conserved in chemical reactions (for example, the element carbon) were extremely complex.

Of particular interest to philosophical accounts of the development of science are the interactions between chemical theories of bonding and atomic physics during the first quarter of the century. To chemists the discovery of the electron immediately provided the key to an understanding of all sorts of electrochemical phenomena, and a theory of ionization could account for properties of solutions of weak and strong electrolytes. Bohr's theory of discrete electron orbits, each containing a maximum number of electrons, also seemed to account for some of the similarities and trends summarized by the periodic table. However, the Bohr model, which was designed to account for the spectra of isolated atoms, postulated moving electrons, and it was not at all clear how it could be extended to cover the distribution of electrons in compounds. Theoretical chemists struck out on their own and in 1916 G.N. Lewis attempted to give a unified account of ionic and covalent bonding, which was based on the idea of the special stability of electron pairs and octets. According to Lewis, bonding electrons were always arranged at the corners of a cube, atoms could complete their octets by sharing corners or faces of their cubes, and ionic bonds formed when the sharing was unequal.

Lewis' account made a certain amount of sense to chemists although it predicted an inaccurate geometry for carbon compounds. It also served to dramatize the inadequacies of what we now call the old quantum theory.

Although scientists can live for some time with inconsistencies within a theory, or contradictions between mutually relevant theories, the drive for consistency is an enduring one and it was a great triumph when in 1926 Schrödinger proposed a quantum mechanical account of the atom which appeared to be applicable to chemical bonds as well as atomic spectra. By conceptualizing electrons as three-dimensional charged clouds one might hope to show that they were denser in just those regions of space where chemists were accustomed to draw lines or pairs of dots to represent chemical bonds. However, actually carrying out the derivations was a formidable problem. When the physicists Heitler and London set out to calculate the bond energy of H_2, the simplest molecule of all, they were forced to adopt a simplified representation of the interactions between four charged bodies (two protons and two electrons) and as a result were able only to approximate the observed values.

If quantum mechanics was to be of any benefit at all to chemists, more rough and ready methods of modelling complex systems had to be found. And there sprang up two rival schools of quantum chemistry: Linus Pauling's valence bond approach seemed well-suited to describing the geometry of covalent molecules, while Robert Mulliken's molecular-orbital approach was particularly useful in describing the vibrational spectra of molecules. There is still debate about the relative merits of these two approaches and both continue to be presented in contemporary chemistry textbooks.

5 Current philosophical issues

What has most intrigued philosophers about twentieth-century chemical theory is the question of whether chemistry is reducible to physics. While it is obviously desirable that theoretical chemistry be *compatible* with quantum mechanics, the demand that it be *reducible* is a much more stringent and controversial goal, for reduction requires that the entire theoretical and empirical corpus of chemistry be derivable from, or reproducible within, the explanatory structure of physics.

The quest for reduction is a recurring scientific ideal, but it was particularly popular with the logical positivists who founded a unity of science movement (see Vienna Circle §4; Unity of science). The paradigm candidate for a successful reduction pair was classical and statistical thermodynamics, but it is also frequently claimed by both scientists and philosophers that it is no longer possible to do foundational theoretical work in chemistry because, after all, chemistry is really nothing more than the application of quantum mechanics to complicated arrays of electrons and nuclei.

While few would dispute the claim of reducibility in principle, there are many who find the project of analysing the actual relationships between the two disciplines to be a challenging and instructive one. Consider, for example, the puzzles surrounding the *aufbau* principle. It is often claimed that the striking chemical similarities among elements summarized by Mendeleev's periodic table can be given a straightforward explanation in terms of quantum mechanics. One simply 'builds up' the table by placing electrons in the lower energy levels first and demonstrates that elements in the same chemical group turn out to have similar electronic configurations. Thus, the outer electrons of fluorine ($2s^2p^5$) and chlorine ($3s^2p^5$) differ only in their principal quantum number. However, this simple procedure fails when one tries to build up the transition elements because the ordering of orbits by

energy levels is not invariant with nuclear charge. So, for example, in nickel (atomic number 28) the 4s level is filled before the 3d level while in copper (atomic number 29) the 3d level is filled at the expense of the 4s! This ’crossing over’ of energy levels can in theory be derived from first principles, but in practice, quantum chemists resort to *ad hoc* rules of thumb, referring to the increased stability of filled and half-filled sub-levels in order to summarize such deviations from the simple *aufbau* recipe.

Philosophers, theoretical chemists and chemistry educators continue to wrestle with the many puzzling conceptual features of contemporary chemistry. Given the holistic character of quantum mechanics, how can we justify chemical explanations which treat atoms and molecules as discrete entities? It has even been argued that quantum mechanics does not permit talk about the shape of a molecule! Can one defend or at least clarify the eclecticism of chemists who switch from valence-bond thinking to an molecular-orbital account according to convenience?

Current chemistry also serves as a strategic site for philosophers who wish to develop more detailed and realistic accounts of methodological issues which arise in many sciences: how to deal with mathematically intractable representations, the role and evaluation of approximate explanations, the analysis of experimentation and instrumentation. Because of its central role in military, industrial and environmental affairs, chemical research has also come under the scrutiny of philosophers interested in the ethical and social dimensions of science (see Responsibilities of scientists and intellectuals).

See also: Conservation principles; Quantum mechanics, interpretation of; Reduction, problems of

NORETTA KOERTGE

References and further reading

Boscovich, R.J. (1763) *Theoria philosophiae naturalis*, Venice; trans. J.M. Child, *A Theory of Natural Philosophy*, Cambridge, MA: MIT Press, 1966.(Relates matter to the concept of field.)

Brock, W.H. (1992) *The Norton History of Chemistry*, London: Norton.(Useful bibliographic guide.)

Herschel, J.F.W. (1830) *A Preliminary Discourse on the Study of Natural Philosophy*, London: Longmans, Rees, Orme, Brown & Green.(Treats chemistry as a branch of physics.)

Kirk, G.S. and Raven, J.E. (1962) *The Presocratic Philosophers: A Critical History with a Selection of Texts*, London: Cambridge University Press.(Contains quotations from early Greek texts.)

Lavoisier, A.-L. (1789) *Traité élémentaire de chimie*; trans. R. Kerr, *The Elements of Chemistry*, London: Wm. Creech, 1790.(A *very* influential textbook.)

Paneth, F. (1961) ’The Epistemological Status of the Concept of Element’, *British Journal for the Philosophy of Science* 13: 1-14, 144-60.(Clarifies relationships between atoms and elements.)

Post, H.R. (1968) ’Atomism 1900’, *Physics Education* 3: 1-13.(Summarizes the opposition to atomism.)

Primas, H. (1983) *Chemistry, Quantum Mechanics, and Reductionism*, New York: Springer.(Account of technical difficulties in carrying out the reduction.)

Scerri, E. (1989) ’Transition Metal Configurations and Limitations of the Orbital Approximation’, *Journal of Chemical Education* 66: 481-3.(Explains the difficulties in applying the *aufbau* principle.)

Cheng

In early Confucian writings, cheng *describes the quality of authentically realizing or 'completing' a given thing's true nature. It appears together with* xin *(trustworthiness), a character to which it is related in sense.* Cheng *refers primarily to the fulfilment of a thing's true nature, while* xin *refers to the quality resulting from this. With regard to human beings, cheng is the authentic realization of ones nature. In texts such as the* Xunzi *and* Zhongyong*, the idea is related to the role human beings are believed to play in realizing or 'completing' a greater universal pattern. This development becomes centrally important for later neo-Confucian thinkers, who see these as different aspects of a single project.*

Cheng ('integrity' or 'sincerity') is not a central term of art for Confucius, but it takes on greater significance in the thought of Mencius and Xunzi. According to the *Mengzi* (4A12), *cheng* requires an understanding of the good. One cannot be *cheng* unless one understands why one is acting as one does. To be *cheng* is to be true to one's self, that is, one's true nature (hence 'sincerity'). Thus *cheng* is necessary for the cultivation of genuine virtue. Since such self-cultivation results in the most satisfying of lives (see Self-cultivation, Chinese theories of), there is no greater joy than to find that one is *cheng*.

For Xunzi, *cheng* assumes greater prominence: 'Nothing is more beneficial for nourishing one's heart-mind', and 'the extension' of this excellence is the beginning and end of the moral life (*Xunzi* 3). Xunzi introduces the idea that one who is *cheng* has the power to transform both people and things. One who is *cheng* and *xin* (trustworthy) is 'like a spirit'.

All the above ideas are further developed in the Daxue (Great Learning) and Zhongyong (Doctrine of the Mean). The former emphasizes that *cheng* is primarily concerned with one's internal mental states by calling for all to *cheng* (make sincere) their *yi* (thoughts). One is to fulfil but not overstep the proper measure in one's perception of and response to things and events. In the *Zhongyong* we find the most complete development of the notion of *cheng*. Echoing Xunzi, the text gives *cheng* a strongly metaphysical sense: it is each thing's realization of its proper nature as part of a grand cosmic pattern. Bringing this state of affairs about is the highest purpose of human life. Thus, '*cheng* is the way of heaven and to realize *cheng* is the way of human beings'. One is to realize *cheng* by fulfilling (that is, completing) one's own nature. This in turn allows one to complete the nature of others and go on to complete the nature of all things, thus 'forming a triad with heaven and earth'.

The notion of *cheng* was extremely important for neo-Confucian thinkers. Cheng Yi offered what became the standard interpretation. Discussing 'nature' and *cheng*, he explained that the former was a substantial thing while the latter was the quality of that thing's being fully realized as it should be: 'If you compare the nature to this fan, integrity (*cheng*) may be compared to this fan being properly made' (Graham 1992: 67).

See also: Neo-Confucian philosophy; Daxue; Self-cultivation, Chinese theories of; Virtue ethics; Xin (trustworthiness); Xunzi; Zhongyong

PHILIP J. IVANHOE

References and further reading

Graham, A.C. (1992) *Two Chinese Philosophers*, La Salle, IL: Open Court.(An insightful study of two neo-Confucian thinkers, Cheng Hao and Cheng Yi; contains helpful discussions of *cheng*, particularly on pages 67-73.)

Cheng Hao (1032-85)

*Cheng Hao was a pivotal figure in the creation of a Confucian tradition that was to become the basis for intellectual and state orthodoxy in China from the thirteenth century to the twentieth century. His decision to seek the Confucian Way (*dao*) through a direct and personalized reading of the classics was later projected as the beginning of this movement. From a new perspective, he redirected Confucian discourse on such cardinal concepts as humaneness and human nature.*

Born into a family which for three generations had distinguished itself in high offices, Cheng Hao accompanied his father to a succession of posts in central China. At the age of twenty he passed the national civil service examination and, for most of the years until 1180, had a notable official career which culminated in 1169-70 with service at the emperor's court. During audiences with the emperor and in written memorials, he followed the example of Mencius in admonishing his ruler to follow benevolence and refusing even to discuss what would bring profit and advantages. So critical was he of Wang Anshi's utilitarian obsession with maximizing advantages and the happiness of the greatest number of people that Cheng was demoted to a local post and eventually dismissed entirely. The last five years of his life he devoted to teaching an increasing number of disciples, drawn in part by his exceptionally gracious and warm disposition.

Retrospective biographical accounts focused on his youthful fascination with Daoism and Buddhism, until his resolve to discover the Way (*dao*) led him to the classics (see Chinese Classics). Reading the classics in the context of his commitment to the Way, he was credited with a breakthrough in understanding the mind and heart of Mencius and reviving the transmission of the Way that had been lost since this last sage of the classical era.

Cheng Hao was the first to expound the idea of 'principle' (*li*), the normative patterns providing coherence to all things, as a Confucian concept, but his younger brother Cheng Yi was the one who developed it into a Confucian philosophy (see Li). By equating human nature (see Xing) with principle and the Way, neither brother regarded human nature as merely the raw human state. Although the younger brother would refer to human nature, principle and the Way as good, Cheng Hao considered them beyond any opposite, so he did not describe them as good. The opposite of good arose, like dysfunctional behaviour, from imbalance and deviation from the Mean. Cheng Hao also first broadened the Confucian idea of humaneness (*ren*) to encompass not only a profound sympathy with all beings but also a mystical oneness with the universe. Drawing support from a medical text, he noted that paralysis of the human limbs was described as an absence of *ren*; thus, it was the life force running through things. Seeking to teach spiritual cultivation of the self in order to realize the unity with all things, he condemned ego-centredness as a barrier to achieving this sense of oneness (see Confucian philosophy, Chinese §5; Neo-Confucian philosophy §5).

Cheng has been interpreted primarily in two ways. Beginning with Zhu Xi's synthesis of the Confucian Way tradition (see Zhu Xi), Cheng and his younger brother have been presented as a unified voice and as students of Zhou Dunyi. This formulation facilitated Zhu's synthesis of diverse ideas, but even some of his contemporaries questioned his historical and philosophical accuracy. Other scholars, most notably Fung Yu-lan (1953), put Cheng Hao at the head of a subjective wing of Confucianism and his younger brother at the beginning of a rationalistic wing that culminated in Zhu Xi. Both conventional views have, since the 1980s, come to be regarded as exaggerated.

See also: Cheng Yi; Confucian philosophy, Chinese; Dao; Neo-Confucian philosophy; Zhou Dunyi

HOYT CLEVELAND TILLMAN

List of works

Cheng Hao (1032-85) *Er Cheng ji (Collected Works of the Two Chengs)*, Beijing: Zhonghua, 1981, 4 vols. (Cheng Hao's writings are included in this collection, most of which comes from his brother Cheng Yi. Many of the sayings and conversations recorded by their disciples do not make it clear which Cheng brother was talking.)

References and further reading

Chan Wing-tsit (1963) *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press.

(Convenient source of translations, and from the Zhu Xi school of interpretation.)

Zhu Xi [Chu Hsi] and Lü Zuqian (1175) *Reflections on Things at Hand*, trans. Chan Wing-tsit, New York: Columbia University Press.(A good translation of the most influential anthology centring on the works of the Chengs.)

Feng, Youlan [Fung Yu-lan]. (1953) *A History of Chinese Philosophy*, trans. D. Bodde, Princeton, NJ: Princeton University Press, 2 vols.(A dated and controversial, but still useful, interpretation of the Chengs from his synthesis of Chinese and Western philosophy.)

Graham, A.C. (1958) *Two Chinese Philosophers: Ch'eng Ming-tao and Ch'eng Yi-ch'uan*, London: Lund Humphries; repr. La Salle, IL: Open Court, 1992. (Classic study by a significant modern philosopher, with copious translated passages from both Cheng brothers.)

Cheng Yi (1033-1107)

One of the most creative Chinese intellectuals, Cheng Yi was the most systematic and influential of a group of thinkers which channelled Confucian thinking into a new philosophical direction that gradually became dominant in East Asia for several centuries. Buddhism was still the most pervasive and sophisticated religious philosophy of his day; yet he effectively borrowed some of its ideas and methods to formulate a philosophy that would enable Confucian teachers to draw intellectuals away from the Buddhist masters.

Born to a scholar-official family, Cheng Yi lived most of his life around Luoyang, traditionally the secondary capital of China. Although he passed the major part of the highest civil service examination in 1059, he failed the culminating palace examination. Thereafter he lost interest in an official career and devoted himself to learning and teaching the 'Way' (*dao*) of the sages. Though he attracted a few early students, his prominence as a Confucian master came in his mid-fifties when he was appointed a lecturer to the emperor and subsequently a professor in the national academy. Even though his official career was thus limited to educational posts held late in life, his outspoken ideas and uncompromising demeanour eventually provoked the government to confiscate his land and exile him for three years. When he died in 1107, only the most loyal of his students dared to attend his funeral.

His followers developed a fellowship of Confucian intellectuals through which his ideas were disseminated. When in the twelfth century Zhu Xi systematized philosophical reorientations of Confucian traditions for what soon became mainstream thinking and subsequently state orthodoxy, he drew more heavily on the ideas of Cheng Yi than any other eleventh-century master.

The central concept in Cheng Yi's philosophy was 'principle' (*li*), the normative patterns inherent in and providing coherence to each and every individual thing (see Li). While his elder brother Cheng Hao used 'principle' to point to a moral force alive in all things, Cheng Yi postulated it as the metaphysical basis of the harmony of the one and the many. Seeking to refute the Buddhist doctrine of 'emptiness' (*śūnyatā*) (see Buddhist concept of emptiness). Cheng Yi pointed to the linkage between the philosophical, ethical ultimate and the myriad things in daily life in order to fortify Confucian belief in the significance of structures, words and deeds in the sociopolitical realm. For instance, identifying human nature with principle provided a more philosophical basis for Mencius' doctrine of the goodness of human nature (see Mencius; Xing). Moreover, as a way of explaining the existence of dysfunctional behaviour, Cheng Hao pointed to the interaction between principal and the psychological energy (*qi*) that constitutes the 'stuff' of all people and things (see Qi). Ironically, as Yü Ying-shih (1987) shows, Cheng Yi developed Confucian 'this worldly' ethics after gaining insights into Chan disciplinary methods of meditation and spiritual cultivation. Perhaps because of the nature of his borrowings from Buddhism, his most frequent criticism of Buddhists was not their metaphysics, but their alleged selfishness in devaluing relations in family and society.

The concept of principle also provided a more sophisticated paradigm than the correlative cosmology developed over a thousand years earlier for explaining what was known about the natural world. Like that earlier synthesis, however, Cheng Yi was using the science of his day to bolster traditional values and roles associated with Confucian ethics. For instance, he rejected earlier Confucians' equating the cardinal virtue of humaneness (*ren*) with altruistic love (see Confucian philosophy, Chinese §5). Love was just a secondary attribute, because humaneness as principle denoted the oneness of all beings.

In Peter Bol's analysis, Cheng Yi made a radical break with the approach that the literati elite took to learning (Bol 1992). Whereas several centuries of intellectuals had focused on literary culture as the medium for understanding culture and values, Cheng announced that the mind could directly grasp the moral patterns inherent in the natural world. Thus, instead of responding primarily to the Buddhist challenge, Cheng was essentially intent on winning the sociopolitical elite to his view of the grounds of ethical values.

See also: Buddhist philosophy, Chinese; Cheng Hao; Confucian philosophy, Chinese; Li; Neo-Confucian philosophy §5

HOYT CLEVELAND TILLMAN

List of works

Cheng Yi (1033-1107) *Er Cheng ji (Collected Works of the Two Chengs)*, Beijing: Zhonghua, 1981, 4 vols. (Although Cheng Hao's writings are included, most of the four volumes comes from the work of Cheng Yi. Many of the sayings and conversations recorded by their disciples do not record which Cheng brother was talking.)

References and further reading

Bol, P.K. (1992) *'This Culture of Ours': Intellectual Transitions in T'ang to Sung China*, Stanford, CA: Stanford University Press. (Using the paradigm of tension between the Way (*dao*) and literary culture (*wen*), Bol explains Cheng Yi's metaphysical philosophy as a radical departure from the Tang and early Song mainstream of literati thought.)

Graham, A.C. (1958) *Two Chinese Philosophers: Ch'eng Ming-tao and Ch'eng Yi-ch'uan*, London: Lund Humphries, reprinted 1992, La Salle, IL: Open Court. (Classic study by a significant modern philosopher and with copious translated passages from both Cheng brothers.)

Graham, A.C. (1986) 'What Was New in the Ch'eng-Chu Theory of Human Nature?', in Chan Wing-tsit (ed.) *Chu Hsi and Neo-Confucianism*, Honolulu, HI: University of Hawaii Press. (Perhaps the best study in English of Cheng Yi's view of the central Confucian problem of human nature.)

Tillman, H.C. (1992) *Confucian Discourse and Chu Hsi's Ascendancy*, Honolulu, HI: University of Hawaii Press. (A study of the Confucian fellowship and its ideas beginning with Cheng Yi and continuing through its establishment as orthodoxy in 1241.)

Yu Yingshi [Yü Ying-shih] (1987) *Zhongguo jinshi congjiao lunli yu shangren jingshen (The Religious Ethic and Mercantile Spirit in Modern China)*, Taibei: Lianjing. (Refutes Max Weber's findings about Chinese religion, but uses the question poised by Weber about the impact of religious values on merchant class values to explore changes in Buddhism and Confucianism, beginning respectively in the Tang and Song periods. An expanded English version is forthcoming.)

Chernyshevskii, Nikolai Gavrilovich (1828-89)

Nikolai Chernyshevskii was the main theorist of the Russian democratic radicalism of 'the 1860s' or, more precisely, of the period of political 'thaw' and liberal reforms which followed Russian defeat in the Crimean War and the enthronement (in 1855) of Alexander II. He was also the best representative of the non-conformist elements among the raznochintsy*, that is, the educated commoners, who at that time began to figure prominently in Russian intellectual and social life. As such, he exerted a powerful formative influence on the Russian intelligentsia.*

In 1862 Chernyshevskii was arrested, brought to trial and, despite insufficient evidence, condemned to lifetime banishment in Siberia. In exile, preserving his integrity to the end, he stoutly refused to ask for clemency (as a result, he remained in a remote Siberian village until 1883). In prison, waiting for trial, he wrote the novel Chto Delat' (What Is To Be Done?) in which he showed the 'new men' of Russia - 'rational egoists', devoted to the cause of progress, and even a type of ascetic, self-sacrificing revolutionist. Thanks to a strange oversight of the censor the novel was serialized in the journal Sovremennik (The Contemporary) and, despite lack of literary distinction, became a powerful source of inspiration for several generations of Russian progressive youth.

Chernyshevskii's philosophical reputation was created by the Russian Marxists. The 'father of Russian Marxism', G.V. Plekhanov, greatly impressed by Chernyshevskii's combination of Feuerbachian materialism and respect for Hegelian dialectics, described him as an important precursor of dialectical materialism. This view was taken up by Lenin who in Materialism and Empiriocriticism called Chernyshevskii 'the great Russian Hegelian and materialist', the only Russian philosopher before Marxism who was able to defend 'integral materialism' against the agnosticism and subjectivism of Neo-Kantians, positivists, Machists and 'other muddleheads'. Soviet philosophers went even further: Chernyshevskii was treated by them not only as the greatest Russian philosopher before Marxism, but also as the greatest pre-Marxian philosopher of the world, founder of the highest form of pre-Marxian materialism. For several decades this was the obligatory dogma of the Soviet official ideology.

After the breakdown of the Soviet Union, interest in Chernyshevskii's philosophy almost completely disappeared in Russia. It is impossible, however, to deny Chernyshevskii's importance in Russian intellectual history. Hence the need of rethinking his philosophical legacy.

1 Aesthetics

Nikolai Gavrilovich Chernyshevskii, son of an Orthodox priest in the city of Saratov, intended initially to become a priest himself. However, he changed his mind and, after graduating from a theological seminary, entered the faculty of history and philology at St Petersburg University. During his studies in St Petersburg he became acquainted with the works of Hegel and Feuerbach, as well as with the writings of the French utopian socialists. His favourite Russian thinker was the literary critic V.G. Belinskii. Young Chernyshevskii fully shared Belinskii's view on the special mission of literature and literary criticism in autocratic Russia. He decided, therefore, to provide literary critics with a theoretical clarification of the essence of art (with particular emphasis on literature) and of the role of art in society. He fulfilled this task in his master's dissertation 'Ėsteticheskie otnosheniia iskusstva k deistvitelnosti' (The Aesthetic Relations Between Art and Reality) (published and publicly defended in the spring of 1855).

Chernyshevskii's dissertation (as he himself explained in the preface to its third edition, 1888) was in fact a critique of Hegelian aesthetics from the point of view of Feuerbachian materialism. However, the Draconian censorship regulations, introduced in 1848 and still holding at the time of its writing, made it impossible to refer openly to Hegel and Feuerbach. Hence, Chernyshevskii's thesis took the form of a sharp critique of the Hegelian aesthetics of F.T. Visher, seen as a representative specimen of the 'dominant aesthetic theory'.

Against Hegelian idealism, defining beauty as a manifestation of the Absolute Idea, Chernyshevskii set his own conception, which reads as follows: 'Beauty is life; beautiful is that being in which we see life as it should be according to our conceptions'. In accordance with Feuerbach's anthropological materialism, this was to be a materialist and an anthropocentric conception: materialist, because 'life' was seen in it as rooted in nature and the beauty of art was interpreted a mere reflection of natural beauty; anthropocentric because the human being in the

full flowering of his faculties was made the measure of all things and the supreme ideal of beauty. Chernyshevskii conceded that this anthropological ideal could be interpreted in different ways, depending on historical and social circumstances, but strongly insisted that the aesthetic ideas of the privileged classes reflected an artificial life and were, therefore, normatively illegitimate. Only the working people, living in touch with nature, were the true representatives of the human species and thus the living embodiments of the supreme aesthetic ideal. In this way the Feuerbachian 'rehabilitation of nature' was linked with a cult of the common people and with the demand for the widest possible democratization of art.

Chernyshevskii's programmatic naturalism led to the conclusion that art was merely a surrogate for reality. Against aesthetic idealism Chernyshevskii set the view that works of art could only copy real life and that artists, therefore, could not claim to possess genuinely creative power. Nevertheless, he also stressed, somewhat inconsistently, that the function of art is not only to reproduce reality, but also to explain and evaluate it - to 'pass judgment' on the real-life phenomena that have been recreated in it. He made it explicit that because of this art, especially literature, could and should perform a socially useful didactic role, and that artists, especially writers, could and should become spiritual leaders of society.

Chernyshevskii's aesthetic ideas aroused very different reactions. Critics who defended uncommitted 'pure art' were horrified by them; Ivan Turgenev, expressing the views of many liberal-minded writers, reduced them to 'blind malice and stupidity'; in contrast with this Dmitrii Pisarev praised Chernyshevskii's thesis as a total elimination of aesthetics in the name of utilitarian values. On the whole, the acceptance or non-acceptance of the 'Aesthetic Relations Between Art and Reality' marked the difference between Russian liberals and democratic radicals. The latter enthusiastically embraced Chernyshevskii's ideas, treating them as obligatory guidelines for progressive literary criticism.

Soviet literary scholars used to emphasize that in *What Is To Be Done?* Chernyshevskii did not limit himself to a critical explanation and evaluation of the existing reality but also expounded positive ideals - an image of a revolutionary hero and a vision of a socialist society of the future. This allowed them to treat Chernyshevskii as an important precursor of 'socialist realism'.

2 The anthropological principle

Chernyshevskii's philosophical position was best defined in his 'Antropologicheskii printsip v filosofii' (Anthropological Principle in Philosophy) (1860), a work written in connection with Lavrov's *Sketches* (see Lavrov, P.L. §1). (Contrary to most commentaries, Chernyshevskii endorsed Lavrov's anthropologism and only wanted to remove from it what he saw as concessions to idealism and eclecticism).

For Chernyshevskii the 'anthropological principle' supplied the theoretical foundation for the integral wholeness of man, the abolition of the eternal dualism of body and soul. The human individual, he argued, has only one indivisible nature and only as such can he represent an absolute value to other human beings. In Chernyshevskii's interpretation this principle was a cornerstone of an entire philosophical worldview (see Russian Materialism: 'the 1860s').

First, it was conceived as the reinstatement of a distinctively Feuerbachian, anthropocentric materialism, that is to say, a form of materialism capable of resisting all forms of idealism without becoming fully dependent on the dehumanized, mechanistic world of the natural sciences. Chernyshevskii intended his materialism to provide a basis for value-judgments; hence he had to distance himself from the cold abstractions of value-neutral objectivist scientism.

Second, Chernyshevskii's materialism was a theory of the human organism as both knower and object of knowledge - thus positing the indivisibility of matter and consciousness. For man, Chernyshevskii reasoned, the primary datum is not thinking but existence, including bodily existence in the material environment. He saw this standpoint as undermining the foundation of idealistic epistemology and proving decisive arguments against subjectivism and agnosticism.

Finally, the anthropological principle was conceived as a philosophical underpinning of the ethical conception of 'rational egoism' - a conception based on the premise that, ultimately, the guiding principle of conduct is individual interest. In the normative sphere this theory gave preference to utilitarianism, rationalism and

egalitarianism. It postulated that the standard by which human actions must be judged is the benefit they bring - that good is not an autonomous value but only a lasting benefit, ‘a very beneficial benefit’. The rational egoist accepts other people’s right to be egoists because he knows that all people are equal; in controversial issues he is guided by the principle of the greatest good of the greatest number. In interpreting this utilitarian principle Chernyshevskii differed from Bentham’s belief in the self-regulating market; he followed instead Feuerbach’s view that truly rational individuals were communal beings, capable of conscious cooperation in pursuing mutually agreed common ends. His social ideal was therefore a cooperative socialism, presupposing the full liberation of the individual from all sorts of blind faith and authoritarian compulsion. He outlined this vision in *What Is To Be Done?*, emphasizing in particular the full emancipation of women.

3 Political philosophy

Chernyshevskii’s political philosophy concentrated on problems relevant to Russia as an underdeveloped country, a ‘latecomer to the arena of history’.

In many points - in his commitment to the idea of the liberation of the individual, in his conviction that Russia should still learn from the West and humbly recognize the superiority of Western achievements - Chernyshevskii continued the line of Russian Westernism. In the essay ‘O prichinakh padeniia Rima’ (What Caused the Downfall of Rome?) (1861) he criticized all Russian conceptions of the ‘senility’ of the West, including Herzen’s ‘Russian socialism’ (see Herzen, A.I.). And yet there was a time (1856-7) when he thought it appropriate to cut himself off from the epigones of Westernism and to declare that in many respects he stood closer to the Slavophiles. He even proposed to the Slavophiles a kind of alliance for the joint defence of the peasant commune against the liberal economists, who demanded that it be abolished together with feudal bondage. This alliance, however, could not be concluded because the differences between the two partners were too important: the Slavophiles wanted to preserve pre-capitalist social relations, while Chernyshevskii saw the commune as a suitable means for a direct transition to a rational socialist collectivism.

In his articles on political economy (which were criticized but highly appreciated by Marx) Chernyshevskii set against bourgeois liberalism a conception of ‘the political economy of the working people’. He fully recognized the progressiveness of capitalist development in the West, but was aware of its contradictions and wanted for his country a more humane and, at the same time, a more rapid type of progress. In his ‘Kritika filosofskikh predubezhdenii protiv obshchinnogo zemlevladeniia’ (Criticism of Philosophical Prejudices Against the Peasant Commune) (1859) he argued that Russia, and backward countries in general, could benefit from the experience and scientific achievements of the West and ‘skip the intermediate stages of development or at least enormously reduce their length’. His main argument for the commune was based on a dialectical conception of progress: the first stage of any development is, as a rule, similar in form to the third; thus primitive communal collectivism is similar in form to the developed collectivism of a socialist society and can make easier a direct transition to it.

Chernyshevskii’s criticism of the liberals was also levelled against political liberalism. He published in his journal *Sovremennik* a series of articles on the revolutions in France in which he sharply distinguished between liberals, who aimed at ‘merely political’ freedom, and democrats, whose main concern was the social welfare of the people. He did not hesitate to say that from the democratic point of view Siberia was a better country than England because Siberian peasants were better off than English proletarians (‘Iulskaiia monarkhiia’ (The July Monarchy) 1858). This populist conception of democracy exerted a mighty influence on the Russian revolutionary movement. Members of the revolutionary organization Land and Freedom treated Chernyshevskii as their intellectual leader and often asked him for advice.

However, Chernyshevskii himself did not embrace revolutionary solutions. Undemocratic methods used to prepare the abolition of serfdom in Russia, as well as his own unlawful arrest and trial, convinced him that the main obstacle on the way to true progress in Russia was lack of the rule of law and political freedom. Therefore in his ‘Pis’ma bez adresa’ (Letters Without Addressee) (1862) - whose addressee was actually Alexander II - he sided with the gentry liberals of Tver’, who demanded a liberal constitution for Russia. In the penal settlement in Siberia he surprised his fellow prisoners by a total reversal of his previous view of political freedom. It is true, he argued, that political freedom cannot feed a hungry man, but without food people can survive several days, while without air they cannot live more than a few minutes. As air is necessary for the life of the human organism, so political freedom is necessary for the normal functioning of society.

Nevertheless, Russian liberals did not claim Chernyshevskii's legacy, thus allowing the revolutionists to take a monopolistic hold on it and to interpret it accordingly. The reception of Chernyshevskii's political ideas was therefore extremely one-sided. He was seen as an intransigent enemy of liberalism, preaching class struggle, categorically rejecting all half-way reforms and pushing Russia onto the path of agrarian revolution. This interpretation was radicalized by Lenin, in whose eyes Chernyshevskii was the greatest, most consistent representative of pre-Marxian 'revolutionary democracy' in Russia. The Soviet Marxists accepted this view and tried to present 'revolutionary democratism' as the most advanced political position before the emergence of Marxism.

ANDRZEJ WALICKI

List of works

Chernyshevskii, N.G. (1939-53) *Polnoe sobranie sochinenii (Complete Works)*, Moscow: Khudozhestvennaiia literatura, 16 vols.

Chernyshevskii, N.G. (1863) *Chto delat'*, in *Sovremennik*, St Petersburg; trans. B.R. Tucker, *What Is To Be Done?*, New York: Vintage Books, 1961.(Chernyshevskii's famous novel, first published in serial form.)

Chernyshevskii, N.G. (1953) *Selected Philosophical Essays*, Moscow: Foreign Language Publishing House.(A representative selection of Chernyshevskii's philosophical thought in English. Includes 'Aesthetic Relations Between Art and Reality', 'Anthropological Principle in Philosophy' and 'Criticism of Philosophical Prejudices Against the Peasant Commune'.)

References and further reading

Lavretskii, A. (1941) *Belinskii, Chernyshevskii i Dobroliubov v bor'be za realism (Belinskii, Chernyshevskii and Dobroliubov in the Struggle for Realism)*, Moscow: Khudozhestvennaiia literatura.(A representative Soviet work on Chernyshevskii's aesthetics and literary criticism, trying to present him as a precursor of 'socialist realism'.)

Lukács, G. (1956) *Einführung in die Aesthetik Tschernyschevskijs (Introduction to Chernyshevskii's Aesthetics)*, in G. Lukács, *Beiträge zur Geschichte der Aesthetik*, Berlin.(Takes the view that Chernyshevskii's theory of realism represented the highest point in the development of pre-Marxist aesthetics.)

Pereira, N.G.O. (1975) *The Thought and Teachings of N.G. Černyševskij*, The Hague: Mouton.(A good introduction to Chernyshevskii's thought in English.)

Plekhanov, G.V. (1894) *N.G. Tschernischevsky. Eine literar-historische Studie (N.G. Chernyshevskii. A Historical and Literary Study)*, Stuttgart: Dietz; *N.G. Chernyshevskii*, St Petersburg: Shipovnik, 1910.(The first comprehensive monograph on Chernyshevskii's philosophical view, by the 'Father of Marxism'.)

Steklov, Y.M. (1928) *N.G. Chernyshevskii, ego zhizn' i deiatelnost' (N.G. Chernyshevskii, His Life and Activities)*, Moscow-Leningrad: Gosudarstvennoe izdatelstvo, 2 vols.(The crowning achievement of the early Soviet scholarship on Chernyshevskii.)

Chillingworth, William (1602-44)

Chillingworth was one of the most notable English-speaking contributors to debates between Protestants and Catholics in the seventeenth century. His use of a distinction between metaphysical and moral certainty proved extremely influential, as did his rationalist and fallibilist approach to issues of faith and authority.

William Chillingworth was born in England at Oxford, and was educated there at Trinity College, where he became a fellow in 1628. In the same year he renounced his allegiance to the Church of England, resigned his fellowship and became a Roman Catholic. He travelled abroad to a Catholic seminary in the Netherlands, probably at Douai, possibly at St Omer, but soon found the life uncongenial and returned to England. In the early 1630s he had no clear religious allegiance, but was reconciled to the Church of England by 1635. From 1634 he lived at Viscount Falkland's house at Great Tew in Oxfordshire; Falkland's posthumously published *Discourse of Infallibility* (1645) owes much to Chillingworth's arguments. In the autumn of 1637 Chillingworth published his chief work, *The Religion of Protestants a Safe Way to Salvation*, directed against the Jesuit Edward Knott. When the Civil War broke out he sided with the King, and died as a prisoner of war in January 1644, following his capture at Arundel.

As its title indicates, *The Religion of Protestants* is a work of religious controversy, not a philosophical treatise. In this respect it resembles Hooker's *Laws of Ecclesiastical Polity* (1593-1600), but whereas Hooker adopted a strategy of laying down general principles and thereby out-flanking his opponents, Chillingworth chose to fight on his opponent's ground (see Hooker, R.). The structure of the work is therefore dictated entirely by its target, Knott's *Mercy and Truth. Or Charity Maintained by Catholickes* (1634). This choice of tactics reflected the fundamental characteristics of Chillingworth's mind, essentially critical and combative rather than speculative or constructive. At Oxford he had acquired a near-legendary reputation for agility in debate, and he was understandably confident of his capacity to subdue any opponent.

Knott and Chillingworth both unequivocally accepted the infallibility of the Bible, but differed as to whether this was enough. For Knott, infallible Scriptures need to be expounded by an infallible interpreter; otherwise nothing could ever be certain in religion, and faith would degenerate into mere opinion. Chillingworth adamantly denied this. The Scriptures provide an infallible rule by which controversies are to be judged, but no infallible judge of their meaning exists on earth, either in the person of the Pope or anywhere else. No church has therefore the right to claim infallibility for its articles of faith. The Church of England claims no more than 'an Authority of determining Controversies of Faith, according to plain and evident Scripture and Universal Tradition' (Chillingworth 1638: ii.162); neither its articles nor those of the Lutherans or the Calvinists possess the infallibility of the Scriptures from which they are derived. It is this that lies behind Chillingworth's most often quoted dictum - so misleading when taken out of context - that 'The BIBLE, I say, the BIBLE only, is the Religion of Protestants!' (Chillingworth 1638: vi.56).

Chillingworth saw Knott's position as resting on two fundamental confusions - that of infallibility with authority, and of infallibility with certainty. Not all certainties are the same - indeed two quite separate kinds need to be distinguished. Metaphysical certainty belongs to direct revelations from God, to self-evident propositions and their logical consequences, and to the direct testimony of the senses. Faith cannot have this kind of certainty - if it could it would no longer be faith, but knowledge; it can, however, be morally certain - that is, probable to the highest degree. This is the only certainty that historical propositions can ever possess. Chillingworth naturally rejected Knott's scriptural and patristic arguments for papal infallibility, but he also insisted repeatedly that, even if successful, they could give the doctrine moral certainty only, not the infallible, metaphysical certainty that Knott's position required.

Chillingworth's approach to theology was unswervingly rationalist: 'neither God doth nor man may, require of us as our dutie, to give a greater assent to the conclusion, than the premises deserve' (Chillingworth 1638: ii.154). Blind faith is in no way meritorious: 'God hath given us our Reason to discern between Truth and Falshood, and he that makes not this use of it, but beleeves things he knowes not why, I say it is by chance that he believes the Truth, and not by choice: and that I cannot but feare, that God will not accept of this *Sacrifice of fooles*' (ii.113). It is hardly surprising that this did not find favour with Chillingworth's opponents, either Catholic or Protestant.

Their doubts about the reality of Chillingworth's own faith were quite unfounded, as his sermons made plain, but suspicions that he might not be the safest of allies were more understandable. As Hobbes, who had met him at Great Tew, sardonically observed, 'he was like a lusty fighting fellow that did drive his enemies before him, but would often give his owne party smart back-blowes' (Aubrey 1898 vol. 1: 173).

Locke greatly admired Chillingworth, and the two men had much in common (see Locke, J. §7). Both were accused by hostile critics of Socinianism, wrongly in Chillingworth's case (see Socinianism), and both can more fairly be regarded as credal minimalists: 'wee suppose that all the necessary points of Religion are plaine and easie, & consequently every man in this cause to be a competent Iudge for himself' (Chillingworth 1638: ii.16). They shared a deep loathing of religious persecution, not merely for its cruelty, but for its presumption: 'God hath authoriz'd no man to force all men to *Unity of Opinion*' (Chillingworth 1638: ii.85).

A few months before *The Religion of Protestants* appeared, Descartes' *Discourse on Method* was published in Holland. Chillingworth had no knowledge of Descartes' work and their two projects were utterly dissimilar; the fundamental congruence of their epistemology is therefore all the more striking (see Descartes, R. §3). Both take as basic the cognitive sovereignty of the individual. Authority may be accepted, but has first to be judged by reason, and this is something that each one of us has to do alone.

See also: Faith

J.R. MILTON

List of works

Chillingworth, W. (1727) *The Works of William Chillingworth*, London: Benjamin Motte. (This ninth edition contains *Nine Sermons* preached to the King and others, and *Nine Additional Discourses*, all quite short, on matters related to the *Religion of Protestants*.)

Chillingworth, W. (1638) *The Religion of Protestants a Safe Way to Salvation*, Oxford: Leonard Lichfield; Menston: Scolar Press, 1972.(Chillingworth's most important philosophical work. Although published in the autumn of 1637, it is dated 1638 on the title page.)

References and further reading

Aubrey, J. (1898) 'Chillingworth', in *'Brief Lives', chiefly of contemporaries… set down between 1669 &1696*, ed. A. Clark, Oxford: Oxford University Press, 2 vols.(Edited from the author's original manuscripts, this is the source of the description of Chillingworth attributed to Hobbes.)

Cary, Sir Lucius, Viscount Falkland (1645) *Discourse of Infallibility*; repr. as *Sir Lucius Cary, Late Viscount of Falkland. His Discourse of Infallibility with an answer to it: And his Lordships Reply*, London: Gertrude Dawson, 1651.(Posthumously published work which owes much to Chillingworth's arguments.)

Hooker, R. (1593-1600) *Of the Laws of Ecclesiastical Polity*, in W.S. Hill (ed.) *The Folger Library Edition of the Works of Richard Hooker*; vols 1-3, Cambridge, MA: Harvard University Press, 1977-81; vol. 6, Binghamton, NY: Medieval and Renaissance Texts and Studies, 1993.(Books I-V of the *Laws* were published between 1593 and 1600, and Books VI-VIII published posthumously between 1648 and 1662. Volumes 1-3 of this edition provide the main text; volume 6 gives discursive introductions and detailed commentary.)

Knott, E. (Matthew Wilson) (1634) *Mercy and Truth. Or Charity Maintained by Catholickes*, St Omer.(The target of Chillingworth's *Religion of Protestants*.)

Leeuwen, H.G. Van (1963) *The Problem of Certainty in English Thought, 1630-1690*, The Hague: Martinus Nijhoff.(Not always accurate on Chillingworth, but a useful account of his influence.)

Orr, R.R. (1967) *Reason and Authority: the Thought of William Chillingworth*, Oxford: Clarendon Press.(The fullest account of Chillingworth's life and thought, including a list of unpublished manuscripts.)

Trevor-Roper, H. (1987) *Catholics, Anglicans and Puritans*, London: Secker & Warburg.(Chapter 4 gives an excellent account of the Great Tew circle.)

Chinese Classics

The Chinese Classics are a group of texts of divination, history, philosophy, poetry, ritual and lexicography that have, to a significant extent, defined the orthodox Ruhist (Confucian) tradition of China. Since the Song dynasty (960-1279), they have consisted of the following thirteen texts:

(1) The Shujing, or Shangshu (Book of Documents, or Documents), the 'classic' of Chinese political philosophy. Allegedly compiled by Confucius, it contains a variety of historical documents, mostly dating from the fourth century BC.
*(2) The Yijing (Book of Changes), a divinatory work using sixty-four permutations of broken (*yin*) and straight (*yang*) lines in six positions. It has two parts: the 'Zhouyi' (Zhou Changes), an ancient divination manual, and the Shiyi (Ten Wings), a commentary dating from the Warring States period (403-222 BC).*
(3) The Shijing (Book of Songs, or Odes), a collection of 305 poems, ostensibly selected by Confucius, on a wide variety of subjects. It includes songs of farming, feasting and love that are clearly of popular origin. It also contains a variety of court poetry including dynastic hymns, hunting and banquet songs and political satires from the Zhou court (1121-222 BC).
(4) The Yili (Ceremony and Rites), a Warring States ritual text.
(5) The Zhouli (Rites of Zhou), another Warring States ritual text.
(6) the Liji (Book of Rites), a Han work that provides information about early Confucian philosophy and ritual. Together, works (4), (5) and (6) make up the Lijing (Classic of Rites).
(7) The Zuozhuan (Zuo Annals).
(8) The Guliangzhuan (Guliang Annals).
(9) The Gongyangzhuan (Gongyang Annals). Works (7), (8) and (9) are commentaries to the Chunqiu (Spring and Autumn Annals, or simply Annals), a chronicle of the reigns of twelve rulers of the state of Lu; its presentation of diplomatic and political events from 722-481 BC is terse and factual, but the three commentaries provide substantial elaboration and exegesis.
(10) The Analects of Confucius (Lunyu), containing anecdotes and short dialogues between Confucius and his disciples. In this work, Confucius established a new emphasis on humanistic ethics and political and social order.
(11) The Xiaojing (Book of Filial Piety), a short dialogue between Confucius and one of his disciples, concerned with filiality in both private and public life; it discusses children's filiality to their parents and subjects' filiality toward their rulers.
(12) The Erya, a book of glosses of Zhou dynasty terms (the title means 'Graceful and Refined').
(13) The Mengzi, which records a series of dialogues and debates between the philosopher Mencius and his students, several rulers and a variety of rhetorical and philosophical opponents. Mencius elaborated upon the Analects, arguing that human nature was inherently good and claiming that four 'sprouts' of goodness could be educated to create intuitive ability as the correct basis for moral judgments.

*The practice of appealing to authoritative texts appeared as early as the Analects of Confucius, around 500 BC. An explicit classical canon first appeared some four hundred years later during the Han dynasty (206 BC-AD 220), when Emperor Wu institutionalized a set of five classics associated with Confucius. At the same time he established new procedures for recruiting officials, created official chairs for the study of the Five Classics, restricted official academic appointments to those five areas and founded an imperial academy for the study and transmission of those works. In this way he effectively created a new 'Confucian' state religion. The term 'classic' (*jing*) also appears as the first of six categories of literature in the classification system of the bibliographical chapter of the Hanshu (History of the Former Han Dynasty). Classics (*jing*) are distinguished from masters (*zi*), the latter being grouped into nine schools starting with the Ru, or Confucians.*

Since the Han dynasty, the content of the classical canon has grown from the original five (or seven) texts, as established during the Han dynasty. The original group of classical texts that acquired official sanction during the early Han empire was supplemented by additional texts during the Tang (617-907) and Song (960-1279) periods. The Chunqiu (Spring and Autumn Annals) became known under the titles of three commentary editions, the Gongyangzhuan, Guliangzhuan and Zuozhuan, as noted above. The Lijing became known as three separate works on ritual, the Yili, the Zhouli and the Liji, again as noted above. The Erya was added to the classical canon during

the Tang dynasty and the Mengzi during the Song dynasty, bringing the total to what became the standard thirteen texts.

These works functioned as classics in a number of ways. They formed the core education of the bureaucratic elite, they provided an important source for imperial authority and they set the philosophical agenda for the dominant Confucian tradition. The classics are also significant for what they do not contain. Many of what are now considered the greatest philosophical works of the Warring States period are classified as masters, not classics; examples include the Zhuangzi, the Xunzi and (until the Song dynasty) even the Mengzi.

1 Early ideas of textual authority

The practice of citing textual authority can be traced to the Zhou dynasty (1121-222 BC). There is evidence that the practice of poetic quotation of selected portions from the *Shijing (Book of Songs)* was highly valued by the Zhou elite. Quotation from the *Shijing* (and, to a lesser extent, the *Shujing (Book of Documents)*) could be used to lend the authority of tradition to particular utterances. The ambiguity of quotation could also be used in a variety of diplomatic, domestic and political contexts to diffuse tension or criticize a hierarchical superior. In this sense, the origins of the classics involve the notion of criticizing authority. In the *Analects* of Confucius, three texts are commonly appealed to as sources of authoritative utterance: the *Shijing*, the *Shujing* and the *Yijing (Book of Changes)*. While the *Analects* thus contains a distinct notion of authoritative texts, there is no reference to an explicit classical canon of any kind.

Four works mark the beginning of what was later to become the Chinese classical tradition: the *Shujing*, the *Yijing*, the *Chunqiu (Spring and Autumn Annals)* and the *Shijing*. While tradition ascribes the compilation of these works to Confucius, they in fact date from the Zhou court between the ninth and sixth centuries BC. These annals and collections of poetry and ritual reflect the values and concerns of the Zhou aristocracy: sacrifice and warfare. The *Shujing* contains terse accounts of investitures, royal decisions and pronouncements and ritual dances, one of which, for example, celebrates the accession of King Wen of Zhou over the preceding Shang dynasty. The *Shijing* also contains sacrificial and ritual hymns from the Zhou court, as well as a variety of other popular songs. The *Chunqiu* is a series of ritual documents in careful chronological order. These may have originated as an historical archive of fire divinations, performed by heating pieces of tortoise shell or animal bone on which a divinatory inquiry had been inscribed. A second form of divination entailed the manipulation of yarrow stalks, where even and odd numbers corresponded to broken or unbroken lines that represented yin and yang respectively, in a complex scheme of sixty-four permutations. Of the various yarrow-stalk divination systems (*yi*) used in various contexts, only the Zhou system has been preserved; the *Zhouyi (Zhou Changes)* forms the core of the *Yijing* (see Yijing).

2 Definition of the canon

Mention of a canon of six classics first appears in the *Zhuangzi*, itself a particularly non-canonical text (see Zhuangzi). Chapter 14 of this work refers to six classics: *Shijing*, *Shujing*, *Lijing (Book of Rites)*, *Yuejing (Book of Music)*, *Yijing* and the *Chunqiu*. (A separate reference in the previous chapter mentions, but does not name, twelve classics.) The late Warring States philosopher Xunzi also refers to the same six, which he helped establish as canonical works of early Ruhist (Confucian) schools. The First Emperor of the Qin proscribed five of the six in 213 BC; the exception was the *Yijing*, presumably for its divinatory content. This imperial proscription, combined with the destruction of the Imperial Library in 207 BC, caused a serious breach in the transmission of these works, and the *Yuejing* may have been lost during this period.

The category of classics (*jing*) first appeared during the Han dynasty as an effort to create or re-create a textual canon. In 136 BC, under the influence of the philosopher Dong Zhongshu, the Emperor Wu of the Han promulgated a set of five classics associated with Confucius. He established five academic chairs for the specific study of five works, the *Yijing*, *Shijing*, *Shujing*, *Lijing* and *Chunqiu*. He further restricted the appointment of academicians to the specific study of those five classics. In 124 BC, also under the influence of Dong Zhongshu, he founded the Taixue, an imperial academy for the training and examination of students in those works, as a way of recruiting able officials. Students completing work at the academy were examined on the Five Classics in a written examination marking the beginning of the state examination system. These measures also established an explicit concept of a canon of classical texts in China, the importance of which cannot be underestimated. They raised Han

Confucianism to the status of a state religion, and established these works as the basis for the education of officials. Despite this official recognition of both texts and institutions, however, the Han dynasty saw considerable debate about the legitimacy of both texts and editions of texts. While the Five Classics was the most frequent articulation of the Han Confucian canon, a set of seven classics added the *Analects* of Confucius and the *Xiaojing (Book of Filial Piety)* to the original five.

Also during the Han period a classification system of six classes of literature emerged, of which the category of the classics was first and foremost. The bibliographical chapter of the *Hanshu (History of the Former Han Dynasty)*, edited by Ban Gu (AD 32-92) and based on an earlier compilation by Liu Xin (d. AD 23) lists the following six classes of literature: (1) the Six Classics (*jing*), (2) the masters (*zi*), grouped into nine schools, of which the Ru were most prominent, (3) poetry, (4) strategy manuals, (5) the numerical arts and (6) ‘prescriptive techniques’. Much of the philosophical tradition of China is to be found not among the classics but among the masters (*zi*). These works are attributed to individuals to whose surnames the suffix *zi* or ‘Master’ has been added. Recent scholarship has shown that many of these texts are probably of composite authorship and are less the works of individuals than the core texts of ‘schools’ or lineages based on textual transmissions of these works (see Guanzi; Han Feizi;Huainanzi; Mencius; Mozi; Sunzi; Xunzi; Zhuangzi).

By Tang times, the number of classics expanded through the treatment of commentaries and component sections as individual works. The *Yili (Ceremony and Rites)*, *Zhouli (Rites of Zhou)* and the *Liji (Book of Rites)*, works on ritual that had presumably comprised the original *Lijing*, were now treated as separate works. The *Gongyangzhuan (Gongyang Annals)*, *Guliangzhuan (Guliang Annals)* and *Zuozhuan (Zuo Annals)*, three commentary editions of the *Chunqiu*, replaced the original in the classical canon. In addition, the *Erya* was recognized as a classic, as were the *Analects* and *Xiaojing*. During the Song dynasty the *Mengzi* was added, bringing the total to thirteen classics.

3 The thirteen classics

The *Shujing* or *Shangshu* (Book of Documents, or Documents), is a heterogeneous compendium of historical documents. Tradition names Confucius as the editor of this ‘classic’ of Chinese political philosophy, much of which was forged during the fourth century BC. The *Shujing* contains terse accounts of investitures, royal decisions and pronouncements, and ritual dances. It also contains dialogues between kings and ministers, including the advice of ministers to kings, all of which provides some insight into Warring States and Han political philosophy. The *Shujing* contains sections of philosophical exposition on principles of government which could be taken either as a standard of conduct for a discerning and virtuous emperor, or as a guide to conduct for a loyal minister.

The *Yijing (Book of Changes)* consists of an ancient manual of divination of uncertain provenance (the *Zhouyi*, or Zhou Changes) and a group of Warring States period commentaries (the Ten Wings). The divination system employs sixty-four permutations of broken (*yin*) and straight (*yang*) lines combined in six positions. As early as the Han dynasty, the *Yijing* was recognized as foremost among the classics and has functioned as a symbolic language of change and transformation. The *Yijing* provides a theoretical conception of the world as both *sui generis* and in a constant process of change and transformation according to discernible patterns of two complementary forces. The *yin-yang* theory first articulated in the *Yijing* provided the basis for later cosmologies of change and formed part of the foundation of early Daoist thought, medicine and the development of scientific thought in China (see Daoist philosophy; Yijing; Yin-yang).

The *Shijing* is the earliest anthology of poetry in China. According to one tradition, Confucius selected compendium of 305 poems from a larger collection of three thousand poems of popular origin gathered at the Zhou court and dating from the beginning of the Zhou dynasty (tenth century BC) to the middle of the Spring and Autumn period (seventh century BC). It contains four sections of differing geographic origin and contents. The airs (*feng*), a northern collection of songs of community life, farming, feasting and love, preserve a subversive quality of ‘pre-Confucian’ society that at times conflicts with later use by Confucian exegetes. The odes or elegantiae (*ya*) are hunting and banquet songs and political complaints and satires, probably from the Zhou court. The lauds (*song*) are dynastic hymns used in court ceremony and sacrifice. Whatever the authorship of these diverse materials, quotation from the *Shijing* was widely used to support Ruhist argumentation and imperial authority within Han and Warring States texts. The Great Preface to the *Shijing* also contains a remarkable concise formulation of a theory of poetics and aesthetics.

The *Lijing (Classic of Rites, or Three Rites)* is a compendium of Zhou dynasty court religious and social ritual. It consists of three works: the *Zhouli (Rites of Zhou)*, the *Yili (Ceremony and Rites)* and the *Liji (Book of Rites)*. The *Lijing* was the worst preserved of the Five Classics, and the three extant versions of these works on ritual all date from later periods. The *Yili* dates from some time after Confucius, the *Zhouli* dates from the late Warring States period and the extant form of the *Liji* dates from the Han dynasty. The *Yili* provides details of ceremonies from the Spring and Autumn period; it includes detailed descriptions of sacrificial rituals for auspicious occasions, rituals for mourning and funerals, rituals for the reception of guests (including state visits and royal audiences) and rituals for various festive occasions (such as marriage, the capping ceremony for young men, archery contests and the reception of ministers). The *Zhouli* is an idealized description of the Zhou bureaucracy. It describes the ministries that formed the Zhou government, including such topics as the supervision control and appointment of officials, agriculture and marriage, state and religious rites, defense and security, legal administration and public works and civil engineering. It contains many interpolations by Han scholars and is of dubious historical value, but its detailed descriptions provide important insights into early Confucian views of government, law and administration. The *Liji* contains material that ranges in date from the late Warring States to the early Han. It provides information about both ritual and Ruhist philosophy, and contains a variety of Ruhist teachings from the late Zhou through the Han. It also contains essays on general principles of ritual. This section includes two works that the neo-Confucian scholar Zhu Xi established as two of the Confucian Four Books: the *Zhongyong (Doctrine of the Mean)* and the *Daxue (Great Preface)* (see Zhongyong; Daxue). Other sections include descriptions on the regulation of the rites, *yin-yang* theory, mourning, worship and sacrifice, auspicious rites, auspicious occasions and the *Yuejing*.

The *Chunqiu (Spring and Autumn Annals)* is a historical chronicle of the reigns of twelve rulers of the state of Lu. It provides a terse and impersonal factual chronicle of diplomatic and political events from 722-481 BC. According to Mencius, Confucius composed the *Chunqiu* as a historical object lesson of the lawlessness of his own age and his own state of Lu. The Chunqiu was of particular importance because of its use by the Han philosopher Dong Zhongshu, a major formulator of official Han Confucian metaphysics (see Confucian philosophy, Chinese). Dong claimed to have derived his ideas from the *Chunqiu*, and followed Mencius in believing that Confucius, having understood the heavenly patterns that governed the cosmos, applied them to human history. Dong gave Confucius the status of a ‘hidden king’, the moral and metaphysical arbiter whose praise and blame passed correct judgments on the actions of the kings of the past.

The *Chunqiu* was one of the Han dynasty’s Five Classics. From the Tang dynasty onwards it was transmitted through three major commentaries, the *Gongyangzhuan (Gongyang Annals)*, the *Guliangzhuan (Guliang Annals)* and the *Zuozhuan (Zuo Annals)*. The *Guliangzhuan* and *Gongyangzhuan* both use question-and-answer dialogue to provide a morally normative interpretation of the *Chunqiu*. Tradition attributes the origins of the *Gongyangzhuan* to the oral teachings of Zixia, a disciple of Confucius; the *Guliangzhuan* borrows from and elaborates on the *Gongyangzhuan*. Both cover the historical span of the *Chunqiu*, and both employ the ‘praise and blame’ historiography of the latter. The *Zuozhuan* covers a longer period than the other two commentaries (722-453 BC) and contains more historical information. Its date and authorship are the source of considerable controversy.

The *Analects* of Confucius (*Lunyu*) is a collection of anecdotes, sayings and short dialogues, mostly between Confucius and his disciples. While Confucius himself claimed to be a transmitter of Zhou values rather than an innovator, the central concerns of the *Analects* moved away from divination of the will of spirits toward the ethics of the human world and the establishment and maintenance of political and social order. Confucius formulated new notions of humane benevolence (*ren*) and the right conduct of superior individuals (*junzi*) that were to remain hallmarks of later Confucian thought (see Confucian philosophy, Chinese; Confucius).

The *Xiaojing (Book of Filial Piety)* is a short dialogue between Confucius and his disciple Zengzu on filiality towards parents (by their children) and, by analogy, rulers (by their subjects). It is didactic in tone, and describes the virtues of filial piety as ostensibly practised during the golden age of the Zhou. It includes quotations from the *Shijing (Book of Songs)* and is written in question and answer form. The *Xiaojing* was included among the Han seven classics, possibly because its articulation of appropriate behaviour of subjects toward rulers made it an attractive text for imperial patronage. Imperial patronage, as much as philosophical esteem within the tradition, conferred and maintained its status as a classic.

The *Erya* is a dictionary of Zhou terms. Various traditional accounts ascribe it to the Duke of Zhou, Confucius and

the latter's disciple Zixia; recent scholarship dates it to the Qin or early Han dynasty. Its most important influence is in the nineteen categories into which words are grouped. These categories became the basis for the categories of encyclopedias and similar works. While many of the glosses in the *Erya* are cryptic, as a text it contains a variety of 'abstract' terms as well as entries on kinship terms, musical terminology, astronomic and calendrical terms, geographical terms, names of grasses, herbs, trees, wild and domestic animals and legendary animals. It was treated as an authoritative guide to the language of the classics.

The *Mengzi* consists of a series of extended conversations and debates between Mencius and various rulers, opponents and disciples. Mencius developed concepts that had been mentioned or suggested but not developed in the *Analects*. He stressed the inherent goodness of human nature and elaborated a theory of four inherent 'sprouts' of goodness: humane benevolence (*ren*), knowledge or wisdom (*zhi*), ceremony or ritual (*li*) and rectitude (*yi*). Mencius held that these sprouts could, with correct nurture, mature in any individual and create the intuitive ability that is the basis for moral conduct and moral judgment. Mencius also attacked a variety of opponents of Confucius, including the Mohists (see Mohist Philosophy), the followers of Yangzhu and military strategists. The first imperial recognition accorded to the works of Mencius was by the Emperor Wen of the Han dynasty, who established a chair for the study of the *Mengzi*. The *Mengzi* is listed in the Han bibliography among the masters rather than the classics, and only appears in the classical canon during the Song dynasty.

4 Significance

During the approximately fifteen hundred years between Confucius and the Song dynasty, several classical schemata gained and lost sway, including the Six Classics of the Warring States period, the Five Classics of the Han, the Nine Classics of the Tang, and the Thirteen Classics of the Song. Nevertheless, although the content and size of Confucian classical canons changed over time, the notion of classics has been relatively continuous since the beginnings of philosophical thought in China. The classics are of philosophical importance because they articulate central ideas, values and dispositions of Chinese culture. They are also important because the definition of the classical canon effectively excluded other works of considerable philosophical significance.

The Chinese Classics functioned as a classical canon in several senses. First, they provided a textual authority that could serve as a source of intellectual or political authority and legitimation for philosophers and rulers alike. For example, Confucius, Mencius and Xunzi all quote extensively from the *Shijing* and *Shujing*. (Xunzi differs from the other two in that he both quoted poetry as authoritative discourse and composed original verse, and is even listed at the head of a school of poets.)

Second, the classics formed the core of the education system that defined the bureaucratic elite. Despite competing claims from other textual canons and variations in emphasis, these works held a central role in the education (and definition) of the bureaucratic elite. The classical compendia defined the knowledge and the political and moral orientation of a learned scholar or official. Because they formed the root of the state examination system, their mastery was an important prerequisite for political office. Memorizing the classics was a large part of the early intellectual and moral education of boys. Additional 'classics' were developed for the separate education of women. The *Nuxiaojing (Book of Filial Piety for Women)* and the *Nulunyu (Analects for Women)* are both Tang works written by women for the education of women, the latter by a consort of the Tang emperor Dezong in the eighth century. These works were taken as repositories of knowledge, guides to conduct and a repertoire for use in situations that required literary elegance.

Finally, the classics set the philosophical agenda for what came to be known as the early Confucian tradition. They were centrally concerned with two related problems: how to maintain political and social order, and how to persuade rulers to voluntarily limit their own power. As such, all the classics functioned in a strongly hierarchical system and their style of argument or exposition reflects the social context in which they operated, for example, by the prevalence of persuasion over debate.

The thirteen classics are also noteworthy for what they do not contain. Many of the recognized classical works of Chinese philosophy appear in traditional classifications not among the classics but among the masters. Ruhist works include the *Mengzi* (until the Song dynasty) and *Xunzi* (see Mencius; Xunzi). Daoist 'masters' include the *Daodejing* and the *Zhuangzi* (see Daodejing; Zhuangzi). The one surviving Mohist work is the Mohist Canon (see Mohist Philosophy; Mozi); the masters conventionally associated with the so-called Legalist school include Han

Fei, Lord Shang, and Shen Buhai (see Legalist philosophy, Chinese; Han Feizi). Less well-known but important masters include the *Huainanzi* (compiled at the court of Liu An *circa* 140 BC) and the *Lushi chunqiu (Springs and Autumns of Master Lu)*, attributed to Lu Buwei (d. 235 BC) (see Huainanzi).

These works in turn function as the ‘classics’ - in the sense of a master text that defines a school - of several traditions. For example, the *Daodejing*, attributed to the apocryphal sage Laozi but probably dating from about 250 BC, is one of the canonical works of several Daoist traditions and may be the most widely read Chinese text outside China. Still other works, not listed among the masters but titled as classics or masters became the foundational texts for specific schools of expertise, for example, the *Huangdi neijing (Yellow Thearch's Classic of Internal Medicine)*, a medical compendium dating from approximately 100 BC that became the master text for the high textual tradition of Chinese medical practice.

See also: Confucian philosophy, Chinese; Confucius; Daoist Philosophy; Aesthetics, Chinese; Daodejing; Dong Zhongshu; Mencius; Yijing

LISA RAPHALS

References and further reading

The dating of all classic texts is problematic and controversial. The dates of compilation given before are approximations.

Analects (4th-2nd century BC), trans. J. Legge, *The Confucian Analects*, Oxford, Clarendon Press, 1893; repr. in *The Chinese Classics*, vol. 1, Hong Kong: Hong Kong University, 1960; trans. A. Waley, *The Analects of Confucius*, New York: Random House, 1966; trans. D.C. Lau, *Confucius: The Analects (Lun yu)*, London: Penguin, 1979.(Lau is the standard English translation of the *Analects (Lunyu)*. The Legge edition contains facing Chinese text and notes.)

Erya (3rd century BC), in *Erya yinde (Index to the Erya)*, Harvard-Yenching Institute Sinological Index Series, supplement 18, Beijing, 1941; ed. D.C. Lau and Chen Fong-ching, *A Concordance to the Erya and Xiaojing*, ICS Series, Hong Kong: Commercial Press, 1994.(There are no translations of this work.)

Gongyangzhuan (Gongyang Annals) (2nd century BC), trans. G. Malmqvist, ‘Studies on the Gongyang and Guliang Commentaries’, *Bulletin of the Museum of Far Eastern Antiquities* 43, 1971: 67-222; 47, 1975: 19-69; and 49, 1977: 33-215.(Partial translation of the Gongyang and Guliang commentaries to the *Chunqiu (Spring and Autumn Annals)*.)

Guliangzhuan (Guliang Annals) (2nd-1st century BC) trans. G. Malmqvist, ‘Studies on the Gongyang and Guliang Commentaries’, *Bulletin of the Museum of Far Eastern Antiquities* 43, 1971: 67-222; 47, 1975: 19-69 and 49, 1977: 33-215.(Partial translation of the Gongyang and Guliang commentaries to the *Chunqiu (Spring and Autumn Annals)*.)

Liji (Book of Rites) (2nd century BC-2nd century AD), trans. J. Legge, *Li chi*, in *Sacred Books of the East*, vols 26-8, New York: University Books, 1926; repr. 1967; trans. S. Couvreur, *Li Ki: ou mémoires sur les bienséances et les cérémonies*, Ho Kien Fou [Hejianfu]: Imprimerie de la Mission Catholique, 1913, 2 vols.(English and French translations respectively of the *Liji*.)

Mengzi (2nd century BC-2nd century AD), trans. J. Legge, *The Works of Mencius*, in *The Chinese Classics*, vol. 2, l895; trans. D.C. Lau, *Mencius*, New York: Penguin, 1970.(Lau is the standard English translation of the *Mengzi*. Legge contains facing Chinese text and notes.)

Shijing (Book of Songs) (5th century BC), trans. J. Legge, *The She King, or The Book of Poetry*, in *The Chinese Classics*, vol. 4, 1872; trans. B. Karlgren, *The Book of Odes*, Stockholm: Museum of Far Eastern Antiquities, 1950; trans. A. Waley, *The Book of Songs*, New York: Grove Press, 1960.(Karlgren's translation of the *Shijing* is highly literal; that of Waley is a literary translation.)

Shujing (Book of Documents) (5th century BC-4th century AD), trans. J. Legge, *The Shoo King or The Book of Historical Documents*, in J. Legge (ed.) *The Chinese Classics*, vol. 3, 1865; repr. Hong Kong: Hong Kong University Press, 1960; trans. B. Karlgren, *The Book of Documents*, Stockholm: Museum of Far Eastern Antiquities, 1950.(Karlgren translates the verifiably pre-Han portions of the *Shujing*, also known as the *Shangshu*. Legge does not distinguish between pre-Han materials and later interpolations.)

Xiaojing (Book of Filial Piety) (2nd or 1st century BC), trans. J. Legge, *The Hsiao ching*, in F.M. Muller (ed.) *Sacred Books of the East*, vol. 3, Oxford: Oxford University Press, 1899; trans. I. Chen, *The Book of Filial Piety*, London: John Murray, 1908, New York: E.P. Dutton, 1908-9; ed. P.K.T. Sih, trans. M.L. Makra, *The Hsiao ching*, New York: St. John's University Press, 1961.(Translations of the *Xiaojing*.)

Yijing (Book of Changes) (3rd-2nd century BC), trans. J. Legge, *The I Ching*, in F.M. Muller (ed.) *Sacred Books of*

the East, vol. 16, Oxford: Oxford University Press, 1899; trans. R. Wilhelm and C.F. Baynes, *The I Ching or Book of Changes*, Bollingen Series 19, Princeton, NJ: Princeton University Press, 1950.(Wilhelm and Baynes is the standard English translation of the *Yijing*.)

Yili (Ceremony and Rites) (2nd or 1st century BC), trans. J. Steele, *The I li or Book of Etiquette and Ceremonial*, London: Probsthain and Co., 1917; trans. S. Couvreur, *Cérémonial*, Hsien hsien [Xian Xian]: Imprimerie de la Mission Catholique, 1916; repr. Paris: Cathasia, 1951. (English and French translations respectively of the *Yili*.)

Zhouli (Rites of Zhou) (2nd or 1st century BC), trans. E. Biot, *Le Tcheou-li ou Rites des Tcheou*, Paris: Imprimerie Nationale, 1851; repr. Taipei: Ch'eng-wen, 1975.(French translation of the *Zhouli*.)

Zuozhuan (Zuo Annals) (2nd-1st century BC), trans. J. Legge, *The Ch'un Ts'ew with the Tso Chuen*, in *The Chinese Classics*, vol. 5, 1872; repr. Hong Kong: Hong Kong University Press, 1960; trans. S. Couvreur, *Tch'ouen st'ieou et Tso tschouan*, Ho Kien Fou, 1914, 3 vols; repr. Paris: Cathasia, 1951; trans. B. Watson, *The Tso Chuan: Selections from China's Oldest Narrative History*, New York: Columbia University Press, 1989. (Legge and Couvreur are the English and French translations of the *Chunqiu (Spring and Autumn Annals)* and the *Zuozhuan* commentary. Watson is a new partial translation of the *Zuozhuan*.)

Chinese philosophy

Any attempt to survey an intellectual tradition which encompasses more than four thousand years would be a daunting task even if it could be presumed that the reader shares, at least tacitly, many of the assumptions underlying that tradition. However, no such commonalities can be assumed in attempting to introduce Asian thinking to Western readers. Until the first Jesuit incursions in the late sixteenth century, China had developed in virtual independence of the Indo-European cultural experience and China and the Western world remained in almost complete ignorance of one another.

The dramatic contrast between Chinese and Western modes of philosophic thinking may be illustrated by the fact that the tendency of European philosophers to seek out the being of things, the essential reality lying behind appearances, would meet with little sympathy among Chinese thinkers, whose principal interests lie in the establishment and cultivation of harmonious relationships within their social ambiance. Contrasted with Anglo-European philosophic traditions, the thinking of the Chinese is far more concrete, this-worldly and, above all, practical.

One reason for this difference is suggested by the fact that cosmogonic and cosmological myths played such a minor role in the development of Chinese intellectual culture and that, as a consequence, Chinese eyes were focused not upon issues of cosmic order but upon more mundane questions of how to achieve communal harmony within a relatively small social nexus. The rather profound linguistic and ethnic localism of what Pliny the Elder described as a 'stay-at-home' China, reinforced by a relative freedom from intercultural contact, generated traditional radial communities in which moral, aesthetic and spiritual values could remain relatively implicit and unarticulated. By contrast, in the West these norms had to be abstracted and raised to the level of consciousness to adjudicate conflicts occasioned by the complex ethnic and linguistic interactions associated with the development of a civilization rooted almost from the beginning in the confluence of Greek, Hebrew, and Latin civilizations.

The distinctive origins and histories of Chinese and Western civilizations are manifested in a number of important ways. The priority of logical reasoning in the West is paralleled in China by the prominence of less formal uses of analogical, parabolic and literary discourse. The Chinese are largely indifferent to abstract analyses that seek to maintain an objective perspective, and are decidedly anthropocentric in their motivations for the acquisition, organization and transmission of knowledge. The disinterest in dispassionate speculations upon the nature of things, and a passionate commitment to the goal of social harmony was dominant throughout most of Chinese history. Indeed, the interest in logical speculations on the part of groups such as the sophists and the later Mohists was short-lived in classical China.

The concrete, practical orientation of the Chinese toward the aim of communal harmony conditioned their approach toward philosophical differences. Ideological conflicts were seen, not only by the politicians but by the intellectuals themselves, to threaten societal well-being. Harmonious interaction was finally more important to these thinkers than abstract issues of who had arrived at the 'truth'. Perhaps the most obvious illustration of the way the Chinese handled their theoretical conflicts is to be found in mutual accommodation of the three emergent traditions of Chinese culture, Confucianism, Daoism and Buddhism. Beginning in the Han dynasty (206 BC-AD 220), the diverse themes inherited from the competing 'hundred schools' of pre-imperial China were harmonized within Confucianism as it ascended to become the state ideology. From the Han synthesis until approximately the tenth century AD, strong Buddhist and religious Daoist influences continued to compete with persistent Confucian themes, while from the eleventh century to the modern period, Neoconfucianism - a Chinese neoclassicism - absorbed into itself these existing tensions and those that would emerge as China, like it or not, confronted Western civilization.

In the development of modern China, when Western influence at last seemed a permanent part of Chinese culture, the values of traditional China have remained dominant. For a brief period, intellectual activity surrounding the May Fourth movement in 1919 seemed to be leading the Chinese into directions of Western philosophic interest. Visits by Bertrand Russell and John Dewey, coupled with a large number of Chinese students seeking education in Europe, Great Britain and the USA, promised a new epoch in China's relations with the rest of the world. However, the Marxism that Mao Zedong sponsored in China was 'a Western heresy with which to confront the West'. Mao's Marxism quickly took on a typically 'Chinese' flavour, and China's isolation from Western

intellectual currents continued essentially unabated.

1 Chinese thinking as *ars contextualis*

Our traditional Western senses of order are grounded upon cosmogonic myths that celebrate the victory of an ordered cosmos over chaos. Chaos is a 'yawning gap', a 'gaping void'; it is an emptiness or absence, a nothingness; it is a confused mass of unorganized surds. Hesiod's *Theogony* tells how the yawning gap of Chaos separating Heaven and Earth was overcome by Eros - love thereby creating harmony (see Hesiod). The Book of Genesis tells how, from a 'dark, formless void', order was created by divine command. In the *Timaeus*, Plato's Demiurge 'persuades' the disorganized, intransigent matter into reasonable order, providing 'a victory of persuasion over necessity'. Classical Chinese culture, on the other hand, was little influenced by myths which contrasted an irrational Chaos with an ordered Cosmos. The relative unimportance of cosmogonic myths in China helps to account for the dramatically different intellectual contexts from which the Chinese and Western cultural sensibilities emerged.

In the Western tradition, thinking about the order of things began with questions such as 'What kinds of things are there?' and 'What is the nature (*physis*) of things?' This inquiry, which later came to be called 'metaphysics', took on two principal forms. One, which the scholastics later termed *ontologia generalis* (general ontology), is the investigation of the most essential features of things - the *being* of beings. A slightly less abstract mode of metaphysical thinking, *scientia universalis* (universal science), involves the attempt to construct a science of the sciences, a way of knowing which organizes the various ways of knowing the world about us. Both general ontology and universal science interpret the order of the cosmos. Both suppose that there are general characteristics - the being of beings, or universal principles - which tell us how things are ordered.

Neither of these forms of metaphysical thinking were influential in classical China. One reason for their unimportance is reflected in the character of the Chinese language. Simply put, the classical Chinese language does not employ a copulative verb. The Chinese terms usually used to translate 'being' and 'not-being' are *you* and *wu* (see You-wu). The Chinese *you* means, not that something 'is' (*esse* in Latin) in the sense that it exists in some essential way; it means rather that 'something is present'. 'To be' is 'to be available', 'to be around'. Likewise, *wu* as 'to not be' means 'not to be around'. Thus the Chinese sense of 'being' overlaps 'having'. A familiar line from the Daoist classic the *Laozi* or *Daodejing*, which is often translated 'Not-being is superior to Being', should more responsibly be translated as 'Not-having is superior to having' (or as a Marxist-inspired translator has rendered it, 'Not owning private property is superior to owning private property'). The Chinese language disposes those who employ the notions of *you* and *wu* to concern themselves with the presence or absence of concrete particular things and the effect of having or not having them at hand. Even in recent centuries, when the influence of translating Indo-European culture required the Chinese language to designate a term to do the work of the copula, the choice was *shi*, meaning 'this', thus indicating proximity and availability rather than 'existence'. One must assume that the practical, concrete disposition of Chinese thinkers is both cause and consequence of this characteristic feature of linguistic usage.

Perhaps the best designation for the most general 'science' of order in the Chinese tradition would be *ars contextualis*. 'The art of contextualizing' contrasts with both *scientia universalis* and *ontologia generalis*. Chinese thinkers sought the understanding of order through the artful disposition of things, a participatory process which does not presume that there are essential features, or antecedent-determining principles, serving as transcendent sources of order. The art of contextualizing seeks to understand and appreciate the manner in which particular things present-to-hand are, or may be, most harmoniously correlated. Classical Chinese thinkers located the energy of transformation and change within a world that is *ziran*, autogenerative or literally 'so-of-itself', and found the more or less harmonious interrelations among the particular things around them to be the natural condition of things, requiring no appeal to an ordering principle or agency for explanation.

The dominant Chinese understanding of the order of things advertises an important ambiguity in the notion of 'order' itself. The most familiar understanding of order in the West is associated with uniformity and pattern regularity. This 'logical' or 'rational' ordering is an implication of the cosmological assumptions which characterize the *logos* of a cosmos in terms of causal laws and formal patterns. A second sense of order is characterized by concrete particularities whose uniqueness is essential to the order itself. No final unity is possible in this view since, were this so, the order of the whole would dominate the order of the parts, cancelling the

uniqueness of its constituent particulars. Thus, 'aesthetic' order is ultimately acosmological in the sense that no single order dominates.

The crucial difference between these two senses of order is that in the one case there is the presumption of an objective standard which one perforce must instantiate; in the other, there is no source of order other than the agency of the elements comprising the order. In the West, mathematical order has been thought the purest. In China, by contrast, any notion of order which abstracts from the concrete details of this-worldly existence has been seen as moving in a direction of decreasing relevance. Rational order depends upon the belief in a single-ordered world, a cosmos; aesthetic order speaks of the world in much less unitary terms. In China, the 'cosmos' is simply 'the ten thousand things'. The belief that the things of nature may be ordered in any number of ways is the basis of philosophical thinking as *ars contextualis*.

Each of these understandings of order existed at the beginnings of both Western and Chinese cultures, and both have persisted as interpretative options within them. It so happens that in the course of their respective histories, the two cultures made distinctly different choices which led to the variant senses of order as grounds for the organizing of personal, social and cosmic environs. These differing senses of order are reflected in the fact that the nineteenth-century Japanese had to coin the term *tetsugaku* (philosophy) to translate the Western philosophic tradition and to recover its Japanese counterpart, and that this same expression was soon thereafter imported into China as *zhexue* for the same purpose. The need to invent a term to refer to 'philosophy' suggests at least that these cultures had to reorganize patterns of indigenous intellectual experience in ways previously unfamiliar to them.

2 The dominance of correlative thinking

Rational or logical thinking, grounded in analytic, dialectical and analogical argumentation, stresses the explanatory power of physical causation. In contrast, Chinese thinking depends upon a species of analogy which may be called 'correlative thinking'. Correlative thinking, as it is found both in classical Chinese 'cosmologies' (the *Yijing (Book of Changes)*, Daoism, the Yin-Yang school) and, less importantly, among the classical Greeks involves the association of image or concept-clusters related by meaningful disposition rather than physical causation. Correlative thinking is a species of spontaneous thinking grounded in informal and *ad hoc* analogical procedures presupposing both association and differentiation. The regulative element in this modality of thinking is shared patterns of culture and tradition rather than common assumptions about causal necessity.

The relative indifference of correlative thinking to logical analysis means that the ambiguity, vagueness and incoherence associable with images and metaphors are carried over into the more formal elements of thought. In fact, the chaotic factor in the underdetermined correlative order has a positive value as an opportunity for personalization and self-construal. In contradistinction to the rational mode of thinking which privileges univocity, correlative thinking involves the association of significances into clustered images which are treated as meaning complexes ultimately unanalyzable into any more basic components. In the Western tradition we are familiar with correlations such as those present in the humour theory of medicine, Pythagorean numerical correlations, Kepler's correlation of the perfections of the trinitarian God, the world and the soul, and so forth (see Hippocratic medicine; Pythagoras; Kepler, J.). Astrological charts provide the most familiar illustration of correlative thinking. But, as may be seen particularly (although not exclusively) in the discussion of Daoist philosophy, correlativity is not only an anthropocentric mode of signification. The Daoist notion of *de* - 'particular focus' or 'virtuality' - extends the context of signification to all phenomena (see Daoist philosophy; De).

Correlative thinking is the primary instrument in the creation, organization and transmission of the classical curriculum in China, from the *Book of Songs* to the *Analects* to the *Yijing* (see Chinese Classics; Confucius; Yijing). Perhaps the most overt illustrations of the Chinese resort to correlative thinking in the classical period are to be found in the period of the Han dynasty (206 BC-AD 220). During the Han period, vast tables of correspondences were employed to identify and organize the sorts of things in the natural and social world which were thought to provide a meaningful context for one's life. One such set of tables, called 'tables of five', compared 'the five phases' (wood, fire, earth, metal, water), 'the five directions' (north, east, south, west and centre), 'the five colours' (green, red, yellow, white, black), 'the five notes', and so forth. Other types of correlation employed the twelve months, the twelve pitches, the twenty-eight constellations, the heavenly roots and the earthly branches. Such classifications include body parts, psycho-physical and affective states, styles of

government, weather, domestic animals, technological instruments, heavenly bodies and much more.

One of the important devices for making such correlations is the contrast of *yin* and *yang*, literally, 'the shady side' and 'the sunny side' of the mountain (see Yin-yang). These notions were employed to identify alternative patterns of hierarchical relationship. The old teacher, Laozi, is wiser than his young student and hence 'overshadows' him in this respect. Laozi is *yang* and the student is *yin*. The student, however, is stronger physically than the old master, and hence in physical prowess the student is *yang* to Laozi's *yin*. When these various strengths and weaknesses defining the relationship can be balanced to maximum effect, the relationship is most productive and harmonious. It is clear from this illustration that *yang* and *yin* are by no means to be understood as 'cosmic principles' or ontological contrasts rooted in the very nature of things. Rather, they are heuristics helpful in reading and characterizing the world as concretely experienced in a variety of ways.

Though the contents of many correlative schemata are often *prima facie* the same as the subject matters of the Western natural sciences, there is a crucial difference in the manner they are treated. In China, correlations were not employed as a means of dispassionately investigating the nature of things. Correlative descriptions are, in fact, prescriptions. Correlative schemes oriented human beings in a very practical manner to their external surroundings. Thus, the Chinese were concerned less with astronomy than with astrology; they were far more enthusiastic in the development of geomancy than geology. Science was always understood as ultimately subject to prevailing human values.

3 The organization and transmission of knowledge

The importance of correlative modes of thinking in classical Chinese culture is to be found not only in the acquisition of knowledge but also in its organization and transmission. Plato's employment of the 'Divided Line' allowed for the organization of knowledge by appeal to principles of clarity and coherence which were realized with the achievement of systematic unity (see Plato). Aristotle organized the body of the known into theoretical, practical and productive enterprises by appeal to the nature of the *psyche* as a thinking, acting and feeling creature. In each case there is the appeal to objective principles which serve to articulate the realms of knowledge.

There is a stark contrast between these Western models of classification and the one characteristically found in the traditional *leishu* (encyclopedic or classificatory works) of China. Chinese 'categories' (*lei*) are defined not by the presumption of a shared essence defining natural 'kinds', but by an identified functional similarity or association that obtains among unique particulars. Definitions are not framed in the terms of essential features and formal class membership; instead, definitions tend to be metaphorical and allusive, and invariably entail the human subject and human values. It is the earliest reference in the canons of classical literature, rather than principled scientific explanation grounded in the canons of reason and logic, that holds the weight of authority. In these compendia, there is little interest in the objective description of natural phenomena. The concern instead is with the relationship that the various contents of the world have to the social and cultural values which shape the human experience of it. As an example, the contemporary scholar Liang Congjie points out that 'of the fifty-five sections (*bu*) that make up the *Taiping yulan*, six of them - emperors, imperial relatives, officials, human affairs, ancestors, and ceremony - occupy thirty-five per cent of the work' (Hall and Ames 1995).

These Chinese encyclopedic works are hierarchical, with the human being self-consciously at the centre. Further, this 'human being' is no abstraction but is the specific imperial Chinese person, the emperor, who commissioned the work for the benefit of those examination candidates destined to assist him in governing his empire. Again in each category, individual entries begin from the most noble and conclude with the most base: animals begin from 'lion' and 'elephant' and finish with 'rat' and 'fox'; trees begin with 'pine tree' and 'cypress' and end with 'thistles' and 'brambles'. The world is not described objectively through an articulation of exclusive categories and subcategories, but is divided up prescriptively into natural and cultural elements which have an increasing influence on the experience of the Chinese court as they stand in proximity to the centre. In 'naming' (*ming*) his world, the ruler is 'commanding' (*ming*) it to be a certain way.

As for the transmission of knowledge, we are accustomed to the notion that philosophical ideas are disseminated through debate and dialectical interchange in which one theory or vision confronts another and arguments ensue as to the adequacy of each. However, this manner of looking at the transmission and alteration of ideas does not well suit the Chinese context. This is true primarily because such debate is rooted in rationalistic assumptions of the

sort that privilege theories as relatively coherent patterns of belief and practice, which can be articulated apart from the institutions to which they are relevant. Things were quite different in China. Ancient China overcame the threat of the tensions and conflicts attendant upon ethnic and cultural pluralism by employing the contextualizing force of the Chinese language itself as means of transmitting culture. A class of literati developed; a canon of classical works was compiled and instituted along with a continuing commentarial tradition which served to translate and perpetuate the doctrines of these classical works; an examination system based upon these texts was introduced in the early Han period and persisted with relatively little change for two thousand years, being abolished only as recently as 1905.

With the dominance of Confucian orthodoxy in China, methods of adjudicating doctrinal conflicts were refined in a manner which took as its highest value the maintenance of social stability. Beginning in the early Han period, commentaries upon Confucian texts were produced which vied with each other for proximity to the canonical center. The authors of these commentaries were almost never interested in overthrowing the authority of the canon in favour of their own ideas, but sought to enrich the authority of the classic by claiming to better understand its original meaning. Further, since tradition was the sole ready resource for norms and doctrines, critics of a particular doctrine depended as much as proponents upon a shared cultural repository. This is evidenced in the conventional use of canonical allusions to focus social and political critique. The appeal to canonical authority is again a way of reinforcing a sense of shared community, and stands in sharp contrast to dialectical arguments which reference canons of reason or logic (see Chinese Classics).

Characteristic of scholarly dispute after the emergence of Han orthodoxy is a fundamental commitment to mutual accommodation. There is a general distaste for contentiousness and an active cultivation of the art of accommodation. In the exercise of criticism, the ritual basis of order comes into play since rituals serve as patterns of deference which accommodate and harmonize differences in desires, beliefs and actions. Criticism assumes a context of common concern and becomes thereby a cooperative exercise among responsible participants that proceeds to search for alternatives on which all can agree. One important constraint on self-assertiveness, as well as an encouragement for a consensual resolution, is that critics themselves are always implicated in the existing context; hence, any criticism is ultimately self-referential. Contentiousness, by contrast, betrays a concern for personal advantage. Such self-assertion threatens to disrupt rather than reinforce or improve the harmony of the existing context. From the *Analects* on, an appropriateness (*yi*) which respects social context has been advocated as the positive alternative to self-interested benefit (*li*). The proper goal of critical or constructive expression in China, whether it be scholarly, social or political, is the strengthening of communal harmony.

4 Confucius and Confucianism

While there may be some truth to the claim that in the West, every person is born either a Platonist or an Aristotelian, it is A.N. Whitehead’s apothegm, ‘All of Western philosophy is a series of footnotes to Plato’, that, *mutatis mutandis*, resonates best with the Chinese context. For indeed, all of Chinese thinking is a series of commentaries on Confucius. In fact, the importance of Confucius in China may be said to outshine that of Plato in the Western tradition on at least two grounds. First, there is effectively no sort of pre-Confucian philosophic tradition to match that of the Presocratics. Confucius is not a synthesizer of past *thinkers,* but an interpreter and transmitter of past *institutions*, namely, the idealized Zhou rituals and customs which Confucius thought to be the key to social stability. Second, Confucius’ thinking came to ground the tradition of Chinese culture for practically its entire intellectual tradition, from the early phases of the Han dynasty in the second century BC to at least the beginnings of the Republican period in the early twentieth century, and arguably down to the present day in a decidedly Chinese form of Marxism (see Marxism, Chinese).

The philosophy of Confucius begins from some basic assumptions, several of which can be derived from the following passage from the *Analects*:

> The Master said: ‘Lead the people with administrative injunctions and put them in their place with penal law, and they will avoid punishments but will be without a sense of shame. Lead them with excellence and put them in their place through roles and ritual practices, and in addition to developing a sense of shame, will order themselves harmoniously.’

(*Analects* 2/3)

First, Confucius believed in the radical malleability of the nascent human being through education and cultivation (see Self-cultivation in Chinese philosophy). Humanity for Confucius is not defined by anything 'given'; there is no essential human nature. Becoming human is a cultural achievement. Second, the formal instrument for pursuing personal refinement and self-articulation is *li*, often translated as ritual practice or propriety (see Law and ritual in Chinese philosophy). Propriety, which includes everything from etiquette to social roles and institutions to the rites of life and death, is the underlying syntax of community. These formal structures reside in the conduct of those members of the community who serve as models of propriety.

The ultimate value of human experience lies in 'becoming a quality person' (*ren*), where the character which represents this accomplishment is constituted by 'person' and the numeral, 'two', suggesting its fundamentally social nature. It is because a person is shaped and articulated as a specific complex of roles and relationships that the Confucian person is irreducibly social. This social definition of person makes the promise of communal approbation an important encouragement for proper conduct, and the threat of shame an equally effective deterrent against undesirable conduct. Further, communal living becomes increasingly meaningful through the deepening quality of those particular relationships which constitute it: *this* person's son, *that* person's husband, *this* person's neighbour.

The goal of overcoming selfishness, fundamental in classical Confucianism, is not designed to be altruistic. The premise here is that selfishness is the greatest obstacle to the realization of one's social self. Since personal, familial, communal, political and even cosmic order are all coterminous and mutually entailing, commitment to community, far from being self-abnegating, is the road to personal fulfillment.

Excellence or virtue (*de*) achieved by members of the community empowers them as likely models of propriety for succeeding generations. Because the authority of community so constructed is internal to it, the community is self-regulating, dependent for its effectiveness upon authoritative leaders rather than the application of some external apparatus such as law and punishment. It is well worth noting that the importance of exemplary models of propriety in a Confucian society contrasts readily with the stress upon the resort to ethical principles so typical of rationalized societies. In the West, exemplary persons - Socrates or St Francis - are not typically thought to be ends in themselves; their lives point beyond to transcendent realities which ground their virtues. In Confucian societies, however, the sage Kings Yao and Shun, the Duke of Zhou - and Confucius himself - are self-realized individuals who serve as distinctly immanent and historical individuals whose lives constitute models for emulation.

The distinction between a society of principles and a society shaped by models of propriety helps us to understand the distinctly 'aesthetic' quality of Confucian morality. Propriety leads to 'proper' conduct in one's relationships by at once reinforcing traditionally appropriate norms while at the same time insisting that they be internalized and 'made one's own' (*yi*). The notion of propriety or 'rightness' (*yi*) in a Confucian society, since it applies always within a social context, must involve notions of 'harmony'. Proper actions are 'fitting' in the sense that they fit and harmonize with other such actions. The notion of the 'right' action, therefore, has much in common with the artist's choice of the 'right' brush or the 'right' colour in the execution of a painting.

The failure to understand the aesthetic character of Confucian ethics has reinforced the tendency for Western philosophers to understand Confucian ritualization (*li*) as the imposition of external guides to conduct, mere forms imposed upon one from outside. Hegel's depiction of China as a culture without *Geist* in his *Philosophy of History* is representative of interpretations by the best minds of Europe and America. This truncated reading has in turn perpetuated the stereotypical opinion of Confucius as a purveyor of trite moral truisms, rather than as a founder of a social order which, by its dependence upon the sort of balanced complexity associated with aesthetic creations, has lasted longer than any other on the face of the planet.

5 Philosophical Daoism

Daoism is a complex movement in early China (see Daoist philosophy). A proto-Daoist religious sensibility seems to have been a stratum of Chinese popular culture centuries before the emergence of 'religious Daoism' as a formal iconoclastic movement in the second century AD. During the late Warring States period, Daoism developed a sublimated and sophisticated intellectual dimension as a response to rival traditions of thought. Because this school of political and philosophical anarchism was articulated in two primary compilations, the *Daodejing* (or *Laozi*) and the *Zhuangzi*, it came to be known as 'Lao-Zhuang' Daoism (see Daodejing and Zhuangzi).

The central message of this school is captured by the title of the *Daodejing* - literally, 'the classic of *dao* and *de*'. In fact, the name '*dao*-ism' itself is an abbreviation of the earliest designation of this tradition as '*dao-de*-ism', reflecting the core question which pervades the Daoist texts: what is the relationship between *dao* and *de*? This can be interpreted and restated as: what constitutes excellence (*de*) and how does one achieve it within one's particular place (*de*) in the world (*dao*)? Since one's 'place' is both spatial and temporal, de is the excellence achieved as one treads one's 'pathway' (or *dao*) through life (see Dao; De).

A central Lao-Zhuang complaint against rival Confucianism is ecological, denouncing the anthropocentric limits it imposes on personal realization. While early Confucianism argues that human beings are the product of harmoniously orchestrated interpersonal relationships, the Daoists insist that the relational definition of humanity be extended to encompass the world more broadly. Human beings inhere in social, cultural *and* natural environments, and these environments are continuous and mutually shaping. The rhythm and regularity of human community is embedded in, and hence must be responsive to, the cadence and flow of all of nature's complex orders. To ignore the responsibility of humanity to participate fully in the harmony of our non-human surroundings leads inevitably to the distortion of our natural impulses by the imposition of often ossified conventions and institutions on the intuitive ground of human experience. Impositional conduct leads to coercion which, in the human world as our most immediate example, reduces the creative possibilities of community by denying the full participation of some of its participants.

The *Daodejing* expresses this notion of teasing an integrated and dynamic harmony out of available differences: 'The myriad things shoulder *yin* and embrace *yang* and blend their energies (*qi*) together to constitute a harmony (*he*)' (see Qi). The 'myriad things' denotes the natural world as a complex of unique and particular thing-events (*de*). This complex of things lacks the suggestion of a single-ordered unity or coherence carried by terms such as 'cosmos' or 'universe', where the inventory of things is thought to be organized according to unifying natural laws and causal relationships. To the extent that the myriad things achieve 'order', it is constituted by the sum of a contingent set of aesthetic harmonies construed from the perspective of each of the participants as they dispose and express themselves in the world.

The intelligible patterns created by the many different things which collaborate to constitute the world are all pathways or *dao*s which can, in varying degrees, be traced out to map one's own place and its context, and in this mapping, to find coherence and meaning. *Dao* is always reflexive in that there is no final distinction between an independent source of order and that which it orders. One's world and its order are constituted by the collaboration of oneself with a myriad other self-causing and mutually shaping particulars. *Dao* is, at any given time and place, both *what* the world is and *how* it is, as entertained from that perspective. For this reason, from a human point of view, explanation does not lie in the discovery of some antecedent agency or the isolation and disclosure of relevant causes. Rather, any particular event or phenomenon can be understood by mapping out the conditions which collaborate to sponsor it. Once broadly understood, these same conditions can be manipulated therapeutically to anticipate the next moment, and to prescribe for it.

The *Daodejing* is primarily a political treatise. It is by bringing this anarchic and ecological sensibility to the operations of human governance that government in its relationship to community can become *wuwei*, free of any coercive activity and free to orchestrate the full talents of its constituent population (see Daodejing).

The *Zhuangzi*, focusing more on personal than political realization, is without question one of the richest and most celebrated pieces of philosophical literature in the Chinese tradition. While expressing an unrelenting scepticism about those evidential claims for certainty and objectivity which empower competing philosophical voices, this text uses a collage of anecdotes, parables, provocative images and other such rhetorical tropes and strategies to defend creativity as a fundamental value. It is through an appreciation of the role of creativity in the world that one comes to an understanding of an inclusively 'proper' order, where the 'anarchic' harmony, eschewing as it does any sense of determinative principle or *archē*, is 'made mine' by the fullest expression of all participating elements. The functional value of such an anarchic understanding of order in the world lies in overcoming any fear of personal injury or death by recognizing and relying upon the discernible regularity and continuity in the world, and by respecting the contribution that transformative change makes to the quality of life (see Zhuangzi).

Daoism, like Confucianism, becomes porous and eclectic as it enters the Han dynasty, and serves as a freewheeling counterpoint to the Confucian state ideology throughout the two millennia of Imperial China. In the

Huainanzi, an early Han compendium of knowledge representative of this syncretic turn, Daoism serves as a primary ore, being alloyed with the concerns and perspectives of competing schools to produce a more malleable and practical amalgam. The coherence of the *Huainanzi*, however, is one true to the spirit of Lao-Zhuang in that conflicting and divided opinions are happily juxtaposed as necessary for providing the fullest summary of China's cultural achievements. It is richness and intensity rather than some rationalized order that is the signature of Huainanzi's version of syncretic Daoism (see Huainanzi).

There were also 'Huang-Lao' Daoists in late Qin and Eastern Han dynasties who combined the seemingly incompatible bureaucratic and technocratic designs associated with the Yellow Emperor (Huangdi) with Daoistic sensibilities as a strategy for participating effectively in the political order. These thinkers coupled the institutional structures and institutions of centralized government with Daoist notions of sagely government to constitute a kind of instrumental Daoism. Recent archaeological discoveries are providing an increasing amount of evidence from which to bring this movement into clearer focus.

The qualitative, aesthetic concerns which pervade Daoism continued to have an important influence throughout the evolution of Chinese culture, most notably in the productive and literary arts: painting, calligraphy, poetry, ceramics and so on. The vocabulary of Daoism was also instrumental in transforming imported Mahāyāna Buddhism from an exotic religion into a source of spiritual growth with largely indigenous aspirations (see Buddhist philosophy, Chinese).

6 The 'Hundred Schools'

Granted the disposition on the part of the Chinese to promote a harmonious narrative of China's cultural development, a closer look at the actual events yields a slightly greater sense of conflict. In the approximately one hundred years intervening between the death of Confucius in 481 BC and the birth of his most influential disciple, Mencius, in *circa* 380 BC, a complex variety of philosophical schools developed. In the Daoist work the *Zhuangzi*, this growth in diversity is referred to as the period of the 'Hundred Schools'. Far from seeing in this a healthy pluralism of opinion, the *Mengzi* (3B/9), as a representative of the Confucian tradition, described this phenomenon in the most negative of terms: 'Sage-kings do not arise, the various nobles do as they please, scholars without position speak freely on any number of topics, and the words of Yang Zhu and Mo Di fill the empire.'

The period of the conflicting schools began when Mo Di, the founder of Mohism, called Confucian ideas into question. Mohist thinking, generally characterized as a kind of utilitarianism, constituted a significant challenge to the ritually grounded traditionalism of Confucius (see Mohist philosophy; Mozi; Logic in China). Legalism, associated with Shang Yang (d. 338 BC) and Han Feizi, differed from both Confucianism and Mohism by beginning its social thinking not with the people but with positive laws and sanctions presumed to be external devices necessary to bring order to the turmoil of its day. With the Legalists came at least the adumbration of a theory of rational political order (see Legalist philosophy, Chinese; Law and ritual in Chinese philosophy). During the succeeding centuries leading up to the founding of the Han dynasty, a plethora of alternative schools emerged and court-sponsored academies were established in different parts of the empire reminiscent of the great academies of classical Greece. The most famous representative of these academies during the fourth and early third centuries was Jixia at the Qi capital of Linzi, attracting over time a range of such notables as Zhuangzi, Song Xing, Shen Dao, Mencius, Gaozi, Xunzi and Zou Yan.

In the beginning, the competing schools engaged primarily in debates over doctrine, although they were also quite ready to offer commentary on the ever changing political situation. Conservative Confucians who sought the meaning of life by appeal to family and social obligations were opposed by those Daoists who sought to attune the human world to the regular rhythms and patterns of nature. There were fierce debates among the Confucians, Daoists, Mohists, Sophists and Yangists (and many others) concerning the goodness or evil or neutrality of human nature. In due course, as the contest became increasingly complex, the debates took a procedural and logical turn. Thinkers such as Zhuangzi, Sophists such as Hui Shi and Gongsun Longzi and the later Mohist logicians began to argue about the meaning of argument itself, and to worry over standards of evidence. Mohism and the School of Names developed a complex and technical vocabulary for disputation, and puzzled over the linguistic paradoxes which advertise the limits of language (see Logic in China). Thus, as was the case in the history of early Greek thinking, a second-order rationalism developed in China primarily as a means of adjudicating doctrinal conflict.

One of the most puzzling questions in Chinese intellectual history is why the rational and proto-scientific activities illustrated by the disputations of the late fourth and third centuries BC had virtually vanished by the early years of the Han dynasty. Perhaps the most plausible explanation for this phenomenon is one best couched in terms of the sociology of knowledge. As noted above in our discussions on the transmission of knowledge in China, the civilizing process in China was not one of urbanization as in the West, where the very word 'civilization' means 'citification'. Politics was explicitly concerned with the *polis*, and thus had to confront the complex patterns of trade and commerce associated with interactions among diverse languages and ethnicities, along with the growth of a plurality of institutions - banks, universities, trading companies - each with its own ideological axe to grind. Politics then became the art of compromise applied essentially to pluralistic urban centers.

In China, civilization was effected by the contextualizing power of a common written language. Thus, the same logical tools which could effectively serve the cause of adjudicating disputes in a pluralistic urban society such as developed in Europe would be unacceptably disruptive in a society whose stability was dependent upon the existence of communal affect associated with a common literature which both shaped and was shaped by a common language. The art historian George Rowley once remarked that for the Chinese, truth is not truth unless it is subtle. This is but to say that the bare bones logical propositions, bereft of subjective forms of feeling, cannot be 'true'. The affective and connotative features of language must always be present if language is to serve its function of harmonizing and stabilizing social interactions.

Herein lies a productive illustration of the greatest of contrasts between the politics of China and of the West. For the Western cultures, which are characterized by a pluralism of beliefs and practices, logical tools are essential as instruments of conflict resolution. For the Chinese culture, characterized by a far greater homogeneity of language and ethnicity, those same tools threaten the community of affect which guarantees social harmony. Thus in China, the seeds of 'rationalism' and the contentiousness it entails could be expected to have fallen upon rocky soil, and indeed they did.

7 Xunxi and rationalized Confucianism

A decisive figure in both the emergence and ultimate decline of rationalism in early China was the Confucian philosopher Xunzi. Xunzi is often touted as the most 'rationalistic' of the classical Confucians, and in one sense he deserves to be so described. However, because his so-called rationalism is grounded in history and culture without appeal to metaphysical, grammatical or sociological determinants, he is rationalistic in a rather Pickwickian sense. Xunzi's project does not illustrate a development of the incipient rationalism of the later Mohists and the School of Names. His is a distinctly historicist programme - and a historicist rationalism is, strictly speaking, oxymoronic. Thus in an analysis of argumentation in Xunzi, one must distinguish Xunzi's concrete, historical rationality from notions of abstract and impersonal reason familiar in classical Western metaphysical thinking.

There is, in spite of many similarities, also a fundamental difference between Xunzi's rationality and the peculiar kind of reasoning which grounds the later Mohist *Canons*. The rationality shared by Xunzi and the later Mohists is based upon their nominalist stances. However, the almost total congruence between the later Mohists and Xunzi on the nature of language and logic, which among other things allows Xunzi to utilize most of the technical vocabulary of the Mohist disputers, must be qualified by the fact that Mohist nominalism shares with physicalist nominalism of the Western tradition the importance of logical (and causal) necessity (see Nominalism). For the Mohists, necessity (*bi*) is what is 'unending' (*bu yi ye*), a condition of logical and scientific disputation which is invulnerable to time. Thus, even though the later Mohist has no explicit metaphysical theory, there is a belief that the world is comprised by concrete particulars with necessary logical relations one to the other (see Mohist philosophy).

By contrast, there are no metaphysical, linguistic or behaviourial determinants to be found in Xunzi. Rationality for Xunzi is formed dialectically amid cultural, social and natural forces, both shaping and being shaped by them. Valid reasoning is the discovery and articulation of appropriate and efficacious historical instances of reasonableness. 'Reasoning' (*li*) and historical analogy are inseparable (see Li). On the one hand, *li* - the mapping out of patterns - can only operate on the basis of assumed classifications (*lei*); at the same time, it is the 'mapping' operation of *li*, including and excluding on the basis of perceived similarities and differences, that establishes classifications (*lei*) in the first place.

Xunzi's nominalism must be understood, much as that of the Greek Sophists, as a tropic rather than a metaphysical device. Nominalism of the 'rhetorical' as opposed to the 'logical' or 'atomistic' variety does not arise from a conviction that universals do not exist, or that there are no abstract entities, or that there are no such things as non-individuals. Xunzi's rhetorical or linguistic nominalism is essentially an anti-metaphysical and an anti-logical methodology that is quite similar to the sophistic nominalisms of many of the early Greek rhetoricians.

Though Mencius was later to emerge in the Chinese tradition as the most prominent of Confucius's interpreters, it was Xunzi's ritual-centred Confucianism that held sway in the formative years of Confucianism as a state ideology. This was emphatically the case during the first century of the Han dynasty and the founding of empire. The institutionalization of academic positions which the Han dynasty inherited from the Qin helped to perpetuate and galvanize the influence of Xunzi. Several of Xunzi's immediate students were responsible for the transmission of specific classics which comprised the Confucian curriculum, including the *Guliang Commentary* to the *Spring and Autumn Annals* and the *Zuozhuan* historical narrative. Even the 'Mao' orthodox version of the *Book of Songs* was named for a lineage of Xunzi disciples. The preface to Xunzi's collected works written two hundred years after his death by the court bibliographer Liu Xiang (79-8 BC), reported that Lanling, still under the influence of Xunzi, continued to produce fine scholars.

Perhaps Xunzi's greatest and most enduring influence came from the extent to which he continued Confucius' emphasis on ritual practice as an instrument for socializing, enculturating and humanizing the Chinese world. His description of the function of ritual in society is cited extensively in the histories and the many canons of ritual that were compiled during this period, and looms large in the syncretic philosophical literature that was to become the signature of the Han. The *Huainanzi*, for example, is by and large a text representing a variety of often conflicting philosophical positions, but the crown of this Han dynasty work is its final chapter, the 'Greatest Clan' (*taizu*), which develops a philosophical position in many aspects reminiscent of Xunzi, especially with respect to the importance of ritual and learning (see Huainanzi).

It can be argued that Xunzi, by co-opting the philosophic concerns of the early rationalists for the emerging Confucian program, made the formal continuation of these competing schools redundant. Put another way, the nascent rationalism which was emerging in those thinkers interested in argument for its own sake was overwhelmed by the conventionalist rationalism of Xunzi and the large following he attracted in the early years of the Han dynasty. The signature of the Confucian sense of order that persists far beyond the temporal borders of the Han dynasty is typified by Xunzi's ritually constituted community, a movement from contesting diversity to an absorbent and inclusive harmony. The emergence of what we might call 'Han' thinking in this period had a determinative effect on the style which Chinese philosophy was to assume throughout its long history.

8 First millennium syncretism

With the emergence of a Confucian orthodoxy in the Han dynasty based on the Xunzi branch of Confucianism, scholarly dispute was tempered by a fundamental commitment to mutual accommodation. In the exercise of criticism the ritual basis of order comes into play, since rituals serve as patterns of deference which accommodate and harmonize differences in desires, attitudes and actions. Ideally, 'dispute' is a cooperative exercise among responsible participants that proceeds beyond obstinacy to search for alternatives on which all can agree. There is a fundamental dis-esteem for coercion of any kind, because aggressiveness or violence threatens to disrupt rather than reinforce or improve upon the existing social order. The goal of protest is not victory in contest, which is necessarily divisive, but the strengthening of communal harmony.

The interest in logic and rationality as tools of disputation which had emerged briefly in the pre-Qin days of the Hundred Schools, along with the analytical and dialectical modes of discourse attendant upon these methodologies, soon faded into the counter-current of Chinese intellectual culture. With the ascent of a Confucian ideology, China emerged as a culture grounded in the immanent aesthetic order of a ritually-constituted society, in large measure precluding the kind of rational conflict familiar in the Western tradition. With the Han thinkers came the emergence of a fortified Confucian orthodoxy, complete with canon and commentary.

By the beginning of the first century BC, Confucianism had become the clear and enduring victor over all contending voices, a success due in important degree to its ability to accommodate within a ritually-grounded intellectual society the most profound elements of Daoism, Legalism and Mohism, a pattern that would be repeated

in Confucianism's gradual appropriation of Buddhist elements by its medieval adherents. It is generally recognized that the intellectual orthodoxy which came to dominance in the Han dynasty was a river fed by three powerful streams: Confucianism, Daoism and Buddhism. The stress here is upon the harmony of the three traditions. Though Confucianism remains dominant, the three sensibilities provide distinct foci in accordance with which one can construe one's life.

Confucians are often distinguished from Daoists by the observation that though both seek aesthetic harmony, the Daoists seek harmony with nature while Confucians are concerned with harmony in the social sphere. 'Nature' ('the ten thousand things') and 'society', as contexts, are both aesthetic products whose order is a creation of the elements of the contexts (see Aesthetics, Chinese). In the Daoist text the *Zhuangzi*, there is the statement that each of the ten thousand things comes into being out of its own inner reflection and yet none can tell how it comes to be so. The Confucian version of this claim is that 'it is the person who extends order in the world (*dao*), not order that extends the person' (*Analects* 15/29).

In China, the phrase 'the continuity between man and Heaven' (*tianren heyi*) has been construed to mean that personal, societal, political and cosmic order are coterminous and mutually entailing, and that from the human perspective, this order is emergent in the process of one's own self-cultivation and articulation. If we think of the various contexts which are to be harmonized as concentric circles, we can see that there is an interdependence between one's self-realization at the center and world order at the outer extreme. Classically, this is expressed through the notion of the Sage as exemplar both of tradition in its broadest sense and 'the will of Heaven' - that is, the specific environing conditions that set up the viable possibilities in a particular social situation or historical epoch. This is the sense of Mencius' assertions that 'all of the myriad things are complete here in me' and 'one who applies exhaustively his heart-and-mind realizes his character, and in thus realizing his character, realizes Heaven (*tian*)' (*Mengzi* 7A/4 and 7A/1).

There are continuous and dynamic patterns discernable in the developmental flow of Chinese philosophy that provide us with a way of organizing the tradition beyond its consolidation in the Han dynasty. First, the membrane that divides Chinese intellectual culture into 'schools' and 'traditions' is highly porous. Intellectual diversity, like political diversity, follows a pattern of being absorbed and assimilated into a harmony dominated by some central doctrine, and then precipitating out of this same harmony in periods of disunity. When the centre is strong, the dominant school draws into itself and co-opts competing elements, thereby fortifying itself against opposition. When, over time, the centre weakens, disintegration sets in and those intellectual resources that have been marginalized by the dominant centre move in to reshape the core.

As noted above, Confucianism, fortified by precepts and concerns drawn from the 'Hundred Schools' of the pre-Qin period, emerges as the prevailing Han ideology. As the Han dynasty declines in the second century AD, religious Daoism and Buddhism move in from the periphery to recolour the intellectual ideology, transforming state-centred Confucianism into the more esoteric and reclusive 'pure conversation' (*qingtan*) and neo-Daoist (*xuanxue*) movements of the Northern and Southern dynasties. However, the influence is mutual. As Buddhism takes root and flourishes in Chinese soil, it is interpreted through indigenous categories which overwhelm many of its original concerns, gradually translating it into a religion and philosophy consonant with the assumptions of the Sinitic world view. By the time the Huayan and Chan sects of Buddhism appear in the Tang dynasty (see Buddhist philosophy, Chinese), doctrinal affinities with Daoism and Confucianism have transformed an erstwhile foreign doctrine into a Chinese institution.

9 Neo-Confucianism: Zhu Xi and Wang Yangming

As the Empire regrouped and neo-Confucianism (see Neo-Confucian philosophy), referred to as '*dao* learning' in Chinese, began to take shape in the medieval period, the speculative and the practical extremes of Buddhism came to exert an influence on the revival of Confucianism and to set the agenda for rival claims to orthodoxy within the neo-Confucian ranks. On one extreme, the Cheng-Zhu school favoured broad text-based learning and 'the investigation of things' (*gewu*) while the competing Lu-Wang school rejected canonical studies for a more subjective, meditative approach to personal realization. What these contesting traditions shared in common was philosophical ambition encouraged by the presumption that there is a direct line between personal cultivation and an understanding of natural and moral order. This led to extended reflections on the nature and order of all things, and heated discussions about the relationships that obtained among the most abstract distinctions which could be

marshalled in explanation of cosmic regularities. Where these two extremes of neo-Confucianism disagreed most fundamentally was on the most effective method of self-cultivation (see Self-cultivation in Chinese philosophy).

Zhu Xi is representative of the systematic and theoretical wing of neo-Confucianism. His extensive commentaries established the Four Books (the *Analects*, the *Mengzi*, the *Daxue (Great Learning)* and the *Zhongyong (Doctrine of the Mean)*) as the core curriculum for official examinations, an orthodoxy that persisted into the twentieth century. The authority of Zhu Xi's project lay in his claim to retrieve and revive the import of classical Confucianism. In so doing, he made use of the traditional philosophical vocabulary, but augmented it with complex theoretical discussions of *li* (the patterned regularity of existence) and *qi* (the psychosomatic stuff of existence), giving precedence to the former as identical with the grounding principle of Zhou Dunyi's cosmology, the Great Ultimate (*taiji*) (see Li; Qi; Zhou Dunyi). Hence, Zhu Xi's wing of neo-Confucianism is often referred to as '*li* learning'. These abstract distinctions had moral significance. They could be appealed to qualitatively in explanation of both the goodness of humanity and how to realize it. Although Zhu Xi did not rule out introspection as a means to illumination, the emphasis of his programme was clearly on scholarly learning.

The most eminent of the thinkers representing an emphasis upon internal cultivation was Wang Yangming. He rejected the intellectualization of personal realization by identifying the heart-and-mind (see Xin (heart-and-mind)) with *li*, or pattern. For Wang, the human mind is both the locus and the standard of sagehood. Perhaps the most celebrated theme in Wang is his belief in the continuity and inseparability of knowledge and practice.

There is some question among contemporary scholars as to whether neo-Confucian philosophy, launched with the eleventh century cosmological speculations of Shao Yong and Zhou Dunyi, abandoned the tradition of *ars contextualis* and became metaphysical in a more recognizably Western sense. This involves the question as to whether notions of 'transcendence', 'objective essences' and 'natural kinds' were at least tacitly introduced. At the very least, one can say that Western philosophers would find the language of neo-Confucian philosophers more familiar than that of most other Chinese thinkers. Having allowed this surface familiarity, one must consider that the philosophical substance of tradition weighs heavily against any assumption that neo-Confucianism was dualistic, and was thus disposed to move in an essentialist direction.

Historically, the speculative, cosmological turn in Chinese philosophy came under formidable attack with the founding of the Qing dynasty in the seventeenth century. Evidential research (*kaozhengxue*) brought with it an attempt to get behind the 'empty' commentaries of neo-Confucianism and a return to the philologically-centered historical scholarship of 'Han learning' (*Hanxue*). On the premise that new problems require new solutions, the abstract theorizing and universalistic tendencies of Song-Ming '*dao* learning' gave way to the analysis of particular historical events and cultural artifacts as a resource for finding answers to the specific issues of the day. Thinkers such as Wang Fuzhi and Dai Zhen recovered and reaffirmed the correlative and interdependent relationship between historical event and the principles of order. Once again, it can be seen how the pragmatic concerns of most Chinese intellectuals militate against the exercise of philosophical speculations that move too far afield from the concrete problems of human beings, or which could conceivably serve to introduce contentiousness among intellectuals.

10 The modern period

Beginning with the Buddhist incursion into China at the beginning of the Common Era, the question of the degree to which 'Western' influences have effected significant changes in China's cultural life has been vigorously debated. Han Yu, like Mencius a thousand years earlier, railed against the pernicious influence upon the Confucian community of competing ideologies. Yet, as we have argued above, the Chinese genius for realizing social stability through the harmonious integration of novel influences led to the effective transformation of Buddhist ideology and practice into a distinctly Chinese institution.

In modern times, events surrounding the May Fourth movement in 1919 seemed likely to lead to the Chinese acceptance of Western ideological influences. Hu Shi, a student of John Dewey and later a distinguished Chinese philosopher, helped arrange Dewey's twenty-six month lecture tour of China in 1919-21), and later sought to disseminate his mentor's ideas. Bertrand Russell, whose lectures in China overlapped Dewey's, was equally well-received by the intelligentsia. However, the final result of the May Fourth uprising, which occurred just three days after Dewey's arrival in China, was not the realization of democratic reform but the founding of the Chinese

Communist Party in 1921. The widespread assumption that Marxism has effectively westernized Chinese philosophical culture is seriously challenged by Mao's own claim, upon adopting Marxist ideology, that he was using a Western heresy to confront the West (see Marxism, Chinese).

The Chinese transformation of Marxist into Maoist thinking in contemporary China reveals the inertia of Chinese tradition. The single most distinctive change that Mao made to Marxism was a commitment to particularity and site-specificity. Dialectical materialism is revised to reflect a *yin-yang ad hoc* relationship between economic principle and social superstructure. Although hierarchical, these forces are seen as interdependent and mutually determining. Human malleability and the fluidity of social nature goes far beyond the standard Marxian line. Where Marx places emphasis on the uniformity of class-originated nature, Mao emphasizes the importance of those differences derived from ways of living and thinking that must be factored into the evaluation of any particular 'concrete' personality. There is in Mao a basic suspicion of abstract, general claims, and a recurrent return to specific cases and historical examples. The contemporary Chinese view so historicizes the Marxist sensibility as to make room for an almost unlimited flexibility with regard to the shaping of individual personalities and the development of individual skills (see Marxism, Chinese).

There is little evidence to suggest that contemporary China has abandoned any significant elements of its syncretic Confucian orthodoxy. The dynastic leadership of contemporary China maintains many of the same characteristics that have dominated since the Han dynasty: a governing state ideology that gives all people their respective place in their community, an understanding of the nation as a 'family', a programmatic constitution which functions more like a Bill of 'Rites' than a Bill of Rights, a filial respect for the ruler as 'father and mother' of the people, and the consequent sense of rule as a personal exercise. With respect to the personal character of rule, it continues to be the case in China that to object to the policies that articulate the existing order is in fact to condemn the ruler's person.

As a ritually-constituted society, without grounding in the objective principles associated with reason or natural law, contemporary China is defined by the exemplars of its tradition. The members of the society are themselves possessed of their 'humanity' not as a gift from God or a common genetic inheritance, but as created by ritual enactment. The Chinese have no inalienable rights. Citizens have been deemed to possess only those rights granted by China's various constitutions. The Chinese would see the Enlightenment insistence upon the universality of certain values and principles as an instance of ethnocentric dogmatism. Chinese ethnocentrism is, perhaps, more consistent than its Western counterpart since it is grounded in the self-conscious insistence upon the centrality of its peculiar ethos, defined by racial and linguistic identity.

China remains a culture grounded in the model of the family which cultivates filial dependency. Thus, the Chinese have no means of cultivating that 'healthy suspicion' of governmental power which we take for granted without undermining the community of affect that binds ruler and people. As a rational means of organizing social and economic interactions, the technology so prized in the West cannot but erode the ritual grounding of interpersonal relationships. One of the catchwords of the Tiananmen protests in 1989 was 'democracy'. But, in a society where individualism remains a symptom of selfishness and license, and freedom of speech must be qualified by the Confucian understanding that not only saying but *thinking* involves a disposition to act, Chinese democracy must certainly take on an unfamiliar form. Indeed, the inhibition of individualism and freedom of speech is not a modern invention of Chinese communists but a persistent feature of a Confucian society in which ideas are always dispositions to act.

One can hardly look closely at the intellectual culture of contemporary China without coming to respect the power of China's traditions. The intransigent sense of 'Chineseness' which coalesced in the Han dynasty continues to determine the shape of Chinese intellectual culture. For good or ill, the Chinese remain the people of the Han.

See also: Aesthetics, Chinese; Buddhist philosophy, Chinese; Chinese Classics; Confucian philosophy, Chinese; Confucius; Daodejing; Daoist philosophy; Guanzi; History, Chinese theories of; Huainanzi; Law and ritual in Chinese philosophy; Legalist philosophy, Chinese; Logic in China; Lushi chunqiu; Marxism, Chinese; Mohist philosophy; Neo-Confucian philosophy; Self-cultivation in Chinese philosophy; Wang Chong; Yijing

DAVID L. HALL
ROGER T. AMES

References and further reading

Bodde, D. (1991) *Chinese Thought, Society, and Science*, Honolulu, HI: University of Hawaii Press.(Examines the social and intellectual forces that influenced the development of science and technology.)

Chan Wing-tsit (1963) *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press.(A basic resource of original sources in translation.)

de Bary, W.T. *et al.* (1960) *Sources of Chinese Tradition*, New York: Columbia University Press.(A representative selection of philosophical literature translated from original sources.)

Gernet, J. (1985) *China and the Christian Impact: A Conflict of Cultures*, trans. J. Lloyd, Cambridge: Cambridge University Press.(Uses journals, correspondence and other sources to place the Jesuits and Chinese intelligentsia in conversation and to reveal their different presuppositions.)

Graham, A.C. (1978) *Later Mohist Logic, Ethics and Science*, Hong Kong: The Chinese University Press.(A detailed look at the analytic side of classical Chinese philosophy.)

Graham, A.C. (1989) *Disputers of the Tao*, La Salle, IL: Open Court.(The most sophisticated overview of major philosophical developments in China's formative period.)

Granet, M. (1950) *La pensée chinoise (Chinese Thought)*, Paris: Éditions Albin Michel.(A pioneering attempt to characterize Chinese philosophical thinking.)

Hall, D.L., and Ames, R.T. (1987) *Thinking Through Confucius*, Albany, NY: State University of New York Press.(A comparative study which attempts to identify underlying assumptions in Confucian thinking.)

Hall, D.L., and Ames, R.T. (1995) *Anticipating China: Thinking Through the Narratives of Chinese and Western Culture*, Albany, NY: State University of New York Press.(A sequel to *Thinking Through Confucius* which further elaborates differences in thinking, arguing that these differences must be factored into discussions of contemporary issues such as gender, rights and so on.)

Hall, D.L., and Ames, R.T. (1997) *Thinking from the Han: Self, Truth and Transcendence in Chinese and Western Culture*, Albany, NY: State University of New York Press.(A discussion of cultural differences that emerge in the contrasting narratives of the Chinese and Western traditions.)

Henderson, J.B. (1984) *The Development and Decline of Chinese Cosmology*, New York: Columbia University Press.(Traces correlative cosmology in China from its origins to late imperial times.)

Maspero, H. (1981) *Taoism and Chinese Religion*, trans. F. Kierman, Jr, Amherst, MA: University of Massachusetts Press.(Argues for the influence of Daoism, rejecting the assumption that Confucianism was the dominant philosophical orientation in Chinese history.)

Schwartz, B.I. (1985) *The World of Thought in Ancient China*, Cambridge, MA: Harvard University Press.(An accessible history of classical Chinese philosophy emphasizing continuities with Western culture.)

Chinese room argument

John Searle's 'Chinese room' argument aims to refute 'strong AI' (artificial intelligence), the view that instantiating a computer program is sufficient for having contentful mental states. Imagine a program that produces conversationally appropriate Chinese responses to Chinese utterances. Suppose Searle, who understands no Chinese, sits in a room and is passed slips of paper bearing strings of shapes which, unbeknown to him, are Chinese sentences. Searle performs the formal manipulations of the program and passes back slips bearing conversationally appropriate Chinese responses. Searle seems to instantiate the program, but understands no Chinese. So, Searle concludes, strong AI is false.

Searle (1980) argues that, since in the imaginary case he does everything the computer would do and he still does not understand a word of Chinese, it follows that a computer successfully programmed to pass a Chinese Turing test (see Turing, A.M. §3) would not understand Chinese either. The problem, according to Searle, is that computer programs, whether executed by electronic devices or by Searle inside the room, concern only *syntax*, or strings of symbols characterized only by *spelling*, not *meaning*, while thought and understanding require meaning and semantics. And he claims you cannot get semantics from mere syntax, no matter how subtle or complicated it may be.

An underlying assumption of Searle's interpretation is that real mentality and genuine intentional content (what he calls 'intrinsic intentionality') require consciousness. In later publications Searle (1990a) has made this explicit in terms of his 'connection principle', which states that the notion of an unconscious or non-conscious thought can be coherently understood only as a disposition to have a corresponding conscious thought; an unconscious belief that *p* or desire for *x* is nothing more than the disposition to have a conscious belief that *p* or desire for *x* under apt circumstances. He argues that the notion of a thought that is in principle inaccessible to consciousness is incoherent (see Unconscious mental states §5).

Many objections have been raised against the Chinese room. Searle pre-emptively discussed and tried to refute several of them at the time of his original publication. Two have been especially important: those which Searle labels the 'systems reply' and the 'robot reply'.

The systems reply concedes that in the imagined case Searle does not understand Chinese, but insists that the whole organized system consisting of Searle, the rule books, paper, pencil and room *does*. According to the systems reply the example errs in comparing Searle with the whole computer when in fact his role is analogous only to that of the central processing unit (CPU) of the computer, a small part of it, to which there is no need to attribute the understanding that is displayed by the entire system (see Semantics, conceptual role).

Searle offers a two-part response. First, he labels absurd the very suggestion that a system consisting of a person, some books, paper and a pencil could literally understand things or have a mind over and above that of the person alone. Such an ensemble he asserts is just the wrong sort of thing literally to understand or to have a mind. Second, he tries to avoid the objection by altering his example and removing himself from the room to a wide open field and committing all the rule books to memory. In such a case he would perform all the purely formal operations in his head, re-establishing his analogy with the whole computer. He asserts he would still not understand a word of Chinese (though some of his critics claim otherwise).

The robot reply concedes that the original Chinese room might not suffice for understanding Chinese. Since all its behaviour is strictly verbal, it is able to make connections between words and words, but not between words and the world: it might give coherent answers to questions about boiled eggs, but not have any ability to recognize an egg. The robot reply thus imagines a program installed within a robot that allows it to simulate the full range of human behaviour, verbal and non-verbal; and *that*, it is claimed, would count as understanding Chinese and having a mind.

Searle, however, disagrees. He notes that the robot's sensors would supply its internal processors only with more formal symbolic inputs. Were he and his Chinese room installed within the robot with additional rule books specifying how to respond to those additional symbols, and were these responses to drive the robot's apt behaviour, he, Searle, would still not understand a single word of Chinese. Nor, he argues, would the robot. Again,

the assumption that mentality requires consciousness seems in part to drive Searle's intuition. Critics less sympathetic to that assumption and more inclined to analyse intentionality in terms of causal interactive relations have been predictably more persuaded by the robot reply, especially if it is combined with the systems reply (see Semantics, informational).

Other objections to the Chinese room have sought to discredit it as a source of unreliable and misleading intuitions. Some have focused on the enormous disparity between the complexity and speed that would be required in a computer able to pass a Turing test, which the man in the room could not match. Others have stressed that strong AI need not be committed to the Turing test, and that the specific program by which behaviour is produced is crucial to whether something has a mind. In reply, Searle denies that speed or style of program is of any importance; what matters, he argues, is that the operations he performs are merely formal.

See also: Artificial intelligence; Consciousness; Intentionality

ROBERT VAN GULICK

References and further reading

Dennett, D.C. and Hofstadter, D. (eds) (1981) *The Mind's I*, New York: Basic Books.(A diverse and delightful collection with commentary by the editors including critical discussion of the Chinese room.)

Searle, J. (1980) 'Minds, Brains, and Programs', *Behavioral and Brain Sciences* 3: 417-24.(Original publication of the Chinese room.)

Searle, J. (1990a) 'Consciousness, Explanatory Inversion and Cognitive Science', *Behavioral and Brain Sciences* 13: 585-96.(Discussion of the connection principle.)

Searle, J. (1990b) 'Is the Brain's Mind a Computer Program?', *Scientific American* 262: 26-31.(A clear and simple introduction to the Chinese room debate.)

Van Gulick, R. (1988) 'Consciousness, Intrinsic Intentionality and Self-Understanding Machines', in A. Marcel and E. Bisiach (eds) *Consciousness in Contemporary Science*, Oxford: Oxford University Press.(Places the Chinese room in the context of general issues about mind, machines, meaning and consciousness.)

Chinul (1158-1210)

Chinul was the founder of the Korean Chogye school of Buddhism. He sought to reconcile the bifurcation between Kyo (doctrinal) thought and Sôn (Zen) practice that rent the Korean Buddhist tradition of his time, by showing the symbiotic connection between Buddhist philosophy and meditation. He also advocated a distinctive program of soteriology that became emblematic of Korean Buddhism from that time forward: an initial sudden awakening to the nature of the mind followed by gradual cultivation of that awakening until full enlightenment was achieved.

1 Life

Chinul, the National Master Puril Pojo, ordained as a monk when he was seven years old and passed the ecclesiastical examinations in the Sôn (Zen) branch of the Korean Buddhist tradition in 1182. According to his biography, however, Chinul became disillusioned with his colleagues' pursuit of worldly fame and profit, and soon afterwards decided to abandon ecclesiastical advancement for a contemplative life in the south of the Korean peninsula. Unlike many of his peers, Chinul did not make the incumbent pilgrimage to the mecca of the Chinese mainland and, despite his Sôn ecclesiastical affiliation, seems also not to have had close personal contacts with any teacher of meditation in Korea. A virtual autodidact, Chinul made up for this lack of individual instruction by drawing on sources that were readily available to him: the scriptures and scholastic commentaries of Kyo and the discourse records of previous masters of Sôn. His reading of scriptures and of Sôn texts catalyzed three separate awakenings, which led to his realization of the fundamental identity between Kyo and Sôn.

Chinul eventually established a number of monasteries and hermitages in the south of the Korean peninsula, including what became one of the three most important monasteries in the Korean Buddhist tradition, the Songgwang-sa (Piny Expanse Monastery). Chinul's works all date from this period of contemplation, starting with his earliest work, *Kwôn su Chônghye kyôlsa mun (Encouragement to Practice: The Compact of the Samādhi and Prajñā Community)*, written in 1190 upon the establishment of his first retreat society. *Susim kyôl (Secrets on Cultivating the Mind)*, arguably his most popular treatise, was composed between 1203 and 1205, and his *magnum opus*, *Pôpchip pyôrhaengnok chôryo pyôngip sagi (Excerpts from the Dharma Collection and Special Practice Record with Personal Notes)*, was completed in 1209, one year before his death. His remaining works were edited and published posthumously by his successor Chin'gak Hyesim (1178-1234).

The approach to Buddhism that Chinul outlined in his writings was marked by a strong reliance on correct doctrinal understanding at the inception of practice. Only after correct understanding had been forged regarding the justification and rationale of Buddhist practice would the student then go on to formal meditation training. That training could run the gamut from techniques heavily beholden to Kyo contemplative exercises, to the joint cultivation of concentration and wisdom (an approach deriving from one of the earliest stratum in Chinese Chan literature), to a new style of meditation just then developing on the Chinese mainland, the investigation of the 'critical phrase' (*kongan*; in Japanese, *kōan*). By synthesizing several antecedent strands of Buddhism, Chinul produced a uniquely Korean Buddhist school called Chogye (after the Korean pronunciation of Caogi Mountain, the monastic home of the sixth patriarch of the Chinese Chan school, Huineng (638-713)). Chinul's Chogye school became the dominant school of Korean Buddhism from the thirteenth century onward and Chinul's distinctive approach to Buddhist thought and practice became the model followed by most subsequent teachers within the Korean tradition (see Buddhist philosophy, Korean).

2 Buddhist ecumenicalism

Although Chinul was principally a Sôn meditator, he nevertheless retained a strong personal interest in the scriptural teachings of Kyo. Chinul's attempted reconciliation of the two branches of Buddhism sought to authenticate the Sôn approach in the scriptures of Kyo, thereby demonstrating the veracity of both. Chinul reviewed the entire Buddhist canon, seeking scriptural substantiation for the Sôn claim that enlightenment could be achieved solely through meditative introspection. While reading a text from the Flower Garland school, he finally found such substantiation and declared, 'What the Buddha said with his lips is Kyo; what the patriarchs of Sôn transmitted with their minds is Sôn. The mind and the words of the Buddha and patriarchs can certainly not be contradictory'. From that point on Chinul was convinced of the affinities between scriptural testimony and meditative experience, and his main ambition was to bring about a rapprochement between the Kyo and Sôn

strands of the Korean Buddhist tradition.

3 Introspection and the mind

For Chinul, the essential unity of Sôn and Kyo would be recognized once Buddhist adepts perceived their original natures: the quality of sentience that was most fundamental to all 'sentient' beings. Such insight was achieved through introspection, looking into the mind itself in order to verify the truth of one's innate buddhahood. It was through introspection - what Chinul termed 'tracing back the light of the mind' (*panjo*) or 'tracing the radiance emanating from the mind back to its source' (*hoegwang panjo*) - that students would overcome all limited conceptions about their true natures and perceive the vast web of interrelationships connecting themselves to all other things in the universe. This state of enlightenment achieved through Sôn practice was, Chinul claimed, what the Flower Garland school of Kyo called the unimpeded interpenetration between all phenomena.

Introspection was a technique that took the usual propulsion of the mind out into the world of the senses and turned it back in upon itself, until the mind's radiant source was discovered. To trace this inherent radiance of the mind back to its source meant to realize instantaneously that oneself was inherently enlightened, that one's mind was congenitally luminous. This brightness of mind was the faculty that illuminated sense objects, allowing them to be cognized. Hence, by tracing the radiance emanating from the mind back to its source, meditators discovered that core of luminosity that was fundamental to all sentient beings (see Illumination; Illuminationist philosophy). This discovery of their true nature constituted the initial understanding of the experience of enlightenment.

4 Soteriology

Chinul derived from this vision of the inherence of enlightenment a soteriological stratagem that came to be the hallmark of Korean Sôn: sudden awakening followed by gradual cultivation. This soteriological programme advocated that the optimal approach to Buddhist practice involved an initial understanding-awakening (*haeo*) to the true nature of the mind, which generated correct comprehension of both the absolute reality of all things (their nature) and their phenomenal appearances (their characteristics). This awakening, which occurred instantaneously, caused students to know that they were in fact innately enlightened and identical in principle with all the buddhas.

However, in Chinul's schema, simply knowing that one was enlightened was not sufficient to be able to act in an enlightened way. It was through a gradual process of subsequent spiritual cultivation that the understanding achieved through the initial awakening permeated all of one's behaviour, leading ultimately to realization-awakening (*chûngo*), or full enlightenment. The habitual patterns of thought and action developed during innumerable previous lifetimes were so deeply engrained that there was little hope of bringing them to an abrupt end; to change the inertial force of these habits required a long process of gradual cultivation. But, because students understood through the awakening that occurred at the very inception of practice that their minds were inherently pure and free of defilements, they could eliminate mental defilements while knowing that there was really nothing that needed to be eliminated. Similarly, students could also develop wholesome qualities of mind while realizing that there was nothing that needed to be developed. Chinul compared this interpretation of cultivation to the gradual calming of waves (mental defilements) after the wind that has whipped them up (the ignorance of one's innate buddhahood) had stopped: one was simply allowing a natural process to complete itself. Hence at the consummation of training one simply 'realized' in all facets of one's life the same truths that one had 'understood' initially. Chinul's soteriological stratagem of sudden awakening followed by gradual cultivation therefore accommodated a developmental component, while remaining strongly beholden to the subitist orientation that dominated much of Zen thought and practice throughout East Asia.

See also: Buddhist philosophy, Korean

ROBERT E. BUSWELL, JR

List of works

Chinul (1190-1209) *Collected works*, in *Han'guk Pulgyo chônsô (Collected Works of Korean Buddhism)*, ed. Han'guk Pulgyo chônsô p'yônch'an wiwônhoe, Seoul: Tongguk University Press, 1980, vol. 4, 698-869; ed. and trans. R. Buswell, *The Korean Approach to Zen: The Collected Works of Chinul*, Honolulu, HI: University of Hawaii Press, 1983.(The standard Korean edition of all of Chinul's works; Buswell's edition is the standard English edition and translation of Chinul's major works. A paperback abridged version of the latter is R.

Buswell, *Tracing Back the Radiance: Chinul's Korean Way of Zen*, Classics in East Asian Buddhism 2, Honolulu, HI: University of Hawaii Press.)

Chinul (1190) *Kwôn su Chônghye kyôlsa mun (Encouragement to Practice: The Compact of the Samādhi and Prajñā Community)*, ed. and trans. R. Buswell, *The Korean Approach to Zen: The Collected Works of Chinul*, Honolulu, HI: University of Hawaii Press, 1983.(Written on the establishment of his first retreat society.)

Chinul (1203-5) *Susim kyôl (Secrets on Cultivating the Mind)*, ed. and trans. R. Buswell, *The Korean Approach to Zen: The Collected Works of Chinul*, Honolulu, HI: University of Hawaii Press, 1983.(Chinul's most popular treatise.)

Chinul (1209) *Pôpchip pyôrhaengnok chôryo pyôngip sagi (Excerpts from the Dharma Collection and Special Practice Record with Personal Notes)*, ed. and trans. R. Buswell, *The Korean Approach to Zen: The Collected Works of Chinul*, Honolulu, HI: University of Hawaii Press, 1983.(Chinul's major work.)

References and further reading

Buswell, R. (1983) *The Korean Approach to Zen: The Collected Works of Chinul*, Honolulu, HI: University of Hawaii Press.(The standard English edition and translation of virtually all of Chinul's works. Includes complete annotation, an extensive introduction to Chinul's life and thought with a survey of the history of Korean Buddhism, and a thorough bibliography to Western and East Asian sources. Abridged in paper back as *Tracing Back the Radiance: Chinul's Korean Way of Zen*, Classics in East Asian Buddhism 2, Honolulu, HI: University of Hawaii Press, 1991.)

Buswell, R. (1986) 'Chinul's Systematization of Chinese Meditative Techniques in Korean Sôn Buddhism', in P.N. Gregory (ed.) *Traditions of Meditation in Chinese Buddhism*, Honolulu, HI: University of Hawaii Press, 199-242.(The ecumenical approach to practice developed by Chinul.)

Buswell, R. (1988) 'Ch'an Hermeneutics: A Korean View', in D.S. Lopez (ed.) *Buddhist Hermeneutics*, Studies in East Asian Buddhism 6, Honolulu, HI: University of Hawaii Press, 231-56.(Describes the hermeneutical stratagems Chinul uses to reconcile statements in Kyo and Sôn materials.)

Buswell, R. (1989) 'Chinul's Ambivalent Critique of Radical Subitism in Korean Sôn Buddhism', *Journal of the International Association of Buddhist Studies* 12 (2): 20-44.(Chinul's treatment of the soteriological stratagem of sudden awakening/sudden cultivation.)

Buswell, R. (1990) 'Chinul's Alternative Vision of Kanhwa Sôn and Its Implications for Sudden Awakening/Sudden Cultivation', *Pojo sasang* 4: 423-63.(Technical discussion of how Chinul seeks to reconcile a radical-subitist form of meditation with his moderate subitism of sudden awakening/gradual cultivation.)

Keel, H. (1984) *Chinul: The Founder of the Korean Sôn Tradition*, Berkeley Buddhist Studies Series vol. 6, Berkeley, CA: Center for South and Southeast Asian Studies.(A valuable expository study of Chinul's contributions to Korean Buddhist philosophy.)

Park Sung-Bae (1983) *Buddhist Faith and Sudden Enlightenment*, Albany, NY: State University of New York Press.(A provocative monograph on Korean Sôn practice, which focuses especially on Chinul and the tension he posits between faith and doubt in Sôn meditation.)

Park Sung-Bae (1994) 'On the Subitist/Gradualist Debate in Korean Buddhism: Sôngch'ôl's Theory of Sudden Enlightenment and Sudden Practice', in D. Suh (ed.) *Korean Studies: New Pacific Currents*, Pacific Association for Korean Studies I, Honolulu, HI: Center for Korean Studies, 163-74.(Analysis of a contemporary Korean argument against Chinul's preferred soteriology.)

Shim Jae-Ryong (1979) 'The Philosophical Foundation of Korean Zen Buddhism: The Integration of Sôn and Kyo by Chinul (1158-1210)', Ph.D. dissertation, University of Hawaii.(Extensive treatment of Chinul's ecumenical philosophy.)

Chisholm, Roderick Milton (1916-)

Chisholm is an important analytic philosopher of the second half of the twentieth century. His work in epistemology, metaphysics and ethics is characterized by scrupulous attention to detail, the use of a few basic, undefined or primitive terms, and extraordinary clarity. One of the first Anglo-American philosophers to make fruitful use of Brentano and Meinong, Chisholm translated many of Brentano's philosophical writings. As one of the great teachers, Chisholm is widely known for the three editions of Theory of Knowledge, a short book and the standard text in US graduate epistemology courses. An ontological Platonist, Chisholm defends human free will and a strict sense of personal identity.

Roderick Milton Chisholm (born in North Attleboro, Massachusetts, USA) has come as close as any philosopher ever does to actually living the good life: marrying happily, teaching graduate philosophy at his Alma Mater and, in later years, living by the ocean. Upon completing an undergraduate philosophy major at Brown University, Chisholm entered the doctoral programme at Harvard, finishing in 1942. The subsequent three years of military experience, administering psychological tests to recruits, had little influence on his philosophy, nor did his brief stint at the Barnes Foundation lecturing on the philosophy of art. Chisholm returned to Brown in 1947 where he taught for the next forty years, heavily influencing Anglo-American epistemology and metaphysics.

Chisholm remained remarkably consistent throughout his career, opposing a variety of fashionable philosophical movements (phenomenalism, linguistic reductions of any sort and extensionalism) and simultaneously developing and elaborating original views on knowledge and justification, intentionality and ontology. In the early years he was most concerned to resist the dominant - and then popular - strains of positivism. For example, against phenomenalists who proposed to translate material-object statements into sensation statements, Chisholm convincingly argued that the requisite boundary or background conditions themselves could never be adequately stated using only sensation statements (see Phenomenalism §2).

It is easy to see in this pattern of reasoning the beginnings of his view, later known as 'the primacy of the intentional'. This doctrine would be elaborated in greater detail as Chisholm debated with Wilfrid Sellars and others who sought to reduce various human referential or intentional capacities to dispositions to engage in certain linguistic behaviour (see Intentionality §§1, 3). The central tenet of the primacy of the intentional is that the ability of humans to use language to refer to things in the world is to be explicated in terms of their ability to think of these same things. Sellars and others held the reverse.

Chisholm was inspired by the work of Franz Brentano to spend years searching for a mark or criterion of the mental. During this period he translated Brentano's *Wahrheit und Evidenz (The True and the Evident)* and many other works by Brentano and Meinong. He lectured frequently in Graz, Austria, winning in 1972 an honorary degree from the University, which was, according to him, the 'high spot' of his philosophical career.

Chisholm published three editions of his *Theory of Knowledge*. In all three, he was a pure foundationalist for a priori knowing, claiming that any a priori knowledge was either intuitively certain or known to be deducible from what was intuitively certain (see A priori §3). However, his account of empirical knowledge was a much more complicated story, containing a strong coherentist component. While claiming that certain internal mental states were 'self-presenting' and thus not in need of any evidential support, Chisholm consistently required substantial coherence among one's internal beliefs before one was justified in believing any proposition which went beyond the internal.

The foundational or anchoring self-presenting states and their corresponding beliefs consist of a variety of 'seemings'. What one seems to believe, seems to intend, seems to hope for or fear and, most importantly, seems to perceive, are all self-presenting; that is, if one is in any of these states and believes that one is in the state, then that belief is certain. Chisholm, following up a suggestion of his colleague Curt Ducasse, developed an adverbial theory of sensory experiencing. Chisholm urged that when referring to internal experiencing, we replace talk of 'sense-data' with adverbial expressions such as 'being appeared to redly'. In Chisholm's view, having the property of being appeared to redly was self-presenting and, thus, certain. The main advantage of this artificial way of speaking was the avoidance of ontological and epistemic puzzles about sense data. While 'being appeared to

redly' is somewhat comprehensible, the approach requires similar adverbs for all the other possible modes of sensing and suffers some implausibility because of their absence from ordinary language (see Mental states, adverbial theory of).

A fundamental characteristic of Chisholm's approach was to employ a fertile primitive term together with simple logical relationships to define other more complex concepts. For example, in the 3rd edition of *Theory of Knowledge* (1989b) the central foundational category of the 'certain' was defined in terms of the primitive comparative term 'more justified than' as follows:

p is certain for *S* =Df For every *q*, believing *p* is more justified for *S* than withholding *q*, and believing *p* is at least as justified for *S* as is believing *q*.

Definitions such as these occur throughout his work and it would not be inappropriate to label his approach 'Definitionalism'. Because of their clarity and simplicity, Chisholm's definitions generated numerous responses, mostly proposed counter-instances. Chisholm welcomed every example and never tired of modifying or 'chisholming away at' his definitions.

Chisholm handled epistemological scepticism by assuming that we have particular justified beliefs and the ability to examine them critically and thereby to discover principles of justification. These principles enabled Chisholm to construct a fifteen-category epistemic hierarchy, ranging from the certainly false to the certainly true. Each of the categories is defined in terms of the primitive 'more justified than'. Chisholm glosses this primitive in terms of an epistemic requirement to prefer. That is, when *S* is more justified in believing *p* than in believing *q*, *S* is required to prefer *p* to *q*. This epistemic requirement would then compete with all one's other kinds of requirement (including ethical), and from the competitive fray one's overall or real obligation would then emerge.

Chisholm was throughout his career inclined towards Platonism, always regarding attributes as abstract and eternal. The details of his ontology evolved and led him to seek a comprehensive theory of ontological categories. In *The First Person* (1981) Chisholm developed a theory of objective reference based around the primitive notion of 'direct attribution'. Direct attribution usually involves the 'emphatic reflexive' typified by such sentences as 'he believes that he himself is wise'. Chisholm uses his account of attribution to explain all *de re* or objective attributions.

Chisholm approached ethics theoretically, developing a theory of intrinsic value, and practically defending a kind of 'indeterminism'. The theory of intrinsic value employed undefined primitive terms to develop hierarchically arranged categories of both value and obligation. Chisholm's indeterminism, or theory of agent causation, affirmed free choices for which there were no internal or external conditions causally sufficient for that choice. Responsible agents themselves contribute the final condition which then brings about the actual choice and leads to the action.

See also: A posteriori §2; Commonsensism; Foundationalism; Free will; Internalism and externalism in epistemology; Knowledge and justification, coherence theory of; Perception; Scepticism

DAVID BENFIELD

List of works

A full bibliography can be found in Bogdan (1986).

Chisholm, R. (1957) *Perceiving: A Philosophical Study*, Ithaca, NY: Cornell University Press.(Exposition of the adverbial theory of perception, an account of adequate evidence based on an ethical primitive, and a Brentano-like view of intentionality. Carefully argued, yet accessible to the diligent beginning reader.)

Chisholm, R. (ed.) (1960) *Realism and the Background of Phenomenology*, Glencoe, IL: Free Press.(Collection of articles from 1870 to 1940 by German, British and American philosophers and psychologists concerned with the reality of objects of perception and thought. The editor's introduction is exceptionally useful and accessible, the articles themselves are of moderate complexity.)

Chisholm, R. (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 1st edn. (Closely argued exposition of entire epistemic system including a discussion of the metaphilosophical issues relevant to scepticism. Textbook for advanced students)

Chisholm, R. and Swartz, R. (eds) (1973) *Empirical Knowledge: Readings from Contemporary Sources*, Englewood Cliffs, NJ: Prentice Hall.(Collection of influential essays on scepticism, foundationalism, memory and belief. Most are accessible, but the essay by Sellars is difficult.)

Chisholm, R. (1976) *Person and Object: A Metaphysical Study*, La Salle, IL: Open Court.(Expanded version of the Carus Lectures on the self, agency, identity through time, and states of affairs; contains several appendices on justified belief and the part-whole relation. Readable but filled with definitions.)

Chisholm, R. (1977) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 2nd edn. (Significantly revised and lengthened, with greater emphasis on the use of primitive terms to define complex concepts. Textbook for advanced students.)

Chisholm, R. and Haller, R. (eds) (1978) *Die Philosophie Franz Brentanos: Beiträge zur Brentano-Konferenz, Graz, 1977, (The philosophy of Franz Brentano: Papers from the Brentano conference, Graz, 1977)*, Amsterdam: Editions Rodopi N.V.(Contributions to the 1977 Conference on Brentano's philosophy in Graz, Austria.)

Chisholm, R. (1981) *The First Person: An Essay on Reference and Intentionality*, Minneapolis, MN: University of Minnesota Press.(Analysis of all reference, especially the referring capacity of linguistic expressions, in terms of first person reference. Concise and rather technical.)

Chisholm, R. (1982) *The Foundations of Knowing*, Minneapolis, MN: University of Minnesota Press.(Collection of previously published but slightly revised essays and an update of his epistemic system involving three new, short essays. Most parts are highly readable.)

Chisholm, R. (1986) *Brentano and Intrinsic Value*, New York: Cambridge University Press. (Exposition and interpretation of Brentano's theory of value, linking it to his general views on intentionality. Rather technical)

Chisholm, R. (1989a) *On Metaphysics*, Minneapolis, MN: University of Minnesota Press. (Collection of previously published, but revised, essays on action, identity, reference, and ontology; almost all views expressed here are worked out in greater detail elsewhere. Readable.)

Chisholm, R. (1989b) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 3rd edn. (Substantially revised and shortened, with several chapters devoted to the metaphilosophical issues raised by the internalism-externalism and foundationalism-coherentism debates. Textbook for advanced students.)

Chisholm, R. (1996) *A Realistic Theory of Categories: An Essay on Ontology*, New York: Cambridge University Press.(An extremely concise reduction of all ontological entities using eight undefined categorial locutions. Includes all major metaphysical doctrines and central epistemological views, few real-world examples and many technical formulations. For advanced readers only.)

References and further reading

Bogdan, R.J. (ed.) (1986) *Roderick M. Chisholm*, Boston, MA: Reidel.(Contains a complete bibliography and a useful self-profile essay by Chisholm)

Brentano, F. (1889) *Vom Ursprung sittlicher Erkenntnis*, ed. and co-trans. R. Chisholm as *The Origin of Our Knowledge of Right and Wrong*, London: Routledge & Kegan Paul, 1969.(Translation of the third edition of 1934. A compact statement of Brentano's ethics which includes a theory of intrinsic value defined in terms of correct interests and preferences, supplemented by a utilitarian account of instrumental value.)

Brentano, F. (1930) *Wahrheit und Evidenz*, ed. and co-trans. R. Chisholm as *The True and the Evident*, London: Routledge & Kegan Paul, 1966.(Chronologically arranged lectures, letters, fragments and notes on the subject of truth, beginning with Brentano's version of the correspondence theory and ending with his view that true judgments are those which someone judging with evidence would accept.)

Brentano, F. (1976) *Philosophische Untersuchungen zu Raum, Zeit, und Kontinuuum*, ed. R. Chisholm and S. Kourner, Hamburg: Felix Meiner Verlag; trans. B. Smith, *Philosophical Investigations on Space, Time, and the Continuum*, London: Croom Helm, 1988.(Brentano's account of time-consciousness, his anti-Cantorian theory of the continuum and his unorthodox views on boundaries.)

Brentano, F. (1978) *Aristotle and his World-View*, ed. and co-trans. R. Chisholm, Los Angeles, CA: University of California Press.(Brentano's views as to what is essential in Aristotle. The sections on various aspects of Aristotle's Deity serve as an excellent introduction to Brentano's own views.)

Brentano, F. (1981) *The Theory of the Categories*, ed. and co-trans. R. Chisholm, The Hague: Martinus Nijhoff. (Brentano's views on various ontological issues.)

Fichte, J. (1800) *The Vocation of Man*, ed. and co-trans. R. Chisholm, New York: Liberal Arts Press, 1955.(Edited and readable translation which contains a very useful, yet concise, introduction to Fichte's Idealism and general philosophical stance.)

Kim, J. (1986) 'Critical review of *The First Person*', *Philosophy and Phenomenological Research* 46 (3): 483-507.(Sympathetic examination of the central doctrine that all reference is ultimately *de se* reference.)

Meinong, A. (1968-78) *Alexius Meinong Gesamtausgabe (Complete works of Alexius Meinong)*, Graz:

Akademische Druck- und Verlagsanstalt, 8 vols.(All the books and essays that Meinong published, together with his intellectual self-portrait and major reviews. Volumes 4 and 5 are edited by Chisholm. Volume 7 contains a comprehensive index. The supplementary volume contains a selection of posthumous material including lecture notes, notes on other philosophers' writings, and a lexicon of logic and epistemology.)

Chomsky, Noam (1928-)

Fish swim, birds fly, people talk. The talents displayed by fish and birds rest on specific biological structures whose intricate detail is attributable to genetic endowment. Human linguistic capacity similarly rests on dedicated mental structures many of whose specific details are an innate biological endowment of the species. One of Chomsky's central concerns has been to press this analogy and uncover its implications for theories of mind, meaning and knowledge.

This work has proceeded along two broad fronts.

First, Chomsky has fundamentally restructured grammatical research. Due to his work, the central object of study in linguistics is 'the language faculty', a postulated mental organ which is dedicated to acquiring linguistic knowledge and is involved in various aspects of language-use, including the production and understanding of utterances. The aim of linguistic theory is to describe the initial state of this faculty and how it changes with exposure to linguistic data. Chomsky (1981) characterizes the initial state of the language faculty as a set of principles and parameters. Language acquisition consists in setting these open parameter values on the basis of linguistic data available to a child. The initial state of the system is a Universal Grammar (UG): a super-recipe for concocting language-specific grammars. Grammars constitute the knowledge of particular languages that result when parametric values are fixed.

Linguistic theory, given these views, has a double mission. First, it aims to 'adequately' characterize the grammars (and hence the mental states) attained by native speakers. Theories are 'descriptively adequate' if they attain this goal. In addition, linguistic theory aims to explain how grammatical competence is attained. Theories are 'explanatorily adequate' if they show how descriptively adequate grammars can arise on the basis of exposure to 'primary linguistic data' (PLD): the data children are exposed to and use in attaining their native grammars. Explanatory adequacy rests on an articulated theory of UG, and in particular a detailed theory of the general principles and open parameters that characterize the initial state of the language faculty (that is, the biologically endowed mental structures).

Chomsky has also pursued a second set of concerns. He has vigorously criticized many philosophical nostrums from the perspective of this revitalized approach to linguistics. Three topics he has consistently returned to are:

- *Knowledge of language and its general epistemological implications*
- *Indeterminacy and underdetermination in linguistic theory*
- *Person-specific 'I-languages' versus socially constituted 'E-languages' as the proper objects of scientific study.*

1 The aims and principles of linguistic theory

There is an intimate relation between how a problem is conceived and the kinds of explanations one should offer. Chomsky proposes that we identify explanation in linguistics with a solution to the problem of how children can attain mastery of their native languages on the basis of a rather slender database. This is often referred to as 'the logical problem of language acquisition'.

A natural language assigns meanings to an unbounded number of sentences. Humans typically come to master at least one such language in a surprisingly short time, without conscious effort, explicit instruction or apparent difficulty. How is this possible? There are significant constraints on any acceptable answer.

First, a human can acquire any language if placed in the appropriate speech community. Grow up in Boston and one grows up speaking English the way Bostonians do. However, the 'primary linguistic data' (PLD) available to the child are unable to guide the task unaided. There are four kinds of problems with the data that prevent it from shaping the outcome:

(a) The set of sentences the child is exposed to is finite. However, the knowledge attained extends over an unbounded domain of sentences.
(b) The child is exposed not to sentences but to utterances of sentences. These are imperfect vehicles for the transmission of sentential information as they can be defective in various ways. Slurred speech, half sentences, slips of the tongue and mispronunciations are only a few of the ways that utterances can obscure sentence

structure.
(c) Acquisition takes place without explicit guidance by the speech community. This is so for a variety of reasons. Children do not make many errors to begin with when one considers the range of logically possible mistakes. Moreover, adults do not engage in systematic corrections of errors that do occur and even when correction is offered children seem neither to notice nor to care. At any rate, children seem surprisingly immune to any form of adult linguistic intrusion (see Lightfoot 1982).
(d) Last of all, and most importantly, of the linguistic evidence theoretically available to the child, it is likely that only simple sentences are absorbed. The gap between input and intake is attributable to various cognitive limitations such as short attention span and limited memory. This implies that the acquisition process is primarily guided by the information available in well-formed simple sentences. Negative data (the information available in unacceptable ill-formed sentences) and complex data (the information yielded by complex constructions) are not among the PLD that guide the process of grammar acquisition. The child constructing its native grammar is limited to an informationally restricted subset of the relevant data. In contrast to the evidence that the linguist exploits in theory construction, the information the child uses in building its grammar is severely restricted. This suggests that whenever the linguistic properties of complex clauses diverge from simple ones, the acquisition of this knowledge cannot be driven by data. Induction is insufficient as the relevant information is simply unavailable in the PLD.

The general picture that emerges from these considerations is that attaining linguistic competence involves the acquisition of a grammar, and that humans come equipped with a rich innate system that guides the process of grammar construction. This system is supple enough to allow for the acquisition of any natural language grammar, yet rigid enough to guide the process despite the degeneracy and deficiency of the PLD. Linguistic theorizing takes the above facts as boundary conditions and aims both at descriptive adequacy (that is, to characterize the knowledge that speakers have of their native grammars) and explanatory adequacy (that is, to adumbrate the fine structure of the innate capacity) (see Language, innateness of).

Issues of descriptive and explanatory adequacy have loomed large in Chomsky's work since the beginning. Chomsky's objection, for example, to 'Markov models' of human linguistic competence was that they were incapable of dealing with long distance dependencies exemplified by conditional constructions in English and hence could not be descriptively adequate. His argument in favour of a transformational approach to grammar rested on the claim that it allowed for the statement of crucial generalizations evident in the judgments of native speakers and so advanced the goal of descriptive adequacy (Chomsky 1957). Similarly, his influential critique (1959) of Skinner's *Verbal Behavior* consisted in showing that the learning theory presented therein was explanatorily inadequate. It was either too vague to be of scientific value or clearly incorrect given even moderately precise notions of stimulus or reinforcement.

The shift from the early *Syntactic Structures* (1957) theory to the one in *Aspects of a Theory of Syntax* (1965) was also motivated by concerns of explanatory adequacy. In the earlier model the recursive application of transformations allows for the generation of more and more complex sentences from the sentences produced by the 'phrase structure' component of the grammar. In the *Aspects* theory, recursion is incorporated into the phrase structure component itself, and removed from the transformational part of the theory (see Syntax §3). The impetus for this was the observation that greater explanatory adequacy could be attained by grammars that had a level of 'Deep Structure' incorporating a recursive base component. In particular, Fillmore (1963) observed that the various optional transformations in a *Syntactic Structures* theory always applied in a particular order in any given derivation. This order is unexplained in a *Syntactic Structures* theory; in *Aspects* it is deduced. Thus, the move to an Aspects-style grammar is motivated on grounds of greater explanatory adequacy: introducing Deep Structure and moving recursion to the base allows for a more restricted theory of Universal Grammar. All things being equal, restricting UG is always desirable as it advances a central goal of grammatical theory; the more restricted the options innately available for grammar construction, the easier it is to explain how language acquisition is possible, despite the difficulties in the PLD noted above.

The same logic motivates various later additions to and shifts in grammatical theory. For example, a major move in the 1970s was radically to simplify transformational operations so as to make their acquisition easier. This involves eliminating any mention of construction-specific properties from transformational rules. For example, an *Aspects* rule for passive constructions looks like (1), the left-hand side being the Structural Description (SD) and the right hand side being the Structural Change (SC):

to X-NP1-V-NP2-Y→ $(1)\mathrm{X}-\mathrm{NP2}-\mathrm{be}+\mathrm{en}\ \mathrm{V}-\mathrm{by}+\mathrm{NP1}-\mathrm{Y}\mathrm{x}$

This rule would explain, for instance the grammaticality of ‘the ball is kicked by John’ given that of ‘John kicks the ball’. Observe that the SC involves the constants ‘ $\mathrm{be}+\mathrm{en}$’ and ‘by’. The SD mentions three general expressions, ‘NP1’, ‘V’ and ‘NP2’ and treats these as part of the context for the application of the rule. In place of this, Chomsky proposed eliminating the passive rule and replacing it with a more general rule that moves NPs (Chomsky 1977, 1986). The passive rule in (1) involves two applications of the ‘Move NP’ rule, one moving the subject ‘NP1’ to the ‘by’ phrase, and another moving the object ‘NP2’ to the subject position. In effect, all the elements that make the passive rule in (1) specific to transitive constructions are deleted and a simpler rule (‘Move NP’) replaces it.

There is a potential empirical cost to simple rules, however. The simpler a transformation the more it generates unacceptable outputs. Thus, while a grammar with (1) would not derive ‘was jumped by John’ from ‘John jumped’, a grammar eschewing (1) and opting for the simpler ‘Move NP’ rule is not similarly restricted. To prevent overgeneration, therefore, the structure of UG must be enriched with general grammatical conditions that function to reign in the undesired overgeneration (Chomsky 1973, 1977, 1986). Chomsky has repeatedly emphasized the tension inherent in developing theories with both wide empirical coverage and reasonable levels of explanatory adequacy.

A high point of this research agenda is Chomsky’s *Lectures on Government and Binding* (1981). Here the transformational component is reduced to the extremely simple rule ‘Move *a*’ - that is, move anything anywhere. To ensure that this transformational liberty does not result in generative chaos, various additions to the grammar are incorporated, many conditions on grammatical operations and outputs are proposed, and many earlier proposals (by both Chomsky and others) are refined. Among these are trace theory, the binding theory, bounding theory, case theory, theta theory and the Empty Category Principle. The picture of the grammar that Chomsky’s *Lectures* presents is that of a highly modular series of interacting subsystems which in concert restrict the operation of very general and very simple grammatical rules. In contrast to earlier traditional approaches to grammar, *Lectures* witnesses the virtual elimination of grammatical constructions as theoretical constructs. Thus, in Government Binding (GB)-style theories there are no rules of Passive, Raising, Relativization or Question Formation as there were in earlier theories. Within GB, language variation is not a matter of different grammars having different rules. Rather, the phenomena attested in different languages are deduced by variously setting the parameters of Universal Grammar. Given the interaction of the grammatical modules, a few parametric changes can result in what appear on the surface to be very different linguistic configurations. In contrast to earlier approaches to language, variation consists not in employing different kinds of rules, but in having set the parameters of an otherwise fixed system in somewhat different ways (see Chomsky 1983).

The GB research programme has proven to be quite successful in both its descriptive range and its explanatory appeal. Despite this, Chomsky has urged a yet more ambitious avenue of research. He has embarked on the development of a rationalist approach to grammar that goes under the name of ‘Minimalism’ (Chomsky 1995). The theory is ‘rationalist’ both in that it is grounded on very simple and perspicuous first principles, and in that it makes use only of notions required by ‘virtual conceptual necessity’. Chomsky hopes to make do with concepts that no approach to grammar can conceivably do without and remain true to the most obvious features of linguistic competence. For example, every theory of grammar treats sentences as pairings of sounds and meanings. Thus, any theory will require that every sentence have a phonological and an interpretative structure. In GB theories, these sorts of information are encoded in the PF (Phonetic Form) and LF (Logical Form) phrase markers respectively. In addition, GB theories recognize two other distinctive grammatical levels: S-structure and D-structure. A minimal theory, Chomsky argues, should dispense with everything but LF and PF. It will be based on natural ‘economy’ principles and indispensable primitives. Chomsky has suggested reanalysing many of the restrictions that GB theories impose in terms of ‘least effort’ notions such as ‘shortest move’ and ‘last resort movement’. For example, he proposes that the unacceptability of sentences such as ‘John is expected will win’, are ultimately due to the fact that the moved NP ‘John’ need not have moved from the embedded subject position (between ‘expected’ and ‘will’) as it fulfils no grammatical requirement by so moving. This work is still in its infancy, but it has already prompted significant revisions of earlier conclusions. For example, with the elimination of D-structure, the recursive engine of the grammar has once again become the province of generalized transformations. Whatever its

ultimate success, however, Minimalism continues the pursuit of the broad goals of descriptive and explanatory adequacy enunciated in Chomsky's earliest work.

2 Knowledge of language

According to Chomsky, the three fundamental epistemological questions in the domain of language are 'What constitutes knowledge of language?', 'How is knowledge of language acquired?' and 'How is this knowledge put to use?'. The answer to the first question is given by a particular generative grammar. Harold's knowledge of English is identified with Harold's being in a particular mental/brain state. A descriptively adequate grammar characterizes this part of Harold's mental/brain make-up. An answer to the second question is provided by a specification of UG and the principles that take the initial state of the language faculty to the knowledgeable state on exposure to PLD. Harold knows English in virtue of being genetically endowed with a language faculty and having been normally brought up in an English-speaking community. Beyond this, further issues of grounding are unnecessary. Issues of epistemological justification and grounding in the data are replaced by questions concerning the fine structure of the initial state of the language faculty and how its open parameters are set on the basis of PLD. The third question is answered by outlining how linguistic knowledge interacts with other cognitive capacities and abilities to issue in various linguistic acts such as expressing one's thoughts, parsing incoming speech and so on (see Chomsky 1986).

How much does the language case tell us about epistemological issues in other domains? In other words, should knowledge of quantum mechanics be analysed in a similar vein, that is, being in a particular mental state, grounded in specific innate capacities and so on. Chomsky only makes sparse comments on this general issue, but those he advances suggest that he believes that knowledge in these domains should be approached in much the same way they are approached in the domain of language. This suggests that humans have an innate science-forming capacity that underlies our success in the few domains of inquiry in which there has indeed been scientific success. As in the domain of language, this capacity is focused and modular rather than being a general all-purpose tool and this, Chomsky speculates, might well underlie the patchiness of our successes. Where we have the right biological propensities, we develop rich insightful theories that far outpace the data from which they are projected. Where this mind/brain structure is lacking, mysteries abound that seem recalcitrant to systematic inquiry. Stressing our cognitive limits is a staple of Chomsky's general epistemological reflections. If humans are part of the natural world we should expect there to be problems that fall within our cognitive grasp and mysteries that lie outside it. The rich theoretical insights allowed in the natural sciences are the result of a chance convergence between properties of the natural world and properties of the human mind/brain (see Chomsky 1975).

3 Indeterminacy and underdetermination

Knowledge of language, Chomsky has argued, presents a strong argument in favour of traditional rationalist approaches to mind and against traditional empiricist approaches (see Learning §1; Rationalism). In particular, 'learning' is treated as more akin to growth and the course of acquisition is seen more as the unfolding of innate propensities under the trigger of experiential input than as the result of the shaping effects of the environment. This rationalist perspective is now quite common and this is largely due to Chomsky's efforts. Chomsky has consistently warned against empiricist prejudices in philosophy, and in no instance more strongly than in his critique of Quine's methodological remarks on linguistics (for example, see Quine 1960).

Chomsky takes Quine to be arguing that linguistic investigations are beset with problems greater than those endemic to inquiry in general. Whereas empirical investigation in general suffers from underdetermination of theory by evidence, linguistic study is beset with the added problem of indeterminacy (see Radical translation and radical interpretation §§2-3). Indeterminacy differs from standard inductive underdetermination (see Underdetermination) in that where there is indeterminacy 'there is no real question of right choice' among competing proposals. Chomsky interprets Quine as arguing that 'determining truth in the study of language differs from the problem of determining truth in the study of physics' (Chomsky 1975: 182-3).

In reply, Chomsky (1969) argues that Quine's thesis rests on classical empiricist assumptions about how languages are acquired. Quine, he argues, supposes that humans have 'an innate quality space with a built-in distance measure' tuned to certain 'simple physical correlates'. In addition, certain kinds of induction in this space are permitted. Beyond this, however, 'language-learning is a matter of association of sentences to one another and to

certain stimuli through conditioning'. Further, one cannot 'make significant generalizations about language or common-sense theories, and the child has no concept of language or of "common-sense" prior to this training' (Chomsky 1969: 54-5, 63).

Chomsky notes that Quine provides no evidence to support these assumptions. Nor can there be any good evidence to support them if the nature of the learning problem in the domain of language is characterized as Chomsky has argued it must be. Chomsky concludes that 'Quine's thesis of the indeterminacy of translation amounts to an implausible and quite unsubstantiated empirical claim about what the mind brings to the problem of acquisition of language (or of knowledge in general) as an innate property' (Chomsky 1969: 66). Stripped of these tendentious empirical assumptions, Quine fails to show that indeterminacy is anything other than the familiar problem of underdetermination of theory by evidence as applied to linguistics. Chomsky (1996) has since argued that the ultimate source of many critiques of the mental sciences in general and linguistics in particular (including Quine's indeterminacy thesis) is a kind of methodological dualism that takes humans to be separate from the natural world. This dualism is manifest in the a priori constraints that philosophers place on explanations in the mental sciences, which would be regarded as inappropriate if applied to the physical sciences.

In this vein Chomsky asks, for example, why access to consciousness is so often taken to be crucial in substantiating the claim that humans have I-language or follow rules. Suppose, he asks, we had a theory that perfectly described what happens when sound waves hit the ear, stimulating the performance system to access the cognitive system and construct a logical form that interacts with other cognitive systems to yield comprehension, in so far as the language faculty enters into this process. What more could be desired? The insistence that this entire process be accessible to consciousness in order for the account to be credible, he argues, is a demand beyond naturalism, a form of methodological dualism of dubious standing that would be summarily rejected if raised elsewhere.

Or consider the oft-voiced suspicions concerning mentalist approaches in psychology. Many philosophers are ready to accept these as perhaps temporarily necessary but ultimately, the view seems to be, mentalist theories must reduce to physical ones to be truly legitimate. Chomsky argues that this sentiment is another manifestation of methodological dualism and should be rejected. First, it presupposes that there is a tenable distinction between the mental and the physical. However, Chomsky argues that since Newton undermined the Cartesian theory of body by showing that more 'occult' forces were required in an adequate physics, mind-body dualism has lost all grounding. Second, even if reduction were possible, reduction comes in many varieties and there is little reason to believe that the contours of the reducing physical theory would be left unaffected by the process. Since Newton, Chomsky notes, 'physical' has been an honorific term that signifies those areas in which we have some nontrivial degree of theoretical understanding. The relevant scientific question is whether some theory or other offers interesting descriptions and explanations. The further insistence that its primitives be couched in physical vocabulary is either vacuous (because 'physical' has no general connotation) or illegitimate (another instance of methodological dualism).

The general conclusion Chomsky draws is that whatever problems linguistic theory encounters, it is no more methodologically problematic than theories in other domains. He attributes the qualms of philosophers to lingering empiricist dogma or an indefensible epistemological dualism.

4 I-language versus E-language

Given the aims of Chomskian linguistic theory, the proper objects of study are the I-languages internalized by native speakers, rather than public E(xternal)-languages used by populations. Chomsky denies that public E-languages are interesting objects of scientific study. Indeed he denies that E-languages can be coherently specified as they simply do not exist. The proper objects of inquiry are I-languages; 'I' standing for intensional, internal and individual. An I-language is individual in that each speaker has one. This focus turns the common wisdom on its head. E-languages like English, Swahili and so forth are (at best) radical idealizations for Chomsky, or (at worst) incoherent pseudo-objects. At best, E-languages are the intersection of the common properties of various I-languages. Thus, for example, it is not that speakers communicate because they have a language in common; rather wherever I-languages *overlap* communication is possible.

An I-language is internal in the sense of being part of a speaker's individual mental make-up. It is neither a

Platonic object nor a social construct. Also, an I-language is intensional, not extensional. Comprised as it is of an unbounded number of sentences, a language cannot be 'given' except via a specification of the function that generates them, that is a grammar for that language. Thus, it is languages in intension, languages dressed in all of their grammatical robes, not simple concatenations of words, that are the proper objects of scientific interest. One consequence of this is that weak generative capacity (that is, the extensional equivalence of languages generated by different grammars) is of dubious interest. In short, the shift from E-language to I-language turns many long-standing questions around, raising some to prominence that were considered secondary and relegating many that previously were considered crucial to the status of pseudo-questions.

Many philosophers have found Chomsky's focus on I-language problematic. To illustrate, we will consider an important philosophical critique and Chomsky's reply.

Dummett (1986) argues against internalist approaches to language that they fail to provide an account of notions like 'language of a community' or 'community norms' in the sense presupposed by virtually all work in the philosophy of language and philosophical semantics. These notions, Dummett claims, are required to provide a notion of a common public language which 'exists independently of any particular speakers' and of which native speakers have a 'partial, and partially erroneous, grasp' (see Language, social nature of §2).

The naturalistic study of language, Chomsky counters, has no place for a Platonistic notion of language, a notion of language outside the mind/brain that is common to various speakers and to which each speaker stands in some cognitive relation. The reason is that this Platonistic reification rests on notions like 'language' and 'community' that are hopelessly under-specified. Asking if two people speak the same language is, in Chomsky's opinion, to ask a highly context-dependent question - much like asking whether Boston is near New York. What counts as a community depends on shifting expectations of individuals and groups. Human society is not neatly divided into communities with languages and their norms. Thus, what counts as a community is too under-specified to be useful for theoretical purposes. Therefore, it is not a defect of linguistic theory that these notions play no role within it.

From Chomsky's perspective E-languages are epiphenomenal objects, if coherent at all. I-language in its universal aspects is part of the human genotype and specifies one aspect of the human mind/brain. Under the triggering effects of experience a particular grammar arises in the mind/brain of an individual. From this perspective, universal grammar and the steady-state grammars that arise from them are real objects. They will be physically realized in the genetic code and the adult brain. E-language, in contrast, has a murky ontological status. Chomsky (1980) argues that the priority of I-language cannot be reasonably doubted once we observe that languages involve an infinite pairing of sounds and meanings. Given that language is infinite, it cannot be specified except in so far as some finite characterization - a function in intension - is provided. It might be possible to give some characterization to the notion 'a language used by a population' but only indirectly via a grammatical specification of the language. But this concedes the priority of I-language as the claim unpacks into something like: each person in the relevant population has a grammar in their mind/brain that determines the E-language. Thus, at best, an E-language is that object which the I-language specifies. However, even this might be giving too much reality to E-languages, for there is nothing in the notion I-language that requires that what they specify corresponds to languages as commonly construed, that is, things like French, English and so on. It is consistent with Chomsky's viewpoint that I-language never specifies any object that we might pre-theoretically call a language. Whether this is indeed the case, the key point is to realize that the move from grammar to language is a step away from real mechanisms to objects of a higher degree of abstraction. I-language is epistemologically and ontologically hardier than E-language, much philosophical opinion to the contrary.

See also: Language, innateness of; Language, philosophy of; Language, social nature of §3; Nativism; Unconscious mental states §4

NORBERT HORNSTEIN

List of works

Chomsky, N. (1957) *Syntactic Structures*, The Hague: Mouton.(First work on Transformational Grammar.)
Chomsky, N. (1959) 'Review of *Verbal Behavior* by B.F. Skinner', *Language* 35: 26-58.(A critique of behaviourist approaches to learning.)
Chomsky, N. (1965) *Aspects of a Theory of Syntax*, Cambridge, MA: MIT Press.(Outlines the Standard Model.)
Chomsky, N. (1969) 'Quine's Empirical Assumptions', in D. Davidson and J. Hintikka (eds) *Words and*

Objections, Dordrecht: Reidel.
Chomsky, N. (1973) ‘Conditions on Transformations’, in S.R. Anderson and P. Kiparsky (eds) *A Festschrift for Morris Halle*, New York: Holt, Rinehart & Winston.(Begins the move away from rule-based approaches to grammar.)
Chomsky, N. (1975) *Reflections on Language*, New York: Pantheon.(A good non-technical review of the extended standard theory and various philosophical issues related to generative grammar.)
Chomsky, N. (1977) *Essays on Form and Interpretation*, Amsterdam: North Holland.(Essays in the extended standard theory.)
Chomsky, N. (1980) *Rules and Representations*, New York: Columbia University Press.(Essays on linguistics and philosophy.)
Chomsky, N. (1981) *Lectures on Government and Binding*, Dordrecht: Foris.
Chomsky, N. (1983) ‘Some Conceptual Shifts in the Study of Language’, in L. Cauman, I. Levi, C. Parsons and R. Schwartz (eds) *How Many Questions?: essays in honor of Sidney Morgenbesser*, Indianapolis, IN: Hackett.(A description of how linguistic theory has changed from Syntactic Structures to Local Government Binding.)
Chomsky, N. (1986) *Knowledge of Language*, New York: Praeger.(Chapter 3 provides an informal yet challenging overview of Government Binding Theory.)
Chomsky, N. (1995) *The Minimalist Program*, Cambridge, MA: MIT Press.(Chomsky’s best current text on Minimalism.)
Chomsky, N. (1996) *Powers and Prospects*, Boston, MA: South End Press.(More recent philosophical esays on E-language and dualism.)

References and further reading

Dummett, M.A.E. (1986) ‘Comments on Davidson and Hacking’, in E. Lepore (ed.) *Truth and Interpretation*, Oxford: Blackwell.(Argues in favour of the importance of E-languages.)
Fillmore, C.J. (1963) ’The Position of Embedding Transformations in a Grammar’, *Word* 19: 208-31.(A technical critique of Generalized Transformations.)
Haegeman, L. (1994) *Introduction to Government and Binding Theory*, Oxford: Blackwell.(A good textbook on Government Binding Theory.)
Lightfoot, D.W. (1982) *The Language Lottery*, Cambridge, MA: MIT Press.(A good introduction to the logic of linguistic research.)
Pinker, S. (1994) *The Language Instinct*, New York: Morrow.(Combines Darwin and Chomsky to argue that linguistic competence is a human instinct rather than cultural phenomenon. Good introduction to linguistic research.)
Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Argues for the radical indeterminacy of certain aspects of linguistic theory.)
Skinner, B.F. (1957) *Verbal Behavior*, New York: Appleton-Century-Crofts.(Presents the behaviourist account of language that Chomsky influentially criticized.)
Webelhuth, G. (ed.) (1995) *Government and Binding Theory and the Minimalist Program*, Oxford: Blackwell.(A very good advanced text on Government Binding Theory.)

Christine de Pizan (1365-*c.*1430)

Christine de Pizan, France's 'first woman of letters', is primarily remembered as a courtly poet and a propagandist for women. Her extensive writings were influenced by the early humanists, reflecting an interest in education (particularly for women and young people) and in government. Following Aquinas, Christine defined wisdom as the highest intellectual virtue and tried to apply the concept of the just war to contemporary problems. Her works are also noteworthy for their contribution to the transmission of Italian literature to Parisian intellectual circles.

Christine de Pizan was born in Venice. Her father and grandfather, city officials in Venice and graduates of the University of Bologna, would undoubtedly have known Petrarch (see Petrarca, F.), and possibly Boccaccio. Christine's father, an astrologer and medical man of some reputation, was invited to the French court of the cultivated King Charles V, and Christine grew up as part of Charles' entourage. Her father encouraged Christine's early taste for study, and she probably received a basic education alongside her two brothers until her marriage, at the age of fifteen, to Etienne Castel, a young notary who was soon appointed a royal secretary. Widowed at twenty-five, with her father already dead, Christine found herself in desperate financial straits with three children to support. However, she managed to obtain patronage through her taste for study and subsequently for writing. Her attachment to early humanistic interests can be attributed to her programme of self-education during these early years of widowhood, as well as to her father's early influence and possibly to his library.

A debate with young Parisian intellectuals, in which Christine insisted on the superiority of Dante's work over Jean de Meun's *Roman de la Rose (Romance of the Rose)*, strengthened her reputation. Becoming successful as a court poet, Christine was one of the first scholars to introduce Dante to France, using him in her first long work, *L'Epître d'Othéa à Hector (The Letter of Othea to Hector)* (*c.*1400) and her semi-autobiographical poem *Le Chemin de Long Estude (The Long Road of Study)* (*c.*1403). Her *Livre de la Cité des Dames (Book of the City of Ladies)* (1405) was inspired in large measure by Boccaccio's *De Claris Mulieribus (Concerning Outstanding Women)*, and to a lesser degree by his *Decameron*, retelling selected tales from a woman's point of view. In addition, the influence of Boccaccio's concept of poetry (as expressed in his *De Genealogia Deorum Gentilium (A Genealogy of Gentile Gods)*) can be discerned in her *Letter of Othea*. Her debt to Petrarch is more elusive, although she mentions him in *The City of Ladies* and reworks his version of the Griselda story there.

Her long allegorical poem *La Mutacion de Fortune (The Mutation of Fortune)* (1403) led to a commission from the duke of Burgundy to write a biography of his late brother, *Les Fais et Bonnes Meurs du Sage Roy Charles V (The Deeds and Good Customs of the Wise King Charles V)*. In this work she did not just give a historical account of Charles' life, but portrayed him as an exemplary ruler and a philosopher, a model of the philosopher-king. Drawing inspiration from Aquinas' commentary on Aristotle's *Metaphysics* she proposed a definition of wisdom as the highest intellectual virtue and the philosophical life as the prerogative of the free individual.

In common with other early humanists, Christine shows particular interest in the education of the young, notably the French dauphin, Louis de Guyenne. She wrote *Le Livre du Corps de Policie (The Book of the Body Politic)* (*c.*1406) and *Le Livre de la Paix (The Book of Peace)* (1413) for Louis, and her *Les Fais d'Armes et de Chevalerie (Deeds of Arms and of Chivalry)* (1410) was probably intended to form part of his military education. Inspired by the Roman author Vegetius as well as by contemporary needs, *Les Fais d'Armes* was the first professional military treatise to have appeared in some time. The first chapters comprise a significant discussion of the concept of the just war, with Christine attempting to apply traditional ideas to the problems of her own time. The book subsequently played a role in the reform of the French army in the second half of the fifteenth century. Her *Livre des Trois Vertus (Book of Three Virtues)* (1405), devoted to the education of women and dedicated to the dauphin's wife, Marguerite of Burgundy, was also significant: it was reprinted as late as 1536, and influenced European leaders including Margaret of Austria, Anne of France and Louise of Savoy.

See also: Humanism, Renaissance

CHARITY CANNON WILLARD

List of works

Christine de Pizan (*c.*1400) *L'Epître d'Othéa à Hector*, trans. J. Chance, *The Letter of Othea to Hector, with Introduction, Notes, and Interpretive Essay*, Newport, RI: Focus Library of Medieval Women, 1990.
Pizan (*c.* 1403) *Le Livre du Chemin de Long Estude (The Long Road of Study)*, ed. R. Püschel, Berlin: Damköhler and Paris: Le Soudier, 1881; new edn, Berlin: Hettler, 1887; repr. of 1887 edn, Geneva: Slatkine, 1974.(A good text, but limited critical apparatus.)
Pizan (1403) *Le Livre de la Mutacion de Fortune (The Mutation of Fortune)*, ed. S. Solente, Paris: Picard, 1959-66, 4 vols.(An excellent first edition.)
Pizan (1404) *Le Livre des Fais et Bonnes Meurs du Sage Roy Charles V (The Deeds and Good Customs of the Wise King Charles V)*, ed. S. Solente, Paris: Champion, 1936-40.(An excellent edition; fully documented.)
Pizan (1405) *Le Livre de la Cité des Dames*, trans. E.J. Richards, *The Book of the City of Ladies*, New York: Persea, 1982.(Argues for the equality of women.)
Pizan (1405) *Le Livre des Trois Vertus (Book of Three Virtues)*, trans. C.C. Willard, in M. Pelner Cosman (ed.) *A Medieval Women's Mirror of Honor; the Treasury of the City of Ladies*, New York: Bard Hall Press/Persea, 1989.
Pizan (*c.*1406) *Le Livre du Corps de Policie*, trans. and ed. K. Langdon Forhan, *The Book of the Body Politic*, Cambridge: Cambridge University Press, 1994.(Major political work; argues for the interdependence of rulers and subjects.)
Pizan (1410) *Le Livre des Fais d'Armes et de Chevalerie*, trans. W. Caxton, *The Book of Faytes of Armes and Chyvalrye*, ed. A.T.P. Byles, London: Early English Text Society, 1932; revised edn, 1937.
Pizan (1413) *Le Livre de la Paix (The Book of Peace)*, ed. C.C. Willard, The Hague: Mouton, 1958.(An excellent edition with introduction and notes.)
Pizan (1993) *The Writings of Christine de Pizan*, ed. C.C. Willard, New York: Persea.(A wide selection of Christine's writings in English translation with notes and bibliography.)
Pizan (1997) *The Selected Writings of Christine de Pizan*, trans. R. Blumenfield-Kosinski and K. Brownlee, ed. R. Blumenfield-Kosinski, New York and London: W.W. Norton.

References and further reading

Batard, Y. (1973) 'Dante et Christine de Pisan', in *Missions et Démarches de la Critique: Mélanges offerts au Professeur J.A. Vier*, Paris: Klincksieck, 345-55.(Basic study of Christine's influence.)
Bozzolo, C. (1967) 'Il *Decameron* come fonte del *Livre de la Cité des Dames* de Christine de Pisan' (The *Decameron* as the Source of Christine de Pizan's *The Book of the City of the Ladies*), in F. Simone (ed.) *Miscellanea di Studi e Richerche sul Quattrocento Francese*, Torino: Giappichelli, 3-24.(Good analysis of Christine's literary borrowings.)
Davis, N.Z. (1980) 'Gender and Genre: Woman as Historical Writers, 1400-1820', in P.H. Labalme (ed.) *Beyond Their Sex: Learned Women in the European Past*, New York: New York University Press, 153-60.(Brief view of Christine as a writer of history, notably in her biography of Charles V.)
Hicks, E. (ed.) (1977) *Le Débat sur le Roman de la Rose*, Paris: Champion.(Good discussion of the debate over the relative merit of Dante and Jean de Meun; includes work by the participants.)
Ornato, E. (1961) 'La prima fortuna del Petrarca in Francia' (The Early Fortunes of Petrarch in France), *Studi Francesi* 5: 201-17, 401-14.(Basic study of Christine de Pizan's influence in the popularization of Petrarch's work.)
Willard, C.C. (1984) *Christine de Pizan. Her Life and Works*, New York: Persea.(General discussion with modern bibliography.)

Chrysippus (*c*.280-*c*.206 BC)

The Greek philosopher Chrysippus of Soli was the third and greatest head of the Stoic school in Athens. He wrote voluminously, and in particular developed Stoic logic into a truly formidable system. His philosophy is effectively identical with 'early Stoicism'.

Chrysippus was born at Soli in Cilicia (southern Turkey). He came to Athens to study philosophy, initially with the Academic sceptic Arcesilaus. By the time he transferred his allegiance to the Stoic school, its founder Zeno of Citium was dead and had been succeeded by Cleanthes, who became Chrysippus' teacher. For a while he taught philosophy outside the Stoa, but in 232 BC Cleanthes died and Chrysippus succeeded him, holding the office till his own death.

Chrysippus may well have been relatively poor. In his work *On Livelihoods* he tackled the question of how a philosopher might appropriately earn a living. The only three acceptable means, he concluded, were serving a king (if one could not oneself be a king), reliance on friends, and teaching. There is no evidence that Chrysippus, like some other Stoics, adopted the first of these practices.

From the long excerpts of his writings which have survived in other authors, it seems that his Greek was clumsy and often obscure. (The people of his native Soli were so notorious for their poor Greek as to have given the 'solecism' its name.) Nevertheless, he wrote prolifically, his works totalling at least 705 rolls of papyrus, and what he wrote was widely read. Often he returned to the same topic several times in different treatises. His writings were also accused of beng padded out with endless quotations. Not all were doctrinal expositions - some were investigative and open-ended.

'If there had been no Chryippus', it was said, 'there would have been no Stoa'. Chrysippus is generally regarded as the person who built the full Stoic system. (The entry Stoicism is therefore in effect an account of his philosophy.) Whether Chrysippean Stoicism should be, as it often is, called 'orthodox' Stoicism is less clear. Later Stoics did not treat him as altogether authoritative: some, including his successor Antipater as well as Posidonius, were openly critical of him. On the other hand, his writings became and remained the classic Stoic texts to cite and analyse. Exegesis of passages from Chrysippus was a later Stoic teaching method, for example in the school of Epictetus, and conversely his writings were the primary target of anti-Stoic polemicists such as Plutarch and Galen. But Chrysippus presented Zeno rather than himself as the voice of authority, and many of his ideas were developed as defences and interpretations of Zeno's pronouncements, often against the rival interpretations of Cleanthes and others.

Chrysippus contributed extensively to every area of Stoic thought, with the possible exception of epistemology. Above all, he was the school's master logician, and the list of his writings partially preserved by Diogenes Laertius includes an astonishing 118 titles of logical treatises. They include at least seven works, filling fifteen rolls, on the Liar Paradox alone.

See also: Logic, ancient

DAVID SEDLEY

References and further reading

Arnim, H. von (1903-5) *Stoicorum Veterum Fragmenta (Fragments of the Early Stoics)*, Leipzig: Teubner, with vol. 4, indexes, by M. Adler, 1924.(The standard collection of early Stoic fragments, in Greek and Latin, commonly abbreviated as *SVF*. Chrysippus fills almost the whole of volumes 2-3, although only a minority of the fragments include an explicit attribution to him.)

Brehier, E. (1910) *Chrysippe et l'ancien stoïcisme*, Paris: Presses Universitaires de France; 2nd edn, Paris: Alcan, 1951.(A still valuable assessment of Chrysippus' contribution to Stoicism.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols. (Book VII 179-202 is a life of Chrysippus, including an incompletely preserved catalogue of his works.)

Galen (*c*. AD 175) *On the Doctrines of Hippocrates and Plato*, ed. and trans. P.H. de Lacy, *Galeni de Placitis*

Hippocratis et Platonis (Corpus Medicorum Graecorum V 4 2 1), Berlin: Akademie-Verlag, 3 vols, 1978-83. (Includes long excerpts from Chrysippus' writings on psychology.)

Gould, J.B. (1970) *The Philosophy of Chrysippus*, Albany, NY: State University of New York Press.(Lucid survey.)

Church, Alonzo (1903-95)

Alonzo Church was one of the twentieth century's leading logicians. His work covers an extensive range of topics in logic and in other areas of mathematics. His most influential work relates to three areas: the general properties of functions, as presented in his 'calculus of lambda conversion'; the theory of computability and the decision problem, to which he made fundamental contributions, known as Church's thesis and Church's theorem; and intensional logic, developing Frege's theory of sense and denotation. In the last four decades of his life Church continued working mostly in this last area.

1 Life

Alonzo Church was born in Washington, DC. After receiving his Ph.D. from Princeton in 1927 he spent two years as a National Research Fellow at Harvard, Göttingen and Amsterdam. He then held positions in mathematics and philosophy at Princeton (1929-67) and UCLA (1967-90).

Church's work covers an extensive range of topics in mathematics and logic, including the Lorentz transformation, differentials, recursive arithmetic, postulate theory, the law of the excluded middle, non-standard interpretations of propositional logic, higher-order logic, a theory of weak implication and notes on the history of logic. His most influential work, however, relates to three areas: (1) the general properties of functions, in his 'calculus of lambda conversion' (see Lambda calculus), first presented in 1932 and developed in 1941; (2) effective calculability - his influential proposal that the effectively calculable functions can be identified with the recursive functions (see Church's thesis) and his negative solution to the decision problem ('*Entscheidungsproblem*') for predicate logic (see Church's theorem and the decision problem) were first published in 1936; and (3) intensional logic and a development of Frege's theory of sense and denotation (see Church 1951; Frege, G. §3). In the last four decades of his life Church continued working mostly in this last area. Particularly important is his series of papers on its 'revised' formulations (1973, 1974, 1993).

While at Princeton Church became one of the principal founders of the *Journal of Symbolic Logic*, the first volume of which appeared in 1936. He was an editor of the journal for its first forty-four volumes.

2 A theory of functions

The notion of mathematical function dominated Church's early work. He concentrated only on the general properties of functions 'independently of their appearance in any particular mathematical (or other) domain'. In his explanation of the notion of function, Church diverged from the standard view in mathematics and logic. Of importance is his emphasis on the intensional aspect of functions according to which they are not identified with a set of ordered tuples as is usual in mathematics, but are regarded as *operations* - items which are applied to tuples of objects (their arguments) to yield other single objects as values. The general properties of functions were studied in his λ-calculus, which presented a language permitting unambiguous denotation of functions. The principal expression of this language is an abstraction operator which is used to construct an expression for a function from an expression for an arbitrary value. For example, '$\lambda x.x^2$' denotes the function which takes any number to its square. The rules of transformation (conversion) of the calculus provide a guide for derivations among expressions of the above type.

The λ-calculus has had a significant effect on the development of logic. It was shown that the original formulation of the calculus is equivalent to the combinatory logic of Schönfinkel and, for this reason, Church is considered to be one of the founders of combinatory logic. The λ-calculus also played an important role in the development of intensional logic, particularly in that part of it which takes intensions as functions of a certain type.

A study of general properties of functions led Church to two seminal discoveries concerning the problem of whether it is possible to make all of mathematics algorithmic. The first of these was his conjecture concerning a precise demarcation of the class of functions characterized intuitively as algorithmically or effectively computable. This conjecture, which has come to be known as 'Church's thesis', is that this intuitively characterized class of functions is identical with the precisely characterized class of functions known as general recursive (see Computability theory §2). The second is Church's proof that the so-called 'decision problem' is unsolvable. The general decision problem for mathematics asks whether for every class of mathematical problems there is an

algorithm which decides the answer. More particularly, the decision problem for predicate logic asks whether there is an algorithm for deciding logical truth. Church established a negative answer to both questions (see Church's theorem and the decision problem).

3 The logic of sense and denotation

Church's work on functions is partly responsible for his interest in developing a theory of intensions. His λ-calculus utilizes a notion of 'function-in-intension' (as opposed to 'function-in-extension') and he notes that two functions-in-intension *f* and *g* might be different - might, that is, differ in intension or 'meaning' - even if they share the same domain of arguments and for each *n*-tuple of objects $a_1, \ldots, a_n$ of that domain, $f(a_1, \ldots, a_n) = g(a_1, \ldots, a_n)$. In his earlier work, the notion of difference in meaning was left unexplained. The last four decades, however, he devoted to developing a theory of this difference in meaning. The result was his 'logic of sense and denotation', based on Frege's theory of proper names (see Proper names §1; Frege, G. §§2-3). Church follows Frege in assigning two sorts of semantic values to proper names. A proper name is said to *express* its 'sense' and to *name* its 'denotation' (or 'reference'). Frege and Church are of the view that two names having the same denotation might none the less differ in sense.

Another characteristic feature of the Frege-Church theory is that declarative sentences are treated as names. They express their senses (propositions), and denote (or name) their truth-values.

Like Frege, Church also distinguished between two uses of proper names: the ordinary and the oblique. The principal criterion for obliqueness is the failure of substitutivity for equality ($=$), and for material equivalence ($\equiv$). Examples include contexts with modalities (possibility or necessity) and sentences expressing belief or knowledge ('propositional attitudes'). Church states that in oblique contexts a proper name does not have its 'usual' denotation but rather denotes what in ordinary contexts is its sense.

Church also accepted the following Fregean views concerning the role played by names in determining the sense and denotation of their larger contexts: if a name is replaced by another having the same sense, the sense of the whole context remains unchanged; and if a name is replaced by another having the same denotation, the denotation of the whole context remains unchanged.

In spite of these similarities between Church and Frege, there are also significant differences. For example, to avoid antinomies found in Frege's theory by Russell, Church built his logic of sense and denotation on a theory of types (see Theory of types). Also, unlike Frege's informal treatment, Church's was formal. Using this method, he formulated exact criteria of identity for 'concepts' or senses and identified three alternative forms for the theory, depending upon the particular criterion of identity adopted.

One of these (known as 'Alternative (2)') is designed to treat oblique contexts with modalities but is ill-suited for contexts expressing knowledge and belief, for which purpose the so-called 'Alternative (0)' was developed.

Powerful arguments against any Fregean theory of proper names were presented by Kripke and others. Nevertheless, many aspects of his theory can still be salvaged and it retains a strong appeal as a theory of propositional attitudes. Church offered compelling arguments that attitudes such as belief are relations between agents and propositions rather than between agents and sentences. Further, Alternative (0) offers a sufficiently fine-grained notion of proposition, by use of which Church was able to resolve a number of philosophical and logical puzzles relating to propositional attitudes, for example, those relating to the failure of substitutivity of coreferential expressions in propositional attitudes (see Sense and reference §§4-5), the paradox of analysis and Mates' puzzle concerning iterated epistemic operators. Church's analysis of intensions for expressions other than proper names is also suggestive - particularly those parts of it that pertain to his Fregean theory of definite descriptions (including his criticisms of Russell's theory of such descriptions).

See also: Combinatory logic; Descriptions; Intensional logics; Logical and mathematical terms, glossary of

PETER DOLNÍK

List of works

Church, A. (1999, forthcoming) *Collected Papers*, ed. T. Burge, H.B. Enderton and M. Zeleny, Cambridge, MA: MIT Press.(A new two-volume collection of Church's work.)

Church, A. (1932, 1933) ‘A Set of Postulates for the Foundation of Logic’, papers 1 and 2, *Annals of Mathematics* 33: 346-66, 34: 839-64.(Church’s first published formulation of the calculus of lambda-conversion.)

Church, A. (1936a) ‘An Unsolvable Problem of Elementary Number Theory’, *American Journal of Mathematics* 58: 345-63.(A seminal paper of mathematical logic; the first publication of Church’s thesis and Church’s theorem.)

Church, A. (1936b) ‘A Note on the *Entscheidungsproblem*’, and correction, *Journal of Symbolic Logic* 1: 40-1, 101-2.(Church’s theorem applied to the classical predicate calculus to show that it is undecidable.)

Church, A. (1940) ‘A Formulation of the Simple Theory of Types’, *Journal of Symbolic Logic* 5: 56-68. (Includes a version of the calculus of lambda-conversion based on the simple theory of types (which played a major role in his logic of sense and denotation).)

Church, A. (1941) *The Calculi of Lambda-Conversion*, Princeton, NJ: Princeton University Press.(Classic source for the calculus of lambda-conversion.)

Church, A. (1943) ‘Carnap’s *Introduction to Semantics*’, *The Philosophical Review* 52: 298-304. (A review in which Church suggests and ‘proves’ the thesis that the denotation of a declarative sentence is a truth-value.)

Church, A. (1950) ‘On Carnap’s Analysis of Statements of Assertion and Belief’, *Analysis* 10: 97-9.(A classic paper in logical semantics and the philosophy of language; develops Langford’s test of translation and shows that verbs such as ‘believes’, ‘knows’, ‘says’ cannot be treated as expressing relations between agents and sentences but rather as expressing relations between agents and propositions.)

Church, A. (1951) ‘A Formulation of the Logic of Sense and Denotation’, in *Structure, Method and Meaning: Essays in Honor of Henry M. Sheffer*, New York: Liberal Arts Press, 3-24.(Classic source but requires considerable training in formal logic.)

Church, A. (1954) ‘Intensional Isomorphism and Identity of Belief’, *Philosophical Studies* 5: 65-73. (Informal presentation of Church’s notion of synonymous isomorphism; forms a foundation for his Alternative (0).)

Church, A. (1956) *Introduction to Mathematical Logic*, vol. 1, Princeton, NJ: Princeton University Press.(An excellent introduction to mathematical logic. The introduction provides a clear, extensive and informal exposition of Church’s views on the theory of sense and denotation; helpful for those interested in Church’s logistic treatment of intensions.)

Church, A. (1960) ‘Logic and Analysis’, in *Atti del XII Congresso Internazionale di Filosofia*, Florence, 77-81.(A relatively ‘unknown’ paper that presents Church’s views on the relation of logic to philosophy and of the place of logic in analysis of epistemic matters.)

Church, A. (1973, 1974) ‘Outline of a Revised Formulation of the Logic of Sense and Denotation’, parts 1 and 2, *Noûs* 7: 24-33, 8: 135-56.(Incorporates the notion of possible world.)

Church, A. (1974) ‘Russellian Simple Type Theory’, *Proceedings and Addresses of the American Philosophical Association* 47: 21-33. (A re-formulation of Russell’s mathematical logic in terms of a simple theory of types. Church argues that if the re-formulation is to be presented as intensional logic then Frege’s distinction between sense and denotation will have to be introduced.)

Church, A. (1976) ‘Comparison of Russell’s Resolution of the Semantical Antinomies with that of Tarski’, *Journal of Symbolic Logic* 41: 747-60.(Two resolutions of the semantic antinomies are compared: Russell’s based on a ramified theory of types and Tarski’s based on the object-language/metalanguage distinction. Church shows that the former resolution is a special case of the latter.)

Church, A. (1993) ‘A Revised Formulation of the Logic of Sense and Denotation. Alternative (1)’, *Noûs* 27: 141-57. (Presents a system of logical axioms characterizing propositions that might be taken as objects of assertion and belief. The system is based on a simple theory of types and it is shown that it is set-theoretically consistent. Requires a considerable training in formal logic.)

References and further reading

Anderson, C.A. (1980) ‘Some New Axioms for the Logic of Sense and Denotation. Alternative (0)’, *Noûs* 14: 217-34.(A modification of Church’s Alternative (0); informal explanation of some of Church’s important ideas and axioms.)

Anderson, C.A. (1984) ‘General Intensional Logic’, in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 2, 355-85.(Development of intuitive motivations for Church’s logic of sense and denotation as well as an introduction to its formalization. Also provides a useful comparison with Montague’s intensional logic.)

Bealer, G. (1994) ‘Property Theory: The Type-Free Approach v. The Church Approach’, *Journal of Philosophical Logic* 23: 139-71.(A comparison of Church’s view of properties with that of Bealer.)

Church's theorem and the decision problem

Church's theorem, published in 1936, states that the set of valid formulas of first-order logic is not effectively decidable: there is no method or algorithm for deciding which formulas of first-order logic are valid. Church's paper exhibited an undecidable combinatorial problem $\mathcal{P}$ and showed that $\mathcal{P}$ was representable in first-order logic. If first-order logic were decidable, $\mathcal{P}$ would also be decidable. Since $\mathcal{P}$ is undecidable, first-order logic must also be undecidable.

*Church's theorem is a negative solution to the decision problem (*Entscheidungsproblem*), the problem of finding a method for deciding whether a given formula of first-order logic is valid, or satisfiable, or neither. The great contribution of Church (and, independently, Turing) was not merely to prove that there is no method but also to propose a mathematical definition of the notion of 'effectively solvable problem', that is, a problem solvable by means of a method or algorithm.*

The period 1929-37 was very eventful for logic. It began with Gödel's proof that first-order logic is semantically complete, that is, that the set of provable formulas coincides with the set of logically valid formulas. It included the incompleteness theorems of Gödel (see Gödel's theorems §§3-4), Church's theorem showing the undecidability of first-order logic, and the statement by Church and Alan Turing of Church's thesis equating the intuitive notion of effectiveness and the mathematical notion of recursiveness. The paper by Turing, which exhibited a universal Turing machine and introduced the notion of a computer program, was the culmination of the period.

Church's theorem, published in 1936 (see Davis 1965), states that the set of valid formulas of first-order logic is not effectively decidable: there is no method or algorithm for deciding which formulas of first-order logic are valid. Turing's slightly later paper proved the same result in a slightly different way. Both papers followed the same pattern. They exhibited an undecidable combinatorial problem $\mathcal{P}$ and showed that $\mathcal{P}$ was representable in first-order logic. If first-order logic were decidable, $\mathcal{P}$ would also be decidable. Since $\mathcal{P}$ is undecidable, first-order logic must also be undecidable.

Hilbert and Ackermann had formulated the *Entscheidungsproblem* (the decision problem) as the problem of finding a method for deciding whether a given formula of first-order logic is valid, or satisfiable, or neither; that is, whether all, some, or none of its possible interpretations are true. Church's theorem is a negative solution to the *Entscheidungsproblem*.

Hilbert and Ackermann had not proposed a mathematical definition of what constitutes a method, and the great contribution of Church and Turing was not merely to prove that there is no method but also to propose a mathematical definition of the notion of 'effectively solvable problem', that is, a problem solvable by means of a method or algorithm.

The technical result proved by Church only showed that the set of *provable* formulas of first-order logic is not λ-definable, that is, not definable in the λ-calculus, a logical calculus invented by Church (see Lambda calculus). To conclude that the set of *valid* formulas in first-order logic is not 'effectively' decidable, we need two additional facts. We need to equate validity in first-order logic with provability in a particular formal system, a result proved already by Gödel, but about which Church had some qualms due to the non-constructive nature of the proof. We also need to equate the mathematical notion of λ-definability (or the provably equivalent notions of recursiveness and Turing computability) with the *intuitive* notion of effective computability. This second equality is Church's thesis (see Church's thesis). Assuming Church's thesis, Church's theorem constitutes a negative solution to the problem of giving a uniform procedure for deciding which formulas of first-order logic are valid.

The particular version of first-order logic which Church used to prove his theorem included, as part of the language, the equality symbol, a constant and several function symbols. However, such a rich vocabulary is not actually necessary. Even first-order logic with a single dyadic predicate symbol is undecidable. (This is in sharp contrast to the decidability of the special case where the language consists solely of monadic predicates, a result first proved by Löwenheim (1915).)

Also, despite Church's qualms about the non-constructiveness of Gödel's completeness proof, Church's theorem is

rather sturdy. It is known that, for Gödel's completeness result to hold, one must allow formulas with non-recursive models to be considered to be satisfiable. However, even if one restricts the possible models - for example, if one confines oneself to finite models or to constructive (recursive) models - the set of valid formulas remains non-recursive, and Church's theorem remains true, although first-order logic is no longer complete relative to these restricted classes of models.

Hilbert's *Entscheidungsproblem* as formulated by Hilbert and Ackermann inquired about the decision problem for logic as such. This would also include second-order logic, whose undecidability follows of course from that of the (weaker) first-order logic. But here a slightly more complex situation prevails. First-order logic, while not decidable, is none the less semi-decidable in that we do have a procedure to enumerate the set of all valid formulas. What prevents first-order logic from being decidable is that we do not have a procedure to enumerate the complementary set of all formulas which are *not* valid, that is, of all formulas whose negations are satisfiable (have at least one true interpretation). However, second-order logic is not even semi-decidable, a fact which is already implicit in Gödel's incompleteness result for arithmetic, and which pre-dated Church's theorem by about five years (see Gödel's theorems §1). Thus, at the time that Church proved his theorem, only the question for first-order logic was really open.

Note that a positive solution to the *Entscheidungsproblem* would not have required a mathematical definition of effectiveness. It would have sufficed to present a method for deciding which formulas of first-order logic were valid and to rely on our mathematical intuition to see that the method qualified as a genuine algorithm. However, a negative solution requires us to delimit the notion of effective computability to show that validity in first-order logic lies outside it, and such a delimitation necessarily requires that the notion of effectiveness be given a mathematical definition, or at least be mathematically circumscribed.

Church himself did not provide in his paper an independent definition of the intuitive notion of effectiveness. Rather he proposed that his technical notion, λ-definability, be taken to be a definition of the intuitive notion of effectiveness. A more substantial analysis was provided by Alan Turing in his paper (see Davis 1965). Turing analysed the possible mental algorithms that a human being could carry out and claimed that his machines, Turing machines, could also carry them out. A crucial element in Turing's argument was the claim that a human computer (sic) would only be able to carry out algorithms which utilized a finite number of distinct symbols, and that while he was carrying out an algorithm, his mind could only occupy a finite number of distinct states. Turing's argument for the first claim is that if the process of calculation used infinitely many basic symbols, then some of these symbols would be very close to others and confusion could not be avoided. He remarks that the same also holds of states of mind, but offers no argument. (See Turing machines.) Gödel, in fact, describes Turing's remarks in this context as a 'philosophical error'. If we do accept Turing's justification that effectiveness equals Turing computability, then since λ-definability and Turing computability are provably equivalent, Turing's analysis also provides substantial ammunition for Church's own claim.

See also: Church's thesis; Logical and mathematical terms, glossary of

ROHIT PARIKH

References and further reading

Davis, M. (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Hewlett, NY: Raven Press.(Includes a corrected version of Church's original paper, and Turing's article.)

Gödel, K. (1990) 'Some Remarks on the Undecidability Results', in *Collected Works*, ed. S. Feferman *et al*., New York and Oxford: Oxford University Press, vol. 2, 305-6.(An excellent account of Gödel's own views.)

Herken, R. (ed.) (1988) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press. (An excellent collection of papers by contemporary researchers; see especially R. Gandy, 'The Confluence of Ideas in 1936', 55-111.)

Löwenheim, L. (1915) 'Über Möglichkeiten im Relativkalkül', *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, 'On Possibilities in the Calculus of Relatives', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 232-51.(Decidability of the special case of first-order logic with a language consisting solely of monadic predicates.)

Church's thesis

An algorithm or mechanical procedure A is said to 'compute' a function f if, for any n in the domain of f, when given n as input, A eventually produces fn as output. A function is 'computable' if there is an algorithm that computes it. A set S is 'decidable' if there is an algorithm that decides membership in S: if, given any appropriate n as input, the algorithm would output 'yes' if $n \in S$, and 'no' if $n \notin S$. The notions of 'algorithm', 'computable' and 'decidable' are informal (or pre-formal) in that they have meaning independently of, and prior to, attempts at rigorous formulation.

Church's thesis, first proposed by Alonzo Church in a paper published in 1936, is the assertion that a function is computable if and only if it is recursive: 'We now define the notion…of an effectively calculable function…by identifying it with the notion of a recursive function….' Independently, Alan Turing argued that a function is computable if and only if there is a Turing machine that computes it; and he showed that a function is Turing-computable if and only if it is recursive.

Church's thesis is widely accepted today. Since an algorithm can be 'read off' a recursive derivation, every recursive function is computable. Three types of 'evidence' have been cited for the converse. First, every algorithm that has been examined has been shown to compute a recursive function. Second, Turing, Church and others provided analyses of the moves available to a person following a mechanical procedure, arguing that everything can be simulated by a Turing machine, a recursive derivation, and so on. The third consideration is 'confluence'. Several different characterizations, developed more or less independently, have been shown to be coextensive, suggesting that all of them are on target. The list includes recursiveness, Turing computability, Herbrand-Gödel derivability, λ-definability and Markov algorithm computability.

1 Computability and decidability

The algorithms or mechanical procedures envisaged here must determine what is to be done at each step, involving no creativity or intuition on the part of the agent (other than an ability to recognize symbols). The paradigm is a recipe for routine calculation, or a computer program.

An algorithm (or mechanical procedure) *A* 'computes' a function *f* if, for any *n* in the domain of *f*, when given *n* as input, *A* produces *fn* as output. A function is 'computable' if there is an algorithm that computes it. (See Computability theory §2.) There is no fixed limit on the amount of materials or memory required for a computation, nor on the number of steps an algorithm might take. If a function were computed only by an algorithm whose execution required more memory than there are particles in the solar system, it would not be capable of computation in any realistic sense, yet it would still be computable. What is required is that the algorithm have only finitely many steps. A set *S* is 'decidable' if there is an algorithm that 'decides' membership in *S*: if, given any appropriate *n* as input, the algorithm outputs 'yes' if $n \in S$, and 'no' if $n \notin S$.

Strictly speaking, computability should apply only to functions on things like marks on paper and magnetic media, since these are the items that can be manipulated by computers and algorithms. It is common to take computability to apply to functions on 'strings', which are the abstract forms of sequences of characters. For example, the string '$abc23$' is the form of all actual and possible concrete inscriptions of these three letters and two numerals. It is also common to think of computability as applying to functions on natural numbers, via canonical notation.

The root meaning of 'computability' would be something like 'capable of computation', and the root meaning of 'decidability', something like 'capable of decision'. Thus, these are modal notions. Computability is also an 'extensional' notion: it applies to functions themselves, not to descriptions or presentations of functions. Moreover, the quantifier 'there is an algorithm' does not involve (*de re*) knowledge of an algorithm (see De re/de dicto §1). For example, let $fn = 1$ if there is a string of at least *n* consecutive 5's in the decimal expansion of π, and $fn = 0$ otherwise. No one knows of any algorithm *A* such that *A* computes *f*. Nevertheless, the function *f* is computable. If there are indefinitely long strings of consecutive 5's in the expansion of π, then *f* is computed by an algorithm that produces '1' no matter the input. Otherwise, let *N* be the longest string of consecutive 5's in the expansion. Then *f* is computed by an algorithm that outputs '1' for all inputs less than or equal to *N*, and outputs '0' for all other inputs. In all cases, then, *f* is computable.

The notions of recursiveness (see Computability theory §1) and λ-definability (see Lambda calculus §4) arose as part of various foundational programmes earlier in the twentieth century. Church and his students became interested in the extension of the λ-definable functions. They set themselves the task of investigating various (computable) functions, to see if they could be shown to be λ-definable. Shortly after Kleene proved that the predecessor function is λ-definable, Church proposed his thesis: that a function is computable if and only if it is recursive. He published it after Kleene established the connection with recursiveness.

Although algorithms have been constructed since antiquity, it is curious that no one attempted a precise characterization of computability *per se* before the mid-1930s, at which time a number of more or less independent characterizations appeared almost simultaneously. One motivation was the desire, or need, to establish *negative* results about computability and decidability. Notice that the straightforward way to show that a given function is computable is to give an algorithm. This does not require a rigorous definition of computability, since most theorists can recognize an algorithm when they see one. To establish that a given function is *not* computable, on the other hand, one needs to show that *no* algorithm computes the function. This requires the formulation of a property shared by all algorithms or all computable functions. The straightforward route to such a property is a careful characterization of computability. At the time, the construction of algorithms for various purposes was a central item on the agenda of mathematical logic (see Hilbert's programme and formalism). Virtually all of the original characterizations of computability include theorems that certain functions are not computable or that certain sets are not decidable.

A related motivation was the desire to understand the extent of Gödel's incompleteness theorem. The original proof only applied to a particular system, but it is clear that the proof could be generalized to any common axiomatization of arithmetic. This, and the work on Hilbert's programme, led to the problem of characterizing just what a formal deductive system is. One requirement is that the syntax should be 'effective', in the sense that the collections of well-formed formulas, axioms and rules of inference should all be decidable. The natural generalization of the incompleteness theorem is that there is no sufficiently rich, effective and complete deductive system for arithmetic. As above, to establish this generalization, one needs a grasp of the limits of algorithms, and this came with the characterization of computability (see Gödel's theorems; Formal languages and systems).

The modern notion of a function as an arbitrary correspondence between sets has emerged only towards the end of the twentieth century. In the past, the very idea of a 'non-computable' function might have been questioned, as it is today by constructivists and intuitionists. It might be added that from the constructivist perspective, Church's thesis has a sharp formulation. For a constructivist, a formula of the form $\forall x \exists ! y \Phi(x, y)$ states that, given any *x*, *one can find* a unique *y* such that $\Phi(x, y)$ holds. This amounts to an assertion that a certain function is computable. Thus, Church's thesis is the scheme that if $\forall x \exists ! y \Phi(x, y)$ then there is a *recursive* function *f* such that $\forall x \Phi(x, fx)$. In other words, for a constructivist, Church's thesis is a claim that *all* number-theoretic functions are recursive. Among intuitionists and constructivists, Church's thesis is controversial (see Intuitionism; Constructivism in mathematics; Beeson 1985; McCarty 1987).

As noted, however, there is a consensus among (non-constructivist) mathematicians and logicians that Church's thesis is correct. Within recursive function theory, a technique has been developed, sometimes called 'argument by Church's thesis', wherein one concludes that a function is recursive, or that a set is decidable, just because there is an algorithm for it. For example, for each natural number *n*, let $g(n)$ be the sum of the divisors of *n*. Since it is routine to give an algorithm to factor a number and add its divisors, one can conclude, by Church's thesis, that *g* is recursive. One does not have to deal with the details of recursive derivations, Turing machines and so on. By avoiding tedious detail, the perspective is broadened, leading to greater insight (see Rogers 1967).

2 What is Church's thesis?

Despite the consensus on the correctness of Church's thesis, there is not much agreement over what the correctness amounts to. What does Church's thesis mean? Does it have a determinate truth-value? Can it be known and, if so, how?

Computability is a pragmatic, pre-theoretic property of functions. It relates somehow to human abilities, suitably idealized. As such, Church's thesis is an important case study for the philosophical problem of accounting for the relationship between mathematics and material or human reality. Some authors have gone so far as to call

computability a 'psychological' concept, since it relates to competence in grasping and executing procedures. One of the founders of computability, Emil Post, wrote that

> for full generality a complete analysis would have to be made of all the possible ways in which the human mind could set up finite processes…we have to do with a certain activity of the human mind as situated in the universe. As activity, this logico-mathematical process has certain temporal properties; as situated in the universe it has certain spatial properties.
>
> (Davis 1965: 408-19)

This passage also notes the aforementioned modal nature of computability. Notice, however, that recursiveness, Turing computability, λ-definability and so on are at least *prima facie* non-modal. They are defined in terms of quantifiers ranging over abstract entities, such as natural numbers, strings on a finite alphabet and so on. Thus, Church's thesis identifies the extension of an idealized, pragmatic, modal property of functions with the extension of a formal, rigorously defined, *prima facie* non-modal arithmetic property. On the surface, then, Church's thesis proposes an exchange of modality for ontology. How are such proposals to be evaluated?

A closely related matter concerns vagueness. Typically, pragmatic, modal notions do not have sharp boundaries. There are, for example, borderline cases of 'climbable mountains', 'liftable weights' and 'teachable students'. Clearly, there are also borderline cases of functions 'capable of computation', if we do not idealize. Slogans such as 'ignoring limits on memory and materials' sharpen the notion, but it is still not clear that idealized computability, as defined above, is an unequivocal property with *completely* precise boundaries. Recursiveness and Turing computability, on the other hand, do appear to have precise boundaries. *Prima facie*, then, Church's thesis identifies a vague notion with a precise one.

Clearly, a vague property cannot *exactly* coincide with a precise one. So, if computability is in fact vague, Church's thesis does not literally have a determinate truth-value or else it is false. Church's thesis might then be thought of as a proposal that recursiveness be *substituted* for computability, for certain purposes. It would be similar to a pronouncement such as 'For the purposes of this form, a resident is defined to be…'.

The value of 'definitions' or 'precisifications' of vague terms depends on the purposes at hand. Given the extensive idealizations, Church's thesis seems best suited for establishing *negative* results about computability. If one shows, for example, that a function is not recursive, that is conclusive reason to hold that it is not 'capable of computation': one should give up trying to compute it. Suppose, on the other hand, that a function has been shown to be recursive. By itself, this gives no information on the feasibility of an algorithm, and thus it does not establish that the function can be calculated, in any realistic sense. Moreover, it is of little value to anyone interested in actually computing the function. For this purpose, there are other, less idealized models, such as 'finite state' computability, 'push down' computability, primitive recursiveness and computability in polynomial time or space (see Complexity, computational).

A different interpretation of Church's thesis is that even though computability is a pre-theoretic notion, it is nevertheless a precise, mathematical property of functions. One might hold, for example, that there are determinate mathematical structures underlying idealized human abilities to calculate, write programs, or even think. Understood this way, Church's thesis states that this property is coextensive with another precise mathematical notion, recursiveness. Thus, Church's thesis has a truth-value. The 'arguments from confluence' support this interpretation. The fact that a number of *prima facie* different attempts to characterize computability lead to the same extension suggests that recursiveness, Turing computability and so on delimit the boundaries of a 'natural kind' underlying computation.

3 Epistemic status

The Goldbach conjecture (that every even number greater than 4 is the sum of two primes) and the 'twin prime' conjecture (that there are infinitely many numbers *n* such that *n* and $n+2$ are both prime) are long-standing open problems in number theory. They can be settled, if at all, by mathematical proof. It is widely held that Church's thesis is not like this. Its evidence is not mathematical. Because of the connection with human abilities, and perhaps because of the supposed vagueness, Church's thesis has been called an 'empirical' or 'quasi-empirical' matter.

As noted above, one type of evidence cited for Church’s thesis is that every algorithm that has been investigated has been shown to compute a recursive function. This is consonant with the quasi-empirical nature of Church’s thesis (on a certain view of scientific hypotheses). The intuitive ability to recognize algorithms is analogous to ‘observation’, and showing that an algorithm computes a recursive function amounts to ‘confirmation’ of Church’s thesis (see Confirmation theory). This type of evidence was prominent in the thought of Church and Kleene.

Some theorists, on the other hand, do regard Church’s thesis as a mathematical matter, capable of whatever epistemic standards are applicable there, and some regard the thesis as proved. Robin Gandy (1988), for example, suggests that Turing’s direct argument that every algorithm can be simulated on a Turing machine does not merely provide evidence for an inductive, scientific hypothesis; rather, it ‘proves a theorem’. Turing provides a careful analysis of a human executing an algorithm, showing how each ‘step’ can be simulated on a Turing machine. There are lemmas that the number of symbols and the number of ‘states of mind’ must be finite. Gandy regards the analysis to be as convincing as typical mathematical work. Apparently, Gödel was also convinced by Turing’s arguments, and not by the ‘quasi-empirical’ evidence (see Davis 1982).

On these views, the pragmatic nature of computability, and its potential vagueness, do not automatically disqualify Church’s thesis from mathematics. This is consistent with a widely held position that there is no sharp boundary between mathematics and empirical science (see Duhem, P.M.M.; Quine, W.V.O.). Mathematics has its roots in everyday life, and cannot be severed from it. Even a cursory look at the history of the notions underlying recursiveness reveals ambivalence, ambiguity and vagueness, especially in the notion of ‘set’ (see Mendelson 1990). Yet no one doubts that sets and recursive functions belong to mathematics. According to the view at hand, computability is no different.

4 Other theses

There are similar ‘theses’ in the history of mathematics and science. Consideration of these may shed some light on the philosophical issues here.

The notion of a grammatical sentence of a natural language is clearly pragmatic, modal and at least *prima facie* vague. This is contrasted with formal, mathematical characterizations of grammaticality, as proposed by linguists. Unlike Church’s thesis, however, there is no consensus on one of these grammars.

Closer to home, the notion of ‘validity’, or ‘logical consequence’, is *prima facie* pragmatic and modal, if not vague. In typical logic textbooks, validity is related to inference abilities, and an argument is said to be valid if it is not possible for the premises to be true and the conclusion false. Yet, when one reads on, validity is formally defined in terms of models, which are mathematical constructions on sets. Thus, we have another exchange of modality and ontology. The identification of the pre-formal validity with model-theoretic consequence may be called ‘Tarski’s thesis’. Notice that there is a consideration of confluence here. Gödel’s completeness theorem shows that for first-order languages, the model-theoretic notion of consequence coincides in extension with a deductive one (see Consequence, conceptions of; Model theory; Proof theory).

Within more traditional mathematics, consider notions such as ‘area’, ‘volume’ and ‘continuity’. While these do not appear pragmatic and modal, they did have clear meanings before their rigorous formulations in analysis and measure theory. Moreover, the notions do have uses in everyday and scientific life and there seem to be borderline cases of the pre-theoretic notions. In modern mathematics, however, these notions all have rigorous, precise definitions. The various identifications with the pre-theoretic notions may be called Euclid’s thesis, Lebesgue’s thesis, Weierstrass’ thesis, Cauchy’s thesis, and so on. The reader may decide the extent to which these theses have determinate truth-values, whether they are mathematical or ‘quasi-empirical’, and whether they have been (or can be) established with characteristic mathematical rigour.

See also: Computability theory; Complexity, computational; Computability and information; Logic machines and diagrams; Logical and mathematical terms, glossary of

STEWART SHAPIRO

References and further reading

Beeson, M.J. (1985) *Foundations of Constructive Mathematics*, New York: Springer.(Includes an account of

Church's thesis from a constructivist perspective.)
Davis, M. (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Hewlett, NY: Raven Press.(Original papers, or English translations, by Gödel, Church, Turing, Rosser, Kleene and Post; includes some items not published elsewhere.)
Davis, M. (1982) 'Why Gödel Didn't Have Church's Thesis', *Information and Control* 54: 3-24.(Careful historical discussion of what convinced some of the major players of Church's thesis.)
Gandy, R. (1988) 'The Confluence of Ideas in 1936', in R. Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, New York: Oxford University Press, 55-111.(Extensive historical study of the development of computability.)
Kleene, S. (1979) 'Origins of Recursive Function Theory', in *Twentieth Annual Symposium on Foundations of Computer Science*, New York: IEEE, 371-82.(Reminiscent account, by one of the founders of recursive function theory; good source on the development of the λ-calculus and on Church's thought on logic and computability.)
Kleene, S. (1987) 'Reflections on Church's Thesis', *Notre Dame Journal of Formal Logic* 28: 490-8. (Philosophical discussion of Church's thesis.)
McCarty, C. (1987) 'Variations on a Thesis: Intuitionism and Computability', *Notre Dame Journal of Formal Logic* 28: 536-80.(An intuitionistic view of Church's thesis.)
Mendelson, E. (1990) 'Second Thoughts about Church's Thesis and Mathematical Proofs', *Journal of Philosophy* 87: 225-33.(Argument that Church's thesis is provable.)
Myhill, J. (1952) 'Some Philosophical Implications of Mathematical Logic', *Review of Metaphysics* 6: 165-98. (Philosophical account of the argument that computability is partly psychological.)
Rogers, H. (1967) *Theory of Recursive Functions and Effective Computability*, New York: McGraw-Hill. (Comprehensive mathematical development, based on 'argument by Church's thesis'.)
Shapiro, S. (1981) 'Understanding Church's Thesis', *Journal of Philosophical Logic* 10: 353-65. (Discussion of the philosophical status of Church's thesis.)
Shapiro, S. (1983) 'Remarks on the Development of Computability', *History and Philosophy of Logic* 4: 203-20. (Historical account.)
Shapiro, S. (1993) 'Understanding Church's Thesis, Again', *Acta Analytica* 11: 59-77.(A further discussion of the philosophical status of Church's thesis, including its modal status, its vagueness and the view that Church's thesis is a mathematical matter.)
Sieg, W. (1994) 'Mechanical Procedures and Mathematical Experience', in A. George (ed.) *Mathematics and Mind*, Oxford: Oxford University Press, 71-117.(Philosophical account of Church's thesis and its relationships with mechanism.)

Chông Yagyong (Tasan) (1762-1836)

Chông Yagyong was a government official and a scholar of the Sirhak (Practical Learning) school in the late Chosôn dynasty of Korea. He is also known by his literary name Tasan. A man of independent mind, Chông was not satisfied with the conventional interpretation of the Confucian classics. He immersed himself in research on the Six Classics and the Four Books, investigating a whole range of writings by scholars from the Han through the Qing dynasties and searching for the true and original intents of the ancient sages uncorrupted by later interpretations. In the course of clarifying ancient terms and concepts, he frequently challenged the orthodox views of the Song neo-Confucianism that had largely dominated the intellectual climate of Chosôn Korea. Although he frequently praised Zhu Xi, he did not hesitate to point out the shortcomings of the neo-Confucian masters.

In his youth, Chông Yagyong came under the influence of the scholarship of Yi Ik (1681-1763). In 1784 he was first introduced to the teachings of Catholicism by his friend Yi Pyôk (1754-86), who provided him with a copy of Matteo Ricci's *The True Meaning of Our Lord in Heaven*, and thereafter he was exposed to Catholicism and Western learning. Having passed the higher civil examination in 1789, he held various government posts, such as local magistrate, Secret Inspector, Third Minister and Royal Secretary. Because of his association with Catholicism his official career was not smooth, but he nevertheless became one of the favorites of King Chôngjo who, recognizing his brilliance, protected him as much as possible. In 1801, during the persecution of Catholics, Chông was forced out of office and had to spend the next eighteen years in exile. His earlier renunciation of Catholicism may have saved him from certain death, and there are conflicting accounts as to whether his public disclaimer was genuine. While in exile, Chông devoted his energy to scholarship, mostly writing commentaries on the classics. In 1818 he was released from exile and allowed to return to his home at Mahyôn, an eastern suburb of present-day Seoul, where he lived until his death. He wrote extensively on a wide range of subjects, but he is better known for his scholarly commentaries on the Six Classics and the Four Books, and his three studies on government and economy.

Not satisfied with the neo-Confucian cosmology based on principle (*i*; in Chinese, *li*) and material force (*ki*; in Chinese, *qi*), Chông developed his own concept of heaven (see Qi; Tian). He sees heaven or the heavenly ruler as a monotheistic spirit who created the universe and presides over all beings. Unlike plants or animals, every person is endowed by heaven with its spiritual quality, enabling them to transcend all things. At the same time, people are equipped with dual qualities: the autonomy to choose good or evil and the natural inclination 'to take pleasure in the good and to be ashamed of evil'. Chông then explains that the reason why 'the superior man does not wait till he sees things, to be cautious' is because he is aware of the presence of the heavenly ruler even in a dark room. Thus men have the obligation to obey and respect heaven, from whence emanates all ethical and moral value. One can see the influence of Catholicism in Chông's concept of heaven, and similarly in his studies of the Six Classics and the Four Books, he built up his own system of philosophy through an eclectic approach.

Chông also valued practical applications of knowledge, and thus declared that 'the teaching of the School of the Principle is to know the Way through self-awareness so that one can exert his utmost toward realizing the right principle'. Of the many works on government and economy that he wrote, three are best known: *Kyôngse yup'yo(Designs for Good Government)*, *Mongmin simsô (Admonitions on Governing the People)* and *Hûmhûm sinsô (Towards a New Jurisprudence)*. Not unlike John Locke, Chông believed that the first governments were formed when people in a community agreed to let a judicious man settle a dispute that arose. Accordingly, he maintained that government exists on behalf of the people. In place of the chaotic land system that prevailed in Korea at the time, he proposed a 'village land' (*yôjôn*) system whereby land would be owned and cultivated jointly by village members and the harvest distributed according to the amount of work contributed by each individual. His overriding goal for the government was to promote and protect the well-being of the people as a whole. A man of practical mind, he also devised cranes that were used in the construction of the fortifications at Suwôn, set up a pontoon bridge over the Han River and experimented with smallpox and measles inoculations. A versatile scholar, he wrote on a wide range of subjects including history, geography, philology and military science. He is regarded as the greatest of all the Sirhak scholars who stood between the traditional and the modern periods in Korea.

See also: Confucian philosophy, Korean; Neo-Confucian philosophy; Qi; Sirhak

YÔNG-HO CH’OE

List of works

Chông Yagyong [Tasan] (1762-1836) *Chông Tasan chônjip (Complete Works of Chông Tasan)*, Seoul: Munhôn p’yônch’an wiwônhoe, 1960-1, 4 vols.(This is a facsimile reproduction in reduced size of the original, published in 1936.)

Chông Yagyong [Tasan] (1817) *Kyôngse yup’yo (Designs for Good Government)*, in *Chông Tasan chônjip*, Seoul: Munhôn p’yônch’an wiwônhoe, 1960-1.(Embodies Chông Yagyong’s proposals for reform in governmental structure and administration as well as the land and tax system.)

Chông Yagyong [Tasan] (1818) *Mongmin simsô (Admonitions on Governing the People)*, in *Chông Tasan chônjip*, Seoul: Munhôn p’yônch’an wiwônhoe, 1960-1.(A handbook for local magistrates on how to manage local administration, along with views on the problems of the common people.)

Chông Yagyong [Tasan] (1819) *Hûmhûm sinsô (Towards a New Jurisprudence)*, in *Chông Tasan chônjip*, Seoul: Munhôn p’yônch’an wiwônhoe, 1960-1.(Study and proposal for the reform of the handling of penal cases for murder.)

References and further reading

Henderson, G. (1957) ‘Chông Tasan: A Study in Korea’s Intellectual History’, *Journal of Asian Studies* 16 (3): 377-86.(A general survey of his career and association with his Catholic friends.)

Kang Man’gil *et al.* (1999) *Tasanhak ûi t’amgu (In Search of the Tasan School)*, Seoul: Minûmsa.(Seven South Korean scholars examine various aspects of Chông’s views on subjects including criminal law, social structure, the tax system, finance, the classics and Catholicism.)

Kûm Changt’ae (1987) *Han’guk sirhak sasang yôngu (Study of the Thought of the Practical Learning in Korea)*, Seoul: Chimmundang.(Part 2 of this book is devoted to the study of Chông’s philosophy such as view on Heaven and the *Yijing*.)

Kwahagwôn Ch’ôrhak Yônguso (1962) *Chông Tasan yôngu (Studies of Chông Tasan)*, Pyôngyang: Kwahagwôn.(North Korean scholars examine Chông’s views on philosophy, society and economy, language, literature, history, education and natural science.)

Setton, M. (1996) *Chông Yagyong: Korea’s Challenge to Orthodox Neo-Confucianism*, Albany, NY: State University of New York Press.(A penetrating study of the worldview behind Chông Yagyong’s reform ideas, with comparative perspectives of China and Japan.)

Yi Urho (1966) *Tasan kyônghak sasang yôngu (Study of Chông Yagyong’s Thought on the Classics)*, Seoul: Uryu munhwasa.(A pioneering study of Chông focusing on his views on the Confucian classics.)

Yi Urho (1975) *Tasanhak ûi ihae (Understanding Chông Yagyong)*, Seoul: Hyônamsa.(Collection of articles from other sources, dealing mostly with Chông’s study of the Confucian classics.)

Yi Urho *et al.* (1989) *Chông Tasan ûi kyônghak (Classical Studies of Chông Tasan)*, Seoul: Minûmsa.(Four Korean scholars examine Chông’s study of the *Daxue, Zhongyong, Mengzi* and *Analects*.)

Cicero, Marcus Tullius (106-43 BC)

Cicero, pre-eminent Roman statesman and orator of the first century BC and a prolific writer, composed the first substantial body of philosophical work in Latin. Rising from small-town obscurity to the pinnacle of Rome's staunchly conservative aristocracy, he devoted most of his life to public affairs. But he was deeply interested in philosophy throughout his life, and during two intervals of forced withdrawal from politics wrote two series of dialogues, first elaborating his political ideals and later examining central issues in epistemology, ethics and theology. Designed to establish philosophical study as an integral part of Roman culture, these works are heavily indebted to Greek philosophy, and some of the later dialogues are largely summaries of Hellenistic debates. But Cicero reworked his sources substantially, and his methodical expositions are thoughtful, judicious and, on questions of politics and morals, often creative. An adherent of the sceptical New Academy, he was opposed to dogmatism but ready to accept the most cogent arguments on topics important to him. His vigorously argued and eloquent critical discussions of perennial problems greatly enriched the intellectual and moral heritage of Rome and shaped Western traditions of liberal education, republican government and rationalism in religion and ethics. These works also afford invaluable insight into the course of philosophy during the three centuries after Aristotle.

1 Life and writings

Marcus Tullius Cicero, elder son of a locally influential family in the town of Arpinum, moved to Rome in his youth to pursue a career in law and government. There he studied with several Greek philosophers, including the Academic Philo of Larissa, and after a brilliant legal debut he spent two years in Greece studying philosophy and rhetoric with Antiochus and the Stoic Posidonius. Upon his return he won election to a major office that brought lifelong membership of the Senate (Rome's supreme governing body) and soon established himself as the foremost advocate of the age. Elected consul (Rome's chief executive office) in 63 BC, he suppressed an insurrection and was hailed his country's saviour. But opponents contrived his exile in 58 BC, and when he returned the following year, he found his influence severely diminished. Turning to writing, he formulated and defended his political ideals in three pioneering dialogues (see §2). When Julius Caesar precipitated civil war in 49 BC, Cicero sided reluctantly with the opposition as the lesser threat to Roman institutions. Caesar's swift victory brought dictatorship, and Cicero, although granted clemency, was excluded from politics. Returning to writing, he championed free political discussion in a series of rhetorical works, then composed in twenty months a dozen works (nine survive whole or in large part) discussing central problems in Hellenistic philosophy (see §§3-4) (see Hellenistic philosophy). The political turmoil that followed Caesar's assassination in March 44 BC slowed, then halted this astonishing pace as he rallied resistance to Mark Antony's despotic designs. His campaign might well have succeeded had Antony not colluded with Caesar's adoptive heir, the future Augustus: Cicero was assassinated and his head impaled in the Forum where he had spoken so often and so eloquently.

Cicero's extant works, although only part of his enormous output, comprise over fifty speeches, nearly a thousand letters to friends and associates, several works on rhetorical theory and practice, and twelve on philosophical topics. This vast corpus, besides displaying great intellectual range and stylistic virtuosity, embodies Cicero's conviction that philosophy and rhetoric are interdependent and both essential for the improvement of human life and society. His oratory bears the stamp of his theoretical studies, and his treatises and dialogues are richly oratorical. The philosophical works in particular unite the rhetorical techniques and ample style of Roman oratory with the analytical methods and conceptual apparatus of Greek philosophy in a unique fusion of eloquence and insight. All but one of these works are fictional dialogues. Some portray Cicero or eminent Romans of the previous century discoursing at length among friends; others, employing a format that reflects Roman political and legal practice but also the critical spirit Cicero admired in Plato and his sceptical heirs in the New Academy (see §3), present paired speeches for and against Epicurean and Stoic theories. Composed for audiences unused to abstract theory and systematic analysis, the discussions lapse at times into earnest declamation, and the close questioning found in Plato's dialogues is rare. However, they are methodically organized and often incisive, and by presenting opposing views and arguments in clear and engaging terms, they dramatize the significance of fundamental problems and encourage critical reflection.

2 Classical republicanism

Cicero’s first philosophical works are three long dialogues that analyse and evaluate the political institutions and practices of contemporary Rome in the light of Greek theory. Although largely conservative, they provided the first political theory in Latin and remain the most systematic ancient account of Roman government; while others described events, only Cicero advanced a structural analysis. Written when Rome’s republican traditions were collapsing under unprecedented concentrations of economic and military power, these dialogues champion political liberty, rational debate and rule by law. Articulating the principles behind his lifelong goal of harmonizing Rome’s competing interests in a just and stable ‘concord of the orders’, they propound a comprehensive vision of civil society directed by an elected elite schooled in rhetoric and philosophy, devoted to constitutional government and able to shape public opinion through effective oratory.

The first of these works, *On the Orator*, explores the role of rhetoric and philosophy in public life. Oratory had long been a potent tool in Roman politics, and Cicero aims to reinforce its prestige and legitimize its influence by showing that its success requires wide learning and sound reasoning. Much of the discussion focuses on education, as he weighs the merits of the traditional Roman emphasis on history, poetry and practical experience against the Greek disciplines of formal rhetoric and philosophy. His model orator, who clearly reflects Cicero’s own proficiencies, unites thorough knowledge of history and law with complete command of logical method, philosophical theory and rhetorical techniques in a Romanized version of Plato’s philosopher-rulers. Both expect philosophical education to produce the best statesmen; but whereas Plato’s ideal hinges on mathematical training and transcendental metaphysics, Cicero proposes a thoroughly pragmatic programme of instruction designed to foster eloquence and informed civic debate.

Cicero delineates the institutional framework behind his conception of leadership in *On the Republic*, which was almost entirely lost until most of the first third was recovered in 1820. Challenging the utopian bent of Greek political theory exemplified by Plato’s *Republic*, Cicero argues that the best constitution, far from being unattainable, was largely realized in Rome, where a unique blend of monarchy, aristocracy and democracy formed a ‘mixed constitution’ that provided a system of government ostensibly stable and just. This account, which recasts Rome’s narrow oligarchy as a paradigm of aristocratic paternalism, rests on an incisive and apparently original analysis of political legitimacy. Defining a republic (*res publica*) as ‘a people’s affair’ (*res populi*) and a people as a community ‘united by consensus about right and by mutual interest’, Cicero criticizes all other constitutions for contravening the people’s rights and interests, then argues that no political system is legitimate unless it distributes legal rights equally to all, but electoral, legislative and judicial authority proportionally according to merit and wealth. Other extant sections of the dialogue classify types of constitution in classical Greek fashion and survey the development of Roman institutions; lost or poorly preserved sections summarized Hellenistic debates about the nature and rewards of justice, discussed Roman education and measured past Roman statesman against Cicero’s ideal. Ending the work is the famous ‘dream of Scipio’, which interweaves astronomy and eschatology to sketch a theodicy that rewrites Plato’s myth of Er in the light of Stoic cosmology and exalts public service by reserving the finest posthumous rewards for outstanding statesmen.

On Laws, a sequel probably left incomplete and published only after Cicero’s death, fills in his constitutional model by outlining a comprehensive legal system. Vying with Plato’s *Laws* this time, he continues his argument that Rome already embodied much of the ideal. His treatment of religion and political administration, which is all that survives, is deeply conservative, largely an explanation and defence of existing statutes and institutions by appeal to Greek theory and Roman history, with proposals for change limited to streamlining and archaizing reforms rather than extensive revision or thorough codification. The work was extremely influential, especially on Christian and early modern thought, because it contains the fullest surviving ancient account of natural law (see Natural law). Drawing heavily on Stoic ideas, Cicero argues that the natural world exhibits a divinely ordained and rationally intelligible order that can be codified in legislation and provides the ultimate tribunal for all positive law.

Civil war interrupted Cicero’s writing and he never returned to constitutional or legal theory. But his very last book rounds out his political thinking by examining the role of personal morality in public life. In the guise of an extended epistle to his son, *On Duties* maps out a code of conduct for the Roman nobility that emphasizes justice, benefaction and public service. The focus throughout is on men of high station and the problems of integrating personal ambitions and social obligations. Borrowing heavily from the Stoic Panaetius (§2), Cicero argues that virtuous conduct is always expedient as well as morally required, and that apparent conflicts between morality and

personal advantage are illusory because virtuous action is always the best option.

3 Scepticism and natural theology

Cicero embraced the sceptical stance of the New Academy in his youth when he studied under Philo of Larissa. Modern scholars debate whether he shifted soon afterward to positions he learned from Antiochus, or remained an Academic sceptic throughout life. But a likely explanation for this debate is the moderate tenor of Cicero's scepticism, which eschews dogmatic certainty but accepts 'convincing' (*probabile*) arguments. Inclined by profession to adversarial argument, Cicero employed the Academic strategy of arguing on both sides of questions in his earlier as well as his later dialogues. But his aim was never suspension of judgment, as in the early New Academy under Arcesilaus (§2). Rather, adopting a pragmatic fallibilism that derives from Carneades (§4) but probably owes most to Philo, he considered the careful weighing of opposing arguments the most reliable route to truth. On this basis, he endorsed an eclectic synthesis of ideas derived mainly from Antiochus and Stoics but also from Aristotle and Plato (whom he styles the prince of philosophers and quotes often in original translations); and while reserving judgment on many theoretical issues, he held firm views about many practical matters of politics and morality.

Cicero discusses his sceptical stance most fully in *Academics*, which inaugurates his second series of dialogues by recounting a pivotal debate within the later New Academy about the nature and possibility of knowledge. Only parts of the work survive. But one fully preserved book (entitled *Lucullus*) summarizes Antiochus' heavily Stoicizing brand of empiricism, then rehearses a battery of sceptical objections drawn mainly from Carneades and Philo. Debate centres on Antiochus' defence of the Stoic 'criterion of truth' (see Stoicism §§12-13), a class of allegedly self-certifying perceptions that would be the source and foundation of any knowledge. Cicero himself delivers the sceptical rebuttal, arguing forcefully that knowledge requires certainty, but that certainty is neither attainable nor necessary for the rational conduct of life. Both sides are powerfully argued, and in line with Philo's fallibilism, readers are obliged to assess which seems more 'convincing'. Though frequently elliptical, the discussion maintains a level of argument rarely matched in Latin before Augustine.

Cicero continued to use this Academic method of presenting opposing arguments in most of his subsequent dialogues. *On the Nature of the Gods* presents two pairs of speeches expounding, then criticizing, Epicurean and Stoic theology. The central issue is not whether gods exist, but rather their role in nature and human affairs. Epicureanism envisages immortals residing far beyond this world and oblivious to human affairs; Stoicism maintains that the world itself is rational and divine, operating in accord with providence to the benefit of humanity; and the Academic critiques, without denying the existence of gods, argue that both schools rely on highly dubious assumptions and wholly fail to prove their cases. The debate covers questions of cosmology at length, including both an impassioned defence and sustained criticism of Stoic arguments from design; and the beliefs and rituals of Greek and Roman polytheism are criticized and reinterpreted throughout. The whole discussion, in its reliance on canons of rational inquiry and analysis, is a model of natural theology. Refusing to rely on revelation or authority, all sides agree that religious issues must be decided by evidence and argument; Stoic appeals to divination and widespread belief in gods, for example, are advanced as inferences to the best explanation. In the end, Cicero endorses the Stoic conception of the world as governed benevolently by divine reason, which he clearly found more compelling both intellectually and morally than either the stark amorality of Epicurean atomism or the official religion of Rome which his public station obliged him to uphold.

Two subsequent works examine related problems of knowledge and causation. *On Divination* presents a richly illustrated defence followed by a scathing critique of various religious claims to foreknowledge. Highlighting epistemological issues, Cicero constructs remarkably balanced cases. The defence, appealing to Stoic arguments that divination is a natural consequence of divine providence and often based on systematic observation, assembles a vast gallery of empirical evidence; his reply, in keeping with the rigorous rationalism of Academic scepticism, challenges the validity of this evidence and argues that divination is inexplicable and inconsistent with the evident regularity of nature. But Cicero, despite ridiculing the defence and pronouncing much of Roman religion sheer superstition, offers no final verdict and endorses traditional rites on grounds of social utility. *On Fate* analyses the closely connected topic of determinism. Cicero defends a theory of free will which he attributes to Carneades, while criticizing the Epicureans for postulating uncaused atomic 'swerves' (see Epicureanism §12), and arguing that the Stoic conception of fate as a universal and eternal causal web precludes voluntary action and moral

responsibility. Detailed analysis of Stoic theories of causation and logical consequence makes this perhaps Cicero’s most exacting work.

4 Humanist ethics

Informing all of Cicero’s work is a profound faith in the natural goodness of humanity and the power of reason to direct and improve human life. The assumptions behind this humanist outlook, which is probably Cicero’s most constructive synthesis of Greek ideas, are systematically examined in *On the Ends of Good and Evil* (usually known as the *De finibus*). Tackling the central question of ancient ethical theory, the dialogue inquires into the ultimate end of human action and how happiness is attained (see Eudaimonia). Paired speeches expound and criticize Epicurean and Stoic accounts of human nature and the status of moral virtue; ending the work and receiving only brief criticism is a neo-Aristotelian account derived from Antiochus. Cicero speaks throughout as an Academic sceptic, arguing that Epicurean hedonism is incoherent and morally subversive, and challenging the Stoic doctrine that moral virtue is the sole good and hence sufficient for happiness. But he commends the moral austerity and theoretical rigour of Stoic ethics, and while he finds the Aristotelian position intuitively attractive and most conducive to public service, he questions whether Antiochus’ view that non-moral interests are also intrinsically good undermines the supremacy of moral virtue.

Setting aside sceptical worries to address practical questions, Cicero explores some applications of his ethical rationalism in two substantial works and a pair of earnest but elegant moralizing essays. *On Duties* employs a Stoic framework to spell out systematic rules of conduct (see §2). *Tusculan Disputations* similarly uses Stoic theory to analyse problems in moral psychology. Adopting a format used by Carneades, Cicero presents five lengthy disquisitions refuting common beliefs about the emotions; but since each speech upholds a Stoic paradox, the result is a sustained defence of Stoic doctrines. Invoking a legion of philosophical and literary authorities, Cicero argues vigorously that philosophy is the medicine of the soul, and that it alone enables us to scorn death, endure pain, overcome grief and other passions, and lead good lives. Much of the argument rests on an acute Stoic analysis of emotions as governed by beliefs. But Cicero’s ideal of rational restraint and self-control as the source of mental tranquillity and happiness distills ideas central to Greek and Roman culture alike. Two shorter dialogues portray eminent Romans from the previous century as sage advisors on more personal topics. *On Old Age* enumerates the lasting rewards of honourable character and education, including a glorious afterlife. *On Friendship* extols a paradigm of aristocratic male companionship based on mutual benefit but also integrity and loyalty. Both works, while distinctly less systematic than Cicero’s other dialogues, exemplify his ability to illuminate vital human concerns with philosophical insight and graceful eloquence.

5 Influence

Cicero was the most influential writer and intellect of his time, and his impact on Western culture has been lasting and profound. His philosophical writings, by forging expressions essential for theoretical discussion, inaugurated over sixteen centuries of philosophy in Latin. They also fuelled the rise of Christianity in the West, as the Latin Fathers mined his dialogues in their campaigns against pagan religion and philosophy (see Patristic philosophy). Augustine, whose life was transformed by the exhortation to philosophy in Cicero’s lost *Hortensius*, drew on his writings extensively, especially in *Against the Academics* and *City of God*. Through these and other writers, most notably Ambrose, Jerome and the pagan Macrobius (whose Neoplatonic commentary on ‘Scipio’s dream’ was widely studied), Cicero’s ideas shaped medieval thought, especially ethics and theories of natural law. His influence reached its zenith in the Renaissance, when Erasmus and other humanists emulated his critical spirit and reaffirmed his secular outlook and ecumenical ideals (see Renaissance philosophy). By making his writings the foundation of a liberal education, they also increased his moral authority; and in the following centuries his ethical and political works (above all *On Duties*, dubbed ‘Tully’s Offices’ in English) fostered the revival of republicanism and the development of liberalism, while his dialogues on religion were inspirational to deism. Voltaire proclaimed him the model of enlightened reason; and Hume was deeply indebted to Cicero, especially in his critique of religious dogmatism and his conception of ‘mitigated’ scepticism. The rise of idealism in the nineteenth century lowered Cicero’s philosophical reputation considerably. But renewed interest in scepticism and virtue ethics, along with improved understanding of Hellenistic philosophy, has in recent decades stimulated intense discussion of his work yet again.

STEPHEN A. WHITE

List of works

Cicero, M.T. (80s BC) *On Invention (De inventione)*, in two books, trans. H. Hubbell, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1949.(Handbook on rhetoric and informal argument; includes Latin text with translation.)

Cicero, M.T. (55 BC) *On the Orator (De oratore)*, in three books, trans. E. Sutton and H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1942.(Dialogue debating the place of rhetoric and philosophy in politics and education.)

Cicero, M.T. (54-51 BC) *On the Republic (De republica)*, in six books, trans. G. Sabine and S. Smith, *On the Commonwealth*, Columbus, OH: Ohio State University Press, 1929.(Dialogue analyzing the structure and legitimacy of the Roman state; less than a third preserved.)

Cicero, M.T. (late 50s BC) *On Laws (De legibus)*, in three books, trans. C. Keyes, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1928.(Dialogue on the Roman legal system; incomplete.)

Cicero, M.T. (46 BC) *Stoic Paradoxes (Paradoxa Stoicorum)*, trans. M.R. Wright, in *On Stoic Good and Evil*, Warminster: Aris & Phillips, 1991.(Short declamations defending six claims of Stoic ethics.)

Cicero, M.T. (late 46 BC) *Hortensius*, ed. A. Grilli, Milan: Istituto Editoriale Cisalpino, 1962.(Lost dialogue proclaiming the value of philosophy; only fragments and reports survive.)

Cicero, M.T. (early 45 BC) *Academics*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933.(Dialogue on the nature and possibility of knowledge. Substantial parts of two versions survive: the second of two books from the first version, entitled *Lucullus* but often called *Prior Academics* or *Academics* II; and the first of four books from a thorough revision, entitled *Varro* but often called *Posterior Academics* or *Academics* I.)

Cicero, M.T. (mid 45 BC) *On the Ends of Good and Evil (De finibus bonorum et malorum)*, in five books, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1914; book III trans. M.R. Wright, in *On Stoic Good and Evil*, Warminster: Aris & Phillips, 1991; book V trans. S. White, Oxford: Oxford University Press, forthcoming.(Three dialogues summarizing and criticizing Epicurean, Stoic, and Aristotelian ethical theories.)

Cicero, M.T. (late 45 BC) *Tusculan Disputations (Tusculanae disputationes)*, in five books, trans. J.E. King, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 2nd edn, 1945; books I, II and V trans. A.E. Douglas, Warminster: Aris & Phillips, 1985 and 1990. (Five disquisitions on topics in moral psychology.)

Cicero, M.T. (late 45 BC) *On the Nature of the Gods (De natura deorum)*, in three books but incomplete, trans. P.G. Walsh, Oxford: Oxford University Press, 1997.(Dialogue summarizing and criticizing Epicurean and Stoic views about divinity.)

Cicero, M.T. (45 or 44 BC) *Timaeus*, ed. R. Giomini, Leipzig: Teubner, 1975.(Latin translation of part of Plato's *Timaeus*: 27d-47b, with gaps.)

Cicero, M.T. (early 44 BC) *Cato the Elder: On Old Age (Cato maior, De senectute)*, trans. F. Copley, Ann Arbor, MI: University of Michigan Press, 1967.(Dialogue on how to live after retiring from public life.)

Cicero, M.T. (early 44 BC) *On Divination (De divinatione)*, in two books, trans. H.M. Poteat, Chicago, IL: University of Chicago Press, 1950.(Dialogue arguing first for, then against a religious basis for knowledge about the future.)

Cicero, M.T. (mid 44 BC) *On Fate (De fato)*, trans. R. Sharples, Warminster: Aris & Phillips, 1991.(Dialogue on causation and determinism; incompletely preserved.)

Cicero, M.T. (mid 44 BC) *Topics*, trans. H. Hubbell, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1949.(Synopsis of informal reasoning.)

Cicero, M.T. (mid 44 BC) *Laelius: On Friendship (Laelius, De amicitia)*, trans. J.G.F. Powell, Warminster: Aris & Phillips, 1990.(Dialogue on the nature and value of friendship.)

Cicero, M.T. (late 44 BC) *On Duties (De officiis)*, in three books, trans. M. Griffin and E.M. Atkins, Cambridge: Cambridge University Press, 1991.(Treatise examining virtuous conduct and its relation to personal advantage.)

References and further reading

Gawlick, G. and Görler, W. (1994) 'Cicero', in H. Flashar (ed.) *Die Philosophie der Antike* 4, Basle: Schwabe, 991-1168.(Comprehensive survey in German of Cicero's philosophical *oeuvre*, with full bibliography.)

Glucker, J. (1988) 'Cicero's Philosophical Affiliations', in J. Dillon and A.A. Long (eds) *The Question of 'Eclecticism'*, Berkeley, CA: University of California Press.(Influential paper arguing that Cicero initially

embraced Philo's scepticism but adopted views of Antiochus for most of his career and reverted to scepticism only in his sixties.)

Griffin, M. and Barnes, J. (eds) (1991) *Philosophia togata: Essays on Philosophy and Roman Society*, Oxford: Oxford University Press.(Several papers discuss Cicero at length; useful analytical bibliography.)

Laks, A. and Schofield, M. (eds) (1995) *Justice and Generosity: Studies in Hellenistic Social and Political Philosophy*, Cambridge: Cambridge University Press.(Includes studies of Cicero's thought by J.-L. Ferrary, A. A. Long and M. Schofield.)

Powell, J.G.F. (ed.) (1995) *Cicero the Philosopher*, Oxford: Oxford University Press.(A dozen studies on diverse topics, including a rejoinder to Glucker by W. Görler.)

Rawson, E. (1983) *Cicero: A Portrait*, Ithaca, NY: Cornell University Press, 2nd edn. (Thoughtful and engaging biography emphasizing intellectual and ethical aspects.)

Cieszkowski, August von (1814-94)

Cieszkowski is perhaps best known for being the first to revise Hegel's philosophy of history into a basis for future social reform. His introduction of the idea of 'praxis' as a synthesis between abstract theory and undirected practice was fundamental in the development of the Young Hegelian school, which formed shortly after Hegel's death and sought the radical transformation of all then-existing political, religious and economic orders.

August was born of a wealthy, aristocratic Polish family. He received an excellent education, particularly in languages. His concern with democratic reform was first shown by his participation in the Polish anti-Russian revolutions of 1830-1. In 1832, he transferred from the Jagiellonian University of Cracow to the University of Berlin. There he joined the circle of Hegel's more liberal interpreters and soon developed an appreciation of the practical possibilities of Hegelian theory.

At the time, most of Hegel's students, inspired by the French Revolution of 1830, were impatient to begin the project which they took to be implied in Hegelian theory: the practical reformation of their age. To this end, both orthodox religion and politics had to be completely revised. Cieszkowski became particularly close to two young professors at the University: Eduard Gans and Carl-Ludwig Michelet. Unlike some of their colleagues, they understood Hegel as supporting liberal doctrines, and they led Cieszkowski towards developing a Hegelian basis for political action. In 1838, Cieszkowski proceeded to revise theoretical Hegelianism in order to reveal its revolutionary content. The result was *Die Prolegomena zur Historiosophie (Prolegomena to the Wisdom of History)*. This work was immediately and enthusiastically approved by the progressive Hegelians, in particular by the extreme 'Young Hegelians' - a group which included such radicals as Moses Hess, Ludwig Feuerbach, Max Stirner, Arnold Ruge, Karl Marx and Bruno Bauer.

Although David F. Strauss' *Das Leben Jesu (Life of Jesus)* (1835-6) is rightly credited with being the first Young Hegelian work, it was nevertheless Cieszkowski's work which set the general pattern for the Young Hegelian school. Prior to it, a major difficulty confronting liberal activists was Hegel's own evident lack of concern with the future course of world history (see Hegel, G.W.F. §8). Hegel's view, expressed throughout his works, held that philosophy played no role in '*teaching* what the world ought to be'. This attitude did not sit well with the Young Hegelians. Cieszkowski proposed a radical restructuring of Hegel's own fourfold division of historical ages (Oriental, Greek, Roman and Christian-Germanic) into a triadic and so more 'dialectical' structure: past, present and future. The past, which linked Adam to Christ, was characterized by a focus on practice; the present, from Christ to Hegel, by a focus on theory. The past, with its direct engagement with the practical and concrete issues of reality, had been followed by an age of equal and opposite focus, one in which theoretical speculation drew mankind away from everyday concerns into an abstract mental realm. The theoretical age, which began with Christ's proclamation of the 'Kingdom of God', asserted the ascendancy of the soul over the body. This age ended in the perfected theory of Hegel. After his final and conclusive theory, the post-Hegelian future would be characterized by a synthesis of past practice and present theory - a new era of 'praxis'. This new division of history obviously justified the study of Hegelian theory by giving theorists a major role in setting a future agenda. Praxis, as expressed in the 'deed' linking the abstract theory of Hegelianism with practical action, would engage future thinkers in an ongoing process of rationalizing the real. In sum, the final result of Hegelian theory was, in a proper dialectical manner, the return of that theory into the practical world from which it had emerged. Cieszkowski's work not only redirected Hegelian interest from religion to politics, but also established a programme for Hegelians. Henceforth, given that Hegelianism was the perfected philosophical theory, there only remained the practical task of conforming the world to it (see Hegelianism §2).

In 1842 Cieszkowski published another philosophical work, *Gott und Palingenesie (God and Regeneration)*, directed against Michelet's reduction of God to an impersonal ideal. For Cieszkowski this implied a failure to grasp the concreteness of the divine personality, and was a step back from the threshold of future praxis into the earlier age of abstract theorizing. However, due to its opaque style and the obscurity of its subject, the work was unsatisfying even to its author.

In the year following the publication of the *Prolegomena*, Cieszkowski visited France, where he become deeply interested in an alternative to German philosophic theory: that is, French economic practice. Moving in the social

circles of such eminent liberal French economists as Auguste Blanqui, Hippolyte de Passy, and P.L.E. Rossi, as well as with such thinkers as Pierre-Joseph Proudhon, Royer-Collard and Louis Blanc, it was not long before Cieszkowski's first economic work appeared: *Du crédit et de la circulation (Concerning Credit and Circulation)* (1838b). It advocated the reform of the monetary system by introducing a new type of interest-bearing note. The work, based upon several years of research into economic issues, proposed to balance both private and public economic interests. In 1844 he wrote a short work proposing a reform of the decadent French Upper House, *De la pairie et de l'aristocratie moderne (On the Peerage and Modern Aristocracy)*. An indirect critique of the French bourgeoisie, it reasserted the need for a strong legislative body representing universal interests, a 'popular patriciate' defending the public order threatened by the anarchic individualism of the Lower House.

After the abortive revolutions of 1848, Cieszkowski became a member of the Prussian Lower House, and supported the cause of Polish nationalism. In 1866 he retired from politics to his estate in Poland, where, until his death, he pursued varied public-interest projects and worked on a lengthy treatise entitled *Ojcze Nasz (Our Father)*. He considered this his major work. But although begun in the 1830s, it remained unfinished at his death in 1894. It is a grand and cloudy romantic vision of a political order inspired by theological principles - principles which Cieszkowski thought were embedded in the Lord's Prayer. The Prayer was the revelation of God's plan, the divine praxis required for the establishment of a new and harmonious social order - an earthly 'Kingdom of God'.

See also: History, philosophy of

LAWRENCE S. STEPELEVICH

List of works

Cieszkowski, A. von (1838a) *Die Prolegomena zur Historiosophie (Prolegomena to the Wisdom of History)*, Berlin: Viet.(The author's major study revising Hegel's philosophy of history.)

Cieszkowski, A. von (1838b) *Du crédit et de la circulation (Concerning Credit and Circulation)*, Paris: Treuttel et Wurtz.(An economic proposal to reform French government monetary policy.)

Cieszkowski, A. von (1842) *Gott und Palingenesie (God and Regeneration)*, Berlin: E.H. Schröder. (A speculative exercise attempting to personalize the Hegelian absolute spirit.)

Cieszkowski, A. von (1844a) *De la pairie et de l'aristocratie moderne (On the Peerage and Modern Aristocracy)*, Paris: Amyot.(A proposal to liberalize and reform the French Upper House towards making it an effective instrument of government.)

Cieszkowski, A. von (1844b) *Ojcze Nasz (Our Father)*, Paris: Maulde et Renou.(A reading of the Lord's Prayer as a key to social reform.)

Cieszkowski, A. von (1979) *Selected Writings of August Cieszkowski*, ed., trans. and intro. A. Liebich, Cambridge: Cambridge University Press.(Contains a comprehensive bibliographic essay.)

References and further reading

Liebich, A. (1979) *Between Ideology and Utopia: The Politics and Philosophy of August Cieszkowski*, Dordrecht: Reidel.(A comprehensive study of the whole of Cieszkowski's life and works.)

Lobkowicz, N. (1967) *Theory and Practice: History of a Concept from Aristotle to Marx*, Notre Dame, IN: Notre Dame University Press, 193-204.(A study placing Cieszkowski in the context of philosophic history.)

Stepelevich, L.S. (1974) 'August von Cieszkowski: From Theory to Practice', *History and Theory* 13 (1): 39-53. (Relates Cieszkowski to the Young Hegelian school.)

Stepelevich, L.S. (1987) 'Making Hegel Into a Better Hegelian: August von Cieszkowski', *Journal of the History of Philosophy* 25 (4): 263-73.(A critique of Cieszkowski's revision of Hegel's philosophy of history.)

Strauss, D.F. (1835-6) *Das Leben Jesu (Life of Jesus)*, Tübingen: Osiander; trans. M.A. Evans (George Eliot), *Life of Jesus*, Philadelphia, PA, Fortress, 1972.(Argues that the messianic Christ was merely a mythic figure.)

Citizenship

Within political philosophy, citizenship refers not only to a legal status, but also to a normative ideal - the governed should be full and equal participants in the political process. As such, it is a distinctively democratic ideal. People who are governed by monarchs or military dictators are subjects, not citizens. Most philosophers therefore view citizenship theory as an extension of democratic theory. Democratic theory focuses on political institutions and procedures; citizenship theory focuses on the attributes of individual participants.

One important topic in citizenship theory concerns the need for citizens to actively participate in political life. In most countries participation in politics is not obligatory, and people are free to place private commitments ahead of political involvement. Yet if too many citizens are apathetic, democratic institutions will collapse. Another topic concerns the identity of citizens. Citizenship is intended to provide a common status and identity which helps integrate members of society. However, some theorists question whether common citizenship can accommodate the increasing social and cultural pluralism of modern societies.

1 Citizenship and democratic theory

‘Citizenship’ is a term the philosophical meaning of which differs from its everyday usage. In everyday speech, citizenship is often used as a synonym for ‘nationality’, to refer to one’s legal status as a member of a particular country. To be a citizen entails having certain rights and responsibilities, but these vary greatly from country to country. For example, citizens in a liberal democracy have political rights and religious liberties, whereas citizens in a monarchy, military dictatorship or religious theocracy may have neither.

In philosophical contexts, citizenship refers to a substantive normative ideal of membership and participation in a political community. To be a citizen, in this sense, is to be recognized as a full and equal member of society, with the right to participate in the political process. As such, it is a distinctively democratic ideal. People who are governed by monarchs or military dictators are subjects, not citizens.

This link between citizenship and democracy is evident in the history of Western thought. Citizenship was a prominent theme among philosophers of the ancient Greek and Roman republics, but disappeared from feudal thought, only to be revived with the rebirth of republicanism during the Renaissance (see Political philosophy, history of §§2, 4). Indeed, it is sometimes difficult to distinguish citizenship, as a philosophical topic, from democracy. However, theories of democracy primarily focus on institutions and procedures - political parties, elections, legislatures and constitutions - while theories of citizenship focus on the attributes of individual citizens.

Theories of citizenship are important because democratic institutions will collapse if citizens lack certain virtues, such as civic-mindedness and mutual goodwill. Indeed, many democracies suffer from voter apathy, racial and religious intolerance, and significant non-compliance with taxation or environmental policies that rely on voluntary cooperation. The health of a democracy depends not only on the structure of its institutions, but also on the qualities of its citizens: for example, their loyalties and how they view potentially competing forms of national, ethnic or religious identities; their ability to work with others who are different from themselves; their desire to participate in public life; their willingness to exercise self-restraint in their economic demands and in personal choices affecting their health and the environment.

2 The responsibilities of citizenship

In ancient Athens, citizenship was viewed primarily in terms of duties. Citizens were legally obliged to take their turn in public office and sacrificed part of their private life to do so. In the modern world, however, citizenship is viewed more as a matter of rights than duties. Citizens have the right to participate in politics, but also the right to place private commitments ahead of political involvement.

An influential exposition of this conception of ‘citizenship-as-rights’ is T.H. Marshall’s *Citizenship and Social Class* (1950). Marshall divides citizenship rights into three categories: civil rights, which arose in England in the eighteenth century; political rights, which arose in the nineteenth century; and social rights - for example, to education, health care, unemployment insurance and old-age pension - which have become established in the

twentieth century (see Rights §§4-5). For Marshall, the culmination of the citizenship ideal is the social-democratic welfare state. By guaranteeing civil, political and social rights to all, the welfare state ensures that every member of society is able to participate fully in the common life of society (see Social democracy).

This is often called ‘passive’ citizenship, because of its emphasis on passive entitlements and the absence of any civic duties. While it has helped secure a reasonable degree of security, prosperity and freedom for most members of Western societies, most theorists believe that the passive acceptance of rights must be supplemented with the active exercise of responsibilities and virtues. Theorists disagree, however, about which virtues are most important and how best to promote them.

Conservatives emphasize the virtue of self-reliance. Whereas Marshall argued that social rights enable the disadvantaged to participate in the mainstream of society, conservatives argue that the welfare state has promoted passivity and dependence among the poor. To promote active citizenship, we should reduce welfare entitlements and emphasize the responsibility to earn a living, which is the key to self-respect and social acceptance (see Conservatism). Critics respond that cutting welfare benefits further marginalizes the underclass. Also, as feminists note, gender-neutral talk about ‘self-reliance’ is often a code for the view that men should financially support the family while women look after the household and care for the elderly, sick, and young. This reinforces the barriers to women’s full participation in society (see Feminist political philosophy).

Civil society theorists focus on how we learn to be responsible citizens. They argue that it is in the voluntary organizations of civil society - churches, families, unions, ethnic associations, environmental groups, neighbourhood associations, support groups - that we learn civic virtues. Because these groups are voluntary, failure to live up to the responsibilities they entail is met with disapproval rather than legal punishment. Yet because the disapproval comes from family, friends and colleagues, it is often a more powerful incentive to act responsibly than punishment by an impersonal state (see Civil society).

The claim that civil society is the seedbed of civic virtue is debatable. The family teaches civility and self-restraint, but it can also be ‘a school of despotism’ that teaches male dominance over women. Similarly, churches often teach deference to authority and intolerance of other faiths; ethnic groups often teach prejudice against other races, and so on.

Liberal virtue theorists emphasize the importance of citizens’ ability to engage in public discourse. This does not mean simply making one’s views known. It also involves the virtue of ‘public reasonableness’. Citizens must give reasons for their political demands, not just state preferences or make threats. Moreover, these reasons must be ‘public’ reasons, in the sense that they are capable of persuading people of different faiths and nationalities. It is not enough to invoke scripture or tradition; a conscientious effort is required to distinguish those beliefs which are matters of private faith from those which are capable of public defence.

Where do we learn this virtue? Liberal virtue theorists often suggest that schools should teach children to distance themselves from their own cultural traditions when engaging in public discourse and to consider other points of view. However, traditionalists object that this encourages children to question parental or religious authority in private life. Those groups which rely on an uncritical acceptance of tradition and authority are threatened by the open-minded and pluralistic attitudes which liberal education encourages. Hence some religious groups see compulsory liberal education as an act of intolerance towards them, even if it is carried out in the name of teaching the virtue of tolerance (see Toleration).

Civic republicans offer another approach to responsible citizenship. In its most general sense, ‘civic republican’ refers to anyone who believes in the need for active and responsible citizens. But there is a narrower conception of civic republicanism, distinguished by its belief (following Aristotle) in the intrinsic value of political participation. Such participation is, in Adrian Oldfield’s words, ‘the highest form of human living-together that most individuals can aspire to’ (1990: 6). On this view, political life is superior to the merely private pleasures of family, neighbourhood and profession, and so should occupy the centre of people’s lives.

This view conflicts with modern understandings of the good life in the Western world. Most people today find the greatest happiness in their family life, work, religion or leisure, not in politics. Political participation is seen as an occasional, often burdensome, activity necessary to ensure that government respects and supports people’s freedom to pursue their personal projects and attachments. The assumption that politics is primarily a means to

protect and promote private life underlies most modern views of citizenship. This attitude may reflect the impoverishment of public life today, compared to the active citizenship of ancient Greece. Political debate seems less meaningful, and people feel less able to participate effectively. But it also reflects the enrichment of private life, given the increased prominence of romantic love and the nuclear family (with its emphasis on intimacy and privacy); increased prosperity (and so richer forms of leisure and consumption); and modern beliefs in the dignity of labour (which the Greeks despised). The call for active citizenship today must compete with the powerful attractions of private life (see Liberalism §§3-4).

3 Citizenship, identity and difference

Citizenship is not just a status, defined by a set of rights and responsibilities. It is also an identity, an expression of one's membership in a political community. Moreover, it is a shared identity, common to the diverse groups in society. Citizenship therefore serves an integrative function. Extending citizenship rights has helped integrate previously excluded groups, like the working class, into society.

Some groups, however - for example, African-Americans, indigenous peoples, ethnic and religious minorities, gays and lesbians - still feel excluded from the mainstream of society, despite possessing the common rights of citizenship. According to cultural pluralists, citizenship must reflect the distinct sociocultural identity of these groups - their 'difference'. The common rights of citizenship, originally defined by and for white men, cannot accommodate the needs of marginalized groups. These groups can only be fully integrated through what Iris Marion Young calls 'differentiated citizenship' (1989). That is, the members of certain groups should be incorporated into the political community, not only as individuals, but also through the group, and their rights should depend in part on their group membership.

This view challenges traditional conceptions of citizenship, which define citizenship in terms of treating people as individuals with equal rights under the law. This is how democratic citizenship is standardly distinguished from feudal and other pre-modern views that determined people's political status by their religion, ethnicity, class or gender. The idea of differentiated citizenship, therefore, is seen by many as a contradiction in terms. Moreover, if groups are encouraged by the very terms of citizenship to turn inward and focus on their 'difference', how can citizenship provide a source of commonality and solidarity for the various groups in society?

It is important to distinguish here two broad categories of differentiated citizenship. For some groups - such as the poor, women, racial minorities and immigrants - the demand for group rights is usually a demand for greater inclusion and participation in the mainstream society. For example, these groups may feel underrepresented in the political process due to historic barriers, and so seek special group-based representation (see Representation, political §2). Or they may want the school curriculum to recognize their contributions to society's culture and history. Or they may seek exemptions from laws that disadvantage them economically, given their beliefs and practices. These groups share the goal of national integration - that is, that historically disadvantaged groups should become full and equal participants in the mainstream society. They simply claim that recognition and accommodation of their 'difference' is needed to ensure integration.

Other groups demanding differentiated citizenship reject the goal of national integration. They wish to govern themselves, apart from the larger society. This is particularly true of national minorities - that is, distinct historical communities, occupying their own homeland or territory and sharing a distinct language and history. These groups find themselves within the boundaries of a larger political community, but claim the right to govern themselves in certain matters in order to ensure the free development of their culture. What these national minorities want is not primarily better representation in the central government, but rather the transfer of power from the central government to their own communities, often through some system of federalism or local autonomy (see Federalism and confederalism). Rather than seeking greater inclusion into the larger society, they seek greater autonomy from it.

This sort of demand challenges traditional accounts of citizenship identity, which presuppose that people view themselves as members of the same society. If democracy is the rule of the people, demands for self-government raise the question of who 'the people' really are. National minorities claim that they are distinct 'nations' or 'peoples', with inherent rights of self-determination which were not relinquished by their (often involuntary) federation with other nations within a larger country. Self-government rights divide the people into separate

peoples, each with its own rights, territories and powers of self-government; and each, therefore, with its own political community. If citizenship is membership of a political community, then self-government rights give rise to a sort of dual citizenship, and to conflicts about which community - the national group or the state - citizens identify with most deeply. Moreover, if limited autonomy is desirable for a national minority, why not secede completely and have a fully autonomous nation-state?

In effect, countries with national minorities face the problem of conflicting nationalisms (see Nation and nationalism). The state seeks to promote a single national identity through common citizenship; the minority seeks to promote its distinct national identity through differentiated citizenship. Finding a source of social unity in such multination countries is a fundamental question facing theorists of citizenship.

See also: Democracy; Republicanism

WILL KYMLICKA

References and further reading

All of these works are written in nontechnical language, accessible to those with little or no background in political philosophy.

Beiner, R. (1994) *Theorizing Citizenship*, Albany, NY: State University of New York Press.(A collection of prominent contemporary essays on citizenship, covering most of the theories and issues discussed in this article.)

Clarke, P.B. (1994) *Citizenship*, London: Pluto Press.(A collection of historical writings on citizenship, from Aristotle to the present day.)

Galston, W. (1991) *Liberal Purposes: Goods, Virtues, and Duties in the Liberal State*, Cambridge: Cambridge University Press.(A prominent statement of the liberal virtue theory discussed in §2.)

Kymlicka, W. (1995) *Multicultural Citizenship*, Oxford: Oxford University Press.(An evaluation of the cultural pluralist critique of universal citizenship, discussed in §3.)

Kymlicka, W. and Norman, W. (1994) ‘Return of the Citizen: A Survey of Recent Work on Citizenship Theory’, *Ethics* 104 (2): 352-81.(A summary of the recent literature on citizenship, with an extensive bibliography.)

Macedo, S. (1990) *Liberal Virtues: Citizenship, Virtue and Community*, Oxford: Oxford University Press.(A statement of the liberal virtue theory, discussed in §2.)

Marshall, T.H. (1965) *Citizenship and Social Class*, Cambridge: Cambridge University Press.(The classic exposition of the modern theory of citizenship-as-rights, discussed in §2.)

Mouffe, C. (1992) *Dimensions of Radical Democracy: Pluralism, Citizenship and Community*, London: Routledge.(A collection of essays on citizenship theory, covering issues of both civic virtue and cultural pluralism.)

Oldfield, A. (1990) *Citizenship and Community: Civic Republicanism and the Modern World*, London: Routledge. (An exposition of the civic republican view, discussed in §2.)

Spinner, J. (1994) *The Boundaries of Citizenship: Race, Ethnicity and Nationality in the Liberal State*, Baltimore, MD: Johns Hopkins Universtity Press.(A liberal response to the cultural pluralist critique of universal citizenship, discussed in §3.)

Walzer, M. (1989) ‘Citizenship’, in T. Ball and J. Farr (eds) *Political Innovation and Conceptual Change*, Cambridge: Cambridge University Press.(A concise history of the concept of citizenship.)

Walzer, M. (1992) ‘The Civil Society Argument’, in C. Mouffe *Dimensions of Radical Democracy: Pluralism, Citizenship and Community*, London: Routledge; also in R. Beiner *Theorizing Citizenship*, Albany, NY: State University of New York Press, 1994. (A clear statement of the civil society theory, discussed in §2.)

Young, I.M. (1989) ‘Polity and Group Difference: A Critique of the Ideal of Universal Citizenship’, *Ethics* 99 (2): 250-74; repr. in R. Beiner *Theorizing Citizenship*, Albany, NY: State University of New York Press, 1994.(A prominent statement of the cultural pluralist critique of universal citizenship, discussed in §3.)

Civil disobedience

According to common definitions, civil disobedience involves a public and nonviolent breach of law that is committed in order to change a law or policy, and in order to better society. More, those classed as civilly disobedient must be willing to accept punishment. Why is the categorization of what counts as civil disobedience of practical importance? The usual assumption is that acts of civil disobedience are easier to justify morally than other illegal acts. Acts of civil disobedience, such as those committed by abolitionists, by followers of Mahatma Gandhi and Martin Luther King, Jr and by opponents of the Vietnam War, have been an important form of social protest.

The decision as to what exactly should count as civil disobedience should be guided both by an ordinary understanding of what the term conveys and by what factors are relevant for moral justification. For justification, nonviolence and publicness matter because they reduce the damage of violating the law. Tactics should be proportionate to the evil against which civil disobedience is aimed; someone who assesses the morality of a particular act of civil disobedience should distinguish an evaluation of tactics from an evaluation of objectives.

1 What constitutes civil disobedience?

People care about definitions of civil disobedience because they believe that acts of civil disobedience are easier to defend than ordinary illegal acts. To understand that assumption, it helps to see how any approach to civil disobedience fits within a larger sense of the moral wrongness of most law breaking. Reasons against violating the law are founded on moral rights or justice (deontological reasons) or on harmful consequences (consequential reasons) (see Deontological ethics; Consequentialism). Many acts that breach standards of criminal and civil law are independently immoral. Thus, murders breach the moral rights of victims and are undesirable measured by consequences. Whether or not people have some general responsibility to obey the law, simply because it is the law, the existence of duties to other members of the community and reasons of consequence support keeping the law on many other occasions. For instance, at least within a fair and reasonably effective legal system, taxpayers should pay their taxes (see Obligation, political).

For acts of civil disobedience to be morally justified, reasons in their favour must outrank or outweigh whatever moral reasons lie behind obeying the law. The significance of an act's constituting civil disobedience is that reasons in favour of obedience are reduced or reasons for disobeying are strengthened, or both. An approach to what counts as civil disobedience should be responsive to widespread understanding of the term's meaning and to factors that matter morally.

Any effort to isolate the phenomenon of civil disobedience must distinguish between what we may call full disobedience of law and violating some 'lower'-level rule whilst hoping or expecting that one will be regarded as justified by some higher *legal* rule. In regimes with written constitutions, for example, someone may claim a constitutional privilege to disobey a statute. Such an action, based on an accepted constitutional right, is not true civil disobedience. However, classification becomes more troublesome when someone breaks a law believing that the law should be declared unconstitutional, but strongly expecting that courts will disagree. This essay disregards this subtlety and concentrates on persons who lack any serious claim that their behaviour is protected by another, higher legal rule.

People usually disobey laws from self-interest or strong emotion, without any belief that they are justified morally. But not infrequently people think they have good moral reason to disobey. They may believe that because the precise terms of a law are not regarded seriously or are widely disobeyed, they have little moral reason to obey. (Many Americans feel this way about speed limits on highways.) Even when they perceive substantial reasons to obey, people may think disobedience is justified because they cannot conscientiously perform the required acts, because they face conflicting moral claims in the circumstances or because they seek to achieve broader social goals. The first sort of reason is exemplified by the refusal on a religious basis to submit to conscription. Those engaging in these conscientious refusals would welcome legal change, but their disobedience has a more personal point. Someone who lies under oath to protect a friend lacks such a conscientious objection, but may believe duties of friendship are stronger than the responsibility to testify truthfully. Other illegal acts have some broader social objective. Acts of civil disobedience have such broader social objectives.

An extremely expansive definition of civil disobedience is any ‘deliberate violation of law for a vital social purpose’ (Zinn 1968: 39). According to this definition, the poisoning of Rasputin, and even the rape of Muslim women in Bosnia - if done to help impose Serbian rule - would qualify as civil disobedience. Such an understanding strays too far from the basic conception that ‘civil’ in ‘civil disobedience’ relates not only to the civil order but also to the kinds of tactics that are used. On this view, ‘civil’ is opposed to ‘violent’, although the line between the two is hardly transparent (see Violence).

At the other end of the spectrum of definitions is John Rawls’ account (see Rawls, J. §4). Rawls states that someone who engages in civil disobedience ‘invokes the commonly shared conception of justice that underlies the political order’, declaring that principles of justice are not being respected and aiming to make the majority reconsider the justice of its actions (1971: 364-5). This approach is part of Rawls’ comprehensive theory of justice for liberal democracies (and does not indicate what counts as civil disobedience for tyrannies). Although not ruling out all other forms of resistance for democracies, Rawls thinks acts of civil disobedience are easier to justify.

Rawls’ account disqualifies many acts commonly regarded as involving civil disobedience. Vegetarians publicize their moral abhorrence of killing animals for food by blocking entrances to stockyards; pacifists trespass peacefully on an army base to object to a popular war. In countries where dominant views are neither vegetarian nor pacifist, such demonstrations do not qualify as civil disobedience for Rawls. Analysis is more complicated if objectors trespass on nuclear power facilities. A demonstrator convinced that shared principles of justice have not been respected will be engaging in civil disobedience, but one who sees the problem as one of government stupidity will be doing something else. Such shades of evaluation probably should not determine whether someone’s participation constitutes civil disobedience.

Rawls’ idea that civil disobedience makes the majority reconsider the justice of its actions oversimplifies the objectives underlying most such activities. Suppose people block construction work where officials have approved destruction of a forest in order that expensive housing may be build. The demonstrators have a triple message: that the government acted unjustly on the facts available; that the government should now change course in light of the demonstrated intensity of opposition; and that inconvenience and embarrassment will attend continuation of the present policy. Demonstrators often seek to manipulate costs and benefits to persuade officials that the present course is too expensive. *Pure* appeals to justice are rare; civil disobedience usually involves calculated elements of pressure. Some of Rawls’ definitional criteria may be better understood as factors relevant to deciding whether acts of civil disobedience are justified. For deciding that acts count as civil disobedience, an account drawn from common definitions is preferable.

2 Justifying civil disobedience

The main argument in favour of engaging in civil disobedience is that an overt violation of law may be effective in a way that neither ordinary speech nor lawful demonstrations can achieve. If one’s objectives are important enough, reasons in favour of obeying the law may be overridden. Acts of civil disobedience do less damage to the values of law observance than do most other violations of law that are intended to achieve similar objectives.

The requirement of nonviolence is directly relevant to the issue of damage. Acts of civil disobedience typically cause inconvenience, to a greater or lesser degree. Some also involve breaches of property rights, such as trespasses. However, they do not threaten the safety of persons, the most important interest that the law protects.

The publicness of civil disobedience is mainly significant because it ensures that violators can be identified and arrested, and can be punished if authorities proceed against them. Voluntary submission to punishment goes some way towards satisfying one’s duty to be fair to others in respect to law. Ordinary violations of law take advantage of those who ‘play by the rules’. But when people act openly in a way that makes punishment simple, they implicitly declare that they seek no personal advantage from their action and are willing to pay the appropriate penalty. They also openly commit their fate to legal processes. Even if submission to punishment does not *fully* satisfy a duty of fairness in respect to law, it reduces greatly the degree of unfairness (see Crime and punishment §1). The consequences of illegal action are also different when violators act openly and submit to punishment. The frustration, resentment and insecurity people feel when their interests are jeopardized are reduced if they realize that those who threaten them are willing to pay an even more costly price. By acting openly, the violators demonstrate the depth of their convictions, showing that they take their claims very seriously. They also signify

their respect for their fellow citizens and for the law, especially if they not only submit to punishment willingly but also acknowledge its appropriateness. Openness also reduces the destructive potentiality of illegal actions as examples to others with various social grievances.

Since the main significance of publicness or openness is in its connection to identification and punishment, openness should be understood in a way that corresponds with this significance. Suppose war protesters surreptitiously enter a government office at night, pour blood on some military files and openly declare the following morning exactly what they have done and why. This behaviour may count as ‘open’ for the purpose of classification as civil disobedience, even though the violation itself was covert. Some open actions do not involve the element of submission to punishment. If police who know they are easily identifiable block traffic but are counting on their bargaining power as indispensable peace keepers to ensure that no punishment is imposed, their unwillingness to submit to the law is little affected by the openness of their actions. For this reason, their behaviour probably should not count as civil disobedience.

What constitutes the conditions for civil disobedience to be morally justified? Perhaps the most common suggestion is that lawful means of protest have been pursued unsuccessfully. This condition makes obvious sense. If a law or policy may be altered by legal means, resort to more drastic illegal means is not yet appropriate. Thus, speeches, letters to politicians and lawful demonstrations should precede unlawful demonstrations. The principle is sound, although people often disagree over when lawful means have proved unsuccessful. The principle does not apply in its usual form in emergency situations, say when the government undertakes an unexpected invasion of another country, and, perhaps, it does not apply when the unproductiveness of legal means of protest is patently clear in advance.

Former US Supreme Court Justice Abe Fortas once suggested that disobedience is never justified unless the law being violated is the target of the protest (Fortas 1968). Although disobedience is usually easier to justify when a close connection exists between the injustice people protest against and the law they disobey, no absolute principle of the kind Fortas proposed is defensible. Among other objections, the principle might leave some laws and policies - for example, a highly unjust definition of treason or an egregious use of military force abroad - entirely immune from law-violating protest; and it would also preclude such protests for the government’s failure to enact laws to prevent injustice.

Rawls has proposed that ordinarily, justified civil disobedience will be limited ‘to instances of substantial and clear injustice, and preferably to those which obstruct the path of removing other injustices’ (1971: 372). The clearer an injustice, the stronger the justification for illegal protest. But what of parents blocking a street in order to get a traffic light installed? Contained instances of civil disobedience may be warranted to combat less than grave injustices, so long as the tactics used are not disproportionate to the evil that they protest against. Rawls’ insistence that protesters appeal to a shared sense of justice might be taken as a condition for justified civil disobedience in a liberal democracy (even if it is rejected as a criterion of what constitutes civil disobedience). When protesters do appeal to a shared sense of justice, they importantly indicate their fidelity to the existing political system. But one should not rule out the possibility that civil disobedience may sometimes be justified to shift the moral sense of the community, as in the examples of protests by vegetarians and pacifists, or to respond to stupidity and ineptitude, which parents may see as the dominant obstacle to having the traffic light installed.

Those considering whether civil disobedience is justified will need to consider the likely effects of their actions in accomplishing their objectives. They must also ask whether their actions will have a serious negative effect on political stability or will lead to reprisals against persons whose interests they represent but who do not endorse the tactics they employ.

No easy formula presents itself for determining whether any particular exercise of civil disobedience is justified. Moreover, since civil disobedience typically accompanies highly controversial social issues, even clarity of evaluation is difficult. Suppose ‘pro-life’ demonstrators illegally block access to an abortion facility, believing that actions of this kind can help alert people to the horrors of abortion, save unborn children and increase the likelihood that the law will become more restrictive. Is this civil disobedience justified? A person might think it is not: (1) because all opposition to women’s choice is not justified; or (2) because the protesters’ actions will disserve their own objectives; or (3) because civil disobedience is not warranted, even if one assumes the pro-life view is sound and the protest will have the hoped-for effect. Only the third view directly confronts the

demonstrators' judgment about the appropriateness of civil disobedience as a tactic. Too often, ostensible objections to tactics end up being reducible to disagreement with objectives. Careful, dispassionate analysis is required to focus upon genuine differences of views about the technique of civil disobedience itself.

See also: Gandhi, M.K.

KENT GREENAWALT

References and further reading

Adams, J.L. (1970) 'Civil Disobedience: Its Occasions and Limits', in J.R. Pennock and J.W. Chapman (eds) *Nomos X: Political and Legal Obligations*, N ew York: Atherton, 293-331.(A useful overall view.)

Bedau, H. (ed.) (1969) *Civil Disobedience: Theory and Practice*, Indianapolis, IN: Pegasus.(Contains important writings by theorists and by leaders of acts of civil disobedience, with an illuminating introductory essay.)

Fortas, A. (1968) *Concerning Dissent and Civil Disobedience*, New York: Signet Books.(Asserts strong limits on justifiable civil disobedience; popular nonphilosophical treatment.)

Greenawalt, K. (1987) *Conflicts of Law and Morality*, New York: Oxford University Press, 226-43; repr. Oxford: Clarendon Press, 1989.(Relates civil disobedience to more general themes about obligation to obey the law and expands on analysis in this essay.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (In chapter 6, pages 363-91, Rawls develops a powerful theory of how civil disobedience can be justified in a liberal democracy; asserts a restrictive view of what counts as civil disobedience and of when civil disobedience is justified.)

Singer, P. (1973) *Democracy and Disobedience*, Oxford: Clarendon Press.(Considers whether and when it is permissible to disobey decisions arrived at by democratic means.)

Zinn H. (1968) *Disobedience and Democracy: Nine Fallacies on Law and Order*, New York: Vint.(Expansive account of civil disobedience and when it is justified; nonphilosophical treatment.)

Civil society

In modern social and political philosophy civil society *has come to refer to a sphere of human activity and a set of institutions outside state or government. It embraces families, churches, voluntary associations and social movements. The contrast between civil society and state was first drawn by eighteenth-century liberals for the purpose of attacking absolutism. Originally the term civil society (in Aristotelian Greek,* politike koinonia*) referred to a political community of equal citizens who participate in ruling and being ruled*

In the twentieth century the separation of philosophy from social sciences, and the greatly expanded role of the state in economic and social life, have seemed to deprive the concept of both its intellectual home and its critical force. Yet, approaching the end of the century, the discourse of civil society is now enormously influential. What explains the concept's revival? Does it have any application in societies that are not constitutional democracies? From a normative point of view, what distinguishes civil society from both the state and the formal economy?

1 Conceptual history

The concept of *civil society* was introduced into modern European political philosophy through Latin translations of the Aristotelian Greek term, *politike koinonia*. In Aristotle's work, *politike koinonia* was defined as an ethical-political community of free and equal citizens who participate in ruling and being ruled under a legally defined system of public procedures and shared values. The term has since come to refer to very different organizations of the sphere regulated by public law - city republics, estate polities, dualistic structures of prince and country, the society of orders within the absolutist state (Riedel 1975). However, the Aristotelian identification of the political and the civil was maintained until the eighteenth century.

The early liberal polemical juxtaposition of society and state, originating in pre-revolutionary intellectual discourses critical of absolutism, decisively affected the meaning of civil society. Ever since, civil society has referred to a sphere of activity and set of institutions outside the state or government, representing an order of legitimacy that properly constructed political organizations ought to serve. The normative ideal of free and equal citizens comprising the body politic (civil society) was retained, but the body politic itself became differentiated in various ways from the state.

Nineteenth-century liberalism in all its versions maintained the society/state dualism. Yet the category of civil society developed primarily at the conservative, revolutionary and republican margins of liberal thought. Although he insisted upon the ethical dignity of the modern state, Hegel (§8) nevertheless conceived civil society as a complex framework of economic activities, legal institutions, intermediary associations and publics upon which modern political culture and identity were to rest (Hegel 1821). Drawing on Hegel's concept of the 'system of needs', Marx (§4) identified the civil with the bourgeois, and reconceptualized civil society from two points of view: as class society and as economic society. Marx also worked with a differentiated model of state and civil society but, unlike Hegel, he sought a revolutionary reunification of the civil and the political (Marx 1843). Finally, Tocqueville worked with a three-part model that differentiated, albeit unsystematically, between a civil society of economic and cultural associations and publics, a political society of local, provincial and national assemblies, and the administrative apparatus of the state (Tocqueville 1835-40).

The professionalization of philosophy and the emergence of differentiated social sciences in the twentieth century seemed to leave the concept of civil society without an intellectual home. Nevertheless it continued to play a role outside, and even within, the disciplines. In Marxian social philosophy, Gramsci (§3) and his followers made the concept of civil society central to their strategy of maintaining a Hegelianized Western Marxism distinct from Soviet Marxism (see Marxism, Western §2). Albeit without using the term, the Durkheimian-Parsonian tradition in sociology continued to develop the idea of intermediary associations and a community of free and equal citizens in order to promote normative integration and combat the atomizing tendencies of the modern state and the capitalist economy. In political theory, the idea of pluralistic limits on the centralizing impulse of the modern state continued to play a role, from Gierke to the British philosophical and the US empirical pluralists. Yet, with the emergence of the structuralist school in Marxism, and the decline of functionalism in sociology and of pluralism in political science, these efforts to translate the originally philosophical concept of civil society into the language of social science and theory apparently came to an end.

2 Revival of the discourse of civil society

Nevertheless, the concept of civil society began to reappear about fifteen years ago in the milieu of neo-Marxist critics of socialist authoritarianism. The pioneering theorists of this revival were Kolakowski, Mlynar, Vajda and Michnik in the East, Habermas, Lefort, Touraine and Bobbio in the West, Weffort, Cardoso and O'Donnell in the South. All knew the works of Hegel, the young Marx or Gramsci and thus had access to the concept of civil society and the idea of a state/society dichotomy. At an earlier stage, neo-Marxists had sought to deepen Marxian social philosophy by drawing upon non-Marxist theorists such as Weber, Simmel, Croce and Freud. They used concepts like alienation, fetishism, hegemony, reification and rationalization to explain and target the endurance of capitalism in the West, as well as new forms of domination and injustice in the East. The recent revival of the concept of civil society seemed to be an analogous move, since its presence in the young Marx justified a critical appropriation of the ideas of another series of non-Marxist thinkers, from Tocqueville to Carl Schmitt and Hannah Arendt (§1).This time, however, instead of invoking Marx to criticize orthodox Marxism, the theorists of civil society located the conceptual origins of communist totalitarianism in the young Marx's demand that the differentiation of state and civil society be overcome (see Totalitarianism). With this self-critique, neo-Marxists became 'post-Marxist'.

The historical success of the revival of the concept of civil society was due to its anticipation of and convergence with a new radical reformist strategy for the transformation of dictatorships, first in the East and subsequently in Latin America. The underlying project of 'societal self-organization' was geared towards rebuilding solidarities outside the authoritarian state. It appealed to a civil public sphere independent of all official, state or party-controlled communication. As a result, there is now vastly expanded discussion, not only in the East (most recently in China) and in the Middle East, but also in many Western countries with established civil societies, where the focus is on the strategic and normative possibilities of further democratizing existing democracies. Yet here too, the discussion has involved an analysis of the specific types of civic solidarities and civil institutions that secure social integration, provide public spaces for participation and reinforce the vibrant civic cultures which, in Robert Putnam's words, 'make democracy work' (1993).

Success in political discourse has led to revival in political science, and to a lesser extent in philosophy and sociology. It has also led to a more explicit critique of the concept.

3 Critiques and responses

On the methodological and analytical levels the concept of civil society is plagued by ambiguities. Even once it is differentiated from political and economic society, it remains unclear what civil society actually comprises. If it entails civil publics, voluntary associations and families, stabilized by basic rights (to association, assembly, expression, press, privacy), we come up against the heterogeneity of these institutions. There are publics in the differentiated domains of society - in science, art, law, religion - admission to which cannot be open to all. Voluntary associations vary in size, purpose, and political function. These in turn differ from social movements, which are also extremely diverse in form and orientation. Family types vary greatly. The question inevitably arises: what is the point of referring to this whole complex as civil society?

One could, of course, ask the same question about the institutions comprising the 'economy' or the 'state'. But in those cases the concepts of money and (political) power do seem to indicate the outlines of differentiation. The concept of communicative action coordination reconstructed by Habermas (1981) may solve the problem in principle for civil society. Yet on the institutional level, ambiguities remain. Surely political parliaments and corporate boards do not exclude deliberation any more than civil associations and publics exclude money or power. Do state, economic and civil institutions have organizing principles that differ in kind or only in degree?

One solution is to argue the distinctiveness of civil society from the normative point of view. The actors and institutions of the polity and the economy are directly oriented to the exercise of state power, to issues of collective goal setting, efficient administration and/or profit. They cannot afford to subordinate strategic or instrumental criteria to those patterns of normative integration that generate solidarities or collective identities and build civic culture and which are characteristic of civil society. Moreover, the publics in state and economic institutions are constrained by pressures of decision-making that preclude in principle the ideal of *open-ended* communication and contestation, an ideal that serves as a standard, however counterfactual, for interaction in civil and associational

publics.

However, even on the level of normative political theory the coherence of the concept of civil society remains in question. Indeed, each of the core elements of civil society has been articulated by *competing* and apparently incompatible traditions of political philosophy. Concerned with the danger of state absolutism, the liberal tradition has focused on individual rights (see Liberalism §3). The threat of anomie and social disintegration in modern society motivates the communitarian's stress on social solidarity, social integration and 'intermediary institutions' (see Community and communitarianism). Fearing corruption paternalism and/or political apathy, republican and democratic theorists focus on civic virtue, participation and public space (see Republicanism). The question remains: can the gaps between these distinct positions be papered over by a conception of civil society that tries to bring together the categories of rights, social solidarity, participation and public space?

The concept of civil society cannot serve as the centrepiece of a political philosophy that is construed either as an alternative to, or a synthesis of, these normative perspectives. Nor can it serve to harmonize competing political projects. However, it can help us to articulate the contested terrain that is at stake in the competing political theoretical traditions. From this perspective, the question is not whether the concept of civil society is still relevant to contemporary political theory, but rather, which version of civil society a particular theoretical approach presupposes. Once this is made clear, the concept can be useful to efforts to develop specific interpretations of rights, solidarities, participatory forms and public spaces that make liberal, communitarian, democratic and republican insights at least compatible. Nevertheless, it is bound to remain what Gallie (1955-6) once called an essentially contested concept.

JEAN L. COHEN

References and further reading

Aristotle (*c*. mid 4th century BC) *Politics*, trans. and with notes by E. Barker, London: Oxford University Press, 1968.(Earliest normative development of the concept of civil society.)

Cohen, J. (1982) *Class and Civil Society: The Limits of Marxian Critical Theory*, Amherst, MA: University of Massachusetts Press.(Thorough analysis of Marx's treatment of the concept of civil society.)

Cohen, J.L. and Arato, A. (1992) *Civil Society and Political Theory*, Cambridge, MA: MIT Press.(The most comprehensive theoretical statement to date.)

Gallie, W.B. (1955-6) 'Essentially Contested Concepts', *Proceedings of the Aristotelian Society* 56: 167-98. (Discusses what is meant by essentially contested concepts.)

Gramsci, A. (1929-35) *Prison Notebooks*, New York: International Publishers, 1971.(Best twentieth-century Marxist elaboration of the concept of civil society.)

Habermas, J. (1981) *The Theory of Communicative Action*, Boston, MA: Beacon Press, 2 vols.(Theoretical framework for the discursive concept of civil society.)

Hegel, G.W.F. (1821) *Grundlinien der Philosophie des Rechts*, trans. and with notes by T.M. Knox, *Hegel's Philosophy of Right*, New York: Oxford University Press, 1967.(Best nineteenth-century synthetic articulation of the concept.)

Keane, J. (ed.) (1988) *Civil Society and the State*, London: Verso.(Collection of seminal essays on the civil society.)

Marx, K. (1843) 'Contribution to the Critique of Hegel's *Philosophy of Right*', and 'On the Jewish Question', in *The Marx-Engels Reader*, ed. and with notes by R.C. Tucker, New York: W.W. Norton, 1978, 16-52.(Most important nineteenth-century critique of the bourgeois character of civil society.)

Parsons, T. (1971) *The System of Modern Societies*, Englewood Cliffs, NJ: Prentice Hall.(Best twentieth-century analysis of the integrating role of civil institutions.)

Putnam, R. (1993) *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton, NJ: Princeton University Press.(Best recent analysis of the importance of civil society to successful democratic polities.)

Riedel, M. (1975) 'Gesellschaft, bürgerliche' (Civil Society), in O. Brunner, W. Conze and R. Koselleck (eds) *Geschichtliche Grundbegriffe (Fundamental Philosophical Concepts)* , Stuttgart: Klett, vol. 2.(Best work on conceptual history of civil society.)

Tocqueville, A. de (1835-40) *Democracy in America*, Garden City, NY: Doubleday, 1969.(Important analytical distinction between civil and political society.)

Touraine, A. (1981) *The Voice and the Eye*, New York: Cambridge University Press.(Best work on the dynamic social movement aspect of civil society.)

Cixous, Hélène (1937-)

Hélène Cixous, a prolific French author born in Algeria, works between poetry and philosophy. She is part of a larger intellectual community in France that, since the 1960s, has sought a critique of the Western (male) subject, claiming that the 'metaphysical' notion of the subject has for three centuries contributed to the repression of nature, women and other cultures by construing human existence in terms of the separation of mind from body, and more generally of concept from metaphor.

Influenced by Nietzsche and Freud, Cixous privileges the artistic and poetic but all her work has philosophical underpinnings, especially those proposed by Derrida. Like his, her thought champions notions of difference, multiplicity and life over identity, univocity and death. She seeks to displace the unified, narcissistic (male) subject, which in her view is on the side of death. Cixous is also importantly influenced by her critical reception of Hegel and Heidegger.

Though she declares herself primarily a poet and a thinker, Cixous has kept close ties with the academy and has undertaken a variety of political activities. After helping to found, in 1968, the experimental University of Paris VIII at Vincennes - an institution both politically controversial and intellectually scintillating - she established there, in 1974, one of France's few Centres for Women's Studies. Her seminars have attracted students from around the world and have become, not merely intellectual experiences, but spiritual training grounds. Her struggle for a different approach to the other and for giving a voice to others has led her to espouse many political causes. But politics for Cixous are, like philosophy, never separated from poetry: from the invention of new languages without which, in her words, no social change can come about.

The critique of a subject identified as 'male' leads, from Cixous' earliest work on, to broadly psychoanalytical concerns. Her acute perceptions of others' psychic 'vibrations' has made her an excellent reader and critic, as evidenced in the early essays on Freud, Joyce, Kleist, and Poe collected in *Prénoms de personne* (1974). But she has also subjected psychoanalysis to intense, deconstructive scrutiny, in which supposed psychoanalytic 'truths' are rejected in favour of open-ended, personal narratives. Her avant-garde writings of the late 1960s and early 1970s conceived of a subject that is no longer masterful and autonomous but always in dialogue - or transference - with other textual or living subjects or voices. In this conception, the subject constructs itself, or is actively 'born' and 'reborn', at the interstices of various texts and history. The philosophical underpinning for this is a critique of Hegel and his brand of dialectical reasoning, as exemplified in the account of the master-slave relation in the *Phenomenology of Spirit* (see Hegel, G.W.F §5) - a relation which Cixous sees as the basis of much of Western dualistic thinking, including the existential humanism of Sartre and de Beauvoir.

In one of her most widely read books, *La Jeune Née* (*The Newly Born Woman*) (1975) co-authored with Catherine Clément, Cixous began a lengthy detour through the cause of women. In this phase of her work, she 'deconstructively' criticized symbolic practices that organize culture through hierarchically ordered oppositions such as nature/culture, speech/writing and man/woman (see Deconstruction). These oppositions, themselves produced historically, were to be displaced into mere differences. The simultaneous exclusion in metaphysics of the terms 'writing' and 'woman', argued for by Derrida, enabled her to coin the productive but controversial expression, 'feminine writing' (*écriture feminine*). This phrase, read in the derivative sense of a woman writing, was used to raise consciousness. But in a Derridean sense, writing is always related to and occurs within an (unconscious) scene. Thus, Cixous argued that the body is never a brute given but is always already encoded, and disarms accusations of essentialism. '*I*' is more than one, for it is in constant dialogue with others, and any identity is imposed from the outside. Cixous destabilized the fixed sexual identities imposed by a reigning cultural order through her concern with bisexuality.

In Cixous' enterprise of feminine writing, philosophy intersects with anthropology as well as with psychoanalysis. Bataille's reading of Mauss's classic essay, *Essai sur le don (The Gift)* (1925) prompted her to look at cultures that practise modes of exchange different from those of Western retention and accumulation (such as Kwakiutl tribes, known for their ritual of *potlatch*). This led her poetically to invent ways of giving and receiving unconditioned by a restricted (male) economy infused with feelings of guilt and debt. She argued that women, because of their marginal status in society, practise the art of the gift more often than men, who are caught in scenarios of

castration. ‘Feminine writing’ itself therefore exemplified an art of giving at the level of content and form. Her texts, speaking of generosity, exemplify it through an excess of meaning. In this, ‘feminine’ serves to designate, not an enduring attribute, but a transitional step toward other modes of exchanging where it will be replaced with other adjectives.

In the early 1980s, Cixous’ critique of the Hegelian dialectic of recognition and sublation gave way to other approaches. Her reading of the Brazilian novelist, Clarice Lispector, first published in the review *Poétique* in 1979, inaugurated a shift toward Heidegger’s philosophy of language. With it came a change in style. Emphasis on surface gave way to depth; earlier ‘explosions of the subject’ were discarded in favour of a return to ‘man’ (*l’homme*). Heidegger’s notion of poetry, a *technē* that he favours over instrumental technology of repression, was welcomed. In this stage of her work, Cixous combined her devotion to writing more explicitly with philosophy, as a new brand of ‘writing-thinking’ enabled her to work through personal, cultural and historical issues. Her texts of this period are concerned with philosophical questions such as, who leaves and who arrives where? How does one give and receive? How does one have what one has? Her continued dialogue with Derrida also led her to deal with questions similar to his, such as the promise and the secret.

In her recent work, the emphasis on individual pleasure (*jouissance*) fades. Her concern with the cause of women, though not abandoned, opens out to others. This is evident in her writings for the theatre. Her plays on Cambodia (1985) and India (1987), on apartheid and the gulag (*Manne*, 1988) broach issues of history and collective voice in terms of neo-colonialism. All these plays and texts oppose a lofty sphere of combat for truth to a low sphere of political battles and sheer power. In this period Cixous has continued her personal meditations, now focusing on her childhood and her Jewish origins. She is also returning more to the classroom, navigating deftly between national literatures and philosophies, between genres and media.

How such a hybrid writing-thinking fares either as poetry or as philosophy remains to be seen. To date, all of Cixous’ production, whatever its genre, has been critical of metaphysical closure and its repressive conceptual and syntactic structures. Her reading and writing of differences performed through a variety of topics undo totalizing systems based on inclusion, exclusion and negation - or, in psychoanalytic terms, on castration. Yet birth - no longer just anatomical and natural - and *sorties*, ways out, do not simply lead to an outside. They must, for Cixous, be continually re-enacted in an ongoing writing-thinking at the interstices of personal, cultural and social history.

See also: Derrida, J.; Feminism and psychoanalysis

VERENA ANDERMATT CONLEY

List of works

Cixous, H. (1974) *Prénoms de personne*, Paris: Éditions du Seuil.(Collection of early essays on Freud, Joyce, Kleist and Poe.)

Cixous, H. (1975) *La Jeune Née*, with Catherine Clément, Paris: Christian Bourgois; trans. B. Wing, *The Newly Born Woman*, Minneapolis, MN: University of Minnesota Press, 1986.(A militant essay that focuses on the previous exclusion of women from cultural discourses and invents new cultural practices through writing.)

Cixous, H. (1979) ‘L’Approche de Clarice Lispector’, *Poétique* 40; trans. D. Jensen, in *Coming to Writing and Other Essays*, Cambridge, MA: Harvard University Press.(A reading of Lispector’s ‘feminine writing’. A marked shift toward Heidegger in its general use of language and the use of a notion of approach.)

Cixous, H. (1985) *L’Histoire terrible mais inachevée de Norodom Sihanouk, roi du Cambodge*, Paris: Théâtre du Soleil; trans. J. MacCannell, J. Pike and L. Groth, *The Terrible but Unfinished Story of Norodom Sihanouk, King of Cambodia*, Lincoln, NE: University of Nebraska Press, 1994.(King Sihanouk’s generous attempt to save his country from political fanatics and technocrats out to destroy it.)

Cixous, H. (1987) *L’Indiade ou l’Inde de leurs rêves*, Paris: Théâtre du Soleil.(Gandhi’s politics of renunciation, non-violence and ‘maternal’ love triumph over the adversaries’ cravings for personal power that lead to death and destruction.)

Cixous, H. (1988) *Manne*, Paris: Éditions des Femmes; trans. C. MacGillivray, *Manna*, Minneapolis, MN: University of Minnesota Press.(A denunciation of apartheid and the gulag as tribute to the courage of Nelson and Winnie Mandela, Osip and Nadezhda Mandelstam, and the Russian poetess Anna Akhmatova. A study of the relation between poetry and history.)

References and further reading

All these works are expository, non-technical and accessible.
Conley, V.A. (1984) *Hélène Cixous: Writing the Feminine*, Lincoln, NE and London: University of Nebraska Press, expanded edn, 1991.(A critical, introductory study especially of Cixous' early writings. Contains interview.)
MacGillivray, C. (1994) 'Introduction' to *Manna*, Minneapolis, MN: University of Minnesota Press.(A helpful, close reading of Cixous' textual strategies.)
Mauss, M. (1925) *Essai sur le don*, Paris: Alcan, L'année sociologique; trans. I. Cunnison, *The Gift*, New York: W.W. Norton, 1967.(The first systematic study of the custom of exchanging gifts as part of all human, personal relationships between individuals and groups.)
Moi, T. (1985) 'Hélène Cixous: An Imaginary Utopia', in *Sexual/Textual Politics*, London: Methuen.(A critical reading of Cixous' 'essentialism'.)
Sellers, S. (1988) *Writing Differences*, Milton Keynes: Open University Press.(A collection of textual readings by Cixous and her seminar participants. Includes interview.)
Shiach, M. (1992) *Hélène Cixous: The Politics of Writing*, London and New York: Routledge.(Focuses on the relations between feminism and writing. Deals with the theatre.)

Clandestine literature

Clandestine philosophical (anti-Christian) literature of the seventeenth century circulated in manuscript form until its publication by the philosophes *in the later eighteenth century. Since research began, the list of texts has grown and now includes some 260 titles which cover the classical heritage of the Renaissance, the works of La Peyrère and Cyrano, the influence of Spinoza, the growth of rationalism which accompanied the splintering of Protestant churches and sects, and the development of the deist debate in Britain.*

Clandestine philosophical literature has been considered a serious object of study since 1912 when G. Lanson discovered a number of manuscript copies of 'philosophical' (that is, anti-Christian) texts in French public libraries. The search for such texts was pursued by I.O. Wade (1938), stimulated by N.L. Torrey's work on Voltaire's library in St Petersburg. Wade thus put forward a list of 102 philosophical texts distributed between 1700 and 1750. These results have been extended by recent research, in particular by M. Benítez, and some 260 titles distributed in Europe from the sixteenth to the eighteenth century are now known. New discoveries will certainly be made and new texts revealed, as suggested by the catalogues of private libraries of that period. This literature raises questions for philosophy, intellectual history, apologetics, printing and the book-trade; it constitutes the roots of anti-Christian thought in the eighteenth century, and the texts of the prestigious *philosophes* of the Enlightenment must now be related to this intellectual context.

The development of this field of study has raised new problems. Research now covers the sixteenth, seventeenth and eighteenth centuries, and there is great risk of anachronism in the interpretation of early texts in relation to later events. Nevertheless, currents of thought can be defined which act as signposts from one century to the next: the clandestine philosophers of the eighteenth century regarded themselves as inheritors of a long anti-Christian and anti-ecclesiastical tradition.

In France, this anti-Catholic literature is not to be confused with Protestant, Socinian, Anabaptist or even Jansenist polemics, but the multiplication of sects and churches had an obvious effect on clandestine authors, who represent all possible positions in relation to Catholic orthodoxy (see Socinianism). In this sense, the barrier between orthodoxy and heterodoxy becomes vague and confusing. Recent research on Spinoza and the Dutch 'collegiants' (such as Bredenburg) has illustrated how philosophical rationalism was enriched by religious thought. A symmetrical development of philosophical ideas within Christian apologetics is evidence of prevalent uncertainty towards traditional barriers between deism and Christianity, belief and unbelief. The 'inner light' and the philosophical 'Enlightenment' enriched each other.

Similar relations can be discerned in the distribution of clandestine literature. The archives of the Bastille yield few names of clandestine philosophers, but reveal the full breadth of unorthodoxy and the full weight of the censorship prevalent in the seventeenth century. The successive waves of Huguenots and Jansenists are easy to discern: throughout the classical age book-pedlars and printers of political, religious or obscene pamphlets are imprisoned; the diffusion of philosophical ideas thus follows in the footsteps of other unorthodox currents of thought. Philosophical writings are distributed by specialists in political satire, in Marrano, Protestant and Jansenist propaganda. The history of clandestine ideas in Europe is thus closely linked to the history of the book-trade and censorship.

An important aspect of 'scribal publication' in the seventeenth century has been underlined by H. Love (1993). The manuscripts, of course, preceded the printed work, but they survived beyond the printing of the text (or rather, one version of the text) and continued to be amended and altered by subsequent scribes: in this sense, they are evidence of the evolution of philosophical ideas. Clandestine literature can thus be studied from the point of view of textual genetics: variants are to be interpreted not as mistakes, not as lessons unfaithful to the original, but as indications of a living and changing philosophy.

Clandestine philosophical literature reflects the main currents of intellectual history and reveals their significance for anti-Christian thought. Thus, in 1659 an enormous compendium of classical thought was constituted under the title *Theophrastus redivivus*, presenting the classical heritage in the form of extracts from ancient authors on a variety of crucial themes: the existence of the gods, the nature of the material world, the status of religion, the

existence of the soul, the existence of hell, the consequences of physical death, the precepts of life according to Nature and so on. It is an anthology of free thought concocted from Ancient texts. At the same period, various versions of Cyrano de Bergerac's *L'Autre monde* were distributed, and Isaac La Peyrère composed his vast trilogy which leads us from the 'pre-Adamites' to the conversion of the Jews. A number of manuscripts reflect debates around the 'new philosophy' of Descartes and a dozen texts deal with aspects of Spinoza's philosophy. Clandestine philosophical literature of the end of the seventeenth century thus developed in the intellectual context of the rivalry between the 'systems' of Descartes, Gassendi, Hobbes, Malebranche, Spinoza and Locke - and soon of Leibniz. Machiavelli, Naudé, La Mothe Le Vayer, Hobbes, Vanini were systematically exploited, while Montaigne and Bayle were inexhaustible sources of plagiarism and imitation. The English context of debates around Deism weighed heavily on the evolution of French thought: Toland, Collins, Tindal, Mandeville, Blount, Middleton, Bolingbroke and Woolston were soon to be translated, as also were the Marranos Isaak de Troki and Orobio de Castro (see Deism).

Clandestine philosophical literature can thus be interpreted within the context of the history of philosophical scepticism recently traced by R.H. Popkin and others. Central to this history in the seventeenth century was the rivalry between the Pyrrhonism inherited from Montaigne by La Mothe Le Vayer, Gassendi and Pascal and the rationalism developed by Descartes, Malebranche and Leibniz (see Pyrrhonism; Rationalism). The philosophy of the Enlightenment can thus be shown to derive from a crisis in Christian philosophy of the seventeenth century.

See also: Libertins; Pascal, B.

ANTONY McKENNA

List of works

Bergerac, C. de (1657) *L'Autre monde ou les estats et empires de la lune (The other world or the estates and empires of the moon)*, ed. M. Alcover, Paris: Champion, 1977; repr. Paris: Klincksiek, 1996.(Well-known imaginary voyage by a disciple of Gassendi.)

La Peyrère, I. (1643) *Du Rappel des Juifs (Of the recalling of the Jews)*; *Systema theologicum ex Præadamitarum hypothesi (Theological system drawn from the hypothesis of the pre-Adamites)*, 1655.(Original theological system built on the re-interpretation of the Old Testament as being the history of the Jewish nation only, and not of the whole of humanity.)

Theophrastus redivivus (1659), ed. G. Canziani and G. Paganini, Florence: Franco Angeli editore, 1981-2.(A compendium of classical thought in the form of extracts from ancient authors on a variety of crucial themes.)

References and further reading

Aylmer, G.E. (1982) 'Unbelief in 17th-century England', in D. Pennington and K. Thomas (eds), *Puritans and Revolutionaries: Essays in 17th-century History presented to Christopher Hill*, Oxford: O.U.P., 22-46. (Historical evidence to be read in the context of L. Febvre's well-known 1942 work *The Problem of Unbelief in the Sixteenth Century: The Religion of Rabelais*, trans. B. Gottlieb, Cambridge, MA: Harvard University Press, 1982.)

Benítez, M. (1988) 'Matériaux pour un inventaire des manuscrits philosophiques clandestins des XVIIe et XVIIIe siècles', *Rivista di storia della filosofia* 3: 501-31.(An inventory of 148 clandestine manuscripts and bibliography of recent work.)

Benítez, M. (1995) *La Face cachée des Lumières. Recherches sur les manuscrits philosophiques clandestins de l'Age classique*, Oxford: The Voltaire Foundation; Paris: Universitas.(A new inventory of some 260 titles of clandestine works, an introduction dealing with the definition of clandestine philosophy and a collection of articles on materialism and pantheism in clandestine works.)

Berman, D. (1988) *The History of Atheism in England: from Hobbes to Russell*, London: Croom Helm.(Good general account covering Hobbes and the early English Deists.)

Bloch, O. (1982) *Le Matérialisme du XVIIIe siècle et la littérature clandestine*, Paris: Vrin.(Collection of articles dealing with a variety of clandestine manuscripts.)

Canziani, G. (ed.) (1994) *Filosofia e religione nella litteratura clandestina (secoli XVII e XVIII)*, Milan: Franco Angeli.(Collection of articles on clandestine literature, with two editions of clandestine texts.)

Gregory, T., Paganini, G., Canziani, G., Pompeo Faracovi, O. and Pastine, D. (eds) (1981) *Ricerche su letteratura libertina e letteratura clandestina nel Seicento*, Florence: La Nuova Italia.(Collection of articles dealing with seventeenth-century free thought.)

Kors, A.C. (1990) *Atheism in France, 1650-1729, vol. 1: The Orthodox sources of disbelief*, Princeton, NJ: Princeton University Press, 1990; vol. 2: *Naturalism and Disbelief*, Princeton, NJ: Princeton University Press, 1997.(An ambitious analysis of causes of unbelief, traced back to contradictions and uncertainties in Christian thought and followed through to the effects in clandestine literature.)

La Lettre clandestine: Bulletin d'information sur la littérature philosophique clandestine à l'âge classique (1992-), Paris: Saint-Etienne.(This journal, directed by O. Bloch and A. McKenna, appears annually and provides a current bibliography.)

Lanson, G. (1912) 'Questions diverses sur l'histoire de l'esprit philosophique en France avant 1750', *Révue d'histoire littéraire de la France* 19: 1-29, 293-317. (Pioneer work in the discovery of clandestine manuscripts.)

Love, H. (1993) *Scribal Publication in Seventeenth-Century England*, Oxford: Clarendon Press.(Details the prevalence of scribal publications in the seventeenth century in all fields of thought and the conclusions to be drawn by the modern editor from this method of diffusion.)

McKenna, A. and Mothu, A. (eds) (1997) *La Littérature clandestine: Actes du colloque de Saint-Etienne, octobre 1993*, Oxford: The Voltaire Foundation; Paris, Universitas. (Papers from a conference on clandestine literature: a variety of manuscripts are analysed.)

Popkin R. (1979) *The History of Scepticism from Erasmus to Spinoza*, Berkeley and Los Angeles, CA, and London: University of California Press.(The first general account of the influence of Pyrrhonism in the seventeenth century: Pyrrho, Sextus Empiricus, Montaigne and the early editions and translations; reaction of the 'libertins érudits' and Counter-Reformation, leading first to Descartes', then to Spinoza's anti-Pyrrhonist rationalism.)

Popkin R. (1990) *The Third Force in 17th century Philosophy*, Leiden: Brill.(A collection of essays dealing with Millenarianism in the seventeenth century and its influence on contemporary philosophical thought.)

Popkin, R. and Vanderjagt, A. (1993) *Scepticism and Irreligion in the Seventeenth and Eighteenth centuries*, Leiden: Brill.(A collection of essays detailing recent research on the history of free thought.)

Torrey, N.L. (1930) *Voltaire and the English Deists*, Yale University Press; repr. Oxford: Marston Press, 1963. (First systematic study of this important question, through Voltaire's library in Leningrad.)

Wade, I.O. (1938) *The Clandestine Organisation and Diffusion of Philosophic ideas in France from 1700 to 1750*, Princeton, NJ: Princeton University Press; New York: Octagon, 1967.(A first inventory of 102 clandestine manuscript titles in France.)

Clarembald of Arras (1110/20-*c*.1187)

A teacher of philosophy at Laon and commentator on Boethius, Clarembald was a product of the School of Chartres. His principal philosophical work, the Tractatulus (Short Treatise on Genesis), is an attempt to link Platonic theories of creation with the account given in Scripture, and includes a discussion of the nature of form and matter.

Clarembald most likely was born between 1110 and 1120. He studied at Paris under Thierry of Chartres and Hugh of St Victor in the late 1130s, and directed the schools of Laon for a short while around 1160. Most of his life, however, was an active one: he was provost and archdeacon of the diocese of Arras, and was chaplain in Laon during the last years of his life.

Clarembald's principal philosophic work links him to the school of Chartres. He wrote an introductory letter to Thierry of Chartres' *De sex dierum operibus (On the Works of the Six Days of Creation)* and a *Tractatulus (Short Treatise on Genesis)*. In the introductory letter, Clarembald claims credit only for the special effort he has made in his *Tractatulus* 'to reconcile many views of the philosophers with the Christian truth so that the word of Scripture might receive strength and protection even from its adversaries.'

The *Tractatulus* thus attempts to reconcile the portion of Plato's *Timaeus* that was translated and commented on by Calcidius with the message of Genesis as interpreted by Ambrose, Augustine, Isidore of Seville and Bede. The scriptural account, according to Clarembald, does not conflict with philosophical explanations of order in the world. The opening words of Genesis, 'In the beginning God created heaven and earth', bring out the deeper meaning of philosophical arguments. They provide an introduction, even for pagans and non-believers, to a knowledge of the Creator. The world (heaven and earth) is made up of contrary elements: hot ones and cold ones, moist ones and dry ones. It is either nature, chance or a maker that has joined together these opposed elements. Nature cannot produce such an order, since it unites only things that are alike; neither can chance produce such a universe, since chance presupposes already existing causes that come together to produce an unexpected happening. These causes thus would be prior to chance, the agent that allegedly joins them together. The only cause capable of uniting these contrary elements into an ordered whole is their maker, the Creator who brought these elements into existence.

The Genesis account adds that 'the earth was void and empty'. Augustine (§8) speaks of the primordial matter as 'formlessness', or the capacity to receive a form. Plato (§16) describes the same formless matter as that which exists between nothingness and substance. His pupil Aristotle (§12) at times calls this *aptitudo* (ability), and at times *carentia* (lack of form). We grasp it, according to Clarembald, by removing forms. When we remove from a formed object its forms, then nothing remains but the potential to receive these forms. If we take away the form of a brass statue, for example, then only the brass is left; if we strip the brass of its properties, then only the element of earth remains. If we remove the characteristics of earth (dryness and frigidity), then that which remains is not nothing, but formless matter. Such formless primordial matter is called by the philosophers *possibilitas* (potency), since it is capable of receiving all forms. When it receives such forms, then it becomes *possibilitas definita* (specified possibility) (see Matter).

Formless matter does not receive its forms from itself, but from what the philosophers call *necessitas absoluta* (absolute necessity). 'Absolute necessity', according to Clarembald, is what Boethius calls 'Divine Providence', what Augustine names 'Word' or 'Divine Wisdom' and what Pythagoras describes as 'One'. 'Divine Wisdom' implants *rationes seminales* (seminal reasons) in the formless matter and it is by means of them that things reproduce other things of the same kind. These seminal reasons, or what the philosophers call *similitudines nascendi* (reproductive likenesses), account for the normal course of nature (see Augustine).

Clarembald's earlier commentaries on Boethius' *De trinitate (On the Trinity)* and *De hebdomadibus (On the Groups of Seven)* already make many of the philosophic points he develops in the *Tractatulus*, but they also reveal his view of the nature of a properly conceived and formulated *quaestio* (question) (Commentary on *De trinitate* Prol. 5-14), his theory of the categories (Commentary on *De trinitate* 4, 1-46), the distinction between dialectical, demonstrative and sophistical syllogisms (Commentary on *De hebdomadibus* i, 1-2), and his interpretation of

Boethius' different levels of abstraction (Commentary on *De trinitate* 2, 17-9). Although strongly influenced by Thierry of Chartres and Hugh of St Victor, Clarembald was on theological issues also a strong critic of Peter Abelard and a more cautious opponent of Gilbert of Poitiers.

See also: Chartres, School of; Thierry of Chartres

STEPHEN F. BROWN

List of works

Clarembald of Arras (*c.*1157) *Tractatus super librum Boetii De trinitate (Exposition of Boethius' De trinitate)*, in N.M. Häring, *Life and Works of Clarembald of Arras*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1965.(Contains his description of a correctly formulated *quaestio* or question, and his interpretation of Aristotle's ten categories.)

(*c.*1158) *Expositio super librum Boetii De hebdomadibus (Commentary on Boethius' De hebdomadibus)*, in N.M. Häring, *Life and Works of Clarembald of Arras*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1965. (Contains his understanding of dialectical, demonstrative and sophistical syllogisms.)

Clarembald of Arras (*c.*1160s) *Tractatulus (Short Treatise on Genesis)*, in N.M. Häring, *Life and Works of Clarembald of Arras*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1965.(Clarembald's short commentary on Genesis.)

References and further reading

Häring, N.M. (1955) 'The Creation and Creator of the World according to Thierry of Chartres and Clarenbaldus of Arras', *Archives d'histoire doctrinale et litteraire du moyen-âge* 22: 137-216.(Clarembald's account of creation.)

Clarke, Samuel (1675-1729)

Regarded in his lifetime along with Locke as the leading English philosopher, Clarke was best known in his role as an advocate of a thoroughgoing natural theology and as a defender of Newtonianism, most notably in his famous correspondence with Leibniz. His natural theology was set out in his Boyle lectures of 1704 and 1705, but it left little room for revelation, and endeared him to neither side in the quarrel between deists and orthodox Anglicans. A staunch proponent of Newtonian natural philosophy, he defended it against criticisms of its notions of gravity and absolute space.

1 Natural theology

Samuel Clarke was born at Norwich, England, and studied at Gonville and Caius College, Cambridge, where he remained until 1700. He was a fine classicist and while still a student he published a Latin translation of the standard Cartesian natural philosophy textbook of the time, Rohault's *Traité de Physique*, which replaced the inferior translation of Théophile Bonnet (see Rohault, J.). In 1697 he was converted to the cause of Arianism, and from this time until his death in 1729 - after many years as rector of St James's, Westminster and chaplain to Queen Anne - he devoted his main energies to the development of a natural theology.

Among the key theological issues within the Anglican Church at this time were an internal dispute between defenders of the Athanasian Creed (which established the notion of the Trinity) and Arians, and an external dispute between Anglicans and deists. Arianism or unitarianism had significant support within rationalist theology in England in the late seventeenth and early eighteenth centuries. Deism - the notion that there is a natural religion with precedence over any revealed religion - had become a significant force with the posthumous publication of the works of Charles Blount (1654-93) in 1695, and reached its apogee in the 1730s. English deists, unlike French Enlightenment writers, generally considered themselves Christians, with the result that the distance between deists and writers like Clarke, who defended natural theology at the expense of revelation, was sometimes very small.

Clarke's aim was to reconstruct religion and ethics on the basis of natural philosophy and natural theology. His proof of the existence of God is a version of the cosmological argument. Its core is that if something exists now, something always must have existed because things cannot come into existence from nothing; but what it is that always has existed must be 'self-existent', that is, it must depend on nothing else for its existence; and what is self-existent is necessarily existent. All these steps are questionable but, even if supposed valid, the argument establishes a very abstract God rather than a personal God who is moral, just and truthful. Consequently Clarke needed to bridge the gap between this God and the God of Christianity (even in its unitarian version).

His account of God harbours a deep and familiar problem. On the one hand, God is completely free, and orders the world in a way that is constrained by nothing, for nothing could constrain such a perfect being. On the other hand, goodness and truth are not arbitrary, and in particular God's action must reflect natural standards of truth and goodness rather than instantiate such standards arbitrarily. Clarke attempts to reconcile these principles by means of a doctrine of the 'fitness of things', whereby nature has a moral aspect which mirrors its physical aspect, moral and physical aspects both being knowable by reason. The idea is that any rational being will guide its conduct in terms of these moral principles. Since God is completely rational, he follows them of necessity, but we are also influenced by passions which act against reason and cause us to behave immorally on occasion, and it is from this that the need for established religions arises.

2 Doctrine of the soul

The Boyle Lectures were set up - on a bequest of £50 per annum from Boyle - to counter deism, Hobbesianism, atheism and other unorthodox views, and in his Boyle Lectures for 1705 Clarke attacked what he perceived to be the reductionism of Hobbesian and Spinozistic accounts of mind (see Deism §2; Hobbes, T. §3; Spinoza, B. de §§5-6). The problem had come to the fore because of Locke's passing statement that it is not inconceivable that God could 'superadd to matter the faculty of thinking' (see Locke, J. §5). Like other critics of Locke, Clarke believed that this was tantamount to materialism. His response was to deny that perception and intelligence can derive from matter - either reductively or in virtue of a combination of matter and motion - on the grounds that only what possesses a perfection can cause or communicate it: something that lacks the perfection of thought

cannot produce this perfection in something else.

Two critics took Clarke to task - Henry Dodwell and Anthony Collins. Collins, in particular, accuses Clarke of failing to distinguish between the powers of the parts of matter and the powers of organized combinations of parts: the latter powers are different from the former and serve to explain thought and action. Clarke denies that there can be powers in the whole that are not powers of the parts, distinguishing various kinds of genuine from non-genuine properties of things: colours, for example, seem to be properties of the whole but not the part, yet they are in fact not genuine properties of the object at all. Collins accuses Clarke of asserting that the whole cannot have powers lacking in its parts, whereas he should be demonstrating it. The question is not simply one of onus of proof, however, for Clarke himself subsequently admits that something can have qualities or powers, such as roundness, which are not present in its parts. But consciousness, he maintains, is not like this: it is a power which must inhere in the parts if it is to be present at all (see Collins, A. §§3-4).

The dispute is inconclusive, but it hard to avoid the conclusion that Clarke set off on the wrong foot in denying that whole and parts can have different properties, and the claim that consciousness must be in all of the parts is not supported by any compelling argument.

3 Defence of Newton

Clarke is often praised for the way in which his notes to his translation of Rohault undermine Rohault's Cartesianism in favour of a Newtonian account. But in fact the notes to the first edition are very uncritical, accepting the vortex theory and making no mention of gravitation. It was only in later editions that the notes took on a Newtonian edge, and by that time Newton's *Principia* had a wide circulation. Clarke was, however, an able defender of Newtonianism in the 1700s, being the first to distinguish clearly between momentum and kinetic energy, and he was chosen to respond to Leibniz's accusation that the ideas of Newton and Locke were responsible for the decline of religion in England (see Leibniz, G.W. §11).

Some of the questions at issue between Clarke and Leibniz are on points of interpretation, as in Leibniz's rejection of Newton's claim that space is simply God's *sensorium*. Leibniz points out that if God needed an organ to perceive things, those things could not depend on him, and Clarke replies by construing Newton's doctrine as meaning that God is omnipresent and so, far from requiring sense organs, has unmediated perceptual grasp. Other disputes involve substantive questions in natural philosophy, and the fundamentals of Leibniz's philosophy - notably his principles of sufficient reason and the identity of indiscernibles - are called into question. Many of the issues in the dispute are essentially continuations of the clash between Cartesian and Newtonian natural philosophy. Descartes had provided a model in which planets were carried around a centrally rotating sun by the swirling motion of the ether around this sun, an account which dispensed with any need for action at a distance. This is essentially the account that Leibniz defends, and Clarke defends both the existence of a void (which undermines the idea of a swirling ether) and action at a distance, which he points out has been shown to exist, even if its causes have not been discovered.

Other questions in dispute raise new problems, and a key issue is the nature of space and time. Leibniz considered these to be relations, and hence to be relative, whereas Clarke considered them to be real and absolute. Newton had considered that, although we cannot tell of any inertial state whether it is a state of rest or a state of uniform rectilinear motion, there is a way in which we can detect the existence of absolute space, and he believed that the concave deformation of the surface of water in a rotating bucket can only be explained by a rotation relative to absolute space (see Newton, I. §3). Leibniz did not accept that the idea of absolute space made sense, for in such a space no part is different from any other, and there would be no reason for God to create things in one place rather than another. Clarke's response is effectively that God does not need sufficient reason: God creates through an act of will. Moreover, points in space are not merely relational: they have real properties that distinguish them from one another, as the bucket experiment indicates. Clarke is in a strong position here, and the challenge to Leibniz is to indicate how differences in inert forces can be explained except by reference to space and time.

STEPHEN GAUKROGER

List of Works

Clarke, S. (1738) *The Works of Samuel Clarke*, London: John & Paul Knapton, 4 vols; repr. New York: Garland, 1978.(This edition contains everything of any philosophical or theological significance by Clarke - except his

notes on Rohault - and provides useful biographical material.)
Clarke, S. (ed.) (1697) *Jacobi Rohaulti Physica, Latine reddit et annotatiunculis quibusdam illustravit S. Clarke…*, London: Jacob Knapton.(Latin translation of Rohault's *Physics* with notes by Clarke; subsequent editions appeared in 1708 and 1718).

References and further reading

Alexander, H.G. (1956) *The Leibniz-Clarke Correspondence*, Manchester: Manchester University Press.(Contains a useful introduction and notes to the translation of this exchange.)
Blount, C. (1695) *The Miscellaneous Works of Charles Blount*, London, 1697; repr. New York: Garland, 1979. (Blount's attacks, in the 1680s on the necessity for miracles, and his call for a natural religion devised by lay people, are generally seen as marking the beginnings of deism in England.)
Priestly, F.E.L. (1970) 'The Clarke-Leibniz Controversy', in R.W. Butts and J.W. Davis (eds.), *The Methodological Heritage of Newton*, Oxford: Blackwell.(Clear account of the controversy.)
Ray, C. (1991) *Time, Space and Philosophy*, London: Routledge.(Chapter 5 looks at the Leibniz-Clarke controversy with an eye to its implications for later developments in physical theory.)
Reventlow, H.G. (1985) *The Authority of the Bible and the Rise of the Modern World*, Philadelphia, PA: Fortress Press, II, ch. 3.(A good account of the context of Clarke's natural theology.)
Yolton, J. (1984) *Thinking Matter*, Oxford: Blackwell.(Chapter 2 deals with the Clarke-Collins dispute.)

Clauberg, Johannes (1622-65)

Johannes Clauberg was a member of the Cartesian school in the years immediately following Descartes' death, and is extremely important as an early expositor of Descartes and a spirited defender of the Cartesian philosophy. His writings include a number of direct commentaries on Descartes' published works, as well as extensions and elaborations of the master's views.

Clauberg was born to Protestant parents in Solingen, Westphalia. He studied at Bremen, then at Groningen, where the Cartesian Tobias Andraea and the anti-Cartesian Martinus Schoock both taught. In 1646 he visited France and England, and in the late 1640s he studied at the University of Leiden, where he attended the lectures of the Cartesian Johannes de Raey. Clauberg's *Elementa philosophiae sive ontosophia*, published in 1647 although probably written in 1645, was a defence of Aristotelian metaphysics against Ramist attacks, and contains nothing recognizably Cartesian (see Ramus, P.). While it is not impossible that he met Descartes himself, he was probably converted to Cartesianism by either Andraea or de Raey before April 1648, when he transcribed the notes of a conversation between Descartes and the young theologian Frans Burman. In 1649, at the request of Count Louis Henry of Nassau, he was appointed professor of philosophy and theology at the University of Herborn in Germany, where he taught Cartesianism for two years in the face of considerable opposition. The count tried to stabilize Clauberg's position by writing to some Dutch universities for their opinions on the new Cartesian philosophy, and while there were some positive responses, they were too few and too late to make a difference. In 1651, along with the Cartesian Christopher Wittich who was his school friend from Groningen days, Clauberg left Herborn for the University of Duisburg, where he taught until his death.

Clauberg's *Logica vetus et nova* (1654) was an attempt to combine the principles of traditional logic with those of a 'Cartesian' logic, something not unlike what Arnauld was to do in the Port-Royal logic a few years later, despite the fact that Descartes himself rejected the enterprise of formal logic (see Arnauld, A. §2; Port-Royal). His *Metaphysica de ente* (1664), sometimes mistaken for a second edition of his earlier *Elementa philosophiae sive ontosophia*, was in reality a completely separate work, presenting a metaphysics in scholastic style but with Cartesian content. Its subject matter was the science of being as such. Clauberg treated the notions of being, essence, existence, the One, the True, the Good and so on - all found in traditional Aristotelian treatises on metaphysics. However, Clauberg's metaphysics was grounded in the Cartesian distinction between mind and body, which he takes to replace the ten Aristotelian categories as the highest division of things into classes. Again, this is somewhat paradoxical, since Descartes himself seemed to reject the idea of a metaphysics in the sense of a science of being as such. His logic and metaphysics in particular earned Clauberg the reputation of the learned Cartesian, and his presentation of Cartesian ideas in scholastic form, complete with scholarly references, probably eased the way for Descartes' philosophy in the German universities.

Of particular philosophical interest is his *Physica* (1664), a collection of treatises including one on living things and another on the union between mind and body. Like other Cartesians writing in this decade, including the French circle of Cartesians (Cordemoy, La Forge, Clerselier and so on), with whom he was probably in contact, Clauberg put forward a kind of occasionalism (see Occasionalism). First, matter as such is inert, and God is the only genuine efficient cause of motion in the physical world; like Descartes, Clauberg held that the soul or mind, a substance whose essence is thought, is really distinct from body, a substance whose essence is extension. But unlike Descartes, he held that the two are not genuine ('physical') causes of changes in one another, and that mind and body are united simply by virtue of the quasi-causal relations between them. The mind is what he called a 'moral cause' of changes in the body; mind does not create new motion, but only changes the direction of the motion already there, otherwise the Cartesian law of the conservation of quantity of motion would be violated. Conversely, body, being less noble than mind, is only the 'procatarctic cause' of ideas in the mind, giving mind the 'occasion to have these or those ideas, which it always has the power to have within it'.

See also: Descartes, R.

DANIEL GARBER

List of works

Clauberg, J. (1691) *Opera Omnia Philosophica (Collected philosophical works)*, Amsterdam, 2 vols; repr. Hildesheim: Olms Verlag, 1968.(A collection of virtually all Clauberg's philosophical writings, as well as a biographical sketch of the author, all in Latin. Since this edition includes many notes and comments not in the originals, it must be used with care.)

Clauberg, J. (1647) *Elementa philosophiae sive ontosophia (The elements of philosophy, that is, the ontosophia)*, Gronigen.(An early Aristotelian treatise on metaphysics, not reprinted in the *Opera Omnia*.)

Clauberg, J. (1652) *Defensio Cartesiana adversus Jacobum Revium (A defence of Descartes against Jacob Revius)*, Amsterdam; repr. in *Opera Omnia Philosophica*, Hildesheim: Olms Verlag, 1968, vol. 2.(A contribution to a controversy at the University of Leiden over Descartes' philosophy, this also is a close discussion and defence of Descartes' *Discourse on the Method*.)

Clauberg, J. (1654) *Logica vetus et nova (Logic, old and new)*, Amsterdam; repr. in *Opera Omnia Philosophica*, Hildesheim: Olms Verlag, 1968, vol. 2 .(An attempt to combine the principles of traditional logic with those of a 'Cartesian' logic. A second edition was published in 1658.)

Clauberg, J. (1655) *Initiatio Philosophi sive dubitatio cartesiana (An initiation into philosophy, or, Cartesian doubt)*, Duisburg; repr. in *Opera Omnia Philosophica*, Hildesheim: Olms Verlag, 1968, vol. 2.(A commentary on and defence of Cartesian doubt as a starting-place for philosophy.)

Clauberg, J. (1656) *De Cognitione Dei et nostri...Exercitationes centum (One-hundred exercises concerning the knowledge of God and of ourselves)*, Duisburg; repr. in *Opera Omnia Philosophica*, Hildesheim: Olms Verlag, 1968, vol. 2.(Largely a commentary on Descartes' views concerning God and the mind, it also contains assorted lectures and disputations on related subjects.)

Clauberg, J. (1657) *Unterschejd zwischen den Cartesianischen, und der sonst in Schulen gebräuchlichen Philosophie Beschriebene (The difference between the Cartesian philosophy and that normally used in the schools)*, Duisburg; repr. in *Opera Omnia Philosophica*, Hildesheim: Olms Verlag, 1968, vol. 2.(This may have been written originally for the examination of Cartesianism at Herborn in 1651. It appeared in Latin translation in 1680.)

Clauberg, J. (1658) *Paraphrasis in Renati Descartes meditationes de prima Philosophia (Paraphrase of René Descartes' Meditations on First Philosophy)*, Duisburg; repr. in *Opera Omnia Philosophica*, Hildesheim: Olms Verlag, 1968, vol. 1.(A virtual reprint of Descartes' *Meditations*, with extensive commentary.)

Clauberg, J. (1664) *Physica quibus rerum corporearum vis et natura, Mentis ad Corpus relatae proprietates, denique Corporis ac Mentis arcta et admirabilis in Homine conjunctio explicantur (The Physics, in which is explained the force and nature of corporeal things, the properties of the mind as it relates to the body, and finally the close and wondrous connection between the mind and the body in man)*, Amsterdam; repr. in *Opera Omnia Philosophica*, Hildesheim: Olms Verlag, 1968, vol. 1.(A collection of shorter treatises, including the *Physica contracta (The short physics)*, the *Disputationes quibus principia Physica latius explicantur (Physical disputations in which the principles of physics are explained more fully)*, the *Theoria corporum viventium (The theory of living bodies)*, and the *Corporis et animae in homine conjunctio (The connection between body and soul in man)*.)

Clauberg, J. (1664) *Metaphysica de ente quae rectius Ontosophia (Metaphysics concerning being, more properly called ontosophia)*, Amsterdam; repr. in *Opera Omnia Philosophica*, Hildesheim: Olms Verlag, 1968, vol. 1. (There may have been an earlier edition of this published in 1660. Despite the similarity of the titles, this is *not* a second edition of the 1647 work.)

References and further reading

Balz, A.G.A. (1951) 'Clauberg and the Development of Occasionalism', in *Cartesian Studies*, New York: Columbia University Press.(A study of Clauberg with special attention to his occasionalism.)

Bohatec, J. (1912) *Die Cartesianische Scholastik in der Philosophie und Theologie der reformierten Dogmatik des 17. Jahrhunderts*, Leipzig: Deichert.(Classic work on Cartesianism in Protestant thought, in which Clauberg plays an important role.)

Menk, G. (1985) ''Omnis novitas est periculosa': Der frühe Cartesianismus an der Hohen Schole Herborn (1649-1651) und die reformierte Geisteswelt nach dem dreißigjährigen Krieg', in K. Schaller (ed.) *Comenius: Erkennen, Glauben, Handeln*, St. Augustin: Richarz, 1985.(A very good account in German of Clauberg's early years, as well as a discussion of the crisis over Cartesianism at Herborn in 1651.)

Trevisani, F. (1992) *Descartes in Germania*, Milan: FrancoAngeli.(A careful study of Cartesianism in the seventeenth century at the University of Duisburg, where Clauberg taught.)

Verbeek, T. (1992) *Descartes and the Dutch: Early Reactions to Cartesian Philosophy, 1637-1650*, Carbondale, IL: Southern Illinois University Press.(A general study of Dutch Cartesianism, in which Clauberg plays a

significant role.)
Weier, W. (1970) ‘Cartesianischer Aristotelismus im siebzehnten Jahrhundert’, *Salzburger Jahrbuch für Philosophie* 14: 35-65.(A study of Clauberg as an example of the attempt to combine elements of Aristotelian and Cartesian thought.)
Weier, W. (1982) ‘Der Okkasionalismus des J. Clauberg und sein Verhältnis zu Descartes, Geulincx, Malebranche’, *Studia Cartesiana* 2: 43-62.(A study of Clauberg’s occasionalism.)

Cleanthes (331-232 BC)

The Greek philosopher Cleanthes of Assos played a leading role in the formation of Stoicism. He was at once the most physicalist and the most religious of the Stoics. Pupil, and eventual successor (in 262), of the school's founder Zeno, he wrote numerous philosophical works, including some poetry. In particular, he developed the notion of fire as the world's governing principle.

Born at Assos in Asia Minor (modern Turkey), Cleanthes is said to have had his first career as a boxer. Moving to Athens, he joined the newly emerging Stoic school, led by Zeno of Citium. Poverty forced him to work at night, drawing water, and to attend Zeno's classes during the day. He became Zeno's closest ally in the school, and in 262 BC succeeded him as its head. He died at a very advanced age, to be succeeded by Chrysippus.

Cleanthes was a prolific writer: we know the titles of some 57 works. Either in addition to these works, or more probably within them, he composed passages of philosophical poetry, of which his short but very powerful *Hymn to Zeus* survives intact. His philosophical outlook was distinctive in three strikingly different ways. (1) He was the most physicalist of the Stoics, with a special interest in the role of fire. (2) He was perhaps the main logician in the school's first generation, and was probably responsible for developing the key notion of the incorporeal *lekton* (see Stoicism §8). (3) He was the most deeply religious of Stoic thinkers.

Cleanthes, even more than other Stoics, revered Heraclitus as an authority on cosmological matters. From Heraclitus he acquired and developed the doctrine of the primacy of fire, which he viewed as the vehicle of divine 'reason' (*logos*), varying according to its degree of 'tension'. Not only does the world begin and end in fire, as nearly all Stoics agreed (see Stoicism §5), but fire - which includes warmth - is the main creative force throughout nature. Here he was not in actual disagreement with other Stoics, but he was the school's leading fire theorist, while after him the emphasis shifted from fire to *pneuma*, which includes air as well as fire (see Stoicism §3).

Zeno had identified the virtues as various kinds or aspects of wisdom - using a description which left the nature of their unity (see Stoicism §16) unclear. Cleanthes interpreted this as referring to a suitably tensioned 'stroke of fire' in the soul, constituting a kind of mental strength which, in different circumstances, comes to be called 'justice', 'courage' and so on. While not incompatible with the usual Stoic intellectualist view of virtue, this analysis has a strikingly physicalist emphasis. Similarly, where Zeno had called a mental impression (*phantasia*) a 'printing' in the soul, Cleanthes was criticized by his successor Chrysippus for treating this as a literal imprint, as if in wax.

Cleanthes' deeply religious temperament is evident from his *Hymn to Zeus*. There, uniquely among Stoics as far as we know, he enhances god's majesty by absolving him of willing the actions of the bad. Another symptom is his outrage at the contemporary astronomer Aristarchus, who alone in antiquity suggested that the earth might orbit the sun rather than vice versa, while revolving on its own axis. Cleanthes proposed that he be prosecuted for impiety, because his theory 'moved the world's hearth'. Yet he is also the author of some surprisingly rationalistic explanations of the origin of religion (Cicero, *On the Nature of the Gods* II 13-15).

Cleanthes' interest in logic is attested by the solution he provided to the Master Argument (see Diodorus Cronus §5), which was being used to argue for the necessity of the future. He responded by rejecting its premise 'Every past truth is necessary', perhaps on the ground that present and future events determine the truth-value of past predictions.

DAVID SEDLEY

References and further reading

Arnim, H. von (1903-5) *Stoicorum Veterum Fragmenta (Fragments of the Early Stoics)*, Leipzig: Teubner, with vol. 4, indexes, by M. Adler, 1924.(The standard collection of early Stoic fragments, in Greek and Latin, commonly abbreviated as *SVF*; Cleanthes is in volume 1, 103-39.)

Cicero (45 BC) *On the Nature of the Gods*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933. (Latin text with English translation. Book II is an eloquent presentation of Stoic theology; see 13-15, 24, 40-1 for Cleanthes' contributions.)

Cleanthes (331-232 BC) Fragments, in A.C. Pearson, *The Fragments of Zeno and Cleanthes*, Cambridge:

Cambridge University Press, 1891.(Still useful, but mainly for specialists.)
Long, A.A. (1975-6) 'Heraclitus and Stoicism', *Philosophia* 5-6: 133-56; repr. in A.A. Long, *Stoic Studies*, Cambridge: Cambridge University Press, 1996, 35-57.(Includes much on Cleanthes' debt to his favourite Presocratic.)
Solmsen, F. (1968) 'Cleanthes or Posidonius? The Basis of Stoic Physics', in F. Solmsen, *Kleine Schriften*, Hildesheim: Olms, vol. 1, 436-60.(On Cleanthes' theory of vital heat.)

Clement of Alexandria (AD 150-215)

Clement of Alexandria, a Christian Platonist, came to conversion through philosophy. In a series of allusive writings he presented a Hellenized Christianity along with the philosophical syncretism of his age: Stoic ethics, Aristotelian logic and especially Platonic metaphysics. Just as Paul saw the Hebrew prophets and law as a preparation for the Gospel, Clement saw Christianity as making possible a confluence of Plato and the Old Testament, both offering anticipations of Jesus' teaching. Clement's fusion of Platonism and Christianity vehemently opposed the dualism and determinism of gnostic theosophy, and stressed free choice and responsibility as fundamental to moral values. Central to his writing is the vindication of faith as the foundation for growth in religious knowledge by philosophical contemplation and biblical study.

Clement's principal works are *Protreptikos (Exhortation to Conversion)* attacking polytheism, *Paidagogos (Moral Tutor)* on etiquette and Christian ethics, and above all the *Stromateis (Miscellany)*, left incomplete. The eighth and last book of *Stromateis*, Clement's commonplace work on logic and epistemology, was not part of the original plan. He intended to make the third work in his trilogy a systematic statement of religious philosophy entitled *Didaskalos (Teacher)*, but wisely, since it would have been controversial, he never wrote it.

The *Stromateis* touches elusively on profound issues but then changes the subject. The reader is invited to participate in the quest for religious truth, granted to those who strenuously search but also receive divine help. If one had to make a choice between eternal salvation and continuing growth in the knowledge of God, one would certainly choose the latter (a saying plagiarized without acknowledgement by G.E. Lessing in *Duplik*).

Platonic language appears in Clement's theme of the mysteriousness of faith in God. Echoes of Plato's seventh letter are heard in his diffidence about putting advanced theology into writing, and the third book of the *Republic* provides his model for attacks on the clever and conceited who scorn such topics.

Clement's objective is to defend educated Christians who see no reason to jettison good philosophy merely because they have renounced the world, the flesh and the devil. Many in the Church were suspicious of philosophers because of the appeal made to them by gnostic sects claiming to offer a higher understanding (*gnosis*). The 'orthodox' (Clement uses the word with an ironical touch) were confident that faith alone sufficed for salvation. Some asserted (following Enoch 16:3) that philosophy was the invention of the devil, transmitted to humanity through the fallen angels' mating with women. Clement noticed that even this negative notion implicitly conceded that in philosophy, perhaps in confused form, there was heavenly truth. He was sure that Plato on God's transcendence, on the aspiration that the soul be 'as like God as possible' and on the goal of the vision of God (as in *Phaedrus*) was so close to Christian faith as to be virtually indistinguishable. The Stoics, on the therapy of the emotions and on the absolutes of the moral life, were similarly stating what he knew to be true. Here Clement owed much to Philo of Alexandria, especially for the thesis that philosophy prepares the mind for theology just as the liberal arts prepare it for philosophy.

How the philosophers succeeded in getting so much right was a question with more than one possible answer. Clement's main answer is that all human beings, being made in God's image, possess a 'shared mind', a 'natural conception' as the Stoics had put it, and so have a universal intuition for God and for virtue and righteousness. The philosophers' ability to 'reflect the truth' may therefore be ascribed to the exercise of natural endowment. Moreover, there can be special inspiration imparted to selected individuals; that this inspiration is Christian in nature is evident from Jesus' parable of the sower.

Philo and Numenius of Apamea had earlier supposed that Plato had been able to get so much right because of his studies in the writings of Moses, who could be shown by chronographers to have lived centuries earlier. Clement develops at length the theme of Greek plagiarism from barbarian traditions, inverting an argument deployed against Christianity by the pagan Platonist Celsus, whose work *Alethes Logos (The True Account)* Clement probably knew. Celsus also thought the oldest and most venerable religious tradition was to be found among primeval barbarian peoples uncorrupted by civilization. Clement was not therefore implausible to his contemporaries in suggesting that classical philosophy was dependent on older, non-Greek tradition for religious truth; and such a lack of religious originality in Plato or Pythagoras was, for him, no demerit.

The nature of faith was a problem never far from the centre of Clement's stage. It was under attack on one side from gnostics who scorned simple catechetical teaching (see Gnosticism); on the other side, educated pagans thought the Christian exhortation to 'believe' denied rational investigation and asked for mere submission to authority. Clement drily remarked that the latter method was a famous principle of no less a figure than Pythagoras. Religious faith he saw as related to natural assent. We assent, without a full prior inquiry, to the presuppositions of rational demonstration. We use the term 'faith' (*pistis*) for the state of satisfied conviction at the conclusion of a sound argument, or for the immediacy of assent to what our senses perceive. The word also contains some element of the commitment of loyalty to a cause or a person. In Christianity, faith is the necessary precondition of advance in the knowledge and love of God through meditation on the inner, symbolic, meaning of Scripture, linked to contemplation of higher things. Thereby the soul may be gradually freed from the distractions of passion and enabled to rise to the mystical vision of God.

Much technical Stoic and some Aristotelian language enters Clement's description of the conflict between reason and emotion. To live according to nature is to obey the Creator-Logos, Reason for all things, who for Clement is known in Christ and who is able to bring power to moderate the passions and, at the highest level of ascetic contemplation, suppress them. Passionlessness (*apatheia*) is synonymous with Plato's 'likeness to God', and with the perfection of which Jesus spoke in the Sermon on the Mount.

See also: Faith; Gnosticism; Origen; Platonism, Early and Middle; Platonism, medieval

HENRY CHADWICK

List of works

Clement of Alexandria (150-215) *Clemens Alexandrinus*, Leipzig: J.C. Hinrichs. (Standard Greek edition of Clement's works by O. Stählin, 4 vols with index, later reprints with revised text; includes the *Protreptikos*, *Paidogogos* and *Stromateis*.)

Clement of Alexandria (150-215) *Protreptikos (Exhortation to Conversion)*, trans. G.W. Butterworth in *Clement of Alexandria*, Cambridge, MA: Harvard University Press, 1919.

Clement of Alexandria (150-215) *Stromateis (Miscellany)*, vols 1-3 trans. J. Ferguson, Washington, DC: Catholic University Press, 1991.(There is also a translation by W. Wilson, consisting of all but *Stromateis* 3 on sex, in *The Writings of Clement of Alexandria*, Edinburgh: T. & T. Clark, vol. 4; repr. Grand Rapids, MI: Eerdmans, 1979, but this should be regarded as a brave failure; nevertheless, it is the only English translation of much of Clement.)

References and further reading

Bigg, C. (1913) *The Christian Platonists of Alexandria*, 2nd edn, Oxford: Clarendon Press.(A classic portrait of Clement and Origen.)

Chadwick, H. (1966) *Early Christian Thought and the Classical Tradition*, Oxford: Clarendon Press.(Covers the work of Justin, Clement and Origen.)

de Faye, E. (1906) *Clément d'Alexandrie (Clement of Alexandria)*, 2nd edn, Paris: E. Leroux.(Standard liberal portrait.)

Lilla, S. (1971) *Clement of Alexandria: A Study in Christian Platonism and Gnosticism*, London: Oxford University Press.(The best study of Clement's middle Platonism.)

Osborn. E.F. (1957) *The Philosophy of Clement of Alexandria*, Cambridge: Cambridge University Press. (Systematic treatment.)

Spanneut, M. (1957) *Le Stoicisme des Pères de l'Église (The Stoicism of the Church Fathers)*, Paris: Éditions du Seuil.(Clement's transformation of Stoic ethics.)

Wyrwa, D. (1983) *Die christliche Platonaneignung in den Stromateis des Clemens von Alexandrien (Christian Platonism in the Stromateis of Clement of Alexandria)*, Berlin: de Gruyter.(Detailed examination of the appropriation of Platonic themes in the *Stromateis*.)

Cleomedes (*c*. 2nd century AD)

Cleomedes was the author of On the Heavens, a Greek treatise on elementary astronomy surviving from a larger exposition of Stoic philosophy. Its account of measurements of the earth, and its applications of Stoic epistemology and philosophy of science, make it important evidence on poorly documented subjects.

Since Cleomedes refers to Posidonius (*c*.135-*c*.50 BC), and conducts polemics against Peripatetics and Epicureans, he could not have been active much later than AD 200. *On the Heavens*, his presentation of elementary astronomy, is part of a larger exposition to which he himself alludes (see I 1.7, and I 1.94-5). Book I deals with celestial and terrestrial zones, seasonal and climatic differences (I 1-4), the sphericity and centrality of the earth (I 5 and 6) and the earth's transparency in celestial observations (I 8). The account in I 7 of Posidonius' and Eratosthenes' measurements of the size of the earth is indispensable evidence for historians of science. The subjects of book II 1-3 are the size of the heavenly bodies, and in particular the Epicurean view that they are no larger than they appear, and lunar theory (II 4-6). A brief appendix (II 7) deals with planetary latitudes, elongations and periods. The work is non-technical, eschews discussion of the complexities of planetary motion and presents material in the stylized fashion typical of ancient elementary pedagogy.

In Byzantium and Renaissance Europe *On the Heavens* was exclusively used as a textbook of elementary astronomy. In the late nineteenth and early twentieth centuries scholars mined it for specific doctrines of Posidonius. But it is Posidonius' philosophy of science that is of the greatest significance for Cleomedes (see Posidonius §6). Posidonius (fr. 18) had defined astronomy as a science subordinate to physical theory, and as employing the ancillary sciences of mathematics and geometry. Cleomedes' work may not precisely embody this programme, but it does fall within its general orbit. Thus it specifies the role of assumptions from physics and geometry in calculations of the size of the earth, sun and moon (I 7; II 1.282, 304-5). Also, the treatise opens by defining, with detailed supporting arguments, the physical context for spherical astronomy as the standard Stoic cosmology of a finite world in infinite void. Elsewhere the theories of the Stoic continuum play a much larger role than physical theory normally would in astronomical handbooks (see for example I 5.6-9, 126-38; I 6.40-3; I 8.79-99; II 4-). The epistemological problems raised by astronomical observations, to which Posidonius (fr. 18) also alludes, are reflected in references to the Stoic theory of the criterion of truth (I 5.1-6; II 1.1-4), and in the lengthy polemic in book II 1 against the Epicurean claim that the sun is no larger than it appears.

It may be unfortunate that Cleomedes' Stoicism is embedded in a survey of elementary astronomy, but in the absence of his other works, *On the Heavens* at least shows that the Stoicism of late antiquity was not, as has been traditionally thought, exclusively preoccupied with prescriptive ethics.

ROBERT B. TODD

List of works

Cleomedes (*c*. 2nd century AD) *On the Heavens*, ed. R.B. Todd, *Cleomedis Caelestia*, Leipzig: Teubner; 1990; trans. R. Goulet, *Cléomède: Théorie Élémentaire*, Paris: Vrin, 1980.(Respectively, Greek text and French translation with commentary.)

References and further reading

Barnes, J. (1989) 'The Size of the Sun in Antiquity', *Acta Classica* (Debrecen) XXV: 29-41.(A survey of the arguments in *On the Heavens*, II 1 in their wider context.)

Kidd, I.G. (1988) *Posidonius: II. The Commentary*, Cambridge: Cambridge University Press.(Includes important commentary on several Cleomedean texts.)

Posidonius (*c*.135-*c*.50 BC) Fragments, in *Posidonius: I. The Fragments*, ed. L. Edelstein and I.G. Kidd, Cambridge: Cambridge University Press, 1989, 2nd edn.(This contains all the Cleomedean passages in which Posidonius is named, and is the edition from which the fragments (fr.) and testimonia (T) of Posidonius are cited.)

Todd, R.B. (1989) 'The Stoics and their Cosmology in the First and Second Centuries A.D.', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II.36: 3, 1365-78.(On the wider historical context of *On the Heavens*.)

Todd, R.B. (1992) 'Cleomedes', *Catalogus Translationum et Commentariorum*, VII: 1-11.(A survey of Cleomedes' reception, with particular reference to the Renaissance.)

Todd, R.B. (1998, forthcoming) 'Physics and Astronomy in Post-Posidonian Stoicism: The Case of Cleomedes', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II.37: 5.(An analysis of the secondary literature on Cleomedes and an assessment of his relation to Posidonius.)

Cockburn, Catharine (1679-1749)

Catharine Cockburn (Catharine Trotter) was a British moral philosopher who turned to philosophy after a successful career as one of the first woman playwrights. She wrote no substantial systematic treatise of her own, but intervened ably and anonymously in philosophical and theological debates of her day, in particular the debate on ethical rationalism triggered by Samuel Clarke's 1704-5 Boyle lectures. Her adversaries included Thomas Rutherforth, William Warburton, Isaac Watts, Francis Hutcheson and Lord Shaftesbury. Her most famous contribution to the philosophy of her time was her able 1702 defence of Locke's Essay Concerning Human Understanding. Her letters, published posthumously, discuss a range of philosophical topics.

Born Catharine Trotter, in London of Scottish parentage, Cockburn appears to have been largely self-taught in philosophy. She was linked to circles around, first, the playwright William Congreve, and later Gilbert Burnet, Bishop of Salisbury. The latter may have been instrumental in persuading her to give up writing plays in favour of philosophy. After her marriage in 1708 to the Scottish clergyman Patrick Cockburn, the adverse pressures of rearing a family in reduced financial circumstances combined with deteriorating eyesight to hamper her pursuit of philosophy.

Throughout her writings Cockburn is consistent and nondogmatic in her philosophical position. Among the topics she discusses are necessary existence, infinite space, the idea of substance and the nature of spirit, but her main focus is on ethical questions, in particular the nature of virtue and of moral obligation. Opposing the ethical relativism of Hobbes, she argues that morality is neither arbitrary nor conventional in foundation (see Hobbes, T. §5). Against latter-day would-be defenders of morality, she denies that moral virtues exist in the abstract and that the obligation to virtue is founded in innate moral sense. Nor does moral obligation depend on externally imposed sanctions. Rather, moral obligation and the principles of moral conduct can be understood from the 'reason, nature and fitness of things' (1751a, vol. 2: 44). Human beings are rational and social beings, so for human beings to act in accordance with their own nature is to act rationally and for the good of all. By the exercise of reason we arrive at a knowledge of the principles of right and wrong from ideas of sensation and reflection. Having thus acquired a concept of goodness, we deduce that God is good and deserves our obedience, and that it is most fit for a reasonable being to act in conformity with God's moral perfections.

The outlines of the philosophical position from which Cockburn defended Clarke are already evident in her defence of Locke in which she wrote 'I know no philosopher before him that has fixed morality upon so solid a foundation' (1702: 48). It is to her credit that, although she did not agree with Locke in all particulars, she defended him against misunderstandings and misreadings that imputed to him scepticism or deism, exposing the inconsistencies and fallacious arguments of his opponents: as she reminded his critics, 'the question is not what Mr. Locke thinks, but what may be proved upon his principles' (1702: 60). She rallied twice more to Locke's defence in *A Letter* (1826) and Vindication (1751b), defending his philosophy from charges that it undermined Christian orthodoxy.

Catherine Cockburn was recognized as an astute mind by her contemporaries, among them Locke, who gratefully acknowledged her defence of his philosophy. Leibniz, too, praised her writings, and she was dubbed 'Sappho Ecossoise' by Leibniz's patron, Sophie Charlotte of Hanover.

See also: Hutcheson, F.; Shaftesbury, Third Earl of

SARAH HUTTON

List of Works

Cockburn, C. (1751a) *The Works of Mrs Catharine Cockburn*, ed. T. Birch, London: J. & P. Knapton.(Collected philosophical writings, with extracts from letters and a biographical introduction.)

Cockburn, C. (1702) *A Defence of Mr Locke's Essay of Human Understanding*, London: W. Turner & J. Nutt. (Defence of Locke's *Essay*, published anonymously.)

Cockburn, C. (1726) *A Letter to Dr Holdworth…concerning the Resurrection of the same Body. In which the passages that concern Mr Locke are chiefly considered*, London: S. Motte.(Further defence of Locke.)

Cockburn, C. (1743) *Remarks upon Some Writers in the Controversy Concerning the Foundation of Moral Virtue*

and Moral Obligation, London.(Defence of Samuel Clarke.)
Cockburn, C. (1747) *Remarks upon the Principles and Reasonings of Dr Rutherforth's Essay on the Nature and Obligations of Virtue, in Vindication of…the late Dr Samuel Clarke*, London: J. & P. Knapton.(Defends Samuel Clarke.)
Cockburn, C. (1751b) *Vindication of Mr Locke's Christian Principles*, in *The Works of Mrs Catharine Cockburn*, vol. 1, ed. T. Birch, London.(Defends Locke on resurrection.)

References and further reading

Bolton, M. (1994) 'Aspects of the Philosophical Work of Catharine Trotter', *Journal of the History of Philosophy* 31: 565-88.(The only serious study of Cockburn's philosophy to date.)
Gosse, E. (1916) 'Catharine Trotter, Precursor to the Blue Stockings', *Transactions of the Royal Society of Literature* 34: 87-118; repr. in *Some Diversions of a Man of Letters*, London: William Heinemann, 1919. (Largely of literary interest.)

Coercion

Coercion (also called 'duress') is one of the basic exculpating excuses both in morality and in some systems of criminal law. Unlike various kinds of direct compulsion that give a victim no choice, a coercee is left with a choice, albeit a very unappealing one. They can do what is demanded, or can refuse, opting instead for the consequences, with which they are threatened. Sometimes courts find that the coercive threat that led the defendant to act as they did was objectively resistible by any person of reasonable fortitude, especially when the defendant's conduct was gravely harmful to others or to the state.

A proposal is an offer when it projects for the recipient's consideration a prospect that is welcome in itself, and not harmful or unwelcome beyond what would happen in the normal course of events. Coercive offers, according to some writers, are those that force a specific choice from the victim while actually enhancing their freedom. Some argue, however, that genuine coercion requires the active and deliberate creation of a vulnerability, and not mere opportunistic exploitation of a vulnerability discovered fortuitously.

1 Coercion and compulsion

Coercion is a technique for forcing people to act as the coercer wants them to act, and presumably contrary to their own preferences. It usually employs a threat of some dire consequence if the actor does not do what the coercer demands, although it is controversial whether a non-threatening offer might in some contexts be coercive (see §2). There is a contrast between coercion, in which the victim can choose one or the other of the two alternatives allowed by the coercer (for example, 'your money or your life'), and what we might call compulsion, in which the victim has no choice at all, as for example when the victim is knocked unconscious and dragged away. In compulsion, options are closed by physical force, making alternatives to what the coercer demands physically impossible. In contrast, coercion does not destroy the preferred alternative as much as destroy its appeal by increasing its cost. The cost is the alternative specified in the coercer's threat.

2 Threats and offers

If such things as 'coercive offers' exist they must not, by definition, be a species of threat, but rather be in essential contrast to the more typical instances of coercion which involve a demand enforced by a threat. Coercive offers would also be in contrast to the more familiar noncoercive inducement; attractive prospects which function as enticements that one 'cannot resist'. With few exceptions, philosophers have held that allegedly 'irresistible' attractiveness does not negate voluntariness or excuse the attracted party from responsibility.

The common genus of which threats and offers are species is that of preference-affecting proposals - those proposals not likely to be received with perfect indifference. To begin with, an offer is a proposal to contribute to a person something that they want or find welcome, something they would prefer having to not having. A threat, on the other hand, is a proposal to inflict something they would not welcome. But, threats and offers are more than 'proposals.' They also contain a reciprocation condition; a 'demand' in the case of threats, and a 'request' in the case of offers. These elements amount to the coercer's *quid pro quo*: 'what is in it for them'. In the case of coercive threats, at least, far more is in it for them than for the other party. This way of speaking makes it clear that credible threats in support of demands are a kind of contractual transaction, but a transaction so lopsided as to be unconscionable, involuntary on one side, and legally invalid.

One kind of preference-affecting proposal is a demand backed by a threat; another is a request backed by an inducement. Both of these can be construed as typically biconditional in form: the action that *A* requests or demands from *B* is taken to be both a necessary and a sufficient condition of *A*'s reciprocation: 'I will do *x* to (or for) you *if and only if* you will do *y* for me'. *A*'s offer of $11,000 for *B*'s car thus translates:

(1) If you give me your car, I will give you $11,000.
(2) If you do not give me your car, I will not give you $11,000.

The first preference-affecting proposal (1) - 'I will give you $11,000' - is obviously an offer. The second - 'I will not give you $11,000' - is not an offer, but neither is it a threat. 'I will not give you $11,000' would count as a threat only if the speaker paid that amount regularly and was now proposing to withhold it, thus disappointing the

other party’s natural expectation. Since the proposal is part offer and part neither offer nor threat, it is an offer overall.

Such an example is an easy case. It would be more difficult if it were unclear what the ‘normal course of events’ was, or what is in the circumstances a ‘reasonable expectation’. An important element of the analysis then would be some proposed ‘norms of expectability’. Some writers employ for this purpose a kind of statistical test while others employ a test of moral requiredness (Nozick 1969). Suppose the consequent clause in the second biconditional (2) is something like: ‘I will not pull you out of the lake in which you are about to drown’ (unless you promise to sign over all your worldly goods to me). Such a proposal violates a moral norm. We not only expect most people in such circumstances to attempt easy rescue (statistical norm); our morality *requires* it (moral norm). Thus the proposal not to rescue would be a threatened deviation both from statistical and from moral norms. And when the projected behaviour deviates from one of these norms but not the other, it can be sufficient to classify it as a threat. In summary, if *A*’s projected consequences of *B*’s failure to do *y* are worse (or less welcome) than they would be in the normal and expected course of events (where ‘expected’ straddles ‘predicted’ and ‘morally required’), then *A*’s proposal is a threat. If it correctly portrays those consequences as improvements, it is an offer.

3 Coercive and noncoercive proposals

Various examples, actual and hypothetical, have convinced some writers that there are such things as coercive offers. One rather standard example is that of the ‘lecherous millionaire’: *B*’s child will die unless the child receives expensive surgery. *A*, a millionaire, proposes to pay for the surgery if, but only if, *B* will agree to become his mistress (Wertheimer 1987: 229). The millionaire’s proposal, which does not project any harm beyond what would happen without his gratuitous intervention, is clearly an offer; an ‘illicit offer’, but an offer none the less. The controversy over this example has been mainly over whether the offer is coercive or not. On the one side, it has been argued that the example, at least from *B*’s point of view, is coercive in just the sense in which the standard gunman’s threat is coercive (Feinberg 1986). In both examples, the powerful wrongdoer (*A*) uses superior advantages (weapons or money) to manipulate *B*’s options so that *B* has ‘no choice’ but to do what *A* wants. In either case, the option seen from the victim’s point of view is: ‘Sleep with me or your baby dies’.

On the other side, it is frequently argued that it is misleading to label as ‘coercion’ a proposal the effect of which is to create a net *increase* in a person’s options, giving them a valued choice not previously possible (Zimmerman 1981). Thus *B* has been given an option she did not have before, without losing any options that she did have before, so that in a sense her freedom has been increased. Yet to some the suggestion that a coercive proposal could be freedom-enhancing is intolerably paradoxical.

4 Coercion and exploitation

A different solution is proposed by David Zimmerman (1981), who employs a critical distinction between coercion, understood as necessarily restrictive of freedom, and exploitation, understood as not necessarily coercive. In the lecherous millionaire example, *A* is an opportunist who happens to chance upon a vulnerable party, but who played no role himself in creating her vulnerability. In Zimmerman’s terms, that makes *A* an exploiter. For him to be a genuine coercer, it must be the case that he deliberately created the situation in which *B* is helpless. Zimmerman illustrates the distinction between exploitative offers that are, and those that are not coercive, with the following example: *A* kidnaps *B*, brings him to the island where *A*’s factory is located, and abandons him on the beach. All the jobs in *A*’s factory are considerably worse than those available on the mainland. The next day *A* approaches *B* with the proposal ‘Take a job in my factory and I won’t let you starve’.

This example is a genuine coercive offer, according to Zimmerman, but if there were two factories on the island, and the owner of the second learns of *B*’s plight and rushes to the beach before *A* can get there in order to make the same kind of offer to the kidnapped worker, then *A*’s offer is merely opportunistic exploitation: ‘Coercing goes beyond exploiting, however objectionable the latter may be’ (Zimmerman 1981: 122). The effect on the victim appears to be coercive whether the exploiter deliberately created the vulnerability or only took advantage of it opportunistically. ‘Work or starve’ is as coercive in effect as ‘Sleep with me or your baby dies’. However, in the example that Zimmerman uses, *A* inflicts a kind of double blow on *B*, first by undermining *B*’s freedom (the act of kidnapping) and then by taking advantage of his undermined state with a ‘coercive offer’. The lecherous

millionaire, in contrast, was a mere opportunist. Whilst his behaviour may be viewed as morally reprehensible, it can be argued that his unsavoury conduct did not violate the vulnerable party's right; nor did it threaten any harm to *B* beyond what she could expect to occur anyway in the normal course of events.

See also: Freedom and liberty; Responsiblity

JOEL FEINBERG

References and further reading

Feinberg, J. (1986) *Harm to Self*, Oxford: Oxford University Press, 189-268.(Argues against paternalistic justifications for coercive interference with a person's voluntary conduct. Considers the ways in which one person's consent to another's action can fall short of voluntariness, including the case in which 'consent' is acquired through the use of coercive action. Includes a discussion of the 'lecherous millionaire', §3.)

Fletcher, G. (1978) *Rethinking Criminal Law*, Boston, MA: Little, Brown, 759-876.(Discusses the theory of criminal defences, both justifications and excuses, in particular, those of duress and necessity and the relations between the two.)

Frankfurt, H. (1973) 'Coercion and Moral Responsibility', in T. Honderich (ed.) *Essays on Freedom of Action*, London: Routledge & Kegan Paul.(Discusses coercion as an exercise in moral discourse. An influential article.)

Gross, H. (1979) *A Theory of Criminal Justice*, Oxford: Oxford University Press, 276-91.(A systematic theory of criminal law which includes a discussion of coercion as an excuse in criminal law.)

Nozick, R. (1969) 'Coercion', in S. Morgenbesser, P. Suppes and M. White (eds) *Philosophy, Science and Method*, New York: St Martin's Press.(A classic source for discussion of 'norms of expectability' and their use in distinguishing threats from offers.)

Taylor, M. (1982) *Community, Anarchy and Liberty*, Cambridge: Cambridge University Press.(Asks whether a society without state coercion could ever achieve durable stability, and if so, under what conditions. Examines the relations - conceptual and empirical - between coercion and community.)

Wertheimer, A. (1987) *Coercion*, Princeton, NJ: Princeton University Press.(One of the few book-length analyses of the concept of coercion, both in law and everyday life. Contains detailed discussion of the relation between coercive threats and offers, coercion and exploitation, and coercion and voluntariness.)

Williams, G. (1961) *Criminal Law: The General Part*, London: Stevens & Sons Limited, 751-69.(A thorough and authoritative examination of the treatment of coercion in English law. Attempts to resolve various controversies about coercion among criminal law theorists.)

Zimmerman, D. (1981) 'Coercive Wage Offers', *Philosophy and Public Affairs* 10 (12): 121-45. (An influential article defending an analysis of economic exploitation and coercion, and the differences between the two.)

Cognition, infant

In the past thirty years developmental psychologists have developed techniques for investigating the cognitive resources of infants. These techniques show that an infant's initial representation of the world is richer and more abstract than traditional empiricists supposed. For instance, infants seem to have at least some understanding of distance, of the continued existence of objects which are out of sight, and of the mental states of others. Such results have led philosophers to reconsider the idea - to be found in Plato - that there may be innate constraints on the way we view the world, and to examine the extent to which innate 'knowledge' may be revised as a result of learning.

How do we ask prelinguistic infants what they know? Infants' behaviour may initially look random and confused, but with closer observation, and particularly with the aid of videotape, we have discovered that it is quite systematic and structured. Very young infants show distinct emotional expressions. They act on objects in distinct and appropriate ways. Infants imitate the actions of others. We can use these patterns of action and emotion to test experimentally how infants behave when we present them with new events.

But we can also use other techniques that do not depend on spontaneous behaviour. Infants show characteristic shifts of visual attention and these shifts have been widely used to assess infant knowledge. An infant is first shown some event over a period of time. Infants typically look attentively at the event at first and then their attention declines, a process called habituation. We can then show the infant a new event. If the infants think the new event is different from the earlier event their attention will return; they will, once again, look attentively, a process called *dis*habituation. But if they think the new event is only a minor variant of the original event it will not capture their attention.

Philosophers have puzzled over the gap between our ordinary conception of the physical world as three-dimensional enduring objects moving in space and the information that arrives at our senses. For example, one classic question is how we infer a three-dimensional world from a two-dimensional retinal image (see Empiricism; Phenomenalism; Vision). Newborn infants seem to have at least some understanding of distance. Even very young infants will extend their arms towards objects within reach, but not out of reach. Similarly, newborn infants show size constancy. They recognize that an object that has moved away from them is still the same object, in spite of the fact that the retinal image is smaller.

A similar puzzle about object knowledge is the Molyneux problem (see Molyneux problem). How do we link information about the same object that comes from different sensory modalities? Intermodal mapping is found in very young infants. Newborn infants imitate visually perceived facial gestures, although they cannot see their own faces, suggesting a mapping between tactile and visual experience. Similarly, in one experiment, infants sucked on either a bumpy or smooth object without ever being allowed to see it. Then they saw a visual image of either a bumpy or smooth object without touching it. Infants consistently looked more at the visual image that matched their previous tactile experience. Infants show a similar appreciation of the congruence between the temporal pattern of a sequence of noises and a sequence of flashes.

Yet another question is how we understand that objects endure even when they move out of sight. Infants appear to make some predictions about the possible movements of visible objects. Even very young infants reach for objects in the dark, suggesting that they can also represent objects that are out of sight for at least a brief time. However, infants will not search for objects that are hidden behind screens or under cloths until they are about 9 months old. This is in spite of the fact that they have the motor capacity to search under screens, and can remember seeing the objects being hidden. Their actions suggest that they believe the object no longer exists.

In contrast, some visual attention experiments suggest that young infants can draw inferences about objects behind screens. For example, we can show 5-month-old infants a screen that folds backwards to a flat position until they habituate to that event. Then we can show them an object in the path of the folding screen. The screen either folds flat (apparently going through the space that should have been occupied by the object), or the screen stops when it reaches the appropriate angle, as if blocked by the hidden object. Infants look more at the first event than the second, in spite of the fact that the first event is actually more superficially similar to the habituating event. In the

case of distance and intermodal matching, the action and attention experiments give similar results. However, in this case there appears to be a discrepancy between the findings from studies of action and of attention. Children apparently demonstrate one set of beliefs when they act on objects and another contradictory set of beliefs when they look at objects. There is an ongoing debate about whether the attention-effects imply that infants understand enduring objects.

Another classic problem is how we infer that others have mental states like our own, given only the evidence of their behaviour (see Other minds). Very young infants act in ways that seem coordinated with the actions and emotions of others, and at around 9 months show signs of understanding that mental states are directed at objects. At this point infants begin to point at objects and to follow the gaze and pointings of others (see Mind, child's theory of).

What philosophical conclusions can we draw from these empirical results? One theoretical possibility might be the sort of strong nativism advocated in Plato's *Meno*, in which Socrates wonders how it is that we have abstract and complex representations of the world, when the evidence of our senses is so limited and concrete? He answers the question by conducting a psychological experiment on a child, Meno's slave boy, testing whether he has the abstract representations in question (*Meno* 82b-86b). When it turns out that he does, Socrates concludes that the representations are innate, a view later echoed by Descartes and Kant. In its modern version, the nativist view suggests that there are strong and unchanging innate constraints on the way we view the world, a kind of 'core knowledge' (see Innate knowledge; Nativism). On an alternative view, the ability of infants to learn and change their beliefs is as impressive as their innate endowment. On this view infants begin with rich innate representations of the world, but those representations are subject to often quite radical revision as the infant learns more about the world. On this view, the infant is like a scientist who constructs and revises theories, and the innate representations are initial theories (see Learning). On either view, infants are an important and often neglected source of information about basic epistemological questions.

See also: Innate knowledge; Language, innateness of; Cognitive development; Piaget, J.

ALISON GOPNIK
ANDREW N. MELTZOFF

References and further reading

Baillargeon, R. (1993) 'The Object Concept Revisited: New Directions in the Investigation of Infants' Physical Knowledge', in C. Granrud (ed.) *Visual Perception and Cognition in Infancy*, Hillsdale, NJ: Erlbaum.(A review of work investigating very early object knowledge using the habituation technique.)

Bower, T.G.R. (1982) *Development in Infancy*, San Francisco, CA: W.H. Freeman.(A seminal and philosophically sophisticated (if sometimes controversial) look at infant perception and cognition.)

Butterworth, G. (1991) 'The Ontogeny and Phylogeny of Joint Visual Attention', in A. Whiten (ed.) *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading*, Oxford: Blackwell, 223-32.(Work on early understanding of other minds.)

Cohen, L.B. (1979) 'Our Developing Knowledge of Infant Perception and Cognition', *American Psychologist* 34: 894-9.(A classic review of the habituation technique and work using it to investigate infant categorization.)

Gopnik, A. (1988) 'Conceptual and Semantic Development as Theory Change', *Mind and Language* 3: 197-217. (A theoretical statement of the 'theory theory' of infant cognition.)

Gopnik, A. and Meltzoff, A.N. (1997) *Words, Thoughts and Theories*, Cambridge, MA: MIT Press.(Considers several domains of infant knowledge and their philosophical implications. Chapters review infant object knowledge, infant theory of mind and infant categorization, as well as presenting alternative theoretical accounts of knowledge in infancy.)

Meltzoff, A.N. (1993) 'Molyneux's Babies: Cross-Modal Perception, Imitation, and the Mind of the Preverbal Infant', in N. Eilan, R. McCarthy and B. Brewer (eds) *Spatial Representation: Problems in Philosophy and Psychology*, Cambridge, MA: Blackwell.(Implications of imitation and cross-modal matching research for philosophical problems about objects and minds.)

Piaget, J. (1952) *The Origins of Intelligence in Childhood*, New York: Basic Books. (Classic discussion of infant action and cognition.)

Piaget, J. (1954) *The Construction of Reality in the Child*, New York: Basic Books.(Another classic discussion of infant action and cognition.)

Rovee-Collier, C. (1990) 'The "Memory System" of Prelinguistic Infants', in A. Diamond (ed.) *The Development*

and Neural Bases of Higher Cognitive Functions, Annals of the New York Academy of Sciences, vol. 608, 517-42.(A review of work using the ‘conjugate reinforcement’ technique, in which infants learn to act in order to produce particular types of interesting events, with special reference to memory and categorization.)

Spelke, E. (1990) ‘Origins of Visual Knowledge’, in D. Osherson and L. Smith (eds) *Visual Cognition and Action: An Invitation to Cognitive Science*, vol. 3, Cambridge, MA: MIT Press.(A review of new data on early perception and object knowledge.)

Spelke, E., Breinlinger, K., Macomber, J. and Jacobson, K. (1992) ‘Origins of Knowledge’, *Psychological Review* 4: 605-12.(A theoretical statement of the ‘core knowledge’ view of infant cognition.)

Trevarthen, C. (1979) ‘Communication and Cooperation in Early Infancy: A Description of Primary Intersubjectivity’, in M. Bulllowa (ed.) *Before Speech*, New York: Cambridge University Press, 231-347. (Descriptions of very early social behaviour in infants.)

Wellman, H.W. (1993) ‘Early Understanding of Mind: The Normal Case’, in S. Baron-Cohen, H. Tager-Flusberg and D. Cohen (eds) *Understanding Other Minds: Perspectives from Autism*, Oxford: Oxford University Press. (A review of work on infant ‘theory of mind’.)

Cognitive architecture

Cognitive architecture involves the properties of mental structures and mental mechanisms that do not vary when people have different goals, beliefs, precepts or other cognitive states. A serious computational theory of mind (CTM) requires that the architecture be constrained independently of such states. One consequence of taking the distinction between architecture and representation-governed process seriously is that it provides a reply to those who are sceptical about the role of rules in cognition, on the grounds that following rules leads to an infinite regress: in CTM, rules are executed by the causal structure of the architecture, and hence do not require further rules for following rules.

The notion of cognitive architecture in the context of the computational theory of mind (CTM) comes directly from the notion of computer architecture, which refers to the relatively fixed set of computational resources available to a programmer in designing a program for a given computer system. Among other properties, this includes the type of memory that the computer has, the way it encodes information (the system of symbolic codes or language it uses), the basic operations that are available, and the constraints on the application of these operations (as in serial v. parallel sequencing). The architecture is a functional characterization of the computer system on which the program runs (see Mind, computational theories of).

It is important to bear in mind that computer architecture reflects the physical properties or 'hardware' only indirectly, since the architecture visible to the programmer might itself be simulated in software or firmware. For this reason it is sometimes referred to as the 'functional architecture' or even as the 'architecture of the virtual machine'. Someone writing a program in, say, LISP or C, has available the resources (operations, datastructures and programming constraints) of those languages and is not concerned with their physical instantiations: LISP or C in that case*is* the architecture of the relevant virtual machine.

In theories of cognition, the notion of architecture is particularly important because it represents the dividing line between cognitive and non-cognitive aspects of a model. In CTM, mental processes are modelled as computer programs, but the process of executing these programs itself lies outside the domain of CTM. What makes the computer model run, or generate token instances of behaviour, is the causal structure of the underlying computational mechanism, the computational architecture.

The architecture of computational models was often taken for granted in modelling cognitive processes. In the 1950s, information-processing models were typically expressed in terms of whatever architecture was available or seemed intuitively reasonable - usually a so-called 'von Neumann' architecture which uses a serial fetch-execute computational regime and location-addressable register memory. But in more recent years, under the influence of Simon and Newell's literal interpretation of programs-as-theories (see Simon and Newell 1964; developed more fully in Newell 1990), it became clear that the assumptions one made about the underlying architecture strongly influenced which programs it could implement (not which functions it could compute, since with suitable external memory it could compute any computable function) and hence constituted a strong claim about the nature of mind.

Mapping out the cognitive architecture in detail is important if the CTM is to provide a literal scientific hypothesis about the causes of cognitive behaviour. Programs can only be individuated relative to an architecture. According to the way in which computer scientists and cognitive scientists individuate programs, two different architectures cannot execute the same program, although the universality of such machines guarantees that (subject to memory requirements) they can compute the same function. The kind of 'strong equivalence' of computational models with mental processes that many cognitive scientists demand is not achieved just by providing a detailed characterization of the process in the form of a program. Because the form of such a program depends on the architecture on which it runs, the theory must specify an independently motivated architecture, including a specification of the representational system it uses. In such a fully articulated model the individual operations - as well as the symbolic expressions they operate over - the sequencing discipline that is imposed on the program execution, the memory constraints that it must adhere to, and so on, all constitute empirical claims. Only when one has independently specified both the architecture and the representations can one's computational model lay claims to being 'strongly equivalent' to the cognitive process being modelled (see Pylyshyn 1984).

In recent years there has been a great deal of debate about which class of architecture is the right one for modelling cognition. In particular there have been proposals that the class of computational architectures that operate over syntactic strings, as in the Turing machine or in proof theory, are inappropriate for modelling psychological processes (see Turing machines; Proof theory). Instead, some people have proposed so-called 'connectionist' or 'neural net' architectures, consisting of a network of simple threshold elements that compute certain functions (the precise class being unknown) by passing activation among the elements over weighted links (see Connectionism). Such networks, it is claimed, compute without reading or writing symbolic expressions. The debate raises the philosophic issue of whether mind can or should be modelled as a syntactic engine, computing in a 'language of thought', as is generally assumed in the CTM (see Fodor and Pylyshin 1988; Language of thought).

Distinguishing architecture and representations (including representations of processes in terms of programs which run on that architecture) is important to the philosopher for a number of reasons (see Pylyshyn 1996). For one, it represents an attempt to understand the nature of the relatively fixed mental structures that instantiate psychological processes: the basic cognitive *capacities* of the mind within which representation-governed processes are instantiated. Some of these structures are likely to be universal and perhaps even innate (we put aside for now the issue of how architecture changes, except to note that it does not change in response to new knowledge - by definition it is not 'cognitively penetrable').

Recognizing the role of architecture also helps to resolve a problem about rule following raised by Wittgenstein (1953): how is a system to know how to follow a rule? If by using another rule, this invites an infinite regress (see Wittgenstein, L. §10). In CTM the regress does not arise because the architecture executes the rules and it does not do this by using rule-following rules, but, instead, by virtue of its causal structure. In a computer we do not have to have algorithms for executing algorithms; instead the physical instantiation of the algorithm in the machine, together with the structure of the machine, simply causes the behaviour to unfold. The relevant abstract description of these causal-structural properties constitutes a description of the architecture of the system.

See also: Computability theory; Language of thought; Modularity of mind

ZENON W. PYLYSHYN

References and further reading

Fodor, J.A. and Pylyshyn, Z. (1988) 'Connectionism and Cognitive Architecture', *Cognition* 28: 3-71.(Provides a detailed analysis of the requirements on cognitive architecture and why these are not met by 'neural net' or 'connectionist' proposals.)

Newell, A. (1990) *Unified Theories of Cognition*, Cambridge, MA: Harvard University Press.(The definitive last work by Newell on a proposed cognitive architecture.)

Pylyshyn, Z. (1984) *Computation and Cognition: Towards a Foundation for Cognitive Science*, Cambridge, MA: MIT Press.(Much of this book is concerned with explicating the notion of cognitive architecture and the criterion of 'cognitive impenetrability' by which it is identified. Discusses 'strong equivalence'.)

Pylyshyn, Z. (1996) 'The Study of Cognitive Architecture', in D. Steier and T. Mitchell (eds) *Mind Matters: Contributions to Cognitive Science in Honor of Allen Newell*, Hilldale, NJ: Erlbaum.(A general, as well as historical, discussion of the issue of cognitive architecture, with emphasis on Allen Newell's important contribution - and some criticism of his assumption of architectural homogeneity as well.)

Simon, H.A. and Newell, A. (1964) 'Information Processing in Computers and Man', *American Scientist* 52 (3). (An early, readable statement of the project of taking programs seriously as models of mental processes.)

Wittgenstein, L. (1953) *Philosophical Investigations*, New York: Macmillan, esp. §§198-292.(An influential discussion of rule-following, and the fact that at some point a rule must be followed 'blindly', without recourse to any further rule.)

Cognitive development

Psychological research of the last two decades has produced a surge of surprising results regarding cognitive development in children that has challenged a number of traditional philosophical assumptions about the nature of knowledge. The developing cognitive system seems organized in terms of specific domains, rather than as a general, all-purpose processor to which traditional empiricism was committed. Children's conceptual development is often guided by highly abstract principles and parameters that are present in a child's mind long before concrete details of a domain are encountered. Moreover, the data seem to indicate that development consists predominantly of gradual conceptual enrichment over time, rather than radical change.

Research in cognitive development addresses the nature of thought and perception, focusing specifically on the origins of cognition, and its changes over time. Thus, it addresses such questions as what sorts of innate expectations infants possess about the kinds of things to be learned; whether development is continuous or stage-like in nature; whether it proceeds from more concrete to more abstract forms of thought; and finally, the nature of cognition itself as either a global and general learning system, or a set of more specialized and separable domain-specific systems (see Modularity of mind).

The emergence of the experimental study of cognitive development in the twentieth century is largely attributed to a few key figures, most notably Jean Piaget, Lev Vygotskii and Heinz Werner. All three favoured stage-like transitions in cognitive development that had sweeping influences across all domains of thought. Although significant portions of their theories are no longer widely believed, the phenomena they uncovered still remain the focus of most contemporary experimental work: knowledge of, for example, arithmetic, spatial layout, physical object mechanics and social interaction.

The last two decades of empirical research on these phenomena have produced a surge of new results that support increasingly strong claims about the nature of cognitive development. Contrary to the associationist assumptions of traditional empiricism, it seems to be guided by highly abstract principles and parameters about specific domains of inquiry - for example, language, causal relations among material objects, biology, intentional psychology - long before the domain is actually encountered. Indeed, along the lines of traditional Rationalism, many of them are arguably innate (see Empiricism; Nativism; Rationalism).

Much of this recent work has been influenced by the revolution in linguistics brought about by Noam Chomsky. Chomsky provided convincing evidence that, in the normal human environment, language is unlearnable unless the learner has specific sets of innate expectations about the nature of grammar. For example, children make systematic errors they have never heard spoken, and use syntactic rules that they have never been taught. Although the nature of this language-specific endowment is still actively debated, the vast majority of language acquisition researchers now acknowledge the need for an a innate set of predetermined expectations about linguistic structure (see Language, innateness of).

Following this proposal in linguistics, many psychologists now argue that children possess innate or early-developing predispositions that guide the development of their understanding of other types of knowledge domains, such as the mechanics of physical objects (see Cognition, infant) and human social behaviour (see Mind, child's theory of). As with language, this domain-specific approach has been proposed to explain both children's impressive cognitive abilities, and their puzzling defects. For example, young infants clearly see objects as solid three-dimensional things (solidity) that move in predictable ways (continuity), but fail to understand gravity, even though evidence for all three is roughly equally present (and important) in the environment (Spelke *et al.* 1992). Infants seem to have a 'theory' of physical objects that ignores gravity until much later in development. This theory may not be as elaborate or extensive as that used by adults, but clearly it guides children's developing understanding of the world in ways similar to adults'.

At least by the age of 3 or 4, children also seem to have a special 'teleological' framework for interpreting some phenomena. Preschool children are more prone to ask such questions as 'What's this for?' when asking about a property of a living kind than of a non-living natural kind, such as a rock, about which they are more likely to ask merely 'What is that?', or 'What is that *P* [property]?', without any presumption that it serves any purpose. They

seem to recognize that the features of living kinds are interconnected and designed in some manner to be causally linked in a way not true of non-living natural kinds. This bias towards teleological explanation may be a fundamental component of conceptual development (Keil 1995). Moreover, when told that human beings possess specific biological properties such as eating, sleeping and having hearts, preschool children will show highly structured patterns of generalization of these properties to living as opposed to non-living kinds (Carey 1985).

It is important to note that these explanatory frameworks go beyond mere associative principles. For example, although being curved is a property equally associated with both bananas and boomerangs, it is causal and central only in the latter case. A straight banana is clearly still a banana, but a straight boomerang is a stick (see Medin and Shobin 1988; Behaviourism, methodological and scientific; Learning §2).

Contrary to earlier claims, recent research also suggests that development consists predominantly of gradual conceptual enrichment over time, rather than radical change. Rather than undergoing a series of cognitive metamorphoses as Piaget supposed, the infant's world contains many parameters of understanding that continue to play a role across development, and into the conceptual systems of adults. For example, 6-month-olds understand that objects continue to exist even when they cannot see them (Diamond 1991), and infants have a surprising understanding of numerosity and some aspects of counting (Gallistel and Gelman 1992).

Research on cognitive development is currently being conducted in a manner that crosses traditional research boundaries and more closely follows the spirit of cognitive science. Carey (1991) has found striking parallels between conceptual development in children and conceptual developments in the history of science. Anthropologists find exciting commonalities in what people in all cultures regard as fundamental categories such as living kinds (Atran 1995). The relevance of linguistics to conceptual development is clearly apparent. Philosophical contributions to such efforts come from many directions, including logic and the nature of inductive learning devices (see Inductive inference), conceptual change in science (see Concepts), the formal nature of reasoning (see Common-sense reasoning, theories of), clarifying relations between metaphysics and epistemology, and how the nature of thought might be revealed through language (see Language, philosophy of).

See also: Learning; Moral development

F.C. KEIL
G. GUTHEIL

References and further reading

Atran, S. (1995) 'Causal Constraints on Categories and Categorical Constraints on Biological Reasoning Across Cultures', in E. Sperber, D. Premack and A.J. Premak (eds) *Causal Cognition: A Multidisciplinary Debate*, Oxford: Clarendon Press, 205-33.(Explores universals in biological thought across different cultures and relations to causal thinking.)

Carey, S. (1985) *Conceptual Change in Childhood*, Cambridge, MA: MIT Press. (Proposes a model of conceptual change for biology in which biological thought emerges largely out of a native psychology.)

Carey, S. (1991) 'Knowledge Acquisition: Enrichment or Conceptual Change?', in S. Carey and R. Gelman (eds) *The Epigenesis of Mind: Essays on Biology and Cognition*, Hillsdale, NJ: Erlbaum, 257-91.(An examination of the difficulties of certain conceptual changes - for example, distinguishing heat from temperature - in both childhood development and the history of science.)

Diamond, A. (1991) 'Neuropsychological Insights into the Meaning of Object Concept Development', in S. Carey and R. Gelman (eds) *The Epigenesis of the Mind: Essays on Biology and Cognition*, Hillsdale, NJ: Erlbaum, 67-110.(Discusses how developmental changes of the infant's object concept might be related to underlying maturational changes in brain function.)

Gallistel, C.R. and Gelman, R. (1992) 'Preverbal and Verbal Counting and Numerical Cognition', *Cognition* 44: 43-74.(Discusses how aspects of counting might be present in infancy and come to guide the emergence of verbal counting.)

Keil, F. (1995) 'The Growth of Causal Understanding in Natural Kinds', in E. Sperber, D. Premack and A.J. Premak (eds) *Causal Cognition: A Multidisciplinary Debate*, Oxford: Clarendon Press, 234-62.(Discusses how the emergence of causal understandings of natural kinds can be dramatically different from other sorts of entities such as artefacts.)

Medin, D. and Shobin, E.J. (1988) 'Context and Structure in Conceptual Combination', *Cognitive Psychology* 20: 158-90.(Shows how conceptual combination can reveal the importance of theoretical relations in concepts

in ways that override mere frequency-based information.)
Spelke, E.S., Breinlinger, K., Macomber, J. and Jacobson, K. (1992) 'Origins of Knowledge', *Psychological Review* 95: 605-32.(Presents a systematic survey of how the infant's knowledge of the physical world might develop and how it reflects a rich array of cognitive principles from very early on.)

Cognitive pluralism

Descriptive cognitive pluralism claims that different people, or people in different cultures, go about the business of reasoning (that is, forming and revising beliefs) in significantly different ways. If descriptive cognitive pluralism is true, it lends considerable urgency to the venerable philosophical problem of deciding which strategies of belief formation and revision we ourselves should use. Normative cognitive pluralism claims that various quite different systems of reasoning may all be equally good. Epistemic relativism, which claims that different strategies of reasoning are best for different people, is a species of normative cognitive pluralism. Evaluative-concept pluralism claims that different people in different cultures use very different concepts of cognitive evaluation. Their notions of rationality and justification (or the closest equivalents in their culture) are quite different from ours. If this is right, it poses a prima facie *challenge to a central strategy in analytic epistemology which tries to arbitrate between different systems of reasoning by determining which system best comports with our own concepts of epistemic evaluation.*

1 Descriptive pluralism

‘Cognitive pluralism’ might be used to label at least three distinct though related claims. One of these, which I call ‘descriptive cognitive pluralism’, has been much debated in anthropology, comparative psychology and the history of science. Descriptive pluralism maintains that different people go about the business of cognition - the forming and revising of beliefs and other cognitive states - in significantly different ways. For example, it has been urged that people in certain pre-literate societies think or reason very differently from the way modern, Western, scientifically educated people do. Closer to home, it has been suggested that different individuals or groups in our own society (men versus women, artists versus scientists, well-educated versus poorly educated) form beliefs and solve cognitive problems in markedly different ways - ways that indicate differences in underlying cognitive processes (see Feminist epistemology §4).

The denial of descriptive cognitive pluralism is descriptive monism, the thesis that all people exploit more or less the same cognitive processes. The distinction between descriptive monism and descriptive pluralism is best viewed not as a hard and fast one, but as a matter of degree. No one would deny that people differ from one another to some extent in the speed and cleverness of their inferences; nor would it be denied that in attempting to solve cognitive problems, different people try different strategies first. But if these are the only sorts of cognitive differences to be found among people, descriptive monism will be vindicated. If, on the other hand, it should turn out that different people, different groups or different cultures use radically different ‘psycho-logics’ - that the revising and updating of their cognitive states is governed by substantially different principles - pluralism will have a firm foot in the door. The more radical the differences, the further we will be towards the pluralistic end of the spectrum.

If descriptive cognitive pluralism turns out to be true, it is possible that some of the diversity in reasoning strategies may be due to genetic differences among individuals or groups. But it might also be the case that some of the diversity, or all of it, is attributable to environmental variables that differ in important ways from one culture to another. Some of the rules of reasoning that people internalize and use, or all of them, may be analogous to the grammatical rules that subserve the production and comprehension of sentences in natural languages. The possibility that there is a fair amount of acquired diversity in human cognitive processes, and that patterns of reasoning or cognitive processing are to a substantial degree moulded by cultural influences, adds a certain urgency to one of the more venerable problems of epistemology. For if there are lots of different ways in which the human mind/brain can go about ordering and re-ordering its beliefs and other cognitive states, if different cultures could or do go about the business of reasoning in very different ways, it becomes quite pressing to ask which of these ways should we use and which cognitive processes are the good ones? Here the analogy with grammatical rules breaks down in an illuminating way. Most of us are inclined to think that one language is as good as another. The one you should use is the one spoken and understood by the people around you. By contrast, most of us are not inclined to accept this sort of thorough-going relativism about cognitive processes. If pre-literate tribes or pre-modern scientists or members of a contemporary sub-culture or our own distant descendants think in ways that are quite different from the ways we think, few of us would be inclined to suggest that all of these ways are equally good. Some ways of going about the business of belief revision are better than others. But just what is it

that makes one system of cognitive processes better than another, and how are we to tell which system of reasoning is best? These are among the most basic and the most disquieting questions that epistemology tries to answer.

2 Normative pluralism

Normative cognitive pluralism is not a claim about the cognitive processes people do use; rather it is a claim about good cognitive processes - those that people ought to use. It asserts that there is no unique system of cognitive processes that people should use, because various systems that are very different from each other may all be equally good. The distinction between normative pluralism and normative monism, like the parallel distinction between descriptive notions, is best viewed as a matter of degree, with the monist end of the spectrum urging that all normatively sanctioned systems of cognitive processing are minor variations of one another. The more substantial the differences among normatively sanctioned systems, the further we move in the direction of pluralism. Epistemic relativism is a species of normative cognitive pluralism. An account of what makes a system of reasoning a good one is relativistic if it entails that different systems are good for different people or different groups of people (see Epistemic relativism). Not all pluralistic accounts of good reasoning are relativistic, since some accounts entail that different systems of reasoning are equally good for everyone.

Historically, it is probably true that much of the support for normative pluralism among social scientists derived from the discovery (or putative discovery) of descriptive pluralism, along with a certain ideologically inspired reluctance to pass negative judgments on the traditions or practices of other cultures. But normative pluralism was certainly not the only response to descriptive pluralism among social scientists. Many reacted to the alleged discovery of odd reasoning patterns among pre-modern peoples by insisting on monism at the normative level, and concluding that the reasoning of pre-modern folk was ‘primitive,’ ‘pre-logical’ or otherwise normatively substandard. Among philosophers, both historical and contemporary, normative cognitive pluralism is a minority view. The dominant philosophical view is that there is only one good way to go about the business of reasoning, or at most a small cluster of similar ways.

3 Evaluative-concept pluralism

Evaluative-concept pluralism is a descriptive thesis, not a normative one. It maintains that people’s intuitive concepts of cognitive evaluation, concepts like those that we express with terms like ‘justified’ or ‘rational’, vary significantly from culture to culture. If evaluative-concept pluralism is correct, the terms of cognitive evaluation exploited in other cultures or intellectual traditions may differ in both meaning and extension from the terms of cognitive evaluation that are embedded in our own everyday thought and language.

Though the issue is controversial, some philosophers think that evaluative-concept pluralism poses a serious challenge to the tradition of analytic epistemology. That tradition tries to resolve normative questions in epistemology by analysing or explicating our ordinary concepts of epistemic evaluation. If we want to know whether our own system of cognitive processes is better or worse than some alternative system, the analytic epistemologist proposes that we settle the question by determining which system does a better job of producing beliefs that comport with our concept of justification (or rationality) (see Normative epistemology §1). But the analytic epistemologist typically offers us no reason to think that the notions of evaluation prevailing in our own language and culture are any better than the alternative evaluative notions that might or do prevail in other cultures. In the absence of any reason to think that the locally prevailing notions of epistemic evaluation are superior to the alternatives, it is hard to see why we should care whether the cognitive processes we use are sanctioned by those local evaluative concepts.

Imagine that we have located some exotic culture that does in fact exploit cognitive processes very different from our own, and that the notions of epistemic evaluation embedded in their language also differ from ours. Suppose further that the cognitive processes prevailing in that culture accord quite well with their evaluative notions, while the cognitive processes prevailing in our own culture accord quite well with ours. Would any of this be of any help in deciding which cognitive processes we should use? Without some reason to think that one set of evaluative notions was preferable to the other, it seems clear that for most of us it would be of no help at all.

See also: Feminist epistemology; Gender and science; Postcolonial philosophy of science; Rational beliefs

STEPHEN P. STICH

References and further reading

Hollis, M. and Lukes, S. (eds) (1982) *Rationality and Relativism*, Cambridge, MA: MIT Press.(A collection of essays by philosophers, anthropologists and historians of science debating the evidence for descriptive pluralism and the merits of relativism.)

Hutchins, E. (1980) *Culture and Inference: A Trobriand Case Study*, Cambridge, MA: Harvard University Press. (A sophisticated study in cognitive anthropology that supports descriptive cognitive monism.)

Levy-Bruhl, L. (1966) *Primitive Mentality*, Boston, MA: Beacon Press.(Classical anthropological defence of descriptive pluralism and normative monism.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Chapters 1 and 2 offer an important argument against the possibility of descriptive pluralism.)

Stich, S. (1990) *The Fragmentation of Reason*, Cambridge, MA: MIT Press. (Chapters 1, 4 and 6 are especially relevant and expand on the material in this entry.)

Wilson, B. (ed.) (1979) *Rationality*, Oxford; Blackwell.(Collection of essays by philosophers, anthropologists and others exploring various aspects of cognitive pluralism.)

Cohen, Hermann (1842-1918)

Hermann Cohen was the founder of the Marburg School of Neo-Kantianism and a major influence on twentieth-century Jewish thought. Die Religion der Vernunft aus den Quellen des Judentums (Religion of Reason out of the Sources of Judaism) (1919) is widely credited with the renewal of Jewish religious philosophy. Cohen's philosophy of Judaism is inextricably linked with his general philosophical position. But his system of critical idealism in logic, ethics, aesthetics and psychology did not originally include a philosophy of religion. The mainly Protestant Marburg School in fact regarded Cohen's Jewish philosophy as an insufficient solution to the philosophical problem of human existence and to that of determining the role of religion in human culture. Thinkers who favoured a new, more existentialist approach in Jewish thought, however, saw Cohen's introduction of religion into the system as a daring departure from the confines of philosophical idealism.

Cohen identified the central Jewish contribution to human culture as the development of a religion that unites historical particularity with ethical universality. At the core of this religion of reason is the interdependence of the idea of God and that of the human being. Cohen derives this theme from the Jewish canon through a philosophical analysis based on his transcendental idealism.

1 Life and work

Hermann Cohen was the son of a cantor in the provincial town of Coswig in the state of Anhalt. He abandoned his rabbinical studies at the Jewish Theological Seminary (Breslau) after witnessing the rift between reform and orthodox Jews over the divine revelation of the Mishnah (the ancient code of rabbinic law). Seeking orientation in the conflict between religion and modern consciousness, Cohen turned to philosophy and the sciences. After receiving his doctorate at Halle in 1865, he published studies on mythological, philosophical and aesthetic topics (God and the soul, Plato's theory of Forms), influenced by the then compelling idea of a 'mechanism of consciousness'; but he gradually abandoned the methodologies of Helmholtz and Steinthal and immersed himself in the philosophy of Kant. His 1871 exposition of Kant's first critique, *Kants Theorie der Erfahrung (Kant's Theory of Experience)*, brought him to the centre of the Neo-Kantian movement and led to his appointment as a Privatdozent at Marburg. Subsequent publications, especially *Kants Begründung der Ethik nebst ihren Anwendungen auf Recht, Religion und Geschichte (Kant's Justification of Ethics and its Application to Law, Religion and History)* (1877) and *Das Prinzip der Infinitesimalmethode (The Principle of the Method of Infinitesimals)* (1883) made original contributions to Kantian thought. Cohen's mature philosophical system follows the classic structure of Kant's three critiques but projects a fourth part, on psychology, and seeks to establish a 'critical idealism' faithful in broad outline to the tradition of Plato, Descartes, Leibniz and Kant. Cohen ascribed the continuity of this tradition to a perennial philosophical problematic that arises from the idealism inherent in scientific thought.

The Marburg School, represented by Cohen and his colleague Paul Natorp, attracted students from as far away as St Petersburg (Boris Pasternak) and Madrid (Ortega y Gasset). Those who became philosophers - Ernst Cassirer, Nicolai Hartmann, Heinz Heimsoeth and Hans-Georg Gadamer - gradually emancipated themselves from Marburg Neo-Kantianism. But the tendency of some philosophers to ignore Cohen's contributions expressed ideological biases - thus Heidegger's 1929 verdict of having 'overcome' Neo-Kantianism.

In 1876 Cohen succeeded his mentor Friedrich Albert Lange, author of the *Geschichte des Materialismus und Kritik seiner Bedeutung in der Gegenwart (History of Materialism)*, to a highly visible position as practically the only non-baptized Jew to hold a chair in philosophy at a Prussian university. Three years later he found himself compelled to defend Judaism against the charges of Heinrich von Treitschke that it was alien to the values of German culture. From then on, Cohen's public engagement on behalf of Judaism and Judaic studies increased steadily, extending in time to the renewal of Jewish religious philosophy, a field often neglected among the more historically oriented disciplines of *Wissenschaft des Judentums* (the 'Science of Judaism').

From 1903 Cohen was active in the *Hochschule für die Wissenschaft des Judentums* in Berlin, where, after retiring from Marburg, he lectured on the biblical prophets, the Psalms, the Greek philosophical roots of medieval Jewish thought, Maimonides, Descartes, and philosophical psychology (see Maimonides, M.). Shortly before his death, Cohen also helped found an academy dedicated to the science of Judaism, at the suggestion of his former student

Franz Rosenzweig.

2 'Critical idealism' as the basis of the sciences and the humanities

Unprecedented scientific progress in the later nineteenth century speeded the collapse of German Idealism. Empirical studies of brain function eclipsed intellectual inquiries into the nature of thought, and in many philosophy departments experimental psychology displaced traditional philosophy. Opposing the tide of positivism and materialism, Cohen championed idealism both for what he saw as its ethical implications and because he conceived of idealism as the true groundwork of scientific thought. Like his mentor Lange, Cohen did not malign materialism but saw it as a heuristic principle, itself a product of critical thought, a hypothesis meant to make scientific cognition possible. Materialism itself, then, showed the fertility of critical idealism (see Köhnke 1991).

Cohen's logic, set out in *System der Philosophie, Erster Teil: Logik der reinen Erkenntnis (System of Philosophy Part One: The Logic of Pure Cognition)* (1902), constructs transcendental philosophy rather differently from Kant. Beginning not from the senses and the a priori forms of sensibility and thought but from the science found in 'printed books', Cohen finds knowledge grounded in mathematical principles. These are products not of experience but of pure thought. Only such knowledge is relevant for the logic of cognition. The 'being' it discovers is 'given' only in the mathematical abstractions that we metaphorically call laws of nature. Sense perception, then, including the refined perception of the laboratory, must be relegated to the role of a methodological principle. The limited rules of syllogistic logic (based as they are on grammar rather than mathematical ideas) are similarly relegated to the methodology of research. While sense perception and syllogistic logic are tools in the verification of hypotheses, the hypotheses themselves are the 'origin' (*Ursprung*) of the objects of cognition. Hence, the 'logic of the origin' determines the foundation of being in thought: thought as the origin of being. Reflecting the constantly progressing sciences with their shifting paradigms, the system of categories and judgments represented in Cohen's logic is open-ended - quite a contrast to Kant's efforts to find fixed normative patterns in our thinking as earnest of its objectivity.

There are other deviations: most strikingly, Kant's 'thing-in-itself' (noumena) (see Kant, I. §3) is eliminated as a superfluous dogmatic prejudice. In another notable reform, to avoid psychologistic readings of the Kantian a priori that would confuse the a priori forms of consciousness with innate functions of the brain, Cohen identifies the knowing subject as the 'unity of cultural consciousness' (*Einheit des Kulturbewußtseins*). The quest for a unity underlying the distinct 'directions of culture' (science, law, art) was the ultimate goal of Cohen's system. The transcendental conditions of that unity were to be addressed in the final part (on psychology) of Cohen's philosophical edifice, which he did not live to complete.

Cohen's revision of transcendental logic pivots on the 'principle of the infinitesimal method' whose paradigm is the physicist's reliance not on the senses but on mathematical models (1902: 126). Cognition, we find, begins, counterintuitively, in a 'no-thing' and an 'adventurous detour' (1902: 84) in which reason itself constitutes the object: the non-sensory infinitesimal becomes the origin of all finite reality.

The anti-materialist, anti-determinist implications of Cohen's conception of reality come to the fore in the humanities and ethics. Once it is understood that reality is discovered in judgments, not sense experiences, morality seems far less ephemeral than materialism might seem to make it - provided it can be shown that the ethical concept of a human being is the actual operative principle in some valid mode of discourse. Cohen finds such a 'direction of culture' in law, which must presuppose a concept of the human being transcending the biological. Ethics, then, is constructed analogously to logic, with the human being as the analogue of nature and the laws that regulate humanity's historical and political existence as analogues of the laws of nature. Since ethics seeks the principles of a reality that presupposes the idea of a human being, Cohen's *Ethik des reinen Willen (Ethics of Pure Will)* (1904) becomes a philosophy of law.

Ethics, as 'the teaching about the human being', becomes the 'centre of philosophy' and lays the groundwork for all disciplines dealing with the products of human action (law, economics, the humanities). Overawed by the sciences, the humanities have sometimes mistakenly adopted the biological notion of the human being, but what they require is the ethical notion of a human being: ethics is 'the positive logic of the humanities' (1904: 1).

3 Ethics and the philosophy of religion

Will, action and consciousness are the ‘constitutive concepts’ (1904: 389) of an ethics that rests on the ‘methodology of the exact concepts of the science of law’, to which ethics ‘listens attentively for the sake of its problems of person and action’ (1904: vii). Will, action and consciousness are defined in the end so as to unite all ethical concepts. All the actions of a moral subject become transformations of the ‘pure will’. Beyond the bare idea of will, however, the concepts of freedom and autonomy are presuppositions of the realization of any good. For this reason, Cohen introduces into his ethics traditional Jewish concepts of sin and repentance. Religious rather than philosophical, in his view, these concepts introduce a crucial element of morality that ethics alone cannot provide: the concrete individual. The state, the community, and other such social agencies operate in the ideal time of legal progress. But only the individual can originate moral decisions in actual time.

Ethics, strictly conceived, can address the problem of guilt (*dolus*, *culpa*) but not that of sin (*peccatum*). It cannot deal with concrete human failure. For the law has no power over sin in the sense of guilt but only over culpability (1904: 366-). Ethics similarly lacks a principle of self-transformation. But this principle is found in the biblical and rabbinic idea of ‘repentance’ (*teshuvah*). Religion can address the concrete individual; but for ethics, by its very structure, the concrete individual is a mere fiction.

Ethics and religion share the goal of advancing a humane civilization. But ethics deals with individuals only as members of collectivities. Religion is charged with the reconstitution of particular individuals in psychological or liturgical time, redirecting each toward the future, through a regeneration of the individual moral consciousness, whose existence is endangered by transgression of the law.

Ethics orients the philosophy of religion toward interpreting particular truths of faith in so far as they contribute to the ethical goal. Religion must demonstrate that its principles not only do not contradict the progress of the state towards the ethical ideal but contribute to its realization. Such a formal directive cannot by itself generate particular truths of faith. These arise in the self-transformation of mythology into a religion which each faith must achieve, in so far as it has a ‘share in reason’. Religion continuously transforms myths into ethically justifiable truths of faith. Thus a religion of reason can direct the individual and the community toward moral progress by contributing to the generation of moral consciousness. The task of a philosophy of religion is exposition of the truths of faith of a particular tradition, in so far as they can be conceived as transcending its myths (1904: 337, 388, 586).

The correlation of God and man, fully developed only in *Die Religion der Vernunft aus den Quellen des Judentums (Religion of Reason out of the Sources of Judaism)* (1919), serves to clarify both the historical development and the systematic structure of the truths of the Jewish faith: God and man are correlated as coordinates that progressively determine each other. Reason here aims continually to reconstitute the moral direction of the individual and the community by generating the whole set of moral conditions in the language of the Jewish faith. Accordingly, the religious correlation of God and man must construct the whole apparatus of moral responsibility and law within itself as a condition for the regeneration of individual moral consciousness.

Cohen’s thought revolves around this correlational idea of God, which is, for him, the decisive contribution of Judaism to civilization. Even in *Ethik des reinen Willen*, God is the systematic capstone uniting ethics with logic. Addressing the ‘being of the ought’, ethics has no grounding in reality unless the human spirit has recourse to a principle that unites the present reality (nature) with the ideal of its future. The realization of the ethical ideal in the actual world rests on the assumption of a common origin of both nature and man. Cohen (1919) supports the same idea from within the sources of Judaism that are concretized in the speech-acts of the liturgy.

See also: Logic of ethical discourse; Neo-Kantianism

MICHAEL ZANK

List of works

Cohen, H. (1977-) *Werke*, ed. Hermann-Cohen-Archiv am Philosophischen Seminar der Universität Zürich unter der Leitung von Helmut Holzhey, Hildesheim, Zurich and New York: Georg Olms, 18 vols.(Including one published and six forthcoming volumes of minor writings and a forthcoming new edition of *Religion der Vernunft*, the *Werke* constitute the first comprehensive edition of Cohen’s works, with mostly excellent introductions, comparison of editions, notes, and so on.)

Cohen, H. (1871) *Kants Theorie der Erfahrung (Kant’s Theory of Experience)*, ed. G. Edel, Berlin: Dümmler; 3rd

edn, 1918; 1st and 3rd edns repr. in *Werke*, vol. 1.1—1.3, Hildesheim: Georg Olms, 1987.(Volume 1.1 of the *Werke* edition consists of a reprint of the 3rd edition, Berlin, 1918; vol. 1.2 lists the variations between the earlier editions and offers an index; vol. 1.3 is a reprint of the 1st edition. The variations between the volumes indicate the progress from Cohen's first exposition of Kantian thought to his own mature system of philosophy, masterfully analysed by the editor.)

Cohen, H. (1877) *Kants Begründung der Ethik nebst ihren Anwendungen auf Recht, Religion und Geschichte(Kant's Justification of Ethics and its Application to Law, Religion and History)*, Berlin: Dümmler; 2nd edn, 1910.(The 1st edition of this work presented a considerable step beyond a mere exposition of Kantian thought. It is here that Cohen first 'eliminates' the Kantian 'thing-in-itself' and presents the foundation of his own philosophical approach. The 2nd edition is much revised, especially in its programmatic approach to the relation of ethics to the humanities. This edition is to be reprinted in volume 2 of *Werke*.)

Cohen, H. (1883) *Das Prinzip der Infinitesimalmethode und seine Geschichte: Ein Kapitel zur Grundlegung der Erkenntnisskritik (The Principle of the Method of Infinitesimals and its History: A Chapter in the Foundation of the Critique of Cognition)*, Berlin: Dümmler; 4th edn repr. in *Werke*, vol. 5.1, Hildesheim: Georg Olms, with an introduction by P. Schulthess, 1984.(This study first established the link between the method of scientific validity and the principle of the origin of cognition in thought that became the cornerstone of Cohen (1902).)

Cohen, H. (1889) *Kants Begründung der Ästhetik (Kant's Justification of Aesthetics)*, Berlin: Dümmler.(Rounding out the tripartite structure of studies based on Kant's three critiques, this is Cohen's foray into the field of aesthetics. The work found little resonance at the time. It is to be reprinted in volume 11 of *Werke*.)

Cohen, H. (1896) 'Biographisches Vorwort und Einleitung mit kritischem Nachtrag' (Biographical Preface and Introduction with a Critical Postscript), in F.A. Lange (ed.) *Geschichte des Materialismus (History of Materialism)*, Leipzig: Baedecker, 5th edn; repr. in *Werke*, vol. 5.2, with an introduction by H. Holzhey, Hildesheim: Georg Olms, 1984.(F.A. Lange had been Cohen's benefactor in that he was responsible for Cohen's call to the University of Marburg, where Cohen became Lange's successor. Cohen was responsible for several subsequent editions of Lange's best-selling philosophical study of the history of materialism. Yet in his 'Introduction with a Critical Postscript', revised in 1914, Cohen also criticized Lange and gave a précis of his own philosophical programme. The edition of 1896 was hailed by Vaihinger as 'no less than a new foundation of critical idealism'.)

Cohen, H. (1902) *System der Philosophie, Erster Teil: Logik der reinen Erkenntnis (System of Philosophy Part One: The Logic of Pure Cognition)*, Berlin: Bruno Cassirer; repr. in *Werke*, vol. 6, Hildesheim: Georg Olms, 1977.(The first part of the mature philosophical system which is clearly no longer Kantian but original.)

Cohen, H. (1904) *System der Philosophie, Zweiter Teil: Ethik des reinen Willen (System of Philosophy Part Two: Ethics of Pure Will)*, Berlin: Bruno Cassirer, 1904; repr. in *Werke*, vol. 7, with an English introduction by S.S. Schwarzschild, Hildesheim: Georg Olms, 1981.(A philosophy of the relation between state, community and individuals based on the conceptual tools of jurisprudence. Lays the foundations of a moral philosophy of law and of religion.)

Cohen, H. (1912) *System der Philosophie, Dritter Teil: Ästhetik des reinen Gefühls (System of Philosophy Part Three: Aesthetics of Pure Feeling)*, Berlin: Bruno Cassirer; repr. in *Werke*, vols 8 and 9, with an introduction by G. Wolandt, Hildesheim: Georg Olms, 1982.(Cohen's aesthetics aims to unite natural and moral aspects in the sentiments of *Erhabenheit* (sublimeness) and *Humor* (humour). Walter Benjamin was one of the few appreciative readers of this work. Other contemporaries considered it an unreadable book and a collection of professorial ramblings.)

Cohen, H. (1915) *Der Begriff der Religion im System der Philosophie (The Concept of Religion in the System of Philosophy)*, Gießen: Töpelmann; repr. in *Werke*, vol. 10, ed. H. Holzhey, with an introduction by A. Poma, Hildesheim: Georg Olms, 1996.(In this work, dedicated to the Marburg School's discussions on religion, Cohen first argued for an integration of religion into the system of philosophy without 'sublating' it into ethics.)

Cohen, H. (1919) *Die Religion der Vernunft aus den Quellen des Judentums*, Leipzig: Fock; trans. S. Kaplan with an introduction by L. Strauss, *Religion of Reason out of the Sources of Judaism*, New York: Frederick Ungar, 1972.(The title of the first edition included a definite article before the word 'religion' as if to say Judaism was the only religion of reason. Since this seemed contrary to Cohen's opinion, subsequent editions dropped the definitie article, reading *Religion der Vernunft*. Hailed by Julius Guttmann as having renewed the discipline of Jewish philosophy, this work nevertheless failed to have a serious impact until its current rediscovery as a serious interpretation of a classical religion in the spirit of humanistic philosophy.)

Cohen, H. (1924) *Hermann Cohens Jüdische Schriften (Hermann Cohen's Jewish Writings)*, ed. B. Straub, with an introduction by F. Rosenzweig, Berlin: Schwetschke; partial English translation by E. Jospe, *Reason and*

Hope: Selections from the Jewish Writings of Hermann Cohen, New York: W.W. Norton, 1971. (The forthcoming six volumes of *Kleinere Schriften*, published as volumes 13 through 17 of *Werke*, will do away with the unhelpful division between this collection and the one published in 1928. Contains a wealth of material of interest to students of the history of philosophy and religion in Wilhelminian Germany.)

Cohen, H. (1928) *Schriften zur Philosophie und Zeitgeschichte (Philosophical and Political Writings)*, eds A. Görland and E. Cassirer, Berlin: Akademieverlag, 2 vols.(Contains political and philosophical writings that will be reissued in chronological order and integrated with Cohen (1924) in the *Werke* edition.)

References and further reading

Adelmann, D. (1968) *Einheit des Bewußtseins als Grundproblem der Philosophie Hermann Cohens*, Inaugural-Dissertation, Heidelberg: Ruprecht-Karl-Universität, 1968.(The first attempt to understand Cohen's philosophy under the aspect of the 'unity of the cultural consciousness'.)

Altmann, A. (1962) 'Hermann Cohens Begriff der Korrelation', in H. Tramer (ed.) *Zwei Welten: Siegfried Moses zum Fünfundsiebzigsten Geburtstag*, Tel Aviv, 366-99.(The first challenge to the, until then pervasive, interpretation of Cohen's Jewish thought associated with Franz Rosenzweig's 'Einleitung' (1924) in Hermann Cohen's *Jüdische Schriften*.)

Brandt, R. and Orlik, F. (1993) *Philosophisches Denken - Politisches Wirken. Hermann-Cohen-Kolloquium Marburg 1992*, Hildesheim, Zurich and New York: Georg Olms.(A collection of essays representing contemporary European scholarship on various aspects of Cohen's philosophy.)

Brelage, M. (1965) *Studien zur Transzendendentalphilosophie*, Berlin: de Gruyter.(Evaluates Cohen's theoretical philosophy in the context of the approaches of Husserl, Hartmann, Hönigswald and Heidegger.)

Edel, G. (1988) *Von der Vernunftkritik zur Erkenntnislogik: Die Entwicklung der theoretischen Philosophie Hermann Cohens*, Freiburg: Alber.(The most lucid exposition of the development of Cohen's theoretical philosophy to date. With a bibliography of literature on Cohen.)

Holzhey, H. (1986) *Cohen und Natorp*, vol. 1, *Ursprung und Einheit: Die Geschichte der 'Marburger Schule' als Auseinandersetzung um die Logik des Denkens*, and vol. 2, *Der Marburger Neukantianismus in Quellen: Zeugnisse kritischer Lektüre: Briefe der Marburger: Dokumente zur Philosophiepolitik der Schule*, Basle and Stuttgart: Schwabe.(Written by the founder of the Hermann-Cohen-Archiv at Philosophisches Seminar der Universität Zürich, vol. 1 clarifies the basic differences between the theoretical philosophies of the two 'heads' of the Marburg School; vol. 2 contains documents and comprehensive bibliographies.)

Köhnke, K.C. (1991) *The Rise of Neokantianism: German Academic Philosophy between Idealism and Positivism*, trans. R.J. Hollingdale, Cambridge and New York: Cambridge University Press.(Excellent background for the rise of Neo-Kantianism. Weak in its chapter on Cohen.)

Orlik, F. (1992) *Hermann Cohen (1842-1918): Kantinterpret - Begründer der 'Marburger Schule' - Jüdischer Religionsphilosoph*, Schriften der Universitätsbibliothek Marburg 63, Marburg: Universitätsbibliothek. (Excellent biographical resource.)

Poma, A. (1988) *Filosofia critica di Hermann Cohen*, Milan: Ugo Mursia Editore, 1988; trans. J. Denton, *The Critical Philosophy of Hermann Cohen*, Albany, NY: State University of New York Press, 1997.(A well-researched and elegantly written comprehensive introduction to the philosophy of Hermann Cohen.)

Schwarzschild, S.S. (1975) 'The Tenability of Hermann Cohen's Construction of the Self', *Journal of the History of Philosophy* 13: 361-84.(The pioneer of US scholarship on Cohen addressing the key problem of Cohen's philosophy of religion.)

Wiedebach, H. (1997) *Die Bedeutung der Nationalität für Hermann Cohen*, Hildesheim, Zurich and New York: Georg Olms.(A masterful pioneering study on the political dimension in Cohen's thought.)

Winter, E. (1980) *Ethik und Rechtswissenschaft. Eine historisch-sytematische Untersuchung zur Ethik-Konzeption des Marburger Neukantianismus im Werke Hermann Cohens*, Berlin: Duncker & Humblot.(On Cohen's ethics, compared to contemporary philosophies of law.)

Zank, M. (1994) 'Reconciling Judaism and "Cultural Consciousness": The Idea of Versöhnung in Hermann Cohen's Philosophy of Religion', Ph.D. dissertation, Waltham, MA: Brandeis University.(Analyses the development of Cohen's religious philosophy; with documents and bibliographies.)

Zank, M. (1996) '"The Individual as I" in Hermann Cohen's Jewish Thought', *The Journal of Jewish Thought and Philosophy* 5: 281-96.(A study of the ambiguity of Cohen's late philosophy of religion, between idealism and existentialism.)

Coleridge, Samuel Taylor (1772-1834)

Although much of Coleridge's life and his best critical and creative powers were devoted to the attempt to develop a philosophical system, he is less well known as a philosopher than as a romantic poet. This is partly because many of his writings remained unpublished until recent years; they now shed new light on the extent of his knowledge of intellectual history, and on the significance of his philosophical synthesis.

As a young man, Coleridge was attracted by the materialist philosophies and theories of human nature which had become part of the Enlightenment's 'Science of Man'. These coincided with his support for the drive towards progress and human brotherhood which he thought inspired the French Revolution. At Cambridge (1791-4) religious doubt accompanied his radical politics and he turned from orthodox Christianity to Unitarianism.

Gradually, however, he became dissatisfied with the 'mechanistic' reductive principles of British eighteenth-century thought. His visit to Germany (1798-9) and his subsequent study of German ideas convinced him that here was a spectrum of philosophical insights which was more adequate to the whole *of human nature; one through which 'head and heart' might be reconciled.*

Coleridge's work reflects his experience of a world subject to violent revolutionary upheavals and his sense of widespread intellectual and moral confusion. Becoming convinced in the early years of the nineteenth century of both the intellectual and spiritual value of Christianity, he sought to re-establish a unity between religious faith and experience and critical philosophy. His 'ideal Realism' reconciled elements of Greek and German philosophy with reinterpretations of Judaeo-Christian themes and doctrines, and with the moral lessons he believed history provided. Any sound philosophy must, he insisted, do justice to every *aspect of human nature. He declared that he was not concerned to be 'original' but to provide a new synthesis, and boldly claimed to have been the first to have 'attempted to reduce all knowledges into harmony'; although his copious notes intended for an Opus Maximum, the 'Logosophia', were never organized into publishable form.*

1 Metaphysics

German philosophy showed Coleridge that he was not alone in seeking a new philosophical ground beyond either dogmatic empiricism or ungrounded metaphysical speculation. He was deeply influenced by, for example, Kant, F.H. Jacobi, Fichte and Schelling, but also examined them critically. In general, he believed idealist philosophy had failed to discern the true nature of a 'self' and its inherent relationality. Other central focuses in his work include the dynamic and constitutive nature of Ideas, the primacy of absolute Will over Being or Mind, and the relation of words as 'living powers' to thought and being.

Coleridge's epistemology is inseparable from his moral philosophy. His insistence in the 'Essay on Faith' that conscience is a precondition of true self-consciousness, for example, is related to his theory of Ideas. He is often categorized as belonging to the Platonist and idealist traditions of thought; certainly he rejects the view that Ideas are merely abstract concepts opposed to reality or (as they were for Kant) merely regulative; they are constitutive of the world, and so must never be confused with mere images or objects of thought. He agrees with Plato that ideal and real are not opposed, and with Descartes that all ideas (as opposed to beliefs or opinions) are their own evidence; but his own theory of ideas is coloured by his agreement with Fichte, that if knowledge is possible at all, it begins with an 'act of will'. There has to be an affirmation given to the inner witness of experience and feeling, even in the first step of perception; though at this stage the 'yes' is pure act, not a cognitive response. Like Fichte, he recognized that, without this, philosophy was powerless to refute scepticism; if, on the other hand, ideas have a dynamic of will, then conscience and consciousness must be intimately connected.

Coleridge's theory of ideas is linked to his development of the German distinction between *Vernunft* (Reason) and *Verstand* (Understanding) (see Kant, I. §8; Hegel, G.W.F. §4). Reason is the faculty of Ideas, a unifying, uniquely human faculty which, in its unity with the will, is inaccessible to causal explanation. The Understanding, by contrast, is the faculty of 'speculative intellect'; the power of analysis, of logical deduction, of calculation, quantification and explanation. Reason includes the faculty of imagination, the image of divine creativity. Imagination creates and interprets the symbols through which universal principles or Ideas are revealed in particular form. It recognizes that what appear to the Understanding to be contradictions (for example, a reality

which is both universal and particular) are in fact the necessary polarities by which reality is constituted.

An important theme in Coleridge's writings is that of 'distinction-in-unity'. Drawn from a synthesis of the philosophies of Pythagoras, Plato and the Alexandrian Jew, Philo, with the Christian model of Trinity, this principle provides the basis of his contribution to the debate on the nature of the self and self-consciousness. Distinction-in-unity was not merely a logical, regulative concept but represented dynamic relationship; that through which, for example, the self is constituted. It is also the means of moral development by which individuals become persons. Coleridge argued that the necessary conditions of personhood (relationship, free will, reason and love) require a transcendent source of 'personeity'. His Absolute is a primacy of Will, self-realized in Mind (Idea) and Being. Throughout his mature writings the nature and function of will and 'act' in relation to reason is constantly reassessed.

Coleridge found in Christianity a language through which the principle of self-consciousness was perfectly expressed. The Trinity was representable as 'ipseity' ('this-ness' or 'self-ness'), 'alterity' ('other-ness') and 'community' (the former two realized in relationship). 'Alterity' is symbolized by *Logos* as the 'Word' or Idea of God; God-other-and-the-same, the first principle of relationship. The 'Community' (Spirit) of the Godhead is a self-realizing, self-limiting love of the 'Other' as another Self. These relations, Coleridge argued, constituted the principle of Humanity. The human self is only fully realized in relation to a Thou ('Essay on Faith').

Distinction-in-unity also provided the basis of his philosophy of nature. He agreed with the German *Naturphilosophen* that life must be understood not in terms of organization, mechanisms or configurations of atoms, but as polarities of forces and powers, and their products (see Naturphilosophie). Throughout his life he followed the latest scientific theories and developments in academic journals, and his *Notebooks* reveal strenuous efforts to show how moral and natural philosophy might be reconciled.

Coleridge saw philosophical integrity and linguistic veracity as interdependent. The abuse of language, with its lack of attention to etymology and to the relation of grammar and language to the processes of thought, resulted in intellectual and moral corruption (*Aids To Reflection*, 1825). It was necessary to 'desynonymize'; to understand the unique, living quality of each word, and therefore its true relation to others. Words should reflect their true source - the *Logos*, the incarnate Word of God.

The *Logos*, as the unifying principle of the 'Logosophia', is first an epistemological principle representing the birth of language, of self-consciousness and therefore of philosophy. *Logos* is also an aesthetic medium; as 'Word' it is the archetypal symbol of the revelatory power of words and imagination. As the principle of polarity ('alterity') it is also the 'life and light' of nature's powers, forces and products. Finally, *Logos* may be defined as ideal Humanity (universal and individual).

2 Political philosophy

As a young man, Coleridge was influenced by the political radicalism of friends such as John Thelwall and William Godwin, and by the reforming zeal of Joseph Priestley's Unitarianism (see Priestley, J.). He agreed with the latter's view that Jesus was himself a true radical and a free-thinking teacher whose humanity was devalued by the attribution of divinity. By the mid-1790s he had distanced himself from Godwin's radical atheism and from Thelwall's view of the irrelevance of Christianity to reform. Soon after, he was to become dissatisfied with Priestley's materialism. However, his Bristol Lectures (1795) and his journal *The Watchman* (1796) reflect his continuing admiration for the principles (if not the practical reality) of the French Revolution, his criticism of the established Church, and his critique of property.

Coleridge contributed essays and leaders to newspapers on all aspects of social and political thought: on economics, government, foreign affairs, relations of Church and state, on monarchy, and on representation. He consistently supported the antislavery movement and opposed the bills, acts and political movements that he thought endangered 'nationalty' and the principles of true humanity. As his theory of nationhood developed, based on the relation between 'national church', monarch and state (*On the Constitution of Church and State*, 1830) the tone of Coleridge's political and social criticism changed. His later views are all related to his concept of the state in the widest sense of nation or 'body politic'. This, in its ideal form, contained the unity of state and national church. In any particular historical representation of the idea, national church and state are *opposing* poles. Together with the sovereign, they maintain the harmony and balance of the whole.

Coleridge contrasted ‘party-spirit’ and ‘clanship’ to true ‘national spirit’. On this basis, he denounced the ‘false patriotism’ of the Irish movements for independence and Catholic emancipation. While Roman Catholics were not to be excluded from the Christian church, their fealty to Rome amounted to self-exclusion from the *national* church, and therefore from full citizenship. For this reason, he found that he could not support the 1829 bill for emancipation. Catholics could not, for example, belong to the ‘clerisy’; the leaders who would foster the ideal characteristics of humanity. These would ensure the development of Reason, the religious sense, a philosophical and moral understanding of history, attention to the true meaning of words and a recognition of the state as (potentially) a living unity of Will and Reason rather than a Hobbesian ‘Leviathan’.

‘Nation’, Coleridge agreed with Kant, was not to be identified with ‘people’; he was not a democrat, and did not equate the desires of the aggregate of the people with the interests of the nation. He distrusted the foreshadowings of mass labour movements, as his later notebooks show, declaring them to be the result of bad government which had ignored the moral lessons of history and abandoned the fundamental question of what it is to be human. He vehemently criticized leading politicians and statesmen, such as William Pitt and Charles Grey, for failing to preserve the ‘nationalty’; that is, a reserve for the nation in the form of ‘a *wealth* not consisting of lands, but yet derivative from the land, and rightfully inseparable from the same’. This was the opposite pole to the division of land into ‘propriety’, or ‘hereditable estates’ among individuals; these were ‘the two constituent factors, the opposite, but correspondent and reciprocally supporting, counter-weights, of the *commonwealth*’. He rejected the ‘political economy’ of such thinkers as Adam Smith and Thomas Malthus as ‘a science which begins with *abstractions*, in order to exclude whatever is not subject to a technical calculation’. These ‘abstractions’ it assumes ‘as the whole of human nature’.

Coleridge described the nation as ‘a moral unity, an organic whole’. Its archetype was that of the Jewish people in the time of Moses. Despite his criticism of ‘party-spirit’, he often openly declared the superiority of the English nation; of its constitution, and its development through the creative tension of the opposing principles of ‘progression’ and ‘conservation’. He became increasingly fearful that the nation was under threat of a moral vacuum, but abandoned his earlier political radicalism in favour of philosophical combat against an encroaching ‘mechanism’ that ignored not only the human ideal, but also human nature, need and experience.

See also: German idealism

MARY ANNE PERKINS

List of works

Coleridge, S.T. (1969-) *The Collected Works*, ed. various, Princeton, NJ, and London: Princeton University Press and Routledge.(The grand project of this critical edition, together with the *Letters* and the *Notebooks* will, when complete, represent the whole of Coleridge’s vast achievement, much of it for the first time in published form. Begun at the end of the 1960s, its aim has been to give his writings the widest possible circulation by providing accurate readable transcriptions, and by attempting to make sense of fragments, trace references and allusions by detailed cross-referencing and editorial commentary. The *Collected Works* includes *The Watchman* (vol. 2) and the ‘Essay on Faith’, in the volume *Shorter Works and Fragments*, ed. H. and J.R. de J. Jackson, 1995.)

Coleridge, S.T. (1966-71) *Collected Letters of Samuel Taylor Coleridge*, ed. E.L. Griggs, Oxford: Clarendon Press, 6 vols.(E.L. Grigg’s edition of the *Collected Letters* gives access to Coleridge’s correspondence with family, friends, publishers, fellow scholars, poets and other eminent men and women of his time. They reveal the chronology of his preoccupations, his hopes and fears, the details of his daily life and the development of his thought.)

Coleridge, S.T. (1957-) *The Coleridge Notebooks*, ed. K. Coburn and M. Christensen, Princeton, NJ: Princeton University Press, 5 vols.(*The Coleridge Notebooks* reveal not only the vast range of Coleridge’s reading and his intellectual brilliance and profundity, but show his repeated return to struggle with certain topics and the sources of influence on which he drew. Here are fragments of his thought on the physical sciences, on history, religion, psychology, politics, grammar - all witnessing to his continual attempt to show the ultimate unity of nature and mind, philosophy and religion, thought, feeling and experience.)

References and further reading

Colmer, J. (1959) *Coleridge, Critic of Society*, Oxford: Oxford University Press.(Still one of the best studies of the whole range of Coleridge’s political thought.)

Jasper, D. (ed.) (1986) *The Interpretation of Belief: Coleridge, Schleiermacher and Romanticism*, London: Macmillan.(An interesting compendium of short critical studies of religious, literary and philosophical themes in Coleridge.)

McKusick, J.C. (1986) *Coleridge's Philosophy of Language*, New Haven, CT: Yale University Press.(A lively and stimulating account of this aspect of Coleridge's thought; of moderate difficulty in its specialism.)

Modiano, R. (1985) *Coleridge and the Concept of Nature*, London: Macmillan.(A very good and readable account of the relation of Coleridge's thought to German philosophies of nature.)

Morrow, J. (1990) *Coleridge's Political Thought: Property, Morality and the Limits of Traditional Discourse*, London: Macmillan.(Clear and useful in the relationship which it draws between Coleridge's moral philosophy and his political criticism. A good bibliography.)

Muirhead, J.H. (1930) *Coleridge as Philosopher*, London: Allen & Unwin.(One of the best and most enduring attempts to take Coleridge's philosophical claims seriously; of moderate difficulty.)

Newsome, D. (1974) *Two Classes of Men*, London: John Murray.(Clear and stimulating, based on Coleridge's division of thinkers into Platonists or Aristotelians; accessible.)

Perkins, M.A. (1994) *Coleridge's Philosophy: the Logos as Unifying Principle*, Oxford: Oxford University Press.(A thorough, sometimes demanding overview of Coleridge's thought in the light of his search for unity.)

Skorupski, J. (1993) *English Language Philosophy 1750-1945*, Oxford: Oxford University Press.(The first chapter here is useful).

Uehlein, F.A. (1982) *Die Manifestation des Selbstbewusstseins im konkreten 'Ich bin' (The Manifestation of Self-Consciousness in the Concrete 'I am')*, Hamburg: Felix Meiner Verlag.(One of the best analyses of the role of self-consciousness. Again, intellectually demanding in places and unfortunately not yet in translation.)

Collegium Conimbricense

The Collegium Conimbricense ('Coimbra group') or the Conimbricenses were late sixteenth- and early seventeenth-century Jesuit philosophy professors at the University of Coimbra, specifically in the College of Arts, which in 1555 had been placed under the direction of the Society of Jesus. Encouraged by their religious superiors and especially by Pedro da Fonseca, between 1592 and 1606 the Conimbricenses published five volumes containing eight treatises of commentary on Aristotle. Distributed particularly through the Jesuits, these volumes were widely influential in Europe, America, and the Far East, including Japan and China. On this last, Sommervogel (1891) has cited the seventeenth-century Jesuit, Athanasius Kircher, to the effect that by his time all the Coimbra commentaries had been translated into Chinese.

Chief members of the Collegium Conimbricense were Emmanuel de Goes (1542-97), Cosmas de Magelhães (1551-1624), Balthasar Alvarez (1561-1630), and Sebastian de Couto (1567-1639). In order of appearance, their treatises were: (1) *Commentarii Collegii Conimbricensis Societatis Jesu in octo libros Physicorum Aristotelis Stagyritae (Commentaries…on the Eight Books of Aristotle's Physics)* (1592); (2) *Commentarii Collegii Conimbricensis Societatis Jesu in quatuor libros de Caelo (Commentaries…on the Four Books of On the Heavens)* (1592); (3) *Commentarii Collegii Conimbricensis Societatis Jesu in libros meteorum (Commentaries…on the Books of the Meteorology)* (1592); (4) *Commentarii Collegii Conimbricensis Societatis Jesu in parva naturalia (Commentaries…on the Parva Naturalia)* (1592); (5) *Commentarii Collegii Conimbricensis Societatis Jesu in libros Ethicorum ad Nichomachum (Commentaries…on the Books of the Nicomachean Ethics)* (1595); (6) *Commentarii Collegii Conimbricensis Societatis Jesu in libros de generatione et corruptione Aristotelis Stagiritae (Commentaries…on Aristotle's On Generation and Corruption)* (1595); (7) *Commentarii Collegii Conimbricensis Societatis Jesu in libros de anima Aristotelis Stagiritae (Commentaries…on Aristotle's On the Soul)* (1595) (this work was edited by Magelhães, who added his 'Tractatio aliquot problematum ad quinque sensus spectantium' (A Treatment of Some Problems regarding the Five Senses) and attached a treatise by Alvarez called 'De Anima Separata' (On the Separated Soul), because Aristotle had not discussed the separated soul in the previous books); and (8) *Commentarii Collegii Conimbricensis et Societatis Jesu in Universam Dialecticam Aristotelis (Commentaries…on the Whole Logic of Aristotle)* (1606).

This last volume, which was edited and published by de Couto, shows the Conimbricenses at their best. Through its two main parts, following the style of Fonseca's commentaries on the *Metaphysics* of Aristotle (see Fonseca, P. da §3), they have reviewed individual treatises of the Organon. For each of these, they give the Greek of Aristotle, translate it into Latin, summarize and explain its philosophical doctrine, and then comment on it by raising *questions* which it had occasioned among the scholastics of the sixteenth and earlier centuries. Their scholarship, with regard to Greek, Latin, philosophy and the scholastic tradition before themselves, is quite evident. Evident also is the sincerity and depth of their own philosophical interest.

A good example is the first chapter of their commentary on Aristotle's *De interpretatione*. Entitled *De signo (On the Sign)*, it is more than sixty pages long. While some treatment of signs in Aristotelian commentary was common among the scholastics, these pages of the Conimbricenses represent the first really major seventeenth-century treatise on signs. Five principal issues are raised: (1) the nature and conditions common to signs; (2) the divisions of signs; (3) the signification of spoken words and of writing; (4) whether concepts are the same among all, and (going into the question of a natural primitive language) whether spoken words are different; and (5) whether some concepts in our minds are true or false, and others devoid of truth and falsity. Along the way there are sub-questions about the essence of a sign, the possibility of something being a sign of itself, signs as natural or conventional, signs as formal or instrumental, signs as actual or aptitudinal, the relations involved in signs, and so forth. Although their commentary is a work of logic, the Conimbricenses show awareness of many epistemological, psychological, metaphysical and theological questions which can be raised with regard to signs and signification. They also display a remarkable understanding of the breadth and scope of semiotics itself.

Some of the items which they have touched on in different ways are the following: language, syntactical speech, laughing, nodding, coughing, people talking in their sleep, people lying, people emitting words without thought. They consider the signification of negative words, of syncategorematic words such as 'if', nonsense words like

'Blictri', and words like 'chimera' and 'goat-stag' to which no real things correspond. They are interested in the signs involved in writing and reading, especially voiceless reading. Coupled with a short discussion of the physiological bases of speech and hearing, they treat the relation between deafness and an inability to speak or communicate. Relying on Herodotus' account of King Psamittachus, who deliberately reared children with animals to see what language they would naturally speak, as well as on an account, from Jesuit missionaries at the court of Akbar the Great, Mogul of India, of a similar experiment conducted by that prince in 1596, the Conimbricenses discuss the language of what amount to feral children. Probably also relying on accounts of Portuguese Jesuit missionaries, they briefly speak of the difference between written languages which use pictograms (they mention hieroglyphics, as well as Chinese and Japanese characters) and those which use an alphabet. Connected with this they touch upon the signification of arithmetical numbers as well as astronomical signs. Topics occasioned by the Bible are the language of Adam and Eve, the language of Adam when giving 'proper' names to the animals, the sign of Cain (which following St Jerome they regard as a tremor), the rainbow given to Noah, and the phenomenon of different languages at the Tower of Babel (see Language, Renaissance philosophy of §2). In the area of zoosemiosis, they mention Aristotle's study of the song of birds, and they discuss the formation of words by parrots and magpies, as well as different kinds of communication among brute animals. Against a general background of optics, they discuss images as signs, especially images in mirrors. While it does not fully express their semiotic theory, one very revealing sentence is the following: 'There is nothing which leads to the cognition of anything else which cannot be reduced to some sort of sign' (*In universam dialecticam* II q.2 a.3).

See also: Aristotelianism in the 17th century §2; Aristotelianism, Renaissance; Fonseca, P. da; Language of thought; Language, Renaissance philosophy of

JOHN P. DOYLE

List of works

Commentarii Collegii Conimbricensis Societatis Jesu in octo libros Physicorum Aristotelis Stagyritae (Commentaries…on the Eight Books of Aristotle's Physics), 1592; Lyons, 1594; repr. Hildesheim: Olms, 1984. (Contains commentary questions on change, substance, accidents and other topics treated in Aristotle's *Physics*.)

Commentarii Collegii Conimbricensis Societatis Jesu in libros de generatione et corruptione Aristotelis Stagiritae (Commentaries…on Aristotle's On Generation and Corruption), 1595; Mainz, 1606; repr. Hildesheim: Olms, forthcoming.(Contains commentary questions on coming to be and passing away and other topics treated in Aristotle's *On Generation and Corruption*.)

Commentarii Collegii Conimbricensis Societatis Jesu in libros de anima Aristotelis Stagiritae (Commentaries…on Aristotle's On the Soul), 1595; repr. Lyons: Horatio Cardon, 1627.(Contains Greek text, Latin translation, explanation of chapters of Aristotle's *On the Soul*, with commentary questions on Books 2 and 3 of this work.)

Commentarii Collegii Conimbricensis Societatis Jesu in libros Ethicorum ad Nichomachum (Commentaries…on the Books of the Nicomachean Ethics), 1595; repr. Lyons: Horatio Cardon, 1616.(This work, which is without text or translation of Aristotle, is ninety-five columns long and made up of nine disputations each divided into questions on the *Nicomachean Ethics*.)

Commentarii Collegii Conimbricensis Societatis Jesu in quatuor libros de caelo, meteorologicos, et parva naturalia, Aristotelis Stagiritae (Commentaries…on Aristotle's On the Heavens, Meterology and Parva Naturalia), Cologne: Impensis Lazari Zetzneri, 1603.(Contains commentary questions on three works of Aristotle).

Commentarii Collegii Conimbricensis et Societatis Jesu. In Universam Dialecticam Aristotelis Stagiritae (Commentaries…on the Whole Logic of Aristotle), Cologne, 1607; repr. Hildesheim: Olms, 1976.(The treatise *De signo (On the Sign)* is in the second part. The reprinted 1607 edition, unlike the 1606 edition cited in the main text, does not contain the Greek text.)

References and further reading

Doyle, J.P. (1985) 'The *Conimbricenses* on the Relations Involved in Signs', in *Semiotics 1984*, New York: Plenum Press, 567-76.(Introduction to the sign theory of the Conimbricenses.)

Lohr, C.H. (1987) *Latin Aristotle Commentaries. II Renaissance Authors*, Florence: Olschki, 98-9.(Contains bibliographical information on the commentaries of the Collegium Conimbricense.)

Risse, W. (1964) *Die Logik der Neuzeit (The Logic of the Modern Period)*, Stuttgart and Bad Cannstatt: Frommann, 373-8.(Conimbricenses' logic treated within its historical context.)

Solana, M. (1940) *Historia de la filosofia española. Época del renacimiento (siglo XVI) (History of Spanish Philosophy. The Time of the Renaissance (16th Century))*, Madrid: Asociación Española para el Progreso de las Ciencias, vol. 3, 366-71.(Contains some treatment of the Conimbricenses' doctrine on sensation.)

Sommervogel, C. (1891) *Bibliothèque de la Compagnie de Jésus (Booklist of the Society of Jesus)*, Bruxelles: Oscar Schepens, vol. 2, cols 1273-8.(Referred to in the introduction; contains information about the editions of the Conimbricenses' commentaries.)

Stegmüller, F. (1959) *Filosofia e teologia nas universidades de Coimbra e Evora no seculo XVI (Philosophy and Theology in the Universities of Coimbra and Evora in the 16th Century)*, Coimbra: Universidade de Coimbra, 95-9.(The Conimbricenses treated within a wider context.)

Tavares, S. and Bacalar Oliveira, J. (1967) 'Conimbricensi', in *Enciclopedia Filosofica*, Florence: Sansoni, vol. 1, cols. 1590-1.(Contains information about the diffusion and the methods of the volumes of the Collegium Conimbricense.)

Collier, Arthur (1680-1732)

Arthur Collier was an English parish priest who arrived, independently, at a version of immaterialism strikingly similar to that of Berkeley. In his 1713 work Clavis Universalis ('universal key'), Collier contends that matter exists 'in, or in dependence on' the mind. Like Berkeley, he defends immaterialism as the only alternative to scepticism. He admits that bodies appear to be external, but their apparent or 'quasi' externeity is, he argues, merely the effect of God's will, and not a sign of 'real' externeity or mind-independence. In Part I of the Clavis, Collier argues (as Berkeley had in his New Theory of Vision) that the visible world is not external. In Part II he argues (as Berkeley had in both the Principles and the Three Dialogues) that the external world 'is a being utterly impossible'.

1 Life

Collier was born at Langford Magna, near Salisbury in Wiltshire, England. He entered Pembroke College, Oxford in 1697 and transferred to Balliol in 1698. In 1704 he became rector of Langford Magna, a position he held until his death. While at Langford, Collier may have met John Norris whose *Essay towards a Theory of the Ideal or Intelligible World* (1701/4) he admired. Norris argued that the existence of the external world could not be demonstrated, but he did not deny that it exists. In the *Clavis Universalis* (1713) Collier set out to demonstrate this more radical conclusion.

By 'external', Collier writes, 'I understand the same as is usually understood by the words, absolute, self-existent, independent, etc.' ([1713a] 1909: 6). 'All matter, body, extension, etc. exists in, or in dependence on mind, thought, or perception, and…is not capable of an existence, which is not thus dependent' - dependent, that is, not on any mind in particular, but on 'some mind or other'. There is, in fact, a material world in the mind of God that is numerically distinct from the world perceived by any of God's creatures. Collier insists that he makes 'no doubt or question' of the existence of bodies. But their existence, he contends, is an 'in-existence'. He compares it to the existence of objects seen in a mirror, and to the images that float before a fevered imagination.

2 The visible world

The *Clavis* is divided into two main parts: in the first, Collier, 'content for a while to grant that there is an external world' ([1713a] 1909: 33), argues that the visible world is not external; in the second, he argues that an external world is utterly impossible. The first part begins with the argument that we cannot justly infer real externeity from visible or apparent externeity. Sounds, smells, tastes and heat, which seem to exist 'altogether without' the mind, have been proven by Descartes, Malebranche and Norris to exist entirely within it. Light and colours, objects of sight which seem to be distant from or external to the mind, are not really external at all, any more than the visions seen by prophets, or the delusions of drunkards and madmen. If Apelles imagines a centaur, Collier says, the centaur seems to be external, even though it is, as an object of imagination, wholly within the mind. God can turn the imaginary object into an object of sight, he proposes, without creating an external centaur, simply by *intensifying* Apelles' perception. The faint and languid image of a centaur will then become a vivid perception or object of sense. All the difference between imagining and seeing, Collier contends, is a difference of degree.

Collier offers several reasons why visible objects are not and cannot be external. If we press our eye as we look at the moon, we see two moons, not one. Both moons are equally vivid and they seem to be equally external. It follows, he claims, that either both are external or neither is. But 'we are agreed…that they are not both so' ([1713a] 1909: 30). He therefore concludes that neither moon is external, observing that this pattern of reasoning can be extended to any visible object whatsoever. A second argument purports to show that the moon in the heavens is not the moon that humans see. The moon in the heavens is dark, opaque, spherical and large; the moon that humans see is luminous, flat and 'so little, as to be entirely covered by a shilling' ([1713a] 1909: 33). According to a third argument, a thing can be seen only if it is present to our mind. But an object cannot be present to the mind if it is external to it. Hence a visible object cannot possibly be external.

Part 1 of the *Clavis* concludes with a series of objections and replies. According to one objection, Descartes has shown that the existence of an external world can be derived from the goodness of God. Collier replies that although we are inclined to believe in an external world, the judgment that there is such a world is not involuntary.

If it is mistaken, the blame falls not on God but on ourselves.

3 The external world

Part 2 of the *Clavis* is as relentlessly argumentative as Part 1: Collier presents nine separate arguments against an external world and concludes by replying to three objections. He begins by arguing that because an external world must be invisible (as established in Part 1), it must be unknowable. This is, he admits, no 'demonstration' that there is no external world, but it is 'near of kin to demonstration' and serves the same purpose, because it is an acknowledged rule of reasoning that an unknown thing should never be supposed to exist. Later he considers the objection (attributed to Malebranche) that Scripture assures us of an external world. Collier replies that it does not. He expands on his reply in *A Specimen of True Philosophy* (1730), where he develops an immaterialist interpretation of the opening verse of Genesis.

Collier's second argument is that God would have no reason to create an external world. It is therefore 'extrinsically impossible', or impossible in view of its cause. He goes on to claim, in arguments that resemble Kant's first and second antinomies, that it is also 'intrinsically impossible', or impossible in its own nature (see Kant, I. §8). He argues, for example, that the external world is both finite and infinite. The contradiction can be avoided, he claims, only by denying that the world is external to the mind. Arthur O. Lovejoy and Ethel Bowman are among the commentators who have noticed the resemblance to Kant. As Lovejoy (1908) points out, a German translation of the *Clavis* (bound with a translation of Berkeley's *Dialogues*) was published in 1756, but as Lovejoy admits (and as H.J. de Vleeschauwer (1938) argues in more detail) there is no persuasive evidence that Kant was actually influenced by Collier.

Collier also argues that because the external world is a created thing, it must 'subsist on the mind, or will, or power, of God' ([1713a] 1909: 87). He also argues, much as Berkeley had, that philosophical definitions of matter come to nothing, and that if the world is external, there is no theory of vision compatible with our seeing it. In a brief conclusion, Collier surveys the consequences of his doctrine. The survey is meant to fulfil his promise of a 'universal key' - a way of unlocking 'almost all the general questions that [the reader] has been used to account very difficult, or perhaps indissoluble' ([1713a] 1909: 131) - but apart from observing that his doctrine undermines the Catholic doctrine of transubstantiation, Collier discusses no particular question in detail. He admits that ordinary language is almost wholly built on the supposition of an external world. But he is no more disturbed by this than he was by the objection that the external existence of the visible world enjoys universal consent. Indeed the motto on the title page of the *Clavis*, borrowed from the Latin translation of Malebranche's *Search after Truth*, boldly proclaims that 'the popular approval of an opinion on a difficult matter is a certain sign of falsehood'.

4 Collier and Berkeley

Some scholars suspect that Collier deliberately concealed his debt to Berkeley; most accept his report (in the first paragraph of the *Clavis*) that he arrived at his views ten years before he published them. Collier first refers to Berkeley in letters written in 1714-15. In *A Specimen of True Philosophy* (1730), he writes that 'except a single passage or two' in Berkeley's *Dialogues*, there is no other book 'which I ever heard of in the World' on the same subject as the *Clavis* (1730: 21). This is a puzzling remark on several counts, one being that in the Preface to the *Dialogues*, Berkeley describes both of his earlier books. Robert Benson, Collier's biographer, reports seeing among his papers an outline, dated 1708, on the question of the visible world being without us or not, but says no more about it. Benson concludes that Collier's independence cannot reasonably be doubted; perhaps the outline would, if unearthed, establish this. There is, however, internal evidence for Collier's independence. As Charles J. McCracken (1983) shows, Collier's most important arguments have parallels in Malebranche and John Norris (who are among Collier's acknowledged sources) and in Bayle (who is not). If we are determined to find Collier unoriginal, then we can do so without supposing that he borrowed from Berkeley.

See also: Idealism

KENNETH P. WINKLER

List of Works

Collier, A. (1713a) *Clavis universalis: or, a new inquiry after truth. Being a demonstration of the non-existence, or impossibility, of an external world*, London: printed for R. Gosling; ed. E. Bowman, Chicago, IL: Open

Court, 1909; repr. New York: Garland, 1978.(Collier's defence of immaterialism. The Bowman edition has an introduction and notes.).

Collier, A. (1713b) *Christian principles of obedience to the higher powers. In a sermon preach'd…at the cathedral church of Sarum* [Salisbury], London: Printed for R. Gosling.(On the subjection of Christians to civil authority.)

Collier, A. (1716) *Of justification by faith, as in opposition to justification by works: in a sermon… preached at the cathedral church of Sarum* [Salisbury], London: Printed and sold by J. Downing.(Concludes with some remarks on systems of ethics.)

Collier, A. (1730) *A specimen of true philosophy in a discourse on Genesis: the first chapter and the first verse*, Sarum [Salisbury]: Printed by Charles Hooton.(An immaterialist interpretation of the opening verse of *Genesis*.)

Parr, S. (ed.) (1837) *Metaphysical tracts by English philosophers of the eighteenth century*, London: E. Lumley; repr. Hildersheim: Georg Olms, 1974; repr. Bristol: Thoemmes Press, 1992.(Includes both the *Clavis* and *A Specimen of True Philosophy*.)

References and further reading

Benson, R. (1837) *Memoirs of the life and writings of the Rev. Arthur Collier*, London: E. Lumley; repr. Bristol: Thoemmes Press, 1990.(A biography, with long selections from Collier's correspondence. Chapter 2 contains several letters of philosophical interest. Collier's theological views, examined in Chapter 3, are summarized in an unpublished manuscript that appears as Appendix A.)

Eschenbach, J.C. (1756) *Sammlung der vornehmsten Schriftsteller die die Würklichkeit ihres Eignen Körpers und der ganzen Körperwelt Läugnen (A collection of the leading authors who deny the reality of their own bodies and of the whole corporeal world)*, Rostock: A.F. Röse.(Contains German translations of both the *Clavis* and Berkeley's *Dialogues*, with extensive critical notes.)

Johnston, G.A. (1923) *The Development of Berkeley's Philosophy*, London: Macmillan.(For Berkeley's relation to Collier see Appendix I, pages 360-82. Refers to earlier discussions of Collier by Thomas Reid and Dugald Stewart.)

Lovejoy, A.O. (1908) 'Kant and the English Platonists', in *Essays Philosophical and Psychological in Honor of William James*, New York: Longman, Green, 1908, 265-302.(Kant and Collier are discussed on pages 284-90.)

McCracken, C. (1983) *Malebranche and British Philosophy*, Oxford: Clarendon. (Chapter 5 includes an accessible survey of Collier's main arguments, emphasizing his debt to Malebranche and Norris.)

Vleeschauwer, H.J. de (1938) 'Les antinomies Kantiennes et la *Clavis Universalis* d'Arthur Collier' ('The Kantian antinomies and the *Clavis Universalis* of Arthur Collier'), *Mind* 47: 303-20.(Argues against the suggestion that Collier influenced Kant.)

Collingwood, Robin George (1889-1943)

Collingwood was the greatest British philosopher of history of the twentieth century. His experience as a practising historian of Roman Britain led him to believe that the besetting vice of philosophy is to abstract propositions away from the context of the practical problems and questions that gave rise to them. Until we know the practical context of problems and questions to which a proposition is supposed to be an answer, we do not know what it means. In this respect his concern with the living activities of language users parallels that of the later Wittgenstein. Collingwood also believed that the interpretation of others was not a scientific exercise of fitting their behaviour into a network of generalizations, but a matter of rethinking their thoughts for oneself. His conviction that this ability, which he identified with historical thinking, was the neglected and crucial component of all human thought stamped him as original, or even a maverick, during his own lifetime. He also shared with Wittgenstein the belief that quite apart from containing propositions that can be evaluated as true or false, systems of thought depend upon 'absolute presuppositions', or a framework or scaffolding of ideas that may change with time. The business of metaphysics is to reconstruct the framework that operated at particular periods of history. Collingwood had extensive moral and political interests, and his writings on art, religion and science confirm his stature as one of the greatest polymaths of twentieth-century British philosophy.

1 Overview

Collingwood was born in the Lake District of England, and educated until the age of 13 by his father, a painter and friend of Ruskin. It was during his childhood that he absorbed his admiration for practical problem-solving, whether the problems were those of a painter, an archaeologist or a sailor: 'I learned to think of a picture not as a finished product exposed for the admiration of virtuosi, but as the visible record…of an attempt to solve a definite problem in painting, so far as the attempt has gone' (1939: 2). He subsequently attended Rugby School and University College, Oxford, where he had a brilliant career as a classicist. He was the only pupil of F.J. Haverfield, the Romano-British historian, to survive the First World War, and conceived it his duty to carry on Haverfield's work. So he effectively combined two careers, as a teacher of philosophy at Pembroke College, Oxford, from 1912, and as a practising archaeologist and classicist. In the second capacity he made extensive contributions to the corpus of Roman inscriptions recorded in Britain, and wrote the sections on Roman Britain in the *Oxford History of England*. Although in his *Autobiography* (completed in 1938) he liked to portray himself as a philosophical outsider, his distinction was acknowledged by the Waynflete Chair of metaphysical philosophy, to which he was elected in 1936, and which he held until a series of strokes forced him to retire in 1941.

Collingwood characterized the philosophical situation at the Oxford of his time as one of opposition between the lingering 'school of Green' and the 'realists', whose members included John Cook-Wilson, H.W.B. Joseph and H.A. Prichard. Similar tendencies to realism in Cambridge were represented by G.E. Moore and Bertrand Russell. By the 'school of Green' Collingwood referred to what is often called 'absolute idealism', although he is careful to note that the leading representative of that school in his time, F.H. Bradley, repudiated the label of idealist, as Collingwood himself did. Although educated as a realist, Collingwood locates his progressive disenchantment, in the years before the First World War, in his realization that the historical understanding that the realists brought to philosophy was totally inadequate. Moore criticized Berkeley for views which Berkeley never held; Cook-Wilson criticized Bradley for views that Bradley never held. As a historian Collingwood found himself offended by the lack of historical scholarship demonstrated by such performances. He differs from the analytic tradition most fundamentally over history. If one reads Moore, or Wittgenstein, it might seem that philosophy and history have nothing to do with one another, and if Moore made a historical mistake about Berkeley, that need not diminish the importance of his philosophical thought. But Collingwood sees thought as something that is essentially historically embodied: the idea of a worthwhile philosophy with no reflection on its own historical situation would be a contradiction. We can understand where we are only by understanding our own historical embedding: how we got there and the thoughts of those who would wish us to have got somewhere else. For Collingwood, therefore, there is no split, of the kind that has been made much of in recent philosophy, between treating historical philosophers as great, but unlearned, contemporaries and treating them as historical material: in all the important respects, the past lives in the present and the present cannot be understood without it (see History, philosophy of §4).

2 Major themes

According to Collingwood, the realists of his time thought of knowledge as the static, direct apprehension of the truth of a proposition: a mysterious unity that arises when mind confronts fact. Collingwood, making use of his experience as a working historian, insisted that knowledge is never that, but must be the outcome of a moving, active, dynamic process of hypothesis formation and testing. Questioning is one half of the act of which the other half is forming an answer, and it takes both halves to make up a situation of knowing anything. It should be noticed that this is not an opposition to realism on behalf of any kind of idealism, and the label is of little diagnostic value in attempting to understand Collingwood.

Applying this doctrine to the interpretation of historical texts, Collingwood insists that you cannot know what a philosopher meant by a doctrine until you know the question to which the doctrine was intended as an answer and how that question arose. Immediately it follows that you cannot tell whether propositions contradict each other unless you know that they are answers to the same question. This is partly a plea for intelligent appreciation of the space of problems within which different writers work, and in effect Collingwood is highlighting a version of what later became called the principle of charity (see Charity, principle of §1). Indeed, he sometimes embraces one extreme consequence of a method based on the principle of charity, namely that since we have to read the questions to which philosophers address themselves back from their answers, there is no possibility of saying that their answers are inadequate or muddled or mistaken. Using his favourite analogy of naval history he asks: 'How can we discover what the tactical problem was that Nelson set himself at Trafalgar? Only by studying the tactics he pursued in the battle. We argue back from the solution to the problem' (1939: 70). He himself notes that this works at least mainly because Nelson won the battle, and we may legitimately wonder whether it is right to assume that the great dead philosophers won all their battles. But for Collingwood the integrity of the historical quest requires interpreting them substantially as if they did.

The upshot is that the scope for fundamental criticism of a philosopher diminishes markedly, and this leads to the second theme of constant importance to Collingwood: the identification of metaphysics with history. At first blush it may seem that he must believe that metaphysics, at least as it is done in critical discussion of the views of great past thinkers, has been entirely superseded by the historical problem of discovering the questions and answers they are proposing. However, the twist is that solving this historical problem, for Collingwood, requires reliving the problem and rethinking the issue for oneself. It requires making the words of the text into one's own, re-centring one's concerns and seeing how these things are to be said or these actions performed in answer to them. So if the metaphysical problem gives way to a historical one, it is equally true that a historical problem gives way to a metaphysical one: for example, to understand Plato historically requires fully grasping the problem to which the theory of Forms provides an answer.

Collingwood's *rapprochement* between history and metaphysics here depends upon his insistence that historical thinking is in this way a matter of living through the thinking of the person confronted with a problem. This is perhaps the best-known of his doctrines, and he has several interlocking reasons for it. He holds that to study a person in respect of their mental features is to study their own self-understanding, and that means the concepts that determine their plans and activities. Understanding these concepts is not an atomistic project, a matter of finding individual elements, perhaps written in the brain, connected by scientific law with other elements. It is an essentially holistic enterprise that needs to draw on the wider knowledge of the person's human context. When I come to understand why you acted as you did I am not concerned to place you in a law-like causal network, but to see the point of your doings. In the modern jargon, rationalizing you is a distinct normative activity, not reducible to seeing your behaviour just as part of what generally happens, part of a scientifically repeatable pattern. By the normativity of thought Collingwood means not just that as bystanders we can assess the thoughts of others for truth or falsity, rationality or the reverse. He means that thinking itself is essentially a process of which such assessments are a part. The thinker is actively engaged in solving a problem, and is constantly evaluating proposals, withdrawing some and improving others. What is being done cannot be clocked or recorded or understood in terms of a succession of passive occurrences.

In all his works, from the early *Religion and Philosophy* (1916), he opposes this theory of interpretation to 'scientific' psychology. He sees the scientific psychologist as one who treats the phenomena of mental life as natural events, surveyable from the outside, and who seeks to generate laws of association and development. But, he says, the mind regarded in this way ceases to be a mind at all. He has no objection to this treatment of essentially passive mental phenomena, such as he takes feelings to be. It is when the approach is tried upon

thinking that he objects. He is at his most scathing when describing the grotesque failure of psychologists of his day to engage with thought, even when they purport to be giving its scientific nature. His critique goes to the heart of the question of method in gaining human understanding. For although he calls his epistemology a theory of history, it is what is with less dignity called a theory of folk psychology (see Folk psychology). He is quite clear that he is not only concerned with the remote past: ‘If it is by historical thinking that we re-think and so rediscover the thought of Hammurabi or Solon, it is in the same way that we discover the thought of a friend who writes us a letter, or a stranger who crosses the street’ (1946: 219) - and, he adds, our own thoughts of ten years or five minutes ago. The inclusion of self-knowledge is deliberate and impressive, for Collingwood held that there was no division between knowing one’s own mind and knowing that of others. In this respect his approach is more sophisticated than the *verstehen* method of Wilhelm Dilthey. Its closest successor in current philosophy is the view that ascription of belief and desire to others is best seen as an ‘off-line simulation’ of their active processes of thought. Collingwood’s view that the self-controlling normativity of thought unfits it for being the subject of a self-standing science strikingly anticipates the later discussions of interpretation of others given by writers such as Davidson and Dennett. His hostility to the external, objectifying, professedly ‘scientific’ approach to the human world is as radical as anything to be found in more recent debates.

A final theme that can be isolated as a constant in his work is the dependence of all thought upon absolute presuppositions. Collingwood is not here thinking of the a priori, for, this time anticipating Quine, he has no use for the category. Rather, at a particular time the identification of questions and the production of answers in response to them must go on against a largely unnoticed background of presuppositions. These themselves are not posed in answer to any questions, and therefore cannot be assessed as true or false. To use the analogy with which Wittgenstein in *On Certainty* characterized the same doctrine (which he held for the same reason - the absence of a method for raising and answering the question of truth) they are the hinges on which the door swings. Collingwood has in mind something like the paradigms of Thomas Kuhn: the resources for thought and the devices for structuring it that, in a particular period of time, form the framework within which ordinary investigation proceeds. Again, his prescience in isolating such a category and insisting that historical research be directed at uncovering its operations in the history of science and philosophy has been amply confirmed by later writing.

3 Other works and topics

If one thinks of British philosophy of the 1920s and 1930s as mainly concerned with Russellian and Moorean approaches to mind and matter, or as easing itself into some kind of relationship to logical positivism, then Collingwood’s interests will seem unusual. His first attempt at a system of his own is *Speculum Mentis* (1924), which reviews five forms of experience as modes of discovering the truth. These are art, religion, science, history and philosophy. Art and religion stand at the bottom, since neither aims at expressed knowledge, although religion aims at symbolizing our relationship with the world. Science aims at truth, but its categories are inadequate to capture human experience. History suffers because historians can be seen as spectators of the events that they write down, and their own perspectives are distorting influences. Only in philosophy does the possibility of transcending these partial perspectives arise, and with it an understanding of the relative place of the four inferior modes of experience. The hierarchy of different modes of knowledge is an example of what he later explored as a ‘scale of forms’ or dialectically related set of categories whereby the essence of a phenomenon becomes more perfectly instanced. There are echoes here of the idealism to which Collingwood was not formally committed. His later reaction to this work was that it misunderstood the nature of history: historians are not spectators, but by reliving past thought become one with the histories they are writing (see Oakeshott, M. §2).

The theme of artistic experience is the topic of his justly famous *The Principles of Art* (1938). Collingwood considers, and rejects, several views about the nature of art: that it is craft, representation or imitation, magic or amusement. He finds its essence in expression and imagination, which gives definite form to what is hitherto unconscious. Arguably the least successful part of the work is the attempt to examine what is involved in expression, which leads to the paradoxical doctrine that successful artists achieve their success in their own imagination, while externalizing or expressing what has been imagined is mere craft. This separation of thought and expression seems to witness a surprising and naive separation of mind and body, and his full view may be more complex. He certainly held, for instance, that there is no such thing as an unexpressed emotion (1938: 238), which is hardly consistent with a simple-minded dualism.

Throughout his life Collingwood wanted to put his theoretical concerns into close relationship with practical, moral and political activity. His deep hostility to utilitarianism, especially visible in his *Essays in Political Philosophy* (1989) and *The New Leviathan* (1942), probably originated in the work of the 'school of Green' although it is also fertilized by continental moral philosophy. Proper living, for Collingwood, is not the repeated satisfaction of a stream of arbitrary desires or caprices, but an exercise of rational, free agency, in conformity with duty. Rationality and freedom are equated with knowledge. Collingwood here sympathized not only with Kant, but with the tradition of European liberalism of such writers as Giovanni Gentile and Guido de Ruggiero. For such thinkers the conditions for freedom include an especial sort of community, whose members acknowledge the same freedom in each other, and in which the institutions are organized so as to promote this mutual recognition (see Freedom and liberty §2). Like Benedetto Croce, Collingwood believed that, properly understood, liberalism tempers democracy with aristocracy. In any body politic there will be rulers and ruled; the rulers will be objects of emulation, and therefore under an obligation to comport themselves aristocratically, in the sense of possessing full awareness of the dignity due to their station. In a liberal society the ranks of the rulers would be replenished from those of the ruled, and systems of education and freedom of ideas would be the devices for ensuring the ongoing quality of the intake.

Collingwood's writing is undoubtedly infuriating. Along with passages and doctrines of great depth and interest there are casual formulations of argument and a rather donnish delight in pugnacious overstatement and paradox that have left him easy prey to unsympathetic critics. Nevertheless the depth and range of his thought have seldom been equalled, and the years since his death have only slowly revealed the central importance of his problems, and the interest of his discussions of them.

SIMON BLACKBURN

List of works

Collingwood, R.G. (1916) *Religion and Philosophy*, London: Macmillan.(A juvenile work notable for its containing the equation of history and philosophy.)

Collingwood, R.G. (1924) *Speculum Mentis*, Oxford: Oxford University Press.(The first attempt at a systematic metaphysics, relating different areas of knowledge: art, religion, science, history and philosophy.)

Collingwood, R.G. (1933) *An Essay on Philosophical Method*, Oxford: Oxford University Press. (An attempt to interpret philosophy through the different 'forms' of the categories that it treats; the work is vaguely reminiscent of Kant.)

Collingwood, R.G. (1938) *The Principles of Art*, Oxford: Oxford University Press.(The most influential and readable work on aesthetics in English.)

Collingwood, R.G. (1939) *An Autobiography*, Oxford: Oxford University Press.(A slightly paranoid but amusing exercise in self-interpretation.)

Collingwood, R.G. (1940) *An Essay on Metaphysics*, Oxford: Oxford University Press.(A rejection of positivism in favour of metaphysics as the science of the presuppositions of thought at a time.)

Collingwood, R.G. (1942) *The New Leviathan*, Oxford: Oxford University Press.(Collingwood's only treatise on political philosophy, celebrated for its equation of freedom and knowledge.)

(1945) *The Idea of Nature*, Oxford: Oxford University Press.(A superb and under-appreciated exercise in the history and philosophy of science.)

(1946) *The Idea of History*, Oxford: Oxford University Press.(Collingwood's most widely read and influential work, covering history and historiography from Herodotus to Toynbee and Spengler.)

(1989) *Essays in Political Philosophy*, ed. D. Boucher, Oxford: Clarendon Press.(Diverse essays from 1924 onwards, concerned both with the structure of political association and the underpinnings of civilization.)

References and further reading

Boucher, D. (1989) *The Social and Political Thought of R.G. Collingwood*, New York: Cambridge University Press.(A clear and useful guide to Collingwood's political thought.)

Donagan, A. (1962) *The Later Philosophy of R.G. Collingwood*, Oxford: Clarendon Press.(A predominantly hostile presentation of Collingwood from the standpoint of the following generation of Oxford philosophy.)

Dussen, W.J. van der (1981) *History as a Science: The Philosophy of R.G. Collingwood*, The Hague: Nijhoff.(A scholarly and close reading of the major writings on history, including commentary on his work as a historian.)

Mink, L.O. (1969) *Mind, History, and Dialectic*, Bloomington, IN: Indiana University Press.(Probably the best overall introduction to Collingwood. It includes revealing discussions of the 'scale of forms' and

Collingwood's more general metaphysical standpoint.)

Collins, Anthony (1676-1729)

Anthony Collins was an English freethinker, best-known to philosophers for his reconciliation of liberty and necessity, and his criticisms of Samuel Clarke's arguments for the immateriality and immortality of the soul. 'I was early convinced', Collins wrote, 'that it was my duty to enquire into, and judge for my self about matters of Religion' (1727: 4). This is a fair summary of Collins' lifelong project: to judge the claims of religion as an impartial scientist would judge the claims of a theory, not by 'the Way of Authority*', but by reason or 'the* Way of private Judgment*' (1727: 107). Collins took on almost the whole of religion as affirmed and practised in the eighteenth-century Church of England: its philosophical foundations, its theological doctrines, its methods of scriptural interpretation and its views on the politics of church and state. He found most of it wanting. His conclusions were at least deistic and perhaps atheistic (though he remained a practising member of the established church): the irony of his writing and the reactive character of virtually everything he published make it difficult to know for sure.*

1 Life

Collins was born in Middlesex, England. He studied at Eton and at King's College, Cambridge (though he did not stand for a degree), and later studied law at London's Middle Temple. When he married his first wife in 1698, his father gave him an estate in Essex. There Collins met John Locke, who was living in retirement nearby, and in the last two years of Locke's life they became devoted friends. Locke's rationalism was an important early influence on Collins. A second important personal influence was the French refugee intellectual Pierre Desmaizeaux. Collins also knew John Toland, who visited Collins several times at his country home. Collins travelled to Holland in 1711 and 1713, where he met Jean Le Clerc. Between 1717 and 1724 he played an active role in Essex government, serving, for example as county treasurer.

2 The defence of freethinking

In his first book, *An Essay concerning the Use of Reason, in Propositions whereof the Evidence is Human Testimony* which appeared anonymously in 1707, Collins had worked to establish the authority of reason over matters of religion. He had, for example, rejected the prevailing distinction between what is *agreeable to* reason, what is *contrary to* reason and what is *above* reason, insisting that every belief belongs in one of the first two categories. If a report of a miracle is consistent in itself and is compatible with what we know intuitively or demonstratively, then it is, so far, not contrary to reason. It is agreeable to reason if the witness reporting it is credible, and contrary to reason if not.

Collins' full-scale defence of the authority of reason came in his anonymously published *A discourse of free-thinking* (1713). He defines freethinking as 'the Use of the Understanding, in endeavouring to find out the Meaning of any Proposition whatsoever, in considering the nature of the Evidence for or against it, and in judging of it according to the seeming Force or Weakness of the Evidence' (1713: 45). Everything Collins published in his lifetime performs one or more of the three tasks identified in this definition: determining meaning, assessing evidence and judging in accordance with the evidence. According to Collins' theory of judgment (as presented in his 1717 work *A Philosophical Inquiry concerning Human Liberty*), we cannot help but judge according to the 'apparent degrees of light' - the apparent evidence - that propositions possess. His plea for freedom in judgment, then, is really a plea that our pronouncements not be forced to diverge from what the evidence indicates.

Collins' main argument for freethinking is that truth is good (required by God, useful to society, neither forbidden by God nor hurtful to anyone), and that there is no way to discover it except by applying reason. Against this there was one particularly pressing objection: true opinion (not *reasoned* true opinion) is what salvation depends on, and true opinion can be arrived at by revelation. Reasoning would either reinforce revealed truth (in which case it would be superfluous) or subvert it (in which case it would deprive us of the most vitally important thing). It seems that reasoning is hardly worth the risk. Collins' response to this objection is largely implicit. He distinguishes between 'what God speaks plainly to the whole World' - what God says to the reasoning power with which everyone is endowed - and 'what [some] suppose he has communicated to a few' (1713: 72). He then argues that claims to revelation can be evaluated only by reason. (He takes evident pleasure in documenting priestly disagreements about the interpretation of Scripture - disagreements that reason alone can adjudicate.) What God

speaks plainly to everyone, he insists, is that 'he can require nothing of Men in any Country or Condition of Life, but that whereof he has given them an opportunity of being convinc'd by Evidence and Reason in the Place where they are, and in that Condition of Life to which Birth or any other Chance has directed them' (1713: 30). The good of society is, he argues, 'the *Rule* of whatever is to be allow'd or restrain'd' (1713: 90), and it is a mistake to think that removing restraints on reason would breed disorder.

Richard Bentley's *Remarks upon a Late Discourse of Free-Thinking* (1713) is best-known for its criticisms of Collins' classical and Biblical scholarship, but Bentley also objects to Collins' conception and defence of freethinking. Bentley thinks his definition is trivial, 'for it comprehends the whole herd of Human Race, even Fools, Madmen, and Children; for they use what *Understanding* they have; and judge as things *seem*' (Bentley 1713: 10). Collins' definition is really a definition of the understanding; it contains 'not a Syllable about *Freedom*' (Bentley 1713: 10), and in practice, Collins' so-called freedom is nothing but 'an inward Promptness and Forwardness to decide about Matters beyond the reach of [the freethinkers'] Studies, in *opposition* to the rest of Mankind' (Bentley 1713: 11). Collins' oppositional creed, in Bentley's opinion, amounts to atheism:

> that the Soul is material and mortal, Christianity an imposture, the Scriptures a forgery, the Worship of God superstition, Hell a fable, and Heaven a dream, our life without providence, and our Death without hope like that of Asses and Dogs.
>
> (Bentley 1713: 14)

There is, according to Bentley, an 'occult meaning' in Collins' discussions of disputes about scripture: 'from the variety of Scripture to insinuate none is true' (Bentley 1713: 45), (see Bentley, R.).

3 Liberty and necessity

In his first book Collins briefly defended a version of compatibilism. He rejected the view that liberty is self-determination, 'without regard to any extrinsecal Causes' (1707a: 34). It is, instead, merely 'a Power to do or forbear several Actions, according to the Determination of [the] Mind' (1707a: 35). It follows that liberty is consistent with Necessity, and 'ought not to be oppos'd to it, but only to Compulsion' (1707a: 38).

Collins embraced the doctrine of necessity in his first book ('no one action', he wrote, 'could possibly not happen' - 1707a: 36), but a full defence was presented only in *A Philosophical Inquiry Concerning Human Liberty*. There Collins argues as follows (1717: 57): (1) 'All … actions have a beginning' and (2) 'Whatever has a beginning must have a cause'. It follows that every action has a cause. Collins then claims that: (3) 'Every cause is a necessary cause'. It follows that every action has a necessary cause. Collins supports (3) by arguing against 'the *Epicurean System* of chance' (see Epicureanism §§4, 12).

Collins argues in the *Inquiry* that experience is no proof of a liberty incompatible with necessity. We deceive ourselves by considering actions of which we now repent. Because we are now in a repentant humour, we find no present cause of the action. But the circumstances we were in and the causes we were under made the action 'unavoidably determin'd'. It is this determination, in fact, that is a matter of experience: 'for are we not manifestly determin'd by pleasure or pain, and by what seems reasonable or unreasonable to us, to judge or will or act?' (1717: 24). Collins also argues that necessity is 'the sole foundation of morality and laws' (1717: iv). It is, for one thing, a presupposition of effective punishment, for if human beings were not necessary agents determined by pleasure and pain, the threat of pain could not motivate them. He even argues that a grasp of the very distinction between virtue and vice depends on being a necessary agent.

Collins often speaks against liberty, but his real target, as he sometimes makes clear, is liberty inconsistent with necessity. We have, as he emphasizes at the close of the *Inquiry*, 'a truly valuable *liberty* of another kind' (1717: 116) - the power to do as we will, or as we please. Had we this power or liberty in all things, he concludes, we would be omnipotent. Collins also admits there is a difference between the 'moral necessity' (determination by reason and the senses) to which human beings are subject - and the 'absolute, physical, *or* mechanical necessity' to which stones and clocks are subject. Throughout the *Inquiry*, Collins portrays the hypothesis of necessity as an opinion friendly to religion, and the belief in liberty inconsistent with necessity as atheistic. He claims that even 'most of those who assert *liberty* in words' - liberty inconsistent with necessity, that is - 'deny the thing, when the question is rightly stated' (1717: 111), an observation that anticipates Hume.

In his *Remarks upon a Book Entitled, A Philosophical Enquiry concerning Human Liberty* (1717), Samuel Clarke agreed that every action has a cause, but denied that every cause is a necessary cause. The cause of an action is the 'self-moving power' of the person. Motives and reasons *influence* action, but they do not cause it. Collins' final work on the topic, *A Dissertation on Liberty and Necessity* was published in 1729, after Clarke's death. He claimed that Clarke, in his criticisms of the *Inquiry*, simply assumed that the soul is a self-moving power (see Clarke, S. §2).

4 The immateriality and immortality of the soul

In *A Letter to the Learned Mr. Henry Dodwell* (1707), Collins took aim at Clarke's arguments for the immateriality and immortality of the soul. In *A Letter to Mr. Dodwell*, Clarke had argued that because a material thing is divisible, it cannot be an individual consciousness. A mind or individual consciousness must therefore be immaterial, and if it is immaterial, it is naturally immortal: it cannot be destroyed (or robbed of its essential powers) by any natural process. Collins made several objections. First, he claimed, Clarke simply assumes that thinking is an 'individual power' - a power that can belong only to an indivisible thing. But there are, Collins maintained, powers (such as the sweet smell of a rose) that belong to divisible things, and cannot be reduced to the sum or aggregate of the powers of their parts. Thinking may be such a power, and for all we know, a human being may be a system of matter 'fitly disposed' (as John Locke had put it) to think (see Locke, J. §5). Second, Collins asked, even if Clarke managed to show that an immaterial substance must continue to exist, what proof is there that it must continue to think?

Clarke replied to Collins' *Letter*, and Collins to Clarke's reply. (When the controversy ended, each author had contributed four pieces, in addition to the original statement of Clarke's case in his *Letter to Dodwell*.) The central point at issue was the interpretation and application of Clarke's thesis (offered in reply to the first objection above) that 'every Power or Quality, that is or can be *inherent* in any System of Matter, is nothing else than the Sum or Aggregate of so many Powers or Qualities *of the same kind*, inherent in all its Parts' (Clarke [1707] 1738 vol. 3: 759) (see Clarke, S. §2).

5 Analogical predication

In *A Vindication of the Divine Attributes* (1710), Collins argues against the proposal (made by William King in *Divine predestination and fore-knowledge* [1709]) that wisdom, goodness, justice and mercy are 'attributed to God in the same improper analogous Sense' (1710: 14) as parts or passions. Collins objects that if *all* of God's attributes are analogical, we cannot differentiate between theism and atheism. Nor can we make it our duty to imitate the divine perfections.

6 Biblical interpretation

William Warburton regarded Collins' *A Discourse of the Grounds and Reasons of the Christian Religion* (1724) as one of the most plausible arguments ever made against Christianity (Warburton [1738] 1765 vol. 5: 281). In *The Reasonableness of Christianity* (1695), Locke had put forward as the one essential doctrine of Christianity the belief that Jesus is the Messiah. In Part 1 of the *Grounds and Reasons*, Collins argues that it can be established that Jesus is the Messiah only if it can be shown that Jesus, as described in the Gospels, fulfils the Old Testament prophecies. But the Old and New Testaments, Collins argues, do not agree 'according to the literal and obvious sense' (1724: 39). The Old Testament must therefore be interpreted allegorically - 'in a secondary or typical, or mystical, or allegorical sense' (1724: 53). Collins insinuates that no reasonably controlled method of interpretation could arrive at the desired results. He mockingly presents a long list of rules meant to transform 'the mean literal sense of…words into a noble and spiritual sense' (1724: 59); the climactic rule permits a simultaneous change in the order of words, the addition of new words and the deletion of others.

In Part 2 Collins takes up William Whiston's proposal, in his *Essay towards Restoring the True text of the Old Testament* (1722), that the Old and New Testaments once agreed, and that the apostles argued literally. Since that time, however, the text of the Old Testament has been corrupted. Whiston proposes to establish the text as it stood in the days of Jesus and the apostles; Collins argues that this simply cannot be done (see Hermeneutics, Biblical).

7 Toleration

In *A Letter to the Reverend Dr. Rogers* (1727) Collins presents the following argument, among many others, for toleration:

(1) I judged the Christian religion true and joined the Church of England.
(2) I also judged that if I resided in Spain or Scotland, it would be my duty to worship God publicly, according to the way of the Church of England, and that the magistrate ought to protect me in this.

Collins does not infer the judgment in (2) from the judgment in (1). But (2) is, he suggests, a natural accompaniment of (1). And this suggests that magistrates in England should extend the same toleration to Spanish Catholics or Scottish Presbyterians that those who embrace the Anglican Church would expect from magistrates in Spain or Scotland (see Toleration; Latitudinarianism).

KENNETH P. WINKLER

List of works

Collins, A. (1707a) *An essay concerning the use of reason in propositions, the evidence whereof depends upon human testimony*, London.(Argues for the authority of reason in matters of religion.)

Collins, A. (1707b) *A letter to the learned Mr. Henry Dodwell; containing some remarks on a (pretended) demonstration of the immateriality and natural immortality of the soul, in Mr. Clarke's answer*, London; *A reply to Mr. Clarke's defence of his letter to Mr. Dodwell*, London, 1707; *Reflections on Mr. Clarke's second defence of his letter to Mr. Dodwell*, London, 1707; and *An answer to Mr. Clarke's third defence of his letter to Mr. Dodwell*, London, 1708, in *The Works of Samuel Clarke*, vol. 4, London, 1738, 719-909; repr. New York and London: Garland, 1978.(On Samuel Clarke's proofs of the soul's immateriality and immortality. Clarke's *Works* includes both sides of the controversy.)

Collins, A. (1710) *A vindication of the divine attributes*, London: H. Hills.(A criticism of William King's analogical theology.)

Collins, A. (1713) *A discourse of free-thinking, occasion'd by the rise and growth of a sect call'd free-thinkers*, London; repr. New York: Garland, 1976.(Collins' defence of freethinking.)

Collins, A. (1717) *A philosophical inquiry concerning human liberty*, London: R. Robinson; repr. as *Determinism and Free Will*, ed. J. O'Higgins, The Hague: Martinus Nijhoff, 1976.(Argues that liberty is compatible with necessity. The modern edition contains a detailed introduction and notes.)

Collins, A. (1724) *A discourse of the grounds and reasons of the christian religion*, London; repr. New York: Garland, 1976.(Contends that Jesus cannot be shown to fulfil Old Testament prophecies.)

Collins, A. (1727) *A letter to the reverend Dr. Rogers on occasion of his eight sermons concerning the necessity of divine revelation*, London.(Argues in defence of toleration.)

Collins, A. (1729a) *A dissertation on liberty and necessity*, London: J. Shuckburgh.(A reply to Samuel Clarke's criticisms of 1717.)

Collins, A. (1729b) *A discourse concerning ridicule and irony in writing*, London: J. Brotherton; repr. Los Angeles, CA: William Andrews Clark Memorial Library, University of California, Los Angeles, 1970.(Defends the use of ridicule, irony, mockery, and raillery in writing on religion; observes that defenders of the established church often resort to it.)

References and further reading

Attfield, R. (1977) 'Clarke, Collins and Compounds', *Journal of the History of Philosophy* 15: 45-54.(An account of the Clarke-Collins debate on the immateriality of the soul. Assesses Clarke's 'reductionist premiss' - his claim that the powers or qualities of a system of matter must be the sum or aggregate of qualities of the same sort in the parts.)

Bentley, R. (1713) *Remarks upon a late discourse of free-thinking*, London: J. Morphew & E. Curll.(An attack on the *Discourse of free-thinking*.)

Berkeley, G. (1732) *Alciphron*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, vol. 3, London: Thomas Nelson, 1950.(An attack on freethinking; Collins is discussed - but not named - in the Sixth and Seventh Dialogues.)

Berman, D. (1975) 'Anthony Collins: aspects of his thought and writings', *Hermathena* 119: 49-70.(Supplements the 1970 work by O'Higgins; describes Collins' place in eighteenth-century British philosophy, particularly in relation to Locke and Berkeley.)

Berman, D. (1988) *A History of Atheism in Britain: from Hobbes to Russell*, London: Croom Helm.(Chapter 3

argues, against O'Higgins, that Collins was not a deist but an atheist. On pages 78 to 82 Berman attempts to reconstruct a pantheistic argument for atheism attributed to Collins by Berkeley, who claimed, in the 'Advertisement' to Alciphron, that Collins boasted of such an argument in conversation.)

Clarke, S. (1707) A letter to Mr. Dodwell; wherein all the arguments in his epistolary discourse against the immortality of the soul are particularly answered, London, followed by four later pamphlets defending the argument, in *The works of Samuel Clarke*, vol. 3, London, 1738, 719-909; repr. New York and London: Garland, 1978. (Clarke's argument for the immateriality and immortality of the soul, which replies to Collins' criticisms. The *Works* includes both sides of the controversy.)

Clarke, S. (1717) 'Remarks on a book, entitled, a philosophical enquiry concerning human liberty', in *The Works of Samuel Clarke*, vol. 4, London, 1738, 719-35; repr. New York and London: Garland, 1978.(Criticism of Collins' 1717 work.)

King, W. (1709) *Divine predestination and fore-knowledge, consistent with the freedom of man's will*, Dublin: J. Baker.(King's analogical theology is criticized by Collins in his 1710 work.)

O'Higgins, J. (1970) *Anthony Collins: The Man and His Works*, The Hague: Martinus Nijhoff.(A richly documented study of Collins' life, work and influence on, for example, Voltaire and d'Holbach.)

Rowe, W. (1987) 'Causality and Free Will in the Controversy Between Collins and Clarke', *Journal of the History of Philosophy* 25: 1-67.(A very helpful account of the main points at issue. Defends Clarke against some of Collins' criticisms.)

Stephen, L. (1876) *History of English Thought in the Eighteenth Century*, London: Smith, Elder.(Collins is discussed in chapter 4, pages 201-28.)

Warburton, W. (1738) *The divine legation of Moses demonstrated*, London: A. Millar and J. & R. Tonson, 4th edn, 5 vols, 1765.(Book VI §6 is a criticism of Collins' 1724 work.)

Whiston, W. (1722) *An essay towards restoring the true text of the Old Testament; and for vindicating the citations made thence in the New Testament*, London.(Collins' target in the second part of his 1724 work.)

Colour and qualia

There are two basic philosophical problems about colour. The first concerns the nature of colour itself. That is, what sort of property is it? When I say of the shirt that I am wearing that it is red, what sort of fact about the shirt am I describing? The second problem concerns the nature of colour experience. When I look at the red shirt I have a visual experience with a certain qualitative character - a 'reddish' one. Thus colour seems in some sense to be a property of my sensory experience, as well as a property of my shirt. What sort of mental property is it?

Obviously, the two problems are intimately related. In particular, there is a great deal of controversy over the following question: if we call the first sort of property 'objective colour' and the second 'subjective colour', which of the two, objective or subjective colour, is basic? Or do they both have an independent ontological status?

Most philosophers adhere to the doctrine of physicalism, the view that all objects and events are ultimately constituted by the fundamental physical particles, properties and relations described in physical theory. The phenomena of both objective and subjective colour present problems for physicalism. With respect to objective colour, it is difficult to find any natural physical candidate with which to identify it. Our visual system responds in a similar manner to surfaces that vary along a wide range of physical parameters, even with respect to the reflection of light waves. Yet what could be more obvious than the fact that objects are coloured?

In the case of subjective colour, the principal topic of this entry, there is an even deeper puzzle. It is natural to think of the reddishness of a visual experience - its qualitative character - as an intrinsic property of the experience. Intrinsic properties are distinguished from relational properties in that an object's possession of the former does not depend on its relation to other objects, whereas its possession of the latter does. If subjective colour is intrinsic, then it would seem to be a neural property of a brain state. But what sort of neural property could explain the reddishness of an experience? Furthermore, reduction of subjective colour to a neural property would rule out even the possibility that forms of life with different physiological structures, or intelligent robots, could have experiences of the same qualitative type as our experiences of red. While some philosophers endorse this consequence, many find it quite implausible.

Neural properties seem best suited to explain how certain functions are carried out, and therefore it might seem better to identify subjective colour with the property of playing a certain functional role within the entire cognitive system realized by the brain. This allows the possibility that structures physically different from human brains could support colour experiences of the same type as our own. However, various puzzles undermine the plausibility of this claim. For instance, it seems possible that two people could agree in all their judgments of relative similarity and yet one sees green where the other sees red. If this 'inverted spectrum' case is a genuine logical possibility, as many philosophers advocate, then it appears that subjective colour must not be a matter of functional role, but rather an intrinsic property of experience.

Faced with the dilemmas posed by subjective colour for physicalist doctrine, some philosophers opt for eliminativism, the doctrine that subjective colour is not a genuine, or real, phenomenon after all. On this view the source of the puzzle is a conceptual confusion; a tendency to extend our judgments concerning objective colour, what appear to be intrinsic properties of the surfaces of physical objects, onto the properties of our mental states. Once we see that all that is happening 'inside' is a perceptual judgment concerning the properties of external objects, we will understand why we cannot locate any state or property of the brain with which to identify subjective colour.

The controversy over the nature of subjective colour is part of a wider debate about the subjective aspect of conscious experience more generally. How does the qualitative character of experience - what it is like to see, hear and smell - fit into a physicalist scientific framework? At present all of the options just presented have their adherents, and no general consensus exists.

1 Physicalism and objective colour

The question 'What sort of property is the redness of the shirt?' is sensible only against a background of assumptions regarding what counts as an informative answer. Otherwise, the following naïve response seems appropriate. Colour is one of the features physical objects have, and red is a type of colour. What else is there to

say? What makes the question substantive, however, is the background assumption of the doctrine of physicalism.

Physicalism is a doctrine accepted by a wide variety of philosophers, despite wide disagreement about its detailed characterization. There are two basic tenets: (1) that there is a privileged level of basic physical properties, its precise inventory being the business of physical science to determine; and (2) that all causal transactions in nature occur by virtue of mechanisms that are ultimately realized by the properties of this basic class. This doctrine is sometimes expressed by the slogan, ‘no change without a (basic) physical change’ (see Materialism in the philosophy of mind; Supervenience of the mental).

Acceptance of physicalism generates a problem about colour in the following way. The colour of my shirt - its being red - has causal influence; most notably, on the visual experiences of sentient beings. By the second principle of physicalism there must be physical mechanisms that mediate this influence. So how do these mechanisms relate to the colour?

There seem to be four possible answers. First, colour is itself a basic physical property, akin to mass, charge, charm and the like. Second, colour is reducible to other physical properties, in something like the way that heat is reducible to mean kinetic energy or water is reducible to H_2O (see Reductionism in the philosophy of mind). Third, there is no natural physical property with which to identify colour: colours are mind-dependent properties in the sense that they can only be characterized by reference to their effects on human visual systems. Finally, colours are not real properties of objects at all: in a sense they are an illusion (see Colour, theories of; Property theory).

The first alternative is clearly a non-starter, since colour is not a basic physical property. There is controversy about the second. To explain the controversy, we must review some basic facts about colour vision. What we normally think of as perceived colour is a combination of three properties: hue, brightness and saturation. Hue (essentially the shade of colour itself) is determined by the wavelength composition of the light hitting the eye; brightness is determined by the relative intensity of the light; and saturation is a matter of how much hue relative to white light there is in the stimulus.

Now, it might be thought that a specific colour could be identified with light of a specific wavelength, intensity and saturation. Then the colour of a physical object could be identified with a tendency to reflect just that composition of light. However, it turns out that bundles of light with quite different distributions of wavelengths can produce the same effect on the visual system and therefore are perceived as the same colour. (Such distributions are called ‘metamers’.) Also, there is the phenomenon of colour constancy. As illumination changes, say from bright outdoors to indoors, or noon to late afternoon, the composition of the light reflected from my red shirt changes. Yet it continues to look red throughout. One standard explanation of colour constancy is that the visual system takes changes of illumination into account by comparing the light reflected from many objects at once, thus using the contrast as a major determinant of perceived colour. But this makes the identification of colour with any property of the light reflected from an object quite complicated, if not hopeless.

In response to this problem, David Hilbert (1987) proposes that we identify colour with the complex, but quite objective, property of ‘spectral reflectance’. An object’s spectral reflectance is a function that takes specific distributions of light contained in illuminants as input and specific distributions of light reflected as output. He argues that it is precisely this property that is preserved in cases of colour constancy. What we perceive when we perceive the shirt as red throughout changes of lighting conditions is its surface spectral reflectance. There are various problems with this view, the most notable being the problem of metamers. As mentioned above, two very different light distributions, even against the same background illumination, can yield the same perceived colour. Yet, in Hilbert’s view, since the two objects reflecting these two different distributions have different spectral reflectances, they must count as differently coloured. This seems quite counterintuitive, especially if the background illumination in question is broad daylight. (Hardin (1988) dismisses the spectral reflectance view largely for this reason. Hilbert (1987) responds at length to the objection, and the interested reader should consult his discussion.)

In view of the difficulties with the second alternative, it might be thought that colour is best thought of as a disposition on the part of objects to cause certain experiences in us. So what makes the shirt red is its disposition, however grounded in its physical properties, to cause a normal human observer, under normal conditions, to have a

reddish visual experience. (The historical roots of this position can be found in Locke's discussion of primary and secondary qualities; see Primary-secondary distinction; Locke, J. §4.) Note that on the dispositional view, colour is a real, but mind-dependent property of objects. That is, objects are really coloured, though their being so is dependent on properties of us.

Hardin (1988) disputes the dispositional view on the grounds that there is no principled analysis of the phrases 'normal conditions' and 'normal human observer'. Thus there is no unequivocal way to attribute colours to objects, since as we change observer and conditions we change the object's colour. On his view the appropriate locus of colour attribution is experience: there are chromatic experiences, but no coloured objects. In our terms, there is only subjective colour, no objective colour.

2 Subjective, 'qualitative' colour

The puzzle concerning subjective colour is an instance - perhaps the most discussed instance - of the general puzzle concerning the qualitative, or phenomenal, character of conscious mental states: the puzzle of 'qualia'. We will continue to restrict our attention to colour experience, but most of the positions and arguments we review can be applied to the more general issue of phenomenal experience as a whole. Moreover, when it comes to visual experience, colour provides the most compelling challenge to an account consistent with physicalism. Indeed, we now need to ask the same question about subjective that we asked about objective colour: what sort of physical property could it be?

One obvious way of sorting properties is to distinguish intrinsic from relational properties. To a first approximation, intrinsic properties are those that an object possesses in virtue of conditions wholly within the object. Relational properties are those that an object possesses in virtue of its standing in some relation to other objects. So, for instance, my body's mass is an intrinsic property, whereas its weight is relational, since it depends on my body's location relative to the gravitational field generated by another object. (I weigh less on the moon than on the earth, though my mass is the same.)

Let us call the reddish qualitative character of my visual experience of the shirt 'subjective red'. Is subjective red an intrinsic property of my experience? At first blush, it certainly seems to be. But, given the constraints of physicalism, if subjective red is an intrinsic property of my experience, and if my experience enters causal transactions with other states (or events) in virtue of its qualitative character, then subjective red must be somehow reducible to, or composed of, a physical property of my nervous system.

3 The identity theory and its problems

This latter line of reasoning has led many to what has been called the 'central state identity theory'. On this view, an experience is identifiable with a neural state, and its (intrinsic) properties are ultimately neural in character (see Mind, identity theory of). Three sorts of challenges have been mounted against this identification: (1) the conceivability argument, (2) the multiple-realizability argument, and (3) the knowledge argument.

According to the first, since it seems conceivable that one could be having an experience of a reddish sort but not be in the neural state in question, the experience cannot be the same as the neural state. However, put so baldly, this argument is clearly fallacious. It is conceivable that water might not have been H_2O, but that is no reason to deny that it is in fact identical to it (see Intentionality; Propositional attitude statements). However, the conceivability argument seems to many to be getting at something not so easily dismissed, and this is brought out by considering the second argument, about multi-realizability: subjective red cannot be identical to a neural state because it seems possible that there could be creatures with a different sort of physical constitution (say, Martians, or robots) that could nevertheless experience reddish visual sensations like ours. This seems more interesting than mere conceivability, since the possibility seems to survive the discovery that, among all the instances we have observed, subjective red always happens to be realized by a certain neural state. This does not seem to be the case with, for example, water and H_2O: if all the water we have observed is H_2O, that seems to be a good reason to think that all water is (see Kripke 1980; Essentialism).

The last argument, the knowledge argument, is due to Frank Jackson (1982). If subjective red is a physical property, then if one knows all the relevant physical facts pertaining to colour vision one will know all there is to know about subjective red. Imagine Mary, a vision scientist who possesses the complete physical theory of colour

vision, but who has lived in a totally achromatic environment her entire life; she has only seen the world in black and white. Upon release from this restricted environment, she encounters a ripe tomato, and exclaims, 'So that's what red looks like!'. It seems she has learned something new, and yet, Jackson argues, if physicalism is true she should already have known all there was to know about colour experience. Hence, subjective red is not a physical property.

Now, one reply to this argument is the same as to the conceivability argument. 'Knowledge', another intentional term, is referentially opaque: the fact that Mary knows all the physical facts and yet cannot recognize subjective red from its physical description does not show that subjective red is not identical to a physical property; after all, the fact that she might know all the physical facts about H_2O without recognizing that it is water does not show that water is not H_2O. Mary may, in one sense, know everything there is to know when she learns all the physical facts, just not under every possible description. But that constitutes no problem for physicalism.

Still, however, many philosophers feel that these replies are missing the point. What provides plausibility to the three arguments is that appeal to the neural properties of one's visual state does not seem to *explain* its qualitative properties. The fact that these neurons are firing in this way does not really explain the shirt's looking 'reddish'. This is known as the 'explanatory gap' argument (see Levine 1983, 1993; Explanation).

It could be put this way: once we discover the chemical composition of water, and know various chemical laws governing the states and interactions of physical substances, we can explain the features of water that we initially used in picking it out. For instance, we can explain why it freezes and boils at the temperatures it does, why it is necessary for life, and a host of its other properties. In fact, it is not really conceivable that something could be H_2O and yet fail to manifest these various properties, at least so long as we fix basic chemical laws. This seems to be what at least the conceivability and knowledge arguments are getting at: that the physical facts do not 'upwardly necessitate' facts about qualitative experience in the way that chemical facts necessitate facts about water. No matter how detailed our knowledge of the physical mechanisms by which neurons transmit their impulses from one to the other, we are still left with the question of why all this electrochemical activity should constitute an experience with a reddish (as opposed to a greenish, say) qualitative character. It still seems quite conceivable that such electrochemical activity should occur and yet not the experience of subjective red, and this residual conceivability is a manifestation of the explanatory gap. To put the matter in terms of the knowledge argument, the fact that Mary would learn something new upon emerging from her achromatic world is evidence that the physical theory of colour vision does not really explain the qualitative character. Even if she knew the meaning of ordinary 'red' she could not deduce from the rest of her knowledge that tomatoes look red (in the way that she could deduce that water freezes at $32^\circ F$).

One might object that appeals to neurophysiology can indeed provide an explanation of subjective red. For one thing, if we were to find a stable correlation between subjects' reports of reddish sensations and their occupying certain neurophysiological states, then we could predict the occurrence of these sensations from knowledge of their brain states. For another, knowledge of the neurophysiological properties explains the mechanisms by which subjective red performs its cognitive functions. Colour perception involves selective sensitivity to fairly complicated properties of the light reflected from physical surfaces, and we now know a lot about how that sensitivity is implemented in neural hardware.

The problem with the first consideration is that in place of a genuine explanation all we are given is a brute correlation. I may be able to predict with perfect confidence that when someone occupies this particular neurophysiological state they are having a reddish sensation, but my ability to make this prediction does not manifest an understanding of why this brain state should go with reddish, and not bluish or greenish sensations. Brute facts are just that; they are not explicable. Moreover, they do not support counterfactuals: how are we to judge whether something lacking a specific physical property we possess really would not be having a reddish experience? What could satisfy us about the modal claim that not only are reddish experiences correlated with a certain neural state *in fact*, but*must* be so correlated?

Of course sometimes we admit that certain facts are brute facts; for instance, the value of the gravitational constant, the basic laws of physics, and the like. But notice it is a methodological assumption of current scientific practice, a corollary of physicalism, that only at the basic physical level are brute facts to be found. The idea that

something as complicated as the brain should just give rise to reddish experience, and there be no explanation of this fact, is inconsistent with fundamental Physicalist assumptions.

If there is any hope for explaining subjective colour within a Physicalist framework, it would seem to depend on the second consideration: showing how the neurophysiological states of the brain provide the mechanisms by which colour sensations perform their cognitive functions. But to make this work it is necessary to analyse subjective colour in terms of its function, for otherwise explaining that function would not count as a full explanation of subjective colour itself. In other words, we need to abandon the treatment of subjective colour as an intrinsic property of sensation, and treat it instead as a relational, or functional, property.

4 Functionalism

Functionalism is the view that mental states are definable in terms of their causal roles, their causal relations with stimuli, behaviour and other mental states (see Functionalism). Thus, arguably, something is a belief if and only if it is the result of perception or reasoning and the cause, with desires, of action. So a state would be an experience of subjective red just in case it was normally caused by viewing red things, it tended to cause judgments to the effect that something was red, and it generally related to other mental states - in particular through similarity judgments - in the way that is typical of experiences of red. Let us call the functional role in question ‘functional red’.

A more specific form of the functional/relational approach is to treat qualitative states as essentially representational, and qualitative content as representational content. The idea is this. When I have a reddish experience I am in the state of representing the object I perceive as having a certain property, in this case the property of being red. Immediately, of course, the question arises how to distinguish colour experience from mere colour belief. The standard reply is to appeal to the special functional role of the visual system, for example, by attributing to it some distinctive representational system all its own. The point is that to have a reddish experience is to represent an external object as red in this special visual way. This is a relational theory since subjective red’s identity is determined by its representational content, which is a matter of its relation to external objects, and it is a functional theory, since the particular functional organization of the visual system provides the basis for distinguishing experience from mere judgment.

One thorny problem for the representational approach involves the alleged representational content of the sensation. Above we said that subjective red is a representation that some object is red. But what is ‘being red’? As our earlier discussion showed, it is not easy to specify what property objective redness is, so it is correspondingly difficult to specify what the representational content of subjective red is supposed to be. There are various ways a representationalist can respond. If one believes that colour is identifiable with a spectral reflectance function, then that could be the content of the sensory representation. If, however, one denies the existence of objective colour, one could say that our sensory systems represent objects as having a property that they in fact do not have. This is the so-called ‘error theory’. Finally, one might be subjectivist about the content of subjective red itself. That is, one could say that a reddish experience is a report from the visual system concerning its own state - a way of saying, ‘I’m experiencing redly now’ (see Mental states, adverbial theory of).

Note that it is no objection to the first proposal that the alleged content has a complex theoretical structure and therefore could not plausibly be attributed to our sensory system. There are many representations we employ which in their internal structure are relatively simple and yet they refer to properties or objects which have a complex structure, and about which experts possess sophisticated theories. Take the well-worn example of water. According to most philosophers of language and mind, when I think about water I am thinking about a substance with the structure H_2O. But of course I do not have to know this, and it certainly is not plausible to claim that my mental representation of water contains representations of hydrogen and oxygen. Similarly, subjective red could function to detect, or register, a certain complicated property of an object’s surface without itself possessing complex structure.

5 Objections to functionalism

The basic objection to a functionalist account of subjective colour, or of any qualitative experience, is that it just seems intuitively plausible that functional red and subjective red could come apart. The mismatch goes in both

directions. That is, according to the famous absent and inverted qualia hypotheses, it seems quite possible that a creature could satisfy the conditions for functional red even though not experiencing subjective red, or, for that matter, having any qualitative experience at all. On the other hand, it also seems quite possible that a creature could experience subjective red even though most of the causal relations it normally maintains were absent (this is less emphasized in the literature).

An example of the first sort of problem is the famous ‘inverted spectrum’ thought experiment. As we noted above, colour experience is a function of three basic features: hue, brightness and saturation. Colour quality space can be modelled then as a three-dimensional solid; in fact, a cone. The vertical dimension represents brightness, the horizontal represents saturation, and the circular dimension represents hue. Given this model, there seems to be a mapping of points in the space onto their complements around an axis that bisects the cone through the middle. Reds would be mapped onto greens, blues onto yellows, and so on. The resulting cone would be isomorphic to the original in the sense that all the distance/similarity relations among points would be maintained.

Given this characterization of an inversion, take two creatures, one of whom has a colour quality space described by the original cone and the other of whom has one described by the inverted cone. If we are just looking at the relational properties - those involving causal relations to external stimuli, internal mental states, and behaviour - it seems that the two creatures would be functionally isomorphic. Yet, when looking at a red fire engine one would be having a reddish experience and the other would be having a greenish one. In other words, while they would both satisfy the criteria for functional red, one would be having an experience of subjective red and the other of subjective green. So, subjective red cannot be identical to functional red.

The ‘absent qualia’ hypothesis represents an even more extreme possibility. Ned Block (1980) asks us to imagine the entire nation of China (or any similarly large group of individuals) organized, say, by telephone, so as to realize the functional organization of a human brain. Let us also imagine that this vast network is connected to a robot which has a video camera for an ‘eye’. When the robot is stimulated by light from a fire engine the network will go into a state of functional red, by hypothesis. Yet, it seems quite bizarre to claim that the robot, the network, or the system comprising both, is having a qualitative experience. Thus we have a case of functional red without subjective red (or any qualitative character at all).

An example of subjective red without the normal causal connections could occur if someone’s normal functioning were disturbed, so that various relations between their colour experience and memory, belief and the like no longer held. It seems possible that this could happen while the qualitative character of their experience remained unchanged. One concrete case of this is colour blindness. The fact that someone cannot distinguish red from green obviously affects the structure of their colour space, yet it is not obvious that this makes their experiences of blue any different from mine.

6 Replies

There have been three basic responses to these anti-functionalist arguments. The first is to grant their cogency and claim that for experiences such as subjective red, as opposed to cognitive states such as belief, functionalism is wrong and the traditional identity theory is right. But then one must confront the objections to that theory discussed above: the multiple-realizability argument and the explanatory gap.

The second sort of response is to attempt to undermine the intuitive resistance to a relational account represented by the absent and inverted qualia hypotheses. Numerous such attempts have been made, but we will focus on two related strategies in particular: what might be called the ‘asymmetries’ strategy and the ‘subtle role’ strategy. In the end they are both problematic, but they represent serious attempts to meet the challenge and their problems are instructive.

The basic idea behind the asymmetries strategy is to argue that an appropriately chosen and sufficiently rich relational description can uniquely identify a type of qualitative character, and thereby get around the sorts of counterexamples just discussed. Consider the inverted spectrum argument. As presented above, it seems to make the assumption that there exists a natural axis of symmetry dividing the colour cone. However, there is empirical evidence to the effect that this is not so (see Colour, theories of).

First, there is the question of the location of the primaries. It seems evident that some hues are experienced as

combinations of others, while some seem not to be. Contrast a pure red or green with orange or purple. One might have thought that pure examples of each of these primaries would occupy points in the three-dimensional colour solid equidistant from the origin, representing equal amounts of brightness and saturation, and equidistant from complex hues. But not so. Yellow, for instance, occupies a point of higher perceived brightness than do the other primaries. Also, some primaries occupy a larger region than do others; that is, there are more steps to be taken before we describe what we see as a combined hue. Furthermore, there is the anomaly of brown, which results from darkening yellow, yet is not thought of as merely dark yellow but another hue altogether. The point is that if you were to invert along the red-green and blue-yellow axis, it does not look as if all the relational judgments would be the same. Normals would consider dark and light blues to have the same hue, but inverts would not. Normals and inverts would differ in when they considered hues to be combinations, as opposed to primaries. Functional identity between normals and inverts could not be maintained. Hence, they would not be functionally identical to us, and spectrum inversion would not constitute a counterexample to functionalism. (See Harrison 1973; Clark 1993; Hardin 1988 for arguments along these lines.)

The second, ‘subtle role’ strategy is aimed more at the absent qualia hypothesis. The idea here is to attack the notion that a creature could be functionally identical to us and yet lack qualia by demonstrating that qualitative character is itself essential to normal functioning. Van Gulick (1993) notes that there is some evidence, for example, from blindsight cases - in which a patient claims not to see anything within a certain region of their visual field and yet can correctly ‘guess’, upon prompting, whether or not something is there - that consciousness is necessary for carrying out certain functions (see Consciousness §7; Unconscious mental states). For instance, blindsight patients tend not to initiate action with respect to the objects that they can passively detect in their blind field. If, as such cases suggest, consciousness is essential to the performance of certain functional roles, then it is not possible for a non-conscious state to play the same functional role. Hence, absent qualia are not possible.

We said above that there are problems with both strategies. With respect to the asymmetries strategy, there are two basic objections. First, even if the geometric space that characterizes our colour experience is in fact asymmetrical in the way described above, this seems to be a contingent feature of it. Why should this matter? So long as it is not demonstrated that it is essential to qualitative experience that it be immune to inversion, we have no reason to believe that all forms of qualia will have this feature. But if any form of qualitative experience is susceptible to inversion, functionalism is in trouble.

Moreover, as suggested earlier by the case of red-green colour blindness, there could easily be cases of individuals whose colour experience is somewhat, but not completely, different from our own. Suppose that theirs is symmetrical, and therefore subject to inversion. Would it not make sense to say that when such a person looked at a red fire engine they were having an experience that was more like our subjective red than, say, our subjective green? But, if inversion were possible for them, we could not say this solely on the basis of their functional organization.

The second problem with the asymmetrical approach is that it does nothing to address the absent qualia hypothesis. Suppose that any inversion of primaries would be detectable. Still, given cases like Block’s China-head, it seems possible that a creature could realize a structure onto which the complete set of similarity relations definitive of the colour solid could be mapped, and yet there be *nothing* it is like to be that creature. Addressing the inverted qualia argument alone does not save functionalism.

The ‘subtle role’ strategy is, of course, specifically addressed to absent qualia. But here too there is a problem. Suppose we have to grant that there may be jobs, or roles, that only qualitative states can fill. That still does not mean that filling that role is what it is to be a qualitative experience. In fact, that very way of putting it - that being qualitative is essential, or necessary to playing the role - seems to imply just the reverse; that being qualitative is one thing, playing the role quite another. Suppose it turned out that being red was essential to some plant’s playing the ecological role it played; nothing that was not red could do it. We would not say that being red is to play that role, but rather that being red is what makes the plant in question especially suited to play that role. It seems that the same goes for subjective red - the qualitative experience of seeing red - in the scenario Van Gulick envisaged.

Nor is it clear how the ‘subtle role’ strategy addresses Block’s counterexample concerning the nation of China. Either one would have to admit that the entire nation experienced subjective red, or one would have to argue that the lack of qualitative character in this case posed an insuperable obstacle to realizing the relevant functional

description. Both positions are difficult to defend. However, some (see Lycan 1987) have argued that there are other constraints - for instance, extremely fine-grained functional descriptions couched in teleological terms - that would rule out the nation of China as a legitimate realization (see Functional explanation).

In discussing the problems for functional accounts of colour qualia, we have not distinguished among various forms of functionalism. In particular, we have not addressed the representational version of functionalism, according to which colour qualia are functionally specific forms of representations of colour properties. It is sufficient to note here that the absent and inverted qualia hypotheses cause problems for this version of functionalism as much as any other. For suppose that an inverted spectrum is possible. Then two states could differ in their qualitative character and yet carry the same information concerning the reflectance properties of distal stimuli. Or, if absent qualia are possible, then a state could carry information concerning the reflectance properties of a distal stimulus and yet there be nothing at all it is like to occupy the state. If either case is possible, one cannot identify qualitative character with the representation of such properties (see Semantics, informational).

7 Eliminativism

We have seen that functional/relational analyses of colour qualia face serious difficulties. We have also seen that treating colour qualia as intrinsic properties also encounters a serious obstacle, in the form of the explanatory gap. Of course there are possible replies to these difficulties, and many contemporary philosophers hold one or the other of these views. Yet, the persistence of objections to the various accounts of qualia has led some philosophers to embrace eliminativism about subjective colour (whether or not they, like Hardin, are eliminativists about the objective sort). That is, they see the philosophical problem of qualia as our attempting to identify a real phenomenon in the world which corresponds to a conception to which no real phenomenon does, nor even could, correspond. Belief in qualia is something like belief in ghosts. Just as we do not bother to find some real phenomenon to serve as the referent of our idea of a ghost, so too we should abandon the idea that a real phenomenon could serve as the referent of our idea of subjective red (see Eliminativism).

Eliminativism is quite naturally joined to the representational position just described. No one literally denies that we experience sensations of colour. That the fire engine looks red to me is a genuine fact. However, the eliminativist teases apart what is genuine about this fact from what is illusory in the following manner. That I see, and therefore judge, the fire engine to be a certain way, is undeniable. But it is wrong to infer from my perceptual judgment concerning the surface of the fire engine to the existence of a colour-like property of my experience itself - a colour *quale*. It is the existence of this inner property - around which inversion and absence puzzles take hold - that the eliminativist denies.

A major source for eliminativist sentiment is Wittgenstein’s famous attack on the coherence of any notion of inner experience that did not manifest itself in outwardly observable behaviour (see Criteria; Private language argument). While Wittgenstein had many targets in mind with his ‘private language argument’ - his argument against the possibility of a language that only one person could possibly understand - certainly the idea that I could introspect and discover qualia as properties of my inner experiences was one of them. A number of philosophers have extended this Wittgensteinian critique, most notably Daniel Dennett (1991). As Dennett characterizes them, believers in qualia are tied to a picture of the mind as a theatre (the ‘Cartesian theatre’), in which mental entities are on display before the mind’s eye. He has coined the term ‘figment’ to capture the unjustified inference from the reality of colours as properties of physical objects to the reality of colour qualia as properties of internal states.

Suggestive as this diagnosis of the problem of colour qualia is, it alone does not really constitute an argument in favour of eliminativism. However, various further arguments have been presented in the recent literature. First, the believer in qualia is accused of allegiance to a dualist metaphysics. Second, qualia are disparaged as posits of common-sense psychology, a theory that, like many common-sense theories of the world, must give way before its more rigorous and detailed scientific competitors. Third, thought experiments like the inverted and absent qualia hypotheses are attacked as internally incoherent. Finally, belief in qualia is alleged to lead to unsavoury sceptical consequences. For instance, if someone could be functionally identical to me without having qualia, how do I know I really have them? (Of course some of these same arguments could serve to support a reduction of qualia to functional states, rather than their elimination. The line between eliminativism and reductionism, either of the functional or neurophysiological variety, is not always easy to draw.)

Advocates of colour qualia have responded to these arguments. Those who do not straightforwardly embrace dualism deny that their position entails it. Qualia, on their view, are not posits of an outmoded theory, but among the primary data that any adequate psychological theory must explain. The discussion above demonstrates how hard it is really to undermine the inverted qualia hypothesis. As for the charge of scepticism, many dismiss it as just that. Sceptical worries attach to most phenomena, they argue. But whether or not these replies succeed, eliminativist challenges of the sort just described will find adherents so long as a truly explanatory connection between qualitative character and neurophysiological or functional organization is still lacking.

See also: Colour, theories of; Consciousness; Qualia; Secondary qualities; Vision

JOSEPH LEVINE

References and further reading

Block, N. (1980) 'Troubles with Functionalism', in N. Block (ed.) *Readings in Philosophy of Psychology*, Cambridge, MA: Harvard University Press, vol. 1, 268-305.(Good presentation of the varieties of functionalism, and source for the 'China-head' example. The first part employs some technical notation.)

Clark, A. (1993) *Sensory Qualities*, Oxford: Oxford University Press.(Presents a theory of sensory qualities as points in a multidimensional quality space defined in terms of similarity judgments. Very detailed and somewhat technical.)

Dennett, D.C. (1991) *Consciousness Explained*, Boston, MA: Little, Brown. (Makes the case for eliminativism about qualia.)

Hardin, C.L. (1988) *Color for Philosophers: Unweaving the Rainbow*, Indianapolis, IN: Hackett Publishing Company.(Excellent review of the empirical literature regarding colour. Argues both for irrealism about objective colour and against the inverted spectrum argument in support of a functional analysis.)

Harrison, B. (1973) *Form and Content*, Oxford: Blackwell.(One of the earliest attempts to combat the inverted spectrum argument through a close consideration of the actual structure of colour experience.)

Hilbert, D.R. (1987) *Colour and Colour Perception: A Study in Anthropocentric Realism*, Stanford, CA: Center for the Study of Language and Information, Stanford University.(Argues for a realist account of objective colour. Extensive use of empirical findings.)

Jackson, F. (1982) 'Epiphenomenal Qualia', *Philosophical Quarterly* 32: 127-36.(Source for the example about Mary the vision scientist. Argues that sensory qualia are non-physical and do not causally interact with physical states.)

Kripke, S. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.(Classic presentation of the position that the reference of 'natural kind' terms, such as 'water' and 'gold', is determined by scientifically discovered essences.)

Levine, J. (1983) 'Materialism and Qualia: The Explanatory Gap', *Pacific Philosophical Quarterly* 64: 354-61. (Raises the problem of the 'explanatory gap').

Levine, J. (1993) 'On Leaving Out What It's Like', in M. Davies and G. Humphreys (eds) *Consciousness: Psychological and Philosophical Essays*, Oxford: Blackwell, 121-36.(Argues for the claim, presented in §3, that the explanatory gap underlies standard objections to identity theory, as well as the claim that the reduction of water to H_2O is different from the reduction of subjective red to a neural state.)

Lycan, W.G. (1987) *Consciousness*, Cambridge, MA: Bradford Books/MIT Press.(Defends a thoroughgoing functionalist theory of conscious states.)

Rey, G. (1992) 'Sensational Sentences Switched', *Philosophical Studies* 67: 77-103.(Combines a version of the representational theory of sensory qualities with eliminativist arguments, as described in §7, and applies it to the inverted spectrum problem.)

Shoemaker, S. (1981) 'The Inverted Spectrum', *Journal of Philosophy* 74: 357-81.(Ingenious defence of functionalism in the face of the inverted spectrum argument, which Shoemaker, unlike other functionalists, endorses.)

Smart, J.J.C. (1959) 'Sensations and Brain Processes', *Philosophical Review* 68: 141-56.(Classic presentation of the identity theory. The discussion of colour experience foreshadows functionalist theories.)

Van Gulick, R. (1993) 'Understanding the Phenomenal Mind: Are We All Just Armadillos?', in M. Davies and G. Humphreys (eds) *Consciousness: Psychological and Philosophical Essays*, Oxford: Blackwell, 137-54.(Good survey of the various anti-functionalist arguments, with lines of reply to each.)

Colour, theories of

The world as perceived by human beings is full of colour. The world as described by physical scientists is composed of colourless particles and fields. Philosophical theories of colour since the Scientific Revolution have been driven primarily by a desire to harmonize these two apparently conflicting pictures of the world. Any adequate theory of colour has to be consistent with the characteristics of colour as perceived without contradicting the deliverances of the physical sciences.

Given this conception of the aim of a theory of colour, there are three possibilities for resolving the apparent conflict between the scientific and perceptual facts. The first is to deny that physical objects have colours. Theories of this kind admit that objects appear coloured but maintain that these appearances are misleading. The conflict is resolved by removing colour from the external world. Second, it might be that colour is a relational property. For an object to possess a particular colour it must be related in the right way to a perceiver. One common version of this view analyses colour as a disposition to cause particular kinds of perceptual experience in a human being. Since the physical sciences deal only with the intrinsic properties of physical objects and their relations to other physical objects and not their relations to perceiving subjects, the possibility of conflict is removed. A third possible response is to maintain that colour really is a property of external objects and that the conflict is merely apparent. Some theories of this form maintain that colour is identical to a physical property of objects. Others maintain that colour is a property that physical objects possess over and above all their physical properties. Philosophical discussions of colour typically take the form of either elaborating on one of these three possibilities or attempting to show more generally that one of these three types of response is to be preferred to the others.

1 Eliminativist theories of colour

The association between atomistic metaphysics and the denial that colour has any place in the external world is both ancient and common. It can be found in authors as widely separated in time and other views as Democritus and Galileo. Although metaphysics is no longer atomistic in a strict sense, many contemporary scientists hold similar views about the absence of colour from the external world. Colours as perceived are commonly seen by those acquainted with scientific metaphysics to have features that preclude all attempts at the identification of colour with a physical property of external objects.

Once colour has been excluded from the external world there still remains the question of which objects, if any, possess colour. One possible answer to this question is to maintain that colours are genuine properties of some perceptual experiences. This view is a kind of phenomenalism with regard to colour (see Phenomenalism). Although colour phenomenalists maintain that no tomato is red, strictly speaking, it may be that the visual experience obtained by looking at a ripe tomato in good light genuinely does possess the property of being red (Jackson 1977). Something like this answer combined with a reductive account of mental properties may explain the repeated claims by visual scientists of the last two centuries that colours are properties of the brain. This form of subjectivism requires a commitment to the existence of *qualia* or sense-data: in other words, intrinsic properties of visual experience to which we have unmediated conscious access (see Qualia).

Alternatively, it may be argued that nothing, including visual experience, has colour. C.L. Hardin (1988) has argued, first, that there are essential properties of colour as perceived that no physical property possesses, then further that there are no mental items which could be the bearers of colour. The upshot is that although there are objects which *appear* coloured, there is nothing that *is* coloured.

The main difficulties faced by the various forms of colour eliminativism all derive from its failure to produce an adequate resolution of the apparent conflict between the scientific world and the perceptual world. Our visual experience is an experience of coloured objects. The external focus of our visual experience is also reflected in much of our thought and talk about colour. All forms of colour subjectivism are forced to maintain that our visual experience is profoundly misleading and most of our talk and thought about colour, if taken literally, profoundly confused. Colour subjectivism removes any possibility of conflict between perception and science by either moving the location or denying the reality of the world as we perceive it. Colours are properties of visual experience mistakenly projected in perception and thought onto external objects. The main motivation for attributing such widespread confusion to our ordinary thought and language is the failure of any less radical

account successfully to accommodate all the relevant facts.

2 Relational theories of colour

One motivation for adopting eliminativism about colour is the obvious dependence of perceived colour on the characteristics of our visual system. Organisms whose visual systems differ from ours will divide the spectrum in different ways. Objects that look the same in colour to us may look very different in colour to organisms whose sensory apparatus differs in certain ways from ours. In general, neither the categories nor the similarity relations that arise from human colour vision map in a straightforward way onto any physical property. Two coloured surfaces that are an exact perceptual match can be physically very different and, in general, similarity in perceived colour (of whatever degree) is no reliable guide to similarity in any interesting physical property. Eliminativism is not, however, the only possible theory of colour capable of accommodating these facts. Rather than eliminating colour from the world, relational theories of colour take colour to consist in a relation between the external perceived object and the perceiving subject.

The most common form of relational theory has taken colours to be dispositions or powers to cause characteristic kinds of perceptual experience. Dispositional theories of colour date back at least to the seventeenth century and continue to have widespread currency among philosophers. The most famous early exponent of this view of colour is John Locke (see Locke, J. §§3-4; Primary-secondary distinction §1). According to a theory of this kind, an object is red, for example, if and only if it has the power or disposition to cause a characteristic kind of experience in an appropriately situated human being.

Colours are to be thought of, according to relational theories, as analogous in important respects to physical properties like solubility and more precisely to qualities like nutritiousness or poisonousness. A substance is poisonous, for example, if it causes degradation in bodily function if taken into the body. Whether or not a substance is poisonous depends on what kind of organism is under consideration. Some substances cause bodily harm to some organisms while being harmless or even beneficial to other organisms. Thus the very same substance can be poisonous to one kind of organism and nutritious to another. Whether or not a substance brings about degradation in bodily function can also depend on the circumstances in which it is ingested. Substances that are harmless if taken by themselves can be poisonous if taken in conjunction with other substances. According to relational views of colour, an object can be red to one kind of perceiver and green to another in virtue of possessing exactly the same intrinsic physical characteristics. What matters is the kind of causal effect the object has on the perceiver, which depends in turn on the characteristics of the perceiver's visual system. The kind of effect an object has on a perceiver can also vary with changes in the viewing conditions, notoriously the character of the illumination, and consequently the colour of an object can change with viewing conditions. According to relational views of colour, no object has any particular colour if considered independently of how it is related to perceiving organisms.

One difficulty with relational views of colour is that they seem on the surface to be incompatible with the way we ordinarily talk about colours. We often make claims like 'that tomato is red' without offering any further explanation of who it is red for and in which circumstances. In addition, if someone were to claim of a beautiful ripe tomato that it is black, the ordinary response would be denial or puzzlement in spite of the fact that there are clearly viewing conditions under which the tomato would produce the characteristic effect that is associated with black things. The typical response of adherents of relational views of colour to this kind of difficulty is to claim that ordinary colour claims contain an implicit reference to a particular class of perceivers and a particular kind of viewing condition. To say that a tomato is red without qualification is to say that it produces the effect characteristic of red things on *normal* (human) perceivers in *normal* (for humans) viewing conditions. If we fix one term of the relation, then we can talk about colour as if it were an intrinsic property of the object being described.

The main problems that relational views of colour face fall into two categories. As we have seen, the theory attempts to account for ordinary colour talk by making use of the notions of a normal perceiver and normal conditions. There are serious questions, however, about whether a suitable conception of normality actually exists. It is possible to specify precisely the characteristics of a normal perceiver, but then it follows from the substantial variation in human colour vision that there will be extremely few normal perceivers. Thus most people will not be in a position to determine visually the colours of objects. If we offer a more relaxed standard for who counts as a normal perceiver, then there will be variation in the effects of objects within the class of normal perceivers, and the

determinacy in colour attribution that the appeal to normal perceivers was supposed to provide will be lost. Similar problems arise in trying to specify which viewing conditions count as normal. The other main difficulty facing relational theories of colour arises from the necessity of describing the characteristics, the production of which gives an object its colour. In ordinary talk we use words like ‘red’ indiscriminately to describe both the external property and the perceptual experience with which it is associated. Relational theories of colour must find some way of independently characterizing perceptual experiences while avoiding circularity and vacuity.

3 Realist theories of colour

Realist theories of colour hold that colour is a genuine property of physical objects and that objects possess this property independently of their relation to perceiving subjects (see Realism and antirealism §2). Realist theories can be divided into two categories depending on whether or not they take colour to be reducible to physical properties. Physicalist theories of colour hold that colour is a physical property, possibly a complicated one, and that to see the colour of an object is to see that it has a physical property. Physicalist theories of colour bear some analogy to materialist theories of mind and share some of the same advantages and problems (see Materialism in the philosophy of mind). In spite of the resemblance in logical structure between the two types of theory, there is no logical connection between materialist theories of mind and physicalist theories of colour, and it is possible consistently to maintain every combination of affirming and denying the two theories.

The early defenders of physicalist theories of colour were somewhat vague about exactly which physical property they took colour to be identical with and largely confined their attention to defending the possibility that there could be some physical property identical with colour. It is obvious, however, that the relevant physical property must be one that has to do with light, or the reflection of light, since these are the prominent elements in the causal chain leading to the perception of colour. The most suitable candidate, as has been argued by D.R. Hilbert (1987), is the way in which objects reflect light. This property, surface spectral reflectance, is a relatively stable property of the surfaces of objects, which does not vary with changes in the illumination. Unlike its reflectance, the light reaching the eye from an object varies with changes in the illumination in ways that match neither our perceptual judgments of object colour nor our inclination to attribute stable colour properties to objects.

The main problems facing physicalist theories revolve around two connected questions: how is the content of colour experience determined and what is the structure of our colour concepts? Armstrong (1968) clearly saw that any very rich concept of colour is incompatible with physicalism. Most people who use and apply colour concepts are ignorant of the relevant physical and physiological facts. If we build a large amount of colour lore and phenomenology into the concept of colour, then there will be no physical property which will fit the bill. Many of the objections to physicalist theories of colour consist in arguing that some part of our concept of colour is incompatible with the empirical facts. The physicalist’s only response to these arguments is to show that either the empirical facts are misrepresented or the claimed necessary truth about colour is not really necessary at all. Physicalist theories of colour require that colour perception not be transparent, in the sense that any colour perceiver necessarily knows all there is to know about colour. The easiest way to avoid this kind of difficulty is to adopt an externalist theory of content. Externalist theories of content have the effect of breaking any necessary connection between internal perceptual states and what they represent and thereby provide the resources for a defence of physicalism against a class of objections drawn from the assumption of transparency (see Content: wide versus narrow).

Not all colour realists are physicalists, and there have been some philosophers who have claimed that colours are intrinsic properties of physical objects that cannot be identified with any physical property. There are two possibilities available to the defender of such a theory. Either colours are properties over and above all the physical properties and bearing no necessary connection to them, or colours are properties that supervene on physical properties but cannot be reduced to them. These two views bear obvious analogies to dualism and non-reductive materialism in the philosophy of mind. The main difficulty the dualist theory faces is that colours, as perceived, will have no place in the causal chain leading to the perception of colour, which raises serious epistemological problems as to how we could ever know the colour of an object. Taking colour to supervene on the physical properties of objects avoids this difficulty and has the virtue of avoiding many of the other difficulties faced by the theories of colour discussed above. The main difficulties facing the theory of colours as supervenient properties are the general difficulties of understanding the relation of supervenience and what kind of evidence could ever

support a conclusion of this kind.

See also: Colour; Secondary qualities

DAVID R. HILBERT

References and further reading

Armstrong, D.M. (1968) *A Materialist Theory of the Mind*, New York: Humanities Press.(Referred to in §3. Chapter 12 defends a physicalist theory of colour in the context of a materialist theory of mind.)

Byrne, A. and Hilbert, D.R. (1997) 'Readings on Color', in *The Philosophy of Color*, vol. 1, Cambridge, MA: MIT Press.(A collection of recent philosophical work on colour. The referenced works by Campbell, Johnston and Peacocke are reprinted here.)

Campbell, J. (1993) 'A Simple View of Colour', in J. Haldane and C. Wright (eds) *Reality, Representation, and Projection*, Oxford: Oxford University Press.(Defends a version of non-physicalist realism about colour.)

Galileo, G. (1623) *The Assayer*; trans. and ed. S. Drake in *Discoveries and Opinions of Galileo*, Garden City, NY: Doubleday, 1957.(Referred to in §1. A presentation of Galileo's metaphysical views.)

Hardin, C.L. (1988) *Color for Philosophers*, Indianapolis, IN: Hackett; repr. expanded edn, 1993.(Referred to in §1. Argues for an eliminativist theory of colour. Also presents a useful introduction to colour science.)

Hilbert, D.R. (1987) *Color and Color Perception*, Stanford, CA: Centre for the Study of Language and Information.(Referred to in §3. Defends a physicalist theory of colour.)

Jackson, F. (1977) *Perception*, Cambridge: Cambridge University Press.(Referred to in §1. Defends a phenomenalist theory of colour in the context of a general theory of perception.)

Johnston, M. (1992) 'How to Speak of the Colors', *Philosophical Studies* 68: 221-63. (Defends a version of dispositionalism and contains a discussion of the features common-sense attributes to colour.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge, Cambridge University Press.(A standard collection of writings by Presocratic philosophers, Democritus included.)

Locke, J. (1689) *Essay Concerning Human Understanding*; Oxford: Oxford University Press, 1975.(Referred to in §2. The origin of modern philosophical debates about the nature of colour.)

Peacocke, C. (1984) 'Colour Concepts and Colour Experience', *Synthèse* 58: 365-81. (Defends a version of dispositionalism and worries about the problem of circularity that faces dispositional theories.)

Combinatory logic

Combinatory logic comprises a battery of formalisms for expressing and studying properties of operations constitutive to contemporary logic and its applications. The sole syntactic category in combinatory logic is that of the applicative term. Closed terms are called 'combinators'; there is no binding of variables. Systems containing the basic combinators S and K exhibit the crucial property of combinatorial completeness: every routine expressible in the system can be captured by a term composed of these two combinators alone. Combinatory logic is a close relative of Church's lambda calculus. M. Schönfinkel first introduced and defined basic combinators in 1920 in assaying foundations for mathematics that avoid bound variables and take operations, rather than sets, as fundamental. H. Curry later rediscovered the combinators (and coined the term 'combinatory logic') independently of Schönfinkel. Curry constructed various formal systems for combinatory logic and, throughout most of the subject's history, was the central figure in the research. In 1969, D. Scott succeeded in constructing set-theoretic, functional models for the lambda calculus and combinatory logic. Since then semantic studies of combinatory systems, together with research on their applications to computer science and further development as foundational systems, have dominated the field.

1 Combinators and the elimination of variables

In his *On the Building Blocks of Mathematical Logic* (1924), M. Schönfinkel sought to surpass Whitehead and Russell's *Principia Mathematica* and Frege's *Grundgesetze* in isolating notions absolutely fundamental to mathematical logic. In addition to employing a Sheffer stroke-like operator for predicate calculus (Sheffer (1913) showed that all truth-functional connectives could be expressed in terms of 'not…and'), Schönfinkel proposed abandoning the type-stratified logical universe of *Principia Mathematica* in favour of one consisting entirely of type-free functions, mathematical operations presupposed by all others, unlimited in their allowed inputs and, hence, self-applicable. Moreover, he looked to reduce the number of all logical symbols (of both first- and higher-order logic) to three: the stroke-like function, a constant function and a 'fusion' function. These were the building blocks from which Schönfinkel intended to analyse logical expressions as compounds of basic operations. In this way, the ultimate operations on which mathematical logic depends - operations usually left implicit and unformalized - were to be revealed to the foundational gaze. As a bonus, this universal analysis was to require a wholesale elimination from logic of the variable.

For a simple example of variable elimination, consider a binary operation $\circ$ which is 'idempotent'; that is, for all x, $x \circ x = x$. To render this relation without bound variable x, one introduces operation symbols or, to use terminology due to H. Curry, 'combinators'. In this case, let O and I be such that $Ox = x \circ x$ and $Ix = x$. Then, to claim idempotency for $\circ$, one writes merely $O = I$. The variable x becomes redundant, for, if we apply both sides of the combinator equation to x, we return the original: $x \circ x = Ox = Ix = x$. Combinators then serve to express general laws, such as those of function application, without bound variables, and knotty issues of substitutability and illicit variable capture are avoided.

In general, a combinator is a variable-free name for an operation (on operations) defined using an equation. A number of combinators are standard and bear conventional names. Some of the most common are S, K, I, B and W, defined as follows: $Sxyz = xz(yz)$, $Kxy = x$, $Ix = x$, $Bxyz = x(yz)$ and $Wxy = xyy$. Schönfinkel, in effect, introduced combinators S and K for the operations he called 'fusion' (*Verschmelzung*) and 'constant' (*Konstanz*). S performs a generalized composition operation and K, when applied to x, yields Kx, the constantly x-valued function of y. I is a universal identity function. B denotes conventional composition of functions, since Bfg, applied to x, gives $f(gx)$. Finally, W realizes diagonalization: Wf, applied to x, yields fxx. If we think of f as a binary function, Wf is the unary function which is f with arguments identified. (Incidentally, here we are exploiting a device known to Schönfinkel, as well as to Frege, but now called 'currying', after Curry. By that device, combinators denote unary operations but achieve the effect of *n*-ary operations ($n > 1$) when iterated.)

2 Combinatory logic

Curry (1900-82), a student of Hilbert who rediscovered the combinators independently of Schönfinkel, devised the first formal systems for (what he termed) 'combinatory logic'. In Curry's systems, the fundamental syntactic

category is that of (combinatory) term, where basic terms are either variables or constants. Compound terms denote applications: if t and u are terms, then so is tu, expressing the application of t to u. Among the constants are the basic combinators K and S. Any term built up entirely from K and S is also called a combinator.

As with the lambda calculus, we can define appropriate notions of reduction, normal form and equality for combinatory terms (see Lambda calculus). The 'reduction' of one combinatory term to another is conceived by analogy with simplifying computation, for example, the reduction of $(3 + 6) \times (10 - 2)$ to 72 by calculation. A 'normal form' is a term which can be simplified no further. Terms t and u are equal whenever they can be linked by a finite sequence of terms $t_0, t_1, \ldots, t_n$ such that, for every k and l, either t_k reduces to t_l or conversely.

Since operations may be viewed both extensionally and non-extensionally, there are extensional ('strong') and non-extensional ('weak') conceptions of reduction and normality, on either of which Kxy reduces to x and $Sxyz$ to $xz(yz)$. Both kinds of reduction are univocal in that, if t reduces to a normal form, it does so uniquely. This is guaranteed by the 'Church-Rosser theorem', which states that, if t reduces to u and to v, then there is a term w to which u and v both reduce. As in the lambda calculus, select combinators can serve as numerals and an arithmetic can be so developed that every partial recursive function is representable within combinatory logic.

It is readily seen that some combinators are definable from others. For example, I is recovered as SKK, since, for any x, $SKKx = Kx(Kx) = x = Ix$. The principle here is general: S and K comprise a 'complete' set of combinators in that every other combinator is derivable from these alone. Indeed, let t be any term, simple or complex, containing only variables from a finite set, say, x, y, z. Then there is a combinator, denoted $[x, y, z]t$, called 'the abstraction of t with respect to x, y and z', containing no free variables and constructed entirely from S and K such that $(([x, y, z]tx)y)z$ reduces (weakly) to t. In effect, Schönfinkel was the first to recognize the completeness of combinatory logic.

The notation '$[x, y, z]t$' is entirely metatheoretic; it denotes a term containing neither free nor bound variables and connotes that the abstraction $[x, y, z]$-operation plays a role in combinatory logic analogous to that played in the lambda calculus by the $\lambda x\, \lambda y\, \lambda z$ abstractor (see Lambda calculus). This loose analogy is made precise by defining functions which inject the set of combinatory terms into those of the lambda calculus so that relevant equality relations coincide. There are metatheoretic functions Λ and C, where Λ maps combinatory terms into lambda terms and C the reverse, so that, for combinatory terms t and u, $C\Lambda(t)$ is the same term as t and, as far as extensional equality is concerned, $t = u$ in combinatory logic if and only if, in the lambda calculus, $\Lambda(t)$ equals $\Lambda(u)$. (The non-extensional versions of the respective calculi are also comparable but the matter is more complex.) The function C employs abstraction in dealing with lambda terms: $C(\lambda x\, \lambda y.t)$, for example, is $[x, y]C(t)$.

Intuitively, a model for (non-extensional) combinatory logic would be a universe D with a binary operation • - representing application - and containing elements k and s which behave, with respect to •, as K and S do. It is easy to verify that such a model satisfies all the equalities of the weak or non-extensional combinatory logic. However, combinatory logic was originally conceived as a doctrine of operations or functions and real difficulties arise in rendering these functions in set-theoretic terms. For one thing, combinatory logic allows self-application: the combinator SS is perfectly well-formed. But no standard set-theoretic function can be a member of its own domain. Moreover, every combinator, unlike every set-theoretic function, admits fixed points. Curry discovered a 'paradoxical' combinator Y, which may be defined $WS(BWB)$, that computes a fixed point for any term: if t is a term, Yt is equal to $t(Yt)$.

In 1969, D. Scott (1976) was the first to overcome these obstacles and give set-theoretic constructions of models (in the above sense) for combinatory logic in which the terms name functions. A key idea was to let combinatory terms name functions which are not arbitrary but continuous. Here, a continuous function on a partial order with limits preserves the order and the limits which exist. Scott's original construction was quite complex; models less complicated to build were soon forthcoming. These semantic breakthroughs inaugurated a period of intense research on combinatory logic and the lambda calculus, stimulating numerous applications of both to computer science.

3 Combinatory logic as a foundation for mathematics

Schönfinkel's original goal, which Curry nurtured and pursued under the rubric 'illative combinatory logic', was

the construction of a comprehensive system for the foundation of mathematics free of types and bound variables, and highlighting those operations essential to logical thought. To be comprehensive, no significant notion was to be banished to the metatheory. Other than the assertion sign, every concept which bears on the system was to be amenable to combination with combinators inside the system.

The strongest of Curry's early illative systems, which appeared in print in 1934, proved excessively comprehensive: Church's students S. Kleene and J. Rosser demonstrated its inconsistency in 1935. Curry himself refined the core of their argument, obtaining the elegant 'Curry's paradox', which shows plainly that even a simple logic of implication cannot be developed in the presence of combinatorial completeness. Curry and his co-workers then sought to avoid paradox by inventing systems to circumscribe, within the class of legal terms, canonical terms standing for propositions. Unfortunately, consistent versions of these systems were too weak to support a comprehensive mathematics. Throughout the middle and late twentieth century, a number of logicians contributed significantly to achieving the goal of Curry and Schönfinkel (a task known as 'Curry's programme' - see Seldin 1980) for illative systems, among them P. Aczel, M. Bunder, S. Feferman, F. Fitch, J. Hindley, D. Scott and J. Seldin. In 1983, M. Bunder published his system HOPC ('Higher-Order Predicate Calculus'), which is an illative system with internal typing and in which a significant portion of Zermelo-Fraenkel set theory can be interpreted.

See also: Logical and mathematical terms, glossary of

DAVID CHARLES McCARTY

References and further reading

Barendregt, H. (1981) *The Lambda Calculus: Its Syntax and Semantics*, Amsterdam: North Holland, 2nd edn, 1984.(A bible of lambda calculus research. Authoritative and detailed.)

Curry, H. (1934) 'Functionality in Combinatory Logic', *Proceedings of the National Academy of Sciences of the USA* 20: 584-90.(An article crucial to later research in illative combinatory logic.)

Curry, H. and Feys, R. (1958) *Combinatory Logic*, vol. 1, Amsterdam: North Holland.(Once the premier work on the subject; it remains an excellent reference.)

Hindley, J. and Seldin, J. (1986) *Introduction to Combinators and λ-Calculus*, London Mathematical Society Student Texts no. 1, Cambridge: Cambridge University Press.(An excellent introduction, highly recommended for beginners.)

Kleene, S. and Rosser, J. (1935) 'The Inconsistency of Certain Formal Logics', *Annals of Mathematics* 36: 630-6. (A technical application of gödelization to proving the inconsistency of early systems of Church and Curry.)

Schönfinkel, M. (1924) 'Über die Bausteine der mathematischen Logik', *Mathematische Annalen* 92: 305-16; trans. *On the Building Blocks of Mathematical Logic*, ed. H. Behmann; repr. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 355-66.(From a 1920 presentation to the Mathematical Society at Göttingen. Introduction in the 1967 version by W.V. Quine.)

Scott, D. (1976) 'Data Types as Lattices', *SIAM Journal of Computing* 5: 522-87.(A characteristically charming exposition of results by Scott, Plotkin and others on the graph model.)

Seldin, J. (1980) 'Curry's Program', in J. Hindley and J. Seldin (eds) *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, New York: Academic Press, 3-34. (An authoritative presentation of Curry's vision for combinatory logic.)

Sheffer, H.M. (1913) 'A Set of Five Independent Postulates for Boolean Algebras, With Application To Logical Constants', *Transactions of the American Mathematical Society* 14: 481-8.(The first publication of the fact that there is a single connective that is expressively complete for sentential logic.)

Stenlund, S. (1972) *Combinators, λ-Terms and Proof Theory*, Dordrecht: Reidel.(A laudably concise introduction to combinatory logic and its role in proof theory.)

Comedy

In the narrowest sense, comedy is drama that makes us laugh and has a happy ending. In a wider sense it is also humorous narrative literature with a happy ending. In the widest sense, comedy includes any literary or graphic work, performance or other art intended to amuse us. This entry will leave aside theories of humour and concentrate on comedy as a dramatic and literary form.

Comedy began at about the same time as tragedy, and because they represent alternative attitudes toward basic issues in life, it is useful to consider them together. Unfortunately, several traditional prejudices discriminate against comedy and in favour of tragedy. There are four standard charges against comedy: it emphasizes the animal aspects of human life, encourages disrespect for leaders and institutions, is based on malice, and endangers our morality. These charges are easily answered, for none picks out something that is both essential to comedy and inherently vicious. In fact, once we get past traditional prejudices, several of the differences between comedy and tragedy can be seen as advantages. While tragedy tends to be idealistic and elitist, for example, comedy tends to be pragmatic and egalitarian. While tragedy values honour, even above life itself, comedy puts little stock in honour and instead emphasizes survival. Tragic heroes preserve their dignity but die in the process; comic characters lose their dignity but live to tell the tale. Most generally, comedy celebrates mental flexibility and a realistic acceptance of the limitations of human life. The comic vision of life, in short, embodies a good deal of wisdom.

1 The demeaning of comedy

Although comedy and tragedy grew up together, and many dramatists from Sophocles to Shakespeare wrote both, tragedy is usually thought superior to comedy, and is often judged the only important dramatic form. Tragedy is called ‘serious’ drama, comedy ‘light’ drama. The low status traditionally held by comedy is revealed by two meanings that arose for the word ‘comical’: ‘befitting comedy; trivial, mean, low; the opposite of tragical, elevated, dignified’ and ‘of persons: low, mean, base, ignoble or clownish’.

The demeaning of comedy, and of humour generally, began with Plato. Four main charges are traditionally offered. One is that comedy, which had its origins in animal masquerades, phallic processions and similar revelry, emphasizes the animal side of human nature. Plato found the Old Comedy of his time still wild and vulgar. In his mind the licence of comedy encouraged the undermining of our rationality by our lower physical nature. When laying down rules for the education of the young guardians in his ideal state, Plato insisted that they must not be prone to laughter and that the literature they read should not show the heroes and gods laughing too heartily.

Comedy is also charged with encouraging irreverence toward leaders and institutions. A society, like an individual, needs rational control, and that requires respect for leaders and traditions. But comedy can make fun of anything; Greek comedy even lampooned the gods. Plato was probably especially resentful of the ridicule his teacher Socrates suffered in the comedy of Aristophanes.

Throughout history, opposition to comedy and laughter has been strongest in societies which emphasize physical restraint, decorum and conformity. Many medieval monastic orders had statutes forbidding laughter. The Puritan and Victorian eras saw many condemnations of comedy and laughter. The more authoritarian the regime, the greater its suppression of comedy. Hitler even set up ‘joke courts’ to punish those who made fun of his regime - one Berlin cabaret comic was executed for naming his horse Adolf.

The third charge against comedy, and humour generally, is that laughter is inherently mean-spirited. According to Plato, the object of laughter is vice, and specifically people’s ignorance about themselves. Dramatic characters and real people are comic to the extent that they think of themselves as wealthier, better-looking, more virtuous or wiser than they really are. Our laughter at their self-ignorance involves a kind of malice towards them - a ‘pain in the soul’, as Plato called it - that is not only antisocial but harmful to our own character.

Aristotle agreed with Plato that the essence of laughter is ridicule. Most people carry humour too far, he claimed, not worrying about hurting the feelings of those at whom they laugh. This view of laughter was later called the superiority theory. Its most famous proponent was the seventeenth-century philosopher Thomas Hobbes, who said that the cause of laughter is the sudden glory we feel when we judge ourselves to be doing better than someone

else. Those who laugh the most, according to Hobbes, are those who are conscious of the fewest abilities in themselves. They have to search out the imperfections of others in order to feel good about themselves.

The last charge against comedy is that it is full of gluttons, drunkards, liars, adulterers and other base characters, who are bound to have a bad influence on our own morality. Aristotle said that comic characters are worse than real people and warned that children should not be allowed to attend comedies because they would be led to imitate the vices they saw on the stage. The purported danger of comedy to morality has been cited many times. It was part of the English Puritans' rationale for outlawing drama. Rousseau used it against the comedies ofMolière. The weight attached to it can be judged from the number of writers and critics who felt obliged to argue that in laughing at immoral behaviour, we reject it, so that comedy discourages rather than encourages vice.Ben Jonson, Sir Philip Sidney, John Dryden, Henry Fielding, George Meredith, Henri Bergson and dozens of others defended comedy by citing its moral utility in this way.

2 Answering the charges

While the four charges against comedy apply to some plays and comic works, none applies to all of them, and none makes a convincing case that some characteristic is both essentially vicious and essential to comedy.

Comedy's emphasis on the animal side of human nature may be a fault if we share Plato's low opinion of the body and of the physical side of life. Similarly, the irreverence of comedy may be objectionable if we agree that our leaders and institutions deserve reverence and not critical questioning. But we need not share Plato's views on these issues. Indeed, a good case can be made for saying that comedy is valuable precisely because it reminds us of our physicality and because it keeps us thinking critically about our leaders and institutions.

The other two charges against comedy - that laughter is an expression of feelings of superiority and malice, and that the base behaviour of comic characters might rub off on us - do focus on two things that are reasonably considered objectionable. But they fail to show a necessary connection between either of these and comedy.

Although the first of these charges has a long history, it has seldom been carefully examined. If the superiority theory is right, then our laughter is always directed at a person, and in laughing we must be comparing ourselves favourably with that person. But, as Francis Hutcheson showed a century after Hobbes, neither of these consequences is true. We sometimes laugh when no one else is involved, and we sometimes laugh when someone seems superior to us. If I open my front door on a November morning to find a foot of snow where there was grass the night before, I may laugh - not at anyone, not even at the snow (in the sense of ridicule), but simply out of surprise. Similarly, I may laugh at clever rhymes or other wordplay in a comedy without comparing myself to the character speaking the lines or to anyone else. Some action which is better than we expected may also make us laugh in surprise. A stock character in early film comedy, for example, is the plucky hero, such as Charlie Chaplin or Buster Keaton, who gets out of trouble with an ingenious acrobatic stunt that we would never have thought of, much less been able to execute. The stunt makes us laugh, though the character looks superior to us.

The last charge against comedy - that we are likely to imitate its base characters - has probably been made more often than any other. But seldom has any evidence been given for it. Do people who see a lot of comedies have a higher rate of drunkenness or adultery? Are people who laugh at hypocrisy more likely to become hypocrites themselves? These are empirical questions calling for empirical research. Several characters, such as the boor, the windbag and the pompous ass, seem to be comic only if we think that their traits are undesirable - someone who emulated them would not find them funny. It may be that other comic characters, such as the smooth-talking liar, do elicit emulation from some people. But that has never been established, nor has it been shown that these characters outnumber the ones who discourage emulation. In short, like the other charges against it, the claim that comedy threatens our morality is largely an ancient prejudice.

3 Comedy and tragedy

Once we set aside traditional prejudices against comedy, we can compare it more equitably with tragedy. The most general similarity between the two is their focus on the incongruities in human life - the ways in which our experiences do not match our expectations. As William Hazlitt said, humans are the only animals who laugh and weep because they are the only ones who are struck by the difference between what things are and what they ought to be.

It is in their responses to life's incongruities that comedy and tragedy differ. Both see misfortune, vice, folly, and, in general, the gap between the real and the ideal as part of the human condition. But tragedy sees these leading to downfall and death, while comedy sees them as something we can live with and even enjoy.

Comedy and tragedy also have different attitudes towards the physical side of human nature. Comedy accepts the limitations of our bodily existence and celebrates acts like eating and sex. Tragedy bemoans our physical limitations and often identifies the human being with the mind, spirit or soul. In general, comedy is more physical and active, and tragedy more intellectual and contemplative. Falstaff might deliver a monologue while gnawing on a leg of mutton: it is inconceivable that Hamlet would do so.

The idealism and dualism of tragedy carry over to its vision of society. In tragedy only a few people are important and only their lives are of interest. The main characters in tragedy, as in the epic, are heroes, typically male rulers or warriors. In comedy, by contrast, there is a greater variety of characters, women are more prominent, and central characters may come from any social class. While the language of tragedy is elevated, the language of comedy is common speech.

Tragedy usually focuses on the suffering of one elite character in an extraordinary situation; comedy involves several characters from different social classes in ordinary situations. When comedy has a central character, that person, unlike the tragic hero, is not exalted above other human beings. Tragedy emphasizes the dignity and pride of the hero, which are often based on the code of honour of a male-dominated, power-based, militarist ideology. Indeed, it is often just this ideology which gets the hero into the tragic situation. Comic characters, not bound by codes of honour, may lack dignity, but at the end of the comedy they are still alive. Indeed, they are often found attending a wedding or another life-affirming celebration.

Furthermore, because comedy values life - especially the life of the community - over honour, it emphasizes the social support we all need. In tragedy, by contrast, the hero is more of a 'loner'. Many comic plots are based on reconciliation and peacemaking, while no tragic plots are. As Aristotle noted, in comedy enemies sometimes become friends, but in tragedy they never do.

4 Comic wisdom

The popularity and value of comedy lie largely in its vision of human life, which contrasts sharply with the dominant ideologies of Western culture. Those ideologies treat as virtues such traits as respect for authority, duty, honour, single-mindedness, courage and a capacity for hard work. These have been promoted by armies and other patriarchal institutions since ancient times. An important way of inculcating them in society at large has been to celebrate them in epic and tragic art, which are full of military imagery. Indeed, patriarchies try to get us to think of everything in military terms. In the USA, social programmes are called 'the war on poverty'; medical research is called 'the war on breast cancer'; even programmes to stop violence are 'the war on violence'! When military metaphors sink deep enough into our culture, life itself becomes a series of battles.

While blind obedience, single-mindedness, the ability to work constantly, and the willingness to die or kill on command are important for the conduct of war, it is not at all clear that they are virtues in all areas of life. Thus alongside the official ideology promulgated by epics and tragedy there has always existed an alternative ideology of comedy. Instead of promoting military virtues, comedy promotes the questioning of authority, mental flexibility, playfulness and the value of life. All of these threaten institutions of power in various ways, and as a result comedy has been suppressed in most cultures. However, because it addresses deep human needs, it has survived.

Comedies have different kinds of characters. Many serve as negative role models, examples of how not to act. In laughing at the miser, the prude and the pedant, as Henri Bergson pointed out, we are recognizing their mechanicalness, their ineptness at living a human life. But most comedies also have at least one character that we identify with and may even admire. Many of the roles played by Charlie Chaplin, Mae West and Groucho Marx are of this type. These characters are so different from epic and tragic heroes that their usual name, 'comic heroes', is misleading. We can call them comic protagonists.

The attitudes of these characters embody what is most valuable in comedy. Unlike tragic heroes, they play as well as work. They are not unwaveringly committed to any cause; nor are they prepared to die, or kill, to achieve their

goals. Like tragic heroes, they face problems and enemies, but instead of confronting them head on with violence, they use trickery, perhaps by turning the power of the threat against itself, or with reverse psychology. When all else fails, they are not ashamed to run away. As the old saying goes, you're a coward for only a moment, but you're dead for the rest of your life.

Comic protagonists differ most notably from tragic heroes in their mental flexibility, a trait which comedy celebrates. The characters who lose in comedy are rigid creatures of habit; those who succeed are adaptable and think on their feet. Unlike tragic heroes, comic protagonists do not have fixed categories for thinking or acting. They can view situations from several perspectives and see many possibilities. Much of their thinking is lateral rather than vertical, to use Edward de Bono's terms.

When confronted by problems, tragic heroes are given to emotions that make them mentally rigid and even obsessive. Comic protagonists keep an unemotional clearheadedness in the face of misfortune that allows them to think rather than feel their way through challenges. They do not engage in self-pity or curse their fate, but are more likely to laugh at their problems, as tragic heroes never do. As a result, they are more likely to bounce back from their mistakes and learn from them. The contrast here is fittingly generalized by Walpole's maxim, 'This world is a comedy to those that think, a tragedy to those that feel'.

Since emotional disengagement and the ability to imagine alternatives are a big part of human freedom, comic protagonists are considerably freer than tragic heroes. They are often in charge of their lives as tragic heroes are not, and they end up victors, while tragic heroes end up victims.

It is often said that the tragic vision of life embodies wisdom. Solemnity and pessimism are considered hallmarks of wisdom. Western thinkers, with a few exceptions such as Democritus and Nietzsche, have usually thought wisdom to be a kind of seriousness about life. But that is all part of the traditional prejudice against comedy. Judged fairly, the comic vision reveals at least as much wisdom as the tragic.

Indeed, if wisdom includes emotional disengagement, seeing life from a higher perspective than usual and seeing it objectively rather than from a self-privileging position, then comedy seems wiser than tragedy. If wisdom includes a realistic attitude towards life, comedy's tolerance for human limitations and its emphasis on adapting ourselves to an imperfect world seem to make it more realistic than tragedy. More fully than tragedy, too, comedy represents the richness of life - especially social life - in the many ways it may be lived and appreciated.

Comic characters make mistakes and suffer misfortunes, but through it all they are at home in their world, and they get by with a little help from their friends. Tragic characters, with their elitism and idealism, are not satisfied with living a merely human life. The central lesson of comedy is that we are finite and prone to error, but with a sense of humour we can still be happy.

The capacity for happiness seems to need some psychological technique for coping with finitude and fallibility, and humour is easily the most effective. Psychological studies have shown that humour is correlated not only with self-esteem but with creativity and a tolerance for ambiguity, diversity and change. Furthermore, humour has medical benefits - it blocks negative emotions, counteracts stress, boosts the activity of the immune system, reduces pain, and even has a laxative effect!

Both comedy and tragedy are reactions to the human condition, but as a dramatic form, an artistic sensibility, and an attitude toward life itself, comedy seems truer to human nature. The displacement of tragedy by comedy and tragicomedy in the twentieth century seems a step towards the acknowledgement of this fact.

See also: Bergson, H.-L. §6; Humour; Tragedy

JOHN MORREALL

References and further reading

Bergson, H. (1956) 'Laughter', in Wylie Sypher (ed.) *Comedy*, Garden City, NY: Doubleday.(An explanation of laughter as the corrective response of a group to an individual's 'mechanical inelasticity'. Difficult style.)

Gutwirth, M. (1993) *Laughing Matter: An Essay on the Comic*, Ithaca, NY, and London: Cornell University Press.(A synthesis of previous theories of laughter and an exploration of the values of comedy. Moderately difficult.)

Morreall, J. (ed.) (1987) *The Philosophy of Laughter and Humor*, Albany, NY: State University of New York Press.(An anthology of philosophical writings on laughter and humour from Plato to Roger Scruton, including all philosophers cited in this article.)

(1989) 'The Rejection of Humour in Western Thought', *Philosophy East and West*, 39 (3): 244-65.(Traces the development of philosophical theories of humour, and philosophical opposition to humour; suggests overlooked philosophical values of humour, discussing humour in Zen. Easy to read.)

Plato (*c*.380s-367 BC) *Republic*, trans. P. Shorey, Cambridge, MA: Loeb Classical Library, Harvard University Press, 1930, esp. 388e-389a.(Parallel Greek text and English translation.)

Comenius, John Amos (1592-1670)

Comenius (Jan Amos Komensky), a Czech philosopher and theologian, was one of the founders of modern educational theory. As a Protestant minister he had to leave Bohemia during the Counter-Reformation, spending most of his life in various European countries. His greatest work (not published during his lifetime) is De rerum humanarum emendatione consultatio catholica (A General Consultation on the Reform of Human Affairs), whose leading idea is the demand for a harmonious arrangement of human relations on the basis of rational enlightenment, the development of education, and the instruction of all humankind. Comenius builds his philosophy on an idea of human nature understood as grounded in an active creative force perpetually leading to improvement: instruction and education are the tools to fulfil this humanitarian ideal. To lend force to this, Comenius constructs a whole ontological system, in which a harmonious development of the whole of existence leads to human reality as its highest tier.

Comenius was born in southern Moravia; his family belonged to the Czech reformist church known as the Unity of Brethren. After studies at the Union's schools in his own country and at the Universities of Herborn and Heidelberg, Comenius became a preacher of the reformed church, and later its bishop. His activities in his homeland - where he had started working on a great encyclopedia of universal knowledge in the Czech language and on his first educational works - were interrupted by the anti-Habsburg uprising of the Bohemian Estates (1618-20) which ended in the loss of the kingdom's independence. Enforced re-Catholicization and persecution of other religions followed. For a few years Comenius went into hiding while writing his best-known Czech book, the symbolic prose *Labyrint světa a ráj srdce (The Labyrinth of the World and the Paradise of the Heart)*, but in 1628 he left his country for good and settled in Leszno (Lissa) in Poland. Here he created his important educational works, *Didactica* and *Janua linguarum reserata*, where he demands equal rights to education for all people, and also his first pansophic works, *Pansophiae prodromus* and *Conatuum pansophicarum dilucidatio*, where he makes proposals for reforming political relations in the world ravaged by the Thirty Years' War.

Comenius' educational and pansophic work found much support in Europe, and the English Parliament invited him to come to London to reform the education system and pursue his general efforts at universal improvement there. He recruited influential patrons such as James Ussher, John Selden, John Pym, Samuel Hartlib and others. Impulses from Comenius' stay in England contributed to the appearance of his *Via lucis* which contains - in addition to a proposal for an international organization of scientific work - Comenius' historical philosophy: Darkness (feudalism, the rule of the papacy and the Habsburgs, and general backwardness) will be driven away by Light because history is a tiered process where the amount of Light grows continuously out of the struggle with Darkness, in the form of rays of general education and rational enlightenment. Comenius expected that the period of the voyages of discovery, expansion of trade, book printing, science and technology will soon usher in an era of enlightenment, peace, international co-operation and prosperity. This should be helped by the spread of education into all layers of society, and to each individual.

The outbreak of the Civil War curtailed Comenius' activities in England. He left for Sweden to reform its educational system, hoping at the same time that Sweden's policy and participation in the Thirty Years' War might result in the liberation of the Kingdom of Bohemia from the rule of the Habsburgs and facilitate the repatriation of its political emigrants, a hope soon to be proved forlorn. During his stay he continued to expand his pansophic plans but found the Swedes uninterested. The years 1651-4 were spent in Hungary reforming its educational system, and Comenius then returned to Leszno where a fire in the town destroyed his library and all his manuscripts. Due to the change of political climate in Poland he could not stay there much longer, and in 1656 he left for Amsterdam where he spent the remainder of his life.

In the Netherlands Comenius published (with the help of the City of Amsterdam) his educational work *Opera didactica omnia* (1657-8). At the same time he pressed on with *De rerum humanarum emendatione consultatio catholica (A General Consultation on the Reform of Human Affairs)*, a six-volume work which represents a synthesis of his life efforts. He almost succeeded in finishing it: some parts were published, but he found no support for a complete edition. Nor could his heirs arrange publication, for the same reason. For a long time the manuscript was believed lost. Only in 1935 was it discovered in the library of the Pietistic Orphanage in Halle, and

the first complete version was published in Prague in 1966.

The *Consultation*, in addition to its prologue ‘To the Lights of Europe’ which addresses Europe’s educated elite, consists of seven parts: ‘Panegersia’ (Universal Awakening) defines human matters, analyses their disconsolate state and calls for a quest for improvement; ‘Panaugia’ (Universal Enlightenment) explores possible ways to reform, and selects the most effective of them, ‘the bright light of Minds spreading everywhere and to everything’; ‘Pantaxia’ (Universal Order) or ‘Pansophia’ (Universal Science), the fundamental section of this work, outlines Comenius’ ontology by describing different tiers of world affairs right to the top one - an ideal human society; ‘Pampaedia’ (Universal Education) proclaims education for all, including adults, throughout their lives, including even ‘entirely barbaric nations’ which should be freed from ‘the darkness of their ignorance’; ‘Panglottia’ (Universal Eloquence), a general study of languages as a means for spreading the light, including an artificial philosophical language; ‘Panorthosia’ (Universal Improvement) shows how - thanks to all so far discussed - the state of education, religion and public administration can be improved ‘so that by God’s command an Enlightened Pious and Peaceful Age can be brought to Earth’; ‘Pannuthesia’ (Universal Call), a challenge to all educated people, lords spiritual and temporal, ‘as well as all Christians without difference, consciously to press for the bringing about of these desired and desirable things’.

Comenius’ thought is characterized by his effort to achieve synthesis. He accepts all sorts of intellectual ideas from the present and past, transforms them and incorporates them into his conceptual structures. First and foremost he brings to fruition the Czech Reformation tradition, the roots of which go back to Jan Hus. From this tradition stems his emphasis on the activity of humans - free beings who, in striving for perfection, come close to God, and transform the world through their action. In this he does not consider humans to be just a ‘res cogitans’ (a thinking substance), and does not reduce them to mere intellectual subjectivity. He sees humans in their complexity, with a developed emotional component and above all with a will which dominates other forces and attributes of human existence. Such a being has all the prerequisites for working upwards again to human perfection, once lost by succumbing to evil and vice.

Comenius built his early philosophy on the empiricism of Bacon. But this did not entirely satisfy him, and he searched for a universal knowledge which would be not just an agglomeration of empirical data but a logically structured whole where empirical findings would exactly fit into a solid framework of fundamental principles, to be found with the help of metaphysics. Comenius’ image of the world was initiated by Neoplatonism and by Nicholas of Cusa but these impulses are again re-shaped. He combines the dynamic descendancy of Neoplatonic Nature with the non-dynamic ascendancy of Aristotelian-Scholastic Nature, creating the picture of a dynamic ascendant world where practical human activity plays an active part. It is not a purely linear process as it has its regresses, but as a basic direction it points to the highest step - an ideal, harmonious human society. Comenius’ lifelong efforts for a universal reform aim at this ideal, as illustrated by his final project outlined in *Consultation*.

See also: Czech Republic, philosophy in; Medieval Philosophy; Neoplatonism

JOSEF ZUMR

List of works

Comenius, J.A. (1969-) *Johannis Amos Comenii Opera omnia*, ed. M. Steiner, Prague: Academia.(A critical edition of all his works in the original languages. Twenty-seven volumes are planned in total, of which thirteen have been published to date.)

Comenius, J.A. (1957) *John Amos Comenius 1592-1670. Selections*, intro. J. Piaget, trans. I. Urwin, Paris: UNESCO.(Produced in commemoration of the third centenary of the publication of *Opera didactica omnia*. Popular anthology of known Comenius’ philosophical and pedagogic texts.)

Comenius, J.A. (1623) *Labyrint světa a ráj srdce (The Labyrinth of the World and the Paradise of the Heart)*, trans. Count Lützow, London: The Golden Cockerel Press, 1950.(The greatest Czech work of Comenius - a symbolic and critical picture of the society of that time.)

Comenius, J.A. (1633) *Janua linguarum reserata*, Lipsiae: Gohofredi Grosii (Gottfried Gross); trans. as *The Gate of Tongues Unlocked and Opened*, London: T. Cotes, 1633.(The English translation of textbook of new methods of foreign language education.)

Comenius, J.A. (1639) *Panosophiae Prodromus*, London: L. Fawne and S. Gellibrand.(The first sketch of Comenius’ pansophism.)

Comenius, J.A. (1644) *Conatuum pansophicarum dilucidatio*, Lugduni Batavorum (Leyden): David Lopez de

Haro.(New version of pansophic theory.)
Comenius, J.A. (1657-8) *Magna Didactica J.A. Comenii*, ed. F.C. Hultgren, Farnborough: Gregg, 1968; trans. and ed. M.W. Keatinge as *The Great Didactic of John Amos Comenius*, New York: Russel & Russel, 1967.(The Hultgreen edition is a facsimile of the original Amsterdam collected edition. Keatinge's translation contains bibliographical, historical and critical introductions.)
Comenius, J.A. (1659) *Orbis Pictus*, ed. J.E. Sadler, London: Oxford University Press, 1968. (The reprint is a facsimile of the first English edition of 1659.)
Comenius, J.A. (*c*.1660s) *De rerum humanarum emendatione consultatio catholica (A General Consultation on the Reform of Human Affairs)*, ed. J. Červenka and V.T. Miškovská-Kozáková with J. Brambora, D. Čapková-Votrubová, J. Hora, J. Kopecký, J. Kyrášek and J. Moravec, 1st edn preface by J. Váňa with an epilogue by J. Patočka, Prague: Academia, 1966.(The synthesis of Comenius' life work not published in its entirety until 1966 - a critical edition. English translation by A.M.O. Dobbie of individual parts of the Consultation: *Pampaedia*, London: Buckland Ltd, 1986; *Panaugia or Universal Light* (1987), *Panegersia or Universal Awakening* (1990), *Panglottia or Universal Language* (1989), *Pannuthesia or Universal Warning* (1991), all Shipston-on-Stour, Warwickshire: Peter I. Drinkwater; *Panorthosia or Universal Reform* chapters 19-26 (1993), chapters 1-18 (1996), Sheffield: Sheffield Academic Press.)
Comenius, J.A. (1668) *Via Lucis*, Amsterdam: Ch. Cunrad; trans. as *The Way of Light*, Liverpool-London: The University Press, 1938.(A proposal for an international organization of scientific work.)

References and further reading

Acta Comeniana. International Review of Comenius Studies and Early Modern Intellectual History, (1969-), Prague.(Periodical containing papers by Czech and other researchers in the principal European languages.)
Studia Comenia et historica,ed. Musaeum Comenii Hunno-Brodense, Uherský Brod (1971-).(Periodical which brings articles from a wide field of history and philosophy, mostly in Czech or German.)
Pešková, J., Cach, J. and Svatoš, M. (eds) (1991) *Homage to J.A. Comenius*, Prague: Karolinum.(A collection of papers in English, exploring a wide range of Comenius' ideas.)
Sadler, J.E. (1966) *J.A. Comenius and the Concept of Universal Education*, London: Allen & Unwin.(An attempt at systematic interpretation of Comenius' philosophical and pedagogic views.)
Turnbull, G.H. (1947) *Hartlib, Dury and Comenius. Gleanings from Hartlib's Papers*, London: Hodder & Stoughton/University Press of Liverpool.(Profiles of Hartlib, Dury and Comenius, based on materials stored in Hartlib Papers in Sheffield.)

Common law

Common law and custom are features of most enduring legal orders. In English law the concepts have taken on special and interrelated significance, since English law is said to be grounded in common law and that in turn is said to derive from custom. According to classical common law theory, which crystallized in the seventeenth century, common law grew from the customs of the English people. It was not made by legal officials, as statutes are. Change was accommodated in this theory, on the basis not of identity *of elements of law over time but of* continuity*, a continuity of authority and reception of legal customs, and of the traditional legal order which declared them to be law. The role of legal officials - particularly judges - was to interpret and declare legal custom; their judgments provided* evidence *of it. They did not make it or invent it. This mode of development through continual interpretation and reinterpretation of the significance and bearing of the legal inheritance was, according to common lawyers, better adapted to social complexity, change and variety, and also to human epistemological and practical limitations, than attempts to cover any field with legislation.*

This theory was largely eclipsed in nineteenth-century England by the theory of legal positivism, and with it were eclipsed for a time some useful insights into social complexity and institutional limitations. Also lost was a sense of the complex dialectic between continuity and change in legal and institutionalized traditions. In its best moments, common law theory had such a sense.

1 Common law as custom

Common law is a term employed in many legal systems to mark a contrast between law which is in some way general or ‘common’ in scope - though the ways considered relevant differ markedly between different legal orders - and more particular types or branches of law once or concurrently applied in the same legal order. Custom, too, has figured widely as a significant term of art within complex legal orders (Kelley 1990). It typically refers to an element or source of law - unwritten, traditional, old, local - to be distinguished from other, usually newer, more deliberate, more explicit, written sources, which apply on a territorial rather than a communal or ‘personal’ basis, and which are administered by officials (Constable 1994). Both common law and custom are then, in part, contrast terms by reference to other aspects of legal systems; and frequently the contrast in mind is between the two of them.

Today when the common law is mentioned, however, it is usually in connection with the English legal tradition, and that tradition is distinctive in the way in which it has purported to connect, rather than contrast, common law and custom. For while it is acknowledged in this tradition that the institutions that administered common law supplanted those of customary law to form a national law, common lawyers have long insisted that the substance of English common law is drawn, and derives its legitimacy from, custom.

After the eleventh-century Norman conquest of England, the King’s Courts gradually extended their sway throughout the kingdom. Institutionally this was at the expense of a variety of local assemblies, tribunals, customary bodies and practices. Substantively, too, the centralized application of written laws had profound transformative effects on oral customs and traditions, even when the courts claimed to be applying customary law (Constable 1994). The law dispensed by the King’s Courts came to be known as the common law. This has led to one broad and two somewhat narrower uses of the term.

In its now most familiar usage, ‘common law’ is a global term for the legal tradition that was developed in Britain by these courts over centuries, and spread to its legal offshoots, such as the USA, Canada, Australia and New Zealand. The common law, in this sense, is most often contrasted with the civilian tradition (see Roman law §3) of continental Europe (and erstwhile colonies of European countries).

In a second and narrower usage, internal to the common law tradition, the law dispensed by the common law courts was distinguished from the specific ‘equitable’ interventions by the Lord Chancellor and the Court of Chancery to supplement traditional common law remedies (see Justice, equity and law §4).

Third, when statutes enacted by Parliament became prominent sources of English law, they were also distinguished from the common law developed by the courts. This third distinction persists, though statutes are more and more general in scope and effect and in many areas have come to override the common law.

English common law practice had many features that distinguished it from other legal traditions. The common law also spawned a distinctive theory of law. Classical common law theory crystallized most clearly in the seventeenth century and maintained its influence through the eighteenth. It insisted - to the dismay of Hobbes, Bentham and other legal positivists - that the common law, far from being made by any identifiable institutions or persons - kings or parliaments or even judges - grew from the customs of the English people. As Sir William Blackstone put it in the eighteenth century, the common law was an 'antient collection of unwritten maxims and customs' (Blackstone [1765-9] 1979, vol. 1: 17), and '[t]he only method of proving, that this or that maxim is a rule of the common law, is by shewing that it hath been always the custom to observe it' (Blackstone [1765-9] 1979, vol. 1: 68). Although the decisions of judges were acknowledged to be crucial elements in the development of the tradition in this perception, they did not make the common law, but were merely evidence of it.

The claim that the common law is grounded in custom addresses, often without much differentiation, descriptive, explanatory and normative issues, as well as social theory and legal and political philosophy. Several matters need to be unravelled, central among them the following three. First, whether, and if so how, an account of law as custom can make sense of legal change. Second, if law is popular custom, what is the role of legal officials? Third, what are the normative implications of the customary character attributed to the common law; what, if any, are its particular virtues?

2 Change and continuity in the common law

One of the most important interpreters of the language of the common law, Pocock (1987), has argued that the common lawyers' identification of common law and custom involved them in a paradox, stemming from an ambiguity inherent in the concept of custom itself. On the one hand, '[i]f the idea that law is custom implies anything, it is that law is in constant change and adaptation, altered to meet each new experience in the life of the people' (Pocock 1987: 36). On the other hand, 'it is equally possible to regard [custom] as that which has been retained throughout the centuries and derives its authority from its having survived unchanged all changes of circumstances' (Pocock 1987: 37). Pocock took the views of Sir Edward Coke, the seventeenth-century common lawyer and (from 1606-16) Chief Justice, as emblematic of the tradition, and emphasized that Coke and his contemporaries seemed largely in thrall to the second, unhistorical, interpretation. Later scholarship (Burgess 1992), and some clarifications by Pocock himself, suggest than even Coke might have been less single-mindedly ahistorical than Pocock was taken originally to suggest and, more important, that Coke's was never the interpretation widespread among the most distinguished and sophisticated common lawyers.

Such lawyers, among them in the seventeenth century Spelman, Vaughan, Selden and Hale, in the eighteenth Blackstone and Lord Mansfield, had no time for immemorial origins. Since evidence was poor and incomplete, there was no way of knowing such origins with any exactness. In any event it is not in the nature of law to remain unchanged:

> From the Nature of Laws themselves in general, which being to be accommodated to the Conditions, Exigencies and Conveniencies of the People, for or by whom they are appointed, as those Exigencies and Conveniencies do insensibly grow upon the People, so many times there grows insensibly a Variation of Laws, especially in a long tract of Time.
>
> (Hale [1713] 1971: 39)

Indeed, while Hale believed that deliberate legal change should be approached with caution and humility, he also thought it absurd to imagine that the law should remain unchanged: 'The matter changeth the custom; the contracts the commerce; the dispositions, educations and tempers of men and societies change in a long tract of time; and so must their lawes in some measure be changed, or they will not be usefull for their state and condition' (Hale [1665] 1787: 269-70).

If there was all the adding, subtracting, altering and creating that Hale describes, what held things together, in the absence of the glue of identity; what made all these different laws part of the same 'common law'? For Hale, and the common law tradition more broadly, change was an intrinsic characteristic of a law in constant organic evolution. Like a person, the law maintained its continuity through a process of continual change and growth. Or, to change the metaphor as several of them did, the common law survived change just as 'the Argonauts Ship was the same when it returned home, as it was when it went out, tho' in the long Voyage it had successive

Amendments, and scarce came back with any of its former Materials' (Hale [1713] 1971: 40). What made all its elements part of the same law was not their changelessness but their continuity. And this continuity had more to do with the continuing authority and reception of the law than it did with any demonstrable objective longevity of all its elements.

The source and ground of this continuity was twofold. First, it stemmed from the anchoring of the common law, developed by the courts, in popular customs which had been 'handed down by tradition, use and experience' (Blackstone [1765-9] 1979, vol. 1: 17) and had survived by constant transmission, reception, interpretation, refinement, transformation and application, from 'time whereof the memory of man runneth not to the contrary' (Blackstone [1765-9] 1979, vol. 1: 67). These customs had developed over the life of the country, and had indeed become constitutive of 'the people, into whom their ancient laws and customes are twisted and woven as a part of their nature…' (Hale [1665] 1787: 255). As these customs displayed continuity through change, so too did the common law, whose officials continuously drew upon them, interpreted them, rendered them explicit and made them binding.

Second, continuity was embedded in the characteristic practices of the legal institutions themselves, for these practices - and particularly those of the common law courts - were best understood as customary ones. As A.W.B. Simpson has revived and endorsed this aspect of the common law account:

> the common law system is properly located as a customary system of law in this sense, that it consists of a body of practices observed and ideas received by a caste of lawyers…. The ideas and practices which comprise the common law are customary in that their status is thought to be dependent upon conformity with the past, and they are traditional in the sense that they are transmitted through time as a received body of knowledge and learning.
>
> (Simpson 1987: 376)

Given such continuity, it was possible to accept, even to advocate, reform of the law, so long as it was done by common law methods, materials and institutions. And much of the most distinguished common law thought of the eighteenth century was devoted to advocating legal reform of precisely that sort (Lieberman 1989).

3 Judges and the common law

Notwithstanding their insistence that the common law reflects and rests upon the customs of the English people, not every custom was law. Customs became recognizably legal in the common law when legal officials recognized them (though this was not, for common lawyers, the source of their authority). Coke, Hale and Blackstone all paid great attention to the activities of such officials. However, whereas legal positivists insisted upon the primacy of legislators, the common law tradition did not. Instead its focus was on the common law judge. The declaration, determination, refinement, alteration and transformation characteristic of common law development were primarily the work of the judges of the common law courts: 'the depositary of the laws; the living oracles, who must decide in all cases of doubt, and who are bound by an oath to decide according to the law of the land' (Blackstone [1765-9] 1979, vol. 1: 69).

Given the acknowledged centrality of judicial pronouncements in its development, it is not surprising that, increasingly from the eighteenth century, the common law has been referred to as 'judge-made' law, distinguished in this character from statutes, which are made by parliaments. To put things this way, however, is implicitly (and today typically unconsciously) to reject the common law understanding of law altogether, in favour of understandings from within the tradition of legal positivism, which - in legal philosophy - has largely overwhelmed that of the common law. For the common lawyers did not view the judge as lawmaker, but as interpreter and declarer of custom, directly or as filtered through earlier judicial decisions; in Coke's words, 'the mouth of the law (for *judex est lex loquens* [the judge is the law speaking])'. This is not a passive role, but nor is it a legislative one. And while common law theory may not furnish an adequate account as it stands, it is neither unperceptive nor absurd.

The positivist critics of common law theory conceived of law as a system of substantive rules, made deliberately by particular persons or institutions (see Legal positivism). Existing rules would either cover a particular case, or not. If the former, the judge's obligation was to apply the rule to the case. If the latter, the judge would have to

supplement the rule in a legislative or quasi-legislative manner. To common lawyers, this set of alternatives would seem unrealistic and impoverished.

On the one hand, well into the nineteenth century common law *practitioners* tended not to think of the law as a collection of substantive rules at all. Rather it was a system of reasoning, and a collection of procedures and remedies, to be adapted to cases as they came before the court. Substance, on this view, came 'from below, in a constant feeding from society' (Lobban 1991: 79).

On the other hand, common law *theorists* did believe that there were substantive rules of the common law, but did not regard them as simply to be applied or made. In any event, unlike statutes, they did not come in a fixed and delimited textual form, ready to be applied. The meaning and bearing of customs and precedents was rather of continual argument, interpretation and reinterpretation. More generally, common law theorists regarded the law, not as a systematic collection of discrete rules, but as a traditional or customary order, of language, thought, maxims, principles, understandings, values and rules - with its own ways of knowing, thinking and arguing, handed down over generations. Lawyers and judges are trained and participate in this order, and interpret it. The judges are the authoritative guardians of this tradition. They conserve it, bring it to bear on particular cases, adapt, apply and transmit it. New things are said and done, of course, on this 'Argonauts Ship' - but by experts working from inside, not out, with materials already on board, and with a keen and experienced sense of the risks involved in straying too far.

The notion here is one of skilled and experienced participants in a familiar and intricate customary order, who deal with both repeated and new problems and changing circumstances by working with old materials. Moreover they do not merely work with those materials, but think in terms that they provide, making distinctions and analogies that make sense in their terms, terms endogenous to the legal materials rather than imposed on them from without. Here, to vary the metaphor yet again, not merely is the statue within the block of stone, but so too are the tools and even the sculptor as well.

Morever, the common law judges are not merely participants in the legal tradition; they are *experts* in the law. This is an expertise not available to everyone without long experience and training. Rather, as Hale observed, 'men are not borne Comon Lawyers' (Hale 1966: 505). Without training and long experience they lack what Coke called 'the artificial reason and judgment of law, which law is an act which requires long study and experience, before that a man can attain to the cognizance of it' (Coke [1608] 1907: 1342-3).

4 The virtues of the common law

An intensely live question, perhaps even now, was what, if anything, England gained from law which was predominantly of this sort: grounded in custom, and determined by judges. More generally, what might any legal order have to gain from law which, unlike legislation 'is not laid down; customs, we know, grow up' (Simpson 1987: 363)? Simply put, the answer of Jeremy Bentham and his followers to these questions was: nothing much and nothing good. For the early positivists' model for law was overwhelmingly legislative. Legislation was the model both for what law is and for what good it can do. Laws are the general commands of legally unlimited sovereigns to habitually obedient subjects. They can do good if clear-headed, deliberate, purposeful legislators issue explicit and unambiguous general commands to deal with social problems and to reform the law - particularly the common law. On this view, either the 'common law' really is inchoate custom or judicial resolution of particular disputes - and then it is no law at all, for no general commands have been issued - or it is purported general commands wrested from individual cases, but in that case power has been arrogated by the 'sinister interests' of 'judge and co.', as Bentham described the masters of the common law.

The common lawyers, by contrast, believed that the common law owed its moral authority (which they did not always sharply distinguish from its legal validity) to its immemorial reception, rather than to its legally irresistible imposition. And they praised it above all for those features which most distinguished it from legislation. Many of the themes of common law theory were elaborated more famously by Edmund Burke (§3) and more recently by F.A. Hayek. They are elements of what might be called the methodological conservatism of common law theory.

First, there is a stress on the irreducible complexity and interconnectedness of social life. The problems and circumstances of life are extraordinarily various, complex, interdependent and changing. Many of the most successful institutions have been the products of no one's design, and the reasons for their success might be

unknown to anyone in particular. There is too much to know for any individual adequately to understand what makes for success in legal arrangements or social arrangements generally - even those that already exist - let alone confidently to predict the results of legal interventions (see Social theory and law §5).

Given the inherent complexity of the world, then, and the inescapable limitations on what we can know and what we can hope to do, thoughtful, cautious, corrigible elaboration and refinement of solutions is preferable to imperious imposition of them. The incremental methods of the common law are suited to the first; legislation to the second. For unlike legislators, as the eighteenth-century Scottish jurist Lord Kames (see Home, H. (Lord Kames)) insisted, '[a] court of justice determines nothing in general; their decisions are adapted to particular circumstances…. They creep along with wary steps, until at last, by induction of many cases…a general rule is with safety formed' (quoted in Lieberman 1989: 162).

Moreover, common law judges do not creep unguided, with just their puny individual 'natural' reason to light their way. For the common law in which they are immersed provides them with a great deal on which to draw. Given the difficulties in doing so, customs that have survived some long time must, at the very least, be presumed to have weathered and adapted to many trials, both literal and metaphorical. Common law judges have the vast storehouse of cases and decisions to draw upon. This law, in its organic and recorded development, encompasses multitudes of cumulative decisions, and these draw upon social customs since time immemorial. These customs - and the judicial reflection upon and refinement of them - embody, or at the very least should be presumed to embody, wisdom far deeper than could be available to any unaided individual, for 'time is the wisest thing under heaven' (Hale [1665] 1787: 270) and the common law is 'fraught with the accumulated wisdom of ages' (Blackstone [1765-9] 1979, vol. 4: 435).

Some interpreters have suggested that the common lawyers favoured custom over reason, but that is not the way they saw it or put it. On the contrary, they insisted that the common law combined both; indeed that its attention to the former uniquely revealed the dictates of the latter. At its most philosophical, this is a variation on Aquinas' distinction (see Aquinas, T. §13) between *deduction* as one way in which particular conclusions can be derived from the natural law, and *determination*, 'of certain general features… like that of the arts in which some common form is determined to a particular instance…. So the natural law establishes that whoever transgresses shall be punished. But that a man should be punished by a specific penalty is a particular determination of the natural law' (*Summa theologiae*, IaIIae.95.2). Determinations will vary according to the ways in which circumstances, customs, other particular conditions and human ingenuity themselves vary. Determinations cannot conflict with the natural law from which they are derived, but nor are they uniquely correct implications of them. For there are no such uniquely correct implications. The common law, dealing with the complex particularities of social life, had general elements of the first sort, but in its detail was mainly of the second. Indeed it is a remarkable store of such determinations, as adapted to and coloured by the particular, complex and variable conditions in which the law must operate, and refined by the sustained application of the 'artificial reason and judgment of law'.

In this way one could insist on the existence of a law of nature, accessible to reason, which the common law could not violate, and at the same time insist that laws apt for some people at some times could not simply be transplanted to other people and other times. In this way, too, one could, by praising the reason immanent in the common law, deny that the common law was simply whatever judges had decided. One could insist, on the contrary - and the eighteenth-century common lawyers increasingly did - that deeper principles underlay the common law, could be discerned through expert interpretation of it, and furnished grounds for decisions which appeared to depart from particular precedents. Such decisions furthered or were said to further values immanent in the common law - general principles 'which run through the cases', in Mansfield's phrase.

To modern positivists this appears to be a sophistical mask for judicial legislation, but this is not how it appeared then, nor is it necessarily so. Rather, underlying it are conceptions which emphasize the particularity and practicality of the judges' task, the collective nature of the publicly justified reasoning processes in which they indulge, and the wisdom of reflecting in a sustained manner on what the tradition in which one works has done in similar circumstances. It is not necessarily the mere reactionary resistance to change, and obscurantist deference to authority, that Jeremy Bentham took it to be.

5 Conclusions

The common lawyers had many commitments on which most contemporary observers would be at best agnostic. They shared, for example, a hagiographic commitment to the common law as the accumulation of the wisdom of ages. And while they did not claim that all of the common law merely stems from the customs of the people - for one of its sources is old statutes - the common lawyers saw no difficulty in claiming that it was grounded directly in the customs of the people rather than merely in the activities of the lawyers and the decisions of the judges.

Many critics have found extremely questionable this premonition of historical jurisprudence, with its homogenization of official law with social life, and of social life itself (see Jurisprudence, historical). Of course the common law has a continuing link with the common life, if only because some cases are brought by ordinary folk, and lawyers and judges are required to meditate on whatever problems are brought before them. It is one thing, however, to say that as a reactive social institution which is driven by the cases brought before it, the common law is likely to adapt over time to the ‘Conditions, Exigencies and Conveniencies of the People’; although even this is not a simple matter, for the cases that come before the courts - particularly the higher courts - are not random samples, and there are open and difficult questions about how and how much that adaptation occurs, about what law responds to in society and about what in law responds. It is another thing entirely to say that judicial law meshes directly, smoothly and constantly with popular custom. Questions abound not only about the degree of autonomy of official law from other social practices and traditions, but also about the complex and variable relationships between official and lay customs and traditions, and about the significance of social and political power in maintaining elite practices. There are also a host of other questions, for example, about what elements in plural, stratified, heterogeneous societies - including England, notwithstanding its insularity - judges’ law might be related to. While it is obvious and important, then, that law and life are intimately related, the relationships are not simple, unidirectional, unmediated or uniform. To say that the products of specialized and differentiated institutions represent a distillation of national mores and customs is to raise and to beg a host of questions of fact, morals and social theory upon which it is difficult to pronounce in general or even with regard to the long history of the common law.

Nevertheless, the common lawyers had a better appreciation than many of their successors of the complexity of social affairs, of the significance of continuity in legal life, of the need for social grounding of legal institutions, of some of the resources available to schooled participants in a traditional order, and of the inextricable relationships between continuity and continuous change that are characteristic of enduring and complex normative traditions, such as the common law. Not all of these insights were retained when the cloying hagiography of common law partisans was thrown out of philosophical discussions of law.

See also: Law, philosophy of; Savigny, F.K. von; Selden, J.; Stair, J.D.

MARTIN KRYGIER

References and further reading

Blackstone, Sir W. (1765-9) *Commentaries on the Laws of England*, 4 vols, Oxford: Clarendon Press, 1787; Chicago, IL: Chicago University Press, 1979.(The 1979 edition contains valuable editorial introductions to each volume.)

Burgess, G. (1992) *The Politics of the Ancient Constitution*, London: Macmillan. (Excellent account of English political and legal thought in the seventeenth century, with thorough and thoughtful revisions to Pocock’s theses about the ‘common law mind’.)

Coke, Sir E. (1608) ‘Prohibitions del Roy’, in *77 English Reports*, London: Stevens and Sons, 1907, 1342-3.(The case was 1608, but is incorporated in the 1907 *English Reports*.)

Constable, M. (1994) *The Law of the Other*, Chicago, IL: University of Chicago Press.(Study of transformation of the customary institution of the ‘mixed jury’ with interesting reflections on the replacement of customary law by ‘official’ law.)

Hale, Sir M. (1665) ‘Considerations Touching the Amendment or Alteration of Lawes’, in F. Hargrave (ed.) *A Collection of Tracts Relating to the Law of England*, vol. 1, 1787, 249-89; repr. Oxford: Professional Books, 1982.(See §2 above for discussion.)

(1713) *The History of the Common Law of England*, ed. C.M. Gray, Chicago, IL: University of Chicago Press, 1971. (Published posthumously; printed three times, 1713, 1716, 1739, virtually as it stood in manuscript and numerous times thereafter with editorial additions.)

(1966) ‘Sir Matthew Hale’s Criticisms on Hobbes’s Dialogue of the Common Laws’, repr. in Sir W. Holdsworth,

A History of English Law, vol. 5, 2nd edn, London: Methuen, 419-513.(Unpublished in Hale's lifetime.)

Kelley, D.R. (1990) *The Human Measure*, Cambridge, MA: Harvard University Press.(Masterly historical discussion of the role of custom in European legal and social theory.)

Lieberman, D. (1989) *The Province of Legislation Determined*, Cambridge: Cambridge University Press. (Excellent discussion of eighteenth-century English and Scottish, common law and positivist, views on the rival claims of common law and legislation - especially as sources of law reform.)

Lobban, M. (1991) *The Common Law and English Jurisprudence, 1760-1850*, Oxford: Clarendon Press.(Study of the tensions between common law practitioners' remedies-focused assumptions about the nature of the common law and the more rule-based focus of eighteenth-century common law system builders, such as Blackstone, and critics of the common law tradition, such as Bentham.)

Pocock, J.G.A. (1987) *The Ancient Constitution and the Feudal Law: A Reissue with a Retrospect*, Cambridge: Cambridge University Press.(Classic study of the 'common law mind' with extensive reply to critics of the first edition.)

Postema, G.J. (1986) *Bentham and the Common Law Tradition*, Oxford: Clarendon Press.(One of the best available accounts of the legal theory of the common law tradition, and of Jermey Bentham's criticisms of, and alternative to, that theory.)

Simpson, A.W.B. (1987) 'The Common Law and Legal Theory', in A.W.B. Simpson, *Legal Theory and Legal History*, London: Hambledon Press, 359-82. (Excellent critique of positivist accounts of the common law, in favour of a slightly modified common law theory.)

Common Sense School

The term 'Common Sense School' refers to the works of Thomas Reid and to the tradition of Scottish realist philosophy for which Reid's works were the main source. The ideas of the school were carried abroad - to France; and to the USA, where they were highly influential, particularly among leading academics critical of Calvinism. Interest in Reid and the tradition to which he gave rise was revived almost a century later by leading American philosophers and their students.

P. Royer Collard introduced Thomas Reid's work to France, and Victor Cousin and Théodore Jouffroy became disciples of the Scot, though they were sometimes loathe to give full credit to their mentor. Through Cousin's influence Reid's expressionist theory of art had a continuous influence throughout the nineteenth century on French aesthetics, culminating in a modern version of expressionism developed by René Sully-Prudhomme, the first recipient of the Nobel prize in literature in 1901.

John Witherspoon emigrated early in life to the United States and introduced Reid and Stewart to the young republic, where they, along with Cousin, exerted a major philosophical influence for fifty years. Their free-will agency theories were used effectively in undermining the Old School Calvinism of the Congregational and Presbyterian Churches. Among the most effective followers of Reid, Stewart and Cousin - all critics of Calvinism - were Francis Wayland (Brown University, RI), Asa Mahan (Oberlin College, OH), Henry P. Tappan (University of Michigan, MI), and Alexander Campbell (Bethany College, WV). Like Witherspoon and George Campbell before him, James McCosh emigrated from Scotland to be the president of Princeton and the last great light of the Scottish tradition in America, providing what the tradition had hitherto lacked - a realistic analysis of causality. The interest in Reid and the Scottish tradition as a whole was revived in the United States in the twentieth century by the writings of C.J. Ducasse, Roderick Chisholm, Wilfred Sellars and their students.

While it must not be thought that all the members of the tradition held identical views, there was a common core to the Scottish realistic tradition to which most of these people adhered.

According to Reid, a number of concepts and principles in daily use could not in principle be learned from experience. The concept of space, for example, could not be learned from experience since every perception presupposes it. Hence concepts and judgments which cannot be learned must be the result of nativistic epistemic input, since that is the only other alternative. The judgments supplied nativistically are numerous and constitute what Reid called the principles of common sense. The following is a small sampling of these principles: every event must have a cause; people have some degree of power over their own actions and decisions; qualities perceived by the senses must have a subject called body or substance. These principles and numerous others proclaim what is either evident or self-evident and hence require no justification other than rebuttal of criticism. Any judgment may be accepted at face value unless there is a reason to doubt it. But the only acceptable reason for doubting such a judgment would be that it conflicts with other evident judgments. Hence no discursive philosophy can throw doubt on them singly or as a class. Such judgments carry more authority than any discursive arguments that are designed to show them to be false. Moreover, common-sensical principles are inviolable because they are universally held, and they are unavoidable in the sense that denying them is pointless (in that further sense that doing so never gets rid of them). The philosopher who momentarily denies them reaffirms them in the market place or, even while denying them, acts upon them unwittingly. That Hume pointed out this state of affairs (see Hume, D. §2) was, for Reid, a very honest if self-defeating thing for a sceptic to do.

Philosophies which deny the truths of common sense principles must contain a fallacy, and the job of the philosopher is to detect that fallacy. That such philosophies must contain a fallacy follows from the fact that denying them leads to absurd consequences and from the fact that evident and self-evident judgments always constitute a better reason for keeping such principles than discursive reasons can provide for disallowing them. Since philosophers on the whole are clever reasoners we must not usually look for the fallacies in their arguments but for some faulty premise with which they begin.

An example of a faulty premise which leads to absurd consequences by rejecting an evident universal and unavoidable judgment is the premise shared by Locke, Berkeley, and Hume. Locke's premise is that the only

things of which we are directly aware are our own ideas, impressions, sensations, phenomena, or whatever else the alleged non-physical entity of direct awareness insisted upon might be called. Once granted this premise, the slippery slope from Locke to Berkeley to Hume's scepticism is inevitable. Since no sensation can in principle in any way resemble an object, Locke's representative realism is untenable and we are left with the equally untenable alternatives: Berkeley's subjective idealism, which avoids scepticism only by inconsistently allowing direct knowledge of the self, or Hume's scepticism, the claim that we can never know or have any good reason to believe that our senses ever give us reliable information. Reid found no good argument to support Locke's premise - indeed, Locke and most philosophers generally take it for granted - and he argued convincingly that Hume's 'diminishing table' argument to bolster this premise was not only not successful but proved the opposite viewpoint. Having dismissed Locke's premise, Reid offered instead his adverbial-like analysis of sensation, which construed sensation as an act of the mind and dispensed with non-physical entities altogether (Reid 1872: 245-306; 'diminishing table' argument, 303- 5). It was this aspect of the common sense tradition which re-emerged in the work of contemporary adverbial theorists like Ducasse, Chisholm, and Sellars.

See also: Aberdeen Philosophical Society; Beattie, J.; Buffier, C.; Commonsensism; Moore, G.E. §3

EDWARD H. MADDEN

List of works

Cousin, V. (1828) *Cours de l'histoire de la philosophie moderne*, Paris: Pichon & Didier, 3 vols; trans. O.W. Wight *Course of the History of Modern Philosophy*, New York: Appleton 1852.(Cousin's use of the word 'absolute' has misled numerous commentators into characterizing him as an absolute idealist.)

(1853) *Du vrai, du beau, du bien*, Paris: Didier; trans. O.W. Wight *The True, the Beautiful and the Good*, New York: Appleton, 1854.(Shortened version of Cousin's lectures and a nineteenth-century translation of the 3rd edition.)

Jouffroy, T. (1838) *Philosophical Miscellanies*, Boston, MA: Hilliard, Gray.(Collection of philosophical and literary essays which, despite the title, is crucial for understanding Jouffroy's thought.)

(1840) *Introduction to Ethics, Including a Critical Survey of Moral Systems*, trans. W.H. Channing, Boston, MA: Hilliard, Gray.(This book was a standard text for ethics courses in American universities during the long dominance of Reid's philosophy in America.)

Mahan, A. (1881) *A Critical History of Philosophy*, London: Elliot Stock.(Probably Mahan's most important work. He developed his own version of Reid's doctrine in volume 2; note particularly his discussion of Kant.)

McCosh, J. (1867) *Intuitions of the Mind*, revised edn, New York: Carter.(McCosh's most important book, containing a more sophisticated analysis of the concept of cause than had heretofore appeared in the common sense tradition.)

Reid, T. (1785) *Essays on the Intellectual Powers of Man*, ed. B. Brody, Cambridge, MA: MIT Press, 1969.(The best edition since it avoids all of Hamilton's footnotes which are systematically misleading.)

(1788) *Essays on the Active Powers of the Human Mind*, ed. B. Brody, Cambridge, MA: MIT Press, 1969.(The best edition since it avoids all of Hamilton's footnotes which are systematically misleading.)

(1872) *Collected Works*, ed. W. Hamilton, Edinburgh: Maclachlan & Stewart, 7th edn.(Hamilton's footnotes are systematically misleading.)

Stewart, D. (1826) *Esquisses de philosophie morale (Outline of moral philosophy)*, trans. T. Jouffroy, Paris: Johanneau.(This book was influential in the formulation of Jouffroy's own moral philosophy.)

Stewart, D. (1854-60) *Collected Works*, ed. W. Hamilton, Edinburgh: Constable.(Text is good, but the footnotes should be avoided. Strangely, Hamilton never had a clear understanding of the common sense tradition.)

Wayland, F. (1835) *The Elements of Moral Science*, ed. J.L. Blau, Cambridge, MA: Harvard University Press, 1963.(Blau's text is by far the best available; his Introduction throws much light on the Reidian tradition in America.)

References and further reading

Barker, S.F. and Beauchamp, T.C. (eds.) (1976) *Thomas Reid: Critical Interpretations*, Philadelphia, PA: Monograph Series.(Analyses and clarifies numerous aspects of Reid's philosophy.)

Grave, S.A. (1960) *The Scottish Philosophy of Common Sense*, Oxford: Clarendon Press.(Presents a flawed interpretation of Reid's views on sensation and perception, but helpfully promoted the resurgence of the new interest in the Common Sense School.)

Lehrer, Keith (1989) *Thomas Reid*, London: Routledge.(Claims Reid's major contribution is his combining

nativism, psychology and epistemology.)
Madden, E.H. (1968) *Civil Disobedience and Moral Law*, Seattle, WA: University of Washington Press.(Deals with the deontological ethics and moral reforms of American common sense advocates.)
(1986) 'Stewart's Enrichment of the Common Sense Tradition', *History of Philosophy Quarterly* 3 (1): 45-63. (Portrays Stewart as a significant philosopher, by no means a shadow of Reid.)
Manns, J.W. (1993) *Reid and his French Disciples*, Leyden: Brill.(Exhibits the best understanding of the common sense tradition yet written.)
Rowe, W.L. (1991) *Thomas Reid on Freedom and Morality*, Ithaca, NY: Cornell University Press.(Constitutes the most thorough examination of Reid's agency theory.)
Stewart, D. (1803) *Account of the Life and Writings of Thomas Reid*, Edinburgh, William Creech.(Portrays Reid as a gentle, modest and amiable man and casts significant light on the nature and scope of Reid's philosophy.)

Common-sense ethics

'Common-sense ethics' refers to the pre-theoretical moral judgments of ordinary people. Moral philosophers have taken different attitudes towards the pre-theoretical judgments of ordinary people. For some they are the 'facts' which any successful moral theory must explain and justify, while for others the point of moral theory is to refine and improve them. 'Common Sense ethics' as a specific kind of moral theory was developed in Scotland during the latter part of the eighteenth century to counter what its proponents saw as the moral scepticism of David Hume. Thomas Reid, the main figure in this school, and his followers argued that moral knowledge and the motives to abide by it are within the reach of everyone. They believed that a plurality of basic self-evident moral principles is revealed by conscience to all mature moral agents. Conscience is an original and natural power of the human mind and this shows that God meant it to guide our will. A deeply Christian outlook underwrites their theory.

'Common-sense ethics' refers to the pre-theoretical moral judgments of ordinary people (see Moral judgment). Moral philosophers have taken different positions on the significance of such pre-theoretical judgments with respect to moral theory. Some see our pre-theoretical judgments as the 'facts' which any successful moral theory must systematize, explain and/or justify. Others claim that moral theory should improve our common-sense moral judgments which without theory may be distorted by local prejudices and biases. 'Common Sense ethics', however, is a specific type of moral theory, which originated in Scotland during the latter part of the eighteenth century and the early nineteenth century. Thomas Reid (1785; 1788) originally developed the Common Sense moral theory in response to what he perceived as David Hume's sceptical attacks on morality. There are two central features of Hume's moral theory which Reid took to be sceptical. The first is Hume's claim that moral approval is a matter of sentiment. According to Hume (1739/40), moral approval is a sentiment, a calm form of love: when we consider persons' characters impartially, sympathy with the persons themselves and those with whom they usually associate causes us to feel love for their good qualities, and this love is moral approval. Reid interpreted Hume's claim that approval is a matter of sentiment in a way which anticipated the emotivist analysis of moral 'judgments' (see Emotivism; Hume, D. §4). According to Reid, for Hume moral approval and disapproval are merely expressions of agreeable or disagreeable feelings which are neither true nor false. When I condemn a person, I am not making a judgment about the person, but am only expressing some uneasy feeling I have (see Moral scepticism §3).

The second point which Reid found to be sceptical concerns Hume's notorious claim that 'reason is and ought only to be the slave of the passions' (1739/40: 415) (see Reasoning/rationality: practical §1; Practical reason and ethics §2). Hume argued that, strictly speaking, the idea of rational action makes no sense. Reason helps us to discern the means to our ends, but it neither selects nor ranks ends. It does not provide us with motives of either prudence or duty. Reid interpreted this as implying that reason has no say in determining our ends. Reason does not even help us to form a conception of our ends. Our ends are supplied by brute feelings.

Reid addressed both of these points in his theory of conscience: conscience is an original and natural power of the mind, common to all human beings. Reid thought this showed that God meant it to guide our wills. It is both an intellectual and active power. As an intellectual power, it enables us to intuit directly the first principles of morality. Reid thought that moral reasoning, and indeed all reasoning, must start from self-evident first principles which we perceive immediately. If we had to figure out the basic principles of morality by a process of ratiocination, as Locke maintained, morality would not be within the reach of everyone. Since morality is required of everyone, 'the knowledge that is necessary for all, must be attainable by all' (1785: 481). But only mature moral agents, those of 'ripe understanding', and free of interest, passion and prejudice, are in a position to see the self-evident moral principles. Moral education is therefore necessary.

Reid did not object to calling conscience a moral sense since he believed that a correct analysis of sensing shows that it involves judgment and so reason. Hume's mistake, according to Reid, was that he failed to see that approval is a complex act of the mind with feeling as only one component. Reid insisted that we all know the difference between feeling and judging. He granted that when I approve of someone, I experience an agreeable feeling, but claimed that this feeling is different from and dependent upon the prior judgment that the person's conduct merits esteem. Persuade me that the agent was bribed and both my esteem and agreeable feeling vanish (see Moral sense

theories).

Conscience is also an active power. Reid followed Joseph Butler in arguing that there are several distinct sources of motivation. Reid called the motives which do not presuppose judgment or reason ‘animal’ principles of action; these include our appetites, desires and affections. Those which require the use of reason are, in contrast, rational principles of action. One rational principle is our overall good. Reid argued that without reasoning we could not even form a conception of our overall good and, once we arrive at a conception of it, we necessarily seek it. It is a governing principle, to which our animal principles ought to be subordinated. The other rational principle is duty, which conscience reveals to us and which motivates us to do our duty for duty’s sake. Reid thought that the concept of duty is simple and refers to a relation between an agent and an action. Conscience’s authority, its right to govern us, is self-evident and superior to that of interest. God’s wisdom and benevolence guarantees that no agent will be a ‘loser’ by doing his duty.

Reid believed that there is a plurality of equally fundamental moral truths. Some of these concern the preconditions and range of moral judgments, for example, that only voluntary acts deserve praise. Others are substantive self-evident principles and concern our basic duties to ourselves, others and God. Reid’s list of self-evident axioms includes such principles as that we ought to act benevolently towards others, that we should treat others as we would judge it right for them to treat us, and that we owe God veneration and submission. Reid addressed an important problem which the earlier rationalists failed to see and his own followers ignored: if there is a plurality of substantive self-evident moral principles, they may come into conflict. According to Reid, the virtues, as dispositions to act according to moral principles, support and strengthen one another. But in concrete instances one principle may direct us to perform an action, while another principle forbids its performance. To use Reid’s example, ‘what generosity solicits, justice may forbid’ (1788: 639). Reid thought that in such cases we rely on self-evident priority rules, for instance, that ‘unmerited generosity should yield to generosity and both to justice’ (see Moral pluralism).

Reid’s followers, such as James Beattie and Dugald Stewart (1793; 1828) accepted the main elements of his moral theory. They agreed that the first principles of morality are self-evident and that there is a plurality of them (see Stewart, D.). They distinguished among animal and rational principles of action. They shared Reid’s belief that God meant conscience to be our guide and that it is an intellectual and active power of the human mind. On the whole they were noncritical proponents of the Common Sense moral theory, adding little to it.

Kant’s famous reaction to the philosophers of the Common Sense school was a characteristic one. Instead of answering Hume’s sceptical challenge, according to Kant (1783), they merely appealed to the common sense of humankind, without any proper insight into the question. While this may be true of Reid in his more unguarded moments and of his disciples more generally, the Common Sense moral theory was influential not only in France and England but also in the US. By the mid-nineteenth century, William Whewell’s version of rational intuitionism became dominant (1845), overshadowing the Common Sense moral theory.

There is a relation between common-sense ethics and the Common Sense moral philosophers: the Common Sense moral philosophers trusted the conscience of ordinary people. Philosophical ethics, although useful in bringing self-evident moral truths to light and systematizing them, has a limited role to play in revising them. If there is a conflict between ‘the practical rules of morality, which have been received in all ages’ and those advanced by theorists, as Reid says, ‘the practical rules ought to be the standard by which the theory is corrected’ (1788: 646). But the Common Sense moral philosophers’ confidence in the judgments of ordinary people sprang from their conviction that conscience is ‘the candle of the Lord set up within us’ (Reid 1788: 597). It is grounded in the belief that God gave us the capacity not only to know what we ought to do but also to do it.

See also: Intuitionism in ethics; Moral justification; Moral knowledge

CHARLOTTE R. BROWN

References and further reading

Beattie, J. (1790-3) *Elements of Moral Science*, New York: Garland Publishing Company, 2 vols, 1977.(A collection of lectures delivered at Marischal College on psychology, natural theology, moral philosophy, politics and economics.)

Grave, S.A. (1960) *The Scottish Philosophy of Common Sense*, Oxford: Clarendon Press.(The most

comprehensive critical discussion of Common Sense moral philosophy, including coverage of Reid, Stewart and Beattie.)

Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978, books II, III.(Contains Hume's discussion of reason's role in action and his theory of moral approval.)

Kant, I. (1783) *Prologomena zur einer jeden künftigen Metaphysik die als Wissenchaft wird auftreten können*, trans. J. Ellington, *Prolegomena to Any Future Metaphysics*, Indianapolis, IN: Hackett Publishing Company, 1971.(Kant disparages the appeal of the Common Sense moral philosophers to common sense in their attempt to settle metaphysical questions; in the introduction, Kant argues that the Comon Sense philosophers failed to understand Hume's sceptical challenge to our concept of causality and the implications of this challenge for metaphysics as a whole.)

Mill, J.S. (1861) *Utilitarianism*, Indianapolis, IN: Hackett Publishing Company, 1979, esp. ch. 2.(Seeks to make utilitarianism acceptable by working hard at minimizing the conflicts between the deliverances of utilitarian theory and those of common-sense ethics.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Rawls' idea of 'reflective equilibrium' incorporates the idea of testing principles of justice with what he calls our 'considered judgments'. These are the moral judgments we are most confident in since self-interest or other biases have been filtered out.)

Reid, T. (1785) 'Of Reasoning', in *Essays on the Intellectual Powers of Man*; in W. Hamilton (ed.) *Works*, Edinburgh, 1846-63, 2 vols; repr. Hildesheim: Georg Olms Verlagsbuchandlung, 1967.(The page number given for the quotation cited in the above entry comes from Hamilton's edition. In ch. 2 of the essay, 'Whether Morality be capable of demonstration', Reid argues against the Lockean view that we arrive at the first principles of morality by means of demonstrative reasoning.)

Reid, T. (1788) 'Of the Rational Principles of Action' and 'Of Morals', in *Essays on the Active Powers of Man*; in W. Hamilton (ed.) *Works*, Edinburgh, 1846-63, 2 vols; repr. Hildesheim: Georg Olms Verlagsbuchandlung, 1967.(The page numbers given for the quotations cited in the above entry come from Hamilton's edition. This work contains Reid's moral theory. It also includes Reid's libertarian account of freedom. Although Reid often misinterprets Hume, some of his criticisms of Hume are insightful. Reid thought that Hume uncritically accepted the theory of ideas and that his mistakes may be traced to this. Includes an appendix in which Hamilton explains the idea of common sense.)

Schneewind, J.B. (1977) *Sidgwick's Ethics and Victorian Moral Philosophy*, Oxford: Clarendon Press, ch. 2. (Contains a clear, concise and informative discussion of the main elements of Reid's moral psychology, his account of the liberty of the will, and justice. Also included is a short but helpful discussion of Reid's followers who wrote about morality.)

Stewart, D. (1793) 'Of the Active and Moral Powers of Man', in *Outlines of Moral Philosophy*; in W. Hamilton (ed.) *Collected Works*, Edinburgh: T. & T. Clark, Constable 1854-77; repr. Westmead: Gregg International, 1971. (Based on his lectures, Stewart wrote this for the use of students at Edinburgh. This work includes a discussion of the various springs to action, the origin of our moral ideas, and our duties to God, others and ourselves. An appendix contains an outline of his political theory.)

Stewart, D. (1828) *Philosophy of the Active and Moral Powers of Man*, ed. and abridged by J. Walker, Boston, MA: Phillips, Samson and Co., 1868.(An expanded version of the *Outlines*, this work contains a critical survey of British modern moral philosophy. Stewart argues against Hobbes and others that the sense of duty cannot be reduced to self-love. He opposes Paley's associationism and Bentham's utilitarianism. The influence of Reid is especially evident, as well as that of Cudworth, Shaftesbury, Butler and Price. Stewart tends to avoid appeals to a common sense.)

Whewell, W. (1845) *The Elements of Morality, Including Polity*, London: J.W. Parker, and New York: Harper & Bros, 2 vols.(Whewell's account of rational intuition is sometimes closer to that of Kant's than Reid's. He emphasizes that intuition is an activity of reason and not just a passive perception of first principles.)

Common-sense reasoning, theories of

The task of formalizing common-sense reasoning within a logical framework can be viewed as an extension of the programme of formalizing mathematical and scientific reasoning that has occupied philosophers throughout much of the twentieth century. The most significant progress in applying logical techniques to the study of common-sense reasoning has been made, however, not by philosophers, but by researchers in artificial intelligence, and the logical study of common-sense reasoning is now a recognized sub-field of that discipline.

The work involved in this area is similar to what one finds in philosophical logic, but it tends to be more detailed, since the ultimate goal is to encode the information that would actually be needed to drive a reasoning agent. Still, the formal study of common-sense reasoning is not just a matter of applied logic, but has led to theoretical advances within logic itself. The most important of these is the development of a new field of 'non-monotonic' logic, in which the conclusions supported by a set of premises might have to be withdrawn as the premise set is supplemented with new information.

The formal study of common-sense reasoning is a field in which the concerns of philosophy and artificial intelligence converge. From a philosophical point of view, it is motivated by a popular approach to the philosophy of mind - known as the 'language of thought' hypothesis - that postulates a domain of mental representations as the primary bearers of meaning, and then analyses cognition as the rule-governed manipulation of these representations. In order for an approach along these lines to be useful, some account must be provided of the structure of the internal representations, and of the rules governing their manipulation (see Language of thought).

These issues have been studied extensively within cognitive psychology, of course. Here, the idea is to focus on an existing system capable of intelligent behaviour in a wide range of circumstances - a human being - and to attempt to infer the symbolic processes underlying this intelligence. But the issues are studied also within the area of artificial intelligence, and from a perspective that offers a different kind of illumination: the goal here is to build an intelligent system, rather than attempting to discover the structure of a system that already exists. Although the performance exhibited by the artificial systems designed to date is significantly less impressive than that of humans when evaluated across a broad range of circumstances, the symbolic processes underlying the behaviour of these systems are at least well understood.

Among the various research methodologies that have emerged within artificial intelligence, one of the most powerful is the logic-based approach first advanced by John McCarthy; and it is primarily this approach that has motivated the formal study of common-sense reasoning. In fact, the logic-based approach in artificial intelligence has much in common with the project of formalizing mathematical and scientific reasoning that has been carried on throughout the twentieth century. Within artificial intelligence, the idea is that much of the knowledge necessary for achieving intelligent action even in everyday situations can be represented in a machine through the formulas of some logical language, and that the reasoning tasks underlying this intelligence can then be accomplished by means of logical deduction.

Both the idea of using logic as an underlying representation language for artificial intelligence and the emphasis on formalizing common-sense knowledge are present even in McCarthy's very early work (for example, McCarthy 1959), but the project receives its clearest articulation in a paper jointly authored by McCarthy and Patrick Hayes (1969), which isolates the task of defining 'a naïve, common-sense view of the world precisely enough to programme a computer to act accordingly'.

Much of the work now produced within artificial intelligence on the topic of formalizing common-sense reasoning is similar to what one finds in philosophical logic - attempts to adapt or generalize logical techniques to apply to some new area - and there is now a good deal of interaction between these two fields. Still, the research that takes place within artificial intelligence has a somewhat different character from standard philosophical logic. For one thing, the matter of implementation is always in the background, and sometimes in the foreground; but even apart from that, the studies generated within artificial intelligence are often focused on very detailed representational problems that would seem peculiar to a philosopher. It would not be unusual, for example, to find a research project in artificial intelligence with the goal of representing what an agent would have to know about the objects

in their kitchen in order to prepare a meal.

This kind of detailed work is necessary, of course, since the ultimate goal is to encode information in such a way that it could actually be used to drive a reasoning agent. Still, it is not as if the project of providing a logical representation of common-sense knowledge were simply a matter of applied logic. In fact, the problems presented by the task of representing this kind of information explicitly have led to a number of theoretical advances within logic itself: the most important of these is the development of the new field of 'non-monotonic' logic.

In most standard logics, the addition of new information to a given set of premises might lead us to draw new conclusions, but never to withdraw conclusions already reached. The set of consequences of a given set of premises is thus said to grow monotonically as the premise set grows: the consequence set can only increase, never decrease, as the premise set is supplemented with new information. A non-monotonic logic is simply a logic in which this property fails - a logic in which the addition of new information to a given premise set might force us to retract some conclusion drawn from the original set of premises (see Non-monotonic logic).

The study of non-monotonic logics within artificial intelligence was motivated by the realization that much of our common-sense knowledge concerns defeasible information - generalizations subject to exceptions, such as 'Birds fly' or 'Things remain where you put them'. Of course, philosophers had always known that these defeasible generalizations could not be represented naturally in ordinary logic: the statement 'Birds fly', for example, cannot be represented by a formula of the form $\forall x(Bx \supset Fx)$. But it was not until the practical need arose for reasoning with defeasible statements such as these that serious attention was focused on the problem; and it then became apparent that the appropriate notion of defeasible consequence would have to be non-monotonic.

To take a standard example, given only the information that Tweety is a bird and that birds tend to fly, it is natural to conclude defeasibly that Tweety flies. But if this premise set were supplemented, consistently, with the additional information that Tweety does not fly (perhaps Tweety is a penguin), it would then seem reasonable to withdraw our initial conclusion.

The monotonicity property flows from assumptions that are deeply rooted in both the proof theory and the semantics of most ordinary logics. From a proof-theoretic point of view, this property follows from the fact that a proof based on a set of premises also counts as a proof based an expansion of the same set of premises; from a semantic point of view, the property results from the assumption that the models of a set of premises are models also of any of its subsets. Because the features underlying the monotonicity property are so basic to the conception of most standard logical systems, researchers concerned with non- monotonic reasoning have been led to explore fundamentally new ideas in both proof theory and semantics.

The first proof-theoretic treatment of non-monotonic reasoning is found in Raymond Reiter's default logic (1980), which supplements ordinary logic with new rules of inference, known as 'default rules'. The semantic approach to non-monotonic reasoning was first explored in McCarthy's own theory of circumscription (1980).

Because of its intrinsic interest and practical importance, the study of non-monotonic logic has grown into a significant area of research. The fixed-point and minimal model approaches pioneered by Reiter and McCarthy are still the best-known and most widely applied techniques in non-monotonic reasoning, but a number of other approaches have been explored as well; a series of articles surveying these different approaches can be found in Gabbay *et al.* (1994). In recent years, the techniques developed within the field of non-monotonic logic have begun to find applications also to a number of philosophical issues, such as the understanding of *ceteris paribus* clauses in scientific generalizations and the formalization of *prima facie* obligations.

See also: Artificial intelligence

JOHN HORTY

References and further reading

Davis, E. (1990) *Representations of Commonsense Knowledge*, San Mateo, CA: Morgan-Kaufmann.(A textbook in the field of formalizing commonsense reasoning.)

Gabbay, D., Hogger, C.J. and Robinson, J.A. (eds) (1994) *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, *Nonmonotonic Reasoning and Uncertain Reasoning*, Oxford: Oxford University Press.(Definitive and detailed survey articles on a variety of non-monotonic logics.)

Horty, J. (1994) ‘Moral Dilemmas and Non-Monotonic Logic’, *Journal of Philosophical Logic* 23: 35-65.(An application of non-monotonic logic to problems in moral reasoning.)

McCarthy, J. (1959) ‘Programs with Common Sense’, in D.V. Blake and A.M. Uttley (eds) *Proceedings of the Symposium on Mechanization of Thought Processes*, London: H.M. Stationery Office; expanded version repr. in R. Brachman and H. Levesque (eds) *Readings in Knowledge Representation*, San Mateo, CA: Morgan-Kaufmann, 1985.(An early statement of McCarthy’s views.)

McCarthy, J. (1980) ‘Circumscription - A Form of Non-Monotonic Reasoning’, *Artificial Intelligence* 13: 27-39.

McCarthy, J. and Hayes, P. (1969) ‘Some Philosophical Problems from the Standpoint of Artificial Intelligence’, in B. Meltzer and D. Michie (eds) *Machine Intelligence*, vol. 4, Edinburgh: Edinburgh University Press.

Reiter, R. (1980) ‘A Logic for Default Reasoning’, *Artificial Intelligence* 13: 81-132.

Thomason, R. (1988) *Philosophical Logic and Artificial Intelligence*, Dordrecht: Kluwer.(Some essays at the intersection of philosophical logic and artificial intelligence.)

Thomason, R. (1991) ‘Logicism, Artificial Intelligence, and Common Sense: John McCarthy’s Program in Philosophical Perspective’, in V. Lifschitz (ed.) *Artificial Intelligence and Mathematical Theory of Computation*, San Diego, CA: Academic Press.(An article relating the project of formalizing common-sense knowledge to the logicist programme in philosophy.)

Commonsensism

'Commonsensism' refers to one of the principal approaches to traditional theory of knowledge where one asks oneself the following Socratic questions: (1) What can I know?; (2) How can I distinguish beliefs that are reasonable for me to have from beliefs that are not reasonable for me to have? and (3) What can I do to replace unreasonable beliefs by reasonable beliefs about the same subject-matter, and to replace beliefs that are less reasonable by beliefs that are more reasonable? The mark of commonsensism is essentially a faith in oneself - a conviction that a human being, by proceeding cautiously, is capable of knowing the world in which it finds itself.

Any inquiry must set out with some *beliefs. If you had no beliefs at all, you could not even begin to inquire. Hence any set of beliefs is better than none. Moreover, the beliefs that we* do *find ourselves with at any given time have so far survived previous inquiry and experience. And it is psychologically impossible to reject everything that you believe. 'Doubting', Peirce says, 'is not as easy as lying'. Inquiry, guided by common sense, leads us to a set of beliefs which indicates that common sense is on the whole a reliable guide to knowledge. And if inquiry were not thus guided by common sense, how would it be able to answer the three Socratic questions with which it begins?*

1 Background: Reid and Peirce

The term 'commonsensism' was introduced by Charles Sanders Peirce who was concerned to contrast his own 'critical commonsensism' with the views of Thomas Reid and with those of the other members of the Scottish Common Sense school of philosophy. According to Reid, the 'principles of common sense' are intuitive truths that all sane people accept when they are not doing philosophy. They are no less reasonable than the truths of logic and mathematics.

Reid is most persuasive in criticizing those philosophers who reject the principles of common sense. He calls our attention to the extremes of British empiricism, culminating in the absurdities of Hume's empirical system. These alleged absurdities include the following: we cannot know anything about the past; we cannot know whether there are any material things; and we cannot even know that we ourselves exist. Reid observes:

> A traveller of good judgment may mistake his way, but when it ends in a coal-pit, it requires no great judgment to know that he hath gone wrong, nor perhaps to find out what misled him'. The committed empiricist who is not yet prepared to abandon his extreme epistemological views may point out that Reid cannot demonstrate, to the empiricist's satisfaction, that the empiricist is not in a delirium. 'But how does he know that he is not in a delirium? I cannot tell; neither can I tell how a man knows that he exists. But, if any man seriously doubts whether he is in a delirium, I think it is highly probable that he is, and that it is time to seek for a cure.
>
> (Reid [1764] 1854: 107)

A somewhat more patient and difficult response may be found in the critical commonsensism of Peirce, which may be thought of as a refinement upon the views of Reid. Peirce says that his own view 'arises out of a contrite fallibilism, combined with a high faith in the reality of knowledge, and an intense desire to find things out' (Peirce 1931-5: 1.14) (see Fallibilism).

2 Critical commonsensism: a systematic treatment

'The slogans are impressive enough,' one may say, 'but how are they to be applied?' In setting out, one presupposes that, by contemplating various possible beliefs, we can find out that some of them logically imply others, that some contradict others, that some are such that they serve to confirm others (they make the others probable) and that some are such as to disconfirm others (they make the others improbable). Probability, as Peirce conceives it, is 'a thing to be inferred upon evidence'.

Two quite different procedures are involved. The first is that of ridding ourselves of beliefs that we should not have. The second is that of reconstruction - that of 'building anew'. We may call the second procedure 'the road back'. Descartes distinguishes the two steps in the first of his *Meditations*: 'I was convinced that I must once for all seriously undertake to rid myself of all the opinions which I had formerly accepted, and commence to build anew from the foundation'. Critical commonsensism, therefore, is a version of foundationalism (see Foundationalism).

To see how commonsensism is applied in the theory of knowledge, we should consider the ‘building anew’ that is involved in the road back. Having some faith in ourselves, we start out with our native common sense - with what we find ourselves inclined to believe. Where else, after all, could we start out? If, like Peirce, we accept critical commonsensism, we will make this assumption: the mere fact that we find ourselves believing one thing rather than another is itself a *prima facie* reason for believing that thing. One way to improve upon this mass of uncritical beliefs is to sift it down and try to cast away the things that should not be there. If we find a set of beliefs that contradict each other, we will try to sift it down in such a way that the remaining set is not contradictory. If we then find that the remaining set disconfirms some of its members, then we will proceed in analogous fashion, in the hope that the surviving set will not thus disconfirm any of its members.

We also appeal to the experiences of perceiving and remembering. Suppose you think that you are perceiving a sheep. That you think you are perceiving a sheep gives some *prima facie* justification for that belief. And at this stage of the road back, there may be still more to be said for the belief. If the belief is confirmed by the set of other beliefs that has so far survived your critical scrutiny, then it is more reasonable than any belief not so confirmed.

What has been said about perceiving may also be said, *mutatis mutandis*, about remembering. Traditional empiricism, Peirce points out, is not adequate to the epistemic status of memory (see Empiricism). Still another source of epistemic respectability is the possibility of concurrence (also called ‘coherence’ and ‘mutual support’). A set of beliefs may be said to concur, or to be related by mutual support, provided that any of its members is confirmed by the conjunction of all the others. If we find such a set among the beliefs that we still have, then we may say that the whole now has a still greater degree of epistemic respectability (see Knowledge and justification, coherence theory of).

In summarizing his approach to the theory of knowledge, Peirce calls attention to ‘one of the most wonderful features of reasoning and one of the most important philosophemes in the doctrine of science, of which, however, you will search in vain in any book I can think of: namely, that reasoning tends to correct itself, and the more so, the more wisely its plan is laid. Nay, it not only corrects its conclusions, it even corrects its premises’ (Peirce 1931-5: 5.575).

3 Commonsensism and metaphysics

The ‘defence of common sense’ associated with G.E. Moore should not be considered as being primarily an attempt to deal with the problems of the traditional theory of knowledge. It is intended, rather, as a corrective to what Moore felt were some of the excesses of the metaphysicians of his day, particularly those in the tradition of absolute idealism (see Moore, G.E. §§2-3). Some of these philosophers, for example, tried to prove that there is no valid distinction between appearance and reality. In lectures on this topic, Moore would refute such views in the following way. He would hold up his hand, saying ‘Here is a hand’, and he would then point out that this obvious fact was inconsistent with the proposed theory about appearance and reality. If their reasoning were sound, then they would not be justified ‘in believing that this is a hand’. Some of these philosophers had also felt that they could prove that ‘time is unreal’. Moore calls to their attention what, apparently, many of them had not noticed - that their thesis has the absurd consequence that no one ever knows whether or not they had their breakfast before having their lunch. Such refutations met with considerable indignation but not with any very convincing rejoinders.

See also: Contextualism, epistemological; Rational beliefs §2

RODERICK M. CHISHOLM

References and further reading

Chisholm, R.M. (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 3rd edn, 1991.(A contemporary statement and defence of critical commonsensism.)

Chisholm, R.M. (1996) *A Realistic Theory of Categories: An Essay on Ontology*, Cambridge: Cambridge University Press.(A summation of Chisholm’s views on an enormous range of topics.)

Descartes, R. (1641) *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes*, vol. 2, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91.(Descartes’ main metaphysical work, mentioned in §2 above.)

Moore, G.E. (1922) *Philosophical Studies*, London: Routledge & Kegan Paul. (See especially chapter 6 entitled

'The Conception of Reality'.)
Moore, G.E. (1959) *Philosophical Papers*, London: George Allen & Unwin.(Chapter 2 'A Defence of Common Sense', and chapter 7 'Proof of an External World', are perhaps the clearest writings that exist on the relation of common sense to philosophy.)
Peirce, C.S. (1931-58) *The Collected Papers of Charles Sanders Peirce*, vols 1-6 ed. C. Hartshorne and P. Weiss, vols 7-8 ed. A. Burks, Cambridge, MA: Harvard University Press.(The indexes should be consulted, as references to commonsensism abound throughout the volumes.)
Reid, T. (1764) *An Inquiry into the Human Mind on the Principles of Common Sense*, in *The Collected Works of Thomas Reid*, ed. Sir William Hamilton, Edinburgh: Maclachlan & Stewart, 1854.(Reid's clearest statement of his views about common sense can be found in this early work.)

Communication and intention

The classic attempt to understand communication in terms of the intentions of a person making an utterance was put forward by Paul Grice in 1957. Grice was concerned with actions in which a speaker means something by what they do and what is meant might just as much be false as true. He looked for the essence of such cases in actions intended to effect a change in the recipient. Grice saw successful communication as depending on the recognition by the audience of the speaker's intention. Since then there have been many attempts to refine Grice's work, and to protect it against various problems. There has also been worry that Grice's approach depends on a false priority of psychology over semantics, seeing complex psychological states as existing independently of whether the agent has linguistic means of expressing them.

1 Grice's theory

The classic attempt to understand communication in terms of the intentions of a person making an utterance was put forward by Paul Grice in his 1957 article. Grice takes as his subject what he calls 'non-natural meaning' (meaning_{NN}), which he contrasts with the wider class of cases of natural meaning in which we say, for instance, that 'those spots mean measles' or 'those clouds mean rain'. It cannot be true that those spots mean measles but the patient has not got measles; meaning is simply a correlation or signalling. Whereas Grice is concerned with actions in which a speaker (or, in general, an agent) means something by what they do and what is meant might just as well be false as true. Here too he makes a distinction: between a wider class of actions that predictably cause a recipient to believe something, and a narrower class of actions that not only may cause the belief that *p*, but also themselves mean that *p*. The wider class would include such actions as leaving photographs around, expecting the recipient to believe that the events shown had happened. The narrower class should take us somewhere near the domain of linguistic communication.

Grice looks for the essence of such cases in actions intended to effect a change in the recipient. He eventually located the difference in the manner in which the change in belief is expected to come about. In the wider class, the action may simply be a natural sign of the state of affairs, but in the narrower class of non-natural meaning there is an element of reflexivity involved, since the intended mechanism, through which the recipient is expected to come to the belief, is that the agent is recognized as acting with the intention that the recipient come to believe it: it is only if the agent's intention is recognized that the change in belief will ensue. So, for example, deliberately ignoring someone in the street might be said to mean_{NN} that the agent is angry with the recipient, because getting the recipient to believe this depends upon their recognizing the intention with which the action was performed. Grice proposes that '*A* meant_{NN} something by *x*' is roughly equivalent to '*A* uttered *x* with the intention of inducing a belief by means of the recognition of this intention'. He sees successful communication as requiring recognition of the speaker's intention in communicating. Of course, the account could be extended to speech acts other than assertion: for example, intentions to get people to perform actions or supply information can be used to fix what is meant by imperatives and questions.

Armed with an account of the speaker's meaning on an occasion, Grice goes on to propose that 'the utterance *x* meant something' is roughly equivalent to 'someone meant something by *x*'. For a sentence to mean something 'timelessly' is for people in general to use it to mean that thing. We can thus progress from an account of communication in a 'one-off' case to a more general account of the kind of meaning possessed by the sentences of a public language.

2 Reactions

Grice's short article spawned a huge literature, and various modifications were suggested. One is that it may not be quite right to identify the speaker's intention as one of changing the hearer's belief. I may say that *p* without caring whether my hearer accepts or indeed already believes that *p*. A change in belief might be one effect of my utterance, but it is not always one I intend (in the terms introduced by Austin, it is a 'perlocutionary' effect), whereas we are looking for a characterization of what a speaker is doing in uttering anything (the 'illocutionary' act; see Speech acts §1). It seems rather that the only reliable intention behind an utterance is that the speaker intends the hearer to understand something by it, not necessarily to believe it or react to it in any other

way. But already it is a matter of fine judgment whether this is important, and, after all, one can say things with the intention that the hearer not even understand them, for example when showing off one's grasp of some technical term. Certainly, a defence of Grice would go, once the whole social practice of communication with language is in place, we can say things with the most devious of intentions. Nevertheless, it may be true that the central or paradigm case is that of imparting information, and indeed any account that fails to put this function of language at the centre risks distortions of its own. So Grice may simply be following the proper path by concentrating upon the central or basic case and hoping that deviations will become explicable in the light of what is said about it.

A more elaborate debate queried whether Grice's conditions are sufficient for meaning. It concentrates on cases where there is an element of deceit. As Bennett puts it, 'in real communication everything is open and above board' (1976: 126), but a range of cases suggest that Grice's conditions could be met while there are hidden and devious intentions around. Grice's conditions could be satisfied while, further up the hierarchy of intentions, I have the intention that you misunderstand something about the situation. Complex counterexamples of this structure were rapidly concocted. However, what Grice called 'sneaky intentions' can be blocked by a number of strategies. One such is to require that all the speaker's intentions be known to the hearer, and another is that the intentions in question should be 'mutually known', where this means that the hearer knows the speaker's intentions, the speaker knows that the hearer knows, the hearer knows that the speaker knows…and so on.

There is a cost to piling up the reflexive intentions with which speakers are credited, for it is easy to wonder whether there is a psychological reality to these complex layers of intention, or whether they are simply an artefact of the theory. In fact, some unease arises when we realise that Grice's approach requires that the reflexive intention - the intention that the hearer's belief be modified by means of their recognition of my intention in speaking - be typically present throughout the field of linguistic communication. It would seem to be to overload the psychology of the young child or not-very-aware adult to credit them with intentions of such complexity: typical speakers, we might say, just do not know or care *how* their messages get across, so long as they do. Certainly a strategist faced with a hitherto unknown type of problem of communication may do something with the hope that the intention with which they do it is recognized, but this seems to be a sophisticated plan for coping with precisely the kind of situation where normal communication has broken down.

This kind of thought suggests that Grice is too much concerned with one-off cases. At least in typical cases of communication we are not involved in one-off strategies, but can rely upon the conventional meanings of our terms. (For Grice himself, as we have seen, the existence of these conventions would simply be the existence of enough people inclined to use a term to effect communication in enough cases.) While reflexive intentions may be necessary to solve new problems of communication, as when we confront someone with whom we share no language, it may be that they drop out of the picture when a rule-governed or conventional medium of communication is available: it may be that reliance on conventions and rules supplants the mechanism of recognizing the speaker's intention, rather than supplementing it. Once we have language I do not have to know even whether you intended me to understand you in order to know what you said. Similarly, there may be sentences whose typical use is not to convey the messages that their strict and literal meanings suggest, but nevertheless, when uttering one of them, one can be held to have said whatever it is that its strict and literal meaning identifies. In short, a conventional and deeply normative system (that is, one in which the speaker is *liable* to be held to have said things, regardless of what was intended) takes on a life of its own, independent of the detailed case-by-case intentions of participants.

Behind this objection there lies the deeper unease that Grice's approach sets things up so that the speaker has intentions of considerable complexity, and we then try to understand meaning and linguistic communication given so much rich psychology. Whereas a permanent strand in modern philosophy of language has been to try to identify the mastery of language with the ability to have thoughts (or, at least, thoughts of this kind of complexity) at all. From this standpoint it may seem perverse arbitrarily to enrich the psychology of the actors, and on that basis to explain their use of sentences as vehicles of their given, antecedent intentions. You cannot intend, the thought goes, without representing to yourself what you intend, and typically this will be by using the best representative medium known to us, which is our natural language. In forming an intention we typically speak to ourselves, saying that we will do this or that, and if this process is itself seen as a kind of internal communication then intentions are not suitable for giving an analysis of communication in general. So we should be thinking of social communication as one of the activities that makes possible the enriched psychology that we find in

participants in conversational exchanges. The problems in this area are, however, deeply intractable. Some philosophers, notably John Searle (1983), insist that the 'intentionality' of language (that is, its directedness or power to represent absent or merely possible states of affairs) must be understood as being derived from a more fundamental power of the 'mind/brain'; in such a scheme there is no problem with Grice's direction of attack. To others it merely mystifies things to postulate intentional powers in the mind/brain, and our capacity to represent absent states of affairs to ourselves must be seen as essentially linguistic, since language is the only representational system that we actually know anything about.

A final radical criticism is that Gricean proposals try to secure openness in communication by adding to the speaker's intentions. But, it has been argued (Meijers 1994), openness cannot be secured that way. It is a matter of a relationship between the speaker and hearer, not a one-way matter of the speaker having sufficiently open intentions. A speaker must not only have the intention to be open, but there must also be a commitment to being open (and liability to penalty if they are not). Arguably this takes us into the domain of collective intentions; ones based on an understanding between the different parties about what they, collectively, are trying to do. The relation between speaker and hearer, by such an account, is more like that between us when we together plan that *we* sing a duet, than when I unilaterally intend that *I* do something with some effect on you. Again, there are various proposals about the nature of such common intentions, and whether they reduce to a kind of aggregate of individual intentions.

3 Prospects and alliances

Sir Peter Strawson (1971) detected a 'Homeric opposition' between followers of Grice and philosophers of language, such as Davidson, who put the notion of a truth-condition at the centre of their picture. It is fair to say that this opposition has softened in the intervening years. Clearly there is no formal contradiction between supposing that a remark in the mouths of members of some community means that *p* if and only if there exists a practice of using it with the intention of communicating that *p*, and saying that the same remark means that *p* if and only if it represents the state of affairs that *p*, or has the truth-condition that *p*, or would be interpreted by an ideal interpreter as saying that *p*. The opposition is a question of whether we give priority to representation over communication, or vice versa. If we think of sentences possessing meaning by way of some kind of accord with things and facts, then their role in communication will seem a kind of bonus; if we think of the seamless way in which linguistic behaviour is woven into communication (and other activities involving the world) then we will reject any appeal to basic representative powers of elements of the 'mind/brain', in favour of a clearer understanding of the representative powers of the whole person, here meaning the linguistically active, socially embodied person communicating with others of the same nature.

The perennial importance of Grice's work is that intention and communication cannot be separated for long. If I intend to tell you that the cat is sick, we have not communicated unless this is how you take my utterance. If I intend a remark ironically or condescendingly, then we have not communicated unless you understand the irony or the condescension. And there is a limit to the possibility of unintended meaning: although on an occasion someone may mean something by an utterance that they did not intend, or fail to mean something that they did intend, such occasions are essentially parasitic. If they occur too often, then the conventions shift, and the meaning of the remark in the language realigns itself with the way the speakers intend it. There is no separating semantics and psychology, but there is the perennial question of priority. The strange and intriguing interdependency of intention and language seems to be two-way: our words can convey no more than we intend by them, but we ourselves can intend only what our words will carry.

Further work on these topics clearly requires a better understanding of whether meaning is essentially social. Work on rule-following has sometimes issued in the view that it must be so, and that determinate meaning only emerges in a fully social, normative practice in which the applications of words are routinely subject to criticism and correction. If this kind of thought can finally be defended, then Grice's programme will have been thoroughly vindicated, in its direction if not in every detail.

See also: Meaning and communication; Meaning and rule-following

SIMON BLACKBURN

References and further reading

Avramides, A. (1989) *Meaning and Mind*, Cambridge, MA: MIT Press.(A thorough investigation particularly of the issue of reduction of semantics to psychology, and whether Grice's approach requires belief in it.)

Bennett, J. (1976) *Linguistic Behaviour*, Cambridge: Cambridge University Press.(A lively development of Grice's work, especially valuable for its discussion of the relation between full intentionality and more primitive signalling systems.)

Grice, P. (1957) 'Meaning', *Philosophical Review* 66: 377-88; repr. in *Studies in the Way of Words*, Cambridge, MA, and London: Harvard University Press, 1989.(The seminal article in the field.)

Lewis, D.K. (1969) *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.(The classic modern analysis of the concept.)

Meijers, A. (1994) *Speech Acts, Communication and Collective Intentionality*, Utrecht: Leiden University.(An impressive defence of the ineliminable presence of collective intentions in communication.)

Searle, J.R. (1969) *Speech Acts*, Cambridge: Cambridge University Press.(A lucid development of J.L. Austin's work on the actions we perform in conversation.)

Searle, J.R. (1983) *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press. (Searle's work is perhaps the most impressive development of the notion of acts performed in speech, but also controversial through his firm belief that the intentional powers of language are derived from a prior biologically engendered capacity.)

Strawson, P.F. (1971) 'Meaning and Truth', in *Logico-Linguistic Papers*, London: Methuen.(An early reflection upon the relations between communication and representation.)

Communicative rationality

The concept of 'communicative rationality' is primarily associated with the work of the philosopher and social theorist Jürgen Habermas. According to Habermas, communication through language necessarily involves the raising of 'validity-claims' (distinguished as 'truth', 'rightness' and 'sincerity'), the status of which, when contested, can ultimately only be resolved through discussion. Habermas further contends that speakers of a language possess an implicit knowledge of the conditions under which such discussion would produce an objectively correct result, and these he has spelled out in terms of the features of an egalitarian 'ideal speech situation'. Communicative rationality refers to the capacity to engage in argumentation under conditions approximating to this ideal situation ('discourse', in Habermas' terminology), with the aim of achieving consensus.

*Habermas relies on the concept of communicative rationality to argue that democratic forms of social organization express more than simply the preferences of a particular cultural and political tradition. In his view, we cannot even understand a speech-act without taking a stance towards the validity-claim it raises, and this stance in turn anticipates the unconstrained discussion which would resolve the status of the claim. Social and political arrangements which inhibit such discussion can therefore be criticized from a standpoint which does not depend on any specific value-commitments, since for Habermas achieving agreement (*Verständigung*) is a 'telos' or goal which is internal to human language as such. A similar philosophical programme has also been developed by Karl-Otto Apel, who lays more stress on the 'transcendental' features of the argumentation involved.*

1 Instrumental and communicative rationality: Habermas and the Frankfurt School

Jürgen Habermas, the leading thinker of the second generation of the Frankfurt School, introduced the concept of communicative rationality in order to correct what he perceives to be the 'normative deficit' in the work of earlier thinkers in this tradition. This deficit consists in the lack of any philosophically perspicuous grounding for their critique of modern society. The thought of the Frankfurt School before Habermas, classically summarized in Theodor Adorno and Max Horkheimer's *Dialectic of Enlightenment* (1944), viewed social relations within the most advanced capitalist societies as being almost entirely shaped by the demands of 'instrumental reason'. Strongly reminiscent of Max Weber's 'purposive rationality' (*Zweckrationalität*), the concept of instrumental reason refers to the capacity to maximize efficiency in the control of objective processes through a knowledge of the determinants of such processes. For Adorno and Horkheimer the predominant institutionalization of this *aspect* of reason, reflected in its pervasive philosophical equation with reason as such, has socially disastrous consequences. It leaves the ends to be collectively pursued beyond the scope of rational determination, and gives rise to a 'totally administered society'. At the same time however, as Habermas repeatedly complains, the earlier Critical Theorists propose no alternative broader conception of reason, from the standpoint of which the restriction of reason to its instrumental aspect could be coherently criticized (see Frankfurt School; Critical theory).

The contrast between instrumental and communicative rationality, which represents Habermas' attempt to remedy this situation, is anchored in the philosophical anthropology systematically presented in *Knowledge and Human Interests* (1971). Here Habermas argues that any society, in order to reproduce itself, must be capable both of productive exchange with nature (in the form of labour), and of the communicative coordination of collective activities. Labour gives rise to concepts which articulate a 'technical interest' in instrumental control, whereas the need for agreement generates the distinct categorial framework of a 'practical interest' in hermeneutic understanding. In later writings Habermas specifies that participants in 'communicative action' must be capable of regarding the statements raised in the course of discussion from the standpoint of their *validity*, and must possess an implicit knowledge both of the appropriate procedures for settling disputed validity-claims and of the (invariably counterfactual) conditions under which following such procedures would produce an objectively correct result. Such knowledge is mobilized, for example, when discovery of covert coercion renders a previously achieved consensus invalid.

In the Western philosophical tradition, the concept of rationality has long been connected with the ability to reflect on and give grounds for one's beliefs and the actions they inform. However, in modern philosophy, it is often only beliefs capable of guiding goal-oriented action which are considered as candidates for rationality. Against this,

Habermas argues that the type of action oriented towards reaching agreement in language is irreducible to that which is oriented towards successful intervention in the objective world. The illocutionary aims of speech acts can only be achieved through cooperation, based on a free acceptance by others of the validity-claims raised by the speaker, which cannot be reduced to a causally producible effect. Hence the ability to achieve consensus by offering grounds reflects a distinct form of rationality which ‘inhabits everyday communicative practice’, and which Habermas therefore describes as ‘communicative rationality’.

With regard to the problem of grounding a critical social theory, two aspects of Habermas’ proposal are significant. First, Habermas extends the scope of validity beyond ‘truth’ (*Wahrheit*). He argues that claims to normative rightness (*Richtigkeit*), typically moral or legal claims, are in principle susceptible of the same form of resolution through discussion as cognitive claims. There can thus be an equivalent of truth in practical matters, although Habermas admits that the status of a claim to ‘sincerity’ (*Wahrhaftigkeit*) cannot be resolved through discussion. (An additional claim, concerning the ‘intelligibility’ (*Verständlichkeit*) of one’s utterance, is only discussed sporadically by Habermas, and is perhaps best considered as a *precondition* of these principal validity-claims.) Second, canons of argument are themselves not merely relative to specific cultural and institutional contexts. Whenever we engage in argumentation, Habermas proposes, we must implicitly assume (however counterfactually) that the conditions of an ‘ideal speech situation’, in which the ‘unforced force of the better argument’ would indeed ultimately triumph, have been fulfilled, otherwise discussion would lose its point. Such a speech situation would be characterized by the equal right of all participants to raise issues, ask questions, pose objections and so on, and therefore provides a normative yardstick against which current decision-making procedures can be critically assessed. In particular, the widespread dominance of the functional requirements of ‘social systems’ such as the market economy and modern bureaucracy can be seen as embodying a one-sided ‘rationalization’ which suppresses the rational potential of those democratic principles which are equally fundamental to modernity.

2 Objections to Habermas’ approach

The concept of communicative rationality faces objections from a variety of quarters. If such rationality is to represent more than the ability to adhere to culturally specific canons of argumentation, then Habermas must show the universal necessity of those pragmatic presuppositions of communication which form the ideal speech situation. The first step towards this is to argue for an internal relation between meaning and validity. Habermas contends that ‘ we understand a speech-act when we know what makes it acceptable’ ([1981] 1984: 400), where acceptability refers to intersubjective recognition of validity and not to correspondence with a reality whose structure would supposedly be given prior to language. He then draws the conclusion that ‘ an orientation towards the possible validity of utterances belongs not just to the pragmatic conditions of reaching understanding, but to the understanding of language as such’ (1988: 76).

However, Habermas’ account of the preconditions for grasping linguistic meaning may be implausibly strong. Herbert Schnädelbach (1991), for example, has objected that to understand what makes an utterance acceptable does not require one to take a stand on whether it is, in the present instance, valid. We can observe communicative action - the exchanging of reasons - from a third-person perspective without being drawn into the process of judging validity-claims, so that a normative standpoint is not internal to the very application of the concept of communicative rationality. In reply, Habermas has argued that even in apparently unfavourable cases, such as a sudden burst of laughter during a speech, we cannot understand the laughter without knowing if it is genuine or not (and thus a valid ‘utterance’), and this involves *assessing* the grounds for it. The methodological suspension of our spontaneous yes/no attitude towards validity-claims merely indicates that the task of interpretation is being postponed.

A further major objection to Habermas’ views has been raised by Albrecht Wellmer, among others. It is that there may indeed be certain idealizations which function as necessary pragmatic presuppositions of linguistic communication, but that the notion of an ideal speech situation represents an ‘objectivistic misinterpretation’ of these. This is because an ideal speech situation, exempt from the fallibility, opacity and temporality of human language, would be beyond the conditions which make communication itself necessary and possible. To regard the striving for such a situation as morally incumbent upon us would be to regard us as obliged to attain the impossible, which is absurd. According to Wellmer (1986), communication indeed relies upon ‘performative

idealizations' (for example, the conviction that the reasons we have for a claim are sound and will stand the test of time), but to transform these idealizations into an ideal to be practically pursued is to fall prey to a 'dialectical semblance' (*Schein*). Habermas, however, continues to maintain that it is meaningful to idealize forms of communication, just as we idealize physical measurements; in other words, to anticipate the full realization of conditions we must already presume to hold to some degree when engaged in serious argumentation.

Other critics have focused on Habermas' claim that reaching an understanding can be seen as the primary function of language, to which we are implicitly committed as soon as we speak. Thinkers influenced by post-structuralism, in particular, have objected that the commitment to achieving consensual truth cannot, except by begging the question, be considered more fundamental than playful, ironic or fictional uses of language. More generally, it is contended that an essential semantic instability of language, or the relativity of meaning to context, undermines even the assumption that identical meanings are exchanged in communication.

In reply to this, Habermas suggests that joking, playful, ironic and fictional uses of language depend on an intentional confusion of contrasting modes correlated to validity-claims (being/illusion, is/ought, essence/appearance), which is simultaneously seen through as such. These uses therefore depend on a prior recognition of the distinctions between validity dimensions. Habermas also emphasizes that belief in the identity of communicated meanings has no *metaphysical* status, but is merely a pragmatic presupposition which can (indeed must) be unproblematically relied on, up to the point where communication actually breaks down.

3 Apel's approach and his critique of Habermas

Habermas' account of communicative rationality has developed over the years in tandem with that of his Frankfurt colleague Karl-Otto Apel. However, Apel's programme is more strongly connected to the tradition of transcendental philosophy (Apel prefers the term 'transcendental pragmatics' to Habermas' 'universal pragmatics' for the general theory of communicative competence). The key conflict between Apel and Habermas concerns their attitude to the possibility of an 'ultimate grounding' (*Letztbegründung*) for a universalistic moral standpoint. In a critical essay (1992), Apel contends that Habermas is mistaken to play down the 'transcendental difference' between empirically testable reconstructions of communicative competence and philosophical reflection which reveals the unconditional validity of certain basic presuppositions of argumentation. Apel sees here a confusion which leads to paradoxical consequences - for example, that the 'principle of fallibilism' (the principle that no empirical consensus can be immune to revision) must itself be regarded as fallible. Habermas, however, replies that the normative preconditions of communication should be considered only as 'factically indispensable': they cannot be imagined as inoperative given our human socio-cultural form of life. To show this, however, we do not need to set philosophical truths in a position of qualitative precedence over knowledge obtained by scientific methods.

The concept of communicative rationality has thus been, and continues to be, the subject of intense debate. These reflect that fact that the project of defining a concept of rationality more encompassing that instrumental rationality, one internally connected to the idea of *reciprocal* and *egalitarian* relations between human beings, has deep roots in the German philosophical tradition since Kant. Because of this, whatever future transformations this project may undergo, it is unlikely to be entirely abandoned.

See also: Habermas, J.; Rationality: belief; Reasoning/rationality: practical

PETER DEWS

References and further reading

Adorno, T.W. and Horkheimer, M. (1944) *Dialectic of Enlightenment*, London: Verso, 1979.(Referred to in §1. Presents the classic Frankfurt School critique of instrumental reason.)

Apel, K.-O. (1973) *Transformation der Philosophie*, Frankfurt am Main: Suhrkamp Verlag, 2 vols; partial trans. G. Adey and D. Frisby, *Towards a Transformation of Philosophy*, London: Routledge & Kegan Paul, 1980. (Sets out Apel's theory of communicative rationality, based on the regulative idea of an unlimited communication community.)

Apel, K.-O. (1992) 'Normative Begründung der "Kritischen Theorie" durch Rekurs auf Lebensweltliche Sittlichkeit? Ein Transzendental-pragmatisch orientierter Versuch mit Habermas gegen Habermas zu denken' ('Normatively Grounding "Critical Theory" through Recourse to the Lifeworld? A Transcendental-Pragmatic

Attempt to Think with Habermas against Habermas'), in A. Honneth, T. McCarthy, C. Offe and A. Wellmar (eds) *Zwischenbetrachtungen. Im Prozess der Aufklärung (Philosophical Interventions in the Unfinished Project of Enlightenment)*, Frankfurt am Main: Suhrkamp Verlag; trans. B. Rehg, Cambridge MA: MIT Press, 125-72.(Referred to in §3. The fullest statement of Apel's critique of Habermas.)

Cooke, M. (1994) *Language and Reason: A Study of Habermas' Pragmatics*, Cambridge MA: MIT Press.(A detailed, sympathetic but critical analysis of Habermas' concept of communicative rationality.)

Habermas, J. (1968) *Erkenntnis und Interesse*, Frankfurt am Main: Suhrkamp Verlag; trans. J.J. Shapiro as *Knowledge and Human Interests*, London: Heinemann, 1971.(Referred to in §1. Develops the distinction between instrumental action and communicative action as two distinct forms of action.)

Habermas, J. (1976) 'Was heißt Universalpragmatik', in K.-O. Apel (ed.) *Sprachpragmatik und Philosophie*, Frankfurt am Main: Suhrkamp Verlag; trans. T. McCarthy as 'What is Universal Pragmatics?', in *Communication and the Evolution of Society*, London: Heinemann, 1979, 1- 68.(Habermas' most detailed statement of his programme for a universal pragmatics)

Habermas, J. (1981) *Theorie des Kommunikative Handelns. Band 1: Handlungsrationalität und gesellschaftliche Rationalisierung*, Frankfurt am Main: Suhrkamp Verlag; trans. T. McCarthy as *Reason and the Rationalization of Society*, vol. 1 of *The Theory of Communicative Action*, Boston MA: Beacon Press, 1984.(Contains important discussions of the concept of communicative rationality.)

Habermas, J. (1988) 'Handlungen, Sprechakte, sprachlich vermittelte Interaktionen und Lebenswelt', in *Nachmetaphysiches Denken (Postmetaphysical thinking)*, Frankfurt am Main: Suhrkamp Verlag. (Source of the second quotation in §2 above.)

Schnädelbach, H. (1991) 'Transformation der Kritischen Theorie' ('The Transformation of Critical Theory'), in A. Honneth and H. Joas (eds) *Kommunikatives Handeln (Communicative Action)*, trans. J. Gaines and D.L. Jones, Cambridge: Polity Press, 7-22.(Referred to in §2. Criticizes Habermas for suggesting that identifying grounds involves evaluating them, thereby making any description of communicative rationality normative.)

Wellmer, A. (1986) *Ethik und Dialog. Elemente des moralischen Urteils bei Kant und in der Diskursethik*, Frankfurt am Main: Suhrkamp Verlag; trans. D. Midgeley as 'Ethics and Dialogue: Elements of Moral Judgement in Kant and Discourse Ethics', ch. 4 of *The Persistence of Modernity*, Cambridge: Polity Press, 1991, 113-231.(Referred to in §2. An important critique of the idealizations built into Habermas' theory of communicative rationality.)

Communism

Communism is the belief that society should be organized without private property, all productive property being held communally, publicly or in common. A communist system is one based on a community of goods. It is generally presented as a positive alternative to competition, a system that is thought to divide people; communism is expected to draw people together and to create a community. In most cases the arguments for communism advocate replacing competition with cooperation either for its own sake or to promote a goal such as equality, or to free specific groups of people to serve a higher ideal such as the state or God.

The word *communism* appears to have first been used in the above sense in France in the 1840s to refer to the ideas of thinkers such as Françoise Émile Babeuf (1760-97) and Étienne Cabet (1788-1856), both of whom advocated the collectivization of all productive property. The concept is ancient, however. Early versions of a community of goods exist in myths that describe the earliest stages of human culture; it was a major issue in ancient Athens, a key component of monasticism and became the basis for much criticism of industrial capitalism.

The word communism later became associated with the teachings of Karl Marx (§12) and his followers and came to refer to an authoritarian political system combined with a centralized economic system run by the state. This form of communism has roots in the earlier idea because the ultimate goal of communism, as seen by Marx, was a society in which goods are distributed to people on the basis of need. The older usage continues to exist in a worldwide communal movement and as a standard by which to criticize both capitalism and Marxian communism.

The idea of communism as collectively owned property first appears in the Western tradition in classical Greece. Plato’s *Republic* contains a notable early defence (see Plato §14). Prior to the invention of the word, major communist theories can be found in some parts of the Christian Bible, in medieval monasticism and in Thomas More’s *Utopia* (1516). In all these cases the basis for collectively owned property is that members of society are freed from the need to devote their time to earning a living or caring for private property so that they can devote themselves to something more important such as the pursuit of knowledge, God or personal fulfilment. The assumption is that the need to provide for oneself or one’s family gets in the way of matters considered more important. For example, Plato advocates abolishing the family in his *Republic* because he fears that family ties will both distract the individual from higher things and tempt people to favour one group (family members) over others (non-family members). In monastic communism, which also abolishes the family, all property is owned by the community and each individual member of the community owns nothing, not even their clothes. Everything is provided by the community for each monk or nun; they are, thereby, freed from the burden of property to devote themselves to God. In More’s *Utopia* all houses and their furnishings are as near identical as possible and, since location cannot be identical, people move from house to house in regular rotation.

Later, secularized versions of communism stressed the equitable (not necessarily equal) distribution of, or at least access to, resources, but the underlying principle is quite similar. Rather than freeing some or even all people in a society to devote themselves to a higher cause, secular communism is designed to allow everyone to pursue personal fulfilment. It may best be characterized by a slogan adopted by Marx that appears to have been first published on the title page of Cabet’s *Voyage en Icarie* (1840): ‘From each according to his ability, to each according to his need’. In other words, each person contributes to society to the best of their ability in the areas of work for which they are suited; in return, society provides their basic needs. The underlying assumption is that all human beings deserve to have their needs met simply because they are human beings; differential ability and talent does not make one person more deserving than another. A specific case is the assumption by most communist theorists that there should be no difference in treatment of those who contribute to society through physical labour and those who contribute through mental labour. For example, in 1888 Edward Bellamy published *Looking Backward*, a utopian novel that became an instant bestseller and produced a worldwide movement. In it, Bellamy advocated an absolutely equal income for all members of society that could then be used by each individual to meet their own felt needs. An approach adopted by many intentional communities is for the group to make collective or social decisions (either by consensus or by majority rule) about economic matters that affect the community as a whole but to provide each individual member of the community with a discretionary income to use as they wish.

Marx maintained major aspects of this approach in the stage of human development that he called 'full' or 'pure' communism or just communism. The non-alienated people of this future communism will create a world in which income will be distributed on the basis of need (see Alienation §§3-4; Marx, K. §4); since everyone will be a productive labourer, there will no longer be any classes; and, because there will no longer be a need for political power to enforce class dominance, the state will gradually disappear to be replaced with decentralized, non-political administrative agencies. Since everyone will work, there will be high productivity and, therefore, plenty for all. Given the changed social situation, people will begin to think differently and social distinctions between occupations and between city and country will disappear. Thus, this form of communism is, in its essentials, identical to the earliest communist tradition. At least one major twentieth-century Marxist theorist, Ernst Bloch (1885-1977) in *Das Prinzip Hoffnung (The Principle of Hope)* (1959) argued that this utopian goal should be at the centre of Marxist theory (see Utopianism §4).

While Marxism and its version of communism predominantly took a different road, the more fundamental and historically earlier theory did not disappear. Non-Marxist forms of collective property have existed and been defended as part of the communal movement best represented by the Israeli kibbutzim and the US communal movement of the early nineteenth and mid-twentieth centuries, and by some forms of anarchism (see Anarchism §3). Communism as common property and as a vision of a better life for all is still a living tradition.

The arguments against communism take a number of different forms. The simplest rely on assumptions about human nature radically at variance with those of communism's proponents. The assumption is made, with little real evidence, that human beings are 'naturally' competitive and that therefore communism cannot work (see Human nature §1). A more developed analysis argues that communism is necessarily economically inefficient and will, therefore, be unable to provide as high a standard of living as a non-communist system. Some go so far as to argue that communism is impossible to sustain over long periods of time because its inefficiency is so great that the economic system must sooner or later collapse. Of course, economic efficiency is very low in the scale of values held by most supporters of traditional communism. Proponents of communism generally take the position that economic efficiency symbolizes the competitive system that they oppose and stands in the way of the cooperative community they hope to achieve.

Today, many believe that the time of communism has passed, largely because Marxist communism has been discredited in much of the world. But communism is not, and never has been, reducible to the Marxist version. Most fundamentally, it is the economic basis for dreams of complete human fulfilment, whether it is sought in monastic orders, intentional communities or whole societies. That dream persists, and with it the ideals of communism.

See also: Equality; Socialism

LYMAN TOWER SARGENT

References and further reading

Bellamy, E. (1888) *Looking Backward: 2000-1887*, ed. C. Tichi, Harmondsworth: Penguin, 1982.(A popular novel depicting an egalitarian society.)

Bernstein, E. (1930) *Cromwell and Communism*, London: Allen & Unwin.(A study by a prominent Marxist of communism in seventeenth-century Britain.)

Cabet, É. (1840) *Voyage et adventures de Lord Villiam Carisdell en Icarie*, Paris: Bureau au Populaire.(A utopian novel in which Cabet popularized communism.)

Cole, G.D.H. (1953-6) *A History of Socialist Thought*, London: Macmillan, 5 vols in 7. (General history of most variants of collective ownership.)

Dawson, D. (1992) *Cities of the Gods: Communist Utopias in Greek Thought*, New York: Oxford University Press.(Early developments of communism.)

Hillery, G.A., Jr. and Morrow, P.C. (1976) 'The Monastery as a Commune', *International Review of Modern Sociology* 6 (1): 139-54; reprinted as Hillery, G.A. in *Communes: Historical and Contemporary*, ed. R. Shonle Cavan and M. Singh Das, New Delhi: Vikas Publishing House, 1979, 152-69.(The best introduction to monastic communism.)

Kenyon, T. (1989) *Utopian Communism and Political Thought in Early Modern England*, London: Pinter.(A recent study of communism in the UK.)

Lugon, C. (1949) *La République communiste chrétienne des Guaranis (1610-1766)*, Paris: Les Éditions Ouvrières.

(A study of Christian communism under the Jesuits in Paraguay.)

More, T. (1516) *Utopia*, ed. G.M. Logan and R.M. Adams, Cambridge: Cambridge University Press, 1989.(The origin of the word utopia and an early statement of communism and religious toleration.)

Plato (*c*.380-367 BC) *Republic*, trans. D. Lee, Harmondsworth: Penguin, 2nd edn, 1974. (One of the earliest works usually called an utopia.)

Prudhommeaux, J. (1907) *Étienne Cabet et les Origines du Communisme Icarien*, Nimes: Imprimerie ‘La Laborieuse’.(A study of one of the earliest people to call hiimself a communist.)

Surtz, E.L., S.-J. (Society of Jesus) (1957) *The Praise of Pleasure: Philosophy, Education and Communism in More’s ‘Utopia’*, Cambridge, MA: Harvard University Press.(The best discussion of communism in the thought of Thomas More.)

Wiles, P.J.D. (1962) *The Political Economy of Communism*, Cambridge, MA: Harvard University Press.(Includes a good discussion of ‘full’ or ‘pure’ communism.)

Community and communitarianism

Reflections on the nature and significance of community have figured prominently in the history of Western ethics and political philosophy, both secular and religious. In ethics and political philosophy the term 'community' refers to a form of connection among individuals that is qualitatively stronger and deeper than a mere association. The concept of a community includes at least two elements: (1) individuals belonging to a community have ends that are in a robust sense common, *not merely congruent private ends, and that are conceived of and valued as common ends by the members of the group; and (2) for the individuals involved, their awareness of themselves as belonging to the group is a significant constituent of their identity, their sense of who they are.*

In the past two decades, an important and influential strand of secular ethical and political thought in the English-speaking countries has emerged under the banner of communitarianism. The term 'communitarianism' is applied to the views of a broad range of contemporary thinkers, including Alasdair MacIntyre, Charles Taylor, Michael Sandel, and sometimes Michael Walzer (MacIntyre 1981; Sandel 1982; Taylor 1979, 1989; Walzer 1983). It is important to note, however, that there is no common creed to which these thinkers all subscribe and that for the most part they avoid the term.

There are two closely related ways to characterize what communitarians have in common; one positive, the other negative. As a positive view, communitarianism is a perspective on ethics and political philosophy that emphasizes the psycho-social and ethical importance of belonging to communities, and which holds that the possibilities for justifying ethical judgments are determined by the fact that ethical reasoning must proceed within the context of a community's traditions and cultural understandings (Bell 1993: 24-45). As a negative view, communitarianism is a variety of anti-liberalism, one that criticizes liberal thought for failing to appreciate the importance of community.

At present the communitarian critique of liberalism is more developed than is communitarianism as a systematic ethical or political philosophy. Existing communitarian literature lacks anything comparable to Rawls' theory of justice or Feinberg's theory of the moral limits of criminal law, both of which are paradigmatic examples of systematic liberal ethical and political theory. For the most part, the positive content of the communitarians' views must be inferred from their criticisms of liberalism. Thus, to a large extent communitarianism so far is chiefly a way of thinking about ethics and political life that stands in fundamental opposition to liberalism. To some, communitarian thinking seems a healthy antidote to what they take to be excessive individualism and obsessive preoccupation with personal autonomy. To others, communitarianism represents a faliure to appreciate the value - and the fragility - of liberal social institutions. The success of communitarianism as an ethical theory depends upon whether an account of ethical reasoning can be developed that emphasizes the importance of social roles and cultural values in the justification of moral judgments without lapsing into an extreme ethical relativism that makes fundamental ethical criticism's of one's own community impossible. The success of communitarianism as a political theory depends upon whether it can be demonstrated that liberal political institutions cannot provide adequate conditions for the flourishing of community or secure appropriate support for persons' identities so far as their identities are determined by their membership in communities.

1 Community

If I am a member of a community, I conceive of the goals and values I share with my fellows as essentially *our* goals and values, not just as goals and values that happen to be the same for all of us because of a contingent convergence of our individual interests. Each member thinks of furthering the community's ends primarily as gains for *us*, not as a gain for themselves that happens to be accompanied by similar gains for other individuals in the group. In the activities that are the life of the community, individuals think of themselves first and foremost as members of the group. Thus, at least in the course of these activities, the distinction between what is in your interest, as opposed to mine or ours, breaks down or recedes into the background.

In contrast, in an association individuals typically conceive of their interests as distinct, even if their goals are collective; that is, as states of affairs that necessarily involve the group. For example, if they are an association rather than a community, when the employees in a firm work together to secure an important contract for the company (the collective goal), each will cooperate with others only to further their own interest, conceiving of it as

distinct from, although congruent with, that of the others.

A contrast can also be drawn between how the members of a community regard the process of pursuing the common goal and how that process is regarded in an association. In a community, the process itself (or rather the relationships among members of the group that constitute the process) is seen by members of the group to be valuable in itself, independently of whether it is effective in achieving the goal of the process. In an association, the relationships that exist in the process of pursuing the collective goal are viewed as being of only instrumental value for attaining private goods for each of the individuals involved.

Since a community is defined at least in large part by a set of shared values and goals, the second element in our analysis of community, shared identity, is at least partly normative. In other words, if in saying who I am I feel it important to identify myself as, for example, a Catholic, then this characterization of who I am is not a description of myself as being a member of a certain religious group; it also includes the awareness that I have certain commitments, an allegiance to certain values and the perception that I have certain obligations by virtue of being a member of this group.

These familiar ways of contrasting communities with mere associations have the virtue of throwing the distinction between the two into bold relief, but they oversimplify. Being a community is a matter of degree: groups fall along a continuum, from communities in the strict sense, in which the sense of a shared identity is fundamental to a person's most basic conception of themselves and in which allegiance to the common interest eclipses any distinct conception of private interest, to mere associations, whose members are connected by the most tenuous and transitory ties of private interests or tastes and whose identity derives chiefly from sources other than membership of the group (Feinberg 1988: 81-122). For example, in many groups, individuals will be motivated to cooperate with others both by a sense of community, according to which what counts is whether 'we' succeed, and by their own private interests in the outcome of the common endeavour, which they will be able to distinguish quite clearly from the interest of the group. Similarly, the extent to which the members will conceive of their membership as being an important component of who they are (what their values and commitments are, and so on) will vary across different groups. Thus the extent to which a group is a community will depend upon how dominant the sense of common, not merely congruent, goals is, and how large a role the sense of allegiance to the values and ends of the group plays in each individual's sense of who they are.

2 Communitarianism

To determine what, according to communitarians, the true significance of participation in community is, we must see the ways in which communitarians believe liberalism has failed to acknowledge that significance.

The first obstacle to reconstructing a positive communitarian view from the communitarian critiques of liberalism is the fact that communitarians have sometimes been unclear about what liberalism is, and in some cases apparently have made uncharitable assumptions about what liberalism is committed to (Feinberg 1988: 86-8; Buchanan 1988: 853-7). One useful place to begin is with a rather narrow but clear articulation of liberalism as a minimal political philosophy - a thesis about the proper role of the state. What I shall call the *liberal normative political thesis* is simply that the state is to place a priority on the enforcement of the basic individual and civil rights, those rights which, roughly speaking, are found in the US Constitution's Bill of Rights and in Rawls' First Principle of Justice, the Greatest Equal Liberty Principle (Rawls 1971: 61) (see Rawls, J. §1). These include the rights to freedom of religion, expression, thought and association, the right of political participation (including the right to vote and to run for office), and the rights of legal due process and equality before the law.

This fundamental thesis of liberal political philosophy implies another: namely, that the proper role of the state is to protect these basic rights, not to make its citizens virtuous or to impose upon them any particular or substantive conception of the good life or of the common good. By honouring this commitment to protect individual and civil political rights and forgoing any other projects that would violate them, the state recognizes a distinction between the private and the public spheres, and hence a limit on its own authority as well as a domain of autonomy for individuals (see Law, limits of §4).

In marking off a private sphere through the protection of individual rights liberalism not only demarcates a domain of autonomy for individuals, it also makes room for a variety of associations - and communities. For this reason it is misleading to say that liberalism protects individual freedom rather than community.

3 Communitarian criticisms of liberalism

The following four criticisms form the core of the communitarian attack on liberalism. Not all are advanced by every thinker usually labelled communitarian.

(1) Liberalism rests on an individualism that is both ontologically and motivationally false. Liberal ontology asserts that only individuals exist and that all putative properties of groups can be reduced to properties of individuals; liberal motivational theory assumes that individuals are motivated solely by preferences for private goods, and that they desire participation in groups only as a means to achieving such goods.
(2) Liberalism undervalues political life, viewing participation in the political community as a mere instrumental good, valuable only as a means toward the attainment of the various private ends that individuals have, rather than as something valuable in itself.
(3) Liberalism presupposes a defective conception of the self, not recognizing that the self is ‘embedded’ in and partly constituted by communal commitments that are not objects of choice. Thus liberalism fails to understand that what one’s obligations are is determined, to a large extent, by one’s identity as a member of the community and the roles one occupies within the community (see Obligation, political §2).
(4) The liberal emphasis on individual rights devalues, neglects, and ultimately undermines community; yet participation in community is a fundamental and irreplaceable ingredient in the good life for human beings.

Only the fourth communitarian criticism is a direct attack on the liberal normative political thesis. The other three are instances of a characteristic communitarian move, one which is most explicit perhaps in Sandel and Taylor: namely, that of attempting to undermine the liberal normative political thesis by exposing flaws in the system of beliefs upon which it supposedly rests. As we shall see, one problem with this strategy is the tendency on the part of those who employ it to fail to appreciate the range of different considerations that the liberal can invoke to support the liberal normative political thesis. At most, only some of these justifications may include beliefs which communitarianism shows to be false. Other justifications for liberalism - including what appear to be the most plausible - may remain intact.

The first criticism is easily rebutted. The liberal normative political thesis does rest on an individualistic premise, but the premise is neither the ontological individualism nor the motivational individualism which communitarians rightly have attacked. Liberalism is individualistic only in a *moral* sense. Liberals hold that what matters ultimately, morally speaking, is individuals - their autonomy and wellbeing.

Moral individualism does not commit the liberal to ontological individualism: holding that individuals are morally primary does not commit one to denying that groups exist or that they are reducible without remainder to the properties of individuals. Similarly, liberalism can readily admit that in order to describe the interest that individuals have in participating in groups it is necessary to make reference to institutions and practices. Liberals can also agree that the institutional or social concepts used to describe the interests that individuals have in belonging to groups and pursuing shared ends are not reducible to pre-institutional or pre-social concepts (such as that of the ‘atomistic’ individual existing in a state of nature). Liberalism, as a normative political thesis about the priority of the basic civil and political rights (and the proper role of the state and the limits of its authority), does not entail any such individualistic views; nor do the more plausible lines of justification invoked to support the liberal normative political thesis include such views among their premises. The latter justifications are simply attempts to show how a political order that protects the basic civil and political rights, and hence has a rather limited role for the state and a large private sphere in which diverse individuals (and groups) can flourish, will, in the long run, and in the majority of cases, best serve the autonomy and welfare of individuals.

The second communitarian criticism, that liberalism undervalues participation in the political community, is harder to assess. Those who voice it sometimes neglect to distinguish clearly between two quite different theses. The first is that participation in the political community is valuable in itself, as a significant ingredient in the good life for all or most human beings. Liberalism need not, and does not, deny this. The second is that the good life requires that one participate in what may be called an all-inclusive political community or in some rather direct and substantial way in the highest-level political organization in one’s society.

The first thesis is more controversial than the bare claim that participation in some form of community or other is

an essential ingredient in the good life. It is the claim that participation in *political* community is essential. Although clearly in need of systematic defence, the latter claim is not implausible. It can be supported by arguing, as Aristotle seems to have done, that the good life for human beings, or at least the best life, requires the exercise and development of certain powers of will and intellect that can only be achieved through participation in activities in which one rules and is ruled in turn (see Aristotle).

Notice, however, that even if this first thesis is granted, we are still a great distance from the second. For one may be able to achieve the good of participation in political community without participating in any significant way in an overarching political community, or the highest level of political organization. All that is necessary is that one participate in one or more political subcommunities, which might include not only local governments but also many other types of communities, religious and secular, in so far as they include a substantial political dimension in their internal activities.

Given the apparent diversity of the conditions of human flourishing, the sweeping pronouncement that the good life or even the best life for all (or even most) human beings requires significant participation in the most inclusive form or highest level of political organization appears rather dogmatic. To my knowledge no communitarian has shouldered the burden of supporting this second thesis.

The third communitarian criticism is sometimes formulated as the slogan that liberalism - at least the liberalism of theorists such as John Rawls and Ronald Dworkin - conceives of ‘the self as being given prior to its ends’ (Sandel 1984a: 115). Put more positively, this is the thesis that an individual’s identity is to a significant degree constituted by an allegiance to the ends to which they are committed by virtue of their membership of the community, and more specifically by the roles they occupy in it.

Liberals need not and should not deny that when human beings become aware of themselves as moral agents and when they confront questions about what they ought to do, they already recognize themselves as having an identity. Nor should they fail to acknowledge that this identity is constituted to a large extent by the various roles in which they find themselves in the practices of their community (or communities), and that these roles carry with them certain obligations and commitments. Liberalism need not deny this commonsensical truth, which we may call ‘the fact of embeddedness’. There is, however, a more radical interpretation of the thesis that which liberal thinkers do not espouse - but one which appears to be quite implausible. This is the view that the ends that are ‘given’ as part of a person’s socially generated identity are *incorrigible* for that individual, that they cannot meaningfully be objects of criticism by the individual, and hence that they can neither be chosen or rejected by them.

It would be uncharitable to say that communitarians unequivocally hold this more radical interpretation of the embeddedness thesis, but there are passages in their writings that suggest that they do:

> I can only answer the question ‘What am I to do?’ if I can answer the prior question ‘Of what story or stories do I find myself a part’? We enter human society, that is, with one or more imputed characters - roles into which we have been drafted.… I am someone’s son or daughter, someone else’s cousin or uncle; I am a citizen of this or that city, a member of this or that guild or profession; I belong to this or that clan, that tribe, this nation. Hence what is good for me has to be the good for one who inhabits these roles. As such I inherit from the past of my family, my city, my tribe, my nation, a variety of debts, inheritances, rightful expectations, and obligations.
>
> (MacIntyre 1981: 203-5)

If the point is that the self is embedded in the sense that the obligations a person has are given, *irrevocably so*, by their membership of and particular role within a community, that the obligations that are imputed along with membership and roles are not subject to criticism, and that a person cannot rationally choose to reject such roles, then liberalism does deny the embeddedness of the self. For liberalism, to the extent that it emphasizes individual autonomy, stresses the ability of individuals, at least under favourable conditions in which social practices and institutions allow the free flow of information and untrammelled discussion, to exercise critical judgment about the suitability of their roles and the validity of the values of their communities, and hence to question the obligations that are imputed to them.

To put the same point in terms of what may be called the normative significance of identity, communitarians

sometimes appear to hold that an individual cannot critically assess those putative obligations that are constitutive of their identity. On the face of it, however, this seems quite wrong. The fact that I am a racist - that racism is partly constitutive of my identity - does not show that I am obliged to treat white people as if they were superior to black people, nor that it is impossible for me to criticize my racist beliefs and reconstitute my identity as that of one who is dedicated to racial equality. It makes perfectly good sense to say that I ought to disavow my racist identity and strive to construct a new, nonracist identity.

The communitarian might reply that even when criticism of culturally-imputed obligations and values does occur, it must always proceed by drawing on conceptual, factual and moral resources that are themselves available only through the culture in question. This may in fact be what MacIntyre has in mind when, immediately following the passage cited above, he states that the obligations one inherits from one's social roles are one's moral starting point. The implication is that one may not remain at this starting point. One may come to revise one's conception of what is valuable, what one's obligations are, and so on - but only by utilizing tools of criticism that are also given by one's community.

On this interpretation, the embeddedness thesis is the claim that if an individual is able to exercise critical judgment about whether to act on the values and perceived obligations that are imputed to them by their community, this is only possible by drawing on perspectives that are available within the very communal normative structure that imputes the values in the first place.

If this means that all factual assumptions, concepts and moral principles that an individual relies on to criticize the values of his own community must be wholly endogenous to that community, must originate there or must be widely accepted there, then it not only appears to be false, but would seem to rest upon a view of communities which, ironically, is 'atomistic individualism' writ large. On this view, communities and the systems of values they embody are like the self-contained, *sui generis* and impermeable 'bare individuals' which communitarians believe to people the illusory world of liberal theory. But surely it cannot be denied that communities interact and that they borrow from one another's cultures, any more than it can be denied that the values of individuals are shaped by the social forces within which they develop.

In fact, modern societies are characterized by not only a plurality of interpenetrating communal cultures within the boundaries of a given society, but also by increasing interaction among different societies as the global economy and the global culture it brings with it expand. For example, Western principles concerning human rights in general and the equality of women in particular have been used, with some success, to criticize the assigned roles and perceived values and obligations of non-Western societies. Once it is recognized that criticism of acculturated values sometimes relies at least in part on beliefs, concepts or principles that originate outside the culture, then even the more moderate version of the embeddedness thesis is seen to be implausible.

Nevertheless an important truth remains: even if one is able to criticize obligations that are constitutive of one's culturally-assigned identity, all reasoning about what one ought to do must begin with a recognition of who one is and what one's obligations are as determined by one's identity. If this is what is meant by embeddedness, then communitarians have articulated an important truth about the nature of ethical reasoning and critical ethical reflection - one which liberal thinkers have perhaps generally neglected. What is not clear, however, is how taking this fact of embeddedness seriously affects the conclusions about the state and individual rights which liberalism seeks to defend.

The fourth and final communitarian criticism of liberalism is potentially more telling. The claim is that liberalism exalts individual autonomy to the neglect and detriment of community. The communitarians' point, presumably, is both that: (1) the justifications liberals offer to support the normative political thesis concerning the priority of basic individual, civil and political rights rest on an excessively positive valuation of individual choice or autonomy; and that (2) implementing the liberal priority on these individual rights is damaging to communities.

Contemporary liberals have offered strong arguments against both elements of the communitarian attack. First, liberal theorists can and do explicitly recognize the importance of community, and to that extent do not assume that individual autonomy is the sole value to be protected. Some of the most prominent contemporary liberal theorists, including Rawls, Dworkin and Feinberg, are most naturally read as ascribing to a moral individualism that bases the case for individual rights ultimately on both the autonomy and the wellbeing of individuals

(Buchanan 1988: 878-81). If this is the case, then all that is needed to incorporate a fundamental appreciation for community into the heart of the liberal justification for individual rights is the commonsensical thesis: for all or most human beings, participation in communities is an important ingredient in their wellbeing, as something that is intrinsically valuable. Once again, there is nothing in liberal theory, or at least in its more plausible variants, that precludes a recognition of the truth of this thesis.

Some contemporary liberal thinkers have argued that participation in community is also valuable because the community's shared values provide a context for meaningful choice (Kymlicka 1989: 205-19). To the extent that this is so, the opportunity to participate in flourishing communities is valuable not only from the perspective of individual wellbeing but also from that of individual autonomy, since autonomy can flourish only where meaningful choices are possible.

Liberalism, therefore, can include a proper recognition of the importance of participation in communities for individuals and can avoid an exclusive emphasis on individual autonomy that would neglect the importance of community for individual wellbeing. In addition, a number of liberal thinkers have pointed out that liberalism's priority on individual civil and political rights not only provides scope for individual autonomy but also serves to protect communities and to allow them to flourish (Feinberg 1988: 108-12; Buchanan 1994: 11-15).

To understand how individual rights protect communities, consider the rights to freedom of association, expression and religion. Historically these rights, as well as political participation rights and legal due process rights, have provided a strong bulwark against attempts by the state or rival communities to destroy or dominate minority communities. They allow individuals to partake of the essential good of community, and to enjoy the context for meaningful choice that community values provide, both by protecting existing communities from interference and by giving individuals the freedom to unite with like-minded others to form new communities. It is wrong to assume, then, that because the rights which liberals champion are individual rights, the purposes those rights serve and the values upon which they rest are purely individualistic.

This is not to say, of course, that any existing liberal theory provides a fully satisfactory way of accommodating both the importance of individual autonomy and the importance of participation in community. Nor is it to suggest that the particular understandings of the scope and limits of individual rights that liberals typically advocate successfully resolve all conflicts between the interests of individuals *qua* individuals and the interests of communities. Perhaps the most obvious instances of such unresolved conflicts - and of the challenge to liberalism that they represent - lie in the area of tensions between the interest in preserving a community's longstanding culture (or certain central features of it) and the liberal commitment to ensuring that children are educated so as to be able to make their own choices as to how to live. For example, some religious communities in the USA and elsewhere have sought an exemption from public schooling requirements on the grounds that exposure to ideas and values dominant in the majority culture would weaken their own community's traditions. Here there seems to be a direct conflict between liberalism's emphasis on individual autonomy and freedom of expression, and the value of preserving a community. And to say, as I have done, that liberalism ought to and can recognize the importance of participation in community for the wellbeing and autonomy of individuals does not in itself shed any light on how the balance should be struck between the individual rights liberalism has traditionally championed and the need to protect cultures when the former and the latter are in conflict.

If striking an adequate balance requires rejecting what I have taken to be the core of liberalism - the normative political thesis of the priority of individual rights - then the fourth major communitarian criticism of liberalism will be telling: the liberal priority on individual rights will be seen to rest on an inadequate appreciation of the value of community. If this is the case, then perhaps the most fruitful line of development for communitarian thought would be a concerted effort to show how political systems that only recognize individual rights as basic fail to respond satisfactorily to the need to preserve communities.

One important element of this line of inquiry is the question of whether states should recognize *group* rights that are justified according to the protection they provide for communities or their cultural values and that are basic rights in the sense of not being generated through the exercise of the individual rights which liberalism recognizes, such as the right to freedom of association and the legal right to make contracts. (Examples of basic group rights - those not generated through the exercise of individual rights - might include language rights for national minorities and collective land rights for indigenous peoples). If the liberal thesis of the primacy of individual rights rules out

recognition of any basic group rights, then a successful case for basic group rights (as needed to provide sufficient protection for communities) would be a conclusive communitarian objection to liberalism thus interpreted. As a way of developing systematically communitarianism both as a positive view and as a critique of liberalism, this line of research may prove crucial (see Multiculturalism).

4 Conclusions

If communitarianism is understood as being, or as centrally including, a radical critique of liberalism, where the latter is construed narrowly as the normative political thesis that the proper role of the state is to uphold the basic individual civil and political rights, then the communitarian project has thus far proven inconclusive. For liberalism can, and a growing number of contemporary liberal thinkers do, take the value of participation in community seriously - thus acknowledging the central tenets of communitarianism while denying that liberalism cannot incorporate them. However, if liberalism is understood to exclude entirely any role for basic group rights in the protection of cultural communities, then the communitarian critique of liberalism may well turn out to be telling.

Alternately, communitarianism might be viewed more positively, not primarily as a criticism of the liberal normative political thesis but as a rationale for an expanded and enriched research agenda in ethics and political philosophy. Taking the importance of community seriously in political philosophy would mean acknowledging and attempting to resolve systematically the conflicts that can arise in any theory that recognizes the importance of both individual autonomy and participation in communities. And doing so may require rethinking the scope and limits of theories of individual rights and a willingness to entertain the possibility that these rights must be supplemented with special group rights designed to provide added protection to especially vulnerable communities (Buchanan 1994: 11-15).

In ethical theory the communitarian message is to take seriously the fact that to a large extent moral obligations are the product of explicit choices or voluntary agreements made not by autonomous individuals operating in a social and institutional vacuum, but by individuals who regard themselves and are regarded by others as members of communities, as occupants of particular social roles. In addition, communitarianism stresses a fundamental truth about how ethical reflection and argumentation must begin; namely, in the concrete context of social practices in which individuals find themselves and in an awareness that the sense of self and the recognition of moral values are inextricably bound together. Thus, another important item on the agenda for a communitarian ethical theory is a careful articulation of the extent to which the social starting place that supplies one's identity places unsurpassable limitations on the sorts of moral conclusions one can draw and the kinds of moral commitments one can come to make through a process of rational reflection.

See also: Confucian Philosophy, Chinese §5; International relations, philosophy of §3; Nation and nationalism; Rawls, J.

ALLEN BUCHANAN

References and further reading

Avineri, S. and de-Shalit, A. (1992) (eds) *Communitarianism and Individualism*, Oxford: Oxford University Press.(A useful collection of articles from both sides of the debate between communitarians and liberals.)

Bell, D. (1993) *Communitarianism and Its Critics*, Oxford: Clarendon Press.(An imaginative defence, in dialogue form, of communitarian approaches to social and political philosophy.)

Buchanan, A. (1988) 'Assessing the Communitarian Critique of Liberalism', *Ethics* 99 (4): 852-82.(A critical reconstruction and evaluation of communitarian criticisms of liberalism.)

Buchanan, A. (1994) 'Liberalism and Group Rights', in J. Coleman and A. Buchanan (eds) *In Harm's Way*, Cambridge: Cambridge University Press.(A defence of the view that liberalism can recognize group rights.)

Feinberg, J. (1988) *Harmless Wrongdoing*, vol. 4, *The Moral Limits of the Criminal Law*, New York: Oxford University Press.(A theory of the moral principles legislators ought to take into account in formulating criminal law concerning harmless wrongdoings.)

Kymlicka, W. (1989) *Liberalism, Community, and Culture*, Oxford: Oxford University Press.(An account of how liberalism can and ought to take into account the value of culture to individuals.)

MacIntyre, A. (1981) *After Virtue: A Study in Moral Theory*, London: Duckworth.(A critique of post-Enlightenment Western religious thought, from a communitarian perspective.)

Mulhall, S. and Swift, A. (1992) *Liberals and Communitarians*, Oxford: Oxford University Press.(An introduction to the thought of Sandel, MacIntyre, Taylor and Walzer, and the liberal responders, Rawls, Rorty and Raz.)

Plant, R. (1974) *Community and Ideology*, London: Routledge & Kegan Paul.(Examines the concept of community and different conceptions of community in political philosophy.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(The most systematic account available of justice at the level of the basic structure of social institutions.)

Sandel, M. (1982) *Liberalism and the Limits of Justice*, Cambridge: Cambridge University Press.(A critique of liberal theories of justice, focusing on Rawls; taken from the standpoint of a communitarian view of obligation and ethical reasoning.)

Sandel, M. (1984a) 'Justice and the Good', in *Liberalism and Its Critics*, New York: New York University Press.(Argues that the attempt of liberal theorists to develop an account of justice that is prior to and independent of the good must fail.)

Sandel, M. (1984b) 'The Political Theory of the Procedural Republic' in *Liberalism and Its Critics*, New York: New York University Press.(Argues that political decision making requires agreement on a substantive conception of the good.)

Taylor, C. (1979) *Hegel and Modern Society*, Cambridge: Cambridge University Press.(Reflections on Hegelian themes.)

Taylor, C. (1985) 'Atomism', in C. Taylor (ed.) *Philosophy and the Human Sciences*, Cambridge: Cambridge University Press.(Critique of atomistic conceptions of society and of rights-based political philosophy.)

Taylor, C. (1989) 'Cross-purposes: The Liberal-Communitarian Debate', in N. Rosenblum (ed.) *Liberalism and the Moral Life*, Cambridge, MA: Harvard University Press.(Argues that the usual division between liberals and communitarians is confused, and that the real dividing line runs between procedural liberals and their republican critics.)

Walzer, M. (1983) *Spheres of Justice*, New York: Basic Books.(Claims that ideas of justice depend on the meanings attached to social goods by the members of a particular community.)

Complexity, computational

The theory of computational complexity is concerned with estimating the resources a computer needs to solve a given problem. The basic resources are time (number of steps executed) and space (amount of memory used). There are problems in logic, algebra and combinatorial games that are solvable in principle by a computer, but computationally intractable because the resources required by relatively small instances are practically infeasible.

The theory of $\mathcal{NP}$-completeness concerns a common type of problem in which a solution is easy to check but may be hard to find. Such problems belong to the class $\mathcal{NP}$; the hardest ones of this type are the $\mathcal{NP}$-complete problems. The problem of determining whether a formula of propositional logic is satisfiable or not is $\mathcal{NP}$-complete. The class of problems with feasible solutions is commonly identified with the class $\mathcal{P}$ of problems solvable in polynomial time. Assuming this identification, the conjecture that some $\mathcal{NP}$ problems require infeasibly long times for their solution is equivalent to the conjecture that $\mathcal{P} \neq \mathcal{NP}$. Although the conjecture remains open, it is widely believed that $\mathcal{NP}$-complete problems are computationally intractable.

1 Time and space in computation

The theory of complexity analyses the computational resources necessary to solve a problem. The most important of these resources are *time* (number of steps in a computation) and *space* (memory capacity of the computer). This entry is mainly concerned with the complexity of decision problems having infinitely many instances. There is another approach to complexity applicable to individual finite objects, in which the complexity of an object is measured by the size of the shortest programme that produces a description of it (see Computability and information).

The model for computation chosen here is the Turing machine (see Turing machines). The *time* for a computation is the number of steps taken before the machine halts; the *space* is the number of cells of the tape visited by the reading head during the computation. Several other models of sequential computation have been proposed. The time and space complexity of a problem clearly depend on the machine model adopted. However, the basic concepts of complexity theory defined in this entry are robust in the sense that they coincide for any reasonable model of sequential computation.

Let Σ be a finite alphabet, and Σ^* the set of all finite strings in this alphabet. A subset of Σ^* is said to be a ‘problem’. The size $|s|$ of a string *s* is its length, that is, the number of occurrences of symbols in it. A function *f* defined on Σ^* and having strings as its values is computed by a Turing machine *M* if for any string *s* in Σ^*, if *M* is started with *s* on its tape, then it halts with $f(s)$ on its tape. A problem *L* is solvable if there is a Turing machine that computes the characteristic function of *L* (the function *f* such that $f(s) = 1$ if *s* is in *L* and $f(s) = 0$ otherwise). For example, the problem Satisfiability of determining whether a formula of propositional logic is satisfiable or not is solvable by the familiar method of truth tables.

Solvable problems can be classified according to the time and space required for their solution. If *f* is a computable function, then we say that *f* is computable in time $T(n)$ if there is a Turing machine *M* computing *f* and a constant *c* such that for any input *s*, *M* halts with output $f(s)$ in at most $c \cdot T(|s|)$ steps. A function *f* is computable in space $S(n)$ if there is a machine *M* computing *f* and a constant *c* such that for any input *s*, *M* halts after visiting at most $c \cdot S(|s|)$ squares on its tape. A problem *L* is solvable in time $T(n)$ if the characteristic function of *L* is computable in time $T(n)$; *L* is solvable in space $S(n)$ if the characteristic function of *L* is computable in space $S(n)$. For example, the truth table method shows that Satisfiability can be solved in time 2^n and space *n* (we need only enough tape space to evaluate the truth table one row at a time).

One of the most significant complexity classes is the class $\mathcal{P}$ of problems solvable in polynomial time. A function is polynomial-time computable if there exists a *k* for which *f* is computable in time n^k. A problem is solvable in polynomial time if its characteristic function is polynomial-time computable. The importance of this class rests on the widely accepted working hypothesis that the class of practically feasible algorithms can be identified with those algorithms that operate in polynomial time. Similarly, the class $\mathcal{PSPACE}$ contains those problems solvable in polynomial space. The class $\mathcal{EXPTIME}$ consists of the problems solvable in exponential time; a problem is

solvable in exponential time if there is a *k* for which it is solvable in time 2^{n^k}. The class $\mathcal{EXPSPACE}$ contains those problems solvable in exponential space. The problem Satisfiability is in $\mathcal{EXPTIME}$; whether it is in $\mathcal{P}$ is a major open problem.

2 Hierarchies and reductions

The complexity classes defined above represent only a few levels within detailed hierarchies defined by space and time bounds. To state results on the hierarchies, we define a class of functions that includes all the complexity bounds considered here. A function $S(n)$ is said to be fully space constructible if there is a Turing machine that uses exactly $S(n)$ tape cells for any input of length *n*. With this definition we can state the space hierarchy theorem (Hartmanis, Lewis and Stearns 1965):

If $S_1(n)$ and $S_2(n)$ are fully space constructible functions and if

$\liminf_{n\to\infty} \frac{S_1(n)}{S_2(n)} = 0,$

then there exists a problem solvable in space $S_2(n)$ but not in space $S_1(n)$.

In informal terms, the theorem says that if $S_1(n)$ and $S_2(n)$ are 'reasonable' mathematical functions and asymptotically $S_2(n)$ grows faster than $S_1(n)$, then there is a problem solvable in space $S_2(n)$ but not in $S_1(n)$. For example, there is a problem solvable in space n^3, but not in space n^2. A similar but slightly weaker time hierarchy theorem can be proved for complexity classes defined by time bounds.

The problems in these theorems that separate the complexity classes are defined by diagonalization. Applications of the theorems to naturally occurring problems in logic and mathematics are obtained by using the tool of efficient reductions. These are the complexity-theoretic analogue of the notion of Turing reducibility in recursion theory.

Let L_1 and L_2 be problems expressed in alphabets Σ_1 and Σ_2, respectively. L_1 is said to be polynomial-time reducible to L_2 (briefly, reducible to L_2) if there is a polynomial-time computable function *f* from Σ_1^* to Σ_2^* such that for any *s* in Σ_1^*, $s \in L_1$ if and only if $f(s) \in L_2$. Many other notions of reducibility have been considered. The important property of polynomial-time reducibility is that it is compatible with the main complexity classes $\mathcal{C}$ considered in this entry; this means that if L_2 is in a complexity class $\mathcal{C}$ and L_1 is reducible to L_2, then L_1 is also in $\mathcal{C}$.

3 Provably intractable problems

If $\mathcal{C}$ is a complexity class and *L* is a problem in $\mathcal{C}$ such that any problem in $\mathcal{C}$ is reducible to *L*, then *L* is said to be $\mathcal{C}$-complete. Such problems are the hardest problems in $\mathcal{C}$; if any problem in $\mathcal{C}$ is computationally intractable, then a $\mathcal{C}$-complete problem is intractable.

The time hierarchy theorem implies that there are problems in $\mathcal{EXPTIME}$ requiring exponential time for their solution, no matter what algorithm is employed. It follows that the same is true for any problem that is $\mathcal{EXPTIME}$-complete. Such problems can be shown to exist in the area of combinatorial games. For example, we can define generalized chess to be a game similar to standard chess, but played on an $n \times n$ board. The problem of determining of an arbitrary position in generalized chess whether or not the first player has a winning move is $\mathcal{EXPTIME}$-complete (Fraenkel and Lichtenstein 1981). If *M* is a Turing machine that solves this problem, there is a constant $c > 1$ such that infinitely many $n \times n$ chess positions exist for which *M* requires time greater than c^n to determine whether the first player has a win or not. In this sense, chess can be said to be a provably difficult game.

Problems that are $\mathcal{EXPSPACE}$-complete are also inherently intractable. In algebra, the word problem for commutative semigroups has been proved $\mathcal{EXPSPACE}$-complete (Mayr and Meyer 1982). An instance of this problem is a finite sequence of equations of the form $s = t$ where the terms *s* and *t* are constructed from a set of constants using a single binary operation; the problem to be solved is whether or not the last equation in the sequence can be inferred from the earlier equations by the usual rules for equations, assuming the commutative and associative laws for the operation. It follows that any Turing machine solving this problem requires an exponential amount of space for infinitely many inputs.

Logical theories are a source of solvable decision problems for which no feasible algorithm exists. An example is first-order arithmetic, restricted to addition. The language of this theory is a first-order language based on identity, the addition symbol and the constants 0 and 1. The decision problem consists in determining whether or not a sentence in this language is true in the natural numbers. Presburger in 1930 showed that this problem is decidable by the method of eliminating quantifiers. In 1973 Rabin and Fischer showed that the complexity of the problem is doubly exponential. This means that for any machine solving this problem, there is a constant $c > 0$ such that for infinitely many sentences of length *n* the machine takes at least $2^{2^{cn}}$ steps to determine whether it is true or not.

Still more powerful lower bounds can be proved by allowing second-order quantifiers. The weakest such extension allows quantifiers ranging over finite sets. Let WS1S be the weak monadic second-order theory of one successor. This theory is formulated in a second-order language with equality, based on the constant 0 and the successor function. The decision problem consists of deciding whether a sentence of WS1S is true in the natural numbers when the second-order quantifiers range over finite sets of natural numbers. It was shown decidable by Büchi in 1960. Define a function $F(n, m)$ by the recursion

$$F(n, 1) = 2^n$$
$$F(n, m+1) = 2^{F(n,m)} .$$

If *d* is a real number greater than 0, then $f(d, n) = F(n, [dn])$ represents exponentiation by a stack of $[dn]$ 2's, where $[x]$ is the integer part of *x*. Albert Meyer in 1972 showed that an algorithm deciding WS1S must use space $f(d, n)$ infinitely often for some $d > 0$. This result shows that the decidability of WS1S is of a purely theoretical nature, so that the theory may be considered undecidable from a practical point of view.

The conclusion of the previous paragraph could be challenged by pointing out that Meyer's lower bound is an asymptotic result that does not rule out a practical decision procedure for sentences of practically feasible size. However, a further result shows that astronomical lower bounds can be proved for WS1S even if we restrict the length of sentences. A Boolean network or circuit is an acyclic directed graph in which each node is labelled with a variable of one of the logical operators 'and', 'not'. The nodes labelled with a variable have no incoming edges, while those labelled 'and' have two, and those labelled 'not' have one. There is exactly one node (the output node) with no outgoing edge. A circuit computes a Boolean function (truth-function) of the variables labelling its input nodes in an obvious way. Meyer and Stockmeyer showed that any such network that decides the truth of all sentences of WS1S of length 616 or less (encoded as binary strings) must contain at least 10^{123} nodes. Even if the nodes were the size of a proton and connected by infinitely thin wires, the network would densely fill the known universe.

4 $\mathcal{NP}$ and the polynomial-time hierarchy

A very common type of practical computational problem takes the form of searching for a solution to a set of conditions, where a proposed solution can be checked easily for correctness. Problems of this kind appear naturally in fields such as operations research, automated theorem proving, computer algebra, linear programming, graph theory and many other areas of discrete mathematics. The features common to such problems have been formalized in the theory of $\mathcal{NP}$-completeness.

A *k*-place relation holding between finite strings of symbols in a finite alphabet Σ is said to be polynomial-time computable (see §1 above) if its characteristic function is polynomial-time computable. (The characteristic function of a relation *R* is the function *f* such that $f(\langle s_1, \ldots, s_k \rangle) = 1$ if $R(s_1, \ldots, s_k)$, and $f(\langle s_1, \ldots, s_k \rangle) = 0$ otherwise.) Let *L* be a problem expressed in an alphabet Σ. *L* belongs to the class $\mathcal{NP}$ if there is a polynomial-time computable relation *R* and a polynomial $p(x)$ such that for any *s* in Σ^*, *s* is in *L* if and only if there exists a *y* such that $|y| \leq p(|s|)$ and $R(s, y)$.

The abbreviation '$\mathcal{NP}$' stands for 'non-deterministic polynomial-time', and is derived from an alternative definition of the class using non-deterministic Turing machines. The instructions for such a machine give for each pair consisting of a scanned symbol and the current state a finite set of permissible actions. A language is in $\mathcal{NP}$ if and only if it is accepted by a non-deterministic Turing machine the length of whose computations is bounded by a polynomial in the size of the input. (A string is accepted by such a machine if there is a computation consisting of permissible actions at each step that ends in an accepting configuration.)

Informally, we could describe the relation between $\mathcal{P}$ and $\mathcal{NP}$ as follows. The class $\mathcal{P}$ consists of problems with a feasible decision method; the class $\mathcal{NP}$ consists of problems where we can guess a solution, and check in a feasible amount of time if our guess was correct.

The problem Satisfiability belongs to $\mathcal{NP}$ since we can take the relation $R(s, y)$ to be the relation '*y* is a truth-value assignment that satisfies the formula *s*'. Satisfiability occupies a central position in the theory of $\mathcal{NP}$-completeness as perhaps the simplest and most natural example of an $\mathcal{NP}$-complete problem. Cook's theorem of 1971 showing that Satisfiability is $\mathcal{NP}$-complete led to the identification of hundreds of $\mathcal{NP}$-complete problems in computer science and discrete mathematics. The book by Garey and Johnson (1979) includes a partial list of such problems.

If a problem is $\mathcal{NP}$-complete, then a polynomial-time algorithm for the problem would lead to a polynomial-time algorithm for any problem in $\mathcal{NP}$ so that the equation $\mathcal{P} = \mathcal{NP}$ would be true. A polynomial-time algorithm for the satisfiability problem for propositional logic would lead to polynomial-time algorithms for hundreds of important problems in discrete mathematics; this follows immediately from the definition of $\mathcal{NP}$-completeness. Because such algorithms have been sought in vain for several decades, it is commonly conjectured that none exists, although no proof of this conjecture has been found. The problem '$\mathcal{P} = \mathcal{NP}$?' is widely considered the most important open problem in theoretical computer science.

The recursive sets can be identified with the computable sets; if we accept the identification of 'feasibly computable' with 'polynomial-time computable' then $\mathcal{P}$ is the class of problems with feasible solutions. The recursively enumerable sets are obtained from recursive relations by unbounded quantification; similarly, the problems in $\mathcal{NP}$ are obtained from relations in $\mathcal{P}$ by polynomially bounded quantification. Thus the class $\mathcal{NP}$ is the complexity-theoretic analogue of the class of recursively enumerable sets, and the $\mathcal{NP}$-complete problems correspond to problems of the highest degree of unsolvability among recursively enumerable sets, such as the halting problem (see Computability theory §5; Turing machines §3).

The analogy with recursion theory has been extended by defining a complexity-theoretic analogue of the Kleene arithmetical hierarchy of recursion theory. The class Σ_k^p is defined as the class of problems expressible in the form

$$\{s\colon (\exists y_1, |y_1| \leq p(|s|))(\forall y_2, |y_2| \leq p(|s|)) \\ \ldots R(s, y_1, \ldots, y_k)\},$$

where there are *k* alternating bounded quantifiers in the prefix of the defining formula, $p(x)$ is a polynomial and *R* is a polynomial-time computable relation. In particular, $\mathcal{NP} = \Sigma_1^p$. The class Π_k^p is given by a definition dual to that for Σ_k^p, where bounded existential quantifiers are replaced by bounded universal quantifiers, and universal by existential.

It is conjectured that all the classes in the polynomial hierarchy are distinct, but no proof is known. If $\mathcal{P} = \mathcal{NP}$, then all the classes in the hierarchy collapse to $\mathcal{P}$. All of the classes in the polynomial hierarchy are contained in $\mathcal{PSPACE}$. The equality $\mathcal{P} = \mathcal{PSPACE}$ is consistent with current knowledge, even though $\mathcal{PSPACE}$ appears to be a much more extensive class than $\mathcal{P}$.

The space and hierarchy theorems above were proved by diagonal arguments similar to those used to prove Church's theorem (see Church's theorem and the decision problem). There are reasons to think that diagonalization cannot be used to separate $\mathcal{P}$ and $\mathcal{NP}$ or other low-level complexity classes (the analogy with recursion theory breaks down at this point). These reasons are based on results about oracle machines. If *A* is a set of strings then a Turing machine with oracle *A* is defined to be a Turing machine that in addition to the normal type of instruction also has instructions of the form (q_i, σ, q_j, q_k), where q_i, q_j, q_k are machine states and σ is a symbol from the tape alphabet (see Turing machines §3). The instruction is interpreted by the machine as follows: when the current state is q_i and the string of symbols to the right of the reading head (starting with the current symbol σ) is in *A*, then the machine's next state is q_j, otherwise its next state is q_k. Thus a Turing machine with oracle *A* has the ability to answer in a single step questions of the type 'Is this string in the set *A*?'. Time and space of a computation by an oracle machine is computed just as for an ordinary Turing machine, counting the time taken for the answer to the oracle query as one step (the oracle answers any query instantaneously).

If *A* is any set of strings, then by imitating the definitions of the complexity classes above, but substituting 'Turing

machine with oracle *A*' everywhere for 'Turing machine' we can define relativized complexity classes $\mathcal{P}^A$, $\mathcal{NP}^A$ and so on. Then we have the following theorem (Baker, Gill and Solovay 1975):

There is a recursive oracle *A* for which $\mathcal{P}^A = \mathcal{NP}^A$ and a recursive oracle *B* for which $\mathcal{P}^B \neq \mathcal{NP}^B$.

The significance of this theorem lies in the fact that known techniques of diagonalization, such as are used in recursion theory, relativize in the presence of oracles. Thus it provides evidence that standard diagonal techniques are inadequate to settle such questions as '$\mathcal{P} = \mathcal{NP}$?'.

See also: Computability theory; Logical and mathematical terms, glossary of

ALASDAIR URQUHART

References and further reading

Baker, T., Gill, J. and Solovay, R. (1975) 'Relativizations of the $\mathcal{P} =? \mathcal{NP}$ Question', *SIAM Journal of Computation* 4: 431-42.(This paper shows that diagonalization techniques very likely do not suffice to answer the $\mathcal{P} =? \mathcal{NP}$ question.)

Cook, S.A. (1971) 'The Complexity of Theorem-Proving Procedures', in *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing*, New York: Association for Computing Machinery, 151-8.(The fundamental paper introducing the theory of $\mathcal{NP}$-completeness.)

Fraenkel, A.S. and Lichtenstein, D. (1981) 'Computing a Perfect Strategy for $n \times n$ Chess Requires Time Exponential in *n*', *Journal of Combinatorial Theory*, series A, 31: 199-213.(This paper shows generalized chess to be a hard game in the computational sense.)

Garey, M.R. and Johnson, D.S. (1979) *Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$-Completeness*, San Francisco, CA: Freeman.(Includes an extensive list of $\mathcal{NP}$-complete problems from many areas.)

Hartmanis, J., Lewis, P.M., II and Stearns, R.E. (1965) 'Classification of Computations by Time and Memory Requirements', in *Proceedings of the IFIP Congress 1965*, New York: Spartan Books, 31-5.(The first appearance of the space and time hierarchy theorems.)

Hopcroft, J.E. and Ullman, J.D. (1979) *Introduction to Automata Theory, Languages and Computation*, Reading, MA: Addison-Wesley.(A clearly written textbook giving the basic definitions and results of complexity theory.)

Leeuwen, J. van (ed.) (1990) *Handbook of Theoretical Computer Science*, vol. A, *Algorithms and Complexity*, Amsterdam: Elsevier.(A collection of detailed survey articles covering many topics not included in this entry, such as parallel complexity and cryptography.)

Mayr, E.W. and Meyer, A.R. (1982) 'The Complexity of the Word Problem for Commutative Semigroups and Polynomial Ideals', *Advances in Mathematics* 46: 305-29.(One of the earliest papers showing the existence of computationally intractable problems in classical algebra.)

Sipser, M. (1997) *Introduction to the Theory of Computation*, Boston, MA: PWS.(An elegant introduction to the basics of computational complexity theory, written by a major figure in the area.)

Stockmeyer, L. (1987) 'Classifying the Computational Complexity of Problems', *Journal of Symbolic Logic* 52: 1-43.(A very informative survey article.)

Compositionality

A language is compositional if the meaning of each of its complex expressions (for example, 'black dog') is determined entirely by the meanings its parts ('black', 'dog') and its syntax. Principles of compositionality provide precise statements of this idea. A compositional semantics for a language is a (finite) theory which explains how semantically important properties such as truth-conditions are determined by the meanings of parts and syntax. Supposing English to have a compositional semantics helps explain how finite creatures like ourselves have the ability to understand English's infinitely many sentences. Whether human languages are in fact compositional, however, is quite controversial.

It is often supposed that meaning, within a context of use, determines truth-value. Meaning and context are also generally thought to determine what terms refer to, and what predicates are true of - that is, to determine extensions. Because of this, compositionality is generally taken to license a principle of substitutivity: if expressions have the same meaning, and substituting one for another in a sentence does not change the sentence's syntax, the substitution can have no effect on truth. Given this, the idea that meaning is compositionally determined constrains what can be identified with meaning. For example, expressions with the same extension cannot always be substituted for one another without change of truth-value (see Propositional attitude statements §2; Modal logic, philosophical issues in §3).

Natural languages obey a principle of compositionality (PC) only if something more 'fine-grained' than extensions plays the role of meanings, so that expressions with the same extension can still have different meanings. Some say that no matter how fine-grained we make meaning, there will be counterexamples to the claim that meaning is compositionally determined. For example, 'woodchuck' and 'groundhog' are synonymous. But the sentences 'Mae thinks Woody is a woodchuck' and 'Mae thinks Woody is a groundhog' apparently may diverge in truth-value. (Certainly Mae may assert, with understanding, that Woody is a woodchuck, while denying, with understanding, that Woody is a groundhog. For she can understand 'woodchuck' and 'groundhog' but not realize that they are synonymous.) Such arguments lead to the conclusion that meaning is not compositionally determined (or that different forms cannot share meanings).

There is no consensus about such arguments. All agree PCs need *some* restrictions - they do not, for instance, apply to words in quotational constructions. So if 'thinks' is implicitly quotational, the argument loses its force (see Use/mention distinction and quotation). If 'thinks', as Frege held, produces a (more or less) systematic shift in the semantic properties of the expressions in its scope, this perhaps reduces the threat to compositionality (see Sense and reference §§4, 5). Some find such accounts of 'thinks' incredible (Davidson 1984).

Linguists often understand the claim that a language is compositional as asserting an extremely tight correspondence between its syntax and semantics (see Syntax). A (simplified) version of such a claim is that (after disambiguating simple word forms), there is, for each (simple) word, a meaning and, for each syntactic rule used in sentence construction, an operation on meanings, such that the meaning of any sentence is mechanically determined by applying the operations on meanings (given by the rules used in constructing the sentence) to the meanings of the simple parts. (Often a host of extra restrictions are incorporated. For example: the operations may be limited to applying function to argument; the order in which operations are applied may be settled by the structure of the sentence.) Some see such principles as providing significant constraints on semantical theories, constraints which may help us decide between theories which are in other respects equivalent (for example, Montague 1974).

Such strong PCs imply that every ambiguous sentence is either syntactically ambiguous or contains simple expressions which are themselves ambiguous (see Ambiguity). There are putative counterexamples to this. For instance, it has been alleged that 'The women lifted the piano' has two meanings, a 'collective' one (the group lifted it) and a 'distributive' one (each individual lifted it) (see Pelletier 1994).

Context sensitivity makes the formulation of PCs a delicate matter. In some important sense, differing uses of 'that is dead' have like meanings and syntax. But obviously they may differ in truth; presumably such differences may arise even within one context. How can this be reconciled with compositionality?

We might (implausibly) say that demonstrative word forms, strictly speaking, lack meanings (only tokens have such). An alternative (see Kaplan 1989) sees the meaning of the word type ‘that’ as ‘incomplete’, needing contextual supplementation to determine a meaning of the sort other terms have. A third strategy insists that if ‘that is dead but that is not’ is true, its ‘that’s are different demonstrative word types. (Advocates of this view are good at seeing very small subscripts.) This view makes it generally opaque to speakers whether tokens are tokens of the same word type, and thus opaque whether their arguments are logically valid. A fourth strategy sees context as providing an assignment of referents to *uses* of a type in a context. On most ways of working this out, it seems there is no true fleshing out of the principle: the meaning of a sentence type (a) is determined by the meanings of parts and syntax, and (b) determines, in context, the sentence’s truth-value. This approach also seems to undermine the idea that an argument like *that is dead; so that is dead* is formally valid (see Demonstratives and indexicals).

Even if relatively tight PCs should prove untenable (as claims about natural languages), it would not follow that natural languages did not have compositional semantics - finite theories which assign truth-conditions and other semantically relevant properties to sentence types or their uses. It has been argued that only if they do would it be possible for finite creatures like ourselves to learn them (since the languages involve infinite pairings of sounds and meanings); analogous arguments hold that our ability to understand (in principle) natural languages requires the existence of such semantics. Such arguments apparently presuppose that learning and understanding a language requires knowing a theory from which semantic facts (such as the fact that, for any speaker *x* and time *t*, a use of ‘J’existe encore’ by *x* at *t* says that, at *t*, *x* still exists) are deducible. While such a view is not implausible, alternative plausible views of competence do not support this sort of argument. For example, one might identify linguistic competence with the possession of *syntactic* knowledge (or even just with the possession of syntactic abilities) by someone with appropriate social and environmental relations and behavioural dispositions. Whether natural languages have compositional semantics, and whether the meanings of their sentences are determined simply by the meanings of their parts and syntax, is still not settled.

See also: Meaning and truth; Semantics

MARK RICHARD

References and further reading

Davidson, D. (1984) *Inquiries into Truth and Understanding*, Oxford: Oxford University Press.(Essay 1 argues that learnable languages have compositional semantics; essay 7 argues against Fregean accounts of ‘thinks’.)

Grandy, R. (1990) ‘Understanding and the Principle of Compositionality’, in J.E. Tomberlin (ed.) *Philosophical Perspectives*, Atascadero, CA: Ridgeview Publishing Company, vol. 4, 557-72.(Defends compositional semantics against the attacks of Schiffer and others.)

Kaplan, D. (1989) ‘Demonstratives’, in J. Almog, J. Perry and H. Wettstein (eds) *Themes From Kaplan*, Oxford: Oxford University Press.(Gives Kaplan’s views of demonstrative meaning.)

Montague, R.M. (1974) *Formal Philosophy*, New Haven, CT: Yale University Press. (Essays 6-8 give formulations of compositionality principles. Very difficult.)

Partee, B. (1984) ‘Compositionality’, in F. Landman and F. Veltman (eds) *Varieties of Formal Semantics*, Dordrecht: Foris, 281-311.(Evaluates Montague-style principles of compositionality. Difficult but illuminating.)

Pelletier, F.J. (1994) ‘The Principle of Semantic Compositionality’, *Topoi* 13: 11-24. (A good introduction to arguments for and against principles of compositionality.)

Schiffer, S. (1987) *Remnants of Meaning*, Cambridge, MA: MIT Press.(Contains arguments against the existence of compositional semantics for natural languages.)

Computability and information

The standard definition of randomness as considered in probability theory and used, for example, in quantum mechanics, allows one to speak of a process (such as tossing a coin, or measuring the diagonal polarization of a horizontally polarized photon) as being random. It does not allow one to call a particular outcome (or string of outcomes, or sequence of outcomes) 'random', except in an intuitive, heuristic sense. Information-theoretic complexity makes this possible. An algorithmically random string is one which cannot be produced by a description significantly shorter than itself; an algorithmically random sequence is one whose initial finite segments are almost random strings.

Gödel's incompleteness theorem states that every axiomatizable theory which is sufficiently rich and sound is incomplete. Chaitin's information-theoretic version of Gödel's theorem goes a step further, revealing the reason for incompleteness: a set of axioms of complexity N cannot yield a theorem that asserts that a specific object is of complexity substantially greater than N. This suggests that incompleteness is not only natural, but pervasive; it can no longer be ignored by everyday mathematics. It also provides a theoretical support for a quasi-empirical and pragmatic attitude to the foundations of mathematics.

Information-theoretic complexity is also relevant to physics and biology. For physics it is convenient to reformulate it as the size of the shortest message specifying a microstate, uniquely up to the assumed resolution. In this way we get a rigorous, entropy-like measure of disorder of an individual, microscopic, definite state of a physical system. The regulatory genes of a developing embryo can be ultimately conceived as a program for constructing an organism. The information contained by this biochemical computer program can be measured by information-theoretic complexity.

1 The paradox of randomness

Suppose that persons *A* and *B* each give us a string of 32 bits, saying that they were obtained from independent coin flips. If *A* gives the string

$x = 01101000100110101101100110100101$

and *B* gives the string

$y = 00000000000000000000000000000000,$

then we will believe *A* and will not believe *B*: the string *x seems* random, but the string *y* does not. Why? The strings are extremely different from the point of view of regularity. The string *y* has a maximum regularity which allows us to express it in a very compact way - 'only zeros' - while *x* appears to have no shortened definition at all. The distinction between regular and irregular strings becomes sharper as the strings get longer: it is clearly easier to specify the number

$10^{10^{10^{10}}}$

than the first 1,000 digits of π. Classical probability theory is not sensitive to the above distinction, as strings are all equally probable. However, even Laplace (1749-1827) was, in a sense, aware of this paradox; he suggested that non-random strings are strings possessing some kind of regularity, and since the number of all those strings (of a given length) is small, the occurrence of such a string is extraordinary.

Borel (1909), von Mises ([1928] 1919), Ville (1939) and Church (1940) all elaborated on this idea, but a formal model of irregularity was not found until the mid-1960s. Information-theoretic complexity was introduced independently by Solomonoff (1964), Kolmogorov (1965) and Chaitin (1966). The basic idea is to measure the informational complexity of an object by the size of the smallest program computing it. There are two principal applications: (1) to provide a new conceptual foundation for probability theory and statistics by making it possible to define rigorously the notions of random string and random sequence; and (2) to describe the major limitative theorems in mathematical logic in an information-theoretic form.

Let $A = \{a_1, a_2, \ldots, a_Q\}$, $Q \geq 2$, be a finite alphabet, and consider the set of all strings $x_1 x_2 \ldots x_n$ of elements in *A*. We denote by $|x|$ the length of the string *x*.

It is convenient to think of a computer as a decoding device at the receiving end of a communication channel. Its programs can be viewed as code words, or strings over the alphabet *A*, and the result of the computation as the decoded message. It is natural to require that successive programs (for example, subroutines) sent across the channel can be separated. This can be achieved by stipulating that programs are ‘self-delimiting’, that is, no meaningful program is a prefix of another. Self-delimiting programs carry within them information about their size, in the sense that they allow the computer to decide when to stop reading the input; they act as ‘atomic’ programs having no significant sub-programmatic structure. Formally, a ‘computer’ is a self-delimiting partial recursive function *C* which carries strings (programs) into strings (see Computability theory §§2-3). The (Chaitin) complexity $H_C(x)$ of a string *x* induced by the computer *C* is the length of the shortest program for *C* that outputs *x*, and halts:

$$H_C(x) = \min\{|y| \mid C(y) = x\}.$$

This notion of complexity is machine-dependent. It is possible to make it ‘asymptotically’ independent of the choice of the computer. The price paid amounts to at most an $\mathrm{O}(1)$ uncertainty in the numerical values, and is expressed by the following ‘invariance theorem’: there exists a computer *U* such that for all computers *C* one has

$$H_U(x) \leq H_C(x) + \mathrm{O}(1).$$

($\mathrm{O}(f)$ denotes a function $g(n)$ whose absolute value is bounded by a constant times $f(n)$, for sufficiently large *n*.)

A universal self-delimiting Turing machine (see Computability theory §4; Turing machines §2) is a possible candidate for *U*. Indeed, *U* simulates the computation of every computer *C* in the most efficient way: there exists a prefix δ such that the program δx makes *U* simulate the computation of *C* on *x*. The prefix δ depends upon *U* and *C*; it is independent of *x*. From now on we fix a computer *U* (called universal) and put $H_U = H$.

The set of meaningful programs for *U* is an instantaneous code, in the sense that each program can be decoded as soon as we come to the end of the string corresponding to it. In view of the invariance theorem it follows that the set of meaningful programs for *U* is an asymptotically optimal code for the set of all strings: every string can be coded by such a program, with minimum expected length.

In this context the joint complexity $H(x, y)$ of two strings x, y, defined to be the size of the smallest program that makes *U* output both of them, satisfies the classical inequality:

$$H(x, y) \leq H(x) + H(y) + \mathrm{O}(1),$$

that is, the complexity is asymptotically sub-additive. In the original formulation of these definitions the sub-additivity failed to be true. Chaitin (1974) and Levin (1974) independently found the solution: computers have to be self-delimiting.

2 Random strings

Randomness is not absolute, but relative. It makes more sense to ask ‘To what extent is *x* random?’ than ‘What is a random string?’.

To define random strings we try to capture the idea that a string is random if it cannot be algorithmically compressed. Formally, a string is ‘random’ if it has ‘almost’ maximal complexity within the set of all strings of the same length. The maximal value of the complexity of strings of length *n* is given by the formula

$$\max_{|x|=n} H(x) = n + H(\mathrm{string}(n)) + \mathrm{O}(1),$$

where string(*n*) is the *n*th string according to the lexicographic order. (Strings are enumerated in order of increasing length: all strings of length one, followed by all strings of length two, and so on.)

We are naturally motivated to adopt the following definition. A string *x* is (Chaitin) ‘*m*-random’ (*m* a natural number) if

$$H(x) \geq |x| + H(\mathrm{string}(|x|)) - m.$$

x is random if it is 0-random.

To validate the above model of randomness one has to show that random strings defined in this way possess (almost) all properties intuitively associated with randomness. We discuss here only two such properties: (1) the absence of any algorithmic rule for defining random strings; and (2) the equi-distribution of letters and blocks of letters. Property (1) holds true in a strong sense: the set of random strings is 'immune', that is, it is infinite and contains no infinite recursively enumerable subset (see Computability theory §5). The intuition motivating (2) is that in a 'truly random' string each letter has to appear with approximately the same frequency, namely Q^{-1}. Moreover, the same property should extend to substrings. Indeed, one proves that all 'reasonably long' substrings of a random string have approximately the same frequency, which depends only on the length of substrings.

3 Random sequences

An infinite sequence $\mathbf{x} = x_1 x_2 \ldots x_n \ldots$ is uniquely determined by its prefixes

$\mathbf{x}(n) = x_1 x_2 \ldots x_n, \quad n = 1, 2 \ldots .$

Is it possible to define the randomness of a sequence by means of the degree of randomness of its prefixes? The answer is affirmative. A sequence $\mathbf{x}$ is random if and only if there exists a natural number *c* such that $H(\mathbf{x}(n)) \geq n - c$ for all natural $n \geq 1$. This condition is equivalent to the following, apparently stronger, requirement (Chaitin 1974): a sequence $\mathbf{x}$ is random if and only if

$\lim_{n\to\infty} H(\mathbf{x}(n)) - n = \infty.$

Martin-Löf (1966) made use of constructive probability theory in order to define random sequences. The set of all sequences can be equipped with the product probability (induced by the uniform distribution of letters in the alphabet). A constructive null set (of sequences) is a set which can be recursively covered by recursively enumerable sets having arbitrarily small probabilities. The union of all constructive null sets is still a null set, in fact, a maximal one. It can be shown that the set of random sequences coincides exactly with the complement of the maximal null set. It follows that with probability one every sequence is random. However, from a topological point of view, the set of random sequences is of first Baire category, which is a rather small set.

Random sequences share many properties associated with randomness; for instance, they are Borel normal (every string *x* appears in a random sequence with the probability $Q^{-|x|}$, which depends only on the length of the string), they pass almost all stochasticity tests, they are incompatible with the slightest form of computability (the possibility to compute an infinite part of a sequence makes that sequence non-random). Random sequences generate, in a sense, all sequences; that is, every sequence is effectively reducible, via a fixed process, to a random one. Randomness is an invariant of number representations: the usual representations of a real number are either random for all bases, or not so for any base.

4 Chaitin's Omega number

Consider a universal self-delimiting program which, instead of being given a specific program at the beginning of the computation, is fed with a 'random string' of bits. Whenever the universal program requests another input bit we just toss a fair coin and input 1 or 0, according to whether the coin comes up heads or tails. Finally we ask the question 'When the above procedure is begun, what is the probability that the universal program will eventually halt?'. The answer is Chaitin's Omega number, Ω.

The sequence of digits of Ω is random. This means that we will never know more than a handful of digits of Ω. This number appears to be cabalistic: it can be known of, but not itself known.

The fascination of Ω comes from the fact that it is computationally equivalent to the halting problem (see Computability theory §5), but contains the same information in a fantastically compressed form. An oracle supplying a few digits of Ω would be of invaluable help in answering many mathematical questions, such as Goldbach's conjecture (we may construct a computer which halts if and only if it finds an even number greater than 2 which is not the sum of two primes), or Riemann's hypothesis.

From a practical point of view, even having access to all digits of Ω would not make the task of solving problems very easy, as the time (see Complexity, computational) to decompress Ω grows extremely rapidly, that is, non-recursively.

The classical opinion that an axiom should be self-evident appears to be too restrictive. Neither Riemann's hypothesis, nor Schrödinger's equation are self-evident, but both are very useful. The possibility of finding new axioms for the positive integers should perhaps get more attention. This quasi-empirical attitude seems to be consistent with Gödel's own views:

> axioms need not be evident in themselves, but rather their justification lies (exactly as in physics) in the fact that they make it possible for these 'sense perceptions' to be deduced…I think that…this view has been largely justified by subsequent developments, and it is to be expected that it will be still more so in the future.
>
> ([1944] 1990: 121)

5 Randomness in physics

All science is founded on the assumption that the physical universe is ordered and rational. The most powerful expression of this state of affairs is found in the successful application of mathematics to make predictions expressed by means of the laws of physics. Where do these laws come from? Why do they operate universally and unfailingly? Nobody seems to have reasonable answers to these questions. The most we can do is to explain that the hypothesis of order is supported by our daily observations: the rhythm of day and night; the regular ticking of clocks. However, there is a limit to this perceived order. How are we to reconcile these seemingly random processes with the supposed order? How can the same physical process (for example, the spin of a roulette wheel) obey two contradictory laws, the laws of chance and the laws of physics?

As our direct information refers to finite experiments, it is not out of the question to discover local rules, functioning on large, but finite scales, even if the global behaviour of the process is truly random. But, to perceive this global randomness we have to have access to infinity, which is not physically possible. It is important to notice that, consistently with our common experience, facing global randomness does not imply the impossibility of making predictions. Space scientists can pinpoint and predict planetary locations and velocities 'well enough' to plan missions years in advance. Astronomers can predict solar or lunar eclipses centuries before their occurrence. However, we have to be aware that all these results - superb as they may be - are only true within a certain degree of precision, which can hardly be evaluated.

This suggests the idea that the universe might be random (see Calude and Salomaa 1994). Here is an example illustrating this idea. Consider reversible transformations of a string $x, x_{t+1} = T(x_t)$. This equation can be regarded as a deterministic law of motion generating the phase-space trajectory of a simple, idealized system. The complexity $H(x_t)$ tends to increase in the course of evolution, provided the initial state is not too complex. It approaches the equilibrium value consistently with the Boltzmann estimate of the entropy, and fluctuates around the equilibrium. This shows how a reversible, deterministic, dynamic evolution can lead to a significant increase of complexity (see Zurek 1989).

Any physical system may be perceived as a computational process. One could even go further and speculate (with Wheeler, Fredkin, Toffoli) that the universe is a gigantic information-processing system, in which the output is as yet undetermined. In Wheeler's compact form: 'It from bit!', that is, 'it' - every force, particle, and so on - is ultimately present through 'bits' of information. Under this interpretation Ω - which represents the halting probability of the system - acquires physical significance. This constant is not only non-computable, but reflects, in a sense, the indeterministic, chaotic nature of the universe, consistently with Svozil's thesis (1993) according to which chaos in physics corresponds to randomness in mathematics.

6 Logical depth

Charles Bennett's thesis (1988) is that a structure is 'deep' if it is superficially random but subtly redundant; in other words, if almost all its algorithmic probability is contributed by slow-running programs. The value of a message is only slightly related to its information content. For example, a typical sequence of coin tosses has high information content, but little message value. According to Bennett, the 'logical depth' of a message can be measured by the amount of computational work required by its originator, which its receiver is saved from having to repeat. Here we include things which are computable in principle, but only at a considerable cost in resources. Tolstoy's *War and Peace*, Leonardo da Vinci's portrait of *Mona Lisa*, or phase-space sheets that evolve from simple coarse-grained cells via the chaotic dynamics of a simple Hamiltonian are objects with large logical depth.

See also: Logical and mathematical terms, glossary of

CRISTIAN S. CALUDE

References and further reading

Atiyah, M. ***et al.*** (1994) 'Responses to ''Theoretical Mathematics: Toward a Cultural Synthesis of Mathematics and Theoretical Physics'', by A. Jaffe and F. Quinn', *Bulletin of the American Mathematical Society* 30: 178-207.(Enlightening and provocative essays by, among others, M. Atiyah, A. Borel, G. Chaitin, S. Mac Lane, B. Mandelbroit and R. Thom, on the nature of mathematics.)

Bennett, C.H. (1988) 'Logical Depth and Physical Complexity', in R. Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press, 227-58.(Expansion of the material of §6.)

Borel, É. (1909) 'Les probabilités dénombrables et leurs applications arithmétiques' (Enumerable Probabilities and their Arithmetical Applications), *Rendiconti del Circolo Matematico di Palermo* 27: 247-71.(An important paper inaugurating the modern analysis of randomness.)

Brisson, L. and Meyerstein, F.W. (1994) *Inventing the Universe*, New York: State University of New York Press.(Some philosophical consequences of Chaitin's work.)

Calude, C. (1994) *Information and Randomness: An Algorithmic Perspective*, Berlin: Springer.(An informative introduction to basic results in algorithmic information theory in a general setting.)

Calude, C. and Jürgensen, H. (1994) 'Randomness as an Invariant for Number Representations', in H. Maurer, J. Karhumäki and G. Rozenberg (eds) *Results and Trends in Theoretical Computer Science*, Berlin: Springer, 44-66.(A proof of the invariance of randomness with respect to number representations and a discussion of random real numbers.)

Calude, C., Jürgensen, H. and Zimand, M. (1994) 'Is Independence an Exception?', *Applied Mathematics and Computation* 66: 63-76.(A topological proof showing that the set of all true but unprovable statements is large.)

Calude, C. and Salomaa, A. (1994) 'Algorithmically Coding the Universe', in G. Rozenberg and A. Salomaa (eds) *Developments in Language Theory*, Singapore: World Scientific, 472-92.(An informal discussion of the randomness hypothesis discussed in §5.)

Chaitin, G.J. (1966) 'On the Length of Programs for Computing Finite Binary Sequences', *Journal of the Association for Computing Machinery* 13: 547-69.(Chaitin's first contribution to the subject.)

Chaitin, G.J. (1974) 'A Theory of Program Size Formally Identical to Information Theory', in *Abstracts of Papers, 1974 IEEE International Symposium on Information Theory*, Indiana, IN: University of Notre Dame Press; repr. in *Journal of the Association for Computing Machinery* 22: 329-40, 1975.(The first systematic presentation of self-delimiting computers and their complexities; discusses complexity-theoretic characterization of random sequences.)

Chaitin, G.J. (1987a) *Algorithmic Information Theory*, Cambridge: Cambridge University Press.(The classic introduction to the subject.)

Chaitin, G.J. (1987b) *Information, Randomness and Incompleteness: Papers on Algorithmic Information Theory*, Singapore: World Scientific; 2nd edn, 1990.(An anthology of Chaitin's seminal papers.)

Chaitin, G.J. (1992) *Information-Theoretic Incompleteness*, Singapore: World Scientific.(An original and systematic examination of the incompleteness phenomenon from an information-theoretic point of view.)

Church, A. (1940) 'On the Concept of a Random Sequence', *Bulletin of the American Mathematical Society* 46: 130-5.(Develops von Mises' ideas in the framework of computability theory.)

Gács, P. (1992) 'Every Sequence is Reducible to a Random One', *Information and Control* 70: 186-92. (A proof of the reducibility of any sequence to a random one.)

Gödel, K. (1944) 'Russell's Mathematical Logic', in P.A. Schlipp (ed.) *The Philosophy of Bertrand Russell*, Evanston, IL: Northwestern University Press, 123-53; repr. in *Collected Works*, ed. S. Feferman *et al.*, Oxford: Oxford University Press, 1990, vol. 2, 119-41.(Source of the quote in §4.)

Kolmogorov, A.N. (1965) 'Three Approaches for Defining the Concept of ''Information Quantity''', *Problems in Information Transmission* 1: 3-11.(Expounds the considerations leading to the algorithmic definition of random strings.)

Levin, L.A. (1974) 'Randomness Conservation Inequalities: Information and Independence in Mathematical Theories', *Problems in Information Transmission* 10: 206-10.(Discussion of self-delimiting computers and their complexities.)

Li, M. and Vitányi, P.M. (1993) *An Introduction to Kolmogorov Complexity and Its Applications*, Berlin: Springer.(A comprehensive treatise on descriptional complexities.)

Martin-Löf, P. (1966) 'The Definition of Random Sequences', *Information and Control* 9: 602-19.(An important piece of work describing random sequences in the framework of constructive probability theory.)

Mises, R. von (1928) *Wahrscheinlichkeit Statistik und Wahrheit*, Vienna: Springer; trans. *Probability, Statistics and Truth*, London: Allen & Unwin; 2nd revised edn prepared by H. Geiringer, New York: Macmillan, 1961. (Expounds the frequential analysis of randomness, as started by the author in his 1919 paper 'Grundlagen der Wahrscheinlichkeitsrechnung', *Mathematische Zeitschrift* 5: 52-99.)

Rozenberg, G. and Salomaa, A. (1994) *Cornerstones of Undecidability*, Englewood Cliffs, NJ: Prentice-Hall.(An enlightening presentation of undecidability and randomness.)

Schnorr, C.P. (1981) *Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Behandlung der Wahrscheinlichkeitstheorie (Randomness and Probability: An Algorithmic Foundation of Probability Theory)*, Berlin: Springer.(The first monograph on algorithmic information theory.)

Shannon, C.E. and Weaver, C.E. (1949) *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press.(The first and still one of the best books on classical information theory.)

Solomonoff, R.J. (1964) 'A Formal Theory of Inductive Inference, Part 1 and Part 2', *Information and Control* 7: 1-22, 224-54.(The first paper dealing with descriptional complexity.)

Solovay, R.M. (1975) 'Draft of a paper (or series of papers) on Chaitin's work…done for the most part during the period of Sept.-Dec. 1974', unpublished manuscript, Yorktown Heights, NY: IBM Thomas J. Watson Research Center.(A highly technical and original presentation of Chaitin's results.)

Svozil, K. (1993) *Randomness & Undecidability in Physics*, Singapore: World Scientific.(The first systematic presentation of algorithmic randomness in physics.)

Ville, J. (1939) *Étude critique de la notion de collectif (A Critical Study of the Notion of Collective)*, Paris: Gauthier-Villars.(Critical examination of von Mises' model of randomness.)

Zurek, W.H. (1989) 'Algorithmic Randomness and Physical Entropy', *Physical Review A* 40: 4,731-51. (A detailed discussion of the relevance of information-theoretic complexity for physics.)

Computability theory

*The effective calculability of number-theoretic functions such as addition and multiplication has always been recognized, and for that judgment a rigorous notion of 'computable function' is not required. A sharp mathematical concept was defined only in the twentieth century, when issues including the decision problem for predicate logic required a precise delimitation of functions that can be viewed as effectively calculable. Predicate logic emerged from Frege's fundamental '*Begriffsschrift*' (1879) as an expressive formal language and was described with mathematical precision by Hilbert in lectures given during the winter of 1917-18. The logical calculus Frege had also developed allowed proofs to proceed as computations in accordance with a fixed set of rules; in principle, according to Gödel, the rules could be applied 'by someone who knew nothing about mathematics, or by a machine'.*

Hilbert grasped the potential of this mechanical aspect and formulated the decision problem for predicate logic as follows: 'The Entscheidungsproblem *[decision problem] is solved if one knows a procedure that permits the decision concerning the validity, respectively, satisfiability of a given logical expression by a finite number of operations.' Some, for example, von Neumann (1927), believed that the inherent freedom of mathematical thought provided a sufficient reason to expect a negative solution to the problem. But how could a proof of undecidability be given? The unsolvability results of other mathematical problems had always been established relative to a determinate class of admissible operations, for example, the impossibility of doubling the cube relative to ruler and compass constructions. A negative solution to the decision problem obviously required the characterization of 'effectively calculable functions'.*

For two other important issues a characterization of that informal notion was also needed, namely, the general formulation of the incompleteness theorems and the effective unsolvability of mathematical problems (for example, of Hilbert's tenth problem). The first task of computability theory was thus to answer the question 'What is a precise notion of "effectively calculable function"?'. Many different answers invariably characterized the same class of number-theoretic functions: the partial recursive ones. Today recursiveness or, equivalently, Turing computability is considered to be the precise mathematical counterpart to 'effective calculability'. Relative to these notions undecidability results have been established, in particular, the undecidability of the decision problem for predicate logic. The notions are idealized in the sense that no time or space limitations are imposed on the calculations; the concept of 'feasibility' is crucial in computer science when trying to capture the subclass of recursive functions whose values can actually be determined.

1 Primitive recursive functions

Let us start with the class of 'primitive recursive' functions (from *n*-tuples of natural numbers to the natural numbers). A precise definition of this class was given by Gödel in 1931 (see van Heijenoort 1967), although he then called it simply the class of recursive functions. The main definitional schema of primitive recursion used in the generation of the elements of this class was well-known in mathematics, having been presented most clearly by Dedekind (1888), who had given a set-theoretic reconstruction that turned out to be of importance also in computability theory. Skolem established in 1923 that most functions in elementary number theory are primitive recursive. The foundational significance of this function class was emphasized by Hilbert and Bernays (1934, 1939): the values of the functions (for any argument) can be determined in finitely many steps, proceeding purely 'mechanically'. This point is also expressed by saying that the functions are 'effectively calculable'.

The class $\mathcal{PR}$ of primitive recursive functions is specified inductively and contains as its initial functions the zero-function Z, the successor function S, and the projection functions P^n_i for each *n* and each *i* with $1 \leq i \leq n$. These functions satisfy the equations $\mathrm{Z}(x) = 0$, $\mathrm{S}(x) = x'$ and $\mathrm{P}^n_i(x_1, \ldots, x_n) = x_i$, for all $x, x_1, \ldots, x_n$; x' is the 'successor' of *x*. The class is, first of all, closed under the 'schema of composition': given an *m*-place function ψ and *n*-place functions $\varphi_1, \ldots, \varphi_m$ all in $\mathcal{PR}$, the function ϕ defined by

$$\phi(x_1, \ldots, x_n) = \psi(\varphi_1(x_1, \ldots, x_n), \ldots, \varphi_m(x_1, \ldots, x_n))$$

is also in $\mathcal{PR}$; ϕ is then said to be obtained 'by composition' from ψ and $\varphi_1, \ldots, \varphi_m$. $\mathcal{PR}$ is also closed under the 'schema of primitive recursion': given an *n*-place function ψ and an $(n+2)$-place function φ both in $\mathcal{PR}$,

the function ϕ defined by

$$\phi(x_1,\dots,x_n,0)=\psi(x_1,\dots,x_n)$$
$$\phi(x_1,\dots,x_n,y')=\varphi(x_1,\dots,x_n,y,\phi(x_1,\dots,x_n,y))$$

is also a function in $\mathcal{PR}$; ϕ is said to be obtained 'by primitive recursion' from ψ and φ. A function is 'primitive recursive' if and only if it can be obtained from the initial functions by finitely many applications of the composition and primitive recursion schemas.

The primitive recursive functions do not exhaust the class of effectively calculable functions; that was shown by Ackermann and independently by Gabriel Sudan, who constructed functions that are obviously effectively calculable, but not primitive recursive (see van Heijenoort 1967 for detailed references). The 'Ackermann function' can be defined by the following recursion equations.

$$\phi_0(x,y)=\mathrm{S}(y)$$
$$\phi_{n'}(x,0)=\begin{cases}x & \text{if } n=0\\ 0 & \text{if } n=1\\ 1 & \text{otherwise}\end{cases}$$
$$\phi_{n'}(x,y')=\phi_n(x,\phi_{n'}(x,y))$$

Notice that ϕ_1 is addition, ϕ_2 is multiplication, ϕ_3 is exponentiation, and so on; that is, the next function is always obtained by iterating the previous one. For each *n*, the function $\phi_n(x,x)$ is primitive recursive, but $\phi_x(x,x)$ is not!

We can construct another example of an effectively calculable, but not primitive recursive function ψ by effectively listing all primitive recursive unary functions $\varphi_0,\varphi_1,\varphi_2,\varphi_3,\dots$ and then setting $\psi(x)=1+\varphi_x(x)$. Since ψ differs from every φ_i, ψ cannot be primitive recursive; but the effectiveness of our listing of the functions φ_i ensures that ψ can be effectively calculated.

For working with this class of functions it is indispensable to establish further closure properties. To allow, for example, directly general explicit definitions (without the cumbersome detour of projections) of functions such as

$$f(x,y,z)=(xy+z)^2+789,$$

one establishes that variables can be permuted, substitutions can be carried out, and that the constant functions $\mathrm{con}^n_m(x_1,\dots,x_n)=m$ are primitive recursive. Beyond addition, multiplication and exponentiation it is useful to know that the following functions are also in $\mathcal{PR}$.

$x!$	(factorial)
$\mathrm{pred}(x)$	(predecessor)
$x \dot{-} y$	(arithmetic subtraction, $\max(0,x-y)$)
$\overline{\mathrm{sg}}(x)$	$(=1 \dot{-} x)$
$\mathrm{sg}(x)$	$(=1 \dot{-} \overline{\mathrm{sg}}(x))$

Then it is very easy to define the bounded sum of an $(n+1)$-place function ϕ in $\mathcal{PR}$ as follows.

$$\phi(x_1,\dots,x_n,0)=0$$
$$\sum_{x<y'}\phi(x_1,\dots,x_n,x)=\sum_{x<y}\phi(x_1,\dots,x_n,x)$$
$$+\phi(x_1,\dots,x_n,y).$$

The bounded product is defined similarly.

The functions we listed are useful for the investigation of primitive recursive relations; an *n*-place relation *R* between natural numbers is called primitive recursive if and only if its characteristic function χ_R is in $\mathcal{PR}$. (For recursion-theoretic purposes it is convenient to consider 0 to be the truth-value T and 1 the truth-value F.) It is now possible to show directly that the primitive recursive relations are closed under Boolean operations - particularly, '

$\neg$' (not), 'Čamp;' (and), ' $\vee$' (or), ' $\rightarrow$' (if…then) - and bounded quantification, that is, '($\forall x < y$)' (for all *x* less than *y*) and '($\exists x < y$)' (for some *x* less than *y*).

Finally, we discuss bounded minimization: if for some *x* less than *y* $R(x_1, \ldots, x_n, x)$, then the function $\mu x < y.R(x_1, \ldots, x_n, x)$ yields the smallest such *x*, otherwise the function yields *y*. Define first

$$\phi_1(x_1, \ldots, x_n, x) = \begin{cases} 0 & \text{if } (\exists z \leq x) R(x_1, \ldots, x_n, z) \\ 1 & \text{otherwise} \end{cases}$$

and notice that

$$\mu x < y.R(x_1, \ldots, x_n, x) = \sum_{x<y} \phi_1(x_1, \ldots, x_n, x).$$

Thus, if the relation *R* is primitive recursive, bounded minimization leads to a function in $\mathcal{PR}$.

Given this list of functions in $\mathcal{PR}$ and the closure conditions for primitive recursive relations, we have a convenient framework in which we can show the primitive recursiveness of number-theoretic relations; for example, the characteristic function $\chi_<$ of the 'less than' relation ' $<$' is given by

$$\chi_<(x, y) = \overline{\mathrm{sg}}(y \dot{-} x);$$

equality, being a divisor of, and being prime are all defined easily. This indicates that the class is mathematically rather rich.

2 Finite computations

In 1931 and in his Princeton lectures of 1934, Gödel used primitive recursive functions and relations to describe the syntax of particular 'formal' theories - after Gödel-numbering the syntactic configurations that make up the theory. Since he strove to arrive at a general concept of formality through the underlying concept of calculability for functions, there was no reason to focus attention on theories whose syntax could be presented primitive recursively. He viewed primitive recursive definability of formulas and proofs as a precise condition which *in practice* sufficed to describe formal systems, but he was searching for a condition that would suffice *in principle*.

In his Princeton lectures, Gödel considered it a very important property that the values of primitive recursive functions can be calculated by a finite computation for arbitrary arguments. He added in a footnote:

> The converse seems to be true if, besides recursions according to the scheme (2) [of primitive recursion], recursions of other forms…are admitted. This cannot be proved, since the notion of finite computation is not defined, but it can serve as a heuristic principle.
>
> (1934)

Parenthetically, we note that this remark does not contain a formulation of Church's thesis (see Church's thesis); that (prima facie plausible) reading was explicitly ruled out by Gödel in a letter to Martin Davis (1982). It does, however, indicate the motivation for the considerations in the last part of Gödel's lectures: there he used quite general forms of recursion to introduce 'general recursive functions'. (The relation of Gödel's proposal to suggestions of Herbrand is discussed in Sieg (1994).) These functions are obtained as unique solutions of a system *E* of equations, and their values must be computable in an equational calculus with just two obviously mechanical rules: the first rule allows the substitution of numerals for variables in any equation derived from *E*; the second rule allows the replacement of terms $t(\nu_1, \ldots, \nu_n)$ in a derived equation by μ, in case $t(\nu_1, \ldots, \nu_n) = \mu$ is a derived equation. (Lower-case Greek letters stand here for numerals.)

This class of general recursive functions very quickly turned out to be characterizable in a variety of ways: Church and Kleene showed the equivalence to λ-definability, Kleene to μ-recursiveness (discussed below), and Turing to computability by his machines (see Lambda calculus; Combinatory logic; Turing machines). The class introduced by Gödel was used by Church in his first formulation of 'Church's thesis' (see Sieg 1997). The early, pre-Turing attempts to argue for the thesis are captured in a mathematically concise way through the concept of a function 'reckonable according to rules' (*regelrecht auswertbare Funktion*). These functions were introduced by Hilbert and Bernays in the second volume of their *Grundlagen der Mathematik* (1934, 1939); they were characterized as being calculable in deductive formalisms satisfying general recursiveness conditions; the critical

condition required the proof predicate of the formalisms to be primitive recursive. It was shown that the calculations could be carried out in a particular subsystem of arithmetic; Gödel took this fact in 1946 as the basis for his claim that computability is an 'absolute' concept (see Sieg 1994). The analysis of Hilbert and Bernays also clearly revealed the 'stumbling block' all these analyses encountered: they tried to characterize the elementary nature of steps in calculations, but could not do so without referring to recursiveness (Church), primitive recursiveness (Hilbert and Bernays), or to very specific rules (Gödel).

Only Turing was able, in his ground-breaking 1936 paper 'On Computable Numbers' (see Davis 1965), to circumvent the stumbling block by focusing on the calculations of a human 'computor' proceeding mechanically (see Turing machines §1). He formulated general boundedness and locality conditions for such a computor and showed that any number-theoretic function whose values can be calculated by a computor satisfying these conditions can actually be computed by a Turing machine (see Mundici and Sieg 1995). The latter fact is sometimes called 'Turing's theorem'. The restricted formulation of Turing machines allows a uniform and simple description of computations; its adequacy for linear computations is guaranteed by Turing's theorem. The starting point of Turing's analysis was, however, the mechanical behaviour of a human computor operating on finite configurations in the plane. This behaviour can be described generally and mathematically precisely. To arrive at such a more general description, we take a preliminary step and replace the states of mind by what Turing described as 'physical and definite counterparts'. This is done by considering 'state of mind' not as a property of the working computor, but rather as part of the configuration which is operated on. Turing discussed this replacement very vividly in his 1936 paper:

> It is always possible for the computer [in our terminology, the computor] to break off from his work, to go away and forget all about it, and later to come back and go on with it. If he does this he must leave a note of instructions (written in some standard form) explaining how the work is to be continued. This note is the counterpart of the 'state of mind'. We will suppose that the computer works in such a desultory manner that he never does more than one step at a sitting. The note of instructions must enable him to carry out one step and write the next note.
>
> ('On Computable Numbers', in Davis 1965)

This was achieved quite beautifully by Post in 1947, and Post's approach of describing Turing machine computations is used by Davis (1958). The configurations on which the machine works are instantaneous descriptions (IDs). These are finite sequences in the alphabet of a Turing machine containing exactly one state symbol; the position of the state symbol indicates which symbol is being scanned. A program of a Turing machine can now be viewed as a set of Post production rules operating on (a single symbol of) such IDs. If in this way of describing Turing machines one replaces finite sequences by finite graphs (with a few well-motivated properties) and the simple Post-Turing operations on one symbol at a time by operations on a fixed finite number of distinguished graphs, then one arrives at the notion of a 'Kolmogorov machine'. This latter notion, or rather a general concept of algorithm, was introduced by Kolmogorov and Uspensky (1958): Kolmogorov machines compute exactly the Turing computable number-theoretic functions.

Which number-theoretic functions can be computed by Turing machines or other computing devices? Kleene's analysis of the equational calculus led to the introduction of the 'regular minimization' operator and to an inductive characterization of Gödel's class of general recursive functions. Suppose the $(n+1)$-place function ϕ has the property that for every $x_1, \ldots, x_n$ there is a *y* such that $\phi(x_1, \ldots, x_n, y) = 0$. Denote by $\mu y.\phi(x_1, \ldots, x_n, y) = 0$ the least such *y*. Under these conditions, the function ψ given by

$$\psi(x_1, \ldots, x_n) = \mu y.\phi(x_1, \ldots, x_n, y) = 0$$

is said to be obtained from ϕ 'by regular minimization'. A function is '(μ-)recursive' if and only if it can be obtained from the initial functions by a finite number of applications of composition, primitive recursion and regular minimization. That is, the class $\mathcal{R}$ of recursive functions is obtained from $\mathcal{PR}$ by closure under regular minimization.

Up to now we have considered total (that is, everywhere defined) functions and operations that do not lead outside the class of total functions. Gödel (after Kleene) emphasized the importance of partial functions defined only for a subset of $\mathbb{N}$ or $\mathbb{N} \times \ldots \times \mathbb{N}$. For example, while $\mu y.xy = 0$ is an equivalent definition of the zero function, an

expression such as $\mu y.x + y = 0$ is undefined for each $x > 0$. For partial functions ϕ we liberalize the regularity condition for minimization:

$u = \mu y.\phi(x_1, \ldots, x_n, y) = 0$

holds if and only if the following two conditions are satisfied:

(1) For any $i = 0, 1, \ldots, u - 1,$ $\phi(x_1, \ldots, x_n, i)$ is defined and different from 0;

$\phi(x_1, \ldots, x_n, u) = 0.(2)$

Given $x_1, \ldots, x_n$, at most one number *u* can satisfy both (1) and (2). If no such *u* exists then the *n*-tuple $(x_1, \ldots, x_n)$ is outside the domain of the function

$\mu y.\phi(x_1, \ldots, x_n, y) = 0$

or, stated differently, this function is undefined at $(x_1, \ldots, x_n)$. A function is 'partial recursive' if and only if it can be obtained from the initial functions by a finite number of applications of composition, primitive recursion and (unrestricted) minimization. Note that the partial recursive functions can be enumerated by a partial recursive function; that is obviously not possible in the case of total functions: a standard diagonal construction leads immediately to a contradiction. This fact (for partial recursive functions) and its significance are described in §3 and the beginning of §4.

One can explicitly construct Turing machines for the initial functions and for those operations on programs that correspond to composition, primitive recursion and regular, as well as unrestricted, minimization. Thus, these operations do not lead outside the class of Turing computable functions. This immediately yields the following theorem, showing the power of Turing machines: all partial recursive functions are Turing computable.

3 Effective descriptions

To prove the converse of the theorem just stated, we follow Pythagoras' ontological prescription and use Gödel's device to assign numbers *e* and *y* as 'codes' to Turing machines *M* and finite sequences of IDs. The coding introduced by Gödel for formal theories in 1931 (and adapted for Turing machines in Davis' book) will do. Once a machine is coded as a number, the code can be supplied as an input to any machine; this adumbrates the metamorphosis of hardware into software, culminating in the stored-program computer. Programming the first computers virtually amounted to inserting wires into plugboards, but von Neumann - influenced by Turing's treatment of his paper machines - realized that programs can be coded and stored as strings of symbols in the same way as data.

For any effective Gödel-numbering it is easy, though somewhat tedious, to establish that simple syntactic notions and operations concerning machines and their IDs can be represented by number-theoretic, indeed primitive recursive, predicates and functions. For instance, the ternary predicate $T(e, x, y)$ expressing that *y* is (the code of) a computation of Turing machine (*M* with code) *e* for input *x* is primitive recursive; also, the unary 'result-extracting' function *U* is primitive recursive, where $U(y)$ is either the number on the tape of the last ID of the computation *y* or, in case *y* is not the code of a computation, $U(y) = 0$. (Our choice of the default value 0 is only to guarantee that *U* is a total function.)

Given a Turing computable function ϕ and a machine *M* with code *e* computing ϕ, there are two possibilities for any input *x*:

Case 1. M terminates. Then there is a (first) computation y^* of *M* for input *x*, that is, y^* is the smallest number *y* such that $T(e, x, y)$; the number $U(y^*)$ read on the tape in the final configuration of y^* coincides with the value $\phi(x)$; writing $\mu y.T(e, x, y)$ as an abbreviation for $\mu y.\chi_T(e, x, y) = 0$, we have

$\phi(x) = U(\mu y.T(e, x, y)).(3)$

Case 2. M does not terminate. Then for any *y*, the predicate $T(e, x, y)$ will be false, that is, $\chi_T(e, x, y) = 1$ for all *y*, and condition (2) in the definition of the unrestricted μ-operator will never hold. Thus, both $\mu y.T(e, x, y)$ and $U(\mu y.T(e, x, y))$ are undefined.

This establishes (3) for both cases, and we have proved 'Kleene's normal form theorem' for unary Turing computable functions:

Let φ be an *n*-place number-theoretic function that can be computed by a Turing machine with code *e*; then for each $x_1, \ldots, x_n$ in the domain of φ we have
$\varphi(x_1, \ldots, x_n) = U(\mu y.T(e, x_1, \ldots, x_n, y)).$

Kleene's theorem holds also, with the same proof, for functions having any finite number of arguments. Kleene established this normal form theorem not for Turing computable functions, but for total functions that can be calculated in Gödel's equational calculus. He concluded that Gödel's general recursive functions are recursive, a conclusion we can draw now for Turing computable partial functions: every Turing computable function is partial recursive.

Using the theorem formulated at the very end of §2, it follows that Turing computability for number-theoretic functions is equivalent to partial recursiveness and, indeed, to all the other characterizations discussed in that section. Kleene's theorem has most interesting additional consequences, to be discussed in the next section. This is done not just to present elegant mathematical results, but to reinforce the conceptual analysis, as these results (are taken by some to) lend additional support to Church's thesis (see Church's thesis).

4 Basic results

Kleene's normal form theorem has important consequences for the theory of computability. Consider, first of all, the two-place function

$\psi(e, x) = U(\mu y.T(e, x, y));$

ψ is partial recursive and provides an enumeration of all unary partial recursive functions. This is 'Kleene's enumeration theorem'. This theorem - together with the equivalence of partial recursiveness and Turing computability - guarantees the existence of a 'universal' Turing machine. Indeed, consider a Turing machine $M[\psi]$ computing ψ. For any pair of arguments *e* and *x*, $M[\psi]$ interprets its first argument as the code of a Turing machine $M[\varphi]$ (computing a unary partial recursive function φ) and its second argument as the input for $M[\varphi]$. Then $M[\psi]$ proceeds to compute $\varphi(x)$ following the program of $M[\varphi]$. Since $M[\varphi]$ was arbitrary, $M[\psi]$ is able to simulate any Turing machine. Similarly, for each $n = 2, 3, 4, \ldots$, there is a partial recursive $(n+1)$-place function $\psi^{(n)}$ such that every *n*-place partial recursive function φ can be written as

$\varphi(x_1, \ldots, x_n) = \psi^{(n)}(e, x_1, \ldots, x_n)$

for a suitable index *e* for φ.

Given a function $\varphi(p_1, \ldots, p_m, x_1, \ldots, x_n)$ of $m+n$ variables, let us 'parametrize' φ, that is, assign values $p_1^*, \ldots, p_m^*$ to the first *m* variables. We then obtain a new *n*-place function ϕ as follows:

$\phi(x_1, \ldots, x_n) = \varphi(p_1^*, \ldots, p_m^*, x_1, \ldots, x_n).$

Intuitively, ϕ is the 'slice' of φ with coordinates $(p_1^*, \ldots, p_m^*)$. By Kleene's enumeration theorem we can write, for suitable indices *e* and e^*:

$$\varphi(p_1, \ldots, p_m, x_1, \ldots, x_n) = \psi^{(m+n)}(e, p_1, \ldots, p_m, x_1, \ldots, x_n)$$

and

$\phi(x_1, \ldots, x_n) = \psi^{(n)}(e^*, x_1, \ldots, x_n).$

The following 'parameter (or S_n^m-) theorem' tells us that the slicing operation leading from *e* to e^* can be performed uniformly and effectively:

For each $m, n > 0$ there is a primitive recursive $(n+1)$-place function S_n^m such that

$$\psi^{(m+n)}(e, p_1, \ldots, p_m, x_1, \ldots, x_n)$$
$$= \psi^{(m)}(\mathrm{S}_n^m(e, p_1, \ldots, p_m), x_1, \ldots, x_n).$$

This theorem was established by Kleene at first for Gödel's general recursive functions; it amounts in that context to replacing effectively the first *m* variables of the function φ computed from equations with code *e* by numerals for $p_1, \ldots, p_m$. This theorem is used to prove the following central result, also due to Kleene, the 'recursion theorem':

Let ϕ be a partial recursive function with $(n+1)$ places. Then there is a number *e* such that for all $x_1, \ldots, x_n$

$$\psi^{(n)}(e, x_1, \ldots, x_n) = \phi(e, x_1, \ldots, x_n).$$

The theorem finds many uses in establishing that implicitly defined functions are actually partial recursive. For example, letting $n = 1$ and writing ψ instead of $\psi^{(1)}$, an application of the recursion theorem to the function $g(y, x) = y$ immediately yields the 'fixed point theorem':

There is a constant e^* such that for all *x*

$$\psi(e^*, x) = e^*.$$

Intuitively, for every possible input *x*, the program with number e^* uses *x* to output a copy of itself - just like a self-replicating organism. More generally, for every recursive function $r(z)$, we obtain by applying the recursion theorem to the function $\psi(r(z), x)$ a number *e* such that for all *x*, $\psi(e, x) = \psi(r(e), x)$.

Kleene's enumeration theorem allows us to reformulate properties of partial recursive functions as number-theoretic properties of their codes. A set *P* of natural numbers is called an 'input-output property' (of unary partial recursive functions) if and only if, whenever e' is in *P* and e'' codes the same function as e' (in the sense that $\psi(e', x) = \psi(e'', x)$ for all *x* in their common domain), then e'' is also in *P*. The following theorem, 'Rice's theorem', states that non-trivial input-output properties are not recursive: if *P* is an input-output property such that neither it nor its complement is empty, then *P* is not recursive.

There are important non-trivial input-output properties of programs. For instance, the set of codes of Turing machines computing total functions is one such property; as a consequence of Rice's theorem we know that there is no Turing machine that decides whether or not a given Turing computable function is total. We will discuss additional undecidable problems in the next section; however, undecidability will be established there in quite different ways.

5 Undecidable problems

'Problems' are identified with (characteristic functions of) one-place predicates or, equivalently, with subsets of $\mathbb{N}$. Let *K* be the problem corresponding to the question 'Does the computation of machine *M* with code *e* terminate, if the input is the natural number *e*?'. This problem was formulated by Turing and is known as the 'halting problem' (see Turing machines §3). A diagonal argument (using the existence of a universal machine) settles the undecidability of *K* without resorting to Rice's theorem.

The halting problem is the progenitor of a multitude of undecidable problems, arising in virtually all fields of mathematics. Consider, for instance, the following problem. Let there be a presentation *P* of a group *G* via generators and relations (*P* looks like a multiplication table for *G*) and two expressions *v* and *w* built up from these generators, using the multiplication operation and its inverse. Do *v* and *w* represent the same element of *G*? This is known as the 'word problem for groups'. Novikov and, independently, Boone proved that the problem is undecidable.

Markov proved the undecidability of the following problem for four-dimensional manifolds. Consider a suitable combinatorial description of two manifolds. Are they homeomorphic? In the two-dimensional case the problem becomes decidable; it is not known whether the three-dimensional problem is decidable or not.

Yet another logic-free undecidable problem is given by elementary functions, that is, those functions *f* of a real variable *x* which can be built up using exponentials, logarithms, *n*th roots, trigonometric functions and their

inverses by addition, multiplication, and composition. While the derivative operation does not lead outside the realm of elementary functions, the integral of, say, $(\sin x)/x$ is not elementary. Indeed, Richardson proved that the following problem is undecidable: is the integral of an elementary function elementary?

Finally, the tenth problem in Hilbert's list presented at the International Congress of Mathematicians in 1900 asks for an algorithm to decide of a Diophantine equation - that is, a polynomial equation $p(x_1, \ldots, x_n) = 0$, with integer coefficients - whether or not it has integer solutions (see Hilbert 1900). Following important work by J. Robinson, Davis and Putnam, Matijasevic proved in 1970 that this problem is undecidable (see Matijasevic 1993).

Let us go back to the halting problem, *K*. Recalling the discussion in §3, it is clear that the predicate $K(e)$ can be defined by the statement 'There exists a *y*, such that $T(e, e, y)$'. *K* is thus an example of a 'recursively enumerable' (RE) set of natural numbers: it can be defined by prefixing one existential quantifier to a decidable binary predicate. Trying for each *n* in $\mathbb{N}$ all possible pairs (e, y) such that $e + y = n$ and singling out *e* whenever $T(e, e, y)$ is true, we can effectively list all members of *K*. Recursively enumerable sets were named because of the following equivalent characterization: a subset of $\mathbb{N}$ is RE if and only if it is either empty or the range of a (primitive) recursive function. Furthermore, a set *R* is recursive if and only if both *R* and its complement are RE. This latter fact allows a different route into computability theory: recursive enumerability is taken as the basic concept, and recursive sets are defined as those RE sets whose complements are also RE; this is done in Post's approach.

The importance of *K* among RE sets stems from its being 'complete', that is, maximally difficult among all RE sets; by this we mean that every RE set can be effectively reduced to *K*. A problem *P* is effectively reducible to *Q* if and only if there is a recursive function *f* such that for all natural numbers *x*, $x \in P$ if and only if $f(x) \in Q$. Thus, if we want to see that *x* is a solution of problem *P* we compute $f(x)$ and test whether $f(x)$ solves *Q*! Clearly, *Q* is at least as difficult as *P*, in the sense that effective solutions for *Q* yield effective solutions for *P*. By effectively reducing the halting problem *K* to the *Entscheidungsproblem* (decision problem) *E*, Turing concluded from the undecidability of *K* that also *E* cannot be decidable. Turing's reduction is nothing but a transcription of the ternary predicate $T(e, e, y)$ into a first-order arithmetical theory with sufficient demonstrative power.

The halting problem *K*, the most complex among RE problems, is the simplest in an infinite sequence of manifestly undecidable problems, the so-called 'jumps'. In order to obtain the jump hierarchy, the concept of computation is relativized to sets of natural numbers whose membership relations are revealed by 'oracles'. The jump K' of *K*, for example, is defined as the halting problem, when an oracle for *K* is available. This hierarchy can be associated in a most informative way with definability questions in the language of arithmetic: all jumps can be defined by increasingly complex arithmetical formulas, and all arithmetically definable sets are reducible to some jump. The above construction underlies the arithmetic hierarchy introduced by Kleene and Mostowski in the fifties.

Certain interesting sets of natural numbers are not definable by arithmetic formulas; one example is the set of all Gödel numbers of true arithmetic statements. If this set were arithmetically definable, one could formulate arithmetically the liar sentence that expresses its own falsity. This observation of Gödel and Tarski (see Gödel's theorems) is the cornerstone for proving the incompleteness of every formal theory of arithmetic that contains symbols for addition and multiplication and has semantically sound axioms and rules of inference: no matter which (true) statements we choose as axioms, and no matter which inference rules we adopt (provided they are truth-preserving) there are statements which, albeit true for the natural numbers, are not provable in the theory. Arithmetic truth can be defined, however, in second-order languages that allow quantification over functions.

6 Physical steps

Turing appealed in his analysis of mechanical calculability to the limitations of the human sensory apparatus; he claimed that the justification for his thesis lies ultimately 'in the fact that the human memory is necessarily limited'. This remark is not expanded on at all, and we can only speculate as to Turing's understanding of this 'fact': did he have in mind more than the spatial limitations for 'encoding' finite configurations? If not, the restrictions can be motivated by physical considerations. The question of whether there are ways of getting around the effect of such physical limitations is discussed with references to the literature by Mundici and Sieg (1995); and Mundici (1981).

Assume that a Kolmogorov machine operates on configurations containing *z* different symbols, each symbol being physically 'coded' by at least one atom - an altogether reasonable assumption. Then there must be at least *z* pairwise disjoint regions containing the codes (of the symbols). Otherwise, the electron clouds of two different codes might overlap, making the codes indistinguishable. Let c and a denote the speed of light and Bohr's radius of the hydrogen atom, respectively, so that $\mathrm{a}/\mathrm{c} = 0.176 \times 10^{-18}$ seconds. It follows that the codes will be contained in a volume of at least $z(4/3)\pi \mathrm{a}^3$ cubic metres; that forces the diameter of the machine to be larger than $2\mathrm{a}z^{\frac{1}{3}}$ metres - the diameter being the largest possible distance between two codes in this volume. Let *f* be the frequency of our machine; thus, $1/f$ is the time available for each computation step. Since signals cannot travel faster than light and since a computation step involves the whole configuration, it follows that *f* cannot exceed $\mathrm{c}/(2\mathrm{a}z^{\frac{1}{3}})$ steps per second. Thus, we obtain the inequality

$$fz^{\frac{1}{3}} < 2.828 \times 10^{18} \text{ steps per second},$$

which points out a fundamental incompatibility between high number of codes (that is, size of configurations) and high computational speed. The operations of Kolmogorov machines are thus restricted in complexity, as they have to lead from distinguished graphs to distinguished graphs; and within the given physical boundaries only finitely many different graphs are realizable.

If we focus on physical devices and analyse machine, not human mechanical, computability we have to take into account the possibility of parallel procedures as incorporated, for example, in cellular automata. Gandy (1980) provided the first conceptual analysis and a general description of parallel algorithms. These algorithms are thought to be carried out by 'discrete deterministic mechanical devices', that is, machines satisfying the physical assumptions explicit in our discussion above; in Gandy's words,

> The only physical assumptions made about mechanical devices…are that there is a lower bound on the linear dimensions of every atomic part of the device and that there is an upper bound (the velocity of light) on the speed of propagation of changes.
>
> (1980: 126)

He formulated axiomatic principles for these devices and proved that whatever can be calculated by devices satisfying the principles ('Gandy machines') is also computable by a Turing machine.

See also: Logical and mathematical terms, glossary of

DANIELE MUNDICI
WILFRIED SIEG

References and further reading

All the classic papers referred to in this entry have been reprinted in either van Heijenoort (1967) or Davis (1965); careful discussions of the history of the subject are given by Gandy in Herken (1988) and by Sieg (1994).

Cutland, N. (1980) *Computability - An Introduction to Recursive Function Theory*, Cambridge: Cambridge University Press.(An informative introduction to basic results.)

Davis, M. (1958) *Computability and Unsolvability*, New York: McGraw-Hill; repr. New York: Dover, 1982.(A classic, detailed presentation of the basic parts of computability theory.)

Davis, M. (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Hewlett, NY: Raven Press.(Anthology of the fundamental papers on the subject by Gödel, Church, Turing, Kleene, Rosser and Post.)

Davis, M. (1982) 'Why Gödel Didn't Have Church's Thesis', *Information and Control* 54: 3-24.(Provides fascinating historical information on the developments in Princeton in 1933 to 1936.)

Dedekind, R. (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; repr. in *Gesammelte mathematische Werke (Collected Mathematical Works)*, ed. R. Fricke, E. Noether and Ö. Ore, Braunschweig: Vieweg, 1932, vol. 3, 335-90; trans. W.W. Beman (1901), 'The Nature and Meaning of Numbers', in *Essays on the Theory of Numbers*, New York: Dover, 1963; trans. *What are Numbers and What Should They Be?*, Orono, ME: Research Institute for Mathematics, 1995.(*The* 'modern' set-theoretic foundation of number theory.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge,

MA: Harvard University Press, 1967, 1-82.(This small booklet includes the most significant step from Aristotelian to modern logic.)

Gandy, R. (1980) 'Church's Thesis and Principles for Mechanisms', in J. Barwise, H.J. Keisler and K. Kunen (eds) *The Kleene Symposium*, Amsterdam: North Holland, 123-45.(Axiomatic characterization of machine, not human mechanical, computability including parallel computations.)

Gödel, K. (1934) Notes of Princeton Lectures, in *Collected Works*, ed. S. Feferman *et al.*, vol. 1, *Publications 1929-1936*, New York and Oxford: Oxford University Press, 1986.(Quoted in §2. German originals and English translations.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879- 1931*, Cambridge, MA: Harvard University Press.(Anthology of fundamental papers in logic starting with Frege's *Begriffsschrift* (1879) and ending with Gödel and Herbrand.)

Herken, R. (ed.) (1988) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press. (This book includes informed, systematic, enlightening, historical and provocative essays by, among others, Kleene, Gandy, Davis, Feferman and Penrose.)

Hilbert, D. (1900) 'Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris 1900', *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen. Geschäftliche Mitteilungen* 253-97; repr. in *Archiv der Mathematik und Physik* 1: 44-63, 213-37, 1901; trans. 'Mathematical Problems', *Bulletin of the American Mathematical Society* 8: 437- 79, 1902.(The most influential list of open mathematical problems in the twentieth century.)

Hilbert, D. and Bernays, P. (1934, 1939) *Grundlagen der Mathematik (Foundations of Mathematics)*, Berlin: Springer, 2 vols; 2nd edn, 1968, 1970.(A most comprehensive presentation of the work in proof theory done in the Hilbert school and of related developments; for example, volume 2 includes the first detailed proof of Gödel's second incompleteness theorem.)

Kolmogorov, A.N. and Uspensky, V.A. (1958) 'On the Definition of an Algorithm', trans. in *American Mathematical Society Translations* 2 (21): 217-45, 1963.(This paper gives a very interesting analysis of 'algorithms'.)

Matijasevic, Y.V. (1993) *Hilbert's Tenth Problem*, Cambridge, MA: MIT Press.(A beautiful and thorough presentation of the solution to Hilbert's tenth problem, originally solved by Matijasevic in a paper of 1970, 'Enumerable Sets are Diophantine'.)

Mundici, D. (1981) 'Irreversibility, Uncertainty, Relativity and Computer Limitations', *Il Nuovo Cimento, Europhysics Journal* 61 B (2): 297-305.(Discusses physical limitations of computer steps, arising from relativity and quantum mechanical principles, such as Pauli indistinguishability.)

Mundici, D. and Sieg, W. (1995) 'Paper Machines', *Philosophia Mathematica* 3: 5-30.(A nontechnical essay on the development of the main ideas of computability, also covering physical computation.)

Neumann, J. von (1927) 'Zur Hilbertschen Beweistheorie' (On Hilbert's Proof Theory), *Mathematische Zeitschrift* 26.(This paper includes the central contribution of von Neumann to the development of proof theory: he proves the consistency of a fragment of number theory.)

Odifreddi, P. (1992) *Classical Recursion Theory*, Amsterdam: North Holland, 2 vols.(A comprehensive modern treatise on computability theory.)

Rogers, H. (1967) *Theory of Recursive Functions and Effective Computability*, New York: McGraw-Hill.(The standard reference on recursion theory before Odifreddi (1992) and Soare (1987).)

Sieg, W. (1994) 'Mechanical Procedures and Mathematical Experience', in A. George (ed.) *Mathematics and Mind*, Oxford: Oxford University Press, 71-117.(Examination of the conceptual analyses underlying computability theory, in particular in Turing's fundamental paper; presents the historical development in detail and emphasizes connections to investigations in the foundations of mathematics.)

Sieg, W. (1997) 'Step by Recursive Step: Church's Analysis of Effective Calculability', *Bulletin of Symbolic Logic*.(Describes in detail the approach Church took to the conceptual issues underlying his work on computability and undecidability. New historical material is taken into account to complement (and correct) the discussion in Davis (1982).)

Soare, R.I. (1987) *Recursively Enumerable Sets and Degrees*, Berlin and New York: Springer.(Brings the reader to the frontiers of research on computable functions, featuring Post's problem, oracles, priority methods and lattices of RE sets.)

Computer science

At first sight, computers would seem to be of minimal philosophical importance; mere symbol manipulators that do the sort of things that we can do anyway, only faster and more conveniently. Nevertheless, computers are being used to illuminate the cognitive abilities of the human and animal mind, explore the organizational principles of life, and open up new approaches to modelling nature. Furthermore, the study of computation has changed our conception of the limits and methodology of scientific knowledge.

Computers have been able to do all this for two reasons. The first is that material computing power (accuracy, storage and speed) permits the development and exploration of models of physical (and mental) systems that combine structural complexity with mathematical intransigence. Through simulation, computational power allows exploration where mathematical analysis falters. The second reason is that a computer is not merely a concrete device, but also can be studied as an abstract object whose rules of operation can be specified with mathematical precision; consequently, its strengths and limitations can be systematically investigated, exploited and appreciated. Herein lies that area of computer science of most interest to philosophers: the theory of computation and algorithms. It is here where we have learned what computers can and cannot do in principle.

1 Turing machines and computability

In his classic paper 'On Computable Numbers, with an Application to the *Entscheidungsproblem*' (1936) Alan Turing investigated the problem of characterizing the computable real numbers. The real numbers are those numbers that correspond to the points on a straight line: the positive and negative integers, the fractions or rational numbers, and finally, the irrational numbers, such as π and the square root of 2. Whereas the values of the integers and fractions can be specified using a finite set of symbols and computed exactly, the same is not true for the irrationals. Some of the irrationals like π or the square root of 2 can be computed to as many decimal places as we like; other irrationals, however, cannot be computed at all. Intuitively, the difference between the two kinds of numbers lies in the fact that the former can be determined to arbitrarily high degrees of precision, in Turing's words, 'using finite means', whereas to obtain an arbitrarily precise value of the latter sort of number would require infinite means.

Turing proceeds to give a precise account of what it means to compute 'using finite means'. Imagine a machine that at any given time is in one of a finite number of states $Q = \{q_0, q_1, \ldots, q_i\}$. At the same time, the machine is scanning a square of a finite but extendible tape on which is printed one of a finite number of symbols $S = \{s_0, s_1, \ldots, s_k\}$. The machine is capable of doing just three things. First, it writes a symbol on the square (possibly the same symbol that it is scanning). Next, it moves one square to the left or one square to the right. And finally, it enters a (possibly) new state. The key idea is that what the machine does is determined only by its current state and the symbol currently being scanned. Exactly what the machine is to do in a given situation is specified by a rule which is stated in the form of a 'quintuple' such as the following:

$q_2 \; * \; \# \; R \; q_3,$

which may be read, 'if in state q_2 and scanning the symbol *, then write the symbol #, go to the right and enter state q_3'. A Turing machine is formally defined as a set of such instructions, subject to the proviso that no two quintuples begin with the same two items. (This ensures that there is no ambiguity as to what the machine is to do when in a certain internal state and scanning a given symbol.) Often, the state q_0 is designated as the starting state of the machine, and the symbol s_0 is considered to be the 'blank' symbol *B*. A Turing machine halts when it is in a certain internal state q_i, scanning a symbol s_i, and there is no quintuple beginning with $q_i s_i$. (What to do, what to do…?) Using these conventions, and assuming that the machine is started on the right-most symbol of a finite string of 1s, then the machine described by the quintuples:

$q_0 \; 1 \; 1 \; L \; q_0$

$q_0 \; B \; 1 \; L \; q_1$

will (as the reader can verify easily) move to the left leaving all of the 1s intact until it finds the first blank ('*B*') which it will then replace by a 1, move one more square to the left, and then halt (see Turing machines).

Let us call the symbols $\mathbf{S}^* = \{s_1, s_2, \ldots, s\}$ the (proper) *alphabet* of the Turing machine. (Notice that the 'blank' symbol q_0 is hereby excluded.) The *words* of this alphabet are now defined as the set of all the finite strings of the members of the alphabet. A function on this alphabet is simply a map that takes any word to a unique word; that is, a map from the set *words* to itself. (For simplicity's sake, only functions taking a single argument will be considered here.) Such a function is said to be *Turing computable* just in case there is a Turing machine *M* which, when started on the right-most letter of a word, eventually halts leaving on its tape the string that is the function's value on the original word. Roughly, a Turing computable function is a function that can be computed by some Turing machine.

Of all the functions on an alphabet, how many are Turing computable? With the aid of Cantor's theory of infinite sets (see Cantor's theorem), we can answer this question fairly easily. Basically, since a Turing machine is determined by (and can thus be regarded as defined by) a finite set of quintuples, Turing machines can (in principle) be ordered 'alphabetically'. So there are only a countable infinity of Turing machines and hence at most a countable infinity of Turing computable functions (since each Turing computable function requires at least one Turing machine to compute it). But there are an uncountable infinite number of functions from (any) set of words to itself. Hence, although there are an infinite number of Turing computable functions, there are an uncountable number of functions that are not Turing computable. So to the question, 'of all the functions from words to words, how many are computable?', the answer is 'a countable infinity, that is, almost none of them'. Later, we shall examine an interesting function - the Halting function - that is not computable.

Although we have considered functions from and to the words of any finite alphabet of symbols, the same results may be obtained by confining our attention to functions from the set *N* of natural numbers $0, 1, 2, \ldots$, to itself. To see this, think of the symbols $\mathbf{S}^* = \{s_1, s_2, \ldots, s\}$ as the digits $\{0, 1, \ldots, i-1\}$. Then, corresponding to a finite string of letters in $\mathbf{S}^*$, we have a string of digits, which can then be regarded as representing a natural number using base *j*. In this way, the words of $\mathbf{S}^*$ become mapped to numbers in *N*, and functions from *words* to *words* become functions from *N* to *N*. This allows us to restrict our attention to the Turing computable functions from *N* to *N* without any important loss of generality.

2 The Church-Turing thesis

In work carried out prior to Turing's 1936 paper (presented to the American Mathematical Society in 1935), Alonzo Church (1936) had defined a class of computable functions from the set $N = \{0, 1, 2, \ldots\}$ of natural numbers to itself which he called the 'recursive' functions. The recursive functions are defined, roughly, as those that can be obtained from a few basic computable functions (in the sense that we have algorithms for them) by applying 'computable' operations such as functional composition or recursion. For example, one of the basic functions is the *successor* function which takes a number *n* to $n+1$. Another is multiplication. Composing these two functions we obtain $(n+1)m$. Church's basic functions, Turing showed, were Turing computable, and the operations for their combination could also be mimicked by Turing machines. For example, the function $mn+1$ can be computed by a Turing machine that chains together the Turing machines that compute *mn* and $n+1$. In this way, Turing showed that the recursive functions are all Turing computable. In addition, he also proved the converse: any Turing computable function from *N* to *N* is recursive. As a result, the Turing computable functions and the recursive functions are identical (Boolos and Jeffrey 1982, chs 7 and 8).

In his 1936 paper, Church had conjectured that the recursive functions were just the computable functions from and to the natural numbers *N*. In his paper of the same year, Turing also conjectured that the functions computable by his machines were also just the computable functions. Since the two sets of functions are identical, these conjectures have since been joined and are known as the *Church-Turing thesis*, namely,

Any computable function from N *to* N *is recursive, that is, Turing computable* (see also Church's thesis).

Let us see what reasons there are for holding this thesis to be correct.

First of all, notice that in the above statement of the thesis, the first occurrence of the term 'computable' has not been given a precise definition. The intuitive idea is that a function is computable if we can calculate its output using means that are of the same sort that we use to do calculations in arithmetic. But what, exactly, are these 'means'? We have examples galore, but no precise general account of what sort of procedures these examples

embody. We could, of course, simply take Church's recursive functions or Turing's machine to *define* just those procedures. But in that case, the Church- Turing thesis becomes a tautology.

So let us stick with 'computable' as a pre-analytic term and see where that gets us. Consider the following example. Suppose that we consider a new kind of machine model, very much like a Turing machine, except that instead of a one-dimensional tape, the new machine operates on a 'board' of two dimensions, a potentially infinite checkerboard. Such a *2D Turing machine*, as I shall call it, will have quintuples just like a Turing machine, except that the move instructions *R*, *L* are expanded to eight possibilities: *N*, *S*, *E*, *W*, *NE*, *NW*, *SE*, and *SW*. Imagine that we now write out a set of quintuples and proceed to use them in the same way that we previously used a Turing machine - say, to determine what results from putting the string !#* down on the board and starting at the eastmost symbol (*). Clearly, if we could show that such a machine would stop when provided with any such input, then we would agree that a function was thereby *computed* by this 2D Turing machine. In short, the same intuitions which assure us that any function which is Turing computable is *computable* in the ordinary sense - the uncontested converse of the Church-Turing thesis - also work here to yield the conviction that any function that is 2D Turing machine computable is *computable*. But if this be granted, it can now be seen that there is - at least in principle - a way to show the *falsity* of the Church-Turing thesis: produce a machine model, such as a 2D Turing machine, that clearly *computes*, and then prove that there are some functions that *it* can compute that a standard Turing machine cannot.

In this spirit (and perhaps just for the fun of it), there has been a great deal of work on developing alternative models of computation. The 2D Turing machine model given above, along with its *n*-dimensional counterparts, can be shown to be no more powerful than a standard linear tape machine. Adding finitely many tapes with their own read/write heads does not help either: any function computable by these enhanced machines can be proven to be computable (perhaps more slowly) by a lowly standard Turing machine. Other models that are not simple variations of Turing machines have also been explored, such as abacus or register machines, post systems and generalized digital computers. While these alternative models might usefully focus on one or more features of computation systems, they all are incapable of computing a function that cannot likewise be computed by a 'plain vanilla' Turing machine. As a result, although there is no way of proving the Church-Turing thesis to be correct, it is widely held to be so.

The Turing machines that we have considered so far have all been dedicated to some particular computational task. A *universal Turing machine* (UTM) would be capable of computing anything that a Turing machine can. In his 1936 paper, Turing showed how to construct machines of this type, and, roughly, here is how they work. A section of the tape of a UTM is set aside (perhaps surrounded by *s or some other symbol in its alphabet reserved for this purpose) for the program that the machine is to execute. This program will be another Turing machine's quintuples (probably suitably encoded). A suitable input is now provided in some input area of the tape, and the UTM passes back and forth from the program and input areas to the work area, performing the computation in accord with the program on its tape. When one considers that much the same sort of activity is precisely what undergraduates do when, in a purely mechanical way, they work out on paper the behaviour of a Turing machine provided by the instructor, then the existence of a universal Turing machine should not come as much of a surprise. What is more remarkable is that such a machine is described in some detail in Turing's 1936 paper. Later work has produced some very concise UTMs, such as Marvin Minsky's, which uses only four symbols and seven states (Minsky 1967: §14.8). Turing's basic idea - put the program on the tape (in the temporary memory, RAM) of a universal computer - was later taken up by von Neumann, and became one of the pillars of current digital computer architecture.

3 The Halting problem

As we have already seen, 'most' functions from *N* to *N* are not computable. Let us now look at a particular function that bears this affliction, the Halting function. Given a particular Turing machine or more generally, a particular program, it would be useful to be able to know whether that program will halt when provided with a given input. Of course, in some cases, the answer is obvious. The problem here is to find a general method, another program, that can compute the answer to the halting problem for *any* given input program and its input. Putting the matter in terms of Turing machines: is there a Turing machine that, when provided with a (suitably encoded version of a) program on its tape, together with an input to that program, will output, say, a '1' if the program will

halt on that input and a ‘0’ otherwise. (It is assumed here that, after delivering its verdict, the Halter machine itself then halts.)

It turns out that no such Turing machine exists. This result is sometimes referred to as the undecidability of the Halting problem. One reason this is so important is that it is the basis of showing that a number of other such problems are undecidable (by reducing them to the Halting problem); another reason is that it shows that there is a strong connection between the failure of computability and a wide class of unsolved problems in number theory. For example, in 1742 Goldbach conjectured that every even number (greater than 2) can be expressed as the sum of two prime numbers; to this day, his conjecture has been neither proven nor disproven. Now it is a simple matter to write a program that will check an even number to see if it is expressible as the sum of two primes, and do this for successively greater even numbers. If such a program were to find an even number not equal to the sum of two primes, then it could be designed to halt and print out some message like ‘Goldbach has been refuted!’. On the other hand, if Goldbach was right, then the program will go on checking even numbers indefinitely, and thus never halt. If the Halting problem were decidable, then such problems - and there are a great many such problems - could (in principle, at least) be settled according to the Leibnizian saying: ‘Let us compute!’ (For a survey of results on uncomputable functions, see Harel 1992, ch. 8.)

4 Problems that are computationally intractable

Even when we consider only computable functions, there are some problems whose solutions will probably forever remain beyond our reach. A classic example is the travelling salesperson problem. Imagine a set of *N* towns laid out on a map, together with the various roads connecting these towns. The length of each road is given. The problem now is to find the shortest itinerary for the salesperson to take which starts at a given town, returns to that town at the finish, and visits every other town exactly once. Clearly, the solution is computable: enumerate each of the possible paths, computing and recording its length along the way, and return as the answer the path requiring the minimum total distance. The problem is that the number of steps required to execute this algorithm goes up drastically as the number of towns involved gets larger. Thus, for 100 towns, the number of steps required by the best current programs is greater than 2^{100}, a very large number. (Assuming our algorithm executes one million steps per second, such an algorithm would require over 10^{16} years to complete its task; the accepted upper limit on the age of the universe is, however, only around 15 billion, that is, 15×10^9 years.)

The travelling salesperson problem is only one of a wide class of problems, called *NP complete* problems, that require, as far as we know, solutions on the order of *N* steps. (The reason that *NP* complete problems are so-called may be found in Harel 1992, ch. 7.) If any one of the problems in this class could be shown to be solvable by an algorithm that was, say, linear in *N* or polynomial in *N*, then the same would be true for all the problems in this class. While it is widely believed that the *NP* complete problems require algorithms that work in exponential time, this has not yet been proven, and remains one of the most difficult unsolved problems in the theory of computation.

Problems having an input of size *N* that *require* 2^N or more steps for their solution are said to be *intractable*. While there remains some (dim) hope for the tractability of the *NP* complete problems, there are many other problems that are provably intractable, and thus require algorithms operating in exponential time or worse. For example, a complete examination of the game tree of checkers could determine whether or not the side that moves first has a guaranteed winning strategy. Consider now generalized checkers, played not on an 8×8 board, but on an $N \times N$ board. It has been shown that the problem of determining the existence of a guaranteed first-mover winning strategy is intractable in *N*; that is, there is no algorithm for solving this problem which does so in fewer than K^N steps $(K > 1)$. This is not the end of it, however. Some problems can be shown to be solvable in at best doubly exponential time 2^{2N}, or even worse. Similar results have been established for problems that involve unreasonably large amounts of memory space (Harel 1992, ch. 7).

The situation is not altogether bleak. There are tractable algorithms for solving *NP* complete problems that give correct or nearly correct answers most of the time. For example, there are tractable algorithms designed to tackle the travelling salesperson problem that routinely come up with paths that are surely far shorter than most routes. Another possible way around intractability is an algorithm that generates the correct answer (say, to a decision problem) with a certain high probability, regardless of the size *N* of the input. More precisely, suppose that we can show that if the answer to our question is truly ‘yes’, then there is an algorithm that will output ‘yes’ with a

probability $P > K$, regardless of the size *N* of the input; on the other hand, if the correct answer is 'no', then our algorithm will quickly yield that result. Clearly, such an algorithm is almost as good as one that provides answers whose correctness can be guaranteed.

One such probabilistic algorithm was presented by Solovay and Strassen (1977). It is designed to test an arbitrary natural number for compositeness. (Does the number have nontrivial factors?) Number theorists have been able to show that if a number is composite (that is, not prime), then over half of the numbers less than that number will have a quickly computable property *W* that will be a conclusive witness to that fact. Thus if a number is not composite, no such witness will exist. A candidate *N* for compositeness is now input into the algorithm and a number less than *N* is chosen at random. If the randomly chosen number has the 'witness' property *W*, then the original number *N* must be composite, so the program outputs 'Composite!' and halts. On the other hand, if the randomly chosen number does not have the witness property *W*, then the algorithm chooses another number less than *N* at random. If 100 or so random choices fail to turn up a witness to *N*s compositeness, then the program outputs 'Not composite!'. A little thought shows that since over $1\colon 2$ of the numbers less than *N* will have *W* if *N* is composite, then if we fail to find such a witness after 100 random trials, the chances of *N*s being composite are less than $1\colon 2^{100}$, a very small chance indeed. Provided we are willing to take a small risk of error, this method results in a very fast probabilistic algorithm for testing compositeness.

Whether the existence of such an algorithm shows that probabilistic algorithms can tame intractable problems is, however, still unclear. The hitch is that the compositeness problem is not *known* to be intractable. Furthermore, there are no cases of problems *known* to be intractable that have succumbed to a probabilistic algorithm (although some *NP* complete problems have been tamed). The issue of whether or not there are probabilistic algorithms that can solve (provably) intractable problems remains open. It *has* been shown, however, that probabilistic algorithms cannot tame problems that are not computable, a result that follows from the fact that any probabilistic algorithm can be simulated deterministically.

The theory of deterministic algorithms has also contributed to our understanding of what apparently stands in direct opposition to determinism: randomness (see Randomness §4).

See also: Complexity, computational; Computability and information; Computability theory; Turing reducibility and Turing degrees

JOHN WINNIE

References and further reading

Boolos, G.S. and Jeffrey, R. (1974) *Computability and Logic*, Cambridge: Cambridge University Press; 2nd edn, 1982.(A classic, easy to read text on computation theory and its relation to logic.)

Chaitin, G. (1969) 'On the Length of Programs for Computing Finite Binary Sequences: Statistical Considerations', *Journal of the Association for Computing Machinery* 16: 145-59.(Chaitin's improvement on the Kolmogorov definition of algorithmic complexity.)

Church, A. (1936) 'An Unsolvable Problem of Elementary Number Theory', *American Journal of Mathematics* 58: 345-63; repr. in M. Davis (ed.) *The Undecidable*, New York: Raven Press, 1965, 89-107.(Church's original paper on the lambda calculus approach to the theory of computation.)

Davis, M. (ed.) (1965) *The Undecidable*, New York: Raven Press.(A sourcebook of classic papers on computation theory.)

Davis, M., Sigal, R. and Weyuker, E. (1983) *Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science*, Boston, MA: Academic Press; 2nd edn, 1994.(A handbook, perhaps *the* handbook, of the current state of computation theory. A work of moderate technical difficulty.)

Garey, M. and Johnson, D. (1979) *Computers and Intractability*, New York: W.H. Freeman.(A detailed but only moderately technical account of the theory of computational intractability.)

Harel, D. (1992) *Algorithmics*, Reading, MA: Addison-Wesley.(An excellent, easy to read survey of the current state of computation theory.)

Kolmogorov, A.N. (1965) 'Three Approaches to the Quantitative Theory of Information', *Problems in Information Transmission* 1: 1-7.(The paper in which Kolmogorov introduces algorithmic measures of complexity.)

Li, M. and Vitanyi, P. (1993) *An Introduction to Kolmogorov Complexity and its Applications*, New York: Springer.(A thorough, fairly technical, account of Kolmogorov and Chaitin's ideas on complexity.)

Minsky, M. (1967) *Computation: Finite and Infinite Machines*, Englewood Cliffs, NJ: Prentice Hall.(A lively, easy to read survey of computation theory as things stood in the early 1960s.)

Solomonoff, R.J. (1964) ‘A Formal Theory of Inductive Inference, Part 1 and Part 2’, *Information and Control* 7: 1-22, 224-54.(The first of the classic trio of papers (Solomonoff, Kolmogorov, Chaitin) on the foundations of algorithmic information theory.)

Solovay, R. and Strassen, V. (1977) ‘A Fast Monte-Carlo Test for Primality’, *Society for Industrial and Applied Mathematics Journal of Computing* 6: 84-5.(An early probabilistic algorithm for testing a number for primality, discussed in the text.)

Turing, A. (1936, 1937) ‘On Computable Numbers, with an Application to the *Entscheidungsproblem*’, *Proceedings of the London Mathematical Society*, series 2, 42: 230-65; corrections, 43: 544-6; repr. in M. David (ed.) *The Undecidable*, New York: Raven Press, 1966, 116-54.(The classic paper on Turing machines: difficult, brilliant, containing many errors.)

Comte, Isidore-Auguste-Marie-François-Xavier (1798-1857)

The French philosopher and social theorist Auguste Comte is known as the originator of sociology and 'positivism', a philosophical system by which he aimed to discover and perfect the proper political arrangements of modern industrial society. He was the first thinker to advocate the use of scientific procedures in the study of economics, politics and social behaviour, and, motivated by the social and moral problems caused by the French Revolution, he held that the practice of such a science would lead inevitably to social regeneration and progress.

Comte's positivism can be characterized as an approach which rejects as illegitimate all that cannot be directly observed in the investigation and study of any subject. His system of 'positive philosophy' had two laws at its foundation: a historical or logical law, 'the law of three stages', and an epistemological law, the classification or hierarchy of the sciences. The law of three stages governs the development of human intelligence and society: in the first stage, early societies base their knowledge on theological grounds, giving ultimately divine explanations for all phenomena; later, in the metaphysical stage, forces and essences are sought as explanations, but these are equally chimerical and untestable; finally, in the positive or scientific stage, knowledge is secured solely on observations, by their correlation and sequence. Comte saw this process occurring not only in European society, but also in the lives of every individual. We seek theological solutions in childhood, metaphysical solutions in youth, and scientific explanations in adulthood.

His second, epistemological law fixed a classification or hierarchy of sciences according to their arrival at the positive stage of knowledge. In order of historical development and thus of increasing complexity, these are mathematics, astronomy, physics, chemistry, biology and sociology. (Comte rejected psychology as a science, on the grounds that its data were unobservable and therefore untestable.) Knowledge of one science rested partly on the findings of the preceding science; for Comte, students must progress through the sciences in the correct order, using the simpler and more precise methods of the preceding science to tackle the more complex issues of later ones. In his six-volume Cours de philosophie positive (The Positive Philosophy) (1830-42), Comte gave an encyclopedic account of these sciences, ending with an exposition of what he regarded as the most advanced: social physics or 'sociology' (a term he invented). The sociologist's job would be to discover the laws that govern human behaviour on a large scale, and the ways in which social institutions and norms operate together in a complex yet ultimately predictable system.

In his later work, Comte fleshed out his vision of the positive society, describing among other things a Religion of Humanity in which historical figures would be worshipped according to their contribution to society. Despite such extravagances, however, the broader themes of his positivism - especially the idea that long-standing social problems should be approached scientifically - proved influential both in France and, through J.S. Mill's early support, in England.

1 Life

Auguste Comte was born in Montpellier, France. He attended the École Polytechnique, from which he was expelled in 1816, for political reasons. Comte's main concern throughout his life was resolving the political, social and moral problems caused by the French Revolution. To that end, he embarked upon an encyclopedic work, which he first conceived under the inspiration of Henri de Saint-Simon, for whom he worked as secretary from 1817 to 1824. At that time, he proposed several plans for a competition to create an encyclopedia modelled on that of Denis Diderot, but designed to bring together the 'positive ideas' of the period, that is, ideas conceived in their relation to modern science and free from the bonds of traditional theology and metaphysics. Comte's encyclopedic project developed into the famous *Cours de philosophie positive (Course in Positive Philosophy)* (1830-42), a complete system of philosophy in six volumes which aimed to provide the foundations for political and social organization in modern industrial society. Meanwhile, he wrote a series of minor works in social philosophy, which became known as the 'opuscules'. The third, *Plan des travaux scientifiques nécessaires pour réorganiser la société (Plan of the Scientific Operations Necessary for Reorganizing Society)* (1822), which is often called 'the fundamental opuscule', presented the first outline of the concepts which would become central to Comte's positivism - the 'law of three stages' and the classification of the sciences (see §3 below).

While pursuing his intellectual career, Comte earned his living first as a mathematics tutor at the École

Polytechnique, then as admissions examiner at the same school. When he lost the latter job , he was forced in 1848 to seek financial support from his disciples in order to survive. All his life Comte regretted his failure to be appointed a tenured professor; he accused François Arago, among others, of deliberately blocking his academic career. He was also unsuccessful when he requested the creation of chairs at the Collège de France: in 1832 the chair of the General History of the Physical and Mathematical Sciences, and in 1846 the chair of the General History of the Positive Sciences.

Comte always linked his theoretical research to the practical aim of moral, social and political reorganization. He considered this reorganization from the theoretical aspect of political science - 'sociology' or 'positive politics' - and from the practical aspect of the union of the social classes. Moreover, it was for the benefit of the proletarians that he gave public courses in popular astronomy between 1831 and 1848. He considered these lectures the prelude to the reorganization of the intellectual system, which, thanks to the scientific knowledge of society, would finally result in 'a politics finally freed from the arbitrary and the utopian' (Larizza-Lolli 1993: 76). Comte wrote the *Discours sur l'esprit positif (A Discourse on the Positive Spirit)* (1844) as a basic introduction to this reorganization. The *Discours sur l'ensemble du positivisme (A General View of Positivism)*, which appeared in 1848, derived from a similar intention, that of presenting to the workers a synthesized vision of positivist philosophy. From 1847, Comte devoted his lectures to the refutation of communism.

The epistemological foundation of Comte's political project tends to be problematic. For Comte, 'epistemology' involved the history, philosophy and methodologies of the sciences, as well as their basic concepts and theories. His positivist politics is based, on the one hand, on historical and logical law, the law of the three stages (theological, metaphysical and positive stages) and, on the other hand, on an epistemological law: the classification or hierarchy of the sciences, fixed according to the order of their arrival at the positive stage of knowledge (mathematics, astronomy, physics, chemistry, biology, sociology). Thus, as a fundamental epistemological concept, 'positivism' designates an intellectual attitude founded on the practice of rational and experimental scientific methods. At the same time, Comte's scientific positivism is a philosophy of the positive sciences, though initially it is also a philosophy of scientific creativity. Comte allowed certain new sciences that had not yet been fully regularized to be classed as positive: for instance, he confirmed the advance of biology to the level of a positive science. The most advanced of all positive sciences - social physics or sociology - allowed for the transformation of society based on the progress of the sciences.

Under the aegis of his Religion of Humanity, first described in the sixth volume of the *Cours de philosophie positive (The Positive Philosophy)* (1842), Comte articulates his vast political project in the *Système de politique positive (System of Positive Polity)* (1851-4). In this work, he gives his project a social and moral goal, founding it on altruism - a term he originated, according to Emile Littré. Comte's last book was *La Synthèse subjective (Subjective Synthesis)* (1856). Never completed, it re-evaluates the role of the sciences from the perspective of a positivist education. Comte died, probably of stomach cancer, on 5 September 1857.

2 Speculation and action

Following the example of Francis Bacon, who stressed the efficacy of combining knowing (*scire*) and doing (*posse*), Comte held that power is proportionate to knowledge: 'From science, comes foresight; from foresight, comes action' ([1830-42] 1975 I: 45). Even though he clearly separates theory and practice in the scientific study of phenomena, Comte orients his positive philosophy towards a constant interrelationship between speculation and action, while keeping it equidistant between rationalism and empiricism.

The pragmatic aspect of Comte's thought goes back to ideas he first expressed in 1816, while still a student at the École Polytechnique. He affirmed the necessity of pragmatism in a letter of 28 September 1819 to his best friend, Pierre Valat, declaring that he could not conceive of a scientific work that would have no useful goal for humanity. Conversely, he added that political research would have to be intellectually challenging in order to interest him; otherwise, it would have no validity in his eyes. This also explains why he underlined the necessity of basing political studies on a scientific foundation.

On the one hand, Comte was interested in a discipline that would soon be known as 'epistemology'. For him, it encompassed the history and philosophy of the sciences, the various methodologies which they used, fundamental scientific concepts and theories, and what he as well as Bacon called 'primary philosophy', which constituted the

synthesis of the regular means of scientific knowledge and their principal, universally valid results. His epistemological concerns later led Comte to establish a 'Table of the Fifteen Universal Laws' in the fourth volume of the *Système de politique positive* (see §§2, 3 below), settling three groups of universal laws about (1) laws formation, (2) static and dynamic theories of understanding, and (3) movement/existence, action/reaction classification and relation. Yet he never worried about defining 'facts', 'scientific verification' or 'observation'. Nor did the problem of the origin of knowledge concern him. Robert C. Scharff rightly emphasizes Comte's insistence 'that such issues [could] not even be understood as issues without recourse to philosophy's history' (1991: 193).

On the other hand, Comte was very conscious of the political, social and cultural problems posed by the French Revolution. As the third opuscule clearly states, he placed himself politically between 'the people' and 'the kings' (1970a: 57), between the 'retrograde' ideas of the latter and the 'critical' ones of the former; in effect, he had rebelled against the royalism and Catholicism of his parents, but at the same time opposed the destructive spirit of the revolutionaries. Similarly, he did not favour the mixture of the retrograde and critical currents advocated by the 'doctrinaires' such as Pierre Royer-Collard, François Guizot, Charles de Rémusat, and Prosper Duvergier de Hauranne, whom Comte grouped in the 'stationary school' (1970a: 69; 1975 II: 43), characterized by 'organic emptiness'. From 1822, after the fading of the spiritual power of the Church, Comte reproached the French body politic for neglecting to create an organization with analogous power. In spite of French politicians' efforts to recast a temporal power, they had, according to Comte, also failed here. He pointed to the absence of any explicit positive theory as proof of the double failure of French politics.

Before his encyclopedic enterprise was sufficiently developed to become the *Cours de philosophie positive*, Comte, in his role of secretary to Saint-Simon and with his approval, edited some research projects which he called 'Programmes'. One example is the *Programme d'un travail sur les rapports des sciences théoriques avec les sciences d'application (Programme of work on the relations between the pure and applied sciences)*, which he published in 1817. From 1817 to 1820, he began a certain number of preparatory works, which were at least initially encouraged, if not conceived, by Saint-Simon. They would continue after 1820, but in a totally different form.

Just as in the fundamental opuscule of May 1822, Comte resolutely insisted on linking his theoretical work to the practical goal of social reorganization in the *Cours de philosophie positive*, the two *Discours* of 1844 and 1848, the *Système de politique positive* and his last writings. Very early, in fact, he had made a definitive and fundamental observation that the development of human intelligence was closely tied to the history of societies, which he formulated as follows: 'intelligence arrives at a higher stage of development when altruism itself is more developed' (1851: 693). It closely connected the development of intelligence with that of the individual's interest in his or her peers, intelligence and altruism both being clear signs of the progress of humanity.

Comte considered the theoretical discipline of sociology to be concerned with the real nature of humanity, and so he gave a higher status to its method, which was originally objective, defining it instead as 'subjective', that is, as situated beyond objective cosmological observation. The objective method employed in the *Cours de philosophie positive* had permitted the passage from the world to humankind. Now it was necessary to go from humankind to the world; hence the designation of this method as 'subjective', that is, as dependent on the real nature of humankind.

3 Comtean epistemology

Comte's epistemology focused on considering and evaluating the social phenomena that the fundamental sciences recognized as subject matter for a 'positive science' - a science based on observation with a legitimate place in the system of recognized sciences. His epistemology reverses the accepted hierarchy of the sciences - astronomy, physics, chemistry and biology, all of which were traditionally subordinated under mathematics - and places them under the domination of sociology. Comte thought that once the principle of sociality was accepted by scientists, the new science should be called 'anthropology'; he presents this in the *Discours sur l'esprit positif* (1844), where he defines the underlying nature of the scientific approach as necessarily social. First, a 'semiotic principle' justifies it as social, because it originates in the constitution of the different systems of signs produced by society, especially the society of scientists. Second, it is social because it obeys a 'principle of homology' (Kremer-Marietti 1980: 53-69), which applies to all relations and then expresses the possibility of applying the

scientific norm of theoretical unification - a norm that Comte shared with Henri Poincaré. The principle of ‘classification’, which would also be recognized by Pierre Duhem, is justified by the principle of homology; the latter belongs logically to a more general thesis (which Poincaé would later formulate) according to which ‘it is illegitimate to take into consideration a single isolated hypothesis in order to verify it’ (Duhem 1914: 393; Kremer-Marietti 1992: 372). In Comte’s case, it gives rise to the conception of the ‘Table of the Fifteen Universal Laws’; the theoretical unity achieved by Comtean epistemology thanks to this table is due to the organization of all laws through a classification of scientific facts. Comte distrusted a logic isolated from science, a fact isolated from theory, a doctrine isolated from method, and a method isolated from its object. It was necessary to have a theory in order to locate the observations which, if isolated, would have no scientific value. Comte never stopped saying that the ‘crudest’ phenomena explain the ‘noblest’ ones.

Oddly enough, without this holistic type of explanation ordering the homology of the concepts of the world and man, abstract morality could not gain access to the positive system. For Comte, the reference to a system obeying the principle of homology was a universally practicable theoretical necessity - a criterion of verification which he could use to assess the value of experiments and conclusions.

In his ‘Oral Course in Positive Philosophy’, given from 1826, Comte developed his general review of the positive sciences, which he subjected to the historical and logical law of three stages. His teaching plan followed the classification or hierarchy of the sciences, which he saw as a fertile classificatory model. Henceforth he would elevate this model to the status of a key or grid of concepts. Not only did he depend on it to determine, according to the law of three stages, the historical and logical necessity of sociology, a new discipline dealing with social, historical and political points of view, but even more important, this classification permitted him to proceed in certain areas where observation was not yet possible. It is on this principle of classification that Comte established his ‘cerebral theory’ (see §4 below). He also drew the necessity of ‘abstract morality’ (the seventh fundamental science) from his view that the order of the individual, concerned with abstract morality, is at the very heart of sociology, where it is subordinate to the social order, exactly as the social order is subordinate to the vital order and as the latter in its turn is subordinate to the material order. In Comte’s eyes, these successive subordinations do not exclude either the specificity or the originality peculiar to the different orders of phenomena and their respective sciences. The seventh degree of the series of sciences ‘arrives’, according to Comte, ‘at man envisaged in the most precise fashion’. Because the double weight of the material and vital orders is in reality borne by the social order, which helps to modify them, the ‘order of the individual’, the most dependent of all, becomes the ‘immediate regulator of our destinies’ at the same time that it experiences the pressure of all the orders through the social order, to which it is subordinate.

4 The logic of systems and the semiotic principle

Comte highlighted a history of human logics by emphasizing the necessary and reciprocal connection between the systems and institutions of signs, which are languages, and social systems. He saw social history as being affected by the systems of logic which he recognized as existing at three levels: the ‘logic of sentiments’, the ‘logic of images’ and the ‘logic of signs’. This tripartition into sentiments, images and signs contains the semiotic elements that we can recognize in Charles Peirce’s semiotic as Firstness, Secondness and Thirdness (Peirce 1966: 241) (see Peirce, C. §7).

Language appeared as a dynamic system relating political life to domestic life. It allowed for the creation of a positive science whose basis was observation but whose ideal reference was mathematical language, which arose from the logic of signs. Comte’s semiotic position explains his mathematical views. He called attention to the fact that physical science depended as much on observation guided by theory as on the proper use of a model of language, which, to him, was the language of mathematics.

As early as 1819, Comte maintained a general theory of language and signs in his *Essais sur la philosophie des mathématiques (Essays on the Philosophy of Mathematics)*. This logic of the systems of signs in both science and aesthetics was clearly confirmed in 1852 in the second volume of *Système de philosophie positive*. Written long after the first and long before the second of these works, the *Cours de philosophie positive* contains an epistemology concerned with revealing the scientific means of demonstrating the final necessity of the human and social sciences. These sciences are ruled by a logic of systems, of which the law of three stages is the determining operational element.

The ‘positive logic’ of the *Système de philosophie positive* is nothing other than a combination of systems of signs (affective, imaginary and intellectual), subjected to the unification of the three systems governing sentiments, images and linguistic or mathematical symbols. Comte wanted every ‘logic of signs’ to be expressly connected with the ‘logic of images’ and the ‘logic of sentiments’. He reaffirmed the semiotic principle in 1856, in *Synthèse subjective*, and this led to a complete formulation showing how ‘the regular conjuncture of sentiments, images and signs’ was able ‘to inspire in us conceptions that meet our moral, intellectual and physical needs’ (1856: 27).

The logic of the systems of signs arises from ‘the interior function of language’ because, according to Hobbes, people communicate by signs, with one phenomenon becoming the sign of another phenomenon. The Table of the Fifteen Universal Laws may be considered the outcome of this logic. The reason is that this table achieves an intellectual unity through its reduction to a precise and definite number of signs, which are specific laws, valid for all phenomena.

Evidently, there is a link between the three Comtean logics and the three-stage law and an even closer link between these logics and the three sub-states of the theological stage. In fact, Comte thought that fetishism, polytheism and monotheism were at the origin of these three logics, whose development was closely connected with the system of society. The logic of sentiments originated in fetishism, the logic of images was created under polytheism, and the logic of signs (linguistic and mathematic) arose from monotheism. It is clear that this semiotic principle has important implications for Comte’s epistemology as well as for his theory of language and art.

5 The moral and political plan

The opuscules of his youth, appended to the fourth volume of the *Système de philosophie positive*, demonstrate Comte’s constant interest in moral and political reconstruction, in which semiotics played a role. As mentioned above, semiotics was connected with the law of three stages and the three theological sub-states. In Comte’s eyes, fetishism was superior to the other two sub-states because it founded human language. For Comte, the ‘fetishist thinker’ was closer to phenomenal reality than the ‘theologicist dreamer’. Yet the theologians taught us to consider purely ideal existences, without which we would not have been able to create the scientific realm. In addition, without monotheism, abstraction would have been impossible.

It is important to remember that mental states have their ontological foundation in the corresponding social state. It is impossible, for example, to speak abstractly of a society that embraces a specific intellectual and mental state. Thus man and society advance in correlation with the development of intelligence. As a result, the positivist project for society studies the different forms of learning, from the most ancient to the most advanced, by means of knowledge of the various forms of civilization.

The study of ‘social statics’ brings out the principle of solidarity that is crucial for the coexistence and cohesion underlying all social systems. The study of ‘social dynamics’ highlights the principle of the continuity and tradition of societies. In brief, scientific positivism demands that we act so that theory and practice are in harmony. This harmony in turn generates the harmony between the knowledge of the environment that surrounds us and our reaction to this environment.

If knowledge leads to power over the outside world, our industrial society should be able to realize an equitable political project only by developing an all-embracing moral project. If Comte had limited himself to considering the use of abstract science in the ‘sciences of application’, or the results of the latter on industry, the industrial activity arising from positive science would not suffice to constitute a project for society. Because of his familiarity with all scientific areas, Comte turned to a complete knowledge of man; he contended that abstract morality, the ‘final science’, could ‘systematize the special knowledge of our individual nature by combining the biological point of view with the sociological’ (1851-4: 2, 438). To this end, Comte provided the essential epistemological basis with the ‘cerebral chart’, established in 1851 from his method of classification. ‘Abstract morality’ usually precedes ‘concrete morality’; from the ‘cerebral chart’ emerges the principle of man’s action in society: ‘Act out of affection, and think in order to act’ (1851-4: 1, 680-733). The phrenology of Franz Josef Gall, which he criticized for separating the brain from the whole nervous system and detaching the individual from his social milieu, permitted Comte to maintain the innateness of human qualities and the importance of the emotions, as well as the distinction between the mind and the heart: their harmony constitutes the soul. For Comte, the rule of positive society will be love, a basic feature of humanity and thus something which it is possible to expect human

beings to develop.

6 The human sciences and the standard conduct of nations

A fundamental question remains: How does Comte's epistemological system make it possible for him to create a political project?

The analysis of the human psyche (§5 above) showed that the need for love transcends natural egoism by means of altruism. The cerebral chart claimed that both speculation and action are dominated by affection and that social consensus depends on the affective life. Comte, who recognized the permanence of affectivity, expressed it by the phrase, 'One tires of thinking and even acting; one never tires of loving', a progressive division in intermediate items between extreme terms. Using a process of binary decomposition, Comte was able to derive the scale of all the intermediary affections between complete egoism and pure altruism. His analytical method was inherent in the taxonomic method that he usually used. It permitted him to maintain that altruism 'when it is energetic, is always better able than egoism to direct and stimulate the intelligence, even among the animals' (1851-4: 1, 693). The cerebral theory thus represents the fundamental order of our natural make-up and can be applied to social existence.

Social statics deals with the play of the permanent social forces revealed by Comte. What is the material power of Western society? Comte answered that it is 'number' (the proletariat) and 'wealth' (the administrators of capital). What is its intellectual power? It is the aesthetic and scientific spirit (its representatives in society). What about sentiment? Sentiment is concerned with the heart, which is both masculine and feminine, for man and woman share masculine and feminine traits. Even if Comte considers women intellectually inferior to men, he recognizes in them a superiority in terms of sensibility and makes them the guardians of universal morality; yet women remain auxiliary to men. No doubt it was from his romantic love for Clotilde de Vaux, which lasted a year (1845-6), that Comte obtained his experience of the feminine condition. Theoretically, Comte applies an Aristotelian principle governing the different forces of society: the double principle of the separation of functions and the combination of efforts.

Property, family and language are the necessary elements of social statics, which owes its unity to 'religion'. To Comte, religion is a synthesis of 'dogma', which represents the philosophical unity of scientific theories; 'worship', which directs sentiments; and 'regulations', which govern behaviour. Worship and regulations form the subjective domain of love, which is subordinated to the objective realm of philosophical dogma. Social dynamics can be summarized in the law of three stages; it is the permanent substratum of human action. The harmony between the social static and the social dynamic expresses itself in the principle of progress, which is conceived as the development of order. The postulate of harmony is evoked in the perspective of that which governs and rallies, that is, religion, which is destined to 'bind the interior and to link it to the outside'.

The *Cours* deals with the creation of the system of scientific systems as does the *Système de philosophie positive*, which also proposes the new construction of a political synthesis inspired by religion. This completes the human unity to which the synthesis tends through loving, thinking and acting. Religion thus becomes a super-theory of the immediately applicable unity; it permits human intervention in the historical and social dynamic, for it puts morality and politics in the service of social progress. The Religion of Humanity is 'proven' because it is founded on cosmological and human knowledge, and is thus the only answer to moral and political questions. Civil society cannot answer these questions even if its mechanisms permitted it, and it would in any case be unable to set its solutions in motion, since it is nothing but a battleground for divergent opinions. The same criticism is found in the works of Hegel, who limits the civil objective to the satisfaction of needs. While the Hegelian state is called upon to transcend the egoistic civil society by an objective moral idea, Comte wants to orient the will towards the superior reality of humanity, a subjective moral idea. For 'Humanity breaks up first into Cities, then into Families, but never into individuals' (1851-4: 4, 31). Morality takes the individual into consideration; families and homelands are, nevertheless, still important to it as the necessary introduction to Humanity. In terms of its composition, the Great Being is defined as 'the continuous totality of converging beings'.

Comte envisages a future where positive morality determines the discipline of existence. His concept of altruism recalls the principle of friendship or philia, which Aristotle made the cement of the ancient cities. While earlier Comte assigned temporal power to the proletarians, he later entrusts it to those who favoured the 'universal union'.

From 1842, he foresees that the spiritual power has to be 'the true philosophical class'; he therefore gives it the title of 'Western Positive Committee' and already imagines it to be a 'Positive Church'. Comte intends it to direct the intellectual and moral regeneration of modern societies, with the assistance of the proletariat, at least after the latter has ended the transitional dictatorship towards positivism. Just as with all women (who do not form a class), the proletariat is the servant of the spiritual power; it watches over the temporal power to see that it respects the general principles of the social regime.

As for the positive economy, Comte hoped that it would be nothing other than the 'universal and continuous systematization of human toil'. He wanted work to be systematized for the sake of future generations. Capitalization supported the sociocracy, which, according to the evidence of those who recognized its inner workings, involved the collaboration of the social classes based on the model of the functions and organs of the body. (Sociocracy was related to regulation, while sociolatry dealt with worship. Both were founded on sociology, which served as their dogma.) Comte's positivist economic theory was summed up in two laws: first, that each man can produce more than what he consumes, and second, that the products obtained can be preserved beyond the time required for their reproduction. The institution of capital is justified by the preponderance of human work over consumption. Comte's detractors, such as Herbert Marcuse, did not understand his view of the origin and social destination of capital funds. Yet Comte maintained that these capital funds made each active citizen the agent of all the others: each one had to function above all for others by applying the positivist motto 'Live for others'. Others' obligations emanated naturally from the social consensus.

Once the altruistic sentiments were recognized and generalized, Comte thought it possible to treat the three principal obligations - moral, judicial and civic - as one. But there was also an economic obligation to work, and to sustain the products of that work as long as possible; for the importance of solidarity among people would only become clear from the perspective of future generations.

7 The complete positivism

Comte had persuaded his disciples that the active class should nourish the contemplative class. After 1852, thanks to an increase in the Positivist Subsidy, he was able to devote himself entirely to his writings and to represent the priesthood of the Religion of Humanity. He concentrated on asserting the ultimate importance of morality and politics, in this way inaugurating the 'second part of the great revolution'. He affirmed the autonomy of the philosopher as well as that of the positivist movement. Since the appearance in July 1851 of the first volume of the *Système de philosophie positive*, Comte felt that he had completed the philosophical part of his work; he was now free to work on the religious part, which was basically a combination of morality and politics. The *Système de philosophie positive* was to conclude the French Revolution. Comte summed up his solution in the ideas 'Order' and 'Progress', which he combined in a unified and radical way. The positivist mission was to attain not a mixture but a 'necessary harmony' between the retrogression of Comte's time and anarchy. Comte loudly claimed to have answered the needs of the people and satisfied the poor, at the same reassuring the rich. He recommended not intervening in the contemporary political scene, but contented himself with constructing a provisional programme which consisted mainly in the spiritual reorganization of the West, while denouncing the powerlessness and instability of the 'incomplete positivists'.

It was impossible to conceive of an isolated people, far from the great human fraternity. Spiritual unity had to make the totality of human affairs prevail over parties and frontiers. Social positivism (the new religion) had to realize and complete intellectual positivism (the new epistemology). The prestige of progress explained the success that intellectual positivism had obtained. Social positivism would be able to achieve its goals only by reconstructing the spiritual order. As an opponent of theoretical materialism, which privileged the cosmological sciences, Comte believed that it was necessary to ensure the preeminence of the human and social sciences - for him, sociology and abstract morality. In 1856, he called attention to the great trilogy that consisted of his three most important works: *Cours de philosophie positive*, *Système de philosophie positive* and *Synthèse subjective*. According to Comte, the reign of solidarity had arrived - that of individuals and of peoples.

8 Influence

Comte's ideas were initially well received in England after the excellent review of the first two volumes of *Cours de philosophie positive* in the *Edinburgh Review* (1838). Later, John Stuart Mill confirmed his support in *System of*

Logic (1843), in which he presents social statics as the science of the 'coexistences of social phenomena' and social dynamics as that of their progressions. Mill saw in social statics a theory of the consensus between the different parts of the social organization, whereas social dynamics was a theory of society in its progressive movement; their combination led to a law of correspondence between the spontaneous stages and their simultaneous transformations, from which one could then discover the scientific law governing the development of humanity. Like Comte, Mill regarded the speculative faculties of human nature as the agents of social progress, and he also advocated the historical method (his 'Inverse Deductive Method') in the social sciences. In politics, he favoured the separation of powers and the establishment of a consensus based on a common public doctrine. In 1841 he began a six-year correspondence with Comte with an enthusiastic letter, but continued it in opposition to him in three areas: psychology, political economy and feminine intelligence. In *Auguste Comte and Positivism* (1865), Mill categorically rejected the Religion of Humanity, though he still accepted Comte's method.

Alexander Bain, the founder of the journal *Mind*, rallied, if not directly to Comtism, at least to the English positivist school. Like Comte, Bain (1855) connected his psychology to the spontaneous activity of the brain. Similarly, though he explicitly differentiated his own philosophy from that of Comte, Herbert Spencer, who wrote *Social Statics* (1851) on the basis of extreme individualism, then applied the doctrine of evolution to sociology, while at the same time rejecting Comte's conclusions. The English positivist Frederick Harrison, founder of *The Positivist Review*, was also the author of a *Social Statics* (1875). He engaged in a polemic with Spencer to force him to recognize the links between his philosophy and that of Comte. Spencer did at least admit that he owed the concept of social consensus to Comte. Other notable supporters are Richard Congreve (1818-99), who devoted himself to the propagation of English positivism, and Harriet Martineau, who edited a summary of Comte's philosophy, *The Positive Philosophy* (Comte 1830-42).

In France, Comte drew support from certain workers, such as Fabien Magnin, the author of *Études Sociales (Social Studies)* (1913), but his main devotee was the academic Pierre Laffitte, who published numerous positivist texts. Many people who did not strictly adhere to positivism nevertheless came under Comte's influence: Claude Bernard advanced the discipline of experimental medicine and studied the internal environment of advanced living organisms; and Pierre Duhem became well-known for his work on the elements of a natural classification, the epistemological independence of the fundamental sciences, and the holistic thesis. On the other hand, many French thinkers were critical of Comte, notably Émile Meyerson.

In general, Comte's positivism was received in two ways: some absorbed his central themes of the importance of scientific method in resolving social issues but rejected his religious movement; others were deeply impressed by his diagnosis of the crises afflicting modern society and were inspired by his vision for its redemption. In France, the former were led by Émile Littré, who was critical but friendly towards Comte's thought; the latter by Pierre Laffitte, a supporter of the positivist religion and from 1892 Professor of History of Science at the Collège de France. Two positivist reviews were published, *La philosophie positive* (1867-82) under the direction of Littré and G. Wyrouboff, and the *Revue occidentale* (1878-1914) created by Laffitte. The *Revue positiviste internationale* (1906-30) was founded later under the direction of Emile Corra. The religious aspect of Comte's positivism was developed by an English-born American, Henry Edger, who published *The Positivist Calendar* in 1856.

See also: Nishi Amane

Translated from the French by Mary Pickering
ANGÈLE KREMER-MARIETTI

List of works

Comte, A. (1830-42) *Cours de philosophie positive (Course in Positive Philosophy)*, Paris: Société Positiviste, 5th edn (identical to the first), 1892, 6 vols; ed. M. Serres *et al.*, Paris: Hermann, 1975, 2 vols; trans. and condensed H. Martineau, *The Positive Philosophy*, London: G. Bell, 1896, 3 vols. (Comte's major work.)

Comte, A. (1844) *Discours sur l'esprit positif (A Discourse on the Positive Spirit)*, Paris: Vrin, 1990.(An attempt to introduce and popularize positivism.)

Comte, A. (1848) *Discours sur l'ensemble du positivisme (A General View of Positivism)*, Paris: Société Positiviste, 1907.(A further introduction to positivism.)

Comte, A. (1851-4) *Système de politique positive ou Traité de sociologie instituant la religion de l'Humanité*, Paris: L. Mathias, 1928, 4 vols; Osnabrück: Zeller, 1967; trans. J.H. Bridges, F. Harrison *et al.*, *System of*

Positive Polity, London, 1875, 4 vols; repr. New York: Burt & Franklin, 1966.(Presents Comte's scheme for a new society.)

Comte, A. (1856) *Synthèse subjective ou Système universel des conceptions propres à l'état normal de l'Humanité (Subjective Synthesis)*, vol. 1: *Traité de philosophie mathématique,* Paris: Société Positiviste. (Comte's last, uncompleted, work.)

Comte, A. (1970a) *Plan des travaux scientifique nécessaires pour réorganiser la société (Plan of the scientific operations necessary for reorganizing society)*, Paris: Subier Montaigne.(The fundamental opuscule, dating from 1822.)

Comte, A. (1970b) *Écrits de jeunesse, 1816-28 (Early Writings, 1816-28)*, ed. P.E. de Berrêdo Carneiro and P. Arnaud, Paris: Mouton.(Includes *Opuscules de philosophie sociale* and *Mémoire sur la Cosmogonie de Laplace* and *Essais sur la philosophie des mathématiques*.)

Comte, A. (1973-90) *Correspondance générale et confessions (General correspondence and Confessions)*, ed. P.E. de Berredo Carneiro, P. Arnaud, P. Arbousse-Bastide and A. Kremer-Marietti, Paris: École des Hautes Etudes en Sciences Sociales, 8 vols.(A scholarly edition of Comte's correspondence.)

Comte, A. (1995) *The Correspondence of John Stuart Mill and Auguste Comte*, trans. with foreword A. Haac, intro. A. Kremer-Marietti, New Brunswick, NJ: Transaction Publishers.(A translation of the correspondence between Comte and Mill.)

References and further reading

Bain, A. (1855) *The Senses and The Intellect*, London.(Applies the findings of physiology to psychology.)

Brewster, D. (ed.) (1838) Review of *Cours de philosophie positive*, *Edinburgh Review* 67.(An unsigned and sympathetic review of the first two volumes of Comte's *Cours de philosophie positive*.)

Duhem, P. (1914) *La théorie physique - son objet - sa structure (Physical theory: its object and structure)*, Paris: M. Rivière.(Famous study by the French physicist.)

Edger, H. (1856) *The Positivist Calendar*, New York.(A guidebook for the practice of Positivism.)

Gouhier, H. (1933-41) *La Jeunesse d'Auguste Comte et la formation du positivisme (The young Comte and the formation of Positivism)*, Paris: Vrin, 3 vols.(A good basis for understanding the evolution of Comte's thought.)

Harrison, F. (1875) *Social Statics*, London.(An essay from one of the leaders of the Positivist movement in England.)

Kremer-Marietti, A. (1980) *Le Projet anthropologique d'Auguste Comte (Auguste Comte's anthropological project)*, Paris: SEDES.(Examines the principle of homology and epistemology.)

Kremer-Marietti, A. (1982) *Entre le Signe et l'Histoire: l'Anthropologie positiviste d'Auguste Comte (Between the sign and history: The positivist anthropology of Auguste Comte)*, Paris: Méridiens Klincksieck. (Expansion of the material in §§2-5 of this entry. Historical and semiotic approach to the theoretical structures of positivist epistemology.)

Kremer-Marietti, A. (1983) *Le Concept de Science positive. Ses tenants et ses aboutissants dans les structures anthropologiques du positivisme (The concept of positive science)*, Paris: Méridiens Klincksieck.(How the concept of positive science is taken as a model: complements §2 of this entry.)

Kremer-Marietti, A. (1987) 'Positivist Anthropology', *Encylopedia of Library and Information Science*, New York: Marcel Dekker INC.(Compares the theories of language and classification of Alexander Bryan Johnson and Comte.)

Kremer-Marietti, A. (1988) 'Auguste Comte et la sémiotique' (Auguste Comte and Semiotics), *RSSI* 8 (1-2): 131-44. (Seeks out the semiotic principle; see §4 above.)

Kremer-Marietti, A. (1992) 'Measurement and Principles. The Structure of Physical Theories', *Revue Internationale de Philosophie* 3 (182): 361-75.(Includes a view on holism.)

Larizza-Lolli, M. (1993) 'Le premier rayonnement en France des idées d'Auguste Comte (1824-48): les milieux, les institutions, les hommes' (The initial reception in France of Auguste Comte's ideas (1824-48): the environment, the institutions, the people), *Bulletin de la Société d'Histoire de la Révolution de 1848 et des Révolutions du XIXème siècle*, Paris, 69-101.(A historical essay on Comte's scientific environment.)

Laudan, L. (1971) 'Towards a Reassessment of Comte's ''Méthode Positive''', *Philosophy of Science* 37: 35-53; also in *Science and Hypothesis: Historical Essays on Scientific Methodology*, ed. L. Laudan, Dordrecht: Reidel, 1981, 141-62.(An essay on Comte's scientific method.)

Magnin, F. (1913) *Etudes Sociales (Social Studies)*, Paris.(The essays of a manual worker inspired by positivism.)

Mill, J.S. (1843) *System of Logic: Ratiocinative and Inductive*, in *Collected Works of John Stuart Mill*, vols 7 and 8, London: Routledge, 1991.(Emphasizes the study of human nature (psychology) and the study of human character (ethology).)

Mill, J.S. (1865) *Auguste Comte and Positivism*, in *Collected Works of John Stuart Mill*, vol. 10, 261-368, London: Routledge, 1991.(Mill criticizes Comte.)

Peirce, C.S. (1966) *Collected Papers*, vols 1-8, ed. C. Hartshorne, P. Weiss and A. Burks, Cambridge, MA: Harvard University Press. (Contributes to the logic of scientific methodology.)

Petit, A. (1991) 'La Révolution occidentale selon Auguste Comte: entre l'Histoire et l'Utopie' (The western revolution according to Auguste Comte: between History and Utopia), *Revue de Synthèse* CXIII (1): 21-40. (Presents utopia as the servant of history.)

Pickering, M. (1993) *Auguste Comte: An Intellectual Biography*, vol. 1, Cambridge: Cambridge University Press. (The newest and very important history of Comte's thought and life.)

Plé, B. (1996) *Die Welte 'aus' den Wissenschaften (The world 'out of' the sciences)*, Stuttgart: Klein-Cotta.(Study of sociology of science: positivism in France, England and Italy from 1848 to 1910.)

Scharff, R. (1991) 'Comte, Philosophy, and the Question of History', *Philosophical Topics* 19: 177-204. (Observations pertinent to Comte's epistemology. Referred to in §1.)

Scharff, R. (1995) *Comte after Positivism*, Cambridge, Cambridge University Press.(A recent study of Comte's thought.)

Schmaus, W. (1982) 'A Reappraisal of Comte's Three-State Law', *History and Theory* 21 (2): 248-66. (An interesting essay on Comte's famous law of three stages.)

Spencer, H. (1851) *Social Statics: or the Conditions Essential to Human Happiness and the First of them Developed*, London: Chapman.(Spencer's first important work, advocating extreme individualism.)

Concepts

The topic of concepts lies at the intersection of semantics and philosophy of mind. A concept is supposed to be a constituent of a thought (or 'proposition') rather in the way that a word is a constituent of a sentence that typically expresses a thought. Indeed, concepts are often thought to be the meanings *of words (and will be designated by enclosing the words for them in brackets: [city] is expressed by 'city' and by 'metropolis'). However, the two topics can diverge: non-linguistic animals may possess concepts, and standard linguistic meanings involve conventions in ways that concepts do not.*

Concepts seem essential to ordinary and scientific psychological explanation, which would be undermined were it not possible for the same concept *to occur in different thought episodes: someone could not even recall something unless the concepts they have now overlap the concepts they had earlier. If a disagreement between people is to be more than 'merely verbal', their words must express the same concepts. And if psychologists are to describe shared patterns of thought across people, they need to advert to shared concepts.*

Concepts also seem essential to categorizing the world, for example, recognizing a cow *and classifying it as a* mammal*. Concepts are also* compositional*: concepts can be combined to form a virtual infinitude of complex categories, in such a way that someone can understand a novel combination, for example, [smallest sub-atomic particle], by understanding its constituents.*

Concepts, however, are not always studied as part of psychology. Some logicians and formal semanticists study the deductive relations among concepts and propositions in abstraction from any mind. Philosophers doing 'philosophical analysis' try to specify the conditions that make something the kind *of thing it is - for example, what it is that makes an act good - an enterprise they take to consist in the analysis of concepts.*

Given these diverse interests, there is considerable disagreement about what exactly a concept is. Psychologists tend to use 'concept' for internal representations*, for example, images, stereotypes, words that may be the vehicles for thought in the mind or brain. Logicians and formal semanticists tend to use it for* sets *of real and possible objects, and* functions *defined over them; and philosophers of mind have variously proposed* properties*,* 'senses'*, inferential* rules *or discrimination* abilities.

A related issue is what it is for someone to possess *a concept. The 'classical view' presumed concepts had 'definitions' known by competent users. For example, grasping [bachelor] seemed to consist in grasping the definition, [adult, unmarried male]. However, if definitions are not to go on forever, there must be* primitive *concepts that are not defined but are grasped in some other way. Empiricism claimed that these definitions were provided by sensory conditions for a concept's application. Thus, [material object] was defined in terms of certain possibilities of sensation.*

The classical view suffers from the fact that few successful definitions have ever been provided. Wittgenstein suggested that concept possession need not consist in knowing a definition, but in appreciating the role of a concept in thought and practice. Moreover, he claimed, a concept need not apply to things by virtue of some closed set of features captured by a definition, but rather by virtue of 'family resemblances' among the things, a suggestion that has given rise in psychology to 'prototype' theories of concepts.

Most traditional approaches to possession conditions have been concerned with the internal *states, especially the beliefs, of the conceptualizer. Quine raised a challenge for such an approach in his doctrine of 'confirmation holism', which stressed that a person's beliefs are fixed by what they find plausible overall. Separating out any particular beliefs as defining a concept seemed to him arbitrary and in conflict with actual practice, where concepts seem shared by people with different beliefs. This led Quine himself to be sceptical about talk of concepts generally, denying that there was* any *principled way to distinguish 'analytic' claims that express definitional claims about a concept from 'synthetic' ones that express merely common beliefs about the things to which a concept applies.*

However, recent philosophers suggest that people share concepts not by virtue of any internal facts, but by virtue of facts about their external (social) environment. For example, people arguably have the concept [water] by virtue of interacting in certain ways with H_2O *and deferring to experts in defining it. This work has given rise to a*

variety of externalist *theories of concepts and semantics generally.*

Many also think, however, that psychology could generalize about people's minds independently of the external contexts they happen to inhabit, and so have proposed 'two-factor theories', according to which there is an internal *component to a concept that may play a role in psychological explanation, as opposed to an* external *component that determines the application of the concept to the world.*

1 Concepts as shareable constituents of thought

Constituents of thought. It is widely thought that 'intentional' explanation in terms of such states as belief, thought and desire affords the best explanation of the behaviour and states of people, many animals and perhaps some machines: someone drinks water because they have a thirst which they think water will quench. By and large, philosophers and psychologists such as Fodor (1975, 1991) or Peacocke (1992) who are interested in intentional explanation take themselves to be committed to the existence of concepts, whereas those sceptical of this form of explanation, for example, Quine (1960), tend to avoid them (see Animal language and thought; Cognitive development; Intentionality).

Suppose one person thinks that water dissolves salt, and another that it does not. Call the thing that they disagree about a 'proposition' - for example, [Water dissolves salt]. It is in some sense shared by them as the object of their disagreement, and it is expressed by the sentence that follows the verb 'thinks that'. Concepts are the constituents of such propositions (in at least one understanding of them; see Propositions, sentences and statements), just as the words 'water', 'dissolves' and 'salt' are constituents of the sentence. Thus, these people could have these beliefs only if they had, *inter alia*, the concepts [water], [dissolves], [salt].

Just which sentential constituents express concepts is a matter of some debate. The central cases that are discussed tend to be the concepts expressed by predicates or general terms, such as 'is water' or '*x* dissolves *y*', terms potentially true of many different individual things. But there are presumably concepts associated with logical words (for example, [not], [some]), as well as with individual things (for example, [Rome], [2]).

Shareability. If we are to make sense of processes of reasoning and communication, and have a basis for generalization in a cognitive psychology, then concepts must be shareable. Consequently, concepts need to be distinguished from the particular ideas, images, sensations that, consciously or unconsciously, pass through people's minds at a particular time. The concept [cat] could not be some individual experience someone has, since in that case no two people could share it and a single person probably could not have the same one twice. Just what kind of shareable object a concept might be is a matter of considerable disagreement among theorists. In much of the psychological literature, where the concern is often with features of actual mental processing, concepts are regarded as mental representations, on such as words or images.

It will be important with respect to this and later proposals to invoke a distinction from the study of language between types and tokens (see Type/token distinction). A linguistic *token*, such as an inscription of 'café' on a door, has a specific spacetime location: one can ask when and where it occurs; whereas a linguistic *type* - the word 'café' - like any 'universal', is an abstract object outside space and time. One can erase a token of the word 'café', but the type word would still exist. Similarly, concepts could be regarded as internal representation types that have individual ideas as their specific tokens. On this view, you and I could share the concept [water] if you and I have tokens of the same representation type in our minds or brains. There is a good deal of discussion in psychology as to whether concepts in this sense are (type) words, phrases, pictures, maps, diagrams or other kinds of representations, for example, 'prototypes' or 'exemplars' (see §7 below; and Smith and Medin (1981); Rips (1995) for reviews of the psychological literature).

But many philosophers take the view that these mental representation types would no more be identical to concepts than are the type words in a natural language. Words in a language are usually individuated syntactically, allowing both spoken and written tokens of a word to be of the same type, and syntactically identical tokens (for example, of 'bank') to be ambiguous, or of different semantic types. Moreover, different syntactic types - for example, 'city', 'metropolis' - can be synonymous, that is, be of the same semantic type. Similarly, one person might express the concept [city] by a mental representation 'city', another by 'metropolis'; still another perhaps by a mental image of bustling boulevards. But, for all that, they might have the same concept [city]: one could believe and another doubt that cities are healthy places to live. Moreover, different people could employ the same

representation to express different concepts: one person might use an image of bustling boulevards to express [city], another to express [pollution]. So, on the standard philosophical reading (which we shall follow here), concepts are to be individuated differently from the representations that express them.

However, although concepts understood in this latter way are arguably also indispensable to psychology, they raise different issues from those concerning representations that standardly interest psychologists. Questions about representations typically involve just the issues that psychologists have tended to investigate: processing time, ease of judgment, susceptibility to errors. If one person represents cities and their relations ‘spatially’, where another represents them by names and descriptions, this may explain differences in how rapidly the two of them can answer questions about cities; but, again, presumably they both still have the concept [city]. Just why they would, what the possession conditions for [city] or any concept might be, is not easy to say (see §5): the point here is that they seem to involve issues different from the issue of identifying a syntactically defined representation.

This difference between the psychologist’s and philosopher’s typical interest is sometimes obscured by ambiguous phrasing. When Kant identifies ‘analytic’ claims (or claims that express the ‘analysis’ of a concept) as those in which one concept is ‘contained in’ another, he glosses this by saying: ‘I have merely to analyse the concept, that is, to become conscious to myself of the manifold which I always think in that concept’ (1781/1787: A7). In our terms, this could be read as a claim about representations, or (presumably what he intended) about the concepts they express. My mental representation of freedom might invariably involve an image of dancing people, but surely neither Kant nor I would want to say that the analysis of my concept of freedom involved dancing. Conversely, there is no reason that a good analysis should serve as a representation in ordinary, rapid reasonings (for example, identifying something as a bird): indeed, vivid images might serve better.

2 Meanings of words

As most of our examples suggest, concepts are presumed to serve as the meanings of linguistic items, underwriting relations of translation, definition, synonymy, antinomy and semantic implication (Katz 1972). Indeed, much work in the semantics of natural languages (see Jackendoff 1987) takes itself to be addressing ‘conceptual structure’. This is partly motivated by Grice’s (1957) proposal to understand linguistic meaning ultimately in terms of the intentions with which speakers produce linguistic tokens: ‘good’ means what it does at least partly because of what users of the word have intended to mean by it; that is, because of the concept they have intended to express (see Meaning and communication; Grice, H.P. §2).

One problem with this role for concepts is that it is by no means clear just what a theory of meaning is supposed to involve. Some of the issues are exactly the issues we are considering here. However, some issues seem peculiar to language: for example, how much of what is understood in the uttering of a sentence is part of its meaning, or semantics, and how much is part of its use, and so an issue of pragmatics? (See Pragmatics.) If I say of someone ‘He is not very good at chess’, is the meaning simply that ‘It is not the case that he is very good’, or ‘He is bad at chess’?

3 Concepts and analysis

Objects of analysis. At least since Plato’s *Euthyphro*, philosophers have been fascinated by a certain sort of question about constitutiveness: in virtue of *what* is something the kind of thing it is - for example, what is ‘essential’ to something’s being good, a piece of knowledge, a free act? Obviously, not just any truth about the target phenomenon will suffice as an answer: to take Plato’s *Euthyphro* example, merely the fact that the gods love the good is no reason to think that that is what makes something good, any more than that all bachelors eat is what makes them bachelors (see Conceptual analysis).

Some philosophers think such questions are answered by natural science. This certainly seems to be true in the case of ‘natural kinds’ such as water or polio, which arguably have ‘real essences’ largely independently of us (see Kripke [1972] 1980; Putnam 1975; Essentialism). But many concepts, such as [magic], [freedom], [soul] may not pick out any real kind of thing at all (much less one studiable by science): in these cases, all that seems shareable by different possessors of the concept is some belief or other. But even in the natural science cases, some conceptual analysis seems to many unavoidable, if only to determine exactly what the science is about (what makes an investigation one about water, or polio, or consciousness; Bealer 1987).

There is a related question that concepts are also sometimes recruited to answer; not a metaphysical question about 'the nature of things', but an epistemological one about how people seem to know a priori (or 'independently of experience') various necessary truths, for example, that there is an infinity of prime numbers or that equiangular triangles are equilateral ones. However, it should be seen as a substantive and controversial hypothesis, to which we will return, whether this epistemological interest should coincide with the above metaphysical one (see A priori; §5 below).

Philosophers have also sometimes hoped that conceptual analysis would help (dis)solve certain philosophical questions about, for example, truth, free will, personal identity, either by clarifying the commitments of the relevant concepts, or by showing that they were somehow defective. A once popular strategy was to show that the application of a concept was not 'verifiable' (§5 below). A more recent strategy is to show that the application of a purported concept would be unintelligible, as in the case of [absolute space] (Peacocke 1992: ch. 8).

Whatever its ultimate philosophical benefits, conceptual analysis does seem to involve the sort of facts that are relevant to the question of concept possession. Returning to our pair of people who represent cites differently: what seems relevant to the question of whether, despite their different representations, they still have the same concept [city] is what sorts of things they believe are essential to being a city; what things they would and would not count as cities (which is not to say that this would be decisive; see §5). This difference can lend an air of unreality to philosophical as opposed to psychological discussions of concepts, depending, as it does, upon difficult questions about what (one would think) is possible in various often very outlandish situations.

Vagueness. One supposed defect of many concepts should be set aside from the start. The belief in constitutive analyses of a concept is often thought to be undermined by the existence of difficult, borderline cases for its application (Wittgenstein 1953: §§66-7; Smith and Medin 1981: 31). Now, it certainly cannot be denied that the world is full of genuine borderline cases: 'Is drizzle rain?', 'Are viruses alive?'. Arguably, the world does not supply determinate answers: all kinds in the world may have vague boundaries, any precise delimitation of which may depend upon human decision. But this does not imply that all applications of concepts are up to human decision, much less that there are no defining essences of the phenomena they pick out. [Unmarried adult male] may be a perfectly good analysis of [bachelor] not only despite hard cases, but because the hard cases for the one correspond exactly to the hard cases for the other. (See Vagueness.)

4 Referential views

One candidate for the common object of people's thoughts has been simply the referents of their representations, that is, the objects in the world picked out by them. Someone might say that two representations express the same concept if and only if they refer to the same thing(s) in the world. Putting aside the difficult problem of explaining 'reference', this seems an appealing suggestion, clearest in the case of (token) proper names such as 'Aristotle' where the name refers to a specific person (see proper names). When we turn to predicates, however, things are not so clear. There are a number of different candidates for what counts as the 'referent' of a predicate and so of the concept it would express: (1) its extension, (2) its intension (as function), and (3) the property that all the (possible) objects satisfying it have in common.

Extensions. The 'extension' is the set of actual objects that satisfy the concept. For example, the extension of [city] might be the set of cities: {Paris, London, Madrid,…}. Russell (1956) proposed an account of 'propositions' according to which they were composed of real objects in the world combined with properties (see also Kaplan 1979). Extensional logicians such as Goodman (1951) and Quine (1960) think that sets of actual objects are all that are needed for serious science: all that needs to be mentioned are actual lions, tigers and quarks. They realize that this suggestion clashes with our ordinary understanding. [Cordate] is not the same as [renate], despite the fact that (let us suppose) all and only actual creatures with kidneys are creatures with hearts ([renate] and [cordate] are 'coextensive'). It seems reasonable to require concepts to cover *possible* cases, for example, possible creatures that are renates but do not have hearts. Goodman would not agree, since, as he famously argues (1951: 5), 'the notion of "possible" cases, of cases that do not exist but might have existed, is far from clear' (see Counterfactual conditionals; Goodman, N. §3). But he and Quine would also be wary of talk of concepts generally; and semanticists such as Kaplan (1979) and Salmon (1986) are anxious to avoid introducing talk of them into talk of the semantics of language.

Intensions as functions. Although one might agree that modal notions such as possibility and necessity are not as clear as one would like, it is by no means agreed that science can actually dispense with them. Many philosophers think that the laws essential to causal explanation in any science require modal and counterfactual talk. But, especially in psychology, it seems doubtful that extensions would perform all the explanatory work concepts are needed to perform. Whether or not biology need worry about the possibility of renates lacking hearts, someone could *think* something is a renate without thinking it has a heart. Consequently, many philosophers have claimed that, in addition to extensions, there must be 'intensions', or entities distinguished more finely than mere extensions permit (see Intensional entities; Intensional logics).

Intensions have been defined in a number of ways. One approach is in terms of 'senses' or 'modes of presentation', to which we shall turn shortly (§5 below). Another approach simply amplifies the extensional characterization to include sets of possible as well as actual objects. Modal logicians and formal semanticists such as Montague (1974), D. Lewis (1972) and Stalnaker (1984), interested in presenting a formal account of the semantics of natural languages, have regarded intensions as functions that map a possible world to the extension of the concept in that world (see Semantics, possible worlds).

However, mere appeals to possibilia may still not cut things finely enough: for there are concepts that are different even though they apply to all the same things in all possible worlds, for example, [equiangular triangle] and [equilateral triangle], or, following Kripke ([1972] 1980), [water] and [H_2O]; or, to take cases of necessarily *empty* extensions, [square circle] and [married bachelor], which both refer to nothing in all possible worlds. Particularly interesting examples of this latter category have been suggested by Kripke ([1972] 1980) and Slote (1975), who argue in different ways that nothing could possibly satisfy the specific demands of [unicorn] or [monster]. How are we to distinguish these concepts by reference to possible objects?

Properties. Some philosophers think the appropriate reference for predicate concepts is not provided by the objects that the concepts pick out (whether in the actual or merely possible worlds), but rather by the properties those objects share. Thus, [city] is not individuated by the set of all actual or possible cities, but rather by the property, 'being a city'.

Historically, concepts have not always been clearly distinguished from properties, both being regarded as 'universals' (see Universals). Thus, the *mortality* one found widespread among men was often assumed to be the same as the concept [mortality] that was a constituent of one's fears. Sometimes this identification seems terminological, as in Frege (1892a) (who uses 'concept' in a quite special way and discusses 'senses' independently), but at other times it is substantive, as in Carnap (1952). Most writers these days would distinguish the two (Putnam 1970; Bealer 1982), even if they also think that there is a property for every concept.

An interesting issue raised by Frege (1892a) that does seem common to both general concepts and properties is the difficulty of specifying exactly what sort of entities they are. As Frege noted, they seem to be incomplete, or 'unsaturated', having places in them awaiting completion by objects, in the way that predicates in language, such as '*x* loves *y*', have variables awaiting substitution with names (such as 'Romeo' and 'Juliet'). This issue becomes important when we try to specify how general and singular concepts combine to form a thought, or how properties and objects combine to constitute facts (Russell 1903: ch. 4; Wittgenstein 1921).

Many worry that appealing to properties to individuate concepts is gratuitous metaphysics, a free invention of a property for every concept, encouraged by the loose presumption that properties, such as 'immortality', can exist even without being instantiated. To answer this charge, philosophers often claim that properties are provided by the actual causal structure of the world: having kidneys and having a heart enter differently into causal relations, as perhaps do equiangularity and equilaterality (Sober 1982). Particularly philosophers of mind such as Dretske (1981, 1988), Millikan (1984) and Fodor (1991), interested in causal interactions between an organism and the world, find this way of thinking about concepts attractive (see Semantics, informational; Semantics, teleological).

However, it is not clear that causal properties will suffice. Are all our concepts really of causally efficacious properties? What about the concept [an inefficacious property]? Or concepts of secondary properties ([red], [sweet]), or ethics and aesthetics ([good], [comical]), which, many have argued, do not pick out genuine causal properties? Or consider, again, necessarily coextensive concepts such as [water] and [H_2O] - which arguably correspond to the same property - or [round square] and [married bachelor], which arguably correspond to none.

How could they differentially enter into the causal structure of the world? Or should we suppose that properties can be distinguished even though they are indistinguishable not only in the actual, but in all possible worlds? Are there any constraints?

Moreover, even if we could distinguish concepts by properties, that would not suffice for conceptual analysis. Plato's *Euthyphro* question - is something good because the gods love it, or do they love it because it is good? - brings this out nicely, since the question remains even if we assume that that gods love the good in all possible worlds (so that [good] and [god-beloved] are necessarily coextensive): the *direction* of analysis still needs to be specified.

5 Possession conditions: external v. internal

Many philosophers might not think that Plato's *Euthyphro* question needs an answer: for purposes of logic, and perhaps even formal semantics, appeal to any of these external phenomena (extensions, intensions or properties) may suffice, 'analysis' be hanged. However, if concepts are to play a role in psychology, then one at least wants to know what sort of relation someone must bear to these external phenomena in order to qualify as a competent user of the concept. This is a question about the possession conditions for a concept to which many have thought analyses are crucial.

In considering possession conditions, special care is needed with the peculiar idiom 'concept of *x*' and the ontology it involves. Psychologists often speak of such things as the child's 'concept of causality'. This could mean the representation the child employs of the concept [causality] that the child shares with adults; or it could mean any of the extension, intension or rule that children associate with the English word 'cause'; or it could mean (as in fact it very often does mean) merely the standard beliefs - what some call the 'conception' - that children associate with the extension, intension: [causality]. Which of these is intended all depends upon what entity one thinks of as the concept and what a mere accompaniment of it. What cannot be seriously intended is the suggestion that a child has a concept [causality] that is both identical to but different from the adults'. (What invites confusion here is the 'of' that implies no relation: just as 'the nation of China' means 'the nation, China', so 'the concept of *x*' often means 'the concept, [*x*]'.)

It might be thought that none of the external identifications of concepts could ever be viable for psychology, since psychology is about what is 'in the mind' (or 'in the head'), not what is in the world external to it (see Jackendoff 1987: 126). This would be an error. We have already seen one way in which it cannot be true: in so far as concepts are shareable, they must be distinguished from individual mental episodes. But still it might be thought that a concept must be a *type* of internally specified mental state, since, after all, surely psychology aims to talk about individual minds, even if it categorizes them in various ways. However, an interest in characterizing what is going on in the mind need not exclude alluding to external objects: the fact that extensions or properties may be external to the head is no reason to think them unsuitable candidates for classifying things that are in the head, just as classifying various words in a book as 'about Vienna' does not prevent those words from existing entirely inside the book.

There have been a variety of external relations that philosophers have proposed that would link internal representations to external phenomena in a way that might constitute concept possession. Influential articles by Kripke ([1972] 1980, 1982), Putnam (1975) and Burge (1979) have given rise to a variety of externalist theories of concepts, which look to such facts as actual causal history (Devitt 1981), various co-variation conditions (Dretske 1981, 1988; Fodor 1991) and evolutionary selection (Millikan 1984; Papineau 1987). However, in so far as they rely on real phenomena in the external world, they are subject to certain limitations that many feel can only be surmounted by appealing to some kind of conditions that are 'in the head'.

Intensions: 'senses'. What argues for the need for some internal condition on concept is the difficulty for any purely external account of capturing psychologically real distinctions. The examples of necessarily unextended concepts, such as [round square], suggest that the mind can somehow make distinctions for which there is no possible external reality. Consequently, many philosophers have argued that, in addition to the referent of a general term, there must also be (following Frege 1892b) its 'sense', or 'mode of presentation' (occasionally 'intension' is used here as well; see Sense and reference). Thus, what really seems to distinguish [equiangular triangles] and [equilateral triangles] is not the actual or possible things to which they refer, but rather the way the

mind conceives them: it is one thing to think of something as (or *qua*) an equilateral triangle, another to think of it *qua* an equiangular triangle (which is why the proof that they are necessarily coextensive is informative). And this obviously helps with the problem of the necessarily coextensive: what distinguishes, for example, [water] and [H_2O] are the different 'ways of thinking', not reflected in any even possible difference in the world. For some (for example, Peacocke 1992) concepts are senses so understood. But, of course, we then need a theory of senses.

6 The classical view and empiricism

One conception of senses is provided by the classical view of concepts. This view has two independent parts that are not always clearly distinguished, one making a claim about the nature of concepts, the other about what is to possess them: (a) concepts have an 'analysis' consisting of conditions that are necessary and sufficient for their satisfaction; and (b) these 'defining' conditions are known to any competent user. An interesting, but problematic, example has been [knowledge], whose analysis was traditionally thought to be [justified true belief], but which has turned out to be far subtler, due to counterexamples raised by philosophers such as Gettier (1963) (see Knowledge, concept of).

The example of [knowledge] brings out an important caveat for the classical view: the proper analysis of a concept need not be readily available to a competent user of it. It was not easy for Athenians to reply to Plato's inquiries about [good], nor for recent philosophers to reply to Gettier. According to a reasonable version of the classical view, a competent user's knowledge of an analysis may be 'tacit' or 'unconscious', rather like the 'knowledge' people have of the grammatical rules of their language, which they seem dependably to obey despite being unable to articulate them (see Evans 1981; Katz 1971; Knowledge, tacit; Unconscious mental states). For Plato, the analyses could only be extracted from someone by a process of 'dialectic', involving consideration of various examples and arguments to a point of 'reflective equilibrium' (Bealer 1987: 322; Jackson and Pettit 1995).

The classical view, however, has always had to face the difficulty of primitive concepts: how are they to be defined? An influential (but not the only possible) answer was provided by seventeenth-century British empiricists, who claimed that all the primitives were sensory. Indeed, the classical view has often been uncritically burdened with this further claim, or, anyway, the claim that all concepts are 'derived from experience'. Locke (1689), Berkeley (1710) and Hume (1739-40) seemed to take this to mean that concepts were somehow composed of introspectible mental items - 'images', 'impressions' - that were ultimately decomposable into basic sensory parts (see Empiricism; Sense-data).

Berkeley ([1710] 1982: 13) noticed a problem with this approach that every generation has had to rediscover: if a concept is a sensory impression, like an image, then how does one distinguish a general concept [triangle] from a more particular one - say, [isosceles triangle] - that would serve in imagining the general one? In any case, images seem quite hopeless for capturing the concepts associated with logical terms (what is the image for negation or for possibility?). Whatever the role of images, concepts and our competence with them involve something more (see Imagery).

Indeed, in addition to images and impressions and other sensory items, a full account of concepts needs to consider issues of logical structure. This is precisely what the logical positivists did, focusing on logically structured propositions and transforming the empiricist claim into their famous 'verifiability theory of meaning': the meaning of a proposition is the means by which it is confirmed or refuted, ultimately by sensory experience; the concept expressed by a predicate is the statement of the (perhaps logically complex) sensory conditions under which people confirm or refute whether something satisfies it (see Meaning and verification). Thus, [acid] might be analysed by reference to tendencies to cause litmus paper to turn red; [belief] by observable behavioural dispositions (see Behaviourism, analytic); [material object] by enduring possibilities of sensation (see phenomenalism).

This once popular position has come under much attack in the last fifty years. Few, if any, successful 'reductions' of ordinary concepts (such as [material object], [cause]) to purely sensory concepts have ever been achieved, and there seems to be a pattern to the failures (Chisholm 1957). There have been four main diagnoses: (1) the classical search for 'necessary and sufficient' conditions is misguided and ought to be replaced by an appreciation of the role of a concept in our reasonings and theories of the world; (2) concepts should be regarded as 'family resemblance' structures, or 'prototypes', (3) because of the 'holism of confirmation', attempts to

analyse concepts in terms of any verification conditions cannot succeed; and (4) we should stop looking for characterizations of concepts in terms of epistemic conditions, but rather, more metaphysically, in terms of the actual phenomena in the world to which people are referring, but about which they might be ignorant (thus, we would abandon clause (b) of the classical view, which requires analyses to be known by competent users). We will discuss each in turn.

7 Inferential roles and prototypes

Inferential roles. The first alternative, inspired by Wittgenstein's famous dictum, 'the meaning of a word is its use' (1953: §43), treats concepts as involving some or other role of a representation, either in a theory or in thought. Thus, many (for example, Kuhn 1962) have argued that someone possesses a concepts such as [witch] or [phlogiston] only if they understand the theories in which they play a role, or can reason with it in certain appropriate ways (see Semantics, conceptual role).

A vexing problem with this approach has been the fact that it is hard to identify just which roles are essential to a concept. It would appear that people can be wrong and/or disagree about almost anything: Berkeley claimed that material objects were ideas, some creationists that people are not animals, some nominalists that numbers are concrete objects. If people are genuinely to disagree with these views, they must share the relevant concepts; otherwise their use of the same words would be equivocal, their disagreement 'merely verbal'. But then it seems very hard to insist upon any specific inferential role being essential to possessing a concept.

Prototypes. Another proposal also inspired by Wittgenstein (1953: §66) is to appeal to 'family resemblances' among the things to which a concept applies: he claimed that games, for example, share no single property, but are similar to each other in various ways that cluster together (some involve winning/losing, others mere entertainment; some are played in groups, others alone). This speculation was taken by psychologists (for example, Rosch 1973; Smith and Medin 1981) to be a testable psychological hypothesis. They showed that people respond differently (in terms of response time and other measures) to questions about whether, for example, penguins rather than robins are birds, in a fashion that suggested that concept membership was a matter not of possessing a classical analysis, but of 'distance' from a 'prototype' or typical 'exemplar'. Thus, a robin satisfies many more of the features of a typical bird than does a penguin and so is a 'better' member of the category; and a malicious lie is a better case of a lie than a well-intentioned one.

It has not always been clear precisely what sort of thing a prototype or exemplar might be. One must take care not to import into the mind procedures, such as comparing one actual bird with another, that make sense only outside of it. Presumably either a prototype or an exemplar is some sort of representation (a list, or an 'image') indicating selected properties, and a metric for determining the distance of a candidate from those properties. Some writers have exploited the resources of 'fuzzy set theory' to capture the intended structure, whereby membership of a category is understood not as an 'all or none' affair, but as a matter of degree: everything satisfies every concept to some degree, however small (see Zadeh 1982; Fuzzy logic).

Quite apart from specifying just what the view involves, there are, however, a number of problems with appeals to prototypicality as a theory of concepts. In the first place, loosening the conditions on a concept's application from 'defining' conditions to mere 'family resemblances' risks leaving that application far too unconstrained. Everything after all bears some resemblance to everything else (Goodman 1970): returning to Wittgenstein's example, anything, *x*, resembles standard games in some way or other (if only in belonging to some arbitrary set that contains all games and that thing *x*!). The question is which resemblances are *essential* to the concept, and which merely *accidental* - a question that returns us to the question the classical view tries to answer (see Essentialism).

Second, prototypes seem poor candidates for handling the crucial phenomenon of conceptual combination: the prototype for [tropical fish] does not seem constructible from the prototypes for [tropical] and [fish], yet someone could grasp [tropical fish] none the less (Osherson and Smith 1982).

Third, prototypicality, which presumably involves distances among a complex cluster of diverse properties, must be distinguished from both vagueness and estimation. As we have already observed (§1), nearly every concept admits of vague cases, in which it is not clear whether the concept applies. But this does not imply that the concept does not have a (correspondingly vague) definition. Similarly, estimation of whether or not something satisfies a

concept is a question that arises with regard to *any* concept. But it is clearly not a *metaphysical* issue of the actual conditions something must satisfy in order to satisfy a concept, but rather an *epistemological* one concerning the belief or epistemic probability that something satisfies the conditions, given certain evidence (see Realism and antirealism). The sight of someone with a toupee may mean that there is a 90 per cent probability that he is actually 50 per cent bald, or a 40 per cent probability that he is actually 95 per cent bald. The question of whether [bald] has a classical analysis is untouched by this issue as well (Rey 1983).

8 Metaphysics v. epistemology

These latter distinctions may turn on the different interests we have already noted (§1) in psychologists' and philosophers' use of 'concept', applying respectively to representations or to their shareable meanings. The fact that people are quicker to say that robins rather than penguins are birds may tell us something about people's representations of [bird], but nothing about the definition of the concept [bird] itself, that is, what is in fact required to satisfy that concept (on reflection, after all, most of us agree that penguins are bona fide birds, despite our initial hesitation). This is not to say that the definitional issue is not relevant to psychology: what people take to be required to be a bird is as much a psychological issue as how they figure out whether those requirements have been met. It is just that prototype theory seems to be addressed largely to the latter issue, the classical view to the former.

However, the differences may be deeper than merely terminological. As the example of estimation shows, it is extremely easy to conflate metaphysical issues about the conditions for something's satisfying a concept with the epistemic ones of estimating whether something actually satisfies those conditions. One reason is that English can encourage running the two together, phrasing the metaphysical question as 'What determines what is what?' and the epistemic one as 'How does someone determine what is what?'. A second reason is that epistemic conditions are as likely to 'come to mind' in thinking with a concept as are its defining conditions (see §1 above).

But a more important reason is that empiricism made a policy of connecting the two: the defining conditions for a concept were to be stated in terms of experiential evidence. As anti-empiricists from Plato on have argued, however, many of our concepts seem to 'transcend experience', in that they seem to be graspable and sometimes applicable in the absence of it. For lack of any genuine Euclidean triangles in the world, it is unclear how our concept of them could be derived from experience. And even instantiated concepts such as [material object], [causation] and [prehistoric] seem to go far beyond mere sensory experience: we seem to be able to think coherently about material objects causally interacting in prehistoric times, even in the absence of any sensory evidence of that interaction. In any case, many of our concepts transcend their stereotypes: most of us understand the concept [female doctor], and recognize that [even number] has a perfectly good definition, despite our demonstrable reliance on stereotypes in both cases (Armstrong, Gleitman and Gleitman 1983).

Moreover, it can often seem arbitrary and unduly restrictive to tie a concept to any particular method of verification (or confirmation). Taking a page from Pierre Duhem (1914), Quine (1953) argued that 'our beliefs confront the tribunal of experience only as a corporate body': litmus paper turning red confirms that a solution is acidic only in conjunction with a great deal of background chemical and physical theory; indeed, Quine claims, only in conjunction with the whole of a person's system of beliefs (a view called 'confirmation holism'; see Analyticity; Confirmation theory; Quine, W.V.). Hence, if a concept is to be analysed as its verification conditions, its meaning would be similarly holistic ('meaning holism'). Given that no two persons' beliefs are likely to be precisely the same, this has the consequence that no two people ever share precisely the same concepts - and no one could, strictly speaking, remember the same thing over any amount of time that included a change of any belief! Fodor and LePore (1992) have recently argued that this sort of conceptual (or semantic) holism would undermine serious psychology, but, fortunately, that the arguments for it are less than compelling (see Holism: mental and semantic; Atomism, ancient).

9 Difficulties for an internalist approach

Even if one distinguishes epistemological from metaphysical issues in determining concept identity, there remain a number of problems for any purely internalist theory of concepts. Whether classical or prototypical, any internalist theory of concepts requires distinguishing internal features - beliefs, inferential roles, prototypes - that are essential to (or defining of) a concept from those that are accidental, and many feel that it really is this distinction that is

undermined by Quine's observations about holism. Indeed, 'sameness of concept' for Quine becomes by and large an 'indeterminate' issue (see Radical translation and radical interpretation). The most one might expect is a similarity of inferential role between symbols in different theories or symbol systems (Harman 1972; Block 1986): which similarities are selected may vary for different explanatory tasks, and may be a pragmatic affair (Bilgrami 1992; Lormand 1996).

A second problem emerged from the externalist approaches of Kripke ([1972] 1980), Putnam (1975) and Burge (1979), and has come to be represented by Putnam's example of 'twin earth' (1975): suppose there were a planet exactly like the earth in every way except that, wherever the earth has H_2O, twin earth has a different, but superficially similar chemical XYZ. Putnam argues that twin-earthlings would not mean by the word 'water' what we mean (for example, their tokens of the sentence 'Water is wet' would not have the same truth-conditions as ours), despite the fact that, *ex hypothesi*, twin-earthlings would have our same internal structure. (Burge (1979) and Stich (1983) present less outlandish examples.) As Putnam famously put it, 'meanings just ain't in the head'; rather, he argues, they depend at least in part upon the relations between internal states and external phenomena (see Content: wide and narrow).

In view of both these examples, and the Quinian worries, Fodor (1991, 1998) opts for an entirely 'atomistic' account of concepts, arguing that concepts have no 'analyses' whatsoever: they are simply ways in which people are directly related to individual properties in their environments, any one of which they might enjoy without the others. In principle, someone might have the concept [bachelor] and no other concepts at all, much less any 'analysis' of it, simply by virtue of having some internal state that is causally connected with bachelorhood in the local environment. Such a view goes hand in hand with Fodor's rejection of not only verificationist, but any empiricist, account of concept learning and construction. Indeed, given the failure of empiricist constructions, Fodor (1975, 1979) argues that concepts are not constructed or 'derived from experience' at all, but are (nearly enough) all innate (see nativism). Devitt (1995) defends a more moderate, 'molecularist' position, allowing that many innate primitives are non-sensory (for example, [cause], [object]) but that others are susceptible to definition, especially in view of the thereby enlarged primitive base.

10 Two-factor theories and a modified classical view

Although externalism does seem to account both for the stability of concepts through variation in belief and for variations in concept due to variations in environment, it still must confront the issues we mentioned earlier (§5) that invite internalism, namely, that there seem to be more distinctions in the mind than are available in the external world. An increasingly popular approach is to separate the internal and external work concepts are asked to perform. According to 'two-factor' theories, 'concepts' should be regarded as having two components: one 'in the head', consisting of an internal representation playing a certain psychological role; and the other, some sort of co-variational law or evolutionary fact that, in a historical context, determines the reference and truth-conditions of the concept. (Sometimes 'concept' is restricted to the internal factor, 'content' to the external; and 'two-factor theory' is sometimes applied only to those views in which the two factors are relatively independent of one another.)

A two-factor theory leaves a place for a modified classical view, as well as for something like philosophical analysis. The internal factor would determine a full semantic content to a conceptual representation only in a particular context, so that the full analysis of (the content of) a concept might await empirical investigation of that context and not be available to its user. But this is perhaps as it should be: philosophical analysis of 'the nature of' a phenomenon may depend both upon the internal rule one is deploying and the actual phenomenon that, in a context, the rule picks out (Bealer 1987; Jackson and Pettit 1995).

See also: Content, non-conceptual; Semantics

GEORGES REY

References and further reading

Armstrong, S., Gleitman, L. and Gleitman, H. (1983) 'What Some Concepts Might Not Be', *Cognition* 13 (3): 263-308.(Discussion of experiments demonstrating prototypicality effects even with concepts, such as [even number], that have obvious classical definitions.)

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz.(The classic defence of the logical positivist

programme, including efforts to analyse all empirically meaningful statements into statements about possible sense experience.)

Bealer, G. (1982) *Quality and Concept*, Oxford: Oxford University Press.(A recent effort to distinguish concepts from properties, and to provide a logical framework for discussing them both.)

Bealer, G. (1987) 'The Limits of Scientific Essentialism', in J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 1, *Metaphysics*, Atascadero, CA: Ridgeview Press.(An argument that externalist theories of meaning and concepts must rely on certain internal conditions, knowable a priori.)

Berkeley, G. (1710) *A Treatise concerning the Principles of Knowledge*, Indianapolis, IN: Hackett Publishing Company, 1982.(Thirteen raises an important problem for an empiricism that derives 'ideas' too directly from sense 'impressions'.)

Bilgrami, A. (1992) *Belief and Meaning: The Unity and Locality of Mental Content*, Oxford: Blackwell.(A strategy for how meaning holism can be made compatible with interpersonal comparisons of meaning and concepts.)

Block, N. (1986) 'Advertisement for a Semantics for Psychology', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 10, *Studies in the Philosophy of Mind*, Minneapolis, MN: University of Minnesota Press.(A defence of an inferential role theory of meaning and concepts as involving merely similarity, not identity among those roles.)

Burge, T. (1979) 'Individualism and the Mental', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 4, Minneapolis, MN: University of Minnesota Press.(Subtle defence of an externalist account of concept possession.)

Carnap, R. (1952) *Fundamentals of Concept Formation in Empirical Science*, Chicago, IL: University of Chicago Press.(Good example of a work in which concepts and properties are identified.)

Chisholm, R. (1957) *Perceiving: A Philosophical Study*, Ithaca, NY: Cornell University Press.(Classic critique of 'phenomenalist' and 'behaviourist' efforts to 'reduce' talk about material objects to talk about sensations, and talk about mental states to talk about external behaviour.)

Devitt, M. (1981) *Designation*, New York: Columbia University Press.(A defence of a causal theory of the reference of names - plausible as part of an externalist account of concepts of individuals.)

Devitt, M. (1995) *Coming to Our Senses*, Cambridge: Cambridge University Press.(A defence of a moderate conceptual molecularism, combining an externalist account of some non-sensory concepts, but also an inferential role account of others.)

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(An effort to base an externalist theory of concepts on the notion of 'information'.)

Dretske, F. (1988) *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: MIT Press.(Further development of an informational semantics, linking informational roles to behavioural consequences.)

Duhem, P. (1914) *The Aim and Structure of Physical Theory*, trans. P. Wiener, New York: Atheneum, 1954.(Early statement of confirmation holism and, thereby, for an empiricist like Quine, for meaning, or conceptual, holism.).

Evans, G. (1981) 'Semantic Theory and Tacit Knowledge', in S. Holtzman and C. Leich (eds) *Wittgenstein: To Follow a Rule*, London: Routledge & Kegan Paul, 118-37.(Discussion of the view that knowledge of the semantics of one's language might be tacit.)

Fodor, J.A. (1975) *The Language of Thought*, New York: Crowell.(Now classic statement of the claim that thought consists in computations on representations encoded in the brain, presumed throughout much cognitive psychology and in this entry; also, independently, initiates the argument that almost all concepts are innate.)

Fodor, J.A. (1979) 'The Present Status of the Innateness Controversy', in *RePresentations: Essays on the Foundations of Cognitive Science*, Cambridge, MA: MIT Press.(Rich discussion of the issues underlying the claim that almost concepts are innate.)

Fodor, J.A. (1991) *A Theory of Content*, Cambridge, MA: MIT Press.(Vigorous defence of a radically externalist theory of meaning and concepts - although he tends to use the latter term for the internal representation of an externally determined 'content'.)

Fodor, J.A. (1998) *Concepts: Where Cognitive Science Went Wrong*, Oxford: Oxford University Press.(Extended critique of conceptual role and other analyses of concepts in current linguistics, philosophy and psychology.)

Fodor, J.A. and LePore, E. (1992) *Holism: A Shoppers' Guide*, Oxford: Blackwell. (A critique of arguments for semantic and conceptual holism.)

Frege, G. (1892a) 'Über Begriff und Gegenstand', *Vierteljahrsschrift für wissenschaftliche Philosophie* 16: 192-205; trans. P.T. Geach, 'On Concept and Object', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980. (Frege himself used

‘concept’ for something like what is more commonly called a ‘property’; but the subtle issues he raises here apply to both.)

Frege, G. (1892b) ‘Über Sinn und Bedeutung’, *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, ‘On Sense and Reference’, in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980.(*Locus classicus* of a ‘sense’ theory of meaning and what many regard as concepts.)

Gettier, E. (1963) ‘Is Justified True Belief Knowledge?’, *Analysis* 23: 121-3.(Short and highly influential article presenting counterexamples to the traditional ‘analysis’ of [knowledge]; excellent as a paradigm of the philosopher’s interest in concepts.)

Goodman, N. (1951) *Structure of Appearance*, Indianapolis, IN: Bobbs-Merrill, 1966.(A classic effort at empiricist analysis of objective into sensory concepts, wholly within an extensional conception.)

Goodman, N. (1970) ‘Seven Strictures on Similarity’, in L. Foster and J. Swanson (eds) *Experience and Theory*, Amherst, MA: University of Massachusetts Press.(Excellent discussion of the perils of appealing to ‘similarity’ in theories of concepts.)

Grice, H.P. (1957) ‘Meaning’, *Philosophical Review* 66: 377-88.(Classic sketch of a programme for deriving the meaning of sentences from the intentions of speakers.)

Harman, G. (1972) *Thought*, Princeton, NJ: Princeton University Press.(Defence of an inferential role theory of meaning and concepts as involving merely similarity, not identity among those roles.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised P.H. Nidditch, Oxford: Clarendon Press, 1975.(Classic development of an early empiricist theory of concepts.)

Jackendoff, R. (1987) *Consciousness and Computation*, Cambridge, MA: MIT Press.(Lively and imaginative internalist account of conceptual structure as it appears to be revealed by the semantics of natural language.)

Jackson, F. and Pettit, P. (1995) ‘Moral Foundationalism and Moral Motivation’, *Philosophical Quarterly* 45 (178): 20-40.(Recent ‘two-factor’ proposal to provide definitions by the role of terms in ordinary thought, which, given a context of utterance, determines a propositional content.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, New York: St. Martin’s Press, 1968.(Classic study of the basic concepts arguably required for any understanding of an objective world.)

Kaplan, D. (1979) ‘Dthat’, in P. French, T. Uehling and H. Wettstein (eds) *Contemporary Studies in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press, 383-400.(Highly influential development of a ‘direct reference’ theory of demonstratives (for example, ‘that’, ‘this’), indexicals (‘I’, ‘her’) and other expressions, according to which ‘concepts’ play much less of a role than traditionally supposed.)

Katz, J. (1971) *The Underlying Reality of Language and its Philosophical Import*, New York: Harper & Row. (Early proposal modelling knowledge of semantics and concepts on the kind of tacit knowledge of syntax postulated by Chomsky.)

Katz, J. (1972) *Semantic Theory*, New York: Harper & Row.(Defence of the classical view on behalf of a semantics of natural language.)

Kripke, S.A. (1972) ‘Naming and Necessity’, in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel, 252-355; expanded version published as *Naming and Necessity*, Oxford and New York: Blackwell, 1980.(One of the original defences of an externalist account of proper names and natural kind terms; suggestive of a similar account of the corresponding concepts.)

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language*, Cambridge, MA: Harvard University Press.(An attack on an internalist account of concept possession, and a suggestion of an externalist ‘social’ theory instead.)

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(Influential defence of a holistic, inferential role theory of concept possession.)

Laurence, S. and Margolis, E. (eds) (1998) *Concepts*, Cambridge, MA: MIT Press.(An excellent collection of leading articles on the topic by philosophers, psychologists and logicians.)

Lewis, D.K. (1972) ‘General Semantics’, in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel.(Systematic presentation of a theory of meaning and concepts as intension-functions on possible worlds.)

Locke, J. (1689) *An Essay concerning Human Understanding*, New York: Dutton (Everyman), 1961.(The original proposal of an empiricist theory of concepts.)

Lormand, E. (1996) ‘How to Be a Meaning Holist’, *Journal of Philosophy* 93: 51-73.(Defence of the view that a token representation may have many different meanings, for different explanatory purposes.)

Millikan, R. (1984) *Language, Thought and Other Biological Categories*, Cambridge, MA: MIT Press.(The original statement of an evolutionary/teleological theory of meaning and concepts.)

Montague, R. (1974) *Formal Philosophy*, ed. R. Thomason, New Haven, CT: Yale University Press.(A systematic development of a theory of meaning and concepts as intension-functions on possible worlds.)

Osherson, D. and Smith, E. (1982) 'Gradedness and Conceptual Combination', *Cognition* 12: 299-318. (Examination of the difficulties of conceptual combination for prototype theories.)

Papineau, D. (1987) *Reality and Representation*, Oxford: Blackwell.(Defence of an evolutionary approach to concept ascription.)

Peacocke, C. (1992) *A Study of Concepts*, Cambridge, MA: MIT Press.(Sophisticated and influential defence of a 'sense' theory of concepts.)

Plato (395-347 BC) *Euthyphro*, in *The Collected Dialogues of Plato including the Letters*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1961.(Raises an important kind of question for conceptual analysis: is something good because the gods love it, or do they love it because it is good?)

Putnam, H. (1970) 'On Properties', in *Philosophical Papers*, vol. 1, Cambridge: Cambridge University Press, 1975, 305-22.(Excellent discussion distinguishing properties from concepts.)

Putnam, H. (1975) 'The Meaning of "Meaning"', in *Philosophical Papers*, vol. 2, *Mind, Language, and Reality*, Cambridge: Cambridge University Press.(Highly influential, lively and readable defence of an 'externalist' theory of meaning and concepts.)

Quine, W.V. (1953) 'Two Dogmas of Empiricism', in *From a Logical Point of View and Other Essays*, Cambridge, MA: Harvard University Press.(Important source of attacks on the scientific significance of conceptual analysis and on the classical view.)

Quine, W.V. (1954) 'Carnap and Logical Truth', in *Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1976.(Superb essay on the history of 'conceptual analysis', as well as important arguments on behalf of his meaning holism and against the scientific significance of analyses.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Classic exposition of an account of mind and language without appeals to concepts or meanings.)

Rey, G. (1983) 'Concepts and Stereotypes', *Cognition* 15: 237-62; repr. in S. Laurence and E. Margolis (eds) *Concepts*, Cambridge, MA: MIT Press, 1998.(Criticism of prototype theories for conflating epistemological with metaphysical issues.)

Rey, G. (forthcoming) 'A Naturalistic A Priori', *Philosophical Studies*.(Argues that the classical view can be defended against Quinian attacks by supposing analyses are sub-doxastic, like the rules of grammar.)

Rips, L. (1995) 'The Current Status of Research on Concept Combination', *Mind and Language* 10 (1/2): 72-104. (Excellent review of psychological research on concepts to that date.)

Rosch, E. (1973) 'On the Internal Structure of Perceptual and Semantic Categories', in T.E. Moore (ed.) *Cognitive Development and Acquisition of Language*, New York: Academic Press.(Influential discussion of the role of 'prototypes' in categorization tasks.)

Russell, B. (1903) *Principles of Mathematics*, New York: Norton, 2nd edn, 1938.(Chapter 4 wrestles with the problem of distinguishing a sentence from mere lists of words, which involves thinking about how a general concept is related to the individual (concept?) that satisfies it; connected to discussions in Frege (1892a) and Wittgenstein (1921).)

Russell, B. (1956) *Logic and Knowledge, Essays 1901-1950*, ed. R.C. Marsh, London: Allen & Unwin; repr. London: Routledge, 1992.(Influential essays presenting a highly 'referential' conception of the constituents of thoughts and propositions.)

Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press/Bradford Books. (Argues against 'sense' theories of linguistic meaning.)

Slote, M. (1975) 'Necessarily Inapplicable Concepts', *Philosophical Studies* 28: 265-71.(Argues that certain familiar concepts, such as [miracle] and [evil], could not possibly be instantiated.)

Smith, E. and Medin, D. (1981) *Concepts and Categories*, Cambridge, MA: Harvard University Press.(Useful discussion of psychological research supporting prototype theories.)

Sober, E. (1982) 'Why Logically Equivalent Predicates May Pick Out Different Properties', *American Philosophical Quarterly* 19: 183-9.(Argues that predicates such as 'equiangular' and 'equilateral' may pick out not only different concepts, but also different properties.)

Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press/Bradford Books.(Clear and readable defence of a theory of meaning and concepts as intension-functions on possible worlds.)

Stich, S. (1983) *From Folk Psychology to Cognitive Science: The Case Against Belief*, Cambridge, MA: MIT Press.(Proposes an essentially pragmatic approach to concept ascription across people; good discussion of intuitions leading us to different descriptions.)

Wittgenstein, L. (1921) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London:

Routledge & Kegan Paul, 1961.(Proposes an interesting ‘picture’ theory of meaning to solve the problem of how predicates combine with names to form sentential representations; suggestive of a similar treatment of mental representations and/or concepts.)

Wittgenstein, L. (1953) *Philosophical Investigations*, New York: Macmillan.(The source of ‘use’ and ‘inferential role’ theories of concept possession, as well as of the suggestion that concepts might have a prototype or ‘family resemblance’ structure.)

Zadeh, L. (1982) ‘A Note on Prototype Theory and Fuzzy Sets’, *Cognition* 12: 291-7.(Application of the formalism of ‘fuzzy set theory’ to a prototype theory of concepts.)

Conceptual analysis

A distinction must be made between the philosophical theory of conceptual analysis and the historical philosophical movement of Conceptual Analysis.

The theory of conceptual analysis holds that concepts - general meanings of linguistic predicates - are the fundamental objects of philosophical inquiry, and that insights into conceptual contents are expressed in necessary 'conceptual truths' (analytic propositions). There are two methods for obtaining these truths:

(1) direct a priori definition of concepts;
(2) indirect 'transcendental' argumentation.

The movement of Conceptual Analysis arose at Cambridge during the first half of the twentieth century, and flourished at Oxford and many American departments of philosophy in the 1950s and early 1960s. In the USA its doctrines came under heavy criticism, and its proponents were not able to respond effectively; by the end of the 1970s the movement was widely regarded as defunct. This reversal of fortunes can be traced primarily to the conjunction of several powerful objections: the attack on intensions and on the analytic/synthetic distinction; the paradox of analysis; the 'scientific essentialist' theory of propositions; and the critique of transcendental arguments. Nevertheless a closer examination indicates that each of these objections presupposes a covert appeal to concepts and conceptual truths. In the light of this dissonance between the conventional wisdom of the critics on the one hand, and the implicit commitments of their arguments on the other, there is a manifest need for a careful re-examination of conceptual analysis.

1 Origins and career of Conceptual Analysis

Many of the elements of Conceptual Analysis are present already in John Locke's *Essay Concerning Human Understanding* (1689) - in his doctrines of general ideas and definitions (decompositions of complex general ideas into sets of simple ideas); in his distinction between 'trifling' and 'instructive' universally certain propositions; and in his closely related distinction between 'intuitive' and 'demonstrative' knowledge (see Locke, J. §§3-4). An even more important source of influence is Immanuel Kant's *Critique of Pure Reason* (1781/1787). There Kant makes three crucial sets of distinctions. The first is between 'analytic' and 'synthetic' propositions, that is, between propositions true by virtue of conceptual content alone, and propositions true by virtue of conceptual content together with a non-conceptual semantic element ('intuition'). The second is between a priori (necessary, experience-independent) and a posteriori (contingent, experience-dependent) truths. The third is a threefold division between proofs by empirical methods, 'constructive' proofs in mathematics, and 'transcendental' proofs. Transcendental proofs establish the truth of non-mathematical synthetic a priori propositions by showing how the natural sciences - and human experience itself - presuppose a set of primitive pure concepts or 'categories' (see Kant, I. §4).

Kant's important idea that conceptual truths can be either analytic a priori or synthetic a priori is effectively erased by Gottlob Frege in his *Foundations of Arithmetic* (1884). Frege's overriding philosophical aim is to put mathematical proof on a firm footing by reducing the truths of arithmetic to analytic truths of logic. In view of this, the proper goal of an analysis is the production of non-circular, explanatory, yet meaning-preserving general definitions of fundamental concepts - as exemplified in Frege's famous definition of a number as a class of equinumerous classes (see Logicism; Frege, G. §§7-8). Analytic a priori truths for Frege are propositions that follow deductively either from the self-evident, unprovable laws of pure logic alone, or else from the laws of logic together with logical definitions.

Frege's method was enthusiastically developed and subtly transformed by G.E. Moore in what may be regarded as the first phase of Conceptual Analysis. Moore supplemented Frege's austere logicism with a Kant-inspired attentiveness to the multiplicity of different sorts of concepts, propositions and logico-semantic relations, and with a predilection for arguments resting on appeals to common sense (see Moore, G.E. §§3-4). Moorean analysis then travelled from Cambridge to Oxford, where J.L. Austin and Gilbert Ryle added to it a special focus on the uses and abuses of ordinary language. This led directly to the vigorous growth in the 1950s of the second phase of Conceptual Analysis, sometimes also called 'Oxford Philosophy'. Conceptual analysis was exported to the USA in

the 1950s and early 1960s, primarily through the writings of H.P. Grice and P.F. Strawson. While it found a niche for a time in many American philosophy departments, it did not ultimately survive. It was attacked on several fronts by leading American philosophers (most damagingly, perhaps, by W.V. Quine - see §3 of this entry for details), and by the end of the 1970s had largely succumbed.

2 The theory and methods of conceptual analysis

The career of Conceptual Analysis was rather brief and embattled, but its underlying philosophical theory, conceptual analysis, should be analysed and judged on its own merits. For simplicity's sake, we can think of conceptual analysis as defined by the conjunction of the following five theses:

(1) *The content thesis.* A concept is a general content possessing intrinsic, individuating structures and relations (an intension), and having a corresponding application either to sets of actual or possible objects (an extension), or to other concepts.
(2) *The linguistic thesis.* A concept is the meaning of a predicate-expression; and all such words have meanings only in the context of whole sentences used (first and foremost) in making statements in ordinary discourse.
(3) *The modal thesis.* Every true proposition expressing conceptual interconnections is necessary and analytic.
(4) *The knowledge thesis.* Purely conceptual inquiry produces important a priori knowledge. This knowledge is expressed in analytic propositions known to be true either by (a) direct definitional analysis of conceptual contents, or by (b) indirect 'transcendental' arguments.
(5) *The metaphilosophical thesis.* All fundamental philosophical errors arise from misunderstandings of concepts, and can be corrected only by proper conceptual analyses.

The first two theses convey a theory of concepts. Being general, concepts play the role traditionally assigned to universals (see Universals). Yet because they are ontologically dependent upon ordinary language, concepts are not otherworldly, Platonic entities. And because concept-possession depends upon linguistic use and mastery, concepts are immediately accessible to all competent speakers.

Concepts bear necessary relations to one another and also have fixed internal structures; these relations and structures are open to the process of analysis; and a capacity for analytical insight is guaranteed by linguistic mastery. Concepts, however, are of two quite different sorts: 'categorematic' and 'syncategorematic'. Categorematic concepts (for example, 'bachelor' or 'being taller than') are 'material' intensional contents that uniquely and independently determine concept-extensions. Syncategorematic concepts, by contrast, are 'formal' intensions that apply in a rule-like way to other concepts or conceptual complexes. These in turn are of two sorts:

(1) 'logical concepts' (such as 'conjunction') expressing logical operations; and
(2) 'categorical concepts' (such as 'objecthood') expressing higher-order conditions of the applicability of lower-order concepts.

Analytical insight into categorematic and syncategorematic concepts permits the capture of both non-logical and logical truths (for example, 'bachelors are unmarried males' and '$\sim(P \;\&\; \sim P)$') within the general class of conceptual truths.

The modal thesis tells us that all conceptual truths are analytic and necessary; such truths reflect conceptual contents alone and bear no connections to the disposition of things in the actual world or any possible world. They are therefore 'topic-neutral'. This makes is relatively easy to see why, as the knowledge thesis asserts, the cognition of analytic propositions is a priori: the insight into conceptual content requires no appeal to empirical facts or individuals. And certainly in the case of such simple definitional propositions as 'bachelors are unmarried males', it appears to be the case that a direct awareness of conceptual identity - guaranteed by linguistic competence and the grasp of word-synonymy - requires no appeal to experience in order to be known. But the very idea of a conceptual identity is not so simple as one might think; nor does insight into conceptual truth always result from definitional inquiries alone. This is manifest in three ways.

In the first place, definitional truths do express conceptual identities or synonymies of words, but the criterion of identity cannot be merely that concepts are identical, or words synonymous, when they share the same extensions necessarily. The concepts 'creature with a heart' and 'creature with a kidney', for example, share actual extensions, but are clearly not identical. And as C.I. Lewis first pointed out in 'Modes of Meaning' (1943-4) there

are also concepts - such as ‘equilateral triangle’ and ‘equiangular triangle’ - that are necessarily co-extensional, but not precisely identical. Hence a stricter criterion of conceptual identity, involving an isomorphism of the concepts’ internal structures, must be invoked.

In the second place, there are analytic propositions expressing conceptual relations that reflect only partial identities of concepts, for example: (A) ’bachelors are males’. And most logical truths appear not to reflect *either* complete *or* partial conceptual identities. Here, however, it is possible to appeal to a criterion of analyticity used by Kant, namely that the denial of an analytic proposition leads to a contradiction. This is closely connected with the idea that when terms in partially definitional propositions *are* replaced by their full definitions, or perfect synonyms, logical truths will result. Thus substituting ’unmarried males’ into (A) for ’bachelors’ produces the logical truth, (A*) ’unmarried males are males’. The denial of (A*) is obviously logically contradictory. So an a priori grasp of conceptual identities and logical concepts appears to be sufficient for knowledge of definitional propositions and logical truths alike.

Third, however, there are conceptually true propositions, such as (B) ’nothing can be simultaneously coloured in two different ways all over’ and (C) ’the world as we experience it contains reidentifiable objective particulars in a single spatiotemporal scheme’, that do not seem to reflect logical truths, or even complete or partial identities of concepts, but rather conceptual connections of a somewhat different sort. Here we are strongly reminded of Kant’s view that some conceptual truths are not analytic, but instead synthetic. And indeed, although conceptual analysts generally eschew the existence of the synthetic a priori, this is precisely where the appeal to transcendental arguments comes in. A transcendental argument aims to show that a proposition *P* (say, (B) or (C)) is conceptually true because it is presupposed by another proposition *Q* (say, ’*a* is red; so it is not green’ or ’*a* is not being perceived by me now; but it is still over there just the same’), which is taken by hypothesis to be perfectly acceptable and a ‘paradigm case’ of some class of statements. Not only is *P* a necessary condition of the truth of *Q*, but more profoundly *P* is a necessary condition of the *real possibility* or *meaningfulness* of *Q* - of its being true or false in the first place. This is because the concepts expressed in *P* are categorial concepts having a ’conceptual priority’ over the concepts expressed in *Q*, which is to say that the concepts in *P* have a central place in the overall ’conceptual scheme’ by which language-using human beings organize their common sense experience in the ways exemplified by *Q*. Thus the conceptual truth (B) expresses an insight about the very nature of human experience of colour; and the conceptual truth (C) expresses an insight about the very nature of human perception of objects. Transcendental arguments thus extend the scope of conceptual analysis from the mere definitional or logical exploration of conceptual contents (sometimes also called ’philosophical grammar’), towards insights into first principles expressing the ‘conceptual geography’ of the common sense world (see Transcendental arguments).

The metaphilosophical thesis follows directly from the other four. Concepts govern the ways we think about all things and other concepts; thus not only all philosophical truths, but also all philosophical errors, are conceptual. The two methods of conceptual analysis - definitional and transcendental - must be employed not merely as means of philosophical insight but also for the unmasking and diagnosis of conceptual confusions.

3 Five fundamental objections

The many different lines and styles of criticism directed against conceptual analysis tend to converge on five basic objections:

(1) *The flight from intensions.* If concepts are linguistic intensions, then obviously any sceptical argument showing that intensions do not exist will undermine the linguistic thesis. Just such an argument has been influentially promoted by Quine, in two parts. First, intensions are said to be either ontologically ’mysterious’ or purely psychological entities that intervene between language (or linguistic behaviour) and reference, and should be ruled out of any properly logical and scientific approach to semantic issues. Second, all the explanatory roles traditionally played by intensions - as what words signify, as truth-vehicles, as grounds of synonymy, as grounds of modality, as objects of the propositional attitudes, and as objects of philosophical analysis - can be functionally mimicked by logical or linguistic devices that make no appeals whatsoever to intensional entities (see Intensional entities).
(2) *The death of analyticity.* Perhaps even more famous than Quine’s attack on intensions is his attack, in ‘Two Dogmas of Empiricism’ (1951), on the very idea of analyticity (see Quine, W.V. §8). Setting aside logical truth, Quine argues that non-logical analyticity is based on the concept of synonymy. But every plausible attempt to give

an explanation of synonymy (by appeal to the notions of definition, linguistic interchangeability, or semantical rules) ends either in circularity or vacuity. In the absence of a clear account of synonymy, no clear boundary between analytic and synthetic (factual, contingent) propositions can be established. If sound, this argument forces the rejection of the modal thesis (see Analyticity).
(3) *The paradox of analysis.* In 'Moore's Notion of Analysis' (1942), C.H. Langford points up a deep difficulty in the conception of a definitional analysis. In order for a proposition expressing the results of such an analysis to be correct or true, it must establish a complete or partial identity between concepts. But if an identity is so established, then the very same concept, wholly or in part, redundantly shows up twice in the same conceptual truth, as expressed by two different words or phrases. Thus every correct definitional analysis of a concept is non-informative and trivial; and the very project of definitionally analysing a concept is epistemically pointless. If true, it follows directly from the paradox that the first part of the Knowledge Thesis, which states that all conceptual truths express important a priori knowledge, is false.
(4) *Scientific essentialism and the contingency of conceptual truths.* Enshrined in the linguistic thesis, the modal thesis and the knowledge thesis, are claims to the effect that the meanings of words are conceptual intensions, and that conceptual truths are analytic, a priori and necessary. But it has been influentially argued by Hilary Putnam (1975) that the extensions of some general words - 'natural-kind' terms such as 'water' or 'cats' - are not in fact determined by their corresponding concepts. The extension of a natural-kind word, says Putnam, is instead determined by a strict relation of identity between the natural kind and the microphysical stuff that locally predominates in the samples used by scientists in their empirical investigations (see Reference §3). The stuff's physical microstructure - say, water's being H_2O - is its scientific *essence*; and propositions expressing this essence - say, 'Water is H_2O' - are necessary and a posteriori. But this immediately implies that the conceptual propositions expressed by the use of sentences including natural-kind terms will not be necessary truths. For example, consider the apparently necessary (because analytic by partial definition) proposition 'Water is a liquid'. The natural-kind word 'water' will pick out only whatever stuff in a given world has the microstructure H_2O. But it is possible that on a different world, under different physical conditions, the stuff that is H_2O and a liquid here on Earth might look very different and have very different surface properties: it might be solid, for example. So the conceptual proposition 'Water is a liquid' is false in that possible world; and thus it is only contingently true in the actual world, even if grasped a priori.
(5) *Transcendental arguments presuppose verificationism.* Even supposing that definitional conceptual truths are empty tautologies, and not always necessary, still conceptual truths gained by transcendental arguments would remain cognitively significant and modally secure. But as Barry Stroud (1968) has pointed out, the theory of transcendental arguments assumes a strongly verificationistic theory of meaningfulness for concepts and propositions. Verificationism, however, is afflicted with insurmountable problems. So transcendental arguments are semantically suspect, and the second part of the Knowledge Thesis would thereby seem to be undermined too.

4 The inescapability of conceptual analysis

On the assumption that these criticisms are sound, things look very bleak for conceptual analysis. And it is true that the movement of Conceptual Analysis did eventually break up under the weight of the criticisms just described. But a closer inspection reveals a striking feature of the philosophical dialectic: In order to gain the acceptance of any argument aimed *against* conceptual analysis, it appears that the critic must finally appeal to the truth of some premises that implicitly invoke concepts and conceptual truths.

To take only one central example. Quine's famous arguments against intensions and analyticity all assume the notion of a logical truth. According to Quine in 'Truth by Convention' (1936), a logical truth is a sentence that contains certain words (logical constants) 'essentially': these words preserve their interpretations across every uniform assignment of values to the non-logical constants in the sentence, ensuring that it 'comes out true' no matter what. Now logical constants, with their 'essential occurrence', are semantically equivalent to the conceptual analyst's *logical concepts*; and in this way Quinean logical truths are (covertly) conceptual truths. Moreover, although Quine suggests in 'Two Dogmas of Empiricism' (1951) that even logical truths are revisable, he states in his later *Philosophy of Logic* that 'every logical truth is obvious, actually or potentially' ([1970] 1986: 82), and that the very attempt to deny a logical truth would involve a change of meaning of the logical constants. In other words, logical constants and logical truths are ineliminable parts of any rational conceptual scheme recognizable as

our own. This recognition is epistemically equivalent to what conceptual analysts mean by the a priori grasp of a conceptual truth; the only difference is that whereas most analysts hold that logical truths are known and justified by direct conceptual insight, Quine persuasively appeals instead to a transcendental proof.

If sound, this argument smoothly generalizes. No philosopher can do without logic; and if logic is itself necessarily such as to contain concepts and conceptual truths that are grasped a priori, and whose existence and validity can be established only via transcendental argument, then no philosopher can ultimately avoid the analysis of concepts. Supposing that conceptual analysis is - even in this minimalistic way - philosophically inescapable, the demand for a re-examination and re-working of its basic theses seems self-evident.

See also: Analytical philosophy; Concepts; Meaning and verification

ROBERT HANNA

References and further reading

Austin, J.L. (1961) *Philosophical Papers*, ed. J.O. Urmson and G.J. Warnock, Oxford: Oxford University Press, 3rd edn, 1979.(Conceptual analyses with a bias towards ordinary language; see especially essays 2, 4-5, 8 and 11-2.)

Flew, A. (ed.) (1956) *Essays in Conceptual Analysis*, London: Macmillan.(Essay 6, by Urmson, discusses the ’paradigm case argument’.)

Frege, G. (1884) *The Foundations of Arithmetic*, trans. J.L. Austin, Evanston, IL: Northwestern University Press, 2nd edn, 1953.(See especially §§1-4, 12-7 and 55-91.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.(A thematically organized collection of Grice’s writings; see especially chapters 1, 10-1, 13, 15, and the ’Retrospective Epilogue’.)

Hempel, C. (1950) ’Problems and Changes in the Empiricist Criterion of Meaning’, *Revue Internationale de Philosophie* 11: 41-63.(A summary of objections to verificationism.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N.K. Smith, London: Macmillan, 1929.(See especially the introduction.)

Langford, C.H. (1942) ’Moore’s Notion of Analysis’, in P.A. Schilpp (ed.) *The Philosophy of G.E. Moore*, New York: Tudor, 2nd edn, 1952.(For Moore’s reply to Langford, see part III.iii.)

Lewis, C.I. (1943-4) ‘Modes of Meaning’, *Philosophy and Phenomenological Research* 4: 236-49.(A sketch of basic principles of conceptual-intensional semantics.)

Locke, J. (1689) *Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1975. (See especially books III and IV.)

Moore, G.E. (1922) *Philosophical Studies*, London: Kegan Paul, Trubner, Trench & Co.(A collection of Moore’s early essays.)

Moore, G.E. (1959) *Philosophical Papers*, London: George Allen & Unwin.(A collection of Moore’s later essays, with special applications of conceptual analyses to anti-sceptical arguments.)

Putnam, H. (1975) ’The Meaning of "Meaning"’, in K. Gunderson (ed.) *Midwest Studies in Philosophy: Language, Mind, and Knowledge*, Minneapolis, MN: University of Minnesota Press, 131-93. (Defends scientific essentialism.)

Quine, W.V. (1936) ’Truth by Convention’, repr. in H. Geigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, 1949, 250-73.(An anticipation of many of the main themes in ‘Two Dogmas’.)

Quine, W.V. (1951) ’Two Dogmas of Empiricism’, repr. in *From a Logical Point of View*, New York: Harper & Row, 1961, 20-47.(Rejects the very idea of analytic truths.)

Quine, W.V. (1970) *Philosophy of Logic*, Cambridge, MA: Harvard University Press, 2nd edn, 1986.(Discusses the nature of logical truth.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(See especially chapters 2 and 6.)

Ryle, G. (1971) *Collected Papers*, vol. 2, London: Hutchinson.(Exercises in conceptual analysis, mostly deflationary in tone and import; see especially essay 12 for a treatment of syncategorematic concepts.)

Strawson, P.F. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen.(Conceptual Analysis takes a ’transcendental turn’.)

Strawson, P.F. (1992) *Analysis and Metaphysics*, Oxford: Oxford University Press.(An introduction to Strawsonian conceptual analysis.)

Stroud, B. (1968) ’Transcendental Arguments’, *Journal of Philosophy* 65: 241-56.(An influential discussion.)

Condillac, Etienne Bonnot de (1715-80)

One of the leading figures of the French Enlightenment period, Condillac is the author of three highly influential books, published between 1746 and 1754, in which he attempted to refine and expand the empirical method of inquiry so as to make it applicable to a broader range of studies than hitherto. In the half-century following the publication of Newton's Principia Mathematica *in 1687, intellectual life in Europe had been engaged upon a fierce debate between the partisans of Cartesian physics, who accepted Descartes' principles of metaphysical dualism and God's veracity as the hallmark of scientific truth, and those who accepted Newton's demonstration that the natural order constituted a single system under laws which could be known through painstaking observation and experiment. By the mid-eighteenth century Newton had gained the ascendancy, and it was the guiding inspiration of the French thinkers, known collectively as the* philosophes, *to appropriate the methods by which Newton had achieved his awesome results and apply them across a broader range of inquiries in the hope of attaining a similar expansion of human knowledge. Condillac was at the centre of this campaign.*

Condillac's first book, An Essay on the Origin of Human Knowledge (1746), bears the subtitle A Supplement to Mr. Locke's Essay on the Human Understanding. While Condillac is usually seen as merely a disciple and popularizer of Locke offering little of any genuine originality, and while he did indeed agree with Locke that experience is the sole source of human knowledge, he attempted to improve on Locke by arguing that sensation alone - and not sensation together with reflection - provided the foundation for knowledge. His most famous book, the Treatise on the Sensations (1754) is based upon the thought-experiment of a statue whose senses are activated one by one, beginning with the sense of smell, with the intention of showing how all the higher cognitive faculties of the mind can be shown to derive from the notice the mind takes of the primitive inputs of the sense organs. Condillac also went beyond Locke in his carefully argued claims regarding the extent to which language affects the growth and reliability of knowledge. His Treatise on Systems (1749) offers a detailed critique of how language had beguiled the great seventeenth-century systems-builders like Descartes, Leibniz and Spinoza and led them into erroneous conceptions of the mind and human knowledge, the influence of which conceptions was as insidious as it was difficult to eradicate.

1 Life and historical context

Etienne Bonnot, Abbé de Condillac, was born in Grenoble. His father was a successful lawyer and the family was prosperous. He was educated for the priesthood and took holy orders in 1740. While he remained a priest and was accustomed to wearing the cassock throughout his life, it appears that he celebrated Mass only once and neglected for the most part the other duties which attended his office. Dissatisfied with the education provided for him, he devoted himself to the study of philosophy, immersing himself in the works of Descartes, Malebranche, Leibniz and Spinoza. He read no English, but acquired his knowledge of Locke through the French translation of the *Essay concerning Human Understanding* (1689) by Pierre Coste and read Voltaire's books on Newton. He was the contemporary of such Enlightenment luminaries as Helvetius, Diderot, Buffon, La Mettrie and Holbach. Voltaire and Diderot both expressed the very highest regard for his writings. Condillac was on friendly terms with Rousseau, and was frequently to be seen at the salons in and around Paris where so much of the intellectual activity of the Enlightenment took place. He spent nine years (1758-67) in Italy as tutor to the prince of Parma, and wrote a thirteen-volume course of studies which encompassed history, grammar, poetics and scientific methodology.

The Enlightenment was a period of intense intellectual ferment whose major currents are both multifarious and complex (see Enlightenment, Continental). It is possible nonetheless to isolate several trends and commitments which define more or less the general tendency of thought and the aspirations of those who worked to articulate them. The *philosophes* thought of themselves, perhaps first and foremost, as the champions of the new methods in the natural sciences. They could see important progress being made in so many areas, not only in physics under the aegis of Newton but in chemistry, in navigational technique, in medicine and so on. Their first commitment was to bring the method of the new sciences to bear on as many different lines of inquiry as possible, their guiding assumption being that there was no field of research that could be placed beyond the purview of scientific investigation. But this commitment to the ways of natural science inevitably brought them into conflict with the authority and power of the religious establishment. Though the prestige and political strength of the Catholic

Church had been eroding in France over the first decades of the eighteenth century, it was still a force to be reckoned with, especially where its authority to pronounce on matters of fundamental truth about the world and about the human place within it were challenged. In their attempts to advance the cause of science, the *philosophes* had to contend at every turn with what they called the powers of ‘superstition’ and ‘enthusiasm’, slightly veiled code words for the doctrinal chauvinism of the Church and what to their minds was its reactionary and stultifying influence upon the political institutions and social conditions of the day. These three currents, the defence and advancement of the cause of natural science, the challenge to the traditional authority and power of the Church and a political agenda directed towards the progress of humanity, define what may be taken as the essential thrust of the Enlightenment as an intellectual movement. Condillac’s own contribution to this movement must be set in this context. His philosophy as a whole may be understood as the attempt to hold science and theology together in a single systematic vision of nature, together with a more benign and accommodating perception of man’s place within the natural order.

2 Epistemology

All Condillac’s philosophy begins in the empiricism of John Locke, but he introduces a further methodological component which, he believes, takes him beyond Locke in his treatment of the nature of human knowledge. Disappointed with what appeared to him the idle speculation and dogmatism of the philosophers of the seventeenth century, he writes:

> It seemed to me that we might reason in metaphysics and in morals with as great exactness as in geometry; that we might frame as accurate ideas as the geometricians; that we might determine, as well as they, the meaning of words in a precise and invariable manner; in short that we might prescribe, perhaps better than they have done, a plain and easy order for the attainment of demonstration.
>
> (1746: 2)

Condillac’s epistemology represents the attempt to bring together Locke’s insistence that all knowledge begins with experience with the rigour of a quasi-geometrical method of inference. With Locke, Condillac rejects Descartes’ doctrine of innate ideas, but he wishes to preserve something of the latter’s deductive logic as we move from the simple ideas received through sensation to the more abstract and complex ideas of scientific theory (see Descartes, R.). This attempt to combine elements of Lockean empiricism with Cartesian rationalism is representative of one of the most important aspirations of the Enlightenment thinkers generally. Condillac’s historical importance lies in the fact that he, more than anyone else, tried to work out the details of such a combination with care and precision, and to formulate once and for all the method by which a sure and steady progress could be attained in virtually all areas of human inquiry. It was his hope, and that of all the French thinkers of the time, that the broader application of such a method would yield the same remarkable results in any number of different domains that Newton had achieved in physics through his combination of painstaking observation and mathematical rigour.

Condillac believed that he had made two important advances on Locke by showing, first, how sensation alone, and not sensation and reflection as Locke held, provided all the ideational resources necessary for knowledge and, second, how ‘signs’ or language provided the essential means for moving beyond our personal awareness of our sensations in the direction of objective knowledge. It is for the sake of demonstrating the first of these claims that Condillac constructed the elaborate thought-experiment of the living statue he discusses in his most famous book, the *Treatise on the Sensations* (1754). He asks the reader to imagine a statue whose senses are to be activated one at a time beginning with the sense of smell. His purpose is to show how all the higher cognitive faculties of the mind are generated out of the sequential occurrence of the most primitive of sensory experiences. The statue is presented with the smell of a rose. It immediately discerns the alteration of its ‘mode’ from the utter lack of sensation to the manifestation of the rose’s scent. This awareness of alteration within itself constitutes what Condillac calls ‘attention’, the first and most basic operation of the mind upon the input provided by sensation. The statue is then presented with the scent of jasmine. It attends to the alteration of its own mode and immediately perceives the difference between the rose and the jasmine. This immediate recognition of the difference between the two, and then again the recognition of identity when presented subsequently with the rose, is a manifestation of the mind’s power of judgment. ‘A judgment’, Condillac writes, ‘is only the perception of a relation between two ideas which are compared’, arising from the mind’s implicit capacity to attend to the ideas separately (1754). The

larger significance here is that identity and difference, two notions of fundamental importance for logic itself, are shown to derive from sensation and not to repose in the mind as innate principles or ‘clear and distinct ideas’ available a priori and suited to provide an initiation point for the construction of knowledge.

The statue is able not only to attend to the single presented sensation but to retain an ‘impression’ of the rose while attending to the jasmine. This is memory, which too, according to Condillac, is only a function of sensation. The mind, then, is both passive, in receiving sensations, and active, in making comparisons between sensations present and remembered. With this much of the mind’s capacities described, Condillac asks what might have occurred if the statue had first been presented with an unpleasant smell. If it lacked all experience of pleasant scents, the statue would not, in the presence of the unpleasant one, immediately have experience of anything we could call discomfort or pain, nor would it desire the removal of the unpleasant smell and a return to its former state of utter privation: ‘Suffering can no more make it desire a good it does not know than enjoyment can make it fear an ill it does not know’ (1754: 5). Pleasure and pain, desire and aversion are thus also shown to arise from primitive sensations, but only through the operation of the mind’s capacity for retention, comparison and judgment. Simply put, pleasure and pain are not themselves primitive sensations but learned responses to the mind’s reflection upon the various sensations which come through to conscious experience by way of the organs of sense.

None the less, Condillac claims, no sensations are completely indifferent, as any two of them will inspire within the statue not only a recognition of their difference but also a preference for one over the other. This immediately discernible preference of one sensation over another engenders, over time, the whole vast range of our desires and aversions; with the experience of desire and aversion the primitive capacities of the mind are magnified. On Condillac’s view, our preference for one sensation over another provides a stimulus to the power of attention, enhancing the mind’s ability to focus more keenly and resolutely on ideas stored in memory and leading in turn to a greater discernment and affinity for those sensations which are preferred. ‘If we bear in mind that there are absolutely no indifferent sensations’, Condillac writes, ‘we shall conclude that the different degrees of pleasure and pain are the law by which the germ of all we are is developed, and that they have produced all our faculties’ (1754: 46). Towards the end of the *Treatise* Condillac reaffirms this claim: ‘As without experience there is no knowledge, so without needs there is no experience, and there are no needs without the alternatives of pleasures and pains. Everything then results from the principle which we laid down at the beginning of this work’ (1754: 209).

In the ensuing pages of the *Treatise* Condillac’s statue is endowed, one by one, with the remaining four senses. Condillac attaches special importance to the sense of touch, arguing that it is only with the addition of this inlet of sensory experience that the statue comes to have ideas of space, extension and external objects, and to acquire awareness of itself as something distinct from its own representations of objects.

3 Knowledge and nature as systems

While the *Treatise on the Sensations* is Condillac’s most famous and influential text, it is only in the *Treatise on Systems* (1749) and several less well known works, especially *The Logic* (1780), that Condillac offers his most thoroughgoing exposition of what he conceived of as a system of human knowledge. Having demonstrated the merits of Lockean empiricism in the *Essay on the Origin of Human Knowledge*, he proceeds, in the *Treatise on Systems*, to offer a critical analysis of the rationalistic systems of thought contrived by the great seventeenth-century philosophers, Descartes, Malebranche, Leibniz and Spinoza. Condillac argues that it was their failure to understand the nature and function of language, and the need to derive even the most general and abstract terms from materials provided through sensation, which led these thinkers astray from the very outset of their labours. Much of their deductive method could be preserved, but each link in the chain of inferences must be subjected to linguistic scrutiny to assure that the developing system of concepts allowed in nothing that could not ultimately be shown to derive from sensory experience. ‘The language of philosophy has been nothing else but a gibberish for many centuries past. At length that gibberish has been banished from the sciences’ (1780: 64).

Condillac believed that the language of science might be brought to a clarity and precision analogous to mathematical symbols and equations through what he called the method of analysis. This method requires that we examine each term in our language and try to trace it back to the point at which it connects up with immediate experience. At a simpler, more primitive point in their history, Condillac conjectures, humans were able to make

an apt assignment of 'signs' to those things in our experience which stood out as particularly relevant to our needs and wants. The human capacity for speech is itself a natural endowment neatly insinuated amidst the various other faculties of the mind where it augments and facilitates their operation. 'There is an innate language, though there are no innate ideas' (1780: 56). Nature itself teaches us the method of analysis. So long as we had hitherto comported ourselves in the construction of our language in accordance with its natural tendency, and to the extent today that we come to understand better the ways in which it was originally adapted to our needs and wants, the knowledge we construct on the basis of our growing stock of abstract terms and concepts can be certified as valid and reliable. 'Error begins when nature ceases to inform us of our mistakes; that is, where judging of things which have little relation to the wants of first necessity, we do not know how to try our judgments in order to discover whether they are true or false' (1780: 6).

Condillac's method of analysis can and must be applied in other directions also. Once the basic terms of our scientific vocabulary are correlated accurately and unambiguously with experientially derived ideas, we are able to exercise the power of judgment over the propositions in which they occur in such a way as to extract and formulate with precision the relations and interconnections our linguistic signs implicitly bear to one another. In this way a systematic structure of concepts begins to emerge which in turn enables us to attend more carefully, especially through the use of controlled experiments, to the continued stream of input from sensation. Our knowledge of the world grows through a sequence of inferential steps, and the emergent system of knowledge could be imbued with the same precision and rigour as is to be found in mathematics itself if we would but attend to the clarity and accuracy of our basic terms.

There is much in Condillac which might plausibly be construed as anticipating certain basic themes in the mid-twentieth century movement of Logical positivism. It would be a mistake, however, to impute too great a prescience to Condillac as a precursor of later philosophical movements because there are other, deeply ingrained aspects of his thought which are wholly characteristic of his historical moment and utterly foreign to the later developments. According to Condillac, all human knowledge will ultimately fall together into one vast system of concepts which reflects with perfect accuracy the systematic structure of nature itself. In using the analytic method to trace the meanings of our words back to their origins in sensory experience we are only following out the order in which our ideas themselves were generated in our encounter with nature. The perfect congruence between human knowledge as a system of concepts and nature itself as a systematic whole is ultimately a reflection of the divine intelligence which designed and brought nature into existence. Human nature, and human intelligence as the highest expression of our nature, have been 'conformed' by the creator as an integral part of the natural order and all our needs, desires and cognitive faculties can be brought to reside in perfect harmony with that order. By coming to understand nature as a system we grow in our knowledge of God.

Condillac, in short, was a deist and his philosophy as a whole gives magnificent expression to the eighteenth-century conception of God as cosmic law-giver (see Deism). The laws of nature which Newton had articulated, and all the other laws of nature which science was in process of discovering, were written into the natural order at its inception. But just as nature had its laws, so too there was a set of laws governing human nature, every bit as objective and every bit as amenable to the method of analysis as any other domain of science.

> The faculties and wants of man being given, the laws are given themselves; and though we make them, yet God who created us with such wants and such faculties, is, in truth, our sole legislator. When, therefore we follow these laws which are conformable to our nature, it is Him we obey, and this accomplishes the morality of actions.
>
> (1780: 27)

It was not only human knowledge which could be brought into perfect systematic harmony between all the parts; human conduct too, all of our various personal and social interrelations could be harmonized in accordance with the law of human nature. It was one of the highest aspirations of the Enlightenment period, and of Condillac as one of its philosophically most astute representatives, to raise humankind to the more civilized and prosperous state that would emerge as we gained better knowledge of God's laws, both natural and human. The 'moral sciences' themselves - including the nascent sciences of economics, psychology and politics - could be fully integrated into the whole vast system of natural laws, and to the extent that they were thus integrated and conceived in such a way as to permit a thoroughgoing application of the method of analysis, it was only to be expected that our knowledge

of the human order and hence our ability to secure for ourselves a greater happiness in this world would be enhanced.

4 Reputation and influence

In his own day, Condillac enjoyed a very high reputation in French intellectual circles as the one writer who had devoted his energies to the elaboration of the theory of knowledge required to support the ideals and political agenda of the Enlightenment movement. His works continued to be published in new editions well into the nineteenth century. His philosophy provided the stimulus and inspiration for the early nineteenth-century thinkers known as the *idéologues*, such as Destutt de Tracy and Cabanis. Condillac is often cited in textbooks on the history of psychology as one of the first to bring methodical scrutiny to bear on the subtle dynamics of such cognitive functions as attention, memory and habit formation.

It remains a matter of scholarly controversy whether Condillac really made any important advances over Locke in his treatment of the empirical basis of knowledge. With regard to the two specific points on which Condillac himself believed he had gone beyond Locke, in reducing the source of ideas to sensation alone instead of sensation and reflection and in his theory of language, there is much that could be argued in Locke's defence. Nonetheless, it can hardly be denied that Condillac brought a much greater coherence and clarity to Locke's often tentative and disorganized discussions in the *Essay concerning Human Understanding*. By doing so he is responsible for bringing the empirical spirit of inquiry into a much broader currency on the Continent that it might otherwise have enjoyed. Condillac has not yet been given the careful attention that he deserves, and the full story of how his philosophy has influenced and shaped the later course of the history of ideas has yet to be written.

See also: Empiricism; Newton, I.; Rationalism

PAUL F. JOHNSON

List of works

Condillac, E.B. de (1947-51) *Œuvres philosophiques*, ed. G. Le Roy, Paris: Presses Universitaires de France, 3 vols.

(1746) *Essai sur l'origine des connaissances humaines: ouvrage où l'on réduit à un seul principe tout ce qui concerne l'entendement humain*, trans. T. Nugent, in J.H. Stam (ed.) *An Essay on the Origin of Human Knowledge*, New York: AMS Press, 1974.(A facsimile reproduction of the original 1756 translation by Thomas Nugent; introduction by J.H. Stam.)

(1749) *Traité des sistèmes, où l'on en démêle les inconvéniens et les avantages (Treatise on Systems)*, La Haye. (Regrettably, there is as yet no English translation of this work.)

(1754) *Traité des sensations*, trans.G. Carr, in H.W. Carr (ed.) *Treatise on the Sensations*, Los Angeles, CA: University of Southern California, 1930.(This is Condillac's best-known and most frequently cited book, containing his famous example of the statue which is endowed one by one with the five senses. The book is frequently acknowledged as a seminal text in the history of psychology.)

(1780) *La logique; ou, Les prémiers dévélopemens de l'art de penser*, Paris: Chez Petit Libraire; repr. in G. Le Roy (ed.) *Œuvres philosophiques*, vol. 2, Paris: Presses Universitaires de France, 1947-51; trans. J. Neef, in D.N. Robinson (ed.), *The Logic*, Washington, DC: University Publications of America, 1977.(The 1997 edition is a reprint of the original 1809 translation by J. Neef. This text offers the most succinct and accessible statement of Condillac's philosophy as a whole, including his conception of scientific method, his religious views and moral theory. Highly recommended.)

References and further reading

Cassirer, E. (1951) *The Philosophy of the Enlightenment*, Princeton, NJ: Princeton University Press.(A classic text in Enlightenment studies. Cassirer acknowledges Condillac's importance and gives him a prominent role as he traces the intellectual developments of the age.)

Frankel, C. (1969) *The Faith of Reason: The Idea of Progress in the French Enlightenment*, New York: Octagon Books.(Though more generally concerned with the French Enlightenment as a whole, Frankel also recognizes Condillac's central importance for the philosophy of the period and keeps him well within view throughout his study. An excellent general introduction to the Enlightenment.)

Hine, E.M. (1979) *A Critical Study of Condillac's Traité des Systèmes*, The Hague: Nijhoff.(An excellent in-depth study of one of Condillac's most important texts. Provides a thorough analysis of Condillac's thought within

the context of eighteenth-century intellectual controversies. Includes a comprehensive bibliography for further research.)

Knight, I.F. (1968) *The Geometric Spirit: The Abbé de Condillac and the French Enlightenment*, New Haven, CT: Yale University Press.(The only book in English devoted to a comprehensive survey of Condillac's life and work. Written in a readily accessible style, it provides an excellent introduction to the philosophy of the Enlightenment in general, and also offers researchers and specialists important insights into the philosophies which influenced Condillac, as well as his influence upon later developments in the history of thought.)

Schaupp, Z. (1925) *The Naturalism of Condillac*.(The first full-length treatment of Condillac in English, this obscure little book is still worth reading. Includes a careful discussion of Condillac's relation to Locke.)

Condorcet, Marie-Jean-Antoine-Nicolas Caritat de (1743-94)

The Marquis de Condorcet belongs to the second generation of eighteenth-century French philosophes. *He was by training and inclination a mathematician, and his work marks a major stage in the development of what is known today as the social sciences. He was held in high regard by contemporaries for his contributions to probability theory, and he published a number of seminal treatises on the theory and application of probabilism. He is best known today for the Esquisse d'un tableau historique des progrès de l'esprit humain (1795), his monumental, secularized historical analysis of the dynamics of man's progress from the primitive state of nature to modernity.*

Condorcet's principal aim was to establish a science of man that would be as concise and certain in its methods and results as the natural and physical sciences. For Condorcet there could be no true basis to science without the model of mathematics, and there was no branch of human knowledge to which the mathematical approach was not relevant. He called the application of mathematics to human behaviour and organization 'social arithmetic'.

The central epistemological assumption, upon which his philosophy was based, was that the truths of observation, whether in the context of the physical or the moral and social sciences, were nothing more than probabilities, but that their varying degrees of certainty could be measured by means of the calculus of probabilities. Condorcet was thus able, through mathematical logic, to counteract the negative implications of Pyrrhonic scepticism for the notions of truth and progress, the calculus providing not only the link between the different orders of knowledge but also the way out of the Pyrrhonic trap by demonstrating man's capacity and freedom to understand and direct the march of progress in a rationally-ordered way.

In his Esquisse Condorcet set out to record not only the history of man's progress through nine 'epochs', from the presocial state of nature to the societies of modern Europe, but in the tenth 'epoch' of this work he also held out the promise of continuing progress in the future. He saw the gradual emancipation of human society and the achievement of human happiness as the consequence of man having been endowed by nature with the capacity to learn from experience and of the cumulative, beneficial effects of the growth of knowledge and enlightenment. Condorcet's Esquisse laid the basis for the positivism of the nineteenth century, and had a particularly significant impact on the work of Saint-Simon and Auguste Comte.

1 Life

Condorcet was one of the outstanding French mathematicians of his time. He was the only eighteenth-century French *philosophe* of stature to have participated in the Revolution and, as a legislator, to have had an impact on events after 1789. Born in Ribemont, his early education took place at Reims, and by 1758 he had entered the University of Paris where he studied ethics, metaphysics, logic and mathematics at the prestigious Collège de Navarre. There he was taught by the Abbé Nollet, a proponent of Newtonian physics, and he worked closely with Georges Girault de Kéroudon on philosophical matters and on the crucial problems of the integral calculus. In later years he also came under the influence of Euler, Fontaine, the Bernouillis and, above all, of the distinguished mathematician and academician, Jean Le Rond D'Alembert, who became his patron. He was elected Perpetual Secretary of the Academy of Sciences in 1773, and in 1782 became a member of the French Academy. An enthusiastic supporter and theorist of the Revolution, he played an important role in the drafting of the *Déclaration des droits* in 1789. Suspected later of being a Girondin, he was denounced, and died, possibly a suicide, in Bourg-la-Reine while awaiting the guillotine.

His first major work, *Du calcul intégral*, was published in 1765 as part of the Academy of Science's proceedings, and was widely acclaimed. This was followed by a series of essays and mathematical papers, published between 1766 and 1769, including important work on the applications of the integral calculus to the still unresolved mathematical obscurities of Newton's *Principia*. The extensions of the methodology of differential calculus, probability (the 'mathematics of hope') and their application to nonscientific areas, particularly the moral, political and social sciences, were to remain at the core of his thinking, especially during and after Turgot's ministry (1774-6). His exploration of the potential of the calculus of probabilities was developed further in the *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* in 1785. The *Essai* is complemented by the *Eléments du calcul des probabilités et son application aux jeux de hasard*, not published in its own right until 1805.

Condorcet was part of that new-wave reformist movement in late eighteenth-century France that included Turgot, the *idéologues*, and the physiocrats, all united in their understanding of how the world of ideas could and must interact with the world of political and social reality. Other major publications include the *Essai sur la constitution et les fonctions des assemblées provinciales* (1788), *Sur l'Instruction publique* (1791-2), *Réflexions sur la jurisprudence criminelle* (1775), *De l'Influence de la Révolution de l'Amérique sur l'Europe* (1786), *Quatres lettres d'un bourgeois de Newh(e)aven á un citoyen de Virginie* (1788), *Lettres sur le commerce des grains* (1775), and the *Réflexions sur l'esclavage des nègres* (1781). In addition, he wrote innumerable pamphlets, drafts of bills and other legislative material for the National Convention. He was also interested in the development of a symbolic logic to give precise expression to intellectual operations and which would be appropriate to the formulation of a universal language of the sciences, although his treatise on this subject, the *Essai d'une langue universelle*, was to remain unfinished. In the non-mathematical area his greatest and most influential work is the *Esquisse d'un tableau historique des progrès de l'esprit humain*, published posthumously in 1795.

2 The science of the probable

Condorcet used mathematics as a model upon which to build a philosophy of social science, and to establish a methodology as applicable to the science of man as it was to the physical sciences. In Condorcet's hands mathematics became an instrument of social and philosophical analysis and, following the lead given by D'Alembert, he set out to integrate the Newtonian view of a rationally determined order of nature into an analagous framework of moral, social and political order. He postulated the view that all human sciences were underpinned by positive fact in the same way as the physical sciences, and open to a rigorous system of analysis made meaningful through the use of a precise, well-determined 'universal' language, capable of unambiguous use across the whole spectrum of scientific enquiry.

Greatly influenced by Locke and Hume, as well as by the French sensationalist philosopher Condillac, Condorcet devoted much of his intellectual life to the development of a concept of 'social arithmetic' based on the calculus of probabilities. He saw probabilism as constituting the essential epistemological link between the social and the physcial sciences. By utilising the calculus of probabilities, the uncertainties and ambivalences inherent in previous attempts to study and evaluate man's behaviour, which had resulted in the case of many philosophers in a profound scepticism, could be dissipated. He was convinced that this 'true philosophy' would provide the foundation for a systematic 'science of man'. The clearest elaboration of this philosophy of probable belief and the methodological principles for its application are to be found in general, tentative outline in the notes to Condorcet's reception speech to the French Academy in 1782, and in more sophisticated mathematical detail in the *Mémoire sur le calcul des probabilités* (1784), in the *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* and in the *Eléments du calcul des probabilités et son application aux jeux de hasard*.

In the preliminary discourse to the *Essai sur l'application de l'analyse*, he postulated two key principles governing the processes of human reasoning: (1) that 'nature follows invariable laws' and (2) that these laws 'are made known to us by observable phenomena'. What leads us to believe in the truth of such a postulation is our phenomenological experience of the facts and of the ways in which that experience accords with these two principles. A perfect and definitive calculation of the probability of their truth can never be fully realized, as it is impossible to take cognizance of the totality of the factors that shape our experience. Condorcet insisted, on the other hand, that such a calculation, were it possible, would indicate a very high degree of probability of the truth of these principles. In the light of this probable truth, Condorcet then added a third working proposition, namely that all human reasoning that informs judgment, decision making, choice and conduct is based ultimately on probability.

'The truths proved by experience are simply probabilities.' For Condorcet this insistence on uncertainty did not, however, lead to the *impasse* of Pyrrhonism. On the contrary, although all knowledge was founded only on probabilities, the value, or degree of probability could be determined with relative precision.

Condorcet fully accepted the Lockean view on epistemological modesty. Uncertainty characterized all human understanding (exception being made for the mathematical model itself), but for Condorcet, as for Locke, uncertainty was not an invincible, action-denying absolute. In the *Essai sur l'application de l'analyse* he sought to

demonstrate, by means of the calculus of probabilities, how the defeatist scepticism of the past could be made to give way before the new positivism.

The calculus of probabilities was applicable in theory to all aspects of human life and behaviour, and in demonstrating the logical foundation for this principle Condorcet developed a view of rational belief that owed as much to Hume as to Locke. Belief in both the moral and physical sciences was in his system simply the representation of things as having to exist in a certain way, based on our experience that what has occurred will tend to recur within a frame of constant laws. Belief was not, however, the result of a raw process of reaction to sense impressions. Man obeys an automatic sentiment that leads him to belief, but in order to avoid

judgment and opinion degenerating into prejudice and irrationality, Condorcet took care to distinguish between the *sentiment* of belief and the actual *grounds* for belief. Reason and experience must play their part if man was to be rescued from the illusions of the senses and the fleeting impressions made upon the senses. To this end, he advanced the view that reason had found a powerful weapon in the form of the calculus of probabilities, which offered a dependable methodology for the estimation of the *grounds* for belief. The calculus would provide the necessary mechanism for the correction of any error arising from the passive, automatic and uncritical *sentiment* of belief, particularly important in the case of the moral and social sciences.

The principles of probabilistic philosophy enabled Condorcet to elaborate a model of calculation that permitted the objective evaluation of man in society, and with it he sought to transform the calculus of probabilities into a mathematically-based language of rational decision-making and action. The *Essai sur l'application de l'analyse* was an attempt to illuminate the ways in which the calculus could work in a practical context, in this case the constitutional process itself, so that the unpredictable and the contingent could be measured and minimized. This particular treatise represents Condorcet's most detailed and sustained attempt to 'discover the probability that assures the validity of a law passed by the smallest possible majority, such that one can believe that it is not unjust to subject others to this law and that it is useful for oneself to submit to it'. The mathematics that he then deployed exemplify the pioneering methodology that he would adopt in other contexts, such as crime, jurisprudence and taxation theory, to locate the human sciences within the realm of the probable, and to attempt to address the otherwise intractable problem of accounting for chance in human behaviour.

3 Progress and the science of man

Condorcet's name has been associated most commonly with the 'idea of progress', and the work in which he developed this idea in depth is the *Esquisse d'un tableau historique des progrès de l'esprit humain*. Based on the empirical observation of data and the statistical analysis of that data, the *Esquisse* traces the trajectory of human achievement using a de-christianized chronology of historical periods or 'epochs'. The tableau starts with primitive man in the state of presocial nature and culminates in the ninth 'epoch', covering the years from Descartes and the late seventeenth century to the birth of the first French republic. A tenth 'epoch' offers a vision of the postmillenium future and holds out the promise of unlimited human perfectibility. Condorcet paid particular attention to two factors in man's advancement: (1) the growth of language as the principal vehicle of social progress and intellectual advancement, and (2) the development of technology and the physical sciences as instruments facilitating the progressive liberation of man from the darkness of past error and servitude.

Lockean sensationalist psychology deeply influenced Condorcet, particularly with regard to his doctrine of moral sentiment. At the start of the *Esquisse* primitive man emerges as the one creature with the faculty of receiving sensations, of reflecting upon them, of analysing them and recombining them. In Condorcet's view, the pleasure-pain principle engendered in early man moral feelings, and eventually relationships, based on controlled self-interest. The sensations facilitated man's difficult, but irreversible, climb out the of the darkness of primitive presocial life into the light of civilisation. Condorcet understood the implications for the moral sciences of Lockean reversion to the origins of knowledge in sense experience, together with its consequential destruction of the myth of innate ideas, and he saw Lockean sensationalism as an intellectual event whose importance was matched only by that of the Newtonian revolution in physics.

Condorcet wanted to show in the *Esquisse* that history was not the creation of random forces, with man cast in the role of passive spectator/victim. The gradual emancipation of man from the limitations imposed upon him by nature, and the consequential liberation of the individual, was itself a natural process, and the reflection of an order

inherent in man's condition that could be made intelligible. Man's progress was enacted within the framework of an exclusively human condition, free from the intervention of transcendental forces. Progress was for Condorcet an entirely secular concept, the fruit of human dynamics interacting with the natural currents of history alone. Evil was not a consequence of man's nature but of the absence of enlightenment, and would recede inevitably as knowledge in the moral sciences caught up with the advances being made in the physical sciences, and extended its beneficial effects.

In linking the pursuit of knowledge, and the inexorable logic of scientific advances, to the mission of progress, Condorcet had to demonstrate necessarily that there was a relationship between advances made in the physical and natural sciences and those made in the moral and social sciences, and that as man learned to order his natural environment by means of the physical sciences he would also learn to order his social environment through the advancement of the moral sciences and their political and sociological extensions. The historical portrait of man in the *Esquisse* is drawn with that demonstration in mind in the context of each successive 'epoch'.

Progress for Condorcet was always a cumulative, collective phenomenon, dependent upon the free pursuit of knowledge and upon the rational application of that knowledge. His view of progress assumed that the laws of nature were constant, and that there was an analogous constancy at work in historical processes to which the calculus of probabilities in relation to the future was relevant. A scientific, mathematically-informed study of history would reveal constant principles, many of which would confirm the truth of human progress, as far as this truth could be defined in probabilistic terms.

The power of mathematics allowed man to rise above the facts of random phenomena and to take advantage of the 'law of calculated observations'. This was the law that permitted a scientific understanding of causes, effects and relationships, that allowed for the determination of those recurring patterns of phenomena in human history that made a given truth probable, and that facilitated the measurement of degrees of certainty, and therefore control, in human affairs. It was the key that would open the way to a rationally-planned application of the 'science of man'. The 'science of man', anchored firmly to what were essentially Baconian traditions of thought - observation, experiment, calculation - and Lockean-Humean epistemology, would establish the basis for a radical reordering of the processes of human understanding to create 'a new understanding admitting only precise ideas, exact notions and truths whose degree of certainty or probability has been rigorously weighed'.

Condorcet argued throughout the *Esquisse* the case for the indefinite perfectibility of human society. His vision entailed the construction of a future in which man's potential for social and political choice of action was theoretically infinite. Condorcet's positivism was not facile, however, nor was his optimism Panglossian. The tenth 'epoch' of the *Esquisse*, in some ways naïvely utopian, is a projection of probabilities set out within a cautiously defined context of preconditions, reservations and contingencies. Condorcet never lost sight of the essential fragility of human civilization; progress remained dependent ultimately on the rational exercise of the human will alone, and without that vital driving-force progress would not take place.

See also: Comte, A.; Human nature, science of, in the 18th century; Positivism in the social sciences; Pyrrhonism; Saint-Simon, C.-H. de R.

DAVID WILLIAMS

List of works

Condorcet, M.J.A.N. (1968) *Œuvres de Condorcet*, ed. A. Condorcet-O'Connor and F. Arago, Stuttgart-Bad Canstatt: Friederich Fromann Verlag, 12 vols.(This is a reprint of the 1847-9 edition of Condorcet's collective works; still the standard edition.)

Condorcet, M.J.A.N. (1785) *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix (Essay on the Application of Analysis to the Probability of Majority Decisions)*, Paris: Imprimerie royale; repr. New York: Chelsea, 1972.

Condorcet, M.J.A.N. (1795) *Esquisse d'un tableau historique des progrès de l'esprit humain*, Paris: Agasse; repr. A. Pons (ed.), Paris: Flammarion, 1988; trans. J. Barraclough, in S. Hampshire (ed.), *Sketch for a Historical Tableau of the Progress of the Human Mind*, London: Weidenfeld and Nicolson.

Condorcet, M.J.A.N. (1805) *Éléments du calcul des probabilités et son application aux jeux de hasard, à la loterie et aux jugements des hommes (Elements of the Calculus of Probabilities and its Application to Games of Chance, to the Lottery and to Men's Judgements)*, Paris: Fayolle.(This treatise first appeared as an appendix to

Euler's correspondence with a German princess (1787-9). There is no modern edition, and it was not included in the O'Connor-Arago edition of Condorcet's works.)

References and further reading

Badinter, E. and R. (1968) *Condorcet 1743-94. Un Intellectuel en politique (Condorcet 1743-94. An Intellectual in Politics)*, Paris: Fayard.(A biographical study with some account of intellectual development. Aimed at the general reader.)

Baker, K. (1975) *Condorcet. From Natural Philosophy to Social Mathematics*, Chicago, IL and London: University of Chicago Press.(The most comprehensive and authoritative analysis of Condorcet's ideas so far published. Particularly useful for the *Essai sur l'application de l'analyse* and the *Éléments du calcul*, as well as for an understanding of the philosophical and historical implications of the *Esquisse*. Excellent bibliography. Suitable for specialists and general readers.)

Granger, G.-G. (1956) *La Mathématique sociale du marquis de Condorcet (The Social Mathematics of the Marquis de Condorcet)*, Paris: Presses Universitaires de France.(Still a very useful study of the originality of Condorcet's contribution to the concept of social science for the general reader.)

Popkin, R. (1987) 'Condorcet and Hume and Turgot', *Condorcet Studies* 2: 47-62.(A closely argued account of the philosophical divisions between Condorcet and Hume, with particular reference to the impact of Hume's break with Turgot. A clear account of the ideas of two of the greatest social scientists of the Enlightenment. Aimed at the specialist.)

Shapiro, J. (1978) *Condorcet and the Rise of Liberalism*, New York: Octagon Books.(A general account of Condorcet's philosophy of progress, with particular reference to the *Esquisse*. A good and lively introduction to Condorcet for the general reader.)

Confirmation theory

The result of a test of a general hypothesis can be positive, negative or neutral. The first, qualitative, task of confirmation theory is to explicate these types of test result. However, as soon as one also takes individual hypotheses into consideration, the interest shifts to the second, quantitative, task of confirmation theory: probabilistically evaluating individual and general hypotheses in the light of an increasing number of test results. This immediately suggests conceiving of the confirmation of an hypothesis as increasing its probability due to new evidence.

Rudolf Carnap initiated a research programme in quantitative confirmation theory by designing a continuum of probability systems with plausible probabilistic properties for the hypothesis that the next test result will be of a certain kind. This continuum of inductive systems has guided the search for optimum systems and for systems that take analogy into account.

Carnapian systems, however, assign zero probability to universal hypotheses. Jaakko Hintikka was the first to reconsider the confirmation of such hypotheses and using Carnap's continuum for this purpose has set the stage for a whole spectrum of inductive systems of this type.

1 Qualitative and quantitative confirmation theory

According to the hypothetico-deductive method a theory is tested by examining its implications. The result of an individual test of a general hypothesis stated in observation terms can be positive, negative or neutral. If it is neutral the test was not well devised; if it is negative, the hypothesis, and hence the theory, has been falsified. *Qualitative* confirmation theory primarily aims at further explicating the intuitive notions of neutral and positive test results. Some paradoxical features discovered by Hempel and some queer predicates defined by Goodman show that this is not an easy task.

Assuming that a black raven confirms the hypothesis 'All ravens are black' and that confirmation is not affected by logically equivalent reformulations, Hempel argued (1965) that not only a non-black non-raven but, even more counterintuitively, also a black non-raven confirms it (see Hempel, C.G. §2). Goodman (1955) argued that not all predicates guarantee the 'projectibility' of a universal hypothesis from observed to non-observed cases. For example, if 'grue' means 'green, if examined before *t*' and 'blue, if not examined before *t*', a green emerald discovered before *t* would not only confirm 'All emeralds are green' but also 'All emeralds are grue', and hence, assuming that consequences are also confirmed, even 'All emeralds not examined before *t* are blue' (see Goodman, N. §3). Whereas Goodman succeeded in formulating criteria for acceptable predicates, in terms of their relative 'entrenchment' in previously successfully projected generalizations, up to now nobody has given a generally accepted qualitative solution to Hempel's riddles (see Underdetermination §2).

Below we treat *quantitative*, more specifically, probabilistic confirmation theory, which aims at explicating the idea of confirmation as increasing probability due to new evidence. Carnap introduced this perspective and pointed confirmation theory towards the search for a suitable notion of logical or inductive probability (see Probability, interpretations of §5). Generally speaking, such probabilities combine indifference properties with inductive properties.

2 The continuum of inductive systems

Mainly in his *The Continuum of Inductive Methods* (1952), Rudolf Carnap started a fruitful research programme centering around the famous λ-continuum. The probability systems in this programme can be described in terms of individuals and observation predicates or in terms of trials and observable outcomes. The latter way of presentation will be used here. Moreover, we will presuppose an objective probability process, although the systems to be presented can be applied in other situations as well.

Consider a hidden wheel of fortune. You are told, truthfully, only that it has precisely four coloured segments, BLUE, GREEN, RED, and YELLOW, without further information about the relative size of the segments. So you do not know the objective probabilities. What you subsequently learn are only the outcomes of successive trials. Given the sequence of outcomes e_n of the first *n* trials, your task is to assign reasonable probabilities, $\mathrm{p}(\mathrm{R}/e_n)$,

to the hypothesis that the next trial will result in, for example, RED.

There are several ways of introducing the λ-continuum, but the basic idea behind it is that it reflects gradually learning from experience. In fact, as Zabell (1982) rediscovered, C-systems were anticipated by Johnson. According to Carnap's favourite approach $\mathrm{p}(\mathrm{R}/e_n)$ should depend only on *n* and the number of occurrences of RED thus far, n_{R}. More specifically, it should be a special weighted mean of the observed relative frequency n_{R}/n and the (reasonable) initial probability 1/4. This turns out to leave room for a continuum of (C-)systems, the λ-continuum, $0 < \lambda < \infty$:

$$\mathrm{p}(\mathrm{R}/e) = (n_{\mathrm{R}} + \lambda/4)/(n + \lambda) =$$
$$n/(n + \lambda) \cdot (n_{\mathrm{R}}/n) + \lambda/(n + \lambda) \cdot (1/4)$$

Note that the weights $n/(n + \lambda)$ and $\lambda/(n + \lambda)$ add up to 1 and that the larger the value of λ the slower the first increases at the expense of the second; that is, the slower one is willing to learn from experience.

C-systems have several attractive indifference, confirmation and convergence properties. The most important are these:

(a) order indifference or exchangeability: the resulting prior probability, $\mathrm{p}(e_n)$, for e_n, does not depend on the order of the results of the trials, that is, $\mathrm{p}(\mathrm{R}/e_n) = \mathrm{p}(e_n^*)$ for any permutation e_n^* of e_n;
(b) instantial confirmation: if e_n is followed by RED this is favourable for RED, that is, $\mathrm{p}(\mathrm{R}/e_n\mathrm{R}) > \mathrm{p}(\mathrm{R}/e_n)$;
(c) instantial convergence: $\mathrm{p}(\mathrm{R}/e_n)$ approaches n_{R}/n for increasing *n*.

However, confirmation and convergence of universal hypotheses are excluded in C-systems. The reason is that C-systems in fact assign zero prior probability to all universal generalizations, for instance to the hypothesis that all results will be RED. Of course, this is desirable in the described situation, but if you were told only that there are *at most* four coloured segments, you would like to leave room for this possibility.

3 Optimum inductive systems

Carnap proved that for certain kinds of objective probability process, such as a wheel of fortune, there is an optimal value of λ, depending on the objective probabilities, in the sense that the average mistake may be expected to be lowest for this value. Surprisingly, this optimal value is independent of *n*. Of course, in actual research, where we do not know the objective probabilities, this optimal value cannot be calculated. Carnap did not raise the question of a reasonable estimate of the optimal value for a specific objective process, for he saw the problem of selecting a value of λ primarily as a choice that each scientist had to make in general. However, the question of a reasonable estimate has attracted the attention of other researchers.

Festa (1993) proposed basing the estimate on 'contextual' knowledge of similar processes in nature. For example, in ecology one may know the relative frequencies of certain species in different habitats before one starts the investigation of a new habitat. For a quite general class of systems, Festa formulates a solution to the estimation problem that relates the research area of confirmation theory to that of truth approximation: the optimum solution may be expected to be the most efficient way of approaching the objective or true probabilities.

Unfortunately, wheels of fortune do not constitute a technological (let alone a biological) kind for which you can use information about previously investigated instances. But if you had knowledge of a random sample of all existing wheels of fortune, Festa's approach would work on the average for a new, randomly drawn, one.

4 Inductive analogy by similarity

Carnap struggled with the question of how to include analogy considerations into inductive probabilities. His important distinction between two kinds of analogy - by similarity and by proximity - was posthumously published (Carnap 1980). Here we will consider only the first kind.

Suppose you find GREEN more similar to BLUE than to RED. Carnap's intuition of analogy by similarity is that, for instance, the posterior probability of GREEN should be higher if, other things being equal, some occurrences of RED are replaced by occurrences of BLUE. The background intuition, not valid for artificial wheels of fortune, is that similarity in one respect (for example, colour) will go together with similarities in other respects (for

example, objective probabilities). However, explicating which precise principle of analogy one wants to fulfil has turned out to be very difficult.

One interesting approach is in terms of *virtual trials*, which is a way of determining how analogy influence is to be distributed after n trials. In the case above, BLUE would get more analogy credits than RED from an occurrence of GREEN. The resulting systems of virtual analogy have the same confirmation and convergence properties as C-systems. Moreover, they satisfy a principle of analogy that is in general more plausible than Carnap's. Roughly, the new principle says that replacement of a previous occurrence of BLUE by GREEN makes less difference for the probability that the next trial will be BLUE than does replacement of a previous occurrence of RED by GREEN for the probability that the next trial will be RED.

Unfortunately, systems of virtual analogy are not exchangeable: the order of results makes a difference. There are, however, exchangeable systems with some kind of analogy by similarity. For example, Skyrms (1993) has sketched a Bayesian approach that uses information about possible gravitational bias of a symmetric wheel of fortune to determine the influence of analogy, and di Maio (1995) follows a suggestion of Carnap.

5 Universal generalizations

Carnap was well aware that C-systems do not leave room for non-analytic universal generalizations. In fact, all systems presented thus far have this problem. Although tolerant of other views, Carnap himself was inclined to downplay the theoretical importance of universal statements.

Hintikka (1966) took up the problem of universal statements. His basic idea was to apply Bayesian conditionalization to Carnap's C-systems with uniform λ, using a specific prior distribution on universal statements containing a parameter α. The systems belonging to the resulting α-λ-continuum have the instantial confirmation and convergence properties and also the desired property of universal confirmation; that is, that the probability of a not yet falsified universal statement increases with another instance of it. Moreover, they satisfy universal convergence: $\mathrm{p}(\mathrm{H_R}/e_n)$ approaches 1 for increasing *n* as long as only RED continues to occur (where $\mathrm{H_R}$ indicates the universal statement that only RED will occur in the long run). For increasing parameter α, universal confirmation is smaller and universal convergence slower.

In fact, systems based on an arbitrary prior distribution and C-systems with arbitrary λ, here called H-systems, already have the general properties of instantial and universal confirmation and convergence. The subclass of H-systems based on an arbitrary prior distribution and C-systems with λ proportional to the corresponding number of possible outcomes is particularly interesting. For this subclass appears to be co-extensive with a class of systems introduced by Hintikka and by Niiniluoto (1988), very differently from H-systems, using principles and parameters related only to finite sequences of outcomes.

There is also a plausible 'delabelled' reformulation of H-systems that can be extended to the very interesting case of an unknown denumerable number of possible outcomes. Presenting the delabelling in terms of exchangeable partitions, Zabell (1996) has studied this case by principles leading to a class of systems with three parameters.

6 Applications

The Carnap-Hintikka programme in confirmation theory, also called inductive probability theory, has applications in several directions. Systems of inductive probability were intended primarily for explicating confirmation as increasing probability. One may even define a quantitative degree of confirmation: namely, as the difference or ratio of the posterior and the initial or prior probability. There are other interesting types of application as well. Carnap (1971) and Stegmüller (1973) stressed that they can be used in decision making, and Skyrms (1991) applies them in game theory (see Decision and game theory). Costantini *et al*. (1982) use them in a rational reconstruction of elementary particle statistics. Festa suggests several areas of empirical science where optimum inductive systems can be used. Finally, for universal hypotheses, Hintikka, Hilpinen (1968) and Pietarinen (1972) use systems of inductive probability to formulate rules of acceptance and Niiniluoto uses these systems to estimate degrees of verisimilitude.

See also: Induction, epistemic issues in; Inductive inference; Learning; Probability, interpretations of; Probability theory and epistemology; Rationality of belief; Statistics

THEO A.F. KUIPERS

References and further reading

Almost all items presuppose some elementary mathematical training.

Carnap, R. (1952) *The Continuum of Inductive Methods*, Chicago, IL: University of Chicago Press.(Introduces the λ-continuum as a way of gradually learning from experience and proves that there is an optimal value of λ, depending on the underlying objective probabilities. See §§2, 3.)

Carnap, R. (1971) 'Inductive Logic and Rational Decisions', in R. Carnap and R. Jeffrey (eds) *Studies in Inductive Logic and Probability*, Berkeley, CA: University of California Press, vol. I, 5-31.(Presents the general aims of inductive logic and in particular its role in rational decision making. §6.)

Carnap, R. (1980) 'A Basic System of Inductive Logic, Part 2', in R. Jeffrey (ed.) *Studies in Inductive Logic and Probability*, Berkeley, CA: University of California Press, vol. II, 7-155.(Continues the foundational analysis of systems of inductive probability, started in Part 1 of Volume I of 1971, with special emphasis on analogy, introducing the important distinction between analogy by similarity and analogy by proximity. §4.)

Costantini, D., Galavotti, M.C. and Rosa, R. (1982) 'A Rational Reconstruction of Elementary Particle Statistics', *Scientia* 76 (117): 151-9.(Presents a rational reconstruction of elementary particle statistics in terms of principles underlying Carnap's λ-continuum. §6.)

Festa, R. (1993) *Optimum Inductive Methods*, Dordrecht: Kluwer.(Gives a detailed analysis of the relation between the generalization of Carnap's λ-continuum, Bayesian statistics, notably Dirichlet distributions, and verisimilitude. Proposes an estimate of the optimal value of λ on the basis of 'contextual' knowledge of similar processes and suggests several areas of empirical science where optimum inductive systems can be used. §§3, 6.)

Goodman, N. (1955) *Fact, Fiction, and Forecast*, Cambridge, MA: Harvard University Press.(Contains discussion of intuitive ideas about qualitative confirmation and argues that they should be restricted to well-entrenched predicates in order to exclude the confirmation of generalizations in terms of queer predicates, like 'grue'. §1.)

Hempel, C.G. (1965) *Aspects of Scientific Explanation*, New York: Free Press. (Contains, among other things, Hempel's most important writings on qualitative confirmation, including the raven paradoxes, induction and rational belief. §1.)

Hilpinen, R. (1968) *Rules of Acceptance and Inductive Logic*, Amsterdam: North Holland.(Gives a detailed analysis of Hintikka's two-dimensional continuum and corresponding rules of acceptance. §6.)

Hintikka, J. (1966) 'A Two-dimensional Continuum of Inductive Methods', in J. Hintikka and P. Suppes (eds) *Aspects of Inductive Logic*, Amsterdam: North Holland, 1966, 113-32.(Presents the two-dimensional α-λ-continuum, leaving room for universal hypotheses, based on Carnap's λ-continuum and a specific prior distribution related to a virtual universe of α individuals. §5.)

Kuipers, T. (1978) *Studies in Inductive Probability and Rational Expectation*, Dordrecht: Reidel. (Studies the systems of Carnap, Hintikka and Niiniluoto, and their generalization, in great detail. Refers also to the related work of Carnap and Hintikka, and to the work of Hilpinen, Pietarinen and Stegmüller. §§2, 5, 6.)

Kuipers, T. (1988) 'Inductive Analogy by Similarity and Proximity', in D.H. Hellman (ed.) *Analogical Reasoning*, Dordrecht: Kluwer, 299-313.(Presents systems with virtual inductive analogy by similarity, and also studies analogy by proximity. §4.)

Maio, M.C. di (1995) 'Predictive Probability and Analogy by Similarity in Inductive Logic', *Erkenntnis* 43: 369-94.(Weakens restricted relevance such that it leads to exchangeable systems which are sensitive to analogy by similarity. §4.)

Niiniluoto, I. (1988) 'Analogy and Similarity in Scientific Reasoning', in D.H. Hellman (ed.) *Analogical Reasoning*, Dordrecht: Kluwer.(Discusses, among others, proposals for inductive analogy by similarity by Costantini, Kuipers, Niiniluoto and Spohn. §4.)

Pietarinen, J. (1972) *Lawlikeness, Analogy and Inductive Logic*, Amsterdam: North Holland.(Gives an overview of the main ideas and problems in the Carnap-Hintikka programme around 1970. §6.)

Skyrms, B. (1991) 'Inductive Deliberation, Admissable Acts and Perfect Equilibrium', in M. Bacharach and S. Hurley (eds) *Foundations of Decision Theory*, Oxford: Blackwell, 220-41.(Applies Carnap's λ-continuum in game theory by letting players update their beliefs according to it. §6.)

Skyrms, B. (1993) 'Analogy by Similarity in HyperCarnapian Inductive Logic', in J. Earman (ed.) *Philosophical Problems of the Internal and External Worlds. Essays concerning the Philosophy of Adolf Grünbaum*, Pittsburgh, PA: University of Pittsburgh Press, 973-82.(Presents a Bayesian approach to analogy that uses information about possible gravitational bias of a symmetric wheel of fortune to determine the influence of

analogy. §4.)

Stegmüller, W. (1973) *Carnap II: Normative Theorie des Induktiven Räsonierens*, Berlin: Springer.(Emphasizing the role in rational decision it presents Carnap's λ-continuum and generalizes it to arbitrary initial probabilities. §§2, 6.)

Zabell, S.L. (1982) 'W.E. Johnson's "Sufficientness" Postulate', *Annals of Statistics* 10 (4): 1,091-9.(Points out that Johnson anticipated C-systems in 1932, including its axiomatic foundation. §2.)

Zabell, S.L. (1996) 'The Continuum of Inductive Methods Revisited', in J. Earman and J. Norton (eds) *The Cosmos of Science*, Pittsburgh, PA: University of Pittsburgh Press, and Konstanz: Universitätsverlag Konstanz, 349-83.(Studies 'delabelled' probability systems in terms of exchangeable partitions, by imposing additional principles, leading to a class of systems with three parameters. §5.)

Confucian philosophy, Chinese

Chinese Confucian philosophy is primarily a set of ethical ideas oriented toward practice. Characteristically, it stresses the traditional boundaries of ethical responsibility and dao*, or the ideal of the good human life as a whole. It may be characterized as an ethics of virtue in the light of its conception of* dao *and* de *(virtue). Comprising the conceptial framework of Confucian ethics are notions of basic virtues such as* ren *(benevolence),* yi *(rightness, righteousness), and* li *(rites, propriety). There are also notions of dependent virtues such as filiality, loyalty, respectfulness and integrity. Basic virtues are considered fundamental, leading or action-guiding, cardinal and the most comprehensive. In the classic Confucian sense,* ren *pertains to affectionate concern for the well-being of fellows in one's community. Notably,* ren *is often used in an extended sense by major Song and Ming Confucians as interchangeable with dao for the ideal of the universe as a moral community.* Yi *pertains to the sense of rightness, especially exercised in coping with changing circumstances of human life, those situations that fall outside the scope of* li. Li *focuses on rules of proper conduct, which have three functions: delimiting, supportive and ennobling. That is, the* li *define the boundaries of proper behaviour, provide opportunities for satisfying desires of moral agents within these boundaries, and encourage the development of noble characters which markedly embody cultural refinement and communal concerns. The* li *are the depository of insights of the Confucian tradition as a living ethical tradition. This tradition is subject to changing interpretation governed by the exercise of* quan *or the weighing of circumstances informed by the sense of rightness (*yi*).*

However, the common Confucian appeal to historical events and paradigmatic individuals is criticized because of lack of understanding of the ethical uses of such a historical appeal. The pedagogical use stresses the study of the classics in terms of the standards of ren, yi *and* li. *Learning, however, is not a mere acquisition of knowledge, but requires understanding and insight. Also, the companion study of paradigmatic individuals is important, not only because they point to models of emulation but also because they are, so to speak, exemplary personifications of the spirits of* ren, yi *and* li. *Moreover, they also function as reminders of moral learning and conduct that appeal especially to what is deemed in the real interest of the learner. The rhetorical use of the historical appeal is basically an appeal to plausible presumptions, or shared beliefs and trustworthiness. These presumptions are subject to further challenge, but they can be accepted as starting points in discourse. The elucidative use of historical appeal purports to clarify the relevance of the past for the present. Perhaps most important for argumentative discourse is the evaluative function of historical appeal. It focuses our knowledge and understanding of our present problematic situations as a basis for exerting the unexamined claims based on the past as a guidance for the present. Thus, both the elucidative and evaluative uses of historical appeal are critical and attentive to evidential grounding of ethical claims.*

*Because of its primary ethical orientation and its influence on traditional Chinese life and thought, Confucianism occupies a pre-eminent place in the history of Chinese philosophy. The core of Confucian thought lies in the teachings of Confucius (551-479 BC) contained in the Analects (*Lunyu*), along with the brilliant and divergent contributions of Mencius (372?-289 BC) and Xunzi (*fl. *298-238 BC), as well as the Daxue (Great Learning) and the Zhongyong (Doctrine of the Mean), originally chapters in the Liji (Book of Rites). Significant and original developments, particularly along a quasi-metaphysical route, are to be found in the works of Zhou Dunyi (1017-73), Zhang Zai (1020-77), Cheng Hao (1032-85), Cheng Yi (1033-1107), Zhu Xi (1130-1200), Lu Xiangshan (1139-93), and Wang Yangming (1472-1529). Li Gou (1009-59), Wang Fuzhi (1619-92), and Dai Zhen (1723-77) have also made noteworthy contributions to the critical development of Confucian philosophy. In the twentieth century, the revitalization and transformation of Confucian philosophy has taken a new turn in response to Western philosophical traditions. Important advances have been made by Feng Youlan, Tang Junyi, Thomé H. Fang, and Mou Zongsan. Most of the recent works in critical reconstruction are marked by a self-conscious concern with analytic methodology and the relevance of existentialism, phenomenology, and hermeneutics. Still lacking is a comprehensive and systematic Confucian theory informed by both the history and the problems of Western philosophy.*

1 The conceptual framework

A student or scholar of moral philosophy, approaching the classical core of Confucian ethics for the first time, is likely to be faced with a major difficulty in the lack of systematic exposition of basic Confucian notions such as

ren (benevolence), *li* (rites), and *yi* (rightness), and their interconnection. Unlike major Western philosophers, Confucian thinkers are not concerned with definitions. As William Theodore de Bary (1970: v) has pointed out, most Chinese thinkers are not interested in the definition and analysis of concepts: 'generally, the more crucial or central the idea, the greater the ambiguity'. Instead, they are concerned with the expansive uses of ideas that suggest the widest possible range of meaning.

From the point of view of contemporary moral philosophy, this pervasive feature of Confucian discourses may appear to be an anomaly, given the classical emphasis on the right use of terms (*zhengming*). A serious student of the works of Xunzi, the one classical Confucian generally considered to be the most rationalistic and systematic philosopher, will be frustrated in the attempt to find definitions, in the sense of necessary and sufficient conditions, for the application of basic Confucian terms. This fact is all the more surprising in view of Xunzi's recurrent employment of certain definitional locutions or quasi-definitional formulas for explaining his theses on human nature and the mind. Like most major Confucians, Xunzi has a pragmatic attitude toward the use of language, that is, the uses of terms that require explanation are those that are liable to misunderstanding in the context of particular discourse. The Confucian explanations of the use of ethical terms are context-dependent and are addressed to a particular rather than universal audience.

Two different assumptions underlying this attitude toward language may account for the absence of Chinese interest in context-independent explanation of the use of ethical terms. First, there is an assumption of the primacy of practice implicit in the Confucian doctrine of the unity of knowledge and action (*zhixing heyi*). Definition, in the sense of meaning explanation, is a matter of practical rather than theoretical necessity. This assumption does not depreciate the importance of theoretical inquiry, but rather focuses on its relevance to the requirements of practice, particularly those that promote unity and harmony among people in the community. Such requirements vary in time and place. A viable ethical theory is thus subject to pragmatic assessment in the light of changing circumstances. In general, ethical requirements cannot be stated in terms of absolute or fixed principles or rules. It is this assumption of the primacy of practice that renders plausible Donald Munro's claim that the consideration important to the Chinese is the behavioural implications of the belief or proposition in question (Munro 1969). What effect does adherence to the belief have on people? What implications for social action can be drawn from the statement? In Confucianism, there was no thought of 'knowing' that did not entail some consequence for action.

Related to the primacy of practice is the assumption that reasoned discourse may legitimately appeal to what Nicholas Rescher calls 'plausible presumptions,' that is, an appeal to shared knowledge, belief or experience, as well as to established or operative standards of discourse (Rescher 1977: 38). For Confucian thinkers, most of these presumptions, at least among the well-educated, represent the shared understanding of a common tradition and a living cultural heritage. These presumptions are often suppressed and mainly form the background of discourse. Thus Confucian reasoning and argumentation appears to be highly vague and inexplicit. From the Aristotelian point of view, Confucian reasoning is 'rhetorical', as it frequently involves enthymemes and arguments from examples.

Given the assumptions of the primacy of practice and appeal to plausible presumptions, any attempt at introducing Confucian ethics to students of Western philosophy involves a basic task in philosophical reconstruction. This is a task in constructive interpretation, not only in providing a general characterization but also in reshaping the basic notions and concerns of Confucian ethics in response to some problems of moral philosophy. Such an exploration serves two important and connected aims: first, a plausible explication of the Confucian outlook in philosophically relevant idioms, and second, a critical development of Confucian ethical thought.

2 An ethics of virtue: *dao* and *de*

While major Confucianists (for example, Mencius and Xunzi) differ in their conceptions of human nature in relation to conduct, most of them accept Confucius' ideal of a well-ordered society based on good government (see Xing; Mencius; Xunzi). Good government is responsive to the basic needs of the people, to issues of wise management of natural resources and to just distribution of burdens and benefits. In this vision of socio-political order, special emphasis is put on harmonious human relationships (*lun*) in accord with *de*, virtues or standards of excellence. This vision is often called *dao*, a term that has been appropriated by different classical schools of Chinese thought, including Daoism and Legalism (see Legalist philosophy, Chinese; Daoist Philosophy). In the

Analects, *dao* is sometimes used as a verb, meaning 'to guide'; sometimes it is used as a concrete noun, meaning literally 'road'. In the latter sense, it can be rendered as 'way'. But in distinct Confucian ethical usage, as commonly acknowledged by commentators, it is *dao* as an abstract noun that is meant, and more especially in the evaluative rather than descriptive sense, that is, as referring to the ethical ideal of a good human life as a whole (see Dao).

Throughout its long history, Confucianism has stressed character formation or personal cultivation of virtues (*de*) (see De). Thus it seems appropriate to characterize Confucian ethics as an ethics of virtue, not an ethics of rules or principles. To avoid misunderstanding, two explanations are in order. In the first place, the Confucian focus on the centrality of virtues assumes that the notion of *de* can be rendered as 'virtue'. In the second place, this focus does not depreciate the importance of rule-governed conduct (see Moral development).

Sinologists differ in their interpretation of the Confucian use of *de*. Some insist that *de* should be construed as 'power', 'force' or 'potency', and in Confucian usage it should be qualified as moral in contrast to physical force. Others argue for 'virtue' in the distinctively ethical sense, as pertaining to the excellence of a character trait or disposition. Interestingly, these two construals of *de* are not incompatible in the light of some English uses of 'virtue' (see Virtue ethics). In *Webster's Third New International Dictionary*, for example, the sixth definition of 'virtue' is: 'an active quality or power whether of physical or of moral nature; the capacity or power adequate to the production of a given effect', and the fifth definition is: 'a characteristic, quality, or trait known or felt to be excellent'. The *Oxford English Dictionary* likewise offers as definitions 'a good quality' and 'efficacy; inherent power'. Both these senses of 'virtue' are found in the classical Chinese uses of *de*.

There is, of course, the value-neutral sense of *de* that leaves open the question whether personal traits or qualities merit ethical approval, and this question is reflected in the distinction, still current in Modern Chinese, between *meide* and *ede*. The former pertains to 'beautiful' or 'commendable' *de*, and the latter to its contrary. *Meide* are those traits that are acquired through personal cultivation. The *Encyclopedia Dictionary of the Chinese Language* offers the following two entries for *de* in the ethical sense, one suggested by an interpretation of its homophone, meaning 'to get' or 'to obtain', found in the *Liji (Book of Rites)*: 'that which is obtained in the heart-and-mind as a result of personal cultivation', and 'the nature that is formed after successful personal cultivation'. Both these definitions involve *meide*, commendable, acquired qualities or traits of character, much in the sense of David Hume's 'personal merits' (see Hume, D.) *Ede*, on the other hand, are personal demerits or 'detestable' qualities of character.

Also important is the sense of *de* as power or force, in view of the Confucian notion of *junzi* (ethically superior or paradigmatic individuals). By virtue of ethical achievement, a *junzi* possesses the power of attraction or influence indicative of effective agency (see Moral agents). As Confucius remarks, the *junzi*, equipped with the virtues (*de*), never stands alone: 'He is bound to have neighbours.' 'The virtue (*de*) of the gentleman (*junzi*) is like the wind; the virtue (*de*) of the small man is like grass. Let the wind blow over the grass and it is sure to bend' (*Analects* 12.19, in Lau 1979: 115-16). In sum, the Confucian notion of *de* can be properly construed as ethical virtues that possess a dual aspect: an achieved condition of a person through self-cultivation of commendable character traits in accordance with the ideal *dao*, and a condition that is deemed to have the peculiar potency or power of efficacy in influencing the course of human life. The difficult problem is to present the Confucian *dao* and *de* as an ethics of virtue with a coherent conceptual scheme.

3 Basic notions and the problem of conceptual unity

The *Analects*, a composite work, is commonly considered the main and most reliable source of Confucius' teachings (see Confucius). It bequeaths to the Chinese tradition a large and complex ethical vocabulary, which contains a significant number of virtue (*de*) terms with implicit reference to the Confucian ideal of *dao*. Terms such as *ren* (benevolence), *yi* (rightness) and *li* (rites) seem to occupy a central position both in the *Analects* and throughout the history of Confucian discourse. Until recent times, however, few philosophical scholars of Confucianism attended to the problem of conceptual explication and the unity of these basic notions, that is, their presumed interconnection or interdependence in the light of *dao* as an ideal, unifying perspective. While most Confucian terms for particular virtues can be rendered into English without the need of elaborate explanation - for example, *xiao* (filiality), *yong* (courage), *wei* (dignity), *zhong* (fidelity), *ci* (kindness), *jing* (respectfulness) - the apparently basic notions (*ren*, *li* and *yi*) are not amenable to the simple expedience of translation and thus pose a

problem for conceptual analysis and interpretation. Moreover, existing translations of these terms ineluctably embody the writer’s interpretation, a sort of implicit commentary, representing the writer’s prior understanding of the translated texts.

Similarly, an explication of basic Confucian notions involves philosophical commentary, a familiar feature in the development of the history of Chinese thought. However, attempts at explication have been beset by a formidable difficulty, especially in defining the basic concepts of the Confucian framework. The pioneering study of the conceptual aspect of Confucian ethics is Chen Daqi’s *Kongzi xueshuo (Teachings of Confucius)* (Chen 1976). Chen reminds us that prior to interpreting the ideas of Confucius, it is essential to inquire into the conceptual status of some recurrent terms in the text. For determining the centrality or basic status of notions or concepts, Chen proposes four criteria, according to which basic concepts are (1) fundamental, (2) leading or guiding, (3) the most important (cardinal) or (4) the most comprehensive.

Fundamental concepts suggest the distinction between basic and derivative concepts. However, it is more plausible and accords better with Chen’s discussion to construe his distinction as one between basic and dependent concepts. Given our characterization of Confucian ethics as an ethics of virtues, this is a conceptual distinction between basic and dependent virtues. A concept may depend on another for its ethical significance without being a logical derivation. For instance, one cannot derive the concept of love from *ren*, yet its ethical significance depends on its connection with *ren*. This is perhaps the principal ground for Zhu Xi’s famous contention that *ren* cannot be equated with love, for it is the rationale of love (*ai zhi li*) (see Zhu Xi).

Leading or guiding concepts recall the purport of ethical terms as guides for action, informing the Confucian agent that the enduring significance of ethical endeavours lies in the pursuit of *dao* or the ideal of the good human life as a whole. *Cardinal concepts* and *comprehensive concepts* are the chief mark of basic ethical concepts. Comprehensive concepts also raise an issue in Confucian scholarship. Perhaps the issue can be settled if comprehensiveness is ascribed to *dao* or *ren* in the broad sense as signifying the holistic, ideal unifying perspective of Confucian discourse. Again, consider Zhu Xi’s thesis that *ren* (in the broad sense) embraces the four: *ren* (in the narrow sense), *yi* (rightness), *li* (rites) and *zhi* (wisdom). In terms of an ethics of virtue, the fundamental distinction is the distinction between cardinal and dependent virtues. Accordingly, Chen proposes that in addition to *dao* and *de*, the Confucian scheme consists of *ren*, *li* and *yi* as the basic, cardinal concepts. This thesis is well-supported by the recurrence of such concepts and their fundamental importance throughout the history of Confucianism.

4 The ethical framework of Confucianism

The foregoing pertains to the question of identifying basic, cardinal concepts as contrasted with dependent concepts. A more formidable problem remains as to how these basic concepts are related to one another. The following discussion presents a sketch of a philosophical reconstruction, which is essentially a conceptual experiment. The sketch offers a general characterization of Confucian ethics as a form of virtue ethics, and provides a sample of how such basic notions as *li* and *yi* can be shaped in response to questions deemed important for the development of a Confucian moral philosophy.

The Confucian ethical framework comprises the five basic concepts: *dao*, *de*, *ren*, *yi* and *li*. The best initial approach is to regard these, with a minimum of interpretation, as ‘focal notions’, that is, terms that function like focal lenses for conveying distinct though not unrelated centres of ethical concern. As generic terms, focal notions are amenable to specification in particular contexts, thus acquiring specific or narrower senses. This distinction is an adaptation of Xunzi’s distinction between generic terms (*gongming*) and specific terms (*bieming*). However, a term used as a specific term in one context may in another context be used as a generic term subject to further specification. In other words, the use of a term in either the generic or specific sense is entirely relative to the speaker’s purpose on a particular occasion, rather than to any theory concerning the intrinsic characters of terms or the essential attributes of things.

As noted earlier, *dao* is an evaluative term. Its focal point of interest lies in the Confucian vision of the good human life as a whole or the ideal of human excellence. Commonly rendered as ‘the way’, *dao* is functionally equivalent to the ideal ‘way of life’. Unlike other basic terms, *dao* is most distinctive as an abstract, formal term in the highest generic sense, that is, subject to general specification by way of such virtue (*de*) words as *ren*, *li* and *yi*. As *de* is an individual achievement through personal cultivation, when a person succeeds in realizing *dao*, they

have attained such basic *de* as *ren, li* and *yi*. The specification of *de*, apart from *ren, li* and *yi*, can take a variety of forms or dependent virtues such as filiality, respectfulness or trustworthiness. In this sense, *de* is an abstract noun like *dao*, but it depends on *dao* for its distinctive character. *De* is thus functionally equivalent to ethical virtue. Thus, the opening remark of the *Daxue (Great Learning)* points out that the way of great learning or adult, ethical education lies in the clear exemplification of the virtues (*ming mingde*) (see Daxue; Moral education; Moral development). With its emphasis on *dao* and *de*, Confucian ethics is properly characterized as an ethics of virtue, but more informatively as an ethics of *ren*, *li* and *yi*, relative to their concrete specification or particularization by terms of dependent virtues (for example, filiality, respectfulness, integrity). As generic, focal terms, *ren*, *li* and *yi* are specific terms relative to *dao* as a generic term. Differently put, implicit in Chen's account is a distinction between basic, interdependent virtues (*ren*, *li* and *yi*) and dependent virtues (filiality, respectfulness and so on) (Chen 1976). As indicated earlier, the latter are dependent in the sense that their ethical import depends on direct or indirect reference to one or more of the basic, interdependent virtues of *ren*, *li* and *yi* (see Virtue ethics).

5 *Ren*

While *ren*, commonly translated as 'benevolence', has a long history of conceptual evolution and interpretation, as a focal notion it centres its ethical interest on love and care for one's fellows, that is, an affectionate concern for the well-being of others - the one persistent idea in the Confucian tradition. For this reason, English translators often render *ren* as 'benevolence', suggestive of the use of 'benevolence' in Francis Hutcheson and David Hume. Perhaps a more literal translation of *ren*, however, is 'human-heartedness' or 'humanity', since *ren* is basically an idea of love or concern for the well-being of human community, a specification of the concrete significance of the Confucian *dao* (see Community and communitarianism §1). This core meaning of *ren* as fellow-feeling is found in the *Analects*; it is reported that Confucius once said to his disciple Zengzi, 'My *dao* has one thread that runs through it.' Zengzi construed this *dao* to consist of *zhong* and *shu*, an interpretation widely acknowledged as a method for pursuing *ren* (*Analects* 4.15, in Lau 1979: 74) While the relation between*zhong* and *shu* has divergent interpretations, *zhong* may be rendered as 'doing one's best', and *shu* as 'consideration' (of other people's feelings and desires). In this light, Confucian ethics displays a concern for both self-regarding and other-regarding virtues.

However, the acquisition of these virtues presupposes a locus in which these particular, dependent virtues are exercised. Thus among the dependent virtues, filiality, brotherhood or sisterhood are primary, for the family is the natural home and the foundation for the extension of *ren*-affection. Thus friendship is viewed as a fundamental human relationship along with those of husband-wife and parent-child. In Song and Ming Confucianism (for example, in Cheng Hao and Wang Yangming), *dao* is sometimes used interchangeably with *ren*. In this manner, *ren* attained the status of a supreme, all-embracing ethical ideal of the well-being of every existent thing, human or non-human, animate or inanimate. Confucius' vision of a well-ordered human society is transformed into the vision of the universe as a moral community. In this conception, anything that is an actual or potential object of human attention is considered an object of human concern. Exploitation of human and natural resources must be subject to evaluation in terms of *dao* as an all embracing ethical ideal of existence. This ideal of *dao* makes no specific demands on ordinary humans. For the most part, conflict of values is left to individual determination, though the welfare of the parents is always the first consideration. Thus, the concrete significance of *dao* is open to the exercise of *yi*, the agent's sense of rightness.

6 *Yi* and *li*

The exercise of *yi*, usually rendered as righteousness or rightness, depends on ethical education based on *daotong* or the tradition of the community of interpretation; that is, the reasoned interpretation of the educated members of the community informed by *sensus communis*, a sense of common interest, a regard for *dao* as the ideal unifying perspective. Disagreement or dispute on the pragmatic import of *dao* is expected, as members of the community of interpretation have their own conceptions of human excellence (*shan*) and possibilities of fulfilment. The ethical solution of conflict of interpretation lies in transforming the disagreement into agreement in the light of *sensus communis*, not in a solution defined by agonistic debate which presumes that there are impartial judges who can render their corporate decision in terms of majority vote. Unlike contemporary democratic polity, the majority rule cannot be reasonably accepted as a standard for setting ethical disagreement. The ethical tradition provides the background and guidelines to ethical conduct. Normative ethical theories have value because they provide different ways of assessing the significance of tradition. Like the basic concepts of Confucian ethics, they are focal

notions for important centres of ethical concern, for example, duty and interest (private or public), or in recent terminology, agent-relative and agent-neutral reasons for action (see Moral agents).

The notion of *li* focuses on the ritual code (see Law and ritual in Chinese philosophy). For this reason, it is commonly rendered as ‘rites’, ‘ritual’, ‘propriety’, or ‘ceremonials’. The ritual code is essentially a set of rules of proper conduct pertaining to the manner or style of performance. As *yi* is incompatible with exclusive regard for personal gain, the *li* set forth the rules of ethical responsibility. For Confucius and his followers, the *li* represent an enlightened tradition. As D.C. Lau has put it:

> The rites (*li*) were a body of rules governing action in every aspect of life and they were the repository of past insights into morality. It is, therefore, important that one should, unless there are strong reasons to the contrary, observe them. Though there is no guarantee that observance of the rites necessarily leads, in every case, to behaviour that is right, the chances are it will, in fact, do so.
>
> (Lau 1979: 20)

Yet, the ethical significance of *li* is determined by the presence of the spirit of *ren*. As Confucius once said, ‘If a man has no *ren*, what has he to do with *li*?’ (*Analects* 3.3, in Lau 1979: 67). Since the ritual code represents a customary practice, the early Confucians, particularly Xunzi and the writers of some chapters in the *Liji*, were concerned with providing a reasoned justification for compliance with *li* or traditional rules of proper conduct. As Arthur Waley aptly remarks: ‘it was with the relation of ritual [*li*] as a whole to morality and not with the details of etiquette and precedence that the early Confucianists were chiefly concerned’ (Waley 1938: 67) The same concern with reasoned justification is evident in Song and Ming Confucianists (for example, Zhu Xi and Wang Yangming) who maintain that the significance of *li* (ritual) lies in its rationale (*li zhi li*).

The ethical significance of *yi*, in part, is an attempt to provide a rationale for the acceptance of *li*. *Yi* focuses principally on what is right or fitting. The equation of *yi* with its homophone meaning ‘appropriateness’ is explicit in the *Zhongyong*, 20 (see Zhongyong), and is generally accepted by Confucianists (for example, Xunzi, Li Gou and Zhu Xi). However, what is right or fitting depends on reasoned judgement. As Xunzi observes: ‘The person concerned with *yi* follows reason’ (*Xunzi* XV, in Li 1979: 328) Thus, *yi* may be construed as reasoned judgement concerning the right thing to do in particular exigencies. Li Gou, a Song Confucian, justly reminds his readers that *yi* is ‘decisive judgement’ that is appropriate to the situation at hand.

In light of the foregoing, the interdependence of basic notions may be stated as follows. Given *dao* as the ideal of the good human life as a whole, *ren*, *li* and *yi*, the basic Confucian virtues (*de*), are constitutive rather than mere instrumental means to its actualization. In other words, the actualization of *dao* requires the co-satisfaction of the standards expressed in *ren*, *li* and *yi*. Since these focal notions pertain to different foci of ethical interest, we may also say that the actualization of *dao* requires a coordination of three equally important centres of ethical interest and endeavour. The connection between these foci is one of interdependence rather than subordination. Thus in the ideal case, *ren*, *li* and *yi* are mutually supportive and adherent to the same ideal of *dao*. When *dao* is in fact realized, *ren*, *li* and yi would be deemed constituents of this condition of achievement. When, on the other hand, one attends to the prospect of *dao*-realization, *ren*, *li* and *yi* would be regarded as complementary foci and means to *dao* as an end. In sum, *ren*, *li* and *yi* are complementary foci of human interest.

7 The scope and functions of *li*

Because of its distinctive character and role in Confucian ethics and its pervasive influence in traditional China and contemporary critique, the notion of *li* requires special attention (see Law and ritual in Chinese philosophy). Implicit in the notion of *li* is an idea of rule-governed conduct. In the *Liji (Book of Rites)*, the subject matter ranges from ritual rules (or formal prescriptions) concerning mourning, sacrifices, marriage and communal festivities to the more ordinary occasions relating to conduct towards ruler, superior, parent, elder, teacher and guest. Because of its emphasis on the form or manner of behaviour, *li* is often translated as ‘religious rites’, ‘ceremony’, ‘deportment’, ‘decorum’, ‘propriety’, ‘formality’, ‘politeness’, ‘courtesy’, ‘etiquette’, ‘good form’ or ‘good behaviour’. These renderings are misleading without understanding the different functions of *li*. At the outset, it is important to note that for Confucians the *li* are an embodiment of a living cultural tradition; that is, they are subject to modification in response to changing circumstances of society. Thus some writers of the *Liji* point out that the *li* are the prescriptions of reason, and that any ritual rule that is deemed right and reasonable (*yi*) can be considered a

part of *li*. On one plausible interpretation, the traditional ritual code represents no more than a codification of ethical experiences based on the concern with *ren* and *yi*. In this light, the *li* are in principle subject to revision or replacement. In the spirit of Zhu Xi, we may say that a Confucian must reject ritual rules that are burdensome and superfluous and accept those that are practicable and essential to the maintenance of a harmonious social order. However, any reasoned attempt to revise or replace *li* presupposes an understanding of their functions. It is this understanding that distinguishes the Confucian scholar from a pedant, who may have a mastery of rules without understanding their underlying rationales. For elucidation, we rely mainly on a reconstruction of Xunzi's view, since we find in some of his essays the most articulate concern for and defence of *li* as an embodiment of a living, cultural tradition (see Xunzi).

With regard to any system of rules governing human conduct, one can always raise questions concerning its purpose. In Confucian ethics, the *li*, as a set of formal prescriptions for proper behaviour, have a threefold function: delimiting, supportive and ennobling. The *delimiting* function is primary, in that the *li* are fundamentally directed to the prevention of human conflict. They comprise a set of constraints that delineate the boundaries of pursuit of individual needs, desires and interests. The *li* purport to set forth rules of proceeding in an orderly fashion, ultimately to promote the unity and harmony of human association in a state ruled by a sage king imbued with the spirit of *ren* and *yi*. This orderliness consists of social distinctions or divisions in various kinds of human relationships (*lun*), namely, the distinctions between ruler and minister, father and son, the eminent and the humble, the elder and the younger, the rich and the poor, and the important and unimportant members of society. In abstraction from the connection with *ren* and *yi*, the delimiting function of *li* may be compared to that of negative moral rules or criminal laws. Like rules against killing, stealing or lying, the *li* impose constraints on conduct. They create paths of obstruction, thus blocking certain moves of agents in the pursuit of their desires or interests. The *li*, in effect, stipulate the conditions of eligibility or permissibility of actions. They do not prejudge the substantive character or value of individual pursuit. They provide information on the limiting conditions of action, but no positive guidance as to how one's desires may be properly satisfied. Put differently, the *li* tell agents what goals *not* to pursue, but they do not tell them how to go about pursuing goals within the prescribed limits of action.

Apart from the delimiting function, the *li* have also a *supportive* function; that is, they provide conditions or opportunities for satisfaction of desires within the prescribed limits of action. Instead of suppressing desires, the *li* provide acceptable channels or outlets for their fulfilment. In an important sense, the supportive function of *li* acknowledges the integrity of our natural desires. So long as they are satisfied within the bounds of propriety, we accept them for what they are whether reasonable or unreasonable, wise or foolish, good or bad. The main supportive function of *li* is the redirection of the course of individual self-seeking activities, not the suppression of motivating desires. Just as the delimiting function of *li* may be compared with that of criminal law, their supportive function may be compared with that of procedural law, which contains rules that enable us to carry out our wishes and desires (for example, the law of wills and contracts). The *li*, like these procedural rules, aid the realization of desires without pronouncing value judgments.

The focus on the *ennobling* function of *li* is a distinctive feature of Confucian ethics and traditional Chinese culture (see Self-cultivation, Chinese theories of). The keynote of the ennobling function is 'cultural refinement', the education and nourishment (*yang*) of emotions or their transformation in accord with the spirit of *ren* and *yi*. The characteristic concern with the form of proper behaviour is still present. However, the form stressed is not just a matter of fitting into an established social structure or set of distinctions, nor is it a matter of methodical procedure that facilitates the satisfaction of the agent's desires and wishes; rather, it involves the elegant form (*wen*) for the expression of ethical character. In other words, the ennobling function of *li* is directed primarily to the development of commendable or beautiful virtues (*meide*). The 'beauty' (*mei*) of the expression of an ethical character lies in the balance between emotions and form. What is deemed admirable in the virtuous conduct of an ethically superior person (j*unzi*) is the harmonious fusion of elegant form and feelings. In the ideal case, a *li*-performance may be said to have an aesthetic dimension. In two different and related ways, a *li*-performance may be said to be an object of delight. In the first place, the elegant form is something that delights our senses. It can be contemplated with delight quite apart from the expressed emotional quality. In the second place, when we attend to the emotion or emotional quality expressed by the action, which we perceive as a sign of an ethical virtue or character, our mind is delighted and exalted, presuming of course that we are also agents interested in the

promotion of ethical virtues in general. Respect for traditional rites of mourning and sacrifices is an expression for the concern with *ren* and *li*, because such practices exemplify the Confucian ideal of humanity; it is a sort of religious concern without endorsing any religious beliefs and practices.

8 The Confucian tradition

If the set of *li*, or rules of proper conduct, is viewed as a repository of traditional insights, a question then arises concerning the Confucians' conception of their own tradition. Like other enduring ethical or religious traditions which extend across history (such as Buddhism or Christianity), the Confucian or Ru tradition, as embodied in the notion of *daotong*, has often been the target of the contemporary critique. According to one familiar appraisal, the Ru tradition is out of tune with our times. Like any cherished cultural artefact, it is best revered as a relic of the past. Moreover, many adherents of the Confucian tradition are dogmatic; they are unwilling to accept reasonable proposal for change or modification of some components of the tradition.

Replying to this charge, a Confucian thinker or scholar might point out that, to a certain extent, the critique is reasonable. Throughout its long history, there is a recurrent tendency of many adherents of the Confucian tradition to institute orthodoxy and uphold their perceived values of the tradition as the true values of the tradition, indifferent to the distinction between perceived and real values. A personal ascription of value to an object cannot be logically equated with the value inherent in the object. Understanding a living tradition requires an appreciation of the distinction between the actual past and the perceived past, and the distinction between a living, robust tradition and a dying, decaying tradition. In the words of Jaroslav Pelikan, 'Tradition is the living faith of the dead, traditionalism is the dead faith of the living' (Pelikan 1984: 65).

Confucianism, in the words of Thomé Fang, is 'a constructive philosophy of comprehensive harmony invested with creative energies of life' (Fang 1981: 33). Although this is a summary statement of the outcome of Fang's metaphysical inquiry, it is an apt characterization of the vitality of the Confucian *dao* or way as an ethical vision of the good human life, an ideal amenable to varying interpretations of its concrete significance in different times and places. *Dao* is an ideal theme rather than an ideal norm or supreme principle of conduct (Cua 1978, ch. 8). It is fundamentally a unifying perspective for harmonizing the diverse elements of moral experience. To adopt the *dao* is to see human life, in its morally excellent form, as possessing a coherence in which apparently conflicting elements are viewed as eligible components of an achievable harmonious order (Cua 1982, ch. 3). It is this vision of *dao* that renders intelligible the idea of a Confucian ethical tradition as a living tradition enriched by its historic past, interpreted by adherents as having a present and prospective significance (see Dao).

9 Xunzi and Zhu Xi

Xunzi and Zhu Xi are perhaps the most articulate proponents of the importance of learning the classics as an indispensable means to inculcating a respect for the historical aspect of tradition. Xunzi points out that learning is an unceasing process of accumulation of goodness, knowledge, and practical understanding: 'If an ethically superior person (*junzi*) studies widely and daily engages in self-examination, his intellect will become enlightened and his conduct be without fault.' In general, the programme of learning 'commences with the recitation of the classics and ends with the study of *li* (rituals)'. Its purpose, however, is 'first to learn to become a scholar and ultimately to become a sage'. The classics are presumed to embody the concrete significance of *dao*. For example, 'the *Odes* give expression to the will (*zhi*) or determination [to realize *dao*]; the *History* or *Book of Documents* to its significance in human affairs; the *Li* or *Rites* to its significance in conduct; the *Music* to its significance in promoting harmony; and the *Spring and Autumn Annals* to its subtleties.' Notably, these classics are not self-explanatory, thus guidance from perceptive teachers is required. For example, 'the *Li* and *Music* present us with models, but no explanations; the *Odes* and the *History* deal with ancient matters and are not always pertinent; the *Spring and Autumn Annals* are terse and cannot be quickly understood' (*Xunzi* I, in Li 1979: 2-14; trans. in Watson 1963: 15-20).

Indeed, for Confucius as for Mencius and Xunzi, the *li* represent an ethical and cultural tradition (see Confucius). Learning to be a moral agent consists in part in appreciating 'the accumulated wisdom of the past'. Compliance with *li* as a set of formal prescriptions of proper behaviour without regard to *ren* and particular circumstances would be deemed unreasonable. It is perhaps for this reason that Confucius stresses *yi* (a sense of rightness) rather than preconceived opinions in coping with novel and exigent situations. Mencius is emphatic that the relevance of

a *li*-requirement must be considered in the light of weighing (*quan*) particular circumstances. In a similar spirit, Xunzi emphasizes the exercise of *yi* in responding to changing circumstances of human life. In general, the enforcement of laws and other rules of conduct depends on superior or exemplary persons (*junzi*) who have not only a knowledge of the subject matter but also an understanding of their underlying purposes, in addition to having a sense of priority and an ability to respond appropriately to changing affairs. Learning to become a sage (*shengren*) must begin with the study of the classics with a view to discerning their actuating significance in living contexts (see Chinese Classics).

The Confucian respect for the authority of the past in relation to the present requires not only a critical understanding of the practical significance of the classics but, more importantly, a critical assessment of the ethical character and achievement of contemporary adherents of the tradition. Thus Xunzi urges his readers to distinguish common people from three sorts of Confucians (Ru): the vulgar or conventional (*suru*), the refined or cultured (*yaru*) and the great or sagely Confucians (*daru*). This distinction between three types of Confucians may be seen to be motivated by Xunzi's desire to respond to the internal challenges within the Confucian tradition - a reminder and a warning to his fellow Confucians against uncritical intellectual and practical affiliation (*Xunzi* VIII, in Li 1979). His critique of Mencius on the relation between morality and human nature is a well-known illustration of an internal conflict between different interpretations of a key aspect of Confucian ethical tradition. Like Mencius, Xunzi is also quite aware of the need to engage in argumentation to answer external challenges to the Confucian tradition (*Xunzi* XXIII, in Li 1979). For Mencius, there was no alternative but to engage in disputation since the *dao* of the sages, embodied in Yao and Shun, had in his own time been obscured by the prevailing teachings of Yang Zhu and Mozi.

In a similar spirit, Xunzi attacks the views of a number of philosophers, for example, Mozi, Hui Shi, Songzi and Shen Dao, as well as Laozi and Zhuangzi. Xunzi considers most of their doctrines pernicious because, though often presented in elegant language and composition, they also present erroneous conceptions of the distinction between right and wrong and between order and disorder. However, it is noteworthy that Xunzi's critique explicitly acknowledges that 'some of what they advocate has a reasoned basis, and thus their statements appear plausible' (*Xunzi* VI, in Li 1979: 94). In the light of his holistic conception of *dao*, many philosophical views are unacceptable not because they are totally erroneous, but because they grasp only partial aspects of *dao* and exaggerate these at the expense of other aspects. In sum, in Xunzi we find a Confucian philosopher who exemplifies a concern with defending the Confucian tradition against both internal and external challenges, and who is at the same time quite capable of critical adoption of non-Confucian views which he deems reasonably acceptable.

Zhu Xi's conception of the Confucian tradition, *daotong*, commonly rendered as 'transmission of *dao*', is a conception central to the orthodox tradition of neo-Confucianism (see Neo-Confucian philosophy). Because of its religious connotation, the notion of orthodoxy quite naturally suggests an uncritical adherence to conventional or currently accepted beliefs as possessing an unquestioned authoritative status. As de Bary has instructively shown, however, *daotong* is better construed as 'repossession of *dao*' or 'reconstitution of *dao*', recalling Pelikan's distinction between a living tradition and traditionalism (de Bary 1981: 2).

As Tu Wei-ming remarks, Zhu Xi's study of the classics represents 'serious efforts to revitalize the tradition [which] involved creative adaptation as well as faithful interpretations' (Tu 1979: 134). For example, in his detailed recommendations of classics for examinations, apart from some commentaries, Zhu Xi stresses the self-critical study of original texts: 'In studying the classics one must have regard for the considered views of former scholars and extrapolate from them, aware that they are not necessarily conclusive but must be weighed as to what they understood and what they missed; then finally all this must be reflected upon in one's own mind to verify it' (*Zhuzi yulei* 3:37, in Li 1962, 1638-42). Zhu Xi also recommends non-orthodox philosophers of early China such as Xunzi, Han Feizi, Laozi (see Daodejing) and Zhuangzi. De Bary rightly reminds us that Zhu Xi's notion of *daotong* should not be construed in a manner that suggests a passive reception and transmission of the Confucian Way, for it involves an activity of revitalization and rediscovery as well as recovery of the meaning of the tradition or, in contemporary language, an activity of constructive or reconstructive interpretation of the tradition: 'In fact, Chu Hsi [Zhu Xi] emphasized the discontinuities in the tradition almost more than the continuities, and underscored the contributions of inspired individuals who rediscovered or "clarified" the Way in new forms' (de Bary 1981: 99).

Both Xunzi and Zhu Xi emphasize the role of inspired and insightful persons regarding classical learning. This emphasis somewhat echoes Confucius' notion of *junzi* or paradigmatic individuals, that is, persons, who through their life and conduct embody *ren*, *li*, *yi* and other ethical excellences (see §2). The *junzi* are the exemplars of the actuating significance of the Confucian tradition, serving as standards of inspiration for committed agents. These paradigmatic individuals are generally ascribed an authority by their contemporary adherents. The authority, however, is based on the acknowledgement of their superior knowledge and achievement, and especially their exercise of yi in interpretive judgements concerning the relevance of rules in exigent situations, or in Xunzi's words, on *dao* and *de* (virtue). However, while possessed of an authoritative ethical status, as standard bearers and interpreters of culture the *junzi* are not arbiters of moral disputes, for their interpretations of the Confucian tradition represent individual efforts toward the repossession or reconstitution of *dao*, thus essentially contestable and subject to reasoned assessment of their contemporaries and posterity.

Implicit in the foregoing characterization of the Confucian tradition as a living tradition is the notion of tradition as an interpretive concept. For the most part, major Confucian philosophers stress the importance of classical education and the role of teachers as *junzi* or paradigmatic individuals, because they are thought to exemplify the best spirit of Confucius. Notably, Confucius' conception of *junzi* focuses on some salient characteristics. In addition to their pursuit of *dao*, the concern with *ren*, *li* and *yi*, they are also persons of integrity, catholicity and neutrality. The integrity is exemplified in their unwavering concern with the harmony of words and deeds (Cua 1978, ch. 5). This stress on the role of *junzi* as a guiding standard of education and moral conduct naturally gives rise to the question on the possibility of modification and transformation of the Confucian tradition. Indeed, if tradition is an interpretive concept, some account must be given on how changes may be viewed as consistent with the conservation of tradition.

10 Transformation of the Confucian tradition

In the Confucian context, this problem of transformation of the tradition is best considered in terms of the distinction and connection between *jing* (the standard) and *quan* (the weighing of circumstances). The former pertains to the operative, established standards of conduct, the latter to the specific circumstances. *Quan* may be regarded as *jing* prior to the determination of its significance in actual, particular situations. In other words, *jing*, the established standards of conduct, depend on *quan* in the sense that their application is undetermined prior to actual circumstances. When such a determination is made, *quan* would become part of *jing*, given the assumption that the judgments are accepted by the Confucian community of interpretation. Concisely put, *jing* is the determinate *quan*, and *quan* is the indeterminate *jing*. Of course, both represent the constant (*chang*) and the changing (*bian*) aspects of *dao*.

More importantly, the exercise of *quan* entails a concern with *yi* or rightness of judgement. Right judgement is necessary not only in dealing with the hard cases of the moral life, but also with normal situations (*jing*) where the intelligent adherent of the tradition confronts the problem of interpretation and application. In both sorts of situation the agent has the same moral objective, that is, to do the right thing. This is perhaps the point of Zhu Xi's saying that *quan* is unavoidably exercised in changing situations that fall outside the scope of the regular practice of *daoli* (moral norm):

> When *quan* attains equilibrium (*zhong*), it does not differ from *jing*', that is, they have achieved the same moral objective. In both cases they are governed by the exercise of *yi* or one's sense of rightness. For Zhu Xi, *yi* constitutes the *guan*, the thread that runs though *jing* and *quan*.
>
> (*Zhuzi yulei* 3:37, in Li 1962, 1638-42)

The foregoing reflections on the distinction and connection between *jing* and *quan* afford us an answer to the question of the possibility of change or modification of the Confucian tradition of *dao*. Intelligent adherents of the Confucian tradition can deploy the sceptical challenge without fear of the corruption of the tradition (see Moral scepticism). The dynamic interplay of *jing* and *quan* in relation to constancy (*chang*) and change (*bian*) shows that the distinction is not a fixed or absolute distinction. As Dai Zhen points out, what is deemed important or unimportant in our assessment will vary in different times and places (Dai 1777 [1975]: 125). Put differently, the substantive content of the tradition of *dao* cannot be considered settled without prejudging the merits of particular circumstances. The exercise of *quan* in any situation requires a careful examination and analysis of all the relevant

factors involved.

Quite naturally, one may raise the question of justification for such judgements in exigent circumstances. In normal situations, the *li* are quite sufficient for action guidance insofar as they are informed by a concern for *ren*. The problem of reasoned justification for such judgments has not received attention from most Confucian thinkers, except Xunzi. However, even in the works of Xunzi we do not find any articulate answer to this problem, though his works do provide materials for constructing a response. In the first place, Xunzi is explicit that any discussion is valuable because there exist certain standards for assessment, and these standards pertain principally to conceptual clarity, respect for linguistic practices and evidence, and the requirement of consistency and coherence in discourse. A philosophical reconstruction along the lines of the theory of argumentation presents an interesting Confucian view of justification in terms of rational and empirical standards of competence, along with certain desirable qualities of participants in ethical discourse. In this reconstruction of Xunzi's works, ethical justification is a phase of discourse, preceded by explanatory efforts in the clarification of normative claims responsive to a problem of common concern among participants. This in turn presupposes that queries concerning the proper uses of terms are understood by participants in ethical argumentation (Cua 1985a).

11 Historical appeal

As widely noted by contemporary writers on Confucian ethics, in argumentative discourse the Confucians are fond of appeal to historical events and paradigmatic individuals. This ethical use of historical knowledge and beliefs is a pervasive feature of Confucianism. A serious student of Chinese philosophy cannot accept the common view that Chinese thinkers are inclined to be uncritical towards their history, or that they believe that history provides the exclusive guidance for present problems and perplexities. A careful examination of the works of Xunzi refutes this thesis. The Confucian respect for tradition and its history, and for exemplary historical persons and events, is best appreciated in terms of four ethical, argumentative uses of the past or historical appeal. These are the pedagogical, rhetorical, elucidative and evaluative uses (see History, Chinese theories of).

For Xunzi, the primary aim of moral education is the transformation of man's native but problematic motivational structure (for example, feelings and desires) by way of knowledge of standards of *ren*, *yi* and *li*. One learns for the sake of doing; knowing the right and the good is for the sake of acting in accord with such knowledge. Learning, however, is not equivalent to the mere acquisition of knowledge but more essentially requires understanding and insight. For Xunzi, the basic philosophical issue in moral education pertains to the rational coordination of the intellectual and volitional activities of the mind (*xin*) by means of the *dao*. Given the autonomy of mind, one can choose to accept or reject its guidance. If the mind is directed by reason and nourished with clarity and is not perturbed by extraneous matters, 'it will be capable of determining right and wrong and of resolving doubts' (*Xunzi* XXI, in Li 1979: 490; trans. in Watson 1963: 131). In sum, the pedagogical use of the historical appeal points not only to models who are worthy or unworthy of emulation, but more significantly to models functioning as reminders in moral learning and conduct that appeal especially to what is deemed in the real interests of the learner. Notably, gentle suasion rather than coercion is involved, as the individual still retains the freedom to accept or reject it. This constitutes the argumentative value of the pedagogical function of the historical appeal rather than the mere exhibition of moral exemplars. The latter's effectiveness lies primarily in the appeal to paradigmatic individuals (*junzi*) (see Moral education; Moral development).

The rhetorical function of the historical appeal is essentially a problem of assurance, of appeal to plausible presumptions or shared beliefs and trustworthiness (*xin*) (see Xin). When I present reasons supporting my interpretation of the Confucian tradition, I will deploy those unquestioned assumptions in discourse. Also, I expect my fellow Confucians to believe me when I cite information that is presumably known to members of my community. Obviously, the historical appeal functions as a technique for assuring the audience that the thesis maintained is consonant with shared historical beliefs, and obviates having to state reasons for their support. It also constitutes a technique of discounting alternative views, shifting the burden of proof to a possible adversary. These uses, of course, are subject to further challenge, for the rational acceptability of the thesis and an assurance that the thesis is not a mere imaginative contrivance to avoid questioning.

The elucidative function of the use of history attempts to clarify the relevance of the past for the present. This is basically an expression of the respect for tradition, clarifying, in interpretation, one's intelligent, critical understanding of the tradition. Of course, the question of evaluation arises in serious Confucian discourse. Inherent

in the use of historical characters for elucidating and demonstrating the applicability of an ethical thesis is the implicit claim that human history is the proper subject of ethical appraisal. However, this claim is explicit in the evaluative use of the historical appeal.

Notably, the elucidative use of the historical appeal is retrospective; that is, it is the use of the appeal to the past for judging the present (*yigu chijin*); while the evaluative use is prospective, that is, for the sake of determining the relevance of the past to the present (*yijin chigu*). In the retrospective use, Xunzi is clearly recommending the adoption of a standpoint based on historical knowledge or beliefs for the purpose of maintaining or assessing the adequacy of current beliefs. The prospective use, however, stresses our knowledge and understanding of our present problematic situations as a basis for assessing the unexamined claims based on the past as a guidance to the present. Both the retrospective and prospective uses of the historical appeal are essentially critical, and can be used either in the positive defence of one's thesis or in the negative evaluation of another's thesis or both. In J.O. Urmson's words, we have here a distinction between *standard-invoking* and *standard-setting* aspects of normative discourse (Urmson 1968: 68). In effect, the retrospective use of the historical appeal invokes an established framework with its operative criteria or standards for justification of ethical judgments. Of course, one can always raise external questions about an established practice.

The prospective use of the historical appeal reverses the standpoint of the retrospective one. Human history is seen as a subject matter rather than as a basis for ethical judgment. In adopting this standpoint, one can no longer avail oneself of the presumption of the truth of historical beliefs without critical examination, nor can one presume the existence of shared historical knowledge. The key issue here lies in the present evidential grounding of ethical claims (for further discussion see Cua 1985b).

In closing, it may be noted that Confucian ethics, like any normative system, presents conceptual problems of interpretation and reconstruction. Attention here has been focused on the Confucian scheme of basic notions and the idea of Confucian tradition (*daotong*). Doing full justice to the range of Confucian concerns also involves a critical consideration of divergent views on such topics as the relation of the basic notions to human nature, the role of emotions in ethical deliberation and justification, the place of law in an ideal social, political union, and the metaphysical grounding of *dao*, particularly in the connection of the human and the natural orders. A serious student of Confucian philosophy must consider these problems with reference both to the divisive efforts and tension within the history of Chinese philosophy, and to the significance of these efforts to of Western philosophical inquiries.

See also: Cheng Hao; Cheng Yi; Chinese philosophy; Buddhist philosophy, Chinese; Chinese Classics; Legalist philosophy, Chinese; Neo-Confucian philosophy; History, Chinese theories of; Confucius; Dai Zhen; Dao; Daoist philosophy; Daxue; De; Dong Zhongshu; Ethical systems, African; Ethics in Islamic philosophy; Family, ethics and the; Han Yu; Confucian philosophy, Japanese; Jia Yi; Confucian philosophy, Korean; Law and ritual in Chinese philosophy; Li; Lu Xiangshan; Moral education; Moral development; Morality and ethics; Self-cultivation, Chinese theories of; Shao Yong; Virtue ethics; Wang Fuzhi; Wang Yangming; Xunzi; Zhang Zai; Zhongyong; Zhou Dunyi; Zhu Xi

A.S. CUA

References and further readings

Chan Wing-tsit (1955) 'The Evolution of the Confucian Concept *Jen* [*ren*]', *Philosophy East and West* 4: 295-319.(A classic study of the concept of *ren* in the history of Confucianism.)

Chan Wing-tsit (1963) *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press.(A translation of and commentary on selections from the Chinese philosophical tradition.)

Chan Wing-tsit (1975) 'Chinese and Western Interpretation of *Jen* [*ren*] (Humanity)', *Journal of Chinese Philosophy* 2: 107-29.(An informative survey of contemporary interpretations of *ren*.)

Chan Wing-tsit and Fu, C. (1978) *Guide to Chinese Philosophy*, Boston, MA: C.K. Hall.(A contemporary bibliography of scholarly work on Chinese philosophy.)

Chang, C. (1962) *The Development of Neo-Confucian Thought*, New York: Bookman, 2 vols.(A standard interpretation of medieval Chinese philosophy.)

Chen Daqi (1976) *Kongzi xueshuo (Teachings of Confucius)*, Taibei: Zhengchong.(A reconstruction of Confucian philosophical terms.)

Chen Daqi (1977) *Pingfan de daode guan (Ordinary View of Morality)*, Taibei: Chonghuan.

Cua, A.S. (1978) *Dimensions of Moral Creativity: Paradigms, Principles, and Ideals*, University Park, PA: Pennsylvania State University Press.(Chapters 4 and 5 deal with Confucian ethics focusing on paradigmatic individuals as embodiments of *ren*, *yi* and *li*.)

Cua, A.S. (1982) *The Unity of Knowledge and Action: A Study in Wang Yang-ming's Moral Psychology*, Honolulu, HI: Unversity of Hawaii Press.(Emphasis on the practical import of moral learning against the background of a Confucian vision of the universe as a moral community.)

Cua, A.S. (1985a) *Ethical Argumentation: A Study in Hsün Tzu's [Xunzi's] Moral Epistemology*, Honolulu, HI: University of Hawaii Press.(A Xunzi-based philosophical study of Confucian ethical argumentation with emphasis on the character of the participants, standards of competence, uses of definition and diagnosis of erroneous ethical beliefs.)

Cua, A.S. (1985b) 'Ethical Uses of the Past in Early Confucianism: The Case of Hsün Tzu [Xunzi]',*Philosophy East and West* 35 (2): 133-56.(Examination of Xunzi's Confucianism.)

Cua, A.S. (1989) 'The Concept of *Li* in Confucian Moral Theory', in R.E. Allison (ed.) *Understanding the Chinese Mind: Philosophical Roots*, Hong Kong: Oxford University Press.(Discussion of the delimiting, supportive and ennobling functions of *li* and their justification.)

Cua, A.S. (1992) 'The Idea of Confucian Tradition', *Review of Metaphysics* 45: 803-40.(Discussion of the concept of living ethical tradition and its application to Confucian ethics.)

Cua, A.S. (1997) *Moral Vision and Tradition: Essays on Chinese Ethics*, Washington, DC: Catholic University of America Press.(Fourteen essays on Chinese ethics dealing with such topics as opposites as complements and 'forgetting morality' in Laozi and Zhuangzi, the Daoist vision and Confucian vision of *ren*, self-deception, tradition and conceptual framework, and ground rules for settling intercultural conflicts.)

Dai Zhen (1777) *Mengzi ziyi shucheng (Commentary on the Meanings of Terms in the Book of Mencius)*, Taibei: Holu, 1975; trans. A. Chen and M. Freeman, New Haven, CT: Yale University Press, 1990.(A commentary on some of the key notions in the works of Mencius and other Confucian classics.)

de Bary, W.T. (ed.) (1970) *Self and Society in Ming Thought*, New York: Columbia University Press.(An anthology of studies of different schools of thought on these subjects.)

de Bary, W.T. (1981) *Neo-Confucian Orthodoxy and the Learning of the Mind-and-Heart*, New York: Columbia University Press.(A study of the neo-Confucian conception of orthodoxy tracing the idea of orthodoxy from Zhu Xi through the early Ming times, and *xinxue*, the development of Zhu Xi's ideas of learning of mind and heart, and their influence in Tokugawa Japan.)

de Bary, W.T. (1983) *The Liberal Tradition in China*, Hong Kong and New York: Chinese University of Hong Kong Press.(A discussion of this topic in terms of the history and contemporary significance of neo-Confucianism, emphasizing such problems as education, individualism, moral cultivation and critiques of political despotism.)

Fang, T.H. (1981) *The Chinese View of Life: The Philosophy of Comprehensive Harmony*, Taipei: Linking.(An important contribution to the unity of Chinese thought by an influential Chinese philosopher in the Chinese-speaking world. More extensive work is found in T.H. Fang, *Chinese Philosophy: Its Spirit and Development*, Taipei: Linking, 1981.)

Fingarette, H. (1972) *Confucius: The Secular as Sacred*, New York: Harper & Row. (A revision of the classical evaluation of Confucius.)

Fung Yu-lan (1952) *A History of Chinese Philosophy*, trans. D. Bodde, Princeton, NJ: Princeton University Press. (The standard history of Chinese philosophy.)

Graham, A.C. (1989) *Disputers of the Tao: Philosophical Argument in Ancient China*, La Salle, IL: Open Court. (A detailed philosophical engagement with the classical period.)

Hall, D.L. and Ames, R.T. (1987) *Thinking Through Confucius*, Albany, NY: State University of New York Press.(A conceptual reconstruction of Confucian terms based primarily on the *Analects*.)

Kao Ming (1986) 'Chu Hsi [Zhu Xi] and the Discipline of Propriety', in W. Chan (ed.) *Chu Hsi and Neo-Confucianism*, Honolulu, HI: University of Hawaii Press.(A discussion of the role of propriety (*li*) in the philosophy of Zhu Xi.)

Knoblock, J. (ed. and trans.) (1988-94) *Xunzi: A Translation and Study of the Complete Works*, Stanford, CA: Stanford University Press.(The new standard translation.)

Lau, D.C. (ed. and trans.) (1979) *Confucius: The Analects*, New York: Penguin.(A standard translation of the *Analects* with a good introduction.)

Li Disheng (ed.) (1979) *Xunzi jishi*, Taibei: Xuesheng.(A good modern Chinese annotated edition of *Xunzi*.)

Moore, C. (ed.) (1967) *The Chinese Mind: Essentials of Chinese Philosophy and Culture*, Honolulu, HI: East-West Center Press.(A selection of papers from the proceedings of four East-West Philosopher's

Conferences (1939, 1949, 1959, 1964). This volume presents informative papers on some general aspects of Chinese ethics, epistemological methods, scientific spirit, metaphysics and so on.)

Munro, D.J. (1969) *The Concept of Man in Early China*, Stanford, CA: Stanford University Press.(A discussion of what it means to be a person in the classical schools of Chinese philosophy.)

Naess, A. and Hannay, A. (eds) (1972) *Invitation to Chinese Philosophy*, Oslo: Universitetsforlaget.(An important anthology of recent comparative Chinese and Western introductions to Chinese philosophy, dealing with such topics as Chinese wisdom, paradigmatic individuals, neo-Confucianism, Daoism, Chan (Zen) Buddhism and so on. Seven of the eight essays by different authors were reprinted from *Inquiry* (1971).)

Pelikan, J. (1984) *Vindication of Tradition*, New Haven, CT: Yale University Press.(A notable, brief defence of tradition as a basis of thought and action by an eminent historian of Christianity.)

Rescher, N. (1977) *Dialectics*, Albany, NY: State University of New York Press.(An excellent introduction to the disputational approach to inquiry from a socially oriented perspective.)

Schwartz, B. (1985) *The World of Thought in Ancient China*, Cambridge, MA: Harvard University Press.(A broad-ranging and critical interpretation of the classical tradition.)

Tu Wei-ming (1979) *Humanity and Self-Cultivation*, Berkeley, CA: Asian Humanities Press.(A collection of interpretive essays on the Confucian tradition.)

Urmson, J.O. (1968) *The Emotive Theory of Ethics*, London: Hutchinson University Library.(A critical discussion of the development of emotive theory from C.K. Ogden and I.A. Richards to C. Stevenson's *Ethics and Language*.)

Waley, A. (ed. and trans.) (1938) *The Analects of Confucius*, New York: Random House.(A readable translation of the *Analects* of Confucius with a good introduction to some of its principal concepts.)

Watson, B. (trans.) (1963) *Hsün Tzu [Xunzi]: Basic Writings*, New York: Columbia University Press.(An accessible and readable translation of ten of Xunzi's major essays.)

Zhu Xi (1130-1200) *Zhuzi yulei (Classified Conversations of Zhu Xi)*, ed. Li Chengde, Taipei: Zhengzhong, 1962, 3 vols.(An important collection of Zhu Xi's discourses on various aspects of his thought; an indispensable source.)

Confucian philosophy, Japanese

Confucian philosophy is said to have arrived in Japan as early as the third century AD, but it did not become a subject of meaningful scholarly inquiry until the seventh century. The 'Confucianism' to which Japanese elites and scholars were first attracted represented fields of knowledge concerned more with ontology and divination than with social ethics and politics. Because of the priority given to birth over talent in official appointments, Confucianism in Japan remained more a gentlemanly accomplishment and never approached the status it had in China, where mastery of its teachings represented a gateway to officialdom. Intellectually, Confucian philosophy was overshadowed both in Japan and on the continent at this time by the teachings of Buddhism, which provided answers both to spiritual and metaphysical concerns.

Confucianism in China was refashioned in the eleventh and twelfth centuries by a number of scholars, of whom Zhu Xi was the most prominent. He revised the curriculum, restored social and ethical concerns to positions of centrality within the tradition and formulated a new rationalistic ontology. His teachings won a broad following among intellectuals in China and eventually earned the government's endorsement as the official interpretation for China's examination system.

From the seventeenth century onwards, Zhu Xi's teachings reached a comparably distinguished position within scholarly circles in Japan, though the government's endorsement of the Hayashi family as official interpreters of Zhu Xi's teachings was the limit of the official authorization of that philosophy in Japan. Though the idealistic Wang Yangming school challenged Zhu Xi's teachings in Japan as it had in China, the more effective challenge was mounted by the classicist teachings known as Ancient Studies. These scholars, of whom the best known was Ogyū Sorai, sought the 'true message of the sages' by emphasizing direct study of the ancient core texts of Confucianism rather than the exegesis on those classics by Zhu Xi and others.

Confucian philosophy contributed to the rationalism, humanism, ethnocentrism and 'historical mindedness' of Tokugawa Japan. The teachings were also responsible for changing fundamental ontological and epistemological assumptions, while also opening intellectual circles to unprecedented pluralism and diversity. Towards the end of the Tokugawa period in the mid-nineteenth century, Confucian philosophy (particularly in the variety fashioned by Wang Yangming) also provided inspiration and justification for those activist reformers who succeeded in overthrowing the old order.

During the modern period, Confucian philosophy has been identified with the Tokugawa tradition which has been at times idealized and and at other times vilified. Nonetheless, a number of the assumptions central to Confucian philosophy continue to characterize much popular and intellectual thought in contemporary Japan, as well as those ethics that tend to be most admired, even though actual knowledge of Confucian philosophy does not appear to be widespread any longer in Japan.

1 Confucian philosophy in early Japan

The earliest extant Japanese histories record that in AD 285 - the actual date was probably a century or so later - Wani, of the Korean kingdom of Paekche, brought copies of the *Analects* (*Lunyu*; in Japanese, *Rongo*) of Confucius and the *Qianziwen* (Thousand Character Classic; *Senjimon* in Japanese) from Korea to Japan (see Confucian philosophy, Korean). Even though most scholarship on Japan tends to identify this introduction of Confucian texts with the introduction of Confucian philosophy, it is nonetheless clear that immigrants from China and Korea who were familiar with the Confucian classics surely preceded the gift of these texts and probably themselves represent the actual introduction of Confucianism to Japan.

Confucianism does not appear to have provoked the same measure of suspicion or antagonism as other forms of continental East Asian civilization being introduced to Japan at this time, and the fact that Japanese interest in Confucianism continued to grow is suggested by the arrival in the sixth century of authorities from Paekche on the Confucian classics. Thereafter, with the coincident introduction of Buddhism (also from Korea), Confucian philosophy came to be regarded in Japan as one component of the richly variegated culture and civilization of continental East Asia, both of which were increasingly welcomed and responsible for major changes in Japan.

The 'Confucianism' to which Japanese elites were first exposed, however, was not the ethical socio-political

teaching represented by texts like the *Analects* or the *Mengzi* (see Confucius; Mencius) which were not at that time regarded as the core of the tradition. Instead, this early continental Confucianism represented the elaboration of metaphysical and cosmological constructions dating from Han dynasty China (206 BC- AD 220) (see Chinese philosophy). According to these teachings, an emperor was the supreme Son of Heaven whose correct performance of his role as an intermediary between Heaven above and Earth below was essential to the harmonious processes of the cosmos (see Tian). According to this doctrine human responsibilities lay principally in the area of governmental administration, which if properly executed would then ensure that everything from seasonal change to plentiful harvest would contribute to terrestrial well-being, and of course the greatest responsibility thus rested with those in the highest positions. Accordingly, those in Japan who aspired to a more powerful Chinese-style imperial institution often invoked these claims to further their own interests. Further, by positing the interrelatedness of all phenomena in the human realm with those of the celestial and terrestrial realms, this early form of Confucianism also taught that an understanding of the whole might be inferred from an understanding of individual particulars, and so it came to include such divinatory exercises as geomancy, astrology and numerology. These too enjoyed considerable appeal in Japan.

However, because long-standing aristocratic principles in Japan gave priority to pedigree over either virtue or administrative performance, Japanese Confucianism differed from that on the continent by marginalizing both discussion of imperial virtue as it was believed to promote effective government, and assessment of administrative performance as it was likewise believed to entitle one to career advancement. Thus, it was characteristic in Japan before the seventeenth century to consult 'Confucian' philosophy on such matters as determining auspicious dates or sites for important events or constructions, but only rarely does one find meaningful engagement of Confucian wisdom on how best to order the affairs of human beings or the state.

One example of this peripheral role for Confucian philosophy may be discerned in Japan's first important statement of goals for the polity in terms of the new political thought from the continent, the so-called 'Seventeen Article Constitution' of 604. In this document, its author Prince Shōtoku exhorts Japanese leaders to accept the continental imperial principle using references to Confucianism as only one possible source of justification, with other often more persuasive arguments drawn from such alternative forms of thought as Buddhism and Legalism (see Shōtoku Constitution). Buddhism in Japan, as in China, succeeded in preempting Confucian claims to authority in the realm of governance by itself asserting its role in 'protecting' the government from both natural and supernatural threats (see Buddhist philosophy, Chinese).

This, however, did not prevent Confucianism from making contributions in other areas. For example, one may discern the influence of Confucian philosophy in the seventh- and eighth-century impulse to record Japanese history for posterity, and it is clear that the early refashioning of the previously oral historical tradition sought to depict at least some emperors in the classic guise of benevolent Confucian-style monarchs. These earliest histories in Japan, however, did not reflect certain other perhaps more fundamental Confucian historiographical principles, such as the notion of dynastic change, the Mandate of Heaven as conferred upon those whose virtue entitles them to rule, and historical evidence of Heaven's rewarding the rectification of names whereby responsibilities were expected to conform to title and position.

Further, from time to time as during the reigns of Emperors Daigo (897-930) and Murakami (946-67), Confucianism did receive attention at the imperial court, but such occasions were the exception rather than the rule. Despite the fact that Confucian texts remained part of the curriculum of those aristocrats who trained for positions in service to the imperial court - positions to which their aristocratic birth gave them first claim - Confucian philosophy *per se* was quickly subordinated to Buddhism and came to be regarded for the most part as a teaching of only secular value whose larger truths might all be found in more developed fashion within Buddhism. Indeed, when the Buddhist pioneer Kūkai sought to represent all religious consciousness in terms of ten stages, he ranked Confucianism near the very bottom and second only to the debased consciousness of animal passions, a perspective that was shared widely during succeeding centuries in Japan.

2 Zhu Xi and the Confucian revival

During the eleventh and twelfth centuries in China, where Confucian philosophy had similarly been subordinated to Buddhism, Confucian teachings were revitalized. The curriculum was refashioned to restore the ethical writings of figures like Confucius and his interpreters to a central position, and metaphysical and cosmological teachings

were restructured in a manner that undermined Buddhism's heretofore hegemonic claims to authority in those areas. The central figure in this development was Zhu Xi (1130-1200) who, without diminishing his achievement, is often described as having synthesized the teachings of a number of earlier figures including Zhou Dunyi, Zhang Zai, Shao Yong and the brothers Cheng Hao and Cheng Yi.

Zhu Xi understood the world to be constructed of a combination of principle (*li*; in Japanese, *ri*) and material force (*qi*; in Japanese, *ki*). Principle, a singular term which included both natural and moral principles, has ontological priority over material force and is complete, unchanging, eternal and good. Material force, by contrast, represents the physical stuff of the universe, is changeable, and contains both good and bad elements (see Qi). Together, principle and material force were thus seen by Zhu Xi as comprising the entire physical universe and as operating according to the same rhythms of change and stasis that govern the cosmos.

Within human beings, according to Zhu Xi, principle corresponds to one's original nature (*benran zhi xing*; in Japanese, *honzen no sei*), or in other words, that disposition which humans share at the moment of birth and which is likewise naturally and equally good among all people (see Xing). Similarly, in humans material force is represented by one's specific nature (*qizhi zhi xing*; in Japanese, *kishitsu no sei*), that disposition which is specific to oneself and which varies in its turbidity from one person to the next. Under ideal circumstances when one's emotions are not aroused to excess, one can respond naturally to one's original nature and one's behaviour accordingly will seemingly spontaneously conform to correct norms. At other times, however, one's emotions and desires are agitated to the extent that they interfere with the ability to respond to others with goodness or in a humane manner. Thus the key to goodness, Zhu Xi taught, was learning how to still one's specific nature so that one's original nature might be manifested in all its excellence. This exercise in turn can be facilitated by either of two ways: first, one can study the external world in order to discern the role of principle within it, recognizing that by apprehending truths concerning principle in the world, one has likewise learned something about one's original nature; alternatively, one may cultivate the quality of seriousness (*qing*; in Japanese, *kei*) through such practices as quiet sitting or study of the canonical literature, thereby learning how to still one's potentially disruptive emotional excesses. This twofold praxis gave Zhu Xi's philosophy both intellectually and spiritually compelling properties.

These teachings, often styled 'neo-Confucianism' in European and North American scholarship, reached Japan in the thirteenth century, but aside from a brief period of vogue at the imperial court of Emperors Hanazono (1308-18) and Godaigo (1318- 39), they were taught almost exclusively in Zen monasteries and hence were not regarded as a subject for serious study independent from Buddhism. Such practises as Confucian quiet sitting were thus represented in the monasteries as simply less well-developed versions of such Zen teachings as 'sitting in meditation' (*zazen*). Thus, even though Confucian texts were also included in the curriculum of the Ashikaga *gakkō* (academy) at which scions of this distinguished family trained for service in the *bakufu* (central government), the teachings had the status more of one of the polite gentlemanly accomplishments than that of political philosophy with its own intellectual integrity. The contrast with the continent could not in this respect have been much greater, for there Zhu Xi's teachings were at the root of a radical change in both ontology and epistemology from a fundamentally Buddhist orientation to one which was grounded in Confucianism (see Neo-Confucian philosophy).

3 Tokugawa Confucianism

The relatively low regard in which Confucian philosophy was held in Japan prior to the seventeenth century changed during the Tokugawa period (1600-1868). Spurred on the one hand by the arrival of new Confucian texts from Korea as part of the booty from Japan's unsuccessful invasion of that country, and on the other hand by the Tokugawa government's eventual understanding of the utility of secular ideology in bolstering its claims to legitimate rule, Confucian teachings for the first time in Japan enjoyed the status of an independent subject of scholarly inquiry. Recent scholarship has demonstrated, however, that Confucianism's displacement of Buddhism's hegemony in these areas was a somewhat slower process than had previously been understood, and required the better part of a century to accomplish.

A pioneering figure in this development was Fujiwara Seika (1561-1617), a former Zen priest who became fascinated with the rich diversity of continental Confucian philosophy reflected in the newly imported texts. Seika was attracted to and taught doctrines from a broad range of Chinese and Korean Confucian scholars, including those who challenged the authority of Zhu Xi's teachings. It was Seika's disciple Hayashi Razan (1583-1657)

who, by contrast, became an early champion of the version represented by the teachings of Zhu Xi. Since these teachings had been sanctioned in China since the early fourteenth century as the correct interpretation for the official state-sponsored exams, they were essential to the curriculum of all who aspired to government service in China, and if anything they enjoyed even greater favour among aristocratic and official elites in Korea (see Confucian philosophy, Korean). Hayashi Razan initiated efforts, later continued by his son Hayashi Gahō (1618-80), to win official endorsement of these teachings by Japan's central military government with the Hayashi themselves as the authorized interpreters of this tradition. These efforts bore fruit later when, in 1691, the Shogun Tokugawa Tsunayoshi donated land for the relocation of the Hayashi family's school in Edo (modern Tokyo) and appointed Razan's grandson Hayashi Hōkō (1644-1732) as its head, a position that the Hayashi maintained within their family through the remainder of the Tokugawa period.

Outside the Hayashi family, Yamazaki Ansai (1618-82) was probably the best known Japanese spokesman for Zhu Xi's teachings. Yamazaki Ansai founded his own school, the Kimon, in which he stressed an uncompromising moral rigour in one's pursuit of sagehood, with particular emphasis on the cultivation of seriousness. Ansai also formulated his own version of Shintō called Suika in which somewhat arbitrary correspondences were drawn between elements in the Confucian metaphysic and Shintō noumena. The Kimon was by far the most successful private academy in Japan devoted to the teachings of Zhu Xi, though Ansai's attempt to reconcile Confucianism with Shintō and his tradition of moralism meant that the academy was never without its share of controversies.

Zhu Xi's teachings were responsible in Japan for a complex intellectual legacy which included humanism, rationalism, ethnocentrism and 'historical mindedness'. The humanism was related to Zhu Xi's conviction that the responsibility for maintaining the delicate balance that lay at the heart of both the individual and the cosmos lay squarely on the shoulders of human beings. The rationalism emerged in response to the Confucian belief that the world and all its phenomena were ultimately understandable, and even though the understandings might vary among Confucian teachers, the mere assumption of intelligibility encouraged a broad range of scholarly enquiries. The ethnocentrism was related to the fact that Confucians in China had traditionally regarded their own heritage and tradition as paramount, and so despite the fact that Confucians in Japan were at times regarded as excessively Sinophilic, Japanese Confucians often manifested patriotic qualities that were in no way secondary to those of their Chinese counterparts. Finally, the historical mindedness derived from the fact that Confucianism taught that all examples or virtue and vice necessary to demonstrate the essentially self-correcting property of historical principles might be found within the copious records of the past.

Furthermore, with the ontological and epistemological shift that Confucian philosophy provoked, one observes the transformation of a host of related fields of study of which Shintō provides perhaps the best example (see Shintō). Shintō had heretofore existed in comfortable equilibrium with Buddhism and had reconciled its teachings to Buddhism, but from the very start of the Tokugawa period, and with increasing frequency through the seventeenth century, one observes Shintō scholars seeking to refashion their theologies, dropping Buddhist epistemes and replacing them with Confucian vocabulary and assumptions. Many Confucian scholars in Japan, in particular those identified with the Zhu Xi tradition, were for their part similarly impelled to reconcile the 'truths' of Confucian philosophy with the traditional spirituality of Shintō.

In China the principal challenge within the Confucian tradition to the teachings of Zhu Xi was identified with the teachings of the scholar and celebrated general Wang Yangming (1472-1528) who identified the mind as the seat of original goodness within human beings. His philosophy emphasized what he styled the 'unity of thought and action' whereby, for example, the concept of filial piety was regarded as the genesis of filial action, and filial action was regarded as the completion of the filial impulse. Further, by emphasizing the mind's intuitive understanding of goodness, Wang Yangming and his followers de-emphasized the study either of texts to glean their truths or the physical world to plumb its principles. This philosophy, known as Yōmeigaku (literally 'Yangming studies') in Japan, had proponents there during the seventeenth century, of whom the two best known were Nakae Tōju (1608-48) and Kumazawa Banzan (1619-91), who was also celebrated as a reformer of the *samurai* class. Nonetheless, Yōmeigaku did not command a large audience in Japan until the closing years of the Tokugawa period when its idealistic orientation proved attractive to a wide range of politically activist reformers.

In Japan, however, a more significant Confucian challenge to Zhu Xi's teachings was mounted by those scholars whose teachings are collectively referred to as Ancient Studies (*kogaku*). Though the various scholars identified

with Ancient Studies differed significantly, they shared the premise that if truths representing a Way lay within the core texts of ancient Confucianism, then those truths might be better apprehended through the study of the ancient texts themselves, than by reading the exegesis on those texts written centuries later by scholars like Zhu Xi. Ancient Studies is said to have begun with the writings of Yamaga Sokō (1622-85), who described how he came to understand that his 'misunderstandings' derived from his reliance on Chinese commentaries, but that once he turned directly to the writings of Confucius and the other ancient sages, he finally understood their message. A similar approach was used by Itō Jinsai (1627-1705) and his son Itō Tōgai (1670-1736) whose school, the Kogidō (Hall of Ancient Meaning), represents the first financially successful private Confucian academy in Japan.

The most prominent of the Ancient Studies scholars, however, was Ogyū Sorai (1666-1728). Where other Confucian scholars in Japan had accepted naturalist ontologies and without exception regarded the sage as a figure of great wisdom who discovered the Way and its truths, Ogyū Sorai applied a historicist perspective to this question. Sorai regarded the Way as a 'comprehensive term' which included those laws, rituals and other practices that ancient rulers invented and applied successfully to rectify the ills of their own ages. If, Sorai argued, the ills of the present are to be similarly rectified, then one must account for the altered circumstances that prevail in another time and place for the application of this redefined 'Way' to be successful. Further, the manner in which one learns of the ancient Way is through the study of ancient texts in which the Way is encoded, but since the words that comprise those texts have themselves changed over time, Sorai insisted that his students become proficient in historical linguistics. Known as the Ken'en, the Sorai school is believed to have been the most popular of all Confucian schools in Japan through the mid-eighteenth century.

Confucian philosophy influenced the Tokugawa polity in a variety of ways. Perhaps the most conspicuous was the tendency within the society to see itself as comprised of four classes with *samurai* (*shi*) at the top, followed by agriculturalists, artisans and merchants in that order. Notwithstanding the fact that this ideal came to deviate significantly from the mercantilist realities of urban Japan, this paradigm was of Confucian origin in China and represented the attempt to identify the Japanese samurai class with the scholar-bureaucrat (*shi*) class of China. Similarly, the study of Confucian philosophy came to be of vocational advantage to *samurai* who found their martial skills no longer as useful to a society that remained essentially at peace for over two centuries. In this way, many *samurai* became sufficiently expert in Confucian philosophy to themselves become instructors of the subject within their own private academies. One benefit that Japanese society later derived from this development in the study of Confucian philosophy was a well-developed literate and patriotic managerial class available to serve the nation in its later modernization efforts.

Throughout the seventeenth and early eighteenth centuries, there was no academic tension between the study of Confucian philosophy and the study of subject material from the native tradition, but during the eighteenth century, scholars devoted to the study of things Japanese criticized Confucians in Japan for not only neglecting but even harming their own tradition by engaging in the study of a 'foreign' humanly constructed Way. Notwithstanding that this argument both applied and owed more to the followers of Ogyū Sorai than other Confucian schools, one does observe a new tension between Confucian and nativist pursuits during much of the eighteenth century, a tension which is essentially overcome through the reconvergence of these academic fields during the nineteenth century.

The scholars who most successfully represented this synthesis of Confucian and native ideals were largely identified with the Mito domain. The *daimyō* (or lords) of Mito had aggressively sponsored scholarship on Japanese history and ancient classics since the seventeenth century, devoting as much as one-third of the domain's revenue to sponsored research. As the shortcomings of the Tokugawa state became steadily more apparent, it was Confucian scholars from Mito who sought to identify within the Japanese traditions paradigms of such Confucian virtues as loyalty and filial piety, arguing that the study of Japanese history revealed those virtues to have existed in more persuasive forms in Japan than in China.

In the same manner, one can observe within Japanese Confucian philosophy itself the fracturing of Confucianism into competing schools in the eighteenth century, followed by the nineteenth-century resurgence of certain syncretic tendencies within the Confucian tradition. For example, even though the popularity of the various Ancient Studies schools of Confucianism contributed unprecedented pluralism and intellectual vitality to Confucian philosophic circles, the often contentious relations between schools represented a scholarly atmosphere

that some regarded as indecorous. Alarmed over what it regarded as an unhealthy degree of intellectual heterodoxy within Confucian circles generally, and the Hayashi school in particular, the Tokugawa government issued a directive in 1790 ordering the head of the Hayashi school to maintain greater fidelity to the teachings of Zhu Xi. This concern with philosophical orthodoxy was, in fact, just one way in which the government at this time was attempting to return to those ‘fundamentals’ which it believed to have been responsible for the success of the Tokugawa state in the seventeenth century.

Numerous Confucian philosophers of the late Tokugawa regarded what they perceived to be the ills of their age as related on the one hand to this contentious intellectual atmosphere and on the other to an excessive tendency toward activism on the part of other Confucians. These scholars sought solace in and believed that they might affect reform through eclectic approaches to the exegesis of traditional Confucian texts. Their scholarship lacks the vitality of that evidenced within Confucian circles of the first half of the Tokugawa period, and in most studies they have been overshadowed by their more activist contemporaries; yet their scholarship provides compelling evidence of the degree to which Confucian philosophy represented an intellectual sanctuary for many late-Tokugawa intellectuals.

4 Confucianism in modern Japan

Confucian philosophy contributed to the rationales used by ideologues who led the Meiji Restoration of 1868. For example, the slogan ‘Loyalty to sovereign (*chū*) and filial piety (*kō*) are one’ argued that Confucian loyalty to one’s lord, here understood to mean the emperor, was no less fundamental than those Confucian virtues that bound family relationships, of which filial piety was always regarded as the foundational virtue. Similarly, the slogan *taigi meibun* used Confucian vocabulary to argue that there was a correct fixed position for each person within an essentially immutable social order, and when each person performed their duty the social order was harmonious. Both of these slogans reflect the fact that Confucianism generally appears both during the Meiji Restoration and thereafter primarily in terms of its social and political ethics, with the emphasis on individual duty and responsibility.

During the Meiji period, when much that was traditional in nature was overwhelmed by the rapid reforms, Confucian philosophy demonstrated remarkable resiliency as its vocabulary and reasoning were used in various ways to reinforce an understanding of the Japanese polity as a quasi-family state. This is particularly evident in the Imperial Rescript on Education promulgated in 1890 under what is believed to be the influence of the Meiji emperor’s Confucian tutor, Motoda Nagazane (1818-91). The Rescript speaks of the Japanese people as subjects who are ‘united in loyalty and filial piety’, representing the ‘fundamental character’ of the Japanese empire and the foundation of its education. Other Meiji scholars such as Nishimura Shigeki (1828-1902) sought to reconcile the ‘truths’ of Confucianism with the ‘truths’ of Western culture and civilization, often under one or another variation of the binary theme ‘Eastern ethics and Western science’.

Despite the fact that Confucian philosophy has been alternatively vilified in Japan as synonymous with all that is old-fashioned and hence bad, and lionized as embodying all that is noble and of enduring value within the traditional culture, it is evident that a number of assumptions that are characteristic of Confucian ontology and ethics remain widely accepted within present-day Japan. Conspicuous among these are the assumptions that there are enduring metaphysical principles that undergird the physical world, and that these principles have a self-correcting quality. For this reason, a fundamental moral optimism is justified. These principles have an analogue within human beings in the form of an originally good disposition which is instinctively capable of moral and ethical behaviour. Human relations are similarly governed by enduring principles, where these relations tend overwhelmingly to be vertical in character with clear superiors and subordinates. One’s most important relationships are those within the family, and it is these domestic relationships that provide the training ground for the more complex relationships one enters into outside the home. Since both the material world and the realm of human society are governed by identical principles, all phenomena and things, all persons and other creatures are interrelated, making it possible to gain insight into the whole from mastery of any of its parts.

There are other ways in which Confucian philosophy is said to have contributed its legacy to the intellectual and social fabric of contemporary Japan. Most prominent among these are the assertions that the high value attached to education, the high levels of respect accorded to those in authority, the reverence shown toward one’s ancestors and forebears, and the willingness to subordinate one’s own interests to the interests of larger collectivities in

modern Japan are all attributable to the influence of Confucian philosophy. More recent scholarship, however, has tended to see these tendencies as features that either have antecedents in pre-Confucian times or that can be traced to other influences, though there is broad agreement that Confucian philosophy at the very minimum contributed to the reinforcement of these values, each of which remains prominent in modern Japan.

See also: Confucian philosophy, Chinese; Fujiwara Seika; Japanese philosophy; Itō Jinsai; Kumazawa Banzan; Neo-Confucian philosophy; Ogyū Sorai; Shintō; Shōtoku Constitution; Wang Yangming; Zhu Xi

PETER NOSCO

References and further reading

de Bary, W.T. (1959) ‘Some Common Tendencies in Neo-Confucianism’, in D.S. Nivison and A. Wright (eds) *Confucianism in Action*, Stanford, CA: Stanford University Press, 25-49.(A valuable essay for understanding Confucianism’s contributions to intellectual circles in Japan and throughout East Asia.)

Hall, J.W. (1959) ‘The Confucian Teacher in Tokugawa Japan’, in D.S. Nivison and A. Wright (eds) *Confucianism in Action*, Stanford, CA: Stanford University Press, 268-301.(Helpful for understanding the institutional context of instruction in Confucianism.)

Inoue, Y. (1992) *Confucius*, trans. R.K. Thomas, London: Peter Owen.(A best-selling historical novel in Japan, that suggests some ways in which Confucius is preceived in popular culture.)

Maruyama, M. (1974) *Studies in the Intellectual History of Tokugawa Japan*, trans. M. Hane, Princeton, NJ: Princeton University Press.(Pioneering study of Ogyū Sorai and his contributions to modern thought in Japan.)

Mercer, R. (1991) *Deep Words: Miura Baien’s System of Natural Philosophy*, Leiden: Brill.(Contains translations of three eighteenth-century works helpful to understanding Confucianism’s impact on one philosopher.)

Nosco, P. (ed.) (1986) *Confucianism and Tokugawa Culture*, Princeton, NJ: Princeton University Press.(Contains nine essays and an introduction on the general theme of responses to Confucianism.)

Spae, J.J. (1967) *Itō Jinsai: A Philosopher, Educator and Sinologist of the Tokugawa Period*, New York: Paragon Reprints.(Initially published in 1948, this remains the standard introduction to a key figure in Ancient Studies.)

Tsunoda, R. *et al.* (eds) (1964) *Sources of Japanese Tradition*, New York: Columbia University Press, 2 vols.(A valuable collection of translated exerpts from the writings of a number of major Japanese Confucian philosophers.)

Tucker, M.E. (1989) *Moral and Spritual Cultivation in Japanese Neo-Confucianism: The Life and Thought of Kaibara Ekken, 1630-1714*, Albany, NY: State University of New York Press.(A thoughtful analysis of a Japanese Confucian philosopher and the intellectual context of his times.)

Confucian philosophy, Korean

Confucianism came to Korea in the late fourth century AD. While Buddhism, which had arrived at the same time, was for centuries the central spiritual and intellectual tradition of Korea, Confucianism was viewed as largely limited to the world of government functionaries. In China during the Song dynasty (979-1279) a creative Confucian movement revitalized and reshaped the tradition, giving rise to what Western Scholars call 'neo-Confucianism'. By the end of the fourteenth century neo-Confucian learning had penetrated deeply among young scholar-officials in Korea, who used it as a lever against the deeply entrenched Buddhist establishment. In 1392, in history's only neo-Confucian revolution, a new dynasty was founded in Korea. The Chosôn dynasty (1392-1910), ruling a country smaller in scale and more centrally unified than China, was to make Korea the most (neo-) Confucian of all East Asian societies.

The scale, control, and temper of Korean society had important consequences for the development of the neo-Confucian tradition. In China the Cheng-Zhu school of neo-Confucian thought held privileged status as the orthodox standard for the all-important civil service examinations. Zhu Xi (1130-1200) was the great creative synthesizer of this school and its foremost authority, but his synthesis drew especially upon the work of the Cheng brothers. The Cheng-Zhu school was rivalled and even eclipsed in popularity later by the school of Wang Yangming (1472-1529), whose more Zen-like approach also found great favour in Japan. Korea, in contrast to both China and Japan, remained almost exclusively devoted to the Cheng-Zhu school.

This exclusive and intensive development of the Cheng-Zhu school of neo-Confucian thought is the most generally distinctive characteristic of Korean neo-Confucianism. When the thinkers of a culture devote themselves for centuries to a single complex body of learning, as was the case for example with Aristotle and medieval Europe, the result is a mode of philosophical discourse described as 'scholasticism'. Scholastic philosophy is renowned for the intricacy and closeness of its argumentation, though this may be an obstacle for the outsider for whom it is often difficult to recapture the intense and absorbing vision which inspired major controversy about seemingly minor differences. Korea, with its exclusive cultivation of Zhu Xi's complex synthesis, produced the most scholastic version of neo-Confucian thought. The writings of scholars or scholar-officials of note were commonly collected and published after their death, so the centuries of 'collected works', written in literary Chinese, are a vast resource in which the twists and turns, the problems and potentials for development in the Cheng-Zhu school are examined with unequalled thoroughness.

To understand the particular contribution this Korean scholasticism made to neo-Confucian thought one must be aware of the scope and complexity of Zhu Xi's synthesis. The intellectual culture within which the Confucian revivalists of the Song dynasty worked had for seven or eight centuries been predominantly shaped by Daoist and Buddhist influences, so the questions in their minds as they returned to the Confucian classics included dimensions neglected by more traditional Confucianism. Read with new eyes, an entirely new level of meaning was uncovered in the ancient texts: they discovered a Confucian foundation for the meditative cultivation of consciousness that had been a particular strength of the Buddhists, and to frame it and provide an account of sagehood equal to Buddhist talk of enlightenment, they found a complete metaphysical system, a Confucian version of the kind of thinking that had been elaborated mainly under Daoist auspices. Thus Zhu Xi's synthesis knit together not just disparate thinkers of the Song dynasty, but contemporary questions with texts well over a thousand years old. It incorporated Daoist metaphysics with Buddhist meditative cultivation in a new structure with Confucian moral values and social concerns at that structure's core. Implicit in this was the conflation of the distinctive world views of India, the origin of Buddhism, and China. This has important metaphysical consequences, for the central paradigm for Indian reflection on the nature of existence was consciousness, while for the Chinese it was the image of a single living physical body.

A synthesis of this scope cannot be a seamless whole, though the conceptual system with which Zhu Xi knit it together achieved a remarkable degree of verbal consistency. This, in fact, is where Korean neo-Confucian thought makes a special contribution. For it is at the seams, where differences and tensions inherent in a synthesis are conceptually masked, that the kind of problems occur that become the source of endless scholastic controversy. Korean neo-Confucianism contains two such controversies; each has occupied minds for centuries. Though the points being debated resist any ultimate solution, the disputes themselves disclose the creative tensions

at the heart of Zhu Xi's synthesis.

The first of these controversies arose in the middle of the sixteenth century and decisively shaped the intellectual agenda for the remainder of the Chosôn dynasty. The protagonists in the controversy, Yi Hwang and Yi I, are the two most famous names in Korean thought, and allegiance to each became the central dividing line of Korean neo-Confucianism. Known as the 'Four-Seven Debate,' this controversy is the most famous philosophical dispute in Korean history. On the surface it involves the question of feelings and how they arise. Some feelings, such as commiseration or shame at doing evil, seem spontaneously human and correct, while others, such as fear, anger, or pleasure, seem more questionable. Are there then two kinds of feelings that arise from different sources? The question is of great philosophical importance, because ultimately it discloses tensions at the heart of the dualistic monism/monistic dualism which is the fundamental structure of Zhu Xi's metaphysical system.

The second great controversy arose among followers of Yi I in the early eighteenth century. Neo-Confucian metaphysics views the entire universe as possessing a single nature, which is manifested differently at different levels of existence due to the differing capacities of the concrete, psychophysical component of various sorts of creatures. The Horak controversy swirled about the question of whether the fundamental or 'original' nature of things is the same or different. The fundamental nature is normative, and it would be absurd to say the norm for a cow is the same as for a human; but the fundamental nature is fundamental and normative precisely because it is considered as anterior to the limitation/distortion of the imperfect psychophysical component. How then can it be considered as differentiated into cow and human? Pulling at these seemingly verbal loose ends leads again deep into Zhu Xi's metaphysics, revealing between its Indian and Chinese (Buddhist and Daoist) motifs tensions which come to a focus more clearly here perhaps than anywhere else in neo-Confucian thought.

1 Historical development

Neo-Confucianism has dimensions similar to Marxism in Western thought: that is, although it can be considered as a philosophical system its scope includes the practice of government, social structure and relationships, and the individual's way of life. The object of the life of the mind is not abstracted as 'truth,' but is integrated with the whole of life: Confucian thought aims beyond intellectual apprehension at the actual practice of a full and integral human life in a harmonious, well-ordered society. (see Confucian philosophy, Chinese; Neo-Confucian philosophy).

Thus the neo-Confucian revolution that inaugurated the Chosôn dynasty in 1392 was also a social revolution. Energies for the first century were largely absorbed in institution building and reform of unacceptable but deeply rooted custom such as matrilocal marriages. Contrary to Confucian theory, Korean society maintained its aristocratic structure, but the Korean aristocracy (*yangban*) in other respects became highly orthodox, correct adherence to neo-Confucian norms becoming a mark of their social distinction. The aristocratic social structure also had the effect of making the Korean ruler much more a first-among-equals, much more subject, as orthodox Confucian theory prescribed, to the remonstrance and persuasion of his ministers than was the case in Ming or Qing-dynasty China.

Towards the end of the first century, remonstrance, the right of government officials to object, criticize, and offer contrary advice to superiors and particularly to the ruler, became a central and bloody issue. Zealous Korean neo-Confucians had enshrined the right to remonstrance in three government ministries, and when these acted in concert they were able to bring the government to a standstill: they could effectively thwart the king and his highest ministers, always on grounds framed in the language of lofty Confucian moral principles. The result of this systemic tension was a series of purges from the end of the fifteenth to the middle of the sixteenth centuries; the purged were the remonstrating officials, who paid with their lives or with exile when the throne and high ministers were pushed beyond the limit. These 'literati purges' left a deep imprint on the Korean neo-Confucian tradition. All Confucian learning was supposed to be for the sake of character formation, but as the established core of civil service examinations it was also the way to get ahead, and the way to secure a high reputation. In the Cheng-Zhu school a movement by earnest scholars who focused on self-cultivation and despised the ambitious motives of 'worldly Confucians' appropriated for itself the label '*daoxue*', (in Korean *tohak*), 'the learning of the true Way'. The effect of the early history of purge and martyrdom was to give special luster to the *tohak* dimension of the tradition, making it a predominant feature in the lives and thought of virtually all major Korean neo-Confucians.

The record of the development of neo-Confucian thought during this early period is somewhat sparse. The major focus initially was on institution building, while the philosophically interesting aspects of neo-Confucian thought are most closely linked with spiritual self-cultivation and its metaphysical framework. Those most likely to produce writings of interest were thus also the most likely purge victims, and what they may have written is lost. What we possess reflects the diversity and unevenness one might expect: during this period Cheng-Zhu orthodoxy had not fully settled in, nor was there a complete and systematic grasp of the integrated whole of this school's complex synthesis. The most famous name of the period, Sô Kyôngdôk (1489-1546), espoused a monism of material force (*ki*; in Chinese *qi*) similar to that of the early Chinese neo-Confucian, Zhang Zai (see Qi). Sô has a great reputation for independence and originality, but unfortunately what remains of his writings is not enough to give a very adequate picture of his ideas. Others, such as the famous martyr of the 1519 purge, Cho Kwangjo (1482-1519), reflect a situation in which those most involved with the self-cultivation side of the tradition had a very inadequate grasp of the delicate balance between theory and practice, study and meditation, that characterizes the Cheng-Zhu school. Cho's writings perished with him, but his students, we are told, often spent the day sitting in meditation rather than studying.

It is generally agreed that a full, balanced, and mature grasp of Zhu Xi's complex synthesis first came with the work of Yi Hwang (1501-70). He and a thinker of the next generation, Yi Yulgok (1536-84), are considered the foremost neo-Confucian philosophers of Korea. Particularly important was the Four-Seven Debate, initiated between Yi Hwang and Ki Taesûng (1527-72) and continued by Yi Yulgok and Sông Hon (1535-98), with Yi I taking up the position advocated by Ki Taesûng (see §§3-4). The debate had the effect of permanently dividing the neo-Confucian intellectual world of Korea into two major schools professing allegiance to Yi Hwang and Yi Yulgok, a division further complicated by complex regional and factional ties as time progressed.

Invasions first by the Japanese under Hideoshi and then by the Manchus devastated the country. The seventeenth century is remembered principally for extreme factional division and conflict by neo-Confucians in government and for the development of a rigid, orthodox tone that caused the persecution of several scholars labelled as 'heterodox' because they questioned elements of Cheng-Zhu learning.

Zhu Xi's synthesis was sufficiently vast, however, to permit major divisions and significantly different interpretations within it. The Yi Hwang-Yi Yulgok split was already evidence of this, and in the early eighteenth century the Horak controversy confirmed it. Named for the regions associated with the principle figures in the debate, the Horak controversy caught the attention and drew in major thinkers of the eighteenth century and beyond. Its participants were all followers of Yi Yulgok, and in some ways, as will be discussed below, the controversy could be considered a final working through of problems deeply embedded in Yi Yulgok's insistence on philosophical consistency in handling the Four-Seven question.

The eighteenth and nineteenth centuries are also celebrated by Korean intellectual historians for the rise of *sirhak* or 'practical learning', a broad and disparate movement among some neo-Confucian scholars towards focusing their scholarly efforts on more concrete issues of political, social and economic concern, bringing the borderless discourse of neo-Confucian thought more within the confines of actual life on the Korean peninsula. The movement historically did not develop a cohesive self-consciousness, and the label '*sirhak*' was applied to it by twentieth century historians, who saw in it the indigenous development of values and types of learning that have become the essence of modernity.

Their attention has brought to light a *sirhak* dimension in a number of significant thinkers of the period, among whom the figure of Chông Yagyong (1762-1836) overshadows all. A true polymath, in addition to extensive writing on practical affairs his works include a voluminous review of the changing interpretation of the Confucian classics over the centuries and a critique of Cheng-Zhu metaphysics from a theistic perspective grounded in the most ancient classics. His restoration of a theistic Confucianism occurred under the influence of the writings of the Roman Catholic missionary to China, Matteo Ricci. The work itself is thoroughly Confucian, however, and since it was worked out by a thinker who also had mastery of the Cheng-Zhu synthesis it is possibly the most thoroughgoing and systematic theistic philosophy produced by a Confucian thinker. In his day Chông had no significant intellectual heirs, but he may well be the most studied thinker of the Chosôn dynasty.

2 The centre of philosophical reflection

Among the East Asian neo-Confucian traditions what is most distinctive about Korea is the depth and continuity of its commitment to the orthodox Cheng-Zhu school. Thus even those like Chông Yagyong who struck out in new directions did so with a deep grounding in Cheng-Zhu thought. For the vast majority of thinkers the fundamental truth of Zhu Xi's vision was a given, and the essential task of every generation was to understand how their particular problems and insights fit into, or 'already were', the original meaning intended by Zhu Xi or the ancient authorities such as Confucius and Mencius as interpreted by Zhu Xi. Over four centuries of Korean reflection in this mode served to develop both the potentialities and the problems inherent in Zhu Xi's synthesis to an unprecedented degree. Korean insight in this respect is captured particularly in the two great controversies mentioned in §1, the Four-Seven and Horak debates.

Both debates focus on the same general area of neo-Confucian thought, *sôngnihak* (in Chinese, *xinglixue*), the study of the nature (*sông*) and principle (*i*) (see Xing; Li). It is no accident that the most complex and philosophically interesting controversies occur in *sôngnihak*, for it is the systematic nexus of every area of neo-Confucian thought: here converge what the West would call metaphysics, philosophy of man, psychology, ethics and ascetical theory or spiritual practice. The motive of Confucian, Daoist and Buddhist thought alike is the full realization of human potential, which demands self-cultivation. But self-cultivation involves an implicit or explicit metaphysical framework, which is necessary to understand in what the full realization of human potential may consist (see Self-cultivation, Chinese theories of). Buddhist meditative cultivation techniques assumed a monistic background. Huayan Buddhism used the indigenous Chinese term *li* (in Korean, *i*, meaning 'truth', 'rationale', 'principle' or 'pattern') to convey this monistic 'Buddha nature', the ultimate single reality of all phenomenal existence. Chan (Zen) Buddhism focused less on such intellectual formulation and stressed instead the direct and immediate realization of this as the reality of one's own and all being (see Buddhist philosophy, Chinese; Buddhist philosophy, Korean). Neo-Confucians, in close touch with this background, were aware of the importance of techniques of cultivating consciousness, but wanted nothing of Buddhist monism; thus they sought a more concrete metaphysics with a more secure grounding in phenomenal reality. Their remedy was to pair *i* with another traditional term, *qi* (in Korean, *ki*), the animating 'life force' of body and psyche and, philosophically, the very 'stuff' of the universe (see Qi). In its most coarse, turbid and condensed form, *ki* thus accounted for the concreteness of the divisible world the West knows as 'matter'; but in its original, pure and rarified condition it is also the stuff of psyche and mind, embracing the characteristics the West incorporates in terms such as 'spirit' and 'soul'.

In neo-Confucian metaphysics, *i* is made concrete and existential through its embodiment in *ki*. The focal point in its spectrum of meanings also shifts somewhat: as a stand-in for Buddha-nature, *i* was more on the side of 'ultimate truth', while in the neo-Confucian context its ancient connections with pattern or order are exploited in a way that makes it virtually interchangeable in many contexts with the much used *dao*, the Way, the inner pattern, the nature that structures all existence and activity. Instead of Buddhist monism, this dualism of *i* and *ki* gives us a modified, organismic monism: *i* is one, but embodied and manifested diversely through the many diverse conditions of *ki*. The many beings of the universe may thus be likened to the members of a single living body, truly different from one another and yet ultimately manifestations of a single unity beyond all name or distinctness: *i* is the single pattern or principle running throughout, while *ki* accounts for the real but not ultimate diversity.

This neo-Confucian metaphysics thus yields a very traditional East Asian organismic vision of the universe, in which every element at every level is interdependently woven together with every other in a patterned, normative whole (see Neo-Confucian philosophy). The cosmic dimensions of this tradition had found clearest manifestation in Daoism, while Confucianism had explored mainly the social ramifications; as united in neo-Confucian thought, these become respectively the framework for and the outcome of self-cultivation. The theory of self-cultivation is related to metaphysics much as epistemology is in western thought. The organismic unity of *i* solves the western problem of how distinct individuals can know 'the other', for in spite of the manifold diversity of their ways of manifesting it, all share in the same *i*. In particular, because of the high purity of the *ki* which constitutes the human psyche, *i*, the pattern of all things, is possessed in its integral fullness by the human heart-and-mind (see Xin). This means we have within ourselves the guiding pattern for appropriate responsiveness to any thing or any situation. If we are out of touch and inappropriately responsive, that is because of some adventitious factor such as an element of turbidity or coarseness in the *ki* of our psychophysical constitution which blocks or distorts our true connectedness. Neo-Confucians compared this to the way a paralyzed limb, even though connected, no longer

responds appropriately to the needs and messages of the body. The ideal sage, endowed with perfect *ki*, responds to all things with spontaneous perfection; being perfectly in touch, he need not be concerned about prolonged thought or self-discipline. Less perfectly endowed individuals, however, need both. They must follow the more laborious path, regaining their apprehension of *i* by the discipline of focused attention externally and clear calmness internally. Meditative practice, 'quiet sitting', could nurture the quiet and calm that self-possession promotes and is a means to our harmonious integration in the *i* that is our nature within and also the structure of the world without.

Sôngnihak, 'the study of the nature and *i*', thus addresses itself to the anthropocosmic core of neo-Confucian learning, where the metaphysics of the universe transmutes into a description of the working of our inner life in relation to every life situation. *I* is not only the pattern of the universe, but the ground of our psyche which prepares us to participate properly in the universe; *ki* is not only the stuff through which all things have their concrete actuality and diversity, but also the factor that distorts the potentially perfect continuity of *i* without and within. Unlike dualisms of spirit and matter which carve out a unique and special place for human beings, the challenge here is to describe all the complexity and conflict of the human heart-and-mind in continuity with the conceptual system used for all other natural phenomena. Korean neo-Confucianism pushes the limits and reveals hidden tensions in the Cheng-Zhu version of this endeavour.

3 The Four-Seven Debate

In the history of philosophy, the Cheng-Zhu metaphysics and anthropology of *i* and *ki* occupy a very special and interesting place. The patterning *i* and the concertizing, activating *ki* bring to mind the Aristotelian dualism of form and matter (see Aristotle §§11-14); but in this case, although *i* is manifested as many forms it is ultimately one form which runs through all things. Thus the universe is truly a single organic unity, rather than a plurality of formally and materially discrete beings as in the Aristotelian case. This is a monistic dualism. Recalling the Buddhist usage of *i* and the vitalistic background of the concept of *ki*, one also sees in the dualism traces of the mind-body split and the spirit-matter dualism of western thought. But here too the parallel is instructive because it breaks down: mind and body alike are comprised of *i* and *ki*, so they too are an organic continuum. As we shall see, each of the great Korean controversies pushes and tests the meaning and resources of the *i*-*ki* dualism along these lines.

The Four-Seven debate arose in 1549, and was carried out in lengthy correspondence first between Yi Hwang and Ki Taesûng, then between Yi Yulgok and Sông Hon. The central question of the debate focused on the metaphysical explanation of good and evil. In a pluralistic western context this is commonly a question of will and choice, premised on the human individual's distinctive characteristic of rationality and self-determination (see Free will; Evil, problem of). But in a culture and philosophy of organicism the question is posed in terms of response rather than choice: as seen in §2, the question of evil is more like why a limb does not respond appropriately than why individuals choose what is wrong. Hence questions of will and choice are secondary in Confucian moral discourse. Of primary concern rather are the feelings and the nature which is manifested in the way the feelings respond. In a famous classical passage, Mencius had argued that human nature is fundamentally or originally good, citing as evidence the spontaneous feelings of alarm and commiseration that arise at the sight of a dire event such as a toddler about to fall into a well (see Xing). He expanded this argument by describing four such spontaneously good feelings or tendencies: in addition to commiseration, there are the feelings of shame and dislike for evil, the tendency to yield and defer to others, and the sense of right and wrong. These innate tendencies, he said, if nurtured properly become respectively the mature characteristics of humanity, righteousness, propriety and wisdom.

The innate tendencies described by Mencius were referred to by later Confucians as the 'four beginnings', and they constitute the 'four' of the Four-Seven debate. Neo-Confucians took the four mature characteristics as descriptions of the composition of human nature, which metaphysically is equated with *i*, the normative inner pattern of all things. Concrete, actual feelings such as the Four Beginnings, are the manifestation of *i* on the level of the active, phenomenal world, the level of *i* combined with *ki*. Because *i*, the norm, is inherently good, so the feelings which manifest it are also naturally good unless imperfections in *ki*, the stuff of our psychophysical constitution, somehow interrupt and distort the way in which it is manifested.

The 'seven' of the Four-Seven debate comes from a different source, the classic *Lijing (Book of Rites)*, which

describes a list of seven feelings: desire, hate, love, fear, grief, anger, and joy. These conventional human feelings are obviously the sort of tendencies that can be either right or wrong, appropriate or inappropriate. They are the object of endless Confucian exhortations to cultivate careful attention and discernment of the movement of the feelings. The initial movement in particular is critical, for a tiny discrepancy at the beginning can become a gap of a thousand miles as things unfold.

On the level of traditional moral discourse none of this poses any special problem. The disruption of feelings and responses from their appropriate course could be accounted for by individualistic self-centredness, long identified as the root of all evil. But the special advance of neo-Confucian thought was that its new dualistic metaphysics could support a metaphysical explanation of the source of evil/disruption as well. This was of critical importance, for the description of the distorting role of *ki* with regard to the inherently perfect *i* provided a precise description of sagehood and the hinderance to sagehood, and hence an understanding of the rationale and effectiveness of approaches to self-cultivation.

In Western thought, spirit-matter dualisms, or rationality versus animal nature, set up clear value hierarchies which become the basis for explaining why it is wrong to subordinate higher values to lower. *I* and *ki*, being associated respectively with the normative and distorting roles, appear to invite a similar value discrimination. There is then an initial plausibility in suggesting that inherently good feelings such as the Four Beginnings are on the side of *i*, while more doubtful ones such as the Seven Feelings must be on the side of *ki*. This is what Yi Hwang did, almost incidentally, in the course of going over a diagrammatic treatise by another scholar. However, Ki Taesûng, a younger contemporary, reproached him. *I* and *ki* are absolutely interdependent, have no separate existence, and cannot separately be the source of different kinds of feeling; rather, he argued, there is really only one kind of feelings, for all feelings arise in precisely the same way, *i* being the formative source and *ki* the element of concrete actualization. There is, he said, actually no real (metaphysical) distinction between the Four Beginnings and Seven Feelings: the Seven Feelings are a general term for all human feelings, and the Four Beginnings are just a subset, singling out feelings when they are appropriate. Yi Hwang fought for a real distinction, even finding a passage in which Zhu Xi himself paired the Four Beginnings with *i* and the Seven Feelings with *ki*. Finally he came up with a formula that attempted to take full account of the inseparability and interdependence of *i* and *ki* and at the same time could support some feelings which are inherently more whole and human than others: in the case of the Four Beginnings, he said, *i* gives issuance and *ki* follows; in the case of the Seven Feelings, *ki* issues and *i* mounts it.

Ki Taesûng eventually yielded to Yi Hwang's formula, but after Yi Hwang's death the debate was resurrected when Yi Yulgok's friend Sông Hon read Yi Hwang's treatise in conjunction with some similarly dualistic remarks by Zhu Xi and was persuaded. Yi Yulgok, perhaps the most brilliant and systematic philosophical mind produced by the Korean neo-Confucian tradition, countered forcefully that such extreme dualism was absolutely wrong. *I* and *ki* could not vary their role and relationship as if they were two independent entities; in the metaphysics of the cosmos, he argued, it was always *ki* that was in the concrete issuing role and *i* 'mounting it' as the formative pattern that made things be and act as they did. If this was the cosmic case, it could be no different when it came to the human psyche. The dualistic expressions frequent in Zhu Xi and other authoritative sources should be understood appropriately in their moralizing context, but not transformed into a metaphysical dualism that would disrupt the delicate one-but-two, two-but-one perfect interdependence and complimentarity of *i* and *ki*.

The Korean neo-Confucian intellectual world subsequently split between followers of Yi Hwang and Yi Yulgok; infrequent thinkers attempted various sorts of syntheses, but the problem has never been resolved. On Yi Hwang's side are the dualistic phenomena common to the moral life; self-cultivation above all was the controlling factor in Yi Hwang's thought. However, Yi Yulgok has the appearance of greater philosophical consistency and rigor; he truly removed insupportable ambiguities in Zhu Xi's system and brought it to a new level of coherence.

4 The Four-Seven Debate (cont.)

What do we learn from this controversy? The two positions and the unresolvable dispute offer a unique window on the Cheng-Zhu philosophical enterprise. Yi Hwang is a faithful reflection of the motivation behind the neo-Confucian project, which is ultimately to bring about good character formation and proper order in society and the world. Yi Yulgok shows us the most careful philosophical crystalization of the deep East Asian assumption that there is no final distinction between humanity and the rest of existence: we are an integral and organic part of

the universe. In the West morality has often been treated as a uniquely human phenomenon, something that marks us off from the rest of the world. As seen through the frame of the Four-Seven debate, the challenge taken up by neo-Confucian metaphysics is whether a philosophy of organic unity can adequately account for the phenomena of human moral experience, particularly the conflicting tendencies to which we are subject and which seem less evident in the life of the rest of nature.

In its historical context the Four-Seven debate concerns the adequacy of the neo-Confucian incorporation of Mencius' explanation that human nature is good. The deep issue is, how can the normative human nature (*i*) be considered a dynamic element assisting in this process? Health and life are normative but not abstract: the body fights for health, injured trees and plants heal themselves if given time. Mencius used such examples in a more than metaphorical sense: he saw human psychic or moral life as part of the same natural system with a similar dynamism, and the neo-Confucians certainly meant to carry forward this vision. Equating the nature with *i*, the normative of each and all things, seemed to be a metaphysical version of Mencius' view. But *ki*, its counterpart, raises questions: it was introduced not only to insure an objectively real world versus the Buddhist monism of consciousness, but also to objectify the moral problem. The latter was perhaps the more important issue in Confucian eyes: Buddhism claimed ultimately to transcend the dualism of good and evil, and neo-Confucians were bitterly critical of the anti-nomian potential in such views. With its degrees of coarseness or turbidity, *ki*, by contrast, objectifies the problem of moral distortion. The neo-Confucian elaboration of a systematic philosophical explanation for a source of evil within the human constitution that does not negate the inherent goodness of human nature was a great triumph. But the solution may have succeeded too well, for the turbidity of *ki* in the psychophysical constitution seemed to put a formidable barrier between the person and the kind of spontaneously good tendencies of which Mencius spoke. In China this resulted in the radical split in which Wang Yangming simply equated the human heart-and-mind with *i* and thus opened the way to immediate intuition and spontaneous perfection in stark contrast to the kind of prolonged book learning and rigorous ascetical practice demanded in the Cheng-Zhu approach. The orthodox followers of Zhu Xi claimed this amounted to another form of Chan (Zen) Buddhism, but Wang's school was nonetheless immensely successful in China and Japan, although not in Korea.

The position worked out by Yi Hwang is an answer to this problem that remains totally within the Cheng-Zhu tradition; in comparison, the Wang Yangming solution appears an extreme and perhaps unnecessary alternative. By granting *i* the possibility of some kind of initiative to set the current in the proper direction before the disrupting influence of *ki* could take over, he was providing grounds for the spontaneous and appropriately human responsiveness Mencius had spoken of: '*I* gives issue and *ki* follows.' But the price is emphasizing the dualistic potential in Cheng-Zhu thought, for perfect complimentarity demands *ki* alone have such a concrete actualizing function. This was Yi Yulgok's objection; but in demanding that Yi Hwang's formula for the more dubious Seven Feelings, '*ki* gives issue and *i* mounts it', be the only one, he is hard pressed to really save Mencius. It would seem that if imperfect *ki* is in at the origin of all feelings in exactly the same sort of way, then the normative goodness of *i* is a mere abstraction, not something that is manifest in the actual tendencies of our inner life. The only way the perfection of *i* can spontaneously manifest in the phenomenal world would seem to be through the agency of perfectly pure *ki*.

The tightness of Yi Yulgok's consistent insistence on the absolute complementarity and interdependence took Cheng-Zhu thought to a new level of systematization. At the same time, the question remains as to whether this systematization could settle the kind of issues solved by Yi Hwang's more dualistic approach. The real test of his system came more than a century later with the Horak debate.

5 The Horak Debate

The Horak debate arose in the early decades of the eighteenth century among scholars belonging to Yi Yulgok's school of thought. Once launched, it was the source of philosophical debate down to the twentieth century, second only to the Four-Seven debate in length, notoriety and intellectual importance among Chosôn dynasty neo-Confucian controversies. Its chief protagonists were Han Wônjin (1682-1751) and Yi Kan (1677-1727). Both were students of the same prominent master, Kwôn Sangha, and both at different times held the prestigous position of Royal Lecturer, reflecting their recognition as leading scholars of their time. The debate takes its name from the geographical regions with which the major participants were associated: the 'ho' group was centred in Ch'ungch 'ôngdo and supported the position of Han Wônjin, while the 'Nak' scholars who supported Yi Kan hailed mainly

from Kyônggido.

The Horak controversy began on grounds that can be considered in some ways a continuation of the Four-Seven debate, driving Yi Yulgok's position to its final ramifications. However, in addition to pushing to the ultimate the question of a non-dualistic metaphysics of *i* and *ki* and its meaning for self-cultivation, the controversy took a new turn that brought out hidden tensions and ambiguities in the concept of *i* itself.

The Four-Seven debate pushed the issue of the ramifications of the metaphysics of *i* and *ki* for the responsive life of the heart-and-mind; the Horak debate revisited the question, but this time with a focus on how to understand the meditative practice that was an important feature of self-cultivation. After centuries of Buddhist predominance, it was natural for early neo-Confucian thinkers to assume that consciousness itself, not just moral habits and self-discipline, is a key object of cultivation. Likewise, long familiarity with meditative discipline affected the way they understood the paradigmatic yin-yang pattern of alternating activity and quiet as it applied to consciousness: quiet was taken to mean not just relative inactivity, but absolute inactivity, that is, the condition of objectless consciousness familiar to trance meditation. Thus classical references to a condition before the feelings were aroused became a warrant for a meditative discipline known as 'quiet sitting': the mind resting in objectless consciousness, in this view, is but the natural compliment to the mind actively engaged with some sort of object. Both of these conditions should have their appropriate role in self-cultivation.

Just as they framed the understanding of the feelings, *i* and *ki* served to interpret the place and meaning of meditation practice. Turbid *ki* translated easily into literal psychic turbidity and tumult; clarifying and calming one's consciousness was also a matter of clarifying and calming one's *ki*. In this conceptualization there is an easy movement from the philosophical usage of the term *ki* in metaphysics to the more popular use of the term in which *ki* is the life force and the force of the feelings. What then would be the meaning of a perfectly clear, calm, objectless state of consciousness? Obviously this would be a condition of pure, undistorted unity with *i*, one's perfect nature which is, as we have seen, the formative grounds for all activity of the heart-and-mind. From this integral contact with the inner pattern for relating to all things, one might expect to gradually develop an ease and spontaneous appropriateness of one's responses in the active life; and this would also be facilitated as the practice fostered increased calm and mental focus in all areas of life. *I* as our own nature and the nature of all things lays the grounds for our appropriate interaction in all circumstances, and this practice seems highly effective as a means of removing the distortions introduced by our turbid *ki*.

One can see here the life context of the *i*-*ki* metaphysics of the neo-Confucians. Clearly, philosophical reflection here grows out of and returns to the task of spiritual cultivation. The importance in practice of *i* as not only an abstract norm but as a formative, dynamic factor in spiritual cultivation is evident. Conceptually this was captured in the idea of the 'original nature', that is, *i* as it is in itself, undistorted by the imperfection that may be present in *ki*. This again, like the Four Beginnings, is a reference to Mencius, for neo-Confucians explain that his argument that human nature is good was in fact pointing to the 'original nature' before there ever was such terminology. Of course the negative side is equally important in understanding self-cultivation. The potentially 'bent out of shape' natures of actual human beings with all the concrete reality of imperfect *ki* was conceptualized, in contrast with the 'original nature', as the 'physical (*kijil*; in Chinese, *qizhi*) nature'.

Neo-Confucians stressed that the double terminology did not mean we have two natures: the 'original nature' terminology was necessary because if one only observed the actual humans with their imperfect 'physical nature', the original perfection of *i* would not be evident. In actuality, of course, there is only one nature: the only nature (*i*) is *i* embodied in *ki*. The point of the dual terminology is that even as embodied and hence manifested in a distorted and imperfect way, the original perfection of the nature is never lost.

But what then is the point? Since all actuality occurs only in the context of *i*-with-*ki*, is the original nature anything but a way of paying lip-service to Mencius' teaching about a good human nature? The same problem we saw in the Four-Seven debate regarding the issuance of the feelings recurs here on the level of the nature, the source of the feelings. The issue is more explosive, however, for the original nature is a central aspect of self-cultivation theory: it is the dynamic base of tendencies to be cultivated and enlarged upon in activity, and meditative quiescence, the 'condition before the mind is aroused', was commonly understood as putting one in touch with the original nature.

6 The Horak Debate (cont.)

Han Wônjin ignited the controversy by remarks he made in 1707 in a short treatise on the original nature and the physical nature. As a typical follower of Yi Yulgok, his main concern was to attack dualistic misunderstandings that sometimes stemmed from this terminology. Almost incidentally, he remarked that, regarding the original nature in the condition before the mind is aroused, since it must still inhere in *ki*, one could also take that fact into account and refer to it as the physical nature. The pure perfection of the original nature could be reached only by an ‘exclusive reference’ that simply did not take *ki* into account; in reality, of course, *ki* had to be there. In quiescence, Han observed, the imperfection that might be present in *ki* could not be observed, but that does not mean it is not there.

To speak of the condition before the mind is aroused and the physical nature in the same breath was novel and provoked a storm of protest, for it deeply challenged the conventional conceptual framework of self-cultivation theory. Han’s chief opponent was Yi Kan, a man deeply steeped in self-cultivation practice who himself had experienced the problems of a turbulent consciousness and the benefits of quiet sitting. The original nature was a reality that functioned in his spiritual practice and explained his positive experience. It could not, he argued, be reduced to a mere matter of ‘exclusive reference’.

In the course of the controversy, Han’s position proved intellectually unassailable and he never modified it, even though more scholars sided with Yi Kan than with him. The logic of a consistent, non-dualistic interpretation of Cheng-Zhu metaphysics supported him: no *i* without *ki*, and hence no *i*-originated actuality without the mediation of *ki*. It was rather Yi Kan who was forced, step-by-step, to modify his position. Finally, in order to support a living contact with the original nature, he introduced the notion of an ‘original mind’, arguing that the turbidity of *ki* applied to the body but that the heart-and-mind remained essentially pure. This is the closest neo-Confucian thought approaches the mind/body dualism common in western thought; but it is a dualism of *ki*, reflecting that the stuff of both, unlike spirit/matter, is ultimately the same.

In the Four-Seven debate, Yi Hwang gave a dualistic twist to *i*-*ki* metaphysics in order to maintain the the status of Mencius’ Four Beginnings. Here Yi Kan, in a similar defence of a Mencius-inspired vision of self-cultivation, when Yi Hwang’s kind of *i*-*ki* dualism is unthinkable, must resort to a division of *ki* itself in terms of mind and body. In the overall structure and intent of Cheng-Zhu thought, the role of *i* goes beyond merely explaining formal intelligibility or grounding an intellectual norm: it is a dynamic, active force in the process of becoming a full human being, but this dynamic side threatens to cast *i* and *ki* as competitive forces in a form of dualism out of tune with the deep structure of neo-Confucian metaphysics. The language of Zhu Xi’s synthesis is continuously ambiguous, sounding in some passages quite dualistic and then, especially in more systematic metaphysical discussion, sounding non-dualistic and emphasizing perfect complementary and interdependence between *i* and *ki*. The history of these two controversies in Korea shows clearly that the ambiguities cannot be removed without a significant and creative recasting of the basic conceptual framework.

The question of the original nature in the condition before the mind is aroused led Han Wônjin into reflection on the whole question of the nature. Cheng Yi’s famous dictum, ‘*I* is one but manifested diversely’, is one of the cornerstones of the Cheng-Zhu understanding of the nature. The comprehensive one *i* of all things is the ‘Supreme Ultimate’, as Zhou Dunyi put it in his famous *Diagram of the Supreme Ultimate*. Zhu Xi’s comment reflects Cheng’s dictum: ‘Each thing has its own nature and all things are the one Supreme Ultimate’. This metaphysics accounts for the diverse manifestation of *i* whereby each thing has its own nature in terms of the varying degrees of the coarseness or turbidity of *ki*. Just as this accounts for the different degrees of intellectual and moral qualities of humans, so on a broader scale it accounts for entirely different species.

In self-cultivation discourse, it is of course the perfectly good original nature which is normative, and this led to a stress on *i*’s transcendence of the concrete limitation of *ki*: the pure and vital original nature remains unsullied, however bent out of shape it may be as manifested in a given individual. A criminal with a turbid psyche thus strives to become a better human, not a better criminal. But how does this apply to other species, similarly differentiated by the imperfection of *ki*? Is the normative, original nature of a cow the same as the original nature of a human? Han observed that according to the classics each creature should follow its nature, which implies that it is the nature *as differentiated* by *ki* that is normative: a good cow acts like a cow, not like a human.

Such reflections led Han Wônjin in 1708 to propose a novel, tripartite way of viewing the nature: (1) the nature as pure *i*, with no reference to *ki*, and hence no differentiation (the nature in this sense is unitary and ineffable); (2)

the nature as 'based on *ki*', that is, as differentiated into species but undistorted and hence normative; and (3) the concrete physical nature according to which every individual, even of the same species, is different from every other. The 'original nature', he suggested, is a differentiated norm, and this demands the distinctive second category, the nature as based on *ki*.

Han's opponents, led by Yi Kan, mocked the novel language of a nature 'based on ki' which somehow was not yet the real, concrete physical nature of an individual being. Yi held firmly to the common understanding of a unitary *i* as the 'single origin', common to and within all things, and *i* as concretely manifested in diverse individuals. With *ki* comes limitation and distortion, and hence separation from the norm; hence the original nature must be identified with the 'single origin,' and its diversity as norm must somehow be likewise contained in that.

This dispute involved the most fundamental elements of Cheng-Zhu thought, the equation of *i* with the nature of things and the use of *ki* to account for both individual (especially moral) differences and the differentiation of various kinds of creatures. It thus was pursued with intensity, but over the years no consensus could be reached. Part of the reason for this lies in a deep ambiguity in Cheng Yi's foundational proposition, '*I* is one and manifested diversely'. Is the diverse manifestation totally the consequence of *ki*? Such was the understanding of Han Wônjin, who demanded the explicit reference to *ki* to account for any kind of diversified norm. Or is *ki* the means of implementing a diversity that has its more profound source in *i* itself? Such was the understanding of Yi Kan, who treated the 'Single Origin' as a self-diversifying, normative *i*. One can find support for both of these interpretations of Cheng's dictum in the writings of Zhu Xi, but the ambiguity runs through most discourse treating of the diversification of *i* through *ki*.

The fundamental ambiguity explored in the Horak debate points ultimately to the seam by which the organismic metaphysics typified by Chinese Daoism is stitched together with the consciousness-oriented participation metaphysics (all creatures are the manifestation of One Mind) which is part of Buddhism's Indian heritage (see Daoist philosophy). *I* as self-diversifying follows the organismic paradigm of the single pattern which accounts for both the unity and the diversity of a living body; *i* as unitary but manifested in varying degrees of fullness through the varying perfection of *ki* represents a kind of monism quite different from this one-body image, bringing to mind more the One Mind which never becomes plural in spite of its manifold manifestation. Mind and body have been subtly united here in a way unparalleled by any other of the world's great traditions. Bodies are differentiated by kinds and consciousness permits of degrees. neo-Confucians utilize *i* in both these ways, explicating through it the distinctness and commonality of all things, the transcendence of a normative nature, and the universal responsiveness of human consciousness which is the highest degree of participation in *i*. As the Horak debate makes clear, this feat involves a varied conceptualization of *i*'s relationship with *ki* as the price of such an accomplishment.

See also: Han Wônjin; Li; Neo-Confucian philosophy; Qi; Self-cultivation, Chinese theories of; Yi Hwang; Yi Kan; Yi Yulgok

MICHAEL C. KALTON

References and further reading

de Bary, W.T. and Haboush, J.K. (eds) (1985) *The Rise of Neo-Confucianism in Korea*, New York: Columbia University Press. (Useful collection of articles on major figures and features of the Korean neo-Confucian tradition.)

Deuchler, M. (1992) *The Confucian Transformation of Korea*, Cambridge, MA: Harvard University Press. (Excellent analysis of process by which Korea became East Asia's 'most Confucian' society.)

Haboush, J.K. (1988) *A Heritage of Kings*, New York: Columbia University Press.(A study of the functioning of neo-Confucian ideology and symbols in the reign of one of Korea's most outstanding kings.)

Kalton, M.C. (ed. and trans.) (1988) *To Become a Sage*, New York: Columbia University Press.(An annotated translation and commentary of the most famous work of Yi Hwang (T'oegye), one of the two leading neo-Confucian thinkers of Korea.)

Kalton, M.C., Kim Oaksook, C., Park Sung-Bae, Ro Youngchan, Tu Weiming and Yamashita, S. (trans) (1993) *The Four-Seven Debate*, Albany, NY: State University of New York Press.(An annotated translation of the most famous controversy in Korean neo-Confucian thought.)

Lee, P.H., Baker, D., Ch'oe Yongho, Kang, H.H. and Kim Han-kyo (eds) (1993) *Sources of Korean Tradition*, New York: Columbia University Press.(An extensive collection of translations from original sources, including

neo-Confucian works.)
Ro Youngchan (1988) *The Korean Neo-Confucianism of Yi Yulgok*, Albany, NY: State University of New York Press.(A study of the thought of one of the two leading neo-Confucian thinkers of Korea.)

Confucius (551-479 BC)

Confucius is arguably the most influential philosopher in human history - 'is' because, taking Chinese philosophy on its own terms, he is still very much alive. Recognized as China's first teacher both chronologically and in importance, his ideas have been the rich soil in which the Chinese cultural tradition has grown and flourished. In fact, whatever we might mean by 'Chineseness' today, some two and a half millennia after his death, is inseparable from the example of personal character that Confucius provided for posterity. Nor was his influence restricted to China; all of the Sinitic cultures - especially Korea, Japan and Vietnam - have evolved around ways of living and thinking derived from the wisdom of the Sage.

A couple of centuries before Plato founded his Academy to train statesmen for the political life of Athens, Confucius had established a school with the explicit purpose of educating the next generation for political leadership. As his curriculum, Confucius is credited with having over his lifetime edited what were to become the Chinese Classics, a collection of poetry, music, historical documents and annals that chronicled the events at the Lu court, along with an extensive commentary on the Yijing (Book of Changes). These classics provided a shared cultural vocabulary for his students, and became the standard curriculum for the Chinese literati in subsequent centuries.

Confucius began the practice of independent philosophers travelling from state to state in an effort to persuade political leaders that their particular teachings were a practicable formula for social and political success. In the decades that followed the death of Confucius, intellectuals of every stripe - Confucians, Legalists, Mohists, Yin-yang theorists, Militarists - would take to the road, attracted by court academies which sprung up to host them. Within these seats of learning, the viability of their various strategies for political and social unity would be hotly debated.

1 Life, works and influence

Although Confucius enjoyed great popularity as a teacher and many of his students found their way into political office, his enduring frustration was that personally he achieved only marginal influence in the practical politics of the day. He was a *philosophe* rather than a theoretical philosopher; he wanted desperately to hold sway over intellectual and social trends, and to improve the quality of life that was dependent upon them. Although there were occasions on which important political figures sought his advice and services, over his years in the state of Lu, he held only minor offices at court.

Early on, however, and certainly by the time of his death, Confucius had risen in reputation to become a model of erudition, attracting attention from all segments of society. As time passed and the stock in Confucius rose, the historical records began to 'recall' details about his official career that had been supposedly lost. Over time, his later disciples altered the wording of his biographical record in his favour, effectively promoting him from minor official to several of the highest positions in the land.

Nor does the story end there. By the time of the Han dynasty (206 BC-AD 220), Confucius was celebrated as the 'uncrowned king' of the state of Lu, and by the fourth century AD, any prefecture wanting to define itself as a political entity was required by imperial decree to erect a temple to Confucius. Gods in China are local cultural heroes who are remembered by history as having contributed meaning and value to the tradition, and of these revered ancestors, that god called Confucius has been remembered best.

Confucius was certainly a flesh and blood historical figure, as real as Jesus or George Washington. But the received Confucius was and still is a 'living corporate person', in the sense that generation after generation of descendants have written commentaries on the legacy of Confucius in an effort to make his teachings appropriate for their own times and places. 'Confucianism' is a lineage of scholars who have continued to elaborate upon the canonical texts passed on after the life of Confucius came to an end, extending the way of living that Confucius had begun.

Although the ascent of Confucius to exalted status began early in the tradition with the continuation of his work by his many disciples, it was not until Confucianism was established as the state ideology during the Han dynasty (206 BC-AD 220) that his school of thought became an unchallenged orthodoxy. By developing his insights around

the most basic and enduring aspects of the human experience - family, friendship, education, community and so on - Confucius had guaranteed their continuing relevance. One characteristic of Confucianism that began with Confucius himself, and made it so resilient in the Chinese tradition, is its porousness and adaptability. Confucius said of himself that he only transmitted traditional culture, he did not create it; his contribution was simply to take ownership of the tradition, and adapt the wisdom of the past to his own present historical moment.

Just as Confucius reinvented the culture of the Zhou and earlier dynasties for his own era, Han dynasty Confucianism drew into itself many of the ideas owned by competing schools in the earlier centuries, and in so doing, fortified itself against their challenge. This pattern - absorbing competing ideas and adapting them to the specific conditions of the time - sustained Confucianism across the centuries as the official doctrine of the Chinese empire until the fall of the Qing dynasty in 1911. In fact, an argument can be made that just as the composite of Buddhism and Confucianism produced neo-Confucianism, the combination of Marxism and Confucianism in this century has created a kind of neo-neo-Confucianism.

As recently as the Cultural Revolution (1966-76), Jiang Qing, Mao Zedong's wife, and her cohorts mounted an Anti-Confucius campaign that swept the country. The great irony of this Anti-Confucius campaign was that during this period one could not acquire a copy of the *Analects*, not because they had been banned or suppressed but because they were sold out; the entire country was put to work reading the teachings of Confucius in order to criticize them.

2 Teachings

There are many sources for the teachings of Confucius that have been passed down to us today. The most authoritative among them are the *Lunyu Analects*. (*Lunyu* literally means 'discourses'; a better translation is 'analects', coming from the Greek *analekta*, which has the root meaning of 'leftovers after a feast'.) It is probably the case that the first fifteen books of these literary 'leftovers' were assembled and edited by a congress of Confucius' disciples shortly after his death. It would seem the disciples concluded that a very special person had walked among them, and that his way - what he said and did - should be preserved for future generations. Much of this portion of the text is devoted to remembering Confucius, a personal narrative of what he had to say, to whom he said it and how he said it. The middle three chapters are like snapshots of his life-habits - Confucius never sat down without first straightening his mat; he never slept in the posture of a corpse; he never sang on a day that he attended a funeral; he drank freely, but never to the point of being confused of mind.

The last five books of the *Analects* appear to have been compiled some time later, after the most prominent disciples of Confucius had launched their own teaching careers and had taken it upon themselves to elaborate on the philosophy of their late Master. Confucius is less prominent in these chapters, yet he is referenced with more honorific terms, while the now mature disciples are themselves often quoted.

In addition to the *Analects*, the other two most important resources for the life and teachings of Confucius are the *Zuozhuan* (the Zuo commentary on the *Spring and Autumn Annals*), and the *Mencius*. The Zuo commentary is a narrative history which purports to interpret the chronicle of the court history of the state of Lu up until the death of Confucius. The *Mencius* is a text named after a disciple who elaborated the doctrines of Confucius some century and a half after his death (see Mencius), and became one of the Four Books in the Song dynasty, which were from then onwards the core of the Confucian classics (see Chinese Classics).

One thing is clear about the *Analects* and these supplementary texts: they do not purport to lay out a formula by which everyone should live. Rather, they provide an account of one man: how he cultivated his humanity, and how he lived a satisfying life, much to the admiration of those around him. The way (*dao*) of Confucius is nothing more or less than the way in which he as a particular person chose to live his life. The power and lasting value of his ideas lie in the fact that they are intuitively persuasive, and readily adaptable.

Confucius begins from the insight that the life of every human being is played out within the context of their particular family, for better or for worse. For Confucius and generations of Chinese to come, it is one's family and the complex of relationships that constitute it, rather than the solitary individual, that is the basic unit of humanity. In fact, for Confucius, there is no individual - no 'self' or 'soul' - that remains once the layers of social relations are peeled away. One is one's roles and relationships. The goal of living, then, is to achieve harmony and enjoyment for oneself and others through acting appropriately in those roles and relationships that constitute one.

Given that we all live within the web of family relationships, it is entirely natural that we should project this institution out onto the community, the polity and the cosmos as an organizing metaphor. The Confucian community is an extension of aunts and uncles, sisters and cousins; the teacher is ‘teacher-father’ and one’s senior classmates are ‘elder-brother students’; ‘the ruler is father and mother to the people, and is the son of "Heaven"’. ‘Heaven’ itself is a faceless amalgam of ancestors rather than some transcendent Creator deity (see Tian). As Confucius says, ‘The exemplary person works hard at the root, for where the root has taken firm hold, the way will grow.’ What then is the root? He continues: ‘Treating your family members properly - this is the root of becoming a person.’

For Confucius, the way to live is not dictated for us by some power beyond; it is something we all must participate in constructing. On one occasion, Confucius said ‘It is not the way that broadens people, but people who broaden the way.’ The way is our passage through life, the road we take. Our forbearers mapped out their way and built their roads, and in so doing, have provided a bearing for succeeding generations. They have given us the culture and institutions that structure our lives and give them value and meaning. But each new generation must be roadbuilders too, and continue the efforts that have gone before.

Confucius saw living as an art rather than a science. There are no blueprints, no formulae, no replications. He once said, ‘The exemplary person seeks harmony, not sameness.’ In a family, each member has his or her unique and particular role. Harmony is simply getting the most out of these differences. Similarly, Confucius saw harmony in community emerging out of the uninhibited contributions of its diverse people. Communal enjoyment is like Chinese cooking - getting the most out of your ingredients.

Confucius was extraordinarily fond of good music, because making music is conducent to harmony, bringing different voices into productive relationships. Music is tolerant in allowing each voice and instrument to have its own place, its own integrity, while at the same time, requiring that each ingredient find a complementary role in which it can add the most to the ensemble. Music is always unique in that each performance has a life of its own.

What Confucius calls *ren* - literally, ‘becoming a person’ - is the recognition that personal character is the consequence of cultivating one’s relationships with others (see Confucian philosophy, Chinese §5). There is nothing more defining of humanity for Confucius than the practical consideration of one human being for another. Importantly, *ren* does not precede practical employment; it is not a principle or standard that has some existence beyond the day to day lives of the people who realize it in their relationships. Rather, *ren* is fostered in the deepening of relationships that occur as one takes on the responsibility and obligations of communal living, and comes fully to life. *Ren* is shared human flourishing. It is the achievement of the quality of relationships which, like the lines in calligraphy or landscape painting, collaborate to maximum aesthetic effect.

Wisdom for Confucius is relevant knowledge: not knowing ‘what’ in some abstract and theoretical sense, but knowing ‘how’ to map one’s way through life and get the most happiness out of it. Happiness for oneself and for others is isomorphic; it is mutually entailing. In discussing knowledge, Confucius says that being fond of something is better than just knowing it, and finding enjoyment in it is better than just being fond of it. Confucius associates *ren* with mountains; it is spiritual and enduring, a constant geographical marker from which we can all take our bearings. Wisdom is like water, pure, flowing and nurturing. The exemplary person is both *ren* and wise, both mountain and water.

A good way to think about ‘the way’ is the notion of passage. On one occasion, Confucius observed while standing on the bank of a river, ‘Isn’t its passing just like this, never ceasing day or night!’ Life is at its very best a pleasant journey, where the inherited body of cultural institutions and the pattern of roles and relationships that locate us within community - what Confucius calls ‘propriety’ (*li*) - are a code of formal behaviours for stabilizing and disciplining our ever-changing circumstances. ‘Propriety’ covers everything from table manners to the three years of mourning on the loss of one’s parent, from the institution of parenthood to the appropriate posture for expressing commiseration. It is a social syntax that brings the particular members of community into meaningful relationships. Propriety is a discourse, which like language, enables people to communicate, and to locate themselves appropriately, one with the other.

What distinguishes ‘propriety’ from rules and regulations is that these cultural norms must be personalized, and are open to refinement. Only I can be father to my sons; only I can be this son to my mother; only I can sacrifice to my

ancestors. And if I act properly, performing my roles and cultivating my relationships so that they are rich and fruitful, other people in community will see me as a model of appropriate conduct, and will defer. It is precisely this power of example that Confucius called ‘excellence’ (*de*). Excellence is the propensity of people to behave a certain way when provided with an inspiring model.

The other side to what Confucius calls ‘propriety’ is the cultivation of a sense of shame. Shame is community-based. It is an awareness of and a concern for how others perceive one’s conduct. Persons with a sense of shame genuinely care about what other people think of them. Self-sufficient individuals, on the other hand, need not be concerned about the judgments of others. Such individuals can thus be capable of acting shamelessly, using any means at all to take what they want when they want it.

3 Disciples

Confucius was tolerant of difference. In fact, on six separate occasions in the *Analects*, he is asked what he means by *ren*, an idea that is at the heart of his teachings, and six times he gives different answers. For Confucius, instructing disciples in *ren* requires that the message be tailored to the conditions of the person asking the question. We have said that, for Confucius, persons are no more than the sum of their specific familial and communal roles and relationships, and that *ren* emerges out of the quality that they are able to achieve in cultivating them. It stands to reason, then, that to know Confucius, we do best to familiarize ourselves with his community of disciples. The Teacher can best be known by his students.

Some of these disciples come to life in a careful reading of the *Analects*. For example, although Confucius was reluctant to use the term *ren* to describe anyone, he did use it of his favourite disciple, Yan Hui, also called Yan Yuan. Living on a bowl of rice and a ladle of water, Yan Hui’s eagerness to learn and his sincerity endeared him to the Master; but he was also possessed of an incomparable character, and was so intelligent that Confucius said of him, ‘When he is told one thing he understands ten’. Although Yan Hui was some thirty years younger than Confucius, it was only him among his many disciples that Confucius saw as his equal. It is no surprise, then, that Confucius was totally devastated by the death of Yan Hui at the young age of thirty-one.

Zilu was another of Confucius’ best-known and favourite disciples. He was a person of courage and action who was sometimes upbraided by Confucius for being too bold and impetuous. When he asked Confucius if courage was indeed the highest virtue, Confucius tried to rein him in by replying that a person who has courage without a sense of appropriateness will be a trouble-maker, and a lesser person will be a thief. Confucius’ feelings for Zilu were mixed. On the one hand, he was constantly critical of Zilu’s rashness and immodesty; and impatient with his seeming indifference to book learning. On the other hand, Confucius appreciated Zilu’s unswerving loyalty and directness; he never delayed on fulfilling his commitments. Being nearer Confucius in age, Zilu with his military temper was not one to take criticism without giving it back; on several occasions, especially in the apocryphal literature, Zilu challenges Confucius’ judgment in associating with political figures of questionable character and immodest reputation. However, Confucius’ enormous affection for the irrepressible Zilu comes through the text.

Confucius was also critical of his disciples. Zai Yu, also called Ziwo, was devoted to Confucius, yet on numerous occasions Confucius criticised him roundly for a lack of character. It was as a metaphorical reference to attempting to educate Ziwo that Confucius said, ’You cannot carve rotten wood, nor can you whitewash a wall made from dry manure’.

Zigong excelled as a statesman and as a merchant, and was perhaps second only to Yan Hui in Confucius’ affections. Confucius was respectful of Zigong’s abilities, and in particular his intellect, but less impressed with Zigong’s use of this intellect to amass personal wealth. Coming from a wealthy, educated home, Zigong was well-spoken, and as such, Confucius’ most persistent criticism of him is that his deeds could not keep pace with his words. Even so, much of the flattering profile of Confucius collected in the *Analects* is cast in the words of the eloquent Zigong.

Zengzi is best remembered as a proponent of filial piety - devotion and service to one’s parents. A natural extension of this affection for one’s family is friendship. Zengzi rose to prominence after the death of Confucius as one of his leading advocates.

These and many other disciples came from around the central states of China, gravitating to the state of Lu to study

with Confucius. In spite of the sometimes severe opinions which Confucius expressed freely about them - and he admonishes almost every one of them - they were devoted to the Master, and responded to him with reverence. There is no greater proof of this enduring respect for Confucius than the fact that they had a hand in recording Confucius’ criticisms of themselves, and then went on to found branch schools based on these same criticisms to perpetuate his teachings.

4 The *Analects*: texts and commentaries

The work known as the *Analects* is mainly a collection of sayings and conversations of Confucius. In the time of Emperor Wu (140-87 BC) of the Han there were three versions of the work, the *Lulun*, the *Qilun* and the *Gulun*. In the first century BC, Zhang Yu taught a version known as *Lunyu according to Marquis Zhang*, which incorporated readings from both the *Lulun* and the *Qilun*. Zheng Xuan (AD 127-200) further adopted readings from the *Gulun*. The text that has come down to us is that of He Yan (AD 190-249) in his *Lunyu jijie (Collected Commentaries on the Lunyu)*.

The extant *Analects* is in twenty books. Of the three early versions, only the *Lulun* had twenty books; the *Qilun* and the *Gulun* both had twenty-two books, though the extra books were not identical. According to the *Xinlun* of Huan Tan (24 BC-AD 56) the order of the chapters in the *Gulun* was different and there were more than four hundred variant readings. Lu Deming (AD 556-637) of the Tang dynasty also remarked that in the *Qilun*, besides the two extra books, ‘the chapters and verses were considerably more numerous than in the *Lulun*’. Some of the variant readings were recorded by scholars before the three versions were lost and these have been collected by textual critics over the centuries, but these consist mainly of variant forms of graphs. Only a handful affect the interpretation of the text. For instance, the text of VII.17 in the traditional text reads, ‘Grant me a few more years so that I may continue to study the *Yijing* at the age of fifty, and I shall be free from major errors.’ According to Lu Deming, the *Lulun* reads *yi* (grammatical particle) in place of *yi* (meaning the *Yijing (Book of Changes)*. This can only be rendered ‘Grant me a few more years so that I may continue to learn at the age of fifty, and I shall, perhaps, be free from major errors.’ Thus the variant reading has a bearing on the substantive point whether Confucius was a keen student of the *Yijing*. Of the two eclectic versions, the *Lunyu according to Marquis Zhang* was based on the *Lulun* while incorporating readings from the *Qilun*. As for Zheng Xuan’s version, it has been the common view that this was likewise based on the *Lulun* but incorporating readings from the *Gulun*. However, this may not be the case, as there is some evidence that Zheng also adopted some readings of the *Qilun*. On their versions of the *Lunyu*, both Zhang and Zheng wrote commentaries. Zhang’s commentary was in an independent work known as *Lu Anchang hou shuo (Marquis Anchang’s Exegesis of the Lu [lun])*. Unfortunately this was lost at a very early age and we do not possess even a single quotation from it.

Zheng’s commentary was attached to his text of the *Analects*, and although the work is lost, numerous quotations, particularly through He Yan’s *Collected Commentaries*, have come down to us. In the present century a number of fragments came to light in Dunhuang and Turfan, the most notable of which is a partial manuscript copy done by a twelve-year-old schoolboy in 710 and discovered in 1969. In all, we now possess over half of Zheng’s total commentary, and this has spurred on the study of this commentary. In Bajiaolang in Ding Xian, Hebei province, a copy of the *Analects* on bamboo slips was unearthed in 1971 in a tomb of the late Western Han, consisting of over half of the whole text. The exciting part of the discovery is that the text antedates the version of Marquis Zhang, and may represent the school of the *Qilun*. Unfortunately, the publication of this text so far has not revealed any major differences.

Of the numerous commentaries on the *Analects*, only a few landmarks can be mentioned. The *Lunyu zhushu (Subcommentary on the Lunyu)* by Xing Bing (932-1010) and *Lunyu jizhu (Collected Commentaries on the Lunyu)* by the great neo-Confucian thinker Zhu Xi (1130-1200) were authoritative works for the educated reader. In the Qing dynasty, as a reaction against the neo-Confucian approach, there were new commentaries on the classics with greater philological emphasis. On the *Analects* in particular, we have Liu Baonan’s *Lunyu zhengyi (The True Interpretation of the Lunyu)*.

There is finally the question of the composition of the *Analects*. First (as noted in §2), the work as we have it was not put into the present form once and for all. The later books were likely to have been added on at a later editing. Second, within a single book, some material must have been taken from existing collections of sayings of Confucius *en bloc* and some chapters added subsequently. Finally, sayings of disciples must have been

incorporated by their own disciples to enhance their standing in the Confucian tradition. This is particularly true of Book I, in which are found sayings by younger disciples such as Youzi and Zengzi who played an important role in the formation of the Confucian tradition.

D.C. LAU
ROGER T. AMES

List of works

Confucius (551-479 BC) *Lunyu (Analects)*, trans. D.C. Lau, *Confucius: The Analects (Lun-yu)*, Hong Kong: Chinese University Press, 1992. (Although the dates given are for Confucius' own life, the *Analects* were in fact compiled over a long period after his death, as noted in the text of the entry. Lau provides a revised bilingual translation of the *Analects*, complete with a philosophical introduction, appendices on the history of the text, events in the life of Confucius and a characterization of his various disciples. The standard English translation.)

References and further reading

Fingarette, H.A. (1972) *Confucius: The Secular as Sacred*, New York: Harper & Row. (A pioneering re-evaluation of the philosophy of Confucius, and its relevance to contemporary philosophy.)

Hall, D.L. and Ames, R.T. (1987) *Thinking Through Confucius*, Albany, NY: State University of New York Press. (A reconstruction of the philosophical insights of Confucius, comparing the presuppositions of his way of thinking with presuppositions that underlie the Western traditions of philosophy.)

Connectionism

Connectionism is an approach to computation that uses connectionist networks. A connectionist network is composed of information-processing units (or nodes); typically, many units process information simultaneously, giving rise to massively 'parallel distributed processing'. Units process information only locally: they respond only to their specific input lines by changing or retaining their activation values; and they causally influence the activation values of their output units by transmitting amounts of activation along connections of various weights or strengths. As a result of such local unit processing, networks themselves can behave in rule-like ways to compute functions.

The study of connectionist computation has grown rapidly since the early 1980s and now extends to every area of cognitive science. For the philosophy of psychology, the primary interest of connectionist computation is its potential role in the computational theory of cognition - the theory that cognitive processes are computational. Networks are employed in the study of perception, memory, learning and categorization; and it has been claimed that connectionism has the potential to yield an alternative to the classical view of cognition as rule-governed symbol manipulation.

Since cognitive capacities are realized in the central nervous system, perhaps the most attractive feature of the connectionist approach to cognitive modelling is the neural-like aspects of network architectures. The members of a certain family of connectionist networks, artificial neural networks, have proved to be a valuable tool for investigating information processing within the nervous system. In artificial neural networks, units are neuron-like; connections, axon-like; and the weights of connections function in ways analogous to synapses.

Another attraction is that connectionist networks, with their units sensitive to varying strengths of multiple inputs, carry out in natural ways 'multiple soft constraint satisfaction' tasks - assessing the extent to which a number of non-mandatory, weighted constraints are satisfied. Tasks of this sort occur in motor-control, early vision, memory, and in categorization and pattern recognition. Moreover, typical networks can re-programme themselves by adjusting the weights of the connections among their units, thereby engaging in a kind of 'learning'; and they can do so even on the basis of the sorts of noisy and/or incomplete data people typically encounter.

The potential role of connectionist architectures in the computational theory of cognition is, however, an open question. One possibility is that cognitive architecture is a 'mixed architecture', with classical and connectionist modules. But the most widely discussed view is that cognitive architecture is thoroughly connectionist. The leading challenge to this view is that an adequate cognitive theory must explain high-level cognitive phenomena such as the systematicity of thought (someone who can think 'The dog chases the cat' can also think 'The cat chases the dog'), its productivity (our ability to think a potential infinity of thoughts) and its inferential coherence (people can infer 'p' from 'p and q'). It has been argued that a connectionist architecture could explain such phenomena only if it implements a classical, language-like symbolic architecture. Whether this is so, however, and, indeed, even whether there are *such phenomena to be explained, are currently subjects of intense debate.*

1 What is a connectionist network?

Units and their activation. A unit in a connectionist network can be an input unit, an output unit or a 'hidden' unit. Input units can directly receive signals from sources external to the network; output units can directly send signals outside the network. Input and output units are sometimes called 'visible' units because of their direct interaction with external factors. The 'hidden' units of a network are those units (if any) that send or receive signals from outside the network only by means of sending or receiving signals from other units, as in the following 'Hamming net':

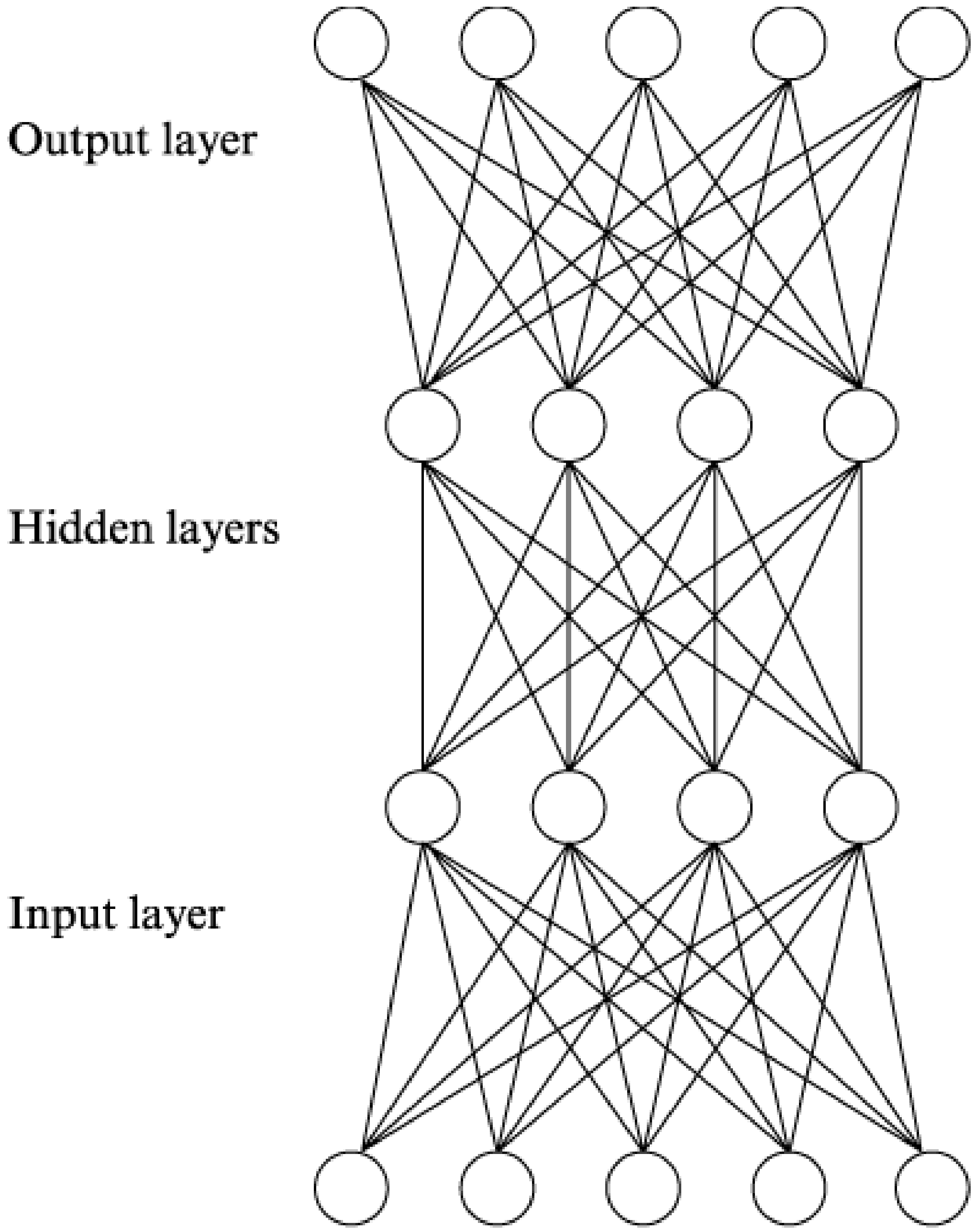

Simple connectionist network (from Dinsmore 1992: 9)

A network can have one or more layers of hidden units, or no hidden units at all.

At any time, a unit is in an activation state. The kinds of activation states units occupy can vary from one network architecture to another, from being merely active or inactive, to being in discrete states of three or more levels of activation, or to having continuous degrees of activation, either bounded or unbounded. The activation state of a unit is indexed by a real number, typically on the interval [0,1], called the 'activation value' of the state.

Each unit computes an output function that maps its current state of activation to an output signal. The signal is not an encoded message such as a data structure, but a degree of activation indexed by a real number. While an output function can be linear, typically it is some threshold function. If a unit has a negative bias, it may send 0 (or the null signal) unless its activation value exceeds a certain threshold; if it has a positive bias, the unit may send a

certain output value unless its activation value falls below a specific threshold. Networks in which units compute non-linear output functions exceed in computational power networks in which units compute only linear ones.

Patterns of connectivity. The activation state of a unit U_i causally influences the degree of activation of a unit U_j by means of the activation output U_i sends to U_j. The direction of such causal influence is the direction of activation spread or flow. There is a connection from a unit U_i to a unit U_j when, but only when, the output of U_i can exert an influence on the activation state of U_j via a direct path within the network - one along which there is no intermediate unit. Connections are thus the direct causal routes along which units send signals to their output units. In some networks, two units can mutually influence each other, and thus a unit can be related to another both as input and output. In 'auto-associative' networks, each unit is connected to every other unit, including itself; in other networks, there are various restrictions on which units are connected.

When units can mutually influence each other, information processing is interactive. In 'recurrent' networks, there are connection patterns that contain loops. That is, a unit is either related to itself as an input unit or there is a series of unit connections from the unit back to itself; thus, the output of a unit at one time can causally influence its activation state at another via activation flow. In 'feed-forward' networks, information processing is non-interactive: activation flows in a single direction from input units to output units via whatever layers, if any, of hidden units the network contains. The Hamming net, for instance, is a feed-forward network with three layers of units, one of which is a layer of hidden units (see diagram above).

The causal influence a sending unit exerts on a receiving unit depends on two intrinsic properties of their connection: its weight (or strength); and whether it acts to increase (excite) or decrease (inhibit) the level of activation of the receiving unit. The notation 'w_{ij}' is used to stand for a real number which indexes the connection from unit U_i to unit U_j by its weight and kind; the number is positive/negative when the connection is excitatory/inhibitory, respectively. The weight of the connection is the absolute value of w_{ij}.

In many networks, the extent and kind of causal influence a unit U_i exerts on a unit U_j is indexed by the product of w_{ij} and the activation value of U_i. The total network activity input to U_j is the sum of all such causal influences: it is the sum of the product of each input activation value and the real number indexing the weight and kind of connection it bears to U_j. (If there are inputs from sources external to the network, they must likewise be indexed by real numbers and included in the summing.) But in some networks, the net input function to a unit - called the 'propagation rule' - is more complex. In 'competitive' networks, units form pools: the units in a pool are all mutually inhibitory - that is, they bear inhibitory connections to each other - while units outside the pool are excitatory to one or more units in the pool. And still more complex patterns of connectivity are possible.

Units respond to the signals they receive either by changing their level of activation or by remaining at the same level. Which response occurs is a function of three factors: the total activity input to the unit, the unit's current activation state, and the unit's bias (if any). The activation function for a unit maps the unit's net input (a sum or weighted sum) and current activation state to an activation state of the unit. In Hopfield networks, units have a sigmoidal (S-shaped) response to net input: their activation levels only increase by a given amount given an increase in net input; after that, they level off and increase no further. But units can also have Gaussian (bell-shaped) activation functions and other sorts of non-linear ones as well.

Local and global effects. Units process information by computing an activation function and an output function. By means of such local interactions among units, global consequences ensue. The behaviour of a network as a whole is a consequence of the pattern of connectivity exhibited by its units at a time and the global activation state of the network at that time. The latter is indexed by a vector (ordered list) of real numbers, the activation value of each unit in the network at that time being indexed by one and only one element of the vector.

Network information processing is characterized as the evolution through time of global patterns of activation. The set of all possible global activation states of a network is its activation space, whose dimensionality is exactly equal to the number of units in the network. The vectors that index possible global states of the network pick out points in the activation space, and the temporal evolution of the network is characterized as a trajectory through that space.

2 Computation, representation and learning

A network itself can behave in a rule-like way to compute functions from one vector to another. The argument of the function is a vector indexing a pattern of activation over input units, and the value of the function for that argument is a vector indexing a pattern of activation over output units. In response to any input activation pattern representing an argument of a particular function, a network with a specific pattern of connectivity can be disposed to produce the output activation pattern representing the value of the function for that argument. The network computes the function when this disposition is activated.

Representations within a network can be of one of two kinds: local or distributed. A representation is *local* if it is (or is realized by) an *individual* unit in a certain activation state. A representation is *distributed* if it is (or is realized by) a pattern of activation over a *group* of two or more units. A network computes functions from one sort of representation to another by computing the corresponding vector-to-vector function in the manner described in the preceding paragraph.

The pattern of connectivity exhibited by the units in a network functions like an 'implicit' look-up table program. A look-up table program looks up (unique) answers to questions posed to it. Patterns of activation over input units function like questions posed to the network (for example, 'To what category does this belong?'), and patterns of activation over output units function like answers to the questions ('To category C.'). Unlike typical classical look-up mechanisms, networks do not *explicitly* store the answers as data structures; rather, the answers are, so to speak, *implicitly* contained in the pattern of connectivity. The patterns of activation over output units can be construed as data structures. But such patterns are produced in response to the network's inputs as a result of the network's pattern of connectivity, rather than retrieved from some storage location in the network. A network's memory, it is said, is in its pattern of connectivity.

A network's pattern of connectivity is not architecturally fixed, for the weights of connections are not. Connectionist networks can programme themselves, via weight change, to execute a given look-up table program. Thus, the look-up table need not be pre-programmed into the network; the network can develop it in response to training. It is taught to recognize a pattern, for instance, by being trained on examples of the pattern: presented with examples of the target pattern as inputs, its outputs are assessed by some evaluation procedure and the weights of connections are then adjusted (see the discussion of learning rules below); the network passes through a number of cycles of this procedure until it acquires the disposition to respond to each input in the right way (or to some desired degree of accuracy). Networks can be trained to associate patterns and to extract features from data sets in response to statistical regularities in them.

During training, weight change occurs in accordance with learning rules. One of the earliest is *Hebb's rule*: when two units are activated at the same time, the strength of the connection between them will be increased (see Hebb 1949). By conforming to this rule a network can learn associations that depend on correlations between activations of units in the network. The generalized *delta rule*, typically called 'back-propagation', is currently the most widely used learning rule in connectionist research. For a given input activation pattern, the actual output pattern is compared with a target output pattern. The difference between the two is then propagated back into whatever connections were used to get the actual output. If the match is good, then the connections among units which contributed to the output are strengthened; if the match is poor, then the relevant connections are reduced in strength. Back-propagation is worthy of special note because it dispelled concerns about the possible computational limitations of network architectures raised by Minsky and Papert's critique of perceptrons - simple two-layered networks (Minsky and Papert 1969).

Back-propagation is a kind of supervised learning rule since it results in weight changes based on an externally generated report on the network's performance - an error signal. Networks can engage in unsupervised learning by monitoring their performance by means of internal feedback via direct or indirect loops.

3 Networks and multiple soft constraint satisfaction

Many kinds of networks are capable of finding optimal (or near-optimal) solutions to problems whose solution involves assessing multiple soft constraint satisfaction - problems Minsky and Papert (1969) called 'best-match' problems. There has been considerable success in analysing a wide range of cognitive problems as such problems. Content-addressable memory - recalling something or someone on the basis of a partial characterization (for example, 'a game played on a court with a ball that players try to drop through a hoop') - is analysed as a

pattern-completion task involving multiple soft constraint satisfaction. And categorization using conceptual schemata has been analysed similarly. This seems to many to capture Wittgenstein's 'family resemblance' account of concepts, whereby objects satisfy a concept by virtue of sharing any of a variety of resemblances, as well as Rosch's related account of concepts in terms of the resemblance of an object to a 'prototype' (see Concepts §7). And recently Prince and Smolensky's optimality theory (1993) outlines how parsers using multiple soft constraint satisfaction might serve for the parsing of natural language.

The basic network approach to such problems is due to Hinton (1977). Each unit represents a hypothesis (note that such representations are local and unstructured), and each connection puts constraints among hypotheses. If, for example, hypothesis H is evidence for hypothesis H', there is a positive connection from the unit representing H to the unit representing H'; and the connection is negative if H is evidence against H'. The tighter the confirming or disconfirming evidential relation, the greater the strength of the connection. If one hypothesis implies another, the connection from the first to the second will be such that the output unit must be on if the input unit is on; if two hypotheses are incompatible, they may be mutually connected in such a way that they cannot both be on. Since input units can also receive external input, the corresponding hypotheses can receive direct confirming or disconfirming evidence from the environment. The fact that different hypotheses may have different a priori probabilities is handled by biasing the relevant units. The overall goodness of fit of a particular hypothesis to the evidence is measured by the sum of the individual constraints on the activation value of the unit representing the hypothesis. The activation values of units range between a minimum and a maximum. The maximum value means that the hypothesis should be accepted, the minimum that it should be rejected, and intermediate values correspond to intermediate states of certainty. The constraint satisfaction problem is thus reduced to one of maximizing this overall goodness of constraint fit.

Implementing such a procedure by a network has proved fruitful in many areas of cognitive science. For example, one of the basic problems in computational vision is to characterize the computational procedure by which the visual system arrives at representations of the scene before the subject's eyes from the subject's retinal images (see Vision). A leading research programme characterizes the computational procedure as a Bayesian decision procedure implemented by a network (see Probability, interpretations of).

One problem that arises for procedures for solving best-match problems is avoiding local maxima of goodness of fit. The Boltzmann machine, developed by J. Hinton and Terence Sejnowski, employs the standard method for avoiding them. Exploiting a computational analogue of the metallurgical process of annealing - whereby metals are heated to a little below their melting point and then cooled very slowly so that all their atoms have time to settle into a single orientation - random noise (heat) is introduced into network activity. The noise functions to jar the network out of local landscape valleys (which correspond to low goodness peaks) so it can explore other parts of the energy landscape to find the global energy minimum (the highest goodness peak). When the network reaches a stable state, it has settled or relaxed into a solution. Given sufficient time, the Boltzmann machine can solve any computational problem for which weights exist.

4 Connectionism and artificial intelligence

Work in connectionist computation has had perhaps its greatest impact in the field of artificial intelligence (AI), which attempts to develop intelligent machines (see Artificial intelligence). AI research into connectionist networks has yielded fruit in many areas of business and industry. (The connectionist machines are, however, virtually always simulated on standard commercial computers, rather than actually manufactured.)

Networks are especially proficient at solving so-called connected problems - ones that do not divide into independently solvable subproblems. Consider, for example, 'the travelling salesman' problem: a salesman must visit each of a number of cities, beginning and ending his journey at the same city, and visiting all the others only once. The problem is to find the shortest route he can take. (If there are, for instance, twenty cities, then there are over 6.5 billion combinations of routes.) Since which city the salesman visits depends on which cities he has already visited, the problem does not divide into independently solvable subproblems. Such problems are decidable, and can thus be solved by classical methods. However, as the connected parameters of the problem increase, there is an exponential increase in the computational resources required to solve them by classical means. Connectionist networks are able to solve such problems much more efficiently. Hopfield's network is able to find a nearly optimal solution to the travelling salesman problem in a small number of training cycles.

Networks are also proficient at pattern-recognition tasks and at learning to perform such tasks. One network that has received a great deal of attention in the popular press is Sejnowski and Rosenberg's NETtalk (1987), which learns to associate letters with phonemes. The network drives a synthesizer that produces pronunciations of the phonemes. After sufficient network training, the synthesizer sounds like a robotic voice literally reading English texts presented to it.

There have been some comparative studies between networks employing back-propagation and classical learning systems employing various members of the family of 'top-down inductive decision trees' (TDIDT), which are kinds of 'production systems' (Quinlan 1987). The studies compared decision trees and networks with respect to speed and predictive accuracy of learning categorization tasks, both from accurate data and from noisy and/or incomplete data (for example, Fisher and McKusick 1989; Mooney *et al.* 1989; Shavlik *et al.* 1991). One data set used in a comparative study is Sejnowski's and Rosenberg's 'NETtalk Full' (Shavlik *et al.* 1991). The TDIDT member that figured in the study was ID3. There was no statistically significant difference in the predictive accuracy of the network using back-propagation and ID3. ID3 was 64.8% accurate; the network was 63.0% accurate. Moreover, both ID3 and back-propagation degraded at roughly the same rate as features were removed from the data set, but ID3 was five times faster than back-propagation. On the NETtalk-A data set, ID3 was 63.1% accurate and the network was 66.4% accurate; and ID3 was 100 times faster. (In studies for some data sets, TDIDTs are as much as 1,000 times faster than networks using back-propagation; but that is unsurprising since back-propagation is notoriously slow.) Speed aside, other such comparative studies indicate that networks using back-propagation are roughly equivalent to various TDIDT members in terms of predictive accuracy and graceful degradation in the face of noisy and/or incomplete data.

New learning algorithms for networks are being explored, as are new TDIDT algorithms. Moreover, in the field of machine learning, there is currently much research into machines that integrate networks and TDIDTs into a single machine architecture. This research proceeds without regard to either neural or psychological plausibility; its aim is not to model how cognizers actually learn, but rather to build better learning machines.

5 Connectionism and neuroscience

The nervous system appears to be a system of neural networks. We are, however, largely ignorant of its workings. By and large, it is unknown what functions any type of neuron computes, and how information generally is encoded in the brain. But enough is known to know that, while connectionist networks were inspired by neural networks, extant connectionist networks are not neurally realistic: no type of connectionist network yet proposed is instantiated by any actual neural network. On the basis of anatomical and immunocytochemical criteria of type individuation, it is estimated that the number of types of cortical neurons is somewhere between fifty and 500 (Sereno 1988). Different types of neurons appear to perform specialized functions. These functions are not well understood, but there is no evidence that any type computes activation functions or output functions computed by units in any extant network model. Further, there is no neural analogue of back-propagation. And while parts of dendrites appear to be able to act as independent information-processing units, this has no analogue in extant connectionist networks (Churchland and Sejnowski 1992).

None the less, work in computational neuroscience with the members of a certain family of connectionist networks, namely artificial neural networks, is playing an important role in increasing our knowledge of the nervous system. Computational neuroscience strives for more and more realistic networks with the aim of someday modelling actual neural-network information-processing (Churchland and Sejnowski 1992). Important pioneer work has already been done. For example, there has been considerable development of artificial neural network models of low-level vision phenomena such as stereopsis (binocular vision) and colour constancy, and of motor-control phenomena such as the vestibular-ocular reflex (that enables eyes to focus on a fixed point as the head moves). Indeed, artificial neural network theory is presently the only approach to understanding information processing in neural networks.

6 Connectionism and the philosophy of psychology

The bulk of the philosophical literature on connectionism has focused on one of two issues: first, the prospects of cognitive architecture proving to be thoroughly connectionist; and, second, the consequences for folk psychology should it prove to be so.

There is nothing approaching a consensus regarding the second issue. Some philosophers argue that connectionism could vindicate folk psychology (for example, Horgan and Tienson 1995); others, that it would entail the elimination of many folk-psychological notions, for example, belief (Ramsey *et al.* 1990). One obstacle to adjudicating such disputes is that it is unclear how wrong folk psychology must be for eliminativism to follow; another is that it is controversial exactly what folk psychology entails (Stich and Warfield 1995). For example, it has been claimed that folk psychology is committed to the view that having beliefs requires having sentence-like mental representations. It is true that according to the *classical computational* theory of belief, a belief that *P* consists in a disposition to compute in certain ways with a mental sentence that means that *P*. However, this account of belief is a scientific hypothesis, not something implied by folk psychology. Similarly, it has been claimed that folk psychology is committed to the view that when an agent acts, some, but not all, of the agent's beliefs and desires play a causal role in producing the action. However, this too is an open question: some philosophers view folk psychology as an interpretive device, rather than as a system of causal hypotheses (see Folk psychology; Eliminativism).

Perhaps the most serious obstacle to adjudicating these disputes is that there is nothing approaching a consensus about what connectionist *cognitive* architecture would be like. For example, some connectionists maintain that a connectionist cognitive architecture would contain sentence-like representations as a phenomenon emerging from the network (Smolensky *et al.* 1992). Indeed, there is no consensus about what kinds of network activity would count as realizing mental activity or what sorts of network states would count as realizing beliefs, desires, intentions or the like. While there has been much interesting connectionist work in cognitive modelling, a connectionist alternative to the classical conception of cognition as rule-governed symbol manipulation has yet to be articulated. The value of discussions of whether folk psychology would be vindicated is that they bring this issue of an alternative conception into bold relief (see Cognitive architecture; Language of thought).

7 Crux of the classical/connectionist debate

Discussions of the prospects of cognitive architecture proving to be thoroughly connectionist have taken place in the context of discussions of whether a connectionist theory of cognition could replace the classical theory.

It is universally acknowledged that cognitive architecture is not a Turing machine architecture (in which a scanner reads and prints symbols on an infinitely long tape - see Turing machines), nor a von Neumann architecture (in which a central processor controls what occurs in different addresses - see Neumann, J. von). The classical theory of cognition is not committed to either of these classical architectures. Neither is it committed to any extant classical programming languages such as LISP (or 'list processor'), nor to rules being explicitly represented, to processing being sequential ('serial' as opposed to 'parallel'), to data structures being explicitly stored in memory locations, nor to there being an executive overseer. Classicism is committed only to our cognitive architecture including (1) a symbol system with a compositional syntax (a 'language of thought') and (2) types of algorithmic and heuristic causal processes in which complex symbols participate in virtue of their logico-syntactic structure (see Language of thought; Mind, computational theories of). The crux of the classicism/connectionist architecture debate is whether an adequate cognitive architecture could fail to include both (1) and (2).

According to proponents of replacement connectionism, connectionism can yield an adequate, alternative theory of cognitive architecture to the classical theory. Jerry Fodor and Zenon Pylyshyn (1988) have raised the most widely discussed challenge to this position. They point out that classicism can explain systematic relationships among capacities to have thoughts involving the exercise of the same concepts in the same intentional mode (for example, belief, desire, intention and so on): for example, that someone able to think that the dog is chasing the cat would normally be able to think that the cat is chasing the dog, and that someone able to want the cat to be on the mat would normally be able to want the mat to be on the cat. They point out, further, that classicism can explain the 'productivity' of thought - that is, the fact that people can have a potential infinitude of thoughts; and that it can explain the inferential coherence of thought processes, for example, the truth-preserving relations between premises and conclusions in a valid argument. They challenge replacement connectionists to show how a connectionist architecture could explain such phenomena without implementing a classical one.

One response to the challenge is to concede that it cannot be met, and to embrace the implementational option, or propose some mixed architecture. Another is to reject the challenge on the grounds that there are no such cognitive

phenomena to be explained. A third is to attempt to meet the challenge: to show how connectionism could explain such phenomena without implementing a classical architecture. While each response has defenders, discussion has focused mainly on the third. It is generally acknowledged that networks with only local representations are inadequate to explain the cognitive phenomena in question. However, some connectionists (most notably Smolensky 1991, 1995) argue that it is possible for a network architecture with distributed representations to yield explanations of systematicity, productivity and inferential coherence without implementing a classical architecture. Whether this is so is a subject of intense debate (see, for example, Fodor and McLaughlin 1990; McLaughlin 1993a, 1993b, 1997). The potential role of connectionism in the computational theory of cognition remains unresolved.

See also: Cognitive architecture; Modularity of mind

BRIAN P. McLAUGHLIN

References and further reading

Aizawa, K. (1992) 'Connectionism and Artificial Intelligence: History and Philosophical Interpretation', *Journal of Experimental Artificial Intelligence* 4: 295-313.(Excellent historical discussion of connectionism.)

Bechtel, W. and Abrahamsen, A. (1991) *Connectionism and the Mind*, Oxford: Blackwell.(Presently the most comprehensive philosophical discussion of the connectionist approach to cognition.)

Broadbent, D. (1985) 'A Question of Levels: Comments on McClelland and Rumelhart', *Journal of Experimental Psychology* 114: 89-192.(Raises the issue of whether connectionism and classicism may be concerned with phenomena at different levels of organization or description.)

Churchland, P.M. (1995) *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*, Cambridge, MA: MIT Press/Bradford Books.(Includes an examination of the role of connectionism in understanding the mind/brain.)

Churchland, P.S. and Sejnowski, T.J. (1992) *The Computational Brain*, Cambridge, MA: MIT Press/Bradford Books.(An excellent primer on the study of neural computation using artificial neural networks; and a key source for §5.)

Clark, A. (1989) *Microcognition*, Cambridge, MA: MIT Press/Bradford Books.(Examines the potential role of connectionism in the understanding of cognition.)

Clark, A. (1993) *Associative Engines: Connectionism, Concepts, and Representational Change*, Cambridge, MA: MIT Press/Bradford Books.(Addresses various problems faced by connectionist models of cognitive phenomena.)

Davies, M. (1991) 'Concepts, Connectionism, and the Language of Thought', in W. Ramsey, S. Stich and D. Rumelhart (eds) *Philosophy and Connectionist Theory*, Hillsdale, NJ: Erlbaum.(Argues that it may be a priori true that thought capacities are systematic; and discusses the prospects for eliminativism concerning beliefs and desires should cognitive architecture prove to be thoroughly connectionist.)

Dietterich, T.G., Hild, H. and Bakirir, G. (1990) 'A Comparative Study of ID3 and Backpropagation for English Text-to-Speech Mapping', *Proceedings of the Seventh International Conference on Machine Learning* 24-31. (Comparative study of a classical learning system and a network using back-propagation on NETtalk data sets.)

Dinsmore, J. (1992) 'Thunder in the Gap', in J. Dinsmore (ed.) *The Symbolic and Connectionist Paradigms: Closing the Gap*, Hillsdale, NJ: Lawrence Erlbaum, 1-23.(The source of the diagram in §1.)

Feldman, J.A. and Ballard, D.H. (1982) 'Connectionist Models and Their Properties', *Cognitive Science* 6: 205-54.(In addition to popularizing the term 'connectionism', this essay proposed the controversial 'one hundred step constraint' on algorithms based on the fact that a neuron can fire only about 100 times in the time it takes a human cognizer to perform various recognition tasks, such as recognizing a familiar word.)

Fisher, D. and McKusick, K.B. (1989) 'An Empirical Comparison of ID3 and Backpropagation', *Proceedings of the Sixth Workshop on Machine Learning* 169-73.(Comparative study of classical learning systems and a network using back-propagation.)

Fodor, J. and McLaughlin, B.P. (1990) 'Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work', *Cognition* 35: 183-204.(Responds to an attempt by Smolensky (1991) to show how a connectionist architecture could explain systematicity without implementing a classical architecture.)

Fodor, J. and Pylyshyn, Z. (1988) 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition* 28: 3-71.(Argues that if connectionism is to provide an adequate, alternative theory of cognition to classicism, a connectionist architecture must be able to explain systematicity, productivity and inferential coherence without implementing a classical (language of thought) architecture.)

Gelder, T. van (1991a) 'Classical Questions, Radical Answers: Connectionism and the Structure of Mental

Representations', in T. Horgan and J. Tienson (eds) *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer, 355-1.(Argues that cognitive psychology is undergoing a Kuhnian paradigm shift from the classical to a connectionist paradigm, and that the new connectionist paradigm need not answer questions framed in terms of the old classical paradigm.)

Gelder, T. van (1991b) 'What is the "D" in "PDP"? A Survey of the Concept of Distribution', in W. Ramsey, S. Stich and D. Rumelhart (eds) *Philosophy and Connectionist Theory*, Hillsdale, NJ: Erlbaum, 33-60.(Attempts to explicate the notion of a distributed representation.)

Hebb, D. (1949) *The Organization of Behavior*, New York: Wiley & Sons.(An important early work on neural networks.)

Hinton, G.E. (1977) 'Relaxation and its Role in Vision', unpublished doctoral dissertation, University of Edinburgh.(Presents the network approach to best-match problems described in §3.)

Horgan, T. and Tienson, J. (eds) (1991) *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer. (Anthology of papers examining the potential role of connectionism in understanding cognition.)

Horgan, T. and Tienson, J. (1995) 'Connectionism and the Commitments of Folk Psychology', *Philosophical Perspectives* 9: 127-52.(Argues that connectionism may well be able to vindicate folk psychology.)

Horgan, T. and Tienson, J. (1996) *Connectionism and the Philosophy of Psychology*, Cambridge, MA: MIT Press/Bradford Books.(Defends the view that the mind is a dynamic system realized in a system of neural networks.)

Marinov, M.S. (1993) 'On the Spuriousness of the Symbolic/Subsymbolic Distinction', *Minds and Machines* 3: 253-70.(Examines comparative studies of networks using back-propagation and classical top-down inductive decision trees; and argues that classicism, like connectionism, can avail itself of 'subsymbols', that is, symbols for microfeatures.)

McClelland, J.L. and Rumelhart, D.E. (1988) *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*, Cambridge, MA: MIT Press/Bradford Books.(An excellent beginning work book in connectionist modelling, which contains a discussion of best-match problems and Hinton's method for solving them that served as a source for §3. Readers who seek a working knowledge of such modelling should begin here.)

McClelland, J.L., Rumelhart, D.E. and the PDP Research Group (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 2, *Psychological and Biological Models*, Cambridge, MA: MIT Press/Bradford Books.(Volume 2 of the two 'bibles' of connectionism. Obligatory reading for students of connectionism.)

McLaughlin, B.P. (1993a) 'The Connectionism/Classicism Battle to Win Souls', *Philosophical Studies* 70: 45-72. (Examines various aspects of the connectionism/classicism debate about the nature of cognitive architecture.)

McLaughlin, B.P. (1993b) 'Systematicity, Conceptual Truth, and Evolution', in C. Hookway and D. Peterson (eds) *Royal Institute of Philosophy*, supplement 34: 217-34.(Argues that the Fodor-Pylyshyn challenge to connectionism would remain even if it is a conceptual truth that thought capacities are systematic; argues that capacities for perceptual representations may be systematic; and argues that the systematicity of thought cannot be explained solely by appeal to natural selection.)

McLaughlin, B.P. (1997) 'Classical Constituents in Smolensky's ICS Architecture', in M.L. dalla Chiara, K. Doets, D. Mundici and J. van Benthem (eds) *The Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence, August 1995*, vol. 2, *Structures and Norms in Science*, Dordrecht: Kluwer, 331-43.(A response to Smolensky (1995).)

McLaughlin, B.P. and Warfield, T.A. (1994) 'The Allure of Connectionism Reexamined', *Synthese* 101: 365-400.(Argues against the widespread misconception that connectionist architectures are better suited to modelling the learning of pattern recognition abilities than are classical architectures.)

Minsky, M.A. and Papert, S. (1969) *Perceptrons*, Cambridge, MA: MIT Press.(A famous critique of perceptrons - simple two-layered networks - which, they argue, cannot compute exclusive 'or'.)

Mooney, R., Shavlik, J., Towell, G. and Gove, A. (1989) 'An Experimental Comparison of Symbol and Connectionist Learning Algorithms', *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* 775-70.(Comparative study of classical learning systems and a network using back-propagation.)

Pinker, S. and Prince, A. (1988) 'On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition', *Cognition* 28: 73-193.(A critique of Rumelhart and McClelland's connectionist model of verbs in the past tense; also contains a discussion of general issues concerning the potential role of connectionism in cognitive science.)

Prince, S. and Smolensky P. (1993) *Optimality Theory: Constraint and Interaction in Generative Grammar*, technical report TR-2, New Brunswick, NJ: Rutgers Center for Cognitive Science, Rutgers University.

(Development of an account of grammar along connectionist lines.)

Quinlan, J.R. (1987) 'Generating Production Rules from Decision Trees', *Proceedings of the Eleventh International Conference on Artificial Intelligence* 304-7.(Argues that top-down inductive decision trees are production systems.)

Raggett, J. and Bains, W. (1992) *Artificial Intelligence from A to Z*, London: Chapman and Hall.(A valuable source of simple explanations for many of the technical terms used in this entry.)

Ramsey, W., Stich, S. and Garon, J. (1990) 'Connectionism, Eliminativism and the Future of Folk Psychology', *Philosophical Perspectives* 4: 499-533.(Argues that connectionist architectures cannot capture the propositional modularity of beliefs and desires: the fact that some but not all beliefs and desires will play a causal role in producing intentional actions; and they argue further that given this, if cognitive architecture is thoroughly connectionist, eliminativism concerning belief and desire will be warranted.)

Ramsey, W., Stich, S. and Rumelhart, D. (eds) (1991) *Philosophy and Connectionist Theory*, Hillsdale, NJ: Erlbaum.(Anthology containing original essays on philosophical issues concerning connectionism.)

Rey, G. (1991) 'An Explanatory Budget for Connectionism and Eliminativism', in T. Horgan and J. Tienson (eds) *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer, 219-40.(Spells out a variety features of propositional attitudes that are explained by classicism and challenges connectionists to say how they can capture them.)

Rumelhart, D.E., McClelland, J.L. and the PDP Research Group (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, *Foundations*, Cambridge, MA: MIT Press/Bradford Books.(Volume 1 of the two 'bibles' of connectionism; a key source for §1. Begin your reading here.)

Sejnowski, J. and Rosenberg, C. (1987) 'Parallel Networks That Learn to Pronounce English Text', *Complex Systems* 1: 145-68.(Presents NETtalk.)

Sereno, M. (1988) 'The Visual System', in I.W. Seelen, U.M. Leinhos and G. Shaw (eds) *Organization of Neural Networks*, Wienheim: VCH, 167-84.(Examination of aspects of the visual system in terms of neural networks.)

Sethi, I.K., and Jain, A.K. (eds) (1991) *Artificial Neural Networks and Statistical Pattern Recognition*, Amsterdam: Elsevier/North Holland.(Essays on neural networks, top-down inductive decision trees and pattern recognition.)

Shavlik, J., Mooney, R. and Towell, G. (1991) 'Symbolic and Neural Learning Algorithms: An Experimental Comparison', *Machine Learning* 6: 111-43.(Includes some of the comparative studies between TDIDTs and networks using back-propagation, described in §4.)

Smolensky, P. (1988) 'On the Proper Treatment of Connectionism', *Behavioral and Brain Sciences* 11: 1-74. (Explicates a connectionist view of cognition.)

Smolensky, P. (1991) 'Connectionism, Constituency and the Language of Thought', in B. Loewer and G. Rey (eds) *Meaning in Mind: Fodor and His Critics*, Oxford: Blackwell, 223-90.(A response to the Fodor-Pylyshyn challenge to explain how a connectionist architecture can explain systematicity without implementing a classical architecture.)

Smolensky, P. (1995) 'Reply: Constituent Structure and Explanation in an Integrated Connectionist/Symbolic Architecture', in C. Macdonald and G. Macdonald (eds) *Connectionism: Debates on Psychological Explanation*, Oxford: Blackwell, 223-90.(Further develops the response in Smolensky (1991) to the Fodor-Pylyshyn challenge; and responds to Fodor and McLaughlin (1990).)

Smolensky, P., Legendre, G. and Miyata, Y. (1992) 'Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition', technical report CU-CS-600-92, Boulder, CO: Department of Computer Science, University of Colorado.(Develops an architecture that integrates features of connectionist and symbolic architectures; an architecture Smolensky (1995) claims can be used to explain systematicity and productivity without implementing a classical architecture.)

Stich, S. and Warfield, T. (1995) 'Reply to Clark and Smolensky: Do Connectionist Minds Have Beliefs?', in C. Macdonald and G. Macdonald (eds) *Connectionism: Debates on Psychological Explanation*, Oxford: Blackwell.(A reply to papers by Clark and Smolensky arguing that, *contra* Ramsey, Stich and Garon (1990), connectionist architectures can accommodate beliefs.)

Conscience

To have a conscience involves being conscious of the moral quality of what one has done, or intends to do. There are several elements under the idea of conscience. First, conscience can signify those very moral convictions persons cleave to most firmly and judge themselves by. Second, the notion may cover the faculty by which we come to know moral truths (assuming there to be such) and apply them to ourselves. Third, conscience can be said to concern the examination by a person of the morality of their desires, actions and so on. Finally, conscience can involve guilt: one can suffer from a 'bad conscience'. In the Christian tradition, conscience can be viewed as 'the voice of God within' each of us. Several of these aspects of conscience are expressed in Milton's lines from Paradise Lost, when God says: 'And I will place within them as a guide/My umpire Conscience' (III: 194-5).

There are many elements comprehended under the idea of conscience, and it is sensible to consider these separately since they do not always occur together. First, someone's conscience may be considered to comprise those fundamental moral convictions by keeping to which they retain a sense of their moral integrity and decency as people. In this sense something is 'a matter of conscience', or raises 'questions of conscience', if it touches on such central personal principles. According to this signification, different people can have markedly different consciences, but, it would be argued, we should still respect each person's conscience since to force them to violate the demands of their conscience is to force them to give up their sense of their own integrity. This line of argument may be resisted, however. We are told that some Nazis saw carrying out the extermination programme as a matter of conscience. To force them not to do this does not seem to involve a moral problem. In reply, it can be said that only consciences which are 'enlightened' require respect. The question whether conscience can be enlightened, or fallible and perverted, leads on to a second strand in thinking about conscience.

According to this strand, a person's conscience comprises their capacity to come to acquire moral standards for their own conduct; or, more specifically, their capacity to come to know moral truths (see Moral knowledge §1). This sense, strictly more basic than *conscientia* itself, is referred to as *synderesis* or *synteresis* by Aquinas (*Summa theologiae*), by which he means the power to grasp fundamental moral principles, a power supposedly common to all persons and one which, if functioning appropriately, results in our all knowing the same basic principles. However these very basic principles (such as 'Do good' and 'Eschew evil') are too general to help us know how to act in particular circumstances. We require also a capacity to derive more concrete principles which will give us moral guidance, and a capacity to apply these appropriately to our own circumstances. These two secondary functions are the province of conscience strictly understood, according to Aquinas. Errors in conscience can arise in deriving these more specific rules of conduct or in their application even if *synderesis* is held to be infallible. Questions can arise, analogous to those referred to earlier, about whether it is better to follow one's conscience even though it may prove to be fallible or to violate one's conscience which, if it is erroneous, may then mean one ends up doing the right thing. Talk of a 'perverted' conscience may mean that a person's ultimate convictions are judged to be perverse, as in the first strand identified; or that their capacity to know good from evil, in general or in the particular case, has been distorted or corrupted.

Building on the above, we may note a third emphasis in the idea of conscience, to do with the care, intensity and frequency with which someone examines the moral credentials of their desires, feelings, actions and omissions. Someone may have an 'overworked' or 'oversensitive' conscience, by making too much a matter of moral self-scrutiny or by being too scrupulous about any and every moral doubt which may arise (if one can be overscrupulous about such things). On the other hand, a person's conscience may be 'fast asleep' and need 'awakening' or 'quickening', as in the famous painting by Holman Hunt, 'The Awakening Conscience' (1852). They do not lack a conscience, but rather the will or disposition to use it. Joseph Butler (1726), in a sensitive and important discussion of the moral significance of conscience, emphasized that conscience sometimes 'without being consulted, without being advised with, magisterially exerts itself' (Sermon II) (see Butler, J. §§2-4). I take this to mean that sometimes without, or against, our will or deliberate intention we find ourselves judging our own intentions and deeds critically and in condemnation. This thought leads on to the last aspect of conscience to be considered.

A person may first become aware that they have done something they feel to be wrong or wicked through

experiencing feelings of uneasiness, guilt, or a vague sense of oppression. It may take some thought to discover what deed these feelings attach to, but they can be regarded as central manifestations of having a ‘bad conscience’ about it. Feelings of remorse, shame, dismay, torment and guilt, all forms of self-punishment following from self-condemnation, are major elements in the functioning of conscience. A clear, easy or happy conscience does not usually bring self-congratulatory feelings with it (that would be more like self-righteousness), merely the absence of the pains of a troubled conscience. Popular morality likes to believe that no-one can escape from the toils of a guilty conscience in the end if they commit some terrible deed, however hard they try.

The self-punitive strand in conscience has particularly attracted the attention of psychoanalysts including Freud (1930). Some people can be crippled in their capacity for active life by the savagery and relentlessness of self-punitive guilt incident to, say, feeling aggressive or sexual impulses. Freud held that children ‘internalize’ powerful parental figures as part of their development and, in doing so, can subject themselves to very severe judgments felt to be emanating from these figures. The ‘super-ego’ thus conceived can confront and harangue the child’s ego, and inhibit its expression (see Freud, S. §8; Psychoanalysis, post-Freudian §2). To gain relief from such savage self-censure is not, of course, to become amoral or conscienceless; it is simply to adopt a less punitive attitude to moral endeavour.

It has been argued that moral regulation via the medium of an inner witness, self-judgment and self-punishment, is not so central to all cultures as it has been to our own. Some anthropologists contrast ‘guilt’ cultures with ‘shame’ cultures; in the latter it is fear of public exposure and loss of ‘face’ which is the principal vehicle of moral regulation (see Moral sentiments §3).

It is inappropriate to ask which of the above four aspects of conscience comprises its essence or makes up what it ‘really’ is. In different contexts one or more of these aspects may be in view and it is more important to appreciate the variety of elements here than it is to determine which of them is definitive of conscience.

See also: Common-sense ethicsIntuitionism in ethics; Moral agents

NICHOLAS DENT

References and further reading

Aquinas, T. (1266-73) *Summa theologiae* (Synopsis of Theology), ed. T. Gilby, London: Eyre & Spottiswoode, 1970, vol. 11, Ia.79.11-13; 1966, vol. 18, IaIIae.19.5-8. (Foundational discussion of profound significance.)

Benedict, R. (1946) *The Chrysanthemum and the Sword*, Rutland, VT: Charles E. Tuttle, ch. 10. (Important discussion of guilt and shame cultures.)

Bennett, J. (1974) ‘The Conscience of Huckleberry Finn’, *Philosophy* 49: 123-34. (A very witty and incisive discussion of erroneous conscience.)

Butler, J. (1726) *Fifteen Sermons Preached at the Rolls Chapel*, Sermons I, II, III, XI, XII; repr. in *Five Sermons Preached at the Rolls Chapel and A Dissertation Upon the Nature of Virtue*, ed.S. Darwall, Indianapolis, IN: Hackett Publishing Company, 1983. (Another foundational discussion of great importance, in which conscience is presented as part of a hierarchical account of the elements of human motivation.)

D’Arcy, E. (1961) *Conscience and Its Right to Freedom*, London: Sheed and Ward. (Through an examination of the historical sources and of contemporary debate, argues for a moral obligation to follow judgments of conscience made in good faith and for a right to religious freedom.)

Freud, S. (1912) ‘Totem Und Tabu’, *IMAGO* 1: 213-27, 301-333; trans. and ed. J. Strachey, ‘Totem and Taboo’, in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 13, London: Hogarth Press, 1955, part II, ‘Totem and Emotional Ambivalence’. (Considers problems of social psychology from the point of view of psychoanalysis, and compares taboo prohibitions with obsessional neuroses.)

Freud, S. (1927) *Die Zukunft Einer Illusion*, Leipzig, Vienna and Zurich: Internationaler Psychoanalytischer Verlag; trans. and ed. J. Strachey, *The Future of an Illusion*, in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 21, London: Hogarth Press, 1955. (Presents an account of conscience in terms of the work of a punitive super-ego.)

Freud, S. (1930) *Das Unbehagen In Der Kultur*, Leipzig, Vienna and Zurich: Internationaler Psychoanalytischer Verlag; trans. and ed. J. Strachey, *Civilisation and its Discontents*, in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 21, London: Hogarth Press, 1955. (A more extensive account of the development of conscience in terms of the renunciation of instinctive gratification and the formation of the super-ego.)

Jones, D.H. (1966) ‘Freud’s Theory of the Moral Conscience’, *Philosophy* 41: 34-57. (Critical assessment of Freud’s theories.)

Lewis, C.S. (1967) *Studies in Words*, Cambridge: Cambridge University Press, ch. 8. (An exceptionally illuminating etymological essay, in which the evolution of the terms related to the nature and exercise of conscience is related to deepening understanding of its significance.)

Potts, T.C. (1982) ‘Conscience’, in N. Kretzmann, A. Kenny, J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press. (A survey of medieval theories of conscience.)

Wallace, J.D. (1978) *Virtues and Vices*, Ithaca, NY: Cornell University Press, ch. 4. (A careful discussion of many aspects of conscience.)

Consciousness

Philosophers have used the term 'consciousness' for four main topics: knowledge in general, intentionality, introspection (and the knowledge it specifically generates) and phenomenal experience. This entry discusses the last two uses. Something within one's mind is 'introspectively conscious' just in case one introspects it (or is poised to do so). Introspection is often thought to deliver one's primary knowledge of one's mental life. An experience or other mental entity is 'phenomenally conscious' just in case there is 'something it is like' for one to have it. The clearest examples are: perceptual experiences, such as tastings and seeings; bodily-sensational experiences, such as those of pains, tickles and itches; imaginative experiences, such as those of one's own actions or perceptions; and streams of thought, as in the experience of thinking 'in words' or 'in images'. Introspection and phenomenality seem independent, or dissociable, although this is controversial.

Phenomenally conscious experiences have been argued to be nonphysical, or at least inexplicable in the manner of other physical entities. Several such arguments allege that phenomenal experience is 'subjective'; that understanding some experiences requires undergoing them (or their components). The claim is that any objective physical science would leave an 'explanatory gap', failing to describe what it is like to have a particular experience and failing to explain why there are phenomenal experiences at all. From this, some philosophers infer 'dualism' rather than 'physicalism' about consciousness, concluding that some facts about consciousness are not wholly constituted by physical facts. This dualist conclusion threatens claims that phenomenal consciousness has causal power, and that it is knowable in others and in oneself.

In reaction, surprisingly much can be said in favour of 'eliminativism' about phenomenal consciousness; the denial of any realm of phenomenal objects and properties of experience. Most (but not all) philosophers deny that there are phenomenal objects - mental images with colour and shape, pain-objects that throb or burn, inner speech with pitch and rhythm, and so on. Instead, experiences may simply seem to involve such objects. The central disagreement concerns whether these experiences have phenomenal properties - 'qualia'; particular aspects of what experiences are like for their bearers. Some philosophers deny that there are phenomenal properties - especially if these are thought to be intrinsic, completely and immediately introspectible, ineffable, subjective or otherwise potentially difficult to explain on physicalist theories. More commonly, philosophers acknowledge qualia of experiences, either articulating less bold conceptions of qualia, or defending dualism about boldly conceived qualia.

Introspective consciousness has seemed less puzzling than phenomenal consciousness. Most thinkers agree that introspection is far from complete about the mind and far from infallible. Perhaps the most familiar account of introspection is that, in addition to 'outwardly perceiving' non-mental entities in one's environment and body, one 'inwardly perceives' one's mental entities, as when one seems to see visual images with one's 'mind's eye'. This view faces several serious objections. Rival views of introspective consciousness fall into three categories, according to whether they treat introspective access (1) as epistemically looser or less direct than inner perception, (2) as tighter or more direct, or (3) as fundamentally non-epistemic or nonrepresentational. Theories in category (1) explain introspection as always retrospective, or as typically based on self-directed theoretical inferences. Rivals from category (2) maintain that an introspectively conscious mental state reflexively represents itself, or treat introspection as involving no mechanism of access at all. Category (3) theories treat a mental state as introspectively conscious if it is distinctively available for linguistic or rational processing, even if it is not itself perceived or otherwise thought about.

1 Pre-Cartesian uses of 'consciousness'

As elsewhere in philosophy, Descartes' writings mark a major shift in philosophical preoccupation with consciousness (see Descartes, R.). Pre-Cartesian philosophers of mind rarely emphasize the terms 'conscious' or 'consciousness' (or clear equivalents). Post-Cartesian philosophers of mind rarely avoid such emphasis. This section compares Descartes' usage with earlier usage.

Descartes typically speaks of being 'conscious' to refer to an allegedly *intimate* source of knowledge about *one's own mental occurrences*. In the 'Conversation with Burman' he says that 'to be conscious is both to think and to reflect on one's thought' (1648: 335), where the term 'thought' extends widely, as in the 'Second Replies', to

'everything that is within us in such a way that we are immediately conscious of it' including 'all the operations of the will, the intellect, the imagination and the senses' (1641: 113). Descartes seems to treat everything mental as introspectively conscious, in these passages and elsewhere (see the 'Fourth Replies', 1641: 171). Elsewhere, however, Descartes seems to deny that introspection is complete, as in the *Discourse on Method*: 'many people do not know what they believe, since believing something and knowing that one believes it are different acts of thinking, and the one often occurs without the other' (1637: 122).

The resulting focus on the scope and limits of introspective knowledge (see §§6-7) is apparently responsible for the modern uses of 'conscious' and 'non-conscious' to mark a potential distinction between two kinds of mental states. Introspected states are (introspectively) conscious, while others, if any, are (introspectively) unconscious. Pre-Cartesian authors do not use the word 'conscious' to mark such a distinction, although some may be committed to the distinction, either implicitly or in other terms (Whyte 1962). To take a rather spectacular example, Socrates claims in Plato's *Meno* that since one's soul 'has been born many times' and 'has learned everything that there is', 'seeking and learning are in fact nothing but recollection' (81c-d). This seems to require that one has latent knowledge of which one can at best become aware with great difficulty (see Plato §11).

Many pre-Cartesian writers share commitment to a special fountain of reflective knowledge, often called 'inner sense'. In *Summa theologiae* Aquinas posits a 'common sense' which enables one, for example, to 'tell white from sweet', and adds that this common sense 'is also able to sense sensation itself, as when somebody sees that he is seeing' (I.78.4). This responds to Aristotle's apparent claim, in *On the Soul*, that it is 'by sight that one perceives that one sees' rather than by another sense (III 2, 425b12). On these views, the subject matter of such inner awareness is more restricted than for Descartes, including sensation but perhaps not other mental processes. It may even be that, for Aristotle, the relevant 'inner' perception - by which one sees that one sees - is directed at one's external sense organs *themselves* and not at anything Descartes would consider strictly 'mental'.

A nearer equivalent to 'introspective consciousness' is the Sanskrit term '*manas*', used widely in Hindu texts for a 'mind-organ' that functions like the external sense organs. For instance, the *Vaiśeṣika Sūtra* (*c.* 3rd century BC) claims that 'Intellect, pleasure, pain, desire, aversion, effort are perceptible by the internal organ' (Radhakrishnan and Moore 1957: 411). This inner perception is carefully distinguished from 'inferential knowledge', and is framed so as not to involve introspection of a 'self'. Indeed, while Descartes holds that he can introspect himself as a 'thinking thing', Hinduism characteristically claims that introspection reveals no 'true self' distinct from one's mental states, and Buddhism characteristically denies that there are 'thinking things' at all (see Momentariness, Buddhist doctrine of; Self, Indian theories of).

The word 'conscious' derives from the Latin words '*cum*' ('together with') and '*scire*' ('knowing'). In the original sense, two people who know something together are said to be conscious of it 'to one another', with the irresistible connotation that they are privy to a scandalous secret. By extension, one can be conscious 'to oneself' of secret shames - whence the original use of 'consciousness' for conscience, the inner accuser silently sharing knowledge of one's transgressions. This archaic moral sense of 'consciousness' is not a concern in this entry (see Conscience).

The Latin conjunction of '*cum*' and '*scire*' also has a use in which the prefix is merely emphatic, so that being 'conscious of' something simply means knowing it, or knowing it well. In this sense the word 'conscious' can also be used as an adjective: a 'knowing' being such as a normal person is a conscious being, while an 'unknowing' being such as a plant or sleeping person is an unconscious being. 'Conscious', like 'knowing', can be used in this way for things *with* minds but not for things with*in* minds, such as mental states. People and animals know things - are conscious of things - but mental states do not themselves know things. Thus, for example, Aquinas uses 'conscious' to describe *bearers* of mental states - such as human beings, animals and God - but not to describe mental states - not even 'seen' seeings. Since the main philosophical problems about consciousness concern the more modern distinction between conscious and unconscious *states*, this entry focuses neither on the distinction between conscious and unconscious *subjects*, nor on the broad 'knowing' sense of 'conscious' (see Knowledge, concept of). In effect, Descartes refashions the 'knowing well' sense of 'conscious', regimenting it for a *particular* source of knowing, introspection. Issues about the specific epistemological status of introspection are a central concern of this entry.

2 Post-Cartesian uses of 'consciousness'

Descartes' use of 'consciousness' for reflective knowledge spread rapidly through the next generation of European philosophers. In *An Essay concerning Human Understanding* Locke writes, as he does of 'reflection', that 'Consciousness is the perception of what passes in a Man's own mind' (1689: II.i.19); Leibniz's term for this is 'apperception'. In his *Critique of Pure Reason* Kant goes on to distinguish between 'empirical apperception' of a 'flux of inner appearances' - mentioning that 'Such consciousness is usually named *inner sense*' - and 'transcendental apperception' which is alleged to be a 'pure original unchangeable consciousness' that reveals a 'fixed and abiding self' (A107; see Kant, I. §6). This entry discusses one's introspective access to the 'flux' of particular events within one's mind, rather than substantive introspective knowledge about one's self - about whether one is made of physical or spiritual components, and whether one persists through time (see Mind, bundle theory of; Introspection, epistemology of §4; Personal identity).

In the broader 'knowing' sense (see §1), a creature is conscious of something just in case it knows something, independently of whether this knowing is itself introspectible. This ancient sense of 'conscious' lingers on, and broadens to cover any kind of belief or cognition (whether or not it is 'knowledge'), and any kind of attitude about something (whether or not it is 'cognitive'). In this sense, a creature has consciousness if it has any kind of 'intentional' mental state. By extension, the state itself can be said to be a state of consciousness, even if it is not introspectible. (This is distinct from the widespread claim that all conscious states are intentional - that 'all consciousness…is consciousness of something' (Sartre 1943: 11). On this broad sense of 'conscious', by definition, all 'ofness' is conscious ofness.) As twentieth-century philosophers of mind and language most often pursue concerns about intentionality using terms other than 'consciousness', intentionality will not be explored in this entry (see Intentionality).

In a still broader sense, 'mind' and 'consciousness' are synonyms, as are 'being mindful of' something and 'being conscious of' it, so that any kind of mental state (whether or not it is an 'attitude') is a state of consciousness. When Hegel (§5), Marx or Lukács speak of 'unhappy', 'false' or 'class' consciousness, or when political activists attempt to 'raise' consciousness, their concerns are usually equally well rendered using a general term such as 'knowledge', 'thinking', 'attitudes' or 'mentality' in place of 'consciousness'. It is not clear that their concern is with introspection, since they refer mainly to thoughts about (or seemingly about) non-mental things, independently of whether these thoughts are themselves introspectively conscious. (When Hegel refers to thoughts explicitly about the mind, he uses 'self-consciousness' rather than 'consciousness'.) Likewise, many scientific writings officially on 'consciousness' are about mentation and mentation-like activity in general, avoiding any question of whether the activity is introspectively conscious. Since these broad uses of 'consciousness' seem to introduce no distinctive philosophical perplexities, this entry puts them aside.

With the dawn of scientific psychology in the late nineteenth century, the central philosophical controversies about consciousness centre around whether consciousness can ever be explained by an objective science of the mind. The biologist Thomas Huxley provides an early attempt to express the sense of mystery: 'How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of Djin when Aladdin rubbed his lamp' (1866: 210). Since that time there have been many scientific advances in understanding the mechanisms of perception, thought and communication, and many philosophical advances in understanding the nature of intentionality and meaning. According to Thomas Nagel and many other philosophers, such advances must leave an unexplained residue, concerning *what it is like* to have phenomenally conscious experiences (as illustrated in the introduction; see Nagel, T. §4). Nagel writes that 'Consciousness is what makes the mind-body problem really intractable', identifying 'subjectivity' as its most troublesome feature:

> Fundamentally an organism has conscious mental states if and only if there is something it is like to *be* that organism - something it is like *for* the organism. We may call this the subjective character of experience…every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view.
>
> (1974: 166-7)

Until quite recently, the agenda for the philosophy of mind has been set more by epistemology than by the sciences of mind. Consequently, introspective consciousness has been the most central and important philosophical notion of consciousness. The remainder of the entry is divided between the explanation of phenomenal consciousness and

the epistemology of introspective consciousness.

3 Subjectivity and the phenomenal

Nagel argues that physical theories cannot explain one's phenomenal consciousness, because they abandon one's point of view (see §2). Science does indeed abandon one's point of view, in so far as one need not be able to understand or defend the theory (if one lacks relevant concepts or evidence), and in so far as subjects with other points of view should be able to understand and defend the theory. But does abandoning a point of view prevent *describing* the point of view, or the features viewed from that point? Even if Nagel is correct that one's point of view and one's phenomenal experience are 'essentially' connected, they may not be *exclusively* connected. An objective physical theory of phenomenal consciousness would presumably allow that phenomenal features are accessible from multiple points of view - for instance, both by some form of introspection and by some form of neurophysiological or psychological observation and theory. Furthermore, if experiences are *necessarily* introspectible or introspected (see §6) this might explain Nagel's claim that phenomenal features are 'essentially connected with a single point of view'.

Frank Jackson (1982) amplifies Nagel's challenge in his 'knowledge argument' against 'physicalism', the thesis that the world is wholly physical. He imagines a super-scientist, Mary, who has never seen anything coloured because she lives her life in a black-and-white room. From a black-and-white television in this room, she learns *all* the objectively specifiable physical (and causal or 'functional') facts in the world. When she finally leaves the room and first sees colour, Jackson argues, she learns a *new* fact about the nature of phenomenal experience; she might exclaim, 'Oh! It is like *this* to see red!' The new fact Mary learns cannot be identical to a physical or functional fact, or else it would be among the facts Mary already knows before leaving the room. So no wholly physicalist account of phenomenal facts can be true (see Colour and qualia; Qualia).

Most responses to Jackson's argument involve denying that Mary learns a new fact upon experiencing red. On some views, she learns *how* to do new things - to imagine experiencing redness or to recognize redness visually - without coming to know *that* any new fact obtains. On others, she learns that an *old* physical fact about experience obtains, but comes to know it *in a new way* - via introspective access or via new concepts. For example, one proposal (Lycan 1990) compares the relevant introspective ways of representing one's experiences with *simple inner-perceptual demonstratives*, ways of knowing colour experience that would be unavailable to Mary from within the black-and-white room. Consider an analogy: when one perceives a banana and thinks of it demonstratively - as 'this' - those who do not perceive the banana cannot think of it in the same way - simply as 'this' (while staring at something else). Likewise, if one's introspections of one's experiences involve some demonstratives of them, this would explain why these representations cannot strictly be shared by someone who does not 'perceive' the same experiences but merely thinks *about* them. (For marks against this strategy, see Raffman 1995; for further elaboration, see Tye 1995: 173-8.)

Even among philosophers who accept that phenomenal consciousness is *some* physical or functional process, there are doubts about the possibility of explaining it. Colin McGinn (1991) suggests that human beings are forever blocked from knowing the 'link' between the brain and consciousness, roughly because introspective consciousness gives no knowledge of brains, while neuroscientific access to brains gives no access to consciousness. Critics respond that one might learn the explanatory link by theoretical inference from *joint* introspective and scientific data, such as correlations between phenomenal features and brain states (see Flanagan 1992: ch. 6).

In slightly different ways, Jackson, Joseph Levine (1993) and David Chalmers (1996) argue that even if one can know what this 'link' *in fact* is, explanations based upon it cannot be as satisfying as other scientific explanations. In essence, the argument is that for any objective, scientific account of phenomenal consciousness, one can *conceive* of a creature that meets the conditions in the account but *lacks* phenomenal consciousness. In the extreme case, it is said, one can conceive of a world that is an exact *physical* duplicate of the actual world - complete with duplicate stars, planets, rocks, plants, animals and philosophers - but which lacks any phenomenal consciousness. All the human-like beings in that world would be non-phenomenal 'zombies'. So the prescientific concept of phenomenal consciousness is not such that scientific premises could necessitate, a priori, conclusions about the phenomenal. In Levine's terms, there is an 'explanatory gap' between physical reality and phenomenal consciousness. By contrast, for example, it is claimed that the prescientific concept of 'water' is such that scientific

premises about H_2O, can necessitate, a priori, conclusions about water. In particular, it is held to be part of our *concept* of water that *if* there is anything roughly unique in the lakes and rivers around us that has a preponderance of features such as boiling, eroding rocks, quenching thirst and so on, then it is water. If science establishes that H_2O meets this condition, then there is no further conceptual possibility that H_2O is not water.

At least three lines of response may be advanced against the explanatory gap argument. One concedes that no scientific premises a priori necessitate conclusions about the phenomenal, but insists that the same is true for scientific explanations of water, and so on. Perhaps someone who believes in water but denies that it boils, is in lakes and so on is making a *false* and *bizarre* claim but are not strictly *contradicting* themselves. The fate of this response presumably depends on the fate of general reservations about a priori or conceptual necessity (see Essentialism; A priori).

The second strategy also concedes the lack of a priori necessity, but attributes it to the idiosyncratic ways in which phenomenal facts are represented - for example, by introspection-based demonstratives (see Tye 1995: 178-81; Demonstratives and indexicals). By comparison, suppose that agent *A* holds banana *B* in front of his face, and comes to accept the demonstrative '*This* is a banana'. This conclusion cannot strictly be deduced from any demonstrative-free *descriptions* of the banana - that *B* is a banana in *A*'s hand, that *A* sees *B*, and so on. The subject could believe all of these premises, and still *conceive* that this is not a banana - perhaps by conceiving that he is not in fact agent *A*. This response to the explanatory gap seems at best to apply only to particular phenomenal conclusions, of the form: it is like *this* to have a given experience. It is an *incomplete* response, since there are phenomenal conclusions without demonstratives, namely, those of the form: there is *something* it is like to have a given experience.

The third strategy is to deny that there is an unbridgeable conceptual gap; in effect, to deny that zombies (non-phenomenal physical duplicates of phenomenal creatures) are even *conceptually* possible. In Nagel-like fashion, Robert Kirk analyses the concept of phenomenality as follows: a creature's phenomenal states are those with different 'characters' that are 'for the creature as a whole' (1994: ch. 5). He explains this in terms of neural states with different 'patterns of activation' that are 'directly available' to 'have different effects on processes of assessment, decision-making, and action initiation'. It follows that if a creature has states with characters for the creature, any physical duplicate of the creature does also. The problem is that Kirk's 'having character' does not express the same concept as Nagel's 'being like something'. Perhaps Kirk demonstrates how states can be 'for a creature', but he does not demonstrate that they are 'like something' for the creature, that is, phenomenal, in Nagel's full-blown sense. For example, *beliefs* may also be realized in different patterns of activation with the relevant direct availability, and so in that sense beliefs may be 'for' a creature, but beliefs are not clearly *themselves* 'like something' for a creature (see §6). Further analysis of the 'like something' idiom would be needed to complete this line of response to the explanatory gap (see Belief).

4 Dualism and the phenomenal

The attraction of subjectivity-based arguments against physicalism makes it important to understand the potential ramifications of various non-physicalist alternatives. The central metaphysical issues concern 'mental causation' and 'emergence' (see Mental causation). The two central epistemological ramifications concern knowledge of other minds and knowledge of one's own mind (see Other Minds).

Physical science promises an explanation of physical events wholly in terms of other physical events. For example, unsupported objects fall, not because they *want* to fall, but because of gravitational forces. Similarly, a large group of neurons cause another neuron to fire, not because of its embarrassment due to peer pressure or its fear of a mob, but because of electrochemical forces. In this way, there should be a purely physical explanation of the activities of brains in producing (reflex or non-reflex) behaviour. Yet *phenomenal consciousness* also seems to have physical consequences on behaviour. When one feels an itch, one scratches; this seems to be in large part *because* of what it is like to itch. But how can an itchy *feel* make a difference, if there is a purely physical explanation of one's hand motions? The physicalist may make room for phenomenal causation by maintaining that the feel of an itch is *wholly constituted by* (or, perhaps, identical with) features of the brain and body (see Mind, identity theory of). The *feel* causes the hand motion, because it is *made of* things that do so, just as the *brain* causes the hand motion by being composed of things that do so (that is, neurons). But this strategy is unavailable to the non-physicalist

who thinks there are mental objects, properties or states that are not wholly composed of physical entities (see Dualism). One possibility for the dualist is to maintain that some physical events are *inexplicable* purely in physical terms - this is Descartes' own position about non-reflex behaviour. Another prominent strategy is to defend 'epiphenomenalism' about the phenomenal, or the idea that, contrary to appearances, phenomenal features are irrelevant to the generation of physical events (see Jackson 1982; Chalmers 1996: ch. 4; Epiphenominalism).

If phenomenal states and features are not wholly constituted by physical states and features, how *are* they constituted? One idea, prevalent among neuroscientists, is that the phenomenal 'emerges' from physical *interactions*. By analogy, water has properties that are not explicable by the properties of hydrogen and oxygen separately. Water dissolves salt, but neither hydrogen nor oxygen does so separately (nor does hydrogen dissolve *part* of salt and oxygen the rest). Analogously, the whole of a phenomenal experience is taken to be more than the sum of its physical parts (that is, the physical entities that help realize it). The initial plausibility of such analogies seems to depend on a too-narrow conception of the 'parts' of a complex entity. A water molecule does have hydrogen and oxygen atoms as parts, but it also has various *relations* among these atoms as other 'parts'. The physical relation of *bonding*, for instance, is as necessary for water as the atoms are. If *all* of the parts of water are counted, it is not clear that water has properties inexplicable by the properties of these parts. This kind of 'emergence' conclusion - one that forgets to 'sum' some of the parts of a whole - is compatible with a wholly physicalist account of water, and of consciousness. In the absence of better analogies, non-physicalist philosophers have more commonly denied that there is *genuine* emergence, emergence of new features that are *inexplicable* in terms of combinations of other features. This leads some non-physicalists to the startling 'panpsychist' claim that mental 'ingredients' of phenomenal consciousness must be present in the tiniest bits of matter capable of comprising brains, and to the claim that phenomenal consciousness is a *fundamental* part of nature - perhaps along with subatomic mass and charge (see Chalmers 1996: 8; Panpsychism).

Epiphenomenalism and panpsychism exacerbate the epistemological problem of other minds, but in opposite directions. It seems that we have excellent, even if not quite *perfect*, justification for believing that other people have experiences. Plausibly, talk of experiences is not simply disguised talk of others' observable behaviour, but is instead to be *inferred* from behaviour with some degree of theoretical risk. Perhaps the inference is in part grounded on drawing an analogy between oneself and others, but inference from a *single* case gives one at best meagre justification for conclusions about others. The residual problem of other minds, then, is to explain how one attains an appropriate level of justification. The physicalist can maintain that one is justified in part because the phenomenal features of experience help explain behaviour - in fact, help cause it. But the epiphenomenalist seems forced to deny this, potentially rendering it *too hard* to attain knowledge of other minds. The panpsychist, on the other hand, renders it *too easy* to attain such knowledge; not only other people but other animals, plants, rocks and protons have mental features.

A final danger with the idea that phenomenal consciousness depends on nonphysical features is that it threatens self-knowledge about one's phenomenal states. The nonphysicalist who believes that zombies are possible believes that two people could be alike in all non-phenomenal respects, while one has phenomenal experience and the other does not. Each could be fully convinced of their own rich, detailed phenomenal experience, but one would be wrong. It is difficult to see how either could justify their belief that they are *not* a zombie, if zombies are possible. Chalmers tries to address this problem by stipulating that non-zombies are *necessarily* distinctively justified:

> To have an experience is automatically to stand in some sort of intimate epistemic relation to the experience - a relation that we might call 'acquaintance.' There is not even a conceptual possibility that a subject could have a red experience like this one without having *any* epistemic contact with it: to have the experience is to be related to it in this way…. [This] is to stand in a relationship to it more primitive than belief: it provides *evidence* for our beliefs, but it does not in itself constitute belief.
>
> (1996: 196-7)

This conclusion anticipates the idea that inner perceptions of one's experiences - taken to be more 'primitive' than reflective beliefs about one's experiences - are necessary for phenomenal consciousness (see §6). But where the non-zombie has inner perceptual states, by hypothesis the zombie has them also. Perhaps what zombies have are *misperceptions*, but this need not lessen their *justification* in accepting these misperceptions at face value, and in believing they have experiences.

Also, the heavy reliance on introspections of experience opens up two alternatives for physicalism. The first is to deny that there is experience at all, but only the favoured kind of introspections *as of* experience. The second - perhaps the more attractive - is to maintain that the favoured kind of introspections are *sufficient* for experience. This is to invert Chalmers' claim that 'to have the experience is to be related to it in this [inner perceptual] way'; if so, this is precisely because to be related to a state in this inner perceptual way is to have an experience. Neither strategy carries obvious commitment to anything nonphysical. (For related objections to 'absent qualia', see Shoemaker (1984: chaps 9, 14); and for a response, see Block (1980). Rey (1986) extends the point, arguing that *non*-psychological requirements on phenomenal consciousness - perhaps biological or neurophysiological ones - would also jeopardize introspective knowledge of experience; see also Chalmers (1996: 7).)

5 Eliminativism and the phenomenal

In addition to subjectivity-based arguments, phenomenally conscious experiences provide other interesting arguments against physicalism. In certain experiences one seems to be aware of phenomenal denizens of an inner mental world: coloured and shaped mental 'images', bodily 'sensations' such as 'pains' that may be throbbing or in one's limb, and inner 'speech' with 'private' volume and pitch. Such alleged mental objects as afterimages, pains and inner speeches, that are naturally reported as having properties of non-mental objects (for example, roundness, throbbingness or loudness), may be called 'phenomenal objects'. The argument against physicalism based on reports of phenomenal objects is simple. In such experiences nothing in one's brain or body or (causally relevant) environment is literally purple and round, literally throbbing and in a limb, or literally soft and medium-pitched. So if phenomenal objects do exist with these properties, they are not among the things in one's brain or body or environment. One must be either a dualist (see §4) or an 'eliminativist' about mental entities with these properties; that is, the physicalist must deny that objects with such properties exist (see Eliminativism).

The challenge for the eliminativist about phenomenal objects is to explain why people are often *tempted* to claims of phenomenal objects, with ordinary perceptible properties. Broadly, the temptation may be attributed to an *ambiguity* or *looseness* in ordinary reports of experiences, or to an *illusion* built into the experiences themselves, which may then be reported strictly and faithfully. Following a suggestion due to Ned Block, Michael Tye (1995) pursues the former strategy by pointing out that people often describe representations as *having* properties that they merely *represent*: the phrase 'a warm thermostat setting' may be used for a thermostat setting *of* warmth, and 'a nude painting' may be used for a painting *of* nudity. It should be no surprise then that people describe experiences *of* colour and shape as themselves *being* coloured and shaped, and so no surprise that, speaking freely, they treat them as images. The other strategy is latent in J.J.C. Smart's suggestion that 'There is, in a sense, no such thing as an after-image…though there is such a thing as the experience of having an image' (1959: 151). On one straightforward construal of 'the experience of having an image', an experience itself *represents* that there is an image with certain features, although there is no such thing (compare with Sartre's 'illusion of immanence', 1940: 5). The illusory-experience view has an advantage over the reporting-based view to the extent that afterimages *look* purple and round, pains *feel* dull or in motion, and inner speech seems to *sound* faint or high-pitched. By contrast, a warm thermostat setting need not itself feel warm, and a nude painting need not itself look nude (any more than other paintings look *clothed*). However, without an account of why experiences misrepresent phenomenal objects, the illusory-experience view does not adequately discharge the eliminativist's explanatory burden (for one account, see Rey 1997).

In addition to phenomenal objects, eliminativists have targeted alleged 'phenomenal properties' of experiences, or 'qualia'. There is not merely 'something' it is like to have a phenomenally conscious experience, but some *particular* 'thing' or things it is like. People sometimes *try* to describe these particular properties, for example, by saying that a given pain is 'sharp' or 'throbbing' to some degree, or that a given visual image is 'blurry' or 'moving'. Even if the eliminativist about phenomenal objects is correct that there are only experiences *as of* sharp throbbing pains and *as of* blurry moving images, descriptions such as 'sharp' and 'blurry' seem in some indirect or nonliteral way to convey particular aspects of what it is like to have these experiences. In a relatively cautious use of the word 'qualia', particular what-it-is-like properties, whatever their nature turns out to be, are qualia.

There are bolder uses of 'qualia' on which the word can apply only to properties of experience that pose challenges to scientific explanation. Some require qualia to be infallibly and completely accessible to introspection, which is puzzling on any plausible scientific explanation of introspection (see §§7-8). Some require

qualia to be inaccessible without introspection - for instance by purely behavioural or neurophysiological tests that one may perform on other people, without relying on an introspective understanding of one's own experience; this seems to preclude explanation of qualia as physical properties discoverable in multiple objective ways (see §3). Perhaps the most controversial philosophical idea about qualia, however, is that they are 'intrinsic' properties of experience. Metaphysicians dispute the correct account of intrinsicality, but the following may serve to convey the intuitive idea, as it applies to experience: for an experience to have a property intrinsically, the experience must have the property solely in virtue of the *spatiotemporal parts* of what realizes the experience. (This is meant to exclude everything that even in part exists when or where the experience does not, for example, stimuli that cause the experience, and behaviour and other mental states that the experience causes.) This seems to preclude explaining qualia in 'functionalist' terms by appeal to the causal role of experiences, or in 'intentionalist' terms by appeal to the representational content of experiences (see Functionalism). Defenders of the intrinsicality of qualia often argue for the possibility of 'inverted qualia', cases in which two experiences differ in qualia even though they have identical causal or representational relations to their mental and non-mental surroundings (see Block 1990).

Qualia in such bold senses have been rejected most forcefully by Dennett (§3). He describes several examples in which changes in whether we like or dislike certain tastes seem to change the tastes themselves, concluding that when someone 'thinks of "that taste" he thinks equivocally or vaguely' and 'need not try - or be able - to settle whether he is including any or all of his reactions' (1988: 61-3). Nevertheless he accepts that, through the changes in likes and dislikes, 'the taste is (sort of) the same', and defenders of intrinsicality may hope to explain such taste similarities by appeal to reaction-independent components of experiences (see Lormand 1994 for further defence of bold qualia).

An argument against the existence of intrinsic qualia can be built upon what G.E. Moore calls the 'diaphanousness' of perceptual experience:

> The moment we try to fix our attention upon consciousness and to see *what*, distinctively, it is, it seems to vanish: it seems as if we had before us a mere emptiness. When we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous.
>
> (1903: 450)

Gilbert Harman argues as follows:

> When you see a tree, you do not experience any features as intrinsic features of your experience. Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict that you will find that the only features there to turn your attention to will be features of the presented tree, including relational features of the tree 'from here'.
>
> (1990: 39)

Defenders of intrinsicality may claim not to satisfy Harman's prediction. This response gains plausibility in cases of degraded perception (for example, blurred or double vision), and it is not clear how Harman's argument is supposed to generalize from perceptual experiences to bodily-sensational, imaginative or thought experiences (see bodily sensations; Imagery). (For discussion of imagery by a philosopher in sympathy with Harman, see Tye (1995).) The prediction may hold for non-degraded perceptual experiences and perhaps 'upgraded' imaginings such as dreams, and in these cases perhaps it shows that no experienced features *seem* intrinsic to experience. One may hold that some experienced features *are* intrinsic to experience, but only by incurring a burden of explaining why these features seem to belong to trees and other non-mental objects. In fact, this burden remains even if qualia are taken to be *relational* features of perceptual experience, since the relevant experienced features are experienced as if they belong *objectively* to trees, and do not seem in experience to depend on the participation of the perceiver. Just as the eliminativist about phenomenal *objects* of experiences would do well to explain the illusory experience of their *presence*, so the *non*-eliminativist about phenomenal *properties* of experience - whether intrinsic or relational - would do well to explain the illusory experience of their *absence* (see Qualia).

6 Introspection and the phenomenal

Here the discussion begins to shift from phenomenality to introspection. To clarify the apparent difference

between phenomenal and introspective consciousness, consider whether a state can be conscious in one but not the other sense. Can there be *nothing* it is like to have a state, even when one is introspectively aware of it? Can there be *something* it is like to have a state, even when one is wholly unaware of it? The *facts* are murky and controversial, but it is important to be clear about the *possibilities*.

The introduction lists four kinds of states that are most clearly phenomenal: perceptual experiences, bodily-sensational experiences, imaginative experiences and streams of thought. There are mental states not explicitly on this list, notably 'propositional attitudes' such as the belief that snow is white (see Propositional attitudes). Usually one's belief that snow is white is latent and unintrospected, though one can raise it to introspective consciousness easily. Normally there seems to be something it is like to have such an introspectively conscious belief (see Belief). However, there is another possibility to explore. What having the belief 'is like' may be completely accounted for by what it is like to have experiences *accompanying* the belief, such as auditory imaginings of asserting the words 'Snow is white' (or 'I believe snow is white', or 'Mon Dieu! La neige! Blanche!'), or visual imaginings of some fictitious white expanse of snow, together with what William James (1890: 287-8) describes as feelings or imaginings of moving eyeballs, eyelids, brow, breath, jaw-muscles and so on as one thinks. Pending evidence of further aspects of what it is like to have the belief, this illustrates how there can be something it is like *when* one has an introspectively conscious state, although the state itself has no phenomenal character (Lormand 1996). Introspective consciousness - at least of the sorts available to beliefs - is unlikely simply to *be* phenomenal consciousness, and is unlikely to be *sufficient* for it.

Might introspection nevertheless be *necessary* for phenomenality? There is a tension between a 'yes' answer and the view that many species of animals can have experiences - that there is something it is like for cats and dogs to feel pain or to see bright lights, for instance. It is implausible that these beings have Cartesian reflective knowledge *that* they feel pain and see. This would require having *concepts* of feeling pain and of seeing, and perhaps a *self*-concept, and all this would seem to involve capacities beyond the reach of most nonhuman animals - for example, the ability to conceive of *others* as feeling pain and seeing, and the ability to *remember* or *envisage* oneself feeling pain and seeing (see Animal language and thought). Notoriously, Descartes himself accepts that nonhuman animals are 'automata' without mental states of any sort. Defenders of a reflective-knowledge requirement may either mimic this strategy, denying that animals have conscious experiences (Carruthers 1992), or else attempt to minimize the conceptual sophistication needed for the reflective knowledge (Rosenthal 1990).

This tension is more commonly taken to be a serious strike against a reflective-knowledge requirement on phenomenality (see, for example, McGinn 1982: 52; Dretske 1995: 18), especially given that a similar tension arises in the case of human infants. Thomas Reid objects against Locke that 'reflection ought to be distinguished from consciousness', since:

> From infancy, till we come to the years of understanding, we are employed solely about external objects. And, although the mind is conscious of its operations, it does not attend to them; its attention is turned solely to the external objects, about which those operations are employed.
>
> ([1785] 1969: 57)

Even for beings with the requisite conceptual capacities, it seems implausible that reflective knowledge must accompany *each* of their experiences. At any given moment one can attend only to a small proportion of the sensory stimuli one encounters. It is also difficult to attend simultaneously to the outside world and to one's experience of it, as Auguste Comte argues (1842: 21): 'The thinking individual cannot cut himself in two - one of the parts reasoning, while the other is looking on' (see Introspection, psychology of). Nevertheless, plausibly, there is something many inattentive perceptions of unattended stimuli are like; experience would be quite impoverished were it not for the contributions of background noises and odours, pressures on one's feet or seat, moisture under one's tongue, peripheral vision and so on. It is possible to maintain that one continually forms reflective *beliefs* about these experiences, but this fits poorly with the difficulty of *remembering* these experiences (after they change, for example).

An introspective requirement on phenomenality can blunt much of the force of these objections by distinguishing inner *perception* from the formation of reflective *beliefs*, and by distinguishing *inattentive* introspection from *attentive* introspection. Just as one might sense a daffodil without having a concept of daffodils, or a tendency to remember the daffodil, so perhaps one can inwardly sense an experience without having a concept of experiences,

or a tendency to remember the experience. Animals and babies might sense even if they cannot form beliefs; likewise, perhaps they can inwardly sense even if they cannot form reflective beliefs (see Belief). Also, according to Locke, just as there can be passive sensation, so reflection need not be done intentionally or with attention (1689: II.i.7). Along these lines, Brentano distinguishes between inner *perception*, which may be automatic and inattentive, and inner *observation*, which is actively guided by purposeful attention (1874: 29). It is true that a creature's most pressing cognitive needs require mental resources to be directed at the *external* world, but if inner perception is normally inattentive, it need not draw resources away from attentive outer perception (see Bodily sensations). It is an open question whether there can be phenomenal experiences without any kind of introspective awareness of them, even of a primitive sort (but see §7 below).

7 Completeness of introspection

Perhaps more thoroughly than Descartes (see §1), Locke identifies the mental with the introspectively conscious, claiming that it is unintelligible 'that any thing *thinks without being conscious of it*, or perceiving, that it does so' because 'thinking consists in being conscious that one thinks' (1689: II.i.19). This completeness claim has been rejected by most subsequent philosophers, with the prominent exception of some in the broadly phenomenological tradition, following Franz Brentano and Edmund Husserl (see §10).

The main evidence for introspectively unconscious mental processes is that they would fill certain theoretical gaps in the scientific explanation of behaviour and introspectible mental activity. This inferential strategy is most prominent in psychoanalytic attempts to explain otherwise bizarre dreams, associations among concepts, apparent slips of the tongue, emotional disorders, neurotic physiological reactions and so on (see Freud, S.; Psychoanalysis, post-Freudian; Unconscious mental states). Although the scientific status of such clinically-based explanations is controversial (see Psychoanalysis, methodological issues in), alternative explanations are elusive; witness Sartre's difficulties in trying to explain self-deception and 'bad faith' without appeal to unconscious mentation (1943).

Furthermore, many replicable psychological experiments lead to parallel conclusions about introspectively unconscious mentation in more mundane settings. In one family of experiments (Lackner and Garrett 1972), subjects are presented with an ambiguous sentence in one ear, and disambiguating words in the other ear, but so quietly as not to be noticed consciously by the subjects - typically the stimulus is reported as a meaningless noise. Nevertheless, the subjects' interpretations of the ambiguous sentence are predictable from the meanings of the disambiguating words. This is evidence that subjects not only identify but *understand* the words, without introspective consciousness of doing so.

Similarly, there is evidence of vision without introspective consciousness of vision in cases of subliminal visual perception (Dixon 1987) and 'blindsight' (Weiskrantz 1988). In these cases subjects act on the basis of information about the visual features of objects, despite denying - sincerely - that they have relevant visual experiences. Blindsight subjects have damage to certain neural pathways connecting portions of the retina to the visual cortex, yet in some sense they have perceptual states sensitive to these stimuli. For example, some ability to discriminate an 'X' from an 'O' is intact. This is evidenced by the preponderance of correct answers they can give to questions about the stimuli. When asked to reach for objects in blindsight regions, also, some subjects reflexively pre-orient their hand and fingers in ways suited to the specific shapes of the objects. What blindsight patients lack most clearly is any ability to *introspect* their perceptual states and abilities: they deny that they are perceiving; they respond to the stimuli only when coaxed to do so; and even then they take themselves merely to be guessing. Nevertheless, their guesses tend to be correct and their behaviour appropriate to the stimuli.

It is not only in unusual cases that there seem to be introspectively unconscious perceptual states. On virtually all detailed theories of normal vision, for example, cells in each retina register the amount of incoming light at various points, and cause further states representing sudden discontinuities of incoming brightness, which cause further representational states and, eventually, introspectively conscious visual experiences (see Vision). These early layers of visual processing seem well beyond the reach of introspection. Although these processes (unlike, say, blood flow in the brain) have many mentation-like features - they are assessable as correct or incorrect in relation to external stimuli; they may increase gradually in stability as evidence for them mounts; they may play a direct role in modulating intentional visuomotor action, and so on - they are at best somewhere in the vague boundary between the mental and the non-mental.

To the extent that introspectively unconscious perceptual states are phenomenal, they present a new threat to the claim that introspection is necessary for phenomenality (see §6). Disagreement arises about whether cases of subliminal perception or blindsight involve phenomenal consciousness - perhaps blindseeing is like something, despite the subject's sincere denials; at any rate the cases are not clear enough to weigh decisively against the necessity claim.

There is more dispute about whether non-introspectible information-bearing states in early visual processing are phenomenally conscious. On the view that they are, it would be difficult to explain why phenomenal visual experiences are not continually like double images, given that one has *separate* left-eye-caused and right-eye-caused early visual states. On behalf of the view, one possibility is that there is something it is like *for one's early visual systems* to have certain states, although there is nothing it is like *for one* to have them. Although this is a possibility, it seems no more likely than the possibility that there is something it is like *for one's neurons* when they fire.

8 Reliability of introspection

Even if introspective consciousness is limited, it may be epistemologically interesting as an especially reliable means of access within its domain (see Introspection, epistemology of). Many philosophers have thought that introspective access to mental facts is more reliable than access to other empirical facts. Augustine writes in *On the Trinity* that 'nothing can be more present to the mind than the mind itself' (X.3.5), and asks rhetorically, 'what is so intimately known as the mind, which perceives that it itself exists and is that by which all other things are perceived?' (VIII.6.9). Descartes devotes his *Second Meditation* to an argument that the mind is 'better known' than the body, and Locke also claims that our knowledge of 'Things without us' is 'not altogether so certain, as our intuitive Knowledge' (1689: IV.xi.3). As with completeness, most subsequent philosophers reject the *infallibility* of introspection, although many would agree that it is *relatively* reliable.

The claim of introspective infallibility is extremely bold. In other empirical domains, at best certain mechanisms keep one's beliefs in rough accord with the facts (for example, mechanisms of perception, reason and memory, and the persistence of facts when one is not continually checking them). But mechanisms fail; a mechanism of this complexity that could never possibly fail would be a miracle 'at least as mysterious as papal infallibility', according to Dennett (1988: 55). The same reason to expect fallibility holds for introspection: if there is the slightest mechanism correlating one's thoughts with one's thoughts about them, it should be breakable, and if there is no mechanism, a perfect correlation between the two would seem to be sheer luck.

Furthermore, scientific investigations of introspection have revealed widespread 'confabulation' in self-access. In identifying one's beliefs and motivations, one systematically but sincerely reports attitudes one thinks rational or statistically normal in the circumstances, even if one does not have them (see Nisbett and Wilson 1977; Introspection, psychology of). For instance, in the 'bystander effect', increasing the number of joint witnesses to a person's need decreases the likelihood that any of them will assist. But bystanders rarely report this as a factor in their decision whether to help, often claiming instead to have reached a decision based solely on their own likelihood of success. According to Nisbett and Wilson, much allegedly 'introspective' access to attitudes consists of self-directed, fallible guesses, based at best on common-sense abilities to rationalize behaviour. Also, since these abilities are at work in one's access to others' mental states, this self-directed guesswork provides some reason to suppose, with Ryle (1949), that introspective consciousness is not interestingly more reliable than access to other minds.

In the face of this evidence, the Cartesian may attempt to identify a restricted domain in which a kind of introspection is comparatively reliable (*Ericsson and Simon 1993*) or even infallible (Lormand 1994). The model of rationalizing or statistical guesswork does not extend easily to introspection of phenomenally conscious experiences. For example, untutored subjects offer consistent and apparently reliable reports of 'feelings' of stinging (rather than throbbing) pain when a limb has restricted blood flow. They seem not to infer these feelings in the way one might form beliefs about another's pain feelings, since no common-sense principles of rationality dictate that one should feel stinging rather than throbbing, and since the subject need know no relevant statistical information about how people feel in these circumstances.

Against any attempt to find a restricted domain safe for infallibility, Dennett argues that one can easily be wrong

about the ‘changes and constancies’ in one’s experience over even brief intervals of time (1988: 59). This weighs against infallible *memory* access to what it was like to have *past* experiences, though not against infallible or especially reliable access to *current* experiences.

9 Introspection as inner perception

How should introspective consciousness be explained? While Lockean inner perception (see §2) has some contemporary defenders, most notably David Armstrong (1980), this view has fallen on hard times in philosophy of mind.

How is inner perception supposed to be distinctively analogous to outer perception? Of course, inner perceptions are not generated by literal inner eyes, ears and their attendant experience-forming processes. Armstrong explains inner perception as being, like outer perception, ‘selective’ (incomplete), ‘fallible’ and ‘causal’. This illuminates little, since probably all cognitive processes have these features, even one’s most theoretical (versus perceptual) scientific beliefs, for example, those about quantum mechanics or cosmology. Another initially tempting idea is the Hindu one (see §1) that inner perception is a causal but *non-inferential* source of evidence about mental states (or, more cautiously, that it is as low as outer perception on flexible, all-things-considered inference). Might a mental state be introspectively conscious only if its bearer inwardly perceives the state, in this sense? One strong objection is that introspective access often involves self-directed theoretical inferences, often confabulatory ones (see §8). It is therefore best to restrict inner-perception views to current phenomenal experience (in accordance with the suggestions at the end of §8).

An influential objection is that inner perception requires phenomenal ‘sense data’ interposed between physical objects and one’s perceptions of them (see Sense-data; Perception). Accepting inner perception may seem to involve accepting that one at best perceives outer objects *indirectly* through perceptions of phenomenal mental entities. But such a mediation theory would have difficulty explaining why inwardly perceiving sense-data did not in turn require perceiving further entities (‘sense-data data’) and so on, infinitely. Such arguments against sense-data have been mounted against inner perception (as in Shoemaker 1994), but with unclear effect. Inner perceptions need not be *interposed* between objects and one’s perceptions of them - the causal chain in perceiving a table need not proceed from the table to an inner perception and then to a perception of the table. Rather, on a more natural view, the causal chain goes directly from the table to a perception of the table, and then (in cases in which the table perception is introspectively conscious) to an inner perception of the perception of the table.

Dennett argues that inner perception of a ‘Cartesian theatre’ of consciousness would be too sharp to explain the vague conscious/unconscious distinction, and that inner perception is wasteful - ‘once a discrimination has been made once [in outer perception], it does not have to be made again’ (1991: 127). Yet it can be vague whether something is inwardly (or outwardly) perceived, and inner perception of an outer perception need not re-represent features the outer perception is *of*, but instead may newly represent features *of* the outer perception (see Lormand 1994, for further defence of inner theatres).

Perhaps the most influential objection to inner-perceptual introspection is based on Moore’s ‘diaphanousness’ claim (see §5). The objection is that, since each outer-perceptual modality (seeing, hearing and so on) makes its own distinctive contribution to what experience is like, an additional modality of inner perception should be expected to make *its* own contribution, to change what it is like. But what it is like to introspect a perceptual experience seems simply borrowed from what it is like to have the experience itself (McGinn 1982: 50-1). When one tries to attend to features of normal experiences, one normally ‘sees through’ the experiences to outer objects. So a fundamental disanalogy between outer perception and alleged inner ‘perception’ is that the former, but not the latter, has its own phenomenology or perceptual quality. This is evidence against inner perception of current phenomenal experiences.

On the other hand, if inner perception is necessary for phenomenality (see §6), then instead of *borrowing* phenomenal qualities from an outer perception, as the diaphanousness objection alleges, inner perception helps *generate* these qualities together with the (otherwise phenomenally unconscious) outer perception. This may explain why inner perception does not add further qualia to an outer-perceptual experience; inner perception may *already* make its phenomenal contribution in the generation of an outer experience with qualia (see Qualia).

10 Alternatives to inner perception

Two rivals to inner-perception accounts treat introspective access as epistemically *less* intimate than inner perception. First, James maintains that introspection is always retrospective (1890: 187-), largely in reaction to Comte's denial that the mind can simultaneously split between ordinary thinking and awareness of that thinking (see §6). James is unhappy with Brentano's response to Comte - that *inattentive perception* can be split between outer and inner domains - because James seeks to defend the reliance on careful, attentive introspective reports in experimental psychology (see James, W. §2). Second, some maintain that introspection is always laden with theoretical (versus perceptual) inferences. The experimental evidence for confabulation and expectation-driven inference (see §8; compare with Lyons 1986) suggests that introspection is *often* theory-laden and retrospective, but does not suggest that *all* cases of introspective consciousness are, including the *seemingly* non-inferential consciousness of phenomenal experiences that *seem* to persist while being accessed.

One theory of consciousness that combines naturally with theory-ladenness is Rosenthal's 'higher-order thought' theory (1990). (A thought or belief is 'higher-order' in virtue of being about mental entities rather than non-mental entities.) According to Rosenthal's view, a mental state is conscious just in case one forms, in a suitably direct way, a thought that one has the state. The state may generate its higher-order thought through inference so long as these inferences are not themselves conscious. This condition is intended to rule out cases in which one comes to think about a state through very indirect inference - say, solely through believing the testimony of a psychologist. On this view, even if introspective access to phenomenally conscious experiences is somehow inferential, it would *seem* non-inferential simply because one lacks higher-order thoughts about the inferences.

At the other extreme from defenders of theory-ladenness and retrospection, many phenomenologists reject inner perception as not being intimate *enough* to explain introspective access. On one suggestion, some introspectively conscious mental states 'reflexively' represent *themselves* (in addition to representing other things). This conclusion is often embraced to avoid an infinite regress threatening the assumption that introspection is complete (see §6): an experience represents itself rather than being represented by a separate introspective state which (assuming completeness) must in turn be represented by a separate introspection of the introspection, and so on. Brentano argues that 'The presentation which accompanies a mental act and refers to it is part of the object on which it is directed' (1874: 128). Husserl also suggests that 'In the case of a perception directed to something immanent [that is, roughly, mental],…perception and perceived form essentially an unmediated unity, that of a single concrete cogitatio' (1913: 112). And Sartre insists that 'the first consciousness of consciousness' - what he calls 'pre-reflective consciousness' - 'is one with the consciousness of which it is [a] consciousness' (1943: 13-14). Given the prevalence of inferential, confabulatory access to one's mental states, reflexivity theories, like inner-perceptual ones, are best restricted to current phenomenally conscious experiences rather than to other mental states. Again, there is a possibility that one's only access even to these experiences is somehow much more confabulatory or inferentially sensitive to expectations than ordinary perception is, but pending evidence for this possibility the restriction to phenomenal experiences is reflexivity's best hope.

Since one often suffers ordinary perceptual illusions, the more analogous introspection is to perception, the more likely it would be that one would suffer naïve introspective illusions about what one's conscious experiences are like. But it rarely if ever happens that one mistakes, say, a dull pain for a sharp pain, in the way that one mistakes a roadside cow for a horse. This may be evidence that introspection is sometimes neither inner perception nor self-directed theoretical inference, but a process with fewer breakable causal links. Some introspective access may be like one's psychologically *primitive* abilities to shift among mental states. Just as the transition from believing *that p and q* to believing *that p* presumably takes place without intermediate inference or inner perception, so might the transition from (say) believing *that p* to believing *that I believe that p*, or the transition from *having a dull pain* to believing *that I have a dull pain*. As Shoemaker proposes:

> Our minds are so constituted, or our brains are so wired, that, for a wide range of mental states, one's being in a certain mental state produces in one, under certain conditions, the belief that one is in that mental state. This is what our introspective access to our own mental states consists in. The 'certain conditions' may include one's considering whether one is in that mental state…. The beliefs thus produced will count as knowledge, not because of the quantity or quality of the evidence on which they are based (for they are based on no evidence), but because of the reliability of the mechanism by which they are produced.

(1994: 268)

A challenge for this view is to explain lawful and systematic *patterns* among introspectible and non-introspectible states, without an *ad hoc* assumption, for each pattern, that it happens to be 'wired' to higher-order beliefs in the right way. For example, why are brain states governing autonomic bodily functions not introspectible? Why are perceptual experiences introspectible but not subliminal perceptions and early perceptual states? Why does it seem easier to introspect one's fleeting thoughts than one's deeply held beliefs?

All of the theories described thus far in this section reject the etymological suggestion that 'introspection' is a kind of inner perception, while retaining the assumption that a subject's introspective 'access' to a mental state is always a matter of somehow *representing* the state. Further alternatives emphasize other forms of access to a state, or the state's functional relations to processes other than a subject's awareness of them. Verbal reportability is the process most frequently appealed to as an alternative requirement for introspective consciousness (Dennett 1978), in keeping with ordinary and scientific reliance on reportability as a symptom of consciousness. One threat is that reportability may only seem relevant to consciousness because it correlates somewhat with inner-directed representation. Normally, in reporting, one *perceives* one's reporting - hears one's speech, feels one's facial motions, and so on - and is thereby in a position to *understand* one's reports - to recognize one's own voice and realize which mental states one's words express. By contrast, if there is speech without any kind of self-perception, perhaps as in some forms of hypnosis or sleeptalking, this may not seem sufficient for, or even relevant to, introspective consciousness.

Dennett's construct is nearer to Block's notion (1995) of 'access consciousness', as distinguished from introspection and from phenomenality. In Block's terms, a mental state is access conscious if it is poised for free rational control of inference and action. The relevant actions may, but need not, include verbal actions, and the relevant inferences may, but need not, include Cartesian reflective knowledge. Although access consciousness has played little role in the history of philosophical discussions of consciousness, many contemporary writers employ comparable notions in attempts to explain consciousness generally (*Dennett 1991*; Tye 1995; Chalmers 1996).

See also: Awareness in Indian thought; Colour and qualia; Dualism; Materialism in the philosophy of mind; Mind, Indian philosophy of; Nirvāṇa; Phenomenological movement; Qualia; Reductionism in the philosophy of mind; Split brains; Supervenience of the mental

ERIC LORMAND

References and further reading

Armstrong, D. (1980) 'What is Consciousness?', in *The Nature of Mind*, Ithaca, NY: Cornell University Press, 55-67.(A sketch of an inner-perception view.)

Block, N. (1980) 'Are Absent Qualia Impossible?', *Philosophical Review* 89: 257-74.(Answers 'no', in response to Shoemaker (1984: ch. 9). Clear but increasingly intricate as possible replies mount.)

Block, N. (1990) 'Inverted Earth', in J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 4, Atascadero, CA: Ridgeview.(The inverted qualia argument mentioned in §5 is a sidelight to a novel argument against functionalism.)

Block, N. (1995) 'On a Confusion about a Function of Consciousness', *Behavioral and Brain Sciences* 18: 227-87. (Distinguishes access consciousness from phenomenal consciousness - see §10.)

Brentano, F. (1874) *Psychology from an Empirical Standpoint*, trans. A. Rancurello *et al.*, London: Routledge & Kegan Paul, 1973.(Study of intentionality and reflexive consciousness.)

Carruthers, P. (1992) *The Animals Issue*, Cambridge: Cambridge University Press. (Uses a higher-order-thought theory of consciousness to argue that nonhuman animals lack phenomenal experience.)

Chalmers, D. (1996) *The Conscious Mind*, Oxford: Oxford University Press. (Appeals to subjectivity in a defence of dualism, epiphenomenalism and panpsychism about phenomenal consciousness.)

Comte, A. (1842) *Introduction to Positive Philosophy*, trans. F. Ferré, Indianapolis, IN: Bobbs-Merrill, 1970. (Rejects introspection as a reliable means of knowledge.)

Dennett, D. (1978) 'Toward a Cognitive Theory of Consciousness', in *Brainstorms*, Cambridge, MA: MIT Press. (Characterizes conscious states as those available for linguistic report.)

Dennett, D. (1988) 'Quining Qualia', in A. Marcel and E. Bisiatch (eds) *Consciousness in Contemporary Science*, Oxford: Oxford University Press.(Challenges qualia claims, including intrinsicality and reliable access.)

Dennett, D. (1991) *Consciousness Explained*, Boston, MA: Little, Brown.(An attack on introspective theories of

phenomenal consciousness, with some leanings towards eliminativism.)

Descartes, R. (1637, 1641, 1648) *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 3 vols, 1984-91.(The boldest epistemological claims for introspection are in the 'Fourth Replies'. I have used 'conscious' in place of the translator's 'aware' to render Descartes' Latin 'conscius' and French 'connaissant' and 'conscient'.)

Dixon, N. (1987) 'Subliminal Perception', in R. Gregory (ed.) *The Oxford Companion to the Mind*, Oxford: Oxford University Press.(Brief overview of the research and controversies.)

Dretske, F. (1995) *Naturalizing the Mind*, Cambridge, MA: MIT Press.(An attempt to explain consciousness in terms of 'information functions'; criticizes introspective theories of phenomenality.)

Ericsson, K. and Simon, H. (1993) *Protocol Analysis*, Cambridge, MA: MIT Press.(A scientific defence of introspective reliability against Nisbett and Wilson (1977) and others.)

Flanagan, O. (1992) *Consciousness Reconsidered*, Cambridge, MA: MIT Press.(A good extended introduction to the central philosophical issues. Responds to McGinn (1991).)

Harman, G. (1990) 'The Intrinsic Quality of Experience', in J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 4, Atascadero, CA: Ridgeview.(Argues against intrinsic qualia using diaphanousness and the rejection of sense-data.)

Husserl, E. (1913) *Ideas*, trans. W. Gibson, New York: Macmillan, 1962.(The classic work of phenomenology as a reflexive study of consciousness.)

Huxley, T. (1866) *Lessons in Experimental Physiology* 8.(Early example of a long line of biologists who accept epiphenomenalism.)

Jackson, F. (1982) 'Epiphenomenal Qualia', *Philosophical Quarterly* 32: 127-36.(Includes the 'knowledge argument' against physicalism and a defence of epiphenomenalism.)

James, W. (1890) *The Principles of Psychology*, repr. Cambridge, MA: Harvard University Press, 1993.(Packed with sensitive but reasonably cautious reports from the heyday of introspective psychology. Defends introspection as retrospection.)

Kirk, R. (1994) *Raw Feeling*, Oxford: Clarendon Press.(Argues that phenomenal 'zombies' are conceptually impossible - see §4.)

Lackner, J. and Garrett, M. (1972) 'Resolving Ambiguity: Effects of Biasing Context in the Unattended Ear', *Cognition* 1: 359-72.(Psycholinguistic experimental paradigm with implications for subliminal perception.)

Levine, J. (1993) 'On Leaving out What it's Like', in M. Davies and G. Humphreys (eds) *Consciousness*, Oxford: Blackwell.(Careful statement of the 'explanatory gap' argument; emphasizes epistemological rather than metaphysical conclusions.)

Locke, J. (1689) *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Oxford University Press, 1975.(Classic work of empiricism; treats consciousness as complete and infallible 'reflection'.)

Lormand, E. (1994) 'Qualia! (Now Showing at a Theater near You)', *Philosophical Topics* 22: 127-56.(A defence of qualia and introspective theories of phenomenal consciousness against Dennett (1988, 1991).)

Lormand, E. (1996) 'Nonphenomenal Consciousness', *Noûs* 30: 242-61.(Argues that conscious moods and attitudes are non-phenomenal and uses this to criticize theories of phenomenal consciousness.)

Lycan, W. (1990) 'What is the Subjectivity of the Mental?', in J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 4, Atascadero, CA: Ridgeview.(A friend of inner perception treats subjectivity.)

Lyons, W. (1986) *The Disappearance of Introspection*, Cambridge, MA: MIT Press.(A valuable historical survey of theories of introspection, and a defence of theory-ladenness.)

McGinn, C. (1982) *The Character of Mind*, Oxford: Oxford University Press.(A brief but rich introduction to philosophy of mind; criticizes inner sense.)

McGinn, C. (1991) *The Problem of Consciousness*, Oxford: Blackwell.(Denies that human beings can ever grasp *how* consciousness is physical, even if it is.)

Moore, G.E. (1903) 'The Refutation of Idealism', *Mind* 12: 433-53.(Introduces diaphanousness by way of arguing against the claim that *esse* is *percipi*.)

Nagel, T. (1974) 'What is it Like to be a Bat?', repr. in *Mortal Questions*, Cambridge: Cambridge University Press, 1979.(Made the 'what it is like' idiom and subjectivity arguments popular in philosophical discussions of qualia.)

Nisbett, R. and T. Wilson (1977) 'Telling More Than We Can Know: Verbal Reports on Mental Processes', *Psychological Review* 84 (3): 231-59.(Influential psychological case for confabulatory and expectation-driven access to attitudes.)

Radhakrishnan, S. and Moore, C. (1957) *A Source Book in Indian Philosophy*, Princeton, NJ: Princeton University Press.(Includes ancient references to inner perception among both defenders of Hindu tradition and

sceptics.)
Raffman, D. (1995) ‘On the Persistence of Phenomenology’, in T. Metzinger (ed.) *Conscious Experience*, Thorverton: Imprint Academic.(Argues that the physicalist cannot use demonstrative or descriptive introspection to explain subjectivity.)
Reid, T. (1785) *Essays on the Intellectual Powers of Man*, repr. Cambridge, MA: MIT Press, 1969.(An attempt to inject ‘common sense’ into empiricism; emphasizes existence of pre-attentive consciousness.)
Rey, G. (1986) ‘A Question about Consciousness’, in H. Otto and J. Tuedio (eds) *Perspectives on Mind*, Dordrecht: Reidel.(Argues that our concept of consciousness is over-simple and incoherent; uses first-person knowledge against spiritual and biological accounts.)
Rey, G. (1997) *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Oxford: Blackwell. (Chapter 11 provides a relatively detailed functionalist theory of some phenomenal states, and attempts to explain mistaken ‘reification’ of phenomenal properties.)
Rosenthal, D. (1990) ‘A Theory of Consciousness’, ZIF Report 40, Bielefeld: Center for Interdisciplinary Research.(A higher-order-thought theory .)
Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.(Attack on Descartes’ dualism and epistemological claims, and defence of philosophical behaviourism.)
Sartre, J.-P. (1940) *The Psychology of Imagination*, trans. New York: Philosophical Library, 1948.(Sustained attack on the illusion of immanence.)
Sartre, J.-P. (1943) *Being and Nothingness*, trans. H. Barnes, New York: Philosophical Library, 1956.(Classic work of existential phenomenology; defends ‘pre-reflexive consciousness’.)
Shoemaker, S. (1984) *Identity, Cause, and Mind*, Cambridge: Cambridge University Press.(Chapter 9 is an intricate defence of functionalism against zombies; chapter 14 is an intricate reply to Block (1980).)
Shoemaker, S. (1994) ‘Self-Knowledge and "Inner Sense"’, *Philosophy and Phenomenological Research* 54: 249-314.(An extended attack on inner perception.)
Smart, J. (1959) ‘Sensations and Brain Processes’, *Philosophical Review* 68: 141-56.(Defence of materialism against phenomenal-object and other objections.)
Tye, M. (1995) *Ten Problems of Consciousness*, Cambridge, MA: MIT Press.(An attempt to explain phenomenality in terms of intentional content that is ‘poised’ for cognitive influence; appeals to demonstrative introspection against subjectivity arguments; also appeals to diaphanousness against intrinsic qualia.)
Weiskrantz, L. (1988) ‘Some Contributions of Neuropsychology of Vision and Memory to the Problem of Consciousness’, in A. Marcel and E. Bisiatch (eds) *Consciousness in Contemporary Science*, Oxford: Oxford University Press.(Brief, directed discussion from the leading authority on blindsight.)
Whyte, L. (1962) *The Unconscious Before Freud*, London: Tavistock.(A compendium of pre-modern and early modern views of ‘hidden’ behavioural influences. Citations are lax.)

Consent

*A concept of central importance in moral, political and legal philosophy, consent is widely recognized as justifying or legitimating acts, arrangements or expectations. In standard cases, a person's consent to another person's acts removes moral or legal objections to or liability for the performance of those acts. Thus, in medical practice the informed consent of a patient to a procedure can justify the physician's actions. In law, the maxim '*volenti non fit injuria*' (the willing person is not wronged) governs a wide range of acts and transactions, from the economic to the sexual. And in politics, it is often supposed that it is 'the consent of the governed' that justifies or makes permissible both governmental policies and the use of official coercion to compel obedience to law. Consent may be given in a variety of more and less direct forms, but its binding force always rests on the satisfaction of conditions of knowledge, intention, competence, voluntariness and acceptability of content.*

1 Meaning and importance

Consent is an act by which one freely changes the existing structure of rights and obligations, typically by undertaking new obligations and authorizing others to act in ways which would otherwise have been impermissible for them. In standard cases, one consents to an act or arrangement in response to a request by others; but consenting is so strongly linked with such acts as promising, contracting, taking oaths and entrusting, that it is best to regard all of these sources of self-assumed obligations as kinds of consent (as in Locke's political philosophy; see Locke, J. §10). In the most general terms, then, consent is any act by which an agent deliberately and suitably communicates to another the intention to undertake by that act new obligations and to convey by that act new rights to others. Given the satisfaction of various conditions, consent then authorizes, permits, justifies or legitimates actions by others or arrangements on which they may rely.

It is important to distinguish consent thus understood as a positive act from the weaker notions of consent - an attitude of approval, acceptance or agreement or as mere passivity, acquiescence or submission. These weaker forms of consent are far less important than and quite different from consent properly so-called. A 'pro-attitude' is neither necessary nor sufficient for true (morally binding) consent. Cases of grudging, off-hand or insincere consent show that a positive attitude is not necessary, while commonplace cases in which we approve of an arrangement without having in any way altered (or attempted to alter) the existing structure of rights and duties show that such an attitude is not sufficient. Similarly, one may passively submit to or even acquiesce in the control of a bully without thereby incurring new obligations of future compliance. Binding consent must be a positive act with the explicit, conventional, or otherwise clear meaning of a voluntary undertaking.

Justification by appeal to consent is especially important within the philosophical structure of liberal thought. Liberalism conceives of persons as self-conscious sources of value, whose choices of plans and actions are morally important and who have rights to govern themselves (within the bounds set by the rights of others) (see Liberalism). Consent is seen as the means by which this individual moral liberty may be limited in a fashion consistent with respect for liberty. For the consensual undertaking of obligation is one kind of use of personal liberty, a use which makes morally possible new and beneficial forms of social interaction. Justification by consent also has the advantage of being a relatively uncontroversial style of justification, since nearly everyone allows that what one has freely agreed to do ought to be done.

2 Forms and conditions

Attempted justifications by consent have appealed to a wide variety of forms of consent. The most fundamental divisions are those between personal and impersonal consent and between actual and hypothetical consent. Consensual justification of an act or policy with respect to some individual usually involves showing that the individual (expressly or tacitly) personally consented to it; but theorists have also appealed to the consent of the majority of an individual's colleagues (for example, fellow citizens) or to the consent of an individual's ancestors (for example, the founders of the society). Justification by appeal to personal consent is by far the most convincing of these options, since the authority of majorities over us seems to depend on our prior personal consent to be governed by them, and our forebears could not possibly have been authorized to consent for us.

Justification by consent also standardly appeals to the actual consent of the affected persons. This actual consent

may be given either by some discrete act or by some ongoing series of acts (such as active participation in the practices of some group). And actual consent may be either express or tacit (implicit, indirect). Express consent is consent given by an explicit verbal or written undertaking or by other direct but nonverbal consensual acts (such as raising one's hand). Tacit consent is given by actions or omissions (such as inactivity or silence) that do not involve an explicit undertaking, but that none the less constitute the making of a morally significant choice in the context of a clear, noncoercive choice situation. Some attempted justifications, however, appeal not to actual but to hypothetical consent. Hypothetical consent can be ideal (what fully rational persons would consent to), counterfactual (what real persons could or would have consented to, had they been pressed to give their consent) or dispositional (what real persons would have consented to, had they been able). Only appeals to the last of these (by which we justify, for example, imposing medical treatment on an unconscious injured person) seem to be genuine justifications by consent. Appeals to the others are really attempts to justify by showing that an arrangement is best or acceptable, independent of people's consent.

Consent of whatever form can only justify acts or arrangements given the satisfaction of a complex set of conditions for binding consent. These conditions can be divided into five categories. First, there are knowledge conditions. Where one is deceived or misled about an arrangement, or where one is otherwise non-negligently ignorant of crucial facts, apparent consensual acts may not bind (because consent is not 'informed'). Second, binding consent must be intentional, not only in the sense that the acts themselves must be intentionally performed, but also that they must involve the intention to give consent and thereby alter existing moral relations. Third, consent can only be given by the competent, which may exclude in various contexts apparent consent given by the insane, severely retarded, emotionally disturbed, immature, intoxicated and so on. Fourth, binding consent must be voluntary, limiting it to cases not involving the extraction of consent by coercion, undue influence, exploitation, unfair bargaining advantage and so on. Finally, consent only binds given acceptability of content. In most legal systems, for instance, agreements you make to commit crimes, become a slave or allow yourself to be killed are not enforceable. The moral basis for such rules could be either the alleged inalienability of the rights at issue or the contradiction supposedly involved in the creation of moral obligations to do the immoral.

3 Philosophical applications

Discussions of consent are most common in three areas of philosophical debate. In medical ethics the idea of informed consent is basic (see Medical ethics). Agreements to undergo some surgical procedure, for instance, are binding (and limit the physician's liability) only if patients understand (as far as they are able) basic facts about the nature of the procedure, its risks, costs and alternatives. In legal philosophy, much discussion of the proper structure of contract law turns on analyses of consent. And what constitutes consent is central to the definition of crimes like rape and sexual assault, where a showing of free consent to an act by a competent adult generally defeats criminal charges.

But by far the most numerous philosophical treatments of consent occur in political philosophy, where theorists argue the merits of consent theories of political obligation and authority (see Obligation, political; Authority). According to consent theory, the obligations of citizens to obey the law and support their governments and the authority or rights their governments have over them derive only from consent. The sort of consent relevant to this political justification is variously claimed to be actual and personal (Locke 1690; Beran 1987; Simmons 1993), impersonal or hypothetical (Rawls 1971; Pitkin 1965-6), discrete or ongoing (Walzer 1970; Pateman 1979). Since express consent is relatively rare in political life, it is commonly argued that consent can be given tacitly by continued residence in a state one is free to leave (Locke 1690), by voting (or possessing the right to vote) in democratic elections (Plamenatz 1968; Weale 1978) or by full and direct participation in political life (Pateman 1979). Many also argue, however, both that the conditions of knowledge, intention and voluntariness rule out many of those alleged sources of political consent and that taking seriously the ideal of government by consent forces the acknowledgement that no existing governments are in fact legitimate under these terms (Simmons 1993; Green 1988; Pateman 1979).

See also: Freedom and liberty; Rights; Liberalism §5

A. JOHN SIMMONS

References and further reading

All of the following involve detailed and sometimes subtle argument, but all are also accessible to readers without formal training in philosophy.

Beran, H. (1987) *The Consent Theory of Political Obligation*, London: Croom Helm.(The most complete presentation of consent theory in political philosophy.)

Feinberg, J. (1986) *Harm to Self*, New York: Oxford University Press.(The best discussion of the conditions on consent is given in chapters 22-6.)

Green, L. (1988) *The Authority of the State*, Oxford: Oxford University Press.(Careful discussion in chapter 6 of the relation between consent and political authority.)

Kleinig, J. (1982) 'The Ethics of Consent', *Canadian Journal of Philosophy* supplementary vol. 8: 91-118. (Thorough treatment of the nature of and conditions on consent.)

Locke, J. (1690) *The Second Treatise of Government*, in *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1960.(The classic statement of the consent theory of political obligation and authority, centred in chapters 7 and 8.)

Pateman, C. (1979) *The Problem of Political Obligation*, Berkeley, CA: University of California Press.(Discusses the history and many varieties of consent theory in chapters 4-7, favouring a Rousseauian version.)

Pitkin, H. (1965-6) 'Obligation and Consent, I and II', *American Political Science Review* 59 (4): 990-9, 60 (1): 39-52.(Develops a hypothetical consent theory of political obligation.)

Plamenatz, J.P. (1968) *Consent, Freedom and Political Obligation*, Oxford: Oxford University Press, 2nd edn. (Discussion of consent theory, with a defence of voting as an expression of consent in the 'Postscript'.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(Powerful defence of political justification by appeal to hypothetical consent, with the implications for political obligation in chapter 6.)

Simmons, A.J. (1993) *On the Edge of Anarchy: Locke, Consent, and the Limits of Society*, Princeton, NJ: Princeton University Press.(Discussion of Locke and the implications of consent theory in chapters 3, 4, 7 and 8; extensive bibliography.)

Walzer, M. (1970) *Obligations: Essays on Disobedience, War, and Citizenship*, New York: Simon & Schuster. (Applies a distinctive conception of consent theory to various practical political problems.)

Weale, A. (1978) 'Consent', *Political Studies* 26 (1): 65-77.(Analysis of consent in terms of inducing reliance.)

Consequence, conceptions of

The idea of one proposition's following from others - of their implying it - is central to argument. It is, however, an idea that comes with a history attached to it, and those who blithely appeal to an 'intuitive' or 'pre-theoretic' idea of consequence are likely to have got hold of just one strand in a string of diverse theories. This entry introduces the main alternatives - to call them rivals would be too strong, since it suggests that they are necessarily in competition with one another. Simply put, consequence may be conceived as a relation that is or is not modal in character, and is or is not formal. Thus for Aristotle, consequence is both necessary and formal; for Chrysippus it is necessary but not formal; for Bolzano and Tarski it is formal but not necessary; and for Philo and Russell it is neither necessary nor formal. Conceptions of consequence that are neither necessary nor formal are also needed if justice is to be done to deduction in science, the law and daily life. Cutting across all these other differences there is a perennial controversy about relevance. Does implication always require a full-blooded connection between premises and conclusion, or may it hold simply because of some property of either separately, for example, because the premises are impossible?

1 Aristotle

Alone among the persons discussed here, Aristotle never defined his conception of consequence. A definition can be supplied, however, that brings out its distinctive features: necessity, formality and relevance. Namely, a number of premises imply a conclusion if, for all premises $P, \ldots$ and conclusions *Q* of the same logical form as the given ones, if $P, \ldots$ are (or were) true, then necessarily *Q* is (would be) true.

The ingredient of necessity here (unlike formality) features in Aristotle's definition of a syllogism. It is required by his demand that proofs should produce 'understanding' (*epistēmē*), coupled with his claim that understanding something involves seeing that it cannot be otherwise. Hence, a proof needs to establish the necessity as well as the truth of its conclusion, which means not only starting from necessarily true axioms but proceeding by steps that preserve necessity as well as truth.

A part of Aristotle's project of axiomatizing science was to separate out those lines of argument that are applicable to all the sciences because they depend purely on topic-neutral terms such as 'every', 'some' and 'not'. These arguments are 'logical' in the special sense in which Aristotle sometimes - not always approvingly - uses the word to signify 'abstract'. The distinction between logical - that is, topic-neutral - and extralogical terms is not clear cut, but he sidesteps the difficulties by selecting a fixed handful of logical terms for study at any one time. Logicians have done the same ever since, splitting consequence in predicate logic, modal logic and so on, rather than lumping them into one broad relation of logical consequence.

Since an extralogical term does no work in making a logical argument valid, the argument remains valid if the term is replaced by any other. In other words, as the definition emphasizes, logical consequence is 'formal'. This feature is exploited by Aristotle's innovation of inference schemes, in which letters take the place of the non-working terms. The combination of formality and an artful selection of working terms creates opportunities for systematization and a mathematical treatment of the subject.

The definition uses counterfactuals because that seems to give the best fit with Aristotle's position on relevance. Where *P* is impossible and *Q* unconnected with it, 'If *P* were true *Q* would be true' is quite impossible to assess, and such cases are duly excluded by his dictum that premises only imply conclusions involving the same terms. But if a scheme is valid whenever its instances can be assessed (as with the syllogistic moods), he assumes that it holds universally; that is, even when the premises are impossible.

2 Chrysippus, Philo and Diodorus

Our knowledge of the fourth century BC dialecticians Chrysippus, Philo and Diodorus is fragmentary and second-hand, but they appear to have made two lasting contributions to the subject of consequence, one distinctly unhelpful, the other of enormous importance (see Philo the Dialectician §2; Diodorus Cronus §3).

Expressions of the form ' follows from ' or ' implies ' need to be completed by sentences embedded in quotation marks or 'that' clauses. As any logic teacher knows, this is not something that comes naturally, and the simpler 'If

P, *Q*' form is a convenient shorthand. The Greek dialecticians seem to have turned it into a regular convention, making the conditional *the* vehicle for discussing consequence. Apart from the other drawbacks of the convention, its blurring of an object language-metalanguage distinction is liable to make a study of consequence look like a contribution to propositional logic. This is especially unfortunate, since their great contribution was to propose the first definitions of consequence and inaugurate a debate about it.

For Chrysippus (the attribution is not certain but strongly probable), premises imply a conclusion if their truth is incompatible with its falsity. 'A man is running' thus implies 'An animal is moving'. As a class, such analytic implications are logical in a broad sense but not in Aristotle's special one; they are necessary without being formal.

For Philo of Megara, an implication holds if - for whatever reason - it is not the case that the premises are true and the conclusion false.

Diodorus Cronus' definition is that 'it neither was nor is possible' that the premises be true and the conclusion false. But since he equates the possible with what is or will be true, this amounts to saying that an implication holds if and only if Philo's condition is satisfied at every time.

Little has survived of the debate itself, alas, but relevance was certainly one issue, for we know that where *P* is impossible, 'If *P*, not-*P*' was acceptable to Diodorus but not Chrysippus. Another issue is likely to have been the extreme weakness of Philo's definition, to which Diodorus' looks like a critical rejoinder. To defend (and perhaps motivate) Philo's definition, one need only ask why there should be a higher standard for the steps of an argument than for its starting point. Premises do not have to be necessarily true, so why should conclusions have to follow *necessarily*? Weak as it is, Philo's definition captures what *Principia Mathematica* (1910-13) calls the 'essential property' of implication; that what is implied by a true proposition is true. It still seems far-fetched to suggest that, for example, all true propositions imply one another, but Philo has been defended even on this point. Russell (1919) distinguishes the presence of an implication from its practical utility for inference. To infer any truths from any others would indeed be *valid* (his word, his italics), but if they are already known to be true there will be no need to infer them, so that inference will in fact only take place if Philo's condition is known to obtain independently, a much stricter requirement.

3 Medieval logic

Medieval logicians tended to follow the ancient dialecticians' expository convention. Thus their 'sequences' (*consequentiae*) tend to be conditional propositions, though their discussion is appropriate to consequence: a sequence is typically said to be 'sound' (*bona*; *valet*) rather than true. The conditional convention lingered into modern times, as witnessed in the ambiguous status of Gentzen's '*Sequenzen*' and Russell's 'material implication'.

As this suggests, medieval logicians were familiar with all the ancient conceptions of consequence. They classified them into 'formal', that is, Aristotelian, and 'material'. Material consequence was subdivided into '*simpliciter*' (absolute), that is, Chrysippian; and '*ut nunc*' (as things are), that is, Philonian, the maxim '*ex falso sequitur quodlibet*' (from a falsehood anything follows) being a reminder of the latter. A few writers created a further subdivision, '*per accidens*' (incidental), to cover a Diodorean conception.

One medieval contribution has left a lasting mark on the subject. We have seen that different ancient conceptions produced different answers about relevance. The medieval development of propositional logic appeared to settle the question by producing a proof that *any* conclusion *Q* really does follow from an impossible premise *P*. The proof runs from *P* to P or Q, and thence, together with the necessary truth not-*P*, to *Q*. Ironically, the credit for this is generally given not to, say, William of Soissons, but to C.I. Lewis, a logician and historian of logic who had no time for the Middle Ages. The topic is still controversial: Bolzano argued that nothing at all follows from an impossibility, and not all modern logicians accept the Lewis proof (see Relevance logic and entailment §1). However, most follow the medieval line, which, when used to reform Chrysippian consequence, produces the familiar textbook formula for validity: 'It is impossible for the premises to be true and the conclusion false'.

4 Bolzano

We saw that Aristotelian consequence is formal, but this does not depend on its being specifically logical: a

comparable relation based on *any* division between working and non-working terms will automatically be formal with respect to the latter. Bolzano exploits this insight in the context of a theory of propositions that are - unlike the propositions mentioned elsewhere in this entry - non-linguistic items, composed of ‘ideas’ (1837). A conclusion is said to follow from a number of premises with respect to an arbitrary choice of ideas if every substitution for those ideas that turns the premises into true propositions also turns the conclusion into a true one. The other logical notions are relativized in a similar way. Bolzano is happy to include ‘logical consequence’ as a special case, that is, consequence when every extralogical idea is treated as substitutable. But his heart is in proving theorems about relative consequence in general. (Since he also broke with the old conditional expository convention, his results are unambiguously intelligible to the modern reader.)

Bolzano’s definition is non-modal. Indeed, a modal relation seems to be ruled out by his theory of ideas. His emphasis on the immutability of ideas and propositions and their reference to the actual, suggests that he cannot allow, as Frege can, that a false proposition *P* could be true in different circumstances. In different circumstances *P* itself would stay false; what would be true would, by hypothesis, be a different proposition. (Moreover, his resolutely non-linguistic stance prevents him from identifying it.) It follows that a modal definition of consequence on the lines of ‘Whatever the circumstances might be, if *P* were true *Q* would be true’ will not work (and cannot be rescued by appealing to ‘the proposition corresponding to *P*’).

5 Tarski

A century after Bolzano, Tarski took a similar approach to logical consequence, but where Bolzano had worked with non-linguistic propositions, he dealt in sentences. Given this choice, the obvious first step would be to replace Bolzano’s substitutions of ideas by substitutions of words. This would produce the following definition (labelled (F) by Tarski, 1936):

(F) A conclusion follows logically from a number of premises if every substitution for the extralogical vocabulary that makes the premises true, makes the conclusion true.

Indeed, Quine takes just this line when he defines a logical truth as a sentence that remains true under every substitution for its extralogical vocabulary. But the problem with such definitions is that in a particular language there may not be enough words of the right kind, so that an inference may come out as logically valid, or a sentence as logically true, simply because the counter-instances happen to be inexpressible. Quine plays down the difficulty by claiming that it will not arise if the language is ‘reasonably rich’, that is, capable of expressing arithmetic. His claim may be correct for logical truth but, as Boolos (1975) has shown, it does not extend to consequence.

Tarski had already come across this problem in his work on the concept of truth in formalized languages (1933). A first shot at the truth-conditions of a universally quantified sentence would be to say that it is true if and only if every particular instance of it is true; but this will not work if there are insufficient constants available, for example, if not every individual has a name. Tarski’s solution was to get round any shortage of genuine constants by treating free variables as roving constants, capable of taking each object of the appropriate type (individuals, relations, and so on) in turn as their ‘value’. Using the idea of a sentential function, in which free variables replace constants, he defined the ‘satisfaction’ (making true) of a sentential function by an assignment of values to its variables. The truth of $(x)Fx$ can then be equated with the universal satisfaction of the sentential function *Fx*.

It was a natural step to apply the same ideas to the definition (F). For brevity, let a ‘model’ of a sentence be an assignment of values that satisfies the sentential function obtained by replacing all the extralogical constants in the sentence by variables. Then, Tarski says, a conclusion follows logically from a number of premises if every model of the premises is a model of the conclusion.

This definition leaves open where to draw the line between logical and extralogical. In particular, it leaves one free to treat ‘thing’ (as in ‘everything *F*s’, ‘something *F*s’) as a logical constant or not. On the second option each assignment of values involves choosing a ‘domain of individuals’ as value for the variable that replaces ‘thing’. The notion of an assignment of values then becomes that of a structure, in the familiar sense of a collection of objects structured by various relations. If too the logical constants are restricted to a mathematically amenable few (for example, $\supset$ instead of ‘if’), the notions of satisfaction and model may be defined in purely set-theoretic terms

(see Model theory). Because of this, the resulting *model-theoretic* conception of consequence has been enthusiastically adopted by many mathematical logicians. Its intimate identification with set theory does, however, expose it to the troubles of that discipline. For, on the prevailing conception, the universe of sets is too large to be a set, and is therefore ineligible to be the domain of any model. For special reasons, first-order logic is unaffected, but the model-theoretic approach to second-order consequence has a question mark over it.

In keeping with the positivism of his day, Tarski wanted nothing to do with modality. (Granted, he uses ‘must’ when formulating (F) as a scheme, but only to indicate that the scheme holds universally; indeed, he stresses that (F) involves no notion more problematic than that of truth.) The non-modal aspect of his definition has been attacked, by Kneale (1961) and subsequently by Etchemendy (1990). A debate is called for, but it will be more fruitful if it asks for what purposes necessity is an essential ingredient of consequence. For example, someone who does not endorse Aristotle’s doctrine of proof and *epistēmē* may well be content with proofs that establish the bare truth of theorems, and it is not obvious that this requires a modal relation of consequence. Again, one motive for developing (say) group theory in the abstract is to derive laws that can be applied to all possible groups. But here a ‘possible group’ does not mean a non-existent structure composed of possibilia, but an actual structure waiting to be recognized as a group (compare with ‘a possible hiding-place’). Once more, a modal conception of consequence is not obviously required.

6 Enthymemes

For a genuinely pre-theoretic notion of consequence one needs to look outside the logic books, and to remember too that a philosopher’s conception of modality is not the only legitimate one: lawyers, scientists and others deal in *real* possibilities, not mere notional logical possibilities. In such circles, arguments take place on the basis of some propositions and some inferences which are common ground and taken for granted; call these an argument’s ‘basis’. But this is not peculiar to everyday argument. A striking example is the predicate calculus, with its built-in presumption that domains of individuals are non-empty (and the corollary that what it holds out as logical consequence fails the test of formality with respect to the extralogical term ‘thing’). Indeed *all* deduction has to proceed on some basis or other, for if every step were challengeable, argument could never get moving.

Historically, adherents of one conception of consequence have tried to exclude any other by the doctrine of the ‘enthymeme’, or incomplete argument. If an argument ‘*P*, so *Q*’ looks sound but does not measure up to the favoured standard, some part of its basis (call it *B*) is declared to be a missing, ‘suppressed’ premise. The original arguer was either incompetent or deluded: their argument should have been, or really was, ‘ P, B, so *Q*’. This expanded version will satisfy the favoured standard, which is thereby vindicated. As well as using this weapon against everyday consequence, logicians have turned it on each other. Traditional logicians use it to claim that all consequence is really syllogistic. Formal logicians of every stripe use it to claim that all necessary consequence is really formal (that is, someone who takes for granted the transitivity of ‘before’ is guilty of suppressing a premise, though someone who takes for granted the transitivity of ‘implies’ is not).

The self-serving character of the enthymeme doctrine is plain. Perhaps what is needed is, rather, an inclusive theory of consequence that will accommodate all kinds of basis for arguments. Let *B* be any basis; then a number of premises imply a conclusion on the basis of *B* if the conclusion belongs to the closure of the premises with respect to *B*. In the spirit of Bolzano, one can go on to develop the general theory of consequence relative to a basis. It needs to be complemented, however, by a quite different kind of discussion, about what bases are appropriate to what purposes and circumstances. In this connection (setting aside any differences over relevance), the broad conceptions of consequence descried in the previous sections can be seen as lumping together a class of eligible bases. Thus Chrysippian consequence covers - more exactly, it is the union of - consequence taken over all bases composed exclusively of analytic propositions and steps. Similarly, Philonian consequence is the union of consequence taken over all bases composed of true propositions and truth-preserving steps, though its paradoxical characteristics suggest that this is not the best way of dealing with so diverse a class.

See also: Logical and mathematical terms, glossary of

TIMOTHY SMILEY

References and further reading

Bolzano, B. (1837) *Wissenschaftslehre*, trans. R. George, *Theory of Science*, Oxford: Blackwell, 1972. (Discussed

in §4. The relevant sections of this readable but wide-ranging book are §§12, 19, 48, 66, 125, 147, 154-5, 182.)

Boolos, G. (1975) 'On Second-Order Logic', *Journal of Philosophy* 72: 509-27.(The appendix includes the technical result cited in §5. Despite the title, Boolos' result, like Quine's, relates to first-order logic.)

Carnap, R. (1934) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul, 1937.(Carnap briefly introduces consequence as closure with respect to arbitrary 'P-rules'. Tarski had anticipated the idea of closure under rules but, as his 'On the Concept of Logical Consequence' reveals, he was unwilling to take it beyond effective, finitary rules, that is, unwilling to use it to define consequence rather than deducibility.)

Etchemendy, J. (1990) *The Concept of Logical Consequence*, Cambridge, MA: Harvard University Press.(Uses a distinction between 'representational' and 'interpretational' semantics to develop Kneale's criticism of Tarski, from a standpoint of uncritical adherence to Chrysippian consequence, relevance and all.)

Kneale, W.C. (1961) 'Universality and Necessity', *British Journal for the Philosophy of Science* 12: 89-102. (Criticizes Bolzano and Tarski.)

Kneale, W. and Kneale, M. (1962) *The Development of Logic*, Oxford: Oxford University Press. (Chapters 2-4 include an invaluable introduction to the early history of consequence.)

Russell, B.A.W. (1919) *Introduction to Mathematical Philosophy*, London: Allen & Unwin.(Cited in §2. Philonian consequence is defended in chapter 14.)

Smiley, T. (1995) 'A Tale of Two Tortoises', *Mind* 104: 725-36.(Expands the theme of §6 of this entry.)

Tarski, A. (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger, 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Oxford: Oxford University Press, 1956, 152-278.(Cited in §5.)

(1936) 'O pojęciu wynikania logicznego', *Przegląd filozoficzny* 39: 58-68; trans. J.H. Woodger, 'On the Concept of Logical Consequence', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Oxford: Oxford University Press, 1956, 409-20.(This non-technical article is where Tarski expounds the idea of formal, non-necessary consequence discussed in §5. Although the option of varying the domain of individuals is not mentioned, it is given a thorough airing in his more technical 1933 piece.)

Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, Cambridge: Cambridge University Press, vol. 1, §1; 2nd edn, 1925.(Cited in §2.)

Consequentialism

Consequentialism assesses the rightness or wrongness of actions in terms of the value of their consequences. The most popular version is act-consequentialism, which states that, of all the actions open to the agent, the right one is that which produces the most good.

Act-consequentialism is at odds with ordinary moral thinking in three respects. First, it seems excessively onerous, because the requirement to make the world a better place would demand all our time and effort; second, it leaves no room for the special duties which we take ourselves to have to those close to us: family, friends and fellow citizens; and third, it might require us, on occasion, to do dreadful things in order to bring about a good result.

Consequentialists standardly try to bring their theory more into line with common thinking by amending the theory in one of two ways. Indirect act-consequentialism holds that we should not necessarily aim to do what is right. We may get closer to making the world the best possible place by behaviour which accords more with ordinary moral thought. Rule-consequentialism holds that an action is right if it is in accordance with a set of rules whose general acceptance would best promote the good. Such rules will bear a fairly close resemblance to the moral rules with which we now operate.

1 Act-consequentialism

Although the term ‘consequentialism’ is a recent coinage - it appears to have first been used in its present sense by Anscombe (1958) - it refers to a type of theory which has a long history. Consequentialism builds on what may seem to be the merest truism, namely that morality is concerned with making the world a better place for all. Consequentialist considerations certainly figure importantly in issues of public policy. Penal, economic or educational programmes are standardly judged by the goodness or badness of their results.

All moral theories offer an account both of the right and of the good. They all tell us, that is, both what makes an action right or wrong, and what kinds of thing are good or valuable. It is characteristic of consequentialist theories to assess whether an action is right in terms of the amount of good it produces (see §4). Deontological ethical theories, by contrast, hold that the right is independent of the good: certain kinds of action are wrong, and others right, independently of the goodness or badness of their consequences (see Deontological ethics; Right and good).

Act-consequentialism, the simplest form of the theory, holds that the right action - the one you should do - is the one which would produce the greatest balance of good over bad consequences; that is, the one which would maximize the good. (Where two or more actions come out equal best, then it is right to do any one of them.) Which action is in fact the right one will depend on what account of the good any particular act-consequentialist theory offers.

A theory of the good is an account of those things which are intrinsically good, good in themselves, and not merely good as a means to something else which is good (see Good, theories of the §2). A visit to the dentist is only extrinsically good, because it leads to healthy teeth and the avoidance of toothache, but it is not in itself a good thing; it is a necessary evil. By far the most popular and influential account of the good within the consequentialist camp is that offered by utilitarianism (see Utilitarianism). On this view, usually known as hedonism or welfarism, the good is pleasure, happiness or wellbeing (see Hedonism; Happiness). The act-utilitarian holds, therefore, that the right action is the one which maximizes happiness.

Many consequentialists reject hedonism. A pioneer in this respect was G.E. Moore (1903), whose theory, somewhat confusingly, used to be referred to as ideal utilitarianism, in contrast to the hedonistic variety (see Moore, G.E. §1). Among the things which have been held to be intrinsically good are knowledge, virtue, beauty, justice, and the flourishing of the environment as a whole. Many of these alternative accounts of the good are pluralist: that is, they claim that there are several different kinds of good thing which cannot all be brought under one head (see Moral pluralism). Pluralist act-consequentialism faces a difficulty. In order to determine which of the possible actions is the right one, agents must be able to rank the outcomes of each action, from the worst to the best. But if there are several distinct values which cannot be reduced to a common measure, how can one kind of value be compared with another in order to produce a definitive ranking? This is the problem of incommensurability of value.

The term ‘consequentialism’, though hallowed by frequent philosophical use, may be misleading since it might naturally be taken to imply that an action itself can have no intrinsic value; its value is all to be found in its consequences. Utilitarianism is indeed committed to this view - for what matters on the utilitarian account is not the nature of the act itself but the pleasure which it produces in anyone affected by it - but it is not an essential feature of consequentialism as such. Some consequentialists wish to leave room for the thought that certain kinds of action, such as lying, cheating, and killing the innocent, are intrinsically bad, while other kinds of act, such as generous, loyal, or just ones, are intrinsically good. Consequentialism can take such values into account in calculating which course of action produces the best results. In deciding whether one course of action is preferable to another, a consequentialist needs to know the total value that would be produced by taking each course of action, and that will include not only the value of the consequences but the value, if any, which attaches to the action itself.

Consequentialism is sometimes described as a teleological theory, because it conceives of a moral theory as setting a goal which we should strive to achieve (see Teleological ethics). The goal which consequentialism sets is to bring about a world containing the greatest balance of good over bad. Such a classification risks confusion, however, since a virtue ethics, such as Aristotle’s, is also usually classified as teleological, yet Aristotle’s theory differs from consequentialism in at least two crucial respects. First, the good at which agents aim, on Aristotle’s view (outlined in *Nicomachean Ethics*), is not the best state of the world, but the good life for humans; agents are to seek to realize distinctively human goods in their own lives. Second, Aristotle’s theory, unlike consequentialism, does not define the right in terms of the good. On the contrary, a full understanding of the good life rests on a prior conception of the right, for an important part of the good life consists in acting rightly (see Aristotle §§21-6; Right and good §2; Virtue ethics).

We also need to distinguish the kind of consequentialism with which we are here concerned from ethical egoism, which is sometimes classified as a consequentialist theory (see Egoism and altruism). Ethical egoism, which holds that the right action is the one which would best promote the agent’s own interests, is structurally similar to consequentialism in that the right action is the one which maximizes a good, in this case, the agent’s own good. What distinguishes egoism from the sort of consequentialism discussed here is that the latter is an *impartial* theory, giving equal weight to each person’s good (see Impartiality).

2 Criticisms of act-consequentialism

How should consequentialists set about deciding what to do? A natural answer is: by calculating, as best they may, what would produce the most good on any particular occasion when they are called upon to act. Of course, lack of time and knowledge limit what they can do by way of calculation, but they must do the best they can. So interpreted, however, act-consequentialism can be criticized for running counter to our intuitive moral convictions in a number of ways.

First, it seems excessively demanding; I shall only be acting rightly in so far as I maximize the good. Given all the bad things in the world, and the fact that few of us do much to improve them, it is clear that, in order to do what act-consequentialism requires, I would have to devote virtually all my energy and resources to making the world a better place (see Help and beneficence §2). This would give me no time or money to pursue my own interests, or even to relax, except to refresh me ready to redouble my moral efforts on the morrow. The degree of self-sacrifice required would make the lives of the saints look self-indulgent. Ordinary morality is surely not as demanding as this; it gives us permission to pursue our own goals, provided that we are not in breach of any of our fundamental duties. Some have proposed, in order to meet this point, that the theory be modified so that an act is right if its consequences are good, or good enough, even if they are not the best. This suggestion has not been widely adopted, for it is usually held that a rational agent will always prefer the greater good to the less.

Second, act-consequentialism appears to leave no place for the duties we take ourselves to have to our family and friends (see Family, ethics and the; Friendship). Such duties are often classified as agent-relative: each of us should help *their own* family and friends, so that the persons to whom the duties are owed vary from agent to agent. Act-consequentialism, however, is an agent-neutral moral theory; the goal at which we should aim does not depend on who the agent is. I should direct my efforts towards those for whom I can do the most good; their relationship to me is irrelevant. Even if act-consequentialism places special value on the cultivation of certain

relationships, such as friendship, this will still not yield a duty of friendship, as traditionally understood. If friendship is a great good, then my duty as a consequentialist is to promote friendship in general between all persons; that will not necessarily require me to give special attention to *my* friends, as distinct from helping others to give special attention to their friends.

Third, if act-consequentialism is too demanding in one respect it seems too permissive in another. For it leaves no room for the thought, central to much ordinary moral thinking, that there are certain constraints on our action, certain kinds of act, such as cheating, torturing and killing, which we ought not to contemplate, even if acting in one of these forbidden ways would maximize the good. The end, as we often say, does not justify the means. Once again, constraints seem to be agent-relative. Each of us is required not to kill or torture the innocent *ourselves* even if, by doing so, we could prevent two such tortures or killings.

3 Indirect act-consequentialism

Because it generates these counter-intuitive results, few consequentialists hold that agents should decide what to do by asking what will produce the best results. There are two theories which offer a less direct link between the overall goal of making things go as well as possible and how one should decide to act on any particular occasion. The first of these is known as indirect act-consequentialism. It retains the claim that the right action is the one with the best consequences, but denies that the virtuous agent need be guided directly by consequentialist thoughts when deciding how to act.

Indirect act-consequentialism builds on the thought that we do not necessarily hit the target if we aim directly at it. The gunner must make allowances for wind, gravity and poorly aligned sights; the moralist may have to direct our thoughts away from the goal if we are to achieve it. Act-consequentialism, on this view, tells us what the target is, but not how to hit it. It is not itself a good guide to action for a number of reasons: the calculations are tricky and time-consuming; we may be tempted to skew the results in our favour; doing the right action may require us to go against dispositions which are both deeply rooted and generally useful. So we may actually do better, in terms of achieving the goals which consequentialism sets us, if we do not aim to do what is right, but follow a few fairly simple moral rules of the traditional type, or encourage within ourselves the development of dispositions, such as kindness and loyalty, which will normally lead us to act in beneficial ways. In adopting such rules, or developing such dispositions, we know that we will sometimes act wrongly when we could, perhaps, have acted rightly. Yet we may still get closer, in the long run, to achieving the consequentialist goal than we would have if we had attempted to aim at it directly.

Some indirect act-consequentialists go further. Since we make better decisions if we eschew consequentialist calculations, it might be best if we rejected consequentialism. It seems possible that agents might behave worse, in consequentialist terms, if they were taught the truth of consequentialism than if they were brought up to believe some other moral theory. In which case consequentialists would do well to prevent its truth being generally known. Opponents see this position as incoherent. If the adoption of consequentialism demands its suppression then in what sense can we adopt it? How could a society be said to be governed by a moral code if no-one in that society believed it?

4 Rule-consequentialism

The second alternative to direct act-consequentialism is rule-consequentialism, which offers a more substantive role for moral rules or principles. Individual acts are judged right or wrong by reference to the rules; the rules, but not the individual acts, are judged by the results of accepting them. The right action is, roughly, the one that is in conformity with a set of moral rules which, if generally accepted, would tend to produce better results than any other set of viable rules we might accept. Rule-consequentialism differs from indirect act-consequentialism in two ways. It maintains that each decision should be guided by thoughts about which action is the right one, and denies that the right action is necessarily the one with the best results. In deciding which rules to accept we should bear in mind that the rules need to be clear, reasonably simple and not too difficult to comply with, given human nature. If they meet these requirements, it is likely that such rules will not be too dissimilar to our present ones.

Rule-consequentialism might be a plausible moral theory, but should it properly be seen as a form of consequentialism? It apparently abandons a central tenet of consequentialism: the claim that our goal should be to

maximize the good. The rule I should follow, on this view, is the one that *would* have better consequences, if generally accepted, than any other rule. If it is not, in fact, generally accepted, then in following it I may not get as close to maximizing the good as I would if I followed some other policy. For that reason, perhaps, act-consequentialism has remained most popular among defenders of the theory, despite its difficulties.

See also: Perfectionism

DAVID McNAUGHTON

References and further reading

Anscombe, G.E.M. (1958) 'Modern Moral Philosophy', *Philosophy* 33: 1-19. (A vigorous rebuttal, among other things, of what she saw as the pervasive strand of consequentialism in English moral philosophy.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. W.D. Ross, Oxford: Oxford University Press, 1972. (Arguably the greatest work on ethics in Western thought; considers what is required to live a full and satisfying life, in which the finest human capacities are exercised properly.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press, 1966, esp. ch. 5. (Argues that to say that an act is right is equivalent to saying that its total results will be the best possible, and claims that by far the most valuable things are personal relationships and the appreciation of beauty.)

Pettit, P. (1993) 'Consequentialism', in P. Singer (ed.) *A Companion to Ethics*, Oxford: Blackwell, 230-40. (A clear introduction to the topic.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press, esp. sections 10-14, 37, 41-4. (Contains one of the most sophisticated recent discussions and defences of consequentialism.)

Scheffler, S. (1982) *The Rejection of Consequentialism*, Oxford: Clarendon Press. (Argues for a 'hybrid' theory in which agents are permitted, but cannot be required, not to maximize the good.)

Scheffler, S. (1988) *Consequentialism and Its Critics*, Oxford: Oxford University Press. (A very useful collection of seminal articles.)

Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan; 7th edn, 1907, esp. I (ch. 9), II (ch. 1), III (chaps 11, 13), IV (chaps 2-5). (The classic source of many of the strategies now discussed by consequentialists and their opponents.)

Smart, J. and Williams, B. (1973) *Utilitarianism For and Against*, Cambridge: Cambridge University Press. (Two fine essays. Williams' piece introduces many of the objections to consequentialism that have figured in subsequent debate, including the notorious example of Jim and the Indians.)

Conservation principles

In antiquity 'self-evident' principles were used to argue for the conservation of certain quantities. The concept of quantitative conservation laws, such as those of mass and energy, is of much later origin. Even prior to the development of modern mechanics, symmetries were employed to solve some dynamical problems. The relation between conserved quantities and symmetries has come to play a central role in the physical sciences. Conservation laws may reflect as much about the way the human mind organizes the phenomena of the world as they do about physical reality itself.

One can distinguish between broad metaphysical principles of conservation and empirical conservation laws. Here we shall be concerned mainly with the latter. Historically, the development of each concept influenced that of the other. In the sixth and fifth centuries BC, several Greek thinkers (Thales, Anaximander, Anaximenes, Empedocles, Leucippus and Democritus) promulgated a range of materialistic philosophies, which included a belief that all substances in the cosmos are derived from some single substrate (or, at most, from a few of them) or that matter has an atomistic structure (see Matter §1). Like these ancient materialists, Antoine Lavoisier (1743-94) in his *Elements of Chemistry* (1789) took it as axiomatic that in natural processes no matter is ever created or destroyed. John Dalton's (1766-1844) atomic hypothesis, based on 'ultimate particles' of matter, also assumed conservation of matter or, in modern terminology, of mass. René Descartes (1596-1650), on basically theological grounds, adumbrated a law of conservation of momentum, while he and Christiaan Huygens (1629-95) had a proper conception of the law of inertia which is closely related to momentum conservation. However, it was only with Isaac Newton's (1642-1727) *Principia* (1687) that mechanics was provided with a sufficiently complete and consistent dynamical scheme from which conservation laws could be derived. In modern terminology, by a conserved quantity we mean one that is a constant in time by virtue of the dynamical equations of motion. The elegant formulations of classical mechanics by Leonhard Euler (1707-83), Joseph-Louis Lagrange (1736-1813) and Pierre-Simon Laplace (1749-1827) allowed one to find the conserved quantity corresponding to an invariance property of a mechanical system (see Mechanics, classical §5). As an example, the invariance of a physical system under rotation about some axis fixed in space implies the conservation of angular momentum about that axis. Given the dynamics, one then derives certain invariances and conservation laws.

The law of conservation of energy is one of the most important in physics. This law is very broad and extends far beyond mechanics. The physicists Benjamin Thompson (1753-1814) and James Joule (1818-89), and the chemist Humphrey Davy (1778-1829), were instrumental in demonstrating the interconvertibility of mechanical energy and heat. The first formulation of the concept of energy conservation was given by the physician and physicist Julius von Mayer (1814-78), although he received proper recognition only after Hermann von Helmholtz (1821- 94) rediscovered his work. Albert Einstein (1879-1955) unified the conservation of mass and of energy into a single law by arguing for the equivalence, or interchangeability, of mass and energy (see Einstein, A. §2). As ever more forms of energy are included in a generalization of the law of conservation of energy, the status of that law becomes almost definitional (but still extremely useful). In fact, Émile Meyerson held that there are principles of lawfulness and of causality which reflect basic and universal characteristics of human thought and understanding. This suggests that our various conservation laws may be rooted as much in the human mind as they are in any external reality.

Much of the fundamental work in twentieth-century theoretical physics is based on an intimate connection between conserved quantities and the symmetries of physical systems. In common usage the term 'symmetry' often refers to a certain pleasing proportion or balance. However, here we are concerned with a technical sense of this term, as in geometrical symmetries or in the crystalline symmetry of a snowflake. As an illustration, a plane equilateral triangle is left invariant under successive rotations of 120 degrees about an axis perpendicular to the plane of the triangle and through its centre. The figure looks the same after the rotation as before. Similarly, we can consider those transformations that leave unchanged the *form* of the mathematical equations describing a physical system. This connection between symmetries and conserved quantities had long been familiar in classical Lagrangian mechanics, but the mathematician Emmy Noether (1882-1935) proved a general result that continues to hold in relativity and in quantum theory. Just as invariance under displacements in space and time yields energy and momentum conservation, so, for example, what is termed the 'internal' symmetry of gauge invariance (a

mathematical operation that leaves Maxwell’s equations for electrodynamics unchanged) leads to the conservation of electrical charge. Another type of gauge invariance (a change of the arbitrary phase of the wave function in the Schrödinger equation) implies conservation of probability in quantum mechanics. What are termed higher symmetries can account for other conservation laws, such as that of baryon (or nucleon) number in elementary-particle physics. In a sense, physics has come full circle, building into theories those symmetries that will ensure certain conservation laws. When we have a complete dynamics, we use it to derive conservation laws. When searching for new theories, symmetries are a guide to discovery (see Electrodynamics; Field theory, quantum).

JAMES T. CUSHING

References and further reading

Elkana, Y. (1974) *The Discovery of the Conservation of Energy*, Cambridge, MA: Harvard University Press.(An examination of the origins of this concept in modern science.)

Gillispie, C.C. (1960) *The Edge of Objectivity*, Princeton, NJ: Princeton University Press.(A history of scientific ideas.)

Goldstein, H. (1950) *Classical Mechanics*, Reading, MA: Addison-Wesley.(A lucid and thorough exposition of the principles of mechanics, including the relation between symmetries and conservation laws.)

Hill, E.L. (1951) ‘Hamilton’s Principle and the Conservation Theorems of Mathematical Physics’, *Reviews of Modern Physics* 23: 253-60.(Noether’s theorem applied to continuous groups as used in physics.)

Sambursky, S. (1975) *Physical Thought From the Presocratics to the Quantum Physicists*, New York: Pica Press. (An intellectual history of Western physical science.)

Weyl, H. (1952) *Symmetry*, Princeton, NJ: Princeton University Press.(A very readable discussion of the role that symmetry has played throughout the ages in various cultures generally and in mathematics.)

Wigner, E.P. (1967) *Symmetries and Reflections*, Bloomington, IN: Indiana University Press.(A collection of essays, both technical and philosophical, on symmetries applied to modern physics.)

Wilson, L.S. (1970) *The Conflict Between Atomism and Conservation Theory 1644-1860*, London: MacDonald & Co.(An internalist history of the evolution of these two traditions.)

Conservatism

Conservatism is an approach to human affairs which mistrusts both a priori reasoning and revolution, preferring to put its trust in experience and in the gradual improvement of tried and tested arrangements. As a conscious statement of position, it dates from the reaction of Burke and de Maistre to the Enlightenment and Revolutionary thought and practices in the eighteenth century. Its roots, however, go far deeper. From Plato, conservatives derive a sense of the complexity and danger of human nature, although they reject emphatically his belief in the desirability of philosophical governance. From Aristotle, conservatives derive their sense of the need for practical experience in judging both moral and political matters, and their understanding of the role of tradition in inculcating habits of virtue and wisdom in the young.

Against Plato, conservatives prefer the limited government advocated by Hobbes, because of their belief in the ignorance and corruptibility of rulers, and because of their wish to encourage the self-reliance of subjects. They do, however, reject any conception of a social contract. In this, they follow de Maistre, who argued that creatures with the institutions and reactions necessary to form a social contract will already be in a society and hence have no need of such a thing.

While de Maistre emphasized the terror underlying political power, more characteristic of modern Anglo-Saxon conservatism is the position of Burke. For Burke, a good constitution is one adorned with 'pleasing illusions' to make 'power gentle and obedience *liberal'. It is also one which dissipates power in a society through autonomous institutions independent of the state. For both these reasons the communist regimes of eastern Europe could not be defended by conservatives, even though for a time they represented a form of social order.*

While conservatism is not antithetical to the free market, and while the market embodies virtues the conservative will approve of, for the conservative the market needs to be supplemented by the morality, the institutions and the authority necessary to sustain it. Human beings are by nature political, and also inevitably derive their identity from the society to which they belong. Our sense of self is established through our family relationships and also through the wider recognition and apportionment of roles we achieve in the public world beyond the family. According to Hegel, who since Aristotle has written most profoundly on the interplay of the private and the public in human life, both family and the public world of civil society need to be sustained through the authority of the state. On the other hand, the distinctions between family, civil society and the state need to be maintained against the characteristically modern tendency to treat them collectively. In his insistence both on authority and on the checks and balances needed in a good society, Hegel may be said to be the most articulate and systematic of conservative thinkers.

Conservatism has been much criticized for its tendency towards complacency and to accept the status quo even when it is unacceptable. However, in its stress on the imperfectibility of human nature and on the dangers of wholesale revolution, it may be said to be more realistic than its opponents. Conservatives can also be quite content with the claim that societies animated by conservative political structures have been more successful morally and materially than socialist or liberal societies. This claim they believe to be true, and it is a fundamental aspect of their position that the dispute between them and their opponents is, at bottom, an empirical one.

1 The conservative temper

Many strands of thought make up conservatism, and thinkers who are regarded as conservative have held a wide variety of opinions on many crucial issues, both philosophical and political. The protean and, to the intellectual mind, untidy character of conservatism is, however, part of its essence. Indeed, it is an important aspect of what distinguishes conservatism from other ideologies, or, as conservatives would say, from ideologies. For conservatism is an approach to human and, more specifically, to political affairs which mistrusts the power of human reason. Its temper is sceptical and empirical. It places more faith in tried and tested arrangements than in ideas, however brilliant. It recognizes the extent to which individuals are situated within their societies, and the extent to which societies are fashioned by their histories. Conservatism is also acutely conscious of the flaws in human nature, as well as its potential for good, and of the way in which civilized life depends on a context of order and tradition.

It has rightly been said that prior to the Reformation, it was impossible to distinguish conservatism in politics, not because none existed but because there was nothing else. However, even between the time of Luther and that of the French Revolution, while we can certainly discern conservative themes in writers such as Hobbes, Swift, Pascal and Hume, it is not until Edmund Burke's *Reflections on the Revolution in France* (1790) and Joseph de Maistre's *Considérations sur la France (Considerations on France)* (1797) that we find comprehensive and self-conscious statements of a conservative position. It is no coincidence that both these impassioned diatribes were conceived as reactions to what their authors perceived as a deadly threat to the established order of society, and in de Maistre's case, to the divine order as well.

The conservative position is one that is better unarticulated not, as its opponents might insist, because it is intellectually threadbare, but because its preference is for the harmony which arises from an unquestioning and untroubled acceptance of settled ways of doing things. In making articulate what is better left unsaid, conservatism is, from its own point of view, in danger of lapsing into ideology, into the erection of principle and dogma over practice and habit. As we will see, neither de Maistre nor certain modern conservatives have proved immune to this temptation.

2 Plato and Aristotle

While explicit conservatism may be a characteristically modern phenomenon, the intellectual roots of much conservatism can be found in the thoughts of Plato and, more particularly, Aristotle. From Plato conservatives derive a keen sense of the dividedness of human nature. For most conservatives would assuredly recognize the image in *Phaedrus* of the soul consisting of a charioteer struggling to harness together a good horse and a bad horse (*c*.366-360 BC; 253c-). For conservatives one incontrovertible lesson of history is the danger of licence and anarchy, and the human and political folly of the romanticism which would do away with political order in the name of freedom or even reason. Individuals do not have reason enough to order their own affairs, nor even do whole societies starting from scratch; and freedom, in practice, is too often freedom for one person's bad horse to do violence both to his or her own better nature and to those around. Thus conservatives would sympathize with the Platonic insistence on the need for order provided by a strong state, as featured in both *Republic* (*c*.380-367 BC) and *Laws* (*c*.360-347 BC) (see Plato §§14,17).

Conservatives would, however, have rather less sympathy with other aspects of Platonic thought. Conservatism is not committed to the view that decline is inevitable, nor indeed to the myth of a past golden age. Being more aware than non-conservatives of the problems involved in change is not synonymous with an obdurate refusal of all reform. Indeed, wholeheartedly rejecting the central Platonic doctrine of the possibility of infinitely enlightened rulers, conservatism must embrace mechanisms for gradual reforms necessitated by errors of policy, errors often visible only after experience of the effects of policy, and particularly to the ruled who suffer most from the effects.

Central to conservatism is the belief that even the best-intentioned action can have unexpected and undesired consequences. The ramifications of the unintended consequences are the greater when the actions in question are policies whose effects will bear upon a whole society. Even the cleverest and most beneficent agencies will inevitably be unable to predict all the effects of their policies in advance. Because of their inevitable ignorance and because, in any case, they are unlikely to remain completely uncorrupted by power, in contrast to Plato conservatives favour a system in which the scope of the actions of rulers is limited, and in which revision and reform is a real possibility.

Overall, in fact, the empiricism of Aristotle is far more congenial to conservatives than the rationalism of Plato. For Aristotle, both ethical enquiry and politics are areas in which there is no substitute for the wisdom which derives jointly from experience and the patient cultivation of good habits. If politics were simply a matter of cleverness, then we would expect to be able to learn about it from sophists, who profess to be able to teach it, but who are far from being able to practise it to good effect. We cannot learn about medicine or painting from people who are not doctors or artists. In like manner, the would-be legislator needs experience of life and politics, over and above intelligence; guidance is needed also from those versed in the intricacies of human affairs, rather than from the merely clever. Moreover:

> The soul of the student must first have been cultivated by means of habits for noble joy and noble hatred, like

> earth which is to nourish the seed. For he who lives as passion directs will not hear argument that dissuades him nor understand it if he does.
>
> (*Nichomachean Ethics* 1179,b,25-6)

Aristotle's assumption (see Aristotle §§27, 28), in contrast to that of moral educators in the romantic and rationalist traditions, who would have even young children submit all their moral beliefs to critical scrutiny, is that moral reasoning can proceed properly only on the basis of good habits, already ingested and ingrained. Aristotle's clear assumption, like that of the conservative, is that moral and political virtues emerge as the fruits of experience, and are passed on to the newcomer brought up within the tradition which recognizes experience as the wisdom of the past. To the newcomer this process of transmission is initially an unthinking one, purely a matter of habit.

As well as his moral conservatism, Aristotle insisted on hierarchies of talent and ability, albeit in a less rigid and doctrinaire manner that Plato. He also argued strongly that justice is not equality, but the giving to each of what each are owed. It is hard to think of any conservative dissenting from Aristotle's inegalitarianism, firmly based as it is in facts about human nature and in the actual development of history and culture. Nor would a conservative be likely to favour Plato's holist state over Aristotle's more limited type of rule in which, although the state is logically prior to the individual, private property, the family and other activities independent of the state are clearly envisaged.

3 Hobbes

Conservatives are dispositionally inclined towards a Hobbesian polity for at least two reasons. First, a strong government is needed in order to protect individuals and communities against injustice and the depredations of the strong, and to turn what would otherwise be a war of all against all into a settled, peaceful existence (see Hobbes, T. §6). Second, limited government is favoured both because of the ignorance and corruptibility of rulers and also to encourage the self-reliance of subjects.

Where, however, conservatives depart from Hobbes and all liberals of a contractarian sort is on the notion of a social contract (see Liberalism §5). A contract, as even Hobbes admits, is, *in extremis*, defeasible (Hobbes 1651). Whereas if, following Burke, conservatives see life as a partnership between the living, the dead and the unborn, it is not a partnership in which we, the living, can be absolved from our duties to our partners, if only because they are not here to release us. In any case, for conservatives the very idea of non-social animals - or as they view them, savages fleeing from the war of all against all - coming together to contract is a nonsense; such beings could never construct a system depending on the mutual recognition of obligations and enforcement of promises. Meanwhile, those creatures who could enact a social contract, with the rewarding, punishing and respecting of each other's rights and freedoms which such a notion requires, are already in a society, and so have no need of a primeval contract (de Maistre 1797) (see Contractarianism §§2-6).

4 De Maistre and Burke

De Maistre, whose criticism of the social contract is given above, held that society is not made by man, weak as he is and his 'reason' a feeble pretence, but by God. Thus, any attack on the social order amounts to a blasphemy. Furthermore, society is held together only by the executioner, 'the terror of human society and the tie that holds it together'. This is conservatism as reactionary ideology. De Maistre takes every cherished doctrine of the Enlightenment and turns it on its head. In so doing, he shows the same cavalier disregard for human life as those he attacks, to such a degree that he is led to compliment the Jacobins on the grounds that they, unlike their liberal-minded predecessors, at least killed people and restored order of a sort.

In so arguing, de Maistre falsifies what, since Burke, has become a central tenet of Anglo-Saxon conservatism; that the main defence against anarchy and against the war of all against all is not the brute power of the state. Burke had no more illusions about human motivation or ability than de Maistre, but he saw that sentiment had to be added to power in order for a state to become a community, tolerable for all to live in. Burke, like de Maistre, believed that 'no great human institution results from deliberation' but is the result of evaluation over time in ways often undreamt of by its founders, and also that 'human works are fragile in proportion to.… the degree to which science and reasoning' have been involved in their construction (1797: 57). Both men pour scorn on the

attempt to formulate a constitution or a bill of rights which is not simply an articulation of existing practice, and even more upon the characteristically modern delusion that liberty, rather than its opposite, can be secured by such a procedure. Furthermore, de Maistre's adage that 'threads which can be broken by a child at play can nevertheless be joined to form a cable capable of supporting the anchor of a great vessel' (1797: 75) is entirely in the spirit of Burkean conservatism, in both its positive and negative aspects; negatively, that seemingly trivial criticism can undermine the spirit of a great enterprise; positively, that what binds a great society need not, indeed must not, seem onerous, or the cause is dead.

And it is here that Burke and de Maistre part company. For Burke, and those of his followers supporting the tradition of Anglo-Saxon conservatism, what is crucial in a successful constitution are 'the pleasing illusions' which make 'power gentle and obedience liberal', which harmonize 'the different shades of life and which, by a bland assimilation' incorporate 'into politics the sentiments which beautify and soften private society'. Revolutions tear off all this 'decent drapery of life' and expose our 'naked shivering nature'.

> On this scheme of things, a king is but a man; a queen is but a woman; a woman is but an animal; and an animal not of the highest order. All homage paid to the sex in general as such, and without distinct views, is to be regarded as romance and folly. Regicide, and parricide, and sacrilege, are but fictions of superstition, corrupting jurisprudence by destroying its simplicity. The murder of a king, or a queen, or a bishop, or a father are only common homicide; and if the people are by any chance, or in any way, gainers by it, a sort of homicide much the most pardonable, and into which we ought not to make too severe a scrutiny.
>
> (Burke 1790: 171)

Burke goes on to contrast this conservatism with the barbarous philosophy, the offspring of cold hearts and muddy understanding, in which 'laws are to be supported only by their own terrors' (see Burke, E.)

De Maistre was certainly as guilty as some of his opponents of subscribing to a barbarous philosophy, and to that extent should be regarded as outside the conservative mainstream. Moreover, the spirit of what Burke says need not require sentiment invested in any particular form of government; only that it is genuinely invested somewhere, or else government - and human life as a whole - will reveal itself in its most repulsive and reductive form. It is thus clear that conservatism does not sanction any form of established power. In particular, it would not sanction a form of established power incapable of sustaining public affection, as was the case in eastern Europe during the Cold War. There the Communist party was the only authority and only power, running every aspect of society through a series of front institutions in whose legitimacy and independence few believed.

Even if the eastern European dictatorships had commanded more public affection than they did, and however long lasting and successful they had been, they would not in any case have been acceptable to conservatives. Following Burke, conservatives put considerable stress on the necessity for autonomous institutions and sources of power, so as to balance state power and to mitigate the consequences of centralized ignorance. It is here that there is an overlap between conservatism and the classical liberalism of Adam Smith and his followers (see Liberalism §2; Smith, A. §4).

5 Conservatism and the market

Adam Smith's model of the free market is based on the assumption that the most efficient way of organizing an economy is one that has evolved without planning, and which runs best without interference from the centre. It is impossible to know how agents in a large society will act economically before they actually engage in market transactions. If there are many entrepreneurs and producers there is a greater likelihood of real needs and desires being met than if there is only one agency of production. Those producers who, by a combination of luck and design, manage to tap into actual demand will benefit; and so will consumers who, from the range of goods and services offered, will pick those which satisfy their wants at the cheapest price. In Smith's view, those of us in capitalist societies have stumbled on a system in which producers and consumers are enabled to benefit each other, and which will produce both a measure of economic stability and productive drive, even though each individual in the system acts only on his or her private desires (to maximize profit, to purchase the best at the cheapest price). The system will, however, break down if central agencies attempt to restrict what is, in effect, the flow of information to and from producers and consumers, and so disturb its natural albeit unplanned tendency towards equilibrium.

Both the evolutionary aspect and the anti-statism of Smithian economic theory is likely to appeal to conservatives, who are also aware of the limitations of planning. But a true conservative would never, as Hayek did, utter any general edict against planning, even planning of a centralized sort (see Hayek, F.A. von §3). Making anti-planning into a principle or economic liberalism into an ideology offends the pragmatic and sceptical temper of the true conservative, who could cheerfully admit a role for state planning and even for economic interference were such things necessary, as in the modern world they no doubt are in order to make one's country, as Burke would have it, 'lovely' to one's compatriots. As Michael Oakeshott put it:

> A plan to resist all planning may be better than its opposite, but it belongs to the same style of politics. And only in a society already deeply infected with Rationalism will the conversion of the traditional resources of resistance to the tyranny of Rationalism into a self-conscious ideology be considered a strengthening of those resources. It seems that now, in order to participate in politics and expect a hearing, it is necessary to have, in the strict sense, a doctrine; not to have a doctrine appears frivolous, even disreputable. And the sanctity, which in some societies was the property of a politics piously attached to traditional ways, has now come to belong exclusively to Rationalist politics.
>
> (Oakeshott 1962: 21-2)

For Oakeshott, conservatism is to be contrasted principally with Rationalism, which is the belief that politics is to be understood in terms of getting things done, of achieving ends laid down by predetermined blueprints. For Oakeshott, this approach embodies the dual fallacy of thinking first that ends can be laid down independently of the means one consider appropriate in the light of one's way of life, and second, of failing to appreciate that the most important things in life are ends in themselves, worth having and cherishing for what they are, not for what they enable one to achieve. One of the deformations of the modern world is the introduction of methods and ideas appropriate to instrumental rationality into areas where they are inappropriate, such as the law, education and the arts. Above all, the introduction of means-end thinking is undesirable in connection with the state itself, not least because we have no choice in regard to membership of it (see Oakeshott, M.J. §3).

For Oakeshottian conservatism, politics should be seen in terms of a people elaborating ways of living together in the light of its particular history and traditions, and because of the compulsion involved, governmental and political intervention should be as light as possible. Political activity should not be seen as an enterprise driven by the pursuit of universally determinable extrinsic goals, such as equality or the elimination of poverty. To those who think otherwise, it might be pointed out that even if what is said is couched in the abstract language of universal rights, what will at any given point count as equality or the elimination of poverty will depend very much on the thought and practice of the time (see Tradition and traditionalism).

Much of the transformation of politics into enterprise and instrument has occurred because of collectivist approaches to institutions; much has happened because of market-led thinking. It is a moot point as to whether classical liberals would have favoured the prevalent instrumentalist view of politics and institutions, which Oakeshott criticizes. Adam Smith himself believed the weakening of traditional allegiances to be one of the 'inconveniences' of market societies. He, along with more recent neo-liberals such as Hayek, also recognized the need for a strong, if minimal, state to enforce contracts and to provide the social stability within which the market can operate. Thus, the supposed conflict between economic liberalism and conservatism often cited as a criticism of Mrs Thatcher's project of a strong national state combined with a free market does not appear as such in the works of her mentors. Nevertheless, from the conservative point of view economic liberalism certainly needs supplementation if it is to provide an adequate account either of man or of society.

6 Hegel

At this point, conservatives will point to a need to generalize de Maistre's criticism of social contract theories of the state. The social contract is a nonsense because it has to assume that what it purports to set up already exists. But equally nonsensical is the idea of the abstract ahistorical individual rational agents envisaged in so much Kantian and post-Kantian moral philosophy, the self who (in Rawls' terms) is prior to the ends affirmed by it. As F.H. Bradley put it: 'the man into whose essence his community with others does not enter, does not include relations to others in his very being, is a fiction' (1876 (1927): 168). Here Bradley is echoing Hegel, who in turn is drawing on de Maistre and Burke and, more remotely, Aristotle (see Bradley, F.H.).

For Aristotle, man is, by nature, a political animal, bound by his very identity to family and city. In Hegel’s hands this involves recognition that we are born into a network of relationships and obligations, which we did not choose, and to which we owe indissoluble debts of piety. Our sense of self arises first in the reciprocity involved in our natural family, and it is then extended into the public world of mutual but impersonal recognition through which we escape the war of all against all. (For Hegel this is civil society, in which is resolved the life and death struggle of master and slave.) As our identity as self-conscious agents is only made possible by this social network, in being disloyal to it, we are disloyal to the very self whose emancipation we vainly seek in our disloyalty. But while it is through the family and the institutions of civil society that we all derive our initial identity and the possibility of developing a life and career in the public world, they in turn owe their existence and wellbeing to an all-protecting state (Hegel 1833) (see Hegel, G.W.F. §8).

Against the liberals, then, Hegel will insist on the priority of the state and its independence from individual, family and the institutions of civil society. Like Burke and de Maistre, he believes that order, family and civil society will collapse without the state, and without cultivation of habits of obedience to the state (and, particularly, of loyalty to the sovereign). On the other hand, collectivists, who include individual, family and civil society within the state, forget the state’s *raison d’être* and the way its role and the respect it warrants demand that it should be above particularity, faction and caprice. While Hegel was not a democrat, neither was he a totalitarian. He argued for the personal, and no doubt in practice unequal, accumulation of capital as a means to protect individual and family and to assure generational continuity; for the welfare state as a means of ensuring cooperation from subjects; and for autonomous institutions against the might of the executive. In this intimate and delicate balancing of individual freedom and state power, the insistence on both authority and on checks and balances, Hegel escapes the one-sidedness of liberals and collectivists alike, and at the same time lays claim to be regarded as the most subtle and far-reaching of conservative thinkers.

7 An assessment of conservatism

While a fully expounded Hegelian conservatism contains within itself the resources to answer many of the objections which can be raised to a theoretical conservatism, in practice those who see their actual society riddled with defects and inadequacies are likely to be impatient with the temperamental resistance to change which they find in Burke and his followers in real politics. Furthermore, the stress on human ignorance and on the wisdom of generations - which would not get a hearing in other fields of endeavour - are too often simply an evasion of the hard task of shaking ourselves out of our complacency and working hard to do things better. To its critics, when faced with what they see as its blindness to pressing human problems, conservatism looks like little more than lazy self-interest bolstered by rhetoric and obscurantism.

It is certainly true that conservatism, in certain circumstances, will tend to uphold hierarchies and social distinctions. Conservatism is certainly not committed to egalitarianism, believing as it does in the right of individuals to make their own way through life and to benefit or not accordingly. Freedom of this sort is an important concomitant of the belief common to conservatives and classical liberals alike that individuals rather than the state are the best judges of their own interests. It is also an aspect of the traditional virtues of self-reliance and responsibility to family, qualities which conservatives also stress. But while not committed to equality, as we saw with Hegel, and in contrast to *laissez-faire* liberalism there is nothing in conservatism *per se* to rule out welfarism up to a point. It may indeed be a required part of life in a society that has any claim to be a community deserving of the allegiance of its members, something which is of great importance to conservatives.

None the less, there is no denying that in so far as it does not aim at egalitarianism and regards the means necessary to produce such an aim as likely to cause far more harm than good (and less equality, if the experiences of the French and Russian revolutions are any guide), conservatism does sometimes uphold hierarchies or inequalities. But it must be emphasized that it does not, and will not, uphold *all* or *any* hierarchies. In particular, precisely because of its belief in political ignorance and the imperfectibility of men, especially rulers, conservatism will not uphold totalitarian hierarchies (see Totalitarianism). It will also, in Aristotelian mode, assert that in matters of morality and politics, in contrast to natural science, there are no special experts. Rather, in these areas the experience of the whole of mankind over generations is the main source of moral knowledge, against which should be balanced the pretensions of particular groups of people to moral expertise, however clever they are in their own fields.

As already stated, conservatism should not be identified with blind or dogmatic reaction. Far from being opposed to reform, the principle of reform is central to conservatism, and follows from its scepticism and empiricism regarding human affairs. If obvious abuses exist, then principles and policies must be modified to accommodate improvement. But the modifications must be suitably cautious and their effects monitored. Ancient institutions should be treated with respect, because they embody a kind of tacit wisdom. Because of their ignorance about the way things work and the effects of change, conservatives, although open to reform, are wary of large-scale upsets of what is working reasonably well. And they tend to prefer the tolerable, and even the tolerably bad, to the promise of an unknown and untried good.

Conservatives have no principled objection to a welfare state, providing it allows plenty of room for autonomous institutions and initiatives. But they are more aware than their socialist opponents of the tendency for welfare systems to stifle initiative and self-reliance, and are also more sensitive to bureaucratic bossiness and the inborn tendency of bureaucracies to serve themselves first and their clients second.

Perhaps in the end the issue between conservatives and their political opponents, whether liberal or socialist, is an empirical one. Against socialists, conservatives argue, plausibly enough, that societies with autonomous institutions and unplanned economies have done rather better for all than those with centralized planning and large bureaucracies. Against classical liberals, conservatives will point to the constant splintering of a society which encourages too much choice and diversity, and to the need in a coherent society for a sense of tradition and a deference to generally recognized authority. If these points can be adequately argued, conservatives can rebut the argument commonly made against them that in their resistance to change they are motivated primarily by self-interest. And if this form of argument seems to make the choice for conservatism rest in part on empirical considerations, this will not unsettle conservatives, for their claim is that their position is rooted in a better understanding of human nature and human history than those of their opponents.

See also: Human nature

ANTHONY O’HEAR

References and further reading

Aristotle (*c*.mid 4th century BC) *Nicomachean Ethics* and *Politics*, in *The Works of Aristotle Translated into English*, ed. W.D. Ross, Oxford: Oxford University Press, 12 vols, 1908-52. (Important here because of the stress on the importance of experience in practical wisdom. Referred to in §2.)

Berlin, I. (1990) ’Joseph de Maistre and the Origins of Fascism’ in *The Crooked Timber of Humanity*, ed. H. Hardy, London: John Murray.(A fascinating study of de Maistre.)

Bradley, F.H. (1876) *Ethical Studies*, Oxford: Oxford University Press, 1927.(Stresses ‘my station and its duties’, and the plain man’s view of right and wrong, against theory-based moral systems; but does not believe morality is completely defined by its conservative underpinnings. Referred to in §6.)

Burke, E. (1790) *Reflections on the Revolution in France*, Harmondsworth: Penguin, 1986. (Classic statement of a conservative position. Referred to in §4.)

Grant, R.A.D. (1992) ’The Politics of Equilibrium’, *Inquiry* 35: 423-46.(Contains an excellent rebuttal of the view that conservatism would conserve *anything*.)

Hayek, F.A. (1988) *The Fatal Conceit*, London: Routledge.(Defends the market in economics and conservatism in morals.)

Hegel, G.W.F. (1833) *The Philosophy of Right*, ed. A.W. Word, Cambridge: Cambridge University Press, 1991. (The most succinct exposition of Hegel’s political philosophy and that most relevant to conservatism. Referred to in §6.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.(In its stress on the limited role of the state and on the need for a state, the text is a source of much conservative thinking. Referred to in §3.)

Honderich, T. (1990) *Conservatism*, London: Penguin.(A blistering attack on conservatism, containing all stock objections, and more.)

Maistre, J. de (1797) *Considerations on France*, trans. R.A. Lebrun, Cambridge: Cambridge University Press. (Shares many conservative themes with Burke. Referred to in §4.)

Oakeshott, M. (1962) *Rationalism in Politics*, London: Methuen.(The best modern version of a Burkean position. Referred to in §5.)

Plato (*c*.380-347 BC) *Republic*, *Phaedrus* and *Laws*, in *The Dialogues of Plato*, trans. B. Jowett, Oxford: Oxford

University Press, 4th edn revised, 1953.(In his assessment of the complexity of human nature, Plato anticipates conservative thinking. Referred to in §2.)

Quinton, A. (1978) *The Politics of Imperfection*, London: Faber.(An elegant defence of conservatism.)

Scruton, R. (1980) *The Meaning of Conservatism*, London: Penguin.(A strong exposition of conservatism.)

Tannsjo, T. (1990) *Conservatism for our Time*, London: Routledge.(Argues that in certain circumstances conservatives should defend a communist regime. Attacked by Grant (1992).)

Constant de Rebecque, Henri-Benjamin (1767-1830)

Benjamin Constant combined the activities of a religious historian, autobiographer and novelist with a career as a political theorist and politician. Constant's intellectual outlook was shaped by French Enlightenment thought and two years spent at Edinburgh University in 1783-5 added experience of observing the British government and constitution at work. Through all of Constant's writings runs a consistent theme: the necessity of safeguarding the freedom of the individual in modern society. At the end of his life he summed up his liberalism thus: 'Freedom in all things, in religion, philosophy, literature, industry and politics. And by freedom I mean the triumph of the individual both over an authority that would wish to govern by despotic means and over the masses who would claim the right to make a minority subservient to a majority' (1957: 835). Constant's political activity and his writings, which some consider prophetic of the growth of modern totalitarian regimes, have been influential in the development of liberal thought in Europe and the USA.

1 Political activity and thought

The thought of Henri-Benjamin Constant de Rebecque, known as Benjamin Constant (born Lausanne, Switzerland, died Paris) needs to be seen against the background both of his complex and subtle mind - ever sceptical and alive to paradox and ambiguity in all human transactions - and his long observation of the abuse of political power during one of the most turbulent periods of French history. Even during his childhood, Constant had experienced political oppression in Lausanne under the rule of the Bernese, and this instilled in him a lifelong libertarianism. This was reinforced by six years as chamberlain at the court of the Duke of Brunswick, during which time his sympathies veered towards Jacobinism. In Paris after *Thermidor* - and under the influence of Germaine de Staël, with whom he had a long and stormy liaison - Constant tried to bring about the establishment of a moderate republican government in France, using pamphlets and, subsequently, his role as a member of Bonaparte's Tribunate. This remained his ideal during his long period of opposition to the despotism of Napoleon. After twelve years of enforced political inactivity and self-imposed exile in Germany, he sought a political career during the first Restoration and then, notoriously, collaborated with Napoleon during the Hundred Days of 1815. Despite criticism of his action, Constant maintained that the personality of a head of state was of little importance; the ex-dictator's declared intention to act henceforth as a constitutional monarch made him preferable to a restored, autocratic and repressive Bourbon. During his last years, Constant became unofficial leader of left-wing liberals in the French Assembly and one of its greatest orators, campaigning for such causes as press freedom and the abolition of slavery. The July Revolution of 1830, in which he was an important figure, saw the (short-lived) triumph of many of Constant's ideals.

Constant's political thought therefore was not shaped by abstract speculation alone. It developed out of opposition to the cult of antiquity prevalent in eighteenth-century France, and in particular to Jean-Jacques Rousseau's idealization of the small Greek city state. For Constant, Rousseau's notion of the supremacy of the general will over the will of the individual as set out in his *Du Contrat social (The Social Contract)* (1762) and his notorious ban on atheism led inexorably to a tyranny such as had been seen during Robespierre's Terror (see General will §2; Rousseau, J.-J. §3). France had experienced a police state that felt justified in interfering in the lives of individual citizens and even in attempting to invade the inner sanctum of private thoughts and beliefs. Such a view of the individual's relationship with society was regressive and anachronistic, ignoring more than 2,000 years of human development. What may have been appropriate for the Greek *polis* in antiquity was certainly no longer so in the modern world, where personal independence and privacy had become essential in large commerce-based societies. These views, set out in *De la Liberté des anciens comparée à celle des modernes (Freedom in Antiquity Compared With Freedom in the Modern Age)* (1820), drew a clear distinction between freedom in the ancient world and modern freedom. In Constant's memory were Robespierre's communitarian rhetoric and exaltation of Spartan virtues, which demonstrated the dangers of re-establishing a form of society and social relationships where there were no clear boundaries between the public and private spheres. Where the supposed demands of the community were allowed precedence over individual rights, anything might be permissible in the name of the general good. He felt the same concern when it came to scientifically organized utopias of the future, such as that of the Saint-Simonians, which promised greater productivity and human happiness (see Saint-Simon, C. de; Utopianism).

Freedom in modern times demanded clearly defined areas where government interference was not allowed. Not only did the essential right to privacy of the individual need to be protected, due legal processes had to be scrupulously respected and preserved at all times as a bulwark for the individual against arbitrary persecution by an authoritarian government. Yet the Constant who steadfastly defended individual liberty and privacy, so abused by the Jacobins during the Revolution, nevertheless held throughout his career to the ideals of 1789: egalitarianism (as far as the rights of individuals, nations and races were concerned), religious toleration (despite the bitter anticlericalism of his own youth), absolute freedom of speech and of the press, and representative institutions which would be a vehicle for the exercise of popular sovereignty.

Constant's opposition to the usurper and dictator Napoleon, so powerfully expressed in *De l'Esprit de conquête et de l'usurpation (On the Spirit of Conquest and Usurpation)* (1814), was based on the same philosophical grounds. Caesarism, that is, rule by a military dictatorship that consciously echoed the republics of Graeco-Roman antiquity, was as much a brutalizing anachronism as Robespierre's Terror. Wars of conquest by a tyrannical demagogue were the norm in antiquity, but were out of tune with a modern world based on industry, banking and commerce, where peace, tolerance and justice were the order of the day. The *res publica* of antiquity had been replaced as the focus of importance by private pursuits and needs. 'The least government [is] the best government', Constant had written under the Directory, anticipating the thinking of modern liberals for whom the legitimate spheres of government activity should be limited to national defence and the maintenance of law and order (see Liberalism §2).

Individual citizens with their private concerns and activities and their inalienable right to own property, to enjoy personal security, and to have freedom of self-expression, of belief and of movement should be expected to play no greater role in the running of the state than that of casting their vote and choosing their representative in a governing assembly. Gone was the claustrophobic world of the small city state so beloved of Rousseau, where every citizen played a direct role in popular democracy and people could legitimately involve themselves in everyone else's views and business. In its place Constant proposed a constitutional monarchy on the British model, with an essentially figurehead monarch and a bicameral assembly comprising an upper hereditary body, limited in its powers and ensuring stability, continuity and moderation, and an elected lower chamber reflecting the views and life of a society in constant evolution.

2 Religious thought

In a well-known fragment of an unpublished Preface to his novel *Adolphe* (1816), Constant wrote that 'all things are connected in nature' and suggested a link between 'religious belief' and 'enthusiasm for liberty'. Behind his lifelong work on the history of world religions, represented by *De la Religion considérée dans sa source, ses formes et ses développements (Religion Considered in its Source, its Forms and its Development)* (1824-31) and *Du Polythéisme romain (Roman Polytheism)* (1833), lay Constant's conviction that two related needs were fundamental to human beings: the need for freedom and the need to believe. For Constant 'le sentiment religieux' was not the illusion excoriated by D'Holbach, Helvétius and other Enlightenment philosophers (see Enlightenment, Continental). It was a real and permanent facet of human nature. Societies had evolved through the stages of theocracy, slavery, feudalism, absolute monarchy to rule by a privileged aristocracy; and at each stage the permanent human longing for freedom had brought resistance to oppression and the casting aside of outworn forms of government. Similarly, men and women of all historical periods had in them a religious sense or feeling, an inner need to rise above mere egoism that expressed itself in self-sacrifice and self-denial. All cultures had seen a constant struggle between the religious feeling of individuals, which was a living thing and needed to evolve and change, and the static outward shape imposed on it by a priestly caste in terms of forms of worship and dogma.

The parallel is clear: as an optimist and believer in human perfectibility, Constant believes that there is a law of movement in history in both in politics and religion that cannot be resisted. The need for liberty inherent in human nature will always ensure that those forms of government or of religious belief which are fixed, stagnant and unchanging are replaced by forms better adapted to the needs of the historical moment. In 'De la Perfectibilité de l'espèce humaine' (On the Perfectibility of the Human Race) (1829), it is unlimited human perfectibility that gives meaning to life - and indeed a form of survival beyond death in the progressive development of intellectual and spiritual qualities from generation to generation. Religious feeling had developed through history from primitive polytheism through theism to Christianity and had reached its highest form in modern Protestantism -

open-minded, tolerant and pluralistic. Yet every form of belief, like every form of government, was a provisional one and would ultimately be outgrown. Political wisdom required an eventual change in institutions to reflect the evolution of society (the ultimate goal being a society that allowed the greatest personal freedom), and likewise a tolerant state would enable religious feeling to be expressed in a form appropriate to a new age.

DENNIS WOOD

List of works

Constant, B. (1957) *Œuvres*, ed. A. Roulin and C. Roth, Paris: Gallimard (Bibliothèque de la Pléiade).(Offers the most comprehensive modern selection of Constant's political and religious writings; another edition of *Œuvres complètes* is in progress (Tübingen: Max Niemeyer, 1993-) but will take many years to complete.)

Constant, B. (1806) *Principes de politique applicables à tous les gouvernements (Political Principles Applicable to All Governments)*, ed. E. Hofmann, Geneva: Droz, 1980.(A summary of Constant's political beliefs.)

Constant, B. (1814) *De l'Esprit de conquête et de l'usurpation (On the Spirit of Conquest and Usurpation)*, Hahn: Hanover; repr. in E. Harpaz (ed.), Paris: Flammarion, 1986; trans. and ed. H. Byrne Lippmann, *Prophecy from the Past: Benjamin Constant on Conquest and Usurpation*, New York: Reynal & Hitchcock, 1941.(An analysis and indictment of military dictatorships inspired by Napoleon's example.)

Constant, B. (1816) *Adolphe: anecdote trouvée dans les papiers d'un inconnu (Adolphe: An Anecdote Found Among the Papers of an Unknown Person)*, London: Colburn and Paris: Tröttel et Wurtz; ed. F. Tilkin, in *Benjamin Constant: Écrits littéraires (1800-1813) (Benjamin Constant: Literary Works (1800-1813))*, Tübingen: Max Niemeyer, 1995, 81-203; trans. with introduction by L.W. Tancock, London: Penguin, 1964. (A first-person novel of close self-analysis centring on the protagonists's inability to free himself from a love relationship.)

Constant, B. (1820) *De la Liberté des anciens comparée à celle des modernes (Freedom in Antiquity Compared With Freedom in the Modern Age)*, vol. IV, in *Cours de politique constitutionnelle* , Paris: Béchet aîné and Rouen: Béchet fils; repr. in the collection, *Benjamin Constant: De la Liberté chez les modernes*, ed. M. Gauchet, Paris: Le Livre de poche, 1980.(Contrasts the notion of freedom in the ancient *polis* with the modern need for individual freedom and privacy.)

Constant, B. (1820-22) *Mémoires sur les Cent-Jours (Memoirs concerning the Hundred Days)*, ed. K. Kloocke and A. Cabanis, Tübingen: Max Niemeyer, 1993.(An explanation of Constant's conduct in rallying to Napoleon during the Hundred Days in 1815.)

Constant, B. (1824-31) *De la Religion considérée dans sa source, ses formes et ses développements (Religion Considered in its Source, its Forms and its Development)*, Paris: Bossange père *et al.*; repr. in P. Deguise (ed.), Lausanne: Bibliothèque romande, 1971 (extracts).(A study of the development of religious feeling through human history.)

Constant, B. (1829a) *Mélanges de littérature et de politique*, Paris: Pichon et Didier.; trans., with introductory and critical notices, by G. Ripley, *Philosophical Miscellanies translated from the French of Cousin, Jouffroy, and B. Constant*, Boston, MA: Hilliard, Gray, and Co., 1838, 2 vols.(A collection of essays on literary and political themes published shortly before Constant's death.)

Constant, B. (1829b) 'De la Perfectibilité de l'espèce humaine' (On the Perfectibility of the Human Race), in *Mélanges de littérature et de politique*, Paris: Pichon et Didier; repr. in P. Deguise (ed.), Lausanne: L'Age d'Homme, 1967.(Good all-round text.)

Constant, B. (1833) *Du Polythéisme romain, considéré dans ses rapports avec la philosophie grecque et la religion chrétienne (Roman Polytheism Considered in Relation to Greek Philosophy and the Christian Religion)*, ed. M.-J. Matter, Paris: Béchet aîné, 2 vols.(A study of Roman belief in many gods that has links with Constant's *De la Religion*.)

Constant, B. (1988) *Political Writings*, ed. B. Fontana, Cambridge: Cambridge University Press.(The only available modern selection in translation.)

References and further reading

Bastid, P. (1966) *Benjamin Constant et sa doctrine (The Political Thought of Benjamin Constant)*, Paris: Armand Colin, 2 vols.(A highly detailed account of Constant's political thought.)

Berlin, I. (1969) *Four Essays on Liberty*, Oxford: Oxford University Press.(Constant as 'the most eloquent of all defenders of freedom and privacy'.)

Deguise, P. (1966) *Benjamin Constant méconnu: Le livre 'De la Religion' (A Neglected Work by Benjamin Constant: On Religion)*, Geneva: Droz.(The standard text covering Constant's work on religion.)

Dodge, G.H. (1980) *Benjamin Constant's Philosophy of Liberalism. A Study in Politics and Religion*, Chapel Hill, NC: University of North Carolina Press.(A clear and accessible account of Constant's work in these related areas.)

Fontana, B. (1991) *Benjamin Constant and the Post-Revolutionary Mind*, New Haven, CT, and London: Yale University Press.(An account of Constant's relation to predecessors such as Montesquieu, Rousseau and Sieyès, and an assessment of his originality.)

Holmes, S. (1984) *Benjamin Constant and the Making of Modern Liberalism*, New Haven, CT, and London: Yale University Press.(The standard exegesis of Constant's contribution to modern liberal thought.)

Constitutionalism

Constitutionalism comprises a set of ideas, principles and rules, all of which deal with the question of how to develop a political system which excludes as far as possible the chance of arbitrary rule. While according to one of the classic sources of constitutionalism, article sixteen of the 1789 French Declaration of the Rights of Man and of the Citizen, 'any society in which rights are not guaranteed, or in which the separation of powers is not defined, has no constitution', the scope of constitutional principles is in fact broader. In addition to these two defining principles, the following are essential: popular sovereignty; the rule of law; rules about the selection of powerholders and about their accountability to the ruled; and principles about the making, unmaking, revision, interpretation and enforcement of a constitution. Despite close affiliations, constitutionalism and democracy are not the same. Whereas democracy is an institutional device which realizes the right of the people to govern themselves, constitutionalism aims to establish institutional restraints on the power of the rulers, even if they are popularly elected and legitimized. Constitutionalism embodies the self-rationalizing and self-restraining principles of popular government.

1 The concept, its historical roots and main variants

Constitutionalism in the modern sense of the term embodies the philosophical and juridical response to man's quest for political freedom. It encompasses the values, principles, reasoning, institutional devices and procedures which shape the idea of an institutional framework by which political freedom is secured. This framework is called constitutional government, and the instrument by which it is established is the constitution. The components which make up constitutional government vary considerably in different countries, to the point that what is regarded as a core element of constitutionalism in one country is considered incompatible with constitutionalism in another. However, there are a small number of properties that over a long period have been taken to define the very essence of constitutionalism, and about which there is a broad consensus.

The following precepts are basic to modern constitutionalism. First, the claim that the arbitrary will of the political ruler(s) must be submitted to rules. A government under law, a rule that governs the governors or, in the most pointed version, the rule of law and not of men, are but different ways of expressing the essence of modern constitutionalism. The term 'rule' means 'legal rule'; however, in particular cases extra-legal (political, moral) rules may also acquire constitutional quality if they are accepted as binding by both the ruled and the rulers (see Rule of law (Rechtsstaat) §1). Second, constitutionalism implies the principle of popular sovereignty. Although in nineteenth-century Germany the 'graciously granted' rules whereby sovereign princes affirmed certain limitations on their sovereign power were called constitutions, the concept of constitutionalism requires the constituent, or constitution-making power of the people. This follows from the premise that in modern societies the 'people' is the ultimate source of sovereign power; since constitutions shape the sovereign power, they can only be ordained by the sovereign (see Sovereignty). The third essential element of constitutionalism is the idea of limited government. This means that both the purposes of government, and the instruments available to power-holders to realize those purposes, are delineated - something implicit in the term 'constituted power'. This entails, as the most important institutional consequence for the form constitutional government is to take, some kind of separation of powers. Furthermore, constitutional government calls for rules which define the accountability of the rulers to the citizenry. Finally, constitutionalism requires that changes to the constitutional rules must be effected according to the rules of the constitution itself. Based on these minimum requirements, constitutional documents normally extend over three main issues: individual rights; the structure of government; and rules about the revision of the constitution.

These characteristics distinguish modern from ancient and medieval constitutionalism, neither of which knew the concept of sovereign power or that of the natural freedom of the individual. The demise of ecclesiastical power, and the resulting disruption of the balance between secular and ecclesiastical authority characteristic of medieval society, gave birth to the idea of supreme power originating from a single sovereign source, paralleled by the idea of natural, equally unlimited, individual freedom. As a new political structure based on the monistic character of power was created, the word 'constitution' was progressively used to counteract the unlimited power of the sovereign prince. Gradually it became associated with the concept of political freedom. Originally appearing in English political discourse at the beginning of the seventeenth century, and signifying the empirical state of the body politic by analogy with the human body, 'constitution' had acquired its modern and narrower meaning by the

end of that century. The term no longer referred to the social structure of a particular country, but to its government as formed by the law of the land. The locus of power and the domestication of power through law became the main point of interest. Moreover, during the seventeenth century, the age when the modern absolutist territorial state was consolidated, it became common to use the word 'constitution' in the sense of an eternal and supreme law.

Constitutional government has turned out to be the only viable alternative to either despotism or anarchy, the two main threats to political freedom. This was essentially Montesquieu's perception that among all states England was the only nation 'whose constitution has political liberty for its direct purpose'. This statement referred to the final triumph of constitutionalism in England's Glorious Revolution of 1688 which imported the three elements characteristic of what has become known as the Westminster model of constitutionalism; namely, parliamentary sovereignty; the rule of law; and the operation of so-called constitutional conventions. The latter element refers to legally non-binding political and moral conventions which ensure that the conduct of parliament (which by virtue of its very supremacy cannot be subject to laws) conforms to the sovereignty of the people, namely, to the will of the majority of the electors (see Montesquieu, C. §3).

Although England is undoubtedly the homeland of modern constitutionalism, since the last quarter of the eighteenth century its notion of constitutionalism and particular form of constitutional government no longer embody the definitive model. In its US colonies the English model became itself the object of a constitutional revolution: the Constitution of the United States was drawn up in 1787, and became effective in 1789, the first year of the French Revolution. Quite different understandings of constitutional government came to prevail in the USA and France, greatly expanding the idea of constitutionalism. While the essentials - the submission of political rule to rules, popular sovereignty, limited government and the accountability of the rulers to the ruled - remained untouched, the ensuing institutions stood in clear, even hostile, opposition to those of the Westminster model. The Americans, who felt that the English version of constitutionalism did not protect them against the arbitrariness of parliament, reinvented, although with a quite different meaning, the medieval distinction between constitution and ordinary laws, a move inconsistent with the notion of parliamentary sovereignty. From this they drew three ideas. First, the constitution is superior to government and ordinary law, hence it creates, defines and limits government, including the legislature itself. Placing the constitution at the top of a legal hierarchy entails, second, that the constitution as the ultimate source of authority of political power be fixed, ordained in an act of extraordinary legislation and written in a document. This in turn generated a third element, namely the precept that the superiority of the constitution over ordinary laws includes the authority of the courts to invalidate laws incompatible with the constitution ('judicial review'). One of the most significant implications of this is the protection of constitutional rights not just against the executive, but against the legislature as well.

In its turn, the French Revolution, in which the US notion of the constitution as a written document enacted by a constituent power and superior to ordinary laws was adopted, gave birth to a notion of rights which stood in stark contrast to the English notion. The latter is based on the inseparable connection between a right and the institutional remedies against its violation so that, in the words of Dicey, 'in England the law of the constitution is little else than a generalisation of rights which the Courts secure to individuals' ([1915] 1982: 119). The French model, in contrast, proclaimed the rights of the individuals as being assured by, and hence dependent upon, the constitution, so that the whole burden of guaranteeing rights was placed on the constitution.

This notion of rights is part of a more comprehensive set of ideas which the French Revolution has contributed to the conceptual development of constitutionalism. Its main characteristic is the understanding of the constitution as the written embodiment of universally valid principles and doctrinal truths which render it a kind of rational plan for an utterly rational and smoothly functioning polity. This constitutional standard of complete rationality entails an incessant quest for the perfection of the polity and has inserted a spirit of political dynamism into the concept of constitutionalism which is lacking in both the English and the US versions. Consequently, in the French variant legal mechanisms to enforce the constitution are less important than political means. Thus what seems paradoxical at first glance, namely that the truly revolutionary and philosophically extremely influential Declaration of the Rights of Man and of the Citizen of 1789 has played only a minor role in the constitutional life of France, is quite understandable. Constitutional rights embody political claims and are primarily satisfied in the sphere of politics. Likewise, French constitutions have never allowed judicial review. They have mostly relied instead on the operation of the political process in general and parliament in particular to enforce the constitution. On the other

hand, the French constitutional model's responsiveness to the political process has facilitated the incorporation of substantive policy goals (like social rights or state goals) in the constitution. Such a purposive constitutionalism is alien to US constitutional reasoning, which views it as a potential risk to the legal normativity of the constitution. Moreover, given the origin of US constitutionalism, Americans trust the courts more than parliaments.

2 Problems, tensions and difficulties

Since the basic idea of constitutionalism is essentially the substitution of reason and rule for the arbitrary will of man, a pivotal element of constitutionalism is the conceptual clarification of the binding force of constitutional rules. If the concept of popular sovereignty includes the supreme power of the people to establish whatever political order it pleases, and if at the same time constitutionalism embodies the idea of forming, and hence restraining, sovereign power, do these not represent inconsistent principles? Moreover, is not the term 'constitutional democracy' an oxymoron? Traditional answers to these questions affirm that indeed there is a tension between constitutionalism and democracy, but they regard constitutional government as a justifiable restraint on democratic rule in order to protect minorities against the will of the majority (see Democracy §3). More recent analyses of constitutionalism emphasize the dialectics of self-binding and point to the fact that constitutions are not only devices against majorities, but serve equally to protect majorities themselves from myopia, moral weakness and a lack of self-restraint. In this understanding both democracy and constitutionalism share the goal of limiting the risks of political liberty and of politics generally, but employ different means: the former protects the individual right to an equal share in the political process, the latter 'lowers the stakes of politics' by enabling the sovereign to exercise governmental self-restraint through institutional devices (Murphy 1993). On this view constitutions are not primarily disabling devices; rather, they serve as a support for democratic self-government in that they enable the people to preserve the conditions of political freedom and democracy, better to recognize the options open to democratic choice, to reduce the risk of irreversible mistakes, to develop the capacity for self-correction and to enhance their democratic capabilities through learning.

There is abundant evidence that the strategy of self-binding need not weaken an agent's capacity but frequently increases their options. The prime example is Ulysses, who had himself bound to the mast in order to be able to listen to the sirens without losing his life. Another instance is the prohibition of contracts to sell oneself into slavery: this does not restrict, but promotes, freedom. Having the collapse of the Weimar constitution in mind, the drafters of the German Basic Law wanted to shield the German people from the temptation of once again embracing totalitarian rule, and thus definitively excluded the option of abolishing basic constitutional institutions through any kind of constitutional amendment. Likewise, the prohibition to print money imposed on the states in article 1, section 10, paragraph 1 of the US constitution can be understood as a mechanism of self-binding against the temptation to pursue policies which induce inflation (Holmes 1988).

3 Future developments

The main political impulse which has given constitutionalism worldwide recognition as the concept of a well-ordered polity has been the desire to 'tame the Leviathan' which was born in the religious struggles of the sixteenth and seventeenth centuries. Despite the manifold refinements which it has undergone, and notwithstanding a growing awareness of the aforementioned enabling functions of constitutions, constitutionalism is still widely associated with the purely negative functions of limited government. It is therefore often wrongly conflated with liberalism. This view does not take sufficiently into account that constitutions, although rarely containing solutions to particular substantive policy problems, can be credited with having resolved difficulties besides those that result from arbitrary state power. Thus, for example, the modern welfare state is the result of amalgamating the working-class demand for social justice with the idea of citizenship, which is a pivotal element of constitutionalism (see Citizenship §§1-2). Likewise, the inherent learning potential of constitutional devices has allowed constitutional states to cope more successfully with the challenges posed by the destabilizing technological, economic, social and cultural changes since the middle of the twentieth century than political systems which for one reason or another lacked constitutional arrangements.

However, approaching the end of the twentieth century, modern societies suffer from difficulties for which the traditional concept of constitutionalism has no answer. Economic inefficiency, civic apathy, a lack of concern for future generations, the seemingly irresistible decay of educational institutions and the continual worldwide impoverishment of urban life are just a few examples of phenomena which until now have hardly been touched

upon by constitutional reasoning, even though they are of utmost importance for the quality of a polity. These problems facing a modern polity, which go beyond the traditional goal of preventing the arbitrary exercise of political power increasingly have been addressed by a school of thought which invokes the idea of a ‘new constitutionalism’. This is characterized by concern for a good polity and hence completely revokes the basic assumption of traditional constitutionalism; namely, that for the sake of civil peace any striving after the common good and common values had to be excluded altogether from the realm of politics. Paradoxically, with this the ‘new constitutionalism’ returns to the Aristotelian point of departure in reasoning about constitutionalism; namely, to the idea of the *politeia* (see Aristotle §§27-28).

See also: Shōtoku Constitution

ULRICH K. PREUß

References and further reading

Dicey, A.V. (1915) *Introduction to the Study of the Law of the Constitution*, reprint of the 8th edn, Indianapolis, IN: Liberty Fund, 1982.(Classic treatise of English constitutional law with broad comparative views on French and US constitutionalism; expounds on the material in §1 of this entry.)

Elkin, S.L. and Soltan, K.E. (eds) (1993) *A New Constitutionalism. Designing Political Institutions for a Good Society*, Chicago, IL, and London: University of Chicago Press. (Thorough and multifaceted reflections about the possible meaning of a new constitutionalism and how it differs from traditional constitutionalism.)

Friedrich, C.J. (1974) *Limited Government. A Comparison*, Englewood Cliffs, NJ: Prentice Hall.(A lucid and comprehensive account of the history and theory of modern constitutionalism from a traditional perspective.)

Greenberg, D., Katz, S.N., Olivero, M.B. and Wheatley, S.C. (eds) (1993) *Constitutionalism and Democracy. Transitions in the Contemporary World*, New York and Oxford: Oxford University Press.(Good account of the role of constitutionalism not only in Europe, but also in other parts of the world such as Latin America and southern Asia.)

Grey, T.C. (1979) ‘Constitutionalism: An Analytic Framework’, in J.R. Pennock and J.W. Chapman (eds) *Constitutionalism*, New York: New York University Press.(Contains a conceptual framework for the analysis of constitutions along particular properties like textuality, quality as supreme law and amendment procedures.)

Grimm, D. (1991) *Die Zukunft der Verfassung (The Future of the Constitution)*, Frankfurt: Suhrkamp. (Contains a historical and analytic perspective on constitutionalism in the framework of social change and development.)

Holmes, S. (1988) ‘Precommitment and the Paradox of Democracy’, in J. Elster and R. Slagstad (eds) *Constitutionalism and Democracy*, Cambridge: Cambridge University Press.(Expansion of the topic of self-binding referred to in §2 of this entry.)

Kasper, G. (1986) ‘Constitutionalism’, in L.W. Levy, K.L. Karst and D.J. Mahoney (eds) *Encyclopedia of the American Constitution*, New York and London: Macmillan.(Contains an elaborate account of the institutions and functions of modern constitutions, largely referring to the US constitution.)

McIlwain, C.H. (1966) *Constitutionalism: Ancient and Modern*, Ithaca, NY: Cornell University Press. (Indispensable for the understanding of the historical emergence of modern constitutionalism; particular emphasis on England.)

Murphy, W.F. (1993) ‘Constitutions, Constitutionalism, and Democracy’, in D. Greenberg, S.N. Katz, M.B.Olivero and S.C. Wheatley (eds) *Constitutionalism and Democracy. Transitions in the Contemporary World*, New York and Oxford: Oxford University Press.(Illuminating conceptual distinctions with respect to the problems raised in §2 of this entry.)

Wood, G.S. (1969) *The Creation of the American Republic, 1776-1787*, New York: W.W. Norton. (Comprehensive and thoughtful narrative of the debates about the US constitution.)

Constructivism

Originally proposed by sociologists of science, constructivism or social constructivism is a view about the nature of scientific knowledge held by many philosophers of science. Constructivists maintain that scientific knowledge is made by scientists and not determined by the world. This makes constructivists antirealists. Constructivism here should not be confused with constructivism in mathematics or logic, although there are some similarities. Constructivism is more aptly compared with Berkeley's idealism.

Most constructivist research involves empirical study of a historical or a contemporary episode in science, with the aim of learning how scientists experiment and theorize. Constructivists try not to bias their case studies with presuppositions about how scientific research is directed. Thus their approach contrasts with approaches in philosophy of science that assume scientists are guided by a particular method. From their case studies, constructivists have concluded that scientific practice is not guided by any one set of methods. Thus constructivism is relativist or antirationalist.

There are two familiar (and related) criticisms of constructivism. First, since constructivists are self-avowed relativists, some philosophers argue that constructivism fails for the same reasons that relativism fails. But many philosophers of science note that relativism can be characterized in various ways and that versions of relativism can be useful in the interpretation of science. Therefore, constructivism's relativism does not by itself render it unacceptable. Second, constructivists are accused of believing that scientists literally 'make the world', in the way some make houses or cars. This is probably not the best way to understand constructivism. Rather, constructivism requires only the weaker thesis that scientific knowledge is 'produced' primarily by scientists and only to a lesser extent determined by fixed structures in the world. This interprets constructivism as a thesis about our access to the world via scientific representations. For example, constructivists claim that the way we represent the structure of DNA is a result of many interrelated scientific practices and is not dictated by some ultimate underlying structure of reality. Constructivist research provides important tools for epistemologists specializing in the study of scientific knowledge.

1 Constructivism and social constructivism

The term 'constructivism' refers to a cluster of related approaches to the study of science. All these approaches involve empirical studies of science emphasizing the social nature of scientific practice. Some constructivists draw the following theoretical conclusion from their empirical studies of science: namely, the truth or falsehood of scientific beliefs derives not from their relation to the world but from the social arrangements of scientists. Social constructivists maintain that social interests or factors in the micro-structure of scientific practice determine which beliefs are held to be true or false. Social constructivist case studies often focus on scientific controversies where resolution does not appear to depend on hard evidence or data, but on 'other factors'. Thus the claim that scientific knowledge is socially constituted results from the generalization of these other factors. This view is similar to a social-determinist view about belief formation. It is best exemplified by the Edinburgh School sociologists of science (see Bloor 1992) and Harry Collins' empirical programme of relativism (Collins and Pinch1993). Another constructivist view is that scientific facts are constituted by arrangements of scientists, and animate and inanimate objects. This view challenges social determinism on the grounds that society is no better understood than nature, and so claiming that to understand the study of nature one needs to understand society is premature. This view is shared by actor-network theorists such as Bruno Latour (1987), and reflexivists such as Steve Woolgar (1988) (see also Latour and Woolgar 1986). They develop the view in different ways, though we will refer to them both as constructivist.

Actor-network theorists, such as Latour, maintain that scientific knowledge results from the establishment of connections between objects, animals and humans. For example, our knowledge of the structure of DNA connects together several different scientific practices, instruments and techniques. Challenging the structure of DNA involves challenging an established set of connections between these domains. Since constructivist case studies aim to show how scientific activity leads to this connectivity, often the case studies have a historical dimension, starting at the time before a particular scientific fact was established.

The terms 'fact' and 'belief' are often conflated in constructivist literature, leading to some confusion. To avoid

further confusion we will define these terms as they are used here. Beliefs are propositions or sentences that are candidates for truth or falsehood. Individual scientists have various beliefs about their world. Facts are established pieces of scientific knowledge, as found in scientific textbooks. Though not the standard philosophical definition of 'fact', this definition will help in understanding constructivism.

Constructivism as discussed here is distinct from constructivism in mathematics or logic (see Constructivism in mathematics), and distinct from constructive empiricism in the philosophy of science. Constructivism shares antirealism with both these views and a kind of empiricism with the latter view. The antirealism in constructivism is expressed in the claim that the truth and falsity of scientific beliefs can be established independently of any evidence from the real world (Collins and Pinch 1993) (see Scientific realism and antirealism §4). The empiricism in constructivism is a methodological tenet: all studies of science should be empirical studies of scientific practice. Versions of constructivism, for example the reflexivist view, are more readily compared to some kinds of scepticism or idealism. These forms of constructivism, claiming that scientists have access to nothing other than representations, are reminiscent of the idealism of Berkeley (Woolgar 1988).

Constructivism is antirealist and relativist: constructivists reject the view that science discovers a determinate structure to reality and they reject the view that scientific knowledge is achieved by following one particular rational method. According to critics, the rejection of these views constitutes a fatal flaw in constructivism. Nonetheless most constructivists enthusiastically embrace the antirealist and relativist consequences of their position.

2 The development of constructivism

Constructivism has a long heritage in philosophy. However, its modern history consists of refinements to the view developed by the Edinburgh School in the sociology of knowledge.

The roots of constructivism are in the sociology of knowledge which can be traced as far back as Plato, who proposed a relationship between a citizen's knowledge and their place in society. Marx provided the starting point for modern sociology of knowledge when he claimed that ' it is not men's consciousness which determines their existence, but on the contrary their social existence which determines their consciousness' (Marx and Engels 1970: 51). Hence, to explain the genesis of human ideas Marx looked to social structures (see Marxist philosophy of science). The discipline addresses issues such as: what is the relationship between a society or social structure and what is known? Does society affect science or the knowledge of nature more or less than it does the knowledge of culture? What kinds of social force produce knowledge and what kind of analysis of society reveals these social forces?

The Edinburgh School of sociologists of scientific knowledge (Bloor 1992) are direct descendants of Marx. Bloor starts out from a critique of the Marxist sociologist of knowledge, Mannheim, who maintained that all knowledge, except for that generated by the physical sciences and mathematics, is socially determined. Bloor expands the sociology of knowledge to include the study of mathematics and scientific knowledge.

As well as responding to Mannheim, Edinburgh School constructivists are reacting to two other strands of thought. The first is American sociology of science, represented by Robert Merton (Merton 1973). Although Merton argues that the rise of science and its success results from social forces, he stops short of claiming any relation between the content of scientific knowledge and society at large. Merton studies science as an institution within society, divided up in particular ways. There are reward systems and hierarchies designed to maintain its social order in accordance with certain norms. Merton is sympathetic towards a logical empiricist account of the status of scientific knowledge. He agrees that philosophical analysis can uncover the method by which scientists achieve knowledge. The Edinburgh School maintains, contra Merton, that scientific knowledge can be examined sociologically.

Edinburgh School sociologists react to a second strand of thought that they call 'traditional philosophy of science'. They reject any philosophical approach that explains the attainment of scientific knowledge in terms of a method or set of methods. They also reject any approach that proposes only one standard of rationality. In Edinburgh School literature 'traditional philosophy of science' includes logical positivism, logical empiricism, Popper, Lakatos and Laudan (see Logical positivism; Scientific method). Further, Edinburgh School sociologists claim that traditional philosophers of science ignore empirical information about scientific practice in their

studies.

The Edinburgh School defends a view known as ‘interest theory’ in the sociology of scientific knowledge. This view can be understood as a set of methodological directives for the study of science. Since they argue that the truth or falsehood of scientists’ beliefs are determined by social interests, they explain the validation of all scientific beliefs using the same causal model. Empirical studies of science are therefore undertaken to establish the derivation of particular scientific beliefs. The results of these studies lead them to conclude that scientific beliefs are not made true by any rational method that aided in their acquisition. Philosophers of science are often confused by the presentation of this position as the Edinburgh School deliberately conflates a method for explaining how they arise and a method for explaining how these beliefs are validated. This is a troublesome conflation, if one takes seriously the distinction between justification and discovery. But in fact, the Edinburgh School attacks this distinction explicitly; the conflation is an important tactical move. It is difficult to decide, however, what a causal explanation of belief validation amounts to (Roth 1994).

Finally, the Edinburgh School argues reflexively that the methodological directives should be applicable to the studies of science that result. In other words, any knowledge obtained about science can be explained by the interest model. In their view such reflexivity ensures that the social study of science is as scientific as the natural science under scrutiny (Bloor 1992).

What all constructivists accept from the Edinburgh School’s account is a method, or loose set of directives, for how to study science. Minimally they accept the injunction to study science empirically as it is practised on a day to day basis. The historical version of this method - which is to move away from a study of scientific texts to the social context the texts arose from - also reveals that science is a messy business, a far cry from the linear and tidy affair implied by rationalist reconstruction.

Despite accepting the methodological directives, other constructivists try to avoid some of the more obviously problematic claims of the Edinburgh School. The Edinburgh School focused on beliefs as the unit of analysis, arguing that scientists’ beliefs are determined by their social interests. Constructivists such as Latour and Woolgar (1986) concentrate more on ‘scientific facts’. This changes the focus of study from knowledge as the possession of individuals to knowledge as a commodity or shared object. This approach introduces a more materialist view of knowledge. An example may help illustrate this point.

A scientific fact is represented in any key content statement from a scientific textbook. For example, a popular neuroscience textbook points out that dopamine-beta-hydroxilase is only present in terminal buttons of nerve cells producing nor-epinepherin. This statement is established in neurobiology and has a key place in research. Latour (1987) refers to such facts as ‘black boxes’: once in place they are never challenged or re-investigated. Fifty years ago the textbook statement would have made no sense, because there was no application for its use. The statement was a conjecture even half-way through the testing of the difference between nor-epinepherin terminal buttons and dopamine terminal buttons. Later it becomes a strong hypothesis, and, finally, is accepted by all researchers in the field.

What constructivists investigate is the ‘hardening’ of such facts, the transformation of conjecture into established background knowledge in a field. Constructivism is the view that such facts are not revealed to scientists, but are constructed by them. But this construction of scientific facts is not merely a social process. Constructivists, such as Latour (1987), challenge the Edinburgh School social-determinist assumption that scientific knowledge is determined solely by social factors, arguing that: the social determinist thesis fails because we do not understand society any better than we do the natural world. For social factors to be a suitable candidate for the sole explananda of scientific knowledge we require a well-developed understanding of society. These constructivists claim, however, that by studying science as the connecting point between society and nature, we improve both our understanding of society and scientific knowledge.

From their empirical discoveries about that nature of scientific practice, constructivists go on to make general claims about science. Importantly, they claim that scientific knowledge is constructed by scientists, not discovered (Woolgar 1988). Constructivists get into difficulty cashing out this claim. Several philosophers criticize careless presentations of constructivism, which imply that scientists literally make the world by negotiating among themselves. What is important about constructivism is the emphasis on the construction of scientific knowledge:

constructivists do not emphasise the construction of the world, but the construction and acceptance of representations of the world.

3 Criticisms and replies

Critics of constructivism charge that the view is necessarily relativist. Next they conclude that since relativism is *prima facie* unsatisfactory, so is constructivism. Can this charge of relativism be answered? Though constructivists have not answered all charges of relativism, they have suggested ways around some of its forms. In what follows we will first survey the many kinds of relativism and then discuss constructivists'responses to this kind of criticism (see Relativism; Epistemic relativism).

Philosophers of science have presented relativism as a menace to science, knowledge, rationality and just about everything else. For example, Larry Laudan says:

> The displacement of the idea that facts and evidence matter by the idea that everything boils down to subjective interests and perspectives is - second only to American political campaigns - the most prominent and pernicious manifestation of anti-intellectualism in our time.
>
> (Laudan 1990: x)

In contrast, many sociologists of science cheerfully embrace relativism. Let us look at some of the forms of relativism that constructivists embrace, before critically assessing them.

Relativists are reacting to two main types of opponent: the realist and the rationalist. We can thus distinguish two broad categories of relativism: antirealism and antirationalism. Laudan characterizes relativism with respect to scientific knowledge as ‘ the thesis that the natural world and such evidence as we have about that world do little or nothing to constrain our beliefs’ (1990: viii). He adds that the relativist slogan is ‘ the way we take things to be is quite independent of the way things are’ (1990: viii). This is antirealist relativism. The negative thesis of antirealism can be formulated as several different positive theses, versions of which are common in the literature. For example:

(a) Scientific knowledge is produced by the whim and fancy of indiviual scientists.
(b) Scientific knowledge is determined by social arrangements of the community of scientists.
(c) Scientific knowledge is determined by the wider political and social interests of individual scientists.

The realist here would argue that scientific knowledge is in some way determined by the real nature of the world. A weaker version of this claim is that knowledge is constrained by the nature of the world.

Another criticism of relativism is that scientific knowledge is further determined by scientists following particular rational methods. This is a thesis about the rationality of scientists and is independent of the realist criticism. The rationalist thesis is opposed by antirationalist relativists who insist that there is no one set of methods guiding scientists to the truth. Critics claim that antirationalism entails that no one set of knowledge claims has any special claim to validity over any others. But this latter position involves an extra step of reasoning and should be distinguished from antirationalist relativism: it can be called judgmental relativism (Knorr-Cetina and Mulkay 1982).

One final distinction is worth making: relativism is not equivalent to subjectivism, although subjectivism is a kind of relativism. A subjectivist believes that the truth of a scientific statement is decided by an individual scientist. Those who hold thesis (a) above are subjectivist and relativist. But there are those who believe that establishing the truth or falsehood of a claim has little to do with individual whim, and yet still has nothing to do with the real world or the scientific method. Those who hold thesis (b) above are in this position. Only thesis (a) above is subjectivist.

Constructivism is not wedded to either subjectivism or judgmental relativism. The assumption that all forms of relativism collapse into one or other of these positions is mistaken. An example of judgmental relativism about scientific knowledge is the claim that it is no more valid than a collection of beliefs about witches. Constructivists will join anyone else in acknowledging that, as a result of advances in scientific knowledge, diseases can be held at bay, we speak to each other across continents and fly around the world. Meanwhile belief in witches' remedies and broomstick travel simply have not caught on throughout society. Constructivists want to answer questions about

why science is more successful than ‘witchcraft’, and so judgmental relativism is not a fruitful theoretical tenet for them. Subjectivism is no more promising as a theoretical basis for addressing such questions.

Critics could still maintain that even if constructivism is not wedded to judgmental relativism or subjectivism, it is still relativist. Why would anyone attempting to give a coherent account of scientific knowledge production be relativist? Constructivists’ defences of relativism derive in part from methodological considerations about the study of scientific practice. Bloor’s symmetry principle (1992) is motivated by these kinds of consideration. The principle states that the causes of a scientific belief can be of the same kind irrespective of that belief’s later acknowledged truth or falsehood. False beliefs, according to Bloor, are not caused by special irrational means. The results of studies guided by this methodological assumption lead Bloor and others to conclude that impartial empirical studies of scientific practice do not lead to rationalism or realism about science. (Paul Feyerabend endorsed similar kinds of conclusion in his historical studies.) This kind of argument amounts to an issue of the burden of proof. Some constructivists, however, reject this argument, resorting to the higher level claim that distinctions between relativism and rationalism or relativism and realism are simply the products of a form of analysis. They suggest that, with a shift of emphasis in the study of science, such distinctions will be rendered obsolete. Latour (1987) adopts this approach.

Latour argues that scientists make claims credible by placing them in a network of objects, other scientists, instruments and other knowledge claims. The following illustrative contrast illuminates this point. There are no widely available textbooks on broom travel, few flying brooms have been seen supporting human weight, and no brooms have challenged aeroplanes as vehicles of commercial air travel. By way of contrast aeroplane travel enjoys much support. In challenging the viability of airplane travel, however, we realize how physically and socially entrenched it is: there are textbooks on fluid dynamics and jet propulsion, and flight schedules are kept daily. Though few scientific facts are as indubitable or widely believed as the fact that 747s fly, all scientific facts are connected to a set of instruments, practices, prominent scientists, observable effects and objects. Latour’s suggestion is that it is possible to tell a constructivist story about how some beliefs have more perseverance than others, without invoking philosopher’s traditional standards of rationality, by accounting for all the connections between the various objects, instruments and social intitutions that constitute the facts that the beliefs pertain to. Notice that on this account, contra antirealist relativism, objects do play a part in determining the validity of scientific beliefs. On this view scientific beliefs acquire authority from the role scientific facts play in connecting social and natural systems. Purported facts supporting other belief systems lack this authority.

This proposed way around relativism rids constructivism of an obviously self-destructive judgmental relativist or subjectivist component; but constructivism is still relativist. Certainly, scientific knowledge is validated by different methods in different fields, and was validated by different methods in earlier epochs; but uncovering these methods is part of the study of science. If this is relativism, then constructivism is relativist. But acquiescence to this kind of relativism is not a disaster.

Though constructivism may not suffer from a disastrous relativism, some argue the view is simply absurd (Devitt 1991). Constructivist sociologists often make an ontological claim that scientists construct objects, and this, according to many critics, is absurd. Of course, scientists do construct objects, thousands of them: telescopes, microscopes, oscilloscopes, micro-pipettes; and, on the smaller side, synthesized enzymes, proteins and chemicals of all kinds. Claiming this is surely not absurd; it is simply stating what goes on in scientific practice. Devitt and others worry that Latour and Woolgar’s (1986) claim that Guillemin and his team *made* Thyrotropin Releasing Factor (Hormone) implies they did not discover it. And Devitt quotes Latour and Woolgar generalizing this point, saying:

> Despite the fact that our scientists held the belief that the inscriptions could be representations or indicators of some entity with an independent existence ‘out there’, we have argued that such entities were construed solely through the use of these inscriptions …objects …are constituted through the artful creativity of scientists.
>
> (Devitt 1991: 256)

A defence against Devitt’s accusation can be made from Latour and Woolgar’s comments about their methodological strategy. They emphasize that their claims about making the world are part of a rhetorical strategy whose methodological point is to detach us from the simple, and predominantly accepted, picture that scientists simply divulge facts about the world that are out there pre-packaged and ready for the taking. But, Latour and

Woolgar's defence aside, many constructivists appear to make the ontological, and not methodological, claim that scientists make up the world (Woolgar 1988). This kind of idealism is not a productive approach to the study of science. Fortunately constructivism does not require this view and nor do constructivists' methodological and epistemological claims require a grounding in any form of idealism (Fine 1996).

Some constructivists tackle ontological questions head on and, rather than endorsing an idealist position, they view nature as a source of resistance to human action (Pickering 1994). The natural world, they argue, is ill-defined and scientific research is a kind of social action that characterizes the world's resistance with various kinds of representation. This view of the world as an as yet undifferentiated source of resistance is the closest constructivists come to a version of realism. It should be noted that critics such as Devitt would not regard this constructivist version of realism as realist.

4 Constructivism as an epistemological thesis

Constructivists claim that a detailed description of what goes into the consolidation of a scientific fact is epistemologically relevant. Scientists consolidate facts with a great deal of effort, moving from a hypothesis about a particular experimental effect to a fact that connects with accepted knowledge in the field. Their effort to consolidate the fact contributes to its authority. Describing how facts accumulate authority helps constructivists understand the distinctive nature of scientific knowledge. This is one version of a story familiar to naturalistic epistemologists (see Naturalized epistemology; Social epistemology). Understood in this way, constructivism is part epistemological and part methodological, and need make no ontological claims.

On the constructivist view, scientific facts acquire authority by passing through several kinds of legitimation strategy. One of the most important of these strategies is for a fact to guide all future practice in a particular area of research. One cannot investigate energy transfer in the mitochondrion without accepting the fact that there are mitochondria, and that they are the loci of energy transfer. Facts accumulate authority by attaching to experimental practice, constraining future experimental practice, being embodied in pieces of equipment, and being unquestioned by members of the scientific community. Constructivists claim to have established empirically that scientific facts do not gain their authority by satisfying standards of certainty or rationality for individual scientists.

Philosophers of science who share the view that empirical studies of scientific practice make an important contribution to understanding science may think that constructivists will contribute to epistemology (Nelson 1994). An epistemology suited to the practicalities of scientific practice has to account for all of science's particularity. Constructivism is most fruitfully interpreted as a contribution to naturalized philosophy of science and as joining in the attempt to give a naturalistic account of the attainment of scientific knowledge.

See also: French philosophy of science; Gender and science; Naturalized philosophy of science; Postcolonial philosophy of science

STEPHEN M. DOWNES

References and further reading

Bloor, D. (1992) *Knowledge and Social Imagery*, Chicago, IL: University of Chicago Press, 2nd edn.(A coherent statement of the Edinburgh School's view. Includes a postscript responding to critics.)

Collins, H. and Pinch, T. (1993) *The Golem*, Cambridge: Cambridge University Press.(A useful introduction to constructivist methodology, which is accessible to all readers. Empirical programme of relativism.)

Devitt, M. (1991) *Realism and Truth*, Oxford: Blackwell, 2nd edn.(This edition includes a chapter criticizing constructivism and which presents a realist alternative.)

Fine, A. (1996) 'Science Made Up: Constructivist Sociology of Scientific Knowledge', in P. Galison and D. Stump (eds) *The Disunity of Science*, Stanford, CA: Stanford University Press, 231-54.(Develops the view that the methodological insights of constructivists are important for philosophy of science. Connects constructivism to a consensus theory of truth.)

Hacking, I. (1989) 'The Participant Irrealist at Large in the Laboratory', *British Journal for the Philosophy of Science* 39: 277-94.(A sympathetic critical appraisal of Latour and Woolgar's work.)

Kitcher, P. (1993) *The Advancement of Science*, Oxford: Oxford University Press. (Contains detailed criticisms of Latour and others' views on realism and rationality.)

Knorr-Cetina, K. (1981) *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*, Oxford: Pergamon Press.(One of the first constructivist laboratory studies.)

Knorr-Cetina, K and Mulkay, M. (1982) *Science in Context*, London: Sage.(A useful collection of case studies and theoretical debates about constructivism.)

Latour, B. (1987) *Science in Action*, Cambridge, MA: Harvard University Press. (Develops the constructivist view via case studies and criticisms of traditional approaches in philosophy of science. Actor-network theory.)

Latour, B. and Woolgar, S. (1986) *Laboratory Life*, Princeton, NJ: Princeton University Press, 2nd edn.(The first constructivist laboratory study published.)

Laudan, L. (1990) *Science and Relativism*, Chicago, IL: University of Chicago Press.(A discussion of the place of relativism in philosophy of science. Introductory level.)

Marx, K. and Engels, F. (1970) *The German Ideology*, ed. C.J. Arthur New York: International Publishers. (Modern sociology of knowledge, background reading.)

Megill, A. (ed.) (1994) *Rethinking Objectivity*, Durham, NC: Duke University Press.(A collection of constructivist, historicist and feminist approaches to defining objectivity.)

Merton, R. (1973) *Sociology of Science*, Chicago, IL: University of Chicago Press. (An influential collection of theoretical papers on sociology of science.)

Nelson, A. (1994) 'How Could Scientific Facts Be Socially Constructed?', *Studies in the History and Philosophy of Science* 25: 535-47.(Argues that there is an important philosophical challenge embodied in constructivist studies of science.)

Pickering, A. (1984) *Constructing Quarks*, Chicago, IL: University of Chicago Press. (A case study from the interest theory perspective. Edinburgh School.)

Pickering, A. (1992) *Science as Cultural Practice*, Chicago, IL: University of Chicago Press. (A collection of theoretical debates about the status of constructivism.)

Pickering, A. (1994) 'Objectivity and the Mangle of Practice', in A. Megill (ed.) *Rethinking Objectivity*, Durham: Duke University Press, 1994.(See Megill 1994.)

Roth, P.A. (1994) 'What Does the Sociology of Scientific Knowledge Explain?: Or, When Epistemological Chickens Come Home to Roost', *History of the Human Sciences* 7: 95-108.(A detailed criticism of the Edinburgh School account of belief causation.)

Shapin, S. and Shaffer, S. (1985) *Leviathan and the Airpump*, Princeton, NJ: Princeton University Press. (Much-cited groundbreaking case study from the interest theory perspective. Edinburgh School.)

Woolgar, S. (1988) *Science: The Very Idea*, London: Tavistock.(Sustained argument for a radical constructivism disguised as an introductory text. Reflexivist approach.)

Constructivism in ethics

There have been many forms of the idea that there are no distinctively ethical properties, and that ethical claims are composed or constructed out of other considerations. In some sociological writing the term is used for the view that ethics is artificial or socially constructed, for example out of social norms or attitudes. However, constructivism in ethics is now identified mainly with certain views not of the source but of the justification of ethics, which have their origin in Kant's work and have been revived and developed by John Rawls. Constructivism in ethics seeks to show how substantive principles, and in particular principles of justice, can be built out of minimal, uncontroversial elements, such as a slender account of action and reason.

1 Social constructivism and constructivism in ethics

The idea that ethical ideas and institutions, like other ideas and institutions, are not basic features of the world but are *made* or *constructed* by human beings has a long history. Protagoras claimed that man is the measure of all things; Hobbes depicted the state and justice as a mortal God made by men; Hume argued that justice is an artificial virtue (see Protagoras §3; Hobbes, T. §§6-7; Hume, D. §4). The twentieth century has seen a wide range of claims that standards of right and wrong, good and bad are constructed. Most of these claims assert that actual ethical norms are the *social constructs* of particular human societies, without relevance or validity beyond the societies in which they arose. Their proponents speak of the social construction of reality (of meaning, of science), the construction of social identity, the construction of political attitudes - and the construction of ethics (see Constructivism). When ethical norms are viewed as social constructions the aim is to explain rather than to justify them; in general social constructivism has interested sociologists more than it has interested philosophers.

A more specific and philosophically significant view of the construction of ethics identifies it with the claim that ethical principles can be justified by appeal to elementary and uncontroversial features of action and reason. Constructivism in ethics can be contrasted both with forms of ethical realism and with many conventionalist or relativist views of ethics (see Moral realism; Moral relativism). Unlike realists, constructivists deny that there are any distinctively moral facts or properties, whether natural or non-natural, which can be *discovered* or *intuited* and which provide the foundations of ethics (see Fact/value distinction; Naturalism in ethics). Unlike conventionalists or relativists, constructivists think that it is possible to justify universal ethical principles (see Universalism in ethics). The distinctive feature of constructivism in ethics is the claim that universal ethical principles can be built out of quite minimal accounts of action and reason.

2 Kantian constructivism: Rawls

At first glance it might seem that most utilitarian and kindred ethical theories are constructivist, because they do not assume that there are objective moral facts and properties, and they aim to justify universally valid ethical claims (see Utilitarianism). However, the term 'constructivist' is not generally applied to utilitarian ethical theories, because their arguments appeal not only to conceptions of action and reason but to individual desires and preferences. It is usually reserved for positions which do not draw on assumptions about the content of desires and preferences in justifying ethical conclusions. For this reason constructivism in ethics is often spoken of as 'Kantian constructivism' (see Kantian ethics).

The leading exponent of Kantian forms of constructivism is John Rawls, who has developed a range of arguments for constructivist approaches to the justification of principles of justice. In *A Theory of Justice* (1971), he uses the term 'constructive' very broadly to refer to ethical positions, including utilitarianism and his own contractualism, which seek to provide procedures for resolving moral problems (see Contractarianism §§7-9). He contrasts constructive positions with theories, such as intuitionism, which provide only a plurality of unranked principles, hence no adequate procedures for resolving ethical problems (see Intuitionism in ethics).

In his later work, beginning with 'Kantian Constructivism in Moral Theory' (1980) and continuing through *Political Liberalism* (1993), Rawls uses the term 'constructivist' more narrowly to refer to methods of reasoning in ethics which not merely offer ways of justifying substantive principles, but do so on the basis of minimal and supposedly uncontroversial claims about reason and action, and without referring to agents' desires or preferences. In these later writings Rawls no longer views utilitarianism as a constructive position and argues only for a

constructive theory of the right, and so of justice. He also concludes that no constructive account of the good or of virtue can be given, and that we must accept that there is no rational way of choosing among conceptions of the good.

The most controversial features of Rawls' subtly differing versions of ethical constructivism are his considerations about the resources out of which an account of substantive ethical principles is to be constructed. His justificatory arguments assume that we can appeal not merely to instrumental but to other conceptions of reason or reasonableness. In *A Theory of Justice*, he had argued that principles of justice were justified if they were in *reflective equilibrium* with our considered moral judgments. In 'Kantian Constructivism in Moral Theory'(1980) he viewed persons as 'agents of construction' who appealed to certain ideals, such as a Kantian ideal of the person, in justifying principles of justice. In 'Justice as Fairness: Political not Metaphysical' (1985), he went on to argue that justifications could appeal to ideals which are constitutive of the identities of citizens of liberal societies. In *Political Liberalism* (1993: 49) he argued that 'persons are reasonable…when…they are ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so' and that this conception of 'public reason is characteristic of a democratic people: it is the reason of its citizens, of those sharing the status of equal citizenship' (1993: 213). Discussions of Rawls' versions of constructivism have questioned whether appeals to the ideals or identities of citizens, or to their conception of fair terms of cooperation, can count as forms of reasonableness.

3 Kantian constructivism: other possibilities

Several other recent writers have proposed alternative versions of ethical constructivism, and have argued that these can be used to justify universally valid ethical principles. Thomas Scanlon (1982) has proposed that moral principles are those which parties similarly motivated could have no reason to reject. Thomas Hill (1989) has suggested that Kant's 'Kingdom of Ends' can be seen as a constructive model for deliberating about moral principles. Onora O'Neill (1989, 1996) has proposed ways of reading Kant and of developing Kantian ethics that not only construct substantive ethical principles out of elementary conceptions of agency and reason, but construct an account of practical reason itself out of the necessary conditions for principles to be adoptable by all relevant parties, so explicating the Kantian identification of practical reason with universalizability.

See also: Practical reason and ethics

ONORA O'NEILL

References and further reading

Hill, T.E, Jr (1989) 'Kantian Constructivism in Ethics', *Ethics* 99: 752-70; repr. in *Dignity and Practical Reason*, Ithaca, NY: Cornell University Press, 1992, 226-50. (Interprets Kant's 'Kingdom of Ends' as model for constructing ethical principles.)

McCarthy T. (1994) 'Kantian Constructivism and Reconstructivism: Rawls and Habermas in Dialogue', *Ethics* 105: 44-63. (Compares the conceptions of practical reason used in the later works of Rawls and Habermas.)

O'Neill, O. (1989) *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge: Cambridge University Press. (Papers on Kant's vindication of reason and on his ethics; emphasis on their constructive character.)

O'Neill, O. (1996) *Towards Justice and Virtue: A Constructive Account of Practical Reasoning*, Cambridge: Cambridge University Press. (A constructive account of practical reason, justice and virtue.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (His magnum opus: the most influential work in political philosophy of the late twentieth century.)

Rawls, J. (1980) 'Kantian Constructivism in Moral Theory' (The Dewey Lectures), *Journal of Philosophy* 77: 515-72. (Substantial revision of *Theory of Justice*; emphasis on procedures by which agents can reach agreement about justice.)

Rawls, J. (1985) 'Justice as Fairness: Political not Metaphysical', *Philosophy and Public Affairs* 14: 223-51. (Further revision; emphasis on the political rather than metaphysical basis of justification in ethics.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press. (Substantial reworking of earlier papers; argues for the political justification of a liberal conception of justice, and the impossibility of justifying any 'conception of the good' in pluralistic societies.)

Scanlon, T.M. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press, 103-28. (Defends a contractualist account of practical

reason: rules are to be accepted if nobody motivated to find agreement could reject them.)

Constructivism in mathematics

Constructivism is not a matter of principles: there are no specifically constructive mathematical axioms which all constructivists accept. Even so, it is traditional to view constructivists as insisting, in one way or another, that proofs of crucial existential theorems in mathematics respect constructive existence: that a crucial existential claim which is constructively admissible must afford means for constructing an instance of it which is also admissible. Allegiance to this idea often demands changes in conventional views about mathematical objects, operations and logic, and, hence, demands reworkings of ordinary mathematics along nonclassical lines. Constructive existence may be so interpreted as to require the abrogation of the law of the excluded middle and the adoption of nonstandard laws of constructive logic and mathematics in its place.

There has been great variation in the forms of constructivism, each form distinguished in its interpretation of constructive existence, in its approaches to mathematical ontology and constructive logic, and in the methods chosen to prove theorems, particularly, theorems of real analysis. In the twentieth century, Russian constructivism, new constructivism, Brouwerian intuitionism, finitism and predicativism have been the most influential forms of constructivism.

1 Orientation

A call for ‘constructive existence’, standardly thought a hallmark of constructivism in mathematics, represents a general, revisionary attitude towards certain crucial statements of existence. That attitude is expressible as E:

(E) A mathematical statement $\exists x M(x)$, if crucial, is admissible only if a well-defined *a* can be constructed such that $M(\mathbf{a})$ is admissible, where a denotes *a*.

For E to be revisionary, the constructivist must indicate which claims are to be crucial, provide accounts of well-definedness and constructibility, and respond to any disruptions in traditional logic incited by E. It is in terms of these requirements that the outlines of influential forms of constructivism are demarcated.

Well-definedness. E need not govern all existential statements. Forms of constructivism are earmarked by choices of crucial statements, statements to which E is to be applied. Predicativists impose a version of E on claims about classes, but sometimes ignore E in elementary number theory. Russian constructivists put all their mathematics under the aegis of a strict reading of E.

As a call for revision, E is nugatory without strictures on well-definedness and constructibility. Conventionally, whenever $\exists n A(n)$ holds of natural numbers, there is a determinate *m* which is the least number for which $A(x)$ holds. However $\exists n A(n)$ is proved, one might then think to ‘construct’ an *n* such that $A(\mathbf{n})$ by using the computable function which outputs the number *m* on any input. So, unless the constructivist supplies further explanation, any proof of $\exists n A(n)$ might seem to satisfy E.

In general, well-definedness is explained in terms of canonical notations. Some constructivists count a natural number canonically denoted once a tally-mark numeral is available for it. Constructivist E. Bishop identified well-defined integers with number descriptions which are convertible into base-ten numerals. A real number may be deemed well-defined once a finitary recipe for approximating it is at hand. A predicativist about classes refuses classes not admitting definition in suitably predicative terms.

If well-definedness depends upon supplies of notations, it is natural to imagine constructive mathematical realms as nominalistic in that everything in them be nameable. It is also common to picture constructive realms as expanding over time as more notations are created. Some see in this image a contemporary rendering of Aristotle’s notion of potential infinities, of domains which are inexhaustible not because there now exist infinitely many elements in them but because, over time, more elements always appear. Explicit treatment of time enters infrequently into constructive mathematics itself, Brouwer’s use of the creative subject affording a celebrated exception (see Intuitionism §3).

Constructions. Constructibility is explicated in functional or operational terms. In general, a value is constructible just in case there is a recipe for computing that value in principle. Again, definite restrictions are needed if E is to

fulfil a desire which often motivates constructivism: that acceptable definitions in mathematics lead to algorithmic computation. This desire is explicit in E *, a generalization of E many constructivists endorse.

(E *) The statement $\forall x \exists y A(x, y)$ is, if crucial, admissible only when there is a construction *o* which computes, given any well-defined value *a* for *x*, a well-defined $o(a)$ such that $A(\mathbf{a}, \mathbf{o}(\mathbf{a}))$ is admissible.

Conventionally, axioms of choice ensure that, when $\forall x \exists y A(x, y)$ holds, there is a function *f* such that $\forall x A(x, \mathbf{f}(x))$ holds too. Hence, talk of 'constructions' must amount to more than talk of conventional functions. Once again, forms of constructivism are differentiated in the kinds of construction allowed. Traditional intuitionists demand that the class of constructions include only those operations given by computing recipes which humans can, in principle, follow. Finitists seem to want more: that recipes be guaranteed to output on any admissible input. Given their nominalistic leanings, certain constructivists identify operations with explicit rules or linguistic expressions of rules.

Constructive logics. Constructivism almost always calls for radical revisions in conventional mathematics. E may require, of any statement, that its admissibility entail the availability of constructions. To see why, take a statement *S* and define a predicate $A(x)$ so that $A(x)$ is true of 0 alone, if *S* holds, and of 1 alone otherwise. Trivially, when *S* holds, there is an *m* such that $A(\mathbf{m})$ and, given E, there must be a construction of such an *m*. But there is more to see from this example. If the law of the excluded middle (LEM) is valid, there is a simple proof of $\exists n A(n)$. But, if the truth or falsity of *S* is unknown, this simple proof appears to violate E. Short of settling *S*, one cannot construct a well-defined *m* such that $A(\mathbf{m})$, if 'construction' and 'well-definedness' are interpreted restrictively, since such an *m* equals 0 just in case *S* is true. Therefore, E is so interpretable as to conflict with the validity of LEM.

It has also been said that constructivism requires changes in logic for the reason that mathematical claims, constructively treated, hold extra information in the form of recipes for construction. Supplying this information is a requirement on any constructively correct proof. Presumably, the proof of $\exists n A(n)$ just described does not bear appropriate information when *S* is unknown. In this regard, it is important to note that the impact of E upon logic should not be thought wholly restrictive. Constructive claims can feature in proofs not only as conclusions but also as premises and, hence, as yielding up extra information. Some constructivists, for example, read E * as underwriting logical principles for deducing, directly from $\forall x \exists y$ premises, axioms of choice.

One way constructivists delimit needed alterations to logic is by establishing canons for correct constructive proof. Given such canons, a principle of inference is to be allowed when it licenses inferences that preserve correct provability. Intuitionists and Russian constructivists, for example, require of disjunctive and universal claims that

(E $\vee$) a constructive proof of $(A \vee B)$ include an effective marker and a further constructive proof. If the marker has value 0, the further proof is one of *A* and, if the marker is 1, the proof is of *B* and
(E $\forall$) a constructive proof of $\forall n A(n)$ involve an operation *f* such that, for any number *m*, $f(m)$ yields a constructive proof of $A(\mathbf{m})$.

The basic idea of using canons of abstract proof to assess logical principles is apparent in the writings of Brouwer. A. Heyting and A.N. Kolmogorov were the first to devise both proof specifications for each form of mathematical statement and formal logics arguably correct with respect to them. (Kolmogorov's logic was not complete, however.) Allegiance to the logic of Heyting, if not to any particular formalization of it, is shared by many constructivists.

Logics obeying E $\vee$ and E $\forall$ may fail to certify LEM. To see this, let $A(n)$ be an undecidable predicate of natural numbers. Then, $\forall n(A(n) \vee \sim A(n))$ plausibly lacks a constructive proof. For, were there such a proof, it would yield a recipe *f* giving, computably in *n*, a marker $f(n)$ signalling, in each case, the truth or the falsity of $A(\mathbf{n})$ and contradicting the supposition that $A(\mathbf{n})$ be undecidable.

2 Russian constructivism

Prominent among Russian (or Markovian) constructivists have been A.A. Markov, who initiated the approach around 1950, N.A. Shanin and G.S. Ceitin. As constructivists, the Russians might be thought most exacting, since E is imposed on existential quantifications over any domain. Entities are well-defined only if they admit encoding

as natural numbers or as words on a finite alphabet. Encodings must be, as Markov wrote, ‘potentially realizable’ as concrete notation (1962: 2). Accordingly, Russian constructivists require operations to be given as Turing machines or Markov algorithms and, hence, as finite words. Sets are identified with arithmetic formulas with one free variable.

In virtue of their views on operations, a belief in (a version of) Church’s thesis (CT) is ascribed to Russian constructivists. CT’s most straightforward version is ‘Every total numerical function is Turing computable’. This CT is not, as in recursion theory, a hypothesis about the meaning of ‘informally computable’, but a powerful mathematical axiom. For example, CT entails, in Russian constructivism, the falsity of the general LEM. The Markovian implementation of E, then, enjoins upon Russian constructivists markedly anti-classical principles of logic and mathematics.

Unlike new constructivists and Brouwerian intuitionists, Russian constructivists adopt (a form of) Heyting’s logic as an official formal system. They employ an understanding of the system’s signs devised by Shanin on which the truth of an existential claim *S* records the existence of an algorithm on whose operation *S* is a specification. In the name of this understanding, mathematical principles including CT get adopted which are permitted by neither intuitionists nor new constructivists. Among these is Markov’s principle:

(MP) Whenever a natural number predicate $A(x)$ is decidable, $\exists n A(n)$ is provable constructively from $\sim\forall n{\sim}A(n)$.

When MP is rendered in algorithmic terms, a motive for adopting it is plainest. The decidability of $A(x)$ then requires a recipe *R* for determining membership in $\{n\colon\ A(n)\}$. If one runs *R* successively on inputs $0, 1, 2, \ldots$ searching for members of $\{n\colon\ A(n)\}$, a proof of $\sim\forall n{\sim}A(n)$ guarantees that this search cannot fail. MP says that, in these circumstances, there is an *m* such that $A(\mathbf{m})$ to which the search procedure will lead.

Adherence to CT lends Russian constructive analysis a computable cast throughout. As O. Aberth wrote, the mathematics ‘may be informally described as an analysis wherein a computation algorithm is required for every entity employed. The functions, the sequences, even the numbers of computable analysis are defined by means of algorithms. In this way definition and evaluation are made inseparable’ (1980: 2). There is no single conception of real number within Russian constructivism. A real is an ‘FR-number’, for example, when it is given in terms of a pair of algorithms, the first yielding a Cauchy sequence of rational approximants and the second a modulus of convergence. (‘FR’ stands for ‘fundamental sequence with regulator of convergence’.) The most famous mathematical theorem of Russian constructivism is that proved by Ceitin (and, independently, by Kreisel, Lacombe and Schoenfield), which implies that every function from the reals to the reals is continuous. In Russian constructivism, this does not entail, as it would conventionally, that every real-valued function on a closed interval is uniformly continuous. Using a result of E. Specker, Russian constructivists prove that there are continuous functions which are unbounded on closed intervals.

3 New constructivism

New constructivists (or Bishop constructivists) comprise a relatively small group of mathematicians inspired by E. Bishop. Bishop’s influential *Foundations of Constructive Analysis* banished lingering doubts that a significant mathematics could survive constructivization. It was consequent to the book’s 1967 publication that new constructivism had its greatest impact upon mathematics.

As for well-defined entities, new constructivists do not require that everything be coded as integers. In addition to numbers, they embrace abstract operations, functions (akin to extensional operations), proofs and sets. There is canonical number notation but no special notations for sets or operations. A set *S* is said to be ‘constructed’ once recipes are given for constructing *S*-elements and conditions are specified under which equalities between those elements can be proved. For example, a new constructivist real number is a triple $\langle r, p, s\rangle$ of operations *r* and *s* and a proof *p* wherein *r* is a sequence of rationals, *s* a sequence of natural numbers and *p* a proof that *s* is a modulus of convergence for *r*.

As far as constructions are concerned, only mathematical operations representable by demonstrably computable recipes are allowed. Even so, new constructivists refuse CT and set no a priori limit on the way recipes must be expressed. Nor do new constructivists endorse a particular formalism for constructive logic. Certain working

limitations are, however, discernible from a metamathematical viewpoint: CT has been proven consistent with systems reflecting Bishop's ideas. Metamathematical experience also suggests that all new constructive theorems are formally provable from suitable axioms via Heyting's logic.

Bishop insisted that new constructive mathematics capture the 'numerical meaning' of mathematical claims. Numerical meaning, he averred, is the sole clear and legitimate meaning for any mathematics. This meaning was to be cashed out in predicting the outcomes of performable computations on the integers, computations that come from realizing constructive theorems as computer programs. Bishop encouraged a comparison between formalizations of new constructivism and high-level specification and programming languages, ones whose proofs could be compiled into implementable code. This general idea of 'programming constructive mathematics' has since been taken up by a number of computer scientists.

The principal effort in new constructivism, which Bishop called 'the most urgent task of the constructivist' (1970: 54), is not a wholly autonomous construction of mathematics under the guidance of E but a painstaking *re*construction of classical mathematics intended to reveal its numerical meaning. The goal is to set as many traditional theorems as possible onto a constructive basis without offending conventional sensibilities. Conflict never erupts between theorems of new constructivism and those of conventional mathematics since new constructivists shun any anti-classical consequences of E.

Without benefit of nonclassical assumptions, new constructivists cannot prove, as their Russian colleagues do, that every real-valued function of a real variable is continuous. Moreover, Specker's theorem shows that they cannot expect to prove that continuous functions are bounded in range, even on closed unit intervals. In order to retain the boundedness theorem for continuous functions on such intervals, new constructivists supplement the traditional statement of the theorem with the extra hypothesis that the function is uniformly continuous. This kind of hypothesis supplementation has become a feature of all branches of new constructive mathematics.

4 Brouwerian intuitionism

The most celebrated of contemporary mathematical constructivisms, at least before 1967, was Brouwerian (Dutch or neo-) intuitionism. To an extent, the celebrity of intuitionism was allied to that of its leading exponent, L.E.J. Brouwer, who first presented the approach in articles appearing between 1907 and 1930 (see Intuitionism §§2-3). Brouwer and his followers maintained a relative liberality in the matter of well-definedness. Besides natural numbers, intuitionists granted full status to non-algorithmic or 'free choice' sequences, kinds of sets, abstract proofs (including infinitary ones) and moments of time. Brouwer refused nominalistic constraints: mathematical existence for him did not await the fashioning of notations. Mathematics was to be constructed as a free creation, unaided by realist metaphysics, linguistic representation or empirical data. Brouwer deemed an entity legitimate if it was apparent to a Neo-Kantian inner intuition. A well-defined entity could be displayed in intuition either as a mental construction from the perception of a passage in time or as a choice sequence or set of items so constructed.

From among a panoply of legitimate operations on numbers, intuitionists denominate as 'law-like' those functions which are computable by recipe. In keeping with his anti-linguistic bent, Brouwer did not establish strict limits on the means by which recipes are to be communicated and CT for law-like sequences is not intuitionistically acceptable. Brouwer expressed a concern (one which both Bishop and Markov later proved unfounded) that the law-like continuum - in which reals are identified with law-like sequences - would not support a satisfactory real analysis. This concern seems to have led him to import into intuitionistic mathematics an analogue of the probabilistic 'choice sequences' of Borel. For Brouwer, choice sequences need not be law-like or even fully determinate. There may be some restrictions set upon sequence terms to appear in future but, beyond that, nothing about the course of a sequence may be presumed known.

Traditionally, there are two intuitionistic analogues to the conventional notion of set: 'species' and 'spread'. Species are properties determined by extensional mathematical predicates and individuated extensionally: pairs of species are identical when they have the same members. Spreads, on which Brouwer laid great emphasis, are either species of choice sequences constrained to obey a 'spread law' or species of sequences obtained from those via continuous mappings. The generators of real numbers comprise a spread in the latter sense.

The quintessential intuitionistic result is 'Brouwer's theorem': every real-valued function of a real variable on a closed, bounded interval is uniformly continuous. Brouwer concluded from this that every real-valued function on

the reals is continuous. Brouwer's proof of the latter is markedly different from that which leads to Ceitin's theorem in Russian constructivism; it relies upon mathematical principles inconsistent with CT (see Intuitionism §3).

5 Finitism

As the name suggests, finitism requires that one eschew reference to infinite collections. In the case of Hilbert's finitism, this requirement was coupled with a desire, epistemically motivated, to replace abstract concepts with concrete, visualizable notation. Hilbert's idea was that a metamathematics confined to the realm of the visualizable would secure conventional mathematics against the paradoxes. Besides Hilbert's finitistic metamathematics, notable forms of twentieth-century finitism are represented by Skolem's primitive recursive arithmetic and Yessenin-Volpin's ultra-intuitionism.

For most finitists, only natural numbers or items encodable as natural numbers count as fully well-defined. Numbers are not conceived as constituting an infinite totality but as individually realizable in concrete notation. Hilbert would have had it that talk of natural number, finitistically construed, comes down to talk of (the intuitable forms of) strings of tally marks. He wrote, 'The subject matter of mathematics is, in accordance with this theory, the concrete symbols themselves whose structure is immediately clear and recognizable' ([1926] 1983: 142) (see Hilbert's programme and formalism).

If the use of unbounded existential quantification is believed to require infinite totalities, as it was by Hilbert, restriction E will be implemented finitistically using constructions carrying explicit bounds. Bounded existential quantifications over the natural numbers are usually finitistically admissible and other existentially quantified claims, when allowed, get treated as 'incomplete statements', abbreviations for bounded quantifications. Universal quantifications are, if admissible, finitistically reconstrued as free variable expressions each instance of which is verifiable or falsifiable by direct calculation.

Constructions do enter into finitistic mathematics, but not as members of an infinite domain of number operations, as in intuitionism. Nor are they identified with members of a category of abstract rules or infinite graphs. An operation on natural numbers is a finitistic construction if realizable as a combinatorial transformation of concrete signs for which explicit bounds on the number of steps required to complete the transformation have been given. Primitive recursive functions are natural candidates for finitistic constructions on the natural numbers since these functions are Turing computable functions definable without unbounded search routines. Whether the class of finitistically admissible functions includes more than the primitive recursive functions is a topic of disagreement.

In Hilbertian finitism, admissible statements are said to capture mathematical claims which are '*inhaltlich*' or contentual. Statements refusing finitistic reconstrual were considered not contentual but ideal, expressions lacking full content. For Skolem, only those statements whose truth-values are fully determinable via primitive recursive calculations are admitted. Arithmetical predicates which are composed of simple equations between primitive recursive terms combined by connectives and bounded quantifiers Skolem treated as primitive recursive functions. Therefore, systems representing this approach are called 'logic-free', since quantifiers and connective signs are not required. Because of the limitations on the kinds of claim permitted, the proofs in Skolem's and Hilbert's finitistic arithmetics obeyed conventional logic.

Finitistic number theory and analysis, in the style of Skolem, has undergone some development, notably by R. Goodstein. Various constructions of real number are examined, among them that of primitive recursive real number: a primitive recursive sequence armed with a primitive recursive modulus. Anti-classical mathematical axioms are refused; hence, there are no finitistic continuity theorems like Ceitin's. However, since finitistic mathematics is intuitionistically acceptable, neither will it be possible to contradict Brouwer's theorems. According to Goodstein, a finitist in the style of Skolem cannot prove that a bounded, monotone sequence of rational numbers is convergent.

The so-called 'ultra-intuitionism' of A.S. Yessenin-Volpin differs radically from the views just described. Standard concepts of natural number and Skolem's 'recursive mode of thought' are not accepted as unequivocally correct. Yessenin-Volpin limits well-definedness to numbers which are feasible rather than potentially realizable: to be well-defined, an ultra-intuitionistic number must be literally displayed as a notation. One might say that the well-defined numbers of ultra-intuitionism do not constitute a single potentially infinite series but form a tree, each

branch of which represents a temporally expanding series of tokens from some numeral system. From P. Bernays and G. Kreisel, commentators have adopted the term ‘strict finitism’ for application to such positions as Yessenin-Volpin’s in distinguishing them from more liberal finitistic conceptions. Among philosophers of language, the viability of strict finitism has proved a subject of controversy. M. Dummett (1975) has maintained that sorites-like paradoxes would so haunt strict finitism that it is revealed as incoherent; C. Wright (1982) has since raised serious questions about the correctness of Dummett’s conclusions.

6 Predicativism

Predicativism took impetus from the writings of Poincaré and Russell on the paradoxes and received influential, but troubled, expression in Whitehead and Russell’s *Principia Mathematica* (1910-13) (see Theory of types). Predicativistic approaches to the foundations of mathematics were later instigated or investigated by C. Chihara, S. Feferman, G. Kreisel, P. Lorenzen, H. Wang and H. Weyl, among others.

Predicativists would have it that all mathematical entities - or all higher-order entities - are logical classes and that their existence is governed by a version of E. A logical class $\{x\colon A(x)\}$ is predicatively well-defined on condition that all quantifiers in $A(x)$ are restricted to a range of classes or entities already known to be well-defined and including neither $\{x\colon A(x)\}$ nor any class presupposing it for its definition. Definitions $\{x\colon A(x)\}$ satisfying this condition are termed ‘predicative’ and others ‘impredicative’. Russell put forward the requirement that all classes be predicative in the form of what came to be known as the ‘vicious-circle principle’. According to Russell, predicative classes must be organized into types so that ‘any expression containing an apparent [bound] variable is of higher type than that variable’ ([1908] 1967: 182). Consequently, the favoured way to implement bans against impredicativity has been to devise hierarchical formal systems (predicative type theories) for class definition and manipulation.

An intended model of predicative formalism, greatly simplified, is segmented into an ascending sequence of orders or levels indexed by natural or ordinal numbers. At the lowest or 0 level is a collection of given elements such as the natural numbers. Level 1 contains those classes whose definitions involve quantifiers ranging only over the given elements, level 2 contains classes specifiable using quantifiers ranging either over the given elements or over classes of level 1, and so on. Each class has a level, so impredicative classes are ruled out. Formalisms embodying an idea of levels include the ‘ramified type theory’ of *Principia Mathematica* and Wang’s systems Σ.

A class is predicative thanks to a suitable definition or notation for it; that notation cannot be grasped before notations for classes of lower level are grasped. As Wang put it, ‘new objects are only to be introduced stage by stage without disturbing the arrangement of things already introduced or depending for determinedness on objects yet to be introduced at a later stage’ (1970: 640). Following Poincaré, one often imagines predicative universes as expanding over time.

As for constructions, we can think of each predicative class as the result of a well-founded construction process involving conventional logical or class operations such as complementation, intersection and union, and quantification which is predicatively restricted. Hence, the imposition of E upon predicative mathematics seems to call for no great alteration in the fundamental inferences of conventional higher-order logic, apart perhaps from variable indexing. Each variable may wear an ordinal index and quantifier rules be so restricted that bound variables share indices with substituted parameters.

Predicativism has not spawned a programme for reworking mathematics as influential as Brouwerian intuitionism or new constructivism. Even so, a wide variety of mathematical results are known. Weyl saw that the theorem that every non-empty bounded class of reals has a least upper bound is unobtainable in complete generality within (a restricted) predicative mathematics if reals get defined in terms of classes. He showed that, by treating reals as sequences, the theorem that every bounded, monotonically increasing sequence approaches a limit is predicatively provable, as are other theorems of analysis. Wang has argued that predicativists cannot retain Cantor’s theorem on the existence of uncountable sets.

See also: Logical and mathematical terms, glossary of

DAVID CHARLES McCARTY

References and further reading

Aberth, O. (1980) *Computable Analysis*, New York: McGraw-Hill.(An attractive technical introduction to analysis in the style of Russian constructivism.)

Beeson, M.J. (1985) *Foundations of Constructive Mathematics*, Berlin: Springer.(An overview of the metamathematics of constructive formal systems. Requires considerable formal experience.)

Benacerraf, P. and Putnam, H. (eds) (1964) *Philosophy of Mathematics: Selected Readings*, Englewood Cliffs, NJ: Prentice Hall, 2nd edn, 1983.(An invaluable collection of classic essays on the foundations of mathematics; most are nontechnical.)

Bishop, E. (1967) *Foundations of Constructive Analysis*, New York: McGraw-Hill.(The bible of new constructivism. A technical work with a lucid nontechnical introduction.)

Bishop, E. (1970) 'Mathematics as a Numerical Language', in J. Myhill, R.E. Vesley and A. Kino (eds) *Intuitionism and Proof Theory*, Amsterdam: North Holland, 53-71.(A clear statement of Bishop's central philosophical ideas.)

Bridges, D. and Richman, F. (1987) *Varieties of Constructive Mathematics*, Cambridge: Cambridge University Press.(A mathematical introduction to constructivism, contrasting theorems of new constructivism with those of intuitionism and Markovian constructivism. A background in conventional analysis is required.)

Brouwer, L.E.J. (1981) *Brouwer's Cambridge Lectures on Intuitionism*, ed. D. van Dalen, Cambridge: Cambridge University Press.(A compact sample of Brouwer's thinking; well edited but quite technical.)

Chihara, C. (1973) *Ontology and the Vicious-Circle Principle*, Ithaca, NY: Cornell University Press. (Includes an exposition, for philosophers, of traditional issues surrounding predicativity.)

Dummett, M.A.E. (1975) 'Wang's Paradox', *Synthese* 30: 301-24; repr. in *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press, 1978, 248-68.(A thought-provoking exploration of issues from philosophy of language which may affect the legitimacy of finitism.)

Dummett, M.A.E. (1977) *Elements of Intuitionism*, Oxford: Clarendon Press, 1990.(An influential text in which great emphasis is placed upon the metalogic of intuitionism.)

Feferman, S. (1964) 'Systems of Predicative Analysis', *Journal of Symbolic Logic* 29: 1-30.(A brilliant introduction to predicativism and its systematic formulation. The second section requires technical knowledge.)

Hazen, A. (1983) 'Predicative Logics', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, vol. 1, *Elements of Classical Logic*, Dordrecht: Reidel.(A helpful overview of and introduction to predicativism and predicative logics.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879- 1931*, Cambridge, MA: Harvard University Press.(An invaluable collection of significant articles by Gödel, Hilbert, Brouwer, Skolem and others. Technical knowledge often required.)

Heyting, A. (1956) *Intuitionism: An Introduction*, Amsterdam: North Holland; 3rd revised edn, 1971. (The first - and still the most charming - popular technical account of Brouwer's intuitionism.)

Hilbert, D. (1926) 'Über das Unendliche', *Mathematische Annalen* 95: 161-90; trans. 'On the Infinite', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983; and in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967.(A classic statement of Hilbert's views on metamathematics; largely nontechnical.)

Markov, A.A. (1962) 'O Konstruktivnoii Matematike', *Trudy Matematicheskogo Instituta imeni V.A. Steklova* 67: 8-14; trans. 'On Constructive Mathematics', *American Mathematical Society Translations*, series 2, 98: 1- 9, 1971.(A brief introduction to Russian constructivism; relatively nontechnical.)

Russell, B.A.W. (1908) 'Mathematical Logic as Based on the Theory of Types', *American Journal of Mathematics* 30: 222-62; repr. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 150-82.(Includes Russell's diagnosis of the paradoxes and an explanation of his ramified type theory.)

Skolem, T. (1923) 'Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich', *Skrifter utgit av Videnskapsselskapet I Kristiania, I. Matematisk-naturvidenskabelig Klasse* 6: 1-38; trans. 'The Foundations of Elementary Arithmetic Established by Means of the Recursive Mode of Thought, Without the Use of Apparent Variables Ranging Over Infinite Domains', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 302-33.(A brilliant and careful formal development of a finitistic arithmetic.)

Tait, W.W. (1981) 'Finitism', *Journal of Symbolic Logic* 78 (9): 524-46.(A lucid exposition of one approach to the foundational issues surrounding finitism.)

Troelstra, A.S. and Dalen, D. van (1988) *Constructivism in Mathematics: An Introduction*, Amsterdam: North

Holland, 2 vols.(A comprehensive technical overview of the current state of constructive mathematics, with emphasis on intuitionism and mathematical logic.)

Wang, H. (1970) *Logic, Computers and Sets*, New York: Chelsea.(Includes several articles by Wang on the historical and technical background to predicativism.)

Weyl, H. (1918) *Das Kontinuum: Kritische Untersuchungen über die Grundlagen der Analysis*, Leipzig: Veit; trans. T. Bole, *The Continuum: A Critical Examination of the Foundations of Analysis*, New York: Dover, 1994.(Historically influential but, at times, idiosyncratic treatment of mathematical and logical foundations for predicative analysis.)

Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press; 2nd edn, 1925.(The classic effort to demonstrate the logicistic thesis; highly technical.)

Wright, C. (1982) ‘Strict Finitism’, *Synthese* 51; repr. in *Realism, Meaning and Truth*, Oxford: Blackwell, 1986, 107-75.(A detailed reconsideration of issues raised in Dummett (1975).)

Content: wide and narrow

A central problem in philosophy is to explain, in a way consistent with their causal efficacy, how mental states can represent states of affairs in the world. Consider, for example, that wanting water and thinking there is some in the tap can lead one to turn on the tap. The contents of these mental states pertain to things in the world (water and the tap), and yet it would seem that their causal efficacy should depend solely on their internal characteristics, not on their external relations. That is, a person could be in just those states and those states could play just the same psychological roles, even if there were no water or tap for them to refer to. However, certain arguments, based on some imaginative thought experiments, have persuaded many philosophers that thought contents do depend on external factors, both physical and social. A tempting solution to this dilemma has been to suppose that there are two kinds of content, wide and narrow. Wide content comprises the referential relations that mental states bear to things and their properties. Narrow content comprises the determinants of psychological role. Philosophers have debated whether both notions of content are viable and, if so, how they are connected.

1 Motivating the distinction

Many of our mental states, including our perceptions, memories and beliefs, represent things in the world and properties of those things (see Propositional attitudes). But this fact seems to produce a dilemma. On the one hand, representational states are what they are in virtue of how they represent things to be. That is, they are most appropriately individuated (distinguished from one another) by their different contents, which are comprised of things in the world (objects and their properties and relations; see Intentionality). On the other hand, in so far as these states play causal roles in our psychologies, they do so because of how the world seems to us, and presumably that is solely a matter of what is ‘in the head’. Their psychological roles would be just the same even if one were a brain in a vat or a victim of Descartes’ evil demon. The dilemma, then, is to reconcile the apparent facts that the representational character of mental states depends on their external relations and that their causal efficacy depends on their internal properties (see Methodological individualism). The distinction between wide and narrow content has been thought to resolve this dilemma.

Is this a genuine dilemma? Leaving aside such sceptical suggestions as that the contents of psychological states have no causal efficacy (see Mental causation), one might argue that the dilemma is illusory because, strictly speaking, contents *are* internal properties. Contrary to what many philosophers have supposed, the mere fact that the notion of content is connected with notions such as reference and truth-conditions does not mean that content is external. For this is compatible with a Fregean conception of content. Frege held that linguistic expressions have sense as well as reference, and the same might be suggested about concepts. Each concept expresses a condition of reference (a sense) and what it refers to is whatever satisfies that condition; and this condition is not dependent on external features of the thinker’s situation and is not otherwise sensitive to contextual factors (see Sense and reference). If concepts are Fregean in this way, then any two thinkers entertaining thoughts with the same conceptual composition are, regardless of their external circumstances, thinking thoughts with the same truth-conditions. This rules out the possibility that, because of sensitivity to circumstances, one could be thinking something true and the other, something false.

One difficulty with this Fregean view is its inability to handle thoughts that are about particular individuals. Thoughts such as ‘That’s a canary’ seem to be essentially indexical, in that their truth-conditions are relative to their contexts of occurrence (see Content, indexical): different people, having qualitatively identical experiences, could think thoughts of that form but be thinking of different birds.

A similar problem for the Fregean picture is posed by recognitional concepts: one might see a certain exotic insect and think, referring to its type, for which one might have no name, ‘That must be indigenous to the Amazon’. One might later observe a similar specimen and think, ‘It must be another one of those’, referring to the same type. One would be mistaken if, no matter how much the second looks like the first, it is actually an insect of a different type. Thus the property of being of the type initially picked out is not a matter of fitting a certain image or conception. Whether or not one’s concept of that type applies to the next specimen depends not on whether it fits one’s conception but on whether it is in fact a thing of the same sort as the one originally picked out. So it seems that the Fregean picture does not apply to recognitional concepts.

2 Twin earth thought experiments

A similar difficulty was revealed by Putnam's celebrated 'twin earth' thought experiments (1975), which introduced the distinction between wide and narrow content (see Putnam, H.). Twin earth is a place where everything is just like it is on earth, except as otherwise specified. In one scenario, two men, Art and Bart (Art's counterpart on twin earth), each have thoughts that they express with the words 'Water quenches thirst'. However, on twin earth the clear liquid that fills the seas and falls from the skies is composed not of H_2O but of some other stuff, XYZ. It is 1750 and Art and Bart, like everyone else, are ignorant of chemistry and could not, even if they had the opportunity, tell the difference between H_2O and XYZ. Nevertheless, Putnam contends, Art and Bart use the word 'water' to express different concepts. If Art were to classify a sample of XYZ as water, he would be wrong, for it would not be a substance of the same kind as water. As for Bart, he does not take XYZ to be water, for he does not have the concept 'water'. So Art and Bart, even though they are not different neurally, have different concepts, with different conditions of correct application. Their concepts are not determined solely by what is in their heads (see Concepts §5; Methodological individualism).

Burge (1979) conducts another thought experiment designed to show that differences in people's social environments can make for differences in mental content. An arthritic patient called Al complains, 'My arthritis has spread to my thigh'. Nothing in his acquisition of the term 'arthritis' has kept him from supposing that this inflammatory disease can occur in the bones as well as the joints. Meanwhile, Cal, his twin earth counterpart, registers a similar complaint. There, however, the term 'arthritis' is used to refer to an inflammatory disease of either the joints or the bones. Cal's exposure to the term 'arthritis' is the same as Al's, but, given how it is used on twin earth, he understands it correctly. Now, according to Burge, both patients are correctly said, on their respective planets, to believe that arthritis can occur in the bones, but, since Cal's belief is true and Al's is false, what they believe is different. However, there is no internal difference between them. Therefore, what they believe is partly an external matter. Contents are not, and do not supervene upon, what is in the head.

These thought experiments have met with considerable enthusiasm but also with neglected criticism (see Unger 1984; Bach 1987; Crane 1991). For one thing, they are conducted selectively: varying their details can yield contrary intuitions, for example, that XYZ is a kind of water. Also, even granting the correctness of the intuition, for example, that XYZ is not a kind of water, they leave it a mystery why the references of different concepts should be determined in different ways. The reference of the concept we express with 'water' depends on the nature (H_2O) of the clear, plentiful liquid around us, but this is not the situation with the concepts we express with the terms 'earth', 'air' and 'fire', whose references are chemically heterogeneous and are determined by the satisfaction of certain phenomenological conditions. Putnam and his followers do not explain why the fact that water is a chemical natural kind, and earth, air and fire are not, should make the concept water a different kind of concept from the concepts earth, air and fire, with its reference determined in a fundamentally different way, even back in 1750.

The arthritis argument depends essentially on the supposition that one can have beliefs with contents one 'incompletely understands'. It assumes, for example, that Al not only misunderstands the word 'arthritis' but operates with the concept arthritis rather than with some broader concept (call it 'tharthritis') that he mistakenly associates with the word. So, it might be objected, Al understands the term 'arthritis' in precisely the same way as Cal does, and has the very same belief, namely that his tharthritis has spread to his thigh. Whatever evidence there is that he also believes that his arthritis has spread to his thigh is overridden by his idiosyncratic understanding of the term 'arthritis' (we are not tempted to say that he believes that he has inflammation of the joints in his thigh).

In any case, we cannot assume that what 'that'-clauses capture is the sort of content relevant to psychology, that is, to characterizing people's perspectives and explaining their actions and inferences. As Loar (1988) and Patterson (1990) have both argued, one can grant that the thought experiments succeed in showing that the truth-conditions of attitude ascriptions are sensitive to aspects of the physical and social environment without granting that 'that'-clauses capture psychological content (see Propositional attitude statements). After all, semantics (of natural language) is not psychology.

3 Two kinds of content?

Burge is satisfied that there is only one kind of content, the externalist kind revealed by the thought experiments

and specified by 'that'-clauses in attitude ascriptions, and that no other kind is needed for psychology. Other philosophers, such as Loar (1988) and Block (1986), accept the thought experiments but propose another kind of content - narrow as opposed to wide - which captures the person's subjective point of view and serves the purposes of psychological explanation. Narrow contents capture what earthlings and their twin earth counterparts have in common, and explanations that ignore narrow contents miss crucial generalizations. If so, what is narrow content and how is it connected to wide content?

Loar and Block take narrow content to be 'conceptual role', which is defined in terms of a concept's inferential connections to other concepts. The main challenge for this view is to find a non-arbitrary way of constraining the relevant connections, so that each psychological state can turn out to possess a determinate narrow content, and to explain how this constrains its truth-condition (see Semantics, conceptual role). Another conception of narrow content, originated by White (1982) and championed by Fodor (1987), is that narrow content is a function from context to wide content. One challenge for this approach is to define the operative notion of context and to specify narrow contents informatively, rather than by abstraction from wide contents. Also, although the distinction between wide and narrow content acknowledges a systematic discrepancy between ordinary attitude attributions and scientific psychological explanation, and is motivated by a respect for both, one might wonder whether there really are two kinds of content or merely one kind described in two different ways.

It is difficult to assess the competing views because none of them has yet been developed in any great detail. A plausible if tentative assessment is that the distinction between wide and narrow content is well-motivated but not well-formulated. It is well-motivated since, in many cases, a thought's truth-condition is not wholly determined by what is in the head, and yet what is in the head does determine, independently of environmental factors, the thinker's perspective. Also, as Frege first recognized, wide content is too coarse to distinguish distinct perspectives on the same state of affairs or otherwise mark relevant differences in cognitive role. Realizing this, opponents of narrow content have suggested syntactic form or computational role as a surrogate for narrow content. However, this suggestion cuts things too finely: it fails to reckon with the possibility that mental representations of different forms or different computational roles might nevertheless embody the same cognitive perspective.

See also: Content, indexical; Holism: mental and semantic; Methodological individualism; Semantics

KENT BACH

References and further reading

Bach, K. (1987) *Thought and Reference*, Oxford: Oxford University Press.(Chapter 13 challenges Putnam's and Burge's twin earth thought experiments.)

Block, N. (1986) 'Advertisement for a Semantics for Psychology', in P. French *et al.* (eds) *Midwest Studies in Philosophy*, vol. 10, Minneapolis, MN: University of Minnesota Press.(Surveys a wide variety of theories and defends a conceptual role theory of narrow content.)

Burge, T. (1979) 'Individualism and the Mental', in P. French *et al.* (eds) *Midwest Studies in Philosophy*, vol. 4, Minneapolis, MN: University of Minnesota Press.(Employs a variety of twin earth thought experiments to argue that mental contents are to a degree socially constituted.)

Crane, T. (1991) 'All the Difference in the World', *Philosophical Quarterly* 41: 1-25.(Challenges Putnam's and Burge's twin earth thought experiments.)

Fodor, J. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.(Explores White's idea that narrow content is a function from context to wide content.)

Fodor, J. (1994) *The Elm and the Expert*, Cambridge, MA: MIT Press.(Argues that wide content is the only notion of content needed for the purposes of psychological explanation.)

Loar, B. (1988) 'Social Content and Psychological Content' and 'Reply to Bilgrami', in R. Grimm and P. Merrill (eds) *Contents of Thoughts*, Tucson, AZ: University of Arizona Press.(Argues, contrary to Burge, that psychological content is not in general what is captured by oblique 'that'-clauses.)

Loewer, B. and Rey, G. (eds) (1991) *Meaning in Mind: Fodor and His Critics*, Oxford: Blackwell.(Articles, with replies by Fodor, addressing main issues in the theory of content.)

Patterson, S. (1990) 'The Explanatory Role of Belief Ascriptions', *Philosophical Studies* 59: 313-32.(Argues that the orthodox interpretation of Burge's twin earth thought experiments does not do justice to the ways in which intentional states are individuated in common-sense psychology.)

Putnam, H. (1975) 'The Meaning of "Meaning"', in K. Gunderson (ed.) *Language, Mind, and Knowledge*, Minneapolis, MN: University of Minnesota Press.(Devises twin earth thought experiments to argue that

meanings cannot both determine reference and be ‘in the head’.)

Stich, S. and Warfield, T. (eds) (1994) *Mental Representation: A Reader*, Oxford: Blackwell.(Includes important articles on major theories of content.)

Unger, P. (1984) *Philosophical Relativity*, Minneapolis, MN: University of Minnesota Press.(Chapter 5 challenges the twin earth thought experiments by showing how varying their details yields conflicting intuitions.)

Wallace, J. and Mason, H.E. (1990) ‘Some Thought Experiments about Mind and Meaning’, in C.A. Anderson and J. Owens (eds) *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*, Stanford, CA: Center for the Study of Language and Information, Stanford University.(Argues that ordinary language provides much richer and subtler ways of attributing mental contents than is allowed by the orthodox interpretation of Burge’s twin earth thought experiments.)

White, S. (1982) ‘Partial Character and the Language of Thought’, *Pacific Philosophical Quarterly* 63: 347-65. (Proposes a conception of narrow content as a function from context to wide content.)

Content, indexical

Many of our thoughts are about particular individuals (persons, things, places, ...). For example, one can spot a certain Ferrari and think that it is red. What enables this thought to latch onto that particular object? It cannot be how the Ferrari looks, for this could not distinguish one Ferrari from another just like it. In general, how a thought represents something cannot determine which thing it represents. What a singular thought latches onto seems to depend also on features of the context in which the thought occurs. This suggests that its content is essentially indexical, contextually variable much as the content of an utterance such as 'I am hungry' depends on who utters it and when. The indexical model of singular thought is not limited to thoughts about individuals one perceives, but applies also to thoughts about individuals one remembers or has been informed of, such as an old bicycle or Christopher Columbus. In each case, a certain contextual relation, based on perception, memory or communication, connects thought to object.

Our thoughts do not merely represent what the world is like. Many of our thoughts manage to latch onto particular individuals (persons, things, places,…), and how this is accomplished needs to be explained. If, for example, one thinks of the bird on the window sill that it is a sparrow (the thought is expressible by 'That bird is a sparrow'), there must be something about the thought which makes it about *that* bird (see Demonstratives and indexicals). It might seem that the object of a thought must be uniquely determined by some component of the thought. This view is analogous to Frege's view that the 'sense' of a linguistic expression determines its reference (see Frege, G. §3; Sense and reference). The idea is that *how* something is represented determines which thing is represented. That is, the subject component of the thought, expressible in this example by the demonstrative description 'that bird', imposes a certain condition whose satisfaction by a certain individual makes it the object of the thought. This condition is the same no matter who entertains the thought, no matter where or when. The trouble with this view is that one might not represent the thing in question with enough specificity to single it out from all other birds. Indeed, one might even misrepresent it - perhaps what one takes to be a sparrow is actually a porcelain figure.

Here is an alternative view. An individual is singled out as the object of the thought not because it satisfies a condition represented in the thought but because the thought is causally connected to it. That is, as Tyler Burge (1977) has suggested, a thought latches onto something by way of bearing a certain contextual relation to that particular individual. If this is correct, then facts about the circumstances in which the thought occurs, and not just atemporal facts about its character, are relevant to the determination of its object. Two widely separated people with similar perceptual experiences (or the same person at different times) could have thoughts of exactly the same character, expressible by 'That bird is a sparrow', and yet be having these thoughts about different birds. One might be correct, the other mistaken. This suggests that singular thoughts are true or false only relative to a context. As John Perry (1993) describes them, they are 'essentially indexical'.

To explain how this can be, we need to distinguish the type of thought that both people have from the particular occurrence (or token) of that type which each has (see Type/token distinction). Tokens of the same thought type can be about distinct objects because singular thoughts are 'token-reflexive', much like indexical expressions. For example, the reference of the indexical 'I' depends on who utters it; of 'here' on where it is uttered; and of 'tomorrow' on when it is uttered. As a result, the sentence 'I will be here tomorrow' is neither true nor false in itself but only relative to a particular tokening of it. That is, a given utterance of it is true if the maker of that utterance will be at the place of the utterance on the day after the utterance. Different people, or the same person at a different place or time, would say different things with these words. Notice that there is also a sense in which everyone who utters 'I will be here tomorrow' says the same thing - utterances of the sentence are type-identical, but their truth-conditional contents differ in the way just described. Singular thoughts are similarly token-reflexive and indexical in content, because their objects are determined only relative to their context of occurrence. The object is not thought of as the unique object of a certain sort, that is, under a definite description (see descriptions), but in some contextually sensitive way. On this indexical view of singular thought, a thought represents a certain object in a context but without representing the context itself. The truth-condition of a particular token of the thought type, such as the one expressible as 'That bird is a sparrow', is not absolute but is contextually variable, in this case varying with the context of perception.

Not only can singular thoughts be about objects of perception, they can also be about individuals one remembers or has been informed of. François Recanati (1993) and Kent Bach (1987) both suggest that these cases involve uses of ‘mental indexicals’. One suggestive idea, developed most fully by Graeme Forbes (1989), is that, when we are introduced to (or otherwise learn of) an individual by name, we form a ‘mental file’ labelled with the name. We can later call up the individual by name. These mental files on persons and things are relatively permanent, but, as Recanati suggests, we seem to operate also with temporary files, labelled with mental counterparts of pronouns, on individuals currently singled out in perception or in conversation.

See also: Content: wide and narrow; Reference

KENT BACH

References and further reading

Bach, K. (1987) *Thought and Reference*, Oxford: Oxford University Press, chaps 1, 2.(Develops a relational account of perception-, memory- and communication-based singular thought, according to which percepts, memories and names [as labels on mental files] function as mental indexicals.)

Burge, T. (1977) ‘Belief *de re*’, *Journal of Philosophy* 74: 338-62.(Contends that singular thoughts are not reducible to descriptional thoughts and are true or false only relative to contexts.)

Burge, T. (1979) ‘Sinning against Frege’, *Philosophical Review* 88: 398-432.(Argues that Frege’s theory of context-insensitive senses, if properly understood and suitably refined, is not vulnerable to certain objections based on indexical thoughts.)

Evans, G. (1982) *The Varieties of Reference*, Oxford: Oxford University Press, chaps 5, 6.(Proposes an information-based account of singular thought.)

Forbes, G. (1989) ‘Cognitive Architecture and the Semantics of Belief’, in P. French *et al.* (eds) *Midwest Studies in Philosophy*, vol. 14, Minneapolis, MN: University of Minnesota Press.(Proposes an account of singular thought, as well as thoughts about natural kinds, according to which the subject term takes the explicitly indexical form, ‘the subject of this dossier’.)

Perry, J. (1993) *The Problem of the Essential Indexical and Other Essays*, Oxford: Oxford University Press. (Collects the author’s influential papers on irreducibly indexical aspects of content.)

Recanati, F. (1993) *Direct Reference: From Language to Thought*, Oxford: Blackwell, chaps 5, 6.(Develops an account of non-descriptive modes of presentation, on which names function as labels on permanent mental files and indexicals as labels on temporary files.)

Content, non-conceptual

To say that a mental state has intentional content is to say that it represents features of the world. The intentional content of a belief can be characterized in terms of concepts: the content of the belief that fish swim is characterized by the concepts 'fish' and 'swimming'. The contents of beliefs are, for this reason, often described as conceptual. One way to explain this idea is to say that to have a belief, one has to possess the concepts which characterize the belief's content. However, some philosophers believe that certain mental states have non-conceptual contents: these states represent the world without the subject having to possess the concepts which characterize their contents. The main examples of these putative states are conscious perceptual experiences and the non-conscious states of cognitive information-processing systems (such as the visual system).

1 Conceptual and non-conceptual content

Beliefs and other propositional attitudes are individuated in terms of their contents. It is natural to think that having these propositional attitudes involves having concepts, construed here as items in the mind. If someone believes that *a is F*, then they must have the concept *F*, and the concept *a* - whatever concepts may be. This can be summed up by saying that beliefs have conceptual contents (see Propositional Attitudes). Some philosophers, for example Evans (1982) and Peacocke (1993), think that not all intentional mental states have conceptual contents. They hold that some intentional states have 'non-conceptual' contents: the contents of these states do not, in some sense, involve concepts.

How can a thinker's state have the content that *a is F* without the state 'involving' the concepts *a* and *F*? The best way to answer this question is to focus on the idea of 'possessing' concepts. In the case of states with conceptual contents - like beliefs - we can say that in order for a subject to be in these states, the subject has to possess certain concepts. Likewise, we can say that in order to be in a state with *non*-conceptual content, a subject does not have to possess certain concepts. But which concepts? It would be too strong to say that being in non-conceptual states does not require having any concepts at all - this issue should not be settled by the very definition of non-conceptual content. Rather, we should say that for a subject *S* to be in a non-conceptual state with content *P*, *S* does not have to possess the concepts which *S would* have to possess if *S were* in a conceptual state with content *P*.

If we call these concepts the concepts which are 'canonical' for *P*, then we can say that a state with non-conceptual content is one of which the following is true:

in order for a subject, *S*, to be in a state with a content *P*, *S* does not have to possess the concepts canonical for *P*.

The idea of concepts canonical for a certain content is just the idea that there are concepts which *essentially* characterize a given content (Cussins 1990: 382-3). The content expressed by the sentence 'snow is white', for example, is essentially characterized in terms of the concepts expressed by the words 'snow' and 'white'.

Why should anyone be interested in the category of non-conceptual content? Why should it be of any more philosophical interest than the category of non-red things? The reason is that intentionality seems closely bound up with having concepts. So it is a substantial thesis that there can be intentionality without possession of (the relevant) concepts (see Intentionality; Concepts). However, simply defining 'non-conceptual content' leaves open the question of whether there actually is any. So are there any states with non-conceptual content?

2 Varieties of non-conceptual content

The definition of non-conceptual content just given applies straightforwardly to some notions of content. States which carry 'information' (in Dretske's (1981) sense) have non-conceptual contents, since a state's carrying information is a matter of its co-varying in a nomic or reliable way with a certain phenomenon. For example: on this view we can say that a weather vane represents or indicates the direction of the wind. But it is obvious that the weather vane does not possess the concept of the direction of the wind.

However, it is debatable whether a state's carrying information is sufficient for the state to have representational content in any interesting sense. For one thing, since nomic co-variation is present wherever there is causation, this theory stretches the notion of representation to the point where it can apply to almost everything in the universe. Another problem is that informational states seem incapable of representing incorrectly; but the possibility of

incorrect representation seems essential to genuine representation. Something more must be added to turn mere informational states into states with genuine representational content (see Semantics, informational).

Two other kinds of state are claimed to have non-conceptual contents: states of the information-processing cognitive systems (for example the visual system) postulated by many psychological theories; and conscious perceptual experiences. Computational psychological theories claim that systems within the brain perform computations: that is, these cognitive systems compute functions by processing representations algorithmically. So the thinker (or the cognitive system) is in certain representational states, or states with content. These representational states are essentially specified in terms of certain concepts, but there is no need for the thinker to possess these concepts in order to be in these states. For example, Marr's (1982) theory of vision analyses visual information processing in terms of complex mathematical concepts; one need not master these concepts in order to process visual information. So these computational states, if they exist, are non-conceptual (see Mind, computational theories of; Vision).

The other kind of states which have been attributed non-conceptual contents are conscious perceptual experiences. The general idea here is that in perception many aspects of the world are presented to the perceiver; yet there is no need to suppose that the perceiver has a distinct concept for each aspect of the world which is so presented. Consider colour experience: is it plausible that each of us has a distinct concept for each precise shade of colour we are able to perceive? If it is not, this can be a reason for holding that perceptual experiences have non-conceptual contents: perception has a 'phenomenological richness' which is not constrained by the concepts the perceiver has (Evans 1982: 229-; other motivations for attributing non-conceptual contents to experiences are offered in Crane 1992 and Martin 1992).

Given that experiences have non-conceptual contents, how should this be explained? One of the most detailed accounts has been given by Peacocke (1993: ch. 3). At one level, the content of a perceptual experience is what he calls a 'scenario': a set of ways of filling out the space around a perceiver with properties and relations - for instance, with colour and shape properties. The experience's representational content is given by the scenario, because the experience is correct just in case the actual distribution of properties and relations around the perceiver belongs in the scenario. However, there is no requirement that the perceiver has all the concepts which essentially characterize the properties and relations in the scenario.

Opinions differ on whether the content of experience is wholly conceptual, wholly non-conceptual, or some mixture. Evans (1982) holds that it is wholly non-conceptual, while Peacocke (1993) holds that experiences can have many layers of content, some of which are conceptual and some non-conceptual (see Perception).

3 Problems with non-conceptual content

The notion of non-conceptual content could be criticized on two fronts. First, the application of the notion to a particular kind of mental state (perception, for example) could be criticized. Second, the very notion of non-conceptual content itself could be criticized.

John McDowell (1994: ch. 3) argues that to treat the content of experience as non-conceptual is to commit oneself to the idea which Wilfrid Sellars criticized as 'the myth of the given': the idea that experience involves being presented with an unconceptualized 'given' which the mind then goes on to conceptualize (Sellars 1956). McDowell argues that accepting this picture renders the relation between mind and world deeply problematic. By contrast, he argues that the content of experience is wholly conceptual. McDowell therefore accounts for the phenomenological richness of experience in a different way: he claims that where we have (for example) discrimination between colours, we may not have a distinct word for each colour, but we do have 'a recognitional capacity, possibly quite short-lived, that sets in with the experience' (1994: 57). And such a recognitional capacity is, he argues, fully conceptual.

Scepticism about the very idea of non-conceptual content can arise out of scepticism about the claim that there are concepts one must have if one is to be capable of having certain intentional states. Accepting this claim amounts to accepting a mentalistic version of a sharp distinction between analytic and synthetic statements - a distinction which has proved highly controversial (see Analyticity).

TIM CRANE

References and further reading

Crane, T. (1992) ‘The Nonconceptual Content of Experience’, in T. Crane (ed.) *The Contents of Experience,* Cambridge: Cambridge University Press.(Defines non-conceptual content and argues that perceptual experience has non-conceptual content, on the basis of a contrast between the different inferential roles of perception and belief. Sections 1-3 provide a straightforward introduction to the topic of non-conceptual content as understood in §1 above.)

Cussins, A. (1990) ‘The Connectionist Construction of Concepts’, in M. Boden (ed.) *The Philosophy of Artificial Intelligence,* Oxford: Oxford University Press.(An attempt to show how states with conceptual content emerge out of states with non-conceptual content.)

Dretske, F.I. (1981) *Knowledge and the Flow of Information*, Oxford: Blackwell.(A physicalist theory of intentional content, based on the mathematical concept of information. Though it presents an original theory, this is a very clearly written book.)

Evans, G. (1982) *The Varieties of Reference*, Oxford: Clarendon Press.(Influential and wide-ranging account of reference and thought. At various points in the book the claim that experiences have non-conceptual content is defended. Difficult in parts.)

Marr, D. (1982) *Vision*, San Francisco, CA: Freeman.(Pioneering and influential computational theory of vision. After the introduction, some knowledge of mathematics is needed.)

Martin, M.G.F. (1992) ‘Perception, Concepts and Memory’, *Philosophical Review* 101: 745-63.(Argues that perceptual experiences have non-conceptual contents, based on an examination of the relation between memory and perception.)

McDowell, J. (1994) *Mind and World*, Cambridge, MA: Harvard University Press. (Chapter 3 argues that experience is fully conceptual. A difficult work, though not technical.)

Peacocke, C. (1993) *A Study of Concepts*, Cambridge, MA: MIT Press.(Lucid though difficult, chapter 3 defends an account of the non-conceptual content of perceptual experience.)

Sellars, W. (1956) ‘Empiricism and the Philosophy of Mind’, in H. Feigl and M. Scriven (eds) *Minnesota Studies in the Philosophy of Science*, vol. 1, Minneapolis, MN: University of Minnesota Press.(Classic statement of Sellars’ argument against the ‘myth of the given’. Difficult.)

Contextualism, epistemological

The idea that norms vary with social setting has long been recognized, but it is only in the late twentieth century that philosophers have developed precise versions of epistemological contextualism, the theory that standards of knowledge and justification vary with context. Ordinary practice seems to support this rather than the 'invariantist' view that epistemological standards are uniform.

Suppose, for example, that having seen my children a minute ago, I assert 'I know my children are in the garden'. My neighbour Harold then says, 'Good, because an escaped prisoner is seeking hostages nearby'. I may then appropriately claim, 'On second thoughts, I do not know, I should check carefully'. Standards for knowledge appear to have shifted, since they now require further investigation.

Contextualism's greatest advantage is its response to scepticism. Sceptics raise radical possibilities, such as that we might be dreaming. The contextualist grants that such doubts are legitimate in the sceptical context, but holds they are illegitimate in everyday situations. Yet contextualism can appear to be an objectionable form of relativism, and may be accused of confusing standards that we apply in practical conversational contexts with the true standards that determine whether someone has knowledge.

1 Elements of the theory

Epistemological contextualism holds that standards of knowledge and justification vary with context. One form of the theory, call it 'subject-based contextualism', holds that epistemic standards depend on the knower's context. A more radical form, 'attributor-based contextualism', treats standards as varying with contexts of attribution. On this latter view, one attributor may legitimately claim that a person knows something while another may legitimately deny that the same person has the knowledge in question. Suppose Susan asserts, 'I know your children are in your garden; I saw them there a minute ago'. When my concern is to call the children to dinner, I attribute knowledge to Susan. But Harold, aware of an escaped prisoner seeking hostages nearby, would not attribute knowledge to her. What Susan knows depends on attributors' contexts, not only on her situation.

Contextualists also vary regarding contextual elements that determine standards. In the examples above, the relative importance of having correct information determines the strength of justification required. Other factors affecting epistemic standards might include: (1) doubts entertained; (2) doubts mentioned; (3) the topic under consideration; and (4) elements of the situation that affect the knower's reliability. David Lewis (1996) suggests that we follow a 'rule of accommodation', such that we tend to adjust our standards so as to count knowledge claims as true. Michael Williams (1991) argues for 'methodological necessities' which exempt propositions from doubt if considering them demands a shift in disciplinary focus. For instance, a molecular biologist may ignore concerns that shift the topic to particle physics.

Some contextualists believe that knowledge has a foundational structure (see Foundationalism). Foundationalists typically claim that knowledge is grounded in basic beliefs which have intrinsic epistemic credibility, and have that credibility as a result of belonging to some type of beliefs, such as beliefs about sensory experience. Contextualists deny both claims. Basic non-inferential beliefs have their status because they are permissible in context, and the belief type that is basic varies with context. We might expect that when discussing where to eat lunch, for example, basic beliefs include beliefs about ordinary material objects, whereas in conversation between physicists, basic beliefs include those about widely accepted theoretical views.

2 Advantages of contextualism

Beyond its compatibility with everyday practice, contextualism provides attractive responses to major philosophical problems. Here, let us consider the regress argument and scepticism. Suppose we hold that knowledge requires justification. The regress argument for foundationalism examines the nature of such justification. Either a belief has intrinsic credibility, and so is 'self-justifying', or it is justified by inference. If the latter, then the belief is ultimately inferred from (1) intrinsically credible beliefs, (2) unjustified beliefs, (3) an infinite regress of justification, or (4) a circular justification in which some beliefs appear in their own support. Foundationalists claim (1) is the only acceptable alternative, but they have had problems explaining and defending the notion of intrinsic credibility.

Those contextualists who allow a foundational structure propose an alternative resolution: justification ends in beliefs that are unjustified in their context. This is made plausible in two ways. First, the contextualist cites contextual factors that make it reasonable for agents to limit justification. For example, where all parties to an investigation agree on certain beliefs, there is no need to provide further justification. The contextualist may also appeal to Williams' concept of methodological necessity, according to which asking for justification can force inquirers to ignore disciplinary constraints. Second, contextualists claim that while justification ends with unjustified belief, demand for further justification can always be legitimated by special doubts about the unjustified belief.

The contextualist response to scepticism is even more attractive. Contextualists claim that sceptical doubts are legitimate, but only in the sceptical context (see Scepticism). Williams (1991) gives an illuminating response along these lines. He contends that scepticism presupposes 'epistemological realism', the view that there is only one correct order of appropriate epistemic inference. Sceptics believe that knowledge of 'the external world' must be justified on the basis of immediate experience. If so, we must argue from beliefs about sense impressions, or how things appear, to beliefs about the material world. Williams claims this order of inference is necessary only within the sceptical context. In other contexts, it is permissible to begin our justifications with beliefs about the material world; and in some situations, such as psychological investigations of perception, we may begin with beliefs about external objects and draw inferences about appearances. Epistemological realism is false, since context determines the proper order of inference.

Keith DeRose (1995) proposes, in contrast, a contextualist 'conditional theory' of knowledge. Suppose *S* has a true belief *p*. Where *S* would not believe *p* if *p* were false, the belief is 'sensitive'; where *S* would believe *p* even if *p* were false, the belief is 'insensitive'. DeRose endorses a 'rule of sensitivity': whenever someone asserts that *S* knows *p*, standards of knowledge are raised to require *S*'s belief *p* to be sensitive and the resulting standard applies for all beliefs in that context.

Suppose the belief *p* is that '*S*'s house is red'. If someone claims *S* knows *p*, the rule of sensitivity requires that if *S*'s house were another colour, *S* would not believe *p*. Philosophers typically evaluate subjunctive conditionals by considering whether, in possible worlds 'close' to the actual world but in which the antecedent of the subjunctive is true, the consequent is also true. 'If not *p*, *S* would still believe *p*', is true if, in worlds similar to our own but in which *S*'s house is not red, *S* can tell the colour of the house. Intuitively, at close worlds, *S* can determine the colour if *S* employs normal perceptual abilities. In contrast, let the belief *q* be 'I am a brain in a vat', where we have in mind the sceptical possibility that a brain is connected to a complex computer which simulates everyday experience. If someone claims *S* knows not-*q*, the rule of sensitivity demands that if *q* were true, *S* would not believe not-*q* (that is, *S* would believe *q*). This requires great epistemic ability, for we evaluate it by considering whether *S* believes *q* in worlds where *S* is a brain in a vat. The closest such world is far from the actual world, since in the actual world we cannot find out whether we are brains in vats. Once the rule of sensitivity raises standards so that we must consider such worlds, it does so for other beliefs in the context, including *p*. But in a world where *S* is a brain in vat, *S* will believe *p* even though *p* is false - in that world, *S* cannot determine the colour of anything! Thus, in contexts where *p* alone is considered, *S* knows *p*. But when *S* entertains *p* and a sceptical hypothesis, the rule of sensitivity raises standards to a level at which ordinary beliefs such as *p* are no longer known.

These responses to scepticism have the consequence, which Williams calls 'instability' and Lewis labels 'elusiveness', that when we raise sceptical doubts about ordinary knowledge, we change context and thereby 'destroy' that knowledge. In the sceptical context, the sceptic wins the debate. But this may be advantageous for contextualism, since it can explain both our judgment that we have knowledge in everyday situations and our feeling that sceptical doubts should be taken seriously.

3 Contextualism as relativism

In the case of contextualism, two objections underlie the charge of relativism. First, contextualism seems to accept statements of this form: '*S* knew *p*, but now, *S*'s evidence and reliability staying the same, *S* does not know *p*'. How can people know something at one time but not know it at a later time if they have the same justification they always had?

The second objection is more pressing. Straightforward relativists about truth - those who claim a proposition may be true for one group yet false for another group - appear to hold that a single proposition can be both true and false. This seems incoherent; the notion of truth is no longer in play but has been replaced by 'truth for' or 'truth from a point of view'. Subject-based contextualism escapes this worry, for while '*S* knows *p*' is true or false depending on *S*'s context, '*S* knows *p* in context *C*' is true from all points of view. The attributor-based contextualist, however, holds that truth of statements having the form '*S* knows *p*', or '*S* knows *p* in context *C*' (where '*C*' refers to the subject *S*'s context) are relative to attributors' context. It seems that there is no truth about whether *S* knows *p*, but only 'truth for' or 'truth from a point of view'.

To answer these objections we must treat 'knows' as an indexical term, analogous to 'I' or the demonstrative 'that' (see Demonstratives and indexicals). Following David Kaplan (1989), we may distinguish the 'character' of an indexical, the part of meaning that is constant, from the 'content', which changes. Roughly, the character of 'I' is a function from context to speaker, and the content of 'I' is the speaker on a given occasion of use. According to DeRose (1992), the character of '*S* knows *p*' requires that *S* should believe *p*, that *p* should be true, and that *S* should be in good enough epistemic position with respect to *p*. But what is 'good enough' varies with context, hence the content of a knowledge attribution is the level of epistemic position that counts as good enough in that context.

For example, the character of 'knows' in (*w*) 'Clara knows there is water in the glass' will require that Clara should believe there is water in the glass, that this belief should be true and that Clara should be in a good enough position to determine that there is water in the glass. But whether Clara is in a good enough position will vary with the attributor's context. Where the attributor's interest is in watering plants, the content of 'know' in (*w*) requires that Clara's position is good enough if she has the typical person's ability to identify water, in which case (*w*) is true. If, however, the attributor is engaged in important experiments, the content of 'know' in (*w*) requires that Clara's position is good enough only if she performed a chemical analysis, in which case (*w*) is false (assuming she has not done the analysis). There is no more mystery about the way truth conditions of (*w*) vary with context than there is about the way truth conditions of 'I lost my keys' and 'That is a rose' vary with context.

The second objection to relativism is easily handled, then: truth is relative to the attributor's context, but the notion of truth is preserved by treating knowledge claims as having an indexical component. What about the first objection? In any context, epistemic standards are set at one level, so we cannot employ two different sets of standards in evaluating a single sentence within that context. Therefore, we will reject statements of the form '*S* knew *p*, but, *S*'s evidence and reliability remaining the same, *S* no longer knows *p*'. Nevertheless, a corresponding 'metalinguistic' statement may be correct: '"*S* knows *p*" was true, but now, *S*'s evidence and reliability remaining the same, "*S* knows *p*" is false'. The contextualist argues that this is unobjectionable, given the indexicality of knowledge attributions.

4 Further objections

Let us consider four other objections:

(1) Knowledge may require greater uniformity than contextualism allows. Edward Craig (1993) has argued that the function of knowledge claims is to tag reliable information. But if epistemic standards regularly shift, we cannot rely on information gained in another context without continually re-evaluating our knowledge in each new situation. At the least, we need uniform standards of evaluation across broad areas of inquiry. The contextualist reply must grant that a core set of standards applies when special contextual factors are absent.
(2) In one context, Harold may say, 'Your children are in the garden', but once he learns about an escaped prisoner seeking hostages nearby, he would not make the same assertion. Assertibility conditions for the claim have changed, yet surely truth conditions have not. Does not the same thing hold for knowledge ascriptions? Assertibility conditions for knowledge claims vary with context, but why does this show that truth conditions for knowledge claims change also? The contextualist answer depends on arguing that speakers legitimately assert *p* only when they have adequate justification for claiming to know *p*. Thus, knowledge claims may not admit of a divide between assertibility conditions and truth conditions.
(3) Contextualism that allows unjustified beliefs is problematic. If there is no ground for non-inferential beliefs, surely we have not really provided justification. Foundationalists and coherentists will claim they can account for our permitting the absence of explicit justification. First, they would endorse the above view that contextualism

conflates assertibility conditions in a practical context with truth conditions that are not contextually determined. Foundationalists might argue that basic beliefs have minimal intrinsic credibility, but must be further justifiable given a legitimate challenge. Some coherentists claim that the most coherent theory provides grounds for accepting initial cognitive output from our sensory mechanisms. This output provides basic beliefs in a context, but is always justifiable in terms of broader theory. Contextualist response to such views requires detailed arguments about the plausibility of foundationalism and coherentism.

(4) Many sceptics believe they point to facts about the human condition that are so important they throw ordinary knowledge claims into doubt. As Barry Stroud (1996) argues, we have an 'objective conception of the world' which entails that almost all our beliefs about it can be mistaken. If this is a fundamental truth, then we cannot avoid scepticism merely by noting that in certain contexts we habitually ignore sceptical worries. The reply to this objection stresses first, that contextualism grants some truth to scepticism and, second, that unlike scepticism, contextualism emphasizes the practical origins of discourse about knowledge.

See also: Justification, epistemic; Knowledge, concept of

BRUCE W. BROWER

References and further reading

Annis, D. (1978) 'A Contextual Theory of Epistemic Justification', *American Philosophical Quarterly* 15: 213-19. (One of the first clear statements of contextualism.)

Aristotle (mid 3rd century BC) *Metaphysics*, ed. W.D. Ross, Oxford: Oxford University Press, 1924, 2 vols. (Although Aristotle does not endorse all aspects of contextualism, a basis for foundational contextualism can be located in his writings: 1011a-14 are especially relevant.)

Cohen, S. (1987) 'Knowledge, Context, and Social Standards', *Synthese* 73: 3-26. (Crucial discussion of how ability as a contextual factor affects the legitimacy of doubts and defeating considerations.)

Craig, E. (1993) 'Understanding Scepticism', in J. Haldane and C. Wright (eds) *Reality, Representation and Projection*, Oxford: Oxford University Press.(Response to contextualism that develops an important alternative way of understanding the social role of knowledge claims.)

DeRose, K. (1992) 'Contextualism and Knowledge Attributions', *Philosophy and Phenomenological Research* 52: 913-29.(Contextualist account of truth conditions for knowledge claims and a good introduction to attributor-based contextualism.)

DeRose, K. (1995) 'Solving the Sceptical Problem', *The Philosophical Review* 104: 1-53. (Sophisticated contextualist response to scepticism.)

Hambourger, R. (1981) 'Justified Assertion and the Relativity of Knowledge', *Philosophical Studies* 40: 241-69. (Contains a careful examination of the relation between assertion and knowledge.

Kaplan, D. (1989) 'Demonstratives', in J. Almog, J. Perry and H. Wettstein (eds) *Themes From Kaplan*, Oxford: Oxford University Press, 481-563.(Important work on the logic of demonstratives and indexicals. The theory developed here may be extended to versions of contextualism that treat 'know' as an indexical.)

Lewis, D. (1996) 'Elusive Knowledge', *Australasian Journal of Philosophy* 74: 549-67.(Investigation of how knowledge loses its status once we reflect on it; contains an important list of contextual factors.)

Stroud, B. (1996) 'Epistemological Reflection on Knowledge of the External World', *Philosophy and Phenomenological Research* 56: 345-58.(Response to Williams by a leading sceptic.)

Unger, P. (1984) *Philosophical Relativity*, Minneapolis, MN: University of Minnesota Press.(Contains an original and carefully developed account of contextualism and invariantism.)

Williams, M. (1991) *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism,* Oxford: Blackwell. (Thorough study of scepticism that develops the contextualist alternative.)

Wittgenstein, L. (1969) *On Certainty*, Oxford: Blackwell.(Contains a foundationalist-contextualist account of justification and certainty. See especially paragraphs 82-110, 151-66 and 224-5.)

Contingency

People are often puzzled about the apparent contingency of the world. To say that something happens contingently is to say that it might not have happened, and to think of the world as contingent is to think that it might not have existed. Is it contingent? Those who ask this are asking whether there might not have been a world at all, or none that was at all like ours. Usually they are not asking whether there could have been a world which differed from ours only in its details: for example, one in which free agents made slightly different choices. That is an important question too, but a separate one. Some people reject the question of the contingency of the world as meaningless because they do not see how any answer to it could be verified. But such verificationism is controversial, and even if there is something wrong with the question it is still worth asking why people find it compelling.

On the classical medieval view, the existence of the world is contingent: it depends on God's choice. God's existence, however, is not contingent but necessary. It belongs to God's essence that he should exist. But there has long been the alternative idea, that the existence of the world is itself necessary. Many Neo-Platonists held (some still do) that the world exists because it is best that it should. If it is best that something should come about, then it must come about, and there is no room for God's choice in the matter. Other philosophers think that the world exists necessarily because it exists in virtue of scientific laws, which are themselves necessary. The idea that it is contingent, they suggest, is due to the confusion of scientific necessity with logical necessity. Logical necessity attaches only to statements - those that are instances of logical laws, often called analytic. The existence of the world itself is (scientifically) necessary; but the statement 'There exists a world like ours' is logically contingent - alternatives can be imagined without contradiction.

Spinoza also held that the world exists necessarily, but for neither of these reasons (see Spinoza, B. §4). For him it is necessary in its own right; existence belongs to the world's essence, as for the medievals it belongs to the essence of God. The necessity here is neither scientific nor logical, but metaphysical. Spinoza (1677) called the world 'God', but it lacks the personality of the orthodox God and is identified with Nature. He infers that because the world itself exists necessarily, there is no room for contingent events within it, but this is a mistake: one could hold, as Hegel did, that the nature of the world requires that there be certain contingencies within it.

It 'exists by the necessity of its nature alone': it is hard to see what this means, either for Spinoza or for his medieval predecessors who had said the same of God. It is not that it exists by definition - that we have formulated a concept of God or Nature in which existence is taken to be a defining characteristic. Spinoza was well aware (as were Anselm and Descartes) that our conceptions are likely to be defective and our definitions consequently irrelevant. It is that existence belongs to its essence as solubility in water belongs to the essence of sugar. So just as sugar is soluble necessarily, so God or Nature exists necessarily. Like Anselm and Descartes, Spinoza inferred that it exists. Aquinas and other medievals were more cautious, and held that this does not follow: what follows is that if there is a God its existence is necessary (see God, arguments for the existence of §§1, 2).

Spinoza also said that God is *causa sui*, here again following many predecessors, including Descartes (1641). This is often translated as 'cause of itself', but 'self-explanatory' might be better. Spinoza equated something's existing necessarily with its being self-explanatory, or requiring no explanation. Unfortunately this is a mistake. If to exist is part of something's essence we can still ask why it exists, just as we can ask why sugar is soluble. No doubt if it is sugar it is necessarily soluble, but we may derive that necessity from more general laws, and we may ask why there should be sugar at all. In general the fact that something is necessary does not remove the need to explain it. Complex mathematical truths can be explained in terms of simpler ones; perhaps elementary logical laws do not need explanation. But Descartes thought they did - they were fixed by God. Among those who hold that the basic laws of physics are necessary, some deny that they need explanation and others insist that they do. Some of these would feel that it strengthens, or at least fills out, their case to say that although these laws are scientifically necessary they are metaphysically contingent; in other words, that this world is itself metaphysically contingent - other worlds with alternative laws could have existed instead. But this is to repeat the mistake of equating what is necessary with what does not need explanation, and what is contingent with what does.

This mistake explains why people puzzle over the contingency of the world. What they really want to know is whether the world's existence requires explanation. It is a difficult question; but if we could solve it, any

explanation we provided would automatically tell us whether the world is contingent or not. Thus if the world is the free creation of God, it would not have existed if God had chosen differently, and it is metaphysically contingent. If, alternatively, as thought, it was created by a God who is bound to create the best of all possible worlds, there is no real possibility of God's choosing differently, and the world is metaphysically necessary (see Leibniz, G.W. §3). Leibniz (1697) refused to put the matter like this, and insisted on calling it contingent, just because it is created by God. There may be various reasons for this refusal, including a wish to continue to call God's choice 'free' even though God's own perfection left no alternative. But one of them is the thought that the world requires explanation, and that to call it necessary would be to imply otherwise.

We always seek explanations wherever possible. Leibniz (and many others) believed in a principle of sufficient reason, which says there is an adequate explanation for everything that exists or occurs. It is natural to hope that there is such an explanation; but then either there must be chains of explanations stretching back infinitely, or else there is some thing or event the nature of which removes the need for further explanation. Some have seen the Big Bang in this way, but it is evidently possible to speculate about why it occurred. There is no incoherence in the idea of something which had no cause and which did not depend on anything, but it would remain natural for us to seek an explanation for it even though there was none to be found. There is room for dispute as to what an explanation is, but if to explain the fact that Φ is to render it intelligible by reference to the fact that Ψ, it would seem that there might be some facts which were, in their own right, as intelligible as possible, like (perhaps) the fact that if *p* then *p*. The demand for further explanations could be terminated if we came to a fact of this kind. Some have thought the existence of God to be maximally intelligible, and the existence of the world not to be; but if 'intelligible' means 'intelligible to us', it seems plain that neither is maximally intelligible, and if that is not what it means, it is unclear what else it means.

A more defensible view would be that God's existence is 'self-explanatory' in that what makes his existence intelligible is his existence itself. It is arguable that we cannot explain anything except by tacitly presupposing the existence of God, because every explanation rests on certain principles of reasoning, and to suppose that these principles yield truth about the world is to suppose the sort of systematic match between the way we think and the way things are that could only be due to a purposive agent. If that were so, God's existence would be required to render anything intelligible, and if anything were proposed as an explanation of God, its explanatory power would derive from God's existence itself. This is as close as we could come, I think, to capturing the idea that God is *causa sui*. But it would not terminate the demand for explanations: we could still ask *why* God exists, and hope to find an answer which makes it more intelligible why there should be a God. And it does not entail the thesis that God's existence is necessary; nor does it follow from it.

RALPH C.S. WALKER

References and further reading

Barrow, J.D. and Tipler, F.H. (1986) *The Anthropic Cosmological Principle*, Oxford: Clarendon Press.(A careful discussion of whether the fact that the universe is such as to allow intelligent life makes it necessary to seek an explanation for it. Combines sophisticated material from physics and biology with intriguing speculation. Not for beginners.)

Descartes, R. (1641) *Meditations on First Philosophy* with *Objections and Replies*, in *The Philosophical Writings of Descartes*, vol. 2, trans. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1984.(The third Meditation and the first and fourth Replies give Descartes's account of God as *causa sui*. The sixth Reply commits him to holding that logical laws depend on God's choice, though some commentators question whether he really meant this.)

Leibniz, G.W. von (1697) 'On the Ultimate Origination of Things' in *Leibniz: Philosophical Writings*, trans. M. Morris and G.H.R. Parkinson, London: Everyman's Library, Dent, 1973; also in *G.W. Leibniz: Philosophical Essays*, trans. R. Ariew and D. Garber, Indianapolis, IN and Cambridge: Hackett, 1989.(A neat statement of Leibniz's position on the contingency of the world - or one of his positions, since he had several, to the annoyance of Leibniz scholars.)

Leslie, J. (1989) *Universes*, London: Routledge.(Argues that the universe requires an explanation. Favours an account described as theistic, but the theism is Neoplatonic: God is said to be 'the world's creative ethical requiredness', and the universe came about because it was good that it should - on this see especially ch. 8.)

Mackie, J.L. (1981) *The Miracle of Theism*, Oxford, Clarendon Press.(Chapters 5, 8 and 13 are particularly relevant. Criticizes the idea that the universe, as a whole, requires explanation, as well as the Neoplatonic

account and Leslie's version of it in particular. Clear and trenchant.)

Prevost, R. (1990) *Probability and Theistic Explanation*, Oxford: Clarendon Press.(See especially chapter 7. Capable defence of the traditional view that a necessarily existent God is needed to explain a contingent world.)

Spinoza, B. (1677) *Ethics*, trans. with notes E. Curley in *The Collected Works of Spinoza*, vol. 1, Princeton, NJ: Princeton University Press, 1985.(Sets out Spinoza's views about God and necessity. The first part is particularly relevant.)

Continuants

There is a common-sense distinction between terms such as 'statue' or 'chair' on the one hand, and 'concert' or 'war' on the other. A long-standing tradition in metaphysics has attached some significance to this distinction, holding that the first kind of term is used to name continuants, whereas the second kind is used to name events or processes. The difference is that continuants can be said to change, and therefore persist through change, whereas events do not. However, the distinction between continuants and events has been challenged on the grounds that no concrete object does, in fact, retain its identity through time. It has been suggested, for example, that unless we give up the notion of identity through time, we are faced with questions that we cannot answer. In addition, the notion that things persist through change is, apparently, threatened by a certain view of time. On this view there is in reality no past, present and future, but rather unchanging temporal relations between events. It has been suggested that such a view is committed to the idea that objects have temporal parts, and these by definition cannot persist through time.

1 Do objects persist?

Intuitively, you are the same individual as someone who existed yesterday, even though you have fewer brain-cells, your body temperature will have fluctuated and some of your desires may have changed. In contrast, a performance of a piece of music does not itself change: rather, it *consists* of a series of changes. An important part of the motivation for a distinction between continuants and events concerns our notion of responsibility. You can only be held responsible for the actions you performed yesterday if they really were your actions: you now and the person who performed those acts must be one and the same. This connection with responsibility might incline us to widen the class of continuants to include football clubs, religious orders and companies. Someone taking a company to court presumes that the company being taken to court today is the same as the one that violated the law last year.

For material objects, identity through time might plausibly be thought to be no more than spatio-temporal continuity. Put briefly, the continuity criterion states that an object, *a*, existing at one time *t* at place *s*, is identical with *b*, existing at a later time *t*′ at place *s*′, if and only if there is a continuous series of spatio-temporal points between *s* at *t* and *s*′ at *t*′ such that no point in that series is unoccupied by an object. This rules out the possibility of an object going out of existence for a period and then coming back into existence again. It also rules out discontinuous spatial motion. Two kinds of problem case arise for this account, however. Consider first a lump of bronze which is moulded into a statue of an old man by Rodin. This later falls into the hands of Henry Moore, who remoulds the statue to represent a pregnant woman. Finally, it comes into the possession of Elizabeth Frink, who turns it into an enormous head. What this obviously fictional example of aesthetic vandalism illustrates is that, although Frink's statue is spatio-temporally continuous with Rodin's statue, they are nevertheless not the same statue. Consider next an amoeba, A, which divides to form two amoebae, B and C. Both B and C are spatio-temporally continuous with A, but they cannot both be identical with A, for then (since 'identical with' is a transitive relation) they would be identical with each other. Clearly, spatio-temporal continuity is not the same relation as identity (see Identity §1).

In the face of this, we could look for a way of strengthening the spatio-temporal continuity criterion, or we could simply regard identity through time as a primitive relation, irreducible to any other. But both these options leave us with the apparently unanswerable question of whether it is B, or C, or neither, that is identical with amoeba A. A third, radical, approach is to reject altogether the notion of identity through time and with it the notion of a continuant. On this view, associated with Quine (1960), things, strictly speaking, do not persist through time. Rather, they are temporally extended in a sense closely analogous to the sense in which they are spatially extended (see Quine, W.V.O. §5). Just as a thing has different spatial parts at different places, so it has different *temporal* parts at different times. There is no question of a temporal part being identical to a later or earlier temporal part, any more than there is a question concerning whether your left foot is identical to your right foot. In the language of temporal parts, we can redescribe the cases above as follows: the collection of temporal parts that is Rodin's statue shares no temporal parts with either the Moore collection or with the Frink collection, though it does share a temporal part with the collection that is the lump of bronze. As for the amoeba, since no temporal part existing before the division can be identical with any part existing afterwards, there can be no question of B or C being the

same as A. What we have is simply a branching collection of temporal parts.

However, the notion of temporal parts is a controversial one. Effectively, it dissolves the distinction between objects and events in treating the former, more or less, as extended processes. So the temporal-parts view needs to explain why it is objects, not events, that are thought to change (see Events). In addition, the view must explain why we are apparently entitled to hold people responsible for past acts if the temporal part held responsible is not the temporal part that performed the act. Indeed, since temporal parts do not have duration, but exist only for an instant, how can the temporal parts of people be said to perform acts at all? The temporal-parts view can preserve the common-sense distinction between such things as chairs and concerts by pointing out that every temporal part of a chair is itself a chair, whereas it is *not* true that every temporal part of a concert is itself a concert. But is this enough to explain the facts about change and responsibility mentioned above?

2 Continuants and tenseless time

The appeal to temporal parts in dealing with the problem cases above is not simply a desperate manoeuvre: further motivation for the temporal-parts view comes from a certain view of time, namely one which denies the passage of time (see Change §2). Just what is involved in this denial should, however, be made more precise.

According to the *tensed theory of time*, tensed expressions such as ‘now’, ‘last year’, ‘in a few days’ time’, reveal a significant feature of time itself. For example, the presentness of a certain event is, on this view, a non-relational property of the time of that event; and also a transient property, in that the event will soon be past. The tensed theory is therefore associated with the view that time passes. According to the *tenseless theory of time*, in contrast, the presentness of an event is a relational feature of it: it is present only with respect to a particular time. Similarly, an event is only past with respect to a time, or group of times. On this view ‘present with respect to’ simply means ‘simultaneous with’, and ‘past with respect to’ simply means ‘earlier than’. So the only facts about presentness are trivial ones: a time is present with respect to (that is, simultaneous with) itself, past with respect to later times, and future with respect to earlier times. Times, therefore, do not on this view change in respect of their presentness. The tenseless theory is thus associated with the *denial* of time’s passage. We can present the contrast between the tensed and tenseless theories in terms of the account they give of what makes tensed assertions true:

Tensed theory

A particular utterance of ‘It is *now* very windy’ is true if and only if it is now very windy (or, equivalently, extreme windiness is present).
A particular utterance of ‘It *was* very windy’ is true if and only if it was very windy (or, equivalently, extreme windiness is past).

Tenseless theory

A particular utterance of ‘It is now very windy’, uttered at time *t*, is true if and only if it is very windy at *t*;
A particular utterance of ‘It was very windy’, uttered at time *t*, is true if and only if it is very windy at some time earlier than *t*;

where *t* is specified in tenseless terms (for example, ‘4 pm on 31 December 1999’).

It has often been assumed that the tenseless theory is committed to temporal parts, and the assumption is a natural one, for tenseless theory does away with what might have seemed a crucial disanalogy between time and space. As a place is ‘here’ to us just because we happen to be at that place, so, on tenseless theory, an event is ‘present’ to us just because we are at a point simultaneous with that event. Since objects have spatial parts, it is natural to assume that the tenseless theorist will suppose them to have temporal parts. There have, however, been attempts to reconcile the tenseless theory with the notion of continuants.

In temporal-part theory, the fact that *x* is *F* at time *t* is represented by the formulation ‘*x*-at-*t* is *F*’, where ‘*x*-at-*t*’ names a temporal part of *x*. There are, however, different ways of reporting the fact in question. For example, instead of qualifying the subject, *x*, with the term denoting a time, we could qualify the predicate: ‘*x* is *F*-at-*t*’. Alternatively, we could treat properties as *relations*, linking an object and a time: ‘*x* bears relation *F* to *t*’. Why are these ways of representing the facts thought to be alternatives to temporal parts? Because it is the whole, temporally extended, object which is *F*-at-*t* (on the first account) or which stands in certain relation to a time (on

the second account).

An objection to these supposed alternatives says that the intuition underlying the notion of continuants is that an object (in contrast to an event) is *wholly present* at each of the instants which compose its temporal extension. There can be no spatial analogue of this: an object is not thought to be wholly located at each of the points which constitute its spatial extension. The idea that something can be wholly present at different times avoids contradiction if we think of time as passing - that is the time at which an object is present as itself changing. Once we give up the idea that times can change in this way, and adopt the tenseless theory of time, we surely also give up the idea that an object can be wholly located at a time. Treating properties as relations between times and a temporally extended object merely disguises this fact (see Time §2).

See also: Processes

ROBIN LE POIDEVIN

References and further reading

Geach, P. (1965) 'Some problems about time', *Proceedings of the British Academy* 51: 321-36.(Accessible, wide-ranging paper. Argues against the notion of temporal parts.)

Johnston, M. (1987) 'Is there a problem about persistence?', *Aristotelian Society* supplementary vol. 62: 107-35. (Discusses alternative formulations of the proposition that a thing has a property at a certain time, and defends an 'adverbial' treatment of the temporal qualifier. Quite technical.)

Le Poidevin, R. (1991) *Change, Cause and Contradiction*, London: Macmillan.(Develops and defends the tenseless theory of time. Ch. 4 expands on the material of §2 above. Mostly quite difficult.)

Mellor, D.H. (1981) *Real Time*, Cambridge: Cambridge University Press.(Defends the tenseless theory of time and attempts to reconcile the theory with the existence of continuants, by treating properties as relations between objects and times. A good but relatively demanding introduction to the tenseless theory of time.)

Noonan, H.W. (1980) *Objects and Identity*, The Hague: Martinus Nijhoff.(A detailed and often quite difficult discussion of synchronic and diachronic identity and the debate over continuants, coming down in favour of Quine's ontology of temporal parts.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: M.I.T. Press.(Referred to in §1 above. See especially page 171 for his view that objects are not to be distinguished from events.)

Simmons, P.M. (1987) *Parts: A Study in Ontology*, Oxford: Clarendon Press.(A detailed and technically demanding application of mereology - the formal study of part-whole relationships - to metaphysical issues such as the nature of continuants. Chapter 5 is especially relevant.)

Continuum hypothesis

The 'continuum hypothesis' (CH) asserts that there is no set intermediate in cardinality ('size') between the set of real numbers (the 'continuum') and the set of natural numbers. Since the continuum can be shown to have the same cardinality as the power set (that is, the set of subsets) of the natural numbers, CH is a special case of the 'generalized continuum hypothesis' (GCH), which says that for any infinite set, there is no set intermediate in cardinality between it and its power set.

Cantor first proposed CH believing it to be true, but, despite persistent efforts, failed to prove it. König proved that the cardinality of the continuum cannot be the sum of denumerably many smaller cardinals, and it has been shown that this is the only restriction the accepted axioms of set theory place on its cardinality. Gödel showed that CH was consistent with these axioms and Cohen that its negation was. Together these results prove the independence of CH from the accepted axioms.

Cantor proposed CH in the context of seeking to answer the question 'What is the identifying nature of continuity?'. These independence results show that, whatever else has been gained from the introduction of transfinite set theory - including greater insight into the import of CH - it has not provided a basis for finally answering this question. This remains the case even when the axioms are supplemented in various plausible ways.

1 The continuous, the discrete and the infinite

Greek mathematics was divided into arithmetic - the science of discrete magnitudes - and geometry - the science of continuous magnitudes. The basis for this division can be found in Aristotle's sophisticated discussions of space and time. The existence of incommensurable magnitudes, such as the side of a square and its diagonal (the ratio of which is $\sqrt{2}$), was already well known, as was the distinction between points and extended parts, however small, of a continuous whole. Aristotle argued that since two points can never be contiguous (either there is a separation between them or they coincide), a continuum can never be built up out of points. He defined a continuously extended whole as one which may be divided at any point into continuously extended wholes and concluded that it is necessary to draw a fundamental distinction between those wholes which are collections of parts and those which are 'given before their parts', that is, between discrete and continuous magnitudes.

So long as this division was in place (that is, until the seventeenth century) a question such as 'What is the (cardinal) number of points in a continuous line?' could not arise because it could not sensibly be asked. 'Continuity' was a primitive geometric concept. The situation changed, however, with the acceptance of Descartes' introduction of algebraic methods into the heart of geometry, allowing geometric curves to be defined by equations of 'motions' of points 'generating' them. This effectively swept aside the division between arithmetic and geometry, but without in any way confronting or resolving the problem of how to model the continuous by the discrete.

Newton and Leibniz explored and exploited the infinite divisibility of continua in their attempts to study continuous motions algebraically. Initially this exploration (infinitesimal calculus, analysis; see Analysis, philosophical issues in §1) moved freely between algebraic expression and geometric intuition, mediated by the concept of continuous motion. However, this led to the discovery of a number of 'pathological' functions, whose graphs cannot be pictured. For example, in a continuous plane, unlimited in extent, there is no limit to the number of wavelike, and hence continuous, oscillations that a point might execute within a given finite interval. So one might ask whether there might be a function which corresponds to a point oscillating with infinite frequency (that is, which is such that no matter how small an interval we take there will be an oscillation contained within it). Weierstrass' everywhere continuous but nowhere differentiable function seemed to 'describe' such a 'motion'. This kind of function contradicts the assumption, based on cases which can be pictured, that continuity and differentiability go together. So an account of continuity was required in terms not based on geometric intuition.

2 Numbering the continuum

Cantor's early attempts to characterize the structure of continua were natural extensions of the kind of work in analysis which led to discovery of these pathological functions. He established that the cardinality of the set $\mathbb{R}$ of real numbers is greater than that of the set $\mathbb{N}$ of natural numbers (see Cantor's theorem). He further proved that

there is a one-one correspondence between the points in a plane (the two-dimensional continuum) and the points in a line. So he already knew (writing $|A|$ for the cardinality of *A*) that

$$|\mathbb{R}\times\mathbb{R}| = |\mathbb{R}| > |\mathbb{N}|.$$

Cantor thus interpreted the problem of how to identify the nature of continuity as that of how to say when a given set of points in an *n*-dimensional space is to be considered a continuum. Having already defined the real numbers in terms of limits of convergent sequences of rational numbers, he made use of the more general notion of a 'limit point'. A point *p* is defined to be a limit point of a set *P* of points if in any neighbourhood of *p* there are infinitely many points of *P*. It can be proved that if *P* is infinite then it has at least one limit point. The set of limit points of *P* is denoted by $P^{(1)}$. The set $\mathbb{R}$ of real numbers (the 'continuum') is such that $\mathbb{R}^{(1)} = \mathbb{R}$. Sets with this property were called 'perfect'. But perfection does not serve to characterize continuity, as Cantor demonstrated with his 'ternary set', also called 'Cantor's discontinuum' (see Set Theory §5). This is the set of points of the closed interval (a 'closed' interval includes the points at either end) $[0, 1]$ which remain after infinitely many repetitions of the following procedure: divide each interval of the set into thirds, and in each case discard ('blank out') the middle third, leaving behind two closed intervals.

This set is dense in no interval of the continuum ($B \subseteq A$ is 'dense' in *A* if it is disjoint from no open interval of *A*), closed, perfect and contains no interior points: it is completely discontinuous.

Nevertheless, Cantor felt that perfect sets held the key since he was able to prove that no *denumerable* point set (that is, a set with the cardinality of the natural numbers) could be perfect. If a set *P* is finite, its first derived set $P^{(1)}$ is empty. For any *P*, either there is some finite *n* such that $P^{(n)}$ is finite and hence $P^{(n+1)}$ is empty, or there is not. For perfect sets there is no such finite *n*. So if this idea is to differentiate between continuous and discontinuous perfect sets it seems that *transfinite* indices of iteration need to be used. This was the context in which Cantor originally introduced transfinite ordinal numbers (see Set theory §2). He divided the ordinals into first and second number classes:

(1) $0, 1, 2, 3, \ldots$
(2) $\omega, \omega+1, \omega+2, \ldots \omega.2,\ \omega.2+1, \ldots \omega.3, \ldots \omega.\omega \ldots.$

(1) is the set of natural numbers, $\mathbb{N}$, which has cardinality $\aleph_0$ ('aleph-nought' or 'aleph-null'). Every ordinal α in (2) is such that the cardinality of $\{x\colon\ x < \alpha\}$ is $\aleph_0$. That is, these numbers represent all the possible orderings of a denumerably infinite set. Cantor proved ([1895, 1897] 1955: 169-73) that these two number classes cannot be put into one-one correspondence and that there can be no set with a cardinality in between the two. This justified calling the cardinality of (2) '$\aleph_1$'.

In this way the extension of the ordinal number sequence into the transfinite generates not only infinite ordinal numbers α but also infinite cardinal numbers $\aleph_\alpha$. It thus provides a scale of cardinal numbers on which one might hope to locate $2^{\aleph_0}$. The 'continuum hypothesis' (CH) expresses the claim that there is a one-one correspondence between the real numbers and the numbers in the second number class (in symbols, $2^{\aleph_0} = \aleph_1$), and another way of interpreting the claim made by the generalized continuum hypothesis (GCH) would be as saying that the two routes to generating higher infinite cardinalities (via power sets and via the number classes) coincide (in symbols, $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ for every ordinal α). Although Cantor believed CH to be true (and Gödel (1939) proved it to be consistent with the axioms of set theory), he realized that there is no immediate guarantee that every cardinality will have a representative among the cardinalities of the number classes unless it can also be assumed that every set can be well-ordered, and hence has an ordinal number. If it were not the case that every set could be well-ordered, the cardinalities of those sets which cannot be so ordered would not be represented by cardinal numbers arising from the construction of the ordinal number sequence. Indeed it can be shown that in

Zermelo-Fraenkel set theory (ZF), GCH entails the axiom of choice and hence that every set can be well-ordered, whereas Cohen's independence result proves that the converse does not hold (see Cohen 1963-4).

3 Reactions to independence

Proof of the independence of CH (GCH) from ZFC (ZF plus the axiom of choice - see Set theory, different systems of §3) stimulated renewed interest in philosophical questions about the status of set theory and the epistemology of mathematics. Is there a universe of sets in virtue of which CH is either true or false, although at present we do not know which? If this is the case, how would we come to know more about it and hence hope to determine the truth-value of CH? The most radical reaction would undercut even these questions by interpreting the independence results as showing that the whole set-theoretic approach to problems of continuity was misguided: CH is undecidable because it does not make any real mathematical sense; it presupposes that we can sensibly talk about actually infinite sets, whereas we cannot. Less radical reactions assume that CH does make mathematical sense. The problem then is not merely how to determine whether or not it should be accepted, but to decide whether there is a basis for acceptance or rejection and, if so, what this might be. It might be argued that the situation of CH is analogous to that of the parallel postulate after it was proved to be independent of the remaining axioms of Euclidean geometry. The result in that case was a widening of the field of mathematical investigation to consider a whole variety of geometries in which the parallel axiom fails. Analogously, in set theory, explorations of systems in which CH holds and in which it fails have been carried out. Gödel (1947), however, contrasts the epistemological situations in geometry and set theory. In geometry the question of the truth or falsity of the parallel postulate retains its sense if the geometric axioms are interpreted as referring to the behaviour of rigid bodies, light rays and so on, but not otherwise - that is, the question of its truth or falsity has become an empirical question to be answered by reference to the contexts in which geometry is used; it is no longer a mathematical question. The question of the truth or falsity of CH, however, retains its sense only if the axioms of set theory are interpreted as referring to mathematical objects (sets), since these axioms have no direct empirical interpretation. Gödel himself believed that sets are mathematical objects which have an independent existence. He would thus interpret the proof of the independence of CH from ZFC as a demonstration that the axioms of ZFC do not contain a complete description of the set-theoretic universe. In this case there is a need to look for new axioms to complete the description. Several such axioms have been suggested and their consequences studied (see Set theory §1), but none has yielded anything approaching consensus on either their acceptability or that of CH.

Gödel's mathematical realism is not the only philosophical position which would underwrite attempts to resolve questions about the status of CH (and other proposed axioms of set theory) by studying the consequences for other areas of mathematics of accepting or denying it. It might be argued that what mathematical sense CH has is derived from its preformal connections with the problem of characterizing the continuous in terms of the discrete. If so, any determination should be based on its impact on the ever-widening area of mathematics connected with this problem. This approach relates philosophical issues about the status of set theory with other areas of mathematics, it does not resolve them.

See also: Logical and mathematical terms, glossary of

MARY TILES

References and further reading

Cantor, G. (1895, 1897) 'Beiträge zur Begründung der transfiniten Mengenlehre', parts 1 and 2, *Mathematische Annalen* 46: 481-512, 49: 207-46; trans. P.E.B. Jourdain (1915), *Contributions to the Founding of the Theory of Transfinite Numbers*, New York: Dover, 1955.(Two of Cantor's most important memoirs on transfinite numbers.)

Cohen, P.J. (1963-4) 'The Independence of the Continuum Hypothesis', *Proceedings of the National Academy of Science, USA* 50: 1,143-8, 51: 105-10; repr. in *Set Theory and the Continuum Hypothesis*, New York: Benjamin, 1966.(The original version of his proof.)

Dauben, J.W. (1979) *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Cambridge, MA: Harvard University Press.(An excellent intellectual biography which traces the development of Cantor's mathematical ideas, while putting them in the context of his philosophy of the infinite and his professional life.)

Gödel, K. (1939) 'Consistency Proof for the Generalized Continuum Hypothesis', *Proceedings of the National Academy of Science, USA* 25: 220-4.(Original publication of the proof.)

Gödel, K. (1940) *The Consistency of the Continuum Hypothesis*, Princeton, NJ: Princeton University Press, 1994. (Revised and more complete version of the original proof.)

Gödel, K. (1947) 'What is Cantor's Continuum Problem?', *American Mathematical Monthly* 54: 515-25; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.(A very valuable discussion which gives some of the context of debate about CH.)

Hallett, M. (1984) *Cantorian Set Theory and Limitation of Size*, Oxford: Clarendon Press.(Includes an integrated account of Cantor's metaphysical and mathematical theories of infinity together with discussion of the relation between Cantorian and modern set theory.)

Tiles, M.E. (1989) *The Philosophy of Set Theory: An Introduction to Cantor's Paradise*, Oxford: Blackwell. (Traces the philosophical and mathematical background to the formulation of CH and considers whether it can be decided.)

Contractarianism

The idea that political relations originate in contract or agreement has been applied in several ways. In Plato's Republic Glaucon suggests that justice is but a pact among rational egoists. Thomas Hobbes developed this idea to analyse the nature of political power.Given the predominantly self-centred nature of humankind, government is necessary for society. Government's role is to stabilize social cooperation. By exercising enforcement powers, government provides each with the assurance that everyone else will abide by cooperative rules, thereby making it rational for all to cooperate. To fulfil this stabilizing role, Hobbes argued that it is rational for each individual to agree to authorize one person to exercise absolute political power. Neo-Hobbesians eschew absolutism and apply the theory of rational choice to argue that rules of justice, perhaps even all morality, can be construed in terms of a rational bargain among self-interested individuals.

John Locke, working from different premises than Hobbes, appealed to a social compact to argue for a constitutional government with limited powers. All men are born with a natural right to equal freedom, and a natural duty to God to preserve themselves and the rest of mankind. No government is just unless it could be commonly agreed to form a position of equal freedom, where agreement is subject to the moral constraints of natural law. Absolutism is unjust according to this criterion.

Rousseau developed egalitarian features of Locke's view to contend for a democratic constitution.The Social Contract embodies the General Will of society, not the unconstrained private wills of its members. The General Will wills the common good, the good of society and all of its members. Only by bringing our individual wills into accord with the General Will can we achieve civic and moral freedom.

In this century, John Rawls has recast natural rights theories of the social contract to argue for a liberal egalitarian conception of justice. From a position of equality, where each person abstracts from knowledge of their historical situations, it is rational for all to agree on principles of justice that guarantee equal basic freedoms and resources adequate for each person's independence.

T.M. Scanlon, meanwhile, has outlined a right-based contractualist account of morality. An act is right if it accords with principles that could not be reasonably rejected by persons who are motivated by a desire to justify their actions according to principles that no one else can reasonably reject.

1 The role of agreement

Contract accounts of justice are intuitively appealing. One explanation of their appeal is promissory obligation; if we agree to something we accept it, and when we agree on condition that others do, there is mutual reliance and grounds for holding all to their commitments. But the obligation of fidelity to promises does not adequately account for the intuitive force of agreement. Not all agreements involve promises. Also, it still must be explained why we should keep our promises, and a contractual promise cannot explain this duty. Hume (1748) explained promissory and contractual obligations by their public utility, and dismissed contractarian accounts of political relations as superficial.

But suppose we see the intuitive force of agreement as deriving not from promising, but from the liberal idea that rules of social cooperation should be based in individuals' consent and for their reciprocal benefit. Hume's argument cannot easily account for these ideas, for there is no clear connection between general utility and free consent; moreover, utilitarian principles do not guarantee institutions that benefit each individually. Here the idea of agreement calls upon deeper associations with democratic ideals of freedom and equality. Assuming that individuals are free and equal by nature or by right, it is natural to think that the political and social requirements they are subject to are contingent on their free agreement. Approached in this way, the ideas of contract and agreement can be applied to elucidate the requirements of freedom and equality.

Contract views differ depending on how the idea of agreement is specified. Who are the parties to the agreement? How are they situated (status quo, state of nature or strict equality)? What are their intentions, capacities and interests? What rights and powers do they have? What is the purpose of the agreement (political obligations, a constitution or social duties and institutions too)? Is the agreement actual or hypothetical, historical or non-historical? There is not one, but several contract views.

Two distinctions help to organize different views. First, *political contractarianism* concerns political and social relations associated with the idea of justice, including the legitimate role and powers of governments, basic rights and duties of citizens, property and its distribution, and the structure of economic relations. Issues of political justice underlie the work of the major representatives of social contract doctrine: Hobbes, Locke, Rousseau, Kant and Rawls. *Moral contractarianism* extends the idea of agreement to other moral relations. The most ambitious view would hold that all our duties and obligations, political and non-political, have their bases in agreement. None of the figures above profess this view.

Next, a *social contract view* maintains that certain basic duties and institutions are grounded in a unanimous agreement among all society's (competent) members. Unanimity illuminates the idea that justice, or morality generally, involves uniform public rules impartially applied to all individuals, which benefit each equally in some way. An *individual contract view* holds that certain basic duties or institutions originate in private or non-social agreement. Robert Nozick's libertarianism is an example (Nozick 1974); he depicts legitimate political power as originating not in social agreement, but in a series of discrete private contracts among isolated individuals (see Libertarianism).

Attention here is devoted to political and moral versions of the social contract tradition. There are two major traditions of social contract thought that deal with issues of political justice: *interest-based contract views* stem from Hobbes; *democratic contract views* derive from the natural rights theories of Locke, Rousseau and Kant.

2 Interest-based contracts and Hobbes

The common aim of interest-based views is to account for justice (or morality generally) in terms of what best promotes each person's enlightened interests. Norms of justice are depicted as the outcome of a mutually advantageous agreement. To carry this idea through, interest-based views contend that the basic desires and interests of individuals are fixed by their nature or circumstances, and are definable without any moral notions. Normally interests are explained as self-regarding or focused on oneself, although they need not be purely so (for example, Hobbes recognized 'conjugal affection' as widespread). What is important is that our final ends conflict. Agreement is then depicted as a rational compromise or bargain among essentially conflicting interests, where each party is willing to qualify the direct pursuit of their interests on condition that others do too. Essential to interest-based views is that all parties to the agreement must be made better off (or at least not worse off) than they would be without it.

Interest-based agreements are *historical* in that the parties have knowledge of their particular desires and circumstances. In Hobbesian interest-based views, historical circumstances are defined by a hypothetical, presocial and non-cooperative state of nature. The purpose of agreement from this baseline is to arrive at conditions of peaceable and efficient social cooperation, on the assumption that humans are mainly concerned with advancing their particular interests, are indifferent to the welfare of strangers and have no concern for morality and justice for their own sake. In interest views based on *convention*, the parties are similarly motivated, but their historical circumstances are defined by the current status quo. Because these views take the status quo as given, they do not so much seek to justify or critically assess prevailing cooperative norms as to explain them. Primary attention in this entry is devoted to Hobbesian views.

Hobbes' social contract has two roles (see Hobbes, T. §§6, 7). Its first purpose is as an analytical device designed to clarify the nature and necessities of political society. Its second purpose is normative: to justify absolute political power (see Absolutism §1). To discover the nature of political power, Hobbes imaginatively dissolves political society into its constituent parts, individuals, and then asks how it could come about from this initial situation. In the absence of effective political power individuals (or families) would inhabit a non-cooperative state of nature. Human nature is such that the primary ends of individuals are self-focused (or family-focused at best) and conflicting. Hobbes contends that the three motivations dearest to us are self-preservation, 'conjugal affection' and the means for commodious living. Given these and other motives, and uncertainty about the future, our desires are insatiable. But resources are scarce, so humankind is naturally competitive. Moreover, we are equal enough in strength of body and mind for the weakest to pose a threat to the strongest. Assuming these conditions, Hobbes' thesis is that a state of nature tends towards a state of war. In the absence of effective political power, it is irrational to be cooperative by keeping one's promises and commitments or trusting in the goodwill of others. To

protect one's person and possessions, each person's most rational strategy is one of 'anticipation', being prepared to attack others whenever circumstances seem favourable. The outcome is a collectively irrational situation where each person does what is best for themselves but all end up worse off than if they (jointly) had acted differently.

Hobbes' state of nature is the first recognition of 'prisoner's dilemma' situations central to modern game theory (see Decision and game theory §3). Hobbes resolves the state of nature dilemma with the social contract. What is needed to deliver those captured in a state of nature, or to avoid falling into one, is: (1) a system of norms which make social cooperation possible ('the laws of nature'); and (2) an effective sovereign to publicize and uniformly enforce positive laws that specify these 'articles of peace'. Hobbes' contract commissions one person to exercise political power sufficient to specify and enforce such laws. The contract is everyone's *mutual authorization* of a sovereign, not an agreement with him (a contract of government), for the sovereign cannot be contractually bound and remain sovereign (see Sovereignty). For Hobbes, to be adequate to its role, political power must be effectively absolute. The sovereign's role is to *stabilize* social cooperation by providing all with the assurance that cooperative norms are being enforced and that others will abide by them.

Locke argued that agreement to absolute political power is irrational. Constitutional regimes that limit power and secure individual rights provide better alternatives. Hobbes contests the stability of constitutionalism: effective sovereign power cannot be limited, divided or subject to the rule of law. History has decided matters differently. What endures in Hobbes is not his argument for absolutism, but the structure of his social contract (see Constitutionalism).

3 Contemporary Hobbesian views

This enduring legacy of Hobbes' social contract becomes apparent in the work of two contemporary Hobbesians, David Gauthier and James Buchanan. Both defend the classical liberalism of the *laissez-faire* economists. Gauthier does not try to justify executive political authority - it is unnecessary among fully rational beings. Instead he aims to (1) give a 'rational reconstruction' or morality, and (2) show how Western market society is grounded in reason (Gauthier 1986). Regarding (1), Gauthier contends that rationality is individual utility maximization. Morality, while it constrains the direct pursuit of utility, can be construed as an extension of principles of rational choice: rules which are rational to follow on condition that others do. Morality is then reduced to enlightened self-interest. It constitutes the restrictions on action that rational persons would mutually agree on to maximize individual utility.

Gauthier exhibits this neo-Hobbesian theme in (2), his defence of the 'system of natural liberty'. For Gauthier (and Buchanan) 'Economic Man' best characterizes human behaviour outside families: 'Economic Man is a radical contractarian in that *all* his free or non-coercive interpersonal relationships are contractual' (Gauthier 1986: 319). Contractual bargains typify market behaviour. Markets are 'an ideal of human interaction' because there we act free of moral constraints and perfect markets are efficient. But due to externalities markets often fail; each acts to maximize individual utility, and the outcome is non-efficient (see Market, Ethics of §1). It seeks to approximate the efficient equilibrium of perfect competition. But this requires cooperation, a joint strategy of mutual adherence to constraints on utility maximization. Gauthier's social contract supplies the cooperative norms needed to approximate perfect market interaction.

Hobbes contends that markets and property are not possible without political power; a sovereign must define property claims and enforce contracts. But Gauthier argues that rational agents in a state of nature would realize that the benefits of market relations outweigh the costs of predation and strife, so natural property and exchange would evolve among physical equals prior to any social contract. Market behaviour and private property are then presocial, part of our natural condition. Gauthier's social contract is not an agreement on political power, on the design of property or on markets as the mechanism for distribution. It is a bargain yielding a principle ('minimax relative concession') for correcting the externalities behind market failures and for distributing the fruits of cooperation (the 'cooperative surplus') arising outside private propertied market relations. This principle mimics the perfect market in that it distributes the benefits and burdens of cooperation according to each person's marginal contribution. Gauthier's social contract has a limited function, but a necessary one he believes, if peaceable and efficient social cooperation are to be possible.

Hobbesian views depict justice and morality as mutually advantageous norms that are rational to observe so long

as others do too. Providing everyone with assurance that cooperative terms will be uniformly complied with is essential for stability. However, even assuming that each benefits from the compliance of others, moral constraints often seem to conflict with our rational interests, and for many this is sufficient reason to disregard them. How can such 'free-riders' be persuaded that it is irrational to disregard moral constraints when circumstances appear favourable? To argue that virtue and justice are their own reward makes no sense on Hobbesians' instrumental view of morals. This problem explains why Hobbes gave the sovereign absolute power: no other way insures against free-riding, which undermines everyone's belief in the rationality of compliance. Gauthier's solution is non-political: free-riders acquire a reputation for unreliability and others refuse to cooperate with them; rationality therefore requires developing a disposition always to comply. This may convince people who continually interact, but what about strangers? The larger a society, the less effective reputation is as a deterrent, so under modern conditions it appears inadequate for stability. (For this reason Buchanan (1975) advocates a 'protective state' as one end of the social contract, to stabilize pre-political market relations.)

Hobbesian views collapse questions of justification into questions of (rational) motivation: to justify norms is to show they promote each person's existing interests. But many who currently benefit from the status quo would be greatly disadvantaged if Gauthier's or Buchanan's *laissez-faire* schemes were enacted (for example, monopolists, oligopolists, beneficiaries of government regulation and subsidies, and recipients of public assistance). What reason do they have for accepting contract arguments for reform? Can it seriously be argued that in mixed Western economies everyone would benefit by reverting to *laissez-faire*? History tells us this shift promises only further adverse consequences for those already naturally and socially disadvantaged, and even greater benefits for most who are already better off. Without resolving this motivational problem, the justificatory task of modern Hobbesian theories cannot be carried through on its own terms.

4 Conventionalism

Hume (§5) attacks Locke's contractarian account of political legitimacy and obligation. But he argues that rules of property and economic exchange, of promising, and of allegiance to governments might be explained as 'a kind of contract' (Hume 1748). A *convention* is a mutually beneficial social rule which people publicly accept and comply with because they believe others will too, and where the rule's benefits can be gained only if there is general compliance. The idea is often used for analytic or explanatory purposes, to clarify institutions or practices commonly accepted as part of the status quo. Conventional explanations of law and morality have been attempted (Harman 1977). The term 'social contract' is often applied, particularly by social scientists, to practices that exhibit the features of Humean conventions. Conventions are, like social contracts, mutually beneficial, but only in the weak sense that each person who participates is better off than they would be in the absence of any such rules. For example, virtually any property scheme can be explained as a convention, or 'social contract' in this weak sense. To call conventional accounts 'contractarian' is acceptable if it is recognized that, standing alone, they are explanatory and not normative. Each person's benefiting from a practice does not justify it, although it may explain why people continue to comply.

5 Democratic contract views and Locke

Freedom and equality are primary democratic ideals. The role of a social contract in Locke (§10), Rousseau, Kant and Rawls is to elucidate the requirements of these ideals in matters of political justice. If free persons all could, or would, agree to something from a suitable position of equality, the standards they would endorse embody requirements of democratic justice applicable to us so far as we aim to cooperate on terms of equal freedom and mutual respect.

Unanimous agreement among free persons equally situated serves different purposes for these thinkers. For Locke the social contract is a test of the legitimacy of existing constitutions; for Rousseau and Kant, it is also a standard for just legislation; and for Rawls it provides criteria for designing the basic institutions of society. Each defines the conditions of equality and the freedom of the parties to the agreement differently. But the general idea shared in common is that the justice of political institutions depends on whether they could, or would, be agreed to from a position of equal political jurisdiction (see Justice §5).

Democratic contract views maintain that, whatever the differences among individuals *de facto*, all persons are free and equal *de jure*, when each has the requisite capacities of reason, intellect and self-control to decide on their

good and comply with social requirements. Democratic contract views are therefore right-based, since they define agreement as subject to antecedent principles of right and justice, which are depicted as not reducible to non-moral interests (or any prior agreement based therein). Among these principles of right may be certain moral rights and duties which cannot be infringed upon or bargained away (as in Locke, Rousseau and Kant). Alternatively, principles of right may be implicit in an ideal of the person as free, equal and endowed with certain moral powers (as in Rawls), that is purportedly implicit in the public culture of a democratic society.

The democratic contract tradition originates with Locke's argument against royal absolutism. Locke contends that all are born with a natural right of equal freedom, by virtue of which no one may legitimately exercise political jurisdiction over another without their consent. Moreover, all have a natural duty to God to preserve themselves and the rest of humankind. Given these rights and duties (the 'fundamental law of nature'), Locke asks, how could legitimate political authority come about from a state of equal right as defined by a non-political but socialized state of nature? Locke contends that no constitution is just unless it could have been contracted into from this initial situation by a series of agreements wherein no one agrees to anything that makes them worse off than in a state of nature, and no one agrees to anything that causes them to violate the rights of others or their own duties. Royal absolutism could not be contracted into compatibly with these conditions, so it is unjust. A constitutional monarchy with a limited franchise could be contracted into, as could other constitutional schemes; each is then, according to Locke's criterion, a legitimate constitution (see Liberalism §5; Legitimacy §2).

Locke's social contract, like those of Hobbes, Rousseau, Kant and Rawls, is hypothetical. Locke's claim is not (as Hume thought) that a constitution is legitimate only if it actually originated in agreement in the past among our forebears: one '*cannot* by any *Compact* whatsoever, bind *his Children* or Posterity' (Locke 1690: §116; original emphasis). Rather, it is legitimate only if it could have originated in agreement from a state of equal political right. Actual agreement is irrelevant for purposes of applying this criterion.

Locke's social contract is also historical, for it envisages agreement from a (hypothetical) state of nature where everyone has full knowledge of their situation and interests there. Unlike Hobbes' state of nature, this situation is moralized - a 'natural community' - for each therein recognizes natural rights and duties as constraints upon everyone's pursuit of their interests, and upon what can legitimately be agreed to.

It is objected that hypothetical contracts have no binding force (see, for example, Dworkin 1977). This assumes that hypothetical agreements provide an account of the political obligations of individuals. But this is not the primary purpose of Locke's social contract (or Rawls', Kant's or Rousseau's either). The main purpose of Locke's agreement is not to explain the duties of individuals, but the justice of political institutions. Promising plays no role in explaining the effectiveness of Locke's criterion when construed as a test of political legitimacy. Locke does give an individual contract account of political obligation. Persons are obligated as citizens and subjects to bear allegiance to a legitimate government in perpetuity, not by birth, but only once they actually give their express consent to join a regime. Those who withhold such express consent to join still consent tacitly to obey the laws upon benefiting from them. Express and tacit consent are actual, not hypothetical; moreover they are individual commitments, not social agreements requiring participation by all society's members. Locke's account of political obligation is subject to many objections (see Consent; Obligation, political §3), but it is not essential to his social contract view.

6 Rousseau's democratic contract view

In Rousseau (§3) and Kant the social contract provides a point of view from which to decide on just legislation and articulates the ideas of moral freedom (autonomy) and the common good (Rousseau 1762; Kant 1797).

All social contract views assume that everyone benefits by social cooperation. Rousseau makes a stronger claim: certain essential human capacities and interests can only be realized in society, and only then where society has been structured democratically. While natural freedom typifies a state of nature, there we are not (in contrast to Hobbes) rationally prudent, but 'stupid, short-sighted animals', moved by instinct and appetite tempered by natural compassion. Our capacities for reason (prudential and moral) are socially interdependent. But human reason, moral sensibilities and deeper human affections have always been distorted by gross inequalities in power and property. Consequently, humankind has been prevented from realizing these great goods and moral and civic freedom. How is it possible to achieve a society where people realize these benefits, without arousing the will to domination,

subservience and other vices of history? Rousseau addresses this problem with his *contrat social*.

The general will expresses the idea that moral and civic freedom are attainable only when institutions satisfy principles free and equal citizens can give to themselves. One's particular will is the interests constituting one's individual good. Hobbes' social contract underscores the inevitable conflict among particular wills, and arbitrates by requiring all to compromise for mutual benefit. Rousseau calls this 'sum of particular wills', the 'will of all'. The general will, by contrast, is what we would jointly will if we subordinated our particular wills and adopted the perspective of free and equal citizens, each motivated by the good of the public and all its members (see General will).

The general will is initially expressed through the social contract. Its object is the common good. For Rousseau (Kant and Rawls too) divergences among individual wills does not mean there cannot be a supreme end all might share which regulates individuals' free pursuit of their particular ends. Such a common good is possible, and is necessary if all are to realize moral and civic freedom. The common good is, first, the institutions and laws that protect the fundamental ('natural') rights of individuals and define a just constitution. Among these are democratic institutions which provide for the civic freedom of citizens and their equal (and for Rousseau direct) participation in legislative deliberation and decision. Beyond this, the common good is whatever laws are willed in legitimate democratic procedures that advance the interests of each member of society in their freedom and security.

For Rousseau (and Kant), the social contract and the general will also express the elusive idea of moral autonomy (see Autonomy, ethical). Natural liberty, or acting as one pleases on the basis of inclination, is a kind of slavery, 'while obedience to a law we prescribe to ourselves is liberty' (Rousseau 1762: I, 8). We have a fundamental interest in realizing such freedom, since then (according to Kant) we fully realize the powers of practical reason constituting our humanity. Moral freedom, or autonomy, is attainable only in society, and (for Rousseau) only then where regulative institutions conform to principles free and equal citizens could give to themselves reciprocally. To act autonomously is to act compatibly with the general will for its own sake: to willingly subordinate the free pursuit of one's particular good to the common good, as articulated by the social contract.

7 Rawls' democratic contract

Rawls (§1) transforms and develops the contract views of Kant, Rousseau and Locke. The object of agreement is principles of justice that apply not just to laws and the constitution, but to other basic social institutions that regulate the distribution of wealth and opportunities to occupy favourable social positions. Rawls aims to provide 'the most appropriate moral basis for a democratic society' (1971: viii).

Consider a problem in Locke's social contract: since the parties have historical knowledge of their differences in a state of nature, it is not irrational for some to agree to alienate their rights of equal political jurisdiction to gain the benefits of political society. Nor is it wrong; for Locke political rights, unlike liberty of conscience and basic freedoms of the person, are not inalienable. If so, it is not unjust to restrict voting rights constitutionally to propertied males, making the rest 'passive citizens'. However democratic Locke's starting position, he does not require a democratic constitution with equal political rights.

Rawls' contract is designed to prevent arbitrary contingencies (for example, prevailing property distributions) from affecting the justice of political and social institutions. A contract is non-historical the more those party to it abstract from knowledge of circumstances (actual or hypothetical) that would advantage or disadvantage people who occupy certain positions. (The idea is already implicit in Rousseau and Kant, for neither envisage social agreement as transpiring in a state of nature.) Rawls' original position is wholly non-historical since contractors are placed behind a complete 'veil of ignorance' regarding their particular situations, so none can take advantage of their social circumstances, their natural talents, their particular conceptions of the good or prevailing historical conditions. By situating his parties symmetrically, Rawls makes them strictly equals, and carries to the limit the ideal of equality behind democratic contractualism.

Rawls argues that the conception of justice that would be agreed to in the original position is neither utilitarian nor perfectionist, but a Kantian account: justice as fairness. Its main principles afford equal rights to certain basic liberties and opportunities and regulate distributions of wealth and access to social positions so that they benefit everyone, with primary attention given to those least advantaged. These principles for institutions, along with individual duties of justice, mutual aid and mutual respect, comprise the general will of free and equal moral

persons. Just laws and institutions constitute their common good. In complying with and acting for the sake of these principles and institutions, citizens are morally autonomous, for they act on and from a law they give to themselves, out of their rational and moral powers.

Hegel (1821) argued that contract doctrines are individualistic and incompatible with community. This criticism may apply to Hobbesian accounts of our relations with strangers but, as for Rousseau and Kant, arguably Hegel misunderstood the role of the general will. Contemporary communitarians (for example, Sandel 1982) restate Hegel's criticism, contending that Rawls' original position presupposes abstract individualism (see Community and communitarianism). Rawls (1993) contends this interpretation is mistaken.

8 Right-based versus interest-based views

Hobbesians and others object that right-based views (especially non-historical ones) are not genuinely contractual, since they involve little or no bargaining. Given their moral assumptions and shared interests in a common good, agreement is superfluous; the same standards can be derived by impartial decision by one individual. Now, one can simply stipulate that contracts and agreements involve bargains among essentially conflicting interests. This has the peculiar consequence that most major proponents of social contract theory (Locke, Kant, Rousseau, Rawls) are not contractarians at all. (Even Hobbes' contract is not a negotiated bargain but a mutual authorization of an effective sovereign.) But not all agreements involve bargains or negotiated compromises among conflicting interests. People with diverse ends often commit themselves to norms on condition that others do, to achieve a shared end of great importance (for example, in religious associations or marital vows). The point of agreement is not to resolve conflict (there may be none), but to tie down the future to keep the parties from deviating from the shared purposes and norms of association. Democratic views involve agreement in this sense of a *shared precommitment* among citizens to relations that maintain their freedom and equal status. This agreement is not effective because of mutual promises. Rather, it establishes principles and institutions which bind citizens perpetually into political relations as equals, and so prevent them from later changing their minds or surrendering to temptations that undermine the good of others or everyone's shared interest in justice.

Being rational bargains, interest-based contracts involve mutual advantage: each person gains, given their historical starting position (status quo or state of nature). Being shared precommitments, right-based agreements involve reciprocity: everyone benefits as measured from a fair baseline of equality. Interest-based agreements between serfs and landlords to increase productivity on condition that serfs get 10 per cent of the surplus are mutually advantageous but not reciprocal, for the terms are unfair and the starting position is unjust. In the reciprocal agreements of democratic views, there is no guarantee that each person will benefit compared to the status quo. The purpose of reciprocal agreements is not to improve on the status quo by preserving its injustices, but to provide a criterion for assessing current injustice and the measures required for reform.

But if there is no guarantee that one will benefit, how can everyone be expected to accept the terms of reciprocal agreements? Many favoured by an unjust status quo will not; if their sense of justice is weak or vacillating, it may be irrational for them to. For democratic theorists this does not count against the validity or reasonableness of right-based contracts. It is a motivation problem belonging to the process of transition from an unjust to a just society. Unlike Hobbesians, for democratic views justice is not defined as a compromise among given and essentially conflicting interests. So justifying a moral conception does not involve showing each person, whatever their desires, that justice benefits them. Rather, a conception is justified when it conforms to reasonable persons' considered judgments of justice and reasonableness.

To Hobbesians this appears circular and unenlightening. They construe moral justifications reductively, as questions of rational motivation given non-moral interests. Hobbesians encounter a different problem; namely persuading those now advantaged by the status quo that it is rational to accept norms that would be agreed to by self-interested members of a hypothetical state of nature. Some sort of motivation problem is then confronted by any contract view that does not take the current status quo as the basis for agreement. And in the history of social contract thought, there are no significant normative doctrines of status quo agreement.

9 Moral contractualism

The idea that not only justice, but morality generally, is based in general agreement has not been worked out in

detail. Rawls and Gauthier suggest it, but their arguments are limited to justice. The primary difficulty with this thesis is that many duties do not seem to involve reciprocity or mutual advantage. The duties of parents to their young children are not contingent upon children observing their duties to honour and obey their parents. And duties forbidding cruelty to animals or the destruction of nature and valuable species are not easily explained in contractual terms.

One significant effort to account contractually for, not all, but many moral duties is T.M. Scanlon's right-based view. Scanlon (1982, 1988) is influenced by the idea of universal acceptability implicit in Kant's claims that as moral beings we should regard ourselves as legislative members of a kingdom of ends. Scanlon depicts morality as the result of 'hypothetical co-deliberation'. It is 'a theory according to which an act is right if it would be required or allowed by principles which no one, suitably motivated, could reasonably reject as a basis for informed, unforced general agreement' (1988: 15). Interpreting this, imagine a community of free and informed agents, each of whom is conscientious, sincere and motivated by a desire to justify their actions, ends and expectations to everyone else similarly motivated. Morality is the set of public norms which such idealized persons would jointly affirm and commit to, aware that these norms are to regulate their activities and serve as the final standards for justification, criticism and settling conflict.

Reasonableness is a central idea in Scanlon, Rawls and right-based views. It is conceived of as a fundamental moral category, not reducible to rationality. One can be fully rational, by acting best to promote one's interests, and still be unreasonable. A reasonable person respects the needs and interests of others and is not prone to overreaching. To be reasonable is to be responsive to the individual reasons of others. It involves a willingness to limit one's expectations and constrain one's actions and ends by rules that respect the legitimate pursuits of others. Reasonableness also includes a willingness to justify one's actions by reasons others could publicly endorse in uncoerced general agreement. Reasonableness is not altruism. One can be unreasonably altruistic, by promoting a majority's interests while undermining the needs of the few. Reasonable persons are equitably fair-minded; they regulate their acts in ways that respect everyone's basic interests.

Scanlon's contractualism yields an account of morality's subject matter, and so of moral objectivity and validity. Many moral theories assume there is a perspective from which to achieve unanimity, at least among rational persons similarly informed. For example, in Locke's intuitionist moral epistemology, the fundamental law of nature is self-evident to the right reason of every informed and unbiased rational being. Contractualism reverses the order of dependence between unanimity of judgment and moral principles. Rather than moral judgments being true when they match a prior moral order, the moral order is specified in terms of ideal judgment and agreement. Moral truth is then defined by referring to the principles to which free persons would agree under ideal conditions. There is no order of moral facts antecedent to hypothetical co-deliberation and agreement. Moral truth is then discoverable by us, not by rational intuition or moral sense, but by deliberating on what free, informed moral agents equally situated would reasonably agree to.

See also: International relations, philosophy of; Liberalism

SAMUEL FREEMAN

References and further reading

Binmore, K. (1994) *Game Theory and the Social Contract*, Cambridge, MA: MIT Press.(An economist's application of game theory to social contract doctrine.)

Buchanan, J. (1975) *The Limits of Liberty*, Chicago, IL: University of Chicago Press.(Hobbesian argument for classical liberalism.)

Cohen, J. (1986) 'Reflections on Rousseau', *Philosophy and Public Affairs* 15 (3): 275-97.(Critical review of recent books on Rousseau.)

Dworkin, R. (1977) 'Justice and Rights', in *Taking Rights Seriously*, London: Duckworth.(Notable assessment of Rawls' earlier work.)

Freeman, S. (1990) 'Reason and Agreement in Social Contract Views', in *Philosophy and Public Affairs* 19 (2): 122-157(Contrasts Hobbesian and democratic contract views.)

Gauthier, D. (1986) *Morals by Agreement*, Oxford: Clarendon Press.(Central work by major modern representative of Hobbesian contract doctrine.)

Gauthier, D. (1990) *Moral Dealing*, Ithaca, NY: Cornell University Press.(Collected papers.)

Gough, J. (1957) *The Social Contract*, Oxford: Clarendon Press, 2nd edn.(Pre-Rawlsian account of the tradition.)

Hampton, J. (1986) *Hobbes and the Social Contract Tradition*, Cambridge: Cambridge University Press. (Prominent interpretation of Hobbesian tradition.)
Harman, G. (1977) *The Nature of Morality*, New York: Oxford University Press, ch. 9.(Covers moral conventionalism.)
Hegel, G. (1821) *The Philosophy of Right*, trans. H.B. Nisbet, ed. A. Wood, Cambridge: Cambridge University Press, 1991.(Criticism of social contract and general will as individualistic.)
Hobbes, T. (1649) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.(Hobbes' primary statement of his doctrine.)
Hume, D. (1748) 'Of the Original Contract' in *Essays Moral, Political, and Literary*, ed. E. Miller, Indianapolis, IN: Liberty Classics, 1985.(Historically effective criticism of Locke's contract doctrine.)
Kant, I. (1797) *Metaphysical Elements of Justice*, trans. J. Ladd, Indianapolis, IN: Bobbs-Merrill, 1965.(Contains Kant's account of the Original Contract.)
Locke, J. (1690) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1963.(The second treatise contains the doctrine that government must be based in consent.)
Nagel, T. (1991) *Equality and Partiality*, New York: Oxford University Press, chaps 4-5.(On political contractualism.)
Nozick, R. (1974) *Anarchy, State and Utopia*, New York: Basic Books.(Individualist contract account of political authority, and libertarian criticism of Rawls.)
Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(The major contemporary statement of democratic contractarianism.)
Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(Develops the Kantian conception of the person informing Rawls' contract view.)
Rousseau, J. (1762) *Du contrat social (Of the Social Contract)*, in *Rousseau's Political Writings*, ed. A. Ritter and J.C. Bondanella, New York: W.W. Norton, 1988.(Account of the General Will and the Social Contract.)
Sandel, M. (1982) *Liberalism and the Limits of Justice*, Cambridge: Cambridge University Press.(Communitarian criticism of Rawlsian liberalism.)
Scanlon, T. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press.(Initial presentation of moral contractualism.)
Scanlon, T. (1988) 'The Significance of Choice', in *The Tanner Lectures on Human Values*, vol. 8, ed. S.M. McMurrin, Salt Lake City, UT: University of Utah Press, 149-216.(Develops contractualism and discusses freedom of the will.)
Simmons, A.J. (1993) *On the Edge of Anarchy: Locke, Consent, and the Limits of Society*, Princeton, NJ: Princeton University Press.(On the role of consent in Locke's political doctrine.)

Conventionalism

How is it known that every number has a successor, that straight lines can intersect each other no more than once, that causes precede their events, and that the electron either went through the slit or it did not? In cases like these it is not easy to find observable evidence, and it is implausible to postulate special modes of intuitive access to the phenomena in question. Yet such theses are relied on in scientific discourse and can hardly be dismissed as meaningless metaphysical excess. In response to this problem the positivists and empiricists (notably Poincaré, Hilbert, Carnap, Reichenbach and Ayer) developed a strategy known as conventionalism. The idea was that certain statements, including fundamental principles of logic, arithmetic and geometry, are asserted as a matter of conventional stipulation, being no more than definitions of some of their constituent terms; consequently they must be true, our commitment to them cannot but be justified, and the facts in virtue of which they are true are simply the facts of our having made those particular decisions about the use of words. This doctrine was a compelling and powerful weapon in the positivist-empiricist arsenal, evolving throughout the 1920s, 1930s and 1940s. But it fell into disfavour under a barrage of serious challenges due mainly to Quine. How are 'conventions' to be identified as such? How could they possibly provide words with meanings, or have the epistemological import that is claimed for them? How could arbitrary, contingent decisions about the use of words result in the existence of necessary facts? In the absence of satisfactory replies to these objections few philosophers these days believe that conventionalism can settle the semantic, epistemological and metaphysical questions that it was intended to answer. However, certain aspects of the view remain defensible and interesting.

1 The scope and nature of conventionalism

Conventionalism is the strategy of supposing that certain theses are implicit definitions - hence mere choices of language - and hoping thereby to explain how their adoption can be justified and to specify that in which their truth consists. It has been applied to the principles of deductive inference, to the claims of mathematics, and to 'bridge laws' specifying the observational consequences of theoretical hypotheses - all domains in which methods of *empirical* verification might seem to be inappropriate. For example, it was said that the axioms and rules of classical deductive logic are nothing more than implicit definitions of the logical constants, 'and', 'not', 'every', and so on; and this was supposed to show both how our reasoning could be justified and to articulate what it is to be a logical truth. Another famous application was to questions about the structure of space. It was maintained by H. Poincaré (1905) that geometrical facts are matters of convention: thus if we decide to use Euclidean geometry, we thereby implicitly define the terms 'point' and 'line' in such a way that the theorems of Euclidean geometry are true. We are therefore justified in employing any geometry we please; for there are no facts about the structure of space that could make some choices correct and others incorrect. A variant form of conventionalism about space and time was urged by Reichenbach (1968). In his view what stand in need of stipulation are certain correlations between spatiotemporal and physical properties (for example, that light travels in straight lines, and that pendulum swings consume equal periods of time). The idea here was that only once such 'coordinative definitions' had been decided upon would it become possible to investigate empirically the structure of space and time; alternate conventions will yield alternate answers, and so there are no facts of the matter as to which answers are correct (see Reichenbach, H. §4; Relativity theory, philosophical significance of §3).

Thus, to whatever domain it was applied, conventionalism in its standard form consisted in the following four elements: one, that some specified, determinate body of sentences (or rules) is accepted as a matter of conventional stipulation; two, that these commitments implicitly define certain constituent terms; three, that the propositions which the sentences express are known a priori; and four that those propositions are true - and the corresponding facts exist - in virtue of our conventional stipulations. In the following sections we shall consider objections to each of these components of the doctrine.

2 The definition of 'convention'

In the ordinary sense of the word a *convention* is an explicit agreement to act in some specified way. But, in that sense, no convention has ever been adopted to go around saying ' $1 + 1 = 2$'; and similarly for the other cases in which conventionalism was deployed. Thus emerges one of the fundamental difficulties faced by this philosophy: that of explaining how to demarcate 'conventions' or 'stipulations' (in the intended technical sense) from other

utterances we are inclined to make. R. Carnap (1950), perhaps the most sophisticated practitioner of the strategy, equated ‘linguistic conventions’ with the analytic (framework) sentences that determine what our words mean and which language we are using. These were in turn identified as the sentences which we decide to accept on the basis of pragmatic rather than epistemological considerations, and it was concluded that the propositions expressed by those sentences are known a priori. However, Quine’s (1935, 1951, 1954) influential critique of conventionalism made a strong case that although these conceptions of ‘convention’, ‘meaning’, ‘language’, ‘pragmatic reason’, ‘analytic’ and ‘a priori’ may well be linked in the way that Carnap supposed, they are all equally obscure and equally in need of clarification, and that it remains to find a way of breaking out of the circle of inter-definable ideas in order to characterize any one of them objectively (see Quine, W.V. §8).

3 The semantic component of conventionalism

The legitimacy of *deciding* that a certain set of sentences (or rules of inference) is to qualify as correct rested on two assumptions: first, that we are at liberty to decide which language to speak, for example, what our words are to mean; and, second, that in stipulating that a sentence be true we are merely fixing the meaning of one or more of its constituent terms. But the second of these assumptions has received considerable criticism, notably in the writings of Frege (1903, 1906), Prior (1960), Belnap (1962) and Dummett (1977). The thesis in question tends to be construed as follows: that in deciding to regard ‘Čnum;F’ as true, one is implicitly deciding to give ‘F’ whatever meaning it would need to have in order that ‘Čnum;F’ be true, given the already established meaning of the rest of the sentence, ‘Čnum;_’. But this thesis is far from obviously correct. In the first place, how can we be sure that there exists *any* meaning which could be assigned to ‘F’ and which would render ‘Čnum;F’ true? Second, how can we be sure that there is only *one* such meaning? Third, even if there is such a meaning, what reason is there to suppose that this is the meaning that ‘F’ acquires? And fourth, even if it is, how does our treating ‘Čnum;F’ as true *bring it about* that ‘F’ acquires that meaning: what account of the nature of meaning could explain how this happens? The difficulty of answering these questions casts doubt on the possibility of implicit definition.

4 The epistemological component of conventionalism

The main point of the doctrine was to show how our views about logic, arithmetic, spatiotemporal structure, and theory-observation relations could be rational, despite the absence of intuitive or empirical access to those domains. The idea was that our claims are justified a priori - independently of intuition or experience - because they merely fix the meanings of terms, which are commitments that are not subject to a posteriori constraint. Now we have just seen how the feasibility of implicit definition may well be questioned. But there is an additional problem concerning the alleged non-empirical character of such conventions. For even if it is conceded that by regarding a sentence as true one can fix the meaning of one of its constituents, there can remain a doubt as to whether the truth of that sentence is known a priori. The grounds for this doubt derive from Quine’s web-of-belief model of theoretical development, according to which our total theory formulation evolves under the constraints of empirical adequacy, global simplicity and conservativism (see Quine, W.V. §3). For if this model is correct then every element in the web earns its keep empirically by contributing towards the overall system that ‘best’ satisfies these three desiderata, and so everything we believe is a posteriori. This is not to deny that our commitment to ‘Čnum;F’ might implicitly define a constituent term ‘F’. The point rather is that given Quine’s model not even such meaning-constituting principles would be known a priori (see A priori).

5 The metaphysical component of conventionalism

On the metaphysical front, the standard conventionalist stance is blatantly antirealist; for insofar as a certain fact is merely a matter of convention, it evidently does not exist independently of us and our practices (see Scientific realism and antirealism). And this metaphysical thesis has provided a glaring target for criticism. For although it is no doubt a fact about us that we decide to commit ourselves to the truth of, for example, ‘ $1+1=2$’, this fact regarding our commitment is not plausibly identifiable with the arithmetical fact that $1+1=2$. The former fact concerns words, the latter numbers. Moreover, the former is contingent - there are possible worlds in which we decide to adopt different conventions - whereas the latter is necessarily the case - there is no possibility that $1+1=3$. Thus the conventionalist’s metaphysical thesis, implying as it does that if we had spoken a different language then perhaps $1+1$ would have been equal to 3, is extremely unattractive and has helped bring the entire

doctrine into disrepute.

6 Prospects for the defence of conventionalism

The above criticisms need not be considered decisive. In each case replies might be devised, involving modifications of the standard four-pronged doctrine. Thus, in response to the challenge to demarcate the class of conventional stipulations, one might abandon the intuitive concept of ‘convention’, with the presumption that it is evident when such things occur, and look to empirical linguistics for a characterization of which regularities in the use of a term (including assertions containing it) are explanatorily basic and hence implicitly define it. As for the difficulties with implicit definition, one might hope to pre-empt them by altering the assumption that, in regarding ‘Čnum;F’ as true, one is giving ‘F’ the meaning it would need to have for ‘Čnum;F’ to be true. Instead, again invoking a use-conception of meaning, one might suppose that, in regarding ‘Čnum;F’ as true, one is giving ‘F’ the meaning that is constituted by that rule of use (Horwich 1997). In that case the four above-mentioned difficulties could not arise. For the existence and uniqueness of a meaning is ensured by the existence of the regularity, and the explanation of how the convention provides meaning would be trivial.

Turning to the epistemological objection, one might resist the thesis that *everything* is a priori, which seemed to follow from the web-of-belief model. For there can surely be different formulations of the same theory - notational variants of each other. Therefore the decision to adopt one rather than another way of expressing one’s beliefs is a convention that is not subject to a posteriori revision. Thus the acceptance of a theory, Tabc, formulated using new terms ‘a’, ‘b’ and ‘c’, may be regarded as the product of two commitments: one being the substantive existential belief that there are unique theoretical properties, x, y and z, satisfying Txyz; and the other being the linguistic decision to call those properties ‘a’, ‘b’ and ‘c’. This conditional thesis, $(\exists!\mathrm{xyz})(\mathrm{Txyz}) \rightarrow \mathrm{Tabc}$, would not be held a priori (Lewis 1970; Horwich 1986).

Finally, the metaphysical critique of conventionalism could be accepted and accommodated by jettisoning the antirealist component of the standard doctrine. It would still be supposed that certain sentences (‘conventions’) implicitly define certain of their constituents; but it would no longer be maintained that the facts expressed by those sentences are reducible to facts about our practices, hence matters of convention.

Thus one might respond to the variety of objections to conventionalism. But this can hardly qualify as a vindication of the doctrine. For the revised defensible position which the objections have compelled solves none of the epistemological or metaphysical problems for which conventionalism was invented. Moreover, there is now reason to suspect that there were no real problems here in the first place. For Quine’s model suggests that logic, geometry, arithmetic and our theory-observation bridge principles have an epistemological and metaphysical status that is no different from the rest of what we believe. This still leaves open the possibility of implicit definition: the conventional a priori decision to use these words rather than those in the articulation of our beliefs, including those of logic, mathematics and science. But this is such a pale shadow of the original positivistic idea, it might be as well to reserve the term ‘conventionalism’ for that doctrine and consider it refuted.

See also: Language, conventionality of; Logical positivism; Necessary truth and convention

PAUL HORWICH

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, New York: Dover.(Forcefully advocates conventionalism and elaborates its extensive philosophical ramifications.)

Belnap, N.D. (1962) ‘Tonk, Plonk and Plink’, *Analysis* 22, 130-3.(Discusses implicit definition and the question of whether a *unique* meaning is determined. See §3.)

Carnap, R. (1950) ‘Empiricism, Semantics and Ontology’, in *Meaning and Necessity*, Chicago, IL: University of Chicago Press; 2nd edn, 1956.(The most detailed and plausible articulation of pre-Quinean logical empiricism. See introduction and §2.)

Dummett, M. (1977) *Elements of Intuitionism*, Oxford: Clarendon Press.(Argues that the logical constants are not implicitly defined by the laws in which they appear, and hence that classical logic cannot be justified by regarding its fundamental principles as implicit definitions. See §3.)

Frege, G. (1895-1903) ‘Correspondence between Frege and Hilbert’, in *Philosophical and Mathematical Correspondence*, Chicago, IL: University of Chicago Press, 1986, 31-52. (Contains Hilbert’s idea - and

Frege's criticism of it - that the axioms of a theory can fix the meanings of its terms.)
Frege, G. (1903, 1906) 'On the Foundations of Geometry', in B. McGuiness (ed.)*Collected Papers on Mathematical Logic and Philosophy*, Oxford: Blackwell, 1983. (1903, series 1, 1906, series 2. Further articulates Frege's critique of implicit definition. See §3.)
Horwich, P. (1986) 'A Defence of Conventionalism', in G. MacDonald and C. Wright (eds) *Fact, Science and Morality*, Oxford: Blackwell.(Argues that Quine's critique leaves open a form of 'global conventionalism' concerning the language with which to express our total theory formulation. See §6.)
Horwich, P. (1997) 'Implicit Definition, Analytic Truth, and A Priori Knowledge', *Nous*. (Argues that objections to the standard conception of implicit definition may be dissolved in the context of a use-theory of meaning. See §§3, 4, 6.)
Lewis, D. (1970) 'How to Define Theoretical Terms', *Journal of Philosophy* LVII (July): 427-66.(Articulates the strategy, suggested by Carnap, Russell and Ramsey in the 1920s, that any theory can be factored into two independent components: a substantive existential claim and an a priori conditional which defines the theoretical terms. See §6.)
Poincaré, H. (1905) *Science and Hypothesis*, London: Walter Scott Publishing. (Maintains that the facts of geometry are mere matters of convention. See introduction and §1.)
Prior, A.N. (1960) 'The Runabout Inference-Ticket', *Analysis* 21, 38-9.(Shows that an alleged implicit definition presupposes what may well be false: that there exists a meaning that, if possessed by the term in question, would render true the sentences defining the term. See §3.)
Quine, W.V. (1935) 'Truth by Convention', in *The Ways of Paradox*, Cambridge, MA: Harvard University Press, 1976. (Argues that logic cannot derive from explicit conventions. See introduction and §2.)
Quine, W.V. (1951) 'Two Dogmas of Empiricism', in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1953.(Suggests that the alleged distinction between analytic and synthetic statements (and related notions such as synonymy and rule of language) resist clarification and are of dubious coherence. See introduction and §§2, 4.)
Quine, W.V. (1954) 'Carnap and Logical Truth', in *The Ways of Paradox*, Cambridge, MA: Harvard University Press, 1976.(Argues against the view that the fundamental principles of logic are conventions of language. See introduction and §§2, 4.)
Reichenbach, H. (1968) *The Philosophy of Space and Time*, New York: Dover. (Maintains that assumptions about the physical manifestation of straight lines and other geometrical objects are matters of convention, hence that geometry, which results from such assumptions, is also conventional. See introduction and §1.)

Conway, Anne (*c*.1630-79)

Anne Conway (née Finch) was the most important of the few English women who engaged in philosophy in the seventeenth century. Her reputation derives from one work published after her death, Principia philosophiae antiquissimae et recentissimae (1690), which proposes a Neoplatonic system of metaphysics featuring a monistic concept of created substance. The work entails a critique of the dualism of both Descartes and Henry More, as well as of the materialism (as she saw it) of Hobbes and Spinoza. In her concept of the monad and her emphasis on the benevolence of God, Conway's system has some interesting affinities with that of Leibniz.

Anne Conway (née Finch) studied philosophy with the Cambridge Platonist, Henry More, who instructed her by letter since, as a woman, she was debarred from attending university (see Cambridge Platonism). She was introduced to philosophy through the study of Cartesianism in the modified - and therefore critical - version accepted by More. Later in life she encountered the physician and thinker, Francis Mercurius van Helmont, who brought her into contact with alchemical thought and the Jewish Kabbalah, and appears to have provided the catalyst for her turning away from the dualism of her Cartesian and Cambridge-Platonist background. Conway's *Principia philosophiae antiquissimae et recentissimae* (1690) is a Latin translation of a manuscript found among her effects after her death and published under the auspices of van Helmont in a collection entitled *Opuscula philosophica*. It was translated back into English and published as *The Principles of the Most Ancient and Modern Philosophy* in 1692. The original manuscript is lost and the printed version is probably incomplete.

Conway posits three orders of being: God, Christ or *Logos*, and 'Creature' or created substance. Where God is essentially unchanging ('a Spirit, Light, and Life, infinitely Wise, Good, Just, Mighty, Omniscient, Omnipresent, Omnipotent, Creator and Maker of all things visible and invisible', [1690] 1982: 149), creatures are characterized by mutability, with Christ as 'Middle Nature' linking God with creation. Created beings or species are modes of one single spiritual substance which comprises an infinite number of hierarchically arranged spiritual particles or monads. What is commonly taken to be matter is in fact less refined spirit - congealed spirit, as it were. All creatures are made up of a combination of both refined and congealed spirit. Created substance may increase in perfection, becoming more spiritual, or fall away from perfection, becoming more like matter. So also creatures themselves can move up or down the ontological scale, transmuting into higher or lower orders of creature: a grain of dust or sand 'through various and succedaneous Transmutations' may come to acquire the noblest attributes of substance, namely 'a capacity of all kind of Feeling, Sense, and Knowledge, Love, Joy, and Fruition, and all kind of Power and Virtue' ([1690] 1982: 225).

Conway developed her scheme in part as an answer to what she saw as the shortcomings of contemporary philosophy, the key error of which was, in her view, the mechanical concept of matter as inert, devoid of life and sense. As a result, mechanism was unable to account for action and life. Although she commends Descartes for his 'Mechanical Skill and Wisdom' she denies that understanding the laws of local motion can explain the vital operations of nature. Hobbes and Spinoza compound the problem by conceiving God in material terms and thereby confounding God with created things. While she accepts an instrumental role for local motion in the operations of nature, Conway conceives motion to entail change in general, defining it more broadly as a vital operation of the 'Strength, Power and Virtue', intrinsically present in substance as modes of substance and capable of extending their influence beyond that substance. Interaction between creatures is by a process analogous to emanation or radiation. Conway calls this 'virtual extension', that is 'a Motion or Action which a Creature hath… proceeding from the innermost parts' ([1690] 1982: 229) and reaching to other creatures through a proper spiritual medium. This spiritual principle of causality may be likened to a field of force, and provides the dynamics of perfectionism in Conway's ontology.

These modal transmutations of substance are also spiritual in the religious sense of that term, since the degeneration away from perfect spiritual substance which is entailed in the monads' becoming more like matter is also a falling away from the goodness of God. By the same token the purification of substance to become more spirit-like involves a return towards goodness and hence towards God. Conway's concept of substance thus underpins her soteriology: she denies the eternity of Hell, regarding the punishment of sin as corrective and ultimately salvific. Thus her overall ontological and spiritual scheme is in the end ameliorative, suffering being

explained as part of the process of moral and substantial purification by which ‘creature’ is restored to a more spiritual and therefore more divine state. Both in her concept of the monad and in her emphasis on God’s benevolence, her system has obvious affinities with that of Leibniz. Yet, although he expressed admiration for her, it is unlikely that she influenced him directly.

See also: Salvation

SARAH HUTTON

List of works

Conway, A. (1690) *Principia philosophiae antiquissimae et recentissimae de Deo, Christo et Creatura id est de materia et spiritu in genere*, in *Opuscula Philosophica*, Amsterdam; trans. as *The Principles of the Most Ancient and Modern Philosophy*, London, 1692; repr. ed. P. Loptson, Dordrecht: Kluwer Academic Publishers, 1982. New trans. by A. Condert and T. Corse, Cambridge: Cambridge University Press, 1996.(Latin translation of a now-lost English original, and an English translation of this. The introduction to the modern reprint is not wholly reliable.)

Conway, A. (1992) *The Conway Letters: the Correspondence of Anne, Viscountess Conway, Henry More and their Friends, 1642-1684*, ed. M. Nicolson and S. Hutton, Oxford: Clarendon Press.(Letters with introduction to the intellectual context, bibliography and biographical commentary.)

References and further reading

Brown, S. (1990) ‘Leibniz and Henry More’s Cabbalistic Circle’, in S. Hutton (ed.) *Henry More (1614-1687): Tercentenary Studies*, Dordrecht: Kluwer Academic Publishers.(Challenges the view that Conway influenced Leibniz.)

Coudert, A. (1975) ‘A Cambridge Platonist’s Kabbalist Nightmare’, *Journal of the History of Ideas* 36: 633-52. (On Henry More, F.M. van Helmont, Anne Conway and the *Kabbala denudata*.)

Gabbey, A. (1977) ‘Anne Conway et Henry More: lettres sur Descartes’ (Anne Conway and Henry More: Letters on Descartes), *Archives de Philosophie* 40: 379-404.(On the discussion of Cartesianism in the early correspondence of More and Conway.)

Hutton, S. (1995) ‘Anne Conway critique de Henry More: l’esprit et la matière’ (Anne Conway, critic of Henry More: spirit and matter), *Archives de Philosophie* 58: 371-84.(On the relationship of Conway’s thought to that of Henry More.)

Merchant, C. (1979) ‘The Vitalism of Anne Conway: its Impact on Leibniz’s Concept of the Monad’, *Journal of the History of Philosophy* 17: 255-69.(Argues that Conway influenced Leibniz by showing parallels between Leibniz and Conway.)

Popkin, R.H. (1990) ‘The Spiritualistic Cosmologies of Henry More and Anne Conway’, in S. Hutton (ed.) *Henry More (1614-1687): Tercentenary Studies*, Dordrecht: Kluwer Academic Publishers.(On Conway’s and More’s alternative to the mechanical philosophy.)

Copernicus, Nicolaus (1473-1543)

Copernicus argued that the earth is a planet revolving around the sun, as well as rotating on its own axis. His work marked the culmination of a tradition of mathematical astronomy stretching back beyond Ptolemy, to the Greeks and Babylonians. Though it was associated with methods and assumptions that had been familiar for centuries, it was also revolutionary because of its implications for the relations between humankind and the universe at large.

Around 1513 Copernicus composed the *Commentariolus*, a brief draft describing a new sun-centred system of astronomy. A few copies were circulated among astronomers, but it was not published until rediscovered in 1878. In the years that followed, he worked on a systematic programme of observation to allow accurate recomputation of the parameters needed for his heliocentric model. Despite his own conviction, however, Copernicus could not find a cogent argument for the counterintuitive claim that the earth *really* moves. It was only in 1540, when his disciple Georg Rheticus gave the first published account of the new system in his *Narratio prima*, that Copernicus was encouraged to complete and publish his long-planned work, the *De revolutionibus orbium coelestium* (1543). His friend, the Lutheran theologian Andreas Osiander, prefaced the work with an anonymous foreword asserting the hypotheses of the work were to be regarded as 'useful mathematical fictions only', thus misleading the first generation of readers into an underestimation of the revolutionary character of Copernicus' proposal. Only later did readers like Bruno and Kepler draw attention to the discrepancy between the message of the foreword and that of the main text.

In the opening lines of the *Commentariolus*, Copernicus describes what prompted him to undertake the immense labour of reconstructing and recomputing the entire system of astronomy. Ptolemy's equants, mathematical devices that allowed planets to move nonuniformly on the circumference of their circles, 'seemed neither sufficiently absolute nor sufficiently pleasing to the mind' (see Ptolemy). Copernicus discovered that placing the sun at the centre not only allowed him to save the phenomena, but also explained a variety of puzzling features of the planetary motions that Ptolemy had to leave as brute fact. In the heliocentric system, these turn out to be 'natural', necessitated by the fact that we are observing the sky, not from the centre, but from a moving vantage-point. Thus he could claim to have brought closer the 'mathematical' and the 'physical' approaches to astronomy that had so long been separated.

Copernicus attributed three motions to the earth. The first was a daily rotation on its own axis, the second a yearly revolution around the sun, the third, an annual motion in declination of the earth's axis (which later turned out to be unneeded). Attributing an annual motion to the earth allowed him to dispense with the larger planetary epicycles of Ptolemy, but the planets still required smaller epicycles in order to eliminate the equants. The first major achievement of Copernicus' system was that the relative distances of the planets could be determined, which the Ptolemaic model could not do. But other advantages were also plain. Making the earth the third planet explained the retrograde motions of the planets: the earth was lapping the outer planets and being lapped by the inner ones. The dimensions and periods of these apparent loops in the planetary paths could now be derived from the relative distances of the planets from the sun and their periods of revolution; in the Ptolemaic model, by contrast, the retrograde motions were an 'add-on', requiring arbitrary stipulation of epicycle parameters. Copernicus also noted why the outer planets are nearest and therefore brightest when in opposition (at the opposite side of the sky from the sun); in Ptolemy's scheme there is no particular reason why this should be the case. Finally, he could explain why Venus and Mercury are 'tied', as it were, to the sun, never appearing more than a certain maximum angular distance from it in the sky. Inner planets in a heliocentric system will *necessarily* appear 'tied' in this way; a puzzle for Ptolemy, because it linked the motions of these planets to the motion of the sun in a way that could be arbitrarily postulated but that defied explanation in his model.

Copernicus was a scientific realist (see Scientific realism and antirealism §1). He knew that the constructions of mathematical astronomers were traditionally regarded as no more than helpful fictions. He knew too that objections to the earth's motion would be urged by those who took a literalist approach to the wording of the Bible. Yet he was confident that he had shown that the earth *really* is in motion. Some recent historians have suggested that he was motivated by a Neoplatonic metaphysical belief in the cosmic primacy of the sun but there is

little real evidence for this. In book one of *De revolutionibus*, he makes the reasons for his realistic stand clear enough. He does not appeal to a superior accuracy of prediction nor to a reduction in the number of circles deployed. Rather, his system displays, he says, a 'harmony' that its rivals lack. And this is not an expression merely of an aesthetic preference. Superior explanatory power was for Copernicus something more than saving the phenomena: it convinced him that he had hit on the truth with regard to the earth's motions.

See also: Cosmology; Explanation; Kuhn, T.S.

ERNAN McMULLIN

List of works

Copernicus, N. (1543) *De revolutionibus orbium coelestium (On the Revolutions of the Heavenly Spheres)*, in J. Dobrzycki (ed.) and E. Rosen (trans. and commentary) London: Macmillan, 1978.(Copernicus' major work, with a detailed commentary.)

Copernicus, N. (1939) *Three Copernican Treatises*, ed. E. Rosen, New York: Dover. (Translations of the *Commentariolus*, Copernicus' *Letter against Werner*, and Rheticus' *Narratio prima*. Also a useful annotated bibliography.)

References and further reading

Armitage, A. (1947) *The World of Copernicus*, New York: Mentor.(Popular biography.)

Kuhn, T. (1957) *The Copernican Revolution*, New York: Random House.(A classic work, containing a particularly useful chapter on Copernican astronomy.)

Rosen, E. (1970) 'Was Copernicus a Hermetist?', in R.G. Stuewer (ed.) *Historical and Philosophical Perspectives of Science*, Minnesota Studies in the Philosophy of Science, Minneapolis, MN: University of Minnesota Press, vol. 5, 163-71.(A decided 'no' to the question posed.)

Swerdlow, N.M. and Neugebauer, O. (1984) *Mathematical Astronomy in Copernicus's De Revolutionibus*, New York: Springer, 2 vols.(A detailed technical analysis of every aspect of Copernicus' work, with a useful short biography.)

Westman, R. (ed.) (1975) *The Copernican Achievement*, Berkeley, CA: University of California Press.(A set of essays marking the fifth centenary of Copernicus' birth. Lively scholarship on a wide array of Copernican topics.)

Westman, R. (1994) 'Two Cultures or One? A Second Look at Kuhn's *The Copernican Revolution*', *Isis* 85: 79-115.(Review of recent scholarship on the Copernican revolution generally.)

Cordemoy, Géraud de (1626-84)

Géraud de Cordemoy was, by profession, first a lawyer, then a tutor to the Dauphin (the future Louis XV). But he was also one of the more important Cartesian philosophers in seventeenth-century France. In Le discernement du corps et de l'ame, Cordemoy defended a strict dualist and mechanist philosophy. But his Cartesianism was unorthodox, since he introduced indivisible atoms into his natural philosophy and was one of the first to argue for occasionalism, the doctrine that God alone is a true causal agent. He also wrote an important work on the nature and origins of speech and language, Le discours physique de la parole.

Géraud de Cordemoy was born in Paris on 6 October 1626. A lawyer by profession, from 1657 onwards he regularly attended occasions at which Cartesian philosophical ideas were discussed, and his *Discours de l'action des corps* was incorporated into the 1664 edition of Descartes' *Le Monde*. His major philosophical work, *Le discernement du corps et de l'ame* (1666), was thoroughly Cartesian in inspiration, and was followed two years later by *Le Discours physique de la parole*. By this time his broad interests included politics, pedagogy and history. In 1673, through the influence of Bossuet, he was appointed *lecteur ordinaire* to the Dauphin and joined the vibrant intellectual circles at the court of Louis XIV. He also began research into the history of France, published posthumously as *Histoire de France*. In 1675 he was elected to the Académie Française, of which he later became director. He died on 15 October 1684.

Most of Cordemoy's philosophical work involved an unorthodox defence and development of Descartes' system. He worked hard, for example, at reconciling Descartes' cosmogony with the biblical account of creation. More important, however, are the modifications he introduced into Cartesian metaphysics and natural philosophy with respect to the nature of matter and causation.

Descartes' matter, or body, understood as pure extension, is by nature divisible *ad indefinitum*. (Descartes preferred to say that the extension of the world is indefinite but not infinite - only God is 'infinite.') Hence, there are no in-principle indivisible atoms. Moreover, whatever is extended is body, so that the universe is a plenum. Cordemoy accepted the Cartesian project of mechanistic physics according to which all natural phenomena are to be explained solely by means of the motion and rest of insensible particles of matter, each of which is characterized by geometrical features alone. But he held that the only consistent form of mechanism is atomism. In the *Discernement*, he distinguished matter from extended substance, or body. Substance is essentially indivisible. Bodily substances, then, are indivisible, impenetrable, extended atoms, each with an invariable figure. Matter, on the other hand, is constituted by collections of atomic bodies, between which there are empty spaces. Cordemoy argued that only an atomism can consistently account in a non-relative way for the individuality and integrity of bodies. Moreover, a material plenum would render motion problematic and rectilinear motion unnatural, contrary to Descartes' own physical laws.

The *Discernement* next turns to the nature of causal relations, both between bodies and between bodies and minds. Cordemoy is generally regarded (along with La Forge) as one of the originators of the modern doctrine of occasionalism (see Occasionalism; Malebranche §3). According to Cordemoy, bodies and minds are devoid of any real causal efficacy. One body cannot cause motion in itself or in another body, nor can it cause an event in the mind. And finite minds cannot cause mental events or move bodies. Only God, an infinite spirit, is a true causal agent. Cordemoy's argumentation proceeds in an a priori manner, starting from definitions and self-evident axioms. In the body-body case, one body cannot be the cause of motion in another body because motion is a state of a body, and states cannot be communicated from one substance to another. Moreover, no body causes motion in itself, since nothing has 'from itself' that which it can lose without ceasing to be what it is (one of Cordemoy's axioms), and a body can lose its motion without ceasing to be a body. It follows that the first mover of bodies is not a body, since such a first mover must have motion from itself. But there are only two sorts of substances: mind and body. Hence, the first mover of bodies must be a mind. Another axiom then states that an action can be continued only by the agent that initiated it. Thus, the mind that first moved bodies continues to move them. It is not *our* minds that move bodies. This is especially clear from the fact that bodies are usually moved independently of the volitions of any finite mind. This is true not just of inanimate bodies, but also of a great deal of the motion that takes place in the human body. What misleads us into attributing causal efficacy to minds and bodies is just

the constant regularity in the succession of states of things. But, Cordemoy says (as Hume would repeat), this regular succession is *all* that experience reveals to us (see Hume, D. §2). The true cause can only be an infinite, all-powerful mind - God - who, acting on the occasion of the antecedent event, brings about the appropriate effect. Taken in a broader context, Cordemoy's occasionalist project is to demonstrate the ultimate metaphysical foundations of the physics of Cartesian (but atomistic) bodies.

Cordemoy's other important philosophical contribution, while not present in the *Discernement* itself, is none the less intimately related to that work's mechanistic science of body and dualist anthropology. In *Le Discours physique de la parole*, Cordemoy's stated aim is an examination of the nature of speech and language. He studies both that part played in linguistic communication by the purely physical mechanisms of the body, and the semantic aspect contributed by the mind. He considers, as well, the process of learning language and certain questions (some ethical) related to eloquence.

Cordemoy's atomist and occasionalist modifications of Descartes' philosophy drew strong critical responses from other Cartesians (such as Dom Robert Desgabets). But he also impressed later thinkers, and Malebranche and Leibniz acknowledged his influence. More recently, the linguist Noam Chomsky identifies Cordemoy as an important contributor to the development of rationalist linguistics.

See also: Atomism, ancient; Descartes, R.; Gassendi, P.

STEVEN NADLER

List of works

Cordemoy, G. (1968) *Oeuvres philosophiques (Philosophical works)*, ed. P. Clair and F. Girbal, Paris: Presses Universitaires de France.(Contains most of the works listed below, along with a 'Bio-Bibliography'.)

Cordemoy, G. (1664) *Discours de l'action des corps (Discourse on the action on bodies)*, incorporated into Descartes' *Le Monde*, Paris: Jacques Le Gras.(A treatise on mechanist physics.)

Cordemoy, G. (1666) *Le Discernement du Corps et de l'Ame en six discours pour servir à l'éclaircissement de la physique (The distinction of the body and the soul in six discourses to serve as the elucidation of physics)*, Paris.(Cordemoy's analysis of body and of its distinction from and relationship to the mind.)

Cordemoy, G. (1668a) *Discours physique de la parole (Physical discourse on language)*, Paris. (An influential study of the nature, origin, and uses of speech and language.)

Cordemoy, G. (1668b) *Lettre écrite au R.P. Cossart de la Compagnie de Jésus (Letter written to Father Cossart of the Jesuits)*, in *Les Oeuvres de feu M. de Cordemoy (The works of the late M. de Cordemoy)*, Paris: Remy, 1704.(Cordemoy's attempt to reconcile Descartes' cosmogony with the biblical account of creation.)

Cordemoy, G. (1691) *Traité de Metaphysique (Treatise on metaphysics)*, Paris.(A brief discussion of human happiness and freedom.)

Cordemoy, G. (1704) *Histoire de France (History of France)*, in *Les Oeuvres de feu M. de Cordemoy (The works of the late M. de Cordemoy)*, Paris: Remy.(Cordemoy's research into the history of France.)

References and further reading

Balz, A.G.A. (1951) *Cartesian Studies*, New York: Columbia University Press. (Contains a useful chapter on Cordemoy's philosophy in general.)

Battail, J.-F. (1973) *L'avocat philosophe Géraud de Cordemoy (The lawyer-philosopher Géraud de Cordemoy)*, The Hague: Martinus Nijhoff.(A first-rate study of Cordemoy's life and thought.)

Chomsky, N. (1966) *Cartesian Linguistics*, Lanham, MD: University Press of America, 1983.(Places Cordemoy's work on language in the history of what the author calls 'rationalist linguistics'.)

Mouy, P. (1934) *Le développement de la physique cartésienne 1646-1712 (The development of Cartesian physics)*, Paris: J.Vrin.(Still the best account of the history of Cartesian physics in the seventeenth century, with some material on Cordemoy's atomism.)

Prost, J. (1907) *Essai sur l'atomisme et l'occasionalisme dans la philosophie cartésienne (Essay on atomism and occasionalism in Cartesian philosophy)*, Paris: Henry Paulin.(Contains several chapters on Cordemoy, focusing on both his atomism and his role in the genesis of occasionalism.)

Corruption

Corruption denotes decay or perversion. The term implies that there is a natural or normal standard of functioning or conduct from which the corrupt state of affairs or action deviates. When we talk of a person becoming corrupt, we mean not just that they have broken a rule, but that the basic norms of ethical conduct no longer have any force for them. Corruption strikes at the root of a thing.

Political corruption involves the decay or perversion of political rule. Broadly, this occurs when a group or individual subverts a society's publicly endorsed practices for conciliating conflicts and pursuing the common good so as to gain illegitimate advantage for their interests in the political process. The precise specification of the nature and dynamics of corruption is inherently controversial. Classical accounts associate it with a collapse of civic virtue and the eventual destruction of the state. Modern theories focus more narrowly on the misuse of public office for private gain.

1 The historical tradition

The corruption of politics, as the decay of the capacity to pursue justice or the common good, was a central concern for political philosophers in antiquity, and it remained so for their republican successors until the end of the eighteenth century. Two paradigmatic accounts of this process are in books 8 and 9 of Plato's *Republic* (*c*.380-367 BC), which sketch a political sociology and social psychology of the decay of the ideal state into ever more debased and less fully human forms (see Plato); and book 3 of Thucydides' *History of the Peloponnesian War* (*c*.424-399 BC), in which the fall of Corcyrus is depicted in terms of the complete subversion of the political, social and moral order of the state (see Thucydides). For classical authors, corruption was only incidentally a case of bribery, malfeasance or profiteering; its core sense concerned the systematic subversion of those practices, institutions and beliefs that provided a society with a shared understanding and a shared set of purposes. Within this tradition we find accounts, of varying degrees of sociological subtlety, of the causal and material preconditions for political rule and for its decay. These centre around the destabilizing effects of economic inequality among citizens, the enervating effects of luxury, concern with the size of states and their populations, the role of military training and combat, the appropriate education for citizens, the threat of foreign contamination and the best constitutional structure for the state. Without attention to these factors, the capacity of citizens to live virtuously and to recognize the demands of the common good becomes corrupted and the state falls.

This part-normative, part-sociological project of theorizing about the material and cultural preconditions for the stability of states exerted a powerful influence on Western traditions of political thought from Polybius to Rousseau, with Machiavelli playing a central role. Through Aristotle, it had still wider influence on thinkers such as the fourteenth-century North African philosopher of history, Ibn Khaldun. In the UK it played a major part in the language of political opposition in the eighteenth century; thence it came to influence both American and French revolutionaries.

From the early modern period, however, there have been two counter-tendencies to this classical tradition. One denies that there is a common good for politics to pursue; instead, politics involves the pursuit of interests within institutional constraints, and claims about corruption are dismissed as emotive and as wholly motivated by factional or individual interests (as in Hobbes). The other narrows the focus, understanding corruption not as a systemic phenomenon but as a form of rule infraction or deviance by holders of public office. A precondition for this view is the separation of state and society, and the belief that civil society functions relatively autonomously from the state - corruption in government becomes seen as unrelated to the activities of social and economic life, and its solution is believed to lie with the construction of systems of scrutiny and mechanisms of accountability, rather than with the creation and sustaining of widespread citizen virtue.

2 Corruption in political science

Modern conceptions of corruption have tended to follow the more restricted ambitions of this view of corruption as individual rule infraction. Political scientists operate with two main, rival definitions: one centred around public office and one centred around public interest. The public office conception of corruption is exemplified by Nye: 'Corruption is behaviour which deviates from the formal duties of a public role because of private regarding

(personal, close family, private clique) pecuniary or status gains; or violates rules against the exercise of certain types of private regarding influence' (1967: 966).

For Nye, questions of the public interest are questions about the effects of corruption and should not be taken as definitional of the phenomenon. Others, however, have sought to define political corruption precisely in terms of a conception of the public interest. Carl Friedrich argues that we have corruption wherever an office-holder or functionary is induced by non-legitimated rewards to act in ways favouring the rewarder and thereby doing damage to the public and its interests (Heidenheimer *et al.* 1989: 10) (see Public interest §3).

Both public office and public interest definitions of corruption, however, must identify the norms by which we are to define the character and scope of public office or the public interest. Both definitions rely on such norms, but it is not obvious which norms we should accept. There may be little consensus within a society; official and popular mores may not coincide; and what people say about these norms may not coincide with how they behave. The law is equally deficient as a definitive source for norms, not least because laws can be the product of corrupt practices. Moreover, laws regulating political conduct rest on prior assumptions about the character of political office and the appropriate rules for its conduct - as such, legal norms are better understood as drawing on, rather than underpinning our understanding of public office.

These difficulties in identifying the norms of public office or the standard for the public interest ensure that the study of political corruption remains an unavoidably normative exercise. They also show that any theory of political corruption will require an understanding of the distinctive character of political activity and an account of the conditions in which this activity can resist subversion by competing values and activities in society.

A similar conclusion can be drawn by pressing the distinction between definitions of public-office and of public-interest. The conduct of public office, for the most part, is not (and cannot be) exhaustively prescribed by systems of rules. Most offices carry considerable discretion and flexibility, so as to allow occupants to respond to the contingencies of political life. But although the rules of office are open in this sense, the end or basic purpose of the office is unquestionably that it be guided by considerations of the public interest. Moreover, in identifying a rule violation as corrupt we are inevitably led to ask in whose interests the rule was broken. Depending on the answer to that question, we may judge an action as foolish, high-handed or misconceived, rather than corrupt.

3 The core of political corruption

These definitional problems, it should be emphasized, do not arise because the term corruption is unclear. On the contrary, its meaning is relatively uncontentious: it is rooted in the sense of a thing being changed from its naturally sound condition. The problem arises in applying this to politics since there is no general consensus on the 'naturally sound condition of politics'. Public office and public interest definitions focus on different aspects of the political; their disagreement is not over corruption, but over which criteria are appropriate to identify the naturally sound condition from which corrupt behaviour deviates.

It is at this point that the classical republican tradition, conceiving corruption in terms of the decay of the distinctive realm of human action which is politics, has renewed appeal (see Republicanism). Classical views identified a natural condition to politics, conceived in teleological terms, which provided the normative basis or standard which corrupt acts were seen as subverting or otherwise destroying. Modern political science is deeply uncomfortable with such normative commitments and yet, if we are to use the term coherently, such normative commitments are essential. They are essential because we cannot identify political corruption - and if we cannot identify it we cannot understand or explain it - without a commitment to some conception of the nature of politics which these acts subvert. Corruption is, then, the 'other' of politics; it is what must be overcome for properly political rule to be possible. The classical republican tradition had the conceptual and normative framework to recognize this relationship, but the scientism and historicism of the nineteenth and twentieth centuries, and the shift in concern from the conditions for attaining a commonality of purpose and interest within the polity to the problem of integrating the potentially insurgent masses into the bureaucratic politics of the modern nation state, served to displace the insights of this older paradigm. Indeed, political scientists have tended to concentrate their efforts on the understanding of corruption in deviant cases, such as in authoritarian states and command economies or as a transitional phenomenon in newly democratized regimes undergoing rapid economic development and social change, and have generally assumed that the modernization and liberalization of the state solves such

problems. Above all, in sharp contrast to the classical model, there has been very little attention paid to the correlation between the maintenance of an active civic culture and the incidence of corruption within the political domain. The result has been the development of models of politics which have, by eschewing questions of the distinctive nature or character of politics and underplaying the relevance of the mores and beliefs of civil society in the incidence of corruption, largely failed to provide an adequate understanding of what remains a widespread phenomenon in many modern states.

MARK PHILP

References and further reading

Dobel, J.P. (1978) 'Corruption of a State', *American Political Science Review* 72 (3): 958-73.(An account of the classical tradition of conceptualizing corruption.)

Euben, J.P. (1989) 'Corruption', in T. Ball, J. Farr and R.L. Hanson (eds) *Political Innovation and Conceptual Change*, Cambridge: Cambridge University Press, 220-46.(Contrasts the classical tradition with more modern traditions in political theory.)

Friedrich, C.J. (1966) 'Political Pathology', *Political Quarterly* 37 (1): 74; and Heidenheimer *et al.* (1989) *Political Corruption: A Handbook*, New Brunswick, NJ: Transaction Press, 10.

Heidenheimer, A.J. (1970) *Political Corruption: Readings in Comparative Analysis*, New York: Holt, Reinhart & Winston.(An early collection of articles, many of which also appear in Heidenheimer *et al.* (1989).)

Heidenheimer, A.J., Johnston, M. and Levine, V.T. (eds) (1989) *Political Corruption: A Handbook*, New Brunswick, NJ: Transaction Press.(The standard work on the subject in politics.)

Ibn Khaldun (1377) *The Muqaddimah: An Introduction to History*, trans. F. Rosenthal, London: Routledge & Kegan Paul, 3 vols, 1958.(See volume 1, chapters 2 and 3, for a succinct statement of Khaldun's understanding of the corruption of the state.)

Nye, J.S. (1967) 'Corruption and Political Development: A Cost-Benefit Analysis', *American Political Science Review* 61 (2): 417-27; reprinted in and cited from Heidenheimer *et al.* (1989) *Political Corruption: A Handbook*, New Brunswick, NJ: Transaction Press, 963-83.(A major exponent of the view that corruption may be beneficial to the economic development of a society.)

Philp, M. (1996) 'Defining Political Corruption', *Political Studies* 45 (3).(An examination of the problems of defining political corruption which links the definition to competing conceptions of the nature of politics.)

Plato (*c*.380-367 BC) *Republic*, trans. P. Shorey, Loeb Classical Library, Cambridge, MA: Harvard University Press, 1930.(See especially books 8 and 9.)

Putnam, R.D. (1993) *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton, NJ: Princeton University Press.(An attempt to show empirically the connection between civic virtue and low levels of corruption.)

Scott, J.C. (1972) *Comparative Political Corruption*, Englewood Cliffs, NJ: Prentice Hall.(A classic study of corruption, focusing primarily on machine politics.)

Thucydides (*c*.424-399 BC) *History of the Peloponnesian War*, trans. C.F. Smith, Loeb Classical Library, Cambridge, MA: Harvard University Press, 1921.(See especially book 3, §§69-85.)

Cosmology

The term 'cosmology' has three main uses. At its most general, it designates a worldview, for example, the Mayan cosmology. In the early eighteenth century, shortly after the term made its first appearance, Christian Wolff used it to draw a distinction between physics, the empirical study of the material world, and cosmology, the branch of metaphysics dealing with material nature in its most general aspects. This usage remained popular into the twentieth century, especially among Kantian and neo-scholastic philosophers. But recent developments in science that allow the construction of plausible universe models have, effectively, pre-empted the use of the term in order to designate the science that deals with the origins and structure of the physical universe as a whole.

Cosmology may be said to have gone through three major phases, each associated with a single major figure - Aristotle, Newton and Einstein. The ancient Greeks were the first to attempt to give a reasoned account of the cosmos. Aristotle constructed a complex interlocking set of spheres centred on an immovable central earth to account for the motions of the heavenly bodies. Newton formulated a theory of gravitational force that required space and time to be both absolute and infinite. Though the laws of nature could, in principle, be specified, nothing could be said about the origins or overall structure of the cosmos. In 1915, Einstein proposed a general theory of relativity whose field-equations could be satisfied by numerous universe-models. Hubble's discovery of the galactic red shift in 1929 led Lemaître in 1931 to choose from among these alternatives an expanding-universe model, which, though challenged in the 1950s by a rival steady-state theory, became the 'standard' view after the cosmic microwave background radiation it had predicted was observed in 1964. The 'Big Bang' theory has since been modified in one important respect by the addition of an inflationary episode in the first fraction of a second of cosmic expansion. As a 'cosmic' theory, it continues to raise issues of special interest to philosophers.

1 Aristotle's cosmos

Cosmology and philosophy came into existence together in the Greek-speaking world of the sixth century BC when some daring minds sought a 'reasoned account' of the origins and nature of the universe. They assumed that the complex must originally have come from the simple, and thus sought clues in the world around them for the elements and processes responsible, resting their claims on argument rather than on the authority of a tradition. Plato was sharply critical of his predecessors' lack of attention to the traces of mind he saw everywhere in the world around; his *Timaeus* took the universe to be the work of a demiurge (artificer) who imposed partial order on a pre-existent and recalcitrant receptacle (see Plato §16). He acknowledged that his account fell short of knowledge proper. But it could still claim the status of a likely story.

Aristotle went much further, and the cosmology he constructed would substantially endure for two millennia. Beginning from the sophisticated geometrical combinations of circles whereby his contemporaries Eudoxus and Callippus had sought to 'save' the motions of the seven 'wandering stars', he proposed a set of fifty-five rotating concentric spheres, the poles of each carried by the sphere next outside it, and each given its own motion of rotation by an immanent mover, akin to a soul. The complex motion of each planet is the resultant of the motions of its own cluster of spheres, the innermost of which carries the planet. The physics of earth does not apply to these spheres, whose eternal circular motions are quite unlike the rectilinear natural motions found on earth. Nothing much further can be said about the nature of the spheres and the celestial bodies they carry except that they are incorruptible (see Aristotle §16).

The aim of Aristotle's cosmology was to give a plausible causal explanation of the planetary motions, and in this it succeeded. It was not of much practical use for the needs of working astronomers, and its concentric spheres could not explain the known regular variations in planetary brightness. The epicycles and eccentrics first introduced by Apollonius and Hipparchus and later welded by Ptolemy into an accurately predictive geometrical scheme were of much more use for practical ends, but could not compete with Aristotle's ingenious model in explanatory appeal. The fact that the two approaches could not be reconciled was troublesome, but medieval philosophers grew accustomed to arguing that the truth of things would be found in 'physics', that is, in causal explanation, even though 'mathematics' would perhaps be more effective for saving the phenomena.

2 Newton's infinites

Aristotle's cosmology fell as the Copernican heliocentric model (see Copernicus) gradually proved itself in the hands of Kepler and Galileo. In his *Astronomia nova* (1609), Kepler argued that a proper cosmology, one that could claim to give a true account of the motions and distances of the planets, would have to elucidate the causes of those motions and not merely save the phenomena. His own attempts to explain the motions in terms of forces acting at a distance were not successful. It was left to Newton to carry this programme through, utilizing a gravitational attraction that varied inversely as the square of the distance. Leibniz and the Cartesian physicists of the day were willing to concede that the new mechanics of Newton's *Principia* (1687) could save the phenomena, but they insisted that action at a distance could not be accounted an explanation, and hence that the *Principia* did not support a genuine cosmology (see Mechanics, classical §3; Newton, I.).

In order to support the reality of the forces he postulated, Newton introduced the idea of 'absolute' motion (changes in which would be the indication of forces acting) and its corollaries, absolute space and absolute time. Absolute space is Euclidean and thus boundless; it exists independently of any material contents, unlike an Aristotelian place. Absolute time 'flows equably without relation to anything external'; it is thus independent of actual physical change (unlike Aristotelian time) and it is without beginning or end. Not much purchase here for questions about the origin and extent of the material universe! A century later, indeed, Kant would argue that the proclivity of human reason to pose cosmological questions such as these, which necessarily (in his view) lead to antinomy, forces a radical critique of pure reason itself (see Kant, I. §8).

Newton made one significant cosmological claim. In Aristotle's universe, terrestrial and celestial bodies were altogether different in nature, so that a simple science of the two was impossible. Newton showed, however, that a single mechanics governed earth and planets, and in the boldly inductivist third rule of reasoning of the *Principia* postulated that qualities characterizing bodies at the local level should be deemed to characterize the universe generally. The inverse square law of gravitation could, on this basis, be portrayed as a cosmological law governing the attraction between any two particles in the universe. Kant was the first to see just how this might explain the origin of such stable cosmic structures as planetary systems or even the Milky Way. Questions about origin or structure at the level of the universe, however, still seemed fruitless.

3 The expanding universe

In 1915, Einstein formulated a general theory of relativity, employing non-Euclidean geometry to describe a unified spacetime in which the curvature of spacetime replaces gravitational force as a cause of motion (see Spacetime §3; Relativity theory, philosophical significance of §§4-7). Universe models could now be constructed as solutions to the field equations of the new theory, though constraints were needed on the large number of possible solutions. A variety of 'natural assumptions' were introduced regarding large-scale cosmic structure, some of them given the status of 'cosmological principles'. Notable amongst these was Einstein's daring conjecture that matter should be distributed uniformly over space on the largest scale. When Einstein discovered that his equations tended to yield unstable models, he added a cosmological constant representing a repulsion that increased with distance, since he assumed that a plausible universe model ought to be static. Alexander Friedmann (1924) and Georges Lemaître (1927) proposed, instead, nonstatic solutions in which space itself could be expanding.

Hubble's discovery of the galactic red shift in 1929 was the next significant development. It indicated that the galaxies are moving away from us, and from one another, at speeds proportional to their distances. Lemaître (1931) realized how neatly this fitted the nonstatic models of general relativity and formulated the theory of the expanding universe, later called the 'Big Bang' theory. From the rate of expansion, one could calculate an upper limit on the length of time since the expansion began from a single enormously condensed volume. An answer to the ancient cosmological query about origins suddenly seemed to be within reach. But a disappointment was in store. Using Hubble's figure for the expansion constant, the limit-age came out as roughly two billion years, much shorter than the estimated age of the solar system or of the oldest stars.

This was one, though not the principal, motivation for the construction of an alternative 'steady-state' model by Bondi and Gold (1947). It rested on what they described as the 'perfect' cosmological principle, requiring a homogeneous distribution of matter in an infinite space and throughout an infinite time. To account for the Hubble expansion, they postulated the continuous creation of hydrogen throughout space at a rate too low to be detectable, sufficient to allow new galaxies to replace the old as the old move out of view. Baade (1952) showed that Hubble's

estimate for galactic distances was too low and that the ‘Big Bang’ age in consequence was a much safer ten to twenty billion years. And in the late 1950s, evidence began to mount that the relative abundance of galaxies of different types varied with distance, contrary to the ‘perfect’ principle.

But it was the discovery by Penzias and Wilson of the uniform cosmic microwave background radiation, reflecting a 3K black-body temperature distribution, that proved decisive (1965). Alpher and Herman had predicted long before (1948) that just such a relic of radiation should be expected from the original Big Bang. The effect of the discovery was not only to bring over most cosmologists to the Big Bang theory but also, to a significant extent, it validated cosmology itself as a respectable branch of science. Fresh data concerning the observed cosmic helium abundance (25 per cent by mass) also fitted the Big Bang account of helium formation in the first minutes of the cosmic expansion.

Still, some serious challenges remained. In the Big Bang model, the universe might be (geometrically) ‘open’, in which case it would be infinite and galaxies would continue to expand indefinitely, or it could be ‘closed’, finite, with the galaxies ultimately collapsing again inwards, or ‘flat’, poised between. The decisive factor is the average cosmic mass-density. Collins and Hawking (1973) showed that for a long-lived universe like ours to develop, one in which the background radiation is isotropic, the early universe would have had to be ‘flat’ to an almost unbelievably exact degree. Cosmologists had always assumed that no particular initial cosmic parameter setting would be needed, so this discovery was both unexpected and unwelcome. Collins and Hawking responded with what Carter (1974) dubbed the ‘anthropic principle’: the fact that we have observed the universe to be long-lived and isotropic is merely a consequence of our own existence. Were it not to be of that sort, we would not be here! To convert this ‘principle’ into an explanation, a further premise was necessary. Their choice was a many-universe hypothesis: if our universe is one among a large number of actual universes, we would naturally find ourselves in one that is suited for human existence. Other writers, most of them philosophers of religion, pointed out on the other hand that such ‘fine tuning’ would be what one would expect were the universe to be the work of a provident Creator (see God, arguments for the existence of §5). Here, then, were two unconventional sorts of anthropic *explanation*, each requiring a much broader context than usual to make it acceptable. (It seems doubtful that there is a nontrivial anthropic *principle*, strictly speaking.) The same two alternatives were also proposed in response to a rather different sort of fine tuning. Were the physical constants, notably the relative strengths of the four fundamental forces, to be even very slightly different, a long-lived galactic universe containing the heavy elements needed for complex life would not have developed.

Alan Guth (1981) took a more conventional approach to the ‘flatness’ problem. He proposed a modification of the Big Bang model that would automatically bring about the critical mass-density needed: a gigantic inflation within the first fraction of a second of the cosmic expansion. This would also eliminate several other troublesome anomalies in the original Big Bang theory; it would, for example, achieve the needed causal coordination across different spatial regions of the early universe. The inflationary hypothesis illustrates the extent to which quantum theory has entered into the cosmological debate. On the one hand, standard quantum ideas have permitted physicists to reconstruct the probable sequence of the early expansion, allowing the derivation of testable predictions concerning, for example, the relative cosmic abundance of the lighter elements. On the other, the enormous energies and densities of the very first moments of the (hypothetical) expansion have become a conceptual testing-ground for quantum field theory of a sort that the particle accelerators of earth could never provide.

Cosmology is, as it has always been, a testing-ground for philosophical ideas, lying at the limits of our notions of space, time and causality. Inductivist doubts as to whether a unique object such as the universe could ever become a legitimate object of knowledge have been to some extent quieted. A theory of the universe can be tested as theory normally is, by such criteria as coherence, predictive novelty and consilience. The vast expansion in our conceptual horizons that Big Bang theory and its inflationary addendum have brought about have led philosophers to be more wary than they were of the sort of appeal to intuition on which philosophies of nature in the past were often based. For the same reason, reliance on ‘cosmological principles’ as anything more than idealizations of a regulative sort has become increasingly suspect. Discussions of anthropic forms of explanation have forced scientists and philosophers alike to clarify what counts as ‘explaining’ and what resources can legitimately be drawn on in the process. The many-universe models, particularly those prompted by the inflationary hypothesis, have raised new questions about what should count as a distinct universe. It is less than a century since Einstein

opened the new era in cosmology. No doubt surprises as great as those already encountered still lie ahead.

See also: Anaximander; Cosmology and cosmogony, Indian theories of; Dao; Li; Pythagoreanism; Space; Xing

ERNAN McMULLIN

References and further reading

Ellis, G.F.R. (1989) 'The Expanding Universe: A History of Cosmology from 1917 to 1960', in D. Howard and J. Stachel (eds) *Einstein and the History of General Relativity*, Boston, MA: Birkhäusen, 367-431.(Good summary history. Detailed bibliography; includes publications mentioned by date in the text.)

Harrison, E.R. (1981) *Cosmology: The Science of the Universe*, Cambridge: Cambridge University Press.(A widely praised textbook.)

Hetherington, N.S. (ed.) (1993) *Encyclopedia of Cosmology*, New York: Garland. (An excellent reference work. Special attention is paid to historical and philosophical aspects on cosmology. Several of the scientific essays were written by those who contributed to the fundamental advances described.)

Isham, C.J. (1993) 'Quantum Theories of the Creation of the Universe', in R.J. Russell, N. Murphy and C.J. Isham (eds) *Quantum Cosmology and the Laws of Nature*, Notre Dame, IN: University of Notre Dame Press, 49-89.(Technical but readable account of various attempts to use quantum theory to illuminate the mode of origination of the universe.)

Kragh, H. (1996) *Cosmology and Controversy*, Princeton, NJ: Princeton University Press.(A careful account of the debate between defenders of the steady-state and Big Bang cosmological models. Includes a helpful chapter on the philosophical issues involved.)

Leslie, J. (1989) *Universes*, London: Routledge.(An extensive review of the evidence for 'fine tuning' in recent cosmology and of the various responses to it. Concludes that either God (in the form of an ethical requirement) is real or that there exist many varied universes.)

Leslie, J. (ed.) (1990) *Physical Cosmology and Philosophy*, New York: Macmillan.(A useful anthology containing some classic papers by major cosmologists, where the authors reflect on such philosophical issues as the propriety of anthropic forms of cosmological explanation.)

McMullin, E. (1994) 'Long Ago and Faraway: Cosmology as Extrapolation', in R. Fuller (ed.) *Bang: The Evolving Cosmos*, Lanham, MD: University Press of America, 105-45.(Can there be a proper science of structures and events as remote from us in space and in time as those that recent cosmology purports to establish? What is to count here as 'proper science'? A review of the credentials of the Big Bang theory, concluding positively.)

Munitz, M.K. (ed.) (1957) *Theories of the Universe from Babylonian Myth to Modern Science*, Glencoe, IL: Free Press.(Selection of readings from those who shaped the science of cosmology, ranging from Aristotle and Lucretius, through Newton and Kant, to Lemaître and Gamow.)

Munitz, M.K. (1986) *Cosmic Understanding: Philosophy and Science of the Universe*, Princeton, NJ: Princeton University Press.(Munitz has done more than any other philosopher writing in English to turn the attention of philosophers to the intriguing issues raised by cosmology, both ancient and modern. A readable account of those issues.)

Rhook, G. and Zangari, M. (1994) 'Should We Believe in the Big Bang? A Critique of the Integrity of Modern Cosmology', in D. Hull, M. Forbes and R.M. Burian (eds) *PSA 1994*, East Lansing, MI: Philosophy of Science Association, vol. 1, 228-37. (The authors argue that the weight of anomalies currently facing the Big Bang theory indicates that the confidence it presently enjoys may be misplaced.)

Cosmology and cosmogony, Indian theories of

Theories of the origin of the universe have been told as stories, riddles and instruction in India since early times. The three prominent religious movements, Hinduism, Buddhism and Jainism each had their own myths and speculations.

In the Hindu tradition there was never one single theory. Among the divergent ideas we can distinguish: an early stage, which included themes such as there being nothing at the beginning, or the universe being created by mutual birth, or creation as the dismemberment of a sacrificial victim, or the gods arriving after the first moment of creation; and a later stage, in which Viṣṇu or Brahmā was regarded as the creator of the universe. Simultaneously, the old Sāōkhya idea of the self-creating universe, in which the original material stuff transforms itself into the different parts of the universe, coexisted with the idea of a god creating the universe.

The early Buddhist tradition neglected questions such as 'Does the universe exist?' The first mention of such topics occurred in the Pāli Canon, where they were condemned. A few centuries later, these cosmological ideas were taken up by Vasubandhu, who collected them and formulated them in a comprehensive way. Without a creator god, the universe is primarily a reflection of meditational experiences of the world, a Single Circular System. There are several other systems, such as the Thousand Universe System, the Immeasurable Universe System and the Pure Land.

The Jaina tradition had a very detailed theory of the spatial arrangement of the universe. This was essential for understanding where all the individual selves travel to after death, given their spiritual accomplishments (or lack of them). From earth they go to heavens or hells, the aim being eventually to reach the place of bliss and thus to gain final freedom.

1 Vedic period

Just as the earliest speculations in many cultures revolved around questions and theories of the origin of the universe and related topics, so in India the earliest available document of the culture, the *Ṛg Veda* (*c.*1200 BC), preserves such ideas. It predates any divergence of the ideological movements. The references are scattered. In the oldest parts, cycles 2-9, the topics centre around separating heaven from earth. The creation of the universe is likened to the building of a house, and measuring is one of the main activities. The doors of the house face east and let in the morning light. Other artistic activities, such as weaving or forging, are used to describe the creation of the universe. But interpretations of the myths and speculations on the origin of the universe can be found only later, in the youngest parts of the *Ṛg Veda*, specifically the tenth cycle.

The tenth cycle contains a number of cosmological hymns. Several themes are discernible: in one hymn there is nothing at the beginning, in another there is mutual birth, another describes the dismemberment of a sacrificial victim, yet another the arrival of gods after the first moments of creation. Early cosmological questions such as 'Where did anything come from?' receive some intriguing treatment; hymn 10.129 says that 'At the beginning there was neither existence nor nonexistence'. The questions were often answered in the form of riddles.

Similarly, the Aditi hymn (*Ṛg Veda* 10.72) is ambiguous. First, Reality or Being arises from Nothing. After this, the skies come into being, born from vast spaces. The earth is born of the vast spaces. Often in the *Ṛg Veda* the sky and the earth are considered to be the parents of the various gods. Then Dakṣa is born of Aditi, the mother goddess. In the same hymn, Aditi is the daughter of Dakṣa. It is something like a mutual birth or an incestuous birth. The very same process of birth is found in the well-known 'Puruṣasūkta' (*Ṛg Veda* 10.90.), where Virāj (who is female) is born of Puruṣa, but in turn Puruṣa is born of Virāj. After the mutual birth in the Aditi hymn the gods come into existence and Aditi gives birth to seven sons, whom she takes to the gods. She also gives birth to an eighth son, whom she sets aside to be the human who procreates and dies.

The hymn of dismemberment of the sacrificial victim is called the 'Puruṣasūkta'. It describes the sacrifice of the Primeval Man, through which the world becomes filled with animals as well as all things and populated with people, who are grouped into four categories. Different parts of the Primeval Man are the sources of the social hierarchy and division of labour in Hindu society. Thus from the head of the Man come the priests, from his arms spring the warriors, from his thighs come the farmers and merchants, and from his feet emerge those who provide

services for the other groups. The dismemberment depicts the different strata of society while justifying class divisions that survive to the present day.

2 Upaniṣads and classical Hinduism

The main background for the variety of often unrelated and even contradictory speculations in Indian cosmology was provided at the time of the late Veda, especially during the time of the *Atharvaveda*, the Brāhmaṇas and the Upaniṣads (1000-600 BC).

The Upaniṣads do not form a uniform ideology, but rather represent a collection of speculations on the nature of things. Here we find, for example, a refutation of the early statement about the universe as born from nothing. In Chāndogya Upaniṣad 6, a father instructs his son, telling him that at the beginning there was only one being and from it all was born (see Monism, Indian §4).

In Chāndogya Upaniṣad 6 and other parts of the early Upaniṣads there is a convention of listing. Frequently we find lists of three items: fire, water, solid matter (called food), each of which in turn becomes threefold. The first sets of ontological categories were presented as lists, and this constituted the scientific method of the early period. The number four often plays an important role in lists of the basic elements: heat, water, the solid matter (called 'food') and the wind. Some thinkers added a fifth, space (*ākāśa*). Before the differentiation of separate religious and cultural movements, there was something of an obsession with these lists. This habit continued in the philosophical schools, such as the School of Logic (Nyāya), the School of Categorization (Sāōkhya) and the School of Particulars (Vaiśeṣika).

After the early suppositions of a god creating the universe in separating the earth from heaven, or something nonexistent willing itself into existence, we find that the large segments of population in India becoming Vaiṣṇavite adopted an evolutionary scheme that most identify with Sāōkhya teachings (see Sāōkhya). The Sāōkhya school vacillated between the numbers two and three: there exists only the material stuff and consciousness, yet the material stuff is of two kinds - undifferentiated, when unmanifest, and manifest when created into separate entities - so together with consciousness we get a list of three. The primeval material stuff is undifferentiated until it is disturbed. Once disturbed, it evolves into twenty-three primal entities, from which the world is formed. According to Sāōkhya, in containing these primal entities within itself in the pre-existent form, the cause contains the effect (see Causation, Indian theories of §3). A distinct entity, consciousness (*puruṣa*), brings about the initial disturbance in the primeval material stuff (*prakṛti*). The exact manner in which this disturbance is effected, or what is the relation of consciousness to this material stuff, has never been satisfactorily explained.

Twenty-three primal entities evolve from the primeval material stuff. These are all material, even though the initial ones comprise what we ordinarily consider to be our mental faculties: reason, ego and mind. Consciousness makes these otherwise material entities conscious through its 'proximity'. These first three entities are followed by five sense faculties, five action faculties, five subtle elements and five gross elements. These evolved parts, the material stuff, together with consciousness comprise everything there is.

But what accounts for the differences among entities within the world, some of them light and active and others heavy and dull? The answer is that the primeval material stuff is composed of three elemental forces: 'intelligence', energy and inertia. Different combinations of these forces account for the differences among worldly things. None of the forces is ever found by itself. Each exists only in different combinations with the others, separating and recombining in differing degrees.

3 Later Hinduism

In the later Hindu tradition, the gods become responsible for the process of the creation of the universe, and its destruction in order to create it again. This is one of the major themes of the Purāṇas. The Purāṇas ('Old Stories'), composed from about the fourth century AD onwards, centre around the creation and dissolution of the universe, time periods and genealogies. They are actually the second round of so-called 'old stories', the first one being collected in the epic of the *Mahābhārata*.

In contrast to the diverse early ideas of the universe, its arrangement and timescale, the Purānic ideas reflect a

synthesizing effort to organize the material into some kind of coherence. There is still no single creation myth, but instead we can recognize several major themes, such as the creation scheme involving *puruṣa* and *prakṛti*. This pair is responsible for the creation of whatever exists. Another theme is that of the universe as an egg. A third theme is that of a god, whether Viṣṇu, Puruṣa or Brahmā, often depicted with very human features, who creates this universe.

The interaction between *puruṣa* and *prakṛti* is often interpreted in terms of an interaction between a male and a female. This produces an egg, but it does not break open to give life; rather the universe comes alive within the unbroken egg. It has seven heavens, where the gods and some higher beings spend their allotted time. The egg also contains seven nether worlds, the dwelling place of the ogres (*asuras*) and other fantastic beings. These nether worlds are lavish places with graceful gardens and comfortable houses. Between the heavens and the nether worlds is the earth. India, the ancient name for which was ‘Jambudvīpa’, is in the centre, surrounded by six other lands, with Mount Meru as the *axis mundi*. It seems that this is all the spatial arrangement there is of the physical world. Hells are described vividly in terms of various punishments, but they do not have a physical existence and appear to be moral spaces. Neither the heavens nor the hells are permanent dwelling places. After one’s merit or demerit is exhausted, the next stop may be in the opposite quarter.

The moral aspect of this cosmology is also reflected in the four ages: *kṛta*, *treta*, *dvāpara* and *kali*. These were the terms used for the throws of a popular dice game much enjoyed at the time. The four ages, which constitute one of several theories of time, are described as involving a gradual loss of grace for humankind. The earliest period, *kṛta*, is posited as the ideal state, the past golden age. Things deteriorate in each successive age, until the fourth age, our present, in which we are more or less miserable. A favourite illustration is the cow or the bull of *dharma*, which stands on all four legs in the golden age of *kṛta*. In each successive age, the cow stands on one leg less, becoming less secure as time goes on. In our time, she stands on a single leg and thus has a hard time standing up at all.

There are other theories of time. One is a calendar, based on observations of the moon and sun, that divides time into days, months and years. This calendar has a magnified counterpart in the lifetime of Brahmā. He is the ultimate creator of the universe and his lifetime, a complicated maze of calendars, spans 12,000,000 divine years. Each divine year is 360 human years. So Brahmā’s lifetime is 4,320,000,000 earthly years. At its end, the universe dies too. This is the dissolution of the universe before its new creation, a cycle that repeats itself endlessly and is referred to as the pulsating universe.

4 Buddhism

Originally, questions about the beginning and the functioning of the world were not considered relevant in Buddhism. Whenever he was asked about such topics, Lord Buddha refused to discuss them, referring to them as vain and useless intellectual endeavour. They did not contribute to liberation.

During the next few centuries, these questions gave rise to speculations and theories which form the basis for the Ābhidharmika schools of Buddhism. Vasubandhu (fourth century AD) examines them in his monumental *Abhidharmakośa (Compendium of Abhidharma*, that is, of theories beyond the teaching of the Buddha). In the third chapter, he describes the world, the universe and the levels through which an accomplished spiritual being will pass on the way to *nirvāṇa*.

Vasubandhu believed in a Single Circular System. At the centre of the universe is Mount Meru. Past this are another seven mountain ranges, all of gold and all circular. Beyond these there are four islands. Yet another mountain range encircles all of this. There are bodies of water between each of the mountain ranges. All but the last consist of fresh water; the last one is salt water. Below Jambudvīpa (India) are the sixteen hells. At the summit of Mount Meru are the thirty-three gods.

If we shift to a nonlocational arrangement, there is a threefold division: the sphere of desire, the sphere of form, and the formless sphere. The sphere of form is further subdivided into four meditational levels. In the formless sphere the beings themselves have no form. Liberation takes place beyond the formless sphere. These three spheres are the different levels in meditation that gradually cultivate the aspirant for the final goal of *nirvāṇa*. This is a vision of a physical universe mixed with a universe known through meditation.

There are other systems in addition to the Single Circular System. We can speak basically of two major types of

cosmology: single-world cosmology, and systems based on the coexistence of multiple worlds. These cosmologies are best viewed as conceptualizations of liberation, as reflected in the different schools. In the single-world cosmology, the mechanism for rebirth involves effort, but can be manipulated: it is possible to avoid hell and go instead to heaven by doing meritorious actions. *Nirvāṇa* is liberation from a single-world cosmology, and is often described as being above the formless heavens (of the nonlocational arrangement). In multiple-world cosmologies, liberation is interpreted differently. It is rebirth in a celestial Buddha's Pure Land. This rebirth is the last rebirth, during which it is possible to attain enlightenment just by listening to Buddha's teaching.

5 Jainism

The Jainas took a keen interest in cosmography, the depiction of the organization of the universe, in order that the destinations of individuals according to their behaviour could be explained. The universe depicted was a representation of mental states experienced in meditational practices, rather than of actual or imaginary physical locations. An important source for the study of Jaina cosmography is Umāsvāti's *Tattvārthasūtra*, of around the fourth to fifth centuries AD. Chapters three and four describe the Jaina universe in painstaking detail.

The Jaina cosmology excludes cosmogony because there is no beginning of time, no beginning or creation of the universe. Time is eternal. Frequently the universe is pictured as a woman, with arms bent outward, although sometimes it is depicted as a male figure. Just as in the Hindu cosmography, the earth is in the middle of the universe. The island of the earth is divided into two parts by two mountain ranges extending north to south. There are two Mount Merus, one in the east and one in the west. Below the earth are the hells, which get progressively wider towards the bottom. The heavens, which are above the earth, first become wider (the arms bent outward), then further on they become narrower. Space, as opposed to time, is finite. The inhabited universe (*loka*) is delimited first by thick air, then by thin air, and finally is surrounded by the uninhabited universe (*aloka*), which is empty of anything. Above the heavens, in a circular region, the liberated selves bask in bliss.

This entry has given representative samples of the major trends in Indian cosmological and cosmogonic thought. There is a large body of literature that deals with a number of other things, but contains many ideas that can be considered cosmology. The *Mahābhārata*, the Purāṇas, and the Mānava Dharmaśāstra are almost inexhaustible sources for the Hindu cosmologies and cosmogonies. As compared to the random, organic growth of cosmological ideas in Hinduism, both the Buddhists and the Jainas formulated their cosmological theories after their traditions were well established, and so they reflect a more deliberate ideology.

See also: Buddhist philosophy, Indian; God, Indian conceptions of; Heaven, Indian conceptions of; Hindu philosophy; Jaina philosophy

EDELTRAUD HARZER CLEAR

References and further reading

Gombrich, R. (1975) 'Ancient Indian Cosmology', in C. Blacker and M. Loewe (eds) *Ancient Cosmologies*, London: Allen & Unwin.(Single, short survey covering all the major facets of cosmologies.)

Kloetzli, R. (1983) *Buddhist Cosmology. From Single World System to Pure Land: Science and Theology in the Images of Motion and Light*, Delhi: Motilal Banarsidass.(A good, comprehensive introduction.)

Ṛg Veda (*c.*1200 BC), trans. W.H. Maurer, *Pinnacles of India's Past: Selections from the Ṛg Veda*, University of Pennsylvania Studies in South Asia 2, Amsterdam and Philadelphia, PA: John Benjamin, 1986.(A literary and precise translation.)

Schubring, W. (1962) *The Doctrine of the Jainas Described after the Old Sources*, trans. W. Beurlen, Delhi: Motilal Banarsidass.(Contains a thorough discussion of Jaina cosmology and cosmography.)

Umāsvāti (4th-5th century) *Tattvārthasūtra*, ed. and trans. N. Tatia, *Tattvārtha Sūtra: That Which Is*, San Francisco, CA: HarperCollins, 1994.(An authoritative edition, containing text in transcription and translation, of a classic Jaina manual.)

Upaniṣads (1000-600 BC), trans. P. Olivelle, *Upaniṣads*, World's Classics Series, Oxford: Oxford University Press, 1996.(The latest, most up-to-date translation of twelve Upaniṣads, with excellent annotations which resolve many hitherto unintelligible passages.)

Counterfactual conditionals

'If bats were deaf, they would hunt during the day.' What you have just read is called a 'counterfactual' conditional; it is an 'If…then…' statement the components of which are 'counter to fact', in this case counter to the fact that bats hear well and sleep during the day. Among the analyses proposed for such statements, two have been especially prominent. According to the first, a counterfactual asserts that there is a sound argument from the antecedent ('bats are deaf') to the consequent ('bats hunt during the day'). The argument uses certain implicit background conditions and laws of nature as additional premises. A variant of this analysis says that a counterfactual is itself a condensed version of such an argument. The analysis is called 'metalinguistic' because of its reference to linguistic items such as premises and arguments. The second analysis refers instead to possible worlds. (One may think of possible worlds as ways things might have gone.) This analysis says that the example is true just in case bats hunt during the day in the closest possible world(s) where they are deaf

1 Classification issues and familiar logical connectives

'Counterfactual' conditionals serve to summarize hypothetical reasoning and to justify or explain action and inaction ('I didn't tell you because, had I told you, everyone in the office would have known'). They serve to state particular causal contingencies ('If the battery were flat, the starter would not turn') and various kinds of necessity ('If he were dead, she would be a widow'). Counterfactuals have also been enlisted for philosophical analyses of dispositions ('*X* is soluble just in case it would dissolve if it were immersed in water'), causality ('Had *A* not occurred, *B* would not have occurred'), freedom of action ('I would have acted otherwise if I had so chosen'), knowledge ('If *p* were false, Mary would not believe that *p*') and laws of nature, which have been characterized as generalizations that are not merely true but support counterfactuals.

Counterfactuals are commonly contrasted with indicative conditionals (see Indicative conditionals). The following pair conveys a feeling for the difference:

(1) If Shakespeare hadn't written *Hamlet*, then someone else would have (counterfactual).
(2) If Shakespeare didn't write *Hamlet*, then someone else did (indicative).

Instead of 'counterfactual conditional', sometimes the term 'subjunctive conditional' is used. The former refers to a property of the statement made by uttering a conditional sentence in an appropriate context, the latter refers to the grammatical form of the sentence uttered. Neither label is satisfactory. A problem with 'counterfactual' is that not all assertions of counterfactuals are, or are meant to be, counter to fact, as, for example, 'If you did the cooking, I would clean up afterwards' (put forward as a suggestion before the cooking has started). 'Subjunctive' is problematic because the grammatical form in question is not the subjunctive in English (where this form is all but extinct) nor in other European languages (where it is widely used). Moreover, 'subjunctive' contrasts with 'indicative', yet indicatives such as (3) below go much more naturally with 'subjunctives' such as (1) than with indicatives such as (2):

(3) If Shakespeare does not write *Hamlet*, then someone else will.

It is surprisingly difficult to characterize the difference between the two kinds of conditional. It is not even agreed at which linguistic level - morphosyntactic, semantic or pragmatic - the distinction is to be drawn. We will content ourselves here with defining counterfactual *sentences* as the likes of (1) above. Our question is then which *statements* these sentences normally serve to express.

In the view of most authors, an analysis must state non-circular and informative truth-conditions. It should also be 'compositional': the truth-conditions of '$A \rightarrow B$' should be a function of the truth-conditions of *A* and *B* (see Compositionality). (Read '$A \rightarrow B$' as 'If it were the case that *A*, then it would be the case that *B*'.) The material conditional of classical logic and the strict conditional of modal logic both meet these demands, but their truth-conditions are unsuitable. A material conditional is true just in case either its antecedent is false or its consequent true. Since most counterfactuals have false antecedents, they would come out true irrespective of their consequents if they were treated as material conditionals. But many are not true (for example, (1) above); hence the material conditional cannot be the right analysis. A strict conditional is true just in case the corresponding

material conditional is necessarily true, or, in possible worlds parlance, just in case in all worlds in which the antecedent is true the consequent is also true. If the material conditional is too easy to satisfy, the strict conditional is too hard. Consider

(4) If fewer people smoked, fewer would contract lung cancer,

which is true if any counterfactual is. Yet there are possible worlds where the antecedent is true and the consequent false, for example, worlds in which lung cancer is caused by eating chocolate. So the strict conditional is too strict. We might confine our attention to less exotic worlds; ones governed by the same natural laws as the actual world. But that still would not do because among the lawful worlds where fewer people smoke there are worlds where more people work in coal mines and the overall incidence of lung cancer is the same. The worlds in which we are interested must resemble the actual world more closely. We imagine that fewer people smoke in these worlds, but that otherwise things are very much as in the actual world. It is this idea of a minimal departure from actuality (or purported actuality) that underlies, in one form or another, all theories of counterfactuals.

2 The metalinguistic view

Nelson Goodman (1954) put forward an influential view in which ‘ $A \rightarrow B$’ is true if and only if (1) there is an appropriate subset of all the true initial conditions that, in conjunction with *A* and the laws of nature, entails *B*; and (2) there is no other subset of the initial conditions that in the same manner entails not-*B*. The minimal departure from actuality lies in the appropriate choice of initial conditions. Two problems beset Goodman’s proposal: the characterization of the relevant initial conditions and the definition of laws. The second problem is serious, but not as serious as the first.

Apart from being true, the relevant initial conditions must be jointly compatible with *A* (as long as *A* is consistent), for otherwise any consequent whatever could be inferred and the corresponding counterfactual would be true. The conditions must also be compatible with not-*B*, for otherwise *A* and the laws would play no role in the inference to *B* - *B* would follow from the initial conditions by themselves. (‘Compatible’ statements are ones which do not jointly entail a contradiction.) But compatibility alone is not a strong enough condition. Consider

(5) If match *m* had been struck, it would have lighted.

Let $O =$ ‘oxygen is present’, $D =$ ‘*m* is dry’, $S =$ ‘*m* is struck’, $L =$ ‘*m* lights’ and $N =$ ‘If *O*, *D* and *S*, then *L*’ (our natural law). Suppose that we start with the situation of *O*, *D* and *N* being true and that the match has not been struck and has not lighted. *O*, *D* and *N* are jointly compatible with *S* and jointly compatible with not-*L*, and together with *S* they imply *L*. Therefore (5) should be true.

However, not-*L* is also true, not-*L*, *O* and *N* are jointly compatible with *S* and jointly compatible with *D*; and *O*, *N* and not-*L* together with *S* imply not-*D*. Thus

(6) If match *m* had been struck, it would not have been dry

should be true as well. This is unacceptable. Hence logical compatibility is not enough. The relevant conditions must be jointly tenable or ‘cotenable’ with the antecedent. (*B* is said to be cotenable with *A* if it is not the case that *B* would be false if *A* were true.) In this case, not lighting is compatible with, but not cotenable with, striking, because, had *m* been struck, it would have lighted; so (5) is true and (6) is not. But now we are caught in a circle or regress because cotenability is defined in terms of counterfactuals, while counterfactuals are defined in terms of cotenability. Goodman saw no way out of this quandary.

3 Possible worlds semantics

In the 1960s, modal logicians turned talk of possible worlds into a powerful analytic tool. Robert Stalnaker and David Lewis applied it to the analysis of counterfactuals. Both saw counterfactuals as devices to locate the actual world within a certain similarity structure in logical space. The intuition they built on is that ‘ $A \rightarrow B$’ is true in the actual world just in case *B* is true in the most similar *A*-world(s) (an *A*-world being a world where *A* is true). Stalnaker (1968) represents the minimal departure from actuality by a selection function on possible worlds, Lewis (1973) by ordering worlds so that given any two worlds *u* and *v*, either *u* is more similar to the actual world than *v*, or vice versa, or else *u* and *v* are equally similar to the actual world. For Stalnaker, ‘ $A \rightarrow B$’ is true if and only if

B is true at $f(A, w^*)$, the world picked out by applying the selection function *f* to antecedent *A* and the actual world w^*. *f* selects a world in which *A* is true and which, intuitively, resembles w^* as closely as the truth of *A* permits. For Lewis, '$A \to B$' is true if and only if either there is no *A*-world, or some *A* & *B*-world is more similar to the actual world than any *A* & not-*B*-world. Lewis' analysis makes '$A \to B$' trivially true when *A* is impossible, which is when there is no *A*-world. So does Stalnaker's analysis because, for impossible *A*, *f* selects 'the impossible world' in which every statement is true.

An argument form is valid if and only if it has no instance with true premise(s) and a false conclusion. The following is an example of an invalid argument:

If Ronald Reagan had been born a Russian, he would have been a communist.
If he had been a communist, he would have been a traitor.
Therefore, if he had been born a Russian, he would have been a traitor.

The premises seem true, but not the conclusion. The argument is an instance of 'hypothetical syllogism':

$A \to B, \quad B \to C \quad \therefore\ A \to C$

The possible worlds theory explains how hypothetical syllogism can fail to be valid. World *c*, the closest world where Reagan was a communist, is closer to the actual world than world *r*, the closest world where he was born a Russian. In *c*, his career is as we know it, except for that traitorous affiliation. In *r*, he grows up in Russia, becomes a loyal party member there and never dreams of being president of the USA. So *c* makes the second premise of the argument true, while *r* makes the first premise true and the conclusion false.

On the Lewis-Stalnaker analyses, a number of argument forms that are valid for the material and various strict conditionals are invalid. Hypothetical syllogism is only one of them. Others are:

$A \to B \therefore\ \text{not-}B \to \text{not-}A$ ('contraposition')
$A \to B \therefore\ C \to B$, where C entails A ('strengthening the antecedent')
$A \to (B \to C) \therefore\ (A\ \&\ B) \to C$ and
$(A\ \&\ B) \to C \therefore\ A \to (B \to C)$ ('import-export')

For practical purposes, the invalid argument forms can often be replaced by similar valid ones, for example, hypothetical syllogism by

$A \to B, \quad (A\ \&\ B) \to C \quad \therefore A \to C.$

The distinction between valid and invalid argument forms ought to match natural language usage. Examples like the one about Reagan therefore carry heavy weight. If they were rejected and any of the argument forms in question added to the logics that emerge from the Lewis-Stalnaker analyses, '$\to$' would collapse into the material conditional. Attacks on the counterexamples are therefore indirect attacks on the core of the analyses. (For more discussion, see Stalnaker 1984.)

4 Similarity

Lewis allows ties in the relative similarity between worlds, as well as no limit to how closely worlds can resemble each other. Stalnaker's theory can be seen as a special case of Lewis' that does not make these allowances. For Stalnaker, there is to be always exactly one most similar antecedent-world. The difference is reflected in a disagreement about the principle of 'conditional excluded middle' which says that '$A \to B$' and '$A \to \text{not-}B$' cannot both be false. The principle holds in Stalnaker's view but not in Lewis'. In Lewis' view, there may be, for any *A* & *B*-world, an equidistant or closer *A* & not-*B*-world and vice versa, so that both counterfactuals can be false. For Stalnaker, there is exactly one closest *A*-world in which either *B* or not-*B* is true. Consider the pair

If Bizet and Verdi had been compatriots, Bizet would have been Italian.
If Bizet and Verdi had been compatriots, Verdi would have been French.

We hesitate to affirm either. Doesn't this show that Lewis is right? Not necessarily. We also hesitate to deny either of the two, and the best explanation may be that the relevant standard of similarity between worlds has not been

made clear, so that we do not know exactly what claims are at issue. Conditional excluded middle holds relative to a given selection function *f*, but it does not apply to cases where *f* has not been determined. Any theory of counterfactuals must allow for a good measure of indeterminacy if it is to account for actual usage.

Goodman's theory can be transposed into the possible worlds framework. With the right definition of cotenability, it becomes equivalent to Lewis'. Does Lewis therefore solve Goodman's problem about cotenability? In some sense, yes, because there is no longer a circle in the truth-conditional schema. In another sense, no, because Goodman's worry was less about the abstract form of the analysis than about how particular cotenability judgments can be grounded. For an answer to this question, the notion of relative similarity between worlds is probably too vague.

The relevant notion of similarity cannot be the one we use when we compare ordinary objects. For otherwise we should reject

If Nixon had pushed the button, there would have been a nuclear holocaust,

because among the worlds where Nixon pushes the ominous button those where nothing happens are more similar to ours, in the everyday sense of similarity, than those where the doomsday machinery goes off. But given that there is (or was) such a button, the counterfactual looks true. Lewis suggests that the similarity relation relevant for counterfactuals has the most similar worlds agree with the actual world throughout the past until shortly before the time of the antecedent. Then a small miracle (a deviation from the laws of the actual world, not from those of the world in question) brings about the antecedent, and from then on history evolves in accordance with the laws of the actual world. The proposal has limited scope. It does not apply to 'counterlegals' about what would be the case if this or that law of nature were different. It also does not deal with disjunctive antecedents pertaining to different times. Maximizing agreement throughout the past would naturally lead here to a selection of worlds in which only the disjunct pertaining to the latest time holds. Other authors favour total agreement in laws over miracles of any size. Miracles prevent a counterfactual extrapolation of what the past would have had to have been in order to bring about the antecedent. Yet such 'backtracking' is perfectly intelligible. Jim and Jack quarrelled yesterday and Jack is still furious. If Jim asked a favour of Jack today, Jack would oblige him none the less. For in order that Jim ask, there would have had to have been no quarrel before. The significance of backtracking and Lewis' proposal about similarity remain contested.

5 Comparisons

Metaphysical and other qualms about the possible worlds analyses rekindled an interest in metalinguistic theories in the 1980s. There have also been proposals for an epistemic analysis based on the idea that '$A \rightarrow B$' is 'acceptable' with respect to a body of beliefs *K* just in case *K*, minimally changed so as to accommodate *A*, entails *B*. Formally, most of these theories are equivalent to Lewis' or slight variants thereof, but they avoid the ontological cost of possible worlds. In so far as they do not provide truth-conditions, the epistemic theories have difficulty explaining the role of counterfactuals embedded in truth-conditional constructions such as conjunctions or disjunctions. Moreover, the idea of revising a body of beliefs seems better suited for evaluating indicative conditionals than counterfactuals. For example, if you add to your present beliefs the assumption that Shakespeare did not write *Hamlet* and make the minimal changes necessary to restore consistency, you will infer that someone else must have written the piece because you have no reason to give up your belief in its existence. This yields the correct judgment for the indicative (2) but not for the counterfactual (1). It is not obvious how to adapt procedures of belief change to counterfactuals.

Goodman and many others believed that an analysis of counterfactuals was indispensable for the philosophy of science. Few still share this opinion. The cotenability problem, the indeterminacy of many counterfactuals, and doubts about possible worlds and their similarities have convinced many that counterfactuals do not belong in the ultimate scientific description of the world. In decision and game theory, by contrast, the peculiarity and importance of counterfactual reasoning has only just begun to be appreciated. The solution of decision problems requires answers to questions about situations that, as a result of the very decision at issue, will be counterfactual. Counterfactuals also continue to play an important role in philosophical analyses, and their use and misuse in daily commerce is largely unhampered by difficulties of interpretation.

See also: Logical and mathematical terms, glossary of; Possible worlds; Relevance logic and entailment

FRANK DÖRING

References and further reading

Edgington, D. (1995) 'On Conditionals', *Mind* 104.414: 235-329.(Gives an excellent overview and synthesis of the various issues surrounding conditionals.)

Goodman, N. (1954) *Fact, Fiction, and Forecast*, Atlantic Highlands, NJ: Athlone Press.(See §3. A non-technical classic.)

Harper, W.L., Stalnaker, R. and Pearce, G. (eds) (1981) *IFS: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht and Boston, MA: Reidel.(A collection of important papers of varying degrees of technicality.)

Jackson, F. (ed.) (1991) *Conditionals: Oxford Readings in Philosophy*, Oxford: Oxford University Press.(A collection of important papers, including Lewis' on the similarity between worlds.)

Lewis, D.K. (1973) *Counterfactuals*, Cambridge, MA: Harvard University Press. (See §4. The philosophical discussion demands no technical expertise, but the details of the semantics and the logic get very technical.)

Sanford, D.H. (1989) *If P then Q: Conditionals and the Foundations of Reasoning*, London and New York: Routledge.(An informative historical overview.)

Stalnaker, R. (1968) 'A Theory of Conditionals',*Studies in Logical Theory: American Philosophical Quarterly Monograph Series* 2: 98-112; repr. in W.L. Harper, R. Stalnaker and G. Pearce (eds) *IFS: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht and Boston, MA: Reidel, 1981, 41-56.(See §4. The first statement of Stalnaker's position; no technical expertise required.)

Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: Bradford Books.(An excellent non-technical discussion of counterfactual and indicative conditionals and their role in reasoning.)

Cournot, Antoine Augustin (1801-77)

Cournot is best known for his work in applying mathematical techniques to economic and social affairs and is generally acknowledged to be the founder of econometrics. His work in philosophy, however, is at least as distinguished. His philosophy may be seen as a meditation on continuity and discontinuity, on law and brute empirical fact, which is unassimilable to law. Empiricism is exclusively preoccupied with the latter phenomenon, rationalism with the former. Cournot affirmed the reality of both. His philosophy thus mediates between empiricism (which, when it is consistent, leads on his view to scepticism) and rationalism (which, when it is consistent, loses contact with reality).

Continuity is real because the world is not a chaos; it is a network of events, forming various series, which reveal necessary and determinable relations. But discontinuity is also real, for the order we discern in the various events is not a single order. The series are independent and the points where they intersect cannot be predicted from within the series themselves. Brute contingency is therefore as real as law. Consequently, though the world may be known, it can never be reduced to a single scheme.

Antoine Augustin Cournot was born in the town of Gray, near Dijon. He was raised in his grandmother's house, being especially influenced by an uncle, who was a Jesuit, and by a maiden aunt, whose views were directly opposed to those of the uncle. At an early age, he learned to balance different views, to form his own conclusions, and to keep his peace. He was educated at Gray and at the École Normale in Paris, specializing in mathematics and science. He spent his working life in the university, holding chairs at Lyons and Grenoble. In 1838, he was named Inspector General and in 1854 was appointed as Rector of the Academie at Dijon. Involved in teaching and administration, he continued to work on fundamental problems in many fields. In 1851, he published his most important philosophical work, *Essai sur les fondements de nos connaissances (An Essay on the Foundations of Our Knowledge)*. Failing eyesight led to his retirement in 1862. But he continued to work at philosophy and in his last years published another important work, *Matérialisme, Vitalisme, Rationalisme (Materialism, Vitalism and Rationalism)*.

Like Kant, Cournot held that knowledge is relative. We know the world not in its ultimate nature but in its relations to ourselves. Cournot, however, rejected what he took to be Kant's subjective idealism, denying that the order we discern in nature is the product of our own minds. The central problem in philosophy is precisely to clarify how the mind makes contact with reality through its grasp of order. Empiricism is inadequate for this purpose, for the senses serve neither as the basis nor as the criterion for knowledge. A simple example will illustrate the point. Consider a law, such as Boyle's law, which is consistent with numerous observations. The observations, however numerous, are a random selection from those that might be made. They represent an element of chance. We accept Boyle's law because we cannot believe that under those chance conditions the law would always be confirmed, unless it represented an order in the gases which is independent of the conditions themselves. It is not the observations which explain the law; it is the law which provides the best explanation for the observations. The point is evident on other grounds. First, observations, being discontinuous, cannot correspond to law, which is continuous; a law holds, whatever the observations. Second, with sufficient ingenuity one can think of many different laws which are consistent with those same observations. Science, according to Cournot, depends on a feeling for order which cannot be quantified and which is akin to the feeling for beauty.

Law in itself, however, is not sufficient for knowledge. The laws of planetary motion will enable us to fix the positions of the planets at any time in the future or the past. But suppose, in the remote past, a comet entered the planetary system, disturbed its arrangement and then disappeared. No trace of this event would appear in our knowledge of astronomy. Consequently, in fixing the positions of the planets, prior to its occurrence, we should be led into systematic error. It is impossible to believe that such events have never occurred. Much in history, therefore, is likely to be not simply unknown but unknowable. In short, scientific rationalism is as inadequate as empiricism.

Cournot anticipated Darwin's theory of evolution (see Darwin, C. §§2-3). He discussed what is virtually the same theory, in his *Essay*, some eight years before Darwin's work appeared, arguing that it could not provide a comprehensive explanation of the evolutionary process. He agreed that the environment has an important role in

selecting organic forms for survival. What the theory cannot explain, however, are the organic forms themselves, which cannot evolve through an accumulation of chance variations. Each organic form is an order, involving an interrelation of its parts. The forms therefore are discontinuous; to get from one to another through continuous variation one would have to pass through stages which do not constitute possible organic forms. Thus a wing which is slightly too weak for flight is entirely useless as a wing. The evolutionary process can be understood only as a phenomenon involving both continuity and discontinuity, order and chance. The pattern of a leaf is definite, so far as its principle veins are concerned, but, in its ultimate ramifications, is so modified by contingent forces that no individual leaf is exactly the same as another. Similarly, the world itself is governed by general laws the detailed fulfilment of which is left to contingent circumstances. Consequently there is no conflict between contingency and finality or purpose. It is through contingency that finality or purpose works. Thus it is not in order to serve the needs of the bee that the flower exists. The flower exists for its own sake, exhibits finality in itself, and simultaneously serves the needs of the bee. For Cournot, the world is a composition, overwhelming in its beauty and complexity, which is executed simultaneously on different planes.

See also: Empiricism; Evolution, theory of; Rationalism

H.O. MOUNCE

List of works

Cournot, A.A. (1838) *Recherches sur les principes mathématique de la théorie des richesses*, Paris; trans. N.I. Bacon, *Researches into the Mathematical Principles of the Theory of Wealth*, London: Economic Classics series, 1938.(His most influential work.)

Cournot, A.A. (1861) *Essai sur les fondements de nos connaissances et sur les caracteres de la critique philosophique*, Paris, 2 vols; trans. M.H. Moore, *An Essay on the Foundations of Our Knowledge*, New York: The Liberal Arts Press, 1956.(His main work in philosophy.)

Cournot, A.A. (1922) *Traité de l'enchaînement des idées fondamental dans les sciences et dans l'histoire (Treatise on the Structure of the Fundamental Ideas of Science and History)*, Paris, 2 vols.(This edition contains a foreword by Lévy-Bruhl.)

Cournot, A.A. (1934) *Matérialisme, vitalisme, rationalisme: Études sur l'emploi des données de la science en philosophie (Materialism, Vitalism, Rationalism: Studies in the Use of Scientific Assumptions in Philosophy)*, Paris.(A work of comparable importance to the *Essai*.)

References and further reading

Moore, M.H. (1934) 'The Place of A.A. Cournot in the History of Philosophy', *Philosophical Review* 43: 380-401.(The best introduction to Cournot in English.)

Poincaré, H. *et al.* (1905) *Revue de metaphysique et de morale* 13: 3.(An edition devoted entirely to Cournot's work. Contains contributions by leading French thinkers.)

Ruddick, C.T. (1940) 'Cournot's Doctrine of Philosophical Probability', *Philosophical Review* 49: 415-23. (Another competent survey.)

Cousin, Victor (1792-1867)

French philosopher, educationalist and historian, Victor Cousin is primarily associated with 'Eclecticism' and the history of philosophy, but his work also includes contributions to aesthetics, philosophy of history and political theory. He was a prolific writer and editor, and a significant figure in the development of philosophy as a professional discipline in France.

Victor Cousin was born in Paris, the son of an artisan. His education was funded by a wealthy sponsor, and culminated in the prix d'honneur and study at the École normale, where his formative influences included Laromiguière, Royer-Collard and Maine de Biran. His subsequent academic career at the École normale and the Sorbonne progressed rapidly, his lectures attracting large audiences and much publicity. In the early 1820s Cousin became a victim of the reaction against liberalism; his university lecturers were suspended and the École normale closed. Employed as a private tutor, Cousin worked on the history of philosophy and studied contemporary German idealism. On a visit to Prussia in 1824, French police reports and his association with members of the Carbonari led to his imprisonment. Hegel, with whom Cousin had established a relatively warm association, intervened on his behalf with the authorities, and he was subsequently released. Restored to his University post in 1828, Cousin's lectures received great acclaim and, especially under the July Monarchy, public recognition followed on a grand scale. He became a member of the Académie Française and of the Académie des Sciences Morales et Politiques, a peer of France, Commander of the Légion d'Honneur, member of the Conseil Supérieur de l'Instruction Publique and, briefly, in Theirs' second cabinet, minister for public instruction. From 1830, Cousin increasingly devoted his energies to administrative and educational affairs. Using his position as director of the École normale, he presided over the development of philosophy as a professional (and non-theological) discipline in France. Cousin also influenced the reform of French primary education, publishing reports on the school systems of Prussia, Saxony and Holland. Following the 'tragic experience' of 1848, he fell out of favour, and with the coup d'état of 2 December 1851, was increasingly removed from public life. His last years were spent quietly, surrounded by his books in rooms in the Sorbonne, writing a series of historical studies, mainly portraits of women notables of the seventeenth century. He died in Cannes.

Cousin was probably the best known exponent of Eclecticism, maintaining that schools of philosophical thought (which, for him, fell into four categories) all contained elements of the truth that, once recovered, could be reconciled. Sensationalism, idealism, scepticism and mysticism, he averred, were 'not false but incomplete'. Cousin denied the charge of facile syncretism, insisting that Eclecticism was grounded in a tripartite model of human psychology revealed by observation and experiment, human nature being composed of the three distinct but complementary faculties of sensibility, will and reason. Various philosophical schools were chided for their neglect of one or more of these capacities; thus Cousin charged the epistemology of Condillac (and of his materialist successors) with mistakenly reducing human experience to the passive faculty of sensation.

For Cousin, the reason revealed by psychological introspection was universal and necessary, and by virtue of this impersonal character - and in a manner not always clear to his critics - it provided the conditions for both the existence of, and our knowledge of, humanity and nature. His apparent assertion of the immanentism of reason led to charges of pantheism or even Neoplatonism (see Pantheism; Neoplatonism). Clerical critics were scarcely reassured when Cousin insisted that, since the masses could not attain philosophical understanding, his system preserved a place for religion as a socially necessary and seemingly inferior (because symbolic) parallel of Eclectic metaphysics.

Art (like religion and philosophy) was a mode in which the infinite, or God, became apparent. It formed an independent sphere in which human nature provided the criteria by which to judge an ideal beauty. For Cousin, the aesthetic hierarchy that resulted was headed by imaginative poetry.

Cousin also discerned a 'manifestation of God' in the historical development of epochs and cultures. History followed a tripartite spiritual progression, as philosophical concerns turned from the infinite, to the finite, to the relation between the two. The resulting Oriental, Greek and Modern worlds were characterized successively by pantheism (and monarchy), polytheism (and democracy), and theism (and constitutionalism). Cousin described this philosophy of history as both speculative and empirical, although critics found it easier to discern evidence of the

former than the latter quality. Furthermore, here as elsewhere in Cousin's work, the distinction between his own opinions and his reproduction of certain themes in contemporary German philosophy is not always apparent.

In rejecting all exclusive doctrines, Eclecticism claimed to have a political parallel in the repudiation of political doctrines and practice based on faction. Cousin was a moderate liberal, committed to constitutional monarchy, the Charter and the interests of the 'juste milieu'. He distinguished a sphere of enforceable individual rights to equal respect and private property, and a subordinate sphere of voluntary duties to protect and promote the wellbeing of others. Cousin attributed a positive role to the state, not least in the provision of moral and educational instruction.

Claiming to recover and reconcile the truth embedded in previous schools of thought, Eclecticism encouraged work in the history of philosophy. Cousin's own efforts in this area were prodigious. His editorial labours included a six-volume edition of Proclus (1820-7); an eleven-volume edition of Descartes (1826); and single-volume editions of Abélard and Maine de Biran. Cousin was also responsible for a thirteen-volume translation of Plato (1822-40). His interpretative studies included works on Aristotle, Locke, the Scottish Enlightenment and Kant, together with an inflammatory contribution to *l'affaire Pascal* (an acrimonious dispute about editorial and interpretative standards in contemporary Pascal scholarship).

The contrast between Cousin's intellectual hegemony in nineteenth-century France and the subsequent neglect of his work is striking. Although Cousin's historical significance is undeniable, claims for the intrinsic philosophical interest of his work remain muted. Eclecticism strikes many as empty or flawed, while Cousin's role as an interpreter and popularizer of post-Kantian German philosophy largely rests on the unsystematic, and often unacknowledged, incorporation of elements of Schelling and Hegel into his own work. Even Cousin's endeavours in the history of philosophy, although not without interest, and unquestionably important in opening up French philosophy to new influences, frequently fall short of modern scholarly standards.

See also: Hegelianism §3; German idealism

DAVID LEOPOLD

List of works

Cousin, V. (1828) *Cours de l'histoire de la philosophie moderne*, Paris: Pichon & Didier, 3 vols; trans. O.W. Wright as *Course of the History of Modern Philosophy*, New York: D. Appleton, 1852, 2 vols.(Published version of Cousin's lectures, and a nineteenth-century translation of a variant of the text.)

Cousin, V. (1833) *De l'instruction publique dans quelques pays de l'Allemagne et particulièrement en Prusse*, Paris: Pitois-Levrault, 3rd edn, 1840; trans S. Austin as *Report on the State of Public Instruction in Prussia*, London: E. Wilson, 1834. (Cousin's influential account of German schooling, and an English translation of the first edition.)

Cousin, V. (1838) *Fragmens philosophiques (Philosophical Fragments)*, Paris: Ladrange, 3rd edn, 2 vols.(A collection of Cousin's early writings.)

Cousin, V. (1848) *Justice et charité*, Paris: Pagnerre; trans. W. Hazlitt as *Justice and Charity*, London: Samson Low, 1858.(A short pamphlet outlining the different demands of justice and charity, and a nineteenth-century translation.)

Cousin, V. (1851) *Discours politiques avec une introduction sur les principes de la Révolution française et du gouvernment representatif (Political speeches with an introduction to the principles of the French Revolution and of representative government)*, Paris: Didier.(Cousin's interpretation of recent French history and his defence of constitutional monarchy.)

Cousin, V. (1853) *Du vrai, du beau, du bien*, Paris: Didier; trans. O.W. Wright as *The True, the Beautiful and the Good*, Edinburgh: T. & T. Clark, 1854.(Shortened version of Cousin's lectures and a nineteenth-century translation of the third edition.)

Cousin, V. (1866) *Fragments philosophiques pour servir à l'histoire de la philosophy (Philosophical fragments to serve as a history of philosophy)*, Paris: Slatkine, 1970. (Reprint of a five-volume edition of Cousin's lectures.)

References and further reading

Boas, G. (1925) *French Philosophies of the Romantic Period*, Baltimore, MD: Johns Hopkins University Press. (Includes a chapter on Eclecticism.)

Kelly, G.A. (1992) *The Humane Comedy*, Cambridge: Cambridge University Press. (A thoughtful study of French

liberalism which includes material on Cousin.)
Ody, H.J. (1953) *Victor Cousin. Ein Legensbild im deutsch-französischen Kulturraum (Victor Cousin. A study in Franco-German culture)*, Saarbrucken: West-Ost-Verlag.(A modern German study of Cousin.)
Simon J. (1887) *Victor Cousin*, trans. G. Masson, London: George Routledge & Sons, 1888.(A general work by a student and contemporary of Cousin.)
Spitzer, A.B. (1987) *The French Generation of 1820*, Princeton, NJ: Princeton University Press.(An interesting historical account of Cousin's age cohort.)
Will, F. (1965) *Flumen Historicum: Victor Cousin's Aesthetic and its Sources*, Chapel Hill, NC: University of North Carolina Press.(A study of the Platonic, neoclassical and German Idealist sources of Cousin's aesthetic.)

Crathorn, William (*fl. c.*1330)

An English scholastic a generation younger than William of Ockham, Crathorn's theological writings confront the central metaphysical and epistemological problems of his day. He is of interest largely because of his willingness to pursue the logical implications of his views to the most extreme conclusions. This characteristic makes Crathorn a provocative and idiosyncratic thinker, although not always a coherent one.

Crathorn was a Dominican friar from northern England who lectured at Oxford on Peter Lombard's *Sentences* during the academic years 1330-2 (see Lombard, P.). These lectures, his *Quaestiones in primum librum Sententiarum (Questions on the First Book of the Sentences)*, are our best guide to Crathorn's philosophy. In them he discusses the most controversial philosophical problems of his day, such as divine foreknowledge and human freedom, the ontological status of the Aristotelian categories and the distinction between intuitive and abstractive cognition. His philosophical agenda owes much to William of Ockham, but he is not in any sense Ockham's disciple; at almost every juncture Crathorn explicitly rejects Ockham's views.

Crathorn's theory of universals, developed in his *Quaestiones*, is central to his thought. He begins by rejecting competing views, including those of Thomas Aquinas and John Duns Scotus as well as Ockham's early *fictum* account. Crathorn's treatment of Aquinas, a fellow Dominican, is particularly interesting. After faithfully describing what Aquinas said about universals, Crathorn denies that this was what Aquinas actually meant. Perhaps, Crathorn suggests, Aquinas advanced such a view 'by way of conforming himself to the way men of his time spoke'. Evidently, any pressure that Crathorn may have felt to adhere to the views of his great *confrère* was not enough to force a substantive change in Crathorn's own thought.

At first glance, Crathorn's own theory of universals appears to be a rather ordinary nominalism. A universal, he says, is anything that either causes, resembles or is conventionally predicated of more than one thing. Only words and concepts can be universals in the ordinary sense of the term. Words can be universal in the third way: by convention, they are predicated of more than one thing. Concepts can be universals in either the second or the third way.

It is here that the peculiarity of Crathorn's position emerges. How can concepts either resemble or be conventionally predicated of external things? The first of these may seem less peculiar, since most scholastics spoke of mental representation in terms of similarity. However, Crathorn is unique in holding this doctrine in a completely literal way. For him, the concept of whiteness really is white. He in effect asks how anything could resemble what is white otherwise than by being white (*Quaestiones* q.1, concl.7). But Crathorn does not hold that all concepts are likenesses of this sort; he notices, for instance, that there is nothing in the external world for the general concept of colour to resemble (*Quaestiones* q.11, concl.4). Here Crathorn faces a problem. Whiteness, like colour, is a determinable property, so there is no one shade of white in the external world for the concept 'whiteness' to resemble. In fact, however, Crathorn does find something in the external world for non-determinate concepts like 'colour' to resemble: they resemble (again, literally) the written or spoken words used to express that concept. Thus our concept of colour is literally a likeness of the word 'colour'. Crathorn says that this concept is conventionally predicated of colours inasmuch as it resembles the word 'colour', which is itself conventionally predicated.

Crathorn explicitly holds that these concepts are the immediate objects of intellective cognition. Likewise, sensory impressions are what we immediately perceive, and they, too, literally resemble features of the external world. This leads him to confront the notorious sceptical problem of how we know anything about the external world if all we directly perceive is internal impressions (see Scepticism). At the outset of his commentary, Crathorn develops on this basis a series of ever more serious sceptical conclusions, culminating in the claim that on the basis of our sensory impressions alone we can have no knowledge of the external world. However, Crathorn averts an epistemological crisis by appealing to the knowledge, available to us a priori, that God would not let us be deceived in this way - at least not regularly.

According to his contemporary and rival Robert Holcot, Crathorn later publicly lectured on the Bible, but no record of these lectures has been found. Richter (1972) describes a manuscript containing what purports to be

quodlibetal questions by Crathorn; some of these questions are from the *Quaestiones*, others not. This additional material, largely philosophical rather than specifically theological, has been neither edited nor studied.

See also: Aristotelianism, medieval; Nominalism; Universals

ROBERT PASNAU

List of works

Crathorn, W. (*c*.1330) *Quaestiones in primum librum Sententiarum (Questions on the First Book of the Sentences)*; ed. F. Hoffmann, *Quästionen zum ersten Sentenzenbuch*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters NF29, Münster: Aschendorff, 1988.(Although unevenly edited, this is the only edition of the entire Latin text of the *Quaestiones*. No translations of Crathorn's work have been published.)

References and further reading

Gelber, H. (1984) 'I Cannot Tell a Lie: Hugh Lawton's Critique of Ockham on Mental Language', *Franciscan Studies* 44: 141-79.(Summarizes Crathorn's account of mental language.)

Nuchelmans, G. (1973) *Theories of the Proposition: Ancient and Medieval Conceptions of the Bearers of Truth and Falsity*, North Holland Linguistic Series 8, Amsterdam: North Holland.(A sophisticated discussion of Crathorn's use of the things signified by a proposition as the objects of knowledge.)

Pasnau, R. (1997) *Theories of Cognition in the Later Middle Ages*, New York: Cambridge University Press.(Gives an extended discussion of Crathorn's likeness theory of mental representation (Chapter 3) and his treatment of scepticism (Chapter 7).)

Richter, V. (1972) 'Handschriftliches zu Crathorn' (Manuscripts Associated with Crathorn), *Zeitschrift für katholische Theologie* 94: 445-9.(A groundbreaking investigation into Crathorn's life and work.)

Schepers, H. (1970) 'Holkot contra dicta Crathorn' (Holkot Against the Claims of Crathorn), *Philosophisches Jahrbuch* 77: 320-54.(Assesses the philosophical dispute between Crathorn and Holcot, with particular reference to their dispute over the objects of knowledge.)

Schepers, H. (1972) 'Holkot contra dicta Crathorn II' (Holkot Against the Claims of Crathorn II), *Philosophisches Jahrbuch* 79: 106-36.(Assesses the philosophical dispute between Crathorn and Holcot, with particular reference to their dispute over the objects of knowledge.)

Tachau, K. (1988) *Vision and Certitude in the Age of Ockham: Optics, Epistemology and the Foundations of Semantics, 1250-1345*, Leiden: Brill.(Contains a good discussion of Crathorn's epistemology.)

Tachau, K. (1995) 'Introduction', in R. Holcot, *Seeing the Future Clearly: Questions on Future Contingents*, Toronto, Ont.: Pontifical Institute of Medieval Studies.(Offers the latest and most persuasive reconstruction of Crathorn's career and works.)

Wood, R. (1988) 'Introduction', in Adam de Wodeham, *Tractatus de Indivisibilibus*, Dordrecht: Reidel.(Outlines Crathorn's view that the continuum is composed of a finite number of indivisibles.)

Cratylus (late 5th/early 4th century BC)

Probably an Athenain, Cratylus was a radical Heraclitean, holding that the world is in constant and total flux. Through this doctrine he had a seminal influence on Plato. Cratylus also, for some time at least, defended the natural correctness of names.

Taking ’Everything is in flux’ to be the core of Heraclitus’ philosophy (see Heraclitus §3), Cratylus concluded (Aristotle, *Metaphysics* 1010a9) that Heraclitus was wrong to deny the possibility of stepping twice into the same river: total flux excludes stepping into the same river even once, and it also implies that one should say nothing. Accordingly, Cratylus ‘finally’ disavowed speech and merely waggled his finger. In the words of Aristotle (Aristotle, *Metaphysics* 987a32), the young Plato ‘became familiar with Cratylus and with the Heraclitean doctrines that all phenomena are always in a state of flux, and that there is no knowledge about them’. These doctrines, Plato himself later espoused.

Both Plato and Aristotle regularly associate Heraclitus with the constant flux of everything. Some scholars take this to be a misinterpretation of Heraclitus, mediated by Cratylus, with little support in Heraclitus’ transmitted words. Yet Aristotle’s remarks about Cratylus only make sense on the assumption that Cratylus took himself to be improving the consistency and implications of Heraclitus’ own flux doctrine. Total flux excludes everything, including a river, from ever being ’the same’. The inference that nothing has an intrinsic identity if everything is in total flux is drawn by Socrates in Plato’s *Theaetetus* (156a-157c) (see Plato §15). There is also a probable echo of Cratylus’ recommendations about refraining from speech when Socrates argues that no names are legitimate that ’make anything stand still’. Plato’s dialogue *Cratylus* ends with Socrates arguing that knowledge is impossible ’if everything is changing and nothing persists’. Plato responded to Cratylus by arguing that even if total flux holds true for the phenomenal world, it does not pertain to the stable and everlasting Forms, which are the only genuine objects of knowledge.

As a linguistic naturalist, Cratylus denied that usage has any bearing on the correctness of names. All names that are really names are correct, and their correctness is due to their capacity, irrespective of particular languages, to signify the nature of things. A phonetic item that is inappropriate to the nature of the thing cannot be that thing’s name. He rejected therefore the possibility of speaking falsely. In Plato’s portrayal Cratylus, as a Heraclitean, approves Socrates’ efforts to explain the phonetics of Greek words as natural signifiers of flux even after Socrates has shown that the hypothesis cannot be valid for all names. By the time Cratylus decided to waggle his finger instead of speaking, had he abandoned this linguistic naturalism? Many have thought so, but the supposition is not necessary. Cratylus may have still believed that the only correct names must be names that reveal the nature of things. His final position was probably not the abandonment of this thesis but the inference that things in total flux have no nature that names could reveal.

A.A. LONG

References and further reading

Allan, D.J. (1954) ’The Problem of Cratylus’, *American Journal of Philology* 75: 271-87.(Develops the theory that Cratylus abandoned his linguistic naturalism in later life.)

Aristotle (*c.* mid fourth century BC) *Metaphysics*, trans in J. Barnes (ed.) *The Complete Works of Aristotle*, Princeton, NJ: Princeton University Press, 1984.(Includes an important survey of earlier thought.)

Cratylus (late 5th/early 4th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 2: 69-70.(The standard edition of the writings of the Presocratic philosophers; includes Greek texts, with translations in German, of the testimonia from Aristotle and a selection from Plato’s *Cratylus*.)

Mouraviev, S.N. (1994) ’Cratylos (d’ Athènes?)’, in *Dictionnaire des Philosophes Antiques*, ed. R. Goulet, Paris: CRNS Éditions, 1994, vol. 2, 503-10.(Lucid guide to the evidence and modern scholarship.)

Plato (*c.*380-367 BC) *Theaetetus*, trans. M.J. Levett, revised M. Burnyeat, *The Theaetetus of Plato*, Indianapolis, IN: Hackett, 1990.

Plato (*c.*380-367 BC) *Cratylus*, trans. in J. M. Cooper (ed.) *Plato: Complete Works*, Indianapolis, IN: Hackett, 1997.(Our main source on Cratylus, especially his linguistc naturalism.)

Creation and conservation, religious doctrine of

The doctrine of the creation of the universe by God is common to the monotheistic religions of Judaism, Christianity and Islam; reflection on creation has been most extensively developed within the Christian tradition. Creation is by a single supreme God, not a group of deities, and is an 'absolute' creation (creation ex nihilo, *'out of nothing') rather than being either a 'making' out of previously existing material or an 'emanation' (outflow) from God's own nature. Creation, furthermore, is a free act on God's part; he has no 'need' to create but has done so out of love and generosity. He not only created the universe 'in the beginning', but he sustains ('conserves') it by his power at each moment of its existence; without God's support it would instantly collapse into nothingness. It is controversial whether the belief in divine creation receives support from contemporary cosmology, as seen in the 'Big Bang' theory.*

1 Creation and polytheism

That the universe has been created by God is one of the most fundamental tenets of Judaism, Christianity and Islam. Creation plays a key role in the theistic worldview and is rich with implications. Every major alternative confronting the theistic worldview has presented a different kind of challenge to the belief in creation. Here we consider several such challenges, taken roughly in historical order.

The biblical belief in divine creation took form in the midst of the conflict between the monotheistic faith of Israel and ancient polytheism. Particularly significant was Mesopotamia, which was both the region from which Abraham came ('Ur of the Chaldees') and the scene of the formative experience of the Babylonian exile. The creation narrative in Genesis 1 has marked parallels with the Babylonian *Enuma Elish* but the theological context and emphasis are vastly different (see Heidel 1951). Here we do not have the numerous gods in their rivalry and occasional strife, but rather the single creator God ordering all things in accordance with his Word. That God simply speaks, with effortless mastery - 'and it was so' - contrasts with the various means of creation in contemporary mythology. There is no trace in Genesis of a struggle in which the earth is fashioned out of the corpse of a slain monster, or of a divine 'begetting' in which some part of the cosmos comes from the deity by sexual generation (or vice versa). Particularly noteworthy is the treatment of the heavenly bodies, which in Babylon as in most ancient cultures were worshipped as divine. Rather than being given divine honours, they are assigned an important but strictly functional role in the creation ('to give light upon the earth, to rule over the day and over the night, and to separate the light from the darkness'); the stars are added almost as an afterthought.

2 Creator or Demiurge?

When the Christians of the second and third centuries began to reflect upon their faith in the light of Greek philosophy, one of the readily available conceptual models for creation was the one presented by Plato (§16) in the *Timaeus*, where the Demiurge in his beneficence 'desired that all things should come as near as possible to being like himself' (*Timaeus* 29e). But certain deficiencies were evident, as compared with what Christians wished to say about their God. The Demiurge is not ontologically supreme, but clearly is subordinate to the Forms. The good intentions of the Demiurge are limited and often frustrated by the recalcitrant matter with which he must work, but which he had no part in creating. In later Gnostic worldviews, the status of matter becomes even more problematic: it is evil in itself, and the fundamental source of evil in human life. The Demiurge, who has imprisoned humans in mortal bodies, is contrasted with and inferior to the High God, who seeks to liberate them from material contamination (see Gnosticism).

In opposition to this, the Church Fathers insisted that God is the supreme source of all concrete existence other than himself: he has created all things *ex nihilo*, 'out of nothing'. The idea is not, of course, that 'nothing' is the name of some negative, peculiarly elusive 'stuff' out of which creation came. On the contrary, creation *ex nihilo* means precisely that there is no pre-existing 'stuff' whatsoever - that things have come to exist solely because of the Word and creative power of God. To be sure, creation *ex nihilo* is not to be found in Genesis; the chaos of Genesis 1: 2 is not said to be created, but is presupposed in the later acts of creation. But creation out of nothing is at least strongly suggested in the New Testament (for example, Romans 4: 17, Hebrews 11: 3), was clearly affirmed by the early Church Fathers (for example, Irenaeus), and has been the consistent teaching of Christian theologians, apart from occasional lapses in the direction of emanationism (see §3). And matter, the physical

realm, is not some negative, inhibiting factor which impedes God in the achievement of his purpose; rather, it is itself created by God and is an integral part of the creation which he pronounced 'very good'.

3 Creation and emanation

Among the philosophical options offered by the ancient world, one of the most attractive to Christian thought was the Neoplatonism of Plotinus (see Neoplatonism; Plotinus §§3-5). This theory emphasized the vast difference between the transcendent God (or 'One') and the mundane realm, and urged human beings, in pursuit of their true good, to seek reunion with the divine source from which they had come. Neoplatonism furnished a rational, philosophical account of the derivation of the world from that source, one that had to be taken seriously by theologians who needed to free themselves from crude, anthropomorphic images of divine 'making'. The idea of evil as 'privation', taken from Neoplatonism, went far towards an understanding of how evil can exist in a world wholly created by God without God's having created evil. Evil, on this account, is not a positive existence in itself, but simply a deficiency in a created being which is, as created, good. And since to be created is as such to be less than God and thus subject to imperfection, there is no occasion for surprise (or for any reproach to God) in the existence of such imperfections in the creation. Finally, some of the 'emanations' from the One were fairly readily assimilated to the emerging doctrine of the Trinity, thus enhancing both the rational appeal of that doctrine and the attractiveness for Christians of the Neoplatonic philosophy.

For Neoplatonism, the fundamental concept for relating the One (that is, God) to the world is that of emanation. In general, emanationist schemes have the following characteristics:

1 God, or the One, is conceived in terms of extreme transcendence, so remote from everyday reality that he (or it) can be described only in negations.
2 God is linked to the world of everyday experience through a graded series of levels. For Plotinus, these are Mind, then Soul (itself divided into an upper and a lower aspect), and then the individual souls of humans and animals, until finally the lowest level, that of material bodies, is reached.
3 Each lower level in the series is derived from the next higher level through the process of emanation. The lower level 'flows out' from the upper level without diminishing it; illustrative analogies include the outflow of a stream from a spring, and light radiating from a lamp. Thus the lower level is inferior in ontological status, yet shares in the nature of the higher level which produces it.
4 The emanation of a lower level from a higher is eternal and necessary; it follows from the nature of the higher level, and does not involve or depend on a decision of will.

This general scheme was adopted with only minor modifications by some Christians, such as Pseudo-Dionysius and Johannes Scottus Eriugena. It was also taken over by the Muslim philosophers al-Farabi (§2) and Ibn Sina (§5). But it was rejected or severely modified by the major theologians of Judaism (see Maimonides §3), Christianity (for example, Augustine and Aquinas>) and Islam (see al-Ghazali §3; Neoplatonism in Islamic philosophy). In general, the following aspects of emanationism are objectionable from the standpoint of biblical or Qur'anic thought:

1 The extreme separation of the Source from the everyday world, and the refusal to apply ordinary predicates to it, make it problematic to equate it with the personal God of theism. In particular, the need for a series of intermediaries linking God and the world is in conflict with the biblical and Qur'anic witness to the immediate presence and activity of God in all aspects of existence.
2 On the other hand, the nature of the process of emanation threatens the fundamental distinction between God and creatures that is essential to theism. 'Creation' in this scheme is really not *ex nihilo* but rather *ex deo* - a conclusion which is hardly welcome to any of the monotheistic religions.
3 Finally, the necessary, nonvolitional character of the process of emanation undermines the freedom of God in creation, and thus also the sense of gratitude to God for the generous gift of existence he has bestowed on us and on the world.

The concept of creation, as developed by theistic philosophers and theologians, agrees with emanationism in strongly emphasizing the dependence of creatures on God. But it insists on a fundamental ontological distinction between the creator and the creatures: creation is *by* God, but not *out of* God. God bestows on creatures a being that is their own and not his, though it is always dependent upon him. Furthermore, it is insisted that creation is a

free and gratuitous act on God's part: there was and is no necessity for God to create a world. To this it might be objected that, since it is obviously better for there to be a universe than for there not to be one, a wise and good God must of necessity have created a world. The answer to this objection is rather startling: God is infinitely greater than creatures, so the value of God plus the entire creation is no greater than the value of God by himself. (Adding a finite quantity, no matter how large, to infinity still leaves you with just infinity, not 'infinity plus'.)

4 Creation as temporal or eternal?

One further question raised by emanationism is not answered by the considerations discussed above. Is creation temporal or eternal? Does the universe have a temporal beginning or is it beginningless? This question was especially acute for medieval thinkers because it was claimed that Aristotle had demonstrated the eternity of the world; on the other hand, the Scriptures seemed to point definitely towards a temporal beginning (see Eternity of the world, medieval views of). In response to this, Maimonides pointed out that even if the eternity of the world had been demonstrated, it would be perfectly possible to give a figurative interpretation of those biblical texts that seem to imply a temporal beginning. Aquinas and other Christian writers had less freedom, because the Fourth Lateran Council (1215) had defined it as a doctrine of faith that the universe had a beginning in time.

The predominant view at the time among both Christians (such as Bonaventure (§5)) and orthodox Muslims was that, contrary to the claims of the Aristotelians, the temporal finitude of the universe could be conclusively demonstrated. This was argued on the basis of certain problems with the idea of an 'actual' (already realized) infinity, such as seems to be involved in the concept of an infinite past. For example: if the past is infinite, then there has occurred a time which was an infinite number of years prior to the present. But if this were so, it would be impossible that, through the passing of years one by one, an infinite number would be traversed and we would find ourselves in the present. So if the past were infinite, the present would never have been reached, which is absurd. (This problem does not arise for an infinite future, because future eternity, unlike past eternity, is only a 'potential' infinity; there will never come a future time separated from the present by an infinite number of years.)

The response of both Maimonides and Aquinas at this point is that neither the finitude nor the infinity of the past has been demonstrated; philosophically, the question remains open. (The answer given to the argument described above was that, even if the past is infinite, it still is not the case that there exists some time which occurred an infinite number of years before the present. Any given time in the past preceded the present by only a finite number of years - and was itself preceded by an already infinite past.) Both Maimonides and Aquinas, then, were free to affirm on the basis of Scripture that the past is finite and that the world has a temporal beginning.

It is important to see, however, that the question of the world's eternity is not settled by the considerations discussed in §3 above. For suppose that, contrary to the conclusion of Maimonides and Aquinas, the world does indeed have an infinite, beginningless past. It could still be said that the universe is created *ex nihilo*, in the sense that it has no existence except that which is wholly dependent on the will and power of God. Its relation to God would then be just what theists suppose the relation to God of the actual universe to be, at every moment subsequent to the first. It could, furthermore, still be the case that God is entirely free in creating the universe - that it is simply by God's free decision that there is eternally a universe distinct from God, rather than God existing alone in self-sufficient perfection.

That is not to say that the issue of a temporal creation is religiously indifferent; there are, after all, the biblical and Qur'anic texts that seem to imply such a beginning. It is clear, furthermore, that the notion of an eternal creation strongly suggests, even if it does not logically imply, some sort of necessity for the creation process - either because God has some kind of need for a creation, or because the universe is produced from the divine nature by some sort of necessity. A belief in the eternity of creation tends towards a denial of the freedom of God in creation, even if it does not force us to that conclusion.

5 Conservation, continuous creation and deism

Belief in God as creator has almost always involved the affirmation that God not only produced the universe in the beginning, but sustains it in being from moment to moment. The New Testament says of Christ that 'in him all things hold together' (Colossians 1: 17), and describes him as 'sustaining all things by his powerful word' (Hebrews 1: 3). And the Psalms are pervaded by the sense of the immediate dependence of the creatures on their

creator: ‘When thou sendest forth thy Spirit, they are created; and thou renewest the face of the ground’ (Psalms 104: 30). So theistic philosophers and theologians have affirmed the divine conservation of creatures, claiming that no creature exists at any moment unless God actively sustains its existence at that moment; without this support it would instantly vanish into nothingness.

There has been debate, however, about the use of the phrase ‘continuous creation’ to describe this process. Descartes wrote that ‘the same power and action are needed to preserve anything at each individual moment of its duration as would be required to create that thing anew if it were not yet in existence’ ([1641] 1984: 33). But this does not seem quite correct. No doubt it is true that a being already in existence is as unable to exist without divine power as one which newly comes into existence this very moment. But it is not strictly true that God does exactly the same thing in the two cases. When he newly creates something *ex nihilo*, he brings into existence something that previously did not exist, either as a whole or in its several parts. When he conserves some being, then he sustains the continued existence of a being that is already in existence. And this point is crucial for countering the argument for occasionalism (see Occasionalism). When God newly creates a being *ex nihilo*, then God alone must determine all that being’s characteristics; there is nothing else that might perform this task. But when God conserves a being that is already in existence, then the previous state and causal powers of that being can play a part - often a large part - in determining the being’s state and characteristics at the time in question.

So it does not seem to be the case that God does exactly the same thing when he creates something anew as when he conserves it in existence. The expression ‘continuous creation’ might still be used as a dramatic way of calling attention to the fact that an already existing being has no more power to maintain itself in existence than a nonexistent and merely possible being has to bring itself into existence. But a straightforward equating of creation with conservation seems to be ruled out.

It should be noted here that most medieval theologians held that, in addition to conserving the world in existence, God ‘concurs’ in each and every instance of creaturely action. This concurrence is not mere permission, but is rather an active causal contribution by God in cooperation with the creaturely action, which apart from this cannot bring about its result. This doctrine of divine *concursus* has not played a major role in recent discussions of creation; it remains to be seen whether a convincing case can be made for it.

The deists rejected the doctrine of conservation, asking in effect why God could not create beings which, once created, could go on about their business without further direct help from him (see Deism). But the religious case for divine conservation is compelling, even aside from scriptural texts. If God did not act to conserve creatures in existence, then his present action in the world would be limited to special miraculous acts of divine intervention - a highly undesirable result, even if one has no objection to the idea of miracle as such. The religious sense of dependence upon God and of being sustained by God, an important part of the religious life for a great many believers, would have to be either rejected as illusory or radically reinterpreted. (It could hardly be maintained that we are able to have, now, an experience of ‘dependence’ on a long-ago act of initial creation.) It could still be said that we are ‘dependent’ on God in the sense that he could annihilate us at any moment - but that is precisely the sense in which hostages become dependent upon terrorists to spare their lives; it is hardly of much use to religion. To be sure, it may be that none of these arguments would have persuaded the deists, who seem to have felt that although religion is somehow necessary, the course of prudence is to get along with as little of it as one can manage.

6 Contemporary issues

For most of the nineteenth and twentieth centuries, such metaphysical reflections as these were out of fashion, and various other issues came to the fore. A problem of some importance in the mid twentieth century was the question of the meaningfulness and intelligibility of the doctrine of creation. It was noted quite correctly that, once we get beyond crude anthropomorphic ideas of ‘making’, the terms used in speaking of creation are employed in ways far removed from their familiar uses. The question, then, is whether the ordinary meaning of the terms has not been eroded away without remainder, and without anything else having been put in its place. There are indeed analogies that can be adduced here: our experience of the instability of many things (for example, of a home constructed in an earthquake zone) may illuminate the ‘ontological instability’ of things and thus their dependence upon God. And we speak of God as ‘planning’ for the world by analogy with our own planning for future contingencies. But the instabilities we are aware of are instabilities *within the world*, remediable by ordinary, intra-worldly means;

our experience of planning presupposes all sorts of limitations on our part which are explicitly excluded in the case of God. So the meaning which our analogies seem to provide may well prove to be illusory.

It is fair to say that this sort of objection became progressively less compelling during the second half of the twentieth century. To speak about the doctrine of creation with some accuracy it is indeed necessary to speak in explicitly metaphysical terms - something which philosophers have increasingly found themselves willing and able to do (for example, Richard Swinburne and Alvin Plantinga). The analogies with ordinary experience are undoubtedly important in enabling us to grasp the significance of the doctrine. But the analogies are abused if we demand that the very same relation which obtains between God and creatures should be exemplified in the relations between creatures.

Most contemporary discussion of creation concerns its relation to natural science. Much of this discussion is over points in contention between the worldviews of theism and scientific naturalism, and cannot be considered here. But there are two points at which science impinges directly on traditional thinking about creation: evolutionary biology, and cosmology.

There is a broad consensus among theologians, including most conservative theologians, that biological evolution as such poses no threat to the doctrine of creation. There is, however, a need to contest philosophical interpretations of evolution in terms of scientific naturalism, interpretations which are sometimes propounded by leading scientists. There is also a need to rethink certain areas of theology (especially theological anthropology) in the light of the acceptance of human evolution. There are differences among theologians over the extent to which evolution should be incorporated into positive theological teaching. Neo-orthodox theologians such as Karl Barth and Emil Brunner accepted evolution, but it was peripheral to their major theological concerns. Teilhard de Chardin, on the other hand, sought to establish a strong positive relationship between Christian hope and evolutionary optimism.

An area of particular interest is the relationship between the doctrine of creation and scientific cosmology (see God, arguments for the existence of §5; Religion and science §6). Here theology seems to be in the unusual position that it cannot be harmed by science but might be in a position to benefit from it. The key theological point at issue is the age of the universe as finite or infinite. It seems theology has nothing to fear from cosmology on this point. There is for one thing the possibility (though not a favoured possibility) for theology to accept an eternal universe (see §4 above). On the other hand, it seems that it would not be possible for science to establish that the universe is infinitely old: what might a scientific proof of that look like? If science were to find no evidence for a ‘beginning’, then at worst theology would be back in the same position as that accepted by Aquinas - able to say that reason cannot settle the issue, but that revelation informs us of a temporal creation. In fact, however, the triumph of ‘Big Bang’ cosmology over the ‘steady state’ theory (with additional help from the second law of thermodynamics) seems to provide substantial support for the idea of a temporal origin of the universe. It would be unwise for theists to cite these scientific developments as ‘proof’ of an absolute beginning, and thus of the need for a First Cause. Such a claim would underestimate the ingenuity of scientists, who continue to produce scenarios which permit an eternal universe. (To be sure, the scenarios available at present are highly speculative, and some of them may be motivated as much by aversion to the idea of an absolute beginning, and therefore a possible creation, as by empirical data.) But one may at least conclude, with Ernan McMullin, that ‘if the universe began in time through the act of a Creator, from our vantage point it would look something like the Big Bang that cosmologists are now talking about’ (1981: 39).

See also: Eternity; Freedom, divine; God, concepts of; Motoori Norinaga; Miracles; Pantheism; Tian

WILLIAM HASKER

References and further reading

Burrell, D.B. (1993) *Freedom and Creation in Three Traditions*, Notre Dame, IN: University of Notre Dame Press.(Comparative analysis of medieval thought on freedom and creation in Judaism, Christianity and Islam.)

Clifford, R.J. (1988) ‘Creation in the Hebrew Bible’, in *Physics, Philosophy, and Theology: A Common Quest for Understanding*, Vatican City State: Vatican Observatory, 151-70.(Concise analysis of the teaching of the Hebrew Bible on creation.)

Craig, W.L. and Smith, Q. (1993) *Theism, Atheism, and Big Bang Cosmology*, Oxford: Clarendon Press.(Debate on the implications of Big Bang cosmology with a finite past for the universe.)

Descartes, R. (1641) *Meditations on First Philosophy*, trans. J. Cottingham, R. Stoothoff and D. Murdoch, *The Philosophical Writings of Descartes*, vol. 2, Cambridge: Cambridge University Press, 1984.(The passage cited in §5 is found in Meditation III.)

Drees, W.B. (1990) *Beyond the Big Bang: Quantum Cosmologies and God*, La Salle, IL: Open Court.(Excellent, accessible summary of the state of cosmology as of the publication date; somewhat limited in philosophical analysis.)

Flew, A. and MacKinnon, D.M. (1955) 'Creation', in A. Flew and A. MacIntyre (eds) *New Essays in Philosophical Theology*, London: Macmillan, 170-86.(Contains a good presentation by Flew of the doubts about the intelligibility of creation language.)

Freddoso, A.J. (1988) 'Medieval Aristotelianism and the Case against Secondary Causation in Nature', in T.V. Morris (ed.) *Divine and Human Action: Essays in the Metaphysics of Theism*, Ithaca, NY, and London: Cornell University Press, 74-118. (Metaphysical analysis of occasionalism, the 'third alternative' to concurrentism and mere conservationism.)

Freddoso, A.J. (1991) 'God's General Concurrence with Secondary Causes: Why Conservation is not Enough', *Philosophical Perspectives* 5: 553-85.(Analysis of the positions of concurrentism and mere conservationism, with Suarez's arguments for concurrentism.)

Freddoso, A.J. (1994) 'God's General Concurrence with Secondary Causes: Pitfalls and Prospects', *American Catholic Philosophical Quarterly* 68 (2): 131-56.(Analyses concurrentism and defends it against the arguments of Durandus.)

Heidel, A. (trans. and ed.) (1951) *The Babylonian Genesis: The Story of Creation*, Chicago, IL: University of Chicago Press, 2nd edn.(Contains *Enuma Elish* and other Babylonian creation narratives, with a discussion of Old Testament parallels.)

Kretzmann, N. (1985) 'Ockham and the Creation of the Beginningless World', *Franciscan Studies*, vol. 45, annual XXIII: 1-31.(Analyses Ockham's responses to the arguments claiming to prove a temporal creation.)

Kvanvig, J.L. and McCann, H.J. (1988) 'Divine Conservation and the Persistence of the World', in T.V. Morris (ed.) *Divine and Human Action: Essays in the Metaphysics of Theism*, Ithaca, NY, and London: Cornell University Press, 13-49.(Argues that the persistence of the world in being requires a divine cause.)

McMullin, E. (1981) 'How Should Cosmology Relate to Theology?' in A.J. Peacocke (ed.) *The Sciences and Theology in the Twentieth Century*, Notre Dame, IN: University of Notre Dame Press, 17-57.(An accurate and very cautious assessment of the relationship between cosmology and theology.)

McMullin, E. (ed.) (1985) *Evolution and Creation*, Notre Dame, IN: University of Notre Dame Press.(Several essays deal with the relationship of evolution to theology; the editor's introduction provides a valuable historical summary.)

Plato (*c.*366-360 BC) *Timaeus*, trans. F.M. Cornford, Indianapolis, IN: Library of Liberal Arts, 1959.(Plato's 'likely story' concerning the origin and structure of the cosmos.)

Quinn, P.L. (1988) 'Divine Conservation, Secondary Causes, and Occasionalism', in T.V. Morris (ed.) *Divine and Human Action: Essays in the Metaphysics of Theism*, Ithaca, NY, and London: Cornell University Press, 50-73. (Argues for mere conservationism, and refutes the charge that the doctrine of divine conservation leads to occasionalism.)

Westermann, C. (1971) *Schöpfung*, Stuttgart: Kreuz-Verlag; trans. J.J. Scullion, *Creation*, Philadelphia, PA: Fortress Press, 1974.(Interesting discussion of the biblical narratives of creation and Fall in the light of ancient Near-Eastern parallels.)

Crescas, Hasdai (*c*.1340-1410)

During the most tragic period of Spanish-Jewish history (1391-1492), Hasdai Crescas wrote a philosophical-theological treatise, Or Adonai (Light of the Lord), seeking to define and fortify the Jewish faith in the face of constant Christian attack. It is a polemical book, aiming to defend a traditional version of Judaism by criticizing the Aristotelian formulations proposed by such Jewish philosophers as Moses Maimonides and Levi ben Gershom (Gersonides). Since they relied on Aristotelian physics, Crescas began his reconstruction of Jewish theology with a demolition of Aristotle's natural philosophy. He then turned to metaphysics in general and Jewish theology in particular. His constructive work was especially novel in its treatment of human choice, divine omniscience and creation. Aiming to defend traditional Jewish ideas, Crescas in the end broached some of the most radical challenges to be found within the medieval philosophical tradition, including the proposal that there might be numerous worlds other than our own, infinite magnitudes, and a void, or vacuum.

1 Crescas' motives

Perhaps the most original and acute of all medieval Jewish philosophers, Hasdai Crescas lived through one of the most difficult periods of Jewish history. From 1391 to 1492, Spanish Jewry, the traditional repository of Jewish philosophy, suffered one defeat after another, culminating in the expulsion of the Jews who remained loyal to their faith in 1492. As chief rabbi of Aragonese Jewry, Crescas was continually called upon to defend both Jews and Judaism against the onslaughts of the Spanish clergy and populace. It is in this context that one must understand the purpose, tone and content of his major work, *Or Adonai* (1410), and his polemical tract, *Bittul 'Iqqarei ha-Notzrim (Refutation of Christian Principles)* (1398). The former work is the focus of our attention here, but it is noteworthy that the latter employs philosophical arguments throughout. One of its main theses is that core Christian dogmas are incompatible with reason.

One of the main threats to Judaism, Crescas believed, was internal: an 'alien woman', namely Aristotelian philosophy, had seduced Jewish thinkers and their followers into a misguided conception of Judaism, which made it easy for the Christians to convert Jews. To defend the Judaism he deemed authentic, Crescas proposed to banish the foreign influence by exposing its pretensions and demolishing its foundations. To this end, he was bold enough to focus his attack on the chief conveyer of the Aristotelian philosophy that had been imported into the sacred domain of the Torah, Moses Maimonides, whose Jewish creed was rooted in the conceptual framework of Aristotelian natural philosophy and metaphysics. *Or Adonai* constructs a new creed on very different conceptual foundations. In elaborating his views, Crescas is concerned to show that the authentic Jewish position is not to be formulated in or defended by Aristotelian language or arguments. Nevertheless, his method is philosophical throughout: to refute the Aristotelians he employs rigorous argumentation and fine analysis worthy of the best of his adversaries.

The credal structure of Judaism, Crescas claimed, is a fourfold system whose foundational axioms are the 'root beliefs' in the existence, unity and incorporeality of God. These three doctrines define religion generically, but only minimally. A higher tier is constituted by the ideas that make up a revealed religion, such as divine omniscience and prophecy. On the 'third floor' are the beliefs taught by the religion Crescas takes to be the true revealed religion - Judaism. These include the creation of the world, the primacy of Mosaic prophecy, and the coming of the messiah. Acceptance of these doctrines is obligatory for all Jews, but their truth is not entailed in the general concept of a revealed religion. Finally, in the 'attic' are views upheld by various Jewish sages which have no binding force, although one may believe in them. Examples include the plurality of universes, the existence of demons, and theories about the nature of the First Mover.

2 Critique of Aristotle

For the Aristotelian philosopher or theologian, metaphysics in general and theology in particular rest on natural philosophy. Thus Crescas' reconstruction of Jewish theology begins with a radical critique of Aristotelian physics. Maimonides had prefaced his exposition of the arguments of the Aristotelian school for the existence, unity and incorporeality of God by laying down twenty-six principles of Aristotelian natural philosophy. Although Maimonides himself had parted company with the Aristotelian tradition, he had explicitly criticized in detail only one of the Aristotelian philosophers' first premises that he listed, the eternity of the cosmos, treating the rest, by

and large, as useful cosmological axioms. Crescas targets the premises of the Aristotelian philosophers at large, aiming, through them, at what he conceives as unreasonable and damaging admissions of Aristotelian views by Maimonides.

The theological thrust of Crescas' 'scientific innovations' can be seen in his critique of several of the standard medieval proofs for the existence of God. Many of these arguments presuppose the impossibility of an actual infinity, an impossibility that Aristotle believed he had proved. There were arguments, for example, which assumed that there cannot be an infinite series of moving causes, so that there must be an Unmoved Mover as the ultimate cause of motion. Or, it was assumed, there cannot be an infinite series of causes and effects existing simultaneously, so there must be a First Cause to produce and sustain the whole series.

But why, Crescas asks, can there not be an actual infinity? After exposing flaws in Aristotle's arguments, Crescas affirms not only the possibility of an actual infinity but its existence. Space, he argues, the vacuum, is actually infinite. Here Crescas compounds his critique by identifying the void as infinite. For Aristotle had denied not only the possibility of an actual infinite, but also the possibility of a vacuum or void, which had seemed to him to involve a variety of absurd implications.

Aristotle and his followers had faced a vexing problem when asked what lies beyond their finite world. Aristotle's answer was that there is nothing - no vacuum, no space, no other world. In effect, he rejected even the question. To Crescas this seemed arbitrary at best. His own answer seemed more sensible: that there is infinite space beyond our world, extension, in which there is a real possibility of other worlds.

Having opened the door to the actual infinite, Crescas begins to open other doors as well: perhaps there is an infinite number of actual individuals or an infinitely large body. None of these hypotheses is absurd or improbable. Crescas' hospitality to an actual infinite was appreciated by Spinoza, who became a powerful advocate in behalf of an idea once widely taken by philosophers to be absurd on the face of it (see Spinoza, B. de §3).

3 Divine omniscience and human choice

The second part of Crescas' treatise discusses the theological dogmas presupposed by any revealed religion. His most penetrating discussions are found in his treatment of divine omniscience and human choice. He analyses these separately, but they can be discussed together since they are closely related in most medieval treatments. Like most medievals, Crescas defends a strong version of divine omniscience: God knows all truths. Many ancient and medieval thinkers had thought God's knowledge of future contingent facts problematic if not impossible. Gersonides (1288-1344), Crescas' 'philosophical litigant', had concluded that God does not know the future doings of human beings as individual events, but only the general laws pertaining to them. Crescas deems this not just false but heretical. After all, if God did not know Moses as an individual, how could God speak to him and give him the Torah? For Crescas, Gersonides' doctrine makes nonsense of the very idea of prophecy and revelation.

Crescas does not rest his case on revelation. He analyses each of Gersonides' arguments and finds them all invalid. He argues, for example, that even if it were true that temporal facts are known by temporal beings through sense-perception, it does not follow that this is the only way they are knowable. Why could a nontemporal being not know them in some nontemporal way? Since everyone agrees that God is eternal, it should also be admitted that God knows everything, including facts about the future, eternally, that is, timelessly. Strictly, there is no future or past to God. Crescas' solution is similar to that reached by many late ancient and medieval thinkers, such as Proclus, Boethius and Aquinas, but he was the first to propose it in the Jewish philosophical literature.

Now if God knows future contingents timelessly, can we still say that our choices are themselves contingent, or free? If not, all legal and moral obligations would be pointless. Crescas' response to this ancient dilemma, as he himself realized, is more novel and daring than his answer to the question about divine knowledge. He declares boldly that in one sense, in so far as they have causes, our choices and the actions based upon them are determined and necessary.

Crescas is an explicit and unabashed defender of causal determinism, not only in nature at large but in human behaviour specifically. If our choice to give charity were uncaused, it would be inexplicable and groundless. But should such an event be the subject of praise and reward? Crescas' answer is that, even though the event is

necessary in so far as it is caused, it is contingent as such, that is, considered in itself, for it is a logically possible state of affairs, not a necessary truth like the theorems of logic or mathematics. My act of charity has a cause, and in this sense is determined and necessary, but it is still free, since it is, in itself, merely a logically contingent state of affairs. As long as I am not subject to external compulsion, such that I feel forced, an action can be deemed voluntary. So what I do tomorrow remains contingent, despite God’s eternal knowledge. Indeed, in different circumstances I could have acted otherwise. And God, in that case, would eternally have known otherwise.

4 Creation

In Part Three of *Or Adonai*, Crescas discusses those beliefs that are not essential to all revealed religions but are taught by the one revealed religion that Crescas takes to be true, that of the Torah. He begins by discussing creation, which is the first belief explained in the Torah. By Crescas’ time, creation had become one of the most controversial issues in philosophy. Three theories were widely debated: (1) the eternalism of Aristotle and his followers; (2) the theory of *formatio mundi* attributed to Plato, creation understood as the formation of an orderly cosmos out of formless matter; and (3) the doctrine by now most generally regarded as orthodox among Jews, Christians and Muslims, *creatio ex nihilo* (see Creation and conservation, religious doctrine of; Eternity of the world, medieval views of).

Crescas focuses chiefly on Gersonides’ modified version of the second theory and Maimonides’ support of the third. He rejects his two predecessors’ rebuttals of Aristotle’s arguments in favour of eternity and against creation, finding them either invalid or dependent on theorems of Aristotelian physics which were shown in Part One of *Or Adonai* to be false or unproved. As in his treatment of divine omniscience, Crescas is especially incensed by Gersonides, whose defence of *formatio mundi* he believes to be a misreading of the Torah and demonstrably false. Gersonides had relied heavily on the claim that *ex nihilo* creation implies the existence of a vacuum before creation and its persistence outside the world even afterwards. But, as we have seen, Crescas is prepared to defend the possibility of a vacuum. Besides, Gersonides’ argument is unsound: in creating the universe God created its spatial dimensions along with it; no empty space was needed before creation.

After disposing of the arguments of Gersonides and Maimonides, Crescas presents his own solution, redefining *ex nihilo* creation to refer to the world’s causal dependence on God. No ‘prior’ matter is needed. Indeed, the whole formulation is temporally neutral: the universe could be eternal albeit created, that is, eternally dependent upon God. A similar view had been defended by the Muslim Neoplatonists al-Farabi and Avicenna (see Ibn Sina). Crescas understands that dependence on their approach might offend some of his readers, so he offers a weaker, more traditional version of the idea: instead of one eternally existing world, an infinite series of worlds might be created by God. Our world, in that case, did have a definite temporal beginning, as tradition teaches, but God’s creativity continues at all times. Both doctrines preserve the important dogma of divine omnipotence.

5 Plurality of worlds

The final part of Crescas’ treatise contains thirteen short disquisitions on various topics about which there is no unanimity among the sages, but where plausible arguments have been adduced in favour of diverse positions. Since these doctrines have no dogmatic status, a Jew is free to affirm or deny them, or to withhold judgment.

The first of these questions is whether this world could be destroyed and replaced. Is ours only one of a series of successive created universes? This was a question of considerable interest to Christian theologians in the fourteenth century. Few Jewish philosophers had devoted much discussion to the question, although there are hints of the idea in the Rabbinic literature, and Maimonides and Gersonides had defended the indestructibility of our universe. Crescas finds none of the arguments for or against a succession of worlds compelling. But his previous suggestion that God can create an infinite series of worlds shows his attraction to this view. It is quite possible, he says, that God could destroy this world and create a more perfect one. But if so, Crescas insists, God would destroy the earth only; the heavenly bodies are everlasting. It remains in doubt whether this last is consistent with Crescas’ views on eternal creation.

In the second question Crescas wonders whether there could be a plurality of coexisting worlds. Gersonides, following Aristotle, had defended the uniqueness of the cosmos: one God, one world. Crescas is not so sure. He seems to lean towards the view that other worlds might coexist with ours. Aristotle had urged (among other

arguments) that since each of the physical elements has a natural place towards which it moves if nothing impedes it (such as fire, which naturally moves upwards towards the outermost region of the earth) then if there were another world, fire *there* would move towards the outermost region in *our* world. Since this seems untenable, our world must contain all the matter there is, and no other universe is possible. Crescas rejects this argument. Even if the theory of natural places is true - and Crescas denies it - the fire in another world would move towards the outermost region of *that* world, not ours. The doctrine of natural places, if true, must be relativized.

There is a certain unintended irony in Crescas' enterprise. He set out to defend a traditional Jewish theology, purified of Aristotelian intrusions. Yet the conclusions he reaches are novel, sometimes radical. He rejects the Aristotelian cosmology that had served for centuries as the backdrop for medieval philosophical speculations. He redefines key theological ideas. Even his pupil, Joseph Albo, criticized and rejected Crescas' deterministic account of choice. And his ill-concealed sympathy for 'eternal creation' is hardly unanimously hailed among later thinkers. The (somewhat Epicurean sounding) proposals about an infinite void and multiple worlds fell largely upon deaf ears, until, in the aftermath of the modern discrediting of the medieval worldview, Crescas' ideas found a sympathetic reader in Spinoza.

See also: Aristotle §§10, 16; Aristotelianism, medieval; Gersonides; Spinoza, B. de

SEYMOUR FELDMAN

List of works

Crescas, H. (1398) *Bittul 'Iqqarei ha-Notzrim (Refutation of Christian Principles)*, trans. D.J. Lasker, *The Refutation of Christian Principles by Hasdai Crescas*, Albany, NY: State University of New York Press, 1992. (Crescas' philosophical critique of the basic dogmas of Christianity.)

Crescas, H. (1410) *Or Adonai (Light of the Lord)*, ed. S. Fischer, Jerusalem: Sifrei Ramot, 1990; 'Choice' (Book 2: 5), trans. S. Feldman, in J.D. Bleich (ed.) *With Perfect Faith*, New York: Ktav, 1983, 472-91; 'Resurrection and Immortality' (Book 3: 1, part 2), trans. W. Harvey, in Bleich (ed.), 663-80; 'Divine Knowledge' (Book 2:1), trans. W. Harvey, in Bleich (ed.), 466-71.(Crescas' systematic philosophical reconstruction of Jewish theology, based upon a radical critique of Aristotelian philosophy and physics.)

References and further reading

Feldman, S. (1980) 'The Theory of Eternal Creation in Hasdai Crescas and Some of his Predecessors', *Viator* 11: 289-320.(Exposition of Crescas' cosmology in its historical-philosophical context.)

Feldman, S. (1982) 'Crescas' Theological Determinism', *Da'at* 9: 3-28. (Philosophical critique of Crescas' doctrine of divine omniscience.)

Feldman, S. (1984) 'A Debate Concerning Determinism in Late Medieval Jewish Philosophy', *Proceedings of the American Academy for Jewish Research* 51: 15-54.(A philosophical analysis and critique of Crescas' determinism using material from Crescas' successors.)

Harvey, W. (1973) 'Hasdai Crescas' Critique of the Theory of the Acquired Intellect', unpublished Ph.D. dissertation, Columbia University, New York.(Contains several translations from Crescas, and an excellent analysis of some major themes in his philosophy.)

Kellner, M. (1986) *Dogma in Medieval Jewish Thought*, Oxford: Oxford University Press.(Includes a useful study of Crescas' creed.)

Pines, S. (1967) *Scholasticism After Thomas Aquinas and the Teachings of Hasdai Crescas and His Predecessors*, Jerusalem: Israel Academy of Sciences and Humanities.(Seeks to show the influence of Christian thought on Crescas and some of his predecessors.)

Ravitzky, A. (1988) *Crescas' Sermon on the Passover and Studies in his Philosophy*, Jerusalem: Israel Academy of Sciences and Humanities.(Detailed historical-literary analysis of the development of Crescas' views on determinism and miracles.)

Rudavsky, T. (1990) 'The Theory of Time in Maimonides and Crescas', *Maimonidean Studies* 1: 143-62. (Challenging, but perceptive and stimulating.)

Schweid, E. (1970) 'The Religious Philosophy of Hasdai Crescas', in *Or Adonai*, Jerusalem: Makor Publishing. (Included as an introduction to the reprinting of the 1555 edition of *Or Adonai*. Excellent and comprehensive study of the major themes.)

Touati, C. (1974) 'Hasdai Crescas et ses paradoxes sur la liberté', *Mélanges d'histoire des religions offerts à Charles H. Puech*, Paris, 573-8.(Good exposure of some of the difficulties in Crescas' doctrine of divine omniscience.)

Touati, C. (1983) 'La Providence divine chez Hasdai Crescas', *Da'at* 10: 15-31.(A fine exposition of Crescas' complex theory of providence.)

Wolfson, H. (1916-17) 'Crescas on the Problem of Divine Atributes', *Jewish Quarterly Review*, n.s. 7: 1-44, 175-221; repr. in *Studies in the History and Philosophy of Religion*, Cambridge, MA: Harvard University Press, 1977, vol. 2, 247-337.(Detailed terminological study of the problem of divine attributes in Crescas and other late medieval Jewish philosophers.)

Wolfson, H. (1929) *Crescas' Critique of Aristotle*, Cambridge, MA: Harvard University Press.(The major study of Crescas' critique of Aristotelian natural philosophy; contains a critical edition, translation and commentary on Book 1 of *Or Adonai*.)

Crime and punishment

An account of how state punishment can be justified requires an account of the state, as having the authority to punish, and of crime, as that which is punished. Crime, as socially proscribed wrongdoing, may be formally censured, and may lead to the payment of compensation to those injured by it - but why should it also attract the kind of 'hard-treatment' punishment which characterizes a system of criminal law? How should we decide which kinds of wrongdoing should count as crimes?

Consequentialists justify punishment by its beneficial effects, notably in preventing crime by deterring, reforming or incapacitating potential criminals. They face the objection that the wholehearted pursuit of such goals would lead to injustice - punishment of those who do not deserve it. Even if that objection is met by imposing non-consequentialist constraints on the system, they also face the objection that a consequentialist system fails to respect criminals as responsible moral agents.

Retributivists hold that punishment must be deserved if it is to be justified, and that the guilty (and only the guilty) deserve punishment. Positive retributivists hold that the guilty should be punished as they deserve, even if this will achieve no consequential good. Negative retributivists hold that only the guilty may be punished, but that they should be punished only if their punishment will be beneficial. The main objection to retributivism is that it fails to explain why the guilty deserve punishment.

Some retributivists have argued that the guilty deserve censure, and that punishment serves to communicate that censure. But why should we use 'hard treatment' such as imprisonment or fines to communicate censure? Does the hard treatment function as a consequentialist deterrent? Or could such punishments serve to reform or educate criminals, thus bringing them to repent their crimes and restoring their relationships with those they have wronged?

A theory of justified punishment must be related to our existing penal institutions. It must, in particular, have something to say about sentencing: about what kinds of punishment should be imposed, and about how sentencers should decide on the appropriate severity of punishment. A central issue concerns the role of the principle of proportionality: the demand that the severity of punishment should be proportionate to the seriousness of the crime.

But we must also ask whether our existing penal practices can be justified at all. We must face the abolitionists' argument that punishment should be abolished in favour of social practices which treat 'crimes' not as wrongdoings that must be punished, but as 'conflicts' which must be resolved by a reconciliatory rather than a punitive process.

1 Punishment, the state and the criminal law

Our focus is on punishment imposed by the state for breaches of the criminal law. Punishment can be initially defined as the deliberate infliction of something meant to be burdensome, by an authority, on an alleged offender, for an alleged offence. It needs justification because it involves doing things (depriving people of life, liberty or money) which are normally wrong. Different moral perspectives, however, generate different accounts of why punishment is morally problematic and thus of what could justify it. Is what matters the infliction of pain, for instance; or the apparent coercive infringement of rights which punishment involves?

A justification of state punishment presupposes a normative theory of the state, as having the authority to punish. Different theories of the state generate different conceptions of punishment: a liberal theory, for example, might set more modest aims for state punishment, and subject it to stricter constraints, than would a communitarian theory (see Liberalism §5; Community and communitarianism §§2, 3).

A justification of punishment also requires an account of crime, since it is crimes that are punished. Crime can be minimally defined as socially proscribed wrongdoing, breaching an authoritative social norm (see Social norms). We require an account of the proper character and scope of such norms (and of what it is to be responsible for breaching them, since crime involves a criminal who can be held responsible for it) (see Responsibility §1). But not all breaches of socially (or legally) authoritative norms count as crimes which merit punishment: we must ask

what kinds of response are appropriate to different kinds of wrongdoing.

Censure is one proper response to breaches of authoritative norms, and the expression of censure may be a further defining feature of punishment; this distinguishes fines, for instance, from taxes (see §3). But censure can be expressed by formal declarations, or by symbolic punishments which are painful only in virtue of their expressive meaning, whereas criminal punishments typically inflict ‘hard treatment’ (the loss of liberty, money or life) which is painful independently of its expressive meaning. Why should such hard treatment be an appropriate response to socially proscribed wrongdoing?

Another response to such wrongdoing is the enforced payment of compensation to those harmed by it; this is a central feature of the civil as distinct from the criminal law. But though punishment may involve the payment of compensation, it also inflicts hard treatment that is not directly compensatory (nor do crimes always harm identifiable victims). Why should such punitive hard treatment ever be appropriate, and for what kinds of conduct? Which should count as crimes, rather than merely as civil wrongs?

Some theorists appeal to the ‘harm principle’ (see Law and morality §2): only conduct which harms or endangers others should be criminal. But this provides at most a necessary, not a sufficient, condition for criminalization: not every kind of (even seriously) harmful conduct is a plausible candidate for criminalization. And, apart from the question of whether paternalistic laws, prohibiting conduct that harms only the agent, can ever be justified, we must ask what counts as ‘harm’. Can we distinguish harmful from merely offensive conduct? Might we count some conduct as ‘harmful’ purely because of its moral character (as, for example, a breach of trust or a denial of rights) rather than because of its material effects?

Whether we talk of conduct that harms interests, or that infringes rights, or that flouts community values, we must ask which interests, rights or values should be protected by the criminal rather than the civil law. Crimes are often said to be public, rather than private or individual, wrongs: wrongs not just against some individual who may then claim damages, but against the community or state. That is why while civil cases are brought (and may be dropped) by individual plaintiffs, criminal cases are brought by the state or community, even when they involve an attack on an individual victim. But can we explain crimes as public wrongs, without distorting the way in which many crimes attack individual victims? To say, for instance, that murder and rape should be crimes not because of what they do to their particular victims, but because they threaten public order, seems to deny the significance of the victim’s suffering. We might suggest that even crimes against individual victims should count as ‘public’ wrongs in that the community should identify itself with the victim, counting the victim’s wrong as ‘ours’. Or we might abandon the idea of crimes as public wrongs (except for those which directly injure the collective rather than any individual, like tax evasion), and portray crimes as attacks on those central rights or interests which the state should protect. Either approach, however, leaves us with the question of which rights or interests should be thus protected, or which wrongs should thus be seen as public wrongs. Or if we say that the criminal law should protect the values which are essential to the identity or existence of the community, we must ask which those values are. (Any account of crime must also explain the distinction between *mala in se*, acts which are wrong independently of any legal rule, and *mala prohibita*, acts which are wrong only because prohibited. *Mala prohibita*, however, include many offences (notably ‘regulatory’ offences, such as minor traffic violations) which some think should not count as true ‘crimes’: they should be dealt with, not by a criminal process which censures and punishes, but by some distinct regulatory procedure.)

Instead of asking directly which kinds of conduct should be criminalized, we might ask what justifies criminal punishment, and found our principles of criminalization on our answer to that question. If the central justifying aim of punishment is deterrence, we can ask which kinds of conduct should be thus deterred; if its proper aim is ‘retribution’, we can ask which kinds of conduct merit such a retributive response.

2 Consequentialism and retributivism

Penal theory has long been a battleground between consequentialists and retributivists. After a period of consequentialist domination, the 1970s saw a revival in retributivist thought, as part of a wider rights-based reaction against consequentialism in social policy.

Consequentialists justify punishment by its instrumental contribution to certain goods: most obviously, the good of crime-prevention. A penal system is justified if its crime-preventive and other benefits outweigh its costs, and no

alternative practice could achieve such goods more cost-effectively. Punishment prevents crime by deterring, incapacitating or reforming potential offenders: by giving them prudential disincentives to crime, by subjecting them to restraints which make it harder for them to break the law, or by so modifying their attitudes that they will obey the law willingly (see Bentham, J. §3; Consequentialism).

It is at most a contingent truth that such effects on potential offenders are efficiently achieved by punishing actual offenders. This generates the familiar objection to any purely consequentialist theory, that it sanctions injustice. A system of deterrent punishment must appear to punish actual offenders: but that leaves open the possibility of framing innocent scapegoats to deter others or to reassure the public. And unless actual offending is the only reliable predictor of the future crimes which reformative or incapacitative measures aim to prevent, such measures might be efficiently (but surely unjustly) inflicted on those who have not yet broken the law but are thought likely to do so. Indeed, since a person's subjection to coercive treatment by the state must depend on the predicted effects of such treatment, rather than on their past conduct, we might wonder whether consequentialists can justify a system of punishment, of measures imposed for a crime, at all.

Some consequentialists do argue that we should replace punishment by other, more efficient methods of dealing with socially dangerous people. Others argue that we should accept the 'injustices' that a strictly consequentialist penal system might perpetrate (noting that we already accept, for instance, the pre-emptive detention of the mentally disordered). Most accept, however, that a justified system of punishment cannot perpetrate the kinds of gross injustice noted above.

Consequentialists might meet this objection by providing a fuller account of the goods to be achieved or protected, and of the methods by which they might practicably be achieved. Thus some argue that individual freedom is an essential good, whose protection precludes the deliberate punishment of those who have not voluntarily broken the law. Others argue that, given the fallibility of human agents, the only safe way to pursue the appropriate goods is to set strict constraints on the penal system: for instance, strictly to forbid the deliberate punishment of an innocent. Such consequentialist defences, however, depend on large empirical claims about the likely effects of penal strategies, which cannot easily be verified. Can the demands of justice to which this objection appeals really be adequately grounded in the contingencies on which this consequentialist argument depends?

Another strategy is to abandon pure consequentialism, and impose non-consequentialist side-constraints of justice on our pursuit of the consequentialist's goals: to insist, for instance, that only those who have voluntarily broken the law may be punished, since responsible agents have a right not to be subjected to such coercive measures unless they voluntarily make themselves liable to them.

One objection to even a side-constrained consequentialist theory concerns the moral standing of those who are punished or threatened with punishment: that a consequentialist system fails to respect its citizens (criminals and non-criminals) as responsible agents. A system of deterrent punishments, Hegel argued, treats all those whom it threatens with punishment like 'dogs': rather than seeking their allegiance to the law by appeal to the moral reasons which justify its demands, it coerces their obedience by threats (see Hegel, G.W.F. §8). A consequentialist system of reform similarly treats those subjected to it as objects to be remoulded, rather than as responsible agents who must determine their own conduct.

Against such objections, some argue that a side-constrained system of deterrent punishments can respect the moral standing of those it threatens and punishes; or that 'rehabilitation' and 'reform' need not be improperly manipulative or coercive. But one stimulus to the retributivist revival in the 1970s was the claim that only retributivism respects the moral standing of criminals: their right to receive 'fair and certain punishment', rather than being 'used merely as means' to the deterrence of others, or being subjected to indefinite terms of reformative 'treatment'.

The central retributivist slogan is that (only) the guilty deserve punishment, and deserve punishments proportionate to the seriousness of their crimes. This demand for 'just deserts' may be interpreted negatively, as forbidding the punishment of the innocent (or the excessive punishment of the guilty); or positively, as requiring that the guilty be punished as they deserve. The negative reading makes guilt a necessary, but not sufficient, condition of justified punishment: it suggests a 'mixed' account, which gives punishment a consequentialist aim but subjects our pursuit of that aim to retributivist side-constraints, requiring that punishment be both deserved and

consequentially beneficial. The positive reading makes guilt a necessary and sufficient condition of justified punishment: the guilty should be punished because they deserve it, whether or not their punishment achieves any consequential good (see Kant, I. §10).

The central task for any retributivist is to explain this supposed justificatory relation between guilt and punishment: what is it about crime that makes punishment an appropriate response to it? The central objection to all retributivist theories is that they fail to discharge this task: they either fail to explain this notion of penal desert, falling back on unexplained intuition or metaphysical mystery-mongering, or offer covertly consequentialist explanations.

The 'new retributivism' of the 1970s offered various accounts of the idea of penal desert. One was that criminals gained by their crimes an unfair advantage over the law-abiding, since they accepted the benefits of the law-abiding self-restraint of others, but evaded that burden of self-restraint themselves: their punishment removed that unfair advantage, thus restoring the fair balance of benefits and burdens which the law should preserve. One objection to this account is that it distorts the nature of crime: what makes rape punishable as a crime is surely the wrong done to its victim, not the unfair advantage the rapist supposedly takes over all those who obey the law.

Another trend in recent retributivist thought has rather built on the idea of punishment as an expressive or communicative practice.

3 Punishment and communication

Expressive accounts of punishment need not be retributivist: since by expressing censure we can modify wrongdoers' conduct, consequentialists can advocate expressive punishments. But the expressive or communicative aspect of punishment can explain the retributivist's slogan that the guilty deserve punishment: if they have broken a law which justifiably claimed their obedience, they deserve censure; and it is a proper task for the state, speaking on behalf of the community, to communicate that censure to them.

We should talk of communication rather than of expression here. For communication is a process which addresses (as expression need not) another as a rational agent; it captures the idea (central to recent versions of retributivism), that we must address criminals as rational and responsible agents.

But even if criminal wrongdoers should be censured, and hard-treatment punishments of the sort imposed by our penal systems can communicate that censure, we must ask why it should be communicated in this way, rather than by formal declarations or purely symbolic punishments (see §1).

Some suggest that hard treatment is necessary if the censure which wrongdoing merits is to be communicated effectively to the criminal, who might not attend to merely symbolic punishments; or that it may be necessary to 'defeat' the claim to superiority which was implicit in the wrongdoer's crime (but do all crimes make such a claim?). But why, if not for the consequentialist reason that this will make the punishment a more effective deterrent, is effective communication so crucial that we must inflict hard-treatment punishments to achieve it?

Others accept that a communicative retributivism cannot by itself justify the use of hard treatment as the communicative vehicle: it must be justified by a consequentialist concern for deterrence. This need not be the kind of 'mixed' account which portrays retributivist values merely as side-constraints on the consequentialist ends which give the penal system its positive aim. The communication of censure can itself be the central justifying aim of punishment, so that the law addresses the citizen as a responsible moral agent, appealing to the moral reasons which justify its demands and the censure that it imposes on those who flout those demands. But recognizing that, as fallible human beings, we will not always be adequately motivated by such moral reasons for obeying the law, we communicate that censure through hard treatment in order to provide an additional prudential incentive for obedience. On one version of this account, the hard treatment should provide only a modest prudential supplement which does not replace or drown the law's moral voice: the question then is whether such modest supplements will be effective. On another version, the hard treatment may be harsh enough to provide by itself an effective deterrent; but this will revive the objections noted earlier to a deterrent conception of punishment.

More ambitiously communicative accounts of punishment portray the hard treatment as a mode of moral communication which aims to reform or educate. Punishment aims to bring wrongdoers to understand and to repent their crimes, and thus to reform their future conduct. Hard treatment assists this purpose by helping to bring

home to them the meaning and implications of what they have done; it can also, if it is willingly undergone, enable them to express their repentance and thus reconcile themselves with their victims and the community. Such accounts are retributivist, since punishment must be focused on the past crime as an appropriate censuring response to it, but they also give punishment a forward-looking purpose: the offender's reform or rehabilitation, the restoration of the relationships which the crime damaged, the making of symbolic (and perhaps material) reparation to the victim and the community. Such purposes, however, are not to be understood in strictly consequentialist terms, as independent ends to which punishment is a contingent means: they can be achieved only through a punitive process which aims to persuade wrongdoers that they must suffer punishment for what they have done.

We must ask, however, whether hard-treatment punishment could ever be an appropriate vehicle for such a communicative, reformative and penitential endeavour; and whether, even if it could (as it might be in, for instance, a religious community that practises penance), the state should take such a coercive interest in the moral condition of its citizens. This conception of punishment might be at home within a communitarian perspective according to which individuals can find their identity and their good only as members of a community united by shared values and mutual concerns; but it seems incompatible with a liberal insistence on the need to protect individual rights and privacy against intrusive state or community power. Liberals can argue that punishment's primary purpose should be the communication of appropriate censure, but may deny that the state should try, by such coercive means, to secure repentance and reform; in which case hard-treatment punishments could be justified only as prudential deterrents which do not seek to invade the criminal's soul.

4 Penal theory and sentencing

Philosophical discussions of punishment are typically conducted at a level of high abstraction, remote both from the actualities of penal practices and from the pressing concerns of penal practitioners. But we must try to relate them to the real penal world.

One central issue is that of sentencing. What kinds of punishment should be available to the courts (capital punishment; imprisonment; fines; community service; probation)? What makes a particular kind of punishment appropriate, either generally or for a particular crime? How should sentencers determine the severity of punishment to be imposed on particular crimes or criminals?

Discussion of the last question often focuses on the principle of proportionality: the severity of punishment should be proportionate to the seriousness of the crime. Some such principle is integral to any retributivist theory, including communicative theories: for if punishment is to communicate an appropriate degree of censure, its severity must be proportionate to the seriousness of the crime. The application of such a principle requires some way of assessing and comparing the seriousness of different crimes, and the severity of different punishments; and it is not clear either just how, or how precisely, this can be done. Furthermore, while such a principle can help to determine the relative severity of sentences, requiring that more serious crimes be punished more severely, and so on, it is not clear whether it can help to fix absolute levels of punishment.

How important is the principle of proportionality? On some views, it is paramount: the primary aim of sentencing is to do justice by assigning proportionate sentences. This means, in practice, that the available range of punishments must be limited, and that the courts should have only very limited discretion in sentencing. Others argue that the demand for proportionality must be weighed against other relevant principles, such as a principle of penal parsimony which requires courts to impose the lightest acceptable sentence, even if that is lighter than is required for strict proportionality; on this view proportionality might be seen as a limiting principle requiring that criminals be punished no more severely than is proportionate to their crimes.

There is also a tension between the demand for proportionality and any ambitious account of punishment as communication. If punishment is given an educative, reformative or penitential aim, courts should seek punishments which are materially appropriate, rather than just formally proportionate, to the crime and the criminal: punishments which will appropriately address the particular criminal. But this would require the courts to be given a more flexible and creative discretion in sentencing, to find or construct sentences appropriate to the particular case: a discretion which might undermine demands for strict and formal proportionality.

Here again we face a conflict between a liberal perspective which emphasizes the demands of formal justice, and

seeks to protect the citizen against the coercive and discretionary power of the state; and a more ambitious conception of the proper role of the state and the criminal law in seeking the moral good of the citizens.

5 Can punishment be justified?

Any plausible normative theory of punishment will show our existing penal institutions to be radically imperfect. The kind and degree of suffering that they inflict cannot be plausibly portrayed as either consequentially cost-effective or retributively just, or well-suited to the aims of a communicative theory of punishment. Nor is it clear that the preconditions of justified punishment are satisfied in our own societies, especially if punishment is portrayed in retributive or communicative terms: can we truly say that most of those who are convicted by our courts have culpably flouted laws which justifiably claimed their allegiance, or that we (in whose name the law speaks) have the moral standing to censure them?

The radical imperfection of our existing penal institutions raises a serious question for any citizen. Should we accept those existing institutions (while also striving for their reform) as necessary to the prevention of yet greater disorder or injustice; or may we have to recognize that they perpetrate such serious injustice, or cause so much harm, that they cannot be justified at all?

The suggestion that, even if a practice of state punishment could in principle be justified, our existing penal institutions may lack any adequate justification, might seem frivolous: can we honestly argue that they should be abolished? But this is just what is argued by ‘abolitionists’, many of whom indeed argue that punishment cannot even in principle be justified: we should work not for the reform of our penal institutions, but for their abolition. Such arguments are not often considered in the philosophical literature, which tends to assume that the key issue is not whether, but how, state punishment can be justified; but they present a challenge that must be taken seriously.

Various themes run through abolitionist writings. One concerns the very concept of ‘crime’ as that which merits a punitive response: we should reconceptualize crimes as ‘conflicts’ that require resolution rather than punishment. Relatedly, we should ‘civilize’ our response to crime, favouring a civil law rather than a criminal model: rather than seeking ‘retributive’ justice by condemning and punishing those judged to have done wrong, we should seek ‘restorative’ justice by striving to reconcile the conflicting parties and (where necessary) negotiating reparation for whatever harm has been done. These themes are often accompanied by an advocacy of ‘informal justice’: rather than allowing the state to ‘steal’ conflicts from the individuals and local communities to whom they properly belong, we should look for informal, participatory modes of conflict-resolution. But punishment (the deliberate infliction of suffering) is never justified: neither as retribution (which is not a proper aim), nor as deterrence (which denies the moral standing of those who are threatened and punished). And while rehabilitative facilities may be offered to those who need and seek them, they can never properly be imposed on citizens.

Against such views it may be argued that some ‘conflicts’ involve the commission of genuine wrongs which should be condemned; that any morally acceptable ‘reconciliation’ must involve the recognition and acceptance of guilt by the wrongdoer (these considerations argue in favour of a communicative conception of punishment as censure); and that a society which truly forswore the whole coercive apparatus of criminal justice would be unable to protect itself or its members against seriously destructive wrongs and social disorder. We might imagine a more perfect society in which the kinds of hard-treatment punishment currently imposed would be unjustified, because unnecessary. We may agree that we now punish too much, too harshly, that our penal institutions do not serve the ends that punishment should serve, and that too often they inflict further suffering on those who are already seriously disadvantaged by the political and economic structures from which many of us benefit. Abolitionists forcibly remind us of these points; but this is not to agree that punishment can never be justified.

See also: Justice; Law, philosophy of

R.A. DUFF

References and further reading

Bianchi, H. and Swaaningen, R. van (eds) (1986) *Abolitionism: Towards a Non-repressive Approach to Crime*, Amsterdam: Free University Press.(A useful collection of abolitionist papers.)

Braithwaite, J. and Pettit, P. (1990) *Not Just Deserts*, Oxford: Oxford University Press. (A consequentialist account of punishment and criminal justice, setting individual freedom under the law as the aim to be pursued,

and focusing on punishment as ‘reintegrative shaming’.)
Christie, N. (1981) *Limits to Pain*, London: Martin Robertson.(An abolitionist critique of punishment.)
Duff, R.A. (1986) *Trials and Punishments*, Cambridge: Cambridge University Press.(A communicative account of punishment as penance.)
Duff, R.A. (1996) ‘Penal Communications: Recent Work in the Philosophy of Punishment’, *Crime and Justice* 20: 1-97.(A critical survey of penal philosophy from 1970-93; includes a full bibliography.)
Farrell, D. (1985) ‘The Justification of General Deterrence’, *Philosophical Review* 94: 367-94.(Argues that deterrence is consistent with proper respect for rights.)
Feinberg, J. (1970) ‘The Expressive Function of Punishment’, in J. Feinberg, *Doing and Deserving*, Princeton, NJ: Princeton University Press, 95-118.(A seminal discussion of punishment as expressive.)
Feinberg, J. (1984-7) *The Moral Limits of the Criminal Law*, Oxford: Oxford University Press, 4 vols.(An exhaustive discussion of the proper scope of the criminal law, focusing especially on the ‘harm principle’.)
Hampton, J. (1984) ‘The Moral Education Theory of Punishment’, *Philosophy and Public Affairs* 13: 208-38. (Punishment as moral education which respects the criminal’s autonomy.)
Hampton, J. (1992) ‘Correcting Harms versus Righting Wrongs: The Goal of Retribution’, *UCLA Law Review* 39: 201-44.(Punishment as communicative retribution which ‘defeats’ the wrongdoer.)
Hart, H.L.A. (1968) *Punishment and Responsibility*, Oxford: Oxford University Press.(A collection of influential papers, developing a side-constrained consequentialist account of punishment.)
Hirsch, A. von (1976) *Doing Justice: The Choice of Punishments*, New York: Hill and Wang.(An important early contribution to the revival of retributivism.)
Hirsch, A. von (1993) *Censure and Sanctions*, Oxford: Oxford University Press.(Punishment as censure, with hard treatment providing a prudential supplement; the role of the principle of proportionality.)
Hirsch, A. von and Ashworth, A .J. (eds) (1992) *Principled Sentencing*, Boston, MA: Northeastern University Press; Edinburgh: Edinburgh University Press.(A useful collection of materials on the principles of sentencing.)
Honderich, T. (1984) *Punishment: The Supposed Justifications*, Harmondsworth: Penguin, revised edn.(A critical discussion of the standard theories of punishment.)
Lacey, N. (1988) *State Punishment*, London: Routledge.(A communitarian, ‘mixed’ account of punishment as aiming to secure both individual autonomy and social goods.)
Morris, H. (1968) ‘Persons and Punishment’, *The Monist* 52: 475-501.(Criticizes ‘treatment’ conceptions of punishment, offers a ‘benefits and burdens’ account.)
Morris, H. (1981) ‘A Paternalistic Theory of Punishment’, *American Philosophical Quarterly* 18: 263-71. (Punishment as moral education and reform.)
Morris, N. and Tonry, M. (1990) *Between Prison and Probation*, New York: Oxford University Press. (Theoretical and practical issues of sentencing, especially in relation to non-custodial penalties.)
Murphy, J.G. (1973) ‘Marxism and Retribution’, *Philosophy and Public Affairs* 2: 217-43; repr. in *Retribution, Justice, and Therapy*, Dordrecht: Reidel, 1979, 93-115.(The ‘benefits and burdens’ theory of punishment; also discusses the problem of the gap between ideal theory and actual practice.)
Murphy, J.G. (1984) ‘Crime and Punishment’, in J.G. Murphy and J. Coleman, *The Philosophy of Law*, Totowa, NJ: Rowman and Allanheld, 113-65.(An introductory discussion of many of the issues covered in this entry, including the distinction between crimes and civil wrongs.)
Primoratz, I. (1989) *Justifying Legal Punishment*, Atlantic Highlands, NJ: Humanities Press.(A useful introductory text, criticizing utilitarian and ‘mixed’ accounts, arguing for retributivism.)
Walker, N. (1991) *Why Punish?*, Oxford: Oxford University Press.(Defends a firmly consequentialist theory of punishment.)

Criteria

The concept of criteria has been interpreted as the central notion in the later Wittgenstein's account of how language functions, in contrast to the realist semantics of the Tractatus. According to this later account, a concept possesses a sense in so far as there are conditions that constitute non-inductive evidence for its application in a particular case. This condition on a concept's possessing a sense has been thought to enable Wittgenstein to refute both solipsism and scepticism about other minds. There are powerful objections to this conception of criteria, which have led some philosophers to look for an alternative account of the role of criteria in Wittgenstein's later philosophy.

1 Criteria and scepticism

Norman Malcolm (1963) sets out the central argument by means of which Wittgenstein allegedly refutes solipsism and scepticism about other minds. Wittgenstein is held to expose the incoherence of the argument from analogy, and thereby to show that if we make the assumption that we know, for example, what pain is only from our own case, then we could never transfer the idea to others. This allegedly forces the defender of the view that 'pain' refers to a private object within a psychological realm into endorsing solipsism. Wittgenstein is then held to use the private language argument to show that solipsism is incoherent (see Wittgenstein, L. §13; Private language argument).

The solipsist has a concept of pain only if it makes sense to talk of identifying a sensation as a sensation of pain correctly or incorrectly. But what does it mean to say that the solipsist has correctly identified a given sensation as one of pain? The solipsist has neither a standard, a sample nor a customary practice of using the word 'pain' against which the inclination to apply the concept can be judged. There is nothing independent of this inclination that enables the solipsist to determine whether the application is correct or incorrect. But in so far as there is nothing to determine whether the application is correct or incorrect, the idea of using the word correctly or incorrectly makes no sense. And in so far as the solipsist lacks a criterion of correct application, the solipsist lacks a concept of pain.

Thus, the assumption that one knows what pain is only from one's own case is shown to lead to a contradiction: if we know about sensations only from our own case, then we cannot know about them even in our own case. It follows, Malcolm argues, that it is a condition of our being able to attribute sensations to ourselves that there should exist *criteria* on the basis of which we are necessarily justified in ascribing them to others. When these criteria are fulfilled, speakers do not merely have good inductive evidence that another is in pain (as the argument from analogy holds); they possess evidence that *establishes beyond question* that another is in pain.

This initially powerful rebuttal of scepticism is weakened, however, when Malcolm is forced to concede that there is a wide range of circumstances (being in a play, being hypnotized, pretending and so on) in which behaviour that would, in other circumstances, constitute criterial evidence for another's being in pain, does not count as fulfilling the criteria. Thus, what looks initially like an argument that establishes that there must be logically necessary and sufficient conditions for another's being in pain - that is, that there must be circumstances in which the behavioural evidence would make it a contradiction for the sceptic to deny that another is in pain - receives a fatal qualification. For given that a doubt may arise over whether the circumstances are ones in which criteria are satisfied (or really satisfied), it seems that the sceptic could always raise a question about any particular case. What looked like an argument to establish that our belief in another's pain is, on particular occasions, logically justified, is now reduced to the much weaker claim that there must be circumstances in which a speaker *accepts* that the criteria for another's being in pain are fulfilled.

There are a number of familiar objections to the argument that Malcolm attributes to Wittgenstein. First of all, it has been seen as amounting to little more than a version of verificationism, and as such it is thought not to warrant the status or the degree of significance that Malcolm ascribes to it. Secondly, it has been objected that the promise to provide a refutation of scepticism is undermined by the qualifications set out in the previous paragraph. It looked as if criteria were to enable us to show that our belief in another's pain is logically justified, but it turns out that the satisfaction of criteria on particular occasions is something that we must ultimately just accept, even though a doubt is, in some sense, still possible. The sceptic could be forgiven for regarding this as vindication,

rather than refutation, for there remains a gap between evidence and truth that must be filled by brute faith that our criteria are satisfied. At this point we seem justified in returning to Wittgenstein's texts to see if there is an alternative interpretation of his use of the concept of criteria.

2 Criteria and grammar

The principal concerns of Wittgenstein's later philosophy are the false pictures of language and of psychological phenomena that plague us in philosophy. These myths and false pictures arise, Wittgenstein believes, when we treat the phenomena that characterize our human form of life - language, consciousness, thought, action and so on - as things whose essence lies hidden from us. We both discover the emptiness of the pictures we construct, and achieve the understanding that we seek, through what Wittgenstein calls a 'grammatical investigation'. His idea of a grammatical investigation is characteristically rich. He describes such an investigation as one in which 'we remind ourselves of the kind of statement that we make about phenomena'. Because the essence of language and psychological phenomena lies open to view in the forms of our ordinary practice, it is through a careful description of our actual employment of expressions that we come to understand the nature of these phenomena. The understanding which a grammatical investigation achieves simply sets before us 'what already lies open to view', but in such a way that it no longer puzzles us or cries out for further explanation. Part of the difficulty in understanding Wittgenstein's later philosophy is that he does not believe that the understanding which a grammatical investigation achieves can be expressed in the form of a clear, unambiguous description of the structure and function of language in general, or of our psychological language game in particular.

It is possible to see Wittgenstein's concept of criteria not as a theoretical term of art but as linked in important ways with this idea of a grammatical investigation. Thus, asking for the criteria that ordinarily establish, in the sense of *identify*, something as an instance of a specific kind of thing (as a mistake, as expecting someone between 4.00 and 4.30, as understanding the order 'Add two' in the way that it was meant and so on) is one particular form of grammatical investigation. Other forms include asking how we teach someone a concept, asking how we would explain it, asking whether we would apply it in certain non-standard cases, asking whether certain facts being different would make it unusable, comparing it with a concept that Wittgenstein invents, and so on. Wittgenstein uses these forms of investigation to induce the clarified vision of the workings of our ordinary language games that he believes is essential to our coming to recognize that everything we need to understand language and psychological phenomena is already there before our eyes.

One of the principal lessons of the grammatical investigation of our ordinary criteria is that the criteria governing our concepts are much more complicated, our language games much more subtle and involved, than at first appears. In the case of psychological concepts this complexity is expressed, not only by the first-person-third-person asymmetry which characterizes these concepts (for example in the case of pain, I do not attribute pain to myself on the basis of observation, the concept of a mistake does not apply in the first-person case, and it makes no sense for me to doubt whether I am in pain), but also by an indeterminacy and uncertainty which Wittgenstein again takes to be a defining feature of our psychological language game. Wittgenstein believes that it is in part these grammatical features of our psychological concepts that prompts the philosopher to form a picture of the mental as an inner realm. For not only is the first-person-third-person asymmetry captured in the distinction between inner and outer, but the idea of the inner allows us to interpret the indeterminacy and uncertainty that is inherent in the language game as merely epistemic: the facts are there alright, but hidden beneath the surface that the other presents to the world.

Wittgenstein's grammatical investigation is directed at a complete overcoming of the philosophical myth of the inner. On the one hand, he tries to show that the picture of a determinate inner realm not only plays no role in our ordinary language game, but that it has no real application. His tactic here is not to try to refute the philosopher's claims, but to develop and explore them in such a way that we come to see, not only that no such thing is ever ordinarily in question, but that the picture of a private inner realm, which initially seems so clear to us, has no real content. He does not argue that no precise system of concepts linked, for example, with exact physiological readings of what is occurring in our brain and nervous system could never be constructed, but he tries to show us that such a system of concepts would not tell us what pain, or thought, or understanding really are; it would in no sense be a refinement of our system of psychological concepts, or constitute an explication of the phenomena that make up our form of life.

On the other hand, Wittgenstein's detailed exploration of the complex, infinitely nuanced language game that the process of acculturation initiates us into, gradually enables us to see how our language game actually functions, and to resist the temptation to misrepresent it in the myth of the inner. He works to overcome our sense that to abandon the idea of a private inner realm is to deny something vital, something without which we are mere machines. He uses the techniques of grammatical investigation to show that our philosophical picture of the inner is grounded in a fundamentally mistaken idea of the relation between our ordinary psychological concepts and the characteristic forms of movement, gesture and expression of the living bodies of humans and other animals. The division which we are mistakenly inclined to draw between an inner and outer realm is shown to be one that is actually grounded in the grammatical distinctions between psychological and non-psychological concepts.

Thus, we gradually come to see that our ordinary psychological language game is distinctive, not in that it describes a hidden realm of facts, but in its grammar. The grammar of our psychological language game reveals the nature of the facts it describes, and in recognizing the distinctiveness of this grammar, we recognize the distinctive form of the phenomena with which we are concerned. Thus, the first-person-third-person asymmetry which characterizes psychological concepts is not something that needs to be explained by (the empty) appeal to a special sort of fact (private facts); rather, this asymmetry itself reveals the distinctive form of mental phenomena.

Likewise, pretence, deceit, betrayal, as well as the ordinary dissemblings that constitute polite behaviour, are part of the form of our psychological language game, and not an unfortunate or incidental addition to it. These phenomena should not lead us to downgrade the 'outer evidence', but to recognize the subtlety and the complexity of the criteria we operate with. Sometimes we are sure of our judgment that another's feeling is genuine, sometimes anyone who is not a lunatic will share our certainty, but at other times we are uncertain and the judgment of different speakers may vary. What Wittgenstein tries to show is that this reflects the essential complexity and ambiguity of our criteria, and not the hiddenness of the facts; our uncertainty arises from the complexity and the subtlety of our relations with others and of the patterns that our psychological concepts require us to discern, and not from the indirectness of our evidence.

Wittgenstein's later philosophy of psychology clearly does constitute a concerted attack on the myth of the inner. This attack does not, however, take the form of a theory of how language functions that imposes on psychological concepts the requirement that what is inner is criterially linked with what is outer. The understanding of psychological phenomena that Wittgenstein's grammatical investigation offers is, in a sense, much more radical than this suggests. What we are gradually brought to see is that the nature of psychological phenomena is not hidden; that the evidence for their existence is not indirect, but involved and complex; that everything that we need to understand the essence of these phenomena is already there in the distinctive grammar of our language game, as this is revealed, for example, by the nature of the criteria with which we operate.

See also: Contextualism, epistemological; Other minds

MARIE McGINN

References and further reading

Albritton, R. (1959) 'On Wittgenstein's Use of the Term "Criterion"', *The Journal of Philosophy* 56: 845-57. (Explores the view that criteria are logically necessary and sufficient conditions.)

Cavell, S. (1979) *The Claim of Reason*, Oxford: Clarendon Press.(This provides the excellent criticisms of Malcolm and the best introduction to the anti-Malcolm view of criteria. See especially part 1, chapters 1 and 2.)

Malcolm, N. (1963) 'Wittgenstein's *Philosophical Investigations*', and 'Knowledge of Other Minds', in *Knowledge and Certainty*, Englewood Cliffs, NJ: Prentice Hall. (Sets out Wittgenstein's alleged refutation of solipsism and scepticism about other minds.)

McDowell, J. (1982) 'Criteria, Defeasibility, and Knowledge', *Proceedings of the British Academy*, 68: 455-79. (Provides good criticism of Malcolm's view.)

Wittgenstein, L. (1978) *Blue and Brown Books*, Oxford: Blackwell.(Pages 23-5 are especially important.)

Wittgenstein, L. (1992) *Last Writings on the Philosophy of Psychology: The Inner and the Outer*, Oxford: Blackwell.(Provides the most sustained treatment of the topic of criteria. But note that it is in the nature of Wittgenstein's philosophical method that remarks on criteria are spread throughout his works and that these remarks bear complex relations both with each other and with other topics discussed.)

Wright, C. (1982) 'Anti-Realist Semantics: The Role of *Criteria*', in G. Vesey (ed.) *Idealism Past and Present*,

Cambridge: Cambridge University Press.(Provides an alternative to Malcolm’s view.)

Critical Legal Studies

Critical Legal Studies first developed in the USA in the latter half of the 1970s. Drawing on the political inspiration of the contemporary New Left, it was an intellectual movement committed to radicalizing legal theory by bringing together US legal realism and modern European social theory. In so doing, it sought to provide a fundamental critique of the nature and place of law in modern capitalist society.

In its first phase, its main target was the liberal positivist theories of law that dominate Anglo-American jurisprudence. Such theories inform both the organization of the traditional legal curriculum and the nature of legal practice. By contrast, Critical Legal Studies saw law as based upon deeply contradictory premises, so that the orthodox positivist claim that law could be in principle rational and coherent was rejected in favour of the 'indeterminacy thesis'. Legal decisions were in truth a matter not of logical deduction but of choice. They could always go one way or the other. Ultimately, therefore, it was an open political decision made by a judge which determined a legal conclusion. The idea of the 'rule of law' operating above politics was rejected, but regarded as important in terms of the political legitimation function it served in Western societies.

While the name 'Critical Legal Studies' has a US provenance, a number of different critical legal projects can be identified. These projects reflect the broader character of the national traditions of which they are a part. European approaches, particularly the German and the British, reveal a more sustained engagement with modern and postmodern social theory. However, as a result of problems in the original project, Critical Legal Studies has entered into a second phase in the USA in which there is an increasing interest in social theory. The result has been a convergence of European and US concerns, but around a highly fragmented group of modern and postmodern social theories. Nietzsche, Foucault, Derrida and Habermas have been introduced into legal theory while Marx and Weber, the original theoretical mainstays of a critical approach to law, have been sidelined. There is a danger in this that Critical Legal Studies will become little more than a group of theorists talking among themselves. While the original US critique of legal doctrine may have run out of steam for want of sufficient theoretical sophistication, it is important not to lose sight of its direct focus on law and legal forms. It is arguable that the recent 'turn to theory' must validate itself in terms of the contribution it is able to make to a critical understanding of law and its practices; also that an important, as yet unaddressed, question concerns the relationship between postmodern forms of criticism and sociological analyses of the development of law.

1 Introduction

Critical Legal Studies as an approach to legal theory is a relatively recent phenomenon beginning in the second half of the 1970s and continuing to the present. It therefore comprises a body of thought that is hard to summarize or judge conclusively. Contributing to the problem is the variety of approaches within it, such that one can talk more easily of a 'movement' than a 'school' or a 'canon'. Notwithstanding the common New Left political heritage, the problem is further compounded by the existence of distinct Critical Legal Studies approaches in the USA and in a variety of west European countries, and by the dynamic, unstable and increasingly fragmentary character of their development. Critical Legal Studies in the early 1990s has a substantially different agenda and set of interests from that of a decade ago. In the light of these problems, it is proposed to begin with a picture of the trends of development in three different countries, the USA, Germany and Britain, before proceeding to offer a general assessment. With regard to this second objective, it is important to consider the relationship between Critical Legal Studies, legal studies in general, and broader movements in social theory. It is only in that context that a deeper perspective on the character, potential and limits of Critical Legal Studies can be achieved. A caveat must however be entered that in relation to so broad, disparate and currently evolving a movement, all generalizations and judgments are more than usually tentative.

2 Critical Legal Studies in the USA

A Critical Legal Studies movement first began in the late 1970s in the USA. The approach was characterized by the marriage of an existing legal realist tradition (see Legal realism §3) with a radical leftist antipathy to law which, while nominally committed to theoretical critique, was in general only weakly informed by an understanding of radical theoretical traditions. A recent survey of Critical Legal Studies (Boyle 1991) refers to the relevance of socialist, structuralist, deconstructionist, feminist, phenomenological and Hegelian theory. Critical

Legal Studies in the USA was always catholic in its sources, but never particularly deeply engaged with theory. It was more comfortable with a relatively narrow critique of the dominant legal tradition and its closer theoretical supports - economic analysis of law or a narrowly philosophical liberalism for example - than with a synthesis of the legal realist tradition and broader social theory (see Social theory and law §4). In recent times, the balance has been redressed through a deep interest in theory, but this has been accompanied by a move away from the earlier critiques of doctrinal analysis, in their time so powerful.

Critical Legal Studies in the USA directed itself against the dominant rationalist and formalist approach to legal studies. Legal realism had already introduced a sceptical critique of the possibility of coherence or consistency within a system of legal rules. Principles could always be opposed by counter-principles or by policies; rules were always subject to exceptions which threatened to swallow them up; the demand for fixed rules always existed in tension with broader standards which could be invoked against them; supposedly clear distinctions such as that between the private and the public were ultimately untenable. Realism's critique flowed from a dissatisfaction with formalistic and positivistic accounts of law (see Legal positivism §3), but its aim was pragmatic: to find a better way of comprehending law in order to put it to better use. Critical Legal Studies by contrast was informed by a more radical political standpoint that saw law not as an instrument to be used in the service of reform but as itself, in its very form, a barrier to social change. In its abstraction and pretence of rationality, law was essentially alienating and disempowering. The critique of law was therefore aimed at breaking the grip of traditional legal thinking upon the consciousness of lawyers and, in an unclearly defined way, 'the people' as a whole. In place of pragmatism, Critical Legal Studies substituted another peculiarly US intellectual current, populism.

A key concept (that was later to be repudiated) was that of a fundamental contradiction in a capitalist society - or, perhaps, all social life - between the individual and the social (Kennedy 1979). The conflict between self and others manifested itself in a variety of forms within law: in the clash of individualist and altruist principles, in the distinction between the private and the public (see Privacy), or between the subjective and the objective. Because this contradiction was fundamental, any attempt to reconcile its two sides was impossible, and legal rationalism was no more than a means of suppressing or hiding it. The existence of contradiction within law also explained the practical adversarial method of the lawyer, and illustrated the central contention of Critical Legal Studies that legal decision-making was essentially a matter of choice that could only be made according to political, as opposed to compelling internal, legal, criteria. The trick of the law was to make this political choice appear to be based upon apolitical rational grounds. The resulting 'legal consciousness' constituted the basis for legitimating forms of ideology in the broader society.

From this brief description, it will be seen that Critical Legal Studies contained a number of strands that did not necessarily add up to a coherent alternative perspective. That law and politics are intertwined is a claim that is not incompatible with more flexible varieties of legal positivism; nor is the argument about the fundamental contradiction within Western thought and law necessarily linked to it. With regard to the latter, it is important to know whether the contradiction is historical or existential. If it is a condition of all sociality or of an evolutionary modernity, then law will be regarded in a different light than if it is seen as a product of capitalist society with its own particular alienating forms. If the resulting legal consciousness is a significant mode of legitimation in the broader society, what are the mechanisms through which this occurs? The problem for Critical Legal Studies was that it remained unclear about the bigger questions that were necessarily inherent in its project. It could lead both to a call to the barricades and to a reconstructive legal project, both to a rejection of law in favour of a radical remaking of the world through informal communitarian modes of regulation and to a commitment to a revised liberal legalism (a 'super-liberalism' in R.M. Unger's 1986 account). It could also lead to individual nihilism, personal angst and intellectual incoherence as two hundred years of liberal legal thought was 'trashed' without much to put in its place.

In its earlier, stronger phase, Critical Legal Studies never moved from what Alan Hunt (1986) has called 'theory adoption' to 'theory construction'. It therefore found itself resting on insufficiently theorized premises that were hard to defend, and which left it vulnerable to counterattack. The 'fundamental contradiction' was eventually, somewhat cavalierly, renounced (Gabel and Kennedy 1984) while the precise scope of the 'indeterminacy thesis' has remained indeterminate. James Boyle (1991) has recently reflected upon the problems of transition from critical legal thought to social theory and of the eclectic quality of the former. Yet it is to be doubted whether his suggestion that Critical Legal Studies should be content with what he calls 'local theory' will assist it in reaching

beyond some of its impasses. This is not to deny the value of much recent work on law and culture, as well as critical race theory and feminist legal analysis (see Feminist jurisprudence), but the more specific direction of this often sophisticated critical work leaves broader issues necessarily unaddressed. The consequence has turned out to be a theoretical turn which frequently leads away from law and into an engagement with modern and postmodern trends in social theory. Compare in the area of feminist critique of law, for example, MacKinnon's (over-)direct analysis of the gendered character of law (1987) with the later, more sophisticated but abstract deconstructionist approach of Cornell (1991). As Goodrich (1993) comments, this route will ultimately be validated only if it leads to new and deeper understanding of law, as opposed to further, more *recherché*, repetitions and (mis)translations of Continental theory.

3 Critical Legal Studies in Germany and Britain

If US Critical Legal Studies in its first phase was reluctant to link theoretical analysis to the critique of legal formalism, the same cannot be said of the parallel German tradition. In Germany, *Politische Rechtstheorie*, the counterpart to Critical Legal Studies in the USA, coupled an interest in the critique of legal formalism with analysis based upon historical materialism, Enlightenment philosophy and post-Weberian sociology of law. Just as Critical Legal Studies drew upon the native critical tradition provided by US pragmatism, populism and legal realism, so German critical legal theory has drawn on the wider traditions of modern classical and post-classical European thought.

The major axes of German thought are provided by the Enlightenment commitment to a philosophical rationalization and legitimation of law, and by the Marxist materialist tradition. Modern debates within these two approaches are also substantially mediated by what might be called the 'Weberian problematic' which involves the post-classical conflict within specifically legal thought between formal and substantive justice. German critical legal thought in this regard has been much concerned with the problem of the 'materialization' of law in the twentieth century, and how legal theory between Hegel and Marx should orient itself to it.

Three primary routes can be identified. The first, fundamentally Marxist, route sought to derive forms of law and the state directly from the logic of capitalist economic development (Holloway and Picciotto 1978). Such an approach came to be regarded as ahistorical and aprioristic and fell into disfavour in the late 1970s, yet it retains a broad residual presence within critical legal thought. It is an approach that compares and contrasts with a second route developed in particular by Jürgen Habermas (§4), which seeks to understand law both as a 'control medium' for capitalism linked with strategic political and economic action, and as a universalistic form linked to the life world and rational communication (see Communicative rationality). In his earlier work on legitimation crises, Habermas (1976) used this emerging dualist perspective in order to analyse critically the social contradictions of law in the welfare state, but his more recent work reveals an endorsement of legal formalism, at the same time as it finds it difficult to recognize the concrete legal forms within which formalism could be instantiated (Habermas 1988). Habermas' work can be seen as an Enlightenment-inspired attempt to establish the possible legitimacy of legal institutions in a post-classical world that has rejected in theory and practice the possibility of a formal justice rooted in positive law.

A third route extends the biological conception of the 'autopoietic' (self-referential and self-sustaining) system to legal and other social systems. This theory responds to the multiple conflicts between formal and substantive law by embracing them and at the same time rejecting the role of law as legitimative. In this account of law, emphasis is placed upon the autonomous and self-reflexive character of law within a social system that is differentiated into heteronomous, self-operating sub-systems. These sub-systems are largely closed to each other, and operate according to their own languages and imperatives. They receive inputs from other sub-systems and provide outputs to them, but all such inputs and outputs must be translated into the language of the sub-system itself. Autopoiesis proposed a radical form of system closure, which has been used to explain the propensity to regulative failure built into legal systems, and to provide a critique of 'juridification' as a system-imperative of the legal sub-system. The problem for autopoietic theory is that, despite its critical rhetoric, it presents a view of law that is little different from that of the self-understanding of the lawyer. Its roots in biological metaphor and its aprioristic method lead it to a sociological view of law's functionality unrelated to social or historical development, teetering on the brink of solipsism. Autopoiesis has rejected the linkage between law and an immanent morality explored by Habermas and his associates, but it has not established a sufficiently critical purchase point from which to understand law. Its

view of law can in some respects be compared with that of Weber, for whom modern legal rationality was part of the 'iron cage' of bureaucracy which entrapped humankind, and from which there was little chance of escape (see Weber, M. §2). It might be said to reflect the proclaimed problems of modernity without possessing the critical resources to deal with them.

British Critical Legal Studies can be seen as characteristically resting somewhere between the USA and the German traditions. Compared to the latter, the British have largely been entrapped within an empirical and utilitarian intellectual tradition that has hardly encouraged a deep interest in or respect for the value of theoretical approaches to law. British Critical Legal Studies lack, the German adherence to a sense of 'schools of thought' or of the necessity of locating oneself within a particular theoretical tradition (other than the positivist, of course). While one should stress a common background in the Marxism of the 1970s, the British approach is now more eclectic, more inclined to draw according to interest or fashion from a variety of Continental theories. Compared to the USA, however, the British critical theorists reject the value of populist or pragmatic modes of thought, and are more deeply committed to theoretical engagement with law. By comparison, they benefit from a deeper engagement with theoretical traditions in the sociology of law which goes back to structuralist, form-derivational and feminist debates of the 1970s. The result has been that, in a comparatively small intellectual community with a common background, there has emerged a relative diversity of approaches of some depth to critical legal analysis.

While it is therefore difficult to speak of a British approach to Critical Legal Studies, a range of individual British critical approaches can be identified. To name three of the more recent authors, Peter Goodrich (1990) has drawn upon Nietzsche and Heidegger to establish a critical counter-aesthetics of law and judgment to weigh against English legal traditions; Peter Fitzpatrick (1992) has brought together a Derridean sense of myth and a critique of colonialism to establish what he describes as the mythical foundations of modern law; and Alan Norrie (1993) has combined form analysis, social history and critical realism to establish a praxiological account of criminal law and punishment. These works share themes, but their intellectual allegiances and methodologies - within the current critical canon - vary.

4 Theoretical overview

Critical Legal Studies in the USA, Germany and Britain are characterized by a wide range of intellectual sources which include elements of Marxism, deconstruction, legal realism and Neo-Kantianism (as in autopoietic theory), as well as attempts to recreate the Enlightenment project (such as that of Habermas). They have in common a political and intellectual desire to transcend the positivist and formalist tradition; but with their diverse starting points, is there any way to analyse Critical Legal Studies approaches save in terms of what they are against?

One possibility is to locate Critical Legal Studies within the 'dialectic of the Enlightenment' and in particular the fragmentation of the Enlightenment project after Hegel and Marx. The breakdown of the Hegelian synthesis between legal and rational thought, together with the impact of Marxism, led in the nineteenth century to a threefold division in thinking about law and philosophy. What followed was a complex development with many overlapping strands, but one can identify schematically three theoretical approaches in order to locate Critical Legal Studies. The first is the development of positivist modes of thought which discarded the need for an external, overall system of philosophical reason with which to legitimate law's operation (the positivist approach) (see Positivism in the social sciences). Second, within ethical philosophy and critical theory, rationalist thought was subjected to increasing scrutiny and subversion (the approach of philosophical critique) (see Critical theory). Third, with Marxism, a historical and materialist method was developed that sought to understand both law and reason as expressions of historically developed social relations (the approach of sociological critique) (see Socialism). These three different approaches to social theory and law constitute the parameters within which Critical Legal Studies operate.

Drawing on the second and third approaches, Critical Legal Studies provides a critique of the idea that law and legal reasoning can be understood independently of broader social, moral and political questions, so that the positivist developments of the nineteenth century (the first approach), together with their modern derivatives, are a primary target. It is this target that gives Critical Legal Studies what unity it has; thereafter, the adoption of any critical perspective is a rather 'free' matter of picking from the range of critical resources available within Marxist and post-Enlightenment political philosophy (the second and third approaches). Because of its leftist political interests and its roots in 1970s social movements, Critical Legal Studies has always drawn significantly upon

Marxism, but more recently there has been a greater emphasis on the method of deconstruction and the work of French post-structuralists such as Foucault, Levinas, Derrida and Lyotard. While the relationship between post-structuralism and Marxism is controversial, these critical approaches have in common a desire to relate law to an ‘outside’ or a ‘beyond’ and to deny its autonomy. In Derrida (1990), the assertion of ‘the other’ as the beyond of law is explicit, while in Marxism, the external is to be found in the social relations of which law is a specific part, but which it also helps to constitute. In contrast to these trends, however, it should be noted that the German approaches as represented by Habermas (at least in his more recent (1988) work; see also Murphy 1989) and autopoietic theory seek to establish a critical basis ‘inside’ the law. Indeed in the latter case it is questionable how much more autopoiesis offers than a sociological variant of the Neo-Kantian positivist tradition in European legal philosophy.

The dialectic of the Enlightenment is both theoretical and historical. The radical rationalism of the Enlightenment became conservative in the work of Hegel, and then collapsed in the context of the ‘well ordered bourgeois state’. Legal positivism derived its strength from precisely this order, which was in important respects juridical. The role of theory came to be the ‘rational reconstruction’ (MacCormick 1993) of actual legal relations and practices, and legal positivism found its place within society and the academy as a result (Cotterrell 1983). Yet positivism was always vulnerable. Its linkage to the rationalization of a particular social practice (the law) created a tension with a range of social theories for which such a link was politically and intellectually problematic (see Legal reasoning and interpretation §3). Critical legal theory brings to bear a critical perspective on legal thought by exploiting this tension between positivistic legal-academic practice and broader social, moral and political concerns. It operates in the conflictual space between the three approaches established by the fragmentation of the Enlightenment project. So doing, it is able to expose legal thought to broader traditions in philosophy and sociology in ways that are at their best creative and critical. It remains on an ‘edge’ of legal-academic theory and practice and is vulnerable to changes in political and intellectual fashion. None the less it provides those who are justifiably dissatisfied with the narrow orthodoxy of traditional legal thought with some critical tools to go beyond the taken-for-granted reality of the mainstream.

5 The future

It is fitting to conclude an account of a new theoretical approach to law with consideration of its possible future directions. I discuss this in the Anglo-American context with reference to two reviews of Critical Legal Studies, one by an original US participant, the other by a recent British commentator.

Mark Tushnet (1991) offers a bleak picture of the current state of Critical Legal Studies in the USA. While he recognizes the need for a credible theoretical programme, he argues that Critical Legal Studies does not in fact have any essential intellectual components. Indeed he recommends that it be treated as a political rather than theoretical location for a left-leaning diaspora of feminists, critical race theorists, postmodernists, cultural radicals and political economists (see Feminist jurisprudence). What is more, the indeterminacy thesis at the heart of the original movement has now been to an extent both abandoned by its proponents and accepted by the mainstream. The gap between the two sides is now mainly one of degree. The only route forward that Tushnet suggests is a return to the classical social theory of Marx and Weber, but the recommendation is lukewarm. Yet if the indeterminacy thesis is partly abandoned by the critics and partly accepted by the mainstream, it is surely the question of how legal systems achieve determinacy under particular social and political conditions that becomes relevant. It is the historical analyses of classical social theory which consider such conditions and recognize the social and political significance of formalist and positivist concepts of law. These approaches therefore hold out the possibility of protecting Critical Legal Studies against absorption into a revised but still orthodox positivism. Tushnet outlines the theoretical process of engagement and disenchantment with Marx and Weber undergone by US critical theorists, but his account only confirms the pragmatic and populist nature and limits of that engagement. It therefore urges by implication a less half-hearted reappraisal of what such theories have to offer a critical legal project.

An alternative orientation to critical legal theory is provided by Peter Goodrich (1993) from within the present current of post-structuralist theory. He argues that critical thought ‘should investigate the conditions of possibility of the text and its interpretation’ (1993: 232). This involves, first, the use of history to reconstruct the intellectual development of the doctrinal tradition, leading to a recognition of the cultural specificity of forms of law, and a

critique of their modes of tradition, representation, repetition and reproduction. Second, such a historical perspective focuses on systems of classification, on conceptual grids and schemata, and on the forms of knowledge which pass as law. This represents a genuine programme for critical legal theory, but Goodrich's account raises questions about how it can be achieved. In particular, the relationship between the social history of legal forms and their critical interpretation remains problematic. In the concrete case of contract law, for example, Goodrich separates the 'historical' from the 'interpretive' stage of his critique. At the historical stage, he cites the existing institutional legal histories of contract, while the interpretive stage draws on literary theory or discourse analysis.

The result is an antinomian form of critique in which the first stage is undertheorized from a critical standpoint, and the second stage is abstractly theorized in a way that is unintegrated with the first. The critical punch of the second stage is provided by Goodrich's discussion of Carole Pateman's feminist criticism (1986) that the political social contract is also a sexual contract that marginalizes women, but nothing is said about the specifically legal institution discussed by the historians of contract. No doubt something could be said to link the two, but what is missing from, or marginalized within, Goodrich's account is any consideration of contract as a social phenomenon along the lines outlined by Marxism and Weber. Such accounts, which are nowadays sometimes treated as *passé*, none the less provide a substantial social and historical context within which the discourses of lawyers and legal historians can be read, and an important starting point for their critique. The point is not negatively to oppose classical social theory to postmodernism, but rather to ask how they can be fruitfully articulated, as surely they ought to be, in any analysis that seeks genuinely to be historically critical.

Treating these two reviews as representative of its current state of development, two dangers confront critical legal theory. The first is that, undertheorized, it will lose its distinctiveness and disappear into the mainstream. The second is that, abstractly theorized, it will lose the grip on law as a specific social and political phenomenon that Critical Legal Studies originally promised, but ultimately failed to deliver. A way beyond both these dangers is provided by a retrieval and integration of, but not an uncritical return to, the traditions of classical social theory represented by Weber and Marx (see Marxism, Western). If Critical Legal Studies does not address these in a more constructive way than it has done over the last decade it risks one of two fates. Either it will dissolve into the orthodox positivist tradition it sought to criticize, or it will lose itself in the abstract ethical theorizing that constitutes a large part of postmodernism.

See also: Critical theory; Law, philosophy of; Postmodernism; Social theory and law

ALAN NORRIE

References and further reading

Boyle, J. (1991) 'Editor's Introduction', *Critical Legal Studies*, Aldershot: Dartmouth.(Useful collection of previously published US Critical Legal Studies essays.)

Cornell, D. (1991) *Beyond Accommodation: Ethical Feminism, Deconstruction and the Law*, London: Routledge. (Important but difficult attempt to provide an ethical basis for feminist legal theory.)

Cotterrell, R. (1983) 'English Conceptions of the Role of Theory in Legal Analysis', *Modern Law Review* 46: 481.(Analyses the role of positivist legal theory in rationalizing and legitimating the work of the legal profession.)

Derrida, J. (1990) 'Force of Law: The "Mystical Foundation of Authority"', *Cardozo Law Review* 11: 919-1045. (Derrida's accessible discussion of law and justice from a deconstructionist standpoint.)

Douzinas, C. and Warrington, R. (1995) *Justice Miscarried*, Brighton: Harvester Wheatsheaf.(Application of Levinas to legal studies.)

Fitzpatrick, P. (1992) *The Mythology of Modern Law*, London: Routledge.(Critical analysis of Western law as 'white myth' (Derrida): as founded on racist and nationalist mythologies.)

Gabel, P. and Kennedy, D. (1984) 'Roll Over Beethoven', *Stanford Law Review* 36: 1-55.(Zany dialogue illuminating the intellectual and existential concerns of two leading critical legal scholars in the USA. The entire volume is devoted to Critical Legal Studies.)

Goodrich, P. (1990) *Languages of Law: From Logics of Memory to Nomadic Masks*, London: Weidenfeld and Nicolson.(A complex and sophisticated postmodern critique of the history, symbols and language of the English common law.)

Goodrich, P. (1993) 'Sleeping with the Enemy', *New York University Law Review* 68: 389-425.(British commentary on the current state of US Critical Legal Studies.)

Habermas, J. (1976) *Legitimation Crisis*, London: Heinemann.(Sophisticated development of the concept of

legitimation in modern capitalist society beyond Weber's conception of legal-rational legitimacy.)
Habermas, J. (1988) 'Law and Morality', ed. S.M. McMurrin, *Tanner Lectures on Human Values*, vol. 8, Cambridge: Cambridge University Press, 219-79.(Argues for 'a rationality of legislative and judicial procedures guaranteeing impartiality': a position which appears to place Habermas within mainstream liberal thought.)
Holloway, J. and Picciotto, S. (1978) *State and Capital: A Marxist Debate*, London: Edward Arnold.(English translation of the German debate which had a considerable contemporary impact.)
Hunt, A. (1986) 'The Theory of Critical Legal Studies', *Oxford Journal of Legal Studies* 6: 1-45.(An early attempt to survey Critical Legal Studies in the USA by a British Marxist.)
Joerges, C. (1989) '*Politische Rechtstheorie* and Critical Legal Studies: Points of Contact and Divergencies', in C. Joerges and D. Trubek (eds) *Critical Legal Thought: An American-German Debate*, Baden-Baden: Nomos.(Useful comparison of US and German critical legal approaches.)
Kelman, M. (1987) *A Guide to Critical Legal Studies*, Cambridge, MA: Harvard University Press.(An insider's analysis and reconstruction of the Critical Legal Studies movement in the USA.)
Kelman, M. (1979) 'The Structure of Blackstone's Commentaries', *Buffalo Law Review* 28: 205.(A seminal Critical Legal Studies essay in the USA.)
MacCormick, D.N. (1993) 'Reconstruction after Deconstruction: Closing in on Critique', in A. Norrie (ed.) *Closure or Critique: New Directions in Legal Theory*, Edinburgh: Edinburgh University Press.(Attempts to defend legal positivism against the Critical Legal Studies attack.)
MacKinnon, C. (1987) *Feminism Unmodified*, Cambridge, MA: Harvard University Press.(Argues that men exploit women through their sexuality, and that law operates to support this.)
Murphy, W.T. (1989) 'The Habermas Effect: Critical Theory and Academic Law', *Current Legal Problems* 42: 135-65.(Salutary reminder of the critical distance - that is not always recognized by 'critical' lawyers - between 'system' and 'life world' in Habermas' project.)
Norrie, A. (1993) *Crime, Reason and History*, London: Weidenfeld and Nicolson.(A critical analysis of English criminal law, focusing on the tensions and contradictions in its general principles.)
Pateman, C. (1986) *The Sexual Contract*, London: Routledge.(A feminist critique of the exclusion of women from the social contract in Western liberal political philosophy.)
Teubner, G. (1992) *Law as Autopoietic System*, Oxford: Blackwell.(Teubner's statement of the autopoietic theory of law.)
Tushnet, M. (1991) 'Critical Legal Studies: A Political History', *Yale Law Journal* 100: 1515-44.(Retrospective account of the history and politics of the Critical Legal Studies movement in the USA by an original member.)
Unger, R.M. (1986) *The Critical Legal Studies Movement*, Cambridge, MA: Harvard University Press.(A virtuoso 'manifesto' for the Critical Legal Studies movement in the USA by one of its leading figures.)

Critical realism

Critical realism is a movement in philosophy and the human sciences starting from Roy Bhaskar's writings. It claims that causal laws state the tendencies of things grounded in their structures, not invariable conjunctions, which are rare outside experiments. Therefore, positivist accounts of science are wrong, but so is the refusal to explain the human world causally. Critical realism holds that there is more to 'what is' than 'what is known', more to powers than their use, and more to society than the individuals composing it. It rejects the widespread view that explanation is always neutral - to explain can be to criticize.

Bhaskar originally called his general philosophy of science 'transcendental realism' (see Bhaskar 1978) and his philosophy of the human sciences 'critical naturalism' (see Bhaskar 1979). The term 'critical realism' arose by elision of these two phrases, but Bhaskar and others in this movement have accepted it since 'critical', like 'transcendental', suggests parallels with Kant's philosophy, while 'realism' indicates the differences from it (see Kant, I. §5).

Critical realism starts from the transcendental question 'how are informative experiments possible?', which includes the question 'why are they necessary?'; that is, why can we not rely on spontaneous observations, but have to set up test-situations artificially, and how can these artificial tests disclose what happens spontaneously outside the laboratory? Bhaskar's answer is that experiments set up, as near as possible, 'closed systems', defined as systems in which something like Humean causation works: whenever A happens B happens.

In closed systems, particular mechanisms of nature can be isolated and studied accurately, but elsewhere in the open systems of which the world is made up, a multiplicity of these mechanisms co-determine events. Experimentally closed systems are of special value to science because they uncover mechanisms of nature, but these mechanisms are only of interest because they also operate in those open systems of which almost all the world consists, outside artificially contrived experimental conditions. Astronomy alone finds approximations to closed systems in nature, due to the lack of forces capable of deflecting heavenly bodies from their courses.

On this basis, critical realism postulates three kinds of 'depth' in nature, and identifies three common fallacies which arise from ignoring them.

(1) Not only may one distinguish, as common-sense realism does, between the contents of experience ('the empirical') and the actual course of events ('the actual'); one may also distinguish 'generative mechanisms' in nature, which are real even if not actualized. Gravity is real, even when the roof is not falling in. Failure to recognize this leads to 'actualism', the attempt to locate laws at the level of the actual (spontaneously occurring constant conjunctions). As against this, laws should be analysed as tendencies: bodies tend to persist in a state of rest or uniform motion in a straight line; hoppy beer tends to make you sleepy; capitalist enterprises tend to get bigger and fewer.
(2) The real multiplicity of natural mechanisms grounds a real plurality of sciences which study them. Even though one kind of mechanism may be explained by another (for instance, biological mechanisms by chemical ones), it may not be reduced to the other. Nature is stratified, with some strata rooted in but emergent from others; the course of events is different from what it would be if the more 'basic' strata alone existed. So the attempt to reduce human to natural or biological to physico-chemical science is ruled out by the nature of the subject-matter.
(3) Science is a social product, the result of experimental work; but the mechanisms it discovers operate in nature prior to and independently of their discovery. So we must distinguish the 'transitive object' of science - the knowledge it has produced - and its 'intransitive object' - the reality it is about, which exists independently of our knowledge, and always has unexplored depths to it. When this is not recognized, we fall into the 'epistemic fallacy', reducing questions about what is to questions about what we can know. Concepts like 'the empirical world' encapsulate this fallacy.

Several conclusions may be drawn for the human sciences.

(1) Most controversies between positivistic and hermeneutic views assume, on both sides, a positivist account of natural science (see Positivism in the social sciences §1). Therefore, positivists expect the human sciences to find constant conjunctions in the human world, though they are scarce enough in the natural; and hermeneuticists

conclude from the absence of such conjunctions that the human sciences are radically unlike the natural ones (see Social science, contemporary philosophy of §§1-3). A critical realist account of science makes possible a non-reductive naturalism in human science.
(2) Closed systems cannot even be established artificially in the human sciences. So experimental method is irrelevant here. More to the point are the detective-like skills of those natural scientists who study open systems: geologists, natural historians, meteorologists.
(3) Since we are humans, we have in our conceptualized social practices a ready-made starting point for the human sciences. This is the truth in hermeneutics. But there are no grounds either for regarding these data as infallible or for treating their operation as non-causal. Once we reject Humean causality and recognize emergence, we can accept that reasons for actions can be their causes - but also that they may be rationalizations of actions whose causes are elsewhere. Human sciences may be interpretive and non-reductive, but at the same time causally explanatory and corrective of agents' conceptions (psychoanalysis, Chomskyan linguistics and certain versions of Marxism seem to fit this model best) (see Chomsky, N.; Social science, history of philosophy of §12). This makes them potentially emancipatory, in that they can correct enslaving illusions (as psychoanalysis claims to), or illusions that subserve social oppression (compare Marx on ideology). In this connection, Bhaskar introduces the notion of an explanatory critique: some beliefs are incompatible with their own true explanation, and since this can apply to moral beliefs, the fact/value gap is bridged.

Critical realist interventions in debates about particular disciplines in the human sciences include Pateman (1987) on linguistics, Will (1980, 1984) on psychoanalysis, Collier (1989) on Marxism, and Soper (1995) on ecology.

See also: Experiment; Experiments in social science; Explanation in history and social science; Naturalism in social science §1

ANDREW COLLIER

References and further reading

Archer, M. (1995) *Realist Social Theory: the Morphogenetic Approach*, Cambridge: Cambridge University Press. (An elaborated defence of the mutual irreducibility of social structure and intentional agency, as propounded by critical realism.)

Bhaskar, R. (1978) *A Realist Theory of Science*, Hemel Hempstead: The Harvester Press.(Basic text on transcendental realist philosophy of science.)

Bhaskar, R. (1979) *The Possibility of Naturalism*, Hemel Hempstead: The Harvester Press. (Basic text on critical naturalist philosophy of the human sciences.)

Bhaskar, R. (1986) *Scientific Realism and Human Emancipation*, London: Verso.(Ramifies the above theories, introduces the concept of explanatory critique, and concludes with a critique of positivism. Very difficult reading.)

Bhaskar, R. (1989) *Reclaiming Reality*, London: Verso.(Essays on various matters and at various levels. The best Bhaskar text to read as an introduction or a sampler.)

Bhaskar, R. (1991) *Philosophy and the Idea of Freedom*, Oxford: Blackwell.(Mainly a critique of Rorty, with some comments on critical realism and Marxism.)

Bhaskar, R. (1993) *Dialectic*, London: Verso.(Further developments of critical realism, focusing on the concepts of negativity and totality. Very difficult reading.)

Bhaskar, R. (1994) *Plato Etc.*, London: Verso.(An essay in the critical realist resolution of the traditional problems of philosophy. Difficult reading.)

Collier, A. (1989) *Scientific Realism and Socialist Thought*, Hemel Hempstead: The Harvester Press.(Suggests critical realist solutions to some problems of Althusserian Marxism.)

Collier, A. (1993) *Critical Realism*, London: Verso.(General introductory exposition of critical realism.)

Lawson, T. (1997) *Economics and Reality*, London: Routledge.(A critique of contemporary economic theory from a critical realist standpoint.)

Manicas, P. (1987) *A History and Philosophy of the Social Sciences*, Oxford and New York: Blackwell.(An account of the development of social scientific theories from a critical realist point of view, concluding with a realist critique of empiricism.)

New, C. (1996) *Agency, Health and Social Survival*, London and Bristol, PA: Taylor and Francis.(A readable critical realist comparison and critique of various theories in psychology and psychoanalysis.)

Pateman, T. (1987) *Language in Mind and Language in Society*, Oxford: Clarendon Press.(Critical realist

interventions in philosophical disputes about linguistics. Some prior knowledge of Chomsky's theories helps.)
Soper, K. (1995) *What is Nature?*, Oxford: Blackwell.(Untangles various concepts of nature, from a standpoint which is at once critical realist, feminist and ecological.)
Will, D. (1980) 'Psychoanalysis as a Human Science', *British Journal of Medical Psychology*, 53: 201-11.(Argues that psychoanalysis conforms to the best current model of science, that is, critical realism.)
Will, D. (1984) 'The Progeny of Positivism: The Maudsley School and Anti-Psychiatry', *British Journal of Psychotherapy*, 1 (1): 50-67.(Uses critical realism to defend psychoanalysis from positivist and anti-scientific criticism.)

Critical theory

The term 'critical theory' designates the approach to the study of society developed between 1930 and 1970 by the so-called 'Frankfurt School'. A group of theorists associated with the Institute for Social Research, the School was founded in Frankfurt, Germany in 1923. The three most important philosophers belonging to it were Max Horkheimer, Theodor Wiesengrund Adorno and Herbert Marcuse.

Horkheimer, Adorno and Marcuse feared that modern Western societies were turning into closed, totalitarian systems in which all individual autonomy was eliminated. In their earliest writings from the 1930s they presented this tendency towards totalitarianism as one result of the capitalist mode of production. In later accounts they give more prominence to the role of science and technology in modern society, and to the concomitant, purely 'instrumental', conception of reason. This conception of reason denies that there can be any such thing as inherently rational ends or goals for human action and asserts that reason is concerned exclusively with the choice of effective instruments or means for attaining arbitrary ends.

'Critical theory' was to be a form of resistance to contemporary society; its basic method was to be that of 'internal' or 'immanent' criticism. Every society, it was claimed, must be seen as making a tacit claim to substantive (and not merely instrumental) rationality; that is, making the claim that it allows its members to lead a good life. This claim gives critical theory a standard for criticism which is internal to the society being criticized. Critical theory demonstrates in what ways contemporary society fails to live up to its own claims. The conception of the good life to which each society makes tacit appeal in legitimizing itself will usually not be fully propositionally explicit, so any critical theory will have to begin by extracting a tacit conception of the good life from the beliefs, cultural artefacts and forms of experience present in the society in question. One of the particular difficulties confronting a critical theory of contemporary society is the disappearance of traditional substantive conceptions of the good life that could serve as a basis for internal criticism, and their replacement with the view that modern society needs no legitimation beyond simple reference to its actual efficient functioning, to its 'instrumental' rationality. The ideology of 'instrumental rationality' thus itself becomes a major target for critical theory.

1 Historical background

In 1923 a group of intellectuals in Frankfurt, Germany, founded an 'Institute for Social Research' as a centre for the interdisciplinary study of social and economic issues in contemporary society from a broadly socialist perspective. The term 'critical theory' is used to designate the approach to social theory developed by the members of the Institute (who came to be known collectively as 'The Frankfurt School') between 1930 and 1970. The philosophers most closely associated with the genesis of critical theory were Max Horkheimer, Theodor Wiesengrund Adorno and Herbert Marcuse. When the National Socialists seized power in Germany in 1933 the Institute moved to New York, where it remained until 1949. Much of the most original work of the Frankfurt School was produced during this period of exile in the USA.

In their early papers, Horkheimer and Marcuse outline the conception of an interdisciplinary social theory that would be guided by an interest in the normative goal of human emancipation. Translations of the most important papers from this period are collected in *Critical Theory* (Horkheimer 1968) and *Negations* (Marcuse 1968).

Although Horkheimer, Adorno and Marcuse took themselves to be rationalists (of a sort), in the 1940s they found themselves becoming increasingly sceptical of the Enlightenment assumption that scientific and technological progress is an unproblematic human good. Empirical science, they thought, was based on a form of rationality - 'instrumental rationality' - which was inherently manipulative and which would have disastrous social and moral consequences if not strictly subordinated to a more encompassing notion of rationality. The 'critique of instrumental reason' is elaborated most fully in The *Dialectic of Enlightenment*, written jointly by Adorno and Horkheimer and published in 1947. The 1940s also saw the completion of Adorno's *The Philosophy of Modern Music* (1949) and, in *Minima Moralia* (1951), his reflections on the impossibility of leading a good life in the contemporary world.

In 1949 Horkheimer negotiated a return of the Institute to Frankfurt. Marcuse, however, elected to remain in the

USA. During the 1950s and 1960s he held positions at various US universities and wrote the two books on which his reputation as a theorist of the New Left mainly rests: *Eros and Civilization* (1955) and *One-Dimensional Man* (1964).

Upon their return to Frankfurt Adorno and Horkheimer became prominent public figures, while continuing to practise their form of cultural criticism. In 1958 Horkheimer retired to Switzerland where he died in 1973. In 1966 Adorno published a lengthy philosophical account of his version of the critical theory, *Negative Dialectics*. He seems to have been surprised by the student movement of the 1960s. Many participants were inspired by his writings and thought of themselves as putting critical theory into practice, however Adorno repudiated their activism. In contrast, Marcuse attempted to the very end of his life to maintain a theoretical commitment to action for radical social change.

2 Conception of ‘theory’

Members of the Frankfurt School agreed in rejecting a widely held view about what a ‘theory’ is. In normal parlance a ‘theory’ is a set of formally specified and interconnected general propositions that can be used for the successful explanation and prediction of the phenomena in some object domain. This conception of theory, the members of the Frankfurt School argued, is extremely misleading because it directs attention away from the social context within which theories necessarily arise, are tested and are applied, and within which alone they are fully comprehensible. The term ‘theory’ should be used in the first instance to designate a form of (ideally social) activity with an especially salient cognitive component, and only derivatively for the propositions that might be formulated in the course of such activity. Human societies are engaged in a constant process of assimilating nature through labour in order to reproduce themselves; they develop forms of cognitive activity in order to make this self-reproduction more secure and more efficient. Horkheimer, Adorno and Marcuse call such cognitive activity ‘traditional theory’. Virtually everything we would in normal parlance call a theory (including all scientific theories) is what the members of the Frankfurt School would call a derivative form of traditional theory. However, they would claim there is another possible kind of cognitive activity, one which is not directed at reproducing society in its present form or making its assimilation of nature more efficient, but rather is directed at changing the existing society radically so as to make it more substantively rational. Essentially it is an attempt to do away with those fundamental features of the society that prevent agents from being able to lead a good life. ‘Critical theory’ is the name given to such inherently oppositional forms of thinking. In a derivative sense the propositions or specific theses brought forward by agents engaged in such oppositional thinking at some particular time may also be called critical theory. A critical theory thus is historically specific, being directed at a particular society that stunts the possible realization of the good life. It is inherently negative, and it depends on a conception of substantive reason.

3 Critique of instrumental reason

Critical theorists argue that in the ancient world the concept of ‘reason’ was an objective and normative one. Reason was thought to refer to a structure or order of what ought to be which was inherent in reality itself and which prescribed a certain way of life as objectively rational. Human beings were thought to have a (subjective) faculty which allowed them to perceive and respond to this objective structure of the world; this faculty could then also be called reason in a derivative sense. Even when ancient philosophers spoke of reason as a human faculty (rather than as a structure of the world), their conception of it was ‘substantive’; humans were thought to be able to use reason to determine which goals or ends of human action were worthy of pursuit.

In the post-Enlightenment world the ‘objective’ conception of reason becomes increasingly implausible. Reason comes to be conceived as essentially a subjective ability to find efficient means to arbitrarily given ends; that is, to whatever ends the agent in question happens to have. The very idea that there could be inherently rational ends is abandoned. Reason becomes subjective, formal and instrumental.

The historical process by which reason is instrumentalized is in some sense inevitable and irreversible. The philosophical position called ‘positivism’ draws from this the conclusion that reason itself should simply be identified with the kind of reason used in natural science. Scientific reason, the critical theorists claim, is a particularly highly developed form of instrumental reason. The point of getting an exact depiction of reality as it is and of the causal laws that govern events is to allow humans to manipulate the world successfully so as to attain

their ends. For this to be possible, the positivists believe, the terms that figure in significant scientific discourse must be clearly defined and their relation to possible confirming or disconfirming perceptual experience must be clearly specified. Reason, the positivists think, can be a guide to life only in a very limited sense. Its role is restricted to discharging three tasks: (1) it can criticize a set of beliefs and ends for failing to satisfy certain minimal principles of logical consistency; (2) it can criticize a given choice of means towards a given end on a variety of possible empirical grounds, such as that the means in question will not actually lead to the envisaged end or will have undesirable side effects, and it can propose more appropriate means; (3) it can unmask inherently non-cognitive beliefs, for instance value judgments, that are presenting themselves as if they had cognitive content. The role of reason in discharging the third of these tasks is especially important in the view of the positivists because any statements that do not belong to the descriptive and explanatory apparatus of science, and in particular any statement about what ought to be the case, stand wholly outside the domain of rational argumentation and can be nothing but the expression of arbitrary choice or personal preference (see Positivism in the social sciences).

Critical theorists think that this line of argument is seductive, but dangerous and false. Although it is true that reason cannot directly prescribe some set of ends as inherently rational and worthy of pursuit for their own sake, it retains an essential function that goes beyond those the positivists would allow it. This further function is that of ‘internal’ or ‘immanent’ criticism.

4 Internal or immanent criticism

To understand the Frankfurt School’s doctrine of internal criticism it is necessary to see it in the context of their general conception of society. Horkheimer, Adorno and Marcuse reject all four components of the positivist approach to the study of human society:

1 the view that human societies are just bundles of separate facts, events and institutions (‘atomism’);
2 the view that social facts and institutions are what they are regardless of what people think of them (‘objectivism’);
3 the view that the concepts we use to give an account of a society are just tools which we are free to define in whatever way seems convenient (‘nominalism’);
4 the view that the concepts we form should be purely descriptive, that is, defined exclusively in terms of purely observable properties with no evaluative component (‘the fact/value distinction’).

Contrary to this, members of the Frankfurt School hold that every society is a ‘totality’ in which each feature is essentially connected with all others and that social reality is partly constituted by the forms of belief, understanding and evaluation that exist in the society in question. Thus atomism and objectivism are false. Furthermore, the actual practices and institutions in a society would make no sense unless they were seen as inherently oriented towards the realization of some socially specified conception of the good life. Thus the more naïve versions of the fact/value distinction are problematic for the study of societies. To use the favoured Hegelian terminology: each institution in a society must be seen in relation to its (objective) ‘concept’, an ideal form of itself which it aspires to approximate, thereby playing its assigned role in the realization of the good life. The concept of an institution is objective in that the agents who participate in the institution need not be fully and explicitly aware of it - it cannot be determined by simply reading off the contents of their beliefs. It is also objective in that if we as researchers fail to formulate the concept correctly, we will have misunderstood the institution in question fundamentally. Thus ‘nominalism’ in the study of society is wrong, too. Although the concept of a social institution is objective in this sense, it is not, of course, objective in the sense in which a natural phenomenon is; it is *finally* constituted by human subjects and their activities, although this process of ‘constitution’ may be a very highly mediated one, lasting through a long historical time. To discover the concept is a very complex, constructive, theoretical activity requiring the social philosopher to enter into the history of the institution and study the ways in which people in the past understood it, the hopes, aspirations and values that were associated with it and their consequences.

A critical theory elicits the concept of a given institution in a given society, formulates it and confronts the actual reality of the institution with this ideal concept. There will be a discrepancy which the critical theory will point out and analyse. This analysis can be called internal or immanent criticism because the standards used in it are derived from the concept of the institution itself. Use of this internal method allows criticism to proceed without it being

necessary for the critic to have an unconditional commitment to or to give an independent justification of the standards of criticism used. The common view held by all the critical theorists until 1969 (the year of Adorno's death) was that substantive reason in the modern world must remain relentlessly negative: one cannot extract from reason the image of a good society or indeed any unconditional set of positive ideals. Reason cannot describe utopia, but can at best specify a 'determinate negation' of some particular feature of contemporary society that has been subject to internal criticism. Just as Marx does not describe the socio-economic formation he believes will succeed capitalism in positive and detailed terms, but only in negative terms as a 'class-*less* society', so similarly the most Adorno is ever willing to assert is that it would be desirable to be able 'to be different without *angst*' (1951: 656). In 1969 Marcuse announced (in the 'Introduction' to *An Essay on Liberation*) that he was breaking with the prohibition on utopian speculation that had been an integral part of the original critical theory; his post-1969 views will be discussed below (see §7) (see Utopianism).

5 The dialectic of enlightenment

Horkheimer and Adorno held that the final framework for social criticism had to be a speculative philosophy of history. In *The Dialectic of Enlightenment* (1947) they set out to provide the global interpretation of human history they believe is necessary.

Human history is a dialectic of 'enlightenment' on the one hand and 'myth/barbarism' on the other. 'Enlightenment' as used by Adorno and Horkheimer means both: (1) a certain theory - that is a specification of goals for society, a set of views about individual morality, the nature of knowledge, rationality and so on; and (2) the actual state of society which results from the massive application of this theory. 'Myth' is the opposite of enlightenment in sense (1); 'barbarism' the opposite of enlightenment in sense (2).

Enlightenment as a theory comprises five central tenets.

1 Commitment to certain ideals; autonomy, individuality, non-coercion, human happiness and so on.
2 Commitment to the view that 'genuine' knowledge is knowledge that is (a) 'objectifying', based on a clear and strict separation between the human subject and nature (as object of knowledge); (b) 'identifying', subsuming individual instances (of whatever it is that is to be known) under unitary general concepts and thus presenting them as in some sense 'the same'; (c) inherently technologically or instrumentally efficacious.
3 Commitment to the view that an increase in genuine knowledge in a society would lead to a greater realization of the ideals of the Enlightenment (as formulated in 1 above).
4 Commitment to a principle of universal criticism; that is nothing is to be taken on faith or authority or because of ' tradition' but every belief must show a warrant that will be recognized by the Tribunal of Reason.
5 Commitment to seeing itself (that is, to seeing enlightenment) as absolutely different from and opposed to barbarism and myth.

Horkheimer and Adorno have a number of criticisms of this theoretical position. First of all they claim that item 5 is false. The enlightenment is wrong to see itself as utterly and radically different from myth. Rather the relation between myth and enlightenment is a 'dialectical' one. Both have a common origin as reactions to the same phenomenon: primeval terror. Human history in fact is nothing but a series of attempts to deal with our overwhelming fear of what is other or unknown. Myth arises from a mimetic reaction to this terror: by making ourselves like that which we fear, by identifying with it, we attempt to do away with its otherness, as primitive hunters might try to deal with their fear of a predator by mimicking its movements in a dance. Another way to react to fear of the unknown is by separating it strictly from the self and subjecting it to a system of identifying categories the better to keep track of it and perhaps eventually control it. This second reaction is that of enlightenment: a rigid fixation on self-preservation as the absolute overriding goal and an incipiently paranoid concern to classify everything so as to be able to subordinate it to the attainment of that goal. Looked at, then, from sufficient distance enlightenment and myth seem similar. Both are attempts to use a form of identity to deal with the *angst* induced by difference. The direction (as it were) of 'identification' runs, however, in the opposite direction in the two cases. In myth we make ourselves like the other; in enlightenment we try to make the other like our category (by subsuming it). Close inspection of myths moreover reveals them to be historically superseded forms of enlightenment. The pantheon of Olympian gods - archetype of mythic thinking in the West - is not just a creation of the mimetic impulse, but must also be seen as a form of enlightenment relative to the religious beliefs

and practices associated with the nameless, chthonic deities of pre-Homeric Greece. What counts as myth and what as enlightenment is not given absolutely once and for all, but is historically relative.

Adorno and Horkheimer's second criticism is that the enlightenment tends to overlook or downplay the price humanity has had to pay for enlightenment. The instrumentally manipulative attitude the enlightenment bids us adopt towards nature will necessarily tend to extend itself to our relations with our fellow humans. Furthermore, effective instrumental control of nature requires that I control and finally repress my own spontaneity. Spontaneity is, however, an essential part of a human's capacity for happiness. Modern subjectivity itself is a result of this process of enlightenment, in which self-preservation is ensured at the cost of impairing our capacity for happiness.

The third line of criticism is that enlightenment has an inherent tendency to destroy itself. The original substantive ideals of the enlightenment (autonomy, individuality and so on) are not themselves exempt from the demands of the principle of universal criticism; that is, from the requirement of giving an account of themselves before the Tribunal of Reason. It becomes increasingly clear, however, that it is not possible to argue from the results of empirical science or from some more general principles of instrumental reason (that is, from some bit of what the enlightenment itself would consider 'genuine knowledge') to the validity of those ideals. In the end the ideals themselves come to look like myths or prejudices which ought to be discarded. The Marquis de Sade, a legitimate child of the enlightenment, finds no 'rational' arguments against cruelty (and much to be said in its favour).

Fourth, the history of the twentieth century has shown that the increase in technological control over the world and the spread of scientific knowledge does not in fact necessarily make people more autonomous, more highly individuated or more happy.

Their final criticism starts from the rigidity and paranoia of the enlightenment project, from its need to encompass *everything* in a single, definitive, closed system of concepts. This means that enlightenment is potentially totalitarian and has an inherent tendency to absolutize itself; that is, (falsely) to declare itself to be not just a given historically relative stage in the global process of enlightenment, but rather the final and definitive form of enlightenment. In thus absolutizing itself and resisting change, the given stage of enlightenment turns itself into a form of myth. This transformation is the final stage of the 'dialectic of enlightenment'.

Despite this battery of criticisms Horkheimer and Adorno do not reject the enlightenment outright. Rather they see their task as one of furthering the underlying enlightenment project by enlightening the enlightenment about itself. By analysing its inherent tendency towards totalitarianism, they hope to save its ideals (even if only in negative form) and prevent it from turning itself into a form of myth and barbarism (see Totalitarianism §4).

6 Negative dialectics

In his *Negative Dialectics* (1966), Adorno tries to give a philosophical account of the kind of thinking that is constitutive of critical theory: a form of substantive reason which is however purely negative.

Both everyday thinking and science, he claims, proceed by subsuming particulars under general concepts. By doing this, however - by claiming that *this* (tree) is a tree and that *that* (tree) is (also) a tree - we are tacitly asserting an identity between the two individuals and between each individual and the concept. To engage in such 'identity thinking' is to be tacitly trying to make identical what is in fact in some sense different - no two trees are exactly alike. Using an identifying concept is a way of trying to crush or suppress difference. The appropriate form of resistance to this reprehensible project is to remain aware of 'non-identity'; that is, of the ways in which instances are not identical with the concepts under which they are subsumed (and with each other). 'Negative dialectics' tacks back and forth between concept and instances, continually pointing out in what concrete ways they are not identical. Such a negative dialectics is a kind of cognition of the non-identical, although the process of moving back and forth negatively between concept and instance has no natural stopping point and will not ever result in some positive, detachable conclusion or new, more adequate concept.

Adorno rejected the usual standards of clarity and communicability for philosophical writing, seeing in them forms of repression, ways of preventing novel thoughts from being thought. He thus consciously wrote in an elusive, convoluted style, and claimed that (his) philosophy could not be summarized. This view makes good sense if one thinks of philosophy as essentially a concrete attempt to specify a 'determinate negation' that cannot in principle be turned into anything positive.

7 Conclusion

The philosophy of art had always played an important part in Adorno's thinking. Art and philosophical reflection on it form one of the few remaining oases for the play of free, spontaneous, human subjectivity in an increasingly regimented world. By the end of the 1960s, however, it had become hard to see how such reflection on high art, negative dialectics or the meditative essays on religion, pessimism and the philosophy of Schopenhauer Horkheimer had begun writing could be seen as part of the 'self consciousness of a revolutionary process of social change', or indeed how the late form of critical theory could be connected with any kind of action at all.

Toward the end of the 1960s Marcuse tried to break out of this impasse by giving up the traditional self-imposed ban among the members of the Frankfurt School on giving a 'positive' theory of any kind or engaging in utopian speculation (see §4). In *An Essay on Liberation* (1969) and *Counter-Revolution and Revolt* (1972) Marcuse claimed that the modern world had brought into existence a 'new sensibility' which he saw expressed most clearly in the emerging student counter-culture. This 'new sensibility', with its demand for aesthetically satisfying forms of immediate experience and its refusal to participate in consumer society, represented a significant new political force in the world. The social change necessary to accommodate the 'new sensibility' had become a vital individual need, and so one could even speak of a 'biological foundation for socialism'.

Since the mid-1960s a group of younger philosophers, most notably Jürgen Habermas, have tried to develop further some of the central components of critical theory. Although there are many similarities between the work of this second-generation of Frankfurt philosophers and the programme of critical theory, there are also some striking and important differences. In a sense the second generation marks a return to the kind of Neo-Kantian philosophy the critical theorists of the 1930s were reacting against. Adorno in particular was uncompromising in his opposition to the idea that philosophy should consist of a closed system of interconnected propositions that rested on a purportedly firm foundation and claimed universal validity. In the work of Habermas and his associates, however, the Kantian themes of finding a fixed universal framework for theorizing, giving firm foundations for knowledge claims of various sorts, and investigating the conditions of the possibility of various human activities, structure much of the discussion.

In retrospect the most important contribution of critical theory to philosophy in the late twentieth century would seem to be their criticism of positivism and their demand that social theory be reflective; that is, that theorists try to be as aware as possible of their own position, the origin of their beliefs and attitudes, and the possible consequences their theorizing might have on what they are studying.

See also: Frankfurt School

RAYMOND GEUSS

References and further reading

Zeitschrift für Sozialforschung (Journal of Social Research) (1932-42).(The house organ of the Institute for Social Research in which the members of the Frankfurt School published their most important theoretical work during the 1930s and the very early 1940s. Most of the essays translated in Horkheimer's *Critical Theory* (1972) and in Marcuse's *Negations* (1968) originally appeared here. The journal has a complicated publishing history. In 1932 it began being published by Hirschfeld Verlag, Leipzig. Between 1933 and 1940 it was published (in German) by Librairie Alcan, Paris. In 1939 it changed its name to *Studies in Philosophy and Social Science* and was published (in English) by The Institute for Social Research, Morningside Heights, New York City. The last issue appeared in 1942. In 1980 the Deutscher Taschenbuchverlag, Munich photomechanically reproduced all of the issues of the *Zeitschrift für Sozialforschung/Studies in Philosophy and Social Science* in a set of nine paperback volumes.)

Adorno, T. (1949) *The Philosophy of Modern Music*, trans. A.G. Mitchell and W.V. Blomster, London: Sheed & Ward, 1973.

Adorno, T. (1951) *Minima Moralia*, trans. E.F.N. Jephcott, London: New Left Books, 1974.

Adorno, T. (1966) *Negative Dialectics*, trans. E.B. Ashton, New York: Seabury Press, 1973.

Adorno, T.W. and Horkheimer, M. (1944) *Dialectic of Enlightenment*, trans. J. Cumming, New York: Herder & Herder, 1972.

Benhabib, S. (1986) *Critique, Norm and Utopia*, New York: Columbia University Press.(An especially full discussion of the Hegelian and Marxist background.)

Geuss, R. (1981) *The Idea of A Critical Theory*, Cambridge: Cambridge University Press. (Discusses concepts of 'ideology' and 'real interests' and the epistemology of a critical theory.)

Held, D. (1980) *Introduction to Critical Theory*, Berkeley and Los Angeles, CA: University of California Press. (Best general introduction.)

Horkheimer, M. (1947) *The Eclipse of Reason*, New York: Oxford University Press.

Horkheimer, M. (1968) *Critical Theory*, trans. M.J. O'Connell, *et al.*, New York: Herder & Herder, 1972. (Translation of the more important of Horkheimer's essays from the 1930s, all originally published in *Zeitschrift für Sozialforschung*.)

Jay, M. (1973) *The Dialectical Imagination*, Boston, MA: Little, Brown.(Historical account of the Institute for Social Research.)

Marcuse, H. (1955) *Eros and Civilization*, Boston, MA: Beacon Press.

Marcuse, H. (1964) *One-Dimensional Man*, Boston, MA: Beacon Press.

Marcuse, H. (1968) *Negations*, trans. J.J. Shapiro, Boston, MA: Beacon Press.(Contains translations of the most important essays from the 1930s originally published in *Zeitschrift für Sozialforschung*.)

Marcuse, H. (1969) *An Essay on Liberation*, Boston, MA: Beacon Press.

Marcuse, H. (1972) *Counter-Revolution & Revolt*, Boston, MA: Beacon Press.

Theunissen, M. (1969) *Gesellschaft und Geschichte (History and Society)*, Berlin: Walter de Gruyter.(Best philosophical analysis.)

Croce, Benedetto (1866-1952)

The leading Italian philosopher of his day, Croce presented his philosophy as a humanist alternative to the consolations of religion. A Hegelian idealist, he argued that all human activity was orientated towards either the Beautiful, the True, the Useful or the Good. These ideals were the four aspects of what, following Hegel, he termed spirit or human consciousness. The first two corresponded to the theoretical dimensions of spirit, namely intuition and logic respectively, the last two to spirit's practical aspects of economic and ethical willing. He contended that the four eternal ideals were 'pure concepts' whose content derived from human thought and action. Spirit or consciousness progressively unfolded through human history as our ideas of beauty, truth, usefulness and morality were steadily reworked and developed.

Croce insisted that his idealism was a form of 'absolute historicism', since it involved the claim that all meaning and value evolved immanently through the historical process. He strenuously denied that spirit could be regarded as some form of transcendent puppet-master that existed apart from the human beings through which it expressed itself. He accused Hegel of making this mistake. He also maintained that Hegel's conception of the dialectic as a synthesis of opposites had paid insufficient attention to the need to retain the distinct moments of spirit. He argued that the Beautiful, the True, the Useful and the Good, though linked, ought never to be confused, and he criticized aestheticism and utilitarianism accordingly.

Croce developed his thesis both in philosophical works devoted to aesthetics, ethics, politics and the philosophy of history, and in detailed historical studies of Italian and European literature, culture, politics and society. Opposition to the Fascist regime led him to identify his philosophy with liberalism on the grounds that it emphasized the creativity and autonomy of the individual. In practical politics, however, he was a conservative.

1 Life and works

Born in 1866 at Pescasseroli, in the Abruzzi in Southern Italy, Croce's family were wealthy landowners; Croce himself never had to pursue a career. In 1883 his parents and sister were killed in an earthquake in which he too was buried. For the next three years he lived in Rome with his uncle, Silvio Spaventa, a prominent liberal statesman of the Cavourian party known as the Historical Right and brother of the prominent Hegelian philosopher Bertrando Spaventa, whose own political thinking also drew on Hegelian themes. Although Croce never took a degree, he attended lectures at Rome university, where he came under the influence of Antonio Labriola who introduced him to the writings of Johann Friedrich Herbart and later Karl Marx. Croce's political views reflected the conservative liberalism of his uncle, but he claimed never to have been attracted by Bertrando Spaventa's brand of Hegelianism or his doctrine of the ethical State. Herbart's Neo-Kantian position, in contrast, inspired his later theory of concept-formation, while his early studies of Marx led him to his view of the Useful as a category distinct from the Good, towards which most practical activity was aimed (see Herbart, J.F.; Marx, K. §11).

Resisting family pressure to enter the law, Croce left Rome in 1886 and settled in Naples, where he determined to become a private scholar. Initially he devoted his energies to numerous antiquarian studies centred on various aspects of southern Italian history in the sixteenth, seventeenth and eighteenth centuries. This research formed the basis of his later books on the Kingdom of Naples and the Baroque era in Italy. This period culminated in his study of Neapolitan theatre from the Renaissance to the Enlightenment, published in 1891. The fusion of cultural and social history in these early writings was highly innovative and contained the germs of many of his later philosophical preoccupations - particularly with respect to the nature of human creativity and action.

These theoretical concerns began to come to the fore when in 1893 he published his first essay in philosophy, 'La storia ridotto sotto il concetto generale dell'arte' (History Subsumed under the General Concept of Art). Croce used an article by the Italian positivist Pasquale Villari entitled 'Is History a Science?' as an excuse both to attack the positivism then dominant in Italy and to enter the contemporary German debate between Windelband and Dilthey over the identity of the human sciences (see Dilthey, W. §3; History, philosophy of §4). Croce argued that while history was like art in representing particular events rather than elaborating general laws, as in the natural sciences, it differed in dealing with what actually occurred, rather than with what might have happened. Prompted by Labriola, he followed up this essay with a number of articles attacking crude quasi-Darwinian materialist interpretations of Marxism, although he went beyond his brief to criticize Marx's economic doctrines as well.

These writings, collected together in book form in 1900 as *Materialismo storico ed economia marxista (Historical Materialism and the Economics of Karl Marx)*, brought him to the attention of an international public including Georges Sorel and Vilfredo Pareto, whose interpretation of Marxism as a secular religion he came to share. His friendship with Giovanni Gentile, then working on a doctoral dissertation on Marx's philosophy, also began at this time (see Gentile, G. §1). Their collaboration over the next twenty years was to transform Italian philosophy, producing a revival of Hegelian Idealism.

In 1903 Croce began to publish his bimonthly journal *La critica*, devoted to reviews of the latest Italian and European books in the humanities and including general surveys of Italian literature and philosophy since unification written by Croce and Gentile respectively. Croce's *Estetica come scienza dell'espressione e linguistica generale (Aesthetic)* had appeared in 1902. This book became the first volume of his *Philosophy of Spirit*, and was followed by his *Logica come scienza del concetto puro (Logic)* (1905, completely reworked in 1909), the *Filosofia della pratica. Economia ed etica (Philosophy of the Practical)* (1909), and the *Teoria e storia della storiografia (Theory and History of Historiography)* (1917). Prompted by Gentile, Croce's philosophy took an increasingly historicist direction from 1905 onwards, a shift that led to the rewriting of the *Logica*). This development also prompted him to write detailed studies of Hegel, *Cio che è vivo e cio che è morto nella filosofia di Hegel (What is Living and What is Dead in the Philosophy of Hegel)* (1907), and of Vico, *La filosofia di Giambattista Vico (The Philosophy of Giambattista Vico*) (1911).

Croce became a life senator in 1910 but only began to write on and participate in politics with the advent of First World War, when he campaigned for Italy to remain neutral. He acted as Minister of Education in the last cabinet of the liberal politician Giovanni Giolitti from 1920 to 1921, putting forward reforms in schools and higher education that he and Gentile in particular had long advocated and which were finally implemented in the Fascist *Riforma Gentile*. Sympathetic to Fascism as long as he felt it was controlled by the old liberal elite as a bulwark against socialism, he never subscribed to its ideology and broke with it after 1924. Relations with Gentile, already strained due to his aversion to his friend's extreme subjectivist 'actual' idealism, finally collapsed with the latter's entry into the Fascist Party, of which he became the self-styled philosopher. Croce became a leading opponent of the regime, penning a famous protest against Gentile's *'Manifesto of Fascist Intellectuals'* in 1925. Opposition gave him a new lease of life and led him to characterize his philosophy as a form of liberalism - most particularly in two series of essays: *Etica e politica (Politics and Morals)* written largely in 1924-8 and collected in 1931, in which he put forward his ethico-political theory of history, and *La storia come pensiero e come azione (History as the Story of Liberty*) (1938), where he defined history as the 'story of liberty'. He also illustrated these theses in four historical studies: *La storia del regno di Napoli (A History of the Kingdom of Naples)* (1925), *Storia d'Italia dal 1871-1915 (History of Italy 1871-1915)* (1928a), *Storia dell'eta barocca in Italia (A History of the Baroque Era in Italy)* (1929) and *Storia d'Europa nel secolo XIX (History of Europe in the Nineteenth Century)* (1932). He made corresponding changes to his aesthetic doctrine, culminating in *La poesia (Poetry)* in 1936. Only the allied invasion caused him to stop publishing *La critica*, and he continued to produce books and essays and occasional *Quaderni della critica* right up to his death in 1952.

2 Aesthetics

Croce's aesthetics were tied both to his activity as a literary critic, which was copious, and developments in his general philosophy. He confessed to having no appreciation of music and wrote comparatively little on architecture or the figurative arts, leading some commentators to suggest that his theory is biased towards poetry. As the major influences on his own thinking, he claimed Francesco De Sanctis, whose *La storia della letteratura italiana (History of Italian Literature)* (1870-1) he continued through to the late nineteenth and early twentieth centuries, and Giambattista Vico, whose aesthetic thought he first brought to prominence.

Commentators commonly identify four phases within Croce's aesthetic doctrine, roughly corresponding to changes in his philosophy, although opinions differ over their compatibility with each other. The first phase, associated with the *Aesthetic* of 1902, consists of Croce's identification of art with intuition. His main targets were positivist or empiricist theories, on the one hand, and intellectualist or rationalist views on the other. Against the first, he claimed that feeling and emotion have to be expressed to have any existence, and that this is a cognitive process. Against the second, he distinguished logical from intuitive knowledge. Whereas the former works through concepts and deals with universal relations between things, the latter is obtained through images of particular

things. Adopting a Herbartian reading of Kant, he argued that we organize the world of experience and sensation through the intuitive faculty of the imagination that gave expression to them. Form and content, intuition and expression were identical. That which was not expressed had not been intuited and remained ‘a mere natural fact’. The logical categorization of this intuitive knowledge was a subsequent stage. He also contended that the theoretical activity of the imagination was distinct from any practical or ethical purpose. True art, therefore, was concerned neither with the True, the Useful or the Good - an argument he deployed against the Italian *verismo* school of naturalist writers influenced by Emile Zola.

The second phase of Croce’s aesthetics came with his theory of the lyrical nature of art. First enunciated in an article of 1908, and subsequently elaborated in the *Breviario di estetica (The Essence of Aesthetic)* (1913), this doctrine attempted to specify further just what artistic expression consisted of: namely the lyrical intuition of ‘intense feelings’ or emotions. Such feelings, he still insisted, could not be described in sensationist terms, but could only be expressed through images. For analogous reasons, he also rejected romantic theories that treated art as no more than a spontaneous outpouring of emotion. Art is a spiritual activity that transforms our bare animal existence. However, he continued to distinguish the images of intuition from the categories of logical thought, and he criticized the artificial canons and rules of classical theorists for ignoring this fact.

While distinct from either philosophy or practice, art was related to them. Drawing on his contemporaneous revision of Hegel and the resulting thesis that spirit evolved via a dialectic of distincts rather than of opposites, Croce argued that intuition tends to give way to perception and so to conceptual thought or judgment. Furthermore, such knowledge leads us to take up a new attitude to life and so affects our practice. This new will in its turn solicits new passions and feelings that find expression in a new lyric and fresh art, constituting a process which Croce termed the ‘circle of spirit’, in which each moment is both independent and dependent, condition and conditioned.

The third phase of Croce’s aesthetics builds on this thesis by insisting on the ‘cosmic’ or ‘universal’ character of art. Put forward in his article ‘Il carattere di totalità dell’espressione artistica’ (The Totality of Artistic Expression) (1918a), this doctrine served to underline the cognitive aspect of his theory by stressing that as an aspect of spirit ‘every genuine artistic representation is itself and is the universe, the universe in that individual form’ (1926: 122). However, intuition formed only a part of the ‘circle of spirit’, and he rebutted criticisms of mysticism or aestheticism which accused him of reducing all knowledge to artistic intuition. Nevertheless, at this time Croce tended to reify spirit and treat human activity as a mere manifestation of its unfolding, a position he later rectified. Croce illustrated his argument in studies of *Goethe* (1919a), *Ariosto, Shakespeare, Corneille* (1920) and *La poesia di Dante (The Poetry of Dante)* (1921).

The final phase was signalled by *La poesia* (1936). Croce regarded this book as incorporating all the subsequent revisions of his aesthetic theory and replacing the *Aesthetic* of 1902. He argued that there were four types of ‘expression’: the ‘sentimental or immediate’, the ‘poetic’, the ‘prosaic’ and the ‘rhetorical’. True poetry, he argued, only arose when these types originated from a ‘lyrical expression’ and had no ulterior utilitarian, moral or philosophical purpose. When lyricism was absent, one had literature, which he further subdivided into the sentimental, moralistic, entertaining or instructive.

3 Logic and history

Croce’s *Logic* centred on his notion of the ‘pure concept’ as a universal idea that could not be confused with the particular representations it encompassed. As such, it was to be distinguished from ‘pseudo-concepts’, which were mere classes of objects. These latter were empirical rather than theoretical categories.

The content of ‘pure concepts’ came through history. The identification of history and philosophy in the second edition of the *Logic* (1909) was the most important revision Croce was to make to his theory. Judgment, he argued, is essentially historical in nature, involving the union of a pure universal concept with a historically given particular object. When we say ‘Peter is a man’ we not only define a characteristic of Peter with a preformed concept of what a man is, we reaffirm and change the concept of man in relation to Peter. Our notions of maleness are historical products of past meetings with other Peters and Pauls, and modifiable by other future encounters. Similarly, Peter’s conception of himself was equally defined by a given notion of man. In this way, Croce argued, philosophy proved both conditioned by history and a means, through conceptual innovation, for the making of new

history. Likewise, theory and practice are mutually related, the one influencing the other.

Croce carried this thesis further in the final volume of the *Philosophy of Spirit*, the *Theory and History of Historiography*. Here he famously claimed that ' All history is contemporary history' because 'it meets a present need' (1917: 5-17). The collection of facts about the past, while a necessary preliminary for historical writing, was not yet history but 'chronicle'. It was a 'dead' past because it had yet to be interpreted and given a meaning. Interpretation involved entering the minds of the historical actors and understanding the inner significance of their actions by reliving these events in the historian's own consciousness. Croce believed that this was possible because we are what our past has made us, so that our knowledge of ourselves is historically conditioned. The past literally lives within us.

Croce was careful to avoid any suggestion of relativism or subjectivism. He denied that history was whatever we took it to be. Indeed, at times he appeared to argue that it was not so much individuals, as spirit 'eternally individualizing itself' that was the true subject of history. This thesis, together with his contention that we are all products of the past, risked denying individuals the possibility for any creative action at all. During the 1920s and 1930s he modified his position on both these fronts. In a series of essays written in 1922-4, he proposed the view that history was 'ethico-political' in nature. In other words, the conflict of different individual ideals was the true dynamic of history, a view he deployed against both Marxism and Fascism. Historicism, therefore, buttressed liberalism. He was not thereby advocating voluntarism, however. As he made clear in his final major work, *History as the Story of Liberty*, successful action depends on a correct appraisal of one's current situation - itself a matter of historical judgment. Knowing the world in thought gives an impetus to the ethical desire to transform it through action.

4 Ethics and politics

Croce's initial concern in the field of practical philosophy was to distinguish the concepts of the Useful and the Good and to combat both utilitarian moralities and moralistic politics as category mistakes. Although this distinction figured in his writings on Marx and formed the basis of an influential debate with Vilfredo Pareto on the relation of economics to ethics, he did not expound it fully until the *Filosofia della pratica: economia ed etica*, where he integrated it into his general theory of the circle of spirit within history.

He did not draw any practical inferences from this doctrine until the First World War. In a series of articles later collected as *Pagine sulla guerra (Writings on the War)* (1919b), he attacked those who maintained that the war was between two rival moralities - the democratic and the autocratic. Croce contended that war, like politics in general, was pre-moral in being orientated towards the Useful. The state had no purpose but that of power and no means other than force, so that the military struggle should be seen as a matter of pure *realpolitik*. He strenuously denied that this view entailed identifying might with right. However, in essays published as *Frammenti di etica (The Conduct of Life)* (1922), he did insist that morality was not something that could be pursued directly, but was a matter of the judgment of history. Accordingly, he argued that each person should simply fulfil the duties of their appointed station and trust in Providence.

The dangers of this deeply conservative position became evident with the rise of Fascism (see Fascism). Croce initially supported it, albeit lukewarmly, as leading to a strengthening of the state. However, Gentile, now a keen supporter of Fascism, made precisely the jump that Croce feared of identifying the force of the state with its moral strength. This move, and his resulting doctrine of the ethical state, led Croce to clarify his position. He accused Gentile and the Fascists of putting forward a 'governmental morality'. While he continued to regard the state in the narrow sense as belonging to the realm of the Useful, he now maintained that politics in general had an ethical dimension in the struggle to realize certain ideals and he endorsed liberal democracy for allowing the widest possible competition between different points of view. However, he claimed his liberalism to be 'metapolitical' rather than narrowly party political. While he believed that liberalism entailed a free market in ideas, he argued that different economic and political policies might be appropriate in different historical circumstances - even communism. This argument provoked a debate with the economist Luigi Einaudi over whether moral and political liberalism entailed libertarianism. Against Einaudi, Croce maintained that economic liberals erred in attempting to categorize everything in terms of the Useful and so failed to see that liberalism was the dialectic of spirit in all of its moments throughout the whole of history. Croce made the same criticism of left-of-centre liberals such as Guido Calogero and Guido de Ruggiero, who campaigned for a social welfare view

of liberalism similar to that of the English New Liberals, such as L.T. Hobhouse. While Croce's 'metapolitical' position proved useful in uniting the various opponents of Fascism, it yielded little except a certain pragmatism when it came to practical politics. After the war, Croce acted as President of a new formed Liberal Party. However, he was unable to forge a broad non-Catholic non-Socialist bloc, as he had hoped. Instead, he found himself presiding over an increasingly right-wing party and ultimately resigned.

The shifts in Croce's politics highlight a constant tension in his philosophy as a whole between historicism and idealism, realism and rationalism, politics and ethics, the Hegelian and the Kantian aspects of his thought. In some respects, his chief claim was to have made a virtue out of this tension, rather than attempting to synthesize these two elements. The constant revisions that his thought underwent, however, suggests that he found himself constantly pulled in different directions, emphasizing one then the other.

See also: Economics and ethics §5; Artistic expression §4; Hegelianism §6; Historicism §1; Poetry §4; Social Democracy §1

RICHARD BELLAMY

List of works

Croce, B. (1906-52) *Opere complete*, Bari: Laterza, 74 vols.(This excludes his correspondence and is not exhaustive. A National Edition is planned, and Adelphi Edizioni of Milan have recently published semi-critical editions of some of his more important works. Croce also made a selection of his works published in 1951 as *Filosofia-Poesia-Storia*, >Milan and Naples: Riccardo Ricciardi; trans. C. Sprigge, *Philosophy, Poetry, History: An Anthology of Essays*, London: Oxford University Press, 1966.)

Croce, B. (1891) *I teatri di Napoli del Secolo XV-XVIII (The Theatres of Naples from the Fifteenth to the Eighteenth Centuries)*, Naples: Pierro.(The culmination of Croce's only archival research into Neapolitan history of this period.)

Croce, B. (1893) 'La storia ridotto sotto il concetto generale dell'arte' (History Subsumed under the General Concept of Art), *Atti dell'Accademia Pontaniana* 24: 1-32. (Croce's first foray into philosophy.)

Croce, B. (1900) *Materialismo storico ed economia marxista*, Palermo: Sandron; trans. C.M. Meredith, *Historical Materialism and the Economics of Karl Marx*, London: Latimer, 1914.(Contains his critique of vulgar Marxism and Marxist economics, and his debate with Pareto.)

Croce, B. (1902) *Estetica come scienza dell'espressione e linguistica generale*, Milan, Palermo, Naples: Sandron; trans. D. Ainslie, *Aesthetic*, London: Macmillan, 1909; 2nd complete edn, 1922.(The first volume of his *Philosophy of Spirit*, which provides both the first version of his aesthetic doctrine and of his philosophy as a whole - each of which were subsequently much modified.)

Croce, B. (1905) *Logica come scienza del concetto puro*, Bari: Laterza; 2nd edn, 1909; trans. D. Ainslie, *Logic*, London: Macmillan, 1917.(The second volume of the *Philosophy of Spirit*. The changes to the second edition were occasioned by criticisms of the first by Gentile, whose influence is duly acknowledged, which led Croce in a more pronounced historicist direction.)

Croce, B. (1907) *Cio che è vivo e cio che è morto nella filosofia de Hegel*, Bari: Laterza; trans. D. Ainslie, *What Is Living and What Is Dead in the Philosophy of Hegel*, London: Macmillan, 1915.(A classic study of the German philosopher, which marked Croce's passage from a Neo-Kantian to a more Hegelian position, albeit with important reservations.)

Croce, B. (1909) *Filosofia della pratica, economia ed etica*, Bari: Laterza; trans. D. Ainslie, *Philosophy of the Practical*, London: Macmillan, 1913.(The third volume of the *Philosophy of Spirit*.)

Croce, B. (1911) *La filosofia de Giambattista Vico*, Bari: Laterza; trans. R.G. Collingwood, *The Philosophy of Giambattista Vico*, London: Latimer, 1913.(One of the earliest studies of the eighteenth-century Neapolitan philosopher, whose thought Croce did much to promote. As with his study of Hegel, his technique is to separate the living from the dead on the basis of those elements which are absorbed or rejected by his own philosophy.)

Croce, B. (1913) *Breviario de estetica*, Bari: Laterza; trans. D. Ainslie, *The Essence of Aesthetic*, London: Heinemann, 1921; first published in *The Book of the Opening of the Rice Institute*, Houston, TX, 1912.(Marks the second phase of his aesthetic doctrine.)

Croce, B. (1917) *Teoria e storia della storiografia*, Bari: Laterza; trans. D. Ainslie, *Theory and History of Historiography*, London: Harrop, 1921.(The fourth and final volume of his *Philosophy of Spirit*, in which he gives his first full exposition of his historicist views.)

Croce, B. (1918) *Contributo alla critica de me stesso*, privately printed in Naples for distribution to friends; trans.

R.G. Collingwood, *Autobiography*, Oxford: Clarendon Press, 1927.(Subsequently extended, and republished as an appendix to *Etica e politica*, it focuses on the evolution of his thought and takes Vico's *Autobiography* as its model. He periodically brought it up to date in later editions of *Etica e politica*.)

Croce, B. (1919a) *Goethe*, Bari: Laterza; trans. E. Anderson, London: Methuen, 1923.(A study of the German poet Goethe in which Croce seeks to illustrate his 'lyrical' account of poetic expression. The Italian edition included translations of his poetry by Croce.)

Croce, B. (1919b) *Pagine sulla guerra (Writings on the War)*, ed. G. Castellano, Naples: Ricciardi; 2nd expanded edn, Bari: Laterza, 1928.(A collection of journalistic pieces in which Croce advocated Italian neutrality and, once Italy had entered the war, argued for a realist view of politics, which he associated with Germany, against the idealist view he linked with the allies.)

Croce, B. (1920) *Ariosto, Shakespeare, Corneille*, Bari: Laterza; trans. D. Ainslie, New York: Holt, 1920. (Originally published separately, Croce employed these studies of three great dramatists to further develop his theory of poetry.)

Croce, B. (1921) *La poesia di Dante*, Bari: Laterza; trans. D. Ainslie, *The Poetry of Dante*, London: Macmillan, 1922.(A study of Dante's poetry.)

Croce, B. (1922) *Frammenti di etica*, Bari: Laterza; trans. A. Livingston, *The Conduct of Life*, New York, 1924. (Collects articles mainly written in 1915-18 which attempt to elaborate a historicist ethics against the background of the First World War.)

Croce, B. (1925) *La storia del regno de Napoli*, Bari: Laterza; trans. F. Frenaye, *History of the Kingdom of Naples*, Chicago, IL: Chicago University Press, 1970.(An attempt at 'ethico-political' history that focuses on the role played by intellectuals.)

Croce, B. (1926) *Nuovi saggi di estetica (New Essays on Aesthetics)*, 2nd edn, Bari: Laterza.(Collects together the various essays in which Croce progressively revised his philosophy of aesthetics. Includes 'Il carattere di totalità della espressione artistica' (1918).)

Croce, B. (1928a) *Storia d'Italia dal 1871 al 1915*, Bari: Laterza; trans. C.M. Ady, *History of Italy 1871-1915*, Oxford: Clarendon Press, 1929.(A defence of the liberal regime of Giolitti, with important chapters on the cultural life of the period.)

Croce, B. (1928b) *Aesthetica in nuce*, Bari: Laterza; trans. R.G. Collingwood, 'Aesthetics', in *Encyclopaedia Britannica*, New York and London, 14th edn, 1929, vol. 7.(The most accessible account of the third phase of Croce's aesthetic doctrine.)

Croce, B. (1929) *Storia dell'eta barocca in Italia (A History of the Baroque Era in Italy)*, Bari: Laterza. (Written in 1924-5 and published in instalments in *La critica* between 1924 and 1928, this was a path-breaking work that prompted study of a relatively neglected period of Italian history.)

Croce, B. (1931) *Etica e Politica*, Bari: Laterza; trans. S. Castiglione, *Politics and Morals*, London: Allen & Unwin, 1946.(Collects essays written in 1924-8, and represents Croce's attempt to defend a specifically liberal political order against Fascism and to revise his historicism in an 'ethico-political' direction that gave greater play to the creative action of individuals.)

Croce, B. (1932) *Storia d'Europa nel secolo XIX*, Bari: Laterza; trans. H. Furst, *History of Europe in the Nineteenth Century*, New York: Harcourt Brace, 1933.(The most 'philosophical' of Croce's histories, which is presented in terms of an apologia for the nineteenth-century liberal culture with which he strongly identified.)

(1936) *La poesia*, Bari: Laterza.(The final version of his aesthetic.)

(1938) *La storia come pensiero e come azione*, Bari: Laterza, trans S. Sprigge, *History as the Story of Liberty*, New York: W.W. Norton, 1941.(A major revision of his earlier historicism that links it firmly with a 'metapolitical' liberalism through its emphasis on the creative role of human thought and action.)

(1949) *My Philosophy, and Other Essays on the Moral and Political Problems of Our Time*, London: Allen & Unwin, selected by R. Klibansky, trans. E.F. Carritt.(A selection from Croce's later works.)

(1981) *Lettere a Giovanni Gentile (1896-1924)*, ed. A. Croce, Milan: Arnaldo Mondadori.(Contains invaluable accounts of his disagreements with Gentile that allows one to trace the development of his thought and the influence of his one-time colleague upon it. Gentile's letters are republished in his Collected Works.)

References and further reading

Bellamy, R.P. (1986) 'Liberalism and Historicism: Benedetto Croce and the Political Role of Idealism in Modern Italy 1890-1952', in A. Moulakis (ed.) *The Promise of History: Essays in Political Philosophy*, Berlin and New York: de Gruyter.(Locates Croce's thought in the cultural and political context of the times.)

Bellamy, R.P. (1987) *Modern Italian Social Theory: Ideology and Politics from Pareto to the Present*, Cambridge: Polity Press, ch. 5.(Focuses on the development of Croce's historicism.)

Bellamy, R.P. (1991) ‘Between Economic and Ethical Liberalism: Benedetto Croce and the Dilemmas of Liberal Politics’, *History of the Human Sciences* 4: 175-95.(An account of Croce’s political ideas.)

Borsari, S. (1964), *L’opera di Benedetto Croce (The Works of Benedetto Croce)*, Naples: Istituto italiano per gli studi storici.(The definitive bibliography of the various editions of Croce’s works, although now a little dated.)

Corsi, M. (1974) *Le origini del pensiero di Benedetto Croce (The Origins of Benedetto Croce’s Thought)*, Naples: Giannini, 2nd edn.(An excellent study of Croce’s writings pre-1903, which traces the origins of his later thought in his earlier work.)

De Sanctis, F. (1870-1) La storia de letteratura italiana (History of Italian Literature), in *Opere di De Sanctis*, Turin: Einaudi, 1958, vols 8 and 9.(Croce produced the first critical edition of this work in 1912. In the volumes of *La critica* preceeding the First World War he sought to continue it with studies of authors and literary schools that had developed in Italy since unification.)

Galasso, G. (1990) *Croce e lo spirito del suo tempo (Croce and the Spirit of His Times)*, Milan: Mondadori.(A historical study of Croce’s philosophy.)

Jacobitti, E.E. (1981) *Revolutionary Humanism and Historicism in Modern Italy*, New Haven, CT: Yale University Press.(Looks at Croce’s writings pre-1914 and relates them to the Italian idealist tradition of the nineteenth century.)

Orsini, G.N.G. (1961) *Benedetto Croce, Philosopher of Art and Literary Critic*, Carbondale, IL: Southern Illinois University Press.(The best account in English of Croce’s aesthetics.)

Roberts, D.D. (1987) *Benedetto Croce and the Uses of Historicism*, Berkeley and Los Angeles, CA, and London: University of California Press.(A defence Croce’s historicism.)

Rossi, P. (1957) ‘Benedetto Croce e lo storicismo assoluto’(Benedetto Croce and Absolute Historicism), *Il Mulino* 6: 322-4.(An excellent critical study of the evolution of Croce’s historicism.)

Sainati, V. (1953) *L’estetica di Croce (Croce’s Aesthetics)*, Florence: Felice Le Monnier.(A fine account of the different phases of Croce’s aesthetic.)

Sasso, G. (1975) *Benedetto Croce: la ricerca della dialettica*, Naples: Morano.(A critical defence of the Crocean project, that attempts to trace an internal logic to the progressive evolution of Croce’s thought.)

Crucial experiments

A 'crucial experiment' allegedly establishes the truth of one of a set of competing theories. Francis Bacon (1620) held that such experiments are frequent in the empirical sciences and are particularly important for terminating an investigation. These claims were denied by Pierre Duhem (1905), who maintained that crucial experiments are impossible in the physical sciences because they require a complete enumeration of all possible theories to explain a phenomenon - something that cannot be achieved. Despite Duhem, scientists frequently regard certain experiments as crucial in the sense that the experimental result helps make one theory among a set of competitors very probable and the others very improbable, given what is currently known.

Francis Bacon was the first to discuss the idea of a 'crucial experiment' at length. He called results of such experiments 'instances of the cross' (hence *experimentum crucis*), borrowing our metaphor from the crosses erected where two roads meet to point out the different directions. Bacon claimed that such decisive instances are to be found both among ones already obtained and, more frequently, among new ones produced expressly for that purpose. He offered several scientific examples, including experiments to decide between competing theories of the tides, of the apparent diurnal rotation of the heavens, of the weight of bodies, of magnetism, and of exploding gunpowder. According to Bacon (1620), such experiments 'afford great light, and are of great weight, so that the course of interpretation sometimes terminates, and is completed in them'.

The most famous attack on the idea of crucial experiments was offered by Duhem, who boldly asserted (1905) that ' a "Crucial Experiment" is impossible in physics'. He gave two arguments. The first, an expression of Duhem's 'holistic' view of science, is that what is tested is not a single hypothesis or theory by itself, but a 'whole theoretical group', including assumptions regarding the laboratory instruments to be used in testing it. Secondly, and more importantly, even if only a 'theoretical group' rather than an individual hypothesis is tested, for a crucial experiment to be decisive:

> it would be necessary to enumerate completely the various [competing] hypotheses which may cover a determinate group of phenomena; but the physicist is never sure he has exhausted all the imaginable assumptions. The truth of a physical theory is not decided by heads or tails.
>
> (Duhem 1905: 190)

What lies behind the idea of a crucial experiment, according to Duhem, is an 'argument by elimination' with a form such as E_1:

(a) Some phenomenon O is observed.
(b) O can be explained by, and only by, exactly one of the theories $T_1, \ldots, T_n$.
(c) But $T_2, \ldots, T_n$ are false, because each entails some observable consequence which the crucial experiment shows to be false.
(d) So T_1 must be true.

In order to use such an argument, one must consider among $T_1, \ldots, T_n$ not just known theories, but all possible ones that could explain the phenomenon; and this, as Duhem points out, is impossible. If crucial experiments depend on such eliminative arguments, they must be impossible.

Yet scientists frequently do offer eliminative arguments, and of a form not subject to Duhem's major objection. In the early nineteenth century, proponents of the wave theory of light, such as Thomas Young and Augustin Fresnel, sought to establish their theory over the then entrenched particle theory by using an eliminative argument consisting of these steps: (1) Because light travels in space with a finite velocity, and the only known methods of communication of motion from one point in space to another are via the transference of bodies or the propagation of a wave disturbance through a medium, it is highly probable that light consists either of waves or of particles. (2) Some optical phenomena, for example, rectilinear propagation of light, reflection and refraction, can be shown to follow from both the wave and particle theories without difficulty. But in explaining certain other optical phenomena, most notably diffraction, the particle theory, but not the wave theory, introduces auxiliary assumptions that are very probable assuming the particle theory is true but very improbable given everything else

that is known. (3) Wave theorists conclude that, given what is known (including all the observed optical phenomena), the wave theory is highly probable (see Achinstein 1991) (see Optics §1).

This argument is not refuted by Duhem's logical claims. Wave theorists do not assert, nor do they need to, that the particle and wave theories are the only *possible* ones (which is false, as was shown by twentieth-century quantum theory). They assert only that, relative to what is known at the time, it is highly probable that one theory or the other is true. And they offer reasons for this claim from known causes of motion. Their eliminative argument has a form expressed not by E_1 but by E_2:

(a) Some phenomenon O is observed.
(b) In the case of phenomena of the same type as O, the only known causes are C_1 and C_2; so probably either T_1 (which postulates cause C_1) or T_2 (which postulates C_2) is true.
(c) But T_2 is probably false because, in order to explain related phenomena, T_2, but not T_1, introduces auxiliary assumptions that although probable given T_2 are very improbable otherwise.
(d) So, given what is known, probably T_1 is true.

How do 'crucial experiments' enter the picture? Experiments that reveal phenomena which one theory but not the other explains without introducing highly improbable assumptions may be considered 'crucial' ones in establishing the high probability of the former theory. Fresnel's 1819 experiments on the diffraction of light showed the positions and intensities of the observed diffraction bands. These results were derivable from the wave theory without difficulty. To obtain them from the particle theory one had to introduce special forces acting at a distance that produce effects completely different from any ever observed in nature. Fresnel considered his experiments crucial in so far as they helped make the wave theory highly probable and the particle theory highly improbable.

PETER ACHINSTEIN

References and further reading

Achinstein, P. (1991) *Particles and Waves*, New York: Oxford University Press. (Co-winner of the Lakatos Award in 1993, this work discusses methodological issues in nineteenth-century physics, and includes an analysis of arguments given by wave theorists of light.)

Bacon, F. (1620) *Novum Organum*; Chicago, IL: University of Chicago Press, 1952.(Book II, section 36, of this classic contains Bacon's discussion of crucial experiments.)

Duhem, P. (1905) *The Aim and Structure of Physical Theory*, Princeton, NJ: Princeton University Press, 1991. (One of the greatest and most readable twentieth-century works in the philosophy of science.)

Glymour, C. (1980) *Theory and Evidence*, Princeton, NJ: Princeton University Press. (Argues, contra Duhem, that individual hypotheses in a theory can be confirmed.)

Crusius, Christian August (1715-75)

Crusius was a pivotal figure in the middle period of the German Enlightenment, linking Pufendorf and Thomasius with Kant. Though sometimes wrongly characterized (for example, by Hegel) as a Wolffian, he was instead an important critic of that position. His system reflected a new alliance between Pietism and Lutheran orthodoxy, offering a comprehensive antirationalist, realist, and voluntarist alternative to the neoscholastic tradition as renovated by Leibniz. Crusius was important in Kant's development and helps us understand the latter's philosophical Protestantism.

Born a pastor's son in Leuna bei Merseburg, in Saxony, Crusius was educated at Leipzig and much influenced there by the Thomasian professor A.F. Hoffmann (1703-41). Interested in both philosophy and theology throughout his career, he accepted a chair as extraordinary professor of philosophy at Leipzig in 1744. In 1750, however, he became ordinary professor of theology, also retaining his teaching post in philosophy until his death. His reputation as a philosopher peaked in Germany during the 1750s and 1760s, mainly on the basis of four scholastic manuals published in German during 1744-9. His greater theological reputation as founder of a 'biblico-prophetic' school emphasizing the inspirational unity of Scripture lasted well into the mid-nineteenth century.

1 Philosophical intentions

Crusius' works divide not only into philosophical and theological, but also into Latin and German. In the latter case, which includes his four scholastic manuals *Anweisung* (1744), *Entwurf* (1745), *Weg* (1747) and *Anleitung* (1749), there is disagreement over which are more important, and one gets differing accounts depending on their respective bases. The Latin essays, including *De usu et limitibus principii rationis determinantis (On the Use and Limits of the Principle of Determining Reason)* (1743), cover much the same ground as the manuals, but they are more focused, less didactic and together yield better insight into Crusius' philosophical concerns. Collected in revised form in the *Opuscula philosophico-theologica (Small Philosophical and Theological Works)* of 1750, most of them received approved German translations within a decade or two. Several works from Crusius' 'theological' period also help to clarify his notion of philosophy and its tasks. They include the *Epistola…de summis rationis principiis (Letter…on the Highest Principles of Reason)* (1752), *Dissertationes II-IV de superstitione (Discourses II-IV on Superstition)* (1766), *Kurzer Lehrbegriff der Moraltheologie (A Short Outline of Moral Theology)* (1772: vol. 1), and the prefaces to later editions of the manuals.

These sources reveal Crusius' opposition to 'the newer philosophy', which is inaccurately characterized as 'the Leibnizian-Wolffian system'. For Wolff himself rejected the term, claiming that he knew little of Leibniz when he wrote his so-called 'German Metaphysics' (1719), and Crusius did not use it until 1753 in the preface to the second edition of his *Entwurf*, having previously described the *praeceptor Germaniae* as a 'propagator of Leibnizian philosophy'. Leibniz and Wolff are often mentioned in the Latin writings (the manuals intentionally omit authorial references) and the similarities between them noted, but they are not regarded as having a joint system. Instead, the 'newer philosophy' consisted for Crusius of Leibnizian doctrines about sufficient reason, pre-established harmony, monadic substances, and the best of all possible worlds, with Wolff playing the role of systematizer and elaborator (see Leibniz, G.W.; Wolff, C.).

Carboncini (1986) suggests that Crusius saw the 'newer philosophy' as a general type. This emerges especially from the theological works, which link it with ancient superstition opposed to sound human understanding, particularly with Stoic fatalism. Beside this objection to its content, Crusius also criticized its syncretistic form and its methodological reliance on sectarian prejudices. His own philosophy, by contrast, emphasized freedom, clarity of concepts (rather than the formal necessity and certainty of arguments) and critical thinking for oneself. These notions link him with the Pietist, 'eclectic' tradition with which he has other, more specific affinities. They also suggest a connection between his philosophical and theological interests and provide a wider context for the needed developmental account of his thought.

2 Basic elements of Crusius' system

Since they were meant to counter Wolff's influence, Crusius' four scholastic manuals are themselves quite

Wolffian: thorough, explicit, systematic and full of definitions, multiple arguments and cross-references. They also address many topics treated by Wolff and clearly complement one another.

Least original among them is the *Anleitung*, which was the first Pietist tract to interpret nonliving nature in mechanistic terms, though Crusius still rejected Newtonian gravity. It also emphasized humans' merely hypothetical knowledge of basic natures and powers, and developed a position about the probable status of empirical generalizations or 'presumptions' (for example, nature makes no leaps) similar to Hume's (see HUME, D.).

A 'logic' in the broad sense of a methodology of knowledge, Crusius' *Weg* covers many topics treated by Wolff under empirical psychology. This is because philosophy was for him not merely a science of the possible but one dealing with really existing things. The general 'thinkability' (*cogitabilitas*) criterion that regulates it was therefore elaborated into three principles instead of one. To the formal principle of (A) non-contradiction from which Wolff had sought to derive all other truths (see Wolff, C. §6), Crusius thought it necessary to add two other, material principles: those of (B) inseparability (*das Nichtzutrennenden*) and (C) unconjoinability (*das Nichtzuverbindenden*). What can or cannot be conceptually separated or joined is not a question of logical possibility only but depends on the nature of the human understanding and the general content of experience. Accordingly Crusius also rejected Wolff's attempt to employ the mathematical method in philosophy, citing no less than nine differences between them, including the fact that mathematics begins with simple ideas whereas philosophy must discover them through an analysis of experience.

Like its counterpart, the will, the human understanding is composed of several basic powers, including sensation, reason, memory, and judgment. We have no clear grasp of these, nor of most other basic concepts, but know them only indirectly or relatively, through comparisons or their effects, in what Crusius calls symbolic as opposed to intuitive knowledge. Symbolic knowledge is regulated by principles (B) and (C) and can be (morally) certain even though it is incomplete. Its conclusions may not be ontologically valid and may, indeed, involve antinomic 'conflicting proofs', but this only reflects the fact that for Crusius some truths are not only beyond but even opposed to our limited human understanding.

Crusius' epistemology combines a Lockean doctrine of simple ideas, discovered through an analysis of experience, with a Leibnizian innatism prefiguring Kant (see Locke, J.; Kant, I. §§9-11). This becomes clearer in the *Entwurf*, which is sectioned into a reordered Wolffian scheme of ontology, theology, cosmology and pneumatology (rational psychology). Here Crusius discusses the (renamed) principle of determining reason, the distinction between the ideal (*ratio cognoscendi*) and the real (*ratio essendi vel fiendi*) ground of things, his doctrine of basic concepts (which he once called 'categories'), particularly space, time and causality, the failure of the ontological argument for God's existence, the nature and interaction of material and spiritual substances, and the two fundamental powers of the human soul, understanding and will.

The fact that the *Anweisung* was the first of Crusius' manuals to appear indicates the primacy of the practical in his thought. Its opening 'thelematology' discussing the human will and its component basic desires and essential freedom is followed by sections on ethics, natural moral theology, natural law, and prudence. These reflect both the standard natural law schema of duties to self, God and others, and the eudaimonist-perfectionist orientation of Leibniz and Wolff (see Natural law). What distinguishes Crusius is his integration of the intellectualist and voluntarist strains in ethics that had uneasily coexisted since Pufendorf into a form for which Kant has received most credit. Yet much of Kant's moral apparatus is Crusian, including the central notions of duty, obligation, law, right, virtue, happiness, the highest good, and the moral postulates (see KANT, I. §§9-11). Indeed, since Crusius clearly distinguished prudential and moral necessity, as well as action in accordance with and for the sake of duty, Kant's classification of him as a 'theological moralist' in the analytic of the *Critique of Practical Reason* (1788) is, while technically correct, also misleading. By not doing justice to Crusius it hides Kant's philosophical debts and thereby distorts the history of ethics.

3 Influence

Kant scholars rediscovered Crusius in the late nineteenth century. Yet even this relationship (most evident in Kant's *Reflexionen* of 1774-5 and lectures) is insufficiently understood, despite the seminal work of Heimsoeth (1956), Wundt (1964), Schmucker (1961) and Tonelli (1967, 1969). Crusius' influence on the *Nova Dilucidatio*

(1755) and the 'Prize Essay' (1764), where Kant distanced himself from Wolffianism, is clear, but his role in Kant's later theoretical and practical philosophy awaits more comprehensive assessment.

Little is known about Crusius' influence beyond Kant. He had numerous disciples into the 1770s, including A.F. Reinhard (1726-83), who received the 1755 Berlin Academy prize on the question of optimism. The Academy itself, led since 1745 by Maupertuis, had broad affinities with Crusianism because of its opposition to Leibniz and Wolff, but these too remain unelaborated. Other links requiring study are those with J.H. Lambert, M. Mendelssohn, J.B. Basedow (who shaped Kant's views on education), and Fichte, who knew Crusius through C.F. Pezold's important 1766 German edition of the *De usu* (see Fichte, J.G.).

See also: Enlightenment, Continental; Pietism; Thomasius, C.

MICHAEL J. SEIDLER

List of works

Crusius, C.A. (1964-) *Die philosophischen Hauptwerke (The Main Philosophical Works)*, ed. Tonelli, G., Hildesheim: Olms, 4 vols.(Volume 4.2, *Namenregister und wissenschaftlicher Apparat zu den Bänden I-IV,1 (Register of names and critical apparatus for vols I-IV.1)* is to appear in 1996.)

Crusius, C.A. (1744) *Anweisung vernünftig zu leben (Guide to Rational Living)*, Leipzig; in Crusius (1964-), vol 1.(Crusius' 'ethics'.)

Crusius, C.A. (1745) *Entwurf der nothwendigen Vernunft-Wahrheiten, wiefern sie den zufälligen entgegen gesetzt werden (Sketch of the Necessary Truths of Reason, insofar as they are Opposed to Contingent Truths)*, Leipzig; in Crusius (1964-), vol 2.(Crusius' 'metaphysics'.)

Crusius, C.A. (1747) *Weg zur Gewissheit und Zuverlässigkeit der menschlichen Erkenntnis (The Way to the Certainty and Dependability of Human Knowledge)*, Leipzig; in Crusius (1964-), vol 3.(Crusius' 'logic'.)

Crusius, C.A. (1749) *Kleinere philosophische Schriften (Minor Philosophical Writings)*, Leipzig; in Crusius (1964-), vol 4.1.(Crusius' most important Latin essays and the philosophically relevant portions of his 'physics', that is, chaps 1-2 of the *Anleitung, über natürliche Begebenheiten ordentlich und vorsichtlich nachzudenken (Guide to Thinking about Natural Events in an Orderly and Careful Manner)*. (See Carboncini and Finster (1987), vol. 4.1)

Schneewind, J.B. (trans.) (1990) 'Christian August Crusius', in J.B. Schneewind (ed.) *Moral Philosophy from Montaigne to Kant*, vol. 2: 568-85, New York: Cambridge University Press.(These excerpts from the *Anweisung* are the only translation of Crusius into English.)

References and further reading

Beck, L.W. (1969) *Early German Philosophy*, Cambridge, MA: Belknap Press, 394-402.(Reliable but limited mainly to Crusius' *Entwurf*.)

Benden, M. (1972) *Christian August Crusius. Wille und Verstand als Prinzipien des Handelns*, Bonn: Bouvier. (Thorough, but repetitive and tedious.)

Carboncini, S. (1986) 'Christian August Crusius und die Leibniz-Wolffsche Philosophie', in A. Heinekamp (ed.), *Beiträge zur Wirkungs- und Rezeptionsgeschichte von Gottfried Wilhelm Leibniz*, Stuttgart: Franz Steiner, 110-25; repr. *Studia Leibnitiana Supplementa* 26.(Referred to in §1. Important for its reconsideration of Crusius' relation to Leibniz and Wolff.)

Carboncini, S. (1989) 'Die thomasianisch-pietistische Tradition und ihre Fortsetzung durch Christian August Crusius', in W. Schneiders (ed.) *Christian Thomasius 1655-1728. Interpretationen zu Werk und Wirkung*, Hamburg: Felix Meiner, 287-304.(Refreshingly different in linking Crusius, and so indirectly Kant, to 'eclecticism'.)

Carboncini, S. and Finster, R. (1987) 'Einleitung', in G. Tonelli (ed.) *Die philosophischen Hauptwerke*, vol. 4.1, Hildesheim and New York: Olms, i-xxxvi.(Essential supplement to Tonelli's 'Introduction' in *Hauptwerke*, vol. 1.)

Heimsoeth, H. (1956) 'Metaphysik und Kritik bei Chr. A. Crusius. Ein Beitrag zur ontologischen Vorgeschichte der *Kritik der reinen Vernunft* im 18. Jahrhundert' (A contribution to the history of the *Kritik der reinen Vernunft* in the eighteenth century), *Studien zur Philosophie Immanuel Kants. Metaphysische Ursprünge und Ontologische Grundlagen*, Cologne: Kölner Universitätsverlag, ch. 3, 125-88.(An acute treatment of Crusius' many anticipations of Kant's metaphysics of experience.)

Krieger, M. (1993) *Geist, Welt und Gott bei Christian August Crusius*, Würzburg: Königshausen & Neumann. (The latest study of Crusius, focusing on his 'special metaphysics'; Chapter 1 contains a full discussion of the

previous Crusius literature.)
Schmucker, J. (1961) *Die Ursprünge der Ethik Kants in seinen vorkritischen Schriften und Reflektionen (The Sources of Kant's Ethics in his Precritical Writings and Reflections)*, Meisenheim am Glan: Anton Hain, 81-98. (Still the most explicit source for Crusius' influence on Kant's mature moral philosophy.)
Tonelli, G. (1967) 'Crusius, Christian August', in P. Edwards (ed.) *The Encyclopedia of Philosophy*, vol. 2: 268-71, New York: Macmillan.(Good general introduction, particularly for English readers.)
(ed.) (1969) 'Einleitung' (Introduction), *Die philosophischen Hauptwerke*, vol. 1, Hildesheim: Olms, vii-lxv.(The standard overview article on Crusius and his works.)
Wundt, M. (1945) *Die deutsche Schulphilosophie im Zeitalter der Aufklärung*, Tübingen: Mohr; repr. Hildesheim: Olms, 1964, 254-64.(Brief treatment based on Crusius' scholastic manuals.)
(1964) *Kant als Metaphysiker*, Stuttgart: Enke, 52-81.(A good account based primarily on Crusius' Latin works.)

Cudworth, Ralph (1617-88)

Ralph Cudworth was the leading philosopher of the group known as the Cambridge Platonists. In his lifetime he published only one work of philosophy, his True Intellectual System of the Universe (1678). This was intended as the first of a series of three volumes dealing with the general topic of liberty and necessity. Two further parts of this project were published posthumously, from the papers he left when he died: A Treatise Concerning Eternal and Immutable Morality (1731) and A Treatise of Freewill (1838).

Cudworth's so-called Cambridge Platonism is broadly Neoplatonic, but he was receptive to other currents of thought, both ancient and modern. In philosophy he was an antideterminist who strove to defend theism in rational terms, and to establish the certainty of knowledge and the existence of unchangeable moral principles in the face of the challenge of Hobbes and Spinoza. He admired and borrowed from Descartes, but also criticized aspects of Cartesianism.

Cudworth's starting point is his fundamental belief in the existence of God, conceived as a fully perfect being, infinitely powerful, wise and good. A major part of his True Intellectual System is taken up with the demonstration of the existence of God, largely through consensus gentium *(universal consent) arguments and the argument from design. The intellect behind his 'intellectual system' is the divine understanding. Mind is antecedent to the world, which is intelligible by virtue of the fact that it bears the stamp of its wise creator. The human mind is capable of knowing the world since it participates in the wisdom of God, whence epistemological certainty derives. The created world is also the best possible world, although not bound by necessity. A central element of Cudworth's philosophy is his defence of the freedom of will - a meaningful system of morals would be impossible without this freedom. Natural justice and morality are founded in the goodness and justice of God rather than in an arbitrary divine will. The principles of virtue and goodness, like the elements of truth, exist independently of human beings. A Treatise Concerning Eternal and Immutable Morality contains the most fully worked-out epistemology of any of the Cambridge Platonists and constitutes the most important statement of innate-idea epistemology by any British philosopher of the seventeenth century.*

1 Life

Cudworth was born in England at Aller, Somerset. Educated at Emmanuel College, Cambridge, he was elected to a fellowship in 1639. During the Civil War, in the wake of the Parliamentary purge of Cambridge dons, he was appointed Master of Clare Hall and Regius Professor of Hebrew. In 1654 he became master of Christ's College, a position which, along with the Regius Professorship, he retained at the restoration of the monarchy and occupied until his death. Cudworth was also one of the most learned classicists of his day. Although his religious background was in predestinarian Calvinism, he rejected his theological roots in favour of a more Arminian theology which stressed the role of reason in religion. This antivoluntarist stance went hand in hand with a tolerant outlook and an interest in millenarianism. He was consulted by Cromwell on the re-admission of the Jews to England.

Cudworth's philosophical antideterminism is the counterpart of his theological antivoluntarism. Common to both is his conception of the deity, which emphasizes divine goodness, wisdom and justice rather than the divine will. His antivoluntarism has fundamental philosophical implications. First, voluntarism leads to scepticism: if truth depends on God's arbitrary will, there can be no certainty. Second, if the will of God determined all human actions, then human beings would not be responsible for their actions and could not act morally. Furthermore, if the arbitrary will of God decided what was right and wrong, there could be no principles of morality. Voluntarism is the theological counterpart of the false intellectual systems of the philosophical determinists (see Voluntarism).

Cudworth was one of the first philosophers to write consistently in the English language. His philosophical reputation rests on a group of incomplete or posthumously-published treatises. The only philosophical work published in his lifetime was his *True Intellectual System of the Universe* (1678). This was originally conceived as part of a larger work on liberty and necessity, which he never completed. His posthumously published *Treatise Concerning Eternal and Immutable Morality* (1731) was originally intended as the epistemological prolegomenon to a projected treatise on ethics (never completed) of which *A Treatise of Free Will* (1838) was intended to form part. The latter prints one of three discussions of the subject which exist in manuscript. Both posthumously

published works of Cudworth as well as the unpublished writings develop philosophical themes sounded in his *True Intellectual System*.

2 *The True Intellectual System of the Universe*

The larger part of Cudworth's *True Intellectual System* is taken up with *consensus gentium* arguments for belief in a supreme deity. Cudworth marshals his immense classical erudition to demonstrate that belief in a deity is natural to humans, by showing that all humans at all times have had some kind of belief in a deity. Cudworth's survey of classical writings was also intended to vindicate corpuscularian natural philosophy from the charge of that it is intrinsically atheistic. Although Cudworth's *System* appears antiquarian, the arguments are relevant to the philosophy of his time and the whole work is underpinned by a concept of philosophy as a unified, timeless whole which has been the same since the beginning of time and which is concerned to discover the same single truth. This *philosophia perennis*, or perennial philosophy, derives from Renaissance models of the history of philosophy propounded by figures such as Agostino Steucho. It is dependent on syncretic and selective readings of philosophical texts, concentrating on noting parallels in doctrines rather than on analysing individual systems. On such a model, contemporary philosophy is not a new development but a recent manifestation of doctrines and arguments for which there are many analogues in antiquity.

For Cudworth, true philosophy is a combination of atomistic, mechanistic natural philosophy and a metaphysics concerned with the immortality of the soul and the existence of God. But this philosophy, originally propounded by Moses, has been corrupted over time. Cudworth distinguishes four distortions of the original atomism in four varieties of atheism: atomical atheism (according to which all things come about by chance), Hylozoic atheism (which imputes life to matter, associated with Strato of Lampsacus), Hylopathian atheism (which is merely materialistic, associated with Anaximander) and Cosmo-plastic atheism (which makes the world soul the highest *numen*). These schools of atheistic philosophy have their contemporary proponents: because he imputes the properties of spirit to matter, Spinoza is a Hylozoist, whereas Hobbes, by denying the existence of spirit, is a Hylopathian atheist.

Cudworth's taxonomy of atomistic philosophies is designed to distinguish between theistic and nontheistic forms of atomism. In this way he attempts to supply contemporary atomism with an impeccably theistic pedigree, clearing it of the imputation that all atomism is materialistic and therefore atheistic. True natural philosophy, he argues, leads inevitably to belief in God. Corpuscularian mechanism supports theism because it is impossible to explain movement, life or thought in terms of the properties attributed to material substance according to the mechanical philosophy, that is in terms of size, shape, position, motion and rest. Since these properties cannot by themselves account for 'Life and Cogitation, Sense and Consciousness, Reason and Understanding, Appetite and Will' (1678: 36), there must be some other cause, necessarily immaterial, namely a spirit or soul. Furthermore, the properties of matter may be deduced from the very idea of it. Matter is therefore intelligible without the useless and unintelligible scholastic apparatus of intentional species and substantial forms. In so far as Cudworth adopted the mechanical theory of matter, he was a modern in his natural philosophy.

3 Plastic Nature

In response to the shortcomings of the mechanical account of motion, Cudworth formulated his distinctive doctrine of Plastic Nature, a formative agency which acts as an intermediary between God and nature, maintaining the orderly day-to-day operations of the physical universe. The concept is indebted to the Platonic doctrine of *anima mundi*, and its functions are associated by Cudworth with the vegetable soul of Aristotelianism. Although he conceives of it as a spiritual agent, he insists that it operates unconsciously, performing the directives of God without question or foresight. Cudworth put forward his theory of Plastic Nature as an alternative to 'Fortuitous Mechanism' on the one hand, which explains all physical phenomena as the result of chance, and occasionalism, on the other, which requires God to intervene in even the most minute details of day-to-day natural occurrences (see Occasionalism). Occasionalist intervention is, in Cudworth's view, unbecoming to God, and it tends to atheism by undermining divine providence, rendering it 'operose, Sollicitous and Distractious'. Even when the mechanistic alternative to occasionalism admits the existence of a deity, it too tends towards atheism by reducing God to an 'idle Spectator'. The doctrine of Plastic Nature is thus designed to rescue divine providence. Its existence shows that nature is not the supreme *numen* but is subordinate to a Perfect Mind. Plastic Nature is the means whereby God imprints his presence on his creation, displaying his wisdom and goodness therein and

rendering it intelligible.

4 Epistemology

Contrary to what its title might suggest, *A Treatise Concerning Eternal and Immutable Morality* is largely a work of epistemology. Cudworth originally intended it to serve as prolegomenon to a work on ethics which he never completed, and of which *A Treatise of Freewill* constitutes a section on moral psychology. The fundamental ethical question posed in the overall work is that of Plato's *Euthyphro*: whether things are good because the gods will them, or whether the gods will them because they are good. The treatise is directed against the epistemology and moral relativism of Hobbes, though some of the discussion takes the form of a commentary on the views of Protagoras from Plato's *Theaetetus*. It is a fundamental tenet for Cudworth that moral values are really existing absolutes founded in the goodness and wisdom of God and that knowledge of moral principles is innate to the mind. He sets out the epistemological basis of this view by arguing that all intelligible things exist independently of human minds and even the world. Truth and the principles of knowledge are contained in the mind of God. Certain knowledge is possible because individual minds participate in the divine mind.

Cudworth's epistemology is constructed round the basic Platonic principles of archetype and ectype (form and copy): divine wisdom and knowledge is imprinted in individual minds as well as being reflected in the make-up of the physical world. Since knowledge is in this way innate to the mind, the process of cognition is an act of recollection. Here Cudworth employs the Platonic theory of *anamnesis*, as well as the Stoic concepts of *koinai ennoiai* (common notions) and *prolepsis* (anticipation) to develop the idea that the mind operates by a kind of foreknowledge or anticipation (see Plato §11; Stoicism §9). This a priori account of cognition accords with his fundamentally Platonic tenet that intellect precedes the world, ideas pre-exist things. But his theory of knowledge is not based on simple correspondence of ideas and things. Mind is not a passive receptor, but an active participant in cognition. Since truth is known from within the mind, not from without, sensory input is insufficient for understanding the external world. Sense data are received passively and piecemeal. Knowledge cannot be derived from them without a comprehending mind able apply 'conceptions or intelligible ideas' to sensory input. The mind focuses on the intelligible aspects of an object, the *scheses* or relationships between its parts, and so penetrates to its essence, understands its function. In the same way, the atomical philosophy can give a meaningful account of the nature of matter, undistracted by sense appearances. It is thus 'the triumph of reason over sense' because it 'solves all the phenomena of the corporeal world by those intelligible principles of magnitude, figure, site and motion, and thereby makes sensible things intelligible' (1731: 544).

5 Free will

In all probability Cudworth's unpublished writings on the subject of free will were a continuation of *Eternal and Immutable Morality*. The three surviving manuscripts on that subject overlap with one another in content, but none is complete (the first was printed in 1831, but the others contain important additions to the discussion). In these papers Cudworth argues that freedom of the will is an essential attribute of rational beings. Without free will moral responsibility would be impossible, and rational beings would be reduced to the level of mere machines. Cudworth conceives of free will as a 'ruling principle' or 'hegemonicon' of the soul which directs the actions of the individual being. More particularly, free will is a power of the soul which directs its actions towards the good. This conception of the beneficent partiality of the will leads Cudworth to repudiate any view that the will is indifferent. In so far as free will entails the power of furthering the perfection of a creature, it properly belongs only to beings capable of improvement. Consequently it is not an attribute of perfect beings and does not pertain to God. Furthermore, in so far as it may misdirect the soul towards spurious goods, it can lead to wrong doing. It is therefore, in Cudworth's phrase, 'a mungrell thing', since it combines the imperfection of the being to which it belongs and the perfection towards which that being aspires (see Free will; Will, the).

The soul is not constrained to act morally by external incentives or disincentives (such as rewards and punishments in the afterlife) but it contains within it the principles of virtue and vice, 'all moral habits' and conscience. This internal moral sense of the soul is superior to intellection and a token of the presence of the divine. Error is a consequence of the imperfection of created beings, who cannot always see the truth clearly and consequently often make judgments about action without clearly and distinctly perceiving the truth.

Cudworth's concept of free will is sustained by a unitary concept of the soul and its powers. The soul is a single,

uncompartmentalized unity, compounded of powers or principles, two higher and one lower. It is the function of the middle principle, or free will, to control the lower appetites and to direct the soul towards the highest principle, the *summum bonum*, just as the charioteer in Plato's *Phaedrus* controls the contrary impulses of his two horses to make them pull in the same direction. The term Cudworth uses to describe the exercise of the power of the controlling principle is 'conatus' (striving).

Cudworth's emphasis on the unitary function of the will, and its integration with the soul as a whole, enables him to side-step what he sees as the irresolvable dilemma of philosophers about the priority of reason over will, or will over reason. Furthermore, the integrative function of his 'middle' principle makes it possible to explain close interaction of mind and body, since the domain of the middle principle extends both to the lower, animal appetites and to the higher principles of the soul. In this respect his conception of free will enables him to avoid the problems associated with Cartesian dualism of mind and body (see Dualism).

6 Relation to Descartes

Cudworth kept abreast of contemporary developments in philosophy and in what we now call science. Indeed, he regarded his 'intellectual system' as the counterpart in philosophy of the physical world system of Copernicus and Galileo. Cudworth was a perceptive critic of contemporary philosophy. His negative assessment of Hobbes and Spinoza was related to his critique of determinism and materialism. His relationship to Descartes was more complex: his philosophy is indebted to Cartesianism in many fundamentals, but he was also critical of Descartes in important particulars. He took over Cartesian dualism of mind and body, and drew on Descartes' theory of the passions. He incorporated the Cartesian account of body into his own philosophy (regarding Descartes as a reviver of Democritean atomism). In epistemology he adopted the Cartesian principle of 'Clear Perceptibility or Intelligibility' as the criterion of truth and used it to underwrite his Platonist identification of being with truth: 'whatsoever is clearly and distinctly perceived to *Be, Is*'. None the less Cudworth put forward a number of criticisms of Descartes, all with sceptical implications. In his *System* (I.iv) he argues that Descartes' proof of the existence of God in the *Meditations* is circular - Descartes first employs the faculties of reason and understanding to prove that God exists, and then proves the truth of those faculties from the existence of God. The argument for the existence of God from the idea of necessary existence, is fallacious because it draws an absolute conclusion from a mere hypothesis. Cudworth offers a number of amendments to Descartes' arguments in an attempt to strengthen them (for example, replacing the argument from necessary existence by one based on possible existence). His other criticisms of Cartesianism are theological, though with important philosophical implications. He argues that Descartes' voluntarism leads to scepticism and his denial of final causes undermines providence. On both these counts Cartesianism is potentially atheistic and Descartes has not been true to his own principles (see Descartes, R. §6).

7 Influence

The weighty humanist erudition which encumbers *The True Intellectual System* is largely responsible for the fact that Cudworth is less well known than he deserves to be. But his philosophy continued to arouse interest in both Britain and the rest of Europe during the last part of the seventeenth and throughout the eighteenth centuries. Cudworth's British legacy remains to be charted in detail, but mention should be made of John Ray, Locke (his critique of innate ideas notwithstanding), Shaftesbury, Ramsay, Price and Reid. Thomas Wise produced a severely pruned edition of *The True Intellectual System* in 1703, designed to focus on theological issues of the early eighteenth century. *A Treatise Concerning Eternal and Immutable Morality* was published by Edward Chandler during the course of the eighteenth-century debate on ethical rationalism sparked by criticisms of Samuel Clarke. Thanks to J.L. Mosheim's Latin translation (1733), Cudworth's writings were known in Germany in the time of Kant. Through extracts published by Jean Le Clerc in the *Bibliothèque choisie* (1703-6), and the controversy between Le Clerc and Bayle which ensued, aspects of Cudworth's philosophy (especially his theory of Plastic Nature) entered enlightenment debates. The doctrine of Plastic Nature was defended by Paul Janet in France as late as the nineteenth-century (Simonutti 1993); and in 1823, Cudworth's philosophy was translated into Italian by Luigi Benedetti.

See also: Cambridge Platonism; Locke, J.; Neoplatonism

SARAH HUTTON

List of works

Cudworth, R. (1678) *The True Intellectual System of the Universe*, London; repr. ed. J. Harrison, London, 1845; repr. Stuttgart-Bad Cannstatt, 1964.(The only major work of philosophy by Cudworth to be published in his lifetime.)

Cudworth, R. (*c*. 1678-88) 'On Liberty and Necessity', British Library Additional MSS 4978-82.(Three separate treatises on free will, only one of which has been published. See *A Treatise of Freewill* (1838).)

Cudworth, R. (1731) *A Treatise Concerning Eternal and Immutable Morality*, London; repr. in *True Intellectual System*, 1845, vol. 2; ed. S. Hutton, Cambridge: Cambridge University Press, 1996.(First published posthumously by Edward Chandler, this contains Cudworth's epistemology.)

Cudworth, R. (1838) *A Treatise of Freewill*, ed. J. Allen, London; ed. S. Hutton, Cambridge: Cambridge University Press, 1996.(British Library Additional MS 4978, one of the three manuscript versions on the subject of free will by Cudworth.)

References and further reading

Aspelin, G. (1943) 'Ralph Cudworth's Interpretation of Greek Philosophy. A Study in the History of English Philosophical Ideas', *Gôteborgs Hôgskolas Arsskrift* 49: 1-47.(Cudworth's interpretation of Greek philosophy.)

Cassirer, E. (1932) *Die Platonische Renaissance in England und die Schule von Cambridge*, Leipzig and Berlin: Studiender Bibliothek Warburg; trans. J.P. Pettigrove, *The Platonic Renaissance in England*, Edinburgh: Nelson, 1953. (Discusses Cambridge Platonism in relation to European philosophy.)

Colie, R. (1957) *Light and Enlightenment*, Cambridge: Cambridge University Press.(On the Cambridge Platonists and the Republic of Letters, including Cudworth, Bayle and vitalism.)

Cragg, G.R. (ed.) (1968) *The Cambridge Platonists*, New York: Oxford University Press.(Selections from *Eternal and Immutable Morality*.)

Darwall, S. (1995) *The British Moralists and the Internal Ought*, Cambridge: Cambridge University Press. (Chapter 5 discusses Cudworth's moral philosophy, especially his writings on free will.)

Gregory, T. (1967) 'Studi sull'atomismo del seicento: III Cudworth e l'atomismo' (Studies on seventeenth century atomism: Cudworth and atomism), *Giornale critico della filosofia italiana* 46: 528-41. (Cudworth and Atomism.)

Hunter, W.B. (1950) 'The Seventeenth-century Doctrine of Plastic Nature', *Harvard Theological Review* 4: 197-213.(Includes discussion of Cudworth's theory of plastic nature.)

Hutton, S. (1995) 'Lord Herbert and the Cambridge Platonists', in S. Brown (ed.) *British Philosophy and the Age of Enlightenment, Routledge History of Philosophy*, vol. 5: 20-42. London: Routledge.(Includes a section on Cudworth's philosophy.)

Passmore, J.A. (1951) *Ralph Cudworth, an Interpretation*, Cambridge: Cambridge University Press.(The only monograph on Cudworth, this deals mainly with posthumous publications and manuscripts. Useful - if dated - bibliography, with incorrect attribution of some manuscripts to Cudworth.)

Popkin, R.H.(1992) 'Cudworth', in *The Third Force in Seventeenth-century Philosophy*, Leiden: Brill, 333-50(Cudworth and Scepticism.)

Rogers, G.A.J. (1988) 'Die Cambridge Platoniker' and 'Ralph Cudworth', *Ueberwegs Grundriss der Geschichte der Philosophie: die Philosophie des 17. Jahrhunderts* 3 (1): 245-6, 267-72, 285-90.(Includes a comprehensive and recent bibliography.)

Sailor, D.B. (1962) 'Cudworth and Descartes', *Journal of the History of Ideas* 23: 133-40.(Cudworth and Descartes.)

Scott, D. (1990) 'Platonic Recollection and Cambridge Platonism', *Hermathena* 149: 73-97.(On the theory of knowledge of Cudworth, More and Whichcote.)

Simonutti, L. (1993) 'Bayle and Le Clerc as readers of Cudworth. Elements of the debate on Plastic Nature,' in the Dutch Learned Journals *Geschiedenis van de Wijsbegeerte in Nederland* 4: 147-165.(Cudworth's fortuna in Enlightenment Europe.)

Cultural identity

If cultural identity means that a person achieves the fullest humanity within an accepted context of traditional symbols, judgments, values, behaviour and relationships with specific others who self-consciously think of themselves as a community, then it must be seen as a great contemporary challenge to many Western philosophical assertions about the person, society, meaning and truth.

This sense of cultural identity, as well as more extreme forms, can amount to classic determinism: individuals are subsumed under the relations with meaning and people surrounding them. Advocates of the centrality of cultural identity often make an argument rooted in an 'authenticity' which purports that, to fully encounter oneself, each of us must grasp and fully accept one's psychological and social location within a specific group of people who interpret life in terms of the particular civilization that contains them.

1 Philosophy and cultural identity

In the twentieth century Western philosophy has endured many assaults on the claims of some of its practitioners that it somehow constitutes a foundational discipline whose concerns are basic to any comprehensive analysis of what it means to be human. In some intellectual quarters the most formidable challenges arise from both scientific practice and the dominant models and images therein. This view is often found among analytic philosophers trained in the Anglo-American traditions. These traditions have emphasized the close study of 'linguistic claims' and sentences whose truth these philosophers seek to ascertain by appealing to logic and 'common language usage'.

Usually quite separate from that tradition have been other continental ones whose fundamental texts were first written in German, French, or more rarely for English-speaking students of philosophy, Russian, Italian, or Spanish. At least some of these European advocates have been influenced by Friedrich Nietzsche and other thinkers (including those commonly called philosophers but also literary figures, artists and social theorists) of an intellectual tradition descended from Nietzsche.

From such a Nietzschean perspective, concepts like cultural identity pose problems of the range and possibility of philosophy quite different from those of science. The Spanish thinker José Ortega Y Gasset has described himself as a human creature who is both 'myself and my circumstance'. In that duality Ortega Y Gasset demonstrates what it means to have become aware of context, or what some existentialists have called 'situationality'. Even when people think it is better to seek truth, they still exist within a situation of having been influenced both consciously and unconsciously by their circumstances in the world.

Thus issues of cultural identity can cause a crisis of legitimacy for any human activity which stresses the search for basic truths, fundamental realities and those thoughts or ideas which serve to ground our lives. If we are creatures with identities rooted in specific cultures, then the ancient Western idealist sense of the humanly universal is either suspect or defunct. This might be because we become products of cultural contexts so that we and the people we encounter across the world may be more fundamentally different than similar.

2 Community and belonging

The decline of empires in the twentieth century has contributed to new senses of this issue of the problematic commonality of individuals and societies. In recent centuries imperial Europeans, the people of the USA, as well as the Japanese, Chinese, Russian, Turkish, Incan, Aztec and various dominant tribes in other continents could confidently view other peoples as still being in need of protection, advancement, development and civilizing enlightenment. Late in this century in numerous areas of the world, the question of what it is to be fully human has arisen, with the conclusion that individuals must define themselves in terms native to the people or community to which they belong. Put radically, 'belonging' within such theories of cultural identity becomes the first condition for achieving full adulthood for members of our species. It is only when we acknowledge to which group we belong that we can discover who we are and what being human truly means.

This perspective has the effect of making most intellectual activity regional. Our symbols and their meaning essentially become products of those others who collectively give us our neighbouring human groups. This is a

process connected to psychology in personal and social senses: adult consciousness is a result of belonging to a group, or community. Self-image necessarily evolved from and continues to comprise elements derived from our personal situations in tandem with others.

Politically this appeal to the region of nearby others and their established meanings for our selfhood has the effect of determining to what group we belong within other neighbourhoods of the world and thus where else we can expect both to be welcome and express the best of ourselves. Such perspectives permit cultural identity to provide people both like and unlike ourselves with the certainty of boundaries. From a psychosocial position, these perspectives of cultural identity can be seen as safe and 'authentic', however, they can also be anti-cosmopolitan, anti-individualistic and relativistic. Their appeal may be seen on two levels: for the relatively powerless there is the possibility of successful initiation and nurture within a tradition of relations and symbolic expression; for the powerful within such a self-defining community there is a means to unify the group and establish both its continuity of power and ready-made 'outsiders' to serve whatever purposes might sustain that power as well as the group's position among other groups.

In the West, within the great democracies and among the ruins of the old empires, appeals to cultural identity have tended to be voiced in three ways. Some groups have insisted on the inevitable alienation their members experience when left as unintegrated, unhappy, unfulfilled, though nominal citizens of such states as the former Soviet Union, Eastern Europe and China. In countries like the USA, Canada, many European countries and other technologically advanced nations like Australia, New Zealand and Japan, some have demanded that the dominant culture admit the presence of and provide significant resources for (sometimes as restitution for past suppressions) groups either fully in possession of their cultural inheritances or struggling before the indifference or hostility of some members of dominant groups. Finally, in some developing nations a combination of two other tendencies has arisen: some communities or peoples wish to escape traditional control by the central metropolitan powers, while others want a better deal from those powers as they continue to be valued components of the nation. This mix can be found in countries such as India, Brazil, Mexico, Malaysia, Indonesia, the Philippines, Nigeria and Zaïre.

In terms of such topics, discussions of cultural identity are difficult to find outside issues arising from the remnants of empire and the countervailing gestures of liberation. In the context of theories of liberation from the old empires, some religious movements seem deeply committed to advancing accompanying cultural identities. Variants of Islam fit this model, as do evangelical Protestants in Catholic countries in Latin America.

More controversially, in the most developed areas of the world cultural identity has sometimes been transformed into a self-defining paradigm for the liberation of women, gays and lesbians. In these cases some have argued that a mix of roles, attitudes, values, personalities, interpersonal styles, language patterns, behaviour, ethics and other expressions constitute a separate origin for identity that is parallel in form to and more powerful than more traditional concepts such as ethnicity or race. From such perspectives being feminist or gay conveys a centrality or origin of personhood that is foundational to personal identity, which manifests itself coincidentally in what these theorists call a separate culture.

However, determinisms tend to crowd and frustrate people when the search for a fundamental essence becomes serious. Most advocates of personhood which focuses on cultural identity do not want to accept gender or sexual preference as competing forms of personal identity. Their usual position is to stress how these aspects of humanity bear out the evidence of the effects of particular cultures, or subcultures within larger social units like nation states and thus provide arenas within behaviour for seeing the power and separateness of cultural identity. Instead of stressing common forms of the feminine, such advocates choose to emphasize the differences of women's roles from one community to another.

Cultural identity, in its purer forms, should be seen as an assertion of an essentialist argument: we are what those among whom we have been raised have made us. We cannot find ourselves fully without embracing these community-based relations: from them we get a sense of what is important and definitional about our lives.

3 Alternative perspectives

There are at least three alternatives which oppose such notions: first, advocates of human generality, often expressed as 'universal humanity', stress intuitions and evidence that common structures arise routinely within the lives of people supposedly living within other, sometimes 'opposed' cultures. From such a perspective, placing a

stress on polarized cultural identities is tangential, narrow and by definition dehumanizing and thus dangerous. Second, cultural identity is a stance rooted more in the politics of the decline of empires than in anything foundational. The reservations inscribed in this stance might be that advocates of cultural identity practise a form of 'local politics' whereby they seek to take advantage of guilt arising from the legacy of empires and stigmatized hegemonic attitudes. This criticism emphasizes the political and social advantages accrued by the leaders of movements who wish to be compensated for past gestures directed at the affected group by dominant cultures and societies. The third opposition to previously discussed notions of community and belonging is that we are individuals whose lives are particularized by the accidents of space, time, biology and history in ways which are unique. Within the terms of this model, cultural origins are extremely important, but they themselves can become critical topics: their elements can become other for us and can be altered or disowned. Options from other human cultures can be substituted or adopted by us.

See also: Community and communitarianism; Persons

JOHN A. LOUGHNEY

References and further reading

Schutte, O. (1993) *Cultural Identity and Social Liberation in Latin American Thought*, Albany, NY: State University of New York Press.(A thorough analysis of the positive features of cultural identity in Latin America.)

Unamuno, M. de (1954) *The Tragic Sense of Life*, New York: Dover.(Argues that the person creates persona, only some of which stress relations to ancestors or neighbours as foundational to personhood.)

Weinstein, M.A. (1976) *The Tragic Sense of Political Life*, Columbia, SC: University of South Carolina Press. (Contains the argument that the individual is badly served by identification with, or belonging to a 'people', or nation-state.)

Weinstein, M.A. (1978) *Meaning and Appreciation*, West Lafayette, IN: Purdue University Press.(Describes human consciousness as existing prior to membership in a people or community.)

Culture

Culture comprises those aspects of human activity which are socially rather than genetically transmitted. Each social group is characterized by its own culture, which informs the thought and activity of its members in myriad ways, perceptible and imperceptible. The notion of culture, as an explanatory concept, gained prominence at the end of the eighteenth century, as a reaction against the Enlightenment's belief in the unity of mankind and universal progress. According to J.G. Herder, each culture is different and has its own systems of meaning and value, and cannot be ranked on any universal scale. Followers of Herder, such as Nietzsche and Spengler, stressed the organic nature of culture and praised cultural particularity against what Spengler called civilization, the world city in which cultural distinctions are eroded. It is difficult, however, to see how Herder and his followers avoid an ultimately self-defeating cultural relativism; the task of those who understand the significance of human culture is to make sense of it without sealing cultures off from one another and making interplay between them impossible.

Over and above the anthropological sense of culture, there is also the sense of culture as that through which a people's highest spiritual and artistic aspirations are articulated. Culture in this sense has been seen by Matthew Arnold and others as a substitute for religion, or as a kind of secular religion. While culture in this sense can certainly inveigh against materialism, it is less clear that it can do this effectively without a basis in religion. Nor is it clear that a rigid distinction between high and low culture is desirable. It is, in fact, only the artistic modernists of the twentieth century who have articulated such a distinction in their work, to the detriment of the high and the low culture of our time.

1 Definition of culture

In human history, culture in its broadest sense is that which is socially rather than genetically transmitted. It is that which children learn by virtue of their being brought up in one group rather than another, and, in its totality, it is that which distinguishes one human group from another. To human culture belong language, customs, morality, types of economy and technology, art and architecture, modes of entertainment, legal systems, religion, systems of education and upbringing, and much else besides; everything, in other words, by virtue of which members of a group endow their activities with meaning and significance. Even from the brief list of elements comprising a culture, it will be evident that there is no clear criterion for identity in the case of human culture. Cultures are characteristically permeable, evolving, open to influence from outside and inside in unpredictable ways, liable to be divided into subcultures, and to generate offspring with their own lives and development. And while individuals from a given culture are formed by it in all sorts of ways, conscious and unconscious, theoretical and practical, individuals are not prisoners within their cultures, but can affect them, react against them and contribute to their development.

2 Herder and the centrality of culture

In view of the all-encompassing nature of culture and of the multiple vaguenesses connected with the notion, one might be tempted to wonder whether it is a useful concept at all. That it is not just useful, but central to the analysis of history, was strongly urged by Herder (§4) in a sustained onslaught on Enlightenment thinking. Against the Enlightenment's conception of the unity of mankind and its belief in progress, Herder stressed the plurality of human societies and the incommensurability of their values (and hence the impossibility of talking of historical progress in any uncontested way). According to Herder, each of us is what we are because of the group to which we belong. Activity and self-expression are valuable in the degree to which they express the personality of the individual agent and that of the group to which they belong. Stressing the organic nature of human societies and the interplay between societies and cultures, Herder rejects centralization, bureaucracy, the elimination of cultural diversity and, above all, imperialism. In his *Briefe zur Beförderung der Humanität (Letters for the Advancement of Mankind)*, he asks:

> Can you name a land where Europeans have entered without defiling themselves forever before defenceless, trusting mankind? …Our part of the earth should not be called the wisest, but the most arrogant, aggressive, money-minded: what it has given these peoples (the colonies) is not civilization but the destruction of the rudiments of their own cultures wherever they could achieve this.

(Herder, quoted in Berlin 1976: 160-1)

The world in which each of us lives is a particular cultural inheritance, binding us to our forefathers and to our descendants, and distinguishing us from members of other cultures. Herder is clearly right to make what is, in effect, the distinction between causal explanations and explanations in terms of meaning. Human beings are not machines, and our activities cannot be accounted for in terms of scientific psychology; while we doubtless have animal needs and urges, our lives are lived for the most part in realms of meaning, even in pursuit of such elementary needs as food, shelter and sex. The human *Lebenswelt* is a world of *intelligibilia*, of norms and normativity, of activities regulated by standards and fashions and criteria demanding our allegiance, on pain of lapse into brutishness, outlawry and, ultimately, incoherence.

If Herder had done no more than draw attention to the importance of culture and meaning in human affairs, and to the equivocal nature of the concept of progress, there would be few who would dissent. His position, however, becomes more problematic when wrestling with the consequences he derives from it. Thus, Herder's dislike of nationalism and repudiation of the idea of a *Favoritvolk* leads him into a cultural relativism which has at least as many difficulties as the nationalist supremacism of successors such as Fichte (see Nation and Nationalism). At the same time, his insistence on the organic nature of culture and his rejection of cosmopolitanism (even if in the form of imperialism) may well have provided encouragement to the very nationalism he deplored. Certainly, it is hard not to see Spengler's celebrated distinction between civilization and culture as owing as much to Herder as to Nietzsche (who was himself no mean exponent of cultural essences). For Spengler, the very idea of a cosmopolis, a world city, is a symptom of decline from the higher state of culture.

3 Culture and civilization

For Spengler (who is in his own way elaborating themes from Nietzsche's *The Birth of Tragedy*), civilization is the movement away from the strong local and unquestioned bonds which constitute organic culture:

> In place of a type-true people, born of and grown on the soil, there is a new type of nomad, cohering unstably in fluid masses, the parasitical city-dweller, traditionless, utterly matter-of-fact, religionless, clever, unfaithful, deeply contemptuous of the countryman, and especially that highest form of countryman, the country gentleman.

(Spengler [1917-22] 1926-8: 32)

Along with the decline of organic culture comes science, rationalism, socialism, internationalism, a preoccupation with trade and luxury, wage disputes and football grounds; all, in other words, which is characteristic of modernity. Like Herder, Spengler does not see any of this as progress, although one wonders whether the deep-seated relativism to which both are committed entitles them to see it as decline either.

4 Cultural relativism

Cultural relativism, indeed, is the pathos of strong theories of culture. What in the hands of Herder and Spengler began as a reasonable reaction against the facile optimism and scientism of the Enlightenment and its offshoots, such as socialism and utilitarianism, can turn all too quickly into an enervating and disabling relativism. This relativism deprives itself of the resources to criticize barbaric societies, old and new. It is noteworthy that for all his invective against both Rome and what he calls its lack of soul, and against the twentieth-century world, Spengler is carefully even-handed in praising the Colosseum and 'the brown massiveness' of Roman brickwork, as well as the aero-engines, the precision lathes and the steel structures of his own day (Spengler 1917-22: introduction). Moreover, there is a worrying determinism in making culture too definitive of the group to which it belongs, in that any reception of outside influence is going to seem either impossible in appealing to values not recognized by insiders or undesirable in diluting the purity and distinctiveness of a culture; or, at different times, both before and after it has happened.

While with regard to culture neither Herder nor Spengler avoid the traps laid by relativism, determinism and prescriptivism, no important thinker who has seen the significance of culture has shown clearly how attendant dangers are to be avoided without lapsing into a kind of universalism which cannot accommodate any significant understanding of culture (see Relativism; Social relativism).

5 Common culture and high culture

So far, this entry has focused on what might be called the anthropological sense of culture: that memorably described by T.S. Eliot as involving all the characteristic activities and interests of a people (Eliot 1948). Speaking of English culture, Eliot lists Derby Day, Henley Regatta, Cowes, the Twelfth of August, a cup final, the dog races, the pin-table, the dartboard, Wensleydale cheese, boiled cabbage cut into sections, beetroot in vinegar, nineteenth-century Gothic churches and the music of Elgar. Eliot, being an American, was rather better positioned to describe *English* culture than most Englishmen. So much of a culture is traditional, unspoken, taken for granted, even unrealized by its adherents. It nevertheless gives form and meaning and moral direction to the experience of those who belong to it. It is partly because of its tacit dimension that theorists of culture fear that too much explicitness, too much reasoning and too much change will unbalance the common culture of a people, and so change or even subvert its identity. Why is the cabbage cut into sections? Isn't Wensleydale cheese hard to make conform to Euro-standards? Shouldn't the cup final be played so as to fit the demands of television? After not very much of this, the whole world, or so it seems, starts to dance to the tune of McDonald's and satellite TV.

After giving his list, Eliot goes on to enunciate what he calls 'the strange idea that what is part of our culture is also part of our *lived* religion' (Eliot [1948] 1962: 31; original emphasis). Eliot's own idea was that a people's culture is the incarnation of its religion, although as he concedes that materialism could be a religion in the required sense, it is hard to assess the force of this claim.

More characteristic of thinking about culture over the last 200 years is the claim associated first with Schiller and, in the English-speaking world, with Carlyle, Ruskin, Matthew Arnold and F.R. Leavis; namely, that culture or, more precisely, high culture can somehow substitute for religions. The thought is that in the absence of an acceptable dogmatic religion, people's highest aspirations and spiritual life are best expressed in works of art.

Leavis' continual complaint was that in so much of contemporary life we find 'no goals having any real authority except those capable of measurement' (Leavis 1972). In such circumstances, we lose our sense of the value and meaning and potential of life; there is 'no wealth, but life', as Ruskin put it, but we no longer have any sure sense of what wealth in life is, or how life should be lived to mine its wealth. The thought is that in some works of art, (in Leavis' view, such as in Eliot's poetry, in D.H. Lawrence and in Blake), we are brought up short against a harder reality than we normally encounter in the utilitarian, hedonistic, calculating world in which we spend most of our lives, and we glimpse possibilities of something higher or deeper.

Certainly Beethoven's symphonies and quartets, the painting of Turner and Van Gogh, the poetry of Rilke and the writing of Proust, are for many today epiphanic or revelatory in the way religion once was. Artists of this calibre seem to Leavis to form a league against the philistine, Benthamite world (which includes among its characteristic products most contemporary analytical philosophy). And art, in this highest sense, spreads messages more terrifying than the insipid 'sweetness and light' which Arnold thought could form a bulwark against anarchy.

It must, however, be a moot point as to whether any of the exponents of the arts mentioned above would have seen themselves as unreligious or in competition with religion as such, rather than as hostile to some very specific forms of religion. It is, in other words, a moot point as to whether art can uplift in the way Leavis and the other critics hope if all doors are closed against religious hope, or whether an explicitly humanistic, atheistical art would not be likely simply to reinforce the very materialism Leavis hopes literature will rebut.

Whether high culture can perform the role foisted on it or not, the very notion of a high culture is increasingly questioned, not least by contemporary artists and critics themselves. Eliot argued that hereditary class divisions were essential for a healthy culture and a healthy art, and however romantically we may look at the groundlings in Shakespeare's theatre or the whole city enraptured by Aeschylus, it is undeniable that there are elements in Shakespeare and Aeschylus which repay and demand the highest levels of attention, ability and concentration, levels clearly beyond the majority of any population known in history.

What, however, is not in doubt is that there have been times and places in which the majority of a culture have been capable of responding in some way to Aeschylus and Shakespeare; as also to Raphael or Rembrandt or Mozart or to the Authorized Version of the Bible. In the same way, 'high' artists, such as Bach, Schubert, Scott, Dickens, Hogarth and David, have drawn on popular forms and created popular forms themselves. The truth is that at least until this century, culture and art have operated at different levels, but in a way in which there were

continuities and overlaps. It is only with the giants of twentieth-century modernism - Schoenberg, Le Corbusier and, to a lesser degree, Picasso and Braque - that exponents of high culture have deliberately cut themselves off from their roots in the surrounding popular culture. The result has been a hermetic and self-referential high culture lacking popular appeal, and a popular culture with few values other than those of advertising and the mass media. The conclusion prompted by this observation is that in cultural matters, the important distinction is not so much that between high culture and popular culture as that between good cultural times and bad cultural times; and that just as good works can be found at both the high and the popular ends of the spectrum, so in bad times works low in ambition and achievement can be found in prestigious galleries and concert halls just as much as in the mass media.

See also: Ibn Khaldun

ANTHONY O'HEAR

References and further reading

Arnold, M. (1869) *Culture and Anarchy*, Cambridge: Cambridge University Press.(Puts the argument for culture's civilizing force.)

Berlin, I. (1976) *Vico and Herder*, London: Chatto & Windus.(An excellent and readable account of Herder.)

Barnard, F.M. (ed.) (1969) *J.G. Herder on Social and Political Culture*, Cambridge: Cambridge University Press.(Classic source on the importance of culture for the history of a people.)

Eliot, T.S. (1948) *Notes Towards the Definition of Culture*, London: Faber & Faber; repr. 1962. (Explores the complexity and necessity of the idea of culture. Referred to in §5.)

Herder, J.G. (1784-91) *Briefe zur Beförderung der Humanität (Letters for the advancement of Mankind)*, trans. T.O. Churchill in *Outlines of a Philosophy of the History of Man*, London, 1800.(Stresses the particularity of different cultures and the absurdity of looking for invariant laws of human behaviour.)

Leavis, F.R. (1972) *Nor Shall My Sword*, London: Chatto & Windus.(Defends a sense of culture against utilitarian consensus. Referred to in §5.)

Scruton, R. (1990) *The Philosopher on Dover Beach*, Manchester: Carcanet.(Much intelligent reflection on the subject matter in §5.)

Spengler, O. (1917-22) *Der Untergang des Abendlanders (The Decline of the West)*, trans. C.F. Atkinson, New York: Alfred A. Knopf, 2 vols, 1926-8.(Significant for its distinction between culture and civilization. Referred to in §3.)

Taylor, C. (1989) *Sources of the Self*, Cambridge: Cambridge University Press.(An excellent, although difficult, discussion of the whole area..)

Culverwell, Nathaniel (*c*.1618-*c*.1651)

Nathaniel Culverwell (or Culverwel) was one of the first natural law theorists in seventeenth-century England, and one of the first moral philosophers to stress the primacy of reason. His aim was to revive the natural law tradition of Aquinas and Suarez, which had fallen into disrepute in English Calvinism. Culverwell's theory is a synthesis of rationalism and voluntarism. It attempts to do justice to both the normative and coercive, to the moral and punitive aspects of law. The emphasis of his theory is, however, strongly rationalist, a reaction against the voluntarist legacy of Calvinism. Culverwell had close connections with some of the leading figures of Cambridge Platonism. He is not, however, a typical member of this school, because of the strong Calvinist strands of his early sermons.

1 Life and work

Not much is known about Nathaniel Culverwell's life. Born into a prominent family of Puritan clergymen, he was christened on 13 January 1619. In 1633 he was admitted to Emmanuel College, Cambridge, where he took his B.A. in 1636 and his M.A. in 1640. Emmanuel College was then notorious as a 'Puritan nursery', but it was also the cradle of Cambridge Platonism (see Cambridge Platonism). Among Culverwell's fellow undergraduates were Ralph Cudworth, John Smith and John Worthington, and it is probable that Culverwell's tutor was Benjamin Whichcote. Culverwell was made a Fellow of Emmanuel in 1642. He died young, probably in 1651.

Culverwell published nothing in his lifetime. His most important work is *An Elegant and Learned Discourse of the Light of Nature* (1652), edited by his colleague William Dillingham. This work was probably written during the university year 1645-6, when it was delivered as a series of sermons in Emmanuel College chapel. Culverwell also wrote eight sermons, which were published with the *Discourse*. The most important of these are 'The White Stone', whose theme is the need to make our election sure, and 'Spiritual Opticks', whose subject is the imperfection of our knowledge of God in this life.

2 Natural law

Culverwell's *Discourse* occupies an important place in the history of English moral philosophy. Appearing before the ethical works of Richard Cumberland, John Locke, Jeremy Taylor, Ralph Cudworth and Henry More, it is one of the first attempts in English Protestantism to stress the role of reason in ethics and to develop a doctrine of natural law. Like Richard Hooker, Culverwell wanted to revive the natural law tradition of Aquinas and Suarez, and to combat the nominalist legacy of Protestantism, which had reduced natural law to divine commands (see Aquinas, T. §13; Cudworth, R. §§4-5; Locke, J. §9; Suarez, F. §4).

The general purpose of the *Discourse* was to determine the proper relationship between reason and faith. Culverwell attempted to strike a middle path between the Socinians, who praised reason at the expense of faith (see Socinianism), and the radical sects, who demoted reason for the sake of the gospel. He maintained that although reason cannot discover or demonstrate the essential mysteries of the gospel - the Trinity, Incarnation and salvation - it can determine the law of nature, which is a reflection of the eternal law of God. Although Culverwell aimed at a *via media*, his chief goal was to defend the powers of reason. He wrote mainly to combat the rise of antinomian sects which taught that all moral laws had been abrogated by the gospel. Their teaching not only posed the dangers of immoralism and anarchism, but it also impugned Calvinism itself, which always stressed the role of grace over reason. In introducing natural law into Calvinism, then, Culverwell's intention was to clear it of the charge of antinomianism (see Calvin, J.).

The basics of Culverwell's doctrine of natural law closely follow Aquinas and Suarez. Like Aquinas, Culverwell holds that the law of nature is the participation of our reason in the eternal law of God. And, true to Suarez, Culverwell argues that the law of nature is essential to a rational being, and cannot apply to animals or nature in general, for 'where there is no Liberty, there's no Law, a Law being nothing else, but a Rational restraint, and limitation of absolute Liberty'. Although he admired the modern natural law theorists John Selden, Hugo Grotius and Claudius Salmasius, Culverwell criticized them because they presupposed but did not examine the general concept of nature. In explaining this concept, Culverwell again returned to the scholastic tradition, defining the nature of a thing as its essence, its characteristic manner of acting.

Although he has been described as a 'voluntarist', Culverwell attempts to avoid the extremes of both voluntarism

and rationalism (see Voluntarism; Rationalism). The foundation of his position is the scholastic distinction between the normative and coercive, or demonstrative and preceptive aspects of law. The normative or demonstrative aspect is its moral justification, the reasons for acting according to it; the coercive or preceptive aspect is the will that decrees it and the power that enforces it. Again following Aquinas and Suarez, Culverwell maintains that both these elements are necessary to an adequate concept of the law of nature. On the one hand, like all laws, the law of nature involves an act of will, a power of enforcement; hence, contrary to rationalism, there could not be a law of nature if God did not exist. On the other hand, the law of nature has a moral validity, a ground of obligation, independent of the will; contrary to voluntarism, then, nothing is right or wrong simply because God commands it. Although Culverwell stresses the importance of the divine will as the source of law, the emphasis of his theory is on reason, which had been underrated in the voluntarist tradition of Calvinism.

For Culverwell, the foundation of the law of nature rests upon the innate faculties of man, which express themselves in self-evident principles such as ‘do unto others as you would have them do unto you’. Culverwell denies, however, that we have pre-existing ideas and argues that only our faculties are innate. Unlike the Cambridge Platonists, he stresses that knowledge requires experience as much as the activity of reason.

3 Cambridge Platonism

Culverwell’s position within Cambridge Platonism has been the subject of dispute. Some regard him as a paradigmatic Platonist, while others place him outside this movement because of his Calvinism. Undoubtedly, there are deeper strains of Calvinism in Culverwell than in most of the Cambridge Platonists, so that it is misleading to regard him as typical of the movement. Some of the early sermons, especially ‘The Act of Oblivion’, ‘The White Stone’ and ‘Spiritual Opticks’, stress some standard Calvinist doctrines: predestination, the necessity of grace, the distinction between the elect and reprobate, and the inadequacy of human reason to understand salvation. These were doctrines that were decisively rejected by Whichcote, Cudworth and More in the 1640s. On the other hand, it is important to recognize that, in the 1640s, many of the rationalist strands of Cambridge Platonism were still only nascent. There are indeed strong Calvinist remnants in Whichcote, Cudworth and More, who stress the necessity of grace no less than Culverwell. Any sharp distinction between Culverwell and Cambridge Platonism therefore runs the risk of anachronism, reading the early years of Cambridge Platonism in the light of its later doctrines. The distinction becomes sharp only by overrating the rationalist strands, and underrating the Calvinist ones, in Cambridge Platonism.

It is more accurate to regard Culverwell as a transitional figure between Calvinism and Cambridge Platonism. It is also likely that Culverwell’s doctrines were in evolution, moving away from Calvinism and towards a greater rationalism, since there is a greater emphasis upon the value of reason in the *Discourse* than in the earlier sermons. It is a sad fact that, given his early death, we can only speculate about the final result of Culverwell’s intellectual development.

See also: Medieval philosophy; Natural law

FREDERICK BEISER

List of works

Culverwell, N. (1652) *An Elegant and Learned Discourse of the Light of Nature, with several other Treatises*, London: Rothwell.(The classic first edition, reissued in 1654, 1661 and 1669. The other treatises included, among others, ‘The White Stone’, ‘Spiritual Opticks’ and ‘The Act of Oblivion’.)

References and further reading

Brown, J. and Caird E. (1857) *The ‘preface’ and ‘critical essay’ to Of the Light of Nature: A Discourse by Nathanael Culverwel*, Edinburgh: Constable.(Still useful.)

De Pauley, W.C. (1937) *The Candle of the Lord: Studies in the Cambridge Platonists*, New York: Macmillan, ch. 7, 163-74.(A helpful analysis.)

Greene, R. (1971) ‘Introduction’ to the critical edition of An Elegant and Learned Discourse of the Light of Nature, *University of Toronto Department of English Studies and Texts* No. 17, ix-lv, Toronto, Ont.: University of Toronto Press. (Invaluable study of Culverwell’s context.)

Passmore, J. (1967) ‘Culverwel, Nathaniel’, *Encyclopedia of Philosophy*, ed. P. Edwards, New York: Macmillan, vol. 2, 276-7.(Useful, but exaggerates Culverwell’s distance from Cambridge Platonism.)

Tulloch, J. (1874) *Rational Theology and Christian Philosophy in England in the Seventeenth Century*, Edinburgh: Blackwood, vol. 2, 410-26.(Old but good.)

Cumberland, Richard (1632-1718)

Richard Cumberland developed his ideas in response to Hobbes' Leviathan. He introduced concepts of aggregate goodness (later used in utilitarianism), of benevolence (used in moral-sense theory), of moral self-obligation, of empirical proofs of providence and of the moral importance of tradition à lá Burke. The philosophical basis for Cumberland's views was a theory of natural law which was strongly anti-voluntarist and committed to objective moral values, but recognizing institutions such as governments of state and church as conventional or traditional. Cumberland was often seen as the third co-founder, with Pufendorf and Grotius, of modern natural law.

The English philosopher Richard Cumberland was educated at Magdalene College, Cambridge, and in several respects was close to the Cambridge Platonists as well as to Nathaniel Culverwell (see Cambridge Platonism). He subscribed to an extreme Erastianism that saw him through both the Restoration of 1660 and the Revolution of 1688-9. He was in fact on the far latitudinarian wing of the Anglican Church, and it is fitting that he was made Bishop of Peterborough as part of the Revolution settlement in 1691. His only philosophical work, *De legibus naturae*, appeared in 1672, the same year as Pufendorf's *De iure naturae et gentium* (see Pufendorf, S.). His other works were in such areas as patriarchal law, Jewish weights and measures, and Phoenician history.

The primary objective of Cumberland's philosophical work was to refute Hobbes by showing that morals has an objective basis and that it cannot simply be the imposition of a sovereign power (see Hobbes, T. §5). This objective basis is to be found in the common good, by which he means the good of the universal moral community of God and humanity through all times. By promoting this common good, we will inevitably promote our own good, but we in fact have a natural inclination (an active power) to pursue the common good, namely benevolence. This natural inclination can be regulated by reason through precepts.

Once we learn through experience, systematized in the science of morals, that benevolence is universal and thus part of God's will, we will realize that the rational precepts which we have developed are, in fact, natural laws prescribed by God; we will see that the natural inclinations to benevolence are also moral virtues; and we will understand that the common good is not just a natural good but a moral good .

Behind all this lies the common idea that morality is a matter of obligation to the law of a superior. Yet, Cumberland avoids a voluntarist view of law and obligation (see Voluntarism). God himself wills as law what is for the common good, but he obliges humanity to obey the law, not by his mere willing it nor by an appeal to the human rational calculation of pleasures and pains, but by an invitation to unify each person's particular will with the general will of God. God is the perfect intuitionist utilitarian who alone is capable of self-obligation, and humans can imperfectly take part in his unity of intuition and will. God's invitation to humanity - revelation apart - is expressed in the rewards and punishments that are naturally connected with compliance with, and transgression of, the law of nature. These are not, as a common utilitarian reading will have it, sanctions for or grounds of obligation to the law of nature; they are grounds of discovery of the law of nature (see Utilitarianism).

While his language is often vague, Cumberland clearly eschews hedonism as an account of the good, including the common good (see Hedonism). He seems to be aiming at a notion of self-perfection as the ideal good, but the perfection in question is, in turn, formalistically understood. It consists in an approximation to the general will of God which, if achieved, would establish a harmonious system of all the individuals making up the moral universe. Like his contemporary, Malebranche, but independently of him, Cumberland is thus coming close to a notion of universalizability (see Malebranche, N. §§5-6).

Cumberland's idea of the harmonious moral system of mutually interdependent individuals was a conscious attempt to transfer Descartes' theory of the full physical world, or *plenum*, to morals. It contributed to the subsequent age's development of an empirical science of morals and thus to the breach with Locke's idea of morals as a demonstrative science (see Descartes, R. §11; Locke, J. §9). The science of morals he regarded as a system of cohering individuals, but the mechanisms that made them cohere were a matter of empirical investigation. Cumberland thus joined Locke in rejecting innate ideas. But one of the empirical factors in accounting for the mind was teleological, namely the mind's possible understanding of, and associated striving for, the common good. By accounting for the universal moral system as a possibility, moral science acquired a

normative function. This role of moral science as self-fulfilling prediction, while rudimentary in Cumberland, became a lasting feature of British moral thought for more than a century.

As with so many natural lawyers, Cumberland has no strong theory of contract. The duties undertaken in contractual relations are, in so far as they are just, nothing but duties imposed by natural law, and rights are simply derivative from the duties. Contracts thus function only to make explicit what is implied by natural law. What is more, Cumberland is keen to acknowledge historically-given moral institutions. In language that often is strikingly Burkean, he maintains that such given institutions provide the best contribution to the common good simply because of the losses to be expected from major reform. Hand in hand with this, he points out that natural benevolence tends to establish social relations, including economic ones, that are independent of political society. When Dugald Stewart, early in the nineteenth century, maintained that Cumberland was an inspiration for the Scottish school of moral thought in the intervening century, he thus had a strong point.

See also: Natural law

KNUD HAAKONSSEN

List of works

Cumberland, R. (1672) *De legibus naturae*, London; trans. J. Maxwell as *A Treatise of the Law of Nature*, London, 1727, repr. New York, 1978; French trans. J. Barbeyrac as *Traité philosophique des loix naturelles*, Amsterdam, 1744.(Barbeyrac's translation has important editorial notes. Another English version is John Tower's 1750 translation, *A Philosophical Enquiry into the Laws of Nature*. James Tyrrell based his *A Brief Disquisition of the Law of Nature*, (London, 1692; rev. 1701; repr. Littleton, CO: Rothman, 1987) on Cumberland, but it does not give a clear idea of the latter's work.)

References and further reading

Haakonssen, K. (1997) 'The Character and Obligation of Natural Law According to Richard Cumberland', in M.A. Stewart (ed.) *Studies in Seventeenth-Century Philosophy*, Oxford: Clarendon Press.(Full interpretation, along the lines indicated above, of the coherence of Cumberland's thought.)

Kirk, L. (1987) *Richard Cumberland and Natural Law: Secularisation of Thought in Seventeenth-Century England*, Cambridge: James Clarke & Co.(The only modern monograph; strong on biography and textual problems.)

Schneewind, J.B. (1995) 'Voluntarism and the Origins of Utilitarianism', *Utilitas* 7: 87-96.(Sets Cumberland in the wider context indicated by the title.)

Sharpe, F.C. (1912) 'The Ethical System of Richard Cumberland and Its Place in the History of British Ethics', *Mind* 21: 371-98.(Classic study of Cumberland as proto-utilitarian.)

Cynics

Cynicism (originating in the mid-fourth century BC*) was arguably the most original and influential branch of the Socratic tradition in antiquity, whether we consider its impact on the formation of Stoicism or its role in the Roman Empire as a popular philosophy and literary tradition. The self-imposed nickname 'Cynic', literally 'doglike', was originally applied to Antisthenes and to Diogenes of Sinope, considered the founders of Cynicism, and later to their followers, including Crates of Thebes and Menippus. It emphasizes one of the most fundamental and controversial features of Cynic thought and practice - its radical re-examination of the animal nature of the human being. Their decision to 'play the dog' revolutionized moral discourse, since humans had traditionally been defined by their place in both a natural (animal → human → god) and a civic hierarchy. By calling such hierarchies into question, Cynicism re-evaluated the place of humankind in nature and the role of civilization in human life.*

*Cynicism includes an innovative and influential literary tradition of satire, parody and aphorism devoted to 'defacing the currency' (that is, the dominant ideologies of the time). It proposes a new morality based on minimizing creaturely needs in pursuit of self-sufficiency (*autarkeia*), achieved in part by physical training (*askēsis*), and on maximizing both freedom of speech (*parrhēsia*) and freedom of action (*eleutheria*) in open defiance of the most entrenched social taboos; and an anti-politics which sees existing governments as a betrayal of human nature, and traditional culture as an obstacle to happiness. In their place, Cynics advocated an immediate relationship to nature and coined the oxymoron* kosmopolitēs *or 'citizen of the cosmos'. However the literary, ethical and political elements of Cynicism are interrelated, all are most easily defined by what they oppose - the inherited beliefs and practices of classical Greek civilization.*

The virtual loss of all early Cynic writings means that the history of Cynicism must be reconstructed from much later sources dating from the Roman Empire, the most important of which is Diogenes Laertius (third century AD*).*

1 The early Cynics

While ancient tradition consistently viewed Antisthenes (*c*.445-365 BC) as the first Cynic and Diogenes of Sinope as his pupil, modern scholarship has systematically questioned this view on chronological, numismatic and other grounds. What is at stake is nothing less than the relationship of Cynicism to Socrates, since Antisthenes was a prominent pupil of Socrates (present at his death) as well as of Gorgias the rhetorician.

Antisthenes wrote a great deal and, unlike later Cynics, his interests included theoretical as well as literary and ethical topics. It is probably more accurate to see him as an important forerunner (rather than as a founder) whose teachings provided Diogenes' practice with some basis in theory. Particularly relevant would be his beliefs that virtue (seeAretē) 'is a matter of deeds and does not needs lots of discourses and learning' and that it is sufficient for happiness 'since happiness requires nothing else except the strength of a Socrates'. If Xenophon follows Antisthenes in his representation of Socrates, then Antisthenes must have laid particular stress on Socrates' 'self-mastery' (*enkrateia*). In his *Symposium* Xenophon represents Antisthenes as actually praising poverty, which certainly resonates with Cynicism. The biographer Diogenes Laertius says that Antisthenes provided the model for Crates' 'self-mastery' and Diogenes' 'imperturbability' (*apatheia*), which he learned by imitating Socrates, thereby inaugurating the Cynic way of life.

Whatever his relationship to Antisthenes, Diogenes of Sinope (412/403-324/321 BC) was the paradigmatic Cynic of antiquity. While Diogenes probably produced written works that do not survive, including a *Republic*, for us his life and thought are inseparable, since the latter is conveyed almost entirely by a tradition of biographical anecdotes that was over 500 years old by the time of our primary source, Diogenes Laertius. Diogenes Laertius preserves over 150 sayings or anecdotes that purport to quote Diogenes verbatim, as well as some brief paraphrases of his philosophy (VI 70- 3), which may be even less reliable in so far as they reflect Stoic and other influences. Their historicity is always problematic, but their representation of Diogenes is the most valuable evidence we possess. Almost one in six of the anecdotes involves some form of word play or allusion to poetry. The form in which Diogenes is transmitted has two crucial features that a summary of his teachings might fail to convey. First, most of the anecdotes involve Diogenes responding wittily to some challenging question or circumstance: he is shown as a satiric provocateur as well as a heterodox moralist, and both sides of his nature

need to be kept in view; if what Diogenes represents is reduced to a simple, practical morality much that made him influential is lost. Second, the biographical form necessarily makes the philosopher the embodiment of his thought; this unity of theory and practice was to become an explicit principle of Cynicism and an important source of Diogenes' authority as a philosopher. It made him (like Socrates) emblematic of the philosophic life in antiquity.

If the anecdotes, and the metaphors and slogans they inspired, are the core of ancient Cynicism, Diogenes' philosophy begins with the fact of his exile from Sinope for (literally) defacing the local currency. When reproached for his exile, he replied: 'But that was how I became a philosopher, you miserable fool!' Asked what good he had got from philosophy, he replied: 'If nothing else, then to be prepared for every kind of luck.' Diogenes describes himself by quoting a fragment from an unidentified tragedy:

Without a city, without a house
without a fatherland
A beggar with a single day's bread.

This is the starting point of Cynicism: it is, among other things, an attempt to show that we are so constituted by nature that, given proper training (*askēsis*), happiness is possible under the most adverse conditions.

It is tempting to see Diogenes as transforming into an act of conscious defiance the exclusion he suffered involuntarily when forced into exile. This defiant stance became his 'philosophy' in a concerted attempt to demonstrate by his own example that happiness does not depend on society or on any contingent circumstance, the domain of Fortune, but wholly on the autonomous self - a self brought into existence, not by years of laborious text-based studies such as those advocated by other philosophers, but by Cynic discipline based on exemplary acts and corporeal training. This is why Cynicism came to be called a 'shortcut to virtue'.

The Cynic shortcut 'defaced' the value philosophers attached to theoretical disciplines as well as the conventional value society attached to such externals as money, status, family and political power. (Yet the most influential Hellenistic schools - Stoicism and Epicureanism - would follow the Cynics in arguing that the happiness of the sage *should be* independent of context.) Thus Diogenes' life becomes an extended counter-example meant to refute his critics, both popular and philosophical, as he turns the cause of his exile into a metaphor for his activity as a philosopher - 'defacing the currency' of conventional wisdom. 'Defacing' took the form of literary parody and satire, provocative acts of free speech (*parrhēsia*) and public exhibitions of the Cynic way of life. The laughter they were meant to provoke is an indispensable element of Diogenes' practice and served to reinforce his independence by distancing him from the rules that everyone else obeyed.

2 Diogenes' conception of the human being

If Diogenes differed from his contemporaries on the sufficient conditions for happiness, it is because they differed in their conception of the human being. Some of the most famous anecdotes dramatize Diogenes' rejection of the prevailing conceptions of the human, such as when he lit a lamp in broad daylight and walked around saying 'I'm looking for a human being', or when Plato defined a human being as a 'featherless biped' and Diogenes produced a plucked chicken, saying, 'Here is Plato's human being'.

Diogenes' positive conception of human nature is given, as always, by his own example, not by definitions. From it we can infer that for the Cynic human beings are animals who have much to learn about freedom and self-sufficiency from their fellow creatures. As well as offering ethical instruction, the use of animals as examples served to illustrate the intrinsic superiority of nature to culture. Theophrastus reports that Diogenes discovered how to adapt to circumstances by carefully observing a mouse. Such adaptability enabled him to test the limits of his species by living like a dog in an abandoned wine-jar (*pithos*). A string of anecdotes develop different aspects of the canine metaphor: among Diogenes' canine characteristics was a shameless indifference to social norms, which enabled him to discard traditional morality based on shame in favour of living naturally and freely - that is, fulfilling his essential creaturely needs for food, sex and shelter without regard for the restraints and prohibitions of culture. Yet unlike sophists such as Callicles, the Cynics never used nature to sanction dominating others (although theft was another matter). Clearly, the Cynic rejection of shame cannot be reduced to a matter of manners as opposed to morals: Diogenes' pursuit of life according to nature led him to question the rational basis of such fundamental dietary and sexual taboos as those against cannibalism and incest, citing examples from

nature and non-Hellenic cultures.

While advocating a rigorous physical training intended to inure one to inevitable hardship, Diogenes showed no aversion to pleasures compatible with his way of life. Indeed, he used ‘any place for any purpose’, exemplifying the Cynic conception of freedom, which applied to sexual activity - from free love to public masturbation - as well as to eating, sleeping and talking. But the ultimate aim of imitating the freedom and self-sufficiency of animals is neither instinctual satisfaction nor mere ‘imperturbability’ (*apatheia*), but the Olympian independence of the gods, who, ‘needing nothing’, live easily free of mortal cares. Hence, Heracles, the one mortal to reach Olympus by triumphing over his famous labours, was adopted and re-interpreted by the Cynics as their mythical prototype. Diogenes explicitly claimed that their life had ‘the same character’ because both ‘deemed nothing more important than freedom’.

Of course Diogenes’ Cynicism is full of paradoxes. While teaching life according to nature and denouncing existing social arrangements, Diogenes went on living in cities, which he made no attempt to reform. While advocating freedom and self-sufficiency as paramount values, he both practised and advocated begging for a living. We must resist the temptation to turn a remarkable experiment, which challenged the most fundamental ideas of Greek civilization, into a mere system.

While Diogenes probably had no pupils as such, his example (whether conveyed orally or by written works) effectively established a philosophical model that invited imitation and interpretation. In the absence of a systematic body of doctrine, the nature of Cynicism was always up for debate; all the Cynics of antiquity as well as those who represent them (for example, Lucian, Dio Chrysostom) were of necessity actively engaged in interpreting the tradition, whether by word or deed.

3 Diogenes’ successors

The most influential Cynic in antiquity after Diogenes was Crates of Thebes (*c*.368-*c*.283 BC), a wealthy landowner and therefore at the opposite end of the social spectrum from a poor exile like Diogenes. He was a hunchback and several anecdotes refer to his comic appearance. He married Hipparchia of Maronea, who, along with her brother Metrocles, became a practising Cynic and, as such, the most famous female philosopher of antiquity. Their Cynic marriage (*kynogamia*), based only on mutual consent, was consistent with Diogenes’ views but radically at odds with Greek custom. The tradition holds that Hipparchia adopted the simple Cynic garb of Diogenes - a rough cloak, knapsack and staff - and lived on equal terms with her husband, attending events usually reserved for men and successfully defending her decision to pursue philosophy instead of weaving. Crates and Hipparchia were also notorious for living and sleeping together in public places, Cynically indifferent to shame and public opinion. It was Crates who described the fruits of philosophy as ‘a quart of beans and to care for naught’.

However Crates came to know Diogenes, his life was a remarkable application of the Cynic’s principles. He clearly regarded himself as a follower, calling himself a ‘fellow citizen of Diogenes’. There are several (probably fictitious) accounts of how Crates became a Cynic, which revolve around the fact that he evidently sold all his possessions and gave the proceeds to his fellow citizens, thereby embracing poverty, as had Diogenes. In contrast to Diogenes with his combative style and acerbic tongue, Crates was remembered as a benevolent figure and, thanks to his role as arbiter of family quarrels, actually revered as a household deity in Athens. But his fragments are clearly informed by a satiric (or seriocomic) perspective, the hallmark of Cynic discourse: ‘He used to say that we should study philosophy until we see in generals nothing but donkey-drivers.’

Crates was one of the most influential literary figures of the fourth century and his writings did much to disseminate Cynic ideology and establish parody as a distinctly Cynic mode of ‘defacing’ tradition. His *oeuvre* is notable both for its originality and its variety. He composed ‘tragedies’, elegies and epistles, and parodies such as his poem in hexameters entitled *Pēra (Knapsack)*, a hymn to frugality, a *Praise of the Lentil*, and an *Ephēmerides (Diary)*.

After Crates came Menippus (first half of the third century BC). The unreliable biographical tradition depicts him as a Phoenician slave who acquired his freedom by begging or usury and hanged himself when his business failed. Be that as it may, Menippus is among the most influential of Hellenistic authors. He is the only Cynic expressly called *spoudogeloios* (‘seriocomic’) in antiquity, and is credited with the invention of Menippean satire, a form of

narrative satire that parodies both myth and philosophy. The imitations and adaptations of his work by Varro (116-26 BC) and Lucian gave Menippean forms a long and influential afterlife in antiquity and the Renaissance, making Cynicism one of the primary sources of satiric literature in Europe.

Bion of Borysthenes (*c*.335-*c*.245 BC) also played an important role in early Cynicism, especially in the domain of literature. Tradition holds that he was sold into slavery as a boy but was bought by a rhetorician and received a rhetorical education. Later, he evidently received an eclectic education in philosophy at Athens: he studied with the Academics (under Xenocrates and Crates), with the Cynics, Cyrenaics and, finally, the Peripatetics. Bion probably originated the literary form of the diatribe - an argumentative monologue with imagined interlocutors, which was an important model for satirists and essayists of the Empire. Because Bion's Cynicism seems less radical and more opportunistic than that of Diogenes, it has sometimes been characterized as a 'hedonizing Cynicism'. References in Horace and other writers suggest an eclectic thinker with remarkable literary talents, as do his witty fragments.

Teles (*fl. c*.235) was a teacher and moralist who quoted extensively from such philosophers as Diogenes, Crates, Metrocles, Stilpo and Bion, his favourite authority and model. The surviving seven excerpts of his *Diatribes* are the earliest examples we possess of this influential Cynic tradition.

The works of Cercidas (*c*.290-220 BC), the last important early Cynic, take a surprising political turn. Cercidas was unusual for a Cynic in being a soldier, politician and lawmaker as well as a poet. He is best-known for his *Meliambi* (written in the Doric dialect), in which he uses a lyric form to mount a serious Cynic critique of contemporary politics.

4 Cynics in the Roman Empire

Whatever the causes of its apparent decline after Cercidas, when Cynicism re-emerges in the Roman Empire, it has changed, as has the world. In a series of confrontations under the emperors Nero, Vespasian and Domitian, Roman aristocrats with republican sympathies were put to death and the philosophers associated with them - both Stoics and Cynics - were banished from Rome. The best known Cynic of the period, Demetrius, a friend and hero of Seneca, is the most conspicuous example of the Cynic involvement in the 'philosophical opposition' to the emperor.

The emergence of Cynicism as a potent political ideology - opposed to hereditary monarchy - in the first century AD is highly significant but potentially misleading: most Cynics did not live in Rome nor in the West but in the Greek-speaking cities of the East from Athens to Alexandria. Most were not politically active but busily engaged in living the Cynic way of life - begging their daily bread and bearing witness to the example of Diogenes. Socially their status had been extremely mixed from the beginning, but as a rule they were not the associates of men of wealth and power such as Seneca. The sight of vagrant Cynics living in groups, travelling from city to city dressed in the Cynic garb, was not uncommon in the Empire. In the centuries since its inception Cynicism (alone among the philosophical sects) had become a popular movement attracting adherents from those strata of society outside the traditional audience for philosophy. It is impossible to quantify the movement, but more than eighty known Cynics have now been identified.

The most prominent Cynics of the Empire - Demetrius, Demonax of Cyprus (second century AD) and Peregrinus Proteus of Parium (*c*. AD 100-65) - were, unlike the early Cynics, teachers, not writers, and what we know of them is filtered through writers with their own philosophic interests, such as the Stoic Seneca and the satirist and sophist Lucian. Indeed, one of the central paradoxes of Cynicism's reception in the Empire is that the influence of Cynic ideology reached its apogee (in the second century AD) when very little in the way of original Cynic literature was being produced by practising Cynics. Oenomaus of Gadara (second century AD) is the only Cynic of the Empire known to us by his written work, *Charlatans Unmasked*, a lively, if not particularly original, attack on the veracity of oracles that survives because it is quoted by Eusebius, but does not seem to have made a great impression on Oenomaus' contemporaries. His other works, including 'tragedies' that scandalized the pious emperor Julian, are lost.

The only literature produced by Cynics in the Empire that we can set beside that of Oenomaus are the *Cynic Epistles*, a collection of fictitious letters attributed to the early Cynics and other sages. The authors of the letters are unknown and their dates of composition may vary considerably (from the third century BC to the second

century AD). The epistles offer an informative survey of the topics, slogans and anecdotes that must have fuelled many a 'diatribe', the term conventionally used to describe the oral performances of the Cynic street preachers so often remarked on by our sources. These oral performances were probably the primary means by which Cynic teachings were disseminated among the general populace. The *Cynic Epistles* confirm the impression that Cynic literary production in the Empire was no longer marked by the innovative parodies and polemics of the classical period and now served primarily to propagate Cynicism as a popular ideology and collective moral praxis.

We must remember, however, that our most important sources (other than Diogenes Laertius and the *Cynic Epistles*), the sophists Lucian (*c*. AD 120-80) and Dio (*c*. AD 40-111), the Stoic Epictetus (*c*. AD 55-120) and the emperor Julian (*c*. AD 332-61) probably had access to classic works by the early Cynics as well as a secondary literature that grew up around them. Both Dio and Lucian draw on these lost traditions to create a contemporary Cynic literature in which Diogenes and other legendary Cynics (Antisthenes, Crates, Menippus) appear as characters. Their works give us the liveliest images we have of what the lost Cynic classics might have been like and (along with Diogenes Laertius) are among the primary means whereby Cynicism became part of the philosophical and literary culture of Europe.

Lucian's use of Cynic traditions is too complex to characterize briefly, but Cynic ideology and the example of the Cynic classics were indispensable to him: they gave him a licence to satirize all things Greek, which now of course included Cynics and Cynicism itself. It is Cynicism as a radical form of cultural criticism - of 'defacing' the idols of the tribe through parody, satire and free speech - that led Lucian to adopt Cynic forms (for example, Menippean satire) and voices in so many works. But his *Life of Demonax* also shows a serious interest in using Cynic (and Socratic) traditions eclectically to construct a contemporary ethical model.

Similarly, Dio - who claims to have lived as a Cynic when in exile - uses the authority of Diogenes' persona in an important series of speeches to advocate Cynic values of particular importance to him as a courtier, particularly autonomy and freedom of speech. While lacking some of the Cynic seriocomic qualities that Diogenes has in Diogenes Laertius or Lucian, Dio's Diogenes is, on the whole, surprisingly true to his persona in the anecdotal tradition. Epictetus, by contrast, offers a distinctly idealized, quasi-religious account of Cynicism as a philosophical vocation and practical philosophy. His emphasis on praxis invites comparison with that of the *Cynic Epistles*. Julian echoes some of Epictetus' religious tendencies while arguing that Cynicism is a universal philosophy founded not by Diogenes nor Antisthenes, but by Apollo and based on the dictum 'Know thyself'.

Yet despite the obvious affinities all these authorities have for Cynic traditions, all denounce contemporary Cynics in the harshest terms as ignorant impostors. It is clear that the Cynic movement had split along the lines of class, wealth and education - the very distinctions Cynicism sought to annul. What we hear in the denunciations of our sources are the traces of a heated argument over the Cynic legacy. What is the ultimate value of Cynicism? As a practical ascetic morality for the have-nots, as a universal ethical model of freedom and autonomy, or as a cultural practice devoted to 'defacing' the false values of the dominant culture? Who has the right to speak out in the name of Diogenes? It is an argument that would resound from the humanists of the Renaissance to the *philosophes* of the eighteenth century, when Diderot wrote his Encyclopedia article on the Cynics and his Cynic masterpiece, *Rameau's Nephew*, exploring the relation between Cynicism, cynicism and the nature of enlightenment.

See also: Ariston of Chios; Socratic schools; Zeno of Citium

R. BRACHT BRANHAM

References and further reading

Billerbeck, M. (ed.) (1991) *Die Kyniker in der modernen Forschung (The Cynics in Modern Research)*, Amsterdam: Grüner.(Includes a selection of modern studies, 1851-1980, and a complete bibliography.)

Branham, R.B. and Goulet-Cazé, M.-O. (eds) (1996) *The Cynics: The Cynic Movement in Antiquity and its Legacy*. Berkeley, CA: University of California Press.(Surveys the history of the movement and its reception; includes a comprehensive catalogue of the ancient Cynics and an annotated bibliography.)

Dio Chrysostom (*c*. AD 100) *Discourses*, vol. 1, trans. J.C. Cahoon and H. Crosby, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1932.(Parallel Greek text and English translation; Diogenes is represented in Orations 4, 6, 8-10, but Cynic (and Stoic) thinking pervades his works.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London:

Heinemann, 1925, 2 vols. (Parallel Greek text and English translation; Book VI is our most important source for the Cynics.)

Downing, F.G. (1992) *Cynics and Christian Origins*, Edinburgh: T. & T. Clark. (Studies the Cynic influence on Christianity.)

Dudley, D.R. (1937) *A History of Cynicism: From Diogenes to the Sixth Century AD*, London: Methuen; repr. Chicago, IL: Ares,1990.(The best general introduction to the ancient Cynics.)

Epictetus (*c*. AD 50-*c*.120) *Discourses and Enchiridion*, trans. W.A. Oldfather, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925-8, 2 vols.(Parallel Greek text and English translation; Cynicism is discussed at III 22, IV 8.30-43.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae (The Fragments of Socrates and the Socratics)*, Naples: Bibliopolis, 4 vols.(Ancient sources in Greek and Latin, with discussion in Italian; includes all the documentation relevant to the Cynics.)

Goulet-Cazé, M.-O. (1986) *L'Ascèse Cynique: Un commentaire de Diogene Laërce VI 70-71 (Cynic Asceticism: A Commentary on Diogenes Laertius VI 70-71)*, Paris: Vrin.(Analyses the central ideas of Diogenes' ethics in the context of the Socratic tradition.)

Goulet-Cazé, M.-O. (1990) 'Le cynisme à l'époque impériale' (Cynicism in the Imperial Era), in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: De Gruyter, II 36 4: 2,720-833.(A lucid survey of Cynicism in the Roman Empire.)

Goulet-Cazé, M.-O. and Goulet, R. (1993) *Le cynisme ancien et ses prolongements (Ancient Cynicism and its Influence)*, Paris: Presses Universitaires de France.(A valuable survey of current work on the Cynics based on an international colloquium (Paris 1991).)

Höistad, R. (1948) *Cynic Hero and Cynic King: Studies in the Cynic Conception of Man*, Lund: Carl Bloms.(A useful introduction vitiated by its rejection of the anecdotal tradition.)

Julian (*c*. AD 140-80) *Discourses*, trans. W.C. Wright, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1913, 2 vols. (Parallel Greek text and English translation; *Orations* 6, 7 and the *Letter to Themistius* treat the Cynics.)

Kindstrand, J.F. (1976) *Bion of Borysthenes: A Collection of the Fragments with Introduction and Commentary*, Uppsala: Acta Universitatis Uppsaliensis, Studia Graeca Uppsaliensa.(Greek text, with commentary in English.)

Kindstrand, J.F. (1980) 'Demetrius the Cynic', *Philologus* 124: 83-98.(Short study of Demetrius.)

Lucian (*c*. AD 140-80) *Works*, trans. A.M. Harmon, K. Kilburn and M.D. Macleod, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1913-17, 8 vols.(Parallel Greek text and English translation; important works for the Cynics include *Demonax*, *The Cynic*, *Peregrinus*, *The Runaways*, *The Fisherman*, *Zeus Refuted*, *Philosophers for Sale! The Double Indictment* and *The Dialogues of the Dead*. Also relevant are *Nigrinus*, *Hermotimus*, *Symposium*, *On Sacrifices*, *On Mourning* and *Zeus the Tragic Actor*.)

Malherbe, A.J. (1977) *The Cynic Epistles: A Study Edition*, Atlanta, GA: Scholars Press.(Parallel Greek text and English translation of most of the Cynic letters.)

Marshall, P. (1922) *Demanding the Impossible: A History of Anarchism*, London: Harper Collins.(Places Cynicism in the history of anarchism.)

Niehues-Pröbsting, H. (1979) *Der Kynismus des Diogenes und der Begriff des Zynismus (The Cynicism of Diogenes and the Modern Concept of Cynicism)*, Munich: Wilhelm Fink.(The best account of the whole tradition (ancient and modern) in one volume.)

O'Neil, E.N. (1977) *Teles: The Cynic Teacher*, Missoula, MT: Scholars Press.(English translation of Teles' fragments.)

Paquet, L. (1988) *Les Cyniques grecs: Fragments et témoinages*, Ottawa, Ont.: Les Presses de l'Université d'Ottawa.(An annotated French translation of the main Cynic fragments and testimonia.)

Relihan, J.C. (1992) *Ancient Menippean Satire*, Baltimore, MD: Johns Hopkins University Press.(The best history of the genre.)

Sayre, F. (1948) *The Greek Cynics*, Baltimore, MD: Furst.(Often eccentric, but surveys much of the ancient evidence in English translation.)

Seneca (*c*. AD 52-62) *On Benefits,* in *Moral Essays*, vol. 3, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1920.(Parallel Latin text and English translation; VII 1.4-7 summarizes Demetrius' teaching.)

Seneca (*c*. AD early 60s) *Letters*, trans. R.M. Gummere, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1917- 25, 3 vols.(Parallel Latin text and English translation; Cynic (and Stoic) themes abound; see letters 20, 24, 47, 62, 67, 90, 91.)

Sloterdijk, P. (1988) *Critique of Cynical Reason*, trans. M. Eldred, London: Verso. (An influential interpretation of Cynicism focusing on its reception in Germany and its significance for critical theory.)

Xenophon (*c*.370s BC) *Symposium*, trans. O.J. Todd, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1923.(Parallel Greek text and English translation.)

Cyrenaics

The Cyrenaic school was a Greek philosophical school which flourished in the fourth and early third centuries BC*. It took its name from the native city of its founder, Aristippus of Cyrene, a member of Socrates' entourage. His most important successors were his grandson, Aristippus the Younger, and Theodorus, Anniceris and Hegesias, the heads of three separate Cyrenaic sects.*

The basis of Cyrenaic philosophy is physiological and psychological. It focuses on the individual feelings of pleasure and pain which are classed as pathē*, experiences produced in a subject by its contact with an object. They are described, respectively, in terms of smooth and rough movements, of the flesh or of the soul. A third category of* pathē*, described as intermediate between pleasure and pain, is also defined as movements and related to one's perception of individual properties or qualities. All* pathē *are short-lived and have no value beyond the actual time of their occurrence.*

These physiological characteristics are encountered both in the ethics and in the epistemology of the school. Although the Cyrenaics differed in their ethical doctrines, all of them attributed a central role in their systems to the individual bodily pleasure experienced in the present moment, and some of them considered it the moral end: it is pursued for its own sake, whereas happiness, conceived as the particular collection of pleasures that one experiences during a lifetime, is sought for the sake of its component pleasures. The goodness of individual pathē *of pleasure is supported by an elaborate epistemological doctrine whose central claims are that we are infallibly and incorrigibly aware of the occurrence and content of our own* pathē*, but that we cannot apprehend the properties of external objects. A striking feature of this doctrine is the neologisms designating the perception of qualities, such as 'I am whitened' and 'I am affected whitely'. This, and other features of Cyrenaic subjectivism, anticipate some modern philosophical analyses of subjective experience.*

1 History of the school

The Cyrenaic school was founded by a close associate of Socrates, Aristippus of Cyrene (see Aristippus the Elder), after Socrates' death (399 BC) and probably after Aristippus' numerous travels. For three successive generations its scholarchs (school heads) were members of the same family: Aristippus was succeeded by his daughter Arete (born *c*.400 BC) and then by his grandson, Aristippus the Younger (born *c*.380 BC). Our sources mention also Antipater of Cyrene, Aethiops and, later, Epitimides and Paraebates, but almost nothing is known about their careers as scholarch. Aristoteles of Cyrene apparently belonged to the same generation as Aristippus the Younger. At the beginning of the Hellenistic period (323-31 BC), the Cyrenaics Theodorus, Anniceris and Hegesias each founded a separate sect. Each sect was identified by the name of its leader ('Theodorians', 'Annicerians', 'Hegesians') but also preserved its Cyrenaic affiliation. It seems likely that, during the same period, there was yet another sect bearing the original name of the school and claiming to represent Cyrenaic orthodoxy. The Cyrenaic sects disappeared in the second half of the third century BC, some time between Arcesilaus' death and Chrysippus' appointment as head of the Stoa.

Of the Cyrenaics' writings, only some titles survive. Aristippus the Elder is credited with several Socratic dialogues on various ethical themes and with a three-volume history of Libya. Theodorus was the author of an influential treatise entitled *On the Gods* which attempted to refute traditional Greek theological beliefs. Hegesias composed a book entitled *The Self-Starver* enumerating the calamities of life, which may have aimed to procure an empirical basis for his moral pessimism.

Some sources maintain that the Cyrenaics devoted themselves exclusively to ethics, disallowing physics and logic as branches of philosophy. However, there is good evidence that they accepted logic as a tool of philosophical argumentation, and that they developed epistemology as a separate subject of inquiry, not as part of ethics or logic.

2 Epistemology

The origins of Cyrenaic philosophy were ethical, the epistemology probably being developed in the third generation of the school by Aristippus the Younger. However, the founder of the school laid the physiological and psychological foundations of the Cyrenaic theory of knowledge, mainly by his analysis of the notions of pleasure and pain, which was endorsed by his successors.

Pleasure and pain are *pathē*. The term belongs to a group of words related to the Greek verb *paschein* (to undergo) and refers to effects upon a subject, usually caused by an external object. The word can be used both regarding an inanimate substance and regarding a perceiving subject: a stone exposed to the sun may turn hot, or a person in a warm room may feel hot. Although the Cyrenaics focused on the *pathē* in connection with perceivers, their analysis preserves physicalist overtones. These are particularly obvious in the definitions of pleasure and pain as smooth and rough movements, of the flesh or of the soul, related to pleasurable and painful feelings. Some sources define that relation as an identity, but there is good evidence that the Cyrenaics drew a distinction between the physical motion associated with pleasure or pain and the subjective experience of *feeling* pleasure or pain. Since the motions associated with the *pathē* disappear with time, the *pathē* are short-lived. The memory of past pleasures and the anticipation of future ones is not really a pleasure, and presumably the same holds for pain as well. There are conditions in which a perceiver experiences neither pleasure nor pain: according to the Annicerians, the states of the absence of pleasure (*aēdonia*) and of the absence of pain (*aponia*) are not *pathē*, for they are not related to motions but are comparable to the state of someone asleep. On the other hand, Aristippus the Younger introduced a third category of *pathē*, which are associated with motions and defined as intermediate between pleasure and pain. They represent the manner in which one is affected by one's perception of the empirical qualities of an object and are reported by verbal or adverbial neologisms, such as 'being whitened' or 'being affected whitely', and 'being sweetened' or 'being disposed sweetly'. Pleasure and pain, as well as the intermediates, are sensed by 'inmost' or 'internal' touch. Although the sources say nothing further about this, it seems likely that the function of internal touch is to register the qualitative alterations of the flesh, and thus to provide a necessary and sufficient condition for one's awareness of one's own *pathē*.

These characteristics of the *pathē* constitute the basis for the Cyrenaic theory of knowledge, whose full formulation probably coincided with the introduction of the intermediates into the doctrine. It is centred on the subjectivist claim that the only knowledge accessible to one is the awareness of one's own *pathē*, and on the dogmatic assertion that the nature of external objects cannot be known.

Although the Cyrenaics draw no sharp distinctions between the physical and mental aspects of the *pathē*, in the context of their epistemology they clearly concentrate on the latter: one's apprehension of the *pathē* concerns primarily the subjective experience that one has. For, first, the Cyrenaics distinguish between the smooth movement associated with pleasure and the sensation or consciousness (*aisthēsis*) of it. Second, their use of adverbial expressions, such as 'being disposed whitely', is applicable to perceivers only, and inapplicable to inanimate objects undergoing purely physical change. And third, the *pathē* are, according to the Cyrenaics, self-presenting states which cannot yield false information about themselves to the person who experiences them. This only makes sense if the *pathē* are taken to be the immediate contents of experience rather than objective physical states.

This is confirmed by their insistence on the privacy of experience: each *pathos* is private (*idion*) to the individual perceiver experiencing it. Autobiographical reports regarding one's own *pathē* are infallible (*adiapseusta*), incorrigible (*anexelenkta*), always firm (*bebaia*) and true (*alēthē*). Thus, one has privileged access to one's own *pathē*, which may be related to the way in which they are detected by 'internal touch'. By contrast, the Cyrenaics stressed the structural and definitional inaccessibility of other people's *pathē* by arguing that people cannot have a common, that is, publicly shared, *pathos* which might be used as a common criterion of truth. The private character of the *pathē* is also said to make them incommunicable to others.

The incorrigible awareness of the *pathē* is contrasted with one's ignorance of external objects. The scope of Cyrenaic scepticism is limited in that it does not question in a radical and systematic way the actual existence of objects and, consequently, does not raise the so-called problem of the external world. It concerns mainly our knowledge of empirical properties or qualities and is supported by the argument that our beliefs about them are in reality beliefs about the content of our *pathē*, and that no one can transcend the barrier between objective properties and our perception of them. Since the causal history of a *pathos* cannot be traced back to the objective property which it is ordinarily supposed to represent, our beliefs regarding the properties of objects can be neither confirmed nor refuted by further evidence of any kind.

The Cyrenaics appear to have focused on specific problems, mostly related to the perception of secondary qualities, and did not try to provide a general epistemological theory. Important criticisms that can be brought

against their position include the following: (1) the claim that only the *pathē* are apprehensible is too narrow to justify our ordinary assumptions about what we know; (2) the theory is either inexpressible or self-refuting; (3) it entails a radical subjectivization of all experience, which raises doubts, unwelcome to the Cyrenaics themselves, about the very existence of the external world and of other people.

3 Early Cyrenaic ethics

It has been commonly held that the Cyrenaics were the only philosophers in antiquity to abandon the traditional position of Greek ethics; that the moral end is happiness (see Eudaimonia), and to identify it instead with individual bodily pleasure experienced in the present moment. However, this position overlooks the value that the Cyrenaics attributed to happiness as well as to individual pleasures, and fails to account for the significant differences between the ethical doctrines of various members of the school.

There are two groups of testimony about the ethical views of Aristippus the Elder. According to one, he was a hedonist (see Hedonism) who defined the moral end as a smooth movement resulting in a short-lived pleasure of the body. He distinguished it from happiness, the particular collection of individual pleasures experienced over a lifetime. But according to the second body of testimony, Aristippus did not define pleasure as the only thing which is intrinsically good, nor did he maintain that it is only pleasure and not happiness that is pursued for its own sake. Although he appears to have conceived of happiness in compositional terms and to have considered pleasure its major constituent, he believed that the enjoyment of pleasure should be conditional upon the capacity to master oneself. This is acquired by means of philosophical education, which also secures spiritual freedom, self-awareness and, more generally, the well-being of one's soul. Provided that these goods are obtained, one may indulge in many and intense pleasures without the fear of enslaving one's soul.

Regardless of the origins of the two traditions, there are good reasons for accepting the latter. First, there is evidence that Aristippus never lectured in defence of a particular moral end and that he detested adopting a didactic tone. This tells against his having argued on behalf of a well-defined hedonistic position and, perhaps, in favour of his having shared the common assumption that the supreme good is happiness, however defined. Second, one source, the Peripatetic Aristocles, flatly denies that Aristippus was a hedonist, and adds that he was only thought to be so on the grounds of his intemperate habits. Third, the importance which he attributed to long-term activities such as philosophy, and his statements that he aims at the easiest and pleasantest life and that he wishes to live a life of freedom leading to happiness (Xenophon, *Memoirs of Socrates* II 1) indicate that he perceived his life as a continuous whole, not as a series of disconnected pleasurable episodes. On that account, too, he seems closer to eudaimonism than to hedonism. Finally, the relationship of Aristippus to Socrates and the claim of his school to a Socratic pedigree make it difficult to believe that he would have held a view which entailed that any means to the pleasure of the present is acceptable and which, therefore, would have verged on the immoral.

The value that Aristippus attributed to a life of freedom and happiness recurs in his political stance. According to Xenophon, he declined to assume the civic roles of ruling or of being ruled which were necessarily attached to the political condition of citizenship and, instead, chose for himself the apolitical stance of 'living as a foreigner in every land'. Thus, he defined freedom not only as internal freedom, but also as freedom from duties and obligations with regard to any particular city. His motivation for assuming this position was eudaimonistic: only the life of a foreigner affords the type of freedom that leads to happiness.

Aristippus the Younger was probably the first Cyrenaic philosopher to single out momentary bodily pleasure as the moral end. He appears to have conceived of the moral end as complete and perfect, for he maintained that bodily pleasures do not differ in intensity or degree: none is more pleasurable than another, presumably because no particular instance of the moral end can be more perfect, and therefore more worthy of choice, than another. He distinguished between happiness and the moral end by defining happiness as a particular collection of individual pleasures experienced over one's lifetime, and by claiming that happiness is not desirable for its own sake but for the sake of the pleasures composing it. However, his hedonistic ethics is not altogether devoid of eudaimonistic considerations. For he believed that the overall moral goal is to live a pleasurable life, thus emphasizing the importance of accumulating the particular pleasures which jointly compose a happy life. Such considerations arguably precluded him from holding the view that all means are permissible for obtaining pleasures. For some means to a particular pleasure may obstruct the achievement of subsequent pleasures or may have painful after-effects, and thus may affect one's enjoyment of a happy life.

4 Later Cyrenaic ethics

The ethics of Anniceris and of Hegesias also rest on a delicate balance between hedonism and eudaimonistic elements. Anniceris and his disciples further clarified the notion of the moral end by emphasizing that pleasure and pain are necessarily related to motion and by arguing that, since the condition of mere absence of pain is not characterized by motion, it cannot be part of the moral end. Their thesis should be contrasted with the position of the rival school of Epicurus, according to which the supreme moral good is static pleasure, characterized by the complete absence of pain (see Epicureanism §10). They broadened Aristippus the Younger's definition of the moral good by attributing intrinsic moral value to mental as well as bodily pleasure, and allowed for circumstances in which sages will act altruistically and deprive themselves of pleasures, but will none the less be happy. Anniceris' concern with the repetition of individual pleasures and the enjoyment of new pleasures over an extended period of time is also indicated by his recommendation of forming good habits in order to correct a naturally bad disposition, and friendships on account of the repeated good feelings that they cause us.

Hegesias was a hedonist in principle, since he too defined the momentary present pleasure as the moral end. However, he added that one scarcely enjoys such pleasures, whereas one suffers numerous pains during one's lifetime. On these grounds, he concluded that the collection of individual pleasures constituting happiness cannot be achieved. The unattainability of happiness led him to maintain that life is a matter of indifference, and that life and death are each desirable in turn. His pessimism is moderated by the intellectualist position that no one errs voluntarily, and by the philanthropic concern that one should not despise other people but teach them.

Theodorus radically modified the hedonism of the other Cyrenaic sects. He introduced the concepts of joy and grief, defining joy as the supreme moral good, and claimed that pleasure and pain are intermediate between good and evil. Although he did not speak openly of happiness, his position has clearly eudaimonistic dimensions. For he viewed particular pleasures instrumentally, as means to joy, and he believed that joy results from the exercise of wisdom and justice. For these reasons, it seems likely that he conceived of joy as a long-term pleasurable state akin to happiness. His doctrine is comparable to the views of Aristippus the Elder: he considered wisdom and justice to be goods and emphasized their importance for achieving the moral end; he maintained that the sage need not be temperate in the enjoyment of pleasure, but unlike Aristippus the Elder did not make it conditional upon the capacity for mastering oneself; and he stressed the self-sufficiency of the sage, advocated cosmopolitanism and considered irrational the performance of military duty for any particular city. On the other hand, Theodorus introduced into Cyrenaic ethics an atheism and an amoralism which are not encountered in the doctrines of his predecessors.

The ethical doctrines of the Cyrenaics, regarding both the central role of pleasure and the importance of happiness, appear to find support in their epistemology. Since any momentary pleasure is a *pathos*, the manner in which it affects one is infallibly apprehended. It could be argued that we infallibly apprehend not only the pleasurable nature of the experience but also its self-evident goodness. Regarding the desirability of a happy life, the Cyrenaics could maintain either that it is itself a *pathos*, so that its content is indubitable, or that the goodness of a pleasurable life is self-evident in so far as the goodness of its component pleasures is self-evident. This connection between their ethics and their epistemology is admittedly no more than a conjecture and invites the objection that moral properties such as goodness are not part of the self-evident content of a *pathos*. Further, even if the goodness of every individual *pathos* of pleasure were self-evident, it would not straightforwardly follow that a collection of such *pathē* must preserve the goodness of the pleasures constituting it.

5 The Socratic connection

Historically, the Cyrenaic school owes its Socratic pedigree to the inclusion of its founder, Aristippus of Cyrene, in Socrates' entourage (see Socratic Schools). Philosophically, there are several elements both in its epistemology and in its ethical doctrines that would justify the claim that it remained close to the spirit of Socrates' teachings.

Concerning their theory of knowledge, the Cyrenaics might argue that, like Socrates, they obeyed the Delphic maxim 'Know yourself!' and came to the conclusion that one cannot know anything but oneself. Thus, both the description of the physical states related to the *pathē* and the epistemic analysis of our apprehension of them could be inscribed in the context of an effort to pursue further the Socratic example of self-knowledge. Further, they could appeal to Socrates' disavowal of knowledge in support of their position that one is necessarily bound to be

ignorant about external objects, thus endorsing, or perhaps even inaugurating, the tradition which made Socrates the founder of scepticism.

As regards their ethical doctrines, the Cyrenaics could make the general remark that even the Platonic Socrates, whether or not a hedonist, considered the enjoyment of mental pleasures compatible with the existence of a virtuous state of soul, and did not disallow all bodily pleasures, but only those which endanger self-mastery and moderation. Finally, most Cyrenaic philosophers appear to have endorsed some version of Socratic intellectualism - perhaps the most important motivation for calling them a Socratic school.

VOULA TSOUNA

References and further reading

Doering, K. (1988) *Der Sokratesschüler Aristipp und die Kyrenaiker (Aristippus the Pupil of Socrates and the Cyrenaics)*, Wiesbaden and Stuttgart.(The most recent book-length study of Cyrenaic philosophy.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae (The Fragments of Socrates and the Socratics)*, Naples: Bibliopolis. (Volume 2 contains the Greek and Latin testimonies for the Cyrenaics.)

Irwin, T. (1991) 'Aristippus Against Happiness', *The Monist* 74: 55-82.(Explores the nature of Cyrenaic hedonism and its epistemological support.)

Tsouna, V. (1998) *The Epistemology of the Cyrenaic School*, Cambridge: Cambridge University Press.(The only full-scale study of this topic, with translations of the relevant texts.)

Xenophon (*c*.360s BC) *Memoirs of Socrates*, trans. H. Tredennick, revised R. Waterfield, Harmondsworth: Penguin, 1990.(Aristippus appears as one of Socrates' interlocutors in II 1 and III 8.)

Czech Republic, philosophy in

The foundation of the University of Prague in 1348 contributed significantly to establishing Bohemia as a centre of philosophical thought. The main philosophers and theologians from the University favoured the Platonic tradition, and from this position they criticised corruption in the Church. The most important representative of this trend was Jan Hus who followed the teaching of John Wyclif in the spirit of rationalism and humanism. His ideas became an ideological base for the anti-feudal Hussite Revolution in the fifteenth century and the later Czech Reformation. Theoreticians on the extreme wing of the Revolution held a natural worldview, opposing the notion of transcendence. Social thinking in this era found expression in Petr Chelčický, who preached a strict pacifism and a classless society. The Revolution broke the power of the Church's ideological monopoly, and had a positive impact on the development of Czech society for the following two centuries. The atmosphere of relative tolerance allowed Renaissance thinking and Czech Reformation rationalism and humanism to enrich each other. This tradition culminated in the work of Jan Amos Comenius, who aimed to improve social relations through rational enlightenment and education, promoting harmony and justice for the development of all humankind.

After 1620, Czech spiritual life was paralysed for many centuries by a forced anti-Reformation and the emigration of many of the country's leading intellectuals. A revival started only at the end of the eighteenth century. František Palacký, inspired by the neohumanism of his era and by the Czech Reformation, formed a new philosophy of Czech history. B. Bolzano achieved impressive results in philosophy of science and logic, while A. Smetana created an independent variant of the philosophy of identity. In the second half of the nineteenth century Herbartism became very influential, contributing to social psychology and aesthetics.

The most important representative of modern Czech thinking is T.G. Masaryk, creator of a philosophical concept of democracy understood in the context of a humanistic world view. Masaryk's philosophy has been followed by many philosophers and theologians in the twentieth century. In the period between the wars, important concepts based on structuralism were created by J.L. Fischer (philosophy, sociology),J. Mukařovský (aesthetics), V. Příhoda (psychology, pedagogy), and many linguists. J. Patočka contributed to the development of phenomenology with his concept of the natural world. In opposition to 'school' philosophy L.Klíma preached extreme subjectivism and individualism. Non-dogmatic Marxists wrote internationally regarded works. Well known in analytical philosophy and modern logic are the achievements of L. Tondl, O. Weinberger, K. Teige, R.Kalivoda and K. Kosík and P. Tichý.

1 Early period

Philosophy in the Czech Republic is connected to national history in a manner unlike any other European 'philosophical' country. It has sometimes had to disregard academic problems and interests to join the social battles of the era, and so it has played a major ideological role. Often it has had to reach beyond itself to develop ideas in other spheres of spiritual life, from the literary and historical to the natural sciences. The beginnings of philosophy in Bohemia go back as far as the thirteenth century to the Prague Cathedral School, which had contact with other European centres of philosophy. The founding of Prague University in 1348 was a positive stage in the development of philosophical scholarship, and important philosophical trends were evident at the University from the beginning. With time, however, realism in opposition to nominalism prevailed, based on the Platonic traditions of European thinking (see Nominalism). The chief representative was Vojtěch Raňkuv of Ježov (Adalbert Rankonis) (d. 1388), a Rector of the Sorbonne who introduced the writings of John Wyclif to Prague University. Others include Jan Milíč of Kroměříž (d. 1374), a severe critic of both secular and Church power who influenced later Hussite thinkers with his chiliastic image of the coming doomsday and the 'era of justice'; Matěj (Mathias) of Janova (d. *circa* 1394), who denounced the institution of the church and demanded the creation of a *communio sanctorum* (community of saints) modelled on the first Christians; and Tomáš of Štítný (d. *circa* 1405), the first scholar to write his critical essays in Czech and the founder of Czech philosophical terminology.

This movement culminated in the early fifteenth century with Jan Hus, whose philosophy built on the achievements of Wyclif to reach humanistic and rationally critical conclusions. Hus believed that every Christian has the right to use their reason to judge the laws of the secular world as well as Church powers, as well as the right to refuse to obey if they find any discrepancies between these and God's law (understood as a Platonic Idea).

The truth wins only when a man is willing to sacrifice his own life for it. Hus attempted to defend his ideas before the Church Council in Konstanz, but failed and was burnt at the stake. After his death, his ideas became an ideological starting point for the Hussite Revolution (1419-34) and the Reformation that followed. His influence was widespread: Jakoubek of Stříbro' (Jacobellus of Mies) (d. 1429), an ideologist of the mild wing of the Hussite Revolution, introduced Holy Communion from the chalice as a symbol of Christian equality for all believers; Mikuláš Biskupec (d. *circa* 1460) represented the radical wing and preached the 'liberal law of the Lord' which was based on the spiritual and social emancipation of human beings; Martin Húska (d. 1421), an extreme left-wing Hussite ideologist, refused to accept any notion of transcendence and denied theist dogmas - for him, God lives in human beings and the Eucharist is a real material feast, all people are equal and sexual life should be free; Petr Chelčický (d. 1460), a social reformer, believed the feudal division of society to be unjust and claimed equal rights for everyone, calling for universal love and condemning all violence (his ideas later influenced L.N. Tolstoi).

The Hussite Revolution inspired the Czech Reformation, the first victorious reformation in Europe, and in 1457 a new Church was founded in Bohemia. Known as *Jednota bratrská* (Unity of Brethren), this new Church, was based on the teachings of Chelčický. Later it became an important centre of spiritual and cultural life in Czech society. (Many of its members emigrated in the seventeenth century before the violent anti-reformation broke out, and one of its branches still exists in the USA as the Moravian Brethren.) The 'Unity' proclaimed the idea of religious tolerance and, combined with other Reformation ideas, contributed to the forming of the 'Czech confession' (1575) which demanded freedom of belief for all - including serfs, thereby going against the feudal principle that they follow the beliefs of their ruler ('*cuius regio, eius religio*'). Among the representatives of the Unity was Jan Blahoslav (d. 1571), an outstanding bishop who wrote many theological and pedagogical essays, as well as the first Czech book on aesthetics and musical theory.

In Bohemia as elsewhere, the Reformation was followed by a surge in Renaissance humanism, both existing simultaneously and having mutual influence. In fact, representatives of both sides aimed to bring the two movements together. Viktorin Kornel of Všehrdy (d. *circa* 1520), a founder of the first Czech law school, worked on the philosophy of natural law, which reflected the self-awareness of the new class of citizens. He was also the creator of a philosophy of Czech history that overemphasized the importance of the Hussite period. Řehoř Hrubý of Jelení (Gelenius) (d.1514), a translator of Cicero, Petrarch and Erasmus of Rotterdam, was influenced by Stoicism and united his patriotism with the idea of Universal Goodness into a universally understood world view. Jan Jessenius (d.1621), a rector of Prague University who was executed for his part in the anti-Habsburg uprising, was a physician and a follower of Renaissance natural philosophy (inspired by Franciscus Patritius). He also supported Heliocentrists, as did the outstanding mathematician and astronomer Tadeáš Hájek (d.1600), whose work was used by Tycho de Brahe, Kepler and Galileo.

Prague at this time was also an important centre of Jewish culture. Its most important representative was Jehuda ben Becalel Liva (d.1609), the legendary Rabbi Löw who, in his chief philosophical and theological document *Be'er ha-gola* (1600) expresses his understanding of a free man opening himself to the outside world through his activity. Rabbi Löw also wanted the Jewish community to engage other nations in dialogue on equal terms.

Czech Reformation thinking, which was enriched by many elements from the Renaissance, culminates in the synthetic work of Jan Amos Comenius, the last bishop of the Unity. In his major work, *De rerum humanarum emendatione consultatio catholica*, he created a philosophical system dominated by the idea of amelioration and the harmonious arrangement of human relations, based on rational enlightenment and general education. He believed in human nature as an active force capable of constant improvement. Didacticism and pedagogy were to serve as an instrument to reach this aim. At he deepest layer of this conception there is an ontology based on a modified Neoplatonist scheme, whose highest degree is the harmonious development of all beings into a new human reality.

The movement in Czech philosophical thinking which culminated in the work of Comenius was interrupted after the Battle of White Mountain in 1620. The Czech State ceased to exist and a forced re-Catholicization of the country followed. The elite of Protestant intellectuals, including Comenius, had to leave the country and their place was taken by Catholic priests, mostly of foreign origin, who brought with them other intellectual traditions. Some of them contributed to the history of seventeenth-century philosophical thinking, including Roderigo Arriaga, a follower of the neoscholastics, who tried to find a compromise between the teaching of Thomas Aquinas

and William of Ockham in a spirit of mild realism; Jan Caramuel of Lobkowicz, considered a father of modern logic of relations; and Valerianus Magni, a Franciscan who developed the Scotistic tradition, opposed neoscholasticism and was close to Descartes' rationalism (see Descartes, R.; Duns Scotus, J.; Rationalism). Jan Marcus Marci of Kronland was a different type of thinker, a natural scientist rooted in Platonism and Neoplatonism who preached the analogical construction of microcosm and macrocosm - through studying the activity of nerves he arrived at ideas anticipating later associative psychology (see Neoplatonism; Platonism, Renaissance). Hieronymus Hirnheim belonged to the anti-Jesuit opposition. He stressed the ethical ideas of original Christianity, the primacy of life-praxis over speculative scholasticism and the importance of subjective belief oriented to a philosophy close to Jansenism.

2 From the National Revival to 1918

After the long period of oppression, Czech intellectual life was slowly revived in the last third of the eighteenth century. It was inspired first by the Enlightenment, later by neohumanism and romanticism, and at the same time by the older national tradition of the Reformation. It paid great attention to the development of the most important branches of scientific research. Consequently Slavic philology, history and natural sciences were studied, and these soon reached European academic levels.

The founder of Czech and Slavic philology, Josef Dobrovský (d.1829), was also a founder of modern critical rationalistic methods of research rooted in the Enlightenment. Bernard Bolzano, the greatest philosopher of this era, developed on the traditions of the Enlightenment and criticized Kantian and post-Kantian philosophy. An inventive mathematician and one of the founders of the philosophy of science and modern logic, he investigated the semantic basis of logical systems. His pursuit of the relationship between idea (*Vorstellung*) and notion, of the notion of consciousness, and of the function of *Vorstellungen* in a sentence, was later expanded by Husserl (see Husserl, E. §6). Jiří Procházka (d.1820), a physiologist, initiated the reflexive theory of spiritual activity. (His 1784 essay *De functionibus systematis nervosi commentatio* was translated into English in 1851, and influenced the theory of Marshall Hall.) Later on, in the spirit of German *Naturphilosophie*, he created a dynamic picture of nature according to the law of polarity derived from the science of electricity. Jan Evangelista Purkyně (d.1869) was also influenced by the German tradition. A founder of the theory of cellular construction of living organisms, he added a historical dimension to his dynamic concept of nature, and held a pantheist ontology (see Pantheism). František Palacký (d. 1876), author of the monumental 'History of the Czech nation', had pursued philosophy in his youth. He was influenced by neo-humanism, and especially by Schiller on whose ideas he formed a model of perfect humanity and thus his humanistic philosophy of Czech history, which for a long time was a basis for Czech policy (see Schiller, F.C.S.).

The revolution of 1848 was philosophically reflected in the work of Augustin Smetana (d. 1851), whose roots were in left-wing Hegelianism. In his philosophy of human spirituality he created an independent variant on the philosophy of identity, based on the relative identity of knowledge (*Wissen*) and existence (*Sein*). He interpreted the Hegelian historical scheme of the development of *Weltgeist* so that it envisaged that the era to come would bring a dominant role and social justice for the Slavs. However, the impact of Hegel's philosophy was not very strong in Bohemia.

The dominant trend in philosophical thinking in the second half of the nineteenth century in Bohemia was Herbartism, introduced to Bohemia by Franz Exner (d. 1853) who opposed Hegel's idealism and criticized his psychology (see Herbart, J.F.). Robert Zimmermann (d. 1898), a student of Bolzano, managed to influence this trend through his historic and synthetic works in aesthetics while working at Prague University. Some followers of Herbartism developed single philosophical disciplines, others originated completely new ideas in the philosophy of religion, the history of philosophy, sociology, social psychology, and logic. Otakar Hostinský (d. 1910) created an aesthetic concept of 'concrete formalism' and the philosophy of art, and some of his ideas were employed by the Prague structuralist school (see Structuralism in linguistics). Herbartism was very influential throughout central Europe, and to some extent it fulfilled the role of positivism in Czech intellectual life (that is why classical positivism appeared relatively late; but it lasted well into the twentieth century). Herbartism tried to react positively to impulses from contemporary sciences (empirical psychology and Darwinism for example) and provided the first steps to a variety of new concepts and inventions such as *Gestaltqualitäten*, the neopositivism of the Vienna Circle, and the relationship of consciousness and subconsciousness in Freud (see Vienna Circle).

Miroslav Tyrš (d. 1884) stood outside this trend. He transformed Schopenhauer's pessimistic concept of the will into a philosophy of human activity of a subject who is both a counterplayer and a co-player in the struggle of life (Darwin's influence) and formulated a new plan for the development of a small nation, working it out in practical terms by founding a mass physical training organisation 'Sokol' which is still in existence (see Schopenhauer, A.). Tomáš Garrigue Masaryk began his philosophical career by opposing Herbartism and German post-Kantian philosophy, and he gradually succeeded in weakening their positions. His philosophical roots were in French and English positivism, but he used these solely as a research methodology to investigate the general crisis of values. He sought to resolve this crisis by using the concept of a human being as a harmonious unity of reason, will and emotion. His thought was based on religious values of the Czech Reformation and on the idea of democracy, understood as a humanistic world outlook. After 1918, this concept became the central idea of the newly formed Czechoslovak state, of which Masaryk was the first president.

3 From the foundation of an independent state

The culture and political climate of the new democratic state was conducive to philosophical investigation, and philosophy was an integral part of the curriculum in the newly formed universities and colleges. Among the many trends, positivism was still in fashion and even became a sort of national philosophy. It was mostly true to J.S. Mill's principles of positive methodology and to Comte's schemes of development, although attempts were made to modernize these (see Comte, A.; Mill, J.S.). Josef Tvrdý (d. 1942) introduced the emergentism of Samuel Alexander to the qualitative understanding of evolutionary processes. However, Josef Král (d.1978) worked out his concept of the history of Czech thought in a traditional spirit. Emanuel Chalupný (d.1958) also pursued the philosophy of Czech history and created his own sociological system. Spiritual fathers of neo-positivism like Ernst Mach (a long-serving professor at Prague University), Rudolf Carnap and Philip Frank, who both worked in Prague for some time, had less direct impact on Czech philosophy.

Masaryk's humanistic philosophy based on religion was adopted mainly by Protestant thinkers, including Emanuel Rádl (a proponent of 'intuitive realism' and critic of positivist scientism), Josef L. Hromádka (an independent follower of so called dialectic theology), and Jan B. Kozák (who was influenced by phenomenology and attempted to solve the problems inherent in the relationship between science and belief, and the question of objective ethical values). Masaryk's influence extended into many different fields and he found followers in a wide range of disciplines.

In the period between the First and Second World Wars, the three basic streams of structuralism, phenomenology and non-orthodox Marxism constituted mainstream philosophy, and their impact was to prove of major importance. About 1930, structuralism became a methodological basis for some of the humanities and social sciences and took a specific dynamic form with an expressive historical aspect. Its noetical, ontological and sociological principles were worked out by Josef Ludvík Fischer who also created a model of structural democracy, freeing all social forces for the benefit of the individual and society as a whole. Theoretical bases for sociology were formed in an analogical way by In. Arnošt Bláha. The *Cercle linguistique de Prague* became an internationally acclaimed propagator of the structural and functional understanding of language, literature and the arts, enriching the philosophy of all three in the works of Jan Mukařovský and Roman Jakobson. Additionally, structural philosophy of education and structuralist methods in psychology were advocated by Václav Přihoda (see Structuralism).

The Czech-German *Cercle philosophique de Prague*, founded on the initiative of the aesthetician Emil Utitz, with J.B. Kozák, became an ideological centre of the phenomenological movement. In 1935 Husserl was invited to Prague and the lectures he delivered were the basis for his final book *The Crisis of the European Sciences*. Jan Patočka, Husserl's pupil, became the most important representative of phenomenology in Bohemia. In the 1930s he worked on the problem of the original, predetermined natural world (*Lebenswelt*), its structure, and the activity of human beings in it. This was a very progressive project and it was some time before it gained a following in the phenomenological literature. Also some structuralists of the era had an interest in phenomenology and duly modified their concept of the function and tasks of art in the context of human existence (see Phenomenological movement).

Marxist philosophy in the first half of the twentieth century had a social basis in a strong left-wing political movement. Apart from the ideologists who took their starting point from the Communist Party, there were those

who took Marxism as a starting point for independent thinking. One such thinker was Karel Teige, an important theoretician of avant-garde art. He regarded artistic activities a means to realize an ideal of a free human being and to apply his creative force in the contemporary social conditions. Jaroslav Kabeš, who was influenced by the young Marx and later on by great representatives of Czech culture, persisted in questioning the purpose of human existence. The problems of a materialistic concept of history, and particularly its relation to Czech history and the history of European revolutions, attracted the attention of Kurt Konrad and Záviš Kalandra. The latter was also one of the first Czech Marxists to consider the importance of Freud's work in understanding the role of the psychological factor in history and the structure of personality.

Other philosophers of the first half of the twentieth century also had significant influence and impact. They include Ladislav Klíma, a radical subjectivist and philosopher of free existence in an absurd world, anticipating later existentialism (see Existentialism); natural scientists František Mareš and Karel Vorovka who argued against positivistic scientism from an idealistic starting point; Vladimír Hoppe who emphasized intuitivism and mysticism; and Karel Engliš who created a philosophical system based on the teleological understanding of social relations (see Functionalism in social science). Among Catholic researchers, neo-scholastic Josef Kratochvil attracted most attention. He called his standpoint 'neo-idealism' and aimed to overcome the positivist view of unknowable transcendence. The group of philosophers and theologians gathered around the journal *Filosofická Revue* was also neo-thomistically oriented, while those German philosophers in Prague who gathered round the Brentano-Gesellschaft, formed from the second generation of Brentano's pupils, were also very active.

During the Nazi occupation and the following six years of the Second World War, the continuity of philosophical investigation in Bohemia was interrupted. Universities and colleges were closed down and severe censorship was introduced. Some important philosophical figures were executed, some imprisoned in concentration camps and others forced into exile. The end of the war did not mean a return to pre-war conditions, the shift towards socialism creating a new situation. In the relatively free political situation of the years 1945-8 the ideological influence of Marxism was increasing. This process was supported by Arnošt Kolman who came back from the Soviet Union, as well as by many philosophers of socialist orientation who accepted Marxism and those who had previously been closer to it. Some non-Marxists continued to hold on to their former views.

This situation changed in February 1948 when the Communist Party came to power. Stalinism became the official ideology and all other philosophical and ideological schools of thought were prohibited. Many philosophers, especially Catholics, were imprisoned, and some of the younger ones left the country. Marxists were also persecuted. It took almost ten years for the ideological opposition to get more space for freer philosophical investigations. Meanwhile, Masaryk continued to have followers at home and abroad, Fischer continued his ontological work based on structuralism, and Patočka proceeded in his phenomenological investigations, trying to synthesize his ideas with those of Heidegger, while also maintaining an interest in the history of philosophy.

In the 1960s, young non-Marxist intellectuals gathered round the cultural journal *Tvář (Face)* - its spiritual agent was a playwright, Václav Havel, and philosophical essayists Ladislav Hejdánek and Jiří Němec were among the contributors. The freer atmosphere was also favourable to the development of non-orthodox Marxism, and the tendency to integrate elements of contemporary European thinking with Marxism became stronger, manifesting itself in many ways. An important experiment aimed at solving phenomenological-existential problems on the basis of Marxism was carried out by Karel Kosík, and similar efforts were made by other philosophers of the same generation. Questions of philosophical anthropology and ethical study also attracted attention. The problems of structuralism and the philosophical message of the artistic avant-garde were considered using Marxist categories by Robert Kalivoda (d. 1989) and simultaneously by other theoreticians in different variants. Other philosophers researched analytical philosophy, cybernetics, semantics and modern logic. In this connection are works written by Ladislav Tondl, Otto Weinberger and Pavel Tichý. At this time, much light was shed on the many periods of the history of philosophy, including that of Bohemia, while at the same time systematic questions raised by the material were widely discussed.

The third catastrophe of Czech philosophy in this century was the defeat of the 'Prague Spring', and the subsequent inauguration of another totalitarian regime. The majority of philosophers in the movements discussed above were dismissed from their posts at universities and research institutions, were prohibited from publishing and their books were removed from public libraries. Those who did not leave the country were compelled to work

as unskilled manual workers. This gave rise to vigorous unofficial intellectual activity. In spite of police repression, lectures and seminars were organized in private houses, sometimes with the participation of philosophers from France, England, Germany and The Netherlands. Dozens of books and some magazines were issued independently (*samizdat*), and many works were published abroad. This situation was not without precedent - the philosopher and poet known as Egon Bondy (real name Z. Fišer), a legendary personality of the Czech underground, had been involved in such activity since 1948. Students of Patočka took part in the preparation of a twenty-seven volume edition of Patočka's Collected Works, and the youngest generation of his pupils continued in his tradition after their teacher's death. This time also saw the formation of an active group of Catholic-orientated philosophers who concentrated their efforts on publishing the illegal magazine *Paraf* (whose name derived from *Para*lelní *f*ilosofie, parallel philosophy, in opposition to the official philosophy). In the 1980s, as the pressure of the regime gradually decreased, many (mostly younger) philosophers were given the opportunity to work in the official institutions.

The fall of the totalitarian regime at the end of 1989 allowed the necessary conditions for the resumption of Czech philosophy to be re-established. Such freedom had existed in modern times only in the period between the two world wars. Many philosophers of the older generation returned to their original profession, as did some younger philosophers who had been forbidden to work in their field of study after graduating. E. Kohák, I. Sviták, J.M. Lochman and K. Chvatík were some of the philosophers living in exile who now renewed contacts with home. Besides these there was the Catholic thinker M. Loblowicz, analytic philosophers and logicians O. Weinberger, P. Tichý and J. Šebestík, phenomenologist I. Šrubař, postmodernist V. Bélohradský and others. Active philosophical life could now resume.

See also: Marxist philosophy, Russian and Soviet; Slovakia, philosophy in

Translated by G.R.F. Burša
JOSEF ZUMR

References and further reading

Comenius, J.A. (*c.*1660s) *De rerum humanarum emendatione consultatio catholica (General consultation on the reform of human affairs)*, Prague: Academia, 1966.(The synthesis of Comenius's life work not published in its entirety until 1966. Mentioned in §1 above.)

Gabriel, Jiří-Nový, Lubomír (eds) (1993) 'Czech Philosophy in the 20th Century', *Czech Philosophical Studies II*, Cultural Heritage and Contemporary Change IV, Washington, DC: The Council for Research in Values and Philosophy.(Summary of separate trends of Czech thinking written by a team of authors.)

Jakovenko, B. (1935) 'La philosophie tchécoslovaque contemporaine' (Contemporary Czech Philosophy), in *Der russische Gedanke* IVe Prague: Supplément de la revue.(Chapter about leading thinkers of the twenties and thirties with bibliographies: texts are in French, German, English and Italian.)

Jehuda ben Becalel Liva (1600) *Be'er ha-gola*, Prague.(Rabbi Löw's chief philosophical and theological document, mentioned in §1 above.)

Král, J. (1934) *La philosophie en Tchécoslovaquie (Philosophy in Czecheslovakia)*, Prague: Bibliothèque des problèmes sociaux.(Digest of history of philosophy in Bohemia and Slovakia with a separate chapter about German philosophy in Bohemia.)

Lobkowicz, Nikolaus (1961) *Marxismus-Leninismus in der ESR, Die tschechoslowakische Philosophie seit 1945*, Dortrecht: Reidel(Detailed analysis of Czech and Slovak Marxism with consideration of other trends.)

Mácha, Karel (1985-9) *Glaube und Vernunft: Die Böhmische Philosophie in geschichtlicher Übersicht*, I-III. Munich, New York, London, Paris: KG Saur.(Broad history. The author's judgment is at times somewhat subjective.)

Procházka, Jiří (1784) *De functionibus systematis nervosi commentatio*, Prague: Academia, 1954. (Treatise on the functions of the nervous system, mentioned in §2 above.)

Zumr, Josef (1971) 'Philosophie der Gegenwart in der Tschechoslowakei', in R. Klibansky (ed.) *Contemporary Philosophy*, Florence: La nuova italia editrice, 455-73.(Digest of Czech and Slovak thinking between 1930 and 1969.)

D’Alembert, Jean Le Rond (1717-83)

Mathematician, scientist and man of letters, Jean D’Alembert is a central figure of the French Enlightenment. As a young man he made significant contributions to the refinement of mathematical techniques, and later was actively engaged in the theoretical controversies which surrounded the gradual assimilation of Newtonian mechanics into the mainstream of European science. For twelve years (1746-58) he was co-editor, with Denis Diderot, of the Encyclopedia, the serial publication of which was one of the defining events of the Enlightenment period as a whole. D’Alembert frequented the various Paris salons where much of the intellectual fervour and high-spiritedness of the age was cultivated and given shape. As Secretary of the French Academy he worked assiduously to advance the cause of human knowledge.

D’Alembert’s philosophy is characterized by an abiding commitment to the clarity and precision which attends mathematical abstraction. He believed that in its essence the natural order is internally structured by laws whose operation can be articulated under the principles of geometry. All natural phenomena are to be explained under the terms of those basic mathematical principles that govern the scientific domain in which they are located (chemistry or astronomy for example), and all scientific domains could be brought ultimately to perfect consistency and systematic order within a comprehensive theory. The events and processes which constitute the natural order reflect the reality of the mathematical structure which underlies them. As he says in the Preliminary Discourse (1751) to the Encyclopedia (1751-65), ‘The universe would only be one fact and one great truth for whoever knew how to embrace it from a single point of view’.

1 Life and works

D’Alembert was the illegitimate son of Mme de Tencin and the chevalier Louis-Camus Destouches. Shortly after his birth he was left by his mother at the steps of the church St Jean Le Rond in Paris, a circumstance which gave the boy his name. While never formally acknowledging paternity, the father took a devoted interest in his son and provided for his education. Jean was enrolled in the Jansenist Collège des Quatre Nations under the name of Dalemberg, which was later altered to the form by which he is known to history. The boy’s intellectual gifts were apparent from the start; he was encouraged to pursue a career in theology, though he never expressed any great interest in it, studied law and medicine for three years, but was increasingly drawn to mathematics. Though his formal instruction in mathematics was limited, by 1739 D’Alembert had attained a sufficient mastery of the subject to submit a paper to the Academy of Sciences in which he pointed out some errors in a popular textbook of the time. On the strength of this and several other mathematical papers he was elected to the Academy of Sciences in 1741.

In 1743 D’Alembert published his *Traité de dynamique (Treatise on Dynamics)*, a book which attracted the attention of the leading scientific authorities and helped secure his reputation as a most promising young savant. In this, his first major publication, he attempted to reconstruct the science of motion (mechanics) on a purely formal and mathematical basis, proscribing appeals to any such quasi-metaphysical notions as force or gravity and assigning observation and experiment to the subordinate role of confirming the basic principles of the theory. The axioms of geometry, together with the assumption of the impenetrability of the objects constituting a system, were all that was needed to establish the precision and certainty that should attend the study of motion. By this time Newtonian mechanics was well on its way toward gaining the ascendancy over the vestiges of Cartesian physics, and D’Alembert was in compliance with the best scientific opinion of the day in rejecting Descartes’ theory of vortices and his conception of nature as a plenum. In certain fundamental respects, however, D’Alembert retained a decided preference for a rigorous, Cartesian rationalism in his approach to scientific theorizing as opposed to the meticulous empirical observation and data-gathering which characterize Newton’s own researches (see EMPIRICISM; RATIONALISM). For D’Alembert priority should always be accorded to the strict logical formulation of the basic principles which grounded any scientific inquiry, and he maintained that such principles could be articulated independently of experience. While many of his contemporaries were impressed by the extent to which D’Alembert had apparently succeeded in preserving and carrying through to completion the best features of the Cartesian approach to physical science, others raised the suspicion that his attempts to mathematicize the foundations of mechanics were based on a confusion regarding the ineluctably empirical dimension of basic mechanical principles. In any case, the *Traité de dynamique* is secure in its place in the history of science for

having clarified the central theoretical issues which remained to be addressed in working out the details of the Newtonian programme, preparing the way, in particular, for the later researches of Lagrange (see DESCARTES, R.; NEWTON, I.).

Over the next few years D'Alembert produced a series of scientific monographs in which he applied the principles worked out in the *Traité de dynamique* to specific problems including the behaviour of fluids, the cause of the winds, precise calculations for the incidence of the equinoxes and the vibration of strings, all impressive scientific achievements in their own right. Beginning in 1746, however, D'Alembert's intellectual energies began to move away from such narrowly focused scientific investigations towards a more literary engagement with the broader currents of the Enlightenment movement that were gaining momentum across Europe. This year marks D'Alembert's earliest involvement in the tempestuous history of the publication of the *Encyclopedia* which he was to co-edit with Denis DIDEROT for the next twelve years. In 1751 he produced the *Preliminary Discourse* to the *Encyclopedia*, the work for which he is best known; he would ultimately contribute, or make substantial editorial additions to some 1,500 articles on an enormous range of topics from mathematics and physics (which make up the largest part of his contribution) to music, philosophy and religion. In many of these articles D'Alembert is unabashedly partisan in his support of the philosophical campaign to secure the claims of science and enlightenment against the resistance of the political authorities and the chauvinism of the religious establishment. By his polemical zeal, for example, in his contentions against the unflagging hostility of the Jesuits to the whole *Encyclopedia* project, he makes known his willingness to be counted among the ranks of the *philosophes* who, in turn, were more than happy to have his considerable reputation as a scientist joined to their cause. The depth of D'Alembert's commitment to the cause was challenged, however, when his article 'Genève' appeared in 1757 in the seventh volume of the *Encyclopedia*. He had intended to present a flattering view of the atmosphere of religious tolerance in Geneva at the time by imputing to the authorities there an advocacy of various doctrines associated with SOCINIANISM, such as the denial of the divinity of Christ. The Genevans were neither flattered nor amused. In the firestorm of criticism and vituperation that came down on his head in the months after his article appeared, D'Alembert resigned his position as editor and withdrew from the public controversies altogether, much to the disappointment of his more vociferous comrades including, notably, Voltaire. His further contributions to the *Encyclopedia* would be confined to relatively uncontroversial topics dealing with scientific matters with which he was especially conversant.

In 1754 D'Alembert was elected to the French Academy, and if the high profile and protracted controversies of the *Encyclopedia* project were too much for D'Alembert's tastes, his membership in the Academy provided the ideal setting for his continued advocacy of the natural sciences. For some years prior to his election, the reputation of the Academy had been diminished in the public's estimation by the perception that its members were perhaps too much beholden to the interests of the political establishment. With D'Alembert things changed. By virtue of his own reputation for scholarly achievement and personal integrity he was able to improve contacts with scientific establishments across Europe and cultivate friendly relations with foreign heads of state. Internally, he worked assiduously in the cause of progressive ideas by setting agendas and seeking to advance the careers of like-minded colleagues. In his speeches before the assembly and public addresses he championed the cause of the philosophic spirit and argued the material and moral advantages to a society in which it was permitted to flourish. In ways that were unprecedented he brought the affairs and proceedings of the Academy before the public and thereby enhanced the reputation of the Academy itself while raising at the same time the status and the influence of scientific and philosophic research. In 1772 he acceded to the office of permanent secretary of the Academy in which capacity he continued to serve until his death.

Throughout this period D'Alembert continued to write and publish books and essays on scientific and philosophical topics. In 1759 he published the *Essai sur les éléments de philosophie*; in the aftermath of the Jesuit expulsion from France in 1762 he published the highly polemical essay *Sur la destruction des Jésuites en France* (1765), in which he argued, in effect, that it was the rising tide of enlightenment which precipitated the downfall of this remaining bastion of religious obscurantism. His technical works in mathematics and physics were published in eight volumes as *Opuscules mathématiques* between 1761 and 1780. His voluminous correspondence, his many speeches and elegies were collected and published posthumously.

2 Philosophical outlook and the *Preliminary Discourse*

D'Alembert, Jean Le Rond (1717-83)

D'Alembert's philosophy is of decisive importance to anyone who would understand the Enlightenment in its intellectual and theoretical (as distinct from its social and political) dimension because it manifests, more clearly perhaps than any other author's, the delicate combination of French rationalism in the tradition of Descartes and Malebranche and British empiricism as represented by Newton and Locke (see Malebranche, N.; Locke, J.). As a mathematician of the very first rank, D'Alembert was entirely sympathetic to Descartes' insistence upon the clarity and precision of basic principles and the need for rigorous deductive logic in establishing the linkages between one theoretical proposition and any others which might be linked with it. Like virtually all his contemporaries, D'Alembert was hugely impressed by the power of Newtonian mechanics to explain and integrate vast tracts of natural phenomena and predict future events, and by the extent to which observational and experimental data could be incorporated into a system constituted at its core by mathematical principles. To understand D'Alembert himself, and to appreciate his importance as a signal representative of the thought of this watershed period in human history, the relationship between the Cartesian and empirical strands of his work must be made out.

D'Alembert was unequivocal in his rejection of metaphysical principles as constituting an appropriate base for inquiry in the natural sciences. He repudiated the Cartesian doctrine of innate ideas and denied that Descartes' *Cogito* had any value as a starting point in constructing the system of human knowledge. Appeals to divine intervention in the natural order had no place in trying to explain the connection between events or the relation between natural objects and the human perception of them. In this respect, D'Alembert adhered closely to the principles of Lockean epistemology as these were modified and expanded in the works of his friend and colleague, CONDILLAC. With Locke and Condillac, D'Alembert insisted that all knowledge begins by attending to the facts and that the way those facts *appear* to us is as close as we can ever hope to get to their reality. But the facts which present themselves to our experience are only isolated and fragmentary pieces of a greater systematic structure in which they are embedded, and for the full articulation of that structure one must have recourse to the principles of mathematics - specifically those of geometry - in order to represent the orderliness of natural events and processes. The basic principles of any science should consist in perfectly unambiguous and precise mathematical formulas together with whatever basic assumptions are necessary to secure their attachment to the objects which constitute the domain under investigation. The various discrete facts disclosed to our observations can then be expressed as formal propositions which could in turn be fitted into the deductive chain of formulas derivable from the basic principles.

D'Alembert rejected Descartes' notion of the universe as a plenum because it was infected by the unwarranted metaphysical doctrine of substances. In his own conception of the universe, however, D'Alembert maintained that all parts of nature are systematically interrelated in a comprehensive structure of laws which evinces the same integrity, consistency and formal precision as any of the more extended and elegant proofs in geometry. Like Descartes, D'Alembert believed that the methodical and painstaking implementation of human reason would gradually serve to disclose this rational structure, and that, as it came into view, the empirical facts could find their place within it and thereby attain theoretical clarity. By setting his first priority on the articulation of the rational structure which informed all of nature, D'Alembert comports with the most fundamental aspects of Cartesian methodology and reaffirms the optimistic faith of his great precursor in the autonomous power of human reason.

In the *Preliminary Discourse* to the *Encyclopedia* D'Alembert presents an interpretation of the contemporary state of scientific research in which his own rationalistic conception of the structure of human knowledge forms the core. The *Encyclopedia* itself is conceived as a compendium of all the scientific truths so far achieved, but the greater utility and value of the work is found to consist precisely in its capacity to arrange and integrate these results so as to facilitate further research by revealing both the systematic interrelations between seemingly remote domains and the lacunas which remained to be filled in. D'Alembert compares the *Encyclopedia* to a road map: certain areas of the terrain of human knowledge are well understood and laid out in careful detail while the expanses between these areas are marked out only by the most rudimentary traces which might none the less provide a researcher with his theoretical bearings as he laboured to extend the boundaries of knowledge ever deeper into the realm of *terra incognita*. All of the articles in the *Encyclopedia* were to be knit together by an extensive network of cross-references so that an arbitrarily selected topic would lead the reader eventually to a comprehensive view of the whole structure in all its intricacies. The *Encyclopedia*, as the systematic embodiment of the whole vast range of human knowledge, would thus reflect the rational structure which constituted the natural order itself.

D'Alembert invokes a second metaphor in the *Discourse*, inspired by the prescient writings of the great English philosopher of science Francis Bacon, in which the 'tree of human knowledge', rooted in such fundamental disciplines as mathematics and logic, branches out into the various scientific subregions and extends ever further to include the most narrowly defined and detailed subject matter. Reconfiguring the Baconian conception of knowledge in accordance with the general principles of epistemology worked out by Condillac, D'Alembert recognizes three essential functions of the human mind: memory, reason and imagination. Under these three rubrics all domains of human inquiry, including history (sacred, civil and natural), literature and art, can be systematically interrelated. This second metaphor expresses D'Alembert's view of knowledge as an organic unity, as something which grows from within when all of the various parts derive their sustenance from the flourishing of the others and contribute in turn to the flourishing of the whole.

To D'Alembert's way of thinking there was no area of human inquiry which could not be incorporated into this unified and integrated structure of knowledge, and in this respect the *Preliminary Discourse* gives powerful and eloquent expression to one of the guiding inspirations of the Enlightenment period as a whole. In his own researches he made significant contributions toward bringing the study of language and aesthetics under the purview of science, and believed that ultimately even politics and morality should have to be subsumed under its methodology if there were to be progress. But here, D'Alembert's attachment to rationalistic principles becomes deeply problematic, in that the more inscrutable and ambiguous aspects of human nature seem unamenable to anything like the geometrical precision that he made the hallmark of all true science. If from our historical vantage point his own philosophical commitments should appear somewhat dogmatic and unwarranted, we must at least acknowledge that the issues concerning the relationships between the various domains of human knowledge to which he was so acutely sensitive continue into our own day, reflected in the uneasy division within institutions of higher learning between the humanities and the 'hard sciences'.

3 Reputation and influence

D'Alembert continues to enjoy a rightly prominent role in virtually all historical treatments of the Enlightenment period. He is one of only a few *philosophes* to have achieved the very highest standing as both a scientist and man of letters. His contributions to mathematics and physics stand as significant landmarks in the history of both disciplines. His indefatigable labours as secretary of the French Academy were profoundly efficacious in advancing the cause of enlightenment across Europe. Some of the most important historical figures of the age, including Voltaire, Frederick the Great and David Hume were proud to count him among their friends, a testament to both his high reputation and the warmth and integrity of his personality. His philosophical work was influential in the world of practical affairs as well, for example, in the formation of economic policy under the ministries of Quesnay and Turgot. And it is by no means the least of his contributions to have served as mentor to and helped to advance the careers of some of the brightest young minds of the rising generation such as Laplace, Lagrange and Condorcet. A close study of D'Alembert's life and works provides an excellent point of entry into the world of the European Enlightenment.

See also: Encyclopedists, eighteenth century; Enlightenment, Continental

PAUL F. JOHNSON

List of works

D'Alembert, J. Le R. (1821-2) *Oeuvres*, ed. A. Belin, Paris: Belin, 5 vols.
(1799) *Oeuvres posthumes*, ed. C. Pougens, Paris: Pougens, 2 vols.
(1743) *Traité de dynamique*, Paris: David.
(1744) *Traité de l'équilibre et du mouvement des fluides pour servir de suite au Traité de dynamique*, Paris: David.
(1747) *Réflexions sur la cause générale des vents*, Paris: David l'aîné.
(1749) *Recherches sur la précession des équinoxes, et sur la nutation de l'axe de la terre dans le systême newtonien*, Paris: David.
(1751-65) *Encyclopédie, ou, Dictionnaire raisonné des sciences, des arts et de metiers, par une société de gens lettrés*, Paris: Briasson.(Contains the *Discours préliminaire* as a preface to the first volume.)
(1751) *Discours préliminaire*, trans. R.N. Schwab, *Preliminary Discourse to the Encyclopedia of Diderot*, Indianapolis, IN: Bobbs-Merrill, 1963.
(1759) *Mélanges de littérature, d'histoire et de philosophie*, Paris: David, 4 vols; vol. 5, 1767.

(1761-80) *Opuscules mathématiques*, Paris: David, 8 vols.
(1765) *Sur la destruction des Jésuites en France, par un auteur désintéressé*, Edinburgh: J. Balfour.
(1766) *An account of the destruction of the Jesuits in France*, Glasgow: R. Urie.

References and further reading

Essar, D.F. (1976) ‘The language theory, epistemology, and aesthetics of Jean Lerond d’Alembert’, *Studies on Voltaire and the Eighteenth Century*, 159.(Contains an excellent discussion of D’Alembert’s researches into the origins and nature of language, and his works on music and the theatre.)

Grimsley, R. (1963) *Jean D’Alembert*, Oxford: Clarendon Press.(Presents a detailed discussion of D’Alembert’s engagement with the *Encyclopedia* and his work in the French Academy; excellent introduction to the life and works of D’Alembert and the French Enlightenment period generally.)

Hankins, T. (1970) *Jean D’Alembert: Science and the Enlightenment*, Oxford: Clarendon Press.(A thorough and more technical examination of D’Alembert’s mathematical and scientific works and sets them in the context of the theoretical controversies of the day. Less suitable as an introductory text, but provides a much more comprehensive analysis of the specific contributions that D’Alembert made to mathematics and mechanics. Includes a very useful bibliography.)

Lough, J. (1968) *Essays on the Encyclopédie of Diderot and D’Alembert*, London: Oxford University Press.(Of use to scholars interested in the publication history of the *Encyclopedia* and the various controversies which attended it. It contains a complete list of the articles that D’Alembert contributed, but on the whole the essays are rather too specialized to be of interest to the general reader.)

Pappas, J.N. (1962) *Voltaire and D’Alembert*, Bloomington, IN: Indiana University Press.(Describes the various ways in which D’Alembert served the cause of the philosophic spirit, with careful discussion of his relations with and influence upon Voltaire.)

Dai Zhen (1724-77)

Dai Zhen, a neo-Confucian philosopher, argues against the received neo-Confucian view of dao *as a metaphysical entity. On the contrary,* dao *is immanent in the world and, in the case of the human world specifically, in the everyday lives of ordinary people irrespective of social status. His philosophical views had important political and social implications.*

Dai Zhen came from Huizhou in Anhui Province. During his early years, Dai often accompanied his father, a cloth merchant, on business trips which brought him in touch with the political and social realities in many parts of China. In 1754 he became entangled in a lawsuit with a powerful clansman who was a friend of the local magistrate and, on being informed that he was about to be arrested, left home in great haste and went to Beijing. This background exerted no small influence on the formulation of his philosophical critique of the dominant Confucian ideology in decades to come.

Dai wrote three important philosophical treatises: *Yuanshan (Inquiry into Goodness)*, *Xuyan (Surviving Words)* and *Mengzi ziyi shuzeng (Commentary on the Meanings of Terms in the Book of Mencius)*. A central concern of all his writings is the immanental status of *dao*. He contends by way of etymology that this is the original meaning of *dao* in early Confucian texts. Understood analytically, *dao* consists of all the principles or patterns (*li*) discoverable in the world. Song dynasty neo-Confucians are wrong, due to the influence of Buddhism, in assuming that *li* is a gift from Heaven and resides in the human mind/heart as if it were a thing. Viewed objectively, *li* is none other than the internal texture and structure in things (see Dao; Li).

True to Confucian humanism, Dai makes an important distinction between what is natural (*ziran*) and what is necessary (*biran*). The former refers to principles or patterns governing all the activities in the world whereas the latter refers to those in the human world exclusively. In this view, not only is the human world continuous with the natural world but the two worlds are also marked by a clear break, thereby distinguishing human beings from all other kinds of beings. He understands the idea of goodness (*shan*) in this light. In his own words: ‘Goodness is what is necessary whereas nature (*xing*) is what is natural (see Xing). What is natural will be fully completed only when it is developed into what is necessary. This is known as developing the natural to the utmost’ (*Mengzi ziyi shuzeng*: 195). This suggests that the world will not be complete without human beings.

Following the same logic, Dai rejects the neo-Confucian bifurcation between moral principles (*li*) on the one hand, and human desires (*yu*) and feelings (*qing*) on the other. Moral norms are capable of regulating human desires and feeling precisely because they are the internal texture and structure in the everyday life of human beings rather than something imposed on it from outside. His objectification of *dao* and *li* naturally leads him to emphasize the central importance of intellectual inquiry known as ‘investigation of things’ (*gewu*) in the Confucian tradition (see Neo-Confucian philosophy §9). Intellect or intelligence (*zhi*) is what distinguishes human beings from other phenomena (see Zhi).

Dai draws important political and social implications from his philosophical views. He believes that the neo-Confucian concept of *li* as Heavenly Principle always serves to justify the oppression of the helpless common people by those in power. The struggle of the common people to satisfy their basic needs in life are denounced as ‘selfish desires’ and anyone who has the audacity to defy the norm imposed on the common people by the ruling elite often ends up being condemned to a moral death. Clearly, he is deeply dissatisfied with the Song dynasty neo-Confucian definition of *li* as ‘heavenly principles’, which has been used by those in power to justify the oppression of the helpless common people. Thus in Dai’s philosophical system, we see an unmistakeable beginning of a modern spirituality uniquely China’s own.

See also: Dao; Neo-Confucian Philosophy; Li; Zhi

YÜ YING-SHIH

List of works

Dai Zhen (1766) *Yuanshan (Inquiry into Goodness)*, in *Dai Zhen guanji (Complete Works of Dai Zhen)*, Beijing: Qinghua University Press, 1991, vol. 1, 9-27; trans. Cheng Chung-Ying, *Tai Chen’s Inquiry into Goodness*,

Honolulu, HI: East-West Center Press, 1971. (Edition and translation.)
Dai Zhen (1769) *Xuyan (Surviving Words),* in *Dai Zhen guanji (Complete Works of Dai Zhen),* Beijing: Qinghua University Press, 1991, vol. 1, 64-116. (No modern translation of this work exists at present.)
Dai Zhen (1777) *Mengzi ziyi shuzheng (Commentary on the Meanings of Terms in the Book of Mencius),* in *Dai Zhen guanji (Complete Works of Dai Zhen),* Beijing: Qinghua University Press, 1991, vol. 1, 149-210; trans. T. Loden, 'Dai Zhen's Evidential Commentary on the Meaning of the Words of Mencius', *Bulletin of the Museum of Far Eastern Antiquities* 60, Stockholm, 1988: 165-313. (Edition and translation.)

References and further readings

Chin Ann-ping and Freeman, M. (1996) *Tai Chen on Mencius*, New Haven, CN: Yale University Press.(An analysis of Dai Zhen's writings on Mencius.)
Yü Ying-shih (1989) 'Tai Chen's Choice Between Philosophy and Philology', *Asia Major*, third series, 11 (1): 79-108.(Article on Dai Zhen's philosophical outlook.)

Damascius (*c*.462-540)

The late Neoplatonist philosopher Damascius was the last head of the Platonist school in Athens. He largely accepted the metaphysical system of the Athenian School of Syrianus and Proclus, but subjected it to acute dialectical scrutiny in a series of commentaries, and especially in his treatise On First Principles. His philosophical position is not comprehensible without bearing in mind that of Proclus, although on certain issues, such as the nature of the first principle and of the soul, he prefers the solutions of the earlier Iamblichus.

Damascius was probably born *c*.462 in Damascus. He studied rhetoric first in Alexandria and then moved to Athens shortly before the death of Proclus in 485. He practised in Athens as a professor of rhetoric and, although he attended lectures in philosophy, he experienced a real ’conversion’ to philosophy only in around 491, under the influence of the Platonist Isidorus, whose life he later wrote. At about this time, following Isidorus, he returned to Alexandria for a spell, where he attended also the lectures of the Platonist Ammonius, son of Hermeas.

Around 510-15, Damascius himself succeeded as head of the Platonist school, and from then until 529 he presided over an intellectual and financial revival of that institution, which had fallen into some obscurity. His students included Simplicius (§§1, 3), who testifies to the acuteness of his intellect and the excellence of his instruction (*On Aristotle's Physics* 624.38).

In 529, the emperor Justinian issued his famous decree against non-Christian teachers. This effectively closed the Platonic school. In its wake (*c*. autumn of 531), following on the accession to the throne of Persia of the young Chosroes, who had a lively interest in philosophy, Damascius and a group of other philosophers decided to emigrate to Persia. They returned after a year, however, under a safe conduct arranged by Chosroes, which enabled them to philosophize in private. Damascius’ last years were probably spent at Emesa in Syria, where he died.

Damascius’ works are only partly preserved, and often in an unsatisfactory state. We have no surviving commentaries by him on Aristotle, but Simplicius, in his commentary *On Aristotle's Physics*, quotes extensively from a special treatise of his *On Number, Space and Time*. His commentaries on Plato have survived somewhat better, but even the surviving ones on the *Phaedo and Philebus* are in the form of students’ notes. His only properly surviving works are the treatise *Problems and Solutions on First Principles*, and his *Commentary on Plato's Parmenides*.

It is chiefly from these two latter works that we can appreciate the particular quality of Damascius’ philosophizing. Situated at what was effectively the end of a long philosophical tradition, he exhibits a remarkable blend of scholastic elaboration and critical acumen. He accepts all the metaphysical convolutions of Syrianus and Proclus, and those of Iamblichus before them. Indeed, one of his distinctive features is a tendency to revive positions of Iamblichus in opposition to those of the later Athenian School, as well as their penchant for syncretistic theology, and such documents as the Chaldaean Oracles and the Orphic Theogonies (see Chaldaean Oracles). In contrast, however, he rigorously puts all the basic principles of Neoplatonism to the test of dialectic, with most stimulating results.

There are four main topics examined in *On First Principles*. First, the nature of the first principle, in which Damascius examines the case for an absolutely ineffable principle prior to the One proper, while analysing the contradictions involved in postulating a first principle which is both entirely simple and transcendent, and the source of all creation and diversity. Second, an analysis of the components of the henadic realm, the triad of One-many (Monad), many-One (Indefinite Dyad) and ‘the unified’, which is the union of the two. Third, arising out of this, a penetrating discussion of the central dynamic of Neoplatonic metaphysics, the triple process of remaining, procession and return (*monē*-*proodos*-*epistrophē*). And finally, the relationship of the intelligible realm as a whole to its parts, the Forms. In his *Commentary on Plato's Parmenides* we find his views on the nature of the soul, which he sees as a radically median entity, subject to change even in its essence.

To ask what is distinctive in Damascius’ philosophy, however, results in a complex answer. His philosophy is unashamedly that of his predecessors. Where he differs from Syrianus and Proclus, as he does on such questions as the nature of the first principle and the economy of the henadic realm, or on the nature and status of the soul, he is often only reverting to the doctrines of Iamblichus. Even his views on space and time, and his postulation of

archetypal forms of both these entities, largely develop positions of Iamblichus. Damascius' particular genius lies rather in the dialectical analysis to which he subjects all these questions, and it is for this that he is valuable.

See also: Neoplatonism §§1, 4

JOHN DILLON

List of works

Damascius' texts cannot be dated relative to each other.

Damascius (*c*.462-540) *Problems and Solutions on First Principles*, ed. L.G. Westerink and J. Combès, *Damascius, Traité des premiers principes*, Paris: Les Belles Lettres, 1986-9, 3 vols.(Contains French translation and notes.)

Damascius (*c*.462-540) *Commentary on Plato's Parmenides*, ed. C.A. Ruelle, *Damascii Successoris Dubitationes et solutiones in Platonis Parmenidem*, Paris: Klincksieck, 1899, 2 vols; repr. Amsterdam, 1966.(Also contains Problems and Solutions on First Principles.)

Damascius (*c*.462-540) *Commentary on Plato's Philebus*, ed. L.G. Westerink, *Damascius, Lectures on the Philebus*, Amsterdam: North Holland, 1959.(Contains text, translation and notes.)

Damascius (*c*.462-540) *Commentary on Plato's Phaedo*, ed. L.G. Westerink, *The Greek Commentaries on Plato's Phaedo* , vol. 2, Amsterdam: North Holland, 1973.(Text and translation.)

(*c*.462-540) *Life of Isidorus*, ed. C. Zintzen, *Damascii vitae Isidori reliquiae*, Hildesheim: Olms, 1976.(Greek text.)

References and further reading

Combès, J. (1989) *Études néoplatoniciennes (Neoplatonist Studies)*, Grenoble: Millon. (Collection of articles in French, mainly on Damascius.)

Gersh, S. (1978) *From Iamblichus to Eriugena*, Leiden: Brill.(Contains much discussion of Damascius' philosophy.)

Sambursky, S. (1982) *The Concept of Place in Late Neoplatonism*, Jerusalem: Israel Academy.(Contains texts and translations relative to Damascius' doctrine on subject of place.)

Sambursky, S. and Pines, S. (1971) *The Concept of Time in Late Neoplatonism*, Jerusalem: Israel Academy. (Contains texts and translations relative to Damascius' doctrine on time.)

Simplicius (after 538) *On Aristotle's Physics*, ed. H. Diels, *Commentaria in Aristotelem Graeca IX-X*, Berlin: Reimer, 1882-5; trans. C. Hagen, *On Aristotle's Physics 7: Simplicius*, London: Duckworth, 1994.

Steel, C. (1978) *The Changing Self, A Study on the Soul in Later Neoplatonism: Iamblichus, Damascius and Priscianus*, Brussels: Paleis der Academien.(An excellent discussion of Damascius' theory of the soul.)

Damian, Peter (1007-72)

Peter Damian is noted for his asceticism, contributions to church reform and literary style, the latter in writings that are primarily religious in character. Because of his hostility to the unbridled use of the disciplines of grammar and dialectic in religious matters, Damian is sometimes depicted as an opponent of philosophy. A more accurate assessment of his attitude is that the liberal arts, including philosophy, must remain subservient to religion. Damian's major work De divina omnipotentia (On Divine Omnipotence) shows that he was willing and able to use philosophical argument in theology.

Peter Damian was born in Ravenna in 1007, entered monastic life with the Benedictine Order at Fonte Avellana around 1036, became prior in 1043, was made cardinal around 1057 and died in 1072. He was actively involved in monastic reform, writing opinions on such topics as simony, married clergy and sexual abuses committed by clergy. Damian's usual mode of writing was the epistle, of which some 180 survive, displaying considerable rhetorical skill on the part of their author.

Damian entered an ongoing controversy over the relation between secular learning - exemplified in the study of grammar, dialectic and philosophy - and divine revelation as contained in the Bible. Although he has sometimes been interpreted as a champion of anti-intellectual fideism, the attitude discernible in Damian's writings is more inchoate than the interpretation suggests. He believes that the Bible contains all the knowledge necessary for personal salvation, and that monks spend their time most profitably in its study. At the same time, however, in insisting that secular learning should be chastened and made subordinate to the study of the word of God, he seems willing to grant secular studies some legitimacy. He offers no delineation of what form that legitimacy might take. The debate about the relation between secular learning and revelation was soon to be subsumed under a larger debate about the respective roles of reason and faith in human understanding. Less than a generation after Damian's flourishing, for example, one can find more sophisticated observations in the writings of Anselm.

In a lengthy letter, now known as *De divina omnipotentia (On Divine Omnipotence)*, to Desiderius, Abbot of Monte Cassino, Damian displays his skills both as philosophical critic and theologian. The question to be determined is whether God's omnipotence includes the ability so to act that what has happened never happened. Damian criticizes Desiderius' claim that God is unable so to act because he does not will so to act. He thinks that Desiderius' claim amounts to (1) God does not will to do *A* if and only if God cannot do *A*, and argues that claim (1), along with the principle (2) God does *A* if and only if God wills to do *A*, entails (3) if God does not do *A*, then God cannot do *A*.

Claim (1) entails that if God does not will to do *A*, then God cannot do *A*. Principle (2) entails that if God wills to do *A*, then God does *A*. Since either God wills to do *A* or God does not will do to *A*, it follows that either God does *A* or God cannot do *A*, which is equivalent to (3). Because (3) means that God is able to do only what God in fact does, Damian thinks that Desiderius' claim (1) must be rejected.

Damian proceeds to offer two separate solutions to the original question. According to the first solution, it is true that everything that has happened necessarily has happened, but that fact is irrelevant to the issue of God's omnipotence. The sentence, 'everything that has happened, necessarily has happened', is trivially true, 'according to the order of speaking'. Although Damian lacked the terminological distinctions that later medieval philosophers had at their disposal, his point can be expressed in their terminology by saying that 'everything that has happened, has happened' is necessarily true *de dicto*, and is not to be confused with the *de re* interpretation, 'everything that has happened has happened necessarily'. Only the latter would impinge on God's omnipotence.

Damian's second resolution maintains that there is a way in which God can still bring it about that Rome, which was founded, never was founded. God lives not in time but in eternity, in which the temporal past, present and future are all equally present to God. Thus, according to Damian, the founding of Rome, although past from our temporal perspective, is eternally present to God's power. As such, God can still cause Rome not to be founded, a situation that we, living after the time of the founding of Rome, describe as God's power to cause Rome never to have been founded.

See also: Eternity; Omnipotence

WILLIAM E. MANN

List of works

Damian, P. (1007-72) *De divina omnipotentia (On Divine Omnipotence)*; in A. Cantin (ed.) *Lettre sur la toute-puissance divine*, Sources Chrétiennes 191, Paris: Éditions du Cerf, 1972.(Contains the complete Latin text with a French translation. The translation by O.J. Blum in *Medieval Philosophy: From St Augustine to Nicholas of Cusa*, eds J.F. Wippel and A.B. Wolter, New York: Free Press, 1969, presents a translation of only a part, with the sections not in sequence.)

Damian, P. (1007-72) Letters, ed. O.J. Blum, Washington, DC: Catholic University of America Press, 1989-.(A part of The Fathers of the Church: Medieval Continuation series. The following volumes have appeared: 1, Letters 1-30, 1989; 2, Letters 31-60, 1990; 3, Letters 61-90, 1992.)

Damian, P. (1007-72) Sermons, ed. G. Lucchesi, *Sancti Petri Damiani Sermones*, Corpus Christianorum, Continuatio Mediaevalis 57, Turnhout: Brepols, 1983.(Contains seventy-five sermons or fragments, twenty-four of which are deemed not to be by Damian.)

References and further reading

Kenny, A. (1979) *The God of the Philosophers*, Oxford: Clarendon Press.(Emphasizes, on pages 100-3, that Damian points out that in the sense in which God cannot alter the past, it would also follow that God cannot alter the future, which is absurd. Also argues that Damian's conception of omnipotence does not commit him to maintaining that God can violate the principles of logic.)

Weinberg. J. (1964) *A Short History of Medieval Philosophy*, Princeton, NJ: Princeton University Press, 60-1. (Claims that Damian maintains that God's omnipotence extends over even the principles of logic.)

Dance, aesthetics of

The aesthetics of dance is the philosophical investigation of the nature of dance, of our interest in it, especially as an art form, and of the variety of aesthetic judgments we make about it - judgments of beauty, grace, line and other aesthetic qualities. Most philosophical issues concerning dance result from considering philosophical questions that arise in other areas: another art form, art in general, or human action. Sometimes the issue for dance can usefully be seen as a combination of issues from other areas. Often, one's response to such issues gives a possible direction for one's thoughts about dance.

A selection of such questions can be taken from the characteristics of dance. Since some dances are works of art, is there a kind of judgment characteristic of an interest in art, and, if so, what are its features? In particular, and in parallel with questions for other arts, does knowledge of the choreographer's intention have any role in understanding the dance? What follows for the understanding of dance from the fact that dance is a multiple art: that a particular dance (like a particular piece of music) can be performed in London at the same time as it is performed in New York? What follows from dance's status as a performing art (like music)? Is any special role for the understanding of dance to be assigned to a notated score in a dance notation? (If so, does this differ from music?) How is the special place of the dancer to fit into accounts of understanding dance? As some dances are regularly thought to be communicative, how does dance differ (if at all) from so-called 'nonverbal communication'? More generally, dance study must address far-reaching philosophical issues: the place of dance as human action; the 'role' of dance, for example, in ritual; the relevance of the history and traditions of dance-forms to the understanding of those forms.

1 Introduction

Dance has not been a major topic for philosophical enquiry for the contemporary aesthetician, because aestheticians have focused on problems in literature, music or painting, and because dance is viewed correctly as having few problems of its own: for example, discussion of the place of the choreographer's intention for criticism or understanding of his work will largely reproduce similar discussions for other artists (see Artist's intention) and theorists might be expected to take similar positions in both cases. Again, any discussion of dance must acknowledge that only some dances can be considered to be works of art, just as only some paintings are considered works of art. For dance, just as for other 'objects of aesthetic appreciation', aestheticians must ask whether the appreciation of art works is importantly different from the rest of 'aesthetic appreciation' or whether, instead, these cases should be understood in terms of one, underlying experience.

More specifically, the status of dance as a multiple art form, which allows the same dance to be performed in two or more places at the same time, and as a performing art are both shared by music. Thus, questioning the appropriateness of, for example, a type/token analysis for the relationship between the work itself and performances of it will be shared with the aesthetics of music, and it will be reasonable to expect analogous answers. Equally, issues resulting from the recording of dance using notation will be shared with those in music. This entry considers a selection from the many issues potentially lying within the ambit of the aesthetics of dance and provides remarks on each.

2 Dance, art and aesthetic appreciation

Does the philosophical aesthetics of dance concern only those dances or dance-forms that are, or could be, art forms, or should its net be cast more widely? In so far as its central interest lies in art-type dance, the aesthetics of dance will reflect concerns beyond the mere appreciation of grace, line, beauty and other aesthetic qualities in human movement. Then any value attributed to dance can begin from ascriptions of whatever value is proper to art, and issues in the aesthetics of dance will align with those in the aesthetics of other arts, especially other performing arts.

3 The distinctiveness of dance?

Dance is characterized by its *physicality*: it is an essentially physical art form and in this it differs from music, literature and visual art. Music and visual art require human action for their creation and existence, but the art objects are not physical (that is, not human action) in quite the same sense. Moreover, most dance - in contrast to

music - has a ‘mixed’ character: a typical dance performance involves not merely movement but costume, music, staging and the like. (In this respect, the typical dance resembles the typical opera.) What exactly are the roles of these other elements, some already having status as art? For example, attending the ballet *The Rite of Spring* involves hearing Stravinsky’s music. How is this to be integrated into our analysis of the art form? Of particular interest here are the ways in which elements of costume - often thought variable across different performances - can sometimes be entirely fixed by the choreographer, for example, in certain works by Alwin Nikolais. Equally, elements normally taken as fixed - for example, the music - may, in certain works, turn out not to be completely determined. For instance, Twyla Tharp’s *The Bix Pieces* (1971) has a variable musical element.

4 Dance as a multiple art and as a performing art

Dance and music share their performing character: one encounters the work of art itself only in some performance of it. So performing arts generate, as it were, two objects for discussion: the work itself and particular performances of it - *Swan Lake* itself and, say, Tuesday night’s performance of it. This fact has interesting implications. First, the dance itself is an abstract object - one never meets it, although, in another way, seeing the performances constitutes seeing it. Second, the dance can be seen in two places at the same time - by groups seeing different performances. So *Swan Lake* could be performed simultaneously in London and New York. Third, London’s *Swan Lake* and New York’s *Swan Lake* will differ, since they use different casts and are performed on stages of different sizes. In London, the producer may emphasize certain characteristics of *Swan Lake*, perhaps to suit the dancers in the cast. For similar reasons, different aspects may be stressed when the dance is staged in New York.

A type/token contrast (see Type/token distinction) that takes the work itself as the type and the particular performance as the token might be introduced here. Such an analysis would generate the three points noted above. Since *Swan Lake* is a type, it follows that it is an abstract object (the first point). As an abstract object, it is encountered only through encountering tokens: hence there can be more than one token at any particular time (the second point). But all the various performances are equally *Swan Lake*, despite their differences (the third point).

However, the type/token distinction cannot itself determine what is to count as a token of a particular type. There is a fine borderline between restaging a dance and re-choreographing to produce a different dance: for example, are some of Nureyev’s productions re-choreographed or merely restaged? This consideration recognizes a difficulty in deciding whether a particular performance is or is not a token of a particular type. For recognizing re-choreography as the production of a different work of art is a way of saying that its performances are tokens of a different type.

Three possible objections to a type/token analysis of dance illustrate the complexity of this issue. First, has the analysis identified the crucial elements? One might think that, as with music, the interpretation of the performer or of the stager was also of crucial importance: yet the proposed analysis leaves no room for it. Second, does the type/token contrast focus on the right characteristics of multiple art form or is Nelson Goodman’s autographic/allographic contrast more appropriate? (See Art works, ontology of §4.) Third, does the analysis tell us enough? Are we not still struggling to decide both when we have a token of a particular type and how the type ‘constrains’ its tokens?

5 Dance notation and understanding

One characteristic feature of dance as art is the possibility of its notation. There are three well-developed notation systems for dance: Labanotation, Benesh and Eshkol Wachman. Goodman has sketched the criteria of adequacy for any notation system for any art form: for example, the symbols must be clearly differentiated in order to pick out different states of affairs. One practical question is, ‘To what degree do the notation systems mentioned above fulfil such criteria?’ Alternatively, a philosophical issue arises concerning the uses to which notated scores might be put. Might they, for instance, play a role in the solution of (numerical) identity questions, such that any performance satisfying a notated score for *Swan Lake* was thereby bound to be an instance of that art work?

Two aspects of such notation systems have direct implications for the aesthetics of dance. First, dance notations radically under-determine the dance, by not guaranteeing that the movements they describe are dance rather than something else; by telling us nothing about costume, staging, music and the like; and by not fully determining the

movements to be performed. Notated scores embody perceptions of the movements in question. The movements are objects that have been analysed, in the choice of what are to count as the key characteristics of a particular pattern of movement, and are not potential objects for analysis. So notation systems offer *interpretative* descriptions rather than neutral documentations of movements. Second, recording a particular dance in different notation systems imposes different sets of constraints on the dancer. The two recordings might be satisfied by slightly different sequences of movements. For instance, a performance which satisfied a notated score for *Swan Lake* in Labanotation might fail to do so for a score in Benesh, and vice versa. All that can be urged is that only reputable notation systems count here: those knowledgable about an art form decide what is reputable.

Further, positive use of the phrase ‘those knowledgable about a particular art form’ is an admission that, in the end, it is a matter for judgment by dance critics whether or not something is an instance of a particular work of art. If Nureyev’s changes mean that his production no longer satisfies a particular notated score for *Swan Lake*, one alternative would be to doubt the appropriateness of that notation. It may, then, be a matter for art criticism, or more precisely for the critics of dance, whether a particular performance genuinely counts as a performance of *Swan Lake*.

6 The role of the dancer

The aesthetics of any performing art must address the role of the performer. The choreographer is the central artist, yet dancers should not be considered mere puppets of the choreographer, even when they have not been specifically involved in the initial creation of the particular dance. (As dances can typically be performed at a later date, there is an inherent possibility of dancers performing previously choreographed pieces.)

Revealingly, the term ‘interpretation’ is used in two distinct ways in performing arts, which I shall call ‘critic’s interpretation’ and ‘performer’s interpretation’. The first characteristically consists of strings of sentences that discuss the structure and value of the work in question. But a pianist’s rendering of a particular piano piece is also an interpretation. Here interpretation typically consists of some set of actions that are performed in producing the witnessable work. For dance, it consists in performing (at least) movements of the body. This second sort - performer’s interpretation - is unique to the performing arts, while critic’s interpretations are possible for all works of art. So any problems we have in understanding the idea of critic’s interpretation, or in seeing the basis of such interpretation are problems for all the arts equally. But concerns about making the interpretation an interpretation of a particular work of art, and of no other, and about ruling out inappropriate interpretations, have two dimensions for the performing arts. They apply not only to critic’s interpretation - a difficulty shared with other art forms - but also to performer’s interpretation.

When one witnesses a work in the performing arts, it has already become a performer’s interpretation. Works in any performing art are under-determined at their creation just as they are under-determined by their notations: they are only brought to determinacy by performance. In this sense, works in the performing arts typically come as recipes, with an instruction to perform them - an instruction to produce a performer’s interpretation. But once that interpretation is complete, the dance in that performance is determinate, and hence in the same position as works of the particular object sort, or non-performing works of the multiple sort, such as works of literature. Each has a physical instantiation which is available for critic’s interpretation and such critic’s interpretation will draw on the traditions and conventions of that art form. One important function of the dancer is thus as the creator of the performer’s interpretation. We can therefore understand the virtues of one dancer over another in terms of their respective abilities to produce interesting and sustainable performer’s interpretations.

7 Verbal and nonverbal communication

The distinctiveness of dance’s expressive character seems to have no parallel in music, and so requires different treatment. How, if at all, does the expressiveness of dance relate to the more general expressiveness of bodily movement: to so-called ‘nonverbal communication’? For the ‘instrument’ of dance is that very same human body in which we recognize, for instance, the sadness of another person through their posture, their words, and so on.

Activities that embody the intention to communicate can be distinguished from those that do not, even though actions of the latter type can teach us something about those who perform them - for example, their state of mind. Such a contrast, if sustainable, clearly places much dance, with verbal language and sign language, in one

category, and much else that is nonverbal in the other. If we distinguish art-type dance from dance that is not art, then any expressive or communicative power of the latter would be that of ‘other movement’. But should the expressive and communicative power of art-type dance be treated in the same way? If we conclude that it should, we have scarcely any reason for drawing the distinction between our interest in art and that in non-art. This in turn has implications for the value of dance, and in particular its educational value.

8 General philosophical issues

Finally, a significant number of general philosophical questions raise important issues for the aesthetics of dance. For instance, dance is a kind of human action: but how should we recognize or value, if at all, the dances of chimps or, in the future, of robots? Tentative solutions to various problems in the philosophy of action may be needed before events can be adequately characterized as dance at all, especially given the diversity of dances. Again, considering ‘dance in other societies’ raises a number of familiar questions concerning the beliefs of other cultures or societies. For instance, having recognized that the North American Indian ghost dances have a specific purpose (namely, to rid the continent of the white man), we will have difficulty concluding that they are art; can we even conclude that they are dance? For philosophical purposes the issue differs little from questions arising from, say, the Lascaux cave paintings.

As with other art forms, a developed aesthetics of dance would explore both the history of dance, at least as an art form, and the history of dance criticism. Thus, the question will arise of when it is appropriate to date the beginnings of dance. Should we, in parallel with Kristeller’s work on the concept ‘art’, think of art-type dance coming into being with the baroque ballet in, say, the court of Louis XIV? Or should we conclude that, since the Greeks had a muse for dance (Terpsichore), there must have been art-type dance in classical Greece? Resolving such questions is partly a matter for history, but also calls for philosophical reflection on the implications of calling the activity *dance* rather than something else.

See also: Action; Artistic interpretation; Music, aesthetics of; Art, performing

GRAHAM McFEE

References and further reading

Copeland, R. and Cohen, M. (eds) (1983) *What is Dance?*, Oxford: Oxford University Press. (A useful anthology, including both the work of philosophers and of dance critics and dance theorists. Especially useful for some classic sources.)

Fancher, G. and Myers, G. (eds) (1981) *Philosophical Essays on Dance*, Brooklyn, NY: American Dance Festival Inc.(An anthology bringing together artists, critics and philosophers, with responses to some of the papers by choreographers and other dance practitioners. Some pieces difficult.)

Goodman, N. (1968) *Languages of Art*, Indianapolis, IN: Bobbs-Merrill.(A classic text on notation, with a distinctive view of the nature of symbolism and of representation in art. Complex and technical.)

Kristeller, P.O. (1951) ‘The Modern System of the Arts: A Study in the History of Aesthetics I’, *Journal of the History of Ideas* 12: 495-527; and (1952) ‘The Modern System of the Arts: A Study in the History of Aesthetics II’, *Journal of the History of Ideas* 13: 17- 46; both repr. as ‘The Modern System of the Arts’, in *Renaissance Thought and the Arts*, Princeton, NJ: Princeton University Press, 1990.(Two important papers arguing for an eighteenth-century beginning to a roughly contemporary conception of the nature of art.)

McFee, G. (1992) *Understanding Dance*, London: Routledge.(An introductory text raising all of the issues discussed here and focusing on the art form of dance.)

(1994) *The Concept of Dance Education*, London: Routledge.(Application to education of ideas from the aesthetics of dance, taking the provisions of the National Curriculum for England and Wales as an example. Introductory.)

Sheets-Johnstone, M. (ed.) (1984) *Illuminating Dance: Philosophical Explorations*, Cranbury, NJ: Associated University Presses.(A collection including work by many of the leading figures in the aesthetics of dance. Some articles technical.)

Sparshott, F. (1988) *Off the Ground: First Steps to a Philosophical Consideration of Dance*, Princeton, NJ: Princeton University Press.(A systematic attempt to address the whole of the aesthetics of dance. Especially good on questions other than those relating to dance as an art form. Some technicality.)

Sparshott, F. (1995) *A Measured Pace: Towards a Philosophical Understanding of the Arts of Dance*, Toronto: University of Toronto Press.(Companion to the above, approaching the art forms of dance with due emphasis

on the perspective of the dancer. Grounded in detailed account of the arts in general. Some passages technical.)

Dao

Dao, *conventionally translated 'the Way', is probably the most pervasive and widely recognized idea in Chinese philosophy. The specific character of Chinese philosophy arises because a dominant cultural factor in the tradition, now and then, has been the priority of process and change over form and stasis, a privileging of cosmology over metaphysics. That the Yijing (Book of Changes) is first among the Chinese Classics in every sense bears witness to the priority of cosmological questions - how, or in what way (*dao*) should the world hang together? - over metaphysical and ontological questions - what is the reality behind appearance, the Being behind the beings, the One behind the many, the true behind the false? The contrast lies in finding a way rather than seeking the truth.*

To contrast the difference between the search for metaphysical 'truth' - the way things really are - and the mapping of an appropriate and productive 'way' (*dao*) in the human world, we might consider the philosophical questions being asked by representative texts. The project in the Confucian *Analects*, for example, is not to speculate on *what* the ultimate source of value in the world might be, but to recount biographically *how* one sensitive person, Confucius, made his way in the world as a possible model for others (see Confucius). The *Daodejing* does not purport to provide an adequate and compelling description of what *dao* and *de* might mean as an ontological explanation for the world around us; rather, it seeks to engage us and to provide guidance in how we ought to interact with the phenomena, human and otherwise, that give us context in the world (see Daodejing). Likewise the *Yijing* is not a systematic cosmology that seeks to explain the sum of all possible situations we might encounter in order to provide algorithmic insight into what to do, but is a resource providing a vocabulary of images that enable us to think through and articulate an appropriate response to the unique and always changing conditions of our lives (see Yijing).

Although the way of living to be discovered in the canonical texts, the *dao*, has historical antecedents, it is not simply to be discovered and walked. As the *Zhuangzi* 2 says, 'The path is made in the walking of it' (see Zhuangzi). *Dao* means both to lead along a path, and to be led along it. In the Confucian *Analects* 15.29 we read, 'It is the human being that can broaden the way (*dao*), not the way that broadens the human being'. The human being must be a roadbuilder because human culture - the human *dao* - is always under construction.

The parts of speech which order Western languages - nouns, verbs, adjectives and adverbs - encourage us to divide up the world in a culturally specific manner. Under the influence of these grammatical determinants, we are inclined to separate things from actions, attributes from modalities, where from when and when from what (see Syntax). However, given the fluidity between time, space and matter assumed in classical Chinese cosmology, these categories do not govern the way in which the world is divided. The categories used to define a Chinese world must be seen as crossing the borders of time, space, and matter. *Dao* is both 'what is' (things and their attributes) and 'how things are' (actions and their modalities). *Dao* has as much to do with the subjects of knowing and their quality of understanding as it does with the object of knowledge and its attributes. There are no clear lines between things and events, and hence we cannot separate 'the Way'as *what* from a way as *how*.

In classical Western metaphysics, the equivocation between one as 'unity' and one as 'uniqueness' tends to be resolved in favour of 'unity'. Thus, in any of the various conceptions of a single-ordered universe assumed by the early systematic philosophers, the many phenomena comprising the world are defined in accordance with unifying first principles which determine the essential reality of the things of the world (see Cosmology). *Dao* is often taken to be this familiar search for the One behind the many.

In classical Chinese reflections on world order, however, the equivocation between one as 'unity' and as 'uniqueness' is usually resolved in favour of the unique. The natural cosmology of classical China does not require a single-ordered cosmos, but invokes an understanding of a 'world' constituted by 'the ten thousand things' (see Chinese philosophy §1). There is no Being behind the beings - only beings are. And *in toto*, these beings are *dao*. Continuity makes *dao* one; difference makes *dao* myriad; change makes *dao* processional and provisional. *Dao* is thus both the One and the many, or better, the field and foci through which it is entertained.

The Chinese 'world as such' is constituted by 'worlding' (*ziran*), a process of spontaneous arising, or literally,

uncaused ‘self-so-ing’, which references no external principle or agency to account for it. The one and the many stand in a holographic relationship: there is the indiscriminate field (*dao*) and its particular focus (*de*). *Dao* as field is always entertained and focused from some perspective or another, from some particular *de* (see De). Just as in a holographic display where each detail contains the whole in an adumbrated form, so each item of the totality focuses the totality as its particular field.

See also: Chinese philosophy; Confucius; Daodejing; Daoist philosophy; De; Qi; Yijing

DAVID L. HALL
ROGER T. AMES

References and further reading

Graham, A.C. (1989) *Disputers of the Tao: Philosophical Argument in Ancient China*, La Salle, IL: Open Court. (A survey of ancient Chinese thought.)

Hall, D.L. and Ames, R.T. (1995) *Anticipating China: Thinking Through the Narratives of Chinese and Western Culture*, Albany, NY: State University of New York Press.(A comparative study of the uncommon assumptions that ground the Chinese and Western philosophical traditions.)

Hansen, C. (1992) *A Daoist Theory of Chinese Thought*, New York: Oxford University Press.(An account of ancient Chinese philosophy that emphasizes language and *dao* as a guiding discourse.)

Daodejing

The Daodejing (or Tao Te Ching) is a brief work probably composed during the period 350-250 BC. It later became the most authoritative 'scripture' in the Daoist religious and philosophical tradition, and in modern times has become among the most often-translated and popular works in world religious literature. It recommends cultivating mental calm, an intuitive, non-conceptual understanding of the world, an integrated and balanced personality, a self-effacing manner and a low-key and non-intrusive leadership style. One who has this spirit has dao *(*tao*), which is also conceived of as a cosmic reality, the origin of the world.*

1 Origin and composition

We have no reliable direct information about the origins of the *Daodejing* (the title means 'Classic of *Dao* and *De*'). Its traditional attribution to the legendary and later deified sage Laozi (Lao Tzu) is now widely questioned (see Chan (1963); Graham (1990)). It is traditionally divided into eighty-one very brief numbered 'chapters', each probably consisting of several oral sayings originally coined in a close-knit community, with unifying editorial additions. The term 'Laoist' is used here to refer to the thought of this book so as to differentiate it from other contemporary and later forms of Daoism (see Daoist Philosophy).

2 Concrete advice

The *Daodejing* repeatedly criticizes egotistic self-assertion and boastful self-promotion, and competing with others for social status. It also criticizes 'working' at impressing others by external refinements such as polished speech and cultivated manners; this is the basis for its criticism of contemporary Confucian self-cultivation. Many paradoxical sayings praise ways of being at the opposite extreme: self-effacing public service, appearing 'empty', 'compromised' or 'dull and stupid', and cultivating softness, 'weakness' and femininity (in contrast to the masculine strength much admired by contemporaries).

The *Daodejing* warns against the mental disturbance caused by external stimulation and the attraction of 'desirable' things. It advocates finding contentment in oneself (*Daodejing* 44, 46) and cultivating mental stillness (16, 52). This will give one an inexhaustible source of energy, ensuring good health and enabling one to 'last long', possibly gaining immortality (33).

The *Daodejing* criticizes confidence in the ability of conceptual ('naming') knowledge to grasp reality, probably a criticism of Mohist logicians and the Confucian programme of 'rectifying names' (see Logic in China). It emphasizes the unpredictability of the world as well as the impossibility of capturing true norms in exact verbal formulas. One must rely instead on the intuitive understandings of a properly cultivated mind. The ultimate basis for norms (*dao*) is 'nameless' and can only be conveyed by 'wordless teaching' (*Daodejing* 2, 43).

Much of the *Daodejing* consists of advice about how to rule, probably reflecting its origins in a group of *shi*, men aspiring to administrative or advisory positions in government (Hsu 1965). It criticizes contemporary rulers who exploit peasant society for their own advantage (53, 72, 75), or who rely on physical violence (war-making and capital punishment) to achieve their ends (30, 31, 69, 74). It also strongly criticizes ruling styles that are in any way overbearing or intrusive. A ruler should not strive to be 'awesome' (72), even by trying to set a high moral example for the people (this is probably an anti-Confucian polemic). Again many paradoxical sayings advocate the extreme opposite of these normal tendencies of rulers. Rather than lording it over his subjects, he should attract their allegiance and cooperation by his deferential attitude, treating them as though he is someone 'lower' than they (61, 66, 68). Rather than searching for ego-gratification by high-profile achievements, the best kind of ruler strives to go unnoticed, giving the people the impression that social order is happening 'naturally' (17). A ruler should not try to mould the society according to his own values and ideas; this is 'working on' society, in contrast to which Laoists advise 'not working' (*wuwei*), 'taking the mind of the people as his mind' (49), and fostering and maintaining the organic social harmony already present in the communities he governs ('helping along [their] naturalness' (64)). Rather than being strict, the ruler should have the 'dull and incompetent' appearance (58) of one with a 'muddled mind' (49). When he has to intrude, he should be as non-confrontational as possible, defeating opposition by 'soft' indirect means, or by nipping bad tendencies in the bud (36, 64).

The *Daodejing* expresses opposition to several contemporary 'progressive' movements, such as the encouragement

of personal ambition or rational-utilitarian thought among the peasantry, as a means of increasing material production and political strength. This causes discontent and contention and upsets social harmony, making it difficult for rulers to maintain a healthy (that is, simple traditional agrarian) social order.

3 Deeper dimensions

Many passages in the *Daodejing* speak about a deeper dimension of reality, most notably those that picture *dao* as the origin of the world; the *Daodejing* and the *Zhuangzi* are largely responsible for giving dao the pregnant meaning it came to have in later Chinese tradition (see Dao; Zhuangzi). However, these passages present a set of overlapping images rather than a system of doctrines, and there is no very explicit indication as to how they relate to each other or to the concrete advice outlined above. Consequently, interpretations of this facet of the *Daodejing*'s thought vary widely, and at present nothing can be regarded as settled.

Philosophically inclined modern interpreters have tended to construe this area of Laoist thought on the model of Western 'metaphysics'. On this view, teachings about *dao* constitute an objective transcendent foundation of the worldview taught by the *Daodejing* as a whole. The implicit model here is very often (Fung 1937; Kaltenmark 1969; Schwartz 1985) speculative mysticism, as found in Neoplatonism or Vedānta (see Neoplatonism; Vedānta), or German Idealism (see German Idealism) as found in Hegel and Heidegger: *dao* is an unmanifest Absolute Being underlying but pervading the phenomenal world of beings. Its 'namelessness' refers to the fact that it is beyond all finite determinations that characterize the world of appearances. One then has to posit also a 'manifest' aspect of *dao* which has some positive and definite content, in order to ground the obvious bias of the *Daodejing* towards one side of the various polarities it poses (softness versus hardness, femininity versus masculinity and so on). There are also, however, several anti-mystical philosophical interpreters who construe *dao* in a more naturalist way as something immanent in the world, or as something like 'natural law' (Needham 1962). Finally, Hansen (1992) thinks the *Daodejing* argues a primarily negative and sceptical thesis, that no socially conditioned value judgments have any objective grounding.

However, it can also be argued that self-cultivation rather than speculative philosophy or mysticism is the background for the thought of the *Daodejing* (Roth 1991), and that its teachings about *dao* are to be construed imaginally rather than as philosophical metaphysics (Girardot 1974). On this view (argued in LaFargue 1994), the *Daodejing* imagines a primordial level of human conscious being that is inactive, unaroused, undifferentiated (*pu* (uncarved), *Daodejing* 15, 28, 32, 19), and not involved in conceptual thought ('nameless' and *tong* (merged), 1, 56). Society in general does not value this, but gravitates instead towards more active, 'excited' states, towards the specialized development of prized and admired 'virtues', towards grasping reality by means of refined conceptual (learned) thought, and towards striving to make one's mark on the world. Each of these disturbs the state of organic harmony, both internal and social, that Laoists regard as 'natural'. Laoists want to reverse this conventional value-orientation, insisting that the inactive, undifferentiated, organically harmonious state is the more valuable, and indeed the source of all that is most valuable in life. Expressed in chronological imagery, this is the 'prior' or 'root' state; people ought to reverse their normal tendencies and 'turn back' to cultivating this level of their being (Chapter 10 contains meditation instruction connected with this self-cultivation practice). The idea is not to remain literally inactive, but to 'dwell in' this layer of one's being and let it express itself in the manner of one's active involvement in the world. This will be a manner different from those whose value-orientation is more conventional: it will make one appear feminine rather than masculine, soft and weak rather than hard and strong, and 'empty' (that is, without a strongly felt, sharply defined presence) rather than solid.

One who cultivates this level of being experiences more mental clarity and also has a different perspective, dominated by a distinctive value-orientation. The world appears different to such a person than it does to conventional society. (This is the view of the world expressed in the more practical aphorisms of the *Daodejing*.) Laoists are aware that the character of this changed world has its origin in the state of mind they cultivate. They imagine this state to be a deeper layer of the world itself, expressed by saying it is 'the world's origin'.

On this view, *dao* was for Laoists primarily an internal spirit to be cultivated, imagined as an independent cosmic force or entity informing an ideal state of mind and expressing itself in certain ideal patterns of conduct. This 'subjective' reality of *dao*-as-experienced served as a basis for the vision of a world with *dao* as its source. The *Daodejing* had not yet transformed this into an objectified doctrine about the world that (as in the metaphysical interpretation) a person could adopt intellectually and 'believe in', even in the absence of self-cultivation and the

concrete experience of *dao* (although such an objectifying transformation did happen rather soon after, and can be seen already in the earliest commentaries of Han Feizi and Wang Bi.)

Laoist 'mysticism' was not speculative - mystical insights did not serve as the basis for a set of doctrines about reality - nor was it unitive - Laoists did not strive for union with *dao* for its own sake. Rather, it was developed in practical directions. Dao was valued as a constantly dependable and nourishing internal 'mother' (*Daodejing* 20), sustaining those whose values deprived them of external social support. It also had a strongly political dimension: *dao* as the highest reality gave whoever possessed it the spiritual status of 'pattern/norm for the world', taking over this religious role which previous tradition had assigned to the Zhou Emperor. When such a person attained an administrative or advisory position in government, the *dao* he possessed would ideally inform the policies he promoted, and above all his leadership style (32, 37). Thus, as *dao* radiated from his person as *de*, subtle 'charisma', it would set the tone for the entire society and become the foundation for an ideal social order. This union of 'spiritual' and political interests is a distinctive feature of the *Daodejing*, at least in its final recension; some scholars argue that early Laoist thought was apolitical, and that applying Laoist ideas to problems of governing represents a later stage of development.

See also: Dao; Daoist philosophy; De; Self-cultivation, Chinese theories of

MICHAEL LAFARGUE

References and further reading

1 *Daodejing*: recommended translations

Addis, S. and Lombardo, S. (trans.) (1993) *Lao Tzu Tao Te Ching*, Indianapolis, IN: Hackett.(A succinct, fairly literal translation.)

Carus, P. (trans.) (1913) *The Canon of Reason and Virtue-being Lao-Tze's Tao te king. Chinese and English*, Chicago, IL: Open Court.(Gives English equivalent of each Chinese character in the text; however, works with an unreliable Chinese text.)

Chan Wing-tsit (trans.) (1963) *The Way of Lao Tzu (Tao te ching)*, Chicago, IL: University of Chicago Press. (Translation with useful commentary and introduction, recommended for novice reader; includes a critical review of traditional evidence concerning 'Laozi' on pages 35-59.)

Duyvendak, J.J.L. (trans.) (1954) *The Book of the Way and its Virtue*, London: John Murray; repr. Boston, MA: Charles E. Tuttle, 1992.(Translation with useful textual analysis and contemporary parallel passages.)

Henricks, R.G. (trans.) (1989) *Lao-Tzu Te-Tao Ching*, New York: Ballantine.(A translation of the newly discovered Mawangdui manuscripts of the *Daodejing* dating from *circa* 200 BC, with useful notes on differences between this and traditional manuscripts.)

Karlgren, B. (trans.) (1975) 'Notes on Lao Tze', *Museum of Far Eastern Antiquities Bulletin* 47: 1-13.(An exceedingly literal translation by a noted authority on ancient Chinese language.)

LaFargue, M. (trans.) (1992) *The Tao of the Tao Te Ching*, Albany, NY: State University of New York Press. (Translation, detailed commentary and essays on main themes; analyzes text into oral sayings and editorial additions; translation recommended for novice reader.)

Lau, D.C. (trans.) (1963) *Lao Tzu Tao Te Ching*, Baltimore, MD: Penguin.(A widely used translation with long introduction.)

Lau, D.C. (1982) *Tao Te Ching*, Hong Kong: Chinese University Press.(Translates both the traditional and the Mawangdui texts, with long introductions to each; Chinese and English on facing pages.)

Waley, A. (trans.) (1958) *The Way and its Power*, New York: Grove Press. (Translation with good historical introduction and some commentary.)

Secondary works

Chan, A.K.L. (1991) *Two Visions of the Way: A Study of the Wang Pi and Ho-shang Kung Commentaries on the Lao-Tzu*, Albany, NY: State University of New York Press.(Detailed comparison of the two earliest Chinese commentaries on the *Daodejing*.)

Creel, H.G. (1970) *What is Taoism?*, Chicago, IL: University of Chicago Press, 1-47. (First proposed the now widely accepted distinction between the 'purposive' Daoism of the *Daodejing* and the 'contemplative' Daoism of *Zhuangzi*.)

Erkes, E. (trans.) (1950) *Ho-Shang-Kung's commentary on Lao-Tse*, Ascona: Artibus Asiae Publishers.(A

translation of an early Chinese commentary (*circa* AD 180), which reads into the *Daodejing* the more complex self-cultivation theories of later Daoism, while still retaining the idea that self-cultivation is preparation for ruling well.)

Fung Yu-lan (1937) *A History of Chinese Philosophy*, Vol. 1, *The Period of the Philosophers*, Peiping: Henri Vetch, 170-91.(A metaphysical interpretation of *dao*, with descriptions also of the *Daodejing*'s practical advice, by a Chinese scholar who trained also in Western philosophy.)

Girardot, N. (1974) *Myth and Meaning in Early Taoism*, Berkeley, CA: University of California Press, 47-76. (Emphasizes the theme of primordial undifferentiated 'chaos' in the *Daodejing*.)

Graham, A.C. (1990) 'The Origins of the Legend of Lao Tan', in *Studies in Chinese Philosophy and Philosophical Literature*, Albany, NY: State University of New York Press, 111-24.(A detailed analysis of the historical development of the legends surrounding 'Laozi', and how the *Daodejing* came to be attributed to him; intricate historical argumentation will be difficult for the general reader.)

Han Feizi (*c.*280-233 BC), bk. 6, chaps 20-1, trans. W.K. Lao, *The Complete Works of Han Fei Tzu*, London: Arthur Probsthain, 1938.(The earliest commentator on selected chapters from the *Daodejing*, showing how the *Daodejing*'s teaching can be integrated into a Legalist framework.)

Hansen, C. (1992) *A Daoist Theory of Chinese Thought: a Philosophical Interpretation*, New York: Oxford University Press, 210-30.(Argues that the purpose of the *Daodejing* is to undermine socially induced value judgments by showing that no way of evaluatively 'naming' things has any objective and unchanging basis; for philosophically advanced readers.)

Hsu Cho-yun (1965) *Ancient China in Transition: An Analysis of Social Mobility, 722-222*, Stanford, CA: Stanford University Press.(The best sociological analysis of China at the time of the *Daodejing*'s composition; see especially the discussion of *shi* (pages 86-106), the class to which the *Daodejing*'s authors probably belonged.)

Julien, S. (1842) *La Livre de la Voie et de la Vertu (The Book of the Way and Virtue)*, Paris: L'Imprimerie Royale. (One of the earliest Western translations of the *Daodejing*; includes interesting excerpts from many traditional Chinese commentators.)

Kaltenmark, M. (1969) *Lao Tzu and Taoism*, Stanford, CA: Stanford University Press, 19-69.(A detailed metaphysical interpretation of *dao*.)

LaFargue, M. (1994) *Tao and Method: a Reasoned Interpretation of the Tao Te Ching*, Albany, NY: State University of New York Press.(Argues the interpretation of the *Daodejing* presented in this article. Detailed discussion of the social background of *Daodejing* and of major methodological problems in its interpretation; detailed philosophical and methodological discussions will be difficult for general reader.)

Lau, D.C. (1958) 'The Treatment of Opposites in Lao-tzu', *Bulletin of the School of Oriental and African Studies* 21 (2): 344-60.(Surveys Chinese interpreters that assimilate *Daodejing* to the 'balancing of opposites' theme of the *Yijing*; argues that the *Daodejing* consistently favors the soft/weak over the hard/strong.)

Needham, J. (1962) *Science and Civilisation in China*, Vol. 2, *History of Scientific Thought*, Cambridge: Cambridge University Press, 33-164.(Argues that the *Daodejing* reflects proto-scientific attempts to discover natural laws, arriving however at a more 'organicist' view that contrasts with the atomism of modern Western science; it also rejects Zhou feudalism in favor of more 'primitive' organic communities. Difficult for the general reader.)

Roth, H. (1991) 'Psychology and Self-Cultivation in Early Taoistic Thought', *Harvard Journal of Asian Studies* 51 (2): 599-650.(Discusses self-cultivation practices reflected in other early 'Daoistic' writings, important for understanding the *Daodejing*'s background in self-cultivation. Difficult for the general reader.)

Schwartz, B. (1985) *The World of Thought in Ancient China*, Cambridge, MA: Harvard University Press, 192-215.(The most persuasive and nuanced version of the mystical/metaphysical interpretation of *dao* in the *Daodejing*; includes also a detailed discussion of its relation to the concrete advice given; for the general reader.)

Wang Bi (AD 226-249) *Commentary on the Lao Tzu by Wang Pi*, trans. A. Rump, Honolulu, HI: University of Hawaii Press, 1979.(Translation of an insightful early Chinese philosophical commentary.)

Daoist philosophy

*Early Daoist philosophy has had an incalculable influence on the development of Chinese philosophy and culture. Philosophical Daoism is often called 'Lao-Zhuang' philosophy, referring directly to the two central and most influential texts, the Daodejing (or Laozi) and the Zhuangzi, both of which were composite, probably compiled in the fourth and third centuries BC. Beyond these two texts we might include the syncretic Huainanzi (*circa *140 BC) and the Liezi, reconstituted around the fourth century AD, as part of the traditional Daoist corpus.*

Second in influence only to the Confucian school, the classical Daoist philosophers in many ways have been construed as both a critique on and a complement to the more conservative, regulatory precepts of their Confucian rivals. Daoism has frequently and unfortunately been characterized in terms of passivity, femininity, quietism and spirituality, a doctrine embraced by artists, recluses and religious mystics. Confucianism, by contrast, has been cast in the language of moral precepts, virtues, imperial edicts and regulative methods, a doctrine embodied in and administered by the state official. The injudicious application of this yin-yang-*like concept to Daoism and Confucianism tends to impoverish our appreciation of the richness and complexity of these two traditions. Used in a heavy-handed way, it obfuscates the fundamental wholeness of both the Confucian and Daoist visions of meaningful human existence by imposing an unwarranted conservatism on classical Confucianism, and an unjustified radicalism on Daoism.*

There is a common ground shared by the teachings of classical Confucianism and Daoism in the advocacy of self-cultivation. In general terms, both traditions treat life as an art rather than a science. Both express a 'this-wordly' concern for the concrete details of immediate existence rather than exercising their minds in the service of grand abstractions and ideals. Both acknowledge the uniqueness, importance and primacy of the particular person and the person's contribution to the world, while at the same time stressing the ecological interrelatedness and interdependence of this person with their context.

However, there are also important differences. For the Daoists, the Confucian penchant for reading the 'constant dao*' myopically as the 'human* dao*' is to experience the world at a level that generates a dichotomy between the human and natural worlds. The argument against the Confucian seems to be that the Confucians do not take the ecological sensitivity far enough, defining self-cultivation in purely human terms. It is the focused concern for the overcoming of discreteness by a spiritual extension and integration in the human world that gives classical Confucianism its sociopolitical and practical orientation. But from the Daoist perspective, 'overcoming discreteness' is not simply the redefinition of the limits of one's concerns and responsibilities within the confines of the human sphere. The Daoists reject the notion that human experience occurs in a vacuum, and that the whole process of existence can be reduced to human values and purposes.*

*To the extent that Daoism is prescriptive, it is so not by articulating rules to follow or asserting the existence of some underlying moral principle, but by describing the conduct of an achieved human being - the sage (*shengren*) or the Authentic Person (*zhenren*) - as a recommended object of emulation. The model for this human ideal, in turn, is the orderly, elegant and harmonious processes of nature. Throughout the philosophical Daoist corpus, there is a 'grand' analogy established in the shared vocabulary used to describe the conduct of the achieved human being on the one hand, and the harmony achieved in the mutual accomodations of natural phenomena on the other.*

The perceived order is an achievement, not a given. Because dao *is an emergent, 'bottom-up' order rather than something imposed, the question is: what is the optimal relationship between* de *and* dao*, between a particular and its environing conditions? The Daoist response is the self-dispositioning of particulars into relationships which allow the fullest degree of self-disclosure and development. In the Daoist literature, this kind of optimally appropriate action is often described as* wuwei*, 'not acting wilfully', 'acting naturally' or 'non-*assertive *activity'.* Wuwei*, then, is the negation of that kind of 'making' or 'doing' which requires that a particular sacrifice its own integrity in acting on behalf of something 'other', a negation of that kind of engagement that makes something false to itself.* Wuwei *activity 'characterizes' - that is, produces the character or ethos of - an aesthetically contrived composition. There is no ideal, no closed perfectedness. Ongoing creative achievement itself provides novel possibilities for a richer creativity.* Wuwei *activity is thus fundamentally qualitative: an aesthetic category and, only derivatively, an ethical one.* Wuwei *can be evaluated on aesthetic grounds, allowing that some*

relationships are more productively wuwei *than others. Some relationships are more successful than others in maximizing the creative possibilities of oneself in* one's *environments.*

This classical Daoist aesthetic, while articulated in these early texts with inimitable flavour and imagination, was, like most philosophical anarchisms, too intangible and impractical to ever be a serious contender as a formal structure for social and political order. In the early years of the Han dynasty (206 BC*-*AD *220), there was an attempt in the Huainanzi to encourage the Daoist sense of* ethos *by tempering the lofty ideals with a functional practicality. It appropriates a syncretic political framework as a compromise for promoting a kind of practicable Daoism - an anarchism within expedient bounds. While historically the Huainanzi fell on deaf ears, it helped to set a pattern for the Daoist contribution to Chinese culture across the sweep of history. Over and over again, in the currency of anecdote and metaphor, identifiably Daoist sensibilities would be expressed through a range of theoretical structures and social grammars, from military strategies, to the dialectical progress of distinctively Chinese schools of Buddhism, to the constantly changing face of poetics and art. It can certainly be argued that the richest models of Confucianism, represented as the convergence of Daoism, Buddhism and Confucianism itself, were an attempt to integrate Confucian concerns with human community with the broader Daoist commitment to an ecologically sensitive humanity.*

1 Classical Daoism

Before attempting to characterize the lineaments of Daoist philosophy, a brief discussion of the meaning of the term 'classical Daoism' is necessary. Unlike Confucianism, the 'school' of Daoism, *daojia*, does not name a tradition constituted by a founding thinker. Indeed, the term 'Daoism' is itself an *ex post facto* creation of subsequent editors and historians.

Daoism is first named as a school by Sima Tan in the *Shiji (Records of the Historian)* in the second century BC; and because Sima Tan's concern was classifying schools in terms of their contributions to 'the art of rulership', there was no expressed connection at this time between Zhuangzi and Laozi, later recognized as the two principal Daoist thinkers. In Sima Tan's classification, the Daoists are listed as one of six schools of thought, the others being the Yin-Yang school (see Yin-yang), the Confucians (see Confucian philosophy, Chinese), the Mohists (see Mohist philosophy), the Legalists (see Legalist philosophy, Chinese) and the School of Names (see Logic in China).

In the first century AD, Ban Gu in the *Hanshu (History of the Former Han Dynasty)* included the writings of Zhuangzi and Laozi together under the heading of 'Daoism' as one of nine philosophical schools. By the first quarter of the fourth century, the principal texts of Daoism were recognized as the *Daodejing* (also called the *Laozi*) and the *Zhuangzi* (see Daodejing; Zhuangzi), along with the respective commentaries by Wang Bi (AD 226-49) and Guo Xiang (d. early fourth century), texts such as the *Guanzi* and *Heguanzi* and writings attributed to the Yellow Emperor (Huangdi). In 1973, an early Han tomb was excavated at Mawangdui and two silk copies of the *Daodejing* were discovered, along with four texts originally attributed to the Yellow Emperor that have a distinctly Legalist flavor. The bundling of these texts together suggests an attempt to borrow the authority of Emperor Huangdi, the legendary founder of the Chinese state, for a Legalist interpretation of Laozi's doctrines. There is some scholarly opinion that it was this Legalistic Daoism, called Huang-Lao, rather than Lao-Zhuang Daoism which actually defined the school during the Han dynasty.

A further complication in explaining the sense of 'Daoism' results from the fact that 'the School of the Way', *daojia*, is to be distinguished from *daojiao*, what may be termed the 'the Doctrine of the Way', with the latter traditionally dating from the middle of the second century. *Daojiao* is the foundation of the organized Daoist religion which has persisted down to the present day. On the surface at least, many of the rituals and doctrines of this religious Daoism do not seem to cohere with the spirit of the more philosophical form of Daoism. Until recently scholars of Daoism, particularly those in the West, have typically drawn a rather sharp distinction between the *daojia* and *daojiao*. More recently, with the translations of Daoist 'religious' texts of the medieval period, there have been attempts to identify a 'mystical philosophy' in the *daojiao* which has its roots in the thought of Laozi and Zhuangzi. It should also be said that many of the contemporary practitioners of so-called religious Daoism distinguish themselves from the philosophical Daoists in terms of the esoteric or 'inner' teachings and practices constituting the central identity of their community versus the exoteric or 'outer' teachings of the *daojia*. Having acknowledged this distinction, there is some contemporary scholarly concern regarding the tendency to over-intellectualize the philosophical tradition at the expense of a meditative and hygenic regimen that

seems implicit in these texts.

Perhaps it is fitting that the historical roots of what has come to be identified as Daoist thinking should be obscure. In fact, relatively fluid identity of the Daoist tradition has in part been responsible for its widespread influence beyond the borders of China. Both the *Daodejing* and the writings of Zhuangzi have become classics of world philosophical and spiritual literature, taking their place as expressions of that *philosophia perennis* which touches the deepest layers of human sensibility. Certainly the hundreds of translations of the *Daodejing* attest to this status.

The usual distinction made between the classical forms of Confucianism and Daoism is in terms of the most general context with which each is concerned. Confucians focus upon issues relevant to a ritualized human society (see Confucian philosophy, Chinese; Confucius), while Daoists concern themselves with the natural ambiance. Such a distinction has but qualified merit, however, since it really characterizes only the overall emphasis of the contrasting models of thinking. The familial or bureaucratic metaphor employed by the broad Confucian tradition is most relevant to the concerns of social structure and harmony, though Han speculations led to the extension of this metaphor to the broadest of cosmological issues. In a similar manner, the more naturalistically inclined Daoists begin with the ‘Six Realms’ - heaven, earth and the four directions - but are quite capable of bringing the model of accommodation, developed with reference to the ten thousand things, to bear upon concrete social and political concerns in a variety of both constructive and critical manners. Indeed, the *Daodejing* is generally considered to be a political manual dealing with the proper guidance of society.

Thus the distinction between Daoism and Confucianism, at what we in the West would term the ‘philosophical’ level, is perhaps not as great as the majority of commentators have claimed. In fact, were we to search for something like a central insight or commitment giving rise to the most general speculations associated with the two visions, we might find that there is a ‘single thread’ running through both Confucianism and Daoism. The central focus of both Confucian and Daoist forms of thinking is the form of engagement which might be termed ‘deference’. In Confucianism one may discuss this deferential activity in terms of the notion of *shu*, a sort of analogical activity which encourages one to put oneself in another’s place as a strategy for determining appropriate conduct. This activity is broadly shaped by a system of rituals (*li*) which characterize the structures of society and the roles through which various human activities are expressed (see Law and ritual in Chinese philosophy).

In Daoism, deference involves a yielding (and being yielded to) grounded in a recognition of the uniqueness of that to which deference is paid. One may best express the meanings of deferential activity in Daoism through the notions of *dao*, *de* and the so-called *wu*-forms: *wuwei*, *wuzhi* and *wuyu*.

2 Fundamental notions: *dao*

> *Dao* engenders one,
> One engenders two,
> Two engenders three,
> And three engenders the myriad things.
> The myriad things shoulder *yin* and embrace *yang*,
> And mix the *qi* to achieve harmony.

(*Daodejing* 42)

Dao is the process of the world itself, the ‘way’ of things. In Daoism, the relevant contrast is not between the cosmological ‘whatness’ of things and the ontological ‘thatness’ of things, but rather between the world as ‘chaos’ (*hundun*) - the sum of all orders - and any given world as construed from some particular perspective (*de*), that is, in terms of any particular one of the orders (see Dao).

Dao is not organic in the sense that a single pattern or *telos* could be said to characterize its processes. It is not *a* whole, but many such wholes. It is not a One to which the Many reduce. Its order is not rational or logical, but aesthetic; there is no transcendent pattern determining the existence or efficacy of the order. The order is a consequence of the particulars comprising the totality of existing things. This interpretation of *dao* makes of it a totality not in the sense of a single-ordered cosmos, but rather as the sum of all potential orders. Any given order is an existing world that is construed from the perspective of a particular element within the totality; but, as a single

world, it is a selective abstraction from the totality of possible orders. Such a limited order cannot serve as fundament or ground. In the Daoist sensibility, all differences are cosmological differences, that is, differences among particular components which, at some moment, comprise the order.

The Daoist understanding of the world is not as an ordered whole but as ‘the ten thousand things’ (*wanwu*). As this term indicates, there is in effect no concept of ‘cosmos’ at all, insofar as that notion could be said to entail a coherent, single-ordered world which is any sense enclosed or defined. The Daoists are therefore primarily ‘acosmological’ thinkers. In the absence of some overarching *archē* or ‘beginning’ as an explanation of the creative process, and under conditions which are thus ‘anarchic’ in the philosophic sense of this term, although nature might indeed be conventionally classified in terms of ‘kinds’, such categories would by no means be natural kinds. Difference is always prior to similarities.

The Chinese binomial most frequently translated as ‘cosmos’ is *yuzhou*, a term that overtly expresses the interdependence between time and space. For the Daoists - indeed for the classical Chinese generally - time pervades everything and is not to be denied; it is not derivative of matter, but a fundamental aspect of it. Unlike traditions which devalue time and change in pursuit of the timeless and eternal, in classical China things are always processional and transforming (*wuhua*) and hence are always provisional. In fact, in the absence of some claim to objectivity that ‘objectifies’ and thus makes ‘objects’ of phenomena, the Chinese tradition does not have the separation between time and entities that would allow for either time without entities or entities without time: there is no possibility of either an empty temporal corridor or an eternal anything (in the sense of being timeless).

What encourages us within a Western metaphysical tradition to separate time and space is our inclination inherited from the Greeks to see things in the world as fixed in their formal aspect, and thus as bounded and limited (see Time §1). But if, instead of giving ontological privilege to the formal aspect of phenomena, we observe them in light of their ceaseless transformation, we are able to temporalize them and perceive them as ‘events’ rather than ‘things’, where each phenomenon is some current or impulse within a temporal flow. In fact, the pervasive capacity of the world to transform continuously is the meaning of time itself. It is because things in the world are reproductive that time is reproductive. Since the world is always entertained from one perspective or another, and since the temporal aspect is never abstract, fictive or replicated, any perspective is an advancing path that is neither linear nor cyclical, but both. Time-space-matter is an advancing spiral. The expression ‘advancing path’ is particularly appropriate, because it is this image of our passage in the world that is captured in *dao* as a grounding metaphor pervasive in the tradition (see Processes).

The most provocative characterization of *dao* in the *Daodejing* is as both ‘nameless’ and ‘nameable’:

> *Dao* can be spoken of, but that is not constant *dao*;
> The name can be named, but that is not the constant name;
> The nameless is the beginning of the heavens and the earth;
> The named is the mother of the myriad things.
> Dwell constantly in nothing in order to observe its mysteries;
> Dwell constantly in something in order to observe its boundaries.

(*Daodejing* 1)

If *dao per se* is the ‘way’ of things, construed as all of the processes of becoming, then nameless and nameable *dao* characterize the functions of ‘not-being present’ and ‘being present’ respectively, as abstracted from the process of becoming-itself. *Dao* is the *that which*: that which *is* and that which *is not* are the polar elements of ‘becoming-itself’.

3 Fundamental notions: *de*

It would be easy to misunderstand this interpretation as claiming a far greater coherence for the Daoist vision than is likely. One must look carefully at the meanings of the explanatory categories *you* and *wu*, conventionally translated as ‘being’ and ‘not-being’, in the Chinese tradition (see You-wu). For example, *wu* means ‘have not’ or ‘there is not’, which readily contrasts with the sense of not-being as ‘nothing’ or ‘no entity’. The sense of *wu* is of the absence of specific concrete things. The correlative sense of *you* in this context is the presence of these same concrete things.

If one looks at Daoism with the appropriately adjusted senses of 'being' and 'not-being', the notion of *dao* as the 'that which' loses its ontological tone. There is only 'this' and 'that'. To translate this idea as 'only beings are' is quite appropriate provided one does not look for any Being standing behind or beneath or beyond these beings, and provided one does not interpret 'are' in an existential sense. The best manner of avoiding such a confusion is simply to say, since there are only 'thises' and 'thats', the locution 'only beings are' is more aptly expressed as 'beings only' or 'these beings'.

Beings are 'thises' and 'thats'. *Dao*, construed as 'that which is' and 'that which is not' - nameable and nameless *dao* - characterizes the process of existence and experience as 'becoming-itself'. This reflexive notion, however, does not name an 'itself' which becomes; rather, the locution 'becoming-itself' refers simply to the unsummed processes of becoming *per se*. It is these processes that are entailed by the notion of *dao*.

What this means, of course, is that Daoism expresses a radical acosmology. If we wish to understand the sense in which 'difference' is to be understood relative to the parity of things, we must restrict ourselves solely to the question of the differences among particular things. Each particular element in the totality has its own *de*. *De* is best understood as a particular focus that orients an item in a field of significances such that it achieves its own intrinsic excellence (see De). The *de* of an element provides the perspective from which it construes all other items in its environs. In this manner each item with respect to its *de* 'names' and creates a world. *Dao* and *de* are related as field and focus respectively. The relations of *dao* and *de* are 'holographic': that is, each element in the totality of things contains the totality in an adumbrated form. The particular focus of an item establishes its world, its environment; the totality as a noncoherent sum of all possible orders is adumbrated by each item.

The *Zhuangzi* contains a passage in which this notion of 'locus' or 'place' is presented as being integral to what it means to know:

> Zhuangzi and Huizi were strolling across the bridge over the Hao river. Zhuangzi observed, 'The minnows swim out and about as they please - this is the way they enjoy themselves.'
>
> Huizi replied, 'You are not a fish - how do you know what they enjoy?'
>
> Zhuangzi returned, 'You are not me - how do you know that I don't know what is enjoyable for the fish?'
>
> Huizi said, 'I am not you, so I certainly don't know what you know; but it follows that, since you are certainly not the fish, you don't know what is enjoyment for the fish either.'
>
> Zhuangzi said, 'Let's get back to your basic question. When you asked '*From where* do you know what the fish enjoy?' you already knew that I know what the fish enjoy, or you wouldn't have asked me. I know it from here above the Hao river.'

(*Zhuangzi* 17)

Zhuangzi is not just depending upon linguistic ambiguity in order to win an argument. He has a more philosophic point to make: he wants to deny the objectivity of knowledge in the sense of the independence of the world as known, from the knower. For Zhuangzi, knowledge is performative and a function of fruitful correlations. It is a 'realizing' of a world in the sense of 'making it real', and the knower and the known are inseparable aspects of this same event. Agency cannot be isolated from action. One and one's posture or perspective is thus integral to what is known; where you are and how you know are mutually entailing. Knowledge is proximity. Zhuangzi's experience with the fish makes his world continuous with the world of the fish, and as such, his claim to knowledge is a claim to having been there. Being continuous with the fish and collaborating with them in the experience does not deny the fish their difference. In fact, it is only through Zhuangzi's deference to their difference - allowing them to be what they are - that the experience can be optimally fruitful and he can really come to know these particular fish (see Knowledge, concept of).

The Daoist belief in the 'the parity of things' celebrates the *differences* that exist among the items of the totality. The manner in which all things are *similar* is easily stated: although the *de* of things, being ever unique, are hierarchical in their relations one to the other, at the same time they have parity in the sense that each *de* is necessary for every other *de* to be what it is. With no Being behind the beings of the world, the way of things is both continuous and radically perspectival. There can be no standards such as the 'great chain of being' or the

'ladder of perfection' establishing ontological hierarchies (see Being §§1-2). What we have instead is both the uniqueness of each perspective, and parity among them.

4 Person and world: the *wu*-forms

Having indicated the sense of parity affirmed by the Daoists, and the deferential activity such an understanding entails, we may now attempt to grasp in greater detail how the Daoists seek to respond to their natural and social environs. As noted earlier, in Confucianism, deference (*shu*) operates within ritual patterns (*li*) shaped in accordance with the familial relationships extending outward into society and culture (see Confucian philosophy, Chinese). By contrast, Daoism expresses its deferential activity through what we shall call the *wu*-forms, *wuzhi*, *wuwei* and *wuyu*: that is, 'no-knowledge', a sort of knowing without resort to rules or principles; 'no-action', or actions in accordance with the *de* (particular perspective) of things, and 'no-desire', or desiring which does not seek to own or control its 'object' (which in effect makes it an 'objectless desire'). In each of these cases, it is necessary to remain sensitive and responsive to the particular perspective (*de*) of what is to be known, what is to be acted in accordance with, or what is to be desired.

Since the *wu*-forms bear some skewed resemblance to the faculties of the tripartite *psychē* which has played such an important role in Western speculations since Plato, we shall begin by noting the contrasting problematics underlying the distinctive interpretations of knowledge, action and desire in classical Daoism and dominant strains of the Western tradition. The Platonic *psychē*, and its permutations throughout the subsequent history of Western thought, has provided us with a wealth of interpretations of the modalities of knowing, acting and desiring which comprise the structure of the self (see Psychē). The heritage of the self construed as a tripartite structure is one of conflict, the soul at war with itself. In Plato, the primary conflict of course is between knowledge and ignorance, reflected in such mottos as 'To know the Truth is to do the truth' and 'You shall know the Truth and the Truth shall make you free'. Allied with this conflict, internal to reason itself, is that tension between the appetitive and the spirited elements on the one hand, and the rational elements on the other. This is essentially a conflict between mind and body, ramified with the confluence of Hebraic and Hellenic sensibilities. It is well expressed in the words of St Paul: 'The good that I would do, I do not do; the evil that I would not do, that I do', and receives perhaps its profoundest expression in Augustine's *Confessiones* in the form of a war between passions and the will, one demanding the satisfactions of desire, the other seeking obedience to the will of God (see Augustine). The conflict is again succinctly stated in a familiar doctrine of David Hume: 'Reason is and always shall be, a slave to the passions' (see Dualism; Emotions, philosophy of; Human nature).

These permutations of the soul at war are so deeply embedded in our own intellectual culture as to shape unconsciously much of our analysis of alternative psychological and cultural sensibilities. It is essential, therefore, to recognize that the model of the tripartite self has little relevance to the dynamics of Chinese models of the person, and the relation of those models to their larger social and cultural context. A signal of this fact is that there is no fundamental mind-body dualism with which to contend; and since this dualism sets up the principal conflict within the self between reason on the one hand and the affective and volitional components on the other, any conflict in the Daoist understanding of person obviously cannot be between heart and mind, or mind and will. In fact, the goal of cultivation of the self for the Daoist is predicated upon a different dynamic altogether: the attempt to move from activities of knowing, feeling and action shaped by construal to those shaped by deference (see Self-cultivation, Chinese theories of).

Having said all this, there are terms such as *zhi*, *wei* and *yu* which initially seem to correlate rather closely with what we in the Western tradition call knowing, acting and desiring. If the Daoist self is not divided in the manner of the Western model of the tripartite soul, how are we to account for these three modalities? The *wu*-forms must be thought of as activities that establish the deferential relations which constitute the self at any given moment.

5 Person and world: the *wu*-forms (cont.)

Wuzhi, as 'no-knowledge', means the absence of a certain kind of knowledge, specifically the sort dependent upon ontological presence (see Zhi). Knowledge grounded in a denial of ontological presence does not presuppose a single-ordered world and its intellectual accoutrements. It is therefore 'unprincipled knowing', the sort of knowing that does not appeal to rules or principles determining the existence, meaning or activity of a phenomenon. *Wuzhi* provides one with a sense of the *de* of a thing, its particular focus, rather than yielding some understanding of that

thing in relation to some concept or natural kind or universal. Ultimately, *wuzhi* is expressed as a grasp of the *daode* relationship of each encountered item, which permits an understanding of the totality construed from the particular focus (*de*) of that item.

Wuzhi, or 'knowing without principles', is tacit and, though inexpressible in literal terms, may be communicated though parabolic and imagistic language. A Confucian critic challenged the Daoist claim which begins the *Daodejing* - '*Dao* can be spoken of, but it is not constant *dao*' - by asking: 'If constant *dao* cannot be spoken of, how is it that the author of the *Daodejing* used several thousand characters in speaking of it?' The Daoist replied: 'I make for you a beautiful embroidery of drakes and pass it along to you for your admiration. I cannot, however, show you the golden needle by which it was made.' Such parabolic language does not presuppose a literal ground. Parabolic language, therefore, is constitutive of discourse itself. Language is, from the beginning, a language of difference and particularity. It is this language that permits the communication of the results of *wuzhi*.

Knowledge, as unprincipled knowing, is the entertainment of the world on its own terms without recourse to rules of discrimination which separate one sort of thing from another sort. The *Huainanzi*, a Han dynasty eclectic text with a strong Daoist colouration (see Huainanzi), reports on this kind of mirror-like 'knowing':

> Therefore, the sage is like a mirror -
> He neither sees things off nor goes out to meet them,
> He responds to everything without storing anything up.
> Thus, he is never injured through the myriad transformations he undergoes…
> It is because the mirror and water do not install cleverness beforehand that when they come into contact with shapes that are square, round, bent and straight, they are unable to escape being shown for what they are.
>
> (*Huainanzi* 6 and 1)

Rules of thumb, habits of action, customs, fixed standards, methods, stipulated concepts, categories, commandments, principles, laws of nature: all these require us to 'welcome things as they come and escort them as they go'. Having stored past experience and organized it in terms of fixed standards or principles, we anticipate, celebrate and recall a world patterned by these discriminations. The sage, however, mirrors the world and 'neither sees things off nor goes out to meet them'. As such, he 'responds to everything without storing anything up'. This means that he mirrors the world *at the moment* in a way that is undetermined by the shape of a world passed away, or by anticipations of a world yet to come.

6 Person and world: the *wu*-forms (cont.)

Wuwei, often misleadingly translated as 'no action' or 'non-action', involves the absence of a specific sort of action, namely, coercive action which interferes with the particular focus (*de*) of those things within one's field of influence. Actions untainted by stored knowledge or ingrained habits are unmediated, unstructured, unprincipled and spontaneous. As such, they are consequences of deferential responses to the item or event in accordance with which, or in relation to which, one is acting. These actions are *ziran*: 'spontaneous' and 'self-so-ing'. They are *nonassertive* actions.

The 'grand analogy' in the *Daodejing* is that *dao* is to world as ruler is to people. *Dao*, as the discernible rhythm and regularity in the world around us, is non-impositional: '*Dao* is constantly non-assertive (*wuwei*) yet there is nothing which is not done' (*Daodejing* 37).

In the Chinese exercise form known as *taijiquan*, there is a practice known as 'push-hands' (*duishou*) which well-illustrates *wuwei*. Two individuals facing one another perform various circular movements of the arms while maintaining minimal hand contact. The movement of each individual mirrors that of the other. *Wuwei* is realized when the movements of each are sensed, by both parties, to be uninitiated and effortless - that is, *spontaneous*. Presumably, one can employ such mirroring responses in 'the art of rulership', providing a sort of nonconstruing model for the guidance of the state. Such is the testimony of the *Daodejing*:

> Thus, the sage says:
> I am nonassertive (*wuwei*)
> And the people transform themselves;

I cherish stillness,
And the people attune themselves;
I do not intervene,
And the people are prosperous of their own accord;
I am objectless in my desires,
And the people retain their natural genuineness of their own accord.

(*Daodejing* 57)

And again:

[The sage] constantly causes the people to seek ‘unprincipled understanding’ and to be objectless in their desires,
And as for the erudite - they wouldn’t dare to do anything.
In simply acting non-assertively,
Everything is properly ordered.

(*Daodejing* 3 (following the Mawangdui text for the last line))

This attitude is carried over into the human world: in government, impositional power is a major concern so the consummate political model in Daoism, corresponding to *dao*, is described as *wuwei* and *ziran*:

The most excellent ruler: the people do not even know that there is a ruler.
The second best: they love and praise him.
The next: they stand in awe of him.
And the worst: they look on him with contempt.
Inadequate integrity in government
Will result in people not trusting those who govern.
So hesitant, the ruler does not speak thoughtlessly.
His job done and the affairs of state in order,
The people all say: ‘We are naturally like this [*ziran*].’

(*Daodejing* 17)

Spontaneous action is a mirroring response. As such, it is action which accommodates the other to whom one is responding on their own terms. Such spontaneity involves recognizing the continuity between oneself and the other, and responding in such a way that one’s own actions, while serving one’s own ends, at the same time promote the interests and well-being of the other. This does not lead to imitation but to complementarity. Handshakes and embraces are actions which both presuppose a recognition of the stance of the other, and complete that stance. When the music starts to play and your partner opens his arms, the dance proceeds as a dyadic harmony of nonassertive actions: provided, of course, you and your partner defer to one another.

7 Person and world: the *wu*-forms (cont.)

Perhaps the best characterization of the term *wuyu* is ‘objectless desire’. Since neither unprincipled knowing nor nonassertive action can in the strict sense objectify a world or any element in it - that is, make discrete and independent objects out of one’s environment - the kind of desiring associated with the Daoist sensibility is in the strictest sense ‘objectless’. The ‘enjoyments’ associated with *wuyu* are possible without the need to define, possess or control the occasion of one’s enjoyment.

Thus, *wuyu*, rather than involving the cessation of desire, represents the achievement of *deferential desire*. Desire, based upon a mirroring understanding (*wuzhi*) and a nonassertive relationship (*wuwei*), thus allowing for the integrity of whatever is entertained, is shaped not by the desire to own, control or consume but simply to celebrate and to enjoy. Desire is for those things desirable in part because they *stand to be desired*. But those things which stand to be desired must themselves be deferential, which means that they cannot *demand* to be desired; for to demand to be desired is to seek a kind of seductive control over the desirer. In a world of events and processes in which discriminations are recognized as conventional and transient, desire is predicated upon the ability at any

given moment to ‘let go’. It is in this sense that wuyu is a nonconstruing, objectless, desire.

The problematic of desire familiar to the Western tradition is expressed by Plato’s definition of Eros as ‘the son of Resource and Need’. Erotic desire is predicated upon a motivation to attain an absent object. Once attained, desire ceases. The resolution of the problem of desire in Plato comes from learning what is truly desirable - namely, knowledge. Complete attainment of knowledge is not possible in this life, so desire continues as an unabated, ultimately harmonizing, motivation to increase one’s understanding, rather than as a movement from one desired object to the next in a frustrated attempt to achieve final satisfaction (see Desire).

The Daoist problematic with respect to desire does not concern what is desired, but rather the manner of the desiring. Objectless desire always allows for letting go. Enjoyment for the Daoist is realized not in spite of the fact that one might lose what is desired, but because of this fact. The world is a complex set of processes of transformation, never at rest. In Plato, the desire for knowledge is the only thing that can define both embodied and disembodied existence; it is the only desire that can be permanent and eternal. In Daoism, transient desire is the only desire that lets things be, that does not construe the world in a certain manner, and that does not seek to apply the brakes on a world of changing things.

The key to understanding *wuyu* - indeed of all these *wu*-forms that are constitutive of the Daoist self - lies in the contrast between ‘objects’ and ‘objectivity’. From a Western perspective, it appears that the thoughts expressed in both the *Zhuangzi* and the *Daodejing* represent something like what would in the West be termed a realism (see Realism and anti-realism). Beyond the confusions introduced by language, and by our own distorted perceptions and tendentious categorizations, there is, with properly Daoist qualifications, an ‘objectively’ real world. Our task is to entertain that world as ‘objectively’ as possible. The problem begins when we believe that the objective world is a world of objects: concrete, unchangeable things which we encounter as over against us, things which announce themselves to us by saying ‘I object!’ For the Daoist, the objective world cannot be objective in this sense. It is a constantly transforming set of events or processes which belie the sorts of discriminations that would permit a final inventory of the furniture of the world:

> Maojiang and Lady Li were beauties for human beings, but fish upon seeing them would seek the deeps, birds on seeing them would fly high, and deer upon seeing them would dash off. Which of these four understands what is really handsome in this world?
>
> (*Zhuangzi* 2)

The moment we begin to discount these other views of Lady Li, we have drained a great deal of significance from the notion of ‘beauty’ by setting up exclusive standards which determine not only the truly beautiful but the unacceptably ugly. Further, these fixed standards - whether of beauty or goodness or justice - are the means of creating a world of fixed objects. Those things or situations which meet the standards are themselves defined and delimited, closed to any other changes than those involved in degradation and decay, that is, failure to realize the demands set by the standard.

Paradoxically, therefore, for the Daoist the objective world is objectless. The sage envisions a world of transforming events which may, for whatever reason, be frozen momentarily into a pattern of discrimination, but which can be recognized, when seen clearly, as beyond such distinctions:

> There is nothing which is not a ‘that’, and nothing which is not a ‘this’. Because we cannot see from a ‘that’ perspective but can only know from our own perspective, it is said that ‘that’ arises out of ‘this’ and ‘this’ further accommodates ‘that’. This is the notion that ‘this’ and ‘that’ are born simultaneously. And even though this is so, being born is simultaneously dying and vice versa; being acceptable is simultaneously being unacceptable and vice versa; accommodating right is accommodating wrong and vice versa. It is for this reason that the sage, illuminating this situation with the way things really are rather than going along with discriminations, is also a case of accommodating what is right and what is ‘this’. But ‘this’ is also ‘that’ and vice versa. And a ‘this’s’ ‘that’ further has one set of right and wrong while ‘this’ has another. In truth, is there really such a thing as ‘this’ and ‘that’ or not?
>
> Where neither ‘this’ nor ‘that’ has an opposite
> Is called the hinge of *dao*,
> And as soon as the hinge is fitted to its bracket

> It can respond endlessly

(*Zhuangzi* 2)

The *wu*-forms - *wuzhi*, *wuwei*, *wuyu* - all provide a way of entertaining and deferring to an object-less world. Thus the sage is concerned with that sort of knowing, acting and desiring which does not depend upon objects. This point is crucial to the Daoist understanding of both self and world. The discriminated, and discriminating, self of the sort recognized in the contemporary West comes into being through encountering a world of things which effectively stand over against and object to him (see Perception). In Daoism, however, personality is forgotten to the extent that discriminated objects no longer constitute the environs of the self.

The consequence of this transformed vision is that knowing, acting and desiring are no longer based upon *construal*. The affirmation of principles as fixed guides for thought and conduct leads us to construe the object of our knowledge by recourse to those principles. An item becomes one of a *kind*, or an instrument for the achievement of an end. Feeling ourselves in tension with objectified others leads us to act in an aggressive or defensive manner to effect our will. Desire motivated by an object leads us to seek possession of that which is desired, allowing it significance only insofar as it meets our need. An individual intoxicated by objects narrows, truncates, and obfuscates the world as it is. Such a person fails to recognize the bottomlessness of each particular phenomenon, leaving off an appreciation of it at the point that the phenomenon ceases to resonate with one's own expectations of it. On the other hand, unprincipled knowing, nonassertive action and objectless desire have this in common: to the extent they are successful, *they leave the world as it is*.

8 The transformation of things: *wuhua*

As we have said, the Daoist self is above all deferential. Deferring to the ways of things as they truly present themselves involves the expression of mirroring responses to environing circumstances. Mirroring such a world in fact involves the mirroring of an indefinite number of worlds since each item is the particular focus of a world. In his commentary on the *Daodejing* 5, Wang Bi notes that 'the myriad things acquire proper order and patterning among themselves'. In ordering themselves, they thereby order a world. Mirroring the individual things in their distinctive particularities therefore involves mirroring a complex set of overlapping orders, each with its own distinctive centre.

Obviously, the construing individual could not but treat such a world as chaotic in the extreme. The deferential self, on the other hand, sees these overlapping orders as a richly spontaneous array of patternings. The famous account of Lord Hundun - Lord Chaos - which closes the Inner Chapters of the *Zhuangzi* is appropriate here:

> The ruler of the North Sea was 'Swift', the ruler of the South Sea was 'Sudden', and the ruler of the Central Sea was 'Chaos'. Swift and Sudden had on several occasions encountered each other in the territory of Chaos, and Chaos had treated them with great hospitality. Swift and Sudden, devising a way to repay Chaos' generosity, said: 'Human beings all have seven orifices through which they see, hear, eat, and breathe. Chaos alone is without them.' They then attempted to bore holes in Chaos, each day boring one hole. On the seventh day, Chaos died.

(*Zhuangzi* 7)

The construing self, relying upon its own seven orifices alone as its strategy for shaping the world, closes down the spontaneity of the unordered totality. But why shouldn't one wish to bring order out of Chaos? A reasonable question indeed, if chaos is the confusion and disarray which Western mythology describes. But if 'chaos' is a noncoherent sum of all orders - perhaps better rendered positively as 'spontaneity' - then imposing order on it means simply selecting one of a myriad candidates for order and privileging that one over the rest.

Obviously something more needs to be said concerning how one might go about actually mirroring the parity of all things. After all, the practical effect of giving each thing its due, when that means allowing it to focus a world from its own peculiar perspective, would seem to be disastrous. We can, however, follow the 'logic' of the Daoist sage at least this far: The most profound experience of the sage is not that of the Oneness *of* all things, or of Oneness *with* all things; it is closer to the experience of the particularity of all things. The experience of unity relevant to the Daoist is of 'this one' and 'that one'. Daoists do not 'become one with all things'; rather, as deferential, they approach becoming one with this or that thing. In effect, the sage becomes one with *all* other things only through

celebrating continuity with other things based upon an intuition of parity.

When the *Zhuangzi* recommends that we become 'one with all things', this is not a Vedānta-like call to surrender one's particularity and dissolve into a unitary and perfect whole. Rather, it is a recognition that each and every unique phenomenon is continuous with every other phenomenon within one's field of experience. Is this an exhaustive claim: are we talking about *all* phenomena? Because the world is processional and because its creativity is *ab initio* rather than *ex nihilo* - a creativity expressed across the careers of its constitutive phenomena - any answer to this question would have to be provisional. Phenomena are never either atomistically discrete nor complete. The *Zhuangzi* recounts:

> With the ancients, understanding had gotten somewhere. Where was that? Its height, its extreme, that to which no more could be added was this: some of them thought that there had never begun to be things. The next lot thought that there are things, but that there had never begun to have boundaries among them. The next lot thought that there are boundaries among things, but that there had never begun to be the acceptable and unacceptable among them…
>
> (*Zhuangzi* 2)

But if the world is not a single, unitary cosmos, with rules and standards and laws finally determining its order - if the world, on the contrary, is the noncoherent totality of all possible orders - how does the deferential sage survive deferring to chaos? Ostensibly, the sage must discriminate as much as we ordinary folk. The difference that makes all the difference is that the sage recognizes the arbitrariness, transitoriness and merely conventional status of such discriminations.

Only 'thises' and 'thats' exist, by virtue of their distinctive *de*. When discriminations are made, as perforce they must be, one treats or 'deems' things in the mode of a transitory discrimination. This involves the utterance of a 'that's it!' which deems something to be the case. The sage discriminates as is deemed necessary, always acting in a deferential mode. However, such discriminations are qualified in two important ways. First, the sage recognizes that such discriminations are conventional and transitory. Secondly, to the extent possible, even these discriminated items are engaged in the modes of unprincipled knowing, nonassertive action and objectless desire.

The Daoist posture of mirroring the world cannot be either dialectical or analytic, since both analysis and dialectic require a putative whole: the former in order to divide into parts, the latter in order to form the opposing parts into some synthetic whole. Neither the 'ten thousand things' nor the self which plies its way among them may be summed to a coherent whole. The Daoist way is *analogical*; that is to say, it involves the correlation of elements with presumed similarities of structure, character or function without the necessity of assuming a holistic context. Such a method emphasizes the *ad hoc* nature of any sorts of discrimination or organization.

The employment of correlativity without the necessity of positing a cosmic whole as background of one's ruminations is a defining characteristic of philosophical Daoism. From that insight follows the irrelevance of dialectical and analytic modes of discourse; and from the irrelevance of the dialectical and analytic modes of discourse follows the implausibility of precise concepts, or of univocal language. From the irrelevance of conceptual language derives the impossibility of having objects or explanations of those objects with any final integrity or comprehensiveness. This means that neither the world nor the self can constitute a coherent unity.

9 The transformation of things: *wuhua* (cont.)

We have seen thus far that the Daoist engages the world in a mirroring manner which accepts order to be a function of the focal activity of each item in the totality of things. In such a world, 'order' is the same as 'chaos'; that is, it is an indefinite set of overlapping orders. While the sage recognizes the hierarchical relationships which obtain among things, the *de* of the individual things must be deferred to because each thing also has parity with the others as an inimical and unsubstitutable defining condition.

We have tried to understand the notions of person and world in classical Daoism by articulating some of the consequences of the sage's act of mirroring which allows things to be seen in their original order(s). One way of doing this is to ask, as we have just done, 'what sort of being is it who engages a world constituted by a complex set of overlapping orders shaped by the individual perspectives or particular foci (*de*), of each of 'the ten thousand things'?' Responding to this question has led us to the notion of the deferential self characterized by the modes of

engagement identified as *wuzhi*, *wuwei* and *wuyu*. Another fruitful route toward understanding the Daoist version of the deferential self is by noting that the employment of the *wu*-forms as a means of mirroring the world as it is, has direct consequences for establishing the 'mood', 'temperament', or 'humour' with which the Daoist engages the ten thousand things.

The Daoist, by virtue of deferential activity, is characterized by a 'light-minded' or 'light-hearted' mood. As a popular translation of one of the chapter titles of the *Zhuangzi* has it, this mood is one of 'free and easy rambling'. Such a mood is often contrasted with the more serious-minded Confucian, and is even more dramatically contrasted with the 'high seriousness' of the Anglo-European culture. The dominant sensibility of Western philosophy is imbued with a commitment to a single-ordered world, hierarchically arranged with human beings near the top of the ladder ('a little less than the Angels'). The high-seriousness of the Western individual is an appropriate response to the fixed and permanent status of his world. Daoists, by contrast, respond to a more horizontal world of things possessed of an indefinite number of orders. Their's is thus a more flexible and light-hearted response. We propose to pursue this contrast between the 'serious' and the 'light-hearted' as means of highlighting the distinctive character of the Daoist world.

The first thing to be said has to do with the rather standard perception of Daoism as a relativistic mode of understanding. In the West, relativism is usually an epistemological doctrine: that is to say, 'relativity' is a theoretical notion. The relativities recognized are not thought to be ontologically constitutive of the way of things, and are seldom thought to involve practical consequences. Relativists do not normally wish to claim that the world itself actually exists in a number of different ways, but only that we have a number of different theories of the way of things, and that there is no satisfactory means of deciding which gets at the way things really are. Secondly, a consistent relativist must claim that nothing practically follows from relativism but inaction. If one tacitly or explicitly commits to one of a number of ways of acting, relativism has been practically abandoned. Thus, the issue of relativism in Western philosophy is, mainly, a red herring (see Relativism).

Philosophical Daoism avoids this sort of relativism by asserting a ground for parity by virtue of the continuity among things, associated with Zhuangzi's 'transformation of things' (*wuhua*). *Dao* is the total process of becoming which constitutes the ways of things. These things form 'worlds' characterizable from the perspective of each and every item in the total process. Given the uniqueness of each thing as a starting point, particulars must with reference to any given issue stand in hierarchical relationship. This is the ground of deference. All things defer, in this way or that, to everything else. This is why a correlative and hierarchical *yin-yang* vocabulary usually works to articulate relationships. We say usually because, given the porous nature and attendant vagueness of particular things and events, there are occasions on which the *yin-yang* language is not functional. As is stated in the *Yijing (Book of Changes)*, 'What *yin-yang* does not fathom is called inscrutability'.

The transformation of all things entails two important consequences. First, there are always myriad alternative postures which challenge the ultimacy of one's present configuration. Second, the processional nature of experience guarantees that one will in fact actually proceed through an indefinite number of such configurations. The sagely recognition of this fact is the source of Daoist light-heartedness.

With Daoism, and specifically the Zhuangzian variety, the issue of relativism is usually seen to be irrelevant. Though the assumption that Zhuangzi is a realist would doubtless be challenged by many scholars who would insist that he is an extremely subtle proponent of a kind of sceptical relativism, Zhuangzi's transformation of things seems to commit him to the belief that the world actually exists as a shifting set of ways of being, a myriad set of overlapping worlds, a chaos of 'thises' and 'thats', a multifarious congeries of orders. 'Chaos' (*hundun*), understood as the totality of all orders, names the way of things. Further, Zhuangzi seems to believe that there are crucial and direct practical consequences of such ontological pluralism. The deferential self, expressed through the modalities of the *wu*-forms, spontaneously engages in mirroring responses which take into account the *de* of the items or events he or she encounters. Thus, Zhuangzi's 'relativism' is seriously moderated at the level of practice since deference entails commitment. There are, therefore, direct practical consequences of mirroring the vast indifferent complex of the ways of things in a manner that allows one to see all things as viable candidates for deferential response.

The paradigm for Daoist light-heartedness turns out to be something like this:

> If people sleep in damp places, they ache at the waist and end up half-paralyzed, but is this the case with the loach? If they live in the trees, they tremble with fear, but is this the case with the ape?
>
> (*Zhuangzi* 2)

The paradigm of high seriousness on the other hand is something like:

What is sauce for the goose is sauce for the gander.

Daoist light-heartedness depends upon the coexistence of a plurality of viable world orders. On this view, the world - if by this we mean that *the* World is one, single-ordered, rational, coherent, internally consistent cosmos - is singularly humourless. Even if the totality is acknowledged to possess many orders, as long as there is any hint of a logical or axiological priority given to one or some orders *vis-à-vis* an other or other orders, the ground of any real light-mindedness or light-heartedness is undermined. This is the case because the notion of a privileged order implies a non-deferential, determinative stance. It is such a notion of superiority which cancels true light-heartedness. However, it is not simply a matter of taking up a variety of sympathetic stances. For Zhuangzi, one actually becomes all of the potential objects and beings with respect to which such stances may be taken:

> Before long, Master Lai fell ill. Wheezing and panting, he was on the brink of death. His wife and children gathered about him and wept. Master Li, having gone to enquire after him, scolded them, saying: 'Get away! Don't impede his transformations!'
>
> Leaning against the door, Master Li talked with him, saying: 'Extraordinary, these transformations! What are you going to be made into next? Where are you going to be sent? Will you be made into a rat's liver? Or will you be made into an insect's arm?…
>
> Now if a great ironsmith were in the process of casting metal, and the metal leapt about saying: 'I must be forged into a Mo Ye sword!' the great ironsmith would certainly consider it to be an inauspicious bit of metal. Now, if once having been cast in the human form, I were to whine: 'Make me into a human being! Make me into a human being!' the transformer of things would certainly take me to be an inauspicious person. Now once we take the heavens and earth to be a giant forge and transformation to be the great ironsmith, where ever I go is just fine. Relaxed I nod off and happily I awake.
>
> (*Zhuangzi* 6)

The transformative processes (*wuhua*) are not to be interfered with, but are to be met with deference. To this extent, the sage is in agreement with Faust who proclaims to Mephistopheles:

If I ever say to the moment, 'Hold! Thou art so fair!' then thou canst require my soul of me.

What distinguishes the life of the Daoist sage from that of the pleasure-seeking Faust lies in the difference between aggressive desiring and *wuyu*. For Faust each moment brings a new object of desire to be seduced, consumed or otherwise enjoyed, and then abandoned.

The Daoist neither demands constant transformation for the sake of the ever-new, nor objects to transformation, attempting to apply the brakes and hold onto the moment. The deferential self yields to the moment, without constraint.

Daoists such as Zhuangzi are light-hearted by virtue of being freed from the serious responsibilities entailed by 'being right', or 'having the truth'. Were one to seek a coherent theory to house Zhuangzi's teachings, he would risk constructing a platform from which he may proclaim (implicitly) the superiority of those teachings. And if we presume that Daoist light-heartedness is used in the defence of a vision of things which necessarily denies the validity of other visions, we have simply made any attempts at light-heartedness secondary to the 'true philosophical enterprise', namely, the defence of a particular way of envisaging things.

The character of light-heartedness lies in a felt incongruity or contradiction. The Daoist treats these incongruities as existing among things and events which are on a par with the others. The pluralist vision of Daoism plays upon the sense of the contradiction between one perspective and a myriad others: between Zhuangzi and a butterfly, among fishes, birds, cicadas and turtle doves, between men and monkeys. Light-heartedness is an essential means of teasing one into the recognition of the variety of perspectives permitted by the totality of existing things. The

need for allusiveness lies in the fact that one cannot fully appreciate the differences without some sense of the reality of those differences. However, one cannot confront these differences in too direct a manner; they may at best be hinted at.

We should note here that though the term ‘irony’ has often been used to name the Daoist form of light-heartedness, it is hardly an altogether appropriate one. There can be no real philosophic irony without something like the dialectic of tragedy and comedy as it developed in ancient Greece, and Zhuangzi’s doctrine of the transformation of things precludes the development of a tragic sense. The tragic sense is predicated upon loss. But the transformation of things guarantees that there is no final loss: there are only processes of transformation which realizes the becoming of what has not yet become.

See also: Aesthetics, Chinese; Confucian philosophy, Chinese; Legalist philosophy, Chinese; Neo-Confucian philosophy; Chinese philosophy; Dao; Daodejing; De; Guanzi; Huainanzi; Qi; Yijing; Yin-yang; You-wu; Zhi; Zhuangzi

DAVID L. HALL
ROGER T. AMES

References and further readings

Ames, R.T. (1989) ‘Putting the *te* Back in Taoism’, in J.B. Callicott and R.T. Ames (eds) *Nature in Asian Traditions of Thought*, Albany, NY: State University of New York Press.(An argument for an ‘aesthetic’ interpretation of Daoism.)

Chang Chung-yuan (1963) *Creativity and Taoism*, New York: Harper & Row.(A discussion of the influence of Daoism on Chinese culture and aesthetics.)

Daodejing, (*c*.350-250 BC), trans. D.C. Lau, *Tao Te Ching*, Chinese Classics: Chinese-English Series, Hong Kong: The Chinese University Press, 1982.(A reliable translation of the text revised on the basis of the Mawangdui archaeological discoveries.)

Graham, A.C. (1989) *Disputers of the Tao*, La Salle, IL: Open Court.(A contextualization of Daoist thinking within classical Chinese philosophy.)

Hall, D.L. (1978) ‘Process and Anarchy - A Taoist Vision of Creativity’, *Philosophy East and West* 28 (3): 271-85.(A new vocabulary for interpreting Daoism.)

Hall, D.L. (1982) *The Uncertain Phoenix*, New York: Fordham University Press.(Daoism as a philosophical resource for contemporary philosophical issues, especially technology.)

Hall, D.L. and Ames, R.T. (1995) *Anticipating China: Thinking Through the Narratives of Chinese and Western Culture*, Albany, NY: State University of New York Press.(A narrative understanding of the difference between Chinese and Western philosophical sensibilities.)

Hall, D.L. and Ames, R.T. (1997) *Thinking from the Han: Self, Truth, and Transcendence in China and the West*, Albany, NY: State University of New York Press.(An exploration of several central philosophical issues that illustrate the distance between Chinese and Western assumptions.)

LaFargue, M. (1992) *The Tao of the Tao Te Ching*, Albany, NY: State University of New York Press.(A translation and commentary on how to read the *Daodejing*.)

Needham, J. (1956) *Science and Civilisation in China*, vol. 2, Cambridge: Cambridge University Press.(A well-known statement of the role of Daoism in classical Chinese thought.)

Waley, A. (1934) *The Way and Its Power: A Study of the Tao Te Ching and its Place in Chinese Thought*, London: George Allen & Unwin.(A pioneering translation of the *Daodejing* with a discussion of many of the central concepts.)

Wu Kuang-ming (1990) *The Butterfly as Companion: Meditations on the First Three Chapters of the Chuang Tzu*, Albany, NY: State University of New York Press.(A Zhuangzian commentary on several core chapters of *Zhuangzi*.)

Zhuangzi (*c*. 4th century BC), trans. B. Watson, *The Complete Works of Chuang Tzu*, New York: Columbia University Press, 1968; trans. A.C. Graham, *Chuang-tzu: The Inner Chapters*, London: George Allen & Unwin, 1981.(Watson is an elegant, literary translation of the *Zhuangzi* influenced by Japanese commentators; Graham is a philosophically sensitive translation of about 85 per cent of the *Zhuangzi*.)

Darwin, Charles Robert (1809-82)

Darwin's On the Origin of Species (1859) popularized the theory that all living things have evolved by natural processes from pre-existing forms. This displaced the traditional belief that species were designed by a wise and benevolent God. Darwin showed how many biological phenomena could be explained on the assumption that related species are descended from a common ancestor. Furthermore, he proposed a radical mechanism to explain how the transformations came about, namely, natural selection. This harsh and apparently purposeless mechanism was seen as a major threat to the claim that the universe has a transcendent goal.

Because Darwin openly extended his evolutionism to include the human race, it was necessary to re-examine the foundations of psychology, ethics and social theory. Moral values might be merely the rationalization of instinctive behaviour patterns. Since the process which produced these patterns was driven by struggle, it could be argued that society must inevitably reflect the harshness of nature ('social Darwinism'). Darwin's book has been seen as the trigger for a 'scientific revolution'. It took many decades for both science and Western culture to assimilate the more radical aspects of Darwin's theory. But since the mid-twentieth century Darwin's selection mechanism has become the basis for a highly successful theory of evolution, the human consequences of which are still being debated.

1 Scientific work

Darwin was born in 1809 in Shrewsbury, England. After a brief spell as a medical student he moved to Cambridge, where he received extra-curricular training in natural history and geology. Travelling on the survey ship HMS *Beagle* (1831-6), he studied the geology of South America and accepted what he took to be Charles Lyell's uniformitarian theory of geological change (see Geology, philosophy of §1). He proposed a new theory to explain the formation of coral reefs, and collected fossils and studied the geographical distribution of South American animals. In particular he was struck by the existence of distinct but related species of birds on the individual islands of the Galapagos group, off the west coast of South America.

Soon after his return to England, he became convinced that the Galapagos species could only be explained as the modified descendants of migrants from the mainland. Evolution theories were highly controversial at the time, but Darwin was soon a complete convert. He studied animal breeding and was led to the theory of natural selection. Having been active in the scientific life of London, in 1842 he moved to Down House in Kent, where he spent the rest of his life suffering from a debilitating illness. He kept up his studies of biogeography and produced a major taxonomic survey of barnacles. This work was intended to throw light on his theory, which remained unpublished until a similar idea was proposed by A.R. Wallace in 1858. *On the Origin of Species* was published at the end of 1859; following an intense controversy, the basic idea of evolution was accepted by the majority of scientists.

In *The Descent of Man* (1871) Darwin extended his theory to the human race. Much of his later scientific work centred on the application of his theory in botany, to explain the activity of climbing and insectivorous plants, and the relationship between flower structure and insect pollination. He also wrote a study of earthworms.

2 The evolutionary worldview

Conventional histories of the Darwinian revolution have assumed that, prior to the publication of the *Origin of Species*, everyone accepted that each species was designed by a wise and benevolent God. Darwin's materialistic theory thus upset a vision of a perfectly ordered universe. In fact, the idea of evolution had been widely discussed in the previous decades. Robert Chambers' *Vestiges of the Natural History of Creation* (1844), promoted the image of life progressing towards higher levels of physical and mental organization under the control of divinely implanted laws. Historians still debate the extent to which Darwin himself saw evolution as necessarily progressive. There was certainly an element of progressionism in his thinking, but his theory was built on different foundations and in later life he was unable to accept that the whole process was designed by the Creator. Darwin may have begun his work believing that the laws governing natural development were instituted by God, but he ended up seeing the universe as a scene of perpetual struggle and suffering.

Darwin argued that each population changes by adapting to its local environment. When members of a population migrate to new areas, they will adapt to the new conditions and thus come to differ from the parent group. If the

degree of difference becomes sufficient, the two populations would find it difficult to interbreed even if they came into contact, and would thus count as distinct species. Local varieties or subspecies are the first steps on the way to full speciation. The overall pattern of evolution must thus be visualized as an irregularly branching tree, not as the ascent of a ladder towards a fixed goal.

Darwin believed that evolution was an immensely slow process, and thus that it was unrealistic to expect ever to observe one species being transmuted into another. Opponents of the theory argued that we only observe minor changes within species, from which it is illegitimate to infer that change spead over a longer timespan will generate a new species distinct from the parent form. Darwin realized that the case for evolution would have to be argued by showing that the theory accounted for a range of otherwise inexplicable phenomena. He showed that the classification of species could be understood as a reflection of the tree-like model of genealogical relationships (see Taxonomy §§3-4). Each major taxon was founded by a single ancestral form which had developed some important evolutionary innovation; the descendants of this ancestor radiated out through the local adaptation of sub-populations, thus creating the genera and species we recognize today. Homologies such as the similar bone structures of the forelimb in humans, horses and bats were explained as adaptive modifications of similar structures inherited from a common ancestor. Relationships between dissimilar adult forms could often be traced in their early embryos, which were not subject to the same degree of modification. Darwin knew that evolutionary change could not be demonstrated from the fossil record, which often gave the appearance of sudden ’leaps’. He argued that the record is very imperfect, the ’leaps’ are in fact gaps concealing gradual evolutionary transformations. Once this factor was taken into account, the fossil record could be reconciled with the tree model. The geographical distribution of organisms could also be explained in terms of species migrating outward from their point of origin, adapting themselves to the new territories they entered (see Evolution, theory of).

3 Natural selection

The basic idea of evolution was accepted quite rapidly in the 1860s, although many people, scientists included, were unwilling to admit that the development of life on earth had no preordained goal. This meant that they were particularly worried about Darwin’s proposed mechanism of natural selection. In this theory, the species could not be regarded as having an underlying ’type’ on which all the individual organisms were modelled. The species was just the breeding population, and if the average structure of the individuals making up the population changed, then so did the species. The natural world could not be composed of a range of fixed, specific types or forms (see Species §2).

There was much debate over whether natural selection could be considered a *vera causa* or true cause capable of producing significant changes in species. Darwin knew that each component of the effect could be demonstrated individually, but doubted that selection could produce a new population incapable of interbreeding with the parental type within an observable timescale. For him, selection was a plausible mechanism of change, and probably the most important one - but he never ruled out the possibility of other mechanisms. Much of the initial opposition to the selection theory was based on an unwillingness to believe that the species does not have some sort of essence which prevents individuals becoming modified beyond a fixed range.

Darwin’s studies of domesticated animals convinced him that the individuals making up any population differ significantly among themselves. This is sometimes called ’random’ variation, because there seems to be no apparent purpose to the differences. Darwin knew that there was an underlying mechanism to explain the production of individual differences and their transmission through heredity, although his own explanation did not stand the test of time. Many of Darwin’s contemporaries found it difficult to accept that the variations which accumulate to give evolution could be produced by a process that imposed no fixed direction. They preferred to believe that the variations were directed along a predetermined channel by forces arising from within the organism, or were produced in response to the organism’s own purposeful activities. If characteristics acquired in response to new behaviour-patterns were inherited, this would give the rival evolutionary mechanism known as ‘Lamarckism’. Darwin himself accepted a subsidiary Lamarckian effect, but stressed the natural selection of random variation. Modern genetics has shown that acquired characteristics cannot be inherited and that mutations produce many different characteristics within a population. In the modern Darwinian theory, selection works by changing the frequencies of genes within the population (see Genetics).

Natural selection depends upon the elimination of those variant characteristics which are maladaptive, and the

superior reproductive fitness of those which confer some adaptive benefit. Following Malthus, Darwin believed that the tendency for a population to breed beyond the capacity of its food supply would lead to a 'struggle for existence' in which the least fit would be eliminated. This was what H. Spencer called the 'survival of the fittest'. Contrary to a popular anti-Darwinian argument, natural selection is not based on a tautology (the survival of those who survive), because fitness is defined in terms of the adaptation which enables an organism to live and breed more effectively. Darwin did, however, argue that there is a different mechanism, sexual selection, capable of enhancing those characteristics which, even if maladaptive, improve the organism's chances of reproducing. Thus the peacock has developed its large tail because this characteristic has become involved with the birds' mating behaviour - a bigger tail attracts peahens more effectively, and the reproductive advantage this confers outweighs the disadvantages of a large tail when escaping from predators.

4 Human origins

From the start, Darwin accepted that the human species would have to be included within the evolutionary worldview. If humans were to be treated as merely advanced animals, the mental and moral faculties traditionally seen as products of the soul would have to be explained as extensions of those faculties already possessed by animals. In his *Descent of Man* (1871) and *Expression of the Emotions in Man and the Animals* (1872), Darwin tried to justify this proposition and to explain how the human mind had developed so much further than those of our closest animal relatives. He did not believe that the mind was a *tabula rasa* or blank slate capable of indefinite modification by experience. Many mental activities are governed by instincts imprinted by evolution. He showed that many aspects of human behaviour, especially the way we express emotions, are relics of our animal ancestry. He also took seriously many stories that interpreted animal behaviour in anthropomorphic terms, thus providing apparent evidence that animals possess rudiments of even the highest mental and moral faculties.

Darwin accepted that the mental faculties were produced by the brain, and that the increased brain size of the 'higher' animals is reponsible for their increased intelligence and more complex behaviour. Many of his contemporaries assumed that evolution would inevitably produce a steady increase in the level of animal intelligence. In *The Descent of Man*, however, Darwin anticipated the modern position in which (since evolution is not inherently progressive) it is necessary to explain why the branch leading to humans has experienced a much greater expansion in intelligence than that leading to our closest cousins, the great apes. He argued that the distant ancestors of humans had stood upright as a means of walking on the open plains, while the apes had stayed in the trees. Human intellengence was a by-product of this change of habitat, produced as a means of exploiting the hand's ability to manipulate the environment.

Darwin argued that the higher animals exhibit social behaviour which provides the foundation for human moral values. Social behaviour is governed by instincts, and natural selection can act upon the variations in such instincts to promote useful behaviour patterns. Since our distant ancestors lived in social groups, we have inherited their social instincts, and our moral values are rationalizations of behavioural tendencies which we feel automatically. Modern efforts to explain some aspects of human behaviour in terms of biologically implanted instincts make use of the same argument (see Sociobiology).

Although Darwin anticipated some modern ideas on human evolution, he shared the prejudices of his own time and was convinced that Europeans were superior to other races and that modern industrial civilization is the highest expression of social evolution. His ideas were taken up by some social evolutionists who argued that here, as in the biological realm, progress would only result if individuals or races were left to struggle among themselves to determine which was fit enough to survive. This social philosophy came to be known as 'social Darwinism' - although one of its leading proponents, Herbert Spencer, was more a Lamarckian than a Darwinian. The claim that struggle was a spur to progress was developed in many other forms besides Darwin's theory of natural selection.

See also: Al-Afghani §2; Evolution and ethics; Evolutionary theory in social sciences; Huxley, T.H.

PETER J. BOWLER

List of works

Darwin, C. (1845) *Journal of Researches into the Geology and Natural History of the Countries Visited during the Voyage of H.M.S. Beagle*, London: John Murray.(A semi-popular account of the voyage, frequently reprinted.)

Darwin, C. (1859) *On the Origin of Species by means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, London: John Murray. (Darwin’s main work on evolution; the much-revised sixth edition of 1872 is the one most frequently reprinted.)

Darwin, C. (1871) *The Descent of Man and Selection in Relation to Sex*, London: John Murray, 2 vols; repr. Princeton, NJ: Princeton University Press, 1981.(Darwin’s main work on human origins, with a long account of sexual selection and an attempt to apply that theory to explain human characteristics.)

References and further reading

Bowler, P. (1984) *Evolution: The History of an Idea*, Berkeley, CA: University of California Press; revised edn 1989.(General survey of the history of evolutionism, with extensive bibliography.)

Bowler, P (1988) *The Non-Darwinian Revolution: Reinterpreting a Historical Myth*, Baltimore, MD: Johns Hopkins University Press.(A reassessment of Darwin’s impact on nineteenth-century thought.)

Bowler, P (1990) *Charles Darwin: The Man and his Influence*, Oxford: Blackwell. (Biography and assessment of Darwin’s influence.)

Desmond, A. and Moore, J. (1991) *Darwin*, London: Micheal Joseph.(Massive biography stressing the social context of Darwin’s thought.)

Ghiselin, M. (1969) *The Triumph of the Darwinian Method*, Berkeley, CA: University of California Press. (Detailed account of Darwin’s scientific work.)

Kohn, D. (ed.) (1985) *The Darwinian Heritage*, Princeton, NJ: Princeton University Press.(Major compilation of Darwin scholarship.)

Richards, R. (1987) *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*, Chicago, IL: University of Chicago Press.(The evolution of mental powers according to Darwin and later writers.)

Young, R. (1985) *Darwin’s Metaphor: Nature’s Place in Victorian Culture*, Cambridge: Cambridge University Press.(Essays on the social context of Darwinism.)

David of Dinant (*fl. c.*1210)

A twelfth- and early thirteenth-century philosopher who may have taught at Paris, David of Dinant was noted for a heretical, pantheistic view that identified God, mind and matter. None of his works survive intact, and we know of them primarily through the works of other authors. His major work, the Quaternuli, was condemned at Paris in 1210. His heretical views were influential enough to receive critical attention in the works of Albert the Great and Thomas Aquinas.

Very little is known for certain about David of Dinant's life. He was born, probably at Dinant, Belgium, but possibly in Dinan in Brittany, in the middle of the twelfth century. He may have taught at the University of Paris. A work of his identified as *Quaternuli*, perhaps the first word of the work's text, was condemned in 1210: all copies of it were to be brought to the Bishop of Paris and burned, and anyone discovered possessing the work after Christmas of that year was to be regarded as a heretic. This work may be identical to one attributed to David with the title *De Tomis hoc est de Divisionibus (On the Volume Known as 'On Divisions')*. David died most likely in the second decade of the thirteenth century, perhaps while a fugitive from France.

Some fragments of his writings appear to have survived, but his most distinctive views are to be gleaned from the accounts of his philosophical opponents. Chief among these are Albert the Great and Thomas Aquinas. Although Nicholas of Cusa also refers to David, he may have acquired his knowledge from reading Albert or Aquinas.

According to Aquinas, David divided reality into three parts, bodies, souls and eternal, separated substances. Bodies are composed of *hyle* or prime matter, souls are constituted from *nous* or mind, and eternal, separated substances are constituted from God. Albert and Aquinas both attribute to David the additional thesis that prime matter, mind and God are one and the same. It follows that God, mind and prime matter constitute all things, or, as Aquinas puts it, that everything is one essence. David's views thus contradict several doctrines of philosophical theology, defended particularly by scholastic philosophers; these include the doctrines that God is an immaterial being, that God is perfectly simple, having no parts or plurality of any kind, that the created world is separate from God, that God created the world out of nothing, and that human souls are not divine (see God, concepts of).

Aquinas claims that David's identification of God and prime matter was based on the argument that, if they were not identical, God would have to differ from prime matter by means of some differentiae. But if they possessed differentiae, God and prime matter would not be metaphysically simple, contrary to the assertion that they are simple. Aquinas, and apparently David, presuppose a theory of predication deriving from Aristotle by way of Porphyry's *Isagōgē* (see Porphyry). On that theory, questions concerning what natural kinds have in common and questions concerning how natural kinds differ from one another are to be answered by specifying the genera and differentiae of the respective natural kinds. Horses and humans, for example, are the same in genus insofar as they are both sensitive, animate corporeal substances, but different in species because horses possess the differentiae of being ungulate while lacking the differentiae of being bipedal and rational, and humans possess the differentiae of being bipedal and rational while not being ungulate. As reported by Aquinas, David's argument begins with the premise that God and prime matter are metaphysically simple, proceeds by inferring that they can thus have no differentiae (for the possession of any differentiae entails that its possessor is metaphysically complex), and concludes that because there is no specific difference between them, God and prime matter are identical.

Aquinas claims that David's argument is fallacious; it confuses *difference* with *diversity*. It is true that two distinct kinds of thing in the same genus must differ by means of some differentiae, but the same does not hold for kinds that share no genus in common. Although horses share something in common with humans, horses and humans share nothing in common with, for example, cerulean blue. There are no differentiae that distinguish colours from animate substances, but it does not follow that colours and animals are identical; they are simply diverse. Similarly, Aquinas argues, God and prime matter are diverse in themselves. In Aristotelian terms, prime matter is pure potentiality, having no actuality of its own, and God is pure actuality, a being having no unactualized potentiality.

See also: Aristotelianism, medieval; Matter; Simplicity, divine

WILLIAM E. MANN

List of works

David of Dinant (before 1210) *Quaternuli*.(This work may also be identical with work attributed to David known as *De Tomis hoc est de Divisionibus*. There are no printed versions of David's work, and only fragments of his writings survive. Knowledge of his work is to be obtained primarily from the work of Albert the Great and Thomas Aquinas.)

References and further reading

Denifle, H. and Chatelaine, E. (eds) (1889) *Chartularium Universitatis Parisiensis (Charters of the University of Paris)*, Paris: Delalain, vol. 1. (Contains the text of the decree of 1210, condemning the *Quaternuli*.)

Albert the Great (1270-80) *Summa theologiae sive de mirabili scientia Dei*, book 1, part 1, questions 1-50A; published in *Opera Omnia*, vol. 34, part 1, Aschendorff: Westphalia Monastery, 1978.(David of Dinant is discussed briefly in treatise 6, question 29, chapter 1, article 2, 'Whether the divine unity admits any plurality'.)

Aquinas, Thomas (1252-73) *Opera omnia (Collected Works)*, ed. R. Busa, Stuttgart-Bad Canstatt: Friedrich Fromman Verlag Günther Holzboog, 1980, 7 vols.(Volume 1 contains the Latin text of Aquinas' commentary on the *Sentences* of Peter Lombard. Aquinas mentions David's threefold division and the thesis that everything is one in essence in book 2, distinction 17, question 1, article 1, 'Whether the human soul is of the divine essence'. Volume 2 contains the Latin texts of *Summa contra gentiles* and *Summa theologiae*.)

Aquinas, Thomas (1259-65) *Summa contra gentiles (Synopsis [of Christina Doctrine] Directed Against Unbelievers)*, trans. A.C. Pegis, *On the Truth of the Catholic Faith*, Garden City, NY: Image Books, 1955-7, 5 vols.(Aquinas offers the argument presented above against David's identification of God and prime matter in book 1, chapter 17, 'That there is no matter in God'.)

Aquinas, Thomas (1266-73) *Summa theologiae*, published as *Summa Theologica*, Westminster, MD: Christian Classics, 1981, 5 vols.(A reprint of the translation of the Fathers of the English Dominican Province. Volume 1 contains the first part of the *Summa theologiae*. Aquinas says that David 'most stupidly' identified God and prime matter (part 1, question 3, article 8, 'Whether God enters into the composition of other things').)

Davidson, Donald (1917-)

Donald Davidson's views about the relationship between our conceptions of ourselves as people and as complex physical objects have had significant impact on contemporary discussions of such topics as intention, action, causal explanation and weakness of the will. His collection of essays, Actions and Events (1980), contains many seminal contributions in these areas. But perhaps even greater has been the influence of Davidson's philosophy of language, as reflected especially in Inquires into Truth and Interpretation (1984). Among the philosophical issues connected to language on which Davidson has been influential are the nature of truth, the semantic paradoxes, first person authority, indexicals, modality, reference, quotation, metaphor, indeterminacy, convention, realism and the publicity of language.

1 Reasons and causes

Since 1963, with the publication of 'Actions, Reasons, and Causes', Donald Davidson's philosophical work has taken centre stage in analytic philosophy. He begins this essay with the question: what is the relation between a reason and an action when that reason explains the action by giving the agent's reasons for performing it? Prior to this essay, something like a consensus had formed in philosophy that this relationship was not causal. Discussion tended to centre around why it could not be causal and the consequences of its not being so.

Davidson's central purpose was 'to defend the ancient - and commonsense - position that rationalization is a species of causal explanation' (Davidson 1980: 3). A large portion of the essay is devoted to refuting various arguments, quite popular then, that purported to show that reasons could not be causes of the actions they rationalize. The arguments are too many and varied to be properly treated here, but in passing I note that they were largely inspired by remarks made by Wittgenstein and/or certain interpretations of Humean strictures on causation. Each argument proposes that a necessary condition for two items to interact causally is not satisfied by reasons and actions. Davidson replies to the leading arguments by showing in each case either that reasons and actions indeed satisfy the necessary condition in question or that the allegedly necessary condition for causal interaction is not necessary at all (see Action).

2 Events

Davidson's chief claim about events is that they are concrete particulars - that is, unrepeatable entities with location in space and time. He regards the mind-body problem as the problem of the relation between mental and physical events. He treats causation as a relation between events and he takes action to be a species of event, so that events comprise the very subject matter of action theory. He argues for the existence of events and for specific claims as to their nature.

In 'Causal Relations' (1980) Davidson argues that the most plausible interpretation of singular causal statements like 'The short circuit caused the fire' treats them as two-place predicate statements with their singular terms (in this case, 'the short circuit' and 'the fire') designating events. In 'The Individuation of Events' (1980), Davidson argues that an adequate theory of action must recognize that we can describe the same action differently. This seems crucial in making sense of perfectly natural claims like 'Jones managed to apologize in saying "I apologize"', wherein apparently one event is described both as an apology and as an utterance (see Events).

3 The mind-body problem

Consider claims (1)-(3):

(1) No mental event is a physical event.
(2) Some mental events cause physical events.
(3) The only causes of physical events are physical events.

Although much can be said in favour of each, together they clearly are inconsistent. The dilemma posed by the plausibility of each of (1)-(3) and by their apparent incompatibility is the traditional mind-body problem. Davidson's resolution of the dilemma is to abandon (1). Davidson's account accepts the following principles about the gap between the mental (*M*) and the physical (*P*) and about causation:

(4) There are no exceptionless psychological or psychophysical laws and in fact all exceptionless causal laws can be expressed in a purely physical vocabulary.
(5) Event *c* causes event *e* only if there is an exceptionless causal law which subsumes *c* and *e*.

It is commonly held that a property expressed by *M* is reducible to a property expressed by *P* only if there is an exceptionless law that links them. So, it follows from (4) that mental and physical properties are distinct. Thesis (5) says that *c* causes *e* only if there are descriptions *D* of *c* and D' of *e* and an exceptionless causal law *L* such that *L* and '*D* occurred' entail '*D* caused D''. Theses (4) and (5) entail that physical events have only physical causes and that all event causation is physically grounded.

Davidson shows that theses (2)-(5) can all be true if (and only if) mental events are identical to physical events. Say that an event *e* is a physical event just in case *e* satisfies a predicate of our basic physical sciences. These are the predicates appearing in exceptionless laws. Since only physical predicates (or predicates expressing properties reducible to basic physical properties) appear in such laws, it follows that every event that enters into causal relations satisfies a basic physical predicate. So, mental events which enter into causal relations are also physical events. The mental and physical are still distinct in so far as mental and physical events *so-described* are not linked by exceptionless law; but still mental events are physical events (see Anomalous monism).

4 Theory of meaning

Davidson's work in the philosophy of language began later than his seminal work in the philosophy of action. I will discuss primarily the first two essays in his 1984 collection, where he identifies an adequacy criterion for theories of meaning for natural languages, and then applies it critically to a number of then-prominent views about natural language. He also sketches a programme in which a Tarski-style truth theory plays the role of a theory of meaning which meets the criterion he initially articulated.

Like the nineteenth-century mathematician/philosopher Gottlob Frege, Davidson requires that we specify what every sentence means by exhibiting its meaning as a function of the meanings of its significant parts (and their arrangement in the sentence). Call any such theory for a language 'a compositional meaning theory' for that language (see Compositionality). That there must be a compositional meaning theory for natural languages, such as English, seems mandatory because natural languages are spoken by finite speakers without magical abilities, but natural languages themselves have an infinity of meaningful (and nonsynonymous) sentences, each one of which, at least potentially, a speaker understands (at a given time). This seems to require that our knowledge of meanings be based on (a finite number of) rules which determine from a finite set of semantical primitives what count as meaningful compositions, where an expression is semantically primitive if the 'rules which give the meaning for the sentences in which it does not appear do not suffice to determine the meaning of the sentences in which it does appear' (Davidson 1984: 9).

According to Davidson, a compositional theory of meaning for a language *L* is such that anyone who knows it is in a position to understand every sentence of *L*. By specifying the meaning of a sentence *S* in *L*, it is clear that what Davidson has in mind requires that the specification enable anyone who understands the language in which the specification is given, to understand *S*. Davidson's idea is that to provide a compositional meaning theory, we can produce a Tarski-like *truth theory* for *L* (see Tarski, A.). A theory is a truth theory for language *L* if and only if for each sentence *S* of *L*, the theory entails a 'T-sentence' of the form:

(T) *S* is true-in-*L* iff *p*,

where the sentence (T) is in a metalanguage *M* used to talk about *L*. An adequate theory of meaning for German, for example, should issue in theorems like (S):

(S) 'Schnee ist weiss' is true in German iff snow is white.

Davidson's twist on Tarski is that, instead of requiring (for a compositional meaning theory) that for each T-sentence the condition that '*p*' translates *S* is satisfied, Davidson merely requires that each T-sentence be true. While we thus can avoid the danger of circularity attendant on using the notion of 'correct translation' in characterizing what it takes to explain meaning, it may seem that the requirement of merely issuing in *true* T-sentences is too weak. All the 'iff' in a T-sentence requires is that the sentences surrounding it are either both

true or both false, so that the following is true, though not at all helpful about meaning:

(S*) ‘Schnee ist weiss’ is true in German iff grass is green.

Davidson points out, however, that the need for a *finitely stated* adequate theory of meaning to contain recursive apparatus plausibly would lead the theory to issue in (S) rather than (S*). Still, he imposes further constraints on an adequate meaning theory, as we will see in the next section (see Meaning and truth).

5 Radical interpretation

Davidson requires of the meaning-giving T-theory that it be empirically warranted under the practice of radical interpretation. This means that empirical considerations must be respected in choosing between different but true truth theories for, say, German, between a theory that issues in (S) and one that issues in the true but non-interpretive (S*). The favoured T-theory is to be selected on the basis of evidence plausibly available to a radical interpreter (RI), ‘someone who does not already know how to interpret utterances the theory is designed to cover’ (1984: 128). More particularly, an RI does not know the language they are trying to interpret and has no access to bilingual informants, prior dictionaries and the like. An RI can generally tell when an informant ‘holds-true’ a sentence even though they do not know the interpretation of the sentence (1984: 135). So, among the primary data for an RI are, for example:

(E) Kurt belongs to the German speech community, Kurt holds-true ‘Es regnet’ on Saturday at noon, and it is raining near Kurt on Saturday at noon.

Data like (E) are collected from a variety of speakers across a variety of times to confirm or support a generalization like:

(GE) For all speakers *x* in the German speech community, for any time *t*, *x* holds-true ‘Es regnet’ at *t* iff it is raining near *x* at *t*.

Sentences like (GE) provide evidence that the speakers of the community take some form of words to express a certain truth. Davidson, remarkably, does not consider anything else as potential evidence for interpreting another (1984: 135).

What licenses an inference from data like (GE) to the corresponding T-sentences? Davidson’s answer is that a principle of charity is presupposed. According to this principle, the favoured truth theory for a language *L* must entail T-sentences according to which most of the sentences that speakers of *L* hold true are true. Under radical interpretation, sentences held true must usually be true because interpretation is partly constituted by this principle of charity. So, once sentences like (GE) are collected, we can infer corresponding T-sentences via a principle of charity (see Charity, principle of §4; Meaning and understanding §2; Radical translation and radical interpretation §§7-10).

6 Indirect quotation

Indirect quotation attributions have long been known to possess features which frustrate the construction of satisfactory theories of meaning. For example, the inference (1)-(3) is an apparent instance of substitutivity of identity but it is invalid:

(1) Galileo said that the earth moves.
(2) The earth = the third planet from the sun.
(3) So, Galileo said that the third planet from the sun moves.

Davidson’s analysis of indirect discourse is driven by the desire to find a way of incorporating contexts like (1) into a truth theory for a language containing them (1984: 176). His proposal is that we analyse utterances of (1) as consisting of two distinct utterances:

(4) Galileo said that. [The earth moves.]

The first part of (4) is just a relational statement; the second part does not function semantically as part of what is said, but rather is there to serve as the referent of the demonstrative ‘that’ in an utterance of the first part. How

does this get the truth-conditions right? According to Davidson, (4) is true just in case what the speaker has referred to as ‘that’ is an utterance which *samesays* some assertive utterance by Galileo. ‘Samesays’ is treated somewhat casually: it is whatever relation between the utterances is required for things to come out intuitively right; two utterances samesay each other just in case they agree in content, import or purport.

Since paraphrase (4) of (1) contains a demonstrative, it is true or false only relative to a context of utterance. A typical utterance of (1) creates a context relative to which ‘that’ refers to an utterance of the bracketed sentence in (4). And since in most contexts the occurrences of ‘that’ in (1) and (3) refer to different utterances, (1)-(3) is invalid. It is also clear that (1)-(3) is not an instance of substitutivity of identity, since, on Davidson’s analysis, the singular terms ‘the earth’ and ‘the third planet from the sun’ do not even occur in (1) and (3).

Despite important virtues, Davidson’s account has not won anything like general acceptance. The literature is replete with objections: for example, Lepore and Loewer (1989). Some authors complain the account is too strong; others that it is too weak, and others that it fails to generalize, for example, to other propositional attitudinal ascriptions or to *de re* constructions (see Propositional attitude statements §3).

7 Adverbial modification

We began with Davidson on events. The strongest argument he advances for the existence of events and for his conception of their nature derives, surprisingly, from his views in the philosophy of language. Any theory of meaning for a language must embody a view of the relationship between language and reality. Davidson’s conviction is that a theory of meaning, by providing a view about this relationship, offers substantive answers to the various metaphysical questions about reality. In particular, it will require events in order to explain the semantic (logical) form of action, event and causal statements.

Consider the English sentence (1). One obvious candidate for its truth-condition is (2):

(1) John hit Bill.
(2) ‘John hit Bill’ is true if and only if John hit Bill.

In (2) language is both mentioned and used, and in this sense (2) ‘hooks up’ language to reality. This hook-up remains silent on the nature of reality. It simply tells us that the English sentence (2) requires for its truth that John hit Bill. Since an adequate theory of meaning must be finite (1984: 4-15), if we try to construct a theory for, say, English, we are forced to read structure into English sentences (see §4). There does not seem to be another way to generate infinitely many sentences from a finite vocabulary. But now consider sentences like (3)-(5):

(3) John hit Bill at six.
(4) John hit Bill at six in the bedroom.
(5) John hit Bill at six in the bedroom with the stick.

There are no specifiable limits upon the number of kinds of adverbial modifiers which can sensibly attach to these sorts of sentences. Therefore, treating each distinctively modified sentence as attributing between John and Bill a distinct *primitive* relation (such as hitting-at-six-in-the-bedroom) threatens to offend against the condition that the theory be finite. On the basis of considerations of this sort, Davidson proffers a proposal which reveals the common elements in these sentences, issues in the correct semantic truth-conditions, and validates the requisite implications - for instance, that (4) implies (3), and so forth. He takes sentences like (1) and (3)-(5) to harbour existential quantifiers ranging over events. The thesis that there are events is true because the best semantics for English requires quantification over them, as witnessed by their counterparts in first order logic ($1'$), and ($3'$)-($4'$) respectively:

($1'$) There is an event *e* and *e* is a hitting of Bill by John.
($3'$) There is an event *e* and *e* is a hitting of Bill by John and *e* occurs at six.
($4'$) There is an event *e* and *e* is a hitting of Bill by John and *e* occurs at six and *e* occurs in the bedroom.

This technique for discerning ontological commitments extends to all cases where quantification and predication are required in order to construct a satisfactory semantics for natural language. The method is a general method for doing ontology (see Adverbs; Ontological commitment).

8 Animal thought

According to Davidson, the capacity to think requires facility with language, so that only creatures with a language can think. He defends this controversial thesis in 'Thought and Talk' (1984) and 'Rational Animals' (1985). His argument begins by noting that ascriptions of propositional attitudes - belief, desire, intentions and the like - exhibit semantic opacity. In attributing, for example, to a dog the belief that the cat went up the tree we might wonder whether the propriety of this ascription would be affected were we to substitute for 'the tree' another expression that refers to the tree. If not, this would disclose that our attribution of belief to the dog falls short of literalness. If a dog holds a belief about a tree, it must do so under some description or other (1985: 475).

Davidson's main contention is that semantic opacity could be had only by connecting thought to language. He advances two different lines of argument to this conclusion. The first appeals to holism and it is of the form that since we could never have grounds for ascribing the required background beliefs to creatures without a language, we could never be warranted in ascribing to such creatures any thoughts at all. The argument is that since ascriptions of belief exhibit semantic opacity and since semantic opacity requires that we regard beliefs as possessing some definite intentional content and since the possession of a belief with a definite content presupposes 'endless' further beliefs, it follows that a creature to whom we are warranted in ascribing a belief is one possessing a sophisticated behavioural repertoire; and only linguistic behaviour exhibits the sort of complex pattern that might warrant such ascriptions.

Even if sound, this argument at most establishes that we are unlikely ever to have decisive evidence that a speechless creature has beliefs. Davidson, however, draws the stronger conclusion that 'unless there is actually such a complex pattern of behavior, there is no thought' (1985: 476). He is aware of this shortcoming in his argument and offers another. He argues that propositional attitudes require a dense network of beliefs, that 'in order to have a belief, it is necessary to have the concept of belief', and that 'in order to have the concept of belief one must have language', that is, one must be a member of a 'speech community' (1985: 478). How, it will be asked, does Davidson get from the ubiquity of belief and the claim that beliefs in turn require second-order beliefs to the conclusion that 'a creature must be a member of a speech community if it is to have the concept of belief' (1984: 170)?

Davidson argues as follows: the possibility of belief or thought generally depends on the concept of a representation that might be true or false; and a concept of truth and falsity includes some notion of an objective, public domain. And this, in turn, is possible only for an interpreter (1984: 170; 1985: 480). Davidson seems to hold that only utterances can afford the fine-grained structure required for attributing thought; for only a creature whose behaviour exhibits the kind of structure implied by a theory of meaning is a creature in which semantically opaque representations can make an appearance (see Animal language and thought).

9 Conceptual schemes

In 'On the Very Idea Of a Conceptual Scheme' (1984), Davidson argues that no good sense can be made of the idea that different people, communities, cultures or periods view, conceptualize, or make the world (or their worlds) in different ways; or of the distinction between conceptual scheme and empirical content. He associates conceptual schemes with sets of intertranslatable languages (1984: 191). Different conceptual schemes then correspond to non-intertranslatable languages. He then argues that it is hard to make sense of a total failure of translatability between languages (1984: 185). No one could be in a position to judge that others had concepts or beliefs radically different from their own (1984: 197).

In part, Davidson holds this view both because strong pressures arise from the very nature of radical interpretation (RI) and because 'all understanding of the speech of another involves radical interpretation' (1984: 125). Since RI, and therefore, all interpretation, has a holistic character, it makes no sense to ascribe a single belief to a person except against the background of a very large number of other beliefs. Furthermore, 'belief is in its nature veridical'. That it is so, can be seen 'by considering what determines the existence and contents of a belief' (1986: 432). It all comes back to Davidson's views about the conditions of correct attribution of beliefs and other propositional attitudes, that is, the possibility of RI. In interpreting another, an interpreter ventures hypotheses as to what in the circumstances in question causes the speaker to hold-true the sentence in question and this is supposed to provide them (normally) with the meaning of that sentence. In every case there will be many different

causal chains leading to the same utterance. An interpreter must choose one cause and does so by responding to something in the environment ('triangulating') and so converging on something that is a common cause both of their own response and of the utterance of the speaker, thereby correlating the two and thus giving content of the speaker's utterances (Davidson 1991: 159-60).

One immediate consequence is that an RI cannot find speakers to have beliefs if they themselves have no opinion as to their general truth and falsity. Given what beliefs are and how their contents are determined on this story, Davidson is committed to its being impossible that all a person's beliefs about the world might be false. An RI must therefore have beliefs about the world in order to succeed in ascribing beliefs about the world to others. But they also must find that others - if they have beliefs at all - largely agree with them in those beliefs. If Davidson is right, then the crucial aspect of his theory of RI is the importance of causality in determining what someone means or believes. We cannot 'in general fix what someone means independently of what he believes and independently of what caused the belief…. The causality plays an indispensable role in determining the content of what we say and believe' (Davidson 1986: 435). So, it is the central role of causation in the fixing of the contents of beliefs that ensures the truth of everything we believe is not in general 'logically independent' of having those very beliefs; and that others cannot differ too much from us about what it is they believe. The way in which the contents of beliefs are determined puts limits on the extent of falsity and diversity discriminable in a coherent set of beliefs. The method of RI 'enforces' on any successful interpreter the conclusion that a speaker's beliefs are largely true and largely like their own. Thus, the possibility of global scepticism is ruled out.

10 Against facts

In recent writings, for example, 'The Myth of the Subjective' (1989) and 'The Structure and Content of Truth' (1990), Davidson attempts to refute the claim that there are representations of reality. In the latter he argues against 'the popular assumption that sentences, or their spoken tokens, or sentence-like entities or configurations in our brains can properly be called "representations", since there is nothing for them to represent' (Davidson 1990). If facts do not make sentences true, then in what sense are they representations? On Davidson's approach to meaning no entities (facts, objects or events) correspond to sentences.

Davidson's main argument against facts and correspondence theories of truth appears in his 'True to the Facts' (1984). There we find the Great Fact argument that given certain plausible assumptions there is at most one fact. The assumptions are 'that a true sentence cannot be made to correspond to something quite different by the substitution of coreferring singular terms, or by the substitution of logically equivalent sentences' and from these assumptions one can prove that 'if true sentences correspond to anything, they all correspond to the same thing' (Davidson 1990: 303). Rightly or wrongly, Davidson takes these assumptions to embody traditional wisdom about facts (see Neale 1995). The main point is that if a context satisfies these two assumptions, then that context is truth functional. So, if a sentence is made true by some fact *f*, then every materially equivalent sentence is made true by *f* as well, and thus, *f* is the Great Fact (see Facts §2).

See also: Akrasia §2; Anomalous monism; Indirect discourse §2; Intention §3; Meaning and understanding §2; Language, social nature of §4; Radical translation and radical interpretation §§6-8

ERNIE LEPORE

List of works

Davidson, D. (1980) *Actions and Events*, Oxford: Oxford University Press.(Contains all of Davidson's essays on action and events mentioned in §§1-2: 'Actions, Reasons, and Causes', 'Causal Relations' and 'The Individuation of Events'.)

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press. (Includes essays on the philosophy of language ('Thought and Talk' and 'On the Very Idea of a Conceptual Scheme') and 'True to the Facts'. Others mentioned in the article are listed under separate titles.)

Davidson, D. (1985) 'Rational Animals', in E. Lepore and B. McLaughlin (eds) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell, 1985.(Argues against the idea of nonlinguistic thinkers.)

Davidson, D. (1986) *'A Coherence Theory of Truth and Knowledge'*, in E. Lepore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell. (Elaborates a theory of the nature and content of beliefs.)

Davidson, D. (1989) *'The Myth of the Subjective'*, in M. Karusz (ed.) *Relativism: Interpretation and Confrontation*, Paris: University of Notre Dame Press, 1989.(One strand of Davidson's argument that there are no representations of reality.)

Davidson, D. (1990) 'The Structure and Content of Truth', *The Journal of Philosophy* 87: 279-328.(Argues that there is nothing in reality for sentences to represent.)

Davidson, D. (1991) 'Epistemology Externalized', *Dialectica* 45: 191-202.(Includes discussion of 'triangulation', contributing to Davidson's argument against the coherence of non-intertranslatable languages.)

References and further reading

Fodor, J. and Lepore, E. (1992) *Holism: A Shopper's Guide*, Oxford: Blackwell.(An extended consideration of arguments, including Davidson's, that the contents of thoughts, concepts and words are fixed only 'holistically' - in concert - rather than each in isolation.)

Lepore, E. (ed.) (1986) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.(Useful essays on this aspect of Davidson's work.)

Lepore, E. and Loewer, B. (1989) 'You Can Say That Again', *Midwest Studies in Philosophy* 14: 338-56. (Explores criticisms and defences of Davidson's account of propositional attitude ascriptions.)

Lepore, E. and McLaughlin, B. (eds) (1985) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.(Useful essays on this aspect of Davidson's work.)

Neale, S. (1995) 'Gödel and the Best of All Possible Slingshots', *Mind* 104: 761-825.(A detailed look at arguments by Davidson and others for denying that sentences stand for entities more specific than truth-values.)

Daxue

Originally a chapter in the Liji (Book of Rites), one of the Five Classics in the Confucian tradition, the Daxue (Great Learning) has for centuries attained the status of a canon, arguably the most influential foundational text in East Asian Confucian humanism. When the great neo-Confucian thinker Zhu Xi grouped the Daxue with the Zhongyong (Doctrine of the Mean), another chapter in the Liji, the Confucian Analects and the Mengzi as the Four Books, its prominence in the Confucian scriptural tradition was assured. Since the Four Books with Master Zhu's commentaries became the required readings for the civil service examinations in 1313, and since Master Zhu insisted that the Daxue must be studied first among the Four, it has been widely acknowledged as the quintessential Confucian text.

Intent on presenting the case of Confucian moral and political agenda, the *Daxue (Great Learning)* establishes the three basic principles: enlightening the enlightened virtue, loving the people and abiding in the highest good. These neatly outline the basic concerns of the Confucian project: self-cultivation, social responsibility and ultimate human flourishing (see Confucian philosophy, Chinese). Enlightening the enlightened virtue is based on the premise that although human nature is naturally endowed with 'enlightened virtue' (*mingde*), a life-long moral self-cultivation is necessary so that our innate virtue will continue to be enlightened (see Self-cultivation, Chinese theories of). This paradox is central to Confucian teaching: while ontologically we are already sages, existentially we must always try to become sages. The second principle, 'loving the people', specifies that self-cultivation, far from being the quest for inner spirituality as a lonely struggle, necessarily involves communication with others. The self as a centre of relationships entails an ever-expanding network of human-relatedness. Loving the people, comparable to the Christian dictum of 'loving thy neighbour', is, according to Wang Yangming (1472-1529), a constitutive aspect of self-cultivation. However, Zhu Xi interpreted this particular principle significantly differently, as 'renew the people'. He was able to do so by changing one single word *qinmin* (to be close in the sense of loving and caring) to *xinmin* (to renew the people). This conflict of interpretation has profound implications for the style of governance implicit in the *Daxue*. The third principle, 'abiding in the highest good', does not raise philological issues, but the semantic significance is difficult to determine. Does it mean that the loftiest moral aspiration must serve as guidance for ordinary human existence, or that we must strive to approximate that which is always beyond our reach?

However, it is the eight steps in the *Daxue* that offer a holistic vision and a practical programme of self-cultivation, a way of learning to be human. These eight steps are: (1) investigation of things, (2) extension of knowledge, (3) authentication of the will, (4) rectification of the heart-and-mind, (5) cultivation of the person, (6) regulation of the family, (7) governance of the state and (8) universal peace. All are centred around self-cultivation (step 5). Thus, steps 1-4 constitute the inner dimension and 6-8 symbolize the outer manifestation of self-cultivation. The centrality of self-cultivation is unequivocal:

> From the Son of Heaven down to the common people, all must regard cultivation of the person as the root. There is never a case where the root is in disorder and yet the branches are in order. There has never been a case when what is treated with great importance becomes a matter of slight importance and what is treated with slight importance becomes a matter of great importance.
>
> (*Daxue*)

The sense of priority and sequence is crucial in the *Daxue*. The life orientation implicit in the text demands that a sense of direction (toward the highest good) can provide an anchorage for our moral self-realization. With this anchorage, we can have a fixed purpose which leads to tranquility of the mind. With tranquility of the mind, we can attain serene repose and only in serene repose can we cogitate. Only through cogitation can we fully embody the virtue that is originally ours. The Confucian project so conceived esteems learning, both as an acquisition of knowledge and as a process of self-transformation. However, it seems obvious that the acquisition of knowledge through the investigation of things is not empirical work for the sake of enhancing our cognitive intelligence; rather, it is the cultivation of moral knowledge as an integral part of our self-understanding.

The relationship between knowledge and morality raises challenging questions in Chinese philosophy. Is it possible that the investigation of things and the extension of knowledge naturally lead to the authentication of the

will? Must we begin the process of self-cultivation by making up our mind that the knowledge of the good is not empirically acquired but experientially confirmed? Should the investigation of things be understood as learning about the external world or should it be conceived as a form of self-reflection in the lifeworld? Understandably, the *Daxue* has become a basic reference and a primary source for numerous fruitful scholarly and philosophical debates in Chinese intellectual history.

However, it is beyond dispute that this approach to learning presupposes that the primary purpose of education is character building. The *Daxue*, as a form of adult education, offers a comprehensive programme for human flourishing. The programme, built upon the 'Elementary Learning', a highly ritualized guide to self-improvement, focuses on political leadership. In the modern context, the argument makes sense only if we think of it as addressed to rulers and their ministers, rather than to any ordinary person in search of moral guidance. Nevertheless, the strength of the argument lies in its assumptive reason that rulership, or leadership in general, proves to be identical with the moral self-cultivation of all people in ordinary human existence in the lifeworld. Indeed, how we exercise power and influence in the presumed privacy of our homes is politically meaningful and, thus, ought to be made publicly accountable. The *Daxue*, a brief essay of some 1,750 words, carries a profound message for cultivating politically significant ethical intelligence for the global community.

See also: Chinese Classics; Chinese philosophy; Confucian philosophy, Chinese; Self-cultivation, Chinese theories of; Zhongyong

TU WEI-MING

References and further reading

Chan Wing-tsit (1963) *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press.(A standard translation of the *Daxue*, with commentary.)

Schwartz, B.I. (1985) *The World of Thought in Ancient China*, Cambridge, MA: Harvard University Press.(A discussion of the philosophical content of this short yet influential document.)

Tu Wei-ming (1989) *Centrality and Commonality: An Essay on Confucian Religiousness*, Albany, NY: State University of New York Press.(Although focused primarily on the *Zhongyong*, this monograph provides a philosophical context for the *Daxue*.)

De

Across the corpus of pre-Qin philosophical literature, de*, conventionally translated as 'potency' or 'virtue', seems to have a fundamental cosmological significance from which its other connotations are derived. We begin from the pervasive assumption that existence is an uncaused, spontaneous process. It is* ziran*: so-of-itself. As a total field, this dynamic process is called* dao*; the individuated existents in this field - its various foci - are called* de*.*

In the Daoist literature, *de* denotes the particular as a focus of potency in this process conception of existence. *De* is any particular disposition of the totality. The *Daodejing* - literally, the classic of *dao* and *de* - states: 'The great *dao* is so expansive. It reaches in all directions. All of individuated existence arises because of it' (*Daodejing* 34) (see Daodejing).

When *de* is cultivated and accumulated, so that the particular is integrated utterly with its environments, the distinction between *dao* and *de* - between field and focus - collapses. *De* as an individuating notion is transformed into *de* as an integrating notion. The *Daodejing* 28 observes: 'One who possesses *de* in abundance is comparable to a new-born babe'. In this literature, both Daoist and Confucian, the infant, the uncarved block and the exemplary person are all metaphors for a condition in which one does not distinguish oneself from one's environments. These metaphors illustrate the assumption that any particular, when viewed in terms of its intrinsic relatedness, entails the full process of existence. Such being the case, because the babe is a matrix through which the full consequence of undiscriminated existence can be brought to focus and experienced, it can be used as a metaphor for the *de* which is *dao*.

The early Confucian tradition limited its concerns to the social and political problems of the human community. At this level, the cultivation of *de* is to overcome the discreteness and discontinuity of the inchoate self in the direction of continuity and social harmony. Through achieved excellence, one becomes a model for others and an object of deference. *De* is both 'potency' and 'virtue' in the sense that each person is a meaning-creating and meaning-disclosing focus. As one becomes extended in community through patterns of deference, one's capacity for valuation is proportionately increased. The human being is a world maker, and the greater one's proportions, the great one's efficacy as a maker of values.

The extension of *de* entails both an act of intending and the attraction of the support necessary to effect what is intended. The particular initiating the direction focuses this support as coextensive with its own particularity. At the same time, the particular becomes coextensive with 'other' by overcoming its particularity and accommodating the natural direction and volition of the 'other' to integrate them into its own field of interpretation. As artist, as communal leader, as teacher, one is able to organize one's natural, social and cultural environments and disclose their possibilities for productive harmony. One is able to manifest, interpret and display the local culture.

On the political level, the relationship between ruler and people is described in terms of *de*. To the extent that the ruler reaches out to become coextensive with the prevailing cultural propensities of the people, he expresses the values of the community. To the extent that he becomes culturally coextensive with them through deference, they come to share his values and moral insights, and he wins them over. If we consider agency, *de* means literally 'bounty' or 'beneficence' as it extends down from the ruler, and 'gratitude' as it is disclosed in the deference of the people. If, more appropriate to the Chinese tradition, we think situationally, *de* is the disposition of being both benefactor and beneficiary through productive, non-coercive communal relations. As the *de* of the people is manifest within the ruler's *de*, his potency is enhanced and he becomes the wind that bends the grass, the north star around which the other stars revolve, the maker and transmitter of culture to which other persons subscribe. It is in this sense that the emperor is the empire, and the patriarch is the family.

There is a coextensiveness of ruler and people in the sense that the ruler expresses the people through his government by accommodating their cultural proclivities and thus orchestrating an interpretation of culture that is both his and theirs. There is always, in this conception of *de*, a tension between attempting to maximize the possibilities for harmony provided by attending conditions, and the expression of an interpretation that does justice to one's own uniqueness and sense of appropriateness. It is the sensitive balancing of these forces that produces harmony within all human environments (see Culture; Sovereignty).

See also: Chinese philosophy; Confucius; Dao; Daoist Philosophy

DAVID L. HALL
ROGER T. AMES

References and further reading

Ames, R.T. (1991) 'Nietzsche's "Will to Power" and Chinese "Virtuality" (*De*): A Comparative Study' in G. Parkes (ed.) *Nietzsche and Asian Thought*, Chicago, IL: University of Chicago Press.(A discussion of *de* that seeks to shed light on Nietzschean creativity.)

Hall, D.L. and Ames, R.T. (1987) *Thinking Through Confucius*, Albany, NY: State University of New York Press.(A study of the central vocabulary of classical Confucianism.)

De Man, Paul (1919-83

De Man's work is among the most renowned and influential in American literary theory of the latter twentieth century, especially regarding literary theory's emergence as an interdisciplinary and philosophically ambitious discourse. Always emphasizing the linguistic aspects of a literary work over thematic, semantic or evaluative ones, de Man specifically focuses on the figurative features of literary language and their consequences for the undecidability of meaning. His extension of his mode of 'rhetorical reading' to philosophic texts also participates in the blurring of generic and institutional distinctions between literature and philosophy, a tendency pronounced in French philosophy of the latter twentieth century.

1 Deconstruction and Derrida

Paul de Man, born in Belgium and educated in comparative literature at Harvard, taught comparative literature at Cornell, Johns Hopkins and Yale universities. Associated with Jacques Derrida and deconstruction, and more generally with post-structuralism, de Man already exerted considerable influence through his teaching before the advent of deconstruction and his publication of an essay on Derrida. His association with Derrida and deconstruction is in part due to their subsequent friendship and colleagueship, but it is in part the artefact of the observer. De Man shares with Derrida the unrelenting attention to questions of language, and the rigorous practice of close reading of texts; with regard to deconstruction, he shares that practice's 'reversal' and 'reinscription' of assumed priorities - in de Man's case, the priority of meaning over the linguistic and tropological devices that ostensibly convey meaning, reversed and reinscribed by him as the irreducible textual productivity of tropes (metaphor, chiasmus, prosopopeia and so on) and other linguistic 'materiality' that leaves meaning suspended in undecidability and reading suspended in rhetorical repetitions and 'disfigurations' of the text before it. But he differs from Derrida in that he believes (as he argued in 'The Rhetoric of Blindness', in de Man 1971) that deconstruction is not something that one does to texts - as a philosopher, say, operating against the dominant grain of institutional philosophy - but rather something that texts do by themselves and to their readers. Thus de Man also de-emphasizes the 'institutional' aspect of Derridean deconstruction's 'reinscription' of problematic priorities back upon the institutional loci of their teaching and application, although in his later work issues of authority, ideology and critiques of falsehood are increasingly pronounced.

2 Literary theory

De Man's literary theory developed by way of studies of novels and poetry as well as philosophic texts, but it always displays several dominant features: drawing primarily upon romantic and modern authors, he foregrounds the rhetorical language of texts and explores its consequences for what texts actually, rather than apparently, mean; how they actually, if contradictorily, function; and how, ultimately, they may resist meaning and any satisfied aesthetic sense of their formal and structural workings. In essays from the 1950s and 1960s, borrowing from European phenomenology and its terminologies of consciousness, intentionality, temporality and being, de Man accents the difference and tension between poetic language's power to posit and imagine linguistically determined conditions on the one hand, and literature's false attractions - representational and nostalgic - toward natural objects and ontological being on the other. His highly influential essay 'The Rhetoric of Temporality' (1969), which also signals his turn toward a focus upon rhetorical language, analyses as allegory the perpetual reinstating of a temporal distance between a subject and its origin, and as irony the unbridgeable discrepancy between a linguistic and an empirical self. In his equally influential collection *Blindness and Insight* (1971), de Man studies European writers and critics (Husserl, Heidegger, Binswanger, Lukács, Blanchot, Poulet, Nietzsche, Derrida, Baudelaire, Mallarmé) to demonstrate the priority of a text's 'insight' or self-knowing 'rhetoricity' over the 'blindness' of its readers who unwittingly ('blindly') repeat the text's structure and statement when they believe they are critically most distant from them. De Man here extends the category of the 'literary' to include philosophic texts as well as literature conventionally construed: 'any text that implicity or explicitly signifies its own rhetorical mode and prefigures its own misunderstanding as the correlative of its rhetorical nature' (1971: 136). *Allegories of Reading* (1979) expands the sense of rhetoricity as the undecidability between literal and figurative meanings of a text, in a series of analyses of the 'allegories' or figurative narratives of mis- and non-understanding that generate themselves in response to the textual figures and rhetoric of such authors as Yeats, Rilke and Proust but also Nietzsche and above all, Rousseau. In this book as well as in subsequent essays on

Wordsworth, Shelley, Baudelaire and others (collected in *The Rhetoric of Romanticism*, 1984), de Man increasingly emphasizes what he calls the 'materiality' of language, its 'inscription', meaning the nonsemantic, noncognitive dimensions of language that subtend all linguistic acts not only as grammar, for example, subtends semantics, but also as the sheer positing of linguistic signifiers and their frequent interference with or 'disfiguring' of signification subtend all possible logical, grammatical or semiotic codifications and decodings of language. This latter aspect of his work marks de Man as post-structuralist, in resolutely refusing any reduction of language and signification to code.

3 Philosophical critique

Throughout his work de Man draws with seeming indifference from philosophy as well as literature and criticism to develop his literary theory. This leads to a deliberate blurring of distinctions between the two kinds of discourse, which is perhaps an easy and uncontroversial manoeuvre when he treats authors who themselves challenged or ignored the difference (Pascal, Rousseau, Nietzsche, Benjamin, Heidegger, Blanchot), as, for example, in his reconfounding of Rousseau's confusing differentiations between natural, fictional and historical conditions. But in a number of essays, some left unfinished by him and just recently (1996) published, de Man attempts a more frontal critique of certain assumptions and goals of mainline Western philosophy. Writing on Locke, Kant and Hegel, de Man not only signals the tropological substratum that enables but also destabilizes the establishment of certain philosophic concepts (for example, a 'ground'), he also calls into question the claims that there is thought that is distinguishable from - nobler, more ideal than - language in its materiality (Hegel), and that language ought to or can be modelled on phenomenal and cognitive presentation of meaning from the world (Kant). He subsumes the 'matching' of language, and with it of grammar and logic, to a phenomenal-cognitive model of nature and its human experience, under the names of an 'aesthetic formalization' or 'aesthetic ideology'; and he criticizes it in the name of language's tropological, material and ultimately nonsignifying otherness from all aesthetics and other rational-logical extensions that would establish a truthful continuity between the world and language or vice versa. De Man's work, then, a collection of radical, unique and often incisive readings of literary and philosophic texts that remains (1997) unconsidered by most professional philosophers, offers a view of language and textuality that is an anti-theory of philosophy as well as a controversial theory of literature.

See also: Deconstruction; Post-structuralism

TIMOTHY BAHTI

List of works

Man, P. de (1969) 'The Rhetoric of Temporality', in C.S. Singleton (ed.) *Interpretation: Theory and Practice*, Baltimore, MD: Johns Hopkins University Press; repr. in *Blindness and Insight*, 2nd edn, Minneapolis, MN: University of Minnesota Press. (De Man's single most influential piece of writing.)

Man, P. de (1971) *Blindness and Insight: Essays in the Rhetoric of Contemporary Criticism*, New York: Oxford University Press. (Includes the important critique of Derrida, 'The Rhetoric of Blindness'.)

Man, P. de (1979) *Allegories of Reading: Figural Language in Rousseau, Nietzsche, Rilke and Proust*, New Haven, CT: Yale University Press. (The second part is a sustained deconstructive reading of Rousseau.)

Man, P. de (1984) *The Rhetoric of Romanticism*, New York: Columbia University Press. (Collection of essays, principally on poets, from 1960 to his death.)

Man, P. de (1986) *The Resistance to Theory*, Minneapolis, MN: University of Minnesota Press. (The title essay is especially important for his critique of philosophy.)

Man, P. de (1996) *Aesthetic Ideology*, Minneapolis, MN: University of Minnesota Press. (Collects his most important papers on philosophy.)

References and further reading

Chase, C. (1994) 'De Man, Paul', in M. Groden and M. Kresiwirth (eds) *The Johns Hopkins Guide to Literary Theory and Criticism*, Baltimore, MD: Johns Hopkins University Press. (An authoritative and comprehensive introduction to de Man's literary theory.)

Derrida, J. (1986) *Mémoires for Paul de Man*, trans.C. Linsay, J. Culler and E. Cadava, New York: Columbia University Press. (A profound meditation, at once analytic, speculative and memoiristic, about de Man's work and the place of deconstruction in the USA.)

Gasché, R. (1989) 'In-Difference to Philosophy: De Man on Kant, Hegel and Nietzsche', in L. Waters and W.

Godzich (eds) *Reading de Man Reading*, Minneapolis, MN: University of Minnesota Press. (A rigorously critical analysis of de Man's 'indifference' to some of philosophy's assumptions and procedures.)

Godzich, W. (1983) 'Caution! Reader at Work!', introduction to 2nd edn of P. de Man, *Blindness and Insight*, Minneapolis, MN: University of Minnesota Press. (A useful introduction, distinguishing de Man's 'reading' from other critical and interpretive practices.)

Godzich, W. (1986) 'The Tiger of the Paper Mat', foreword to P. de Man, *The Resistance to Theory*, Minneapolis, MN: University of Minnesota Press. (Especially helpful in sketching the philosophical background and importance of de Man's use of resistance.)

Hamacher, W. (1989) 'LECTIO: De Man's Imperative', in L. Waters and W. Godzich (eds) *Reading de Man Reading*, Minneapolis, MN: University of Minnesota Press. (The most incisive analysis thus far of the active or performative aspects of 'reading' in de Man's writings.)

Jacobs, C. (1989) 'Allegories of Reading Paul de Man', in L. Waters and W. Godzich (eds) *Reading de Man Reading*, Minneapolis, MN: University of Minnesota Press. (The best example of deploying de Man's reading of allegory to read de Man on allegory.)

De Morgan, Augustus (1806-71)

Augustus De Morgan was an important British mathematician and logician. Much of his logical work was directed to expanding the traditional syllogistic theory, and to meeting the objections of Sir William Hamilton and his allies to the techniques he used. More important for the future of logic, though, was De Morgan's work in two areas: the logic of complex (compound) terms, in which he essentially developed the theory of Boolean algebra, and his introduction of the logic of relations as a serious topic for formal logic. His work on probability logic, while flawed, was also significant.

Augustus De Morgan was born in Madura, India. He grew up in England and graduated from Trinity College, Cambridge in 1827. His mathematical abilities were such that he was appointed Professor of Mathematics at the newly founded University of London (now University College) the following year. He resigned on grounds of principle in 1831, but returned in 1836. He then held his professorship for thirty years, resigning, once again for reasons of principle, in 1866.

De Morgan was an amazingly prolific writer for both specialist and general audiences. As a mathematician, he clarified the concepts of mathematical induction and of limits, and he proved an important result about infinite series. He made significant contributions to the foundations and philosophy of algebra, and was deeply interested in the history and bibliography of mathematics. He is best known for the logical works which he published from 1847 to 1862. Their interpretation is difficult, however, for De Morgan's unusually fertile imagination was often constrained by being placed at the service of an essentially conservative attempt to extend the theory of the syllogism.

Much of De Morgan's published work arose from his highly publicized controversy with Sir William Hamilton over the quantification of the predicate. At first this dispute concerned Hamilton's charge that De Morgan had plagiarized from Hamilton's work, but it was soon resolved in De Morgan's favour. It then became a lively controversy over the comparative merits of Hamilton's and De Morgan's modifications of traditional logic. To the modern reader, much of this dispute seems very dated. However, De Morgan's analyses of categorical propositions remain interesting. He introduced into logic the concept of the universe of discourse, and he allowed that for each term, '*X*', there would also be its contrary, '*x*'. With this notation, there are eight possible categorical propositions:

(1) All *X*s are *Y*s
(2) All *x*s are *y*s
(3) All *X*s are *y*s
(4) All *x*s are *Y*s
(5) Some *X*s are *Y*s
(6) Some *x*s are *y*s
(7) Some *X*s are *y*s
(8) Some *x*s are *Y*s

All except (4) and (6) can be expressed as traditional categorical propositions using only '*X*' and '*Y*'; (4), however, must be stated as 'Everything is *X* or *Y*', and (6), its negation, as 'Some things are neither *X*s or *Y*s'. De Morgan devised an ingenious 'spicular' notation for expressing these forms, which allowed him to state concisely the rules of inference for this enlarged system.

If De Morgan's syllogistic modifications look backwards, two facets of his work look forwards. The first is the often neglected logic of complex (or compound) terms in his *Formal Logic* (1847: ch. 6). In this treatment, '*XY*' stands for the 'conjunction' of the classes *X* and *Y*, and '*X*, *Y*' stands for their 'disjunction'. The most familiar feature of this section is 'De Morgan's laws':

$\text{non-}(XY) = x, y$ and $\text{non-}(X, Y) = xy$

However, there is much more to this section than these laws; it contains, in essence, the full theory of Boolean algebra. In fact, De Morgan's system is probably closer to modern Boolean algebra than Boole's (see Boole, G.;

Boolean algebra).

One interesting feature of De Morgan's account is his characterization of conjunction and disjunction. Where '*X*)*P*' stands for 'Every *X* is *P*', the Conjunctive Principle is

$$\text{If } X)P \text{ and } X)Q \text{ then } X)PQ,$$

and the Disjunctive Principle is

$$\text{If } P)X \text{ and } Q)X \text{ then } P,Q)X.$$

These laws for conjunction and disjunction (intersection and union) today provide the basis for one of the two common formulations of Boolean algebra.

De Morgan's other main contribution to logic was his powerful logic of relations. Philosophical considerations led him to give a relational analysis of all propositions; and this resulted, in turn, in his formal logic of relations ('On the Syllogism IV', 1860).

Where '*X*' and '*Y*' are singular terms, and '*L*' and '*M*' are relational expressions, De Morgan used the following notation:

$X..LY$ *X* is an *L* of *Y*
$X.LY$ *X* is not an *L* of *Y*
$X..lY$ *X* is a non-*L* of *Y*
$X..LMY$ *X* is an *L* of an *M* of *Y*
$X..LM'Y$ *X* is an *L* of every *M* of *Y*
$X..L,MY$ *X* is an *L* of none but *M*s of *Y*
$L))M$ *L* is included in *M*

De Morgan stated a great many relational identities, such as

$$\text{not}(L^{-1}) = l^{-1}$$

and

$$(LM)^{-1} = M^{-1}L^{-1}$$

One of De Morgan's most interesting propositions is 'Theorem K':

$$\text{If } LM))N, \text{ then } nM^{-1}))1,$$

for which he provided an elegant demonstration.

De Morgan seemed most interested in applying his logic of relations to the analysis of the syllogism. Thus, while he hinted at some of the elaborations of this logic which would come later, he did not carry them out, nor did he work further on the logic of relations.

Some of De Morgan's other ideas, such as that of the numerically definite syllogism, have had only a minor impact on the history of logic. However, his attempt to include within logic the theory of probability, conceived of as the theory of rational degrees of belief, was of considerable importance. His concern was in calculating the probability of the conclusion of a logically valid argument whose premises are merely probable. Within this domain, he emphasized traditional problems about the uses of authority and testimony in argument. While his analyses of such cases were insightful, they were often flawed by a failure to require the stochastic independence of the elements of a conjunction when replacing the probability of a conjunction by the product of the probabilities of its conjuncts (see Probability theory and epistemology).

See also: Logic in the 19th century §§3-4

DANIEL D. MERRILL

List of works

De Morgan, A. (1838) *An Essay on Probabilities*, London: Longman's; repr. New York: Arno Press, 1981. (Contains a useful introduction to De Morgan's philosophy of probability.)

De Morgan, A. (1847) *Formal Logic*, London: Taylor & Walton; repr. London: Open Court, 1926. (De Morgan's major early work.)

De Morgan, A. (1966) *On the Syllogism and Other Logical Writings*, ed. P. Heath, New Haven, CT: Yale University Press.(A carefully prepared comprehensive edition of De Morgan's other logical writings. His essay 'On the Syllogism IV' (1860) appears on pages 208-46.)

References and further reading

Hailperin, T. (1988) 'The Development of Probability Logic from Leibniz to MacColl', *History and Philosophy of Logic* 9: 131-91.(A critical study of several attempts at devising a probability logic, including De Morgan's.)

Heath, P. (1966) 'Introduction' to *On the Syllogism and Other Logical Writings*, New Haven, CT: Yale University Press, vii-xxxi.(An account of De Morgan's life and works, especially strong on the controversy with Hamilton.)

Mansel, H.L. (1851) 'Recent Extensions of Formal Logic', *North British Review* 15: 90-121.(An extended criticism of the new logical work of De Morgan and others from a more traditional point of view.)

Martin, R.M. (1979) 'De Morgan and the Logic of Relations', in R.M. Martin, *Peirce's Logic of Relations and Other Studies*, Lisse: Peter de Ridder Press, 46-53.(An interpretation of De Morgan's logic of relations using the language of contemporary symbolic logic.)

Merrill, D.D. (1990) *Augustus De Morgan and the Logic of Relations*, Dordrecht: Kluwer.(A study of the philosophical and technical aspects of De Morgan's logic of relations.)

De re/de dicto

*'*De re*' and '*de dicto*' have been used to label a host of different, albeit interrelated, distinctions. '*De dicto*' means 'of, or concerning, a dictum', that is, something having representative content, such as a sentence, statement or proposition. '*De re*' means 'of, or concerning, a thing'. For example, a* de dicto *belief is a belief that a bearer of representative content is true, while a* de re *belief is a belief concerning some thing, that it has a particular characteristic.*

Consider the following example:

John believes his next-door neighbour is a Buddhist.

This statement is ambiguous. Construed de dicto*, it is true in the following circumstance. John has never had any contact with his next-door neighbour. Nevertheless, John believes that his next-door neighbour is bound to be a Buddhist. Construed in this* de dicto *fashion, the statement does not attribute to John a belief that is distinctively about a particular individual. In contrast, construed* de re*, it does attribute to John a belief that is about a particular individual. For example, construed* de re*, the statement is true in the following circumstance. John encounters his next-door neighbour, Fred, at a party without realizing that Fred is his next-door neighbour. On the basis of his conversation with Fred, John forms a belief about the individual who is in fact his next-door neighbour to the effect that he is a Buddhist.*

1 The different distinctions

'*De dicto*' means 'of, or concerning, a dictum'. '*De re*' means 'of, or concerning, a thing'. Roughly, a *de dicto* belief is a belief that a bearer of representative content (that is, a statement, sentence, proposition or such like) is true, while a *de re* belief is a belief concerning some thing, that it has a particular characteristic. (I)-(IX) below are a number of different distinctions that have been described as *de re*/*de dicto* distinctions in the literature.

(I) In the case of modality and belief we can draw scope distinctions. Consider the following sentences:

(1) Every effect necessarily has a cause.
(2) Joan believes her husband is an axe murderer.

(1) and (2) manifest a scope ambiguity. (1) is ambiguous between:

(1′) [squ] (if *e* is an effect then *e* has a cause),

in which the entire sentence 'If *e* is an effect then *e* has a cause' falls within the scope of 'Čsqu;'; and:

(1*) If *e* is an effect, [squ] (*e* has a cause),

in which only the part '*e* has a cause' falls within the scope of 'Čsqu;'. (2) may, likewise, be disambiguated as:

(2′) Joan believes that her husband is an axe murderer

or:

(2*) Joan's husband is believed by Joan to be an axe murderer.

(1′) and (2′) give the *de dicto* readings of (1) and (2). (1*) and (2*) give their *de re* readings. According to (1′), it is impossible for something to be an effect without having a cause - something few would dispute. According to (1*), it is impossible for anything which is an effect not to have been caused - a much more controversial claim. (2′) implies that Joan has the disconcerting belief she could express by uttering, 'My husband is an axe murderer'. (2*) implies that Joan has a less disconcerting belief. According to (2*), she believes that someone, whom, perhaps unbeknownst to her, she is married to, is an axe murderer (see Propositional attitudes §2).

(II) A sentence expresses a *de re* modality provided that some object essentially features in its truth-condition in the following sense: its truth-value depends on the characteristics that the object might have. For example, the

sentence ‘Possibly Fred is a Buddhist’ expresses a *de re* modality because its truth-value depends on whether a certain individual, Fred, could have the characteristic of being a Buddhist.

(III) Some hold that predicates such as ‘is necessarily human’ or ‘is possibly six feet tall’ attribute a special kind of property known as a modal property. A *de re* modal attribution attributes a modal property. For example, ‘John is necessarily human’ expresses a *de re* modal attribution if it attributes to John the modal property of being necessarily human.

(IV) Another *de re*/*de dicto* distinction depends on the distinction between rigid and non-rigid designators. Proper names such as ‘Napoleon’ are said to be rigid designators because they designate the same thing in any possible situation in which they designate at all (see Proper names §2). According to this way of drawing the distinction, a modal sentence expresses a *de re* modality if it contains a rigid designator falling within the scope of a modal operator. For example, ‘It might have been the case that Napoleon spoke Chinese’ expresses a *de re* modality, in contrast to ‘It might have been the case that the most famous French general spoke Chinese’.

(V) Quine (1953) insists that modalities are invariably relativized to a description in the following sense. We cannot say that something necessarily or possibly has a characteristic absolutely: we can only say that something necessarily or possibly has a characteristic relative to a description. Bill Clinton is necessarily a president relative to the description ‘42nd President’. He is contingently a president relative to the description ‘husband of Hillary’. In Quine’s view all modalities are *de dicto*. *De re* modalities, if, contra Quine, they existed, would be those not relativized to a description.

Some further distinctions taken to be distinctions between the *de re* and the *de dicto* are illustrated in application to belief.

(VI) It has been argued that we should distinguish between two kinds of belief. Having a *de dicto* belief amounts to standing in a two place relation to a dictum. Having a *de re* belief amounts to standing in an at least three place relation to an object and a property. Suppose John believes that all prominent politicians are wealthy. In consequence, without knowing who is president, he believes that the president is wealthy. In that case, John will have a *de dicto* belief in virtue of standing in the two place belief relation to the sentence ‘The president is wealthy’.

Now suppose that Sally meets Bill Clinton, and judges from the cut of his clothes that he is wealthy. Whether or not she believes he is president, Sally believes of the president that he is wealthy. According to this way of distinguishing *de re* from *de dicto* beliefs, Sally stands in the three place belief relation to the president and the characteristic of being wealthy.

(VII) In another way of distinguishing between *de re* and *de dicto* beliefs, a *de re* belief is about a particular individual or individuals, whereas a *de dicto* belief is general. Sally’s belief about the president that he is wealthy is a belief about a particular individual to the effect that he is wealthy. On the other hand, John’s belief that the president is wealthy is the general belief that whoever occupies the presidency is wealthy.

(VIII) A different epistemic distinction between *de re* and *de dicto* beliefs is as follows. In contrast with a *de dicto* belief, in order to have a *de re* belief one must be epistemically acquainted with the subject of one’s belief: one must know who, which or what it is.

(IX) A final distinction is that between object-dependent and object-independent beliefs. Sally sees a cat and forms the belief that she expresses by uttering, ‘That cat has mange’. Some say that she cannot hold this belief about the cat unless it exists. If so, her belief is object-dependent. It is sometimes held that *de re*, as opposed to *de dicto*, beliefs are object-dependent in this sense.

2 Relationships between the different distinctions

Consider the following examples:

(3) [squ] (the subject of Gance’s best-known film was human).
(3′) The subject of Gance’s best-known film is such that [squ] it was human.

Abel Gance’s best-known film was about Napoleon. According to distinction (I) above, sentence (3) expresses a

de dicto modality and (3′) a *de re* modality. ‘The subject of Gance’s best-known film’ falls within the scope of ‘Čsqu;’ in (3), and outside the scope of that modal operator in (3′). So Napoleon does not essentially feature in the truth-condition of (3). Hence, distinctions (I) and (II) above would agree in counting (3) as a *de dicto* modality. Since ‘the subject of Gance’s best-known film’ falls outside the scope of ‘Čsqu;’ in (3′), (3′) is true if and only if it is impossible for Napoleon to have been nonhuman. So (I) and (II) likewise agree in counting (3′) as a *de re* modality.

Distinction (III) has it that sentences expressing *de re* modalities attribute modal properties. The truth-value of (3) does not depend on Napoleon having the modal property of being necessarily human. Arguably, the truth-value of (3′) does depend on his having that modal property. So (III), like (I) and (II), decrees that (3) is a *de dicto* modality and (3′) is *de re*.

‘The subject of Gance’s best-known film’ is a non-rigid designator because it designates different things in different possible situations. Since ‘the subject of Gance’s best-known film’ falls within the scope of the modal operator in (3), (IV) deems (3) to be a *de dicto* modality. In (3′), ‘the subject of Gance’s best-known film’ falling outside the scope of ‘Čsqu;’ has the same effect as replacing ‘the subject of Gance’s best-known film’ with a rigid designator such as ‘Napoleon’ in (3).

(3*) [squ] (Napoleon was human).

The truth-value of (3*) depends on whether the actual referent of ‘Napoleon’ might have been nonhuman. According to (IV), (3*) expresses a *de re* modality for basically the same reason that (3′) does according to (II) and (III).

Distinction (V) implies that a modality is *de dicto* if it is relativized to a description. Otherwise, it is *de re*. (3) is true if and only if nothing could be nonhuman which satisfies the description ‘the subject of Gance’s best-known film’. In Quine’s sense the modality in (3) is relativized to a description, and is an example of a *de dicto* modality. In contrast, since the definite description in (3′) falls outside the scope of the modal operator, the truth of (3′) requires that the individual who actually satisfies that definite description must be human. (3′) is true if and only if Napoleon, the individual actually referred to by ‘the subject of Gance’s best-known film’, must be human. (V) implies that (3′) is a *de re* modality.

Now consider:

(4) John believes that the president is wealthy.
(4′) The president is such that Sally believes he is wealthy.

According to (VI) above, a *de dicto* belief relates a believer to a proposition, sentence or statement, whereas a *de re* belief relates a believer to something that may be extralinguistic. John’s *de dicto* belief that the president is wealthy relates John to the sentence ‘The president is wealthy’. The *de re* belief attributed to Sally by (4′) relates her to Bill Clinton.

Since Sally’s *de re* belief about the president relates her to Clinton, she could not have that belief unless Clinton exists. Hence, the belief attributed to Sally by (4′) turns out to be object-dependent in conformity with distinction (IX). Moreover, if Sally’s *de re* belief about the president relates her to Clinton, it is about a specific individual. Hence, it is *de re* in conformity with (VII). Since Sally knows who it is she believes to be wealthy, her belief is also *de re* in the terms of distinction (VIII).

See also: Essentialism; Logical and mathematical terms, glossary of; Use/mention distinction and quotation

ANDRÉ GALLOIS

References and further reading

Burge, T. (1977) ‘Belief *De Re*’, *Journal of Philosophy* 74: 317-38.(A classic attempt to articulate the notion of a *de re* belief, and argue that it is more fundamental than the notion of a *de dicto* belief.)

Dennett, D. (1982) ‘Beyond Belief’, in A. Woodfield (ed.) *Thought and Object*, Oxford and New York: Oxford University Press.(A sustained attempt to call into question the distinction between *de re* and *de dicto* beliefs.)

Fine, K. (1989) ‘The Problem of *De Re* Modality’, in J. Almog, J. Perry and H.K. Wettstein (eds) *Themes From*

Kaplan, Oxford and New York: Oxford University Press, 197-272.(A somewhat difficult, but rewarding, discussion of key distinctions related to *de re* modalities.)

Kneale, W. (1960) 'Modality *De Dicto* and *De Re*', in E. Nagel, P. Suppes and A. Tarski (eds) *Logic, Methodology and the Philosophy of Science: Proceedings of the 1960 International Conference*, Stanford, CA: Stanford University Press, 1962, 622-33.(Particularly useful in providing a historical background to the distinction between *de re* and *de dicto* modalities.)

Kraut, R. (1983) 'There are No *De Dicto* Attitudes', *Synthese* 54: 275-94.(A defence of scepticism about the distinction between *de re* and *de dicto* attitudes.)

McDowell, J. (1984) '*De Re* Senses', *Philosophical Quarterly* 34: 283-94.(Includes a discussion of the alleged object dependence of *de re* thoughts.)

Plantinga, A. (1969) '*De dicto* et *de re*', *Noûs* (3): 235-58.(An attempt to reduce *de re* to *de dicto* modalities.)

Quine, W.V. (1953) 'Reference and Modality', in *From a Logical Point of View*, New York: Harper, 139-59. (Includes Quine's argument against the meaningfulness of *de re* modal attributions.)

Quine, W.V. (1976) 'Quantifiers and Propositional Attitudes', in *Ways of Paradox and Other Essays*, Cambridge: Cambridge University Press, 185-96.(A classic statement of one way of distinguishing between *de re* and *de dicto* beliefs.)

Quine, W.V. (1979) 'Intensions Revisited', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press, 268-74.(Quine's second thoughts about the *de re*/*de dicto* distinction.)

Death

Reflection on death gives rise to a variety of philosophical questions. One of the deepest of these is a question about the nature of death. Typically, philosophers interpret this question as a call for an analysis or definition of the concept of death. Plato, for example, proposed to define death as the separation of soul from body. However, this definition is not acceptable to those who think that there are no souls. It is also unacceptable to anyone who thinks that plants and lower animals have no souls, but can nonetheless die. Others have defined death simply as the cessation of life. This too is problematic, since an organism that goes into suspended animation ceases to live, but may not actually die.

Death is described as 'mysterious', but neither is it clear what this means. Suppose we cannot formulate a satisfactory analysis of the concept of death: in this respect death would be mysterious, but no more so than any other concept that defies analysis. Some have said that what makes death especially mysterious and frightening is the fact that we cannot know what it will be like. Death is typically regarded as a great evil, especially if it strikes someone too soon. However, Epicurus and others argued that death cannot harm those who die, since people go out of existence when they die, and people cannot be harmed at times when they do not exist. Others have countered that the evil of death may lie in the fact that death deprives us of the goods we would have enjoyed if we had lived. On this view, death may be a great evil for a person, even if they cease to exist at the moment of death.

Philosophers have also been concerned with the question of whether people can survive death. This is open to several interpretations, depending on what we understand to be people and what we mean by 'survive'. Traditional materialists take each person to be a purely physical object - a human body. Since human bodies generally continue to exist after death, such materialists presumably must say that we generally survive death. However, such survival would be of little value to the deceased, since the surviving entity is just a lifeless corpse. Dualists take each person to have both a body and a soul. A dualist may maintain that at death the soul separates from the body, thereby continuing to enjoy (or suffer) various experiences after the body has died.

Some who believe in survival think that the eternal life of the soul after bodily death can be a good beyond comparison. But Bernard Williams has argued that eternal life would be profoundly unattractive. If we imagine ourselves perpetually stuck at a given age, we may reasonably fear that eternal life will eventually become rather boring. On the other hand, if we imagine ourselves experiencing an endless sequence of varied 'lives', each disconnected from the others, then it is questionable whether it will in fact be 'one person' who lives eternally.

Finally, there are questions about death and the meaning of life. Suppose death marks the end of all conscious experience - would our lives be then rendered meaningless? Or would the fact of impending death help us to recognize the value of our lives, and thereby give deeper meaning to life?

1 Analysis of death versus criterion of death

The most fundamental philosophical question about death is the question concerning its nature or essence - 'What is death?'. When philosophers offer answers to this question, they may be said to be defining death. It is important to recognize that two distinct projects may be confused under the single name 'defining death'. The first is a project in conceptual analysis which tries to give an account of the nature of death. Since death is the event that takes place when an organism dies, one way to explain the nature of death would be to formulate a definition of '*x* dies at *t*'. If successful, such a definition would tell us what we mean when we say that something dies, and thus reveal the nature of death. Among the most popular proposals is:

(D1) *x* dies at $t =_{df.} x$ ceases to be alive at *t*

A definition of this sort is successful if it is true. A fully adequate definition of death would display the structure of the concept of death.

The second (or 'criterial') project is one of public policy. For many practical purposes, it is important to have agreement on the question of whether a person is dead or alive. Furthermore, for practical purposes, it is important that there should be agreement about the time of death. Even if we agreed that (D1) is true, we might still be in grave doubt about whether certain people are dead. For example, consider a person whose brain has been

irreparably destroyed in an accident, but whose blood is being oxygenated and circulated by life-support mechanisms. Since it is unclear whether people such as this have ceased to be alive, it is unclear whether they are dead. Those who engage in the criterial project try to formulate a criterion of human death that is relatively easy to apply. Such a criterion would pick out an observable change that occurs around the time on which most would agree that humans die. The proposal would then be that this change (for example, cessation of electrical activity in the brain, cessation of respiration, cessation of heartbeat, and so on) should be taken to be the legal mark of death. If the proposal were accepted, then medical personnel and undertakers, among others, could appeal to this criterion as a legal defence.

Conceptual analysis would yield a necessary truth about the structure of the concept of death. The search for such analyses fits into a philosophical tradition going back to Plato and Aristotle. It resembles attempts to analyse such concepts as knowledge, causation, goodness and truth, so the analytical project is classically philosophical. The criterial project, if successful, would lead to a contingently useful criterion of human death. If it were accepted by the courts, it might remain in use for decades (until medical technology made it obsolete). But at best it would be a contingent principle with just temporary and practical value. The project itself requires knowledge of medical details and legal precedents, but does not seem to be a project for which philosophers are especially well qualified.

2 Death as separation of body from soul

In the *Phaedo*, Plato has Socrates say:

> Is it [death] not the separation of soul and body? And to be dead is the completion of this; when the soul exists in herself, and is released from the body and the body is released from the soul, what is this but death?
>
> (*Phaedo* [Jowett, 1937 vol. 1]: 447)

These remarks suggest a view about the nature of death:

(D2) *x* dies at $t =_{df.}$ x's soul separates from *x*'s body at *t*

This is problematic, one problem being that (D2) entails that if a thing is to die, it must have a soul. Yet many find it hard to believe that plants (which clearly can die) have souls. Similarly, many find it hard to believe that every living cell has a soul, and yet cells can die.

Many philosophers accept a materialist conception of people. They think that people are material objects - their bodies. Materialists of this sort think that people have no souls. Unless they want to say that people never die, (D2) is clearly unacceptable to such materialists (see Materialism).

3 Death as the cessation of life

The most popular analysis of the concept of death is expressed by (D1). One problem with (D1), however, is its obscurity. There is great controversy about the concept of life. Some claim that to be alive is to be able to engage in life processes, such as nutrition, respiration, and reproduction. Others say that life requires the presence of a soul or some other animating substance. Still others define life by appeal to the notion that living things are able to resist the force of entropy. A number of incompatible accounts of the nature of life have been proposed, but none enjoys universal acceptance. Hence, it is not entirely clear that we know precisely what we mean when we say that something is alive. Since in (D1) death is defined by appeal to the concept of life, (D1) inherits the obscurity of the concept of life.

Furthermore, (D1) seems to be inconsistent with certain empirical facts. One fact concerns suspended animation. Freezing, drying and certain other procedures may be used to arrest the life functions of formerly living entities. Viruses, bacteria and other micro-organisms are placed in suspended animation in laboratories as a matter of course. Sperm, eggs and blastulas of horses, cows and even human beings may be held in this state for months or years. Since all the life functions of such entities have been suspended, it seems that they have ceased to live. But since they can return to life again later, it seems that they have not died. (D1) therefore fails, since it implies that when such organisms go into suspended animation, they die.

In light of this difficulty with suspended animation, it might appear that it would be better to define death not as the mere cessation of life, but as the *permanent* cessation of life:

(D3) x dies at $t =_{df.} x$ ceases permanently to be alive at t.

If an organism goes into suspended animation but will later return to life, (D3) implies that it has not died as its loss of life was not permanent. Thus, (D3) seems to be an improvement over (D1).

Nevertheless, (D3) is still problematic. Suppose two similar organisms go into suspended animation at some time. Suppose one is later brought back to life, whereas the other is not. Then during the period when both were in suspended animation, the first was not dead (since it had not *permanently* ceased to live) but the second was already dead (since it had *permanently* ceased to be alive). This may seem odd, since the two organisms might have been cell-for-cell duplicates during the period of suspended animation, yet according to (D3) one was already dead and the other was not. This shows that (D3) conflicts with the intuitively plausible notion that the life and death of an organism depends upon the intrinsic character of that organism. Thus none of the traditional analyses of the concept of death is definitively correct. The fundamental question about death remains unanswered: we do not know precisely what death is.

4 The mystery of death

A recurrent theme about death in popular thought (as well as in some philosophical circles) is the idea that death is mysterious. As we have seen, it is difficult to formulate a satisfactory philosophical analysis of the concept of death. If it is impossible to analyse the concept of death, then it is impossible to explain precisely what we mean when we say that something dies. It might be said therefore that, in virtue of this fact, death is mysterious. Of course, death is not *distinctively* mysterious - all other unanalysable concepts are equally mysterious in this way.

Yet it is widely thought that there is a special mystery about death. Some seem to take the mystery to be that we cannot know or even conceive of what being dead will be like. This might be thought to follow from the fact that most of us who are living have no recollection of ever having been dead, and thus we lack first-hand experience of what death is like. Furthermore, since there is considerable doubt about the veracity of the testimony of those few who claim to recall having been dead, none of us has a reliable second-hand report of what death is like.

A deeper reason for thinking death difficult to imagine becomes evident if we suppose that death is an ‘experiential blank’ - that is, if we suppose that the dead have absolutely no experiences. It has been claimed that being dead is impossible to conceptualize precisely because being dead is like this. However, while it must be admitted that it is quite difficult to form any clear conception of what an experiential blank would feel like, this may be a bogus problem. Perhaps it is difficult to imagine this feeling not because it is mysterious or hidden, but rather because there is no such feeling. If the dead have no experiences, then it is no wonder and no mystery that we cannot imagine what the experiences of the dead feel like.

5 The evil of death

If we cease to have experiences at death then we cannot experience pain or any other sort of misfortune while dead. How then can death harm us? And if death cannot harm us, how can it be reasonable for us to fear death? Lucretius claimed that it would be more reasonable for us to view death with the same calm indifference with which we view the infinite stretch of time prior to our creation. These are ancient questions: they were discussed by Plato and by Epicurus and his followers, just as they are still discussed today (see Plato §§7, 13; Epicurus §2).

Dualists who believe that our souls continue to live after our bodies die have an easy reply to these questions. They can say that we do not cease to have experiences at death. If we go to hell, we will suffer eternal torment. Thus, death can harm us and (if we have been bad) it is quite reasonable for us to fear death. Our deaths will mark the beginning of the worst and longest period of misery we will ever experience.

The questions about the evil of death are more puzzling for materialists and others who accept the notion that the dead cannot experience pains or other misfortunes. How can such a person explain the evil of death? Some, such as Thomas Nagel, reject the principle that we cannot be harmed by something that we do not experience. They cite the harms of deprivation: suppose a person suffers a brain injury and is reduced to the mental state of a contented infant - they may experience no pain, and may not be aware of any misfortune, yet they have been seriously harmed simply by the deprivation of their mental capacities. According to a more extreme view, a person may be harmed by falling into nonexistence. Consider a girl who dies painlessly in her youth. Suppose that if she had

lived, she would have been quite happy. Her death therefore seems to deprive her of a lifetime of happiness. Some (for example, McMahan or Feldman) see this as a grave harm - though of course a harm of which the victim has no conscious experience. These philosophers claim that it may be reasonable for us to view death as a great evil - even if we will have no experiences while dead. Perhaps the fear of death is not entirely irrational.

6 The survival of death

For a variety of reasons, many are deeply troubled by the question of whether we will survive death. The question is open to several interpretations. Different interpretations arise from different conceptions of the metaphysics of persons, and from different conceptions of death and survival.

Materialists of one traditional sort maintain that people are just physical objects - their own bodies. On this view, there are no souls. People have psychological properties simply because their brains are functioning properly. A materialist of this sort might take the question about survival to be ‘will I (= my body) continue to exist after I die?’. On this interpretation, the person is taken to be the human body, the survival of which is taken to be continued existence. If we interpret the question in this way, the answer is fairly obvious. In the vast majority of cases the human body does continue to exist for at least a few months after it dies. However, since the brain is no longer functioning, it presumably has no experiences. It is difficult to understand how anyone could take this sort of survival to be of any value to the deceased.

Such a materialist might reinterpret the question to mean ‘will I continue to *live* after I die?’. Almost certainly, the answer to this question must be no. If you are just your body, and your body becomes dead and ceases forever to live when it dies, then you become dead and cease forever to live when you die.

Dualists maintain that each person has both a body and a soul. In classical forms of dualism, the soul is taken to be a non-physical object - not made of atoms or molecules. During life, the soul and body are intimately associated. Some would say that during life the soul ‘animates’ the body. Dualists of one tradition take each person to be a unified complex of body and soul. Descartes sometimes seems to endorse this view of persons (‘ I am not lodged in my body merely as a pilot in a ship, but so intimately conjoined, and as it were intermingled with it, that with it I form a unitary whole’ - *Meditations* VI (1958): 239). It would be natural for such a dualist to adopt the Platonic conception and take death to be the dissolution of the soul-body compound. For such a philosopher, the question about survival may be put thus: ‘Will I (= this complex of body and soul) continue to exist after I am dissolved by death and my parts have gone their separate ways?’. Once we have accepted the metaphysical assumptions of this sort of dualism, the answer to this question seems to be no. If the person is the complex of body and soul, and the complex is destroyed by death, then the person is destroyed by death. Though the parts may continue to exist, the compound itself must cease to exist and to live at death (see Descartes, R. §8) .

Dualists of another tradition take persons to be embodied souls. On this view, each living person is a soul that happens to be attached to a body. Descartes seems in other passages to endorse this view of persons (‘ I have a body with which I am very closely conjoined, yet…I am truly distinct from my body, and can exist without it’ - *Meditations* VI (1958): 239). A philosopher who accepts this metaphysical conception of persons might understand the question about survival in this way: ‘Will I (= my soul) continue to have experiences after my body has died?’. Although some have denied it, the correct answer to this question is conceivably yes, and it is somewhat easier to understand why someone who accepts this metaphysical conception of persons might be interested in this sort of survival.

It is interesting to note that on this second dualistic conception, it is not quite clear that *people* actually die. Of course, *human bodies* die. However, on the view in question, no person is a human body. Each living person is just a soul - albeit a soul that happens to be ensconced in a mortal body. That which continues to live after the death of the body is a thing that does not die - the soul. That which dies - the body - typically deteriorates and eventually goes out of existence after death (see Dualism §1).

7 Immortality

We might think that the concept of immortality is the concept of never dying. However, this would be slightly misleading since it would imply that non-biological objects such as rocks and bricks, atoms and planets, numbers and properties, are all immortal: since they never live, they never die. It would also be misleading to suppose that

the concept of immortality is the concept of living at some time but never dying, since this would also have odd implications. Consider a living thing that goes into eternal suspended animation. It never dies, but this sort of immortality is hardly better than death. It is better, therefore, to take the concept of immortality to be the concept of living forever. This seems more interesting, and is probably closer to the concept that has been discussed.

Some say that eternal life would be a great blessing - something of unsurpassable value. Perhaps they reason as follows: life is good; more of a good thing is always better than less; therefore eternal life is exceedingly good. But others (most notably Bernard Williams) have argued that eternal life could not possibly be desirable. Williams considers several possible ‘models’ of eternal life. On one, a person (identified as ‘E.M.’) remains eternally at the same biological age (in his example, the age is forty-two). Williams claims that after a few hundred years, this person would inevitably become bored with life. The boredom is inevitable, he insists, in virtue of the fact that ‘everything that could happen and make sense to one particular human being of 42 had already happened to her’ (1973: 90).

On another model we imagine that E.M. does not remain at a constant biological age of forty-two, and does not retain a certain character. Rather, we imagine that E.M. lives out an endless succession of different lives, each with a new character and personality. In this way, E.M. avoids the tedium of the first model. Williams suggests that this is not a model on which a single person lives forever. Rather, it is a model on which a series of distinct persons live. Thus, it has no bearing on the question of the desirability of eternal life.

It could be argued, however, that Williams has neglected certain important possibilities. One is the possibility that there are activities whose pleasure does not decrease with repetition. Another is that people may grow and change gradually over time, and thereby come to have new interests without losing their identities as individuals. Such growth might make it possible for a single individual to live forever without falling into the tedium that might result from steadfast pursuit of just one interest over an endless stretch of time.

8 Death and the meaning of life

A number of thinkers have suggested that there is some important connection between death and ‘the meaning of life’. A person who thinks that death is not followed by any sort of afterlife may think that death makes life meaningless. Schopenhauer sometimes seems to have expressed this view (see Schopenhauer, A. §6). Others who believe in God and immortality may see it this way: God placed us here on earth in order that we may either sin or achieve our salvation. If we sin, we are punished with eternal damnation. If we achieve salvation, we are rewarded with eternal bliss. A person who accepts this picture might say that if there were no God and no afterlife in which to receive reward or punishment, then life would be (to quote Shakespeare) ‘a tale told by an idiot, full of sound and fury, signifying nothing’. In other words, if there were no God and no afterlife, our lives would not be components of a larger and purposeful scheme - we would live for a while and then simply die. For many, if it were like this, life would be meaningless. Those who believe in an afterlife may take comfort in thoughts of this afterlife, and think that its existence serves to make ordinary life here on earth meaningful.

Yet it appears that sense can be made of the idea that life is meaningful even if it ends in death. If people have worthwhile goals and exert themselves to achieve these goals, taking some pleasure in both the exertion and the achievement, then their lives may be said to be meaningful - at least in what Paul Edwards calls ‘the terrestrial sense’. Death, of course, may bring an end to such meaningfulness, but the fact that they will someday die seems not to be able to rob people’s lives of this sort of meaningfulness while they are alive.

According to an even more extreme view, life is made *more* meaningful by the recognition that it will end with death. According to this view, we gain a deeper appreciation for the common satisfactions of our everyday experience when we fully realize that someday we will die, and will then have nothing at all.

See also: Life and death; Life, meaning of; Soul, nature and immortality of

FRED FELDMAN

References and further reading

Descartes, R. (1641) *Meditations on First Philosophy*, in N. Kemp Smith (ed. and trans.) *Descartes: Philosophical Writings*, New York: Random House, 1958.(Descartes’ presentation of his dualistic conception of persons.)

Donnelly, J. (ed.) (1994) *Language, Metaphysics, and Death*, 2nd edn, New York: Fordham University Press. (Useful and well-organized collection of essays on the nature and meaning of death, the nature of the soul, and immortality. Includes papers by Nagel, Williams, Rosenbaum, Edwards, Feldman and others.)

Edwards, P. (1967) 'Life, Meaning and Value of', in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York and London: Macmillan Publishing Company & The Free Press. (Clear exposition of main views on meaning of life. Explains and criticizes Schopenhauer's arguments for pessimism.)

Edwards, P. (ed.) (1992) *Immortality*, New York: Macmillan Publishing Company. (Anthology of thirty-four readings on immortality, including selections from Plato, Lucretius, Aquinas, Descartes, Hume, Reid, Mill, Broad and others. Extensive, detailed annotated bibliography.)

Epicurus (*c*.300 BC) 'Letter to Menoecceus', trans. C. Bailey in W. J. Oates (ed.) *The Stoic and Epicurean Philosophers*, New York: The Modern Library, 1940, 30-4.(Classic statement of the view that since we will not exist once we are dead, death cannot harm us. Claims that the fear of death is wholly irrational.)

Feldman, F. (1992) *Confrontations with the Reaper: A Philosophical Study of the Nature and Value of Death*, New York: Oxford University Press.(Defends a materialist conceptual scheme for death and associated concepts; argues for a version of the deprivation approach to the evil of death; presents a modified utilitarian theory about the wrongness of killing.)

Fischer, J.M. (1993) *The Metaphysics of Death*, Stanford, CA: Stanford University Press. (Clear and well-organized introduction by Fischer, followed by sixteen papers on death. Includes papers by Nagel, Williams, Parfit, McMahan, Feldman and others. Excellent bibliography.)

Kamm, F.M. (1993) *Morality, Mortality, Volume 1: Death and Whom to Save From It*, Oxford: Oxford University Press.(Extensive discussion of the evil of death, focusing on the question why non-existence after death seems so much worse than non-existence before birth. Also discusses moral questions concerning the allocation of transplantable bodily organs.)

Nagel, T. (1973) 'Death' in *Mortal Questions*, Cambridge: Cambridge University Press, 1-10.(Discusses whether it is a bad thing to die. See §5 above.)

Plato (*c*.330 BC) *Phaedo*, trans. B. Jowett in *The Dialogues of Plato*, New York: Random House, 1937. (Introduction by R. Demos. Represents a dialogue about death, the nature of the person, and the role of the philosopher. Socrates is the leading character in the discussion, which takes place in his cell in jail, on the day of his death.)

President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research (1981) *Defining Death: Medical, Legal and Ethical Issues in the Determination of Death*, Washington, DC: President's Commission.(Report of U.S. Presidential Commission directed to investigate ethical and legal implications of proposed and enacted legal criteria of death. Contains a review of actual statutes as well as thoughtful discussion of a proposed new uniform statute.)

Williams, B. (1973) *Problems of the Self*, New York: Cambridge University Press. (Contains 'The Makropoulos Case: Reflections on the Tedium of Immortality', and other essays by Williams.)

Deconstruction

Although the term is often used interchangeably (and loosely) alongside others like 'post-structuralism' and 'postmodernism', deconstruction differs from these other movements. Unlike post-structuralism, its sources lie squarely within the tradition of Western philosophical debate about truth, knowledge, logic, language and representation. Where post-structuralism follows the linguist Saussure - or its own version of Saussure - in espousing a radically conventionalist (hence sceptical and relativist) approach to these issues, deconstruction pursues a more complex and critical path, examining the texts of philosophy with an eye to their various blindspots and contradictions. Where postmodernism blithely declares an end to the typecast 'Enlightenment' or 'modernist' project of truth-seeking rational enquiry, deconstruction preserves the critical spirit of Enlightenment thought while questioning its more dogmatic or complacent habits of belief. It does so primarily through the close reading of philosophical and other texts and by drawing attention to the moments of 'aporia' (unresolved tension or conflict) that tend to be ignored by mainstream exegetes. Yet this is not to say (as its detractors often do) that deconstruction is a kind of all-licensing textualist 'freeplay' which abandons every last standard of interpretive fidelity, rigour or truth. At any rate it is a charge that finds no warrant in the writings of those - Jacques Derrida and Paul de Man chief among them - whose work is discussed below.

1 Is there a logic of deconstruction?

What is 'deconstruction' and what does it do? There is a problem here in that Derrida often goes out of his way to disown any summary treatment of the topic or any attempt to define (and so delimit) deconstruction in 'adequate' conceptual terms. Nevertheless one can offer at least some attempt at a working definition. What typically occurs in a deconstructive reading is that the text in question is shown to harbour contradictory logics which are standardly ignored - or concealed from view - on other, more orthodox accounts. Very often it is a matter of locating certain clearly-marked binary oppositions (as for instance between nature and culture, speech and writing, concept and metaphor, or philosophy and literature) and showing that their order of priority is by no means as stable or secure as the text seeks to maintain. That is to say, there is a counter-logic at work whereby those distinctions can be shown to break down, or to generate a reading markedly at odds with the author's overt intent. Not that intention is simply ruled out as irrelevant for the purposes of a deconstructive reading. On the contrary, it offers an 'indispensable guardrail' (Derrida [1967b] 1975: 158) which saves interpretation from running wild in endless subtleties of its own ingenious devising. However, this leaves open the possibility that texts may mean something other - and more - than is allowed for by any straightforward appeal to the warrant of authorial intention.

Thus in each of the above-mentioned cases one term (the first) is conventionally assigned a positive or superior value while the other is construed - by author and mainstream interpreters alike - as self-evidently standing in a dependent, derivative or supplementary relation to it. For Rousseau, nature both precedes culture and represents a better, more authentic way of life, while writing is regarded as a bad 'supplement' to speech, one that involves all manner of corrupting artifice and which infects the very sources of spontaneous 'natural' expression. Yet he cannot describe that idealized 'state of nature' except in terms that perforce acknowledge the *impossibility* of any such state having ever existed. For there is clearly a logical or conceptual problem about an argument that sets up this rigid opposition between nature and culture, but which then proceeds to identify the best (that is, the most primitive, least 'civilized') cultures with an early stage of social evolution that would somehow *not yet* have embarked upon the path to civilization and ruin. The case with speech and writing is somewhat more complex but can be stated along much the same lines (see Rousseau, J. §2).

For Rousseau, speech - in its 'natural' condition - is a mode of utterance that expresses sentiments so directly (with so little room for civilized pretence) that it allows human beings to communicate face-to-face without any need for the kinds of elaborate convention that characterize other, more articulate forms of linguistic exchange. What is bad about writing is precisely its dependence on a graphic notation which involves such artifice as the very condition of its possibility, and which thus stands at the furthest remove from the passional origins of speech. However this argument fails to take account of the conventional character of *all* language, spoken and written alike; that is to say, the extent to which speech partakes of various signifying systems (phonetic, semantic, syntactic and so on) in the absence of which it would simply not qualify as language. Furthermore it is just this reliance on conventions -

a matter of structural necessity - that Rousseau has to recognize (against his own wishes) as an intrinsic condition of language-in-general. For it would then be the case that writing (not speech) might offer the best model for explaining those constitutive features of language - its structural properties, conventional character, capacity to signify in various contexts irrespective of self-present utterer's intentions, and so forth - which necessarily pertain to *all* language, whether written or spoken (Derrida 1989).

There is a parallel here with Wittgenstein's argument against the possibility of a private language, that is to say, his contention that any such 'language' would not in fact constitute a language since it would lack the criteria for shared understanding and hence be wholly unintelligible even to its solitary user (see Private language argument; Wittgenstein, L.). What is distinctive about Derrida's approach to such issues is also what sets deconstruction apart from most work in the 'other' (analytic or Anglo-American) tradition. That is, he treats them as issues specific to the reading of certain problematical passages in certain texts, passages whose overt or manifest sense is shown to be at odds with the structural logic that governs the argument as a whole. Hence Derrida's claim with regard to Rousseau that 'his declared intention is not annulled by this but rather *inscribed* within a system which it no longer dominates' (Derrida [1967b] 1975: 243). However this emphasis on textual close-reading goes along with a keen analytical grasp of the philosophic issues involved and should not, therefore, be taken as evidence that Derrida is practising literary criticism (or a mode of 'merely' rhetorical exegesis) under the guise of linguistic philosophy. Rather he argues from the *conditions of possibility* for raising such questions about meaning, context, authorial intention and so forth, questions that clearly point beyond the sphere of localized interpretative insight or thematic commentary. Thus he is not so much concerned with particular instances of paradox, aporia or self-contradiction in Rousseau's texts. What chiefly engages Derrida's interest is that pervasive 'logic of supplementarity' whereby the second (supposedly inferior or derivative) term in each pair turns out to be always already presupposed in any definition of the first (supposedly original and self-sufficient) term. This logic may resist formalized statement through its complexities of modal structure - which Derrida brings out to striking effect in his reading of Rousseau - and its apparent suspension of classical axioms like bivalence or non-contradiction. But this should hardly place his work beyond the pale for philosophers with a knowledge of developments in recent post-classical (for example, intuitionist or many-valued) logic.

2 Paul de Man and 'aesthetic ideology'

Paul De Man is another thinker whose writing occupies the currently contested zone between philosophy and literary theory. In de Man's case also, the suspicion is often voiced that his practice of minute close-reading - especially when applied to philosophic texts - must go along with a lack of genuine analytic rigour or merely a desire to make trouble by inventing all manner of pseudo-problems. I should not wish to claim that this characterization is entirely wide of the mark, at least as regards de Man's more tortuous attempts to outdo Nietzsche in deconstructing all the concepts and categories of logic, epistemology, causal explanation and other 'metaphysical' notions (de Man 1979). All too often these claims have been taken on trust or treated as a deconstructive *fait accompli* by literary theorists keen to upstage their colleagues in philosophy or the natural sciences. At its best, however, de Man's work cannot be dismissed as simply a product of the turf-battle between rival academic disciplines. In late essays like 'The Epistemology of Metaphor' (1978) on Locke, Condillac and Kant, and 'Pascal's Allegory of Persuasion' (1981) on the relation between suasive or performative language and the language of logical proof-procedures, his thought exhibits an exceptional degree of analytic acuity and rigour. Certainly he is far from advancing the sorts of blanket 'textualist' claim - such as 'all concepts are metaphors', 'all philosophy is a kind of literature', 'all talk of "truth" is just a species of rhetorical imposture', and so forth - that are often attributed to him (and to Derrida) by disciples and opponents alike.

Above all, de Man insists that textual exegesis in the deconstructive mode can at times produce readings sharply at odds with established or mainstream critical opinion. This produces an extreme - some would say perverse - attentiveness to just those kinds of aberrant textual detail (tropological swerves, displacements, substitutions or elisions) which complicate the reading-process beyond any notion of straightforward fidelity to the author's presumed intent. Of course this description might also apply to that mode of 'rational reconstruction' by which some philosophers have lately seen fit to treat the texts of past thinkers with a view to their own, mainly analytic concerns. However it is the aim of this approach to discover some viable (constructive) solution to problems whose importance is taken to justify the strong-revisionist line. For de Man, conversely, it is the purpose of a deconstructive reading to respect the very letter of the text even where this leads to a stage of aporetic deadlock

where the text turns out to contest or subvert its own statements of intent.

Hence his insistence on the stubborn *literality* of deconstruction, that is to say, its power to hold out against the seductive ('eudaimonic') appeal of other, more compliant and therefore - he would argue - less rigorous modes of understanding. For de Man much depends on the way that we read certain passages in the work of philosophers and literary theorists who have reflected - in a more or less self-conscious or critical mode - on these issues of language, knowledge and representation. What he often discovers is an alternating rhythm of co-implicated 'blindness' and 'insight', a rhythm characterized by moments of apparent naïvety (for example, with regard to the supposed organic or consubstantial relation between language and nature) which none the less exist in close proximity to moments of un-self-deceiving rigour (de Man 1983). His aim is to expose the workings of a deep-laid 'aesthetic ideology' whose effect - so he argues - is chiefly manifest in the discourse of post-Kantian philosophy and literary theory. Most often it results from a confusion between linguistic (especially metaphorical) structures and those forms of phenomenal or sensory perception that characterize our knowledge of natural-world objects and processes. Thus : 'literature is fiction not because it somehow refuses to acknowledge "reality", but because it is not *a priori* certain that language functions according to principles which are those, or which are *like* those, of the phenomenal world' (de Man 1986: 11). To suppose otherwise is to fall into a way of thinking that can all too easily be co-opted for the purposes of ideological mystification.

Such is, for instance, the Cratylist delusion - revived by Romantic poet-philosophers like Schiller and Coleridge - according to which it is possible for language (especially the language of symbol and metaphor) to achieve a condition of organic form beyond all the Kantian antinomies of subject and object, mind and nature, or intuitive and conceptual knowledge. Aesthetic ideology can thus be seen as the source of some far-reaching errors, among them - most crucially for de Man - the idea that particular (national) languages and cultures can likewise manifest the kind of predestined, organic development and growth that grants them an inherent superiority over other, more fractured or dissociated forms of cultural-linguistic expression. De Man's notorious wartime journalism, published in Nazi-occupied Belgium, often attributed just such a natural, predestined eminence to German thought and culture. It seems plausible to claim that his later writings were concerned in various ways to challenge or resist what he subsequently saw as the seductive power of an aesthetic creed with large implications for our thinking about issues of culture, politics and moral responsibility. Thus: 'what we call ideology is precisely the confusion of linguistic with natural reality, of reference with phenomenalism'. And again, more pointedly, those who charge deconstruction with a culpable indifference to issues of history and politics 'are merely stating their fear at having their own ideological mystifications exposed by the tool they are trying to discredit. They are, in short, very poor readers of Marx's *German Ideology*' (de Man 1986: 11).

3 Some analytic bearings

Gilbert Ryle lost patience with the so-called 'psychologism' of Husserl, despite having previously published some detailed and constructive (albeit critical) studies of his work (Ryle 1971 - see Ryle, G. §1) . So it was - or so the story is often told - that a rift emerged between phenomenology (along with various derivative 'continental' schools of thought) and later developments in the Anglo-American analytic tradition. This view fitted in with the received idea - initiated by Russell and Frege, and erected into a full-scale programme by logical empiricists like Rudolf Carnap - that philosophy should work to clear away those errors and confusions of natural language that could best be revealed by analysis in the formal or logico-semantic mode (see Analytical philosophy §2). It would also help to resist the kinds of metaphysical bewitchment by language to be seen at their worst (so Carnap thought) in Heidegger's obscurantist appeal to the primordial wisdom supposedly enshrined in Greek and German etymology. Derrida's work has indeed been much influenced by Heidegger, especially as regards its questioning of those various deep-laid 'logocentric' priorities and values whose presence in the Western philosophical tradition is a constant theme of his writing, early and late. However he has also criticicized some aspects of Heidegger's thought, among them precisely the harking-back to a lost originary plenitude of Being (Derrida 1967b). For it is here that Heidegger himself falls prey to a potent irrationalist mystique, an appeal to language - or to certain privileged *national* languages and cultures - as the source of a wisdom supposedly concealed throughout the history of so-called 'Western metaphysics' (see Heidegger, M. §6).

In de Man's work also, Heidegger figures as a prime example of that alternating pattern of 'blindness' and 'insight' which in turn can enable the reader to attain a less prejudiced (if never fully adequate, transparent or

demystified) level of understanding (de Man 1983). Such arguments will scarcely be convincing to those for whom ‘deconstruction’ is pretty much synonymous with the obscurantist ‘jargon of authenticity’ that Heidegger promoted and that Derrida has supplied with its latest fashionable twist. However this notion cannot stand up to the sort of close reading which Derrida brings to his commentary on Heidegger and others, and which his own texts properly require on the part of responsible critics. Here again what has most often got in its way is the fixed preconception - handed down from the logical empiricists - that analytic (that is, genuine or serious) thought should have no truck with such extravagant forms of ‘continental’ pseudo-philosophy.

If the prospects look fair for a mending of this rift, then one clear sign is the current reappraisal (by Michael Dummett (1993) among others) of the strong continuities that exist within and between these two lines of descent. Another is the willingness of some commentators to look again, for example, at Derrida’s reading of J.L. Austin - the target of a well-known hostile response by John Searle (1977) - and to discover not only points of genuine philosophical interest but also a certain affinity between the thought of Austin and Derrida (Austin 1961, Derrida 1989). This emerges, for instance, in their shared alertness to the ways that language (so-called ‘ordinary’ language) can prove more complex and resistant to codification than anything allowed for by speech-act theory in its systematized form. If ‘undecidability’ is in question here - as so often in Derrida’s work - then it is the upshot of a reading closely attentive to the detail and the logic of Austin’s arguments.

See also: Post-structuralism

CHRISTOPHER NORRIS

References and further reading

Austin, J.L. (1961) *How to Do Things With Words*, London: Oxford University Press.(A classic exposition of speech-act theory, ‘deconstructed’ with relish - though by no means attacked or dismissed - by Derrida in *Limited Inc.*)

de Man, P. (1978) ‘The Epistemology of Metaphor’, *Critical Inquiry* 5 (1): 13-30. (On metaphor in the texts of Locke, Condillac and Kant.)

de Man, P. (1979) *Allegories of Reading: figural language in Rousseau, Nietzsche, Rilke, and Proust*, New Haven, CT: Yale University Press.(Essays in his typical late mode of deconstructive textual exegesis.)

de Man, P. (1981) ‘Pascal’s Allegory of Persuasion’, in S.J. Greenblatt (ed.) *Allegory and Representation*, Baltimore, MD: Johns Hopkins University Press, 1-25.(Analyses the relationship between constative and performative language in Pascal’s writings on geometry, mathematics and religious faith.)

de Man, P. (1983) *Blindness and Insight: essays in the rhetoric of contemporary criticism*, London: Methuen, 2nd edn.(Examines the blind-spots of presupposition and ideological prejudice in various twentieth-century critics and philosophers.)

de Man, P. (1986) *The Resistance to Theory*, Minneapolis, MN: University of Minnesota Press.(On ‘aesthetic ideology’ and the resistance to it through a deconstructive reading of various - mainly literary-critical - texts.)

Derrida, J. (1967a) *La Voix et le Phénomène*; trans. D. Allison as *Speech and Phenomena and Other Essays on Husserl’s Theory of Signs*, Evanston, IL: Northwestern University Press, 1973. (Represents Derrida’s early, intensive, and critical engagement with Husserl.)

Derrida, J. (1967b) *De la Grammatologie*; trans. G. Spivak as *Of Grammatology*, Baltimore, MD: Johns Hopkins University Press, 1975.(On Rousseau, speech/writing, nature/culture, and other key themes of Derrida’s early work.)

Derrida, J. (1967c) *L’Écriture et la Différence*; trans. A. Bass as *Writing and Difference*, Chicago, IL: The University of Chicago Press, 1978.(Essays on Husserl, Levinas, Foucault, structuralism and the human sciences.)

Derrida, J. (1972a) *Marges de la Philosophie*; trans. A. Bass as *Margins of Philosophy*, Chicago, IL: University of Chicago Press, 1982.(Contains some of Derrida’s most important and powerfully-argued essays, among them ‘White Mythology: metaphor in the text of philosophy’.)

Derrida, J. (1972b) *La Dissémination*; trans. B. Johnson as *Dissemination*, London: Athlone Press, 1981. (‘Plato’s Pharmacy’ and other essays in the border-zone between philosophy and literature.)

Derrida, J. (1989) *Limited Inc*, ed. G. Graff, Evanston, IL: Northwestern University Press. (Includes Derrida’s essay on Austin; also his response to Searle (1977) and further reflections on speech-act philosophy.)

Dummett, M. (1993) *On the Origins of Analytic Philosophy*, Cambridge, MA: Harvard University Press. (Comparative/contrastive account of work in the two traditions, with reference mainly to Frege and Husserl.)

Norris, C. (1987) *Derrida*, London: Fontana.(Introductory work for students of philosophy and literary theory.)

Norris, C. (1988) *Paul de Man and the Critique of Aesthetic Ideology*, London: Routledge. (Critical survey of de Man's writings; develops some of the arguments outlined above.)

Ryle, G. (1971) *Collected Papers*, vol. 1, London: Hutchinson.(Includes Ryle's early review-articles on Husserl and Heidegger.)

Searle, J.R. (1977) 'Reiterating the Differences', *Glyph*, vol. 1, Baltimore, MD: Johns Hopkins University Press. (Highly critical response to Derrida on the topic of Austinian speech-act philosophy.)

Dedekind, Julius Wilhelm Richard (1831-1916)

Dedekind is known chiefly, among philosophers, for contributions to the foundations of the arithmetic of the real and the natural numbers. These made available for the first time a systematic and explicit way, starting from very general notions (which Dedekind himself regarded as belonging to logic), to ground the differential and integral calculus without appeal to geometric 'intuition'. This work also forms a pioneering contribution to set theory (further advanced in Dedekind's correspondence with Georg Cantor) and to the general notion of a 'mathematical structure'.

Dedekind's foundational work had a close connection with his advancement of substantive mathematical knowledge, particularly in the theories of algebraic numbers and algebraic functions. His achievements in these fields make him one of the greatest mathematicians of the nineteenth century.

1 Career

Richard Dedekind was born in Braunschweig, Germany. In 1848 he entered the Collegium Carolinum in that city to study 'mathematics and science', but it was for mathematics that he had already decided. In 1850 he matriculated at the University of Göttingen, and two years later received his doctorate, under the sponsorship of the great C.F. Gauss. In 1854 he 'habilitated' (qualified as a private lecturer) at Göttingen, a few weeks after Bernhard Riemann, who had been a fellow student and who was one of the two greatest influences upon Dedekind's mathematical - and philosophical - career. The second was J.P.G. Lejeune Dirichlet, who succeeded Gauss on his death in 1855. On Dedekind's later testimony, Dirichlet had 'inaugurated a new era for the study of mathematics at Göttingen', and had 'made a new man' of him.

In external details, the story of Dedekind's career is brief: he became a professor at the Polytechnic Institute of Zurich in 1858, then, in 1862, accepted a professorship at the Collegium Carolinum (itself by then a Technical Institute). This position he held until his retirement in 1894, having several times declined offers of university appointments (preferring to remain in his home town and near his ancestral family). After formally retiring he continued to offer occasional courses of lectures until shortly before his death.

Dedekind's style of thought in general was to ponder slowly over the subjects that engaged him; he tended not to publish until he had brought a piece of work to a degree of classic perfection and could display for his readers the heart of the matter with a maximum of clarity. (A word Dedekind uses to characterize this quality he valued in mathematical exposition is 'quietness'. The motto of his work might well have been that explicitly invoked by Gauss: *pauca sed matura* (few, but ripe).)

2 Algebraic numbers and algebraic functions; new conception of mathematical 'objects'

A considerable part of Dedekind's energy during the early years of his career was devoted to the posthumous publication of works by his friends: Dirichlet's lectures on number theory; then Riemann's papers, both published and unpublished.

It was in the elaboration of themes arising from Dirichlet's number-theoretic lectures that Dedekind found the central theme of his own great contributions to mathematics - in the first place, a very far-reaching generalization (and clearer grounding) of the then new theory of algebraic numbers. Dedekind's major publication on this subject took the modest form of a supplement, among a whole series of supplements designed to clarify points arising from the main text, in his edition of the Dirichlet lectures. This work first appeared in the second edition (1871) as 'Über die Komposition der binären quadratischen Formen' ('On the Composition of Binary Quadratic Forms'). It was revised in subsequent editions (1879, 1894), appearing as 'Über die Theorie der ganzen algebraischen Zahlen' ('On the Theory of Integral Algebraic Numbers').

Characteristic of this work from the first was the introduction of a series of notions that have come to form the basis of modern algebra - among them, those of 'field', 'module', 'ideal' (and 'algebraic integer' itself) - on the basis of what we would now describe as structural axiomatic characterizations: thus, for instance, Dedekind defines not only a particular field, but the conditions to be satisfied for any system to be a field. Thus Dedekind effected a move towards regarding mathematics itself as concerned with species of structure explicitly

characterized, rather than with particular kinds of ‘objects’, grasped by ‘intuition’. (In this one may see the influence of Riemann, whose characterization of geometry was quite the same.) It was also characteristic that Dedekind made extensive use of sets (‘systems’ in his terminology) as ‘mathematical objects’ on which operations might be performed; and that he introduced as fundamental the consideration of mappings, both within and between particular structures.

The generality of Dedekind’s formulations in this work opened the way for another significant development: his establishment (in collaboration with Heinrich Weber) of a rigorous foundation for the theory, initiated by Riemann, of algebraic functions of one variable (‘compact Riemann surfaces’), bringing to light a deep algebraic analogy between this theory and that of algebraic integers. The resulting paper (1882) has become a cornerstone of twentieth-century algebraic geometry.

3 Real and natural numbers; their basis in logic

Dedekind’s theory of real numbers and continuity, published in 1872, was motivated, he tells us, by his first obligation to teach the differential calculus (in 1858), and his dissatisfaction with the ‘geometrical’ reasoning upon which one traditionally relied. Beginning from the system of rational numbers, Dedekind remarks that this ordered system contains what we may call ‘gaps’, in the following sense: there exist certain splittings of the system into two non-empty, exhaustive parts, with each element of the one less than each element of the other, such that the ‘lower’ part contains no greatest element and the ‘upper’ part no least (for example, negative numbers and those positive numbers with squares less than 2, on the one hand; and positive numbers with squares greater than 2, on the other). He calls a system *continuous* if such gaps never occur; remarks that, whereas ordinary geometry tacitly assumes continuity for the system of points on a line, this is by no means a conceptually necessary property of space - it should rather be postulated explicitly when desired; and shows how, for a non-continuous system such as that of the rational numbers, one can construct a continuous extension by introducing a new element for each ‘gap’, so placed in the ordering as to ‘fill’ the gap. Most crucially, Dedekind (1) proves that the new system so constructed automatically fulfils the requirement of continuity, and (2) defines, in the numerical case, arithmetic operations for the new (‘irrational’) numbers (and between them and the rationals), and demonstrates that all the properties needed for traditional mathematics hold.

Dedekind’s work on the natural numbers was done between 1872 and 1878, but published only in 1888. It is based on the definition of a ‘simply infinite system’: a set *N*, with a one-to-one mapping ϕ of *N* into itself, and an element (called ‘1’) in *N* that is not taken on as a value by ϕ, such that every element in *N* belongs to what Dedekind calls the ‘chain’ of 1, that is, the intersection of all sets that contain 1 and are ‘closed under ϕ’ - that is, contain, together with any element *x*, its image $\phi(x)$. Dedekind shows how this definition justifies the principle of mathematical induction for any simply infinite system (see Logicism §3), and how it also justifies the powerful technique of recursive definition. According to Dedekind, ‘the system of the natural numbers’ is just any simply infinite system, whose elements are regarded without respect to anything but their position in that system. This characterization (from which Giuseppe Peano directly borrowed his axioms for arithmetic) has been crucial for subsequent foundational work in mathematics (see Arithmetic, philosophical issues in §4).

Besides his published work, Dedekind made significant contributions, in his correspondence with Georg Cantor (1937), to set theory and to point-set topology, which the latter was then developing.

See also: Cantor, G.; Kronecker, L.; Logical and mathematical terms, glossary of;Logicism

HOWARD STEIN

List of Works

Dedekind, R. (1930-2) *Gesammelte mathematische Werke (Collected Mathematical Works)*, ed. R. Fricke, E. Noether and Ö. Ore, Braunschweig: Vieweg, 3 vols.(Dedekind’s collected works. Volume 3 reproduces several of the works given here.)

Dedekind, R. (1871) ‘Über die Komposition der binären quadratischen Formen’ (‘On the Composition of Binary Quadratic Forms’), in P.G. Lejeune Dirichlet, *Vorlesungen über Zahlentheorie (Lectures on Number Theory)*, ed. R. Dedekind, Braunschweig: Vieweg, 2nd edn, supplement 10; passages significantly altered in subsequent edns are repr. in Dedekind (1932), vol. 3, 223-61.(The first version of Dedekind’s fundamental work on the theory of algebraic integers. Later versions were published in 1879 and 1894 (see below).)

Dedekind, R. (1872) *Stetigkeit und Irrationalzahlen*, repr. in Dedekind (1932), vol. 3, 315-34; trans. W.W. Beman, ‘Continuity and Irrational Numbers’, in *Essays on the Theory of Numbers*, New York: Dover, 1963. (His classic work on the construction of the real numbers and the notion of ‘continuity’ or ‘Dedekind completeness’.)

Dedekind, R. (1879, 1894) ‘Über die Theorie der ganzen algebraischen Zahlen’ (‘On the Theory of Integral Algebraic Numbers’), in P.G. Lejeune Dirichlet, *Vorlesungen über Zahlentheorie (Lectures on Number Theory)*, ed. R. Dedekind, Braunschweig: Vieweg, 3rd and 4th edns, supplement 11; repr. in Dedekind (1932), vol. 3, 1-222, 297-313; 4th edn repr. New York: Chelsea, 1968.(The later versions - each very significantly reworked - of the theory first published in 1871.)

Dedekind, R. and Weber, H. (1882) ‘Theorie der algebraischen Funktionen einer Veränderlichen’ (‘Theory of the Algebraic Functions of One Variable’), *Journal für reine und angewandte Mathematik* 92: 181-290; repr. in Dedekind (1930), vol. 1, 238-349.(A fundamental contribution to algebraic geometry.)

Dedekind, R. (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; repr. in Dedekind (1932), vol. 3, 335-90; trans. W.W. Beman, ‘The Nature and Meaning of Numbers’, in *Essays on the Theory of Numbers*, New York: Dover, 1963; trans. *What are Numbers and What Should They Be?*, Orono, ME: Research Institute for Mathematics, 1995.(His classic work on the foundations of the theory of the natural numbers. Characterized by Dedekind’s usual clarity, but a little laborious in its working-out of basic set-theoretic relations. The title is somewhat misconstrued in translation as it should more properly be rendered as ‘What are Numbers and What are They For?’ or ‘The Nature and Function of Numbers’ - see Stein (1988: 246-7).)

Dedekind, R. (1890) *Letter to H. Keferstein*, trans. H. Wang and S. Bauer-Mengelberg, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 99-103.(A masterly presentation, in a short space, of the basic ideas of Dedekind (1888). A very useful guide for the beginner.)

Dedekind, R. and Cantor, G. (1937) *Briefwechsel Cantor-Dedekind (Cantor-Dedekind Correspondence)*, ed. E. Noether and J. Cavaillès, Paris: Hermann; French trans. in J. Cavaillès, *Philosophie mathématique*, Paris: Hermann, 1962.(Includes most of the correspondence of Dedekind and Georg Cantor. The surviving portion of the correspondence not included in this collection is published in Dugac (1976).)

References and further reading

Dugac, P. (1976) *Richard Dedekind et les fondements des mathématiques (Richard Dedekind and the Foundations of Mathematics)*, Paris: Vrin.(Includes biographical material, commentary, otherwise unpublished correspondence, preliminary notes for the works of 1872 and 1888, and manuscript material.)

Landau, E. (1917) ‘Richard Dedekind’, *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen. Geschäftliche Mitteilungen* 50-70; repr. in *Collected Works*, ed. P.T. Bateman *et al*., Essen: Thales, 449-65.(A biographical eulogy.)

Stein, H. (1988) ‘Logos, Logic, and Logistiké: Some Philosophical Remarks on the Nineteenth-Century Transformation of Mathematics’, in W. Aspray and P. Kitcher (eds) *History and Philosophy of Modern Mathematics*, Minneapolis, MN: University of Minnesota Press, 238-59.(Discusses the work of Dedekind in relation to the general development of mathematics in his century, and especially his relation to Dirichlet, Riemann, Kronecker and Hilbert.)

Deductive closure principle

It seems that one can expand one's body of knowledge by making deductive inferences from propositions one knows. The 'deductive closure principle' captures this idea: if S knows that P, and S correctly deduces Q from P, then S knows that Q. A closely related principle is that knowledge is closed under known logical implication: if S knows that P and S knows that P logically implies Q, then S knows that Q. These principles, if they hold, are guaranteed by general features of the concept of knowledge. They would form part of a logic of knowledge.

An influential argument for scepticism about knowledge of the external world employs the deductive closure principle. The sceptic begins by sketching a logically possible hypothesis, or counter-possibility (for example, that one is a brain in a vat, with computer-induced sense experience) which is logically incompatible with various things one claims to know (such as that one has hands). The proposition that one has hands logically implies the falsity of the sceptical hypothesis. Supposing that one is aware of this implication, the deductive closure principle yields the consequence that if one knows that one has hands, then one knows that one is not a brain in a vat. The sceptic argues that one does not know this: if one were in a vat, then one would have just the sensory evidence one actually has. It follows that one does not *know that one has hands. Some philosophers have sought to block this argument by denying the deductive closure principle.*

1 Closure principles

According to the 'deductive closure principle', if S knows that *P*, and S correctly deduces *Q* from *P*, then S knows that *Q* (call this principle 'Closure'). The set *X* of propositions which one knows is not closed (in the mathematical sense) under logical implication *simpliciter*. This is because there are countless propositions which are logical consequences of what one knows which one does not recognize as such. Since one does not believe these logically implied propositions, one does not know them, and so they are not members of *X*. A similar point holds for the closure principle for known logical implication: it seems possible for one to know that *P*, know that *P* implies *Q*, and yet fail to deduce *Q*. According to Closure, those logical consequences which one deduces from what one knows are members of *X*. Such deduced consequences meet two of the conditions for knowledge: they are true (since they follow from known, and therefore true, propositions), and they are believed. If these deduced consequences are to be known, as Closure says, then they must satisfy whatever further necessary conditions for knowledge there might be.

Suppose that *having justification for believing that P* is a necessary condition for knowing that *P* (see Knowledge, concept of §2; Justification, epistemic). Then it is plausible to suppose that Closure will hold only if *closure for justification* holds: if S has justification for believing that *P*, and S correctly deduces *Q* from *P*, then S has justification for believing that *Q*. This principle is used in constructing Gettier examples in which one starts with a justified belief of a false proposition (such as that Jones has ten coins in his pocket and will win the race) and deduces a proposition (say, that someone who has ten coins in his pocket will win the race) which just happens to be true. If closure for justification holds, then one will have a justified, true belief that does not amount to knowledge (see Gettier problems).

S may have justification for believing that *P* without actually believing that *P* (for example, S may fail to realize that S's evidence justifies a belief about the identity of the killer). Thus it is plausible to hold that justification is closed under logical implication *simpliciter*: if S has justification for believing propositions which logically imply *Q*, then S has justification for believing that *Q*. This principle is employed in generating the lottery paradox. One holds a ticket in a very large, fair lottery which one justifiably believes will have just one winner. Assume that the very high probability that one's ticket will not win justifies one in believing that one's ticket will not win. Since similar reasoning will apply to every other ticket, one has justification for believing that the first will not win, one has justification for believing that the second will not win, and so on. Given that these propositions logically imply that no ticket will win, the closure principle now under consideration implies that, paradoxically, one has justification for believing that no ticket will win (see Paradoxes, epistemic §1).

2 Justification and relevant alternatives

Fred Dretske (1970) describes a case which appears to be a counterexample to both closure for justification and

Closure itself (assuming that justification is a necessary condition for knowledge). Here we see again that the assessment of Closure depends on one's views about the analysis of the concept of knowledge. At the local zoo, one sees some striped animals in an enclosure marked 'Zebras'. One's evidence justifies a belief that these animals are zebras. However, one's evidence does not count towards their *not* being cleverly disguised mules (since one would have exactly similar evidence were it a hoax). Therefore, one's evidence does not justify one in believing the proposition that these animals are not cleverly disguised mules, even though one recognizes (we will suppose) that that proposition is logically implied by the zebra proposition for which one has justification.

A sceptically minded philosopher might well object that it is far from clear that one's evidence does justify a belief that the animals are zebras. To the extent that one's evidence does not count against the cleverly disguised mule hypothesis, the sceptic would say, it does not count in favour of the zebra hypothesis.

Dretske sketches a general theory of knowledge which bolsters his appeal to brute intuition about the proper analysis of the zebra case (intuition which might beg the question against scepticism). He holds that when S knows that *P* in some particular context, there is a range of *relevant alternatives* (counter-possibilities) to *P* which is a subset of the set of *all* alternatives to *P*. To know that *P*, it is only required that S be able to rule out the relevant alternatives to *P*. On this theory, Closure fails, because there will be situations in which S knows that *P*, knows that *R* is an alternative to *P* (knows that not-*R* is logically implied by *P*), and yet is unable to rule out *R* (fails to know that not-*R*). These are situations in which *R* is an *irrelevant* alternative to *P*.

The main motivation behind the relevant-alternatives theory is to provide a way of blocking the Cartesian sceptical argument. Therefore, we need an account of *relevance* according to which the sceptic's bizarre alternatives to ordinary propositions about the external world turn out to be irrelevant. This will make it possible for one to know that one has hands while lacking knowledge that one is not a disembodied brain in a vat. There is little agreement among relevant-alternative theorists on the analysis of the concept of relevance. One view is that an alternative *R* is relevant with respect to a particular claim to know that *P* if and only if the objective probability of *R* meets some specified level. If there are hoaxing zoos in one's vicinity, then this could render the cleverly disguised mule possibility a relevant alternative in the zebra case. Another view is that alternative relevance is determined by various features of the conversational context in which a knowledge attribution is made (a context which might not include the putative knower). The mentioning of alternatives might be such a contextual feature, so that the brain in a vat alternative can become relevant in a philosophical discussion of knowledge.

It has been argued (Klein 1981) that closure for justification holds even though (as Dretske claims) one's evidence *E* may justify a belief that *P* without justifying a belief of a deduced proposition, such as that one is not in a vat. On this view, even though *E* does not justify one's belief that one is not in a vat, one still has an adequate reason for believing that proposition. This is because *P* itself becomes available as evidence for the deduced proposition, given that *E* justifies one's belief that *P*.

3 Closure and reliabilist theories of knowledge

Let us consider the *reliabilist* approach to the analysis of the concept of knowledge, according to which knowing that *P* is a matter of having a reliably produced true belief that *P*. On many such accounts, the justification condition for knowledge is either analysed as, or replaced by, a reliability condition (see Reliabilism). Whether Closure fails on such an analysis depends on how reliability is conceived. On one conception (Goldman 1976), in order for S's belief that *P* to count as reliably produced, S must be able to discriminate the actual situation in which *P* is true from those *relevant possible situations* in which *P* is false. This is close to the relevant alternatives theory, and Closure will accordingly fail. If a possible situation in which *P* is false and S is envatted is not a relevant one, then S can know that *P* without being able to discriminate the actual situation from the one in which S is envatted. Now consider S's claim to know the logically implied proposition that S is not envatted. A possible situation in which S *is* envatted is plausibly regarded in this case as a relevant alternative possibility. Since, as we assumed, S is unable to discriminate the actual situation from a vat situation, it follows that S does not know that S is not in a vat. Thus Closure fails, since S can know that *P* without knowing that S is not envatted.

Reliability is sometimes conceived as having an explicitly counterfactual dimension. For example, Robert Nozick's (1981) 'tracking' analysis of knowledge contains this counterfactual condition: if *P* were false, then S would not mistakenly believe that *P*. Suppose that S claims to know that S is sitting, on an occasion when S is

sitting in a perfectly normal environment. If S were not sitting, then, presumably, S would be standing, or lying down, or the like. In such a counterfactual circumstance, S would *not* mistakenly believe that S is sitting. So S's claim to know that S is sitting satisfies Nozick's counterfactual condition (as well as the other conditions for knowing, we may suppose). Now consider S's claim to know the logically implied proposition that S is not in a vat. If S *were* in a vat, then, presumably, S *would* mistakenly believe that S is *not* in a vat. So S does not satisfy the counterfactual condition for knowing that S is not in a vat. Thus on the present analysis, S knows that S is sitting but fails to know that S is not a brain in a vat.

According to an alternative counterfactual conception of reliability, whether a belief that *P* is reliably produced does not necessarily depend on what one would believe in counterfactual circumstances in which *P* is false. Instead, we focus on the process which issues in the belief. A belief-forming process will count as reliable just in case the process actually yields a sufficiently high ratio of true beliefs and also would yield the required ratio in counterfactual circumstances similar to one's actual circumstances (which might not include circumstances in which *P* is false). Closure will hold on this version of reliabilism. Suppose that one forms the correct belief that one has hands via an ordinary perceptual process. Then one forms the further belief that one is not disembodied in a vat, via the belief-forming process of deductive inference. This inferential belief issues from a process that is obviously reliable in the sense under consideration. Thus if the original belief issues from a process which is reliable in that sense and therefore amounts to knowledge, then so will the inferential belief.

In the end, we see that even though the denial of the closure principles we have discussed would aid in the refutation of scepticism, whether these principles hold is a controversial matter which depends on the nature of knowledge and justification.

See also: Epistemic logic; Knowledge, concept of §§7-10; Scepticism

ANTHONY BRUECKNER

References and further reading

Brueckner, A. (1985) 'Skepticism and Epistemic Closure', *Philosophical Topics* 13 (3): 89-117.(Critical overview of the issues discussed in this entry.)

Dretske, F.I. (1970) 'Epistemic Operators', *Journal of Philosophy* 67 (24): 1007-23. (*Locus classicus* of the relevant alternatives theory discussed in §2 above, and includes the zebra case discussed in that section; contains a sketch of a view like Nozick's, considered above in §3.)

Dretske, F.I. (1981) 'The Pragmatic Dimension of Knowledge', *Philosophical Studies* 40 (3): 363-78.(Useful discussion of the concept of a relevant alternative.)

Gettier, E. (1963) 'Is Justified True Belief Knowledge?', *Analysis* 23 (6): 121-3.(The origin of the Gettier problem discussed above in §1.)

Goldman, A.I. (1976) 'Discrimination and Perceptual Knowledge', *Journal of Philosophy* 73 (20): 771-91. (Reliability as discrimination, as discussed in §3 above; contains a useful discussion of the concept of a relevant alternative.)

Klein, P. (1981) *Certainty: A Refutation of Scepticism*, Minneapolis, MN: University of Minnesota Press. (Discussion of the argument for the closure of justification discussed in §2 above; discussion of the relation between closure principles and the problem of scepticism.)

Klein, P. (1987) 'On Behalf of the Skeptic', in S. Luper-Foy (ed.) *The Possibility of Knowledge: Nozick and his Critics*, Totowa, NJ: Rowman & Littlefield.(Contains an argument that reliabilism can provide support for Closure, as discussed above in §3; the book contains many trenchant critical articles about the anti-sceptical strategy of Nozick, as well as a helpful bibliography.)

Kyburg, H.E. (1970) *Probability and Inductive Logic*, Toronto, Ont.: Macmillan. (Chapter 13 gives a presentation of the lottery paradox discussed in §1 above.)

Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press.(Chapter 3 contains a discussion of the anti-sceptical strategy discussed in §3 above.)

Williams, M. (1991) *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism*, Oxford: Blackwell. (Chapter 8 contains arguments that reliabilism does not induce the failure of Closure, as discussed above in §3.)

Definition

A definition is a statement, declaration or proposal establishing the meaning of an expression. In virtue of the definition, the expression being defined (the ‘definiendum’) is to acquire the same meaning as the expression in terms of which it is defined (the ‘definiens’). For example, ‘Man is a rational animal’ determines the meaning of the term ‘man’ by making it synonymous with ‘rational animal’. Classical theory maintains that a good definition captures the ‘real nature’ of what is defined: ‘A ‘‘definition’’ is a phrase signifying a thing’s essence’ (Aristotle). Historically, philosophers have come to distinguish these ‘real’ definitions from ‘nominal’ definitions that specify the meaning of a linguistic expression rather than signify the essential nature of an object, ‘making another understand by Words, what Idea, the term defined stands for’ (Locke).

A further distinction can be drawn between contextual or implicit definitions, on the one hand, and explicit definitions, on the other. Often a definition fixes meaning directly and explicitly: for example, the definition of a proper name might well take the form of an explicit identity statement (‘Pegasus = the winged horse’) and a definition of a predicate is usually given (or can be re-cast) in the form of an equivalence (‘For every x: x is a man if and only if x is a rational animal’). But sometimes the meaning of a term is specified in context, by way of the meaning of larger expressions in which the term occurs. A paradigmatic example of this is Bertrand Russell’s analysis of the meaning of the definite article.

1 The role of definitions

A definition is a statement, declaration or proposal (the ‘definiens’) establishing the meaning of an expression (the ‘definiendum’). Explicit definitions can be viewed as providing useful abbreviations for complex expressions. It follows that, strictly speaking, they are theoretically dispensable, in so far as we can always replace the definiendum by the definiens in any context. This was already noted by Whitehead and Russell in *Principia Mathematica*, according to whom definitions are ‘mere typographical conveniences’ ([1910] 1925: 11). However, they also point out that from a more pragmatic point of view, the use of a definition (whether explicit or not) conveys additional information besides the literal meaning of the sentence in which the definiendum occurs.

This can happen in two ways. First, a definition calls attention to the fact that the (complex) concept expressed by the definiens plays an important role in the discourse in which the definiendum is used: for instance, agreeing to mean by ‘prime’ a number that can be divided without remainder only by 1 and itself calls attention to the importance of this notion for number theory. ‘Hence the collection of definitions embodies our choice of subjects and our judgment as to what is most important’ ([1910] 1925: 11-12). Second, a definition may sometimes supply an analysis of an already existing concept that had not previously been made precise. In this case the newly introduced concept shares most or all of the traits of its informal counterpart, although realizing those traits in a more specific and precise manner. It should be observed that in general many competing, equally precise and sometimes formally equivalent analyses of the same informal idea are possible, and the choice among them is then inspired by considerations such as simplicity, elegance and adaptability to the particular context of application.

A related view of definitions was put forward by Carnap (1947), who characterizes certain definitions as ‘rational reconstructions’ of less precise concepts derived from everyday life or earlier stages of scientific development. Such a definition is called an ‘explication’, and the newly defined term the ‘explicatum’; the earlier, less precise concept is referred to as the ‘explicandum’. In giving an explication, one ‘sharpens’ a vague, pre-existing concept, making it precise; or even replaces it with a new, more exact one. In general it is not required that the explicandum have the same meaning as the explicatum. All that is required is some sort of correspondence between the two concepts, in such a way that the latter can be used instead of the former. This task of rational reconstruction is ‘one of the most important…of logical analysis and logical construction’ ([1947] 1956: 7-8).

2 Formal conditions on definability

From a purely formal point of view, a definition given in the context of some theory behaves just like any other axiom of the theory, in that it can be used to infer propositions (theorems). However, as already noted by Whitehead and Russell, a definition should not convey any new information: the theory T' obtained by supplementing some theory T by a definition should, in some sense, be equivalent to T.

It is possible to make this requirement precise by laying down two criteria that a definition should satisfy: eliminability and non-creativity. It appears that these criteria were first formulated by the Polish logician S. Leśniewski (1931), but the presentation given here follows P. Suppes (1957).

In order to explain the two criteria, we will suppose that we are working within a formalized background theory T, that is, a set of axioms expressed in some formal language. In such a language, as is customary, free variables are to be construed as implicitly universally quantified; moreover, T is assumed to be consistent. A definition then is just a sentence ϕ that involves some new symbol not already in T. In particular, ϕ is allowed to occur in inferences from the axioms of T. The two criteria can be formulated as follows. A sentence ϕ introducing a new symbol satisfies the criterion of eliminability if and only if whenever η is a formula in which the new symbol occurs, then there is a formula θ in which the new symbol does not occur and such that $(\eta \leftrightarrow \theta)$ can be inferred from T augmented with ϕ. Similarly, we say that ϕ satisfies the criterion of non-creativity if and only if whenever θ is a sentence not involving the new symbol and such that $\phi \rightarrow \theta$ can be inferred from T, then θ can already be inferred from T; in other words, the introduction of the definition does not make possible the derivation of any new theorems that do not involve the new symbol. (One then says that $T + \phi$ is a 'conservative extension' of *T*.) Notice that the second criterion implies, in particular, that the theory supplemented by the definition be free of contradictions: inconsistency is a strong form of creativity.

Although the two criteria apply to explicit and implicit definitions alike, in the case of explicit definitions it is possible to lay down certain constraints on the syntactic form of the sentence ϕ ensuring that the two criteria are automatically satisfied. For simplicity, we will review these constraints only for the case in which the definiendum is a relation symbol or an individual constant (name), but a similar constraint can be formulated for the definition of function symbols.

A formula ϕ is an acceptable definition of an n-place relation symbol R if and only if ϕ has the form $R(x_1, \ldots, x_n) \leftrightarrow \theta$, where $x_1, \ldots, x_n$ are all distinct variables and θ is a formula all of whose free variables are among $x_1, \ldots, x_n$ and in which R does not occur. The requirement that the variables $x_1, \ldots, x_n$ all be distinct is necessary to ensure that the eliminability criterion is satisfied, as is the further requirement that the definiendum R not be allowed to occur in the definiens; jointly, the two requirements are also sufficient for the satisfaction of the eliminability criterion.

The requirement that the definiens not contain any variables that do not already occur in the definiendum is meant to ensure non-creativity. To see that the constraint expresses a necessary condition, consider, for instance, the definition '$G(x) \leftrightarrow x > y$'. Since free variables are implicitly universally quantified, the definition is equivalent to the conjunction of '$G(x) \rightarrow \forall y(x > y)$' and '$\exists y(x > y) \rightarrow G(x)$'; by transitivity of implication we obtain that if $\exists y(x > y)$ then $\forall y(x > y)$, which, when interpreted as a statement of arithmetic, is false. The definition is therefore creative. (The constraint on variables is also sufficient for non-creativity if and only if the definiendum is not allowed to occur in the definiens.)

The constraint on the explicit definition of individual constants is slightly more complex, due to the existential import of such constants in ordinary logic. Rather than considering definitions in the form of an explicit identity '$c =$ the so-and-so', we will give a constraint using the more general form of an equivalence. A sentence ϕ is an acceptable definition of an individual constant c if and only if ϕ has the form $c = x \leftrightarrow \theta$, where θ is a formula not containing c with x as its only free variable, and such that $\exists! x\theta$ (that is, there is one and only one x such that θ) is derivable from the theory.

3 Contextual definitions

Sometimes it is neither possible nor convenient to give a definition in the explicit form of an equivalence or identity. Indeed, given that formally definitions are just further axioms, any sentence containing a new symbol could be taken to be a candidate for a definition of that symbol. Such definitions are called 'contextual', in that the meaning of the new symbol is specified (at least partially) by giving the meaning of some larger expression in which it occurs; the symbol itself, when occurring in isolation, need not be endowed with any meaning at all. Perhaps the first and most conspicuous example of contextual definition is given by Russell (see, for example, Whitehead and Russell [1910] 1925: 66; but see also Frege1884, 1903), when specifying the meaning of the definite article by giving explicit truth-conditions for sentences in which it occurs. As is well known, Russell

proposes to analyse sentences of the form 'The ϕ is ψ' as the conjunction of the three following sentences in which the definite article no longer occurs: (1) There is at least one ϕ; (2) There is at most one ϕ; and (3) Whatever is a ϕ is a ψ. Other examples are the contextual definition of the membership symbol '$\in$' given by the axioms of set theory, and the usual recursive definition of '$+$' in the context of arithmetic (the latter is given by the conjunction of the two equations $x + 0 = x$ and $x + y' = (x + y)'$, where x' denotes the successor of x).

Of course, in general, one would not expect the criterion of eliminability to be satisfied, for a contextual definition need only tell us how to find an equivalent not involving the new symbol for a restricted class of sentences, namely those in which the definiendum occurs in the context given by the definition. For instance, the usual recursive definition of '$+$' does not tell us how to eliminate this symbol in the context $\forall x \forall y(x + y = y + x)$. But the criterion of non-creativity still applies, along with the weaker requirement that the theory augmented by the definition still be consistent.

As a related question one might consider whether, within the framework of a particular theory, a contextual definition can be replaced by an explicit one. A useful tool in answering this question is a method invented by A. Padoa (1902) and further developed by Tarski (1956: 296-320). The method relies on the fact that if a symbol α is explicitly definable within a given theory T, then any two interpretations on which all sentences of T are true must agree on the interpretation of α. Therefore, in order to show that α is not explicitly definable in T it suffices to exhibit two interpretations of T that assign different values to α.

Padoa's method also has a converse of sorts, which is a deep result about the expressive power of first-order logic. Let us say that a symbol α is 'implicitly definable' in T if any two interpretations of T agree on the value they assign to α. (So implicit definitions are a particular case of contextual definitions.) The proposition that any symbol implicitly definable in T is also explicitly definable is known as Beth's definability theorem (see Beth's theorem and Craig's theorem). This is a powerful theorem, and one of the results of model theory with the most far-reaching philosophical implications.

4 'Circular' definitions

We have seen that one of the constraints on the syntactic form of definitions meant to ensure the satisfaction of the criterion of eliminability is that the definiendum itself not occur in the definiens. That is, the requirement rules out 'circular' definitions. This is in keeping with the idea that a definition, among other things, should also provide a way to determine the extension of the defined term, that is, the set of entities that fall under it. Indeed, a way of determining whether an object x falls in the extension of, say, a predicate P is to ascertain whether x satisfies the definiens of P: if the definiens itself is allowed to have occurrences of P, then it might be impossible to determine whether anything satisfies it without knowing the extension of P. So circularly defined terms would lack a definite criterion of applicability.

However, there are circumstances in which it appears that no non-circular definition is available for a given concept. Consider, for instance, the concept of truth, construed as a predicate of sentences of a given language. If the language in question itself contains the truth-predicate then we can have circularity as follows. Tarski (1944) suggested that for any sentence ϕ, the following equivalence (known as a 'T-biconditional') might be regarded as a *partial* definition of truth: t is true if and only if ϕ, where t is a term denoting ϕ (for instance, t could be obtained by enclosing ϕ in quotes). If ϕ itself contains occurrences of 'is true', the T-biconditional for ϕ exhibits a form of circularity. As is well known, this circularity can lead to paradoxes, as in the case of the so-called liar paradox arising from the consideration of the T-biconditional for the sentence λ: λ is not true. As another example, consider the kind of (harmless) circularity that arises when we are trying to specify a function f by means of the two equations $f(x, 0) = x$ and $f(x, y') = (f(x, y))'$, where as before x' is the successor of x (this is harmless because, as is well known, the two equations above identify a unique function over the natural numbers, namely addition).

A general theory of possibly circular definitions was put forward by A. Gupta and N. Belnap (1993). They regard a definition not as supplying a way to fix the extension of a term, but as providing a 'rule of revision', that is, a way of improving upon hypothetical extensions. Suppose we have a possibly circular definition of a 1-place predicate P, which has the form $P(x) \leftrightarrow \phi(P)$, where the definiendum P occurs in the definiens. This can be viewed as an operation defined over the space of the possible extensions of P (that is, over the set of all subsets of the domain):

given a possible extension E as input the rule of revision returns the value of $\phi(P)$ when P is interpreted as E. In turn, a revision rule can be used to give rise to a revision process, that is, an infinite sequence of possible extensions, defined as follows. We begin with an arbitrary extension E as a 'bootstrapper' or initial guess; we then let E_1 be the result of applying the revision rule to E, E_2 the result of applying the revision rule to E_1, and so on. In this way we obtain an infinite sequence. Say that an extension E' 'coheres' with such a sequence if it contains all items that are eventually in every member of the sequence, that is, all items x such that for some E_n, x is in E_i for every $i \geq n$. We can then extend the sequence into the transfinite by letting E^∞ come after each member of the sequence, where E^∞ is an arbitrary extension that coheres with the sequence, and start applying the revision rule again.

So each revision rule specifies a class of revision processes, each one of which is characterized by a different 'bootstrapper' and different 'guesses' at limit stages. Should a revision process culminate in a fixed point, that is, an extension that is returned unchanged by the revision rule, we have found a 'good' candidate for the extension of P. But it should be noted that this is a by-product of the definition: the meaning of the definition is not given by the extension (which might not exist), but by the pattern of variation exhibited by the various revision processes to which it gives rise.

See also: Frege, G. §§6, 8, 9; Logical and mathematical terms, glossary of; Paradoxes of set and property §3

G. ALDO ANTONELLI

References and further reading

Belnap, N.D. (1993) 'On Rigorous Definitions', *Philosophical Studies* 72: 115-46.(Expanded treatment of the issues covered here; rigorous yet accessible both from a philosophical and technical point of view; emphasizes definition as an act of conceptual clarification.)

Carnap, R. (1947) *Meaning and Necessity: A Study in Semantics and Modal Logic*, Chicago, IL: University of Chicago Press; 2nd, enlarged edn, 1956.(Presents Carnap's notion of explication.)

Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, 2nd edn, 1980.(Numbers are defined contextually - as extensions of particular concepts - by giving their identity conditions.)

Frege, G. (1903) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, Jena: Pohle, vol. 2; fragments, including 'Frege on Definitions', in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980.(Gives the 'sharp boundary' condition, emphasizing that definitions should be complete and not leave any cases undecided. Also, the inconsistent axiom (V) can be viewed as a (creative!) contextual definition of the extension of a concept.)

Gupta, A. and Belnap, N.D. (1993) *The Revision Theory of Truth*, Cambridge, MA: MIT Press. (Includes the first full presentation of revision theory as a general theory of definition.)

Leśniewski, S. (1931) 'Über Definitionen in der sogenannten Theorie der Deduktion', *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie* (Classe 3) 24; trans. in *Collected Works*, ed. S.J. Surma, J.T. Srzednicki and D.I. Barnett, with an annotated bibliography by V.F. Rickey, Boston, MA: Reidel, 2 vols, 1981, 629-48.(The original source for the theory of definition in first-order logic.)

Padoa, A. (1902) 'Un nouveau système irreductible de postulats pour l'algebre', in *Deuxiême Congrès International des Mathematiciens*, Paris, 249-56.(First appearance of a necessary condition for implicit definability, later developed by Tarski.)

Robinson, R. (1950) *Definition*, Oxford: Oxford University Press.(A good source for the nontechnical issues in the theory of definition. Includes some history.)

Suppes, P. (1957) *Introduction to Logic*, Princeton, NJ: Van Nostrand.(Possibly the best treatment of definitions in first-order logic from an elementary point of view.)

Tarski, A. (1944) 'The Semantic Conception of Truth', *Journal of Philosophy and Phenomenological Research* 4: 341-76; repr. in L. Linsky (ed.) *Semantics and the Philosophy of Language: A Collection of Readings*, Champaign, IL: University of Illinois Press, 1969.(Tarski's more philosophical treatment of the concept of truth, including the partial definition of truth as given by the T-biconditionals.)

Tarski, A. (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press; repr. and ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(Tarski's development of Padoa's method. See also his remarks on formal and material correctness in 'The Concept of Truth in Formalized Languages'.)

Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press, 2nd edn, 1925.(The source for contextual definitions - sums up Russell's work published elsewhere.)

Definition, Indian concepts of

*Definitions in Indian philosophy are conceived very differently from definitions in Western philosophy. In Western philosophy and logic, it is usual to define a term or a linguistic expression. A definition here consists of a 'definiens', typically a longer expression, statement or proposal, and a 'definiendum', a shorter expression or term whose meaning is established by the definiens. Definitions permit the definiendum to be put in place of the definiens and are thus 'abbreviations' (for example, 'father' is an abbrevation of 'male parent'). In India, definitions in the sense of abbreviations were regularly used in grammar from the earliest times, as in the work of Pāṇini (*c.*800 BC).*

*In Indian philosophy, however, definitions are not conceived of as abbreviations. We may have direct acquaintance with an object; this is one way of knowing it. We may also know an object or many objects through their properties or features; this is another way of knowing them. These properties or features are the modes under which objects are cognized. If we know objects through the properties that belong to all of them and only to them, then the objects are collected together through their properties to form a group. A group is nothing real; it is a way of collecting objects by knowing them under one mode. When we know a group of objects through properties common to all of them and only to them, we may also want to know another set of properties or features which also belongs to all the objects and only to them. The second set of properties is the defining mark (*lakṣaṇa*), or, simply, the definition, of the objects collected together into a group by being known under one mode. The objects themselves are the definienda of the definition. The first set of properties through which the definienda are collected together to form a group is called 'the limiting properties of being the definienda of the definition'. The defining mark, that is, the definition, is not an essential property of the definienda, but is only a property (or set of properties) common to all of them and only them.*

1 General features of definition

There are two stages in the Indian theory of definition: (1) the objects to be defined must all be known under one mode ('the limiting property of being the definienda'); and (2) the definition of those objects must be a property coextensive with the limiting property. The limiting property and the definition cannot be identical, although they must be coextensive. A single object, too, can be defined. Its defining mark, that is, its definition, depends upon the prior act of identifying it. The object's criterion of identity is the limiting property of being the definiendum, which has to be different from the defining mark.

A sentence stating the defining mark of a group of objects is called 'the sentence of definition'. It has some special features which should be noted. A sentence of definition has to be a general sentence applicable to all the definienda. In Indian philosophy, variables are not used; pronouns are used to attain generality. For example a true cognition (not 'a true cognition') is defined as a cognition having *that* as a predicate in a *that*-possessor. For example, if a thing has potness (that is, is a pot), then the cognition that the thing has potness (that it is a pot) is a true cognition. Using the pronoun 'that' twice in the sentence makes it a general sentence. For we may write any property term for 'that'.

As objects belonging to very different ontological categories can be the definienda of a definition, the defining mark may belong to the different kinds of definienda by very different ontological relations. This is a peculiarity of the Indian theory which distinguishes it from the theories of traditional Western logic.

2 Navya-Nyāya theory of definition

Only Navya-Nyāya ('New Logic'), created by Gaōgeśa (*c.*1325) and developed by Raghunātha (*c.*1530), Jagadīśa (*c.*1620), Gadādhara (1604-1709) and later philosophers, has a fully developed theory of definition.

There are two fundamental ways in which the Navya-Nyāya theory of definition differs logically from Western theories, apart from the ontological point of view. First, it differs from Plato's theory in that whereas Plato tries to find out the essence or the universal of, say, justice, in Navya-Nyāya, definitions are properties, not necessarily universals or essences, and may even be individual properties belonging to single individuals.

In Plato, Aristotle and Porphyry, there is a hierarchy of genus and species, and definition is *per genus et*

differentiam. Thus every definition must have at least two components - the generic property and the differentia; the defining property of 'human', for example, is animality-and-rationality. The Navya-Nyāya theory of definition is fundamentally different because the structure of the defining property is not always composite, especially when it is a proper universal in the sense of Nyāya. A universal is an unanalysable property, even though it may be subsumed under a wider universal. No universal is a composite of a generic property and differentia. Thus the defining property of humans is just humanity, even though it is subsumed under the generic property of animality. According to Navya-Nyāya, humanity as a property is as simple as animality and cannot be obtained by adding differentia to the generic property. Analysis of a concept is possible only in the case of composite properties, which are not proper universals. Universals, when not referred to by an expression, are known in and through themselves and not under the mode of some other property, and therefore do not stand in need of analysis; indeed, they cannot be analysed at all. It has been objected that although it is correct to say that what games have in common is their being games, it is not at all enlightening. But enlightenment, if sought by further analysis, is not possible in this case.

3 Navya-Nyāya theory of relations

The second fundamental point of difference between the Navya-Nyāya theory and traditional Western theories of definition stems from the Navya-Nyāya doctrine that a defining property need not be a universal. A proper universal is related to all its instances by one relation, inherence, according to Navya-Nyāya. The defining property is very often a property different from a proper universal and is often related to different types of objects by different types of relations. Navya-Nyāya therefore developed a very elaborate technique of showing how the defining property can be said to be 'common' to objects which very often belong to ontologically different categories. This is done by determining the specific nature of the relation by which the defining property can be said to be 'common to' the different types of objects. This specific relation between the defining property and the different types of objects is very often the sum of the different relations by which the defining property is related to different types of objects. This sum of relations is necessary, for it is by this relation that the defining property can be said to be 'common to' all the different types of objects. As an example of this technique of showing how a defining property is 'common to' all the objects to be defined, we may explain the definition of *being a positive object* (*bhāvatva*). According to Navya-Nyāya, there are six ontological categories of positive reals. There is one proper universal common to the first three categories of positive reals, namely existence (*sattā*), which is the highest universal (*parā jāti*). This universal is related to six categories: (1) substance, (2) quality, (3) motion, (4) universal, 5) ultimate differentia, and (6) inherence (see Nyāya-Vaiśeṣika §§4-5; Ontology in Indian philosophy).

Existence is a property common to the first three categories in the sense of being related to them by inherence. Now, according to Navya-Nyāya, the remaining three categories of objects do not have existence inhering in them. Thus, if existence is to be the defining property of being a positive object it must be related to the last three categories of objects by other relations. Navya-Nyāya shows how existence can be indirectly related to the fourth and fifth categories of objects. Universals and ultimate differentia inhere in, say, substances, in which existence also inheres. In Navya-Nyāya terminology, existence is related to universals (and ultimate differentia) as both inhere in the same object.

Still there is the problem of showing how existence can be related to inherence itself, for inherence is the sixth category of positive reals. Existence cannot inhere in inherence, for according to Navya-Nyāya, existence inheres in only the first three categories of positive reals. According to Navya-Nyāya, existence is related to, is located in, inherence by a self-linking relation which is ontologically inherence itself. There is no need to postulate an ontologically different relation between existence and inherence. Inherence, which is the second term of the relation, also functions as the relation. Thus, although the relation between existence and inherence is not *inherence as inherence*, but only *inherence as a self-linking relation*, still it is inherence which is ontologically the same as this relation.

In Western logic, universal laws are formulated by means of variables. These variables are signs bound by some operators, and also act as place-holders for names of their values. Because of the convention that the same name in to be written at all places where a variable occurs (within the scope of one quantifier, or in the same formula if it occurs freely), the variables perform the function of relative pronouns.

Although variables have been used in Indian algebra for a long time, they were not used in Nyāya. In algebra

before the introduction of variables, the pronouns *yāvat* and *tāvat* were used. In Nyāya, there is a very popular method similar to this practice which uses the relative pronouns *yat* and *tat* ('that' and 'which') and *sva* ('its', or 'one's own'). Consider the following definition: 'Pervasion is the co-presence of that with that which is not limited by the limiter of the counterpositiveness to the constant absence that has a common locus with it and has no common locus with its counterpositive.' In this definition the first 'that' stands for the probans, the second 'that' for the probandum, and the 'its' for the constant absence. (The 'its' occurs only in the translation; in Sanskrit it is eliminated by a compound.)

4 First type of definition

Two different types of definition are admitted in Navya-Nyāya, corresponding to two different purposes for which defining marks of objects are determined. The first purpose for finding a defining mark of objects collected under a mode is to find out a mark by which all the objects can be distinguished from everything else. Thus the defining mark in this case becomes the probans (corresponding to the middle term of a syllogism) for inferring that every one of the objects to be defined is different from everything else. This inference has to be for the sake of others; for according to Navya-Nyāya, the defining mark is used to prove to others that it excludes the objects to be defined, *P*s, from every non-*P*. This proof, which takes the form of an inference for the sake of others, has to be put in the negative form (see Inference, Indian theories of §5).

This point may be explained with an example. The defining mark of earth is smell. By smell all earthen things are differentiated from non-earth. The inference is of the form: 'Earth is different from non-earth, the reason being smell.' This can be put in the form of a syllogism thus: 'Whatever is not different from non-earth has no smell. Earth has smell; therefore earth is different from non-earth.' Here earth is *P* (the minor term), difference from non-earth is the probandum (the major term) and smell is the probans (the middle term).

This example shows a fundamental difference between Western theories and Indian theories in general, and Navya-Nyāya in particular. In Western philosophy, 'earth is different from non-earth' ('earth is not non-earth') is regarded as an analytical sentence. In traditional Western logic, 'earth is not non-earth' is the obverse of 'earth is earth', which is an identity sentence. But in Navya-Nyāya, no analytic sentence is significant. Although one can know (always truly) that earth is earth, there is no point in saying it, for a sentence expressing it will not communicate any information. So, according to Navya-Nyāya, there is analytical judgment, but there is no analytical sentence. But the sentence 'earth is different from non-earth' ('earth is not non-earth') is not an analytical sentence; it can be used as the conclusion of an inference. This is not true a priori; it needs to be proved. Every definition in this sense is the probans of a conclusion of the form: *P* is different from non-*P*.

5 Second type of definition

The second type of definition is used to explain why a word applies to whatever it applies to. According to Nyāya, a word becomes applicable to its referents if and only if there is a sufficient reason for its application to them. This reason must be a property of the referents, and this property is the defining property of the referents being the referents of that word. The defining mark is the nature of the referents of the word, and is denoted by the abstract noun formed from the word. For example, jarness is the nature of jars and is denoted by the abstract noun 'jarness' formed from the word 'jar' by the abstraction suffix '-ness'.

There may be different properties which belong to all and only those objects which are denoted by a word, and the problem is to decide which of them is to be regarded as the defining mark of the objects considered as the referents of the word. Not all coextensive properties are necessarily identical, even though two coextensive universals are. For example, all jars and only jars have the two properties of (1) having a neck like a conch shell and (2) jarness. Now these two properties, although coextensive, are still different, for (1) is an analysable property while (2) is not. It is to be noted that jarness cannot be used as a probans to infer that jars are different from non-jars, for here the limiting property of being objects to be defined, jarness, becomes the same as the defining mark. Thus jarness cannot be regarded as the defining mark of jars, in the first sense of the term. Jarness can, however, be the defining mark of jars in the sense of being the sufficient reason for the application of the word 'jar' to jars. In this case, the limiting property of being objects to be defined does not become identical with the defining mark, jarness, for the limiting property is not, in this case, jarness, but the property 'being the referent of the word "jar"'. In this type of definition we are considering jars only as referents of the word 'jar', and are trying to discover the sufficient

reason for its application to them.

6 Relation between the two types

There are two different motives for finding defining marks. First, when one hears the word ‘earth’, one may want to know the distinguishing mark of earth. Here the limiter of being objects to be defined is a property of the objects denoted by the word. Second, one may want to know why the word ‘earth’ is applicable to whatever it is applicable. Here the limiter of being objects to be defined is being a referent of the word ‘earth’.

There are three different theories of the relation between the two kinds of defining marks. First, the most commonly accepted theory is that these are of two different kinds. The second theory is that a defining mark in the proper sense is only a distinguishing mark; the so-called usage-explaining marks, too, can be used as probantia for inferring difference from all others by only reinterpreting the limiter of being objects to be defined as the property of being referents of the words. If this property is regarded as the limiter, then all usage-explaining marks can be used as distinguishing marks. The third theory is that a distinguishing mark can also be used as a usage-explaining mark by conceiving a usage-explaining mark as a mark whose cognition enables one to infer the correct applicability of the word. Thus there are only distinguishing defining marks, and no special type of usage-explaining defining marks.

See also: Definition

SIBAJIBAN BHATTACHARYYA

References and further reading

Bhattacharyya, S. (1987) *Doubt, Belief and Knowledge*, New Delhi: Allied Publishers, esp. chaps 17-18.(General introduction to the Navya-Nyāya theory of inference.)

Bhattacharyya, S. (1990) *Gadādhara’s Theory of Objectivity*, New Delhi: Motilal Banarsidass, esp. part 1, ch. 5. (For detailed discussion of the Navya-Nyāya theory of definition.)

Bhattacharyya, S. and Potter, K.H. (eds) (1993) *Encyclopedia of Indian Philosophies*, vol. 6, *Indian Philosophical Analysis: Nyāya-Vaiśeṣika from Gaōgeśa to Raghunātha Śiromaṇi*, Princeton, NJ: Princeton University Press, and New Delhi: Motilal Banarsidass.(For a general introduction to Navya-Nyāya theory; see especially the introduction.)

Potter, K.H. (ed.) (1974) *Encyclopedia of Indian Philosophies*, vol. 1, *Bibliography*, Princeton, NJ: Princeton University Press, and New Delhi: Motilal Banarsidass.(For the dates and works of Navya-Nyāya authors mentioned in the present entry.)

Potter, K.H. (ed.) (1977) *Encyclopedia of Indian Philosophies*, vol. 2, *Indian Metaphysics and Epistemology: The Tradition of Nyāya-Vaiśeṣika up to Gaōgeśa*, Princeton, NJ: Princeton University Press, and New Delhi: Motilal Banarsidass.(For a general introduction to Nyāya theory.)

Deism

In the popular sense, a deist is someone who believes that God created the world but thereafter has exercised no providential control over what goes on in it. In the proper sense, a deist is someone who affirms a divine creator but denies any divine revelation, holding that human reason alone can give us everything we need to know to live a correct moral and religious life. In this sense of 'deism' some deists held that God exercises providential control over the world and provides for a future state of rewards and punishments, while other deists denied this. However, they all agreed that human reason alone was the basis on which religious questions had to be settled, rejecting the orthodox claim to a special divine revelation of truths that go beyond human reason. Deism flourished in the seventeenth and eighteenth centuries, principally in England, France and America.

1 The meaning of 'deism'

Deism in the popular sense asserts that a supreme being created the world but then, like an absentee landlord, left it to run on its own. Deism in the proper sense affirms a divine creator of the world but denies any divine revelation, holding that human reason alone is sufficient to provide us with whatever knowledge is necessary for a correct moral and religious life. (The *Oxford English Dictionary* defines 'deist' as 'one who acknowledges the existence of a God upon the testimony of reason, but rejects revealed religion'.) What is the connection between the popular and proper senses of deism? Someone who believed that God created the world but left it and its inhabitants to shift for themselves would surely deny that God acted in the world to provide us with a special revelation. So a deist in the popular sense will be a deist in the proper sense. But one may be a deist in the proper sense without being a deist in the popular sense. A person who believes in God but denies divine revelation may hold, nevertheless, that human reason is sufficient to establish that God exercises providential control over his creation. Indeed, a deist in the proper sense may even insist that reason shows us that the supreme being who created the world not only exercises providential control over his creation but is also perfectly good and just and has provided for rewards and punishments for human beings in a life to come. Thus, the basic, common point that separates deists from traditional theists is the issue of divine revelation. The deist denies that God reveals to us truths that are important for us to believe but which human reason cannot discover on its own. (Some deists were prepared to allow the existence of a revelation provided that its contents were not utterly mysterious to reason, claiming that God would have no interest in revealing things to us that we could not comprehend.) This is basic deism. On top of this one may add additional items that distinguish more radical deists from the traditional theist who believes that God has provided by special revelation truths that are important for us to believe but which our reason is insufficient to comprehend or establish on its own. The deist who sees God as creator of the world but uninterested in its future and the welfare of its human inhabitants is a basic deist who has added denials of other important ideas in traditional theism.

2 Four kinds of deist

In the second series of his Boyle Lectures, delivered in 1705, the philosopher and theologian Samuel Clarke (§§1-2) distinguished four kinds of deist. The most important kind of deist, according to Clarke, believes in the existence of God in the sense of a supreme being who is infinitely powerful, all-knowing, perfectly good and the creator of the world. Furthermore, this deist believes that God exercises providential control over the world, has created human beings who have moral and religious duties, and has provided them with a future life in which the good will be rewarded and the wicked punished. But this deist rejects the idea of divine revelation, accepting only what is discoverable by natural reason. The next kind of deist holds the same view of God's attributes and providential control over the world, but in addition to denying divine revelation also denies human immortality and any future state of rewards and punishments. A further kind of deist believes that the supreme author of the world exercises control over the natural operations of the world but, lacking the moral attributes of goodness and justice, is indifferent to human welfare. It is understood that this deist also rejects divine revelation and a future state of rewards and punishments. The most radical kind of deist not only denies the moral attributes of the supreme being and any providential concern for human welfare, but also denies that the supreme being exercises control over the natural operations of the world. According to this deist, the supreme being, although infinitely powerful and intelligent, just created matter in motion without causing any particular natural happenings in the world or even establishing the natural laws according to which these happenings take place.

It is apparent from Clarke's discussion that the denial of divine revelation is common to all four kinds of deist. In his view, the most plausible kind of deism affirms that God is morally perfect and has provided for a future state of rewards and punishments. Against this kind of deist Clarke argues that it is eminently reasonable to believe that God would provide a revelation detailing our duties towards him and towards one another, particularly our need to repent of sin and establish a right relationship with him. Clarke also contends that fulfilled prophesies and the evidence for miracles rationally justify us in taking the Bible to contain this divine revelation. The next kind of deist also ascribes moral perfection to God but is encumbered with the difficulty of explaining why such a being would not provide for a future state of rewards and punishments. The last two kinds of deist deny that the creator has any moral attributes. Against both kinds Clarke argues that there are objective moral truths independent of any divine will and that, consequently, an infinitely intelligent, all-powerful being could not fail to have the moral attributes of goodness, justice and truth. For, he argues, knowing what is best to be done is itself a motive for doing it. And since God knows perfectly the principles of duty and has no needs that might conflict with his acting in accordance with duty, he always does what is best and, therefore, is perfectly good.

3 Deism's period of influence

Deism could emerge as an important movement only in a period when belief in a divine creator is not only virtually unquestioned but taken to be provable by human reason. Deism also requires a climate of enormous confidence in the power of human reason to disclose the basic truths about God, the world, morality and human destiny. And finally, deism can flourish best, if at all, only when there is general agreement that the traditions and authority of the established religion should be subject to the test of reason, and when there is some degree of toleration for different views on religious questions. In the latter half of the seventeenth century these conditions came to prevail in England, with the result that deism took root there, spreading then to Europe and America. But by the late eighteenth century its influence in England and elsewhere had declined.

The father of deism in England, Lord Herbert of Cherbury (§§3-4) (1583-1648), argued that from the very beginning of the human race God implanted in our minds a disposition to believe in the existence of a supreme being, in the obligation to worship this being by pursuing a life of piety and virtue, and in the existence of a future life in which the good will be rewarded and the wicked punished. These basic ideas he held to be at the heart of all religious traditions, constituting a universal, natural religion apprehended solely by human reason. Later deistic writers followed Herbert's lead in taking human reason as the sole source of a basic religion of nature, viewing with scepticism, if not disbelief, appeals to the authority of the Bible, the authority of the Church and its traditions, and the testimony of the inner light. Among these deistic writers in England were John Toland (1670-1722), Anthony Collins (1676-1729), Thomas Woolston (1660-1724) and Matthew Tindal (1657-1730). Toland, an Irishman, published *Christianity not Mysterious* (1696), in which he argued that nothing in Christianity that is true is either contrary to reason or above reason, attributing the mysteries in Christianity to paganism and the interests of priests. He also expressed doubts about the authenticity of parts of the Bible. His book raised such a storm that it was condemned by the Irish parliament, and he fled to England to avoid imprisonment. He professed to be a disciple of Locke, and it is instructive to compare their views. In his opposition to enthusiasm, Locke had advocated the use of reason in religion and was prepared to reject anything in Scripture that was contrary to reason. But he allowed for divine revelation of truths that were beyond reason as well as truths that were supported by reason. Thus he accepted the mysteries of the virgin birth, the resurrection and the divinity of Christ. Toland, and deists generally, departed from Locke in rejecting anything in Scripture that was mysterious or above reason (see Locke, J. §7).

The case for Christianity rested in large part on two claims: that the Old Testament prophecies regarding the messiah were fulfilled in Christ and that the miracles reported in the New Testament, particularly the resurrection of Christ, are well confirmed. If neither claim could be sustained in the court of reason, deism would appear to be intellectually superior to traditional Christianity. At least this is how the deists saw it. Consequently, they directed their attacks at the Old Testament prophecies and the New Testament miracles. In *A Discourse of the Grounds and Reason of the Christian Religion* (1724), which received thirty-five critical replies, Collins argued in some detail that on any literal interpretation of the prophecies it is simply unreasonable to see their fulfilment in the New Testament. In *Discourses on the Miracles of Our Saviour* (1727-29), Woolston attacked the argument based on the New Testament miracles. Arguing that taken literally the miracle stories were full of absurdities, Woolston claimed that even the Church Fathers recognized that they could only be taken figuratively. His work sold widely,

receiving numerous critical replies. He was fined and imprisoned for his efforts.

Tindal's influential work *Christianity as Old as the Creation* (1730) reaffirmed and developed the deists' favourite claim that God had instilled in all human beings sufficient reason to discover the simple truths expressing our duties both towards himself and our fellow human beings. Tindal argued that the fundamental teachings of Christianity express these duties and, by doing so, make Christianity as old as the creation. He proceeded to argue that this religion of nature which God wrote in our hearts from the dawn of creation is the criterion against which the teachings of all religions must be measured. Any purported new revelation must be judged by reason in light of the original religion of nature.

Voltaire, the most prominent of the French deists, came into contact with English deism during his extended stay in London (1726-9). Returning to France, he attacked both the superstition and deceit he felt to be rampant in the Catholic Church and the materialism and atheism of younger philosophers. His own deism involved a belief in God, in absolute moral values and in personal immortality. Rousseau moved from Calvinism to Catholicism and finally to deism. His deism was similar to Voltaire's in stressing belief in God and a future life, but he particularly emphasized the role of conscience in establishing a personal relationship with God.

In eighteenth-century America, the works of English religious thinkers, including deists, were well known. In the latter half of the century Voltaire and Rousseau were read. Franklin, Jefferson, Washington and Thomas Paine were deists, although Paine was far and away the most outspoken. Paine believed deeply in the existence of God and the practice of virtue. Although less concerned with human immortality, he accepted the idea of a future life. He regarded deism as the rational replacement for revealed religion and viewed the rise of atheism as a reaction to the irrationality, fanaticism and political conservatism of revealed religion. To promote deism he undertook to destroy the authority of traditional Christianity by attacking its foundation in the Bible as the revealed word of God. In *The Age of Reason* (1794) he mounted the traditional arguments against the general idea of a supernatural revelation and then focused on the claims for the Bible, arguing that Moses was not the author of the first five books of the Old Testament and that the Gospels were not written by the apostles. He denied that Jesus was divine, claiming that the life of Christ depicted in the Gospels is a fable. The great importance of Paine's book is that it brought the ideas of deism to the general public.

4 The significance of deism

In an age of reason, deism had some advantages over orthodoxy. That there is a God, that we have duties to him and our fellow humans, and that there is a future state of rewards and punishments appeared to many to be truths firmly supported by reason. Indeed, many followers of orthodoxy also held that these three points are well supported by reason. But in the fight over the need for and the existence of special revelation, orthodoxy was put on the defensive. For it was not overly difficult for the deists to point out serious weaknesses in the case for special revelation built on Old Testament prophecies and New Testament miracles. Thus, so long as reason was the only testing ground, deism appeared to have the advantage. But deism was not without problems and excesses of its own. First, deists never could agree on which beliefs were essential. Some rejected immortality, whereas others took it to be central to the religion of nature. Second, the idea that there is a fundamental, universal religion of nature common to all human beings from the dawn of creation turned out to be a myth, wholly unsupported by historical studies. And, finally, as reason's ability to settle such ultimate questions came under serious attack, deism itself was called into question. For, unlike orthodoxy, deism rejected any religious claim that could not be established by reason. Thus, although deistic beliefs (such as belief in a divine reality, belief in moral duties towards others, and so on) survive, deism's belief in the power of reason to disclose to us a universal religion of nature is largely a thing of the past.

See also: Miracles; Natural theology; Providence; Religion and science; Religion, history of philosophy of

WILLIAM L. ROWE

References and further reading

Byrne, P. (1989) *Natural Religion and the Nature of Religion: The Legacy of Deism*, London and New York: Routledge.(Contains a valuable discussion of deism.)

Clarke, S. (1706; 9th edn 1738) *A Discourse Concerning the Being and Attributes of God, the Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation*, British Philosophers and

Theologians of the 17th and 18th Centuries, New York: Garland, 1978.(Clarke's introduction contains his careful classification of four kinds of deist; the Garland reprint is of the ninth (1738) edition.)

Collins, A. (1724) *A Discourse of the Grounds and Reason of the Christian Religion*, British Philosophers and Theologians of the 17th and 18th Centuries, New York: Garland, 1976.(Important historical attack against revelation and prophecies as adequate rational support for Christianity.)

Herbert of Cherbury, E. (1624) *De Veritate*, trans. and ed. M.H. Carre, Bristol: J.W. Arrowsmith, 1937.(Argues for an original, natural religion common to the human race.)

Morais, H.M. (1960) *Deism in Eighteenth Century America*, New York: Russell & Russell.(Useful summary of the views of a number of eighteenth-century American deists.)

Paine, T. (1794) *The Age of Reason*, Indianapolis, IN: Bobbs-Merrill, 1957. (Popular deistic work attacking what Paine took to be the foundations of Christianity in superstitions and myths.)

Stromberg, R.N. (1954) *Religious Liberalism in Eighteenth-Century England*, Oxford: Oxford University Press. (Contains an illuminating discussion of the rise of deism in England.)

Tindal, M. (1730) *Christianity as Old as the Creation*, British Philosophers and Theologians of the 17th and 18th Centuries, New York: Garland, 1978.(Argues that whatever is true in Christianity is apparent to natural reason. A major work in English deism that occasioned considerable controversy.)

Toland, J. (1696) *Christianity not Mysterious*, British Philosophers and Theologians of the 17th and 18th Centuries, New York: Garland, 1978.(Widely read work insisting that the true gospel can contain nothing contrary to reason, nor above it.)

Woolston, T. (1727-9) *Discourses on the Miracles of Our Saviour*, British Philosophers and Theologians of the 17th and 18th Centuries, New York: Garland, 1979.(Popular work arguing that a literal interpretation of the miracles attributed to Jesus leads to absurdities and inconsistencies.)

Deleuze, Gilles (1925-95)

Although grounded in the history of philosophy, Gilles Deleuze's work does not begin with first principles but grasps the philosophical terrain 'in the middle'. This method overthrows subject-object relations in order to initiate a philosophy of difference and chance that is not derived from static being; a philosophy of the event, not of the signifier-signified; a form of content that consists of a complex of forces that are not separable from their form of expression; the assemblage or body without organs, not the organized ego; time, intensity and duration instead of space; in short, a world in constant motion consisting of becomings and encounters with the 'outside' that such concepts do not grasp.

This radical philosophical project is rendered most clearly in Deleuze (and his collaborator Guattari's concept of the 'rhizome'). The rhizome is a multiplicity without any unity that could fix a subject or object. Any point of the rhizome can and must be connected to any other, though in no fixed order and with no homogeneity. It can break or rupture at any point, yet old connections will start up again or new connections will be made; the rhizome's connections thus have the character of a map, not a structural or generative formation. The rhizome, then, is no model, but a 'line of flight' that opens up the route for encounters and makes philosophy into cartography.

1 Histories of philosophy and philosophical works

Born in Paris, Gilles Deleuze spent the period of the Second World War studying philosophy at the Sorbonne. His teachers included Ferdinand Alquié, Georges Canguilhem, Maurice de Gandillac and Jean Hyppolite. Deleuze particularly admired Alquié and Hyppolite, but found them both to be stuck in the history of philosophy. For him, only Sartre escaped that history, and introduced a breath of fresh air into the post-Liberation situation of the intellectual in France. Deleuze's academic career proceeded in the usual manner: Professeur de Lycée, Professeur de l'Université de Provence, researcher at the Centre National des Recherches Scientifiques, Professeur de l'Université de Paris VIII. His public life is perhaps defined more by what did not happen than what did. He seldom travelled outside France, never joined the Communist Party, never embraced existentialism, phenomenology, or Heidegger; but he never renounced Marx, nor was Deleuze himself denounced during the upheavals of May 1968 in France. He died in 1995, by suicide, under conditions of deteriorating health.

By his own account, all the authors in the history of philosophy who interested Deleuze had something in common: they all escaped from the history of philosophy in some respect. Beyond this they had few relationships with one another: the Stoics, Hume, Bergson, Nietzsche, Leibniz, and especially Spinoza. Their likeness had to be produced by a technique that explored what happens between them, in an encounter for which no single philosopher served as a model. In Deleuze's rewritings of the history of philosophy, as well as in his two early philosophical works, *Différence et répétition (Difference and Repetition)* and *Logique du sens (The Logic of Sense)*, what results is the production, not of history at all, but of 'philosophical geography', the line between two ideas forming an ideal space that is neither the one nor the other, but is in flux. Deleuze carries this out by grasping each philosophy 'in the middle' rather than by seeking its first principle.

An example of Deleuze's monstrous production can be found in his engagement with Kant in *La Philosophie critique de Kant (Kant's Critical Philosophy)* which he called 'a book on the enemy'. If Kant harmonizes the faculties in order to ground the architectonic of rationality, Deleuze focuses on how Kant makes possible the *disjunction* of the faculties. This incites the terrible struggle between imagination and reason and understanding and inner sense, a struggle whose importance lies in the notion that this discord produces an accord, even though the faculties are no longer determined by their succession in time or their contiguity in space, and are unregulated, as well, by any law.

Such a reading is no doubt indebted to Deleuze's admiration for Hume: not the Hume whose empiricism is claimed by the binary first principle: sensible ideas-intelligible ideas, but the Hume who substitutes the external and changing relation A *and* B, for the internal and essential relation A *is* B. Such a move realizes philosophical geography by undermining the verb 'to be' and establishing the moving series:*and, and, and*. Of equal importance, the substitution of 'and' for 'is' makes way for the concept of systems, unities or wholes that are open, as opposed to unities that transcend their parts. Finally, since philosophers have by and large begun with the concept of unity and then derived multiplicity from it as its contrary, to think multiplicity as what is, without a closed unity,

Deleuze insists that we have to learn to think in terms of time instead of space. This is Deleuze's view of Bergson's contribution to philosophy, that in the world (and equally in cinema) matter, the movement-image, is produced by the expansion or relaxation of memory-duration, the time-image. Bergson - upon whom, Deleuze remarks, so much hatred was focused within the French university - argued, against Riemann and Einstein, that duration is what differs with itself originally. It is internal difference, a world of intensive magnitudes preserved, virtually coexistent in memory as the repetition of multiple planes, and so not as a numerically multiple exteriority.

Even before Bergson, Deleuze found in Nietzsche the guarantee that memory is the return, not of 'to be', of what is the same, but of becoming and difference. Following Hume, Deleuze argues that, for Nietzsche, all bodies consist of multiple relations between forces; will to power wills becoming and difference when it experiments with forces. When a force extends its power as far as it is able, its will is the expression of the power that it is and, in willing, it affirms difference and chance.

Connecting Hume, Nietzsche, the Stoics, and especially, Spinoza, Deleuze discovered what he called 'a secret link which resides in the critique of negation, the cultivation of joy, the hatred of interiority, the exteriority of forces and relations, the denunciation of power'. What guides this is, in part, a profound and consistent anti-Hegelianism. In *Le Bergsonisme (Bergsonism)* this anti-Hegelianism is oriented by Bergson's objection to the imprecision and generality of Hegel's notions of unity and multiplicity. In *Nietzsche et la philosophie (Nietzsche and Philosophy)* it is guided by Nietzsche's critique of negative slave morality, values based upon a logic that states, 'You are evil; therefore, I am good', and upon the false conception of action as external rather than internal to power. This critique is clarified in Deleuze's work on the Stoics and Spinoza, where it is played out in terms of a conception of the body as multiple and active.

For the Stoics, there are two planes of being: bodies and events. Bodies are beings in depth; they are real, existing in space and temporally, in the present. On the other plane, we find 'incorporeal acts' - sense, the formalization of expression, ideational events at the surface of bodies. Mixtures deep inside bodies are the *causes* of incorporeal events, that is, of sense: becoming green, becoming poisoned. These becomings are the effects of bodily mixtures and are not reducible to bodies. Deleuze's logic of becoming undermines propositional logic, for the attribute is not a quality related to a subject by the indicative 'is'; any verb in the infinitive operates as limitless becoming. This has the additional effect of referring to no subject, no I claims the event which is always referred to simply as 'it'.

Deleuze wrote two books on Spinoza: *Spinoza et le problème de l'expression (Expressionism in Philosophy: Spinoza)* which served as part of his doctoral thesis, and *Spinoza: philosophie pratique (Spinoza: Practical Philosophy)*. For Deleuze's project of geographical philosophy, Spinoza's importance lies in the fact that, practically alone among philosophers, Spinoza asks what a body, understood as an assemblage of affects or becomings, can do, and how this leads to the practical or ethical question of how to increase our power to think, exist and act. The answer is, on Deleuze's account, a matter of a body's capacity to be affected in encounters among bodies, each of which is a dynamic and open multiplicity, an assemblage. Bodies that are compatible increase one another's power to act; incompatible bodies decrease power for one or both of them. As a decrease in power is the normal situation for human beings, so the 'sadness' that characterizes the human condition is the first point of attack. On the practical or ethical level, Deleuze-Spinoza recommends a two-pronged approach: first, devalue sad passions; second, carry out an analysis of the systems of relations of the parts of bodies to determine which relations are compatible. This sparks the production of a geography that moves from merely joyful passive affections, to compatible relations as the cause of joy and finally, to the Spinozan ethical imperative: become active!

2 Collaborations with Félix Guattari

In 1969, Deleuze met Félix Guattari, a practising psychoanalyst and political activist. They collaborated on explosive and radical books of philosophy: *Capitalisme et schizophrénie (Capitalism and Schizophrenia)*, *L'Anti-Oedipe (Anti-Oedipus)*, *Mille Plateaux (A Thousand Plateaus)* and *Qu'est ce que la philosophie? (What is Philosophy?)*. In this collaboration, an assemblage was constructed: 'Since each of us was several, there was already quite a crowd', and the geography of the encounter was practised: 'We are no longer ourselves', 'We have been aided, inspired, multiplied.' *Anti-Oedipus* is less an argument against the idol of psychoanalysis than a

demonstration of the relation between, on the one hand, the herd instinct and the desire to be led that motivates those who profess their faith in Oedipus or in the state; and on the other, state fascism and the fascism we all carry with us.

Deleuze and Guattari address the key question posed by Wilhelm Reich: 'How could the masses be made to desire their own repression?' They hypothesize that fascism is lying in wait wherever the judgment 'He is evil, so I am good' is made and wherever what they call *lines of escape* or deterritorialized flows of desire, have been reduced to state, family or religious hierarchies. In the first case, we recognize what Nietzsche condemned as base valuations that can end only in nihilism. In the second, we find state institutions, hierarchies committed to binary thinking and dedicated to limiting or breaking the connections that allow for the construction of the multiplicity known as the assemblage, as well as heading off encounters between assemblages.

A Thousand Plateaus operates on many planes at once. Thus it serves as an embodiment of the open whole, which Deleuze and Guattari now call the 'rhizome'. Each chapter or 'plateau' of this book can be read independently of any other, and each is accompanied by a date, though in no particular order and not always of importance in any traditional history: the date of Freud's analysis of the Wolf-man, the date the *Reichsmark* was declared to be no longer money, the date of the destruction of the Jewish Temple, the date of vampires. These dates or events take their place in various *semiotic* systems, some of which are linguistic and some not. But they never appear in the despotic linguistics of the signifier-signified relation, wherein all aspects of life are reduced to language. For, even when Saussure's linguistics released the signified (concept) from any constant relation to the word or sound image, it did so only by elevating the signifier, insofar as the relations between signifiers always determine the value (meaning) of any signified (see Saussure, F. de).

If meaning-events are the effects of bodies, then an entirely different relation between bodies and linguistic systems has to be established. If, following Nietzsche, a thing has as many meanings as there are forces capable of seizing it, and if each force consists of a complex of other forces, then in place of the old dualism of substance and form Deleuze and Guattari assert two substance/form complexes: one of content and one of expression. These two assemblages of forces cannot be separated from one another. A distinction between form of expression (an order and organization of functions) and form of content (an order and organization of qualities) can only arise for purposes of analysis: for example, in the encounter between a nomad war machine (form of expression) and itinerant metallurgy (form of content). Nor is the encounter necessarily or even primarily linguistic - linguistics is merely one of many semiotics, and not the most important.

The role of language and the task of philosophy regarding language are also crucial to *What is Philosophy?*, the end of the experiment between Deleuze and Guattari, an end sealed by Guattari's death in 1992. In this book, the Stoic treatment of bodies and events implicit throughout *A Thousand Plateaus* re-emerges forcefully, as Deleuze and Guattari argue that the task of philosophy - as opposed to science - is the creation of concepts, the extraction of events from beings and things so as to create a plane of immanence, a plane of consistency. For philosophy does not seek an external plane of reference or truth, as science does. The distinction between bodies and events is important here, for while philosophical concepts are constituted from immanent variations of consistent events, scientific functions refer to states of affairs or mixtures in bodies. Thus, while philosophical concepts are enunciated by conceptual personae (the friend, the stammerer) that are not the philosopher, but that trace out the line of becoming that is the concept, scientific functions are enunciated by the scientific observer situated within the perspective of a particular state of affairs (the relativity traveller).

3 Minor literature and nomad arts

Given Deleuze's interest in the history of philosophy and in social and political philosophy, his careful attention to literature and the arts may be unexpected. Yet, his two books on cinema, books on Sacher-Masoch and Sade, Kafka, Proust and Francis Bacon, as well as articles and constant references to a host of literary, musical, and artistic works and persons all evidence the central role of literature and the arts to philosophical geography. Thus, there is *Le Pli: Leibniz et la baroque (The Fold, Leibniz and the Baroque)* celebrating the pleat, the curve, the twisting surface of science and mathematics, but also of music, architecture, the line and the word; a sensuous view of the world that lives in the work of art. Nomad arts are blocs of sensations or a 'fold'. Each is a compound of percepts and affects wrested from the subject and its states by means of the material of each art. By means of its material, words and syntax, affective writing is the becoming-multiple that deterritorializes a major language and

produces minor writing, which is the composing and decomposing of saturated percepts, rather than a record of memories, fantasies or travels. Thus, everything that is written takes its place within the open whole which is the social world. Likewise, cinema is sensation conditioned by cinema's plastic mass of visual material that puts into operation the open whole as both incessant flux and instantaneous disjunction. In place of language, art gives us percepts, affects and blocs of sensation put into motion. This is not philosophy, yet it makes philosophy happen by giving the event a body, a life, a universe.

See also: Alterity and identity, postmodern theories of §3; Selfhood, postmodern critique of

DOROTHEA E. OLKOWSKI

List of works

Deleuze, G. and Cresson, A. (1952) *David Hume: sa vie, son oeuvre, avec exposé de sa philosophie*, Paris: Presses Universitaires de France.(Deleuze's early account of Hume's empiricism and Hume's life.)

Deleuze, G. (1953) *Empirisme et subjectivité: essai sur la nature humaine selon Hume*, Paris: Presses Universitaires de France; trans. C.V. Boundas, *Empiricism and Subjectivity: An Essay on Hume's Theory of Human Nature*, New York: Columbia University Press, 1991.(Approaches Hume's empiricism as a doctrine of ideas that are external to one another thus always differing from one another while qualifying and making a subject of the mind that contemplates them.)

Deleuze, G. (1956) 'La conception de la différence chez Bergson', *Les Études Bergsoniennes* 4: 77-112. (Differentiates Bergson's ontology of difference from Hegel and Plato by separating differences of degree from difference in nature or kind and by arguing that repetition is difference.)

Deleuze, G. (1962) *Nietzsche et la philosophie*, Paris: Presses Universitaires de France; trans. H. Tomlinson, *Nietzsche and Philosophy*, New York: Columbia University Press, 1983.(Reads Nietzsche as the overthrow of Platonism that affirms becoming and change and rejects Ideas as a priori norms or concepts.)

Deleuze, G. (1963) *La Philosophie critique de Kant: doctrines des facultés*, Paris: Presses Universitaires de France; trans. H. Tomlinson and B. Habberjam, *Kant's Critical Philosophy: The Doctrine of the Faculties*, Minneapolis, MN: University of Minnesota Press, 1984.(Rejects the Kantian faculty's regulation of the senses, thought and morality in favour of the Kantian sublime as the unregulated exercise of all the faculties.)

Deleuze, G. (1966) *Le Bergsonisme*, Paris: Presses Universitaires de France; trans. H. Tomlinson and B. Habberjam, *Bergsonism*, New York: Zone Books, 1988.(Argues that duration is the qualitative and heterogeneous becoming operating in all life and the virtual and creative process of unconditioned change.)

Deleuze, G. (1968a) *Différence et répétition*, Paris: Presses Universitaires de France; trans. P. Patton, *Difference and Repetition*, New York: Columbia University Press, 1994.(A sustained reading of the history of philosophy that defines repetition as difference and not as representation.)

Deleuze, G. (1968b) *Spinoza et le problème de l'expression*, Paris: Éditions de Minuit; trans. M. Joughin, *Expressionism in Philosophy: Spinoza*, New York: Zone Books, 1990.(Finds Spinoza's concept of substance to be a concept of the One as an open-ended and differentiated whole that expresses itself by means of an infinity of attributes in modes that unfold from the One.)

Deleuze, G. (1969) *Logique du sens*, Paris: Éditions de Minuit; trans. M. Lester and C. Stivale, ed. C.V. Boundas, *The Logic of Sense*, New York: Columbia University Press, 1990.(Revives the Stoic conception of logic which articulates the event as an effect of bodily mixtures and as the surface between bodies and language.)

Deleuze, G. (1970) *Spinoza: philosophie pratique*, Paris: Éditions de Minuit; trans. R. Hurley, *Spinoza: Practical Philosophy*, San Francisco, CA: City Lights Books, 1988.(Opposes Spinoza's Ethical philosophy to moral thought.)

Deleuze, G. and Guattari, F. (1972) *Capitalisme et schizophrénie*, vol. 1, *L'Anti-Oedipe*, Paris: Éditions de Minuit, 2nd enlarged edn, 1980; trans. R. Hurley, M. Seem and H.R. Lane, *Capitalism and Schizophrenia*, vol. 1, *Anti-Oedipus*, New York: Viking Press, 1977, reprinted Minneapolis, MN: University of Minnesota Press, 1983.(A revolutionary reconception of desire as productive and positive social force that reconfigures the Oedipal psychoanalysis of lack as a schizoanalysis of creative flows.)

Deleuze, G. and Guattari, F. (1980) *Capitalisme et schizophrénie*, vol. 2, *Mille Plateaux*, Paris: Éditions de Minuit; trans. B. Massumi, *A Thousand Plateaus*, Minneapolis, MN: University of Minnesota Press, 1987. (Fifteen chapters, called plateaus, traversing traditional disciplines and analyses. Each plateau develops its own concepts to construct new thematics in place of the traditional ones.)

Deleuze, G. (1981) *Francis Bacon: logique de la sensation*, Paris: Éditions de la Différence, 2 vols; trans. D. Smith, *Francis Bacon: The Logic of Sensation*, Cambridge, MA: MIT Press, forthcoming.(A study of the work of the artist Francis Bacon as a painter who frees the figure from representation to render sensation in and of

itself.)

Deleuze, G. (1986) *Foucault*, Paris: Éditions de Minuit; trans. S. Hand, *Foucault*, Minneapolis, MN: University of Minnesota Press, 1988.(Reads the work of Michel Foucault as a new functionalism that produces the topology of diffuse and local rather than globalized power.)

Deleuze, G. (1988) *Le Pli: Leibniz et le baroque*, Paris: Éditions de Minuit; trans. T. Conley, *The Fold: Leibniz and the Baroque*, Minneapolis, MN: University of Minnesota Press, 1993.(Develops the fold as an anti-extentional concept of the multiple, an anitdialectical concept of the event and an anti-Cartesian concept of the subject.)

Deleuze, G. and Guattari, F. (1991) *Qu'est-ce que la philosophie?*, Paris: Éditions de Minuit; trans. H. Tomlinson and G. Burchell, *What Is Philosophy?*, New York: Columbia University Press, 1994.(The last work co-written by Deleuze and Guattari, differenciates philosophy from science, logic and art.)

References and further reading

Bogue, R. (1989) *Deleuze and Guattari*, New York: Routledge. (An introduction to Deleuze's best-known work including *A Thousand Plateaus*.)

Boundas, C.V. (1993a) 'Introduction', in C.V. Boundas (ed.) *The Deleuze Reader*, New York: Columbia University Press. (An outline of all of Deleuze's work, including the historical texts.)

Boundas, C.V. (ed.) (1993b) *The Thought of Gilles Deleuze*, special issue of *Journal of the British Society for Phenomenology* 24 (1). (Selected critical essays.)

Boundas, C.V. and Olkowski, D. (1994) *Gilles Deleuze and the Theatre of Philosophy*, New York: Routledge. (A collection of critical and explanatory essays covering all phases of Deleuze's work by philosophical, literary and cultural critics.)

Foucault, M. (1977) 'Theatrum Philosophicum', in *Language, Counter-Memory, Practice*, ed. D. Bouchard, Ithaca, NY: Cornell University Press. (Focuses on Deleuze's innovative thinking in *The Logic of Sense* and *Difference and Repetition*.)

Goodchild, P. (1994) *Deleuze and Guattari: An Introduction to the Politics of Desire*, London: Sage.

Hardt, M. (1993) *Gilles Deleuze, An Apprenticeship in Philosophy*, Minneapolis, MN: University of Minnesota Press. (A detailed examination of Deleuze's work on Nietzsche, Bergson and Spinoza articulating Deleuze's anti-Hegelianism and the influence of scholastic philosophy.)

L'Arc (1972) *Deleuze*, special issue of *L'Arc* 49; revised edn, 1980.(Essays on Deleuze by French scholars including Clément, Gandillac and Klossowski, and an interview with Foucault by Deleuze.)

Martin, J.-C. (1993) *Variations: la philosophie de Gilles Deleuze*, Paris: Éditions Payot; trans. C.V. Boundas, *Variations: The Philosophy of Gilles Deleuze*, Atlantic Highlands, NJ: Humanities Press, 1997. (A complex and knowledgeable treatment of multiplicity from an aesthetic point of view.)

Massumi, B. (1992) *A User's Guide to Capitalism and Schizophrenia: Deviations from Deleuze and Guattari*, Cambridge, MA: MIT Press. (A practical elaboration of the two volumes from a postmodern cultural perspective.)

Olkowski, D. (forthcoming 1998) *The Ruins of Representation*, Berkeley, CA: University of California Press. (A feminist-oriented account of Deleuze's ontology of becoming through Hume and Bergson.)

SemioText(e) (1977) *'Anti-Oedipus'*, special issue of *SemioText(e)* 2 (3).(Special issue devoted to essays on *Anti-Oedipus* including several translations of the commentaries of French scholars.)

SubStance (1984) *Gilles Deleuze*, special issue of *SubStance* 44/5. (Special issue devoted to essays on Deleuze, focuses especially on literary criticism.)

Delmedigo, Elijah (*c*.1460-93)

Throughout the treatises and translations commissioned by his many patrons in Italy, Elijah Delmedigo championed Aristotle and Ibn Rushd (Averroes). In Latin texts prepared for Pico della Mirandola, Delmedigo affirmed such cardinal Averroist notions as the absolute unity of all human minds and the role of God as the unmeditated principle of intelligibility in the universe. In the Hebrew Behinat ha-Dat (The Examination of Religion), Delmedigo urges the superiority of a rationalistic Judaism over other religions, especially Christianity, and over the nonphilosophic, improperly philosophic and antiphilosophic versions of Judaism. Sections of this work amount to a nuanced critique of Kabbalah. Combining ardent Averroism with qualified admiration for Maimonides, Delmedigo repeatedly argued for the compatibility of Judaism with secular philosophic speculation.

Born to a distinguished family of rabbinic scholars and intellectuals in Candia on the island of Crete, Delmedigo migrated to Italy in 1480. He spent the next ten years expounding and defending Ibn Rushd's controversial version of Aristotelian philosophy (see Ibn Rushd; Averroism, Jewish). Residing mainly in Padua, Venice, Bassano and Florence, Delmedigo was commissioned by such figures as Girolamo Donato, Domenico Grimani, Antonino Pizzamano and the Neoplatonic enthusiast of Kabbalah, Giovanni Pico della Mirandola, to translate Hebrew versions of Ibn Rushd's texts into Latin and to write systematic accounts of the latter's views. Seeking a wider audience, Delmedigo translated at least three of his own Latin works into Hebrew: two treatises dealing with the unicity of the hylic intellect and conjunction with the 'active intellect' - *Shete she'eloth 'al ha-Nefesh* or *Shene derushim 'al Hanefesh (Two Questions on the Soul)* - and *Ma'amar 'al 'etzem ha-Galgal*, his exposition of Ibn Rushd's *De Substantia Orbis (On the Substance of the Celestial Sphere)*. Slavishly literal, these Hebrew translations differ from Delmedigo's Latin originals only by the interpolation of autobiographical data and discussions of religious topics that Delmedigo considered relevant to the Jewish community.

In the late 1490s, Delmedigo returned to Crete and began writing *Behinat ha-Dat*. He faulted Christianity for its irrational dogmas and excoriated Jewish mystics for their excessively Neoplatonic metaphysics and 'idolatrous' interpretations of Jewish ritual observance. He also chided Maimonides and other unspecified rationalist theologians for confounding religious faith with philosophic speculation. Framing his criticism of the revered Maimonides diplomatically, he blamed him and the others for undermining social unity with their rationalist interpretations of religion. To avoid sectarian strife, Delmedigo counselled against publication of written doctrinal interpretations of Scripture and the law. He also argued that it was imprudent to convey the mistaken impression that religious belief and practice are rationally grounded rather than prophetically instituted.

Delmedigo was nevertheless eager to establish the compatibility of Judaism and philosophy. He emphasized the congenial parallel between Talmudic logic and philosophic argumentation, explaining that just as the rabbis use a set of well-defined principles for drawing inferences from Scripture, so the philosophers use syllogisms for reaching demonstrative truths. Delmedigo was intent on mitigating conflicts within the Jewish community. He balanced his trenchant criticisms of the Kabbalah by lacing *Behinat ha-Dat* with conciliatory remarks addressed to the Jewish mystics, highlighting the accords of Kabbalah, rational belief and popular religion.

Averroists are notoriously difficult to pin down; their utterances often seem inconsistent or contradictory. Their explications of Aristotle can be read as detached scholarship or partisan advocacy. They are known for tolerating what they deem to be mistaken popular beliefs that preserve the social order. Their overriding concern sometimes seems to be to make society safe for philosophy. The elusive Delmedigo was no exception. Most of his writings are not systematic statements of his own philosophy. Rather, they are secular commentaries and polemical treatises expounding Aristotle as interpreted by Ibn Rushd, who was, in turn, often misunderstood by the likes of John of Jandun, a Latin Averroist whose many mistakes in interpretation Delmedigo seldom failed to mention and correct. Extracting Delmedigo's own convictions regarding physics, metaphysics and political philosophy from exegetical sources of this kind is difficult. *Behinat ha-Dat* is not as direct a help as we might wish. It was motivated as much by a diplomat's eirenic 'love for the words of our Torah, for our sages, and for the people of our nation' as it was by a medieval philosopher's austere devotion to logic, reason and truth. Knowing how anxious Delmedigo was to keep the peace makes it hard to determine where his devotion to Aristotelian philosophy ends and his love for Judaism begins.

Regardless of motive, Delmedigo unequivocally declared Aristotle the supreme philosopher and Ibn Rushd his only trustworthy commentator. With equal tenacity he affirmed philosophy and Judaism to be compatible, but only ‘when they are properly understood’. These philosophical commitments put him at loggerheads with most of his contemporaries. Expounding Ibn Rushd’s controversial monopsychism and its corollary, impersonal and unindividuated immortality, Delmedigo affirmed two of the most notorious doctrines against which Thomas Aquinas and other theologians had argued vehemently. Defending Ibn Rushd’s understanding of God as a self-sufficient mind whose rationality constitutes the unmediated principle of intelligibility and regularity in the universe, Delmedigo staked out his position among a heroic minority. Arrayed against him were Italian Neoplatonists and Kabbalists who embraced the metaphysics of emanation. Opposed to him also were obscurantists, who ridiculed philosophy, and scholastic theologians of all creeds, who defended radically voluntaristic doctrines of God’s utter transcendence, unfettered will and direct awareness of particulars.

Delmedigo’s enduring contributions to philosophy are twofold: he defended the secular autonomy of philosophy, its right to follow its own course, free of theological preconditions; and, if we take him at his word, he legitimized philosophy by stressing its power to establish and maintain civil order - without which philosophy itself would be unattainable.

See also: Averroism; Averroism, Jewish; Ibn Rushd; Kabbalah; Maimonides, M.; Messer Leon, J.; Pico della Mirandola, G.

KALMAN BLAND

List of works

Delmedigo, E. (1482) *Shete she’eloth ‘al ha-Nefesh* or *Shene derushim ‘al Hanefesh (Two Questions on the Soul)*, MS Milan: Biblioteca Ambrosiana, 128, and MS Paris: Bibliothèque Nationale, Hebrew 968.(Delmedigo’s Hebrew translations of his own Latin treatises (no longer extant), composed originally for Pico della Mirandola, the first dealing with the unicity of intellect, the second dealing with the conjunction of the human intellect with the active intellect.)

Delmedigo, E. (1485) *Ma’amar ‘al ‘etzem ha-Galgal (Commentary on De Substantia Orbis)*, MS Paris: Bibliothèque Nationale, Hebrew 968.(Delmedigo’s translation of and supercommentary to Ibn Rushd’s treatise.)

Delmedigo, E. (1486) *Letter to Pico della Mirandola*, MS Paris: Bibliothèque Nationale, Latin 6508.(Delmedigo’s personal letter to Pico on various personal and philosophic issues, with an important section of the purchase of Kabbalistic books and their doctrinal meaning.)

Delmedigo, E. (1488) ‘Quaestiones: De Primo Motore, De Mundi Efficientia, De Esse et Uno’ (‘Questions: On the Prime Mover; On the World’s Efficiencies; On Being and Unity’), in John of Jandun, *Super octo libros Aristotelis de physico auditu subtilissimae quaestiones (On the Eight Books of Physics by Aristotle)*, Frankfurt: Minerva, 1969.(Delmedigo’s analytic discussions and paraphrases of themes in Aristotelian metaphysics and physics.)

Delmedigo, E. (*c*.1480-90) *Behinat ha-Dat (The Examination of Religion)*, ed. with notes by J.J. Ross, *Sefer Behinat Ha-Dat of Elijah Del-Medigo*, Tel Aviv: Chaim Rosenberg School of Jewish Studies, 1984.(Critical edition of the Hebrew original. Delmedigo’s treatise on the philosophy of religion, critique of Christianity and Kabbalah, and the role of rationality in revealed traditions.)

Delmedigo, E. (*c*.1480-90) *Parafrasi Della Repubblica Nella Traduzione Latina Di Elia Del Medigo (Latin Paraphrase by Elia Del Medigo of the Republic)*, ed. A. Coviello and P.E. Fornaciari, *Delmedigo’s Latin Translation of Ibn Rushd’s Paraphrase and Commentary to Plato’s Republic*, Florence: Leo S. Olschki Editore, 1994.(Delmedigo’s Latin translation of Ibn Rushd’s paraphrase and commentary to Plato’s *Republic*.)

Delmedigo, E. (*c*.1480-90) *In Meteorologica Aristotelis (Epitome of Aristotle’s Meteorology)*, MS Vatican 4550.(Delmedigo’s Latin translation of Ibn Rushd’s *Epitome* of Aristotle’s *Meteorologica*.)

Delmedigo, E. (*c*.1480-90) *Aristotle’s Book Lambda of the Metaphysics*, MS Paris: Bibliothèque Nationale, Latin 6508.(Delmedigo’s Latin translation of the introduction to Book 12 of *Mah she’ahar ha-Tev‘a*, Ibn Rushd’s Hebrew version of Aristotle’s *Metaphysics*.)

Delmedigo, E. (*c*.1480-90) *Averrois Commentatio in Metaphysica Aristotelis (Epitome of Aristotle’s Metaphysics)*, MS Paris: Bibliothèque Nationale, Latin 6508.(Delmedigo’s Latin translation of Ibn Rushd’s *Epitome of Mah she’ahar ha-Tev‘a (Metaphysics)*.)

References and further reading

Bland, K.P. (1995) ‘Elijah Del Medigo, Unicity of Intellect, and Immortality of Soul’, *Proceedings of the American Academy for Jewish Research* 61: 1-22.(A description of Delmedigo’s literary style, attitude toward faith and reason, and treatment of Averroist psychology.)

Bland, K.P. (1991) ‘Elijah del Medigo’s Averroist response to the Kabbalahs of fifteenth-century Jewry and Pico della Mirandola’, *The Journal of Jewish Thought and Philosophy* 1: 23-53.(Focuses on the often discussed relationship to Pico and on Delmedigo’s diplomatic critique of Jewish mysticism; footnotes cite the modern scholarship on Delmedigo.)

Geffen, D. (1973/4) ‘Insights into the Life and Thought of Elijah Medigo Based on his Published and Unpublished Works’, *Proceedings of the American Academy for Jewish Research* 61/2: 69-86.(Based on an unpublished doctoral dissertation which contains an English translation of *Behinat ha-Dat*; the footnotes lead the reader deeply into Delmedigo’s world.)

Demarcation problem

The problem of demarcation is to distinguish science from nonscientific disciplines that also purport to make true claims about the world. Various criteria have been proposed by philosophers of science, including that science, unlike 'non-science', (1) is empirical, (2) seeks certainty, (3) proceeds by the use of a scientific method, (4) describes the observable world, not an unobservable one, and (5) is cumulative and progressive.

Philosophers of science offer conflicting viewpoints concerning these criteria. Some reject one or more completely. For example, while many accept the idea that science is empirical, rationalists reject it, at least for fundamental principles regarding space, matter and motion. Even among empiricists differences emerge, for example between those who advocate that scientific principles must be verifiable and those who deny that this is possible, claiming that falsifiability is all that is required.

Some version of each of these five criteria - considered as goals to be achieved - may be defensible.

1 What is the problem and why is it raised?

'The problem of demarcation' is an expression introduced by Popper to refer to ' the problem of finding a criterion which would enable us to distinguish between the empirical sciences on the one hand, and mathematics and logic as well as "metaphysical" systems on the other' ([1934] 1959: 34). Although Popper mentions mathematics and logic, other writers focus on distinguishing science from metaphysics and pseudoscience.

Some, including Popper, raise the problem because of an intellectual desire to clarify this distinction. Logical positivists had in addition the aim of overthrowing non-scientific disciplines such as metaphysics and theology that purport to describe the physical world but, being unverifiable, are (they claimed) lacking in meaning. Others have more practical aims. In a country such as the USA, which officially attempts to separate church and state, religion is not to be taught in the public schools, but science is. So the practical question becomes what to count as science (for example, is 'creation science' appropriately named?).

2 'Science is empirical'

Metaphysics, philosophy and theology are not. What this means, however, can vary from one view to another. On one typical account, science consists of propositions (such as the law of conservation of energy) that are either directly verifiable by sensory observation, or can be inferred from such observations by the use of inductive reasoning from observed members of a class to all members, or are derivable from either of these using deduction. By contrast, the propositions of metaphysics (for example, 'Platonic universals exist') fail to satisfy these criteria. Newton, who held such an empiricist view, claimed to derive his three laws of motion and law of gravity from observed phenomena using induction and deduction (see Newton, I. §4). Similar views have been held by many empiricists, including J.S. Mill and the logical positivists. A significantly different empiricist view is espoused by Popper, who, following Hume, regards inductive reasoning as lacking in logical justification. For Popper, although scientific laws cannot be verified inductively, they can be falsified by observing a single negative instance; accordingly, falsifiability by observation, rather than verifiability, is what makes science empirical (see Popper K.R. §2).

Although the vast majority of influential writers on science have held that scientific propositions are empirical, there are significant exceptions, especially as regards fundamental principles involving space, matter, and motion. Descartes believed that the most basic principles of physics (no less than those of metaphysics) are knowable a priori by the use of rational intuitions and deductions from them (see Descartes, R. §7). Other a priorists include Kant, who asserted that these principles are synthetic a priori (see Kant, I. §7), and Poincaré, who defended them as conventions of language (see Poincaré, H. §§3-4). Whewell suggested a mixed view according to which, although scientists must first arrive at propositions empirically, as their scientific ideas become clarified they realize that these propositions are necessary and knowable a priori (see Whewell, W.).

3 'Science seeks certainty'

In this respect it is like mathematics and unlike metaphysics, theology and astrology, which, it is alleged, can never

be more than speculative. In his *Regulae ad directionem ingenii (Rules for the Direction of the Mind)* (1620-*c*.9) Descartes' second rule is that 'we must occupy ourselves only with those objects that our intellectual powers appear competent to know certainly and indubitably'. Empiricists are more divided. Newton, who rejected Descartes' idea that science should aim at indubitable propositions, recognizes in his fourth rule of scientific method that any proposition, no matter how strongly supported, is liable to exception as further phenomena become observed. Nevertheless, scientists should always aim at the highest certainty possible in an empirical endeavour, which is obtainable by 'deducing propositions from phenomena and rendering them general by induction'.

Among empiricists at the other extreme from Newton are Popper and Laudan. For Popper a high degree of certainty is not possible in science since the use of any inductive generalizations that might generate such certainty is unjustified. Nor is it desirable, since scientists should make the strongest possible and hence the least likely generalizations. For Laudan (1977) science aims at providing 'adequate' solutions to 'interesting' problems, for which questions of truth, certainty and even probability are irrelevant.

An empiricist view that lies between these extremes is taken by Carnap and other probabilists. Scientists should aim at providing empirical evidence for a theory that increases its probability without necessarily rendering it high (see Achinstein 1983 for a critique).

4 'Scientists follow a scientific method'

Practitioners of non-sciences do not. A scientific method is a set of rules scientists should follow in discovering and testing laws and theories. Whether there are such rules and if so, what forms they take, whether they are universal for all sciences or even within a science, whether they change from one period to another, are all hotly disputed questions (see Scientific method).

On one view, there are rules for testing scientific theories that hold for all science during all periods. This view was espoused by Descartes, who proposed twenty-one such rules; it was also championed by the empiricist Newton (1687), who proposed four 'Rules of Reasoning in [natural] philosophy', comprising two for inferring causes of things, and two for making inductive generalizations from observed phenomena.

The two most prominent empiricist positions committed to a universal scientific method are hypothetico-deductivism and inductivism. According to the former, the scientist confronted with data and a problem begins by proposing an hypothesis, which is not inferred deductively or inductively from the data or anything else, but is simply offered as a conjecture. From it, and possibly other assumptions, observational conclusions are generated deductively, using logic and frequently mathematics. If the conclusions are established by observation the hypothesis is tentatively accepted. If they are found to be false, the hypothesis is rejected and a new one proposed. This is Popper's view.

By contrast, inductivists require one further step, an inductive argument giving independent support to the hypothesis or theory. This involves arguing from observed instances of a law to all cases or from similar effects to similar causes. Inductivists such as Newton and Mill reject the hypothetico-deductive method on the grounds that numerous conflicting hypotheses can entail the same data; what is required is independent inductive support for one of these.

The existence of a universal scientific method has been challenged by several twentieth-century writers, particularly since the 1960s. Thomas Kuhn (1962), arguing for a historical, relativistic approach, asserts that scientists at a given time work within a 'paradigm' which consists of a set of concepts, practices, standards of evaluation, rules of reasoning and methods of observation that vary considerably from one science and period to another. The paradigm sets the problems to be solved and the methods for doing so. There is no scientific method common to all paradigms.

Finally, a sociological approach to science has been advocated (see, for example, Pinch 1986), a strong version of which rejects appeal to methodological rules in explaining how scientists proceed. Theories, being underdetermined by the data (see Underdetermination), are not inferred from those data by using rules. One must look instead at social factors within the scientific community that explain how a scientific theory is developed and how the group 'negotiated' to accept it (see Constructivism).

5 ‘Science describes only the observable world’

Metaphysics, theology and even mathematics describe worlds underlying, or beyond, or independent of, what can be observed. This is a view upheld by instrumentalists and other ’antirealists’ and rejected by realists. Antirealists claim that the aim of science is to ’save the phenomena’ by formulating theories that will correctly predict what is observable. Some antirealists, for example, Duhem and van Fraassen, believe that an unobservable world exists, but deny that the aim of science is to describe it. For Duhem that is the aim of metaphysics and theology (see Duhem, P.M.M. §4). Other antirealists, for example, logical positivists of the 1950s and 1960s, held a strong form of instrumentalism, according to which claims about unobservables in science are to be construed, not as speaking of real entities, but as linguistic devices in a theory for generating claims about what can be observed. For these writers both science and metaphysics introduce terms for unobservables that do not denote objects in the world. The difference between the two disciplines is that scientists but not metaphysicians tie these terms to observables in such a way as to allow observational predictions.

Realists, by contrast, claim that there is a physical world that exists independently of our minds, our observations and our theories. This world contains not only things and events scientists can observe (for example, the planets), but ones they cannot (for example, quarks). The job of the scientist is to describe both the observable and the unobservable worlds. To accept a theory involves a commitment that what it says about both is true. For the realist the distinction between science and metaphysics is not to be drawn by saying that science, but not metaphysics, describes the observable (see Scientific realism and antirealism).

6 ‘Science is cumulative and progressive’

Metaphysics, theology and philosophy are not. The history of philosophy, for example, is just the history of one theory after another; the more different the new theory from the old, the more it conflicts with it, the better its chance of being taking seriously. Science, by contrast, develops in an incremental, progressive way, so as to retain and build upon much of what has come before. In addition, as science advances it yields more and more truths about the world (or saves more and more phenomena); philosophy does not.

This idea, forcefully defended by logical positivists, came under strong challenge in the 1960s particularly from Feyerabend and Kuhn. Both argued that a predecessor theory is usually not derivable from a successor. Sometimes, as Feyerabend (1993) noted, the two are not even logically compatible. Sometimes the two are ‘incommensurable’: they use concepts that completely depend for their meanings on different theories and that cannot be translated from one theory to the other. For Kuhn this occurs when, as in Newtonian and relativistic mechanics, the theories are parts of different paradigms (see Incommensurability).

7 Is a criterion of demarcation possible?

One might suppose not, in view of such controversies over what constitutes science. Indeed, even the need for a demarcation criterion has been challenged by some philosophers, especially ones who think that philosophy itself is or ought to be construed as part of science (for example Quine 1973).

Nevertheless, we do tend to distinguish science from other disciplines. Newton’s mechanics is a scientific theory, but his theory of scientific method is not; it is a philosophical theory. One helpful step might involve taking science, or even a particular science such as physics, at a given historical time and saying something general about the goals of practitioners at the time, while recognizing that these may change and that not all scientists accept them. With this in mind, we might say that the previously quoted claims reflect scientific goals to be achieved at some point.

For example, in modern physics as well as other sciences, a goal is to present empirically testable ideas - even if at the moment (as with superstring theory) it is not known how to test the ideas empirically. For J.J. Thomson, who in 1897 claimed that cathode rays are charged particles not waves, it was not enough simply to postulate that this is so. He wanted to show how in principle this claim could be tested empirically by demonstrating that the rays can be electrically deflected, something that had not previously been accomplished. The goal of making hypotheses empirically testable is one generally shared by physicists but not by philosophers. Physicists can be just as speculative as metaphysicians, but they cannot be so for too long.

Another goal is to achieve certainty, at least as much as possible, by actually carrying out such tests. Although philosophers, theologians and mathematicians may also crave certainty, their route to it is a priori argumentation, not empirical testing. (For a contrary viewpoint regarding the status of philosophy of science, see Donovan *et al.* 1988) This is related to two further goals. Even if both physicists and metaphysicians postulate 'unobservable' entities, one of the aims of physics, but not metaphysics, is to eventually render them 'observable' - to detect and measure them. It was not sufficient for Thomson simply to postulate the existence of the particles (electrons), or even to describe an experimental procedure for electrically deflecting them and measuring their mass to charge ratio. He also attempted (successfully) to perform the experiment. Nor was this the end of the matter. Thomson sought to *argue* for his theory of charged particles from the results of his experiments. He did so, as Newton and Mill suggest, using both deduction and inductive generalization (see Achinstein 1991). Achieving as much certainty as possible through empirical testing requires the use of logical arguments in accordance with a 'scientific method'. Without such arguments mere 'negotiation' within a group is not enough.

Finally, physics today is much more cumulative and progressive than metaphysics or other fields of philosophy. A physicist may present completely new ideas, but in working out consequences it is desirable to make use of general principles, such as conservation laws, that are empirically established. Otherwise deductions to testable conclusions would be much rarer than they are. Even in contemporary philosophy, by contrast, few if any general principles are considered by most philosophers to be established.

Assuming these represent goals of many physicists (but not of philosophers or other 'non-scientists'), a job for philosophers of science is to provide a critical, systematic account of them. Philosophers of science may also try to determine the extent to which these goals, so clarified and systematized, can be attributed to scientists in other fields and at different times.

See also: Logical positivism §5

PETER ACHINSTEIN

References and further reading

Achinstein, P. (1983) *The Nature of Explanation*, New York: Oxford University Press.(A defence of a pragmatic concept of explanation; chapters 10 and 11 discuss the concept of scientific evidence. See §3.)

Achinstein, P. (1991) *Particles and Waves*, New York: Oxford University Press.(This work discusses methodological issues in nineteenth-century physics, including the discovery of the electron by J.J. Thomson. See §7.)

Carnap, R. (1962) *Logical Foundations of Probability*, Chicago, IL: University of Chicago Press.(An important and influential work that defends an a priori account of probability and an increase-in-probability definition of evidence. See §3.)

Descartes, R. (1988) *Selected Philosophical Writings*, New York: Cambridge University Press.(See particularly the work 'Rules for the Direction of the Mind'. See §§2-4.)

Donovan, A., Laudan, L. and Laudan, R. (eds) (1988) *Scrutinizing Science*, Dordrecht: Kluwer.(Contributors to this volume offer empirical tests for the validity of various scientific methods.)

Duhem, P. (1991) *The Aim and Structure of Physical Theory*, Princeton, NJ: Princeton University Press.(A modern classic in the philosophy of science.)

Feyerabend, P.K. (1993) *Against Method*, London: Verso.(A controversial but always interesting attack on traditional views of rationality in science.)

Fraassen, B. van (1980) *The Scientific Image*, Oxford: Oxford University Press. (Defence of antirealism.)

Kant, I. (1786) *Metaphysical Foundations of Natural Science*, Indianapolis, IN: Bobbs-Merrill, 1970.(See particularly chapter 3 'Metaphysical Foundations of Mechanics'.)

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(One of the most influential works on the nature of science in the twentieth century.)

Laudan, L. (1977) *Progress and its Problems*, Berkeley, CA: University of California Press.(Very readable defence of a problem-solving approach to science.)

Mill, J.S. (1843) *A System of Logic*, London: Longmans, 1959.(Book III of this classic contains Mill's defence of induction and his attack on the hypothetico-deductive method.)

Newton, I. (1687) *Philosophiae Naturalis Principia Mathematica*, Berkeley, CA: University of California Press, 1966.(Book III begins with Newton's four rules of scientific method.)

Pinch, T. (1986) *Confronting Nature: The Sociology of the Solar-Neutrino Detection*, Dordrecht: Kluwer.(A

sociological approach to scientific discovery and change.)
Poincaré, H. (1902) *Science and Hypothesis*, New York: Dover, 1962.(Classic defence of a conventionalist approach to science.)
Popper, K. (1934) *The Logic of Scientific Discovery*, London: Hutchinson, 1959. (The most famous twentieth-century defence of hypothetico-deductivism and the idea that science is falsifiable but not verifiable.)
Quine, W.V. (1973) *The Roots of Reference*, La Salle, IL: Open Court.(The 1971 Paul Carus Lectures.)
Snyder, L. (1994) 'It's *All* Necessarily So: William Whewell on Scientific Truth', *Studies in History and Philosophy of Science* 25, 785-807.(A novel interpretation of Whewell's a priorism.)
Whewell, W. (1847) *The Philosophy of the Inductive Sciences*, New York: Johnson Reprint Corporation, 1967. (One of the great works of nineteenth-century philosophy of science.)

Democracy

Democracy means rule by the people, as contrasted with rule by a special person or group. It is a system of decision making in which everyone who belongs to the political organism making the decision is actually or potentially involved. They all have equal power. There have been competing conceptions about what this involves. On one conception this means that everyone should participate in making the decision themselves, which should emerge from a full discussion. On another conception, it means that everyone should be able to vote between proposals or for representatives who will be entrusted with making the decision; the proposal or representative with most votes wins.

Philosophical problems connected with democracy relate both to its nature and its value. It might seem obvious that democracy has value because it promotes liberty and equality. As compared with, for example, dictatorship, everyone has equal political power and is free from control by a special individual or group. However, at least on the voting conception of democracy, it is the majority who have the control. This means that the minority may not be thought to be treated equally; and they lack liberty in the sense that they are controlled by the majority.

Another objection to democracy is that, by counting everyone's opinions as of equal value, it considers the ignorant as being as important as the knowledgeable, and so does not result in properly informed decisions. However, voting may in certain circumstances be the right way of achieving knowledge. Pooling opinions may lead to better group judgement.

These difficulties with democracy are alleviated by the model which concentrates on mutual discussion rather than people just feeding opinions into a voting mechanism. Opinions should in such circumstances be better formed; and individuals are more obviously equally respected. However, this depends upon them starting from positions of equal power and liberty; rather than being consequences of a democratic procedure, it would seem that equality and liberty are instead prerequisites which are needed in order for it to work properly.

1 What democracy is

Democracy means rule by the people. It is a form of decision making or government whose meaning can be made more precise by contrast with rival forms, such as dictatorship, oligarchy or monarchy. In these rival forms a single person or a select group rules. With democracy this is not so. The people themselves rule and they rule themselves. The same body is both ruler and ruled.

Philosophical accounts of democracy analyse its nature and discuss its value. The two cannot be completely separated. Any account which explains the value of democracy has to provide or presuppose an account of what it is holding to be of value. Conversely, supposedly neutral analyses of the nature of democracy are influenced by values. For example, someone who thinks that democracy is a good thing is liable to analyse it in terms of other features also thought to be good.

The concept of democracy therefore may naturally be thought of as what W.B. Gallie called an essentially contested concept. Such concepts are concepts whose analysis is unresolvable because different analysts read into it their favoured values. For example, before the reunification of Germany, both East and West Germany called themselves democracies. Yet each had very different political systems, one being a Marxist single-party state, the other having economic and political competition with several parties and contested elections. A dispute about which one was really a democracy would be irresolvable.

This account and this example presuppose that democracy is desirable, so there is a competition to lay claim to the honorific title. However, for most of the time since the invention of the concept of democracy it has not been taken to be a term of honour. A kind of democracy did exist in ancient Athens. But, this was a form of government criticized by the leading Greek thinkers of the time, Plato and Aristotle. For most of the time since this early democracy ended, democracy has neither existed nor been thought to be desirable.

Much later, with the creation of the USA, we reach a system which most people today would take to be a paradigmatic example of democracy. Unsurprisingly it was defended by its founding fathers. However, what might surprise us more today is that in one of the most famous of these defences, James Madison was careful not to use

‘democracy’ as the name for the system he supported. He identifies things called democracies and does not support them; the description he uses instead for the fledgling USA is ‘republic’ (Madison, Hamilton and Jay 1787-8).

What Madison means by a republic is ‘a government in which the scheme of representation takes place’, and by a democracy ‘a society consisting of a small number of citizens, who assemble and administer the government in person’ (Madison, Hamilton and Jay [1787-8] 1987: 126). It might be thought that the central question here is one of size. Commentators writing just before Madison, such as Rousseau (§3) (1762) and Montesquieu (§3) (1748), held that democracy was only possible in small states; and Madison can be taken to be marking the transition to the modern world, with large states rather than small ones; and a corresponding move from direct democracy to representative government. What is today standardly called democracy is very different from what was standardly so called in the ancient world.

However, size is not the only important distinction here. Individuals in very large modern political units can now be so linked together by modern technology that they can relate to each other much as if they all met together. On the other hand, political decisions by and for small groups are still made in the modern world. It can still be asked of these whether they should be made democratically; and, if so, which sort of democracy is appropriate. So, whatever the size of political unit, questions can arise about the importance of participation or discussion before decisions are made. It can be asked whether democracy should be seen primarily as a mechanism in which people vote for policies or representatives without assembly, participation or discussion. At one extreme (as with Joseph Schumpeter (§3) (1943)), we could analyse democracy as a competition for votes between professional politicians. At the other extreme, we could analyse democracy as a system in which unanimous decisions are reached after a prolonged discussion which respects the equal autonomy and participation of everyone involved. The former seems more practical, but may not uphold any (other) ideals; the latter seems impressively ideal, but may be ineffective in practice.

It has just been said that voting and representation is at least practical. However, this ignores one prominent problem. This is that the collective view which results from voting may not be related in the way we would wish to the individual views expressed in the votes. In particular this applies if there are three or more options to be arranged in order of preference and there are three or more such individual orderings (see Social choice). These problems will not be discussed further here; although it should be recognized that many people (such as W.H. Riker (1982)) think that they are an insuperable objection to democratic decision making.

2 The value of democracy

Once we have an idea of democracy, the next question is why, or whether, it is of value. The Greek historians identified the original introduction of democracy with the advance of liberty and equality. Since both liberty and equality are usually thought to be of value, this would seem to be a natural answer to the question. Democracy is of value because it produces liberty and equality. With dictatorship or other forms of special leadership, a particular person or group has more power than others. By contrast, in democracy everyone is equal. Everyone has the same (political) power. So democracy is egalitarian as compared with other forms of government or decision making.

Similarly for liberty. A democracy introduced by the overthrow of a dictator increases political liberty. People have been freed from the control of the dictator. Hence democracy promotes liberty. There are several connected terms here: liberty, freedom, autonomy. However, whichever term is used, this argument seems to work. Consider autonomy. It means, literally, self-rule. Yet this is exactly what happens in democracy, as opposed to other forms of government. The people rule themselves.

However, as always, further inspection makes matters less obvious. Suppose decisions are made by majority vote and someone is in the minority. This person is outvoted and so their wants will not be put into effect. Therefore we can question whether in this (democratic) situation, this person is really autonomous. They are being made to do something which they do not want to do. Hence they are not really autonomous. Similarly for equality. Not everyone is treated equally when majority decisions are adopted, because only the views of some people (the majority) are put into effect. The minority’s views are disregarded. Hence they are not treated equally. The winner takes all, and hence winners and losers are not equally treated.

If a community is divided into two parts living in mutual antipathy, this becomes even more obvious. The majority

community could, by democratic vote, bear heavily down on the minority community, restricting or removing things it holds to be of fundamental value. In such circumstances the members of the minority community could hardly be said to be at liberty; nor could it be said that they were being equally treated. Hence the phrase, used by Tocqueville (1835) and taken over by J.S. Mill (1859), 'the tyranny of the majority'. The initial contrast between democracy and dictatorship has now been left behind. If democracy is really the dictatorship of the majority, then it is not so obvious that democracy promotes freedom and equality.

3 The paradox of democracy

If we examine democracy from the standpoint of the minority, as in the last section, this helps to focus the problem of its value. A democrat thinks that the majority view ought to be enacted. But in voting they also declare their own view about what ought to be enacted. When they are in a minority these diverge and they seem to be caught in a contradiction. There are two incompatible policies, A and not-A. Yet the minority democrat seems to think both that A ought to be enacted (because that is what the majority want) and also that not-A ought to be enacted (because that is their own view). Richard Wollheim (1962) called this the 'paradox of democracy'.

However, if democracy can be given a value, the paradox is resolvable. For what we then have is a simple (and familiar) conflict of values. The democrat's direct view of the matter indicates the value of the course of action for which they voted. But once it is defeated by the majority this rival course of action also possesses value. For it inherits the value of democracy. If, for example, democracy is taken to be an egalitarian procedure then adopting this rival course of action has egalitarian value.

An example: four of us in a car have to decide to go either to the beach or to the town. There is only one car and we can go to only one place. We agree to decide democratically, by vote. The vote is taken. I vote for the beach, and am outvoted by three to one. The beach is of value to me. This is shown by my vote. However, I am also a democrat. After the vote, the town also has value to me. With the town three people's views are respected; with the beach only one. If I hold that people are of equal value, then I have a reason for the car to go to the town.

4 Democracy and knowledge

In his *Republic* Plato says that 'it is not in the natural course of things for the pilot to beg the crew to take his orders' (*c*.380-367 bc: 489b). The implication is that if we want as a group to go to the right place, it is not sensible to assume that everyone has an equally valid opinion. Instead we should follow the lead of those who know. Hence democracy, which treats everybody's opinions equally, is inefficient as a means of determining the right thing to do (see Plato §14).

This argument makes several presuppositions and can be resisted by contesting them. Some people can only know more than others about something if there is indeed something to be known. That is, if there is a truth about the matter independent of people's opinions. But this is precisely what might be contested when the question is what the state should do. This being a matter of value, it might be held that no independent truth, and hence no knowledge, is available. More precisely, it might be thought that a line can be drawn between areas in which knowledge is available, and which, for example, might be handled by a professional, trained civil service; and areas for which no knowledge is obtainable, and which should be left to democratic, untrained, amateur decision. Benjamin Barber (1984), for example, takes the area of politics to be one of action, not truth; and for him democracy takes over in the areas where metaphysics fails, creating its own epistemology.

It should be noticed, however, that an argument for the goodness of democratic decision making cannot simply be made on the basis of a complete scepticism about values. For if no truths about values are available, then no truths about the value of democracy are available either. Hence a valid argument cannot be made from this premise to a conclusion that it is true that democracy is of value.

Conversely, even if it is allowed that there are independent truths about value, it does not directly follow that democracy is an inappropriate way of discovering these truths. For it is quite possible that the truth about what in general the state should do is the kind of truth about which people have a roughly equal capacity. Furthermore, even if people do not have equal capacity, as long as it cannot be told which ones are superior, democracy may still be the appropriate method to use. The Platonic argument assumes that there is a truth about what should happen; that this truth is better known to some people than others; and that it is possible to tell independently of

their views which these people are. All these assumptions could be resisted.

If people are of roughly equal capacity (or it cannot be told who is superior) then, as long as everyone is more likely than not to be right, voting and adopting the majority view is an efficient method to use. For the majority decision has a higher probability of being right than any individual decision, as Condorcet (§2) was the first to show. In other words, if I have to make a sequence of decisions about the truth of something and I am in a group each of whose members gets the answer right more often than not, then I do much better systematically following the majority view of the group than my own initial views.

Even if some people clearly have better informed views than others, it still does not follow that democratic decision making is inefficient. For if it is obvious who the experts are, then people with an interest in discovering what is right will generally follow their views. In other words, the same answers will be arrived at as would happen if, as in Plato, the better informed were made dictators. Democracy will not be inferior in discovering of the truth, and will have other advantages.

On the other hand, if it is not obvious who the experts are, then it is indeed the case that the majority view may not follow expert opinion. But, if some people are dictators, it may also be the case that the people who are made dictators are not the ones who are better informed. The dangers of mistake in following majority opinion are matched by the dangers of mistake in making the wrong people the dictators. The Platonic argument only works if the experts can be recignized in advance, for example (as in Plato’s *Republic*), because they are educated in a way which ensures that they will have expertise.

5 The use of democracy

Other justifications for democracy are possible. One standard device for justification, for many areas, is utilitarianism (see Utilitarianism). Something is justified if it promotes general happiness or utility. It can be asked of a form of government, just like anything else, whether it does tend to promote this. The answer, at least of the classical utilitarians such as Bentham (§6) and James Mill, is that democracy does.

This argument is expressed most simply in James Mill’s short essay, *Government* (1820). He starts with an evaluative and a factual premise. The evaluative premise is utilitarianism. Actions are right in so far as they promote the general happiness. The factual premise is universal self-interest. People seek those things which promote their own interests. The problem is to find the form of government in which both of these premises can be true together, to find the form in which people seeking their own interest will nevertheless do those things which promote the general happiness. It is not difficult to show that representative democracy is the answer. Kings will promote the interest of kings, dictators of dictators, oligarchies of oligarchies. In all cases the interest promoted is that of the ruling group, not that of the people as a whole. However, if the people as a whole are put in charge, they will promote the interests of the people as a whole. Seeking their own interests, they will produce general happiness. Hence both premises are satisfied simultaneously.

It is perhaps unsurprising that the greatest happiness of the greatest number results if the majority (the greatest number) are put in charge. However the answer does depend upon certain presuppositions. It assumes that people act in their own interests. Even if this is what they generally intend to do (which might be disputed), it does not follow from this that they are successful. For they may not know their interests. For example, it is often held that people discount the future too severely, so they prefer less important immediate interests to more important long-term interests. If this is so, then democratic decisions will lead to too short-term results, which are not even in the interests of those voting (see Needs and interests; Public interest).

A related point is that this model takes preferences as they are, without allowing them to be changed by the democratic process. Yet if people are ill informed about what is good for them, it would be better to operate on the preferences before counting votes. Democracy treats all votes equally. But people may not be equally informed about their interests. So the result may be that some interests are catered for better than others. Analogously, treating votes equally means that strongly held and weakly held views are considered of equal importance. Yet if the goal is to maximize utility, it may be wrong to follow the weakly held view of a majority rather than the strongly held view of a significant minority. If the utility of getting something is supposed to be roughly proportional to the strength of the desire for it, then it could be that the total of less people multiplied by a greater utility per person is higher than the total of more people multiplied by less utility per person.

6 Other consequences

The idea of utilitarianism as a mere preference-satisfying machine, in which antecedently given preferences are satisfied, has often been criticized. One alternative is to treat the values more objectively. Democracy can then be shown to be good in terms of these independently specified consequences. Such was the approach of J.S. Mill (1861) and more recent defenders such as William Nelson (1980). Democracy is justified as a form of education or development; it is taken to be a political system in which individuals are made to think for themselves and are therefore improved. Even if the decisions they make are not the best decisions, it is better for individuals if they try and take part in such decisions.

Another consequence which might justify democracy is the supposed promotion of dynamic economic activity; as opposed to the sluggish effects supposedly emanating from more centralized planning and control. Yet even if democracy does correlate with such beneficial economic circumstances, it is not clear that this by itself can be used as an argument for promoting democracy. Jon Elster (1986) identifies the questionable role of such arguments based on indirect effects. For it may be that these other effects only happen if people are attached to democracy for more direct reasons (such as thinking that it is a just form of government). If people were only to support democracy because they thought that it encouraged economic dynamism, then the democracy would not work, and so the economic dynamism would not follow either.

This relates to another familiar problem with starting from antecedent preferences and then taking democracy to be a sort of market mechanism in which these preferences are traded. If people only act through self-interest, trying to get their antecedently given preferences fulfilled, then it is not obvious why they should vote at all. For the advantages of voting will come to them if the others vote and their own vote seems to be merely a cost. At the national government level, it is exceedingly unlikely that any one vote will be decisive. So they would be better off not voting at all.

One answer to all of these problems is to dispense with the idea of democracy being a mechanism for satisfying antecedently given preferences. Instead of taking these as given, democracy should be held as a device in which people develop and discover their views about what is right. And, in thinking about what is right, they should think about what is right for the group as a whole, and not just themselves. People should therefore participate in a form of decision making in which they share their ideas, discuss together and, with luck, eventually reach general agreement.

The form of arguments people can use in such discussions is naturally constrained, as people seeking agreement should look as if they are appealing to general principles rather than merely appealing to self-interest. The condition of publicity (that is, of what can be said publicly) imposes constraints. If people think from the general point of view rather than in terms of their own individual interest, the forms of reasoning and the antecedent judgements will be different. Discussion rather than voting becomes the central feature of democracy, and it is important that people can meet and talk together before decisions are made.

These ideas promoting discussion and participation have several presuppositions. They presuppose that people will be better able to work out the truth (about what is good for the group) by working in groups rather than individually. This may be the case if they are all independently motivated by the same desire to discover the truth. It is less obviously the case if there are deep conflicts of interest (such as capital against labour; or country against town; or this world against the next). The supposition is that group discussion leads to more rationality; but in some circumstances group dynamics merely increase and inflame passion, so that people behave badly together in a way that they never would separately.

The values considered at the start of this discussion - namely, liberty and equality - now reappear; only now not as the consequence of democratic activity but as its prerequisites. For if discussion is to reach the right answer, it needs to start with roughly equal power between the discussants. Otherwise discussion will be forced in the interests of the stronger. Hence the idea that democracy needs circumstances of roughly equal wealth (held by Montesquieu and Rousseau). Hence the Marxist criticism that Western liberal democracy works on the fiction of an idealized equality when the real situation is one of greatly unequal economic power. Hence John Rawls' argument (1971) that political parties should be paid for by the state to avoid the economically powerful buying votes (see Rawls, J. §2). Hence also the objections of feminist theorists. If men and women are antecedently in a

situation of different power, then the supposed equality of democracy will only result in most of the power remaining with the men. Discussion, yes: but only if the forum is subject to powerful antecedent control and regulation. Otherwise we return to the bad old world of bargaining between antecedently given preferences from which this optimistic espousal of discussion was meant to save us.

See also: Constitutionalism; General will; Representation, political; Social democracy

ROSS HARRISON

References and further reading

Barber, B. (1984) *Strong Democracy*, Berkeley, CA: University of California Press. (Referred to in §4.)

Copp, D., Hampton, J. and Roemer, J.E. (eds) (1993) *The Idea of Democracy*, Cambridge: Cambridge University Press.(Full collection of studies on truth, antecedent preferences and publicity.)

Dahl, R.A. (1989) *Democracy and its Critics*, New Haven, CT, and London: Yale University Press.(Presentation of criticisms and reply; useful on equality.)

Elster, J. (1986) ‘The Market and the Forum’, in J. Elster and A. Hylland (eds) *Foundations of Social Choice Theory*, Cambridge: Cambridge University Press. (Mentioned in §6.)

Estlund, D. (1989) ‘The Puzzle of the Minority Democrat’, *American Philosophical Quarterly* 26: 143-51. (Contains an account and full bibliography of the paradox of democracy.)

Gallie, W.B. (1956) ‘Essentially Contested Concepts’, *Proceedings of the Aristotelian Society* 56: 167-98. (Discussed in §1.)

Graham, K. (1986) *The Battle of Democracy*, Brighton: Wheatsheaf.(Analyses democracy in terms of various normative grounds; good defence in terms of autonomy and also useful on Marxist theories.)

Harrison, R. (1993) *Democracy*, London: Routledge.(Historical account and analysis of value in terms of such values as equality, knowledge and autonomy.)

Madison, J., Hamilton, A. and Jay, J. (1787-8) *The Federalist Papers*, ed. I. Kramnick, Harmondsworth: Penguin, 1987.(Paper number ten, written by Madison, is quoted in §1.)

Mill, J. (1820) *Government*, reprinted in J. Lively and J. Rees (eds) *Utilitarian Logic and Politics*, Oxford: Oxford University Press, 1978.(Referred to in §5.)

Mill, J.S. (1859) *On Liberty*, in *Utilitarianism/On Liberty/Considerations on Representative Government*, London: Dent, 1910.(Chapter 1 refers to the ‘tyranny of the majority’, mentioned in §2.)

Mill, J.S. (1861) *Considerations on Representative Government*, in *Utilitarianism/On Liberty/Considerations on Representative Government*, London: Dent, 1910.(Discussed in §6.)

Montesquieu, C.L. de S. (1748) *De l’esprit des lois (The Spirit of the Laws)*, trans. A. Cohler *et al.*, Cambridge: Cambridge University Press, 1989.(The statement about size, referred to in §1, is in Book 8, chapter 16.)

Nelson, W.N. (1980) *On Justifying Democracy*, London: Routledge & Kegan Paul.(Mentioned in §6; criticizes various supposed bases.)

Phillips, A. (1991) *Engendering Democracy*, Cambridge: Polity Press.(Feminist analysis and critique of democracy.)

Plato (*c*.380-367 BC) *The Republic*, trans. with introduction and notes by F.M. Cornford, Oxford: Oxford University Press, 1941.(Referred to in §4.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Section 23 is on publicity; sections 53 and 54 are on majority rule; the proposal to pay political parties, mentioned in §6, occurs on page 226.)

Rousseau, J.-J. (1762) *Du contrat social*, ed. and trans. G.D.H. Cole in *The Social Contract and Discourses*, London: Dent, 1973.(That states should be small and fortunes equal as preconditions for democracy is claimed in Book 3, chapter 4; however, it is the first two books which have been most influential, particularly about the general will.)

Riker, W.H. (1982) *Liberalism against Populism*, San Francisco, CA: W.H. Freeman. (Mentioned in §1.)

Schumpeter, J.A. (1943) *Capitalism, Socialism, and Democracy*, London: Allen & Unwin.(Part IV criticizes the classical account of democracy and proposes instead that democracy should be seen as the competition between various elites bidding for votes.)

Singer, P. (1973) *Democracy and Disobedience*, Oxford: Oxford University Press. (Argues that participation in a democratic procedure places the participants under an obligation to accept its results.)

Tocqueville, A. de (1835) *De la democratie en Amerique (Democracy in America)*, ed. J.P. Mayer and M. Lerner, London: Fontana, 1966.(Part I, chapter 15, is the section on the ‘tyranny of the majority’, referred to in §2.)

Wollheim, R. (1962) ‘A Paradox in the Theory of Democracy’, in P. Laslett and W.G. Runciman (eds)

Philosophy, Politics and Society, 2nd series, Oxford: Blackwell.(The article referred to in §3.)

Democritus (mid 5th-4th century BC)

A co-founder with Leucippus of the theory of atomism, The Greek Philosopher Democritus developed it into a universal system, embracing physics, cosmology, epistemology, psychology and theology. He is also reported to have written on a wide range of topics, including mathematics, ethics, literary criticism and theory of language. His works are lost, except for a substantial number of quotations, mostly on ethics, whose authenticity is disputed. Our knowledge of his principal doctrines depends primarily on Aristotle's critical discussions, and secondarily on reports by historians of philosophy whose work derives from that of Aristotle and his school.

The atomists attempted to reconcile the observable data of plurality, motion and change with Parmenides' denial of the possibility of coming to be or ceasing to be. They postulated an infinite number of unchangeable primary substances, characterized by a minimum range of explanatory properties (shape, size, spatial ordering and orientation within a given arrangement). All observable bodies are aggregates of these basic substances, and what appears as generation and corruption is in fact the formation and dissolution of these aggregates. The basic substances are physically indivisible (whence the term atomon*, literally 'uncuttable') not merely in fact but in principle; (1) because (as Democritus argued) if it were theoretically possible to divide a material thing* ad infinitum*, the division would reduce the thing to nothing; and (2) because physical division presupposes that the thing divided contains gaps. Atoms are in eternal motion in empty space, the motion caused by an infinite series of prior atomic 'collisions'. (There is reason to believe, however, although the point is disputed, that atoms cannot collide, since they must always be separated by void, however small; hence impact is only apparent, and all action is at a distance.) The void is necessary for motion, but is characterized as 'what*-is-*not', thus violating the Eleatic principle that what-is-not cannot be.*

Democritus seems to have been the first thinker to recognize the observer-dependence of the secondary qualities. He argued from the distinction between appearance and reality to the unreliability of the senses, but it is disputed whether he embraced scepticism, or maintained that theory could make good the deficiency of the senses. He maintained a materialistic account of the mind, explaining thought and perception by the physical impact of images emitted by external objects. This theory gave rise to a naturalistic theology; he held that the gods are a special kind of images, endowed with life and intelligence, intervening in human affairs. The ethical fragments (if genuine) show that he maintained a conservative social philosophy on the basis of a form of enlightened hedonism.

1 Life and works

Democritus lived from approximately the middle of the fifth century BC to well into the following century. Though he was a co-founder of atomism with Leucippus, the exact relation between the two is obscure; Aristotle and his school agree in treating Leucippus as the originator of the theory, but also in assigning its basic principles to both, while later sources treat the theory as the work of Democritus alone. He may have collaborated with Leucippus, and almost certainly developed the latter's theory into a universal system.

Very little is known about Democritus' life. He came from Abdera, on the north coast of the Aegean (also the native city of Protagoras). He describes himself as being young in the old age of Anaxagoras, that is, probably in the 430s BC, and was traditionally reported to have lived to a very great age. He is said to have visited Athens and to have had some slight acquaintance with Socrates. The list of his works preserved by the third-century AD biographer Diogenes Laertius is long and encyclopedic in scope, including a complete account of the physical universe, the *Lesser World-System* (so called to distinguish it from his predecessor Leucippus' *Great World-System*), and works on astronomy and other natural sciences, epistemology, ethics, mathematics and literature. None survives. Ancient sources preserve almost 300 purported quotations, the great majority on ethics, from two different, although overlapping, collections of ethical maxims, one by the fifth-century anthologist Stobaeus, the other from a collection, probably earlier than that of Stobaeus, traditionally known as 'The Sayings of Democrates'. The authenticity of the ethical fragments is disputed. Sextus Empiricus preserves some important quotations on epistemology, and Plutarch, Galen and others preserve a few other miscellaneous quotations. For our knowledge of the physical doctrines we are reliant on the doxographical tradition stemming ultimately from Aristotle, who discusses atomism extensively (see doxography).

2 Physical doctrines

According to Aristotle, the atomists attempted to reconcile the observable data of plurality, motion and change with the Eleatic denial of the possibility of coming to be or ceasing to be. Accordingly they postulated unchangeable primary substances, and explained apparent generation and corruption as the formation and dissolution of aggregates of those substances. Since the theory had to account for an assumed infinity of phenomena, it assumed an infinite number of primary substances, while postulating the minimum range of explanatory properties, specifically shape, size, spatial ordering and orientation within a given arrangement. All observable bodies are aggregates of basic substances, which must therefore be too small to be perceived. These corpuscles are physically indivisible (*atomon*, literally 'uncuttable'), not merely in fact but in principle; Aristotle reports an (unsound) atomistic argument, which has some affinities with one of Zeno of Elea's arguments against plurality, that if (as Anaxagoras maintained) it were theoretically possible to divide a material thing *ad infinitum*, the division must reduce the thing to nothing (see Zeno of Elea §4; Anaxagoras §2). This Zenonian argument was supported by another for the same conclusion; atoms are theoretically indivisible because they contain no void. On this view, bodies split along their interstices; hence, where there are no interstices, as in an atom, no splitting is possible. (The same principle accounts for the immunity of the atoms to other kinds of change, such as reshaping, compression and expansion; all require displacement of matter within an atom, which is impossible without any gaps to receive the displaced matter.) It is tempting to connect the assumption that bodies split only along their interstices with the Principle of Sufficient Reason, which the atomists appealed to as a fundamental principle of explanation (arguing, for example, that the number of atomic shapes must be infinite, because there is no more reason for an atom to have one shape than another (Simplicius, *On Aristotle's Physics* 28.9-10), and that there must be infinitely many worlds because there is no more reason for one to have been formed here than elsewhere. Given the total uniformity of an atom, they may have thought that there could be no reason why it should split at any point, or in any direction, rather than any other. Hence, by the Principle of Sufficient Reason, it could not split at all.

Atoms are in a state of eternal motion in empty space; the motion is not the product of design, but is determined by an infinite series of prior atomic collisions (whence two of Aristotle's principle criticisms of Democritus: that he eliminated final causation and that he made all atomic motion 'unnatural', that is to say, imposed by external force). Empty space was postulated as required for motion, but was characterized as 'what-is-not', thus violating the Eleatic principle that what-is-not cannot be. We have no evidence of how the atomists met the accusation of outright self-contradiction. As well as explaining the possibility of motion, the void was postulated to account for the observed plurality of things, since the atomists followed Parmenides (fr. 8.22-5) in maintaining that there could not be many things if there were no void to separate them (see Parmenides §6). The theoretical role of the void in accounting for the separation of atoms from one another has an interesting implication, recorded by Philoponus (*On Aristotle's On Physics* 494.19-25, *On Aristotle's On Generation and Corruption* 158.26-159.7). Since atoms are separated from one another by the void, they can never strictly speaking come into contact with one another. For if they did, even momentarily, there would be nothing separating them from one another. But then they would be as inseparable from one another as the inseparable parts of a single atom, whose indivisibility is attributed to the lack of void within it (see above); indeed, the two former atoms would now be parts of a single larger atom. But, the atomists held, it is impossible that two things should become one. Holding atomic fusion to be theoretically impossible, and taking it that any case of contact between atoms would be a case of fusion (since only the intervening void prevents fusion), they perhaps drew the conclusion that contact itself is theoretically impossible. Hence what appears to be impact is in fact action at an extremely short distance; rather than actually banging into one another, atoms have to be conceived as repelling one another by some sort of force transmitted through the void. Again, although no source directly attests this, the interlocking of atoms which is the fundamental principle of the formation of aggregates is not strictly speaking interlocking, since the principle of no contact between atoms forbids interlocking as much as impact. Just as impact has to be reconstrued as something like magnetic repulsion, so interlocking has to be reconstrued as quasi-magnetic attraction. If this suggestion is correct (and it is fair to point out that no ancient source other than Philoponus supports it), it is a striking fact that, whereas the post-Renaissance corpuscular philosophy which developed from Greek atomism tended to take the impossibility of action at a distance as an axiom, the original form of the theory contained the a priori thesis that all action is action at a distance, and consequently that impact, as far from giving us our most fundamental conception of physical interaction, is itself a mere appearance which disappears from the world when the description of reality is pursued with full rigour (see Blasius of Parma §4; Bruno, G. §5).

3 Epistemology and psychology

Democritus seems to have been the first philosopher to recognize the observer-dependence of the secondary qualities. Perception of the secondary qualities reveals merely how things seem to us (and these qualities are said to be merely *nomōi*, or 'convention'), as opposed to how they really (*eteēi*) are - aggregates of unobservable corpuscles. The perception of any secondary quality is a response, not to the properties of an individual atom, but to a physical stimulus constituted by a continuous bombardment of a series of arrays of atoms. Thus visual perception, including the perception of colour, is a response to the impact on the visual organs of a continuous succession of films of atoms (*eidōla*) flowing from the surface of the perceived object. For an object to be red, for example, is thus for it to emit films of atoms of such a nature that, when those films collide with an appropriately situated perceiver, the object will look red to that perceiver.

It is therefore a necessary consequence of the atomists' account of the secondary qualities not merely that secondary qualities are among the causally significant properties of the atoms, but, more strongly, that atoms have no secondary qualities. Colour, sound, taste and smell are qualities of aggregates, not of single atoms. To ascribe one of these qualities to a subject is to describe that subject as so structured as to prompt a certain sensory response in a certain kind of perceiver; and the structure in question as described in terms of the intrinsic properties (shape and size) and relational properties (arrangement and orientation) of its constituent atoms. The persuasiveness or otherwise of the account of the secondary qualities cannot be separated from that of the whole theory of perception of which it is part, and that in turn cannot be separated from the theory of human nature, and ultimately of the natural world as a whole. As presented by the atomists, the theory is entirely speculative, since it posits as explanatory entities microscopic structures of whose existence and nature there could be no experimental confirmation. Modern developments in sciences such as neurophysiology have revised our conceptions of the structures underlying perceptual phenomena to such an extent that modern accounts would have been unrecognizable to Leucippus or Democritus; but the basic intuitions of ancient atomism - that appearances are to be explained at the level of the internal structure of the perceiver and of the perceived object, and that the ideal of science is to incorporate the description of those structures within the scope of a unified and quantitatively precise theory of the nature of matter in general - have stood the test of time.

According to some sources, Democritus used the contrast between unobservable atomic structures and the appearances which they present to us to demonstrate the unreliability of the senses. He then faced the problem of the justification of his theory, which was founded on data provided by the senses; in a famous fragment quoted by Galen the senses say to the mind, 'Wretched mind, you get your evidence from us, and yet do you overthrow us? The overthrow is a fall for you' (fr. 125). It is disputed whether Democritus drew a sceptical conclusion, or whether his position was rather that, since there can be no ground for preferring one piece of sensory information to another, the task of theory is to provide an account of how all appearances are appropriate representations of an underlying unobservable reality (that is, all appearances are true, a thesis which Aristotle attributes to him, and which was later explicitly maintained by the Epicureans; see Epicureanism (§6)).

Democritus' uncompromising materialism extended to his psychology. Although there is some conflict in the sources, the best evidence is that he drew no distinction between the rational soul or mind and the non-rational soul or life principle, giving a single account of both as a physical structure of spherical atoms permeating the entire body. This theory of the identity of soul and mind extended beyond identity of physical structure to identity of function; Democritus explained thought, the activity of the rational soul, by the same process as that by which he explained perception, one of the activities of the sensitive or non-rational soul. Both are produced by the impact on the soul of extremely fine, fast-moving films of atoms (*eidōla*), constantly emitted in continuous streams by the surfaces of everything around us. This theory combines a causal account of both perception and thought with a crude pictorial view of thought. The paradigm case of perception is vision; seeing something and thinking of something alike consist in picturing the thing seen or thought of, and picturing consists in having a series of actual physical pictures of the thing impinge on one's soul. While this assimilation of thought and experience has some affinities with classical Empiricism, it differs in one crucial respect; whereas the basic doctrine of empiricism is that thought derives from experience, for Democritus thought is a form of experience, or, more precisely, the categories of thought and experience are insufficiently differentiated to allow one to be characterized as more fundamental than the other. Among other difficulties, this theory faces the problem of accounting for the distinction, central to Democritus' epistemology, between perception of the observable properties of atomic

aggregates and thought of the unobservable structure of those aggregates. We have no knowledge of how, if at all, Democritus attempted to deal with this problem.

4 Theology

Another disputed question is whether Democritus' materialistic account of the universe left any room for the divine. According to most of the ancient sources he believed that there are gods, which are living, intelligent, material beings (of a peculiar sort), playing a significant role in human affairs. They are atomic compounds, and like all such compounds they come to be and perish. They did not create the physical world (of which they are part), nor, although they are intelligent, do they organize or control it. They are as firmly part of the natural order as any other living beings. Specifically, Democritus believed the gods to be living *eidōla*, probably of gigantic size, possessing intelligence, moral character and interest in human affairs. While some sources suggest that these *eidōla* emanate from actual divine beings, the majority of sources agree that they are themselves the only divine beings which Democritus recognized. Some modern scholars (for example, Barnes 1982) interpret this as amounting to atheism, taking Democritus to have held that the gods are nothing more than the contents of human fantasy. But for Democritus *eidōla* are not intrinsically psychological; they are not contents of subjective states, but part of the objective world, causing psychological states through their impact on physical minds. In this case his theory must explain their source and their properties, notably their being alive. Since the gods were of human form, it is plausible to suggest that their source is actual human beings, possibly giants living in the remote past. They are themselves alive in that, flowing from beings permeated with soul-atoms, they contain soul-atoms themselves. Consistently with this naturalistic theology Democritus gave a naturalistic account of the origin of religion, identifying two types of phenomena as having given rise to religious belief: first, the occurrence of *eidōla* themselves, presumably in dreams and ecstatic states; and second, celestial phenomena such as thunder, lightning and eclipses.

Democritus' theology thus contrives to incorporate some of the most characteristic features of the gods of traditional belief, notably their anthropomorphism, power, longevity (although not, crucially, immortality), personal interaction with humans and interest (for good or ill) in human affairs, within the framework of a naturalistic and materialistic theory. It is thus, despite the bold originality of its account of the divine nature, notably more conservative than some of its predecessors (especially the non-anthropomorphic monotheism normally attributed to Xenophanes) and than that of its Epicurean successor, whose main concern is to exclude the gods from all concern with human affairs (see Xenophanes §3; Epicureanism §§8-9).

5 Ethics and politics

The evidence for Democritus' ethical views differs radically from that for the areas discussed above, since while the ethical doxography is meagre, our sources (see §1) preserve a large body of purported quotations on ethical topics. While the bulk of this material is probably Democritean in origin, the existing quotations represent a long process of excerpting and paraphrase, making it difficult to determine how close any particular saying is to Democritus' own words. Various features of style and content suggest that Stobaeus' collection of maxims contains a greater proportion of authentically Democritean material than does the collection which passes under the name of 'Democrates'.

Subject to the limitations imposed by the nature of this material, we can draw some tentative conclusions about Democritus' ethical views. He was engaged with the wide-ranging contemporary debates on individual and social ethics of which we have evidence from Plato and other sources. On what Socrates presents as the fundamental question in ethics, 'How should one live?' (Plato, *Gorgias* 500c, *Republic* 352d), Democritus is the earliest thinker reported as having explicitly posited a supreme good or goal, which he called 'cheerfulness' or 'wellbeing', and which he appears to have identified with the untroubled enjoyment of life. It is reasonable to suppose that he shared the presumption of the primacy of self-interest which is common both to the Platonic Socrates (§6) and to his immoralist opponents, Callicles and Thrasymachus. Having identified the ultimate human interest with 'cheerfulness', the evidence of the testimonia and the fragments is that he thought that it was to be achieved by moderation, including moderation in the pursuit of pleasures, by discrimination of useful from harmful pleasures and by conformity to conventional morality. The upshot is a recommendation to a life of moderate, enlightened hedonism, which has some affinities with the life recommended by Socrates (whether in his own person or as representing ordinary enlightened views is disputed) in Plato's *Protagoras*), and, more obviously, with the

Epicurean ideal of which it was the forerunner.

An interesting feature of the fragments is the frequent stress on individual conscience. Some fragments stress the pleasures of a good conscience and the torments of a bad one (frs 174, 215) while others recommend that one should be motivated by one's internal sense of shame rather than by concern for the opinion of others (frs 244, 264, 'Sayings of Democrates' 84). This theme may well reflect the interest, discernible in contemporary debates, in what later came to be known as the question of the sanctions of morality. A recurrent theme in criticisms of conventional morality was that, since the enforcement of morality rests on conventions, someone who can escape conventional sanctions, for example by doing wrong in secret, has no reason to comply with moral demands. Proponents of such a belief include Antiphon (fr. 44), Critias (fr. 25) (see Physis and nomos) and Plato, in Glaucon's tale of Gyges' ring in the *Republic* (359b-360d). A defender of conventional morality who, like Democritus and Plato, accepts the primacy of self-interest therefore faces the challenge of showing, in one way or another, that self-interest is best promoted by the observance of conventional moral precepts.

The appeal to divine sanctions, cynically described in Critias' fragment 25, represents one way of doing this, and there are some traces of the same response in Democritus. While his theory of the atomic, and hence mortal, nature of the soul admits no possibility of postmortem rewards and punishments, the theory allows for divine rewards and punishments in this life. Fragment 175 suggests a complication: the gods bestow benefits on humans, but humans bring harm on themselves through their own folly. Is the thought that the gods do not inflict punishment arbitrarily, but that humans bring it on themselves? Or is it rather that the form which divine punishments take is that of natural calamities, which humans fail to avoid through their own folly? The latter alternative would make the pangs of conscience one of the forms of divine punishment, while the former would see it as a further sanction. Either way (and the question is surely unanswerable), we have some evidence that Democritus was the earliest thinker to make the appeal to 'internal sanctions' central to his attempt to derive morality from self-interest, thus opening up a path followed by others including Bishop Butler (§§2-4) and John Stuart Mill (§11).

The attempt, however pursued, to ground morality in self-interest involves the rejection of the antithesis between law or convention (*nomos*) and nature (*physis*) which underlies much criticism of morality in the fifth and fourth centuries BC (see Physis and nomos). For Antiphon, Callicles, Thrasymachus and Glaucon, nature prompts one to seek one's own interest while law and convention seek, more or less successfully, to inhibit one from doing so. But if one's long-term interest is the attainment of a pleasant life, and if the natural consequences of wrong-doing, including ill health, insecurity and the pangs of conscience, give one an unpleasant life, while the natural consequences of right-doing give one a contrastingly pleasant life, then nature and convention point in the same direction, not in opposite directions as the critics of morality allege. (We have no evidence whether Democritus had considered the objections that conscience is a product of convention, and that exhorting people to develop their conscience assumes that it must be.) Although the texts contain no express mention of the *nomos*/*physis* contrast itself, several of them refer to law in such a way as to suggest rejection of the antithesis. Fragment 248 asserts that the aim of law is to benefit people, thus contradicting Glaucon's claim (Plato, *Republic* II 359c) that law constrains people contrary to their natural bent. Fragment 248 is supplemented and explained by fragment 245: laws interfere with people living as they please in order only to stop them from harming one another, to which they are prompted by envy. So law frees people from the aggression of others, thus benefiting them by giving them the opportunity to follow the promptings of nature towards their own advantage. The strongest expression of the integration of *nomos* and *physis* is found in fragment 252: the city's being well run is the greatest good, and if it is preserved everything is preserved, while if it is destroyed everything is destroyed. That is to say, a stable community is necessary for the attainment of that wellbeing which is nature's goal for us. This quotation encapsulates the central point in the defence of *nomos* (emphasized in Protagoras' myth at (Plato, *Protagoras* 322a-323a) that law and civilization are not contrary to nature, but required for human nature to flourish, a point also central to the Epicurean account of the development of civilization.

See also: Atomism, ancient

C.C.W. TAYLOR

References and further reading

Barnes, J. (1982) *The Presocratic Philosophers*, London: Routledge & Kegan Paul, 2nd edn.(Scholarly and argumentative; excellent bibliography; contains English translations.)

Barnes, J. (1987) *Early Greek Philosophy*, Harmondsworth: Penguin, 244-88.(Fragments and testimonia of Democritus in English translation.)

Democritus (mid 5th-4th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidmann, 6th edn, 1951, vol. 2, 81-230.(The standard collection of the ancient sources, includes Greek texts with translations in German. *Antiphon*, fragment 44, can be found in vol. 2 pages 346-55 (English trans. Barnes (1982), 509-12) and *Critias* fragment 25, in vol. 2 pages 386-9; English translations of the latter two can be found in Barnes (1982: 509-12 and 451-2).)

Furley, D. (1987) *The Greek Cosmologists* vol. 1, Cambridge: Cambridge University Press, 115-68.(Deals with principles of atomism, cosmology and anthropology.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press 6 vols. (Volume 2, pages 386-507 offer a comprehensive survey.)

Luria, S. (1970) *Demokrit*, Leningrad: Nauka.(A fuller selection of texts than Diels and Kranz (1951), with Russian commentary.)

Makin, S. (1993) *Indifference Arguments*, Oxford: Blackwell.(Atomistic arguments using the Principle of Sufficient Reason.)

McKirahan, R.D., Jr (1994) *Philosophy Before Socrates*, Indianapolis, IN: Hackett, 303-43.(Translation and commentary, covering both Leucippus and Democritus.)

Salem, J. (1996) *Démocrite: Grains de poussière dans un rayon de soleil (Democritus: grains of dust in a ray of sunlight)*, Paris: Vrin.(Comprehensive review of ancient evidence and modern literature; original texts in footnotes.)

Taylor, C.C.W. (1967) ‘Pleasure, Knowledge and Sensation in Democritus’, *Phronesis* 12: 6-27.(A sceptical rejoinder to Vlastos (1945, 1946).)

Vlastos, G. (1945, 1946) ‘Ethics and Physics in Democritus’, *Philosophical Review* 54: 578-92 and 55: 53-64, repr. in D.J. Furley and R.E. Allen (eds) *Studies in Presocratic Philosophy*, London, Routledge, 1975, vol. 2. (Argues for connection between ethical and physical theories.)

Demonstratives and indexicals

Demonstratives and indexicals are words and phrases whose interpretations are dependent on features of the context in which they are used. For example, the reference of 'I' depends on conditions associated with its use: as you use it, it refers to you; as I use it, it refers to me. In contrast, what 'the inventor of bifocals' refers to does not depend on when or where or by whom it is used. Among indexicals are the words 'here', 'now', 'today', demonstrative pronouns such as 'this', reflexive, possessive and personal pronouns; and compound phrases employing indexicals, such as 'my mother'. C.S. Peirce introduced the term 'indexical' to suggest the idea of pointing (as in 'index finger').

The phenomenon of indexicality figures prominently in recent debates in philosophy. This is because indexicals allow us to express beliefs about our subjective 'place' in the world, beliefs which are the immediate antecedents of action; and some argue that such beliefs are irreducibly indexical. For example, my belief that I am about to be attacked by a bear is distinct from my belief that HD is about to be attacked by a bear, since my having the former belief explains why I act as I do (I flee), whereas my having the latter belief explains nothing unless the explanation continues 'and I believe that I am HD'. It seems impossible to describe the beliefs that prompt my action without the help of 'I'. Similarly, some have argued that indexical-free accounts of the self or of consciousness are necessarily incomplete, so that a purely objective physicalism is impossible. In a different vein, some (such as Putnam 1975) have argued that our terms for natural substances, kinds and phenomena ('gold', 'water', 'light') are indexical in a way that entails that certain substantive scientific claims - for example, that water is H_2O - are, if true, necessarily true. Thus, reflection on indexicality has yielded some surprising (and controversial) philosophical conclusions.

1 Semantic data

An indexical is context-dependent: its interpretation varies with changes in aspects of the context of use. Yet almost any expression exhibits variation in interpretation with respect to some parameter or other. For example, a contingent sentence is one which is possibly true and possibly false, so we may think of it as varying in truth-value with changes in possible worlds. But mere contingency is not a species of indexicality. Similarly, tensed verbs induce variation in interpretation with respect to the passage of time, but tense is not usually viewed by linguists and logicians as a kind of indexicality (although there is some confusion about the matter). The difference between true context dependence and mere 'parameter dependence' can be seen as follows. Consider the sentence

(1) Jones will remember what Smith said just now.

The truth or falsity of this sentence at a moment in time *t* depends on the truth or falsity of the present tense sentence

(2) Jones remembers what Smith said just now

at moments t' future to *t*. But the referent of the occurrence of 'now' in (2) must refer back to *t* and not to any of the future moments t'. This means that the semantic rule for 'now' cannot be simply the rule which states that at any time *t*, 'now' refers to *t*. Instead, the rule must say that a use of 'now' at a time *t* will refer at time *t and at any other time* to *t*. Thus, the rule for 'now' must be 'double-indexed'. In general, the semantic rules for indexicals must involve the pairing of two sorts of parameter: contextual and circumstantial. The contextual parameters (speaker, time of utterance, place of utterance, world of utterance, addressee and so on) fix the referent of an indexical, so that its interpretation is unaffected by any circumstantial parameter (possible world, moment of time) required by the presence of tense or modal elements. Hereafter, the phrase 'context-dependent' will refer to the dependence of indexicals on both context and circumstance.

Indexicals are 'referential' in that they invariably have wide scope when embedded in tensed, modal or propositional attitude constructions. For example, consider the two statements 'The mayor might not have been a Democrat' and 'That person over there, the Mayor, might not have been a Democrat'. The former statement could be interpreted to mean merely that the office of mayor might have been occupied by a non-Democrat; but the latter statement can only be interpreted to mean that a certain person (who happens to be the mayor) might not have been

a Democrat. This property of indexicals is an effect of their context dependence. The referent of the indexical phrase ‘that person’ is fixed by the context and not by circumstances required to interpret the modal element ‘might have been’.

Though the interpretation of an indexical will vary with the context, indexicals are not in general ambiguous. Your use of ‘I’ and mine are linked by a common meaning. The language supplies a rule - the same rule for both of us - for the use of ‘I’. The rule is roughly this: ‘As *x* uses "I" in such and such a context, "I" refers to *x* in all circumstances’. But this sort of meaning is not like the Fregean sense of a definite description or proper name; it is not a concept of some fixed object and it cannot be combined with the sense of a (non-indexical) predicate to yield a complete proposition (see Frege, G. §§2-3). Some theorists believe that this fact requires the introduction into semantics of a whole new level of meaning distinct from both sense and reference. Kaplan, for example, draws a three-way distinction between the ‘character’ (linguistic meaning), content (roughly, sense) and referent of an indexical expression such as ‘my mother’ (1970). Kaplan’s view may not be inevitable, however. It may be that ‘I’ has an ‘incomplete sense’, in the way that the sense of a functional expression ‘the square root of *x*’ or ‘the author of *x*’ is incomplete. If so, that would explain why the linguistic meaning of ‘I’ cannot be combined with a predicate to yield a complete proposition and there would be no need to postulate a kind of meaning in addition to sense and reference.

Some subscribe to the principle that for one to understand an ‘object-dependent’ proposition - a proposition that such and such an object has such and such a property - one must be able to identify the object in question; one must be able to say who or what it is. This principle seems to be directly contradicted by propositions involving indexicals. I can know that I am hungry or that it is raining now without having any substantive information about who I am or what time it is. Similarly, if I point to something and say ‘I don’t have the slightest idea what that is’ I might well speak the truth! Yet these propositions are object-dependent in that they would not be available if the objects they are about did not exist. Let us mark the fact that speakers can use indexicals with total understanding without having any substantive knowledge of their referents by saying that indexicals are ‘transparent’.

Some indexical sentences have the property of being true in any context of use. For example, one who says (or thinks) ‘I exist’ or ‘I am here now’ or ‘I am saying (or thinking) something now’ must be right. These sentences, then, are somewhat similar to logical truths. But if I say ‘I exist’, then what I say (that HD exists) is not a *necessary* truth, for there are possible circumstances in which I do not exist. So a sentence such as ‘I exist’ is contingent in one respect and yet non-contingent in another. Any theory of indexicality must explain how this is possible.

Perhaps the most puzzling aspect of indexicals is that some of them seem to be ineliminable and their ineliminability seems to have to do with the baffling distinction between subjectivity and objectivity (see Content, indexical). Perry gives the following example (1979). Suppose a meeting is set for Tuesday at noon. All week you believe that the meeting will start at noon on Tuesday. At a certain point, you leave to go to the meeting. Obviously, you must have acquired some new information, a new belief, distinct from the belief that the meeting starts at noon on Tuesday, for you now act in a way you have not acted all week (that is, you leave to go to the meeting). What belief is this? It is the belief that the meeting starts *now*, or, more elaborately, the belief that the meeting starts at noon on Tuesday and it is *now* noon on Tuesday. The use of ‘now’ in describing your new belief seems essential, and not eliminable in favour of non-indexical designations of noon on Tuesday. Moreover, your new ‘indexical belief’ places you within the fabric of time in a way that the beliefs you acquire when you look at your appointment book to determine the time of the meeting do not. An observer ‘outside of time’ could have the belief that the meeting starts at noon on Tuesday, but not the belief that the meeting starts ‘now’. In this way, your indexical belief seems linked to your subjective place in the world. Perry calls such beliefs ‘self-locating’ beliefs. Yet all this is quite puzzling. Is your belief that it is ‘now’ noon on Tuesday something that you can believe only ‘now’? Certainly, you can believe truly that it is now noon on Tuesday only at noon on Tuesday. But can you not at a later time continue to hold precisely the belief you expressed earlier by saying ‘It is now noon on Tuesday’? Yet if you believe at the later time exactly what you believed at the earlier, then why do you not, at the later time, do just what you did at the earlier time - namely, head off to the meeting? Similarly, I believe that I am HD. Can you not believe that too? If you believe just what I do when I believe that I am HD - that is, if you believe exactly the proposition I do, it would appear to follow that our mental states *vis-à-vis* this belief must be identical, then why do you not have any inclination, as I do, to identify yourself as HD? Let us call this ‘the puzzle about

indexical belief'. The puzzle has two parts: first, one argues on the basis of behavioural changes (for example, your getting up to go to the meeting) that an indexical belief (assertion, thought) is not equivalent to any non-indexical counterpart of it. Second, however, similar behavioural changes - or the lack of any - seem to show that such beliefs are 'isolated' - that they cannot be retained over time or shared by different persons: only I can know that I am HD, and that seems counterintuitive. Any adequate theory of indexicality must solve this puzzle about indexical belief.

2 The Kaplan-Perry theory

The most influential theory that has had some success explaining the semantic data about indexicals as outlined above was developed by Kaplan in a series of papers originating in the late 1960s and early 1970s. In the mid-1970s, Perry applied Kaplan's theory to explain the data about the ineliminability of indexicality and the puzzle about indexical belief - data that were first noticed and systematically studied by Castañeda (1967).

According to Kaplan-Perry theory (KPT), each indexical expression has a 'character' or linguistic meaning (Perry's term is 'role') and, given a context, a 'content' in that context. The content in a context of a (declarative) sentence containing an indexical is a 'singular proposition' - a proposition containing the very referent of the indexical as a 'constituent' rather than a sense or concept of that referent. So the content of my use of 'I am hungry now' at a time *t* is a proposition to the effect that HD is hungry at time *t* - with HD right there 'in' the proposition, in at least a metaphorical sense. Kaplan takes the character of an indexical to be a function (in the mathematical sense) that takes context into content. Contexts are treated as *n*-tuples $c = \langle a, t, p, w, \ldots \rangle$ with coordinates for the 'agent of the context' (to handle 'I'), the 'time of the context' (to handle 'now'), the 'place of the context', the 'world of the context' and so on (for example, there should be a coordinate for the 'day of the context' to handle 'yesterday', 'today' and 'tomorrow'). To deal with pure demonstratives such as 'this', Kaplan adds a coordinate for a 'demonstration'. This is a kind of abstract counterpart of an act of pointing. For example, the sentence 'That is mine now' will be true at a parameter $\langle a, t, p, w, d, \ldots \rangle$ if the demonstrated object *d* belongs to agent *a* at *t* in place *p* and world *w*.

Kaplan distinguishes between 'contexts' and 'circumstances'. Formally, the latter are pairs $\langle t, w \rangle$ with *t* a time and *w* a possible world. Strictly, expressions are evaluated relative to parameters of the form $\langle \text{context}, \text{circumstance} \rangle$. So the semantic rules are double-indexed. In addition, Kaplan claims that indexicals are 'directly referential'. By this he means that an indexical refers without the aid of a Fregean sense. In this respect indexicals are like proper names. But proper names are, in Kaplan's view, unlike indexicals in that the latter possess characters whereas the former do not (see Proper names §§2-3).

By appealing to the content/character and context/circumstance distinctions, KPT can readily explain the fact that indexicals are context-dependent, referential and univocal. Moreover, the character/content distinction helps to explain the transparency of indexicals. One need only grasp the character of an indexical to use it with understanding. On occasion, its content may be hidden, as in the case of amnesiacs, who, despite their affliction, can certainly use the first person pronoun perfectly competently. KPT can also explain the special status of sentences such as 'I exist'. The character of 'I exist' always produces a true content in any context. But that content may turn out to be false in further circumstances. For example, if you now write 'I exist' in your diary, the content you express is true now. But, inevitably, that content will be false years hence in circumstances in which you no longer exist and someone then reads these words in your diary. According to KPT, 'I exist' is a logical truth, although its necessitation, 'Necessarily, I exist', and its temporal necessitation, 'It is always the case that I exist', are not. KPT thus involves a novel conception of logical truth. For it is usually assumed that the class of logical truths contains 'Necessarily, *P*' and 'It is always the case that *P*' if it contains *P*. Finally, KPT solves the problem of indexical belief by appealing to the notion of character. KPT holds that belief is linked to action via character, not content. If you and I have beliefs with the same content but different character we may well behave differently. But if we have beliefs with the same character and different content we may well behave in the same way. When each of us thinks 'I am about to be attacked by a bear' (same character, different content) we both run for our lives, but when I think 'I am about to be attacked by a bear' and you think 'You are about to be attacked by a bear' (same content, different character) I run for my life and you do not. Similarly, you can retain the belief that the meeting starts now but you can express it at a later time only by using words having a character distinct from that of 'The meeting starts now', and it is the difference in character that accounts for the fact that, at the later

time, you do not do what you did at the earlier time. Likewise, KPT appeals to character to account for the difference between the thought that I am about to be attacked by a bear and the thought that HD is about to be attacked by a bear. The former has a certain character which the latter does not, and that makes all the difference. Similarly, consider the difference between the thought that I am HD and that HD is HD. The former is informative whereas the latter is not. Again, the difference can be traced to the fact that 'I' has a certain character which 'HD' does not.

KPT needs only the character/content and context/circumstance distinctions to account for the semantic data. But it goes further in claiming that indexicals are 'directly referential' and that indexicals give rise to 'singular propositions'. Proponents of KPT argue in several ways for the direct reference thesis. First, Kaplan argues that 'I' does not have the sense of 'the speaker' or any description such as 'the utterer' or 'the thinker'. He would argue, for example, that since 'I am not the speaker' is not contradictory, whereas 'The speaker is not the speaker' is contradictory, 'I' and 'the speaker' are not synonymous. But this argument seems equivocal. Any candidate synonym for 'I' must be context-dependent (and not just parameter-dependent). If we associate a context-dependent rule with 'the speaker' - so that it refers in any circumstance to the 'speaker of the context' - then 'I am not the speaker' is indeed contradictory. It is only by treating 'I' as context-dependent and 'the speaker' as merely parameter-dependent (as it is in one sense) that there is a discrepancy. Second, Kaplan argues in the following way. The descriptive meaning of 'now' is given by the following rule: at *t*, 'now' refers to *t*. Thus, the descriptive meaning of 'now' is 'either inapplicable or irrelevant to determining a referent with respect to a circumstance of evaluation' (1989: 500). Hence, 'now' is directly referential - its descriptive meaning does not serve as the route to its referent in a circumstance. This argument does not work either. One can take the view, as explained above, that the descriptive rule for 'now' is not the one Kaplan assumes. (So the first premise of Kaplan's argument is false.) And if we substitute the double-indexed rule for 'now' it cannot be argued that descriptive meaning (that is, the double-indexed rule) of 'now' is 'irrelevant or inapplicable' to determining its referent in a circumstance.

See also: Logical and mathematical terms, glossary of;Reference §5

HARRY DEUTSCH

References and further reading

Braun, D. (1995) 'What is Character?', *Journal of Philosophical Logic* 24: 227-40.(A critical and relatively nontechnical discussion of some aspects of Kaplan's notion of character.)

Castañeda, H.-N. (1967) 'Indicators and Quasi-Indicators', *American Philosophical Quarterly* 4: 85-100.(This is a representative example of several pioneering papers by this author exploring the behaviour of indexicals in propositional attitude constructions.)

Forbes, G. (1989) *Languages of Possibility*, Oxford: Oxford University Press.(This work includes a detailed neo-Fregean theory of indexicality. In contrast, the Kaplan-Perry theory discussed above may be described as 'anti-Fregean'.)

Kamp, H. (1971) 'Formal Properties of "Now"', *Theoria* 37: 227-73.(This paper develops a double-indexed semantics for certain tense logics with an operator representing 'now'. Kamp was the first to recognize the need for the double-index technique.)

Kaplan, D. (1970) 'On the Logic of Demonstratives', *Journal of Philosophical Logic* 8: 81-98.(The pioneering paper in which the key distinctions between character and content and context and circumstance were first introduced.)

Kaplan, D. (1989) 'Demonstratives', in *Themes from Kaplan*, ed. J. Almog, J. Perry and H.K. Wettstein, Oxford: Oxford University Press.(This is a classic meditation on the anti-Fregean aspects of the Kaplan-Perry theory.)

Perry, J. (1977) 'Frege on Demonstratives', repr. in P. Yourgrau (ed.) *Demonstratives*, Oxford: Oxford University Press, 1990.(This paper attempts to demonstrate the inability of Frege's semantics to handle indexicality and the inadequacy of Frege's own attempts to deal with the phenomenon.)

Perry, J. (1979) 'The Problem of the Essential Indexical', *Noûs* 13: 3-21.(By means of several ingenious examples this paper uncovers and attempts to solve the problem of indexical belief.)

Putnam, H. (1975) 'The Meaning of "Meaning"', in *Mind, Language and Reality: Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press.(This paper is the main source of the view according to which our terms for natural substances, kinds and phenomena are indexical.)

Yourgrau, P. (ed.) (1990) *Demonstratives*, Oxford: Oxford University Press.(This is a collection of essays on

indexicality which emphasize not only semantic issues but also metaphysical and epistemological themes.)

Dennett, Daniel Clement (1942-)

A student of Gilbert Ryle and a connoisseur of cognitive psychology, neuroscience and evolutionary biology, American philosopher Daniel Dennett has urged Rylean views in the philosophy of mind, especially on each title topic of his first book,Content and Consciousness (1969). He defends a broadly instrumentalist view of propositional attitudes (such as belief and desire) and their intentional contents; like Ryle and the behaviourists, Dennett rejects the idea of beliefs and desires as causally active inner states of people. Construing them in a more purely operational or instrumental fashion, he maintains instead that belief- and desire-ascriptions are merely caculational devices.

Dennett offers a severely deflationary account of consciousness, subjectivity and the phenomenal or qualitative character of sensory states. He maintains that those topics are conceptually posterior to that of propositional-attitude content: the qualitative features of which we are directly conscious in experience are merely the intentional contents of judgments.

1 Propositional attitudes

Rejecting the idea of beliefs and desires as inner causes, Dennett maintains instead that ascriptions of belief and desire are merely calculational devices, that happen to have predictive usefulness (see Ryle, G.; Behaviourism, analytic). In interpreting other creatures, which is something we do constantly without knowing a thing about those creatures' insides, we start by trusting that they will believe what they should believe given their perceptions and desire what they should desire given their needs; then we correct empirically for observed departures from those normative ideals. Such ascriptions are often objectively true, he grants, but not in virtue of describing inner mechanisms, any more than references to centres of gravity, kinematic vectors and the Equator describe inner mechanisms.

Thus Dennett is an instrumentalist of sorts about propositional attitudes, such as belief and desire. To ascribe a 'belief' or a 'desire' is not to describe some segment of physical reality, Dennett says, but is more like performing part of a calculation. But he also protests that his own interpretation-based view does not rob his propositional-attitude concepts of their objectivity, for he understands the truth of an attitude ascription as being determined by whether someone's taking a certain interpretive stance *would in fact* achieve good predictive results. Still, for Dennett, attitude content is conceptually constructed out of purpose-relative interpretation; his objectivity does not amount to determinacy (he explicitly endorses Quine's indeterminacy thesis for content); and for him, a thing's innards (including the matter of whether it has any) are conceptually irrelevant to its having intentional states.

Dennett has always used mathematical abstractions, such as centres of gravity, to illustrate the sort of view he favours. It should be noted, however, that he has recently used this comparison to moderate his line. In an important paper, 'Real Patterns' (1991a), he emphasizes the reality of mathematical abstractions and de-emphasizes interpretation.

Dennett offers five grounds for his rejection of the common-sensical inner-cause thesis. (1) He thinks it ludicrously unlikely that any science will ever turn up any distinctive inner-causal mechanism that would be shared by all the possible subjects that had a particular belief. (2) He offers numerous objections to 'language-of-thought psychology', the most popular inner-cause theory (see Language of thought). (3) He compares the belief/desire interpretation of human beings to that of lower animals, chess-playing computers and even lightning rods, arguing that in their case we have no reason to think of belief- and desire-ascriptions as other than mere calculational or predictive devices, and then that we have no more reason for the case of humans to think of belief- and desire-ascriptions as other than such devices. (4) Dennett argues that the verification conditions of belief- and desire-ascriptions are basically a matter of extrapolating rationally from what subjects ought to believe and want in their circumstances, and then he boldly identifies the truth-makers of those ascriptions with these verification conditions, challenging inner-cause theorists to show why instrumentalism does not accommodate all the actual evidence. (5) He argues that in any case if a purely normative assumption (the 'rationality assumption', that people will generally believe what they ought to believe and desire what they should desire) is required for the licensing of an ascription, then the ascription cannot itself be a purely factual description of a plain state of affairs (see

Propositional attitudes).

2 Freedom of the will

In *Elbow Room* (1984) Dennett defends a common-sensical compatibilism regarding free will and determinism against attacks by metaphysicians and also against other common-sensical 'intuitions' that he takes to be confused. Here his style as well as his substance is Rylean: he does not engage serious contemporary incompatibilist arguments, but only jollies the reader into thinking that such arguments must be misguided. However, he does offer positive compatibilist accounts of practical reason, control over one's actions, deliberation and the appropriateness of holding people responsible (see Free will).

3 Consciousness

Dennett's mature view of consciousness and related topics is expounded at length in *Consciousness Explained* (1991b). His main claims are that there is nothing to the human mind that is not already constituted by the brain; that the brain itself is but a bag of tricks, an only very loose improvised association of little specialists whose individual actions contribute piecemeal and somewhat haphazardly to a subject's operative psychological state; that linguistic utterances do not normally express determinate pre-existing thought contents; and that there is no single 'whole me', or self or Boss Unit or even CPU within the brain, serving as chief executive of my utterings and other actions (see Consciousness).

Dennett is especially concerned to deny that there is any determinate stage of information processing that is the locus of conscious mental states or events; there is no such locus at all, however physically characterized. He calls his target 'Cartesian materialism', that being the usually tacit assumption that there is a physically realized spatial or temporal turnstile in the brain, a stage where 'it all comes together' and the product of pre-processing is exhibited 'to consciousness'.

Dennett gives two arguments against Cartesian materialism. First, given his naturalism about brains and the unplanned nature of their organization, it is unlikely that nature has furnished the human brain with any central viewing-room or single monitor to do the viewing; nor is there any positive neurophysiological sign of such organs. Second, Dennett adverts to the famous 'temporal anomalies' of consciousness discovered by psychophysical research, such as color phi, the cutaneous rabbit and Libet's 'backward referral' of sensory experiences, in which, for example, the mind seems to detect a stimulus before that stimulus could possibly have been processed by the brain. Dennett argues at length that those curious phenomena are anomalous only as long as Cartesian materialism is being assumed; jettison the assumption, and the phenomena are readily explained.

Dennett's recommended alternative to Cartesian materialism is the 'Multiple Drafts' model:

> What we actually experience is a product of many processes of interpretation - editorial processes, in effect. They take in relatively raw and one-sided representations, and yield collated, revised, enhanced representations, and they take place in the streams of activity occurring in various parts of the brain.…Once a particular 'observation' of some feature has been made, by a specialized, localized portion of the brain, the information content thus fixed does not have to be sent somewhere else to be rediscriminated by some 'master' discriminator.
>
> (1991b: 112-13)

The resulting stream of contents, subject to continual editing by many processes distributed around in the brain, is not altogether like a narrative, because of its multiplicity; at any point in time there are 'multiple drafts' of narrative fragments at various stages of editing in various places in the brain.

Dennett grants that there is something of the sort often called 'the stream of consciousness': 'Conscious human minds are more-or-less serial virtual machines implemented - inefficiently - on the parallel hardware that evolution has provided for us' (1991b: 218). They are 'Joycean' virtual machines that formulate synthesized reports of their own passing states, though the reports are almost never entirely accurate.

By way of *explaining* consciousness as advertised, Dennett offers a 'method for phenomenology', called 'heterophenomenology'. He says the method is designed to 'do justice to the most private and ineffable subjective experiences, while never abandoning the methodological scruples of science' (1991b: 72). It is to listen carefully

to a subject's introspective reports, respect them as perfectly sincere, and take them as serious descriptions of the subject's 'heterophenomenological world'. That world will be 'a stable, intersubjectively confirmable theoretical posit' (1991b: 81). But the cognitive scientist is to take the subject's narrative as fiction, or, less pejoratively, as a narrative that may or may not be true, just as an anthropologist takes the narratives of a primitive people. The job of the psychologist is then to discover whatever it is in the brain's great chaos of information processing that makes subjects produce the narratives that they do. To make that discovery is, in one sense, to explain the subject's consciousness.

Some readers have insisted that Dennett has simply avoided the real issues. If all the scientist need do is explain propositional sayings in terms of internal propositional states and episodes, then we know already that neither the scientist nor Dennett will find or even seek anything about phenomenal character, subjectivity, qualia, or any other feature of consciousness that makes it interesting over and above bare intentionality or aboutness *per se*. It may seem that Dennett has simply changed the subject. But he does squarely address the questions of subjectivity and phenomenal character in particular. He rejects the thesis of Thomas Nagel and Frank Jackson that an omniscient objective science would necessarily leave the introspectible first-person aspects of sensory experiences unexplained and even unrecorded, and he treats phenomenal properties of sensations essentially as representational contents, properties ascribed truly or falsely to external objects or to parts of one's own body (see Qualia).

WILLIAM G. LYCAN

List of works

Dennett, D.C. (1969) *Content and Consciousness*, Routledge & Kegan Paul.(Dennett's early views on intentionality, conscious awareness and the mind-body problem.)

Dennett, D.C. (1978) *Brainstorms*, Montgomery, VT: Bradford Books.(Most of Dennett's classic papers, on propsositional attitudes, consciousness and personhood.)

Dennett, D.C. (1984) *Elbow Room*, Cambridge, MA: Bradford Books/MIT Press.(Dennett on freedom of the will.)

Dennett, D.C. (1987) *The Intentional Stance*, Cambridge, MA: Bradford Books/MIT Press. (Dennett's mature theory of intentionality and propositional attitudes.)

Dennett, D.C.(1991a) 'Real Patterns', *Journal of Philosophy* 88: 27-51. (A shift, at least of emphasis, in Dennett's theory of propositional attitudes.)

Dennett, D.C. (1991b) *Consciousness Explained*, Boston, MA: Little, Brown.(Dennett's mature theory of consciousness and the qualitative character of experience.)

Dennett, D.C. (1994) 'Self-Portrait', in S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*, Oxford: Blackwell.(A summary of Dennett's philosophy of mind.)

Dennett, D.C. (1995) *Darwin's Dangerous Idea*, New York: Simon & Schuster.(An interpretation of Darwin's theory of evolution by natural selection, and a drawing of philosophical morals for the general reader.)

Dennett, D.C. (1996) *Kinds of Minds*, New York: Basic Books.(A new summary of Dennett's philosophy of mind, with some changes of emphasis and some extra attention to nonhuman animals.)

References and further reading

Dalhbom, B. (ed.) (1993) *Dennett and His Critics: Demystifying Mind*, Oxford: Blackwell.(Essays on Dennett and a collective reply. Full bibliography of Dennett's writings.)

Philosophical Topics (1994) Special journal issue 22 (1, 2).(An issue devoted to Dennett's work; essays and a collective reply.)

Stich, S.P. (1981) 'Dennett on Intentional Systems' *Philosophical Topics* 12 (1): 39-62; repr. in W.G. Lycan (ed.) *Mind and Cognition*, Oxford: Blackwell, 1990. (A very clear exposition and critique of Dennett's intentional instrumentalism.)

Symposium (1988) *The Intentional Stance, Behavioral and Brain Sciences* 11 (3): 495-546.(Dennett's own précis of the book, twenty-four brief commentaries by philosophers and various cognitive scientists, and Dennett's collective response.)

Denys the Carthusian (1402/3-71)

Denys de Leeuwis was born in the village of Rijkel, in modern Belgium. In 1421 he matriculated at the University of Cologne, where he received the Master of Arts degree in 1424. There he followed 'the way of Thomas Aquinas', whom he calls his 'patron' in his early works. Later Denys adopted 'Albertist' against 'Thomist' positions on a number of philosophical issues. After leaving the University, he entered the Carthusian monastery in Roermond, where, save for brief periods, he spent the rest of his life. He corresponded with Nicholas of Cusa and dedicated two or three works to him. Denys was a voracious reader of the ancient and medieval philosophers whose writings were available in Latin, and of scholastic theologians. Because of his extensive references to authorities, historians often call him 'eclectic'. Yet from his sources he educes his own distinctive philosophy. Like Albert the Great, Denys practised philosophy and theology by paraphrasing and analysing their histories.

Denys was one of the most prolific writers of the Middle Ages. His writings include scholastic, mystical, moral and pastoral commentaries and treatises, commentaries on each book of Scripture, over 900 sermons, and paraphrase/commentaries on works by Aquinas. His major philosophical works are *De lumine christianae theoriae (The Light of Christian Theory)*, encyclopedic commentaries on Peter Lombard's *Sentences* (see Lombard, P.) *compendia* of philosophy and theology that distill materials from his commentaries on the *Sentences* and summarize his own positions, a commentary on Boethius' *Consolation of Philosophy* (see Boethius, A.M.S.) and commentaries on all of the writings of Pseudo-Dionysius. In Book I of *De lumine*, Denys analyses the teaching of many 'Peripatetic' and 'Platonic' philosophers on a wide range of metaphysical questions. Likewise, on each question of the *Sentences* he reports and interprets the teaching of many scholastic doctors. He discounts the arguments of 'nominalists' as more concerned with 'terms' than 'realities', returning instead to the teachings of thirteenth-century masters.

Like many fifteenth-century philosophers, Denys was preoccupied with the question of the soul's immortality. Influenced especially by Avicenna (see Ibn Sina) and Albert the Great, he maintains that the immortality of the soul can be demonstrated by reason alone (see Soul, nature and immortality of). In this context, he argues against Aquinas that the human mind need not have recourse to phantasms in every act of knowing and that it is capable of knowing separated substances. The mind's ability to know the essences of things without phantasms is the surest evidence of its immortality, and its natural capacity for pure abstraction is the foundation for the imageless contemplation of mystical theology.

Denys' understanding of the soul's self-subsistence and its ways of knowing affects other philosophic doctrines. He argues that the soul knows some first principles by immediate self-reference, and that the individuation of persons is not effected by matter alone, but by the whole human composite, most powerfully by its spiritual form. He disputes a 'real' distinction between essence and existence in creatures, arguing instead - in a way reminiscent of Henry of Ghent - for an 'intentional' distinction between them (see Existence). His noetic theory admits the possibility of a natural, philosophic beatitude, wherein the soul contemplates separated substances in the participated light of the lowest Intelligence in the hierarchy of celestial beings. In accord with these conclusions, Denys draws a sharp distinction between 'natural, philosophic wisdom' and 'supernatural, theological wisdom', which are analogically coordinated and touch at the points of God's existence, unity and attributes. True philosophical conclusions never contradict divine revelation. Here Denys' 'eclecticism' serves him well; among the philosophers he discovers those teachings most accordant with the 'analogy of faith', reconciles them among themselves, and brings them into rapport with the divinely revealed standard. Denys' departure from Thomistic conceptions was prompted by his reading of Pseudo-Dionysius, whom he considered to be the greatest authority of Christian wisdom.

Nevertheless, Denys retained many of the insights of Aristotle and Thomas Aquinas. He states that Aristotle's understanding of God as 'the pure act of being' is one of the most sublime of all philosophical teachings. He affirms that the intellectual soul is the single, substantial form of the human body, grants that human knowing ordinarily requires phantasms, and argues a 'real distinction' among the faculties of the soul. He maintains the priority of the intellect over the will in every human act, and teaches that love in the will is exactly commensurate to light in the intellect.

Denys conceived his philosophical enterprise as a remedy to attitudes common among his monastic brethren, who often scorned scholastic teaching and exalted the affections and the will in the spiritual life. Beyond that, he sought to reconcile the varying doctrines of the philosophers and the scholastic theologians by means of the higher, unified light of mystical theology.

See also: Aristotelianism, medieval

KENT EMERY, JR

List of works

Denys the Carthusian (1402/3-71) *Opera omnia*; ed. monks of the Carthusian Order, *Doctoris ecstatici D. Dionysii Cartusiani Opera omnia*, Montreuil/Tournai/Parkminster: Typis Cartusiae S.M. de Pratis, 1896-1935, 42 vols.(Essentially a reprinting of sixteenth-century editions, with excellent indexes of names and topics.)

Denys the Carthusian (1440-5) *De contemplatione (On Contemplation)*; in *Opera omnia* 41: 133-289.(This is Denys' major work on contemplative theory and practice; it treats the different kinds of contemplation, recites the teaching of the most notable patristic and medieval masters, and resolves disputed questions concerning the nature of mystical union.)

Denys the Carthusian (*c*.1451) *De lumine christianae theoriae (The Light of Christian Theory)*; in *Opera omnia* 33: 233-513.(In the first book of this work, Denys recites and analyses philosophical doctrines about the 'going out' of all things from the first principle; in the second book, based largely on Aquinas' *Summa contra gentiles*, he treats the 'return' of human creatures to God through Christian grace and perfection.)

Denys the Carthusian (*c*.1464) *Commentaria in IV libros Sententiarum (Commentaries on the Four Books of the Sentences)*; in *Opera omnia* vols 19-25.(In these massive commentaries, Denys arranges dialectically, recites and comments upon or criticizes the opinions of many scholastic thinkers on each question of Peter Lombard's *Sentences*.)

Denys the Carthusian (*c*.1465) *Elementatio philosophica (Elements of Philosophy)*; in *Opera omnia* 33: 21-231.(This is an epitome of the strictly philosophic matters and questions extracted from the commentaries on the *Sentences* and arranged in a proper philosophical order.)

Denys the Carthusian (*c*.1465) *Elementatio theologica (Elements of Theology)*; in *Opera omnia* 33: 21-231.(In this work, Denys extracts from his commentaries on the *Sentences* strictly theological matters and questions and states his own conclusions.)

Denys the Carthusian (*c*.1466) *Ennarationes in V libros De consolatione philosophiae B. Severinii Boetii (Commentary on Boethius' Consolation of Philosophy)*; in *Opera omnia* vol. 26.(In this work, cast in the form of a dialogue between a master and a disciple, Denys gives a 'literal' philosophic commentary on Boethius' famous text, to which he joins 'allegorical', theological readings.)

Denys the Carthusian (1466-7) *Commentaria in libros S. Dionysii Areopagitae (Commentaries on the Writings of Dionysius the Areopagite)*; in *Opera omnia* vols 15-16.(With Albert the Great, Denys was the only Latin writer to comment on each work in the Dionysian corpus, including the letters; his commentaries are line-by-line, and include articles about disputed points of interpretation.)

References and further reading

Beer, M. (1963) *Dionysius' des Kartausers Lehre vom 'desiderium naturale' des Menschen nach der Gotteschau (Denys the Carthusian's Teaching on the Human Being's 'Natural Desire' for the Vision of God)*, Münchener theologische Studien II: Systematische Abteilung 28, Munich: M. Hueber.(A study of Denys' teaching on the human being's natural desire for beatitude, which involves important questions on the relations between nature and grace, philosophy and theology. On this topic, Denys criticized the teachings of Thomas Aquinas and Giles of Rome.)

Emery, K., Jr (1988) 'Twofold Wisdom and Contemplation in Denys of Ryckel (Dionysius Cartusiensis, 1402-1471)', *Journal of Medieval and Renaissance Studies*, 18: 99-134.(A study of Denys' hierarchical noetic ordering of philosophic, scholastic and mystical wisdom.)

Emery, K., Jr (1990) 'Theology as a Science: The Teaching of Denys of Ryckel', in R. Työrinoja, A.I. Lehtinen and D. Føllesdal (eds) *Knowledge and the Sciences in Medieval Philosophy: The Proceedings of the Eighth International Congress of Medieval Philosophy (S.I.E.P.M.)*, Helsinki: Publications of Luther-Agricola Society, series B 19, 376-88.(The scientific foundation of scholastic theology is the vision of the blessed, accessible in part to transient mystical intuition.)

Emery, K., Jr (1991) *Dionysii Cartusiensis Opera selecta 1: (Prolegomena) Bibliotheca manuscripta 1A-1B: Studia bibliographica*, Corpus Christianorum Continuatio Mediaeualis 121-1a, Turnhout: Brepols.(A study of

the surviving manuscripts of Denys' writings; a newly established list of the authentic works, and studies of the transmission of spurious works.)

Emery, K., Jr (1992a) 'Denys the Carthusian and the Doxography of Scholastic Theology', in M.D. Jordan and K. Emery, Jr (eds) *Ad litteram: Authoritative Texts and their Medieval Readers*, Notre Dame, IN: University of Notre Dame Press, 327-59.(A study of the composition and method of Denys' commentaries on the Sentences, and his evaluation of the scholastic tradition on questions concerning the divine attributes, individuation, and so on.)

Emery, K., Jr (1992b) '*Sapientissimus Aristoteles and Theologicissimus Dionysius*: The Reading of Aristotle and the Understanding of Nature in Denys the Carthusian', in A. Speer and A. Zimmerman (eds) *Mensch und Natur im Mittelalter*, Miscellanea Mediaevalia 21/2, Berlin: de Gruyter, 572-606.(A study of Denys' philosophy of nature, on questions concerning issues such as the immortality of the soul, the composition of human beings, eternity of the world, emanation, and the authority of Aristotle and Pseudo-Dionysius.)

Emery, K., Jr (1994) 'Denys the Carthusian and the Invention of Preaching Materials', *Viator* 25: 377-409.(A study of the organization of Denys' moral writings and their philosophic foundation.)

Macken, R. (1984) 'Denys the Carthusian, Commentator on Boethius's *De consolatione philosophiae*', *Analecta Cartusiana* 118: 1-70.(A study of Denys' treatment of noetic questions in the commentary.)

Maginot, N. (1968) *Der 'Actus humanus moralis' unter dem Einfluss des Heiligen Geistes nach Dionysius Carthusianus(The 'Moral Act' Under the Influence of the Holy Spirit According to Denys the Carthusian)*, Münchener theologische Studien II. Systematische Abteilung 35, Munich: M. Hueber.(A study of Denys' analysis of the requirements for consciously moral action, in relation to his teaching on the gifts and influence of the Holy Spirit in human action. This is a classic topic in the tradition of neo-scholastic philosophy.)

Stoelen, A. (1957) 'Denys le Chartreux', in *Dictionnaire de spiritualité ascétique et mystique, histoire et doctrine*, Paris: Beauchesne, vol. 3: 430-49.(An excellent introduction to Denys' life, works and doctrine, with a chronology of the works and a good bibliography of older scholarship.)

Teeuwen, P. (1938) *Dionysius de Karthuizer en de philosophisch-theologische Stroomingen aan de Keulsche Universiteit (Denys the Carthusian in the Philosophical and Theological Cross-Currents of the University of Cologne),* Brussels/Nijmegen: Standaard-Boekhandel.(An important study, in Dutch, of Denys' relation to the philosophic disputes in the fifteenth-century University of Cologne.)

Wassermann, D. (1996) *Dionysius der Kartäuser: Einführung in Werk und Gedankenwelt(Dionysius the Carthusian: An Introduction to His Work and World of Thought)*, Analecta Cartusiana 133, Salzburg: Institut für Anglistik and Amerikanistik.(Wassermann argues that Denys, like the Carthusians in general, followed an older, monastic pattern of learning, and he emphasizes Denys' concern for personal and social religious reform.)

Deontic logic

Deontic logic is the investigation of the logic of normative concepts, especially obligation ('ought', 'should', 'must'), permission ('may') and prohibition ('ought not', 'forbidden'). Deontic logic differs from normative legal theory and ethics in that it does not attempt to determine which principles hold, nor what obligations exist, for any given system. Rather it seeks to develop a formal language that can adequately represent the normative expressions of natural languages, and to regiment such expressions in a logical system.

The theorems of deontic logic specify relationships both among normative concepts (for example, whatever is obligatory is permissible) and between normative and non-normative concepts (for example, whatever is obligatory is possible). Contemporary research beginning with von Wright treats deontic logic as a branch of modal logic, in so far as (as was noted already by medieval logicians) the logical relations between the obligatory, permissible and forbidden to some extent parallel those between the necessary, possible and impossible (concepts treated in 'alethic' modal logic).

1 Standard deontic logic (SDL)

Deontic logic is the logic of normative reasoning. It is common in deontic logic to take one of the three concepts - 'obligatory', 'forbidden', 'permissible' - as primitive and to define the others in terms of the chosen one, the most popular approach defining the permissible and forbidden in terms of the obligatory. In 'standard deontic logic' (SDL) a modal operator *O* is added to the language of propositional logic so that if *p* is a sentence of the language, then so is *Op*, which informally expresses 'It is obligatory that *p*'. (Alternative approaches treat *O* as a predicate applied to names of act types, or as an operator on what Castañeda (1981) calls 'practitions'.) The standard definitions of 'permissible' and 'forbidden' are $Pp = \sim O \sim p$ and $Fp = O \sim p$, respectively.

Possible worlds semantics for SDL assumes that relative to any given normative system there is for each possible world *w* a non-empty set of 'deontic alternatives' for *w*. This set functions as the normative standard for *w*, so that *Op* is said to be true at *w* (relative to the given normative system) just in case *p* is true at each of *w*'s deontic alternatives.

This is put more formally in terms of model structures $\langle W, R \rangle$, where *W* is a set of possible worlds and *R* a two-place relation on *W*. An SDL model is a model structure $\langle W, R \rangle$ plus a function *I* which assigns truth-values to each $\langle p, w \rangle$, for atomic sentences *p* and worlds *w*, and obeys the usual conditions for truth-functional compounds as well as the following condition for *O*: *Op* is true at *w* in $\langle W, R, I \rangle$ just in case *p* is true in $\langle W, R, I \rangle$ at all worlds w^* such that Rww^*. A sentence *p* is valid in SDL iff it is true at all worlds of all models.

Just as there are different systems of alethic modal logic (see Modal logic), different systems of deontic logic are determined semantically by imposing various requirements on *R*. Consider, for instance, a chief respect in which the parallel between alethic and deontic notions breaks down. 'If it is necessary that *p*, then *p*' is logically valid but, to our daily chagrin, 'If *Op*, then *p*' is not. This difference is reflected in SDL by *not* requiring that *R* be 'reflexive', that is, not requiring that *w* be a deontic alternative for itself (whereas an alethic modal system would require reflexivity on its alethic-accessibility relation). For SDL we require only that *R* is 'serial'; that the set of deontic alternatives for each world is non-empty. It is this requirement that validates $Op \supset \sim O \sim p$. SDL is characterized axiomatically by tautologies *T*, *OT*, $Op \supset \sim O \sim p$ and $O(p \supset q) \supset (Op \supset Oq)$. Rules of proof are *modus ponens* and substitution. SDL can be extended by imposing additional constraints on R, such as 'reflexivity once removed' (which validates $O(Op \supset p)$) and 'transitivity' ($Op \supset OOp$), among many others.

A number of so-called 'paradoxes' in SDL motivated the development of alternative systems. The most important inadequacies of SDL were revealed by thought experiments such as Chisholm's 'contrary to duty imperative' example (1974). Normative systems often specify what should happen in case certain requirements are violated, for example, that one should apologize or pay reparations or not cry (while mopping up). A set of simple rules of this type may say, for instance, that you ought to tell neither Ann nor Bert a certain secret; and yet were you to go ahead and tell either Ann or Bert then you should tell the other as well. This set obviously is consistent. How can it be represented in SDL? If *a* abbreviates 'You tell Ann', and *b*, 'You tell Bert', $O \sim a$ and $O \sim b$ together plausibly represent the primary requirements not to tell either one the secret. What about the conditionals which tell you

what to do if you violate one of these rules? It is here that SDL gets into trouble.

A 'wide scope' representation $O(a \supset b)$ of the conditional 'You ought to tell Bert if you tell Ann' is a poor choice since it follows in SDL from $O{\sim}a$ (whereas intuitively the rules are logically independent). Moreover, there is no way to detach expression of the obligation to tell Bert, after you tell Ann: $O(a \supset b)$ and *a* together do not entail *Ob* in SDL. (The conditional $a \supset b$ may be true in all deontic alternatives and *a* true in the actual world, even though *b* is not true in all the alternatives.) A 'narrow scope' representation $a \supset Ob$ of the rule is no better since it follows from ${\sim}a$ (and yet the rule itself is logically independent of *a*). Moreover, while this option permits detachment of *Ob* given *a* by means of *modus ponens*, *Ob* is not consistent in SDL with $O{\sim}b$ (because of the validity of $Oq \supset {\sim}O{\sim}q$). So detachment generates an inconsistency that is counterintuitive, since the original rules are logically consistent with the fact that *a*. This may suggest rewriting the original unconditional obligations as ${\sim}b \supset O{\sim}a$ and ${\sim}a \supset O{\sim}b$. No inconsistent detachment is possible now, but the suggestion is not viable because the revised rule set would be wholly complied with if you told both Ann and Bert, whereas the rules obviously are violated in some sense if both *a* and *b* are true. Theorists seeking formal solutions to these problems proposed two types of response, corresponding broadly to refinements on the wide- and narrow-scope proposals.

2 Two alternatives to SDL

Von Wright proposed a new type of dyadic (two-place) deontic operator to represent expressions of conditional obligation. $O(q/p)$ is read informally as 'It ought to be that *q* given p'. Danielsson, Hansson, Lewis (1973), van Fraassen, Åqvist and others developed semantics in terms of rankings $\leq$ of worlds according to their relative acceptability (similar to the way in which orderings of worlds are used to interpret subjunctive conditionals). Let $\leq$ be a dyadic relation, on a set of worlds, which is transitive and strongly connected (that is, for any two worlds *u* and *v* in the set, either $u \leq v$ or $v \leq u$). $O(q/p)$ is true relative to $\leq$ iff some p & q-world is ranked more highly in $\leq$ than any p & ${\sim}q$-world. The unconditional *Oq* is defined as $O(q/T)$, for tautology *T*, so that *Oq* holds just in case *q* is true in all most acceptable worlds. The rules in the Ann/Bert example may be represented consistently as $\{O{\sim}a, O{\sim}b, O(a/b), O(b/a)\}$ and, significantly, it would be consistent to add further rules of the form $O({\sim}b/a \;\&\; c)$ and $O(b/a \;\&\; c \;\&\; d)$ and so on. No member of the set is entailed by any other member, and all are independent of factual circumstances. So this language is well suited for representing the content of normative systems with rules and principles that have exceptions or are 'defeasible' like $O{\sim}b$ in our example (there are circumstances, such as when *a* is true, in which $O{\sim}b$ should not be used to guide action). Since typically the rules of a system or practice are used to justify actions or states of affairs, a shortcoming of this approach taken alone is that there is no means even for expressing the unconditional 'all-out' obligations that may hold in non-ideal circumstances. For instance, in circumstances in which the non-ideal *a* is true obviously one ought (in some sense) to tell Bert because $O(b/a)$, but this obligation is not entailed by $O(b/a)$ and *a*. All most acceptable *a*-worlds may be *b*-worlds, and *a* may be true in the actual world, but nothing follows from this because the actual world may not be a most acceptable *a*-world.

A second approach, developed by Montague, Chellas (1984), Thomason (1981), van Eck (1982), Feldman (1986) and others, ties the semantics of *O* to specific factual circumstances by interpreting *O* relative to both worlds *and times*. One way to do this is to index *O* to a time *t*, so that '$O_t q$' is read 'It ought at time *t* to be that *q*' and is interpreted relative to what is 'historically necessary' (or 'settled') at *t* in *w* (see Tense and temporal logic). The set of deontic alternatives at *t* in *w* is the set of most acceptable worlds (according to a given normative system) of those that are compatible with what is settled at *t* in *w*; and $O_t q$ is true at *w* just in case *q* is true at all the deontic alternatives at *t* in *w*. So 'all-out' obligations at *t* are relative to what is settled at *t*. Let the alethic '$N_t q$' say '*q* is settled at *t*' and be true in *w* just in case *q* is true at all worlds historically possible at *t* in *w*. At times *t* when neither *a* nor *b* is settled at *w* (as in our example), $O_t{\sim}a$ and $O_t{\sim}b$ are true. But if $N_{t^*} a$, for a later time t^*, then $O_{t^*} b$.

What about the 'rules' of the system? There is no straightforward way to formalize them on this approach, though it can be done indirectly by quantifying over time: for example, $(t)({\sim}N_t a \supset O_t{\sim}b)$ replaces $O{\sim}b$, and $(t)(N_t a \supset O_t b)$ replaces $O(b/a)$. This approach works for the example and permits detachment by *modus ponens* of appropriate all-out obligation statements. The paradoxes of material implication reappear, however, so a first difficulty is distinguishing 'genuine' from spurious rules. More importantly, the formal representation of 'You should not tell Bert' as a conditional is awkward compared with the English.

3 Combining the two approaches

In any case, the elegant treatment of rules in the first approach is perfectly compatible with the second's temporal treatment of all-out obligation, and combining the two approaches also solves the main problem associated with the first. Given a ranking of worlds, all-out obligations at *t* are determined by restricting the ranking to worlds that are possible at t. If S expresses all and only what is 'settled' at *t* in *w*, then $O_t q$ is true at *w* iff $O(q/S)$ is true according to $\leq$. This means $O_t q$ is true at *w* just in case *q* is true in all most acceptable worlds, according to $\leq$, of those that are still accessible at *t* in *w*. (An agent-relative O_t can be formulated in terms of the best worlds within an agent's control.) To detach $O_t q$ from $O(q/p)$ when *p* does not express all that is settled at *t*, it is useful to express the idea that there is no *r* such that $N_t r$ and $\sim O(q/p \mathbin{\&} r)$. Let $R_t(q,p)$ say as much. This enables us to reason with $O(a/b)$ even if other rules like $\sim O(a/b \mathbin{\&} c)$ also hold. The schema $O(q/p) \mathbin{\&} N_t p \mathbin{\&} R_t(q,p) \supset O_t q$ can be validated in the combined semantics and a family of normative concepts defined precisely in its terms.

There is an 'all-out' (conclusive) reason for *q* at *t* (according to a certain normative system) iff $O_t q$ holds (for that normative system), while *q* is *prima facie* 'obligatory' at *t* iff there is some *p* such that $O(q/p)$ and $N_t p$. $O(q/p)$ is 'undercut' at *t* iff there is some *r* such that $N_t r$ and $\sim O(q/p \mathbin{\&} r)$; and $O(q/p)$ is 'overridden' at *t* iff there is some *r* such that $N_t r$ and $O(\sim q/p \mathbin{\&} r)$. $O(q/p)$ is 'defeated' at *t* iff it is either undercut or overridden. A *prima facie* reason for *q* based on $O(q/p)$ is defeated if $O(q/p)$ is defeated. There is a *prima facie* conflict at *t* iff there are p, q, r and *s* such that $N_t(q \mathbin{\&} s)$, $O(p/q)$, $O(r/s)$ and $N_t{\sim}(p \equiv r)$.

The combined system does not solve all of the problems that SDL encounters. In SDL and in the combined approach, *Op* entails *Oq* if *p* entails *q*. So, for instance, since *h*: 'You help the man who was robbed' entails *r*: 'The man was robbed', $O_t h$ entails $O_t r$, a counterintuitive result (hence the 'Good Samaritan paradox').

Also in the combined system, as in SDL, no all-out conflict (for example, of the form $O_t p \mathbin{\&} O_t q$, where *p* and *q* are contraries) is consistent, but arguably there are coherent normative systems in which genuine conflicts are a possibility, and if so it is not the business of deontic logic to legislate otherwise.

So the combined system, like SDL, is arguably too 'strong' in the sense that it validates too much. It can be weakened semantically by interpreting $O(q/p)$ not in terms of $\leq$, but rather in terms of a 'neighbourhood' selection function *f* that assigns to each proposition (set of worlds) *q* a set of subsets of *W* (the set of propositions that are obligatory given *q*). $O(q/p)$ is true iff $q \in f(p)$. Let $R_t(q,p)$ be true iff for all *r*, if $N_t r$, then $q \in f(p)$ iff $q \in f(p \mathbin{\&} r)$; and $O_t q$ is true iff there is some *p* such that $O(q/p)$, $N_t p$ and $R_t(q,p)$. $O_t p$ does not entail $O_t q$ even if *p* entails *q*, and $O_t p \mathbin{\&} O_t q$ is consistent even for contraries *p* and *q*. (In general, there is an 'all-out conflict' at *t* iff there is a *prima facie* conflict at *t*, as described above, and neither $O(p/q)$ nor $O(r/s)$ is defeated at *t*.) There is so little logical structure here (the only rule of proof is substitution) that it is difficult to envisage complaints grounded in its excessive strength. None the less, it can be used, exactly as above, to define all of the concepts defined earlier. While the weakness of this very minimal system probably is excessive, at least for applications to many normative systems, it is an appropriate starting point for developing stronger systems, and for structuring discussion of the appropriate strength of deontic logics for particular applications.

See also: Logical and mathematical terms, glossary of

MARVIN BELZER

References and further reading

Castañeda, H.-N. (1981) 'The Paradoxes of Deontic Logic: The Simplest Solution to All of Them in one Fell Swoop', in R. Hilpinen (ed.) *New Studies in Deontic Logic*, Dordrecht: Reidel, 37-85. (Develops a distinction between propositions and practitions, and uses this distinction to address the problems of SDL in a unique way.)

Chellas, B.F. (1980) *Modal Logic*, Cambridge: Cambridge University Press, esp. 190-203, 272-7. (Discusses SDL and its extensions, the dependence of obligation on time, and conditional obligation.)

Chisholm, R. (1974) 'Practical Reason and the Logic of Requirement', in S. Korner (ed.) *Practical Reason*, New Haven, CT: Yale University Press, 1-16.(Presents an informal account of many of the definitions in §3.)

Eck, J. van (1982) 'A System of Temporally Relative Modal and Deontic Predicate Logic and its Philosophical Applications', *Logique et Analyse* 25: 249-90, 339-81.(Develops a temporally relative logic as in the 'second

approach' with quantifiers and a subjunctive conditional, and offers alternative definitions of some of the terms defined in §3.)

Feldman, F. (1986) *Doing the Best We Can*, Dordrecht: Reidel.(Application to an act-utilitarian ethics of a temporally relative deontic logic that uses world rankings and defines an agent-relative conception of 'settled'.)

Føllesdal, D. and Hilpinen, R. (1971) 'Deontic Logic: An Introduction', in R. Hilpinen (ed.) *Deontic Logic: Introductory and Systematic Readings*, Dordrecht: Reidel, 1-35.(Good introduction to SDL and its history.)

Humberstone, I. (1983) 'The Background of Circumstances', *Pacific Philosophical Quarterly* 64: 19-34. (Develops a time- and agent-relative all-out ought.)

Lewis, D.K. (1973) *Counterfactuals*, Oxford: Blackwell.(Discusses the defeasibility of expressions of conditional obligation and presents logics for the dyadic operator discussed in §2.)

Lewis, D.K. (1974) 'Semantic Analyses for Dyadic Deontic Logic', in S. Stenlund (ed.) *Logical Theory and Semantic Analysis*, Dordrecht: Reidel, 1-14.(Compares different semantics for the dyadic operator discussed in §2.)

Loewer, B. and Belzer, M. (1983) 'Dyadic Deontic Detachment', *Synthèse* 54: 295-319.(Expands on the 'combined approach' discussed in §3.)

Thomason, R. (1981) 'Deontic Logic as Founded on Tense Logic', in R. Hilpinen (ed.) *New Studies in Deontic Logic*, Dordrecht: Reidel, 165-76.(Discusses the dependence of obligation on time as in the 'second approach'.)

Deontological ethics

Deontology asserts that there are several distinct duties. Certain kinds of act are intrinsically right and other kinds intrinsically wrong. The rightness or wrongness of any particular act is thus not (or not wholly) determined by the goodness or badness of its consequences. Some ways of treating people, such as killing the innocent, are ruled out, even to prevent others doing worse deeds. Many deontologies leave agents considerable scope for developing their own lives in their own way; provided they breach no duty they are free to live as they see fit.

Deontology may not have the theoretical tidiness which many philosophers crave, but has some claim to represent everyday moral thought.

Deontology (the word comes from the Greek *deon* meaning 'one must') typically holds that there are several irreducibly distinct duties, such as promise-keeping and refraining from lying (see Duty; Moral pluralism). Some deontologists, such as W.D. Ross (1930), maintain that one of these duties is a duty to do as much good as possible. Most deny that there is such a duty, while conceding that there is a limited duty of benevolence, a duty to do *something* for the less fortunate (see Help and beneficence). All agree, however, that there are occasions when it would be wrong for us to act in a way that would maximize the good, because we would be in breach of some (other) duty. In this respect they are opposed to act-consequentialism (see Consequentialism §§1, 2).

Most deontologies include two important classes of duties. First, there are duties which stem from the social and personal relationships in which we stand to particular people. Parents have duties to children, and children to parents; people have duties in virtue of their jobs and the associations to which they belong; debtors have a duty to repay their creditors, promisors to keep their promises and borrowers to return what has been lent to them (see Family, ethics and the; Friendship; Professional ethics; Promising; Solidarity). Some of these social relationships are ones we enter voluntarily, but many are not. The second kind take the form of general prohibitions or constraints. We should not lie to, cheat, torture or murder *anyone*, even in the pursuit of good aims (see Truthfulness).

Deontology is often described as an agent-relative moral theory, in contrast to act-consequentialism, which is an agent-neutral theory. According to act-consequentialism the identity of the agent makes no difference to what their duty is on any particular occasion; that is determined solely by which of the courses of action open to them will produce the best consequences. In deontology, by contrast, a reference to the agent often plays an ineliminable role in the specification of the duty. This is especially clear in the case of duties which stem from social relationships. I have a duty to help *this* person. Why? Because he or she is *my* friend, or *my* child. I have a duty to pay *my* debts and to keep *my* promises.

Constraints also involve agent-relativity, though in a slightly different way. The duty not to murder does not take the form of enjoining us to minimize the number of murders. The rule tells me not to commit murder myself even if I could thereby prevent something worse being done, such as two murders being committed. Proponents of deontology think of this as moral integrity; their opponents refer to it disparagingly as 'keeping one's hands clean'.

Many deontologists hold that our duties, though sometimes very onerous, are quite limited in scope. Provided I am in breach of no duty, I am morally at liberty to devote quite a large part of my time and effort to pursuing my own projects in whatever way I please. This latitude leaves room for acts of supererogation: heroic or saintly acts that clearly go beyond the call of duty, and deserve high praise (see Supererogation).

There is a sharp division in the deontologist camp over the status of constraints. Some, such as Fried, think of them as absolute: they have no exceptions and may not be breached in any circumstances which we are likely to encounter. Others regard the fact that an act would breach a constraint as providing a weighty objection to it, but one which could be overcome if there were a sufficiently pressing duty on the other side. Conflicts between two duties which are not absolute must be settled by determining which duty is the more pressing in the circumstances.

Deontology gains much of its appeal from the fact that it seems to capture the essential outlines of our everyday moral thinking, but it is open to several objections. First, its claim that there is a plurality of distinct duties runs counter to the theoretician's search for simplicity. The deontologist will reply, of course, that a theory must do

justice to the complexity of the phenomena. Second, many deontologists further defy the supposed canons of good theorizing by denying that there is any overarching explanation of why there are the duties there are; they record our conviction that there are such duties without seeking to justify them. Others, usually inspired by Kant (1785), do attempt such an explanation based on some broader precept, such as respect for persons (see Kant, I. §§9-11; Respect for persons; Kantian ethics). Third, those who hold that some kinds of action, such as lying, are absolutely prohibited have to provide clear and detailed criteria for determining the boundary between lying and some supposedly less nefarious activity, such as 'being economical with the truth'. Such casuistry can appear both excessively legalistic and incompatible with the spirit of morality (see Casuistry). Fourth, deontology provides no procedure to settle conflicts of duty (though some might think that an advantage). Finally, from a consequentialist perspective, the notion of a constraint seems perverse. If what is wrong with murder is that it is a bad thing, how can it be rational to forbid an agent to commit one murder in order to prevent two? If deontology is to answer this challenge, it must show how it can be that one's duty does not rest (wholly) on the goodness or badness of the results of acting in that way.

See also: Double effect, principle of

DAVID McNAUGHTON

References and further reading

Dancy, J. (1993) *Moral Reasons*, Oxford: Blackwell, chaps 10-12. (A challenging and idiosyncratic defence of a broadly deontological moral theory.)

Darwall, S. (1986) 'Agent-Centred Restrictions from the Inside Out', *Philosophical Studies* 50: 291-319. (A lucid and stimulating attempt to defend the broadly Kantian project of showing that constraints can be justified if we start with the considerations that should weigh with a person of good character.)

Davis, N. (1993) 'Contemporary Deontology', in P. Singer (ed.) *A Companion to Ethics*, Oxford: Blackwell, 205-18. (An accessible discussion of deontologists of an absolutist stripe, such as Fried.)

Fried, C. (1978) *Right and Wrong*, Cambridge, MA: Harvard University Press. (A readable and vigorous defence of an absolutist deontology.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(A classic work of moral philosophy, which defends absolute constraints. It has become fashionable to deny that Kant is a deontologist, but he certainly qualifies as a member of the tradition. Difficult but rewarding.)

McNaughton, D. and Rawling, P. (1991) 'Agent-Relativity and the Doing-Happening Distinction', *Philosophical Studies* 63: 167-85. (An attempt to provide a satisfactory account of the distinction between the agent-neutral and the agent-relative using formal logic, and to address the relationship between this distinction and that between consequentialism and deontology.)

McNaughton, D. (1996) 'An Unconnected Heap of Duties?', *Philosophical Quarterly* 46 (185): 433-47. (Argues that deontology, in the form advocated by Ross, can offer a systematic account of our distinct duties.)

Nagel, T. (1986) *The View from Nowhere*, Oxford: Oxford University Press, ch. 9. (A fascinating attempt to find a place for constraints by appeal to the importance of the personal point of view.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Clarendon Press, ch. 2. (The finest modern systematic exposition of a moderate deontology.)

Depiction

How do pictures work? How are they able to represent what they do? A picture of a goat, for example, is a flat surface covered with marks, yet it depicts a goat, chewing straw, while standing on a hillock. The puzzle of depiction is to understand how the flat marks can do this.

Language poses a similar problem. A written description of a goat will also be a collection of marks on a flat surface, which none the less represent that animal. In the case of language the solution clearly has something to do with the arbitrary way we use those marks. The word 'leg', for example, is applied to legs, but any other mark would do as well, providing we all use it in the same way. In the case of pictures, however, something different seems to be going on. There is not the same freedom in producing a picture of a goat on a hillock chewing straw - the surface must be marked in the right *way, a way we are not free to choose. So what is the right way?*

A helpful thought is that the surface must be marked so as to let us experience it in a special way. With the description, we merely need to know what the words it contains are used to stand for. With the picture, we must instead be able to see *a goat in it. However, although this does seem right, it is difficult to make clear. After all, we do not see a goat in the same way that we see a horse in a view from a window. For one thing, there is no goat there to be seen. For another, it is not even true that looking at the picture is like looking at a goat. It is partly because of the differences that, as we look at the picture, we are always aware that it is merely a collection of marks on a flat surface. So what is this special experience, seeing a goat in the picture? This is the question that a philosophical account of depiction must try to answer.*

1 The question posed

How do pictures represent? Like many other things, they represent in a variety of ways. Consider Picasso's *Guernica*. It pictures a certain scene - horses, mutilated people, a light bulb, an explosion. It expresses the horror of war, and exemplifies that art which is engaged with the moral and political world. In all these respects the picture may be said to represent something, but the forms of representation involved are not the same. Only one of these is distinctively pictorial, a form that may be dubbed 'depiction'. What is this form of representation?

One way to focus this question is to contrast depiction with other ways of representing. The most useful comparison is with words. Words and pictures may perform many of the same tasks, and may represent many of the same things; yet, at least at first glance, they seem fundamentally different. A description of the scene depicted in *Guernica* would differ considerably from the painting, and it is tempting to think that this reflects differences in the forms of representation involved. Perhaps exploring the contrast with linguistic representation will illuminate depiction itself.

2 Resemblance and its problems

Philosophers have offered many different accounts of the nature of depiction; indeed the variety can be bewildering. But there is one answer which, in the light of its venerable age and undeniable appeal, deserves early consideration. This is the thought that pictures represent by *looking like* what they depict. This appears to capture neatly the difference between the ways that words and pictures represent - *Guernica* looks like a scene of mayhem and destruction, whereas a description of such a scene does not. Moreover, the suggestion implies that depiction is a peculiarly visual form of representation - and that idea too has considerable initial appeal.

Unfortunately this view faces many difficulties, some very serious. It is natural to take the claim to be stating that one thing depicts another only if the two resemble each other in respect of some visible property. So understood, its two most important drawbacks are as follows.

First, resemblance makes demands on the world which depiction does not. For resemblance to hold, two things must exist - the thing resembling and the thing resembled. By contrast, depiction does not require there to be two things; one depicting, the other depicted. The picture alone suffices, since it may depict what does not exist. For example, it may depict a horse, but no horse in particular. (Horses exist, but none of those that do is depicted by the picture.) Alternatively, the picture may depict something of a type of which there are no instances, say a 53-sided regular polyhedron. In view of these facts, the resemblance account must do two things. It must find a

way to accommodate depiction where there is no thing depicted; and it must explain how that depiction is related to the depiction of things that do exist. This last condition is necessary, since it is now clearly possible that the account will offer two distinct analyses of depiction.

Second, the intuition that pictures look like what they depict is undermined by the difficulty of saying *in what respect* the two resemble. A picture need not resemble what it depicts in shape - *Guernica* is flat, the scene depicted is not. Nor need the two resemble in colour - a simple line drawing is black and white, but need not depict a black-and-white scene. Nor need picture and depicted match in the materials of which they are made. In fact, for any respect which might provide the resemblance, difference is as common as similarity. It seems, then, that the resemblance view may have to surrender the intuition on which it drew. Unless it can locate the putative resemblance, the thought that there must be one, and that depiction turns upon its presence, cannot be maintained.

3 Goodman’s radical alternative

The most radical reaction to these difficulties would be to abandon not merely the idea of resemblance, but also the strong contrast between depiction and linguistic representation, and the idea that the former is peculiarly visual. The key exponent of such an approach, and indeed of the criticisms of resemblance that motivate it, is Nelson Goodman (§2).

Goodman locates the difference between the ways in which pictures and words represent things in formal features of the representational systems in question. To do this, he develops a set of theoretical tools. Marks on surfaces are grouped as inscriptions of different characters. For example, you are currently looking at marks, each of which is an inscription of a character in the Roman alphabet. Characters are grouped so as to form symbol schemes - the Roman alphabet is one such scheme, while the first eight characters in it form another. If the characters are correlated with a field of reference, the result is a symbol system. Examples include the system for naming the eight notes in a musical octave, and, at a higher level of complexity, written English.

What is special about symbol systems that are pictorial?Goodman identifies three important features. First, such systems are ‘syntactically dense’. That is, they involve infinitely many characters, and are ordered so that between any two characters lies a third. Second, they are ‘semantically dense’. The field of reference of the characters is so ordered that between any two referents there is a third. Finally, they are ‘relatively replete’ - a relatively wide range of properties of the mark determines which character it inscribes. Roughly, this amounts to the following. Pictorial systems are ones in which, for a wide range of properties of the mark on the surface, the tiniest differences in that property affect what is represented. This is not so for linguistic systems - consider, for instance, the English names for the constellations. More generally, Goodman claims that all three features are exhibited by pictorial systems, and that they are not all exhibited by those which represent linguistically. Beyond this simple differentiation claim he is reluctant to tread.

Many of Goodman’s critics have sought to attack his position by counterexample. A graph might be used to track the various properties of a colourless gas over time. With ‘time elapsed’ represented by distance along the x-axis and temperature along the y-axis, thickness of the plotted line might represent the density of the gas, colour might represent its radioactivity, and so on. Such a graph could be a symbol in a system that is both syntactically and semantically dense, and relatively replete. No difference in the graph is in principle too small to affect the precise state of the gas represented. Despite this, the graph would not depict anything.

It has not always been recognized that these examples prove little as they stand, save that Goodman was right not to claim to provide conditions sufficient for depiction. None the less, they may be used to make a stronger case against Goodman since they raise the question *why* it is so plausible that, in this example, the graph does not depict the gas.

One plausible answer is that this representation lacks certain features common to all depiction. For one thing, the graph does not represent the gas from any particular point of view. Pictures, by contrast, always represent what they depict from some perspective - sometimes several. Further, the graph does not represent any properties of the gas that are properties of its visual appearance. The gas has no visual appearance, and is not represented as having one. But, again, depiction is always the representation of things as having an appearance. Finally, although the graph has a comparatively complex content, it might easily not have done so. Had every one of the above features fed in some weighted manner into the representation of the temperature of the gas, the graph would still have met

Goodman's conditions for belonging to a pictorial symbol system, but would have represented nothing more than temperature. Yet pictures always have comparatively detailed contents, even when relatively schematic. This, after all, is the truth of the common saying 'a picture paints a thousand words'.

The force of the example can now be made clear. It not only shows that the account is incomplete, but also indicates how seriously it is so. The graph, at least in its second form, meets Goodman's differentiation conditions without having any of the three features characteristic of depiction. An adequate theory of depiction must be a theory of a form of representation with those features, and should ideally explain why they hold. As it stands Goodman's account plainly does not meet either of these requirements. Moreover, it does not seem that it could readily be developed to meet them. For what might be added to the specification of formal features of the symbol system so as to obtain this result? It seems that the only emendations to suffice would amount to gerrymandering. Thus the only course, if these features are to be accommodated, seems to be to return from Goodman's radicalism towards views which bind depiction more closely to the visual.

4 Our experience of pictures

One such course which has very often been proposed in some form or other is an appeal to the idea that depictions are *experienced* in a special way. Suppose yourself confronted by a picture, but one you do not yet understand at all. You can see quite clearly what colour lies where on its surface. Perhaps you can even see *that* it is a picture of something or other; certainly your experience may give you good grounds to suspect this. But you cannot see what is depicted by the picture. Then, in a moment, you 'get the point'. You can see that the picture is of a horse, that the strange shaped lump that had puzzled you depicts its head, those straggly lines of colour depict its legs, and so forth. It does not seem to you that anything in your environment has altered. You still see the same patches of colour in the same locations. And yet in some sense things now look very different. You see the surface organized in a way you could not see before.

There are two experiences of the picture here, one accompanying the successful interpretation, the other preceding it. The second experience has a distinctive phenomenology, or subjective character, for anyone who enjoys it. It is a way of seeing the coloured surface, but a way somehow involving the thought of what is absent - in this case, a horse. And it is, in some sense, an integrated whole; unlike, for instance, the experience of seeing a castle while visualizing a horse.

Some, most notably Richard Wollheim (1987), have argued that every picture can in principle be experienced in two ways, parallel to those above. Moreover, they have claimed that experiences like the second hold the key to depiction. The idea is that in those experiences one's awareness of the picture's surface involves the thought of something else, and that thing is just what the picture depicts. In the standard terminology, one *sees* in the picture what is depicted. For such philosophers, the key task is to characterize this experience more fully. After all, what has been said thus far fails to distinguish 'seeing-in' from many other experiences, those forming one sort of aspect perception (see Wittgenstein, L.).

Wollheim's own account of the experience of seeing one thing in another is as follows. It is a visual experience of a differentiated surface, and has two 'folds', or aspects. One aspect is in some way analogous to the experience of seeing the surface without seeing anything in it. The other is in some way analogous to the experience of seeing (face-to-face) whatever is seen in the surface. It is not possible to say anything illuminating about the nature of the analogies here involved; indeed to ask for such clarification would be a sign of confusion.

Of those sympathetic to this broad approach, many have thought that more than this can be said. This response is vindicated by the need to offer an account of depiction that accommodates the features mentioned in §3. For unless Wollheim can say more about seeing-in, he cannot hope to offer explanations of those features. Consider, for example, the fact that depiction is always from a point of view. It is natural for the defender of seeing-in to try to account for this by tying what is depicted to what is seen in a surface, and then by establishing that seeing something in a surface itself necessarily involves a point of view on that absent item. But what reason can there be to assert this last claim? The answer must come from the fact that one fold of seeing-in is analogous to an experience, that of seeing the absent item face-to-face, which itself necessarily involves a point of view. However, this is only the sketch for an explanation until that analogy is clarified - what features of the analogy guarantee the presence of perspective in the fold, given its presence in the analogous experience?

There have been other attempts to characterize seeing-in, but the obstacles to success are considerable.Kendall Walton, for instance, makes skilful use of an appealing idea: that seeing-in involves the visual imagination (1990). In essence, his claim is that when I see a horse in a picture, I imagine of my seeing the picture's surface, that it is my seeing a horse. The fundamental difficulty with any such view is that it is ambiguous, and that on either reading it is equally problematic. The ambiguity turns on whether or not the imagining in question involves visualizing. Visualizing has its own distinctive phenomenology. For this reason, it is plausible to claim that any experience with which it is involved will have a distinctive character informed by that phenomenology, but hard to see how that experience can form an integrated whole. On the other hand, imagining that is not visualizing lacks any distinctive phenomenology. It is therefore easy to integrate into a unified experience, but ill-suited to contributing to a distinctive character. Since seeing-in is both phenomenologically distinctive and an integrated whole, the prospects for characterizing it using the imagination are not good (see Imagination).

5 Interpreting depiction

The task of characterizing seeing-in is one of the most challenging in this area. Given this, and scepticism about some of the central notions which any such characterization must deploy, there have been attempts to clarify in other ways the thought that depiction is peculiarly visual. For example, some have attempted to analyse depiction using the epistemic resources required to interpret it. The idea is that what is special about pictures, as opposed to words, is what one needs to know in order to interpret them. For language, one must know the conventions governing the words' use, conventions specific to individual words. Understanding pictures, in contrast, requires a rather different sort of knowledge. Flint Schier, for example, claims that one must know what the depicted item looks like (1986). Moreover, this is in essence *all* one needs to know to interpret pictures, provided that one has a general competence with depiction, that is, the ability to interpret any pictures at all. Since this is so, Schier speculates that a depiction of something engages our ability to recognize that thing, and that it is definitive of depiction to do so.

Schier's view has many advantages. In particular, it finally allows for accommodating and explaining the features of depiction noted in §3. Pictures engage our abilities to recognize what they depict, but those abilities are relative to points of view. I may be able to recognize you seen from the front, while being unable to do so from directly above you. So what individual pictures must engage is the ability to recognize what they depict from a particular point of view - hence the perspectival nature of depiction. Further, visual recognitional abilities are engaged in clusters, rather than singly - I recognize you because I recognize a person with such and such features standing before me. So if a picture is to engage a visual recognitional ability of mine, it will engage others at the same time. Thus there is no depicting something without depicting it as having a certain appearance, and there is no depiction which is not fairly complex in content.

Despite this, there is one form of attack to which Schier's position is vulnerable. For, as he himself admits, it is tempting to think that there is a deeper explanation for why depiction engages our recognitional capacities - that pictures do after all resemble what they depict. If this line can be made good, explanations can not only be provided for the features of §3, but forSchier's view too, along with his observations on the epistemic resources depiction requires. However, the problems facing the resemblance view were severe. Can solutions be found for them?

6 Resemblance revisited

The intuition that pictures look like what they depict was clarified above in the claim that depicting and depicted resemble in respect of some visible property. The discussion of seeing-in provides an opportunity to present the intuition another way. Depiction should be understood by characterizing our experience of it. Thus the better way to understand the intuition is as advocating a certain view of seeing-in: that it is the experience of likeness.

So understood, the view fares somewhat better with the first problem presented in §2. Resemblance may require there to be two things, one resembling, one resembled; the experience of resemblance does not. For instance, I may experience a sound I hear as resembling another, even if that other sound is one I have merely on occasion conjured in my imagination, and even if I am fully aware of this. This opens the way for a single account of depiction, whether or not there exists something depicted. In either case, the picture is experienced in a special way. It is seen in a way somehow involving the thought of the thing depicted (existing or not), and the proposal is

to understand that way of seeing the picture as an experience of resemblance.

However, this still leaves the second problem entirely untouched. In what respect is resemblance experienced? For it is no more plausible to say that a picture need be *experienced as* resembling what it depicts in shape, colour, material composition and the like than it is to say that the two must resemble in those respects.

The answer lies in a property of things that we regularly perceive, but rarely articulate. Consider the thought that in one way a circular disk seen at an oblique angle looks elliptical, yet for all that its 3-D shape is clearly unchanged. What is picked out here is a shape property of an unusual kind. For it is relative to a point - in this case, the point from which the disk is seen. In fact, it is the property of subtending a certain solid angle at that point. For convenience, call this property the ‘outline shape’ the disk has at that point. This is not, however, simply the shape of the object’s silhouette. For if the disk is marked with concentric circles, each of those will also subtend a certain solid angle at the point, and the outline shape of the whole may be taken to include these nested sets of angles too. So understood, it is at least clear that many pictures which do not resemble what they depict in 3-D shape may do so in outline shape.

Seeing-in, then, is essentially the experience of likeness in respect of outline shape. Depiction may then be understood as that representation which works through the deliberate generation of this experience. This leaves many details to work through, and many objections to counter. None the less it is a position with great promise. For it offers a way to perform all the most pressing tasks encountered above - accommodating our intuitions about depiction, explaining various of its features, and characterizing the experience to which it gives rise.

See also: Fictional entities; Photography, aesthetics of

R.D. HOPKINS

References and further reading

Gombrich, E.H. (1977) *Art and Illusion*, 5th edn, Oxford: Phaidon.(An influential and seminal work, deploying immense knowledge of art history, philosophy and psychology.)

Goodman, N. (1969) *Languages of Art*, 2nd edn, Oxford: Oxford University Press. (Difficult, technical, but brilliant treatment of this and several other problems concerning the arts.)

Goodman, N. and Elgin, C.Z. (1988) *Reconceptions in Philosophy*, London: Routledge. (Goodman’s later thoughts on the matter - and some useful replies to criticism.)

Hopkins, R. (1995) ’Explaining Depiction’, *Philosophical Review* 104 (3): 425-55.(For more on the notion of outline shape, and its relevance to depiction; and for some sense of what other ingredients a complete account of the latter requires. Moderately difficult.)

Peacocke, C. (1987) ‘Depiction’, *Philosophical Review* 96: 383-410.(A difficult attempt to put forward something very close to a resemblance view, and the source of the original gas example of §3.)

Schier, F. (1986) *Deeper Into Pictures*, Cambridge: Cambridge University Press.(A fascinating full-length exploration of pictures, although sometimes moderately difficult.)

Walton, K. (1990) *Mimesis as Make-Believe*, London: Harvard University Press.(A highly influential and readable account, rooted in a grand theory of representation.)

Wollheim, R. (1977) ‘Representation: The Philosophical Contribution to Psychology’ *Critical Inquiry* 3 (4): 709-23.(A useful and non-technical introduction to the range of views available - although it has its axe to grind.)

(1987) *Painting As An Art*, London: Thames & Hudson.(An impressive and elegant exploration of depiction and many issues more directly related to the appreciation of pictorial art.)

Derrida, Jacques (1930-)

Jacques Derrida is a prolific French philosopher born in Algeria. His work can be understood in terms of his argument that it is necessary to interrogate the Western philosophical tradition from the standpoint of 'deconstruction'. As an attempt to approach that which remains unthought in this tradition, deconstruction is concerned with the category of the 'wholly other'.

Derrida has called into question the 'metaphysics of presence', a valuing of truth as self-identical immediacy which has been sustained by traditional attempts to demonstrate the ontological priority and superiority of speech over writing. Arguing that the distinction between speech and writing can be sustained only by way of a violent exclusion of otherness, Derrida has attempted to develop a radically different conception of language, one that would begin from the irreducibility of difference to identity and that would issue in a correspondingly different conception of ethical and political responsibility.

1 Historical context and influences

To identify all of the philosophical problematics at stake in Derrida's writings, it would probably be necessary to have some familiarity with the works of Kant, Hegel, Marx, Nietzsche, Husserl, Heidegger, Levinas, Saussure, Lévi-Strauss, Austin, Foucault, Freud, Lacan, Nicolas Abraham and Maria Torok. Lacking such familiarity, readers of Derrida are often not in a position to recognize - let alone assess - the philosophical moves he makes. Ill-equipped commentators have ascribed to Derrida theses that have nothing to do with his actual philosophical concerns. Some of the confusion can be traced to the borrowing of Derrida's term 'deconstruction' by the literary critic Paul De Man and others for whom the term has come to refer to a style of reading texts. For Derrida, deconstruction is less a reading technique than it is a way of approaching the 'wholly other' and of attempting thereby to conjure 'an experience of the impossible'.

Derrida's work can be situated within the broad constellation of post-Hegelian efforts to consider whether a rigorous critique of the language of metaphysics can free philosophical thinking from its regularly repeated dogmatic gestures. The theme of avoiding metaphysical dogmatism is one that could be traced back to Kant and the British empiricists, and even to Descartes in certain respects. But criticizing metaphysical *dogmatism* and criticizing metaphysical *language* are not necessarily the same enterprise, and it is only in the post-Hegelian world that the project of interrogating language becomes a steady theme for Western philosophers. Defined in this way, the constellation of figures to which Derrida belongs would include members of the pragmatist and logical positivist movements as much as it would those thinkers associated with continental philosophy. But it is primarily out of the continental tradition that Derrida's own philosophical concerns have developed.

In his earliest writings, Derrida is drawn to the Husserlian phenomenological project as a rigorous attempt to develop an account of signification (see Husserl, E.). Noting, however, that Husserl retains the philosophical ideal of univocal language, Derrida argues that Husserl's phenomenological method remains motivated by a traditional metaphysical appeal to the value of truth as self-presence. To some extent, Heidegger had already criticized Husserlian phenomenology on similar grounds. Appreciative of Heidegger's arguments but also hesitant about some of his conclusions, Derrida takes Heidegger's attempt to define a 'task for thinking' at the 'end of philosophy' to pose a number of crucial questions to which he has repeatedly returned in his own work.

In his later writings, Heidegger critiques the claim that language is something which 'belongs to man', substituting for this the attempt to show how 'man belongs to language'. Were he simply advancing a proposition thereby, Heidegger would seem to be saying that thought does not determine language; rather, language determines thought. But it would be a mistake to read Heidegger as making an empirical claim of the Whorfian variety. If the beings that comprise the world manifest themselves as such only by way of language, then language is that which opens up the world *as such* - that is, as a sphere of particular beings which can be studied empirically. Such a claim would be analogous to the Kantian thesis that the basic categories of human thought are not mere empirical concepts which we get from our observation of objects; rather, they constitute the very objectivity of the world. Kant's thesis is transcendental. For Heidegger, the fact that different languages constitute the world differently for different speakers renders problematic the Kantian appeal to universal concepts. If thinking is always limited by a finite language which it cannot master, it follows that thinking can at most raise regional problems about the ways

in which it is so determined.

2 Deconstruction

It is out of his reflections on the implications of this position that Derrida comes to clarify the task of deconstruction. Derrida first proposed the French *déconstruction* as a way of translating Heidegger's use of the German word *Destruktion*, a term which appears in *Being and Time* (1927). According to Heidegger, the task of thinking time as the horizon of Being requires a *Destruktion* of the history of metaphysical concepts, especially the concept of time as developed in Aristotle, Descartes, and Kant. Heidegger does not equate *Destruktion* with obliteration. For Heidegger, the history of metaphysics consists in the tightening of a structure - what he later calls the *Gestell* - and it is his aim to 'loosen' the hold that this structure has on thinking. Considered in this way, *Destruktion* would be a freeing-up: a de-structuring. It is partly in order to capture this particular sense of the Heideggerian term that Derrida invokes the word *déconstruction*.

But Derrida takes his work of translation to be one which also transforms the Heideggerian conception of 'freeing-up', for Derrida sees Heidegger as recapitulating certain gestures of the metaphysics of presence. Derrida is especially suspicious of those places where Heidegger seems to suggest that the *Destruktion* of the history of metaphysics can bring thought back to some sort of 'original' relation to language as a poetic 'Saying' of 'the Being of beings'. For Derrida, there can be no 'original' position from which thinking begins or to which it could return.

For Derrida, the history of Western metaphysics consists in a series of repeated efforts to affirm self-presence as the paradigm of truth. At the heart of this tradition is the definition of 'man' as that being who can signal his self-presence to himself through language. Against this pretension of 'man's' unique position among beings - and in crucial respects his position here is closer to Hegel's than is usually acknowledged - Derrida argues that no thought, not even that of an 'I think', can ever be immediately present to itself (see Selfhood, postmodern critique of). To understand this position, we need to see why Derrida is drawn away from the theme of language to that of writing.

In order to maintain the value of self-presence, Western philosophers have traditionally tended to ignore the degree to which thinking is dependent on language. Yet the fact of such dependence could not be dismissed entirely. Accordingly, philosophers from Plato to Rousseau to Hegel distinguish between an ideal language in which thought would be immediately transparent to itself, and a secondary language into which this original and univocal 'language of thought' could then be translated. *Spoken* words - because they exist only in the disappearing moment in which they are spoken, and because they can be heard 'in the head' of the speaker - would seem to be directly expressive of thought. By contrast, *written* words, because they can function even in the absence of their producer, would seem to be exterior to thought. Speech immediately embodies thought; writing is merely a sign of speech: such would be the founding claim of philosophy's pretension to self-presence through a forgetting of the materiality of signification. On the basis of this flimsy argument, Western philosophers writing alphabetically have insisted that their own 'phonetic' writing is more intelligible - that is, a sign of greater intelligence - than the 'non-phonetic' writing of non-Europeans - this despite the contravening fact that, as in Leibniz, the ideal of a formal language has necessarily found its exemplar in the model of non-phonetic writing.

Obviously, there are all kinds of empirical reasons why it should seem impossible to say what counts as phonetic writing and what counts as non-phonetic writing. More significantly, Derrida argues, it is impossible to make sense of the ideal of a purely expressive language. For in its dependence on iterable signifiers, all language requires that very non-expressive element which has traditionally been ascribed to non-phonetic writing. Accordingly, there is simply no basis for drawing a rigorous distinction between 'speech' and 'writing'.

In this way, Derrida purports to deconstruct the 'logocentric' or 'phonocentric' basis of the metaphysics of presence. The hierarchical privileging of speech over writing turns out to be but one of an indefinite series of hierarchical oppositions upon which traditional metaphysics founds itself. In each case, that which would function as something 'other' to the pretension to self-presence is excluded by being accorded the same 'fallen' status that non-phonetic writing is said to have. Derrida's aim is not to 'reverse' these hierarchical oppositions - as it would be if he were interested in privileging writing over speech - but to deconstruct the very logic of such exclusionary founding gestures.

On Derrida's account, then, the untenability of the speech/writing distinction makes the concept of writing just as suspect as the concept of speech. On the other hand, writing continues to function as an exemplary figure of otherness. For this reason, it is tempting to say that, much as Heidegger suggests that the task of thinking is 'to bring language as language to language', so deconstruction tries 'to bring writing as writing to writing'. To deconstruct the metaphysics of presence would be to disclose thinking *as* writing - rather than writing as a way of recording pure thoughts. Derrida stresses this point by describing 'arche-writing' - the archival character of a thought without *arkhe*, or origin - as *différance*, a term he introduces to name the impossibility of naming a 'first' or 'central' term of any sort. As a way of opening up the possibility for thinking otherwise than metaphysically, it is necessary to free up writing from its metaphysical interpretation as language intended to express a self-present 'meaning' or 'truth'.

Toward this end, Derrida has developed a number of different writing strategies. One strategy is to produce irreducibly multiple 'texts' which resist being read as unified 'books'. Derrida's *Glas* (1974; the word can be translated as 'death knell') is fragmented in this way. Its basic structure consists of two separate 'columns', one of which examines a number of issues related to Hegel's conception of a crypt, and the other of which juxtaposes to this first column a multi-faceted examination of the signature of Jean Genet as a kind of crypt that is regularly inscribed in Genet's writings. Among other things, *Glas* can be read as a deconstruction that would 'free up' the writings of Genet from the quasi-Hegelian 'summing-up' of Genet which Sartre presented in his *Saint Genet* (1952). Another strategy of Derrida's is to produce textual marks - such as *différance* - that are specifically designed to render problematic the traditional concept of a signifier. Such marks are not quite 'neologisms', since they are neither words nor signifiers, but deconstructive 'surds' which call attention to their own resistance to intelligibility.

3 Early reception and subsequent clarifications

The thematics of *différance* have given rise to a number of misunderstandings about Derrida's work, and some of his writings of the 1980s and 1990s are attempts to specify the task of deconstruction more precisely. The fact that *différance* indicates the un-sayability of a 'final word' about something like 'Being' has been read by some as resonating with the language of negative theology. Derrida shows why the apparent affinity is unavoidable but warns that to equate deconstruction with negative theology would be to reaffirm the eschatological orientation of the metaphysics of presence. Instead, Derrida has argued that every experience is structured by an undeconstructible messianic promise which interrupts the presence of the here-and-now without being reducible to the prophecy of a determinate future. Perhaps the Heraclitean conception of 'becoming' comes closest to indicating what Derrida is after with his writing of *différance* - provided that we read Heraclitus' affirmation of becoming as compatible with Derrida's account of the experience of the promise.

Such a reading of Heraclitus can perhaps be found in Nietzsche's writings, the deconstructive aspects of which Derrida takes Heidegger to overlook. According to Heidegger, Nietzsche's texts represent the single thought of a single thinker at a singular moment in the history of Being. Derrida counters this suggestion by calling attention to aspects of Nietzsche's writing styles. Rather than express a meaning or truth, Nietzsche's writings introduce textual problematics meant to rule out recourse to the traditional conceptual resources of interpretation. Derrida demonstrates this, in *Éperons: les styles de Nietzsche (Spurs: Nietzsche's Styles)*, by showing the difficulties that would arise for any interpretive decision concerning a scrap of paper on which Nietzsche has scribbled (in German) 'I have forgotten my umbrella.' At issue is not only the question of whether a text can have a single meaning, but whether it makes sense to speak of any 'thing' as having a single identity. Rather than speak of *Nietzsche*'s texts, Derrida writes of those texts that are signed with the *signature* - or, as it will turn out, signatures - of 'Nietzsche'. The problematic of the signature thereby becomes an important issue which arises in many of Derrida's texts, and the question of his 'own' signature features prominently in these considerations. Derrida is especially attentive to what he calls the 'iterability' of signatures - their ability, as marks, to function in the absence of any determinable addressee. On Derrida's analysis, iterability (which is the condition for the possibility of a written mark's ability to function as a signifier) disrupts the concept of a fixed context that would determine a set of rules governing signification.

Although he has seemed, to some critics, merely to duplicate the writing strategies of Nietzsche, Derrida expresses some reservations about certain of Nietzsche's gestures. For example, while he notes that the term 'woman'

functions in Nietzsche's writings as a trope for non-truth - and thus as an exemplary textual 'site' for deconstruction - he also calls attention to the violence of such gestures. Derrida seeks a strategy that would exploit the subversive potential of marginal subject positions without reifying the logic of marginality. In this way, deconstruction becomes a way of thinking about the political. From his earliest writings Derrida has been concerned with the relationship between justice and violence. Taking off from Heidegger's essay on the Anaximander Fragment (1946), Derrida engages Levinas and others on the question of whether or not it is possible to thematize a purely non-violent conception of justice.

What exactly a more fully developed deconstructive politics might look like is a topic that has received a great deal of attention in recent years. The fact that *différance* implies an endless deferral of what, from the standpoint of metaphysics, would be the 'meaning' or 'truth' of beings (or of Being), has led some critics to suggest that deconstruction undermines the very possibility of making ethical and political determinations. Such interpretations overlook the constant concern with justice that informs Derrida's works. Working both with and against the Kantian conception of regulative ideals, Derrida seeks to delineate a conception of justice as an *aporia* which both calls for, and is called forth by, the work of deconstruction.

Despite evidence to the contrary, Habermas, Gadamer and their followers insist that deconstruction implies an affirmation of indeterminacy and thus the complete inability to make political judgments. These critics argue for a politics based on an ideal of consensus, and they take to task philosophers such as Derrida who do not share their consensus that consensus is a goal that no one needs to be coerced to accept. By calling nearly every position substantively different from his own a 'performative contradiction', Habermas exhibits the very sort of normative violence that deconstruction questions.

Derrida has countered such criticisms by more explicitly addressing ethical and political issues. Toward this end, he has drawn on a wide variety of traditional texts, among them Aristotle and de Montesquieu on friendship; Aristotle on the difference between economics and chrematistics; Marx on commodity fetishism; Kant on international politics, and Benjamin on the messianic. A constant concern of Derrida's has been that of developing a rigorously deconstructive conception of responsibility, a theme he often develops with reference to the work of Levinas.

Derrida's contributions to social theory cannot be fully understood, I would argue, without some sense of his response to Nicolas Abraham and Maria Torok's recasting of psychoanalytic problematics along lines that connect Husserlian and Freudian conceptions of subjectivity. Abraham and Torok's reworking of such psychoanalytic concepts as incorporation, introjection, mourning and haunting, are frequently invoked in Derrida's writings. In *Spectres de Marx (Specters of Marx)*, Derrida develops a conception of 'hauntology' which uses many of these terms to work through questions concerning the texts of Marx and the political programmes associated with Marxism. Making explicit a complex problematic of 'the spectral' in Marx's texts, Derrida juxtaposes the force of Marx's account of the spectrality involved in commodity fetishism with Marx's resistance to the concept of spectrality. By way of considering Marx's conflicted relationship with Stirner, Derrida calls for a return to a Marxist thinking that would not banish the spectral dimension which Stirner opens up. In so doing, Derrida implicitly suggests that Marx's attempt to banish Stirnerian concerns parallels the attempt of contemporary social theorists to 'conjure away' the spectre of deconstruction.

See also: Deconstruction; Postmodernism; Post-structuralism; Tel Quel school

ANDREW CUTROFELLO

List of works

Derrida, J. (1962) 'Introduction', in E. Husserl, *L'Origine de la géometrie*, trans. J. Derrida; trans. D. Allison, *Edmund Husserl's 'Origin of Geometry': An Introduction*, Pittsburgh, PA: Duquesne University Press, 1978. (Critique of the value of univocity in Husserl's account of ideality.)

Derrida, J. (1967a) *La Voix et le phénomène*, trans. D. Allison, *Speech and Phenomena and Other Essays on Husserl's Theory of Signs*, Evanston, IL: Northwestern University Press, 1973. (Argues that the indicative, or written, element of language cannot be reduced in Husserl's account of meaning.)

Derrida, J. (1967b) *De la grammatologie*, trans. G. Spivak, *Of Grammatology*, Chicago, IL: University of Chicago Press, 1974. (Shows that the history of philosophy has been governed by a privileging of speech over writing.)

Derrida, J. (1967c) *L'Écriture et la différence*, trans. A. Bass, *Writing and Difference*, Chicago, IL: University of Chicago Press, 1978. (Includes the important essay, 'De l'économie restreinte à l'économie générale: un hégélianisme sans réserve' (From Restricted to General Economy: A Hegelianism Without Reserve) as well as essays on Foucault and Levinas.)

Derrida, J. (1972) *Marges - de la Philosophie*, Paris: Éditions de Minuit; trans. A. Bass, *Margins of Philosophy*, Chicago, IL: University of Chicago Press, 1982. (Includes 'Le Puits et la pyramide: introduction à la sémiologie de Hegel' (The Pit and the Pyramid: Introduction to Hegel's Semiology) in which he discusses Hegel's theory of signs, and essays on Heidegger and other philosophers.)

Derrida, J. (1974) *Glas*, Paris: Éditions Galilée; trans. J. Leavey and R. Rand, Glas, Lincoln, NE: University of Nebraska Press, 1986.(Difficult but richly textured readings of Hegel and Genet.)

Derrida, J. (1978) *Éperons: les styles de Nietzsche*; trans. B. Harlow, *Spurs: Nietzsche's Styles*, Chicago, IL: University of Chicago Press, 1979. (A short, and in some ways exemplary, deconstructive reading of Nietzsche.)

Derrida, J. (1987) *De l'esprit: Heidegger et la question*; trans. G. Bennington and R. Bowlby, *Of Spirit: Heidegger and the Question*, Chicago, IL: University of Chicago Press, 1989. (Examines the use of the German term *Geist* and its derivatives in Heidegger.)

Derrida, J. (1992) 'Force of Law: The "Mystical Foundation of Authority"', in D. Cornell, M. Rosenfeld and D.G. Carlson (eds) *Deconstruction and the Possibility of Justice*, New York: Routledge. (Widely cited essay in which Derrida distinguishes between justice and law.)

Derrida, J. (1993) *Spectres de Marx*; trans. P. Kamuf, *Specters of Marx*, Chicago, IL: University of Chicago Press, 1994. (Juxtaposes readings of *Hamlet*, Heidegger and Marx as a way of conceiving the relationship between justice, mourning, responsibility and spectrality.)

References and further reading

Abraham, N. and Torok, M. (1994) *The Shell and the Kernel*, vol. 1, trans. and ed. N. Rand, Chicago, IL: University of Chicago Press. (An important collection of essays that have influenced Derrida's attempt to think together certain phenomenological and psychoanalytic themes.)

Cornell, D. (1992) *The Philosophy of the Limit*, New York: Routledge. (An original interpretation of the philosophical implications of deconstruction, emphasizing its relevance to ethics and social theory.)

Gasché, R. (1994) *Inventions of Difference: On Jacques Derrida*, Cambridge, MA: Harvard University Press. (Essays which examine specific aspects of Derrida's philosophical problematics in detail.)

Harvey, I. (1986) *Derrida and the Economy of Différance*, Bloomington, IN: Indiana University Press. (A good account of the relationship between deconstruction and continental philosophy beginning with Kant.)

Heidegger, M. (1927) *Sein und Zeit*, trans. J. Macquarrie and E. Robinson, *Being and Time*, New York: Harper & Row, 1962. (Heidegger's magnum opus, an important touchstone for all his later work.)

(1946) 'Der Spruch des Anaximander', trans. D. Farrell Krell and F. Capuzzi, 'The Anaximander Fragment, in *Early Greek Thinking*, New York: Harper & Row, 1975. (Philological and theoretical reflection on the use of the Greek term *dike* (justice), in the fragment of a pre-Socratic philosopher.)

Jameson, F. (1995) 'Marx's Purloined Letter', *New Left Review* 209. (Review essay of Derrida's *Specters of Marx*.)

Kofman, S. (1984) *Lectures de Derrida*, Paris: Éditions Galilée. (Discusses a number of important themes in Derrida's writings.)

Laclau, E. (1991) *New Reflections on the Revolution of our Time*, London: Verso. (An investigation of deconstruction as articulating a logic of the political.)

Mallet, M. (ed.) (1994) *Le Passage des frontières: autour du travail de Jacques Derrida*, Paris: Éditions Galilée. (A collection of essays presented by various authors at a conference focusing on the work of Derrida.)

Sartre, J.-P. (1952) *Saint Genet*, trans. B. Fechtman, New York: Braziller.

Spivak, G. (1993) *Outside in the Teaching Machine*, New York: Routledge. (A unique philosophical statement, but one influenced by the works of Derrida.)

Descartes, René (1596-1650)

René Descartes, often called the father of modern philosophy, attempted to break with the philosophical traditions of his day and start philosophy anew. Rejecting the Aristotelian philosophy of the schools, the authority of tradition and the authority of the senses, he built a philosophical system that included a method of inquiry, a metaphysics, a mechanistic physics and biology, and an account of human psychology intended to ground an ethics. Descartes was also important as one of the founders of the new analytic geometry, which combines geometry and algebra, and whose certainty provided a kind of model for the rest of his philosophy.

After an education in the scholastic and humanistic traditions, Descartes' earliest work was mostly in mathematics and mathematical physics, in which his most important achievements were his analytical geometry and his discovery of the law of refraction in optics. In this early period he also wrote his unfinished treatise on method, the Rules for the Direction of the Mind, which set out a procedure for investigating nature, based on the reduction of complex problems to simpler ones solvable by direct intuition. From these intuitively established foundations, Descartes tried to show how one could then attain the solution of the problems originally posed.

Descartes abandoned these methodological studies by 1628 or 1629, turning first to metaphysics, and soon afterwards to an orderly exposition of his physics and biology in The World. But this work was overtly Copernican in its cosmology, and when Galileo was condemned in 1633, Descartes withdrew The World from publication; it appeared only after his death.

Descartes' mature philosophy began to appear in 1637 with the publication of a single volume containing the Geometry, Dioptrics and Meteors, three essays in which he presented some of his most notable scientific results, preceded by the Discourse on the Method, a semi-autobiographical introduction that outlined his approach to philosophy and the full system into which the specific results fit. In the years following, he published a series of writings in which he set out his system in a more orderly way, beginning with its metaphysical foundations in the Meditations (1641), adding his physics in the Principles of Philosophy (1644), and offering a sketch of the psychology and moral philosophy in the Passions of the Soul (1649).

In our youth, Descartes held, we acquire many prejudices which interfere with the proper use of our reason. Consequently, later we must reject everything we believe and start anew. Hence the Meditations begins with a series of arguments intended to cast doubt upon everything formerly believed, and culminating in the hypothesis of an all-deceiving evil genius, a device to keep former beliefs from returning. The rebuilding of the world begins with the discovery of the self through the 'Cogito Argument' ('I am thinking, therefore I exist') - a self known only as a thinking thing, and known independently of the senses. Within this thinking self, Descartes discovers an idea of God, an idea of something so perfect that it could not have been caused in us by anything with less perfection than God Himself. From this he concluded that God must exist which, in turn, guarantees that reason can be trusted. Since we are made in such a way that we cannot help holding certain beliefs (the so-called 'clear and distinct' perceptions), God would be a deceiver, and thus imperfect, if such beliefs were wrong; any mistakes must be due to our own misuse of reason. This is Descartes' famous epistemological principle of clear and distinct perception. This central argument in Descartes' philosophy, however, is threatened with circularity - the Cartesian Circle - since the arguments that establish the trustworthiness of reason (the Cogito Argument and the argument for the existence of God) themselves seem to depend on the trustworthiness of reason.

Also central to Descartes' metaphysics was the distinction between mind and body. Since the clear and distinct ideas of mind and body are entirely separate, God can create them apart from one another. Therefore, they are distinct substances. The mind is a substance whose essence is thought alone, and hence exists entirely outside geometric categories, including place. Body is a substance whose essence is extension alone, a geometric object without even sensory qualities like colour or taste, which exist only in the perceiving mind. We know that such bodies exist as the causes of sensation: God has given us a great propensity to believe that our sensations come to us from external bodies, and no means to correct that propensity; hence, he would be a deceiver if we were mistaken. But Descartes also held that the mind and body are closely united with one another; sensation and other feelings, such as hunger and pain, arise from this union. Sensations cannot inform us about the real nature of things, but they can be reliable as sources of knowledge useful to maintaining the mind and body unity. While many of Descartes' contemporaries found it difficult to understand how mind and body can relate to one another,

Descartes took it as a simple fact of experience that they do. His account of the passions is an account of how this connection leads us to feelings like wonder, love, hatred, desire, joy and sadness, from which all other passions derive. Understanding these passions helps us to control them, which was a central aim of morality for Descartes.

Descartes' account of body as extended substance led to a physics as well. Because to be extended is to be a body, there can be no empty space. Furthermore, since all body is of the same nature, all differences between bodies are to be explained in terms of the size, shape and motion of their component parts, and in terms of the laws of motion that they obey. Descartes attempted to derive these laws from the way in which God, in his constancy, conserves the world at every moment. In these mechanistic terms, Descartes attempted to explain a wide variety of features of the world, from the formation of planetary systems out of an initial chaos, to magnetism, to the vital functions of animals, which he considered to be mere machines.

Descartes never finished working out his ambitious programme in full detail. Though he published the metaphysics and the general portion of his physics, the physical explanation of specific phenomena, especially biological, remained unfinished, as did his moral theory. Despite this, however, Descartes' programme had an enormous influence on the philosophy that followed, both within the substantial group that identified themselves as his followers, and outside.

1 Life

René Descartes was born on 31 March 1596 in the Touraine region of France, in the town of La Haye, later renamed Descartes in his honour. In 1606 or 1607 he was sent to the Collège Royal de La Flèche, run by the Jesuit order. Here he received an education that combined elements of earlier Aristotelian scholasticism with the new humanistic emphasis on the study of language and literature. But the core of the collegiate curriculum was the study of Aristotelian logic, metaphysics, physics and ethics. Descartes left La Flèche in 1614 or 1615, and went to the University of Poitiers, where he received his *baccalauréat* and his *licence en droit* in late 1616. In Part I of the *Discours de la méthode (Discourse on the Method)* (1637), he discusses his education in some detail, explaining why he found it increasingly unsatisfactory. In the end, he reports, he left school, rejecting much of what he had been taught there. He chose the life of the military engineer, and set out across Europe to learn his trade, following the armies and the wars. On 10 November 1618, in the course of his travels, he met Isaac Beeckman. An enthusiastic scientific amateur since his early twenties, Beeckman introduced Descartes to some of the new currents in science, the newly revived atomist ideas, and the attempt to combine mathematics and physics (see Atomism, ancient). Despite the fact that they only spent a few months together, Beeckman put Descartes on the path that led to his life's work. A number of discussions between them are preserved in Beeckman's extensive notebooks (1604-34), which still survive, and include problems Beeckman set for Descartes, as well as Descartes' solutions. It was for Beeckman that Descartes wrote his first surviving work, the *Compendium musicum*, a tract on music theory, then considered a branch of what was called mixed mathematics, along with other disciplines such as mathematical astronomy and geometric optics. Exactly a year after first meeting Beeckman, this new path was confirmed for Descartes in a series of three dreams that he interpreted as a call to settle down to his work as a mathematician and philosopher.

During the 1620s, Descartes worked on a number of projects including optics and the mathematics that was eventually to become his analytic geometry. In optics, he discovered the law of refraction - the mathematical law that relates the angle of incidence of a ray of light on a refractive medium, with the angle of refraction. Though some claim that Descartes learned this law from Willebrod Snel, after whom the law is now named, it is generally thought that Descartes discovered it independently. In his mathematical programme, he showed how algebra could be used to solve geometric problems, and how geometric constructions could be used to solve algebraic problems.

Descartes' most extensive writing from this period is the *Regulae ad directionem ingenii (Rules for the Direction of the Mind)*, a treatise on method that he worked on between 1619 or 1620 and 1628, when he abandoned it incomplete. He continued to travel extensively throughout Europe, returning to Paris in 1625, where he was to stay until spring 1629. In Paris, Descartes became closely associated with Marin Mersenne who later became a central figure in the dissemination of the new philosophy and science in Europe, the organizer of a kind of scientific academy and the centre of a circle of correspondents, as well as Descartes' intellectual patron (see Mersenne, M.). Through his voluminous correspondence with Mersenne, Descartes remained connected to all the intellectual currents in Europe, wherever he was to live in later years. An important event in this period took place at a

gathering at the home of the Papal Nuncio in Paris in 1627 or 1628, where Descartes, responding to an alchemical lecture by one M. Chandoux, took the occasion to present his own ideas, including his ‘fine rule, or natural method’ and the principles on which his own philosophy was based (letter to Villebressieu, summer 1631; Descartes 1984-91 vol 3: 32). This attracted the attention of the Cardinal Bérule, who in a private meeting, urged Descartes to develop his philosophy.

In spring 1629 Descartes left Paris and moved to the Low Countries, where he set his methodological writing aside and began his philosophy in earnest. The winter of 1629-30 was largely occupied with the composition of a metaphysical treatise, which, as we shall later see, represents the foundations of his philosophy. The treatise is now lost, but Descartes told Mersenne that it had tried to ‘prove the existence of God and of our souls when they are separated from the body, from which follows their immortality’ (letter to Mersenne, 25 November 1630; Descartes 1984-91 vol 3: 29). This was followed by the drafting of *Le Monde (The World)*, Descartes’ mechanist physics and physiology, a book intended for publication. In the first part, also called the *Traité de la lumière (Treatise on Light)*, Descartes begins with a general account of the distinction between a sensation and the motion of tiny particles of different sizes and shapes that is its cause, followed by an account of the foundations of the laws of nature. After then positing an initial chaos of particles in motion (not *our* cosmos, but one made by God in some unused corner of the world), Descartes argues that by means of the laws of nature alone, this cosmos will sort itself into planetary systems, central suns around which swirl vortices of subtle matter which carry planets with them. He concludes the *Traité de la lumière* with an account of important terrestrial phenomena, including gravity, the tides and light, showing how much like our cosmos this imaginary mechanist cosmos will appear. The second part, the *Traité de l’homme (Treatise on Man)*, begins abruptly by positing that God made a body that looks exactly like ours, but which is merely a machine. Presumably missing - or never written - is a transition between the two treatises that shows how by the laws of nature alone this human body could arise in our world. (This part of the argument is noted in Part V of the later *Discourse on the Method*.) In the text that we now have, Descartes then went on to argue that all phenomena that pertain to life (thought aside) can arise in this body in a purely mechanical way, including nutrition and digestion, the circulation of blood, the movement of the muscles and the transmission of sensory information to the brain.

By 1633 Descartes had in hand a relatively complete version of his philosophy, from method, to metaphysics, to physics and biology. But in late 1633, he heard of the condemnation of Galileo’s Copernicanism in Rome, and cautiously decided not to publish his *World*, which was evidently Copernican (see Galileo Galilei). Indeed, he first decided never to publish anything at all. But the despair did not last. Between 1634 and 1636, Descartes collected some of the material he had been working on, and prepared three essays for publication, the *Géométrie*, the *Météors* and the *Dioptrique*. These scientific essays were preceded by a general introduction, the *Discours de la méthode pour bien conduir sa raison et chercher la vérité dans les sciences (Discourse on the Method for Properly Conducting Reason and Searching for Truth in the Sciences)*. The *Discourse* presents itself as autobiography, an account of the path the young author (the book was published anonymously) followed in his discoveries, including a summary of his method (Part II), of his early metaphysical speculations (Part IV) and of the programme of *The World* (Part V). In the scientific essays, Descartes presented some of his most striking results, hiding the foundational elements (such as his apparent Copernicanism and his rejection of scholastic form and matter) that would be most controversial.

While not uncontroversial, the *Discourse* and *Essays* were very successful, and induced Descartes to continue his programme for publishing his philosophy. The next work to appear was the *Meditationes (Meditations)* of 1641, which included an extensive selection of objections to the *Meditations* from various scholars in learned Europe, including Hobbes, Gassendi, Arnauld, and Mersenne himself, along with Descartes’ responses, a total of seven sets in all (these are cited in this entry as the *First Objections*, *Second Replies*, and so on). This was followed in 1644 by the publication of the *Principia Philosophiae (Principles of Philosophy)* in which, after a review of his metaphysics, Descartes gives an exposition of his physics adapted and expanded from *The World*. French translations of the *Meditations* and the *Principles* done by others, but with important variants from the original Latin (presumably introduced by Descartes himself), appeared in 1647.

By the late 1630s, Descartes’ work had entered the Dutch universities, and was taught at the University of Utrecht by Henricus Reneri and, following him, by Henricus Regius. Descartes’ un-Aristotelian views called down the wrath of Gisbertus Voëtius, who started a pamphlet war against Descartes and Regius that raged for some time.

Descartes supported Regius, and gave him advice as to how to respond and contain the affair. Eventually, however, Descartes broke with him when Regius wrote and in 1646 published his *Fundamenta physices*, about which Descartes had severe reservations. Regius responded with a broadsheet, a kind of summary of his main theses, emphasizing their differences. Descartes, in turn, responded in 1648 with the *Notae in programma quoddam (Comments on a Certain Broadsheet)*. There was a similar incident in Leiden, where Descartes had disciples (François du Ban, Adriaan Heereboord) as well as an influential enemy (Revius).

In the late 1640s Descartes was working on drafting and publishing more of his philosophy. Two additional parts of the *Principles* were planned, extending the work to cover elements of human biology. While notes remain in the form of an incomplete treatise on the human body (*La description du corps humain* - *Description of the Human Body*) and on the foetus (*Prima cogitationes circa generationem animalium* - *First Thoughts on the Generation of Animals*), the larger work was never finished. There are also important works concerning morals and moral psychology dating from these years. Some of this material is found in the letters to the Princess Elisabeth of Bohemia, with whom he had a long and important correspondence, starting in 1643. Descartes' account of the passions is found in the last work he published in his lifetime, the short *Passions de l'âme (Passions of the Soul)*, which appeared in 1649.

With the exception of a few short trips to Paris in 1644, 1647 and 1648, Descartes remained in the Low Countries until October 1649, when he was lured to Stockholm to be a member of the court of Queen Christina. There he fell ill in early 1650, and died on 11 February of that year.

2 The programme

Descartes' thought developed and changed over the years. But even so, there are a number of threads that run through it. Like most of his lettered contemporaries, Descartes was educated in a scholastic tradition that attempted to combine Christian doctrine with the philosophy of Aristotle. Indeed, at La Flèche, where he first learned philosophy, Aristotle as interpreted by Aquinas was at the centre of the curriculum. What he learned was an interconnected system of philosophy, including logic, physics, cosmology, metaphysics, morals and theology.

On his own account, Descartes rejected the Aristotelian philosophy as soon as he left school. From the notes Beeckman took on their conversations, it is probable that what dissatisfied him most in what he had been taught was natural philosophy. For an Aristotelian, the understanding of the natural world was grounded in a conception of body as composed of matter and form. Matter was that which remained constant even during the generation and corruption of bodies of different kinds, and that which all bodies of all sorts have in common; form was that which was responsible for the characteristic properties of particular sorts of bodies. For example, form was to explain why earth falls and tends to be cold, and why fire rises and tends to be hot. In contrast, though he came to reject Beeckman's rather strict atomism, Descartes seems to have been attracted to the kind of mechanistic view of the world that his mentor espoused. Descartes held from then on that the manifest properties of bodies must be explained in terms of the size, shape and motion of the tiny parts that make them up, and rejected the appeal to innate tendencies to behaviour that lay at the foundation of the Aristotelian view (see Aristotle §10).

But even though he rejected much of the philosophy of the schools, there was one element that remained with him: like his teachers at La Flèche, Descartes always held that knowledge has a kind of systematic coherence. In Rule 1 of the *Rules* Descartes wrote that 'everything is so interconnected that it is far easier to learn all things together than it is to separate one from the others… All [sciences are] connected with one another and depend upon one another'. Later, when he read Galileo's *Two New Sciences* (1638), Descartes dismissed the Italian scientist because his work lacked that kind of coherence (letter to Mersenne, 11 October 1638; Descartes 1984-91 vol 3: 124-8). His own project was to build his own interconnected system of knowledge, a system comprising an account of knowledge, a metaphysics, a physics and other sciences. This ambition is summarized in one of his last writings, the Preface to the French edition of the *Principles*, where he wrote that 'all philosophy is like a tree, whose roots are metaphysics, whose trunk is physics, and whose branches, which grow from this trunk, are all of the other sciences, which reduce to three principle sciences, namely medicine, mechanics, and morals'. In this way, Descartes saw himself as reconstituting the Aristotelian-Christian synthesis of the scholastics, grounded not in a natural philosophy of matter and form, but in a mechanist conception of body, where everything is to be explained in terms of size, shape and motion.

Certain important features of the Cartesian programme are worth special mention. The Aristotelian-Christian synthesis is founded in a variety of kinds of authority: the authority of the senses, the authority of ancient texts and the authority of his teachers. Descartes wanted to ground his thought in himself alone, and in the reason that God gave him. Since, Descartes claimed, reason gives us genuine certainty, this means that true knowledge is certain. In Rule 3 of the *Rules* he wrote that 'concerning things proposed, one ought to seek not what others have thought, nor what we conjecture, but what we can clearly and evidently intuit or deduce with certainty; for in no other way is knowledge acquired'. The rejection of the authority of the senses, texts and teachers shaped Descartes' thought in fundamental ways. Because of it, his philosophical system began with the Cogito Argument, which establishes the self as the starting-place of knowledge. Moreover, his two most influential works, the *Discourse* and the *Meditations*, were written in the first person so as to *show* the reader how Descartes did or might have come to his own state of knowledge and certainty, rather than *telling* readers what they are to believe, and thus setting himself up as an authority in his own right. Despite his rejection of authority, however, Descartes always claimed to submit himself to the authority of the Church on doctrinal matters, separating the domain of revealed theology from that of philosophy.

Another important feature of Descartes' tree of knowledge was its hierarchical organization. Throughout his career he held firmly to the notion that the interconnected body of knowledge that he sought to build has a particular order. Knowledge, for Descartes, begins in metaphysics, and metaphysics begins with the self. From the self we arrive at God, and from God we arrive at the full knowledge of mind and body. This, in turn, grounds the knowledge of physics, in which the general truths of physics (the nature of body as extension, the denial of the vacuum, the laws of nature) ground more particular truths about the physical world. Physics, in turn, grounds the applied sciences of medicine (the science of the human body), mechanics (the science of machines) and morals (the science of the embodied mind).

3 Method

Before beginning an account of the individual parts of Descartes' tree of knowledge, it is necessary to discuss his method. Method was the focus of his earliest philosophical writing, the *Rules*, and appeared prominently in his first published writing, the *Discourse on the Method*. But what exactly that method was is somewhat obscure.

In the second part of the *Discourse*, the method is presented as having four rules: (1) 'never to accept anything as true if I did not have evident knowledge of its truth: that is, carefully to avoid precipitate conclusions and preconceptions'; (2) 'to divide each of the difficulties I examined into as many parts as possible'; (3) 'to direct my thoughts in an orderly manner, by beginning with the simplest and most easily known objects in order to ascend little by little… to knowledge of the most complex'; and (4) 'throughout to make enumerations so complete and reviews so comprehensive, that I could be sure of leaving nothing out'. Given the general nature and apparent obviousness of these rules, it is not surprising that many of Descartes' contemporaries suspected him of hiding his *real* method from the public.

But Descartes' account of the method in the *Rules* is somewhat fuller. In Rule 5 he says: 'We shall be following this method exactly if we first reduce complicated and obscure propositions step by step to simpler ones, and then, starting with the intuition of the simplest ones of all, try to ascend through the same steps to a knowledge of all the rest'. This method is illustrated with an example in Rule 8. There Descartes considers the problem of the anaclastic line, the shape of a lens which will focus parallel lines to a single point. The first step in the solution of the problem, Descartes claims, is to see that 'the determination of this line depends on the ratio of the angles of refraction to the angles of incidence'. This, in turn, depends on 'the changes in these angles brought about by differences in the media'. But 'these changes depend on the manner in which a ray passes through the entire transparent body, and knowledge of this process presupposes also a knowledge of the nature of the action of light'. Finally, Descartes claims that this last knowledge rests on our knowledge of 'what a natural power in general is'. This last question can, presumably, be answered by intuition alone, that is, a purely rational apprehension of the truth of a proposition that has absolute certainty. Once we know what the nature of a natural power is, we can, Descartes thought, answer one by one all the other questions raised, and eventually answer the question originally posed, and determine the shape of the lens with the required properties. These successive answers are to be connected deductively (in a way outlined in the *Rules*) with the first intuition, so that successive answers follow intuitively from the first intuition.

The example of the anaclastic line suggests that Descartes' method proceeds as follows. One starts with a particular question, q_1. The reductive moment in the method then proceeds by asking what we have to know in order to answer the question originally posed. This leads us from q_1 to another question, q_2, whose answer is presupposed in order for us to be able to answer q_1; it is in this sense that q_1 is said to be reduced to q_2. This reductive process continues until we reach a question whose answer we are capable of knowing through intuition, say q_n. At that point, we begin what might be called the constructive moment of the method, and successively answer the questions we have posed for ourselves, using the answer to q_n to answer q_{n-1}, the answer to q_{n-1} to answer q_{n-2}, and so on until we arrive at q_1, the question originally posed, and answer that.

Understood in this way, the method has some very interesting properties. First, it results in knowledge that is completely certain. When we follow this method, the answer to the question originally posed is grounded in an intuition; the answers to the successive questions in the series are to be answered by deducing propositions from propositions that have been intuitively grasped as well. Second, the method imposes a certain structure on knowledge. As we follow the series of questions that constitute the reductive step of the method, we proceed from more specific questions to more general, from the shape of a particular lens to the law of refraction, and ultimately to the nature of a natural power. The answers that are provided in the constructive stage follow the opposite path, from the metaphysically more general and more basic to the more specific.

The *Rules* was written over a long period of time, and there are numerous strata of composition evident in the work as it survives. In a passage from Rule 8 that is probably in one of the last strata of composition, Descartes raises a problem for the method itself to confront, indeed the *first* problem that it should confront: 'The most useful inquiry we can make at this stage is to ask: What is human knowledge and what is its scope?… This is a task which everyone with the slightest love of truth ought to undertake at least once in life, since the true instruments of knowledge and the entire method are involved in the investigation of the problem'. While it is not entirely clear what Descartes had in mind here, it is not implausible to interpret him as raising the problem of the justification of intuition itself, the epistemological foundation of the method. In framing the method in the *Rules*, Descartes takes for granted that he has a faculty, intuition, by which he is capable of grasping truth in some immediate way, and what he knows by intuition is worthy of trust. But why *should* we trust intuition? This, in essence, is one of the central questions in the *Meditations*, where Descartes argues that whatever we perceive clearly and distinctly is true.

Method was a central concern of Descartes' earlier writings, in both the *Rules* and the *Discourse*. In later writings it seemed to play little explicit role in his thought, but the hierarchical structure of knowledge with which the method is closely connected - the idea that knowledge is grounded in a structure of successively more metaphysically basic truths, ultimately terminating in an intuition - remained basic to his thought. In his later work, that ultimate intuition is not the nature of a natural power, as it was in the anaclastic line example, but the intuition that establishes the existence of the knowing subject, the Cogito Argument.

4 Doubt and the quest for certainty

In the *Rules for the Direction of the Mind*, Rule 2 reads: 'We should attend only to those objects of which our minds seem capable of having certain and indubitable cognition'. While, as we shall later see, Descartes seemed to relax this demand somewhat in his later writings, the demand for certainty was prominent throughout many of his writings. Historically, this can be seen as a reaction against important sceptical currents in Renaissance thought, the so-called 'Pyrrhonist Crisis'. In the face of the rapidly expanding boundaries of the European world in the sixteenth century, from new texts to new scientific discoveries to the discoveries of new worlds, contradictions and tensions in the intellectual world abounded, making it more and more attractive to hold, with the classical sceptics, that real knowledge is simply beyond the ability of human beings to acquire (see Scepticism, Renaissance; Pyrrhonism). Against this, Descartes asserted that real, certain knowledge is possible; though his name is associated with scepticism, it is as an opponent and not an advocate.

But though certainty was central to Descartes, the path to certainty begins with doubt. In Meditation I, entitled 'What can be called into doubt', Descartes says that 'I realised that it was necessary, once in the course of my life, to demolish everything completely and start again right from the foundations if I wanted to establish anything at all in the sciences that was stable and likely to last'. Following that, he presents a series of three sceptical arguments designed to eliminate his current beliefs in preparation for replacing them with certainties. The strategy is to

undermine the beliefs, not one by one but by undermining ‘the basic principles’ on which they rest. While at least some of these arguments can be found in versions in the *Discourse* and in other writings by Descartes, they receive their fullest exposition in the *Meditations*.

The first argument is directed at the naïve belief that everything learned via the senses is worthy of belief. Against this Descartes points out that ‘from time to time I have found that the senses deceive, and it is prudent never to trust completely those who have deceived us even once’. The second, the famous dream argument, is directed against the somewhat less naïve view that the senses are at least worthy of belief when dealing with middle-sized objects in our immediate vicinity: ‘A brilliant piece of reasoning! As if I were not a man who sleeps at night… I plainly see that there are never any sure signs by means of which being awake can be distinguished from being asleep’. But even if I doubt the reliability of what the senses seem to be conveying to me right now, Descartes supposes, the dream argument still leaves open the possibility that there are some general truths, not directly dependent on my present sensations, that I can know. Descartes replies to this with his deceiving God argument.

This complex argument has two horns. Descartes first supposes ‘the long-standing opinion that there is an omnipotent God who made me the kind of creature that I am’. Because God is omnipotent, he might have made me in such a way that I go wrong in even the simplest and most evident beliefs that I have - for example, that $2 + 3 = 5$, or that a square has four sides. Though God is thought to be good, the possibility that I am so deeply prone to error seems as consistent with his goodness as the fact that I go wrong even occasionally, at least at this stage of the investigation. But what if there is no God, or what if I arose by ‘fate or chance or a continuous chain of events, or by some other means’? In this case, Descartes argues, the less powerful my original cause, ‘the more likely it is that I am so imperfect as to be deceived all the time’.

With this, the sceptical arguments of Meditation I are complete: ‘I am finally compelled to admit that there is not one of my former beliefs about which a doubt may not properly be raised’. But, Descartes notes, ‘my habitual opinions keep coming back’. It is for that reason that Descartes posits his famous evil genius: ‘I will suppose therefore that not God, who is supremely good and the source of truth, but rather some evil genius of the utmost power and cunning has employed all his energies in order to deceive me’. The evil genius (sometimes translated as the ‘evil demon’) is introduced here not as a separate argument for doubt, but as a device to help prevent the return of the former beliefs called into doubt.

These arguments have a crucial function in Descartes’ project. As he notes in the introductory synopsis of the *Meditations*, these arguments ‘free us from all our preconceived opinions, and provide the easiest route by which the mind may be led away from the senses’. In this way the sceptical doubt of Meditation I prepares the mind for the certainty to which Descartes aspired. But in the *Third Replies*, responding to criticisms from Hobbes, Descartes notes two other roles that the sceptical arguments play in his thought. Descartes remarks that they are introduced ‘so that I could reply to them in the subsequent Meditations’. As considered below, the deceiving God argument is answered in Meditations III and IV, and the dream argument is answered in the course of his discussion of sensation in Meditation VI. (Since Descartes, quite rightly, continued to maintain that sensation is not entirely trustworthy as a guide to how things really are, the first sceptical argument is never fully answered, though in Meditation VI he carefully sets out the conditions under which we can trust the senses.) Finally, he notes that the arguments are there as a kind of standard against which he can measure the certainty of his later conclusions: ‘I wanted to show the firmness of the truths which I propounded later on, in the light of the fact that they cannot be shaken by these metaphysical doubts’. In all these ways, Descartes presented himself as addressing the sceptic, and defending a kind of dogmatic philosophy.

5 The Cogito Argument

Descartes’ philosophy begins in doubt. The first step towards certainty, the Archimedean point from which the whole structure will grow, is the discovery of the existence of the self. At the beginning of Meditation II, reflecting on the evil genius posited at the end of Meditation I, Descartes observes: ‘Let him deceive me as much as he can, he will never bring it about that I am nothing so long as I think that I am something… I must finally conclude that this proposition, *I am*, *I exist*, is necessarily true whenever it is put forward by me or conceived in my mind’. In the earlier *Discourse* (Part IV) and the later *Principles of Philosophy* (Part I §7), this proposition has the more familiar form, ‘I am thinking, therefore I exist,’ or, ‘ego cogito, ergo sum,’ in its Latin formulation. Here, it is called the Cogito Argument.

There is considerable discussion about how exactly Descartes thought this argument functions. There are two strains of interpretation that derive directly from his texts. In the *Second Replies*, Descartes observes that 'when we become aware that we are thinking things, this is a primary notion which is not derived by means of any syllogism'. This suggests that the Cogito Argument is known immediately by direct intuition. In the *Principles* (Part I §10), however, Descartes notes that before knowing the Cogito, we must grasp not only the concepts of thought, existence and certainty, but also the proposition that 'it is impossible that that which thinks should not exist'. This suggests that the Cogito *is* a kind of syllogism, in which I infer my existence from the fact that I am thinking, and with the premise that whatever thinks must exist. Recent analytic philosophers have also been attracted to the Cogito, trying to understand its obvious allure through speech act theory and theories of demonstratives (Hintikka 1967; Markie 1992). These accounts, however, are distant from anything that Descartes himself conceived.

There is also some confusion about what the conclusion of the Cogito Argument is supposed to be. In the body of Meditation II, Descartes clearly establishes the existence of the self as a thinking thing or a mind. But the title of Meditation II, 'The nature of the human mind, and how it is better known than the body' suggests that Descartes believed that he had established that the nature of the human mind is thought. Further still, in parallel texts in the *Discourse* (Part IV) and the *Principles* (Part I §7-8), Descartes suggests that the Cogito establishes the existence of a thinking substance distinct from the body, though in the text of Meditation II, this seems to be denied.

Though most closely associated with Descartes, the Cogito Argument may not be altogether original. A number of Descartes' contemporaries, both during his life and afterwards, noticed the connection between the Cogito and similar formulations in Augustine (see Augustine §2). However, what was important to Descartes about the Cogito is the foundational role it plays in his system. For Descartes, it is 'the first thing we come to know when we philosophise in an orderly way' (*Principles*: Part I §7). Common sense might think that the physical world of bodies, known through sensation, is more accessible to us than is the mind, a thinking thing whose existence is established, even though we have rejected the senses. But, as Descartes argues in Meditation II using the example of a piece of wax, despite our prejudices, bodies are not conceived through the senses or the imagination but through the same process of purely intellectual conception that gives us the conception of ourselves as thinking things. Furthermore, knowledge of the external world is less certain than knowledge of the mind, since whatever thought could lead us to a probable belief in the existence of bodies will lead us to believe in the existence of the self with certainty.

The project, then, is to build the entire world from the thinking self. It is important here that it is not just the mind that is the foundation, but *my* mind. In this way, the starting place of philosophy for Descartes was connected with the rejection of authority that is central to the Cartesian philosophy. In beginning with the Cogito, we build a philosophy detached from history and tradition.

6 God

The next stage in the system, as outlined in the *Meditations*, seeks to establish that God exists. In his writings, Descartes made use of three principal arguments. The first (at least in the order of presentation in the *Meditations*) is a causal argument. While its fullest statement is in Meditation III, it is also found in the *Discourse* (Part IV) and in the *Principles* (Part I §§ 17-18). The argument begins by examining the thoughts contained in the mind, distinguishing between the formal reality of an idea and its objective reality. The formal reality of any thing is just its actual existence and the degree of its perfection; the formal reality of an idea is thus its actual existence and degree of perfection as a mode of mind. The objective reality of an idea is the degree of perfection it has, considered now with respect to its content. (This conception extends naturally to the formal and objective reality of a painting, a description or any other representation.) In this connection, Descartes recognized three fundamental degrees of perfection connected with the capacity a thing has for independent existence, a hierarchy implicit in the argument of Meditation III and made explicit in the *Third Replies* (in response to Hobbes). The highest degree is that of an infinite substance (God), which depends on nothing; the next degree is that of a finite substance (an individual body or mind), which depends on God alone; the lowest is that of a mode (a property of a substance), which depends on the substance for its existence. Descartes claims that 'it is manifest by the natural light that there must be at least as much reality in the efficient and total cause as in the effect of that cause'. From that he infers that there must be as much *formal* reality in the cause of an idea as there is *objective* reality in the idea itself.

This is a bridge principle that allows Descartes to infer the existence of causes from the nature of the particular ideas that are in the mind, and thus are effects of some causes or another. In Meditation III, Descartes discusses various classes of ideas, one by one, and concludes that, as a finite substance, he can conceivably be the cause of all the ideas he has in his mind except for one: the idea of God. Since the idea of God is an idea of something that has infinite perfection, the only thing that can cause that idea in my mind is a thing that formally (actually) has the perfection that my idea has objectively - that is, God himself.

Descartes used two other arguments for the existence of God in his writings. In Meditation III, following the causal argument, he offers a version of the cosmological argument for those who, still blinded by the senses, may be reluctant to accept the bridge principle that his causal argument requires. (Versions of this argument are also found in *Discourse* Part IV, and in *Principles* Part I §§20-1.) This argument begins with the author's own existence, as established in Meditation II. But, the author might ask, what could have created me? It will not do, Descartes argues, to suggest that I have been in existence always, and thus I do not need a creator, since it takes as much power to sustain me from moment to moment as it does to create me anew. I could not have created myself because then I would have been able to give myself all the perfections that I so evidently lack. Furthermore, if I could create myself, then I could also sustain myself, which I do not have the power to do; being a thinking thing, if I had such a power, I would be aware of having it. My parents cannot be my creators, properly speaking, since they have neither the ability to create a thinking thing (which is all I know myself to be at this stage of the *Meditations*), nor to sustain it once created. Finally, I could not have been created by another creature of lesser perfection than God, since I have an idea of God, an idea I could not acquire from a lesser being. (Here one suspects that this cosmological argument really collapses into the first causal argument.) From this Descartes concludes 'that the mere fact that I exist and have within me an idea of a most perfect being… provides a very clear proof that God indeed exists'.

These first two arguments for the existence of God play a central role in the validation of reason, as discussed below. But after reason has been validated on theological grounds, Descartes presents in Meditation V a version of the ontological argument (see God, arguments for the existence of §§2-3). After reflecting on the basis of geometric reasoning, the fact that 'everything which I clearly and distinctly perceive to belong to that thing really does belong to it', Descartes concludes that this applies to the idea of God as well. Hence he concludes that 'it is quite evident that existence can no more be separated from the essence of God than the fact that its three angles equal to two right angles can be separated from the essence of a triangle, or than the idea of a mountain can be separated from the idea of a valley'. Though apparently circular in so far as its validity seems to depend on the prior arguments for the existence of God, it is not; Descartes' point is that 'even if it turned out that not everything on which I have meditated in these past days is true, I ought still to regard the existence of God as having at least the same level of certainty as I have hitherto attributed to the truths of mathematics'. As with the other two arguments, Descartes' ontological argument is also found in the *Discourse* (Part IV) and in the *Principles* (Part I §§14-16); indeed, in the *Principles* it is the first argument he gives.

As noted above, the existence of God plays a major role in the validation of reason. But it also plays a major role in two other parts of Descartes' system. As we shall later see in connection with Descartes' physics, God is the first cause of motion, and the sustainer of motion in the world. Furthermore, because of the way he sustains motion, God constitutes the ground of the laws of motion. Finally, Descartes held that God is the creator of the so-called eternal truths. In a series of letters in 1630, Descartes enunciated the view that 'the mathematical truths which you call eternal have been laid down by God and depend on Him entirely no less than the rest of His creatures' (letter to Mersenne, 15 April 1630; Descartes 1984-91 vol 3: 23), a view that Descartes seems to have held into his mature years. While it never again gets the prominence it had in 1630, it is clearly present both in correspondence (for example, letter to Arnauld, 29 July 1648; Descartes 1984-91 vol 3: 358-9) and in published writings (for example, in the *Sixth Responses*).

Various commentators have proposed that Descartes was really an atheist, and that he includes the arguments for the existence of God as window dressing. While this is not impossible, the frequent appeal to God in philosophical contexts, both in private letters and in published work, suggests that it is rather unlikely.

7 The validation of reason

With the existence of God established, the next stage in Descartes' programme is the validation of reason. At the

beginning of Meditation III, before proving God's existence, Descartes notes that the uncertainty that remains is due only to the fact that the meditator does not know whether or not there is a God and, if there is, if he can be a deceiver. This suggests that all one must do to restore reason and defeat the third and most general sceptical argument presented in Meditation I is to prove that there is a benevolent God. And at the end of Meditation III, after two proofs for the existence of God, Descartes concludes directly that this God 'cannot be a deceiver, since it is manifest by the natural light that all fraud and deception depend on some defect'. But this is not enough. In the course of the deceiving-God argument of Meditation I, Descartes notes that if some deception is consistent with divine benevolence, then total deception would be as well. Since it is undeniable that we do make mistakes from time to time, and are thus deceived, this raises a problem for Descartes: what, if anything, does God's benevolence and veracity guarantee?

Descartes answers this question by way of an account of error in Meditation IV. Roughly speaking, the mistakes I make are due to myself and my (improper) exercise of my free will, while the truths I come to know are because of the way God made me. More exactly, Descartes asserts that judgments depend on two faculties of the mind, 'namely, on the faculty of knowledge [or intellect] which is in me, and on the faculty of choice or freedom of the will'. A judgment is made when the will assents to an idea that is in the intellect. But the intellect is finite and limited in the sense that it does not have ideas of all possible things. On the other hand, the will is indefinite in its extent, Descartes claims: 'It is only the will or freedom of choice which I experience within me to be so great that the idea of any greater faculty is beyond my grasp'. It is in our free will that we most resemble God. In certain circumstances, Descartes held, 'a great light in the intellect is followed by a great inclination in the will', and in this way the intellect determines the will to assent. This, he thought, is a proper use of the will in judgment. In this situation, where the intellect determines the will to assent, Descartes talks of our having a clear and distinct perception of a truth. In this case, God has made us in such a way that we cannot but assent. (Clear and distinct perceptions are very close to what he calls 'intuitions' in the *Rules*, as discussed above.) But because the will has a greater extent than the intellect, and is not restrained by it, sometimes things outside the intellect move the will to assent. This is where error enters: 'The scope of the will is wider than that of the intellect; but instead of restricting it within the same limits, I extend its use to matters which I do not understand. Since the will is indifferent in such cases, it easily turns aside from what is true and good, and this is the source of my error and sin.' In this way, I am responsible for error by extending my will beyond where it belongs. God can in no way be held accountable for my mistakes any more than he can be responsible for my sins. I cannot reproach my maker for not having given me more ideas in the intellect than I have, nor can I fault him for having made me more perfect by giving me a free will. But as a result of a limited intellect and a free will, it is possible for me both to make mistakes and to sin.

As a result of this analysis of error, Descartes is able, in Meditation IV to assert his famous principle of clear and distinct perception, an epistemological principle to replace the principles that were rejected as a result of the sceptical arguments of Meditation I: 'If I simply refrain from making a judgement in cases where I do not perceive the truth with sufficient clarity and distinctness, then it is clear that I am behaving correctly and avoiding error. But if in such cases I either affirm or deny, then I am not using my free will correctly'. With this, reason is validated, and the deceiving-God argument answered. Yet, this does not end Descartes' engagement with the sceptical arguments of Meditation I, and in Meditation VI he also addresses the question of the reliability of the senses, presents a limited validation of sensory knowledge, and answers the dream argument.

The validation of reason, central as it is to Descartes' project in the *Meditations*, has one apparent flaw: it seems to be circular. The validation of reason in Meditation IV depends on the proof of the existence of God which, in turn, depends on the proof of the existence of the self as a thinking thing. But evidently we must assume that clear and distinct perceptions are trustworthy in order to trust the Cogito and the proofs for the existence of God that ground the validation of reason - the so-called 'Cartesian Circle'. Two of the objectors to the *Meditations* noticed this point, and elicited responses from Descartes, in the *Second* and the *Fourth Replies*. Descartes' answer is not altogether clear. In the *Second Replies* he remarks, in answer to one such objection, that 'when I said that we can know nothing for certain until we are aware that God exists, I expressly declared that I was speaking only of knowledge of those conclusions' deduced by long arguments, and not of 'first principles', such as the Cogito Argument. This suggests that Descartes would exempt immediately intuitable (self-evident) propositions from the scope of the doubt of Meditation I, and use them as tools for establishing the premises of the argument that leads to

the validation of reason in Meditation IV. There are serious problems with this approach. For one, it seems arbitrary to exempt self-evident propositions from the scope of doubt. Such propositions would seem to fall quite naturally among those most obvious of things that Descartes calls into doubt there; if God could create me in such a way that I go wrong when I add two and three, he could create me in such a way that I go wrong with any other self-evident belief. Furthermore, even if those propositions that are immediately evident are outside the scope of doubt, Descartes' proofs for the existence of God, necessary premises of his validation of reason, are not self-evident. These apparent problems might be either weaknesses in Descartes' response, or reasons to doubt that we have understood Descartes correctly in his responses.

The problem of circularity and the obvious problems in Descartes' apparent answer have elicited numerous examinations of the issue in the commentary literature. It is not clear just what Descartes' own solution was, nor whether or not there is a good response to the Cartesian Circle. But whatever the answer, the problem is not a superficial oversight on Descartes' part. It is a deep philosophical problem that will arise in some form or another whenever one attempts a rational defence of reason.

8 Mind and body

One of Descartes' most celebrated positions is the distinction between the mind and the body. Descartes did not invent the position. It can be found in various forms in a number of earlier thinkers. It is a standard feature of Platonism and, in a different form, is common to most earlier Christian philosophers, who generally held that some feature of the human being - its mind or its soul - survives the death of the body (see Plato §13). But the particular features of Descartes' way of drawing the distinction and the arguments that he used were very influential on later thinkers.

There are suggestions, particularly in the *Discourse* (Part IV) and in the *Principles* (Part I §§7-8) that the distinction between mind and body follows directly from the Cogito Argument, as discussed above. However, in the *Meditations* Descartes is quite clear that the distinction is to be established on other grounds. In Meditation VI he argues as follows: I have a clear and distinct idea of myself as a thinking non-extended thing, and a clear and distinct idea of body as an extended and non-thinking thing. Whatever I can conceive clearly and distinctly, God can so create. So, Descartes argues, the mind, a thinking thing, can exist apart from its extended body. And therefore, the mind is a substance distinct from the body, a substance whose essence is thought.

Implicit in this argument is a certain conception of what it means to be a substance, a view made explicit in the *Principles* (Part I §51) which defines a substance as 'nothing other than a thing which exists in such a way as to depend on no other thing for its existence', no other thing but God, of course. In so far as the mind can exist independently of the body, it is a substance on this definition. (God is the third kind of substance, along with mind and body, though because of his absolute independence, he is a substance in a somewhat different sense.) On Descartes' metaphysics, each substance has a principal attribute, an attribute that characterizes its nature. For mind it is thought, and for body it is extension. In addition, substances have modes, literally ways of instantiating the attributes. So, for Descartes, particular ideas, particular volitions, particular passions are modes of mind, and particular shapes, sizes and motions are modes of body.

Descartes' conception of mind and body represents significant departures from the conceptions of both notions in the late scholastic thought in which he was educated. For the late scholastics, working in the Aristotelian tradition, body is composed of matter and form. Matter is that which remains constant in change, while form is that which gives bodies the characteristic properties that they have. For Descartes, however, all body is of the same kind, a substance that contains only geometric properties, the objects of geometry made concrete. The characteristic properties of particular forms of body are explained in terms of the size, shape and motion of its insensible parts (see §11 below). For the late scholastics, the mind is connected with the account of life. On the Aristotelian view, the soul is the principle of life, that which distinguishes a living thing from a dead thing; it is also taken to be the form that pertains to the living body. The mind is the rational part of the soul, that which characterizes humans, and not usually considered a genuine substance, though by most accounts, with divine aid, it can survive the death of the body (see Nous; Psychē). For Descartes, the majority of the vital functions are explained in terms of the physical organization of the organic body. The mind, thus, is not a principle of life but a principle of thought. It involves reason, as does the rational soul of the Aristotelians, but it also involves other varieties of thought, which pertain to other parts of the Aristotelian soul (see Aristotle §17). Furthermore, it is a genuine substance, and

survives the death of the body naturally and not through special divine intervention.

Mind and body are distinct because they can exist apart from one another. However, in this life, they do not. In Meditation VI Descartes observed: 'Nature also teaches me, by these sensations of pain, hunger, thirst and so on that I am not merely present in my body as a sailor is present in a ship, but that I am very closely and, as it were, intermingled with it, so that I and the body form a unit'. He sometimes went so far as to say that the human mind is the form of the human body, the only kind of form that he recognizes in nature, and that the human being - the union of a mind and a body - constitutes a genuine substance, though the context of these statements suggests that they may be made more for orthodoxy's sake than an expression of his own views (see, for example, the letter to Regius, January 1642 (Descartes 1984-91 vol 3: 208), where he is advising Regius on the best way to answer the attacks made by the more orthodox Aristotelian Gisbertus Voëtius in connection with the controversy at Utrecht.) But be that as it may, he was clearly committed to holding that the mind and the body are united. Some of his contemporaries found it difficult to understand how two such different substances could interact and be joined. Sometimes Descartes dismissed this objection by saying that it is no more difficult than understanding how form and matter unite for the Aristotelians, something that everyone learns at school (letter to Arnauld, 29 July 1648; Descartes 1984-91 vol 3: 358). But elsewhere, particularly in an important exchange of letters with the Princess Elisabeth of Bohemia, Descartes offered a different explanation, remarking that it is simply an empirical fact that they do unite and interact, something that we learn from everyday experience, and suggesting that just as we have innate notions that allow us to understand the notions of thinking and extended substance, we also have an innate notion that allows us to comprehend how mind and body interact, and how together they can constitute a unity (letters to Elisabeth, 21 May 1643 and 28 June 1643; Descartes 1984-91 vol 3: 217-20, 226-9).

According to Descartes, the mind is joined to the body in one specific place: the pineal gland, a single gland in the centre of the brain, between the two lobes. This is the spot in which interaction takes place. The mind has the ability to move the pineal gland, and by doing so, to change the state of the brain in such a way as to produce voluntary motions. Similarly, the sensory organs all transmit their information to the pineal gland and, as a result of that, sensation is transmitted to the attached mind. However, because of the interconnection of the parts that make up the organic body, by virtue of being connected to the pineal gland, the mind can properly be said to be connected with the body as a whole (*Passions*: §§30-2) (see Persons).

9 The external world and sensation

The argument for the distinction between mind and body in Meditation VI establishes the nature of body as extension, but it does not establish the real existence of the world of bodies outside of the mind. This is the focus of the last series of arguments in the *Meditations*. The argument begins in Meditation VI with the recognition that I have 'a passive faculty of sensory perception', which would be useless unless there was also an 'active faculty, either in me or in something else' which produces the ideas of sensation. Descartes has already established in Meditation IV that the mind has only two faculties - a passive faculty of perception, and the active faculty of will. Since it is passive, perception cannot be the source of my ideas of sensation, and since sensations are involuntary, they cannot be the product of my will. So, the ideas of sense must come from somewhere else. God 'has given me a great propensity to believe that they are produced by corporeal things', and no means to correct my error if that propensity is deceptive. So, Descartes concluded, God would be a deceiver if my sensory ideas come from anything but from bodies. This argument does not prove that everything we sense about bodies is reliable, but only that 'they possess all the properties which I clearly and distinctly understand, that is, all those which, viewed in general terms, are comprised within the subject-matter of pure mathematics'. (In the *Principles* Part II §1 there is also an argument for the existence of the external world, but it is somewhat different.)

The proof of the existence of the external world tells us that, in general, bodies are the causes of our sensations and it tells us, in general, what the nature of body is. But it does not seem to tell us much about what we can (and cannot) learn about specific bodies in the world around us in specific circumstances. These questions are addressed at the end of Meditation VI in a general discussion of the reliability of sensation, the most extensive such discussion in Descartes' writings. He argues there that the senses are given to me 'simply to inform the mind of what is beneficial or harmful for the composite of which the mind is a part; and to this extent they are sufficiently clear and distinct'. That is, while they cannot tell me anything about the real nature of things - that is for the intellect or reason to determine - they can inform me about specific features of my environment that relate to

maintaining the union of my mind and body. So, for example, when the senses tell us that some particular apples are red and others green, this can give us reliable information that some may be ripe (and thus nutritious) and others not, but it cannot tell us that the one is, in its nature, red, and the other really green. Similarly, when I feel a pain in my toe, this tells me that there is damage to my toe, not that there is something resembling the sensation that is actually in the toe. Even in this, the sensation may be misleading. As Descartes points out, people sometimes feel pain in limbs even after they have been amputated.

Given the nature of the extended body, and the causal process by which pains (and other sensations) are transmitted through the body to the pineal gland, where the non-extended mind is joined to the extended body, such misleading sensations are inevitable; similar sensations in the mind can be the result of very different causal processes in the body. For example, a sensation of pain-in-the-toe can be caused either by a change in the state of the toe itself, or by an appropriate stimulation of the nerve connecting the toe to the brain at any point between the two. But, Descartes claims, though sensation is fallible, 'I know that in matters regarding the well-being of the body, all my senses report the truth much more frequently than not'. Furthermore, I can use multiple senses and memory, together with the intellect, 'which has by now examined all the causes of error' in order to weigh the evidence of the senses and use it properly. And with this, Descartes is finally able to answer the dream argument of Meditation I. For my waking experience is interconnected in a way in which my dreaming experience is not; the things I see in waking life, unlike those in dreams, come to me through all my senses, and connect with my memory of other objects. I can use this interconnectedness of waking experience, together with my intellect and my knowledge of the causes of error, to sort out veridical sensations and distinguish them from the deceptive sensory experiences of dreams. Sometimes even my waking experiences will be deceptive, of course, but we are capable of determining specific circumstances in which the senses are worthy of our trust. And so, contrary to the original doubts raised by the dreaming argument in Meditation I, there is no general reason to reject waking experience as such.

Though subordinated to reason, sensation, cast into doubt in Meditation I, re-enters as a legitimate source of knowledge about the world by the end of Meditation VI.

10 Philosophical psychology and morals

Morality was a concern of Descartes' in a variety of texts. In the third part of his *Discourse* he presents what he calls a 'provisional morality', a morality to govern our behaviour while we are in the process of revising our beliefs and coming to certainty. In the tree of philosophy in the Preface to the French edition of the *Principles*, morals is listed as one of the fruits of the tree, along with medicine and mechanics. It is also a theme in the letters he exchanged with the Princess Elisabeth of Bohemia in the mid-1640s, together with another concern - the passions, what they are and, more importantly, how to control them. These themes are intertwined again in Descartes' last major work, the *Passions of the Soul* (1649).

In one of the letters that serves as a preface to the *Passions*, Descartes announces that he will treat the passions 'only as a natural philosopher [*en physicien*], and not as a rhetorician or even as a moral philosopher'. Accordingly, the bulk of the *Passions* is taken up with detailed accounts of what the passions are, and how they arise from the connection between the human body and the human mind. As Descartes conceived them, the passions are grouped with sensation and imagination, perceptions of the mind that arise from external impulses. In this respect, Descartes differed radically from the Aristotelian scholastic philosophers who attached the passions to the appetitive faculty rather than the perceptive. But though grouped with other perceptions, the ones that concern Descartes in this treatise are a special group of perceptions, 'those whose effects we feel as being in the soul itself, and for which we do not normally know any proximate cause to which we can refer them', those 'which are caused, maintained, and strengthened by some movement of the spirits' (*Passions*: §§25, 27). (The 'spirits' in question are the animal spirits, a fluid matter that played a major role in Descartes' biology.) The principal effect of the passions is to 'move and dispose the soul to want the things for which they prepare the body. Thus the feeling of fear moves the soul to want to flee, that of courage to want to fight' and so on (*Passions*: §40). As with sensations, the passions of the soul play a role in the preservation of the mind - body union: 'The function of all the passions consists solely in this, that they dispose our soul to want the things which nature deems useful for us, and to persist in this volition; and the same agitation of the spirits which normally causes the passions also disposes the body to make movements which help us to attain these things' (*Passions*: §52).

For the schoolmen, the passions pertained to the appetitive faculty, and were principally organized around a distinction between the ‘irascible’ and the ‘concupiscent’ appetites. Descartes, however, was attempting to fashion a conception of the passions based on a very different conception of the soul, one in which there is no distinction among appetites (*Passions*: §68). His categorization of the passions was based on a list of six primitive passions, which pertain to the perceptive rather than to the appetitive faculty: wonder, desire, love and hatred, joy and sadness - ‘all the others are either composed from some of these six or they are species of them’ (*Passions*:§69). Much of his attention in the short book is directed at accounts of what each of these basic passions is, what it feels like and its physiological causes and effects in the body, and how all the other passions can be understood in terms of the six basic ones.

But although Descartes presents himself as examining the passions ‘*en physicien*’, there is a moral dimension to the discussion as well. Part of the motivation for the examination of the passions is their control. While the passions, like everything given to us by God, can contribute to our well-being, they can also be excessive and must be controlled (*Passions*: §211). While the passions are not under our direct control, by understanding what they are and how they are caused we can learn indirect means for controlling them (*Passions*: §§45-50; 211). This, Descartes asserts, is the ‘chief use of wisdom, [which] lies in its teaching us to be masters of our passions and to control them with such skill that the evils which they cause are quite bearable, and even become a source of joy’ (*Passions*: §212). Important in this process is what Descartes calls *génénerosité*, best translated as ‘nobility’. *Générosité* is the knowledge that all that belongs to us, properly speaking, is our own free will, and the resolution to use it well, ‘that is, never to lack the will to undertake and carry out whatever one judges to be best’ (*Passions*: §53). Understood in this way, *générosité* is both a passion (an immediate feeling) and a virtue (‘a habit of the soul which disposes it to have certain thoughts’) (*Passions*: §§160-1). The person who has *générosité* ‘has very little esteem for everything that depends on others’, and as a result, Descartes claims, is able to control their passions (*Passions*:§156).

11 Physics and mathematics

To his contemporaries, Descartes was as well-known for his system of physics as he was for the metaphysical views that are now more studied. Indeed, as he indicates in the Preface to the French edition of the *Principles*, metaphysics constitutes the roots of the tree of philosophy, but the trunk is physics.

Descartes’ physics was developed in two main places. The earliest is in the treatise *Le Monde (The World)* which he suppressed when Galileo was condemned for Copernicanism in 1633, though summarized in Part V of the *Discourse*. Later, in the early 1640s, he presented much of the material in a more carefully worked-out form, in Parts II, III and IV of the *Principles*. Like the physical thought of many of his contemporaries, his physics can be divided into two parts - a general part, which includes accounts of matter and the general laws of nature, and a specific part, which includes an account of particular phenomena.

The central doctrine at the foundations of Descartes’ physics is the claim that the essence of body is extension (discussed in §8 above). This doctrine excludes substantial forms and any sort of sensory qualities from body. For the schoolmen there are four primary qualities (wet and dry, hot and cold) which characterize the four elements. For Descartes, these qualities are sensations in the mind, and only in the mind; bodies are in their nature simply the objects of geometry made real. Descartes also rejected atoms and the void, the two central doctrines of the atomists, an ancient school of philosophy whose revival by Gassendi and others constituted a major rival among contemporary mechanists. Because there can be no extension without an extended substance, namely body, there can be no space without body, Descartes argued. His world was a plenum, and all motion must ultimately be circular, since one body must give way to another in order for there to be a place for it to enter (*Principles* II: §§2-19, 33). Against atoms, he argued that extension is by its nature indefinitely divisible: no piece of matter in its nature indivisible (*Principles* II: §20). Yet he agreed that, since bodies are simply matter, the apparent diversity between them must be explicable in terms of the size, shape and motion of the small parts that make them (*Principles* II: §§23, 64) (see Leibniz, G.W. §4).

Accordingly, motion and its laws played a special role in Descartes’ physics. The essentials of this account can be found in *The World*, but it is set out most clearly in the *Principles*. There (*Principles* II: §25), motion is defined as the translation of a body from one neighbourhood of surrounding bodies into another. Descartes is careful to distinguish motion itself from its cause(s). While, as we have seen, motion is sometimes caused by the volition of a

mind, the general cause of motion in the inanimate world is God, who creates bodies and their motion, and sustains them from moment to moment. From the constancy of the way in which God sustains motion, Descartes argues, the same quantity of motion is always preserved in the world, a quantity that is measured by the size of a body multiplied by its speed (*Principles* II: §36). To this general conservation law he adds three more particular laws of nature, also based on the constancy by which God conserves his creation. According to the first law, everything retains its own state, in so far as it can. As a consequence, what is in motion remains in motion until interfered with by an external cause, a principle directly opposed to the Aristotelian view that things in motion tend to come to rest (*Principles* II: §§37-8). According to the second law, bodies tend to move in rectilinear paths, with the result that bodies in circular motion tend to move away at the tangent (*Principles* II: §39). The first and second laws together arguably constitute the first published statement of what Newton, later called the law of inertia. Descartes' third law governs the collision between bodies, specifying when one body imposes its motion on another, and when two bodies rebound from one another without exchanging motion. The abstract law is followed by seven specific rules covering special cases (*Principles* II: §§40-52). Though the law of collision turns out to be radically inadequate, it casts considerable light on Descartes' conception of the physical world. One of the determinants of the outcome of a collision is what Descartes calls the 'force' of a body, both its force for continuing in motion, and its resistance to change in its motion (*Principles* II: §43). The role of such forces in Descartes' mechanist world has generated much discussion, since they would seem to be completely inconsistent with Descartes' view that the essence of body is extension alone.

These general accounts of matter and motion form the basis of Descartes' physical theories of particular phenomena. The *Principles* goes on to explain how the earth turns around the sun in an enormous fluid vortex and how the light that comes from the sun is nothing but the centrifugal force of the fluid in the vortex, with ingenious explanations of many other particular phenomena in terms of the size, shape and motion of their parts. Other works contain further mechanistic explanations, for example of the law governing the refraction of light (*Dioptrics* II) and the way colours arise in the rainbow (*Meteors* VIII).

Descartes' hope was that he could begin with an assumption about how God created the world, and then deduce, on the basis of the laws of motion, how the world would have to have come out (*Discourse* V, VI; *Principles* III: §46). But this procedure caused some problems. It is not easy to specify just how God might have created the world - whether the particles that he first created were of the same size, for example, or of every possible size. Furthermore, any hypothesis of this sort would seem to be inconsistent with the account of creation in Genesis (*Principles* III: §§43-7). These difficulties aside, it seemed obvious to Descartes how to proceed. For example, from his denial of the vacuum it would seem to follow that bodies in motion would sort themselves out into circular swirls of matter, the vortices which were to explain the circulation of the planets around a central sun. Similarly, Descartes used the tendency to centrifugal motion generated by the circular motion of the vortex to explain light, which, he claimed, was the pressure of the subtle matter in the vortex. But the very complexity of the world militates against the full certainty that Descartes originally sought, particularly when dealing with the explanation of particular phenomena, such as the magnet. Indeed, by the end of the *Principles*, it can seem that he has given up the goal of certainty and come to accept the kind of probability that he initially rejected (*Principles* IV:§204-6).

Central to Descartes' physics is his rejection of final causes: 'When dealing with natural things we will, then, never derive any explanations from the purposes which God or nature may have had in view when creating them. For we should not be so arrogant as to suppose that we can share in God's plans' (*Principles* I: §28). The emphasis on efficient causes was to prove very controversial later in the century.

One especially curious feature of Descartes' physics, however, is the lack of any substantive role for mathematics. Descartes was one of the great mathematicians of his age. While it is, perhaps, anachronistic to see modern analytic geometry and so-called Cartesian coordinates in his *Geometry* (1637), there is no question but that it is a work of real depth and influence. In it he shows how one could use algebra to solve geometric problems and geometry to solve algebraic problems by showing how algebraic operations could be interpreted purely in terms of the manipulation and construction of line segments. In traditional mathematics, if a quantity was represented as a given line, then the square of that quantity was represented as a square constructed with that line as a side, and the cube of the quantity represented as a cube constructed with that line as an edge, effectively limiting the geometric representation of algebraic operations to a very few. By demonstrating how the square, cube (and so on) of a given

quantity could all be represented as other lines, Descartes opened the way to a more complete unification of algebra and geometry. Also important to his mathematical work was the notion of analysis. Descartes saw himself as reviving the work of the ancient mathematician, Pappus of Alexandria, and setting out a methodology for the solution of problems, a methodology radically different from the Euclidean style of doing geometry in terms of definitions, axioms, postulates and propositions, which he regarded as a method of presentation rather than a method of discovery. According to the procedure of analysis, as Descartes understood it, one begins by labelling unknowns in a geometric problem with letters, setting out a series of equations that involve these letters, and then solving for the unknowns to the extent that this is possible.

Unlike his contemporary, Galileo, or his successors, Leibniz and Newton, Descartes never quite figured out how to apply his mathematical insights to the physical world. Indeed, it is a curious feature of his tree of knowledge that, despite the central place occupied by mathematics in his own accomplishments, it seems to have no place there.

12 Life and the foundations of biology

The last part of the Cartesian programme was his biology. First presented in the *Treatise on Man*, part of *The World* project which was abandoned in 1633 when Galileo was condemned, Descartes intended to rework some of that material and publish it - as Parts V and VI of the *Principles* under the title ‘De Homine’. Although he never finished this rewriting, it is clear from the notes left behind that it was very much on his mind in years that preceded his sudden and premature death.

His hope seems to have been to show how, from matter and the laws of motion alone, life would arise spontaneously as matter came to organize itself in an appropriate way (*Discourse* V). Unfortunately he never worked out this view, suggestive of later theories of evolution, in any detail. Yet, he was quite clear that all the functions of life (with the exception of thought and reason in humans) are to be explained not in terms of the soul, the principle of life, but in terms of matter in motion. Accordingly, in the *Treatise on Man*, he accounts for a variety of phenomena, including digestion, involuntary motion, the action of the heart, and sense perception, in purely mechanical terms. (Summarized in *Discourse* Part V, with special emphasis on the circulation of the blood.)

While Descartes’ biology was controversial among his contemporaries, one aspect was especially so. According to his account, there is only one kind of soul in the world, the rational soul, which humans and angels have and animals do not. Humans are organic machines, collections of matter organized so as to be able to perform vital functions, attached to rational souls. Animals, on the other hand, are just machines: their behaviour is purely mechanical and they are, strictly speaking, incapable of conscious experience of any sort (*Discourse* V).

13 The Cartesian heritage

It is difficult to overestimate the influence of Descartes. In philosophy, the Cogito Argument signalled the centrality of the self and the rejection of authority from without, the authority of both texts and teacher. For physics, Descartes represented the rejection of the scholastic physics of matter and form, and its replacement by a mechanistic physics of matter and motion. So in biology, he stood for mechanism and the rejection of Aristotelian vitalism.

Descartes had many followers who took his ideas (as they understood them) as dogma, and developed them as they thought he would have wanted them to do. The most important centres of Cartesian thought were France, where he was remembered as a countryman, despite his long absence, and the Netherlands, where he had lived. In France, his thought was carried on by a circle around Claude Clerselier, who gathered and published his letters as well as other works. Louis de La Forge commented on Descartes’ physiology, and wrote a Cartesian treatise on the mind, extending Descartes’ ideas. Gerauld de Cordemoy, tried to blend Cartesian philosophy with atomism, to the puzzlement of most of his contemporaries. Jacques Rohault was influential in Cartesian physics well after Newton had published the work that would eventually eclipse such theories. Other followers, mainly in the Low Countries, include Henricus Regius, considered Cartesian by many despite Descartes’ public rejection; Adriaan Heereboord, one of Descartes’ partisans in Leiden; Johannes de Raey, one of those who attempted to reconcile Descartes with the true philosophy of Aristotle; and Johannes Clauberg, who recast Cartesianism into more scholastic garb. There were many more minor Cartesians of various nationalities. Late seventeenth-century Europe was flooded with

paraphrases of and commentaries on Descartes' writings.

Other more independent thinkers were strongly influenced by Descartes without explicitly being followers. The best-known such figure is probably Nicolas Malebranche. While his thought owes much to other influences, particularly to seventeenth-century Augustinianism, in his *Recherche de la vérité (Search after Truth)* (1674-5) he follows Descartes in offering a critique of the senses, rejecting the authority of tradition, and appealing to clear and distinct perceptions. Descartes was also an important influence on the Cambridge Platonist Henry More, who regarded Descartes' philosophy, in particular his distinction between mind and body, as support for his own attacks on materialism (see Cambridge Platonism). Spinoza, too, was influenced by Descartes. His first published book was a commentary on Descartes' *Principles*, and although he later moved well outside the Cartesian camp, Descartes' doctrines helped to structure his mature thought. Spinoza's metaphysical vocabulary (substance, attribute and mode) is borrowed from Descartes, as is the centrality of the attributes of thought and extension in his metaphysics.

While many of Descartes' partisans tried to remain orthodox, there is at least one doctrine characteristic of later Cartesianism that Descartes himself probably did not hold, namely, occasionalism (see Occasionalism). Malebranche and the Flemish Cartesian Arnold Geulincx are most often associated with the doctrine, but it appears in Cartesian writings long before theirs. According to occasionalism, God is the only active causal agent in the world; finite minds and bodies are not real causes, but only occasions for God to exercise his causal efficacy. Motivated by the picture of divine sustenance from moment to moment that underlies Descartes' derivation of the laws of motion, together, perhaps, with general worries about the efficacy of finite causes and specific worries about mind - body interaction, occasionalism became a standard doctrine. Though often also attributed to Descartes himself, the grounds for doing so are rather slim.

Descartes' mark can also be seen among his opponents. He was clearly a target of Hobbes' materialism and sensationalism in, for example, Part I of *Leviathan* (1651). His epistemology and treatment of God were explicitly targeted by Pascal in the *Pensées* (1658-62, published 1670). Leibniz, too, attacked his physics, his rejection of formal logic, his conception of body and his conception of the mind, among many other things. The inadequacy of the Cartesian philosophy is a constant subtext to Locke's *Essay Concerning Human Understanding* (1689), particularly in his discussion of our knowledge of mind and his rejection of the dogmatic claim to know the essences of substances. In natural philosophy, Newton's early writings show a careful study of Descartes' writings, particularly those on motion, and book II of his *Principia* was devoted to a refutation of the vortex theory of planetary motion. Between around 1650 and the eclipse of Cartesian philosophy some time in the early eighteenth century, it was simply impossible to write philosophy without reacting in some way to Descartes.

See also: Aristotelianism in the 17th century; Certainty; Doubt; Dualism; Locke, J.; Aristotelianism, medieval; Perception, epistemic issues in; Rationalism; Régis, P.-S.; Scepticism; Substance

DANIEL GARBER

List of works

Descartes, R. (1964-74) *Oeuvres de Descartes*, ed. C. Adam and P. Tannery, Paris: CNRS/Vrin, new edn, 11 vols. (Originally published 1897-1913, and more recently updated, this is still the standard edition of Descartes' writings in the original languages. It contains editions of all of the writings listed below, as well as his letters. The incomplete *La description du corps humain* and *Prima cogitationes* are included in volume 11.)

Descartes, R. (1984-91) *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 3 vols.(The now-standard English translation of Descartes' writings. It contains the entire *Rules*, *Discourse*, *Meditations* and *Passions*, as well as selections from his other writings and letters.)

Descartes, R. (1620-*c*.28) *Regulae ad directionem ingenii (Rules for the Direction of the Mind)*, in vol. 1 of *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91.(Descartes' early treatise on method, left unpublished at his death. Manuscripts were widely circulated, and it was published first in Dutch translation in 1684 and in the Latin original in 1701.)

Descartes, R. (c.1630-3) *Le Monde (The World)*, excerpted in vol. 1 of *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91.(Descartes' first draft of a scientific system, including general physics, cosmology, terrestrial

physics, and human physiology. Withdrawn from publication when Galileo was condemned, the *Treatise on Man* was first published in Latin translation in 1662, then in the French original in 1664; the *Treatise on Light* was first published in French in 1664.)

Descartes, R. (1637) *Discours de la méthode pour bien conduir sa raison et chercher la vérité dans les sciences plus la dioptrique, les meteores, et la geometrie, qui sont des essais de cete methode (Discourse on the Method for Properly Conducting Reason and Searching for Truth in the Sciences, as well as the Dioptrics, the Meteors, and the Geometry, which are essays in this method)*, in vol. 1 of *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91.(Descartes' first publication: three scientific treatises, together with an introduction that was to become more famous and widely read than the essays that it introduced.)

Descartes, R. (1641) *Meditationes de prima philosophia (Meditations on First Philosophy)*, in vol. 2 of *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91.(Descartes' main metaphysical work, it was published with a series of objections together with Descartes' replies, six sets in the 1641 edition; a seventh was added in the 1642 edition. Also important is the 1647 French translation, published with some changes.)

Descartes, R. (1644) *Principia philosophiae (Principles of Philosophy)*, excerpted in vol. 1 of *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91. (After Part I, which deals with metaphysics, this mainly deals with Descartes' physics. Also important is the 1647 French translation, with a new preface and some significant changes.)

Descartes, R. (1649) *Les passions de l'âme (The Passions of the Soul)*, in vol. 1 of *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91. (Descartes' last published work, this deals with philosophical psychology and morals.)

References and further reading

The journal *Archives de Philosophie* annually publishes a report from the Centre d'études cartésiennes (Université de Paris IV (Sorbonne)), which contains full bibliographies of recent work on Descartes and Cartesianism, together with selective reviews.

Baillet, A. (1691) *La Vie de M. Descartes (Life of Descartes)*, Paris: Daniel Horthemels, 2 vols.(The most important source on Descartes' life. It was his 'official' biography, commissioned by a circle of Descartes' followers. Baillet had access to numerous papers that are no longer extant.)

Beck, L.J. (1952) *The Method of Descartes: A Study of the Regulae*, Oxford: Oxford University Press.(Detailed study of Descartes' method, as presented in the *Rules*.)

Beck, L.J. (1965) *The Metaphysics of Descartes: A Study of the Meditations*, Oxford: Oxford University Press.(A careful study of Descartes' *Meditations*.)

Beeckman, I. (1604-34) *Journal tenu par Isaac Beeckman de 1604 à 1634*, ed. C. de Waard, The Hague: Nijhoff, 4 vols, 1939-53.(Excerpts from the journals, which include discussons between Descartes and Beeckman, are also included in *Oeuvres de Descartes*, ed. C. Adam and P. Tannery, Paris: CNRS/Vrin, 1964-74, vol 10: 41-78, 151-69.)

Bouillier, F. (1868) *Histoire de la philosophie cartésienne (History of Cartesian philosophy)*, Paris: Durand, 3rd edn, 2 vols; repr. New York: Garland, 1987.(After more than a century, this remains the best study of Descartes' influence on his contemporaries and followers.)

Chappell, V. and Doney, W. (eds.) (1987) *Twenty-five years of Descartes Scholarship 1960-1984. A Bibliography*, New York: Garland.(An update of Sebba's earlier bibliography.)

Clarke, D.M. (1982) *Descartes' Philosophy of Science*, Manchester: Manchester University Press.(A study of Descartes' science and its philosophy, from the point of view of Descartes' actual practice.)

Cottingham, J.G. (1986) *Descartes*, Oxford: Blackwell.(Good introductory study of Descartes' philosophical thought.)

Cottingham, J.G. (ed.) (1992) *The Cambridge Companion to Descartes*, Cambridge: Cambridge University Press. (This collection has good, up-to-date overviews of current thought on a variety of topics in Descartes' philosophy.)

Curley, E.M. (1978) *Descartes against the Skeptics*, Cambridge, MA: Harvard University Press.(A study of Descartes' philosophical writings as a response to classical scepticism.)

Des Chene, D. (1996) *Physiologia: Natural philosophy in late Aristotelian and Cartesian thought*, Ithaca, NY: Cornell University Press.(An up-to-date study of Descartes' relation to scholastic philosophy, progressively

becoming a central theme in Cartesian studies.)
Doney, W. (ed.) (1967) *Descartes: A Collection of Critical Essays*, Garden City, NY: Doubleday.(Collection of classic and often-cited articles on Descartes' philosophy.)
Frankfurt, H. (1970) *Demons, Dreamers and Madmen: The Defense of Reason in Descartes' Meditations*, Indianapolis, IN: Bobbs-Merill.(An imaginative and highly influential reading of Descartes' epistemology, particularly the Cartesian circle.)
Galilei, G. (1638) *Discorsi e dimostrazioni matematiche, intorno à due nuove scienze*, trans. S. Drake, *Two New Sciences*, Madison, WI: University of Wisconsin Press, 1974.(The work dismissed by Descartes as lacking systematic coherence.)
Garber, D. (1992) *Descartes' Metaphysical Physics*, Chicago, IL: University of Chicago Press.(This concentrates on Descartes' natural philosophy and its relation to his more philosophical interests.)
Gaukroger, S. (ed.) (1980) *Descartes: Philosophy, Mathematics and Physics*, Sussex: Harvester Press.(This collection contains a number of now-classic essays on Descartes' physics and mathematics in relation to his more philosophical concerns.)
Gaukroger, S. (1995) *Descartes: An Intellectual Biography*, Oxford: Oxford University Press. (This biography also serves as a historically sensitive commentary on Descartes' thought. It is especially good on Descartes' earlier period.)
Gilson, É (1930) *Études sur le rôle de la pensée médievale dans la formation du système cartésien (Studies on the role of Medieval thought in the formation of the Cartesian system of philosophy)*, Paris: Vrin.(A classic discussion of Descartes' relations to the scholastic thought which dominated earlier philosophy.)
Grosholz, E. (1991) *Cartesian Method and the Problem of Reduction*, Oxford, Oxford University Press.(Study of Cartesian method, particularly good for its discussion of Descartes' mathematics.)
Gueroult, M. (1984) *Descartes' Philosophy Interpreted According to the Order of Reasons*, trans. R. Ariew, Minneapolis, MN: University of Minnesota Press, 2 vols. (Though Gueroult has certain interpretative biases, this remains one of the great close commentaries on Descartes' philosophy in the French tradition.)
Hintikka, J. (1968) '*Cogito, Ergo Sum*: Inference or Performance', in W. Doney (ed.) *Descartes: A Collection of Critical Essays*, Garden City, NY: Doubleday, 108-139.(The main adherent of the approach to the Cogito through speech act theory and theories of demonstratives, mentioned in §5 above.)
Kenny, A. (1968) *Descartes*, New York: Random House.(Classic introduction to Descartes' thought.)
Marion, J.-L. (1975) *Sur l'ontologie grise de Descartes (On Descartes' hidden ontology)*, Paris: Vrin, 2nd edn, 1981.(A study of Descartes' *Rules* in opposition to Aristotle, arguing that Descartes smuggles in a kind of secret ontology. This and the following two books constitute a trilogy, offering a somewhat idealistic interpretation of Descartes' thought. They have been widely read, and enormously influential on recent Cartesian studies.)
Marion, J.-L. (1981) *Sur la théologie blanche de Descartes (On Descartes' blank theology)*, Paris: Presses Universitaires de France.(A study of Descartes' conception of God and the role he plays in the Cartesian system, focusing on the conception of God as the creator of eternal truths.)
Marion, J.-L. (1986) *Sur le prisme métaphysique de Descartes. Constitution et limites de l'onto-théo-logie dans la pensée cartésienne (On Descartes' metaphysical prism: the constitution and limits of onto-theo-logy in Cartesian thought)*, Paris: Presses Universitaires de France.(A Heideggerian interpretation of the metaphysics of the *Meditations*, concentrating on the relation between God and the self.)
Markie, P. (1992) 'The Cogito and its importance', in J. Cottingham (ed.) *The Cambridge Companion to Descartes*, Cambridge: Cambridge University Press.(A general survey of analytic interpretations of the Cogito.)
Regius, H. (1646) *Fundamenta physices (Foundations of physics)*, Amsterdam.(Regius' main presentation of his natural philosophy, about which Descartes had serious reservations.)
Rodis-Lewis, G. (1971) *L'oeuvre de Descartes (The work of Descartes)*, Paris: Vrin, 2 vols.(A chronological study of Descartes' thought and writings that also serves as an intellectual biography. Elegantly written, it shows an unusually broad knowledge of Descartes' writings. Also useful for its encyclopedic discussion of the French literature on Descartes before 1971.)
Rorty, A.O. (ed.) (1986) *Essays on Descartes' Meditations*, Berkeley, CA: University of California Press.(This collection of essays, all on themes from the *Meditations*, serves as a collective commentary on the work.)
Sabra, A.I. (1961) *Theories of Light from Descartes to Newton*, London: Oldbourne; 2nd edn, Cambridge: Cambridge University Press, 1981.(Study of seventeenth-century optics, particularly good on Descartes contribution to the subject.)
Sebba, G. (1964) *Bibliographia Cartesiana. A Critical Guide to the Descartes Literature 1800-1960*, The Hague:

Matinus Nijhoff.(The standard source for the older literature on Descartes. In addition to a listing by author, it contains some commentary on the more important items.)

Shea, W.R. (1991) *The Magic of Numbers and Motion: The Scientific Career of René Descartes*, Canton, MA: Science History Publications.(This study of Descartes' scientific thought is particularly good on his early years.)

Verbeek, T. (1992) *Descartes and the Dutch: Early Reactions to Cartesianism (1637-1650)*, Journal of the History of Philosophy Monograph Series, Carbondale, IL: Southern Illinois University Press.(Excellent study of Descartes' intellectual and personal relations with the people among whom he lived for the better part of his adult life.)

Voss. S. (ed.) (1993) *Essays on the Philosophy and Science of René Descartes*, New York and Oxford; Oxford University Press.(Collection of articles that gives a good idea of current/recent work.)

Vuillemin, J. (1960) *Mathématiques et métaphysique chez Descartes (Descartes' mathematics and metaphysics)*, Paris: Presses Universitaires de France.(A very valuable study of Descartes' mathematics understood in the context of his philosophy.)

Williams, B. (1978) *Descartes: the Project of Pure Enquiry*, Hassock: Harvester. (Classic introduction to Descartes' philosophical thought from an analytic point of view.)

Wilson, M. (1978) *Descartes*, London: Routledge.(An excellent and influential commentary, centred on the *Meditations*.)

Descriptions

'Definite descriptions' are noun phrases of the form 'the' + noun complex (for example, 'the finest Greek poet', 'the cube of five') or of the form possessive + noun complex (for example, 'Sparta's defeat of Athens'). As Russell realized, it is important to philosophy to be clear about the semantics of such expressions. In the sentence 'Aeschylus fought at Marathon', the function of the subject, 'Aeschylus', is to refer to something; it is a referential noun phrase (or 'singular term'). By contrast, in the sentence 'Every Athenian remembers Marathon', the subject noun phrase, 'every Athenian', is not referential but quantificational. Definite descriptions appear at first sight to be referential. Frege treated them referentially, but Russell held that they should be treated quantificationally in accordance with his theory of descriptions, and argued that certain philosophical puzzles were thereby solved.

1 Frege

Gottlob Frege provided the first systematic account of quantification in natural language and the first systematic theory of reference (1892). The class of 'singular terms' (referential noun phrases), for Frege, was delimited by a set of logical tests (for example, the licensing of existential generalization) and was recursive. It included ordinary proper names and definite descriptions. Thus '5', 'the cube of 5', 'Aeschylus' and so on were all singular terms.

If a description $\ulcorner \text{the } F \urcorner$ is referential, then it is natural to take its reference to be the unique entity satisfying '*F*'; a sentence $\ulcorner \text{The } F \text{ is } G \urcorner$ is true if and only if that entity is *G*. But what if no entity, or more than one entity, satisfies the descriptive condition, as in (1) or (2)?

(1) *The largest prime number* lies between 10^{23} and 10^{27}.
(2) *The man who landed on the moon* was American.

Such descriptions are said to be 'improper'. Frege considered it a defect of natural language that it permits the possibility of improper terms. As far as his own logical system was concerned, he thought it essential that every formula have a truth-value, and so he insisted that every singular term have a reference (or meaning): he stipulated that a specified object in the range of the quantifier(s) serve as the referent of every improper description. While this stipulation proved useful for his formal system, Frege recognized that some other account of improper descriptions in ordinary language was needed. Once he had made his distinction between the 'sense' and 'reference' of an expression, he suggested that an improper term has a sense but no reference (see Frege, G. §3; Sense and reference §1). The main problem with this proposal is that it predicts, rather counterintuitively, that any sentence containing an improper term (in a transparent context) lacks a truth-value. For example, it predicts that (1) and (2) above lack truth-values.

2 Russell's theory of descriptions: informal characterization

Like Frege, Bertrand Russell thought it important to explain how a sentence such as (1) or (2) could be meaningful. At one time he entertained the idea of a realm of non-existent entities to serve as the referents of descriptions such as 'the largest prime number' and 'the round square'; but by 1905 he thought this idea conflicted with a 'robust sense of reality' and his theory of descriptions came about, in part as an attempt to purify his ontology (see Existence §2).

On Russell's account, descriptions are not singular terms at all but phrases that logical analysis reveals to be quantificational: if $\ulcorner \text{the } F \urcorner$ is a definite description and $\ulcorner \ldots \text{is } G \urcorner$ is a predicate phrase, then the proposition expressed by an utterance of $\ulcorner \text{The } F \text{ is } G \urcorner$ is equivalent, says Russell, to the proposition expressed by an utterance of $\ulcorner$There is exactly one *F*, and everything that is *F* is $G \urcorner$. That is, $\ulcorner$The *F* is $G \urcorner$ is analysed as

(3) $(\exists x)((\forall y)(Fy \leftrightarrow y = x) \;\&\; Gx)$.

The proposition expressed by $\ulcorner \text{The } F \text{ is } G \urcorner$ is 'general' ('object-independent') rather than 'singular' ('object-dependent') in the sense that there is no object for which its grammatical subject stands; upon whose existence that of the proposition expressed depends. Unlike a singular term, a definite description, even if it is in fact satisfied by a unique object, does not actually *refer* to that object. It is as wrong, on Russell's account, to inquire into the referent of $\ulcorner \text{the } F \urcorner$ as it is to inquire into the referent of $\ulcorner \text{every } F \urcorner$ or $\ulcorner \text{no } F \urcorner$.

On Russell's account, sentences containing improper descriptions have truth-values. For example, (1) above is false as it is not the case that there exists a largest prime number. Similarly, sentence (2) is false as it is not the case that there exists exactly one man who landed on the moon.

Russell's theory opens up the possibility of accounting for certain *de dicto*/*de re* ambiguities in terms of scope permutations (see De re/de dicto). For example, (4) below may be analysed as either (5) or (6), according to whether the description 'the largest prime number' is given large or small scope with respect to 'John thinks that':

(4) John thinks that the largest prime number lies between 10^{23} and 10^{27}.
(5) $(\exists x)((\forall y)(\text{largest-prime } y \leftrightarrow y = x)$ & John thinks that: *x* lies between 10^{23} and 10^{27}).
(6) John thinks that: $(\exists x)((\forall y)(\text{largest-prime } y \leftrightarrow y = x)$ & *x* lies between 10^{23} and 10^{27}).

(5) is false; but (6) may be true. Thus Russell is able to explain the intuitive ambiguity of (4) and avoid positing an ontology that includes such things as a largest prime number. Smullyan (1948) points out that Russell's theory similarly explains *de dicto*/*de re* ambiguities in modal contexts, for example, in 'The number of planets is necessarily odd' (see De re/de dicto §2; Modal logic).

Russell came to treat ordinary proper names as 'disguised' or 'truncated' descriptions. For example, the name 'Cicero' might be analysed as the description 'the greatest Roman orator' while the coreferential name 'Tully' might be analysed as the description 'the author of *De Fato*'. On the face of it, this provided Russell with accounts (not dissimilar to Frege's) of why 'Cicero was bald' and 'Tully was bald' differ in informativeness, and of why (7) and (8) need not agree in truth-value:

(7) John believes Cicero was bald.
(8) John believes Tully was bald.

In the light of Kripke's work on names and necessity (1980), it is widely held that descriptive analyses of proper names cannot succeed (see Proper names). There is good reason, however, to think that at least some pronouns anaphorically linked to (referring back to), but not bound by, quantified noun phrases are understood in terms of definite descriptions (see Neale 1990).

3 Russell's theory of descriptions: formal characterization

On Russell's account, descriptions are 'incomplete' symbols; they have 'no meaning in isolation', that is, they do not stand for things. In *Principia Mathematica* (1910-13), descriptions are represented by quasi-singular terms of the form '$(\iota x)(Fx)$', which can be read as 'the unique *x* which is *F*'. Superficially, the iota-operator is a variable-binding device for forming a term from a formula. A predicate symbol '*G*' may be prefixed to a description '$(\iota x)(Fx)$' to form a formula '$G(\iota x)(Fx)$', which can be expanded in accordance with a suitable 'contextual definition'. (To define an expression ζ contextually is to provide a procedure for converting any sentence containing occurrences of ζ into an equivalent sentence that is ζ-free.)

The analysis in (3) above does not constitute a final contextual definition of '$G(\iota x)(Fx)$' because of the possibility of scope ambiguity where a formula containing a description is itself a constituent of a larger formula (see Scope). Scope ambiguity is conveniently illustrated with descriptions in the context of negation. For a genuine singular term α, there is no difference between wide and narrow scope negation: α is not-*F* just in case it is not the case that α is *F*. For a description, however, there is a formal ambiguity. Let '*Kx*' represent '*x* is a king of France' and '*Wx*' represent '*x* is wise'. Then the formula '$\neg W(\iota x)(Kx)$' ('The king of France is not wise') is ambiguous between (9) and (10):

(9) $(\exists x)((\forall y)(Ky \leftrightarrow y = x)$ & $\neg Wx)$
(10) $\neg(\exists x)((\forall y)(Ky \leftrightarrow y = x)$ & $Wx)$.

These are not equivalent: only (10) can be true when there is no king of France. In *Principia Mathematica*, the scope of a description is specified by appending a copy of it within square brackets to the front of the formula that constitutes its scope. Thus (9) and (10) are represented as (11) and (12) respectively:

(11) $[(\iota x)(Kx)]\neg\{W(\iota x)(Kx)\}$
(12) $\neg\{[(\iota x)(Kx)]W(\iota x)(Kx)\}$.

In (11) the description has what Russell calls a ‘primary occurrence’ by virtue of having scope over the negation; in (12) the description has a ‘secondary occurrence’ by virtue of lying within the scope of the negation. Where a description has smallest possible scope, it is conventional to omit the scope marker; thus (12) can be reduced to ‘ $\neg W(\iota x)(Kx)$’.

With the matter of scope behind us, the theory of descriptions can be stated exactly:

$[(\iota x)(\phi x)]G(\iota x)(\phi x) =_{\mathrm{df}} (\exists x)((\forall y)(\phi y \leftrightarrow y = x) \;\&\; Gx),$

where ϕ is a formula. On Russell’s account, there is no possibility of a genuine referring expression failing to refer, so no predicate letter in the language of *Principia Mathematica* stands for ‘exists’. Russell introduces a symbol ‘E!’ (‘E shriek’) that may be combined with a description to create a well-formed formula. Thus

$\mathrm{E!}(\iota x)(\phi x) =_{\mathrm{df}} (\exists x)(\forall y)(\phi y \leftrightarrow y = x).$

‘E!’ allows a treatment of negative existentials. (According to Russell, an utterance of ‘The king of France does not exist’ made today would be true precisely because there is no king of France.) Successive applications will allow any well-formed formula containing a definite description to be replaced by an equivalent formula that is description-free.

It is often objected that Russell’s theory, which substitutes complex quantificational structure for ‘the’, is unfaithful to surface syntax. The objection is engendered by an insufficiently keen appreciation of the distinction between a theory and its formal implementation. The extent of the mismatch between ‘The king is wise’ and its analysis

(13) $(\exists x)((\forall y)(\text{king } y \leftrightarrow y = x) \;\&\; \text{wise } x)$

has nothing to do with descriptions *per se*. In order to characterize the logical forms of even ‘some philosophers are wise’ and ‘every philosopher is wise’ in the predicate calculus we have to use formulas containing sentence connectives, no counterparts of which occur in the surface forms of the sentences:

(14) $(\exists x)$(philosopher *x* & wise *x*)
(15) $(\forall x)$(philosopher $x \rightarrow$ wise *x*).

And when we formalize sentences such as ‘Just two philosophers are wise’, we find much more complexity than there is in surface syntax:

(16) $(\exists x)(\exists y)$[philosopher *x* & philosopher *y* & wise *x* & wise *y* & $(\forall z)$((philosopher *z* & wise *z*) $\rightarrow z = x \vee z = y$)].

The supposed problem about descriptions, then, is in fact a symptom of a larger problem involving the application of first-order logic to sentences of ordinary language.

Work on ‘generalized’ quantification provides a solution to the larger problem (as well as treatments of quantifiers such as ‘most’ that cannot be handled within first-order logic; see Quantifiers, generalized). Natural language quantification is normally restricted: we talk about all philosophers or most poets, not about all or most entities. A simple modification of the predicate calculus yields a language - call it ‘RQ’ - that captures this fact while retaining the precision of regular first-order logic. In RQ, a determiner such as ‘some’, ‘every’ or ‘no’ combines with a formula to create a restricted quantifier such as ‘[every *x*: philosopher *x*]’. And such a quantifier may combine with a formula to form a formula:

(17) [every *x*: philosopher *x*] (wise *x*).

The viability of such a language shows that the language of *Principia Mathematica* is not essential to the theory of descriptions. Since the word ‘the’ is a one-place quantificational determiner (as are ‘some’, ‘every’, ‘no’ and so on), RQ can treat ‘the’ as combining with a formula ‘king *x*’ to form a restricted quantifier ‘[the *x*: king *x*]’. The sentence ‘The king is wise’ will then be represented as

(18) [the *x*: king *x*] (wise *x*).

Different scope possibilities are easily captured. For instance, ‘The king is not wise’ is ambiguous between (19) and (20):

(19) [the *x*: king *x*] ¬(wise *x*)
(20) ¬[the *x*: king *x*](wise *x*).

Using a formal language in which descriptions are treated as restricted quantifiers does not mean abandoning Russell’s view that descriptions are ‘incomplete symbols’ that ‘disappear on analysis’. Rather, treating descriptions as restricted quantifiers results in an explanation of where his theory of descriptions fits into a systematic account of natural language quantification, a theory in which ‘every’, ‘some’, ‘most’, ‘a’, ‘the’ and so on are members of a unified syntactic and semantic category.

4 Strawson’s theory and criticisms of Russell

As part of a broad critique of the idea that the semantics of formal languages can be used to analyse the meanings of statements of natural language, P.F. Strawson argued against Russell’s theory of descriptions on the grounds that (1) it fails to recognize that referring is something done by speakers and not expressions, (2) it fails to do justice to the way speakers ordinarily use sentences containing descriptions to make statements (speakers use descriptions to refer, not to quantify) and (3) it rides roughshod over important distinctions, such as the distinction between the meaning of a sentence σ and the statement made by a particular use of σ (see Propositions, sentences and statements).

Using as an example ‘The present king of France is wise’, Strawson argues that Russell’s theory is thwarted because the same sentence can be used to say something true on one occasion and something false on another. It is certainly true that Russell paid little attention to the distinction between the linguistic meaning of a sentence type and the proposition expressed by a particular dated utterance of that sentence type; but it was the latter that actually concerned him, and Strawson could get no mileage out of Russell’s inattention to the distinction. The fact that a description (or any other quantified noun phrase) may contain an indexical component (‘the *present* king of France’, ‘every man *here*’, and so on) illustrates that some descriptions are subject to both the theory of descriptions and a theory of indexicality (see Demonstratives and indexicals). Thus contextual features play a role in fixing the proposition expressed. And this can be true also if the overt indexical element is missing, as in ‘The king of France is wise’.

This appreciation of contextual factors forms the basis of the Russellian response to a second Strawsonian objection. According to this, someone who uses a description $\ulcorner \mathrm{the}\ F \urcorner$ typically intends to refer to some object or other and say something about it; there is no question of claiming that some object uniquely satisfies *F*. Someone who says ‘The table is covered with books’, for instance, does not express a proposition that entails the existence of exactly one table. But, Strawson claimed, it is a part of the meaning of $\ulcorner \mathrm{the}\ F \urcorner$ that such an expression is used correctly only if there is an *F*. If this condition is not satisfied - if the ‘presupposition’ that there is an *F* is false - a use of $\ulcorner \mathrm{The}\ F\ \mathrm{is}\ G \urcorner$ cannot be considered to express a proposition that is either true or false. The Russellian response to Strawson is that descriptions such as ‘the table’ are often understood as elliptical uses of fuller descriptions such as ‘the table over there’, ‘the table in front of me’ and so on; or else they are evaluated with respect to a restricted domain of discourse. Again the phenomenon is not confined to descriptions, but is found with quantified noun phrases more generally.

Strawson’s original statement (1950) of his own theory contains an interesting ambiguity. He can be understood as claiming either that no proposition is expressed, or that a proposition which is neither true nor false is expressed, when someone uses a sentence containing an empty description. A second ambiguity comes with the notion of ‘presupposition’. This can be viewed as an epistemological or pragmatic relation between a person and a statement, or as a logical relation between two statements (see Presupposition). An epistemological or pragmatic notion of presupposition appears to have no bearing on the semantic issues Strawson wanted to address when he challenged Russell.

The Strawsonian position faces some serious obstacles. If someone were to utter (21) right now, they would unquestionably say something false.

(21) The king of France shot my cat last night.

But on Strawson's account, the speaker will have expressed no proposition because the presupposition that there is a unique king of France is false. Descriptions occurring in the context of attitude verbs create a similar problem. For example, someone might utter a true statement using (22):

(22) Ponce de León thought the fountain of youth was in Florida,

so the presence of an empty description does not always result in a failed speech act. This is something Strawson (1964) came to concede. In order to reduce the number of incorrect predictions made by his earlier theory, he suggested that the presence of an empty description sometimes renders the proposition expressed false and at other times prevents a proposition from being expressed at all (sometimes $\ulcorner$The F is $G\urcorner$ *entails* the existence of a unique *F*, and at other times it (only) *presupposes* it).

5 Ambiguity theories

Consideration of the behaviour of descriptions in non-extensional contexts and the possibility of misdescribing an individual as the *F*, but successfully communicating something about that individual, has led some philosophers (for example, Donnellan 1966) to suggest that descriptions are sometimes quantificational, at other times referential. When $\ulcorner$the $F\urcorner$ is used in the Russellian way, the proposition expressed is general (object-independent); when it is used referentially the proposition expressed is singular (object-dependent). Donnellan considers examples such as the following: (1) A detective discovers Smith's mutilated body but has no idea who killed him. Looking at the body, he exclaims, 'The murderer is insane'. (2) Jones is on trial for Smith's murder; I am convinced of his guilt; hearing Jones ranting in court, I say, 'The murderer is insane'. On Donnellan's account, in case (1) the description is being used attributively; in case (2) it is being used referentially. Cases such as (2), it is argued, cannot be treated in accordance with Russell's theory. Following Grice (1969), however, many have argued (1) that so-called referential uses of descriptions can usually be accommodated within Russell's theory by invoking a distinction between the proposition expressed by (an utterance of) a sentence on a given occasion and the proposition the speaker primarily intends to communicate on that occasion; (2) that the phenomenon of referential usage is not specific to definite descriptions, but arises with quantified noun phrases quite generally; (3) that the referential/attributive distinction is neither exclusive nor exhaustive; and (4) that no such distinction can do the work of Russell's notion of the scope of a description. It would seem, then, that something very close to Russell's theory will probably form a component of any finally acceptable theory.

See also: Free logics, philosophical issues in §3; Logical and mathematical terms, glossary of; Reference §6

STEPHEN NEALE

References and further reading

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Discussion of views presented in §§1-2.)

Davies, M. (1981) *Meaning, Quantification and Necessity*, London: Routledge. (Discussion of views presented in §5.)

Donnellan, K. (1966) 'Reference and Definite Descriptions', *Philosophical Review* 77: 281-304.(Discussion of views presented in §5.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980.(Source of views presented in §1. The page numbers of the original publications appear in the text.)

Grice, H.P. (1969) 'Vacuous Names', in D. Davidson and J. Hintikka (eds) *Words and Objections: Essays on the Work of W.V. Quine*, Dordrecht: Reidel, 118-45. (Discussion of views presented in §§2, 5.)

Kaplan, D. (1978) 'Dthat', in P. Cole (ed.) *Syntax and Semantics*, vol. 9, *Pragmatics*, New York: Academic Press, 221-43.(Discussion of views presented in §5.)

Kripke, S.A. (1979) 'Speaker Reference and Semantic Reference', in P.A. French, T.E. Uehling and H.K. Wettstein (eds) *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press, 6-27.(Discussion of views presented in §§2, 5.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.(See §2.)

Neale, S. (1990) *Descriptions*, Cambridge, MA: MIT Press.(Discussion of views presented in §§2-5. Extensive bibliography.)

Neale, S. (1994) 'Grammatical Form, Logical Form, and Incomplete Symbols', in A.D. Irvine and G.A. Wedeking (eds) *Russell and Analytic Philosophy*, Toronto, Ont.: University of Toronto Press, 97-139.(Discussion of views presented in §2.)

Russell, B.A.W. (1905) 'On Denoting', *Mind* 14: 479-93; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992.(Source of views presented in §2.)

Russell, B.A.W. (1911) 'Knowledge by Acquaintance and Knowledge by Description', in *Mysticism and Logic*, London: Routledge, 1986.(Source of views presented in §2.)

Sainsbury, R.M. (1979) *Russell*, London: Routledge & Kegan Paul.(Discussion of views presented in §§2, 5.)

Sellars, W.S. (1954) 'Presupposing', *Philosophical Review* 63: 197-215.(Discussion of views presented in §4.)

Smullyan, A.F. (1948) 'Modality and Description', *Journal of Symbolic Logic* 13: 31-7; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971, 35-43.(Discussion of views presented in §2.)

Strawson, P.F. (1950) 'On Referring', *Mind* 59: 320-44.(Source of views presented in §4.)

Strawson, P.F. (1954) 'Reply to Mr. Sellars', *Philosophical Review* 63: 216-31.(Source of views presented in §4.)

Strawson, P.F. (1964) 'Identifying Reference and Truth-Values', *Theoria* 30: 96-118.(Source of views presented in §4.)

Strawson, P.F. (1972) *Subject and Predicate in Logic and Grammar*, London: Methuen. (Source of views presented in §4.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7, esp. vol. 1, *14. (Source of views presented in §3.)

Desert and merit

The ideas of desert and merit are fundamental to the way we normally think about our personal relationships and our social institutions. We believe that people who perform good deeds and display admirable qualities deserve praise, honours and rewards, whereas people whose behaviour is anti-social deserve blame and punishment. We also think that justice is in large part a matter of people receiving the treatment that they deserve. But many philosophers have found these ways of thinking hard to justify. Why should people's past deeds determine how we should treat them in the future? Since we cannot see inside their heads, how can we ever know what people really deserve? How can we reconcile our belief that people must be responsible for their actions in order to deserve credit or blame with the determinist claim that all actions are in principle capable of being explained by causes over which we have no control?

1 Desert and merit

We use the concepts of desert and merit when we are thinking about how it is appropriate to treat people in a wide range of contexts. We say that Einstein deserved the Nobel prize for his work in theoretical physics, that Nixon deserved to be impeached for his role in the Watergate cover-up, that Jane deserves more thanks from her children than she is getting, that nurses deserve to have their pay increased. All of these claims have essentially the same structure (see Feinberg 1970): there is a subject, either an individual like Einstein or a set of people like the nurses, who are said to be deserving; there is a characteristic or an activity in virtue of which they are said to be deserving, which I shall call the *desert base* - Einstein's researches, or the work that nurses do; and there is something that is deserved, a burden or a benefit that it is appropriate or fitting for the subject to have on account of the desert base - the Nobel prize, a pay award (see Praise and blame).

Both the treatment deserved and the desert base vary a great deal from case to case, and mistakes in analysing the concept of desert often arise from considering only one kind of example. We can say in general that the characteristic or activity that forms the desert base must be something that we value positively or, in the case of deserved blame or punishment, negatively (see Values). But the valuing can be of many different kinds. Sometimes people are said to deserve rewards for activities that are morally valuable - we might say, for instance, that the charitable work done by a great philanthropist deserves to be honoured. In other cases - such as those of athletes or scientists who are awarded prizes - we value someone's accomplishments without giving them moral credit. To suppose that all desert must be moral desert is to shrink the range of possible desert bases unwarrantably.

The concepts of desert and merit are linked closely, and some languages have only a single word where English has two. The main difference is that we tend to use desert when talking about someone's performance, and merit when talking about their personal capacities and abilities (see Lucas 1980). Thus when we say that appointments should be made on the basis of merit, we mean that the candidate who shows the greatest capacity to fill the position should be awarded it. On the other hand, when somebody has performed a courageous act or written an excellent essay, we usually say that they deserve something in return, a decoration or a grade. Desert also conveys the sense that the subject is responsible for what he does, whereas in pre-modern systems of thought, including that of ancient Greece, people were judged meritorious according to the quality of their actions, and with little regard to individual responsibility as we understand it (see Adkins 1960).

2 Desert and personal responsibility

What are the implications of saying that people deserve rewards, punishments and other modes of treatment only when they are responsible for the performance that forms the basis of desert (see Responsibility)? It means first of all that they must have intended to do what they did. If I gun down a desperado thinking that he is the sheriff, I do not deserve the price that has been put on his head. Moreover my deserts are lessened or even removed entirely to the extent that my performance depends on luck. I may win a competition through a lucky series of guesses, but the person who really deserved to win was the person who knew the answers all along. This link between desert and responsibility is particularly tight when punishment is at stake (see Crime and punishment). When legal systems punish people for occurrences they could have done nothing to prevent, we regard these punishments as undeserved and unfair. The driver who runs over a child even though he was taking good care and could have done

nothing to avoid the accident does not deserve blame or punishment, but sympathy.

Problems arise, however, when we reflect that people’s performances depend in nearly every case not only on their intentions and efforts but on circumstances beyond their control, including the natural talents that they bring to the performance. Einstein’s achievement in developing the theory of relativity required considerable innate mathematical ability as well as a stimulating early education in science. Many philosophers have concluded that genuine desert must be based only on those parts of people’s actions or characteristics for which they are responsible, for instance the efforts they make or the talents they have chosen to develop. Whatever stems from natural causes beyond their control is irrelevant.

But this conclusion, too, faces difficulties. For willingness to make an effort, or a decision to develop a talent, is likely to be affected by initial circumstances and natural abilities. Would Einstein have decided to work on the theory of relativity if he had not known that he had a talent for theoretical physics? As John Rawls puts it, the character that enables someone to make an effort ‘depends in large part upon fortunate family and social circumstances for which he can claim no credit’ (1971: 104).

It seems, then, that the notion of desert collapses if we insist that a deserving person must be personally responsible for all of the conditions necessary for their performance. To keep the link between desert and personal responsibility we must substitute a weaker understanding of responsibility, such that a person can be responsible for a performance even though that performance depends upon circumstances and capacities for which they were not responsible. So long as they intended to perform the activity, and were sufficiently in control of it (it wasn’t a lucky accident), they may deserve benefits on account of it, even though their performance depended on natural capacities that others might lack (see Moral luck).

3 Critics of desert

An alternative response is to insist that desert requires responsibility in the stronger sense, and then to say that, because no one is ever fully responsible in this sense for what they do, the concept of desert as ordinarily understood should be scrapped. This is often thought to follow from determinism (see Free will §1). If every human action and decision has ultimately a set of causes beyond the agent, the idea of desert must either be abandoned or be given a new meaning. Utilitarians have argued that to say of someone that they deserve something is only to say that it is socially useful to reward them for performing a certain service (see Utilitarianism).

The same conclusion - that desert is at best a secondary idea - has been arrived at in other ways too. As we have seen, desert as ordinarily understood very often looks to the past as a reason for assigning a benefit or burden now: criminals deserve punishment for their past misdeeds; Einstein deservedly won the Nobel prize in 1921 for work done fifteen years earlier. But this appears to collide with the principle that in deciding how to act, we should consider only the future effects of our actions, which many have taken to be a requirement of practical rationality. On this view, we must reinterpret desert in forward-looking terms - we must estimate people’s deserts by considering the consequences of different ways of allocating our resources, rather than looking back at what they have already done - if we are to have any good reason to give people what they deserve.

A third line of attack is to suggest that we cannot establish what people deserve until we know which institutions are fair. It is an illusion to suppose that people have deserts in advance of social institutions like competitions, honours systems, or property systems, which are conventional in nature (see Property). When we say that people deserve things, what we mean is that they are legitimately entitled to them; these are the benefits they would expect to get, or the burdens they would expect to suffer, if the institutions were fair and working as they should. Once again the effect is to relegate desert to the status of a secondary moral idea.

We can now see what is required in order to defend something like our ordinary concept of desert. First, we must either reject determinism, or else argue that determinism is compatible with personal responsibility in a sense strong enough to support desert. Second, we must defend what Sher (1987) has called ‘antecedentialism’, the view that an act or occurrence may be called for because it would stand in a certain relation to some antecedent act or event. Third, we must show that our judgments about desert are sufficiently independent of institutional conventions for desert to function as a critical standard which we can appeal to when deciding which institutions to adopt.

4 Desert and justice

Justice often requires us to give people what they deserve (see Justice). Where it does not, this is usually because it is beyond our power to bring about the deserved outcome, or because attempting to do so would have unjust side effects (we cannot ensure that the best athlete wins the race or that good people find the partners they deserve). But if a society allows women to be paid less than men for doing the same job, or punishes lesser crimes more severely than greater crimes, it is to that extent unjust, and we have an obligation to change it.

The assumptions underlying this claim have been challenged both by conservatives and by radicals, who converge in thinking that 'desert' and 'merit' are too subjective to stand as independent criteria of distribution. Hume argued that merit might seem the most natural basis of distribution but 'so great is the uncertainty of merit, both from its natural obscurity, and from the self-conceit of each individual, that no determinate rule of conduct would ever result from it' (1751: 193). More recently, Hayek (1960) has argued along similar lines (see Hume, D.; Hayek, F.A.). They both make two errors. One is to suppose that desert must mean moral desert; the difficulty of making objective estimates of moral deservingness does not imply that other measures of desert - such as economic productivity - are equally obscure. The other error is to suppose that desert must supplant all other forms of entitlement, rather than serving as a critical standard by which existing entitlements can be judged, and new ones created.

Radicals claim that standards of desert and merit are socially constructed to support existing practices and institutions (see Young 1990). We can see this claim deployed in debates over affirmative action. When it is argued that giving preference to women or racial minorities in the allocation of jobs or college places conflicts with the principle of merit, the critics reply that 'merit' here merely refers to qualities which existing (male, white) society has chosen to designate as relevant. This criticism is likely to backfire, however, because if merit is abandoned altogether, it appears to open the door to practices that discriminate more openly and directly against disadvantaged groups (see Discrimination).

Justice is more than the requital of desert. It sometimes demands rule-following and respect for existing entitlement, sometimes equal treatment, or distribution according to need (see Equality; Needs and interests). But to divorce justice and desert entirely would be to abandon some of our most deeply held convictions for reasons which may be less than compelling.

See also: Rectification and remainders; Moral sentiments

DAVID MILLER

References and further reading

Adkins, A.W.H. (1960) *Merit and Responsibility*, Oxford: Oxford University Press. (Explores ideas of merit in classical Greek thought.)

Feinberg, J. (1970) 'Justice and Personal Desert' in *Doing and Deserving*, Princeton, NJ: Princeton University Press. (A valuable analysis drawing attention to the variety of claims we make about people's deserts.)

Hayek, F.A. (1960) *The Constitution of Liberty*, London: Routledge & Kegan Paul, ch. 6. (Argues that distribution according to merit cannot be reconciled with a market economy.)

Honderich, T. (1990) *The Consequences of Determinism: A Theory of Determinism*, Oxford: Clarendon Press, vol. 2, ch. 4. (Considers carefully the implications of determinism for desert as a ground of punishment, and other kinds of desert.)

Hume, D. (1751) *An Enquiry Concerning the Principles of Morals, in Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 3rd edn, 1975. (Argues that merit has no place among the rules of justice, which are merely useful conventions.)

Lucas, J. (1980) *On Justice*, Oxford: Clarendon Press. (Defends desert, and claims that it is central to our thinking about justice.)

MacIntyre, A. (1981) *After Virtue*, London: Duckworth. (Claims that justice must be understood in terms of desert, but that modern societies lack the practices that sustain desert.)

Rawls, J. (1971) *A Theory Of Justice*, Cambridge, MA: Harvard University Press, sections 17, 48. (Dismisses desert as basic moral idea, and reinterprets it as legitimate entitlement.)

Sher, G. (1987) *Desert*, Princeton, NJ: Princeton University Press. (The best recent discussion of desert, exploring

possible justifications for a wide range of desert claims.)
Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan; 7th edn, 1907, book III, ch. 5, book IV, ch. 3. (A classic review of ordinary thinking about desert, together with proposals for reconstructing the concept along utilitarian lines.)
Young, I.M. (1990) 'The Myth of Merit' in *Justice and the Politics of Difference*, Princeton, NJ: Princeton University Press. (Radical attack on the principle of merit.)

Desgabets, Robert (1610-78)

Although he is now little known, Desgabets was an important seventeenth-century French philosopher, theologian, scientist and mathematician. An early defender of Cartesian philosophy, his physical explication of transubstantiation created such an uproar that he complied with a public order and renounced his views. He defended the essential union and interaction of soul and body, and the free creation of the eternal truths. The latter view led him to an empiricist epistemology: all ideas have a sensory basis and are essentially related to existing objects. Despite his originality, he is best known for his polemic with Foucher.

Robert Desgabets, Benedictine monk and early defender of Cartesian philosophy, was born in Ancemont, France. He was active in the Cartesian circles of Paris and Toulouse, well known to Clerselier, Rohault, Malebranche, Cordemoy, Retz and Régis. Despite his busy ecclesiastical career, he was *au courant* with the latest scientific discoveries. His pioneering work on blood transfusion inspired the French physician Jean Denis to attempt the procedure for the first time on a human subject (the subject lived, but the procedure was not successful). During his lifetime, only three small works were published: *'Discourse de la communication ou transfusion du sang' (Discourse on the communication or transfusion of blood)* (1668), *Considérations sur l'état présent de la controverse touchant le Très Saint-Sacrement de l'autel (Considerations on the present state of the controversy concerning the Holy Sacrament)* (1671) and *Critique de la Critique de la Recherche de la vérité (Critique of the Critique of the Search After Truth)* (1675). However, many of his philosophical manuscripts are known to have circulated widely in Cartesian circles; Régis was greatly influenced by his ideas.

Desgabets' writings deal with a broad range of philosophical, scientific and theological issues. His *Traité de l'indéfectibilité des créatures (Treatise on the indefectibility of created beings)* (*c*.1654), contains 'the indefectibility thesis', his most original thesis, drawn from Descartes' doctrine that God is the free and indifferent cause of everything, including the eternal truths and essences. Desgabets reasoned that if everything was created by an immutable, indivisible act of God's will then substances were created in their essence as well as their existence. There is no priority of intelligible essence over material existence since both are equally contingent on the omnipotent will of God. Thus, every substance is equally indestructible, necessary and real in both its essence and existence. The epistemological upshot of favouring divine omnipotence over divine wisdom is empiricism, since if God's will is not constrained by rational considerations then any a priori basis for knowledge is upset, leaving experience as the only means for knowing what God actually created. Desgabets thus lays the metaphysical foundation for the Cartesian empiricism which he subsequently developed over twenty years.

His *Le guide de la raison naturelle (Guide to Natural Reason)* (1671) is an examination of the principles and limits of reason. On the assumption that substances are created, he argues that conceivability itself has its source in created substance. He concludes that to conceive of an object is to have an idea of an actually existing object. It is the essential intentionality of ideas which ensures that simple conceptions are always true.

His important *Supplément à la philosophie de M. Descartes (Supplement to the philosophy of Descartes)* (1675) is an examination of the basic principles of Cartesianism, with corrections of Descartes' alleged errors. In this work his empiricism is fully developed. From Descartes' doctrines of soul-body union and interaction, he derives the empiricist claims that all that is in the soul passes through the senses, that all thoughts are accompanied by sensible signs or words, and that our thoughts have succession and duration which are inseparable from bodily movement. Even the *Cogito* involves the succession of ideas, which itself presupposes the idea of extension or body. Thus, the mind is not better known than body. Nor is the first principle of philosophy the *Cogito*, but rather, that all simple conceptions are always true and conform to their objects.

In physics, Desgabets' fidelity to Descartes' ideas is readily evident. He praises Descartes' system of physics against that of Gassendi, arguing that the essence of matter is extension, that sensible qualities belong to mind, and that a void is impossible. Moreover, he rejects Cordemoy's atomist revision of Cartesianism as one based on a false conception of the continuum and infinite divisibility: if atoms have extension, then they have parts, and if they have parts then they are divisible.

Critique of the Critique of the Search After Truth is a polemical defence of Malebranche against the sceptic

Foucher. Desgabets challenges Foucher's 'most odious' doctrine that ideas must resemble their objects in order to represent them. Both Malebranche and Desgabets follow Descartes in the belief that resemblance is not necessary for representation. However, Desgabets opts for a different account to that of Malebranche, explicating the relation between ideas and their objects in terms of an 'essential and intentional relation' (see Malebranche, N. §3).

Desgabets did not believe that reason and faith are opposed, but thought that theological revelation should be made consistent with philosophy. In *Considerations on the present state of the controversy concerning the Holy Sacrament* he argues that the new philosophical reformation is a product of joining mathematics to philosophy, which has enabled the explication of theological mysteries such as that of the Eucharist. That matter and quantity are the same thing means that there cannot be quantity without substance. This entails that if the body of Christ is present in the Host then it is actually present, that is, locally extended. Not surprisingly, Desgabets' views were judged by fellow ecclesiastics - notably Arnauld - to be completely against tradition. After a public trial and retraction, which resulted in a future ban on his writings, he retreated from public life to a monastery at Breuil, France. Fortunately, under the protection of Cardinal de Retz, he produced some of his best philosophical writings and participated in Cartesian conferences at Commercy.

Desgabets' system is erected on two basic and connected theses: the free creation of the eternal truths and essences, and the principle that simple conceptions contain a demonstrative proof of the reality and existence of their objects. His significance lies in his development and defence of a form of empiricism that, despite its anti-rationalist tendencies, has a clear basis in the work of Descartes himself and constitutes a respectable strain of Cartesianism.

See also: Descartes, R.

PATRICIA A. EASTON

List of works

Desgabets, R. (1983) *Oeuvres philosophiques inédites*, Analecta Cartesiana 2, ed. J. Beaude, intro. by G. Rodis-Lewis, Amsterdam: Quadratures.(An essential but not widely available collection of Desgabets' philosophical writings, including works previously available only in manuscript. It includes *Traité de l'indéfectibilité des créatures (Treatise on the indefectibility of created beings)* (1653-74), *Le guide de la raison naturelle (The Guide to Natural Reason)* (1671) and *Supplément à la philosophie de M. Descartes (Supplement to the Philosophy of Descartes)* (1675), as well as rich bio-bibliographical information superseding all previous compilations.)

Desgabets, R. (1668) *'Discourse de la communication ou transfusion du sang' (Discourse on the communication or transfusion of blood)*, Paris: J.B. Denis.(This scientific piece describes an apparatus and procedure for blood transfusion, invented by Desgabets.)

Desgabets, R. (1671) *Considérations sur l'état présent de la controverse touchant le Très Saint-Sacrement de l'autel (Considerations on the present state of the controversy concerning the Holy Sacrament)*, Holland.(This work, published anonymously, stirred up controversy for Cartesians because of its theologically sensitive thesis that the body of Christ is locally extended.)

Desgabets, R. (1675) *Critique de la Critique de la Recherche de la vérité (Critique of the Critique of the Search After Truth)*, Paris.(A response to Foucher's sceptical critique of the first edition of Malebranche's *Search After Truth* (1674) - although Malebranche did not welcome Desgabets' defence. Desgabets attacks Foucher's position that ideas must resemble their objects in order to represent them.)

References and further reading

Armogathe, J.-R. (1977) *Theologia Cartesiana: L'Explication physique de l'Euchariste chez Descartes et dom Desgabets (Cartesian theology: the physical explication of the Eucharist according to Descartes and dom Desgabets)*, The Hague: Martinus Nijhoff.(A detailed history of the controversy surrounding Cartesian explications of the mystery of the Eucharist, with an examination of Descartes' and Desgabets' physical explication of transubstantiation.)

Beaude, J. (1974) 'Desgabets et son oeuvre' (Desgabets and his work), *Revue de sythèse* 95:7-17.(A solid bio-bibliographical piece.)

Beaude, J. (1979) 'Cartésianisme et anticartésianisme de Desgabets' (The Cartesianism and anti-Cartesianism of Desgabets), *Studia Cartesiana 1*, Amsterdam: Quadratures, 1-24.(An insightful analysis of the Cartesian and anti-Cartesian tendencies of Desgabets' thought.)

Beaude, J. (1980) 'Le Guide de la raison naturelle dans l'oeuvre de Desgabets' (The guide to natural reason in the writings of Desgabets) *Recherches sur le XVIIe siècle IV*, Paris: Centre National de la Recherche scientifique. (This work, establishing Desgabets as author of the unpublished manuscript *The Guide of Natural Reason* (1671), includes a useful discussion of some of Desgabets' primary precepts.)

Cousin, V. (1852) *Fragments de philosophie cartésienne (Fragments of the Cartesian philosophy)*, Paris: Didier; repr. Geneva: Slatkine Reprints, 1970.(Volume 3 includes selections from unpublished exchanges between Retz, Malebranche and Corbinelli on Desgabets. Due to incompleteness and emphasis on Retz, these fragments do not serve as an introduction to Desgabets, but the work includes rare manuscript material and a good bibliography.)

Lemaire, P. (1901) *Le Cartésianisme chez les Bénédictins: Dom Robert Desgabets son système, son influence et son école, (Cartesianism according to the Benedictines: Dom Robert Desgabets, his system, his influence, and his school)* Paris: Alcan. (Despite some bio-bibliographical errors and unreliable inferences about the significance of Desgabets' ideas, this remains an important full-length study with many quotations from unpublished manuscripts and a comprehensive bibliography.)

Robinet, A. (1974) 'Dom Robert Desgabets, le conflit philosophique avec Malebranche et son l'oeuvre métaphysique' (Dom Robert Desgabets, his philosophical conflict with Malebranche and his metaphysical work), *Journée Desgabets, Revue de synthèse* 95: 65-83.(This article examines Desgabets' influence on Malebranche, found in Malebranche's views on occasionalism, the non-materiality of thoughts and the nature of the eternal truths.)

Rodis-Lewis, G. (1993) 'Der Cartesianismus in Frankreich' (The diffusion of Cartesianism in France), in J-P. Schobinger ed. *Grundriss der Geschichte der Philosophie, Die Philosophie des 17. Jahrhunderts*, Band II, Basel/Stuttgart, 398-445.(One section of this chapter is devoted to Desgabets, including a summary of his main philosophical theses.)

Rodis-Lewis, G. (1981) 'Polémiques sur la création des possibles et sur l'impossible dans l'école cartésienne' (Polemics in the Cartesian School on the creation of possibles and on the impossible), *Studia Cartesiana 2*, Amsterdam: Quadratures, 105-23.(Discusses Desgabets' role in the Cartesian debate on Descartes' doctrine of the free creation of essences and truths.)

Rodis-Lewis, G. (1966) *The Downfall of Cartesianism 1673-1712,* The Hague: Martinus Nijhoff.(Includes an examination of Foucher's critique of Desgabets' account of ideas as intentional representations.)

Determinism and indeterminism

Over the centuries, the doctrine of determinism has been understood, and assessed, in different ways. Since the seventeenth century, it has been commonly understood as the doctrine that every event has a cause; or as the predictability, in principle, of the entire future. To assess the truth of determinism, so understood, philosophers have often looked to physical science; they have assumed that their current best physical theory is their best guide to the truth of determinism. It seems that most have believed that classical physics, especially Newton's physics, is deterministic. And in this century, most have believed that quantum theory is indeterministic. Since quantum theory has superseded classical physics, philosophers have typically come to the tentative conclusion that determinism is false.

In fact, these impressions are badly misleading. The above formulations of determinism are unsatisfactory. Once we use a better formulation, we see that there is a large gap between the determinism of a given physical theory, and the bolder, vague idea that motivated the traditional formulations: the idea that the world in itself is deterministic. Admittedly, one can make sense of this idea by adopting a sufficiently bold metaphysics; but it cannot be made sense of just by considering determinism for physical theories.

As regards physical theories, the traditional impression is again misleading. Which theories are deterministic turns out to be a subtle and complicated matter, with many open questions. But broadly speaking, it turns out that much of classical physics, even much of Newton's physics, is indeterministic. Furthermore, the alleged indeterminism of quantum theory is very controversial: it enters, if at all, only in quantum theory's account of measurement processes, an account which remains the most controversial part of the theory.

1 Consensus

Over the centuries, the doctrine of determinism has been understood, and its truth or falsity assessed, in different ways. (We follow the nearly universal practice of taking 'indeterminism' as simply the negation of determinism; so our discussion can focus on determinism.) Since the rise of modern science in the seventeenth century, it has been commonly understood as the 'law of universal causation', that every event has a cause; or as the predictability, in principle, of all of the future, given full knowledge of the present.

What evidence a philosopher takes to count for or against this doctrine varies immensely from one philosopher to another, according to their philosophical project. For example, many assess determinism in the light of their opinions about such metaphysical topics as free will or God (see Free will; Omniscience §3). Others see connections between determinism and broadly logical topics about time (see Many-valued logics, philosophical issues in §1). But this entry is restricted to formulating determinism, and assessing whether it is true, by considering the deliverances of physical theory. Of course, this restriction does not mean that our discussion only applies to wholly secular philosophers: many theistic philosophers, for example Kant, have discussed determinism in terms of the physics of their day. Some have even endorsed it, as part of their philosophy of nature; again Kant provides the outstanding example (see Kant, I. §7).

Making this restriction, we can say that since the seventeenth century, philosophers have typically taken their current best physical theory as their guide to the truth of determinism. And during the second half of this period, there has been a remarkable consensus about what that best theory is, and what it indicates about determinism.

During the nineteenth century, most of the educated public took Newtonian mechanics, and especially the Newtonian theory of gravitation, to be their best physical theory. Indeed, many took it to be an unrevisable foundation for physical theorizing. At its simplest, the idea was that Newton had laid down in his mechanics a schema for the mechanical explanation of the physical world. The schema was encapsulated in Newton's second law, that the force on a body is equal to its mass times its acceleration. Knowing the force and the mass, one could calculate the acceleration, and thus how the body moved. So to get a mechanical explanation of a given phenomenon, one had only to 'fill in the schema' by finding the forces involved. The paradigm case was of course gravitation; here Newton himself had discovered the nature of the force, and had calculated with stunning success how the planets and other celestial bodies move. Accordingly, many believed that Newtonian mechanics could in principle describe any phenomenon, perhaps by postulating strange forces (see Mechanics, classical §2).

They also believed that all the theories that could arise by thus filling in the schema would be deterministic; for the motion of a body would be determined by the forces on it (together with its initial position and velocity). The *locus classicus* for this view is a passage by Laplace, in which he not only states the doctrine that Newtonian mechanics is deterministic, but also provides formulations of determinism - first, in terms of causation, and then in terms of prediction; we shall later have reason to criticize both the doctrine and the formulations.

> We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it - an intelligence sufficiently vast to submit these data to analysis - it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. The human mind offers, in the perfection which it has been able to give to astronomy, a feeble idea of such an intelligence.
>
> (Laplace [1820] 1951: 4)

During the twentieth century, quantum theory and relativity theory became our best physical theories; by 1930, they had superseded classical physics. Since these theories are comparatively new and technically demanding, they have not become part of 'educated common sense' in the way in which Newtonian theories did (at least eventually, say by the mid-nineteenth century). But most philosophers who have addressed the topic have concluded that while relativity theory is deterministic, quantum theory is indeterministic. Indeterminism is taken to be the lesson of the much-cited uncertainty principle. This conclusion also has authority on its side: the great majority of the discoverers of quantum theory endorse it. So philosophers have typically come to the tentative conclusion that determinism is false.

2 Controversy

But this consensus has been badly misleading. First of all, formulations of determinism in terms of causation or predictability are unsatisfactory. And once we use a correct formulation, it turns out that much of classical physics, even much Newtonian physics, is indeterministic; and that parts of relativity theory are indeterministic (owing to singularities). Furthermore, the alleged indeterminism of quantum theory is very controversial - for it enters only, if at all, in quantum theory's account of measurement processes, an account which remains the most controversial part of the theory.

Formulations of determinism in terms of causation or predictability are unsatisfactory, precisely because of philosophers' interest in assessing determinism by considering physical theories. That interest means that determinism should be formulated in terms that are clearly related to such theories. But 'event', 'causation' and 'prediction' are vague and controversial notions, and are not used (at least not univocally) in most physical theories. Prediction has the further defect of being an epistemological notion - hence Laplace's appeal to an 'intelligence'; while the intuitive idea of determinism concerns the ontology or 'world-picture' of a given theory (see Causation; Events).

Fortunately, the intuitive idea of determinism can be formulated quite precisely, without these notions. The key idea is that determinism is a property of a theory. Imagine a theory that ascribes properties to objects of a certain kind, and claims that the sequence through time of any such object's properties satisfies certain regularities. In physics, such objects are usually called 'systems'; the properties are called 'states'; and the regularities are called 'the laws of the theory'. Then we say that the theory is deterministic if and only if for any two such systems: if they are in exactly the same state as one another at a given time, then according to the theory (for example, its laws about the evolution of states over time), they will at all future times be in the same state as one another. (Montague 1974, pioneered this kind of formulation.)

We can make determinism even more precise in the context of specific physical theories, by using their notions of system, state and law (regularity). But the classification of theories as deterministic or indeterministic is not completely automatic. For the notions of system, and so on, are often not precise in a physical theory as usually formulated. So various different formulations of determinism are often in principle legitimate for a given theory; and there is room for judgment about which formulation is interpretatively best.

However, it is well-established that the main conclusions are as reported above. The philosopher who has done most to classify physical theories in this way is Earman (1986), who upholds these main conclusions. We shall just briefly support these conclusions with two points that his book does not cover. (The details of the above formulation of determinism will not be needed for these points.)

First, much Newtonian physics is indeterministic. Indeed, indeterminism is lurking in the paradigm case discussed by Laplace: point-masses influenced only by their mutual gravitational attraction, as described by Newton’s law of gravitation.

But surely the motion of each point-mass is determined by thus forces on it, in this setting the gravitational force (together with its initial position and velocity)? (See Mechanics, classical §2.) Indeed it is, *locally*. That is: given the initial positions, velocities and forces, the motion of each point-mass is determined, for *some* interval of time extending into the future. But it might be a very short interval. (Technically, the equations of the theory have a unique solution for some interval of time, perhaps a very short one.) Furthermore, as time goes on, the interval of time for which there is such a solution might get shorter, shrinking to zero, in such a way that after some period of time, the solution does not exist any more. In effect, determinism might hold locally in time, and yet break down globally.

One way this might happen is by collisions: in general, Newtonian mechanics is silent about what would happen after two or more point-masses collide. But more interestingly, it seems that it might also happen *without* collisions. Thus it was conjectured in 1897 that one might somehow arrange for one of the point-masses to accelerate in a given spatial direction, ever more rapidly and at so great a rate, during a period of time, in such a way that it does not exist in space at the end of the period! By that time, it has disappeared to ‘spatial infinity’. (The source for the energy needed by the acceleration is the infinite potential well-associated with Newton’s inverse-square law of gravitation.) That this can indeed happen with just Newtonian gravity was finally proved true by Xia in 1992 (using a system of five point-masses). So now we know that, even setting aside collisions, Laplace’s vision of Newtonian determinism is only valid for local intervals of time.

Second, quantum theory can be interpreted as being deterministic. De Broglie and Bohm showed that such an interpretation of elementary quantum theory is possible, despite the alleged proofs that it was impossible (given in the 1930s by some of the discoverers of quantum theory). The basic idea is that a quantum system consists of both a wave and a particle. The wave evolves deterministically over time according to the fundamental equation of quantum theory (the Schrödinger equation) and it determines the particle’s motion, which therefore also moves deterministically, given the wave (hence this interpretation is also called the pilot wave interpretation). This contrasts with the orthodox interpretation. Roughly speaking, the orthodox interpretation accepts only the wave, and accommodates particle-like phenomena by having the wave evolve indeterministically (violating the Schrödinger equation) during processes of measurement (see Quantum mechanics, interpretation of §3; Quantum measurement problem). In recent years, the de Broglie-Bohm approach has been greatly developed so as to yield a deterministic interpretation of more and more of advanced quantum theory, including quantum field theory (see Cushing 1994). Suffice it to say, a deterministic interpretation of quantum theory is entirely coherent.

There remain two other controversial matters; which return us to general metaphysics and philosophy of science. First, should we apply the idea of determinism, as we have formulated it, to theories of the whole universe, that is, cosmologies? If so, then the ‘systems’ in question will be universes or ‘possible worlds’, that is, total possible courses of history. So one will in general not require that the systems whose states one compares must lie in the same possible world (see Possible worlds).

Second, should we accept the idea, for a given kind of system, of a complete theory, a theory that in some sense describes the whole truth about the systems? Some philosophers hold that this idea is incoherent: at least if it is filled out as allowing that such a theory is never formulated by humans; or at least if it allows that humans might be in principle incapable of formulating such a theory (see Scientific realism and antirealism §1). But if we accept some version of this idea, then we can reasonably talk of the systems, or perhaps the kind of system, being deterministic: namely, if and only if the systems’ final theory is deterministic.

So the cautious answer to these questions is No. To answer Yes is to be bold: (some would say, incoherent). In particular, if we answer Yes to both questions, we are in effect accepting that it is meaningful to talk of the whole

universe being deterministic. For we are accepting the idea of a complete theory of a given possible world (a total possible course of history). So we can reasonably call this theory the 'theory of the world', and its general propositions 'the laws of nature' (see Laws, natural §1). (Again, humans are unlikely to have much idea of this theory or its laws.) And then we can say that the given world is deterministic if and only if its theory is. That is, the given world is deterministic if and only if any two worlds, obeying this theory, that have exactly the same state (in the sense of this theory) as one another at a given time also have exactly the same state at all later times.

The rest of this entry is restricted to discussing determinism for given physical theories. It thereby answers No to the second question; and, cosmological theories apart, it also answers No to the first question. But before embarking on this cautious strategy, we should briefly note that, historically, the bold (perhaps incoherent) idea of the entire world being deterministic, irrespective of any theory, has been very important; it has been the focus of countless philosophers' discussions of determinism (both for and against it).

3 Defining determinism

In §2 we said that a theory is deterministic if and only if for any two systems of the kind described by the theory: if they are in exactly the same state as one another at a given time, then according to the theory, they will at all future times be in the same state as one another. But this formulation is still rough.

The main problem is that, whatever theory one considers, its systems are continually interacting with their environment: as physicists put it, no system is 'completely isolated'. For example, each system feels the gravitational pull of other objects. These interactions make determinism, as just formulated, an impossibly tall order. For, first, it will be very rare for two systems to be in exactly the same state at a given time. And even if they are, it will be virtually impossible that their subsequent interactions with their environments match so exactly that they are also in the same state as one another at all future times. But surely, determinism should not be so formulated that it will fail because of the vagaries of interactions with other systems: whether it fails or holds should be a matter internal to the theory considered.

The remedy is clear enough. To set aside such interactions, we need to formulate determinism in terms of completely isolated systems. But we cannot just think in terms of two systems in an otherwise empty universe. For in general the theory will take these two systems to interact with each other; so that again determinism can fail in a spurious way. That is to say, even if we suppose that at a given time the two systems are in the same state, at some future time they may well not be: their interaction, as described by the theory, might lead to their states differing. (The problem is of course aggravated if we think of the systems as also interacting in other ways, not described by the theory.)

To avoid this kind of spurious failure of determinism, we need to think of the theory as describing single completely isolated systems, each one alone in its universe. Let us say that a sequence of states for such a single system, that conforms to the laws of the theory, is a *model* of the theory. So a model contains a system of the theory's kind, undergoing a history allowed by the theory: the model is a 'toy universe' or 'toy possible world', according to the theory. (This use of 'model' is common in general philosophy of science. In particular, it is often useful to consider a scientific theory as the class of its models in this sense, rather than in the traditional manner of logicians - as a set of sentences closed under deduction; see Models; Theories, scientific.) Using this notion of model, we can give a better definition of determinism, which avoids the problem of interactions. We say that a theory is deterministic if and only if: any two of its models that agree at a time *t* on the state of their objects, also agree at all times future to *t*.

(This definition returns us to the first of the two questions at the end of §2: namely, should we apply the idea of determinism to theories of the whole universe, for example cosmologies? We now see that our strategy for avoiding spurious violations of determinism, due to interactions between systems, commits us to answering Yes to this question. For by taking a model of any theory to describe a single completely isolated system, alone in its universe, we are in a sense treating any theory as a cosmology. But since each of a theory's possible universes contains just one system of the kind treated by the theory, it is typically a humble, even a dull, cosmology!)

This definition is still a bit vague: precisely how should we understand a single time *t* in two models, and two models 'agreeing' on their states at *t*? The answers to these questions lie in the idea of isomorphism of models, or parts of models; in the usual sense used by logicians. (There is no need for a 'meta-time' outside the two models,

in terms of which their time series can be compared: thank goodness, since that would be very questionable!). Thus we can speak of an ‘instantaneous slice’ of one model (that is, the part describing the system at a single time) being isomorphic to an instantaneous slice of another model. And similarly, we can speak of isomorphism of ‘final segments’ of two models: that is, isomorphism of parts of two models, each part describing the system at all times future to some time within the model. Determinism is then a matter of isomorphic instantaneous slices implying that the corresponding final segments are isomorphic (where ‘corresponding’ means ‘starting at the time of the instantaneous slice’). That is: we say that a theory is deterministic if, and only if: for any two of its models, if they have instantaneous slices that are isomorphic, then the corresponding final segments are also isomorphic.

4 The notion of state

To a philosopher, our definition of determinism looks very formal. And indeed, it is closely related to purely mathematical questions. For a physical theory is often presented as a set of equations, so-called ‘differential equations’, governing how physical magnitudes (for example, numerically measurable quantities like distance, energy and so on) change with time, given their values at an initial time. Our definition then corresponds to such a set of equations having a unique solution for future times, given the values at the initial time; and whether a set of differential equations has a unique solution (for given initial values) is a purely mathematical property of the set.

But we should beware of identifying determinism with this purely mathematical property: there are conceptual, indeed metaphysical, matters behind the mathematics. The reason lies in the notion of state. We have taken states to be simply the properties ascribed by a theory to objects of a certain kind (the theory’s ‘systems’); and so as varying from one theory to another. But there are two general features which the notion of state needs to have if our definition of determinism is to be sure of capturing the intuitive idea. These features are vague, and cannot be formalized: but without them, there is a threat that our definition will be intuitively too weak.

First, states need to be intrinsic properties. It is notoriously hard to say exactly what is meant by ‘intrinsic’, but the idea is to rule out properties which might code information about how the future just happens to go, and thus support a spurious determinism. Thus, to take an everyday example, ‘Fred was mortally wounded at noon’ implies that Fred later dies. But the property ascribed at noon is clearly extrinsic: it ‘looks ahead’ to the future. And so this implication does not show that there is any genuine determinism about the processes that led to Fred’s later death.

Second, states need to be ‘maximal’, that is they need to be the logically strongest consistent properties the theory can express (compatibly with their being intrinsic). For in an intuitively indeterministic theory, there might well be some properties (typically, logically weak ones) such that models agreeing on these properties at one time implies their agreeing on them at all later times.

Do physical theories’ notions of state have these two features? The question is vague because there is no agreed analysis of the ideas of an intrinsic, or a maximal, property. Perhaps ‘maximal’ can be readily enough analysed in terms of logical strength, as just hinted. But it is notoriously hard to analyse ‘intrinsic’. But, by and large, the answer to this question is surely Yes. Physics texts typically define, or gloss, ‘state’ and similar words as a system’s maximal (or ‘complete’) set of intrinsic (or ‘possessed’) properties; and in philosophy, the most commonly cited examples of intrinsic properties are the magnitudes figuring in the states of familiar physical theories, such as mass or electric charge. So it seems there is no widespread problem of spurious satisfactions of determinism.

But although there is not a problem, the need for these features brings out the main point: determinism is not a formal feature of a set of equations. Indeed, there are many examples of a set of differential equations which can be interpreted as a deterministic theory, or as an indeterministic theory, depending on the notion of state used to interpret the equations.

The idea of states as intrinsic also brings out two other points, one philosophical and one technical. The philosophical point concerns the ideas at the end of §2 about laws of nature, and the whole universe (as against a given theory) being deterministic. One of course expects that making sense of the idea of laws of nature will involve the theory of properties. But we now see a more specific point: that making sense of the universe being deterministic will involve the general analysis of ‘intrinsic’.

The technical point concerns theories that treat all the states up to the given time, taken together, as contributing to

determining the future states. (There are a few such theories. It does not matter here whether all these earlier states taken together do determine all the future states: as, one might say, whether there is determinism of the future by the whole past.) At first sight, it looks as if such theories add to the usual intrinsic notion of state, a highly extrinsic notion - for which the state at a given time encodes some of the information in all earlier intrinsic states. What is going on?

In fact, in all such theories (so far as this author knows) the extrinsic notion is a technical convenience, rather than a new notion of state. The theory refers to the arbitrarily distant past (typically in some time-integral from minus infinity to the given time) just as a mathematically tractable way of stating information about the state at the given time, information that contributes to the future development of the system. (Wanting to state this information of course reflects the idea of states as maximal.) For example: in statistical physics, some such time-integrals define correlations in the present state; and in theories that study systems interacting with their environment, the past states of the system yield useful information about the present influence of the environment (which is otherwise not represented in the formalism).

Note that this explanation accords with a familiar tenet about causation: that past states influence the future, but only via their influence on the current state - there is thus no ‘action at a temporal distance’. This tenet is widely held by philosophers; and to the extent that one can talk about causation in physical theories, it is upheld in physics. This is especially true of relativity theory; for relativity both unifies space and time, and upholds the principle of contact-action (see Relativity theory, philosophical significance of §3; Spacetime). The tenet is also closely related to a very common property (being Markovian) of probabilistic theories, both in physics and beyond. Indeed, according to some probabilistic theories of causation, the tenet is equivalent to the theory being Markovian.

Theories that refer to past states are relevant to our final topic: the fact (mentioned in §2) that there are some uncontroversial variations on our definition of determinism. So far we have for simplicity assumed that there is a single intuitive idea of determinism: the idea of the present state determining future states. But as we have just seen, there is an analogous idea: that all the states up to the present, taken together, determine future states. It just so happens that (using an intrinsic, maximal notion of state!) this idea is not obeyed in known physical theories: they have no ‘action at a temporal distance’. But that is no reason to deny to the idea the name ‘determinism’; or, more clearly, ‘determinism of the future by the past’ (rather than by the present).

This point is reinforced by other analogous ideas, ideas which *are* obeyed in known physical theories. Thus in general relativity, and in quantum field theory, diverse technical reasons make it much easier to define a state on an interval of time (called a ‘sandwich’ of spacetime!) than at an instant of time (a ‘slice’ of spacetime). There is no hint here of action at a temporal distance: the interval can be arbitrarily short - it is just that for technical reasons it must have some duration. But such states prompt rather different definitions of determinism, requiring (roughly speaking) that for any interval, no matter how short, states to the future of that interval are determined by the state on it. And these definitions are often satisfied.

One can instead strengthen the definition of determinism, requiring the state at the given time to determine not only future states, but also past states. Many important physical theories have a property called ‘time-symmetry’ or ‘time-reversal invariance’, which implies that they satisfy this stronger definition, if they satisfy our first one. A famous example is Newtonian mechanics. Indeed, it may well be that Laplace had in mind this point (rather than just the intelligence having a memory), when he said in the quotation above ‘as the past’ (see Mechanics, classical; Thermodynamics §§4-5).

JEREMY BUTTERFIELD

References and further reading

Bohm, D. (1957) *Causality and Chance in Modern Physics*, London: Routledge & Kegan Paul.(Discussion of determinism and chance, by one of the two inventors of a deterministic interpretation of quantum theory; see end of §2.)

Cushing, J. (1994) *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*, Chicago, IL: University of Chicago Press.(Discussion of historical, philosophical and physical aspects of the de Broglie-Bohm interpretation.)

Earman, J. (1986) *A Primer on Determinism*, Dordrecht: Reidel.(The best single book on determinism in the

physical sciences. It wages a fine campaign against the false consensus, discussed in §§1, 2.)

Laplace, P. (1820) Essai philosophique sur les probabilités; forming the introduction to his *Théorie analytique des probabilités*, Paris: V Courcier; repr. F.W. Truscott and F.L. Emory (trans.) *A Philosophical Essay on Probabilities*, New York: Dover, 1951.(The page reference is from the reprint. This work is famous as a treatise on probability theory and its philosophy, quite apart from its statement of determinism in the introduction.)

Montague, R. (1974) ‘Deterministic Theories’, in R. Thomason (ed.) *Formal Philosophy*, New Haven, CT: Yale University Press.(Technical paper pioneering the kind of definition given in §2-3. This collection was posthumous published.)

Xia, Z. (1992) ‘The Existence of Noncollision Singularities in Newtonian Systems’, *Annals of Mathematics* 135: 411-68.(Mathematical paper proving indeterminism in Newtonian gravitation.)

Development ethics

Development ethics is ethical reflection on the ends and means of socioeconomic change in poor countries and regions. It has several sources: criticism of colonialism and post-Second World War development strategies; Denis Goulet's writings; Anglo-American philosophical debates about the ethics of famine relief; and Paul Streeten's and Amartya Sen's approaches to development.

Development ethicists agree that the moral dimension of development theory and practice is just as important as the scientific and policy components. What is often called 'development' - economic growth, for instance - may be bad for people, communities and the environment. Hence, the process of development should be reconceived as beneficial change, usually specified as alleviating human misery and environmental degradation in poor countries.

Development ethicists do not yet agree on whether their ethical reflection should extend to destitution in rich *countries or aspects of North-South relations apart from development aid. Other unresolved controversies concern the status and content of substantive development norms. Finally, agreement does not yet exist as to how the benefits of and responsibilities for development should be distributed within and between countries.*

1 The nature of development ethics

National policy makers, project managers and international aid donors involved in development in poor countries often confront moral questions in their work. Development scholars recognize that social-scientific theories of 'development' and 'underdevelopment' have ethical as well as empirical and policy components. Development philosophers and other ethicists formulate ethical principles relevant to social change in poor countries, analyse and assess the moral dimensions of development theories and seek to resolve the moral quandaries raised in development policies and practice: In what direction and by what means should a society 'develop'? Who is morally responsible for beneficial change? What are the obligations, if any, of rich societies (and their citizens) to poor societies?

2 Sources

There are several sources for moral assessment of the theory and practice of development. First, beginning in the 1940s, activists and social critics - such as Gandhi in India, Raúl Prébisch in Latin America, and Frantz Fanon in Africa - criticized colonial and orthodox economic development (see Gandhi, M.K.; Fanon, F.). Second, since the early 1960s, American Denis Goulet, influenced by French economist Louis-Joseph Lebret and social scientists such as Gunner Myrdal, has argued that '"development" needs to be redefined, demystified, and thrust into the arena of moral debate' (1971: xix). Drawing on his training in continental philosophy, political science and social planning as well as on his grassroots experience in numerous projects in poor countries, Goulet was a pioneer in addressing 'the ethical and value questions posed by development theory, planning, and practice' (1977: 5). One of the most important lessons taught by Goulet (1971) is that so-called 'development', owing to its costs in human suffering and loss of meaning, can amount to 'anti-development' (see Berger 1974).

A third source of development ethics is the effort of Anglo-American moral philosophers to deepen and broaden philosophical debate about famine relief and food aid. Beginning in the early 1970s, often in response to Peter Singer's utilitarian argument for famine relief (1972) and Garrett Hardin's 'lifeboat ethics' (1974), many philosophers debated whether affluent nations (or their citizens) have moral obligations to aid starving people in poor countries and, if they do, what are the nature, bases and extent of those obligations (see Aiken and LaFollette (eds) 1976). By the early 1980s, however, moral philosophers, such as Nigel Dower, Onora O'Neill and Jerome M. Segal, had come to agree with those development specialists who for many years had believed that famine relief and food aid were only one part of the solution to the problems of hunger, poverty, underdevelopment and international injustice. These philosophers have argued that what is needed is not merely an ethics of aid but a more comprehensive, empirically informed, and policy relevant 'ethics of Third World development'. The kind of assistance and North-South relations called for depend on how (good) development is understood.

A fourth source of development ethics is the work of Paul Streeten and Amartya Sen. Both of these economists have addressed the causes of global economic inequality, hunger and underdevelopment, and attacked these problems with, among other things, a conception of development formulated explicitly in terms of ethical

principles. Building on Streeten's 'basic human needs' strategy (1981), Sen argues that development should be understood ultimately not as economic growth, industrialization or modernization, which are at best means (and sometimes not very good means), but as the expansion of people's valuable capabilities and functionings: 'what people can or cannot do, e.g., whether they can live long, escape avoidable morbidity, be well nourished, be able to read and write and communicate, take part in literary and scientific pursuits, and so forth' (1984: 497) (see Nussbaum and Sen (eds) 1993; Nussbaum and Glover (eds) 1995).

These four sources have been especially influential in the work of Anglo-American development ethicists. When practised by Latin Americans, Asians, Africans and non-Anglo Europeans, development ethics often draws on philosophical and moral traditions distinctive of their cultural contexts (see, for example, Luis Camacho (1993) on Costa Rica, and Godfrey Gunatilleke (1988) on Asia).

3 Areas of consensus

Although differing on a number of matters, development ethicists exhibit a wide consensus about the commitments that inform their enterprise, the questions they are posing and the unreasonableness of certain answers. Development ethicists typically ask the following related questions. What should count as (good) development? Should we continue using the concept of development instead of, for example, 'progress', 'transformation', 'liberation', or 'postdevelopment alternatives to development' (Escobar 1995)? What should be a society's basic economic, political and cultural goals and strategies, and what principles should inform their selection? What moral issues emerge in development policy making and practice and how should they be resolved? How should the burdens and benefits of development be conceived and distributed? Who or what should be responsible for bringing about development? A nation's government, civil society or the market (see Market, ethics of the)? What role - if any - should more affluent states, international institutions, and nongovernmental associations and individuals have in the self-development of poor countries? What are the most serious local, national and international impediments to good development? To what extent, if any, do moral scepticism, moral relativism, national sovereignty and political realism pose a challenge to this boundary-crossing ethical inquiry? Who should decide these questions and by what methods (see Moral expertise)?

In addition to accepting the importance of these questions, most development ethicists share ideas about their field and the general parameters for ethically-based development. First, development ethicists contend that development practices and theories have ethical dimensions and can benefit from explicit ethical analysis and criticism. Second, development ethicists tend to see development as a multidisciplinary field with both theoretical and practical components that intertwine in various ways. Hence, they aim not merely to understand development, conceived generally as desirable social change, but also to argue for and promote specific conceptions of such change. Third, although they may understand the terms in somewhat different ways, development ethicists are committed to understanding and reducing human deprivation and misery in poor countries. Fourth, a consensus exists that development projects and aid givers should seek strategies in which both human wellbeing and a healthy environment jointly exist and are mutually reinforcing (Engel and Engel (eds) 1990). Fifth, these ethicists are aware that what is frequently called 'development' - for instance, economic growth - has created as many problems as it has solved. 'Development' can be used both descriptively and normatively. In the descriptive sense, 'development' is usually identified as the processes of economic growth, industrialization, and modernization that result in a society's achieving a high (per capita) gross domestic product. So conceived, a 'developed' society may be either celebrated or criticized. In the normative sense, a developed society, ranging from villages to national and regional societies, is one whose established institutions realize or approximate (what the proponent believes to be) worthwhile goals - most centrally, the overcoming of economic and social deprivation. To avoid confusion, when a normative sense of 'development' is meant, the noun is often preceded by a positive adjective such as 'good' or 'ethically justified'.

A sixth area of agreement is that development ethics must be conducted at various levels of generality and specificity. Just as development debates occur at various levels of abstraction, so development ethics should assess (1) basic ethical principles, (2) development goals and models such as 'economic growth', 'growth with equity', 'basic needs' and, in the1990s, 'sustainable development', 'structural adjustment' and 'human development' (United Nations Development Programme), and (3) specific institutions and strategies.

Seventh, most development ethicists believe their enterprise should be international in the triple sense that the

ethicists engaged in it come from many nations, including poor ones; that they are seeking to forge an international consensus; and that this consensus emphasizes a commitment to alleviating worldwide deprivation.

Eighth, although many development ethicists contend that at least some development principles or procedures are relevant for any poor country, most agree that development strategies must be contextually sensitive. What constitutes the best means - for instance, state provisioning, market mechanisms, civil society and their hybrids - will depend on a society's history and stage of social change as well as on regional and global forces.

Ninth, this flexibility concerning development models and strategies is compatible with the uniform rejection of certain extremes. For example, most development ethicists would repudiate two models: (1) the maximization of economic growth in a society without paying any direct attention to converting greater opulence into better human living conditions for its members, what Sen and Jean Drèze call 'unaimed opulence', and (2) an authoritarian egalitarianism in which physical needs are satisfied at the expense of political liberties.

4 Controversies and agenda

In addition to these points of agreement, there are several divisions and unsettled issues. A first unresolved issue concerns the scope of development ethics. Development ethics originated as the 'ethics of Third World Development'. There are good reasons to drop - as a Cold War relic - the 'First-Second-Third World' trichotomy. There is no consensus, however, on whether or not development ethics should extend beyond its central concern of assessing the development ends and means of poor societies.

Some argue that development ethicists should criticize human deprivation wherever it exists and that rich countries, since they too have problems of poverty, powerlessness, and alienation, are 'underdeveloped' and, hence, fall properly within the scope of development ethics. Perhaps the socioeconomic model that the North has been exporting to the South results in the underdevelopment of both. Others contend that since development ethicists address questions of rich country responsibility and global distributive justice, they should not restrict themselves to official development assistance but should also treat international trade, capital flows, migration, environmental pacts, military intervention, and responses to human rights violations committed by prior regimes. The chief argument against extending the boundaries in these ways is that development ethics would thereby become too ambitious and diffuse. If development ethics grew to be identical with all social ethics or all international ethics, the result might be that insufficient attention would be paid to alleviating poverty and powerlessness in *poor* countries. Both sides agree that development ethicists should assess various kinds of North-South (and South-South) relations with respect to their effects on economic and political inequality in poor countries. What is unresolved, however, is whether development ethics also should address such topics as the ethics of trade, military intervention and international institutions.

Development ethicists are divided also on the status of the moral norms that they seek to justify and apply. Three positions have emerged. Universalists, such as utilitarians and Kantians, argue that development goals and principles are valid for all societies (see Universalism in ethics §3; Utilitarianism; Kantian ethics). Particularists, especially communitarians and postmodern relativists, reply that universalism masks ethnocentrism and (Northern) cultural imperialism (see Community and communitarianism; Postmodernism). Pro-development particularists either eschew all universal principles or affirm only the *procedural* principle that each nation or society should draw only on its own traditions and decide its own development ethic and path. (Anti-development particularists, rejecting both change brought from the outside and public reasoning about social change, condemn all development discourse and practice.) A third approach - advanced, for example, by Martha Nussbaum, Jonathan Glover, Seyla Benhabib and David Crocker - tries to avoid the standoff between the first two positions (see Nussbaum and Glover (eds) 1995). On this view, development ethics should forge a cross-cultural consensus in which a society's own freedom to make development choices is one among a plurality of fundamental norms and in which these norms are of sufficient generality so as not only to permit but also to require sensitivity to societal differences (see Moral pluralism).

Next is a question related to the universalism-particularism debate: to what extent, if any, should development ethicists propose visions committed to a certain conception of human wellbeing, and how 'thick' or extensive should this vision be? There is a continuum here: at one end, there is more commitment to the values of individual choice, tolerance of differences, and public deliberation about social ends and means; and, at the other, more

normative guidance about the good human life but less room for individual and social choice.

Supposing that development principles should have some substantive content (beyond the procedural principle that each society or person should decide for itself), there are disagreements about that content. Assuming that social development concerns human development, with what moral categories should it be conceived? Candidates include: utility (preference satisfaction); social primary goods, such as income; negative liberty; basic human needs; autonomy; valuable capabilities and functionings; and rights. Although some think that a development ethic ought to include more than one of these moral concepts, development ethicists differ with respect to which ones to embrace and how to relate them. One alternative would be to work out a concept of human wellbeing that combines, on the one hand, a Neo-Kantian commitment to autonomy, critical dialogue and public deliberation with, on the other hand, neo-Aristotelian beliefs in the importance of physical health and social participation. Development duties might then flow from the idea that all humans should have a right to at least a minimal level of wellbeing.

There is also an ongoing debate about how development's benefits, burdens and responsibilities should be distributed within poor countries and between rich and poor countries. Utilitarians prescribe simple aggregation and maximization of individual utilities. Rawlsians advocate that income and wealth be maximized for the least well-off (individuals or nations) (see Rawls, J.). Libertarians contend that a society should guarantee no form of equality apart from equal freedom from the interference of government and other people. Capabilities ethicists defend governmental responsibility to enable everyone to be able to advance to a level of sufficiency with respect to the valuable functionings.

Development ethicists also differ with respect to whether (good) social development should have - as an ultimate goal - the promotion of values other than the present and future human good. Some development ethicists ascribe intrinsic value, equal to or even superior to the good of individual human beings, to human communities of various kinds, for instance, family, nation or cultural group. Others argue that nonhuman individuals and species, as well as ecological communities, have equal and even superior value to human individuals (see Animals and ethics). Those committed to 'ecodevelopment' or 'sustainable development' do not yet agree on what should be sustained as an end in itself and what should be maintained as an indispensable or merely helpful means (see Environmental ethics). Nor do they agree on how to surmount conflicts among intrinsic values.

See also: Future generations, obligations to; Justice, international; Postcolonialism

DAVID A. CROCKER

References and further reading

Aiken, W. and LaFollette, H. (eds) (1976) *World Hunger and Morality*, Upper Saddle River, NJ: Prentice Hall; 2nd edn, 1996.(The first edition comprises philosphical debates in the early 1970s over the ethics of food aid; as well as retaining three of these articles, the second edition contains later pieces that either address responsibilities for aid or, especially in parts III and IV, recast the ethics of food aid to an ethics of development and international justice.)

Aman, K. (ed.) (1991) *Ethical Principles for Development: Needs, Capacities or Rights?*, Upper Montclair, NJ: Institute for Critical Thinking.(Outlines diverse approaches to the question of whether development ethics should emphasize needs, capabilities or rights.)

Attfield, R. and Wilkins, B. (eds) (1992) *International Justice and the Third World*, London: Routledge.(British essays on development ethics and international justice.)

Berger, P. (1974) *Pyramids of Sacrifice: Political Ethics and Social Change*, New York: Basic Books. (Challenging essay linking 'political ethics' and Third World poverty.)

Camacho, L. (1993) *Ciencia y tecnología en el subdesarrollo*, Cartago: Editorial Tecnológica de Costa Rica.(A leading Latin American philosopher addresses ethical issues concerning science and technology in developing countries.)

Crocker, D.A. (1991) 'Toward Development Ethics', *World Development* 19: 457-83.(An introductory survey and bibliography of the nature, methods, and value of development ethics.)

Dower, N. (1988) 'What is Development? A Philosopher's Answer', *Centre for Development Studies Occasional Paper Series* 3, Glasgow: University of Glasgow.(Argues for the role of philosophers in development studies and for the normative meaning of development as socioeconomic change that ought to occur.)

Engel, J.R. and. Engel, J.G. (eds) (1990) *Ethics of Environment and Development: Global Challenge,*

International Response, Tucson, AZ: University of Arizona Press.(An international collection that explores different ways of relating environmental and development ethics.)

Escobar, A. (1995) *Encountering Development: The Making and Unmaking of the Third World*, Princeton, NJ: Princeton University Press.(Postmodernist ‘deconstruction’ of development theory and practice as imperialist and destructive of traditional life.)

Gasper, D. (1994) ‘Development Ethics - An Emergent Field? A Look at Scope and Structure, with Special Reference to the Ethics of Aid’, in R. Prendergast and F. Stewart (eds) *Market Forces and World Development*, London: Macmillan, New York: St Martin’s Press.(A helpful stocktaking that defends development ethics as a ‘multidisiciplinary field’ that closes the gap between abstract philosophy and practical experience in order to promote desirable change in poor countries.)

Goulet, D. (1971) *The Cruel Choice: A New Concept in the Theory of Development*, New York: Athenaeum.(In this seminal work, the pioneer in development ethics applies ethical reflection to development theory, policy and practices.)

Goulet, D. (1977) *The Uncertain Promise: Value Conflicts in Technology Transfer*, New York: IDOC/North America.(With many concrete examples, this work shows that technology transfer can be both an obstacle and an aid to authentic development.)

Gunatilleke, G., Tiruchelvam, N. and Coomaraswamy, R. (eds) (1988) *Ethical Dilemmas of Development in Asia*, Lexington, MA: Lexington Books.(Accessible international collection addressing ethical, empirical and political aspects of Asian development.)

Hardin, G. (1974) ‘Lifeboat Ethics: The Case Against Helping the Poor’, *Psychology Today* 8: 38-43, 123-6. (Argues on utilitarian and prudential grounds that affluent nations are morally obliged not to help famine-stricken countries.)

Nussbaum, M. and Glover, J. (eds) (1995) *Women, Culture and Development*, Oxford: Clarendon Press.(A valuable international collection of advanced essays explaining, assessing and applying the capabilities ethic to the issue of gender equality in developing countries.)

Nussbaum, M. and Sen, A. (eds) (1993) *The Quality of Life*, Oxford: Clarendon Press.(Advanced essays that clarify and evaluate the capabilities and other approaches to quality of life in both rich and poor countries.)

O’Neill, O. (1980) ‘The Moral Perplexities of Famine Relief’, in T. Regan (ed.) *Matters of Life and Death: New Introductory Essays in Moral Philosophy*, New York: Random House; repr. as ‘Ending World Hunger’, in T. Regan (ed.) *Matters of Life and Death: New Introductory Essays in Moral Philosophy*, New York: McGraw-Hill, 1993, 3rd edn.(A critique of utilitarian approaches to world hunger and an accessible statement of the author’s Kantian duty-based ethic of aid and development.)

Segal, J.M. (1991) ‘What is Development?’, in C.V. Blatz (ed.) *Ethics and Agriculture: An Anthology on Current Issues in World Context*, Moscow, ID: University of Idaho Press.(An influential paper that argues for an ethically based concept of development and evaluates three development models: growth, growth with equity, and basic needs.)

Sen, A. (1984) *Resources, Values and Development*, Oxford: Blackwell, Cambridge, MA: Harvard University Press.(Contains several of Sen’s important papers on the capabilities approach to development.)

Singer, P. (1972) ‘Famine, Affluence and Morality’, *Philosophy and Public Affairs* 1: 229-43; repr. in W. Aiken and H. LaFollette (eds) *World Hunger and Morality*, Upper Saddle River, NJ: Prentice Hall, 2nd edn, 1996. (Argues on utilitarian grounds affluent individuals are morally obliged to contribute to famine relief.)

Streeten, P. with Burki, S.J., Haq, M., Hicks, N., and Stewart, F. (1981) *First Things First: Meeting Basic Needs in Developing Countries*, Oxford: Oxford University Press.(The classic statement of the basic needs approach to change in poor countries.)

Dewey, John (1859-1952)

The philosophy of John Dewey is original and comprehensive. His extensive writings contend systematically with problems in metaphysics, epistemology, logic, aesthetics, ethics, social and political philosophy, philosophy and education, and philosophical anthropology. Although his work is widely read, it is not widely understood.

Dewey had a distinctive conception of philosophy, and the key to understanding and benefiting from his work is to keep this conception in mind. A worthwhile philosophy, he urged, must be practical. Philosophic inquiry, that is, ought to take its point of departure from the aspirations and problems characteristic of the various sorts of human activity, and an effective philosophy would develop ideas responsive to those conditions. Any system of ideas that has the effect of making common experience less intelligible than we find it to be is on that account a failure. Dewey's theory of inquiry, for example, does not entertain a conception of knowledge that makes it problematic whether we can know anything at all. Inasmuch as scientists have made extraordinary advances in knowledge, it behoves the philosopher to find out exactly what scientists do, rather than to question whether they do anything of real consequence.

Moral philosophy, likewise, should not address the consternations of philosophers as such, but the characteristic urgencies and aspirations of common life; and it should attempt to identify the resources and limitations of human nature and the environment with which it interacts. Human beings might then contend effectively with the typical perplexities and promises of mortal existence. To this end, Dewey formulated an exceptionally innovative and far-reaching philosophy of morality and democracy.

The subject matter of philosophy is not philosophy, Dewey liked to say, but 'problems of men'. All too often, he found, the theories of philosophers made the primary subject matter more obscure rather than less so. The tendency of thinkers is to become bewitched by inherited philosophic puzzles, when the persistence of the puzzle is a consequence of failing to consider the assumptions that created it. Dewey was gifted in discerning and discarding the philosophic premises that create needless mysteries. Rather than fret, for instance, about the question of how immaterial mental substance can possibly interact with material substance, he went to the root of the problem by challenging the notion of substance itself.

Indeed, Dewey's dissatisfaction with the so-called classic tradition in philosophy, stemming at least from Plato if not from Parmenides, led him to reconstruct the entire inheritance of the Western tradition in philosophy. The result is one of the most seminal and fruitful philosophies of the twentieth century.

1 Life

Dewey was born in Burlington, Vermont in 1859. He earned his BA degree at the University of Vermont, where he was particularly drawn to Scottish common-sense realism. He took his Ph.D. at the Johns Hopkins University, writing his dissertation on the psychology of Kant. At Hopkins he was much influenced by the experimental psychology of G. Stanley Hall, but he was taken above all by the philosophy of Hegel, taught by George Sylvester Morris. Hegel's philosophy, Dewey later said, answered his craving for a unified vision of reality, as distinguished from dualistic and reductive philosophies and those that were content with providing isolated analyses of particular phenomena.

After holding appointments at the University of Michigan and the University of Minnesota, he moved to the University of Chicago in 1894, where he was head of the Department of Philosophy, Psychology and Pedagogy. There he founded the University Elementary School (better known as the Laboratory School or the Dewey School), where he introduced experimental educational practices that both stemmed from and modified his philosophy of education. During this period Dewey began a decided shift away from Hegelian idealism in the direction of his naturalistic experimentalism. In 1904 he moved to Columbia University in New York City, where he developed his mature philosophy. He retired in 1930 but continued to teach as Professor Emeritus at Columbia until 1939. Thereafter he continued to be active, mostly in New York City, until his death in 1952.

Dewey's production of philosophic works was prodigious: forty books and more than 700 articles. After the death of William James in 1910, he was without peer in philosophy in the USA and became a figure of exceptional national and international prominence. He was frequently invited to lecture abroad; he did so in Mexico, China,

Japan, Turkey, France, Scotland and the USSR. He was called to several major lectureships, including the Carus Lectures in 1922, the Gifford Lectures in 1929, the William James Lecture in 1930 and the Terry Lectures in 1934. He received honorary degrees from several major universities world-wide and was a founder and the first President of the American Association of University Professors and President of the American Philosophical Association.

Inasmuch as he took seriously the notion of philosophy as a guide to conduct, it was fitting that he devoted much of his time to practical causes. He took a leading role in numerous associations devoted to political and educational reforms, and at the same time contributed frequently to public discourse in such journals of opinion as *The New Republic*, *Commentary*, *The Nation* and *Christian Century*. In 1937, at the age of 78, he accepted the chairmanship of the Commission of Inquiry to investigate the charges of Josef Stalin against Leon Trotsky. The question of Trotsky's guilt or innocence was secondary to the question of whether Stalin was the leader of a great democratic and economic revolution or the perpetrator of massive and brutal oppression. Dewey and the Commission found Trotsky not guilty and Stalin guilty, for which he was viciously criticized by a host of intellectuals sympathetic to Stalin.

2 Philosophical anthropology

Dewey's thinking about human nature is profoundly indebted to the theory of biological evolution as well as to James' *Principles of Psychology* (1890), which itself owed much to the Darwinian view. The classical idea of species implies changeless essence. It was also taken to imply that there is a perfectly definite end and station in life that each human being ought to achieve in order to fulfil their nature. Darwin's theory that species come into existence and undergo change was, therefore, truly revolutionary: the potentialities of human nature are not fixed, and there is no predetermined outcome or definite social role towards which each and every person ought to aim (see Darwin, C.R. §§2-3; James, W. §2).

According to the Darwinian conception, the human being is not understood primarily as a spectator of nature, but as a participant in it. Like any living organism, we are caught up in processes of change whose tendencies are fateful for our weal and woe. All human practices and aims are conceived with this assumption foremost in mind. Any form of human endeavour - including moral, cognitive and artistic activity - is elucidated as a mode of practice in which the individual acts with multiple agencies of the surrounding environment. Any experience is a conjoint effect of the organism and environment acting together, rather than a passive registration of external stimuli.

Dewey's notion of habit is fundamental to his analysis of human nature. A habit is not the private possession of an individual. It is the product of interactions inclusive of organism and environment, in which the resultant habit is determined as much by particular features of the environment as by the impulses of the organism. The habits appropriate to agricultural practice, for example, could not arise in the individual in isolation from the conditions in which farming occurs. The soil, the seasons and the elements dictate forms of action as much as the physiological and intellectual capacities of the farmer. The habits of farming are a collaborative result of organism and environment.

An impulse by itself does not constitute a form of action. It is plastic, and a formed habit of thought or action can only result from interaction. Dewey believed that a person's nature can take a plurality of possible forms, depending on environmental conditions. While he acknowledged that there are native differences in human beings, Dewey regarded such variations as highly malleable and subject to change as a consequence of learning. For everyone, then, there are multiple opportunities and possible directions for growth. Growth implies that our basic dispositions, our virtues, intelligence, talents and traits of personality are subject to the formations of home, school, church and every manner of human association. It is no surprise, then, that Dewey devoted so much study and experiment to the practice of education.

There is no mind or self antecedent to the formation of habit. The self, according to Dewey, is a dynamic complex of habits. It is an outcome, not an original existence. Mind consists of habits. Every idea is a habit of action; it is an anticipation of definite ways of behaving. The idea of a table, for example, consists of all the ways one can act with the object so designated. To have an idea of a table is not necessarily to have an image, nor is it to envisage the inherent essence of tablehood. It is to be prepared to act with such objects in ways to bring about predictable outcomes. Such outcomes are originally learned by means of active participation with objects, hence mind is

derivative of behaviour. But the meanings which constitute mind are mostly (perhaps exclusively, Dewey sometimes seems to say) provided by language. Language, too, is derivative of conduct, but it is necessarily social conduct. The public meanings conveyed by language could not arise except in shared activity with shared features of an environment. According to Dewey's thinking, consequently, mind can arise only within a social condition. Actual thinking is a function of acquired habits and the promptings of a specific situation.

In accordance with these conceptions, Dewey distinguished a form of conduct that could be common to all human activities. This form he called, generically, art. It is a deliberate practice with the environment, by means of which an individual transforms conditions that are initially disordered and troubling into a situation that is unified and directed. The individual must identify the nature of the problematic conditions, formulate possible plans of action that would utilize the instrumentalities of the particular situation, and execute the plan that seems to hold most promise for success. In conduct as art, individuals move from an initial condition of alienation from the environment to one in which they have effectively engaged their powers with those of their surroundings and have brought the two to consummation. Dewey thought that conduct of this sort characterizes all forms of effective activity. This type of conduct is natural to a living being striving to cope with processes of change, and it provides a consummation of effort that human beings find intrinsically fulfilling (see Pragmatism §1; Anthropology, philosophy of).

3 Metaphysics

Dewey did not think of metaphysics as a foundational discipline, conjuring up principles to certify value judgments and cognitive claims. On the contrary, he was very critical of metaphysics which had such pretensions; such endeavours were part of the obnoxious heritage of the classic tradition.

The classic tradition originating in ancient Greek philosophy is above all characterized by the assumption of disparate realms of being, Dewey contended. These realms are conceived in notably different ways, but they have in common the belief that there is a world of 'true' being, the most real being: that of changeless perfection, which is juxtaposed to a world of becoming, where all things come into existence, undergo change and pass away. The classic notion of being is not only false but pernicious, for it provides the rationale for moral absolutes, whose actual effect is to demand a conformity that is injurious to human flourishing (see Plato §10).

A universal assumption of the reigning tradition, Dewey observed, was that true being is conceived as the object of rational knowledge as such. That alone is the really real. Prior philosophies accordingly had been typically reductive, declaring with the idealists, for example, that reality *per se* is nothing but a rational order of ideas, or with the materialists that nature is nothing but matter in motion, possessing none of the other traits commonly experienced: such fateful events as immediate quality, disorder and ends. With these reductive assumptions, philosophers attributed to subjective mind within itself the plural and varying properties that would otherwise be predicated of nature (see Idealism §1; Materialism §3).

Dewey's metaphysical thought is an attempt to restore to the world the integrity of which philosophers had divested it. He provides a characterization of the most prominent features of the natural world in their continuities with human experience. This is a world of becoming, in the classic vernacular; there is no changeless being. Such phenomena as experience, human nature, mind, knowledge, art and value are conceived as natural functions of the biological organism interacting with an environment marked by change, order and disorder, obstacles, instrumentalities and episodic fulfilments.

The restoration of integrity of vision is consequent upon the continuity of experience and nature. We experience real events in their relations to ourselves and to each other. We accept our experience as disclosing real traits of the world. The varying qualities of events are natural functions. Not only are sensations of colours and sounds the outcomes of natural processes, but also the full panoply of characteristics that we refer to with such predicates as 'lovely', 'hateful', 'wicked' and 'admirable'. Genuine potentialities of nature are thus exhibited. The potentiality is neither in the organism alone nor in the object alone, but in the situation inclusive of both.

With such assumptions, Dewey insists, for example, on the reality of what he calls the precarious: events that are threatening, disordered, baffling and the like, which were formerly excluded from the really real. We experience such things as terrorist attacks, avalanches, lost keys, droughts and stubbed toes. These are not simply mental occurrences, unrelated to nature. They are public phenomena, and if we cannot accord them appropriate

ontological status, philosophy can offer no clue as to how we can contend with them.

Qualitative change, likewise, is as real as anything else and it is one of the traits of nature that Dewey discerned and analysed with telling effect. He calls these changes ‘histories’. Phenomena that had been regarded in the classic tradition as original entities or substances are treated by Dewey as the outcomes of histories. Mind, intelligence, knowledge, human nature and moral and aesthetic values are such outcomes. Dewey frequently analysed them in detail, believing that to possess knowledge of such change is invaluable in contending deliberately and effectively with the processes in which we are mortally implicated.

A further significant trait is what Dewey calls ‘the stable’. This expression refers not only to conditions that are relatively constant and enduring, but, fundamentally, to the correlations between processes of change. So-called laws of nature are not self-subsistent beings; they are the correlations between one process of change and another, as are those involving pressure and temperature in an enclosed volume of gas, for example. Both pressure and temperature undergo change, and the change in one is correlated in a measurable and predictable way with change in the other. As such correlations are determined, it becomes possible to introduce deliberate variations in natural events for the sake of directing them to predictable outcomes. Knowledge, then, is neither a direct grasp of essences nor a summary of antecedently received sensations. It is an instrument for directing processes of change.

The purpose of Dewey’s metaphysics is practical. Its main intent, he declared, is to exhibit the nature and function of intelligence. Contrast the classic model with Dewey’s: if the nature of the world is thought to be morally and cognitively determined by a pattern of changeless perfection, then the role of intelligence is to attain a direct grasp of such principles. Subsequently, events in the world of becoming are classified as conforming or not conforming to the antecedent pattern. This is a classificatory and deductive logic, formalized by Aristotle. According to such a logic, moral action consists of the attempt to constrain processes of change to conform to the antecedent absolute. If, however, the object of knowledge is the correlation between processes of change, and if the constituents of these processes can be varied, then it becomes possible to direct such processes towards welcome fulfilments. This procedure Dewey calls ‘experimental logic’. These are distinctly different procedures of practical action, and if the assumptions of the classic tradition are false, then conduct is pointlessly straitjacketed. When Dewey says it is the task of philosophy to characterize the nature of things in a way that illuminates possible harmonies between our ambitions and aspirations and the realities of the world, he has in mind the substitution of an experimental logic for the stultifying anachronisms of the classic tradition (see Processes §4; Whitehead, A.N. §4).

4 Theory of inquiry

The heart of Dewey’s pragmatism - or ‘instrumentalism’, as he preferred - is his theory of inquiry. It is most significant that the explicit guiding assumption of this theory is that logical distinctions and methods of inquiry develop out of the process of problem-solving activities. The logic of inquiry is not a set of norms existing independently of and prior to our cognitive efforts. Rather, procedures and logical distinctions introduced by inquirers prove themselves in their successes in resolving cognitively problematic situations. ‘Successful’ is not to be defined by philosophers, but by the aims of modern science itself, which Dewey took to be the prediction and control of natural phenomena. Ultimately at stake in disputes about the nature and possibility of knowledge is the capacity of human beings to act in concert with ongoing events. Such action is dependent upon the determination of how changes in one process are correlated with changes in another.

Science, then, does not get its credentials from philosophy. Instead, the philosopher studies science for a number of worthy purposes: to understand the nature of effective inquiry, the nature of the valid object of knowledge and the possible extensions of scientific methods to domains of problem-solving that are not exclusively scientific. Dewey also appropriates scientific results as indispensable constituents in his particular analyses of human nature, social processes and methods of education. (This is not to say that scientists themselves do not sometimes misunderstand the nature of their own activity, as Dewey on occasion points out.)

According to any philosophy at the turn of the twentieth century, all experience is exclusively subjective: it takes place within the mind, rather than being a consciousness of events external to the sentient being. Traditional dualistic epistemology, then, treats experience as a private, self-enclosed domain that is only problematically (if at all) related to anything beyond itself. Epistemologists laboured to conceive how a purely subjective mental event might possibly correspond to something external to it. If experience is continuous with nature, however, the real

problem is of an entirely different sort: to determine how variations in one natural and public event (or set of events) are related to variations in other such events.

This is a problem not for philosophers, but scientists, and scientists have proved very successful in reaching such determinations. They perform experiments and make observations in accordance with a hypothesis. The hypothesis states that if specified changes are introduced under controlled conditions, predictable effects will ensue. If the inquiry is conducted as the hypothesis prescribes, and if the effects occur as predicted, then the hypothesis is tentatively confirmed. Dewey always insisted that warrant for the putative knowledge is dependent upon confirmation from the community of relevant inquirers. Such warrant is always a matter of degree and it is always fallible.

Inquiry involves the deliberate manipulation of the conditions within which it occurs. The very subject matter of the investigation might be controlled, as in laboratory work; or the conditions of observation might be controlled, as in astronomy. In any case, the inquirer introduces deliberate change in the cognitive situation in order to produce the object of knowledge.

The object of knowledge as such is the outcome of a process; it does not exist antecedently to inquiry. Natural events, with their own potentialities, powers and limitations, exist prior to inquiry, of course, but they do not in that condition constitute an object of knowledge. When such an object is produced, its character is dependent upon the specific conditions of inquiry that have been introduced, including the subject matter under scrutiny. The investigator has responded to a specific problematic situation, has contrived an inquiry of a particular sort, and makes observations and measurements in terms of concepts that are largely characteristic of the existing state of the science. Inquiry is inevitably theory-laden - or, in Dewey’s parlance, meaning-laden. Inquiry constitutes, in identifiable and specifiable ways, a transformation of antecedent subject matter.

Inasmuch as Dewey regards the object of knowledge as a deliberate construction, some readers have concluded that he is an antirealist. This conclusion is mistaken. Dewey repudiated the archaic idea that knowledge is a correspondence between an object and a mental image, but that does not make him an antirealist. Indeed, whatever ‘realism’ is taken to mean, Dewey’s study of inquiry concludes that our ideas are determined by the nature of real events to a far greater extent than was recognized in any pre-pragmatic philosophy. The meaning of any object consists of the conception of all that object can do in relation to other objects. In other words, we know how objects function as constituents of processes of change; and in that form alone can we utilize them for practical purposes. Mere images cannot do so; nor can Platonic forms or Cartesian essences. Our conceptual systems, moreover, have arisen from the very processes of effective inquiry. Although our concepts are subject to revision, it can hardly be said that they are arbitrary impositions on nature. While it must be insisted that the successes of science are always open to refinement and revision, it is unmistakable that our ability to function with formerly inscrutable and overwhelming natural powers has undergone astounding advancement. Dewey calls this shift a revolution from arts of acceptance to arts of control (see Scientific realism and antirealism §4).

5 Moral philosophy: the construction of good

The classic tradition has supposed the existence of something essentially changeless and perfect in the nature of things, ranging from Platonic forms to natural law, and from Kantian reason to the nature of the Absolute. Even utilitarianism shares in that tradition in supposing a single criterion of good, as does Marxism in holding that there is but a single pattern that describes all historical change. Many philosophers have thought that they espy a moral absolute; and many have believed that such an absolute is of supreme value in human conduct. Many others would like to agree, but they have despaired of discerning such a rationally incontrovertible standard.

The classic tradition assumes a changeless principle or goal for all moral action, yet the real world is marked by change. It presents the continuing occurrence of unique situations, each of which can yield a variety of possible outcomes. There is likewise an indefinite plurality of human possibilities. Adherence to the classic tradition, then, constitutes a demand for unyielding conformity to prevailing cultural requirements, which are given a specious authority by religious or philosophic speculations. The tradition enforces conditions in which human possibilities are denied for the sake of maintaining precedent, custom or arbitrary authority. Historically, the feudal system has been the most prominent example. The contemporary demand that women and men should not be confined to their traditional roles is, in effect, a repudiation of the classic tradition.

As Dewey conceived it, the moral life is suffused with innumerable possibilities of enjoyment and happiness, as well as of disaster. Ordinary life revolves around familiar attachments, ambitions and fears. The philosophic task, Dewey believed, is to place at the disposal of human beings the assumptions and methods that would facilitate the efforts in which they will be engaged in any case. It is a project of enabling and liberating. Each situation might be approached to identify its unique opportunities for fulfilment. The precarious values of the situation might be clarified and united into more inclusive wholes and made more secure against obstacles and threats.

This procedure Dewey calls the 'construction' of good. No phrase encapsulates so well both his opposition to the classical tradition and his alternative to it. Rather than conform to an antecedently given pattern of right conduct, one actively constructs a new situation, intending to discriminate its particular tendencies and to contrive a consummation of the energies of the organism and environment - the practice referred to in §3 as 'conduct as art'. Dewey can be eloquent in characterizing such experience as one of deeply enriched qualities and cherished meanings. It is a cumulative experience, enhancing human powers. The continuation of this process over time Dewey denominates 'growth'.

There is no absolute consummation in life, nor any exact station determined a priori by one's nature. Rather, one may treasure the very process of conduct as art. Dewey says that growth itself is the only moral end, and he identifies it primarily as a social process. Our behaviour is characteristically interpersonal; as such, it is the source of most of our learning, and our participation with others is the source of our most profound satisfactions. As a constituent of growth, 'shared experience is the greatest of human goods'. The construction of good is typically a shared activity, greatly facilitated by deliberate cooperation.

The activity requires appropriate habits of thought and conduct, including those of communication and cooperation. Above all, Dewey believes, it requires the habits of experimental inquiry. The situations to be transformed are often of a complex and comprehensive nature, such as, for example, the reconstruction of educational institutions; so we must possess reliable knowledge of how such practices actually work. We must also possess the imaginative habits of the experimenter to formulate hypotheses about how such conditions might be reconstructed. We might hypothesize that certain innovations in practice would eventuate, say, in more alert, reflective and well-motivated students. The hypothesis guides conduct by prescribing the changes that would effect such results: the introduction of specified procedures would transform educational activity in such a way as to produce the desired outcomes. This is the manner in which scientific hypotheses guide conduct, not - by contrast - in deducing a prescription for action from an alleged moral absolute. That would be a reversion to the logic of the classical tradition (see Pragmatism in ethics §3).

6 Moral philosophy: the democratic way of life

Dewey was convinced that modern society is swamped in the modes of thought inherited from the classic tradition. Instead of using experimental logic, we depend on putative moral absolutes. Our conduct is governed, in effect, by precedent, which largely represents, according to Dewey, the entrenched prerogatives of privileged groups. The most beneficent change that could be introduced into society, consequently, would be the teaching and widespread adoption of experimental habits of thought (see Education, history of philosophy of §9). We would not just address our several problematic situations with greater promise of success, but we would re-evaluate our economic and political behaviour with a view to what that behaviour really does and to what any alternatives to it might do.

Dewey sometimes called deliberation and action of this sort 'social intelligence' and sometimes 'democracy as a way of life'. Democracy as a way of life, he said, could be a norm for all forms of human association. Surprisingly, Dewey never provided a systematic exposition of the norms of this life, but its constituents can be extracted from his writings. It consists of several habits or virtues, foremost among them experimental habits of inquiry. These include fallibilism, rather than dogmatism, and an insistence that we follow the evidence of inquiry wherever it leads, rather than cling to preconceived ideas. Experimentalism implies both a willingness to entertain novel hypotheses and the personal flexibility to try out unaccustomed modes of action. The dissemination of the results of inquiry as widely as possible is also indispensable to the democratic ethos.

Democratic life, as Dewey conceives it, also implies respect for persons. Respect of this sort does not entail policy prescriptions, but it means that every person has a right to participate in the formation of goods. It implies that individuals should be seriously receptive to each other's concerns, taking them into account - at least initially - as

deserving an honest hearing. It also implies that persons communicate with each other freely and honestly to convey their concerns and to propose their tentatively preferred plans of action. Out of such virtue and discourse, leavened with scientific knowledge, shared proposals for action would emerge and would be honoured. It should go without saying that democratic virtue excludes deliberately antisocial behaviour.

On a small scale, these procedures are familiar. In our personal relations - friendships, marriages and voluntary associations of all kinds - we do not presume to enforce our views on other participants in a unilateral way. We assume that there will be differences of opinion, great and small, and we typically assume that we will talk these problems over in a civil manner and try to arrive at some more or less amicable agreement about what to do. We do not always succeed in doing so, but when we fail, we do not generally bolt the association. Our common activities are usually too precious for anything so shortsighted. Our treasured relationships do not survive such intransigence. Dewey believed that the virtues exhibited in such relations might be extended to incorporate much wider groups.

No philosopher before Dewey had conceived moral discourse as essentially communicative. The democratic virtues, after all, are impertinent to the various incarnations of the classic tradition. If the good and the right are known absolutely, the discourse of social intelligence is superfluous. One must concede that such discourse does not guarantee that consensus will always occur. Conflict and frustration are inevitable; but Dewey believed that they would be substantially less than they are when immovable oppositions are generated by moral absolutes.

The disciplines of the democratic way of life are demanding, and they can easily be corrupted. Dewey urged that schooling of all kinds incorporate the procedures of cooperative intelligence as a matter of routine. Indeed, all forms of association might be conducted in a manner to make democratic values habitual. Were such favourable conditions to exist, he thought, this way of life could be reasonably approximated.

It is a way of life that can be criticized for providing insufficient finality to moral judgment. It is true: the democratic/scientific virtues do not ensure the unanimity craved by adherents to the classic tradition. On the other hand, their exercise precludes the source of most of our torments: deliberately hurtful conduct. And it promises a release of human possibilities from needless subservience to moral absolutes. In fact, the verdict on Dewey's moral philosophy might be that it makes excessive demands of human flesh, not that it is too permissive. In any case, if the classic tradition is indeed a relic, then perhaps Dewey's thought points the way to a more pertinent philosophy than any other currently offered to our candid judgment (see Social democracy §2).

JAMES GOUINLOCK

List of works

Dewey, J. (1969-90) *The Early Works of John Dewey, 1882-1898*; *The Middle Works of John Dewey, 1899-1924*; *The Later Works of John Dewey, 1925-1953*, ed. J.A. Boydston, Carbondale, IL: Southern Illinois University Press, 37 vols. (Contains the entirety of Dewey's published work. Does not include his correspondence, which is being edited for eventual publication.)

Dewey, J. (1887) *Psychology*; repr. in *Early Works*, vol. 2, Carbondale, IL: Southern Illinois University Press, 1969-90.(Dewey's text draws on Helmholz, Wundt, Lotze, Herbart, Bain, Spencer, James, Hall and others in a defence of idealism; criticized by James and others for its Hegelianism, it was eclipsed three years later by the publication of James' *The Principles of Psychology*.)

Dewey, J. (1896a) 'The Reflex Arc Concept in Psychology', *Psychological Review* 3: 357-70; repr. in *Early Works*, Carbondale, IL: Southern Illinois University Press, 1969-90, vol. 5, 96-106.(Important transitional work between Dewey's early Hegelianism and his later experimentalism and pragmatism.)

Dewey, J. (1896b) 'Interest in Relation to the Training of the Will', *Second Supplement to the First Yearbook of the National Herbart Society*, 209-46; repr. in *Early Works*, Carbondale, IL: Southern Illinois University Press, 1969-90, vol. 5, 111-50.(Dewey's revolutionary educational philosophy.)

Dewey, J. (1916) *Democracy and Education*, New York: Macmillan.(Dewey's most systematic statement of his philosophy of education. Good introduction to his entire philosophy. Contained in *Middle Works*, vol. 9.)

Dewey, J. (1922) *Human Nature and Conduct*, New York: Holt.(A comprehensive statement of Dewey's philosophy of human nature, including its bearing on moral conduct. Contained in *Middle Works*, vol. 12.)

Dewey, J. (1925) *Experience and Nature*, Chicago, IL, and London: Open Court.(Dewey's metaphysics. Contained in *Later Works*, vol. 1.)

Dewey, J. (1927) *The Public and Its Problems*, New York: Holt.(Principally on the nature of democratic life.

Contained in *Later Works*, vol. 2.)
Dewey, J. (1929) *The Quest for Certainty*, New York: Minton Balch.(Sustained critique of the classic tradition in philosophy. Contrasts absolutistic and experimental approaches to ethics. Contained in *Later Works*, vol. 4.)
Dewey, J. and Tufts, J.H. (1932) *Ethics*, New York: Holt.(An account of the distinctive values of the moral life, their sources, functions and interrelations. Contained in *Later Works*, vol. 7.)
Dewey, J. (1934) *Art as Experience*, New York: Minton Balch.(Superb statement of the nature of art and its significance for all human conduct. Contained in *Later Works*, vol. 10.)
Dewey, J. (1938) *Logic: The Theory of Inquiry*, New York: Holt.(Magisterial elaboration of Dewey's instrumentalism. Contained in *Later Works*, vol. 12.)

References and further reading

Alexander, T. (1987) *John Dewey's Theory of Art, Experience, and Nature: The Horizons of Feeling*, Albany, NY: State University of New York Press.(Elucidates the prominence of aesthetic experience in all conduct. Excellent antidote to the treatment of Dewey as technocrat.)
Boisvert, R. (1988) *Dewey's Metaphysics*, New York: Fordham University Press. (Clearly written exposition displays the sophistication of Dewey's metaphysics.)
Campbell, J. (1995) *Understanding John Dewey*, Chicago, IL: Open Court.(The best general introduction to Dewey's philosophy.)
Gouinlock, J. (1972) *John Dewey's Philosophy of Value*, New York: Humanities Press.(Exhibits Dewey's moral thought as organic to his philosophical anthropology and metaphysics.)
Gouinlock, J. (1986) *Excellence in Public Discourse: John Stuart Mill, John Dewey, and Social Intelligence*, New York: Teachers College Press.(Systematic exposition of Dewey's philosophy of social intelligence.)
James, W. (1890) *The Principles of Psychology*, Cambridge, MA: Harvard University Press, 1981, 3 vols.(A brilliant and exhaustive study of psychological theory as it existed in the late nineteenth century. Includes the extensive elaboration of James' own views.)
Rockefeller, S. (1991) *John Dewey: Religious Faith and Democratic Humanism*, New York: Columbia University Press.(A rich account of Dewey's moral philosophy and philosophy of religion in the context of Dewey's own life experience.)
Sleeper, R. (1986) *The Necessity of Pragmatism*, New Haven, CT: Yale University Press.(Excellent study of Dewey's theory of inquiry.)

Dharmakīrti (*c.*600-60)

Dharmakīrti represents the philosophical apex of the Buddhist contribution to Indian thought of the post-systematic period. On the basis of Dignāga's late works he developed a system of epistemology with a strong emphasis on logic, propounding it both as an explanation and defence of Dignāga's thought. His logic, particularly, was new; in order to create a system of Buddhist logic proper, it clearly established the general Indian idea that logical relations are founded in reality. Buddhist epistemology as shaped by Dharmakīrti became a strong and influential rational tradition in late Indian Buddhism and has been studied and continued in Tibet up to the present time.

1 Writings

According to legend, Dharmakīrti was born a South Indian Brahman and taught at the monastic 'university' of Nālanda. He was a pupil of Īśvarasena and, in the words of the Chinese pilgrim Yijing, 'brought further progress in the field of logic after Dignāga' ([671-95] 1966: 182). He was also an acknowledged poet. The later Buddhist tradition followed Dharmakīrti's interpretation of Dignāga, and he strongly influenced the Brahmanical logicians as well. His theories have been studied and transmitted by Tibetan epistemologists and logicians up to the present time and remain part of the traditional higher education curriculum in Tibet.

Possibly all Dharmakīrti's philosophical works, four major and three minor, are extant in their original Sanskrit in China, but due to the inaccessibility of these sources, in the case of the *Pramāṇaviniścaya* and the *Hetubindu* we are still forced to rely on Tibetan translations, fragments and reconstructions. Research on Dharmakīrti has shifted from an earlier, more doxographic attitude towards an investigation of his work in context, focusing on the problems he was concerned with and the answers developed in his major works. An early attempt at interpretation by Stcherbatsky (1932) provided a fascinating overview, but was hampered by a too limited textual basis and a strong neo-Kantian bias in interpretation.

Dharmakīrti's first work, a treatise consisting of verses and partly explanatory, partly elaborative prose, is an audacious statement of his new logical theory, which he openly presents as a new interpretation of Dignāga's teaching, superior to that of his teacher Īśvarasena. This early work (called *Hetu prakaraṇa* by Frauwallner) is basically a study of the definition of logical reason (*hetu*) which may have been intended to supersede a similar work by Dignāga that is now lost (the *Hetumukha*), although it is not, in character, a commentary on Dignāga's treatise. It was incorporated by Dharmakīrti into the great verse-compendium of the *Pramāṇavārttika (Commentary on Valid Cognition)* as its first chapter ('On Inference'), and the parts in prose were regarded as Dharmakīrti's own commentary on this chapter. The other chapters of the *Pramāṇavārttika* are, in fact, conceived as a commentary on Dignāga's *Pramāṇasamuccaya (Collected Writings on Valid Cognition)*, the second chapter dealing with the proof of the Buddha's authority, the third with perception and the fourth with proof. Much is new in this work but, because it is structurally unbalanced, Dharmakīrti left it unfinished.

Drawing heavily on the earlier work, Dharmakīrti chose to compose a new statement of his ideas as an independent text in three chapters (on valid cognition and perception, on inference, and on proof), the *Pramāṇaviniścaya (Analytical Determination of Valid Cognition)*, of which the *Nyāyabindu (Drop of Logic)* is an abstract. Dharmakīrti returned to his first subject with the *Hetubindu (Drop of Logical Reason)*, treating it in a more formal way, with digressions devoted to the proof of momentariness, to a detailed presentation of his theory of causality, and to a final refutation of his teacher's attempt at developing an alternative logic in order to 'save' Dignāga's theory (see Momentariness, Buddhist doctrine of §§1, 5). In his last work, the *Vādanyāya (Rule for Disputations)*, he applied the results of his work on the logical reason to interpretations of the forms and rules governing public disputations, and refuted the canon of the Nyāya school that had seemingly been in use for that purpose up to his time (see Nyāya-Vaiśeṣika §6). In the context of Indian polemical practice, Dharmakīrti is unique in emphasizing that disputations have their purpose in the search for truth. Two small works, the *Sambandhaparīkṣā (Examination of Relations)* and the *Santānāntarasiddhi (Proof of Other Mental Continua)*, have not yet been attributed to a definite period of production.

2 Epistemology and logic

Except for the second and third chapters of the *Pramāṇavārttika* and the two small treatises just mentioned, Dharmakīrti's work is mainly concerned with developing Dignāga's epistemology and logic. He presents new ideas in the traditional way as the correct interpretation of the latter's thought. Epistemological thought, in this context, is an attempt at providing a rational frame for human practice, meaningful both in everyday life and in the pursuit of the ultimate goal indicated by the Buddha. In the second chapter of the *Pramāṇavārttika*, the Buddha is proven to be an authority on the value of all human activity and the final goals of spiritual aspiration. The third chapter, 'On Perception', is a critical analysis of the nature and possibility of perception. While elsewhere in his work, especially when dealing with epistemological issues, Dharmakīrti takes his stand on an ontology of 'critical realism' (the Sautrāntika tradition) and develops empirical models and categories for his theories, when analysing perception the option of a reality independent of cognition is dissolved, and an idealistic ontology (of the Yogācāra kind) remains (see Epistemology, Indian schools of §1; Buddhism, Yogācāra school of §§1-4).

Dharmakīrti's theory of meaning is nominalistic. Words and concepts refer only to the difference from others (*anyāpoha*) that is common to individual realities. This common character is unreal, but in human experience realities evoke a like judgment of their capacities, and thereby a link is provided between the false realm of beginningless constructs and the momentary instants of reality experienced (see Universals, Indian theories of §3).

Like Dignāga, Dharmakīrti accepts two kinds of valid cognition which hold good in practice: perception and inference. Perception is a cognition devoid of conceptual construction and is nonerroneous. Inferential cognition, being conceptual, is by nature erroneous, but if attained under the strict formal conditions of logic may be necessarily true. Verbal cognition based on authority or scripture is also inferential when it can be established as being trustworthy with regard to its specific, totally imperceptible object.

In the field of logic, Dharmakīrti transforms Dignāga's theory, which emphasized the formal aspects of logical reason, by insisting, like pre-Dignāga Sāõkhya logic, on the need for the necessary logical relation to have a foundation in reality. In fact, he interprets Dignāga's logic through his own, arguing that the former is not possible without implying the latter. The three marks of a logical reason required by Dignāga are fulfilled only by Dharmakīrti's three kinds of logical reason. Two possible connections in reality, real identity and causality, restrict the kinds of concepts from which others can be inferred to three: 'essential property' (*svabhāva*) and 'nonperception' (*anupalabdhi*), based on real identity, and 'effect' (*kārya*), based on causality. When these real connections are ascertained, inference is a valid cognition, although because it is conceptual by nature it is essentially problematic. However, when inference is correctly applied it may serve to remove or correct erroneous concepts and replace them with others more appropriate to reality (see Inference, Indian theories of §§5-6).

In Dharmakīrti's thought, logical theory and application are tightly interwoven. For example, he develops a new form of the proof of universal momentary destruction (*kṣaṇabhaõga*), deriving it, in the *Pramāṇaviniścaya*, from the concept of being (*sattva*). In elaborating this proof for the first time, he applies a new method for ascertaining the logical relation between the two concepts involved. Subsequently, in the later *Hetubindu*, this new method is also incorporated into the theory of logic.

See also: Buddhist philosophy, Indian §§4-5; Nominalism, Buddhist doctrine of

ERNST STEINKELLNER

List of works

Most translations and studies are in German and Japanese (by E. Frauwallner, T. Vetter, H. Tosaki, E. Steinkellner, C. Oetke).

Steinkellner, E. and Much, M.T. (1995) *Texte der erkenntnistheoretischen Schule des Buddhismus (Texts of the Buddhist Epistemological Schools)*, Göttingen: Vandenhoek & Ruprecht, 23-44.(Bibliographical survey of the Dharmakīrti's texts in Sanskrit or Tibetan translation, and their translations into European languages and Japanese.)

References and further reading

Bijlert, V.A. van (1989) *Epistemology and Spiritual Authority*, Vienna: Arbeitskreis für Tibetische und Buddhistische Studien Universität Wien.(An annotated translation of *Pramāṇavārttika* ch. 2, verses 1-7, containing the definition of validity in cognitions.)

Dreyfus, G.B. (1997) *Recognizing Reality, Dharmakīrti's Philosophy and its Tibetan Interpretations*, Albany,

NY: State University of New York Press.(A comprehensive study of the Tibetan philosophical work in continuation of Dharmakīrti's ontology, philosophy of language and epistemology.)

Franco, E. (1997) *Dharmakīrti on Compassion and Rebirth*, Vienna: Arbeitskreis für Tibetische und Buddhistische Studien Universität Wien.(An annotated translation of *Pramāṇavārttika* ch. 2, verses 34-72, together with Prajñākaragupta's Commentary and five relevant studies.)

Stcherbatsky, T. (1930, 1932) *Buddhist Logic*, Leningrad: Izdatel'stvo Akademii Nauk SSSR, 2 vols; Osnabrück: Biblio Verlag, 1970; Delhi: Motilal Banarsidass, 1992. (An introduction to Buddhist logic on the basis of the *Nyāyabindu* and its commentary by Dharmottara; written with a heavy Neo-Kantian bias, it is a largely out-of-date, but still useful, survey and has been reprinted many times.)

Steinkellner, E. (ed.) (1991) *Studies in the Buddhist Epistemological Tradition. Proceedings of the Second International Dharmakīrti Conference, Vienna, June 11-16, 1989*, Vienna: Verlag der Österreichischen Akademie der Wissenschaften.(Conference proceedings giving a clear impression of the state of Dharmakīrti studies.)

Vetter, T. (1984) *Der Buddha und seine Lehre in Dharmakīrtis Pramāṇavārttikam (The Buddha and his Theory in Dharmakīrti's Pramāṇavārttikam),* Vienna: Arbeitskreis für Tibetische und Buddhistische Studien Universität Wien, 1990.(An annotated translation of *Pramāṇavārttika* ch. 2, verses 131-285 with a concise introduction on Dharmakīrti's religious thought.)

Yijing (671-95) *A Record of the Buddhist Religion as Practised in India and the Malay Archipelago*, trans. J. Takakuru, London, 1896; repr. Delhi, 1966.(Referred to in §1.)

Dialectical materialism

Dialectical materialism is the official name given to Marxist-Leninist philosophy by its proponents in the Soviet Union and their affiliates elsewhere. The term, never used by either Karl Marx or Friedrich Engels, was the invention of the Russian Marxist Georgii Plekhanov, who first used it in 1891. Engels, however, favourably contrasted 'materialist dialectics' with the 'idealist dialectics' of Hegel and the German idealist tradition, and the 'dialectical' outlook of Marxism with the 'mechanistic' or 'metaphysical' standpoint of other nineteenth-century materialists.

Dialectical materialism proclaims allegiance to the methods of empirical science and opposition to all forms of scepticism which deny that science can know the nature of reality. Dialectical materialists reject religious belief generally, denying the existence of non-material or supernatural entities (such as God or an immortal human soul). Unlike other forms of materialism, however, dialectical materialists maintain that the fundamental laws governing both matter and mind are dialectical *in the sense in which that term is used in the philosophy of G.W.F. Hegel.*

Although dialectical materialism is supposed to constitute the philosophical underpinnings of Marxism, Marx's only major contribution to it was his materialist conception of history. The more fundamental philosophical views of dialectical materialism have their main source in the writings of Engels, especially Anti-Dühring (1878), Dialectics of Nature (1875-82) and Ludwig Feuerbach and the End of Classical German Philosophy (1886). To this last work Engels appended the eleven 'Theses on Feuerbach' written by Marx in 1845, which contrasted the 'old' or 'contemplative' materialism with the practically oriented materialism which was to be the basis of the proletarian movement. Further developments of dialectical materialism are found in writings by V.I. Lenin and subsequent Soviet writers. Lenin's chief additions were his critique of 'empirio-criticism' (the empiricist phenomenalism of certain Russian followers of Ernst Mach, who argued that matter was to be reduced to sense data), and his conception of the 'partisanship' of all philosophical views.

1 Materialism

According to Engels, the fundamental question of all philosophy is 'Which is primary, matter or consciousness?' The question of 'primacy' is also put this way: 'Which, matter or consciousness, is the source of the other?' Materialism holds to the primacy of matter, idealism to the primacy of consciousness. Theism, which maintains that matter was created by a supernatural (divine) consciousness, is taken to be the chief form of idealism; under the title 'objective idealism' this is sometimes distinguished from 'subjective idealism,' the view that the material world exists only for the individual mind (see Idealism). Although these two versions of idealism do not appear to make consciousness the 'source' of matter in the same sense, it is even less clear in what way materialism takes matter to be the 'source' of consciousness. Dialectical materialists apparently mean to endorse whatever account of mind results from empirical scientific investigations; but they are confident that we already know enough to exclude supernaturalist, dualist or idealist accounts.

Dialectical materialists sometimes also give the 'primacy of matter over consciousness' an epistemological interpretation. Idealists are charged with a tendency to scepticism concerning knowledge of the material world, whereas materialists maintain that the material world is knowable through empirical science. This confidence is supported by appeal to the practical successes of empirical science, by which is meant both the results of experimentation (which involve the experimenter's practical interaction with the world) and the technological fruits of empirical science. Practice is asserted to be the sole criterion of truth; doubts and questions which cannot be given a practical significance are to be dismissed. The sceptical doubts of idealistic philosophy are held to be refutable in this way.

Dialectical materialists also insist that thought bears a certain determinate relation to matter, serving as its 'image' or 'reflection'. The world of consciousness is the material world 'translated into forms of thought'. The point of this last phrase seems to be that thought is given in certain determinate forms, which bear determinate relationships (especially developmental ones) to each other, whose subject matter is 'dialectics'. The epistemology of dialectical materialism opposes some traditional forms of empiricism in that it emphasizes the independent importance of determinate (dialectical) forms of thinking in the theoretical process through which science comes to know reality.

Engels, for example, argues that Marx's theory of capitalism is 'dialectical' by emphasizing structural parallels between Hegel's *Science of Logic* and Marx's *Capital*.

2 Dialectics

If the opposition of idealism and materialism concerns the fundamental question of philosophy, the opposition between metaphysics and dialectics concerns the fundamental issue of method. The 'metaphysical' method (as Engels calls it, borrowing the term from Hegel's critique of pre-Kantian philosophy) is identified with the mechanistic programme of early modern science, which is taken to have been discredited by such nineteenth-century discoveries as electromagnetic field theory. But following Engels, dialectical materialism upholds (at least a modified version of) the critique of early modern science presented by German idealism and its 'philosophy of nature', which opposes formalism and reductionism and emphasizes phenomena of organic interconnection and qualitative emergence. Early modern mechanism was a form of atomistic reductionism, which takes the world to consist fundamentally of material corpuscles interacting and combining in various ways according to unchanging laws. Essences or natural kinds no longer have any fundamental role to play in science, since all qualitative differences are ultimately to be reduced to quantitative differences at the fundamental level of explanation.

This picture was rejected by German idealists such as Schelling and Hegel. Under the title 'philosophy of nature' (*Naturphilosophie*) they attempted to develop a comprehensive picture of nature as the expression of an absolute reality transcending nature. Although Engels could not accept the anti-naturalist form of idealist philosophy of nature, he shared many of its aims and wanted to vindicate its 'dialectical' principles against contemporary materialists such as Ludwig Büchner and Erich Haeckel, whom he regarded as still fettered by the old mechanism or 'metaphysics'. From the idealists, Engels retained the idea that the natural world involves forms of organization that are essentially different in quality, each with its own distinctive regularities; they are irreducible to any fundamental form of matter or unchanging set of laws. Engels' new materialism holds that new essences or forms of organization may emerge in the course of history. Thus the commonest charges dialectical materialists bring against metaphysical materialism are that it ignores the fundamentally developmental nature of matter, that it tries to reduce all change to quantitative change, and that it fails to recognize internal contradictions in the nature of material things as the fundamental source of change. The antidote is to recognize and employ the dialectical laws of thought.

3 Dialectical laws

First law: the unity of opposites. The nature of everything involves internal opposition or contradiction. This is probably the most obscure dialectical idea, because it is the most wilfully paradoxical, seeming at times to amount to a flat denial of the logical law of non-contradiction. Echoing Hegel's 'resolution' of Zeno's paradoxes, for example, Engels writes: 'Motion itself is a contradiction… coming about through a body at one and the same moment of time being both in one place and in another place, being at one and the same place and also not in it'. Regarded as an attempted solution to Zeno's paradoxes, such remarks are at best extremely obscure; they do apparently intend to challenge the canonical status of traditional logic, and put in question even the law of non-contradiction.

Hegel, however, sometimes draws a distinction between the way the principles of traditional logic as they are 'meant' (so understood, he regards them as trivially true and of no philosophical interest) and a deeper 'metaphysical' significance that may be discovered in them (see Hegel, G.W.F. §6). Understood in the latter way the laws are false, but importantly so, because they represent the standpoint of the abstract 'understanding' which fails to comprehend the essential interrelatedness of things and their necessary development. Understood as a denial of traditional logic on this metaphysical interpretation, the law of the unity (or interpenetration) of opposites claims something which, if not always entirely clear, is at least plainly not (in the traditional sense) logically self-contradictory: namely, that the essences of things include differing and opposed, often complementary, elements or processes, and that essential change or development in the world occurs through the conflict of these opposed moments. It is not hard to see how an emphasis on the organic could give rise to such a view, since the life of an organism often consists in a dynamic equilibrium of opposed processes (Engels' example is nutrition and excretion). Further, it is not difficult to see how a representation of changes in the world as resulting from conflicting local tendencies deemed essential to a certain organization constitutes a clear alternative to a reductive

mechanistic explanation in terms of unchanging basic elements following general laws. Some later dialectical materialists distinguish 'non-antagonistic' contradictions from 'antagonistic' ones: that is, opposed processes that harmonize and reinforce one another from those which tend towards the essential destruction of the complex they constitute. They also distinguish 'internal' contradictions from 'external' ones, the former constituting the essential nature of a phenomenon and the latter involving the contingent collision of things with different natures.

Second law: quantity and quality. Quantitative change always eventually leads to qualitative change or development. Engels cites such phenomena as the suddenness of the transition between the solid, liquid and gaseous states of substances as their temperatures change. The important thing in such examples is that purely quantitative changes lead to the emergence of new qualities which cannot be adequately grasped in terms of the concepts in which the purely quantitative changes can be formulated. Once again, the rejection of mechanistic reductionism is evident both in the insistence on the existence and importance of qualitative differences and in the idea that what is ultimately important, even about the quantitative changes studied by the mechanists, is the way they lead to qualitative differences.

Third law: negation of the negation. Change negates what is changed, and the result is in turn negated, but this second negation leads to a further development and not a return to that with which we began. Marx, in Volume 1 of *Capital* (1865), famously described capitalist property as the 'negation' of individual property based on the labour of the proprietor, and communist property as the 'negation' of capitalist property, hence as the 'negation of the negation'. The point once again is to emphasize the emergence of what is qualitatively new, and the fact that this emergence is to be understood in terms of the specific essence of that which is in process rather than in terms of general laws applying to simple elements of which it is composed.

The three dialectical laws are often seen to operate in close association, as when 'negation' is the result of antagonistic or 'contradictory' tendencies essential to something, which thereby change it into something qualitatively different. In general, dialectical thinking tries to explain change as the outcome of conflicts which emerge from the essence of something, which are resolved when its essence is transformed and it attains a new stage of its development. This last idea is sometimes presented by popular expositors of 'dialectics' in the jargon of 'thesis-antithesis-synthesis'. This particular bit of grotesque jargon, however, is almost never found in the writings of dialectical materialists themselves. It was never used by Hegel or Engels, and was used by Marx only once, solely for the purpose of ridicule. Its employment nearly always betrays either ignorance of dialectical thinking or contemptuous hostility to it - usually both these at once.

4 Historical materialism

The most prominent application of dialectical materialism is to the historical development of society, in the shape of Marx's 'materialist conception of history'. It has sometimes been made a matter of controversy how far Marx's economic theory of the history of capitalism and his shrewd analyses of contemporary political events actually embody his theoretical programme, the main lines of which are stated in the *German Ideology* (1845) and the Preface to *Toward a Critique of Political Economy* (1859). The fundamental thing in human history is the productive powers of society and their tendency to grow. Productive powers at a given stage of development determine the nature of human labouring activity because labour consists in the exercise of precisely those powers. A given set of productive powers also thereby favours certain 'material relations of production', forms of human cooperation or division of labour, which are not directly part of them, but facilitate their employment to a greater degree than rival forms would do. They thereby also favour certain 'social relations of production', systems of social roles relating to the control of the production process and the disposition of its fruits. It is this system of social relations of production which Marx calls the 'economic structure of society' which forms the 'real basis' of social life on the materialist theory, conditioning 'superstructures' such as the political state and the 'ideological' forms of consciousness found in religion, philosophy, morality and art.

Within the framework of any system of social relations of production, society's productive powers expand at a greater or lesser rate, depending on the historical circumstances, including the social relations themselves. Eventually, however, a given set of social relations of production are outgrown or rendered obsolete by the productive powers. The prevailing relations either make it difficult to employ the existing powers or else fetter the further development of these powers. Powers and relations of production thus come into conflict or 'contradiction'; an 'epoch of social revolution' begins. The outcome of the conflict is the transformation of the

relations of production to bring them into line with the productive powers so as to facilitate the further expansion of these powers. Changes in the superstructure of society, including its political and legal institutions, are to be explained by the required changes in the social relations of production. The mechanism by which these adjustments are to be carried out is the class struggle. At a given stage of history, that class is victorious whose class interests consist in the establishment of that set of production relations which best suits the productive powers at that stage.

Historical materialism exemplifies the principles of dialectical materialism: each mode of production has its own essential laws and tendencies; each has inherent tendencies to undermine itself and change into something new, and these tendencies operate through oppositions or contradictions essential to it. Although Marx occasionally applies it to pre-modern societies, this theory of history was obviously suggested to him by the rise of capitalism; Marx envisages the overthrow of capitalism and the rise of socialism as following the same pattern of historical development.

5 Materialism and practice

Beginning with Marx's 'Theses on Feuerbach' (1845), one prominent claim made by dialectical materialists is that their view alone provides for an adequate relation of philosophy to 'practice'. In the 'Theses', Marx meant to side with Hegelian philosophy against the empiricism of Feuerbach and eighteenth-century mechanistic materialism, holding that its conception of our relation to nature is exclusively contemplative, while until now only idealism has grasped the 'active side' of this relation (albeit inadequately). An adequate grasp of practice, in his view, is based on the understanding of philosophy's involvement in the fundamental social struggles brought to light by historical materialism. Only in this way do we become fully conscious of ourselves as natural, and naturally historical, beings.

Lenin's theory of 'partisanship' was an attempt to develop this important side of Marxism: the recognition that all philosophical thinking must be understood in its social significance and its relation to historical struggles. The 'partisanship' of philosophy is all too easy for us to stigmatize as simply an excuse for constraining philosophical thought through party dogma. But the philosophical import of the doctrine (as distinct from its political abuses) is rather to liberate thinking by making it more self-aware as regards its class basis and social situatedness (see Lenin, V.).

6 The fate of dialectical materialism

The plausibility of any theory as programmatic as dialectical materialism must finally depend on how fruitful it is in its actual employment. As the official philosophy of Soviet Marxism-Leninism, however, dialectical materialism was doomed to stasis and sterility by the repressive and authoritarian political conditions under which its practitioners were forced to operate. Thus a philosophy whose spirit was to challenge traditional religious authority and to exalt the fact of qualitative novelty and ceaseless progressive development has become our century's most notorious example of ossified dogmatism, incapable either of internal development or of response to external change, often degenerating into the mechanical repetition of empty phrases borrowed from the science and philosophy of an earlier century. Yet this historical irony too easily obscures the important fact that the basic aims and principles of dialectical materialism are very much in harmony with any outlook which perceives a fundamental opposition between scientific theories and religious myths, tries to address the scientific challenges posed by the failure of the seventeenth and eighteenth century mechanistic programme, and seeks a scientific metaphysics as the basis for an enlightened view of the world.

See also: Engels, F; Marx, K; Materialism; Plekhanov, G.V.

ALLEN W. WOOD

References and further reading

Acton, H.B. (1955) *The Illusion of the Epoch: Marxism-Leninism as a Philosophical Creed*, London: Cohen & West.(An informed but critical discussion of Marxist philosophy.)

Afanasyev, V.G. (1980) *Marxist Philosophy*, 4th edn, Moscow: Progress.(Standard exposition of Soviet Marxist-Leninist philosophy.)

Bochenski, J.M. (1963) *Dialectical Materialism*, Dordrecht: Nijhoff.(A very polemical discussion of Soviet

dialectical materialism.)
Cohen, G.A. (1978) *Karl Marx's Theory of History*, Princeton, NJ: Princeton University Press.(An influential interpretation/defence of Marxian historical materialism.)
Cornforth, M. (1971) *Dialectical Materialism*, New York: International Publishers.(A standard survey of Marxist-Leninist philosophy by an important British exponent.)
Engels, F. (1878) *Anti-Dühring*, Moscow: Progress, 1962.(The major philosophical work by Engels.)
Engels, F. (1875-82) *Dialectics of Nature*, New York: International Publishers, 1973.(The chief source of Engels' philosophy of nature.)
Engels, F. (1886) *Ludwig Feuerbach and the End of Classical German Philosophy*, in K. Marx and F. Engels, *Selected Works in One Volume*, London: Lawrence & Wishart, 1968, 584-622. (First published as a review, this is a history of German philosophy, placing Marx's materialism at its culmination.)
Hegel, G.W.F. (1812-16) *Science of Logic*, trans. A.V. Miller, Atlantic Highlands, NJ: Humanities Press, 1993. (Mentioned in §1 above. Engels parallelled this work to Marx.)
Lenin, V.I. (1960) *Collected Works*, Moscow: Progress.(Vol. 38 contains the *Philosophical Notebooks* which contain his sympathetic reflections on Hegelian speculative logic.)
Lenin, V.I. (1909) *Materialism and Empirio-Criticism*, trans. A. Fineberg, New York: International Publishers, 1948.(Lenin's critique of the Russian Machists contains some of his main discussion of philosophical questions.)
Marx, K. and Engels F. (1975-) *Marx Engels Collected Works*, New York: International Publishers.(The comprehensive edition in English.)
Oizerman, T. (1984) *The Main Trends in Philosophy*, Moscow: Progress.(A survey of philosophy from a Marxist-Leninist standpoint during the last years of the Soviet period.)
Plekhanov, G. (1961) *Selected Philosophical Works*, trans. A. Rothstein, New York: International Publishers.(The chief Russian source of much of Marxist-Leninist thought.)
Ruben, D. (1979) *Marxism and Materialism*, 2nd edn, Brighton: Harvester.(A sympathetic but critical discussion of Marxist materialism.)
Wetter, G. (1959) *Dialectical Materialism*, trans. P. Heath, New York: Praeger.(A critical study of Marxism-Leninism.)
Wood, A. (1981) *Karl Marx*, London: Routledge & Kegan Paul.(A comprehensive study of Marx's philosophical thought.)

Dialectical school

An offshoot of the Megarian school, and active c.*350-250* BC*, the Dialectical school was an important precursor of Stoic logic. Its leading members were Diodorus Cronus and Philo. Its interests included modality, conditionals and logical puzzles. Its primary allegiance was probably Socratic, but with some additional debt to the Eleatic Zeno.*

The Dialecticians have regularly been assumed to be identical with the Megarians (see Megarian school), but several testimonies show them to be a distinct and rival school. The Megarians are even reported to have filched pupils from them (Diogenes Laertius, II 113). Although linked by a common interest in sophisms, by a shared Socratic background tinged with Eleaticism and by the fact that Clinomachus of Thurii, official founder of the Dialectical school, had been a pupil of the Megarian founder Euclides, the two schools exhibited largely different philosophical outlooks. Where the Megarians emphasized and developed Socrates' specifically moral teachings, the Dialecticians seem to have stressed the intrinsic value of dialectical activity - argument by question and answer. Since Socrates himself had called dialectical discussion the greatest human good (Plato, *Apology* 38a), the school's stance can be seen as a bid for Socrates' mantle, in direct competition with the Megarians and other Socratic schools (see Socratic schools). The school's appearance in a list of 'ethical' sects (Diogenes Laertius, I 18) can therefore be taken seriously. If Diodorus, in devising his own four arguments against motion, was also indicating allegiance to Zeno of Elea, the acknowledged founder of dialectic, the move was perhaps legitimized by Plato's *Parmenides*, where the young Socrates welcomes a lesson in Zenonian dialectic. What united the school was its commitment to dialectical virtuosity as such. On individual points of logical theory they differed widely.

Clinomachus is reported as 'the first author of treatises on propositions, predicates etc.' (Diogenes Laertius, II 112). The school's pioneering work on the nature of predicates is confirmed by Seneca (*Letters* 117.12), but it is also clear from Diodorus' and Philo's work on conditionals and modal logic that they anticipated the Stoics, and differed from Aristotle, in treating entire propositions rather than terms as the basic units in logic on this and on Diodorus' other work, see Diodorus Cronus and Philo the Dialectician). Their theory of sign inference also paved the way for Stoicism, as did their analysis of inferential validity if, as Ebert (1991) has argued, a number of references to 'the dialecticians' in Sextus Empiricus are to this school. Zeno of Citium, founder of Stoicism, studied under Diodorus.

Diodorus' Master Argument, defending his definition of 'possible' as 'what is or will be true', was their most celebrated product. A less known companion piece, also attributed to the school, is the Mowing Argument, which tries to devalue the word 'perhaps' by arguing that every prediction must be either simply true or simply false. Although both arguments lent themselves to deterministic interpretations, the different modal views of Philo and the little-known Panthoides did not, and the school as such clearly had no commitment to determinism.

Another area of dialectical activity, with strong Socratic antecedents, was the framing of definitions. A Dialectical definition of 'expertise' (*technē*) is recorded, and Aristotle (*Topics* VI 10) criticizes that of 'life' by the early Dialectician Dionysius of Chalcedon.

DAVID SEDLEY

References and further reading

Aristotle (*c.* mid 4th century BC) *Topics*, trans. in J. Barnes (ed.) *The Complete Works of Aristotle*, vol. 1, Princeton, NJ: Princeton University Press, 1984.(Aristotle's handbook on dialectics.)

Diogenes Laertius (*c.* early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(Book II 106-12 in volume 1 covers the main Dialectical philosophers.)

Ebert, T. (1991) *Dialektiker und frühe Stoiker bei Sextus Empiricus (Dialectians and Early Stoics in Sextus Empiricus)*, Göttingen: Vandenhoeck & Ruprecht.(Pioneering attempt to recover evidence on the Dialectical school; an earlier version of Part 1, on signs, had appeared in English in *Oxford Studies in Ancient Philosophy* 5 (1987): 83-126.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae (The Fragments of Socrates and the Socratics)*, Naples: Bibliopolis.(Volume 1 includes the fullest available collection of Greek and Latin testimonies on the

Dialecticians, although there classed as ‘Megarians’; many testimonies on the school are still missing.) P
Primavesi, O. (1992) ‘Dionysios der Dialektiker und Aristoteles über die Definition des Lebens’ (Dionysius the Dialectician and Aristotle on the Definition of Life), *Rheinisches Museum für Philologie* 135: 246-61. (Discovers a Dialectical definition in Aristotle, I VI 10.)
Sedley, D. (1977) ‘Diodorus Cronus and Hellenistic philosophy’, *Proceedings of the Cambridge Philological Society* 203: 74-120.(On the separation of the Dialectical from the Megarian school, and on Diodorus’ influence.)
Sextus Empiricus (2nd century AD) *Works*, trans. R.G. Bury, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933-49, 4 vols.(The main passages in which Ebert (1991) has detected reports of the Dialectical school are Outlines of Pyrrhonism II 110-11, 146-50, 229-35 and Against the Professors VIII (Against the Logicians II) 93-8, 108-17, 124-8.)

Dialogical logic

Dialogical logic characterizes logical constants (such as 'and', 'or', 'for all') by their use in a critical dialogue between two parties: a proponent who has asserted a thesis and an opponent who challenges it. For each logical constant, a rule specifies how to challenge a statement that displays the corresponding logical form, and how to respond to such a challenge. These rules are incorporated into systems of regimented dialogue that are games in the game-theoretical sense. Dialogical concepts of logical consequence can then be based upon the concept of a winning strategy in a (formal) dialogue game: B is a logical consequence of A if and only if there is a winning strategy for the proponent of B against any opponent who is willing to concede A. But it should be stressed that there are several plausible (and non-equivalent) ways to draw up the rules.

1 Main characteristics

Logic is to provide us with viable concepts of logical consequence, logical truth, consistency and so on, and to that end to elucidate the meaning of logical constants (see Logical constants). Dialogical logic, which is the logical part of dialogue theory, goes about this by assigning to each logical constant a clear 'meaning-in-use', where the context of use is one of critical dialogue arising from a conflict of opinions about a statement: the initial thesis. This thesis - the question at issue, one might say - is challenged by the opponent (*O*) and defended by the proponent (*P*). The opponent may or may not have granted some statements in advance: the initial concessions or hypotheses of the dialogue. The dialogical approach is to be compared to, and contrasted with, semantic or model-theoretic characterizations of the logical constants (see Model theory) and deduction-theoretic characterizations (see Proof theory; Natural deduction, tableau and sequent systems).

Each logical constant is characterized by the specific modes of challenge of which the participants may avail themselves to attack or question a statement with that particular logical constant as its principal operator; as well as by the specific responses to these challenges. For instance, a conjunctive statement '*A* and *B*' can be challenged in two ways: by questioning *A* and by questioning *B*. The response to the first type of challenge consists of *A*, whereas the response to the second type of challenge consists of *B*. In the case of a disjunctive statement '*A* or *B*', however, the only type of challenge pertains to this statement in its totality, but the responses are the same, with the distinctive difference that in this case the choice between *A* and *B* is the respondent's, rather than the challenger's. A survey of these so-called logical rules is given in the table.

The logical rules are not sufficient, for the meaning of each logical constant also depends on global characteristics of the dialogues. Structural rules stipulate who may challenge what and how often, whether responses may be repeated, and so on. Finally, rules for winning and losing determine the situations in which a participant may be declared to be the winner (or loser) of the dialogue. A full set of rules of all three types (see the table and lists of rules below) defines a 'dialogue game', or 'dialectic system'.

If a dialogue game makes use of an interpreted nonlogical vocabulary, the game and its dialogues are called material; otherwise they are called formal. In material games, typically, elementary statements can be defended by pointing out their (extra-dialogical) truth. Thus observation and other ways of fact-finding are connected with material dialogues. In formal dialogue games these material modes of defence are not available.

2 An example

The following formal dialogue game yields intuitionistic (constructive, effective) predicate logic (see Intuitionism). It is based on a first-order language for predicate logic with an infinity of individual constants *a*, *b*,…. Formulas are denoted as *A*, *B*,…. The result of substituting an individual constant *a* for the free variable *x* in the formula *A* is denoted as $A[a/x]$.

This game is asymmetric in the following sense: at all times *P* has a statement to defend, whereas *O* never has a thesis to defend. However, *P* must have some means to exploit the initial concessions. To that end *P* is allowed to question the concessions in order to generate more detailed concessions. The corresponding moves of question (by *P*) and answer (by *O*) are formally identical to the moves of challenge (properly so called) by *O* and response (or, defence) by *P*, but the intuitive motivation is different. Thus the asymmetry of the game remains hidden as far as the logical rules are concerned. It is clearly borne out, however, by the structural rules and by the very different

ways *P* and *O* can win a dialogue.

Logical Rules

	Statement by proponent	**Challenge/question by opponent**	**Defence/answer by proponent**
Rule $_{\rightarrow}$	$A \rightarrow B$	A	B
Rule $_{\neg}$	$\neg A$	A	$\perp$ (elementary statement of absurdity)
Rule $_{\vee}$	$A \vee B$	?	A or B (proponent chooses)
Rule $_{\wedge}$	$A \wedge B$	A?	A
		or B? (opponent chooses)	B
Rule $_{\forall}$	$\forall x A$	a? (opponent selects a constant)	$A[a/x]$
Rule $_{\exists}$	$\exists x A$	?	$A[a/x]$ (proponent selects a constant)
Rule $_{\mathrm{El}}$	elementary statement	?	none

Structural Rules

1 There are two participants: the proponent, *P*, who defends a thesis, and the opponent, *O*, who may or may not have granted a number of concessions.
2 The participants move alternately. *O* makes the first move: a challenge of the thesis.
3 Each move by *O* is a challenge or an answer according to a logical rule.
4 Each move by *P* is a question or a defence according to a logical rule, or a winning remark, that is, either '*Ipse dixisti!*' ('You said so yourself!') or '*Absurdum dixisti!*' ('You said something absurd!').
5 *P* is not allowed to question *O*'s elementary statements.
6 A winning remark '*Ipse dixisti!*' is allowed only if the most recently challenged statement of *P*'s can also be found among *O*'s concessions.
7 A winning remark '*Absurdum dixisti!*' is allowed only if $\perp$ can be found among *O*'s concessions.
8 The only statement *P* is allowed to defend is the one that was most recently attacked by *O*.
9 Except for the initial challenge, each move by *O* is to consist of a reaction to the immediately preceding move by *P* (answering *P*'s question, or challenging a newly introduced statement).
10 Before the onset of the dialogue proper, *P* is to announce a limit on the number of *P*-moves. *P* is not allowed any moves beyond that number.

Rules for Winning and Losing

1 *P* wins by making a winning remark.
2 *O* wins whenever it is *P*'s turn to make a move but no legal move is available.

3 Motivations and applications

The connection between dialogues and logic, which goes back to Aristotle's *Topics*, was reintroduced by Paul Lorenzen in 1958 (see Lorenzen and Lorenz 1978). Originally, the motivation was to be found in the philosophy of mathematics: the dialogues were to yield a criterion of constructivity and to justify a constructive logic without

recourse to Brouwerian solipsism (see Intuitionism). But studies by Kuno Lorenz and others were not slow to point out that with a little fiddling of the structural rules the dialogue games could be forced to yield classical logic instead, although constructive logic seemed more 'natural' from the dialogical point of view.

In the meantime it had become clear how logical concepts were to be defined dialogically. We have seen above how a definition of logical consequence can be framed in terms of the existence of a winning strategy for the proponent in a formal game. The case of an empty set of concessions gives us a definition of logical truth. Consistency of a set of statements can be defined as the existence of a winning strategy for an opponent who concedes these statements against a proponent who defends $\perp$. This strategy shows how *O* can maintain this position without conceding absurdity. Clearly, different formal dialogue games may give rise to different concepts of logical consequence; to different logics, one might say.

The challenges and responses given in the logical rules of dialogical logic are closely similar and often identical to those given in Hintikka's rules for game-theoretic semantics (see Semantics, game-theoretic). But in game-theoretic semantics one tries to exploit the concept of a game to construct a semantic theory (for natural language), whereas dialogical logic is rather to be conceived as providing an alternative to the semantic approach. Even so, the two approaches are closely related.

Later on dialogical logic came to be part of a programme for a normative reconstruction of the language of science, ethics, philosophy and politics that was worked out by Paul Lorenzen and others see Lorenzen and Schwemmer 1973) in the so-called Erlangen school of German constructivism (see Lorenzen, P.). The dialogue rules contribute to this programme by giving us reconstructed versions of the logical constants.

More generally, dialogical logic has been influential in the development of argumentation theory, including the theory of argumentative discussions and fallacy theory.

ERIK C.W. KRABBE

References and further reading

Barth, E.M. and Krabbe, E.C.W. (1982) *From Axiom to Dialogue: A Philosophical Study of Logics and Argumentation*, Berlin and New York: De Gruyter.(Chapters 3 and 4 provide normative foundations, from the point of view of a theory of argumentative discussions, for a number of dialectic systems. Chapters 5, 7 and 11 discuss the metatheory of dialectic systems and present proofs of equivalence with model-theoretic and deduction-theoretic systems. A standard introductory course in logic is presupposed.)

Felscher, W. (1986) 'Dialogues as a Foundation for Intuitionistic Logic', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, vol. 3, *Alternatives to Classical Logic*, Dordrecht: Reidel, 341-72.(A useful and concise introduction to the subject, with a useful bibliography. A bit more technical.)

Haas, G. (1984) *Konstruktive Einführung in die formale Logik (A Constructive Introduction to Formal Logic)*, Mannheim: Bibliographisches Institut, Wissenschaftsverlag.(This is both an introduction to logic and an introduction to dialogical logic, with many examples, detailed explanations and proofs.)

Krabbe, E.C.W. (1985) 'Formal Systems of Dialogue Rules', *Synthèse* 63 (3): 295-328.(Includes a survey of options in constructing a dialectic system, with a closer study of constructive systems and equivalence proofs. Only the last part is highly technical. Useful bibliography.)

Lorenzen, P. (1969) *Normative Logic and Ethics*, Mannheim: Bibliographisches Institut.(An early and brief exposition in English.)

Lorenzen, P. (1986) *Lehrbuch der konstruktiven Wissenschaftstheorie (Textbook of Constructive Philosophy of Science)*, Mannheim: Bibliographisches Institut, Wissenschaftsverlag.(Contains the first part of (Lorenzen and Schwemmer (1973).)

Lorenzen, P. and Lorenz, K. (1978) *Dialogische Logik*, Darmstadt: Wissenschaftliche Buchgesellschaft.(The essential writings of Lorenzen and Lorenz from the genesis of contemporary dialogical logic in 1958 to 1973.)

Lorenzen, P. and Schwemmer, O. (1973) *Konstruktive Logik, Ethik und Wissenschaftstheorie (Constructive Logic, Ethics and Philosophy of Science)*, Mannheim: Bibliographisches Institut, 2nd edn, 1975.(The 'bible' of Erlangen constructivism. The first part has a mature exposition of Lorenzen's dialogical logic and the way it functions in the Erlangen programme mentioned in §3.)

Walton, D.N. and Krabbe, E.C.W. (1995) *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*, Albany, NY: State University of New York Press.(Pursues the integration of dialogical logic with other dialogue theory, especially Charles Hamblin's formal dialectic.)

Dicey, Albert Venn (1835-1922)

A.V. Dicey, who held the Vinerian Chair of English Law at Oxford University between 1882 and 1909, is widely regarded as the high priest of orthodox constitutional thought in Britain. Living through a period of great political and economic change, Dicey recognized that the duty of the constitutional lawyer could no longer be one of simply venerating Britain's ancestral, unwritten constitution. He appreciated the need to try to lay bare its legal foundations and to identify its fundamental principles. In embarking on this exercise, Dicey, a follower of John Austin, employed an analytical method which treated law as a datum to be analysed and classified and which served to furnish a descriptive account of how law's various divisions fit together to provide an ordered whole. Dicey was the first to apply this legal positivist method to the study of constitutional law in Britain. The method became so ensconced in legal thought in twentieth-century Britain that today lawyers scarcely acknowledge any work in constitutional law which predates Dicey's work. It is as though he invented the subject.

In his classic text, *The Law of the Constitution* (1885), Dicey defined the subject of constitutional law as the study of the rules which affect the distribution or the exercise of the sovereign power of the state. These rules comprise two sets: laws in the strict sense which are enforced by courts; and conventions, understandings, habits and practices which are not so enforced but which regulate the conduct of the members of the sovereign power. Through this formalistic and positivistic distinction between law and convention, Dicey was able authoritatively to demarcate the boundaries of the subject.

Dicey's influence, however, has not been confined to his method. His view of the character of constitutional law was shaped by a particular political outlook which, on the face of it, seemed far removed from an Austinian method. This outlook was a form of 'diffused Burkeanism' (see Burke, E.) which was a characteristic of much of nineteenth-century English political thought. He believed that the English system - resting on a love of order, a Conservative spirit, and a belief in ordinary morality as a guide to public life - represented the climax of political achievement. These qualities were to be found, above all, in the common law heritage and, owing to the fact that it is a judge-made constitution, within the British constitution itself.

In *The Law of the Constitution* Dicey identified three guiding principles underpinning the British constitution: the legislative sovereignty of the Queen-in-Parliament; the universal rule throughout the constitution of ordinary law (otherwise understood as the rule of law); and the ultimate dependence of constitutional conventions on the law of the constitution. His objective was to show how, once their interlocking nature is understood, these principles were conducive to the promotion of liberty. Parliamentary sovereignty, the right of Parliament to make or unmake any law whatsoever, was the dominant characteristic of the political institutions. Though an instrument well adapted for the establishment of democratic despotism, Dicey argued that the sovereignty of Parliament, as contrasted with other forms of sovereign power, favoured the supremacy of law: first, because the constituent parts of Parliament provided checks and balances (the idea of 'self-correcting democracy'); and, second, because the rule of law implies the predominance of rigid legality throughout institutions and thus increases the authority of Parliamentary sovereignty (see Sovereignty §§2-3).

Dicey's formulation of the idea of the rule of law has been viewed as the most influential restatement of the principle since the eighteenth century. Nevertheless, it is not without its ambiguities. Dicey suggested that the concept had three meanings: the absolute supremacy of regular law as opposed to the influence of arbitrary power; the principle of equality before the law; and, finally, that the constitution is a product of ordinary law and thus that constitutional law is not the source but the consequence of individual rights. Running throughout these meanings, there lies a relatively specific conception of law as a universal body of general rules which have evolved through time and which are enunciated by judges. Also implicit within, but fundamental to, the conceptual structure of Dicey's constitutional thought is the idea of the separation of powers. Dicey's view that Parliamentary sovereignty favours the supremacy of law seems to hold sway only in so far as we regard the legislative function as that of laying down general rules for society rather than incorporating the vesting of discretionary powers in executive agencies. Similarly, Dicey's notorious view that administrative law undermined the rule of law can readily be understood once administrative law is viewed as an infringement of the principle of the separation of judicial and executive powers (see Rule of law (Rechtsstaat) §§1-2).

The Law of the Constitution, though masquerading as an introductory treatise of positivist legal science, is in reality a work of considerable intellectual imperialism. The Whig interpretation of history which underpins it was later laid bare in *Law and Opinion in England during the Nineteenth Century* (1905), a work much criticized by historians and which in actuality simply explicates Dicey's intellectual and political assumptions. These assumptions, however, did not survive Dicey's own times. In 1914 he wrote a long introduction to the eighth edition of *The Law of the Constitution*. Taking the form of a lament, it mourned a loss of faith in Parliamentary government and a decline in veneration for the rule of law which Dicey believed had occurred since the mid-Victorian era and which he suggested were threatening the principles of the British constitution which he had first sought to formulate in 1885. However, Dicey's actual concerns mattered little. By this stage, the historical figure had more or less disappeared, to be replaced by an entity known as 'Dicey' which has lived on through successive generations of lawyers as the expositor of the fundamental principles on which the British constitution is founded.

See also: Austin, J.; Constitutionalism; Law, philosophy of; Legal positivism

MARTIN LOUGHLIN

List of works

Dicey, A.V. (1885) *An Introduction to the Study of the Law of the Constitution*, 8th edn, London: Macmillan, 1915.(The last edition which Dicey edited of his classic work.)

Dicey, A.V. (1905) *Lectures on the Relation between Law and Opinion in England during the Nineteenth Century*, London: Macmillan.(Based on lectures delivered at Harvard on the history of English law in the nineteenth century.)

References and further reading

Collini, S. (1991) *Public Moralists: Political Thought and Intellectual Life in Britain 1850-1930*, Oxford: Clarendon Press, ch. 7.(Subtle interpretation of Dicey's thought.)

Cosgrove, R.A. (1980) *The Rule of Law: Albert Venn Dicey, Victorian Jurist*, Chapel Hill, NC: University of North Carolina Press.(Competent intellectual biography.)

Craig, P.P. (1990) *Public Law and Democracy in the United Kingdom and the United States of America*, Oxford: Clarendon Press, ch. 2.(On Dicey and the idea of 'self-correcting democracy'.)

Lawson, F.H. (1959) 'Dicey Revisited', *Political Studies* 7: 109-26, 207-21.(General and orthodox appraisal.)

Loughlin, M. (1992) *Public Law and Political Theory*, Oxford: Clarendon Press, chaps 1, 7.(On Dicey's method, principles and values.)

Diderot, Denis (1713-84)

Chief editor of the great eighteenth-century Encyclopédie *(1751-72), Diderot set out a philosophy of the arts and sciences which took the progress of civilization to be a measure of mankind's moral improvement. He did not regard that progress as having produced universal benefits, however, and perceived the Christian religion which had accompanied it as morally harmful to those who subscribed to it and even more dangerous to societies thus far untouched by it. Religious dogmas tended to pervert the organic development of human passions, and secular education which presumed that all minds were equally receptive to instruction threatened to thwart the natural evolution of human faculties in other ways.*

Like Rousseau, Diderot subscribed to a philosophy of education which encouraged curiosity rather than promoted truth. He stressed the need for the adaptability of moral rules to the physiological characteristics of the individuals to whom they applied, pointing to a connection between human cultures and biology in a manner that would influence fresh outlooks upon the sciences of man at the end of the Age of Enlightenment.

1 Encyclopedic polymath

One of the most central figures of the European Enlightenment, Diderot was chiefly renowned in his own lifetime as the principal editor of the *Encyclopédie* (1751-72), constructed as a monument of human reason and invention, like the Crystal Palace that would be built in London a century later in celebration of an age of industry and commerce. More than any other philosopher of the eighteenth century, he shaped his intellectual career around the production and circulation of books, which in progressive circles were held to be both the vehicle and measure of enlightenment for an epoch of greater literacy and prosperity than Europe had ever known before.

Appointed with Jean Le Rond D'Alembert as joint editor of the *Encyclopédie* in 1747, Diderot initially undertook to write more articles himself than any other contributor to this reasoned dictionary of the arts and sciences, whose publishing history from 1751 would eventually extend over a period of thirty years, and whose thirty-five volumes would come to embrace 20 million words and more than 2,000 plates. As the guiding spirit of that collective enterprise of men of letters, Diderot often reassembled the ideas of other thinkers, drawing upon ancient and modern sources alike, from Lucretius to Shaftesbury. Having begun his literary career around 1743 as a translator, he proved the Enlightenment's most expert literary ventriloquist, sometimes passing off borrowed material as if he had written it himself, sometimes ascribing his own ideas to others.

In part because so much of his best work was to appear posthumously, often in the form of fictional dialogues, Diderot came to be regarded as the Enlightenment's most eclectic thinker, displaying a special fondness for the pluralist dimensions of eclecticism in general in his own entry on that very subject for the *Encyclopédie*, itself indebted to other authors. In *Le Neveu de Rameau (Rameau's Nephew)* (1762-74) in particular, deemed by Hegel, Goethe and Marx to be the finest work of the whole Enlightenment, he portrays the character of the most explosively original and romantic personality of eighteenth-century European literature as at odds with itself and also with Diderot, who figures both as the author of the dialogue and a personage within it, depicted as conventional in upholding a bourgeois code of morals. As distinct from some of his equally celebrated contemporaries - such as Montesquieu, Voltaire or Rousseau - Diderot elaborated no single overriding doctrine, nor did he devote himself to one great intellectual crusade.

Especially well versed in the history of philosophy, he extolled its ancient and modern achievements alike, from Epicurean doctrines of continual flux, to Spinozistic conceptions of selfsufficiency, to the atheistic materialism of La Mettrie and Maupertuis. In his *Pensées philosophiques (Philosophical Thoughts)* of 1746, largely inspired by Shaftesbury, he put forward a set of aphorisms that expressed notions of natural law and ideals of natural religion, which could be engraved in the human heart without need of Christian scruples. In his *Lettre sur les aveugles (Letter on the Blind)* of 1749, he set out a case for the adaptability of the senses as evidenced by the fact that mental images could be formed without sight or conversations pursued without sound, believing that powers of perception were transitive across human faculties and did not themselves spring from sensory experience. The late eighteenth-century teaching of sign languages to the deaf, and the deep structural grammars of modern Chomskian linguistics, were to draw inspiration from such claims that fundamental human capacities are not determined by the contingencies of their exercise. At the time of its publication, however, Diderot's *Lettre sur les aveugles* only

managed to invite the suspicion that he was a materialist and atheist, on which grounds he was to be imprisoned in Vincennes for over three months.

In the longest of the contributions to the work of which he was chief editor, the article '*encylopédie*' itself, he expounded themes of practical or applied philosophy that could bring the arts and sciences of civilization together, by encouraging the interpenetration of the liberal with the mechanical arts and the cooperation of specialists across disciplines, along lines that bridged what we have come to term 'the two cultures'. In showing how science had been rendered useful to the public interest or common good through such inventions as the compass and the printing press, Diderot sought to provide tangible illustrations of a Platonic ideal of promoting virtue through knowledge. His goal, like that of so many other *philosophes* of the eighteenth century, was to cast off the veil of ignorance and lift the yoke of precedence, thus overcoming dead dogma. Almost as much as D'Alembert's preliminary discourse to the *Encyclopédie*, Diderot's own article on that very subject could be read as a general manifesto of the Enlightenment.

2 Speculating about politics

His political philosophy was first propounded in the *Encyclopédie*, particularly in articles such as *autorité politique*, which offers a theory of the social compact that binds both the prince and his subjects to one another without recourse to theological sanctions; and *droit naturel* (natural law), which sets out a theory of universal justice with respect to mankind's 'general will', a term of scant significance in the history of political thought before Diderot gave it fresh impetus. In 1774, in his *Observations sur le Nakaz*, or commentary on a new code of law for Russia as proposed by the empress Catherine the Great, he called for a programme of secular education to stimulate the civilization and selfreliance of that country through indigenous and small-scale enterprises, opposing the importation of Western physiocratic doctrines and their schemes of wholesale agrarian reform within a cultural environment to which they were ill-suited. Catherine had purchased Diderot's library and had invited him to St Petersburg, but she failed to convince him of the wisdom of her efforts to reform the political system of her nation along lines she imagined had already been achieved in the West. While advocating the emancipation of serfs, which was not to her liking, Diderot was convinced that material progress in Russia should proceed apace with its intellectual development rather than in accordance with alien precepts of modernity.

As editorial spokesman for the cause of enlightenment and a major contributor on political subjects to the *Encyclopédie*, Diderot stood for principles which appear to differ sharply from those of Rousseau, for fifteen years his closest companion, who had conceived the central idea of his own philosophy of history - which forms a moral critique of the progress of the arts and sciences - en route to visit his friend in prison. Whereas Rousseau decried the moral consequences of the progress of the arts and sciences, Diderot applauded them. Whereas Rousseau subscribed to ideals of spartan republicanism, Diderot was cosmopolitan and willing to encourage monarchical regimes so long as their authority was tolerant and benign, directed to the promotion of the public interest. Rousseau's notion of the general will, as an ideal of public engagement only conceivable to the citizens of selfgoverning states, was largely designed to contradict Diderot's conception of the same term as pertaining to the common humanity of individuals even when motivated by self-interest, in so far our notions of what is right and good do not, Diderot supposed, depend upon our political identities.

If, for Rousseau, men's morals were shaped by their politics, Diderot was instead convinced that politics were fundamentally shaped by morality. In his article *hobbisme* for the *Encyclopédie*, he contrasted Rousseau's idea that society rendered mankind malicious with a Hobbesian notion that war was our species' natural state, rejecting them both (see Hobbes, T.). Believing that conflict was not inescapable either in nature or in society, he upheld an alternative doctrine, which he drew ultimately from Pufendorf's *De iure naturae et gentium* (1672), that human frailties had rendered civil society necessary, while the natural sociability of selfishly motivated persons had rendered it possible and, at least in principle, beneficial as well (see Pufendorf, S.).

3 On primitivism, religion and sexuality

In his *Supplément au Voyage de Bougainville (Supplement to the Bougainville Journey)*, dating from the early 1770s but largely unpublished until after his death, Diderot was to portray the virtues of a primitive society in colours so attractive as to seem to shed doubt upon his allegiance to the moral splendours of civilization. This text, purporting to serve as an appendix to one of the earliest descriptions of Tahiti, contrasts an exotic world

characterized by the zealous satisfaction of bodily appetites with a familiar world wherein it is dictated that they should be repressed. Diderot's depiction of a society in which sexuality is celebrated, and within which incest, adultery and fornication are not held to be either criminal or sinful, forms a critique of monogamy, celibacy and sexual repression in the West at the same time as it offers tribute to the promiscuity of Polynesians.

The *Supplément*'s fictional opposition of two cultures - one stirred naturally by libidinous desire, the other artificially held in check by religion and hypocrisy - owes a debt to the *Lettres persanes (The Persian Letters)* (1721) of Montesquieu and more generally to eighteenth-century travellers' tales, either real or imaginary, which found European, and most particularly Christian, civilization to be at fault in the light of Oriental alternatives. In granting tangible presence and geographical specificity in the modern world to a population that still enjoyed the freedom of its state of nature, Diderot also seems to have drawn inspiration from Rousseau's *Discours sur l'inégalité (Discourse on the Origin of Inequality)* (1755). Like Rousseau, he judged that the moral values which prevailed in the West were often contrary to human nature, which they distorted through artifice, subterfuge and selfdeception.

Especially with respect to the sexual longings of individuals in Christian communities, and above all the warped personalities of their monastic priests, Diderot was convinced that in Western cultures there had arisen a great gulf between human physicality and morality. But he did not share Rousseau's view that such a gulf was unbridgeable. He imagined that Tahitian society showed how it could be overcome, in promoting the public interest out of individuals' sexual proclivities, thereby stimulating a socially welcome growth in population. The irreverent, sensuous, liberal, anti-colonial and utilitarian themes of the *Supplément au Voyage de Bougainville* pursue, in different idioms, a philosophy which springs from the same point of view that informed Diderot's contributions to the *Encyclopédie* and to which he was to subscribe virtually all his life - the belief that great public benefits may spring from unconstrained selfish sociability. In holding to such a doctrine, more often associated in diverse formulations with Bernard Mandeville and Adam Smith, Diderot espoused one of the most central moral philosophies of the whole Enlightenment.

4 Theory of education

At least some of the tensions which are manifest in his writings reflect the clash of principles and personalities that he perceived in the world at large. At once a materialist and an atheist, Diderot was determined to describe the behaviour of individuals empirically, in all their diverse, contrasting and irregular forms. Such diversity and contrast were features of human nature as much as of human history, he supposed and, in his *Réfutation de l'ouvrage d'Helvétius intitulé 'L'Homme' (Refutation of Helvétius' Work Man)* of 1774, he accordingly took issue with the famous dictum of Helvétius' *De l'homme*, published posthumously in 1773, that 'l'éducation peut tout' (see Helvétius, C.-A.). The proposition that education can achieve everything was false, according to Diderot, because it presumed that men and women were everywhere identical and that differences in character depended entirely upon the circumstances of a person's upbringing.

Because this supposition appealed to a notion of human equality, it was tempting to believe, in the light of it, that all individuals at birth shared the same capacities. But that was a delusion, Diderot contended, based on the false premise that each person was motivated only by his or her sensations which, through instruction and refinement, would become the grounds of every sort of rational and moral judgment. It was impossible, Diderot insisted, to pass directly from sensation to judgment or, in effect, from animal reflexes to human design. Intraspecifically, it had to be noted that only humans possess the capacity to reason and thus to formulate ideas and combine them into judgments. Interspecifically, it was necessary to remark that individuals exhibit their reason in different degrees, mental capacity as a whole being unequally distributed. In allowing with Helvétius that vices do not spring from human nature but are generally attributable to harmful social practices and teachings, there should be no presumption that all individuals have good minds, endowed with a latent natural genius just awaiting cultivation and training.

Diderot's critique of the philosophy of education of Helvétius bears a striking resemblance to Rousseau's challenge, made in *Émile* a decade earlier, of much the same theory of the formation of the human intellect and character already sketched in Helvétius' *De l'esprit* (1758). Rousseau and Diderot were each convinced that genius cannot be instilled by tuition and that education ought to nourish the natural dispositions and qualities of children rather than impose their tutors' designs and expectations upon them. Their common objections to Helvétius

nevertheless took different forms. While Rousseau decried materialism in general, Diderot took issue only with that crude variant which, on his understanding, reduced all thought to pure sensation. Whereas Rousseau mistrusted teachers who either believed or pretended that they spoke on God's behalf, Diderot was persuaded that God himself had been invented by man. He judged that Rousseau had fundamentally confused unsociability for selfreliance, and he showed no interest in promoting any scheme of negative education that might be fit for children freed from all dependence on others, like Robinson Crusoe learning to fend for himself.

Diderot's conception of human nature differs from Rousseau's in attributing more dynamic potentialities for change and self-improvement to each individual's seminal personality. If he had coined the term *perfectibilité*, which was in fact invented by Rousseau, it would have encapsulated his understanding of the spark of genius, the impulse of creativity and the moral formation of character. Yet both men agreed, in contrast with Helvétius, that education ought fundamentally to accord with human nature instead of recasting it. Each subscribed essentially to a belief in self-education, aiming to encourage children to follow their intuitions and thereby attain intellectual and moral maturity by being true to themseves. They also shared a profound mistrust of the superstitious idolatry that passed for a Christian education, proffered in priestly interpretations of the mysteries of Holy Writ. As opposed to the arcane dogmas of a revealed religion, they were drawn to the spectacle of creation and to the open book of nature. They both thought that education ought to be pursued endogenously, out of natural curiosity, accompanying the development of the human faculties, and never by indoctrination.

5 Vitalist materialism

It is hardly surprising that a philosopher who advocated that moral principles should be true to Nature, which he believed to be in a state of continual flux, should have found himself suspected of lack of consistency. Like his colleague and co-editor of the *Encyclopédie*, D'Alembert, he saw his own Age of Enlightenment as committed to a systematic spirit but not the spirit of system which had distinguished the principal contributions of seventeenth-century European philosophy. Although he was the author of a treatise on mathematics and geometry, he was more interested in matters of style, particularly with respect to painting, music and the theatre. His *Salons* (1759-81) comprise perhaps the principal contribution to art criticism in the eighteenth century, and his posthumously published *Essai sur la peinture (Essay on Painting)* (written in 1765) was to stir the imagination of Baudelaire and other admirers in the Age of Romanticism.

He is better described as an empiricist than as a rationalist, indebted more to the philosophical methods of Bacon and Locke than to those of Descartes and Leibniz. He was a materialist in denying the existence in man of an intangible spirit or soul, and over the course of his life he became a progressively more outspoken atheist, in rejecting all notions of either a transcendent or immanent God. As is most conspicuous from his so-called *Lettre à Landois* (1756) and his novella *Jacques le fataliste* (written in 1773) he was at once a philosophical determinist, convinced that Nature left no room for unpredictability or caprice, and at the same time a libertarian, who identified the fundamental dignity of human conduct in terms of autonomous agency and choice, in poignant recognition of the bittersweet paradox that to choose is to act from a motive beyond one's control.

Diderot's materialism drew more inspiration from the sciences of physiology and medicine of his day than from physics, since he perceived matter not as inert but as inherently active and self-animating, as a collecton of atoms in a seminal fluid in continual fermentation. He was struck by Haller's research on the sensitivity and irritability of organic matter and was most particularly impressed by Bordeu and the Montpellier School of physiological medicine, which excited his interest in the internal mechanisms that gave impetus to such matter and regulated its health. Like Montesquieu, he sought to investigate the dynamic forces and tensions which shape the conduct of both living bodies and social systems. To the extent that a number of human sciences would around the end of the eighteenth century come to address such issues as apparently more fundamental than the constitutional and legislative programmes of Enlightenment political theorists, Diderot may be regarded as a philosophical precursor of such change.

Michel Foucault's depiction in *Les mots et les choses: une archéologie des sciences humaines (The Order of Things)* (1966) of certain epistemic metamorphoses within the sciences of man that may be dated from around 1770 to 1800 could have embraced the materialist philosophy of Diderot in addition to the examples offered from economics, linguistics and biology (see Foucault, M.). Diderot's reflections on the organic forms and forces of social life and on the growth and decadence of civilizations might in the twentieth century have come to be termed

'structural-functionalism'. In his own day they illustrate an approach to the science of society which would progressively displace speculations about the moral dimensions of human conduct with investigations into its underlying, bodily or material, causes. They herald the advent of 'sociology' in the nineteenth century in so far as it was modelled upon eighteenth-century 'physiology'.

6 Central tenets in oblique articulations

Diderot's reflections on the arts, human nature and society were not only to strike his readers among romanticists and scientists alike as particularly modern. By virtue of his irregular and sometimes fitful literary style, his preference for the cut and thrust of the dialogue over the discursive finality of the treatise, and his pursuit of anonymity as the mysteriously unidentified author of his own texts, he anticipated some of the philosophical issues raised by postmodernist thinkers today, often in criticism of the so-called 'Enlightenment Project' as a whole. No major writer of the eighteenth century left so many of his principal works unpublished in his lifetime, even though several, including the best among them, had been drafted long before his death. No encyclopedist ever managed better to convey his own thoughts by way of plagiarism, insinuation, multiple cross-references and deliberately misleading notes and commentaries. In assuming the personae of historical figures whose ideas he transcribed in his own fashion, and of characters he invented in order to contradict himself, he succeeded more than any other eighteenth-century writer in bestowing independent life upon his work, in the act of creation releasing it from the pen of its author. Like Socrates in Plato's *Phaedrus* or Derrida in *De la grammatologie*, he sought to ensure that the dead hand of script would not count as his last word. Even through his writings Diderot's voice appears to speak aloud, weaving a conversation without end.

If coming to final conclusions was not his chief aim either as a philosopher or as an encyclopedist, a number of unsettled doctrines may nevertheless be ascribed to him as among his most significant contributions to eighteenth-century thought across several disciplines. First, he supposed that civilizations were like socially organic forms of life, marked by cyclical patterns of growth, maturity and decay, and by institutions which reflected the dynamic or repressive character of their moral or religious principles. In these beliefs, which above all inform his contributions, dealing with global history, to Abbé Raynal's *Histoire des Deux Indes (History of Two Indias)* (1772-80), he was guided both by Buffon's *Histoire naturelle (Natural History)* (1749-88) and by Montesquieu's *Esprit des lois (The Spirit of the Laws)* (1748).

Second, he held that biological attributes of human nature - including instincts for survival and companionship but also bodily weaknesses and mental stupidity endemic to solitary existence - made society both necessary and possible. His *Apologie de l'abbé de Prades*, together with several of his articles on political subjects for the *Encyclopédie*, make plain the influence upon him with respect to this subject of Pufendorf, as well as the double-edged nature of his rejection of Hobbes's and Rousseau's differing perspectives on mankind's isolated state of nature.

Third, he was convinced that moral rules which dictate human conduct could promote the general interest of a society only if they accorded with and gave free rein to the physical desires of its members. Both in *Le Rêve de d'Alembert* (1769) and the *Supplément au Voyage de Bougainville*, he put a case for the ultimate physicality of human nature, whose constituent matter he believed to be moved by vital forces which would ensure its natural development towards an appropriate morality unchecked by extraneous rules. Convinced that liberty and necessity went hand in hand, he saw certain features of Western civilization as injurious to freedom when they attempted to obstruct the determinate mechanisms of natural law, as was the case most particularly with religions whose codes of duty stifled the expression of usefully pleasurable passions, thereby harming both individual wellbeing and public peace. In the light of his philosophy, he would have emblazoned our churches and their attendant institutions with warnings of the hazards they pose to our health. He thought Christianity was not only the opiate of the people but a spiritually carcinogenic poison.

Fourth, he believed that political and social institutions needed to correspond as well to a population's national or regional temperament and to its level of development. Just as punishments would be tailored to crimes, so laws should fit the communities they serve and direct. In his *Observations sur le Nakaz* he claimed that the character of the Russian peasantry was ill-suited to the drastic reforms proposed by French physiocrats, while in the *Histoire des Deux Indes* he argued that the fresh freedoms won in America were unlikely to take root in the acidic old soil of Europe. If he would not categorically exclude a favourable meaning for the term 'enlightened despotism',

Diderot believed that good government was never harsh and that no people ought to be governed too much.

All of these principles of organic growth and adapatibility led Diderot to perceive human liberty in much the same way that Montesquieu described laws - that is, as a set of necessary relations deriving from the nature of things. The character of individuals must not be forced, he thought, but, in passing gradually and without impediment through the seasonal changes of their constitutions, whole populations naturally change themselves. As their nature changes, so too should their political systems. The metamorphoses of human development ought only to record the various stages of our species' spontaneous generation.

See also: Education, history of philosophy of; Enlightenment, continental; Encyclopedists, eighteenth-century; Materialism; Natural law

ROBERT WOKLER

List of works

Diderot's numerous works include: *Pensées philosophiques* (1746); *Les Bijoux indiscrets* (1748); *Lettre sur les aveugles* (1749); *Lettre sur les sourds et muets* (1751); the *Encyclopédie* (1751-72); *Pensées sur l'interprétation de la nature* (1754); *Lettre à Landois* (1756); *Le Fils naturel* (1757); *Salons* (1759-81); *Le Neveu de Rameau* (1762-74); *Essai sur la peinture* (1765); *Le Rêve de d'Alembert* (1769); *Supplément au Voyage de Bougainville* (1772); contributions to *Histoire des Deux Indes* (1772-80); *Observations sur le Nakaz* (1773-4); *Réfutation de l'ouvrage d'Helvétius intitulé L'Homme* (1774); *Éléments de physiologie* (1774-80); *Essai sur la vie de Sénèque* (1778). Listed below are the most notable collections of his work.

Diderot, D. (1875-7) *Oeuvres complètes*, ed. J. Assézat and M. Tourneux, Paris.(The best more or less complete edition.)

Diderot, D. (1956) *Oeuvres philosophiques*, Paris: Garnier.

Diderot, D. (1963) *Oeuvres politiques*, ed. P. Vernière, Paris: Garnier.

Diderot, D. (1975-) *Oeuvres complètes*, ed. H. Dieckmann, J. Proust and J. Varloot, Paris: Hermann.(Still in progress, the most accurate edition.)

Diderot, D. (1982) *The Irresistible Diderot*, ed. J. Hope Mason, London: Verso.(Extensive extracts in translation, with introductions.)

Diderot, D. (1992) *Diderot's Political Writings*, ed. J. Hope Mason and R. Wokler, Cambridge: Cambridge University Press.

References and further reading

Benot, Y. (1970) *Diderot: De l'athéisme à l'anticolonialisme*, Paris: François Maspero.(A notable interpretation of Diderot's politics, with a detailed chronology.)

Chouillet, A.-M. (ed.) (1985) *Colloque International Diderot (1713-1784)*, Paris: Aux amateurs de livres.(A major collection of essays reflecting the state of Diderot scholarship today.)

Chouillet, J. (1973) *La Formation des idées esthétiques de Diderot*, Paris: Armand Colin.(The fullest treatment of Diderot's philosophy of art.)

Chouillet, J. (1977) *Diderot*, Paris: Société d'édition d'enseignement supérieur.(An introduction to Diderot's thought across the spectrum of his interests and enthusiasms.)

Crocker, L. (1966) *Diderot: The Embattled Philosopher*, New York: Free Press.(A well-informed intellectual portrait intended for a popular readership.)

Duchet, M. (1971) *Anthropologie et histoire au siècle des lumières*, Paris: François Maspero.(Contains a chapter on Diderot's anthropology.)

Gordon, D. and Torrey, N. (1947) *The Censoring of Diderot's 'Encyclopédie' and the Re-established Text*, New York: Columbia University Press.(On the tribulations of the publishing history of Diderot's chief work.)

France, P. (1983) *Diderot*, Oxford: Oxford University Press.(A brief survey in English of Diderot's literary career.)

Furbank, P.N. (1992) *Diderot: A Critical Biography*, London: Secker & Warburg. (Locates Diderot's interest in the arts at the centre of his life.)

Lough, J. (1968) *Essays on the 'Encyclopédie' of Diderot and d'Alembert*, London: Oxford University Press.(A detailed commentary on the compilation of a work of reference.)

Mauzi, R. (1960) *L'idée du bonheur dans la littérature et la pensée française au XVIIIe siècle*, Paris: Armand Colin.(Treats Diderot's ideas on happiness in the context of eighteenth-century French philosophy.)

Proust, J. (1967) *Diderot et l'Encyclopédie*, Paris: Armand Colin.(The most comprehensive discussion of Diderot's contributions.)
Strugnell, A. (1973) *Diderot's Politics*, The Hague: Nijhoff.(Includes a commentary on Diderot's contributions to the *Histoire des Deux Indes*.)
Wilson, A. (1972) *Diderot*, New York: Oxford University Press.(The fullest intellectual biography in any language.)

Dietrich of Freiberg (*c*.1250-after 1310)

In his work on the rainbow, De iride et radialibus impressionibus (On the Rainbow and Radial Impressions), Dietrich makes extensive use of experimental observation. He also wrote a number of other, more theoretical works including De esse et essentia (On Existence and Essence) and De intellectu et intelligibili (On Intellect and the Intelligible). In these works, Dietrich's emphasis varies; his theological works tend to be heavily Neoplatonic, while his more secular philosophical works are more Aristotelian. Dietrich disagreed with Aquinas on certain metaphysical issues, and seems to have written in opposition to particular works by Aquinas.

A Dominican from Germany, Dietrich studied at the University of Paris *c*.1275-7 and became master of theology at St Jacques in Paris before 1303. He was named provincial of Teutonia in 1293 and vicar provincial in 1310. He is best known for his investigations on the rainbow, put into writing by special request of the meeting of the general chapter of the Dominican order at Toulouse in 1304. The resulting work, *De iride et radialibus impressionibus (On the Rainbow and Radial Impressions)* draws on the optical theories of Alhazen, but in addition takes a new approach to researching the causes of the rainbow. Dietrich's innovation, as Boyer (1959) explains, was to see the value of investigating the reflective and refractive properties of individual raindrops rather than of the rain cloud taken as a whole. Extensive experimental observations of the paths of light rays through glass spheres and beakers filled with water enabled him to deduce how light passes through raindrops to produce both the lower and upper bows of rainbows. This part of his theory would not be further refined until the time of Descartes. His explanation of how colours are produced in the rainbow was less enduring, but is well thought out and closely argued (see Natural philosophy, medieval).

Dietrich also wrote on a wide range of other topics in natural philosophy, logic, psychology, metaphysics and theology. Other works in natural philosophy are less oriented toward experimental evidence: they include short works on the four elements and their compounds, on light and colour, on time and on astronomy. In his less experimental works, and particularly in his theological works, Dietrich displays extensive knowledge of and interest in Neoplatonic as well as Aristotelian theories. In the theological work *De intellectu et intelligibili (On Intellect and the Intelligible)*, for example, Dietrich describes creation as intellectual emanation through a hierarchy of Intelligences beginning with the primary intelligence - the soul of the first firmament - and progressing down to the world of man. Gilson (1955) has characterized the cosmology Dietrich sets forth here as a restatement of Proclus using Aristotelian terminology. This aspect of Dietrich's thught was to influence German Dominican mystics such as Meister Eckhart.

In his psychology, Dietrich rejects the Averroistic notion of a separate, shared agent intellect (see Averroism). Instead, Dietrich argues that there is an individual agent intellect always in action within each person. Thus far most thinkers of his time would have agreed, but Dietrich's account of the role of the agent intellect in human knowledge is once more strongly Neoplatonic. In his view, intellectual knowledge cannot be explained by the process of abstraction alone; an act of the agent intellect is required to make the object of cognition intelligible. The agent intellect is able to carry out this operation of intuitive cognition because it is made in the image of the divine intellect, proceeding from the divine ideas as an image and likeness of total being. Since, as an image of the divine, the human agent intellect comprehends all being, it is capable of knowing all other things by virtue of its knowledge of itself and as a feature of its orientation toward God (see Platonism, medieval).

Dietrich disagrees with Aquinas on certain important metaphysical issues. Maurer (1956), in his edition of Dietrich's *De quidditatibus entium (On The Essential Natures of Things)*, has argued that both this work and *De esse et essentia (On Existence and Essence)* were written in opposition to Aquinas' *De esse et essentia*. Dietrich denies Aquinas' argument that there is a distinction between essence (*essentia*) and existence (*esse*) in reality. The fact that we can understand a rose without knowing if one exists does not force us to acknowledge a real distinction between the rose's existence and its essence, by Dietrich's account, but merely shows that there are two ways of knowing about the rose. The only difference between essence and existence that Dietrich allows is in their manner of signifying being: *esse* signifies being as actuality, while *essentia* signifies it as disposition to actuality. Several other differences with Aquinas follow from this important fundamental disparity. To give one important example, Dietrich denies that it is in any way possible for accidents to exist without a subject, an argument that

puts in question Aquinas' explanation of transubstantiation.

FIONA SOMERSET

List of works

Dietrich of Freiberg (*c.*1280-1310) 'Scientific Works', in W.A. Wallace (ed.) *The Scientific Methodology of Theodoric of Freiberg*, Studia Friburgensia, new series 26, Fribourg: Fribourg University Press, 1959, 306-76. (Includes texts of three minor scientific works (*De miscibilis in mixto*, *De luce et eius origine* and *De coloribus*) and parts of two others (*De accidentibus*, excerpts from chapters 1-5 and 8-15; *De elementis corporum naturalium inquantum sunt partes mundi*, excerpts from chapters 1-8), plus a full list of minor works available in modern editions up to 1959.)

Dietrich of Freiberg (*c.*1280-1310) *De esse et essentia (On Existence and Essence)*, ed. E. Krebs, 'Le Traité *De Esse et Essentia* de Thierry de Fribourg', *Revue néoscolastique de philosophie* 18: 516-36, 1911.(Philosophical work, possibly written in opposition to Aquinas.)

Dietrich of Freiberg (*c.*1280-1310) *De quidditatibus entium (On The Essential Natures of Things)*, ed. and trans. A. Maurer, 'The *De quidditatibus entium* of Dietrich of Freiberg and its Criticism of Thomistic Metaphysics', *Mediaeval Studies* 18: 173-203, 1956.(A further work on essence, also possibly written in opposition to Aquinas.)

Dietrich of Freiberg (*c.*1280-1310) *De intellectu et intelligibili (On The Intellect and the Intelligible)*, printed as *Tractatus de intellectu et intelligibili* in E. Krebs, *Meister Diestrich (Theodoricus Teutonicus de Vriberg). Sein Leben, seine Werke, seine Wissenschaft*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters XII, 5-6, Munster: Aschendorff, 1914, 119-206.(Dietrich's major work on the intellect.)

Dietrich of Freiberg (1304-10) *De iride et radialibus impressionibus*, ed. and trans. J. Würschmidt, *Dietrich von Freiberg: Über den Regenbogen und die durch Strahlen erzeugten Eindrücke*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters XII, 5-6, Munster: Aschendorff, 1914.(Dietrich's best known work. Excerpts are also translated in W.A. Wallace (ed. and trans.) 'On the Rainbow and Radiant Impressions', in E. Grant (ed.) *A Source Book in Medieval Science*, Cambridge, MA: Harvard University Press, 1974.)

References and further reading

Boyer, C.B. (1959) *The Rainbow: From Myth to Mathematics*, New York: T. Yoseloff, 110-24.(Explains the importance and the influence of Dietrich's work on the rainbow.)

Gilson E. (1955) *History of Christian Philosophy in the Middle Ages*, London: Sheed & Ward, 433-7, 753-5. (Discusses three of Dietrich's strongly Neoplatonic theological works.)

Wallace, W.A. (1981) 'Dietrich von Freiberg', in C.C. Gillispie (ed.) *Dictionary of Scientific Biography*, New York: Scribner, vol. 4, 92-5.(A more detailed introduction to Dietrich's scientific writings.)

Digby, Kenelm (1603-65)

Seventeenth-century English Catholic, original member of the Royal Society, and one of the first philosophers to produce a fully developed system of mechanical philosophy, Sir Kenelm Digby cut a dashing figure as a poet, privateer and philosopher. As a Catholic and royalist, he spent much of his life in semi-exile on the continent where he conversed with many of the political and intellectual leaders of his time; as a philosopher, he was favourably compared to René Descartes and John Locke. He attempted to wed the philosophy of Aristotle to the new mechanical physics, which maintained that every event in the material world is reducible to matter in motion. His interests and writings cover a wide range, from religion and magic to vegetative growth and literary commentary. The explicit goal of his most significant book, Two Treatises (1644), was to prove the immortality of the human soul. To this end, the first treatise constitutes an exhaustive study of bodies and their features. By showing that all corporeal qualities are to be explained in strictly material terms, he prepares the way for a thorough discussion of the soul. Digby argues that the soul must be immaterial (and hence immortal) because otherwise its features cannot be explained. He went on to apply the mechanical principles which he developed in this work to a variety of topics, including some traditionally associated with the occult. His works on alchemical, medical and religious topics were also widely read.

1 Life

Sir Kenelm Digby was one of the most colourful and impressive figures of his generation. He was not yet three when his father, Sir Everard Digby, was tortured and executed for his role in the Gunpowder Plot. As a Catholic at Oxford (1618-20), the young man was forced to remain on the periphery of the university. He studied there with the mathematician and astronomer, Thomas Allen, who probably introduced him to alchemical experiments. He travelled to France, Spain and Italy (1620-3) where a prince considered him ‘the just measure of perfection’ and Marie de Medici made immodest advances. A friend to such writers and artists of his era as Ben Jonson and Anthony van Dyck, and to members of the English royal family, Digby was knighted by James I, was a successful privateer for Charles I, and chancellor to Queen Henrietta Maria. In 1625 he married Venetia Stanley, famous for her beauty and intellectual achievements, with whom he had five children, and whose sudden death in 1633 turned his attention from public concerns to philosophical contemplation. The first two years of a lengthy period of mourning were spent in secluded study at Gresham College, London, in whose laboratories he conducted experiments. Digby converted to Anglicanism in 1630 (possibly for political reasons), but reconverted before he moved to France in 1635, where he felt more comfortable as a Catholic.

Although under suspicion in England as a Catholic and ardent royalist, and briefly imprisoned there (1642-3), he often travelled from France and even negotiated with Cromwell in 1648 and 1654-5 for toleration of Catholicism in England. After his reconversion, he turned his attention to religious and ecumenical issues and published *A Conference with a Lady about Choyce of Religion* (1638) and a response to Thomas Browne’s popular *Religio Medici* entitled *Observations upon Religio Medici*, the latter written in 1643 while he was in prison. During his many years in Paris, he was part of the Mersenne circle, studied anatomy and chemistry, published his *Two Treatises* (1644), and was one of the early admirers of Descartes’ *Discourse on Method*. He conversed and corresponded with Hobbes and Descartes, both of whom seem to have thought well of him. Along with his mentor and fellow English Catholic, Thomas White, Digby was one of the first Englishmen to produce a fully developed system of mechanical philosophy. He applied the mechanical principles which he developed in his *Two Treatises* to a variety of topics. He collected medical, chemical, and household remedies as well as books and manuscripts. Many of his recipes and remedies, which he exchanged with other natural philosophers including Robert Boyle, were published after his death. Digby went to Montpellier in 1657 to take the waters and there gave his famous account of the powder of sympathy. This led to his most popular book, *A late Discourse Made in a Solemne Assembly… touching the Cure of Wounds by the Powder of Sympathy* (1658 - originally in French), which was reprinted often and translated into a number of languages. Digby travelled to Germany, the Low Countries and Scandinavia; in 1660, after the Restoration, he returned to England, where he was one of the original members of the Royal Society. He presented his *A Discourse Concerning the Vegetation of Plants* to the Society in 1661. Although Digby’s works were soon superseded, he was highly regarded in his day. As a contemporary wrote, he was: ‘This Ages Wonder for his Noble Parts, Skilled in six Tongues, and learned in all the Arts’.

2 The *Two Treatises*

Digby published his *Two Treatises* in Paris in 1644 with the professed goal of proving the immortality of the soul. In the dedicatory epistle he makes plain that he desires to analyse bodies because ‘the knowledge of them serveth to the knowledge of the soul’, that is, through an exhaustive account of the nature of bodies and their qualities, he hopes to show that features of the human soul are not ‘effected by corporeall agents’, and hence must have something immaterial as their source. In his preface, Digby is particularly clear about his method. Against the scholastics who achieve neither science nor certitude, but attribute to bodies ‘hidden qualities that mans witt cannot reach unto’, Digby intends to begin with ‘the commonest thinges’ in nature so as ‘to shew from what principles, all kindes of corporeall operations do proceed; and what kind of operations all these must be, which issue out of these principles’. By arguing in his first treatise that corporeal qualities can be explained wholly in terms of the material constituents of bodies and without any reference to occult powers, Digby prepares the way in his second treatise for his argument for the immortality of the soul. As he puts it in his preface, he intends to ‘raise’ his discussion in the second treatise ‘to the contemplation of the soule; and shew, that her operations are such, as cannot proceed from those [mechanical] principles; which being adequate and common to all bodies, we may rest assured, that what cannot issue from them, cannot have a body for its source’.

In the *Two Treatises* Digby interweaves ancient and modern principles into an Aristotelian fabric. He has unmitigated scorn for the scholastics and high esteem for contemporaries like Descartes and Galileo. But whereas the latter mechanists explicitly reject the Aristotelian philosophy, Digby embraces it (see Aristotle). He asserts in the conclusion to the first treatise that his proposals ‘are built upon the same foundations’ as Aristotle whom he describes as: ‘the greatest Logician, Metaphysitian and Universall scholler… that ever lived’. Aristotle’s ‘name must never be mentioned among schollers, but with reverence, for his unparalleled worth; and with gratitude for the large stocke of knowledge he hath enriched us with’. But Digby is equally explicit about the fact that for the true science, we must inform Aristotelian principles with the new discoveries: ‘Let us then admire him for what he has deliver’d us: and where he falls short or is weary in his search… let us seek to supply and relieve him’ (1644: ch. XX).

Defining body in terms of quantity, and quantity in terms of divisibility, Digby offers an account of bodies and their qualities in mechanical terms. He uses rarity and density, which he describes as the different ways in which something is divisible, to explain the four elements of Aristotle. He maintains that his analyses of Aristotle’s four elements are ‘the proper notions of the four elements… which are, the notions of Quantity’ (1644: ch. IV). He applies the principles of force and velocity to these elements and then ‘deduces’ the principles ‘which govern Mechanics’ (1644: ch. IX). From the features of these ‘most simple bodies’ or elements, Digby intends to explain the qualities and operations of all the ‘compounded’ bodies of nature, both animate and inanimate. He acknowledges in his conclusion that he departs from Aristotle on a ‘few points’, but insists that his fundamental principles are Aristotle’s: ‘all the difference between us is, that we enlarge ourselves to more particulars then he hath done’.

In his criticism of the scholastic philosophy, Digby nicely exemplifies the transformation under way in seventeenth-century thought. The mechanical physics had brought about a radical re-evaluation of a number of doctrines which the scholastics had promulgated for centuries. Digby was particularly articulate concerning the ontology of the objects of perception. On this topic, he shares many of the same doctrines, if not the same depth of thought, of John Locke (see Locke, J. §§4-5). According to the scholastic tradition, sensible qualities were treated as irreducibly distinct attributes of the object of perception. The sensible quality of colour, for instance, was believed to exist in essential relation to sight and was not considered reducible to any other attribute in the object. Digby rejected as unintelligible this tendency ‘to give actual Beings to the quantity, figure, colour, smell, taste and other accidents’ of the object, ‘each of them distinct one from another, as also from the substance which they clothe’. His point was that just ‘because I find the notions of them really distinguished… in my mind’ I should not act ‘as if there were different Entities’ (1644: ch. I). According to Digby (and Locke), the sensible accidents of the scholastics in fact have no separate or distinct existence in reality. Rather, ‘what is but one entire thing in it self, seemeth to be many things in my understanding’. However separate these qualities may be in the mind, they are but aspects of a single, unitary thing in reality. According to Digby, this unitary thing consists of corpuscles of matter in motion which in turn cause motions in the surrounding matter. Some of these motions pass through both the medium and the sense organs until they reach the brain of the perceiver and there cause an appropriate idea in

the mind. The multiplicity of the qualities which we perceive are therefore due not to a multiplicity in the object but rather to one in the sense organs. According to Digby, the senses ‘make severall and distinct pictures of what entereth by their doors’ and thereby constitute a different way in which the object can cause motions and hence ideas in us (1644: ch. I).

Having made an exhaustive study of bodies and their features in his first treatise, Digby turns his attention to matters immaterial in his second. By showing that all corporeal qualities are to be explained in strictly material terms, he has prepared the way for his discussion of the rational soul and its qualities. At this point he acknowledges that we ‘have passed the Rubicon of experimental knowledge’ so that ‘henceforth, we must in all our searches and conclusions rely upon the single evidence of Reason’. Digby claims that when we are stripped of everything material ‘there still remaineth a *substance*, a *thinker*, an *Ego* or *I*’ (1644: ch. IX). His argument for immortality is fairly standard. He defines the soul as that which is immaterial and indivisible and notes that only something material can undergo the decay and dissolution which mortality requires. That is, since ‘no change at all can happen to an abstracted soul’, it is incapable of mortality (1644: ch. XII).

Digby’s *Two Treatises* was widely read and generally well received by his contemporaries. His physical proposals, however, did not make a lasting contribution. Digby himself offers the most likely explanation for their failure in his dedicatory epistle: ‘a gallant man, whose thoughts flye att the highest game… deemeth it farre too meane for him, to dwell upon the subtilest of their mysteries for science sake’. In the end, Digby’s mechanical physics was too imprecise and incomplete; the treatises could not compete with the more fully articulated physical systems of Galileo, Descartes and Gassendi.

3 Other works

Digby applied himself to topics ranging from alchemy and plant growth to religion and literature. Of special interest to philosophers is the fact that he was prepared to extend the mechanical principles developed in the *Two Treatises* to all of his scientific and philosophical studies, even to topics traditionally associated with the occult. He practised alchemy, but agreed with many of his contemporaries that its language and method should be made more precise. He studied and conversed with Nicolas le Fèvre and other chemists-alchemists on the continent and, like them, was inclined to turn alchemical experiments into precise chemical events. Some of the results of his studies appear in his *A Choice Collection of Rare Chymical Secrets and Experiments in Philosophy*, published posthumously in 1682.

Digby also studied medicine: he both collected and experimented with medicinal recipes and, like many of his contemporaries, he was interested in the Paracelsian tradition of sympathetic medicine (see Paracelsus). Digby agreed with his Paracelsian predecessors that a wound could be healed by treating an object that contained blood from it (say, the sword that had caused the wound or bandage that had bound it). He was convinced that the Paracelsians were correct, since he had successfully cured a serious wound by such means. He diverged significantly from the tradition, however, in his explanation of the effectiveness of the treatment: whereas the Paracelsian explanation was based on occult virtues and hidden sympathies, Digby’s explanation relies on mechanical principles. In his book on the powder of sympathy, Digby presents the seven physical principles from which his explanation is supposed to follow and then proposes that because the medicine ‘being incorporated with the blood cannot choose but make the same voyage together with the atoms of blood’, and because ‘there will be a kind of current of air drawn round about the wound’, it follows that the blood on the object treated with the medicine will find ‘the proper source and original root whence they [the atoms of blood] issue’ and ‘will joyntly be imbibed together [with the medicine] within all the corners, fibres, and orifices of the Veins which lye open about the wound; when it must of necessity be refresht, and in fine imperceptibly cured’. Thus Digby explains the traditional medical ‘fact’ by modern means.

While some of his alchemical and medical opinions placed him in a tradition that was soon supplanted, Digby was at the forefront of the science of biology. In his *A Discourse Concerning the Vegetation of Plants*, which he presented to the nascent Royal Society in 1661, he attempts to explain plant growth by chemical and mechanical means. In his analysis of the growth and decay of plants, he relies on his notions of the four elements and their characteristic ‘actions’. He gives a very clear description of the germination, development, and reproduction of plants and uses an observational technique very accurate for the time. While neither so colourful nor so popular as his book on the powder of sympathy, this work was probably his most significant scientific offering.

See also: Alchemy; Aristotelianism in the 17th Century

CHRISTIA MERCER

List of works

Digby, K. (1638) *A Conference with a Lady about Choyce of Religion*, Paris.
Digby, K. (1643) *Observations upon Religio Medici*, London.
Digby, K. (1644) *Two Treatises, In the one of which the Nature of Bodies; in the other, the Nature of Mans Soule; is looked into: in the way of discovery, of the Immortality of Reasonable Soules*, Paris; repr. New York: Garland, 1978.
Digby, K. (1652) *A Discourse, Concerning Infallibility in Religion*, Paris.
Digby, K. (1658) *Discours fait en une célèbre assemblée, par le Chevalier Digby… touchant la guérison des playes par la poudre de sympathie*, Paris; trans. E. White as *A late Discourse Made in a Solemne Assembly… touching the Cure of Wounds by the Powder of Sympathy*, London, 1658.
Digby, K. (1661) *A Discourse Concerning the Vegetation of Plants*, London.
Digby, K. (1668) *Choice and Experimented Receipts in Physick and Chirurgery… Collected by the Honourable and truly Learned Sir Kenelm Digby Kt. Chancellour to her Majesty the Queen Mother*, London.
Digby, K. (1682) *A Choice Collection of Rare Chymical Secrets and Experiments in Philosophy, as also Rare and unheard-of Medicines*, ed. G. Hartman, London.

References and further reading

Ayers, M. (1991) *Locke*, London: Routledge, 2 vols.(Contains an account of Digby on the ontology of the object of perception and on other matters related to Locke and the generation of British philosophers after Digby.)
Ayers, M. and Garber, D. (eds) (1997) *The Cambridge History of Seventeenth-Century Philosophy*, Cambridge: Cambridge University Press.(Contains several discussions of Digby.)
Bligh, E.W. (1932) *Sir Kenelm Digby and his Venetia*, London: Sampson Low & Co., Ltd.(This also includes the ‘expurgated passages’ from Digby’s memoirs, and a bibliography.)
Browne, T. (1643) *Religio Medici*, London; repr. ed. W.A. Greenhill, La Salle, IL: Open Court, 1991.(A reprint of an 1881 imprint, this edition also includes correspondence between Brown and Digby, notes and an index.)
Dobbs, B.J. (1971-4) ‘Studies in the Natural Philosophy of Sir Kenelm Digby’, *Ambix* 18: 1-25; 20: 143-63; 21: 1-28.(The most thorough summary available of Digby’s work in natural philosophy and alchemy.)
Henry, J. (1982) ‘Atomism and Eschatology: Catholicism and Natural Philosophy in the Interregnum’, *British Journal for the History of Science* 15: 211-39.(A discussion of the ecumenism of Digby and his mentor, Thomas White.)
Krook, D. (1993) *John Sergeant and his Circle*, Leiden: Brill.(A discussion of Digby’s Aristotelianism and his relation to Sergeant and to other progressive English Catholics. Chapter 3 is especially relevant.)
Petersson, R.T. (1956) *Sir Kenelm Digby: The Ornament of England, 1603-65*, Cambridge, MA: Harvard University Press.(A thorough and colourful account of Digby’s life and work, with a complete bibliography of his works, including some not listed above.)

Dignāga (*c*.480-*c*.540)

A logician and epistemologist, Dignāga is traditionally regarded as the founder of a Buddhist school that sought to avoid divisive controversies over which Buddhist writings were authentic by emphasizing logic and epistemology rather than the study of scriptures and their commentaries. His principal contributions consisted of refining the theory of inference and tightening the forms of argument commonly used in debate and polemics. His theories became the basis on which the influential philosopher Dharmakīrti built his system, which became the standard Buddhist scholastic system in India and later in Tibet.

1 Epistemology

According to traditional hagiographies, Dignāga was born around 480 into a Brahman family near Kāñcīin south India, where he was ordained a Buddhist monk as a young man. Dissatisfied with his teachers in the south, he is supposed to have travelled north and become a disciple of Vasubandhu. Celebrated for his debating skills, he was nicknamed the 'Bull in debate'.

At the beginning of his principal work, *Pramāṇasamuccaya*, Dignāga stated that his purpose was to resolve several controversies that other philosophers had generated about the means of acquiring knowledge. Whereas his predecessors had enumerated several methods by which knowledge may be acquired, Dignāga took the position that there are in fact only two methods. These two methods are distinguished from one another in virtue of the kinds of object that can be their subject matters. The first method of securing new knowledge is described as pure sensation (*pratyakṣa*), a form of cognition that is free of all judgment (*kalpanā*). The subject matter of this type of cognition is particular instances (*svalakṣana*) of colour, sound and other sensible properties. The second method is described as inferential reasoning. The subject matter of this type of cognition is universals (*sāmānyalakṣaṇa*). In contrast to most Brahmanical philosophers WHO had preceded him, Dignāga held that only the senses can be aware of particular sensations, and only the intellect can be aware of universals. The view that there are exactly two types of knowledge, and that there is no subject matter common to both of them, came to be a hallmark of Buddhist doctrine, since it was accepted by most Buddhist philosophers writing in Sanskrit and rejected by the majority of Brahmanical and Jaina philosophers (see Epistemology, Indian schools of §1).

Before Dignāga, most Indian treatises dealing with reasoning were primarily devoted to outlining the rules of debate between opposing parties. Consequently, discussions of the formal properties of correct lines of argument were mingled with discussions of which errors on a discussant's part would result in defeat. Dignāga is usually given credit for being the first Indian philosopher to make a clear distinction between the formal properties of correct reasoning and the rules of debate. Debate, he said, is merely the articulation for the benefit of another person of a conclusion that one has arrived at by oneself through correct reasoning.

A correct inference, said Dignāga, is one that makes use of an observable property that serves as a sign (*liōga*) of an unobserved property. The object to which the signifiable property can be inferred to belong is called the subject (*pakṣa*) of the inference. One property can serve as the sign of a second property only if three conditions are met. First, the sign must be observed to be a property of the subject of the inference. Second, the sign must be known to occur together with the signified property in objects other than the subject of the inference. And third, the sign must not be known to occur in objects in which the signified property is absent (see Inference, Indian theories of).

2 Theory of language

In contrast to the Brahmanical philosophical traditions, Buddhist and Jaina philosophers denied the authority of the Veda, the sacred texts on which the Brahmanical religion was based. According to Brahmanical philosophers, these sacred texts came from a nonhuman source and were infallible. Brahmanical philosophers tended to argue that human perception and reasoning are always liable to error, and human beings must therefore appeal to superhuman sources of knowledge if they are ever to have certainty. Moreover, human knowledge is ultimately limited to what can be known through the senses, and it requires a special source of knowledge such as the Veda to make people aware of duty, the merit that attends doing one's duties, and the rewards of earning that merit. The language of the Veda was regarded as different from all ordinary human languages in that it is eternal (see MīmāṃsāČsect;3). All of these Brahmanical doctrines were repudiated by Dignāga, who held that all sentences,

including those of the Veda, are merely a special kind of sign that function in essentially the same ways as any other kind of inferential sign. Since language is merely a species of inference, information communicated through language is no more reliable than information gained through reasoning. And like all other inferential knowledge, the knowledge communicated through language does not convey positive information about things but merely eliminates certain states of affairs from consideration. Dignāga argued that language cannot express the particulars that are known through the senses; rather, verbal symbols express only concepts generated by the mind. These concepts are formed not by observing universals or similarities that are sensed in particulars, but by ignoring the differences in what is observed.

3 Influence

Although Dignāga was honoured for centuries in India and Tibet as the founder of the Buddhist epistemological movement, his works were eventually superseded by his followers, the most notable of whom were Dharmakīrti, Dharmottara, Śāntarakṣita and Kamalaśīla in India, and Tsong kha pa Blo bzang grags pa and mKhas grub dge legs dpal bzang po in Tibet. Although a few of Dignāga's minor works were translated from Sanskrit into Chinese, none of his most important writings, and virtually none of the writings of his followers, were studied in China. Consequently, the epistemological school of Buddhism never developed in China and other parts of East Asia. After the demise of Buddhism in Northern India, the Sanskrit originals of most of Dignāga's works were gradually lost and nearly forgotten. In the early twentieth century, historians of Indian philosophy began to take a renewed interest in studying his works (mostly in Tibetan translations) and assessing his place in the evolution of Indian theories of inference.

See also: Buddhist philosophy, Indian §4; Knowledge, Indian views of §3; Nominalism, Buddhist doctrine of; Universals, Indian theories of

RICHARD P. HAYES

List of works

Over twenty philosophical writings have been attributed to Dignāga. Of these, seven are no longer extant. The remaining works are extant only in Tibetan or Chinese translation. The following list includes only those for which studies or translations in European languages are available.

Dignāga (*c.*520) *Ālambanaparīkṣa (Examination of the Object of Awareness)*, trans. S. Yamaguchi, 'Examen de l'objet de la connaissance (*Ālambana-parīkṣa*): textes tibétains et chinois et traduction des stances et du commentaire éclaircissements et notes d'après le commentaire tibétain de Vinītadeva' (Examination of the Object of Awareness: Tibetan and Chinese texts, with translation of illuminating stanzas and commentary, and notes after Vinītadeva, *Journal Asiatique* (January-March): 1-59, Paris: Imprimerie Nationale, 1929; also trans. F. Tola and C. Dragonetti, 'Dignāga's *Ālambanaparīkṣavṛtti*', *Journal of Indian Philosophy* 10: 105-34, 1982.(A brief work in eight verses that investigates various possibilities for what the exact nature of the content of a sensory cognition is.)

Dignāga (*c.*525) *Nyāyamukha (Introduction to Systematic Reasoning)*, trans. G. Tucci, *The Nyāyamukha of Dignāga*, Heidelberg: Materialien zur Kunde des Buddhismus, 1930.(An introductory work on the methods of correct reasoning.)

Dignāga (*c.*530) *Pramāṇasamuccaya (Collected Writings on the Acquisition of Knowledge)*, first chapter translated in Hattori, M., *Dignāga, On Perception*, Cambridge, MA: Harvard University Press, 1968; excerpts of chapters two and five translated in Hayes, R.P., *Dignāga on the Interpretation of Signs*, Dordrecht: Kluwer Academic Publishers, 1988.(A collection of writings on epistemology, with sections on sense perception, inference, debate and theory of language.)

References and further reading

Frauwallner, E. (1959) 'Dignāga, sein Werk und seine Entwicklung' (Dignāga, his Work and its Development), *Wiener Zeitschrift für die Kunde Süd- und Ostasiens* 3: 83-164.(An attempt to place Dignāga's writings in chronological order.)

Hattori, M. (1968) *Dignāga, On Perception*, Cambridge, MA: Harvard University Press.(Introduction includes biography of Dignāga and a complete list of works attributed to him.)

Hayes, R.P. (1988) *Dignāga on the Interpretation of Signs*, Dordrecht: Kluwer Academic Publishers.(A study of several of Dignāga's works on epistemology, with an attempt to place his theories in the context of earlier

Buddhist philosophers.)

Herzberger, R. (1986) *Bhartṛhari and the Buddhists*, Dordrecht: Kluwer Academic Publishers.(A study of the relationship between the Brahmanical philosopher Bhartṛhari and the Buddhists Dignāga and Dharmakīrti. Rather complex and technical, but informative.)

Matilal, B.K. and Evans, R.D. (1986) *Buddhist Logic and Epistemology: Studies in the Buddhist Analysis of Inference and Language*, Dordrecht: Reidel.(A collection of papers, some quite technical. Matilal's introduction is accessible to the nonspecialist and offers a good overview of issues addressed by Dignāga and his followers.)

Mookerjee, S. (1935) *The Buddhist Philosophy of Universal Flux: an Exposition of the Philosophy of Critical Realism as Expounded by the School of Dignāga*, Delhi: Motilal Banarsidass, repr. 1975.(Although somewhat dated, this is still one of the best introductions to the philosophical issues at the centre of the epistemological school founded by Dignāga.)

Shastri, D.N. (1964) *Critique of Indian Realism:A Study of the Conflict Between the Nyāya-vaiśeṣika and the Buddhist Dignāga School*, Agra: Agra University.(A thorough study of the Brahmanical reactions to the Buddhist criticisms of their ontological and epistemological positions.)

Warder, A.K. (1980) *Indian Buddhism*, Delhi: Motilal Banarsidass.(A good survey of Buddhism in classical India that has a substantial chapter on Dignāga and the epistemological school of Buddhism.)

Dilthey, Wilhelm (1833-1911)

Wilhelm Dilthey saw his work as contributing to a 'Critique of Historical Reason' which would expand the scope of Kant's Critique of Pure Reason by examining the epistemological conditions of the human sciences as well as of the natural sciences. Both kinds of science take their departure from ordinary life and experience, but whereas the natural sciences seek to focus on the way things behave independently of human involvement, the human sciences take account of this very involvement. The natural sciences use external observation and measurement to construct an objective domain of nature that is abstracted from the fullness of lived experience. The human sciences (humanities and social sciences), by contrast, help to define what Dilthey calls the historical world. By making use of inner as well as outer experience, the human sciences preserve a more direct link with our original sense of life than do the natural sciences. Whereas the natural sciences seek explanations of nature, connecting the discrete representations of outer experience through hypothetical generalizations and causal laws, the human sciences aim at an understanding that articulates the fundamental structures of historical life given in lived experience. Finding lived experience to be inherently connected and meaningful, Dilthey opposed traditional atomistic and associationist psychologies and developed a descriptive psychology that has been recognized as anticipating phenomenology.

Dilthey first thought that this descriptive psychology could provide a neutral foundation for the other human sciences, but in his later hermeneutical writings he rejected the idea of a foundational discipline or method. Thus he ends by claiming that all the human sciences are interpretive and mutually dependent. Hermeneutically conceived, understanding is a process of interpreting the 'objectifications of life', the external expressions or manifestations of human thought and action. Interpersonal understanding is attained through these common objectifications and not, as is widely believed, through empathy. Moreover, to fully understand myself I must analyse the expressions of my life in the same way that I analyse the expressions of others.

Not every aspect of life can be captured within the respective limits of the natural and the human sciences. Dilthey's philosophy of life also leaves room for a kind of anthropological reflection whereby we attempt to do justice to the ultimate riddles of life and death. Such reflection receives its fullest expression in worldviews, which are overall perspectives on life encompassing the way we perceive and conceive the world, evaluate it aesthetically and respond to it in action. Dilthey discerned many typical worldviews in art and religion, but in Western philosophy he distinguished three recurrent types: the worldviews of naturalism, the idealism of freedom and objective idealism.

1 Foundation of the human sciences

Dilthey initially studied theology at Heidelberg and Berlin, but increasingly devoted his attention to history and philosophy. Before completing his dissertation on Schleiermacher's ethics in 1864, he had already written a long prize-essay, 'Das hermeneutische System Schleiermachers in der Auseinandersetzung mit der älteren protestantischen Hermeneutik' (Schleiermacher's Hermeneutic, A System in Relation to Earlier Protestant Hermeneutics) (see Schleiermacher, F.D.E.). After teaching at Basel, Kiel and Breslau, Dilthey was appointed in 1882 to the chair in philosophy in Berlin that Hegel had once occupied. His more empirical approach to the understanding of history departs from Hegel in being radically multidisciplinary. Dilthey's contributions to cultural and social history, to literary criticism and to the history of the human sciences in general found their philosophical grounding in a major, although unfinished, theoretical work, *Einleitung in die Geisteswissenschaften* (*Introduction to the Human Sciences*) (1883, 1982).

In 1883, Dilthey published the first of two projected volumes of this work. Book One of the *Einleitung* was devoted to an overview of the human sciences; Book Two to a history of the rise and fall of metaphysics in relation to the project of grounding the natural and human sciences. At the end of this history he maintains that both metaphysics and the modern natural sciences have established false models for the human sciences by constructing abstract intelligible worlds independent of lived experience. A new epistemology is needed which will show that the modern scientific conception of nature is a mere phenomenal abstraction from the more inclusive reality of life. Dilthey introduces the idea of a life-nexus (*Zusammenhang des Lebens*) as the original matrix of reality, not only for the human sciences, but also for the natural sciences. If the natural sciences find it useful to explain phenomena

by means of elemental entities and universal mechanistic laws, this does not entail that they possess a more ultimate reality than the human sciences or that the latter should be constrained to drop the idea of purposiveness from their understanding of history and society.

Dilthey shows that in the past the human sciences were blinded by the success of natural science to adopt its mechanistic approach. Inspired by the ideas of Hobbes, Spinoza and Hume, a natural system of the human sciences was formed (see Hobbes, T. §2; Hume, D. §1; Spinoza, B.). A theoretical analysis of human nature was to establish ‘a few general psychic elements to explain the facts of human historical life’. These psychological elements and a finite number of non-teleological laws, like that of self-preservation, were supposed to explain the mysteries of life, the intricacies of human action, and the complexities of social interaction.

Even more problematic than the reductive character of this natural system of the human sciences is the fact that from the traditional epistemological perspective the separate foundations of the natural and the human sciences mark a division within theory alone. This has the unfortunate effect of relegating practical philosophy to a secondary concern. Dilthey disapproves of the abstract way in which a purely theoretical approach to the human sciences seeks the explanative basis for the value judgments and the imperatives that regulate the life of the individual and society. The proper foundation of the system of the sciences must be located in ‘self-reflection’ in contrast to ‘theory of knowledge’. Self-reflection based on our experience of life provides the foundation for action as well as for thought. This philosophical self-reflection renders theory and practice equiprimordial.

The real task of founding both the natural and the human sciences was left for the second volume of the *Einleitung*, but Dilthey did not publish it during his lifetime. He did, however, work out a large part of it, which first appeared in 1982 in Volume 19 of the *Gesammelte Schriften (Collected Works)*. Dilthey examines the conditions of consciousness involved in our prescientific awareness of reality. He begins Book Four with a first principle, that of phenomenality, according to which everything real is accessible as a fact of consciousness without being reduced to a mere representation of consciousness. It is important to point out that Dilthey’s conception of phenomenality should be contrasted to phenomenalism, which reduces all reality to facts of consciousness (see Phenomenalism). Nor is it to be confused with the view that facts of consciousness are mere phenomena as distinct from external reality. Dilthey shows how the facts of our consciousness possess a primordial reality that already contains a reference to things beyond consciousness. When we become aware of something *as* a fact of consciousness we possess what Dilthey calls *Innewerden* or ‘reflexive awareness’. This reflexive awareness is pre-reflective and involves a felt self-givenness but no explicit sense of self. *Innewerden* is thus not to be equated with an objectifying self-consciousness, for it precedes any subject-object, act-content distinction. Reflexive awareness is proto-intentional in that it is oriented towards the world even if the world is not yet thematized as an external objective domain. It is this sense of being already a part of the world that is lost in the natural sciences, but must be preserved in the human sciences. Dilthey’s initial descriptions of the facts of consciousness constitute an empirical phenomenology and are at the level of prescientific *Wissen* or immediate knowledge.

Dilthey’s second principle, that of self-reflection, allows him to move to the level of scientific *Erkenntnis* or conceptual knowledge. In reflexive awareness we possess the felt reality of the facts of consciousness, but this does not entitle us to posit their reality conceptually. The move from the first level of prescientific reflexive awareness to conceptualization is one of explication, where the reflexive becomes reflective. The principle of self-reflection explicates the facts of consciousness into an overall nexus that encompasses all aspects of psychic life.

Within this framework of total consciousness, self-reflection then begins to differentiate between facts of self-consciousness and facts of the world. Facts that are perceived as existing in my consciousness are grasped as part of inner perception. Those facts that are independent of my self are considered facts of outer perception. But the distinction between self and world is not one that can be derived from the intellect alone. The reality of the external world is not an inference based on causal hypotheses, but is felt primarily through resistance to the practical impulses of the will.

Dilthey proceeds to show that a reflexive or felt awareness accompanying acts of will provides a crucial step in the process of differentiating self and world. Through the felt tension between efficacy of the will and resistance to it we learn to distinguish the reality of inner perception from outer perception. This dynamic relation to the external world is more fundamental than the static epistemological relation of a representation to an object. Whereas

representational consciousness projects the world as a theoretical horizon for objects of natural science, reflexive awareness possesses the world as a temporal nexus in which I participate, but which is also full of things and persons actively resisting my will. A new, reflective epistemology of the human sciences must reclaim our original access to life and thus cannot be merely an extension of the epistemology of the natural sciences.

Book Five again does not restrict itself to the human sciences, but considers the logical conditions that determine all acts of representing reality. It provides Dilthey's logic, not as a formal logic that abstracts from reality, but as an intermediary between the epistemological and methodological approaches to all the sciences that represent reality. Being representational, thought leaves behind the immediate certainty of reflexive awareness, but gains the capacity to apprehend truth. Thought is the attentive 'positing of reality' (*das Wirkliche*) not simply as that which momentarily resists, but as that which 'realizes a constant effect' (*das Wirkende*). Whereas perception posits as the real the shifting contents of what happens to be present, thought creates a nexus of representations that is in agreement with a system of reality that remains more or less constant.

The sixth and last book of the *Einleitung* begins with a general discussion of method and how it makes explicit the procedures for grasping reality inherent in the logical operations of thought. Dilthey distinguishes methods recurring in every domain of human knowledge from those peculiar to particular problems. The latter kind of method constitutes what Dilthey calls a higher logic which establishes 'rules of procedure that arise when a particular set of real conditions is introduced'. Dilthey is still speaking of both natural and human sciences, for he claims that 'aspects of mathematics, higher criticism, hermeneutics, jurisprudence [and] statistics belong to this higher logic'. It is only at the level of higher logic that we can begin to differentiate methodologically between the natural and the human sciences in specific ways.

The task of analysis and synthesis in the sciences is to simplify reality conceptually and to reorganize it in more systematic terms. Related to the process of analysis is the process of abstraction. The difference is that abstraction starts with a complex whole, 'singles out one fact and disregards the others, whereas [analysis] seeks to apprehend the majority of the facts that make up the factors of a complex whole'. Abstraction can occur in both the natural and the human sciences, but applies especially to the former. To the extent that analysis moves from a whole to its parts without isolating them from this whole, it can be said to engender understanding in the human sciences.

2 Descriptive psychology

Dilthey's 'Ideen über eine beschreibende und zergliedernde Psychologie' ('Ideas Concerning a Descriptive and Analytic Psychology') (1894) aims to wean psychology away from naturalistic models and redefine it as a human science. This new psychology would be 'first' within the system of the human sciences in the sense that it provides neutral descriptions of experience. It must suspend the general hypotheses of traditional psychology as much as possible, whether they be psychophysical or associationist. Such hypotheses strive to explain all psychic processes as different constructions from certain basic mental elements, such as representations. However, these elements cannot be well-determined or measured, with the result that hypotheses relating them have remained largely untestable.

Most psychological hypotheses about the association of representations are dispensable because inner experience is already interconnected. Whereas outer experience presents us with many unconnected phenomena which can only be related through the hypotheses of the natural sciences, psychology must consult the data of inner experience which are given as parts of a real continuum. This means that connectedness in psychic life does not need to be explained hypothetically, but can be experienced directly.

However, Dilthey does not altogether rule out hypotheses from his descriptive psychology. They are merely prevented from assuming the fundamental role they play in the natural sciences. It is clearly necessary to construct particular hypotheses when the continuity that normally exists among psychic processes is broken or interrupted. Also in cases where we were not fully attentive, questions of detail may remain problematic and thus require hypotheses to clarify what might have happened. Whereas explanative psychology *qua* natural science begins with general hypotheses, descriptive psychology may end with particular explanative hypotheses.

So far psychology has been considered mainly in relation to the description of inner experience. It should be pointed out that for Dilthey inner experience is not purely inwardly directed or introspective. Inner experience is often about external objects, but focuses on our attitude towards them. It is therefore in some way more

encompassing than outer experience, which excludes subjective attitudes. To move beyond this paradox Dilthey developed a new conception of experience that replaces the Kantian term *Erfahrung* with *Erlebnis* (lived experience). Lived experience is experience in its most inclusive sense and ‘contains a relation of inner and outer’.

Because lived experience discloses the original continuum of psychic life, it becomes necessary to reformulate the traditional conception of understanding which had been oriented primarily to our discrete outer experiences of natural phenomena. According to Kant, our experience of nature involves a discursive faculty of understanding (*Verstand*) that proceeds synthetically from partial representations to construct objective wholes (see Kant, I. §§6-7). Dilthey’s attack on faculty psychology undermines Kant’s pure intellectual conception of *Verstand*, which constructs the world in terms of fixed, abstract categories. When Dilthey speaks of understanding he means a very different process of *Verstehen*, which is concrete and develops historically. In so far as the *Verstehen* (understanding) of psychic life is based on lived experience it can be intuitive and proceed from the whole to the parts.

Kant’s *Verstand* as intellectual faculty was really geared to the scientific explanation of natural events. It allows us to relate the phenomena of nature, but provides no insight into their underlying reality. Dilthey accepts this limit for our knowledge of nature, but not for the understanding of ourselves and others. We cannot really understand nature, because it is not of our own making. Dilthey aligns himself with Vico by claiming that we can only truly know what we have ourselves made or contributed to (see Vico, G. §6). We thus have an access to psychological, social and historical reality - a recognition of ourselves in others, so to speak - which is impossible in relation to nature. Understanding constitutes the main goal of the human sciences in the way that explanation defines the natural sciences.

The fact that Dilthey points to the limits of scientific explanation has given some the mistaken impression that understanding is irrational. While explanation may be conceived as a purely intellectual process, understanding should not be characterized antithetically as a product of mere feeling or empathy. Lived experience does provide an immediate sense of the whole, but understanding, in appealing to all the powers of the psyche, does not overlook the intellectual processes. Sometimes understanding and explanation converge. Since understanding is contextual, it may need to bring external factors to bear which will then play a subordinate explanative role. Similarly, the description of lived experience must go over into analysis to bring out implicit structural relations more clearly.

There is an interdependence between the cognitive, emotive and volitional aspects of our consciousness which is at once structural and dynamic. While Dilthey begins with a discussion of representations, they are not considered as the underlying elements, with feelings and acts of will reduced to mere functions of representations. In some cases, representations may produce certain feelings which lead to a disposition to act. But in turn, representations themselves receive retroactive influences from our feelings and volitions. Clearly no aspect of psychic life can be understood in isolation or as most basic.

Dilthey regards development as a process of articulation whereby an indeterminate psychic continuum is differentiated into more distinctly related parts of a structural whole. The psychological concept especially formulated to define this formation of the self is that of the ‘acquired nexus of psychic life’ (*erworbener Zusammenhang des Seelenlebens*). This acquired psychic nexus embodies the history of the development of an individual and reveals the structural ordering of past experience. Encompassing knowledge of the world, evaluations and dispositions to act, it orients all present and future experience.

The individuality of the self is defined in terms of the structural articulation of the acquired psychic nexus. No qualitative uniqueness need be posited to explain individuality. What serves to distinguish individuals from one another is that the commonly held traits are manifested with differing forcefulness. In a given person, some qualities may be so faintly exhibited as to be, in effect, unobservable, others with such strength that we tend to notice them alone.

Dilthey’s discussion of individuality reflects one of the most pervasive themes of his philosophy - the understanding of human individuality as an essential goal of history and the human sciences. Dilthey meant to further such understanding with a comparative psychology, but he never completed it as he went through a transition in which he reconsidered some of his assumptions about the role of psychology and began to develop his

later hermeneutic approach to the human sciences.

3 The formation of the historical world

With his ‘Entstehung der Hermeneutik’ (‘The Rise of Hermeneutics’) in 1900, Dilthey began to sketch out a position which would define his work until his death. While he does not abandon the psychological description of lived experience, Dilthey comes to view its ability to capture the meaning of our life as more limited. Much of the meaning of our experience remains unconscious until it is expressed. Thus the description of the life of the subject cannot be done without the interpretation of its expressed objectifications. Dilthey resumed the task of a Critique of Historical Reason from this hermeneutical perspective, publishing *Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften* (*The Formation of the Historical World in the Human Sciences*) in 1910.

There is a further problem which occasioned Dilthey’s move to hermeneutics and underscores the limitations of description. The descriptions of psychology seem to function primarily on the scientific level of representational consciousness which - however well-integrated - stands apart from the world. How do we deal with prescientific reflexive awareness according to which we are already part of the world? Can description also be applied to its more inclusive life-nexus? Here Dilthey found inspiration in Husserl’s *Logische Untersuchungen* (*Logical Investigations*) and its theory of intentionality. Dilthey had given a seminar on this work in the Winter Semester 1904/5. Husserl in turn visited Dilthey and wrote that this meeting inspired him to occupy himself extensively with the problems of the human sciences (see Husserl, E. §2).

In Dilthey’s ‘Ideen’, understanding was used to describe the connectedness of facts of consciousness. According to his later hermeneutical writings, understanding must clarify these connections as meaning-relations. Since the meaning-relations of our experience and expressions are not necessarily measurable in terms of the framework of the psychic nexus, Dilthey turns to historical life as the ultimate framework for interpretation. Whereas the acquired psychic nexus incorporated the objective sphere into the life of the individual subject, in the *Aufbau* Dilthey becomes equally concerned to conceive the inverse way in which subjects are intentionally related to objective and public spheres. To do so Dilthey appropriates the Hegelian term ‘objective spirit’ as the overall historical context for understanding. He rejects Hegel’s particular definition of objective spirit as the socio-historical stage of the self-realization of absolute spirit in favour of a concept consistent with a reinterpretation of spirit as human activity. ‘Objective spirit’ designates the whole range of human objectifications, whether they be expressions in language and other media meant to communicate, or practices and deeds meant to influence. This objective spirit is at once the embodiment of human thought and action as well as the medium within which they occur. It includes the contexts we share to make interaction possible: not only the sociopolitical institutions originally considered by Hegel, but also the cultural institutions of art, religion and philosophy which he had classified as absolute spirit (see Hegel, G.W.F. §8).

Objective spirit provides the kind of overall framework for Dilthey’s hermeneutics that tradition comes to provide in Gadamer’s hermeneutics. But objective spirit is not as dominant as tradition in Gadamer, for Dilthey articulates it into more specific *Wirkungszusammenhänge* (systems of reciprocal influence), whether they be historical epochs or social and cultural systems. This is relevant to Dilthey’s distinction between elementary and higher understanding. The former orients an expression to the common context of objective spirit. That is, the elementary understanding of a sentence focuses only on what it explicitly asserts and is commonly assumed to mean. Problems concerning the implicit meaning of expressions call for higher understanding, which requires us to refer to more specific systems of reciprocal influence. Thus we attempt to determine any ambiguous expressions in a legal document by considering the particular legal system of the period in which it was drawn up. Only after having exhausted what the appropriate public contexts of expressions can do to clarify their objective meaning, should we turn to the subjective or psychological context. Now psychology is no longer the first of the human sciences, but the last. The highest or last mode of understanding is the *Nacherleben* or re-experiencing of an expression of lived experience. This re-experiencing does not involve reproducing the state of mind of the author. Hermeneutics expects us to understand authors not as they understood themselves, but better.

Although his hermeneutics ultimately focuses on the relation between lived experience, expression, and understanding (*Erlebnis*, *Ausdruck*, *Verstehen*), Dilthey admits that many expressions such as mathematical formulas and handshakes expressing agreement can be understood apart from relating them to lived experience. But when an expression does articulate our lived experience, as in a work of art, it can enrich our understanding of

life in immeasurable ways. This is why Dilthey’s writings on the imagination of artists bear importantly on his theory of interpretation.

4 Anthropological reflection on life and types of worldview

Dilthey’s aesthetic writings play a central role in his philosophy. They not only contribute to his conception of interpretation as it functions in the theory of the human sciences, but they also explicate the way in which we reflect about the meaning of life in everyday existence. There are many levels at which we try to understand what happens in life. Both prescientific anthropological reflection and scientific psychological description contribute to the understanding of order in our lives. Similarly, the categories whereby we establish order in the world are by no means purely scientific, but can in some cases be traced back to the ‘syntactical articulation of language’. Sometimes the same category (for example, Aristotle’s acting and suffering) can receive varying formulations as it functions at the level of ordinary experience (for example, efficacy), of the human sciences (for example, influence), and of the natural sciences (for example, causality). Other categories such as value, purpose and meaning are distinctive of the human sciences. The peculiar fascination of the arts, especially the literary arts, is that they can somehow relate all these levels in the search for order, without, however, producing a conceptual system. Like philosophers and religious thinkers, certain great poets have the capacity to articulate a comprehensive worldview.

One of Dilthey’s last essays was *Die Typen der Weltanschauung und ihre Ausbildung in den metaphysischen Systemen* (*Types of Worldview and their Development in Metaphysical Systems*) (1911). Whereas the kinds of worldview found in art and religion are quite diverse, worldviews in Western philosophy have received a more conceptual, metaphysical formulation that allows them to be distinguished into three main types: naturalism, the idealism of freedom and objective idealism. Naturalism as found in Democritus, Hobbes and others, reduces everything to what can be perceived or determinately cognized and is pluralistic in structure; the idealism of freedom as found in Plato, Kant and others insists on the irreducibility of the will and is dualistic; objective idealism as found in Heraclitus, Leibniz and Hegel affirms reality as the embodiment of a harmonious set of values and is monistic. Each of these metaphysical types of worldview is a perspectival interpretation of reality respectively emphasizing either our cognitive or representational capacities, our volitional ends, or what is felt to be valuable. Despite being totalistic, these types of worldview cannot attain absolute knowledge, according to Dilthey.

Because Dilthey analyses three incommensurable types of worldview that recur, he is often considered a relativist. In his essay ‘Philosophie als strenge Wissenschaft’ (‘Philosophy as a Rigorous Science’), Husserl quotes some passages from this essay that also seem to lead Dilthey down the path of historicism. In one of them, Dilthey uses ‘the development of the historical consciousness’ to question the universal validity of any metaphysical worldview that claims to comprehend conceptually how everything in the world is interconnected. By appealing to historical consciousness, Dilthey is not, as Husserl thinks, making a mere factual claim about the inability of past metaphysical speculative systems to gain universal acceptance. Historical consciousness is for Dilthey a broadening perspective that takes claims out of their actual local contexts and locates them in the sphere of universal history. It is a product of the Enlightenment and could even be considered the counterpart of the transcendental point of view. Indeed, Dilthey’s stance here is analogous to Kant’s in rejecting metaphysical speculation and is no more relativistic or historicist than Kant’s standpoint.

Dilthey’s historical consciousness makes possible a critical analysis of worldviews, and it is this kind of analysis that shows that metaphysical worldviews cannot be scientifically adjudicated. Any effort to provide a comprehensive scientific account of reality would have to synthesize the results of the natural sciences and the human sciences. But ultimately the approaches of these two kinds of science are so different that they cannot in principle be synthesized. Dilthey’s appeal to historical consciousness is thus not at all a challenge to the objective validity of science. It is instead an attempt to preserve the objectivity of scientific *Erkenntnis*. Worldviews, in turn, have their main value for Dilthey as reflective articulations of the meaning of our own life-nexus as given in prescientific *Wissen*.

See also: Hegelianism §6; Hermeneutics; Historicism §1; History, philosophy of §4; Phenomenological movement; Social science, history of philosophy of §8

RUDOLF A. MAKKREEL

List of works

Dilthey, W. (1914-90) *Gesammelte Schriften (Collected Works)*, Göttingen: Vandenhoeck & Ruprecht, 20 vols. (The standard complete works.)

Dilthey, W. (1985-) *Selected Works*, ed. R.A. Makkreel and F. Rodi, Princeton, NJ: Princeton University Press, 6 vols.(Vol. 1 (1989), vol. 4, *Hermeneutics and the Study of History* (1996), and vol. 5, *Poetry and Experience* (1985), are the three volumes now in print.)

Dilthey, W. (1883, 1982) *Einleitung in die Geisteswissenschaften*, trans. R.A. Makkreel and F. Rodi, *Introduction to the Human Sciences*, in *Selected Works*, ed.R.A. Makkreel and F. Rodi, Princeton, NJ: Princeton University Press, vol. 1, 1989.(Includes all of Book One and materials from Books Two, Four, Five, and Six.)

Dilthey, W. (1894) 'Ideen über eine beschreibende und zergliedernde Psychologie', trans. R. Zaner and K. Heiges, 'Ideas Concerning a Descriptive and Analytic Psychology', in *Descriptive Psychology and Historical Understanding*, with intro. by R.A. Makkreel, The Hague: Martinus Nijhoff, 1977.(The translated volume also contains 'The Understanding of Other Persons'.)

Dilthey, W. (1900) 'Entstehung der Hermeneutik', trans. R.A. Makkreel and F. Rodi, 'The Rise of Hermeneutics', in *Hermeneutics and the Study of History*, vol. 4 of *Selected Works*, ed. R.A. Makkreel and F. Rodi, Princeton, NJ: Princeton University Press, 1996.(This volume of the *Selected Works* also contains 'Das hermeneutische System Schleiermachers in der Auseinandersetzung mit der älteren protestantischen Hermeneutik' (Schleiermacher's Hermeneutic, A System in Relation to Earlier Protestant Hermeneutics) (1960).)

Dilthey, W. (1910) *Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften*, trans. R.A. Makkreel and F. Rodi, *The Formation of the Historical World in the Human Sciences*, in *Selected Works*, ed. R.A. Makkreel and F. Rodi, Princeton, NJ: Princeton University Press, vol. 3, (forthcoming).(Here Dilthey comes closest to working out his Critique of Historical Reason.)

Dilthey, W. (1911) *Die Typen der Weltanschauung und ihre Ausbildung in den metaphysischen Systemen*, trans. W. Kluback and M. Weinbaum, *Types of Worldview and their Development in Metaphysical Systems*, in *Dilthey's Philosophy of Existence*, New York: Bookman Associates, 1957.(Relates worldviews in poetry, religion and philosophy.)

Dilthey, W. (1976) *Selected Writings*, ed. H.P. Rickman, Cambridge: Cambridge University Press.(Contains excerpts from various works, including *The Formation of the Historical World in the Human Sciences* (1910).)

References and further reading

Ermarth, M. (1978) *Wilhelm Dilthey: The Critique of Historical Reason*, Chicago, IL: University of Chicago Press.(A comprehensive account of Dilthey's thought with a good historical background.)

Husserl, E. (1900-1) *Logische Untersuchungen*, trans. J.N. Findlay, *Logical Investigations*, London and New York: Routledge, 1977.(Classical formulation of the theory of the intentionality of consciousness and of such phenomenological concepts as act, content and object, expression and meaning.)

Husserl, E. (1910-11) 'Philosophie als strenge Wissenschaft', trans. G. Laver, 'Philosophy as a Rigorous Science', in *Phenomenology and the Crisis of Philosophy*, New York: Harper & Row, 1965.(An attack on worldview philosophy as lacking in rigour.)

Makkreel, R.A. (1975, 1992) *Dilthey: Philosopher of the Human Studies*, Princeton, NJ: Princeton University Press.(A developmental examination of Dilthey's philosophy with special attention to its relation to Kant;updated to deal with recently published posthumous writings.)

Makkreel, R.A. and Scanlon, J. (eds) (1987) *Dilthey and Phenomenology*, Washington, DC: Center for Advanced Research in Phenomenology and University Press of America.(Exploration of Dilthey's relation to phenomenology by ten international scholars.)

Orth, E.W. (ed.) (1985) *Dilthey und die Philosophie der Gegenwart (Dilthey and Contemporary Philosophy)*, Freiburg, Alber Verlag.(Exploration of Dilthey's relevance to a variety of topics by twelve scholars, including Apel and Gadamer.)

Owensby, J. (1994) *Dilthey and the Narrative of History*, Ithaca, NY: Cornell University Press.(A topical study focusing on recently published posthumous writings.)

Rodi, F. (ed.) (1983-) *Dilthey-Jahrbuch für Philosophie und Geschichte der Geisteswissenschaften (Dilthey-Yearbook for Philosophy and the History of the Human Sciences)*, Göttingen: Vandenhoeck & Ruprecht.(Published essays on Dilthey and related figures and themes. Also provides bibliography updates.)

Rodi, F. and Lessing, H.-U. (eds) (1984) *Materialien zur Philosophie Wilhelm Diltheys (Materials Relevant to the Philosophy of Wilhelm Dilthey)*, Frankfurt am Main: Suhrkamp.(A collection of classical essays on Dilthey by such thinkers as Scheler, Landgrebe, Bollnow, Plessner, Marcuse, Misch, Habermas and Gadamer.)

Diodorus Cronus (late 4th-early 3rd centuries BC)

The most famous member of the Dialectical school, the Greek philosopher Diodorus Cronus maintained various paradoxical theses. He argued that any attempt to divide space, time or matter must end with little regions, periods or bodies that cannot further be divided; hence, he inferred, things cannot be in motion. Diodorus also contributed to the contemporary debate on conditionals: one proposition implies another, he held, if and only if it never has been possible, and is not now possible, to have the former proposition true and the latter proposition false. Diodorus is however most famous for inventing the master argument. The master argument relied on two assumptions: that every past truth is necessary, and that the impossible does not follow from the possible. It concluded, on these assumptions, that no proposition is possible unless it either is true or will be. The master argument was designed to support Diodorus' definition of possibility: a proposition is possible if and only if it either is or will be true. This definition is not exactly tantamount to the fatalist doctrine that all truths are necessary, but it was felt to come too close to fatalism for comfort.

1 Life and reputation

Born in Iasos, in Asia Minor, during the last half of the fourth century BC, Diodorus died around 284 BC. The abusive nickname 'Cronus' ('Old Fogey') he inherited from his teacher Apollonius Cronus. Diodorus was the greatest showman among the Dialecticians (see Dialectical school), a school to which, it is said, his five daughters also belonged. He held that words get their meaning not from nature nor even from convention, but instead from the intention with which they are uttered; and that in consequence none of his words were ambiguous unless uttered with the intention of saying more things than one. His views on meaning he illustrated by giving a daughter the male name 'Theognis', and by naming a slave with the Greek equivalent of 'In fact however'. He helped refine various already current paradoxes such as the liar paradox ('What I am saying is false'), the veiled paradox ('You don't know who the man behind the veil is. He is your father. So you don't know who your father is') and the paradox of the Sorites (which in Diodorus' hands became 'If two in a trio of atoms are in motion, then the trio are predominantly in motion; if the trio are predominantly in motion, then a quartet to which that trio belong are also predominantly in motion; and so on; so a myriad of atoms will be predominantly in motion, even if only two of them are actually moving'). Diodorus' most celebrated contributions were his arguments against change, his theory of logical consequence and the views on modality which he defended with his master argument. It was said that 'even the crows on the roof-tops caw' about the issues Diodorus raised; and conversation about the master argument was a notoriously effective way of blighting dinner parties. In modern times, the master argument has become the subject of a massive literature, much of it highly speculative.

2 Atomism and change

Diodorus held that all things consist of 'minimal and partless bodies' so small as to be indivisible, but big enough for there to be only finitely many of them in any finitely sized body. He held similar views about time and space: any larger stretch of time may be divided into short periods, each of length greater than zero but with no smaller parts into which it can be divided; and each finite space consists of finitely many indivisible volumes.

Diodorus concluded that nothing is ever moving. Sextus Empiricus gives the fullest single report of Diodorus' many arguments on these matters:

> If something is moving, it is moving now. If it is moving now, it is moving in the present time. If it is moving in the present time, it is moving in a partless time. For if the present time is divided, it will be divided completely into the past and future, and so will no longer be present. If something is moving in a partless time, it traverses partless places. If it traverses partless places, it is not moving. For when it is in the first partless place, it is not moving; for it is still in the first partless place. And when it is in the second partless place, again it is not moving, but has moved. So nothing is moving.
>
> (*Against the Professors* X 119-20)

Although Diodorus denies that things move, he concedes that they are not now where they were, and that they therefore have moved. If we wonder how a thing can have moved without ever moving, Diodorus points out that Helen had three husbands, even though the present-tensed 'Helen has three husbands' was never true; for her

husbands were successive, not simultaneous.

3 Conditionals and consequence

It was widely agreed that the relation between antecedent and consequent in a true conditional is the same as that between premises and conclusion in a valid argument: the conclusion (or consequent) was said to ‘follow from’ the premises (or antecedent). This was unfortunate. It implies that in an argument like:

It is day.
If it is day, it is light.
So, it is light.

the conditional premise is either false or redundant. The parties to the unfortunate agreement that these two relations are identical did however dispute about what the single relation was (see Philo the Dialectician §2).

Diodorus was party to the agreement, and also to the dispute. He declared that one proposition follows from another if and only if it neither was nor is possible for the former to be false while the latter is true.

An immediate consequence of Diodorus’ definition is that if one proposition now follows from another, then it has always done so. This consequence is a defect if Diodorus’ definition is taken as an account of which conditionals are true. For by holding out my right arm I can make true a previously false conditional: ‘If I hold out my left arm, both my arms will be held out’. This consequence is however a virtue if Diodorus’ definition is taken as an account of which arguments are valid. For it would be astonishing if a previously invalid argument should become valid.

On Diodorus’ definition, everything follows from a proposition that is and always has been impossible; and a proposition that is and always has been necessary follows from everything. Diodorus’ atomism was invoked to give a piquant illustration of both points simultaneously: on Diodorus’ account, ‘There are minimal and partless bodies’ follows from its own negation, ‘There are no minimal and partless bodies’.

Diodorus’ explanation of following can be explained further. The result is: one proposition follows from another if and only if there is no time when the former is false and the latter is true. This result is reached by applying, to Diodorus’ original definition of following, his definition of possibility.

4 The definition of possibility

Diodorus defined the possible as what is or will be true (for example, it is possible for me to be in Corinth if and only if I am or will be there), and the necessary as what, being true, will never be false. Diodorus therefore disagreed both with the Aristotelian idea that the necessary is what is true at every time, and with the Megaric idea that the possible is what is true now (see Megarian school).

One curious consequence concerns change of modalities. On Diodorus’ definitions, what was once contingent can become either necessary or impossible. The reverse change cannot take place: once something is either necessary or impossible it remains so for ever. As time passes, previously open possibilities therefore get closed off. Moreover, the only way a previously open possibility can get closed off is by being actualized: if something once was possible but is no longer so, then it has been actually true in the meantime.

A second curious consequence concerns ‘future contingents’, propositions of the form ‘So-and-so will happen’ that are neither necessary nor impossible. Suppose that ‘Adam will sin’ is such a proposition. If ‘Adam will sin’ is not impossible, then, on Diodorus’ definition, it either is true now or will be true in the future; either way, Adam will sin. If ‘Adam will sin’ is not necessary, then, on Diodorus’ definition, it either is or will be false. But since Adam will sin, the first alternative can be eliminated. Hence ‘Adam will sin’ will be false at some time in the future. Moreover, once such a proposition is false, it will remain false for ever thereafter. Hence, on Diodorus’ definition, ‘Adam will sin’ will become and remain impossible. This reasoning is generally applicable: on Diodorus’ definition, each future contingent is now true and will eventually become permanently impossible.

A third curious consequence concerns past necessities. If a proposition has been true, then it always will have been so. By Diodorus’ definition, past truths are therefore necessary. This applies only to simple past tensed truths, like ‘Diodorus was born in Iasos’. The more complex truth ‘Diodorus’ fragments have never yet been properly edited’ is contingent if his fragments will be properly edited in future. And the metric truth ‘Diodorus was born nearly 25

centuries ago' can also be false, for it will be.

Why should Diodorus' definition be accepted? It is obvious enough that if something is or will be true, then it is possible. More controversial is the converse, that if something is possible, then it is or will be true. To establish the converse, Diodorus devised the master argument.

5 The master argument

According to Epictetus:

> The master argument seems to have been put forward from some such starting points as these. There is a mutual inconsistency of these three propositions with one another:
>
> Every past truth is necessary;
>
> The impossible does not follow from the possible;
>
> Something is possible which neither is true nor will be.
>
> Diodorus saw this inconsistency and used the plausibility of the first two propositions to establish that nothing is possible which neither is true nor will be.

(*Discourses* II 19.1)

The word translated 'past' is the most general term for verbs in any of Greek's various past tenses. Hence Diodorus' first premise presumably means that each true proposition in a past tense is necessarily true. This interpretation is that of Chrysippus, who accepted the first premise (Epictetus, *Discourses* II 19.2-5) and who defended it on the grounds that past truths cannot change from true to false (Cicero, *On Fate* 14): Chrysippus' reasoning is entirely cogent if by 'past truth' we mean only simple past tensed truths, like 'Diodorus was born in Iasos'.

How did Diodorus reason from his first two premises to the negation of the third? We learn of the first two premises because they were disputed by Stoics, who rejected Diodorus' conclusion but who could find nothing other than these two premises to contest (see Cleanthes). Hence our reconstruction should not make obvious use of other premises more likely to be contested by the Stoics than these first two. Here is one reconstruction to meet this standard.

Let us take the proposition 'Diodorus snores'. From this proposition, it follows that 'Diodorus snores' never was at any past time going to be false for ever thereafter. Hence, if this proposition is possible, then it is possible also that 'Diodorus snores' never was at any past time going to be false for ever thereafter (for the impossible does not follow from the possible). If however our proposition now is and always will be false, then 'Diodorus snores' was at some past time going to be false for ever thereafter; indeed, it is necessary that 'Diodorus snores' was at some past time going to be false for ever thereafter (for every past truth is necessary). Hence if the same proposition is and always will be false but is nevertheless possible, we have a contradiction: that the proposition will always be false has necessarily been so at some time, and also has possibly never been so.

See also: Logic, Ancient §5; Prior, A.N.; Tense and temporal logic

NICHOLAS DENYER

References and further reading

Denyer, N. (1981a) 'The Atomism of Diodorus Cronus', *Prudentia* 13: 33-45.(An expansion of the material in §2.)

Denyer, N. (1981b) 'Time and Modality in Diodorus Cronus', *Theoria* 47: 31-53.(An earlier version of the material in §§3-5.)

Döring, K. (1972) *Die Megariker (The Megarians)*, Amsterdam: Grüner.(Collection of ancient sources, together with some German commentary.)

Gaskin, R. (1995) *The Sea Battle and the Master Argument*, Berlin and New York: de Gruyter.(Contains ample discussion of rival interpretations, together with a massive bibliography.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae (The Remains of Socrates and the Socratics)*, Naples: Bibliopolis. (Volume 1 pages 413-35 contains an almost complete collection of the ancient evidence about

Diodorus.)
Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press. (Contains some of the main sources, together with English commentaries and translations.)
Prior, A. (1967) *Past, Present and Future*, Oxford: Clarendon Press.(Chapter 2 axiomatizes Diodoran modal logic.)
Sedley, D. (1977) 'Diodorus Cronus and Hellenistic Philosophy', *Proceedings of the Cambridge Philological Society* 203, new series, 23: 74-120.(Comprehensively sets Diodorus in his historical context.)
White, M.J. (1992) *The Continuous and the Discrete: Ancient Physical Theories from a Contemporary Perspective*, Oxford: Clarendon Press.(Expounds, sometimes rather technically, Diodoran atomism and its ancient rivals.)

Diogenes Laertius (*c.* AD 300-50)

Diogenes Laertius is the author of a famous work entitled Lives of the Philosophers consisting of nearly one hundred accounts of individual philosophers. These contain mainly biographical information, but sometimes also include doctrinal summaries. The work is extremely valuable because it preserves much information on Greek philosophers from sources now lost.

Diogenes Laertius is the author of the only work to come down to us from antiquity that gives a comprehensive account of the Greek philosophical tradition. It consists of ten books, which contain eighty-four accounts of individual Greek philosophers, varying in length from an entire book to a few lines. Very little is known about the author. A disputed text (IX 109) suggests that his home town was Nicaea in Bithynia (not far from Byzantium). This provincial location might explain the somewhat dated quality of his erudition. Diogenes scarcely mentions philosophers of the Roman period except for a list of Sceptics at IX 116. The last-named of these is a pupil of Sextus Empiricus, which indicates that he wrote the book no earlier than the first half of the third century AD. Diogenes himself does not belong to any philosophical school, but does appear to have sympathy for the philosophy of Epicurus (X 29).

An older title of Diogenes' work, known since the Middle Ages simply as *Lives of the Philosophers*, gives a more accurate indication of its content: *Lives and Maxims of Those who Gained Fame in Philosophy and the Doctrines of Each School*. The overall structure of the work is determined by the theory of successions in Greek philosophical historiography (see Doxography), as the following summary shows:

Book I the seven sages, including Thales;
Book II the Ionian succession, including Anaximander, Anaxagoras, Socrates and the Socratics;
Book III Plato;
Book IV Plato's successors in the Academy;
Book V Aristotle and his successors in the Lyceum;
Book VI Antisthenes, Diogenes and the Cynic philosophers;
Book VII Zeno and the Stoic philosophers (some twenty chapters have gone missing at the end);
Book VIII the Italian succession, headed by Pythagoras and Empedocles;
Book IX the 'scattered' philosophers, including Heraclitus, Parmenides and Democritus;
Book X Epicurus.

Only one woman philosopher receives a separate chapter: Hipparchia the Cynic (VI 96-8).

The work is in an unfinished state. Individual chapters or 'biographies' contain a remarkable variety of material, giving the work an uneven and compilatory quality. Doctrines are only presented in detail when the chapter deals with the founder of a school (for example, the extremely valuable doxography on the Stoics in Book VII). A typical chapter includes the origin of the philosopher and who his teachers were, the date when he flourished (*akmē*), anecdotes illustrating his life, maxims, lists of works (extremely valuable), other documents such as wills and letters (the latter usually fictitious), an epigram by the author on the philosopher's death, and a list of other persons with the same name. Diogenes does not write a history of philosophy in the modern sense. He combines biographical and doctrinal material because he is convinced, like most writers on philosophy in antiquity, that there is a intrinsic relation between the *bios* (life) and the *logos* (doctrine) of every philosopher.

The value of the information supplied by Diogenes is entirely dependent on the quality of his sources. The nature and extent of these has long been the subject of debate. It is now thought that he has excerpted and stitched together material from widely scattered sources. These date mainly from the Hellenistic period, and include many important works in the historiography of ancient philosophy which are no longer extant. Since the Byzantine period Diogenes' work has been extensively used. No other work provides us with so much information on ancient philosophers. Despite its undeniable mediocrity, it is truly priceless.

DAVID T. RUNIA

List of works

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols; revised H.S. Long, 1972.(English translation with Greek text facing; outdated edition, but not yet superseded.)

References and further reading

Knoepfler, D. (1991) *La vie de Ménédème d'Érétrie de Diogène Laërce: contribution à l'histoire et à la critique du texte des Vie des Philosophes (Diogenes Laertius' Life of Menedemus of Eretria: A Contribution to the History and Evaluation of the Text of the Lives of the Philosophers)*, Basle: Friedrich Reinhardt.(Important remarks on the state of the textual tradition.)

Mansfeld, J. (1986) 'Diogenes Laertius on Stoic Philosophy', *Elenchos* 7: 295-382. (Includes important general remarks on Diogenes' method, together with a detailed analysis of how he proceeds in Book VII on the Stoics.)

Mejer, J. (1978) *Diogenes Laertius and his Hellenistic Background*, Hermes Einzelschriften 40, Wiesbaden: Franz Steiner.(Very good on Diogenes' mediated use of Hellenistic sources.)

Diogenes of Apollonia (5th century BC)

Diogenes was the last of the early Greek physicists. He claimed that interactions between things would be impossible unless all were forms of one basic substance. Adapting ideas of Anaximenes and Anaxagoras, he identified the basic substance as air, which in its optimal form possesses intelligence and thereby controls the universe at large and animal life in particular. Diogenes worked out a detailed psychology and physiology, explaining sense perception as an exercise of intelligence due to interaction between air in the region of the brain and atmospheric air. This theory was mocked by Aristophanes, but influenced various Hippocratic writings.

1 Life and work

Almost nothing is known of Diogenes' life. His birthplace Apollonia was probably the Milesian colony of that name on the Black Sea. Diogenes Laertius (IX 57) dates him to the same period as Anaxagoras (500-428 BC). Aristophanes' allusion to his ideas (*Clouds* 227-33) and Theophrastus' statement that he was pretty much the last of the Presocratics (A5) suggest someone younger. Simplicius attributes the surviving fragments to a general work on nature (A4), but the sequence of topics would seem to point to another of the treatises by Diogenes - *On the Nature of Man* - as their probable source.

Diogenes is elsewhere credited like the atomists with belief in an infinite number of worlds, and with a cosmogony and account of the earth and the heavenly bodies which largely recapitulate Anaxagoras' theories (A1, A6, A12) (see Anaxagoras §4). Indeed, Theophrastus explicitly represented him as drawing eclectically from these two major fifth-century systems (A5). But the work from which Simplicius quotes evidently had a quite different theme. It began with metaphysical argument for the fundamental identity of all physical things (fr. 2). Then, consequent upon proofs that the universe is permeated with intelligence (frs 3 and 4), it contended that the seat of intelligence is air, which is a single substance, despite variations in temperature, moisture and mobility etc. (fr. 5). Simplicius assures us that this substance is the eternal imperishable body whose existence was derived a priori in the opening argument (frs 7 and 8).

There followed a sequence of proofs designed to substantiate the dependence of animal life and intelligence upon air. Simplicius mentions demonstrations that sperm is aerated and that acts of intelligence occur when air via blood takes possession of the whole body. He refers to an anatomy of the human veins, which is preserved more or less verbatim by Aristotle (*History of Animals* III 2, see Diogenes fr. 6). Diogenes probably gave his account of sense perception, extensively reported by Theophrastus (A19), at this point in the book, along with discussions of physiological phenomena like sleep and digestion.

These detailed explanations appear to have constituted the real focus of the work. Its structure is just what one might expect of a fifth-century book 'on the nature of man': Polybus, author of the Hippocratic treatise bearing this title, complains of the way some writers on human physiology preface their accounts with grandiose claims, supported by meaningless proofs, about a single basic substance.

2 Monism

Ancient and modern historians of philosophy alike see Diogenes' theory as reviving the system of Anaximenes a century earlier (see Anaximenes §1). But it is a revival informed by reflection on more recent philosophy. Thus the argument for Monism found in fragment 2 has clear affinities with theses of Anaxagoras and Democritus, although Democritus may depend on Diogenes; there are also verbal echoes of Melissus and Empedocles. Diogenes proposes that two objects cannot mix with or benefit or harm each other, nor can one grow or be born from another, unless they are really - 'in their nature' - the same. Since he assumes that the four elements all do interact with one another, he concludes that all of them, and therefore all things whatsoever, are forms of the same basic substance. Diogenes' monism therefore allows for a world of variety and change. Nor does he acknowledge any need to rebut Melissus' proofs of the logical impossibility of alteration, even if the wording of fragment 2 betrays his familiarity with them.

Diogenes' arguments that there is 'much intelligence' in the basic substance recall Heraclitus as well as Anaxagoras. He appeals principally to the optimal disposition of day and night and the seasons, but also to the role of soul and intelligence in sustaining animal life (frs 3 and 4). Yet this is only preparatory to enunciation of his

main thesis: the identification of air as what possesses intelligence. Once again Diogenes echoes Anaxagoras (fr. 12: see Anaxagoras §4), but this time to subvert his insistence that mind (*nous*) is something pure from admixture with any other kind of substance. For Diogenes it is not mind, but the air that is its physical seat which permeates all things and thereby organizes and controls them. It is accordingly air that is to be regarded as what is strong and eternal and 'knowing many things' (fr. 8), and as thereby deserving ascription of the divinity Anaxagoras effectively predicated of mind. Mind is in fact an attribute, not a substance at all. Hence no doubt the substitution of *noēsis*, 'intelligence', for Anaxagoras' *nous*.

Only air in a particular state has the power of intelligence; namely, when hotter than atmospheric air, but cooler than that in the vicinity of the sun (fr. 5). The surviving fragments say nothing of how air can take the form of, for example, earth or water (although the doxography speaks of compression; A1, A6). We might wonder why evidence that air pervades and controls everything should suggest that it is the basic substance. No doubt the underlying thought is simply that no substance *could* be dominant without being basic.

3 Psychology and physiology

If Diogenes conceived his chief general contribution to philosophy as rescuing Anaxagorean metaphysics from dualism, his work in psychology and physiology was, as noted, more influential. Diogenes' systematic treatment of these topics challenges comparison with those of Anaxagoras and Empedocles. Following the argument given in fragment 2, he made sense perception a product of interaction between air within and air without (Theophrastus, *On the Senses* 39-45). His exposition began unusually but understandably with the best cases for his theory: smell and then hearing. A weaker explanation is provided for sight and taste however. Diogenes' fanciful analysis of the vascular system allowed him to argue that air was conveyed by blood throughout the body to various destinations, including notably the head. Following Alcmaeon he conceived of the brain as the principal seat of sensations, which were apparently construed as functions of intelligence: Diogenes observed that when we have our minds on other things, we neither see nor hear. In general, the finer the air within and the more delicate the passages through which it is conveyed, the more acute the associated perceptions.

This explanatory scheme was applied to a great variety of phenomena. To take just one, pleasure is viewed as the result of efficient mixture of air and blood in the veins permeating the whole body. Many undesirable conditions are put down to the effect of moisture or *ikmas*. Evidently a key Diogenean notion, this was used to explain such phenomena as magnetism and the flooding of the Nile (A18, A33). Its action upon internal air inhibits intelligence, as manifested in sleep and drunkenness, and may also (to Aristophanes' amusement) be evidenced in animals which breathe air close to the ground.

See also: Pneuma §1

MALCOLM SCHOFIELD

References and further reading

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge & Kegan Paul.(The final chapter is an unorthodox but philosophically stimulating account of Diogenes.)

Diogenes of Apollonia (5th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 2, 51-69.(The standard collection of the ancient sources both fragments and testimonia, the latter designated by 'A'; includes Greek texts with translations in German.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols.(The most detailed and comprehensive English-language history of early Greek thought; a serviceable treatment of Diogenes can be found in volume 2 pages 362-81.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(A valuable survey of Presocratic philosophy, including texts and translations; contains an engaged but judicious evaluation of Diogenes.)

Laks, A. (1983) *Diogène d'Apollonie*, Lille: Presses Universitaire de Lille and Paris: Éditions de la Maison des sciences de l'homme.(The fullest modern account of Diogenes, fundamental for further scholarly work; includes all the texts in Diels and Kranz (1952) with French translations.)

Diogenes of Oenoanda (*c.* 2nd century AD)

The Epicurean philosopher Diogenes came from the Greek town Oenoanda in Lycia (Turkey). He is known exclusively for his massive philosophical inscription, erected in a colonnade there. Its remains contain sayings of Epicurus, plus Diogenes' own writings, mainly on physics and ethics. His Epicureanism is largely traditional, but possible innovations include talk of a future Epicurean golden age.

Diogenes is known for the inscription he erected in a colonnade in the town of Oenoanda, Lycia (Turkey), his birthplace. Most probably the inscription belongs to the second century AD. Diogenes informs the reader that he was already old and sick when he set up the inscription. We also learn that he was in contact with Epicurean friends in Athens, Chalcis, Thebes and Rhodes. The fact that he could afford to set up this enormous inscription in a public place testifies that he was a wealthy and influential man in Oenoanda.

Since 1884, 212 blocks or fragments have been discovered. The inscription probably covered a wall about 80 metres in length in the town's south stoa or colonnade, on the so-called esplanade. The text was arranged in either five or seven courses, carrying treatises on ethics, physics and old age. The ethics section (frs 28-61) was the lowest inscribed course and was underwritten by a course of maxims of Epicurus. They support Diogenes' argument, and are clearly meant to establish its Epicurean authenticity. The physics treatise, on the next course, starts with an introduction to the whole inscription. Higher courses carried letters of Diogenes (to his friends Antipater, Dionysius, Carus and Menneas) and his maxims, as well as letters attributed to Epicurus. Above stood Diogenes' treatise *On Old Age* (frs 137-79). Each section of the text looks like an opened papyrus scroll.

Diogenes wants the physics part to be read first, and at the beginning he announces the aim of his inscription: to help those who are 'well-constituted'. Like Epicurus he regards philosophy as a therapy, and intends to offer his philosophy as a medicine not only for his fellow citizens but also for foreigners: they are all citizens of one country, the world. The inscription, closely following traditional Epicureanism, attests its author's humanity and cosmopolitanism.

The physics and ethics treatises deal mainly with standard Epicurean topics. However, one remarkable fragment of the ethics treatise describes an Epicurean golden age, a future in which justice and friendship will render the protection of citizens by institutions unnecessary (fr. 56).

In the treatise on old age (frs 137-79) Diogenes addresses young readers and repudiates their negativity about old age, in which, he maintains, the mind remains active. Homer shows that old men are excellent speakers. If slow like the elephant, the old also resemble it in intelligence. The absence of sensual desires is not a disadvantage. Poverty is more valuable than wealth.

Among the letters and maxims, the letter to Antipater (imitating Epicurus' *Letter to Pythocles* in form and content; see Epicurus §2) sheds some light on Diogenes' life and his contacts with Epicurean friends in Greece. Philosophically, it deals with the infinite number of worlds. Other of Diogenes' fragments contain the remains of a letter to Dionysius (and Carus?) whose topics were epistemological, ethical and biographical. It speaks of Epicurus 'the herald who brought you complete salvation' (fr. 73 III 13), and we learn of Epicurus' narrow escape from death after a shipwreck. Most important are the fragments of a letter by Epicurus to his mother, which deals with visions of absent persons, the author's happiness and the fear of death, along with a plea not to send money any more since the author has sufficient (frs 125-6).

See also: Epicureanism §1

MICHAEL ERLER

References and further reading

Clay, D. (1990) 'The Philosophical Inscription of Diogenes of Oenoanda: New Discoveries 1969-1983', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36: 4, 2446-559.(Valuable survey of Diogenes' Stoa, the inscribed wall and individual documents of the inscription.)

Diogenes of Oenoanda (*c.* 2nd century AD) Fragments, ed. M.F. Smith, *Diogenes of Oinoanda, The Epicurean Inscription*, Naples: Bibliopolis, 1993.(Up-to-date edition, with English translation and lucid commentary, by

the scholar who has discovered numerous new fragments of the inscription.)
Puech, B. and Goulet, R. (1994) 'Diogène d'Oinoanda', in R. Goulet (ed.) *Dictionnaire des Philosophes Antiques*, Paris: CNRS Éditions, vol. 2, 803-6.(Guide to the modern scholarship, with extensive bibliography.)

Diogenes of Sinope (412/403-324/321 BC)

*Diogenes of Sinope was considered, along with Antisthenes, the founder of Cynicism. His nickname 'Cynic', literally 'doglike', reflects the highly unconventional lifestyle he lived and advocated. Radically re-evaluating mankind's relation to both nature and civilization, Diogenes redefined the individual's freedom and self-sufficiency, advocating a training (*askēsis*) for achieving both.*

Although a historical figure, Diogenes quickly became a literary character - probably in his own lost works, certainly in those of others. Hence his life, lost writings and oral teachings are intertwined in a complex tradition comprising: (1) a biographical strand transmitted by Diogenes Laertius - itself a collage of literary and oral traditions about Diogenes; (2) the more overtly literary representations by writers such as Lucian and Dio Chrysostom (*Discourses*). Contemporary evidence for Diogenes is virtually nil.

Critical assessment of the biographical tradition is more than usually important in Diogenes' case, since his thought is transmitted to us primarily in the form of pointed anecdotes and aphorisms (*chreiai*). Diogenes was exiled from Sinope (on the southern coast of the Black Sea) for defacing the city's coins. 'Defacing the currency' was to become a central metaphor for the Cynics' critique of the tradition - driving out the counterfeit coin of conventional wisdom to make room for the authentic Cynic life lived 'according to nature'. Surprisingly, numismatic evidence discovered in the twentieth century appears to confirm the story of Diogenes' exile, but that may be the only trustworthy part of the biography. The tradition also claims that Diogenes (1) studied with Antisthenes, (2) discovered his vocation ('defacing the currency') by consulting an oracle and (3) was sold into slavery and spent the rest of his life as a private tutor to his master's children. Claim (1) may be chronologically impossible, and was perhaps fabricated by Stoics eager to give their school a Socratic pedigree via the Cynics. Claim (2) is suspiciously reminiscent of stories told of Socrates and of Zeno of Citium. Claim (3) conflicts with the tradition that Diogenes grew old spending his summers in Corinth and winters in Athens living in his *pithos*, a large wine-jar. In all probability, both (2) and (3) are based on literary works by or about Diogenes that were later treated biographically.

Similarly, there are several conflicting versions of Diogenes' literary activity. Two authorities denied that Diogenes had left anything in writing, yet we have two lists of works attributed to him. The first consists of thirteen dialogues (including a *Republic*), some epistles and seven tragedies. A second list, probably of Stoic origin, consists of twelve dialogues (eight of which are absent from the first list) and some letters and sayings (*chreiai*), thus implicitly denying the authenticity of Diogenes' *Republic* and of his tragedies.

Diogenes' philosophical significance was a product of the manifold traditions purporting to represent him as much as it was of the facts of his life, of which we know few. Diogenes' unconditional pursuit of happiness in the face of exile and poverty led him to challenge the most fundamental ideas and taboos of Greek civilization and to valorize nature as a source of moral insight greater than that of custom or of the existing philosophical schools. The consequences of his experiment were remarkable and lasting, not only for philosophy - Stoicism and Epicureanism emerge in the philosophical context Diogenes helped to create - but for the literary and social traditions that antiquity transmitted to Europe.

See also: Cynics

R. BRACHT BRANHAM

References and further reading

Dio Chrysostom (*c*. AD 100) *Discourses*, vol. 1, trans. J.C. Cahoon and H. Crosby, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1932.(Parallel Greek text and English translation; Diogenes is represented in Orations 4, 6, 8-10.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols. (Parallel Greek text and English translation; Book VI is our most important source for Diogenes of Sinope.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae*, Naples: Bibliopolis, 4 vols.(Ancient sources in Greek

and Latin, with discussion in Italian; includes all the documentation relevant to Diogenes.)
Paquet, L. (1988) *Les Cyniques grecs: Fragments et témoinages*, Ottawa, Ont.: Les Presses de l'Université d'Ottawa.(An annotated French translation of the main Cynic fragments and testimonia.)

Discourse semantics

Discourse and its interpretation have interested philosophers since ancient times, and have been studied in different areas of philosophy such as rhetoric, the philosophy of language and the philosophy of literature. Discourse has attracted interest from philosophers working in the continental tradition, and it received considerable attention in the 1980s from analytic philosophers, philosophers of language, linguists, cognitive scientists and computer scientists; within these fields a formal, logical analysis of discourse interpretation, or discourse semantics, has emerged.

Discourse semantics arose in an attempt to solve certain problems that affected formal theories of meaning for single sentences. These problems had to do with the interpretation of pronouns and other 'anaphoric' elements in language. A detailed examination of the data showed that the meaning of an individual sentence in a discourse could depend upon information given by previous sentences in the discourse. To analyse this dependence, discourse semantics developed a formal analysis of a discourse context and of the interaction between the meaning of a sentence and the discourse context in which it is to be interpreted. The essential idea of discourse semantics is that the meaning of a sentence is a relation between contexts. The 'input' discourse context furnishes the information needed to interpret the anaphoric elements in the sentence; the information conveyed by the sentence when added to the input context yields a new, or 'output', discourse context that can serve to interpret the next sentence in the discourse.

1 Static semantics for discourse and its difficulties

To understand the development of discourse semantics, one must look at the problems that motivated its development. To do this, one must know something about the interpretation of the constituents of discourse - the constituent sentences and their constituent phrases as treated in standard formal semantics. In the 1960s, Richard Montague developed a very influential framework, Montague Grammar, for analysing how the meaning of a complex expression depends on the meanings of its constituents (see Compositionality; Possible worlds; Semantics). One might think that there is nothing more to the interpretation of a discourse than simply building up the meanings of its constituent sentences and then combining them. But once a clear and precise proposal for the semantics of sentences was given by Montague Grammar, several problems emerged that showed that the interpretation of discourse would not be so simple. There are two related phenomena in semantics that cause difficulty for Montague Grammar, namely pronominal anaphora and temporal anaphora, and these have spurred advances in the analysis of discourse.

Pronominal anaphora. One problem for the Montague Grammar approach to discourse is the interpretation of anaphoric pronouns. Pronominal anaphora occurs when a pronoun refers back to some word or phrase in the preceding discourse. In the discourses below, the anaphoric pronouns 'he' and 'it' refer back respectively to a 'farmer' and a 'donkey'.

(1) A farmer owns a donkey and he beats it.
(2) A farmer owns a donkey. He beats it.
(3) If a farmer owns a donkey, he beats it.

In semantics the meaning of a discourse like (1) can be given by a formula of some logical language with a clear model-theoretic interpretation - this formula is known as the logical form of the sentence. Montague Grammar uses the language of intensional, higher-order logic, but for our purposes the only thing we need to know about this language is that indefinite noun phrases like 'a farmer' are translated as existential quantifiers and pronouns are translated as bound variables. Thus, a sentence such as (1) has the translation:

$$\exists x \exists y(\text{farmer}(x)\ \&\ \text{donkey}(y)\ \&\ \text{owns}(x,\ y)\ \&\ \text{beats}(x,\ y))$$

The first problem for Montague Grammar is that since it gives the meaning of a discourse sentence by sentence, it is unclear how to connect the pronouns in the second sentence of (2) with the existential quantifiers that bind them in the first sentence. Another problem emerges with the attempt to treat (3), which is an example of the 'donkey sentence problem' originally discovered by Geach (1963). The meaning of (3) is captured by the formula:

$$\forall x \forall y((\text{farmer}(x) \ \& \ \text{donkey}(y) \ \& \ \text{owns}(x,\ y)) \ \rightarrow \ (\text{beats}(x,\ y))$$

But Montague Grammar has no uniform translation of the indefinite determiner *a* that yields a correct treatment of (1) and (3) satisfactorily; it does not provide any explanation of why the translation of indefinites is context-sensitive in this way, nor does it handle inter-sentential anaphora (see Anaphora).

Temporal structure of discourse. Observations by Barbara Partee and people working on the analysis of French tense in discourse first brought attention to a facet of the meaning of tenses that was missing from the best analyses of tense of the day. Those analyses took tenses to be temporal operators of the sort found in tense logic. In a French discourse like:

(4) Pierre entra dans le salon. Il s'assit sur la banquette. Il s'endorma.
(Pierre entered the room. He sat down on the sofa. He fell asleep.)

the three events introduced (Pierre's entering, his sitting down, and his falling asleep) all occur in the past if (4) is true. This much the analyses of tense were able to capture. But they miss a crucial generalization about texts like (4). The events described there are portrayed as occurring not only in the past but also in a definite sequence - that in which they are introduced in the discourse. We naturally understand the story as telling us that Pierre's entering the room occurred prior to his sitting down which in turn occurred prior to his falling asleep. On the operator view of tenses, a past-tensed sentence of the (logical) form $P\varphi$ is true just if φ holds for some time prior to the moment of speech; this view is incapable of capturing the contextual sensitivity shown in (4). Similarly incapable are views on which verbs are treated as predicates of events and the past tense introduces a relation of 'earlier than' between the event and the moment of speech (see Tense and temporal logic).

2 Dynamic semantics for discourse

Motivated by the problems with the interpretation of pronouns in discourse, Hans Kamp (1981) and Irene Heim (1982) independently proposed not only a solution to the interpretation of anaphoric elements but a new way of looking at discourse semantics. They redefined the semantic contributions in a discourse of a sentence and its constituents. In Montague Grammar, among other accounts of discourse interpretation, the contribution of a sentence is a proposition, or, formally, a set of possible worlds in which the sentence is true. Such a proposition contributes to the content of a discourse in a simple way: the meaning of a discourse is just the intersection of all the propositions that are the meanings of the discourse's constituent sentences. For Kamp and Heim, a sentence *S* when interpreted in a discourse *D* no longer simply yields a set of worlds; *S* may also yield a set of discourse entities or 'discourse referents' to which elements of subsequent discourse may refer. This approach came to dominate semantics in linguistics during the 1980s and 1990s. What follows develops Kamp's approach to discourse semantics, known as 'Discourse Representation Theory' or DRT.

DRT assigns a truth-conditional meaning to a natural language discourse in two steps: the 'DRS construction procedure' and the 'correctness definition'. The first step is to construct a representation of the content of the discourse (known as a 'Discourse Representation Structure' or DRS). The ontological status of a DRS has been subject to some debate: some view it as a partial model of the discourse, others view it merely as an alternative logical form for English, still others as some sort of conceptual structure. Nevertheless, the structure of a DRS is precisely defined. It consists of a pair of sets: (1) a set, or 'universe', of discourse referents, and (2) a set of 'conditions'. Discourse referents are analogous to the domain of a partial model - they are objects talked about in the discourse - while conditions are property ascriptions to these discourse individuals. The technical definition of a DRS and DRS condition is given through a doubly recursive definition. In thinking of DRSs as partial models, DRSs as abstract structures should be distinguished from the language used to describe them. The DRS language uses a box notation - the upper part of the box lists the discourse referents in the universe of the DRS, while the bottom part of the box describes the conditions. Kamp's notation is in fact reminiscent of Peirce's existential graphs. For instance, the DRS for 'a farmer owns a donkey' is given in (K1); it tells us that the discourse speaks of two entities, one a farmer (*x*), one a donkey (*y*), and that, according to the discourse, the farmer owns the donkey.

(K1)

$$\begin{array}{|c|}\hline x \quad y \\ \hline \text{farmer}(x) \\ \text{donkey}(y) \\ \text{owns}(x,y) \\ \hline \end{array}$$

While Kamp develops a top-down approach to DRS construction, Wada and Asher (1986), Zeevat (1989) and Asher (1993) develop a construction procedure whereby each lexical element contributes some sort of DR-theoretic structure. These then combine following the syntactic structure of the sentence to produce the DRS for the sentence in a manner that is in the tradition of compositional Montague semantics. Noun phrases like ‘a farmer’ or ‘a donkey’ introduce a novel discourse referent into the DRS as well as a condition on that discourse referent; verbs introduce conditions on the discourse referents introduced by the noun phrases that are their syntactic arguments. Anaphoric pronouns, following the original Kamp treatment of DRT (see Kamp 1981), are treated as bound variables; they introduce new occurrences of discourse referents that have been introduced by noun phrases in antecedent discourse.

The construction of a DRS for a discourse proceeds incrementally, sentence by sentence, exploiting the syntactic parse of each sentence, and building a DRS for the discourse as a whole from DRSs for each constituent sentence. More precisely, if $\mathbf{K}_m$ is the DRS derived from the first *m* sentences and K_{m+1} is the DRS derived from the $m+1$st sentence, then the DRS for the discourse of $m+1$ sentences is just the DRS that combines the universes of $\mathbf{K_m}$ and K_{m+1} and their condition sets:

$$\langle(\mathrm{U}_{\mathbf{K}_m} \cup \mathrm{U}_{K_{m+1}}),\ (\mathrm{Con}_{\mathbf{K}_m} \cup \mathrm{Con}_{K_{m+1}})\rangle$$

So in constructing a DRS for (2), for instance, we would add to the DRS in (K1) simply the condition ‘ $\text{beats}(x,\ y)$’ to get a correct representation for (2):

(K2)

$$\begin{array}{|c|}\hline x \quad y \\ \hline \text{farmer}(x) \\ \text{donkey}(y) \\ \text{owns}(x,y) \\ \text{beats}(x,y) \\ \hline \end{array}$$

Let us now briefly look at DRT’s treatment of sentence (3) given in §1. We use the same analysis of indefinite noun phrases and pronouns as introduced above, together with construction procedures for conditionals and universal quantifiers (such as the quantifier expressed by ‘every farmer’). Both constructions introduce what is known as a ‘complex condition’ in a DRS; complex conditions consist of DRSs as arguments to some operator. For instance, ‘if…, then… ’ introduces the following relation on DRSs, where *P* and *Q* are variables for DRSs.

$$\lambda P \lambda Q \quad \begin{array}{|c|}\hline \\ \hline P \quad \Rightarrow \quad Q \\ \hline \end{array}$$

So when we have two clauses linked by ‘if…, then… ’, the first clause gives us a DRS bound by *P*, while the second gives us a DRS which replaces *Q*. For example, ‘if Mary is happy, then Fred is happy’ would yield the following DRS:

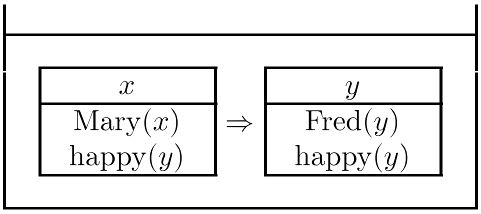

If we choose the right discourse referents for the pronouns ‘he’ and ‘it’, we get the right DR-theoretic translation for (3) in a uniform way from the treatment of the indefinite noun phrase and pronouns:

(K3)

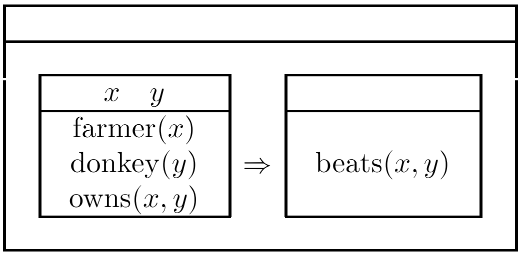

Several forms of complex conditions have been proposed in DRT to handle various connectives and quantifiers. Most prevalent are those of the form $K \Rightarrow K'$ used to represent sentences that contain a conditional or universal quantifier, and those of the form $\neg K$ which result from processing sentences in which a verbal or sentence-level negation occurs.

In order to understand what these complex conditions mean we must look at the second component of DRT, the correctness definition. The correctness definition tells us what conditions must obtain in order for a DRS, thought of as a partial model, to be properly embedded in a Tarskian model of the sort familiar from first-order logic (see Model theory). If the DRS for a discourse can be properly embedded in a model *M*, then we say that the discourse is true in *M*. A ‘DRS model’ is a pair <*D*, *I*> where *D* is a non-empty set (a domain of individuals) and *I* is a function that assigns *n*-tuples from D^n to atomic *n*-ary conditions of DRSs. Atomic conditions are those conditions that are derived from natural-language nouns, verbs and some adjectives - the sort that are contained in the DRSs for (1) and (2). A DRS *K* is properly embedded in a DRS model *M* if and only if there is a function *f* from the universe of *K* into the universe of *M* such that all the conditions of *K* are satisfied in *M*. What we have to do now is to define satisfaction of a condition in *K*. Intuitively the idea is simple: for a condition in the DRS to be satisfied in a model, there must be a mapping or embedding of the universe of *K* in the universe of *M* such that the objects $f(x_1),\ldots, f(x_n)$ bear the relation that the DRS condition predicates of $x_1,\ldots, x_n$. For an atomic DRS condition of the form $\Psi(\alpha_1, \ldots, \alpha_n)$, where $\alpha_1, \ldots, \alpha_n$ are discourse referents, $\Psi(\alpha_1, \ldots, \alpha_n)$ is satisfied in a DRS model *M* relative to an embedding function *f* if and only if $\langle f(\alpha_1), \ldots, f(\alpha_n)\rangle$ is an element of the interpretation of Ψ in *M*. A complex condition of the form $K \Rightarrow K'$ is satisfied in a DRS model *M* relative to an embedding function *f* if and only if for every function *g* that extends *f* to a proper embedding of *K* there is an extension *h* of *g* that is a proper embedding of K'. A complex condition of the form $\neg K$ is satisfied in a DRS model *M* relative to an embedding function *f* if and only if there is no function *g* that extends *f* to a proper embedding of *K*. Intuitively, we see how the interpretation of these DRS conditions captures the definition of a material conditional, a universal quantifier and negation. Technically we have a simultaneous recursive definition of proper embedding of a DRS in a model and of the satisfaction of a DRS condition with respect to a model. (For details see Kamp and Reyle 1993.)

In applying this definition of truth to (K3), we get the intuitively right truth-conditions for the sentence: for every pair of objects such that one is a farmer and the other is a donkey and the farmer owns the donkey, the farmer beats the donkey. These truth-conditions are the same as those for the universally quantified sentence, ‘every farmer who owns a donkey beats it’. While some problems in the interpretation of pronouns have been uncovered in further studies on anaphora in a dynamic semantic setting, the dynamic semantic approach represents a considerable advance over Montague Grammar’s approach to the interpretation of pronouns in discourse.

DRT and temporal anaphora. Following Davidson (1967), DRT takes verbs to introduce conditions on

eventualities (events or states) (see Davidson, D. §2). DRT postulates a complex algorithm that serves to connect by means of temporal relations the eventuality introduced by the main verb of a sentence with the eventualities already present in the DRS built up from previous discourse. In declarative form, we formulate the results of the algorithm as follows. If α and β are two sentences that follow one after another in the discourse and both introduce eventualities with a verb in the *passé simple*, we have (where $<$ is the 'earlier than' relation):

$$\mathrm{PS}(\alpha)\ \&\ \mathrm{PS}(\beta) \rightarrow \mathrm{e}_\alpha < \mathrm{e}_\beta$$

If past tense also implies that the event described by the verb occurs prior to the speech time, this rule in effect predicts just the right temporal interpretation of discourse (4). The complex DRS construction procedure has been applied to basic tense forms of French and English (see Kamp and Reyle 1993 for an extensive discussion and bibliography).

Limitations of dynamic semantics. DRT makes an important contribution to our understanding of how the discourse context affects pronominal and temporal interpretation. But DRT does not do justice to the complex interaction of pragmatic and semantic factors in discourse, and this leads to incorrect predictions about the way anaphora, both temporal and pronominal, is treated in DRT.

A clear example of where DRT goes wrong is with its analysis of temporal anaphora. DRT is right to make the contributions of tense depend upon the discourse context. But it attempts to make the temporal structure of a text almost completely dependent on the tense forms used in the text. In most natural languages, however, the temporal structure of the events introduced in a text is underdetermined by the sequence of tense forms. In particular, the rule above, which is a consequence of the DRS construction procedure, is false for French - a point of which some of the earliest workers on tense in DRT were aware - or English. Consider the following examples (from Lascarides and Asher 1993):

(5a) John entered the room. Fred greeted him.
(5b) John fell. Fred pushed him.

These two discourses employ the same sequence of tense forms, yet they suggest different temporal structures. DR-theorists have been forced to revise the construction procedure and to abandon the view that the tense forms and the order of the sentences in a discourse alone determine temporal structure. This conclusion follows not only from an examination of the English simple past but also from a careful look at the data concerning the French *plus que parfait* (see Asher and Bras 1992) and the English pluperfect (Lascarides and Asher 1992).

One might ask, what in combination with tense sequences determines the temporal structure? One proposal is that a more developed view of discourse structure is needed. Originally suggested by the computer scientist and artificial intelligence (AI) researcher Jerry Hobbs (1985), this thesis has been worked out in the context of a formal discourse semantics by Lascarides and Asher (1993). This approach to tense is part of a more general view: discourse structure has systematic effects on discourse interpretation, which it is the task of discourse semantics to study. As we shall see, both temporal and pronominal anaphora require a more elaborate discourse semantics.

3 More elaborate discourse semantics

When interpreters read a text, they naturally understand it as divided into parts related in various ways. Some parts form a narrative, others furnish a background to a narrative or an explanation of some action, still others may elaborate other parts of the text. Sometimes these divisions and the relations between them are signalled by punctuation and the presence of certain words in the text. But often interpreters infer the appropriate divisions and discourse relations without these cues. This sort of structure is completely missing from the treatment of discourse in DRT, but it has been recognized as an important aspect of discourse in other disciplines - for example, literary theory and the natural language processing (NLP) area of AI (see Artificial intelligence). In the mid-1970s, computer scientists like Roger Shank claimed text-understanding systems and text-information retrieval systems to be one of the central goals of NLP research in AI. This early interest led to a substantial amount of research on discourse structure in the mid to late 1980s (Polanyi 1988; Hobbs 1985; Grosz and Sidner 1985). This work has concentrated on the discourse structure of written texts, though analyses of highly structured, task-oriented dialogues have yielded important insights. Many divergent positions have been taken, but several uncontroversial conclusions have been drawn from these studies: (1) discourse is structured and can be represented by a set of

objects (propositions or bits of text) related by discourse relations; (2) this structure is hierarchical and recursive. For discourse semantics, we must integrate the view of discourse structure elaborated by these authors with a theory of meaning. This connection between semantics and discourse structure, however, is lacking in most of the discourse theories in AI.

A richer notion of discourse structure is incorporated into the framework of dynamic semantics in Asher (1993), resulting in a nontrivial extension of DRT known as ‘segmented Discourse Representation Theory’ or SDRT. SDRT is one way to integrate discourse structure for texts and semantics. Another approach can be found in the works of Scha and his colleagues, but it exploits a different version of dynamic semantics.

According to SDRT, every discourse is analysed into a representational structure called a ‘segmented DRS’ or SDRS, and this is in turn embedded into a complete model that furnishes truth-conditions for the discourse. The model theory for SDRSs is somewhat involved, because the models must allow for quantification over and reference to dynamic propositions (for details see Asher 1993 or Asher 1996). An SDRS, like a DRS, is a pair of sets; but unlike a DRS the elements of the first set are either DRSs or SDRSs, while the second pair consists of a set of conditions in which discourse relations are predicated of elements in the first set. In comparing an SDRS for a narrative text with a DRS for the same text, we would see that the DRS is segmented into elements of the first set in the SDRS with discourse relations holding between these elements. These elements or ‘constituents’ of an SDRS serve several useful functions. First of all, they constitute referents for anaphoric pronouns (like ‘this’ in the example below) whose antecedents are facts or propositions:

(6) One plaintiff was passed over for promotion. Another didn’t get a pay increase for five years. A third received a lower wage than men doing the same work. But the jury didn’t believe this (any of this).

In the example above, we appear to have two possible antecedents (depending to some degree on how this is stressed): the proposition formed by the discourse as a whole and the proposition given by the last sentence. In SDRT these two anaphoric possibilities and only these two anaphoric possibilities are predicted; they result from two different, acceptable places to which the new information contained in the last sentence can be attached to the SDRS built up from the previous three sentences. Below we have a pictorial representation of the two SDRSs that could be constructed for (6). $K_1,\ldots, K_4$ represent the DRSs derived from the four sentences of the discourse; K_0' is a topic that results from the SDRS construction and the constraints on Narration (every Narration must have a nontrivial topic); the shadowed letters represent complex constituents or constituents that are SDRSs. The lines represent membership, while $\Downarrow$ represents a particular type of discourse relation (like Elaboration) that introduces a hierarchical structure into an SDRS. Narr stands for Narration, Elab for Elaboration, and Contr for Contrast. Notice the two different positions of attachment for K_4.

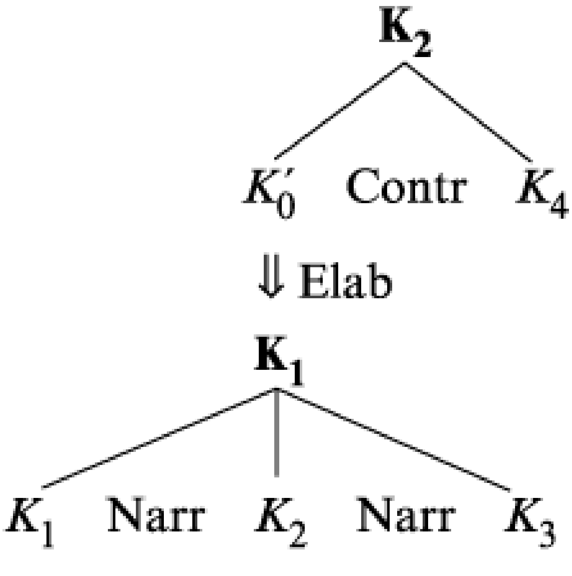

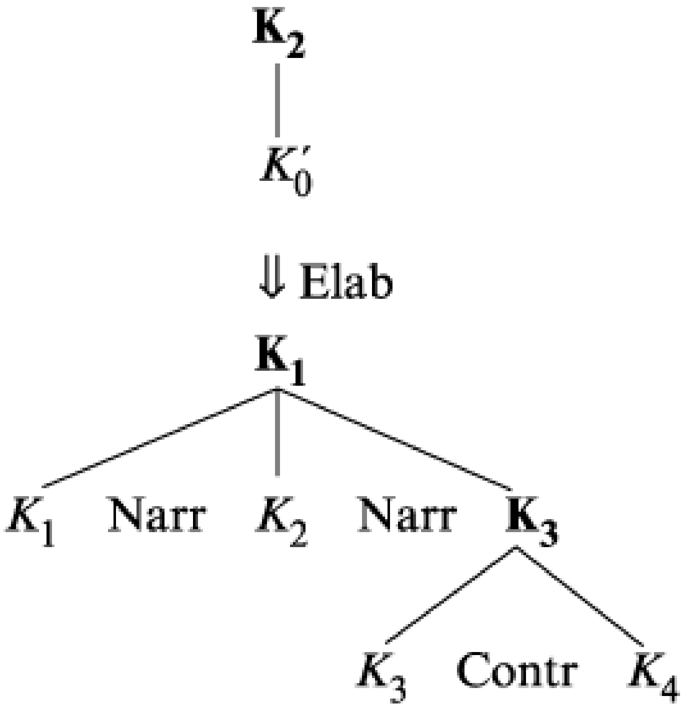

The crucial task for a discourse semantics like SDRT is to specify a procedure that will enable us to construct a discourse representation. In DRT the discourse representation is built up from the syntactic structure and a translation procedure for syntactic elements in the parse tree. SDRT exploits the DRS construction procedure to build DRSs for clauses, and it also makes use of the constraints DRT places on anaphora resolution. But in SDRT we need more information to be able to deduce where to attach new information in the antecedently built-up representation and by means of which discourse relation to attach it. The constraints on appropriate attachment sites are defined configurationally, following much of the work on discourse structure in AI. But the difficult part is to infer the appropriate discourse relation. Sometimes certain cue words or phrases indicate the presence of particular discourse relations, but those discourse relations may be present even when there are no cues. For instance, consider the following pairs:

(7a) John went to jail. He embezzled funds from the pension account.
(7b) John went to jail, because he embezzled funds from the pension account.

(8a) Mary went to school. She met with her students. She went home.
(8b) Mary went to school. Then she met with her students. Afterwards she went home.

Discourses (7a) and (7b) on the one hand and (8a) and (8b) on the other, lend themselves to the same interpretation. In the examples in (7), the second clause or sentence gives us the explanation for what happened in the first; while in (8), we read the clauses in each as narrating a sequence of events that are in the order that they are introduced in the text. Yet only in (7b) and (8b) are these interpretations fully determined by the semantic contributions of special ‘clue’ words like ‘because’ or the temporal adverbials in (8b). In (7a) and (8a), we infer the presence of an Explanation relation or a Narration relation based on pragmatic principles and the contents of the constituents the discourse relations tie together in the SDRSs for these discourses. However, this inference is not deductively valid; within a richer discourse context, we can cancel, for instance, the inference that the second sentence in (7a) offers an explanation of what happened in the first:

(7c) John went to jail. He embezzled funds from the pension account. But that is not why he went to jail.

But it also seems that the explanation relation is undeniably there in (7a). Thus, we must have a means of inferring discourse relations by default; the default or ‘non-monotonic’ inference of a discourse relation *R* can be overridden in cases where we have information in the context that is inconsistent with our conclusion that *R* holds. SDRT uses facts about the discourse context, the contents of the constituents including the lexical contents of particular words, world knowledge and pragmatic principles, to infer discourse relations non-monotonically. An attempt to code all

of this knowledge explicitly for certain simple examples is given in Lascarides and Asher (1993).

Once we have inferred that some discourse relation *R* relates new information to some available attachment point in the SDRS constructed from the preceding discourse, we can update this SDRS with the new information. But this updating does not consist in simple addition. It also requires in many cases the revision of the contents of the constituents in the SDRS. It is in the process of constituent revision that anaphora resolution and the determination of temporal structure are accomplished in SDRT. For instance, consider again the discourse in (4). The *passé simple* of the verbs in this text all introduce the simple DRS condition saying that the eventuality described by the verb occurred prior to the speech point. There is no attempt in the semantics to link up the eventuality with the temporal structure built up antecedently, as there is in DRT. But now the underlying logical inference mechanisms of SDRT discussed above lead us to conclude that Narration relates the constituents α and β derived from the first two sentences and the constituents β and χ derived from the second and third sentence. According to SDRT, Narration has certain temporal effects: $\mathrm{Narration}(\alpha, \beta) \rightarrow \mathrm{e}_\alpha < \mathrm{e}_\beta$. So once we have attached the three constituents in (4) with Narration, we immediately conclude that the text gives us a sequence of events that follows the order in which these events are introduced in the text. The effect of Narration is one way to formulate precisely part of Grice's maxim 'be orderly'. The order of presentation of the events in the text reflects their temporal order (see Implicature §4). But the pragmatic principles appealed to in the formalization of Lascarides and Asher (1993) and used to infer that Narration holds between these constituents are a good deal more specific than the general Gricean maxims of conversation, and they are expressed in logic.

In (5b) additional information contained in the two clauses leads us to defeat the inference to Narration and to conclude that a different discourse relation - Explanation - holds between the two constituents of the SDRS. Explanation has different temporal consequences from Narration: $\mathrm{Explanation}(\alpha, \beta) \rightarrow \neg(\mathrm{e}_\alpha < \mathrm{e}_\beta)$ and, where '$>$' is a weak conditional used to represent default patterns, $\mathrm{Explanation}(\alpha, \beta) > (\mathrm{e}_\beta < \mathrm{e}_\alpha)$. So we conclude that in (5b) the pushing came *before* the falling, as intuitions would dictate. Thus, while DRT attempted to derive the temporal structure of a text from the semantics of the sequence of tenses in a discourse alone, SDRT exploits not only the compositional semantics but also many other sources of information that contribute to discourse structure to determine temporal structure.

4 Conclusions

Because of the rapidity with which the field of discourse semantics has evolved, it is hard to say exactly where the field will develop or even whether it will continue as a field. It is quite unclear whether DRT or SDRT will remain active research areas. SDRT has been applied not only to the central discourse semantic concerns of temporal and pronominal anaphora, but it has also helped in analysing the spatiotemporal structure of texts (Asher *et al.* 1994), the role of discourse structure in lexical disambiguation (Asher and Lascarides 1995), and the interactions between discourse structure and speakers' and interpreters' cognitive states - in particular their beliefs and discourse goals (Asher and Lascarides 1994). To advance further, SDRT and other formalisms for discourse semantics require a better understanding of the organization of the lexicon and of the role of general world knowledge. As discourse semantics moves to give a formally precise analysis (capitalizing on empirical studies) of discourse interpretation in dialogue, other factors like the interaction between semantics and speech acts will have to be taken into account (see Speech acts).

See also: Pragmatics

NICHOLAS ASHER

References and further reading

Asher, N. (1993) *Reference to Abstract Objects in Discourse*, Dordrecht: Kluwer.(This book (referred to in §3) introduces SDRT to take account of rhetorical relations and their effects on truth-conditional content. It applies SDRT to propositional anaphora and VP ellipsis.)

Asher, N. (1996) 'Mathematical Treatments of Discourse Contexts', *Proceedings of the Tenth Amsterdam Conference on Formal Semantics*, Amsterdam: ILLC Publications, vol. 1, 21-40.(This article gives a revised, improved version of SDRT. It is a much shorter introduction to SDRT than Asher 1993.)

Asher, N., Aurnague, M., Bras, M., Sablayrolles, P. and Vieu, L. (1994) 'Computing the Spatiotemporal Structure of Discourse', in H. Hunt, R. Muskens and G. Rentier (eds) *Proceedings of the First International Workshop on Computational Semantics*, Tilburg: Tilburg University Publications, 11-20.(This article uses

SDRT to model the spatio-temporal structure of discourse.)

Asher, N. and Bras, M. (1992) 'The Temporal Structure of French Texts within Segmented Discourse Representation Theory', in M. Aurnague, A. Borillo and M. Bras (eds) *Semantics of Time, Space, Movement and Spatio-Temporal Reasoning*, Working Papers of the 4th International Workshop on Time and Space, IRIT Working Papers, Toulouse: Université Paul Sabatier, 203-18.

Asher, N. and Lascarides, A. (1994) 'Intentions and Information in Discourse', *Annual Proceedings of the American Computational Linguistics Meetings*, 34-41.(This paper addresses the information flow between discourse structure and cognitive states in a couple of examples. Referred to in §4.)

Asher, N. and Lascarides, A. (1995) 'Lexical Disambiguation in a Discourse Context', *Journal of Semantics* 12 (1): 69-108.(This paper applies SDRT to lexical disambiguation and shows how discourse structure can affect this task. Referred to in §4.)

Davidson, D. (1967) 'The Logical Form of Action Sentences', in N. Rescher (ed.) *The Logic of Decision and Action*, Pittsburgh, PA: The University Press, 81-95.(In this article, Davidson proposes that verbs are predicates of events.)

Geach, P.T. (1963) *Reference and Generality*, Ithaca, NY: Cornell University Press.(This book is a collection of Geach's influential articles, in which among other things the puzzle about donkey sentences (referred to in §1) is introduced.)

Grice, H.P. (1967) *Logic and Conversation. William James Lectures*, repr. in *Studies in the Ways of Words*, Cambridge, MA: Harvard University Press, 1989.(The introduction of conversational 'maxims', which are referred to in §3.)

Groenendijk, J. and Stokhof, M. (1990) 'Dynamic Predicate Logic', in H. Kamp (ed.) *Dyana Deliverable*, 2.5a. (This paper introduces an alternative approach to dynamic semantics from DRT as presented in Kamp 1981.)

Grosz, B. and Sidner, C. (1985) 'The Structure of Discourse Structure', Palo Alto: SRI, SRI Technical Note 369. (This paper gives a very influential view of discourse structure and argues for a parallel between discourse structure and the intentions of the speaker. Referred to in §3.)

Heim, I. (1982) *The Semantics of Indefinite and Definite Noun Phrases*, Amherst, MA: University of Massachusetts Press.(Referred to in §2; Heim's Ph.D. thesis offers an alternative way of developing dynamic semantics to Kamp 1981 and is close to the views presented in Groenendijk and Stokhof 1990.)

Hobbs, J. (1985) 'On the Coherence and Structure of Discourse', Stanford, CA: Center for the Study of Language and Information, report no. CSLI-85-37.(This paper introduces and argues for the idea that various anaphoric phenomena may be dependent upon discourse structure. Referred to in §§2-3.)

Kamp, H. (1981) 'A Theory of Truth and Semantic Representation', in J. Groenendijk, T. Janssen and M. Stokhof (eds) *Formal Methods in the Study of Language*, Amsterdam: Mathematisch Centrum Tracts, 277-322; repr. in J. Groenendijk *et al.* (eds) *Truth, Interpretation and Information*, Dordrecht: Foris, 1984, 1-41.(This article (referred to in §2) contains the original presentation of DRT. It is difficult to read but is the principal source for many of the developments in dynamic semantics.)

Kamp, H. and Reyle, U. (1993) *From Discourse to Logic; Introduction to Model-Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Dordrecht: Kluwer.(This book is an introductory textbook to DRT that extends considerably the semantic coverage of the theory from that given in Kamp 1981. Referred to in §2.)

Lascarides, A. and Asher, N. (1992) 'The Pluperfect in Narrative Discourse', in M. Aurnague, A. Borillo, M. Borillo and M. Bras (eds) *Semantics of Time, Space, Movement and Spatio-Temporal Reasoning*, Working Papers of the 4th International Workshop on Time and Space, IRIT Working Papers, Toulouse: Université Paul Sabatier, 183-217.(A condensed version of this paper called 'A Semantics and Pragmatics of the Pluperfect' appears on pages 250-9 of *Proceedings of the 6th Conference of the European Chapter of Computational Linguistics*, published in Utrecht in 1993. These papers discuss the pluperfect in narrative discourse from the perspective of SDRT.)

Lascarides, A. and Asher, N. (1993) 'Temporal Interpretation, Discourse Relations, and Commonsense Entailment', *Linguistics and Philosophy* 16: 437-93.(This paper criticizes the analysis of tense within DRT and proposes a new simplified account using the SDRT framework. Referred to in §§2 and 3.)

Partee, B. (1973) 'Some Structural Analogies between Tenses and Pronouns in English', *Journal of Philosophy* 70: 601-9.(In this article Partee argues that tenses like the simple past exhibit a contextual sensitivity in interpretation similar to that of anaphoric pronouns.)

Polanyi, L. (1988) 'A Formal Model of the Structure of Discourse', *Pragmatics* 1-88.(This article contains another influential approach to discourse structure, with powerful treatments of interruptions and discourse repairs. Referred to in §3.)

Prust, J. and Scha, R. (1990) 'A Discourse Perspective on Verb Phrase Anaphora', in M. Stokhof and L. Torenvliet (eds) *Proceedings of the Seventh Amsterdam Colloquium*, Amsterdam: ITLI Publications, 451-74. (This article develops the theory of discourse parallelism in detail and points to the use of dynamic semantics as developed by Groendijk and Stokhof as the interpretive component.)

Scha, R. and Polanyi, L. (1988) 'An Augmented Context Free Grammar for Discourse', *Proceedings of COLING*, Budapest.(This article proposes a 'discourse syntax', in which trees model discourse structures, together with an interpretation of the leaves of those trees. The interpretation uses Montague Grammar, however, and is not dynamic.)

Wada, H. and Asher, N. (1986) 'BLDRS: A Prolog Implementation of LFG and DR Theory', *Proceedings of the 11th International Conference on Computational Linguistics*, Bonn, 540-5.(This was one early implementation of DRT and suggested a compositional treatment of dynamic semantics.)

Zeevat, H. (1989) 'A Compositional Approach to Discourse Representations', *Linguistics and Philosophy* 12: 95-131.(This paper gives a detailed treatment of how to build up DRSs in a compositional way, similar to the construction of formulas representing meanings of expressions in Montague Grammar. Referred to in §2.)

Discovery, logic of

Bacon, Descartes, Newton and other makers of the Scientific Revolution claimed to have found and even used powerful logics or methods of discovery, step-by-step procedures for systematically generating new truths in mathematics and the natural sciences. Method of discovery was also the prime method of justification: generation by correct method was something akin to logical derivation and thus the strongest justification a claim could have. The 'logic' of these methods was deductive, inductive or both. By the mid-nineteenth century, logic of discovery was yielding to the more flexible and theory-tolerant method of hypothesis as the 'official' method of science. In the twentieth century, Karl Popper and most logical positivists completed the methodological reversal from generativism to consequentialism by setting their hypothetico-deductive method against logic of discovery. What is epistemologically important, they said, is not how new claims are generated but how they fare in empirical tests of their predictive consequences. They demoted discovery to the status of historical anecdote and psychological process. Since the late 1950s, however, there has been a revival of interest in methodology of discovery on two fronts - logical and historical. An earlier explosion of work in symbolic logic had led to automata theory, computers, and then artificial intelligence. Meanwhile, a maturing history of science was furnishing information on science as a process, on how historical actors and communities actually discovered or constructed their claims and practices. Now, in the 1980s and 1990s, liberal epistemologists once again admit discovery as a legitimate topic for philosophy of science. Yet attempts to both naturalize and to socialize inquiry pose new challenges to the possibility of logics of discovery. Its strong associations with 'the' method of science makes logic of discovery a target of postmodernist attack, but a more flexible construal is defensible.

1 Logic and method

A logic of discovery in the narrowest sense would be a valid reasoning procedure that leads to new knowledge. More broadly, 'logic of discovery' includes routine methods and research techniques that contain inductive or heuristic steps and sometimes produce new knowledge claims that need further justification. A heuristic procedure is a shortcut that compromises the reliability of an algorithm for greater speed or efficiency. In the broadest sense of discovery as an achievement, 'logic of discovery' or the discovery process includes the entire research effort from the first tentative posing of a problem to the 'final' testing and accepting of a solution as adequate.

Does deductive logic constitute a logic of discovery in the narrow sense? The usual objection - that deductive argument is not ampliative - confuses logical with epistemic novelty and 'amplification'. True, the conclusion contains no more logical content than the premises collectively, but it may contain more epistemic content, relative to the current state of knowledge. Strictly deductive arguments in mathematics, logic and the sciences have led to astounding (epistemically novel) conclusions. Since deductive arguments may have empirical premises, and hence empirical conclusions, this reply also blunts the related objection that deductive methodology is a priori and so cannot generate new empirical claims. Deductive reasoning from what we already 'know' can yield empirical conclusions that are epistemically novel, even revolutionary; for example, the Lorentz transformation is derivable from a mathematically generalized form of the Galilean transformation plus the 'empirical fact' that the speed of light is constant (see Relativity theory, philosophical significance of §1).

A more serious objection to deductive logic of discovery is that it is not a method or procedure and hence provides no guidance to discovery. It is a logic of justification that tells us whether a proof is valid, but it does not show us how to find such proofs.

Relaxing deductive strictures to allow inductive methods does yield low-level logics of discovery. An example is Reichenbach's inductive method, based on repeated self-correcting applications of his 'straight rule' (see Reichenbach, H. §3). Factor analysis and other probabilistic and statistical methods furnish inductive discovery procedures, but we cannot expect these methods alone to produce novel, deep-structural theories.

Another way in which to 'relax' deductive logic into a useful method is to add heuristic rules, for example, rules that instruct us to search first for an answer here or to try A, B and C, in that order. Heuristic rules in the form of 'production rules' (if the situation is X, then do Y) are central to the work of Allen Newell, Herbert Simon and others in artificial intelligence (AI) (see Artificial intelligence §2). Their motivating idea is that discovery is problem-solving behaviour; hence, what we already know about problem solving can be transferred to

methodology of discovery. And we know that problem solving typically involves heuristic search through spaces or ’trees’ of possible solutions. They claim that their BACON and KEKADA programs have rediscovered laws of Kepler, Black, Ohm, Krebs and others (see Langley *et al.* 1987).

While Simon’s group prefers to employ very general heuristics, others employ heuristics that are highly content- and context-specific. In effect they answer the above a priori objection by employing heuristics laden with theoretical and empirical content specific to the discipline or even the particular problem at hand. In AI this corresponds to the move from general problem solvers to knowledge-based or ’expert’ and case-based systems. This move is motivated by the quest for greater problem-solving power. Roughly speaking, the more powerful the program, the more specific its sphere of application.

Developments in AI represent the ’logical’ wing of methodology of discovery. The ‘historical’ wing was inspired by a post-war ’internalist’ history of science, which showed that typical discoveries are discursive, structured in time. Accordingly, historical philosophers reject as overly romantic the view that discoveries are momentary, irrational inspirations. But finding no uniform logic of discovery, some defend the epistemic interest of discovery by retreating from logic to claiming the rationality of discovery and by contending that mode of discovery can be relevant to justification and hence to epistemology. Some writers interpose a third stage - prior appraisal, preliminary evaluation or heuristic appraisal - between original generation of ideas and ’final’ justification. Few historicists fully embrace AI accounts of discovery. Historicists doubt whether we do know enough about problem solving in general to capture the richness of historical investigations in the form of AI programs.

AI advocates retort that the historical approach is too vague to explain much. An adequate explanation of historical discoveries must be ’computational’: it must furnish a program capable of actually re-making the discovery in question, given the historically relevant problems and data. On strong versions of this view, an adequate discovery explanation posits something akin to a constructive proof already available to the historical actors and used by them to make that discovery! Historicists reply that this demand puts the cart before the horse and imputes foreknowledge to the actors; that, typically, routinized problem-solving procedures are later, refined products or parts of the total discovery rather than antecedently available explainers of it. Historicists complain that the AI programs employ cleaned-up text-book versions of Kepler’s and others’ problems and data and thus illegitimately reverse the order of knowing by imposing present conceptions on the past. They whiggishly sacrifice historical reality to AI production values.

2 Is there a logic of discovery?

The question is ambiguous and has been understood in at least three ways.

(1) Does there exist, in the abstract mathematical sense, one or more logics (routine methods) of discovery?
(2) Do/did scientists actually possess one or more logics of discovery?
(3) If so, does their knowledge of methods of discovery explain any historical discoveries?

To those who engage in highly formal studies of type (1), for example machine learning theory, historical issues are simply irrelevant. For them the question of existence is mathematical-ontological, not historical. However, a formal existence proof does not in itself produce a method actually usable on the frontier of research. Accordingly, most methodologists have been more interested in (2) and (3). (2) puts the question in epistemological rather than ontological terms, while (3) concerns historical and historiographic application.

Is there a logic or method of discovery in senses (2) or (3), a systematic and routine way of attaining new knowledge? Can the Scientific Revolution be partially explained in terms of new methods of discovery? Most methodologists agree that there is no single method of discovery, let alone a single ‘method of science’, understood as something that gives precise direction to inquiry. However, in a weaker and more local sense, we surely do possess numerous routinized problem-solving methods, procedures that would immensely impress a Descartes or a Newton. While these procedures do not, all by themselves, generate deep new theories, they do help us organize in fruitful ways both extant knowledge and our knowledge-productive powers; and the semi-automated (computer-aided) modelling techniques available in many sciences can guide the construction of serious, deep theory candidates (see Models).

Actually, the epistemic importance of discovery, here understood as initial generation of a claim, would seem to be

secure, whether or not there is a method or even a rationality of discovery in some pre-specified sense. For if we are ever to find true, correct or adequate claims and techniques, they must first be generated in some way, whether by a rational, methodical process or not. A correct claim has no chance of being selected (justified) unless a discovery process first puts it on the table for further consideration. Thus, manner of generation is relevant to epistemology. But this point must not be overrated, for the epistemic importance of discovery in this broad sense does not entail the existence of a method or logic, or even the rationality, of discovery.

The principal objection to there being rules for discovery is that ’creative routine’ is a contradiction-in-terms. Romantic poets long ago raised this objection: originality, by definition, cannot be a product of rules. (They do have a point, but do they not also confuse epistemic and logical novelty?) More recently, from a different quarter, evolutionary epistemologists such as Karl Popper (1959) and Donald Campbell (1960) have contended that ’logic of discovery’ is an oxymoron. They in effect raise the a priori objection of §1 from the opposite side: a genuine logic of discovery is impossible, for it would require prescience or clairvoyance - knowing now what we shall only find out later. On the frontier of knowledge, says Campbell, we can only proceed blindly. At the absolute frontier of research, there can be no method to guide us. All genuine creativity in the arts and sciences amounts to blind variation plus selective retention $(\mathrm{BV}+\mathrm{SR})$, which is basically the mechanism of biological evolution. This claim threatens both AI and historical methodologies of discovery.

Methodologists of discovery respond that actual research, even on the frontier, looks more directed than this; hence, there remains some scope for method, with its concern for economy of research. From the perspective of methodology as a sort of managerial science of our limited epistemic resources, $\mathrm{BV}+\mathrm{SR}$ is the most wasteful form of production there is. It introduces no ’quality control at the source’ but instead tests and discards almost everything at the justificatory ’selection’ stage. Reducing discovery to blind luck is precisely what Bacon and Descartes wished to avoid. And, in fact, historical investigators have nearly always been guided by powerful constraints at the generation stage. Evolutionary epistemologists reply that guiding heuristics are often available but that we must consider them, in turn, as ultimately products of previous hierarchies of $\mathrm{BV}+\mathrm{SR}$. Thus it would be wrong to see Campbell and Simon as being totally opposed.

3 Some problems and prospects

Many critics dismiss logic of discovery as a historically dated subject; yet, understood as a broad topic area rather than a specific method, it engages several intellectual developments. We can conceive logic of discovery as the study of knowledge amplification in all its forms; whence it embraces all the ways in which scientific claims are constructed. We need not cast the issues in terms of false dilemmas such as history versus logic, social constructivism versus realism and logic versus rhetoric. On the contrary, discovery logics (routinized modes of problem solving, model construction, and data analysis) are themselves discovered or constructed and maintained as succeeding generations reconstruct and refine the work of their predecessors. And the rhetoric of inquiry (including problem representation, modelling, learning by example, and point-of-view in problem solving) is crucial to creative cognition. Rhetoric helps both to familiarize the unfamiliar and to de-familiarize the familiar. Kuhn (1962) and current AI and cognitive psychological work on learning by example, by exemplars rather than rules, suggests that the rhetoric of inquiry is, in part, a sort of fuzzy logic of discovery. Kuhnian exemplars constitute a system of rhetorical places or concrete models for ongoing work at the frontier. It is noteworthy that nearly every philosophical account of learning, theory structure, theory reduction, modelling, and so on, ultimately depends on analogy, metaphor, similarity relations, and other rhetorical tropes.

Heuristic appraisal, the evaluation of the promise or fertility of a problem, theory, technique or research proposal is a ’discovery’ topic that deserves more study. It is immensely important, as scientists and funding committees must base nearly all of their decisions on heuristic appraisal. But how is it possible to go beyond an item’s ‘track record’ to evaluate its future promise? Here discovery intersects the central epistemological problems of futurology: what can we know about possible futures, and how?

A related issue, the more important insofar as successful logics of discovery are local and content-specific, is to what extent these methods can be applied in other areas. This is one form of the problem of transfer of technology and connects with problems of standardization. There may be no scientific method that is at once universal and deeply instructive, but scientists nonetheless want to generalize the more modest results that they do have, to make as much of the world as possible scientifically intelligible. Far from presupposing the existence of a fixed scientific

method, logics of discovery are themselves products of discovery/construction. Even the logical and rhetorical discovery and constitution of new sciences falls under discovery, broadly understood.

See also: Hanson, N.R.; Scientific method

THOMAS NICKLES

References and further reading

Baird, D. (1992) *Inductive Logic: Probability and Statistics*, Englewood Cliffs, NJ: Prentice Hall.(Accessible text with a chapter on factor analysis and discovery.)

Bechtel, W. and Richardson, R. (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton, NJ: Princeton University Press.(A study of research strategies in the biological and behavioural sciences.)

Campbell, D.T. (1960) 'Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes', *Psychological Review* 67: 380-400.(A classic, readable presentation of an evolutionary epistemological account of creativity.)

Darden, L. (1991) *Theory Change in Science: Strategies from Mendelian Genetics*, Oxford: Oxford University Press.(Theory construction in twentieth-century biology, informed by AI.)

Kelly, K. (1996) *The Logic of Reliable Inquiry*, Oxford: Oxford University Press.(An excellent, technical treatment of the formal learning theory approach to discovery.)

Kantorovich, A. (1993) *Scientific Discovery: Logic and Tinkering*, Buffalo, NY: SUNY Press.(An attempt to work out a naturalistic account in detail.)

Kleiner, S. (1993) *Scientific Discovery: A Theory of the Rationality of Science*, Dordrecht: Kluwer.(Logic of questions applied to Darwin's work.)

Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press; 2nd edn, 1970.(The contemporary classic that marks the transition from logical positivism and Popper to historical philosophy, sociology and psychology of science.)

Langley, P., Simon, H.A., Bradshaw, G. and Zytkow, J. (1987) *Scientific Discovery*, Cambridge, MA: MIT Press.(An accessible presentation of largely data-driven AI discovery programs such as BACON.)

Laudan, L. (1981) *Science and Hypothesis*, Dordrecht: Reidel.(Essays on the emergence of 'the' method of hypothesis, in opposition to methodologies of discovery.)

Newell, A. and Simon, H.A. (1972) *Human Problem Solving*, Englewood Cliffs, NJ: Prentice Hall.(A classical, reconstructed summary of their work since the 1950s.)

Nickles, T. (1980) 'Scientific Discovery, Logic, and Rationality' and 'Scientific Discovery', in *Case Studies*, Dordrecht: Reidel.(Conference papers, mostly taking an historical approach; extensive introduction and references.)

Nickles, T. (1998) 'Discovery', in W. Newton-Smith (ed.) *A Companion to the Philosophy of Science*, Oxford: Blackwell.(Expands on several points in this entry.)

Popper, K.R. (1959) *The Logic of Scientific Discovery*, New York: Basic Books. (Popper defends his hypothetico-deductive method against inductivist and other methods of discovery.)

Shrager, J. and Langley, P. (eds) (1990) *Computational Models of Scientific Discovery and Theory Formation*, San Mateo, CA: Morgan Kaufmann.(A good sample of 1980s work in AI, including Kulkarni and Simon's KEKADA.)

Wimsatt, W. (1980) 'Reductionistic Research Strategies and their Biases in the Units of Selection Controversy', in T. Nickles (ed.) *Case Studies*, 213-59.(A biological case study of heuristics and their limits; combines insights of Simon and Campbell.)

Discrimination

A principle forbidding discrimination is widely used to criticize and prohibit actions and policies that disadvantage racial, ethnic and religious groups, women and homosexuals. Discriminatory actions often rely on unfavourable group stereotypes and the belief that members of certain groups are not worthy of equal treatment. A prohibition of discrimination applies to the distribution of important benefits such as education and jobs, and says that people are not to be awarded or denied such benefits on grounds of characteristics such as race, ethnicity, religion or gender. Attempts have been made to expand this principle to cover institutional *discrimination. Discrimination is morally wrong because its premise that one group is less worthy than another is insulting to its victims, because it harms its victims by reducing their self-esteem and opportunities, and because it is unfair.*

1 The anti-discrimination principle

Prejudice and ethnic hostility, and the discrimination and other forms of intolerance that they generate, are found in all parts of the world. Prejudice typically involves a negative attitude towards a group based on a belief that members of the group are undeserving of equal treatment because they are somehow inferior or morally deficient. Prejudice often relies on stereotypes - simplified and negatively slanted conceptions of the typical characteristics and activities of members of the group. Racism is one species of prejudice; other species include bigotry, sexism, extreme nationalism and homophobia.

The moral norm that makes discrimination wrong can be formulated as follows:

The anti-discrimination principle. When distributing educational opportunities and jobs [*list of items*], never exclude whole groups of persons or choose one person over another on grounds of race, ethnicity, religion, or race [*list of excluded characteristics*] unless the use of these characteristics in particular circumstances is demonstrably legitimate and important.

This anti-discrimination norm is presupposed by most condemnations of discrimination, and is found in most civil rights legislation (Brest 1976). It rejects ways of thinking that assume the inferiority or depravity of certain groups, and forbids some of the most important ways of acting and deliberating that flow from such assumptions. Because it is concerned with the grounds on which decisions are made (and thus with people's reasons and intentions), proving that discrimination has occurred is often difficult.

The anti-discrimination principle forbids those distributing certain benefits and burdens to 'choose … on the basis of …' certain excluded characteristics. This regulates the process of selection, but only indirectly regulates its outcomes. It does not forbid an employer to deny employment to a Muslim, if that is the outcome of a non-discriminatory selection process, but it blocks the employer from using an applicant's Muslim beliefs as a reason in deciding whether or not to employ the applicant. Formal or informal statements of qualifications cannot mention religion or other excluded characteristics. The anti-discrimination principle says how *not* to allocate certain important benefits, but it does not say how they should be allocated. It does not, for example, prescribe that these benefits should be allocated in accordance with competency or qualifications. Goldman (1979) has defended the additional principle that employment and selection for advanced education should always be on the basis of competence.

The formulation of the anti-discrimination principle given above is too narrow. The two bracketed clauses (the 'list of items' and the 'list of excluded characteristics') mark slots that need to be filled in more adequately. The list of items could be expanded by listing, along with educational opportunities and jobs, matters such as housing, public accommodation, legal rights, opportunities for political participation, public assistance and military service. Perhaps the underlying idea here is to focus on benefits and burdens that have a major impact on people's prospects, long-term welfare and self-respect, and whose distribution can be regulated without deeply invading people's liberties. The anti-discrimination principle does not apply to the distribution of all benefits and burdens. It does not forbid, for example, choosing someone who shares one's religion as a friend, or using gender as a factor in choosing a spouse.

The list of excluded characteristics could be expanded beyond race, ethnicity, religion and gender to include national origin, political beliefs, being a non-citizen, sexual orientation, income and age. It is sometimes suggested

that an anti-discrimination principle should exclude reliance on characteristics that are ‘irrelevant’ or ‘arbitrary’. What makes this plausible is the fact that one’s religious beliefs, for example, have little bearing on one’s qualifications to work in a construction company. Choosing people for construction work on the basis of their religion just seems irrational. But if the owners of a large construction company seek to employ only fellow Christians because they think they are more likely to be honest and hard-working, or because they want to create a certain religious atmosphere within the company, they cannot be faulted for making arbitrary choices or using irrelevant criteria. The problem is rather that these selection procedures are inappropriate for a large company in a diverse country because they are unfair to non-Christians.

2 Institutional discrimination

The anti-discrimination principle is narrow in its scope and the occurrence of discrimination is often hard to prove. Thus it is unlikely that widespread acceptance and legal enforcement of the principle will quickly stop discrimination or eliminate the consequences of past discrimination. Affirmative action programmes, which go beyond anti-discrimination, are intended to do these things more rapidly (see Affirmative action). But many have attempted to expand the scope of the anti-discrimination principle to cover *institutional* discrimination (Banton 1991; Ezorsky 1991). There are at least five sorts of phenomena that a norm addressing institutional discrimination might cover.

(1) *Unconscious discrimination*. To engage in discriminatory behaviour one does not have to be aware that one is doing so. One can discriminate by relying on stereotypes one is unaware of or by unconsciously following old customs. This sort of discrimination can be brought under the anti-discrimination principle by defining ‘choose … on grounds of …’ in a broad enough way to cover decisions that unconsciously rely on stereotypes and discriminatory customs.
(2) *Statistical discrimination*. Characteristics that are relevant to selection for a position often are statistically correlated to some degree with characteristics such as race, ethnicity and gender. An employer may refuse, for example, to employ a woman in a warehouse because they believe correctly that most women are unable to lift the heavy boxes stored in the warehouse. Although the likelihood that one can do a certain kind of job is generally an appropriate basis for selection, it is discriminatory to exclude all women from jobs requiring heavy lifting on the grounds that most women are unable to do it. Lifting abilities are easy to test on an individual basis, and hence it is fairer to require that people be selected as individuals rather than on the basis of some class to which they belong. Statistical discrimination can be excluded by the anti-discrimination principle if we specify that ‘choose … on the basis of …’ covers and thus forbids using the excluded characteristics as statistical indicators of the presence or absence of qualifications.
(3) *Reaction qualifications*. When the customers of a business are prejudiced it may be rational for the business to conform to those prejudices in choosing its employees. It is not in general wrong to select people for positions on the basis of the preferences of the people they will serve. If customers like travel agents with pleasant personalities there is nothing wrong with making the possession of such a personality a job qualification for travel agents. Thus we cannot say that it is arbitrary to select people on the basis of reaction qualifications, but in some contexts it is discriminatory to do so. Wertheimer (1983) argues that relying on reaction qualifications based on race and gender often leads to injustices. It may be possible to bring reaction qualifications under the anti-discrimination principle by extending ‘choose … on the grounds of …’ to cover choosing on the basis of one’s customers’ prejudices.
(4) *Personal connections*. Ezorsky (1991) argues that finding new employees by word-of-mouth advertising within the social networks of one’s present employees is discriminatory towards blacks in contexts where blacks have been excluded from employment by past discrimination and where residential segregation persists. Although this method of recruiting employees might be unobjectionable in some circumstances, it is discriminatory in societies with a history of racism. To prohibit this sort of discrimination we probably have to go beyond the negative anti-discrimination principle to a supplementary requirement that jobs be advertised publicly and that employment through personal connections be avoided.
(5) *Legacy of discrimination*. Allocative decisions can transmit a legacy of discrimination even though they themselves are not discriminatory in the standard sense. Black people growing up in the southern states of the USA in the 1950s were likely to have had educational opportunities inferior to those of whites and to have faced substantial discrimination in seeking employment when they left school. Thirty years later, when such a person sought a job from a company that had fully non-discriminatory hiring policies, they could well have had a less

good record of work experience than a white who was not disadvantaged early on by discrimination in education and employment. Choosing a white person because of better work experience in such a case would not be directly discriminatory, but it would transmit the legacy of discrimination. How to address this sort of case seems more a matter of affirmative action than anti-discrimination.

3 What makes discrimination wrong?

Consider a clear violation of the anti-discrimination principle, such as an employer prejudicially refusing to hire or even consider blacks for work in a restaurant. What is wrong with such an action? First, it is insulting to blacks because of its underlying premise that some racial groups are less worthy than others. Second, it is harmful to blacks because it often reduces self-esteem and produces a sense of inferiority, and because it deprives people of opportunities that would have allowed them to live better lives and make greater contributions to society. Discrimination in education and employment will tend to reduce a group's aspirations and productivity by reducing the payoff of investments in education and work experience. Third, discrimination is often irrational. It frequently relies on unjustified beliefs and stereotypes, and selects people on irrelevant and arbitrary grounds. Although it is not in general morally wrong to act irrationally, irrational actions that distribute important goods in ways that harm and disadvantage other people may indeed be wrong because they are unfair. Finally, the most important reason why discrimination is wrong is that it is unfair. Victims of racial discrimination in employment, for example, can legitimately complain that as full members of society who have done their part in preparing to participate in and contribute to society it is unfair for them to be handicapped in areas such as education, employment and politics by discrimination flowing from other people's prejudices.

See also: Linguistic discrimination

JAMES W. NICKEL

References and further reading

Alexander, L. (1992) 'What Makes Wrongful Discrimination Wrong?: Biases, Preferences, Stereotypes, and Proxies', *University of Pennsylvania Law Review* 141 (November): 149-219.(A clear but complicated exploration of the morality of various types of discriminatory preferences.)

Banton, M. (1992) 'The Nature and Causes of Racism and Racial Discrimination', *International Sociology* 7 (1): 69-84.(Contrasts discrimination, which is 'universal and normal', with racism, which is 'historically and geographically specific'. Also discusses the United Nations 1965 *Convention on the Elimination of all Forms of Racial Discrimination*.)

Boxill, B. (1984) *Blacks and Social Justice*, Totowa, NJ: Rowman & Allenheld.(Chapter one criticizes views that equate non-discrimination with colour blindness.)

Brest, P. (1976) 'Foreward: In Defense of the Anti-discrimination Principle', *Harvard Law Review* 90 (2): 1.(An attempt to define, limit and defend a principle prohibiting discrimination based on race.)

Ezorsky, G. (1991) *Racism and Justice: The Case for Affirmative Action*, Ithaca, NY: Cornell University Press. (Chapter one contains a good discussion of 'overt and institutional racism'.)

Goldman, A. (1979) *Justice and Reverse Discrimination*, Princeton, NJ: Princeton University Press.(Chapter two constructs a philosophical defence of awarding positions by competence.)

King, M. (1969) 'Letter from Birmingham City Jail', in H. Bedau (ed.) *Civil Disobedience*, New York: Pegasus. (The late civil rights leader's indictment of racism and racial discrimination.)

Thomas, L. (1980) 'Sexism and Racism: Some Conceptual Differences', *Ethics* 90 (2): 239-50.(Argues that racist attitudes are 'relatively easier to give up' than sexist ones.)

Wertheimer, A. (1983) 'Jobs, Qualifications, and Preferences', *Ethics* 94 (2): 99-112.(A thorough discussion of qualifications based on client preferences.)

Dissoi logoi

Dissoi logoi ('Twofold Arguments') is the title scholars apply to a short anonymous collection of arguments for and against various theses. The work, in Greek, is (questionably) dated around 400 BC, and regarded as an interesting, if second-rank, product of the Sophistic age.

The opening sentence of *Dissoi logoi* announces 'twofold arguments' about good and bad: one set to show that good and bad are the same, another to show they are different. Chapters 2-4 follow the same pattern for the opposites seemly and shameful, just and unjust, true and false, respectively. Chapters 5-7 argue for and against a thesis, respectively: 'The demented and the sane, the wise and the ignorant, both do and say the same things', 'Wisdom and virtue cannot be taught or learned' and 'Public offices should be assigned by lot'. Chapter 8 argues positively for the idea that a good speaker will know everything. Chapter 9 recommends the art of memorization, but the manuscripts break off leaving both chapter and treatise incomplete.

This is a mere summary of the document. Nothing can be said with certainty about it. Not only its author, but its text, date, overall purpose and intellectual affiliations remain matters for scholarly speculation. Speculation has run riot.

To begin with the text: its Doric dialect is marred by intrusions of Attic and some Ionic forms. Is this due to the tendency of scribes to normalize the texts they copy? To a non-Doric speaker writing for a Doric-speaking audience? Or did a Doric author rewrite an earlier text? Imperfect Doric betrays numerous forgeries ascribed to early Pythagoreans but produced long *after* 400 BC. The antique feel to the text could be misleading.

What of the standard dating? This rests largely on a mention of the Peloponnesian war (431-404 BC) as the most recent of a series of notable wars going back, via the Persian and Trojan wars, to the battle of the gods and giants. Was there a comparable war between 404 BC and Rome's conquest of the Hellenistic kingdoms? If not, the *Dissoi logoi* could have been written centuries after 404 BC.

Concerning the overall purpose of the piece, on the standard dating, the best guide is Aristotle's description (*Sophistical Refutations* 183b36-184a1) of earlier methods of teaching rhetoric and dialectic: the idea was that memorizing specimen arguments would somehow equip the student to produce, at the appropriate juncture, new instances of a typical pattern. On this hypothesis, the author's intellectual affiliations are not discernible from first-person remarks like 'I assert this latter view' (4.2). These belong to the argumentative strategies a student must memorize. That sometimes the author appears to endorse both sides of the dispute must be disregarded: 'I' does not refer to the author but to the speaker in some imaginary debate.

It is not a high-powered debate. Many of the arguments for and against do not even manage to contradict each other. Good and bad are the same because the same thing is good in some circumstances, bad in others; they are different because 'good' and 'bad' mean different things. 'I am an initiate' is both false (in your mouth) and true (in mine); yet 'false' and 'true' mean different things.

Scholars seek influences and sources for particular arguments in the collection: Protagoras and Hippias are especially popular. But most of what we 'know' about individual Sophists is itself speculation (see Hippias; Protagoras; Sophists). Besides, time moves slower in philosophy than in life: arguments and examples get repeated for centuries after their first appearance. The arguments in chapter 6 for and against the teachability of wisdom and virtue show numerous points of contact with Plato's *Protagoras* (319a-328d) and *Meno* (89e-96d). Is Plato drawing on themes already broached in the *Dissoi logoi* or did our author draw on Plato? Neither conclusion is safe. Philosophy being what it is, statements like 'before now people have frequented sophists and gained no benefit' (6.5) could easily be written when the 'Sophists' were a phenomenon of the distant past.

So why were these feeble arguments preserved? The *Dissoi logoi* survives in fact as an appendix to the manuscripts of Sextus Empiricus, who regularly argues for and against a thesis to induce suspension of judgment about it. This suggests the collection was of interest to Pyrrhonian sceptics, either for historical reasons or because, according to these sceptics, even feeble arguments are useful as the appropriate cure for milder cases of dogmatism (Sextus Empiricus, *Outlines of Pyrrhonism* III 280-1). A sceptic called Zeuxis, a friend of aenesidemus, wrote *On*

Dissoi Logoi (Diogenes Laertius, IX 106). Both Pyrrho himself and his follower Aenesidemus came from Doric-speaking regions. Could Zeuxis be commenting on the *Dissoi logoi*, supposing or pretending it was antique?

Like all suggestions about the *Dissoi logoi*, the above is speculation. Sober readers will suspend judgment on every question about the work.

M.F. BURNYEAT

References and further reading

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge & Kegan Paul.(The account of the *Dissoi logoi* in volume 2, pages 214-20 improves some of the feebler relativistic arguments.)

Diels, H. and Kranz, W. (eds) (1952) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidmann, 6th edn.(The standard collection of the ancient sources. Full Greek text of *Dissoi logoi* is in volume 2, pages 405-16.)

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, vol. 3, 316-19; part 1 is repr. as *The Sophists*, Cambridge: Cambridge University Press, 1971.(Helpful account from the standard viewpoint.)

Robinson, T.M. (1979) *Contrasting Arguments: An Edition of the Dissoi Logoi*, New York: Arno Press.(The first and only full-length edition, with translation, commentary and bibliography of previous work.)

Ryle, G. (1966) *Plato's Progress*, Cambridge: Cambridge University Press, ch. 4.(Entertaining account of the background and point of the work on the conventional dating.)

Sprague, R.K. (1968) 'Dissoi Logoi or Dialexeis', *Mind* 77: 155-67; repr. in *The Older Sophists*, Columbia, SC: University of South Carolina Press, 1972.(English translation of the text from Diels and Kranz (1952).)

Dodgson, Charles Lutwidge (Lewis Carroll) (1832-98)

Dodgson, an Oxford teacher of mathematics, is best known under his pseudonym, Lewis Carroll. Although not an exceptional mathematician, his standing has risen somewhat in the light of recent research. He is also of note as a symbolic logician in the tradition of Boole and De Morgan, as a pioneer in the theory of voting, and as a gifted amateur photographer. His literary output, ranging from satirical pamphleteering, light verse and puzzle-making to an immense correspondence, is again largely amateur in nature, and would hardly have survived without the worldwide success of his three master-works, Alice's Adventures in Wonderland (1865), Through the Looking-Glass (1871) and The Hunting of the Snark (1876). Together with portions of his two-volume fairy-novel Sylvie and Bruno (1889/93) they are the only writings, ostensibly for children, to have attracted or deserved the notice of philosophers.

Dodgson was the eldest son of a North-of-England clergyman. From Rugby School he went on to Christ Church, Oxford, where he obtained a first in mathematics and, in due course, a lectureship which he held for the rest of his working life. A prim, sentimental bachelor clergyman, Dodgson himself was no philosopher; his diaries and letters record the banalities of his London theatre-going, child friendships and contacts with the art world, but reflect throughout a busily inventive rather than a deeply thoughtful mind. Mainly through photography, he got to know a number of eminent people, from Tennyson, Ruskin and the Rossettis to Lord Salisbury, the Prime Minister, but never himself figured as a literary lion, still less a Victorian sage. As Dodgson the mathematician, he would even disclaim authorship of the *Alice* books, never suspecting that these would be the works to bring him philosophical fame, not merely as a fount of allusion for logical analysts but also as an ample wabe for the gyrings and gimblings of literary theorists across the English Channel.

To the analytical philosopher, *Wonderland* provides a wealth of bad arguments on such themes as the criteria of reality and personal identity, the relativity of space and size, the effects of arrested time and causality-reversal, the theory of types, the status of nonentities and the paradoxes of solipsism and idealism. Although nobody imagines that the Mad Hatter, Tweedledee or the White Queen have any serious light to shed on these matters, their dicta remain perennially useful as a model of how not to think, and are freely cited by the likes of Russell, Moore, Broad and even Wittgenstein.

By French neo-structuralists, on the other hand (notably Deleuze and Lecercle), the Carrollian corpus is elevated to a far more solemn significance, and taken to afford genuine insight into the logic of dreams and madness, the nature of nihilism and nonsense, the collapse of all differences between names and things, the deconstruction of language by puns and portmanteau-words, and its ultimate subjugation under the stipulative dictatorship of a Humpty-Dumptyan subject. Like the Freudian commentators of an earlier generation, who found cannibalism and psychosis everywhere they looked in *Alice*, the postmodern critics have discovered in it themes and purposes that would have horrified the author, supposing that anyone could have explained them to him.

That *Alice* has influenced a host of modernist writers, from Kafka and Joyce to Borges, Ionesco, Nabokov, Beckett, Artaud and the entire surrealist movement, it would be idle to dispute. But before Mr Dodgson, the pedantic Oxford logician, is thereupon canonized as an apostle of anarchy and irrationalism, it needs to be stated in his defence that the Carrollian absurd, whatever its consequences, is so far from being morbidly antirational in origin that it issues directly from the nature of logic. Formal logicians care nothing for the content of their propositions; they only want to know the consequences. As arbitrary class-names, snarks and boojums, cabbages and kings, bats and tea-trays, beavers and butchers and things that begin with an 'M' are all equally serviceable as terms for the syllogisms that Dodgson delighted in, and what is the harm, then, in animating them with a touch of fictional life? Carroll, so regarded, is in no way a deconstructor of the established order. He is simply one more in the long line of logical humorists, stretching from Zeno, Euthydemus and Buridan to his own mentor, De Morgan, and continuing thereafter via F.C.S. Schiller to Russell, Jourdain, Quine and Raymond Smullyan. It should be no surprise to anyone, therefore, that all five of the last-mentioned are known devotees of *Alice*.

PETER HEATH

List of works

For details of the following works, the reader is advised to consult *The Lewis Carroll Handbook*, ed. S. Williams *et al.*, listed in References and further reading.

Carroll, Lewis (1865) *Alice's Adventures in Wonderland*, London: Macmillan, 1866.
Carroll, Lewis (1871) *Through the Looking-Glass*, London: Macmillan.
Carroll, Lewis (1876) *The Hunting of the Snark*, London: Macmillan.
Dodgson, C.L. (1879) *Euclid and his Modern Rivals*, London: Macmillan.
Carroll, Lewis (1885) *A Tangled Tale*, London: Macmillan.
Carroll, Lewis (1886) *The Game of Logic*, London: Macmillan.
Carroll, Lewis (1889) *Sylvie and Bruno*, London: Macmillan.
(1893) *Sylvie and Bruno Concluded* , London: Macmillan.
Carroll, Lewis (1896) *Symbolic Logic*, Part I, London: Macmillan.

References and further reading

Abeles, F. (ed.) (1994) *The Mathematical Pamphlets, etc., of C.L. Dodgson*, New York: The Lewis Carroll Society of North America.(The sole collection of its kind, indispensable to the study of Dodgson the mathematician.)

Bakewell, M. (1996) *Lewis Carroll*, London: Heinemann.(Well-written general biography by a competent media professional.)

Bartley, W.W., III (ed.) (1977) *Lewis Carroll's Symbolic Logic*, New York: Clarkson Potter, 1986.(Carroll's unfinished treatise on logic, in an adroitly edited modern reconstruction.)

Clark, A. (1979) *Lewis Carroll: A Biography*, London: Dent, and New York: Schocken.(A sensible account of the life, providing new details, with relatively little speculation.)

Cohen, M.N. (ed.) (1979) *The Letters of Lewis Carroll*, London and New York: Oxford University Press, 2 vols. (Fourteen-hundred letters, annotated with immense care and learning.)

Cohen, M.N. (1995) *Lewis Carroll: a Biography*, London: Macmillan, and New York: Knopf. (Full-length portrait, over-solicitous in places, but still a work of great authority. The treatment of Dodgson's religious opinions is particularly worthy of note.)

Collingwood, S.D. (1898) *The Life and Letters of Lewis Carroll*, London: Fisher Unwin, and New York: Century. (The authorized biography, by Dodgson's nephew. Although written in haste, it has lasted well and supplies a foundation for all the rest.)

Deleuze, G. (1969) *Logique du sens*, Paris: Editions de Minuit; trans. M. Lester as *The Logic of Sense*, New York, Columbia University Press, 1990.(None too much of either for the uninitiated reader; but its post-Freudian ruminations on language are heavily indebted to *Alice*.)

Gardner, M. (1960) *The Annotated Alice*, New York: Clarkson Potter, and London: Penguin; revised edn, 1993. (Along with Hudson's biography, this work is largely responsible for the widespread modern-day interest in *Alice*.)

Gardner, M. (1962) *The Annotated Snark*, New York: Simon & Schuster, and London: Penguin.(Lively and informative guide.)

Gardner, M. (1990) *More Annotated Alice*, New York: Random House.(Another gleaning of Alician lore, much of it supplied by correspondents.)

Green, R.L. (ed.) (1953) *The Diaries of Lewis Carroll*, London: Cassell, and New York: Oxford University Press, 2 vols.(Full but inevitably selective edition, since it omits mathematical and logical entries. Wakeling's new version will supersede it.)

Guiliano, E. (ed.) (1976) *Lewis Carroll Observed*, New York: Clarkson Potter. (Useful anthology of critical essays and reviews.)

Guiliano, E. (ed.) (1982) *Lewis Carroll: A Celebration*, New York: Clarkson Potter.(Critical essays and reviews by various hands; worthwhile, but somewhat leaden as a birthday tribute.)

Guiliano, E. (1980) *Lewis Carroll: An Annotated Bibliography, 1960-1977*, Charlottesville, VA: University of Virginia Press.(For the period it covers, this is a great improvement on the *Handbook*.)

Guiliano, E. and Kincaid, J. (eds) (1982) *Soaring with the Dodo*, New York: The Lewis Carroll Society of North America.(A further selection of original articles and reviews.)

Heath, P. (1974) *The Philosopher's Alice*, New York: St Martin's Press, and London: Academy Editions.(A not-entirely-serious vade mecum to the logico-philosophic undertones of *Alice*.)

Hudson, D. (1954) *Lewis Carroll: An Illustrated Biography*, London: Constable; new edn, New York: Clarkson Potter, 1977.(An updated and much embellished version of a life generally considered the most level-headed of its day.)

Lecercle, J.-J. (1985) *Philosophy through the Looking Glass*, London: Hutchinson, and La Salle, IL: Open Court. (More lucid and coherent foray into the territory frequented by Deleuze, and a useful means of access to his thought.)

Lecercle, J.-J. (1994) *Philosophy of Nonsense,* London and New York: Routledge.(*Alice* under deconstruction once more, in the vain hope of finding a philosophy of language in the wreckage.)

Phillips, R. (ed.) (1971) *Aspects of Alice,* New York: Vanguard.(Reprints a wide range of articles from the earlier critical literature.)

Sutherland, R.D. (1970) *Language and Lewis Carroll*, The Hague: Mouton. (Academic in tone but a very thorough and perceptive study, by a well-read professional linguist.)

Thomas, D. (1996) *Lewis Carroll*, London: John Murray.(A darker portrait than that of Bakewell, lit by stage fire from the Victorian underworld.)

Wakeling, E. (ed.) (1993) *The Oxford Pamphlets, etc., of C.L. Dodgson*, Charlottesville, VA: University of Virginia Press.(Assembles virtually the entire literary output of Dodgson the Oxford don, college worthy and public nuisance.)

Wakeling, E. (ed.) (1993-) *Lewis Carroll's Diaries*, Luton: The Lewis Carroll Society.(An ongoing multivolume edition of material not hitherto published in full.)

Williams, S. and Madan, F. (1924) *The Lewis Carroll Handbook*, Folkestone: Dawson, and Hamden, CT: Archon Books; revised edn, ed. R.L. Green, 1962; revised edn, ed. D. Crutch, 1979.(Long past its prime, despite revisions. Rambling and undependable, it remains, for all that, the chief source of bibliographical information and a necessity for the serious student or collector.)

Dōgen (1200-53)

Dōgen Kigen, the founder of Japanese SōtōZen Buddhism, is most noted for his argument that meditation is the expression or enactment of enlightenment, not the means to attaining it. Dōgen believed that even a novice might achieve insight, however fleeting. The difficulty, however, is in expressing that insight in one's daily acts, both linguistic and non-linguistic. In developing his position, Dōgen articulated a phenomenology of incarnate consciousness and a sophisticated analysis of meaning. His theories of mind-body unity, contextualized meaning, temporality and theory-praxis influenced many prominent modern Japanese philosophers such as Watsuji Tetsurō, Tanabe Hajime and Nishitani Keiji.

Tradition states that Dōgen was of aristocratic background and orphaned at the age of seven. A few years later he entered the Tendai Buddhist monastery on Mount Hiei, one of Japan's great spiritual and intellectual centres of the time. He left the monastery at seventeen because he could not resolve an apparent contradiction in Tendai doctrine, the idea that spiritual practice to achieve enlightenment is necessary even though we are all already somehow inherently enlightened. His quest took him to China in 1223 where he studied under the Chinese Chan (Zen) master Rujing (1163-1228). Dōgen returned to Japan in 1227 with Rujing's verification that he had achieved enlightenment and was a Zen master in his own right. In Japan, Dōgen wrote a series of innovative essays, many on philosophical subjects, later collected as *Shōbōgenzō(Treasury of the True Dharma Eye)*. He spent the final decade of his life establishing a comprehensive monastic community in the isolated mountain regions of today's Fukui prefecture. Sōtōeventually became the most popular form of Zen Buddhism in Japan with a strong following among the laity.

Dōgen was a philosophical realist insofar as he maintained we have access to 'presencing of things as they are' (*genjōkōan*). However, that presence is inherently meaningless. Meaning only arises within context, and context always involves an act of human consciousness. Hence, a simple correspondence between what-is and what-is-thought is not possible (see Truth, correspondence theory of). The nature and function of context must also enter into the consideration of truth and falseness.

Dōgen sometimes drew on the traditional Buddhist example of the ocean. To a fish the ocean is a translucent palace; to a person at sea, it is a great circle extending to the horizon in all directions; to a heavenly being, it is a shimmering string of jewel-like reflections. What is the ocean in itself? Dōgen maintained it is intrinsically none of those things, yet has the capacity to be all of them. The meaningless presence is expressed by such vacuous Buddhist terms as 'suchness', 'thusness' or 'emptiness'. Yet, in the fish's context, the sea truly becomes a palace; in the person's it is a circle, and in the heavenly being's it is jewel-like reflections. In itself, the presence (the ocean) is simply the not-yet-conceptualized, meaningless reality that can be contextualized and made meaningful in an infinite number of ways.

For Dōgen, therefore, error is possible by either overly absolutizing or overly relativizing meaning. Fixed in only one context of understanding, one may absolutize its meanings, assuming them to exist independently of consciousness as the things themselves. For example, although the fish would be correct to think in its context that the presence is a palace, it would be wrong to think that it can *only* be a palace. Alternatively, one may err in the direction of extreme relativism. Because there are a limitless number of contexts giving meaning to the presence as, for example, 'palace', 'circle' or 'jewels', this does not imply the presence can assume any meaning whatsoever. Such extreme relativism would be wrong on two counts. First, in each specific context there is only one correct meaning; to the person at sea, for example, the presence is not a palace. Second, there can be some attributed meanings that cannot be correct in any context. Dōgen gave the example of time as 'flying away'. Since the idea itself is logically incoherent (at what rate does time fly?), it can never be true in any context.

For Dōgen these ruminations on presence, context and meaning had implications for Zen practice. He believed the anxiety of everyday life arises from misunderstanding meaning and misreading contexts. We become fixed in interpreting the world as if there were only a very limited number of possibilities. We force our interpretations to fit that limited repertoire of contexts. This inhibits our capacity to adjust to change, and in fact encourages us to deny the reality or importance of both change and context. Zen practice addresses this predicament in two ways.

First, Dōgen believed that Zen meditation, 'just sitting', is a method for disengaging the normal meaning-bestowing function of consciousness. In the non-affirming, non-negating mode he called 'without thinking', one can encounter phenomena in their contextless, meaningless state. In this respect, insight into presence is immediately available to anyone. Yet, human life requires meaning; thus the second function of Zen meditation is to be aware of the contextualizing, meaning-bestowing process, watching it structure the formerly inchoate 'presence'.

It is extremely difficult to distinguish whether one is watching a spontaneous unfolding of contextual meaning or a projection of a previous meaning on to the phenomena. Therefore, the student benefits from a master who continually shifts the context of discourse and action, requiring the student to express new meanings appropriate to the shifts. To the extent the master does the unexpected, the student becomes aware of having had an expectation, a contextual presupposition about what is meaningful in the situation. The master's actions deconstruct those expectations, those fixed meanings. In Dōgen's terminology, the 'master and student practise together' by opening each other to new vistas of expressive meaning, each true to the newly emergent context. This type of practice, Dōgen maintained, is the expression or enactment of enlightenment, not the test or method used to achieve it. In this responsive openness to meaning, one breaks free of the delusory idea that truth is fixed and reality static.

See also: Buddhist philosophy, Japanese; Japanese philosophy; Meaning and truth; Truth, deflationary theories of

THOMAS P. KASULIS

List of works

Dōgen (1200-53) *Dōgen zenji zenshū(Complete Works of Zen Master Dōgen)*, ed. Ōkubo Dōshū, Tokyo: Chikuma Shōbō, 1969-70, 2 vols.(The standard critical edition with manuscript variants noted.)

Dōgen (1200-53) *Shōbōgenzō(Treasury of the True Dharma Eye)*, trans. Kim Hee-jin, *Flowers of Emptiness: Selections from Dōgen's Shōbōgenzō*, Studies in Asian Thought and Religion 2, Lewiston, NY: The Edwin Mellen Press, 1985; trans. Yokoi Yūhō, *The Shōbō-genzō*, Tokyo: SankibōBusshorin, 1985.(Yokoi is currently the best complete translation of Dōgen's *magnum opus*, but flawed somewhat by lack of philosophical subtlety and inadequate explanatory apparatus. Kim's edition is a translation of the bulk of the text with the aim of capturing in English some of the philosophical wordplay of Dōgen's distinctive style of writing.)

Dōgen (1231) *Bendōwa*, trans. N. Waddell and Abe Masaro, 'Dōgen's Bendōwa', *The Eastern Buddhist* 5, 1970 (1): 124-57.(Technical, excellent translation of one of Dōgen's more philosophical essays.)

Dōgen (1233) *Genjōkōan*, trans. N. Waddell and Abe Masaro, 'ShōbōgenzōGenjōkōan', *The Eastern Buddhist* 6, 1973 (2): 115-28.(Technical, excellent translation of one of Dōgen's more philosophical essays.)

Dōgen (1241) 'ShōbōgenzōBuddha-nature', *The Eastern Buddhist* 8 (2): 94-112; 9 (1): 87-105; 9 (2): 71-87. (Technical, excellent translation of one of Dōgen's more philosophical essays.)

References and further reading

Abe Masao (1992) *A Study of Dōgen: His Philosophy and Religion*, ed. S. Heine, Albany, NY: State University of New York Press.(A collection of essays on Dōgen's philosophy by an important Japanese scholar writing in a comparative context mainly for a Western audience.)

Bielefeldt, C. (1988) *Dōgen's Manuals of Zen Meditation*, Berkeley, CA: University of California Press. (Meticulous, historical study comparing the technical details of Dōgen's theory of meditation with his Zen Buddhist predecessors and rivals.)

Heine, S. (1985) *Existential and Ontological Dimensions of Time in Heidegger and Dōgen*, Albany, NY: State University of New York Press.(Detailed analysis of Dōgen's theory of temporality in comparison with that of Heidegger.)

Kasulis, T.P. (1981) *Zen Action/Zen Person*, Honolulu, HI: University of Hawaii Press. (Chapters 6-7 explain more fully the aspects of Dōgen's thought presented in this entry.)

Kim Hee-jin (1975) *Dōgen Kigen Mystical Realist*, Association for Asian Studies Monograph XXIX, Tucson, AZ: University of Arizona Press.(Pioneering work in making Dōgen's philosophical perspective available to Western readers.)

LaFleur, W.R. (ed.) (1985) *Dōgen Studies*, Honolulu, HI: University of Hawaii Press. (Good collection of essays by several Dōgen scholars on a variety of points, many philosophical.)

Nagatomo Shigenori (1992) *Attunement Through the Body*, Albany, NY: State University of New York Press.(A provocative theory of mind-body unity, drawing on traditional and modern Japanese philosophical ideas. Chapters 4-7 deal specifically with Dōgen.)

Shaner, D.E. (1985) *The Bodymind Experience in Japanese Buddhism: A Phenomenological Study of Kūkai and Dōgen*, Albany, NY: State University Press of New York.(Using Western phenomenological categories, a good analysis of Dōgen's theory of the oneness of body and mind.)

Tanahashi Kazuaki (ed.) (1985) *Moon in a Dewdrop: Writings of Zen Master Dōgen*, San Francisco, CA: North Point Press.(A readable translation of a broad sampling of Dōgen's writings with copious annotations for technical terms; useful introduction to Dōgen's writings.)

Dong Zhongshu (195-115 BC)

Tradition hailed Dong Zhongshu as the 'father of Han Confucianism' because of his influential theories that posit a perfect congruence between divine and human realms kept in balance by the true king, who functions as mediator, moral exemplar and lawmaker. Undoubtedly the most famous exegete in the 'Gongyang' commentarial tradition to the Chunqiu (Spring and Autumn Annals), Dong is also credited by convention with the composition of the Chunqiufanlu (Luxuriant Dew of the Annals), though recent scholarship questions this attribution.

Major revisions in the conventional assessment of Dong Zhongshu are under way for several reasons. First, no early evidence supports the view of Dong as architect responsible for adapting pre-Qin Confucian thought to the needs of the centralized imperial Chinese state. Second, many theoretical innovations once tied to Dong can now be shown to either predate (sometimes by millennia) or postdate Dong, requiring a systematic reconsideration of Dong's distinctive contributions. For example, it now seems that Dong did not introduce either Five Phases theory or the Red-White-Black 'three ages' historical cycle into Confucianism. Third, the *Chunqiufanlu* is now more often seen as a composite text that may represent some 150 years of Gongyang scholarship, compelling scholars to begin the arduous task of isolating those passages assignable to Dong alone.

A review of all the extant material that is indisputably authentic shows the central importance for Dong of Heaven's relation to the king (see Tian). According to Dong, Heaven originates and embodies the impulse towards perfection in all things, then works through impersonal natural forces (specifically, the cycles of *yinqi* and *yangqi* (see Qi; Yin-yang)) to foster good rule by suasive virtue, thereby obviating the need for harsh punishments. A perfect congruence between the natural (*tiandi*) and human worlds means that change in one sphere produces comparable change in the other, so that human acts (especially those of the ruler and his immediate circle) influence cosmic harmony for good or for ill. This triadic correspondence is exemplified by the Chinese character for 'king', in which three parallel horizontal lines (for heaven-earth-man) are connected by a single vertical signifying the mediating 'king'. Implicit in Dong's simple vision are three important presumptions: (1) in cosmos and society, hierarchy is 'natural'; (2) rule by moral suasion can be perfectly effective because it partakes of the divine; and (3) unusual events (such as floods, droughts and fires) are to be analysed according to set rules imbedded in the *Chunqiu*, for they convey Heaven's encoded messages to the ruler.

Dong's thought may be reconstructed in three additional areas for which sufficient reliable material survives, namely his theories on law, human nature and economic policy. As jurist, Dong argued that the Han statutes should be modified in light of the Confucian principle of 'rectifying names' (see Logic in China §1; Confucian Philosophy, Chinese). In essence, Dong asserted that subordinates are obligated only to those superiors who have properly fulfilled their ascribed societal roles. (This assertion, of course, had important implications for legitimacy theory.) With regard to human nature, Dong believed that all humans at birth have the potential for both good and evil because Heaven endows both *yinqi* and *yangqi*; the transforming influence of the ruler and of ritual is needed if ordinary persons are fully to realize Heaven's intention. At one stroke, Dong thereby resolved the quarrel between Mencius and Xunzi over human nature and the origin of evil (see Xing). In his memorials on the economy, Dong urged the state to induce commoners to return to the 'basic occupation' of agriculture, while minimizing the gross inequalities in wealth that inevitably prove divisive to society. The ruler must insure his subjects' economic and physical security, and then moral education, if he is to fulfill his sacred obligations as Son of Heaven.

Internal contradictions abound in Dong's thought. For example, both Heaven and the sage seem at once omnipotent and sharply constrained, not only by great cosmic and historical cycles but also by the others' will to take the moral initiative. Thus it remains unclear which of several factors (such as the reigning monarch's character, receipt of the Mandate from Heaven or the operation of impersonal cycles) is primarily responsible for the state of the empire at any one time. Moreover, Dong's call for the total abolition of punishment for the common people is at odds with his firm belief that even sage-rulers provide less than perfect moral models for their subjects. Given such apparent confusion, Dong is often remembered best for three memorials urging the Han throne to reserve imperial patronage for Confucianism alone, since continued competition among contending ways of thought would preclude any attempts to forge a truly unified political and social system.

See also: Chinese Classics; Confucian philosophy, Chinese; Tian; Yin-yang

MICHAEL NYLAN

List of works

Dong Zhongshu (195-115 BC) *Dong Zhongshu jueshi (Legal Decisions of Dong Zhongshu)*, in Ma Guohan, *Yuhanshan fangji yishu*, Taibei: Shijie shuju, 1967, vol. 2, 1180-1.(This contains fragments of Dong's legal pronouncements. The bulk of the *Dong Zhongshu* (123 chapters) is now lost. It has been erroneously identified since the 6th century AD with the extant *Chunqiufanlu*, which includes material from the earlier *Gongyang* interpretive tradition, as well as material postdating Dong.)

References and further reading

Arbuckle, G. (1991) 'Restoring Dong Zhongshu (BCE 195-115): An Experiment in Historical and Philosophical Reconstruction', Ph.D. thesis, University of British Columbia.(An excellent overview of Dong's life and work.)

Dai Jun-ren. (1968) 'Dong Zhongshu bushuo wuxing kao', *Guoli zhongyang tushuguan guankan* 2 (2): 9-19.(An early challenge to the authenticity of the *Chunqiufanlu*.)

Keimatsu Mitsuo (1959) 'ShunjūHanrigogyōshohen gisaku kō', *Kanazawa Daigaku hōbun gakubu ronshū* 6: 25-46.(The first systematic argument alleging the spurious nature of the Five Phases chapters of the *Chunqiufanlu*.)

Pankenier, D. (1990) 'Sandai Astronomical Origins of Heaven's Mandate', paper prepared for the 6th International Conference on the History of Science in China, Cambridge, 2-7 August 1990.(Discusses the early origins of Five Phases and the Black-White-Red cyclical scheme of history.)

Queen, S. (1996) *From Chronicle to Canon: The Hermeneutics of the Spring and Autumn, According to Tung Chung-shu*, Cambridge: Cambridge University Press.(A fine introduction to Dong, with special emphasis on his elaboration of *Gongyang* theories.)

Dooyeweerd, Herman (1894-1977)

A Dutch philosopher and legal theorist, Herman Dooyeweerd challenged the Enlightenment ideal of autonomous rational thought and sought a religious foundation for philosophy. Religious self-knowledge, he contended, is the necessary condition of all knowing. Dooyeweerd proposed an elaborate ontology of physical and social reality based on the Christian doctrine of creation. His thought has been influential among Calvinists in The Netherlands, North America and South Africa.

Dooyeweerd was a Dutch neo-Calvinist philosopher and professor of law at the Free University of Amsterdam (1926-65). His most notable achievement was his lifelong effort to develop a distinctly Christian philosophy that included an epistemology which directly challenged modern Kantian rationalism and an ontology rooted in the Christian doctrine of creation. With D.H.T. Vollenhoven he founded the international Society for Calvinist Philosophy in 1935. His full philosophical statement, *A New Critique of Theoretical Thought*, anticipates a number of late-twentieth-century philosophical interests, including postmodern critiques of Enlightenment rationalism.

Dooyeweerd's work self-consciously builds on the late-nineteenth-century Dutch Calvinist revival movement inspired and directed by Abraham Kuyper (1837-1920). This neo-Calvinism had as its goal the full rechristianizing of Dutch culture and society, and proceeded from a strong Augustinian sense of spiritual antithesis between Christian and secular culture. For Dooyeweerd, as for Kuyper, all theorizing was rooted in religious presuppositions; the Enlightenment prejudice against prejudice was itself rooted in personal commitments. All thought is engaged and interested thought. Dooyeweerd's method of exposing the Enlightenment bias - his 'transcendental critique' - is one of his truly creative and significant accomplishments.

Unlike some Christian philosophers who attempt transcendent or external critiques of the pretended autonomy of modern thought and simply *assert* bias, Dooyeweerd tried to show that an immanent examination of thought itself reveals its religious presuppositions. Analogously to Kant's critical method, Dooyeweerd, who was decisively influenced by Neo-Kantian philosophy as well as by the phenomenology of Husserl, argued that a consideration of the necessary conditions of thought discloses a supratheoretical or religious starting point. Philosophy's concern about meaning directs all thought to questions of origin, and these questions cannot be answered by theoretical thought itself. This awareness of the boundary limits of theoretical thought forces the philosopher to confront the integrating capacity of the human ego or, in biblical terms, the heart. Self-knowledge is thus the basis of all epistemology, and pure presuppositionless rationalism is an impossibility.

Religion is thus the integrative reality of all thought. Dooyeweerd did not restrict his understanding of 'religion' to concrete, particular, historical religions, but, like Tillich (§4), understood it to be whatever human beings might give ultimate allegiance or value to. In his judgment, four religious, integrative 'ground-motives' dominate the history of Western thought: the Greek 'form-matter' motive, the medieval synthesis of 'nature-grace', the modern dialectic of 'nature' ('necessity') and 'freedom', and the biblical motive of 'Creation-Fall-Redemption'. For Dooyeweerd, as for neo-Calvinism, redemption is the restoration of creation.

Dooyeweerd's own ontology is rooted in Christian convictions about creation, particularly the conviction that creation is meaningfully diverse and structured by God's law. In Dutch his philosophy is called *Wijsbegeerte der Wetsidee* ('Philosophy of the Law-Idea'). Here Dooyeweerd also built on Kuyper's thought, this time on the notion of 'sphere-sovereignty': the idea that physical and social reality exhibits an irreducible diversity which is grounded in divine ordinances. While Kuyper utilized this doctrine primarily for political purposes (to free social structures such as the family and the school from the hegemony of the state), Dooyeweerd expanded it to a full ontology of 'modal aspects'. These modal aspects of created reality (including numerical, spatial, biotic, psychical, lingual, economic, jural-legal) are the investigative areas of the special sciences. The task of philosophy is to examine and explain how the modalities relate to each other. Dooyeweerd also developed a complex theory of whole 'entities' - termed 'individuality structures' - each of which in a unique configuration exhibits all the modal aspects.

Emphasizing the doctrine of creation also led Dooyeweerd to a rather positive attitude towards human culture and history. Strongly affirming Calvin's and Kuyper's doctrine of 'common grace' - God's non-saving favour towards all humanity - Dooyeweerd articulated a philosophy of history that was very appreciative of cultural development

(see Calvin, J. §§2-3). Cultural products and social institutions were perceived as organically unfolding and differentiating from their creational roots. Both this differentiation and cultural diversity are to be respected.

In addition to his native country, Dooyeweerd has had significant attention and influence in North America and South Africa. A ‘Dooyeweerd Center for Christian Philosophy’, which will concentrate on publishing Dooyeweerd’s works in English and stimulating the study of his thought, was opened at Redeemer College, Canada, in 1994.

JOHN BOLT

List of works

Dooyeweerd, H. (1953-8) *A New Critique of Theoretical Thought*, Amsterdam: H.J. Paris and Philadelphia, PA: Presbyterian & Reformed Publishing Company, 3 vols. (The full and mature expression of Dooyeweerd’s epistemology and modal ontology.)

Dooyeweerd, H. (1960) *In the Twilight of Western Thought*, Philadelphia, PA: Presbyterian & Reformed Publishing Company.(A series of lectures given by Dooyeweerd on a North American tour, this remains the most accessible introduction to Dooyeweerd’s writing.)

Dooyeweerd, H. (1979) *Roots of Western Culture*, Toronto, Ont.: Wedge.(This is a more elaborate introduction to Dooyeweerd’s cultural analysis, including the four ground-motives.)

References and further reading

Kalsbeek, L. (1975) *Contours of a Christian Philosophy: An Introduction to Herman Dooyeweerd’s Thought*, Toronto, Ont.: Wedge.(A very readable entry into Dooyeweerd’s often complex thought.)

Kuyper, A. (1931) *Lectures in Calvinism*, Grand Rapids, MI: Eerdmans.(Originally delivered as the Stone Lectures at Princeton University in 1898, these are indispensable background reading for understanding Dooyeweerd’s neo-Calvinist vision.)

McIntire, C.T. (1985a) ‘Herman Dooyeweerd in North America’, in D. Wells (ed.) *Reformed Theology in America*, Grand Rapids, MI: Eerdmans, 172-85.(Traces Dooyeweerd’s philosophical influence among Calvinists in North America.)

McIntire, C.T. (ed.) (1985b) *The Legacy of Herman Dooyeweerd*, Lanham, MD: University Press of America. (Contains a series of explorations of Dooyeweerd’s thought by the faculty of the Institute for Christian Studies, Toronto, Canada. It also contains a thorough bibliography.)

Dostoevskii, Fëdor Mikhailovich (1821-81)

Dostoevskii, regarded as one of the world's greatest novelists, is especially well known for his mastery of philosophical or ideological fiction. In his works, characters espouse intriguing ideas about theology, morality and psychology. Plots are shaped by conflicts of ideas and by the interaction of theories with the psychology of the people who espouse them. Indeed, Dostoevskii is usually considered one of the greatest psychologists in the history of Western thought, not only because of the accounts of the mind his characters and narrators elucidate in detail, but also because of the peculiar behaviour betraying the depths of their souls. Dostoevskii is particularly well known for his description of the irrational in its many modes.

Deeply engaged with the political and social problems of his day, Dostoevskii brought his understanding of individual and social psychology to bear on contemporary issues and gave them a lasting relevance. His predictions about the likely consequences of influential ideas, such as communism and the social theory of crime, have proven astonishingly accurate; he has often been regarded as something of a prophet of the twentieth century.

His reputation rests primarily on four long philosophical novels - Prestuplenie i nakazanie (Crime and Punishment) (1866), Idiot (The Idiot) (1868-9), Besy (The Possessed, also known as The Devils) (1871-2) and Brat'ia Karamazovy (The Brothers Karamazov) (1879-80) - and on one novella, Zapiski iz podpol'ia (Notes From Underground) (1864). In his day, Dostoevskii was as famous for his journalistic writing as for his fiction, and a few of his articles have remained classics, including 'Mr. D-bov and the Question of Art' (1861) - a critique of utilitarian aesthetics - and 'Environment' (1873).

Dostoevskii's works have had major influence on Western and Russian philosophy. In Russia, his novels inspired numerous religious thinkers, including Sergei Bulgakov and Nikolai Berdiaev; existentialists, such as Lev Shestov; and literary and ethical theorists, most notably Mikhail Bakhtin. In the West, his influence has also been great. Here, too, his writings are repeatedly cited (along with Kierkegaard's) as founding works of existentialism. Perhaps because of a misreading, they influenced Freud and Freudianism. Directly and through the medium of Bakhtin, his ideas have played a role in the rethinking of mind and language. And his rejection of utopianism and socialism has been repeatedly cited in twentieth-century political debates and theories.

Dostoevskii's influence has been diverse and at times contradictory, in part because of the different genres in which his ideas are expressed. Not only the overall meanings of his novels but also the views of his characters, including those he meant to refute, have been attributed to him. Moreover, his essays sometimes express ideas at variance with his novels. Most recently, philosophical significance has been discovered not only in the content but also in the very form of his novels. Their odd plot structure has been shown to have implications for an understanding of authorship, responsibility and time.

1 Life

The son of a military surgeon, Dostoevskii grew up in a pious household to which he attributed his (rather tortured) religious faith. Celebrated at a young age for his early fiction, he became involved in a secret socialist society, the Petrashevskii Circle, which was infiltrated by the tsarist secret police. In 1849 he was arrested and, after spending months in prison, was sentenced to be executed, a sentence reduced at the last possible moment. The experience of mock-execution shaped both his sense of how the human mind works in extreme situations and his obsession with temporality. Dostoevskii then spent years in a Siberian prison camp, which he later described in his short novel *Notes From the House of the Dead* (1860-1), and several more years in internal exile before being allowed to return to European Russia. By his own account, his time in Siberia changed his views decisively. He rediscovered his religious faith and decisively rejected Western utopian socialism. Imprisonment itself led him to appreciate the importance of privacy and of individual choice, which he felt were rejected by socialists. Attending to the psychology of criminals, he came to see more clearly the irrational in everyone and rejected rationalist and utilitarian models of the psyche. Above all, his experience living among common people developed in him a deep distaste for the condescension of intellectuals who professed to save 'the people', whom they hardly knew. In the process, he came to reject the intellectuals' principal tool, abstract theory, as inadequate to the complexity of real human needs and motives.

Upon returning to Russia, Dostoevskii worked alternately as a journalist and a novelist. During this time, he suffered from a variety of misfortunes, including the epilepsy that first struck him in Siberia, compulsive gambling and constant indebtedness. He was forced to travel abroad to escape creditors and developed a deep distaste for the West. During his epileptic seizures, he experienced an extraordinary sense of universal harmony and of the cessation of all temporal processes, topics that appear prominently in his fiction (especially *The Idiot*) and fed his interest in time. His gambling led him to explore the human need for the radically uncertain and (a favourite problem) the nature of self-destructive behaviour performed by people who recognize it as such. His indebtedness and poverty, which reached extremes, sometimes forced him to write and publish his novels serially without an advance plan. Consequently, he was at times as uncertain as his characters about their destiny. He eventually made a virtue of this process, and by so doing discovered new ways to represent the openness of the future and the palpability of choice. He managed to transfer the excitement of the creative process to the novels themselves as he wrote them and as the readers read them.

His politics grew increasingly conservative, or rather anti-revolutionary, as he wrote novels about people possessed by extreme and radical ideologies. Although for most of his life Dostoevskii was severely sceptical about the possibility of a substantive philosophy of history, for a period of about two years (1876-7), he reversed his stance and imagined he had discovered the key to history and the signs of an imminent apocalypse, after which Russia would assume a messianic role and lead the world. These views appear in *Dnevnik pisatelia* (*A Writer's Diary*) (published intermittently from 1873 to 1881), a work that also experimented with turning a one-person periodical into a peculiar literary genre open to the contingencies of current events. Some articles written during this period, which occasionally included anti-Semitic diatribes, remain the greatest embarrassment to his contemporary admirers. Having come to his senses, he produced his greatest and probably most humane work, *The Brothers Karamazov*, shortly before his death in 1881.

2 Core ideas

Perhaps the core idea of Dostoevskii's works is human freedom. Famous as a psychological writer, he declared that he was not in fact a psychologist, if by that term one meant a person who regards human behaviour as governed by laws that make it in principle predictable. He described the historical process as not subject to laws and decisively rejected any theory guaranteeing progress (or any other outcome). In his theology, he insisted that the essence of Christianity was free choice, and so any religion that offered 'miracle, mystery, and authority' (a phrase in *The Brothers Karamazov*) as ways of allaying doubt were in fact anti-Christian. Thus uncertainty and striving without guarantees was essential, not detrimental, to Christian belief. For a similar reason, utopian politics promising to offer a final solution to ensure happiness would deprive us of our humanity, which is essentially a matter of effort, of choices that can be mistaken, and of processes without guaranteed outcomes.

Dostoevskii's novels consequently develop an antideterministic theory of time in which each moment possesses the potential for multiple futures. Whatever did happen at a given moment something else might have. In his articles on crime he argued, and in his novels he illustrated, that human intentions are often essentially processual: there may never be a moment when a decision is made, even subconsciously. Rather, a shifting and always incomplete process governs behaviour from moment to moment. In such cases, the ascription of an intention in the usual sense is a retrospective activity oversimplifying the facts. If a crime committed in this way could be repeated, the result might well be different, because the process allows for many possibilities at each moment of choice. One meaning of the image of the devil, who appears as a character in *The Brothers Karamazov*, is the personification of the forces introducing radical uncertainty into the world.

It followed for Dostoevskii that all ethical philosophy aspiring to iron-clad rules for moral action must be mistaken; for if such rules could be provided, action according to them would lack the quality of choosing the good in the face of uncertainty. Christ's image, not an abstract philosophy, must be our guide. The unknowability of the future also explains the fallacy behind utilitarian views of art. Dostoevskii agrees that art should be useful, but he contends that utility should apply to an uncertain future, when our purposes may be far from those of the present. Great art lasts because it is capable of affecting us in radically different times and places. Among the qualities that allow art to be unforeseeably useful is beauty.

Insisting on human unpredictability, Dostoevskii rejected utopian schemes because they were based on too great a

confidence in the needs of humanity as a whole and too little concern for individuals to make their own varying choices. Indeed, thinkers who place too much faith in an abstract theory are unlikely to recognize that people are essentially processual. Nevertheless, Dostoevskii believed that utopians with an unjustifiable faith in salvationist theories would likely seize control. As a result, he was probably unique among nineteenth-century writers in predicting that the twentieth century would be the era of what we have come to call totalitarianism, the features of which he described (especially in *The Possessed*) in remarkably accurate detail.

Dostoevskii's distrust of theory, especially his rejection of materialist and rationalist definitions of humanity, led him to another prescient prediction, this one pertaining to Western liberal societies. Whereas others typically assumed that increasing wealth and expanding freedoms would lead to reduced crime, Dostoevskii predicted the opposite. What would swamp liberal progress would be a decline in religious belief and the conviction that moral norms transcend utilitarian purposes. The more people come to accept that good and evil are merely social constructs, the less restraint they will feel. If God is dead, then all is permitted, as Ivan Karamazov says (in *The Brothers Karamazov*). No amount of law enforcement could ever compensate for the inner conviction that crime is simply morally wrong in the eyes of God or from the perspective of some transcendent value. Moreover, the very theory that crime is merely the product of the social environment is itself part of the social environment. Its acceptance is bound to increase crime. Dostoevskii repeatedly performs this self-reflexive act on theories, a move that fits well with his novelistic technique of examining ideas by their lived consequences.

3 Major works before *The Brothers Karamazov*

Zapiski iz podpol'ia (*Notes From Underground*) (1864) consists of two parts, a paradoxical philosophical diatribe by the narrator and a series of reminiscences of his earlier life. Part I argues, and both parts illustrate, the complexity of psychological processes that transcend all theories presuming human rationality. His critique of all rational choice models can still stand as a useful corrective today. These models fail for two important reasons. First, they are false to actual experience. The underground man cites witty examples from daily life and history that only the most tortured or tautological reasoning could subsume under a model of rationality. Second, the most important human preference does not resemble the others and cannot be listed along with them: the desire to choose unpredictably. Any model that makes human action in principle predictable deprives us of our very selves. Satisfy all our desires, and we will desire to harm ourselves, simply to show that we are people and not 'organ stops' or 'piano keys'. That is, our greatest desire is a meta-desire, a longing to desire unpredictably. That is the source of the underground man's 'spite', a concept to which he gives metaphysical significance.

In the course of his tirade, the underground man also rejects the existence of any historical laws and all models discerning an inner rationality to history. Alluding to Hegel, he observes: 'The only thing one cannot say [about history] is that it is rational'. Polemicizing with Buckle and liberal faith in a law of progress, he observes that civilization has given humanity more efficient and crueller ways to kill and torture each other (another passage anticipating the twentieth century). Thinkers invent historical and social laws because they have decided in advance that only such an approach is truly scientific, which is ironic because science involves fidelity to the facts. 'But man is so fond of systems and abstract deductions that he is ready to distort the truth intentionally, he is ready to deny what he can see and hear just to justify his logic' ([1912-20] 1960: 21).

If determinism were correct and if scholars really could know the true laws of history and psychology, then human life would turn into 'extracting square roots' and behaving according to 'a table of logarithms'. Life would then be simply a finished product, but it can be meaningful only as a process. All utopian schemes, with their image of a finished and perfected world, would, if realized, become prisons, houses of the dead. Nevertheless, Dostoevskii also rejects the underground man's way of opposing such rationalism by pure spite. Spite, the novel demonstrates, also has its own iron-clad, if perverse, logic. Real freedom cannot be found simply by inverting determinism, as the underground man does. Rather, the work implies, it demands the Christian concept of humanity as free and the self as developing in relation to others.

In *Crime and Punishment* (1866), Dostoevskii explores the psychology of an intellectual who wants to live according to various theories justifying murder. The novel shows each of these theories to be inadequate to the way people actually live and feel. Indeed, the theories motivating the hero contradict each other, as one justifies killing as moral on utilitarian grounds while another denies that morality is anything but an empty social construct. Moreover, the hero finds himself unable to account, after the crime, for his decision to commit it, the reason being

that he never actually made a decision. He was, rather, possessed by a state of mind that guided him processually, moment by moment, in a dream-like state. Taken as a whole, the novel refutes ideology - not just particular theories, but the ideological cast of mind itself - by psychology.

One of the hero's theories is that humanity is divided into two groups, a majority of ordinary people who serve as mere breeders and a minority of innovators who have the right to transgress all moral codes. Dostoevskii suggests that the desire to be a superman in this way is typical of any intelligentsia that takes its theories too seriously. Even, or especially, if such schemes promise to make people happy, they will lead to killing and cause more harm than good. But they exert an almost irresistible appeal for intelligentsias.

The Idiot (1868-9) written without an advance plan with the author discovering the course of events along with his characters, may be taken as a kind of thought experiment. The author creates a perfectly good man and sees what effect he would have on real people, whose psychology includes the perverse and irrational as well as more normal human motivations. People often react to the Idiot's goodness by feeling resentful at their own moral inferiority, which leads them to still worse behaviour. The result is that, because of his very goodness, he destroys both himself and almost everyone around him.

In effect turning its own processual composition into a theme, this novel also explores the nature of time, arguing (in what is probably the work's most famous passage) that 'Columbus was happy not when he had discovered America, but when he was discovering it.… It's life that matters, nothing but life - the process of discovering, the everlasting and perpetual process, not the discovery itself, at all' ([1912-20] 1962: 21).

The Possessed, often considered the greatest European political novel, explores the mentality of revolutionaries and fashionable liberals. Almost as critical of liberals as of radicals, Dostoevskii saw the former as the progenitors of the latter, even though they are often horrified by them. Both share the mentality of the intelligentsia, which views itself as entrusted with the salvation of humanity. The novel criticizes the tendency of utopians to reason down from abstract principles, which leads them to absurd but dangerous social conclusions. Arguably the most quoted line of the novel belongs to one radical theoretician: 'Starting from unlimited freedom, I arrive at unlimited despotism'. In a passage that seems to anticipate the Chinese Cultural Revolution, the novel attacks radical egalitarianism as an ideal. Under socialism, one revolutionary declares, 'All are slaves and equal in their slavery… the great thing about it is equality… Cicero will have his tongue cut out, Copernicus will have his eyes put out, Shakespeare will be stoned' ([1912-20] 1963: 424).

4 *The Brothers Karamazov*

More extensively than any other work of Dostoevskii, *The Brothers Karamazov* explores the nature of Christian truth. Developing an idea in *The Idiot*, Dostoevskii demonstrates that goodness consists of small, kindly actions, rather than great and heroic deeds. Small actions may have a concatenating effect no one can foresee. Evil, too, is conceived as consisting primarily of day-to-day actions. Thus in one chapter, when the devil appears as a character, he is not grand, Satanic and alien, but ordinary and banal. This 'petty devil' is one of Dostoevskii's most famous creations.

The height of Dostoevskii's achievement appears in a sequence of three chapters, in which the intellectual brother, Ivan, tries to seduce his pious younger brother, Alësha, from religion by refuting Christianity on moral grounds. It is remarkable that these arguments, coming from the pen of a Christian author, have often been considered to be the strongest pages ever written against faith. In 'Rebellion', Ivan contrasts all conceivable theodicies with the reality of innocently suffering children. Ivan has also composed 'The Grand Inquisitor', a story in which Christ returns to earth during the Spanish Inquisition and is arrested by the Grand Inquisitor, who contends that Christ should have accepted the temptations of the devil just as (he says) the Catholic Church actually has. The Inquisitor reasons that people do not want to be free, only to think they are free. They can be happy only when their consciences are held by others and their choices are made for them, so they will not feel guilt or regret whatever happens. The Inquisitor's ideals evidently represent socialism as well as Catholicism, two apparently antagonist movements that, in Dostoevskii's view, were really the same because they share an anthropology.

Christ never replies to the Inquisitor, and the Inquisitor's arguments appear irresistible, and yet the reader, for reasons that are hard to specify, adheres all the more strongly to the Christian ideal of freedom. Ivan has 'refuted' God the Father in 'Rebellion' and God the Son in 'The Grand Inquisitor', but he has unwittingly adhered

to the Holy Spirit in ‘The Brothers Make Friends’. Here his sheer love of ‘the sticky green leaves that open in the spring’ expresses a love of God’s world. As elsewhere in the novel, Dostoevskii’s point is that faith comes not from doctrine but from how one lives moment to moment. It is not doctrine that leads to faith, but faith, born of good living, that may lead to acceptance of doctrine, which is not the content of true religion but its expression.

The novel’s plot turns on Ivan’s theory that moral responsibility, if it exists at all, pertains only to actions, not to wishes. Ivan will do nothing to harm his thoroughly noxious father, but ‘in my wishes I reserve myself full latitude’. When his father is killed, Ivan comes to feel guilty precisely for wishing his father’s death. He is right to do so, because the novel demonstrates that most evil happens indirectly, because a climate of evil is created by our evil wishes. In such a climate, crime is bound to happen one way or another. The proximate cause of such crimes is less important than the climate. Because all of us indulge such wishes, ‘everyone is responsible for everyone and for everything’, as the novel’s holy Father Zosima contends ([1912-20] 1950: 244). This logic also explains why the devil, as the personification of evil, is thoroughly ordinary.

We are reminded of the Sermon on the Mount, where Jesus says that we are responsible for our wishes. In this way, the novel demonstrates the truth of the Gospel psychologically. Ivan’s doctrines, whatever their strength as dialectic, run counter to the human spirit. Of course, Dostoevskii’s psychological ‘proof’ of Christianity does not demonstrate the existence of God or the doctrine of the Trinity. But it does show that faith is grounded not in some philosophical propositions, but existentially and anthropologically, in the depths of our being.

Like *Notes From Underground* and *The Possessed*, *The Brothers Karamazov* explores the psychology of ‘the insulted and the humiliated’, a trade mark theme of Dostoevskii. If one understands this spiritual state, one will see that those afflicted with the psychology of their own victimization are precisely the ones who understand best how to inflict pain on others and who are most likely to seek revenge by doing so. It follows that such people, for all their claims to the moral high ground, are the last ones in whom one should entrust political power.

5 Time, form and ethics

Deeply concerned with describing time as open, Dostoevskii remained suspicious that the well-structured plots of traditional novels implicitly endorse determinism or fatalism. The problem is that in these works everything fits and events take place not only because of prior causation, as in life, but also in order to fit a pattern visible when the work is over. Thus, contingency is ruled out. Dostoevskii therefore developed a method of writing ensuring that the reader senses at each moment that whatever happens, something else easily could have. Time therefore becomes a field of possibilities. Nothing is tending to a pregiven ending - which, in fact, Dostoevskii himself sometimes did not know. The unstructured quality of some of his works, which often contain numerous loose ends, results from this technique. On the other hand, the extraordinary suspense of his novels derives from this method of representing time the way it is experienced in life, as truly open. Thus the very form of Dostoevskii’s novels became an instrument of his argument for freedom in an open world containing ‘an infinite multitude of ramifications’ (as one character in *The Idiot* says) and countless possibilities, shaped by chance and by our own actions.

The sense of open time is also reflected in Dostoevskii’s ethical tenets. If people are fundamentally free, then they always have the ability to be surprising. It followed for Dostoevskii that one thing we must never do is treat others as if they were fully known quantities, as if we could be sure what they are going to do, as the various schools of psychology in principle do. In rejecting determinism, materialism, nihilism, historicism and psychologism, while rethinking the nature of narrative art, Dostoevskii explored the meaning of human freedom.

See also: Bakhtin, M.M.; Berdiaev, N.A.; Bulgakov, S.N.; Existentialism; Shestov, L.

GARY SAUL MORSON

List of works

Dostoevskii, F.M. (1972-90) *Polnoe sobranie sochinenii F.M. Dostoevskogo (The Complete Works of F.M. Dostoevskii)*, Leningrad: Nauka, 30 vols.(The standard Russian edition of Dostoevskii’s works.)

[Dostoevsky] (1912-20) *The Novels of Fyodor Dostoevsky*, trans. C. Garnett, 12 vols; repr. New York: Modern Library, 1950; repr. New York: Dutton, 1960. (The best translation of Dostoevskii’s novels, reissued and revised many times.)

[Dostoevsky] (1860-1) *Zapiski iz mërtvogo doma in Vremia*, trans. C. Garnett, *Notes From the House of the Dead*, New York: Dutton, 1960.(Describes Dostoevskii's years in a Siberian prison camp.)
[Dostoevsky] (1864) *Zapiski iz podpol'ia*, trans. C. Garnett, revised R. Matlaw, *Notes From Underground*, New York: Dutton, 1960.(Examined in §3.)
[Dostoevsky] (1866) *Prestuplenie i nakazanie*, in *Russkii vestnik*; trans. C. Garnett, *Crime and Punishment*, New York: Dutton, 1960.(Exploration of the psychology of an intellectual who wants to live according to various theories justifying murder.)
[Dostoevsky] (1868-9) *Idiot*, in *Russkii vestnik*; trans. C. Garnett, *The Idiot*, New York: Modern Library, 1962. (Examined in §3.)
[Dostoevsky] (1871-2) *Besy*, trans. C. Garnett, *The Possessed*, New York: Modern Library, 1963.(Often considered to be the greatest European political novel. Explores the mentality of revolutionaries and fashionable liberals.)
[Dostoevsky] (1873, 1876-7, 1880, 1881) *Dnevnik pisatelia*, trans. K. Lantz, *A Writer's Diary*, Evanston, IL: Northwestern University Press, 1993, 1994, 2 vols.(Referred to in §1.)
[Dostoevsky] (1879-80) *Brat'ia Karamazovy*, in *Russkii vestnik*; trans. C. Garnett, *The Brothers Karamazov*, New York: Modern Library, 1950. (Examined in §4.)
[Dostoevsky] (1963) *Dostoevsky's Occasional Writings*, ed. and trans. D. Magarshack, New York: Random House.(A collection of Dostoevskii's nonfictional works.)

References and further reading

Bakhtin, M. (1929) *Problemy poetiki Dostoevskogo*; expanded 2nd edn 1963; ed. and trans. C. Emerson, *Problems of Dostoevsky's Poetics*, Minneapolis, MN: University of Minnesota Press, 1984.(The most influential book on Dostoevskii, this theoretical study explores the significance of his invention of 'the polyphonic novel', develops a theory of language and relates Dostoevskii's work to ethics, psychology and authorship.)
Frank, J. (1986) *Dostoevsky: The Stir of Liberation, 1860-65*, Princeton, NJ: Princeton University Press.(Part of Frank's five-volume biography, this book situates Dostoevskii's ideas about nihilism, socialism and materialism in the intellectual milieu of his time. It discusses *Notes From Underground*.)
Frank, J. (1995) *Dostoevsky: The Miraculous Years, 1865-71*, Princeton, NJ: Princeton University Press.(This volume of Frank's biography deals with *Crime and Punishment*, *The Idiot* and *The Possessed*.)
Jackson, R.L. (1966) *Dostoevsky's Quest for Form: A Study of His Philosophy of Art*, New Haven, CT: Yale University Press.(Explores Dostoevskii's aesthetics as expressed in his articles and embodied in his work.)
Morson, G.S. (1994) *Narrative and Freedom: The Shadows of Time*, New Haven, CT: Yale University Press.(An exploration of the philosophy of time and narrative form in Dostoevskii and other writers.)

Double effect, principle of

'Double effect' refers to the good and bad effects which may foreseeably follow from one and the same act. The principle of double effect originates in Aquinas' ethics, and is supposed to guide decision about acts with double effect where the bad effect is something that must not be intended, such as the death of an innocent person. The principle permits such acts only if the bad effect is unintended, not disproportionate to the intended good effect, and unavoidable if the good effect is to be achieved. The principle has wide relevance in the moral evaluation of acts which have foreseen double effects. Controversy arises over the identification of the agent's intention in difficult cases, and over the use of the principle to resolve issues such as abortion, euthanasia, the use of pain-relieving drugs which hasten death, self-defence, and the killing of certain sorts of non-combatants in war.

1 The role of the principle

The principle, or doctrine, of double effect is important in the application of moral theory to practical cases. The principle asserts the moral relevance of a distinction between those effects of our acts which are intended and those effects which are foreseen but unintended. Sometimes in acting for the best we bring about an unavoidable yet foreseen bad effect which is not part of what we are aiming to achieve. For instance, a doctor prescribes chemotherapy for cancer, knowing that this treatment also causes hair loss; a driver whose brakes fail swerves in order to avoid a group of children on a pedestrian crossing, foreseeing that the car will hit someone standing at the side of the road; a teacher fails an essay that does not deserve to pass, knowing that the student will be upset. 'Double effect' refers to the two foreseen effects of acts such as these: the good effect (eliminating cancer, saving lives, maintaining academic standards) which the agent intends, and the bad effect (hair loss, injuring or killing someone, causing distress) which the agent foresees but does not act in order to bring about.

Many acts with double effect can be characterized as instances of bringing about the lesser of two evils: opting for what is morally more important, at the expense of foreseeably causing a bad effect which is outweighed by the intended good effect. Where the foreseen bad effect is itself morally serious (such as the death of an innocent person), the principle of double effect holds that the bad effect must be unintended: it must not be what the agent aims to achieve in the circumstances nor a means to bringing about the intended good effect. The principle could permit the driver to swerve in order to avoid hitting the children on the crossing, even though the person standing at the side of the road is thereby seriously endangered. But the principle does not permit an agent to kill an innocent person as a means of saving others; the driver must not aim to kill the person at the side of the road as part of a plan.

2 Origin, development, conditions and application

The principle is traceable to Aquinas' justification, in *Summa theologiae*, of homicide in self-defence. Aquinas held that a private person must never kill anyone intentionally; hence his appeal to 'double effect' in justifying foreseen killing in self-defence. The use of lethal force in self-defence has two foreseen effects: saving oneself (good effect) and the death of the aggressor (bad effect). Aquinas maintained that, provided one's intention is only to use necessary and proportionate force in warding off an unjust attack, the aggressor's death is unintended.

The principle of double effect has been developed subsequently as a general precept of nonconsequentialist ethics. The principle is central to absolutist moral theories which hold that some intended acts are always intrinsically evil and that it is never permissible to use an evil means to achieve a good end. The combination of these two tenets can make moral decision very difficult in some cases in which agents will foreseeably bring about a serious bad effect whatever they do. If the driver whose brakes fail does not swerve, the children on the crossing will probably be killed; if the driver does swerve, the person at the side of the road will probably be killed. As a guide to decision in such cases, the principle holds that under strict conditions it is permissible foreseeably to bring about an effect of a type that it is never permissible to intend. These conditions are: that the act itself (in this case, swerving the car) be morally good or indifferent; that the bad effect (killing the one person) be an unavoidable, unintended effect of the act which also achieves the good effect (saving the children); and that the good effect be sufficiently weighty to warrant causing the bad effect. The last condition, proportionality, must always be taken into account. The agent's good intention, and the fact that the bad effect is unintended, are on their own insufficient to permit foreseeably causing serious harm. Evaluation of the relative seriousness of good and bad

foreseen effects is sometimes reasonably straightforward, but sometimes can be difficult.

The distinction between intended and unintended effects is also more widely invoked as part of nonabsolutist, nonconsequentialist moral theories which hold that the agent's intention is important to the moral evaluation of acts which have both good and bad foreseen effects.

3 Criticisms, difficulties and further applications

Critical discussions of the principle focus on the distinction between incidental and intended effects and on its relevance to practical moral problems. According to the principle, a foreseen bad effect is unintended provided it is incidental to what the agent aims to achieve in the circumstances and to the means of bringing about the intended good effect. If we return to the example of the swerving car, incidental and intended effects are distinguished easily: the driver does not swerve the car in order to hit the person at the side of the road, but in order to avoid hitting the children; the children are saved not by that person's being harmed but by the car's swerving. Exponents of the principle often invoke the so-called 'test of failure' in maintaining that particular foreseen bad effects are incidental both to the agent's aim and to the means of its achievement in the circumstances. This test, which Duff (1990) points out must be applied in the light of the agent's beliefs at the time of acting, asks: would agents fail to achieve what they intend if, contrary to expectation, the foreseen bad effect does not occur? To this question, the swerving driver can answer sincerely, 'No, I intend to harm as few people as possible; my aim will be achieved if the person at the side of the road somehow escapes injury.' However, both the identification of the agent's intention and the distinction between the foreseen good and bad effects, on which the principle relies, can be difficult. For instance, if I can stop an aggressor only by shooting through an innocent person whom the aggressor is using as a shield, do I intend harm to the innocent shield or not? If the 'test of failure' is plausibly to identify particular foreseen bad effects as incidental, such effects must be distinguishable in a defensible way both from the good effect and from the means of achieving it in the circumstances. As has been suggested by Foot (1967), the principle cannot allow one to maintain that one does not intend to kill someone, but only to blow them to pieces. The characterization of the agent's intention in the context of some traditional applications of the principle, such as termination of ectopic pregnancy and homicide in self-defence, is problematic. These important difficulties have been addressed by both advocates and critics of the principle.

The principle of double effect is relevant to critical legal thinking about, for instance, the mental element of murder, and the appropriateness of the defence of necessity in cases of justified risk taking. In paradigm acts of double effect the agent foresees both the good and bad effects as certain or highly probable. The distinction between intended and foreseen effects, however, is also applicable to acts of justifiable risk taking, in which the good and bad effects are incompatible possible outcomes; in such cases the bad effect can even be the more likely to occur. For example, a surgeon might justifiably operate in a desperate attempt to save someone's life, foreseeing that the surgery or its effects may well kill the patient; a parent might justifiably throw a child out of a burning building as its only hope of survival, realizing that the fall may well kill it.

See also: Intention; Responsibility

SUZANNE UNIACKE

References and further reading

Anscombe, G.E.M. (1961) 'War and Murder', in W. Stein (ed.) *Nuclear Weapons: A Catholic Response*, London, Merlin Press; repr. in *Collected Philosophical Papers*, Oxford: Blackwell, 1981, vol. 3, 51-61.(Clear absolutist defence of the principle; critical discussion of its application to killing non-combatants in war.)

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, London: Eyre & Spottiswood, 1966, vol. 38, IIa.IIae.64, esp. answer 7.(Formulation of the principle in context of self-defence.)

Connell, F.J. (1967) 'Double Effect, Principle of', in *New Catholic Encyclopedia*, New York: McGraw-Hill, vol. 4, 1020-2.(Clear traditional statement of the principle and its conditions.)

Duff, R.A. (1990) *Intention, Agency and Criminal Liability*, Oxford: Blackwell, chaps 3-5.(Detailed discussion of the distinction between intention and foresight and its legal significance.)

Finnis, J. (1991) 'Intention and Side-Effects', in R.G. Frey and C.W. Morris (eds) *Liability and Responsibility*, Cambridge: Cambridge University Press, 32-64. (Considers the legal significance of distinction between intention and foresight.)

Foot, P. (1967) 'The Problem of Abortion and the Doctrine of Double Effect', *Oxford Review* 5; repr. in *Virtues*

and Vices, Oxford: Blackwell, 1978, 19-32. (Critique of the principle and its application to abortion.)
Grisez, G. (1970) 'Towards a Consistent Natural Law Ethics of Killing', *The American Journal of Jurisprudence* 15: 64-96.(Complex interpretation of the principle, in terms of its application to abortion and self-defence.)
Mangan, J. (1949) 'An Historical Analysis of the Principle of Double Effect', *Theological Studies* X (1): 41-61. (Outlines the historical beginnings of the principle, and its development through to modern times.)
Quinn, W. (1989) 'Actions, Intentions, and Consequences: The Doctrine of Double Effect', *Philosophy and Public Affairs* 18 (4): 334-51.(Nonabsolutist defence of the principle.)
Uniacke, S. (1984) 'The Doctrine of Double Effect', *The Thomist* 48 (2): 188-218. (General critical discussion of the principle.)
Uniacke, S. (1994) *Permissible Killing: The Self-Defence Justification of Homicide*, Cambridge: Cambridge University Press, ch. 4.(Detailed critique of the principle's application to self-defence.)

Doubt

Doubt is often defined as a state of indecision or hesitancy with respect to accepting or rejecting a given proposition. Thus, doubt is opposed to belief. But doubt is also contrasted with certainty. Since it seems intelligible to say that there are many things we believe without being completely certain about them, it appears that we may not have a unitary concept of doubt.

Although doubt is often associated in philosophy with scepticism, historically the relation between the two is complex. Moreover, some philosophers deny that sceptical arguments have any essential connection with inducing doubts.

Sceptical doubts, as philosophers understand them, differ from ordinary doubts in their depth and generality. We all have doubts about some things. But the philosophical sceptic wonders whether we ever have the slightest reason to believe one thing rather than another. However, the reasonableness of such doubts - and even their intelligibility - remains controversial. The various attitudes philosophers adopt with respect to the status of sceptical doubts characterize the main approaches to epistemological theory.

1 Doubt, belief and certainty

Dictionaries typically run together concepts such as indecision and hesitancy, contrasting doubt with both belief and certainty. But as we can evidently believe something without being certain about it, it seems that our concept of doubt itself may not be clear-cut. It is tempting to speculate that this seeming duality is a result of competing epistemological traditions: an older tradition that identifies knowledge with demonstrative knowledge, hence with rational certainty (see Certainty), and a newer tradition for which probabilistic justification is sufficient. But it is also worth noting that ‘belief’ has become a term of art. Outside philosophy, reference to a person’s ‘beliefs’ is likely to mean their fundamental convictions and not just anything they might be said to accept. Even in philosophy, much discussion of doubt has had in the background questions about religious faith, which might also contribute to a dictionary’s associating belief with certainty.

Thinking of belief in the modern philosopher’s wide sense, how should we understand the relations between belief, certainty and doubt? One question concerns whether belief is an all-or-nothing matter or whether it admits of degrees. For Bayesians, although there are some beliefs to which we assign the highest possible strength, most belief is partial (see Probability theory and epistemology). So from a Bayesian standpoint, although there is a natural contrast between doubt and certainty, doubt is not opposed to belief.

What if we insist that belief is all-or-nothing? It is widely held that there are two types of certainty: subjective (which applies to persons), and objective (which applies to propositions). So we might say that, when we assent to propositions on inconclusive evidence, we recognize *them* to be less than certain but *we* have made up our minds. To assent is to eliminate subjective uncertainty - that is, doubt.

This is not conclusive. Even if belief is all-or-nothing, some beliefs are more firmly rooted than others. If we think of these degrees of entrenchment as degrees of subjective certainty, ‘doubt versus belief’ and ‘doubt versus certainty’ will draw different distinctions. In reply, we might say that a shallowly rooted proposition is one I regard as doubt*ful* - that is, open to reasonable doubt - which does entail its being doubt*ed*. But it seems equally natural to say that this would be something I accept, while still having doubts about it. Our concept of doubt just may not be unitary.

2 Philosophical doubt

In philosophy, doubt has often been connected with scepticism. However, for the ancient sceptics, suspension of judgment is not doubt but its cure (see Pyrrhonism). This makes sense if we think of doubt as an *uncomfortable* state of indecision, the discomfort arising from a persisting desire to decide. If systematic suspension of judgment attenuates this desire, it eliminates the anxiety of doubt.

Scepticism can vary in strength. A fairly weak form of scepticism denies that anything is certain (doubt versus certainty again). By contrast, radical scepticism attacks the possibility of our being able to justify a belief to even the slightest degree. Scepticism can also vary in scope: it can be narrowly focused; it can pertain to certain broad

classes of beliefs (for example, all beliefs about the external world); it can be universal. The sceptic's 'philosophical' doubts are distinguished from ordinary doubts by both their unusual depth and their generality: they extend to matters that are not ordinarily doubted, or even considered open to doubt.

Philosophers have been interested mainly in *reasonable* doubt - doubt for which one has a good reason - and particularly with the *limits* of reasonable doubt. Since sceptical arguments suggest that the scope for reasonable doubt is much wider than we normally suppose, exploring the limits of reasonable doubt is very much a matter of exploring the limits of scepticism.

The fundamental argument for universal scepticism goes back to the classical Greek sceptics and centres on a deadly trilemma: that any attempt to justify a belief leads to a vicious regress, reasoning in a circle, or reliance on brute assumptions. This argument would be defeated if we could identify beliefs which could not rationally be doubted. The quest for such 'basic' beliefs is characteristic of foundationalism (see Foundationalism). However, philosophical opinion has moved steadily away from regarding any beliefs as strictly indubitable. Contemporary foundationalists mostly favour a more modest conception of basic beliefs, according them some intrinsic but defeasible credibility (see Knowledge, defeasibility theory of).

It would be desirable to defeat universal scepticism without resort to the idea of intrinsically credible basic beliefs. Beliefs that can plausibly be accorded this privileged status tend to be unambitious with respect to their content, which makes problematic their adequacy as a basis for reasonable belief. Thus foundationalism meets one kind of scepticism only at the cost of threatening us with others. Consider scepticism with respect to knowledge of the external world. The argument for this form of scepticism famously occurs in Descartes, who claims that, since his entire experience of the world could be a kind of systematic hallucination induced by an evil demon, all his beliefs about external reality can reasonably be doubted, at least until he sees how the possibility of such deception can be eliminated (see Descartes, R. §4). But the thought that experiential knowledge underdetermines what it is reasonable to believe about the world makes a sceptical point only if, in the last analysis, experiential beliefs are intrinsically 'epistemologically prior' to beliefs about the world. Generalized doubts depend on generalized relations of justificatory dependence.

If this is correct, 'Cartesian' doubts depend essentially on foundationalism. However, many philosophers, including Stroud (1984b) and Nagel (1986), deny that any sceptical arguments depend on such highly specific and controversial epistemological commitments. They see scepticism as 'natural' or intuitive in the sense of arising out of the most ordinary and everyday epistemological considerations; but this conception of scepticism is more difficult to defend than is often thought.

Since by ordinary standards the error-possibility Descartes cites as a ground for doubt is remote, how reasonable are his doubts? The answer to this question is complicated by the fact that the 'reasonableness' of doubt is affected by other than purely epistemological considerations. For example, in a situation where the costs of making a mistake are very high, it might be reasonable to remain in doubt about something even though in possession of evidence that would ordinarily be considered sufficient (see Contextualism, epistemological). This suggests a defence of the reasonableness of doubts based on the remote error-possibilities that figure in sceptical arguments. True, such possibilities are not to be taken seriously in ordinary contexts of inquiry. But Descartes' project - a systematic, reconstruction of human knowledge from the ground upwards - is far from ordinary. This project, which requires setting aside all practical concerns, enforces a higher standard of certainty than would be reasonable in everyday contexts. However, this argument, because of its emphasis on certainty, is not clearly a defence of the reasonableness of radical scepticism, which is the most interesting type.

Hume claims that sceptical 'doubts' are not just unreasonable but entirely fictitious. According to Hume, no one seriously doubts that there is an external world because it is psychologically impossible to doubt such 'natural' beliefs. Having a theoretical answer to scepticism is neither here nor there. (This line of thought has recently been endorsed by Strawson (1985).) There are two replies. One draws on the point made above: that sceptical doubts arise only in the context of a distinctive kind of theoretical inquiry. To be sure, no one ordinarily doubts the things the sceptic questions; but why does this show that they are always and everywhere immune from doubt? Even Hume admits to being overwhelmed by sceptical doubts when he withdraws from ordinary concerns and reflects on their rational basis. The second reply (see, for example, Harman 1973) defends the interest of sceptical arguments by denying that they have any essential connection with doubt. We do not need to refute the sceptic to

assuage anyone's actual doubts. Rather, the hope is to learn something about knowledge by seeing how such arguments go wrong. At bottom, these two replies are not all that different. Not even Descartes claims to be able permanently to strip himself of his ordinary beliefs. His doubting them in the course of philosophical inquiry consists in not taking them for granted. In this sense, even the most purely methodological sceptic may be said to doubt his common-sense convictions.

3 Philosophical doubt (cont.)

Highlighting the context-sensitivity of sceptical doubts raises interesting and unresolved issues about the relationship between the results of philosophical reflection and everyday practices of inquiry. The traditional view has been that philosophical reflection brings into view fundamental epistemological facts which 'practical' concerns lead us to ignore. So, if sceptical arguments are irrefutable, we have discovered, under the conditions of philosophical inquiry, that no belief about, say, the external world can be adequately justified. But if philosophical inquiry imposes extraordinary constraints on knowledge and justification, we may only have discovered that knowledge is impossible under the (self-imposed) conditions of philosophical inquiry.

There is a deep disagreement here. Is sceptical doubt rooted in fundamental, context-invariant facts, like the supposed epistemological priority of experiential beliefs over beliefs about the world, so that it points to something about knowledge in general? Call the view that it is 'epistemological realism' and note that it is far from obvious that ordinary epistemic practices commit us to such a conception of justificational relations. Or does what it is reasonable or even possible to doubt depend on contextually variable factors, most notably the kind of inquiry in which we are involved, so that sceptical doubt may be an artefact of a certain conception of epistemological investigation, hence (again) not at all 'natural'? This much is sure: if doubt is strongly contextually constrained, it is not a merely psychological fact that doubts which seem so compelling in the study dissipate in the street.

Traditional epistemological theories such as foundationalism and the coherence theory take sceptical doubts at face value and try to provide reassurance (see Knowledge and justification, coherence theory of; Foundationalism). But some philosophers question the very intelligibility of universal doubt. This results from philosophy's having taken 'the linguistic turn'. Whereas language used to be regarded as a medium for the communication of thought, which was conceived to exist independently of language, the currently dominant view is that language is a *vehicle* of thought. This conception leads many philosophers to see an essential connection between truth and meaning: what our words mean - hence what thoughts we have - is a function of what sentences we hold true (see Charity, principle of). To entertain all the thoughts we currently have while universally doubting their truth may not be as straightforward as once was supposed. Such ideas mesh nicely with a contextualized conception of doubt and justification. *Anything* may be dubitable, but not *everything* at once. We can keep doubt as a spur to inquiry without worrying that it is in permanent danger of getting out of hand.

See also: Error and illusion, Indian conceptions of; Fallibilism; Rational beliefs; Scepticism

MICHAEL WILLIAMS

References and further reading

Burnyeat, M. (1983) 'Can the Sceptic Live his Scepticism?', in M. Burnyeat (ed.) *The Skeptical Tradition*, Berkeley, CA: University of California Press.(Argues that the Pyrrhonian ideal of a life without belief cannot be realized.)

Cavell, S. (1979) *The Claim of Reason*, Oxford: Oxford University Press.(Part 2 is especially relevant. Claims that although the sceptic seems to adapt ordinary doubting procedures, his doubts are less than fully intelligible.)

Clarke, T. (1972) 'The Legacy of Skepticism', *Journal of Philosophy* 69 (20): 754-69.(Influential paper exploring a distinction between 'plain' and 'philosophical' doubts.)

Descartes, R. (1641) *Meditations on First Philosophy*, in J. Cottingham, R. Stoothoff and D. Murdoch (eds) *The Philosophical Writings of Descartes*, Cambridge: Cambridge University Press, 1975, vol. 2.(The most important discussion of sceptical doubts in early modern philosophy.)

Dewey, J. (1929) *The Quest for Certainty: Gifford Lectures 1929*, New York: Capricorn, 1960.(Fundamental text of US pragmatism, tracing sceptical worries to a misguided 'spectator' conception of knowledge.)

Fogelin, R. (1995) *Pyrrhonian Reflections on Knowledge and Justification*, Princeton, NJ: Princeton University Press.(Argues that Pyrrhonian doubts admit of no theoretical resolution.)

Harman, G. (1973) *Thought*, Princeton, NJ: Princeton University Press.(Chapter 1 argues that 'refuting the

sceptic' is purely a matter of understanding where sceptical arguments go wrong.)
Hume, D. (1751) *Enquiries Concerning Human Understanding and the Principles of Morals*, ed. L.A. Selby-Bigge, 3rd edition revised by P.H. Nidditch, Oxford: Oxford University Press, 1975.(A pivotal text in the sceptical tradition, claiming that our most fundamental beliefs are groundless but unshakeable.)
Moore, G.E. (1925) 'A Defence of Common Sense', in *Philosophical Papers*, London: Allen & Unwin, 1959. (Much-discussed paper, though critics disagree sharply over its intent and significance.)
Nagel, T. (1995) *The View from Nowhere*, Oxford: Oxford University Press.(See especially chapter 10 where Nagel argues that sceptical doubts arise inevitably from our tendency to take an 'objective' view of our epistemic situation.)
Sextus Empiricus (*c*.200) *Outlines of Pyrrhonism (Sextus Empiricus Vol. 1)*, trans. R.G. Bury, London: Heinemann, 1933.(The most important source for ancient Greek scepticism.)
Strawson, P. (1985) *Skepticism and Naturalism: Some Varieties*, London: Methuen. (See especially chapter 1. Defends the view that our resistance to sceptical doubts is at bottom psychological because our most fundamental beliefs are not a matter of 'reason'.)
Stroud, B. (1984a) *The Significance of Philosophical Scepticism*, Oxford: Oxford University Press.(Important critical examination of the main anti-sceptical strategies.)
Stroud, B. (1984b) 'Skepticism and the Possibility of Knowledge', *Journal of Philosophy* 81 (10): 545-51.(Denies that sceptical problems reflect commitment to epistemological foundationalism.)
Williams, M. (1988) 'Scepticism without Theory', *Review of Metaphysics* 41 (3): 547-88.(Argues that the difference between 'Pyrrhonian' and 'Cartesian' scepticism has been widely misunderstood.)
Williams, M. (1992) *Unnatural Doubts*, Oxford: Blackwell; corrected edn, Princeton, NJ: Princeton University Press, 1996.(Claims that sceptical problems depend on an antecedent commitment to contentious philosophical ideas.)
Wittgenstein, L. (1969) *On Certainty*, Oxford, Blackwell.(Argues that doubt can arise only within contexts in which other beliefs are held certain: hence global doubt is impossible.)

Doxography

*Doxography is a term describing the method of recording opinions (*doxai*) of philosophers frequently employed by ancient Greek writers on philosophy. It can also refer to texts or passages consisting of such accounts. The ancient tradition of doxographical writing finds its origin in the dialectical method of Aristotle and Theophrastus. Later works by authors such as Aëtius and Arius Didymus record much valuable material on ancient philosophers, although usually with little analysis or argumentation. Doxographical passages are also found in other ancient philosophical works, usually as a prelude to the discussion of a theme.*

The term ’doxography’ is derived from the neologism *doxographus* introduced by the German scholar Herman Diels. Literally it means ‘writer of opinions (*doxai*)’. In his great work *Doxographi Graeci* (1879) Diels collected various ancient documents which summarize the philosophical doctrines of ancient philosophers (particularly in the area of natural philosophy). He meant the term to contrast with *biographus*, ‘writer of lives’, although in practice it is difficult to make an absolute distinction between the two kinds of writings (see Diogenes Laertius).

If we examine the documents collected by Diels and others more closely, we can see that at least three different kinds of writing are found:

(1) In the so-called ‘Placita’ literature the organizing principle is systematic. Groups of *placita* (views) or *doxai* (opinions) are collected on particular subjects, each attributed to a philosopher or school: for example, ‘Plato declared that the cosmos is unique. Democritus and Epicurus affirmed that there are infinite worlds’, etc. The most famous example of such a collection is by Aëtius, first reconstructed by Diels on the basis of a Pseudo-Plutarchean *epitome* and extracts preserved by the anthologist Johannes Stobaeus.
(2) In the so-called ‘On the Schools’ (*Peri haereseōn*) literature, the doxographical information is collected in order to present the thought of a school of thought (for example, the Stoa) or an important philosopher (for example, its founder Zeno). Good examples of such doxographies are found in the surviving fragments of Arius Didymus and in Diogenes Laertius.
(3) A third genre that is related to doxography is the so-called ‘Successions’ (*diadochai*) literature, in which the generations of philosophers are linked to each other by way of a teacher-pupil relation, both prior to and during the development of the philosophical schools in the Hellenistic period: for example, Anaximander was the pupil of Thales in the so-called Ionian succession. The best surviving examples are the works on the Academy and the Stoa found in the library of Philodemus at Herculaneum. Although in these works relatively little information is given on the doctrines of the philosophers, the principle of the succession is a feature shared with other doxographical accounts.

Apart from these specific doxographical writings, doxographies are frequently found in authors when they deal with particular philosophical topics. A famous example is found in Cicero’s work *On the Nature of the Gods* (I 18-41), which begins with a long doxographical survey of theological opinions inserted in a piece of Epicurean polemic (see Cicero §3). This reflects the origin of the practice, which, as Mansfeld (1992) has shown, is to be located in the dialectical method introduced by Aristotle. When embarking on the examination of a philosophical theme, Aristotle often first collects and examines the so-called *endoxa* (reputable opinions). For example at the beginning of *On the Soul* (I 2.403b20) he states: ‘For our study of soul it is necessary, when formulating the problems of which in our further advance we are to find the solutions, to summon the opinions (*doxai*) of our predecessors, so that we may profit by whatever is sound in their suggestions and avoid their errors’.

Aristotle’s practice was continued and extended by his colleague and successor Theophrastus (§8). Especially influential appears to have been his large collection of *Physical Opinions* in eighteen books, which gave a systematic presentation of the views of the philosophers up to Plato on natural philosophy. This lost work is thought to have been the source of much material in later doxographical works and accounts, although the precise contours of its influence have proved difficult to trace.

From its origins in the Peripatetic school, the doxographical method became widely used in ancient philosophical writings. It is particularly common in Patristic sources, since it gives a quick overview of pagan thinking. A defect of doxography is that in the course of time it tends to become rather schematic, eschewing both argument and

analysis, and concentrating chiefly on the juxtaposition of views or on a thin skeleton of doctrine. This lack of argumentation has given the doxographical tradition a bad name. Nevertheless it should be recognized that it has distinctive methods (especially involving the use of disjunction and diaeresis), which give us important insights into the way that ancient philosophers handled their own history. Moreover it has preserved a great amount of invaluable information on the views of philosophers which would otherwise have been lost.

In modern scholarship on ancient philosophy the term ‘doxography’ is also often used in a broader, less technical sense. In such cases it describes all forms of presentation of philosophers’ views by means of summaries without direct reference to the philosopher’s own words. Because this broader usage invites confusion with the doxographical tradition proper as described in the main body of this article, it is not to be recommended.

DAVID T. RUNIA

References and further reading

Diels, H. (1879) *Doxographi Graeci (Greek Doxographers)*, Berlin: de Gruyter; repr. 1976.(Main collection of doxographical writings; includes Aëtius, Arius Didymus and fragments of Theophrastus.)

Hahm, D.E. (1990) ‘The Ethical Doxography of Arius Didymus’, in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt Part II: Principate*, vol. 36.4: 2935-3055.(Fine analysis of an important doxographical source on Peripatetic and Stoic ethics.)

Lachenaud, G. (1993) ’Plutarque’, in *Œuvres morales*, vol. 12 (2): *Opinions des Philosophes*, Paris: Les Belles Lettres, 2nd edn.(Useful for the writings of Aëtius, although Diels’ reconstruction of the original, for all its imperfections, is still indispensable.)

Mansfeld, J. (1992) ’*Physikai doxai* and *Problēmata physica* from Aristotle to Aëtius (and Beyond)’, in W.W. Fortenbaugh and D. Gutas (eds) *Theophrastus: His Psychological, Doxographical and Scientific Writings*, Rutgers University Studies in the Classical Humanities 5, New Brunswick, NJ, and London: Transaction Publishers, 63-111.(Valuable research on the philosophical background of the doxographical method.)

Mansfeld, J. and Runia, D.T. (1996) *Aëtiana: The Method and Intellectual Context of a Doxographer*, volume 1: *The Sources, Philosophia Antiqua* 73, Leiden: Brill.(First volume of a thorough analysis of the Diels hypothesis and the entire doxographical tradition.)

Runia, D.T. (1989) ‘Xenophanes on the Moon: A *Doxographicum* in Aëtius’, *Phronesis* 34: 245-69.(Gives an example of how doxographical material needs to be analysed.)

Du Bois-Reymond, Emil (1818-96)

Emil du Bois-Reymond conducted pioneering research in electrophysiology which established him as a major figure in German science in the second half of the nineteenth century. His influence extended further through his more general writings in politics and philosophy of science, in which he argued for theoretical restraint and for the recognition that certain problems (for example, the mind-body problem) fell beyond scientific modes of inquiry and explanation.

Du Bois was born in Berlin and educated there in medicine and physiology. His distinctions in experimental science resulted in election to the Berlin Academy of Sciences; he was later honoured as its perpetual secretary from 1867. Holder of the chair of Physiology at Berlin from 1858 and co-editor of *Archiv für Anatomie* from 1857 to 1877, du Bois exercised great influence in the biological sciences throughout the second half of the nineteenth century.

Du Bois is of interest as a result of two distinguishable but interconnected projects. The first is that body of research and theory concerned with basic principles in neurophysiology. The second is his more general and accessible writings in philosophy of science. It was through his standing within the scientific community that these latter productions came to enjoy great authority.

At the start of his career, Du Bois was a researcher in Johannes Müller’s laboratory, as one of a group of young scientists who in a matter of years would become leading figures, namely Carl Ludwig,Ernst Brücke and Herman von Helmholtz. These four entered into something of a pact: they were to reject any theory in biology that was not grounded in the basic sciences of physics and chemistry. The graduation thesis du Bois had submitted to the Berlin faculty reported the results of his research on electric fish. In 1841, Müller passed along Carlo Matteucci’s recent ‘Note sur les phénomènes électrique des animaux’ (*C.R. Acad. Sci.*, Paris, 1841, (13): 540-), in which Matteucci had shown that electrical currents could be recorded between a point on the surface of a muscle and one deep within it. Taking this as his starting point, du Bois proceeded to investigate electrophysiological phenomena under a wide variety of conditions, publishing the results and their theoretical implications in his influential two-volume *Untersuchungen über Thierische Elektricität* (*On Animal Electricity*) (1848-9). In later years, du Bois criticized Matteucci’s methods and resented his failure to cite those of du Bois’ studies on whichMatteucci’s work depended. Du Bois explained his findings in terms of a wave of polarization created by what he called ‘electromotive molecules’ of positive and negative charge. He scrupulously avoided taking a firm position on the actual physical nature of these hypothetical entities, stating only that they should be understood as ‘foci of chemical change’. Subsequent developments in neurophysiology disconfirmed some of the details of du Bois’ theory, but his general perspective was redeemed.

As a statesman of science eager to establish its proper place within the larger contexts of culture and thought, du Bois contributed two books of considerable interest and importance. In his *Über die Grenzen des Naturerkennens (On the Limits of the Knowledge of Nature)* (1872) a modest and restrained positivism can be identified, the echo of the early pact barely audible. He acknowledges the inability of experimental science to settle issues of an essentially metaphysical nature, and is prepared to accept the possibility that even certain natural phenomena will never disclose their most fundamental principles. The mind-body problem illustrates this well. This problem, as he understood it, exemplifies the general problem of reconciling different ‘substances’. A distinguished neurophysiologist, du Bois fully appreciated the dependence of mental life on the integrity and functional capacities of the nervous system. But on the question of just how these functions do give rise to mental life - just how physicochemical processes influence, determine or regulate mental life - du Bois offered his famous one-word answer: *‘Ignorabimus!’* (‘We will be ignorant!’). It is interesting that du Bois would illustrate the general problem further by noting the (substantial) difference between matter and energy.

A later edition of this work, published in 1880, includes his *Die Sieben Weltrathsel (The Seven Riddles of the Universe)*. Here du Bois defends the overarching perspective of the scientist, which includes not only the commitment to disinterested and objective inquiry, but also the modesty of the fallibilist. In acknowledging at once the power and the limits of scientific knowledge, he preserved room for philosophy to bring its quite different methods of analysis and inquiry to bear upon vexed questions, even as he liberated science from the burden of

having to explain what in the end may well be inexplicable.

Du Bois also expressed himself publicly on matters of political and social importance. On the Franco-Prussian War, for example, his 1870 published address at the University of Berlin, *Über den Deutschen Krieg* (*The German War*), was widely cited and discussed.

See also: Science, 19th century philosophy of §6

DANIEL N. ROBINSON

List of works

Du Bois-Reymond, E. (1848-84) *Untersuchungen über Thierische Elektricität*, Berlin: G. Reimer, 2 vols; trans. *On Animal Electricity*, H.B. Jones (ed.), London, Churchill, 1852.(In this work du Bois takes early steps towards what would become the ionic theory of the formation and propagation of the neural impulse.)

Du Bois-Reymond, E. (1870) *Über den deutschen Krieg*, trans. G. Solling, *The German War*, Berlin, J. Sittenfeld, 1870.(Here, du Bois as citizen expresses his judgment of the Franco-Prussian War.)

Du Bois-Reymond, E. (1872) *Über die Grenzen des Naturerkennens (On the Limits of the Knowledge of Nature)*, Leipzig: Veit.(The limits and power of natural science are discussed with a calm and seasoned criticality.)

Du Bois-Reymond, E. (1880) *Über die Grenzen des Naturerkennens. Die sieben Welträthsel (On the Limits of the Knowledge of Nature; The Seven Riddles of the Universe)*, Leipzig: Veit.(One of the riddles, the connection between mind and body, is recognized as a source of enduring perplexity. '*Ignorabimus*!' - we will be ignorant.)

Du Bois-Reymond, E. (1883) *Sitzungsberichte der Königlich-preussischen Academie der Wissenschaften zu Berlin* 16; trans. L. Brunon, as *On Secondary Electromotive Phenomena in Muscles, Nerves and Electrical Organs*, in *Translation of Biological Memoirs*, Oxford: Oxford University Press, 161-225, 1887.(The electrical theory of nerve and muscle excitability is developed in this work and illustrated by way of experimental measurements.)

References and further reading

Cranefield, P.F. (ed.) (1982) *Two Great Scientists of the Nineteenth Century: Correspondence of Emil Du Bois-Reymond and Carl Ludwig*, trans. S.L. Ayed, Baltimore, MD: Johns Hopkins University Press.(This volume presents letters collected by Estelle Du Bois-Reymond and contains a foreword by Paul Cranefield as well as notes and indexes.)

Ducasse, Curt John (1881-1969)

Ducasse was a highly systematic philosopher and scarcely any field or topic escaped his attention. He criticized Hume's account of causality, advocated 'soft determinism' and developed an 'adverbial' realist account of our knowledge of the external world. He was a dualist on the mind-body relation, a 'progressive hedonist' in ethics, a defender of the 'will to believe', an expressionist in the philosophy of art, a scourge of art critics and a critic of theism. He also wrote on propositions, truth, signs, liberal education, linguistic metaphilosophy and paranormal phenomena. His influence on younger philosophers has been greatest, however, in the areas of causality, adverbial realism, progressive hedonism, the will to believe and aesthetics. Ducasse died in 1969, but his work remains significant, especially through his influence on Roderick Chisholm and Wilfred Sellars.

1 Life

Born in Angoulême, France, in 1881, Ducasse was educated at the Bordeaux Lycée, Abbotsholme School in England and, after emigrating to the USA, at the University of Washington and Harvard University. He received the Ph.D. from Harvard in 1912, having done his dissertation under the direction of Josiah Royce. He taught at University of Washington from 1912 to 1926, then at Brown University until 1958, when he retired. Ducasse helped establish the *Journal of Symbolic Logic*, and was the first president of the Association for Symbolic Logic. He was one of the clearest philosophical writers in the United States and always sought to avoid obscurity.

2 Causality and necessity

Ducasse was a perceptive critic of Hume's analysis of causality, according to which, objectively considered, 'cause' means only regularity among contiguous sequences of logically independent impressions: having observed without exception that one object follows another object, we thereby acquire a habit of mind, namely, the ' propensity, which custom produces to pass from an object to the idea of its usual attendant' (Hume 1739-40: 165). Ducasse was unimpressed with Hume's view. Two of his many criticisms are particularly effective. First, Hume's phenomenalism deprives him of the right to use the distinction between 'perceptions' and 'objects', without which his own rendition of the causal relation could not even be stated. Second, Hume's own position undermines his use of the concept of propensity. In Hume's world there can be only perceptions (impressions) which *are* but cannot *act*, while a propensity refers to what a thing or substance is capable of doing or is likely to do. Hence perception is too narrow a concept to sustain the ideas of propensity (see Hume, D. §2).

As an alternative Ducasse referred to the method of single difference: the strict experiment. He defined causation as the observable relation that holds between the three terms of a strict experiment - the complete state of affairs (S) wherein occur two and *only two* changes in a given time-period, one of which occurs immediately after and adjacent to the other, the second qualifying as proximate effect (E) of the first, the proximate cause. Both Hume and Ducasse, it would seem, confused the reasons we have for saying that x and y are causally related with the meaning of the term 'cause'. Hume saw the method of agreement as the primary way causal propositions are confirmed and so interpreted the meaning of 'cause' as 'constant conjunction'. On the other hand, Ducasse saw the method of single difference as the most reliable way causal propositions are known to be true and hence interpreted the meaning of 'cause' as 'the only change in S immediately prior to E'. It is clear that both methods under different conditions provide good reasons for saying that x is the cause of y, but if so they cannot also be given as definitions of the concept 'cause'.

For Ducasse, there are other necessities besides logical ones - the blow of a hammer physically necessitates breaking the glass, the death of John's wife psychologically necessitates his depression, and so on. His inclusive concept of necessity he labelled 'etiological necessity'. Though usually clear, Ducasse wrote ambiguously about this concept, sometimes portraying it as an ontological ingredient of the world and at other times denying it. On one hand, he wrote that when King Charles lost his head the swing of the axe *forced* his head to be severed from his body; that people directly perceive a bird perching on a branch to *cause* the branch to bend; and so on. Moreover he often referred to the distinction between agents and patients among the particulars of the worlds. On the other hand, he wrote that any 'power' analysis of causality would introduce a mysterious, ineffable element into causality and thus must be rejected. Ducasse had apparently caught something of the Humean virus when he least expected it. He certainly believed in the agent-patient view of causality, yet there was no possibility of

justifying this distinction unless he conceived some particulars to have the power to produce certain effects.

3 Adverbial realism

Ducasse's greatest contribution to philosophy is probably his 'adverbial' view of perceiving physical objects, a view developed by his student Roderick Chisholm and others (see Mental states, adverbial theory of §2). According to Ducasse, in everyday life, as we all agree, we directly perceive physical objects and some of their properties and we cannot avoid believing in their independent existence. Many philosophers, however, point out that we sometimes experience illusions and hallucinations and hence, at least in some cases, must be directly aware of our own sensations and impressions and are not apprehending physical objects. Further, since perception is the result of a long causal chain of physical and physiological stimulation which ends in the brain as a sensation, it would seem that we are *always* directly aware of our own sensations and impressions, and never of physical objects. Like Thomas Reid before him, Ducasse dismissed these arguments because they spawned an unsolvable dilemma: either the physical world becomes unknowable, or 'physical object' takes on a meaning radically different from what people ordinarily do in fact mean by the term.

It was to escape from this dilemma that Ducasse developed his adverbial view of sensation. He proposed that we reject sensations as entities and think of them instead as ways of sensing an object, expressed by adverbs, not nouns: 'I see bluely', not 'I see this blue colour'. Thus when a person sees something which looks blue, they are not directly aware of a blue sensation but are being appeared to bluely by an object. The adverbial view, Ducasse claimed, not only avoids the traditional dilemma; it also shows that it is impossible to state the sensationalist view without referring to physical objects as basic epistemic entities. How an object appears to people must make some reference to the physiological states of the observer and the physical conditions of observation; such reference to specific people and conditions is required *before* one can construct a substantive vocabulary of sensations. Ducasse also used the concept of variable appearings to counter the sensationalist argument from illusion and hallucination by accounting for them in terms of variability in the state of the observer and the physical conditions of observation.

4 Progressive hedonism

According to Ducasse various social contexts exemplify different moral codes, but none of the codes can itself be established as the right one and the others as wrong ones, because every code is equally tenable in its context, each being as logically arbitrary as the others. Ducasse called his view 'ethical liberalism' and, though it sounds like straightforward relativism, it is somewhat distinct from this traditional doctrine.

Since no ethical code can be proved to be the correct one, there is no way of establishing either egoism or altruism as the right attitude. Nevertheless there are other ways of changing attitudes than by argumentation; egoists may have their horizons broadened and come to empathize with other people. The point of a liberal education is precisely that of widening horizons by enlarging the range of ideas to be apprehended by students and by making them familiar with a wider range of people with whom they can learn to identify. According to Ducasse egoists have not been convinced by arguments in any shift from egoism to altruism; the point is that widening horizons and the knowledge of what it is possible to enjoy or care about have changed the egoist into a person who *in fact* cares about others. This view has come to be called Ducasse's 'progressive hedonism' (see Hedonism §1).

Progressive hedonism is a form of universal hedonism and hence, minus the relativism, is akin to utilitarianism (see Utilitarianism §5). Numerous criticisms have been made of utilitarianism, one of which is particularly relevant to Ducasse: a person may have to be dishonest in order to avoid starvation, but how is this situation compatible with progressive hedonism? Ducasse answered that, instead of saying that one should contribute to the general welfare, the progressive hedonist should recommend that one commit the smallest amount of evil commensurate with continued existence. Ducasse, then, became an early advocate of what has come to be called 'negative utilitarianism'.

5 The will to believe

What are the moral requirements in epistemology when there is insufficient evidence to decide an issue? Dickenson Miller (1942) argued that it is a moral duty to suspend judgment until further evidence becomes available. His friend Ducasse, however, believed that the only *moral* responsibility is to consult all available

evidence impartially; if the evidence is in balance, then one has the right, but not the duty, either to suspend judgment or to accept whatever alternative has life-sustaining value.

Ducasse insisted that this was the legitimate core of William James' article 'The Will to Believe' (1897), which Ducasse thought should have been called 'The Right to Believe'. Against Ducasse, Miller argued that believing without adequate evidence would relax one to the point of believing propositions for practical reasons when the evidence was sufficient to substantiate the denial of that belief. But Ducasse countered by pointing out an opposite spreading effect. He feared that Miller's 'sceptical' condemnation of belief where there is no preponderance of evidence might spread to areas where there is a supporting preponderance in the form of probabilities. Should we insist on near certainties and ignore acting on probabilities?

In later correspondence, however, Ducasse made the same mistake that James had made earlier and committed himself to the strong view that one *ought* to accept life-sustaining beliefs if the evidence is inconclusive - indeed, it is one's duty to do so. In Ducasse's words a 'fool's paradise' is preferable to a 'fool's hell'. He even suggested that in order to avoid the latter a person is justified in using psychological devises to strengthen commitment to the former. Miller was horrified at this suggestion, since such psychological conditioning, he claimed, would make it impossible to look for new evidence and evaluate it properly if encountered. Miller's criticism can be forcefully paraphrased thus: 'Is there a great difference between self-inflicted brainwashing and political brainwashing?'

6 Aesthetics

Ducasse formulated an expressionist theory of art according to which an artist's feelings or emotions, some quite subdued and nameless, are embodied in words, lines, colours, forms, sounds and motion. This self-expression of artists is objective in the sense that they create works of art which when contemplated yield back the feelings of which they are the attempted expression. This objectification of feelings must be consciously done and critically controlled to distinguish it from expressive activities such as sighing and yawning. The critical control usually involves much trial and error as artists experiment within their different media. The trial-and-error procedure helps to clarify and refine the feeling, and in this respect also gives to art creation an objective dimension. The feeling that a work of art expresses, then, is not identical with the original feeling of the artist but is a 'feeling image', a concept not clearly articulated by Ducasse but one which seems to be a recollection of a feeling rather than the feeling itself (see Artistic expression §5). There are many definitions of fine art which Ducasse flatly rejected, including Tolstoi's view that communication of feeling is an essential part of fine art (see Tolstoi, L.N.). For Ducasse it is no essential aspect of a work of art that consumers of it should have its feeling dimension communicated to them. Ducasse also separated art from the concept of beauty; for him art was wholly defined as objective self-expression. No doubt he established one sense of art which is independent of such traditional aesthetic concepts as beauty and communication; however, restricting the meaning of 'art' to objective self-expression seems as unnecessarily restrictive as it would be to limit the meaning of 'religion' to varieties of monotheism.

Outside academic circles Ducasse was well known for his criticism of art critics. They give biographical facts about artists, collate their pieces in chronological or stylistic orders and refer to previous artists who influenced them and those whom they subsequently influenced. Though this may be interesting discursive information, most of it does not help in forming artistic judgments. None the less Ducasse never denied that discursive scholarship can be an aid to appreciation and evaluation, if it serves as a guide for perceiving more acutely and for discriminating in greater detail. The proper role of the critic, Ducasse thought, is that of a guide in a foreign city who, on the basis of their thorough knowledge of the city, points out what onlookers might easily have missed on their own.

EDWARD H. MADDEN

List of works

Ducasse, C.J. (1924) *Causation and the Types of Necessity*, Seattle, WA: University of Washington Press.(The most complete and thorough presentation of Ducasse's non-Humean analysis of causality. Unlike Hume's constant conjunction view, Ducasse argues that propositions are singular in nature.)

Ducasse, C.J. (1930) *The Philosophy of Art*, New York: The Dial Press.(Ducasse's classic statement of his expressionist view of art in which a work of art embodies some subtle nuance of feeling that can be captured by

a viewer as well as recaptured by the artist. This and his other popular writings on art led to his election to the presidency of the American Society of Aesthetics.)

Ducasse, C.J. (1944) *Art, the Critics and You*, New York: Oskar Piest.(A popular presentation of his aesthetic theory, but more importantly a stinging criticism of art critics who think that scholarly information about works of art is necessary for appreciation of art.)

Ducasse, C.J. (1951) *Nature, Mind, and Death*, The 1949 Carus Lectures, La Salle, IL: Open Court.(Ducasse's most elaborate and profound book, touching, as it does, on all the major areas of his interest from technical epistemology and metaphysics through a host of other issues culminating in a section on paranormal phenomena.)

Ducasse, C.J. (1968) *Truth, Knowledge and Causation*, London: Routledge & Kegan Paul.(A republication of the most important articles written by Ducasse in epistemology and metaphysics. It is a fruitful book for anyone who wants the technical philosophy of which Ducasse was such a master.)

References and further reading

Dommeyer, F.C. (ed.) (1966) *Current Philosophical Issues: Essays in Honour of Curt John Ducasse*, Springfield, IL: Charles C. Thomas.(Includes the editor's charming biography of Ducasse and a thorough bibliography of his work, as well as essays related to Ducasse's philosophy by Alice Ambrose, Charles Baylis, Brand Blanshard, Roderick Chisholm, Marvin Farber, Charles Hartshorne, Morris Lazerowitz, E.H. Madden, A.I. Melden, H.H. Price and Vincent Tomas, among others. Many of Ducasse's articles and reviews are available only in the journals in which they originally appeared, and Dommeyer's bibliography of Ducasse's work is indispensable.)

Hare, P.H. and Madden, E.H. (1975) *Causing, Perceiving and Believing: An Examination of the Philosophy of C.J. Ducasse*, Dordrecht: Reidel.(A full-scale examination of Ducasse's life including bibliographical references to his own work and to the secondary literature. Contains a guide to Ducasse's unpublished writings in the Archives of Brown University, John May Library.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge and P. Nidditch, Oxford: Oxford University Press, 1975.(Hume argues here that the concept of natural necessity is a function of repeated associations that lead to a mental expectation of B when presented with A. In this way he eliminates power as a basic metaphysical category.)

James, W. (1897) 'The Will to Believe', in *The Will to Believe and Other Essays in Popular Philosophy*, Cambridge, MA: Harvard University Press.(The definitive edition of the works of William James.)

Miller, D. (1942) 'James's Doctrine of the "Right to Believe"', *Philosophical Review* 51.(Miller's most powerful attack on James's will to believe doctrine.)

Murphy, G. *et al.* (1970) 'A Tribute to C.J. Ducasse', *Journal of the American Society for Psychical Research* 64: 139-47.(Tributes to Ducasse by Brand Blanshard, C.D. Broad, Antony Flew and Gardner Murphy.)

Du Châtelet-Lomont, Gabrielle-Émilie (1706-49)

The name of this aristocratic woman, a true intellectual with a passion for mathematics, philosophy and science, is linked to that of Voltaire, whose life and interests she shared for fifteen years. In the rural retreat of Cirey they studied, disagreed, wrote and published. She was introduced to Newtonian science by Maupertuis and Voltaire but departed from their views, when she tried to combine Leibnizian metaphysics with Newton's empirically based science in her Institutions de physique (1740). She translated Newton's Principia from Latin into French and added a commentary on the system of the world (1745-9, published 1756).

Gabrielle-Émilie de Breteuil was born into an aristocratic family in Paris and showed an early interest in intellectual pursuits. As an adolescent she translated Virgil's *Aeneid* into French. Later she learned Italian and English. She developed a passion for astronomy and physics, and studied mathematics seriously in order to understand these sciences.

In 1725 she married Florent-Claude, Marquis du Châtelet and Comte de Lomont, by whom she had three children. The marquis, often away with his regiment, was tolerant of his wife's behaviour. She was able to combine an intellectual life with the worldly, social life of the aristocracy. Among her lovers were the Duc de Richelieu, the mathematician and early French Newtonian Pierre-Louis de Maupertuis, who taught her mathematics, and Voltaire, who, risking arrest after the publication of his *Lettres philosophiques* in 1734, took refuge at her family property at Cirey-sur-Blaise in Champagne. There they created a centre of learning and pleasure. They set up a well-equipped laboratory and performed experiments on light, fire, optics, and so forth. Among their visitors were the popularizer of Newton's optics, Francesco Algarotti, Maupertuis, Johann (II) Bernoulli and Samuel König, a Swiss mathematician and philosopher, who spent most of 1739 with the Cirey household and taught Madame du Châtelet mathematics and the metaphysics of Leibniz and Wolff.

At Cirey Voltaire composed his *Éléments de la philosophie de Newton* and *Traité de métaphysique* (published 1738 and 1784 respectively). When he submitted an essay on the nature of fire for the 1738 competition of the Academy of Sciences, her disagreement with his views led her to submit, without Voltaire's knowledge, her own entry. She especially criticized his material view of light and fire. For her, fire - which had no weight - justified Descartes's subtle matter, which Voltaire had consistently attacked. Although neither Voltaire nor Madame du Châtelet won, their entries were published by the Academy.

Madame du Châtelet's most lasting contribution to learning was her translation into French of Sir Isaac Newton's *Philosophiae Naturalis Principia Mathematica*, to which she devoted the last years of her life. It was published posthumously in 1756, seven years after her death. She also wrote an original work, her *Institutions de physique (Instruction in Physics)* (1740), a textbook written to introduce her son to science, in which she presented metaphysics as a necessary introduction to physics. She admitted her debt to Christian Wolff's *Ontologia* (1729) and to summaries by König of works by Wolff. The manuscript, in the Bibliothèque Nationale de France, shows she had prepared an earlier version before König's arrival which was not Leibnizian, except for the chapter on *vis viva* (kinetic energy). König converted her to the Leibniz-Wolff philosophy, for her the only satisfactory metaphysics. She rewrote the metaphysical chapters accordingly. Voltaire liked to believe she never was really convinced by this philosophy. However, she was very firm in her Leibnizian views. She had adopted the Leibnizian view of the force of moving objects and was reading Wolff's Latin works in 1738 before König's arrival (see Leibniz, G.W.; Wolff, C.).

Her disagreements with Voltaire led to a sort of public dialogue between them. She reviewed his *Éléments de la philosophie de Newton* in the *Journal des savants* (September 1738); he published an article on her *Dissertation sur la nature du feu* in the *Mercure de France* (June 1739) and a critical review of her *Institutions de physique* also in the *Mercure de France* (June 1741). Voltaire kept reaffirming his Newtonian views against her attacks. The Leibnizian principle of sufficient reason had caused her to abandon attraction at a distance, empty space and atoms, all part of Voltaire's Newtonianism.

While never denying that the world behaved according to Newtonian principles, she felt that science had to discover mechanical causes for all phenomena. She followed Leibniz and Wolff in adopting non-material monads,

which contained force, the sufficient reason for change in the material world. Force was conserved in a perfectly functioning universe. Voltaire remained very sceptical of metaphysics, and was particularly hostile to monads, since he was committed to attraction at a distance as the mysterious but divine force that kept the universe in operation.

Her stand on *vis viva* in the final chapter of her *Institutions* gave rise to a public dispute with Dortous de Mairan and indirectly with Voltaire. Was the force of moving bodies equal to the mass times the velocity squared, as Leibniz claimed, or to the mass times the simple velocity, the traditional view, supported aggressively by Mairan? Both Mme du Châtelet and Voltaire became involved in the dispute.

In her *Institutions* she attacked Mairan's view. Mairan, obviously piqued, about the time he became secretary of the Academy of Sciences (January 1741) replied in an open *Lettre à Madame la marquise Du Chastellet, sur la question des forces vives…*. She then published her *Réponse… à la lettre de M. de Mairan sur la question des forces vives* (1741), an attack both personal and insulting. Finally Voltaire submitted to the Academy his *Doutes sur la mesure des forces motrices* (1741), in support of Mairan, and attacking the Leibnizian view of his mistress, without mentioning her name.

Voltaire, like the periodicals of the period, admitted that Madame du Châtelet had given the French a clear, elegant account of the Leibniz-Wolff philosophy, which was little known in France.

Madame du Châtelet was also the author of a *Discours sur le bonheur*, written about 1747 and first published in 1779, and is considered the author of an unpublished manuscript of biblical criticism entitled *Examen de la Genèse*, in another area of interest she shared with Voltaire.

In 1748 Mme du Châtelet fell in love with the minor poet, Jean-François de Saint-Lambert, became pregnant by him at the age of 43 and died of childbed fever in Lunéville at the court of Lorraine.

See also: Newton, I.; Voltaire

ROBERT L. WALTERS

List of works

du Châtelet-Lomont, G.-É. (1739) *Dissertation sur la nature et la propagation du feu (Dissertation on the Nature and Propagation of Fire)*, in *Pièces qui ont été présentées à l'Académie des sciences, pour concourir au prix de l'année 1738 (Entries presented to the Academy of Sciences for the competition of 1738)*, Paris: Imprimerie royale.

du Châtelet-Lomont, G.-É. (1740) *Institutions de physique (Instruction in Physics)*, Paris: Prault fils; *Institutions physiques*, Amsterdam: Au dépens de la compagnie, 2nd edn, 1742; repr. of 2nd edn, Hildesheim: Olms, 1988.

du Châtelet-Lomont, G.-É. (1741) *Réponse de madame la marquise du Châtelet, à la lettre que M. de Mairan, secrétaire perpétuel de l'Académie des sciences, &c. lui a écrit le 18 février 1741 sur la question des forces vives (Reply from Mme du Châtelet to the letter of M. de Mairan, permanent secretary to the Academy of Sciences, of 18 February 1741, on the question of vis viva)*, Brussels: Foppens.

du Châtelet-Lomont, G.-É. (*c.*1747) *Discours sur le bonheur (Discourse on Happiness)*, ed. R. Mauzi, Paris: Les Belles Lettres, 1961. (First published 1779.)

Newton, I. (1745-9) *Philosophiae Naturalis Principia Mathematica (Mathematical Principles of Natural Philosophy)*, French trans. G.-É. du Châtelet-Lomont, *Principes mathématiques de la philosophie naturelle*, Paris: Desaint et Saillant, partial edn, 1756; complete edn, 1759; repr. Paris: Blanchard, 1966; repr. Paris: Jacques Gabay, 1990.(du Châtelet-Lomont's translation of Newton's *Principia*.)

References and further reading

Barber, W.H. (1955) *Leibniz in France from Arnauld to Voltaire*, Oxford: Clarendon Press.(A French enthusiast; Mme du Châtelet and the general French hostility to Leibniz.)

(1967) 'Mme du Châtelet and Leibnizianism: The Genesis of the *Institutions de physique*', in W.H. Barber, J.D. Brumfitt *et al.* (eds) *The Age of the Enlightenment, Studies Presented to Theodore Besterman*, Edinburgh: Oliver & Boyd, 200-22.

Janik, L.G. (1982) 'Searching for the Metaphysics of Science: The Structure and Composition of Madame du Châtelet's *Institutions de physique*, 1737-1740', *Studies on Voltaire and the Eighteenth Century*, 201: 85-113. (A thorough analysis of the manuscript in the Bibliothèque Nationale de France and the stages that led Mme du

Châtelet to present Leibnizian metaphysics as an introduction to science.)
Taton, R. (1969) 'Madame du Châtelet, traductrice de Newton', *Archives internationales d'histoire des sciences* 22: 185-210.
Vaillot, R. (1978) *Madame du Châtelet*, Paris: Michel.(A readable biography, which examines the various aspects of her life.)
Wade, I.O. (1947) *Studies on Voltaire with some unpublished papers of Mme du Châtelet*, Princeton, NJ: Princeton University Press.(Contains her translation of Mandeville's *Fable of the Bees*, her fragmentary *Essai sur l'optique*, and three chapters of a *Grammaire raisonnée*; analysis of these texts and Voltaire's *Traité de métaphysique*.)
(1969) *The Intellectual Development of Voltaire*, Princeton, NJ: Princeton University Press.(Part III, pages 253-570, examines the intellectual activity at Cirey.)

Dühring, Eugen Karl (1833-1921)

Versatile and prolific, Eugen Dühring constructed a metaphysical system uniting naturalism with a priori principles, such as a 'law of definite number' which asserts that everything countable must be finite; hence, the natural world must be limited, and past time must have a beginning. Value judgments are based on natural drives and feelings: in particular, the concept of injustice arises from the resentment produced by injury. Since criminal law is 'a public administration of revenge', the deterrent function of punishment is irrelevant to its rightness. In politics, Dühring combined his socialism with a fervent racism, chauvinism and anti-Semitism.

Dühring was born in Berlin, the only child of a civil servant. He studied law, but when an eye complaint led to complete blindness, shifted to philosophy and political economy. Dühring gained his doctorate in 1861 and became *Privatdozent* at the University of Berlin in 1863, but his campaign against 'spiritual corruption' in universities and his attacks on professors led to his dismissal in 1877. Quarrelsome and litigious, Dühring continued to feud with editors and publishers, produced books on many topics, and championed Robert Mayer's claim for priority in establishing the conservation of energy. Having outlived his enemies, he died at Nowawes, near Potsdam, in 1921.

Dühring constructed a realistic and naturalistic system which he called *Wirklichkeitsphilosophie* ('philosophy of reality'), based on 'principles', which are not really universal propositions but indications of the conditions for gaining knowledge. The most important task for metaphysics, Dühring claims, is to eliminate false conceptions of infinity. The only meaningful concept of infinity is that of a succession which never reaches an end: a completed infinity is an absurdity or 'chimaera', even a self-contradiction. Dühring's finitism is expressed in what he terms 'the law of definite number'. Everything that can be counted must be finite in number. Since this mathematical axiom must apply to reality, the number of world-bodies at any given moment is a particular determinate number.

The most striking consequence of this principle is a denial of the eternity of the world. Because an infinite series of past events would be a completed infinity, there must be a first state of the world, a timeless stability or original chaos, though how this gave rise to a succession of change is a problem. Only for future time is infinitude meaningful, as it is no more than the absence of any final state. Similarly, space can be thought of as infinite only negatively, as the unlimited possibility of going further; and infinite divisibility cannot be imagined as a process leading to infinitesimal quantities.

Dühring's moral philosophy is a version of naturalism. Since thinking by itself can never produce an 'ought' out of an 'is', he argues, all ideas of right and wrong must be based on drives and feelings. In particular, our notion of injustice arises out of the natural impulse to retaliate against those who have harmed us. The reactive feeling of *ressentiment*, as Dühring terms it, which finds expression in the drive for revenge, is not produced by some moral judgment made on other grounds; on the contrary, it is what supplies the material content needed for any concept of justice. Civilized society has taken over the task of acting against offenders, rather than leaving it to individual vengeance; yet however universal and impersonal our conceptions of right and wrong may become, their underlying impulse remains the same. Criminal justice, even in its ideal form, is a public administration of revenge. Hence the death penalty is appropriate for murder, whether or not it discourages further crime: the employment of legal punishment for deterrence is a 'terroristic system', having nothing to do with justice (see Crime and punishment §§1-2).

More broadly, Dühring proposes an alternative to both pessimism and optimism by surveying the feelings and emotions that give value to life. Injustice is the greatest evil in the world, while the greatest good is harmonious interaction with others: here Dühring shows the influence of Ludwig Feuerbach in emphasizing human sympathy, love and mutual respect. He condemns asceticism for devaluing the natural side of human nature, and separating the spiritual from the sensuous. Satisfaction comes from striving for real goals, and trusting in the value of life; but religion is harmful, since it diverts attention from actuality to an imaginary world.

As a political economist, Dühring followed Henry Carey in suggesting a 'harmony' of labour and capital, and advocating measures to improve the worker's condition. His own 'socialitarian' programme envisaged future society as a network of egalitarian communes. However, Dühring opposed the International and its 'Jewish

ringleader' Karl Marx. A fervent racist, he called for the exclusion of Jews and the elimination of 'asiatism' in all its forms - including Christianity - from German culture, while rejecting Darwinism as a brutal, immoral and typically English doctrine.

Ironically, Dühring's most important influences were negative ones. Interest in his ideas among German socialists impelled Karl Marx and Friedrich Engels to construct a comprehensive philosophy of their own. Encouraged by Marx, Engels subjected Dühring's system to a detailed critique in *Herr Eugen Dühring's Revolution in Science* (1878), commonly known as *Anti-Dühring*, which became an official text for later Marxism. Engels rejects Dühring's apriorism in favour of an empiricist and inductivist approach and, perhaps provoked by Dühring's attack on the Hegelian aspects ofMarx's *Kapital*, defends the 'laws' of dialectic as general descriptions of natural processes, even affirming the presence of contradictions in reality (see Engels, F.).

Another contemporary to oppose Dühring was Friedrich Nietzsche. In Nietzsche's view, Dühring is right in tracing our familiar concepts of good and evil back to vengefulness, but wrong in overlooking other drives which are spontaneous and active rather than reactive, embody strength rather than weakness, and belong to conquerors and rulers rather than the oppressed. Nietzsche proposes to establish new values which express this 'will to power'; and when he condemns socialism as a politics of *ressentiment* it is, rightly or wrongly, Dühring that he has in mind.

See also: Naturalism in ethics

ROBIN SMALL

List of works

Dühring, E. (1865a) *Capital und Arbeit (Capital and Labour)*, Berlin: A. Eichhoff. (Programme for a reformist socialism.)

Dühring, E. (1865b) *Carey's Umwälzung der Volkswirthschaftslehre und Socialwissenschaft (Carey's Revolution in Economic Theory and Social Science)*, Munich: E.A. Fleischmann.(Defence of the American writer's economic and political ideas.)

Dühring, E. (1865c) *Natürliche Dialektik (Natural Dialectic)*, Berlin: E.S. Mittler.(Theory of the principles of thinking.)

Dühring, E. (1865d) *Der Werth des Lebens (The Value of Life)*, Breslau: E. Trewendt. (Dühring's answer to pessimism.)

Dühring, E. (1869) *Kritische Geschichte der Philosophie von ihren Anfängen bis zur Gegenwart (Critical History of Philosophy from its Beginnings to the Present)*, Berlin: L. Heimann.(A survey from Thales toFeuerbach and Mill.)

Dühring, E. (1871) *Kritische Geschichte der Nationalökonomie und des Socialismus (Critical History of Political Economy and Socialism)*, Berlin: T. Grieben.(An attempted synthesis of economic theory and socialist practice.)

Dühring, E. (1873a) *Cursus der National- und Socialökonomie (Course of Political and Social Economy)*, Berlin: T. Grieben.(Systematic presentation of economic theory.)

Dühring, E. (1873b) *Kritische Geschichte der allgemeinen Principien der Mechanik (Critical History of the General Principles of Mechanics)*, Berlin, T. Grieben.(Interpretation of modern physics as based on axiomatic universal principles.)

Dühring, E. (1875) *Cursus der Philosophie (Course of Philosophy)*, Leipzig: E. Koschny. (Exposition of Dühring's metaphysical system.)

Dühring, E. (1880) *Robert Mayer der Galilei des neunzehnten Jahrhunderts (Robert Mayer, the Galileo of the Nineteenth Century)*, Chemnitz: E. Schmeitzner.(Biographical defence of the controversial scientist.)

Dühring, E. (1881) *Die Judenfrage als Racen-, Sitten- und Culturfrage (The Jewish Question as a Question of Race, Morality and Culture)*, Karlsruhe: H. Reuter.(A violently anti-Semitic polemic.)

Dühring, E. (1882) *Sache, Leben und Feinde (Cause, Life and Enemies)*, Karlsruhe: H. Reuter.(Dühring's autobiography.)

Dühring, E. (1883) *Der Ersatz der Religion durch Vollkommeneres (The Replacement of Religion by Something Higher)*, Karlsruhe: H. Reuter.(Proposes a 'spiritual leadership' more suited to the German race than Christianity.)

References and further reading

Binder, H. (1933) *Das sozialitäre System Eugen Dührings*, Jena: G. Fischer.(A study of Dühring's version of socialism.)

Bois, H. (1909) 'Le finitisme de Dühring', *L'Année philosophique* 20: 93-124. (Discusses Dühring's finitism.)

Engels, F. (1878) *Herrn Eugen Dühring's Umwälzung der Wissenschaft*, Leipzig: Genossenschaftsdruckerei; 3rd edn, 1894, trans. E. Burns *Herr Eugen Dühring's Revolution in Science*, in F. Engels and K. Marx, *Collected Works* vol. 25, London: Lawrence & Wishart, 1987.(Engels' polemic against Dühring.)

Small, R. (1990) 'Nietzsche, Dühring, and Time', *Journal of the History of Philosophy* 28: 229-50.(Analysis of debate about infinite time.)

Vaihinger, H. (1876) *Hartmann, Dühring und Lange*, Iserlohn: J. Baedeker.(A balanced contemporary assessment.)

Venturelli, A. (1986) 'Asketismus und Wille zur Macht. Nietzsches Auseinandersetzung mit Eugen Dühring', *Nietzsche-Studien* 15: 107-39.(Surveys Dühring's influence onNietzsche.)

Duhem, Pierre Maurice Marie (1861-1916)

Duhem was a French Catholic physicist, historian of science and philosopher of science. Champion of a programme of generalized thermodynamics as a unifying framework for physical science, he was a pioneer in the history of medieval and renaissance science, where he emphasized a continuity between medieval and early modern science. Duhem was also one of the most influential philosophers of science of his day, thanks to his opposition to mechanistic modes of explanation and his development of a holistic conception of scientific theories, according to which individual empirical propositions are not tested in isolation but only in conjunction with other theoretical claims and associated auxiliary hypotheses. Such a view of theory testing entails that there are no 'crucial experiments' deciding unambiguously for or against a given theory and that empirical evidence therefore underdetermines theory choice. Theory choice is thus partly a matter of convention. Duhem's conventionalism is similar in kind to that later advocated by Otto Neurath and by W.V. Quine.

1 Life

Born in Paris and educated at the École Normale Supérieure, Duhem took his doctorate (in mathematics) from the Sorbonne in 1888 with a thesis on the theory of magnetization by induction. He had made a first attempt at the doctorate (in physics) in 1884, while only a third-year student at the École Normale, with a thesis on the concept of the 'thermodynamic potential'. This concept was later to become the centrepiece of most of his work in physics. Although published two years later as Duhem's first book, the 1884 thesis was rejected mainly due to the opposition of Marcelin Berthelot, whose principle of maximum work - which wrongly asserts that chemical changes tend spontaneously to produce maximum heat - was criticized by Duhem. Undaunted by this rejection, Duhem took first place in the *concours d'agrégation*, the competitive examination for university teaching positions in France, in 1885. Two years later he began work as a lecturer in physics at Lille, where he met his wife, Marie-Adèle Chayet, whom he married in late October 1890 in Paris. A daughter, Hélenè, was born in September 1891, but the great tragedy of Duhem's life occurred in July 1892, when Marie-Adèle died after giving birth prematurely to a second daughter. Duhem never remarried, keeping household in later years with his devoted daughter and future biographer, Hélenè, who also arranged for the posthumous publication of the last five volumes of Duhem's ten-volume historical magnum opus, *Le système du monde*.

Duhem's conservative, Catholic orientation, his stubborn and contentious manner and the continuing opposition of politically influential, anti-clerical, senior French scientists like Berthelot all combined to prevent his ever winning an academic position in Paris. In 1893, he left Lille for Rennes, moving the following year to Bordeaux, where he remained until his death in 1916. Duhem deeply resented this lack of recognition. When asked whether he would be interested in a newly created chair in the history of science at the Collége de France in 1893, he replied that he would go to Paris as a physicist or not at all, but he was pleased when, in 1913, he was finally elected a non-resident member of the Académie des Sciences.

Few thinkers can match Duhem's breadth and versatility, for he made contributions of fundamental importance in three different areas - thermodynamics and physics, philosophy of science and history of science. His work in these areas divided roughly into three overlapping phases. First, from the mid-1880s to 1900, thermodynamics and physics were his dominant interest, but with treatises on elasticity and energetics still being published as late as 1906 and 1911, respectively. In 1892, Duhem began publishing a series of essays on questions of methodology that would culminate in the publication in book form of his most influential work in the philosophy of science, *La théorie physique. Son objet et sa structure* in 1906. It was also in the mid-1890s that Duhem published his first essays in the history of science, starting on the path that would lead him to such important historical works as his *Études sur Léonard de Vinci* (1906-13) and *Le système du monde* (1913-59).

A curious denouement to Duhem's prodigious literary outpouring was his publication a few months after the outbreak of the First World War, and shortly before his death, of a series of lectures under the title *La science allemande*. He regarded this work as his patriotic contribution to the French war effort. Drawing upon Pascal's distinction between *l'esprit géométrique* and *l'esprit d'finesse* (see Pascal, B. §5), Duhem excoriated what he characterized as the axiomatizing, deductive excess of the German scientific intellect, contrasting this with the healthy good sense (*bon sens*) displayed by the French mind in its choice of first principles. This good sense is

alleged to provide French thinkers with an anchor in intuitively apprehended truth not to be found in German science. However much one might lament this dubious excursion into national scientific typology, *La science allemande* can and should also be read as an intellectual autobiography, as Duhem's mature statement about his own preferred way of doing science.

2 Physics

Duhem's most important contributions to physics were in the areas of thermodynamics and physical chemistry (see Thermodynamics; Chemistry, philosophical aspects of §3). Indeed his entire scientific programme was driven by the conviction that a generalized thermodynamics would be the proper unifying scheme for physical theory, with all of chemistry and physics, including mechanics, electricity and magnetism, being derivable from thermodynamic first principles. The intellectual style evinced in this project could be characterized as a balanced combination of the abstract, axiomatizing geometric spirit and a *bon sens* not too rigidly controlled by empirical constraints in the choice of axioms, along with a total unconcern for mechanical models and a thoroughgoing scepticism about any underlying atomistic ontology.

Strongly influenced by the work of J.W. Gibbs and H. von Helmholtz, Duhem started from the concept of the 'thermodynamic potential', related to the Gibbs and Helmholtz free energies, which he deployed in a manner formally analogous to potentials in mechanics for the purpose of representing all physical and chemical changes. The programme, which finds its mature statement in Duhem's *Traité d'énergétique* of 1911, was not ultimately successful, but along the way Duhem made a number of enduring contributions to thermodynamics and physical chemistry. Among these were the Duhem-Margules and Gibbs-Duhem equations, the first clear definition of a reversible process in thermodynamics as a quasi-static limiting process, and the first general proof of the Gibbs phase rule. That these results were obtained in the context of a programme of generalized thermodynamics was sufficient to earn Duhem's work a warm reception among late-nineteenth-century German energeticists, such as Wilhelm Ostwald and Georg Helm. Duhem was also one of the first continental scientists to appreciate the work of Gibbs, penning the earliest critical examination of Gibbs' 'On the Equilibrium of Heterogenous Substances' (1876, 1878) in 1887 and promoting the French translation of Gibbs' works with his pamphlet, *Josiah Willard Gibbs, à propos de la publication de ses mémoires scientifiques* (1907).

3 History of science

In the mid-1890s Duhem began publishing some shorter papers on the history of mechanics and thermodynamics in the *Revue des deux mondes* and the *Revue des questions scientifiques*. In 1902 he published two major historical studies, *Le mixte et la combinaison chimique. Essai sur l'évolution d'une idée* and *Les théories électriques de J. Clerk Maxwell. Étude historique et critique*, the first of a series of such works that was also to include *L'évolution de la mécanique* (1903a), *Les origines de la statique. Les sources des théories physiques* (1905-6), *Études sur Léonard de Vinci, ceux qu'il a lus et ceux qui l'ont lu* (1906-13), and his greatest historical work, the ten-volume *Le système du monde. Histoire des doctrines cosmologiques de Platon à Copernic* (1913-59).

It was in the history of medieval and Renaissance science that Duhem made his most important contribution. Here his name is associated with the thesis that most of the important innovations of early modern science had roots in the medieval period sufficient to warrant our viewing the science of that later period as a continuous outgrowth of the earlier one and not as a radical, revolutionary break with a scholastic dark age, as had standardly been assumed by previous historians of science, such as William Whewell.

As late as the time of his 1903 *L'évolution de la mécanique*, Duhem still subscribed to the dominant view of the Middle Ages as a fallow period in the history of science between the Greeks and the revival of science in the late Renaissance. It was in the course of his work on his *Origines de la statique* (1905-6) that he first came to suspect otherwise, when his tracking down a few references to the then all but unknown Jordanus de Nemore led to his discovery of a rich medieval tradition, centred at the University of Paris, that was known to and influenced later authors. Most importantly, in his *Études sur Léonard de Vinci* (1906-13) and *Le système du monde* (1913-59), Duhem began to reconstruct the neglected science of the Middle Ages, including the work of such figures as Jean Buridan, Albert of Saxony, and Nicole Oresme, attempting to document the manner in which the pivotal contributions of Leonardo, in particular, drew upon these earlier thinkers.

A theological motivation was never far from the surface in Duhem's historical writings. In addition to seeking to undercut the popular anti-clerical modernism of the Third Republic by disputing claims to revolutionary novelty on behalf of modern science, Duhem also sought to reveal the role in intellectual history of divine providence in the form of what he termed an 'ideè directrice' leading science forward to a 'natural classification' as the endpoint of inquiry.

4 Philosophy of science

Duhem's philosophy of science is a species of conventionalism (see Conventionalism). It was first developed in a series of essays appearing in the early to mid-1890s in the *Revue des questions scientifiques*, but is best known from the presentation in Duhem's *La théorie physique. Son objet et sa structure*. This first appeared in 1904 and 1905 as a series of articles in the *Revue de philosophie*, and then in book form in 1906. The essential content of Duhem's position is also to be found adumbrated in his *ΣΩΖΕΙΝ ΤΑ ΦΑΙΝΟΜΕΝΑ, Essai sur la notion de théorie physique, de Platon á Galilée* (1908).

According to Duhem: 'A physical theory is not an explanation. It is a system of mathematical propositions, deduced from a small number of principles, which aim to represent as simply, as completely, and as accurately as possible a set of experimental laws'. Moreover, from a purely logical point of view, there will always be a multiplicity of different physical theories equally capable of representing a given set of experimental laws. The argument for this conclusion starts from the doctrine of epistemological holism, according to which a scientific hypothesis is tested not in isolation, but only as part of an entire body of scientific theory (see Theories, scientific). Duhem exemplified this through his critique of the Newtonian idea of the *experimentum crucis*, an experiment the outcome of which would, by itself, determine unambiguously the truth or falsity of a given scientific hypothesis. There can be no crucial experiment, according to Duhem, because in order to derive a testable prediction from a hypothesis we must draw upon a set of auxiliary assumptions concerning such matters as the functioning of the experimental apparatus and the absence of extraneous physical influences (see Crucial experiments). It follows that a negative experimental result could be explained by rejecting either the hypothesis in question or one or more of the auxiliary assumptions. Hence there will always be a multiplicity of empirically-equivalent total theories for representing any set of experimental laws, meaning that theory choice is underdetermined by empirical evidence, and that the community's choice of one among these alternative theoretical representations will have the status of a convention (see Underdetermination §3).

Nevertheless, there is progress in science, the history of a field like optics or mechanics revealing continuous convergence towards what Duhem terms a 'natural classification'. An analysis of scientific method proves that science does not provide explanations; it is an 'act of faith' that convinces us that such an order in our theoretical representations corresponds to the real order in nature. Duhem quotes his favourite philosopher, Pascal: 'We have an impotence to prove, which cannot be conquered by any dogmatism; we have an idea of truth which cannot be conquered by any Pyrrhonian scepticism'.

Here lies the explanation for the generally anti-metaphysical tendency of Duhem's thinking about scientific method. Thus, he criticizes what he alleges to be a typically English penchant for seeking mechanical models of physical phemonena, as illustrated by Maxwell's search for mechanical models of the electromagnetic ether, arguing that the search for models is as often as not an impediment to the progress of scientific inquiry (see Electrodynamics; Models §3). By contrast, he commends what he regards as the typically French geometric or axiomatic/formal approach to science, as exemplified in his own work in thermodynamics and other areas.

Duhem's avowed purpose in denying to empirical science the capacity to determine unambiguously the deepest truth about nature was thereby to reserve a place for theology as the arbiter of truth in metaphysics. This emerges with special clarity from his 1905 essay, 'Physique de Croyant', which he added as an appendix to the second edition of *La théorie physique*. But whatever Duhem's conservative, Thomistic, theological aims might have been, it was the anti-metaphysical bent of his philosophy of science that was prized by many of his postivistically inclined scientific and philosophical contemporaries, from Édouard Le Roy, who characterized Duhem's philosophy as an instance of the 'new positivism', to Ernst Mach, who was so taken by *La théorie physique* that he promoted its translation into German in 1908.

Duhem's philosophy of science differs in crucial respects from the conventionalism of H.J. Poincaré. Most

importantly, Duhem did not follow Poincaré in characterizing conventions in science as merely 'apparent hypotheses' that are, in fact, 'disguised definitions', to be distinguished from genuine empirically verifiable hypotheses. Duhem's holism entails that all of the propositions constituting a body of scientific theory are on an epistemic par, any possible differences in susceptibility to revision in the face of new evidence being a matter, not of principle - as in the difference between analytic definitions and synthetic empirical propositions - but rather a contingent, practical matter of the community's historically and sociologically conditioned degree of commitment to different pieces of the total body of scientific theory. This difference between the conventionalism of Poincaré and that of Duhem is reflected in the different varieties of conventionalism that grew up in later twentieth-century philosophy of science. Moritz Schlick and Hans Reichenbach adopted an essentially Poincaréan conception of the conventionality of coordinating definitions, while Otto Neurath, Albert Einstein and W.V. Quine adopted Duhem's holism, the latter two thinkers extending holism about the empirical content of theories to a holism about the meanings of individual terms in science.

See also: Crucial experiments; Logical positivism; Scientific method

DON HOWARD

List of works

Duhem, P.M.M. (1886) *Le potential thermodynamique et ses applications à la mécanique chimique et à la étude des phénomènes électriques (The Thermodynamic Potential and Its Applications to Mechanics, Chemistry and the Study of Electrical Phenomena)*, Paris: A. Hermann.(Duhem's dissertation.)

Duhem, P.M.M. (1892a) 'Quelques réflexions au sujet des théories physiques' (Some Reflections on the Subject of Physical Theories), *Revue des questions scientifique*, 2nd series, 31: 139-77.(Duhem's first paper on the philosophy of science.)

Duhem, P.M.M. (1892b) 'Notation atomique et hypothèse atomistique' (Atomic Notation and the Atomistic Hypothesis), *Revue des questions scientifique*, 2nd series, 31: 391-454.

Duhem, P.M.M. (1893a) 'Physique et métaphysique' (Physics and Metaphysics), *Revue des questions scientifique*, 2nd series, 34: 55-83. (On the difference between the methods and objects of physics and metaphysics.)

Duhem, P.M.M. (1893b) 'L'école anglaise et les théories physiques, à propos d'un livre de W. Thomson' (The English School and Physical Theories: apropos a Book by W. Thomson), *Revue des questions scientifique*, 2nd series, 34: 345-78.(Critique of mechanistic explanation.)

Duhem, P.M.M. (1894) 'Quelques réflexions au sujet de la physique expérimentale' (Some Reflections on the Subject of Experimental Physics), *Revue des questions scientifique*, 2nd series, 36: 179-229.(The first clear statement of Duhem's underdeterminationism.)

Duhem, P.M.M. (1895a) 'Les théories de la chaleur. I. Les précurseurs de la thermodynamique' (Theories of Heat. I. The Precursors of Thermodynamics), *Revue des deux mondes* 129: 869-901.(On the history of thermodynamics.)

Duhem, P.M.M. (1895b) 'Les théories de la chaleur. II. Les créateurs de la thermodynamique' (Theories of Heat. II. The Creators of Thermodynamics), *Revue des deux mondes* 130: 379-415.(Continuation of 1895a.)

Duhem, P.M.M. (1895c) 'Les théories de la chaleur. III. Chaleur et movement' (Theories of Heat. III. Heat and Motion), *Revue des deux mondes* 130: 851-68.(Continuation of 1895a.)

Duhem, P.M.M. (1896) 'L'évolution des théories physiques du XVII^e siècle jusqu'à nos jours' (The Evolution of Physical Theories from the Seventeenth Century to Our Day), *Revue des questions scientifique*, 2nd series, 40: 463-99.(On the history of physical theory in the modern period.)

Duhem, P.M.M. (1902a) *Le mixte et la combinaison chimique. Essai sur l'évolution d'une idée (Mixture and Chemical Combination: An Essay on the Evolution of an Idea)*, Paris: C. Naud; repr. Paris: Fayard, 1985.(On the history of the idea of chemical combination.)

Duhem, P.M.M. (1902b) *Les théories électriques de J. Clerk Maxwell. Étude historique et critique (The Electrical Theories of J. Clerk Maxwell: A Historical and Critical Study)*, Paris: A. Hermann.(Critial history of Maxwell's work on electrodynamics.)

Duhem, P.M.M. (1903a) *L'évolution de la mécanique*, Paris: A. Joanin; German trans. P. Frank *Die Wandlungen der Mechanik und der mechanischen Naturerklärung*, Leipzig: J.A. Barth, 1912; trans. M. Cole *The Evolution of Mechanics, Monographs, and Textbooks on Mechanics of Solids and Fluids*, Alphen aan den Rijn, The Netherlands and Germantown, MD: Sijthoff & Noordhoff, 1980.(Duhem's major study of the history of mechanics.)

Duhem, P.M.M. (1903b) 'Analyse de l'ouvrage de Ernst Mach: *Le mécanique. Étude historique et critique de son*

développement' (Analysis of the Ernst Mach's Work: Le mécanique. Étude historique et critique de son développement), *Bulletin des sciences mathématiques*, 2nd series, 27: 261-83.(Review of the French translation of Mach's *Science of Mechanics*.)

Duhem, P.M.M. (1905) 'Physique de croyant', *Annales de philosophie chrétienne* 155: 44-67, 133-159; repr. in *La théorie physique*, 2nd edn, 413-47; trans. as 'Physics of a Believer', included in *La théorie physique*, 2nd edn, 1914, 273-311.

Duhem, P.M.M. (1905-6) *Les origines de la statique. Les sources des théories physiques*, Paris: A. Hermann, 2 vols; trans. G.F. Leneaux, V.N. Vagliente and G.H. Wagener, *The Origins of Statics: The Sources of Physical Theory*, Boston and Dordrecht: Kluwer, 1991.(On the history of statistics.)

Duhem, P.M.M. (1906a) *La théorie physique. Son objet et sa structure*, Paris: Chevalier et Rivière; 2nd edn, 1914; repr. Paris: Vrin, 1989; German trans. F. Adler, intro. by E. Mach *Ziel und Struktur der physikalischen Theorien*, Leipzig: J.A. Barth, 1908; trans. P.P. Wiener, *The Aim and Structure of Physical Theory*, Princeton, NJ: Princeton University Press, 1954.(Originally published in the *Revue de philosophie* 4, 1904; 5, 1904; and 6, 1905. Duhem's major essay on the philosophy of science.)

Duhem, P.M.M. (1906b) *Recherches sur l'élasticité (Investigations on Elasticity)*, Paris: Gauthier-Villars.(One of Duhem's last major scientific works.)

Duhem, P.M.M. (1906-13) *Études sur Léonard de Vinci, ceux qu'il a lus et ceux qui l'ont lu (Studies on Leonardo da Vinci: Those Whom He Had Read and Those Who Have Read Him)*, Paris: A. Hermann, 3 vols.(Historical study of Leonardo da Vinci.)

Duhem, P.M.M. (1907) *Josiah Willard Gibbs, à propos de la publication de ses mémoires scientifiques (Josiah Willard Gibbs: On the Occasion of the Publication of His Scientific Writings)*, Paris: A. Hermann.(Duhem's brief on behalf of Gibb's approach to thermodynamics.)

Duhem, P.M.M. (1908) ΣΩ*ZEIN TA* Φ*AINOMENA, Essai sur la notion de théorie physique, de Platon á Galilée*, Paris: A. Hermann; repr. Paris: Vrin, 1990; trans. E. Dolan and C. Maschier, *To Save the Phenomena: An Essay on the Idea of Physical Theory from Plato to Galileo*, Chicago, IL: University of Chicago Press, 1969.(Briefer historical statement of Duhem's philosophy of science.)

Duhem, P.M.M. (1911) *Traité d'énergétique ou de thermodynamique générale (Treatise on Energetics or General Thermodynamics)*, Paris: Gauthier-Villars, 2 vols.(Late statement of Duhem's programme of generalized thermodynamics.)

Duhem, P.M.M. (1913-59) *Le système du monde. Histoire des doctrines cosmologiques de Platon à Copernic (The System of the World: A History of Cosmological Doctrines from Plato to Copernicus)*, Paris: A. Hermann, 10 vols; selected portions, mainly from vol. 7, trans. R. Ariew *Medieval Cosmology*, Chicago, IL: University of Chicago Press, 1985. (Duhem's most important study of the history of ancient and medieval cosmology.)

Duhem, P.M.M. (1915) *La science allemande*, Paris: A. Hermann; trans. J. Lyon, *German Science*, La Salle, IL: Open Court, 1991.(Wartime critique of the German way of doing science.)

Duhem, P.M.M. (1996) *Essays in the History and Philosophy of Science*, trans. R. Ariew and P. Barker, Indianapolis, IN and Cambridge: Hackett.(A collection of essays in translation, includes 1892a, 1893a, 1893b, 1894, 1903b and selections from 1908.)

References and further reading

Ariew, R. and Barker, P. (eds) (1990) 'Pierre Duhem: Historian and Philosopher of Science', special issue of *Synthèse* 83 (2-3).(A very rich collection representative of the best recent work on Duhem's history and philosophy of science.)

Brenner, A. (1990) *Duhem: Science, réalité et apparence. La relation entre philosophie et histoire dans l'œuvre de Pierre Duhem (Duhem: Science, Reality and Appearance. The Relation between Philosophy and History in the Work of Pierre Duhem)*, Paris: Vrin.(A reliable and accessible recent critical study that stresses the systematic unity of Duhem's work.)

Brouzeng, P. (1987) *Duhem, 1861-1916. Science et providence (Duhem, 1861-1916: Science and Providence)*, Paris: Belin.(A recent biographical study.)

Diederich, W. (1974) *Konventionalität in der Physik. Wissenschaftstheoretische Untersuchungen zum Konventionalismus (Conventionality in Physics: Investigations in the Philosophy of Science Concerning on Conventionalism)*, Berlin: Duncker & Humblot.(A little-known but still very helpful study of the conventionalist tradition, emphasizing Poincaré and Duhem and their reception by logical empiricists.)

Duhem, H.-P. (1936) *Un savant français: Pierre Duhem (A French Scholar: Pierre Duhem)*, Paris: Librairie Plon. (Biography by Duhem's daughter.)

Jaki, S.L. (1987) *Uneasy Genius: The Life and Work of Pierre Duhem*, Dordrecht: Nijhoff.(The most comprehensive recent biography; also contains the most complete bibliography.)

Lowinger, A. (1941) *The Methodology of Pierre Duhem*, New York: Columbia University Press.(The first study to introduce Duhem's philosophy of science to an American audience.)

Maiocchi, R. (1985) *Chimica e Filosofia, Scienza, Epistemologia, Storia e Religione nell' Opera di Pierre Duhem (Chemistry and Philosophy, Science, Epistemology, History and Religion in the Work of Pierre Duhem)*, Florence: La Nuova Italia Editrice.(A provocative reading of Duhem's philosophy of science as aiming to defend the place of theory in science rather than a statement of instrumentalistic conventionalism.)

Martin, R.N.D. (1991) *Pierre Duhem: Philosophy and History in the Work of a Believing Physicist*, La Salle, IL: Open Court.(A very helpful study that pays special attention to the religious side of Duhem's thinking.)

Picard, E. (1921) *La vie et l'oeuvre de Pierre Duhem (The Life and the Work of Pierre Duhem)*, Paris: Gauthier-Villars.(One of the earliest biographies.)

Dummett, Michael Anthony Eardley (1925-)

For Michael Dummett, the core of philosophy lies in the theory of meaning. His exploration of meaning begins with the model proposed by Gottlob Frege, of whose work Dummett is a prime expositor. A central feature of that model is that the sense (content) of a sentence is given by a condition for its truth, displayed as deriving from its constituent structure. If sense so explicated is to explain linguistic practice, knowledge of these truth-conditions must be attributed to language users by identifying features of use in which it is manifested. Analysis of truth suggests we seek such manifestation in patterns of assertion. But scrutiny of those patterns shows that there is no distinction between use which manifests knowledge of classical truth-conditions, and use which manifests knowledge of a weaker kind of truth - for example, one which holds whenever we possess a potential warrant for a statement.

Such considerations motivate reconstruing sense as given by conditions for this weaker kind of truth. But rejigging Fregean semantics in line with such a conception is highly nontrivial. Mathematical intuitionism, properly construed, gives us models for doing so with mathematical language; Dummett's programme is to extend such work to everyday discourses. Since he further argues that realism consists in defending the classical semantics for a discourse, this programme amounts to probing the viability of antirealism about such things as the material world, other minds and past events.

1 Sense

Michael Dummett, English logician and philosopher of language, was educated at Oxford, where he spent almost all his academic career, and occupied the Wykeham Chair of Logic from 1979 to 1992. His numerous papers in books and journals began appearing in the early 1950s; major longer works stretch from *Frege: Philosophy of Language* (1973) to *The Origins of Analytical Philosophy* (1993a).

While at odds with most Oxford philosophers of his own generation and of its predecessor in the importance he attaches to formal techniques, Dummett preserves a fundamental Oxford perspective by insisting on the centrality of the theory of meaning. Pure philosophy, as an a priori discipline, he argues, can do no more than clarify thought. But we cannot take our access to thoughts as unproblematic. (Ignoring this is the error of 'code' conceptions of language, which in Dummett's view include that of H.P. Grice (§§2-3); for these take such access for granted, and then try to explain meaning by correlating utterances with thoughts as thus (mis)conceived.) For we need to explain the compositional nature of thoughts, and so to display them as structures of thought-elements (concepts). There is, however, no mental process consisting of the bare apprehension of a concept. Rather, thought for us requires a vehicle; and our only access to the structure of thought is through the structure of some such vehicle. Further, though images and so on may serve as vehicles for simple thoughts, language is the paradigmatic thought-vehicle, and essential for sophisticated thought. So a theory of thoughts is best based on a theory of a language apt to express them.

Dummett derives his guiding conception of such a theory of language from Frege, in a profound interpretation of the work of that philosopher, to the elaboration and defence of which many of his publications are devoted. This paradigm puts a theory of sense at the core of the theory of meaning for a language, charged with articulating how sentences express thoughts; it supplements this with a theory of force, which spells out the conventions enabling utterances to be classified as assertions, orders, questions and so on. The dichotomy, Dummett argues, is an essential one; the later Wittgenstein repudiated it at the cost of neglecting the systematicity of language (see Wittgenstein, L.J.J. §9).

The core theory of sense is founded on a syntax and a semantic theory. The former reads into each sentence of the language a structure out of finitely-many primitive syntactic elements. Then, relative to a given domain of objects, the semantic theory specifies, for each kind of primitive expression the syntax discerns, a sort of entity apt to count as the semantic value of expressions of that kind. Finally, rules of the theory explain how complex expressions, and in particular sentences, systematically assume semantic values in a way determined by their syntactic structure and the semantic values of their syntactic parts. Frege provided the detail of a model syntax apt for a wide range of languages, and of a semantic theory suited to such a syntax. Crucially, his account takes the semantic value of a singular term to be a specific object from the domain (its referent), and that of a sentence to be a truth-value,

classically conceived.

Given a semantic theory, an ‘interpretation’ for a language consists of a domain along with an assignment of semantic values to primitive expressions. With some sentence value(s) distinguished as ‘designated’, these interpretations fix a logic: *C* follows from premises *X* if and only if on any interpretation in which all members of *X* take a designated value, so does *C*. Specifying an interpretation, and so presenting the semantic values of primitive expressions in a specific way, gives them a sense; ‘the’ sense of the primitives of a functioning language is given by a specification sensitive to the antecedent meanings. Senses for complex expressions are then given by the specifications of their derivative semantic values which the semantic theory generates on the basis of their constituent structure; in particular, on the Fregean model at least, the sense of a sentence is given by that condition for its truth which is derived on the basis of its constituent structure.

How does sense, thus explicated, relate to linguistic ability? Dummett argues that the latter is not merely practical, like the ability to swim, but that it involves a substantial theoretical component, which the theory of sense spells out. Thus speakers are to be credited with knowledge of the truth-conditions of sentences as deriving from constituent structure, and it is in such knowledge that their grasp of the sense of a sentence consists. Attribution of such knowledge is unproblematic where the speaker can supply a non-circular verbal explanation of the truth-condition of a sentence, so that the knowledge concerned can be regarded as explicit and actual. But, on pain of circularity, not all cases can be like this. For the rest, we must reckon the knowledge involved either actual but implicit, that is, unconscious; or else explicit but hypothetical (so that our theory articulates only what *would* suffice to account for linguistic capacity, were it known explicitly, saying nothing of the mechanisms which actually underpin the capacity it characterizes). Either way, the theory of sense must justify such attributions, by showing how they meet but do not exceed what is required to account for features of the linguistic practice they purport to explain. Indeed, by pointing to the way the knowledge it attributes is manifested in use, it explains in what this knowledge consists (see Sense and reference §§1-3).

2 Truth, use and inference

Scrutiny of the relationship between key notions in the theories of force and of sense, Dummett thinks, reveals the point of contact between the speaker’s knowledge of sense and linguistic practice. For the archetype of the sort of force which an utterance may carry is the basic one of assertion, so a major task for the theory of force will be to articulate the conventions governing that practice, and in particular to delineate the conditions under which an assertion is correct, that is, unimpeachable *qua* assertion (even if still liable to more general criticism, for example, *qua* obscenity). In doing so, it links assertion with truth: for an assertion is correct if and only if made by a speaker possessed of what is conventionally counted an adequate warrant for its truth. This connection, suitably communalized, might almost serve to characterize the most primitive notion of truth: truth at its most primitive is the property a sentence has if and only if among us we possess a warrant which would render an assertion of it correct.

Almost; for in fact the condition for truth even at its most minimal must extend beyond our actual possession of a warrant. This follows from the logical behaviour sentences exhibit when embedded in more complex contexts, such as the antecedent of a conditional: ‘if some even number exceeding 2 is not the sum of two primes, Goldbach was in error’ says that Goldbach was in error under a condition including, but more generous than, that in which we possess a warrant for the antecedent because we can actually produce an instance of an even number of the relevant sort. Again, truth is what deductive reasoning preserves; and we can learn something new by drawing a deductive inference. But that would be impossible if reaching a true conclusion always meant reaching one for which we already possessed a warrant.

Classical truth is the concept obtained by widening the core notion of actual warranted assertibility until a statement is counted true if and only if a warrant for it could be obtained by a being able to transcend the limitations which restrict our observations to the local and the gross, and our surveys to finite domains. While Dummett sometimes reserves the word ‘true’ for classical truth, he also makes it clear that less radical widenings may produce concepts deservedly called truth-concepts. The one he discusses most is verificationist truth, which retains the human measure of warrants, but counts a statement true when our warrant for it is merely potential, not actual. Another is pragmatist truth, under which a statement is reckoned true if we potentially possess a warrant for everything following from the statement as a consequence. Evidently, a logic based on either of these alternative

truth-concepts would differ from the classical one, in ways it remains for theory to detail, though some points of difference are obvious immediately. In particular, the existence of undecidable statements - statements for which we possess neither a potential warrant, nor a guarantee that we will never possess one, such as statements about the remote past or future, or ones involving counterfactual conditionals or quantification over infinite domains - means that neither will follow the classical conception in validating the Bivalence Principle (that every proposition is either true or false).

The primordial connection of truth-concepts with assertion suggests that we look to patterns of assertive practice for behavioural expression of a speaker's grasp of truth-conditions. If, moreover, the grasp of truth-conditions is to be constitutive of attaching sense, the pattern of assertive practice associated with grasp of the truth-condition of a given sentence must be so structured as not to include assertions already presupposing understanding of the sentence itself. Thus Dummett rejects meaning holism, and postulates instead the existence of a hierarchy of sentences, with understanding of a given sentence presupposing a prior grasp only of sentences occupying a lower place in the hierarchy.

Now let a 'canonical' warrant for a statement at the lowest level of this hierarchy be some cognitive state - paradigmatically an observational one - determined by the semantic structure of the statement, and certified by the theory of force as sufficient for its correct assertion; and let a canonical warrant for a higher-level statement be a set of lower-level statements jointly providing adequate grounds for the statement in question, by the lights of the deductive logic of the underlying semantic theory, or such looser standards as the theory of force approves. Then a speaker may manifest grasp of verificationist truth-conditions by a disposition to assert a lowest-level statement when possessed of a canonical warrant for it, and a higher-level one when disposed to assert lower-level sentences constituting one of its canonical warrants. Similarly, a grasp of the pragmatist truth-conditions of a statement is manifested by a disposition to assert such lower-level statements as are guaranteed to be canonically warranted if the given one is. (And to act in ways consonant with such warrants, we may add. But such action too is reflected in patterns of assertion, because acting warrants the assertion that one has acted.)

Critically, however, there is no additional assertive pattern which unambiguously certifies that a speaker's grasp is of a *classical* truth-condition, as distinct from one of these weaker substitutes. The sole plausible candidate is that it is manifested in the additional propensity to reason with the full strength of classical logic. This is, however, an unreliable indicator, since we allow our deductive practices to be shaped by our semantic theorizing - including, perhaps, an illegitimate attribution to ourselves of a grasp of classical truth-conditions.

The inability of a theory of sense based on classical truth to meet the manifestation requirement suggests founding the theory on an alternative truth-concept. The matter of choosing between the verificationist and pragmatist concepts for this role is less substantial than may appear, Dummett argues; for the aspects of linguistic practice which each emphasizes - justification of and commitment by assertion - are so intertwined that, to the extent our practice forms a whole which is harmonious (in a sense apt for formal characterization), the same ultimate end will be reached from either beginning. What is unclear is how to frame a semantic theory apt to underpin a theory of sense based on either concept. Construing the semantic value of a sentence as a truth-value works only in the classical case. We can still take the sense of a sentence to be given by a condition for nonclassical truth; but the elaborating theory must proceed indirectly, and assign to sentences semantic values not themselves truth-values, but in whose terms truth may be defined. One paradigm is provided by Heyting's semantics for mathematical sentences, which takes as the semantic value of a sentence an effective division of mathematical constructions into those which do, and those which do not, prove the sentence, and defines verificationist truth in terms of such values. The urgent task for the theory of meaning is to find analogous nonclassical semantics apt to found theories of sense for less rarified areas of discourse. In the absence of these, our most detailed account of language rests unsatisfactorily on a suspect attribution to its users of a shady concept.

3 Realism and antirealism

Whereas, in Dummett's view, further philosophical issues typically arise through applying insights of the core theory of meaning to the language of specialist disciplines, the metaphysical issue of realism is distinctive inasmuch as, properly diagnosed, it is directly concerned with themes in the theory of meaning, and hence a semantic issue (see Realism and antirealism §4).

The leading idea is that the essence of a realist thesis is not the independent existence of a range of entities, but the objectivity of the truth of a set of sentences. The earlier Dummett seeks to cash objectivity in terms of bivalence, but baulks at classifying all rejections of bivalence as rejections of realism, citing in illustration Strawson's doctrine that sentences containing vacuous singular terms lack truth-value (see Strawson, P.F. §2). Hence his earlier writings seek to distinguish deep and shallow rejections of bivalence, Strawson's typifying the shallow; and regard realism as compromised only when rejection goes deep. His later view is that even Strawson's rejection of bivalence does repudiate a realism, namely Meinong's unlovely realism concerning nonexistent objects (see Meinong, A. §3); but further, that rejecting *any* feature of the classical semantic theory, not just bivalence, marks a departure from realism. So Russell's alternative to Strawson's treatment of vacuous singular terms also involves antirealism, since his Theory of Descriptions treats (apparent) singular terms nonclassically in not giving their semantic values by an assignment of referents (see Descriptions §2). In so far as the analysis of the theory of meaning sketched in preceding sections militates against classical semantics, therefore, it constitutes a case for global antirealism, though an incomplete one pending details of a suitable antirealist alternative.

In Dummett's terms, reductionism involves repudiating realism only when the semantics of the reducing theory, or the details of the reduction, preclude reading classical semantics into sentences of the reduced theory; notably, when they determine that truth as applied to such sentences is not bivalent, or that the semantic value of a singular term is other than its referent. Central-state materialism thus involves no antirealism: sentences and singular terms of mental discourse assume respectively the determinate truth-values and referents of the neurophysiological expressions to which they are reduced. A common reductive pattern which *does* entail antirealism arises out of subjunctive conditionals. Let 'Were A the case, B would be' and 'Were A the case, B would not be' be called 'opposites'. Then there is no general guarantee that one of a pair of opposites must always hold. Yet often reduction proceeds by reducing a sentence to a subjunctive conditional, and its negation to the opposite; so bivalence for the reduced sentences is flouted and realism rejected, unless special circumstances can be pleaded. Behaviourism and phenomenalism both exemplify this pattern and, lacking grounds for special pleading, qualify as antirealisms, though the commitment to rejecting bivalence was not recognized by historical phenomenalists.

Sundering the link between antirealism and reductionism allows for non-reductionist antirealisms, an exemplar of which Dummett finds in mathematical intuitionism. True, many formulations of intuitionism link it with reductionist doctrines of mathematics as the study of solipsistic mental constructions. But these, Dummett argues, are best construed metaphorically; he construes intuitionism as advocating a semantics for mathematical language based upon verificationist truth, embodying suggestions for the detail of such a semantics, and a mine of ideas for its extension beyond mathematics. These intuitionist sympathies fall short of open espousal, though Dummett is committed to mathematical antirealism in some form - his careful examination of Frege's philosophy of mathematics concludes that paradox is inevitable when classical logic combines with 'indefinitely extensible concepts' such as that of number. Tensed discourse constitutes another area for which any antirealism must be non-reductionist. Rejection of bivalence for future-tensed statements is familiar enough; a further Dummettian theme is that the apparent fixity of the past is not the barrier it may seem to a similar rejection for past-tensed ones.

See also: Antirealism in the philosophy of mathematics §2;Intuitionist logic and antirealism §4; Many-valued logics, philosophical issues in §4-5; Meaning and verification §5; Knowledge, Tacit

BARRY TAYLOR

List of works

This list comprises only books on philosophical themes; these include anthologies, in which almost all of Dummett's philosophical essays are collected.

Dummett, M.A.E. (1973) *Frege: Philosophy of Language*, London: Duckworth; 2nd edn, 1981.(Detailed discussion of Frege's philosophy of language, and themes arising therefrom.)

Dummett, M.A.E. (1977) *Elements of Intuitionism*, Oxford: Clarendon Press.(Text on intuitionist logic and mathematics, with much philosophical motivation.)

Dummett, M.A.E. (1978) *Truth and Other Enigmas*, London: Duckworth.(Anthology containing almost all philosophical papers published before 1976. See especially Essay 1, on truth; Essay 14, on philosophical motivations of intuitionism; Essay 21, on antirealism about the past.)

Dummett, M.A.E. (1981) *The Interpretation of Frege's Philosophy*, London: Duckworth.(Further defence of the

interpretation of Frege put forward in Dummett 1973.)
Dummett, M.A.E. (1991a) *Frege and Other Philosophers*, Oxford: Clarendon Press.(Anthology of papers entirely or largely about Frege and not previously anthologized.)
Dummett, M.A.E. (1991b) *Frege: Philosophy of Mathematics*, London: Duckworth.(Detailed examination of Frege's philosophy of mathematics.)
Dummett, M.A.E. (1991c) *The Logical Basis of Metaphysics*, Cambridge, MA: Harvard University Press.(The most systematic statement of Dummett's own overall position. Not for beginners.)
Dummett, M.A.E. (1993a) *The Origins of Analytical Philosophy*, London: Duckworth.(Traces the origins of the analytical movement, emphasizing the contribution of Frege.)
Dummett, M.A.E. (1993b) *The Seas of Language*, Oxford: Clarendon Press.(Anthology of almost all papers on metaphysics and philosophy of language not previously anthologized. See especially Essay 2 on meaning; Essay 8, on truth; Essay 18, on mathematics; Essay 20, the best introduction to realism and antirealism; Essay 11, for more detail on the same theme.)

References and further reading

Heck, R. (ed.) (1998, forthcoming) *Logic, Language, and Reality: Essays in Honour of Michael Dummett*, Oxford: Oxford University Press.(Collection of papers on all aspects of Dummett's work.)
McGinn, C. (1980) 'Truth and Use', in M. Platts (ed.) *Reference, Truth and Reality*, London, Boston, MA, and Henley: Routledge & Kegan Paul, 19-40.(Lucid exposition of Dummett's position combined with critical evaluation.)
McGuinness, B. and Oliveri, G. (eds) (1994) *The Philosophy of Michael Dummett*, Dordrecht: Kluwer. (Collection of papers on Dummett's work, with replies by Dummett.)
Peacocke, C. (1986) *Thoughts: An Essay on Content*, Oxford: Blackwell.(See especially Chapters 1-3. An illuminating if at times difficult attempt to reconcile the manifestability requirement with classical truth.)
Taylor, B. (ed.) (1987) *Michael Dummett*, Dordrecht: Martinus Nijhoff.(Collection of papers on Dummett's work, with replies by Dummett.)
Wright, C. (1993) *Realism, Meaning and Truth*, Oxford and Cambridge, MA: Blackwell, 2nd edn.(Collection of papers elaborating and expounding a position deeply influenced by Dummett's work.)

Duns Scotus, John (*c.*1266-1308)

Duns Scotus was one of the most important thinkers of the entire scholastic period. Of Scottish origin, he was a member of the Franciscan order and undertook theological studies first at Oxford and later at Paris. He left behind a considerable body of work, much of which unfortunately was still undergoing revision at the time of his death. Notable among his works are questions on Aristotle's Metaphysics*, at least three different commentaries on the Sentences of Peter Lombard (the required text for a degree in theology) and a lengthy set of university disputations, the quodlibetal questions. A notoriously difficult and highly original thinker, Scotus was referred to as 'the* subtle *doctor' because of his extremely nuanced and technical reasoning. On many important issues, Scotus developed his positions in critical reaction to the Parisian theologian Henry of Ghent, the most important thinker of the immediately preceding generation and a severe Augustinian critic of Aquinas.*

Scotus made important and influential contributions in metaphysics, epistemology and ethics. In metaphysics, he was the first scholastic to hold that the concepts of being and the other transcendentals were univocal, not only in application to substance and accidents but even to God and creatures. In this, Scotus broke with the unanimous view based on Aristotle that being could not be predicated of both substance and accident, much less of, except by analogy, God and creature. Scotus argued in general that univocity was required to underwrite any natural knowledge of God from creatures or of substance from accidents. Given univocity, he concluded that the primary object of the created intellect was being, rejecting Aquinas' Aristotelian *view that it was limited to the quiddity of the sense particular and Henry of Ghent's* Augustinian *view that it was God. That is, Scotus argues that even the finite intellect of the creature is by its very nature open to knowing all being.*

Scotus' proof of the existence of God is the most ambitious of the entire scholastic period. Prior efforts at demonstrating the existence of God showed little concern with connecting the eclectic body of inherited arguments. Scotus' proof stands apart as an attempt to integrate logically into a single demonstration the various lines of traditional argument, culminating in the existence of God as an actually infinite being. As a result, his demonstration is exceedingly complex, establishing within a sustained and protracted argument God as first efficient cause, as ultimate final cause and as most eminent being - the so-called triple primacy - the identity within a unique nature of these primacies, and finally the actual infinity of this primary nature. Only with this final result of infinity is Scotus prepared to claim he has fully demonstrated the existence of God. Notable features of the proof include Scotus' rejection of Aristotle's argument from Physics VIII (the favoured demonstration of Aquinas), the reduction of exemplar cause to a species of efficient cause, important clarifications about the causal relations at issue in arguments against infinite regress, an a priori proof constructed from the possibility of God similar to that proposed by Leibniz, and the rejection of the traditional argument that the infinity of God can be inferred from creation ex nihilo.

Scotus is a realist on the issue of universals and one of the main adversaries of Ockham's programme of nominalism. He endorsed Avicenna's theory of the common nature, according to which essences have an independence and priority to their existence as either universal in the mind or singular outside it. Intepreting Avicenna, Scotus argued that natures as common must have their own proper unity which is both real and less than the numerical unity of a singular; that is, natures are common prior to any act of the intellect and possess their own real, lesser unity. They are accordingly not of themselves singular, but require a principle of individuation. Rejecting the standard views that essences are individuated by either actual existence, quantity or matter, Scotus maintained that the principle of individuation is a further substantial difference added to the species. This 'individual difference' is the so-called haecceitas *or 'thisness', a term used seldom by Scotus himself. The common nature and individual difference were said by Scotus to be really identical in the individual, but 'formally distinct'. The 'formal distinction,' developed by Scotus chiefly in connection with the Trinity and the divine attributes, is an integral part of his realism and was as such attacked by Ockham. It admits within one and the same thing a distinction between realities, formalities or entities antecedent to any act of the intellect to provide an objective foundation for our concepts. These formalities are nonetheless really identical and inseparably united within the individual.*

In epistemology, Scotus is important for his demolition of Augustinian illumination, at least in the elaborate defence of it given by Henry of Ghent, and the distinction between intuitive and abstractive cognition. Scotus

rejected Henry's defence as leading to nothing but scepticism, and set about giving a complete account of certitude apart from illumination. He grounded certitude in the knowledge of self-evident propositions, induction and awareness of our own states. After Scotus, illumination never made a serious recovery. Scotus' other epistemological contribution was the allocation to the intellect of a direct, existential awareness of the intelligible object. This was called intuitive cognition, in contrast to abstractive knowledge, which seized the object independently of whether it was present to the intellect in actual existence or not. This distinction, credited to Scotus by his contemporaries, was invoked in nearly every subsequent scholastic discussion of certitude.

*While known primarily for his metaphysics, the importance and originality of Scotus' ethical theory has been increasingly appreciated. Scotus is a voluntarist, holding for example that not all of the natural law (the decalogue) is absolutely binding, that prudence and the moral virtues are not necessarily connected and that the will can act against a completely correct judgment of the intellect. It is Scotus' theory of will itself, however, that has attracted the most attention. He argues that the will is a power for opposites, not just in the sense that it can have opposite acts over time but in the deeper sense that, even when actually willing one thing, it retains a real, active power to will the opposite. In other words, he detaches the notion of freedom from those of time and variability, arguing that if a created will existed only for an instant its choice would still be free. In this, he has been heralded as breaking with ancient notions of modality that treated contingency principally in terms of change over time. Scotus argued that the will, as a capacity for opposites, was the only truly rational power, where the rational was opposed to purely natural agents whose action was determined. In this sense, the intellect, as a purely natural agent, was not a rational power. Finally, Scotus endowed the will with an innate inclination to the good in itself apart from any advantage it might bring to the agent. This inclination or affection for the just (*affectio iustitiae*, as it was termed by Anselm), enabled the will to escape the deterministic inclination of natures toward their own perfection and fulfilment.*

1 Career

Despite his importance and influence, little is known with certainty about the life and career of Duns Scotus. The commonly reported details of his family origins, early education and entry into the Franciscan order are now regarded as suspect owing to their origin in a partially fabricated eighteenth-century chronicle. More reliable is the date of his ordination to the priesthood in 1291, from which it is inferred that he must have been born about 1266. Scotus probably began his theological studies at Oxford about 1288, although it is debated whether he also studied at Paris prior to 1302. In any event, it is certain that he was present in Oxford in 1300 as a bachelor in theology, at which time he was participating in disputations and beginning to revise his lectures given there on the *Sentences* of Peter Lombard.

Scotus, however, never became master of theology at Oxford. At the recommendation of the English provincial, he was instead sent to Paris to lecture on the *Sentences* for a second time, which he began to do in the autumn of 1302. His Paris lectures were interrupted in June 1303 when King Philip the Fair required declarations of allegiance from religious houses at the university during his dispute with Pope Boniface VIII over taxation of church property. Scotus was among some eighty Franciscans from the Paris convent expelled from France by Philip for siding with the pope. During his exile from Paris Scotus is thought to have returned to Oxford, at which time he may have held his Oxford *Collations*. Scotus was back in Paris to resume his lectures on the *Sentences* by the autumn of 1304, when he was nominated by the Franciscan minister-general, Gonsalvus of Spain, as next in line for the Franciscan chair in theology. In his recommendation, Gonsalvus attested that Scotus' reputation had already 'spread everywhere'. Scotus is accordingly thought to have incepted as master by early 1305.

As regent in theology at Paris, Scotus disputed one set of quodlibetal questions and perhaps his Paris *Collations*. For reasons that are unclear, he was soon transferred from Paris to the Franciscan house of studies in Cologne, where a document dated February 1307 names him as a lector. Nothing is known of his activities at Cologne, where he appears to have remained until his early death, traditionally given as 8 November 1308. Remarkably, Scotus produced the bulk of his substantial writings in a period of barely ten years.

2 Works

As with the details of his career, Scotus' works suffered greatly in transmission. It is fair to say that no scholastic thinker of his stature has been so burdened by misattribution and textual confusion. While much progress has been

made in untangling Scotus' corpus, fundamental questions remain, particularly concerning chronology.

Scotus' works can be divided into philosophical and theological writings, with the latter generally regarded as later and more definitive. The philosophical writings consist first of all of questions on Porphyry and on the *Categories*, *Peri hermeneias* and *Sophistical Refutations* of Aristotle. These logical works are all presumed to be early products of his arts training, and they appear to have exercised little influence. Much more important are Scotus' lengthy questions on the *Metaphysics* (only books I-IX are authentic), also traditionally regarded as an early work dating from his arts career. The questions on the *Metaphysics* are notorious for their difficulty, arising in part from the hundreds of revisions, additions and intrusions made to the text. Their traditionally early dating has been somewhat tempered in light of research indicating that certain sections appear to have been later. Finally, a much shorter set of questions on Aristotle's *On the Soul* is attributed to Scotus, but its chronology is uncertain.

Scotus' reputation, however, rests on his longer and more developed theological writings, and principal among these are his commentaries on the *Sentences*. A major advance of textual research on Scotus has been to tease apart the various versions of his *Sentences* that had been conflated even by his earliest disciples. At least three commentaries are now recognized: the *Lectura*, which are his earliest lectures on the *Sentences* at Oxford; the *Ordinatio*, a greatly expanded revision of the *Lectura*; and the *Reportatio parisiensis*, which are students' reports of his Parisian lectures. Of capital importance for the interpretation of Scotus is the chronological relationship of the *Ordinatio* and *Reportatio parisiensis*, because the treatment of important issues in the Paris lectures differs markedly from that in the Oxford commentaries. A governing thesis of the critical edition of this work has been that the *Ordinatio* formed the latest and most definitive of Scotus' commentaries, incorporating both his early Oxford *Lectura* and his Paris lectures. A revised tendency, however, has been to see the first book of the *Ordinatio* as earlier than the *Reportatio parisiensis*. In other words, it is increasingly thought that Scotus must have begun work on the *Ordinatio* before he left Oxford for Paris in 1302. Resolution of this must await further study of the Paris reports, which remain unedited for the first book.

In addition to his commentaries on the *Sentences*, Scotus left two sets of theological disputations. The first, his *Quaestiones quodlibetales (Quodlibetal Questions)*, certainly date from his regency at Paris and should be regarded as the mature work of a master at the height of his career. A second set of university disputations, the *Collationes (Collations)* are also important but have been little studied. As with the *Sentences*, Scotus has *Collations* from both Oxford and Paris. Finally, Scotus wrote two treatises, the *De primo principio (On the First Principle)* and the *Theoremata*. The former is a lengthy and systematic deduction of the existence and nature of God according to axiomatic method. Nearly two-thirds of the *De primo principio*, however, comes directly from the *Ordinatio*, which suggests that Scotus may not have finished the treatise himself. The authenticity of the *Theoremata* has been contested owing to a section which argues, among other things, that natural reason cannot demonstrate the existence of God. The *Theoremata* is a work nonetheless attributed to Scotus both by manuscripts and by his contemporaries.

3 Sources and method

After nearly fifty years of publication and research, the critical edition of Scotus' works has resulted in two general findings important for the exegesis of his thought. The first is that Scotus' single, most important source by far was Henry of Ghent, the leading theologian at the University of Paris in the generation after Aquinas. On one major issue after another, Scotus begins with an extensive analysis and criticism of Henry's position only to develop his own view in reaction to it. At least for his Oxford commentaries, Scotus' real textbook was not Peter Lombard's *Sentences* but, in effect, Henry of Ghent's *Summa*. The relation of Scotus to Henry, however, is complex and does not simply consist in the former rejecting Henry's conclusions. Scotus' own position is often indebted to Henry's vocabulary, distinctions and general philosophical framework. Even when he does not have Henry's opinion under direct consideration, he will presume Henry's prior discussion of the matter. Accordingly, Henry should be properly viewed as the major intellectual influence on Scotus as much as his principal adversary.

A second finding is that Scotus revised his works heavily by way of additions, annotations and insertions, termed *extra*, or 'outside Scotus' original text'. This method is especially apparent in the first book of the *Ordinatio* and in the questions on the *Metaphysics*. Scotus' additions and insertions to his initial text can run to pages and typically record further objections and replies, often lodged by a contemporary, and note cross-references to related arguments elsewhere. Consequently, at these places Scotus' texts cannot be read as finished products but as work

in progress containing several chronological layers.

4 The place of Scotus in medieval philosophy

Scotus occupied a pivotal place in scholastic thought, closing the thirteenth century and opening the fourteenth intellectually as well as chronologically. First of all, Scotus' focus on Henry of Ghent, quite apart from its obvious exegetical importance, was otherwise significant for the period. By in effect making a contemporary work the object of his commentary on the *Sentences*, Scotus fundamentally changed the programme and form of scholastic literature itself. While this change was already underway in the previous generation, Scotus nonetheless marks a clear divide between the thirteenth-century project of incorporating Greek and Arabic sources, as exemplified by Albert the Great, Bonaventure and Aquinas, and the fourteenth-century focus on contemporary opinion evident in William of Ockham.

However, Scotus' greater contribution lies in his philosophical innovations, which not only became frequently discussed opinions - Scotus is one of the most cited authors in the fourteenth century - but defined the very issues and terms of analysis for the next century. Among the important concepts introduced by Scotus must be considered the following: in metaphysics, the univocity of the transcendental concepts, proofs for the existence of God and the principle of individuation; in epistemology, the distinction between intuitive and abstractive cognition; and in ethics, the will as a rational power for contraries and the interpretation of Anselm's distinction between affection for justice (*affectio iustitiae*) and affection for what is advantageous (*affectio commodi*).

5 Univocity of the transcendental concepts

One of Scotus' most striking metaphysical positions was that being and the other transcendentals could be conceived as univocally common to God and creatures, substance and accident. In this, he broke with the unanimous view of the thirteenth century that being could not be predicated univocally of substance and accident, much less of God and creatures. The common scholastic opinion, based directly on Aristotle, was that being was predicated of God and creatures neither univocally nor equivocally but according to analogy (see Being). Univocal predication was taken to violate God's transcendence over creatures and equivocity to render natural knowledge of God impossible. Rather, being was said to be predicated according to analogy, which meant that it was asserted of God in a primary sense and of creatures in a related but derived sense. Analogy was therefore construed as a middle way between the extremes of univocity and equivocity, balancing the competing demands of God's transcendence and knowability.

In general, Scotus' position was that some univocal concept of being common to God and creatures was presumed by the traditional conviction that knowledge of the divine nature or attributes was naturally attainable. Scotus singled out Henry of Ghent for sustained attack on this score as part of a comprehensive and critical appraisal of his entire theory of knowledge, which was in general Augustinian (see Augustinianism). In his treatment of analogy, Henry was much more emphatic and explicit than previous discussions had been in claiming that being, when conceived in its utmost generality, did not form a single, common notion (*ratio*), but only two exclusive and proper concepts, one applicable only to God and the other only to creatures. He repeatedly stressed that there could be no separate notion of being distinct from those proper to either God or creature, so that being could be conceived only as either finite or infinite, created or uncreated. To admit an absolute concept of being apart from these two would simply be an admission of one common to both. Scotus argued from several angles that Henry could not consistently maintain that being resolved only into two proper notions, having no conceptual element in common, and at the same time uphold the possibility of natural knowledge of God.

Of Scotus' several arguments for univocity - at one point he outlines ten - the one labelled 'from certain and doubtful concepts' was regarded by his contemporaries as the most compelling. This argument was aimed squarely at Henry's repeated insistence that there could be no concept of being apart from the analogous and proper notions of infinite and finite being applicable exclusively to God and creature (see Henry of Ghent). An intellect certain about one concept but doubtful about others, has a concept about which it is certain that is different from the concepts about which it is doubtful. We can be certain that God is a being, while doubting whether God is infinite or finite being. Therefore, the concept of being is different from, and hence univocal to, the concepts of infinite or finite being. Scotus takes the first premise to be evident, for a given intellect cannot be both certain and doubtful of the same concept. The second premise is true *de facto* because past thinkers, such as the Presocratics, disagreed as

to whether the first principle was finite or infinite. Yet, in attempting to establish one of these alternatives, no philosopher ever doubted that the first principle was a being. Being must therefore have a separate, distinct concept.

Put more generally, Scotus' point is that prior to demonstration, the intellect is doubtful whether God is an infinite or finite being. Yet, such a demonstration must be based upon something certain about God, for otherwise it would proceed from premises doubtful in all respects. Thus, unless the concept of being is admitted as certain, apart from the doubtful concepts of infinite and finite which are themselves the object of demonstration, no certain reasoning about God will be possible. Henry's refusal to admit a concept of being distinct from any proper to God therefore entails that the intellect is either certain and doubtful of the same notion or certain of none at all.

In a similar line of attack, Scotus is more explicit still that Henry's denial of a univocal concept of being renders natural knowledge of God impossible. A creature causes a concept that is either common to it and God or proper to God alone. Since Henry denies the former, he must hold the latter to explain the natural origin of our concepts of God. Scotus, however, argues that it is impossible for a creature to cause directly any concept wholly proper to God. In general, an object can only produce a concept of what it contains either as an essential part or an essential property, as is evident from the traditional division of essential predicates. Obviously, the creature can contain nothing proper to God as either a part or property of its essence without violating divine transcendence. Thus, if a creature can directly produce any concept applicable to God at all, it must be one that is common rather than proper.

Finally, Scotus applies these same criticisms to the very foundations of all scholastic accounts for natural knowledge of God, the Anselmian doctrine of pure perfections and the Pseudo-Dionysian procedure of removal and eminence. According to Scotus, both of these presuppose univocity, so that 'All the masters and theologians seem to use a concept common to God and creature, although they deny this verbally when they apply it' (*Lectura* 1 d.3 n.29). Anselm's doctrine holds that we attribute to God those perfections found in creatures which are pure in the sense that, conceived in themselves, they entail no imperfection, such as will, intellect or wisdom. These perfections are defined by Anselm generally as 'what absolutely is better to be than not' (see Anselm of Canterbury). But by this definition, something is first determined to constitute a pure perfection and then on that basis attributed to God, not the reverse. Pure perfections abstracted from creatures must therefore have some meaning that is prior to any they have as proper to God alone. Scotus makes the same point regarding the Pseudo-Dionysian methods of removal and eminence (see Pseudo-Dionysius). According to Scotus, all metaphysical inquiry about God proceeds by taking some formal notion (*ratio formalis*) and removing from it all imperfections with which it is found in a creature. For example, we take the formal notion of the will - the power for opposites - and remove any limitations connected with its existence in a creature, such as variability. We then attribute it to God by conceiving of it not just as lacking imperfection, but as possessing the greatest degree of perfection, such that it is omnipotent. This process presumes that the formal notion of the will stripped of creaturely limitations is the same notion of will assigned the highest degree of perfection; otherwise the first step of the procedure would simply have no relevance to the second. If nothing of the notion abstracted from creatures remains when we attribute it to God, then perfections in creatures have nothing to say about the perfection of God.

The outcome of the above arguments is Scotus' revision of the structure of the concept of transcendental being. In place of Henry's scheme, where being taken in its ultimate generality resolves into a pair of simple notions proper to God and proper to creatures, Scotus admits a single, simple concept which is common to both. As a result, the analogous concepts of being proper to God and proper to creatures, which for Henry were simple and irreducible, became for Scotus composite, comprising the common notion of being and the determining concepts of infinite and finite.

It is important to stress that in these arguments Scotus does not entirely set aside the received doctrine of analogy. He of course admitted that the concepts of being proper to God and proper to creature were analogous. His fundamental point is rather that unless there is some underlying concept of being common to these analogous ones, then they will in fact turn out not to be analogous at all but purely equivocal, thus rendering natural knowledge of God impossible. What Scotus did set aside was reliance on the analogous relationship itself as sufficient to account for any proper concept of God. Since knowledge of a relation is posterior, not prior, to any knowledge of the terms related, analogy does not explain but presupposes a knowledge of being as proper to God. Accordingly, some

univocal, conceptual community between God and creatures is demanded by the traditional project of natural knowledge of the divine nature or attributes.

6 Primary object of the intellect

Given univocity, it seems to follow directly that being is the primary object of the intellect, for the concept of being is common to, and hence more primitive than, any notion proper to either God or creatures. Scotus in fact draws this conclusion, arguing that unless the concept of being is univocal, there can be no object encompassing all that the intellect can know. Thus, in addition to ensuring the possibility of natural knowledge of God, the other important epistemological function of univocity is to provide the intellect with a primary or defining object.

To avoid what he saw as equivocation on the issue, Scotus distinguished three different viewpoints from which an object of the intellect could be considered primary: generation or origin, perfection and adequation. The first two were relatively uncontroversial, since there was broad agreement that all knowledge originated from the senses and that God constituted the highest knowable object. Accordingly, Scotus concluded that the specific nature of the sense particular was first in terms of the generation or origin of knowledge. In a similarly conventional position, Scotus put God as the first object in the order of perfection absolutely speaking, while the most perfect object proportioned to our intellect was the sensible nature. (Scotus, like most, refined these positions with various qualifications and distinctions.) The order of adequation, however, was more disputed, for here the primary object defined the nature of the intellectual power as such.

By ‘adequate’ Scotus meant what Aristotle called the ‘commensurately universal’, which formed one of the conditions of strict demonstration outlined in the *Posterior Analytics*. The adequate object of a power is that which is coextensive and commensurate with all objects over which that power ranges. It is in this sense that the primary object circumscribes the scope of a power and hence marks it off as distinct from other powers. In general, Scotus recognizes two ways in which an object may be adequate: either because it forms a universal nature or aspect (*ratio*) found in all things which a power surveys, such as colour in the case of sight, or because it is a single, most perfect object that includes within itself all the other objects governed by a power. In the case of the intellect, Scotus says that an object is adequate in the first way by community or predication, for it is common to and hence predicable of all that is intelligible; in the second way, an object is adequate virtually, for by understanding it, the intellect is moved to understand all else that is intelligible.

Scotus reports two competing opinions on the adequate object of the intellect, representing broadly Aristotelian and Augustinian theories of knowledge. The first is the well-known position of Aquinas, taken to represent the Aristotelian orientation, that the adequate object of the human intellect is the essence or quiddity of the sense particular. Scotus argues against Aquinas’ position on both theological and philosophical grounds, maintaining in each case that the adequate object concerns the nature of the power as such. Theologically, Scotus rejects Aquinas’ position because, in limiting the scope of the created intellect in its nature to the material quiddity, it rendered the knowledge of the immaterial essence of God promised in beatitude impossible. On the same grounds, Scotus rejects Aquinas’ explanation that, since ‘grace perfects nature’, the intellect will be elevated by a supernatural quality that will enable it to attain an immaterial object. While Scotus of course holds that a supernatural grace is required for our intellect to have direct vision of God, he denies that any supervening quality can modify a power so as to change its adequate object. In that case, the power is not simply perfected but, by definition, transformed into a power of an altogether different nature. Philosophically, Scotus argues that on Aquinas’ position the science of metaphysics would be impossible, for the intellect cannot acquire a science whose object exceeds the scope of the primary object of the intellect itself. But the object of metaphysics, being *qua* being, is more universal than material natures.

The second opinion, which represents an Augustinian approach, is that of Henry of Ghent, who posited God as the primary object of the intellect. As indicated above, Henry held that being and the other transcendentals taken in their utmost generality resolved into two distinct notions, one proper to God and the other to creatures. Henry had to designate one of the two as primary as regards our intellect, and he argued that those proper to God were prior. This followed from his strong commitment to an Augustinian theory of illumination, according to which the essence of a creature was truly known only by reference to its eternal archetype or idea in the mind of God (see Augustinianism). Scotus replies that, as indicated, if God is the adequate object of the created intellect, the divine nature must either be common to or virtually include all that is intelligible. The divine essence is obviously not

universally common to all intelligible objects, since God cannot be predicated of creatures. Although the divine essence does virtually include all that is intelligible, God is not on this account the adequate object of any created intellect. If this were the case, then the human intellect would be moved to understanding all intelligibles by a single object, namely the divine essence, rather than directly by those intelligibles themselves. The divine essence, of course, can function as an object in this way only for the divine intellect, which is to say that God is the adequate object of the divine mind alone. By way of a corollary to his refutation of Henry, Scotus also excludes on similar grounds substance as the adequate object of the intellect in the sense that it virtually contains accidents. This would mean that accidents could only be known through substance, which is false, since accidents themselves can move the intellect as intelligible objects.

Having excluded both God and substance in his refutation of Henry, Scotus concluded that no single object can be primary for the created intellect in the sense that it virtually contains all else that is intelligible. Therefore, if there is a primary, adequate object of the intellect, it must be such owing to its community. Since nothing is more common than being, it must be the primary object of the intellect. Scotus notes that this presupposes a univocal concept of being, so that if univocity is denied the intellect can have no adequate object.

In making being the primary object of the intellect, Scotus was in fact tacitly advancing yet a third opinion, that of Avicenna (see Ibn Sina). In central passages of his *Metaphysics*, Avicenna had made being both a primary conception of the mind and, in explicit contradistinction to either God or substance, the proper subject of a universal science of metaphysics. These texts strongly implied that Avicenna had seen being as the primary object of the intellect, prior to both God and creatures. Henry of Ghent had already explicitly raised just this interpretation of Avicenna in order to reject it forcefully on the grounds that it entailed a univocal concept of being. Henry's analysis was not lost on Scotus, who adopted the Avicennian position as a consequence of his doctrine of univocity.

7 Proofs for the existence of God

Duns Scotus' argument for the existence of God was perhaps the most ambitious of the scholastic period. Running to hundreds of pages and comprising dozens of interim conclusions and corollaries, it exists in at least four significantly modified versions, one in each of his three commentaries on the *Sentences* and separately as the treatise *De primo principio*. Among its distinctive features are the demonstration of the so-called 'triple primacy', the rejection of Aristotle's proof from motion, the definition of essentially ordered causes, the argument from possibility and the demonstration of God as infinite being.

According to Scotus, the highest naturally attainable concept of God is that of an actually infinite being. Consequently, Scotus holds that a complete argument for the existence of God can demonstrate nothing less than that some being is actually infinite. For Scotus a proof for a first efficient cause, such as Aquinas' second way, would not itself fully constitute a demonstration that God exists, but, as will be evident, would be merely a preliminary step in such a demonstration. Thus whereas most scholastics, like Aquinas, first establish that God exists and only later derive infinity as a divine attribute, Scotus requires the demonstration of infinity as logically necessary to establish God's existence itself. As a result, the structure of Scotus' proof is exceedingly complex, involving a good portion of his entire natural theology.

The overall structure of Scotus' demonstration comprises three large, principal steps divided into two main articles. The first step establishes that there is a first efficient and final cause and a most perfect being, the second that these three coincide in a unique nature, and the third that this nature is actually infinite. The first two steps together constitute the first article, which Scotus says establishes God according to the relative properties of causality and eminence (in other words, perfection), and the third step forms the second article, which reaches God according to the absolute property of actual infinity. The establishment in the first half of the proof of something primary according to each of the relative properties of efficiency, finality and eminence is referred to as the demonstration of the 'triple primacy.' All of these steps are intricately argued and supported by sometimes large preliminary results, such as that God has both an intellect and will as preparatory to the proof of actual infinity.

As in other areas of his thought, Scotus' proof reveals the influence of Henry of Ghent. Scotus took over the structure of the triple primacy directly from Henry's attempt to schematize the eclectic body of received arguments for the existence of God according to the Pseudo-Dionysian ways of causality and eminence (see

Pseudo-Dionysius). By subdividing causality into efficient, formal or exemplar, and final causes and interpreting eminence in terms of degrees of perfection, Henry sought to reconcile the disparate arguments from the Aristotelian (efficient and final cause) and Augustinian (exemplar cause and eminence) traditions. In an otherwise important metaphysical step, Scotus maintained against Henry that exemplar cause was simply a species of efficienct cause, and therefore eliminated it as a separate class of argument for the existence of God on the grounds of logical economy. He thereby streamlined Henry's original divisions of causality and eminence to the three of the triple primacy.

8 The proof from efficient causality

Each of Scotus' proofs for a primacy in efficient cause, final cause and eminence contains three main conclusions: that there is a first in that order, that it is uncaused and that it actually exists. Because Scotus establishes these three results principally for efficienct causation and then applies them to the other two cases, it will be sufficient to examine each of these three conclusions leading to a first efficient cause.

Before beginning the proof from efficienct causation proper, Scotus explains that he is demonstrating a first efficient cause of being and explicitly discards Aristotle's argument for a prime mover in *Physics* VIII. Scotus does so not because he thinks Aristotle's proof invalid but, again, on grounds of economy. To prove a first cause of motion is not necessarily to prove a first efficient cause of being. While the two may coincide in reality, further demonstration is required to establish this identity. Including Aristotle's physical proof would therefore necessitate a further step. Prior to Scotus, Aristotle's approach had been increasingly seen as inferior to Avicenna's strictly metaphysical proof based on necessity and possibility, especially by Henry of Ghent. Scotus, however, represents the final step where *Physics* VIII is omitted altogether from the standard corpus of arguments for God.

The first conclusion under efficienct causation, then, is that there is some efficient cause absolutely first, so that it neither exists nor exercises its causal power by virtue of some prior cause. Scotus' main argument for this first conclusion is brief. He formulates the initial premise broadly to include even possible effects: some being can be caused efficiently (*aliquod ens est effectibile*). It is therefore either caused by itself, by nothing or by another. Since it cannot be caused by nothing or by itself, it is caused by another. This other is either a first efficient cause in the way explained or it is a posterior agent, either because it can be an effect of, or can cause in virtue of, another efficient cause. Again, either this is first, or we argue as before, and some prior cause is required. Thus, there is either an infinity of efficient causes, so that each has some cause prior to it, or there is a first cause posterior to none. Since an infinity of causes is impossible, there must be a primary cause that is posterior to none.

This argument immediately encounters two objections, and Scotus' replies to them contain the bulk of his proof. The first is that the argument begs the question because it assumes that an infinite regress of causes is impossible. Here is it observed that the philosophers (that is, Aristotle and Avicenna) admitted an infinite series of causes, for they held that the generation of individuals could proceed to infinity, and hence every generating agent would have some prior cause. The second objection is that Scotus' argument is not strictly demonstrative because it is based on premises that are contingent, namely, that some effect exists. As such, the proposed proof lacks the necessity required of Aristotelian demonstration in the proper sense.

9 Essentially-ordered causes

In response to the first objection, Scotus defines the precise nature of causal relations at issue in arguments against infinite regress. According to Scotus, they do not concern simply essential as opposed to accidental causes, but rather essentially-ordered as opposed to accidentally-ordered causes. The former concern only the relationship that a single cause bears to its given effect, namely, that the effect arises from the nature of the cause rather than from something incidental to it. Essentially-ordered causes, however, concern the relationship of several causes to each other in jointly producing an effect. As defined by Scotus, there are three features of essentially-ordered as opposed to accidentally-ordered causes. The first is that the posterior cause depends upon the prior for the very exercise of its causality and not just for its being, which can be the case in accidentally-ordered causes. The second is that essentially-ordered causes always differ in nature so that the prior cause is more perfect in kind. This is a consequence of the first feature, for given two causes of the same nature, either is sufficient to produce the same effect. Finally, essentially-ordered causes must be simultaneously present to produce their effect, for otherwise from the second feature, some perfection in causality required for that effect would be missing.

Given this distinction, Scotus excludes the counterexample drawn from the philosophers, since it concerns an infinite series of temporally successive, generating causes. The causes in such a series are therefore not essentially but only accidentally ordered to each other in producing a given effect, for the posterior cause does not depend upon the prior for its causal action itself, but only for its existence. This is clear, for all the generating causes in the series are individual agents of the same nature or species, and thus not all are required simultaneously to produce the same effect. For instance, parents can produce a child whether or not their own parent or grandparent is alive. Rather, Scotus directs his argument against an infinite series of causes upon which the entire succession of individual agents itself would depend. Scotus claims that no philosopher admitted an infinite series of such essentially ordered or 'ascending' causes.

Having so defined the notion of causality operative in the proof, Scotus remains content to give five brief arguments against infinite regress, based in part upon the received reasoning of Aristotle's *Metaphysics* II and Avicenna. An exception is the fifth of these arguments, for here Scotus establishes the possibility of a first efficient cause, the actual existence of which he deduces later. Scotus argues that since efficient causality does not of itself imply imperfection, it is possible for it to exist in some nature without imperfection. That is, efficient causality, like wisdom or intellect, is a 'pure perfection'. However, if there is an infinite regress in efficient causes, then all would be dependent on some prior cause and efficiency could never be found without imperfection, contrary to assumption. Therefore, a first efficient cause in the sense defined must be possible. While seemingly weaker than the other arguments against infinite regress in that it establishes only the possibility of a first efficient cause, this result enables Scotus to construct a strict demonstration which he claims is necessary.

10 Proof from possibility

To the second objection, that the proof began from contingent premises and thus was not a true demonstration, Scotus replies that his argument for a first efficient cause can be formulated with either existential or modal premises. In the first way, the argument begins with the actual existence of some effect or change and argues directly to the existence of a cause owing to the correlative nature of effect and cause. Formulated in this fashion, the argument is based on contingent but evidently true premises. In the second way, the argument takes its premise from the possibility of some effect and concludes to the possibility of a cause. The actual existence of a first efficient cause is then deduced from its possibility. In this way, the argument can be recast so as to be a necessary demonstration, for the premises are statements not about the actual existence of some effect but about its nature or possibility. Scotus draws out the necessary argument from possibility in the last conclusion concerning efficient cause.

Having answered these two objections, Scotus proceeds to the remaining two of three conclusions necessary to establish a primacy in efficient causality. The second is that the first efficient cause is uncausable, with respect to both its own existence and its ability to cause. As Scotus indicates, this conclusion simply makes explicit the notion of 'first' already demonstrated in the arguments against infinite regress. The third and final conclusion is that an efficient cause first in this sense actually exists. As established in the fifth argument against infinite regress, a first efficient cause is possible. Scotus then argues that if such a cause is possible, it must actually exist. If it does not exist, then it could only be possible if some other cause was able to bring it into existence. But such a first efficient cause is absolutely uncausable, so that if it does not actually exist, it is impossible for it to exist. Therefore, if the first efficient cause can exist, it does exist. Alternatively, Scotus says, the same conclusion can be reached by the other traditional arguments recorded against infinite regress, although, as indicated in the above reply to the second objection, they begin from contingent premises.

11 Actual infinity of first efficient cause

After establishing the existence of a first efficient and final cause and most perfect being, and their identity in a unique nature, Scotus then moves to the actual infinity of that nature. As indicated, Scotus does not consider the existence of God to have been proven until actual infinity has been demonstrated. He derives the infinity of this primary nature from its properties of efficienct causality, final causality and eminence. The demonstrations based on efficient causality are notable.

Scotus considers two standard arguments for the infinity of the first efficient cause, both treated in some detail by Henry of Ghent in his question on divine infinity, and then constructs a third of his own. The first, based on

Aristotle's *Physics* VIII, is that since the first mover causes an infinite (that is, beginningless) motion, it must be infinite in power. Scotus, however, sees the argument as needing considerable expansion to establish infinity in the sense desired. First of all, to avoid basing the argument on the false assumption of an eternal world, he argues that the antecedent could be changed to the weaker claim that it is possible for God to cause an infinite motion. (Scotus, like Aquinas, regarded the eternity of the world as factually false but not impossible (see Eternity of the world, medieval views of).) Scotus claims that the consequent equally follows if God can, but in fact does not, produce an eternal motion. Second, he recognizes that a cause is not infinite in power simply because it produces an effect or succession of like effects - in this case the uniform rotations of the outermost heaven - for an infinite duration. Given a finite effect of infinite duration or an infinite succession of such effects, it only follows that their cause is also infinite in duration, not necessarily in power or perfection. Thus, Scotus revises Aristotle's original reasoning by arguing that the prime mover is causally responsible for the entire infinite succession of motions and derived effects taken together and in their totality. This he does in several ways, all of which depend upon recognizing that the prime mover, since it is the first efficient cause, depends on nothing else for its causal power. As such, the prime mover must possess within itself all at once the total power required to produce all of its effects realizable over an infinite time, for it can receive no power from any external cause. Since these effects are infinite in number, it must be infinite in power.

In addition to the Aristotelian approach based upon the prime mover as the first efficient cause of motion, Henry advanced a second argument for divine infinity based on God as the first efficient cause of being, that is, on God as creator. Henry argued that since the distance traversed from nothing to being in the act of creation *ex nihilo* was infinite (for a finite distance presupposes two finite beings), God had to be infinite in power. Whereas Scotus found the first approach based on Aristotle's *Physics* salvageable, he rejected outright this second argument based on creation. First of all, he noted that the argument requires that creation be taken in the revealed sense of a temporal beginning of the world, which is a matter of religious belief, rather than in the philosophically demonstrable sense of causal dependence. Secondly, Scotus denies that there is an infinite 'distance' between being and nothing. Contradictories are not distant in the sense that there is some interval between them, for they are immediate, but only in the sense that one extreme is more perfect than the other. Thus, two opposed extremes cannot be more 'distant' than the more perfect of the two. But the more perfect extreme in creation is merely finite. Thus, while a creature is infinitely distant from God, since the more perfect extreme is infinite, it is not so distant from nothing.

Scotus' third and preferred argument for infinity is drawn from exemplar causation, that is, from the efficient cause considered as an intellective agent. Prior to this proof, Scotus first establishes three necessary preliminary results: that the first cause has an intellect and will, that its intellectual and voluntary acts are identical to its essence and that it knows all that can be known both distinctly and actually. From this he argues that since the divine intellect knows distinctly and actually all that can be known, it knows these things all at once, for an intellect knows successively only if it moves from confused to distinct or from potential to actual knowledge. The things that can be known are infinitely many; therefore, since the intellect of the first efficient cause knows infinitely many things at once, it is actually infinite.

In sum, then, Scotus accepts proofs for infinity based upon the first efficient cause as prime mover and exemplar but not as creating cause. After completing further proofs for infinity based on finality and eminence, he concludes that the existence of God has been demonstrated according to the highest attainable concept (see God, arguments for the existence of).

12 Universals and individuation

Generally speaking, Scotus is regarded as a realist on the issue of universals because he admits that the universal has some reality and unity prior to any act of the intellect and accordingly that it has some sort of real distinction - the so called formal distinction - from the individual. On this score, he was attacked by William of Ockham, who, committed to a thorough nominalism, denied any sort of distinction within the individual that would grant the universal a reality of its own (see Nominalism; Universals). As such, Scotus and Ockham are typically viewed as poles in the fourteenth-century strain of the realist-nominalism debate. Scotus, however, was not an extreme realist. For instance, he argued at length against Henry of Ghent's theory which accorded the essences of things a real being in the mind of God antecedent to their creation (see Henry of Ghent). Even Ockham places Scotus next

to last in his series of opinions ranked according to their degree of realism.

Scotus does not directly treat the problem of the reality of universals, as one finds it treated in the commentaries of Boethius and Abelard, but rather addresses it in the course of determining the principle of individuation. By Scotus' time, the thirteenth-century discussion of individuation had become highly involved, leading Peter Olivi to remark that there was 'an endless forest of opinions on the matter'. Scotus reaches his own position after a lengthy examination and rejection of five possible views: a common opinion that there is no need to posit a separate principle of individuation, followed by four specific views: negation (Henry of Ghent), actual existence (a common view), quantity (Giles of Rome) and matter (a common view attributed to Aristotle). Scotus devotes a separate question to the elimination of each of these opinions, leaving him to conclude in a sixth view what the principle of individuation must be.

The greatest burden of Scotus' entire discussion is the refutation of the first view that there is no need to posit a distinct principle of individuation. The issue is whether a common nature, such as equinity or humanity, is of itself individual. This first view holds that the nature is individual of itself, so that there is no need to account for its individuality by any other factors than those that bring the nature itself into actual existence, namely, the generating causes themselves. It is not, in this view, that the nature is first produced as universal and then some intervening causes are required to contract it to a singular instance, for the nature is produced and exists only as singular. To the contrary, what is required is an explanation of the nature's universality, for this does not belong to the nature as it exists absolutely and in reality but only in relation to the intellect.

In essence, what Scotus is combating is the nominalistic position that reality is thoroughly singular and hence individuality requires no explanation. Scotus mounts two main arguments against this view, both of which bring out his realism. The first is that the object of a power, insofar as it functions as its object, is naturally prior to the act of that power. The reason is that a cause is prior to its effect, and the object is a cause of the act of a power. However, if the nature is of itself singular insofar as it is prior to an act of the intellect, then the intellect in its act of understanding will grasp its object in a way contrary to the very nature of that object itself, namely, as common rather than singular. Therefore, the nature cannot be of itself singular but must be the common antecedent to an act of the intellect.

Scotus' second argument issues in his doctrine of a lesser or 'minor' unity. He maintains that the nature must have its own proper unity which is both real and less than the numerical unity of the singular. Otherwise, every real unity would be numerical, which is false. The reason is that many relationships are recognized as real in the sense that they are not mind-dependent, yet are not based on things numerically one but on species, genera and other common classes. Therefore, these must have unity which is less than numerical but nonetheless real. For instance, the basis of all physical change, which is real, is contrariety; but things are not contrary insofar as they are numerically one, for then there would be as many contraries as individuals. Thus contraries - hot and cold, up and down - must each be one by a unity that is real but less than numerical. Or again, the relation of similarity is real in the sense that it is not simply the product of the mind. It cannot be based on what is numerically one, for then all things would be equally similar. Conversely, there are degrees of diversity that are not merely mind-imposed, so that Socrates differs more from rock than from Plato. This would not follow, however, if all unity were numerical, for then all individuals would be equally diverse.

Accordingly, Scotus concludes that the natures of things, such as equinity and humanity, are not of themselves individual but are the common antecedent to any act of the intellect. From this it follows that natures taken in themselves are not, strictly speaking, universal either, for universality in the strict sense is a relation of reason resulting from an act of the intellect. Quoting Avicenna's famous text on the common nature (*natura communis*), Scotus says that the nature taken in itself is neither universal nor singular: 'Equinity is nothing else but equinity alone. Of itself it is neither one nor many, neither universal nor particular' (*Ordinatio* 2 d.3 n.31) That is, according to Scotus, although the nature is never found except as universal in the mind or as singular outside the mind, it cannot of itself be either. If equinity were in itself universal, so as to include the note of universality in its definition, it could not be predicated of any singular instance, for no individual is a universal. If it were singular of itself, it could be asserted of only one instance. Thus, in order to be capable of realization in either state, the nature taken in itself must be neutral with respect to both. It is this common nature that forms the proper object of the intellect, functions as the predicate in true, universal statements, and has the real, lesser unity demonstrated above.

Given that the natures or essences of things are not of themselves singular, they must be made such by some individuating factor, just as they are made universal in the strict sense by an intervening act of the intellect. As indicated, Scotus rejects four candidates for this individuating principle. The first is Henry of Ghent's position that a nature is individual when it cannot be pluralized further (for example, Socrates cannot be multiplied) and is not identified with another (Socrates is not Plato).Thus, Henry concluded that a nature is individuated by a twofold negation. Scotus argues that an individual is something positive and thus cannot be caused by negation. In any event, Henry's theory does not give an exact cause, for every negation presupposes something positive. What is sought then is the positive factor that causes these two negative properties of an individual.

The next opinion is that actual existence individuates, which is based upon the principle that actuality distinguishes. Since existence is the ultimate act, it must cause the ultimate distinction, namely, individuality (see Henry of Ghent). Against this, Scotus argues that while existence is an act, it is not an act relevant to individuation. At issue is what makes some substance, such as a horse or stone, individual. Existence is an act outside of and posterior to the whole predicamental line of substance; it is in this sense that existence is often said to be 'accidental', for it lies outside the essence or natures of things. One of the more common views was that individuation was caused by quantity, since a form was taken to be pluralized insofar as it was found in an extended, material substrate. Scotus replies that if accidents are posterior to substance, this holds *a fortiori* of individual substance, since Aristotle identifies this as substance in the primary sense. Therefore, quantity as a accident is posterior to whatever makes a substance individual. Elsewhere, Scotus deploys this same line of reasoning against the Boethian theory that a collection of accidents individuates (see Boethius, A.M.S.). Finally, he rejects the standard Aristotelian view that matter individuates, since matter is in itself indeterminate and indistinct. It cannot therefore be a principle of distinction.

From all of this, Scotus concludes that the principle of individuation must be something real and positive in the substantial order as opposed to any kind of accident, whether existential or categorical, and, while not a substantial form itself, the ultimate reality or perfection of that form. In other words, the principle of individuation is a further substantial difference added to the specific nature; indeed, Scotus calls it an 'individual difference'. While the individual difference is of course not a further specific difference, Scotus depicts it as functioning metaphysically in a closely analogous way. Thus, just as the specific difference renders the nature of which is it a part incapable of division into any further species, so the individual difference renders the singular absolutely indivisible. Further, the specific difference is a reality formally distinct from, and actual with respect to, the reality of the genus. Also, the individual difference is actual with respect to the reality of the specific nature and formally distinct from it. Finally, the individual difference is irreducibly simple and hence wholly diverse from any other individual difference. In this it is comparable to ultimate specific differences, which are absolutely simple and diverse. The individual difference, however, is unlike any specific difference, because it adds no further quidditative or essential reality. If the nature or essence of a thing be considered its form, then the individuating difference may be considered 'material' in the extended sense that it contributes no common essence or nature but rather contracts such a nature to an ultimate subject. Posterity has labelled this individual difference *haecceitas* or 'thisness', a term used sparingly by Scotus himself and then usually to mean the state of being singular (*singularitas*) rather than the principle of individuation itself.

The significance of Scotus' theory of individuation is that it breaks with the fundamental Greek conception of the species as the principal locus of both being and intelligibility, codified in the Latin tradition by the Boethian dictum that 'The species is the entire being of an individual.' As schematized in the so called Porphyrian tree, differences proper in the category of substance were seen to end with the final species and individuation was explained through something extrinsic, whether by way of accidents or matter. Scotus, however, extends the process of division and differentiation in the substantial line past the species and down into the constitution of the individual itself, going so far as to place the species and individual difference in a relation of potency and act akin to that of genus and specific difference. He thus accords the individual a true reality and admits as a consequence that the individual is *per se* intelligible. In so elevating both the reality and intelligibility of the individual, Scotus' realism on the issue of universals is decidedly un-Platonic.

13 Intuitive and abstractive cognition

In the area of epistemology, Scotus' most influential contributions were the distinction between intuitive and

abstractive cognition and the demolition of Augustinian illumination, at least in the highly sophisticated form given to it by Henry of Ghent. While the latter is of broader philosophical interest, virtually every scholastic discussion in epistemology after Scotus utilizes his distinction between abstraction and intuition, which contemporaries claimed originated with him. Scotus' notion especially of intuitive cognition was, of course, subjected to refinement and revision in subsequent discussions, but always with Scotus' original definition in mind.

As defined technically by Scotus, intuitive cognition is knowledge of an object insofar as it is actually existing and present to the intellect. Abstractive cognition is knowledge of the object insofar as it is abstracted from actual existence or non-existence. A number of clarifications are in order. First of all, as will be clear from Scotus' argument for the distinction, both intuitive and abstractive cognition are acts of the intellect proper and do not differ in that intuition grasps the sense particular and abstraction the universal. Both types of cognition have as their object the essence or quiddity as opposed to the sense particular. In intuition, the quiddity is known as being caused by what is existing and present, in abstractive cognition by the intelligible species residing in the intellect as surrogate for the existing object itself. In this context, then, 'abstractive' does not for Scotus refer to Aristotelian abstraction of the universal. Second, Scotus is specific that 'intuitive' is not here equated with 'non-discursive', the common epistemological sense associated with the Augustinian term *intuitus* (glance), particularly in the context of divine knowledge or the beatific vision (see Augustinianism). Some abstractive knowledge can be 'intuitive' in this sense, since it can be non-discursive. Scotus says that he is here taking 'intuitive' absolutely, as when we say that we 'see (*intueri*) a thing as it really is'.

Scotus argues that the intellect must possess both types of cognition based upon its commonly admitted functions. Thus, the intellect must be capable of abstractive cognition, for otherwise scientific knowledge in the strict Aristotelian sense would be impossible. The reason is that an object is contingent insofar as it is actually existent and present to the intellect. If therefore the intellect cannot grasp an object in abstraction from its existence, all knowledge would be contingent. In other words, the intellect could know no statements about an object as true or false independent of that object's existential state. Conversely, the intellect must be capable of intuitive cognition, for a perfection found in a lower power must be found in a higher power of the same type. However, the senses, which are cognitive faculties like the intellect, seize the sensible particular as present and existing. Therefore, the intellect must also have this capacity. As Scotus explains, the particular senses have intuitive, sensible cognition of the particular, while the imagination knows the same object abstractively by means of the sensible species which can remain in the absence of the sensible thing itself. The same twofold cognitive capacity must, by parity, be found in the intellect. At the level of sense, however, two separate powers are required for these two different cognitive acts because the sense powers are distinguished by having different material organs. Owing to its greater perfection as an immaterial power, the intellect possesses both capacities in a united way. Furthermore, intuitive cognition is also required to account for the beatific vision, where the divine essence will be known, according to scripture, 'face to face'; that is, as existentially present to the intellect.

For Scotus, then, the intellect has a direct apprehension of an intelligible object insofar as it is the actually existing and present cause of its cognitive act. The chief philosophical use to which Scotus puts intuitive cognition is to supply certitude for contingent propositions. For example, he claims that by means of intuitive cognition we are as certain about our own acts as we are about necessary, self-evident propositions. After Scotus, the entire fourteenth-century preoccupation with certitude was regularly cast in terms of intuitive cognition. For instance, a common problem discussed was whether God could cause an intuitive cognition of a non-existent object (see William of Ockham §4).

14 Theory of will

While perhaps better known for his metaphysics than his ethics, Scotus' ethical theory has attracted increasing attention for being innovative and even radical. For example, he departs from fundamental thirteenth-century positions by holding that not all of the natural law (the decalogue) is absolutely binding, that prudence is not necessarily connected to moral virtue and that the will can act contrary to a fully correct moral judgment of the intellect. These and other such conclusions arise from Scotus' strong notion of will itself, which is complex and the focus of much of the attention given to his ethical theory. Three features of Scotus' conception of will have been seen as particularly important: the will as a power for opposites, the will as rational power and the dual

inclination or ‘affection’ (*affectio*) of the will.

Scotus holds that there is a twofold freedom arising from the will as a power for opposites. The will is free in an evident way, says Scotus, since it is capable of opposite acts successively, such as loving and hating. This type of freedom, however, is not a perfection, since it pertains to the will as changeable and variable. In a second, less evident way, Scotus argues, the will is also free apart from any succession or change, for at the very moment at which it is willing an act, it remains a real, active power to will the opposite. Obviously Scotus does not mean by this that the will is capable of willing contrary acts simultaneously. Rather, he means that, if there is to be a contingent and free cause called the will, an act must be consistent with the real possibility of its opposite at the same time. (Scotus goes to considerable lengths to clarify the logical ambiguities of his position.)

Scotus argues for this ‘less evident’ sense of freedom and contingency by means of his famous hypothetical case of an instantaneously existent will, which in fact derived from the standard scholastic question of whether an angel could have sinned at the first instant of its creation. Consider a created will that has been brought into existence only for an instant and at that instant has a determinate act of willing. Scotus argues that, despite existing only for an instant, this will cannot produce its volition necessarily, but must do so freely and contingently. The reason is that a cause, when it actually causes, must do so either necessarily or contingently. That is, a cause is not now contingent because it existed previously and then, at that previous time, was able either to cause or not, but only because it is such at the moment when it actually operates. Thus, if a will existing at an instant causes necessarily, it would cause in that way at every instant, and thus never be a free or contingent cause. Therefore, since the will causes contingently and freely at that instant, it must have a real power for the opposite at that same instant. The will is thus a power for opposites apart from any succession, for there is no succession at an instant.

In arguing that the will is a power for opposites apart from any succession or change, Scotus departed from a long standing conception of free choice, such as represented by the standard discussion of freedom in Lombard’s *Sentences*. There, Lombard states that choice is not free with respect to what is past or present, but only with respect to the future. The reason is that what is present is already determined, and it is not within our power to make what already is not be the case. Rather, we are only free to change what will be in the future. Scotus denies this on the above grounds that it would render the will a necessary rather than contingent and free cause, for the causal nature of the will is determined only when it operates as a cause. To be free, therefore, the will must be contingently related to its act of volition even at the moment of that act. This means that the will must have a real power for the opposite of what it wills at that very moment (see Free will).

The notion of contingency resulting from this particular aspect of Scotus’ doctrine of will is regarded by some as his most important philosophical contribution of all. In this connection, Scotus is regularly portrayed as breaking with Aristotelian conceptions of modality that persisted until the scholastic period. As the above account indicates, something is contingent according to Scotus if, at the moment it occurs, there is a real possibility for its opposite. This is in contrast to Aristotle’s construction of contingency, where something is contingent if its opposite can actually occur at some other time. In this, Scotus is seen as ushering in a modern conception of possibility previously thought to have begun with Leibniz. While Scotus’ originality on this score has been overstated - the basic doctrine of will behind this new notion of contingency is found in Peter Olivi - there can be little doubt that the extended analysis given to it by Scotus ensured its influence.

In a position related to the above conception of the will, Scotus maintained that the will was a rational power. Commenting on the text of Aristotle in which rational powers are defined as those capable of producing contrary effects, Scotus made the primary division of all active powers the natural versus the voluntary. A natural agent is one that is of itself determined to act. That is, a natural power will issue in a determinate act necessarily and to its greatest capacity unless impeded. A voluntary or free power is not determined of itself to act, so that it may issue in a contrary act or no act at all. By this Scotus really means that the will is self-determining. Its indeterminacy to act is not a defect owing to an insufficiency of power but a perfection that results from an abundance of power capable of contrary effects. Given this primary division of nature and will, Scotus places the intellect on the side of natural powers so that, in Aristotle’s definition, it is not strictly speaking rational. The will consequently became the only truly rational power, where ‘rational’ was contrasted with ‘naturally determined’. In a complete reversal of the intellectualist and Aristotelian model accepted by Aquinas, Scotus concluded that the intellect was rational only in the qualified sense that it is required as a precondition for the action of the will.

The will, however, is not only an active power, but an appetite with inclinations. Here too, Scotus sought to protect the will from natural determinism by adopting Anselm's distinction between an affection or inclination for the advantageous (*affectio commodi*) and an affection for justice (*affectio iustitiae*) (see Anselm of Canterbury §6). As interpreted by Scotus, the former is the inclination to self-fulfilment characteristic of natural desire. What is sought is the perfection of the agent. The latter is an inclination not for the good of the agent but for the good in itself. Scotus claims that the will has an 'innate' affection for the just and that this is the basis of its liberty. The affection for the just enables the will to transcend the determination of natural appetite to self-fulfilment by loving the supreme good, God, for its own sake or other lesser goods for their own worth (see Right and good §2).

See also: Aristotelianism, medieval; Augustinianism; Being; Francis of Meyronnes; God, arguments for the existence of; God, concepts of; Henry of Ghent; William of Ockham

STEPHEN D. DUMONT

List of works

Duns Scotus, John (*c.*1290-1308) *Opera omnia. Editio nova iuxta editionem Waddingi XII tomos continentem*, Paris: Vivès, 1891-5, 26 vols; selections ed. and trans. A.B. Wolter, *Duns Scotus on the Will and Morality*, Washington, DC: Catholic University, 1987; ed. and trans. A.B. Wolter and W.A. Frank, *Duns Scotus, Metaphysician*, Lafayette, IL: Purdue University Press, 1995.(The Vivès edition is a typographically enlarged reprint of the Wadding edition (Lyons, 1639). Contains many spurious works. For the certainly authentic works, see C. Balić (1966) *John Duns Scotus. Some Reflections on the Occasion of the Seventh Centenary of his Birth*, Rome: Scotistic Commission, 29-44. Since the critical Vatican edition is far from complete, this is still the only text for many of Scotus' writings. Even for those texts which have been critically edited, the edition remains valuable for the scholia, parallel citations and commentaries by later Scotists. Wolter (1987) contains extensive selections with facing Latin and substantial introductions; Wolter and Frank provide a re-translation of selected texts with commentary.)

Duns Scotus, John (*c.*1290-1308) *Opera omnia studio et cura Commissionis Scotisticae ad fidem codicum edita*, Vatican City: Typis Polyglottis Vaticanis, 1950-, 11 vols to date. (Planned critical or Vatican edition of Scotus' writings. Published to date are vols 1-7 (*Ordinatio* to bk 2, dist. 3), vols 16-19 (*Lectura*).)

Duns Scotus, John (before 1300) *Lectura*, in *Opera omnia studio et cura Commissionis Scotisticae ad fidem codicum edita*, Vatican City: Typis Polyglottis Vaticanis, 1950-, vols 16-19.(The questions on divine foreknowledge are translated by A. Vos, *Contingency and Freedom: John Duns Scotus, Lectura I 39*, Dordrecht: Kluwer, 1994.)

Duns Scotus, John (after 1300) *Ordinatio*, in *Opera omnia studio et cura Commissionis Scotisticae ad fidem codicum edita*, Vatican City: Typis Polyglottis Vaticanis, 1950-, vols 1-7.(The prologue is translated by A. Wolter in 'Duns Scotus on the Necessity of Revealed Knowledge', *Franciscan Studies* 11: 231-71, 1951. The question on whether God's existence is self-evident is translated in E. Fairweather, *A Scholastic Miscellany: Anselm to Ockham*, Philadelphia, PA: Westminster Press, 1956, 428-39. The questions on the principle of individuation are translated by P.V. Spade in *Five Texts on the Medieval Problem of Universals*, Indianapolis, IN: Hackett Publishing Company, 1994, 57-113. *Duns Scotus, Philosophical Writings*, trans. A. B. Wolter, Indianapolis, IN, and Cambridge: Hackett Publishing Company, 1987, is a translation of selections from the *Ordinatio* with facing Latin text of the Vivès edition.)

Duns Scotus, John (*c.*1302-5) *Reportatio parisiensis*, part trans. A. Wolter and M. Adams, 'Duns Scotus' Parisian Proof for the Existence of God', *Franciscan Studies* 42: 248-321, 1982.(Translation and edition of the Parisian verison of the proof for the existence of God.)

Duns Scotus, John (*c.*1306) *Quodlibetal Questions*, ed. F. Alluntis, *Obras del Doctor Sutil, Juan Duns Escoto: Cuestiones Cuodlibetales*, Madrid: Biblioteca De Autores Cristianos, 1968; trans. F. Alluntis and A.B. Wolter, *God and Creatures: The Quodlibetal Questions*, Princeton, NJ: Princeton University Press, 1975; repr. Washington, DC: Catholic University of America Press, 1987.(Alluntis' edition is a revision of Vivès text of the *Quodlibetal Questions*. Alluntis and Wolter's translation includes a helpful glossary of Scotistic vocabulary.)

Duns Scotus (*c.*1308) *De primo principio (On First Principle)*, ed. and trans. A.B. Wolter, *John Duns Scotus: A Treatise on God as First Principle*, Chicago, IL: Franciscan Herald Press, 1966; 2nd revised edn, 1983. (An edition and translation of *De primo principio*. The revised edition adds an extensive commentary.)

References and further reading

In addition to the works noted below, the proceedings of the International Scotistic Congress (*Congressus Scotisticus Internationalis*) contain many articles on all aspects of Scotus' thought. See *De doctrina Ioannis Duns Scoti*, Rome: Cura Commissionis Scotisticae, 1968, 4 vols; *Deus et homo ad mentem I. Duns Scoti*, Rome: Societas Internationalis Scotisticae, 1972; *Regnum hominis et regnum Dei*, ed. C. Bérubé, Rome: Societas Internationalis Scotistica, 1978, 2 vols; *Homo et Mundus*, ed. C. Bérubé, Rome: Societas Internationalis Scotistica, 1981.

Adams, M.M. (1982) 'Universals in the Fourteenth Century', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 411-39. (Includes comparison of Scotus and Ockham on universals.)

Boler, J. (1993) 'Transcending the Natural: Duns Scotus on the Two Affections of the Will', *American Catholic Philosophical Quarterly* 67: 109-22.(Discussion of *affectio commodi* and *affectio iustitiae*.)

Brown, J. (1976) 'Duns Scotus on Henry of Ghent's Arguments for Divine Illumination: The Statement of the Case', *Vivarium* 14 (1): 94-113.(Detailed examination of Scotus' rejection of illumination.)

Brown, S. (1965) 'Avicenna and the Unity of the Concept of Being: The Interpretations of Henry of Ghent, Duns Scotus, Gerard of Bologna and Peter Aureoli', *Franciscan Studies* 25: 117-50.(History of the problem of univocity.)

Cress, D. (1975) 'Toward a Bibliography on Duns Scotus on the Existence of God', *Franciscan Studies* 35: 45-65. (Scholarship on this topic in Scotus is nearly the size of that for Descartes.)

Dumont, S.D. (1984) 'The *quaestio si est* and the Metaphysical Proof for the Existence of God according to Henry of Ghent and Duns Scotus', *Franziskanische Studien* 66: 335-67.(General metaphysical background in Henry of Ghent assumed by Scotus' proof.)

Dumont, S.D. (1987-9) 'The Univocity of Being in the Fourteenth Century', *Mediaeval Studies* 49: 1-75; 50: 186-256; (with S. Brown) 51: 1-129.(Technical and detailed history of univocity among Scotus' followers. Contains previously unedited texts.)

Dumont, S.D. (1988) 'The Necessary Connection of Prudence to the Moral Virtues according to John Duns Scotus - Revisited', *Recherches de théologie ancienne et médiévale* 55: 184-206.(Refutes Lottin's interpretation that Scotus subscribes to the Aristotelian theory of the connection of the virtues.)

Dumont, S.D. (1989) 'The Scientific Character of Theology and the Origin of Duns Scotus' Distinction between Intuitive and Abstractive Cognition', *Speculum* 64 (3): 579-99. (Argues primary application of the distinction is theological.)

Dumont, S.D. (1992) 'Transcendental Being: Scotus and Scotists', *Topoi* 11: 135-48. (Interpretation given to univocity by Scotus' followers.)

Dumont, S.D. (1995a) 'The Origin of Scotus' Theory of Synchronic Contingency', *The Modern Schoolman* 72 (January-March) 149-68.(Peter Olivi's influence on Scotus' doctrine of will.)

Dumont, S.D. (1995b) 'The Question on Individuation in Scotus' *Quaestiones in Metaphysicam*', in L. Sileo (ed.) *Vita Scoti. Methodologica ad mentem Joannis Duns Scoti*, Rome: Edizioni Antonianum, vol. 1.193-227. (Against the traditional view that the version of the theory in the *Metaphysics* is earlier than that of the *Sentences*.)

Frank, W.A. (1992) 'Duns Scotus on Autonomous Freedom and Divine Co-Causality', *Medieval Philosophy and Theology* 2: 142-64.(Interesting analysis of an acute difficulty for Scotus' notion of free will.)

Gilson, É. (1952) *Jean Duns Scot. Introduction à ses positions fondamentales (John Duns Scotus: Introduction to His Fundamental Positions)*, Paris: Vrin.(The most recent comprehensive book on Scotus' philosophy, but of limited value owing to its failure to take account of Henry of Ghent.)

Honnefelder, L. (1979) *Ens inquantum ens. Der Begriff des Seienden als solchen als Gegenstand der Metaphysik nach der Lehre des Johannes Duns Scotus (Being qua Being: The Concept of Being as such as the Object of Metaphysics in the Thought of John Duns Scotus)*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters, Neue Folge 16, Münster: Aschendorff.(Extensive work on Scotus' conception of the science of metaphysics.)

Ingham, M.E. (1989) *Ethics And Freedom: An Historical-Critical Investigation Of Scotist Ethical Thought*, Washington, DC: University Press of America.(Argues that Scotus' ethics emphasizes efficient rather than final cause.)

Kent, B. (1996) *Virtues of Will: The Transformation of Ethics in the Late Thirteenth Century*, Washington, DC: Catholic University of America Press.(Excellent study referring to Scotus throughout. Argues that Scotus is more moderate and indebted over his doctrine of will than commonly portrayed.)

King, P. (1992) 'Duns Scotus on the Common Nature and the Individual Difference', *Philosophical Topics* 20 (2): 51-76.(Detailed philosophical analysis of theory of individuation.)

Marenbon, J. (1987) *Later Medieval Philosophy (1150-1350)*, London: Routledge, 154-68.(Contains chapter on

Scotus' epistemology addressing issue of 'intellective memory'.)
Marrone, S. (1983) 'The Notion of Univocity in Duns Scotus' Early Works', *Franciscan Studies* 43: 347-95. (Traces evolution of the doctrine in Scotus, who appears to have initially denied univocity.)
Marrone, S. (1988) 'Henry of Ghent and Duns Scotus on the Knowledge of Being', *Speculum* 63 (1): 22-57. (Argues than Henry and Scotus are closer than usually admitted.)
Owens, J. (1948) 'Up to What Point is God Included in the Metaphysics of Duns Scotus?', *Mediaeval Studies* 10: 163-77.(Analysis of Scotus' difficult opening question in his *Metaphysics* on the subject of metaphysics.)
Owens, J. (1957) 'Common Nature: A Point of Comparison between Thomistic and Scotistic Metaphysics', *Mediaeval Studies* 19: 1-14.(Influential article on the treatment of Avicennian common nature by Aquinas and Scotus.)
Philosophy of John Duns Scotus in Commemoration of the 700th Anniversary of His Birth, (1965) special edition of *The Monist* 49.(Collection of articles.)
Prentice, R. (1968) 'The Voluntarism of Duns Scotus as seen in his Comparison of the Intellect and the Will', *Franciscan Studies* 28: 63-103.(Detailed examination of Scotus' analysis of Henry of Ghent's position that the will is higher than the intellect and Aquinas' position to the opposite.)
Ryan, J.K. and Bonansea, B. (eds) (1965) 'John Duns Scotus, 1265-1965', in J.K. Ryan and B. Bonansea (eds) *Studies in Philosophy and the History of Philosophy* 3, Washington, DC: Catholic University of America.(Very useful and now standard collection of articles.)
Schaefer, O. (1955) *Bibliographia de vita operibus et doctrina Ioannis Duns Scoti, Saec. XIX-XX (Bibliography of the Life, Works and Doctrine of John Duns Scotus)*, Rome: Orbis Catholicus-Herder.(Standard bibliography that includes anything remotely to do with Scotus.)
Schaefer, O. (1967) 'Resenha abreviada da bibliographia escotista mais recente (1954-1966)' (A Short List of More Recent Scotistic Bibliography), *Revistas Portuguesa de Filosofia* 23: 338-63.(Addendum to above.)
Sileo, L. (ed.) (1995) *Via Scoti. Methodologica ad mentem Joannis Duns Scoti (The Way of Scotus: Methodology in the Thought of John Duns Scotus)*, Rome: Edizioni Antonianum, 2 vols.(Massive collection of articles on Scotus in all languages.)
Wolter, A.B. (1946) *The Transcendentals and Their Function in the Philosophy of Duns Scotus*, St. Bonaventure, NY: The Franciscan Institute.(A study still regarded as the best introduction to Scotus' metaphysics.)
Wolter, A.B. (1990) *The Philosophical Theology of John Duns Scotus*, ed. M. Adams, Ithaca, NY: Cornell University Press.(Collection of many of Wolter's excellent and clear articles on Scotus covering main topics in his epistemology, metaphysics and theory of will.)
Wolter, A.B. (ed.) (1993) *Duns Scotus*, special issue of *American Catholic Philosophical Quarterly* 67.(Dedicated issue of the journal. Contain excellent and up-to-date introduction to Scotus' life and works by Wolter.)
Wood, R., Honnefelder, L. and Dreyer, M. (eds) (1996) *John Duns Scotus: Metaphysics and Ethics*, Leiden: Brill.(Collection of articles.)

Duran, Profiat (d. *c*.1414)

Duran, known also as Efodi, produced a wide variety of works displaying a considerable understanding of Christian culture, which he then used to criticize Christianity from a Jewish perspective. Heavily influenced by both Maimonides and Abraham ibn Ezra, he stressed the dual nature of the Torah as a system of belief and practice. His work includes Neoplatonic and astrological ideas, and sought to emphasize the salvific force of the Torah.

Profiat Duran's given name was Isaac ben Moses Levi, but he is perhaps best known as Efodi. He was a Spanish-Jewish theologian and polemicist of the later fourteenth and early fifteenth centuries. An active anti-Christian polemicist, he was also the author of two philosophical works, the 'Introduction' to his grammatical work, *Ma'aseh Efod (The Making of the Ephod)* and a popular commentary to Maimonides's *Guide to the Perplexed*. In the former, Duran maintains that the Torah teaches true science (logic, physics, metaphysics) and is the best way to learn it. Such knowledge is important, but true felicity in this world and in the next depends upon proper fulfillment of the Torah's commandments (*mitzvot*). Duran deals with the different ways in which thinkers have analysed how the Torah is to be observed. Some had argued that all that is important is carrying out the *mitzvot*. It is not necessary to know why these commandments ought to be followed beyond recognizing that they are commandments, on the model of medical treatment where someone may not understand the workings of the treatment, yet undertake it and as a result be cured of the ailment. Duran disagrees with this view, arguing that the Torah has two parts, knowledge and practice. These parts are intimately connected, since it is necessary to have knowledge to understand how to carry out the commandments. Duran criticizes the view that religion and philosophy are in opposition to each other. Carrying out one's religious obligations, he argues, serves to attract the emanation of heaven, along the Neoplatonic pattern. Worship achieves the heights of perfection only in the Land of Israel, since it lies closer to the powerful forces of grace than anywhere else in the world. Nonetheless, the commandments must be obeyed wherever one is, and their fulfillment brings about happiness in this life.

Duran considers the claims of the Kabbalists and concludes that they are closer to the spirit of the Torah than are the philosophers, even more so if they are right in their claims to be able to affect the course of nature through their practices. Yet they also disagree among themselves a great deal, which is hardly an indication of a secure repository of truth. So, although philosophy and Kabbalah both have merits, the proper object of study is really the Torah, which is the only sure route to eternal happiness. In the Torah one is working with a holy language, and in studying it one participates in the holiness of the names which are to be found there. When the Temple existed, it was the conduit through which the divine emanation flowed to the people of Israel, and when it was destroyed (because the Jews had stopped studying Scripture) it was replaced by the Torah. The Torah is the only path to the highest happiness, and the only route to God is through carrying out its commandments. Theoretical knowledge in itself is incapable of attaining this eternal and complete happiness, and only the Jews are really able to carry out the proper role of human beings in serving God through carrying out his commandments. The Torah represents the source of divine providence for the Jews and guarantees their survival, despite the many dangers and disasters which they undergo. Duran was no doubt reflecting from his own experience here, for he had been forced to convert to Christianity and was able to practice openly as a Jew only later in his life. Although he defended stoutly the study of the Torah, he also accepted the importance of the natural sciences so long as they do not contradict the foundations of the faith.

Duran's *Commentary on the 'Guide to the Perplexed'* is printed in the standard editions of that work. It is brief, literal and very conservative, trying to present Maimonides as a thinker well within the mainstream of received Judaism. In places where Maimonides seems to interpret Judaism in too thoroughly Aristotelian a manner, Duran sands down the rough edges, pulling Maimonides back in the direction of traditionalist orthodoxy. On the other hand, he frankly acknowledges the dangers of some of Maimonides's arguments as far as religion is concerned. His conservatism is in many ways typical of the Jewish philosophers of the late fourteenth and early fifteenth centuries. He did, after all, have to respond both to the continual spiritual crises in the Spanish Jewish community as a result of the conversionary pressures of the Christians, and to the impact of the Jewish Averroist movement (see Averroism, Jewish). On the other hand, his arguments for the role of the Torah in reconciling the physical and spiritual aspects of humanity was to play a large role in Jewish philosophy in the fifteenth century on its approach

to modernity.

See also: Ibn Ezra, A.; Maimonides, M.

MENACHEM KELLNER
OLIVER LEAMAN

List of works

Duran, P. (before *c*.1414) *Commentary on the 'Guide to the Perplexed'*, Venice, 1551.(Duran's treatment of the *Guide*.)

Duran, P. (before *c*.1414) 'Introduction' to *Ma'aseh Efod*, ed. D. Raffel in *Sinai* 100, 1987: 749-95.(Duran's defence of the Torah.)

References and further reading

Lasker, D. (1977) *Jewish Philosophical Polemics Against Christianity in the Middle Ages*, New York: Ktav.(An excellent account of Duran's anti-Christian writings.)

Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Cambridge: Cambridge University Press, 352-7.(General discussion of Duran's life and works, linking him closely with Abraham ibn Ezra and Moses Maimonides.)

Duran, Simeon ben Tzemach (1361-1444)

Simeon Duran was chiefly a religious thinker who incorporated a variety of philosophical traditions into his thought. He argues that revelation is the only certain route to knowledge. Following this principle, he criticizes those who argued that some of the principles of religion are more important or basic than others. To reject any aspect of religious law is to abandon the whole, and on this point Duran sets himself against Maimonides.

1 Life and works

Simeon ben Zemah Duran, Spanish *halakhist* and theologian, fled the anti-Jewish outbreaks of 1391 and spent the rest of his life in North Africa. Duran is best known as the author of numerous *responsa* (written answers to *halakhic* inquiries), from which much can be learned concerning the history of the Jews of Spain and North Africa in general and in particular of conversos (Jews forcibly converted to Christianity). He is also the author of a number of theological and polemical works that reflect a moderate Maimonideanism: Duran accepts the basic naturalism of Maimonides's Aristotelian view of the world while moderating its intellectualism. Duran's polemic against Christianity and Islam (*Keshet u-Magen*) is not philosophical in character. His most important contribution to Jewish thought is in the area of dogmatics, found primarily in two commentaries, *Ohev Mishpat (Lover of Justice)* - a work on Job and *Magen Avot (Shield of the Fathers)* - a work on Tractate Avot. The introduction to the former includes an extensive discussion of divine providence.

2 Dogma and the nature of Judaism

The issue of dogma had been introduced into medieval Judaism by Moses Maimonides (1138-1204), who formulated thirteen articles of belief in his commentary on the Mishnaic tractate Sanhedrin. Maimonides's enormously influential account of these doctrines of Judaism was presented, in effect, as an answer to the question: who is a Jew? In his discussions, Duran approaches the issue from a number of perspectives, not one of them that of Maimonides. In one place, Duran asks a structural question: construing Judaism as a deductive science, what are its axioms? His answer is: creation and providence. Providence, however, can be derived from creation, so in fact creation is the only basic principle of Judaism. In a second place, Duran seeks to explicate and defend Maimonides's account of the Jewish creed, reducing his thirteen dogmas to three: God's existence, Torah from heaven (that is, divine revelation as manifested in the giving of the Torah), and divine retribution. This account seems to have decisively influenced Joseph Albo (fifteenth century) in his *Sefer ha-'Iqqarim (Book of Principles)*. This second account, however, is not merely a summary of Maimonides, since it depends upon and even derives from Duran's third discussion, which diverges crucially from Maimonides. In this discussion Duran makes the claim that Judaism has only one principle - to believe what the Torah teaches - or, alternatively, that Judaism has as many principles as there are commandments in the Torah. It is in the context of this discussion that Duran defines heresy, not as false belief, but as the conscious rejection of a belief known to be taught by the Torah or the conscious adoption of a belief known to be denied by the Torah. The relative liberalism represented by Duran's departure from the idea of a fixed set of dogmas constitutive of Jewish faith is extremely rare in medieval religion.

Duran's three accounts answer three different questions:

(1) Construing Judaism as a deductive science, what are its axioms?
(2) Why did Maimonides posit the beliefs he did in his 'Thirteen Principles'?
(3) What is heresy?

Many of Duran's ideas were new and unprecedented but his influence was modest at best, both because his views were not adequately understood in his own day and because he wrote near the end of the period of active medieval Jewish philosphical discussion.

See also: Maimonides, M.

MENACHEM KELLNER

List of works

Duran, S. ben Tzemach (1361-1444) *Ohev Mishpat (Lover of Justice)*, Venice, 1590.(A commentary on the Book of Job.)

Duran, S. ben Tzemach (1361-1444) *Keshet u-Magen (Bow and Shield)*, Livorno, 1762.(A defence of Judaism against the attacks of Christianity and Islam.)

Duran, S. ben Tzemach (1361-1444) *Magen Avot (Shield of the Fathers)*, Livorno, 1785.(A commentary on the mishnaic tractate *Avot*.)

References and further reading

Albo, J. (*c*.1425) *Sefer ha-'Iqqarim (Book of Principles)*, ed. and trans. I. Husik, *Sefer ha-'Iqqarim: Book of Roots*, Philadelphia, PA: Jewish Publications Society, 1946, 5 vols.(Presentation of the dogmas of Judaism according to Albo.)

Bleich J. (1929) 'Duran's view of the Nature of Providence', *Jewish Quarterly Review* 69: 208-25.(Summary of Duran's theory of providence.)

Kellner, M. (1986) *Dogma in Medieval Jewish Thought*, Oxford: Oxford University Press, ch. 3.(Contains translations of Duran's writings on dogma and analyses of them; the notes contain full references to studies on Duran.)

Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Cambridge: Cambridge University Press. (Chronological history of Jewish philosophy in the Middle Ages.)

Durandus of St Pourçain (1275?-1334)

Although strongly Aristotelian in outlook, Durandus rejected certain classic points of Thomist doctrine such as the speculative 'scientific' and unique nature of theology, the theory of the active intellect and the doctrine of species. After carrying out a detailed examination of the Thomist theses, Durandus decided firmly in favour of theology being faith in revelation, something aenigmatica *and therefore meritorious. In this way, in the opinion of some modern scholars, he anticipated the position of William of Ockham; the position later known as 'Ockham's razor' appears more than once in his writings.*

The date of Durandus' birth is uncertain, but is believed to be about 1275. Early references speak of his presence in the convent of St Jacques in Paris, where he obtained a *licentia docendi* in 1312, and of the fact that as early as 1306 he had been criticized by a brother Dominican, Hervaeus Natalis, for an assertion contained in his *Commentariorum* (commentary on the *Sentences* of Peter Lombard), a work that was probably unfinished at the time but was no doubt already in circulation among the 'curious', as Durandus himself complained. In 1313, Pope Clement V made Durandus lector of the Holy Palace, and four years later, Pope John XXII made him bishop of Limoux, thus freeing him from his vow of obedience to the Dominican Order. In 1318 he became bishop of Le Puy and, in 1326, bishop of Meaux. He died in Meaux in 1334.

Following further criticism of his *Commentariorum* in 1314 and 1316, the Dominican Order carried out two formal inquiries, which concluded that certain theses contained in the work were dangerous and likely to impair that unity of doctrine which legitimized the intellectual independence of the Dominicans. At this time the teaching of the doctrines of Aquinas was becoming increasingly dominant; Thomist doctrine was defended repeatedly as being 'the soundest', while anyone who 'acted in a contrary manner and in spite of warnings failed to withdraw from his position' was severely censured. The disagreement between Durandus and his order continued, and the second edition of his commentary on the *Sentences* appeared in a reduced and expurgated form in 1310-11. In a later third edition (1317?-27), the censured points were to some extent restored.

In addition to three editions of his commentary on the *Sentences*, Durandus' works include the *Quodlibeta*, the *Quaestiones de habitibus (Questions About States)*, *De libero arbitrio (On Free Will)* and *De natura cognitionis (On the Nature of Cognition)*, written in Paris, three further *Quodlibeta* written in Avignon, and the *De origine jurisdictionum (The Origins of Jurisdictions)*, written while he was bishop of Meaux. In 1333, after being consulted by the pope and asked to express his opinion on the theory of the beatific vision, Durandus wrote *De visione Dei (The Vision of God)*, but this failed to win the approval of John XXII, who submitted it to a committee of censors.

The originality of Durandus as a theologian lies in the fact that he remained fundamentally faithful to Aristotelian-Thomist subjects and terminology, while diverging from that tradition in his conclusions. At the same time, he appears to have stood outside the anti-Thomist Augustinian tradition of the Franciscans (see Augustinianism). The chief problem for the historian is how to interpret the accusations of 'aversion to Thomism' which his contemporaries launched against him and which were later repeated in sixteenth century editions of the *Commentariorum* (Lyon 1558, Venice 1589). There is a theoretical consistency in Durandus' opposition to Aquinas, which emerges particularly over the question of the scientific nature of theology discussed in the commentary, showing that Durandus possessed his own clear-cut epistemology. Articles of faith, which Aquinas likened to the principles of knowledge, in Durandus' view are not 'rational', nor can they be referred to other self-evident principles, but 'only believed'. Faith is unconnected both with knowledge based on reasoning and with that obtained by way of experience. The subject of theology is supernatural and therefore not to be examined by man with his natural cognitive faculty; and it follows that the theological process itself cannot be rational. ' It is unreal to postulate a supernatural light that does not exist in all men and cannot be given to those who do not possess it' (*Commentariorum* 1, f.5 v.17).

Such statements are linked with fundamental points of Durandus' metaphysics and epistemology. The intellectual act itself is able to apprehend reality without the mediation of images (the 'species' of the Thomists), and *in se* can perceive individuals. Universals do not exist since everything that exists is individual. The formal distinction drawn by Duns Scotus between real beings and conceptual beings is 'frivolous' and leads to the conclusion that

man's knowledge of God is purely nominal. Truth may be defined as the formal condition of the object of knowledge, the relation between being and being perceived. It may be added that more than once in the course of discussion Durandus appealed to what was later to be known as Ockham's razor by asserting that *frustra ponuntur plura ubi pauciora sufficiunt* (it is futile to do with more what can be done with fewer) (see William of Ockham).

Durandus' principal thesis is that the theology of a human being who is a *viator* (one who has not yet attained the beatific vision) is based on belief in Scripture, which speaks in the language of metaphor capable of imparting not knowledge but moral teaching. Theology, Durandus remarks, is like rules of navigation which must be observed by sailors if they are to voyage in safety even 'if the stars are not of interest to them in themselves'. If the theologian exceeds these limits and launches into demonstrations that are no more than apparent, 'he assumes the capacity of the philosopher and loses his definition of theologian'. For Durandus, theology was an enigmatic body of knowledge and therefore meritorious for men, since ' it is not true that theologians or simple believers know the will of God and from this draw further knowledge' (*Commentariorum* f.11 r.10).

See also: Aristotelianism, medieval

MARIATERESA FUMAGALLI BEONIO-BROCCHIERI

List of works

Durandus of St Pourçain [Durandus de Sancto Porciano] (before 1313) *Quaestiones de habitibus (Questions About States)*, ed J. Koch, *Durandi de Sancto Porciano, O.P. Tractatus de habitibus, quaestio quarta [De subiectis habituum]*, Opuscula et textus historiam ecclesiae eiusque vitam atque doctrinam illustrantia. Series scholastica 8, Monasterii: Typis Aschendorff, 1930.(Edition of Durandus' *Quaestiones*.)

Durandus of St Pourçain [Durandus de Sancto Porciano] (1313?-17) *Quodlibeta (Quodlibetal Questions)*, ed. P.T. Stella, *Qudolibeta avenionensia tria, additis correctionibus Hervei Natalis supra dicta Durandi in primo quolibet*, Textus et studia in historium scholasticae, cura Pontificii Athenaei Salesiani I, Zurich: Pas-Verlag, 1965; *Quodlibeta Paris I (Q.I-Q.IV)*, ed. T. Takada, Series of Hitherto Unedited Texts of Medieval Thinkers IV, Kyoto, 1968.(Editions of Durandus' quodlibetal questions.)

Durandus of St Pourçain [Durandus de Sancto Porciano] (1317?-27) *Sententias theologicas Commentarium (Commentary on the Sentences)*; Latin edition published as *D. Durandi a Sancto Porciano In Petri Lombardi Sententias theologicas Commentarium libri 4*, Venice, ex typographia Guerraea, 1571; facsimile reprint, Ridgewood, NJ: Gregg, 1964.(This is an edition of the later third edition of the *Commentarium*.)

Durandus of St Pourçain [Durandus de Sancto Porciano] (1333) *De visione Dei (The Vision of God)*, ed. G. Cremascoli, 'Il "Libellus de visione dei" di Durando di S. Porziano', *Studi medievali* Ser. III 25, 1984, 393-443. (Durandus on beatific vision.)

References and further reading

Fournier, P. (1938) 'Durand de S. Pourçain', Histoire littéraire de la France, vol. XXXIV, Paris: Imprimerie Nationale.(Study of his life and works.)

Fumagalli Beonio-Brocchieri, M. (1969) *Durando di S. Porziano. Elementi filosofici della terza redazione del 'Commento alle Sentenze' (Durandus of St Pourçain: Philosophical Elements in the Third Redaction of the Commentary on the Sentences)*, Florence: La Nuova Italia.(An account of the principal philosophical and theological conceptions in the *Commentary*.)

Godet, P. (1924) 'Durand de S. Pourçain', *Dictionnaire de théologie catholique*, vol. IV, Paris: Letouzey et Ane, 1964-6. (Short biography.)

Hervaeus Natalis (*c*.1308) *De relatione contra Durandum et Durandi de S. Porciano, O.P., Quol. Paris. I. Q. 2 (3)*, ed. T. Takada, Kyoto, 1966.(Hervaeus' criticism of Durandus' commentary on the Sentences.)

Hinnebusch, W.A., and Bedouelle, G. (1990) *Brève histoire de l'ordre dominicain (Short History of the Dominican Order)*, Paris: Éditions du Cerf.(History of the Dominican Order including passages on Durandus.)

Koch, J. (1927) *Durandus de S. Porciano, O.P.: Forschungen zum Streit um Thomas von Aquin zu Beginn des 14. Jahrhunderts (Introduction to the Opposition to Thomas Aquinas at the Beginning of the Fourteenth Century)*, vol. 1: *Literargeschichtliche Grundlegung*, Beiträge zur Geschichte der Philosophie des Mittelalters: Texte und Untersuchungen 26, Münster: Aschendorff.(A fundamental study of Durandus' life, works and thought).

Mandonnet, P. (1947) 'Frères prêcheurs', *Dictionnaire de théologie catholique*, Paris: Letouzey et Ane, vol. 6, 863-924.(Short history of the Dominican Order.)

Martin, R.M. (1930) *La controverse sur le péché originel au début du XIV^e^ siècle. Textes inédits (The Dispute About Original Sin at the Beginning of the Fourteenth Century)*, Louvain: 'Spicilegium sacrum lovaniense'.

(Includes the views of Durandus.)
Müller, E. (1941) 'Quaestionen der ersten Redaktion von I und II Sententiarum des D. de S.P.' (Questions about the First Version of Books I and II of the *Sentences* by Durandus of St Pourçain), *Divus Thomas*, Freiburg: Albertinum, 435-40.(Study of Durandus' commentary on the *Sentences*.)
Synan, E.A. (1990) Sensibility and Science in Mediaeval Theology: the Witness of Durandus of Saint-Pourçain and Denis the Carthusian, in R. Työrinoja, A.I. Lehtinen and D. Føllesdal (eds) *Knowledge and the Sciences in Mediaeval Philosophy: Proceedings of the Eighth International Congress of Mediaeval Philosophy (SIEPM) Helsinki 24-9 August 1987*, Helsinki: Finnish Society for Missiology, vol. III.(Durandus' views on knowledge.)
Tabarroni, A. (1990) *'Paupertas Christi et apostolorum'. L'ideale francescano in discussione (1322-1324) (The Poverty of Christ and the Apostles: The Dispute about the Franciscan Ideal 1322-4)*, Rome: Istituto Storico Italiano per il Medio Evo.(The early fourteenth-century dispute on apostolic poverty.)

Durkheim, Émile (1858-1917)

Émile Durkheim is generally recognized to be one of the founders of sociology as a distinct scientific discipline. Trained as a philosopher, Durkheim identified the central theme of sociology as the emergence and persistence of morality and social solidarity (along with their pathologies) in modern and traditional human societies. His distinctive approach to sociology was to adopt the positivistic method in identifying and explaining social facts - the facts of the moral life. Sociology was to be, in Durkheim's own words, a science of ethics.

Durkheim's sociology combined a positivistic methodology of research with an idealistic theory of social solidarity. On the one hand, Durkheim forcefully claimed that the empirical observation and analysis of regularities in the social world must be the starting point of the sociological enterprise; on the other hand, he was equally emphatic in claiming that sociological investigation must deal with the ultimate ends of human action - the moral values and goals that guide human conduct and create the essential conditions for social solidarity. Accordingly, in his scholarly writings on the division of labour, on suicide, on education, and on religion, Durkheim sought to identify through empirical evidence the major sources of social solidarity and of the social pathologies that undermine it.

1 Intellectual biography

Émile Durkheim was born in Epinal (France) in 1858 to a three-generation family of Jewish rabbis. From an early age he was being trained to become a rabbi, but in his youth he first had a passing religious crisis, and later broke away from Judaism and became agnostic. Durkheim was trained in philosophy at the École Normale Supérieure in Paris, influenced by such thinkers as Charles Renouvier and Émile Boutroux, and obtained his *agregation de philosophie* in 1882. He taught philosophy at various Parisian *lycées* early in his career. In 1887 he first taught and later became professor of social sciences at the University of Bordeaux, a position he held until 1902 when he became first professor of education and later professor of education and sociology at the prestigious Sorbonne in Paris until the year of his death.

During his tenure at the University of Bordeaux, Durkheim published in quick succession three major works that qualified him as one of sociology's founders: *The Division of Labor in Society* (1893), *The Rules of Sociological Method* (1895), and *Suicide: A Study in Sociology* (1897). During that same period Durkheim established the *Année sociologique,* France's first sociology journal, and directed a social science research group that gravitated around the journal. Durkheim's years at the Sorbonne were characterized by a focus on the broad philosophical and ethical themes he recognized to be at the centre of the sociological enterprise. He lectured extensively on his philosophy of human nature, on value judgments in the social sciences, on epistemology, on the history of pedagogy, on moral education, and on religion. His last major monograph, *The Elementary Forms of the Religious Life* was published in 1912. Greatly saddened by the loss of his own son and close friends during the First World War, Émile Durkheim died in 1917, aged 59.

2 The sociological method

In his attempt to establish sociology as an academic discipline, Durkheim sought to identify early in his work the distinctiveness of the sociological approach and its unique methodology of sociological research. First, Durkheim established that society was not simply equivalent to the sum of its individual members; using a chemical metaphor, he viewed it as an organic compound endowed with qualities that surpassed the qualities of its constituting elements. While advocating that sociology should follow the path of the natural sciences and life sciences in developing a scientific standing, Durkheim realized that social facts - the phenomena of the moral life - were very different from the phenomena observed in the natural world. Social facts were representations of the collective life of society: laws, religious beliefs, norms, myths, customs. But the sociologist should treat them with the same empirical scrutiny natural scientists afforded to the phenomena of nature. According to Durkheim, we must treat social facts as things, and to identify social facts we must rely on the two characteristics they display: they are external to individuals and they constrain them. Social facts are obligatory, and their transgressions are punished with appropriate social sanctions.

Social facts are independent of individuals, and as such they are not affected by individuals' behaviour. The causes

of social facts, instead, must be sought in antecedent social facts; for instance, the social fact of punishment is causally explained by the strong collective sentiments felt by a society against a particular kind of behaviour. At the same time, social facts also perform a function in society - they are the means for the achievement of societal ends. The social fact of punishment, for instance, seeks to maintain and reinforce the collective sentiments that sanction the improper conduct of society's members. Durkheim acknowledged the importance of investigating both the causes and the functions of social facts, but he emphasized that functional analysis is distinctively sociological in that it seeks to relate a social fact to the larger social organism.

Following through his metaphor of society as an organism, Durkheim classified social facts as normal or pathological, in terms of their relation with the larger social environment - with the social organism as a whole. The adequacy of a social fact like a social norm must be evaluated in terms of the social ends such a norm fosters - this requires that the sociologist evaluate the desirability of social ends. Durkheim argued that value judgments are inevitable in the social sciences, since the facts of society are the values and beliefs of the collectivity (see Value judgments in social science §1). Thus, a scientific sociology must be able to identify desirable social goals before it can identify the best means to achieve them. Here, I believe, Durkheim's positivistic methodology and idealistic social theory show their greatest strain as he tries to make value judgments amenable to direct scientific investigation (see Positivism in the social sciences). Such strain is reflected in Durkheim's substantive sociological research on the division of labour and on suicide.

3 Social solidarity and social pathologies

Durkheim's empirical research on the division of labour and on suicide demonstrated the potential of sociology as an academic discipline by taking up topics that had been previously analysed by other social sciences, and showing that sociology's distinctive approach shed a new light on them.

The division of labour was considered an economic phenomenon that led to increased production and wealth. Durkheim argued that the division of labour was also an important moral phenomenon that described the changing foundations of social solidarity from traditional to modern societies. On the one hand, traditional societies exhibited relatively undifferentiated roles of individuals in economic life; individuals were able to perform a variety of tasks and were basically interchangeable with each other. Social solidarity in traditional societies emerged from the unmediated identity of its members. Durkheim called this traditional form of social solidarity mechanical solidarity. On the other hand, modern societies are characterized by a high degree of differentiation of tasks in economic life, and individuals are highly interdependent in their economic activities. Social solidarity in modern societies emerges from the functional interdependence of individuals' differentiated roles; this modern form of social solidarity he calls organic solidarity. Thus, the increase in the division of labour was not detrimental in modern societies, but rather provided a new organic source of solidarity to replace the waning mechanical solidarity of traditional societies. But Durkheim was also concerned with the social pathologies manifested in highly differentiated societies; specifically, he identified a condition of 'anomic' division of labour whereby the norms that bind the economic activities of individuals are greatly weakened. The anomic division of labour led to a loss of social solidarity, and society as a whole suffered from it. Guided by his idealistic standard of social solidarity, Durkheim failed to consider the possibility that the weakening of social norms was in fact a normal and inevitable feature of modernity.

Suicide was another topic Durkheim claimed as rightfully sociological. While psychological, biological, and ecological explanations had been provided, Durkheim argued that suicide was also a significant social phenomenon to be explained sociologically. Suicide rates illustrated how insufficient or excessive regulation and integration in the social organism led to a predictable quota of voluntary deaths, year after year. Durkheim was particularly concerned with the social pathologies of modernity, as he identified two pertinent types of suicide: egoistic suicide and anomic suicide. Egoistic suicide was prompted by the insufficient integration and attachment of individuals to society, demonstrated by the higher rates of suicide observed in Protestant countries as compared to Catholic countries. Anomic suicide was prompted by society's weakened regulation of individuals' conduct, especially in economic life, as demonstrated by the higher rates of suicide found in industrial societies as compared to agricultural societies. Egoistic and anomic suicide shared the spotlight in Durkheim's study because they both illustrated his concern with the underlying pathology of modernity: the loosening of individuals' attachment to society and the breakdown of the social norms that should restrain individuals' asocial behaviour. As

with his theory of the anomic division of labour, Durkheim's theory of egoistic and anomic suicide failed to consider suicide as, possibly, a normal phenomenon of our times, deriving from the collectivity's value of the freedom and autonomy of individuals as the cornerstone of modernity.

4 Moral education and religion

In the last two decades of his life Durkheim turned his attention away from the pathologies of modernity, to focus on the central role of education and religion in creating, renewing, and sustaining social solidarity.

In his lectures on moral education (published posthumously in 1925) Durkheim argued that education should provide individuals with discipline, attachment to society, and autonomy of judgment. The first element guaranteed society's successful role in regulating the conduct of its members; the second gave individuals a sense of belonging and of successful integration with the collectivity; the third equipped the competent member of society with the ability autonomously to discern the desirability of society's moral values and to subscribe to them spontaneously. The themes of insufficient regulation and social integration that characterized Durkheim's work on suicide were recast here in a positive form: how to achieve successful regulation and social integration in modern societies. The theme of autonomy, then, echoed Durkheim's repeated emphasis that the dignity of the individual was unquestionably the central moral value of modernity.

Durkheim's work on religion brought to full circle his characterization of society as centred on collective representations. Criticizing the Marxist approach, Durkheim rejected the view that economic life preceded religious phenomena; instead he argued that in primitive societies everything was religious. In *The Elementary Forms of the Religious Life* Durkheim demonstrated the centrality of religion in providing the cognitive, intellectual and moral foundations of social life. He argued that since religious phenomena were the earliest expression of social life they were also the source of our rules of logic and of our system of thought. The categories of time and space, of class and number, and of genus and species must have originated in the early experience of the collectivity, manifested through religious life. Thus, for example, the idea of time and of its division of years and seasons emerged, in Durkheim's view, from the religious distribution of ceremonies, rituals and holidays in primitive societies.

By demonstrating that cognitive and moral phenomena found their roots in primitive religious life Durkheim sought to prove that everything is social in its origins. This was, in the end, Durkheim's master thesis, that society is the source of everything that makes us human - moral rules, altruism, rules of logic, cognitive categories, and ultimate beliefs and values. Accordingly, he constantly downplayed the role of psychology, of economics, of history, and constantly emphasized the role of collective representations in shaping the life of society and of its members. Perhaps Durkheim's philosophy was exceedingly sociocentric; it was certainly unambiguous and forceful in making the case for sociology as the science of moral life.

See also: Religion and political philosophy; Social science, methodology of §2

MARCO ORRÚ

List of works

Durkheim, É. (1893) *De la division du travail social*, Paris: Alcan; trans. G. Simpson, *The Division of Labor in Society*, New York: Macmillan, 1933.(Discussed in §3 of this entry.)

Durkheim, É. (1895) *Les regles de la methode sociologique*, Paris: Alcan; trans. S.A. Solovay and J.H. Mueller, *The Rules of Sociological Method*, Chicago, IL: University of Chicago Press, 1938.(Discussed in §2 of this entry.)

Durkheim, É. (1897) *Le Suicide: etude de sociologie*, Paris: Alcan; trans. J.A. Spaulding and G. Simpson, *Suicide: A Study in Sociology*, Glencoe, IL: Free Press of Glencoe, 1951.(Discussed in §3 of this entry.)

Durkheim, É. (1912) *Les Formes elementaires de la vie religieuse*, Paris: Alcan; trans. J.W. Swain, *The Elementary Forms of the Religious Life*, New York: Macmillan, 1915.(Discussed in §4 of this entry.)

Durkheim, É. (1925) *L'Education morale*, Paris: Alcan; trans. E.K. Wilson and H. Schnurer, *Moral Education*, New York: Free Press, 1961.(Discussed in §4 of this entry.)

References and further reading

Fenton, S. (1984) *Durkheim and Modern Sociology*, Cambridge and New York: Cambridge University Press.

(Discusses Durkheim's sociology within its disciplinary context.)
Hall, R.T. (1987) *Émile Durkheim: Ethics and the Sociology of Morals*, New York: Greenwood Press.(Systematic analysis of Durkheim's thought on morality and ethics.)
Jones, R.A. (1986) *Émile Durkheim: An Introduction to Four Major Works*, Beverly Hills, CA: Sage.(An introduction to *Division of Labor*, *Rules*, *Suicide* and *Elementary Forms*.)
LaCapra, D. (1972) *Émile Durkheim: Sociologist and Philosopher*, Ithaca, NY: Cornell University Press. (Discusses Durkheim's major works and highlights their philosophical framework.)
Lukes, S. (1973) *Émile Durkheim: His Life and Work*, New York: Harper & Row. (Exhaustive presentation of Durkheim's work within the context of his intellectual biography.)
Parkin, F. (1992) *Durkheim*, Oxford and New York: Oxford University Press.(A short introduction to Durkheim.)
Pickering, W.S.F. (1984) *Durkheim's Sociology of Religion*, London: Routledge & Kegan Paul.(A detailed treatment of themes and theories in Durkheim's sociology of religion.)
Pope, W. (1975) *Durkheim's Suicide: A Classic Analyzed*, Chicago, IL: University of Chicago Press.(Analyses *Suicide*'s theory and data.)
Thompson, K. (1982) *Émile Durkheim*, New York: Tavistock.(An introductory overview of Durkheim.)
Turner, S.P. (ed.) (1993) *Émile Durkheim: Sociologist and Moralist*, London: Routledge.(A collection on Durkheim's science of morality.)
Wallwork, E. (1972) *Durkheim: Morality and Milieu*, Cambridge, MA: Harvard University Press.(Describes the evolution and shifts in Durkheim's science of ethics.)

Duty

To have a duty is, above all, to be subject to a binding, normative requirement. This means that unless there are exculpating reasons someone who has a duty is required to satisfy it, and can be justifiably criticized for not doing so. Having a duty to do something is like having been given a command by someone who has a right to be obeyed: it must be done.

Sometimes we speak as if we have duties to individuals (such as persons and institutions). So, for example, if Jones makes a promise to Smith, then Jones has a duty to Smith to keep the promise. We also talk as if we have duties to perform, or to refrain from performing, types of actions, for instance, a duty to help those in need. Even if the performance of such a duty involves treating an individual in a certain way, the duty may not be to that individual. For example, a duty to be charitable might not be a duty to anyone, not even the recipient of the charity.

An important feature of duties is that they provide some justifying reason for action. If we explain why we did something by saying that it was our duty, we are offering a justification for the action. Such a justifying reason does not depend on the entire nature of the action. For example, if we make a promise, we have some justifying reason for keeping it, regardless of what was promised, or to whom the promise was made. Again it is like a command. If we are given a command by someone with a right to be obeyed, we have some justification for obeying it, no matter what we are commanded to do. On some views, however, the justifying reason we have for doing something because it is required by duty may not be decisive: we may have an even better reason for doing something else. None the less, that something is required by duty provides some justifying reason for doing it.

Talk about duties is found in many areas; we speak, for example, of legal duties, moral duties, professional duties, the duties of a scholar and, even, matrimonial duties. This discussion will focus on moral duty, but may have wider application.

1 Duty and other normative concepts

Related to the concept of duty are a number of other important normative concepts. Among these are the concepts of obligation, right, permission and prohibition. The relations between permissions, prohibitions and duties are fairly straightforward: if we are permitted to do something, then we do not have a duty not to do it; if we are prohibited from doing something, we have a duty not to do it (see Deontic logic). The relationships between duties, obligations and rights are not as obvious.

According to most modern use, obligation and duty are taken to be coextensive, if not identical. So, to have a duty to do something is the same as to have an obligation to do it. However, there is a narrower view concerning duty. According to this, duties must result from roles, namely, from status, occupation or position. For example, subjects are said to have duties to their sovereign, physicians to their patients and parents to their children. The duty arises from the role (being a subject, physician or parent). Obligations are distinct from duties on this view. Obligations result only from voluntary actions; for instance, an obligation might result from the making of a promise. It may be that the distinction between duties and obligations is currently seldom recognized because many of our roles are taken on voluntarily. Although there is much to be said for the narrower view, for the rest of this essay current usage will be followed and it will be assumed that duties and obligations are substantially the same.

One important view about the relationship between rights and duties is that there is a correlation between, at least, some rights and duties. A version of this thesis is expressed in the two following principles Ross (1930: 48).

1 A right of *A* against *B* implies a duty of *B* to *A*.
2 A duty of *B* to *A* implies a right of *A* against *B*.

It should be noted that accepting these principles does not force us to accept the less plausible view that there is a general correlation between rights and duties, such that every right implies a duty and every duty implies a right. As has been pointed out, there may be duties concerning individuals which are not duties to those individuals. Likewise, individuals may have rights which are not rights against anyone in particular. About such cases the above principles are silent. Rather, the proposal is that there is a correlation between, for example, Smith’s having a duty to Jones, and Jones’ having a right against Smith (see Rights).

An evaluation of these principles requires that we examine the following questions. (1) What can have duties? (2) To what can there be duties? (3) What can have rights? (4) Against what can there be rights?

The answers to (1) and (4) are fairly clear. It is reasonable to think that in order for an entity to have a duty, or for there to be a right against it, that entity must be able to recognize that it is subject to a normative requirement, a cognitive ability lacked by beings that are not rational (see Moral agents). Consequently, only persons (rational beings), such as humans, can have duties and we can have rights only against persons. This still leaves open whether non-persons can have rights and whether we can have duties to non-persons. Can we, for example, have duties to nonhuman animals, or, perhaps, the biosphere? Can such things have rights against us? (See Animals and ethics; Environmental ethics.)

One important view is that non-rational beings cannot have rights and we cannot have duties to them. Again, the thought is that a required cognitive capacity is lacking. It is sometimes argued that having any right, let alone a right against someone, requires the ability to claim it. Even according to this view, however, we might have duties concerning non-persons, or with respect to them. We could, for example, have a duty not to cause non-persons unnecessary pain, without its being a duty to anyone. Accepting this view would not require rejection of either of the principles.

An alternative view argues that for an individual to have a right, or for us to have a duty to that individual, does not require that individual to have anything like the cognitive capacity needed to have a duty, or to be the sort of thing against whom others can have rights. The standards are different. Perhaps all that is needed is being sentient, or having interests. If so, then we could accept that some non-persons can have rights against us and that we can have duties to some non-persons. This also is consistent with both of the principles.

One possible reason for rejecting the principle that a duty of *B* to *A* implies a right of *A* against *B* is the thought that we can have duties to ourselves, but cannot have rights against ourselves. The idea that persons can have rights against themselves is difficult to accept. That persons can have duties to themselves, such as that of self-improvement, seems much more plausible. However, if we take it that a duty to a person involves what is owed to that person, then the idea is much less plausible: can we really owe anything to ourselves? Alternatively, it might be that we have duties that concern ourselves, without their actually being duties to ourselves. If this is so, then no reason has been offered for rejecting the two principles (see Moral standing).

2 Kinds of duties

A number of distinctions have been offered between types or kinds of duties. The most important of these distinctions are those between positive and negative duties, *prima facie* and ‘all-in’ duties or ‘all things considered’ duties, perfect and imperfect duties, and subjective and objective duties.

The distinction between positive and negative duties is supposed to rest on a more fundamental difference between acting and refraining from acting. The idea is that positive duties concern what we are required to do, while negative duties concern what we are required to refrain from doing. Examples of positive duties might be the duties to help others and to develop our talents. Examples of negative duties might be the duties not to lie and not to kill.

A concentration on negative duties is often associated with deontological theories (see Deontological ethics). According to such theories, morality consists largely of constraints on action, none of which we can morally violate, but which may be relevant only in a limited number of situations. Deontologists often hold that the distinction between positive and negative duties can be morally significant: the negative duties are more important or more stringent than the positive ones. Indeed, it is sometimes held that there can be a duty not to bring about something, while there is no duty to prevent it from occurring (see Double effect, principle of). For example, there might be a very stringent duty not to kill someone, while there is, at best, a much less stringent duty to save that person’s life (see Life and death).

Teleological theories, especially consequentialist ones, even if they recognize a distinction between positive and negative *actions*, generally do not consider it to underpin a difference between negative and positive *duties* that has any moral significance (see Teleological ethics; Perfectionism; Consequentialism). Such theories are concerned with the promoting of values, and it does not matter whether the values are promoted by acting or refraining from

acting. For example, if the value is respect for life, then it can be promoted equally well either by refraining from killing or by saving lives.

While it seems reasonable to think that there is some difference between positive and negative actions, that this involves a morally significant difference is much more contentious. Views about this divide roughly along the same lines as do those concerning the relative advantages of deontological and teleological theories. No doubt this will remain controversial as long as these approaches to duty compete.

An equally controversial and perhaps equally important distinction can be drawn, as it was by Ross (1930), between *prima facie* duties and 'all-in' duties (see Ross, W.D.). (*Prima facie* duties are sometimes called 'pro tanto duties' or 'duties other things being equal', while all-in duties are sometimes called 'duties proper' or 'duties all things considered'.) A *prima facie* duty is not, contrary to its literal meaning, something that merely *appears* to be a duty. Rather, to say that we have a *prima facie* duty to perform some action expresses the idea that the action involves an important moral consideration, but one that can be (morally) outweighed, or overridden, by other moral considerations. For example, assume that an agent has a choice between keeping a promise and helping someone in need, but cannot do both. Keeping the promise is a *prima facie* duty because it involves the morally important consideration of fidelity; helping the person in need is also a *prima facie* duty as it involves the morally important consideration of beneficence. The agent's all-in duty depends on the nature of all the relevant *prima facie* duties. That is, an all-in duty is what duty requires all things considered, in particular, given all *prima facie* duties.

Part of the motivation for this distinction is the idea that (all-in) duties can always be satisfied: there are never any real conflicts of duties. Using this distinction, we can attempt to explain apparent conflicts of duties in terms of conflicts of *prima facie* duties. When it seems that we have a duty to do two things, and we cannot do both, this distinction allows us to say that they are not really competing duties, but merely competing *prima facie* duties. (In some cases, of course, we would be permitted to do either, but would not have a duty to do either.) However, that all apparent conflicts of duties or moral dilemmas can be explained away in this fashion is not obvious. There seem to be situations where agents have to choose between courses of action that are so horrible that none is acceptable and the agent should feel regret no matter what is done. Whether such cases involve real conflicts of duties depends on two things. First, it depends on whether we take dilemmas to exclude there being a correct answer to the question of what duty requires. It may be that we call something a dilemma simply because the choices are so awful. Second, it depends on whether there is some inconsistency in feeling regret for doing something that was required by duty.

Regardless of the attitude that is taken to possible conflicts of duty, the notion of a *prima facie* duty faces difficulties. Foremost among these is the lack of a persuasive account of how all-in duty results from diverse *prima facie* duties. No satisfactory account has been given of the conditions under which one *prima facie* duty 'overrides' another.

The view that there is a distinction between perfect and imperfect duties is attributed to Kant (§10) (see Kantian ethics). He said that a perfect duty 'allows no exception in the interest of inclination' ([1785] 1903: 421). Although it is not clear exactly what he had in mind, one plausible view is that the distinction is between duties owed to particular individuals and duties that are not owed to anyone but which merely concern them. If we have a duty to someone, then that person has a right against us. (Kant thought that we can have duties to ourselves and may even have thought that we can have rights against ourselves.) If we have a duty that is not to anyone, then that duty does not involve a right against us. This, arguably, gives us more latitude in how to satisfy the duty. A duty not to kill is owed to each person, while a duty of charity is not owed to anyone, and can be satisfied by helping any of a number of persons. Perfect duties are usually associated with negative, stringent duties, while imperfect duties are associated with positive, less stringent duties.

Finally, if a theory allows for a difference between what a person sincerely considers to be required by duty and what is actually required by duty, then it can allow for a distinction between subjective and objective duty. Subjective duty is that which one sincerely takes to be one's duty, while objective duty is that which is actually required by duty. There are other notions closely related to subjective obligation, which are sometimes offered as accounts of subjective duty. One is the notion of that which duty probably requires; another is that which it would be most reasonable to think that duty requires. Each of the latter is distinct from subjective duty, as I have

characterized it. One may be sincere but incorrect about what duty probably requires. Also, one may have a sincere belief about what duty requires, which is not the most reasonable belief one can have.

3 The grounds of duties

Given that to have a duty is to be subject to a binding normative requirement, we might naturally wonder how it is that we could become subject to such a requirement. Who or what binds us? This, perhaps, is the most difficult and most interesting question concerning duties.

One tradition says that only God could be the source of our duties (see Religion and morality §1; Natural law). Given God's existence, it might be thought that God has authority over us and can impose duties on us. God would then have a right to our obedience. Giving us the Ten Commandments can be thought of as one way that God has imposed duties on us. If this is correct, then we have a duty to obey them. A problem with this view is that it has the implication that if God does not exist, then there are no duties. Another problem with it is that it is unclear why we are required to do as God commands. That is, it is unclear why God has authority over us. Indeed, a similar question can be asked about anyone or anything else, whether a person, institution or society, that is supposed to be able to impose duties on us. Why should we think that they have the authority to impose duties on us, or what gives them the right to tell us what to do?

Perhaps instead of having duties imposed on us by other humans, it is the structure of the universe itself that imposes the duties. On this view, duties (or moral laws) would be analogous to physical laws. Just as we are subject to physical laws, such as Newton's law of gravitation, we are subject to binding normative requirements, for example, the duty to help others who are in need. It is not that the universe has authority over us and we must obey its command, rather it is that normativity is part of the very fabric of the universe.

Various problems with this view have been raised. One is that the features of the universe that produce the duty would be very strange, unlike anything else about which we know (see Naturalism in ethics; Moral realism §6). A second is that it is difficult to see how we could learn about these duties. We do not have any perceptual experience of them, and perceptual experience is widely thought to be the way that we gain knowledge about the world (see Moral knowledge). The third objection rests on the action-guiding nature of duty. The idea is that it is a necessary feature of accepting something as being required by duty that we are to some extent motivated to do it. How is it that duties imposed on us by the structure of the world motivate us (see Moral motivation)?

If duties are not imposed by things independent of us, perhaps we impose them on ourselves. Indeed, as Kant argued, it might be that a duty can only be imposed by the person having it (see Kant, I. §9). A non-Kantian expression of a similar idea is this. Only moral agents can have moral duties. Moral agents are responsible for their actions, and being responsible for an action requires that the agent does it (see Responsibility). However, freedom is not sufficient for moral agency. For example, it seems that children are free, but not moral agents. Moral agency also requires that agents act rationally, that they act on reasons that they accept as adequate. This means an action is performed because of the nature of the action. If duties could be imposed on us by others - God, society, or the universe - then we would not be moral agents. We would not be acting on adequate reason because our reason for acting would not be the nature of the action. Instead, we would be acting because the action was imposed as a duty. It would be like obeying a command simply because it was a command and not because what was commanded was appropriate for the situation. So for something to be someone's duty, it must result from that person - only we can impose duties on ourselves. A similar point is this: moral agents have authority over themselves, but are subject to no other authority (see Autonomy, ethical).

Numerous questions can be raised about the idea that we impose duties on ourselves. Why is giving ourselves reasons for actions imposing duties on ourselves? After all, if we have authority over ourselves, can we not change our duties at will? Indeed, what is it to have authority over ourselves? A possible answer to these questions is that the second condition of moral agency, that we act rationally or act from adequate reason, provides additional normative constraints on what we can do as moral agents. In particular, this condition may require a sort of impartiality: a reason for a rational agent acting in a certain way in a particular situation must be a reason for any rational agent to act in the same way in similar situations (see Impartiality; Universalism in ethics §5).

If we were to accept that others cannot impose duties on us and we cannot impose duties on ourselves, would there be any use for the notion of duty? Is the idea of a duty, of morality itself, a fiction to be discarded? Not

necessarily. There could be other explanations of why we think that the notion of duty is relevant to our actions.

One position is that we cannot do without some notion of duty: it is a useful instrument. On such an account, the reasons for accepting a system of morality would not be moral reasons. Instead, the reasons would be ones of self-interest, or a combination of self-interest and interest in others (see Prudence §2). The idea is that people will generally be better off living in a society where there are social institutions and conventions regulating behaviour (see Contractarianism). For example, killing might be proscribed because it is to everyone's mutual advantage not to have to worry about being killed. A problem for this view is that there may be some people in a society who contribute such a small amount to the society that there is no reason to cooperate with them. Indeed, some may contribute so little that there is not even any reason to refrain from killing them.

Another approach suggests that what seem to be claims about duties, for example, that there is a duty to refrain from killing, actually are not claims at all. Since they are not claims, they cannot be true or false. They are merely expressions of attitudes. For example, if someone says, 'We have a duty to help those in need', no claim is being made; the utterance is merely the expression of a positive attitude towards helping those in need. It is as if the person said, 'Hooray for helping others' (see Emotivism). The main challenge for this view is to explain why much of moral conversation seems like reasoned argument. One strategy for dealing with this challenge is to argue that there is a logic of attitudes that mirrors the logic of claims or propositions.

In addition to the difficulties in giving an account of how we come to be subject to duties, the entire project of trying to characterize ethical concerns in terms of duties has come under criticism. One main objection is that the idea of duty is so closely connected to the idea of God as a lawmaker or imposer of duties that the concept has no place in secular philosophy (see Anscombe 1958). Another is that concentrating on duty blinds us to the rich diversity of considerations that are relevant to ethics. Instead of asking what our duties are, we should be asking how we should live (Williams 1985) (see Virtue ethics).

See also: Conscience; Duty and virtue, Indian conceptions of

ROBERT L. FRAZIER

References and further reading

Some of these items involve intricate argument but few are technical.

Anscombe, G.E.M. (1958) 'Modern moral philosophy', *Philosophy* 33: 1-19.(A powerful attack on the use of the concept of duty in secular philosophy.)

Bennett, J. (1995) *The Act Itself*, Oxford: Clarendon Press.(A discussion of the nature of action, including whether there is a morally significant difference between positive and negative actions.)

Frazier, R.L. (1995) 'Moral relevance and *ceteris paribus* principles', *Ratio* 8: 115-27.(A modern account of *prima facie* duties.)

Gauthier, D. (1986) *Morals By Agreement*, Oxford: Oxford University Press.(A modern defence of contractarianism.)

Gowans, C. (ed.) (1987) *Moral Dilemmas*, Oxford: Oxford University Press.(A collection of articles concerning the possibility of moral dilemmas.)

Helm, P. (ed.) (1981) *Divine Commands and Morality*, Oxford: Oxford University Press.(A collection containing articles on the proposal that God imposes duties on us.)

Hobbes, T. (1651) *Leviathian*, ed. J.P. Plamenatz, London: Collins, 1962.(The classic statement of contractarianism.)

Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978.(The classic account of the view that what seem to be moral claims are actually expressions of moral sentiments.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, in *Kants gesammelte Schriften*, ed. Königlichen Preußischen Akademie der Wissenschaften, Berlin: Reimer, vol. 4, 1903; trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(References made to this work in the entry give the page number from the 1903 Berlin Akademie volume; these page numbers are included in the Paton translation. This classic work contains the argument that we impose duties on ourselves.)

Korsgaard, C.M. (1996) *The Sources of Normativity*, Cambridge: Cambridge University Press.(A Neo-Kantian account of the grounds of duty, with a good discussion of alternative proposals.)

Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin. (Contains a good introductory discussion of the problems with thinking that duty can be imposed on us by the nature of the universe.)

Sayre-McCord, G. (ed.) (1988) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press.(A collection of articles concerning whether the nature of the universe can provide the grounds of duty. It contains a very good introductory chapter.)

Sinnott-Armstrong, W. and Timmons, M. (eds) (1996) *Moral Knowledge? New Readings in Moral Epistemology*, Oxford: Oxford University Press.(A collection of articles on the possibility of moral knowledge, containing a very good annotated bibliography.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(A very influential defence of a position which is an interesting combination of a contractarian and a Neo-Kantian approach to ethics.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Oxford University Press.(The source of a notion of a *prima facie* duty.)

Thomson, J.J. (1990) *The Realm of Rights*, Cambridge, MA: Harvard University Press.(Includes a discussion of the relationship between duties and rights, and a discussion of who (or what) can have rights and duties.)

White, A.R. (1984) *Rights*, Oxford: Clarendon Press.(Includes a discussion of the relationship between duties and rights, and a discussion of who (or what) can have rights and duties.)

Williams, B. (1985) *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press and London: Fontana.(A collection of articles supporting the view that duty-based moral systems are not sensitive to a wide range of important ethical considerations.)

Zimmerman, M.J. (1996) *The Concept of Moral Obligation*, Cambridge: Cambridge University Press.(A very general and comprehensive account of obligation (or duty), which gives close consideration to many of the topics discussed in this entry.)

Duty and virtue, Indian conceptions of

*Two principal strains of ethical thought are evident in Indian religious and philosophical literature: one, central to Hinduism, emphasizes adherence to the established norms of ancient Indian culture, which are stated in the literature known as the Dharmaśāstras; another, found in texts of Buddhism, Jainism and Hinduism alike, stresses the renunciation of one's familial and social obligations for the sake of attaining enlightenment or liberation from the cycle of rebirth. The Dharmaśāstras define in elaborate detail a way of life based on a division of society into four 'orders' (*varṇas*) - priests, warriors, tradesmen and servants or labourers - and, for the three highest orders, four 'stages of life' (*āśramas*). Renunciation is valid only in the final two stages of life, after one has fulfilled one's responsibilities as a student of scripture and as a householder. The various traditions that stress liberation, on the other hand, advocate total, immediate commitment to the goal of liberation, for which the householder life presents insuperable distractions. Here, the duties of the householder are replaced by the practice of yoga and asceticism. Nevertheless, specific ethical observances are also recommended as prerequisites for the achievement of higher knowledge through yoga, in particular, nonviolence, truthfulness, not stealing, celibacy and poverty. The liberation traditions criticized the system of the Dharmaśāstras for being overly concerned with ritual and external forms of purity and condoning - indeed, prescribing - the killing of living beings in Vedic sacrifices; but it was only in the Dharmaśāstras that the notion of action solely for duty's sake was appreciated. The Hindu scripture the Bhagavad Gītā (Song of God) represents an effort to synthesize the two ideals of renunciation and the fulfilment of obligation. It teaches that one should integrate yoga and action in the world. Only when acting out of the state of inner peace and detachment that is the culmination of the practice of yoga can one execute one's duty without regard for the consequences of one's actions. On the other hand, without the cultivation of inner yoga, the external forms of renunciation - celibacy, mendicancy, asceticism - are without significance. It is inner yoga that is the essence of renunciation, yet yoga is quite compatible with carrying out one's obligations in the world.*

1 Introduction

Ethics was not a prominent topic of sectarian controversy in classical Indian philosophy. While Hindu, Buddhist and Jaina philosophers debated such things as the existence of universals, the reality of the external world and the existence of a self, they did not often argue directly about the nature of the good or right action. Each religious-philosophical tradition had its own body of sacred texts to which it appealed in defining *dharma* - that is, duty or righteousness, the way of conducting one's life so as to achieve happiness on earth and salvation after death. Given allegiance to a particular scripture, the question of how one should live was largely moot; debate centred, rather, on the epistemological question of how one can be confident that a particular scripture is authoritative. At the same time, two distinct ethical ideals are found espoused, explicitly or implicitly, throughout Indian religious and philosophical literature - the observance of duty, and renunciation. The ideal of renunciation cuts across all three major traditions. Especially in the Hindu tradition, where both ideals are represented, their conflict is an important theme and an attempt was made, in the *Bhagavad Gītā (Song of God)*, to reconcile them.

2 Duty

The emphasis on duty is represented primarily, if not exclusively, by texts of the Hindu tradition. In the *Ṛg Veda*, the oldest stratum of text (*c.*2000 BC), virtue is conformity to cosmic law (*ṛta*). That involves not only the performance of various obligatory rites on behalf of the gods, but also the observance of fidelity and rectitude in human relationships. In the Brāhmaṇas, later texts devoted to analysing and interpreting rituals, virtue is narrowly conceived as ritual excellence. The sacrifice is understood to yield its result through its own magic, by directly effecting what it symbolizes, not by honouring or ingratiating the gods to whom it is addressed. Transgression is thus merely a ritual mistake, a blunder or negligence - not an offence against a personal being to whom one owes an obligation - either in the context of the sacrifice itself or in human affairs seen as an extension thereof; and it produces a defilement, materially conceived as a taint or miasma, that can only be ritually removed. The Brāhmaṇas declare, for example, that the impurity of the most heinous deeds, even the killing of a Brahman (priest), can be wiped away by performing a horse sacrifice.

This ritualistic conception of virtue was significantly modified in the Dharmaśāstras, 'treatises on *dharma*'

(composed from about 600 BC onward), the most famous of which is the *Mānavadharmaśāstra (Laws of Manu)*. The correct performance of rituals is still important, but the highest *dharma*, according to this literature, is 'good conduct' in the broadest sense: 'Neither austerities nor [the study of] the Veda, nor [the performance of] the Agnihotra [rite], nor lavish liberality [to priests] can ever save him whose conduct is vile and who has strayed from the path [of *dharma*]' (*Vasiṣṭha dharmasūtra* 6.3). Nevertheless, *dharma* still fundamentally involves *doing* (and avoiding) certain things; it is implicitly defined as the adherence to certain norms, specifically, the norms of the ancient Indian society that referred to itself as 'the Aryas', which transmitted the Vedic literature. The Dharmaśāstras lay down the occupations of the various classes or 'orders' (*varṇas*, commonly referred to as 'castes') of Aryan society. Brahmans serve as priests and teachers; *kṣatriyas* as princes and warriors, *vaiśyas* as tradesmen and *śūdras* as labourers and servants. For the three highest, or 'twice-born', orders they prescribe four 'stages of life' (*āśramas*) - studentship, marriage, forest residence and renunciation - and specify that they are to be observed only in succession. Thus, although renunciation is part of the scheme, it is still to be undertaken only at a certain time and in a certain way, and so falls under the rubric of duty. (A well-known verse of the *Mānavadharmaśāstra* states, 'A twice-born man who seeks final liberation without having studied the Vedas, without having begotten sons, and without having offered sacrifices, sinks downwards' (6.37).) For the Brahman order in particular, the Dharmaśāstras spell out numerous restrictions intended to preserve its unique 'purity' - the avoidance of polluting agents, especially contact with people of lower orders, the avoidance of certain foods, and so forth. They also specify rules bearing on marriage and inheritance and other matters of civil law. They prescribe various sacraments to be undergone at certain junctures of life, the most important being the investiture of the sacred thread at the beginning of the period of studentship. They go into a plethora of matters of etiquette, even prescribing how to have sex and answer the calls of nature.

Western scholars have criticized the Brahmanical system of the Dharmaśāstras (so called because it was composed by Brahman literati in ancient times and describes a society in which Brahmans have ultimate authority) for considering virtue to be mere conformity to established practices without attempting to arrive at more fundamental principles by which such practices could be justified or criticized. But the Dharmaśāstras speak of, besides the *dharma* of the specific *varṇas*, 'the *dharma* common to all orders' (*sarvāśramadharma*). That includes such things as forgiveness, self-control, not stealing or lying, nonviolence, restraint of the senses, compassion, patience, freedom from anger, envy and avarice, and so on. Thus, while there is indeed a lack of reflection on the nature of virtue in general in the Dharmaśāstras, or for that matter in Indian thought as a whole, a step in that direction can be seen in the attempt to identify these more general duties - which, however, are considered to apply only to Aryas; other 'barbarian' societies are condemned simply for being non-Aryan. At the same time, the influential Mīmāṃsā school of Hindu philosophy explicitly asserted that, *dharma* being supersensible and thus unknowable by unaided reason, we can learn about it only from revelation. Moreover, one's state of mind is clearly recognized as essential to right conduct in the Dharmaśāstras; mere external conformity is not enough. The psychological sources of transgression - desire, greed, anger and delusion - are identified, and intention is said to add weight to the deed. Although mechanical expiation is still provided for - misconduct is still conceived to produce a defilement that can be washed away, and numerous purification rituals are described - remorse is also part of expiation.

The various Hindu schools of philosophy engaged in an extensive discussion of the nature of the scriptural injunction (*vidhi*) that commands one to carry out or avoid a certain act. Nyāya philosophers held that the injunction makes one aware of a means (the dharmic act) by which a certain desirable purpose can be achieved; that is, it motivates the hearer to act by implicitly appealing to the hearer's self-interest, whether it be spiritual or hedonistic. In general, a dharmic, pious life was thought to be conducive to happiness and prosperity on earth and the attainment of heaven after death; but, as we shall see, it was also considered instrumental in attaining ultimate salvation in the form of liberation from the cycle of rebirth. Bhāṭṭa Mīmāṃsā philosophers, on the other hand, believed that the *vidhi*, specifically, the ending of the optative verb of the injunctive sentence, actually *impels* the agent toward acting by virtue of its own peculiar efficacy (referred to as *śābdabhāvanā*). At the same time, an awareness of some purpose to be achieved through the act is also necessary. Prābhākara Mīmāṃsā philosophers, finally, held that a *vidhi* motivates one to act neither by *causing* one to act through the potency of its language, nor by *persuading* one to do so by suggesting the achievement of a desired end, but simply by making one aware of an obligation (*niyoga*) that one feels bound to carry out; it is followed for its own sake, without regard for the result of the act it prescribes. Understood in this way, the scriptural injunction resembles the Kantian categorical

imperative. This theory also implies that the will of the agent who carries out an injunction is free (see Mīmāṃsā).

3 Renunciation

The ideal of the renunciation (*sannyāsa*) of all familial and social obligations for the sake of enlightenment or liberation (*mokṣa*, *nirvāṇa*) from the cycle of rebirth was espoused in various classical Indian traditions and represents a rejection of the Brahmanical ideal. Already in the late Vedic texts called the Upaniṣads (800-300 BC) the sacrificial cult is called into question. Residence in heaven, to be achieved through ritual acts, is declared undesirable because it is impermanent: after the merit of one's good deeds (karma) is exhausted the soul must return to the physical world, where one again experiences suffering. The only possibility of fulfilment is to transcend embodied existence altogether; this is conceived in the Upaniṣads and the Vedānta school of philosophy, which based itself on the Upaniṣads, as a merging of the individual soul with Brahman, the ultimate reality of the universe. For the Hindu Nyāya-Vaiśeṣika school, it was the dissociation of the self from the mind and body, and hence from all conditions of consciousness. The Buddha refused to characterize *nirvāṇa* at all, suggesting that it is simply beyond human conception (see Vedānta §§1, 3; Nyāya-Vaiśeṣika §7; Nirvāṇa). Almost all liberation traditions emphasized techniques of yoga and asceticism along with renunciation - withdrawal from society and life as a celibate monk, hermit or wandering mendicant - as means of reducing the factors that determine rebirth, variously identified as ignorance, karma or desire. Even the commission of good acts was believed to bind the soul further in *saṃsāra* (the cycle of rebirth); for every action, whether bad or good, must eventually yield a retribution in the form of pleasure or pain, both of which require a body. (An exception is the case of an enlightened being, whose actions have no consequences because such a being has overcome the state of individuality.) Jainism, especially, recommended asceticism and self-mortification as means of directly reducing the load of accumulated karma (see Jaina philosophy §3).

However, the liberation traditions also acknowledged that the observance of certain moral precepts is essential for spiritual advancement. The texts of Buddhism, Jainism and Hinduistic yoga state the same pentad of commandments: nonviolence, truthfulness, not stealing, celibacy and nonpossession. It was generally believed that such observances reduce the tendencies towards passion that serve as the conditions of the fruition of karma. Overall, this list of virtues represents a critique of the ethical system of the Dharmaśāstras; it recommends universal principles over traditional, caste-specific norms and stresses intention over overt performance. (For each of the five virtues there corresponds a mental attitude to be cultivated; one should, for example, not only refrain from taking what does not belong to oneself, but also not have covetous thoughts.) Buddhist and Jaina texts in particular ridicule the *dharma* of the Brahmans because it not only reflects an obsession with external forms of purity but also actually involves injury to living beings. It is puzzling how a man could think himself good just because he bathes every time he comes into contact with a *śūdra*; it is even more incomprehensible that he could think so because he slaughters an animal in a sacrifice! One can see, however, that this critique of Brahmanical ethics is not entirely fair. Though they emphasize specific caste obligations, the Dharmaśāstras also recommend the cultivation, by all members of society, of most of the same virtues stressed in the liberation traditions (the notable exception being nonpossession, which precludes existence as a householder; celibacy in the Brahmanical scheme was reinterpreted as sexual restraint in marriage). They even advocate nonviolence, though typically they recommend it *outside* the context of the sacrifice; and it is the killing of a Brahman that is especially condemned.

It should also be kept in mind that the virtues recognized by the liberation traditions were understood chiefly as means to an end - enlightenment or liberation. It was only within the tradition of the Dharmaśāstras, specifically in the Mīmāṃsādiscussions of *vidhi* and, as we shall see, in the *Bhagavad Gītā*, that an appreciation of the ideal of action simply for duty's sake was attained. However, the emphasis Buddhism placed on nonviolence (*ahiṃsā*) and compassion (*karuṇā*) indicates that at least those virtues had more than instrumental value there. The Mahāyāna writings, in particular, stress that compassion is not to be practised out of any interest - certainly not for the sake of worldly delights, but also not even for the sake of liberation. The saints of Mahāyāna Buddhism, the Bodhisattvas, postpone disappearance into *nirvāṇa* until they have helped to enlighten all other beings: 'By his own person, like a baited hook, without partaking of any enjoyment himself, he attracts others and saves them. Thus it is said… "Charity is the enlightenment of the Bodhisattva"' (*Śikṣāsamuccaya*, BST edn, 22). The Buddhists came close to deriving nonviolence explicitly from the Golden Rule. While nonviolence is sometimes treated purely instrumentally in Indian thought - it is said in certain Jaina texts that violence must be avoided because it causes injury to *oneself*, in so far as the passions that engender it prevent the soul from attaining liberation - there is none

the less a marked tendency to view it as a duty grounded on a law of reason or insight into one's kinship with all creatures.

4 Renunciation in action

The Hindu scripture the *Bhagavad Gītā*, perhaps the single most important Indian text that relates to ethics, can be seen as an attempt to reconcile the two ideals of duty and renunciation. Contained in the epic poem the *Mahābhārata* (*c*.200 BC-AD 200, traditionally attributed to Vyāsa), the *Gītā* relates a dialogue between Kṛṣṇa (Krishna), an incarnation of the Supreme Being Viṣṇu, and the warrior Arjuna. The issue of the dialogue is whether Arjuna should participate in the great battle that is about to take place, for his own cousins lead the opposing side. Arjuna knows that his cause is just - his cousins are violent and despotic; they have deprived him and his brothers of their due portion of the kingdom - but he nevertheless feels that it would be a violation of *dharma* to be an agent of the destruction of his own friends and relatives, no matter how wrong they are. Filial duty has come into conflict with Arjuna's duty as a prince (*kṣatriya*) to preserve order in society and defend it against its enemies. In response to Arjuna's quandary, Kṛṣṇa does not enter into a discussion of the nature of virtue or universal principles of right action (as perhaps Socrates would have done) but tells Arjuna to cultivate yoga, which will enable him to rise above his situation. Yoga will provide Arjuna with clarity and equanimity, so that he is no longer confused about his duty - which Kṛṣṇa suggests is obvious, anyway: to defend society against evil - or overcome with remorse about the pain he must inflict in doing what he must do. It will bring about a state of calm and centredness from which he can carry out right action undistracted by passion; 'Yoga is skill in action', says Kṛṣṇa (2.50). Moreover, it will transform the personality, so that Arjuna will no longer act as an individual agent but yield to the forces of nature - ultimately, to God - acting through his body: 'He who thinks this self is a killer and he who thinks it is killed, both fail to understand; it does not kill, nor is it killed' (2.19). The yogin is not 'bound' by what he does; his actions do not necessitate further rebirth.

Thus the *Gītā* works out a synthesis of the ideals of duty and renunciation, proposing a new ideal of renunciation in action. Especially in the third, fourth and fifth chapters the theme of renunciation is taken up. Merely external forms of renunciation are criticized: one who physically refrains from acting but still yearns after sensual pleasures in his mind is a hypocrite (3.6). Thus, the tables are turned on the liberation traditions, which had criticized Brahmanical ethics for being formalistic. True renunciation is the state of inner yoga, a condition of complete calm, contentment and detachment. This is compatible with acting so as to carry out one's responsibilities in the world, even with observing the specific prescriptions of the Dharmaśāstras - indeed, the *Gītā* upholds the *varṇa*-*āśrama* system. In reality, all creatures are driven to act in so far as they are part of nature, which is subject to constant change - the fluctuations of 'the three *guṇas*' (basic ingredients of material nature). Kṛṣṇa, the Supreme Being, refers to his own case: 'These worlds would collapse if I did not perform action' (3.24). It is stressed, however, that action should be undertaken only for its own sake, out of a sense of duty: 'You have a right to action alone, not the fruits [of action]' (2.47). Thus, the state of yoga is essential for right action, because only in that state is one not driven by inclination. It is the detachment of yoga that makes it possible for one to act only for duty's sake. The *Gītā* also describes this state of mind in terms of considering one's actions as a sacrifice and surrendering one's actions to God.

The *Bhagavad Gītā* contains many religious teachings that do not relate directly to ethics but reflect the sectarian (Vaiṣṇava) origins of the poem. It stresses above all that Viṣṇu is the ultimate reality and that salvation can be attained only by worshipping Viṣṇu in his true form. Yet in the end, this metaphysical knowledge provides the solution to Arjuna's ethical dilemma. At a certain point Kṛṣṇa, who until then has appeared just as a fellow warrior, reveals himself to Arjuna as God. Arjuna sees the infinite within his body - innumerable gods, creatures and universes; the physical and psychological constituents of all beings; the past and the future, including even the outcome of the battle he is about to fight. Overwhelmed by Kṛṣṇa's infinite splendour, he now sees his own concerns as trivial. Realizing that his own actions are part of a vast panorama of life directed by God, he surrenders himself completely to Kṛṣṇa and prepares to fight.

See also: Fa; Fatalism, Indian; Karma and rebirth, Indian conceptions of

JOHN A. TABER

References and further reading

Bhagavad Gītā (Song of God) (*c*.200 BC), trans. F. Edgerton *The Bhagavad Gītā*, Harvard Oriental Series 38-9, Cambridge, MA: Harvard University Press, 1944. (A readable and accurate translation of the most important text of Hinduism.)

Bühler, G. (ed.) (1879-92) *The Sacred Laws of the Aryas: As Taught in the Schools of Āpastamba, Gautama, Vasiṣtham, and Baudhāyana*, Sacred Books of the East 2, 14, Oxford: Clarendon Press.(A collection of important Dharmaśāstras.)

Dhammapada (6th-5th century BC), trans. F.M. Müller, *The Dhammapada*, Sacred Books of the East 12, New York: Scribner, 1901.(A foundational treatise on Buddhist practice.)

Hopkins, E.W. (1924) *Ethics of India*, New Haven, CT: Yale University Press.(Still the best survey of Indian ethics.)

Kane, P.V. (1968-77) *History of the Dharmaśāstras*, Poona: Bhandarkar Oriental Research Institute, 5 vols.(A comprehensive guide to Hindu legal and ethical literature.)

Maitra, S.K. (1978) *The Ethics of the Hindus*, Delhi: Asian Publication Services. (Contains in-depth discussions of the views of the Hindu philosophical schools on various topics.)

Mānavadharmaśāstra (*c*.600 BC), trans. W. Doniger and B.K. Smith, *Laws of Manu*, New York: Viking Penguin, 1992.(The most influential of the Dharmaśāstras.)

Dworkin, Ronald (1931-)

Ronald Dworkin's early, highly controversial, thesis that there are right answers in hard cases in law, coupled with his attack on the idea that law is simply a system of rules, gained him a prominent and distinct place in the anti-positivist strand of legal theory. He has developed and enriched his earlier insights by tying his notion of law-as-interpretation to the ideals of community and equality. Dworkin is an influential representative of liberal thought, who combines clear and analytical thinking with political involvement expressed in decisive and timely interventions in many of the important political debates of our time.

Ronald Dworkin, born in Massachusetts, read philosophy and law at Harvard and Oxford; after a brief period in legal practice he became a professor of law at Yale in 1962. In 1969 he succeeded H.L.A. Hart as Professor of Jurisprudence at Oxford University and since 1975 has also been Professor at New York University. His ideas have been profoundly influential in developing British and US jurisprudential thought.

Since very early in his career Dworkin has queried the positivist (and particularly the Hartian) concept of law as a system of rules. Dworkin criticizes positivism for its limited and impoverishing account of law that necessarily entails judicial law-making and retrospectivity in hard cases. Dworkin argues that every judicial decision requires discretion understood as the exercise of judgment, as interpretation, and yet none requires discretion of the unrestricted type envisaged by positivists. The right legal answer cannot be 'read off' a rule but requires a justification (that subsumes the rule under it) in terms of discussing principles embodied in the law (see Legal positivism §5; Legal reasoning and interpretation §§2-4).

To his first distinction, 'rules/principles' - that underlies his attack on positivism - Dworkin adds a second, that of 'principles/policies', that underlies his attack on utilitarianism. Principles are propositions that describe rights, and to 'take rights seriously' (1978) means to always allow them to 'trump' policies. In every case a requirement of justice or fairness embodied in principle should override any competing consideration of policy associated with the pursuit of collective goals.

Both distinctions, both attacks, are present in *Law's Empire* (1986), which integrates the earlier insights in a theory of law-as-interpretation fundamentally enriched by the notion that the law is a practice and occurs in a community of interpreters. Legal interpretation requires us to appreciate that law is a practice that involves us as participants in arguing its meaning. For the interpretive attitude to take hold people must try to view the practice 'in its best light' and reconceive its requirements in the light of what would most fully realize its implied purpose. This is a requirement performed on objective ground: not by imposing upon the practice outside moral or personal purposes, but by retrieving purpose from within the practice, as it is intelligible to the people participating in the common form of life, the community.

Dworkin's theory of law-as-interpretation focuses precisely on the junction of law with the notion of community. First, as horizon and inventory, community circumscribes the ambit of possibilities of what a practice may require, locating each attempt at the best interpretation of the requirements of the practice within the context we share of its possible meanings. But debating the best among possible understandings involves the participants in an argument over the purpose or point of the practice; interpretation is always justification for Dworkin, the imputation of justificatory principle, but a principle that is already embodied in the past record of the practice, retrieved not posited. It is a kind of reflexive equilibrium between what fits the community's legal record and what morally justifies it - a balancing undertaken against the background of the whole body of the law - that yields for each case in law the right answer, which Dworkin calls 'integrity'. With neither unreflective adherence nor contempt for past legal decisions - a contempt of which Dworkin accuses his prime adversaries, the critical legal scholars - Dworkin's ideal judge secures that the community act in a consistent manner in applying its conception of justice (see Critical Legal Studies). Through the notion of integrity Dworkin can construct a concept that is true to demands of interpretation and community. Interpreting a set of social practices under the demands of integrity sustains the unity of community, elevates it from a bare to a true community and allows the fraternal attitude within to flourish.

Dworkin's work has attracted a great deal of criticism, of which we may identify four axes. The first contests his

account of interpretation. It has been argued here that Dworkin falsely assumes that interpretation ought to be a ‘best’ reconstruction when the point of interpretation is to show the practice as it actually is; that making the law more attractive than it is performs a mystifying, ideological function; and that it allows the meaning of laws to become relativized since ‘best justifications’ are - to some extent at least - subjective. The second questions the ‘dialectical’ relationship between community and the law. Why assume with Dworkin that law is constitutive of the texture of our communities, and provides the terms in which we argue our normative commitments? A third line contests the fundamental distinctions that Dworkin employs. The rules/principles distinction, it has been argued, cannot serve to refute positivism because nothing in positivism precludes principles understood in a justificatory role and as rationalizations behind rules. The privileging of principle over policy runs into the frequent difficulty of keeping the two types of political aim apart in practice. Finally it has been argued that fit cannot be meaningfully divorced from justification - a ‘fit’ is always already justified as such - and as an integral part of justification it could not at the same time constrain it. The fourth line questions Dworkin’s theory’s liberal underpinning, one that is best captured in his advocacy of law as an institution that elevates conflicts from the battleground of politics to the forum of principle (1985) in such a way that with its insistence on the maintainance of unity and of ‘fit’ within the system precludes the possibility of radical critique and change.

Dworkin’s suggestion about what it means to treat people as equals informs his intervention in the debate over reverse discrimination. His theory of distributive justice is explored in a series of articles on equality (1981a, 1981b, 1991) where he argues for equality of resources rather than equality of welfare. Dworkin is known for his interventions in many heated controversies of our time over civil disobedience, pornography, freedom of speech (1985) and abortion and euthanasia (1993). The latter is an attempt to shift the debate from the seemingly intractable disputes in the battlefield of politics towards the common ground that we share: a commitment to the sanctity of life. Thus recasting the debate unifies our political communities around a common commitment while allowing principled disagreement on the morality of abortion and euthanasia that in turn yields greater tolerance for individual choice and renews Dworkin’s image as one of the most important liberal theorists of our time.

See also: Justice, equity and law §6; Law and morality §4; Legal idealism

EMILIOS A. CHRISTODOULIDIS

List of works

Dworkin, R. (1978) *Taking Rights Seriously*, London: Duckworth.(A collection of early articles, including his attack on positivism.)

Dworkin, R. (1981a) ‘What Is Equality? Equality of Welfare’ (part I), *Philosophy and Public Affairs* 10: 185-246. (In both this article and the following article (1981b) Dworkin argues for an ‘endowment-insensitive’ equality of resources rather than for equality of welfare. There are two further parts to his theory of equality, also in 1991.)

Dworkin, R. (1981b) ‘What is Equality? Equality of Resources’ (part II), *Philosophy and Public Affairs* 10: 283-345.(See Dworkin 1981a.)

Dworkin, R. (1985) *A Matter of Principle*, Cambridge, MA: Harvard University Press.(A collection of articles on, among others, ‘the political basis of law’, the nature of interpretation, the economic view of law, reverse discrimination and censorship.)

Dworkin, R. (1986) *Law’s Empire*, London: Fontana.(Dworkin’s major work in legal theory on the nature of law as an interpretive practice and his prescription as to how legal interpretation should be undertaken.)

Dworkin, R. (1991) ‘Foundations of Liberal Equality’, *Tanner Lectures of Human Values*, Salt Lake City, UT: University of Utah Press, 3-119.(Restatement of the theory of equality.)

Dworkin, R. (1993) *Life’s Dominion*, London: HarperCollins.(Dworkin’s thesis on abortion, euthanasia and the sanctity of life.)

Dworkin, R. (1996) *Freedom’s Law*, Harvard, MA: Belknap.(A collection of interventions on political and constitutional issues.)

References and further reading

Cohen, M. (ed) (1983) *Ronald Dworkin and Contemporary Jurisprudence*, London: Duckworth.(A collection of important critiques; see especially Fish, Postema. Includes a ‘Reply’ by Dworkin.)

Guest, S. (1992) *Ronald Dworkin*, Edinburgh: Edinburgh University Press. (Sympathetic and comprehensive account of the life and the works; includes full bibliography.)

Hunt, A. (ed) (1992) *Reading Dworkin Critically*, New York: Berg.(An uneven collection, containing some interesting critiques; see especially Hunt.)
Kerruish, V. (1991) *Jurisprudence as Ideology*, London: Routledge.(A feminist critique of legal liberalism that contains a critique of Dworkin's work.)

Dynamic logics

Dynamic logics have been designed by Pratt as formal systems for reasoning about computer programs. The main ingredients discussed are programs, operations on programs, states and properties of states. In particular one can formalize that every execution of a program p *starting in state* s *terminates in a state with a given property. Thus correctness statements for programs can be dealt with. According to Segerberg (1980) programs might be viewed more generally as actions of some agent so that certain aspects of human action theory can also be formulated and studied in these systems.*

1 Motivation

In Pratt's (1976) setting the meaning of a program *p* with respect to an abstract machine $\mathcal{M}$ is taken as the transition relation $\mathcal{M}_p$ between states of $\mathcal{M}$, so that $\mathcal{M}_p$ holds for states *s* and s' (symbolized by $s\mathcal{M}_p s'$) if and only if there is a machine run according to *p* starting in state *s* and terminating in state s'. The execution of programs can be non-deterministic - another run of $\mathcal{M}$ according to *p* might well lead from the same initial state to quite a different final state. Deterministic programs are therefore special programs where the transition relation is actually a (partial) function. From already given programs one can obtain other programs by applying operations like 'first do p_1 and then p_2' (*composition*), 'do p_1 or p_2' (*nondeterministic choice*) or 'apply *p* a finite number of times' (*iteration* or *looping*). Given a certain stock of basic propositions, i.e. properties of states, one can build new propositions by taking boolean combinations of already given propositions or asking modal questions like 'does every terminating run of a program *p* starting in state *s* necessarily end in a state with property *Q*?'. Thus to every program *p* naturally corresponds a modal operator $[p]$ … to be read: 'after (each terminating execution of) *p* it is the case that … ' The proposition '*not*: $[p]$ *false*' expresses for instance at *s* that at least one run of *p* starting in *s* will terminate. Every proposition *Q* induces a test programme 'if *Q* proceed, otherwise skip' (*test*). The corresponding relation holds for *s* and s' if and only if $s = s'$ and *s* satisfies *Q*. Formalizing this situation leads immediately to propositional dynamic logic.

2 Propositional dynamic logic

The above remarks suggest the following formal language where formulas and program terms (for short: programs) are simultaneously built up from propositional and program variables with the aid of the usual propositional connectives, say $\neg$ (not) and $\rightarrow$ (implies), the *p*-necessity operator $[p]$ for programs *p*, and the program forming operations $;$ (composition), $\cup$ (choice), $*$ (iteration) and $?$ (test). The formation rules for formulas and programs are as follows:

• propositional variables are formulas and if φ and ψ are formulas and *p* is a program variable then $\neg\varphi$, $(\varphi \rightarrow \psi)$ and $[p]\varphi$ are formulas
• program variables are programs and if *p* and *q* are programs and φ is a formula then $(p;q)$, $(p \cup q)$, $p*$ and $\varphi?$ are programs.

Other propositional connectives and the *p*-possibility operator $\langle p\rangle$ are introduced by definition. For instance $\langle p\rangle$ stands for $\neg[p]\neg$. A Kripke-semantics is now easily established. A model $\mathcal{M} = (S, I)$ for this language consists of a nonempty set *S* (the set of states or 'possible worlds' of $\mathcal{M}$) and a function *I* which assigns to every propositional variable a property in *S*, i.e. a subset of *S*, and to every program variable a two-place relation in *S*, an accessibility relation. Then the truth of formulas φ in $\mathcal{M}$ at state *s* (symbolically: $\mathcal{M} \models_s \varphi$) and the interpretation of programs *p* (symbolically: $\mathcal{M}_p$) can be simultaneously defined. The interesting cases are:

- $\mathcal{M} \models_s [p]\varphi$ iff for every s', $s\mathcal{M}_p s'$ implies $\mathcal{M} \models'_s \varphi$
- $\mathcal{M}_{(p;q)} = \mathcal{M}_p \circ \mathcal{M}_q$, $\mathcal{M}_{(p\cup q)} = \mathcal{M}_p \cup \mathcal{M}_q$, $\mathcal{M}_{p*} = (\mathcal{M}_p)*$
- $\mathcal{M}_{\varphi?} = \{(s,s) \mid \mathcal{M} \models_s \varphi\}$

where $R_1 \circ R_2$ denotes the product of R_1 and R_2, i.e. $sR_1 \circ R_2 s'$ holds iff there is an s_1 such that $sR_1 s_1$ and $s_1 R_2 s'$, $R_1 \cup R_2$ denotes the set theoretical union of R_1 and R_2, and $R*$ denotes the reflexive and transitive closure of *R*, i.e. $sR * s'$ holds iff either $s = s'$ or for some natural number *n* there are $s_1, \ldots, s_n$ such that sRs_1 and $\ldots$ and $s_n Rs'$ hold. A lot of constructs known from ordinary programming languages can now be formalized,

for instance 'if φ then *p* else *q*' is $((\varphi?;p)\cup(\neg\varphi?;q))$ or 'while φ do *p*' is $((\varphi?;p)*;\neg\varphi?)$. The set of valid formulas, i.e. formulas which are true in every model at all states, can be generated (in Hilbert-style) from certain axioms by *modus ponens* and the necessititation rule. The set of all tautologies together with the formulas given by the schemas below could be chosen (after Segerberg 1980) as axioms:

- $[p](\varphi\to\psi)\to([p]\varphi\to[p]\psi)$
- $[(p;q)]\varphi\leftrightarrow[p][q]\varphi$
- $[(p\cup q)]\varphi\leftrightarrow([p]\varphi\wedge([q]\varphi)$
- $[\varphi?]\psi\leftrightarrow(\varphi\to\psi)$
- $[p*]\varphi\to(\varphi\wedge[p][p*]\varphi)$
- $[p*](\varphi\to[p]\varphi)\to(\varphi\to[p*]\varphi)$

Correctness of the system is easily verified and completeness has been proved by Gabby, Parikh and Segerberg (see Goldblatt 1987). There are also correct and complete Gentzen-type systems. The question of general validity of a formula turns out to be decidable (Fischer and Ladner 1979).

3 Dynamic predicate logic

Viewing states as the content of certain storage units $x_0,x_1,x_2,...$ they can be identified with functions assigning to each x_n its contents $s(x_n)$. Thus $s(x_n)$ denotes the object (for instance a string over an alphabet) that occupies in state *s* the *n*-th storage unit. The set of objects that are allowed to appear in some storage uni-constitutes a one- (or many-) sorted domain in which certain basic first- (or higher-) order properties and functions are fixed. Using these initial data a first- (or higher-) order possibly many-sorted modal language, known as dynamic predicate logic, can be used to describe transition phenomena initiated by programs in state sets. Simple assignments of the form $x{:=}t$ (to be read: 'set *x* equal to *t*') are often taken as the only basic programs from which all others are derived by program forming operations. A run of such a simple assignment $x{:=}t$ starting in state *s* has the effect that only the content of the storage unit *x* is 'updated' by the value of *t* and every other unit is left untouched. The resulting state s' will be denoted by $s(x/t)$. Simple assignment programs are obviously deterministic and universally terminating. Fixing a first order language with individual variables $x_0,x_1,x_2,...$ and certain relation and function symbols, a model $\mathcal{M}=(D,I)$ for such a language consists of a nonempty domain *D* and an interpretation function *I* which associates with every relation and function symbol of the language an appropriate relation or function in *D*. Every formula φ then defines a property in *S*, the set of all assignments to the variables - that is the set of all functions from $\{x_0,x_1,x_2,...\}$ to *D*. An element *s* in *S* has the φ-property iff φ is true in $\mathcal{M}$ under the assignment *s*, symbolically $\mathcal{M}\models_s\varphi$. The essential new case in defining truth is:

- $\mathcal{M}\models_s[x{:=}t]\varphi$ iff $\mathcal{M}\models_{s(x/t)}\varphi$

Therefore $[x{:=}t]\varphi$ is (under certain syntactical conditions) equivalent to $\varphi(x/t)$ which arises from φ by substituting *t* for every free occurrence of *x* in φ. Using the iteration operation it is possible to express things which in general cannot be said by any fixed number of applications of the substitution operation. For instance $[(x_1{:=}f(x_1))*]Q(x_1)$ expresses in $\mathcal{M}$ at *s* that all of

$$s(x_1),f^{\mathcal{M}}(f^M(s(x_1))),...$$

have the property corresponding to *Q* in $\mathcal{M}$. In such a language the natural numbers can now be characterized up to isomorphism by a formula. The main part of this formula says that for every *y* the program

$$(x{:=}0;\ \text{while } x\neq y \text{ do } x{:=}x+1)$$

terminates (at least once). This means that every element can be reached in finitely many steps from 0 by adding 1 each time. Here iteration is used essentially. From the characterizability of the natural numbers it follows in particular that the set of valid formulas is not recursively enumerable so that no usual proof system can be correct and complete. However, by adding an infinitary rule (a sort of ω-rule) correctness and completeness can be re-established.

Dynamic predicate logic can be viewed as a sublogic of an extended predicate logic in which countable

conjunctions and disjunctions are allowed. That is because a formula $[p*]\varphi$ is obviously equivalent of the disjunction over the set of formulas.

$\varphi, [p]\varphi, [(p;p)]\varphi, [((p;p);p)]\varphi, [(((p;p);p);p)]\varphi, \ldots$

Many variations of the core formalism are possible; for instance tests are often restricted to Boolean combinations of atomic formulas. This seems to be more realistic since tests of the form $(\forall x\varphi)?$ might involve a search through an infinite domain. Also, random assignments (written as $x:=?$) are sometimes introduced. In a model $\mathcal{M} = (D, I)$ such a basic random assignment denotes the relation which holds between two assignments *s* and s' if and only if $s' = s(x/d)$ for some $d \in D$. The formula $\forall x\varphi$ can then be expressed by $[x:=?]\varphi$.

See also: Logical and mathematical terms, glossary of

ULF FRIEDRICHSDORF

References and further reading

Fischer, M.J. and Ladner, R.E. (1979) ‘Propositional dynamic logic of regular programs’, *Journal of Computer and System Sciences* 18, 194-211.(Presents the first proof of decidability for propositional dynamic logic.)

Goldblatt, R. (1987) ‘Logics of Time and Comutation’, *CSLI Lecture Notes Number 7*, Stanford.(Good general reference for modal, tense and dynamic logic.)

Harel, D. (1984) ‘Dynamic logic’, in D. Gabbay and F. Guenthaer (eds) *Handbook of Philosophical Logic* vol. 2, Dordrecht: Reidel, 497-60.(Extensive handbook article presenting many results and some proofs. Contains valuable bibliographic information.)

Kozen, D. and Tiuryn, J. (1990) ‘Logics of programs’, in J. v. Leeuwen (ed.) *Handbook of Theoretical Computer Science* vol. B, Amsterdam: Elsevier, 789-840.(Another, perhaps more accessible, handbook of articles.)

Pratt, V.R. (1976) ‘Semantical considerations on Floyd-Hoare logic’, *Proceedings of 17th IEEE Symposium on Foundations of Computer Science*, 109-21.(Work initiating the subject.)

Segerberg, K. (1980) ‘Applying modal logic’, *Studia Logica* vol. 39, 275-95.(A good introduction to dynamic logic.)

East Asian philosophy

Sinitic civilization, which includes the Chinese-influenced cultures of Japan and Korea, established an early lead over the rest of the world in the development of its material culture - textiles, iron casting, paper, maritime arts, pottery, soil sciences, agricultural and water technologies, and so on. For centuries after the first sustained incursions of Europe into East Asia, there were more books printed in the classical Chinese language - the 'Latin' of East Asia - than in all of the rest of the world's languages combined. As recently as the beginning of the industrial revolution in the eighteenth century, it was China rather than Europe which, by most standards, was the arbiter of science and civilization on this planet.

If 'philosophy'- the pursuit of wisdom - is an aspiration of high cultures generally, why then was it not until the late nineteenth century, in response to a growing relationship with Western learning, that an East Asian term for 'philosophy' was coined, first by the Japanese (*tetsugaku*), and then introduced into Chinese (*zhexue*) and Korean (*ch'ôlhak*)? If it would be absurd to suggest that East Asian cultures have no history, no sociology, no economics, then how do we explain the fact that Asian philosophy is a subject neither researched nor taught in most Anglo-European seats of higher learning?

1 Uncommon assumptions, common misconceptions

The prominent French sinologist Jacques Gernet (1985) argues that when the two civilizations of China and Europe, having developed almost entirely independently of each other, first made contact in about 1600, the seeming resistance of the Chinese to embracing Christianity and, more importantly, the philosophic edifice that undergirded it was not simply an uneasy difference in the encounter between disparate intellectual traditions. It was a far more profound difference in mental categories and modes of thought, and particularly, a fundamental difference in the conception of human agency. Much of what Christianity and Western philosophy had to say to the East Asians was, quite literally, nonsense - given their own philosophic commitments, they could not think it. In turn, the Jesuits interpreted this difference in ways of thinking quite specifically as ineptness in reasoning, logic and dialectic.

The West has fared little better in its opportunity to appreciate and to appropriate Sinitic culture. In fact, it has fared so badly that the very word 'Chinese' in the English language, found in illustrative expressions from 'Chinese revenge' and 'Chinese puzzle' to 'Chinese firedrill', came to denote 'confusion', 'incomprehensibility' or 'impenetrability', a sense of order inaccessible to the Western mind. The degree of difference between a dominant Western metaphysical sense of order and the historicist 'aesthetic' order prevalent in the radial Sinitic world view has plagued the encounter between these antique cultures from the start. When seventeenth-century European savants such as Leibniz and Wolff were looking to corroborate their universal indices in other high cultures - the one true God, impersonal rationality, a universal language - China was idealized as a remarkable and 'curious land' requiring the utmost scrutiny. In the course of time, however, reported on by philosophers such as Kant, Hegel, Mill and Emerson, Western esteem for this 'curious land' plummeted from such 'Cathay' idealizations to the depths of disaffection for the inertia of what, in the context of the Europe-driven industrial revolution, was recast as a moribund, backward-looking and fundamentally stagnant culture.

In classical Chinese there is an expression: 'We cannot see the true face of Mount Lu because we are standing on top of it.' Although virtually all cultural traditions and historical epochs are complex and diverse, there are certain fundamental and often unannounced assumptions on which they stand that give them their specific genetic identity and continuities. These assumptions, extraordinarily important as they are for understanding the cultural narrative, are often concealed from the consciousness of the participants in the culture who are inscribed by them, and become obvious only from a perspective external to the particular tradition or epoch. Often a tradition suspends within itself competing and even conflicting elements which, although at odds with one another, still reflect a pattern of importances integral to and constitutive of its cultural identity. These underlying strands are not necessarily or even typically logically coherent or systematic, yet they do have a coherence as the defining fabric of a specific and unique culture.

Looking at and trying to understand elements of the East Asian cultural narrative from the distance of Western traditions, then, embedded as we are within our own pattern of cultural assumptions, has both advantages and

disadvantages. One disadvantage is obvious and inescapable. To the extent that we are unconscious of the difference between our own fundamental assumptions and those that have shaped the emergence of East Asian philosophies, we are sure to impose upon this geographical area our own presuppositions about the nature of the world, making what is exotic familiar and what is distant near. On the other hand, a clear advantage of an external perspective is that we are able to see with greater clarity at least some aspects of 'the true face of Mount Lu': we are able to discern, however imperfectly, the common ground on which the Confucian and the Buddhist stand in debating their differences, ground which is in important measure concealed from they themselves by their unconscious assumptions.

2 One-world natural cosmology

In the dominant world view of classical East Asia, we do not begin from the dualistic 'two-world' reality/appearance distinction familiar in classical Greek metaphysics, giving rise as it does to ontological questions such as: 'What is the Being behind the beings?' Rather, we begin from the assumption that there is only the one continuous concrete world that is the source and locus of all of our experience, giving rise to cosmological and ultimately ethical questions such as: 'How do these myriad beings best hang together?' Order within the classical East Asian world view is 'immanental' and 'emergent', an indwelling regularity in things themselves. It is the always unique yet continuous graining in wood, the distinctive striations in a piece of jade, the regular cadence of the surf, the peculiar veining in each and every leaf. The power of creativity resides in the world itself. The order and regularity this world evidences is neither derived from nor imposed upon it by some independent, activating power, but inheres in the world itself. Change and continuity are equally 'real'; time itself is the persistence of this self-transformation.

The 'one' world, then, is the efficient cause of itself. Situation takes priority over agency; process and change take priority over form and stasis. The context itself is resolutely dynamic, autogenerative, self-organizing and, in a real sense, alive. This one world is constituted as a sea of *qi*, psychophysical energy that disposes itself in various concentrations, configurations and perturbations (see Qi). There is an intelligible pattern (see Li) that can be discerned and mapped from each different perspective within the world (see De) that is its *dao*, a 'pathway' which can, in varying degrees, be traced out to make one's place and one's context coherent (see Dao). *Dao* is, at any given time, both *what* the world is and *how* it is, always as entertained from some particular perspective or another. In this tradition, there is no final distinction between some independent source of order, and what it orders. There is no determinative beginning or presumptive teleological end. The world and its order at any particular time is self-causing, 'so-of-itself' (*ziran*) (see Chinese philosophy; Daoist philosophy; Daodejing; Zhuangzi). Truth, beauty and goodness as standards of order are not 'givens': they are historically emergent, something done, a cultural product. Given the priority of situation over agency, there is a continuity between nature and nurture, a mutuality between context and the human being. In such a world, it is not unexpected that the *Yijing (Book of Changes)* is the first among the ancient classics (see Yijing).

3 *Ars contextualis*: the art of contextualizing

The 'two-world' metaphysical order inherited out of classical Greece has given the Western tradition a theoretical basis for *objectivity* - the possibility of standing outside and taking a wholly external view of things - a 'view from nowhere'. Objectivity is not only the basis for such universalistic claims as objective truth, impersonal reason and necessity, but further permits the decontextualization of things as 'objects' in our world. It is the basis on which we can separate objective description from subjective prescription.

By contrast, in the 'one world' of classical East Asia, instead of starting abstractly from some underlying, unifying and originating principle, one begins from one's own specific place within the world. Without objectivity, 'objects' dissolve into the flux and flow, and existence becomes a continuous, uninterrupted process. Each person is invariably experiencing the world as one perspective within the context of many. Since there is only the one world, we cannot get outside of it. From the always unique place one occupies within the cosmos of classical East Asia, one construes and interprets the order of the world around one as contrasting 'thises' and 'thats'- 'this person' and 'that person' - more or less proximate to oneself. Since each and every person or thing or event is perceived from some position or other, and hence is continuous with the position that entertains it, each thing is related to and a condition of every other.

In the human world, all relationships are continuous from ruler and subject to friend and friend, relating everyone as an extended ‘family’. Similarly, all ‘things’, like all members of a family, are correlated and thus interdependent. Every thing is holographic in entailing all other things as conditions for its continued existence, and is what it is at the pleasure of everything else. Whatever can be predicated of one thing or one person is a function of a network of relationships, all of which combine to give it its role and to constitute its place and its definition.

There is no strict notion of identity that issues forth as some essential defining feature - a divinely endowed soul, rational capacity or natural locus of rights - that makes all human beings equal. In the absence of such equality, the various relationships which define one thing in relation to another are qualitatively hierarchical and contrastive: bigger or smaller, more noble or more base, harder or softer, stronger or weaker, more senior or more junior. Change in the quality of relationships between things always occurs on a continuum as movement between such polar oppositions.

The general and most basic language for articulating such correlations among things is metaphorical: in some particular aspect at some specific point in time, one person or thing is ‘overshadowed’ by another; that is, made *yin* to another’s *yang*. Literally, *yin* means ‘shady’ and *yang* means ‘sunny’, defining in the most general terms those contrasting and hierarchical relationships which constitute indwelling order and regularity (see Yin-yang).

It is important to recognize the *interdependence* and correlative character of the *yin-yang* kind of polar opposites, and to distinguish this contrastive tension from the dualistic opposition implicit in the vocabulary of the classical Greek world, where one primary member of a set such as Being transcends and stands *independent* of, and thus is more ‘real’ than the world of Becoming. The implications of this difference between dualism and correlativity contrast are fundamental and pervasive.

To continue the ‘person’ example, generally in East Asian philosophy, a particular person is not a discrete individual defined in terms of some inherent nature, but is a centre of constitutive roles and relationships. These roles and relationships are dynamic, constantly being enacted, reinforced and ideally deepened through the multiple levels of natural, cultural and social discourse. By virtue of these specific roles and relationships, a person comes to occupy a place and posture in the context of family, community and world. The human being is not shaped by some given design which underlies natural and moral order in the cosmos, and which stands as the ultimate objective of human growth and experience. Rather, the ‘purpose’ of the human experience, if it can be so described, is more immediate; it is to coordinate the various ingredients which constitute one’s particular world here and now, and to negotiate the most productive harmony out of them. Simply put, it is to get the most out of what you have here and now.

4 Radial harmony

A major theme in Confucianism, foundational throughout East Asia, is captured in the phrase from *Analects* 13.23, ‘the exemplary person pursues harmony (*ho*), not sameness’ (see Confucian philosophy, Chinese; Confucian philosophy, Japanese; Confucian philosophy, Korean; Neo-Confucian philosophy). This conception of ‘harmony’ is explained in the classical commentaries by appeal to the culinary arts. In the classical period, a common food staple throughout northern Asia was *keng*, a kind of a millet gruel in which various locally available and seasonal ingredients were brought into relationship with one another. The goal was for each ingredient - the cabbage, the radish, the bit of pork - to retain its own colour, texture and flavor, but at the same time to be enhanced by its relationship with the other ingredients. The key to this sense of harmony is that it begins from the unique conditions of a specific geographical location and the full contribution of those particular ingredients readily at hand - *this* piece of cabbage, *this* fresh, young radish, *this* tender bit of pork and so on - and relies upon artistry rather than recipe for its success.

The Confucian distinction between an inclusive harmony and an exclusive sameness has an obvious social and political application, underscoring the fertility of the kind of harmony that maximizes difference. This ‘harmony’ is not a given in some preassigned cosmic design, but is the quality of the combination at any one moment created by effectively correlating and contextualizing the available ingredients, whether they be foodstuffs, farmers or infantry. It is not a quest of discovery, grasping an unchanging reality behind the shadows of appearance, but a profoundly creative journey where the quality of the journey is itself the end. It is the attempt to make the most of

any situation.

In summary, at the core of the classical East Asian world view is the cultivation of radial harmony, a specifically ‘centre-seeking’ or ‘centripetal’ harmony which is productive of consensus and orthodoxy. This harmony begins from what is most concrete and immediate - that is, from the perspective of any particular human being - and draws through patterns of deference from the outside in toward its centre. Hence there is the almost pervasive emphasis on personal cultivation and refinement as the starting point for familial, social, political and cosmic order (see Self-cultivation in Chinese philosophy). A preoccupation in classical East Asian philosophy, then, is the cultivation of this centripetal harmony as it begins with oneself, and radiates outward.

The East Asian world view is thus dominated by this ‘bottom-up’ and emergent sense of order which begins from the coordination of concrete detail. It can be described fairly as an ‘aestheticism’, exhibiting concern for the artful way in which particular things can be correlated efficaciously to thereby constitute the *ethos* or character of concrete historical events and cultural achievements. Order, like a work of art, begins with always unique details, from ‘this bit’ and ‘that’, and emerges out of the way in which these details are juxtaposed and harmonized. As such, the order is embedded and concrete - the colouration that differentiates the various layers of earth, the symphony of the morning garden, the wind piping through the orifices of the earth, the rituals and roles that constitute a communal grammar to give community meaning. Such an achieved harmony is always particular and specific, and is resistant to notions of formula and replication.

5 Philosophical syncreticism

As one might expect in a cultural narrative which privileges interdependence and the pursuit of radial harmony, orthodoxy is neither exclusive nor systematic. Rather, traditions are porous and syncretic. In the Han dynasty, for example, Confucianism is first fortified by elements appropriated from the competing schools of pre-Qin China such as Daoism and Legalism (see Legalist philosophy, Chinese). Later it absorbs into itself an increasingly Sinicized Buddhist tradition, evolving over time into a neo-Confucianism (see Neo-Confucian philosophy). At the same time, the *shuyuan* acadamies established by the great neo-Confucian syncretist Zhu Xi are modelled on Buddhist monastic schools. In more recent years the Western heresy, Marxism, and other elements of Western learning such as the philosophy of Kant and Hegel, are being appropriated by China and digested to produce what today is being called the ‘New Confucianism’ (see Marxism, Western; Marxism, Chinese).

The indigenous shamanistic tradition of Korean popular religion absorbed first Buddhism and then Confucianism from China, reshaping these traditions fundamentally to suit the uniqueness of the Korean social and political conditions (see Buddhist philosophy, Korean; Confucian philosophy, Korean). Native Japanese Shintoism emerges as a distinction made necessary by the introduction of first Buddhism and then Confucianism, where each tradition assumes a complementary function within the culture (see Shintō; Confucian philosophy, Japanese; Buddhist philosophy, Japanese). More recently, in the work of Kyoto School thinkers such as Nishida, Tanabe and Nishitani, German idealism is mined and alloyed with the Japanese Buddhist tradition to produce new directions.

Although Confucianism, Buddhism and Daoism - the dominant traditions of East Asia - have certainly been rivals at one level, it has been characteristic of the living philosophical traditions defining of East Asian culture to pursue mutual accommodation through an ongoing process of encounter and appropriation; hence the familiar expression *sanjiao weiyi*, ‘the three teachings (Confucianism, Buddhism and Daoism) are as one’. A continuation of this process is presently underway with the ongoing East Asian appropriation of Western philosophy.

See also: Ancient philosophy; Chinese philosophy; Indian and Tibetan philosophy; Japanese philosophy

ROGER T. AMES

References and further reading

Gernet, J. (1985) *China and the Christian Impact*, Cambridge: Cambridge University Press.(A reconstructed conversation between the Jesuits and Chinese intellectuals on their first encounters in the sixteenth and seventeenth centuries.)

Hall, D.L and Ames, R.T. (1995) *Anticipating China: Thinking Through the Narratives of Chinese and Western Culture*, Albany, NY: State University of New York Press.(A comparative study of the uncommon assumptions that ground the Chinese and Western philosophical traditions.)

Hall, D.L and Ames, R.T. (1997) *Thinking From the Han: Self, Truth, and Transcendence in China and the West*, Albany, NY: State University of New York Press.(An examination of several fundamental themes that distinguish the Chinese cultural narrative from the Western philosophical tradition.)

Mungello, D.E. (1985) *Curious Land: Jesuit Accommodation and the Origins of Sinology*, Honolulu, HI: University of Hawaii Press.(A discussion of the inquiry of the seventeenth-century European intellectuals into Sinitic culture.)

Eberhard, Johann August (1739-1809)

A German philosopher and theologian, Eberhard was trained in the rationalist tradition of Christian Wolff, but was also influenced by the more empirical 'popular philosophy' of the Enlightenment. Although a prolific author, who wrote on virtually every area of philosophy, he is best known today for his controversies with the two foremost German thinkers of his time, Gotthold Ephraim Lessing and Immanuel Kant.

Born in Halberstadt, Eberhard studied theology at Halle and entered the ministry in 1763. In 1778 he was appointed professor of philosophy at Halle and in 1786 he became a member of the Berlin Academy.

The controversy with LESSING was initiated by the publication of his first work, *Neue Apologie des Socrates (New Apology of Socrates)* (1772-8). Here Eberhard defended the heretical doctrine of the salvation of virtuous pagans (such as Socrates) and attacked the orthodox doctrines of original sin and eternal punishment. He also accused Leibniz of insincerity for his defence of the latter doctrine in his *Theodicy*. The attack on Leibniz occasioned Lessing's reply, which consists of an ironical defence of both Leibniz and the doctrine of eternal punishment, as well as a critique of what he took to be Eberhard's superficial rationalism. Eberhard responded to Lessing in the second volume of the *Neue Apologie*.

After this, Eberhard turned his interest to philosophy and published his *Allgemeine Theorie des Denkens und Empfindens (General Theory of Thinking and Sensing)* (1776). Although deeply influenced by Leibniz's *Nouveaux essais (New Essays)* (1704), he criticized Leibniz's account of innate ideas and defended Locke's view that all ideas derive from experience (see LOCKE, J.). The work was immediately successful, resulting in a prize from the Berlin Academy and the appointment at Halle. He was also quite successful as a teacher and exerted considerable influence on the young Friedrich Schleiermacher. He also published several handbooks on various aspects of philosophy, based largely on his lectures.

Eberhard initiated the controversy with Kant through his editorship of the *Philosophisches Magazin* (1789-91) and the *Philosophische Archiv* (1792-5). In these journals, which were created expressly to attack the Kantian philosophy, he published numerous articles, mostly written by himself and his colleague, J.G. Maass, criticizing major tenets of the *Critique of Pure Reason* (1781/1787) from a Leibnizian standpoint. The central theme of these articles is the superiority of the Leibnizian to the Kantian philosophy. It is claimed that whatever is true in Kant is already to be found in Leibniz and that wherever Kant departs from Leibniz he is wrong. In the course of arguing for this, however, Eberhard and Maass also raise important objections to Kant's distinction between analytic and synthetic judgments, the doctrine of the ideality of space and time, and the concept of an unknowable thing in itself. Kant, who usually refrained from engaging in direct polemics, made an exception in this case and responded in his *On a Discovery According to which Any New Critique of Pure Reason is Rendered Dispensable by an Older One* (1790). Eberhard is thus to be credited with occasioning this very significant clarification of Kant's views.

See also: Enlightenment, Continental

HENRY E. ALLISON

List of works

Eberhard, J.A. (1772-8) *Neue Apologie des Socrates (New Apology of Socrates)*, Berlin, 2 vols.
Eberhard, J.A. (1776) *Allgemeine Theorie des Denkens und Empfindens (General Theory of Thinking and Sensing)*, Berlin.
Eberhard, J.A. (1783) *Theorie der schönen Künste und Wissenschaften (Theory of Fine Arts and Sciences)*, Berlin.
Eberhard, J.A. (1787) *Vernunftlehre der natürlichen Theologie (Rational Doctrine of Natural Theology)*, Berlin.
Eberhard, J.A. (ed.) (1789-91) *Philosophisches Magazin*, 4 vols.

References and further reading

Allison, H.E. (1966) *Lessing and the Enlightenment*, Ann Arbor, MI: The University of Michigan Press.(Contains analysis of both Eberhard's theological views and Lessing's critique.)
Allison, H.E. (1973) *The Kant-Eberhard Controversy*, Baltimore, MD and London: Johns Hopkins University

Press.(Contains an English translation of Kant’s response to Eberhard and an analysis of the controversy.)
Beiser, F. (1987) *The Fate of Reason, German Philosophy from Kant to Fichte*, Cambridge, MA: Harvard University Press.(Chapter 7 contains an account of the controversy with Kant.)

Ecological philosophy

In the early 1970s a small number of academic philosophers in the English-speaking world began to turn their attention to questions concerning the natural environment. Environmental philosophy initially encompassed various types of inquiry, including applied ethics oriented to issues such as nuclear power and the deployment of toxic chemicals; more abstract extrapolations of traditional ethical theories, such as Kantianism and utilitarianism, into environmental contexts; and, also, a far more radical project involving the reappraisal of basic presuppositions of Western thought in the light of their implications for our relation to the natural world. The first two were basically extensions of existing areas of philosophy, and it is arguably the third project - often described as 'ecological philosophy' or 'ecophilosophy' - which constitutes a distinctively new branch of philosophy. It is to environmental philosophy in this third sense that the present entry is devoted.

Although the ecophilosophical project was explicitly normative in intent, it was quickly found to entail far-reaching investigations into the fundamental nature of the world. Indeed it was seen by many as entailing a search for an entirely new ecological paradigm - a worldview organized around a principle of interconnectedness, with transformative implications for metaphysics, epistemology, spirituality and politics, as well as ethics. Moreover, the process of elaborating a new ecological view of the world was found to uncover the contours of an already deeply embedded worldview, organized around a principle of separation or division, underlying and shaping the traditional streams of modern Western thought.

1 Origins of ecophilosophy

In its earlier phases ecophilosophy was primarily (though not exclusively) associated with a movement known as 'deep ecology', a term coined in 1973 by the Norwegian philosopher Arne Naess. Naess contrasted 'shallow environmentalism', which sought environmental reform within existing philosophical and ideological parameters, with the 'deep, long-range ecology movement', which involved challenging prevailing philosophical and ideological assumptions about our relation to the natural world. Naess provided his own versions of these challenges and gave his own responses to them, but arguably his most important contribution to environmental philosophy has been to open up this new horizon and to reveal the radical potential of the ecological approach.

Ecophilosophy can no longer be encompassed within the various 'platforms' of deep ecology initially enunciated by Naess and his followers. However, although the 'deeper' questions that ecology poses are now being explored in a variety of ways, many ecophilosophers agree on the broad outlines of the hitherto dominant worldview and its emerging ecological alternatives.

2 The classical scientific worldview

The dominant worldview is sometimes called Newtonian, sometimes Cartesian. It was forged during the scientific revolution of the seventeenth and eighteenth centuries, though it is often seen as having antecedents in both classical Greek thought and Christianity. It is Newtonian insofar as it is atomistic and mechanistic, and Cartesian insofar as it is dualistic. From a Newtonian point of view, the natural world is made up of *atoms* - discrete particles, units that exist independently of one another. It is built like a machine out of aggregates held together not by any mutual affinity or intrinsic dynamism but purely *mechanically*, by blind, external laws of motion. (Gravitation was an anomaly for Newton' an 'occult' force which vitiated the simplicity of the mechanistic picture.) Such an atomistic world is fully quantifiable and hence describable in mathematical terms, since it can be reduced to nothing but countable units in measurable spatial relations. It is nothing but the mechanical sum of its inert parts.

To characterize the natural world in this way drained it of all of the attributes associated with mind - attributes such as agency, the power of self-mobilization, subjectivity, soul, meaning and purpose. It followed either that mind was illusory, a phenomenon which, as Hobbes argued, could be reduced to mechanism, or that mind was no part of the natural system, but, as Descartes claimed, a distinct category of substance, ontologically independent of matter (see Substance §2).

It was the Cartesian *dualism*, rather than the Hobbesian reduction, which came to inform the worldview that emerged from classical science. Mind, banished from matter' took up its dwelling place in us. Because the world

in which we exist was considered to be made up of discrete parts, none of which was intrinsically connected with any other, mind could indeed be sealed away in human brains in this way, and, so to speak, not leak out into the surrounding world. This cultural preference for the Cartesian over the Hobbesian view might perhaps be attributed to the fact that the nihilism implicit in Hobbes was relatively unlivable, whereas the Cartesian approach was ideologically appealing.

For if humanity was the repository of all meaning and purpose, then it was also the exclusive locus of value; matter as the inert and dead, the purposeless and blind, possessed only the meaning and hence value that we projected onto it This flattering image of humanity as categorically elevated above the rest of nature had an obvious appeal in an expansive and ambitious age. Moreover, the natural world, divested of intrinsic value, was rendered a fit object for human use, significant only as a reservoir of resources for humankind. This dissolution of traditional constraints on the exploitation of the natural environment obviously suited the mercantile ends of the early modern era.

In reinforcing the assumption that humanity is the only proper object of moral concern and the only yardstick of meaning and value in life, the scientific worldview is seen by ecophilosophers as entrenching the human centredness or *anthropocentrism* that was to a certain extent already, in the pre-modern period, characteristic of Western thought, with its Judaeo-Christian and classical Greek and Roman origins.

Ultimately, the classical scientific worldview and the anthropocentrism which accompanies it may be seen to rest on a fundamental principle of division or separation: reality at every level, indeed thought itself, is separated out into units, or divided into opposing categories. Not only is the natural world divided into atoms, and mind divorced from body, but God, as spirit, stands apart from and above the realm of matter; the epistemic subject is constituted by its separation and strict 'detachment' from the object; society is seen as the sum of the discrete and autonomous individuals which are its units; thought proceeds by analysis, by the consideration of parts.

3 Ecological worldviews

If the 'dominant worldview' emphasizes separation, alternative, ecological worldviews rest on a principle of interconnectedness. 'Ecology' is not intended here in its literal sense, as denoting merely a branch of biological science. The science of ecology, with its representation of the living world as a system of relationships, is in itself neutral with respect to its deeper metaphysical implications. In ecophilosophy, however, this science serves as a *model* - an appropriate theoretical image for a set of irreducible ontological interdependencies.

From this ecological perspective, the world is not seen as divided into mutually independent parts and mutually exclusive attributes: everything is seen as implicating, and being implicated in, the identities of other things, reality being a relational system of shared, interpenetrating essences. Since the properties of a given individual are a function of its relations with individuals of countless other kinds, these properties do not belong exclusively to it: each individual owes its nature to others in the network. It follows that meaning-giving and value-endowing qualities such as mind or subjectivity or soul, and the value that accompanies them, cannot be monopolized by human beings, but must be diffused throughout the systems of the natural world.

In restoring to reality its intrinsic interconnectedness, the ecological perspective implies that since non-human others are implicated in our identity, they are also implicated in our ends and interests, and hence in our conception of the good. In reanimating the world with a quality analogous to mind, this perspective furthermore implies that nature is morally significant in its own right, and is not merely a stockpile of resources for humankind. Such an *ecocentric* perspective situates humanity *in* nature rather than apart from and above it, and thereby removes the traditional justification for anthropocentric attitudes. The idea of interconnectedness can function as an organizing principle not only within metaphysics and ethics, but in many different theoretical and practical contexts. For example, from the interconnected perspective, the epistemic subject is not constituted by way of separation or detachment from its object. Rather, the subject seeks understanding of 'the other' through empathy, affinity, a sense of kinship - through dialogue rather than the objectifying gaze of traditional science. In the theological context, spirit is seen as immanent in matter, body, Nature. Spirituality thus involves celebration of and reverence for our corporeal existence, and affirmation of the natural world and our enmeshment in it. In the sociopolitical context, society is seen not as an aggregate of social atoms - essentially mutually independent individuals - but as a collective, in which individuals are constituted through their relations with one another.

The idea of interconnectedness, however, has been interpreted in different ways by different authors. Perhaps the most important divergence is that between broadly holistic and relational interpretations. According to the more holistic interpretations, interconnectedness implies that individuals exist or are significant only at the level of appearance: individuality lacks ultimate ontological significance. In this case, the individual may be seen as ’one with’ the greater whole.

The holistic interpretation can itself be read in two ways. Either I see nature as an extension of myself, so that its interests become indistinguishable from my interests, and ’eco-defence’ becomes a matter of self-defence. Or I see myself as an outgrowth or expression of nature, a purely natural phenomenon, so to speak, without independent identity or interests In this case I substitute nature’s interests for my own. From this holistic point of view then, I either absorb nature conceptually into myself, or lose myself in nature. Holistically oriented ecophilosophers have tended to opt for the former reading.

According to the more relational interpretations of interconnectedness, individuals are real, but their reality is a function of their relationships: it is through their interactions that they maintain themselves and define their boundaries. The natural world thus appears not as a field or continuum, but as a community, in which individuals are internally related, in the manner of a family, but are nevertheless distinct. It is incumbent on us, from this point of view, to recognize our kinship with other individuals, but also to respect their otherness. Empathy is required, but not outright identification.

The holistic interpretation of interconnectedness has been identified, by certain critics, with deep ecologists. Some feminists have argued that metaphysics recapitulates psychology, in the sense that different views of the structure of the world emanate from different ego psychologies. In particular, the atomistic outlook is taken to reflect a psychology of separateness associated, in certain (‘object relations’) branches of feminist psychoanalytic theory, with the masculine self, while a relational outlook is taken to reflect a relational psychology associated with the feminine self. Some feminists argue that when ecophilosophers postulate an holistic metaphysic, in an attempt to escape the alienation of the atomistic self, they inadvertently perpetuate this alienation, by replacing the isolated local self with an inflated but still isolated global self. From this point of view, then, true interconnectedness calls for relationality rather than holism, for self-in-relation rather than the expansion of self into a greater unity.

Alhough there is much ongoing ambivalence among ecophilosophers as to the exact nature of ecological interconnectedness, the holistic and relational interpretations (and the relational form of individualism that the latter interpretation implies) may be construed as complementary: interconnectedness may be seen as entailing the identities of both wholes and (relational) individuals. Individuals may be seen as energy configurations or sub-systems which maintain their integrity by continuous give and take with the wider system within which they are nested. Their relationality thus entails their individuality, but the relational constitution of the wider system ensures that that system, too, possesses an inherent unity and integrity which makes it an indivisible whole.

4 Three streams of ecophilosophy

Three broad streams of ecophilosophy have emerged to date, though a number of individual ecophilosophers remain unaffiliated with any of them. These streams are deep ecology, ecofeminism and social ecology, and though all are broadly ecocentric in their outlook, their critiques of the dominant worldview and their elaborations of the idea of ecological interconnectedness differ to a certain extent. They each impute the contemporary environmental crisis to the anthropocentric underpinnings of Western thought; but they offer somewhat competing analyses of this anthropocentric premise and differently nuanced remedies for it.

While deep ecologists have been most effective in bringing the problem of anthropocentrism to light, they have generally been less concerned than either ecofeminists or social ecologists with analysing its origins in the Western tradition. This relative lack of concern reflects a shying away, in deep ecology, from systematic approaches generally. Early in the development of deep ecology, Naess drew a distinction between deep ecology as a ’platform’, on the one hand, and deep ecology as a collection of ’ecosophies’, on the other. The deep ecology platform, which served as the manifesto for deep ecology as a radical political movement, consisted of an (evolving) set of principles on which all who counted themselves deep ecologists could agree. These principles have in the past included such propositions as that all beings have an equal right to live and blossom (’biocentric egalitarianism’), and that the wellbeing and flourishing of non human life on earth has intrinsic value.

’Intrinsic value’ is here understood as value which is independent of the usefulness of non human life for narrow human purposes (see Values). An ecosophy is any philosophical theory or tradition which can ground this set of principles (see Drengson and Inoue 1995).

Naess’ own ecosophy is an amalgam of various influences, including Spinoza, Gandhi, logical positivism and the nature-centredness of the Norwegian folk tradition. It is articulated by Naess as a philosophy of ‘self-realization’, according to which individuals can overcome their culturally acquired alienation from the natural world by recognizing the ultimate ontological seamlessness of reality, and gradually identifying themselves with wider and wider circles of being, until their sense of self encompasses not only the personal ego, but family, community, bio-region, land, society at large and eventually the planet as a whole. As they realize this larger Self, individuals will perceive their own interests as converging with those of the rest of life.

Ecofeminists tend to trace the human/nature split that defines anthropocentrism to a mind/body dualism which has its earliest origins in antiquity. Like ecophilosophers generally, they recognize the importance of resolving this schism, of making the partitioned world whole again, and of re-animating a ‘dead’ nature. However, ecofeminists identify the mind/body divide as part of a wider, indeed comprehensive system of dualistic thought, the purpose of which is to divide the world conceptually into mutually opposing and hierarchically ranked categories. This system revolves around certain core dichotomies, which, in addition to mind/body, include spirit/matter, subject/object, culture/Nature, reason/emotion and reason/instinct. Under the influence of these dualistic categories, Western culture has, over approximately the last 2000 years, developed a view of the world as divided into things which possess mind or reason and things which lack it, where the former are set above the latter, and the moral significance of the latter is discounted. Reason is paradigmatically aligned with humankind, and unreason with a mechanical nature, which can accordingly then be subjected to the kinds of instrumental and exploitative attitudes that have led to the current environmental crisis.

Ecofeminists, unlike deep ecologists, offer a political analysis of the dualistic system of thought: this system serves as the naturalizing and legitimating framework for hierarchical political regimes. Theorists such as Val Plumwood (1993) show that the reason/nature dichotomy is used not only to elevate humanity above nature, but to rank one group above another within society: different social groups are assessed according to their perceived relative identifiability, or otherwise, with reason or mind. Thus, in Western cultures, men have traditionally appropriated reason (and hence mind, spirit, intellect and the subject position), while women have been consigned to nature (and hence to the body, matter, emotion, instinct and the object position). White colonizers have identified themselves with reason (civilization, science, enlightenment) and adopted the masculine subject position, while identifying the colonized with nature (primitivism, superstition, barbarism), thereby relegating them to the object position. So, too, with the ruling classes in capitalist societies; they have defined themselves in terms of reason (mental work), while defining workers in terms of the body or nature (manual work). Other groups of oppressors characterize themselves, and those they oppress, according to the same self-serving dualistic logic.

The dualistic construction of nature, with its disastrous consequences for the environment, can thus be seen as an ideological instrument of political oppression. From this point of view nature cannot be ideologically rehabilitated until the political agenda that is served by its being rendered inferior is exposed and addressed. At the same time, however, political oppression cannot be eliminated until the assumption that there exists anything corresponding to the idea of a pure object, to which subordinated groups can be conceptually assimilated, is rejected. In other words, we must respect and engage with nature as a subject in its own right, before we can respect and engage with the social ’other’ whose subjectivity has also been denied.

From this kind of ecofeminist perspective, environmentalism is inextricably linked with other social movements, such as those against sexism, racism and class oppression: ideologies of gender and race and class cannot be recast without an overhaul of our understanding of nature, and the ideological rehabilitation of nature cannot be achieved without the concurrent rehabilitation of women, colonized races, and other oppressed groups.

The approach of social ecology to the problem of anthropocentrism closely parallels that of the kind of ecofeminism just outlined. Murray Bookchin, the main architect of social ecology, argues (1982) that the human impulse to dominate Nature is formed in hierarchical societies, and will not be overcome until the social conditions which generate ’epistemologies of rule’ or psychologies of domination are corrected. This will not occur, according to Bookchin, until hierarchy in all its forms - familial, gerontological, institutional as well as political -

is eradicated from society.

Like ecofeminists and deep ecologists, Bookchin regards nature as a field of relations holistically imbued with an active attribute which confers upon it meaning and moral significance. He describes this attribute as ’subjectivity’. However, unlike ecofeminists and deep ecologists, Bookchin emphasizes the evolutionary aspect of ecology: Nature is not just an assemblage of interrelated parts but a process whereby new and more complex unities are continually being generated through increased differentiation of their elements. In other words, an ecosystem is constituted not only through its internal relations, but also through its history and its telos. In light of this recognition of the evolutionary drive towards ever greater complexity, Bookchin draws a qualitative line between humankind and nature. While human subjectivity (’second nature’) is ontologically continuous with the subjectivity of the non human realm (’first nature’), inasmuch as it derives from and recapitulates the latter, it nevertheless transcends it, in the sense that subjectivity has in us attained new levels of complexity and self-organization. For Bookchin then, harmony with the natural environment is achieved not by a simple identification with Nature, such as deep ecologists prescribe, nor merely by a resolution of dualistic opposites - putting mind back into matter - as ecofeminists recommend. From his evolutionary point of view, we achieve harmony with nature, and save the planet, by allowing our own nature the freedom to unfold in its own essential way. Since our nature was formed in a thoroughly ecological system, and since the telos of such a system is the continual elaboration of greater complexity and further diversity, this is also the telos of our nature. We can accordingly only flourish in a society which preserves or recreates, or even enhances, at both social and biological levels, the ecological dynamic of our original matrix. Social relations as well as relations with the natural environment will have to be ecological, and therefore nonhierarchical, if we are to continue to evolve more complex and diverse forms of subjectivity or mentality. Our own self-realization, as individuals, and the further evolution of humanity, thus in fact call for the very kind of nonhierarchical society which, by eliminating psychologies of domination, ensures that our relations with the natural world are reciprocal and benign.

5 Foundations of the ecological view

The intuition of ecological interconnectedness has been epistemologically grounded and philosophically elaborated in ways beyond those already canvassed. For some authors any attempt to provide a theoretical foundation for this intuition is seen as misguided, since a sense of ecological interconnectedness is, according to them, properly derived from direct experience of nature in the field. Other authors appeal for justification to certain ’minority traditions’ within Western philosophy, such as pantheism, Romanticism, the monism of Spinoza or the process philosophy of Alfred North Whitehead. Eastern traditions, such as Daoism and Buddhism, and the spiritual traditions of indigenous peoples, are also invoked. Developments in disciplines other than philosophy, including the ’new physics’ (Capra 1983; Bohm 1980), systems theory (Bateson 1979; Lovelock 1979), biology (Maturana and Varela 1988) and cultural studies (Roszak 1973; Berman 1981) are cited by yet other authors. In all these areas, traditional assumptions about nature are being overturned and reorganized along more holistic or relational lines. Naess’ reminder that many different traditions and theories can provide the starting point for an ecological view of the world can be considered an essential aspect of the ecophilosophical project. His invitation to people to construct their own ’ecosophies’, drawing on direct experience and philosophical ideas that are familiar and congenial to them, implies that premises couched in terms as different as mythology, religion, science and mathematics may yet lead to conclusions convergent in relevant respects. Commitment to an ecological worldview, then, is not inconsistent with a high degree of epistemological pluralism and a rich diversity of cultural expression.

The idea that everything to a certain extent shares the fate of everything else is in fact the spiritual axiom of many pre-modern and non-Western societies. It may have been the rejection of this idea, in favour of the idea of division or separation, that helped to bring about the estrangement between Western philosophy and spirituality in the modern period. Perhaps - though some ecophilosophers would disagree - ecophilosophy represents a first step toward the reunion of these two pathways to understanding.

See also: Environmental ethics; Green political philosophy; Monism; Naturphilosophie; Pantheism

FREYA MATHEWS

References and further reading

Bateson, G. (1979) *Mind and Nature*, London: Wildwood House.(Systems approach to mind and nature.)
Berman, M. (1981) *The Re-enchantment of the World*, Ithaca, NY: Cornell University Press.(Study of the place of science in the history of consciousness.)
Bohm, D. (1980) *Wholeness and the Implicate Order*, London: Routledge & Kegan Paul.(A holistic cosmology based on the new physics.)
Bookchin, M. (1982) *The Ecology of Freedom*, Palo Alto, CA: Cheshire Books.(A source book for the metaphysical foundations of social ecology.)
Brennan, A. (1988) *Thinking About Nature*, London: Routledge.(Includes account of the metaphysical implications of scientific ecology.)
Callicott, J.B. (1989) *In Defense of the Land Ethic*, Albany, NY: State University of New York Press.(Part II includes discussion of metaphysical implications of ecology; Part IV examines material relevant to §§2 and 5.)
Capra, F. (1983) *The Turning Point*, London: Fontana.(An influential popular account of the new physics.)
Drengson, A. and Inoue, Y. (eds) (1995) *The Deep Ecology Movement*, Berkeley, CA: North Atlantic Books.(An anthology of seminal deep ecology articles by Arne Naess and others. For Naess bibliography see Rothenberg, 1993.)
Fox, W. (1990) *Toward a Transpersonal Ecology*, Boston, MA and London: Shambhala.(Overview of deep ecology literature.)
Lovelock, J. (1979) *Gaia: A New Look at Life on Earth*, Oxford and New York: Oxford University Press.(A systems account of the biosphere.)
Mathews, F. (1991) *The Ecological Self*, London: Routledge.(Expansion of material in §§2-3.)
Maturana, H. and Varela, F. (1988) *The Tree of Knowledge: the Biological Roots of Human Understanding*, Boston, MA: Shambhala.(A systems account of living things.)
Plumwood, V. (1993) *Feminism and the Mastery of Nature*, London: Routledge.(An ecofeminist analysis of dualistic patterns of thought and critique of traditional philosophies of nature.)
Rolston, H. (1986) *Philosophy Gone Wild*, Buffalo, NY: Prometheus Books.(See especially Section I.)
Rolston, H. (1988) *Environmental Ethics*, Philadelphia, PA: Temple University Press.(This seminal environmental philosopher interweaves ethical and ontological questions, and offers a distinctively evolutionary perspective on ecology.)
Roszak, T. (1973) *Where the Wasteland Ends: Politics and Transcendence in Postindustrial Society*, Garden City, NY: Anchor/Doubleday.(Nature from the viewpoint of the cultural history of the West.)
Roszak, T. (1979) *Person/Planet: The Creative Disintegration of Industrial Society*, London: Victor Gollancz. (Cultural study of human identity in light of environmental crisis.)
Rothenberg, D. (1993) *Is it Painful to Think? Conversations with Arne Naess*, Minneapolis, MN: University of Minnesota Press.(Extensive Naess bibliography.)
Taylor, P. (1986) *Respect for Nature*, Princeton, NJ: Princeton University Press.(Chapter 3 is relevant to material in §3 on individual identity from the ecological viewpoint.)
Whitehead, A.N. (1929) *Process and Reality: An Essay in Cosmology*, New York: Macmillan, and Cambridge: Cambridge University Press; corrected edn, with comparative readings and detailed index, by D.R. Griffin and D.W. Sherburne, New York: The Free Press, 1978.(Whitehead's central work in speculative metaphysics, presented in axiomatized form with extensive applications.)

Ecology

Philosophers of science have paid relatively little attention to ecology (compared to other areas of biology like evolution and genetics), but ecology poses many interesting foundational and methodological problems. For example, the problems of clarifying the differences and causal connections between the various levels of the ecological hierarchy (organism, population, community, ecosystem…); the issue of how central evolutionary biology is to ecology; long-standing issues concerning the extent to which the domain of ecology is more law-governed or more a matter of historical contingency, and the related question of whether ecologists should rely more on laboratory/manipulative versus field/comparative methods of investigation.

1 The ecological hierarchy

Ecology is most often characterized as the study of the ’distribution and abundance of organisms’. For example, ecologists ask, what biotic factors (for example, predation, interspecific competition for resources) and/or abiotic factors (for example, climate) are responsible for the presence or absence of a species in a particular location, or for the diversity of species in a particular area? More specifically (to pose a few famous ecological questions), why are there so many different species on the face of the earth? Are there just as many species as there are niches (and what are ’niches’, anyway)? Why are there so many different species of plankton in a lake? Why are big, fierce animals rare?

Ecology is also characterized in terms of the hierarchy of entities that constitute its domain: organisms, populations (for example, the population of a particular species of bass in a lake), communities (for example, the interacting populations of bass, minnows and plankton in a lake), ecosystems (for example, the lake, including the bass, minnows and plankton), … the biosphere. (For the sake of comparison or contrast, the genetic/evolutionary hierarchy consists of genes, organisms, populations, species, genera, families, orders, classes … kingdoms.)

The ecological hierarchy imposes or affords two directions of causal analysis for each level. Thus, for example, ecology includes the study of changes in population size as influenced by traits of the organisms that compose the populations (’upward’ causation, in the sense of ’up’ the hierarchy), and the study of changes in population size in the context of the communities to which the populations belong (’downward’ causation). Ecologists study changes in community structure as influenced by the populations that compose the communities, and in the context of the ecosystems to which the communities belong. And so on. Among the most basic foundational issues that arise in ecology are the problems of clarifying the differences and causal connections between the different levels of the ecological hierarchy.

The diversity of spatial and temporal scales associated with the variety of ecological entities and processes is also an important consideration in many of the foundational and methodological issues that arise in ecology. Consider for instance the difficulty in characterizing the nature of ’communities’ - a maple-birch forest together with the deer that inhabit it may be considered one community, and the various interacting species of microorganisms that inhabit the gut of any one of those deer may be considered another community. To make matters more difficult, the species’ composition of a community may change over time as new species colonize the area in question and outcompete other species, and so on. At what point do we say that one community has been replaced by another?

2 Superorganisms

The upper levels of the ecological hierarchy have from time to time been conceived quite holistically, even as higher level organisms or ’superorganisms’. For instance, just as organisms of a particular species undergo a fairly regular sequence of developmental stages to adulthood, so too the species’ composition of a disturbed area has been thought to undergo a fairly regular sequence of ’successional’ stages back to the ‘mature’ state (for example, from the weedy annuals that first colonize a forest clearing, to the perennial herbs, to the shrubs, and ultimately to the trees that reconstitute the mature forest). Just as individual organisms display self-regulatory abilities (like the ability to maintain body temperature within a narrow range), so too populations, communities and ecosystems have been conceived as having self regulatory properties (like the ability of a population to regulate its size). Just as organisms of different species have distinctive metabolisms, so too different communities and ecosystems have distinctively different energy flows and nutrient cycles. Still more analogies have been pursued.

The superorganism concept was also motivated by methodological considerations. Ecology emerged as a self-conscious discipline in the early twentieth century when the methodological model for biology was experimental physiology, with its emphasis on the repeatable, the manipulable and the verifiable. This was considered the best antidote to excessive interest in evolution, which seemed bound-up with the historically unique, the nonmanipulable and the nonrecoverable. Early ecologists advocated the investigation of superorganisms in the manner in which experimental physiologists investigated organisms.

That is not to say that the superorganism concept is, or was, inherently anti-evolutionary. It has played an important role in 'group selection' theories (see Evolution, theory of; Sociobiology). For example, just as Darwinians have traditionally explained the prevalence of a type of organism in terms of natural selection for organisms of that type, so too some ecologists have explained the prevalence of some types of populations, communities and ecosystems in terms of the selective persistence of systems of those types. Superorganismal, group-selectionist concepts in ecology have been the target of considerable criticism from some evolutionary biologists, who argue that the concepts reflect a lack of understanding of evolution. However, these issues have been a fruitful source of discussion concerning the conceptual relationships of ecology to evolutionary biology (more about the connections between ecology and evolutionary biology later).

Many who have rejected superorganism concepts in ecology have nonetheless also rejected the idea that the higher level entities are merely interacting collections of the lower level components. They have held instead that the lower level components interact in such a way as to contribute to the integrity of the upper level entity. So for example, communities and ecosystems are understood by many to be integrated into food webs, which are in turn structured largely by predator-prey relations and interspecific competition for available prey/resources. This integrity has been characterized most generally as 'stability'. But this stability has proved difficult to characterize in a general way. Numerous variants of the concept have been elaborated: neighbourhood stability, global stability, dynamic stability, bounded stability, cyclic stability, trajectory stability, resistance, resiliency, constancy, elasticity, persistence, inertia, and more. Compounding the difficulty of specifying the nature of the integrity of higher level ecological entities is the scale - temporal and spatial - at which, and during which, the supposed stability is maintained.

3 Methodological issues

Proponents of the importance of evolutionary thinking in ecology have emphasized the role of natural selection in contributing to community and ecosystem integrity and stability. That is, to the extent that natural selection results in the mutual adaptation of the member species of a community, the community becomes a more integrated and stable whole. Arguments for the importance of evolutionary thinking in ecology have hinged on several issues, one of which concerns 'evolutionary time' versus 'ecological time'. Ecological time is often considered much shorter - too short for ecological processes to be influenced by evolutionary changes. Proponents of the importance of evolutionary thinking in ecology have challenged the distinction, arguing, for example, that evolution by natural selection occurs in a short enough time period to have a significant effect on the outcome of an ecological process. Consider for instance an ecological process like interspecific competition for resources. One possible outcome would be extirpation of one of the species from the area - a purely ecological phenomenon. But this is not the only possible outcome: one of the species could evolve in a direction that allows its members to make use of a resource not utilized by the other species. In this case, evolution would allow both species to coexist within a community.

The connections between ecology and economics may be as significant as the connections between ecology and evolutionary biology. Those connections are well evidenced by the importance within ecology of such concepts as resource, competition, consumer, producer, productivity, efficiency, energy budget, guild, and so on.

One of the major conceptual divides in ecology has been between those who view upper level entities of the ecological hierarchy as real entities having some characterizable integrity, and those who view the upper levels more as combinations of populations that coincidentally migrated to the same area and that coincidentally have the same range of environmental tolerances. Of course, this is a continuum, not a strict dichotomy of outlooks. Not only might any particular community lie between the extremes, but some communities might lie more towards one end of the spectrum, while other communities lie towards the other end. But many ecologists have pursued (and continue to pursue) the extremes.

Associated with this difference in perspective are other conceptual and methodological differences. Proponents of the more holistic viewpoint have tended to emphasize and focus on optimal equilibrium states in which higher level entities persist, and/or towards which they develop, and/or towards which they return when disturbed. Proponents of equilibrium viewpoints tend to belittle the importance of historical contingency, inasmuch as equilibrium states can be reached by different routes. Proponents of the alternative viewpoint tend to question the reality of equilibrium states, and stress the importance of historical contingencies. Although, again, this is a continuum of perspectives, not a strict dichotomy. And once again, the temporal and spatial scale of the investigation is relevant to the apparent importance of historical accident. While ecology distanced itself from history in its early days, and while leading ecologists continued to stress the differences between ecology and the historical sciences through the 1970s, there are now calls for a ’new ecology’ that takes historical contingency more seriously.

Discussions of the merits and implications of historical contingency versus equilibrium approaches often involve discussions of the predictiveness and nomic status of ecological theory. That is, to the extent that historical contingency prevails, it has been suggested, ecological theory will be less predictive and less law-like. Equilibrium theories are not the only way to bring predictiveness and law-like order to ecology; but they have long been held as the best hope. After all, equilibrium points are by definition predictable/inevitable as long as specifiable ’disturbing’ conditions do not obtain. Equilibrium specifications thus have just the right form for *ceteris paribus* laws.

A re-evaluation of the importance of historical contingency is also occurring in evolutionary biology. In ecology, as in evolutionary biology, concerns about history have led to discussions of methodology, and to criticism of the idea that the only proper methodological models are laboratory-based sciences such as particle physics and molecular biology. While ecology aligned itself methodologically with laboratory experimental physiology in its early days, it has recently been suggested that it might as well or better model itself on historical disciplines like astronomy and geology (see Geology, philosophy of). A wide range of experimental methodologies, from laboratory to field, and from manipulative to comparative, is employed in such disciplines and seems to have an important place in ecology as well.

See also: Ecological philosophy; Environmental ethics; Holism and individualism in history and social science; Species; Taxonomy

JOHN BEATTY

References and further reading

Diamond, J. and Case, T.J. (eds) (1986) *Community Ecology*, New York: Harper & Row.(Still a fairly state-of-the-art collection of essays addressing and exhibiting the ‘plurality’ of theories and methods appropriate to understanding this central level of the ecological hierarchy. Included are excellent essays on the equilibrium/law versus the nonequilibrium/history continuum of theories, and the laboratory manipulation versus field/comparative continuum of methodologies.)

Hagen, J.B. (1992) *An Entangled Bank: The Origins of Ecosystem Ecology*, New Brunswick, NJ: Rutgers University Press.(An historical introduction to foundational issues in community and ecosystems ecology.)

Keller, E.F. and Lloyd, E.A. (eds) (1992) *Keywords in Evolutionary Biology*, Cambridge, MA: Harvard University Press.(Includes entries for a number of ecological concepts, including ’Community’, ’Competition’, ’Environment’, ’Group Selection’, ’Niche’, ’Resource’, ’Unit of Selection’.)

Kingsland, S.E. (1985) *Modelling Nature: Episodes in the History of Population Ecology*, Chicago, IL: University of Chicago Press.(A historical introduction to foundational and methodological issues in population ecology; the ’Afterword’ to the second edition, 1996, includes a discussion of the new emphasis of history in ecology.)

Kormondy, E.J. (ed.) (1965) *Readings in Ecology*, Englewood Cliffs, NJ: Prentice Hall.(A nice collection of classic papers, with short commentaries.)

McIntosh, R.P. (1985) *The Background of Ecology: Concept and Theory*, Cambridge: Cambridge University Press.(A comprehensive survey - somewhat historical, somewhat philosophical - of foundational and methodological issues in ecology.)

Peters, R.H. (1991) *A Critique for Ecology*, Cambridge: Cambridge University Press. (A somewhat old-fashioned but comprehensive criticism of - rather than an attempt to understand - the level of generality and predictiveness in ecology.)

Real, L.A. and Brown, J.H. (eds) (1991) *Foundations of Ecology: Classic Papers with Commentaries*, Chicago, IL: University of Chicago Press.(Just as the title says.)

Shrader-Frechette, K.S. and McCoy, E.D. (1993) *Method in Ecology: Strategies for Conservation*, Cambridge: Cambridge University Press.(Discussion of difficulties surrounding the concepts of ’community’ and ’stability’, as well as the problems of generalization in ecology.)

Economics and ethics

Unlike many other sciences, economics is linked both to ethics and to the theory of rationality. Although many economists regard economics as a 'positive' science of one sort of social phenomena, economics is built around a normative theory of rationality, and has a special relevance to policy making and the criticism of social institutions. Economics complements and intersects with moral philosophy in both the concepts it has constructed and in its treatment of normative problems.

Fundamental to modern economics is its conception of human beings as rational agents, whose choices are determined by complete and transitive preferences. Although economists stress the usefulness of this notion of rationality in explaining human behaviour, rationality is clearly also a normative notion. The mathematical tools economists have developed to represent and study the implications of rational action in collective and interactive contexts are thus of immediate relevance to moral philosophers.

Also of interest to moral philosophy is the problematic attempt in welfare economics to fashion a normative theory of economic institutions and policies around the goal of helping people satisfy their subjective preferences. This project relies, controversially, on equating people's wellbeing with the degree of satisfaction of their subjective preferences; an individual's 'utility' on this view is no more than an index of how well their subjective preferences are satisfied. Furthermore, since most welfare economists assume that there is no meaningful way to compare degrees of preference satisfaction across people, the project also requires a scheme for weighing the effectiveness of alternative economic arrangements in satisfying preferences without weighing the comparative satisfaction levels of different individuals. Central to the project is Pareto optimality - the notion of an 'efficient' arrangement as one in which no individual can achieve higher preference satisfaction without someone else undergoing a reduction in their satisfaction level.

Economic policies and institutions can be appraised in terms of a variety of values other than efficiency. Notable both in historical and contemporary discussions are the values of liberty, justice and equality. Since a large part of economics is carried out with a view to its possible application to policy, ethics has a significant part to play in economics. By the same token, economics may be of great importance to ethics, both through its exploration of consequences and through the development of mathematical and conceptual tools.

1 Rationality

At the foundation of both positive and normative economics lies a normative theory of individual rationality (see Rational choice theory). The theory is very thin, because it raises no questions about the rationality of ultimate ends and few questions about the rationality of beliefs. The standard view of rationality concerns only the internal completeness and consistency of an individual's preferences and the connection between preference and choice. In 'revealed-preference theory' preferences are derived from choices, and rationality is in effect taken to be a condition on choices (actions) only. In the standard theory of rationality (and implicitly in revealed-preference theory, too), preferences are subjective states that lead to actions.

An agent *A* chooses (acts) rationally if *A*'s preferences are rational, and *A* never prefers an available option to the option chosen. *A*'s preferences are rational only if they are complete and transitive. *A*'s preferences are transitive if and only if, for all options *x*, *y*, and *z*, if *A* prefers *x* to *y* and *y* to *z*, then *A* prefers *x* to *z*. The same stands where *A* is indifferent. Transitivity is a plausible, although not completely uncontroversial, requirement of rationality. *A*'s preferences are complete if, for all options *x* and *y*, either *A* prefers *x* to *y*, or *A* prefers *y* to *x*, or *A* is indifferent between *x* and *y*. If *A*'s preferences are complete, then *A* is never unable to rank *x* and *y*. It is questionable whether completeness is a condition of rationality.

If an agent's preferences are rational, one can assign numbers to the objects of preference. These numbers, which are arbitrary apart from their order, merely indicate preference ranking. They are 'ordinal utilities', and the theory of rationality may be re-stated: agents are rational if and only if their preferences may be represented by ordinal utility functions, and their choices maximize utility. (Representation by a continuous utility function presupposes an additional technical condition.) One should avoid saying that they act 'in order to maximize utility'. In contemporary economic theory, utility is merely an index or indicator. It is not a substantive aim or an object of

preference. One cannot sensibly pursue utility or take more or less utility in exchange for more or less of some genuine object of preference. Maximizing utility is just doing what one most prefers to do.

This standard theory of rationality is silent concerning what to do in circumstances of risk or uncertainty. Someone who prefers $10 to $5 but who prefers a gamble with a $5 prize to an identical gamble with a $10 prize violates none of the axioms of the standard theory (see Risk). To extend the standard theory requires (apart from additional technical conditions) one further substantive axiom, the so-called 'independence condition' or 'sure-thing principle'. This condition says that when an individual faces a choice between two bets that are identical except for one prize, then they will prefer the bet with the preferred prize. When one's preferences satisfy this additional independence condition, then they can be represented by an 'expected utility function'. Such functions have three important features. First, the expected utility of a lottery, $U(L)$, that pays off R with probability p and S with probability $(1-p)$ is $pU(R)+(1-p)U(S)$. Second, if U and U' are expected utility functions representing the preferences of an agent, then $U'=aU+b$, where a is a real number greater than zero and b is any real number. Expected utility functions are thus said to be 'unique up to a positive affine transformation'. Third, the requirement that subjective degrees of belief conform to the axioms of the probability calculus is implicit in the application of expected utility theory in conditions of uncertainty. Hence expected utility theory also contains a thin theory of rational belief. Expected utilities, like ordinal utilities, only provide information concerning preference and are not themselves goals or objects of preference. In expected utility theory, as in ordinal utility theory, choice is determined by preference.

The standard views of rationality and the theorems that permit their mathematical representation provide a framework in which to pose questions about the relations between morality and rationality. Their importance to moral philosophy depends in addition on the links between this view of rationality and a particular view of welfare.

2 Utility theory and welfare

All plausible moral views assign an important place to individual wellbeing. This is most obviously true of utilitarianism, which takes morality as maximizing some function of the welfare of individual members of society (see Utilitarianism §§1-3). Non-utilitarian views that emphasize notions of rights, fairness and justice also need a conception of human wellbeing. These views not only recognize the virtue of benevolence, which requires some notion of human good, but even their core notions make reference to wellbeing. Thus, for example, justice is understood in terms of treating the interests of different persons properly, and acting rightly often involves avoiding harm to other individuals. Notions of harm and of interest are plainly connected to notions of human wellbeing.

Given the commitments of economists to utility theory in explaining human choices, it is natural that they would look also to levels of utility or of preference satisfaction as the fundamental measure of human wellbeing for evaluative purposes. And indeed it is easy to reason as follows. Rational people have complete and transitive preference rankings and choose from among the available alternatives what they most prefer. Rational people also choose from the available options whatever they believe to be best. If one assumes in addition (as is common in positive economics) that individuals have correct beliefs and that they are self-interested, then it follows that (among the available options) what rational people most prefer is what rational people choose, which is the same thing as what is best for them. In other words, wellbeing is the satisfaction of preference.

Philosophers are less happy than economists with the idealization involved in assuming correct beliefs, and they are less convinced that people are always self-interested. Hence, they generally reject a view of welfare as the satisfaction of actual preferences. A view of welfare as the satisfaction of 'informed' and self-interested preferences is however common in contemporary philosophy and suffices to make the tools and explorations of economists relevant to philosophical concerns (see Welfare §1).

As both economists and philosophers have noticed, preference-satisfaction theories of wellbeing are problematic, even if one insists that the preferences be informed and self-directed. For preferences that are not based on indisputably false beliefs may nevertheless be based on highly contestable beliefs. Even if the preferences of people who beg for money to make a sacrifice to their god are well-informed, others may question whether satisfying them will make these people better off. If the notion of individual wellbeing is to be employed in

weighing the strength of claims for social provision, it must permit reasoned social agreement on basic components of wellbeing and on the relative 'urgency' of claims to different goods (Scanlon 1975). One might acknowledge a moral obligation to feed such people, but not to finance their worship.

There are other reasons, too, to question whether the conception of wellbeing as the satisfaction of preferences is suitable for assessing claims to scarce resources. For example, a person who has cultivated a taste for exotic foods may be miserable without them and, in one sense, worse off than someone in similar circumstances who delights in junk food. But one may question whether social policy should be responsive to such preferences. Similarly, it is questionable whether satisfying racist, sadistic and other antisocial preferences should count as contributing to individual wellbeing. Also of concern are 'adaptive preferences' and the effects of culture on preferences (Elster 1983; Sen 1987). For example, women who are systematically denied roles in public life, or equal shares of consumption goods, may learn not to want what they have not got. Their utility may be reasonably high, but objectively their level of wellbeing may be low.

Despite these difficulties, most economists prefer to cling to a preference-satisfaction view as a reasonable approximation. Philosophical proponents of these views (for example, Griffin 1986) have worked hard to deflect some of the objections and have defended the view that wellbeing is the satisfaction of 'informed' or 'rational' preferences. The commitment most economists have to a preference-satisfaction view derives from the links between a preference-satisfaction view of welfare and the explanatory uses of utility theory, and it is strengthened by the strong aversion many economists feel to anything that smacks of paternalism. Even though recognizing that individuals may be wrong about what is good for them does not imply that one should support paternalistic policies; it permits the question of paternalism to come up.

3 Alternative views of wellbeing

On the other hand, the difficulties with a subjective preference-satisfaction view of wellbeing have led many philosophers and a few economists to propose different conceptions of wellbeing: traditional hedonistic views and hybrids such as Sidgwick's (1901) view that certain mental states are intrinsically good yet are good because they are intrinsically desired, 'perfectionist' views in which the desires that count and their weight depend on the objective value of their objects (Griffin 1986; Parfit 1984; Raz 1986), Rawls' 'primary goods' (1971) and Amartya Sen's 'capabilities' (1987, 1992).

In Sen's view, a capability is the ability to achieve a certain sort of 'functioning' - literacy is a capability; reading is a functioning. People may value capabilities for their own sake as well as for the functionings they permit - you are glad to know you can walk around even if you are inclined to stay put. In Sen's view, preference satisfaction is too crude and too insensitive to the variety of values to serve as the basis for social evaluation. Rawls' primary-goods approach also fails to include a wide-enough range of relevant 'capabilities'. In focusing on primary social goods, Sen argues that it leaves out more internal features of persons' circumstances, such as physical handicaps, which may profoundly affect wellbeing. Rawls' account also may not work well for comparing the wellbeing of persons who are much above the minimum.

Sen's alternative is to define wellbeing in terms of the set of functionings a person achieves. Being well nourished is an example of a functioning whose achievement may be impaired either by internal obstacles (such as a digestive disorder) or external ones (a lack of money to buy food). Functionings are of course a vector (being well nourished, reading books, sleeping well) and combining them into a measure of wellbeing requires settling on the relative weight to be given to different capabilities and functionings. Obviously the problems of weighing different capabilities and functionings are serious. But even if they are insuperable and it turns out that focusing on capabilities and functionings sometimes leaves one unable to rank social alternatives, such an approach still leads one to focus on the morally relevant features of social alternatives. Particularly in the context of poor people and poor countries, measuring the extent of certain basic capabilities, like literacy or health-related capabilities, may be an effective way of gauging levels of wellbeing.

One important advantage of more objective approaches is that they avoid the problems of interpersonal comparison that derive from identifying wellbeing and utility (construed as an index of preference satisfaction). Formerly, economists such as Pigou (1920) argued that overall welfare would be maximized by equalizing incomes as much as was consistent with retaining incentives to produce. Citing diminishing marginal utility of

income, they maintained (not implausibly), that $100 contributes less to the wellbeing of someone with an income of $50,000 than to the wellbeing of someone with an income of $5,000. Other things being equal, a more equal distribution of income would increase total welfare.

This argument supposes that one can compare the contributions income makes to the wellbeing of different people. Once one takes seriously the preference-satisfaction view of wellbeing, these comparisons become problematic. Lionel Robbins (1932) argued compellingly that since preference rankings are entirely subjective and introspective, there is no way to compare where *A* is in *A*'s preference ranking with where *B* is in *B*'s ranking. The fact that people nevertheless make interpersonal comparisons of wellbeing depends, Robbins argues, on the value judgment that the preferences of people in similar circumstances are similarly satisfied. (One might instead question whether these comparisons of wellbeing are comparisons of preference satisfaction.)

Robbins' argument did not convince everyone. One approach has been to build up interpersonal welfare judgments from individual judgments of the form: 'I would prefer to be person *P* with consumption bundle *x* than *Q* with consumption bundle *y*' (see Kolm 1972 and Harsanyi 1977). If there is widespread agreement among individuals on such extended preferences, they can be used to construct interpersonally comparable welfare measures. Alternatively such 'extended preferences' may summarize what a perfected psychological theory would say about the relationship between the causal factors that lead people to have the preferences they have and the states of the world that satisfy preferences.

Although preference-satisfaction views of welfare do permit one to make use of the technical apparatus economists have developed, one should note the many problematic features of such views. The identification of wellbeing with the satisfaction of preferences is questionable in itself. It gives rise to serious problems of interpersonal comparisons. It is biased by adaptive preferences, and it fails to link up with the normative terms of policy debate.

4 Efficiency and Pareto optimality

Economists have given the common-sense notion of 'efficiency' a precise technical meaning: adopt the metric of utility (preference satisfaction) to measure a person's wellbeing. Consider some allocation of resources which yields a certain distribution of utility. A different allocation is a 'Pareto improvement' if and only if it makes at least one person better off (in terms of preference satisfaction) and makes no one worse off. An allocation is 'efficient' or 'Pareto optimal' if no other allocation is a Pareto improvement compared to it. In other words, in a Pareto optimal allocation, every alternative that satisfies someone's preferences better leads to someone else's preferences being less well satisfied. The great attraction of this notion for economists is that it permits a partial ranking of resource allocations or of social states which is responsive to individual preferences without employing any interpersonal comparisons of utility.

If one is minimally benevolent and favours making people better off, then, other things being equal, one should favour prospects that are better for someone and worse for no one. If in addition one identifies being better off with having a higher utility, then (other things being equal) one will endorse Pareto improvements and Pareto optimal allocations. Moreover, one can offer an argument in terms of these concepts for regarding a perfectly competitive equilibrium as a sort of moral ideal, other things being equal. What economists call 'the first welfare theorem' shows that competitive equilibria under certain idealized conditions (no externalities, no public goods, no informational limits, and so on) are Pareto efficient. Minimal benevolence then implies that competitive equilibria are (other things being equal) morally good economic states, and that market failures are bad. Furthermore, the 'second welfare theorem' shows that an efficient economic outcome producing any desired distribution of welfare can be attained by the workings of a competitive market, given the right initial distribution of endowments to agents. So with no ethical premise more controversial than minimal beneficence, albeit with a controversial conception of individual wellbeing, economists can offer moral evaluations along one significant dimension.

However, the Pareto principle may be unappealing when it involves honouring people's intrusive preferences regarding others' consumption patterns or levels of wellbeing. Moreover, it has been shown that the Pareto principle is inconsistent with the conjunction of individual and collective rationality in conditions of uncertainty (Seidenfeld *et al.* 1989). The problem lies in the fact that unanimity in individual rankings may rest on disagreements in subjective probability judgments and disagreements in preferences among options involving no uncertainty. Should unanimity in individual preference for *A* over *B* require that *A* be judged better, if the

unanimity depends on disagreements in values and probability judgments? Quite apart from these problems, the Pareto principle has very limited applicability, since economic changes usually involve winners and losers. Economists have been reluctant to accept these severe limitations and have struggled for many years over the issue of whether, without invoking interpersonal utility comparisons, some device could be found that would permit judgments in cases involving trade-offs among individuals. The hope was that one could find a way to endorse ‘potential Pareto improvements’, in which the gains to the winners were sufficient to compensate the losers in utility terms, without actually making the compensation (Hicks 1939; Kaldor 1939). The conclusion has been that no such blanket endorsement is possible, partly owing to technical failings in the proposed devices for identifying ‘improvements’, but more fundamentally because there simply is no way to judge changes that affect distributions while remaining neutral on distributive questions (Little 1957; Streeten 1953; Samuelson 1950). But to stop here would leave welfare economists with little to say, and practical applications of welfare economics in the form of cost-benefit analysis need to go on (see Pareto principle).

The difficulties seem to have been dealt with in two ways. One is by increasing modesty: one cannot always endorse potential Pareto improvements, but at least economists can identify them and leave the endorsing to politicians. Second, it seems that welfare economists have moved implicitly towards reuniting considerations of efficiency and equity. Blanket endorsements of potential Pareto improvements are unjustifiable, but since there are various theoretical ways to adjust such measures for equity concerns, one can suggest that equity-adjusted potential Pareto improvements can consistently be endorsed. (Stimulated partly by an interest in making inequality comparisons and equity judgments more precise, economists have devoted extensive energy to developing a variety of statistical techniques for measuring and describing inequality (Sen 1992).) Economists remain queasy about the final step of actually endorsing any one way of adjusting for equity concerns.

5 Economics and ‘moral mathematics’

There are currently many areas in which the tools and theories of economists have contributed to moral philosophy. The two areas of economics that have been of the most immediate relevance to moral philosophy have been social choice theory and game theory (see Social choice theory §3; Rational choice theory §§2-3). Economists and philosophers have used social choice theory and game theory mainly to investigate questions concerning the aggregation of individual preferences and judgments and concerning justice in social coordination and bargaining. But the tools of economics may apply to a much wider range of ethical issues. For example, there has been an energetic debate concerning conceptions of equality that has depended on formal concepts and arguments developed by economists (see Equality). When two individuals do not have exactly the same bundles of commodities, what could economic equality mean? One answer is ‘welfarist’: economic equality is equality of preference satisfaction. But economists seek to avoid making interpersonal utility comparisons, and it is questionable whether egalitarians should be concerned to equalize welfare. Another way to give sense to the notion of equality is to suppose that individuals acquired their different commodity bundles by means of voluntary exchanges beginning from a situation in which all had identical bundles (Dworkin 1981). In that case the allocation will be ‘envy-free’ in the sense that no one will prefer the commodity bundle of anyone else. It is then possible to investigate formally a number of questions: when will envy-free allocations exist? When will they be Pareto optimal? What sorts of events will upset or preserve the envy-freeness of a distribution? Is there any alternative formal characterization of equality? Here is an area in which inquiries of moral philosophers and economists are closely interwoven. There are others, too, ranging from John Roemer’s (1988) formal explorations into concepts of exploitation to recent work attempting to give a formal characterization of freedom.

6 Conclusion

Economists and moral philosophers share interests in rationality and in the evaluation and criticism of social institutions. Although the theoretical precommitments and starting points of scholars trained in these disciplines are often different, the two subjects have a great deal to offer one another. To the extent that moral philosophers want to bring their work to bear on the criticism and reform of social institutions, they have much to learn from economists, who have long been concerned with working out the consequences of alternative policies and with finding operational measures of theoretical concepts. Economists offer moral philosophers a range of powerful tools. At the same time, economists concerned with evaluating institutions and policies can hardly avoid a concern with ethics.

See also: Decision and game theory; Economics, philosophy of; Guanzi; Market, Ethics of the

DANIEL HAUSMAN
MICHAEL S. McPHERSON

References and further reading

Buchanan, A.E. (1985) *Ethics, Efficiency, and the Market*, Totowa, NJ: Rowman & Allanheld.(A discussion of ethics and economics by a contemporary moral philosopher.)

Dworkin, G., Bermant, G. and Brown, P.G. (eds) (1977) *Markets and Morals*, Washington, DC: Hemisphere, and New York: Halsted Press.(An anthology of essays on issues in ethics and economics.)

Dworkin, R. (1981) 'What is Equality? Part 2: Equality of Resources', *Philosophy and Public Affairs* 10 (4): 283-345.(A seminal discussion of equality of resources.)

Elster, J. (1983) *Sour Grapes: Studies in the Subversion of Rationality*, New York: Cambridge University Press. (A study of the notion of rationality and of failures of rationality.)

Elster, J. and Roemer, J. (eds) (1991) *Interpersonal Comparisons of Wellbeing*, Cambridge: Cambridge University Press.(An anthology of essays on the problems of making interpersonal utility comparisons.)

Gauthier, D. (1986) *Morals by Agreement*, Oxford: Oxford University Press.(An ambitious and influential attempt to derive a theory of justice from bargaining theory.)

Griffin, J. (1986) *Wellbeing: Its Meaning, Measurement and Moral Importance*, Oxford: Clarendon Press.(A major restatement of utilitarianism arguing for an informed-preference view of wellbeing.)

Harsanyi, J. (1977) *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge: Cambridge University Press.(A general development of rational choice theory and game theory with a systematic discussion of interpersonal utility comparisons.)

Hausman, D. and McPherson, M. (1993) 'Taking Ethics Seriously: Economics and Contemporary Moral Philosophy', *Journal of Economic Literature* 31 (2): 671-731.(A comprehensive survey of literature on ethics and economics.)

Hausman, D. and McPherson, M. (1996) *Economic Analysis and Moral Philosophy*, Cambridge: Cambridge University Press.(A nontechnical introduction to major issues in economics and ethics.)

Hicks, J. (1939) 'The Foundations of Welfare Economics', *Economic Journal* 49 (196): 696-712.(A classic statement of 'new welfare economics' based on the compensation criterion.)

Kaldor, N. (1939) 'Welfare Propositions of Economics and Interpersonal Comparisons of Utility', *Economic Journal* 49 (195): 549-52.(The first presentation and application of the notion of a potential Pareto improvement.)

Kolm, S.-C. (1972) *Justice et équité (Justice and Equality)*, Paris: Éditions du Centre National de la Recherche Scientifique.(A groundbreaking work, with the first systematic treatment of extended sympathy judgments.)

Little, I. (1957) *A Critique of Welfare Economics*, Oxford: Oxford University Press, 2nd edn.(A classic discussion of the difficulties of welfare economics without interpersonal comparisons.)

McClennen, E. (1990) *Rationality and Dynamic Choice: Foundational Explorations*, Cambridge: Cambridge University Press.(A general development of the theory of rational choice, including a critique of the standard theory and the development of an alternative theory.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Oxford University Press.(A major study of relations between rationality, personal identity and consequentialism.)

Pigou, A.C. (1920) *The Economics of Welfare*, London: Macmillan. (The classic statement of traditional utilitarian welfare economics.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(A development of a contractualist theory of justice and the most important work of social philosophy of the last century.)

Raz, J. (1986) *The Morality of Freedom*, Oxford: Oxford University Press.(A defence of liberalism and perfectionism.)

Robbins, L. (1932) *An Essay on the Nature and Significance of Economic Science*, London: Macmillan; 2nd edn, 1935.(A major methodology overview with a classical argument in chapter 6 against the possibility of interpersonal comparisons.)

Roemer, J. (1988) *Free to Lose*, Cambridge, MA: Harvard University Press.(An accessible introduction to Roemer's analytically sophisticated re-statement of themes in Marxian economics and social philosophy.)

Samuelson, P.A. (1950) 'Evaluation of Real National Income', *Oxford Economic Papers*, new series, 2 Jan (1): 1-29.(A technical demonstration of the impossibility of disentangling efficiency and distributional questions.)

Scanlon, T. (1975) 'Preference and Urgency', *Journal of Philosophy* 72 (22): 655-70.(An argument that questions

of justice cannot be decided only by considerations involving preference satisfaction.)
Seidenfeld, T., Kadane, J. and Schervish, M. (1989) 'On the Shared Preferences of Two Bayesian Decision Makers', *Journal of Philosophy* 86 (5): 225-44.(Proof of the inconsistency of individual and social Bayesian rationality in circumstances of uncertainty.)
Sen, A.K. (1982) *Choice, Welfare and Measurement*, Cambridge, MA: MIT Press.(This collection of essays contains many of Sen's principal earlier contributions to ethics and welfare economics.)
Sen, A.K. (1987) *On Ethics and Economics*, Oxford: Blackwell.(A concise and lucid but demanding introduction to the subject.)
Sen, A.K. (1992) *Inequality Reexamined*, Cambridge, MA: Harvard University Press.(An introduction to Sen's more recent work on inequality, capabilities and welfare.)
Sen, A. and Williams, B. (eds) (1982) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press.(An influential and useful anthology of essays on economics and ethics, with particular reference to utilitarianism.)
Sidgwick, H. (1901) *The Methods of Ethics*, London: Macmillan, 6th edn.(One of the most important works in ethics of the past two centuries, containing a classical restatement and defence of utilitarianism.)
Streeten, P. (1953) 'Appendix: Recent Controversies', in G. Myrdal *The Political Element in the Development of Economic Theory*, trans. P. Streeten, London: Routledge & Kegan Paul, 208-17.(A brief overview of the interdependence of questions of efficiency and distribution.)

Economics, philosophy of

People have thought about economics for as long as they have thought about how to manage their households, indeed Aristotle compared the study of the economic affairs of a city to the study of the management of a household. During the two millennia between Aristotle and Adam Smith, one finds reflections concerning economic problems mainly in the context of discussions of moral or policy questions. For example, scholastic philosophers commented on money and interest in inquiries concerning the justice of 'usury' (charging interest on money loans), and in the seventeenth century there was a great deal of discussion of policy concerning foreign trade. Economics only emerged as a distinct field of study with the bold eighteenth-century idea that there were 'economies' - that is autonomous law-governed systems of human interaction involving production, distribution and exchange. This view was already well developed in Adam Smith's Wealth of Nations, from which much of economics derives.

Economics is of philosophical interest in three main regards: it raises moral questions concerning welfare, justice and freedom; it raises foundational questions concerning the nature of rationality; and it raises methodological or epistemological questions concerning the character and possibility of knowledge of social phenomena. The fundamental theory of standard orthodox economics is of particular epistemological interest because of its resemblance to theories in the natural sciences coupled with its uneven empirical performance.

More than 150 years ago John Stuart Mill confronted the problem of how to reconcile his high regard for economics (despite its empirical adequacies) with his commitment to empiricism. His solution, which was accepted by most economists until the 1930s, held that the basic principles of economics are well established by introspection or everyday experience. One can thus justifiably have confidence in economics, despite the inexactness of its implications, which is only to be expected, since economics deals only with the most important determinants of economic phenomena.

In the 1930s Mill's views were rejected as too dogmatic and insufficiently empirical. But the views that succeeded Mill's during the generation after the Second World War, most notably the position of Milton Friedman and views deriving from the work of Karl Popper and Imre Lakatos, were less able to deal satisfactorily with the empirical inadequacies of economic theories than were Mill's. Since the mid 1980s there have been several new approaches, ranging from Donald McCloskey's rejection of methodological assessment at one extreme to Alexander Rosenberg's conclusion that economics cannot be a successful empirical science at the other.

1 What is economics?

Contemporary economics is divided into many schools and branches. The main 'orthodox', 'neoclassical', or 'neo-Walrasian' school models economic outcomes as equilibria in which individuals have done as well for themselves as they could given their preferences and the constraints on their choices. Common to all branches of orthodox economics are standard assumptions concerning technological possibilities and individual choice. Production is subject to diminishing returns to additional units of any given input, other inputs held constant, and is assumed not to show increasing returns to scale. The rational choices of individual agents are constrained by their initial endowment with resources and goods and by the technological possibilities. Agents are rational in the sense that their choices are determined by their preferences, which are complete and transitive. An agent's preferences are complete if they can rank all alternatives. An agent's preferences are transitive if for all alternatives *x*, *y* and *z*, they prefer *x* to *z*, whenever they prefer *x* to *y* and *y* to *z* (and similarly for indifference). Agents are typically assumed to want more consumption or leisure for their households and greater net returns for their firms. In competitive markets it is assumed that firms and individuals cannot influence prices, but economists and game theorists are also interested in strategic interactions, in which the rational choices of separate individuals are interdependent.

Orthodox economics has four main branches and many sub-specialities. The view of economic phenomena as equilibria resulting from constrained rational individual choice has a greater prominence in some branches and subspecialities than in others. It is central in the most theoretical branch, which consists of microeconomics, general equilibrium theory and game theory. Econometrics, in contrast, is as much statistics as economics. It uses data and statistical techniques to determine the values of parameters and to test specific models. Macroeconomics,

the third branch, is concerned with whole economies and particularly with the causes of the business cycle and economic growth. In John Maynard Keynes' *General Theory* (1936) its links with individual choice are tenuous. Contemporary economists in contrast have defended alternative macroeconomic theories in which the view of economic phenomena as equilibria resulting from constrained rational choices is central. The fourth branch, consisting of orthodox work in 'applied economics' (which is what most economists do), is consistent with the general theoretical picture of economic equilibrium, but applied theories make further simplifications to facilitate application. Within particular sub-specialities, such as international trade theory, labour economics or financial economics, one may find work lying along all of the four main branches.

There are also many other schools of economics. Austrian economists accept orthodox views of choices and constraints, but they emphasize uncertainty and question whether one should regard outcomes as equilibria. Institutionalist economists question the value of abstract general theorizing. They emphasize the importance of generalizations concerning norms and behaviour within particular institutions. Applied work in orthodox economics is sometimes very similar to applied institutionalist economics. Marxian economists traditionally articulated and developed Karl Marx's economic theories, but recently many Marxian economists have revised traditional Marxian concepts and themes with tools borrowed from orthodox economic theory. There are also socioeconomists, behavioural economists, chaos theorists, post-Keynesians, and neo-Ricardians. Within orthodox economics itself, there are also many different schools or approaches, such as agency theorists, the Chicago school, constitutional political economy, new institutional economics, and public choice theory. Economics is not one homogeneous enterprise. This entry will focus on neoclassical economic theory because the orthodox neoclassical school is the best known and most influential and because its central theory has attracted the most philosophical attention.

2 Why is economics of philosophical interest?

Economics has been of philosophical interest in three main regards. First, it raises moral questions concerning freedom, social welfare and justice. Although economists often deny that their theories have ethical content, they are ready with advice about how to make life better. Markets, which are the central institutions with which economics has traditionally been concerned, involve voluntary interactions, yet they are simultaneously mechanisms that regulate individual activities and allocate goods to people. They thus raise intricate moral questions concerning coercion, voluntary action, and social justice (see Economics and ethics). All of the leading figures in contemporary social and political philosophy comment on and are influenced by work in economics Hausman and McPherson (1996).

Second, contemporary theoretical economics is largely a theory of rational choice. This may seem surprising, since economics is supposed to be an explanatory and predictive science of the actual interactions among people rather than a normative discipline studying how people ought rationally to choose, but it is indeed a fact. This fact joins the interests of economists to the interests of those philosophers concerned with rational choice (see Rational choice theory; Social choice).

Third, economics raises important questions in philosophy of science. In part this is because all significant cognitive enterprises raise questions for epistemology or philosophy of science. But orthodox theory is of particular methodological interest for seven reasons.

Positive and normative. The extent to which economics appears to be permeated with normative concerns raises methodological questions about the relationships between a positive science (of 'what is') and a normative science (of 'what ought to be'). The standard view is that the positive science of economics, like engineering, helps policy-makers to choose the means to accomplish their ends, but it has no bearing on the choice of ends itself. This view is questionable, because economists have to interpret and articulate the incomplete specifications of goals and constraints provided by policy makers (see Value judgments in social science).

Reasons and causes. It is of philosophical interest that orthodox theoretical economics is as much a theory of rational choices as it is a theory that explains and predicts economic outcomes. Although economists are more interested in the aggregate results of individual choices than in the choices themselves, their theories offer both causal explanations for why individuals choose as they do and accounts of the reasons for their choices. Embedded within orthodox economics is a specific variant of 'folk psychology', and orthodox economics provides a specific

context in which to question whether folk-psychological explanations in terms of reasons can also be causal explanations (see Psychology, theories of; Explanation in history and social science).

Naturalism. Of all the social sciences, economics most closely resembles the natural sciences. Economic theories have been axiomatized, and essays and books of economics are full of theorems. Of all the social sciences, only economics boasts a Nobel Prize. Economics is thus a test case for those concerned with the extent of the similarities and differences between the natural and social sciences (see Naturalism in social science §2).

Abstraction and idealization. Economics raises questions concerning the legitimacy of severe abstraction and idealization. For example, economic models often stipulate that everyone is perfectly rational and has perfect information or that commodities are infinitely divisible. Such claims are exaggerations, and they are clearly false. Can good science make such false claims (see Scientific realism and social science)?

Ceteris paribus clauses. Because economists attempt to study economic phenomena as constituting a separate domain, influenced only by a small number of causal factors, the claims of economics are true only *ceteris paribus* - that is, they are true only if there are no interferences or disturbing causes. What are *ceteris paribus* clauses, and when if ever are they legitimate in science (see Social laws)?

Causation. Many important generalizations in economics make causal claims. For example, the law of demand asserts that a price increase will (*ceteris paribus*) diminish the quantity demanded. Yet economists are wary of causal language because of its suggestion that outcomes have single causes and because of difficulties in integrating talk of causation and talk of mutual determination. Econometricians have also been deeply concerned with the possibilities for determining causal relations from statistical evidence and with the relevance of causal relations to the possibility of consistent estimation of parameter values (see Causation §3).

Structure and strategy. During the past generation philosophers of science have been concerned to comprehend the larger theoretical structures that unify and guide research within particular research traditions or research programmes. Since orthodox economics is systematically unified, though not in quite the way that Kuhn (1970) or Lakatos (1970) discuss, it poses interesting puzzles about what guides research. Since the success of orthodox economics is controversial, this ‘research tradition’ also poses questions about how unified and constrained research ought to be.

These are the seven most significant philosophical issues concerning neoclassical economic theory, and many of these issues arise concerning all schools of economics.

3 Classic discussions of philosophical foundations

Explicit methodological reflection on economics dates back to the 1830s to the work of Nassau Senior (1836) and John Stuart Mill (1836). Their methodological reflections must be understood against the backdrop of the economic theory with which they were familiar. That theory (‘classical economics’) said comparatively little concerning the choices of consumers and supposed that people seek more wealth and are overly inclined to reproduce. With diminishing returns in agriculture and an increasing population, the rate of return in agriculture (and, given the mobility of capital, elsewhere) will diminish, and the ultimate result will be a stationary state in which profits are low, workers receive subsistence wages, and landlords receive large rents. On account of this view of the future, economics was called ‘the dismal science’. Classical economics, like contemporary theory, relied on bold simplifications and it had empirical difficulties. Although populations grew considerably, the rate of return in the nineteenth century did not fall sharply, and wages increased dramatically.

Mill was firmly committed to the economics of his day, yet he was a strict empiricist (see Mill, J.S.; Empiricism §5). Since economics faced such major empirical difficulties, it might appear that Mill would have to change his epistemology or disavow his economics. Call this conflict between empiricism and economics, which arises from the apparent disconfirmations of economics and the difficulty of testing it, ‘Mill’s problem’. Mill attempted to solve it by maintaining that the basic premises of economics are empirically well established by introspective psychology or by experimental testing of technical claims such as the law of diminishing returns. These well-supported premises state how specific causal factors operate. If the only causal factors influencing economic phenomena were those specified in these premises, then the predictions of economic theory would be correct. But economic phenomena depend on many causal factors that are left out of economic theories. Consequently, the

implications are inexact. They are always imprecise, and when the factors left out are of particular importance, the predictions of the theories may be completely mistaken. This inexactness explains why the implications of economic theories are so poorly confirmed, and consequently the problems do not show that there is anything mistaken in the fundamental generalizations of economics. In Mill's view, the empirical confirmation of economic theories is indirect and 'deductive'. It derives from the confirmation of their premises. The inductive method of 'specific experience' cannot be employed because of the multiplicity of causes. Furthermore, since there is no way to incorporate a much larger number of causal factors without destroying the 'separateness' of economics and subsuming it into a general social science, this inexactness is an inevitable feature of economics as a distinct discipline. Economics is unavoidably a science of 'tendencies' only.

Mill's view was tremendously influential until the 1930s, although there were always critics who argued for a more directly empirical approach to assessing economic theories. The most important methodological works of the late nineteenth century, J.E. Cairnes' *The Character and Logical Method of Political Economy* (1875) and John Neville Keynes' *The Scope and Method of Political Economy* (1891) defend Mill's view, and one can still find it clearly expressed in Lionel Robbins' classic *An Essay on the Nature and Significance of Economic Science* (1932). Some have also argued that Mill is, in fact, essentially right Hausman (1992, chaps 8, 12). But by the 1940s most of those writing on economic methodology had repudiated Mill's view of economics as a separate and inexact science justified by a deductive method.

Methodological challenges to Mill's view before the 1930s came mainly from those working outside the orthodox mainstream. The most important challenge came from the German Historical School in the nineteenth century and from the American Institutionalists in the twentieth century. These critics argued that theories should apply more directly to specific historical circumstances and should be tested more directly through these applications. Members of the German Historical School also questioned whether it is possible to separate a positive economic science from normative issues of policy making. Mill accepts the positive/normative division, though not in quite its modern form, and it is explicit in later followers such as J.N. Keynes. Members of the German Historical School wanted economic theorizing to be more explicitly normative and oriented toward specific historical circumstances and policy recommendations.

Although these objections did not shake the views on theory assessment accepted by mainstream economists, views which remained stable for the century after Mill, changes in philosophical conceptions did accompany the significant changes that occurred during that century in economic theory itself. Neoclassical economics, unlike its classical predecessor, focuses on individual choices, which unavoidably reflect subjective preferences and beliefs. This fact disturbed critics of neoclassical economics, such as Thorstein Veblen, who saw it as a departure from the general trend of science away from teleological concerns and towards objective causal relations (1898). Defenders such as Frank Knight, Ludwig von Mises and Lionel Robbins agreed that the role of subjective factors sets economics apart from the natural sciences. But they denied that the methodological distinctiveness of economics demonstrates any methodological error. Indeed von Mises and his followers saw this peculiarity of economics as a virtue that provides a special certainty to the conclusions of economics (1949). In their view (though not in the view of other so-called 'Austrian' economists such as Hayek), the conclusions of economics are beyond rational dispute, because they result from the privileged access people have to their preferences and beliefs or because they derive from the meaning of rationality (see Certainty).

The neoclassical focus on individual choice also led to a redefinition of the discipline. Mill and the classical economists had seen economics as concerned with specific causal factors - essentially acquisitive motives, diminishing returns, and the propensity to reproduce - which preponderate in a specific domain of social life (1836: 323). Robbins, in contrast, defined economics as 'the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses' (1932: 15). In Robbins' view, economics is not concerned with production, consumption, distribution or exchange as such. It is instead concerned with an aspect of all human action. The causal factors with which economics is concerned - rational self-interest - have a role in all human affairs. This vision of economics has been vindicated to some extent by applications of economic models to matters such as decisions to marry Becker (1981) and by the development of game theory (see Decision and game theory). To regard the subject matter of economics as an aspect of all human action is nevertheless an exaggeration, since economists continue to focus on a particular domain of social phenomena and on causal factors - such as diminishing returns - that are of particular importance only in that domain.

The methodological reflections of Knight, Robbins, and the Austrians reinforced Mill's views of economics as an inexact science, whose conclusions are credible because they follow deductively from well-established premises. Yet this view of theory assessment in economics, which was virtually unchallenged within mainstream economics in 1930, was almost without defenders by the end of the 1950s. During this short period economists came to believe that Mill's view of confirmation was insufficiently empirical. Economics could be scientific only if its theories were subjected to empirical testing, and its theories could be rationally acceptable only if they passed these tests. It looks as if economists became familiar with the work of the logical positivists (see Logical positivism) and applied their views to economics. But there is little solid evidence that the methodological changes were the result of the influence of logical positivism. The general changes in intellectual climate that gave rise to logical positivism may have separately caused this revolution in economic methodology. Books by Lesley Fraser (1937), Terence Hutchison (1938) and Felix Kaufmann (1944) made the case for such a methodological reorientation. Hutchison's was the most influential of these books (probably because of Frank Knight's lengthy attack on Hutchison's views in the *Journal of Political Economy* (1940). Hutchison had studied in Germany and was familiar with the work of the logical positivists and of Karl Popper, and he complained that statements qualified with vague *ceteris paribus* clauses cannot be refuted.

Paul Samuelson's work on the theory of revealed preference (1938) was especially influential. Samuelson showed that if consumer choices in simplified models satisfy a consistency condition, then it is possible in principle to construct from the choices a preference ordering over commodity bundles. His work thus suggested that it might be possible to reformulate economic theory without relying on hard-to-test claims about subjective states of choosers. In teasing out the 'operationally significant' content of economics from its mentalistic formulation, one might be able to make economics conform to demanding empiricist strictures. Samuelson was himself more influenced by the work of scientists such as Gibbs and Bridgman than by the logical positivists, and his own view of methodology (1963), which essentially repudiates all generalizations that are not themselves separately and directly testable, is more extreme than are the views of the logical positivists.

Whatever the cause, Mill's views were eclipsed. Methodologists no longer defended them, although it is arguable that economists continued to rely on them in practice. But Mill's problem remained. Neoclassical economic theory, like its classical predecessor, is hard to test and often apparently disconfirmed. Unless one is prepared to reject the theory or to deny that theories ought to fit the data, one needs to provide some defence of the empirical credentials of economics.

4 Milton Friedman's views

Since the mid twentieth century a large literature devoted to economic methodology has emerged. That literature explores many methodological approaches and applies its conclusions to many schools and branches of economics. Most of the literature focuses on the fundamental theory of mainstream economics - the theory of the equilibria resulting from constrained rational individual choice. This section focuses on Milton Friedman's extremely influential views. Two particularly significant alternatives to Friedman not discussed here can be found in Machlup (1978) and Samuelson (1963).

In his 1953 essay, 'The Methodology of Positive Economics', Friedman argued that the only relevant test of an economic theory is its success in predicting the phenomena that economists are concerned with. He believes that standard microeconomic theories and the quantity theory of money pass such tests well. He responds to criticisms of theories that point out that they contain 'unrealistic assumptions' (such as the assumption that firms are profit maximizers) by arguing that the criticisms presuppose mistakenly that theories can be tested by their assumptions. In Friedman's view, Mill's problem evaporates as soon as one recognizes that scrutinizing the realism of assumptions is a methodological mistake.

Friedman begins his essay by distinguishing sharply between positive and normative economics (see Economics and Ethics §5). He dismisses the possibility of rational argument concerning values, but maintains that disagreements about economic policy result more from disagreements concerning the consequences of alternative policies than from disputes about values. The goal of positive economics is, in Friedman's view, exclusively predictive: 'The ultimate goal of a positive science is the development of a "theory" or ‘‘hypothesis" that yields valid and meaningful (i.e., not truistic) predictions about phenomena not yet observed' (1953: 7). Friedman thus defends an instrumentalist view of science. Note that 'predictions' here are simply implications of a theory whose

truth is not yet known.

Friedman’s central claim that the realism of ‘assumptions’ is irrelevant to the assessment of a theory does not follow from a standard instrumentalist perspective, in which the correctness or incorrectness of all the observable consequences of a theory bear on its acceptability. For the assumptions of economics (for example that firms maximize profits) are testable, and a standard instrumentalist would not dismiss apparent disconfirmations as irrelevant. To make sense of Friedman’s position, one must recognize that he rejects a standard instrumentalist concern with all the predictions of a theory. Friedman denies that the goal of economics is predictive success in general. The practical point of an economic theory is correct prediction only for ‘the phenomena it purports to explain’ (1953: 30). A good tool need not be an all-purpose tool. The fact that survey results are inconsistent with predictions derived from the hypothesis that firms are profit maximizers is irrelevant, because survey results are not among the phenomena that the theory of the firm is intended to explain.

From his thesis that the goal of economics is such ‘narrow’ predictive success, Friedman jumps to the conclusion that narrow predictive success is the only relevant test. Although tempting, this inference is a mistake. It is like arguing that the only relevant way to check a computer program is to run it and see whether it does what it is supposed to. If it is possible to tell for sure by running a program whether it will always do what it is supposed to, then there is indeed no point to studying the code (though studying the code might be cheaper and easier than investigating what happens when one runs the program). But with economic theories, as with complicated computer programs, one can only look at a small sample of their performance, and success in the sample is no guarantee of success in general. One cannot look and see how well the theory performs with respect to the full range of phenomena it was designed to account for. Indeed, the point of a theory is precisely to provide guidance when one does not yet know how the theory’s predictions come out. Just as one can assess computer programs by studying their code or examining how they work in uninteresting applications, so one can assess theories by examining their assumptions and attending to the success or failure of uninteresting predictions. Such scrutiny of the ‘realism of assumptions’ is of particular importance when extending a theory to new circumstances or when modifying a theory in the face of predictive failure.

Friedman’s narrow view of the goals of economics is also contestable. Even though economics has a crucial policy mission, many economists are concerned to explain economic phenomena, and many simply want to know the truth about various aspects of economies. Those economists obviously cannot accept Friedman’s view that the realism of assumptions does not matter. The argument above shows that economists who accept Friedman’s narrow view of the goals of economics cannot dismiss criticisms of the realism of assumptions either. Friedman does not solve Mill’s problem.

5 Popperian perspectives

A second modern approach to Mill’s problem derives from Karl Popper’s philosophy of science and is more critical of economics (see Popper, K. §§2-3). Popper defends a ‘falsificationist’ methodology. Scientists should formulate theories that are ‘logically falsifiable’ - that is, inconsistent with some possible observation reports. ‘All crows are black’ is logically falsifiable, since it is inconsistent with (and would be falsified by) an observation report of a red crow. Second, Popper maintains that scientists should subject theories to harsh tests and should be willing to reject them when they fail the tests. Third, scientists should regard theories as, at best, interesting conjectures. Passing a test does not confirm a theory or provide one with reason to believe it. All it does is to justify continuing to employ it (since it has not yet been falsified) and devoting increased efforts to attempting to falsify it (since it has thus far survived testing). Popper has also written in defence of what he calls ‘situational logic’ (which is basically rational choice theory) as the correct method for the social sciences. There appear to be serious tensions between Popper’s falsificationism and his defence of situational logic, and his discussion of situational logic has not been as influential as his falsificationism. For discussion of this aspect of Popperian economic methodology, see Hands (1985).

Given Popper’s philosophy of science, there seems little hope of solving Mill’s problem without rejecting contemporary economic theory and condemning its practitioners as behaving in scientifically illegitimate ways. For specific economic theories are rarely logically falsifiable. When they are, they are rarely subjected to tests, let alone harsh tests. When they fail the tests, they are rarely repudiated. Worst of all, economic theories, which have not even been well tested, are taken to be well-established, perhaps even unquestionable guides to policy, rather

than merely conjectures. Some critics of neoclassical economics have made these criticisms Eichner (1983). But most of those who have espoused Popper's philosophy of science have not repudiated mainstream economics and have not been harshly critical of its practitioners.

Mark Blaug (1992) and Terence Hutchison (1938), who are the most prominent Popperian methodologists, criticize particular features of economics, and they both call for more testing and a more critical attitude. But both understate the radicalism of Popper's views and take his message to be merely that scientists should be critical and concerned to test their theories. Blaug's and Hutchison's criticisms have sometimes been challenged on the grounds that economic theories cannot be tested, because of their *ceteris paribus* clauses and the many subsidiary assumptions required to derive testable implications Caldwell (1984). But this response ignores Popper's insistence that testing requires methodological decisions not to attribute failures of predictions to mistakes in subsidiary assumptions or to 'interferences'. For views of Popper's philosophy and its applicability to economics, see de Marchi (1988) and Caldwell (1991).

Applying Popper's views on falsification literally would be destructive. Not only neoclassical economics, but all known and probably all imaginable economic theories would be condemned as unscientific, and there would be no way to discriminate among economic theories. One major problem is that one cannot derive testable implications from theories by themselves. To derive testable implications, one also needs subsidiary assumptions or hypotheses concerning distributions, measurement devices, proxies for unmeasurable variables, the absence of various interferences, and so forth (see Duhem, P.; Quine, W.V. §1). These problems arise generally and Popper proposes that they be solved by a methodological decision to regard a failure of the deduced testable implication to be a failure of the theory. But in economics the subsidiary assumptions are extremely uncertain and in many cases known to be false. Making the methodological decision that Popper requires is unreasonable and would lead one to reject all economic theories.

Imre Lakatos (1970), who was for most of his philosophical career a follower of Popper, offers a broadly Popperian solution to this problem (see Lakatos, I.). Lakatos denies that one tests individual theories. When theories face empirical difficulties, as they always do, one attempts to modify them. Scientifically acceptable (in Lakatos' terminology 'theoretically progressive') modifications always have some additional testable implications and are thus not purely *ad hoc*. If some of the new predictions are confirmed, then the modification is 'empirically progressive' and one has reason to reject the unmodified theory and to employ the new theory, regardless of how unsuccessful, in general, either theory may be. Though progress may be hard to come by, Lakatos' views do not have the same destructive implications as Popper's. Lakatos appears to solve Mill's problem by arguing that what matters is empirical progress or retrogression rather than empirical success or failure. It is thus easy to see why Lakatos' views have become more widespread among economic methodologists than Popper's. Developing Kuhn's notion of a 'paradigm' (1970) and some hints from Popper, Lakatos also developed a view of the global theory structure of whole theoretical enterprises, which he called 'scientific research programmes'. Lakatos emphasized that there is a 'hard core' of basic theoretical propositions that are not to be questioned and that substantial heuristics guide members of a research programme in the articulation and modification of specific theories. These views were also attractive to economic methodologists, since theory development in economics is so sharply constrained and directed, and since economics appears at first glance to have a hard core. The fact that economists do not give up basic theoretical postulates that appear to be false might be explained and justified by regarding them as part of the hard core of the neoclassical research programme.

Yet Lakatos' views do not provide a satisfactory solution to Mill's problem. For it is questionable whether the development of neoclassical economic theory has demonstrated empirical progress. For example, the replacement of 'cardinal' utility theory by 'ordinal' utility theory in the 1930s, which is generally regarded as a major step forward, involved the replacement of one theory by another that was strictly weaker and which had no additional empirical content. Furthermore, despite his emphasis on heuristics as guiding theory modification, Lakatos still emphasizes testing. Science is for Lakatos a much more empirically driven enterprise than is contemporary economics Hands (1991). It is also doubtful whether research enterprises in economics have hard cores (Hoover 1991; Hausman 1992, ch. 6). For attempts to apply Lakatos' views to economics, see Latsis (1976) and Weintraub (1985). As is apparent in de Marchi and Blaug (1991), writers on economic methodology have in recent years become increasingly disenchanted with Lakatos' philosophy.

There is a second major problem with Popper's philosophy of science, which plagues Lakatos' views as well. Both defend the startling thesis that there is no such thing as empirical confirmation. Popper and Lakatos deny that the results of testing can give one reason to believe that statements are true, and both deny that results of tests can justify relying on statements in practical endeavours or in theoretical inquiry. They would accordingly regard as mistaken any claim that there is better evidence for one unfalsified proposition than for another. Someone who questions whether there is enough evidence for some proposition to justify relying on it in theoretical studies or for policy purposes would be making the methodological 'error' of supposing that there can be evidence in support of hypotheses. With the notable exception of Watkins (1984), few philosophers within the Popperian tradition have faced up to this radical consequence.

6 Contemporary directions in economic methodology

The failure of Friedman's, Popper's, and Lakatos' views to solve Mill's problem has had little impact on most economists, who are little concerned with methodology and who, when pressed, typically defend a position like Friedman's. But those specifically concerned with methodology have turned in many new directions, of which only three will be discussed here. Especially notable among the projects not discussed here are Lawrence Boland's neo-Popperian views (1989), Bruce Caldwell's 'methodological pluralism' (1982), Uskali Mäki's realist philosophy of economics (1990), Phillip Mirowski's exploration of the metaphorical structure of economics (1990), and E. Roy Weintraub's social constructivism (1991). There has also been a substantial effort to apply structuralist views of scientific theories to economics Stegmüller *et al.* (1982; Balzer and Hamminga (1989), and recently there have been some discussions of economics from feminist perspectives Ferber and Nelson (1993).

One radical reaction to the failure to solve Mill's problem is to deny that it can be solved. In Alexander Rosenberg's view (1992), economics can only make imprecise 'generic' predictions, and it cannot make progress, because it is built around folk psychology, which is a mediocre theory of human behaviour which (owing to the irreducibility of intentional notions) cannot be improved. Complex economic theories ought to be valued as applied mathematics, not as empirical theory. Since economics, despite its many well-trained practitioners, does not show the same consistent progress as the natural sciences, one cannot dismiss out of hand Rosenberg's suggestion that economics is an empirical dead end. But his view that it has made no progress and that it does not permit quantitative predictions is hard to accept. For example, it seems that contemporary economists can do a much better job of predicting the revenue consequences of a change in tax rates than could economists of even a century ago.

An equally radical but opposite reaction is Donald McCloskey's. He would solve Mill's problem by repudiating methodology. In McCloskey's view, the only relevant and significant criteria for assessing the practices and products of a discipline are those accepted by the practitioners. Apart from a few general standards such as honesty and a willingness to listen to criticisms, the only justifiable criteria for any conversation are those of the participants. The pretensions of philosophers to judge the discourse of scientists are arrogant and may be dismissed. Mill's problem dissolves when economists recognize that philosophical standards of empirical success may be safely ignored. Those who are interested in understanding the character of economics and in contributing to its improvement should eschew methodology and study instead the 'rhetoric' of economics - that is, the means of argument and persuasion that succeed among economists.

McCloskey's studies of the rhetoric of economics have been valuable and influential (1985, esp. ch. 5-7), but much of his work consists not of such studies but of philosophical critiques of the pretensions of methodology. These are more problematic, because the position sketched in the previous paragraph is hard to defend and potentially self-defeating. It is hard to defend, because epistemological standards for good science have already infected the conversation of economists. The standards of predictive success which give rise to Mill's problem are standards that economists already accept. Mill's problem can be dissolved only if economists can be persuaded to surrender the standards that gave rise to it. But the position sketched in the previous paragraph makes it difficult to argue for any changes in standards. Furthermore, as Alexander Rosenberg has argued, it seems that economists would doom themselves to irrelevance if they were to surrender standards of predictive success, for it is upon such standards that policy decisions are made. McCloskey does not, in fact, want to preclude all 'external' criticisms, which maintain that economists are sometimes persuaded when they should not be or are not persuaded when they should be. For he criticizes the conflations of statistical significance with economic importance, which are, as he

documents, nevertheless extremely common (1985, ch. 9). Sometimes McCloskey characterizes rhetoric descriptively as the study of what in fact persuades, but sometimes he characterizes it normatively as the study of what ought to persuade (1985, ch. 2). And if rhetoric is the study of what ought to persuade, then it is methodology, not an alternative to methodology. Mill's problem cannot be conjured away.

A third approach is to return to Mill's own solution. Many of the basic principles of economics are plausible and are borne out in everyday experience. Although such plausibility does not place these principles beyond question, it does provide some warrant for them and some warrant for what may be deduced from them. Given the weakness of tests involving market data, in which there is an uncontrolled multiplicity of causal factors, it may be reasonable to hang on to orthodox theory in the face of disconfirmation. This thought can be rigorously supported in terms of Bayesian confirmation theory Hausman (1992, ch. 12).

Yet this restoration of something like Mill's solution is in many ways as close to Rosenberg's and McCloskey's views as it is to Mill's. For the recognition that market tests will generally be too weak to shake the initial plausibility of the basic principles is simultaneously a recognition that there is little to be learned from market data. Although not quite as hostile to the empirical claims of economics as is Rosenberg's position, this view holds out few hopes for major improvements in economics. (By seeking higher quality data from observational studies and particularly experiments (see Roth 1988), economists may however be able to find a way out of this impasse.) On the other hand, in its insistence that one study the specific problems and constraints that economists face, this view joins McCloskey's in calling for less philosophically blinded studies of the rhetoric of economics (although it does not, of course, accept McCloskey's condemnation of epistemology).

7 Conclusions

Contemporary economic methodology is in turmoil. It has become an active area of research engaging major efforts by dozens of economists and a small group of philosophers. Economics presents a tantalizing list of puzzles, with Mill's problem at the core. And these problems have a special philosophical interest because of the connections between economics and ethics and between economics and the theory of rationality. There is much that is unknown and puzzling about how much people can learn about the character of their interactions and about what methods are most likely to help them to learn more. Economics is one exciting venue in which these general problems can be raised, but it has not, as yet, provided any easy solutions.

See also: Guanzi; Social science, history of philosophy of

DANIEL HAUSMAN

References and further reading

Backhouse, R. (ed.) (1994) *New Perspectives on Economic Methodology*, London: Routledge.(An accessible collection of recent essays.)

Balzer, W. and Hamminga, B. (eds) (1989) *Philosophy of Economics*, Dordrecht: Kluwer.(Contains mainly applications to economics of the structuralistic view of theories; cited in §6.)

Becker, G. (1981) *A Treatise on the Family*, Cambridge, MA: Harvard University Press.(Applications of economic theory to the family; cited in §3.)

Blaug, M. (1992) *The Methodology of Economics: Or How Economists Explain*, Cambridge: Cambridge University Press, 2nd edn.(An influential discussion from a Popperian perspective; cited in §5.)

Boland, L. (1989) *The Methodology of Economic Model Building: Methodology after Samuelson*, London: Routledge.(One of a series of books developing a distinctive neo-Popperian view of economic methodology; cited in §6.)

Buchanan, J. and Vanberg, V. (1991) 'The Market as a Creative Process', *Economics and Philosophy* 7: 167-86. (A classic statement of an 'Austrian' perspective.)

Cairnes, J.E. (1875) *The Character and Logical Method of Political Economy*, New York: A.M. Kelley, 2nd edn, 1965.(One of the major nineteenth-century works on economic methodology by a follower of John Stuart Mill; cited in §3.)

Caldwell, B. (1982) *Beyond Positivism: Economic Methodology in the Twentieth Century*, London: Allen & Unwin.(An extremely influential introduction to economic methodology; cited in §6.)

(ed.) (1984) *Appraisal and Criticism in Economics*, London: Allen & Unwin.(A useful collection of roughly positivist or Popperian essays on economic methodology; cited in §5.)

(1991) ‘Clarifying Popper’, *Journal of Economic Literature* 29: 1-33.(A major summary of Popper’s views as applicable to economics; cited in §5.)
(ed.) (1993) *The Philosophy and Methodology of Economics*, Cheltenham: Edward Elgar.(A three-volume collection of essays on economic methodology.)
Dugger, W. (1979) ‘Methodological Differences between Institutional and Neoclassical Economics’, *Journal of Economic Issues* 13: 899-909.(A further discussion of the character of institutionalist economics; mentioned in §3.)
Eichner, A. (1983) ‘Why Economics is not yet a Science’, in A. Eichner (ed.) *Why Economics Is not yet a Science*, Armonk, NY: M.E. Sharpe, 205-41.(An example of a Popperian critique of orthodox economics; cited in §5.)
Ferber, M. and Nelson, J. (eds) (1993) *Beyond Economic Man: Feminist Theory and Economics*, Chicago, IL: University of Chicago Press.(A collection of essays examining economics from a feminist perspective; cited in §6.)
Fraser, L. (1937) *Economic Thought and Language. A Critique of Some Fundamental Concepts*, London: A. & C. Black.(Calls for a more positivistic view of economics; cited in §3.)
Friedman, M. (1953) ‘The Methodology of Positive Economics’, in *Essays in Positive Economics*, Chicago, IL: University of Chicago Press, 3-43.(The most influential single essay on economic methodology; discussed in §4.)
Hands, D. (1985) ‘Karl Popper and Economic Methodology’, *Economics and Philosophy* 1: 83-100.(A discussion of the relevance of Popper’s situational logic to economic methodology; cited in §5.)
Hands, D. (1991) ‘The Problem of Excess Content: Economics, Novelty and a Long Popperian Tale’, in N. de Marchi and M. Blaug (eds), *Appraising Modern Economics: Studies in the Methodology of Scientific Research Programs*, Cheltenham: Edward Elgar, 38-75.(A discussion of empirical emphasis in Lakatos’ philosophy of science; cited in §5.)
Hausman, D. (1992) *The Inexact and Separate Science of Economics*, Cambridge: Cambridge University Press.(A general account of the philosophy of economics that defends a reformulation of J.S. Mill’s views; cited in §§5, 6.)
Hausman, D. (ed.) (1984) *The Philosophy of Economics: An Anthology*, Cambridge: Cambridge University Press. (A collection of both classic and recent essays in the philosophy of economics.)
Hausman, D. and McPherson, M. (1996) *Economic Analysis and Moral Philosophy* Cambridge: Cambridge University Press.(An introduction to work at the boundaries of economics and ethics; cited in §2.)
Hoover, K. (1991) ‘Scientific Research Program or Tribe? A joint Appraisal of Lakatos and the New Classical Macroeconomics’, in N. de Marchi and M. Blaug (eds) *Appraising Modern Economics: Studies in the Methodology of Scientific Research Programs*, Cheltenham: Edward Elgar, 364-94.(A discussion of the structure of new classical economics with a critique of Lakatos’ philosophy of science; cited in §5.)
Hutchison, T. (1938) *The Significance and Basic Postulates of Economic Theory*, New York: A.M. Kelley, repr. with a new preface, 1960.(Probably the most important methodological work calling for a more positivistic or Popperian view; cited in §3.)
Kaufmann, F. (1944) *Methodology of the Social Sciences*, London: Oxford University Press.(Call for a more positivistic approach; cited in §3.)
Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*, London: Macmillan.(The work that announced the Keynesian revolution in economics; cited in §1.)
Keynes, J.N. (1891) *The Scope and Method of Political Economy*, New York: A.M. Kelley, 4th edn, 1917.(A classic work on economic methodology heavily influenced by J.S. Mill; cited in §3.)
Klamer, A., McCloskey, D. and Solow, R. (eds) (1988) *The Consequences of Economic Rhetoric*, New York: Cambridge University Press.(Essays containing further discussion of McCloskey’s views on the rhetoric of economics; discussed in §6).
Knight, F. (1940) ‘What is "Truth" in Economics?’, *Journal of Political Economy* 48: 1-32.(An influential review of Hutchison (1938) that defends an old-fashioned Millian view of economic methodology; cited in §3.)
Koopmans, T. (1957) *Three Essays on the State of Economic Science*, New York: McGraw-Hill.(The most important book on economic methodology of the 1950s; defends a moderate empiricist position.)
Kuhn, T. (1970) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press, 2nd edn.(A recent classic in philosophy of science defending the importance of ‘paradigms’; cited in §2.)
Lakatos, I. (1970) ‘Falsification and the Methodology of Scientific Research Programmes’, in I. Lakatos and A. Musgrave (eds), *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press, 91-196. (The major statement of Lakatos’ philosophy of science; cited in §§2, 5.)

Latsis, S. (ed.) (1976) *Method and Appraisal in Economics*, Cambridge: Cambridge University Press.(A collection of essays applying Lakatos' philosophy to economics; cited in §5.)

Leijonhufvud, A. (1973) 'Life Among the Econ', *Western Economic Journal* 11: 327-37.(A humorous commentary on economic methodology.)

Leontief, W. (1971) 'Theoretical Assumptions and Nonobserved Facts', *American Economic Review* 61: 1-7.(An influential critique of the abstractness and empirical weakness of contemporary economics.)

Little, D. (ed.) (1993) *On the Reliability of Economic Models: Essays in the Philosophy of Economics*, Boston, MA: Kluwer.(A recent collection of essays on economic methodology.)

McCloskey, D. (1985) *The Rhetoric of Economics*, Madison, WI: University of Wisconsin Press.(This contains the major statement of McCloskey's views; cited in §6.)

Machlup, F. (1978) *Methodology of Economics and Other Social Sciences*, New York: Academic Press.(A collection of Machlup's essays on economic methodology; cited in §4.)

Mäki, U. (1990) *Studies in Realism and Explanation in Economics*, Helsinki: Suomalainen Tiedeakatemia.(Essays clarifying notions of realism and defending its importance in understanding economics; cited in §6.)

Marchi, N. de (ed.) (1988) *The Popperian Legacy in Economics*, Cambridge: Cambridge University Press. (Contains essays and debate concerning Popperian economic methodology; cited in §5.)

(ed.) (1992) *Post-Popperian Methodology of Economics: Recovering Practice*, Boston, MA: Kluwer.(A collection with examples of new directions in economic methodology.)

Marchi, N. de and Blaug, M. (eds) (1991) *Appraising Modern Economics: Studies in the Methodology of Scientific Research Programs*, Cheltenham: Edward Elgar.(A collection devoted to essays appraising Lakatosian economic methodology; cited in §5.)

Marx, K. (1857-8) *Grundrisse*, trans. M. Nicolaus, New York: Random House, 1973.(The introduction contains Marx's most extensive discussion of economic methodology.)

Menger, C. (1883) *Problems of Economics and Sociology*, ed. L. Schneider, trans. F. Nock, Urbana, IL: University of Illinois Press, 1963.(A critique of the views of the German Historical School discussed in §3.)

Mill, J.S. (1836) 'On the Definition of Political Economy and the Method of Investigation Proper to it', in *Collected Works of John Stuart Mill*, vol. 4, Toronto, Ont.: University of Toronto Press, repr. 1967.(The first presentation of Mill's influential views on economic methodology; cited in §3.)

Mill, J.S. (1843) *A System of Logic*, London: Longmans, Green & Co., 1949.(Book VI contains a significant reworking of Mill's views on economic methodology; discussed in §3.)

Mirowski, P. (1990) *More Heat Than Light*, Cambridge: Cambridge University Press.(A recent discussion of the history and methodology of economics that emphasizes the use of analogies to physics; cited in §6.)

Mises, L. von (1949) *Human Action. A Treatise on Economics*, New Haven, CT: Yale University Press.(Mises' major work, which defends a view of the basic principles of economics as known a priori; cited in §3.)

Morgan, M. (1990) *History of Econometric Ideas*, Cambridge: Cambridge University Press.(A historical account of econometrics that highlights its methodological problems.)

Popper, K. (1968) *The Logic of Scientific Discovery*, London: Hutchinson & Co, rev. edn.(Popper's major work discussed in §5.)

Redman, D. (1989) *Economic Methodology: A Bibliography with References to Works in the Philosophy of Science, 1860-1988*, New York: Greenwood Press.(A useful source for further references.)

Robbins, L. (1932) *An Essay on the Nature and Significance of Economic Science*, London: Macmillan, 1935. (One of the most important books on methodology; cited in §3.)

Roscher, W. (1874) *Geschichte der National-oekonomik in Deutschland* (The History of the National Economy in Germany), Munich: R. Oldenbourg.(One statement of the views of the German Historical School mentioned in §3.)

Rosenberg, A. (1992) *Economics - Mathematical Politics or Science of Diminishing Returns?*, Chicago, IL: University of Chicago Press.(Denies the possibility of empirical progress in economics; cited and discussed in §6.)

Roth, A. (1988) 'Laboratory Experimentation in Economics: a Methodological Overview', *Economic Journal* 98: 974-1031.(A survey emphasizing methodological issues; cited in §6.)

Samuels, W. (ed.) (1980) *The Methodology of Economic Thought: Critical Papers from the Journal of Economic Thought (Issues)*, New Brunswick, NJ: Transaction Books.(A collection of methodological essays from an institutionalist perspective.)

Samuelson, P. (1938) 'A Note on the Pure Theory of Consumer's Behavior', *Economica* 5: 61-71.(An influential presentation of revealed-preference theory that led economists to take a more positivistic view of economic methodology; cited in §3.)

Samuelson, P. (1963) 'Problems of Methodology - Discussion', *American Economic Review Papers and Proceedings* 53: 232-6.(A presentation of Samuelson's views on economic methodology; cited in §4.)
Senior, N. (1836) *Outline of the Science of Political Economy*, New York: A.M. Kelley, repr. 1965.(Contains one of the first discussions of economic methodology. Senior's position is similar to Mill's and cited in §3.)
Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, Oxford: Oxford University Press, 1976.(One of the most important economics books ever written; cited in introduction.)
Stegmüller, W., Balzer, W. and Spohn, W. (eds) (1982) *Philosophy of Economics: Proceedings, Munich, July 1981*, New York: Springer.(Applications of the structuralist view of scientific theories to economics; cited in §6.)
Veblen, T. (1898) 'Why is Economics not an Evolutionary Science?', *Quarterly Journal of Economics* 12: 373-97.(A classic institutionalist critique of neoclassical economics.)
Watkins, J. (1984) *Science and Scepticism*, Princeton, NJ: Princeton University Press.(A sophisticated account of philosophy of science in the Popperian tradition; cited in §5.)
Weber, M. (1975) *Roscher and Knies: The Logical Problem of Historical Economics*, trans. G. Oakes, New York: Macmillan.(A discussion of the German Historical School; cited in §3.)
Weintraub, E. (1985) *General Equilibrium Analysis: Studies in Appraisal*, Cambridge: Cambridge University Press.(An influential application of Lakatos' philosophy of science; cited in §5.)
Weintraub, E. (1991) *Stabilizing Dynamics*, Cambridge: Cambridge University Press.(A history of recent work on dynamics defending postmodern, social constructivist views.)Essays on economic methodology are conveniently indexed in *The Journal of Economic Literature* and in the *Index of Economic Articles in Journal and Collective Volumes* under the number, 036 prior to 1991 and under the number B4 since then. Redman (1989) is an annotated bibliography of essays on economic methodology published before 1988.)

Education, history of philosophy of

The philosophy of education may be considered a branch of practical philosophy, aimed ultimately at the guidance of an important aspect of human affairs. Its questions thus arise more or less directly from the features of educational practice and the role of education in the promotion of individual and social wellbeing, however much its answers may be conditioned by the larger philosophical and historical settings in which they are posed. Philosophers have concerned themselves with what the aims of education should be, and through what forms of instruction, inquiry and practice those aims might be attained. This demands attention to the contents of instruction and who shall have authority over it. It demands attention to the nature of instruction itself, its epistemic dimensions and what is entailed by its reliance on language; the nature of learning and human development, both moral and intellectual; and how all of these are interrelated. The philosophy of education thus stands at the intersection of moral and political philosophy, epistemology, and the philosophy of mind and language, as they bear on the foundations of educational practice.

The philosophy of education began in classical antiquity with the challenges posed by Socrates to the educational claims of the sophists. Plato and Aristotle developed systematic theories of education guided by an ethic of justice and self-restraint, and by the goal of promoting social harmony and the happiness or wellbeing of all citizens. The Stoic descendants of Socrates were expelled from Rome and the oratorical model of higher education given official sanction, but Augustine re-established the philosophical model through a synthesis of Platonism and Christianity, and in his mature educational thought brought elements of the oratorical and Platonic models together in his account of the Christian teacher's training.

The religious wars of the Reformation inspired several philosophical stances toward the relationships of Church, state, school and conscience. Hobbes argued for a consolidation of ecclesiastical and civil authority, with full sovereign authority over education; Locke for liberty, religious toleration, and private education aiming at self-governance in accordance with reason; and Rousseau not just for the free development and exercise of the full array of human faculties, but for the establishment of a civic religion limited to the core of shared Christian beliefs which Enlightenment figures from Descartes onward had thought evident to natural reason.

The Enlightenment's embrace of science and reason yielded efforts towards the development of a science of learning and pedagogy in the nineteenth century, but Rousseau's romantic reaction to it and defence of democracy were also powerful influences. In the twentieth century, Dewey produced a new synthesis of Enlightenment and Rousseauian themes, drawing on Hegel, the experimentalism of Mill, evolutionary theory and psychology, and aspects of the substance and intent of Rousseau's pedagogy.

1 Socrates and the sophists

Educational philosophy began in the Greek classical period with the examination of the educational claims of the sophists undertaken by Socrates. The sophists brought higher education to the democratized Athens of the fifth and fourth centuries BC, offering those who aspired to political leadership a training in political *aretē* (the goodness, excellence or virtue required for success in pursuing appropriate ends) or *phronēsis* (sound judgment or practical wisdom) (see Aretē; Sophists). This form of education suggested that most citizens lacked the virtue and judgment required for a life in public affairs, and one can detect a concession to the political dangerousness of this in the claims of Protagoras that cities and their citizens do indeed teach virtue to the young, but that his own teaching could refine and develop it by degrees. In Plato's *Protagoras*, Socrates exposes the tensions in this view by distinguishing between the habitual virtue of good or obedient citizens, and true virtue which involves intellectual insight and sound judgment (that is, *phronēsis*), noting that a skill which merely refines and enlarges the former cannot yield the latter.

More generally, the Socratic response to the sophists was above all cautionary. As we encounter Socrates in the *Protagoras* and other early dialogues of Plato, he dedicated himself to showing through his method of questioning (*elenchus*) that those who claimed to be teachers of *aretē* lacked the expert knowledge of it which its teaching would require. The possession of such knowledge would allow one to defend and explain the truths one believes through a reasoned account (*aitias logismos*), and Socrates denied that he was himself a teacher, apparently on the grounds that he was unable to give such an account of his own beliefs. He advocated the individual care of one's

own soul, and embraced an ethic of justice, wisdom and self-restraint, in opposition to the competitive ethic of the warrior heroes portrayed by Homer and embraced by Greek popular morality. How far he thought his own elenctic method would carry one in this care of the soul, or in the search for the best way to live, is unclear (see Socrates).

2 Plato and Aristotle

The adequacy of the Socratic *elenchus* was evidently one of Plato’s concerns in the *Meno*, for the work presents us with a Socratic interlocutor who is purged of misplaced confidence and ready to join Socrates in a search for the nature of virtue, but asking how the search is to go forward without the benefit of any knowledge of the object of the search. This prompts the introduction of Plato’s theory that all learning is recollection, and a demonstration of an *elenchus* arriving, without recourse to prior knowledge, at the discovery of a geometrical fact. The theory of recollection underwrites the hope that a search will yield truth, by suggesting that one can count on having enough true beliefs to drive out the false, as one’s beliefs become more coherent under questioning. Plato implies that one ascends from there to knowledge as one reasons through the relationships among all the truths in the relevant domain and can thereby understand and provide an explanatory account of any one of them.

In the *Republic*, Plato envisages the inculcation of self-restraint through a comprehensive system of compulsory education for each social class, and develops an elaborate defence of the Socratic thesis that justice is a virtue indispensable for a happy life. Plato was evidently concerned that the education of his time, which was dominated by the study of Homer, was inconsistent with the inculcation of these virtues, and regarded an education which would eliminate unnecessary desires for contested goods as the one way to eliminate wars of conquest and the political instability so prevalent in Greek life. He envisages a craft training for members of the labouring class, which will free them from the dominance of desires for unnecessary consumption and prepare them for a happy life of efficient money making. Members of the ‘guardian’ class are to be given a communal upbringing which will eliminate conflict arising from clan loyalties, and a primary education in music, athletics and stories which will enable them to be happy in their pursuit of honour by making them courageous and gentle towards each other and those they protect. The philosopher-kings must know what is good for the city, and Plato proposes an education which will not only elevate their minds through the study of abstract sciences to dialectic and apprehension of the transcendent ‘Form’ of goodness, but also provide many years of experience in the affairs of the city.

Plato’s *Laws* describes a city which resembles that of the *Republic* in aiming at the virtue and happiness of all citizens, but is quite unlike it in being a constitutional rule of law grounded in reason and informed consent. The insistence on consent, and an unwavering commitment to the idea that every institution, including legislation and its enforcement, is educative, leads to the proposal that all laws be prefaced with ‘preludes’ that are both moving and explanatory, and that these ‘preludes’ should serve as models for the literary substance of state-operated day schools. Two noteworthy aspects of the theory of learning and instruction at work in the *Laws* are the prominent role assigned to guided practice, and its rich use of the model of a good doctor who communes with and earns the trust and cooperation of a patient before prescribing a remedy. It shares with the *Republic* the view that rational self-control and the ability to perceive what is good can only develop under the influence of good upbringing and education.

Aristotle’s *De Partibus Animalium (On the Parts of Animals)* opens with a characterization of a complete education as one which enables a person to form reasonable judgments of the goodness or badness of reasoning in all branches of knowledge. This requires an understanding of the manner and canons of investigation in the various branches of knowledge, and it is fair to say that Aristotle dedicated himself not only to the discovery and elaboration of those canons of investigation, but also to making them known through his lectures and the wide-ranging scientific investigations of his ‘school’, the Lyceum.

Of the works of Aristotle that have survived, it is the *Politics* and *Nicomachean Ethics*, the works of ‘political science’ intended as guides to political practice, that are most revealing of his educational philosophy. They suggest that he thought a higher education for political *phronēsis* should provide leaders and their advisors with systematic knowledge of how to pursue the proper aim of the state (namely, the best kind of life for all its citizens), and a knowledge of the causes of dissolution of states which would provide corrupt leaders with the motivation to pursue reforms. Plato had argued that unjust regimes tend to be short-lived, and Aristotle offered his own comparative study of 158 constitutional histories as confirmation.

Aristotle shared Plato's view that justice and human wellbeing require systematic educational efforts to make citizens virtuous and to create social unity, but his accounts of education for good character and unity are interestingly different. His view of *aretē*is that what is essential to it, beyond being disposed to desire the right ends and take pleasure in their attainment, is judging and acting from a well-rounded grasp of the morally important particulars of the situations one faces. The beginnings of *aretē* must be established through exposure to good models and habitual obedience to good law, but its refinement requires practice guided by coaching and correction. This progress in perception, responsiveness and judgment reaches its highest development through conversation with trusted and exemplary companions who can best enable one to know oneself.

Aristotle was critical of the communism of the *Republic*, and followed the Plato of the *Laws* in proposing the establishment of state-sponsored day schools. He seems to have regarded common schooling, and its proposed education for virtue, as the natural culmination of reforms by which a city can be unified through the cultivation of goodwill, common non-competitive aspirations, and friendships which bridge the divides between different social classes.

3 Hellenistic and Roman education

The conquests of Alexander the Great brought an end to any hope that the fulfilment of human nature could be achieved through the educational and political efforts of an autonomous *polis* or city-state. Philosophy of education thereby lost its connection to public life in the Hellenistic Age, and what remained of it in Stoicism was the doctrine that an understanding of divine reason allows one to be happy, good and free of the tyranny of unnecessary desires, by enabling one to accept the inevitability of lacking what one does not have and cannot possess. The Stoics took divine reason to be manifested in the order of nature, and Chrysippus and others apparently held that an adequate understanding of nature is best attained through a study of logic and physics (see Stoicism §3).

Rome had no need for a philosophy withdrawn from public life, and when the Emperor Vespasian granted the first public subsidy for higher learning in the first century, it went to the school of rhetoric operated by Quintilian. His work, the *Institutio Oratoria (The Training of the Orator)*, is both a textbook of rhetoric and a treatise on the principles of education, and it exercised a powerful influence on Western educational ideals well into the nineteenth century. Quintilian took the highest aim of education to be the formation of a perfect orator, a person made both eloquent and good through exercises in composition and declamation, and a broad study of the liberal arts and rhetorical theory.

4 Augustine and Aquinas

Augustine is regarded, with justice, as not only the primary architect of the synthesis of Christianity and Platonism on which Christian theology was built, but also the foremost philosopher of education of late antiquity and of Catholicism generally. He was trained as an orator and held teaching posts in rhetoric for twelve years, but upon his conversion to Christianity in 386 he renounced the shallow worldliness of the oratorical ideal, setting himself and the philosophy of education upon a new course. Fundamental to Augustine's view of education is the idea that its proper goals are first of all conversion and repentance, or the acceptance on faith of Christian beliefs and a commitment and effort of will to live without sin; and secondly, for the few who can achieve it, a wisdom consisting of direct knowledge of the soul and God through reason, or the insight of an intellect directed inwardly.

Augustine's view of the course of higher studies preparatory for the attainment of this wisdom, and for its use by the Christian teacher, underwent a marked development between the educational works written in the first years after his conversion and those finished after he was consecrated bishop of Hippo in 395. The mainsprings of this development were a dramatic growth in his respect for biblical Scriptures as the word of God, which led him to revise his view of the role and value of liberal studies, and an appreciation for the value of rhetoric which grew with his responsibilities in the Church.

In his *De ordine (On Order)*, Augustine affirms the existence of an all-embracing Divine order, knowable through liberal studies which allow one to grasp the pervasive reasonableness in things emanating from God's intellect or Logos. It is dialectic or philosophical reasoning which teaches one how to learn and how to teach, he says, and this is compatible both with his view that the soul must prepare itself to see God through the exercise of its own reason,

and his view in *De magistro (The Teacher)* that knowledge of a thing is achieved in *seeing* it, not in hearing someone name or describe it. He argues there through a theory of signs and language that teaching through words cannot enable us to know things by displaying them, but can only remind or prompt us to consult our own divinely illuminated reason, or else induce us to believe things on trust without understanding or knowledge.

Augustine's *De catechizandis rudibus (Catechizing the Uninstructed)* develops a curriculum for the uneducated which is intended to induce moral behaviour and faith in God sufficient for salvation. His *De doctrina christiana (Christian Instruction)* explicates the training and methods through which Christian teachers may properly interpret biblical Scripture and teach the truths discovered in them. He prescribes a course of studies encompassing languages, history, geography, natural science, technology, logic and mathematics, all as an aid to understanding scripture, and offers canons of interpretation by which the Christian teacher may distinguish the literal from the allegorical, and find the hidden meaning in the latter. Rhetoric returns in the service of faith and love of God and neighbour, in the training of the Christian teacher and in detailed guidance on how the truths discovered in scripture are best presented.

The embrace of Aristotle by Thomas Aquinas produced a second major synthesis of classical and Christian thought, and a philosophy of teaching which owes much to both Aristotle and Augustine. Aquinas argues that acquiring knowledge by discovery and by instruction are alike in that a teacher can only lead a student to knowledge by making manifest to them the discursive reasoning through which discoveries are made. 'Outward' teaching may thus produce knowledge by aiding the student in reasoning through the application of self-evident general principles to particular matters, though it is God who teaches 'inwardly' by endowing the mind with the 'light of reason' and a knowledge of these first principles.

5 The Renaissance and Reformation

The Renaissance and Reformation are noteworthy in educational history on many grounds, including the incorporation of the educational ideals and programme of Quintilian into the Christian humanism of Desiderius Erasmus and others; Martin Luther's translations and promotion of universal, publicly-funded elementary education, so that the common people might read and thereby consult for themselves the Scriptures (see Luther, M.); and the founding of the Jesuit order and its system of colleges by Ignatius Loyola. What inspired the most weighty philosophical response, however, was the turbulent course of the Reformation itself, and the urgent questions it raised about the relationships between church, state, school and individual conscience.

Thomas Hobbes has been neglected by histories of philosophy of education, but a central concern of his *Leviathan* was to establish the desirability of uniting civil and ecclesiastical authority in a sovereign empowered to use public education to inculcate correct religious and moral beliefs. In both *Leviathan* and *Behemoth*, his history of the English Civil War, he attributed social disorder, including conflict and civil war arising from the exercise of individual judgment in religious matters, to faulty instruction from the pulpit and in the universities. This instruction not only incited conflict over theological questions without answers, he argued, but spread doctrines contrary to the natural law of civil and moral duty. Because these doctrines shaped a conception of religious interests linked to the expectation of an afterlife, Hobbes regarded attempts to suppress rebellion by force as futile, and argued that sovereign authority over education and the interpretation of scripture was desirable and necessary to ensure peace.

6 Descartes and Locke

René Descartes' remarks on education are continuous with the classical tradition in identifying universal wisdom as the greatest good and proper aim of instruction, but he turned a deeply Augustinian account of reason against the curricula of Augustine, Aquinas and the scholastic university, and championed the free and methodical exercise of individual reason or self-instruction as the surest path to wisdom and the progress of science and human emancipation. He regarded the unrestrained use of reason in the pursuit of understanding as a natural and desirable expression of human nature, and held that the individual intellect can easily discover within itself a knowledge of God, the soul and good actions. Thus, he regarded the inward gaze of reason itself as the best guide for the conduct of life, and the means through which theological disagreements and the conflict arising from them may be resolved. However, he held that the purging of prejudice through doubt or the suspension of belief is a necessary prerequisite to the effective employment of reason.

With regard to the curricular proposals of Augustine, Descartes maintained that the revealed truths of Scripture are beyond the intelligence of even the best trained interpreter, and that in general it takes more skill to find what is true in books than to discover the truth for oneself. In answer to Aquinas and the scholastic tradition, he argued that instruction through syllogistic deductions undermines both the freedom and power of reason, because syllogistic deduction does not ensure any intellectual grasp of the propositions involved and does not enable one to discover new knowledge. Exercise with inquiries conducted on his own more natural method of reasoning is the recommended alternative (see Descartes, R. §2).

Descartes laid the foundations for the educational philosophy of the Enlightenment, but it remained to John Locke to provide an influential articulation of it. Locke deepened Descartes' attack on second-hand knowledge by combining the latter's view that knowledge can only be obtained through the perception of 'clear and distinct' ideas, with his own systematic development of the claim that one can only acquire ideas through one's own experience. With this epistemic individualism as his starting point, he took the fundamental and humanizing goal of education to be the development of rational abilities and the habits of doubt, reflection and foresight required to form children into adults who will judge and act in accordance with the dictates of reason. This is consistent with his having put virtue, wisdom, and 'breeding' (that is, respect for self and others) before learning or the acquisition of knowledge, since he regarded virtue as a disposition to follow one's desires only when doing so accords with the dictates of reason or natural moral law. Autonomy of judgment and action are thereby reconciled with virtue and sociability through the concept of self-mastery, and education is to be first and foremost a process of habituation to self-mastery.

Locke held that parents have a natural duty to educate their children to be self-governing and able to provide for themselves, but he proposed public action to ensure that even children of paupers would receive instruction in Christian morality and a trade: those who will never have the leisure for knowledge or science would have little use for more than this, he suggested. For the sons of landed gentry he counselled a general education through private tutelage, aimed to produce gentlemen ready to contribute to the wellbeing of the community. This education should encourage a breadth of experience and inquiry reaching down to the foundations of beliefs, and promote reflection on desires by refusing to gratify those that are excessive. It should exercise firm but gentle authority over the child, appeal early and often to the child's reason, and make education enjoyable since autonomous learning cannot be coerced (see Locke, J. §9).

7 Rousseau and Kant

The concern that animates the educational and political thought of Jean-Jacques Rousseau is the preservation of the freedom and goodness which he took to define human nature, in opposition to the Enlightenment conception of humanity as essentially rational, progressing towards happiness and the perfection of its nature through the cultivation of reason. Rousseau argued in his first and second *Discourses* that what the arts and sciences encourage are sentiments of self-love and ambition which are destructive of our goodness and freedom, leading us to an insatiable pursuit of private gain at the expense of the common good. To this image of fallen humans, rooted in the orthodox identification of pride as the 'root of all evil', he coupled a further challenge to the hope that progress might be achieved through the ascendancy of reason, observing that in such a state of society the voice of individual reason will counsel free-riding, the unjust enjoyment of the rights of citizenship without the discharge of corresponding duties.

Rousseau held in his *Émile*, the most influential educational tract of the modern era, that in order to educate a boy both for himself and for society, one must preserve his natural freedom and goodness by conforming the education of 'humans' and 'things' to nature's timetable for the development of his faculties and motives. Children are moved first by the expectation of pleasure and pain, later by a grasp of what is useful, and finally by reason; and they should learn at each stage through experience and feeling, not through books or the imposition of discipline or teaching. The tutor need do little but shield the child from the corrupting influence of society and organize his activities (and thereby his experiences and feelings) in order to nurture the growth of moral sentiments and conscience, of a self-sufficiency grounded in prudence and mastery of a craft, and of his faculties and knowledge of the world, God and moral law. Freedom from material reliance on others and from the domination of excessive desires is thus assured.

By contrast with Locke, Rousseau held that freedom from domination by others requires that the child's behaviour

be corrected solely by the experience of the natural consequences of actions, and that adults make law for themselves through a form of democracy in which the citizens retain and exercise direct legislative sovereignty. Good citizenship requires, however, that they aim at the common good when it conflicts with their private ends, and Rousseau envisages in *On The Social Contract* a 'Great Legislator' who proposes good laws and persuades the public to enact and obey them through neither force (which would negate freedom) nor rational argument (which wouldn't solve the free-rider problem). The legislator must teach by example and through a 'natural' civil religion, established not as dogma but as sentiments of sociability (see Rousseau, J.-J. §3).

Immanuel Kant's philosophy of education is like Rousseau's in being grounded in a philosophy of history, but it reaffirms not only the essential rationality of human nature, but also original sin and the naturalness of discord and competitive self-love. His conception of the end for which we were created is that our rational nature should find full expression, that we should achieve moral perfection through effort and the grace it occasions, and that we should thereby become both happy and worthy of that happiness. This future kingdom of heaven on earth is what children should be educated for, but the stages of their education must first recapitulate the stages of human history already traversed. The first of these is nurture; the second discipline, to counterbalance natural unruliness; the third is culture (information and instruction), to develop ability and prudence; the last (for the world order yet to come) is moral training, to encourage respect for moral law and an acceptance of its priority over self-love as the principle of one's will.

8 The nineteenth century

The nineteenth century was a period of unprecedented innovation in the theory of learning and the theory and practice of pedagogy, particularly in Germany. From a philosophical standpoint this activity can be described most simply as a playing-out of Enlightenment ideas and of the romantic and aesthetic reactions to them initiated by Rousseau and Friedrich von Schiller. Johann Heinrich Pestalozzi (1746-1827) and Friedrich Wilhelm Froebel (1782-1852) were the most prominent of those strongly influenced by Rousseau, the former being known for a pedagogy grounded in the analysis of tasks into simple components, his advocacy of universal educational rights and efforts to educate the poor 'organically' in a family-like setting; the latter for founding the kindergarten as a place of growth in unity with God and nature. Johann Friedrich Herbart is noteworthy for his attempts to develop the experimental science of pedagogy envisaged by Kant, and for a theory of ideas and learning which held that it is 'apperceptive masses' formed by past experience that structure new experiences, allowing the assimilation of ideas that are compatible but driving from consciousness those that are not.

Though not a philosopher of education himself, Karl Marx has figured importantly in the subsequent history of the field through, among other things, his theory of ideology, an exceedingly influential if unsystematic account of the way in which socioeconomic roles and the ideas that structure them are learned through contact with a society and its institutions. Contraposed to the radical egalitarianism of Marx was the radical elitism of Friedrich Nietzsche, the other great immoralist of the nineteenth century, who condemned the democratization of the universities on the grounds that the reversal of cultural decadence and advancement of human wellbeing through the free play of creativity requires that higher education remain the exclusive province of 'higher types'.

9 John Dewey

John Dewey is regarded with justice as the pre-eminent philosopher of education of the twentieth century. Like Marx and Nietzsche, he understood the function of philosophy to be social and cultural reconstruction, and like Marx he aspired to reconstruct philosophy itself by bringing science into it. But he rejected the teleological conceptions of history embraced by Marx, Hegel and Kant in favour of a kind of evolutionary naturalism. He took this naturalism to preclude the existence of absolute norms of conduct, and to warrant recognition of a plurality of goods and moral criteria. If there is any 'end' entailed by human nature it is growth itself, he argued and, following Rousseau, he held that a good society is one in which growth is maximized, harmonized with that of others, and reconciled with the collective work of citizenship. He argued, much as Rousseau did, that democracy is the one form of political life in which the convergence of these goals is possible, and he adopted and developed John Stuart Mill's idea that social progress is possible through the freedom of individuals to engage in social experimentation, or the application of the methods of science to the problems of social existence (see Mill, J.S.). Dewey insists that this experimentation or application of 'social intelligence' must be cooperative, governed by the norms of respectful communication, and guided by the plurality of goods and criteria valued by those whose

interests are at stake. He thought that in this way the Enlightenment promise of social progress through the application of scientific method might be achieved simultaneously with enacting democracy as 'a way of life' in which human growth and quality of experience can be optimized.

Experience became for Dewey both the goal and means of education: the richness of future experience its goal, and present experiences which engage the student's activity and improve the prospect of desirable experiences in the future its means. His methodological Hegelianism, or strategy of undercutting the dualisms dividing one school of opinion against another, is as evident in his philosophy of education as elsewhere. He resisted the duality of thought and action, book learning and vocationalism, pedagogy driven by a preconceived curriculum and pedagogy driven by the antecedent interests of the child. Above all, he held that the social world of the school and of a properly democratic society must coincide, and that the exercise of intelligence and engagement in adaptive conduct are inseparable. Classrooms must therefore be devoted to the collaborative activity of inquiry, he concluded (see Dewey, J.).

See also: Education, philosophy of; Kaibara Ekken

RANDALL R. CURREN

References and further reading

Aquinas (*c*.1257) 'De Magistro' (The Teacher), question 11 in *Quaestiones Disputatae de Veritate (Truth)*, vol. 2, trans. J.V. McGlynn, Indianapolis, IN: Hackett Publishing Co., 1994. (Aquinas' treatise on teaching.)

Augustine (386) *De ordine (On Order or Divine Providence and the Problem of Evil)*, trans. R.P. Russell in *The Fathers of the Church: The Writings of Saint Augustine*, vol. 1, ed. L. Schopp, New York: Cima Publishing Co., 1948, 229-332.(An early work on liberal studies and knowledge of God.)

Augustine (386-9) *Contra Academicos (Against the Academicians)* and *De magistro (The Teacher)*, trans., intro. and notes by P. King, Indianapolis, IN: Hackett Publishing Co., 1995. (A critique of scepticism and an account of knowledge and teaching.)

Augustine (397-426) *De doctrina christiana (Christian Instruction)*, trans. J.J. Gavigan in *Fathers of the Church: The Writings of Saint Augustine*, vol. 4, ed. R. Deferrari, Washington, DC: The Catholic University of America Press, Inc., 1947-66, 19-235.(A treatise on the Christian teacher's training and methods of teaching.)

Augustine (400) *De catechizandis rudibus (Catechizing the Uninstructed)*, trans. J.P. Christopher, Washington, DC: Catholic University Press, 1926.(Proposes a course of instruction for the uneducated.)

Aristotle (*c*.330 BC) *De Partibus Animalium (On the Parts of Animals)*, Greek text with trans. A. Peck in *Aristotle*, vol. 12, Loeb Classical Library, Cambridge, MA: Harvard University Press.(Opens with remarks on the nature of education.)

Aristotle (*c*.330 BC) *Nicomachean Ethics*, trans. and notes T. Irwin, Indianapolis, IN: Hackett Publishing Co., 1985.(Aristotle's main ethical work and an important source of his educational thought.)

Aristotle (*c*.330 BC) *Politics*, trans., intro. and notes by C. Lord, Chicago, IL: University of Chicago Press, 1984. (Aristotle's lectures on politics and the place of education in a good city.)

Ashcraft, R. (ed.) (1991) *John Locke: Critical Assessments*, vol. 2, London: Routledge. (Includes articles on Locke's educational thought and related topics such as reason, natural law and toleration.)

Bacon, F. (1605-10) *The Advancement of Learning and New Atlantis*, ed. A. Johnston, Oxford: Clarendon Press, 1974.(Seminal works of the modern era on the condition and advancement of knowledge and human progress - they shaped the pedagogical proposals of J.A. Comenius and others.)

Beck, F.A.G. (1964) *Greek Education*, London: Methuen.(A historical account of Greek education and its leading theorists.)

Beck, L.W. (1978) 'Kant on Education', in *Essays on Kant and Hume*, New Haven, CT: Yale University Press.(A valuable introduction and guide to the relevant texts.)

Curren, R. (1997) *Aristotle on the Necessity of Public Education*, Savage, MD: Rowman & Littlefield.(A study in Aristotle's educational politics which gives close attention to the influence of Plato's *Laws*.)

Descartes (1628-49) *The Philosophical Writings of Descartes*, vols 1 and 2, trans. J. Cottingham *et al.*, Cambridge: Cambridge University Press, 1984-5.(See his *Rules For the Direction of The Mind*; *Discourse on Method*, parts 1 and 2; *The Search For Truth By Means of the Natural Light*; and the 'Author's Letter' or preface to the *Principles of Philosophy*.)

Dewey, J. (1899) *The School and Society*, Chicago, IL: University of Chicago Press.(A good introduction to Dewey's educational thought.)

Dewey, J. (1902) *The Child and the Curriculum*, Chicago, IL: University of Chicago Press. (Another early and

accessible statement of Dewey’s educational philosophy.)
Dewey, J. (1909) *Moral Principles in Education*, Boston, MA: Houghton Mifflin.(On moral education in the schools.)
Dewey, J. (1916) *Democracy and Education*, New York: Macmillan.(Dewey’s most systematic work on education.)
Dewey, J. (1927) *The Public and Its Problems*, New York: Henry Holt.(On social intelligence and related ideas.)
Dewey, J. (1938) *Experience and Education*, New York: Macmillan.(An introductory restatement of his educational views which looks back critically on the progressive education movement.)
Epictetus (*c*.120) *The Discourses as Reported by Arrian, The Manual, and Fragments*, Greek text with trans. by W. Oldfather, 2 vols, Cambridge, MA: Harvard University Press, 1966-7. (An example of late Stoicism with extended commentary on moral instruction.)
Flower, E. and Murphey, M.G. (1977) *A History of Philosophy in America*, vol. 2, New York: Putnam.(The long chapter on Dewey provides an excellent overview of his thought.)
Froebel, F. (1826) *Die Menschenerziehung* (*The Education of Man*), trans. W.N. Hailmann, Clifton, NJ: Keeley, 1974.(The earliest and most systematic of Froebel’s educational works.)
Garforth, F.W. (1980) *Educative Democracy: John Stuart Mill on Education in Society*, Oxford: Oxford University Press.(A detailed account of Mill’s views on education and its role in society.)
Herbart, J.H. (1804-6) *Über die ästhetische Darstellung der Welt als das Hauptgeschäft der Erziehung; Allgemeine Pädagogik aus dem Zweck der Erziehung abgeleitet* (*The Science of Education, Its General Principles Deduced From Its Aim, and the Aesthetic Revelation of the World*), trans. with intro. H.M. Felkin and E. Felkin, Boston, MA: Heath, 1892.(A work of practical pedagogical theory.)
Herbart, J.H. (1836) *Umriß von pädagogischen Vorlesungen* (*Outlines of Educational Doctrine*), trans. A.F. Lange, notes C. De Garmo, London: Macmillan, 1901.(A theoretical enquiry into how education is possible.)
Hobbes, T. (1651) *Leviathan: Or The Matter, Form, and Power of a Commonwealth Ecclesiastical and Civil*, ed., intro. and notes E. Curley, Indianapolis, IN: Hackett Publishing Company, 1994.(The most important of Hobbes’ works of political philosophy.)
Hobbes, T. (1668) *Behemoth*, ed. W. Molesworth, Source Works Series No. 38, New York: Burt Franklin Research, 1962.(Reveals Hobbes’ view of the role of the universities in the English Civil War.)
Hook, S. (1946) *Education For Modern Man: A New Perspective*, New York: Alfred A. Knopf.(An elaboration and defence of Dewey’s philosophy of education.)
Kant, I. (1775-80) *Eine Vorlesung Kants Über Ethik im Auftrage der Kantgesellschaft* (*Lectures on Ethics*) trans. L. Infield, Indianapolis, IN: Hackett Publishing Company, 1980.(See especially ‘The Ethical Systems of the Ancients’; ‘Conscience’; ‘Self-Mastery’; ‘Duties Towards Particular Classes of Human Beings’; ‘Duties Arising From Differences of Age’; and ‘The Ultimate Destiny of the Human Race’.)
Kant, I. (1784-98) *Political Writings*, ed., intro. and notes H. Reiss, trans. H.B. Nisbet, Cambridge: Cambridge University Press, 1991.(See especially the essays ‘Idea For a Universal History With a Cosmopolitan Purpose’; ‘An Answer to the Question: "What is Enlightenment?"’; ‘The Contest of Faculties’; and ‘Conjectures on the Beginning of Human History’.)
Kant, I. (1803) *Über Pädagogik* (*Kant on Education*), trans. A. Churton, Bristol: Thoemmes Press, 1992.(A compilation of Kant’s notes and published remarks on education.)
Kraut, R. (ed.) (1992) *The Cambridge Companion to Plato*, Cambridge: Cambridge University Press.(Many important aspects of Plato’s educational philosophy are examined in this accessible guide to his thought.)
Lloyd, S.A. (1992) *Ideals as Interests in Hobbes’s Leviathan*, Cambridge: Cambridge University Press.(An interpretation which takes seriously the important role of education in the *Leviathan*.)
Locke, J. (1689) *Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1975.(A seminal work in the theory of ideas and knowledge.)
Locke, J. (1690) *Second Treatise of Government*, ed. and intro. C.B. Macpherson, Indianapolis, IN: Hackett Publishing Company, 1980.(A political work which includes remarks on the educational duties of parents.)
Locke, J. (1693) *Some Thoughts Concerning Education*, ed., intro. and notes J.W. Yolton and J.S. Yolton, Oxford: Clarendon Press, 1989.(Locke’s principal work on education.)
Locke, J. (1706) *Of the Conduct of the Understanding*, ed., intro. and notes T. Fowler, New York: Burt Franklin, 1971.(Reprint of 1882 edition. A treatise on the cultivation of the intellect.)
Maritain, J. (1943) *Education at the Crossroads*, New Haven, CT: Yale University Press.(An influential elaboration of a twentieth-century Catholic philosophy of education.)
Marrou, H. (1956) *A History of Education in Antiquity*, trans. G. Lamb, London: Sheed & Ward.(A readable account of educational theory and practice down through the Hellenistic, Roman and early medieval periods.)

Marrou, H. (1957) *St. Augustine and His Influence Through the Ages*, New York: Harper & Brothers.(Standard interpretation of the educational aspects of Augustine's thought and influence.)

Marx, K. (1845-6) *Die deutsche Ideologie* (*The German Ideology*), pt 1, with selections from pts 2 and 3, ed. C.J. Arthur, trans. W. Lough, C. Dutt and C.P. Magill, New York: International Publishers, 1974.(The only work by Marx expressly concerned with ideology.)

Marx, K. (1867) *Das Kapital* (*Capital*), vol. 1, ed. F. Engles, trans. S. Moore and Aveling, New York: International Publishers, 1967.(An important source of Marx's ideas about ideology.)

Masters, R.D. (1968) *The Political Writings of Rousseau*, Princeton, NJ: Princeton University Press.(A general and very readable commentary offering an extended examination of *Émile* as its point of departure.)

Mill, J.S. (1859) *On Liberty*, ed. E. Rapaport, Indianapolis, IN: Hackett Publishing Company, 1978. (An important source of liberal, educational and cultural theory.)

Mill, J.S. (1867) 'Inaugural Address at the University of St. Andrews', in *John Stuart Mill on Education*, ed., intro. and notes F.W. Garforth, New York: Teachers College Press, 1971. (This volume also contains the parts of Mill's *Autobiography* which are of educational interest.)

Mill, J.S. (1873) *Autobiography*, ed. R. Howson, New York: Columbia University Press, 1944. (Includes a lengthy account of Mill's own education and commentary on it.)

Nehamas, A. (1992) 'What Did Socrates Teach And To Whom Did He Teach It?', *Review of Metaphysics* 46 (4): 279-306.(An examination of Socrates' denial that he was a teacher, with useful discussion of other interpretations.)

Nietzsche, F. (1872) *Über die Zukunft unserer Bildungsanstalten* (*On the Future of Our Educational Institutions*), in *The Complete Works of Friedrich Nietzsche*, vol. 3, ed. O. Levy, trans. J.M. Kennedy, New York: Russell & Russell, 1924.(A critique of the democratization of the universities.)

Nietzsche, F. (1874) *Schopenhauer als Erzieher* (*Schopenhauer as Educator*), trans. J.W. Hillesheim and M.R. Simpson, Chicago, IL: Regnery Gateway, Inc., 1965.(On what it means to be a teacher of mankind.)

Painter, F. (1889) *Luther on Education*, St. Louis, MO: Concordia Publishing House.(A readable but not impartial history with translations of Luther's two notable educational works.)

Pestalozzi, J. (1777-1826) *Pestalozzi's Educational Writings*, ed. J.A. Green and F. Collie, London: Longmans, 1912.(A selection of his diverse educational works.)

Plato (390s-350s BC) *Complete Works*, ed. J. Cooper, Indianapolis, IN: Hackett Publishing Company, 1996.(The dialogues of most direct relevance are the *Euthydemus*, *Protagoras*, *Meno*, *Gorgias*, *Symposium*, *Republic*, *Theaetetus* and *Laws*.)

Price, K. (1967) *Education and Philosophical Thought*, Boston, MA: Allyn & Bacon, 2nd edn.(An introductory but dated history, with selections from major figures.)

Quintilian (Marcus Fabius Quintilianus) (*c*.92) *Institutio Oratoria* (*The Education of an Orator*), vols. 1-4, text and trans. H.E. Butler, London: Heinemann, 1920-2.(On rhetoric and the principles of education.)

Reeve, C.D.C. (1988) *Philosopher-Kings*, Princeton, NJ: Princeton University Press.(An excellent treatment of the *Republic* and its educational proposals.)

Rousseau, J.-J. (1750-62) *The Basic Political Writings of Jean-Jacques Rousseau*, trans. D.A. Cress, Indianapolis, IN: Hackett Publishing Company, 1987.(Includes the first and second *Discourses*, *Political Economy*, and *On the Social Contract*.)

Rousseau, J.-J. (1762) *Émile*, (or *On Education*), trans. A. Bloom, New York: Basic Books, 1979. (Rousseau's principal work on education.)

Schiller, F. (1795) *Briefe über die ästhetische Erziehung des Menschen* (*On The Aesthetic Education of Man: In a Series of Letters*), trans. E.M. Wilkinson and L.A. Willoughby, Oxford: Clarendon Press, 1967.(A seminal defence of aesthetic education as the means to develop both sensibility and rationality. See §8 above.)

Schouls, P. (1989) *Descartes and the Enlightenment*, Edinburgh: Edinburgh University Press. (A detailed discussion of the aspects of Descartes' thought bearing on his concept of good instruction, and their influence on the Enlightenment.)

Sherman, N. (1989) *The Fabric of Character*, Oxford: Oxford University Press.(A sophisticated account of Aristotle's theory of virtue and character education.)

Sullivan, R. (1989) *Immanuel Kant's Moral Theory*, Cambridge: Cambridge University Press. (A comprehensive and lucid interpretation, with an appendix on moral education.)

Ulich, R. (ed.) (1954) *Three Thousand Years of Educational Wisdom: Selections From Great Documents*, Cambridge, MA: Harvard University Press.(Useful as an introduction to texts and traditions in educational thought which could not be included here.)

Whitehead, A.N. (1929) *The Aims of Education*, New York: Macmillan.(Apart from Dewey, possibly the most

significant work in philosophy of education from the first half of the twentieth century.)

Woodward, W. (1904) *Desiderius Erasmus: Concerning the Aim and Method of Education*, Cambridge: Cambridge University Press.(A historical and expository essay and translations of the major educational writings.)

Education, philosophy of

The philosophy of education is primarily concerned with the nature, aims and means of education, and also with the character and structure of educational theory, and its own place in that structure. Educational theory is best regarded as a kind of practical theory which would ideally furnish useful guidance for every aspect and office of educational practice. Such guidance would rest in a well-grounded and elaborated account of educational aims and the moral and political dimensions of education, and also in adequate conceptions and knowledge of teaching, learning, evaluation, the structure and dynamics of educational and social systems, the roles of relevant stake-holders and the like.

Philosophers of education often approach educational issues from the vantage points of other philosophical sub-disciplines, and contribute in a variety of ways to the larger unfinished project of educational theory. These contributions may be divided into work on the nature and aims of education, on the normative dimensions of the methods and circumstances of education, and on the conceptual and methodological underpinnings of its methods and circumstances - either directly or through work on the foundations of other forms of research relied upon by education theory.

Philosophical analysis and argument have suggested certain aims as essential to education, and various movements and branches of philosophy, from Marxism and existentialism to epistemology and ethics, have suggested aims, in every case controversially. Thus, one encounters normative theories of thought, conduct and the aims of education inspired by a broad consideration of epistemology, logic, aesthetics and ethics, as well as Marxism, feminism and a host of other '-isms'. In this mode of educational philosophizing, the objects of various branches of philosophical study are proposed as the ends of education, and the significance of pursuing those ends is elaborated with reference to those branches of study.

A second form of educational philosophy derives from substantive arguments and theories of ethics, social and political philosophy and philosophy of law, and concerns itself with the aims of education and the acceptability of various means to achieve them. It revolves around arguments concerning the moral, social and political appropriateness of educational aims, initiatives and policies, and moral evaluation of the methods, circumstances and effects of education. Recent debate has been dominated by concerns about children's rights and freedom, educational equality and justice, moral and political education, and issues of authority, control and professional ethics.

The philosophy of education has also sought to guide educational practice through examining its assumptions about the structure of specific knowledge domains and the minds of learners; about learning, development, motivation, and the communication and acquisition of knowledge and understanding. Philosophy of science and mathematics have informed the design of curriculum, pedagogy and evaluation in the teaching of science and mathematics. Philosophy of mind, language and psychology bear on the foundations of our understanding of how learning occurs, and thus how teaching may best promote it.

1 Philosophical analysis and theory

Philosophy of education emerged in the second half of the twentieth century as a distinct branch of philosophy, devoted to the systematic study of education through the methods of conceptual analysis and the methods and findings of other, more established philosophical sub-disciplines and movements. This was evident in the creation of journals and societies of philosophy of education, but more importantly in the work of Richard Peters, Israel Scheffler and others who recast philosophy of education in an analytical mould and brought educational philosophizing in the schools and departments of education of the English-speaking world into closer contact with mainstream philosophy.

This process of recasting the methods of philosophy of education had begun in the 1940s with the work of C.D. Hardie, but at mid-century what mostly passed for educational philosophy was a lengthy menu of 'educational philosophies' consisting of little more than restatements of a common stock of wisdom about schooling in the vocabularies of various 'schools' of thought. These pragmatist, idealist, realist and other philosophies of education were grounded in the conviction that a philosophy of education can be derived from the tenets of any

‘philosophical school’, a mistake whose continuing influence is evident in more recent attempts to develop existentialist, phenomenological, hermeneutical, postmodernist and other philosophies of education.

As it developed from the late 1950s to the 1970s, analytical philosophy of education was centrally concerned with the analysis of educational concepts such as ‘teaching’, ‘learning’, and ‘understanding’. A good example of this kind of work is Scheffler’s (1960) analysis of teaching as involving an intent to bring about learning, the use of instructional strategies reasonably calculated to engender it, and a manner respectful of the learner’s rationality. An extensive literature developed around this ‘standard thesis’, generating alternative analyses and accounts of quality in teaching, the boundaries between indoctrination and teaching, and the nature of learning. Some of this was illuminating and may have helped to clear away confusions and false assumptions about teaching in the literature of educational studies and reform, but conceptual analysis was a tool of limited efficacy in a domain lacking the sorts of metaphysical problems that Wittgenstein, Austin and Ryle had been at pains to resolve. The analysis of educational concepts went forward with neither the larger philosophical purposes of the formal and ordinary language movements themselves, nor the normative resources which the great educational theories of philosophy’s past had brought to bear on practice (see Analytical philosophy; Ordinary language philosophy, school of).

Work on this core of distinctly educational concepts had declined so substantially by the 1980s that one might reasonably judge that philosophy of education had lost its identity as a distinct sub-discipline and been reabsorbed into the parent discipline and its sturdier progeny. In any case it has been relatively inconspicuous in an era when other domains of practical philosophy have enjoyed spectacular growth and development, even as its scattered remains have been reinvigorated by the richer methods and bodies of work that have developed in ethics, political philosophy, informal logic and philosophy of the sciences, and by an outpouring of scholarship on the history of philosophy that is both philosophical and seriously historical. Philosophers of education had never ceased to engage the history of their field, but they are now in a position to do so more profitably.

The rich and largely untapped opportunities at hand are best pursued with a reasonable conception in view of the ways in which the philosophy of education can appropriately contribute to the larger enterprise of educational theory. This requires a model of the structure of a theory of education and of the place of philosophy of education within it.

Frankena (1965) and others have defended the view that philosophy of education must be regarded as principally an offshoot of ethics, and Hirst (1983) has formulated a plausible view of educational theory as a systematic body of practical principles generated, tested and justified by practice as much as through research in the disciplines. But a full account of the ways in which philosophy can contribute to the development of educational theory has yet to be synthesized. Broadly speaking, one can say that a fully elaborated theory of education would span education’s aims and means, its patients, practitioners and beneficiaries, its manner, mode and circumstances of provision, providing practical guidance which both explains and justifies what is to be done. The subdivisions of philosophy contribute to this enterprise through suggesting and illuminating the character of possible goals of education (see §§2-4), justifying educational aims and providing normative guidance for various aspects of the provision of education (§§5-7), and through guidance regarding assumptions of fact about learning, development and the character of specific subject domains (§§8-9).

2 Epistemology and the curriculum

Philosophical theories of the curriculum have generally been guided either by ethical or political-theoretic arguments about the proper aims and constraints of education (see §§5-6), or by an embrace of the concerns of one or another of the philosophical sub-disciplines or movements. Existentialists have proposed authenticity and freedom as the aims which should guide teaching and the curriculum, Marxists and critical theorists propose social reconstruction or the abandonment of hierarchically imposed curricula, feminists a more girl-centred curriculum (see Existentialism; Feminism; Marx, K.). Epistemology commends knowledge as a worthy object of instruction, and provides some basic guidance in how to make it one.

Peters (1966) held in his early writings that education consists in initiation into the forms of knowledge evolved by society, both to develop the mind and to transmit a valuable heritage. Dewey had rejected knowledge-transmission models of education in favour of an emphasis on experimentation, problem-solving and critical thinking rooted in

the child’s prior interests (see Dewey, J.); Peters argued in response that the disciplines themselves supply the various forms of critical thinking and problems to be addressed, and that it is the teacher’s job to lead children beyond their present interests to a love of knowledge. Hirst (1973) developed a similar view of *liberal* education, arguing that the development of a rational mind is inseparable from initiation into the forms of knowledge, and that this initiation should involve an ‘immersion’ in the concepts and methods of the various forms sufficient to yield an understanding of how each works and to allow the learner to experience the world through their distinctive structures.

Epistemology clearly lacks the wherewithal to warrant such a narrow view of the curriculum, as Peters later recognized, but it may still provide guidance for education to the extent that knowledge or the development of an ability to acquire it is an acknowledged aim. Scheffler (1965) has perhaps made the most of what analytical epistemology has to offer, by developing the distinction between knowing-that and knowing-how, and examining the implications for teaching of the *belief* and *evidence* conditions for knowing-that. The former demands that classrooms engage students’ beliefs, rather than merely train them to give the right answers; the latter demands that they encourage internalization of the methods and forms of evidence by which knowledge is generated if students are to be epistemically autonomous.

Most of the recent literature has taken its cue not so much from philosophical epistemology as from a variety of movements built on unwitting adherence to remnants of logical positivism, confusion over the significance of debates about foundationalism, and epistemological scepticism rooted in theories of ideology and the incommensurability of paradigms (see §8). The dominant concern is that authority in school and society has rested illicitly on claims of epistemic authority.

3 Logic and critical thinking

Logic would seem to commend rationality as an educational goal or ideal, much as epistemology does, and to suggest that there are skills of reason or rational inference that can be isolated and taught independently of other aspects of the curriculum. Since the 1960s, however, there has developed a large body of work on critical thinking predicated on the inadequacy of formal deductive logic as an account of quality in inference, argumentation and reasoning. Formal logic has thus been largely set aside as a model for the teaching of critical thinking, though something of its spirit has been preserved in the idea of subject-neutral principles of reason assessment advocated even by those who regard critical thinking as no less epistemic than logical. This critical thinking movement has produced not only one of the more important streams of theoretical analysis undertaken in recent philosophy of education, but also innovations in curriculum and evaluation procedures at all levels of instruction.

Debate about the nature, teaching and evaluation of critical thinking was strongly influenced through the 1960s and 1970s by the work of Ennis (1962), who first conceived of critical thinking as correctly assessing statements, later as thinking that is reasonable, reflective and concerned with assessing not only beliefs, but also actions. His focus on discrete skills of reasoning has drawn a variety of criticisms.

McPeck (1990) has argued, against not only Ennis but the critical thinking movement generally, that there is no significant body of general critical thinking skills which can be taught. Citing Hirst, he argues that the subject-specific forms of inquiry we engage in have too little in common for us to extract a common logic which can be profitably taught as a subject in its own right.

Other philosophers have rebutted McPeck’s grounds for denying the existence of a robust set of subject-neutral critical thinking skills, while also arguing that there is more to being a critical thinker than possessing such skills. Paul (1982) has insisted that the mastery of atomic thinking skills only makes one a ‘weak sense’ critical thinker with the ability to criticize the views of others and reinforce one’s own. Exchanges of arguments involve clashes between opposing perspectives or world views, so to be a ‘strong sense’ critical thinker involves being able to take a sympathetic view of perspectives and world views one does not hold, he argues. Resisting the potential relativism in Paul’s view (see Relativism), Siegel (1988) has argued that it is ‘atomistic’ criteria of appraisal that we must rely on in judging worldviews. He holds that a critical thinker is one who is appropriately moved by reasons, one who possesses not only the abilities of rational assessment but a willingness, desire and disposition to follow reason. Lipman (1991) has argued that philosophy itself affords the best training in reasonableness and judiciousness guided by the ideal of rationality, and that it requires that we convert classrooms at all grade levels

into 'communities of inquiry' through conversation governed by norms of reasonableness and mutual respect.

These developments lay useful groundwork for an integration of work in philosophy of education with broader philosophical investigations of the nature of rationality and good judgment.

4 Aesthetics, perception and creativity

The precarious position of the arts and the aesthetic in contemporary education has lent urgency to questions about the aims and value of education in the arts. Philosophers of education have sought to answer these questions through engagement with philosophical aesthetics and the philosophy of art and literature. Their primary concern has been to elaborate a compelling vision of the value of a proper education in the arts; their assertions of value have rested sometimes in direct appeals to the quality of human experience, but more often in accounts of the contributions of the aesthetic to the cognitive, practical and productive aspects of human existence.

In work first published in the 1970s, Scheffler held that aesthetic experience is not simply emotional but also cognitive, and that in interpreting both art and the situations which confront us in life our feelings guide our perceptions (see (Scheffler 1991). Greene 1995) and Nussbaum (1990) have taken related lines of thought farther in arguing that education in the arts develops a perceptiveness which equips one well for life. The former has argued that the visual and performance arts embody distinct representational symbol systems whose mastery 'releases' the learner's perception and imagination, thereby liberating both judgment and action. The latter has brought an Aristotelian account of the development of virtue and practical judgment in defence of literary art, and the novel especially, for its value in cultivating the capacity to discern and respond to what is morally salient in concrete situations.

Philosophical interest in the productive and performance aspects of arts instruction has focused on the nature of creativity and how best to promote it. In opposition to the widespread belief that creative and rational thought are mutually exclusive, the tendency of recent philosophical work, such as Bailin (1988), has been to view significant achievement as the measure of creativity, and productive imagination as resting on knowledge and skill.

5 Ethics

Ethics holds a central place in philosophy of education through its role in justifying not only the aims and priorities of education, but education itself. Peters ((1966) shaped the course of debate with an argument for eschewing talk of aims and a non-instrumental justification of education, but by the 1970s the landscape was crowded with 'de-schoolers' and child liberationists who denied that imposing education on children is justified. The liberationists argued that children must be assigned the same liberty rights as adults, and held that the consequences of doing so would be benign. Purdy 1992) has refuted this most effectively, with a review of the evidence against the liberationists' optimistic assumptions about children's rationality and development in the absence of constraint, and with a defence of the legitimacy of moral education in government-run schools. She, Callan (1988) and others have nevertheless defended the importance of autonomy as a central goal of education, arguing in various ways that the development of autonomy is promoted by upbringing and teaching which nurture self-control, critical thinking, self-examination and related traits. Others have appealed to children's rights and wellbeing, the duties of adults and other forms of moral premises in defending educational aims ranging from various conceptions of the individual student's good to promotion of environmental responsibility.

R.M. Hare stimulated interest in moral education with his view, first adumbrated in *The Language of Morals* (1952), that what is important is that children should learn the *form* of morality so that they may find and embrace a set of moral principles of their own. Other prominent moral theorists followed his lead, generating a debate largely defined by the tensions between Kantian and Aristotelian views (see Kantian ethics). Kurt Baier (1973) and others took positions sympathetic to Lawrence Kohlberg's conception of moral development as a sequence of stages toward mature moral reasoning. Resisting both 'moral indoctrination' and treating the child's intuited 'values' as authoritative, they advocated instruction in moral reasoning that would enable a child to become morally autonomous or able to identify and apply rationally acceptable moral principles. Taking a more Aristotelian view, Pincoffs ((1986) defended moral habituation from the charge of indoctrination with the argument that it is only through such habituation that one can perceive and appreciate the moral qualities of things (see Aristotle §§21-3). He and Nussbaum have developed virtue-centred accounts of moral education which regard

the development of perception, affect and judgment as interrelated. Recent work has sought to join such approaches with accounts of care, trust and the psychosocial foundations of moral learning.

Educational ethics has become increasingly visible as a branch of professional and practical ethics not unlike medical or business ethics. This is particularly true of the ethics of higher education, which has flourished in eras of crisis in the universities such as the 1960s and 1990s. The topics that dominate current debate include curricular multiculturalism; university neutrality in the face of injustice or political controversy; university-business partnerships and other arrangements that may compromise academic integrity; the proper management of student life in the face of sexual harassment and violence, alcohol and drug abuse, and racism. Multiculturalism has also become an important focus of debate in the ethics of primary and secondary education, joining such topics as the ethics of student evaluation, discipline and classroom management, and the ethics of educational research (see Multiculturalism).

6 Political philosophy

Political philosophy has been a perennial source of reflection on the significance of education for civic life, and on the proper role of the state in overseeing and sponsoring education. The classical tradition took education to be an important instrument of statesmanship and civic well-being, while liberals such as John Stuart Mill regarded state sponsorship of education as a threat to liberty. Modern liberal democracies have neither fully accepted nor fully rejected either of these views in practice, nor have they fully accepted or rejected the liberal egalitarian idea that equal educational opportunity is essential to social justice. The tensions and tradeoffs involved in these compromised commitments generate ongoing debate, and provide an important context for assessing the import of the ethical endorsements of education for autonomy noted in §5.

A ‘liberal education’ is defined by ‘comprehensive liberals’, such as Ackerman 1980), as one that would provide children with the resources to develop their own self-conceptions and choose their own courses in life. This requires that it be nondirective with respect to substantive conceptions of the good, and promote autonomy grounded in rational scrutiny of the alternatives. ‘Political liberals’, such as Macedo (1995), argue in response that this requirement is not easily reconciled with liberalism’s own acceptance of diversity. Can one coherently endorse individual autonomy while denying parents the right to bring their children up in their own traditions and religion? Another line of argument which rejects comprehensive liberalism concerns itself with a perceived decline in civic commitment and a sense of sharing in a common good and belonging to a common social order (see Community and communitarianism; Liberalism).

Gutmann’s liberal democratic view of education (1987) is built on an uneasy alliance between this communitarian view and comprehensive liberalism. She holds that democracy requires that communities be able to shape their futures through democratic control of education, subject to the limits set by principles which demand, among other things, that all children be enabled through education to participate ‘effectively’ in democratic deliberations. In the face of widely divergent views of the prospects for creating an ‘educated public’ able to rationally engage in such deliberations, MacIntyre (1987) has derived an account of the conditions that are required, from a historical instance in which they were arguably present.

Marxist critiques of schooling and liberal meritocracy begin by noting the persistence of inequality, and historical connections between the rise of government-run school systems and the factory system. They conclude that the perpetuation of inequality has been essential to the role of schools in capitalist societies, their function being to sort students for different economic positions, in a way that masks the intergenerational transmission of class behind an ideology of individual effort and desert. Many variants of this view have developed around different conceptions of the assets which schools provide to those bound for economic and social privilege, but their recommendations for practice seem invariably to follow Gramsci’s vision of offering whatever resistance to the dominant ideology one can, in whatever institutional setting one can (see Gramsci 1971).

The most recent work in the Marxist tradition attempts to overcome its preoccupation with class stratification, by emphasizing the importance of other dimensions of social cleavage and oppression, such as race, gender and sexual orientation. A related feminist criticism of liberal educational theory and reform, which bears importantly on the care and upbringing of children, is that they have opened up greater opportunity for women to choose traditionally male paths in life, but have done little to elevate domestic work or prepare anyone to do it well (see

Feminist political philosophy).

7 Philosophy of law and education policy

In philosophy of education there are two kinds of work which involve a direct concern with law and draw significantly upon work in the history and philosophy of law. One addresses the general relationships between law and education, the other treats educational issues as problems of public policy whose resolution may be aided by legal argument and reform.

Ancient Greek thought was acutely attuned to questions about the relationships between law and education, and understood both that compliance with law must rest for the most part in an informed acceptance of its demands, and that the laws themselves and all the institutions of a city are educative and must be so regarded to ensure civic well-being. The re-evaluation of these views is evident in Curren's examination of the role of education in the foundations of just law (1995b), in connection with the ideal of rule by consent, and in the work of Sandel (1996) and other communitarians who have reasserted the claim that national policy and law should be shaped by a sensitivity to their educative import. This latter development is particularly important for educational philosophy, inasmuch as there are compelling reasons to think that schooling is not the whole of education, and that its prospects for success are limited by the broader educational forces in a society.

Education is an important arena of public policy debate, and an appropriate though underdeveloped object of philosophical public policy studies. Some of the most important topics of current discussion are educational equality, freedom of speech and religion in government-run education, proposals to create arrangements which would allow parents a choice of schools at the public expense, and school safety and security. These debates bring a consideration of common law principles and fundamental constitutional rights to bear on some of the issues at stake in debates about liberal and democratic education.

Debate over educational equality has focused on alternative conceptions of its form, the educational goals it should be concerned with, and the principles which demand it. Is full equalization the appropriate aim, for instance, or is the relative equality inherent in assuring that all students reach some specified threshold acceptable? Or should one only be concerned, as the courts of the USA are, to ensure that race, native language, learning disabilities and other 'suspect' factors are not barriers to equal educational opportunity? With regard to the justification of equality (see Equality), does the principle of 'equal protection' in the free enjoyment of fundamental political rights provide grounds for insisting on equality, as some, including Gutmann (1987) and Curren (1995a), have argued?

An important line of argument regarding children's rights begins with the conviction that the free speech rights of adults are meaningless if the state can control through education what one thinks as a child. If this is correct, then one might conclude that inculcation of community values beyond a core of essential civic commitments is a breach of fundamental rights. One might also conclude, as Arons (1986) does, that taking freedom of speech and religion seriously requires that parents be provided with a financially feasible choice among schools. An important issue in the philosophy and policy of higher education at present, is how constitutional protection of free speech should be understood to apply to attempts to draft speech codes to suppress racist 'hate speech'.

A debate which invokes common-law principles, rather than constitutional ones, is how we should understand the framework of fundamental legal principles bearing on the school's responsibility for the safety of students. How should one regard the elements of sovereign immunity to liability still present in the legal systems of common-law countries, given the role of compulsory attendance laws in placing children in a position of danger, if school personnel fail to provide adequate security against assault by third parties?

8 Philosophy of science

In order to teach a subject with sensitivity to its epistemic structure, teachers must grasp and be guided by the same kind of understanding which philosophers of science, mathematics, history or economics aim at - that is, a secure (if nontechnical) sense of the methods of discovery or investigation, of the methods and logic of confirmation, and of the forms of explanation and theory which define it as a mode of inquiry, and they must attend to these in both teaching and evaluating students. Philosophy of science in particular has been called into service in efforts to reshape pedagogy, curriculum and methods of assessment, though the lessons drawn could be applied in the teaching of other subjects as well.

An international ‘crisis’ in science education was declared in the 1980s in a series of reports citing dismal levels of both learning and interest in science, and further concern about the inadequacy of science instruction has been generated since then by neo-romantic attacks on science by postmodernists and some feminists. Government calls for revision of science curricula have become more receptive to the idea that students should not only learn science, but learn *about* science, giving encouragement to the idea that introducing history and philosophy of science into the science curriculum would give students not only a more accurate appreciation of science as an enterprise, but a far more engaging one as well.

Two important related topics in current debates about science instruction are constructivism and multiculturalism. Constructivists promote ‘discovery’ learning in science in the belief that scientific thinking develops naturally in children, a belief rooted in a form of empiricism according to which science finds order in experience, and (in the radical versions of constructivism) has access to nothing beyond immediate experience. Versions of relativism originating in selective readings of Kuhn’s work have also become pervasive in educational theory and found their way into proposals for multicultural curricula which would present ethno-sciences as no less rationally confirmed than Western science. Matthews (1994) and others have responded with observations about the distinctiveness of Western science, reminders of Kuhn’s insistence that science *is* rational (though not in a way that can be adequately represented as rule following), and by identifying the empiricist and positivistic premises relied upon by constructivists, even as they denounce positivism (see Kuhn, T.).

Philosophers of education have also made use of work in philosophy of science in offering methodological critiques of social science research in education, and have relied on models of explanation drawn from action theory and philosophy of the social sciences in evaluating and offering alternatives to the explanatory models underlying educational critiques and policy proposals. Green (1980) has been among the most ambitious in this regard in developing what amounts to a rational-choice model of the dynamics of educational systems.

9 Philosophy of mind, language and psychology

It is widely assumed within the educational research community that psychology, cognitive science, psychometrics and social learning theory have largely displaced philosophy from its historical role in formulating theories of the mind, language and human development, and determining what implications (if any) those theories have for educational practice (see Cognitive development; Learning). Nevertheless, at least two roles related to this empirical aspect of educational theory remain appropriate for philosophy, one foundational and critical, the other interpretative and synthetic.

Gilbert Ryle gave encouragement in *The Concept of Mind* to the possibility of a ‘philosophy of learning’, but the analyses of mental concepts which this inspired in the 1950s and 1960s have given way more recently to interest in the foundations and adequacy of the bodies of psychological and measurement research which educational researchers have looked to for guidance in the conduct of educational practice. Behaviourist learning theory and cognitive and moral stage theories have been subjected to critical scrutiny by a number of philosophers. Wren (1982) has examined social learning theory and found its accounts of self-regulation inadequate. Block and Dworkin (1974) have produced a notable critique of ‘general intelligence’ as a theoretical construct and identified deficiencies in the validation of IQ tests as a measure of it. Norris (1995) and others have contributed in a variety of ways to conceptual, experimental and development efforts pertaining to validity, test formats and the measurement of critical and ‘higher-order’ thinking in classroom and standardized tests of achievement.

Even when psychological findings can be accepted at face value, however, their relevance to educational practice can only be settled by recourse to a theory of education concerned with the kinds of persons we should want students to become. A psychological theory is not itself such a theory, despite the recent dominance of psychology in educational studies. There remains the interpretive and synthetic task of building a cogent educational theory, drawing on whatever facts and practical know-how may be available, which a philosopher is as well-equipped as anyone to undertake. It is a task as important as it is difficult, and if the results of past efforts often fall short, both philosophically and practically, that is all the more reason for capable philosophers to turn their attention to it.

See also: Education, history of philosophy of

RANDALL R. CURREN

References and further reading

Ackerman, B.A. (1980) *Social Justice in The Liberal State*, New Haven, CT: Yale University Press.(Referred to in §6. Includes a 'comprehensive liberal' theory of education.)
Arons, S. (1986) *Compelling Belief*, Amherst, MA: University of Massachusetts Press. (Referred to in §7. A good introduction to issues of free speech, religion and values, and parental versus community control of education.)
Baier, K. (1973) 'Moral Autonomy as an Aim of Moral Education', in G. Langford and D.J. O'Connor (eds) *New Essays in the Philosophy of Education*, London: Routledge & Kegan Paul.(Referred to in §5. Defence of a Kantian view by a leading moral theorist.)
Bailin, S. (1988) *Achieving Extraordinary Ends: An Essay on Creativity*, Dordrecht: Kluwer Academic Publishers. (Referred to in &4. An essay on the nature and teaching of creativity.)
Block, N. and G. Dworkin (1974) 'IQ, Heritability and Inequality' pts I and II, *Philosophy and Public Affairs* 3 (4): 331-409; 4 (1): 40-99.(Referred to in §9. Combines philosophy of psychology and ethics.)
Burgess, R.G. (ed.) (1989) *The Ethics of Educational Research*, New York: The Falmer Press.(Topics in educational research ethics.)
Cahn, S. (ed.) (1990) *Morality, Responsibility, and the University*, Philadelphia, PA: Temple University Press. (Essays in the ethics of higher education.)
Callan, E. (1988) *Autonomy and Schooling*, Montreal, Que.: McGill-Queen's University Press.(Referred to in §5. A defence of autonomy as an aim of education.)
Cooper, D.E. (1986) *Education, Values and Mind: Essays for R.S. Peters*, London: Routledge & Kegan Paul. (Provides a good picture of Peters' formative role in the field and a bibliography of his work.)
Curren, R. (1995a) 'Justice and the Threshold of Educational Equality', in M. Katz (ed.) *Philosophy of Education 1994*, Urbana, IL: Philosophy of Education Society.(Referred to in §7. On the form of equality, problems in Gutmann's view, an alternative to hers linked to Curren 1995b.)
Curren, R. (1995b) 'Punishment and Inclusion', *Canadian Journal of Law & Jurisprudence* 8 (2): 259-74. (Referred to in §7. On the role of education in the foundations of law.)
Curren, R. (1995c) 'Coercion and the Ethics of Grading and Testing', *Educational Theory* 45 (4): 425-41.(A moral argument for changes in both standardized and classroom testing.)
Ennis, R. (1962) 'A Concept of Critical Thinking', *Harvard Educational Review* 32 (1): 81-111.(Referred to in §3. An influential conception of critical thinking and its role in education.)
Frankena, W. (1965) *Philosophy of Education*, New York: Macmillan.(Referred to in §1. An account of the structure of a philosophy of education derived from J.S. Mill, and a sampling of twentieth-century theories.)
Fuller, T. (ed.) (1989) *The Voice of Liberal Learning: Michael Oakeshott on Education*, New Haven, CT: Yale University Press.(A prominent conservative's major writings on education.)
Fullinwider, R. (ed.) (1996) *Public Education in a Multicultural Society*, Cambridge: Cambridge University Press.(On the philosophical issues pertaining to multicultural schooling.)
Gramsci, A. (1971) *Selections From the Prison Notebooks*, London: Lawrence & Wishart. (Referred to in §6. A classic of twentieth-century Marxism.)
Green, T.F. (1980) *Predicting The Behavior of the Educational System*, Syracuse, NY: Syracuse University Press. (Referred to in §8. Develops an explanatory model inspired by Anscombe's work on the explanation of action.)
Greene, M. (1995) *Releasing the Imagination: Essays on Education, the Arts, and Social Change*, San Francisco, CA: Jossey-Bass Publishers.(Referred to in §4. Essays by a leading US proponent of aesthetic education.)
Gutmann, A. (1987) *Democratic Education*, Princeton, NJ: Princeton University Press. (Referred to in §§6-7. An influential work and useful guide to debates in education politics and policy.)
Hardie, C.D. (1942) *Truth and Fallacy in Educational Theory*, Cambridge: Cambridge University Press.(Referred to in §1. A pioneering work in analytical philosophy of education.)
Hare, R.M. (1952) *The Language of Morals*, Oxford: Clarendon Press.(Referred to in §5. A major meta-ethical work of the mid-twentieth century.)
Hare, R.M. (1992) *Essays on Religion and Education*, Oxford: Clarendon Press.(Essays on education and religion by a leading moral theorist.)
Hirst, P.H. (1973) 'Liberal Education and the Nature of Knowledge', in R.S. Peters (ed.) *The Philosophy of Education*, Oxford: Oxford University Press.(Referred to in §2. An influential theory of the curriculum. The volume is a landmark in the emergence of the field and contains a helpful discursive bibliography.)
Hirst, P.H. (1983) 'Educational Theory', in *Educational Theory and Its Foundation Disciplines*, London: Routledge & Kegan Paul.(Referred to in §1. On the character and development of educational theory.)
Hoekema, D. (1994) *Campus Rules and Moral Community: In Place of In Loco Parentis*, Lanham, MA: Rowman

& Littlefield.(On the ethics of college student life.)
Kaminsky, J.S. (1993) *A New History of Educational Philosophy*, London: Greenwood Press. (A controversial but informative history of educational philosophy as a field of study in the education schools of Australasia, the UK and the USA.)
Kleinig, J. (1982) *Philosophical Issues in Education*, London: Croom Helm.(A good survey of issues with an excellent bibliography.)
Lipman, M. (1991) *Thinking In Education*, Cambridge: Cambridge University Press. (Referred to in §3. Good judgment as an aim of education, and Lipman's community of inquiry model for teaching it.)
Macedo, S. (1995) 'Liberal Civic Education and Religious Fundamentalism: The Case of God v. John Rawls?' *Ethics* 105 (3): 468-96.(Referred to in §6. Lead article in a symposium on citizenship, democracy and education.)
MacIntyre, A. (1987) 'The Idea of an Educated Public', in G. Haydon (ed.) *Education and Values*, London: Institute of Education, University of London.(Referred to in §6. Identifies the requirements for creating an educated public.) Martin, J.R. (1994) Changing the Educational Landscape: Philosophy, Women, and Curriculum, London: Routledge. (Essays in feminist philosophy of education.)
Matthews, M.R. (1994) *Science Teaching: The Role of History and Philosophy of Science*, London: Routledge. (Referred to in §8. The best piece of work on science and education to date.)
McPeck, J. (1990) *Teaching Critical Thinking*, New York: Routledge.(McPeck's essays and an exchange between him and other leading figures.)
Norris, S. (1995) 'Format Effects on Critical Thinking Test Performance', *Alberta Journal of Educational Research* 41 (4): 378-406.(Referred to in §9. A philosopher's multidisciplinary work on alternative measures of performance in critical thinking.)
Nussbaum, M. (1990) *Love's Knowledge: Essays on Philosophy and Literature*, Oxford: Oxford University Press. (Referred to in §§4-5. Essays on moral learning, the place of literary art in education and related topics.)
Passmore, J. (1980) *The Philosophy of Teaching*, London: Duckworth.(A comprehensive philosophy of pedagogy.)
Paul, R.W. (1982) 'Teaching Critical Thinking in the "Strong" Sense', *Informal Logic Newsletter* 4 (2): 2-7. (Referred to in §3. A concise statement of Paul's view of critical thinking.)
Peters, R.S. (1966) *Ethics and Education*, London: Allen & Unwin.(Referred to in §§2 and 5. Major statement of his early view, progressively modified in later works. For a list of his works see Cooper 1986.)
Phillips, D.C. (1995) 'The good, the bad, and the ugly: The many faces of Constructivism', *Educational Researcher* 24 (7): 5-12.(A survey of the varieties of constructivist learning theory.)
Pincoffs, E. (1986) *Quandries and Virtues*, Lawrence, KS: Kansas University Press. (Referred to in §5. Develops an account of moral education within a virtue-ethical framework.)
Purdy, L. (1992) *In Their Best Interests? The Case Against Equal Rights For Children*, Ithaca, NY: Cornell University Press.(Referred to in §5. A feminist response to liberationist theories of education.)
Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson University Library.(Referred to in §9. A major work of the mid-twentieth century, and a significant influence on the course of philosophy of education.)
Sandel, M. (1996) *Democracy's Discontent*, Cambridge, MA: Harvard University Press. (Referred to in §7. Takes a broad interdisciplinary view of the educative significance of public institutions and policies.)
Scheffler, I. (1960) *The Language of Education*, Springfield, IL: Charles C. Thomas. (Referred to in §1. An influential work in the analytic mode.)
Scheffler, I. (1965) *Conditions of Knowledge: An Introduction to Epistemology and Education*, Glenview, IL: Scott, Foresman & Co.(Referred to in §2. Remains a good overview of the lessons of philosophical epistemology for education, even though its references are dated.)
Scheffler, I. (1991) *In Praise of the Cognitive Emotions*, New York: Routledge.(Referred to in §4. Scheffler's essays in philosophy of education from the 1970s and 1980s.)
Sellers, M. (1994) *An Ethical Education*, Oxford: Berg Publishers.(Papers on the curriculum and aims of the university, campus speech and affirmative action.)
Siegel, H. (1988) *Educating Reason: Rationality, Critical Thinking and Education*, New York: Routledge. (Referred to in §3. An important account of critical thinking as a goal of education.)
Strike, K. (1989) *Liberal Justice and the Marxist Critique of Education*, New York: Routledge.(A useful guide to Marxist theories of education.)
Wren, T. (1982) 'Social Learning Theory, Self-Regulation, and Morality', *Ethics* 92 (3): 409-24.(Referred to in §9. Lead article in a symposium on moral development.)

Edwards, Jonathan (1703-58)

Jonathan Edwards' work as a whole is an elaboration of two themes - God's absolute sovereignty and the beauty of his holiness. God's sovereignty is articulated in several ways. Freedom of the Will (1754) defends theological determinism. God is the complete cause of everything that occurs, including human volitions. Edwards is also an occasionalist, idealist and mental phenomenalist. God is the only real cause of events. Human volitions and 'natural causes' are mere 'occasions' upon which God produces the appropriate effects. Physical objects are collections of sensible 'ideas' of colour, shape, solidity, and so on, and finite minds are collections of 'thoughts' or 'perceptions'. God's production of sensible ideas and thoughts in the order which pleases him is the only 'substance' underlying them. God is thus truly 'being in general', the 'sum of all being'.

The beauty or splendour of God's holiness is the principal theme of two late works - End of Creation *and* True Virtue *(both published posthumously in 1765). The first argues that God's end in creation is the external manifestation of his internal splendour. That splendour primarily consists in his holiness and its most perfect external expression is the holiness of the saints, which mirrors and depends upon it.* True Virtue *defines holiness as 'true benevolence' or 'the love of being in general' ([1765b] 1957-, vol. 8: 546), and distinguishes it from such counterfeits as rational self-love, instincts like parental affection and pity, and natural conscience. Since beauty is defined as 'agreement' or 'consent' and since true benevolence consents to being in general, true benevolence alone is truly beautiful. Natural beauty and the beauty of art are merely its image. Only those with truly benevolent hearts, however, can discern this beauty.*

Edwards' projected History of Redemption would have drawn these themes together, for it is in God's work of redemption that his sovereignty, holiness and beauty are most effectively displayed.

1 Life

Edwards was born into a family of Congregationalist clergymen in East Windsor, Connecticut in 1703. He entered Yale in 1716, where he read Locke and Newton and began 'The Mind' (1829a) and 'Natural Philosophy' (1829b). In 1725 the church in Northampton chose Edwards to succeed his grandfather, Solomon Stoddard. The most notable events of his tenure were the revivals of 1734 and 1740, the latter of which came to be known as the 'Great Awakening'. Edwards' defence of the revivals and criticisms of their excesses culminated in his first major treatise, the *Religious Affections* (1746). Worsening relations with his congregation came to a head in a dispute over qualifications for church membership, with Edwards rejecting the less rigorous standards of his grandfather and insisting on a public profession of saving faith. He was dismissed in 1750 by a margin of one vote. Refusing invitations to pulpits in North America and Scotland, Edwards retreated to the Indian mission at Stockbridge, where he completed his last major works - *Freedom of the Will* (1754), *Original Sin* (1758), *End of Creation* and *True Virtue* (both published posthumously in 1765). Edwards accepted an appointment as President of the College of New Jersey (now Princeton) in 1757. He died from smallpox in 1758, less than four weeks after his arrival.

2 Theological determinism

Edwards believed that indeterminism is incompatible with our dependence on God and hence with his sovereignty. If our responses to God's grace are contra-causally free, then our salvation partly depends on us and God's sovereignty is not 'absolute and universal'. *Freedom of the Will* is Edwards' defence of theological determinism. He first attempts to show that indeterminism is incoherent. For example, he argues that by 'self-determination' the indeterminist must mean either that one's actions (including one's acts of willing) are preceded by an act of free will or that one's acts of will lack sufficient causes. The first leads to an infinite regress while the second implies that acts of will happen accidentally. Edwards also contends that indeterminism is inconsistent with ordinary moral concepts. If, for instance, the necessity of sinning wholly excuses, then a bias to sin should partially excuse. But it does not; a person who acts from settled habits of maliciousness is deemed 'so much the more worthy to be detested and condemned' ([1754] 1957-, vol. 1: 360). Since indeterminism implies that necessity excuses, it is inconsistent with the way we attribute blame.

Edwards' principal reasons *for* theological determinism are God's sovereignty, the principle of sufficient reason (which requires that everything that begins to be have a complete cause), and God's foreknowledge (see

Omniscience §3). God's foreknowledge is discussed at length. Call a decision I make at time t_n 'D'. God is omniscient and has therefore always believed that D occurs at t_n. Since God cannot be mistaken, the occurrence of D at t_n is entailed by his believing at some earlier time t_{n-m} that D occurs at t_n. But God's forebelief is past (in relation to t_n); therefore, it is 'now necessary' in the sense that nothing can be done at t_n to alter it. Now what is entailed by a necessary fact is itself necessary. Hence, D could not fail to occur at t_n. The conclusion cannot be evaded by appealing to God's timelessness and denying that God's forebeliefs precede their objects. For even though God's forebeliefs may be timeless, divinely inspired prophecies are not. Yet divinely inspired prophecies are necessarily connected with the acts of will they foretell and are clearly past (and hence necessary) in relation to them. *Freedom of the Will* concludes by arguing that necessity is not incompatible with the ascription of praise and blame or with moral responsibility. We are responsible for our actions in the ordinary sense when we do what pleases us, and determinism does not deny that we often act as we please. And though God and Christ necessarily act for the best, their actions are eminently praiseworthy.

3 Idealism, mental phenomenalism, occasionalism and views on identity

Edwards' idealism, mental phenomenalism and occasionalism provide a philosophical interpretation of the doctrine of God's absolute sovereignty. If Edwards is correct, God is the only true substance and the only true cause.

True or real causes meet three conditions: (1) they cannot exist at different times or places from their effects; (2) they are total causes, in that nothing else is needed for their effects; and (3) they necessitate their effects. Now God is spatially and temporally omnipresent; nothing is separated from him by spatial or temporal distance. Furthermore, his activity is a necessary and fully sufficient condition for the occurrence of any spatiotemporal effect. Finally, God's will is necessarily effective. His will, therefore, is a true cause. Nothing else meets these conditions. Hence, God is the only real cause. Apparent causes (for example, striking a match) are simply the *occasions* upon which God produces effects (for example, the match's ignition) according to 'established methods and laws' (see Occasionalism).

In an early paper ('Of Atoms') Edwards argued that the concept of a material substance underlying solidity, resistance and other physical properties is either empty or denotes God's causal activity. If material substance were real, it would be something which 'subsisted by itself, and stood underneath and kept up solidity and all other [physical] properties' ([1829b] 1957-, vol. 6: 215). God alone meets these conditions and is therefore the only 'substance' underlying physical properties. Deploying arguments analogous to (but apparently uninfluenced by) Berkeley's, Edwards further argues that a 'body is nothing but a particular mode of perception' and that 'the material universe exists only in the mind' ([1829a] 1957-, vol. 6: 368; 1957-, vol. 6: 398) (see Berkeley, G. §§3, 6-7).

Edwards' mental phenomenalism is a natural extension of his occasionalism and views on material substance. God is the only real cause of 'thoughts' or 'perceptions'. If a substance is what 'subsists by itself', 'stands underneath' and 'keeps up' a set of properties, then the concept of mental substance too is either empty or denotes God's causal activity. A mind, therefore, 'is nothing but a composition and series of perceptions [mental events]… connected by… laws' ([1829a] 1957-, vol. 6: 398). Mental and physical substance are thus God's producing mental events and their objects (sensible ideas or 'sensations') 'according to… methods and laws' which he has freely established ([1829a] 1957-, vol. 6: 344).

God's sovereignty extends to criteria of identity. 'Species' (kinds or natures) are the ways we classify things. But our classifications depend on our needs and interests, and on the character of the world we live in. In determining these things, God has thus determined what counts as a 'species' or kind. Now criteria of identity are determined by a thing's nature. God, therefore, is the ultimate ground of these criteria (and hence can if he wishes treat Adam and his posterity as one thing, holding the children responsible for the sin of their progenitor).

4 God as being in general

As the only true substance and only true cause, God is 'being in general'. He 'is the sum of all being and there is no being without His being. All things are in Him and He in all' (1955: 87). Edwards does not mean that God is the power of being or being as such rather than an omniscient mind and omnipotent will. God 'is properly' a

necessarily existing 'intelligent willing agent, such as our souls, only without our imperfections, and not some inconceivable, unintelligent, necessary agent' ([1957] 1957-, vol. 13: 452).

True Virtue associates being with capacity or power and asserts that 'degree of existence' is a function of 'greater capacity and power' ([1765b] 1957-, vol. 8: 546). 'Miscellany 94' (1890) assimilates perfect entity and perfect activity. 'An Essay on the Trinity' (1903) identifies God's perfect activity with his outflowing love or holy will. Other passages identify entity with mind or 'perceiving being'. The drift of these scattered observations is an identification of being with mind in action and of degree of being with degree of consciousness and the comparative perfection of the activity in which it is engaged. Because God's power and consciousness are unlimited, so too is his being.

In speaking of God as 'being in general', Edwards means more than that finite beings are totally dependent on God for their existence and properties. Because he is the only true substance and only true cause, created beings are no more than God's 'images' or 'shadows'. God is the 'head' of the system of beings, its 'chief part', an absolute sovereign whose being, power and perfection are so great 'that the whole system of created beings, in comparison of Him, is as the light dust of the balance' (1955: 142). 'Being in general' thus refers to the system of beings - principally to God but to 'particular beings' as well in so far as they depend upon and reflect him.

5 God's end in creation

End of Creation defines God's glory as 'the emanation and true external expression of God's internal glory and fullness'. It includes (1) 'the exercise of God's perfections to produce a proper effect', (2) 'the manifestation of his internal glory to created understandings', (3) 'the communication of the infinite fullness of God to the creature' and (4) 'the creature's high esteem of God, love to God, and complacence and joy in God; and the proper exercises and expressions of these' ([1765a] 1957-, vol. 8: 527).

The first and third 'parts' of God's glory are not ontologically distinct since the principal effect of God's exercise of his perfections is 'his fullness communicated'. And the third includes the second and fourth, as God's internal fullness or glory is the 'fullness of his understanding consisting in his knowledge' of himself 'and the fullness of his will consisting in his virtue and happiness'. God's 'external glory… consists in the communication of these', that is, in bringing it about that 'particular minds' know and love God and delight in him. One through four are thus 'one thing, in a variety of views and relations' ([1765a] 1957-, vol. 8: 527).

Edwards never doubted that God's end or aim must be that which is supremely good and excellent, namely himself. But he also thought that the essence of goodness is the communication of good for its own sake and that this implies that the creature's happiness is God's ultimate end. *End of Creation* reconciles these convictions by including human happiness (which consists in knowing and loving God, and rejoicing in him) in God's ultimate end, namely the communication of his internal glory '*ad extra*'.

Now God is primarily glorious for his holiness and his will is holy because it exhibits true benevolence - the love of being in general in which 'true virtue' consists. The communication of his internal glory therefore principally consists in the holiness of the saints (the elect), which is identical with their love of being in general, that is, of God and their neighbour.

6 Ethics

Edwards' reflections on ethics culminate in *True Virtue* (1765b). True virtue's aim is the general good. Those who love the general good, however, also prize the disposition that promotes it. Truly benevolent people thus love two things - being and benevolence. But truly virtuous people not only value benevolence because it promotes the general good; they also relish it for its own sake. Hence, while virtue 'most essentially consists in benevolence to Being in general' ([1765b] 1957-, vol. 8: 540), there is a wider sense in which it includes not only benevolence but also a delight or 'complacence' in benevolence's intrinsic excellence or beauty.

It is God that is 'infinitely the greatest being' and 'infinitely the most beautiful and excellent'. True virtue thus principally consists 'in a supreme love to God, both of benevolence and complacence' ([1765b] 1957-, vol. 8: 550-1). It follows that 'a determination of mind to union and benevolence to a *particular person or private system* [whether one's self, one's family, one's nation or humanity], which is but a small part of the universal system of

being… is not of the nature of true virtue' unless it is dependent on, or 'subordinate to, benevolence to Being in general' ([1765b] 1957-, vol. 8: 554).

Things that ordinarily pass as virtue - rational self-interest, such natural instincts as parental affection and pity, and conscience - are counterfeits. These simulacra prompt us to promote the good of others and to approve virtue and condemn vice. But they fall 'short of the extent of true virtuous benevolence, both in… nature and object' ([1765b] 1957-, vol 8: 609). There is no more true virtue in loving others from self-love than in self-love itself. Natural instincts such as parental affection or pity are defective because 'they don't arise from any temper of benevolence to Being in general' ([1765b] 1957-, vol. 8: 610). (Thus I may pity the misfortune of someone whose good fortune would displease me.) Conscience is the product of a power of placing ourselves in the situation of others (which is necessary for mutual understanding), a sense of the natural fitness of certain responses (injury and punishment or disapproval, benefit and approval or reward), and self-love. Placing ourselves in the situation of those we have injured, we recognize that we would resent being treated in the same way and that we are therefore inconsistent in approving of treating others in ways we would not wish to be treated ourselves. The resulting sense of 'inconsistence' or 'self-opposition' makes us 'uneasy' since 'self-love implies an inclination to feel and act as one with ourselves' ([1765b] 1957-, vol. 8: 589).

7 Theological aesthetics

According to Edwards, beauty or 'excellency' 'consists in the similarness of one being to another - not merely equality and proportion, but any kind of similarness…. This is an universal definition of excellency: The consent of being to being' (*[1829a] 1957-, vol. 6: 336*). One who loves others, for example, or wishes them well 'agrees' with them or 'consents' to them. But love's scope can be narrower or wider. Agreement or consent is 'comprehensive' or 'universal' only when directed towards being in general. True benevolence alone, therefore, is truly beautiful. The beauty of a well-ordered society, of the natural fitness between actions and circumstances, 'of a building, of a flower, or of the rainbow' is an 'inferior, secondary beauty' which is a mere image of this. Secondary beauty consists in order, symmetry, harmony or proportion, that is, in 'uniformity in the midst of variety' ([1765b] 1957-, vol. 8: 561-2).

Only God's benevolence is perfect. Hence, God alone is (truly) beautiful without qualification. The fitness of God's dispensations, the harmony of his providential design, and so on, also exhibit the highest degree of secondary beauty. God, therefore, is 'infinitely the most beautiful and excellent' ([1765b] 1957-, vol. 8: 550).

God is also the 'foundation and fountain… of all beauty'. 'All the beauty to be found throughout the whole creation is… the reflection of the diffused beams of that Being who hath an infinite fullness of brightness and glory' ([1765b] 1957-, vol. 8: 550-1). God's world is saturated with beauty - not only the 'harmony of sounds, and the beauties of nature' to which Edwards was especially sensitive but also (and primarily) the beauty of the gospel, of God's providential work in history and of the saints. The saints alone, however, can discern true beauty.

8 The sense of the heart

Because the Holy Spirit dwells in them, the saints love being in general. Their benevolence is the basis of a new 'spiritual sense' whose 'immediate object' is the 'beauty of holiness' - a 'new simple idea' that cannot 'be produced by exalting, varying or compounding' other ideas and that truly 'represents' divine reality ([1746] 1957-, vol. 2: 205, 260; [1765b] 1957-, vol. 8: 622).

Edwards sometimes identifies beauty with the pleasure that holy things evoke in the saints or the tendency they have to evoke it, and at other times with the consent of being to being. His view, however, appears to be as follows. An idea of true beauty is similar to a Lockean idea of a primary or secondary quality. Feelings of spiritual delight are (to use Locke's words) simple 'sensations or perceptions in our understanding' like our ideas of colour or solidity. The tendency holy things have to evoke these ideas in the saints is what Locke calls a 'quality', that is, 'the power to produce those ideas in us' ([1689] 1894, vol. 1: 169). Benevolence is the objective 'mechanism' underlying this power. Like simple ideas of primary and secondary qualities, the spiritual sensation 'represents' or is a 'perception' of its object. Just as 'extension' can refer to the idea, the power or the configuration of matter which is the base of the power, so 'true beauty' can refer to the spiritual sensation, the relevant dispositional property or to benevolence.

Edwards calls the new mode of spiritual understanding a ‘sense’ because the apprehension of spiritual beauty is (1) immediate (non-inferential) and (2) involuntary. Furthermore, (3) it involves relish or delight and (following Locke) being pleased or pained is a kind of sensation or perception. Finally, (4) it is the source of a new simple idea and (as Locke says) all simple ideas come ‘from experience’.

The saints alone are in a proper epistemic position to discern the truths of religion. Only those who are struck with the beauty of holiness can appreciate the ‘hatefulness of sin’, for example, and thus be convinced of the justice of divine punishment and our inability to make restitution. The new sense also helps us grasp the truth of the gospel as a whole. A conviction of the gospel’s truth is an immediate inference from its beauty or splendour.

Edwards’ defence of the objectivity of the new spiritual sense has four steps. (1) Benevolence agrees with the nature of things. The world is an interconnected system of minds and ideas in which the only true substance and cause is an infinite and omnipotent love. Human benevolence is thus an appropriate or fitting response to reality. (2) Benevolence delights in benevolence. Since benevolence is an appropriate response to reality, so too is delight in benevolence. (3) But delighting in benevolence just is perceiving its spiritual beauty. It follows that (4) the saints’ spiritual perceptions are true ‘representations’ of something ‘besides what [is] in [their] own minds’ (*[1765b] 1957-, vol. 8: 622*).

9 *The History of Redemption*

In a letter of 1757 to the trustees of the College of New Jersey, Edwards said that he had long contemplated ‘a great work, which I call a *History of the Work of Redemption*, a body of divinity in an entire new method, being thrown into the form of a history’ ([1829-30, vol. 1: 569-70] 1957-, vol. 9: 62). Edwards’ untimely death aborted the project, but a sermon series delivered in 1739 ([1774] 1957-, vol. 9) which traces the work of redemption ‘from the fall of man to the end of the world’ would have undoubtedly provided its basis. His proposed history would have been the natural culmination of the project begun in *True Virtue* and *End of Creation*. For the redemption is ‘the *summum* and *ultimum* of all the divine operations and decrees’, the manifestation of God’s internal glory in time ([1829-30, vol. 1: 569-70] 1957-, vol. 9: 62). It would have also provided a fitting climax to Edwards’ intellectual career as a whole. For it is in his work of redemption that God most fully displays his sovereignty, holiness and splendour. Whether Edwards’ work would have anticipated modern historiography, as some claim, is more doubtful. For the sermon series is essentially a doctrinal work. (The section on Christ’s earthly ministry, for example, is not a life of Jesus but a discussion of the incarnation and atonement.) Nor does Edwards restrict himself to natural causes, but freely appeals to divine decrees and to typology to explain events. Whatever novelty the sermon series has consists in the rich skein of images by which Edwards connects the events of redemption history and in an emphasis on the objective side of God’s act of redemption; this is comparatively rare in a Puritanism which stressed the redemption’s application. (Edwards treats the subjective side in *Religious Affections*.)

WILLIAM J. WAINWRIGHT

List of works

Edwards, J. (1957) *The Works of Jonathan Edwards*, ed. P. Miller (vols 1-2), J. Smith (vols 3-9) and H.S. Stout (vols 10-), New Haven, CT: Yale University Press, 1957-.(The earlier Dwight edition is a widely available edition of Edwards’ work. Miller *et al.* will supersede earlier editions when completed; the extensive introductions are especially helpful.)

Edwards, J. (1746) *Religious Affections*, in *The Works of Jonathan Edwards*, New Haven, CT: Yale University Press, 1957-, vol. 2.(An analysis of the nature of holy affections and the criteria for distinguishing them from counterfeits.)

Edwards, J. (1754) *Freedom of the Will*, in *The Works of Jonathan Edwards*, New Haven, CT: Yale University Press, 1957-, vol. 1.(A philosophically sophisticated defence of determinism.)

Edwards, J. (1758) *Original Sin*, in *The Works of Jonathan Edwards*, New Haven, CT: Yale University Press, 1957-, vol. 3.(Part 3 (351-433) contains interesting discussions of inherited liability, identity and occasionalism.)

Edwards, J. (1765a) *End of Creation*, in *The Works of Jonathan Edwards*, New Haven, CT: Yale University Press, 1957-, vol. 8, 405-536.(Discusses God’s reasons for creating the world.)

Edwards, J. (1765b) *True Virtue*, in *The Works of Jonathan Edwards*, New Haven, CT: Yale University Press,

1957-, vol. 8, 539-627.(Edwards' major treatise on ethics.)
Edwards, J. (1774) *A History of the Work of Redemption*, in *The Works of Jonathan Edwards*, New Haven, CT: Yale University Press, 1957-, vol. 9.(The sermon series delivered in 1739.)
Edwards, J. (1829a) 'The Mind', in *The Works of Jonathan Edwards*, New Haven, CT: Yale University Press, 1957-, vol. 6, 332-93.(Notes for a proposed treatise on the mind.)
Edwards, J. (1829b) 'Natural Philosophy', in *The Works of Jonathan Edwards*, New Haven, CT: Yale University Press, 1957-, vol. 6, 192-295.(A collection of papers and notes that contains the important 'Of Being', 'Of the Prejudices of the Imagination' and 'Of Atoms'.)
Edwards, J. (1890) 'Miscellany 94', in *The Works of Jonathan Edwards*, New Haven, CT: Yale University Press, 1957-, vol. 13, 256-63.(Edwards' first discussion of the Trinity.)
Edwards, J. (1903) 'An Essay on the Trinity', in *Treatise on Grace and Other Posthumously Published Writings*, Cambridge: James Clarke & Co., 1971, 99-131.(Written in 1730. Edwards' second major discussion of the Trinity.)
Edwards, J. (1817, 1847) *The Works of President Edwards*, ed. E. Williams and E. Parsons, New York: B. Franklin, 1968.(A reprint of the eight-volume 1817 edition and the two supplementary volumes published in 1847.)
Edwards, J. (1829-30) *The Works of President Edwards*, ed. S.E. Dwight, New York: G. & C. & H. Carvill, 10 vols.(A widely available edition of Edwards' work.)
Edwards, J. (1955) *The Philosophy of Jonathan Edwards from His Private Notebooks*, ed. H.G. Townsend, Eugene, OR: University of Oregon Monographs.(Contains philosophically important selections from the 'Miscellanies', a series of over 1,300 entries on philosophical and theological topics begun in 1722.)
Edwards, J. (1971) *Treatise On Grace and Other Posthumously Published Writings*, ed. P. Helm, Cambridge: James Clarke & Co.(Contains papers not included in the collected works, including 'Observations Concerning the Scripture Oeconomy of the Trinity and Covenant of Redemption' (Miscellany 1062).)

References and further reading

Delattre, R.A. (1968) *Beauty and Sensibility in the Thought of Jonathan Edwards*, New Haven, CT: Yale University Press.(The best study of Edwards' theological aesthetics.)
Fiering, N. (1981) *Jonathan Edwards' Moral Thought and its British Context*, Chapel Hill, NC: University of North Carolina Press.(An important study of Edwards' ethics and its sources.)
Jenson, R.W. (1988) *America's Theologian: A Recommendation of Jonathan Edwards*, New York: Oxford University Press.(An examination of Edwards' complex relations to Enlightenment thought.)
Locke, J. (1689) *An Essay Concerning Human Understanding*, Oxford: Clarendon Press, 1894, 2 vols.(An important influence on Edwards' epistemology, philosophy of language and philosophical anthropology.)
Miller, P. (1949) *Jonathan Edwards*, New York: W. Sloane Associates.(A pioneering work that emphasizes Edwards' debt to Locke.)
Smith, J.E. (1992) *Jonathan Edwards, Puritan Preacher, Philosopher*, London: Geoffrey Chapman and Notre Dame, IN: University of Notre Dame Press.(A useful introduction.)
Wainwright, W.J. (1980) 'Jonathan Edwards and the Language of God', *Journal of the American Academy of Religion* 48: 520-30.(Examines Edwards' theory of images and types.)
Wainwright, W.J. (1982) 'Jonathan Edwards, Atoms, and Immaterialism', *Idealistic Studies* 12: 79-89.(A reconstruction of Edwards' argument for immaterialism.)
Wainwright, W.J. (1988) 'Original Sin', in T.V. Morris (ed.) *Philosophy and the Christian Faith*, Notre Dame, IN: University of Notre Dame Press.(Examines Edwards' views on responsibility and identity.)
Wainwright, W.J. (1995) *Reason and the Heart*, Ithaca, NY: Cornell University Press.(Chapter 1 expands the material of §8 of this entry and discusses Edwards' views on reason.)
Wainwright, W.J. (1996) 'Jonathan Edwards, William Rowe, and the Necessity of Creation', in J. Jordan and D. Howard-Snyder (eds) *Faith, Freedom and Responsibility*, Lanham, MD: Rowman & Littlefield.(Argues that Edwards' theistic metaphysics entails that God must create the actual world.)
Winslow, O.E. (1940) *Jonathan Edwards 1703-1758: A Biography*, New York: Macmillan.(Still the best biography.)

Egoism and altruism

Henry Sidgwick conceived of egoism as an ethical theory parallel to utilitarianism: the utilitarian holds that one should maximize the good of all beings in the universe; the egoist holds instead that the good one is ultimately to aim at is only one's own. This form of egoism (often called 'ethical egoism') is to be distinguished from the empirical hypothesis ('psychological egoism') that human beings seek to maximize their own good. Ethical egoism can approve of behaviour that benefits others, for often the best way to promote one's good is to form cooperative relationships. But the egoist cannot approve of an altruistic justification for such cooperation: altruism requires benefiting others merely for their sake, whereas the egoist insists that one's ultimate goal must be solely one's own good.

One way to defend ethical egoism is to affirm psychological egoism and then to propose that our obligations cannot outstrip our capacities; if we cannot help seeking to maximize our own well being, we should not hold ourselves to a less selfish standard. But this defence is widely rejected, because psychological egoism seems too simple a conception of human behaviour. Moreover, egoism violates our sense of impartiality; there is no fact about oneself that justifies excluding others from one's ultimate end.

There is, however, a different form of egoism, which flourished in the ancient world, and is not vulnerable to this criticism. It holds that one's good consists largely or exclusively in acting virtuously, and that self-interest properly understood is therefore our best guide.

1 Definitions of 'egoism'

The term 'egoism' was introduced into modern moral philosophy as a label for a type of ethical theory that is structurally parallel to utilitarianism (see Utilitarianism). The latter theory holds that one ought to consider everyone and produce the greatest balance of good over evil; egoism, by contrast, says that each person ought to maximize their own good. Both theories are teleological, in that they hold that the right thing to do is always to produce a certain good (see Teleological ethics). But the utilitarian claims that the good that one is to maximize is the universal good - the good of all human beings and perhaps all sentient creatures. The egoist, on the other hand, holds that the good one is ultimately to aim at is only one's own (see Good, theories of the).

This way of classifying ethical theories is due to Henry Sidgwick, who regarded the choice between utilitarianism and egoism as one of the principal problems of moral philosophy. In *The Methods of Ethics* (1874), Sidgwick frames the issue in terms that assume that the good is identical to pleasure (a doctrine called 'hedonism') (see Hedonism). He uses 'utilitarianism' for the view that one is to maximize the amount of pleasure in the universe, and holds that the only form of egoism worth considering is hedonistic egoism. Since few philosophers now accept the identity of pleasure and the good, the terms of the debate have changed. 'Egoism' is applied to any doctrine, whatever its conception of the good, that advocates maximizing one's own good.

Often this doctrine is called 'ethical egoism', to emphasize its normative status. By contrast, the term 'psychological egoism' is applied to an empirical hypothesis about human motivation. It holds that whenever one has a choice to make, one decides in favour of the action one thinks will maximize one's own good. It is possible to agree that we are inevitably selfish in this way, but to regard this as an evil element in our nature. Conversely, it is possible to hold that although people ought to maximize their own good, they seldom try to do so. (In the remainder of this entry, 'egoism' will refer to ethical egoism, unless otherwise indicated.)

2 Egoism's treatment of altruism

A defender of egoism need not frown upon attachments to others, feelings of compassion, or beneficent acts. For it is open to the egoist to argue that these social ties are an effective means to one's own ends. For example, it is a matter of common sense that altruistic behaviour - behaviour intended to help others - is often advantageous, when it motivates others to respond in kind (see Reciprocity). What little one loses in simple acts of kindness may be more than compensated when others reciprocate.

Although egoists may argue that benefiting others is generally in one's interests, they give a controversial justification for such beneficence. It is widely agreed that one should at times benefit others for *their* sake

(Aristotle, for example, in *Nicomachean Ethics*, considers this essential to the best kind of friendship). To act for the sake of others is to take their good as a sufficient reason for action. But this is exactly what egoists cannot accept. They hold that ultimately the only justification for acts of beneficence is that they maximize one's own good. Ultimately, one is not to benefit others for their sake, but for one's own. If 'altruism' is used (as it often is) to refer to behaviour that not only benefits others, but is undertaken for their sake, then egoism is opposed to altruism.

3 Arguments for and against

Philosophers have sometimes tried to refute egoism by showing that it contains a contradiction or is in some way self-undermining. The best known attempt is that of G.E. Moore in *Principia Ethica* (1903), but he has had few followers. Instead, Sidgwick's opinion that egoism is rational is generally accepted. But even if one agrees, one may ask whether there are good reasons for choosing egoism over other alternatives. Why must it always be a mistake to sacrifice one's good for the greater good of others? If a small loss in one's wellbeing can produce great gains for others, what is wrong with accepting that loss?

The egoist might at this point take refuge in psychological egoism. Although it is possible to affirm psychological egoism and reject ethical egoism - to agree that by nature we are ultimately self-seeking, and to condemn such behaviour as evil - few philosophers regard this as an appealing mix of theories. For what plausibility can there be in a standard of behaviour that we are incapable of achieving? The egoist may therefore respond to our question 'Why should we not sacrifice our good for the sake of others?' by urging us not to impose impossible standards upon ourselves. We do not in fact make such sacrifices, and should not blame ourselves for being the way we are.

The problem with this strategy is that psychological egoism has come under heavy attack in the modern period. Hobbes (1651) and Mandeville (1714) have been widely read as psychological egoists, and were criticized by such philosophers as Hutcheson (1725), Rousseau (1755) and Hume (1751), who sought to show that benevolence, pity and sympathy are as natural as self-love (see Hobbes, T.; Mandeville, B.; Hutcheson, F.; Rousseau, J.-J.; Hume, D.). Kant held (1788), against psychological egoism, that the rational recognition of moral principles can by itself motivate us and overcome self-love (see Kantian ethics). Perhaps the most influential critique of psychological egoism is that of Butler (1726), who argued that by its nature self-love cannot be the only component of our motivational repertoire (see Butler, J.). He also pointed out that even if we feel gratification when we satisfy our desires, it cannot be inferred that such gratification is the object of those desires. The combined force of these attacks has left psychological egoism with few philosophical defenders.

At this point, an important challenge to ethical egoism should be noticed: although my circumstances, history, or qualities may differ from yours in morally significant ways, and these differences may justify me in seeking my good in preference to yours, the mere fact that I am myself and not you is not by itself a morally relevant difference between us. That my good is mine does not explain why ultimately it alone should concern me. So, if my good provides me with a reason for action, why should not your good, or the good of anyone else, also provide me with a reason - so long as there are no relevant differences between us? The ideal of impartiality seems to support the conclusion that we should have at least some concern with others (see Impartiality). In fact, egoists implicitly accept a notion of impartiality, since they say that just as my ultimate end should be my good, yours should be your good. So they must explain why they accept this minimal conception of impartiality, but nothing stronger. There is nothing morally appealing about excluding all others from one's final end; why then should one do so?

4 An ancient form of egoism

The kind of egoism we have been discussing can be called 'formal', in that it makes no claim about what in particular is good or bad for human beings. It holds instead that whatever the good is, it is one's own good that should be one's ultimate end. This is the conception one arrives at when one begins from Sidgwick's pairing of egoism with utilitarianism, then abstracts from his hedonism. A different kind of egoism, which might be called 'substantive', first proposes a concrete conception of the good, and then urges each of us to maximize our own good, so conceived. It is this form of egoism that flourished in the ancient world. Plato, Aristotle and the Stoics do not accept the formal principle that whatever the good is, we should seek only our own good, or prefer it to the good of others (see Plato; Aristotle; Stoicism). Instead, they argue for a specific conception of the good, and

because the social virtues play so large a role in that conception, they regard self-love not as the enemy of virtue and the larger community but as an honourable motive, once it is developed in the proper direction (see Eudaimonia; Virtues and vices).

Even if psychological egoism is too simple a conception of human nature, it is undeniable that we normally have a deep concern for our own welfare. If self-love is a force that often conflicts with moral duty and inherently resists education, then human beings are necessarily and deeply divided creatures. This is the Augustinian and Kantian picture. By contrast, the dominant strand of ancient ethics proposes a more optimistic conception of the human situation. It does not claim that one should seek one’s own good, come what may for others; rather, by arguing that acting virtuously and acting well coincide, it seeks to undermine the common assumption that at bottom the self must come into conflict with others.

See also: Buddhist concept of emptiness; Moral motivation; Moral scepticism; Morality and identity §2; Prudence

RICHARD KRAUT

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985, books VIII, IX. (Discusses friendship.)

Butler, J. (1726) *Fifteen Sermons Preached at the Rolls Chapel*, Sermons I, II, III, XI, XII; repr. in S. Darwall (ed.) *Five Sermons Preached at the Rolls Chapel and A Dissertation Upon the Nature of Virtue*, Indianapolis, IN: Hackett Publishing Company, 1983, esp. Sermon XI.(Argues that self-love cannot be the only human motivation.)

Gauthier, D. (ed.) (1970) *Morality and Rational Self-Interest*, Englewood Cliffs, NJ: Prentice Hall.(Selections by historical figures, contemporary essays and a bibliography.)

Hobbes, T. (1651) *Leviathan*, ed. E. Curley, Indianapolis, IN: Hackett Publishing Company, 1994, part I, chaps 6-16.(Often read as a work of psychological egoism.)

Hume, D. (1751) *An Enquiry Concerning the Principles of Morals*, ed. J.B. Schneewind, Indianapolis, IN: Hackett Publishing Company, 1983, sections 5, 9. (Seeks to show the naturalness of sympathy.)

Hutcheson, F. (1725) *The Original of our Ideas of Beauty and Virtue*, excerpted in J.B. Schneewind (ed.) *Moral Philosophy from Montaigne to Kant*, vol. 2, Cambridge: Cambridge University Press, 1990.(Attempts to refute psychological egoism.)

Kant, I. (1788) *Critik der practischen Vernunft*, trans. L.W. Beck, *Critique of Practical Reason*, New York: Macmillan, 1993, 36-8.(Argues that recognition of moral principles can overcome self-love.)

Mandeville, B. (1714) *The Fable of the Bees: or, Private Vices, Publick Benefits*, ed. F.B. Kaye, Indianapolis, IN: Liberty Fund, 1988.(A portrait of human selfishness.)

Moore, G.E. (1903) *Principia Ethica*, ed. T. Baldwin, Cambridge: Cambridge University Press, revised edn, 1993, sections 58-61, 63.(An attempt to refute egoism.)

Nagel, T. (1970) *The Possibility of Altruism*, Oxford: Clarendon Press.(A difficult but widely discussed attack on egoism.)

Plato (*c*.380-367 BC) *Republic*, trans. A.D. Lindsay, revised by T.H. Irwin, London: Dent, 1992.(The most elaborate attempt to show that it is in one’s interest to be just.)

Rousseau, J.-J. (1755) *Discours sur l’origine et les fondements de l’inégalité parmi les hommes*, trans. D.A. Cress, *Discourse on the Origin of Inequality*, Indianapolis, IN: Hackett Publishing Company, 1992.(Contains a discussion of self-love and compassion.)

Rousseau, J.-J. (1762) *Émile: ou, de l’éducation*, trans. A. Bloom, *Emile: or, On Education*, Harmondsworth: Penguin, 1991, esp. book IV.(Distinguishes healthy from destructive self-love, and describes an education in which healthy self-love flourishes.)

Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan; 7th edn, 1907. (Argues for the plausibility of both egoism and utilitarianism.)

Egyptian cosmology, ancient

Ancient Egypt has left us no systematic philosophy in the modern sense. However, there is abundant evidence that the Egyptians were concerned with all the usual problems of existence. The answers to these questions were mostly expressed through the use of myth, or commentary upon myth. Though complex and polytheistic, Egyptian religion provided a subtle means of commentary upon a range of theological, ethical and psychological questions. The range and quality of Egyptian technical achievements presupposes a degree of theoretical knowledge, some of which has survived and some of which can be reconstructed, either from Egyptian texts themselves, or from commentaries in the classical authors. Until recently much of the latter has been dismissed as inaccurate, but modern scholars are increasingly inclined to agree with the high value which Greek commentators placed on Egyptian thinking.

1 Sources

Although Egyptian civilization lasted for almost thirty-five centuries, many texts survive only in fragments, or are lost completely. The small fraction which has come down to us is biased in favour of royal inscriptions or monumental texts. The loss of temple libraries, which are known to have contained technical manuals and catalogues, is especially unfortunate. The large quantity of administrative texts which survive are of more value to the social and economic historian than to the philosopher. On the other hand, a strong tradition of scribal education existed which has ensured the transmission of ethical and didactic works in some numbers. As a result, the broad principles of Egyptian thinking can be reconstructed with some confidence.

A good starting point can be found in the so-called *Memphite Theology*. In its present form this is a hieroglyphic inscription, now in the British Museum. It is dated to the reign of Shabaqo (*c*.710 BC), although the original text could date from 500 years earlier and a date in the third millennium BC has even been suggested. In the *Memphite Theology*, the god Ptah, patron of Memphis and of craftsmanship, is described as the creator of the world, which he does by means of his heart (the traditional seat of thought and feeling) and his tongue. In other words, the creator first conceptualizes the forms to be created and then brings them into being by means of articulate language.

The Egyptians attached great importance to names. They believed names contained the essence of the object or person to which they referred. They were also drawn to etymologies based on puns. These were considered not to be accidental, but to contain vital information about the meaning and purpose of words. Egyptian dictionaries were based on this principle and the hieroglyphic script itself arises from similar use of paronomasia, or play on words. Creation legends were common in Egypt as in other civilizations, but the sophistication and clarity of the *Memphite Theology* are remarkable. Parallels with the opening of Genesis are occasionally made, although these should be treated with caution.

Similar sophistication can be found unusually in the *Book of the Dead*. This is a mixture of spells, myths and apotropaic formulae (intended to prevent evil), designed to secure the survival of the deceased beyond the divine judgment which awaited them and into the afterlife. Although the ideas underlying the *Book of the Dead* are complex and sublime, the composition itself is not normally considered an intellectual triumph. Chapter 75 of the text contains impressive material which takes the form of a dialogue between the spirit of the deceased, often identified with the god Osiris, and a supreme creator named Atum. The dead man finds eternity incomprehensible because it lacks food and drink, human love and familiar landmarks of time or space. Atum’s reply is that all these have been superseded by the eternal contemplation of his own face, amounting to a communion with his presence beyond physical description. Here too there is evidence of serious questioning about the nature of existence and the problem of comprehending other dimensions of reality.

2 Doctrines

Egyptian religion was complex and its only modern parallel would be Hinduism. A multiplicity of gods was subdivided into local divinities. These were arranged in triads, broadly comparable with human families consisting of god, goddess and child (normally male, but sometimes female). The same gods could rearrange themselves into different families in differing localities. Most gods could be assimilated into the solar deity, which had various names, while some gods combined in pairs for certain functions. The god Osiris was thought of as an earthly

god-king who had undergone death and been restored to immortality in the underworld. In many ways, Osiris and the sun-god are counterparts. This concept introduces us to one of the characteristics of Egyptian thought: the emphasis on duality and the need to reconcile opposing principles. The force which reconciled these opposites was *ma'at*, a concept embracing truth, justice and harmony, and which could be personified as a goddess in its own right.

A similar duality can be seen in the theology of kingship. Unlike the rulers of neighbouring states, the Pharaoh was both divine and human. The divine aspect of the ruler was expressed in the prenomen given to him (or occasionally her) at the coronation. This name was borne by him even after his death, when he became Osiris. His personal name, or nomen, was equally revered, but was not part of his immortal identity. (Conventionally, this duality was expressed in a double title, but wrongly translated as 'King of Upper and Lower Egypt'.) It is possible although not certain that this pattern of thought survived into the early Christian period, where it may have predisposed the Egyptians towards Monophysitism, the belief that the person of Christ has only one nature, which is divine and possesses human attributes.

A corresponding pattern of duality and immanence can be seen in the concept of the *ka*, essentially a birth-spirit similar to the Latin *genius*. However, the idea also encompassed the notion of the immortal essence humans were thought to possess and the individuality which distinguished them. Other living things possessed *kas*, a belief which conveniently allowed offerings of food or flowers to be presented to the gods, who would happily partake of their *kas* while leaving the physical objects to be consumed by the priesthood. However, the *ka* was also dependent on the survival of the earthly body after death. This tenet underwrote the entire death industry characteristic of the culture, although in the main it was concerned with immortality rather than morbidity. There is also a psychological link between the *ka* and the name. In later texts these two terms can be used interchangeably.

Later Neoplatonists, notably Iamblichus, were fascinated by Egyptian religion, at least by the aspects which were accessible. Iamblichus saw the entire system as monotheistic and he and others were convinced that it foreshadowed the Platonic Theory of Forms (see Plato §10). Most Egyptologists have been dismissive of these ideas, although they have begun to gain some credibility over time. Monotheism is compatible with much of what is known about Egyptian gods. The idea itself, and the broad outlines of the theory of Forms, might have been acceptable to an educated Egyptian at the earliest period. The Egyptian would have rejected the exclusive nature of such ideas as the sole explanation of thinking. Those who did try to impose such exclusivity upon Egyptian religion, such as the Pharaoh Akhenaten, were condemned. Some of these ideas, combined with themes from Hellenistic philosophy, resurfaced in the body of texts known as the *Corpus Hermeticum* (second century AD), where they are ascribed to the Graeco-Egyptian divine sage known as Hermes Trismegistos (see Hermetism).

Egyptian technology was impressive as was the thinking which went with it. The formula for the area of a circle, $A = (8/9\ d)^2$, gives an excellent working value for π, and there are textual references to nanoseconds, the dependence of moonlight on sunlight and possibly the speed of light itself. The fact that such things could not be measured did not mean that they could not be envisaged. Astronomical methods of timekeeping were fairly accurate, as was the use of the water-clock. Astrology, however, was a late import from Mesopotamia, although one that soon became thoroughly naturalized.

Egyptian wisdom literature is a genre of its own. Primarily it is concerned with social and ethical problems of some complexity. As the genre developed, there was a growing emphasis away from questions of purely social behaviour and towards the problem of fate and its relation to individual guilt and responsibility. One answer to this dichotomy was to contrast the psychological impact of good behaviour with that of bad, thus internalizing moral responsibility. In a late text, *Papyrus Insinger*, good is externalized and seen as an absolute to be pursued regardless of consequences or the vicissitudes of fortune. Here too the complexity and range of Egyptian writings are seen at their best. Egyptian thought, like Egyptian art and architecture, tolerates diversity and is characterized by both sophistication and simplicity.

See also: Egyptian philosophy: influence on ancient Greek thought

JOHN D. RAY

References and further reading

Assmann, J. (1995) *Egyptian Solar Religion in the New Kingdom*, London: Kegan Paul International.(Not an easy

read, but an excellent example from one of the best authorities.)
Copenhaver, B.P. (1992) *Hermetica*, Cambridge: Cambridge University Press. (Translations of the *Corpus Hermeticum* and a very full commentary.)
Hornung, E. (1983) *Conceptions of God in Ancient Egypt: the One and the Many*, London: Routledge & Kegan Paul.(Thoughtful analysis of the main patterns of Egyptian thought.)
Griffiths, J. Gwyn (1995) *Triads and Trinity*, Cardiff: University of Wales Press.(Important study in comparative religion.)
Lichtheim, M. (1973-80) *Ancient Egyptian Literature*, 3 vols, Los Angeles, CA: University of California Press. (Clear translations and extracts from standard works.)
Ray, J.D. (1995) 'Egyptian Wisdom Literature', in J. Day *et al*. (eds) *Wisdom in Ancient Israel: Essays in Honour of J.A. Emerton*, Cambridge: Cambridge University Press.(Survey of Egyptian ethical writing.)
Shafer, B.E. (ed.) (1991) *Religion in Ancient Egypt*, London: Routledge. (A series of good essays with wide a variety of approaches.)
Shaw, I. and Nicholson, P. (1995) *The British Museum Dictionary of Ancient Egypt*, London: British Museum Press.(Useful individual entries.)

Egyptian philosophy: influence on ancient Greek thought

Before the decipherment of hieroglyphics (a process only completed in the 1830s), it was widely believed that many famous Greek philosophers had studied in Egypt and that Greek philosophy was ultimately derived from a lost 'Egyptian mystery system'. This belief was derived in part from ancient sources, which described how certain Greek philosophers had studied with Egyptian priests. The notion that these individual sessions were part of an extensive formal programme of education derives from an historical novel, Séthos (1731), by the Abbé Jean Terrasson. This book, which pretended to be based on lost original sources, offered a detailed portrait of a complex Egyptian university system. It was translated into several European languages and widely popularized in the rituals and mythology of Freemasonry.

The existence of such a formal Egyptian mystery system of education was not confirmed by actual Egyptian sources once they could be read and translated. The myth has been given a new lease of life in revisionist histories of the ancient world composed by writers whose ancestors had been brought to the New World as slaves. These writers sought to show that Greek philosophy was derived from Egyptian philosophy and that what has been recognized as Western civilization stems from Africa. This entry reviews the evidence for these claims and concludes that although the Greeks had great respect for Egyptian wisdom and piety, what has always been known as Greek philosophy derives from the original work of Greeks. It finds moreover that if the Greek philosophers who lived in Ionia were influenced by any outside ideas, these came to them through the monotheistic religions of other peoples living in the Near East.

1 Greek philosophers in Egypt

Nothing in surviving Egyptian literature from the period before the founding of Alexandria in 332 BC resembles the dialectical methods and argumentative structures invented and used by Greek philosophers. Even after Greeks had been living in Egypt for some centuries, Egypto-Greek philosophical treatises, such as those preserved in the Hermetic Corpus, have a declarative format and theological character (see Hermetism). If Greek philosophy were closely dependent on Egyptian models, we would expect to find evidence of parallel texts or direct echoes, as we do in the case of the Roman poets who from early times adapted and quoted Greek sources. However, we do not find this, in Egypt or in any other contemporary civilization, in the area of Greek philosophy. This is not to say that the Egyptians did not possess a profound theology and concept of justice, or that this theology and other wisdom is not impressively expressed in surviving Egyptian literature. It is simply that Egyptian (or Hebrew) wisdom literatures are different in nature from the philosophical writings of Plato and Aristotle. The writings of these philosophers are nontheological in character and reach their conclusions through argument and discussion. Nor is it likely that there was an Egyptian philosophy, similar to Greek philosophy, that is now lost. If there had been such an extensive body of literature some Greeks would have known about it and there would not have been any reason not to acknowledge it. Greeks tended to be so respectful of Egyptian learning that they were eager to refer to it whenever they could. The kind of Egyptian learning to which Plato refers in his works takes the form of wisdom narratives, the kind of moral tales of which we possess Egyptian examples. In the *Phaedrus* and the *Philebus*, Socrates tells a story about the god Theuth's invention of letters. In the *Timaeus* and the *Critias*, Plato's incomplete dialogue, Critias tells the story of Atlantis, which his ancestor Solon learned from an Egyptian priest during his visit to the Egyptian pharaoh Amasis. These stories are not told in the usual question-and-answer dialogue form, but as long didactic narratives almost certainly of Plato's own invention.

The ancient biographers who describe the philosophers' sojourns in Egypt never discuss in detail what philosophical ideas the Greeks might have learned from Egyptian priests. Rather, they imagine that the Greeks acquired from the Egyptians practical learning, often of a scientific nature. In the minds of the biographers, the acquisition of such Eastern wisdom provided a motivation for Greek philosophers to travel to Egypt and to the Near East. Foreign travel also helped to explain why particular philosophers contrived to be distinctive and original. When Thales (sixth century BC) conjectured that the first element was water he could have been drawing on Babylonian mythology, or reflecting on the importance of the river Maeander to his native city of Miletus. But his biographers supposed that he had also seen the most remarkable of all rivers, the Nile. In order to account for the originality of his ideas of mathematics and religion Greek writers assumed that Pythagoras visited Egypt and the Near East before settling in southern Italy. A century after Pythagoras' death, historian Herodotus conjectured

that Pythagoras’ theory of the transmigration of souls, which has no parallel in early Greek thought, had an Egyptian origin. In fact Egyptian notions of the fate of the soul after death are completely different from the Greek. In the first century BC Egyptian priests told Diodorus that Pythagoras shared with the Egyptians the notion that animals had souls. They also claimed that Pythagoras learned about geometry in Egypt, without mentioning anything specific. Both Egyptians and Greeks were interested in geometry and this common interest was seen to be evidence of dependence. Still more was discovered about Pythagoras as time went on. The Greek Iamblichus specifies that Pythagoras spent twenty-two years in Egypt during which time he studied all aspects of Egyptian religion. He was also initiated into all the rites and learned astronomy and geometry. Iamblichus provides no specific information about what he actually learned. Rather, the point of the journey seems to be that he studied whatever it was he learned abroad. Before going to Egypt he visited Syria and after Egypt he studied with the Magi in Babylon. For ancient Greek biographers, foreign study was the most natural and convenient way to account for the originality of Pythagoras’ ideas.

The priests told Diodorus that everything for which these Greeks were admired was brought from Egypt. They said that Greek philosopher Democritus of Abdera (fifth century BC) spent five years studying astrology in Egypt. This connection seems tenuous because the Egyptians were interested in astronomy, the motion of the stars and not in astrology, predictions about human fate that might be derived from astronomical observation. Like many Greek writers Democritus was interested in the causes of the inundation of the Nile. However, according to astronomer Ptolemy (second century AD), Democritus did his research on weather indications in Macedonia and Thrace. Nothing that he is reported to have said suggests that he had a detailed personal knowledge of Egypt.

In the case of other Greek philosophers, interest in the Nile or geometry counts as ‘evidence’ of a visit to Egypt. The priests told Diodorus that the Greek astronomer Oenopides of Chios (fifth century BC) learned from Egyptian priests about the obliquity of the ecliptic of the sun. Diodorus does not indicate and perhaps did not know that the Pythagoreans already knew about the ecliptic before Oenopides discovered it, or that it had been recognized by the Babylonians around 700 BC. Earlier in his account of Egypt Diodorus gives Oenopides’ explanation of the inundation of the Nile. He explained Oenopides’ deduction that the temperature of well water feels warm in the winter and cold in the summer indicating that the Nile’s subterranean waters maintain similar temperatures in the summer when there are no rains in Egypt. Again there is no reason to imagine that Oenopides needed to study in Egypt in order to form this false hypothesis. His reliance on analogy rather than on empirical evidence seems characteristically Greek.

Diodorus reports that the fourth-century Greek philosopher Eudoxus of Cnidus, like Democritus, learned astrology from the Egyptian priests, although it was the Alexandrian Greeks rather than the Egyptians who were interested in that subject. No doubt Eudoxus could have learned geometry in Egypt. However, it was also possible and more likely that he learned the theory of axiomatic mathematics from his teacher the fouth-century Greek philosopher Archytas of Tarentum. In fact nothing in the surviving fragments suggests that Eudoxus had a highly specialized knowledge of Egypt judging from the fragments. He could have learned what he knew from Greek writers like Hecataeus and Herodotus without ever having visited Egypt. None the less, new details about his visit were added as time went on. The priests at Heliopolis showed the Greek geographer Strabo (first century BC) statues of Plato and Eudoxus and pointed out the places where they had studied, although no one was sure how long they had stayed there as there were several different versions of the story. Diogenes Laertius supplies the name of Eudoxus’ Egyptian teacher, saying that while in Egypt Eudoxus shaved his beard and eyebrows like an Egyptian priest and that while there an Apis bull licked his cloak. Egyptian priests understood this to be an omen that he would be famous, but only for a short time. These anecdotes tell us nothing about what Eudoxus studied or did in Egypt. Instead, they portray him as a late antique holy man, in much the same way as Iamblichus described Pythagoras.

In view of the way Plato uses Egyptian lore, the presence of Egyptian tales in his dialogues does not necessarily indicate that he studied in Egypt, especially since the narration of these tales does not require the use of dialectic and argumentation. Plato never says in any of his writings that he went to Egypt and there is no reference to such a visit in the semibiographical *Seventh Letter*. The superficial knowledge of Egypt and vague chronology displayed in his dialogues resemble historical fiction more than history. Anecdotes about his visit to Egypt appear in the work of writers after the Hellenistic period (*c*.300 BC to the mid first century BC). Later biographers added details to the story of Plato’s Egyptian travels in order to provide aetiologies for the Egyptian references in his writings. In a letter purportedly sent by Phaedrus to Plato in Sais, Plato studies the question of the ‘all’ and Phaedrus asks

for information about the pyramids and unusual Egyptian animals. Plato was supposed to have heard from Sechnupis of Heliopolis the story of Theuth told in the *Phaedrus* and Diogenes Laertius says he went to Egypt to study with Egyptian seers.

The Church Father Clement of Alexandria (AD 150-215) suggested that Plato studied in Egypt with the Hermes Trismegistus, or 'thrice greatest Hermes'. But in fact the collection of works attributed to Hermes could not have been written without the conceptual vocabulary developed by Plato and Aristotle. It is deeply influenced by Plato and the writings of Neoplatonist philosophers in the early centuries AD (see Neoplatonism). Plato seems never to have learned from these Egyptian teachers anything characteristically Egyptian, so far as is known about Egyptian theology from Egyptian sources. His notion of Egyptian theology and life remains similar to that of other Athenians. He did not so much change the Athenian notion of Egyptian culture as enrich and idealize it so that it could provide a dramatic and instructive contrast with Athenian customs in his dialogues.

2 The myth of an Egyptian mystery system

The Egyptians had an urgent reason for wanting to assert their priority over the Greeks. In Herodotus' day the country was under Persian domination and then was ruled by Greeks after Alexander's conquest, therefore one of the few remaining ways for them to maintain national pride was through their history. Similar assertions of priority over the Greeks were made by Jews living in Alexandria. These Jews were determined to show that despite being subject to Greeks, they not only understood Greek culture but had provided the inspiration for the authors of the sacred writings and cherished literature of their conquerors' civilization: the mythical Greek singer Musaeus was none other than Moses. Such claims were taken up by the Church Fathers who sought to show that Greek philosophy owed a considerable debt to Hebrew scripture. There was support for the notion of Egyptian influence on Greek thought in the syncretism of the early centuries AD, most particularly in the cult of Isis, which combined Egyptian ritual practices with the Greek notion of *mysteria*, or initiations.

Terrasson drew on these sources (and his own ideas about French Christian education) for his depiction in his novel *Séthos* (1731) of an elaborate series of Egyptian initiations. The process he described inspired the contemporary initiation rituals of Freemasonry and were perhaps parodied in Mozart's opera *Die Zauberflöte*. The belief persists that these fundamentally Greek ideas could have been Egyptian according to revisionist histories of writers who call themselves Afrocentrists. These writers have insisted that Greek philosophy originated in Egypt, although no direct connection can be found between Egyptian wisdom literature and work of writers like Heraclitus, Plato, or Aristotle.

One of the most influential of these writers, Marcus Garvey (1887- 1940) stated that white Europeans had concealed the truth about the priority of Egyptian (or African) culture because of their hostility towards black people. M. Bernal supported these claims in his study *Black Athena* (1987). C.A. Diop in his influential *Civilisation ou Barbarie (Civilization or Barbarism: an Authentic Anthropology)* (1981) sought to show that there was widespread African influence on Western thought. G.G.M. James in *Stolen Legacy* (1954) and his pupil Y.A.A. ben-Jochannan in *Africa, Mother of Western Civilization* (1971) stated that Aristotle stole books from the library at Alexandria, which in fact was not built until after Aristotle's death. Moreover, no other ancient author says that Aristotle went to Egypt with Alexander. Such claims have no more substance than the assertions of the Hellenistic Egyptians and Jews who claimed that the Greeks borrowed all their ideas from their ancestors. There is no evidence that the Greeks stole or borrowed their philosophy from Egypt, or indeed from other peoples, such as the Phoenicians, Hittites, or Medes, with whom they came into contact. However, there was some Egyptian influence on Greek medicine, science and mathematics from the middle of the second millennium BC which continued after the Greeks re-established contact with the Egyptians in the sixth century BC. But, influence cannot be thought of as theft or borrowing and is usually a two-way process. If the early Greek philosophers drew their ideas from any foreign cultures, it was principally from the civilizations of the Near East.

See also: Egyptian cosmology, ancient

MARY R. LEFKOWITZ

References and further reading

ben-Jochannan, Y.A.A. (1971) *Africa, Mother of Western Civilization*, Baltimore, MD: Black Classic Press, repr. 1988.(Purports to show that Greek civilization was stolen from Egypt.)

Bernal, M. (1987) *Black Athena: The Afro-Asiatic Roots of Classical Civilization*, vol. 1, *The Fabrication of Ancient Greece*, vol. 2, *The Archaeological and Documentary Evidence*, New Brunswick, NJ: Rutgers University Press.(Vol. 1 attempts to demonstrate that because of racism and anti-Semitism, European scholars failed to recognize the Egyptian and near Eastern origins of Greek civilization. Vol. 2 supplies factual evidence for the theories advanced in vol. 1.)

Brisson, L. (1987) 'L'Égypte de Platon' (Plato's Egypt), *Les études philosophiques*, new series, 2-3: 101-31.(A brief account of the treatment of Egypt in Plato's philosophy.)

Burkert, W. (1987) *Ancient Mystery Cults*, Cambridge, MA: Harvard University Press.(Discusses the religious and social purposes of ancient Greek mystery cults.)

Copenhaver, B.P. (1992) *Hermetica*, Cambridge: Cambridge University Press.(An annotated translation of the treatises attributed to Hermes Trismegistus.)

Curto, S., Donadoni, S. and Donadoni Roveri, A.M. (eds) (1990) *L'Egitto dal mito all' Egittologia*, trans. E. Poore and F.L. Rossi, *Egypt from Myth to Egyptology*, Turin: Istituto Bancario San Paolo.(Describes how Egypt has been seen by Europeans from ancient times to the present day.)

Dent, E. (1960) *Mozart's Operas*, Oxford: Clarendon Press, 2nd edn. (Offers a detailed account of the Masonic elements in Mozart's opera *The Magic Flute*.)

Diop, C.A. (1981) *Civilisation ou Barbarie*, Paris: Présence Africaine; trans. Y.-L. Meema Ngemi, *Civilization or Barbarism: an Authentic Anthropology*, Brooklyn, NY: Lawrence Hill Books, 1991.(Attempts to discover the origins of European thought in ancient Africa.)

Froidefond, C. (1971) *Le mirage égyptien dans la littérature grecque d'Homère à Aristote (The Egyptian Mirage in Greek Literature From Homer to Aristotle)*, Aix-en-Provence: Ophrys.(Examines the Greeks' knowledge of Egypt before Alexander's conquest.)

Iversen, E. (1961) *The Myth of Egypt and its Hieroglyphs in European Tradition*, Princeton, NJ: Princeton University Press; repr. 1993. (Describes how the nature of the Egyptian system of writing was misunderstood by Europeans and was used by them to create a mythical Egypt.)

James, G.G.M. (1954) *Stolen Legacy*, New York: Philosophical Library.(Claims that Greek philosophy was stolen from Egypt.)

Lefkowitz, M. (1996) *Not Out of Africa: How Afrocentrism Became an Excuse to Teach Myth as History*, New York: Basic Books.(Examines Afrocentric myths about Greek culture and the origins of Greek civilization.)

Morenz, S. (1973) *Egyptian Religion*, trans. A.E. Keep, Ithaca, NY: Cornell University Press.(A general account of Egyptian ritual practices and beliefs.)

Palter, R. (1996) 'Black Athena, Afro-Centrism, and the History of Science', in M.R. Lefkowitz and G.M. Rogers (eds) *Black Athena Revisited*, Chapel Hill, NC: University of North Carolina Press, 209-66. (Examines the extent of Egyptian influence on Greek science.)

Riginos, A.S. (1976) *Platonica: The Anecdotes Concerning the Life and Writings of Plato*, Leiden: E.J. Brill. (Analyses the ancient legends about the life of Plato.)

Rutherford, R.B. (1995) *The Art of Plato*, London: Duckworth.(DiscussesPlato's use in his dialogues of stories about Egypt.)

Terrasson, J. (1731) *Séthos*, Paris: Jacques Guérin; trans. T. Lediard, London: J. Walthoe, 1732.(An historical novel, based on Greek and Roman sources, about the education of an Egyptian prince.)

Einstein, Albert (1879-1955)

Albert Einstein was a German-born Swiss and American naturalized physicist and the twentieth century's most prominent scientist. He produced the special and general theories of relativity, which overturned the classical understanding of space, time and gravitation. According to the special theory (1905), uniformly moving observers with different velocities measure the same speed for light. From this he deduced that the length of a system shrinks and its clocks slow at speeds approaching that of light. The general theory (completed 1915) proceeds from Hermann Minkowski's geometric formulation of special relativity as a four-dimensional spacetime. Einstein's theory allows, however, that the geometry of spacetime may vary from place to place. This variable geometry or curvature is associated with the presence of gravitational fields. Acting through geometrical curvature, these fields can slow clocks and bend light rays.

Einstein made many fundamental contributions to statistical mechanics and quantum theory, including the demonstration of the atomic character of matter and the proposal that light energy is organized in spatially discrete light quanta. In later life, he searched for a unified theory of gravitation and electromagnetism as an alternative to the quantum theory developed in the 1920s. He complained resolutely that this new quantum theory was not complete. Einstein's writings in philosophy of science developed a conventionalist position, stressing our freedom to construct theoretical concepts; his later writings emphasized his realist tendencies and the heuristic value of the search for mathematically simple laws.

1 Life

Albert Einstein was born in Ulm in southern Germany. He entered the Federal Polytechnical School in Zurich, Switzerland, in 1896 to study physics, after failing the entrance exam the previous year. Upon completion of his studies in 1900, he was unable to find an academic position. Having renounced German citizenship in 1896, he acquired Swiss citizenship in 1901. From 1902 to 1909 he worked as a patent examiner in Bern. He then moved rapidly through a series of academic positions in Zurich and Prague. Finally in 1914 Einstein moved to a prestigious professorship without teaching obligations at the University of Berlin and also to the directorship of the new Kaiser Wilhelm Institute for Physics. In 1919, observations of the solar eclipse provided confirmation of his general theory of relativity. The success was trumpeted in the popular press and Einstein became an unwilling celebrity. In 1922, he was awarded the Nobel Prize. After 1920, Einstein became known as an advocate of cultural Zionism, also a socialist and (until the rise of Hitler) a pacifist. After fleeing Germany, from 1933 until his death he was at the Institute for Advanced Study in Princeton, New Jersey, adopting citizenship of the USA in 1940.

2 Special relativity

Classical electrodynamics posited the ether as a medium for carrying light and other electromagnetic waves (see Maxwell J.C. §2; Electrodynamics; Optics §§1, 4). Einstein based his 1905 special theory of relativity on the conviction that the state of rest of this medium was superfluous and could be eliminated from an otherwise unaltered electrodynamics by revising classical notions of space and time. These revisions were developed as the consequences of two postulates: the principle of relativity, which required the equivalence of all uniform or inertial states of motion, and the light postulate, which required that the speed of light remain the same constant, independent of the motion of the source, for observers in any inertial motion.

At first, Einstein's two postulates seemed incompatible. As an observer proceeds through a sequence of ever-faster inertial states of motion, the observer's speed comes closer to that of light. However the light postulate requires that they would see no reduction in the speed of light. Einstein showed that the postulates are compatible if rods and clocks do not behave classically when they move close to the speed of light. Similarly, he showed that no material system could be accelerated up to the speed of light. As the system approaches that speed its inertial mass would increase without limit, precluding further acceleration. The mass gained is proportional to the energy of motion and provides one example of the inertia of energy, a result that became famous under the rubric '$E = m\mathrm{c}^2$'.

In his celebrated 1905 analysis, Einstein explained the apparent incompatibility of the postulates as resulting from an arbitrary conception about the simultaneity of distant events. He described a procedure involving light signals

that would by definition synchronize distant clocks. From this definition and the postulates of his theory, Einstein showed that inertially moving observers in relative motion must disagree over which distant events are simultaneous, in contradiction of classical expectations (see Relativity theory, philosophical significance of §§1-3).

3 General relativity

Einstein's general theory of relativity is a relativistic theory of gravitation extending Minkowski's spacetime formulation of special relativity (see Spacetime §2). Its principal novelty is that gravitational fields are not represented as entities like electric fields distinct from the background spacetime. Rather they appear as curvature in the background spacetime geometry. Since Einstein's theory relegated gravitation to this background structure, it could readily account for the most distinctive property of gravitation: the trajectories of bodies in free fall are essentially independent of their masses and internal constitutions.

The theory agrees observationally with Newtonian theory almost exactly in common domains. Marked differences emerge only in very strong gravitational fields or on cosmic scales. Early empirical successes of the theory depended on very small effects: minute deviations in the motion of the planet Mercury from that expected classically, a small deflection of starlight grazing the sun and the slight reddening of light from a massive star due to the gravitational slowing of the atom-clocks emitting the light.

Einstein's development of the theory over the period 1907-15 was inspired by the hope that a relativistic theory of gravitation might extend the principle of relativity to acceleration. He hoped the final theory would satisfy a principle he attributed to Ernst Mach: inertial forces acting on accelerating bodies are not due to acceleration through space but to an interaction with all other masses (see Relativity theory, philosophical significance of §6). He also postulated the principle of equivalence, which required that the inertial field of uniform acceleration is equivalent to a homogeneous gravitational field. While Einstein's general theory remains without serious challenge as a relativistic theory of gravitation and cosmology, debate continues over all his heuristic principles.

General relativity is 'generally covariant': it employs arbitrary spacetime coordinates. From 1913 to 1915, on the basis of his 'hole argument', Einstein had doubted that such freedom was physically admissible. He answered these doubts with the 'point coincidence argument'. It supposed that the physical content of his theory was fully exhausted by a catalogue of spacetime coincidences, which remain unaltered under arbitrary changes of coordinate system. As early as 1917, Erich Kretschmann objected that general covariance cannot express a generalized principle of relativity since, under the suppositions of the point coincidence argument, any spacetime theory can be recast in a generally covariant form (see Relativity theory, philosophical significance of §4).

One of Einstein's earliest applications of his new theory was his 1917 cosmological model. Its spatial geometry is spherical, so that space has a finite volume and no edge (see Cosmology §3). Einstein's motivation in devising the model was to satisfy Mach's principle, which he felt was violated by stipulating the limiting cosmic geometry at the infinity of space. From the 1920s until his death, Einstein's work on general relativity was increasingly absorbed into his attempts to produce a unified field theory which would treat not just gravitational but also electromagnetic fields geometrically.

4 Quantum theory: early contributions

By 1905 Einstein was convinced of the failure of classical electrodynamics to account for black body radiation, and blamed the wave picture of radiation. His light quantum hypothesis proposed that high-frequency heat radiation behaves thermodynamically as though its energy were spatially localized in quanta whose size depends on Planck's constant. This hypothesis was at the heart of his Nobel prize-winning explanation of the photo-electric effect. The following year, Einstein treated the non-classical behaviour of specific heats at low temperatures with a similar assumption (implicit in Planck's work of 1900) of discreteness for the energy levels of atoms (see Planck, M.).

Einstein never integrated these discontinuities into a coherent theory. Such a theory could not merely replace the wave picture by a particle view, for, in 1909, he had demonstrated that both wave and particle views together were required for a complete account of radiation fluctuations. His 1916 theory of the emission and absorption of radiation was largely independent of the classical picture. It associated momentum as well as energy with the

quanta and introduced stimulated emission, which governs the operation of lasers. His last contribution, before the quantum revolution of 1925-7, was to treat radiation as a quantum gas of indistinguishable particles by means of Einstein-Bose statistics.

Throughout, Einstein employed statistical methods, even representing emission and absorption of radiation as random processes. Yet he never broke with his belief that his treatment was defective since ‘it leaves the time and direction of elementary processes to "chance"’.

5 Quantum theory: critique of the standard interpretation

Einstein reacted strongly against complementarity, the positivist interpretation of the new quantum theory that became standard (see Bohr, N.; Logical positivism §4; Heisenberg, W.), in particular to its indeterminism and irrealism. Einstein felt that no essentially statistical theory could be fundamental in physics (his determinism) and that every fundamental theory should allow one to describe individual objects and processes independently of acts or conditions of observation (his realism). Moreover, he thought that the collapse of the quantum wave function introduced a kind of non-local action that could not be reconciled with relativity and that one could not recover classical mechanics as a suitable limiting case of quantum mechanics.

The impressive empirical success of quantum mechanics led him to envisage a better relativistic theory from which both the quantum theory and ordinary mechanics would emerge as limiting cases. He believed that this new (realist and deterministic) theory would not employ the dynamical variables of mechanics but would require different concepts. Thus Einstein’s vision was not that of a hidden variables’ extension of the quantum theory (see Quantum mechanics, interpretation of §3), as some have supposed. Instead, he worked to build a unified theory of the electromagnetic and gravitational fields that might be the starting point for the new physics. Toward the end of his life, with a unification programme not proceeding well, he may have abandoned his hope for a spacetime (’field’) theory, looking instead to what he sometimes referred to as an ’algebraical physics’.

Einstein’s scientific fame made him the quantum theory’s most important critic. He held that the theory was descriptively incomplete: that the quantum state function describes only ensembles of systems and, at best, provides an incomplete or partial description of the individual case. Einstein argued for this conclusion in a famous paper (’EPR’) co-authored with his then assistants, Podolsky and Rosen (1935). The structure of EPR is complex and has given rise to a whole literature attempting to identify its errors and challenge its assumptions (see Bell’s theorem §1). Ironically, it turns out that EPR was actually written not by Einstein but by Podolsky, and that Einstein himself was critical of Podolsky’s text, which he said ’buried’ the main point. In subsequent publications Einstein gave his own versions. One version assumes that spatially separated quantum systems possess ’real physical states’ (separability) and that these states are not immediately influenced by distant events (locality). Another version discharges these twin assumptions by focusing on macroscopic systems, for which separability and locality seem secure. This argument, developed in a correspondence that produced the similar ’cat paradox’ of Erwin Schrödinger, challenges whether the quantum theory can accommodate gross macroscopic events, like an explosion due to radioactive decay (see Quantum measurement problem).

6 Philosophy of science

The founders of Vienna Circle logical empiricism claimed Einstein as a supporter, arguing, for example, that Einstein’s critique of absolute distant simultaneity implicitly assumed an empiricist criterion of meaningfulness (see Logical positivism §2; Vienna Circle §3). Actually, from the 1920s, Einstein distanced himself from the anti-metaphysical tendencies of logical empiricism. Only in the 1960s, when logical empiricism’s influence waned, did a balanced interpretation of Einstein’s philosophy of science begin to emerge.

Einstein acknowledged Mach’s influence, especially Mach’s example of the historical and critical analysis of received scientific concepts. But from his early work on statistical physics to his late work on unified field theory, Einstein evinced no doubt about the reality of his underlying physical ontology, often representing proof of that reality as his aim, as in his claim that his work on statistical physics and Brownian motion aimed to establish the reality of atoms and molecules.

Einstein’s realism is not, however, the scientific realism of the later twentieth century, nor the convergentist realism of Peirce or Popper (see Scientific realism and antirealism §1). Einstein’s mature philosophy of science

combines a working physicist's realism with Duhemian underdetermination and conventionalism (see Duhem, P.M.M. §4). For Einstein, there is, in principle, no unique determination of theory by empirical evidence, although, in practice, theory choice seems uniquely determined. Because of this underdetermination in principle that Einstein believed pervaded all physical theory, he held that scientific theories are 'free creations of the human intellect', whose predictions must conform to empirical evidence. After the early 1930s, Einstein's rationalist sentiments grew, for he increasingly viewed mathematical or logical simplicity as a guide in choosing among empirically equivalent theories, although he doubted that the relevant sense of simplicity could be precisely defined since it involved 'a reciprocal weighing of incommensurable qualities'. How an underdetermined deep ontology can be considered 'real' is clarified by the recognition that, for Einstein, realism was less a general semantic thesis about the interpretation of all theories and more a demand that spatio-temporal theories satisfy a few, specifically physical, requirements, chiefly determinism and separability.

Einstein's conventionalist realism was refined during the 1920s as he, Schlick and Reichenbach sought to construct a new empiricism opposing the Neo-Kantianism of thinkers like Ernst Cassirer, who tried to save Kant's doctrine of the a priori in the face of general relativity's assertion that the global geometry of spacetime is non-Euclidean. Einstein disputed Schlick and Reichenbach's declaration that only coordinating definitions are conventional, arguing that the coordinating definition- empirical proposition distinction was arbitrary (see General relativity, philosophical responses to).

One of Einstein's most original contributions to scientific methodology was his distinguishing between 'principle theories' and 'constructive theories'. The former are based upon well-confirmed high-level empirical generalizations, like the light and relativity principles in special relativity, or the first and second laws of thermodynamics, which function as regulative principles constraining the choice of constructive models. Ultimate understanding comes via a constructive theory, but early investigations in new domains will progress more rapidly by first seeking principle theories to guide the search for constructive models.

See also Poincaré, H.; Space; Spacetime; Time

ARTHUR FINE
DON HOWARD
JOHN D. NORTON

List of works

Einstein's numerous works are available in a series of collections. The more accessible are referenced here.

Einstein, A. (1905a) 'Über einem die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt' (On a Heuristic Viewpoint Concerning the Production and Transformation of Light), *Annalen der Physik* 17: 132-48.(Translated by A.B. Arons and M. Peppard, (1965) *American Journal of Physics* 33: 367-74. The photo-electric effect and the light-quantum hypothesis. See §4.)

Einstein, A. (1905b) 'Zur Elektrodynamik bewegter Körper' (On the Electrodynamics of Moving Bodies), *Annalen der Physik* 17: 891-921.(First presentation of special relativity; Einstein's most famous paper. See §2.)

Einstein, A. (1907) 'Plancksche Theorie der Strahlung und die Theorie der spezifischen Warme' (Planck's Theory of Radiation and the Theory of Specific Heat), *Annalen der Physik* 22: 180-90, 800.(The first paper - and a correction - on specific heats. Submitted November 1906. See §4.)

Einstein, A. (1909) 'Zum gegenwaertigen Stand des Strahlungsproblems' (On the Present Status of the Radiation Problem), *Physikalische Zeitschrift* 10: 185-93.(Develops wave particle duality for radiation. See §4.)

Einstein, A. (1916a) 'Die Grundlage der allgemeinen Relativitätstheorie' (The Foundation of the General Theory of Relativity), *Annalen der Physik* 49: 769-822.(First review of complete general theory of relativity. See §3.)

Einstein, A. (1916b) 'Ernst Mach', *Physikalische Zeitschrift* 17: 101-4.(Obituary, with comments on Mach's philosophy. See §6.)

Einstein, A. (1917a) 'Quantentheorie der Strahlung' (Quantum Theory of Radiation), *Physikalische Zeitschrift* 18: 121-8.(Introduces momentum and spontaneous emission for radiation processes. Here Einstein expresses his reservation, cited in §4, about the breakdown of determinism.)

Einstein, A. (1917b) 'Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie' (Cosmological Considerations on the General Theory of Relativity), *Preussische Akademie der Wissenschaften, Sitzungsberichte* 1917: 142-52.(Einstein's first paper in relativistic cosmology. See §3.)

Einstein, A. (1918) 'Motive des Forschens' (Principles of Research), in *Zu Max Plancks sechzigstem Geburtstag. Ansprachen, gehalten am 26. April 1918 in der Deutschen Physikalischen Gesellschaft*, Karlsruhe: C.F. Müller,

29-32.(Distinguishes underdetermination in principle from in-practice univocal determination of theory by evidence. See §6.)

Einstein, A. (1919) 'Time, Space, and Gravitation', *Times* (London), 28 November: 13.(Distinguishes constructive and principle theories. See §6.)

Einstein, A. (1921) *Geometrie und Erfahrung, (Geometry and Experience) Erweiterte Fassung des Festvortrages gehalten an der Preussischen Akademie der Wissenschaften zu Berlin am 27. Januar 1921*, Berlin: Springer. (Discusses empirical interpretation of geometry and physics along holist and conventionalist lines. Describes theories as 'free creations of the human spirit'. See §6.)

Einstein, A. (1924) 'Quantentheorie des einatomigen idealen Gases' (Quantum Theory of Monatomic Ideal Gases), *Preussische Akademie der Wissenschaften, Sitzungsberichte* 1924: 261-7.(See 1925.)

Einstein, A. (1925) 'Quantentheorie des einatomigen idealen Gases 2' (Quantum Theory of Monatomic Ideal Gases 2), *Preussische Akademie der Wissenschaften, Sitzungsberichte* 1925: 3-14.(The two preceding papers treat Bose-Einstein statistics and condensation for a quantum gas. See §4.)

Einstein, A. (1933) *On the Method of Theoretical Physics. The Herbert Spencer Lecture, delivered at Oxford, 10 June 1933*, Oxford: Clarendon Press.(Einstein praises the heuristic of mathematical simplicity. See §6.)

Einstein, A. (1936) 'Physik und Realitaet' (Physics and Reality), *Journal of The Franklin Institute* 221: 313-47. (Most extensive discussion of structure and interpretation of scientific theories. One version of argument for incompleteness of quantum mechanics. Speculation about 'algebraic, alternative to continuum field theories'. See §§5 and 6.)

Einstein, A. (1948) 'Quanten-mechanik und Wirklichkeit' (Quantum Mechanics and Reality), *Dialectica* 2: 320-4. (Distinguishes locality and separability principles in argument for incompleteness of quantum mechanics. See §§5. and 6.)

Einstein, A. (1949) *Autobiographical Notes*, La Salle, IL: Open Court, 1979.(Einstein's only extensive autobiographical recollections, devoted almost exclusively to his scientific work.)

Einstein, A. (1954) *Ideas and Opinions*, New York: Crown.(A collection of nonscientific and popular scientific writings. Includes writings and translations of central importance to §6.)

Einstein, A. (1974) *Meaning of Relativity*, Princeton, NJ: Princeton University Press, 5th edn.(Einstein's textbook development of relativity theory.)

Einstein, A. (1977) *Relativity: The Special and the General Theory*, London: Methuen, 15th edn.(A popularization of relativity theory.)

Einstein, A. (1987-) *The Collected Papers of Albert Einstein*, J. Stachel, M. Klein, R. Schulmann, *et al*. (eds) Princeton, NJ: Princeton University Press.(Ongoing publication of a standard edition of Einstein's published and private papers with companion English translation volumes.)

Einstein, A. (1990) *Out of My Later Years*, New York: Bonanza.(A collection of nonscientific and popular scientific writings.)

Einstein, A. and Infeld, L. (1938) *The Evolution of Physics: The Growth of Ideas from the Early Concepts to Relativity and Quanta*, Cambridge: Cambridge University Press.(Survey of the historical development of physics.)

Einstein, A., Podolsky, B. and Rosen, N. (1935) 'Can Quantum Mechanical Description of Physical Reality be Considered Complete?', *Physical Review* 47: 777-80. (The 'EPR' paper, written by Podolsky, discussed in §5.)

References and further reading

Beller, M. and Fine, A. (1993) 'Bohr's Response to EPR', in J. Faye and H. Folse (eds) *Niels Bohr and Contemporary Philosophy*, Dordrecht: Kluwer, 1-31.(The impact of EPR on Bohr's complementarity. See §5.)

Cassirer, E. (1921) *Zur Einsteinschen Relativitaetstheorie. Erkenntnistheoretische Betrachtungen*, Berlin: Bruno Cassirer.(Neo-Kantian response to general relativity. See §6. Translation as second part of *Substance and Function and Einstein's Theory of Relativity*, New York: Dover, 1953.)

Clark, R.W. (1971) *Einstein: The Life and Times*, New York: World Publishing Company.(Most detailed of Einstein's many non-scientific biographies.)

Duhem, P. (1906) *La Theorie physique: son objet et sa structure*, Paris: Chevalier & Rivière.(Classic statement of Duhem's philosophy of science. See §6.) Translated as *The Aim and Stucture of Physical Theory*, Princeton, NJ: Princeton University Press, 1954.)

Fine, A. (1986) *The Shaky Game: Einstein, Realism and the Quantum Theory*, Chicago, IL: University of Chicago Press.(Realism in general and Einstein's realism (see §6) and expansion of material of §5.)

Furth, R. (ed.) (1961) *Investigations of the Theory of the Brownian Movement*, New York: Crown.(Translation by A.D. Cowper of five papers by Einstein on Brownian motion, written during 1905-8, with notes by R. Furth.)

Hoffmann, B. and Dukas, H. (1972) *Albert Einstein: Creator and Rebel*, New York: Viking Press.(Biography by individuals close to Einstein.)

Holton, G. (1968) ’Mach, Einstein, and the Search for Reality’, *Daedalus* 97: 636-73; repr. in *Thematic Origins of Scientific Thought: Kepler to Einstein*, Cambridge, MA: Harvard University Press, 1973, 219-59.(Significant early treatment of Einstein’s realism. Compare with later treatments in Fine 1986 and Howard 1993. See §6.)

Howard, D. (1985) ’Einstein on Locality and Separability’, *Studies in History and Philosophy of Science* 16: 171-201.(Discussion of role of locality and separability principles in Einstein’s argument for incompleteness of quantum mechanics. See §5.)

Howard, D. (1993) ’Was Einstein Really a Realist?’ *Perspectives on Science* 1: 204-51.(Argues for significant realist moment even in Einstein’s early philosophy of science, but distinguishes Einstein’s marriage of conventionalism, holism and realism from scientific realism of later twentieth century; examines role of physical principles in Einstein’s realism. See §6.)

Howard, D. (1994) ’Einstein, Kant, and the Origins of Logical Empiricism’, in W. Salmon and G. Wolters, (eds) *Logic, Language, and the Structure of Scientific Theories*, Pittsburgh, PA: University of Pittsburgh Press, and Konstanz: Universittsverlag, 45-105.(Traces divergent forms of realism developed by Einstein, Schlick and Reichenbach in responding to Neo-Kantians. See §6.)

Howard, D. and Stachel, J. (series eds) (1989-) *Einstein Studies*, Boston, MA: Birkhaueser.(Series devoted to Einstein and physics in his tradition. Volumes 1 (1989), 3 (1992) and 5 (1993) treat the history of general relativity.)

Lorentz, H.A., Einstein, A., Minkowski, H. and Weye, H. (eds) (1952) *The Principle of Relativity*, notes by A. Sommerfeld, trans. W. Perrett and G.B. Jeffrey, New York: Dover.(Collection of original papers in relativity theory by Einstein and others including the most-used English translations of *On the Electrodynamics of Moving Bodies* (1905) and *The Foundation of the General Theory of Relativity* (1916). See §§2 and 3.)

Miller, A. (1981) *Albert Einstein’s Special Theory of Relativity: Emergence (1905) and Early Interpretation (1905-1911)*, Reading, MA: Addison Wesley.(Detailed treatment of the early history of Einstein’s special theory.)

Nathan, O. and Norden, H. (1968) *Einstein on Peace*, New York: Schocken Books. (Comprehensive study of Einstein’s social and political activities and writings, with lengthy excepts from writings. See §1.)

Norton, J.D. (1993) ’General Covariance and the Foundations of General Relativity: Eight Decades of Dispute’, *Reports on Progress in Physics* 56: 791-858.(Review of the debate over the physical significance of general covariance. See §3.)

Pais, A. (1982) *Subtle is the Lord… The Science and the Life of Albert Einstein*, Oxford: Clarendon Press. (Biographical account of Einstein’s scientific work.)

Paty, M. (1993) *Einstein Philosophe: La Physique comme philosophie practique* (Einstein, Philosopher: Physics as Practical Philosophy), Paris: Presses Universitaires de France.(Comprehensive study of Einstein’s philosophy of science. See §6.)

Reichenbach, H. (1928) *Philosophie der Raum-Zeit-Lehre*, Berlin: Springer.(Classic study of philosophy of space and time in relativity theory from logical empiricist point of view. See §6. Translated as *Philosophy of Space and Time*, New York: Dover, 1958.)

Schilpp, P.A. (ed.) (1951) *Albert Einstein: Philosopher-Scientist*, New York, Tudor, 2nd edn.(Discussion of Einstein’s work by leading scientists and philosophers of Einstein’s later years; includes Einstein’s Autobiographical Notes, Einstein’s ’Reply to Criticisms’ and bibliography of Einstein’s writing.)

Schlick, M. (1922) *Raum und Zeit in den gegenwaertigen Physik. Zur Einfuehrung in das Verstaendnis der Relativitaets- und Gravitationstheorie*, Berlin: Springer, 4th edn.(Classic early analysis of general relativity with emphasis on its philosophical implications, written before Schlick’s turn to Vienna Circle logical empiricism. See §6. Translated as *Space and Time in Contemporary Physics: An Introduction to the Theory of Relativity and Gravitation*, New York: Oxford University Press, 1920.)

Seelig, C. (1960) *Albert Einstein. Leben und Werk eines Genies unserer Zeit* (Albert Einstein. Life and Work of a Genius of our Time), Zurich: Europa Verlag.(This biography draws on an extensive correspondence between the author, Einstein and others.)

Electrodynamics

Electric charges interact via the electric and magnetic fields they produce. Electrodynamics is the study of the laws governing these interactions. The phenomena of electricity and of magnetism were once taken to constitute separate subjects. By the beginning of the nineteenth century they were recognized as closely related topics and by the end of that century electromagnetic phenomena had been unified with those of optics. Classical electrodynamics provided the foundation for the special theory of relativity, and its unification with the principles of quantum mechanics has led to modern quantum field theory, arguably our most fundamental physical theory to date.

As James Clerk Maxwell (1831-79) pointed out in the preface to his *A Treatise on Electricity and Magnetism* (1873), the ancients were aware that certain bodies, such as amber, once rubbed, attract other bodies. These and similar phenomena are termed electric after the Greek for amber, *ēlectron*. Similarly, the lodestone has long been known to attract pieces of iron at a distance. These and related phenomena are termed magnetic after Magnesia (Greek, *magnēs*), an ancient city in Asia Minor where the lodestone was found. A full historical background to electrodynamics would include the development of electric circuit theory and of optics (see Optics).

The modern science of electricity and magnetism can be conveniently dated by the publication in 1600 of William Gilbert’s (1540-1603) great work on magnetic and electric phenomena. To him we owe the terms ’electric attraction’, ’electrical force’ and ’magnetic pole’. Gilbert associated electricity with a tenuous effluvium (or vapour) that was liberated from a body by the action of friction. René Descartes (1596-1650) attempted to explain magnetism in terms of vortices in a fluid that surrounded magnets. These ideas were consonant with a corpuscular theory of light and the caloric theory of heat. The electric effluvium was even conjectured to be associated with the effect of gravity. Such a plenum or ether would play a central role in the development of physical theory into the beginning of the twentieth century.

Stephen Gray (d. 1736) and Benjamin Franklin (1706-90) were among those who postulated various qualitative theories based on electrical fluids. Joseph Priestley (1733-1804), using an analogy with Newton’s law of gravitation, suggested an inverse square law of attraction or repulsion between electric charges. Although there were earlier approximate verifications, by 1785 Charles-Augustin de Coulomb (1736-1806) made a fairly accurate direct experimental test of this law, which now bears his name.

A concise mathematical formulation of electrostatic theory was presented in 1812 by Siméon-Denis Poisson (1781-1840). Both he and Pierre-Simon Laplace (1749-1827) recognized that the vector electrostatic force (like the gravitational one) could be calculated from the potential energy of a system of charges. By 1828, the theory of the electric potential had been formalized by George Green (1793-1841). In 1820 Hans Christian Ørsted (1777-1851) discovered a new type of effect that arose when charges were set in motion to produce an electric current. These magnetic effects were soon given precise mathematical formulations by Jean-Baptiste Biot (1774-1862), Félix Savart (1791-1841) and, most elegantly, by André-Marie Ampère (1775-1836). In 1845, Franz Neumann (1798-1895) discovered a vector potential function for the magnetic field. Action at a distance remained the dominant mode of thought in this area until the time of Michael Faraday (1791-1867), who introduced the concept of electric and magnetic fields. Faraday’s field, and his 1831 law of electromagnetic induction, formed the basis for Maxwell’s unification of electric and magnetic phenomena. Maxwell’s theory demonstrated that electric and magnetic phenomena are governed by a common set of equations and that optical phenomena are electromagnetic in nature (see Field theory, classical; Maxwell, J.C. §2). Heinrich Hertz (1857-94) contributed significantly to the acceptance of Maxwell’s theory, not only by experimental verification in the 1880s of the electromagnetic nature of light, but also, in the 1890s, with a series of papers that gave a concise presentation of the axioms of Maxwell’s theory and a clear explanation of electromagnetic phenomena by rigorous deduction from those principles. Hertz’s often-quoted dictum that the essence of Maxwell’s theory is the equations themselves would become a fair characterization of the attitude of one school of modern physicists toward theories in general (see Hertz, H.). Even into the twentieth century ‘continental’-style physicists prized theories of generality and mathematical abstractness, whereas models of specific phenomena were favoured in Great Britian.

So dominant had been Newtonian mechanics and the mechanical worldview, that even Maxwell and many of his

contemporaries attempted to underpin electromagnetic phenomena with mechanical models. This was the programme of the electromagnetic ether, brought to its most complete mathematical formulation by Hendrik Lorentz (1853-1928) at the end of the nineteenth century. In fact, electromagnetic phenomena came to play such a central role in physics that, in the early years of the twentieth century, attempts were made to replace the mechanical worldview with an electromagnetic one, whereby mechanical properties such as mass were accounted for in terms of electromagnetic properties such as charge. It was out of these foundational struggles in electrodynamics that Albert Einstein's special theory of relativity emerged in 1905 (see Einstein, A. §2). Subsequent efforts by Paul Dirac (1902-84) to apply the principles of quantum mechanics to electrodynamics led in 1928 to the first formulation of quantum electrodynamics and, through the work of Werner Heisenberg (1901-76) and Wolfgang Pauli (1900-58), which began in 1929, to modern quantum field theory. Electrodynamics was central in quantum field theory coming to the fore and in replacing the old particle-field dualism with the quantum-field ontology (see Field theory, quantum).

See also: Mechanics, classical; Relativity theory, philosophical significance of

JAMES T. CUSHING

References and further reading

Buchwald, J.Z. (1985) *From Maxwell to Microphysics*, Chicago, IL: University of Chicago Press.(A history of electromagnetic theory during the last quarter of the nineteenth century.)

Faraday, M. (1839-55) *Experimental Researches in Electricity*, London: R. & J.E. Taylor.(An exhaustive account of Faraday's seminal work in electricity and magnetism.)

Gilbert, W. (1600) *De Magnete*, London: Peter Short.(An early study of magnetism.)

Jackson, J.D. (1975) *Classical Electrodynamics*, New York: John Wiley & Sons, 2nd edn.(An authoritative modern textbook on the technical aspects of the subject.)

Maxwell, J.C. (1873) *A Treatise on Electricity and Magnetism*, Oxford: Clarendon Press.(The classic text on electromagnetic theory.)

Sommerfeld, A. (1952) *Electrodynamics*, New York: Academic Press.(A technical presentation, accompanied by useful comments on the history of the subject.)

Tricker, R.A.R. (1965) *Early Electrodynamics: The First Law of Circulation*, London: Pergamon Press.(An elementary and detailed account of the discovery of the basic laws governing the interactions between electric currents and magnetic fields.)

Whittaker, E. (1973) *A History of the Theories of Aether and Electricity*, New York: Humanities Press.(A classic on the history of the subject from antiquity to the discovery of quantum mechanics.)

Eliade, Mircea (1907-86)

Eliade was educated as a philosopher. He published extensively in the history of religions and acted as editor-in-chief of Macmillan's Encyclopedia of Religion (1987). The influence of his thought, through these works and through thirty years as director of the history of religions department at Chicago University, is considerable.

Eliade's analysis of religion assumes the existence of 'the sacred' as the object of worship of religious humanity. It appears as the source of power, significance and value. Humanity apprehends 'hierophanies' - physical manifestations or revelations of the sacred - often, but not only, in the form of symbols, myths and rituals. Any phenomenal entity is a potential hierophany and can give access to nonhistorical time, what Eliade calls illud tempus *('that time'). The apprehension of this sacred time is a constitutive feature of the religious aspect of humanity.*

1 Life

Eliade was born in Bucharest, Romania. Despite a childhood interest in entomology and botany (which doubtless first attracted his attention to Goethe, a lifelong role model and inspiration), he developed an interest in world literature and was led from there to philology, philosophy and comparative religion. As a youth he read extensively in Romanian, French and German, and around 1924-5 he learned Italian and English to read Raffaele Pettazzoni and James George Frazer in the original.

In 1925 Eliade enrolled at the University of Bucharest, where he studied in the department of philosophy. The influence of Nae Ionescu, then an assistant professor of logic and metaphysics and an active journalist, was keenly felt by the young Eliade and the shadow which fell on the older scholar because of his involvement with the extreme right in interwar Romania has darkened Eliade's reputation.

Eliade's Master's thesis examined Italian Renaissance philosophers from Marcilio Ficino to Giordano Bruno, and Renaissance humanism was one of his major influences when he turned to India in order to 'universalize' the 'provincial' philosophy he had inherited from his European education. Finding that the Maharaja of Kassimbazar sponsored European scholars to study in India, Eliade applied and was granted an allowance for four years. In 1928 he sailed for Calcutta to study Sanskrit and philosophy under Surendranath Dasgupta, a professor at the university there.

He returned to Bucharest in 1932 and successfully submitted an analysis of yoga as his doctoral thesis in 1933, which was published in French as *Yoga: essai sur les origines de la mystique Indienne (Yoga: A Treatise on the Origins of Indian Mysticism)*. This was extensively revised and republished as *Yoga, Immortality, and Freedom* (1958a). As Ionescu's assistant, Eliade lectured on, among other things, Aristotle's *Metaphysics* and Nicholas of Cusa's *Docta ignorantia*. From 1933 to 1939 he was active with the Criterion group, which gave public seminars on a wide range of topics. They were strongly influenced by the philosophy of 'trairism' the search for the 'authentic' in and through lived experience (Romanian, *traire*), which was seen as the only source of 'authenticity'.

After the Second World War, during which he served with the Romanian Legation in the UK and Portugal, Eliade was unable to return to the newly communist Romania because of his connection with the right-wing Ionescu. In 1945 he moved to Paris, where his acquaintance with George Dumézil, an important scholar of comparative mythology, secured him a part-time post teaching comparative religion at the École des Hautes Études at the Sorbonne. From this time on almost all of Eliade's scholarly works were written in French.

At the prompting of Joachim Wach, director of the history of religions department at the University of Chicago, Eliade was invited to give the 1956 Haskell Lectures on 'Patterns of Initiation' at Chicago. These were later published as *Rites and Symbols of Initiation* (1958b). In 1958 Eliade was invited to assume the chair of the history of religions department at Chicago. There he stayed until his death in 1986, publishing extensively and writing largely unpublished fiction. He also launched the journals *History of Religions* and *The Journal of Religion*, and acted as editor-in-chief for Macmillan's *Encyclopedia of Religion* (1987).

2 Thought

Despite his focus on the history of religions, Eliade never relinquished his philosophical agenda. That said, he never fully clarified his philosophy. There has been radical disagreement over his thought, some seeing it as a crucial contribution to the study of religion, and some seeing him as an obscurantist whose normative assumptions are unacceptable.

In *Cosmos and History: The Myth of the Eternal Return* (1954), a book which he was tempted to subtitle *Introduction to a Philosophy of History*, Eliade distinguishes between religious and nonreligious humanity on the basis of the perception of time as heterogeneous and homogeneous respectively. This distinction will be immediately familiar to students of Henri Bergson (§2) as an element of that philosopher's analysis of time and space. Eliade contends that the perception of time as a homogeneous, linear and unrepeatable medium is a peculiarity of modern and nonreligious humanity. Archaic or religious humanity (*homo religiosus*), in comparison, perceives time as heterogeneous; that is, as divided between profane time (linear) and sacred time (cyclical and reactualizable). By means of myths and rituals which give access to this sacred time, religious humanity protects itself against the 'terror of history', a condition of helplessness before the absolute *data* of historical time, a form of existential anxiety.

In the very process of establishing this distinction, however, Eliade undermines it, insisting that nonreligious humanity in any pure sense is a very rare phenomenon. Myth and *illud tempus* are still operative, albeit concealed, in the world of modern humanity and Eliade clearly regards the attempt to restrict real time to linear historical time as finally self-contradictory. He squarely sets himself against the historicism of Hegel (§8).

'The sacred' has also been the subject of considerable contention. Some have seen Eliade's 'sacred' as simply corresponding to a conventional concept of deity, or to Rudolf Otto's *ganz andere* (the 'wholly other'), whereas others have seen a closer resemblance to Émile Durkheim's socially influenced sacred (see Otto, R.; Durkheim, É. §4). Eliade himself repeatedly identifies the sacred as the real, yet he states clearly that 'the sacred is a structure of human consciousness' (1969: i; 1978: xiii). This would argue more for the latter interpretation: a social construction of both the sacred and of reality. Yet the sacred is identified as the *source* of significance, meaning, power and being, and its manifestations as hierophanies, cratophanies or ontophanies accordingly (appearances of the holy, of power or of being). Corresponding to the suggested ambiguity of the sacred itself is the ambiguity of its manifestations.

Eliade states that believers for whom the hierophany is a revelation of the sacred must be prepared by their experience, including their traditional religious background, before they can apprehend it. To others the sacred tree, for example, remains simply a tree. It is an indispensable element of Eliade's analysis that any phenomenal entity could be apprehended as a hierophany with the appropriate preparation. The conclusion must be that all beings reveal, and at the same time conceal, the nature of Being. A reprise of Nicholas of Cusa's 'coincidence of opposites' is evident here, as is a possible explanation of the systematic ambiguity of Eliade's writings (see Nicholas of Cusa §2).

Finally, religion, systematically understood as the apprehension of relative worth conferred through nonhistorical realities (including all abstract and imaginary entities), but revealed and confirmed through historical phenomena, is seen as a unifying human universal. It is characteristic of Eliade's style of writing, both in his fictional and in his academic work, that this conclusion is nowhere clearly stated. Leading assertions are scattered throughout his publications on the history of religions, alchemy, symbolism, initiation, myth and so on inviting his readers either to make an immediate interpretation or to pursue the question further into the thicket of his *oeuvre*.

See also: Mysticism, nature of; Phenomenology of religion; Religious experience §1

BRYAN STEPHENSON RENNIE

List of works

Eliade, M. (1954) *Cosmos and History: The Myth of the Eternal Return*, trans. W. Trask, Princeton, NJ: Princeton University Press.(Probably Eliade's most crucial and approachable short work. Contains his analysis of heterogeneous and homogeneous time, and his conception of the 'terror of history' and the ability to 'reactualize' religious time.)

Eliade, M. (1958a) *Yoga, Immortality and Freedom*, trans. W. Trask, London: Routledge & Kegan Paul.(First published in French as *Yoga: essai sur l'origine de la mystique Indienne* in 1933, this informative and scholarly

work analyses yoga as a concrete search for freedom from human limitations.)
Eliade, M. (1958b) *Rites and Symbols of Initiation (Birth and Rebirth)*, trans. W. Trask, London: Harvill Press. (Eliade's 1956 Haskell Lectures at the University of Chicago. His analysis of initiatory themes implies their ubiquity and structure as a symbolic death and rebirth.)
Eliade, M. (1958c) *Patterns in Comparative Religion*, trans. R. Sheed, London: Sheed & Ward.(An attempt to delineate the morphology of the sacred, this is frequently criticized for its cross-cultural and ahistorical approach, which organizes religious phenomena by structural similarities regardless of time or place of origin. A valuable source of data nevertheless.)
Eliade, M. (1959) *The Sacred and the Profane: The Nature of Religion*, trans. W. Trask, London: Harcourt Brace Jovanovich.(Picking up where Otto's *The Idea of the Holy* left off, the sacred is explicated through its relation to its binary counterpart, the profane, and the complex dialectic of the two is outlined.)
Eliade, M. (1960) *Myths, Dreams and Mysteries: The Encounter between Contemporary Faiths and Archaic Realities*, trans. P. Mairet, London: Harvill Press.(Eliade's understanding of myth in the modern world and the mythic prestige of origins, and his analysis, among other things, of the symbolism of ascension, flight, the labyrinth and swallowing by a monster.)
Eliade, M. (1961) *Images and Symbols: Studies in Religious Symbolism*, trans. P. Mairet, London: Harvill Press. (More on symbolism, particularly the symbolism of the centre, knots, shells and pearls. Symbolism and history are discussed, as is method.)
Eliade, M. (1963) *Myth and Reality*, trans. W. Trask, New York: Harper & Row.(The structure of myths. More on the prestige of origins and on the survival of myths and mythic themes in modern thought.)
Eliade, M. (1964) *Shamanism: Archaic Techniques of Ecstasy*, trans. W. Trask, London: Routledge & Kegan Paul.(Long a standard work in the study of shamanism, this is a detailed and valuable source of information on the phenomenon.)
Eliade, M. (1965) *The Two and the One*, trans. J.M. Cohen, Chicago, IL: University of Chicago Press.(An important analysis of the *coincidenta oppositorum* (coincidence of opposites), or binary oppositions, in religious ideas. Androgyny is explored, as are cosmogony and eschatology, the birth and death of the cosmos or worldview.)
Eliade, M. (1969) *The Quest: History and Meaning in Religion*, London: University of Chicago Press.(This attempt at a more methodological work pulls together articles previously published on Eliade's methodological and theoretical presuppositions, including his 'new humanism' and his response to the quest for the 'origins' of religion.)
Eliade, M. (1978) *A History of Religious Ideas*, vol. 1, *From the Stone Age to the Eleusinian Mysteries*, trans. W. Trask, Chicago, IL: University of Chicago Press.(Originally projected as a complete history of religion in one volume, this was an attempt to give Eliade's understanding of the entire history of religion from a unified perspective. It is a useful reference work, potentially readable in its entirety. Many of Eliade's categories survive in this mature work: the terror of history, the *coincidenta oppositorum*, the symbolism of the centre, and the *hieros gamos*, or symbolic heavenly marriage.)
Eliade, M. (1982) *A History of Religious Ideas*, vol. 2, *From Gautama Buddha to the Triumph of Christianity*, trans. W. Trask, Chicago, IL: University of Chicago Press. (See note to volume 1 (1978).)
Eliade, M. (1985) *The History of Religious Ideas*, vol. 3, *From Muhammad to the Age of the Reforms*, trans. A. Hiltebeitel and D. Apostolos-Cappadona, Chicago, IL: University of Chicago Press.(See note to volume 1 (1978).)
Eliade, M. (ed.) (1987) *Encyclopedia of Religion*, New York: Macmillan. (Sixteen volumes of articles on every aspect of religion by leading scholars in the field. Currently the standard reference encyclopedia on religion.)

References and further reading

Allen, D. (1978) *Structure and Creativity in Religion: Hermeneutics in Mircea Eliade's Phenomenology and New Directions*, The Hague: Mouton.(An interesting philosophical study of Eliade's thought that perhaps overemphasizes Eliade's methodology as phenomenology of religion.)
Allen, D. and Doeing, D. (1980) *Mircea Eliade: An Annotated Bibliography*, New York and London: Garland.(An excellent polyglot bibliography, although a definitive posthumous bibliography is in preparation by M.L. Ricketts and M. Handoca.)
Baird, R.D. (1971) 'Phenomenological Understanding: Mircea Eliade', in *Category Formation and the History of Religion*, The Hague: Mouton, 74-91.(A critique of Eliade as making unwarranted ontological assumptions and normative judgments that should, however, be verified against a thorough reading of the primary sources.)
Cave, J.D. (1992) *Mircea Eliade's Vision for a New Humanism*, Oxford: Oxford University Press.(An accessible

interpretation of Eliade’s work in terms of humanism, which was one of Eliade’s major foci.)

Dudley, G., III (1977) *Religion on Trial: Mircea Eliade and His Critics*, Philadelphia, PA: Temple University Press.(This readable and broadly favourable work reconceives Eliade’s thought in terms of a research programme for the history of religion.)

Idinopulos, T.A. and Yonan, E. (eds) (1994) *Religion and Reductionism: Essays on Eliade, Segal, and the Challenge of the Social Sciences for the Study of Religion*, Leiden: E.J. Brill.(A variety of challenging articles on Eliade’s status, and implications for the study of religion. Of particular interest is W. Paden’s ‘Before "The Sacred" Became Theological: Rereading the Durkheimian Legacy’.)

Olson, C. (1992) *The Theology and Philosophy of Eliade*, New York: St Martin’s Press.(A general and readable consideration of Eliade’s thought.)

Rennie, B. (1996) *Reconstructing Eliade: Making Sense of Religion*, New York: State University of New York Press.(A thorough explication of Eliade’s work, which assumes that it has an internal coherence for the critic to uncover; includes an extensive bibliography.)

Ricketts, M.L. (1988) *Mircea Eliade: The Romanian Roots*, New York: Columbia University Press, 2 vols.(A massive and thorough piece of research into Eliade’s life, thought and works up until 1945; a useful bibliography is provided.)

Strenski, I. (1987) ‘Mircea Eliade’, in *Four Theories of Myth in Twentieth-Century History*, London: Macmillan. (One of the earliest critiques to raise Eliade’s political background; it attempts to trace the influence of that background in his theoretical constructs.)

Eliot, George (1819-80)

George Eliot is the pseudonym of the English writer Mary Ann Evans, whose mind was strongly influenced by the main philosophical currents of the time and who made a distinctive contribution of her own through her critical essays and fiction.

Although George Eliot was early seen to be intellectually gifted, she received little education of an institutional kind after leaving school at 16; from the age of nine, however, she was a prodigious reader, her appetite for learning being strengthened further by the religious conflicts of her youth. After a phase of intense Evangelical devotion she translated Strauss' *Life of Jesus* and turned to humanists such as Spinoza, finding Wordsworth's tempering of pantheism with moral insistence even more congenial (see Strauss, D.F. §1). She was well versed in traditional philosophy and at an early stage spoke of her aspiration that she might 'live to reconcile the philosophies of Locke and Kant'. Her primary concern, however, was with scientific naturalism and its implications. Her spell as assistant editor of the *Westminster Review* from 1851 to 1854 and then as contributor gave her the opportunity to meet and debate with some of the leading writers of the day and to write on topics such as human physiology and critical approaches to the Bible. She worked intimately with Herbert Spencer for a time, respecting him particularly for his acknowledgment that the universe was ultimately unknowable. She admired Darwin's *Origin of Species* but declared herself to be more impressed by 'the mystery that underlies the processes'. She was also fascinated by the associationist psychology of Alexander Bain, its leading exponent, and by the positivism of Auguste Comte, of whom G.H. Lewes, with whom she lived from 1854, was a devoted if critical follower. Her strongest sympathies were perhaps with the intellectual position of Ludwig Feuerbach, who saw religion as essentially an anthropological phenomenon, with Christianity providing the key to the subjective and moral nature of human beings (see Feuerbach, L. §2).

George Eliot thought of herself primarily as an aesthetic teacher, arguing at the same time that 'if such teaching lapses anywhere from the picture to the diagram it becomes the most offensive of all teaching'. Feuerbach's example provided an attractive way of viewing her own role as a novelist, using her longer fictions both as a repository for observations tested by experience and a crucible for the exposition and testing of her own ideas. In *Middlemarch* (1871-2) she carried on this process in a study of provincial life forty years before she was writing, when the implications of the new biblical criticism and developments in the study of physiology were throwing former certainties into doubt, yet also creating exciting prospects. In her most complex fiction, *Daniel Deronda* (1876), current work in physiology and psychology is caught up into the detail of her human observations, while her aspirations towards moral single-mindedness are reflected in the hero's embracing of the Zionist ideal. Her attempts to find a stance for him which could combine the cool rationality of an empiricist with scope for his fiery ardour suggests the ultimate contradiction within which her moral thinking was wrought.

JOHN BEER

List of works

Eliot, G. (1955, 1978) *The George Eliot Letters*, ed. G.S. Haight, New Haven, CT, and London: Yale University Press, vols 1-7, 1955; 8-9, 1978.(The comprehensive edition of the letters, with some extracts from the letters and journals of G.H. Lewes.)

Eliot, G. (1984) *Daniel Deronda*, ed. G. Handley, Oxford: Clarendon Press.(The standard edition of this novel.)

Eliot, G. (1986) *Middlemarch*, ed. D. Carroll, Oxford: Clarendon Press.(The standard edition of this novel.)

Byatt, A.S. and Warren, N. (eds) (1990) *Selected Essays, Poems and Other Writings*, Harmondsworth: Penguin. (The best selection from the essays, poems and other writings, with an introduction by A.S. Byatt.)

References and further reading

Ashton, R. (1980) *The German Idea: Four English Writers and the Reception of German Thought*, Cambridge: Cambridge University Press.(Chapter 4 deals with George Eliot.)

Beer, G. (1986) *George Eliot*, London: Routledge; Bloomington, IN: Indiana University Press.(On her feminism.)

Dodd, V.A. (1990) *George Eliot: An Intellectual Life*, London: Macmillan.(An account of her intellectual development.)

Haight, G.S. (1968) *George Eliot: A Biography*, Oxford and New York: Oxford University Press.(The standard

biography.)
Karl, F. (1995) *George Eliot: A Biography*, London: HarperCollins.(A fuller and updated account.)
Myers, W. (1984) *The Teaching of George Eliot*, Leicester: Leicester University Press.(A good general guide.)
Newton, K.M. (1981) *George Eliot: Romantic Humanist*, London: Macmillan; Totowa, NJ: Barnes & Noble.(On her Romantic heritage as a thinker.)
Paris, B. (1965) *Experiments in Life: George Eliot's Quest for Values*, Detroit, MI: Wayne State University Press. (The impact of empirical philosophy on George Eliot's thought.)
Shaffer, E.S. (1975) *'Kubla Khan' and The Fall of Jerusalem*, Cambridge: Cambridge University Press.(Chapter 6 is 'George Eliot and the Higher Criticism of the Bible'.)
Shuttleworth, S. (1984) *George Eliot and nineteenth-century science: the make-believe of a beginning*, Cambridge: Cambridge University Press.(On George Eliot's attitude to physiology.)
Strauss, D.F. (1835-6) *Das Leben Jesu*, Tübingen: Osiander; trans. M.A. Evans (George Eliot, 1846), *Life of Jesus*, Philadelphia, PA: Fortress, 1972.(Strauss's 'quest' for the historical Jesus.)
Wright, T.R. (1986) *The Religion of Humanity: The Impact of Comtean Positivism on Victorian Britain*, Cambridge: Cambridge University Press.(Chapter 5 deals with George Eliot and positivism.)

Elisabeth of Bohemia (1618-80)

Elisabeth of Bohemia, Princess Palatine, exerted an influence on seventeenth-century Cartesianism via her correspondence with Descartes. She questioned his accounts of mind-body interaction and free will, and persuasively argued that certain facts of embodiment, the unlucky fate of loved ones, and the demands of the public good, constitute serious challenges to Descartes' neo-Stoic view of the happy life of the autonomous will.

Eldest daughter of the exiled 'Winter King and Queen' of Bohemia, Princess Elisabeth was educated at the Prisenhof in Leiden. In 1644, Descartes dedicated his *Principles of Philosophy* to her, and two years later she attempted to introduce Descartes' work at the German courts. In 1670, as Abbess at Herford, she provided a haven for her correspondent, Anna Maria van Schurman, and the persecuted Labadists. She corresponded and met with Quaker leaders William Penn and Robert Barclay. Near the end of her life, she was interested in the views of Malebranche and the mystical philosopher, Jacob Boehme; she corresponded with Leibniz via her sister Louise, Abbess of Maubuisson in France.

On 6 May 1643, Elisabeth began her correspondence with Descartes by arguing that voluntary motion is unintelligible in terms of Descartes' mechanical philosophy. For, determination of movement is due to (1) the impulsion of the moved object, (2) the manner in which the mover impels the object, and (3) the qualification and figure of the surface of the moving object. (1) and (2) require contact; (3) requires extension. But souls are not extended and contact seems incompatible with their immateriality. Elisabeth urges that we need a more comprehensive definition of 'the soul' than simply a 'thinking thing'; we need to know what other properties the soul has, if we are to make its causal powers intelligible. She notes that while we suppose thought to be essential to the soul, this is difficult to prove in the cases of foetuses and deep faints - a challenge to Cartesianism that the Empiricists subsequently developed. To Descartes' rejoinder that we have a *per se* intelligible primitive notion, known by sense, of how an incorporeal soul can move a corporeal body, Elisabeth - prefiguring the position of Locke and Hume - responds that she has no such notion. She acknowledges that the senses show that the soul moves the body, but she denies that they, any more than imagination or understanding, show *how* this happens. She suggests that there might be unknown properties in the soul - even extension, which might belong to a faculty like sensation.

It is Elisabeth's personal problems, especially 'the weakness of my sex' (poor health due to afflictions of the soul) and the suffering of her family, that lead to discussion of *De Vita Beata (On the Happy Life)* by Seneca. Descartes' neo-Stoic solution requires that happiness depend only on what follows from the will guided by reason. But Elisabeth finds that her passions cannot be *immediately* controlled by her will and this often leads to her body becoming disordered. In addition, there are illnesses that diminish or deprive us of our reasoning power. Finally, regret, 'one of the principal obstacles to happiness', appears unavoidable when attempting to weight one's own goals along with the competing goods of others in society: 'To know all those [goods] about which one is constrained to make a choice in an active life, it would be necessary to possess infinite knowledge' (13 September 1645). Descartes replies that while we cannot know everything, we need only follow certain neo-Stoic principles for a happy life. One such principle is: the infallibility of God's decrees teaches us to accept in good spirit everything that happens to us, since it comes from God. Elisabeth replies that this principle does not help in the case of free decisions of the will; in such cases God's predetermination is not intelligible. She remained puzzled by Descartes' retort that the independence which we feel, which makes our actions sanctionable, is not incompatible with the dependence all things have on God. She argues that it is 'as impossible for the will to be at the same time free and attached to the decrees of Providence, as for the divine power to be both infinite and limited' (30 November 1645). Elisabeth's focus is on the will's liberty of indifference, which Descartes had stressed in his *Principles*, and its incompatibility with determination of any kind.

Some recent commentators have taken Elisabeth's use of her own body and female social role in her criticisms of the privileged, disembodied will as prefiguring contemporary feminist critiques of Cartesianism (Bordo 1998; Harth 1992; Thompson 1983).

See also: Stoicism

EILEEN O'NEILL

List of works

Elisabeth left no published treatises and her firm wish was that her letters should not be published. In consequence, none of her letters was published until 1862, and to date they have not been collected in a single volume; they may be found in the published works of her correspondents.

Barclay, R. (1812) *A Genealogical Account of the Barclays of Urie*, London: J. Herbert.

Barclay, Colonel D. (1870) *Reliquiae Barclaianae: Correspondence of Colonel D. Barclay and Robert Barclay of Urie and his son Robert, including Letters from Princess Elisabeth of the Rhine, the Earl of Perth, the Countess of Sutherland, William Penn, George Fox and others*, London: Winter & Bailey.(Contains Elisabeth's 1676-9 correspondence in English with Robert Barclay about religious matters and political events affecting the Quakers; also contains one letter in English from Elizabeth to Benjamin Furly.)

Blom, J. (ed. and trans.) (1978) *Descartes: His Moral Philosophy and Psychology*, New York: New York University Press.(Contains a translation of fourteen philosophical letters of Elisabeth to Descartes and his responses.)

Chevalier, J. (1935) *Lettres sur la morale: Correspondance avec la princesse Elisabeth, Chanut, et la reine Christine*, Paris: Boivin.

Descartes, R. (1936-63) *Descartes: Correspondance publiée avec une introduction et des notes*, ed. C. Adam and G. Milhaud, Paris: Felix Alcan and Presses Universitaires.

Descartes, R. (1964-74) *Oeuvres de Descartes*, vols 3, 4 and 5, ed. C. Adam and P. Tannery, nouvelle présentation, Paris: J. Vrin.(The standard edition of the Elisabeth-Descartes correspondence.)

Descartes, R. (1978) *Descartes: His Moral Philosophy and Psychology*, ed. and trans. J. Blom, New York: New York University Press.(Contains English translations of fourteen philosophical letters of Elisabeth to Descartes and his responses.)

Foucher de Careil, L. (ed.) (1879) *Descartes et la Princesse Elisabeth*, Paris: Germer-Baillière.(The earliest edition of the Descartes-Elisabeth correspondence, followed by the publications listed as Foucher de Careil 1909, Chevalier 1935 and Descartes 1936-63.)

Foucher de Careil, L. (1909) *Descartes, la princesse Elisabeth, et la reine Christine d'apres des lettres inedites…*, Paris: Felix Alcan.

Hauck, K. (ed.) (1908) *Die Briefe der Kinder des Winterkönigs (The letters of the children of the Winter King)*, Heidelberg: G. Koester.(Letters of Elisabeth and her siblings, which provide historical information about her family.)

Malebranche, N. (1958-84) *Oeuvres complètes de Malebranche*, vols 18 and 19, ed. A. Robinet, Paris: J. Vrin. (Contains synopses of the lost Elisabeth-Malebranche correspondence as well as discussion of Elisabeth in the Leibniz-Malebranche correspondence.)

References and further reading

Bordo, S. (ed.) (1998) *Rereading the Canon: Descartes*, University Park, PA: Pennsylvania State University Press.(Contains several essays exploring Elisabeth's philosophical contributions.)

Foucher de Careil, L. (1862) *Descartes et la Princesse Palatine, ou de l'influence du cartésianisme sur les femmes au XVIIe siècle (Descartes and the Princess Palatine, or on the Cartesian influence seventeenth-century women)*, Paris: Auguste Durand.(One of the early discussions of Elisabeth and the *Cartésiennes*.)

Garber, D. (1983) 'Understanding Interaction: What Descartes Should Have Told Elisabeth', *Southern Journal of Philosophy* 21, supplement: 15-32.(Examines Descartes's use of the 'primitive notion of union' to answer Elisabeth's worries about mind-body interaction.)

Godfrey, E. (1909) *A Sister of Prince Rupert: Elizabeth Princess Palatine and Abbess of Herford*, London and New York: John Lane.(Written under a pseudonym by Jessie Bedford, this is one of the earliest book-length biographies of Elisabeth in English.)

Harth, E. (1992) *Cartesian Women: Versions and Subversions of Rational Discourse in the Old Regime*, Ithaca, NY: Cornell University Press.(Focuses on Elisabeth's criticisms of Descartes' neo-Stoic view of the will and examines gender issues in their correspondence. Pages 67-78 are especially relevant.)

Néel, M. (1946) *Descartes et la princesse Elisabeth*, Paris: Éditions Elzévir.(One of the few treatments of the Descartes-Elisabeth correspondence that focuses on Elisabeth's contributions.)

O'Neill, E. (1983) *Mind and Mechanism: An Examination of Some Mind-Body Problems in Descartes' Philosophy*, Ph.D. diss., Princeton University, Ann Arbor, MI: University Microfilms.(Contains a discussion of the Descartes-Elisabeth correspondence on mind-body interaction and free will versus divine providence.)

Penn, W. (1981) *The Papers of William Penn*, vol. 1, ed. M. Dunn and R. Dunn, Philadelphia, PA: University of Pennsylvania Press.(Includes biographical information about Elisabeth's meetings with Quaker and Labadist leaders.)

Thompson, J. (1983) 'Women and the High Priests of Reason', *Radical Philosophy* 34: 10-14.(A feminist critique of Descartes' picture of rationality, in which Elisabeth's letters are read as an early version of this line of criticism.)

Zedler, B. (1989) 'The Three Princesses', *Hypatia* 4 (1): 28-63.(Provides bibliographies and a discussion of the philosophical correspondences of Elisabeth, her sister Sophie, Electress of Hanover, and the latter's daughter, Sophie Charlotte, Queen of Prussia.)

Emerson, Ralph Waldo (1803-82)

The American philosopher and poet Ralph Waldo Emerson developed a philosophy of flux or transitions in which the active human self plays a central role. At the core of his thought was a hierarchy of value or existence, and an unlimited aspiration for personal and social progress. 'Man is the dwarf of himself', he wrote in his first book Nature (1836). Emerson resented a dire portrait of humankind's condition: 'Men in the world of today are bugs or spawn, and are called "the mass" and "the herd"'. We are governed by moods which 'do not believe in one another', by necessities real or only imagined, but also,Emerson held, by opportunities for 'untaught sallies of the spirit' - those few real moments of life which may nevertheless alter the whole.

Emerson's lectures drew large audiences throughout America and in England, and his works were widely read in his own time. He influenced the German philosophical tradition through Nietzsche - whose The Gay Science carries an epigraph from 'History' - and the Anglo-American tradition viaWilliam James and John Dewey. Emerson's major works are essays, each with its own structure, but his sentences and paragraphs often stand on their own as expressions of his thought.

1 Life

Ralph Waldo (known to friends and family as 'Waldo') was the fourth of eight children born in Boston to the Reverend William Emerson and Ruth Haskins Emerson. He lost his father to tuberculosis just before his seventh birthday and was brought up by his mother and his father's sister Mary Moody Emerson. After four undistinguished years at Harvard, he became a schoolteacher and studied theology there, preparing for the ministry.

In 1829,Emerson was ordained pastor of the Second (Unitarian) Church of Boston and married Ellen Tucker. Ellen died of tuberculosis in 1831, and the following year Emerson resigned his position on the grounds that he could no longer administer the sacrament of the Last Supper, which he considered a 'dead form'. On Christmas Day, 1832, he set sail for Europe, where he toured Malta, Italy, France, Switzerland, England and Scotland. In Britain, he visited the ageing poets William Wordsworth and Samuel Taylor Coleridge, and the historian and philosopher of culture Thomas Carlyle, who became a lifelong friend.

Soon after his return in November 1833, Emerson gave his first lecture, 'The Uses of Natural History', at the Masonic Temple of Boston, embarking on a career as an orator that would continue for the next half-century. He married Lidian Jackson in 1835 and settled in Concord, Massachusetts, from which he set out on lecture tours through the northeastern United States and later to England (1847-8 and 1872-3) and the American midwest.

Emerson's published works, derived from his lectures and journals, include *Nature* (1836), *Essays, First Series* (1841), *Essays, Second Series* (1844), *Representative Men* (1850), *English Traits* (1856), *The Conduct of Life* (1860) and *Society and Solitude* (1875).

2 Emerson as philosopher

There is no one to whom commentators, philosophers or not, are more apt to *deny* the title of philosopher than Emerson. Most twentieth-century discussions of his work have been by literary critics, with such notable exceptions as John Dewey's 1903 address 'Ralph Waldo Emerson - Philosopher of Democracy' and a series of papers by Stanley Cavell.Emerson is no system-builder in the mould of Descartes, Kant or Hegel, and his use of the essay form - inherited from his hero, Montaigne - corresponds to the radical epistemological and metaphysical openness of his thought: towards the end of his great essay 'Experience', he writes, 'I know better than to claim any completeness for my picture. I am a fragment and this is a fragment of me.'

Emerson has a broadly Kantian outlook, according to which the world is in some way our construction. But like other Romantics, he finds that the 'lenses which paint the world' include our passions: 'Life is a train of moods like a string of beads, and as we pass through them they prove to be many-coloured lenses which paint the world their own hue, and each shows only what lies in its focus.' These moods 'do not believe in one another': each comes with 'its own tissue of facts and beliefs'. In one moment, we find ourselves bound by fate, in another real possibilities open up; in one moment we see a picture or read a book with a sense of adventure and understanding,

in another we cannot see what interested us before. ‘Our life’, Emerson states, ‘is March weather, savage and serene in one hour’.

At the centre of the series of moods lies the self - or, it would be better to say, the problem of the self. Emerson, like Kierkegaard and Nietzsche, finds the existence of the self to be a major issue. If the mass of men are ‘bugs or spawn’ and even the great or representative person is ‘partial,’ then the achievement of a fully developed human self is an enormous task. Emerson presents himself as having undertaken this task in ‘Self-Reliance’ (1841): ‘Few and mean as my gifts may be, I actually am, and do not need for my own assurance or the assurance of my fellows any secondary testimony.’

Emerson’s philosophy is a blend of classical and incipiently ‘postmodern’ or ‘pragmatist’ notions. With its references to the ‘Unity’ or ‘Over-Soul, within which every man’s particular being is contained’, or to ‘the Ideal journeying always with us, the heaven without rent or seam’, Emerson’s writing exhibits a strong Neoplatonic streak. Yet in a world of shifting moods and things that ‘slip through our fingers… when we clutch hardest at them’, Emerson finds no foundations, but only ‘a house founded on the sea’. Even our language, far from reflecting permanent forms, is ‘fluxional’, ‘vehicular and transitive… good, as ferries and horses are, for conveyance, not as farms and houses are, for homestead’. These Neoplatonic and pragmatic tendencies come together in Emerson’s statement that ‘the one thing in the world, of value, is the active soul’. Whether in history, philosophy or conversation, Emerson stresses the expansions or transitions of thinking the individual undergoes. Each of his major essays offers a series of such transitions.

3 Early works

Emerson’s philosophy is often taken as starting with his first book *Nature* (1836), where he expresses a sense of ‘decorum and sanctity in the woods’ and of vast ‘prospects’ for a culture of new thought and ‘new men’. The new culture can be achieved, and the beauty of the world restored, ‘by the redemption of the soul’, but this redemption takes place, Emerson emphasizes, according to no formula or model, but through ‘untaught sallies of the spirit, by a continual self-recovery, and by entire humility’.

Emerson’s distinctive philosophical voice emerges in ‘The American Scholar’ (1837) where, in one of the reversals characteristic of his thinking, he writes of the scholar less as a man in a library than as a complete ‘Man Thinking’, whose ‘dictionary’ is a life of free action. Influenced by but not ‘warped out of his own orbit’ by past writing, the scholar is an original source rather than ‘the parrot of other men’s thinking’. Emerson calls us back to ordinary life: to ‘the literature of the poor, the feelings of the child, the philosophy of the street, the meaning of household life’. Although he is often termed a ‘transcendentalist’, Emerson does not wish to transcend the common world. ‘I embrace the common, I explore and sit at the feet of the familiar, the low. Give me insight into today and you may have the antique and future worlds.’

‘An Address Delivered Before the Senior Class in Divinity College, Cambridge’, commonly known as the ‘Divinity School Address’ (1838), contains a fierce attack on institutional religion, but a defence of such ‘holy bards’ as Moses and Jesus. The ‘eastern monarchy of a Christianity’ thatEmerson finds around him treats the revelation as something that happened ‘long ago… as if God were dead’. But, Emerson insists, ‘God is; not was’.

Emerson’s first series of twelve essays contains some of his best-known work, including ‘History’, ‘Self-Reliance’, ‘The Over-Soul’ and ‘Circles’, as well as ‘Friendship’, ‘Spiritual Laws’, ‘Intellect’ and ‘Compensation’. Emerson thinks of history, like scholarship, primarily as a matter of the personal and the present, as ‘the desire to do away this wild, savage and preposterous There or Then, and introduce in its place the Here and the Now’. ‘The Over-Soul’ teaches a religion of the here and now to go along withEmerson’s present-oriented history and scholarship: ‘The simplest person, who in his integrity worships God, becomes God.’ ‘Circles’ expresses Emerson’s vision of flux and incompletion, in which ‘permanence is but a word of degrees’. Especially in morality, ‘there is no virtue which is final; all are initial’. Yet Emerson presents his own set of initial or experimental virtues, including especially ‘abandonment’ and ‘enthusiasm’.

‘Self-Reliance’ offers an indictment of the crowd or public - a ‘mob’ of ‘timorous, desponding whimperers’ - and a radical defence of the individual. ‘Whoso would be a man’, Emerson states, ‘must be a nonconformist’; and the healthy attitude of human nature is displayed in ‘the nonchalance of boys who are sure of a dinner’. Anticipating Nietzsche’s idea of the human creation of higher values, Emerson brazenly asks ‘What have I to do with the

sacredness of traditions if I live wholly from within?'

4 Mature philosophy

Emerson's 'Experience' dominates the second series of essays (1844), building an interpretation of human experience around the writer's grief at the death of his five-year-old son. The essay opens with the depiction of a series of stairs whose top and bottom we cannot see. This, Emerson tells us, is 'where we find ourselves'. Like many Emersonian essays, 'Experience' tells the story of 'the fall of man' and of rebirth or renewal - not through a foreign power but through contemporary 'men and women'. How can we make our way through 'the system of illusions' or the 'train of moods' in which we find ourselves, when there is no final, best view, no 'anchorage'? A newly pragmatic Emerson maintains that we can learn to skate over the surfaces of life; or, in a related metaphor, to find that 'everything good is on the highway'. The second series also includes 'Manners', where Emerson develops a philosophy of social relations that, anticipating Nietzsche, stresses the distance between individuals. In 'Nominalist and Realist' the final essay in the series, he develops a perspectival metaphysics that complements his epistemology of moods.

Emerson's preoccupation with the heroic develops most fully in Representative Men (1850), which includes essays on Plato, Napoleon, Montaigne,Shakespeare, Swedenborg and Goethe. The sceptic in the Montaigne essay takes a position between those of 'the abstractionist and the materialist', each of whom treats the world as more solid than it is. We are in fact 'spinning like bubbles in a river… bottomed and capped and wrapped in delusions'. Montaigne, the wise sceptic, develops a philosophy of 'fluxions and mobility… a ship in these billows we inhabit… tight, and fit to the form of man'.

The greatest of Emerson's late essays, 'Fate' (1860), dwells on the biological, physical and psychological forces controlling our experience. Our individual fortunes are fated, Emerson holds, in that the events that 'seem to meet' us are as much 'exuded' from our character as encountered. Yet he insists that there is also liberty or freedom, and that this liberty rests on our powers of thinking: 'if there be irresistible dictation, this dictation understands itself'. As in all his work, from *Nature* onward, Emerson records both our subjection to necessity and our powers of overcoming it. The ideal life, he suggests, can be achieved through a controlled oscillation or balance between 'Nature' and 'Thought'.

See also: American philosophy in the 18th and 19th centuries §2; Neoplatonism; Pragmatism

RUSSELL B. GOODMAN

List of works

Emerson, R.W. (1971-) *The Collected Works of Ralph Waldo Emerson*, ed. R. Spiller *et al.*, Cambridge, MA: Harvard University Press.(The new standard edition of Emerson's writings.)

Emerson, R.W. (1903-4) *The Complete Works of Ralph Waldo Emerson*, ed. E.W. Emerson, Boston, MA: Houghton Mifflin, 12 vols.(Former standard edition.)

Emerson, R.W. (1836) *Nature*; repr. in *Collected Works*, Cambridge, MA: Harvard University Press, 1971, vol. 1, 1-45.(Emerson's first book, advocating an 'original relation' to the universe.)

Emerson, R.W. (1837) 'The American Scholar'; repr. in *Collected Works*, Cambridge, MA: Harvard University Press, 1971, vol. 1, 52-70.(Calls for active souls rather than slavish scholars or 'bookworms'.)

Emerson, R.W. (1838) 'The Divinity School Address'; repr. in *Collected Works*, Cambridge, MA: Harvard University Press, 1971, vol. 1, 76-93.(Controversial graduation address at Harvard Divinity School, in which Emerson attacks the 'Monster' of institutional Christianity.)

Emerson, R.W. (1841) *Essays, First Series*, in *Collected Works*, Cambridge, MA: Harvard University Press, 1971, vol. 2.(Contains 'Self-Reliance', 'Circles', 'The Over-Soul' and 'Intellect'.)

Emerson, R.W. (1844) *Essays, Second Series*, in *Collected Works*, Cambridge, MA: Harvard University Press, 1971,vol. 3.(Contains 'Experience', 'The Poet' and 'Nominalist and Realist'.)

Emerson, R.W. (1850) *Representative Men*, in *Collected Works*, Cambridge, MA: Harvard University Press,1971, vol. 4.(Essays on Montaigne, Plato, Napoleon and others.)

Emerson, R.W. (1860) *The Conduct of Life*, New York: Harcourt Brace, 1960.(Contains 'Fate', 'Power' and 'Illusions'.)

Emerson, R.W. (1910-14) *The Journals of Ralph Waldo Emerson*, ed. E.W. Emerson and W.E. Forbes, Boston, MA, and New York: Houghton Mifflin, 10 vols.(Records of his reading and thinking from 1819 onwards, and

the source for much in his essays.)
Emerson, R.W. (1960-82) *The Journals and Miscellaneous Notebooks of Ralph Waldo Emerson*, ed. W. Gillman *et al.*, Cambridge, MA: Belknap Press, Harvard University Press, 16 vols.(The standard edition.)
Emerson, R.W. (1961-72) *The Early Lectures of Ralph Waldo Emerson*, ed. S.E. Whicher, R.E. Spiller and W.E. Williams, Cambridge, MA: Harvard University Press, 3 vols. (Includes Emerson's first lectures of 1833; and a course on 'The Present Age' given in 1839-40.)
Emerson, R.W. (1964) *The Letters of Ralph Waldo Emerson*, ed. R.L. Rusk, New York: Columbia University Press, 6 vols.(Correspondence with family and friends, including Mary Moody Emerson,Margaret Fuller, Henry David Thoreau, Henry James Sr, Oliver Wendell Holmes Sr.)
Emerson, R.W. and Carlyle, T. (1964) *The Correspondence of Emerson and Carlyle*, ed. J. Slater, New York: Columbia University Press.(Transatlantic correspondence of two great figures of Romanticism.)

References and further reading

Allen, G.W. (1981) *Waldo Emerson*, New York: Viking Press.(A fine, readable biography of Emerson.)
Cavell, S. (1981) 'Thinking of Emerson' and 'An Emerson Mood', in *The Senses of Walden, An Expanded Edition*, San Francisco, CA: North Point Press.(Sets out Emerson's relation to Heidegger through discussion of his 'epistemology of moods'.)
Cavell, S. (1988) *In Quest of the Ordinary: Lines of Skepticism and Romanticism*, Chicago, IL: University of Chicago Press.(Traces lines of thought from Kant to Coleridge and Emerson, and offers an existentialist reading of 'Self-Reliance'.)
Cavell, S. (1989) 'Finding as Founding: Taking Steps in Emerson's "Experience"', in *This New Yet Unapproachable America*, Albuquerque, NM: Living Batch Press.(Argues thatEmerson offers a proto-Heideggerian criticism of 'thinking as clutching', and a conception of foundation as 'finding' or 'taking the open road'.)
Cavell, S. (1990) 'Introduction' and 'Aversive Thinking' in *Conditions Handsome and Unhandsome: The Constitution of Emersonian Perfectionism*, Chicago, IL: University of Chicago Press, 1990.(Emerson's thought in relation to issues discussed by Kant, Nietzsche, Heidegger and Rawls.)
Ellison, J. (1984) *Emerson's Romantic Style*, Princeton, NJ: Princeton University Press.(A close reading of Emersonian texts stressing the 'defensive and aggressive functions' of his irony, repetitions and contradictions.)
Firkins, O.W. (1915) *Ralph Waldo Emerson*, Boston, MA: Houghton Mifflin.(A superbly written general account ofEmerson's life and writing.)
Goodman, R.B. (1990) *American Philosophy and the Romantic Tradition*, Cambridge: Cambridge University Press.(Discusses Emerson's relation to European Romanticism and American pragmatism.)
Kateb, G. (1995) *Emerson and Self-Reliance*, Thousand Oaks, CA: Sage Publications. (Emerson's social and political philosophy.)
Packer, B.L. (1982) *Emerson's Fall*, New York: Continuum.(The best contemporary literary treatment of Emerson's essays.)
Poirier, R. (1992) *Poetry and Pragmatism*, Cambridge, MA: Harvard University Press.('Emersonian pragmatism' in William James, Robert Frost, Gertrude Stein and Wallace Stevens.)
Richardson, R.D., Jr (1995) *Emerson: The Mind on Fire*, Berkeley, CA: University of California Press.(An excellent, comprehensive biography of Emerson.)
Whicher, S. (1953) *Freedom and Fate: An Inner Life of Ralph Waldo Emerson*, Philadelphia, PA: University of Pennsylvania Press.(A classic discussion ofEmerson's dialectic of freedom and fate.)

Emotion in response to art

The main philosophical questions concerning emotion in response to art are as follows. (1) What kind or type of emotions are had in response to works of art? (2) How can we intelligibly have emotions for fictional persons or situations, given that we do not believe in their existence? (This is known as 'the paradox of fiction'.) (3) Why do abstract works of art, especially musical ones, generate emotions in audiences, and what do audiences then have these emotions towards*? (4) How can we make sense of the interest appreciators have in experiencing empathetically art that is expressive of* negative *emotions? (A particular form of this query is 'the paradox of tragedy'.) (5) Is there a special* aesthetic *emotion, raised only in the context of experience of art? (6) Is there an irresolvable tension between an emotional response to art and the demands of aesthetic appreciation? Answers to these questions depend to some extent on the conception of emotion adopted.*

1 Introduction

Having emotions for art works or responding emotionally to them is a familiar enough occurrence, and hardly seems puzzling when recalled at that level of generality. Why should not works of art, in company with people, animals, natural objects and political events, produce emotions in us? Philosophers have, however, raised questions about emotional responses to art in particular contexts, where, viewed from certain angles, there is indeed something puzzling about them.

One such context is that of response to fictions (whether literary, dramatic or cinematic ones), where emotions appear to be had, not only for the work or representation itself, but for the fictional characters or situations represented therein, even though these are perfectly well understood not to exist. A second such context is that of abstract, or nonrepresentational art (music being the example *par excellence*), where it is unclear both what could elicit such a response and what its object could be. A third context is that in which art works expressive of negative emotion (for example, tragedies, requiems and tales of horror) engender parallel responses in perceivers without thereby being avoided or disapproved of by them. And a fourth context in which emotional response to art has struck philosophers as problematic is where the proper appreciation of art is at issue, when it is asked whether such appreciation is compatible with undergoing emotions of the familiar sort that art seems capable of raising in us.

2 The nature of emotions

In order to assess fruitfully the varieties of emotional response to art it is obviously of use to have some account of exactly what emotions are - in the occurrent, as opposed to the dispositional, sense. Philosophical debate on the nature of emotions, informed to greater or lesser degree by available work in psychology, has in the past thirty years or so revolved around an opposition between feeling-based (or sensation-based) approaches and thought-based (or cognition-based) approaches. The former holds that at the core of an emotion is an internal feeling or set of sensations, while the latter holds that at its core an emotion is a particular kind of thought, judgment or evaluation. While the feeling approach has trouble with the intentionality (or object-directedness) and amenability to reason of many emotions, the thought approach has trouble with the experiential aspect of emotions (that is, with what it is to feel them, as opposed to merely having the beliefs or entertaining the thoughts that may be associated with them), with the evident inertia and passivity of many emotional conditions, and with states of desire, whose connection with many emotions seems more than contingent. Still, while the feeling approach can be faulted for too 'mindless' a picture of emotions, it is right to insist on bodily response and inner affect of some sort as a *sine qua non* of emotion, and while the thought approach can be faulted for too 'mindful' a picture of emotions, it is right to emphasize that many emotions include cognitive elements essentially: for example, thoughts with specific contents, which contents are in many cases socially shaped.

At present there appears to be some consensus that, in perhaps the majority of cases, an emotion is best thought of as a bodily response with a distinctive physiological, phenomenological and expressive profile, one that serves to focus attention in a given direction, and which involves cognition to varying degrees and at various levels. The level of cognitive involvement runs from mere registering of presence, to ways of seeing or regarding that which is registered, to propositional conceptions of the object responded to, to articulate beliefs about or attitudes toward the object of response. Alternatively put, an experienced emotion can be said to have as its core a bodily reaction - comprising physiological sensations, feelings of comfort and discomfort, and orientings of attention - which

reaction is often caused or modified by, and is sometimes necessarily bound up with, cognitions of various sorts and strengths, depending on the type of emotion involved. Note that on such a view of emotion, in which cognitive representations on the order of beliefs (or for that matter, desires) are seen as characteristic of, but not essential to, experienced emotion, the intentionality or directedness of emotions (as opposed, say, to moods) is preserved by the root feature of orientation of attention to or focusing of concern on that which the subject registers as significant.

On the other hand, there is also a growing acknowledgment that the pre-theoretically recognized emotions constitute an irreducibly heterogeneous class, that is, that they do not form a 'natural kind'. It seems reasonable to recognize a spectrum of emotional states experienced by humans, from the startle reaction, involving minimal cognition, at one end, to pride, envy, shame, jealousy, grief, remorse, embarrassment and the like, involving complex and often morally conditioned cognitions, at the other end, with hunger, surprise, lust, fear, anger, joy, sorrow and so on filling the vast middle ground. The emotional responses typical of engagement with art, though, tend to be of a sort that has a moderate or high cognitive involvement - a fact relevant to some recent attempts to dissolve too quickly the paradox of fiction by appeal to what need not be true of all cases of emotion.

Although it is convenient to speak of emotions having elements or components of various sorts (for example, thoughts, sensations, desires, feelings, pleasures, pains, shifts of attention), these should not be thought of as merely bundled together, and the emotion as a mere conglomeration. The truth is rather that an emotion is an *ordered* complex or structure of the elements it is taken to comprehend, with *causal* relations prominent among those in which this order consists. My anger at my daughter for having carelessly misplaced my keys, for example, is a bodily response, rooted in physiology and reflected in countenance, involving a focusing of attention on her, and feelings of agitation and displeasure, which result jointly from my thought of her action and my desire that she should not have so acted, while fuelling, perhaps, my desire that she in some way pay for having so acted.

3 Emotional response to representational art: the paradox of fiction

The much-discussed paradox of fiction can be formulated as a set of three propositions, to each of which we seem to have strong allegiance, but which are jointly inconsistent, and thus impossible to maintain coherently as a set. Solutions to the paradox, then, typically take the form of a rejection of one or more of the propositions, with a reasoned justification for doing so. The propositions are these:

(a) We often have emotions for fictional characters and situations known to be purely fictional;
(b) Emotions for objects logically presuppose beliefs in the existence and features of the objects in question;
(c) We do not harbour beliefs in the existence and features of objects known to be fictional. In the extensive discussion in the literature of this conundrum, almost every possible solution to it has been aired. The following comprise most of the solutions that have found adherents.

(1) The Non-Intentionalist solution: emotional responses to fictions are not, despite appearances, instances of emotions as such, but rather of less complex states, such as moods (cheerfulness, for example) or reflex reactions (shock, for example), which lack the full intentionality and cognitivity of emotions *per se*. As is evident, this solution involves the denial of (a). But the diagnosis it offers seems to apply comfortably to only a small portion of the full range of sustained responses to fictions.

(2) The Suspension-of-Disbelief solution: while caught up in fictions, consumers thereof temporarily allow themselves to believe in the nonexistent characters and situations of the fiction, and thus to have bona fide emotions for them, reverting to standing beliefs in their nonexistence once the fiction no longer actively engages them. Such a solution, turning on a denial of (c), though popular in the nineteenth century, unacceptably depicts consumers of fiction as having both a rather tenuous grip on reality and an amazing ability to manipulate their beliefs at will.

(3) The Surrogate-Object solution: emotional responses to fictions take as their real objects not known-to-be-nonexistent persons and events in fictions, but other existent and believed-to-be-existent objects. This solution, in one way or another, thus calls (a) into question.

On one version of this solution, the object of response is simply the fictional work or artistic representation itself, or parts thereof. On another version, the objects of response are rather the descriptions, images, propositions or thought contents afforded *by* the fiction or representation. And on a third version, different enough from the

preceding two to deserve a separate label - the Shadow-Object proposal - the objects of response are real individuals or phenomena from the subject's life experience, ones resembling the persons or events of the fiction, of which the fiction puts the subject covertly or indirectly in mind.

The Surrogate-Object solution in its first two guises distorts the logic and phenomenology of emotional response to fictions. Whatever the nature or status of our response to fictional characters or situations, it is an emotional response to *them*, not to something else. Our responses, however ultimately analysed, have those characters and situations as their evident objects, and not the vehicles that bring them to us or the thoughts through which they are delineated. Much the same complaint can be brought against the Shadow-Object proposal, though here it is clear that the sort of response to which the proposal draws attention does indeed often accompany and underlie the emotional response to fictional matters *per se*. Still, despising a fictional character, say, is not simply reducible to despising people of that sort generally, or to despising some actual similar individual of one's acquaintance.

(4) The Anti-Judgmentalist solution: emotional responses to objects do not logically require beliefs concerning the existence or features of such objects, but only weaker sorts of cognitions, for example, seeing a certain way, conceiving in a certain manner or regarding as if such and such; thus, there is no good reason not to categorize the emotional responses had towards fictions as standard emotions, since they satisfy the demands of a more relaxed cognitivism about emotions. This approach to the paradox, which directly challenges (b), has a growing number of proponents, and merits extended discussion.

The instances of emotional response that challenge judgmentalism - the view that the cognitive element involved in all emotions is a judgment or belief - are mostly of two types. The first type is where there is insufficient time for cognition as such, so that no real representation of the object responded to is formed, there being only a virtually instantaneous reaction, instinctive or reflexive in nature, unmediated by conscious thought (examples: apprehension at a suddenly looming shape, disgust at an accidentally felt slug). A second type is where, though cognition is involved in generating the response, the representation thus formed is either not propositional in nature, or else does not have the status of a judgment, or both (examples: phobic fear of garter snakes, unfounded resentment of female superiors).

As noted earlier, the emotions involved in responding to fictions (for instance, pity, sorrow, love, admiration, anger, hate, hope) lie in the main in the middle and upper ranges of cognitive complexity for emotions. It thus seems undeniable that, whether or not they involve beliefs, such emotions are centrally mediated by representations of various sorts, such as views, conceptions or evaluations, which serve to *characterize* the object of response.

However, even if emotions at this cognitive level do not necessarily involve beliefs of a *characterizing* sort about their objects, such emotions, it seems, must still involve *existential* beliefs in regard to those objects, or something very close to that; that is to say, attitudes or stances of the order of 'taking to exist' or 'regarding as existent'. Otherwise, the state attributed becomes unintelligible, whether as an emotion or anything else. How can one be said to pity, fear, admire or hate something that one does not, concurrently with one's emotion, at least take or regard as existing, now or at some other time? If indeed that cannot be said, then the problem resurfaces, despite what is right in the critique of judgmentalism: since sane consumers of fiction do not regard fictional characters as existing, even when fully engaged with them appreciatively, they cannot really be in the full-fledged emotional states they are casually said to inhabit. The paradox of fiction is proof against anti-judgmentalist dissolution, even if we grant that emotions can occur without characterizing beliefs.

The sticking-point of the paradox is the dimension of existence and nonexistence, as this connects to the cognitive characterization that emotions of the sort in question minimally require. When we view or conceive an object as having such and such properties, whether or not we strictly believe that it does, we must, on pain of incoherence, be taking it to exist or regarding it as existent. For nothing can coherently be viewed or conceived as having properties without at the same time being treated as existent. A case of genuine emotion of a cognitively mediated sort, unlike a corresponding emotional response to a fictional character, involves at least viewing or conceiving an object as having such and such features, which thus in turn presupposes regarding it as existent or taking it to exist.

But I do not, when reading Dostoyevsky's *The Brothers Karamazov*, take Smerdyakov to exist, and so cannot

strictly be viewing or conceiving him as having properties, such as being base or being a murderer. And though my evaluative Smerdyakov-thoughts, generated as I read, may largely be what causes my hate response, directed ostensibly at him, for that response to strictly have *him* as its object, and so count clearly as an instance of hatred *of* Smerdyakov, requires, once more, that I take him to exist - which I clearly do not. It may, however, be the case both that I imagine or make-believe that someone to that effect exists, and that as a result I imaginarily, or make-believedly, experience for him an emotion of hate.

(5) The Surrogate-Belief solution: certain emotional responses to fictions (for instance, that of pity) require only beliefs that, *in the fiction*, the character exists and is or does such and such, and those beliefs are indeed widely held by rational consumers of fiction. This solution thus rejects (c), though not in the manner of Suspension-of-Disbelief theorists.

However, the beliefs this proposed solution highlights, beliefs about what is fictionally the case, can only ground the truth of one's fictionally or imaginarily pitying a character, not of one's literally doing so. Furthermore, that such beliefs play a role in generating emotional responses to fictions does not touch the heart of the paradox, which is that intelligible emotions for objects of the sort typical of engagement with fiction conceptually require beliefs in the existence of such objects, or at a minimum, an existential stance towards them. Beliefs about how things are fictionally can *cause* emotional reactions of some sort, to be sure, but they cannot logically ground intelligible emotions for entities whose existence is denied. Even where the emotion in question is such as to constitutively require beliefs, they are the wrong sort of beliefs partly to constitute such emotions. The beliefs I have 'about' Anna Karenina, say, cannot coherently make her the proper object of any pitiful reaction I might have. Pity involves concern for the welfare of some creature and distress at its suffering. If one does not believe such welfare or suffering is actual, what can one be concerned *for* or distressed *about*? Pity may likewise involve wishes or desires with respect to the thing pitied, but in the absence of a belief in the thing, or more loosely, an existential commitment to it, there cannot coherently be any such wishes or desires.

(6) The Irrationalist solution: while caught up in fictions, consumers of fiction become irrational, responding emotionally to objects that they know do not exist and thus do not have the features they are represented as having. Irrationalists either implicitly deny (c), proposing that we do in some manner endorse the existence of fictional characters and events - while apparently at the same time disavowing them - which qualifies as irrational, in the sense of being inconsistent, or else they implicitly deny (b), holding that we can have emotions for such as fictional characters and events, towards which we lack the usual beliefs, but qualifying such emotions consequently as irrational, in the sense of being unwarranted.

On the first construal the Irrationalist solution closely approaches that of Suspension-of-Disbelief, with the difference, perhaps, that no attempt is made to mitigate the clash of existential stances involved by suggesting that they are not simultaneously in full force. On the second construal the Irrationalist solution holds appreciators of fiction at fault, not for believing what they already believe the negation of - that fictional characters and events exist - but for emotionally responding to such characters and events in ways contraindicated by their beliefs.

It might seem that the Irrationalist solution, on this second construal, is saved from being a non-starter by the rejection of judgmentalism, since otherwise it could be held to be simply impossible, rather than just possibly irrational, to experience full-fledged emotions in the absence of certain beliefs. But as suggested earlier, if the critique of judgmentalism, applied to emotions of the sort that fiction typically elicits, shows only that characterizing beliefs, as opposed to existential ones, may be absent in such cases, the logical space this construal hopes to occupy may not be available. In any event, in the judgment of most commentators, portraying the normal consumer of fiction as fundamentally enmeshed in irrationality, however this be understood, is too high a price to pay for this to be an acceptable solution to the paradox.

(7) The Make-Believe, or Imaginary, solution: emotional responses to fictions cannot, despite appearances, be instances of the ordinary emotions with whose names we tend to label them, but are instead instances of imaginary or make-believe emotions. For, first, the standard emotions of life arguably have belief or belief-like presuppositions, notably existential ones, that are not fulfilled in normal engagement with fictions; and second, such emotions have motivational or behavioural consequences that are not in evidence in the course of such engagement.

The proposal is that in our interactions with works of fiction we experience make-believe emotions, or make-believedly experience emotions, for fictional characters and situations; it is thus (a) that is rejected on this solution. Make-believedly experiencing fear, say, is sufficiently similar to really experiencing fear, especially internally, that it is easily confused with it, and yet make-believedly experiencing fear can be reconciled (while really experiencing fear cannot) with the absence of existential endorsement and motivational upshot *vis-à-vis* the fictions that are feared. In this way the paradox is finally resolved.

In considering this solution, it is important to distinguish the claim that what we feel for fictional characters is *some* kind of emotion, or constitutes emotional response in the *broad* sense, from the claim, here disputed, that what we feel for fictional characters and describe with some ordinary emotion-word is literally an example of such emotion. We are indeed moved, this solution affirms, but not strictly to the standard emotions whose names come to our lips. Note also that to qualify our emotions for fictions as imaginary is to say that they are ones we imagine ourselves to be having, on the basis of experiences contributory to emotion that we are *actually* having, but does not imply that such emotions are illusory or unreal.

Though the Make-Believe proposal probably provides the best resolution to the paradox of fiction as such, a full account of our emotional responses when engaged with fictions - as opposed to our emotions for fictional characters *per se* - will want to acknowledge what is called to our attention by the Non-Intentionalist and Surrogate-Object proposals as well. And even the Irrationalist proposal, on the first construal, may contain a grain of truth; for perhaps we are, at least at moments of maximum involvement, in the incoherent states of mind it postulates as ours throughout.

4 Emotional response to abstract art: music and feeling

Emotional response to abstract art is puzzling, principally because the strategies that provide obvious explanations of both why we respond emotionally, and what we are responding to, in the case of representational art, seem not to be available. A novel, film or Impressionist landscape gives me the image of a human world, elements of which I can empathize or identify with, react to sympathetically or antipathetically, or even mirror unthinkingly, by a sort of natural contagion. But with a symphony, sonata, minimalist sculpture or Abstract Expressionist painting such explanations appear to have no purchase. Human beings and their predicaments are notably absent, at least as far as representation is concerned. So why or how does perception of such art works raise emotion, and on what is such emotion directed? Concentrating for brevity's sake on the art of music, rough answers are as follows.

In so far as music is capable of eliciting emotions in listeners, this appears to work through two different routes or mechanisms, typically operating in tandem. The first we may label the sensory or cognitively unmediated route, and the second the perceptual-imaginative, or cognitively mediated route. It seems undeniable that music has a certain power to induce sensations, feelings and even moods in virtue of its basic musical properties, virtually without any interpretation or construal on the listener's part. Particular timbres, rhythms, intervals, dynamics and tempi exemplify this power most clearly. Such properties need only be registered, as it were, to have their effect, at least for one acclimatized to a given musical culture. The rise in heartbeat caused by rapid tempo, the discomfort occasioned by dissonant intervals, the kinetic impulses induced by dancing rhythms, the excitement produced by quick alternations of soft and loud, the relaxation engendered by a certain tone colour or manner of articulation, are all familiar phenomena. But if the capacity of music to elicit emotion were exhausted by the direct effects of sensing basic musical features, it would be a poor thing, falling far short of the evocation of emotions proper, or even the semblance of such. The gap is filled by the second, or cognitively mediated, route to such evocation.

In addition to presenting an array of sonic features, simultaneously and successively, much music offers the appearance of human emotion, or of persons outwardly manifesting emotional states; arguably, that is what the expressiveness of music largely consists in (see Expression, artistic). In other words, music is often heard as, or heard as if, or just imagined to be, the expression of emotion by an unspecified individual, whom we may call the music's 'persona'. The degree of resemblance between the shape of music and the behaviour through which emotions are commonly expressed in life will have something, though not everything, to do with our being disposed to hear music in such ways. In any event, once this occurs, the mechanisms mentioned above and familiar from appreciation of representational art - mirroring, identification, empathy, sympathy, antipathy - can come into play, resulting in the arousal in the auditor of those same emotions, or else the feelings characteristic of them, or else those emotions on an imaginary plane. The sensory aspect of music alone indeed seems capable of inducing in

us at least a number of simple states of arousal ingredient in many emotions. But it is the perceptual-imaginative aspect, manifested in our disposition to hear emotion or emotional expression in music, that is surely primarily responsible for the complex, more robustly emotional responses to music, whether mirroring or reactive, that so many listeners report.

It remains to add that these mechanisms do not operate in total isolation from each other. The emotion which I hear a passage as expressing may soften or accentuate the particular psychological effect some basic musical feature produces on me, while the effect induced in me largely unthinkingly by some basic musical feature may influence and constrain the emotion of which I am disposed to hear an image in the music.

But if emotions are often produced in listeners in virtue of listening to emotionally expressive music, towards what are such emotions directed? Music neither supplies any objects, nor appears itself to be an appropriate object, for at least the vast majority of such emotions as are putatively aroused. In addition, music does not seem to provide anything that would justify the beliefs or attitudes towards objects that many objects can be held to require. Among the ways of responding to this difficulty are the following.

It can be held that music produces only moods in listeners, moods which intrinsically lack intentionality (anxiety or elation, for example), or else that it produces ‘objectless’ emotions, ones characteristically taking objects but somehow lacking them when aroused by music, for example, sadness or joy directed on nothing, or on nothing in particular.

Alternatively, it can be held that music produces in listeners just the *feeling* component of an emotion, together with the sense of focus or directedness inherent in the bodily response at the emotion’s core, but not the cognitions which characteristically accompany or even partly constitute the emotion.

Finally, it could be maintained that what music occasions in many listeners are states of *imaginary* emotion. The idea is that listeners readily erect, upon a basis of feelings produced in them by music whose expressiveness they empathetically grasp, imagined emotions of a corresponding sort, and that they do this through imagining, usually tacitly, objects and thoughts suitable to the emotions in question. The object of musical emotion, then, is not missing, but merely indefinitely posited in imagination, or is perhaps logically appropriated, as it were, from the emotion imaginarily ascribed to the music’s persona.

5 Emotional response to negatively emotional art: the paradox of tragedy

The paradox of negative emotion in art - of which the paradox of tragedy is a classical illustration - is this. Art that is negatively emotional, that is, art that represents, expresses or otherwise deals with emotions such as shame, grief, horror, sorrow, anger, remorse, despair and the like, seems to have a propensity to elicit parallel responses in appreciators. But if that is so, one would expect appreciators to avoid, or at any rate, judge as inferior, art of this nature. Yet not only do they not do so, but often they hold such art to be the highest or most rewarding of all.

A number of possible explanations have been given for how it is that persons rationally desire or value the empathic experience of negatively emotional art, given the ostensibly negative character of that experience. Here is a general categorization of such explanations.

(1) Compensatory explanations: negative emotion aroused by negatively emotional art is, as such, unpleasant, but undergoing it offers other rewards that compensate for this. (2) Conversionary explanations: negative emotion, which is initially or ordinarily a disagreeable response, is transformed, in the context of artistic appreciation, into something that is in fact agreeable, or at any rate, capable of being enjoyed. (3) Organicist explanations: negative emotion aroused by negatively emotional art is an essential element in a total experience, an organic whole, that is desired or valued. (4) Revisionary explanations: neither negative emotions, nor the feelings they include, are intrinsically unpleasant or undesirable, and thus there is nothing odd about appreciating art that induces such emotions or feelings. (5) Deflationary explanations: despite appearances, neither negative emotions, nor the feelings they include, are really aroused in us by negatively emotional art.

Compensatory explanations include Aristotle’s doctrine of catharsis, understood as a purging or purification of excess or unruly emotions of pity and fear through engagement with tragic drama, which justifies the raising of such emotions in the course of that engagement (see Katharsis). Another such explanation appeals to the value of knowledge of important truths of human existence that emotional engagement with negative art is said to afford. A

third explanation endorses such engagement, not for the knowledge of life it may afford, but rather for the knowledge of the art work it facilitates, emotional engagement with a work being seen as a necessary cost, in many cases, of fully understanding it. A fourth such explanation invokes the moral exercise that is provided, or the moral deepening that results, as a benefit of engagement with negatively emotional art. And a fifth explanation appeals to purely aesthetic pleasures in the beauty, lifelikeness or virtuosity of the representation or expression itself, positing these as enough to outweigh whatever negative emotion is undergone in their appreciation.

Conversionary explanations include Hume's explanation of the appreciation of tragedy; like that just noted, Hume's explanation highlights the pleasure in artistic representation and expression as such, but premises that this pleasure, being greater than the pain of the negative emotions concomitantly raised, does not simply offset that pain, but rather overwhelms and absorbs it, leaving an experience of uniformly positive character (see Hume, D. §3). A rather different conversionary explanation proposes that since the negative emotions raised by a work of art have no life implications for spectators, calling for no actions and betokening no real harms, such emotions must evidently be so altered by the artistic conditions under which they issue that, though still recognizable as this or that negative emotion (disagreeable affects intact), they are yet capable of being relished or enjoyed for experience's sake.

An example of an organicist explanation would be one invoking a satisfaction in the raising of some negative emotion by a work of art, perhaps because the emotion strikes one as appropriately raised in such circumstances, and oneself as admirably human for being thus susceptible. Such a satisfaction would obviously be inseparable from the negative emotion raised, in the fact of which satisfaction is taken. Another such explanation would appeal to the value of working through negative emotions in connection with a work of art, via immersion in its formal, narrative or dramatic structure, the emotions raised thus being an essential element in the experience valued as a whole.

Revisionary explanations are something like this: the experience of negative emotions is not intrinsically unpleasant; the affects involved (that is, sensations and feelings) are not in themselves disagreeable, and can be unproblematically savoured as such, in appropriate contexts. What is negative about negative emotions is only the *evaluation* of their objects that is central to such emotions. Thus, there is no special difficulty about people seeking these emotions from art.

Deflationary explanations come in at least three varieties. One of these hypothesizes artistic analogues of the life emotions, distinct from them in hedonic tone, conative connectedness and behavioural implication, and proposes that only these are raised in us by engagement with emotional art, and not the life emotions themselves. Another deflationary explanation simply flatly denies that anything like the ordinary emotions are evoked in subjects in the course of engaging with emotional art, and suggests that the subject's response, in so far as it is emotional, is exhausted by properly *appreciative* reactions, such as being moved by a work's beauty of expression. A third deflationary explanation maintains that spectators are always only make-believedly in states of negative emotion in virtue of engaging with a work of art, and that on the assumption that make-believe emotions of the negative sort are not inherently displeasing, there is no special problem about people tolerating or even actively pursuing such experiences.

Lack of space prevents a detailed assessment of these proposals, but there would seem to be more merit in compensatory and organicist explanations, and in the second of the conversionary explanations sketched above, than in revisionary or deflationary ones.

6 Emotion for art and aesthetic appreciation of art

Are there emotions unique to the appreciation of art, or aesthetic emotions *per se*, had when and only when a work is apprehended aesthetically? Past theorists, notably Clive Bell, have posited something of this sort, but such a posit has not lately found favour, nor does it appear to answer to any pressing theoretical problem about art. On the other hand, there may be an interesting category of positive emotions that, if not had uniquely for art, are both distinctive of the appreciation of art and not of the sort that typically figure in the content of art. Candidates would be admiration for a work's skill, fascination with a work's form, delight in a work's beauty, or awe at a work's depth of insight or expression. What might also figure here are experiences, remarked by many, of momentary will-lessness or self-transcendence occasioned by intense absorption in a work of art.

The question may also be raised as to the appropriateness of emotional responses to art of the ordinary sort. One form of this question concerns an apparent tension between the familiar picture of an emotion as a disturbing derangement of the psyche and the image of aesthetic appreciation as a state of calm and unclouded attention to a work of art.

The traditional notion of the aesthetic attitude, whose roots are in Kant, Schopenhauer and the eighteenth-century theorists of taste, depicts a frame of mind characterized by disinterestedness, detachment and disengagement from the practical (see Aesthetic attitude §3). But, charitably construed, such a notion demands only that one's personal situation or condition not be what primarily drives or directs one's response to a work, but instead, the humanly significant material that the work presents. In other words, such a notion need not call for suppression of emotional receptivity generally. So long as one's emotional response is a way of connecting to a work, of tracing its expressive outline or grasping its dramatic import, rather than a means of being distracted from it, or a springboard to wallowing in one's private concerns, then there is no conflict between responding to a work with a range of ordinary emotions on the basis of one's life experience and individual sensibility, and appreciating a work in an aesthetically appropriate manner as the specific embodiment of human content it is. By contrast, disinterestedness or detachment understood not as a principle for maintaining focus on a work rather than one's own circumstances, but as a desired end-state of impassivity or imperturbability, is nothing an account of artistic appreciation need embrace.

See also: Abstract art; Art, value of §6; Emotions, philosophy of; Erotic art; Fictional entities; Humour; Music, aesthetics of §7; Sublime, the; Tragedy

JERROLD LEVINSON

References and further reading

Boruah, B. (1988) *Fiction and Emotion*, Oxford: Oxford University Press.(An assessment of prevailing theories of fiction and emotion.)

Budd, M. (1985) *Music and The Emotions*, London: Routledge.(A masterly and largely critical survey of influential theories of musical expression and evocation.)

Budd, M. (1995) *Values of Art*, Harmondsworth: Penguin.(Advances in its fourth chapter an important positive account of the relationship of music and emotion, and in its third chapter, a nuanced solution to the paradox of tragedy.)

Carroll, N. (1990) *The Philosophy of Horror*, London: Routledge.(A ground-breaking approach to issues of our response to horror fictions, including discussion of both the paradox of fiction and the paradox of tragedy.)

Currie, G. (1990) *The Nature of Fiction*, Cambridge: Cambridge University Press.(Develops a general theory of fictions and our response to them, one influenced by, but not coincident with, that of Walton.)

Davies, S. (1994) *Musical Meaning and Expression*, Ithaca, NY: Cornell University Press.(A painstaking and comprehensive survey of its topics, covering some of the same ground as Budd (1985).)

Deigh, J. (1994) 'Cognitivism in the Theory of Emotions', *Ethics* 104: 824-54.(A valuable survey article consolidating recent critiques of the cognitivist paradigm of emotions.)

Eaton, M.M. (1982) 'A Strange Kind of Sadness', *Journal of Aesthetics and Art Criticism* 41: 51-63.(Addresses the enjoyment of negative emotions from art, emphasizing the element of control of such emotions by the art work as key to such enjoyment.)

Feagin, S. (1983) 'The Pleasures of Tragedy', *American Philosophical Quarterly* 20: 95-104.(Proposes an approach to the paradox of tragedy in terms of pleasurable meta-responses had to intrinsically unpleasant first-order responses.)

Hjort, M and Laver, S. (eds) (1997) *Emotion and the Arts*, Oxford: Oxford University Press.(A collection of new essays on the present topic.)

Kivy, P. (1989) *Sound Sentiment*, Philadelphia, PA: Temple University Press.(Containing 'The Corded Shell' and supplementary essays, this is a well-known brief for cognitivism about musical expression and musical appreciation, denying that normal listeners respond to expressive music by feeling ordinary emotions.)

Kivy, P. (1990) *Music Alone*, Ithaca, NY: Cornell University Press.(Further development of the theory in Kivy (1989), with an emphasis on appreciative emotions, for instance, being moved by a work's beauty.)

Lamarque, P. (1981) 'How Can We Fear and Pity Fictions?', *British Journal of Aesthetics* 21: 291-304.(Proposes a form of the Surrogate-Object solution to the paradox of fiction, to the effect that thought-complexes are the objects of our emotional responses to fictional characters.)

Levinson, J. (1990) 'Music and Negative Emotion', in *Music, Art, and Metaphysics*, Ithaca, NY: Cornell

University Press.(Proposes a pluralistic solution to the problem posed by appreciation of negatively emotional music, and explores the nature of a listener's imaginative engagement with expressive music).

Matravers, D. (1997) *Art and Emotion*, Oxford: Oxford University Press.(A sophisticated defence of the arousal theory of emotional content in art, with reference to both fiction and music.)

Moran, R. (1994) 'The Expression of Feeling in Imagination', *Philosophical Review* 103: 75-106.(Emphasizes the importance of emotional modes of imagining, as opposed to the imagining of emotions, in understanding our emotions for art.)

Morreall, J. (1985) 'Enjoying Negative Emotions in Fiction', *Philosophy and Literature* 9: 95-103.(Addresses the enjoyment of negative emotions from fictions, suggesting that the subject's control of such experience is crucial to its enjoyability.)

Morreall, J. (1993) 'Fear Without Belief', *Journal of Philosophy* 90: 359-66.(Critiques judgmentalism about emotions and suggests that this dissolves the paradox of fiction, obviating solutions such as that of Walton, which posit make-believe emotions centrally.)

Neill, A. (1991) 'Fear, Fiction, and Make-Believe', *Journal of Aesthetics and Art Criticism* 49: 47-56.(Discusses and criticizes Walton's approach to the problem of fearing fictions.)

Neill, A. (1993) 'Fiction and the Emotions', *American Philosophical Quarterly* 30: 1-13.(Develops a Surrogate-Belief solution to the paradox of fiction.)

Neill, A. (1995) 'Fear and Belief', *Philosophy and Literature* 19: 94-101.(Responds to Morreall (1993).)

Oatley, K. (1994) 'A Taxonomy of the Emotions of Literary Response and a Theory of Identification in Fictional Narrative', *Poetics* 23: 53-74.(Title is self-explanatory.)

Packer, M. (1989) 'Dissolving the Paradox of Tragedy', *Journal of Aesthetics and Art Criticism* 47: 212-19.(Critically reviews a number of proposals on the paradox and defends a form of compensatory theory in which the satisfaction of apprehending necessity and universality in human affairs is brought to the fore.)

Radford, C. (1975) 'How Can We Be Moved by the Fate of Anna Karenina?', *Proceedings of the Aristotelian Society* Supplement 49: 67-93.(The initiating paper in the modern series of responses to the paradox of fiction, which claims that in such responses we inescapably reveal ourselves as irrational.)

Radford, C. (1989) 'Emotions and Music: A Reply to the Cognitivists', *Journal of Aesthetics and Art Criticism* 47: 69-76.(A reply to Kivy (1989), arguing for the power of emotional music to induce at least simple emotions in listeners.)

Radford, C. (1995) 'Fiction, Pity, Fear, and Jealousy', *Journal of Aesthetics and Art Criticism* 53: 71-5.(Responds to Neill (1993).)

Ridley, A. (1995) *Music, Value, and the Passions*, Ithaca, NY: Cornell University Press. (Defends a view of emotion in music that mediates between cognitivist and arousalist positions.)

(1992) 'Desire in the Experience of Fiction', *Philosophy and Literature* 16: 279-91. (Argues that we can and do have desires regarding fictional characters, thus making it more plausible that we straightforwardly have emotions for them.)

Robinson, J. (1994) 'The Expression and Arousal of Emotion in Music', *Journal of Aesthetics and Art Criticism* 52: 13-22.(Emphasizes the importance of feelings of a small-scale sort in the analysis of music's emotionality.)

(1995) 'Startle', *Journal of Philosophy* 92: 53-74.(Develops a bodily-response based conception of emotions, arguing convincingly that recent philosophy has overemphasized the cognitive dimension of emotions.)

Schaper, E. (1978) 'Fiction and the Suspension of Disbelief', *British Journal of Aesthetics* 18: 31-44.(Argues that suspension of disbelief cannot coherently explain our reactions to fictional characters, and endorses a variety of Surrogate-Belief solution to the paradox of fiction.)

Sparshott, F. (1994) 'Music and Feeling', *Journal of Aesthetics and Art Criticism* 52: 23-35.(A review of the topic particularly sensitive to the variety of items that can be ranged under the term 'feeling'.)

Walton, K. (1990) *Mimesis as Make-Believe*, Cambridge, MA: Harvard University Press.(A defence of the role of making-believe and make-believe emotions in the understanding of fictions; widely influential, both in this form and in that of the articles that preceded it.)

Yanal, R. (1994) 'The Paradox of Emotion and Fiction', *Pacific Philosophical Quarterly* 75: 54-75.(Offers a categorization of different positions regarding the paradox of fiction, and endorses mainly a Surrogate-Belief solution to it.)

Zemach, E. (1996) 'Emotion and Fictional Beings', *Journal of Aesthetics and Art Criticism* 54: 41-8.(In response to the paradox of fiction, proposes that we have appropriate and rational beliefs toward nonexistents such as fictional characters.)

Emotive meaning

Emotive meaning contrasts with descriptive meaning. Terms have descriptive meaning if they do the job of stating facts: they have emotive meaning if they do the job of expressing the speaker's emotions or attitudes, or exciting emotions or attitudes in others. Emotivism, the theory that moral *terms have only or primarily emotive meaning, is an important position in twentieth-century ethics. The most important problem for the idea of emotive meaning is that emotive meaning may not really be a kind of meaning: the jobs of moral terms supposed to constitute emotive meaning may really be performed by speakers using moral terms, on only some of the occasions on which they use them.*

There are two components in emotivist accounts of the function and meaning of moral terms. One is a matter of relations to the speaker: moral assertions serve to express the speaker's emotions or attitudes. The other is a matter of relations to the audience: moral assertions serve to commend things, or to arouse emotions or attitudes in the audience.

The most celebrated accounts of emotive meaning were developed by A.J. Ayer and C.L. Stevenson. Ayer (1936) argued on general metaphysical and epistemological grounds that moral terms can only express and excite emotions. Stevenson (1937) developed a more detailed theory, relying more on distinctively ethical considerations (see Emotivism). These theories are liable to make moral discussion seem irrational, and to make no distinction between moral argument and propaganda. R.M. Hare developed a theory designed to remedy these defects. He argued that sentences using paradigm moral terms like 'good', 'right' and 'ought' are really disguised imperatives. Since there is a logic of imperatives, there is room for rational moral argumentation, and moral argument can be distinguished from propaganda, even though moral assertions do not primarily state facts (see Prescriptivism).

The most important difficulties for the idea of emotive meaning can be raised by asking whether the emotive meaning of moral terms is a matter of the speech act performed by someone using these terms, and, if so, what kind of speech act? One can distinguish between locutionary, illocutionary, and perlocutionary acts (or force). Locutionary acts are simply a matter of uttering certain words with certain senses and referents. Illocutionary acts are done *in* saying things; what illocutionary act one performs in uttering a sentence is determined together by the senses and referents of the words in the sentence and the context. Perlocutionary acts are done *by* saying things. That one performs a certain perlocutionary act is not guaranteed by performing an appropriate illocutionary act: it depends on further variable features of the context. Suppose Bob utters the words 'Down with the aristocrats!' before a large crowd in a revolutionary situation. Bob performs a locutionary act just by saying words with that meaning. In this context, Bob also performs the illocutionary act of *inciting* revolution. Bob's words may also, in that context, have the perlocutionary effect of *provoking* revolution; but this is not guaranteed by the locutionary or the illocutionary force of Bob's utterance. If Bob is sufficiently unpopular, his advocating revolution may actually dampen revolutionary enthusiasm. This classification suggests that the order of explanation typically goes from the locutionary to the illocutionary and the perlocutionary: it is in virtue of the sense and reference of the words one utters and the context that one performs a certain illocutionary act, and that one performs certain perlocutionary acts. The presumption is that an account of meaning will begin with sense and reference; if it begins with illocutionary or perlocutionary acts, it may begin in the wrong place (see Speech acts).

Some early accounts of emotive meaning seem to identify the meaning of moral assertions with perlocutionary acts (for example, arousing emotions). But this seems both to be the wrong place to begin and to raise a further concern. If we think of meaning as a matter of convention, not (mere) causal variation, then causal correlations between utterances and the production of certain effects are not really meanings. Perhaps then emotive meaning is a matter of illocutionary force. Unfortunately, seeing emotive meaning as illocutionary force is also problematic; in addition to the general problem that an account of meaning apparently should not begin with the illocutionary, there is the specific problem that moral terms do not seem *always* to have the right kinds of illocutionary force. Consider the suggestion that 'good' is used to commend. While it may be true that the term 'good' as used in 'This is a good tennis racket', uttered in a sports shop has the illocutionary force of commending a tennis racket, 'good' as used in 'If you can't get a good one there, try the shop down the street', does not obviously seem to commend anything.

Moral arguments raise especially acutely a version of the same problem for the view that emotive meaning is illocutionary force. Consider the argument: 'Telling the truth is good; if telling the truth is good, getting your little brother to tell the truth is good; so getting your little brother to tell the truth is good.' This argument looks valid. But while 'telling the truth is good' in its occurrence as the first premise commends telling the truth, in its occurrence in the second premise it does not seem to commend anything. So, the suggestion is, emotivists cannot account for the validity of some moral arguments, because if meaning is understood as illocutionary force, it is not the same between different occurences of the same words in arguments. This problem is often called 'the problem of unasserted contexts' (in the second premise, 'telling the truth is good' occurs unasserted).

Fans of emotive meaning can respond to these criticisms. Hare (1952), Blackburn (1984) and Gibbard (1990) have all offered solutions to the problem of unasserted contexts. It is certainly true that the meanings of *some* words (like 'promise') seem well explained by explaining the illocutionary act one performs in using them. Moreover, it is not an accident that, for instance, 'good' is often used to commend, while it is an accident if 'fast' is used to commend. Still, an attractive alternative to the theory that 'good' has emotive meaning is that 'good' *means* something like 'meets the relevant standards'. It is by virtue of this meaning that 'good' is often used to commend. Emotive force is then explained by meaning, not vice versa.

See also: Emotivism; Prescriptivism

DAVID PHILLIPS

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, New York: Dover.(Chapter 6 is Ayer's classic argument for and presentation of emotivism.)

Alston, W. (1967) 'Emotive Meaning', in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York: Macmillan and the Free Press, vol. 2, 486-93.(A trenchant critique of the idea of emotive meaning.)

Blackburn, S. (1984) *Spreading The Word*, Oxford: Oxford University Press.(Chapters 5 and 6 defend a theory giving moral words emotive meaning.)

Gibbard, A. (1990) *Wise Choices, Apt Feelings*, Cambridge, MA: Harvard University Press.(A sophisticated and wide ranging recent defence of, *inter alia*, a kind of emotive meaning.)

Hare, R.M. (1952) *The Language of Morals*, Oxford: Oxford University Press.(Classic defence of the view that moral terms are disguised imperatives.)

Stevenson, C.L. (1937) 'The Emotive Meaning of Ethical Terms', *Mind* 46: 14-31.(Stevenson's initial presentation of emotivism.)

Stevenson, C.L. (1944) *Ethics and Language*, New Haven, CT: Yale University Press.(Book-length development of Stevenson's view.)

Urmson, J.O. (1968) *The Emotive Theory of Ethics*, London: Hutchinson & Co.(Balanced consideration and development of emotive theory.)

Ziff, P. (1960) *Semantic Analysis*, Ithaca, NY: Cornell University Press.(Chapter 6 is an early and trenchant development of the kind of critique of emotive meaning given above.)

Emotivism

Emotivists held that moral judgments express and arouse emotions, not beliefs. Saying that an act is right or wrong was thus supposed to be rather like saying 'Boo!' or 'Hooray!' Emotivism explained well the apparent necessary connection between moral judgment and motivation. If people judge it wrong to lie, and their judgment expresses their hostility, then it comes as no surprise that we can infer that they are disinclined to lie. Emotivism did a bad job of explaining the important role of rational argument in moral practice, however. Indeed, since it entailed that moral judgments elude assessment in terms of truth and falsehood, it suggested that rational argument about morals might be at best inappropriate, and at worst impossible.

In the early part of the twentieth century, under the influence of logical positivism, a new view about the nature of morality emerged: emotivism (see Logical positivism). Emotivists held that when people say, 'It is wrong to tell lies', they express their hostility towards lying and try to get others to share that hostility with them. Moral claims were thus supposed to be very different from claims expressing beliefs. Beliefs purport to represent the world, and so are assessable in terms of truth and falsehood. Emotions, by contrast, do not purport to represent the world, so moral claims were supposed to elude such assessment (see Analytic ethics §1; Moral judgment §1). Judging acts right and wrong was thus rather like saying 'Boo!' and 'Hooray!'

Emotivism had evident appeal. It is widely agreed that there is a necessary connection of sorts between moral judgment and motivation. If someone judges telling lies to be wrong then they are motivated, to some extent, not to lie. But what people are motivated to do depends on what they approve of, or are hostile towards, not simply on what they believe (see Moral motivation). Imagine, then, that someone's judgment that telling lies is wrong expressed a belief. In order to know whether they are inclined to lie or not we would then need to know, in addition, whether they approve of, or are hostile towards, telling lies. But we need to know no such thing. Knowing that they judge lying wrong suffices to know that they are disinclined to lie. This fits well with the idea that the judgment itself simply expresses hostility.

Emotivism also had its difficulties, however. Though emotivists admitted that rational argument about morals had an important role to play, their view entailed that this role was strictly limited. Since they agreed that less fundamental moral claims are entailed by more fundamental claims along with factual premises, and since they agreed that factual premises could be criticized rationally, they held that less fundamental moral claims must be rationally-based. Someone who judges lying wrong because they think that lies are harmful must, they thought, change their mind on pain of irrationality if shown that lying is harmless. But at the same time they insisted that fundamental moral claims - those that are not so derived like, perhaps, the claim that it is wrong to cause harm - are immune from such rational criticism. This was the so-called 'fact/value gap' (see Fact/value distinction; Logic of ethical discourse).

It is unclear whether emotivists were consistent in allowing even this limited role for rational argument, however.

1 If it is wrong to cause harm and lying causes harm then it is wrong to tell lies
2 It is wrong to cause harm
3 Lying causes harm
Therefore, it is wrong to tell lies

This argument is valid only if 'It is wrong to cause harm' in premises (1) and (2) means the same thing. If this phrase means different things then there is an equivocation and the argument is straightforwardly invalid. Emotivism entails that someone who asserts (2) expresses hostility towards causing harm. Yet whatever 'It is wrong to cause harm' means in (1), it most certainly does not serve to express such hostility. In (1) the phrase appears in the antecedent of a conditional. Someone who asserts (1) may thus even deny that it is wrong to cause harm. They need therefore have no hostility to express towards causing harm.

Philosophers sympathetic to emotivism have tried to rescue it from this objection. There is a real question whether emotivists themselves should ever have been interested in preserving an important role for rational argument about morals, however. If the function of moral judgment is simply to express emotions and arouse like emotions in others then it follows that rational argument is at best one way, and perhaps not a very good way, of achieving

these aims. We might be more effective if we distracted people from the facts and used rhetoric, humiliation and brainwashing instead. It is hard to see how emotivists could find fault with the idea that a practice in which the use of such technologies was widespread could still constitute a perfectly proper *moral* practice.

The best emotivists could say at this point was, ‘Boo for persuasion and brainwashing!’ Philosophers who thought this response failed to acknowledge the central and defining role played by rational argument in moral practice concluded that emotivism extracted too high a price for its explanation of the necessary connection between moral judgment and motivation. Subsequent theorists have focused on whether an alternative explanation of the necessary connection is available, one which also accommodates the idea that rational argument plays such a central and defining role. No consensus on this issue has emerged, however.

If nothing else, emotivism succeeded in making clear how difficult it is to explain the necessary connection between moral judgment and motivation, together with the idea that rational argument plays a central and defining role in moral practice, if the emotions that cause our actions are assumed to be beyond rational criticism. Much recent work about the nature of morality proceeds by calling this assumption into question.

See also: Ayer, A.J.; Expression, artistic; Moral knowledge; Moral realism; Morality and emotions §§1-2; Prescriptivism; Stevenson, C.L.

MICHAEL SMITH

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz; 2nd edn, 1946, ch. 6.(Contains a classic statement of emotivism by a logical positivist.)

Blackburn, S. (1984) *Spreading the Word*, Oxford: Oxford University Press, ch. 6.(Shows how modern versions of emotivism attempt to avoid the problems faced by their ancestor.)

Smith, M. (1994) *The Moral Problem*, Oxford: Blackwell.(Argues that, contrary to the standard assumption, emotions can be rationally criticized. Ch. 2 contains a critical discussion of Ayer’s emotivism and more modern versions.)

Stevenson, C.L. (1944) *Ethics and Language*, New Haven, CT: Yale University Press.(Another classic statement of emotivism and explanation of the difference between disagreements about values and disagreements about facts.)

Warnock, G. (1967) *Contemporary Moral Philosophy*, London: Macmillan, ch. 3.(Contains a critical discussion of emotivism.)

Empedocles (*c.*495-*c.*435 BC)

Empedocles, born in the Sicilian city of Acragas (modern Agrigento), was a major Greek philosopher of the Presocratic period. Numerous fragments survive from his two major works, poems in epic verse known later in antiquity as On Nature and Purifications.

On Nature sets out a vision of reality as a theatre of ceaseless change, whose invariable pattern consists in the repetition of the two processes of harmonization into unity followed by dissolution into plurality. The force unifying the four elements from which all else is created - earth, air, fire and water - is called Love, and Strife is the force dissolving them once again into plurality. The cycle is most apparent in the rhythms of plant and animal life, but Empedocles' main objective is to tell the history of the universe itself as an exemplification of the pattern.

The basic structure of the world is the outcome of disruption of a total blending of the elements into main masses which eventually develop into the earth, the sea, the air and the fiery heaven. Life, however, emerged not from separation but by mixture of elements, and Empedocles elaborates an account of the evolution of living forms of increasing complexity and capacity for survival, culminating in the creation of species as they are at present. There followed a detailed treatment of a whole range of biological phenomena, from reproduction to the comparative morphology of the parts of animals and the physiology of sense perception and thinking.

The idea of a cycle involving the fracture and restoration of harmony bears a clear relation to the Pythagorean belief in the cycle of reincarnations which the guilty soul must undergo before it can recover heavenly bliss. Empedocles avows his allegiance to this belief, and identifies the primal sin requiring the punishment of reincarnation as an act of bloodshed committed through 'trust in raving strife'. Purifications accordingly attacked the practice of animal sacrifice, and proclaimed prohibition against killing animals to be a law of nature.

Empedocles' four elements survived as the basis of physics for 2,000 years. Aristotle was fascinated by On Nature; his biology probably owes a good deal to its comparative morphology. Empedocles' cosmic cycle attracted the interest of the early Stoics. Lucretius found in him the model of a philosophical poet. Philosophical attacks on animal sacrifice made later in antiquity appealed to him as an authority.

1 Life and work

The first lines of Empedocles' poem *Purifications* give a flavour of the man:

> Friends, who live in the great city of the yellow Acragas, up on the heights of the citadel, caring for good deeds, I give you greetings. An immortal god, mortal no more, I go about honoured by all, as is fitting, crowned with ribbons and fresh garlands.
>
> (fr. 115)

Men and women followed him in their thousands, Empedocles says, wanting prophecies or remedies for diseases. Not surprisingly there accumulated in antiquity a huge conglomeration of fact and fantasy about the life and death of such a figure, summarized by Diogenes Laertius (VIII 51-75). A cautious sifting yields the following picture.

Empedocles was born of aristocratic family, a little after Anaxagoras. He died aged sixty. He was active in the political life of Acragas as a fierce opponent of oligarchy and tyranny. He had a reputation as an orator: Aristotle even makes him the inventor of rhetoric. He is described as a physician, but despite his profound interest in human physiology, anecdotes of his miracle working and the supernatural powers he claimed his teaching would impart (fr. 111) both strongly suggest a practitioner of magic, an activity doubtless to be seen in the context of his Pythagorean religious beliefs.

Among various writings ascribed to him, the two most important were *On Nature* and *Purifications*, hexameter poems probably in three and two books respectively. *On Nature* was at the time of its composition very likely the longest work of philosophy ever written and Empedocles' fragments constitute the largest corpus of original extracts to survive of any Presocratic. Scholars disagree about which of the fragments belong to which of the two major works - the sources seldom supply specific information on this point. In the age of Victorian rationalism it was supposed that *On Nature* presented a sober materialist philosophy of nature, subsequently abandoned in

Purifications for the intoxications of mystery religion. Fragments were assigned to the two poems accordingly. The basis for this division of the material has long since collapsed, and more recent study (notably by Kahn 1960) has suggested that *On Nature* itself draws religious morals from a philosophy which was always conceived in religious terms. Indeed, it has been argued that the great majority of the fragments, including those on religious themes, belong to *On Nature*, *Purifications* being simply a collection of oracles and ritual prescriptions designed to satisfy the desire for healing and salvation Empedocles mentions in its opening verses.

2 Pythagoreanism

Empedocles praises Pythagoras as a man of surpassing knowledge and 'wise works', with powers of foresight extending to ten or twenty generations ahead (fr. 129). What prompted this extravagant admiration was evidently Pythagoras' analysis of the human condition: to atone for sin the soul is subject to a cycle of reincarnations into a variety of living forms (since all life is akin) until release is eventually achieved (see Pythagoras §2). On this view animal sacrifice counts as unwitting slaughter of one's kin. Empedocles dramatizes the implication in some gothic hexameters:

> Father lifts up a beloved son changed in form, and butchers him with a prayer, helpless fool…. Similarly son seizes father and children their mother, and tearing out the life they consume the flesh of those they love.
>
> (fr. 137)

Further verses spell out the penalty bloodshed incurs, invoking an 'oracle of Necessity' which condemns guilty spirits (*daimones*) to wander apart from the blessed for 30,000 years, in all manner of mortal forms. They conclude with a dramatic confession: 'Of these I too am now one, an exile from the gods and a wanderer, having put my trust in raving Strife' (fr. 115).

Fragment 115 is said by Plutarch to have formed part of the preface to Empedocles' philosophy. Is this a reference to *Purifications* (the usual supposition)? 'Philosophy' rather suggests *On Nature*; and Empedocles announces the *recovery* of his divinity at the start of *Purifications* (see §1). On the basis of comparison with the proem to Lucretius' *On the Nature of Things* Sedley (1989) argues persuasively that the entire sequence of fragments just summarized helped to launch *On Nature*. Whether this is right or not, Empedocles' Pythagoreanism is the best clue we have to his intentions in *On Nature*. The poem should be seen as an attempt to exhibit the cycle of incarnation as an instance of a general pattern of repetition governing all change: plurality is converted into unity by the power of Love and unity is then broken into plurality by Strife, until the process is reversed and conversion of plurality into unity begins once again. What *On Nature* works out in detail is the realization of this pattern in the rhythms of plant and animal life and the design and dissolution of the body, but above all in the history of the universe itself (see Pythagoreanism §3).

3 The cycle of change

The cycle of change is announced in fragment 17.1-2: 'A twofold tale I shall tell: at one time they grew to be one alone out of many, at another again they grew apart to be many out of one'. These lines deliberately echo and defiantly contradict the assertion of Parmenides (§3) that there is only *one* tale to tell, that of changeless and timeless being and unity (fr. 8.1-6). The language of growth is not accidental. Empedocles has already indicated that the subjects of the dual process are what he calls the 'roots': earth, air, fire and water. Subsequent philosophy would speak of these as 'elements', but Empedocles chooses a designation which captures the idea that they are not merely the basis of everything else, but themselves have a potentiality for development. He is emphatic in his agreement with the Parmenidean thesis that nothing comes into being from non-existence or perishes into it: mixture and separation of roots are what humans mistakenly call birth and death (frs 8-12).

After amplification of the twofold tale Empedocles continues: 'And these things never cease their continual interchange, now through love all coming together into one, now again each carried apart by strife's hatred' (fr. 17.6-8). Mixture and separation would not occur without the agency of forces bringing them about. Are Love and Strife physical or psychological forces? For us it is counterintuitive to envisage earth or fire as capable of psychological responses. But *On Nature* is permeated with expressions of the hatred the roots have for each other, of their desire for one another. Empedocles does not write as though he wants the reader to construe them metaphorically. On the other hand, the operation of Love in creating mixtures of the roots is often also described in

the language of craftsmanship: she welds (frs 34, 96) and rivets (fr. 87) and fires like a potter (fr. 73). Here Love seems to represent whatever physical force makes for the assimilation of things. However we are to conceive of Love and Strife, they are certainly treated as existing independently of the roots. But there is a difference between them: Love is ‘among’ the roots, Strife ’apart from’ them (fr. 17.19-20). The implication is presumably that Love works with the grain, Strife against it: roots have a natural tendency to join together, whether earth with earth or in mixture with air, fire and water, whereas their separation is unnatural (fr. 22). Aristotle, however, found Empedocles thoroughly confusing on this issue (*Generation and Corruption* II 6).

Fragment 17.9-13 sums up the two key features of change - its oscillating duality and its ceaseless repetition - in a surprisingly Heraclitean conclusion (compare Heraclitus §3):

> So insofar as they have learned to grow one from many, and again they grow many as the one grows apart, thus far do they come into being and have no stable life; but insofar as they never cease their continual interchange, thus far they exist always, changeless in the cycle.

The implication is that Parmenides looked for changeless existence in the wrong place: it is not to be found in being (not even the being of the roots), but in the regularity of the cycle of unending flux.

4 The biological paradigm

The initial presentation of the cycle of change in fragment 17 is entirely general and abstract in its formulations. The evidence for the structure and content of the following section of *On Nature* is inadequate, although newly discovered fragments may clarify the matter (see Martin and Primaveri 1997). However Empedocles seems to have taken the creation of plants and animals as a paradigm of the way mixture of roots generates a huge variety of other forms (fr. 21). He appealed to the analogy of a painter, using a few pigments with many potentialities:

> When they seize pigments of many colours in their hands, mixing in harmony more of some and less of others, they produce from them forms resembling all things, creating trees and men and women, beasts and birds and water-bred fish, and long-lived gods, too, highest in honour.
>
> (fr. 23)

If this is how art achieves its effects, why should we look for any other explanation of the way nature produces the originals copied by the painter?

It may be that the paradigm of animal creation, once established, was subsequently invoked in later sections of the poem. For example, the following lines from fragment 20 might well have been written to support the notion expressed in fragment 31 of the disintegration of the limbs of the cosmic sphere:

> This is well-known in the mass of mortal limbs: at one time, in the maturity of a vigorous life, all the limbs that are the body’s portion come into one through love; at another time again, torn asunder by evil strifes, they wander, each apart, on the shore of life. So it is too for shrubs and water-housed fish and mountain-laired beasts and wing-progressing gulls.

5 Cosmic history

Mention of the cosmic sphere brings us to Empedocles’ boldest application of the concept of a cycle of change: to the history of the universe. He conceived that just as harmonization by Love and decomposition by Strife constitute the common pattern for the biological development of plants and animals, there must be a similar story to tell about the universe as a whole. But here he supposed that the unity achieved at one pole of the cycle and the division at the other took more radical and absolute forms than in the biological realm. When - as he described first as usual - the influence of Love is at its strongest, all distinctions disappear as reality becomes a perfect divine sphere (frs 27-9). Strife for its part generates a vortex which not only breaks the sphere down into its constituent roots, but achieves their complete separation. It is under Strife’s domination that the world as we are familiar with it comes into being (A37, 42).

Empedocles’ account of this development was evidently extensive, and included full discussions of all the topics by now traditional in philosophical cosmogony: notably, the formation of the earth, the sea, and the heavenly bodies, and the evolution of life. Strife divides the roots into four great isolated masses. What happens next is

obscure, but the key process was physical interaction between these masses: some of its effects inevitable, some pure chance (see, for example, fr. 53). In particular, misty air heated by fire rises up and forms a nocturnal hemisphere balanced by a diurnal counterpart of fire. The sun is a reflection of this fire, the moon compacted air. The earth too sweats under the heat of the sun, which is the origin of the sea (A30, 49, 66).

The power of Strife is clearly not what it was, but we hear nothing yet of Love. Its influence begins to be apparent with the emergence of life. A clear example is the creation of bone, ascribed to *harmonia*, one of Empedocles' synonyms for Love: 'And kindly earth received in its broad melting-pots two parts of the gleam of Nestis [water] out of eight, and four of Hephaestus [fire], and they became white bones, marvellously joined by the gluing of harmonia' (fr. 23). The creatures to which bodily parts such as this belonged were described subsequently, in Empedocles' memorable theory of evolution:

> Empedocles held that the first generations of animals and plants were not complete, but consisted of separate limbs not joined together; the second, arising from the joining of these limbs, were like creatures in dreams; the third was the generation of whole-natured forms. The fourth arose no longer from the homogeneous substances such as earth or water, but by intermingling, in some cases as the result of the compacting of their food, in others because female beauty excited the sexual urge. And the various species of animal were distinguished by the quality of the mixture in them.
>
> (A72)

Surviving fragments give vivid details of the bizarre beings of the first three stages (frs 57-62). From these it is clear that in the first two especially chance was made responsible for a great deal. Empedocles may have talked in this context of the survival of the fittest: a remarkable anticipation of Darwinism, although criticized by Aristotle for its inability to account for the regular teleological patterns of nature (*Physics* II 8). It is only with the animals of the fourth (that is, present) stage that Love begins to exercise a control over the whole structure and pattern of life (fr. 71). The evidence suggests that at this point in the poem Empedocles included a full-scale comparative biology of plant and animal species, focused on explanation of the formation and function of the parts of the body.

Ultimately all this diversity would be reabsorbed into the divine sphere. *On Nature* may have concluded with some verses which foreshadow this, speaking of god as a mind, 'holy and beyond description, darting through the whole universe with swift thoughts' (fr. 134).

Aristotle sometimes talks as if he thinks two cosmogonies were envisaged: one occurring as Strife grew more powerful, another as its influence ebbed. He then complains that Empedocles said and could say nothing about the second (A42). Some modern commentators have reinforced this idea of a double cosmogony, adding a double zoogony and appealing to some obscure lines which speak of a 'double birth and double failing of mortal things' (fr. 17.3-5). But it seems likely that this is just a reference to the growth of unity, then of plurality, both 'mortal' conditions persisting only for a while. In any event, the reduplicating interpretation is hard to reconcile with the shape of the cosmic history as it emerges from the rest of the fragments.

Another way of reading Empedocles' history of the universe is as an arbitration and synthesis between two approaches to cosmogony adopted by different among his philosophical predecessors. Broadly speaking, the Ionians from Anaximander on explain the emergence of a world or worlds as the outcome of separation from an original undifferentiated condition, and of the consequent interaction between the physical forces released by the separation. By contrast, in the cosmological part of his poem Parmenides had posited an original duality of fire and night, and explained the development of cosmic order and life on earth as the work of Aphrodite mixing the two basic forms (Parmenides fr. 13). Empedocles implies that each strategy is half right: separation produces a differentiated universe, but mixture the biosphere.

There is in fact reason to think that Empedocles wanted to be read as among other things an encyclopedist of previous thought. For example, the forms generated in the early phases of evolution are not just the stuff of dreams, but also recapitulate and rationalize myth. Most notable of these is the figure of the Minotaur, surely the model of Empedocles' 'ox-headed man-natured' being (fr. 61). At the opposite extreme, his verses describing the divine sphere deliberately recall Xenophanes' god (see Xenophanes §3) - 'No twin branches spring from his back, no feet, no nimble knees, no fertile parts' (Empedocles, fr. 29) - and that of Parmenides' being - 'Thus he is held fast in the close obscurity of *harmonia*, a rounded sphere rejoicing in joyous solitude' (Empedocles, fr. 27).

6 Biological explanations

The power of Empedocles as a thinker and his gifts as a poet are most happily married in the surviving extracts of his biology, where he shows an extraordinary capacity to get the reader to *see* the kinship of all nature. Sometimes parts of animals with homologous functions are just listed: 'The same things are hair and leaves and close-packed wings of birds and scales, coming into being on sturdy limbs' (fr. 82). Sometimes metaphor makes the point: 'Thus do tall trees lay eggs: first olives… ' (fr. 79). Other animals are armed with horns or teeth or stings, but 'Sharp-speared hairs bristle on hedgehogs' backs' (fr. 83).

Perception was explained by an ingenious theory of pores and effluences. It occurs when pores in the sense organ are just the right size to admit effluences of shape, sound, etc. given off by things (A92). The theory was employed to account for other phenomena also, such as chemical mixture (frs 91-2, A87) and magnetism (A89). Only like things (or things made like by Love) could have symmetrical pores and effluences. Empedocles posits fire in the eye for perception of light colours, a resounding bell in the ear to account for hearing, and breath in the nose for smelling (A86). Little of this material survives in his own words, except for an extended simile:

> As when someone planning a journey through the wintry night prepares a light, a flame of blazing fire, kindling for whatever the weather a linen lantern, which scatters the breath of the winds when they blow, but the finer light leaps through outside and shines across the threshold with unyielding beams: so at that time did she [Aphrodite] give birth to the round eye, primeval fire confined within membranes and delicate garments, and these held back the deep water that flowed around, but they let through the finer fire to the outside.
>
> (fr. 84)

Empedocles made thinking a function of blood around the heart (fr. 105). Of all physical substances blood comes closest to an equal blending of all the elements, even though its origin still owes something to chance (fr. 98). This is what equips it to grasp the nature of things: earth with earth, and so on - but also Love with Love and Strife with Strife (fr. 109).

7 Sacrifice

Empedocles relates his guilt over bloodshed to 'trust in raving strife' (fr. 115). As all creation of plants and animals is due to the power of Love, so Strife is the invariable cause of their dissolution and death. It is in his account of blood that Empedocles describes the mixture of the roots as 'anchored in the perfect harbours of Kypris [Love]' (fr. 98). Hence to kill by spilling blood is to act against the principle that makes for all that flourishes and enjoys harmony. It is no surprise that Empedocles conceives it as madness.

Theophrastus tells us that in his account of 'sacrifices and theogony', usually taken to belong to *Purifications*, Empedocles portrayed a golden age when humans recognized only Love as a god: 'They did not count Ares a god nor Battle-cry, nor was Zeus their king nor Kronos nor Poseidon, but Kypris [that is, Aphrodite] was queen' (fr. 128). There was no animal sacrifice:

> Her they propitiated with holy images, with paintings of living creatures, with perfumes of varied fragrance and sacrifices of pure myrrh and sweet-scented frankincense, throwing to the ground libations of yellow honey. The altar was not drenched with the unspeakable slaughters of bulls, but this was held among humans the greatest defilement - to tear out the life from noble limbs and eat them.
>
> (fr.128, continued)

Empedocles envisages this golden age as a time when humans were in fact friends with the rest of animal creation: 'All things were tame and gentle to humans, both beasts and birds, and friendship burned bright' (fr. 130). Is the image of harmony painted in these passages intended as part of the cosmic history? According to Empedocles' theory of evolution Strife was more, not less, dominant in the past. We should infer that like most pictures of primal bliss, this too is designed to function principally as an ideal measure of contemporary misery and wickedness.

What Empedocles lays down for all time is a moral rule against killing living things, whose expression led Aristotle to treat it as a paradigm of how natural law is to be conceived: 'That which is the law for all extends unendingly throughout wide-ruling air and the boundless sunlight' (fr. 135). Particular Pythagorean injunctions

against touching beans and laurel leaves are recorded as Empedoclean (frs 140-1). As for the future, other verses promise the repentant sinner eventual release from the burden of reincarnation entailed by committing bloodshed:

> At the end they come among humans on earth as prophets, bards, doctors and princes; and thence they arise as gods highest in honours, sharing with the other immortals their hearth and table, without part in the sorrows of men, unwearied.
>
> (frs 146-7)

Empedocles no doubt found the killing of other humans as horrific as animal sacrifice - the lines in which he represents sacrifice *as* infanticide etc. trade on a particular form of revulsion on the part of the reader. But he directs his outrage at violence to life as such, and more specifically at the assumption that blood-shedding can form part of a proper form of worship. In other words, his protest aims both to broaden our moral horizons and to reform religion. It should be seen as a radical challenge to the entire cultural framework of the ancient Greek city-state.

MALCOLM SCHOFIELD

References and further reading

Bollack, J. (1965-9) *Empédocle*, Paris: Les Éditions de Minuit, 4 vols.(An edition of fragments of *On Nature* and associated testimonia, with Greek text, French translation and commentary; champions the interpretation of the cosmic cycle adopted here; sometimes eccentric, but contains a wealth of insight, particularly on Empedocles' vocabulary.)

Empedocles (*c*.495-*c*.435 BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 1, 276-375.(The standard collection of the ancient sources, both fragments and testimonia, the latter designated by 'A'; includes Greek text of the fragments with translations in German.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols.(The most detailed and comprehensive English-language history of early Greek thought; the long and sympathetic account of Empedocles, in volume 2 pages 122-265, is still the best treatment in English.)

Kahn, C.H. (1960) 'Religion and Natural Philosophy in Empedocles' Doctrine of the Soul', *Archiv für Geschichte der Philosophie* 42: 3-35; repr. in A.P.D. Mourelatos (ed.) *The Pre-Socratics*, Garden City, NY: Doubleday, 1974, 397-425.(A seminal article on the coherence of Empedocles' thought.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(A valuable survey of Presocratic philosophy, including texts and translations; the account of Empedocles presented in the biographical entry largely follows its interpretations.)

Kingsley, P. (1995) *Ancient Philosophy, Mystery and Magic: Empedocles and Pythagorean Tradition*, Oxford: Clarendon Press.(On Empedocles' physical system and its connection with Pythagorean traditions.)

Martin, A. and Primaveri, O. (1997) *L'Empédocle de Strasbourg*, Berlin: de Gruyter.(An edition of papyrus fragments of Empedocles in the possession of the library of the University of Strasbourg.)

Sedley, D.N. (1989) 'The Proems of Empedocles and Lucretius', *Greek, Roman and Byzantine Studies* 30: 269-96. (An attractive reconstruction of the proem to *On Nature*; revises the allocation of fragments between *On Nature* and *Purifications*.)

Solmsen, F. (1965) 'Love and Strife in Empedocles' Cosmology', *Phronesis* 10: 123-45; repr. in R.E. Allen and D.J. Furley (eds), *Studies in Presocratic Philosophy*, London: Routledge & Kegan Paul, vol. 2, 221-64.(The best argument for the single-cosmogony interpretation.)

Wright, M.R. (1981) *Empedocles: The Extant Fragments*, New Haven, CT, and London: Yale University Press. (An edition with translations, commentary and glossary; an indispensable aid to deeper study; contains a useful statement of the double-cosmogony interpretation.)

Zuntz, G. (1971) *Persephone*, Oxford: Clarendon Press.(A study of western Greek mystery religion; part 2 pages 181-274 re-edits *Purifications* and offers a powerfully suggestive interpretation.)

Empiricism

In all its forms, empiricism stresses the fundamental role of experience. As a doctrine in epistemology it holds that all knowledge is ultimately based on experience. Likewise an empirical theory of meaning or of thought holds that the meaning of words or our concepts are derivative from experience. This entry is restricted to epistemological empiricism. It is difficult to give an illuminating analysis of 'experience'. Let us say that it includes any mode of consciousness in which something seems to be presented to the subject, as contrasted with the mental activity of thinking about things. Experience, so understood, has a variety of modes - sensory, aesthetic, moral, religious and so on - but empiricists usually concentrate on sense experience, the modes of consciousness that result from the stimulation of the five senses.

It is obvious that not all knowledge stems directly from experience. Hence empiricism always assumes a stratified form, in which the lowest level issues directly from experience, and higher levels are based on lower levels. It has most commonly been thought by empiricists that beliefs at the lowest level simply 'read off' what is presented in experience. If a tree is visually presented to me as green I simply 'register' this appearance in forming the belief that the tree is green. Most of our beliefs - general beliefs for example - do not have this status but, according to empiricism, are supported by other beliefs in ways that eventually trace back to experience. Thus the belief that maple trees are bare in winter is supported by particular perceptual beliefs to the effect that this maple tree is bare and it is winter.

Empiricism comes in many versions. A major difference concerns the base on which each rests. A public version takes beliefs about what we perceive in the physical environment to be directly supported by experience. A phenomenalist version supposes that only beliefs about one's own sensory experience are directly supported, taking perceptual beliefs about the environment to get their support from the former sort of beliefs. The main difficulties for a global empiricism (all knowledge is based on experience) come from types of knowledge it is difficult to construe in this way, such as mathematical knowledge.

1 Versions of epistemic empiricism

There are broad distinctions within epistemology that affect empiricism as well as other positions. One can think of *knowledge* or of *justified belief* as based on experience. To simplify the discussion we concentrate on the latter whenever we get into the details, though it is sometimes convenient to speak in terms of knowledge. If, as is often supposed, knowledge is justified belief that meets further conditions, then the necessity of an empirical basis for justified belief will extend to knowledge as well. There is also the question of whether the justification of belief depends on what the belief is *based on* (what gives rise to it), or whether it simply depends on what the believer *has* in the way of possible grounds, reasons or evidence, whether made use of or not. We take the former as a basis for discussion. Thus the empiricism under consideration here holds that all justified beliefs acquire their justification from being based, directly or indirectly, on experience.

Although empiricism typically concentrates on sense experience, this is not the only possibility. It is plausible to suppose that one's introspective knowledge of one's feelings, desires and other mental states derives from one's experience of those states (see Introspection, epistemology of). There is also a religious empiricism that takes certain beliefs about God to be justified by being based on a (frequently non-sensory) experiential presentation of God (see James, W. §4; Religious experience). This entry is restricted to sensory empiricism.

Empiricism comes in stronger and weaker forms. One such distinction has to do with scope - whether the view takes all knowledge to be based on experience, or restricts this claim to knowledge of the physical universe, excluding, for example, mathematical and/or religious knowledge. There are also differences regarding the strength of support lower levels must give higher levels in order that the beliefs at the higher levels be justified.

2 A phenomenalist empirical base

A common-sense form of empiricism takes perceptual beliefs about the physical environment (external beliefs) to be directly supported by experience - such beliefs as 'that house is on fire', 'a rabbit just ran across the yard' or 'a car is parked in front of the house'. But there are considerations that have driven philosophers to retreat to phenomenal beliefs about one's own sensory experience as constituting the base. For one thing, external beliefs are

often not solely based on experience, but rest, at least in part, on other beliefs. Thus I typically recognize a book as my copy of *Principia Ethica* not just by the way it looks (many other copies look just like that), but also by the fact that it is on a shelf in a study that belongs to me. For another thing, external beliefs can be mistaken even if confidently based on sense experience. I may unhesitatingly form the perceptual belief that the car in front of the house is a Pontiac when it is a Buick. Many philosophers have felt that if empirical knowledge is to be worthy of the name, at least the foundations on which it rests must consist of absolutely certain knowledge that cannot possibly be mistaken (Lewis 1946). Finally, it has been held that since, when I form an external perceptual belief, I would, if I reflected on the matter, take the justification of that belief to be the sensory experience on which it is based, that shows that beliefs about sensory experience are more basic in our empirical knowledge. All this can drive one to take phenomenal beliefs as the only ones that are directly supported by experience, with external beliefs supported by the phenomenal beliefs.

And yet these considerations do not require that conclusion. It remains to be shown that beliefs must be mistake-proof to constitute an empirical base. And the fact that external perceptual beliefs are based on sense experience should not be confused with the claim that they are based on beliefs about sense experience. Finally, even if my belief that this is my copy of *Principia Ethica* is partly based on other beliefs, that would prevent it from figuring in the empirical base only on an extreme form of empiricism, one that requires knowledge to be based on beliefs that are justified solely by experience. There are also more moderate forms that take the empirical base to include beliefs that are partly based on experience. An alternative way of handling this point would be to restrict the empirical base to those external beliefs that are justified solely by experience, leaving others for the superstructure.

It is just as well that those considerations are not conclusive, for there are serious difficulties in resting empirical knowledge of the world on a purely phenomenal base. Despite centuries of strenuous effort, no one has succeeded in showing how knowledge of the public physical world can be derived from knowledge of one's own experience, at least if we confine ourselves to generally recognized modes of inference. It has been widely recognized at least since the time of Hume that if we try to base an inference from sensory appearances to external objects on an empirically established correlation between them, we cannot establish that the correlation holds unless we already have knowledge of each side of it (see Hume, D. §2). And how can we get knowledge of the external side without already having established some such correlations to go on? Some have tried to side-step this difficulty by arguing that our experience is best explained by supposing that it is due to public physical objects in the ways we usually suppose (BonJour 1985). But that argument has never been developed in a thoroughly convincing way. Finally, some have turned in desperation to phenomenalism, the view that physical object statements are to be analysed in terms of what experiences one would have under certain circumstances. To say that there is a car parked in front of my house is to say something about what sensory experiences a sentient subject would have under certain conditions (Lewis 1946). The hope of the phenomenalist is that if the physical world is not radically different in nature from sense experience, it will not be impossible to infer the former from the latter. But apart from problems that attach to even these inferences, the programme runs afoul of the fact, classically pointed out by Chisholm (1948), that when we try to give a phenomenalist interpretation of, for example, 'There is a car parked in front of my house', we cannot specify the conditions in which a subject would have the relevant experiences without using physical-object language to do so. For example, we must include in those conditions the physical orientation and physical condition of the subject. We have to presuppose what we are trying to analyse in order to give the analysis (see Phenomenalism).

3 A public empirical base

These considerations drive us back to taking external perceptual beliefs to be, at least in part, directly justified by sense experience. There are many ways of spelling this out. The major differences come from differences in the analysis of sensory experience. We can distinguish three traditional views on this:

(1) Sense-datum theory. It consists of immediate awareness of non-physical entities that are, so to speak, reifications of the ways external objects appear to our experience. Thus, on this view, a typical visual experience of a round, red ball would involve being directly aware of a round, red sense-datum, which is distinct from the ball itself in being non-physical and existing only as an object for sensory awareness (Price 1932).
(2) Theory of appearing. A sense experience is an awareness of an object (in veridical perception an external

object) appearing to one in a certain way (looking, sounding, tasting… a certain way).
(3) Adverbial theory. Sense experience should not be construed as the direct awareness of any object - internal or external. It is rather a way or mode of awareness, a way of being conscious. When I see a red round ball, my consciousness is a matter of sensing redly, roundly, and, perhaps, 'ball-ly' (Chisholm [1966] 1977).

It has been held that sense experience is to be understood as a process of acquiring perceptual beliefs or inclinations to such beliefs (Armstrong 1961). Others have advocated physicalist construals in terms of the stimulation of sense organs. But these innovations have not had much effect on empiricist epistemology.

One might think that on the sense-datum theory one could not suppose external perceptual beliefs to be directly justified by sense experience. For this theory postulates the awareness of an internal object that, so to speak, stands between the subject and the external object. Does that not imply that any beliefs about the external object would have to be based on beliefs about the sense datum? Not necessarily. Several of the most prominent sense-datum theorists in the first part of this century - G.E. Moore, C.D. Broad and H.H. Price - emphatically denied that any inference from sense-data to external objects is involved. Instead they maintained that when one forms the usual external belief upon becoming aware of a certain sense-datum, that belief is justified just by virtue of being so formed. This is, in effect, to take the belief to be justified by the sense experience in question.

In considering the accounts of direct empirical justification one gets on these different construals of sense experience it will be helpful to recognize that in every case we need to draw a connection between features of the sense experience and the content of the belief. A belief will be justified by an experience only if there is the right kind of 'match' between the two. The details of the match will vary, depending on the account of sense experience. The sense-datum theory will have to work out some way of 'projecting' characteristics of external objects from features of sense-data. No one has ever done this in a convincing way for a realist (non-phenomenalist) account of physical objects. With phenomenalism it is a different ball game, for there the only entities referred to even in external beliefs are sense-data. The other views of sense experience have an easier time of it here. Since the theory of appearing construes sense experience as a matter of the way objects (usually external objects) appear, there can be a direct match between what an object appears to be and what it is. If an object looks like a two-storey wooden house, one is justified in believing it to be a two-storey wooden house. The only complexity here comes from complete hallucinations - Macbeth supposing himself to see a dagger in front of him when there is no dagger there. The theory can deal with such cases by saying that one is justified in the perceptual belief that *X* is *P* if and only if it is just as if *X* appears to one as *P*. (To Macbeth it was just as if a dagger appeared before him, the handle towards his hand.) The adverbial theory does not have this problem, since it does not construe sense experience as consisting in some *object* appearing in a certain way. It can read the content of the belief directly off the way of sensing, assuming, as may not be the case, that it is intelligible to convert all perceivable features of external objects into *ways* of sensing.

The complexity involved in formulating principles of justification that relate belief content to experiential content is so staggering that more than one philosopher has sought to cut the Gordian knot by simply granting *carte blanche* to all perceptual beliefs (Chisholm [1966] 1977). To be sure, one cannot deny that some perceptual beliefs are ill formed, and even among those that are not some can be shown to be false. We must not forget the many contradictions between witnesses to automobile accidents, as well as the proverbial drunkard who 'sees snakes'. A great help at this point is the distinction between *prima facie* and unqualified justification. To say that a subject is *prima facie* justified in believing that *p* is to say that this belief will be justified provided there is nothing to invalidate that status - either sufficient reasons for regarding the belief as false or sufficient reasons for supposing that the justifying grounds do not do the job in this instance (see Knowledge, defeasibility theory of). If I have a visual presentation of an elephant sitting on my lawn, I am justified in supposing that there is an elephant on my lawn, unless I have sufficient reasons for supposing that there are no elephants in the vicinity, or sufficient reasons for supposing that my visual experience was produced in such a way as to make it an unreliable indication of what is before me. Thus we can reasonably hold that all perceptual beliefs are *ipso facto prima facie* justified, even though this status may be overridden in some cases (see Perception, epistemic issues in).

4 Problems about the superstructure

As we have been characterizing empiricism, it would seem to be committed to foundationalism, the doctrine that

all knowledge (justified belief) rests on a foundation of beliefs that are justified otherwise than by other beliefs, for example, by experience (see Foundationalism). The ‘empirical base’ of which we have been speaking is simply a special case of the ‘foundations’ of foundationalism. It is only a special case because there are other possibilities for foundations, for example, rationally self-evident truths. But remember that we have acknowledged forms of empiricism that do not restrict the empirical base to beliefs justified solely by experience. Do these forms count as foundationalism? Well, we can also recognize stronger and weaker forms of foundationalism along the same lines (BonJour 1985). Strong foundationalism holds that all knowledge rests on beliefs that are wholly justified by something other than beliefs. Weaker forms require of foundational beliefs only that they are at least partially justified by something other than beliefs. Thus a weaker empiricism will also be a weaker foundationalism.

Continuing in the foundationalist vein, an empiricist epistemology will include not only a doctrine of the empirical base, but also a doctrine of the ways in which other beliefs are justified by their relations to that base. Traditional accounts have concentrated on deductive and inductive inference. Any beliefs that can be deductively or inductively inferred from that base are, or can be, justified. How this works out depends on the scope of inductive inference. If it is restricted to simple enumeration (inferring a generalization from instances), much of what we ordinarily consider to be empirical knowledge will fail to pass the test. Consider testimony: much of what we think we know has been acquired by taking someone’s word for it. Are we always or usually in possession of inductive evidence that the testifier is reliable? That is, have we checked out the person’s testimony in a sufficient number of cases and found it to be usually accurate? Obviously not. With most of the information we glean from books we have no evidence worthy of the name for the reliability of the author. We simply take it that the author is to be trusted unless we have reason to the contrary (see Testimony). Here too we give *prima facie* credence to certain beliefs, supposing them to be justified in the absence of sufficient reasons to think otherwise. Something similar can be said of the arguments to the best explanation that we constantly employ in a variety of contexts, for example, in arriving at an explanation of Jim’s recent unfriendliness towards me. Once again we take the fact that this seems to us the best explanation of the empirical facts in question to justify us in accepting it, even though, in most cases, we lack any significant deductive or inductive reasons for that acceptance (see Inference to the best explanation). Thus, if empiricism is to be at all plausible, it will have to recognize modes of building up the superstructure other than the traditional deductive and enumeratively inductive modes (Chisholm [1966] 1977).

5 Criticisms of empiricism

Criticisms of empiricist epistemology have mostly been of two types. First, there are areas of generally accepted knowledge that seem not to be accounted for by empiricism. Most prominently and most obviously, there is a priori knowledge, knowledge based on something other than experience (see A priori). The least controversial examples of this are logic and mathematics. To take a simple example, it seems for all the world as if our knowledge that ‘2 + 2 = 4’ does not rest at all on sense experience. It seems that we can know this to be true just by considering the proposition. The proposition is *self-evident*. There is no need to support it by empirical evidence, nor could it be overthrown by empirical evidence. If we seem to perceive two apples and two more apples making a sum of five apples rather than four, we would reject the supposed perception rather than the arithmetical truth. Other empirically recalcitrant areas include high-level theory in science, aesthetic knowledge and religious knowledge. But perhaps these latter cases can be handled by sufficiently extended ways of getting the superstructure from the base (high-level theoretical science) or by recognizing modes of justifying experience other than sense experience (aesthetic and religious experience). Still, logic and mathematics remain a stumbling block.

Some empiricists, like John Stuart Mill, have sought to show that mathematical knowledge, contrary to first impressions, rests on empirical evidence after all (see Mill, J.S. §§2-4). But a more common empiricist tack in the twentieth century has been that such knowledge is a matter of tracing out the logical implications of the meanings of the constituent terms. We know that ‘2 + 2 = 4’ just by virtue of realizing that its truth follows from what is meant by ‘2’, ‘Čplus;’, ‘Čequals;’ and so on. We know that ‘p’ logically follows from ‘p and q’ just because of what is meant by ‘and’ (Ayer 1936). But how does this show it to be *empirical* knowledge? It would seem that knowledge that owes its status to our grasp of the meanings of words is as far from being supported by *experience* as knowledge of self-evident propositions. The answer is that those who take this line do not suppose themselves to have shown that the knowledge in question is empirical, but rather that it is not knowledge ‘of the world’, not knowledge of what things are like independent of our conceptual arrangements. They take this line to imply that

so-called logical and mathematical knowledge is restricted to the consequences of the way we conceptualize the world, and hence not the sort of thing we should expect to be based on experience. In the most radical version of this position, the claim is that this (so-called) knowledge falls outside the empiricist net because it is not really knowledge at all.

The second criticism is of an internal kind. Whereas the first objection was that there are areas of knowledge that empiricism cannot accommodate, the second objection is that even on the empiricist's home field knowledge claims rest, in part, on non-empirically based principles. The most widely advertised example concerns induction. Many philosophers have concluded that we cannot rationally infer generalizations from instances without assuming some principle of regularity. Suppose we have examined 5,000 samples of copper all of which are ductile. How does that warrant us in inferring that copper is always (or even almost always) ductile? How do we know that the next 5,000 samples will not fail to exhibit ductility? When we make inferences like this, are we not assuming some such general principle as that 'Regularities that hold in a large and varied sample will hold universally', or 'The future will resemble the past', or 'Properly chosen samples are representative of the whole class of which they are samples'? And are such principles themselves justified by experience? If we were to suppose they are, would that not require us to assume such principles in order to validate the inference? And that would make the empirical justification circular. Hence it seems that empirical induction depends for its validity on principles that cannot be empirically justified. And that would seem to be an absolute limit on the extent to which we can take knowledge, even in the 'empirical' sphere, to be wholly justified by experience (Russell 1948) (see Induction, epistemic issues in).

See also: A posteriori; Epistemology, history of; Innate knowledge; Rationalism

WILLIAM P. ALSTON

References and further reading

Alston, W.P. (1991) *Perceiving God: The Epistemology of Religious Experience*, Ithaca, NY: Cornell University Press.(A defence of the thesis, mentioned in §1 above, that religious experience is a source of epistemic justification for certain kinds of beliefs about God.)

Alston, W.P. (1993) *The Reliability of Sense Perception*, Ithaca, NY: Cornell University Press.(A fairly technical criticism of attempts to establish the reliability of sense perception without relying on sense perception to do so.)

Armstrong, D.M. (1961) *Perception and the Physical World*, London: Routledge & Kegan Paul.(General account of perception.)

Ayer, A.J. (1936) *Language, Truth, and Logic*, London: Victor Gollancz Ltd, 2nd edn,1948.(Classic presentation of logical empiricism.)

Bealer, G. (1992) 'The Incoherence of Empiricism', *The Aristotelian Society*, supplementary vol. 66: 99-138. (Powerful internal criticism of empiricism.)

BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.(An accessible presentation of a coherentist account of empirical knowledge.)

Chisholm, R.M. (1948) 'The Problem of Empiricism', *Journal of Philosophy* 45 (19): 512-17.(The classic critique of phenomenalism.)

Chisholm, R.M. (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 2nd edn, 1977.(The second edition provides a very influential, concise presentation of a foundationalist epistemology.)

Lewis, C.I. (1946) *An Analysis of Knowledge and Valuation*, La Salle, IL: Open Court.(The definitive account of a phenomenalist theory of empirical knowledge.)

Moser, P. (1989) *Knowledge and Evidence*, Cambridge: Cambridge University Press. (A difficult but comprehensive presentation of an empiricist foundationalism.)

Pollock, J. (1974) *Knowledge and Justification*, Princeton, NJ: Princeton University Press.(An advanced presentation of a foundationalist account of empirical knowledge.)

Price, H.H. (1932) *Perception*, London: Methuen, 2nd edn, 1950.(A classic presentation of the sense-datum theory of perception.)

Russell, B.A.W. (1948) *Human Knowledge: Its Scope and Limits*, London: Allen & Unwin.(The distillation of decades of Russell's thought about empirical knowledge.)

Encyclopedists, 18th-century

The Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers was published in seventeen folio volumes (about 20 million words) between 1751 and 1765, accompanied by eleven volumes of engravings (1762-72). Its chief editor was Diderot, with D'Alembert acting as co-editor for the first seven volumes. The work was an expression of the Enlightenment belief in improvement through knowledge.

From the outset the *Encyclopédie* attracted hostile criticism, especially from ecclesiastical circles; publication was halted twice and the last ten volumes had to be collected by subscribers from outside Paris. Because of its comprehensive scope, the wide range of its contributors, and the prominence which it owed as much to its critics as its supporters, the work came to be regarded as an embodiment of the French Enlightenment. Numerous editions (and volumes of extracts) were published in the 1770s and 1780s, as well as four supplementary volumes and two volumes of indices.

In his article 'Encyclopédie' in volume V (1755), Diderot expressed an intention 'of changing the general way of thinking'; in his *Discours préliminaire* to volume I (1751), D'Alembert provided the fullest account of the principles guiding this ambition. They were epitomized by the phrases 'esprit philosophique' or 'esprit systématique', by which was meant a critical, enquiring, rational approach, opposed to prejudice, obscurity, superstition, intolerance and dogmatism (whether that emanating from established institutions or that evident in the 'esprit de système' of a priori philosophical doctrines). The empirical method of Locke, the scientific principles of Newton, and the practical benefits of knowledge advocated by Francis Bacon, were held up as models; the writings of Bayle, Fontenelle and Montesquieu were significant influences. The practical application of thought, in terms of both social reform and the mechanical arts, received particular emphasis; a humanitarian concern and a desire for freedom (intellectual, civil, economic, political) were recurring themes, although the work was not politically radical. The philosophical articles were not notable for their originality, (although the occasional discussion of materialist ideas was seen as dangerously new by some); many took the form of a historical survey, written by Diderot but drawing heavily on J.J. Brucker's *Historia critica philosophiae*.

The title-page announced that the *Encyclopédie* was written by 'une société de gens de lettres', over a hundred and thirty of whom have been identified. Apart from contributions from Voltaire, who wrote over forty articles of minor interest, and some by figures who later became famous - such as the editors themselves, Rousseau, Turgot, d'Holbach, Quesnay, or Marmontel - most were by individuals who were little known or of whom we otherwise know nothing (see Rousseau, J.-J.). Generally speaking they seem to have been experts in their fields: professional men, academicians, civil servants, scientists, doctors, men of letters, engineers, manufacturers, and so on. Almost a quarter of the work was written by one dedicated, indefatigable and (regrettably) undistinguished contributor, the Chevalier de Jaucourt. The diversity of these figures, the variety of views expressed and the unevenness of the articles prevent any easy categorization of these writers, but their participation in the enterprise itself demonstrated a widespread faith in the power of ideas to bring about human improvement.

See also: Enlightenment, Continental

JOHN HOPE MASON

List of works

Diderot, D. and D'Alembert, J. Le R. (eds) (1751-65) *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers*, Paris; trans. N. Hoyt *et al.*, *Encyclopedia*, New York, 1965.(The translation contains thirty-nine articles.)

D'Alembert, J. Le R. (1751) *Discours préliminaire*, intro. and trans. by R.N. Schwab, *Preliminary Discourse*, New York, 1963.

Hope Mason, J. (1982) *The Irresistible Diderot*, London.(Contains extracts from sixty articles by Diderot, in translation.)

References and further reading

Kafker, F.A. and Kafker, S.L. (1988) *The Encyclopedists as individuals: a biographical dictionary of the 'Encyclopédie'*, Studies on Voltaire and the Eighteenth Century 257, Oxford: The Voltaire Foundation.

(Comprehensive study of the contributors.)
Lough, J. (1968) *Essays on the Encyclopédie of Diderot and D'Alembert*, Oxford. (Detailed study of certain articles.)
(1971) *The Encyclopédie*, London.(Best general survey.)
Proust, J. (1962) *Diderot et l'Encyclopédie*, Paris.(Masterly account of Diderot's thought in relation to his articles.)
(1965) *L'Encyclopédie*, Paris.(Good introduction.)
Schwab, R.N. *et al.* (1971-84) *Inventory of the Encyclopédie*, Studies on Voltaire and the Eighteenth Century 80, 83, 85, 91-3, 223, Oxford: The Volatire Foundation.
Wilson, A.M. (1972) *Diderot*, New York.(Detailed biography.)

Encyclopedists, medieval

The modern encyclopedic genre was unknown in the classical world. In the grammar-based culture of late antiquity, learned compendia, by both pagan and Christian writers, were organized around a text treated as sacred or around the canon of seven liberal arts and sciences, which were seen as preparatory to divine contemplation. Such compendia, heavily influenced by Neoplatonism, helped to unite the classical and Christian traditions and transmit learning, including Aristotelian logic, to the Middle Ages.

Writers in the encyclopedic tradition include figures such as Augustine and Boethius, both of whom were extremely influential throughout the medieval period. Other important writers included Macrobius, whose Saturnalia spans a very wide range of subjects; Martianus Capella, whose De nuptiis Philologiae et Mercurii (The Marriage of Philology and Mercury) covers the seven liberal arts and sciences; Cassiodorus, who presents the arts as leading towards the comtemplation of the heavenly and immaterial; and Isidore, whose Etymologies became one of the most widely referred-to texts of the Middle Ages. These writers also had a strong influence which can be seen later in the period, particularly in the Carolingian Renaissance and again in the twelfth century.

1 Learned compendia and the encyclopedic tradition

The classical world produced nothing like the modern encyclopedia with its universal range and disconnected entries in alphabetical order. Instead, ancient scholars preferred to unify their knowledge by literary or philosophical means. By the first century BC the term 'encyclopedia' was often used, but it generally denoted a set of interconnected disciplines fit for upper class freemen (hence 'liberal') and propaedeutic in nature. The philosophers might see such disciplines as leading the mind from the material to the immaterial world, and thence to metaphysical philosophy. For most, the set included both mathematical and linguistic arts. With much fluctuation, there evolved a canon of seven interrelated liberal arts and sciences: grammar, rhetoric and dialectical logic (the medieval trivium), and arithmetic, geometry, music and astronomy (the medieval quadrivium). During the first three centuries AD, Middle Platonist and Neoplatonist philosophers strongly influenced the solidifying of this canon; indeed, it has been seen (controversially) as primarily their invention (Hadot 1984) (see Neoplatonism; Platonism, early and middle).

Various types of learned compendium existed, and it is possible to distinguish three particular types or genres. Genre A can be described as miscellaneous information, presented haphazardly, perhaps in the attractive and memorable dialogue form, as a gentleman's 'learned baggage' (to use Macrobius' phrase) and handbook of civilized conversation. An example is the *Attic Nights* of Aulus Gellius, written in the second century AD. In genre B, information is organized by disciplines, whether in a sequence of treatises or as a unitary work. Varro (116–-27 BC) probably composed an encyclopedic sequence of volumes on the seven arts and sciences, with medicine and architecture added. Varro allegedly saw the mathematical four as leading to contemplation of the immaterial; in this he may have been influenced by Neopythagorean philosophy (see Neo-Pythagoreanism). In the first century AD, in his *Natural History*, the elder Pliny united his subject by themes of nature's divinity and human partnership with it; this allowed expansion into a wide range of information.

In the culture of late antiquity, dominated by the grammarian, a third type of compendium developed. In genre C, compendia were centred around some revered text, classical or biblical. These texts, which were often seen as inspired, could be allegorically interpreted and used as a source of cosmological, spiritual and metaphysical wisdom. Isidore, for example, drew extensively for his encyclopedia on both the *Hexaemeron (Commentary on Genesis)* of Ambrose (AD 339-97) and the commentaries on Vergil of Servius (*fl. c.*AD 400), whom Macrobius depicted as the model grammarian. To understand such sacred texts, encyclopedic learning, especially the skill in literary analysis owed to grammar and rhetoric, was seen as necessary and the texts themselves might serve as educational exemplars (see Patristic philosophy).

2 Augustine

As a young man, Augustine was strongly influenced by the Neoplatonists. Following his conversion to mainstream Christianity, he proposed a sequence of works on the seven arts and sciences, perhaps substituting philosophy for

astronomy. He saw these as propaedeutic to a theology of Christian Neoplatonist type Hadot 1984). He also increasingly saw them as handmaids to biblical exegesis (see his very influential *De doctrina christiana (On Christian Instruction)*). He apparently completed works on grammar and music only (the latter in dialogue form), but his *De civitate Dei (The City of God)* transmitted a vast range of learning in its replies to polytheists and pagan Neoplatonists.

3 Macrobius

Macrobius' *Saturnalia*, written *c*.AD 430, but set in the 380s, belongs partly to genre A. In their form, and as an exemplar of scholarly manners, these prandial conversations between Roman gentlemen owe much to Cicero and Aulus Gellius. However, their core (Books III-VI) consists of discussions of Vergil, seen as universally learned and inspired. Partly through and around these discussions, Macrobius transmitted views and information on literary style, allegorical technique, pagan ritual and theology, astronomy, physiology, etymology and general antiquarian topics.

For Macrobius, the other three inspired authors were Homer, Plato and Cicero. His Platonizing commentary on *The Dream of Scipio* (from Cicero's *Republic*) helped to transmit classical cosmological science to the Middle Ages, integrating it with Neoplatonist theology. This work is also important for its numerology, classification of the virtues and presentation of Platonist and Aristotelian disputes on the nature of the soul. It owed much to commentaries on Plato's *Timaeus* and *Republic* by Porphyry, and notably influenced the thirteenth-century encyclopedist Vincent of Beauvais.

4 Martianus Capella

Like Macrobius, Martianus was probably a pagan; he wrote at Carthage, perhaps around AD 470. His *De nuptiis Philologiae et Mercurii (The Marriage of Philology and Mercury)* belongs partly to genre B, systematically expounding and summarizing the Varronian canon of arts and sciences (omitting medicine and architecture). However, the work (a 'Menippean' mixture of prose and verse, like Varro's satires) also depicts, in an allegorical myth, the Neoplatonic apotheosis of the learned human soul, which returns to its heavenly home to become the bride of a god; on this marriage the arts and sciences attend as bridesmaids. Thus Martianus audaciously improved on Macrobius by creating his own sacred text: ' With its description of the kosmos, the fall and ascent of the soul, theurgical initiation, and an apotheosis, the work can be seen as a summa of pagan knowledge, religious and secular' (Shanzer 1986: 28).

Paradoxically, this was a work which, in terms of both form and content, had a profound influence on the Christian Middle Ages. In the ninth century, Johannes Scottus Eriugena drew extensively on it for his cosmology; in the tenth century, Gerbert of Aurillac made much use of the books on the quadrivium. Its influence was especially strong on the Platonists and cosmographers associated with the twelfth-century School of Chartres (see Chartres, School of). Alan of Lille owed much to it in his *De planctu naturae*, and also in his *Anticlaudianus*, a summa of the seven arts and sciences which depicts the ascent of Phronesis, with their help, to the Empyrean. Bernard Of Tours similarly has an ascent of Natura in his Menippean *De mundi universitate*.

Book IV of *De nuptiis*, on dialectic, was especially influential in the ninth to eleventh centuries. Martianus' logic was basically Aristotelian and syllogistic, depending much on Porphyry's *Isagoge*, which he perhaps knew in the fourth-century translation by Marius Victorinus; he may well have used a *Perihermeneias* (perhaps that of Apuleius in the second century) for categorical propositions, and the categorical syllogism. He was also affected by Stoic development of the hypothetical syllogism and by Stoic propositional logic, derived ultimately from Varro (see Stoicism). Like Boethius and Cassiodorus, Martianus transmitted the so-called *logica vetus* (old logic); this was partly superseded during the twelfth century by the *logica nova* (new logic), through the discovery of more sophisticated works of Aristotle (see Aristotelianism, medieval; Language, medieval theories of; Logic, medieval).

5 Boethius

Boethius, writing in sixth-century Italy, was much influenced by the Neoplatonism of Macrobius and the fifth century Alexandrians. He planned, but never finished, complete translations of Plato and Aristotle and a series of introductions and translations for the seven arts and sciences; he is therefore another encyclopedic author. His

Institutio arithmetica (Introduction to Arithmetic) and *Institutio musica (Introduction to Music)* survive, as do a commentary on Cicero's *Topics*, translations and commentaries for Porphyry's *Isagōgē* and Aristotle's *Prior Analytics*, *Categories* and *Perihermeneias*, and translations of Aristotle's *Topics* and *Sophistic Refutations*. In five treatises, he gave his own introduction to Peripatetic logic.

Eriugena, Gerbert of Aurillac, Abbo of Fleury, Anselm and Abelard all owed much to Boethius. His logic influenced the debate between realists and nominalists, and remained popular until the end of the fifteenth century. He also transmitted Pythagorean musical theory and Euclidean geometry.

6 Cassiodorus

Boethius' kinsman Cassiodorus, writing also in the sixth century, was a rhetorician rather than a philosopher and was an encylopedist in all three genres. His *Variae* (letters drafted for barbarian monarchs) digress into a wide range of learning offering a model of cultivated discourse between ruler and subject, and were exploited later in the Middle Ages. His *Expositio Psalmorum (Exposition on the Psalms)* is partly conventional exegesis, but also developed the Psalms for Christians as an encyclopedic exemplar of the arts and sciences, especially rhetoric.

Cassiodorus' *Institutiones*, partly an intellectual rule for his monastery of Vivarium, were also meant for a wider audience, superseding a project for a Roman school of Christian higher education. Book I is a kind of patristic and monastic encyclopedia. It gives an expanded bibliography of biblical texts, commentaries and works of relevant theology, as well as advice on exegesis, a list of church councils and information on history, geography, medicine, gardening and textual copying. Book II, which achieved a separate and wider circulation, is devoted to the seven arts and sciences. Like Augustine, Cassiodorus regarded these as essential to scriptural study; he also presents them as leading the soul towards the heavenly and immaterial, to the point where the Bible will deepen awareness of God. He treats the seven methodically, analysing and summarizing their salient points and advising on further reading. His canon is that of Varro and Martianus Capella, but he owes something also to the tradition of the fifth century Neoplatonists (especially, perhaps, Ammonius). From the latter he derives the order of his treatment, his influential distinction between the arts of the trivium and the disciplines of the quadrivium (the former deal with contingent matters, the latter with necessities) and his analysis of the subjects of philosophy: the disciplines belong to doctrinal or mathematical philosophy, which is in turn a branch of speculative philosophy.

His section on dialectic is based partly on the translations and commentaries of Boethius and Marius Victorinus, but perhaps also on Ammonius' commentary on Porphyry's *Isagōgē. Institutiones* II proved a highly valuable handbook in the early Middle Ages, and specially influenced two major figures of the Carolingian Renaissance, Alcuin and Hrabanus Maurus. The latter's *De institutione clericorum* integrated the arts and sciences into clerical education, as Cassiodorus had done (see Carolingian Renaissance).

7 Isidore of Seville

Isidore's *Etymologies* or *Origins*, written in the early seventh century, was a new - though still grammar-based - departure, organizing knowledge around the histories and definitions of individual words. Indebted ultimately to Stoic theories of etymology, but more directly to Augustine's remarks on semiotics (especially in *De doctrina Christiana* II), Isidore assumed that the name of a thing gives a key to its nature. On this principle (partly rational, but also reflecting Neo-Pythagorean and biblical mysticism) he founded a twenty-book compendium, encyclopedic in range even by modern standards. Only the first four books deal with the seven arts and sciences (medicine is added as a second philosophy, healing the body as philosophy does the soul), and the dictionary method precludes extensive expositions.

Following the tradition, Isidore sees the canon as steps to divine contemplation. From Book V onwards the work is less systematic, though by no means random in construction. Isidore follows trains of thought in which (for example) Christian studies (Books VI-VIII) conclude with a catalogue of heresies; heresies suggest pagan philosophies, philosophies suggest classical poetry and poetry suggests sibyls, mages and diviners. Book II, on rhetoric and dialectic, is based mainly on Cassiodorus, but Isidore made additions: there is a section on law, based on Tertullian, and another on opposites (ultimately Aristotelian, and also treated by Martianus Capella). Etymologies in this book yielded useful translations of Greek technical terms.

Second-hand Stoicism influenced Isidore's views on cosmology and morality (see Stoicism). Although

condemning Epicureanism, he retained the atomic theory, and extended it to numbers and language (Book XIII). He gave a brief survey of ancient philosophy (Book VIII), valuable despite its crudities. Also valuable was his exclusion of Christian theology and revelation from his logical definitions, and his strong assertion of the worth of dialectic (Book II); he affirmed the continuing relevance of philosophy, despite its dangers to the Christian (Books IV and VIII). The simple Latin, brief entries and lists of contents made the *Etymologies* ideal for reference; its survival in over 1,000 manuscripts attests its popularity. Curtius (1953: 23) has called it the basic book of the entire Middle Ages. Hrabanus Maurus adopted Isidore's method, and much of his contents, in his encyclopedia *On the Nature of Things*.

8 Conclusion

Late antiquity was a period of translations, florilegia, abridgements and other short guides to a variety of disciplines. As Boethius saw, these were vital when Greek east and Latin west were drifting apart, and educational institutions declining. It was also a time of tension between Christianity and the classical tradition. In the preservation of learning and eventual combination of the two roads to wisdom, the encyclopedists, pagan and Christian, played an essential part, thanks largely to the adaptability and spiritual attractions of the Neoplatonic philosophy which so strongly influenced their subject matter.

See also: Marius Victorinus; Patristic philosophy

SAMUEL BARNISH

References and further reading

Augustine (386) *De musica (On Music)*, ed. J.-P. Migne, *Patrologia Latina*, vol. 32, Paris: J.-P. Migne, 1861; trans. R.C. Taliaferro in *Fathers of the Church*, vol. 4, New York: Fathers of the Church Inc., 1947.

Augustine (*c*.395-8) *De doctrina Christiana (On Christian Doctrine)*, ed. J. Martin, *Corpus Christianorum*, Series Latina 32, Turnhout: Brepols, 1962; trans. J.J. Garrigan, Christian Instruction, in *Fathers of the Church*, vol. 2, New York: Fathers of the Church Inc., 1947.(A Bible-centred treatise on Christian education and rhetoric.)

Augustine (413-27) *De civitate Dei (The City of God)*, ed. B. Dombart and A. Kalb, Leipzig: Teubner, 1928-9; trans. H. Bettenson, *The City of God*, Harmondsworth: Penguin Books, 1972.(A wide-ranging response to pagan criticism of Christianity, and exposition of Christian political theology.)

Boethius, A.M.S. (*c*.500-*c*.522) *Collected Works*, ed. J.-P. Migne, *Patrologia Latina*, Paris: J.-P. Migne, 1860, vols 63-4.

Boethius, A.M.S. (*c*.500-*c*.522) *Institutio arithmetica (Introduction to Arithmetic)*, ed. G. Friedlein, Leipzig: Teubner, 1867; trans. M. Masi, *Boethian Number Theory: A Translation of the 'De Institutione Arithmetica'*, in Studies in Classical Antiquity, vol. 6, Amsterdam: Rodopi, 1983.

Boethius, A.M.S. (*c*.500-*c*.522) *Institutio musica (Introduction to Music)*, ed. G. Friedlein, Leipzig: Teubner, 1867; trans. C.M. Bower, New Haven, CT: Yale University Press, 1989.

Boethius, A.M.S. (*c*.500-*c*.522) *Commentarii in librum Aristotelis Peri hermenias (Commentary on Aristotle's Perihermeneias)*, ed. K. Meiser, Leipzig: Teubner, 1887-90.

Boethius, A.M.S. (*c*.500-*c*.522) *In Isagogen Porphyrii Commenta (Commentaries on Porphyry's Isagōgē)*, ed. S. Brandt, Corpus Scriptorum Ecclesiasticorum Latinorum 48, Vienna and Leipzig: Tempsky & Freytag, 1906.

Boethius, A.M.S. (*c*.500-*c*.522) *De hypotheticis syllogismis (On Hypothetical Syllogisms)*, ed. L. Obertello, *Logicalia*, vol. 1, Brescia: Paideia editrice, 1969.

Boethius, A.M.S. (*c*.500-*c*.522) Aristotelian translations and translation of Porphyry's *Isagoge*, ed. L. Minio-Paluello and B. Dod, *Aristoteles Latinus*, Brussels, Leiden, Bruges and Paris: Desclée De Brouwer, Brill, 1961-75.

Boethius, A.M.S. (*c*.500-*c*.522) *De topicis differentiis (On Topical Differences)*, trans. and commentary by E. Stump, *Boethius's De topicis differentiis*, Ithaca, NY: Cornell University Press, 1978.

Cassiodorus (506-37) *Variae*, ed. T. Mommsen, Berlin: Weidmann, 1894; selection trans. S.J.B. Barnish, *Variae*, Liverpool: Liverpool University Press, 1992.(State documents with learned digressions, written for the Ostrogothic kings of Italy.)

Cassiodorus (*c*.540-*c*.555) *Expositio Psalmorum (Exposition of the Psalms)*, ed. M. Adriaen, Corpus Christianorum, Series Latina 97-8, Turnhout: Brepols, 1958; trans. P.G. Walsh, *Explanation of the Psalms*, New York and Mahwah, NJ: Paulist Press, 1990-1.(A commentary on the Psalms which uses them as a textbook of rhetoric, logic and miscellaneous learning.)

Cassiodorus (*c*.562) *Institutiones (Institutes)*, ed. R.A.B. Mynors, Oxford: Clarendon Press, 1937; trans. L.W.

Jones, *An Introduction to Divine and Human Readings by Cassiodorus Senator*, New York: Columbia University Press, 1946.(A Christian guide to sacred and secular learning.)

Cahiers d'Histoire Mondiale (1966), 9 (3).(A useful volume of articles on the history of encyclopedias; see especially articles by P. Grimal (459-82), M. de Gandillac (483-518) and J. Fontaine (519-38).)

Courcelle, P. (1969) *Late Latin Writers and their Greek Sources*, Cambridge, MA: Harvard University Press. (Translation from a French original. The work has chapters on Macrobius, Augustine, Boethius and Cassiodorus; influential, but to be used with caution.)

Curtius, E.R. (1953) *European Literature and the Latin Middle Ages*, London: Routledge & Kegan Paul. (Translation from a German original by W.R. Trask. Useful on the role of rhetoric and etymology, and the afterlives of the encyclopedists; a literary rather than philosophical treatment.)

Fontaine, J. (1984) *Isidore de Séville et la Culture Classique dans l'Espagne Wisigothique (Isidore of Seville and Classical Culture in Visigothic Spain)*, 2nd edn, Paris: Études Augustiniennes.(The fundamental study of Isidore in his own time.)

Gibson, M.T. (ed.) (1981) *Boethius: His Life, Thought and Influence*, Oxford: Blackwell.(Valuable chapters by J. Barnes, J. Caldwell, O. Lewry, and D. Pingree on Boethius as philosopher, musicologist and geometer, and his influence.)

Hadot, I. (1984) *Arts Libéraux et Philosophie dans la Pensée Antique (The Liberal Arts and Philosophy in Ancient Thought)*, Paris: Études Augustiniennes.(Controversially ascribes the canon of the seven arts and sciences to the Neoplatonists and Augustine; highly critical of Marrou (1938).)

Irvine, M. (1994) The Making of Textual Culture: 'Grammatica' and Literary Theory, 350-1100, Cambridge: Cambridge University Press.(An imaginative study, putting the encyclopedists in the context of a grammarians' world.)

Isidore (*c*.620-35) *Etymologiae (Etymologies)*, ed. W.M. Lindsay, Oxford: Clarendon Press, 1911.(An encyclopedia organized around the histories and definitions of words. Translations of the separate books in various languages are in progress under the auspices of the Centre National des Recherches Scientifiques; among these note P.K. Marshall's English translation of Book II, Paris: Éditions des Belles Lettres, 1983.)

Laistner, M.L.W. (1957) *Thought and Letters in Western Europe, A.D. 500-1900*, London: Methuen.(Has material on Hrabanus Maurus, Cassiodorus and Isidore. Standard work, but should be used with caution.)

Macrobius (probably *c*.430) *Saturnalia*, ed. J. Willis, Leipzig: Teubner, 1970; trans. P.V. Davies, New York: Columbia University Press, 1969.(A symposiastic dialogue whose core is devoted to Vergil.)

Macrobius (probably *c*.430) *Commentarii in Somnium Scipionis (Commentary on the Dream of Scipio)*, trans. W.H. Stahl, New York: Columbia University Press, 1952.(A Platonizing commentary on a myth from Cicero's *Republic*.)

Marrou, H.-I. (1938) *Saint Augustin et la Fin de la Culture Antique (Saint Augustine and the End of Ancient Culture)*, Bibliothèque des Écoles Françaises d'Athènes et de Rome, vol. 45, Paris: Éditions de Boccard. (Fundamental on the evolution of the canon of the seven arts and sciences in relation to Augustine.)

Martianus Capella (perhaps *c*.470) *De nuptiis Philologiae et Mercurii (The Marriage of Philology and Mercury)*, ed. J. Willis, Leipzig: Teubner, 1983; trans. W.H. Stahl and E.L. Burge, *The Marriage of Philology and Mercury*, vol. II, trans. W.H. Stahl and E.L. Burge, *Martianus Capella and the Seven Liberal Arts*, New York: Columbia University Press, 1977.(Encyclopedic learning presented in an allegorical myth. Volume I of this work (with R. Johnson) gives a full introduction to Martianus Capella.)

O'Donnell, J.J. (1979) *Cassiodorus*, Berkeley, CA: University of California Press.(Useful, but uninspired.)

Shanzer, D.R. (1986) *A Philosophical and Literary Commentary on Martianus Capella's De Nuptiis Philologiae et Mercurii*, Book I, Berkeley, CA: University of California Press.(As an introduction to Martianus, less comprehensive but more intelligent than Stahl and Burge's first volume.)

Engels, Friedrich (1820-95)

Until the 1970s the most influential framework for understanding Marx's career and ideas was the one established by Engels. This framework was crucially related to his understanding of philosophy and its supposed culmination in Hegel's systematic and all-encompassing idealism.

Engels claimed that Marx had grounded Hegel's insights in a materialism that was coincident with the physical and natural sciences of his day, and that Marx had identified a dialectical method applicable to nature, history and thought. With respect to history, Marx was said to have formed a 'materialist conception', from which his analysis of capitalist society and its 'secret' of surplus value were derived. Together these intellectual features were the core of the 'scientific socialism' which, Engels argued, should form the theory, and inform the practice, of the worldwide socialist or communist movement. This was to abolish the poverty and exploitation necessarily engendered, he claimed, by modern industrial production.

Philosophically the tenets of dialectical and historical materialism have been defended and modified by orthodox communists and non-party Marxists, and expounded and criticized by political and intellectual opponents. The three laws of dialectics, and the doctrine that history is determined by material factors in the last instance, have been attacked as tautologous and indeterminate. Engels's view that scientific socialism is a defensible representation of Marx's project has also been challenged by textual scholars and historians.

1 Life and works

Engels rightly described himself as an 'autodidact in philosophy', and neither planned nor realized a career as a philosopher in his lifetime. Yet his works represent the founding texts of Marxism and of Marxist philosophy, in that they set an authoritative context through which to interpret the ideas of Karl Marx in a specifically philosophical way.

Born into a wealthy Rhineland family, and growing up amidst the disruption of rapid industrialization, the teenage Friedrich embarked on a paradoxical career. All his life he was reluctantly associated with the family's business interests, though politically he sympathized with the plight of the industrial working classes. Pursuing his political interests as a revolutionary populist required him to address intellectual elites and doctrinaires, particularly those within the burgeoning socialist movement and tiny communist parties. Yet the issues addressed in those works became ever more rarified and philosophical, despite the scorn that he always evinced for 'mere' philosophers and dabbling cranks, and his enthusiasm for political action.

Most paradoxically of all, Engels pronounced himself 'junior partner' to Marx, yet for many years the only way to identify, explicate and defend the *philosophical* content of Marx's work was through *Herrn Eugen Dührings Umwälzung der Wissenschaft* (*Anti-Dühring*), *Socialisme utopique et socialisme scientifique* (*Socialism: Utopian and Scientific*) and *Ludwig Feuerbach und der Ausgang der klassischen deutschen Philosophie* (*Ludwig Feuerbach and the End of Classical German Philosophy*), all written by Engels and published under his name alone. The claim that those works represent the views of Marx rests on retrospective textual exegesis, theories of 'partnership' and 'division of labour' between the two, and negative 'evidence' that Marx, who was Engels's economic dependent from the 1850s until his death in 1883, had opportunities to repudiate what Engels was saying. Engels himself contributed to this narrative, establishing his own works as the authoritative interpretation of Marx. Hence his self-description as 'second fiddle' is in reality somewhat paradoxical.

Bizarrely, Engels is also given credit for 'assuming paternity' for the illegitimate son of the Marx family's housemaid, supposedly to save Marx embarrassment and his wife the scandal. But the documentary evidence for this tale is highly suspect, and the view that paternity of this child was an issue for anyone at all before the 1960s, when the story first surfaced, rests on reading this lurid narrative back into the ambiguities of contemporary correspondence (Carver 1989: 162-71).

Engels had the image of a lifelong bachelor in public and a discreet bohemian in private, and his successive liaisons with the working-class Burns sisters (Mary and 'Lizzie') are well attested. Yet they now seem to fit all too easily into the nineteenth-century pattern wherein bourgeois males could 'keep' women of an inferior class in the suburbs or employ them in a respectable function as housekeeper, whilst presuming in every way on their

economic dependence. Engels married ‘Lizzie’ only when it was clear that she was dying and could not be Mrs/Frau Engels in his Anglo-German milieu, nor survive him as a claimant on his estate. These liaisons were childless, and it is likely that both women were lifelong illiterates, or nearly so.

Engels himself was an accomplished writer from his early twenties in German and English, fluent in French, and later a student and correspondent in numerous other languages. His library and papers, including the collection of manuscripts and other materials inherited from Marx, represented an important intellectual resource for the socialist and communist movement from his death until the dissolution of Soviet-style communism in the 1980s, not least the notebooks posthumously published as *Dialektik der Natur* (*The Dialectics of Nature*), from which an ‘official’ philosophy of science was derived (see Dialectical materialism).

2 Political theory

By the time he was nineteen, Engels’s intellectual and political interests had moved from a heady combination of revolutionary liberalism and romantic nationalism to the Young Hegelian perspective then current in the universities, though he was never officially able to be a student. From a twentieth-century perspective this movement seems an intellectual cabal, but Engels’s apprehension of these debates was overwhelmingly political. In a regime where politics itself was almost wholly an illegitimate activity, any form of criticism was by definition subversion and inevitably confined to a small circle of writers, publishers and readers.

Engels can be distinguished from others in the school by his unusual interest in modern technology and applied sciences, and in his preoccupation with the industrial working classes, existing and prospective. Within a short time he was contributing to the *Rheinische Zeitung*, most notably on the social science of political economy, as yet little appreciated in Germany, and on attempts to discern its place within Hegel’s overall system. Thus Engels first met the earnest Dr Marx and the communist Moses Hess on a brief visit to the editorial collective at Cologne in 1842, whilst *en route* to a further posting with the family manufactures at Manchester.

In England Engels wrote in both German and English on ‘the social question’ and the politics of class conflict, giving particular attention to socialism or communism (the terms were largely indistinguishable then). This was conceived as a system of cooperative ownership of productive resources and an egalitarian, non-monetary system of distribution applicable to society at large, combined with institutions for democratic decision-making. Engels’s *Umrisse zu einer Kritik der Nationalökonomie* (*Outlines of a Critique of Political Economy*) argued that a developing rationality in history could be discerned in both economic theory and practice. Applying a Hegelian dialectic of successive negations to the concepts ‘free trade’ and ‘competition’, he derived the necessity of worsening economic crises as his result, and the tenets of communism and the politics of class struggle as a resolution. Moreover in *Die Lage der arbeitenden Klasse in England* (*The Condition of the Working Class in England*) he produced an impassioned documentary and eyewitness survey of proletarian poverty and exploitation in modern industry.

On his return to Germany in 1844 Engels continued his career as a political agitator and socialist pamphleteer, this time in particular association with Marx, and the two produced three collaborative works: *Die heilige Familie* (*The Holy Family*) (by Engels and Marx, though with separately signed chapters), the manuscript *Die deutsche Ideologie* (*The German Ideology*) (posthumously published as the work of Marx and Engels), and their joint, though unsigned, masterpiece *Manifest der kommunistischen Partei* (*Manifesto of the Communist Party*).

During the revolutionary events of 1848-9 Engels assisted Marx, who had assumed editorship of the *Neue Rheinische Zeitung* in Cologne. Although Engels took advantage of the need to flee political repression from resurgent monarchists by making a walking tour through Burgundy in the autumn of 1848, he also saw active service in the spring of 1849 as liberal insurgents retreated in a more or less orderly way to the Swiss border. Making his way to England, Engels settled gradually but firmly into a life away from Germany, from partisan politics in either country, and from direct engagement with economic theory or social conditions, other than through a role, largely self-assumed, as publicist, popularizer and eventually editor for Marx.

3 Scientific socialism

Philosophically, Engels’s distinctive contributions were the terms ‘materialist conception of history’ and the ‘laws of dialectics’, which were assigned to Marx but explicated in various reviews, letters, introductions and texts. The

assignment of these terms to Marx is itself questionable, and the explications by Engels are neither as full nor as tightly written as his reputation suggests.

Marx formulated and employed an ‘outlook’, occasionally self-identified as ‘materialist’ (in opposition to both idealism and ‘traditional’ materialism), and used terms such as ‘science’, ‘law’ and ‘dialectic’ in his works (though much more rarely than most commentaries suggest). But it was Engels who insisted that Marx’s presuppositions included a matter-consciousness dichotomy, that his works were part of a system of Hegelian proportions, that his politically engaged social science was inclusive of the natural sciences of the day (including Darwinian biology as well as physical chemistry), and that a methodology derived from Hegel’s presumed ‘dialectic’ was crucial for Marx and for understanding his achievements.

In Engels’s commentaries, Marx’s ‘materialist conception of history’ included both the quoted view that ‘the mode of production of material life conditions the social, political and intellectual life-process in general’, and the gloss that ‘political action and its results originated in material causes’. Thus Marx’s ‘guiding thread’ was said to be the ‘great law of motion in history’, but in Engels’s version it emerged with a further layer of philosophical ambiguity. Engels’s qualifications - that ‘the production and reproduction of real life’ is the ‘ultimately’, but not the only, ‘determining element’ in history - could never allay the additional difficulties that he imposed on what were already problematic, and arguably for Marx, quite marginal formulations.

The ‘laws of dialectics’ assigned to Marx were formulated by Engels as reworkings of Hegel’s insights into a modern ‘science of interconnections, in contrast to metaphysics’. He suggested three: ‘transformation of quantity into quality and vice versa’, ‘interpenetration of opposites’ and ‘negation of the negation’. He described these as causal and invariable laws of motion forcing their way through the innumerable changes that can be observed in natural phenomena, human history, and ‘thought’ or logic, and maintained that they were validated by deduction. The circularity of such observations and the tautologous character of such laws were not apparent to Engels, and his more extravagant claims to link those ideas to a Marxian ‘system’ were not in fact published until after Marx’s death (Carver 1983: 129-41).

Engels’s ‘scientific socialism’ grounded the intellectual wing of the socialist and communist movement firmly in German philosophy, which he presumed had culminated in Hegel. The overarching project of his life was to link the world of progressive politics with the truths that he believed could be discerned from a philosophical system spanning the material, social and conceptual worlds, conceived developmentally. Among those truths, he alleged, was Marx’s concept of ‘surplus value’. This and other truths identified by Marx were said by Engels to be critical elements in a revolutionary political strategy. But Engels was not himself expert at the practicalities of getting from philosophical ideas to political action, nor was his philosophy capable of generating a concept of human agency that was philosophically defensible or politically useful.

See also: Hegelianism §3; Marxism, Western §1; Marxist philosophy of science; Marxist philosophy, Russian and Soviet; Socialism

TERRELL CARVER

List of works

Engels, F. (1844) *Umrisse zu einer Kritik der Nationalökonomie*, trans. M. Milligan, *Outlines of a Critique of Political Economy*, in K. Marx and F. Engels, *Collected Works*, London: Lawrence & Wishart, 1975, vol. 3, 418-43.(Pioneering attempt to subject political economy to philosophical analysis.)

Engels, F. (1845) *Die Lage der arbeitenden Klasse in England*, trans. F. Kelly-Wischnewestzky, *The Condition of the Working Class in England*, in K. Marx and F. Engels, *Collected Works*, London: Lawrence & Wishart, 1975, vol. 4, 295-583.(A survey, still very readable, of the horrors of early nineteenth-century industrialization.)

Engels, F. (1878-9) *Herrn Eugen Dührings Umwälzung der Wissenschaft*, trans. E. Burns, *Anti-Dühring*, in K. Marx and F. Engels, *Collected Works*, London: Lawrence & Wishart, 1987, vol. 25, 5-309.(Polemical work incorporating Engels’s influential interpretation and defence of Marx.)

Engels, F. (1880) *Socialisme utopique et socialisme scientifique*, trans. *Socialism: Utopian and Scientific*, in K. Marx and F. Engels, *Selected Works in One Volume*, London: Lawrence & Wishart, 1968, 375-428.(Three chapters of *Anti-Dühring* revised, published as a pamphlet and very widely translated and circulated.)

Engels, F. (1884) *Der Ursprung der Familie, des Privateigentums und des Staats*, trans. *The Origin of the Family,*

Private Property and the State, in K. Marx and F. Engels, *Selected Works in One Volume*, London: Lawrence & Wishart, 1968, 449-583. (Engels's attempt to account for the origin of women's oppression and the evolution of private property in prehistory by using a social Darwinian theory of sexual selection as well as gendered explanations of economic development.)

Engels, F. (1886) *Ludwig Feuerbach und der Ausgang der klassischen deutschen Philosophie*, trans. *Ludwig Feuerbach and the End of Classical German Philosophy*, in K. Marx and F. Engels, *Selected Works in One Volume*, London: Lawrence & Wishart, 1968, 584-622.(First published as a review, this is a history of German philosophy, placing Marx's materialism at its culmination and identifying it as an inversion of Hegelian idealism and arguing its coincidence with natural science.)

Engels, F. (1925) *Dialektik der Natur*, trans. C. Dutt, *The Dialectics of Nature*, in K. Marx and F. Engels, *Collected Works*, London: Lawrence & Wishart, 1987, vol. 25, 311-588.(Notes comprising a philosophical analysis of the subject matter and methods of nineteenth-century natural science, including the 'laws of dialectics'.)

Engels, F. and Marx, K. (1845) *Die heilige Familie*, trans. R. Dixon and C. Dutt, *The Holy Family*, in K. Marx and F. Engels, *Collected Works*, London: Lawrence & Wishart, 1975, vol. 4, 5-209.(An attack on romantic, utopian and philosophical socialists of the 1840s, but incorporating interesting early formulations of Marx's 'outlook'.)

Engels, F. and Marx, K. (1932) *Die deutsche Ideologie*, trans. W. Lough, *The German Ideology*, in K. Marx and F. Engels, *Collected Works* vol. 5, London: Lawrence & Wishart, 1976, 19-581.(Foundational manuscript, still not accessible in a properly edited version, nor adequately interpreted yet as an integrated work, but essential reading as an important exposition of Marx's 'outlook' in Part One.)

Engels, F. and Marx, K. (1848) *Manifest der kommunistischen Partei*, trans. S. Moore, *Manifesto of the Communist Party*, in K. Marx and F. Engels, *Collected Works*, London: Lawrence & Wishart, 1976, vol. 6, 477-519.(Still the best introduction to Marx's 'outlook', particularly the centrality of class struggle, and innocent of Engels's later interpretive framework, save in the footnotes.)

References and further reading

Arthur, C. J. (ed.) (1996) *Engels Today: A Centenary Appreciation*, London: Macmillan.(Collection of original articles covering aspects of Engel's work that are currrently of interest.)

Carver, T. (1983) *Marx and Engels: The Intellectual Relationship*, Brighton: Harvester Wheatsheaf.(A detailed textual study of similarities and differences between the works of the two men).

Carver, T. (1989) *Friedrich Engels: His Life and Thought*, Basingstoke: Macmillan. (Intellectual biography focusing on Engels's career before he met Marx, and including a detailed guide to further reading.)

Henderson, W.O. (1976) *The Life of Friedrich Engels*, London: Frank Cass, 2 vols. (Valuable for factual detail.)

Mayer, G. (1934) *Friedrich Engels*, The Hague: Martinus Nijhoff, 2 vols; repr. Cologne: Kiepenheuer & Witsch, 1971.(Inspirational but out of date and somewhat uncritical 'classic' biography in German.)

Rubel, M. (1981) 'The "Marx Legend", or Engels, Founder of Marxism', in *Rubel on Marx*, ed. Joseph O'Malley and Keith Algozin, Cambridge and New York: Cambridge University Press, 15-25.(Courageous paper dating from 1970 questioning Engels's own account of his role *vis-à-vis* Marx, and thus casting doubt on all 'standard' accounts of Marxism.)

Stedman Jones, G. (1973) 'Engels and the End of Classical German Philosophy', *New Left Review* 79: 17-36. (Brief but promising inquiry into Engels's appropriation of idealist philosophy.)

Stedman Jones, G. (1977) 'Engels and the Genesis of Marxism', *New Left Review* 106: 79-104. (Brief but detailed study of Engels's framework for interpreting Marx.)

Steger, M and Carver, T. (eds) (1997) *Engels after Marx*, College Park, PA: Pennsylvania University Press. (Collection of original articles covering 'identity', 'philosophy', 'politics' and 'legacy'.)

Engineering and ethics

Engineering ethics is that form of applied or professional ethics concerned with the conduct of engineers. Though engineers do many different things, they share a common history, which includes codes of ethics. Most codes explicitly declare public health, safety and welfare to be 'paramount'. Many questions of engineering ethics concern interpretation of 'public', 'safety' and 'paramount'. Engineers also have important obligations to client and employer, including confidentiality, proper response to conflict of interest, stewardship of resources, and honesty (not only avoiding false statements but volunteering certain information). Each engineer also has obligations to other engineers and to the profession as a whole.

1 The profession of engineering

'Engineering' as an honorific term suggesting both precision and success is often used where it does not belong. The pyramids are sometimes described as 'works of engineering', though engineers had nothing to do with them. 'Social engineering', 're-engineering' and even 'genetic engineering' - though technologies or applied sciences - are no more engineering than architecture or medicine is.

What, then, is 'engineering'? Unfortunately, there is no wholly satisfactory answer. It is, of course, what engineers do, but engineers do many things. They design machinery, chemical plants, harbours, bridges, office buildings, electrical grids, and other complex systems, manage their construction, oversee their operation, and plan their disposal. Though other occupations may do such things too (for example, architects also design buildings), engineers differ from these others in how they do it. Engineering is a certain way of doing such things. This way of doing things has a history; it is that history, rather than any abstract idea, that defines engineering.

The English word 'engineer' comes from French. The first people to be called 'engineers' were soldiers associated with 'engines of war' (such as cannons and siege towers). They were engineers only in the sense that they operated (or otherwise worked with) an 'engine' (that is, a complex device for some useful purpose). In 1676, the French organized these 'engineers' into a special unit, the *corps du génie*. Within two decades, the *corps* was known all over Europe for unusual achievements in military construction. When another country borrowed the French word 'engineering' for use in its own army, it was for the sort of activity the *corps du génie* engaged in.

At first, the *corps du génie* was more like an organization of masters and apprentices than a modern profession. Only during the 1700s did the French slowly come to understand what they wanted in an *officieur du génie* and how to get it by formal education. By the end of the 1700s, they had a curriculum from which today's engineering curriculum differs only in detail; they had also invented engineering as an occupation distinguished from that of other builders by its knowledge of modern physics, chemistry and mathematics, its skill in organization, and its concern with utility rather than beauty, and from the sciences by its focus on making rather than knowing. Civilian engineering is a branch of this (originally) military tree.

So far, this is a history of an occupation. The history of a profession tells how people engaged in a certain occupation organized to hold each other to standards beyond what law, market, and morality would otherwise demand. The history of a profession is the history of organizations, standards of competence, and standards of conduct. For engineering, that history tells of a slow shift from granting membership based on connection with large construction projects, practical invention, or other technological achievements to granting it based on two more demanding requirements. One requirement - a certain sort of knowledge - is occupational. This requirement is now typically identified with a degree in engineering. The other requirement - a commitment to use that knowledge in certain ways (that is, according to engineering's code of ethics) - is professional (see Professional ethics).

2 Ethics and engineering

'Ethics' (in this context) can be used in three senses: (a) as a synonym for ordinary morality; (b) as the name of a philosophical study attempting to understand morality as a rational undertaking; or (c) as the name of certain special standards of conduct governing members of a group in virtue of their membership (see Morality and ethics §§1-2).

In the first sense, engineering ethics is the application of ordinary moral rules, principles, or ideals to circumstances involving engineering. Neither philosophers nor codes of professional ethics have more than an educational or heuristic role. In the second sense, engineering ethics consists of attempts to offer a reasoned understanding (a theory) of how ordinary morality should guide engineering. In this sense, engineering ethics is just a subdivision of moral theory ('philosophical ethics'). But in the third sense, engineering ethics resembles law. Just as a law (or legal system) applies only to certain moral agents, those within its jurisdiction, so engineering ethics would apply only to certain moral agents, engineers. And just as law includes the interpretation, application, and justification of particular laws, so engineering ethics would include the interpretation, application and justification of engineering's special standards. The philosophical contribution to engineering ethics (in this sense) resembles legal philosophy's contribution to law, more a sorting of concepts and arguments than an application of moral theory.

Formal codes of engineering ethics did not appear in England before the mid-1800s, in the United States before 1900, or in Germany before the 1950s. While many professions, especially law and medicine, make a commitment to the profession's code of ethics a formal requirement for admission, engineering has not, except for licensed Professional Engineers. Instead, the expectation of commitment reveals itself when an engineer is found to have violated the code of ethics. The defence, 'I'm an engineer but I didn't promise to follow the code and therefore did nothing wrong', is never accepted. The profession answers, 'You committed yourself to the code when you claimed to be an engineer.'

While engineering ethics (in our third sense) dates from the 1800s, its philosophical study (ethics in our second sense) dates only from the 1970s. The field is still taking shape, working out its relation with moral theory, philosophy of technology, and philosophy of the professions.

Engineering ethics does not yet have a settled place in the college curriculum. In philosophy departments, classes in engineering ethics are often called 'Moral Problems in Engineering'; in engineering departments, 'Issues in Professionalism'. A course in engineering ethics, whether taught in a philosophy or an engineering department, will cover much about the history and organization of engineering.

3 Topics in engineering ethics

One topic of engineering ethics is the status of the code of ethics itself. (There are, in fact, several codes, though differences are relatively small.) Is a code just a statement of what morality would require of engineers, code or no code, or does it demand more of engineers than morality does? If it demands more, on what basis, if any, can it claim obedience from engineers? Is the claim legal or moral? If moral, does it rest on an implicit promise or on some other moral consideration (see Promising)?

Almost any code of engineering ethics provides a good checklist of major topics of engineering ethics. We may divide the other topics into four large categories. The first category concerns obligations to the public. Though the first codes of ethics emphasized personal honour and loyalty to client or employer, today most emphasize the public health, safety and welfare. Indeed, many codes make that consideration 'paramount'. Among questions dealt with under this obligation are: Who are the public? (Citizens of other countries or just fellow citizens? Employees of one's own company or just 'ordinary people'?) How safe is 'safe' (see Risk)? What is the public welfare? Does treating the public health, safety or welfare as paramount mean that an engineer must sometimes blow the whistle on his employer?

The second category concerns obligations to employer or client. Engineers are supposed to be faithful agents or trustees of their employer or client. Not only should they protect trade secrets, maintain confidentiality, avoid (or, at least, reveal) conflicts of interest, make clear the limits of their expertise, and try to use resources efficiently, they should also make sure that an employer or client understands the full implications of any decision in which an engineer is involved. Where an employer or client is a large corporation, there are special problems of determining *who* must understand for the employer or client to understand.

The third category concerns obligations to other engineers. Not only should engineers not discriminate against one another or compete unfairly; they have an obligation to provide engineers under their supervision with opportunities for professional development.

The fourth category, obligations to the profession, is a miscellany of questions. Engineers have an obligation to increase the competence and prestige of the engineering profession. Are they then sometimes obliged to reveal their employer's trade secrets to the profession? Engineers have an obligation to make public statements on engineering matters only in an objective and truthful manner. How forceful can an engineer's manner become without ceasing to be objective? Do engineers have an obligation to participate in engineering societies? What is the relation of this obligation to that of faithful agent and trustee of the employer or client?

See also: Responsibilities of scientists and intellectuals; Technology and ethics

MICHAEL DAVIS

References and further reading

Board of Ethical Review, National Society of Professional Engineers (1965-93) *Opinions of the Board of Ethical Review*, Arlington, VA: NSPE Publications, National Society of Professional Engineers, vols I-II. (Several hundred ethics cases, indexed by subject and with (official) discussion, applying NSPE Code.)

Davis, M. (1991) 'Thinking Like an Engineer: The Place of a Code of Ethics in the Practice of a Profession', *Philosophy & Public Affairs* 20 (2): 150-67.(Defends role of codes in engineering ethics, using the case of the 'Challenger' explosion (January 1986); example of what is being done.)

Elbaz, S.W. (1990) *Professional Ethics and Engineering: A Resource Guide*, Arlington, VA: National Institute of Engineering Ethics.(Designed for someone thinking of teaching engineering ethics; includes useful bibliography.)

Harris, C.E., Pritchard, M.S., and Rabins, M.J. (1995) *Engineering Ethics: Concepts and Cases,* Belmont, CA: Wadsworth Publishing.(Code-based approach to field; good bibliography.)

Johnson, D.G. (1991) *Ethical Issues in Engineering*, Englewood Cliffs, NJ: Prentice Hall.(Standard collection of readings, useful for introduction to issues.)

Ladenson, R.F., Choromokos, J., d'Anjou, E., Pimsler, M., and Rosen, H. (1980) *A Selected Bibliography of Professional Ethics and Social Responsibility in Engineering*, Chicago, IL: Center for the Study of Ethics in the Professions at Illinois Institute of Technology.(Old, but extensive and wide-ranging bibliography.)

Martin, M.W. and Schinzinger, R. (1983) *Ethics in Engineering*, New York: McGraw-Hill, revised 2nd edn, 1989.(Treats engineering as social experimentation.)

Schlossberger, E. (1993) *The Ethical Engineer*, Philadelphia, PA: Temple University Press.(Emphasizes morality, rather than codes, in analyzing issues in engineering ethics.)

Enlightenment, Continental

The Enlightenment is frequently portrayed as a campaign on behalf of freedom and reason as against dogmatic faith and its sectarian and barbarous consequences in the history of Western civilization. Many commentators who subscribe to this view find the Enlightenment's cosmopolitan opposition to priestly theology to be dangerously intolerant itself, too committed to uniform ideals of individual self-reliance without regard to community or diversity, or to recasting human nature in the light of science. Modern debates about the nature of the Enlightenment have their roots in eighteenth-century controversies about the arts and sciences and about ideas of progress and reason and the political consequences of promoting them. Even when they shared common objectives, eighteenth-century philosophers were seldom in agreement on substantive issues in epistemology or politics. If they were united at all, it was by virtue only of their collective scepticism in rejecting the universalist pretensions of uncritical theology and in expressing humanitarian revulsion at crimes committed in the name of sacred truth.

1 Modern assessments

The eighteenth-century Enlightenment is the only period of modern history described, most often by its detractors but sometimes also by its admirers, as a single intellectual movement or campaign, orchestrated around a common set of themes. Its leading thinkers, comprising Europe's first intelligentsia, are perceived as jointly committed to liberating mankind from the tyrannies of dead dogma and blind faith. They held that the barbarism of European culture in its Dark Ages had been due to ignorance and bigotry, and that the Crusades, the Inquisition and sectarian wars had been responsible for retarding the moral development of individuals and fragmenting their political communities. Unlike earlier philosophers who had sought to interpret the world, the advocates of Enlightenment were determined to change it. Their frequent engagement in public policy or status as civil servants and advisors to princes, their esteem for science over tradition, their empiricism and apparent preference for applied philosophy over metaphysics, have made them seem wedded to a belief in the unity of theory and practice. These same features of their doctrines or professions have in turn inspired the opposition of critics who, like Edmund Burke among their contemporaries, judged Enlightenment philosophy responsible for the French Revolutionary overthrow of the *ancien régime*, or who, like Alasdair MacIntyre today, blame it for inaugurating the monolithic philosophical trappings of modernity. Such charges against a so-called 'Enlightenment Project' have figured centrally in hostile interpretations of its significance over the past 200 years. On the one hand, both its conservative and liberal critics, such as Hegel at the turn of the nineteenth century or Jacob Talmon in the mid-twentieth century, have condemned its universalist philosophy of the rights of man, which they see as having been unsheathed in the Jacobin Terror and subsequently in the totalitarian regimes of our own day, whose vast schemes of social engineering are said to be inspired by its supposition that human nature may be recast like pliant clay. On the other hand, communitarian and postmodernist opponents of the Enlightenment, such as Charles Taylor in the first instance and Michel Foucault in the second, have equally decried its moral atomism and superficial understanding of the self, or alternatively its espousal of uniformitarian doctrines which leave scant room for individual diversity or cultural pluralism (see Community and communitarianism; Postmodernism). Most hostile of all are those commentators who have followed the contention of Max Horkheimer and Theodor Adorno that the Enlightenment's purely instrumental grasp of scientific rationality paved the way to the organization of genocide in the Second World War.

Much of the criticism of the Enlightenment Project has been concentrated upon two of its most seminal figures - Adam Smith and Immanuel Kant - the first because his liberal philosophy appears to remove all notions of absolute value from interpersonal relations, the second because his account of moral duty and autonomy seems to disregard the complex motives which actually inspire human action. In his attachment to universal rules legislated in the light of abstract reason; in his insistence (in modern parlance) upon preferring the right over the good; in his impartiality and essentially cosmopolitan indifference to difference, Kant in particular has come to be regarded as the principal spokesman for the vacuous subjectivism and empty formalism of the whole Enlightenment Project. The philosophical hermeneutics of Hans-Georg Gadamer takes up a fundamentally Hegelian position in criticizing Kant's, and therefore the Enlightenment's, abstract notion of individual self-awareness on account of its deracination from real experience. But Jürgen Habermas, in reaffirming the critical stance of Kant's philosophy, equally upholds the emancipatory aims of that same Enlightenment in its subjection of traditional beliefs to the

scrutiny of reason, which cannot but undermine tradition's authority.

Among modern social philosophers, Habermas is particularly well-disposed to what he takes to be mainstream Enlightenment thought, but the *Oxford English Dictionary* suggests an anti-Enlightenment bias within the English language itself, in identifying much the same perspective which Habermas applauds as 'superficial intellectualism' marked by 'insufficient respect for authority and tradition', adding that a *philosophe* - in the French language associated above all with the Enlightenment - is 'one who philosophizes erroneously'. Among eighteenth-century figures, only Hume and Kant, and occasionally Berkeley or Condillac, have been acknowledged by historians of philosophy as genuinely first-rate thinkers. The contributions of Hume and Kant to modern empiricism and rationalism, respectively, have nevertheless been considerable. When MacIntyre takes issue with Alfred Ayer and Richard Hare, he challenges contemporary Oxford exponents of two perspectives of an Enlightenment Project which he claims was bound to fail. The liberal principles of John Rawls' influential *Theory of Justice* (1971) also spring above all from the moral philosophy of Kant (see Rawls, J.).

2 Appraisals in the eighteenth century

Debates of the nineteenth and twentieth century about the Enlightenment's significance in European intellectual history take up themes already pursued in the eighteenth century by both its protagonists and critics, particularly in France in the 1750s in connection with the *Encyclopédie* of Diderot and D'Alembert (1751-72) and in Prussia in the 1780s and 1790s with regard to the limits of reason and the political implications of promoting its use. More than any other work of the eighteenth century the *Encyclopédie* may be regarded as the centrepiece of the Enlightenment, representing the self-image of an age of reason and invention much like the Crystal Palace would exhibit the character of an age of science and industry a century later. Assembled by contributors who formed an international republic of letters or party of humanity, the *Encyclopédie* was conceived as a reasoned dictionary of the arts and sciences, a modern monument to a classical ideal, inherited above all from Plato, that the promotion of knowledge and the pursuit of virtue should go hand in hand (see Encyclopedists, eighteenth-century).

In his preliminary discourse published in its first volume in 1751, D'Alembert, with Diderot the *Encyclopédie*'s joint overall editor, delivered what has sometimes been described as the manifesto of the Enlightenment as a whole. Taking stock of the achievements of the human mind through works of individual genius, D'Alembert lavished praise upon his precursors, including Bacon, Descartes and Leibniz among the philosophical and scientific giants of the seventeenth century. But no previous age had made such great progress along the path of knowledge as his own, he argued, in part because men of genius had turned their attention to the liberal and mechanical arts with the same enthusiasm as others had embraced pure science, in part because the human mind's productions had come to form a veritable library of useful information for a world in which prejudice could be overcome through reading. The *Encyclopédie* was an agent of general enlightenment, he thought, because it was accessible to a literate audience, transmitted in a vulgar rather than arcane language.

Some of these notions were borrowed from Fontenelle and other progressively-minded contributors to the early eighteenth-century *Querelle des anciens et des modernes*, and Diderot was to develop them further in his own article, 'encyclopédie', which appeared in volume five, of comparable length to D'Alembert's discourse, and equally designed to pay tribute to human industry and scientific inventiveness, which pushed back the frontiers of knowledge and thereby roused mankind from the dogmatic slumbers of its fundamentalist faiths. Such principles articulated one of the most central doctrines of French *philosophes* in particular, who like their patron saint, Voltaire, warmed to the achievements of English science, civility and toleration, which they characteristically took to be more robust guides to enlightenment than the predominantly classical, more archaic, forms of French culture and taste.

They were to figure as well as in the celebrations of the progress of the human mind produced by Turgot around the same time as D'Alembert's discourse, and above all, in the course of the French Revolution by Condorcet, D'Alembert's successor as permanent secretary of the French Académie des Sciences. Condorcet's sketch of the nine historical epochs through which civilization has passed, and of the tenth epoch to come, is particularly notable for its anticlericalism, its devotion to science, its faith in the power of reason to bring about societal change, and its optimistic tone adopted at the darkest moment of the French Revolution. For the past 200 years it has been regarded as the most characteristic expression of the idea of progress in Enlightenment thought, turning D'Alembert's focus upon books and treatises to more political ends and wider social objectives.

Just before D'Alembert's preliminary discourse was published, another work appeared, however, which he cites because it pursues an altogether different line of argument about the connection between virtue and knowledge. This was Rousseau's *Discourse on the Arts and Sciences* (1750), in which he contends that the trappings of civilization are but garlands of flowers around the iron chains by which mankind is weighed down (see Rousseau, J.-J.). The philosopher of nature as opposed to culture thus launched the first of his many attacks upon his own age of enlightenment. Although neither Nietzsche nor Foucault would ever recognize a debt to Rousseau in their own conjunctions of *savoir* with *pouvoir*, their critiques of the Enlightenment followed a path already mapped by him, which challenged the main thrust of the *Encyclopédie* at the moment of its birth.

The German debate on the meaning of enlightenment was launched more than thirty years later, by way of an article by the theologian Johann Friedrich Zöllner which appeared in the *Berlinische Monatsschrift* in December 1783. To the question he put, 'What is enlightenment?', numerous replies were submitted by correspondents in the course of the next several years, some published in other journals. The most important of these were by Moses Mendelssohn, who was one of Rousseau's German translators, and by Immanuel Kant. Mendelssohn sought to meet the objections to enlightenment posed by those most anxious to protect human society from overwrought change, and he argued that the discretion of each individual was a better censor of dangerous ideas than the full force of the law. To keep the inventiveness of enterprising persons in check on account of the hazards they raised was to legislate against human progress, which was a matter not for man but for Providence, he maintained, admitting that there were nevertheless certain truths useful to individuals but harmful to citizens.

In his short essay 'Was ist Aufklärung?', which appeared in the *Berlinische Monatsschrift* in December 1784, Kant introduced a similar distinction between the individual and the citizen, claiming that enlightenment was more difficult to achieve in the private than in the public domain, because of each person's ingrained immaturity or lack of courage to rely upon their own understanding. It was in the public sphere, however, rather than in private that it was most imperative for enlightenment to be gained, he insisted, and what was needed above all for its triumph was freedom. The century of Frederick the Great was not yet an enlightened age, but it was, Kant remarked, an age of enlightenment in which the obstacles to mankind's self-reliance and self-guidance were being gradually lifted. Over a decade later, he was to draw a still more optimistic portrait of the disinterested sympathy which would be felt by all mankind as it sought to live under republican constitutions which could have no interest in wars. In his essay *The Contest of the Faculties* (1798) he argued, in prophetic voice, that the human race was continually improving and that even apparent reversals like the Jacobin Terror would not arrest its further moral progress.

3 Its diversity and contrasts

Such differences of emphasis betwen French and German asessments of the Enlightenment lend warrant to the proposition, most often put forward by historians against philosophers, that any notion of a coherent and collective 'Enlightenment Project' is little more than a myth. French *philosophes*, Scottish moralists and German pietists, even when they managed to agree amongst themselves, were not committed to the same cause, it is suggested - French thinkers in particular proving characteristically more radical than their counterparts elsewhere, not least because, for the most part writers without university appointments or official duties as state servants, they were more free to speak their minds. The undeniable truth of that contention points towards greater specificity in identifying Enlightenment philosophies without obscuring the genuinely collective enterprise in which their authors were often engaged.

Thanks to a variety of moral weeklies, literary periodicals and eventually newspapers, and to an international book trade, both authorized and clandestine, philosophical essays, discourses, treatises and encyclopedias circulated on a scale unprecedented in human history and at a speed which enabled Scottish writings indebted to Montesquieu and German commentaries on those same works to cross-fertilize one another in appearing within a matter of months, as if electronically linked to the Internet today. Political thinkers of the Enlightenment served as ministers in the courts of Maria Theresa, Joseph II, Leopold II, George III, Gustav III, Frederik VI, and Louis XVI, while kings corresponded with philosophers, sought their company and solicited their advice. Literacy made the spread of enlightenment possible, granting to experts in all subjects a political presence and power in society they had not earlier enjoyed. This command of publicity by the friends of the Enlightenment makes the notion of their collective project credible, even allowing for the diverse and sometimes conflicting objectives of its adherents.

It is nevertheless imperative that due weight be placed on the substantive differences between those objectives, however much they were pursued in concert. Condillac, the most influential of the French empiricists of the Enlightenment, came to revise his own views substantially in the course of his intellectual career, attempting at first to correct Locke's epistemology by way of refining the doctrine of the association of ideas so as to explain the generation of self-awareness without recourse to intuition, while later, in his account of the statue-man, putting forward a more physiological notion of psychic life as rooted in the senses, particularly the sense of touch. The French materialists, in turn - principally Maupertuis, La Mettrie and d'Holbach - put forward a number of claims about the physical determination of morality which were similar to those of the later Condillac, although their works were received differently and came to be regarded, especially by the Church, as the most incendiary of the philosophies of the Enlightenment on account of their denial of the existence of an immaterial and immortal soul (see La Mettrie, J.O. de).

When political thinkers of the Enlightenment agreed that their aim should be to devise a set of rules for the proper management of human affairs in the light of the known tendencies of human nature, they were sometimes drawn to utilitarianism in the manner of Hutcheson, Helvétius, Beccaria or Bentham, who advocated policies which promoted the greatest happiness or least pain of the greatest number, by way of constitutional reforms which abolished physical torture and all civil retribution for moral crimes (see Utilitarianism). But from much the same premises other philosophers advocated instead sumptuary laws in order to promote economic equality and the collective ownership of property such as were proposed by Dom Deschamps, Morelly and Mably, while still others adopted the wholly contrary stance of Turgot, Antonio Genovesi and Giuseppe Palmieri, who perceived the social benefits of economic inequality and held that the state could best promote economic development by exercising as little control as possible over the liberties and initiative of its citizens.

In criticizing monolithic notions of a science of the human mind, perceived as subject to the laws of nature in a universe which it reflects, critics of the Enlightenment Project often overlook the richness and diversity of its various themes. No major thinker of the eighteenth century was more persuaded of physicalist explanations of social behaviour and culture than Montesquieu, who plainly fits the description offered by Richard Rorty, for instance, of a philosopher convinced that objective knowledge of all things, including human conduct, is possible and that mind is a mirror of nature. Yet from the single dimension of his belief in the scientific study of both matter and mind, there springs no universalism or cosmopolitanism of any kind, since above virtually all his contempories Montesquieu was specially sensitive to local variety, specificity and the uniqueness of social institutions, customs and mores.

Enlightenment philosophers - empiricists and rationalists together - are perhaps best described as sceptics who collectively framed fundamentally liberal objections to what they took to be the universalist bigotry of sacred truth. If their principal battlecry was Voltaire's 'Écrasez l'infâme', they were, on account of their commitment to toleration, drawn by an ideal of peaceable assembly such as Voltaire himself in his *Philosophical Letters* (1734) had found among the traders of the London Stock Exchange, where no church was supreme and individuals of different faiths could deal with one another as if they were of the same religion (see Voltaire). In opposing the ethnic cleansers of their day, the philosophers of the Enlightenment contributed to our understanding of what we now call crimes against humanity.

See also: Clandestine literature; Enlightenment, Jewish; Enlightenment, Russian; Enlightenment, Scottish

ROBERT WOKLER

References and further reading

Cassirer, E. (1932) *Philosophie der Aufklärung*, Boston, MA: Beacon Press; trans. A. Koelin and J. Pettegrove, *The Philosophy of the Enlightenment*, Princeton, NJ: Princeton University Press, 1951.(Originally published in German in 1932, a high-minded account of critical philosophy in the eighteenth century, addressing its systematic spirit as distinct from the spirit of system of seventeenth-century metaphysics.)

Darnton, R. (1982) *The Literary Underground of the Old Regime*, Cambridge, MA: Harvard University Press.(An account of Grub Street hacks, publishers and peddlers of forbidden books in eighteenth-century France, who pursued wordly success more often than philosophical truth.)

Fox, C., Porter, R. and Wokler, R. (eds) (1995) *Inventing Human Science: Eighteenth-Century Domains*, Berkeley, CA: University of California Press.(Commentaries on the Enlightenment sciences of natural history,

medicine, anthropology, psychology and politics.)

Gay, P. (1966, 1969) *The Enlightenment: An Interpretation*, vol. 1, *The Rise of Modern Paganism*, vol. 2, *The Science of Freedom*, New York: Knopf.(Assesses the Enlightenment as an intellectual movement of diverse political reformers committed to secularism, humanity, cosmopolitanism and freedom.)

Hampson, N. (1968) *The Enlightenment*, Harmondsworth: Penguin.(A general study of eighteenth-century thinkers within their particular backgrounds.)

Havens, G.R. (1955) *The Age of Ideas: From Reaction to Revolution in Eighteenth-Century France*, New York: Holt.(A detailed study of leading thinkers of the French Enlightenment from Bayle and Fontenelle to Diderot and Beaumarchais.)

Hazard, P. (1954) *European Thought in the Eighteenth Century*, London: Hollis & Carter.(Originally published in French in 1946, a comprehensive account of Christianity, unbelief and educational and moral doctrines in the Enlightenment's republic of letters.)

Manuel, F.E. (1959) *The Eighteenth Century Confronts the Gods*, Cambridge, MA: Harvard University Press.(A study of the place of pagan mythology, polytheism and deism in the Enlightenment, and of eighteenth-century critiques of religious fanaticism.)

Outram, D. (1995) *The Enlightenment*, Cambridge: Cambridge University Press.(A general treatment, addressing recent commentaries on the social production and marketing of ideas, with chapters on science, religion, exoticism, gender and government.)

Porter, R. and Mikuláš, T. (eds) (1981) *The Enlightenment in National Context*, Cambridge: Cambridge University Press.(Assesses literary and professional elites of the eighteenth century in their diverse social and political environments.)

Reill, P.H. (1975) *The German Enlightenment and the Rise of Historicism*, Berkeley, CA: University of California Press.(Interprets the German Aufklärung as an intellectual movement shaped by Leibnizian philosophy, the Ständestaat tradition and the contending claims of Pietism and rationalism.)

Schmidt, J. (ed.) (1996) *What is Enlightenment? Eighteenth-Century Answers and Twentieth-Century Questions*, Berkeley, CA: University of California Press.(Reassembles eighteenth-century German interpretations of and twentieth-century commentaries on the meaning of enlightenment, with a notable introduction.)

Scott, H.M. (ed.) (1990) *Enlightened Absolutism: Reform and Reformersin Later Eighteenth-Century Europe*, London: Macmillan.(Addresses the importance of the particular national and provincial environments in which philosophical kingship was pursued.)

Venturi, F. (1972) *Italy and the Enlightenment: Studies in a Cosmopolitan Century*, London: Longman.(Learned essays on Italian contributors to the Enlightenment theory and practice of cosmopolitanism, patriotism and socialism, from Radicati and Paoli to Filangieri and Pagano.)

Wade, I.O. (1938) *The Clandestine Organization and Diffusion of Philosophic Ideas in France from 1700 to 1750*, Princeton, NJ: Princeton University Press.(A study of the circulation of philosophical and theological manuscripts which prefigured the predominant philosophical currents of the period 1750-89.)

Yolton, J.W., Porter, R., Rogers, P. and Stafford, B.M. (eds) (1991) *The Blackwell Companion to the Enlightenment*, London: Blackwell.(An illustrated encyclopedic treatment of political and literary doctrines and their authors, and of the arts and sciences in the eighteenth century.)

Enlightenment, Jewish

The eighteenth century in Europe saw the beginnings of Jewish emancipation, and this led to an intellectual development which came to be known as the Jewish Enlightenment or Haskalah. This movement emphasized the rational individual, the notion of natural law, natural religion and toleration, and natural rights. The effect of this form of thought was to provide a justification for the equality of the Jews with other citizens of national entities. The most important exponent of this movement was Moses Mendelssohn, who dominated the debate on the role Jews should play in the state and the rationality of Judaism as a religion. Ultimately the Jewish Enlightenment moved east and became connected with such movements as Zionism. In Germany it led to the development of the Reform movement. The Jewish Enlightenment very much set the agenda for the next two centuries of debate about Jewish ideas by seeking to analyse the links between religion and reason in Judaism.

1 Haskalah and philosophy

Haskalah (Jewish Enlightenment) is the name given to a series of movements at times resembling and often modelled on the various Enlightenment movements in western and central Europe. The Haskalah first took root in the major Prussian cities of Berlin and Königsberg in the latter half of the eighteenth century. It spread to other German cities and to Vienna; and from there, throughout the nineteenth century, to Bohemia, Galicia, Podolia, Lithuania and the southern Ukraine, especially the city of Odessa. Although each area of Haskalah had its own unique features, the *maskilim* (the adherents of Haskalah) of Europe were all committed to the social and intellectual revitalization of the Jewish communities of Europe. Their programme included the learning of European languages and developing familiarity with European culture, including philosophy, as well as a transformation of the Jewish economic profile and educational system. While *maskilim* were often few in number the Haskalah had an important impact on Jewish culture in central and eastern Europe in the nineteenth century.

The Haskalah in Prussia produced one major philosopher: Moses Mendelssohn (1729-86). Mendelssohn became one of the most important philosophers of the German *Aufklärer*, even while he remained central to the development of Jewish thought and ideological modernization. In his *Briefe über die Empfindungen (Letters on Feeling)* (1755b) and *Philosophische Schriften (Philosophical Writings)* (1761) he redirected aesthetic discussion by insisting that judgments about beauty are independent of ethical purpose or logical criteria. In 1763 he published *Abhandlung über die Evidenz in metaphysischen Wissenschaften (Treatise on Evidence in Metaphysical Knowledge)* in which he tried to shield metaphysics from the claim that its proofs for the existence of God and morality do not attain the certainty of mathematics. His reputation as a major German philosopher was assured with the publication of his Platonic dialogue *Phädon, oder über die Unsterblichkeit der Seele (Phaedo, or on the Immortality of the Soul)* in 1767. The dialogue was a philosophical best-seller and adumbrated the quintessential Enlightenment statement of belief in the immortality of the soul. Even after the publication of Kant's critical works, Mendelssohn remained a staunch defender of the demonstrability of the existence of God, as is shown in his last work, *Morgenstunden, oder Vorlesungen über das Dasein Gottes (Morning Hours, or Lectures on the Existence of God)* (1786), a work intended, among other things, to defend Mendelssohn's friend Lessing from charges of atheism and Spinozism.

Even as he took full and active part in the European Enlightenment, Mendelssohn remained committed to the development of Jewish culture and thought, as can be seen by his numerous Hebrew works, especially his commentary on Exodus, and his book *Jerusalem, oder über religiöse Macht und Judentum (Jerusalem, or on Religious Power and Judaism)* (1783). In this work, Mendelssohn applied theories of natural right to the political status of the Jews and to European societies generally. He insisted on a clear delineation of the prerogatives of Church and state in (what ought to be) a secular society. In the second half of the book he developed a philosophy of Judaism that defended his ancestral faith against charges of immorally excluding non-Jews from salvation. He insisted that Judaism embodied a revealed legislation, rather than a revealed religion and argued that Judaism does not coerce its adherents in matters of religion any more than it excludes non-Jews from salvation.

Mendelssohn used his Bible commentary to disseminate some of his ideas regarding metaphysics and politics to a larger Jewish audience. In addition, we find in his Bible commentary a distinct contribution to the emerging field of philosophical linguistics, although this side of Mendelssohn's work remained largely unknown beyond the

Jewish community, as it was written in Hebrew. Here Mendelssohn tried to develop the notion that language was a vehicle for mediating more than one intended significance. The search for authorial intention, then, although a standard concern at the time, could not proceed simply on a philological and literal level. Texts needed to be interpreted more broadly to appreciate fully the intended significance of their words, phrases and sentences. In practice, for Mendelssohn, this meant the submission of the interpreter to the ancient interpretative traditions. The recognition of the polyvalency of meanings in a text might have led Mendelssohn to a more historically rooted view of texts and their appropriate interpretations. But he tended to rest content with the traditional subtleties of the rabbinic midrashic method (see Midrash).

Mendelssohn was the only important Jewish philosopher in German-speaking lands who remained committed to Haskalah throughout his life. But mention should be made of two other members of the Berlin Jewish community who were associated with the Haskalah movement for a time. Salomon Maimon (1753/4-1800), a Polish Jew, came to Berlin in quest of opportunities for the free pursuit of knowledge. He was associated for a time with the *maskilim* of Berlin and wrote a Hebrew work designed to reconcile scientific principles with the study of the Torah. He translated Mendelssohn’s *Morgenstunden* into Hebrew and wrote a philosophical commentary, also in Hebrew, on Maimonides’ *Guide to the Perplexed* (see Maimonides, M.). Eventually, Maimon rejected the notion that Jewish culture could be reconciled with modern philosophy and devoted the rest of his career to writing several philosophical works critical of Kant’s transcendental philosophy.

The stimulus of Kantian thought was to lead yet another, albeit far less original, philosophical voice away from Haskalah, that of Marcus Herz (1747-1803). Herz was briefly associated with the Haskalah movement, but went on to become one of Kant’s earliest disciples - he was chosen by Kant to be a respondent in the defence of his inaugural dissertation. Herz thanks Kant for helping to liberate him from a life of ‘dragging a burden of prejudices’, a reference to his Jewish cultural background. His *Betrachtungen aus der spekulativen Weltweisheit (Meditations Drawn from Speculative Wisdom)*, written in 1771, represents an early commentary on Kant’s inaugural dissertation, which, like the work of so many commentators and popularizers, shows distinct misunderstanding of the Kant’s work. Ultimately, Herz made a name for himself as an important scientist and physician. As with Maimon, the Haskalah made possible his encounter with European philosophy. But unlike Maimon, he did little to advance the cause of Jewish enlightenment.

The various other centres of Haskalah produced only one other important philosopher, Nachman Krochmal of Galicia (1785-1840). Unlike the Prussians, Krochmal advanced philosophical discussion within the Jewish community alone. Although the leading intellectual of the Haskalah in Galicia, far from the German ‘intellectual marketplace’ as he put it, he became an astute observer of the German philosophical environment. Krochmal provided one of the first Jewish responses to, and adaptations of, elements of the philosophical work of Spinoza, Kant, Herder, Schelling and Hegel. In his posthumously published *Moreh nevukhei ha-Zeman (Guide to the Perplexed of Today)* (1851), Krochmal adapted Kantian epistemological concepts to demonstrate the need to interpret religious sources with an eye to their deeper philosophical significance. He then proceeded to show that the deeper philosophical significance of traditional Jewish sources, both ancient and medieval, corresponded to the most recent discoveries of the German Idealist philosophers. Using his interpretative method, Krochmal argued that a range of traditional Jewish concepts - such as belief in a personal God who created the world and revealed a desired way of life - remained philosophically coherent, for, when properly interpreted, they were shown to be concrete representations of the abstract metaphysics of Hegel and Schelling.

Krochmal responded to the regnant philosophy of history that considered Jewish culture - among many others - to be sublated in the onward march of history, by arguing that Jewish religion, because of its apprehension of the absolute, stood outside the historical ‘laws’ that mandated the eventual cultural demise of all nations and states. Although he was an important model for aspiring Jewish intellectuals in eastern Europe, his philosophical work was of limited influence. His lasting contribution would seem to lie in his implicit exposure of the unstated and culturally biased assumptions of modern idealist philosophy. His work showed that different cultural assumptions would lead to the application of idealist concepts in a way that yielded philosophical and religious conclusions that diverged significantly from those reached by the German Idealists themselves.

2 The Haskalah and the history of philosophy

The Haskalah, not unlike the Enlightenment movements of western and central Europe, produced few original

philosophical minds (see Enlightenment, Continental). But there were many lesser lights and still more followers of their guidance in the quest for an enlightened way of life and thought. For this reason at least, the Haskalah remains an important, although largely unmined, resource for the student of the history of philosophy. The hundreds of *maskilim* throughout Europe were avid consumers of French and German philosophy. They were also eager readers of Spinoza's works as well (see Spinoza, B. de). Knowing what they read and how they understood it is vital to an understanding of the spread and impact of the modernizing trends of European thought. The earliest Prussian *maskilim* emerge at a time when Christian Wolff was one of the leading philosophers of Germany. In the works of these early *maskilim*, including Mendelssohn, one can discern Wolff's distinct influence. Similarly, the ethical writings of *maskilim* in Germany display familiarity with Enlightenment ethics. One result is a turning away from the pessimistic tone of many traditional Jewish ethical writings. The influences that we observe show that the German *maskilim* were the most acculturated, and most at home, linguistically and culturally, in the world of European philosophy.

Yet the Haskalah moved eastward just as European philosophy was undergoing a revolution reflected in the works of the post-Enlightenment Idealists, and the works of Herder as well. The derivative thinking of the Eastern *maskilim* remained rooted in Enlightenment thought even as it embraced the new idealism. This often difficult combination produced fascinating tensions, facilitated by the fact that, with the exception of Krochmal, Jewish readers in central and eastern Europe tended to read popularizations of Kant rather than Kant himself. What this meant in practice was that the eastern European Jewish writers and thinkers of the Haskalah often developed 'Kantian' ideas in ways that were quite consistent with central Enlightenment ideals - with assumptions and consequences that would have been unrecognizable to the master. Similarly, Jewish writers embraced certain Rousseauian notions regarding education and ethics (see Rousseau, J.-J.). Again, though, they tended to receive this material second-hand. Especially important as a popularizer of Kant and Rousseau was the Swiss writer, thinker and historian Heinrich Zschokke (1771-1848). The dependence on Zschokke for knowledge of Kant and Rousseau is particularly manifest among the *maskilim* of Lithuania.

In examining the *maskilim* as consumers of European thought, we find several matters especially worthy of note to the historian of philosophy. The first is that post-Kantian *maskilim* tended to find Kant's ethical thought thoroughly compelling, despite its overtly anti-Judaic component. Yet these *maskilim* separated Kant's ethics from his critical metaphysics. They had little use for the latter, and tended to reduce even Kant's ethics to the principles of duty and autonomy. They then went about showing how such ethics were compatible with the biblical *mitzvot*, despite the fact that these divine commandments might readily seem thoroughly heteronomous, and certainly seemed so to Kant. Here Kant's ethical thinking was removed from its place within his thought and enlisted on behalf of a very different, modernizing cum apologetic cultural project.

Post-Hegelian *maskilim* tended to have a very limited understanding of Hegel. Here again Krochmal is the exception. Yet thinkers who had no apparent understanding of Hegel's metaphysics, ethics or politics were drawn to his philosophy of history. Once again we find these thinkers drawing selectively from Hegel's thought. Even when some paid rhetorical lip service to the dialectical movement of history, what interested the *maskilim* most was the teleological view of history that Hegel's thought seemed to legitimate. This view in turn allowed the *maskilim* to conceptualize observed changes in Jewish history as part of a larger drama that they could interpret as pointing towards the fulfilment of their own cultural aspirations. The continued influence of Enlightenment ideals is manifest here.

Eclectic borrowing from Spinoza may be discerned in the work of some of the more religiously radical *maskilim*. They wanted to re-enact the drama of Spinoza's liberation from the Jewish religious tradition. Yet they sought to do so without adopting Spinoza's universalism.

3 The afterlife of the Haskalah

Throughout the encounter with post-Enlightenment European thought in the various centres of Haskalah, the movement in some ways remained very much modelled on the European Enlightenment. Everywhere and always committed to Enlightenment rationalism, the *maskilim* remained thoroughly optimistic regarding the ability of the Jewish community to generate a rational culture even as they expressed deep contempt for the Jewish masses, and often for its rabbinic elite. But European thought and politics moved on, and for some the efforts of the *maskilim* to graft more recent thinking onto the tree of rationalism and optimism came to appear hopelessly naïve. Many

others, like Maimon and Herz, came to see the effort to reconcile European thought - however well or poorly understood - with Judaism, traditional or reformed, as irrelevant at best, dishonest at worst. Thus, in each of its centres, the Haskalah, like the Enlightenment itself, tended to be of limited duration. Yet in its own limited way it helped bring about fundamental intellectual change in the Jewish communities of Europe. It promoted and legitimized new areas of inquiry among Jews and contributed mightily to the emergence of new philosophical and political ideologies as Jews entered the twentieth century.

See also: Enlightenment, Continental; Jewish philosophy in the early 19th century; Mendelssohn, M.

JAY M. HARRIS

References and further reading

Altmann, A. (1973) *Moses Mendelssohn: A Biographical Study*, Philadelphia, PA: Jewish Publications Society. (Magisterial study of all aspects of his life, including his personal and intellectual relationships to many of the leading philosophical voices of the day.)

Arkush, A. (1994) *Moses Mendelssohn and the Enlightenment*, Albany, NY: State University of New York Press. (An effort to situate Mendelssohn within the context of the thought of Leibniz, Wolff, Kant and Lessing, based almost exclusively on Mendelssohn's German writings.)

Dubin, L. (1997) 'The Social and Cultural Context: Eighteenth-Century Enlightenment', in D. Frank and O. Leaman (eds) *History of Jewish Philosophy*, London: Routledge.(A comprehensive guide to the period, including a detailed bibliography.)

Harris, J. (1991) *Nachman Krochmal: Guiding the Perplexed of the Modern Age*, New York: New York University Press.(An intellectual biography which treats his complex relationship with the major schools of German Idealism, as well as the thought of Herder and Spinoza.)

Krochmal, N. (1851) *Moreh Nevukhei ha-Zeman (Guide to the Perplexed of Today)*, ed. with Hebrew introduction by S. Rawidowicz, Berlin: Einot, 1924; repr. Waltham, MA: Ararat, 1961.(Adapts Kantian epistemological methods to interpret the Jewish religious sources with an eye to discovering their inner, philosophical meaning. Krochmal argues that at their deeper level these sources, both ancient and medieval, anticipated the discoveries of the German Idealist philosophers.)

Mendelssohn, M. (1755a) *Philosophische Gespräche (Philosophical Speeches)*, Berlin.(In this work, Mendelssohn declares himself a disciple of the school of Leibniz and takes sides with Spinoza.)

Mendelssohn, M. (1755b) *Briefe über die Empfindungen (Letters on Feeling)*, Berlin.(Contains a philosophy of the beatiful and forms a chief basis of all philosophical-aesthetic criticism in Germany.)

Mendelssohn, M. (1763) *Abhandlung über die Evidenz in metaphysischen Wissenschaften (Treatise on Evidence in Metaphysical Knowledge)*, Berlin: Mande und Spener, 1786.(Essay comparing the demonstrability of metaphysical propositions with that of mathematical ones.)

Mendelssohn, M. (1767) *Phädon, oder über die Unsterblichkeit der Seele (Phaedo, or on the Immortality of the Soul)*, Wien: C.F. Schade, 1812.(His chief philosophical work, dealing with the immortality of the soul, following Plato's dialogue of the same name.)

Mendelssohn, M. (1783) *Jerusalem, oder über religiöse Macht und Judentum (Jerusalem, or on Religious Power and Judaism)*, trans. with notes by A. Arkush, *Jerusalem*, Hanover, NH: University Press of New England, 1983.(His epoch-making work calling for the emancipation of the Jews and setting out the conditions of religious freedom in a pluralistic state.)

Mendelssohn, M. (1785) *Morgenstunden, oder Vorlesungen über das Dasein Gottes (Morning Hours, or Lectures on the Existence of God)*, Berlin: C. Voss, 1786.(Sets forth Mendelssohn's fundamental metaphysical beliefs and argues against pantheism.)

Rotenstreich, N. (1984) *Jews and German Philosophy*, New York: Schocken.(A useful and thorough consideration of the relationship between Jewish and German thought in the eighteenth to twentieth centuries.)

Sorkin, D. (1996) *Moses Mendelssohn and the Religious Enlightenment*, Berkeley, CA: University of California Press.(Complements and challenges Arkush by relying on both the Hebrew and Jewish writings of Mendelssohn within the context of theological Wolffianism.)

Enlightenment, Russian

When Russia embraced secular European ways of thought under Peter the Great, its educated elite came into contact first with the German Enlightenment, which combined the rationalism of Descartes and Leibniz with the emotionalism of Protestant pietism. With its acknowledgement of established authority, and emphasis on a person's responsibility to the community rather than individual rights, this strand of early Enlightenment thought suited the state-building of Peter.

By the second half of the eighteenth century, the universality of the French language, the influence of geniuses such as Montesquieu and Voltaire, and the formation of a conscious Enlightenment party among the French philosophes *meant that their polemical writings carried the mainstream of progressive thought. It was this Enlightenment that was embraced by Catherine the Great, who professed its tenets: tolerance and the conviction that perfecting social organization would ensure human happiness. She encouraged the growth of an intellectual elite to spread the ideas of the* philosophes *and form an enlightened public opinion.*

Russia's tradition of absolutism and the institution of serfdom, however, proved inimical to Enlightenment values. Nevertheless, the Russian Enlightenment engendered an elite of individuals eager to act as critics and moral leaders of their society who would determine the future course of Russian social thought.

1 State Enlightenment

The ground was prepared for the Russian Enlightenment when Peter the Great reshaped Muscovy, with its Orthodox religious culture, into a modern, secular European state. In order to achieve his reforms, the Russian service elite were obliged to immerse themselves in the contemporary currents of European thought, science, social progress and literature. Some were sent abroad for training, but most learned from the technicians and educators imported from the West to modernize the state. The majority came from Germany and brought with them the philosophical outlook of their homeland. This outlook, forged in the West's recent scientific revolution, also underpinned the modern technology imported by these specialists. Consequently, Petrine Russia sought advice on education from Leibniz and Christian Wolff, and it was their pupils who were recruited for the new Academy of Sciences: G.F. Müller, Schlözer and Stählin, the historians; Euler, the mathematician; and Pallas, the biologist, geographer, linguist and explorer - these were the leading scholars and scientists in eighteenth-century Russia.

The intellectual framework which these teachers brought with them was constructed from the rationalism of Descartes and the natural-law ideas of Leibniz and Pufendorf. The rationalism and legalism were imbued, however, with the emotionalism of Protestant pietism, whose proponent was Christian Wolff. Both the pietist emphasis on the acceptance of properly constituted authority, whether religious or political, and its stress on a person's obligation to serve the community rather than on selfish individual rights, suited the requirements for state-building initiated by Peter. The influence of this particular German strand of Enlightenment was reinforced by the prestige of the eighteenth-century polymath M.V. Lomonosov (1711-65) whose work and writings reflected the scientific education he had received in the universities of Marburg and Freiburg, with their strong pietist tradition. A constant promoter of the 'common good', he was instrumental in the founding in 1755 of Moscow University, where the employment of a number of influential German professors ensured the persistence of the German outlook.

Despite the influx of Western patterns of thought the essence of Enlightenment - and accumulation of knowledge permitting the critical appraisal of social relationships in order to improve human life - was not assimilated by Russia in the first half of the eighteenth century. It was in fact negated when Peter exploited the means of Enlightenment as an instrument to transform society from above in order to augment the state's military, economic and administrative efficiency. Nevertheless, some of the prerequisites for Enlightenment, such as the founding of scientific, educational and publishing institutions, had been established. And crucially, a secular conceptual framework had been created outside traditional Orthodoxy.

2 Enlightenment of society

The mainstream of the European Enlightenment, represented by the works of the French *philosophes*, penetrated

into Russia from the 1740s, but the progress and development of the Enlightenment in Russia is intimately connected with the person and the reign (1762-96) of Catherine the Great, who declared her allegiance to the ideals and aims of the Enlightenment in her memoirs. In 1767 she published her *Nakaz*, or the *Instruction* provided to guide the deliberations of the deputies summoned to a Legislative Commission. Although the rational precepts of Bielfeld and Justi and the pietist morality of Christian Wolff are present in the document as reminders of the German influence, the *Instruction* was mainly composed of liberal ideas taken from the *Encyclopédie*, from Beccaria's *Dei delitti e delle pene* (*On Crimes and Punishments*) (1764) and above all from Montesquieu, whose *De l'esprit des lois* (*The Spirit of the Laws*) (1748) was called by Catherine her 'prayer book'. The *Instruction* demonstrated the sovereign's respect for Enlightenment thought and gave it imperial sanction.

Another Enlightenment initiative of Catherine's was the 1766 essay competition on the peasant question organized by the Free Economic Society, Russia's first independent cultural institution, and inspired by the economic arguments of the *Encyclopédie* and the physiocrats. Few Russians competed, but among the 164 entries were essays by Voltaire and Marmontel.

The Free Economic Society competition and the *Instruction* linked Russia with the wider European Enlightenment in two ways. First, it provided Europe with another example of an 'enlightened despot' in the person of Catherine. It was important in the eyes of the *philosophes* that sovereigns themselves should understand the lack of justice in the existing social order. In Catherine the Great the French *philosophes* discovered an empress prepared to play the part of an 'enlightened monarch', eager to buy their books, to correspond with them, to invite them to visit her, to give them patronage and even to carry out the reformation of society they proposed. Proof of this was her declaration of admiration for Montesquieu in a letter to d'Alembert where she explained that she had plagiarized him 'for the good of twenty million people, which must come of it'.

Second, the consultative assembly called to consider the *Instruction* was an example of the involvement of broader sections of society in the Enlightenment. It supplied a precondition for the spread of the Enlightenment: the establishment of public forums for debate and publicity, permitting individuals to participate in the process of social reform. One of the striking propositions in the *Instruction* (article 58) was the need 'to prepare the Minds of People' before the implementation of improved legislation.

European Enlightenment was no longer exploited merely as an instrument to strengthen the efficiency of the state; the welfare of society and that of its members was now the objective. 'The happiness of one and all' was the motto on the medal given to the deputies summoned to respond to the *Instruction*, and one of its closing articles (521) stated that its propositions were 'the Choice of those Means, whereby the Russians may be rendered the most happy possible of Mankind'.

3 Means of Enlightenment

After the suspension of the Legislative Commission, as a result of the outbreak of the Turkish War, other channels were found for the diffusion of Enlightenment thought. In 1768 Catherine founded a 'Society for the Translation of Foreign Books' which by 1772 had published over forty titles, including extracts from the *Encyclopédie*, Montesquieu and Voltaire. Meanwhile she had called into being the publication of moral weeklies, modelled on the *Spectator* of Addison and Steele, that popularized Enlightenment attitudes. She opened state archives to aid private historiographical enterprises. Licensing requirements for presses and publishing houses were abolished in 1783. In Catherine's approval for private charitable initiatives in popular education, and particularly for Ivan Betskoi's plans for foundling homes that would produce a new type of citizen, support was given to the institutionalization of the social vision of the Enlightenment.

Although these means were all under imperial tutelage, there seems to have been a policy of encouraging a class of intellectuals who would guide society, correct its defects and help eradicate the obstacles frustrating the creation of an enlightened society. Eventually, with the emergence of a social elite that was not an official part of state authority, public opinion began to take on a corrective function and became a power factor with which the state had to contend.

In this a leading role was played by writers, aware of themselves as forming a 'republic of letters', as Mikhail Kheraskov, Director of Moscow University's Press and Library, called the group of young men who had gathered around him. No original contribution, however, was made by them to European Enlightenment philosophy. For

these Russian writers, Enlightenment meant initially the reception and later the adaptation of Western developments, and the participation of Russia in the general cultural development of Europe. Their own function was that of the mediator, spreading the ideas of the French *philosophes* of the previous generation. This insinuation of Enlightenment ideas into the consciousness of Europeanized Russians was as essential a part of the Enlightenment as the composition of the key works had been.

The most prominent director of this process was Nikolai Novikov (l714-1818), originally a minute-taker at the Legislative Commission where he would have noted the shaping of a rudimentary public opinion. Following the Commission's disbandment, Novikov appealed to that newly aware public as editor of a series of moral weeklies which castigated obscurantism and promoted the image of the truly enlightened citizen. The 1770s saw Novikov making a significant contribution to the development of publishing, and in 1779 he obtained the lease of Moscow University Press, which dominated Russian publishing for the next decade. Simultaneously he created the Typographical Company, a significant independent publishing concern.

Novikov was also a leading figure in Russian Freemasonry, one of the principal channels for the dissemination of Enlightenment values. In the 1770s the outlook of the Russian Freemason was broadly that of the approved *philosophes* and for Russians at that time the terms 'Voltairean' and 'Freemasonic' were synonymous. But later, Russian Freemasonry, under the influence of pietism, shunned the radical trends of the *philosophes* and nurtured a moderate Enlightenment which refused to become anti-Christian.

The fusion of Enlightenment values with a heightened moral awareness and religious and spiritual sympathies could prove a powerful amalgam. This was shown by another significant figure of the Russian Enlightenment, Aleksandr Radishchev (1748-1802). Radishchev studied at Leipzig, a centre of pietism, where he was also influenced by the later radical generation of *philosophe* writers: Mably, Raynal, Helvetius and Rousseau, who argued that there was a need to change institutions in order for humanity to be able to realize its true potential. This combination of spiritual enthusiasm, moral commitment and a desire for institutional change, particularly for the abolition of serfdom, was crystallized in Radishchev's *Puteshestvie iz Peterburga v Moskvu* (*Journey from St Petersburg to Moscow*) (1790), the most radical statement of the Russian Enlightenment.

4 Consequences of Russian Enlightenment

For publishing his *Journey from St Petersburg to Moscow* Radishchev received a death sentence, commuted to ten years of exile. Novikov was also sentenced in 1792 to fifteen years of imprisonment for his involvement with Freemasonry. This repression signalled the end of Catherine's enthusiasm for the Enlightenment, which had been blighted by the Pugachev rebellion and the French Revolution. At the end of her reign it seemed the Enlightenment had left little trace on Russia's official institutions. However, a cultural elite among the nobility had emerged as a self-conscious civil society, steeped in Enlightenment values and yearning to be free from the shackles of the state. The Enlightenment had taught them to question Russia's traditional political structure, particularly autocracy and serfdom. The Enlightenment belief in the perfectibility of society, the pursuit of happiness, and the prospect of a rational code of laws persisted. It was a faith protested with fervour by men nurtured in the Masonic lodges which imparted a particular spiritual earnestness to the Russian Enlighteners. Here was the breeding ground for the Russian intelligentsia of the nineteenth century.

See also: Enlightenment, Continental; Leibniz, G.W.; Montesquieu; Natural law; Pietism; Voltaire; Wolff, C.; Liberalism, Russian §1

W. GARETH JONES

References and further reading

Beccaria, C.B. (1764) *Dei delitti e delle pene*, ed. F. Venturi, Turin: Einuadi, 1965; R. Bellamy (ed.), trans. R. Davies, V. Cox, R. Bellamy, *On Crimes and Punishments and Other Writings*, Cambridge: Cambridge University Press, 1995.

Billington, J.H. (1966) *The Icon and the Axe: An Interpretive History of Russian Culture*, London: Weidenfeld & Nicolson.(The chapter 'The Troubled Enlightenment' (213-68) examines the Enlightenment in the context of Russia's general cultural history.)

Catherine the Great (1767) *Nakaz (Instruction)*; P. Dukes (ed.) *Catherine the Great's Instruction (Nakaz) to the Legislative Commission, 1767*, Newtonville, MA: Oriental Research Partners, 1977.(A contemporary

eighteenth-century English translation of the *Nakaz* with a balanced introduction and bibliography.)

Dukes, P. (1981) 'The Russian Enlightenment', in R. Porter and M. Teich (eds), *The Enlightenment in National Context*, Cambridge: Cambridge University Press, 176-91.(Examines the Russian national variant of the European Enlightenment.)

Jones, W.G. (1984) *Nikolay Novikov: Enlightener of Russia*, Cambridge: Cambridge University Press.(Expansion of §3.)

McConnell, A. (1964) *A Russian Philosophe: Alexander Radishchev 1749-1802*, The Hague: Martinus Nijhoff. (Expansion of §3.)

Montesquieu, C. de S., Baron de (1748), *De l'esprit des lois*, ed. G. Truc, Paris: Garnier, 1949; trans. T. Nugent, *The Spirit of the Laws*, New York: Hafner, 1949.

Raeff, M. (1966) *The Origins of the Russian Intelligentsia: the Eighteenth-Century Nobility*, New York: Harcourt, Brace & World.(The contribution of the nobility to the Enlightenment; expansion of §4.)

Raeff, M. (1973) *'The Enlightenment in Russia and Russian Thought in the Enlightenment'*, in J.G. Garrard (ed.) *The Eighteenth Century in Russia*, Oxford: Clarendon Press, 25-47.(Particularly useful for its expansion of §1.)

Robel, G. (1992) 'Zur Aufklärung in Adelsgesellschaften: Rußland und Polen' (On the Enlightenment in Nobility Societies: Russia and Poland), in S. Juttner and J. Schlobach (eds) *Europäische Aufklärung(en): Einheit und nationale Vielfalt*, Hamburg: Felix Meiner Verlag, 152-71.(An excellent survey, the comparison with Poland being used to highlight the particular nature of the Russian experience.)

Enlightenment, Scottish

This term refers to the intellectual movement in Scotland in roughly the second half of the eighteenth century. As a movement it included many theorists - the best known of whom are David Hume, Adam Smith and Thomas Reid - who maintained both institutional and personal links with each other. It was not narrowly philosophical, although in the Common Sense School it did develop its own distinctive body of argument. Its most characteristic feature was the development of a wide-ranging social theory that included pioneering 'sociological' works by Adam Ferguson and John Millar, socio-cultural history by Henry Home (Lord Kames) and William Robertson as well as Hume's Essays (1777) and Smith's classic 'economics' text The Wealth of Nations (1776). All these works shared a commitment to 'scientific' causal explanation and sought, from the premise of the uniformity of human nature, to establish a history of social institutions in which the notion of a mode of subsistence played a key organising role. Typically of the Enlightenment as a whole this explanatory endeavour was not divorced from explicit evaluation. Though not uncritical of their own commercial society, the Scots were in no doubt as to the superiority of their own age compared to what had gone before.

1 Overview

The Scottish Enlightenment may be conveniently dated from the publication of Hume's *Treatise of Human Nature* (1739-40; see Hume, D.) to the revised sixth edition of Smith's *Theory of Moral Sentiments* (1790; see Smith, A.) It can be fairly described as a 'movement' since it was not confined to a handful of authors but encompassed upwards of fifty participants, who all either knew or were aware of each other. Unlike in France, the Enlightenment in Scotland had strong institutional and establishment roots being based in the universities and the various satellite groups and societies to which even non-academics like Hume and James Hutton or law-lords like Henry Home (Lord Kames) and James Burnett (Lord Monboddo) belonged. As a movement it was also multifaceted. Scientific theory and experiment were well represented. Some of Isaac Newton's earliest disciples were Scottish, including, notably, Colin Maclaurin. Important work was undertaken by William Cullen and Joseph Black in chemistry and at the end of the period Hutton laid the foundation for uniformitarian geology. Literary theory was also prominent with treatises on rhetoric by George Campbell (1776); and Hugh Blair (1783); disquisitions on taste by Alexander Gerard (1759), and many others, plus pioneering works of literary criticism by Thomas Blackwell (notably Blackwell 1735). In philosophy a number of Scots, led by Reid, developed their own Common Sense School, which was stimulated by Hume's perceived scepticism and was critical of the entire 'way of ideas' (see Common Sense School; Reid, T.).

2 Social theory

Perhaps the most characteristic aspect of the Scottish Enlightenment was its social theory and it is upon that aspect that this entry focuses, taking the term to include historical and moral theory.

The Scots took human sociality to be an evidentially warranted fact. A common explanation of it was that humans possessed a social instinct or appetite. This was seen as 'social' rather than simply familial, and Kames (1774), among others, undertook a detailed comparison between human and animal association. Despite some similarities, the distinctiveness of human sociality was insisted upon especially in the form of non-instrumental social ties such as friendship and loyalty. Part of the significance of this commitment to evidence was the deliberate rejection, most notably by Hume (see Hume, D. §§4-5), of the conceptions of a state of nature and a social contract.

Not only were these notions fanciful but they over-emphasized the role of purposive rationality. Here the Scots' views differed from those typical of the French Enlightenment and of English thinkers like Priestley. For these latter thinkers Lockean epistemology (see Locke, J.) had laid the foundations for an essentially optimistic 'perfectibilist' philosophy. Since to mould experience is to mould human character, then, informed by the findings of reason, it becomes possible to set humans irrevocably on the right track. Accordingly, the more rational society becomes, the more rational will be the experience that it passes on to the next generation. For the Scots this was simplistic. In practice, the scope of reason was circumscribed by habit and social convention; social norms, for example, were the product of socialization not rational insight. In the case of government this meant that although they all had their origin in violence, 'time by degrees… accustoms the nation to regard as their lawful or native princes, that family which at first they considered as usurpers' (Hume [1741-77] 1987: 474-5). Legitimacy in

practice was the work of sentiments over time rather than the correspondence (or not) to some rational principle, like consent.

This link between social experience and normative judgments was the linchpin of Smith’s *Theory of Moral Sentiments* (1757). To Smith the effects of social intercourse teach what behaviour is acceptable and in due course individuals internalise these social judgments as conscience, viewing their own actions and motives as an ’impartial well-informed spectator’ would (see Smith, A. §3). The assumptions in Smith’s moral philosophy were widely shared. Following the lead of Francis Hutcheson they sought on the one hand to dismiss ’rationalist’ accounts of morality (it was ’more properly felt than judg’d of’ (Hume 1739-40: III.1.2)) and, most especially, on the other to undermine the egocentric assimilation of ’morality’ to self-interest as in the work of Thomas Hobbes and Bernard Mandeville. Against the pretended naturalism of Hobbes and Mandeville, all the Scots insisted on a true naturalism, the facts of human nature testifying in their view to the presence of truly moral experience, a disinterested concern for the wellbeing of others. Hutcheson invoked a special ’moral sense’ (as did, though not identically, Kames (1751) to explain this concern, while both Hume and Smith (without being in total agreement) rejected this recourse to a direct sense, referring instead to a concurrence of sentiments by means of a principle of sympathy (see Hume, D. §3; Smith, A. §2).

This sensitivity to socialization also meant the Scots were sceptical of the supposed achievements of Great Legislators like Lycurgus in moulding social institutions. Instead the Scots drew attention to the accidental and gradual formation of these institutions; in a phrase of Adam Ferguson’s, co-opted by later critics of rationalism like, for example, Hayek (1960), ’nations stumble upon establishments which are indeed the result of human action, but not the execution of any human design’ (Ferguson 1767); see Ferguson, A.). This is an exemplification of the ’law of unintended consequences’, of which the other classic formulation in the Scottish Enlightenment was Smith’s ’invisible hand’, whereby pursuit of self-interest by individuals furthered the public interest. While this has a clear bearing on the operation of markets, it was also used more generally - by John Millar, for example (1787, volume 1), to account for the development of the jury system.

This diminution of the role of rationalistic individualism was part of the Scots’ endeavour to explain social institutions as the effects of correspondingly social causes. They sought to go beyond the mere cataloguing of facts and provision of indiscriminate historical narrative to the tracing of the chain of causes and effects (see the Preface to Kames (1758)). This was typically executed by writing ’natural histories’ (as in Hume (1757)). The key assumption was that human nature was constant and uniform in its principles and operations (see Human nature, science of in the 18th century). From that premise the theorists traced the development of key institutions, such as property, government, ranks, the status of women and religion, from simplicity to complexity or from rudeness to civilization. As a means of organizing this development, emphasis was laid, as Robertson said, on different ’modes of subsistence’ (1777: Book 4). With Smith’s account seemingly seminal, social development was seen to fall into four stages - hunting, herding, farming and exchanging (see Smith, A. §4). In outlining this development, evidence from contemporary ethnography and classical history was combined in a deliberate utilization of the comparative method. Such comparisons, as well as being a way of checking the validity of any particular report, enabled them both to sift out ’chance’ or random associations from truly causal conjunctions and to supply the missing links in the causal chain. This permitted them self-consciously to write universal histories. The title of James Dunbar’s *Essays on the History of Mankind in Rude and Cultivated Ages* (1780) captured what was typical of this general enterprise. The overall aim of these inquiries was to render the ’amazing diversity’ (as Millar termed it (1779: 3rd edition, Introduction) of social life explicable. While later philosophers like Collingwood (see Collingwood 1961) have heavily criticized this view of the past, for the Scots of the Enlightenment, history was (in Hume’s phrase) a branch of the ’science of man’. It was also openly evaluative.

3 Civilized society

The Scots regarded their own society as belonging to the fourth commercial stage. Given that ’industry, knowledge and humanity’ are indissolubly linked (see Hume (1752) 1741-77), and present in commercial societies, then, in contrast, earlier savage societies were indolent, ignorant and cruel. Commercial society was ’polished’; moral sentiments were refined; taste was delicate; religion had lost much of its superstitious overlay. These features fitted together to characterize not only their own ’civilized’ society but also their theoretical comprehension of it. In this way their social theory meshed with their moral, aesthetic and religious thought. A nation of savages would live in

small scattered groups, worship many gods, have a dull moral sense and a crude taste. But refinement in all these linked spheres develops as humans gradually succeed in triumphing over the dictatorship of needs.

What was seen especially to distinguish a civilized society was the operation of the rule of law, and this was reflected in the salience the Scots attached to justice. Hume, Smith and Kames, though differing in details, all agreed in regarding justice as indispensable to social life. What made this fourth stage 'commercial' was that within it everyone was 'in some measure a merchant' (Smith 1776: Book 1, Chapter 4). As Smith argued, the decisive ingredient of this stage was the extent of the division of labour, which, in its turn, depended on the extent of the market (see Smith, A. §4). Although bartering was a natural disposition it required the stability wrought by the rule of law to become decisively effective. Once the division of labour was established it greatly increased the material wellbeing of the inhabitants, but that it also had deleterious effects was widely recognised. Ferguson and Kames evaluated these using the long-standing vocabulary of 'corruption'. For Ferguson these effects were part of a more pervasive weakness of commercial societies. He was most especially concerned about the passivity - the preoccupation with 'private' affairs - that these societies engendered. Liberty, he believed, could not be secure if all it meant was obeying just laws; it also needed active involvement. It was to this end that he wished to see a revival of public spirit in the form of militias. Smith, like Hume, did believe that 'modern liberty' came from rule-following, since a stable framework, provided by the inflexible administration of justice, permitted individuals to pursue their own interests in their own way. He did nonetheless recognize, as an unintended consequence of the specialization attendant on the division of labour, that it affected adversely the labourer's virtues. His remedy was the provision, at the public's expense, of primary education. This mixture of analysis and judgment by the Scots in their treatment of their own society is characteristic of the Enlightenment in general and demonstrates that for all their undoubted distinctiveness there truly was a Scottish Enlightenment.

See also: Aberdeen Philosophical Society; Commensensism; Enlightenment, Continental; Moral sense theories; Naturalism in ethics; Naturalism in social science

CHRISTOPHER J. BERRY

References and further reading

Berry, C. (1997) *The Social Theory of the Scottish Enlightenment*, Edinburgh: Edinburgh University Press. (Provides both a detailed analysis and a synoptic overview, including a survey of interpretations.)

Blackwell, T. (1735) *An Enquiry into the Life and Writings of Homer*, repr. Menston: Scolar Press1972. (Investigates the reciprocal relationships between the poet, his poetry and his socio-cultural context.)

Blair, H. (1783) *Lectures on Rhetoric and Belles-Lettres*, ed. H.F. Harding, Carbondale, IL: Southern Illinois University Press, 2 vols, 1965.(A largely derivative but for that reason valuably indicative assembly of contemporary views on literary and aesthetic topics.)

Campbell, G. (1776) *The Philosophy of Rhetoric*, ed. L. Bitzer, Carbondale, IL and Edwardsville, IL: Southern Illinois University Press.(Investigates how language, widely understood, is linked to the faculties of human nature.)

Campbell, R. and Skinner, A. (eds) (1982) *The Origins and Nature of the Scottish Enlightenment*, Edinburgh: Donald.(A collection which deals both with the institutional setting, precursors and aspects of the Scots' thought).

Collingwood, R. (1961) *The Idea of History*, Oxford: Oxford University Press.(Outlines his idealist philosophy of history and criticizes alternative views.)

Dunbar, J. (1780) *Essays on the History of Mankind in Rude and Cultivated Ages*, Bristol: Thoemmes Press, 2nd edn, 1781.(Develops a distinctive view of psychological and social development.)

Ferguson, A. (1767) *An Essay on the History of Civil Society*, ed. D. Forbes, Edinburgh: Edinburgh University Press, 1966.(Frequently interpreted as a pioneering work of sociology, it is a wide-ranging moralistic assessment of commercial society in a broad historical sweep.)

Gerard, A. (1759) *An Essay on Taste*, facsimile repr. of 3rd edn (1780), Gainsville, FL: Scholars' Facsimiles & Reprints, 1963.(Applies associationist principles to literary criticism.)

Hayek, F. (1960) *The Constitution of Liberty*, London: Routledge & Kegan Paul.(Important statement of Hayek's philosophy where his debt to the Scots is acknowledged.)

Hont, I. and Ignatieff, M. (eds) (1983) *Wealth and Virtue: The Shaping of Political Economy in the Scottish Enlightenment*, Cambridge: Cambridge University Press.(A collection of articles focusing on the Scots' analyses of commerce.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge; revised P.H. Nidditch, Oxford: Clarendon Press, 1978.(One of philosophy's great books that Hume himself later dismissed in preference for his subsequent writing)

(1741-77) *Essays Moral Political and Literary*, ed. E.F. Miller, Indianapolis, IN: Liberty Press, revised edn. 1987. (Important collection, especially for Hume's economic and political thinking.)

(1757) *Natural History of Religion*, ed. W. Colver, Oxford: Clarendon Press, 1976.(Traces the development of religious belief from its roots in human nature)

Hutcheson, F. (1725) *Inquiry into the Original of our Ideas of Beauty and Virtue*, London and Dublin.(Contains his first formulation of the philosophy of moral sense.)

Kames, Lord (Home, H.) (1751) *Essays on the Principles of Morality and Natural Religion*, Edinburgh: Kincaid & Donaldson; revised London: Hitch & Hawes; Dodsley, Rivington, Fletcher & Richardson, 1758; Edinburgh: Bell & Murray, 1779.(Contains critiques of Hutcheson, Hume and Smith, late editions amended the necessitarianism of the first edition.)

(1758) *Historical Law Tracts*, Edinburgh: Kincaid & Bell; London: Millar, 2 vols; revised and enlarged, Edinburgh: Bell & Creech; London: Cadell, 1776, 1 vol.(Traces the evolution of legal thinking, important for its evocation of stages of social development.)

(1774) *Sketches on the History of Man*, Edinburgh: Creech; London: Strahan & Cadell; 3rd edn with Kames' last additions and corrections, Edinburgh: Creech; London: Strahan & Cadell, 1778.(A rambling comendium that deals idiosyncratically with most aspects of social life.)

Millar, J. (1779) *The Origin of the Distinction of Ranks*, 3rd. amended edn, repr. in W. Lehmann (ed.), *John Millar of Glasgow*, Cambridge: Cambridge University Press, 1960.(The *Ranks* is another work that has been identified as a significant early sociological treatise.)

(1787-1803) *Historical View of the English Government*; excerpts repr. in W. Lehmann (ed.), *John Millar of Glasgow*, Cambridge: Cambridge University Press, 1960.(The last 4 vols were published posthumously.)

Reid, T. (1764) *Inquiry into the Human Mind, On the Principles of Common Sense*, Edinburgh.(The seminal statement of Common Sense philosophy.)

Robertson, W. (1777) *The History of America*, ed. D. Stewart, in 1 vol., London: William Ball, 1740.(Book 4 contains an important discussion of Amerindian society.)

Sher, R. (1985) *Church and University in the Scottish Enlightenment*, Edinburgh: Edinburgh University Press. (Contains a lengthy bibliography.)

Smith, A. (1757) *The Theory of Moral Sentiments* ed. A. Macfie and D. Raphael, Indianapolis, IN: Liberty Press, 1982.(A major work that outlines a philosophical psychology of moral judgment.)

— (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, ed. R. Campbell and A. Skinner, Indianapolis, IN: Liberty Press, 1981.(Epoch-making work that laid the foundation for analytical economics and advocated free trade.)

Stewart, M.A. (ed.) (1990) *Studies in the Philosophy of the Scottish Enlightenment*, Oxford: Clarendon Press.(A mix of general and specific - especially on Hume - essays.)

Enthusiasm

For much of the seventeenth and eighteenth centuries, enthusiasm denotes a state of (claimed) divine inspiration. The claimed inspiration is almost always seen by those who employ the term as delusory, and enthusiasm is almost always seen as bad, akin to fanaticism, irrationality, and madness. The term is most commonly applied to Protestants outside the Church of England and, at times, to Catholics and pre-Christian mystics. Throughout the period, enthusiasm much less often denotes devotion and zeal or poetic inspiration. In the nineteenth century, concern with enthusiasm declines.

There are various debates about enthusiasm, understood as claimed divine inspiration. One concerns its causes. Thomas Hobbes (1651) suspected that much enthusiasm was hypocritical, a political ploy. Others linked enthusiasm with melancholy. Henry More (1662) argued that melancholy causes both overheated passions and an overheated imagination; enthusiasts suppose that anything extraordinary, such as the state of their passions, must be due to God, and so conclude that they are inspired. Later, Isaac Taylor (1823) drops the reference to melancholy but leaves an excessive imagination, trespassing on the terrain of reason, as the etiology. Of the cases where no explanation is obvious, Meric Casaubon (1656) argues that the cause is likely still natural; there are many extraordinary yet natural events. All note that attention and admiration from others discourages critical thought about one’s claims to inspiration.

More, Casaubon and Taylor proceed by cataloging varieties of enthusiasm, treating each as a natural and medical rather than supernatural and possibly heretical phenomenon. Although they see that their approach could, in other hands, furnish a thoroughly sceptical account of religious experience, they do not intend to rule out true divine inspiration. They merely quarrel with contemporary claims to it.

Another debate concerns why enthusiasm is bad. Enthusiasm can be criticized on epistemic grounds. Enthusiasts justify their claims, not by reason or faith, but by citing direct communication from God. John Locke (1700) asks why we should believe that such communication has occurred. This cannot be proven by citing the feeling that the claim is from God, nor by noting that we cannot explain why the claim is made, nor by citing our strength of conviction, nor by showing by reason that the claim is true. One needs reason to show that the claim comes from God. Locke thinks this is possible: Moses’s doubts were dispelled by the burning but unconsumed bush and the transformation of the rod into a serpent. But Locke supposes contemporary enthusiasts have no such assurance.

Enthusiasm can be criticized on religious grounds. It leads to schisms in the Church, and an opposition to church government, since mediators between oneself and God are unnecessary. It leads one to read, or even ignore, Scripture in light of one’s own fancy, rather than in light of scholarly interpretation. Jeremy Taylor (1854) worries that enthusiasts bypass the established ways to grace. More’s attack on enthusiasm is motivated by his desire to defend Christian belief and experience as rational. Enthusiasts, by abandoning reason for an ‘inner light,’ make this impossible.

Enthusiasm can also be criticized on moral and political grounds. Enthusiasm was often seen as a leading cause of the Anabaptist Peasants’ Revolt and the English civil war. On the standard view, enthusiasts foist their views on others with force and intolerance, since they are convinced that they act in God’s name. They are rarely up for reasoned discussion with others, particularly given the private nature of their inspiration. They ignore common morality for ‘higher’ concerns. And in opposing church government, they bring anarchy, since church government is vital to civil government.

Against this, David Hume (1741) gives an idiosyncratic analysis: religions based on enthusiasm have violent beginnings, given the presumptuous character of those prone to enthusiasm. But in time, these religions become calm and innocuous, since there is little church government and ritual to support their practices. In their pride and resistance to church government, Hume argues, enthusiasts become friends to civil liberty.

A third debate centres on the best way of combating enthusiasm. Some favour state coercion. The Third Earl of Shaftesbury (1711) replies that religious enthusiasm flourishes under oppression. He favours free public debate, featuring ridicule of enthusiasts. (Swift’s *A Tale of a Tub* (1704) is the most notable example.) To defend *oneself* against enthusiasm, Shaftesbury recommends self-knowledge sufficient to identify the true causes of any supposed

inspiration. For those infected, More recommends temperance, humility, and reason. Jeremy Taylor recommends prudence.

A fourth debate centres on distinguishing good and bad enthusiasm. More, Locke and Isaac Taylor suggest various tests for true divine inspiration. In addition to public phenomena such as the burning bush, the beliefs and actions which proceed from the inspiration must be moral, reasonable and consistent with Scripture. Worries about what happens should only some of these tests be fulfilled, like worries about Biblical instances of enthusiasm that do not meet these tests, are largely ignored. The emphasis is on exposing the bad, not identifying the good.

Here Shaftesbury is an exception. He argues in favour of enthusiasm, largely because he treats any divine inspiration not as a matter of supernatural prophesying, quaking, or speaking in tongues, but as a natural raising of the imagination in the presence of beauty and goodness. Enthusiasm, he argues, is needed for any extraordinary achievement. Without it, life is dull and brutish.

In the nineteenth century, enthusiasm comes to be usually understood as mere devotion and zeal, without implying divine inspiration as their cause. The term ‘enthusiasm’ largely ceases to name an object of philosophical debate, since whether devotion and zeal are good or bad depends largely on the ends to which they are put. Debate about divine inspiration continues in literature on the epistemology of religion. But in the nineteenth century the issue becomes less pressing - perhaps because the Evangelical movement brought enthusiasm into the Church of England, perhaps because enthusiasts were seen as less of a political threat, or perhaps because increased scepticism about defending Christianity on rational grounds made the contrast between rational and enthusiastic believer less sharp.

See also: Mysticism, nature of; Religion and epistemology; Religious experience

ROBERT SHAVER

References and further reading

Alston, W. P. (1991) *Perceiving God*, Ithaca, NY: Cornell University Press.(A sympathetic, analytic treatment of the epistemology of religious experience.)

Casaubon, M. (1656) *A Treatise Concerning Enthusiasme*, Gainesville, FL: Scholars’ Facsimiles and Reprints, 1970.(A naturalistic etiology and taxonomy. A shorter edition was published in 1655.)

Hobbes, T. (1651) *Leviathan*, ed. C.B. Macpherson, Harmondsworth: Penguin, 1968.(For Hobbes’s scepticism concerning divine possession, see especially chapters 8, 12, 32 and 34.)

Hume, D. (1741) ’Of Superstition and Enthusiasm’, in *Essays: Moral, Political, and Literary*, ed. E.F. Miller, Indianapolis, IN: Liberty, 1985.(A contrast between enthusiasm and superstition, with respect to etiology and consequences.The essay appeared in the first (1741) edition of Hume’s *Essays*.)

Knox, R.A. (1950) *Enthusiasm*, Oxford: Clarendon Press.(A detailed history of the subject, concentrating on the seventeenth and eighteenth centuries.)

Locke, J. (1700) *An Essay Concerning Human Understanding*, 2nd edn, ed. P.H. Nidditch, Oxford: Clarendon Press, 1979.(See ’Of Enthusiasm’, IV.xix. Epistemic tests for divine inspiration. Locke added this chapter to the second (1700) edition of the *Essay*, first published 1689.)

More, H. (1662) *Enthusiasmus Triumphatus*, Los Angeles, CA: Augustan Reprint Society, 1966.(A shorter early edition was published in 1656. Another naturalistic etiology and taxonomy. Snappier than Casaubon.)

Shaftesbury, A., 3rd Earl of (1711) *Characteristics of Men, Manners, Opinions, Times, etc.*, ed. J.M. Robertson, Glouster, MA: Peter Smith, 1963, 2 vols.(For defence of enthusiasm, understood as a quality necessary for any extraordinary achievement, see especially *A Letter Concerning Enthusiasm*, *The Moralists*, part 3, section 2, and *Miscellaneous Reflections* II, chapter I.)

Steffan, T.G. (1941) ’The Social Argument Against Enthusiasm (1650-1660)’, *Texas Studies in English* 21: 39-63. (An excellent summary.)

Swift, J. (1704) *A Tale of a Tub and other works*, ed. A. Ross and D. Woolley, Oxford: Oxford University Press, 1986.(The most famous instance of enthusiasm ridiculed.)

Taylor, I. (1823) *The Natural History of Enthusiasm*, London: Holdsworth.(The emphasis is on errors of religion rather than errors of rationality.)

Taylor, J. (1854) *The Whole Works of the Right Reverend Jeremy Taylor*, ed. R. Herber and C.P. Eden, London: Longman, Brown, Green, and Longmans, 10 vols.(Taylor wrote in the 1640s through the 1660s. His references to enthusiasm are scattered; Steffan (1941) gives a short account and references.)

Tucker, S.I. (1972) *Enthusiasm*, Cambridge: Cambridge University Press.(A clear history of the word ‘enthusiasm,’ with a fine bibliography.)

Environmental ethics

Theories of ethics try to answer the question, 'How ought we to live?'. An environmental ethic refers to our natural surroundings in giving the answer. It may claim that all natural things and systems are of value in their own right and worthy of moral respect. A weaker position is the biocentric one, arguing that living things merit moral consideration. An ethic which restricts the possession of moral value to human persons can still be environmental. Such a view may depict the existence of certain natural values as necessary for the flourishing of present and future generations of human beings. Moral respect for animals has been discussed since the time of the pre-Socratic philosophers, while the significance to our wellbeing of the natural environment has been pondered since the time of Kant and Rousseau. The relation of the natural to the built environment, and the importance of place, is a central feature of the philosophy of Heidegger. Under the impact of increasing species loss and land clearance, the work on environmental ethics since the 1970s has focused largely on one specific aspect of the environment - nature in the wild.

1 Respect for life and the 'natural'

Living things are characterized by a tendency to preserve and maintain themselves in the face of environmental challenge. Aristotelian and Stoic thinkers regarded such self-preservation as an 'end' or *telos* at which living things aim. A weaker conception of the endeavour to maintain life is found in Spinoza's conception of *conatus* (see Spinoza, B. de§§9-11). The term 'autopoeitic' has been coined more recently to refer to those processes which are self-producing and self-maintaining. Are living things - characterized as having *telos*, *conatus*, or being autopoeitic - worth respecting for themselves (see Moral standing §§1-2)? And would such respect be moral? Only in the present century has an affirmative answer to both questions been argued in detail (see Taylor and Mathews, in Brennan (ed.) 1995; O'Neill 1993).

This contrasts with moral concern for animals, which can be traced back to the pre-Socratic philosophers of ancient Greece (see Presocratic philosophy). The Renaissance essayist Montaigne referred to them as our 'brothers and companions', while Jaques, in Shakespeare's *As You Like It*, made a striking plea for the rights of deer to graze in the forest undisturbed by hunters. A peak of concern for animals was reached in the eighteenth century when there was great debate over whether orang-utans should be classified as a kind of human. Since then, various forms of animal liberation have been proposed on both consequentialist and deontological grounds (see Consequentialism; Deontological ethics).

It was also in the eighteenth century that the Swiss philosopher Jean-Jacques Rousseau urged extending our moral concern beyond animals to include care for plants and all living things. Deploring the utilitarian preoccupation with medicinal uses of herbs, he quips: 'It is no use seeking garlands for shepherdesses among the ingredients of an enema' (1782: 110). At roughly the same time, Immanuel Kant's *Kritik der Urteilskraft* (*Critique of Judgment*) (1790) emphasized the importance of aesthetic experiences of nature to human wellbeing (see Kant. I. §12). Kant does not, however, suggest that animals or other natural things have any moral worth in their own right. If it is morally wrong to beat a dog or damage a beautiful landscape, this will be because of the damage these acts do to human sensibilities and character.

Rousseau was the inspiration for French romanticism and can claim to be the first environmental ethicist. Two centuries later, Aldo Leopold, professor of forestry at the University of Wisconsin unconsciously echoes him: 'Examine each question in terms of what is ethically and aesthetically right, as well as what is economically expedient. A thing is right when it tends to preserve the integrity, stability and beauty of the biotic community' (Leopold, in Zimmerman *et al*. (eds) 1993). Leopold's land ethic has been philosophically influential since its revival by American writers in the 1970s and 1980s.

2 Radical environmental ethics

The best-known radical ethic, Arne Naess' 'deep ecology', argues that human self-realization depends on identification with nature (see Self-realization). Deep ecology started as a doctrine of biospheric egalitarianism - that all living things have the same claim to live and flourish. It evolved into a platform meant to embrace all those who recognize the inherent value of natural things and who share a concern to preserve natural diversity whatever

their differences in underlying philosophies (see Zimmerman *et al*. (eds) 1993). A key point of deep ecology is that all living things are members of larger biotic or ecological communities. The larger community may then be regarded as a place of value, with individual needs and projects assessed in terms of their contribution to the good of the larger whole.

Other radical positions are already aligned with well-known political standpoints. Murray Bookchin's 'social ecology' is a type of Green anarchism, while some ecofeminists regard the destruction of nature as intimately linked to the oppressive structures of patriarchy (see Anarchism). Although these latter positions are associated with what Michael Zimmerman (1994) calls 'countercultural' movements, they need not be morally radical in the sense of advocating an ecocentric or biocentric ethic. Most eco-anarchists hold anthropocentric value theories, and different feminist theories give contrasting accounts of the value of nature and the position of women with respect to it. As a result, there is no generally agreed radical platform. Instead, there is often strife among competing interests. For example, deep ecology has been attacked by ecofeminists as a form of patriarchal imposition (Salleh, in Brennan (ed.) 1995) or as embodying masculine egoism (Plumwood 1993).

3 Applications and environmental politics

In an applied turn for the subject, several writers have studied the philosophy of policy and the limits of economic methods in application to environmental valuation. Typical work in this area is by Mark Sagoff, Bryan Norton and Kristin Shrader-Frechette. For the applied workers, issues about nuclear wastes, pollution, cost-benefit analysis and the working of the global economy are of central importance (see Technology and ethics). However, for many of these writers, the 'environment' in question is still the wilderness. Others, such as Henry David Thoreau and Martin Heidegger have found nature closer to home. Heidegger's work has been taken up and applied by some writers on architecture, especially Christian Norberg-Schulz (1988).

For Heidegger human life takes place 'between earth and sky' (see Norberg-Schulz 1988: 38-42). Architecture and engineering are means of bringing a landscape into being: 'The bridge gathers the earth as landscape around the stream…. It does not just connect banks that are already there. The banks emerge as banks only as the bridge crosses the stream.' The concepts of place, orientation and identification are central to his account of humans as dwelling in a land. It can also be argued that an ethic of care for the earth is derivable from Heidegger's work: the contribution of the individual to the greater totality is to protect and articulate the place they have been given to take care of. This may be the meaning of Heidegger's statement that 'mortals dwell inasmuch as they save the earth… ' (1971). This account accepts the inevitability of, and necessity for, human intervention in nature. Despite Heidegger's problematic association with German National Socialism, Norberg-Schulz provides a detoxified interpretation which permits his ideas to resonate with those of ecofeminists, Leopoldians and 'deep' ecologists.

Bertrand Russell comments in his *History of Western Philosophy* that Hitler is an outcome of Rousseau. Certainly, there are well-known connections between ecologistic thinking and the political right, not just confined to Heidegger. Ernest Haeckel,who coined the term 'ecology', was associated with the extreme right, the Hitler youth were taught to value nature and the SS training manual declared the forests of Germany to be of special value (Pois 1986). Although followers of the deep ecology platform are typically vague about the political solutions they put forward, the position has on occasion been accused of supporting 'eco-fascism'. The fear behind this accusation is that biocentrism or ecocentrism may motivate the state to be unacceptably coercive towards individuals for the sake of some larger environmental good. Usually, however, supporters of deep ecology take pains to distance themselves from fascistic or totalitarian solutions to environmental problems. It should be noted that one of the best-known published defences of environmental totalitarianism is in the work of Hans Jonas (1985), an avowed anthropocentric theorist.

4 A contested project

Since the 1970s, acceptance that living things, and possibly all natural things, have value in their own right has been the touchstone to distinguish those who are 'deeper' Green theorists from those of a 'shallower' style (see Sylvan and Bennett 1994). It very soon appeared that the animal protection movement and the 'new' environmental consciousness were on a collision course (see Animals and ethics). While animal liberationists focused on freeing animals from various human oppressions, the new ecological sensibility was less concerned about the evils of the intensive farm than with the fate of its waste stream. Extreme animal advocates thought up

schemes for reducing the total amount of animal suffering by the abolition of predation and parasitism; meantime the new ecological thinkers recognized the naturalness of suffering and death in the struggle for existence.

Behind this difference lay a lack of clarity about the concept of nature. Should humans in some sense ‘follow nature’ despite the presence of pain and harm throughout the living world (see Mill 1874; Rolston, in Brennan (ed.) 1995)? The French theorist Michel Serres (1995) diagnosed the environmental crisis as stemming in part from the inability to state coherently the distinction between ‘nature’ and ‘culture’ (see French philosophy of science §7). In a related vein, Jean-François Lyotard has suggested that animals can be wronged in a way that forbids restitution: their difference from humans means that the rules of human justice, and for the assigning of damages, cannot be applied legitimately to them. So animals are forever barred from the possibility of bearing witness according to human rules (see Lyotard, J.-F.). If Lyotard’s suggestion were extended to nonhuman objects in general, then there would arguably be no viewpoint from which a comprehensive ethic of the environment could be formulated. Notice that the grounds for such pessimism seem overstated. Even if the rules of human justice do not, literally, apply to dogs, cats and rabbits, we have no difficulty in describing actions towards animals as ‘kind’, ‘unfair’, ‘cruel’ and so on.

Maintaining a narrow perspective, the agenda set by many American, Australian and Norwegian writers was focused firmly on issues concerned with wilderness and the reasons for its preservation. Virtually no attention was paid to the built environment. Outside the cities, the environment of most people is the savannah. But it was seldom mentioned either. Instead, most environmental philosophers were concerned with the forest and the mountain, places of recreation for the elite and of permanent abode of only a few humans. Perhaps the new sub-discipline should have been called ‘the philosophy of wild places’ or ‘jungle ethics’.

Writers like Naess and Rolston have followed the eighteenth-century romantics - especially Kant and Rousseau - by emphasizing the importance of wilderness experience to the human psyche. Lifestyles in which such enthusiasms can be indulged demand a standard of living far beyond the dreams of most of the world’s population. Travel to wild places often uses the jet aircraft and motor transport. It has been argued that the motor car is at the heart of many of the most serious environmental problems, including pollution and resource depletion. The economic structures which support elite access to wilderness seem to be implicated in the destruction and pollution which provoked environmental concern in the first place. Yet there is seldom much discussion of car-dependency in the literature.

This is not the only area of silence. The exploitation, oppression and war associated with securing oil supplies and other resources are not normally discussed. A similar silence hangs over toxic waste dumping, forced movements of populations, civil war, the causes of world poverty and the other social features that have damaging environmental impact. In the absence of a discussion of the unequal use of resources and energy across different groups of people and different nations, the focus on wilderness preservation led to a suspicion that Western-style radical environmentalism had a bias against the developing countries. Perhaps, indeed, it was another aspect of the imperialist or colonizing projects of the rich countries (as suggested in Guha, in Brennan (ed.) 1995) (see Development ethics).

The project of environmental ethics involves intense scrutiny of the foundations of value theory. As a result, there is the potential for engagement with mainstream moral philosophy as well as with rational actor theory and economics. Questions about the aesthetics of nature resonate with many topics in literary and art theory, including the significance of originals as opposed to fakes (see Elliot, in Elliot (ed.) 1995). Finally, the project provides a route into discussions of the general predicaments of modernity and postmodernism, as evidenced by Zimmerman (1994) and Serres (1995) (see Postmodernism).

See also: Agricultural ethics; Ecological philosophy; Engineering and ethics; Future generations, obligations to

ANDREW BRENNAN

References and further reading

Brennan, A. (ed.) (1995) *The Ethics of the Environment*, London: Dartmouth.(A selection of essays aimed at the researcher and advanced student, covering the central areas of the field. Among the articles included are:‘The Ethics of Respect for Nature’, by Paul Taylor; ‘Conservation and Self-Realization: A Deep Ecology Perspective’, by Freya Mathews; ‘The Ecofeminism/Deep Ecology Debate’, by Ariel Salleh; ‘Can and Ought

We to Follow Nature?', by Holmes Rolston; and 'Radical American Environmentalism and Wilderness Preservation: A Third World Critique', by Ramachandra Guha.)

Elliot, R. (ed.) (1995) *Environmental Ethics*, Oxford: Oxford University Press.(This collection includes material critical of radical and 'deep' approaches. Includes the editor's own essay, 'Faking Nature'.)

Heidegger, M. (1971) *Poetry, Language, Thought*, trans. and with an introduction by A. Hofstadter, New York: Harper & Row.(Three essays in this volume - 'The Origin of the Work of Art', 'The Thing' and 'Building, Dwelling, Thinking' - outline Heidegger's theory that our buildings can both disclose and appropriate our natural surroundings. When we build poetically, we dwell in and belong to the places we occupy.)

Jonas, H. (1985) *The Imperative of Responsibility: In Search of An Ethics for the Technological Age*, Chicago, IL: University of Chicago Press, 128-38.(Although arguing that natural things have a good of their own, Jonas insists that we only have a responsibility to care for nonhuman things to the extent that their good impinges on human flourishing.)

Kant, I. (1790) *Kritik der Urteilskraft*, trans. W.S. Pluhar, *Critique of Judgment*, Indianapolis, IN: Hackett Publishing Company, 1987.(Places emphasis on the importance of aesthetic experiences of nature to human wellbeing.)

Mill, J.S. (1874) 'Nature', in *Three Essays on Religion*, London: Longmans, Green, Reader and Dyer; repr. in *Collected Works of John Stuart Mill*, ed. J.M. Robson, London: Routledge and Toronto, Ont.: University of Toronto Press, 1991, vol. 10.(Poses a classic dilemma: the injunction to follow nature is either empty or wicked. Anyone who advocates 'following nature' needs to consider Mill's arguments.)

Norberg-Schulz, C. (1988) *Architecture, Meaning and Place*, New York: Electa/Rizzoli.(Provides a reading of Heidegger enriched by his own striking turns of phrase. An easy introduction to difficult subject-matter.)

O'Neill, J. (1993) *Ecology, Policy and Politics*, London: Routledge.(A useful study of environmental ethics and policy informed by an Aristotelian conception of human wellbeing.)

Plumwood, V. (1993) *Feminism and the Mastery of Nature*, London: Routledge.(A sophisticated and philosophically-challenging study of ecofeminism.)

Pois, R.A. (1986) *National Socialism and the Religion of Nature*, London: Croom Helm.(A cautionary historical survey of how ideas about nature were partly constitutive of the myth of the *Herrenvolk*.)

Rousseau, J.-J. (1782) *Rêveries du promeneur solitaire*, trans. P. France, *Reveries of the Solitary Walker*, Harmondsworth: Penguin, 1979.(In this neglected work, Rousseau dreams of touching nature, identifying with it and so becoming free from 'the memory of men and the attacks of the wicked'. The solitariness of the walker reminds the reader that such identification requires withdrawal from society and the social contract.)

Serres, M. (1992) *The Natural Contract*, trans. E. MacArthur and W. Paulson, Ann Arbor, MI: University of Michigan Press, 1995.(Stimulating, if obscure, attempt to explain why a global environmental problem may elude both human description and social legislation.)

Sylvan, R. and Bennett, D. (1994) *Greening Ethics*, Cambridge: White Horse Press (A defence of a radical environmental ethic with attention to its political consequences.)

Zimmerman, M., Callicott, J.B., Sessions, G., Warren, K.J., Clark, J. (eds) (1993) *Environmental Philosophy: From Animal Rights to Radical Ecology*, Englewood Cliffs, NJ: Prentice Hall.(A typical introductory collection focusing on often-quoted material. Includes Aldo Leopold's 'The Land Ethic'.)

Zimmerman, M. (1994) *Contesting Earth's Future*, Berkeley, CA: University of California Press.(This work discusses countercultural movements in the United States in the context of the postmodern predicament.)

Epicharmus (*c*. early 5th century BC)

One of the earliest Greek dramatists, Epicharmus wrote mostly comedies with mythological content in Sicily around 500 BC. A number of philosophical passages were attributed to him in antiquity; their authenticity, sometimes doubted even then, has remained controversial, but the quotations were widely influential and remain of interest.

Although Epicharmus' comedies are lost, fragments provide tantalizing hints of their character: short farces filled with slapstick, stock characters, lively language and sententious maxims; mythological burlesques, often involving Hercules and Odysseus. His appreciation for the comic possibilities of all aspects of contemporary Greek culture led him on occasion to introduce and parody such recent developments as rhetorical theory and philosophical speculation; for example, Aristotle reports that he attacked Xenophanes, and one of his mythological plays was entitled *Logos and Logina ('Mr and Mrs Reason')*.

In later antiquity, Epicharmus' frequent aphorisms and occasional (pseudo-) philosophical passages encouraged the popular view that he was a Pythagorean sage. Treatises on physics and medicine were attributed to him and 'authenticated' by acrostics, and collections of useful but largely spurious maxims circulated under his name (one introduction to such a collection survives on a papyrus); as early as the fourth century BC the Aristotelian Aristoxenus recognized that at least one philosophical work attributed to Epicharmus was not genuine. Doubters remained a minority however.

Also in the fourth century BC, the Sicilian author Alcimus accused Plato of having plagiarized the theories of Epicharmus (whom Plato twice cites with respect); four passages cited as evidence by Alcimus and said by him to derive from Epicharmus are transmitted by Diogenes Laertius. Only the first one (fr. 1) seems likely to be by Epicharmus. Its high-flown, seemingly serious contrast between the constant mutability of human beings and the eternal changelessness of the gods in fact probably served to make a ludicrous comic point: yesterday's borrower owes no money today, and the guest invited yesterday may be refused entrance today, since in the meantime each has become a different person; a man who promised payment yesterday but refuses payment on these grounds today can be beaten by his creditor with impunity - the defaulter's complaint can be refuted on these very same grounds. On the basis of this passage, Epicharmus was later said to have invented the so-called 'growing argument': if growth is the addition of new matter, whoever grows is constantly becoming a different individual, in which case there is no enduring individual who can be said to have grown (see Stoicism §6).

The other fragments distinguish the Good in itself from good people, who show by their actions that they have learned what it is, attribute wisdom to animals, and assert that beauty is in the eye of the beholder. It is unlikely that Epicharmus is the author, and they are certainly not the source of Plato's doctrines; but they do provide fascinating evidence of the degree to which philosophical discourse could find a place on the comic stage.

GLENN W. MOST

References and further reading

Cassio, A.C. (1985) 'Two Studies on Epicharmus and his Influence', *Harvard Studies in Classical Philology* 89: 37-51.(Discusses Alcimus' use of passages from Epicharmus to attack Plato and Aristoxenus' Pythagorizing interpretation of Epicharmus.)

Epicharmus (*c*. early 5th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidmann, 1952, 6th edn, vol. 1, 190-210.(The standard collection of the ancient sources; includes Greek texts with translations in German.)

Freeman, K. (1957) *Ancilla to The Pre-Socratic Philosophers*, Cambridge, MA: Harvard University Press, 34-40. (An English translation of the fragments in Diels' collection.)

Kaibel, G. (ed.) (1899) *Comicorum Graecorum Fragmenta. I.1: Doriensium Comoedia Mimi Phlyaces (Fragments of the Greek Comic Poets. I.1: The Comedy of the Dorians, Mimes, Phlyacean Farces)*, Berlin: Weidmann, 88-147.(Provides the fullest collection of surviving reports and fragments.)

Pickard-Cambridge, Sir A. (1927) *Dithyramb Tragedy and Comedy*, revised T.B.L. Webster, Oxford: Clarendon Press, 2nd edn, 1962.(Chapter 4 pages 230-88, especially 247-55, presents an excellent survey of the material and its difficulties.)

Sedley, D. (1982) ‘The Stoic Criterion of Identity’, *Phronesis* 27: 255-75.(An important and stimulating analysis of the growing argument.)

Epictetus (AD *c*.50-*c*.120)

*Epictetus was a Greek Stoic philosopher of the late first and early second centuries AD. He developed Stoic ideas of responsibility into a doctrine of autonomy and inner freedom based on his concept of moral personality (*prohairesis*). Ethics and practical moral training are central to his thought, but he was also responsible for innovations in epistemology. He emphasized the need to achieve freedom from the passions and to maintain equanimity in the face of a world determined by a providential, though often inscrutable, fate. He frequently treats the Stoic Zeus as a personal deity, and his distinctive combination of personal piety and stringent rationalism (together with his pungent style) have contributed to his enduring influence.*

1 Life and works

Born in Hierapolis of Phrygia, Epictetus was a slave owned by a powerful freedman at the court of Nero. He became a follower of the Stoic Musonius Rufus and a philosopher in his own right. At around the age of 40 he was banished and moved to Nicopolis on the Adriatic coast of Greece; there he taught until his death some time after 120. His thought owed most to early Stoicism, especially its third head Chrysippus, on whom he lectured. Epictetus wrote nothing for publication, but his student Arrian (a Roman aristocrat) recorded and published his informal lectures, mostly on ethics (the *Discourses*, of which four books survive). Epictetus' formal teaching consisted of the exegesis of early Stoic texts and possibly those of other philosophers. Arrian also compiled a *Handbook* of Epictetus' teaching (the *Enchiridion*). The Neoplatonic commentator Simplicius wrote a commentary on it, and it has been widely read since its revival in the Renaissance.

2 Teachings

The central idea of Epictetus' moral teaching is the distinction between what is in our power (*eph' hēmin*) and what is not. This contrast goes back to early Stoic discussions of determinism and moral responsibility (see Stoicism §21). For Epictetus, only our mental life (thoughts, beliefs, decisions, emotions) is in our power and so 'free'; hence it and the moral state dependent on it (virtue or vice) are the key to happiness. Everything else, including bodily pleasure and pain, is subject to control by external forces and so irrelevant to genuine moral welfare.

Epictetus did not organize his teaching around the traditional triad of logic, physics and ethics. Rather (*Discourses* III 2) he developed a scheme of three topics or areas of practice. First, desires and aversions must be managed, so as never to desire the unattainable nor to flee the inevitable. Two mental techniques are recommended to achieve this: the rational anticipation of possible negative outcomes and 'reservation' (*hupexairesis*, a restriction of desires with the proviso 'if that is what Zeus wills for me'). Since the only truly valuable things are in our power, this goal can be achieved by learning the difference between what is good and what is merely 'preferred'. Second, one must learn to manage impulses and choices and learn what the appropriate thing to do is in different circumstances; one must also learn the importance of living in an orderly, well-thought-out manner. The third topic aims at gaining control over one's own assent so that all error and precipitancy are avoided.

The first topic will free us from irrational passions (*pathē*). The second will guide us in our dealings with others. The third topic - based essentially on logic and epistemology - is reserved for those who have progressed through the first two. The intellectual strength it yields should be put to use in supporting a moral life, not indulged in for its own sake - a conception of the purpose of logic which goes back to the earliest generations of the Stoic school. In the *Discourses* Epictetus stresses that logic is of no value in its own right, but that it must serve ethical needs. (He makes the same point about virtually every form of learning, whether the expertise of professional literary critics and grammarians or even his own ability to expound the classical texts of early Stoicism.) The role of logic, then, is to help us understand the workings of the world and our place in it - which requires the ability to make reliable inferences from careful observations.

Epictetus made one important innovation, which concerns the so-called 'preconceptions' (*prolēpseis*), the antecedent notions which most humans share. Earlier Stoics too were committed to the task of examining them and rendering them consistent with each other, with their experience, and with the common conceptions of other people. But for earlier Stoics the origin of these preconceptions lay in the normal experience of the world which

virtually everyone shares. Epictetus converts these preconceptions into something approaching innate ideas. The impact of this change on ethics was negligible, but it foreshadowed important epistemological developments and the openness to Platonism which one senses in Marcus Aurelius.

We may form an impression of Epictetus' competence in logic and dialectic from his account of the master argument of Diodorus Cronus §5, which shows first-hand familiarity with several different Stoic views as well as with Diodorus, and from his many references to the technicalities of Stoic logic (*Discourses* II 19).

Physics, the second traditional branch of philosophy, is largely taken over from earlier Stoic theory. The key point is the providential organization of the world by nature (which Epictetus often regards as a personal god, Zeus). Nature is a rational organizing force and the principles which guide it are similar in kind to our own rationality. Hence, like earlier Stoics, Epictetus regards it as the perfection of our own rationality to accommodate ourselves willingly to the inevitable operations of the world. Where Chrysippus had spoken of the goal of life as learning to live in accordance with an understanding of what happens by nature, Epictetus speaks more often of following the will of Zeus. This more religious tone may be due to the nature of the *Discourses*, addressed as they were to a non-specialist audience; there is no reason to believe that Epictetus' views on the importance of physics or theology to ethics differed much from those of Chrysippus or Cleanthes (whose *Hymn to Zeus* he quotes at the end of the *Enchiridion* alongside Plato's *Crito* and *Apology*, and Euripides).

3 Ethics

Ethics is clearly the core of Epictetus' teaching. Following a trend which is also apparent in the work of Seneca and Musonius, and which culminates perhaps in the *Meditations* of Marcus Aurelius, Epictetus emphasizes the importance of inner mental life. All philosophers who claimed Socrates as their basic inspiration held that happiness (Eudaimonia) was the goal of life and that it depended on the care of the soul. But later Stoics put more emphasis on the autonomy of our inner life, its independence of the contingencies of our bodily and social experience. Several features of Epictetus' thought flow from this:

(1) The contrast between what is in our power (*eph' hēmin*) and what is not: 'In our power are belief, impulse to action, desire, aversion - in a word, everything that *we* do; not in our power are our body, possessions, reputation, political office - in a word, everything that is not our own doing' (*Enchiridion* 1.1).
(2) The focus on *prohairesis*, or moral personality. Earlier Stoics seldom mention this term (which is important for Aristotle (§20), though in a different sense); Epictetus adopted it to express the idea that their moral identity is something which rational agents can control and for which they are wholly responsible. In this respect it recalls Seneca's novel emphasis on will (Latin *voluntas*). Epictetus does not refer to anything which could not be expressed by referring to the commanding faculty (*hēgemonikon)* and to assent as earlier Stoics did, but the new term permits an emphasis on inner mental life.
(3) The polarization between our impressions (*phantasiai*) and the critical use which we make of them. Impressions can be external, such as the appearance of a possible source of pleasure or pain, or internal, such as our own opinions and notions, but all must be subjected to critical scrutiny before we accept them. This process, the 'use of impressions', is, like the Socratic *elenchos* (see Socrates §§2-3), internalized, and it is the most important moral practice which Epictetus urges on his audience.

Other philosophical schools appear frequently in Epictetus' lectures; he does not merely preach to the converted. He takes an anti-Epicurean stance, attacking Epicurus' neglect of logic, his denial of providence, his hedonism, and most of all the anti-social implications of his egoism (see Arcesilaus; Carneades; Epicureanism §§10-11). The Academics are also attacked for their scepticism. As one might expect, Epictetus has an ambivalent attitude to Cynics. He rejected the extremes of their life-style. But from another point of view, their moral autonomy made them an ideal for which to strive. It was inevitable that Epictetus should look up to the Cynics, whose very extremism turned them into symbols of the values which he held most dear. All human beings, he thought, must strive for inner freedom; the independent and even anti-social behaviour of the Cynics expressed in a socially visible form the ultimate goal of moral life for Epictetus.

See also: Stoicism

BRAD INWOOD

List of works

Epictetus (*c*. early 2nd century AD) *Discourses*, trans. W.A. Oldfather, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925-8; trans. J. Souilhé, *Epictète: Entretiens*, Paris: Les Belles Lettres, 1941.(Greek text with, respectively, English translation and French translation; the latter edition is the more up to date.)

Epictetus (*c*. early 2nd century AD) *The Handbook (Enchiridion)*, trans. W.A. Oldfather, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925-8; trans. N. White, *The Handbook of Epictetus*, Indianapolis, IN: Hackett, 1983.(The former includes Greek text; the latter is the best available translation, with a good introduction to Epictetus.)

References and further reading

Arnold, E.V. (1911) *Roman Stoicism*, Cambridge: Cambridge University Press. (Standard work, somewhat dated but still useful.)

Bonhöffer, A. (1890) *Epictet und die Stoa*, Stuttgart: Enke.(Still fundamental.)

Bonhöffer, A. (1894) *Die Ethik des Stoikers Epictet*, Stuttgart: Enke.(Still fundamental.)

Hershbell, J.P (1989) 'The Stoicism of Epictetus: Twentieth Century Perspectives', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin: de Gruyter, II 36.3: 2, 148-63.(Excellent literature review; includes discussion of Epictetus' relation to other philosophers.)

Hijmans, B.L. (1959) *Askēsis: Notes on Epictetus' Educational System*, Assen: Van Gorcum.(Epictetus' life and work; focuses on moral education.)

Long, A.A. (1991) 'Representation and the Self in Stoicism', in S. Everson (ed.) *Psychology*, Cambridge: Cambridge University Press, 102-20.(Emphasizes the philosophical interest of Epictetus' psychology.)

Epicureanism

Epicureanism is one of the three dominant philosophies of the Hellenistic age. The school was founded by Epicurus (341-271 BC*). Only small samples and indirect testimonia of his writings now survive, supplemented by the poem of the Roman Epicurean Lucretius, along with a mass of further fragmentary texts and secondary evidence. Its main features are an anti-teleological physics, an empiricist epistemology and a hedonistic ethics.*

Epicurean physics eveloped out of the fifth-century atomist system of Democritus. The only per se *existents are bodies and space, each of them infinite in quantity. Space includes absolute void, which makes motion possible, while body is constituted out of physically indissoluble particles, 'atoms'. Atoms are themselves further measurable into sets of absolute 'minima', the ultimate units of magnitude. Atoms are in constant rapid motion, at equal speed (since in the pure void there is nothing to slow them down). Stability emerges as an overall property of compounds, which large groups of atoms form by settling into regular patterns of complex motion. Motion is governed by the three principles of weight, collisions and a minimal random movement, the 'swerve', which initiates new patterns of motion and obviates the danger of determinism. Atoms themselves have only the primary properties of shape, size and weight. All secondary properties, for example, colour, are generated out of atomic compounds; given their dependent status, they cannot be added to the list of* per se *existents, but it does not follow that they are not real. Our world, like the countless other worlds, is an accidentally generated compound, of finite duration. There is no divine mind behind it. The gods are to be viewed as ideal beings, models of the Epicurean good life, and therefore blissfully detached from our affairs.*

The foundation of the Epicurean theory of knowledge ('Canonic') is that 'all sensations are *true' - that is, representationally (not propositionally) true. In the paradigm case of sight, thin films of atoms ('images') constantly flood off bodies, and our eyes mechanically register those which reach them, neither embroidering nor interpreting. These primary visual data (like photographs, which 'cannot lie') have unassailable evidential value. But inferences from them to the nature of external objects themselves involves judgment, and it is there that error can occur. Sensations thus serve as one of the three 'criteria of truth', along with feelings, a criterion of values and psychological data, and* prolēpseis*, naturally acquired generic conceptions. On the basis of sense evidence, we are entitled to infer the nature of microscopic or remote phenomena. Celestial phenomena, for example, cannot be regarded as divinely engineered (which would conflict with the* prolēpsis *of god as tranquil), and experience supplies plenty of models adequate to explain them naturalistically. Such grounds amount to consistency with directly observed phenomena, and are called* ouk antimarturēsis*, 'lack of counterevidence'. Paradoxically, when several alternative explanations of the same phenomenon pass this test, all must be accepted as true. Fortunately, when it comes to the foundational tenets of physics, it is held that only one theory passes the test.*

*In ethics, pleasure is the one good and our innately sought goal, to which all other values are subordinated. Pain is the only bad, and there is no intermediate state. Bodily pleasure becomes more secure if we adopt a simple lifestyle which satisfies only our natural and necessary desires, with the support of like-minded friends. Bodily pain, when inevitable, can be outweighed by mental pleasure, which exceeds it because it can range over past, present and future enjoyments. The highest pleasure, whether of soul or of body, is a satisfied state, 'static pleasure'. The short-term ('kinetic') pleasures of stimulation can vary this state, but cannot make it more pleasant. In striving to accumulate such pleasures, you run the risk of becoming dependent on them and thus needlessly vulnerable to fortune. The primary aim should instead be the minimization of pain. This is achieved for the body through a simple lifestyle, and for the soul through the study of physics, which offers the most prized 'static' pleasure, 'freedom from disturbance' (*ataraxia*), by eliminating the two main sources of human anguish, the fears of god and of death. It teaches us that cosmic phenomena do not convey divine threats, and that death is mere disintegration of the soul, with hell an illusion. Being dead will be no worse than not having yet been born. Physics also teaches us how to evade determinism, which would turn moral agents into mindless fatalists: the indeterministic 'swerve'* doctrine *(see above), along with the logical doctrine that future-tensed propositions may be neither true nor false, leaves the will free.*

Although Epicurean groups sought to opt out of public life, they respected civic justice, which they analysed not as an absolute value but as one perpetually subject to revision in the light of changing circumstances, a contract between humans to refrain from harmful activity in their own mutual interest.

1 The Epicureans

For the foundation and nature of the Epicurean school, and for Epicurus' own writings, see Epicurus. After Epicurus' death in 271 BC the school continued to flourish, at Athens and in other centres around the Mediterranean, for at least five centuries. Although there were numerous developments and schisms, these were always moderated by appeals to the scriptural authority of Epicurus' own writings and those of his close collaborators Metrodorus, Polyaenus and Hermarchus.

The surviving writings of Epicurus include *Letter to Herodotus*, *Letter to Pythocles*, *Letter to Menoeceus*, *Kyriai doxai* (*Key Doctrines*) - all preserved in Book X of Diogenes Laertius - and parts of *On Nature*. In addition, we have from the first century BC the Epicurean poem of Lucretius, dealing mainly with physics, and numerous fragmentary papyri of the Epicurean Philodemus, excavated from a villa at Herculaneum. Finally, there is the second-century AD philosophical inscription of the Epicurean Diogenes of Oenoanda, which contains much supplementary information.

2 Foundations of physics

Whatever the universe consists of must be permanent. Nothing comes into being out of nothing or perishes into nothing - two fundamental principles widely regarded as indubitable, but defended by the Epicureans on empirical grounds (Lucretius I 159-264): literal generation would be incompatible with the observed regularities of natural processes, such as the dependence of organic growth on the right material conditions; and literal annihilation would, given the infinity of past time, by now have ensured that nothing at all was left. Nor can the universe (the 'all') be changed by addition or subtraction, since by definition it has nothing outside it that could enter or into which bits could escape.

Next comes a statement of what the universe does consist of: bodies and space (*Letter to Herodotus* 39-40). This is taken to be self-evident, that is, underivatively known and unchallengeable. The reasoning is no doubt that bodies are the things which most obviously have independent existence; and that since that independence is most evident in their ability to move in space, the bits of space which they vacate as they move must exist independently of *them*.

Space is at this stage simply presented as what the bodies are in, and what they move through - a three-dimensional extension which persists whether occupied or unoccupied. When occupied it gets called 'place', when unoccupied 'void' (*kenon*, literally the 'empty'), and when things move through it 'room' (*chōra*, etymologically linked by Epicurus with *chōrein*, 'go'). Epicurus prefers to use these familiar terms rather than a generic one for 'space', for which Greek has no precise word (the concept itself being one which had emerged only gradually, with Epicurus perhaps the first to isolate it clearly), and which he manages to capture only by his invented phrase 'intangible nature' (*anaphēs physis*). That there is void, that is, that contrary to the majority ancient view some space is altogether empty, is argued by appeal to such phenomena as motion and permeation. Motion cannot be accounted for, it is argued (Lucretius I 370-83), by the alternative hypothesis of complementary redistribution of matter on the model of a fish exchanging places with the water when it swims. Even the fish would be stuck fast unless there were some void to break the deadlock. Otherwise it would be unable to move until the water had cleared a path for it, or the water until the fish had cleared a path for it.

As for body, it remains for now largely unanalysed, beyond a set of arguments to show that it must exist microscopically as well as macroscopically: its underlying atomic structure cannot be demonstrated until it has been shown that body and space are the sole constituents of the universe. And the next move is to show just that. First, body and space are analysed as contradictory opposites: only three-dimensional things exist *per se*, and if these are resistant they must be body, if non-resistant void (Lucretius I 430-9). This is the positive proof that body and space are not only irreducibly distinct but also jointly exhaustive. There then follows a supplementary argument (Lucretius I 449-82), in which all other contenders for *per se* existence - including properties and time - are written off as secondary attributes, parasitic on body and/or space. None of these could exist independently of bodies and/or space. Time is dependent on change, that is, on moving bodies. And even facts about the past (for example, that there was a Trojan War), which might seem to outlive the bodies (Agamemnon and others) of which they are true and therefore to acquire independent existence, are truths about places which still persist (Troy, Mycenae and so on), or, if you prefer, about the universe.

Only now that it is fully established can the body-space dualism be deployed to show that at the lowest level of analysis there will be not only portions of empty space uninterrupted by body but also portions of body uninterrupted by empty space - and therefore, since there is no third thing, totally uninterrupted. Being perfectly solid, these are 'atoms', literally 'uncuttables'. The capacity of a body to disintegrate is directly correlated to the amount of void within it; therefore a body containing no void at all is altogether indestructible (Lucretius I 511-39). That there are such bodies is confirmed both by the evident fact that matter is not completely annihilated by fragmentation, and by the observed regularities of nature, which imply that something altogether unchanging underlies them.

Having mapped the universe out into space occupied by discrete portions of body, Epicurus adds that both space and body are infinite in extent (*Letter to Herodotus* 41-2, Lucretius I 958-97). First, the infinity of the universe itself (a highly controversial thesis in antiquity, rejected by both the Platonist and the Aristotelian tradition) is argued by appeal to the notion of a limit: it could only be limited if there were something beyond it to limit it, and the notion of the universe (the 'all') precludes that. (Lucretius adds the time-honoured argument: what if I go to the supposed edge of the universe and throw something?) Second, it is argued that each of the two constituents of the universe must also, taken on its own, be infinite: finite body in infinite space would be too dissipated to form compounds, while infinite body in finite space would simply not fit.

3 Minima

The fifth-century atomists (see Atomism, ancient; Leucippus; Democritus §2) had first introduced atoms at least partly to circumvent the puzzles that Zeno of Elea (§§4-6) had derived from the supposition of infinite divisibility: any magnitude will, as the sum of infinitely many parts, be infinitely large; and motion will be impossible, since it will require the traversal of infinitely many sub-distances. But if atoms, as seems inevitable given their varying shapes and sizes, are not the smallest possible magnitudes but simply indissoluble lumps, it is hard to see that this kind of indivisibility can do anything to thwart Zeno: each atom will consist of infinitely many parts, and threaten to be infinitely large and/or untraversable.

Epicurus (*Letter to Herodotus* 56-9) starts by making the first clear distinction in ancient thought between (1) things which are physically indivisible, and (2) those which modern scholarship sometimes calls theoretically, conceptually or mathematically indivisible, but which in ancient usage are called either 'minimal' (*elachista*) or 'partless' (*amerē*). (For Epicurus' main predecessor in a theory of type (2), see Diodorus Cronus §2.) Not only, Epicurus says (*Letter to Herodotus* 56), (1) can things not be 'cut' to infinity (a reference to atoms, which are physically 'uncuttable'), as proved earlier, 'but also (2) we must not consider that in finite bodies there is *traversal* [moving along something part by part] to infinity, not even through smaller and smaller parts [that is, not even in a convergent series such as $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$]'. Otherwise, he warns, Zenonian consequences will ensue.

By showing that a finite body cannot contain an infinite number of parts, Epicurus considers that he has established the existence of an absolutely smallest portion of body, which henceforward he calls 'the minimum (*elachiston*) in the atom'. Clearly it cannot be larger than an atom, or it would not be a minimum, so it must be either an entire atom or part of one. (For reasons which emerge later, it must in fact be the latter.) The idea that an extended magnitude like an atom consists of minimal or partless units is designed to avoid Zenonian paradoxes, but, as Aristotle had pointed out in *Physics* VI, it generates paradoxes of its own. In particular, it is hard to see how two partless items can be adjacent, since they cannot touch either part to part (they do not have parts) or whole to whole (or they would be co-extensive, not adjacent). Epicurus answers brilliantly with the analogy of the 'minimum in sensation' - the smallest magnitude you can see. Any larger visible magnitude consists of a precise number of these minima, which are seen as adjacent without touching in either of the two ways offered by Aristotle, but, as Epicurus puts it, 'in their own special way'. This provides a model which is readily transferable to the 'minimum in the atom'.

But it also follows, of course, that *any* extended magnitude will consist of an exact number of minima, since after analysis into minima there could not be a fraction of a minimum left over. This makes all magnitudes commensurable, and conflicts with the geometers' recognition of incommensurable lengths, for example, the side and diagonal of a square. So much the worse for geometry, was the conclusion of Epicurus and his collaborator Polyaenus, himself formerly a distinguished geometer. They concluded, at least provisionally, that geometry is false.

In one respect the analogy between the sensible and actual minimum fails. Visible minima are capable of independent motion (for example, a falling object viewed at a suitable distance might be a sensible minimum), but actual minima are not, says Epicurus. This perhaps reflects Aristotle's argument in *Physics* VI 10 that a partless entity (he is thinking especially of a geometric point) could not move except incidentally to the motion of a larger body. At all events, it follows that, since atoms move, no atom consists of just one minimum.

Atoms must vary widely in shape and size in order to account for the large range of phenomena (Lucretius II 333-477). But they cannot vary indefinitely, since out of a finite set of minima only a finite number of shapes can be formed. Being partless, minima cannot be partly adjacent, so two minima can only be arranged in one shape, three minima in only two shapes, and so on (Lucretius II 478-531).

4 Motion

Surprisingly, atoms never stop moving, even within a compound object, since the medium through which they move is void, which can offer them no resistance. More surprisingly, for the same reason they move at a vastly greater speed than any familiar motion through an obstructive medium such as air; even than sunlight, which is seen to spread from horizon to horizon virtually instantaneously (Lucretius II 142-64). More surprisingly still, they all move at equal speed, since in a vacuum, unlike air, there is no resistance from the medium to slow down the lighter ones more than the heavier ones (*Letter to Herodotus* 61). In stating all these claims, Epicurus is accepting paradoxical consequences of the hypothesis that void exists, consequences which Aristotle had drawn (*Physics* IV) in the belief that they were sufficiently absurd to discredit the hypothesis. Moreover, the equal speed of atoms was confirmed by another objection Aristotle thought he had found to atomism (*Physics* VI 2): if there is a minimal magnitude, there can be no differences of speed, because then in the time the faster object took to travel one minimum the slower one would, impossibly, have to travel less than one minimum. Epicurus welcomed this argument, along with the conclusion Aristotle thought absurd, because his theories of void and minima now offered two independent grounds for the same conclusion, that atoms move at equal speed.

The apparent lack of fit between these findings about atoms and the variable speed of macroscopic motions is explained as follows (*Letter to Herodotus* 62). Even in a compound object the individual atoms are perpetually moving, but in tight and regular cyclical patterns which make the complex as a whole stable. Phenomenal differences of speed, say between two runners, represent merely the aggregate motions of the atoms in each over an observed period of time.

There are three causes of an atom's motion. The first is its own weight, interpreted as an inherent tendency to move downwards (see §8). The second is collisions with other atoms, which can deflect an originally downward motion along any number of new rectilinear trajectories, thus generating the patterns of motion of which compounds are born.

The third cause of atomic motion is the 'swerve' (*parenklisis*), whereby an atom may shift from its rectilinear trajectory onto an adjacent one - a displacement sideways by a distance of one minimum (there being no smaller distance). This happens 'at no fixed place or time', meaning that the occurrence of a swerve is causally undetermined. The theory, derided by Epicurus' opponents but now recognized as comparable in its implications to modern quantum indeterminism, looks like a drastic solution requiring a drastic problem. Two such problems are recorded (Lucretius II 216-93). First, since all atomic motion starts out as vertical and equal in speed, without a swerve no collisions would ever have started, and hence no world could have been formed. It may be doubted whether this was a sufficiently pressing problem to motivate an abandonment of universal causality: given the infinite past history of the universe, Epicurus had no need to posit a very first collision; in which case every collision could have been explained as the effect of previous ones. The second problem seems to have been the real motivation of the swerve: if all atomic motion is causally determined, free will becomes impossible (see §12).

5 Qualities

Atoms themselves have only the primary or ineliminable features of body: size, shape and weight. They lack the secondary properties of colour, flavour and so on. The ground for this parsimony is (*Letter to Herodotus* 54-5) that secondary properties are in their nature changeable, whereas atoms have been posited as the enduring entities which underlie change. Atomism instead treats colour and the like as purely macroscopic properties, caused by the

individual shapes and overall arrangement of the constituent atoms in a compound body, and changed if that arrangement changes. Lucretius frequently compares the explanatory economy of this system to that of the alphabet, capable of generating endless different words simply by rearranging a modest stock of letters. In practice the crudely sketched atomistic explanations of macroscopic properties owe more to the shapes than to the arrangement of a thing's constituent atoms (fluids consist of smaller and rounder atoms, sweet things of smoother ones…), although the degree of separation between atoms plays some part.

There are two kinds of properties (*symbebēkota*). 'Permanent accompaniments' (Greek *ta aidion parakolouthounta*, Latin *coniuncta*) are essential to a thing's very existence, for example, tangibility for body and heat for fire, while accidents (*symptōmata*) are non-essential, for example, slavery or poverty for a human being. It is crucial to note that no ontological priority is implied for the properties of atoms over those of phenomenal bodies. The earlier atomist tradition had tended to treat atoms and void alone as real, with phenomenal properties mere arbitrary constructions placed upon them by the human mind and sense organs. This had led atomism in the direction of epistemological scepticism, a tendency which Epicurus himself strenuously resists. In discussing the status of properties (*Letter to Herodotus* 68-71), he emphasizes that although undoubtedly different from the *per se* existents, body and void, they are no less real for that. (The second-generation Epicurean Polystratus adds that even relative properties such as 'beneficial' and 'harmful', often decried as unreal by sceptics, have clear causal effects.) Accidents, in fact, have no existence at the microscopic level (*Letter to Herodotus* 70), and yet are real. This is a clear indication that Epicurus is consciously opposed to the atomist reductionism of his predecessors: colours and other accidents are real, yet irreducibly different from atomic structures. Although atomic structures are *causally* prior to the phenomena which they generate, they are in no way *ontologically* privileged over them.

6 The criteria of truth

Since, then, Epicurus' version of atomism allows the reality of sensible properties, he does not inherit Democritus' motive for casting doubt on the veracity of sense-perception (see Democritus §3). And indeed for Epicurus the primary 'criterion of truth' is the senses: 'All sensations are true.'

The soul is an atomic structure spread throughout the body, but with a command centre which houses rational thought (these can be considered functionally, although not anatomically, equivalent to the nervous system and brain, respectively). A sensation is the soul's mechanical but conscious registering of the phenomenal properties of an external body. In the paradigm case of sight, this occurs because the outer layers of bodies are constantly streaming off them in all directions, taking the form of 'images' (Greek *eidōla*, Latin *simulacra*), atom-thin films of matter which more or less preserve their colour and shape in transit (*Letter to Herodotus* 49-52, Lucretius IV 26-268). Streams of these can enter the eye, producing vision; isolated ones can also directly enter the mind, asleep or awake, and enable it to visualize objects (Lucretius IV 722-822, 962-1036). Importantly, *any* act of picturing, by the eyes or the mind, is the registration of one or more images arriving from outside. The infallibility of the senses consists in their mechanical registering of these images, without adding, subtracting, embroidering or interpreting. A useful analogy is photography, which 'cannot lie' because it merely records mechanically the patterns of light arriving at the camera lens, and leaves it to us to interpret what they represent. All sensations are bona fide evidence about the external world. All error lies in the 'added opinion' (*to prosdoxazomenon*) by which the mind interprets these data. Note that truth here is representational, not propositional: the sensation is true because it accurately represents the physical data reaching the sense organ. The opinion based on it may be regarded as true or false according to whether or not it succeeds in representing accurately the state of affairs which caused the sensation. Even in a case of outright delusion it is the mind which, because deranged, misinterprets the perfectly accurate impressions which reach it. As for the more important counterexamples to the theory, those of optical illusions, the paradigmatic case is the square tower which looks round at a distance. The images start out square, but because of the distance which they must travel through obstructive air they are rounded on arrival at the eyes (if this is too crude to be credible, substitute the image of a coin, oval due to perspective). Therefore vision is accurately registering the images as they are on arrival. This is a correct, not a misleading, view of a distant tower, and to call it a case of mis-seeing, says Epicurus (Sextus Empiricus, *Against the Professors* VII 206-10), is like saying that you are mishearing someone across the room just because you are not hearing their voice as it would sound inside their mouth.

The veracity of the senses is not, as many hold, impugned by their ability to conflict with each other. Strictly

speaking, no two sensations are commensurable. If sight and touch differ about the shape of an object, that is because sight reports the shape of a colour patch, touch the shape of a body. If two visual sensations of the same object conflict, as in the case of the tower, that is because they are reporting two different kinds of colour patch, mediated by different quantities of intervening air.

The photographic reliability of the senses is what makes them not only true but a criterion of truth. However, when Epicurus appeals to the evidence of the senses as his ultimate criterion, it is not usually to these raw sensations, still awaiting interpretation by the mind, but to facts directly attested by them, for example, that things move, or that round objects move more readily than jagged ones. He regularly talks as if these interpreted sensations acquire the self-evident veracity which initially belongs only to the raw impressions.

The second criterion of truth is *prolēpsis*, or ‘preconception’ - a key term in ancient epistemology, first introduced by Epicurus. The *prolēpsis* of a thing is an instinctively acquired generic grasp of its nature, which enables you to recognize instances of it and is available for analysis in conceptual inquiries. Your *prolēpsis* of a human being, acquired unreflectively by accumulated past experience, enables you both to recognize humans when you meet them, and in the course of an inquiry to establish their essential features, for example, rationality and mortality. (When *prolēpseis* are said to be built up from past sensory experiences, these must again be interpreted sense-impressions, not raw ones.) Other *prolēpseis* appealed to in Epicurean arguments include those of god, body, utility and responsibility. *Prolēpseis* are taken to be common to all human beings, and therefore to function like a set of shared intuitions which we can hope to rediscover beneath our acquired false beliefs and to use as common ground for joint philosophical inquiry

The third criterion of truth is feelings (*pathē*), which generically divide up into just two kinds: pleasure and pain. As such, feelings certainly function as criteria of *value*, since Epicurus equates good with pleasant, bad with painful. But they also provide the introspective data on which Epicurus founds his psychological theory (*Letter to Herodotus* 63).

7 Scientific method

Starting from the direct data of the senses, broader and less accessible truths can be established. Epicurus apparently divided the accessing process into two classes, ‘attestation’ (or ‘witnessing’, ‘confirmation’, and so on: *epimarturēsis*) and ‘non-contestation’ (or ‘lack of counterevidence’, ‘non-confirmation’, and so on: *ouk antimarturēsis*). Attestation plays a relatively minor part in scientific inference, since it seems confined, if not to truths directly supplied by the senses, at most to inductive generalizations derived from these. The main concern of Epicurean science is the extension of such knowledge beyond the realm of direct experience, into inaccessibly remote regions of the universe such as the heaven or other worlds, and into the microscopic realm of atoms and void. Here theories cannot be directly confirmed, but are tested instead by their *consistency* with empirical data: hence ‘non-contestation’ by phenomena.

Not that *mere* consistency with experience is enough to establish truth. A theory is only entertained in the first place if it has some explanatory power or recommends itself in some comparable way. But there may be several alternative theories of equal merit in this regard, and if so each must be tested for ‘non-contestation’ by phenomena, to see which survives. There are many intrinsically credible general theories of matter, for example, including the four-element theory and the monistic fire theory of Heraclitus, and likewise many intrinsically plausible theories of vision, and so on. But all except the Epicurean theory fail somehow to be consistent with the entire range of phenomena against which they are tested.

In some other cases, however, especially astronomical ones, Epicurus admits that two or more rival theories survive the test. This might be considered an indication that the consistency test can at best prove a theory possible, not true. But Epicurus insists that in such cases all the successful theories are indeed true. Sometimes this means that several different explanations of the same phenomenon - for example, thunder, or the generation of perceptual ‘images’ (see §6) - operate concurrently in our world. In other cases it means no more than that, given the intrinsic possibility of the hypothesized causal process, it must operate *somewhere* in the infinite universe, even if not here.

All such ‘sign-inferences’ (*sēmeiōseis*) are from something evident to something non-evident. Often this is from the macroscopic to the microscopic: either causal inference from a macroscopic effect to a microscopic cause (for

example, from the observed regularities of nature to the existence of unchangeable elements; see §2), or analogical inference (for example, from the mobility of observed spheres to that of spherical atoms, see §5). Others are from the macroscopic to the macroscopic: either from accessible to inaccessible entities or events (for example, from the mechanism of the water-wheel to the rotation of the stars), or from the present to the past or future (for example, from the current structure of a social institution like law or language to its historical origin, or from the impermanence of familiar compounds to the future destruction of the world).

The nature and validity of such sign-inferences is debated between the Epicureans and the Stoics in Philodemus, *On Signs*, including an important controversy about the justification of induction.

8 Cosmology

Epicurus argues that there can be no creating or controlling divinity (see §9), and that our world, one of infinitely many, is an accidental and temporary product of large-scale atomic collisions (for the role of the 'swerve' in this, see §4). Apparent evidence of divine creation can be explained mechanistically. Animal parts, for instance, however well suited to their uses, came into existence accidentally before those uses were conceived. In the early days of life on earth many non-viable creatures were generated, but did not survive (Lucretius V 837-77; a widely admired anticipation of Darwinian survival of the fittest). Even human institutions such as language and law, often attributed to divine benefactors, are formalized versions of modes of behaviour with purely natural origins in human need and instinct.

Epicurus' account of the world's structure is largely Presocratic in inspiration. Since atoms have no inherent attractive powers, he cannot accept the geocentric cosmologies of Plato, Aristotle and others, in which heavy stuffs tend not strictly downwards but inwards towards the centre. The direction 'down' is itself taken to be an absolute one, so that throughout the universe objects fall parallel to each other, rather than towards the centre of a spherical earth. For him the earth's stability depends not on its being at the centre (in infinite space there is no centre) but on the cushioning effects of the air beneath it. In a notorious passage (I 1052-82) Lucretius dismisses as ill-conceived the geocentrists' impressively accurate description of the antipodes.

9 God

That our world cannot be a product of divine craftsmanship is argued on several grounds (especially Lucretius V 156-234). Quite apart from the world's obvious imperfection, and the difficulty of finding a motive for its creation by already blissfully happy beings, the true conception of a god is incompatible with the role of cosmic administrator. A god is a supremely tranquil being, whereas the burdens of government include attitudes of anger, favour and worry.

How do we know that god is like this? In Epicurus' view there is a natural conception (*prolēpsis*: see §6) of god as a blessed and immortal anthropomorphic being, a conception shared by all human beings, even though in most it has been obscured by a veneer of false beliefs, for example, that the gods are vengeful, or that they govern our lives, turn the heavens and so on. People tend to endow god with their own moral values, especially the competitive values of political society, and by the same token the Epicurean reversion to the true conception of divinity as tranquil and detached is also a rediscovery of the natural human goal, tranquillity (see §10). Epicurus is insistent that 'there are gods', and even that they should be worshipped, but as an act of veneration for a life to which we ourselves aspire, not in the hope of appeasement.

But how can there be gods, if that means literally (that is, biologically) immortal beings? If according to Epicureanism nothing exists independently except bodies and void, and a god can hardly be either void or a single atom, a god must be an atomic complex. But it is a cardinal tenet that no compound body can be everlasting (Lucretius III 806-18).

Here scholarly opinion divides. The majority seek special ways in which an Epicurean god can nevertheless be literally immortal, by living in sheltered regions beyond our world and by consisting of constantly replenished streams of visual 'images' (on which see §6). Others (following Sextus Empiricus, *Against the Professors* IX 43-7) favour an idealist interpretation of Epicurean theology, whereby god, properly understood, just is our own idealization of a happy human life. On this view, when god is said by the sources to consist of visual 'images', this is simply because according to Epicurean cognitive psychology (§6) *all* imagination consists in the apprehension

of images which enter us from outside. The images which provide the raw material for our conception of god need not flow to us from any objectively real divine being, but may be ordinary locally generated human images. Epicurus' main point in identifying god with a stream of images was apparently the negative one of denying that god is a 'solid body' at all (Cicero, *On the Nature of the Gods* I 49).

Epicurus seems to have spoken repeatedly about how we should *think* of god as being, behaving and so on, which in itself is equally compatible with the realist and the idealist interpretation. Equally powerless to settle the dispute is the insistence of Epicurus and his followers that their position is theistic, or that of his critics that it is atheistic. But on either reading it should be clear that the importance of god in Epicurus' system is not cosmological but ethical. Having a correct conception of god is identical with moral enlightenment.

10 Hedonism

Another 'innate' human attitude, along with the *prolēpsis* of god, is the pursuit of pleasure as the only positive value, or 'end' (*telos*). Epicurus argues that the behaviour of the newborn, and even of non-human animals, confirms that to maximize pleasure and minimize pain is the natural and primal drive (Cicero, *On Ends* I 30). Just as physics began with the crude mapping out of *per se* existents into bodies and space (§2), so ethics starts with the mapping out of all intrinsic values into pleasure and pain. And again as in physics, the next move consists in showing these two items to be jointly exhaustive: the absence of pain is itself pleasure, and there is therefore no intermediate state (Cicero, *On Ends* I 37-9).

This controversial thesis goes to the heart of Epicurus' ethics. In his view, most human misery results from ignorance of how to quantify pleasure. Where some hedonists, especially the contemporary Cyrenaic School (see Cyrenaics §4), had recommended the constant renewal of pleasure through self-indulgence, Epicurus observes that this accumulation does not increase the total of pleasure beyond that achieved when all pain has gone, but only 'varies' it. Freedom from pain is itself already a supremely pleasant state. The pursuit of luxury, far from increasing pleasure, enlarges your desires and leaves you needlessly vulnerable to the whims of fortune.

In his physics, Epicurus had completed the division of *per se* existents into bodies and void by showing that other claimants to *per se* existence, such as properties and time, are in fact parasitic on bodies for their existence. Likewise in ethics, he now examines the non-hedonic values which others assert, such as virtue, and argues that they are in fact valued not for their own sake but as instrumental means to pleasure (Cicero, *On Ends* I 42-54).

It remains to fill out the prescription for the maximization of pleasure, that is, to sketch the ideal Epicurean life. This involves calculating the relative roles of bodily and mental pleasures, and of static and 'kinetic' pleasures. Bodily feeling is in a way focal, since mental pleasure and pain consist ultimately in satisfaction and dissatisfaction, respectively, about bodily feeling. For instance, the greatest mental pain, namely fear, is primarily the expectation of future bodily pain (which is the main ground, and a mistaken one, for the fear of death). But although mental feelings ultimately depend on bodily ones, and not vice versa, mental feelings are a more powerful ingredient in an overall good life. Someone in bodily pain - which may be unavoidable - can outweigh this by the mental act of reliving past pleasures and looking forward to future ones. It is this ability to range over past and future that gives mental feeling its greater power. But misused, especially when people fear everlasting torture after death, it can equally well become a greater evil than its bodily counterpart.

Static pleasure is the absence of pain. The bodily version of it is called 'painlessness' (*aponia*), the mental version 'tranquillity' (*ataraxia*, literally 'non-disturbance'). Tranquillity depends above all on an understanding of the universe, which will show that contrary to the beliefs of the ignorant it is unthreatening. (This is, strictly speaking, the sole justification for the study of physics.) Kinetic pleasure is the process of stimulation by which you either arrive at static pleasure (for example, drinking when thirsty) or 'vary' it (for example, drinking when not thirsty). There are mental as well as bodily kinetic pleasures, for example, (perhaps) the 'joy' of resolving a philosophical doubt or holding a fruitful discussion with friends. Although kinetic pleasures have no incremental value, Epicurus does apparently consider them an essential part of the good life. This is particularly because the mental pleasure which serves to outweigh present pain will inevitably consist in reliving past *kinetic* pleasures and anticipating future ones. So a successful Epicurean life cannot be monotonous, but must be textured by regular kinetic pleasures. In the letter written on his deathbed, Epicurus claimed that despite the intense bodily pains this was the happiest day of his life, because of all the past joys of philosophical discussion that he could relive.

At the same time, these kinetic pleasures must be carefully managed. Some desires are natural, others empty. The latter - for example, thirst for honours - should not be indulged, because their satisfaction will bring either no pleasure or a preponderance of pain over pleasure. Even of the natural ones, some are non-necessary. For instance, the desire for food is necessary, but the desire for luxurious food is not. In order to be maximally independent of fortune, it is important to stick primarily to the satisfaction of natural and necessary desires. But occasional indulgence in those kinetic pleasures which are natural but non-necessary has a part to play, so long as you do not become dependent on them. True to this principle, Epicurean communities lived on simple fare, and even trained themselves in asceticism, but held occasional banquets (see Epicurus §3).

11 Social values

Ancient ethics does not problematize altruism as such, but it does seek the moral foundations of two specific forms of altruism: justice, that is, respecting the interests of your fellow citizens; and friendship. Given that Epicurean hedonism is egoistic - that all your choices as an agent aim at your own pleasure - is it possible to put someone else's pleasure before your own?

Epicurus analyses justice not as an absolute value but as a contractual relation between fellow citizens, its precise character engendered by current social circumstances (*Key Doctrines* 31-7). Sometimes it proves mutually advantageous to abstain from forms of behaviour which harm others, in return for a like undertaking from them. So long as such a contract proves socially advantageous, it is correctly called 'justice'. It imposes no moral obligation as such, and the ground for respecting it is egoistic - that even if you commit an injustice with impunity the lingering fear of being found out will disrupt your tranquillity (*Key Doctrines* 17, 35).

With regard to his own philosophical community, Epicurus attached positive value to justice and to the specific laws which enforced it not because philosophers need any restraint from wrongdoing but because they need protection from the harm that others might inflict. 'Do not take part in politics' was an Epicurean injunction, political ambition being a misguided and self-defeating quest for personal security. But the school nevertheless upheld the need for legal and political institutions, and sought to work within their framework.

Where the political life fails to deliver personal security, friendship can succeed. The very foundation of the Epicurean philosophical community was friendship, of which the mutual dealings of Epicurus and his contemporaries were held up as an ideal model by their successors. Unlike justice, friendship is held to have intrinsic value - meaning not that it is valuable independently of pleasure, but that it is intrinsically pleasant, not merely instrumentally pleasant like justice. Moreover, the pleasure lies in altruistic acts of friendship, not merely in the benefits received by way of reciprocation.

Later Epicureans were pressed by their critics for a more precise reconciliation of friendship with egoism, and developed the position as follows (Cicero, *On Ends* I 66-70). According to one group, it is indeed for our own pleasure that we form friendships, and it is as a means to this, not ultimately for our friends' sake, that we share their pleasure and place it on a par with our own. A second group veered away from egoism: although friendship starts out as described by the first group, the outcome is something irreducibly altruistic, whereby we come to desire our friends' pleasure purely for their sake. A third group sought to restore egoism: the second group is right, but with the addition that friendship is a symmetrical contract, analogous to justice: *each* friend is committed to loving the other for the other's own sake.

12 Free will

Epicurus was arguably the first to make free will a central philosophical issue (see Free will). He takes it that determinism must be false. It is incompatible with basic moral attitudes; the data of self-awareness falsify it; and it is a position which cannot even be argued coherently, since to enter an argument is to assume that both parties to the debate are responsible for their beliefs and could adopt others (*On Nature* XXV).

Determinism is arrived at by two routes, each of which must therefore contain a false supposition. The first route is via a logical law, that of bivalence: since all propositions, including those about the future, are either true or false, if it is already false that a given event will occur, it cannot occur, and if true, it cannot fail to occur. Therefore everything that occurs occurs of necessity. Epicurus' response is (like Aristotle's in *De interpretatione* 9, as widely interpreted) to reject the law of bivalence as regards future-tensed propositions. Predictions whose accuracy

depends on human decisions yet to be made are, at the time of utterance, neither true nor false.

The second route is physical. We - our souls as well as our bodies - consist entirely of atoms, which move according to mechanical laws. How then can anything be genuinely 'up to us'? Are we not automata, our actions the outcome of infinite atomic causal chains? This time the Epicurean response is that the laws of atomic motion are not after all entirely deterministic. Atoms' motion through weight and blows is mechanical and invariable, but there must be a third, indeterministic, aspect of their motion, the minimal 'swerve' (see §4).

But how do swerves help explain free will? The question is much debated by scholars. Epicurus may appear to be merely substituting an unpredictable mechanism for a predictable one, not putting *us* in charge. Perhaps, as with the denial of bivalence, he is merely attempting to remove an obstacle to free will, not to explain it. Whether the swerve enters more directly than this into his account of volition will depend partly on whether he thinks that mental events such as volition are reducible to atomic changes in the soul, in which swerving atoms could play a part. There is in fact strong evidence (*On Nature* XXV) that he regarded mental events as irreducibly different from the soul atoms underlying them, and even as having their own causal efficacy *on* the atoms. (For his rejection of atomic reductionism, see §5 above.) If so, it may be safest to conclude that volition is already by its own nature autonomous - not part of an antecedent causal chain at all - and that this is the primary explanation of free will. Unfortunately its autonomy would be rendered impotent if either the laws of logic or those of physics had already determined our future actions independently of it. Therefore both sets of laws must be so rewritten as to circumvent this danger, and to keep alternative possibilities genuinely open. In the case of physics, the indeterministic swerve just is the most economical realization of that requirement.

13 Death

That the soul must itself be composed of atoms has, in Epicurus' judgment, one very cheering implication. It can easily be shown to perish with the body. Lucretius (III 417-829) presents a whole battery of arguments for this, based not only on the evidence for the soul's dissolubility but also on other indications, such as the continuous parallelism between its development and decay and those of the body. From the finding that there is no conscious survival after death Epicurus concludes that 'death is nothing to us'. To fear your own future non-existence is as groundless as to regret the time when you had not yet been born, and involves the absurdity of imagining yourself being there to witness and lament your own non-existence. To regret that your pleasures will be cut short is to make the computational mistake of supposing that a finite lifespan is *ipso facto* less pleasant than an infinite one. In a brilliant diatribe against the fear of death, Lucretius (III 830-1094) interprets the myths of punishment in the afterlife as allegories for moral malaise in this life, and portrays much human unhappiness, manifested in the vain search for security through wealth and power, as subconsciously nourished by the fear of death.

14 Influence

Epicureanism enjoyed exceptionally widespread popularity, but unlike its great rival Stoicism it never entered the intellectual bloodstream of the ancient world. Its stances were dismissed by many as Philistine, especially its official rejection of all cultural and intellectual activities not geared to the Epicurean good life. It was also increasingly viewed as atheistic, and its ascetic hedonism misrepresented as crude sensualism (hence the modern use of 'epicure'). The school nevertheless continued to flourish down to and well beyond the end of the Hellenistic age. The poets Virgil and Horace had Epicurean backgrounds, and other prominent Romans such as Cassius, the assassin of Julius Caesar, called themselves Epicureans. In the first three centuries of the Roman Empire many writers show some debt to Epicurean thought, including not only the novelist Petronius but even the Stoic Seneca and the Platonist Porphyry. When Marcus Aurelius (§1), Roman emperor AD 161-80, established four official chairs of philosophy at Athens, a chair of Epicureanism was among them. In later antiquity Epicureanism's influence declined, although it continued to provide a target for thinkers, both Christian and pagan, in search of a godless philosophy to attack. Serious interest in it was revived by Renaissance humanists, and its atomism was an important influence on early modern physics, especially through Gassendi.

See also: Hellenistic philosophy

DAVID SEDLEY

References and further reading

Aristotle (*c*. mid 4th century BC) *Physics*, trans. in J. Barnes (ed.) *The Complete Works of Aristotle*, Princeton, NJ: Princeton University Press, 1984.(Book VI, especially chapters 1 and 10, contains challenges to 'partless' constituents of magnitude, to which Epicurus seems to be responding.)

Asmis, E. (1986) *Epicurus' Scientific Method*, Ithaca, NY: Cornell University Press. (The fullest account of Epicurean epistemology.)

Cicero (45 BC) *On Ends*, trans. H. Rackham in *De finibus bonorum et malorum (On the Ends of Good and Evil)*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1914.(Latin text with English translation; Epicurean ethics is expounded in Book I and criticized in Book II.)

Cicero (45 BC) *On the Nature of the Gods*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933.(Latin text with English translation; Book I is a report and critique of Epicurean theology.)

Diogenes Laertius (*c.* early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols. (Book X, in volume 2, contains the life of Epicurus, along with complete texts of his three letters and the *Key Doctrines*.)

Festugière, A.J. (1946) *Epicure et ses dieux*, Paris: Presses Universitaires de France;; trans. C.W. Chilton, *Epicurus and his Gods*, Oxford: Blackwell, 1955.(Brilliant and accessible portrayal of Epicureanism's philosophical mission.)

Furley, D.J. (1967) *Two Studies in the Greek Atomists*, Princeton, NJ: Princeton University Press.(Classic investigation of Epicurus' theories of minima and free will; fairly technical.)

Jones, H. (1989) *The Epicurean Tradition*, London: Duckworth.(Useful guide to the influence of Epicureanism in antiquity and after.)

Long, A.A. (1974) *Hellenistic Philosophy*, London: Duckworth.(Includes the best introductory account of Epicureanism.)

Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press, 2 vols.(Volume 1 contains translations of the principal texts, with commentary, volume 2 the original Greek and Latin; Epicureanism is in §§4-25.)

Lucretius (*c*.55 BC) *On the Nature of Things*, trans. W.H.D. Rouse, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1975.(Latin text with English translation, revised by M.F. Smith.)

Mitsis, P. (1988) *Epicurus' Ethical Theory*, Ithaca, NY: Cornell University Press. (Lucid philosophical critique.)

Philodemus (*c*.40 BC) *On Signs*, ed. P. De Lacy and E. De Lacy, *Philodemus on Methods of Inference*, Naples: Bibliopolis, 1978.(With English translation and notes; a classic record of a Stoic Epicurean debate on scientific inference.)

Usener, H. (1887) *Epicurea*, Stuttgart: Teubner.(Still the major collection of Epicurean texts and testimonia, but without translation.)

Epicurus (341-271 BC)

Epicurus of Samos founded the Epicurean school of philosophy. Initially a Democritean, he overhauled Democritus' atomism so radically that his system was soon considered an independent one. He formed three Epicurean communities, the final one at Athens where his school, the Garden, became synonymous with Epicureanism. He and his three leading colleagues wrote voluminously, and their collective works became the school's canonical texts in later generations. Those of Epicurus alone amounted to 300 books (scrolls), including his seminal treatise On Nature. Most were long and technical, but he also composed short digests as an aide memoire*.*

1 Life

Although born on the Aegean island of Samos, Epicurus was the son of an Athenian, Neocles, and hence an Athenian citizen; this entitled him to own land at Athens, where he later established his principal school. While still in Samos he took lessons with a Platonist, Pamphilus, and much of his eventual philosophy can be seen as a reaction against Platonism. Equally formative was his period of study at Teos, although he violently disapproved of Nausiphanes, his Democritean teacher there, and may have been reacting against him in setting out to reshape the atomist tradition (see Atomism, ancient; Democritus). Epicurus famously professed to be self-taught. This was a denial that he had learnt anything from his teachers, Pamphilus and Nausiphanes, and he certainly never disguised his debt to Democritus and other forerunners. He also, in these formative years, developed an admiration for Pyrrho, later to become the patron saint of Scepticism, and may have acquired from his model of practical conduct the ideal of tranquillity which was to become the primary ethical goal in his own system (see Epicureanism §10).

At the age of 31 Epicurus set up his first school, in Mytilene on the island of Lesbos. Soon after that, he moved to Lampsacus (on the west coast of modern Turkey), where he established a second school. There he acquired many of the pupils who were to remain his closest associates for the remainder of his life, including Polyaenus, Metrodorus and his eventual successor Hermarchus. These communities continued to flourish even after 307 BC, when Epicurus and a number of his colleagues moved on to Athens to establish the Garden, thereafter the headquarters of Epicureanism, on a plot of land which he bought just outside the city walls.

2 Writings

Epicurus' writings were voluminous, but now only a few short works survive, along with fragments of others and a great mass of secondary testimonies. *On Nature*, of which fragments survive on papyrus, was his major treatise, written in thirty-seven books over many years, and covering much more than the physical topics which the title suggests. Books I-XIII, probably written before the move to Athens, covered the principal subject matter of physics. Books XIV-XV (dated 301-300 BC) seem to have contained a refutation of other physical theories. Another book (perhaps XXV) is devoted to psychology. Book XXVIII (dated 296 BC) discusses the philosophical role of language. These texts, although difficult and stylistically clumsy, show Epicurus actively at work on his philosophy, responding to opponents and colleagues and entertaining doubts.

Just four short works by Epicurus survive entire, preserved verbatim by the historian of philosophy Diogenes Laertius. The *Letter to Herodotus* is an epitome of *On Nature* I-XIII, on basic physics. The *Letter to Pythocles* epitomizes those parts of his work which dealt with astronomy, meteorology and kindred phenomena. The *Letter to Menoeceus* is an ethical epitome. Finally, the *Kyriai doxai* - 'Key Doctrines' - is a set of maxims on ethics and epistemology, apparently excerpted from individual works. The first four of these were held to sum up the aims of the system, and ran, in the abbreviated form known as 'the fourfold remedy' (*tetrapharmakos*): 'God presents no fears, death no worries. And while good is readily attainable, evil is readily endurable'. The object of these short digests was to aid Epicurus' pupils not so much in understanding his teachings as in memorizing them. They are nevertheless often the best evidence we have to go on. (Another preserved anthology of maxims, the *Vatican Sayings*, less structured than the *Kuriai doxai*, is someone's compilation from the works of Epicurus and other early school members.) None of these digests has a firm date, but there are reasons for placing the *Letter to Herodotus* around 305 BC, and the *Letter to Pythocles* shortly after.

Of his lost works, the most important seem to have been the *Canon*, on the principles of epistemology, and *On the End*, on ethics.

3 The school

In later generations, Epicureans worked from a set of school texts with virtually biblical status. These were the work of four founding fathers - 'the men' (*hoi andres*), as they were known in the school, meaning roughly 'the great men'. The infallible four were Epicurus, Metrodorus, Hermarchus and Polyaenus. Their texts were endlessly scrutinized in disputes about orthodoxy, and their correspondence, along with that of other early Epicureans, was anthologized and studied as a model of the philosophical life. Hermarchus is particularly notable for his detailed account of the origins of law, preserved by Porphyry (*On Abstinence* I 7-12), and Polyaenus was a former mathematician who, in joining the school, came to renounce all geometry as false (see Epicureanism §3).

The Epicurean community was a close-knit one, based on the cult of friendship and the communal study of philosophy, detached from political society without actively opposing it. Unusually for the time, women, children and even slaves and prostitutes participated on equal terms with men. Epicurus taught philosophy to his slave Mys, and Leontion was a prostitute who became a leading member of the group and wrote a work against Theophrastus.

In later generations Epicurus' portrait was ubiquitous, along with the often strikingly similar images of his co-founders. In the manner of the 'hero cults' of classical Greece, regular celebrations were held in their memory - a ritual which Epicurus had already instituted in his own lifetime in remembrance of Metrodorus, Polyaenus and others. It was a cardinal tenet that dead friends should be remembered and celebrated, never mourned. Aside from the feasting on these occasions, the Epicurean lifestyle was an ascetic one, belying the modern meaning of 'epicure'.

See also: Epicureanism

DAVID SEDLEY

References and further reading

Arrighetti, G. (1968) *Epicuro, opere*, Turin: Einaudi, 2nd edn, 1973.(At present the only comprehensive collection of Epicurus' surviving writings with Italian translation and commentary.)

Bailey, C. (1926) *Epicurus, The Extant Remains*, Oxford: Clarendon Press.(Unfortunately still the best English edition of Epicurus' main surviving writings; includes Greek text.)

Clay, D. (1983) 'Individual and Community in the First Generation of the Epicurean School', in ' *SUZHTHSIS. Studi sull'epicureismo greco e romano offerti a Marcello Gigante*, Naples: Macchiaroli, I, 255-79.

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Cambridge, MA: Loeb Classical Library, Harvard University Press and London: Heinemann, 1925, 2 vols.(Book X, in volume 2, contains the life of Epicurus, along with complete texts of his three letters and the *Kyriai doxai*.)

Frischer, B. (1982) *The Sculpted Word*, Berkeley, CA: University of California Press. (Intriguing investigation of Epicurean portraiture as a means of recruitment.)

Sedley, D. (1976) 'Epicurus and his Professional Rivals', in J. Bollack and A. Laks (eds) *Études sur l'epicurisme antique (Cahiers de Philologie I)*, 119-59.(On Epicurus' attitude to other philosophers.)

Epistemic logic

Modern treatment of epistemic logic began in the 1950s when some philosophers noticed (as scholastics had done before them) certain regularities in the logical behaviour of the concept of knowledge (for example, that knowing a conjunction is equivalent to knowing all its conjuncts) and began to systematize them. Initially these regularities were presented in the form of an axiomatic-deductive system, as in other branches of logic. Later, questions began to be asked concerning the model theory on which such an 'epistemic logic' is based. Still later, the concrete interpretation of this model theory has become an issue. In this way, gradually a bridge has begun to be forged from purely logical questions to such central epistemological questions as those concerning the objects of knowledge, different kinds (or even senses) of knowledge (and their interrelations), the intensional character of knowledge, the de dicto *versus* de re *distinction, and so on.*

1 Propositional epistemic logic

Epistemic logic began as a study of the logical behaviour of the expression 'knows that' (see Propositional attitude statements). One of the main aims of this study is to be able to analyse other constructions in terms of 'knows' by means of '*b* knows that'. This basic notion will be expressed in the notation used here by 'K_b'. (This symbolization is slightly misleading in that, in a formula of the form K_bS, the term '*b*' for the agent (knower) is intended to be outside the scope of '*K*', not inside, as the notation might suggest.) When the knower is the same throughout a sentence (or an argument), the '*b*' is often omitted.

The logic of '*K*' ('knows that') is based on certain simple, pragmatic and epistemological insights. First, what is the use of the notion of knowledge? In what way are you better off when you come to know that *S*? The basic answer is clear. When you know that *S*, you can legitimately omit from consideration all possibilities under which it is not the case that *S*; in other words, you can restrict your attention to the situations in which it is true that *S*. (What is meant by 'legitimately' here is a major philosophical problem. It is essentially tantamount to the well-known problem of defining knowledge (see Knowledge, concept of). For the purposes of epistemic logic, however, we do not have to solve this problem. The structural characterization just given suffices for our purposes, no matter how legitimacy is defined.)

This characterization amounts to making a distinction between two kinds of alternatives (situations; possible worlds) among all the relevant ones. Some possibilities are excluded by what *b* knows. They are the ones that can legitimately be disregarded. Others are not so excluded. They are the ones *b* will have to heed. It should be emphasized that the specification of such a set of admitted situations can be made independently of the psychological, cognitive or sociological peculiarities of the knower in question. Hence epistemic logic does not necessarily involve any subjective or even pragmatic element.

Thus we can see what kinds of structures are to be used in the model theory of epistemic logic (see Model theory). They will be called 'model structures'. They are classes of models of the underlying language minus *K*. Let us consider one of them, say, w_0, in which a potential knower, *b*, exists. Each such w_0 is associated with a number of other models ('possible worlds') w_1 which are not excluded by what *b* knows in w_0. They will be called the epistemic *b*-alternatives to w_0. Thus the basic truth-condition of epistemic logic can be formulated as follows:

K_bS is true in w_0 iff *S* is true in every epistemic *b*-alternative w_1 to w_0.

This definition is of course relative to a model structure. If the *b*-alternativeness relation is called R_b and Ω is the set of all possible worlds, the truth-condition can be expressed as follows:

K_bS is true in $w_0 \in \Omega$ iff *S* is true in each $w_1 \in \Omega$ such that $w_0R_bw_1$.

We can also introduce a notion of epistemic possibility, *P*. Then P_bS means, roughly, 'It is possible, for all that *b* knows, that *S*'. Obviously the following pairs of sentences are logically equivalent: $\sim K_bS$ and $P_b{\sim}S$; $\sim P_bS$ and $K_b{\sim}S$. The truth-condition for epistemic possibility is the obvious one:

P_bS is true in $w_0 \in \Omega$ iff *S* is true in some $w_1 \in \Omega$ such that $w_0R_bw_1$.

Normally it is required that each $w_0 \in \Omega$ be an alternative to itself (for each *b* in w_0). This amounts to requiring

that *b* can know (in w_0) that *S* only if it is true (in w_0) that *S*. Combined with the usual truth-conditions for propositional connectives, we can in this way obtain a semantic basis, complete with a truth-definition, for propositional epistemic logic.

The question of whether the alternativeness relation is transitive or not is equivalent to the question of the logical truth of sentences of the form:

(1) $(\forall x)(K_x S \supset K_x K_x S)$.

Unfortunately, in ordinary discourse, saying that one ‘knows that one knows’ has a number of different meanings, most of which have little to do with (1). Hence the question of whether ‘knowing implies that one knows that one knows’ needs a separate examination which falls only partially within the purview of epistemic logic.

On the basis of these semantic ideas, an epistemic propositional logic is easily developed. It admits of a complete Gentzen-style axiomatization. In such an axiomatization, the rules for ‘Čamp;’ and ‘ $\vee$’ are the usual cut-free ones, while no transfer of formulas between the left-hand side and the right-hand side is allowed. (For simplicity, it is assumed here and in the following that the formulas considered are in negation normal form, that is, that all negation signs in them are prefixed to atomic formulas.) The epistemic rules can then be formulated as follows:

$$\frac{\Gamma, K(S_1 \& S_2) \to \Delta}{\Gamma, KS_1, KS_2 \to \Delta}$$

$$\frac{\Gamma \to \Delta, P(S_1 \vee S_2)}{\Gamma \to \Delta, PS_1, PS_2}$$

$$\frac{S_1, S_2 \to S_3}{\Gamma, KS_1, PS_2 \to \Delta, PS_3}$$

$$\frac{S_1 \to S_2, S_3}{\Gamma, KS_1 \to \Delta, KS_2, PS_3}$$

We also need, of course, suitable structural rules and axioms. The following can serve as such axioms.

$S_1, {\sim}S_1 \to$
$\to S_1, {\sim}S_1$
$S \to S$

A game-theoretic truth-definition is better suited for the purposes of epistemic logic than an ordinary (Tarski-type) one (see Semantics, game-theoretic). In such a definition two players are involved: the proponent and the opponent. At each stage of the game, the two players are considering a sentence and a possible world. The game starts from the world w_0 in which a sentence *S* is to be evaluated. Such a game $G(S; w_0)$ is defined by the following rules:

(G.$\sim$) $G({\sim}S; w_0)$ is like $G(S; w_0)$ but with the roles of the two players reversed. (That is, the proponent of $G({\sim}S; w_0)$ is the opponent in $G(S; w_0)$, and vice versa.)
(G.&) $G((S_1 \ \& \ S_2); w_0)$ begins with a choice by the opponent of S_i ($i =$ 1 or 2) and the rest of the game is as in $G(S_i; w_0)$.
(G.$\vee$) $G((S_1 \vee S_2); w_0)$ likewise, except that the choice of S_1 or S_2 is made by the proponent.
(G.K) $G(K_b S; w_0)$ begins with a choice by the opponent of an epistemic *b*-alternative w_1 to w_0. The rest of the game is as in $G(S; w_1)$.
(G.P) $G(P_b S; w_0)$ likewise, except that the choice is made by the proponent.
(G.atom) Any play of a semantic game will come to an end in a finite number of moves with an atomic sentence or identity *A* and a model (world) *w*. Since the underlying language has been interpreted in *w*, *A* is either true or false. If the former, the proponent of *A* has won and the opponent lost. If the latter, vice versa.

The game-theoretic definition of truth and falsity says that *S* is true (false) in w_0 iff there exists a winning strategy for the proponent (opponent) in $G(S; w_0)$.

2 The semantics of quantified epistemic logic

When we move to *quantified* epistemic logic, we need individuals to serve as values of bound variables. Since we

are dealing with a multiplicity of models (worlds), individuals must in principle be identifiable in more than one of them. ('No entity without identity', to quote Quine.) Hence an individual is, model-theoretically speaking, a 'world line' connecting the roles or embodiments of one and the same individual in several worlds. Less metaphysically expressed, given a model system Ω, an individual is logically speaking a function (an 'individuating function') from models $w \in \Omega$ to the members of the domain $do(w)$ of *w* (see Intensional logics §2). (Since we want to keep our discussion as general as possible, we must allow individuating functions to be partial. In other words, an individual might not be identifiable in all possible worlds.)

A singular term likewise has as its value (bearer), in a given model *w*, a member of $do(w)$. However, these might not be the same individual in two different models. In other words, the values of a singular term *c* in two worlds might not be connected by a world line. The bearers of a singular term *c* in different models are said to be the values of the 'meaning function' associated with *c*. Individuating functions are a special case of meaning functions. Obviously, a meaning function is, like an individuating function, a (possibly partial) function from models $w \in \Omega$ to their respective domains $do(w)$. (In the interests of generality, a singular term must be allowed to be empty in some models.)

It cannot be assumed in epistemic logic that individuating functions are expressed in ordinary language by singular terms of any particular linguistic category, such as proper names, in the sense that the individuating function would coincide with the meaning function of such a 'rigid designator'. This impossibility is shown by the fact that one can meaningfully ask, using a proper name, for example, 'Do you know who John is?' or 'Who is John?' The idea that there are in natural language 'rigid designators' that express individuating functions hence does not work in epistemic contexts (see Reference §3).

On the basis of these considerations, we can now formulate the additional game rules needed in epistemic first-order logic:

($\mathrm{G}.\forall$) $G((\forall x)S(x); w_0)$ begins with the choice by the opponent of an individual from the domain of w_0. If it is stipulated that a new singular term *c* has the same meaning function as the identifying function of the chosen individual, the rest of the game is as in $G(S(c); w_0)$.

Even though it seems to restrict our treatment somewhat, we can also require that the chosen individual exists in w_0, that is, that its identifying function is defined for the argument w_0.

($\mathrm{G}.\exists$) $G((\exists x)S(x); w_0)$ likewise, except that the proponent makes the choice.

In the epistemic first-order logic so far formulated, many interesting logical relationships can be expressed. However, a fully satisfactory logic of the concept of knowledge can be obtained only by extending our language in one important respect. For one thing, the interrelationship between our formal language and knowledge statements in ordinary language is made much more perspicuous by this extension.

The new element in our epistemic language is a notation for informational independence, slightly extended. Moves are sometimes made by a player in ignorance of what has happened in certain specified earlier moves. Such moves are said to be independent of the earlier ones. This idea applies naturally to our semantic games. In them an application of the rules ($\mathrm{G}.\&$), ($\mathrm{G}.\vee$), ($\mathrm{G}.\forall$) or ($\mathrm{G}.\exists$) can be informationally independent of an earlier application of one of the rules (G.K) or (G.P). Actually, in much of epistemic logic, we need to consider only the possible independence of ($\mathrm{G}.\vee$) and ($\mathrm{G}.\exists$) of (G.K).

Independence will be indicated by a slash notation, for example, in ($\vee/K$) and ($\exists x/K$). However, the possibility of moving from one model to another necessitates a specification as to what the model *w* is in relation to which a move connected with, say, ($\exists x/K$) is made. Here the obvious stipulation is that it must be made in relation to the model (world) which the rule application connected with *K* moved us away from. In brief, the slash indicates that the moves connected with *K* and with the existential quantifier $\exists x$ are made in the reverse order.

These stipulations imply that there is a difference in meaning, for example, between sentences of the following forms: $K(\forall x)S(x)$ and $(\forall x)KS(x)$. In the former, the choice of the values of *x* is between the individuals existing in an epistemic alternative chosen by the opponent. In the latter, the choice is between individuals existing in the world in which the sentence is evaluated.

It is our reliance on the notion of informational independence that forces us to use game-theoretic semantics. The reason is that the logic of informational independence cannot be handled by means of the usual Tarski-type truth-definitions.

By means of the slash notation, most of the usual types of knowledge, and most of the logical laws holding for these different types of knowledge, can be formulated. The value of a quantifier such as '($\exists x/K_b$)' is chosen independently of the epistemic alternative selected by the opponent in applying the rule (G.K). Hence the value of *x* must be the same in all the situations compatible with what *b* knows, that is, must be known to *b*. Yet this value may depend on that of a universal quantifier which depends on *K*, wherefore no linear ordering can capture the dependence relations in question. Thus we have, for instance:

(2) ' b knows whether S' $= K_b(S(\vee/K_b){\sim}S)$
(3) ' b knows whether S_1 or S_2' $= K_b(S_1(\vee/K_b)S_2)$
(4) ' b knows who satisfies the condition $S(x)$' $= K_b(\exists x/K_b)S(x)$
(5) ' b knows of everybody whether they satisfy the condition $S(x)$' $= K_b(\forall x)(S(x)(\vee/K_b){\sim}S(x))$
(6) ' b knows to whom each person bears the relation $S(x, y)$' $= K_b(\forall x)(\exists y/K_b)S(x, y)$.

Even more complex types of knowledge can be expressed in our notation. In these examples, it is assumed that the range of quantifiers consists of persons.

Of (2)-(6), the first three have equivalents in a slash-free notation, namely:

(7) $K_bS \vee K_b{\sim}S$
(8) $K_bS_1 \vee K_bS_2$
(9) $(\exists x)K_bS(x)$.

However, (5) and (6) do not have slash-free equivalents. For instance, (6) is not equivalent to:

(10) $(\forall x)(\exists y)K_bS(x, y)$.

This is seen from the following second-order equivalents to (6) and (10), respectively:

(11) $(\exists f)K_b(\forall x)S(x, f(x))$
(12) $(\exists f)(\forall x)K_bS(x, f(x))$.

This reduction of a large number of other constructions with 'knows' to the 'knows that' construction is evidence for the propositional character of knowledge.

3 Rules for quantified epistemic logic

In terms of the slash notation, deductive rules for quantified epistemic logic can be formulated.

First, the vaunted intensionality of knowledge can be understood and accounted for without postulating any special intensional entities to serve as 'objects of knowledge' (see Intensional entities). Intensionality is a matter of how meaning functions and identifying functions behave. It means that an identity which holds in one situation (model) need not hold in another one. In other words, we cannot have in our epistemic logic an unrestricted rule of the following form:

(13) $$\frac{\Gamma, S(a), a = b, S(b) \to \Delta}{\Gamma, S(a), a = b \to \Delta}$$

We can have such a rule only when *S* is required not to contain any intensional operators.

Second, rules for existential generalization and universal instantiation have to be changed. For instance, we cannot have the rule:

(14) $$\frac{\Gamma, (\forall x)S(x), S(a) \to \Delta}{\Gamma, (\forall x)S(x) \to \Delta}$$

for this inference presupposes that *a* is the same individual in all the relevant models. If *x* occurs in $S(x)$ within

the scope of only one epistemic operator, say, *K* (while $S(x)$ is in negation normal form), this rule can be restored by means of an additional premise.

(15) $$\frac{\Gamma, (\forall x)S(x), K(\exists y/K)(a = y), S(a) \rightarrow \Delta}{\Gamma, (\forall x)S(x), K(\exists y/K)(a = y) \rightarrow \Delta}$$

This modification of the usual rules is eminently natural both ‘intuitively’ and model-theoretically. If I know that *a* did something, it does not follow that I know who did it, unless I know who *a* is. At the same time, the need of the extra clause shows that in a sense the objects of knowledge are not world-bound individuals, but individuals that can be identified in several possible situations. If this is what is meant by saying that knowledge is intentional, the saying is true.

Third, rules for existential instantiation and universal generalization also have to be changed. They no longer do their whole job. We cannot have merely:

(16) $$\frac{\Gamma, S(a) \rightarrow \Delta}{\Gamma, (\exists x)S(x) \rightarrow \Delta}$$

where *a* does not occur anywhere in $S(x)$, Γ or Δ. Instead, we need a stronger rule of the form:

(17) $$\frac{\Gamma, S_0(S_1(f(y_1, y_2, \ldots))) \rightarrow \Delta}{\Gamma, S_0((\exists x)S_1(x)) \rightarrow \Delta}$$

where S_0 is in negation normal form and $(\forall y_1), (\forall y_2), \ldots$ are all the universal quantifiers within the scope of which $(\exists x)S_1(x)$ occurs in $S_0 = S_0((\exists x)S_1(x))$, and where *f* is a function symbol which does not occur anywhere in S_0, Γ or Δ. Clearly (16) can be thought of as a special case of (17). Parallel sequent rules have to be modified analogously.

Fourth, negation behaves in quantified epistemic logic in an unwonted way. For instance, the law of excluded middle no longer holds. This is almost predictable on the basis of our definitions of truth and falsity in §1 above. They imply that the law of excluded middle is a determinacy assumption, that is, that it says that one or the other player always has a winning strategy in one of our epistemic games. From game theory it is known that such determinacy assumptions are very strong, and hence likely to fail.

In general, quantified epistemic logic thus exhibits several phenomena that have no counterpart in ordinary first-order logic. In spite of the failure of the law of excluded middle, one can formulate a complete disproof procedure for epistemic logic. It can also be shown that many standard metalogical results, such as compactness, and the separation and Löwenheim-Skolem theorems, can be extended to epistemic logic.

4 Wider perspectives

The epistemic logic outlined here can be extended and applied in several different directions.

First, epistemic logics have proved relevant to such branches of computer science as artificial intelligence and database theory. A growing number of results have been reached by scientists working in this area.

Second, an interpretational rule (game rule) must be associated also with nonlogical constants of different kinds, for instance, with singular terms such as names. Applications of such rules can be independent of epistemic rules, which create distinctions between sentences such as the following:

(18) $K_b S(d)$
(19) $K_b S((d/K))$

This distinction is nothing but the so-called *de re-de dicto* distinction (see De re/de dicto). This famous distinction turns out to be another independence phenomenon and not an irreducible contrast between two kinds of knowledge nor two kinds of reference.

Third, epistemic logic is the cornerstone of the logic of knowledge-seeking developed by Jaakko Hintikka and his associates and called by them the interrogative model of inquiry (Hintikka *et al.* 1997). In fact, we obtain the rules

for such an inquiry by reversing Gentzen-type rules for epistemic logic and adding to them the following additional rule:

$$\text{(R.Q)} \quad \frac{\Gamma, S_0 \rightarrow \Delta}{\Gamma, S_0, S_1 \rightarrow \Delta}$$

where S_0 is the presupposition of any one question and S_1 the desideratum of the same question. In such interrogative inquiry, it must also be specified which questions are answerable. The applicability of (R.Q) then presupposes the answerability of the question.

Fourth, in epistemic contexts quantifiers have been shown to depend on a method of identifying individuals as members of different possible situations. Now it can be seen that individuals can be identified in two different ways. Then we have two sets of identifying functions and, corresponding to them, two pairs of quantifiers. The rules governing each of them separately are the same as have been explained. In one kind of identification, the knower uses their direct cognitive relations to objects (for example, their visual space) as a framework of identification. Such identification has been called 'perspectival'. In ordinary language, perspectival identification shows up in the form of the direct-object construction with 'knows', for example, 'Jack knows Jill', as distinguished from 'Jack knows who Jill is'. In the other, an impersonal (public) framework of identification is relied on. It can be shown that this duality is indeed realized in our actual conceptual practice with epistemic concepts. This results in an extremely important application of epistemic logic. In this application, questions concerning the interaction of the two kinds of quantifiers belong to the province of epistemology and cognitive science rather than epistemic logic.

This distinction between two kinds of identification is approximately instantiated by Bertrand Russell's distinction between knowledge by description and knowledge by acquaintance (see Knowledge by acquaintance and description), by neuroscientists' distinction between the what-system and where-system in cognition, by Endel Tulving's distinction between semantic and episodic memory, and so on. It is important to realize that these distinctions are not at bottom distinctions between two kinds of knowledge (or memory-knowledge), or between two kinds of objects within some one world, but between two kinds of principles of identification of individuals from one world to another. It is nevertheless tempting and even natural to speak of 'two kinds of objects' here. For instance, so-called phenomenological objects (including Russell's 'objects of acquaintance') are in effect individuals identified in a certain way.

Fifth, in natural language we are dealing with more than just one domain of individuals. Rather, we have at the very least a different domain corresponding to each different 'wh- question'. This distinction between different ranges of quantifiers (for that is what wh- words in effect express) is closely related to Aristotle's distinction between different categories.

Finally, a model-theoretic treatment of knowledge seems to lead to the paradox of logical omniscience, that is, to the consequence that everyone knows all the logical consequences of what they know. One natural explanation is that in their actual thinking humans articulate the possible worlds they are tacitly considering only to a certain depth. This notion of a world articulated only to a given depth can be captured by means of Rantala's (1975) notion of urn models, which hence offers a natural way out of the paradox.

See also: Logic and mathematical terms, glossary of

JAAKKO HINTIKKA
ILPO HALONEN

References and further reading

Boh, I. (1993) *Epistemic Logic in the Later Middle Ages*, London: Routledge.(The prehistory of epistemic logic.)

Fagin, R., Halpern, J.Y., Moses, Y. and Vardi, Moshe Y. (1955) *Reasoning about Knowledge*, Cambridge, MA: MIT Press.(Synthesis of computer scientists' work on epistemic logic.)

Hintikka, J. (1962) *Knowledge and Belief*, Ithaca, NY: Cornell University Press.(First fully fledged treatment of the logic of knowledge.)

Hintikka, J. (1992) 'Independence-Friendly Logic as a Medium of Information Representation and Reasoning About Knowledge', in S. Ohsuga *et al*. (eds) *Information, Modelling and Knowledge Bases*, Amsterdam: IOS Press, vol. 3, 258- 65.(Develops the themes of §§2-3 of this entry.)

Hintikka, J. (1996) ‘Knowledge Acknowledged: Knowledge of Propositions vs. Knowledge of Objects’, *Philosophy and Phenomenological Research* 56: 251-75.(Knowledge of truths and knowledge of objects distinguished from each other.)

Hintikka, J., Halonen, I. and Mutanen, A. (1997) ‘Interrogative logic as a general theory of reasoning’, in R. Johnson and J. Woods (eds) *Handbook of Applied Logic*, Dordercht: Kluwer.(Survey of the uses of epistemic logic in knowledge acquisition by questioning.)

Hintikka, J. and Hintikka, M.B. (1989) *The Logic of Epistemology and the Epistemology of Logic*, Dordrecht: Kluwer.(Essay 2 is an alternative survey of epistemic logic; essay 4 applies epistemic logic to the problem of the objects of knowledge; essay 5 treats the problem of logical omniscience; and essay 8 explores the connections between different modes of identification and their manifestations in cognitive science.)

Hintikka, J. and Sandu, G. (1989) ‘Informational Independence as a Semantical Phenomenon’, in J.E. Fenstad, I.T. Frolov and R. Hilpinen (eds) *Logic, Methodology and Philosophy of Science*, Amsterdam: North Holland, vol. 8, 571-89.(Introduces the idea of informational independence into logical semantics.)

Lenzen, W. (1978) *Recent Work in Epistemic Logic*, Acta Philosophica Fennica, vol. 30, number 1, Helsinki: Societas Philosophica Fennica.(Survey of the literature up to 1978.)

Rantala, V. (1975) ‘Urn Models’, *Journal of Philosophical Logic* 4: 455-74.(Urn models defined.)

Wright, G.H. von (1951) *An Essay in Modal Logic*, Amsterdam: North Holland.(Von Wright began the first systematic treatment of epistemic logic.)

Epistemic relativism

An account of what makes a system of reasoning or belief revision a good one is relativistic if it is sensitive to facts about the person or group using the system. It may then turn out that one system is best for one person or group, while a quite different system is best for another. Some of the most popular accounts of how systems of reasoning are to be assessed, including those based on reflective equilibrium and those based on the system's truth-generating capacity, appear to be relativistic. It is sometimes claimed that epistemic relativism leads to nihilism or that it severs the connection between good reasoning and true belief.

1 Relativism defined

The term 'epistemic relativism' has been used in a bewildering variety of ways. Here, we focus on an account that takes epistemic relativism to be a species of normative cognitive pluralism (see Cognitive pluralism). Normative cognitive pluralism claims that there is no unique system of reasoning (or of forming and revising beliefs) that people ought to use, because various quite different systems can all be equally good. An account of what makes a system of reasoning a good one is relativistic if the assessments of cognitive systems it offers are sensitive to facts about the person or group using the system. If systems of reasoning are evaluated in this way, then in general it will make no sense to ask whether one system is better than another: rather, we must ask whether one system is better than another for a given person or group.

2 Two relativistic accounts of cognitive assessment

Though it often goes unnoticed, some of the most popular accounts of how systems of reasoning are to be assessed are, or at least might well turn out to be, relativistic. Here, two such accounts are considered: one based on reflective equilibrium, the other based on a system's truth-generating capacity.

Nelson Goodman claimed that general principles of inference were justified by their conformity with the particular inferences we make and accept, and that our acceptance of particular inferences was justified by their accord with general inferential principles. This, he noted, looked 'flagrantly circular' but, he continued:

> this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either.
>
> (Goodman 1965: 64)

John Rawls (1971) introduced the term 'reflective equilibrium' to label the endpoint of the process of 'delicate … mutual adjustments' that Goodman describes.

Although Goodman did not discuss the matter, other authors have noted that there is no guarantee that everyone who uses the process will end up at the same point. If two people begin with significantly different judgments rejecting or accepting particular inferences, or with different views about which rules they are willing to amend (or both), then it seems entirely possible that they will end up with quite different sets of rules, though each set will be in reflective equilibrium. If, as Goodman insists, the process of mutual adjustment is all that is needed for rules and inferences to be justified, then these people may end up reasoning in very different ways, each of which is justified for the person who reasons in that way.

Reliabilist accounts of how to assess systems of reasoning or belief revision link the assessment to the truth-generating capacity of the system (see Reliabilism). Other things being equal, the better a system is at producing true beliefs and avoiding false one, the more highly a reliabilist will rank it. Though it is not often emphasized by reliabilists, this sort of assessment is quite sensitive to the environment in which people using the system find themselves. Thus it may well turn out that a given system of reasoning does an excellent job for one person and a very poor job for another. Imagine a pair of people who suddenly fall victim to Descartes' demon, and are from that time provided with systematically misleading or deceptive perceptual data. Suppose that one of the victims has been using cognitive processes quite like our own, and that these have done a good job in

generating truths and avoiding falsehoods, while the other victim's cognitive processes have been (by our lights) quite mad, and have produced far more falsehoods and far fewer truths. In their new demon-infested environment, however, the 'normal' system of cognitive processes will yield a growing fabric of false beliefs. The other system, by contrast, may now do a much better job at generating truths and avoiding falsehoods, since what the evil demon is doing is providing his victims with radically misleading evidence - evidence that only a lunatic would take to be evidence for what actually is the case. So on an account of cognitive evaluation in which generating truths and avoiding falsehoods plays a central role, our system would be preferable in one environment, the mad system in another. Which system a person ought to use will depend on which environment the person is in.

Invocation of evil demons to make the point might suggest that this is a very peripheral phenomenon that is hardly worth worrying about. However, the Cartesian demon case is just the very small tip of a very large iceberg. Any reliabilist evaluation of cognitive processes is going to be acutely sensitive to the cultural, technological and epistemic setting in which the processes are to function. The likelihood that one system of cognitive processes will do a better job than another at generating truth, I suspect, will depend on such factors as the existence of a system of writing, the existence and the structure of disciplinary communities, and the relation of those communities to the political and economic arrangements of the wider society. It will also often depend on the level of conceptual, mathematical, scientific and technological sophistication that has been achieved. If these conjectures are right, it follows that reliabilist accounts of cognitive or epistemic evaluation will have a certain post-Hegelian historicist flavour. There will be no one ideal method of inquiry, no cognitive system that excels in all historical settings. Rather, we can expect that the assessment of a cognitive system will vary as its historical setting varies, and that, just as with technologies (and indeed with genes), it will sometimes happen that a successful system will undermine its own success by changing the environment in such a way that competing systems will now be more successful.

3 Is epistemic relativism problematic?

Many philosophers consider epistemic relativism a dangerous or troubling doctrine. It is, however, not easy to find plausible arguments justifying this negative attitude. This section briefly sketches two lines of argument that might motivate opposition to relativism, although I do not think either argument very persuasive.

The first charge against relativism is that it is nihilistic because it simply gives up on the project of distinguishing good reasoning from bad, and embraces a sort of epistemic anarchy. From our previous discussion, however, it should be clear that the 'anything goes' slogan is a singularly inappropriate one for many relativistic accounts of cognitive assessment. Many versions of reliabilism are relativistic. But reliabilists are certainly not epistemic anarchists - quite the contrary. Reliabilism offers an extremely demanding account of cognitive evaluation. For a given cognitive agent in a given historical setting, it will typically be the case that a reliabilist evaluation will rank one system of reasoning higher than another. Rarely will it be the case that reliabilism ranks all contenders on a par.

A second complaint against relativism is that it threatens the connection between cognitive inquiry and truth. For if the epistemic relativist is right, then there may be a pair of people whose systems of reasoning are very different from one another, though each system is optimal for the person using it. We can expect that on being exposed to essentially the same data these people will sometimes end up with very different sets of beliefs. When this happens it is unlikely to be the case that both sets are true; at least one set of beliefs will be substantially mistaken. Since at least one person will end up with false beliefs, and since *ex hypothesi* they are both using optimally good cognitive systems, it can not be the case that good cognition always leads to true beliefs.

What this argument shows is that if the epistemic relativist is right, then good reasoning does not guarantee truth. But it does not show that good reasoning and truth are unconnected. If, for example, we adopt a reliabilist account of cognitive evaluation, then people who reason well will do the best job possible at producing truths and avoiding falsehoods. To expect more than this seems unreasonable.

See also: Rational beliefs; Relativism

STEPHEN P. STICH

References and further reading

Daniels, N. (1980) ‘Wide Reflective Equilibrium and Archimedean Points’, *Canadian Journal of Philosophy* 10: 83-103.(Discussion of the virtues and shortcomings of reflective-equilibrium accounts of justification.)
Foley, R. (1987) *The Theory of Epistemic Rationality*, Cambridge, MA: Harvard University Press.(A sophisticated statement of an epistemic position that has some features of relativism.)
Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA: MIT Press. (Important statement of reliabilism.)
Goodman, N. (1965) *Fact, Fiction and Forecast*, Indianapolis, IN: Bobbs-Merrill. (Referred to in §2 above. The classic statement of a reflective-equilibrium account of justification.)
Hollis, M. and Lukes, S. (eds) (1982) *Rationality and Relativism*, Cambridge, MA: MIT Press.(Collection of essays by philosophers, anthropologists and historians of science debating the evidence for descriptive pluralism and the merits of relativism.)
Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(Classic work in which reflective equilibrium is used to defend an account of justics.)
Stein, E. (1996) *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science*, Oxford: Clarendon Press.(Systematic study of the empirical and philosophical literature relevant to relativism.)
Stich, S. (1990) *The Fragmentation of Reason*, Cambridge, MA: MIT Press.(Expansion of the material of this entry; see especially chapters 1, 4 and 6.)

Epistemology

Epistemology is one of the core areas of philosophy. It is concerned with the nature, sources and limits of knowledge (see Knowledge, concept of). There is a vast array of views about those topics, but one virtually universal presupposition is that knowledge is true belief, but not mere true belief (see Belief and knowledge). For example, lucky guesses or true beliefs resulting from wishful thinking are not knowledge. Thus, a central question in epistemology is: what must be added to true beliefs to convert them into knowledge?

1 The normative answers: foundationalism and coherentism

The historically dominant tradition in epistemology answers that question by claiming that it is the quality of the reasons for our beliefs that converts true beliefs into knowledge (see Epistemology, history of). When the reasons are sufficiently cogent, we have knowledge (see Rational beliefs). This is the normative tradition in epistemology (see Normative epistemology). An analogy with ethics is useful: just as an action is justified when ethical principles sanction its performance, a belief is justified when epistemic principles sanction accepting it (see Justification, epistemic; Epistemology and ethics). The second tradition in epistemology, the naturalistic tradition, does not focus on the quality of the reasons for beliefs but, rather, requires that the conditions in which beliefs are acquired typically produce true beliefs (see Internalism and externalism in epistemology; Naturalized epistemology).

Within the normative tradition, two views about the proper structure of reasons have been developed: foundationalism and coherentism (see Reasons for belief). By far, the most commonly held view is foundationalism. It holds that reasons rest on a foundational structure comprised of ‘basic’ beliefs (see Foundationalism). The foundational propositions, though justified, derive none of their justification from other propositions. (Coherentism, discussed below, denies that there are foundational propositions).

These basic beliefs can be of several types. Empiricists (such as Hume and Locke) hold that basic beliefs exhibit knowledge initially gained through the senses or introspection (see A posteriori; Empiricism; Introspection, epistemology of; Perception, epistemic issues in). Rationalists (such as Descartes, Leibniz and Spinoza) hold that at least some basic beliefs are the result of rational intuition (see A priori; Rationalism). Since not all knowledge seems to be based on sense experience, introspection or rational intuition, some epistemologists claim that some knowledge is innate (see Innate knowledge; Knowledge, tacit; Kant, I.; Plato). Still others argue that some propositions are basic in virtue of conversational contextual features. That is, some propositions are taken for granted by the appropriate epistemic community (see Contextualism, epistemological).

Foundationalists hold that epistemic principles of inference are available which allow an epistemic agent to reason from the basic propositions to the non-basic (inferred) propositions. They suggest, for example, that if a set of basic propositions is explained by some hypothesis and additional confirming evidence for the hypothesis is discovered, then the hypothesis is justified (see Inference to the best explanation). A notorious problem with this suggestion is that it is always possible to form more than one hypothesis that appears equally well confirmed by the total available data, and consequently no one hypothesis seems favoured over all its rivals (see Induction, epistemic issues in; Goodman, N.). Some epistemologists have argued that this problem can be overcome by appealing to features of the rival hypotheses beyond their explanatory power. For example, the relative simplicity of one hypothesis might be thought to provide a basis for preferring it to its rivals (see Simplicity (in scientific theories); Theoretical (epistemic) virtues).

In contrast to foundationalism, coherentism claims that every belief derives some of its justification from other beliefs (see Knowledge and justification, coherence theory of; Probability theory and epistemology; Bosanquet, B.; Bradley, F.H.). All coherentists hold that, like the poles of a tepee, beliefs are mutually reinforcing. Some coherentists, however, assign a special justificatory role to those propositions that are more difficult to dislodge because they provide more support for the other propositions and are more supported by them. The set of these special propositions overlaps the set of basic propositions specified by foundationalism.

There are some objections aimed specifically at foundationalism and others aimed specifically at coherentism. But there is one deep difficulty with both traditional normative accounts. This problem, known as the ‘Gettier

Problem' (after a famous three-page article by Edmund Gettier in 1963), can be stated succinctly as follows (see Gettier problems): suppose that a false belief can be justified (see Fallibilism), and suppose that its justificatory status can be transferred to another proposition through deduction or other principles of inference (see Deductive closure principle). Suppose further that the inferred proposition is true. If these suppositions can be true simultaneously - and that seems to be the case - the inferred proposition would be true, justified (by either foundationalist or coherentist criteria) and believed, but it clearly is not knowledge, since it was inferred from a false proposition. It is a felicitous coincidence that the truth was obtained.

One strategy for addressing the Gettier Problem remains firmly within the normative tradition. It employs the original normative intuition that it is the quality of the reasons that distinguishes knowledge from mere true belief. This is the defeasibility theory of knowledge. There are various defeasibility accounts but, generally, all of them hold that the felicitous coincidence can be avoided if the reasons which justify the belief are such that they cannot be defeated by further truths (see Knowledge, defeasibility theory of).

2 The naturalistic answers: causes of belief

There is a second general strategy for addressing the Gettier Problem that falls outside of the normative tradition and lies squarely within the naturalistic tradition (see Quine, W.V.). As the name suggests, the naturalistic tradition describes knowledge as a natural phenomenon occurring in a wide range of subjects. Adult humans may employ reasoning to arrive at some of their knowledge, but the naturalists are quick to point out that children and adult humans arrive at knowledge in ways that do not appear to involve any reasoning whatsoever. Roughly, when a true belief has the appropriate causal history, then the belief counts as knowledge (see Knowledge, causal theory of).

Suppose that I am informed by a reliable person that the temperature outside the building is warmer now than it was two hours ago. That certainly looks like a bit of knowledge gained and there could be good reasons provided for the belief. The normativists would appeal to those good reasons to account for the acquisition of knowledge. The naturalists, however, would argue that true belief resulting from testimony from a reliable source is sufficient for knowledge (see Social epistemology; Testimony).

Testimony is just one reliable way of gaining knowledge (see Reliabilism). There are other ways such as sense perception, memory and reasoning. Of course, sometimes these sources are faulty (see Memory, epistemology of). A central task of naturalized epistemology is to characterize conditions in which reliable information is obtained (see Information theory and epistemology). Thus, in some of its forms, naturalized epistemology can be seen as a branch of cognitive psychology, and the issues can be addressed by empirical investigation.

Now let us return to the Gettier Problem. Recall that it arose in response to the recognition that truth might be obtained through a felicitous coincidence. The naturalistic tradition ties together the belief and truth conditions of knowledge in a straightforward way by requiring that the means by which the true belief is produced or maintained should be reliable.

3 Scepticism

The contrast between normative and naturalized epistemology is apparent in the way in which each addresses one of the most crucial issues in epistemology, namely, scepticism (see Scepticism). Scepticism comes in many forms. In one form, the requirements for knowledge become so stringent that knowledge becomes impossible, or virtually impossible, to obtain. For example, suppose that a belief is knowledge only if it is certain, and a belief is certain only if it is beyond all logically possible doubt. Knowledge would then become a very rare commodity (see Certainty; Doubt).

Other forms of scepticism only require good, but not logically unassailable, reasoning. We have alluded to scepticism about induction. That form of scepticism illustrates the general pattern of the sceptical problem: there appear to be intuitively clear cases of the type of knowledge questioned by the sceptic, but intuitively plausible general epistemic principles appealed to by the sceptic seem to preclude that very type of knowledge.

Another example will help to clarify the general pattern of the sceptical problem. Consider the possibility that my brain is not lodged in my skull but is located in a vat and hooked up to a very powerful computer that stimulates it to have exactly the experiences, memories and thoughts that I am now having. Call it the 'sceptical hypothesis'.

That hypothetical situation is clearly incompatible with the way I think the world is. Now, it seems to be an acceptable normative epistemic principle that if I am justified in believing that the world is the way I believe it to be (with other people, tables, governments and so on), I should have some good reasons for denying the sceptical hypothesis. But, so the argument goes, I could not have such reasons; for if the sceptical hypothesis were true, everything would appear to be just as it now does. So, there appears to be a conflict between the intuition that we have such knowledge and the intuitively appealing epistemic principle. Thus, scepticism can be seen as one instance of an interesting array of epistemic paradoxes (see Paradoxes, epistemic).

Of course, epistemologists have developed various answers to scepticism. Within the normative tradition, there are several responses available. One of them is simply to deny any epistemic principle - even if it seems initially plausible - that precludes us from having what we ordinarily think is within our ken (see Commonsensism; Chisholm, R.M.; Moore, G.E.; Reid, T.). Another response is to examine the epistemic principles carefully in an attempt to show that, properly interpreted, they do not lead to scepticism. Of course, there is always the option of simply declaring that we do not have knowledge. Whatever choice is made, some initially plausible intuitions will be sacrificed.

Within the naturalistic tradition, there appears to be an easy way to handle the sceptical worries. Possessing knowledge is not determined by whether we have good enough reasons for our beliefs but, rather, whether the processes that produced the beliefs in question are sufficiently reliable. So, if I am a brain in a vat, I do not have knowledge; and if I am not a brain in a vat (and the world is generally the way I think it is), then I do have knowledge. Nevertheless, those within the normative tradition will argue that we are obliged to withhold full assent to propositions for which we have less than adequate reasons, regardless of the causal history of the belief.

4 Recent developments in epistemology

Some recent developments in epistemology question and/or expand on some aspects of the tradition. Virtue epistemology focuses on the characteristics of the knower rather than individual beliefs or collections of beliefs (see Virtue epistemology). Roughly, the claim is that when a true belief is the result of the exercise of intellectual virtue, it is, ceteris paribus, knowledge. Thus, the virtue epistemologist can incorporate certain features of both the normative and naturalist traditions. Virtues, as opposed to vices, are good, highly prized dispositional states. The intellectual virtues, in particular, are just those deep dispositions that produce mostly true beliefs. Such an approach reintroduces some neglected areas of epistemology, for example, the connection of knowledge to wisdom and understanding (see Wisdom).

In addition, there are emerging challenges to certain presuppositions of traditional epistemology. For example, some argue that there is no set of rules for belief acquisition that are appropriate for all peoples and all situations (see Cognitive pluralism; Epistemic relativism). Others have suggested that many of the proposed conditions of good reasoning, for example ‘objectivity’ or ‘neutrality’, are not invoked in the service of gaining truths, as traditional epistemology would hold, but rather they are employed to prolong entrenched power and (at least in some cases) distort the objects of knowledge (see Feminist epistemology).

In spite of these fundamental challenges and the suggestions inherent in some forms of naturalized epistemology that the only interesting questions are empirically answerable, it is clear that epistemology remains a vigorous area of inquiry at the heart of philosophy.

See also: Charity, principle of; Criteria; Epistemic logic; Hermeneutics; Phenomenalism; Phenomenology, epistemic issues in; Rorty, R.M.; Solipsism

PETER D. KLEIN

References and further reading

Chisholm, R. (1966/1977/1989) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 1st, 2nd and 3rd edns.(The successive editions contain a general introduction to many issues in epistemology and increasingly complex foundationalist accounts of knowledge, along with versions of the defeasibility account. The first edition is a good place to begin a study of contemporary epistemology.)

Gettier, E. (1963) ‘Is justified true belief knowledge?’, *Analysis* 23 (6):121-3.(This article was responsible for

focusing attention on the inadequacy of characterizing knowledge as true, justified belief. Many of the most interesting contemporary issues in epistemology can be traced directly or indirectly to this article.)

Lehrer, K. (1990) *Theory of Knowledge*, Boulder, CO: Westview Press.(An accessible introduction to the fundamental questions in epistemology that defends a version of coherentism as supplemented by the defeasibility account.)

Lucey, K. (1996) *On Knowing and the Known*, Buffalo, NY: Prometheus Books.(A collection of contemporary, fairly accessible articles on a wide variety of epistemic issues.)

Moser, P. and Vander Nat, A. (eds) (1987) *Human Knowledge: Classical and Contemporary Approaches*, New York and Oxford: Oxford University Press.(An accessible collection of classical and contemporary essays on a wide variety of issues in epistemology.)

Sosa, E. (ed.) (1994) *Knowledge and Justification*, vols 1 and 2, Brookfield, VT: Ashgate Publishing Company.(A comprehensive set of contemporary essays in epistemology.)

Epistemology and ethics

Epistemology and ethics are both concerned with evaluations: ethics with evaluations of conduct, epistemology with evaluations of beliefs and other cognitive acts. Of considerable interest to philosophers are the ways in which the two kinds of evaluations relate to one another. Philosophers' explorations of these relations divide into two general categories: examination of potential analogies between the two fields, and attempts to identify necessary or conceptual connections between the two domains.

There is little doubt that there are at least superficial similarities between ethics and epistemology: one might say that ethics is about the appraisal of social behaviour and agents, while epistemology is about the appraisal of cognitive acts and agents. On the other hand, the widely held view that behaviour subject to moral evaluation is free and voluntary while beliefs are not, suggests one important disanalogy between the two fields.

1 Epistemic and ethical evaluations

People regularly evaluate actions as right or wrong, justified or unjustified, obligatory or prohibited, and they evaluate individuals as good or bad, virtuous or immoral. Ethics is the branch of philosophy that studies evaluations of these sorts. While ethicists do at times evaluate particular acts and agents, their concern is more typically with the general principles governing these evaluations and with the meaning and nature of the evaluations themselves. People make comparable judgments about beliefs and other cognitive acts, sometimes using the same evaluative language. They say that beliefs are justified or unjustified, that someone ought or ought not to believe a certain proposition. They also evaluate people as reasonable or unreasonable, as rational or irrational. Epistemology is the branch of philosophy that studies these sorts of evaluations of cognitive acts. Like ethicists, epistemologists are usually primarily concerned with the general principles governing epistemic evaluations and with their meaning and nature rather than with the status of particular beliefs or believers.

The evaluative nature of the central concepts of ethics and epistemology gives rise to intriguing philosophical questions. To say that an action is right, for example, seems to be to endorse or approve of the action, and it also seems to say something descriptive about the action, something that could be true or false. This suggests that there is a property or characteristic of actions, rightness, just as there are other causal, spatial or temporal properties of objects. But what sort of property is moral rightness? What connection does it have to the properties to which the natural sciences appeal? Are there universal standards of moral rightness or are standards in some sense determined individually or culturally? Can a person have an obligation to perform an action even if they do not see anything of value in the action? Analogous questions arise in epistemology. When we say that a belief is justified or that a body of evidence is good evidence for some conclusion, we seem to suggest that there is a real property - being justified - and a real relation - being good evidence for something. But what sort of property and relation are these? What connection do they have to the properties and relations of the natural sciences? Are there universal standards of justification and rationality or are standards determined individually or culturally? Can a person have good reasons to believe something even if they do not see the evidential merit of those reasons? Many philosophers have thought that there are important analogies between ethics and epistemology and that the answers to these questions in one domain may shed light on the analogous questions in the other.

2 Naturalism

There has been considerable discussion concerning the connections of ethics and epistemology to the sciences. The debate has often focused on the plausibility of doctrines known as 'naturalism'. There are at least two distinct issues taken up under this heading, one of which has received a great deal of attention from ethicists while the other has been more prominent in epistemology. The discussion of naturalism in ethics has concentrated on the plausibility of giving naturalistic definitions of ethical terms. The discussion of naturalism in epistemology has focused on the connection between empirical studies of cognition and epistemological theorizing.

Naturalistic definitions of ethical terms purport to define - or reduce - the concepts or terms of ethics to those of science. To make sense of reductionism, we first must acknowledge a distinction between two categories of properties or concepts: the evaluative properties of ethics on the one hand and the allegedly more fundamental properties of the sciences. These latter would, perhaps, include causal, structural, spatial and temporal properties.

Reductionists about ethics think that the descriptive role of moral predicates - for example 'right' - is to express moral properties - the property 'moral rightness' - and that these moral properties are simply complexes of basic scientific properties. Further, they think that it is a conceptual or philosophical job to say which complexes of natural properties moral properties reduce to. To take a simple (and not very plausible) example, an ethicist might hold that the property of being a right action is the same as the property of having the approval of the majority of people in its agent's society. This amounts to a proposed reduction of the ethical notion of a right action to a social and psychological notion of majority approval. Reductionists need not say that the entire meaning of evaluative terms is captured by these reductive definitions. They can say that there is also a prescriptive component of their meaning as well. Their contention is that their definitions bring out the core descriptive meaning of the terms, that they spell out in scientifically acceptable terms what moral properties are.

The open question argument, due to G.E. Moore, is a widely discussed objection to all proposed naturalistic definitions of ethical terms. It begins with the observation that for any proposed definitional equivalent of a term such as 'good' or 'right', it is an open (or significant) question whether things having the reductive property are always good or right. However, if the properties were identical, then it would not be an open question whether things having one would have the 'other'. Hence, no such definition is correct (see Moore, G.E. §1).

Among ethicists who reject naturalistic definitions, on the basis of the open question argument or for other reasons, there are some who contend that moral predicates do not have descriptive meaning and are instead used to express approval or disapproval or to prescribe or proscribe behaviour. On this view, known as non-cognitivism, evaluative sentences do not say anything that can be true or false. Rather, they express the speaker's attitude (moral indignation or approbation). Others contend that moral properties are natural properties of objects, though they deny the possibility of any sort of definitional reduction to the concepts of, say, physics. Some versions of this non-reductive naturalism draw on the causal theory of reference. They hold that there is a natural property of objects which is causally related to our use of words such as 'right' in some appropriate way and that an action is right just in case it has that property. There is no need to assume, however, that philosophical reflection can reveal just what this property is or that there are any other terms of our language that refer to this property and would constitute a definition of 'right'. Another view, not widely endorsed, is non-naturalism, according to which ethical properties exist and are genuine properties of objects, but they are in some sense independent of the properties studied in the sciences. It strikes many philosophers as mysterious how there could be any moral knowledge if non-naturalism is true.

One can see in the debate about naturalism in epistemology analogues of most of the views just described. There have been some attempts at naturalistic definitions of epistemic terms. An example is reliabilism, which defines 'justified belief' roughly as 'belief caused by a reliable belief-forming process'. Non-cognitivism in epistemology has found some support. For example, J.L. Austin (1946) held that to say that I know some proposition is not to describe some striking cognitive feat, nor is it to describe anything else. It is, he held, to 'give one's word' that the proposition is true. It is something like promising that the proposition is true. Notably, however, non-cognitivism in epistemology has received far less attention and support than it has in ethics.

The view that epistemic properties are not reducible to naturalistic properties is not uncommon. On this view there are fundamental epistemic facts, facts that cannot be reduced to logical or empirical facts. To say that *p* is evidence for *q* is not, according to this view, to say that *p* implies *q* or that *p* has any other logical relationship to *q*, and it is not to say that it is an empirical fact that typically, when *p* is true, *q* is also true. Thus, for example, non-naturalists might say that certain sorts of perceptual experiences simply do provide evidence for certain truths about the world or that certain data provide evidence for a particular theory. According to a similar view, defended by Chisholm (1991), epistemic concepts are all ultimately to be understood in terms of the epistemic relation 'being more reasonable than', as in 'believing *p* is more reasonable than believing *q*'. As is the case in ethics, critics of this view wonder how we could have any access to epistemological properties and relations if they are not, in some sense, complexes of natural properties.

Another element of the debate about naturalism concerns the relationship of ethics and epistemology to the sciences. A primary source for discussions of this aspect of naturalism in epistemology is Quine's 'Epistemology Naturalized' (1969). There, Quine appears to hold that traditional epistemology ought to be replaced by empirical investigation of the ways in which people form beliefs. The analogous view in ethics has scarcely received

attention from philosophers. While few epistemologists endorse Quine's replacement thesis, many agree that epistemology would benefit from closer ties to cognitive science. Epistemologists, who see epistemology as largely devoted to helping people to reason better, understandably think that their task can best be accomplished by paying attention to empirical discoveries about the ways people do reason and about their cognitive capacities. Lacking that sort of empirical information, epistemologists might recommend futile strategies that people cannot follow and attempt to correct errors that do not actually occur. The epistemological enterprise that seeks to make realistic recommendations about reasoning is in some ways comparable to efforts in applied ethics to ascertain the status of particular kinds of actions. Few would deny that empirical results have some implications for those efforts. For example, information about the deterrent effects of capital punishment surely has some bearing on its moral status. There is less clearly room for empirical input into debates about the conceptual analysis of epistemic or ethical terms. Accordingly, epistemologists and ethicists whose primary aim is to provide such analyses are less inclined to endorse naturalism (see Naturalism in ethics; Naturalized epistemology).

3 Universality, objectivity and subjectivity

A second area in which the analogies between ethics and epistemology warrant careful scrutiny concerns the universality and objectivity of the judgments under discussion. It is common to wonder whether there are standards of proper conduct or rational belief that transcend particular cultural norms, and whether there are standards that exist independently of the values and practices of any particular culture. Negative answers to these questions have long been common in ethics, and they are becoming increasingly common in epistemology. Some philosophers have tried to make use of the more developed debate in ethics about these matters to shed light on the epistemological questions (see Cognitive pluralism; Epistemic relativism).

People often wonder whether ethical evaluations are in any interesting sense 'objective'. Some of our conflicting intuitions about this possibly can be resolved by distinguishing two senses of evaluative terms. Consider a physician faced with a choice between two incompatible treatment plans for a patient with a serious disease. It seems clearly right for the physician to select the treatment plan that the evidence indicates would best serve the patient. If, however, that evidence is for some reason misleading and the other approach would in fact prove more successful, it also seems that it would be right for the physician to chose that other alternative. These judgments appear to conflict, but a plausible way out is to say that the former judgment concerns a subjective sense of 'right' and the latter concerns an objective sense of the term. More generally, we might say that the desired end of our actions is the overall well-being of those affected. An action is objectively right or good when it actually would have the best consequences among all available alternatives. It is subjectively right (for a person) when that person reasonably thinks that it would have the best consequences among the available alternatives. Some philosophers would recognize an even more radically subjective notion, according to which an action is right (for a person) provided they think, with or without good reason, that it would produce the most good. Note that the less extreme subjective notion makes at least one important ethical concept analysable in epistemic terms.

Some philosophers have suggested that there are analogous distinctions in the senses of epistemic terms, and this too may help resolve some apparent controversies in epistemology. It is unclear, however, exactly how to conceive of the distinction in the epistemic realm. One might say that the epistemic goal is true belief and that a belief is objectively epistemically justified when it achieves this goal. However, this has the plainly mistaken result that all true beliefs are objectively epistemically justified. A belief held solely on the basis of, say, wishful thinking, is not justified in any sense, even if it happens to be true. An alternative conception of the objective-subjective distinction in epistemology takes an objectively justified belief to be one held for what in fact are good reasons and a subjectively justified belief to be one held for what the believer has good reasons to think are good reasons. Again, one can say that there is an even more radically subjective notion, according to which a belief is justified provided the believer thinks, with or without good reason, that the reasons for it are good. This way of drawing the objective-subjective distinction makes even objective epistemic justification depend upon an individual's evidence or reasons, and thus it differs from objective ethical justification which has no such dependence on evidence. Hence, it seems that objective epistemic justification is perspectival or dependent upon individual factors in a way that objective ethical justification is not. In this area the analogy may be weaker than some philosophers have thought.

4 Internalism and externalism

Epistemology and ethics alike have seen extensive debate about views known as internalism and externalism. Internalists in ethics contend that the factors which determine moral rightness must be in some sense internal factors that are accessible to agents by reflection or introspection alone and that these factors must to some extent motivate agents. Externalists hold that a factor can make an action morally right for an agent even if the agent is not motivated by it or would not be motivated by it upon becoming aware of it (see Moral motivation §1). Epistemology has experienced a similar split. Internalists hold that the factors which determine epistemic justification must be internal and accessible by reflection. Perceptual experiences, memories and other beliefs would thus qualify as potential sources of justification. Further, on at least some internalist views, the relation between evidence and what it justifies is in some important sense internally accessible. Externalists, in contrast, think that factors external to, and not directly accessible to, a mind can determine what is justified (see Internalism and externalism in epistemology; Reliabilism).

5 The 'ought-implies-can' principle

An area in which the analogy between ethics and epistemology seems to break down has to do with the voluntariness of the behaviour they are about. We regularly speak of how we ought to act or believe, and of our ethical and epistemic obligations. It is natural to slide readily from the claim that an act is justified or right to the claim that it ought to be done. Similarly, it is natural to take as equivalent the statements that a belief is justified and that it ought to be held or adopted. A closer look uncovers puzzles. A widely held principle in ethics is the 'ought-implies-can' principle: a person ought to perform an action only if the person can perform the action. This principle forms the basis for the powerful inclination to think that the inability to do things excuses people from doing those things and absolves them of charges of having failed to act as they ought.

Carrying the ought-implies-can principle into epistemology is problematic. In general, our beliefs do not seem to be things over which we have any sort of direct control. At best, we can set out to expose ourselves to influences that we think are likely to lead us to form desired beliefs. If the ought-implies-can principle is true and we have only this limited sort of control over our beliefs, then it less than clear how we can be justified in believing things that we do not believe or how any actual beliefs can be unjustified. Yet, it seems clear that there are examples of both kinds. A person presented with conclusive evidence for the occurrence of some dreaded event might be psychologically incapable of believing the conclusion justified by that evidence. They are then justified in believing something but cannot believe it. Considerations of this sort lead some philosophers to conclude that the ought-implies-can principle does not hold in the epistemological realm, others to conclude that the proper subject matter of epistemic evaluation is voluntary behaviour such as evidence gathering, yet others to conclude that talk of obligation about beliefs is misguided and that epistemic justification has nothing to do with obligations. Analogues of the first and third of these responses are less common in ethics.

There are, of course, numerous other areas in which one can explore similarities and differences between ethics and epistemology. Although much of ethics has explored the evaluation of individual actions, there has also been considerable discussion of the moral virtues. Here, emphasis is placed on identifying and understanding virtuous character traits, not on identifying conditions of right and wrong action. Some epistemologists have followed suit, investigating what they call the 'epistemic virtues' and developing a brand of epistemology known as 'virtue epistemology'. Epistemic virtues might include such traits of mind as a concern for the truth, impartiality, a desire to acquire relevant evidence and so on. The least revisionary versions of virtue epistemology attempt to analyse justification and knowledge in terms of epistemically virtuous belief. More radical versions argue that epistemologists ought to focus on the study of the intellectual virtues themselves, even though their connection to knowledge and justification might be indirect. Whether further investigation of the virtues reveals close analogies or notable differences between ethics and epistemology, such investigation is likely to prove fruitful (see Virtue epistemology).

6 Conceptual connections

Some philosophers have argued that the relation between epistemology and ethics is not simply one of analogy, but rather of reducibility or interdefinability, claiming that the concepts of one domain can be defined or analysed in terms of the concepts of the other. Although the idea that the fundamental ethical concepts can be defined in epistemic terms has not received much support, it is widely held that there is an important ethical concept that can be partially defined in epistemic terms. Thus, many philosophers have held that a person is subjectively justified in

performing a particular action provided they are epistemically justified in believing that the action satisfies some appropriate ethical standard. Philosophers will differ about exactly what that standard is, depending upon their basic perspective in ethics. This sort of connection between ethics and epistemology does not help to resolve any central ethical controversies. It simply acknowledges that there is a notion of ethical justification that features in an agent's epistemic perspective.

Perhaps of greater interest to philosophers have been questions about whether epistemic concepts can be understood in terms of ethical concepts. Roderick Firth (1959) and Roderick Chisholm (1991) long debated this question, Firth arguing that the connection between epistemic justification and ethical justification is merely that of analogy, and Chisholm contending that epistemic justification is a kind of ethical justification.

The simplest possible account of epistemic justification in ethical terms holds that '*S* is justified in believing *p*' means something like 'It is morally right for *S* to believe *p*' or '*S* has a moral duty to believe *p*'. On this view, beliefs and other actions are all appraised using the same standards, whether these be consequentialist, deontological or otherwise. Epistemic appraisal is thus taken to be ethical appraisal applied to belief. This simple theory is unsatisfactory. Assuming for the sake of argument that beliefs are the sort of thing for which ethical appraisal is appropriate, there seem to be numerous cases in which ethical evaluations of beliefs depart from epistemic evaluations. Sometimes, there are decisive moral considerations against a belief that is epistemically justified. For example, it may be that one is morally justified in believing the testimony of family and friends, even when the evidence seems to go against them. Perhaps the value of optimism in the face of adversity morally justifies beliefs that go against the evidence. Examples such as these convince many that epistemic justification is not simply ethical justification applied to beliefs. Of course, the failure of the simplest analysis of epistemic justification in ethical terms does not guarantee the failure of more complex analyses. However, Firth argues that no such 'categorical' analysis of epistemic justification in ethical terms can succeed. He claims that while the epistemic fact is determined by considerations of evidential support for the belief, these considerations always leave open the moral evaluation of the belief.

Another approach to analysing epistemic concepts in ethical terms appeals to the notion of *prima facie* ethical justification. The simplest version of such an analysis says that '*S* is epistemically justified in believing *p*' means the same as '*S* has a *prima facie* moral duty to believe *p*'. This implies that if a person is epistemically justified in believing a proposition, then they have a moral duty to believe it unless some other considerations override that duty. It also implies that if a person has a *prima facie* moral duty to believe a proposition, then they are epistemically justified in believing the proposition. This latter implication opens the analysis to convincing counterexamples. One can have *prima facie* moral duties to believe propositions arising from non-epistemic sources (on the assumption that we can have duties to believe at all). In such cases, there will be the *prima facie* duty, but no epistemic justification.

In addition to efforts to define epistemic terms in ethical terms, philosophers have argued for a weaker sort of necessary connection. The best known claim along these lines is due to W.K. Clifford: 'It is wrong, always, everywhere, and for every one, to believe anything upon insufficient evidence' (1866: 346). Given two assumptions, Clifford's assertion has implications for the connection between epistemology and ethics. One is that by 'wrong' he means 'ethically wrong'. The other is that one has epistemic justification for a belief if and only if one has sufficient evidence for it. Given those assumptions, Clifford's remark implies that if a person is not epistemically justified in believing a proposition, then it is morally wrong for them to believe the proposition. What Clifford says here does not amount to a full reductive analysis, since, for one thing, it does not make a claim about the meaning of epistemic statements. Furthermore, his words do not imply that if a person has sufficient evidence for a proposition, then it is morally wrong for them to fail to believe.

Most commentators take Clifford's thesis to be too strong, for reasons such as those mentioned above. Non-epistemic factors can make belief morally wrong, even though it is supported by the evidence. A more modest thesis along the lines of Clifford's suggestion holds that if a person is justified in believing a proposition, then they have a *prima facie* moral duty to believe it. It is difficult to envisage a decisive refutation of this claim, since it only asserts the existence of a *prima facie* moral duty. Objectors must contend that epistemic justification can fail to carry any moral significance at all. Some would agree with the general line suggested by Clifford, and by Chisholm, which holds that as rational beings we have a moral duty to believe as our evidence indicates, provided

there are no stronger moral considerations imposing a conflicting obligation.

See also: Confucian philosophy, Chinese; Justification, epistemic; Normative epistemology

RICHARD FELDMAN

References and further reading

All these items involve detailed argument but are not technical. All focus primarily on epistemological issues.

Alston, W. (1978) 'Meta-Ethics and Meta-Epistemology', in A.I. Goldman and J. Kim (eds) *Values and Morals*, Dordrecht: Reidel.(Explores the analogies between the two fields, with emphasis on issues concerning naturalism.)

Alston, W. (1989) *Epistemic Justification: Essays in the Theory of Knowledge*, Ithaca, NY: Cornell University Press.(Includes essays on internalism and externalism, subjective and objective conceptions of justification, the ought-implies-can principle, and other related topics.)

Austin, J.L. (1946) 'Other Minds', in J.O. Urmson and G.J. Warnock (eds) *Philosophical Papers*, New York: Oxford University Press, 3rd edn, 1979, 76-116. (Defends a non-cognitivist account of sentences about knowledge.)

Chisholm, R. (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 2nd edn, 1977. (Defends a non-naturalistic account of epistemic concepts.)

Chisholm, R. (1991) 'Firth and the Ethics of Belief', *Philosophy and Phenomenological Research* 51 (1): 119-28. (Defends the thesis that epistemic evaluations are a species of ethical evaluation.)

Clifford, W.K. (1866) *Lectures and Essays*, London: Macmillan, 2nd edn.(Defends the thesis that it is always wrong to believe on the basis of insufficient evidence.)

Feldman, R. (1988) 'Epistemic Obligations', in J.E. Tomberlin (ed.) *Philosophical Perspectives* 2, Atascadero, CA: Ridgeview, 235-56.(Discusses the ought-implies-can principle in epistemology and defends an account of epistemic obligations.)

Firth, R. (1959) 'Chisholm and the Ethics of Belief', *Philosophical Review* 68 (4): 493-506.(Firth's earliest defence of his view that epistemic concepts can not be defined in ethical terms.)

Firth, R. (1978) 'Are Epistemic Concepts Reducible to Ethical Concepts', in A.I. Goldman and J. Kim (eds) *Values and Morals*, Dordrecht: Reidel.(A later and more detailed account of Firth's view.)

Goldman, A. (1992) *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge, MA: MIT Press. (Collection of essays discussing several issues related to naturalism.)

Kvanvig, J.L. (1992) *The Intellectual Virtues and the Life of the Mind*, Savage, MD: Roman & Littlefield.(Argues that the cognitive virtues have a central place in epistemology.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press, 1968.(Contains the classic statement of Moore's open question argument.)

Quine, W.V.O. (1969) 'Epistemology Naturalized', in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 69-90.(Classic statement of Quine's account of naturalistic epistemology.)

Sosa, E. (1991) *Knowledge in Perspective*, Cambridge: Cambridge University Press. (Defends an analysis of knowledge in which epistemic virtues figure prominently.)

Zagzebski, L. (1996) *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, Cambridge: Cambridge University Press.(Develops a unified virtue theory in ethics and epistemology.)

Epistemology in Islamic philosophy

Muslim philosophers agree that knowledge is possible. Knowledge is the intellect's grasp of the immaterial forms, the pure essences or universals that constitute the natures of things, and human happiness is achieved only through the intellect's grasp of such universals. They stress that for knowledge of the immaterial forms, the human intellect generally relies on the senses. Some philosophers, such as Ibn Rushd and occasionally Ibn Sina, assert that it is the material forms themselves, which the senses provide, that are grasped by the intellect after being stripped of their materiality with the help of the divine world. However, the general view as expressed by al-Farabi and Ibn Sina seems to be that the material forms only prepare the way for the reception of the immaterial forms, which are then provided by the divine world. They also state that on rare occasions the divine world simply bestows the immaterial forms on the human intellect without any help from the senses. This occurrence is known as prophecy. While all Muslim philosophers agree that grasping eternal entities ensures happiness, they differ as to whether such grasping is also necessary for eternal existence.

1 Nature of knowledge

Muslim philosophers are primarily concerned with human happiness and its attainment. Regardless of what they consider this happiness to be, all agree that the only way to attain it is through knowledge. The theory of knowledge, epistemology, has therefore been their main preoccupation and appears chiefly in their logical and psychological writings. Epistemology concerns itself primarily with the possibility, nature and sources of knowledge. Taking the possibility of knowledge for granted, Muslim philosophers focused their epistemological effort on the study of the nature and sources of knowledge. Their intellectual inquiries, beginning with logic and ending with metaphysics and in some cases mysticism, were in the main directed towards helping to understand what knowledge is and how it comes about.

Following in the footsteps of the Greek philosophers, Muslim philosophers consider knowledge to be the grasping of the immaterial forms, natures, essences or realities of things. They are agreed that the forms of things are either material (that is, existing in matter) or immaterial (existing in themselves). While the latter can be known as such, the former cannot be known unless first detached from their materiality. Once in the mind, the pure forms act as the pillars of knowledge. The mind constructs objects from these forms, and with these objects it makes judgments. Thus Muslim philosophers, like Aristotle before them, divided knowledge in the human mind into conception (*tasawwur*), apprehension of an object with no judgment, and assent (*tasdiq*), apprehension of an object with a judgment, the latter being, according to them, a mental relation of correspondence between the concept and the object for which it stands. Conceptions are the main pillars of assent; without conception, one cannot have a judgment. In itself, conception is not subject to truth and falsity, but assent is. However, it should be pointed out that *tasdiq* is a misleading term in Islamic philosophy. It is generally used in the sense of 'accepting truth or falsity', but also occasionally in the sense of 'accepting truth'. One must keep in mind, however, that when assent is said to be a form of knowledge, the word is then used, not in the broad sense to mean true or false judgment, but in the narrow sense to mean true judgment.

In Islamic philosophy, conceptions are in the main divided into the known and the unknown. The former are grasped by the mind actually, the latter potentially. Known conceptions are either self-evident (that is, objects known to normal human minds with immediacy such as 'being', 'thing' and 'necessary') or acquired (that is, objects known through mediation, such as 'triangle'). With the exception of the self-evident conceptions, conceptions are known or unknown relative to individual minds. Similarly, Muslim philosophers divided assent into the known and the unknown, and the known assent into the self-evident and the acquired. The self-evident assent is exemplified by 'the whole is greater than the part', and the acquired by 'the world is composite'. In *Kitab at-tanbih 'ala sabil as-sa'ada (The Book of Remarks Concerning the Path of Happiness)*, al-Farabi calls the self-evident objects: 'the customary, primary, well-known knowledge, which one may deny with one's tongue, but which one cannot deny with one's mind since it is impossible to think their contrary'. Of the objects of conception and assent, only the unknown ones are subject to inquiry. By reducing the number of unknown objects one can increase knowledge and provide the chance for happiness. But how does such reduction come about?

2 Sources of knowledge

In Islamic philosophy there are two theories about the manner in which the number of unknown objects is reduced. One theory stresses that this reduction is brought about by moving from known objects to unknown ones, the other that it is merely the result of direct illumination given by the divine world. The former is the upward or philosophical way, the second the downward or prophetic one. According to the former theory, movement from the known objects of conception to the unknown ones can be effected chiefly through the explanatory phrase (*al-qawl ash-sharih*). The proof (*al-burhan*) is the method for moving from the known objects of assent to the unknown ones. The explanatory phrase and proof can be either valid or invalid: the former leads to certitude, the latter to falsehood. The validity and invalidity of the explanatory phrase and proof can be determined by logic, which is a set of rules for such determination. Ibn Sina points out that logic is a necessary key to knowledge and cannot be replaced except by God's guidance, as opposed to other types of rules such as grammar for discourse (which can be replaced by a good natural mind) and metre for poetry (which can be replaced by good taste).

By distinguishing the valid from the invalid explanatory phrase and proof, logic serves a higher purpose, namely that of disclosing the natures or essences of things. It does this because conceptions reflect the realities or natures of things and are the cornerstones of the explanatory phrase and proof. Because logic deals only with expressions that correspond to conceptions, when it distinguishes the valid from the invalid it distinguishes at the same time the realities or natures of things from their opposites. Thus logic is described as the key to the knowledge of the natures of things. This knowledge is described as the key to happiness; hence the special status of logic in Islamic philosophy.

3 Logic and knowledge

We are told that because logic deals only with the known and unknown, it cannot deal with anything outside the mind. Because it is a linguistic instrument (foreign in nature to the realities of things), it cannot deal with such realities directly, whether they exist in the mind or outside it, or are external to these two realms of existence. It can only deal with the states or accidents of such realities, these states comprising links among the realities and intermediaries between the realities and language. Logic therefore deals with the states of such realities, as they exist in the mind. Such states are exemplified by 'subject' or 'predicate', 'universality' or 'particularity', 'essentiality' or 'accidentality'. In other words, logic can deal with realities only in that these realities are subjects or predicates, universal or particular, essential or accidental and so on.

Because the ultimate human objective is the understanding of the realities, essences or natures of things, and because the ultimate logical objective is the understanding of conceptions, logicians must focus on the understanding of those conceptions that lead to the understanding of the essences if they intend to serve humanity. Ibn Sina points out that since the essences are universal, such expressions are also universal in the sense of representing universal conceptions such as 'human being', not in the sense of being universal only in expression, such as 'Zayd'. A universal expression can be applied to more than one thing, as the last two examples show, but one must keep in mind Ibn Sina's distinction between these two types of universal expressions: the former represents reality, although indirectly, the latter does not. It is only the former with which the logician should be concerned (see Logic in Islamic philosophy).

Considering that the discussion of universals occupies a central place in Arabic logic, it is important to focus briefly on this subject to ensure understanding of the proper objects of the knowledge of the natures of things. Muslim philosophers divide universal expressions into five types, known together as the five predicables: genus, species, difference, property and common accident. Genus refers to the common nature of all the species that fall under it, such as 'animality' for 'human being', 'dog', 'cat' and so on. As such, it tells us what the general nature of a thing is. Species refers to the common nature of all the individuals that fall under it, such as 'human being' for 'John', 'George' and 'Dorothy'. As such, it tells us what the specific nature of a thing is. Difference refers to that which differentiates the members of the genus, such as 'rational', which differentiates the species of being human from other animal species; it tells us which thing a being is. These three universals are essential to a thing; that is, without them the essence will not be what it is. Property and common accident are accidental, in that they attach to the thing but are not part of its essence. Property refers to something that necessarily attaches to one universal only, such as 'capacity for laughter' for 'human being'. Common accident refers to a quality that attaches to more than one universal, either in an inseparable manner, such as 'black' for 'crow', or in a separable manner, such as 'black' for 'human being'. The inseparability of the common accident, however, is only in existence.

Only the first three of the above universals constitute the essences of things. If one is to understand the essence of a thing, one must first understand its genus, species and difference or differences. The understanding of these three universals takes place through the explanatory phrase and proof, of which these universals are simple elements. The explanatory phrase is either definition or description. The definition is a phrase which mirrors the essence of a thing by indicating its general and specific essential qualities, that is, its genus, species and difference; the description is like the definition except that it indicates the property instead of the difference. Thus the description does not give a complete picture of the essence of a thing as does the definition. The proof is a set of propositions, which consist of conceptions joined or separated by particles. The proof that helps in the understanding of the essences of things is that which moves from known universal judgments to an unknown universal one.

The important question that concerned Muslim philosophers is how the universals or forms that are essential to the natures of things arrive at the human mind before it has the chance to employ the explanatory phrase and proof to compose known conceptions and known judgments from them. In order to answer this question, Muslim philosophers first discussed the structure of the human soul and then the steps through which the universals pass on their way to the place of knowledge (see Soul in Islamic philosophy). As stated above, conceptions come to the mind through either the philosophical way or the prophetic way. The philosophical way requires one first to use one's external senses to grasp the universals as they exist in the external world, mixed with matter. Then the internal senses, which like the external senses are a part of the animal soul, take in these universals and purify them of matter as much as possible. The imagination is the highest internal sense, in which these universals settle until the next cognitive move. It is from this point to the next step in the philosophical journey that the details seem particularly unclear.

4 The role of the mind

All Muslim philosophers believe that above the senses there is the rational soul. This has two parts: the practical and theoretical intellects. The theoretical intellect is responsible for knowledge; the practical intellect concerns itself only with the proper management of the body through apprehension of particular things so that it can do the good and avoid the bad. All the major Muslim philosophers, beginning with al-Kindi, wrote treatises on the nature and function of the theoretical intellect, which may be referred to as the house of knowledge.

In addition to the senses and the theoretical intellect, Muslim philosophers include in their discussion of the instruments of knowledge a third factor. They teach that the divine world contains, among other things, intelligences, the lowest of which is what al-Kindi calls the First Intellect (*al-'aql al-awwal*), better known in Arabic philosophy as the 'agent intellect' (*al-'aql al-fa''al*), the name given to it by al-Farabi (§3), or 'the giver of forms' (*wahib as-suwar*). They contend that the world around us is necessary for the attainment of philosophical knowledge. Some, such as Ibn Bajja, Ibn Rushd and occasionally Ibn Sina, say that the mixed universals in the imagination that have been derived from the outside world through the senses are eventually purified completely by the light of the agent intellect, and are then reflected onto the theoretical intellect.

Al-Farabi's and Ibn Sina's general view, however, is that these imagined universals only prepare the theoretical intellect for the reception of the universals from the agent intellect that already contains them. When expressing this view, Ibn Sina states that it is not the universals in the imagination themselves that are transmitted to the theoretical intellect but their shadow, which is created when the light of the agent intellect is shed on these universals. This is similar, he says, to the shadow of an object which is reflected on the eye when sunlight is cast on that object. While the manner in which the universals in the imagination can prepare the theoretical intellect for knowledge is in general unclear, it is vaguely remarked by al-Farabi and Ibn Sina that this preparation is due to the similarity of these universals to the pure universals, and to the familiarity of the theoretical intellect with the imagined universals owing to its proximity to the imagination. In other words, the familiarity of this intellect with what resembles its proper objects prepares it for the reception of these objects from the agent intellect.

5 Philosophical and prophetic knowledge

The prophetic way is a much easier and simpler path (see Prophecy). One need not take any action to receive the divinely given universals; the only requirement seems to be the possession of a strong soul capable of receiving them. While the philosophical way moves from the imagination upward to the theoretical intellect, the prophetic way takes the reverse path, from the theoretical intellect to the imagination. For this reason, knowledge of

philosophy is knowledge of the natures of things themselves, while knowledge of prophecy is knowledge of the natures of things as wrapped up in symbols, the shadows of the imagination.

Philosophical and prophetic truth is the same, but it is attained and expressed differently. Ibn Tufayl's *Hayy Ibn Yaqzan* is the best illustration of the harmony of philosophy and religion (see Ibn Tufayl). The so-called double truth theory wrongly views these two paths to knowledge as two types of truth, thus attributing to Ibn Rushd a view foreign to Islamic philosophy. One of the most important contributions of Islamic philosophy is the attempt to reconcile Greek philosophy and Islam by accepting the philosophical and prophetic paths as leading to the same truth.

Muslim philosophers agree that knowledge in the theoretical intellect passes through stages. It moves from potentiality to actuality and from actuality to reflection on actuality, thus giving the theoretical intellect the respective names of potential intellect, actual intellect and acquired intellect. Some Muslim philosophers explain that the last is called 'acquired' because its knowledge comes to it from the outside, and so it can be said to acquire it. The acquired intellect is the highest human achievement, a holy state that conjoins the human and the divine realms by conjoining the theoretical and agent intellects.

Following in the footsteps of Alexander of Aphrodisias, al-Farabi, Ibn Bajja and Ibn Rushd believe that the theoretical intellect is potential by nature, and therefore disintegrates unless it grasps the eternal objects, the essential universals, for the known and the knower are one. Ibn Sina rejects the view that the theoretical intellect is potential by nature. He argues instead that it is eternal by nature because unless it is, it cannot grasp the eternal objects. For him, happiness is achieved by this intellect's grasping of the eternal objects, for such grasping perfects the soul. Muslim philosophers who believe that eternity is attained only through knowledge also agree with Ibn Sina that knowledge is perfection and perfection is happiness.

See also: Epistemology; Ethics in Islamic philosophy; al-Farabi (§3); Ibn Bajja (§3); Ibn Rushd (§6); Ibn Sina (§3); al-Kindi; Logic in Islamic philosophy; Meaning in Islamic philosophy

SHAMS C. INATI

References and further reading

Davidson, H.A. (1992) *Al-Farabi, Avicenna and Averroes on Intellect*, London: Oxford University Press. (Discusses the link between Greek and Arabic understanding of intellect and the various transformations the concept of intellect underwent in Islamic philosophy.)

Fakhry, M. (ed.) (1992) *Rasa'il Ibn Bajja al-ilahiyya (Ibn Bajja's Metaphysical Essays)*, Beirut: Dar al-Jil. (Includes the most important of Ibn Bajja's philosophical treatises, *Tadbir al-mutawahhid (Management of the Solitary)*, *Risalat al-ittisal al-'aql al-fa''al bil-insan (Essay on the Conjunction of the Intellect with Human Beings)* and *Risalat al-wada' (Essay on Bidding Farewell)*.)

al-Farabi (*c*.870-950) *Risala fi al-'aql (Essay on the Intellect)*, ed. M. Bouyges, Beirut: al-Maktab al-Katulikiyya, 1939.(One of the best known and most influential treatises on intellect in Islamic philosophy; it gives the different senses of 'intellect' known to al-Farabi.)

al-Farabi (*c*.870-950) *Kitab at-tanbih 'ala sabil as-sa'ada (The Book of Remarks Concerning the Path of Happiness)*, ed. J. Al-Yasin, Beirut: Dar al-Manahil, 1985. (Includes al-Farabi's definition of the self-evident objects.)

Ibn Rushd (1126-98) *Talkhis kitab an-nafs (Epitome of Aristotle's On the Soul)*, ed. A.F. al-Ahwani, Cairo: Maktabat an-Nahda, 1950.(This edition also includes three other essays: Ibn Bajja's *Risalat al-ittisal (Essay on Conjunction)*, Ishaq ibn Hunayn's *Kitab fi an-nafs (Book on the Soul)* and al-Kindi's *Risalat al-'aql (Essay on Intellect)*.)

Ibn Sina (980-1037) *al-Shifa' (Healing)*, ed. F. Rahman, London: Oxford University Press, 1959.(Standard account by Ibn Sina of his views on the soul, including the essays *at-Tabi'iyyat (Physics)* and *an-Nafs (Psychology)*.)

Ibn Sina (980-1037) *al-Isharat wa'l-tanbihat (Remarks and Admonitions)*, part translated by S.C. Inati, *Remarks and Admonitions: Part One, Logic*, Toronto: Pontifical Institute for Mediaeval Studies, 1984.(The most comprehensive of Ibn Sina's logic and best representation of Arabic logic.)

Ibn Tufayl (before 1185) *Hayy Ibn Yaqzan (The Living Son of the Vigilant)*, ed. L. Gauthier, Beirut: Catholic Press, 1936; trans. L. Goodman, *Ibn Tufayl's Hayy Ibn Yaqzan, A Philosophical Tale*, New York: Twayne Publishers, 1972.(Expresses the harmony between reason and revelation in a literary form).

Nuseibeh, S. (1996) 'Epistemology', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 49, 824-40.(Analysis of the main concepts of epistemology, along with discussion of how some of the main thinkers take up different positions.)

Rida, A. (ed.) (1950) *Rasa'il al-Kindi al-falsafiyya (al-Kindi's Philosophical Essays)*, Cairo: Dar al-Fikr al-'Arabi.(These two volumes include four essays relevant to al-Kindi's theory of knowledge: *Risalat al-Kindi fi al-qawl fi an-nafs (Al-Kindi's Essay on the Discourse Concerning the Soul)*, *Kalam lil-Kindi fi an-nafs (Words for al-Kindi Concerning the Soul)*, *Risalat al-Kindi fi mahiyyat an-nawm war-ru'ya (Al-Kindi's Essay on Sleep and Vision)* and *Risalat al-Kindi fi al-'aql (Al-Kindi's Essay on the Intellect)*. The last of these is the best known and seems to have been the first in a long and influential series of Arabic works on the intellect.)

Rosenthal, F. (1970) *Knowledge Triumphant: The Concept of Knowledge in Medieval Islam*, Leiden: Brill.(By far the best work on epistemology in Islamic thought, authoritative and always interesting.)

Epistemology, history of

Epistemology has always been concerned with issues such as the nature, extent, sources and legitimacy of knowledge. Over the course of western philosophy, philosophers have concentrated sometimes on one or two of these issues to the exclusion of the others; rarely has a philosopher addressed all of them. Some central questions are:

(1) What is knowledge - what is the correct analysis or definition of the concept of knowledge?
(2) What is the extent of our knowledge - about what sorts of things is knowledge actually held?
(3) What are the sources of knowledge - how is knowledge acquired?
(4) Is there any genuine knowledge?

Concern with the first question has predominated in philosophy since the mid-twentieth century, but it was also discussed at some length in antiquity. Attention to the second question seems to have begun with Plato, and it has continued with few interruptions to the present day. The third question was also important in antiquity, but has also been a central focus of epistemological discussion through the medieval and early modern periods. The fourth question raises the issue of scepticism, a topic which has generated interest and discussion from antiquity to the present day, though there were some periods in which sceptical worries were largely ignored.

Various attempts to answer these questions throughout the history of philosophy have invariably served to raise additional questions which are more narrow in focus. The principal one which will be treated below can be stated as:

(5) What is a justified belief - under which conditions is a belief justified?

There has been but occasional interest in this last question in the history of philosophy; however, it has been a crucial question for many philosophers in the twentieth century.

1 Ancient philosophy

The extant writings of the Presocratics primarily address issues in metaphysics and cosmology; epistemological concerns appear to arise first in Plato. In the *Meno* Plato tells the story of a slave boy who has had no formal education and in particular has never studied geometry. In a conversation with Socrates, the boy is led to answer questions about a geometrical figure and his answers turn out to be correct. The boy is led to assert that when given a square with side S, so that its area is S^2, then a square of exactly $2S^2$ is formed by taking as its side the diagonal of the original square. The boy could hardly have learned this earlier, since he is uneducated. Plato takes this example to show that the boy knew the geometrical truth all along and, more generally, that the boy's soul existed earlier in a state of knowledge. Indeed, he held that the boy's soul earlier knew all truths but had since forgotten them. What the boy was really doing in his conversation with Socrates was recollecting something he had forgotten. And this Plato takes to hold for everyone: what we think of as coming to know is really recollecting.

The soul or knower may have come into existence and so it would not have always had knowledge; or it may have always been in existence but at some time acquired its stock of knowledge; or it may have always been and always had knowledge. Plato certainly rejects the first option, especially in the *Phaedo* where he argues for the indestructibility and indefinite prior existence of the soul. He also appears to reject the second option (*Meno* 86b) so that his view would be that the soul always exists and earlier had a great deal of knowledge without having acquired this knowledge at some time.

The *Meno* also contains a distinction between true belief and knowledge (97d-98b). Knowledge, Plato says, is 'tied down' or tethered in a way that true opinion is not. This view, which seems to suggest that knowledge is justified true belief, is taken up again in the *Theaetetus*, where Plato suggests that knowledge is true belief plus an account or *logos* (201d). However, several attempts to explicate the notion of an account are rejected, and the dialogue ends inconclusively. It is not clear whether Plato rejects this account of knowledge outright, or whether he is best construed as rejecting this definition given the defective notions of an account (*Theaetetus* 210a-b).

In the *Republic*, especially in Book V, Plato addresses a version of our question pertaining to the extent of

knowledge. There he distinguishes between knowledge at one extreme and ignorance at the other, and he roughly identifies an intermediate state of opinion or belief. Each of these states of mind, Plato says, has an object. The object of knowledge is what is or exists; the object of ignorance is what does not exist; and the object of belief is some intermediate entity, often taken as what is becoming or the sensible physical world of objects and their qualities (*Republic* 508d-e; *Cratylus* 440a-d). What truly exists for Plato are unchanging Forms, and it is these which he indicates as the true objects of knowledge. Moreover, knowledge is infallible, while belief is fallible (*Republic* 477e). In thus identifying knowledge with infallibility or certainty, Plato is departing widely from the view of knowledge given in *Meno* and *Theaetetus*. And, his account of the extent of our knowledge is also severely restricted: genuine knowledge is had only of the higher realm of immutable, ideal Forms (see Certainty; Fallibilism; Plato §§11, 15).

Aristotle discusses a special form of knowledge, scientific knowledge, in the *Posterior Analytics*. A science, as Aristotle understands it, is to be thought of as a group of theorems each of which is proved in a demonstrative syllogism. In the first instance, a demonstrative syllogism in science *S* is a syllogistic argument whose premises are first principles of *S*. These first principles, in turn, must be true, primary, immediate, better known than and prior to the conclusion, which is further related to them as effect to cause (*Posterior Analytics* 71b 21-22). First principles are primary and immediate when they are not themselves demonstrable. Still, such principles are known; indeed, they are better known than the demonstrated conclusion, a contention which may mean either that they are more familiar than the conclusion, or perhaps more certain than the conclusion. The first principles are also said to be prior to the conclusions in an epistemic way: knowledge of the conclusion requires knowledge of the first principles, but not conversely (*Taylor 1990: 121*). And the first principles must explain why it is that the demonstrated conclusion is true.

The science can be extended by taking theorems proved from first principles as premises in additional demonstrative syllogisms for further conclusions. Here, too, the premises must explain the truth of the conclusion. A science will be the sum of all such theorems, demonstrated either from first principles or from already-demonstrated theorems in appropriate syllogisms. And a person who carries through all these syllogisms with relevant understanding has knowledge of all of the theorems.

The first principles, however, are also known though they are not demonstrated as theorems. On this point Aristotle gives what may be the first statement of a regress argument in favour of a kind of foundationalist position (see Foundationalism). Some might hold that even the first principles must be demonstrated if they are to be known. This would lead to an infinite regress, as these first principles would themselves be conclusions of syllogisms whose premises were other first principles which, to be known, would have to be demonstrated. To avoid the infinite regress, one would need either to allow for circular demonstration, or to agree that the first principles are not themselves known but are mere suppositions. Aristotle rejects all these options in favour of the foundationalist view in which the first principles are known even though they are not demonstrated. For him, one has an immediate, intuitive grasp of the first principles. However, his foundationalism is to be distinguished from those discussed below, because his foundations are made up of fundamental principles of special sciences.

In *De Anima*, Aristotle discusses perception and perceptual knowledge. Among perceptible objects, he distinguishes between proper and common sensibles. Common sensibles are those objects that are perceivable by more than one sense, for example, the shape of a box which can be both seen and touched. Proper sensibles are those objects that are only perceivable by one sense, for example, colour can only be seen. With respect to these proper objects, Aristotle says that one cannot be in error, or one cannot be deceived (*De Anima* 418a 9-13; 428b 17-21). So, if a person sees a white cat, they can be deceived about whether it is a cat, but not in thinking that there is white present. The same would apply to proper objects of other senses. If we then assume, as Aristotle seems to have done, that the impossibility of being deceived about *X* is sufficient for having knowledge about *X*, then we reach the conclusion that we have certain perceptual knowledge about the proper objects of each sense. What is not clear is whether Aristotle felt that one could not be mistaken about actual qualities of physical objects, such as their colours; or rather about an object's perceived qualities. Clearly the former is much less plausible than the latter (see Aristotle §6).

2 Hellenistic philosophy

The Hellenistic phase of philosophy occupies several centuries after Aristotle's death (322 BC), and is notable for

its three schools of philosophy: Epicureanism, Stoicism, and Scepticism (see Scepticism). The sceptical tradition, however, continued well into the second century AD.

The Epicurean school supported an even more thoroughgoing empiricism than we find in Aristotle, and is best known for its doctrine that all perceptions are true. In perception, Epicurus says, thin layers of atoms are emitted from external physical objects (*eídōla*) and reach our senses, which passively receive and register these *eídōla* exactly as they are. But this, *per se,* is not to have knowledge of the external causes of our experiences of *eídōla*. For that, we need to make well-grounded inferences to the existence and nature of the external objects. Epicurus, however, maintains that these inferences, doubtless causal ones, can be legitimately made, and thus that there is genuine perceptual knowledge of physical objects. The fact that these inferences may fare poorly under sceptical scrutiny is not a matter to which Epicurus paid special attention, probably because he did not think that the defeat of sceptical concerns was a necessary project within epistemology (see Epicureanism §§6-7).

The Stoic position is much less optimistic. Their central concept is that of a cognitive impression. In normal conditions a red object appears red, and one has the thought (cognitive impression) ‘This is red’. Such a cognitive impression, Stoics held, cannot fail to be true. It is not, by itself, knowledge of the red object however, because the person might not assent to this impression. One has knowledge only if one assents to a cognitive impression and this assent is firm, the sort of assent that one cannot be persuaded to withhold. Ordinary people fall short of assent of this firm sort, and so really have mere opinions about objects. Only the wise man typically engages in firm assent to only cognitive impressions; so only the wise man truly has knowledge of such objects (see Stoicism §12).

In thus restricting knowledge, the Stoic position is actually close to a sceptical doctrine. The two schools of ancient scepticism, the Academic and the Pyrrhonian, had notable differences and each had points of development over nearly five centuries, ending with Sextus Empiricus in the late second century AD. A common feature of each school, though, is an attack on claims to knowledge. For any argument towards a conclusion which goes beyond sensory appearances, sceptics maintained that an equally strong counter-argument could be given. Other sceptical arguments point to the relativity of all perception, depending on changes in the percipient or in the observation conditions or perspectives, and conclude that we do not gain knowledge of external physical objects via perception. Also, if a criterion for knowledge-acquisition is relied upon - such as perception or causal inductive inference as in Epicurus - then this criterion could be questioned as itself far from evidently reliable. These sceptical arguments were properly taken to lead to the suspension of belief (*epochē*), rather than to the assertion that there is no knowledge. Moreover, the Pyrrhonian sceptics noted that the ultimate goal of their arguments was a non-epistemic one, that of *ataraxía* or being undisturbed. This calm state is presumed achievable once beliefs have been suspended and one is content to carry on one’s life dealing only with appearances (see Pyrrhonism; Sextus Empiricus).

Scepticism was challenged in the early medieval period by Augustine in his *Contra Acadcemicos*, in which he dealt critically with the arguments of Cicero, the last of the great academic sceptics (see Augustine §2; Cicero). However, scepticism was not a major concern in the Middle Ages, and did not receive special philosophical attention again until the Renaissance.

3 Medieval philosophy

Medieval philosophy is concerned primarily with issues in metaphysics, logic and natural theology, less with epistemological topics. Aquinas and Ockham, however, were two thinkers for whom epistemological questions were of great interest and importance.

Aquinas closely followed Aristotle on many points, including Aristotle’s account of scientific knowledge (see §1 above). Thus, *scientia* or genuine scientific knowledge is restricted to propositions proved in demonstrative syllogisms whose premises are themselves known. And, as in Aristotle, Aquinas holds that this account of *scientia* requires as premises in some demonstrative syllogisms, first principles that are known *per se*, immediately and without inference. That there must be such first principles is shown by the fact that otherwise one is faced with either an infinite regress of items of knowledge or with circularity. The former is ruled out because nobody can achieve an infinite number of inferential steps; and circularity, wherein one knows *p* on the basis of *q* and knows *q* on the basis of *p,* is dismissed because it allows that a single proposition, *p*, can both be and not be epistemically prior to *q*. For both Aquinas and Aristotle, the premises of a demonstrative syllogism are epistemically prior to

their conclusions in the sense that one cannot know the conclusion without first knowing the premises, though not vice versa.

First principles of demonstrative syllogisms, for Aquinas, are necessary truths, propositions in which there is a necessary connection between the predicate and subject concepts. To grasp or know a first principle, then, requires that one grasp this necessary connection. To achieve this, one must first possess the general concepts expressed by the subject and predicate terms. Accordingly, Aquinas gives an account of how we may acquire such concepts. To do this, one must abstract the intelligible species or forms of objects from the sense impressions received in perception of them. Ultimately, then, knowledge of the first principles depends on perception, though it is not epistemically based on perception.

William of Ockham makes an interesting break with Aquinas' conception of knowledge. Part of the break concerns the fact that Ockham allows for knowledge of contingent truths, as opposed to restricting knowledge to necessary truths. Another difference concerns perception. Aquinas held that, in perception, one has an experience of an image or phantasm from which by some cognitive mechanism (called an agent intellect) one abstracts an intelligible species or form of the perceived object. By means of this abstracted item one knows some universal aspect of the perceived object (see Aquinas, T. §11). Ockham disagrees with a number of these points. For him, in perception of an external physical object, there is no intermediary such as a phantasm or sensible species. Instead, the object is itself perceived directly, an experience which Ockham regards as an intuitive cognition. In this respect, Ockham's account is close to that of perceptual direct realism (see Perception, epistemic issues in). But he also holds epistemic direct realism, because he argues that the direct visual awareness of the external object suffices for one to acquire immediate and certain knowledge of the existence of the object and of some of its qualities. In this regard, one is 'knowing the singular' rather than something universal; and the proposition one comes to know in perception is a contingent truth. So, in perceiving a red box, one might come to know that there is a red box before one.

Sceptical worries could intrude, even here. One might see a red box and still be mistaken in one's resulting belief. One might think, therefore, that forming the belief as a result of a red-box visual experience does not suffice for knowledge. Ockham is aware of this objection, but he has a two-part answer to it. First, he notes that the mere possibility of mistaken belief does not rule out knowledge. He also notes that aligning the concept of certainty with the impossibility of mistaken belief is itself an error. Certainty, he says, requires only the absence of actual doubt or grounds for doubt. Certainty of this weaker sort, he holds, is all that is needed for knowledge (see William of Ockham §§4-5).

This is a decisive break with earlier views, both with respect to the concept of certainty and with respect to its relation to the concept of knowledge. But, with one or two notable exceptions, these ideas were not taken up by many writers until much later.

4 Modern philosophy: Descartes

It is customary to begin the story of modern philosophy with Descartes, but we need to start a little farther back with a discussion of scepticism. We have noted that ancient scepticism was hardly known during the Middle Ages. In the sixteenth century, however, the old sceptical texts of Cicero were re-published and works by Sextus Empiricus were translated into Latin and thus made available to scholars. These texts and their arguments became very important to those on both sides of disputes over the legitimacy and extent of religious knowledge, an issue given great currency by the Reformation and the Counter-Reformation. Under the direct influence of Sextus Empiricus, Montaigne published his *Apology for Raimond Sebond* (1576) in which he set forth sceptical arguments and recommended suspension of belief on practically all topics. His disciple, Pierre Charron, popularized sceptical doctrines even further. This sceptical climate was well known to Descartes in the first half of the seventeenth century. Still later, Pierre Bayle's *Dictionary*, which contained a number of sceptical entries, was to have a great deal of influence on Berkeley and Hume (see Bayle, P.).

Descartes was thoroughly aware of the sceptical writings and debates of his time, and of the development of sceptical literature since Montaigne. But Descartes himself was no sceptic; on the contrary, he set out to defeat scepticism on its own terms, that is, by finding some knowledge which is completely certain and thus immune to sceptical criticism.

To accomplish this, Descartes used the method of doubt, a method wherein a proposition is considered false provided there is even the slightest possible ground for doubting it. Whole classes of propositions would then be excluded as not known: everything which one believes on the basis of the senses is dubitable by this criterion, and so is not knowledge. The many propositions of science also qualify as dubitable, and so are not items of knowledge. Indeed, it is possible, Descartes reasons, that an evil demon systematically deceives us all, even with respect to the necessary truths of mathematics. If such a demon is even possible, then there is at least the possibility of grounds for doubt, and so virtually nothing would qualify as knowledge.

Descartes contends, however, that such an evil demon cannot deceive him in one case, namely when he thinks in any way. Even when the thinking in which he engages is a case of doubting, whenever Descartes thinks he must exist, and thus he affirms as certain '*Cogito, ergo sum*' - I think, therefore I exist. This is an item about which he cannot be deceived, and it is thus for Descartes indubitable or certain, and assuredly a case of knowledge.

Descartes' epistemological project then becomes one of seeing whether any other genuine certain knowledge can be derived from this very slender base. He finds first a criterion for certainty: those thoughts or ideas which are clear and distinct are also true. In fact, he says that clarity and distinctness of a thought or idea suffices to assure him of its truth. Using this criterion together with his certain knowledge that he exists, Descartes constructs a complex causal argument for the existence of God. The clarity and distinctness of the thoughts that God is not a deceiver and that God would not allow wholesale deception is then put to work to try to derive propositions formerly excluded as dubitable by the method. Especially important here are propositions concerning the existence of external physical objects.

Descartes' project is thus a foundationalist one of an austere sort. For him, the foundations are restricted to the propositions that he himself exists, that he has certain ideas, and that God exists. From these, utilizing the criterion of clarity and distinctness, the foundations can be augmented to include propositions about immediately experienced sensations. However, derivations of other propositions from these foundational ones have to be restricted to deductions that themselves can be seen to be clear and distinct. If the derivations were inductive, then grounds for doubting the conclusions would be possible. And even if the derivations were deductive, if one did not see that they were validly made from individually indubitable premises, once again the possibility of grounds for doubting the conclusions would arise. Only if the possibility of such grounds are eliminated can these derived conclusions count as items of knowledge.

Descartes, thus, perpetuates and even emphasizes the close conceptual connection between knowledge and the strictest sort of certainty (see Certainty). He also gives currency to the problem of the external world, that is, the problem of deriving propositions concerning external physical objects from foundational propositions made up mostly of propositions concerning sensations. Of course, Descartes has propositions about a non-deceiving God in his foundations, unlike later writers who grappled with this problem. So armed, Descartes claims in the sixth Meditation that he can derive the general claims that there are external physical objects and that they have at most the so-called primary qualities. But even if these claims count as items of knowledge, so that to some extent scepticism is vanquished, it does not seem that Descartes secures knowledge of individual propositions about physical objects and their qualities. For he concedes that, with respect to these, error is possible in the best of circumstances, even with God's help (Williams 1978: 234, 249ff) (see Descartes, R. §§3-5).

5 Modern philosophy: Spinoza and Leibniz

It is customary to classify Spinoza and Leibniz along with Descartes as rationalists. In epistemology, rationalism is the view which stresses the role of reason in the acquisition of knowledge, and correspondingly downplays the role of experience or observation. A limiting case of rationalism, then, would be a position which held that only reason is operative in knowledge acquisition. It is perhaps Spinoza who comes closest to a rationalist position of this sort (see Rationalism).

For Spinoza, a true idea is one which must agree with its object (*Ethics* I: Ax.6). An adequate idea is one which, considered by itself, has an internal sign or intrinsic mark of a true idea (*Ethics* II: Def.4). Having an adequate idea, then, suffices to recognize it as true. There is no need of a clarity-and-distinctness criterion for determining which ideas are true. In this respect, Spinoza differs sharply from Descartes.

Spinoza distinguishes three levels of knowledge. The first is that which we receive in sense perception or from

what he calls ‘signs’, as when the sight of some printed words causes one to remember something. First-level knowledge is not strictly knowledge, however, but rather opinion or imagination. Second-level knowledge or reason (*ratio*) is knowledge of the properties of objects and of relations between properties. Third-level knowledge is intuitive science, which Spinoza says ‘advances from an adequate idea of the formal essence of certain attributes of God to the adequate knowledge of the essence of things’ (*Ethics* II: pr.XL, schol. 2). Third-level knowledge proceeds from one thing to another in the sense that a person who has an adequate idea of the formal essence of one of God’s attributes may logically infer to adequate knowledge of the essence of things.

Knowledge is adequate when one may logically infer, merely from having an adequate idea of *x*, to some general truth about *x* (second level), or to some truth about *x*’s effects (third level). Thus, on the second level, from an adequate idea of body one may infer that all bodies are capable of motion, and thus knowledge of this proposition is adequate. And from the adequate idea of the essence of a divine attribute, one may infer to the essence or nature of objects, and thus the proposition concerning the essence of the objects is adequately known.

Spinoza certainly thinks we have adequate ideas, and so have adequate knowledge (*Ethics* II: pr.XXXIV). And he holds that the propositions known at the second and third levels are necessarily true (*Ethics* II: pr.XLI). So, it looks very much as if Spinoza is committed to the view that second- and third-level knowledge is a priori, that is, knowledge that need not rely on experience, and to this degree he would qualify as a rationalist (see Spinoza, B. de §§7-8).

Leibniz, the other great philosopher usually classified as a rationalist, did not develop a systematic view in epistemology. His classification as a rationalist is no doubt tied to two important strands of his thought. For simple subject-predicate propositions, fundamental for Leibniz, he proposed the predicate-in-notion principle. This is the thesis that the concept of the predicate in such a proposition is contained in the concept of the subject. It seems as though this principle implies that all subject-predicate propositions are necessarily true. For the conceptual-containment doctrine amounts to the claim that such propositions are true in virtue of their meanings or are conceptually true, and this would make them necessarily true. This has the twofold result that all truth is necessary truth, given that subject-predicate propositions are fundamental; and, that all knowledge is or can be a priori, the latter on the assumption that if a proposition is a necessary truth, then it is a priori knowable. If Leibniz held these views, his status as a rationalist is secure.

Leibniz strove to ward off these consequences, however, by an account of analysis. He held that in a necessary proposition, the concept-containment feature allows for the proposition to be analysed or reduced to an identity proposition in a finite number of steps. Contingent truths, however, cannot be so analysed, despite the concept-containment thesis. Instead, in an infinite number of steps of analysis, such propositions would converge on an identity proposition. (Sometimes Leibniz suggests that such propositions can be analysed into identity propositions by God.) So, not all truths are necessary, and thus neither is all knowledge a priori.

There is a strong rationalist side to Leibniz, however, which emerges in the second strand of his thought, namely, his defence of innate truths. In a dispute with Locke, Leibniz contended that there are numerous innate concepts and principles in pure mathematics, logic, metaphysics and ethics. These innate truths are all necessary truths, and they are all knowable a priori. The senses, Leibniz says, merely function as the occasions by and on which these truths are brought to attention (see Leibniz, G.W. §§8-9).

6 Modern philosophy: Locke and Berkeley

Locke provides a strong empiricist contrast to both Spinoza and Leibniz. For Locke, the fundamental items of all cognitions are ideas, which divide into those of sensation and those of reflection. The former are acquired in perception, the latter in introspective attention to the contents and workings of one’s own mind. Perception and reflection, for Locke, make up experience, and the fundamental empiricist thesis is that all ideas and all knowledge derive from experience. It follows from empiricism so construed that no ideas are innate. For Locke, the mind is a ‘blank tablet’ at birth, and it is only by experience that it acquires its stock of ideas.

Locke defines knowledge as the perception of the agreement or disagreement of two ideas (1689: IV, I, 1 and 5). This definition has the immediate effect of restricting all knowledge to ideas, something Locke recognizes and appears to accept (1689: IV, II, 1). It also seems to have the effect of restricting knowledge to relations between ideas. The definition and the restriction accord well with most of what Locke says about knowledge.

Intuitive knowledge is the perception of the agreement or disagreement between two ideas 'immediately by themselves, without the intervention of any other' (Locke (1689): IV, II, 1). Perception that white is not the same as black, for example, is immediate and requires no intermediate idea between those of white and black. Intuitive knowledge, for Locke, is the most certain: it is both irresistible and infallible.

Locke seems to desert his definition of knowledge in three important cases, however, and in two of these cases intuitive knowledge is at issue. One has, for instance, intuitive knowledge of individual ideas, as when one knows that some pain is very sharp (Locke 1689: IV, II, 1). Locke also maintains that one has intuitive knowledge of oneself. In such a case, even if an idea of reflection is had, self-knowledge is not a perception of an agreement or disagreement of two ideas. Moreover, it is knowledge about the self, which is not an idea or group of ideas. In this case, Locke departs not merely from his definition, but also from his explicit claim about the extent of our knowledge.

Demonstrative knowledge, for Locke, requires that each step in the demonstration be intuitively known, and that the relation between the premises and the conclusion also be intuitively known. Meeting these constraints on demonstrative knowledge assures that it is virtually as certain as intuitive knowledge. But meeting these constraints is not easy, especially in long demonstrations where one must keep in mind inferences made earlier. In such cases, Locke indicates, one's degree of certainty with respect to the conclusion will drop and one will not have demonstrative knowledge properly speaking.

Locke's account of sensitive knowledge marks a third point at which he seems to depart from his official definition of knowledge and the restriction of knowledge to our ideas. Sensitive knowledge is knowledge of the existence of external physical objects. It is not as certain as intuitive or demonstrative knowledge, yet it is still knowledge. And Locke clearly thinks that we have such knowledge, at least in those cases when an external physical object is actually present to one's senses (1689: IV, III, 5). Locke conceives of sensitive knowledge of presently perceived physical objects as inferential knowledge. From knowledge of presently experienced ideas one infers that there is an external physical object present as the cause of those ideas. Locke is untroubled by sceptical worries over whether such inferences can be legitimately made.

The distinction between intuitive and certain knowledge of ideas, and sensitive knowledge of physical objects, with the latter knowledge inductively based on the former, is indicative of Locke's foundationalist position. It differs from that proposed by Descartes, however, in two important ways. First, the propositions making up the foundations are different. For Locke, these are confined to propositions about individually experienced ideas, or to propositions describing a perceived agreement or disagreement between ideas. Thus, Locke's commitment to empiricism dictates what the foundations shall be. Another difference comes in the inferences from the foundational propositions which Locke finds acceptable. He allows for both deductive and inductive inferences, whereas for Descartes permissible inferences may only be deductive. Locke thus marks a liberalization of the foundationalist strictures imposed by Descartes (see Locke, J. §§2-3).

Berkeley was critical of Locke's account of knowledge of physical objects, as was Hume (though, unlike Berkeley, Hume did not mention Locke by name). Locke notes that inductive inferences from currently experienced ideas to physical objects will succeed only when there is a conformity between the ideas and the physical object (1689: IV, IV, 3). Ideas, that is, have to represent the physical objects in some way. Berkeley denies that ideas can serve this role. An idea, he says, can only be like or similar to another idea, not a physical object. Moreover, even allowing for this similarity, the needed inductive inferences depend on and so require that one establish that some ideas do adequately represent objects. To accomplish this, Berkeley argues, one must be in a position to compare the ideas and the physical object. However, as Berkeley notes, this is a position one cannot occupy given Locke's account of perception, which restricts immediate perception to ideas, and so never allows for immediate perception of physical objects. Locke's overall theory, according to Berkeley, really leads to scepticism about physical objects.

To avoid this, Locke could drop the demand that currently experienced ideas conform to or represent objects. Berkeley suggests that this manoeuvre is no help because Locke's theory still requires inductive inferences from the ideas to the physical objects. He notes that the inferences at issue would be explanatory - the supposition that there are physical objects present causally explains the ideas one experiences - but denies their cogency (Berkeley

1710: 19-20, 53). A simpler and thus better explanation of our ideas, Berkeley argues, would be the supposition that they are caused by a single powerful being such as God.

Locke's empiricist version of foundationalism is often attributed to Berkeley. However, Berkeley seems to reject such a theory in favour of a foundationalism both more expansive and more modest. It is more expansive because Berkeley allows that we have immediate and certain knowledge of physical object propositions as well as propositions about currently experienced ideas. Hence, while Berkeley accepts an empiricist version of foundationalism, the propositions he is willing to count as foundational include many more than are countenanced on Locke's theory. Berkeley's theory is more modest in regard to the concept of certainty. For him, a proposition is certain provided that one has no actual grounds for doubting it. It is not further required that mistaken belief is logically impossible. In this way, Berkeley is able to contend that physical object propositions are certain, and he can avail himself of a much more modest criterion of what is to count as a foundational proposition. On this point, Berkeley lines up with Ockham (see §3 above), and with certain twentieth century philosophers (see §8 below).

Berkeley also aimed to refute scepticism regarding the external world. He argues that this may be achieved provided one can find a way to allow for the immediate perception of physical objects. His thought is that if we perceive physical objects immediately, then we also have immediate and certain knowledge of them. He claims that these results are all achieved by abandoning realism regarding physical objects, and embracing instead a thesis which entails that objects exist if and only if they are perceived. Thus, he defends the phenomenalist thesis that a physical object is identical to a collection of ideas (see Phenomenalism). Objects which are collections are immediately perceivable so long as one immediately perceives some of their constituent members. So, the phenomenalist thesis regarding objects allows Berkeley to defend the view that physical objects are immediately perceivable, and hence to argue for the claimed refutation of scepticism and a more expansive foundationalism. In these respects, Berkeley claims, he is merely defending the views of common sense (see Berkeley, G. §§5-9).

7 Modern philosophy: from Hume to Peirce

Both Locke and Berkeley accept the theory that in every perceptual experience, one is immediately aware of at least one idea. Hume follows them in this, but he distinguishes between impressions, which are our more lively and original perceptions, and ideas, which are less lively. In seeing a red cup, one experiences a red impression (or perhaps an impression of red), while in remembering the cup one attends to an idea of the red cup. Hume's fundamental principle is that all ideas are derived from impressions, and in this regard he is a thorough-going empiricist about concepts (ideas).

He also seems to accept epistemic empiricism, at least in the sense that a proposition about an object not currently present to one's senses would count as knowledge only if that proposition were derivable from propositions about currently experienced impressions. Hume denies that physical-object propositions can be deduced from propositions about impressions. He also notes that inductive inference is not something that can be given a non-circular justification. Hence, the inductive inferences from impression propositions to physical-object propositions are not justified, and so scepticism regarding physical objects results.

Hume does note, however, that nature or our psychological make-up does not allow us actually to accept scepticism, or to refrain from making inductive inferences, especially causal ones. He may mean that the fact that we are built in such a way, psychologically, that we make inductive inferences beyond our impressions to beliefs about physical objects itself constitutes being justified in having these beliefs and making these inferences. If so, then Hume is an early externalist about justification and knowledge (see Internalism and externalism in epistemology). Or, he may mean that we can only describe the beliefs we have and the inferences we make; questions about the justification of these inferences and whether the beliefs count as knowledge cannot be settled. In that case, Hume accepts the sceptical results noted above (see Hume, D. §2).

Two very important critics of what they regarded as Humean scepticism were Reid and Kant. Reid argued that Hume's scepticism was generated by acceptance of the theory of ideas (impressions), arguing that no philosopher had ever given any good reason for accepting this theory, and that it gives a mistaken account of perception in any case. The correct account, for Reid, is a complex version of direct realism in which we gain immediate and certain knowledge of physical objects. The beliefs we gain in direct perception of objects are typically irresistible, and it is a first principle for Reid, a matter of common sense, that perception is reliable and so such beliefs are justified and

constitute knowledge (see Reid, T. §§1, 7).

Hume's scepticism did not extend to what he called 'relations of ideas'. These included the necessary truths of mathematics, and of these Hume allowed we can have a priori knowledge (1748: IV, 1). It was only with respect to some statements of matters of fact that Hume was sceptical. For Kant, relations of ideas are analytic statements, while matters of fact are synthetic (see A priori; A posteriori; Analyticity). He felt that there was a third category, however, which Hume had missed, namely synthetic a priori propositions. These are necessary truths in which the meanings of the predicate terms are not contained in the meanings of their subject terms; hence, they are synthetic. But Kant argued that the necessary truths of geometry and arithmetic are synthetic a priori, as are some very general principles of science, and these can all be known a priori. He argues that the a priori concepts he calls categories genuinely apply to objects we experience, and that our experience actually is objective in the sense that it is of real physical objects. Kant also held that having experience of objects suffices for having knowledge of such objects, and so scepticism regarding physical objects is incorrect (see Kant, I. §§4, 6).

Hegel's *Phenomenology of Spirit* (1807) contains an extended criticism of a doctrine often thought to be common to all empiricists, namely that there is immediate knowledge of something *given* in perception (for the classical empiricists, ideas), and that this knowing is passive in the sense that it is unmediated by concepts. This criticism, of course, would apply to any variant of empiricism, including a view which holds that physical objects and not subjective ideas are perceptually given and are objects of passive, immediate knowledge. Hegel's view is that there simply is no knowledge of this sort. Rather, all knowledge is conceptually mediated. Hegel seems to have drawn the conclusion that there is nothing at all which is given, a doctrine later given great currency in the twentieth century (see Hegel, G.W.F. §5).

Charles Peirce was another important critic of foundationalism, both empiricist and Cartesian. Against the former, Charles Peirce held that no empirical belief is certain - we can be mistaken in any empirical belief - and neither is it unrevisable - we can be reasonably motivated to give up any empirical belief in the light of new evidence. These two points make up part of Peirce's fallibilism (see Fallibilism). The Cartesian programme is criticized on the grounds that wholesale doubt is not a psychologically possible action, so that Descartes' method of securing foundations for knowledge does not succeed (see Peirce, C.S. §2).

8 Twentieth century

The empiricist tradition continued into the twentieth century in sense-datum theories of the sort found in Russell, with special attention paid to knowledge of the external physical world. It was argued that, in any perceptual experience, one is immediately aware of sense-data rather than physical objects. Sense-data are taken to be phenomenal objects having qualities such as colour and shape (see Sense-data). Immediate awareness of sense-data is acquaintance, itself a form of certain knowing, namely, knowing objects rather than propositions about objects. Propositional knowledge of objects is knowledge by description, and it is inferential, based upon acquaintance knowledge of sense-data (see Knowledge by acquaintance and description).

The needed inferences were to be underwritten by analytical phenomenalism, that is, the thesis that all physical-object sentences are analysable into, and so equivalent in meaning to, sets of sense-datum sentences. Given this equivalence, it was felt, inferences from sense-datum sentences to physical-object sentences would be secured as legitimate; thus the problem of the external world was solved.

Related theories were defended by Ayer and C.I. Lewis. Ayer dropped the notion of acquaintance. Sense-data were taken as items of immediate awareness, which typically issued in incorrigible propositional knowledge. Lewis dispensed with sense-data; he expressed the foundational sentences using ordinary idioms such as 'This seems red', but he made the same demands on these as Ayer did of sense-datum sentences. They are certain, and the basis of all other empirical knowledge. As in Russell, inferences from these were supposed sanctioned by analytical phenomenalism.

Interestingly enough, G.E. Moore also defended a sense-datum theory of perception, but did not couple it with an empiricist version of foundationalism. Rather, he defended common sense, which for him included the view that there are many particular material-object propositions which are known immediately and with certainty. For instance, Moore claimed to know, immediately and with certainty, that a certain mantelpiece was closer to his body than a specific bookcase. For Moore, knowledge of this proposition and of many other material-object

propositions need not be based on more secure knowledge of sense-data. In this regard, then, Moore’s view is more a version of epistemic direct realism than it is empiricist foundationalism (see Moore, G.E. §3).

The programme of empiricist foundationalism and analytical phenomenalism was widely criticized. The alleged incorrigibility or certain knowledge of sense-data was influentially attacked by J.L. Austin. All empirical sentences, he argued, are corrigible because in forming a belief about an object, such as is expressed by ‘This is red’, one is classifying the object as among the red things and so is relying on one’s wholly fallible memory of other comparably red items. Moreover, certainty or incorrigibility is not needed for knowledge. Many critics argued that certainty in the sense of lack of actual grounds for doubt was a more adequate analysis of this concept, and in this sense many physical-object sentences would count as certain. Analytical phenomenalism was also criticized, principally by Chisholm (see Phenomenalism §2). He showed that physical-object sentences do not entail sense-datum sentences, and hence are not equivalent to them.

Ayer and Lewis also were in rough agreement on the definition of the concept of knowledge. They held that propositional knowledge is justified true belief, an account shared by many others. Edmund Gettier (1963) argued that this definition was incorrect. His idea was that one could have a true justified belief which is not knowledge in a situation in which one reasons from some already justified beliefs to a new belief that, as it happens, is coincidentally true. Since it would then be a matter of coincidence that one’s belief was correct, it would not count as knowledge, even though it was a justified belief because it was knowingly inferred from already justified beliefs (see Gettier problems).

Gettier’s 1963 article generated a great deal of interest. While some argued that his argument was unsatisfactory, a majority assumed that he was more or less right, and many new analyses of knowledge were proposed, including many which incorporated the justified true belief analysis as a part. What has emerged as perhaps the most promising and least prone to new counterexamples is the defeasibility analysis. The key idea is that of defeated justification: where one is justified in believing a proposition *p* on the basis of evidence *e*, one’s justification is defeated when there is a true proposition *q*, such that the conjunction (*e*&*q*) does not justify *p*. The defeasibility analysis would then be that knowledge is justified, true, undefeated belief. Sophisticated versions of the defeasibility analysis have been worked out in detail by a number of authors, including Klein (1971), Lehrer (1974) and Swain (1972) (see Knowledge, defeasibility theory of).

Closely connected to the concept of knowledge is the concept of a justified belief, and a number of important theories of epistemic justification have been developed, the principal ones being foundational theories, coherence theories and reliability theories (see Justification, epistemic). We have already noted Cartesian and empiricist versions of foundationalism. In recent years some philosophers have defended modest versions of foundationalism. That is, they have defended the view that a belief would be justified if and only if either it were a basic, foundational belief, or were inferrable from basic beliefs. The modesty of the theory would then derive from the fact that basic beliefs need not be certain or incorrigible; it would suffice if the basic beliefs were to be non-inferentially justified. Beliefs are non-inferentially justified when their justification need not result from being based on or inferrable from other justified beliefs.

Many philosophers have found even modest foundationalism suspect, primarily because they have found problematic the notion of a basic, non-inferentially justified belief. Some have accordingly avoided this notion altogether, and developed coherence theories of justified belief. The core idea in all such coherence theories is that a belief is justified if and only if it is a member of a system of beliefs, and this system of beliefs is coherent. A number of different accounts of coherence have been proposed, but most favoured has been that of explanatory coherence. On such a view, some beliefs (explainees) in the coherent system are justified because they are explained by other beliefs in the system; the remaining beliefs in the system are justified in virtue of their role in explaining the explainees. A problem for these theories has been to provide a reasonable way of selecting those beliefs within the system which are to be explained (see Knowledge and justification, coherence theory of).

The most widely discussed reliabilist theory has been the reliable-process theory. The core idea here is that a belief is justified if and only if it is caused by, or causally sustained by, a reliable process. A process is reliable when it has a high truth-ratio; that is, when that process produces more true beliefs than false ones. Typical processes selected as reliable belief-forming or belief-sustaining ones are perception, memory, introspection, and inferring or reasoning.

A problem which has proved especially vexing for supporters of the reliable-process theory is that of generality. Any specific belief is produced (sustained) by a process token which is an instance of many different process types. The generality problem is essentially that of fixing how broadly to individuate the process types in question (see Reliabilism).

9 Recent issues

In a reliable-process theory cognizers may have no knowledge or awareness of the processes which cause or causally sustain their beliefs, or of the reliability of these processes. Most foundational and coherence theories, however, construe the notion of justification in such a way that a person's belief is justified only if they have some access to, or awareness of, whatever it is that serves to justify that belief. Theories with this access condition are generally thought of as internalist theories; those which dispense with the access requirement are externalist. Though widely discussed, no fully adequate resolution of the question of whether an access requirement should be imposed on a theory of justification has yet gained general acceptance (see Internalism and externalism in epistemology).

Proponents of reliable-process theories have typically tried to develop a naturalistic theory (see Naturalized epistemology). Minimally, a naturalistic epistemological theory is one in which key epistemic concepts such as knowledge and justification are analysed or explained in a form which makes use only of non-epistemic concepts. A more radical form of naturalism in epistemology, proposed by Quine, dispenses outright with the normative elements of traditional epistemology, and reconceives the subject as a part of empirical psychology (see Quine, W.V. §2). On this view, epistemology becomes a wholly descriptive discipline, one which studies how beliefs are formed, and how they are related to what we take evidence to be. Whether a minimal or more radical form of naturalized epistemology is acceptable is an open question at present.

Issues in social epistemology have also loomed large recently, as have topics in feminist epistemology. Within the former, two important questions are whether social factors play a role in determining whether a person has knowledge or justified belief, and whether non-individuals such as groups or institutions can be said to have knowledge or justified beliefs (see Social epistemology). Within feminist epistemology a leading question has been whether women acquire knowledge in ways that differ from methods of knowledge acquisition open to men. Another important issue has been whether social and cultural factors as they affect women have a bearing on what it is that women know (see Feminist epistemology).

All these recent developments in epistemology are being vigorously pursued and explored. They have served to expand and enrich the field in ways not appreciated just a few decades ago.

See also: Chinese philosophy; Epistemology, Indian schools of; Epistemology in Islamic philosophy; Knowledge, concept of; Knowledge, Indian views of

GEORGE S. PAPPAS

References and further reading

Adams, M. (1987) *William Ockham*, Notre Dame, IN: University of Notre Dame Press.(Authoritative work; chapters 13 and 14 are especially relevant.)

Alston, W. (1989) *Epistemic Justification*, Ithaca, NY: Cornell University Press. (Defends a modest foundationalism.)

Aquinas, T. (1266-73) *Summa Theologiae*, trans. by Fathers of the English Dominican Province, Westminster: Christian Classics, 5 vols, 1981.(Particularly relevant to the matter of abstracting intelligible species from sense impressions, mainly in Questions 84-7.)

Aquinas, T. (1271-2) *Commentary on Aristotle's Posterior Analytics*, trans. F. Larcher, Albany, NY: Magi Books, 1970.(Contains Aquinas' account of genuine knowledge and demonstrative syllogism.)

Aristotle (*c.* mid 4th century BC) *Posterior Analytics*, ed. and trans. J. Barnes, Oxford: Clarendon Press, 1975. (Contains Aristotle's account of scientific knowledge.)

Aristotle (*c.* mid 4th century BC) *De Anima*, ed. and trans. D. Hamlyn, Oxford: Clarendon Press, 1968. (Aristotle's account of perception and perceptual knowledge is found in Books 2 and 3.)

Augustine (386) *Contra academicos (Against the Academicians)*, ed. and trans. P. King, Indianapolis, IN: Hackett Publishing Company, 1995.(Augustine's extended criticism of academic scepticism, mainly that found in

Cicero.)
Ayer, A.J. (1940) *Foundations of Empirical Knowledge*, London: Macmillan.(A defence of empiricist foundationalism and of analytical phenomenalism.)
Ayer, A.J. (1956) *The Problem of Knowledge*, Baltimore, MD: Pelican.(An account of knowledge as roughly justified true belief.)
Ayers, M. (1993) *Locke: Epistemology and Ontology*, London: Routledge. (Authoritative, scholarly treatment of Locke on all matters pertaining to his epistemology.)
Bayle, P. (1695-7) *Historical and Critical Dictionary, Selections*, trans. R. Popkin, Indianapolis, IN: Hackett Publishing Company, 1991.(Some of the *Dictionary* material strongly influenced Berkeley and Hume.)
Berkeley, G. (1710) *A Treatise Concerning the Principles of Human Knowledge*, ed. K. Winkler, Indianapolis, IN: Hackett Publishing Company, 1982.(One of Berkeley's main writings on epistemological topics.)
Berkeley, G. (1713) *Three Dialogues Between Hylas and Philonous*, ed. R.M. Adams, Indianapolis, IN: Hackett Publishing Company, 1979.(Important for Berkeley's discussion of scepticism.)
BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.(Both an influential criticism of foundationalism and a defence of explanatory coherence.)
Broad, C.D. (1975) *Leibniz - an Introduction*, Cambridge: Cambridge University Press.(A fine introduction to Leibniz by a master expositor, but it is somewhat advanced.)
Burnyeat, M. (1980) 'Can the sceptic live his scepticism?', in M. Schofield, M. Burnyeat and J. Barnes (eds) *Doubt and Dogmatism*, Oxford: Clarendon Press. (Deals with the question of whether those Hellenistics professing scepticism can actually withhold belief in the relevant ways.)
Chisholm, R. (1948) 'The Problem of Empiricism', *Journal of Philosophy* 45 (19): 512-17.(A trenchant criticism of analytical phenomenalism.)
Chisholm, R. (1957) *Perceiving*, Ithaca, NY: Cornell University Press.(Contains an anti-sense-datum account of perception.)
Chisholm, R. (1966) *Theory of Knowledge, Englewood Cliffs, NJ: Prentice Hall*, 2nd edn, 1977.(General discussion which includes a defence of foundationalism.)
Code, L. (1991) *What Can She Know? Feminist Theory and the Construction of Knowledge*, Ithaca, NY: Cornell University Press.(Defence of feminist epistemology as a reorientation of the field of epistemology, by a leading researcher in the area.)
Cornman, J. (1980) *Skepticism, Justification, and Explanation*, Dordrecht: Reidel.(Careful discussion of scepticism and many varieties of theories of justification.)
Descartes, R. (1641) *Meditations on First Philosophy*, in J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny (eds) *The Philosophical Writings of Descartes*, Cambridge: Cambridge University Press, 3 vols, 1984-91.(Now the standard English edition of the main philosophical works.)
Everson, S. (1990) 'Epicurus on the Truth of the Senses', in S. Everson (ed.) *Epistemology*, Cambridge: Cambridge University Press.(Especially helpful on Epicurus' relations to sceptical doctrines.)
Feldman, R. (1985) 'Reliability and justification', *The Monist* 68 (2): 159-74.(Penetrating statement of the generality problem for reliabilism.)
Fine, G. (1978) 'Knowledge and Belief in Republic V', *Archiv fur Geschichte der Philosophie* 58: 121-39.(Deals with knowledge of Forms and the contrasts between knowledge and belief in Plato's *Republic*.)
Fogelin, R. (1985) *Hume's Scepticism in the Treatise of Human Nature*, London: Routledge.(Interprets Hume as an epistemological sceptic.)
Frede, M. (1983) 'Stoics and sceptics on clear and distinct ideas', in M. Burnyeat (ed.) *The Sceptical Tradition*, Berkeley, CA: University of California Press.(A very sympathetic examination of Stoic doctrines; particularly helpful in understanding the role of cognitive impressions.)
Fumerton, R. (1988) 'The internalism/externalism controversy', *Philosophical Perspectives* 2: 442-59.(Good discussion of the matters at issue in this dispute.)
Gettier, E. (1963) 'Is justified true belief knowledge?', *Analysis* 23 (6): 121-3.(This brief paper stimulated more than two decades of work on the analysis of factual knowledge.)
Goldman, A. (1979) 'What is justified belief?', in G. Pappas (ed.) *Justification and Knowledge*, Dordrecht: Reidel.(Exceptionally lucid early statement of a reliable-process theory of justification.)
Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press. (Updated version of reliabilism, along with some attempt to solve the generality problem and a detailed look at the relevance of psychological theories and findings to epistemology.)
Hegel, G.W.F. (1807) *The Phenomenology of Spirit*, trans. A.V. Miller, Oxford: Clarendon Press, 1977.(The introduction contains trenchant criticism of empiricism and Cartesian epistemology.)

Hookway, C.J. (1985) *Peirce*, London: Routledge & Kegan Paul.(Good account of Peirce's views of science and scientific method.)

Hume, D. (1740) *A Treatise of Human Nature*, ed. P. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978.(Hume's views on knowledge and scepticism are found in Book 1.)

Hume, D. (1748) *Enquiry Concerning Human Understanding and Enquiry Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, revised P. Nidditch, Oxford: Clarendon Press, 1975.(Hume here expresses his views in a more accessible manner than in the *Treatise*.)

Irwin, T. (1988) *Aristotle's First Principles*, Oxford: Clarendon.(A fine, detailed treatment of Aristotle's epistemological doctrines.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1964.(This difficult work contains Kant's answer to both empiricists and rationalists. Most relevant is the 'Transcendental Analytic'.)

Kim, J. (1988) 'What is naturalized epistemology?', *Philosophical Perspectives* 2: 381-405.(Helpful discussion of Quine's view, and a defence of the idea that traditional epistemology should not be, and perhaps cannot be, given up.)

Klein, P. (1971) 'A proposed definition of propositional knowledge', *Journal of Philosophy* 68 (16): 471-82.(An early defence of the defeasibility theory.)

Lehrer, K. (1974) *Knowledge*, New York: Oxford University Press.(Contains a powerful criticism of explanatory coherence theories, and a defence of a subjective coherence theory.)

Lehrer, K. (1989) *Thomas Reid*, London: Routledge.(A comprehensive discussion of Reid's philosophy. It includes an excellent account of Reid's criticisms of Hume, and of Reid's defence of common sense.)

Lehrer, K. (1990) *Theory of Knowledge*, Boulder, CO: Westview Press.(Contains Lehrer's latest defence of defeasibility and much besides.)

Leibniz, G. (1765) *New Essays on Human Understanding*, trans. P. Remnant and J. Bennett, Cambridge: Cambridge University Press, 1981.(Comments on and responds to Locke; notable for Leibniz' discussion of innate truths.)

Leibniz, G. (1969) *Philosophical Papers and Letters*, ed. and trans. L.E. Loemker, Dordrecht: Reidel. (Contains most of Leibniz' important philosophical writings.)

Lewis, C.I. (1946) *Analysis of Knowledge and Valuation*, LaSalle, IL: Open Court Publishing.(Defence of a complex form of analytical phenomenalism, and of empiricist foundationalism without sense-data.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P. Nidditch, Oxford: Clarendon Press, 1975. (The best modern edition of Locke's classic. Most important are Books I, II, and IV.)

Long, A. and Sedley, D.N. (eds) (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press. (Contains relevant Epicurean and Stoic texts.)

MacDonald, S. (1992) 'Aquinas's Epistemology', in N. Kretzmann and E. Stump (eds) *Aquinas*, Cambridge: Cambridge University Press.(Comprehensive discussion of Aquinas' scattered epistemological remarks.)

Montaigne, M. de (1576) *Apology for Raimond Sebond*, trans. D. Frame in *The Complete Works of Montaigne*, Stanford: Stanford University Press, 1958.(Very influential early modern text.)

Moore, G.E. (1953) *Some Main Problems of Philosophy*, London: Allen & Unwin. (Contains Moore's account of the common sense view of the world.)

Moore, G.E. (1959) *Philosophical Papers*, London: Allen & Unwin.(This collection contains Moore's famous paper 'A Defence of Common Sense'.)

Moser, P. (1989) *Knowledge and Evidence*, New York: Cambridge University Press. (A defence of foundationalism.)

Pappas, G. (1991) 'Berkeley and Common Sense Realism', *History of Philosophy Quarterly* 8 (1): 27-42. (Discusses Berkeley's commitment to epistemic direct realism.)

Pappas, G. and Swain, M. (1978) *Essays on Knowledge and Justification*, Ithaca, NY: Cornell University Press. (Contains many papers on the analysis of knowledge and on foundational and coherentist theories of justified belief.)

Parkinson, G.H.R. (1954) *Spinoza's Theory of Knowledge*, Oxford: Clarendon Press.(Comprehensive treatment of Spinoza's doctrines about knowledge.)

Peirce, C. (1931-58) *Collected Papers of Charles S. Peirce*, ed. P. Weiss, C. Hartshorne and A. Burks, Cambridge, MA: Harvard University Press, 8 vols.(Volume 5 contains papers most relevant to epistemological issues.)

Pitcher, G. (1977) *Berkeley*, London: Routledge & Kegan Paul.(Penetrating analysis of Berkeley's thought, a modern classic which interprets Berkeley as an empiricist foundationalist.)

Plato (427-347 BC) *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1961.(Contains all of Plato's dialogues in English translation. See especially the *Meno*, *Theatetus*, *Cratylus* and *Republic*, discussed in §1 above.)

Popkin, R. (1979) *The History of Scepticism from Erasmus to Spinoza*, Berkeley, CA: University of California Press.(A great source of information, both historical and philosophical.)

Quine, W.V. (1969) 'Epistemology naturalized', in *Ontological Relativity and Other Essays*, New York: Columbia University Press.(Defends abandonment of traditional epistemology and the view that epistemology should be taken as a branch of empirical psychology.)

Reid, T. (1785) *Essays on the Intellectual Powers*, ed. B. Brody, Cambridge, MA: MIT Press, 1969.(Reid's systematic treatment of epistemological themes.)

Russell, B. (1912) *Problems of Philosophy*, London: Williams & Norgate.(Discusses and defends the sense-datum theory and empiricist foundationalism; also contains Russell's famous distinction between knowledge by acquaintance and knowledge by description.)

Russell, B. (1914) *Knowledge of the External World*, London: Allen & Unwin.(As with his 1912 work, this also discusses and defends the sense-datum theory and empiricist foundationalism; analytical phenomenalism is defended under the term 'logical constructions'.)

Russell, B. (1918) *Mysticism and Logic*, New York: Longmans.(Discusses and defends the sense-datum theory and empiricist foundationalism - see especially Essays 7, 8 and 10.)

Schmitt, F. (1994) *Socializing Epistemology*, Lanham, MD: Rowman & Littlefield. (Contains many important recent papers on social epistemology, with notable attempts to map out the field in Schmitt's introduction and in papers by Kornblith, Kitcher and Schmitt.)

Sellars, W. (1963) 'Empiricism and the philosophy of mind', in *Science, Perception and Reality*, New York: Humanities.(A most influential criticism of foundationalism.)

Sextus Empiricus (*c.* AD 200) *Outlines of Pyrrhonism* and *Against the Mathematicians*, ed. R. Bury, London: Heinemann, 4 vols, 1933-49.(Complete offering of Sextus' writings)

Shope, R. (1983) *The Analysis of Knowing*, Princeton, NJ: Princeton University Press. (A most useful in-depth survey of a wealth of material on the concept of knowledge.)

Spinoza, B. (1677) *Ethics* and *Treatise on the Improvement of the Intellect*, in *The Collected Works of Spinoza*, ed. E. Curley, Princeton, NJ: Princeton University Press, 2 vols, 1985.(The primary texts.)

Stroud, B. (1977) *Hume*, London: Routledge & Kegan Paul.(Fine discussion of all Hume's main ideas and arguments.)

Swain, M. (1974) 'Epistemic defeasibility', *American Philosophical Quarterly* 11: 15-25.(The Klein and Swain papers deal with defeasibility and knowledge in roughly the same way; both are very polished complex discussions of the issues.)

Taylor, C.C.W. (1990) 'Aristotle's Epistemology', in S. Everson (ed.) *Epistemology*, Cambridge: Cambridge University Press.(Very clear discussion, easily accessible to the non-specialist.)

William of Ockham (*c.*1287-1347) *Ockham: Philosophical Writings*, ed. P. Boehner, Indianapolis, IN: Bobbs-Merrill, 1964.(Good translations of major writings in philosophy.)

William of Ockham (1322-7) *Quodlibeta Septem*, in R. McKeon (ed.) *Selections from Medieval Philosophers*, New York: Charles Scribner's Sons, 1930.(The seven 'quodlibeta' are directly concerned with intuitive cognition.)

Williams, B. (1978) *Descartes: The Project of Pure Enquiry*, Harmondsworth: Penguin.(Outstanding and detailed treatment of all of the main ideas in Descartes' epistemology; this book is very helpful in understanding Descartes' relation to sceptical arguments.)

Epistemology, Indian schools of

Each classical Indian philosophical school classifies and defines itself with reference to a foundational text or figure, through elaboration of inherited positions, and by disputing the views of other schools. Moreover, the schools have literatures that define them in a most concrete sense, literatures that in some cases stretch across twenty centuries and comprise hundreds of texts. And without exception, every school takes a stance on the nature of knowledge and justification, if only, as with the Mādhyamika Buddhist, to attack the positions of others. A blend of epistemology, ontology or metaphysics, and, sometimes, religious or ethical teachings constitutes the view of most schools, and sometimes only very subtle shifts concerning a single issue differentiate one school's stance from another's.

Relabelling schools of Indian epistemology using terminology forged in Western traditions ('foundationalism', 'coherentism', and so on) risks skewing the priorities of classical disputants and distorting classical debates. Nevertheless, there are positions shared across some of the schools, as well as refinements of position that apparently because of merit received greater attention in classical discussions and appear to deserve it still. Given the broad context of world philosophy, selectivity cannot be free from bias stemming from a sense of reverberation with non-Indian traditions of thought. With these warnings in mind, we may proceed to examine three important approaches within classical Indian philosophy to questions of epistemology.

First, the late Yogācāra Buddhist philosophers, Dignāga (b. circa *480), Dharmakīrti (*c.*600-660) and followers, present a complex first-person approach to questions about knowledge that is constrained by an anti-metaphysical theme (found in earlier Buddhist treatises), along with a phenomenalism that grows out of a vivid sense of the real possibility of* nirvāṇa *experience as the supreme good. Their thought also exhibits an academic strand that is sensitive to non-Buddhist philosophical discussions. Second, a reliabilism identifying sources of veridical awareness is the most distinctive, and most central, approach to epistemology within classical Indian philosophy as a whole. Even the Yogācāra first-person approach gets framed in terms of reliable sources (perception and inference as* pramāṇas, *'sources of knowledge'). Philosophers of diverse allegiance make contributions to what may be called this* field *of thought (as opposed to an* approach*), since, to repeat, it is the philosophical mainstream. However, the Nyāya school (the 'Logic' school) leads in most periods. Finally, the Brahmanical school known as Mīmāṃsā ('Exegesis'), supplemented in particular by centuries of reflection under an Advaita Vedānta flag, develops what can be called an ethics of belief, namely, that we should accept what we see (for example) as real (and the propositional content of perceptual awarenesses as true), what we are told by another as true, what we infer as true, and so on, except under specific circumstances that prove a proposition false or at least draw it into question. The (Nyāya) epistemological mainstream is moved to incorporate a variation on this position; for Mīmāṃsā and Advaita, 'self-certification' (*svataḥprāmāṇya*), or the intrinsic veridicality of cognition, defines an alternative approach to questions about knowledge, awareness and a presumed obligation to believe.*

1 Buddhist pragmatism and coherentism

The most important context for the epistemology of the late Yogācārins (*c*.450 on), sometimes called the Buddhist logicians, may well be non-Buddhist developments, or views about the criteria of successful debate mutually developed in both Buddhist and non-Buddhist circles. But Buddhist soteriology ('mystical enlightenment theory'), including an anti-metaphysical attitude, appears to determine much that is distinctive in Dignāga's school's approach to epistemology. The *summum bonum* is conceived within Buddhism as an immediate experience, a state of awareness attained through concentrated meditation (*samādhi*) that is free from such disturbances as the inner voice of thought and feelings of desire. Possibly through a confusion of thought as an occurrent psychological phenomenon with conceptualization in general, Buddhist teachings of meditative practice embrace, or come to embrace, a broad anti-intellectualism, especially concerning speculative efforts of philosophy. With thought devalued, sceptical arguments found a receptive audience among Buddhist practitioners. Immediate experience, in contrast, is highly regarded. Indeed, *nirvāṇa* is immediate experience purged of distractions, in particular thought and desire. The very designation 'Yogācāra' indicates those devoted to meditation and other yogic practices. The school's focus on phenomenal presentations thus need not be taken to stem from a radically empiricist approach to questions of evidence, as in Western phenomenalism. It appears to have moorings in meditative mysticism instead.

Now sense perception in its raw immediacy and stripped (if this is possible) of all relatings to previous experiences, seems to present unique particulars; at least, this contention is made by Dignāga and other Yogācārins. The world of thought and desire is not the world of immediate experience, especially not the world of immediate experience purged of psychological defilements. The world of immediate experience is a world of ineffably interdependent and strictly unconceptualizable particulars, and perception accessible to thought and discussion is already infected with concepts, with desire-based, imaginational relatings (*kalpanās*) to other experiences. Thus the best we can do concerning generalities and concept formation (all concepts are general) and questions of warrant about the claims that purport to classify what are really unclassifiable particulars is to take a pragmatic attitude underpinned by a broad scepticism about what things are in reality. To best capture the exclusion (*apoha*) of generalities from particulars, Bessie's being a cow, for example, should be expressed as her not being a non-cow. We cognize things in accordance with what we want to do with them, and the success of actions based on our cognitions proves whether those cognitions are adequate (pragmatically true) or not.

Of course, success in action is itself made known to us through our own cognitions. The final touchstone of a cognition's veridicality is its agreement with other cognitions, or, in some formulations, lack of failure to agree. Incoherence is a cognitive dissonance that means that we hold something in error. We may well be in error about almost everything, but we do not notice it until there is a (jarring) cognitive failure of agreement. Now satisfaction of desire is also mediated by sense experience, and so some scholars have seen an implicit foundationalism (knowledge as founded in sense experiences). But Dharmakīrti at least is a coherentist in the sense that no empirical proposition is sacrosanct; any proposition can be defeated and overridden by considerations of overall coherence. Neither he, nor any Yogācārin, is a pure coherentist, however; it seems a judgment would, by virtue of being perceptual, be accorded warrant of at least a minimal sort.

The Yogācāra approach to logic appears to be broadly conventionalist, although sometimes logical principles are discussed as though they were discoveries. In the world of generalities, Bessie's not being a non-cow is always equivalent to her being a cow, for we set up our logic to have invariable rules. Again, the presentations of immediate experience do not obey logical principles. But when we worry about cognitive agreement, we reason according to rules. Dignāga concocts a table of inferential relations that is abstract and formal, unlike the treatment in Nyāya, where inference is thought to be underpinned by natural realities, by relations occurring in the world independently of us as cognizers.

Finally, much in the logical and epistemological texts of the late Yogācārins seems to grow out of a concern with criteria for proper debate, including, to be sure, what makes an argument cogent, but also criteria for appropriate procedure. In this and in many details of logic and debate theory - also even in the epistemology of sense perception - there is a lot of overlap with non-Buddhist texts. The extent to which academicism flourished among classical philosophers is underappreciated by modern scholars, who tend to look for something distinctively Buddhist in everything a Buddhist epistemologist says. But probably much or even most of Yogācāra epistemology was not considered to flow out of peculiarly Buddhist views. Dignāga, Dharmakīrti and company often seem to write simply as epistemologists, not as adherents to a Buddhist epistemological school.

2 Nyāya reliabilism

At the heart of the Nyāya ('Logic') approach to epistemology is the view that a proposition-laden awareness is to be counted veridical - for the purposes of resolving a doubt or an issue in dispute - just in case it is the result of a reliable cognitive process, *pramāṇa*. (According to Nyāya, what the veridicality of an awareness amounts to is, roughly, a reflecting of, or corresponding to, a fact; our concern, however, is not what it is for an awareness to *be* veridical, but why an awareness is to *be regarded as* veridical. This latter is a matter of reliabilism in the Logic approach, although a qualification concerning context of inquiry has to be made: see below.) Nyāya identifies four such reliable processes: perception, inference, vocabulary acquisition through an analogy drawn, and testimony. Other schools identify fewer *pramāṇas* as independent 'means of knowledge' (for example, Yogācārins accept only two, perception and inference); still others accept more (for example, Mīmāṃsā accepts six) (see Knowledge, Indian views of).

Logic's reliabilism is a causal theory of knowing; it is articulated in the *Nyāyasūtra* (*c.* AD 150; although it did not reach its present form until about 400) and developed in commentaries and extra-commentarial literature stretching into modern times. We rely on our cognition to get what we want, and both our getting it and knowing it are

natural, causal processes.

Logic philosophers may also be said to be empiricists in the precise sense of viewing perception as our principal cognitive link with the world. It is the process generating reliable awarenesses *par excellence*. When we acquire a new piece of vocabulary through someone drawing an analogy for us, we must have already acquired analogue terms experientially, by definitions given through ostension. Testimony is reliable only if the testifier (or the first in a chain of testifiers) knows through perception (or through an inference itself perceptually based) what is being testified. Not only may the result of an inference be defeated by perception, but the crucial general rule, or 'pervasion' (*vyāpti*, wherever *F* there *G*) is established by wide perceptual experience. And, of course, through sight of a single counterinstance (an *F* but no *G*) a putative pervasion would be disproved.

An ontological realism and a fallibilism about knowledge are also important in the overall epistemological approach of the Logic school. A pervasion is held to obtain in nature, and for this reason one could, strictly speaking, be grasped from a single exhibited connection. For example, from a single instance of the sight of smoke rising from fire, the pervasion 'wherever smoke there fire' could be grasped; or, from viewing a horse and a cow, one can know that a horse is not a cow and - to express the pervasion known, here a negative one - 'whatever is a cow is not a horse, and conversely'. Repeated experience of the connection (along with no experience of the presence of an *F* without a *G*) increases one's confidence. Perception, inference and the rest are natural causal process, and generally reliable, but on a particular occasion there may be no way of knowing whether we are right or wrong. Knowledge is a factual matter, and we may well not know that we know something or other when in fact we do know it. That is to say, a lower-order fact may be being cognized ('That is a pot' as known), without one's becoming aware that one knows ('I am knowing "That is a pot"' as itself not known), just as there are plenty of other types of fact of which we are unaware. With the exception of apperception (see Awareness in Indian thought §3), certification, according to Nyāya, is a process external to the knowledge episode that is certified - a position giving rise to disputes between Logicians and Exegetes (see §3).

The difference between inquiry about a cognition's source and ultimate certification appears to be a matter of context. Most generally, certification depends on the success of action undertaken on the basis of a (claim-laden) cognition, and sometimes we find out in action that, for example, what we thought was a veridical perception is an illusion (one cannot drink sand). But in a particular context of inquiry it may be enough to trace a cognition to its source, or to turn to the recognized sources for new information. The entire reliabilism programme of Nyāya is said to be grounded in the success of action undertaken on the basis of identified sources of knowledge. We know perception *et al.* are generally reliable by an inference based on the usual success of acts guided by perception *et al.*

3 Mīmāṃsā self-certificationalism

The Mīmāṃsā ('Exegesis') approach to epistemology emerges in the context of a challenge to religious practices and social mores now known as Hindu from, in particular, Buddhism. The teachings of the Vedas and Vedic literature were propagated by a hereditary class of priests, or Brahmans, comprising the highest of the traditional Hindu castes. Buddhism (also Jainism) rejects caste along with the Brahmanical canon. Buddhists claim that their religious teachings are founded on nothing other than experience (albeit mystical), the *nirvāṇa* experience that makes a *buddha* an expert about spiritual matters. Similar moves were made by some with regard to the Vedas. But the genius of Mīmāṃsā is to defend the Brahmanical canon by an appeal to the common practice of taking even unexamined beliefs to be correct. That the Vedas' truth is secured by a Divine Author is a position at odds with the Exegetes' polytheism and their sense of the Vedas as a moral authority superior to the gods. The Vedas are so holy and profound that no one, neither the greatest of mystics nor a god, could have composed them. The Vedas are *apauruṣeya*, 'without a person as author'. The ethics of belief that arises among Exegetes proposes more than a right to believe Vedic teachings; it proposes that we have a duty to follow Vedic injunctions to perform certain sacrifices and lead a certain form of life. Just as in perceiving an object such as a cow we require no certification extrinsic to the perception itself to act appropriately with respect to the cow, so in hearing Vedic injunctions we requires no extrinsic certification with respect to what we know we ought to do.

In classical epistemological debates, Exegetes, led by Kumārila (*c.*670) and Prabhākara (*c.*700), attend, like their rivals, to *pramāṇas*, 'knowledge sources'. The point of this focus, according to Logic philosophers, is to be able to trace a cognition or claim to a reliable source as a way of dispelling doubt and ending debate; Yogācārins turn to

coherence considerations as the test of truth and veridicality, with perception as the source responsible for an initial credibility or warrant that by agreement or disagreement with other cognitions may be strengthened or overruled. In contrast, Exegetes insist that the initial credibility of a cognition is practically absolute. Strictly speaking, no awareness is nonveridical; no cognition is absolutely wrong, but is at worst a confusion. Error can and does occur. But what is called a perceptual illusion, for example, is a confusion, or blending, of a veridical perception and a correct memory in such a way that it appears that there is a single cognition (misrepresenting the world), whereas in fact there are two. Error results from a departure from a *pramāṇa*'s normal functioning and a failure, for example, to distinguish a perception and a memory (the error of seeing mother-of-pearl as silver combines a perception of an object that is indeed there - mother-of-pearl - and a memory of silver that one has indeed previously experienced).

Exegetes promote ideas of demarcation concerning legitimate spheres of *pramāṇa* operation and causal considerations with respect to each sphere. For example, sense experience is not a legitimate source of ethical knowledge, for only an existent object can be the cause of a perception, and ethics concerns future acts.

Such reflection as this is put forth to dispel confusion, not to establish a legitimate pedigree. Inquiry concerning legitimate epistemic pedigree is not itself normally legitimate: the absolutely central thesis of self-certification does not rule out *pramāṇa* inquiry, but it allows us to take a genial, nonsceptical and inclusivist attitude (finding as we do six sources of knowledge), and to get clear about distinctions concerning what is known and how we know it.

There are different versions of self-certification (*svataḥprāmāṇya*) among Exegetes. According to Prabhākara, every veridical cognition is known to be so intrinsically, just by the occurrence of the cognition itself. Nonveridicality, however, is known extrinsically. Thus the same causal nexus that produces a veridical cognition produces knowledge of its veridicality. But when we discover an error, we discover it on extrinsic grounds. On an alternative view deriving from Kumārila, veridicality is known through a process of inference: this is the same process of inference whereby a cognition itself is known as having occurred. A cognition, which is an act, produces a feature in the object it cognizes, namely 'cognizedness', and then from apprehension of this feature, both the original cognition and its veridicality are known. (Similarly, cooking is an act that produces a feature in its object, namely 'being cooked'.) Thus on Kumārila's view, too, there is no difference between knowing a cognition as a fact and knowing it as veridical. Certification is intrinsic to a cognition's occurrence, that is, with cognitions that are veridical. With respect to knowledge of nonveridicality, extrinsic certification is required - as in Prabhākara's view.

Briefly we may note a Logic rejoinder, which is that normally there is no call to wonder about a cognition's veridicality; normally we proceed to act without wondering, and rightly without wondering, whether the cognition guiding an act is reliable. An assumption of veridicality is the epistemological default. But under circumstances that call for reflection about certification, we do not find it in the cognition itself, for that would make nonveridical cognitions impossible. Logic philosophers also hold that there is an excellence (*guṇa*) in the causal process resulting in a veridical cognition that may be said to account for its veridicality, just as there is a reliability-undercutting fault (*doṣa*) in the process resulting in a cognition that is nonveridical. Exegetes have a pretty easy time showing that this is a cumbersome conception. On the score of what cognition reveals, both schools are realist and learn and borrow from one another.

Finally, Advaita Vedāntins, generally speaking, endorse the Mīmāṃsā approach to epistemology, especially that of the Prabhākara camp. Indeed, in disputes with Logic philosophers and others, Advaitins bring out new ramifications of self-certificationalism and interesting attacks on rival positions. However, in contrast to the common-sense realism of Mīmāṃsā concerning the pots and cows of the world, Advaitins are intent upon upholding the sole reality of the self, which is identified with the Absolute, Brahman. For Advaitins, self-certificationalism proves to be an immensely important mainstay in an overall idealist polemic (see Monism, Indian §§1-2).

See also: Buddhism, Yogācāra school of §§1-4; Error and illusion, Indian conceptions of; Mīmāṃsā §§2-3; Nyāya-Vaiśeṣika §6; Sāōkhya §4

STEPHEN H. PHILLIPS

References and further reading

Bhatt, G.P. (1962) *The Basic Ways of Knowing*, Delhi: Motilal Banarsidass, 2nd edn, 1989.(A helpful study of, in particular, Kumārila's Mīmāṃsā epistemology, although Bhatt makes the common mistake of rendering *jñāna* as 'knowledge' - instead of 'cognition' or 'awareness' - and thus speaks of true and false knowledges.)

Bhattacharyya, S. (1987) *Doubt, Belief and Knowledge*, New Delhi: Indian Council of Philosophical Research and Allied Publishers.(Lucid reflection on late developments within the Logic school.)

Dharmarāja (*c.*1600) *Vedāntaparibhāṣā*, trans. and ed. S.S.S. Sastri, *Vedāntaparibhāṣā by Dharmarāja Adhvarin*, Madras: Adyar Library and Research Centre, 1971.(An excellent translation of an important late classical Advaita text with an epistemological slant.)

Hayes, R.P. (1988) *Dignāga on the Interpretation of Signs*, Dordrecht: Kluwer. (Provides a good sense of the overall Yogācāra epistemological approach.)

Kumārila (*c.*670) *Ślokavārttika*, trans. G. Jha, Delhi: Sri Satguru, 1983.(The only available English translation of the work of the important Mīmāṃsaka philosopher; this is a reprint of a 1908 edition.)

Matilal, B.K. (1986) *Perception: An Essay on Classical Indian Theories of Knowledge,* Oxford: Oxford University Press.(Though Matilal defends Nyāya, his discussion ranges penetratingly over the whole of classical Indian theory of knowledge.)

Matilal, B.K. and Evans, R.D. (eds) (1986) *Buddhist Logic and Epistemology*, Dordrecht: Kluwer.(Contains several excellent papers on the Yogācāra logicians and epistemologists.)

Mohanty, J.N. (1966) *Gaōgeśa's Theory of Truth*, Delhi: Motilal Banarsidass, 1989. (Excellent investigation of, especially, the late Nyāya and Mīmāṃsādebate over certification.)

Equality

Equality has long been a source of political and philosophical controversy. A central question about equality is how one might link empirical or moral claims about the extent to which persons are equal to judgements about the moral acceptability or unacceptability of social inequalities, and in particular how far considerations of equality license social action to bring about greater social equality. A traditional liberal argument holds that approximate equality of human strength makes it prudent for humans to place themselves under a common political authority, thus producing a justification for equality before the law. But any generalization of this argument ignores the cases where strength is unequal and the resulting balance of power unjust. Equality of worth is a principle recognized in many philosophical traditions, but its broad acceptance leaves open many problems of interpretation. In particular, it is not clear how far the principle calls for greater equality of social conditions. Persons may derive a sense of worth from enjoying the fruits of their labour, and this will legitimately block some redistribution; certain inequalities may work to everyone's advantage; and the impartial concern of the equality principle may be at odds with the sense of ourselves as persons with specific attachments. In this context, some have wanted to soften the interpretation of equality to mean equality of opportunity or merely that inequalities should not be cumulative, although how far these moves are justified is a matter for dispute. By contrast, challenges to the equality principle from considerations of incentives, desert or difference can more easily be met.

1 The idea of equality

Nearly 2,500 years ago Aristotle remarked in *The Politics* (336-22 BC) that disputes over equality and inequality were generally behind the warfare within states. Nowadays in topics as diverse as the distribution of income and wealth, access to public services, the distribution of work and employment opportunities, the political representation of different social groups and the control of natural resources among nations, the issue of equality plays a central, but controversial, role.

The idea of equality occurs in political philosophy in three main ways. First, it has sometimes been used as a purported description of certain features of human life in society, most notably in the claim that human beings are approximately equal in strength. Second, it is used as a principle of action to the effect that persons should be treated as having equal moral worth. Third, it is used to indicate a supposedly desirable set of social conditions, for example 'one person, one vote' or a more equal distribution of income. A central issue is how far one can employ premises drawn from the first two senses of the term to evaluate social, economic and political inequalities in the third sense. In particular the question arises as to how far social inequalities are just, or at least justifiable (see Justice §5).

To speak of equality without qualification is to speak elliptically. Strictly speaking, equality is a relation that holds between objects or persons in respect of some common characteristic that they share. In most respects, human beings are unequal: they exhibit differences of height, weight, intelligence, manual dexterity, earning capacity and so on. Moreover, if society or governments were to seek to equalize some goods or resources, for example income, they would render unequal some other aspect of social relations, for example hourly rates of pay when people work for different lengths of time.

The upshot of these logical points is that the idea of equality needs to be embedded within a broader theory of politics and society in order to be given a specific content. No political theory aims at equality pure and simple. It aims instead at specific types of equality thought to be morally or socially important.

2 Equality of strength

Given the diversity of human beings, can it ever be said that all human beings are equal in any respect? In *Leviathan* (1651), Hobbes suggested that, although differing in bodily and mental capacities, human beings were equal in that it would be always possible for one person to kill another, an idea picked up by Hume in the eighteenth century and H.L.A. Hart in the twentieth.

According to this tradition, such rough equality of strength makes it prudent or rational for human beings to place themselves under a common body of rules comprising a system of mutual forbearance and agreement. Thus, despite persistent differences in wealth and status in society, equality of obligation under the law flows from

equality of strength in some hypothesized pre-legal situation. Political agents should therefore recognize a common system of authority not merely as wise but also as just.

This line of thought can be carried too far, however. For the equality of strength condition only holds under rather specific circumstances, and where it does not hold the weak are vulnerable. One intuitive idea of justice is that bargaining advantage should not be allowed to prevail where this would threaten certain fundamental interests of persons; political power will thus only be legitimate when it protects the vulnerable (see Justice §4).

3 Equality of moral worth

If we cannot establish a normatively adequate theory of politics solely on the notion of an approximate equality of strength, the idea of equality will have to be given an irreducible moral content. This introduces the second sense of the idea: namely, equality means that all persons are to be thought of as having the same moral worth and are entitled to equal respect and consideration in the treatment of their interests. Although associated with contemporary liberal political thought, both in its utilitarian and its Kantian variants (see Liberalism), this idea has also found a place in many other philosophical and religious traditions, including Jewish and Christian political thought, the Stoic tradition, and the Confucianism associated with Mencius. Currents of thought within Islam, as well as breakaway sects from Hinduism, have also advanced the idea that there is an equal moral worth of all persons. However, can any specific implications about the ordering of social relations be derived from a principle that thinkers of such varying political and philosophical persuasions have endorsed?

One issue here concerns the terms in which we spell out the implications of greater equality. Greater social equality in principle leaves open the question of what exactly is to be equalized: should it be resources, welfare or some other aspect of human life? There are sound arguments for taking any one of these possibilities as the basis of public policy.

However the issue of measurement is resolved, the central justificatory arguments for equality claim that the principle of equal moral worth places limits on what actions a state may pursue (for example, it may not privilege the interests of some at the expense of others), and that a sense of personal worth depends upon the political, social and economic institutions in which members of society participate. Yet, it is easy to see that, although in general this claim may be true, there is no simple way in which the precise nature of these connections can be ascertained.

Suppose it is allowed that people's sense of self-worth depends upon the social institutions in which they participate, so that poverty and social deprivation will undermine a person's sense of self-worth. The extent to which this assumption will license a general equalization of social and economic resources is not obvious. Thus, some libertarians have claimed that such equalization will require practices of redistribution that in effect treat the wealthy as mere means to the attainment of some social goal, in a way that is inconsistent with the respect for persons that the principle of equal worth requires (see Libertarianism §3).

In this simple form the libertarian argument does not carry much conviction, since it does not show why requiring persons to fulfil putative obligations under a principle of justice is to treat them merely as means. However, it does at least indicate that there may be constraints on the way in which the practice of redistribution is accomplished and that persons should, by and large, enjoy the fruits of their own labour.

Quite apart from these libertarian objections to the equalization of property, it has also been argued that the principle of equality it should be recognized that there may be a general advantage in some forms of social inequality. Thus, Rawls (1972) has consistently argued that it would be irrational for persons seeking to advance their own interests to prefer a more equal distribution of resources to an unequal distribution of a larger stock that left the poorest persons better off than they would be under the scheme of equality. By this argument the principle of equal moral worth is seen to coexist with permissible inequalities in the allocation of resources or welfare in society.

Further complications stem from the notion of personality itself. It may be argued that the sense of ourselves as persons derives from the special attachments that we form to other persons, or particular places and institutions. Our sense of moral worth depends therefore on subjectively meaningful attachments that limit the extent to which we can take a general impartial concern in the interests of others. For example, if all people are of equal moral worth, it would seem that large-scale international redistribution is called for to realize the principle, but such a

proposal could cut across the subjective attachments that persons form within the political communities of which they are members.

The distinctive feature of all three of these sorts of argument is that they do not simply counterpose the principle of equality to other principles of social organization, but seek instead to develop different aspects of that principle to cut the inference from equal moral worth to an equal allocation of rights and goods in society.

In response, one option is to weaken the implications of the principle of equality. For example, perhaps the principle of equal worth does not imply an equalization of rights and goods taken as a whole, but merely that there should be no discrimination in their allocation, so that relative advantages are not unfairly achieved. Similarly, it may be argued that the principle of equality merely requires that persons be given an equal opportunity to acquire unequal social and economic advantages. A particularly influential argument in this mode has been the claim that it is not social and economic inequality in itself that is objectionable, but the accumulation of social and economic inequalities across the different dimensions of social life. Related to this argument is the claim that it may be important to reduce or eliminate only certain social inequalities - in particular those that relate to access to health care or education - in order to be consistent with the principle of equal moral worth.

The most natural way to deal with this variety of interpretations is to allow that one can identify not one but a family of political positions, all of which claim to embody the principle of equal moral worth. The choice between these positions in part depends upon broader issues of social theory (for example, the extent to which one believes that inequalities in different dimensions can be kept separate) and in part depends upon the exercise of judgement (for example, the extent to which one judges that a sense of personal worth depends upon persons being able to retain the fruits of their labour as distinct from enjoying access to collectively shared rights and goods).

4 Challenges to equality

There have been a number of challenges to the moral claims of the equality principle. One familiar argument refers to the supposed disincentives to work that would be created by the practice of redistribution from the more productive to the less productive. From one point of view this can be regarded as a pragmatic question, with policy makers having to determine what the balance should be between greater productivity and greater equality. The Rawlsian difference principle, according to which inequalities should be allowed provided they raise the incomes of the least advantaged, can be interpreted here as one possible, but attractive, compromise on this question (see Rawls, J. §1).

However, the issue of incentives also raises matters of principle, since one way to avoid disincentive effects would be to adopt a scheme of taxation based not on labour but on ability. Although difficult to operate, such a scheme would in principle impose a lump-sum tax (an inherited debt) upon the more able that they would have to work to pay off. Clearly, the problem of work incentives in this case leads straight back to some of the problems of combining equality with a sense of personal attachment that were identified in the previous section.

A second principle that is often set against that of equality is the principle of merit or desert. This principle does provide a direct challenge to equality, since it requires that goods be distributed in proportion to the merits of those receiving them - with the best flute players receiving the best flutes, as Aristotle put it. However, the notion of desert is not without its own problems, and it may be argued that its application only makes sense within the context of an ongoing practice, like an orchestra with its own standards of performance, and not across society as a whole (see Desert and merit).

The final challenge to the principle of equality has come from the appeal to the notion of difference. Here the argument is that equality implies a uniformity of treatment, whereas interests in society are in fact plural. Thus, instead of basing public decision making on the principle of one person, one vote, a principle of difference would require some groups to be given a special say in certain matters of public policy, for example women a veto over changes to the law on abortion.

The extent to which the appeal to difference mounts a serious challenge to the principle of equality is a matter for dispute, however. Requiring equal consideration of the interests of all persons is compatible with recognizing that over some questions the interests of certain groups are more deeply engaged than others, as can quite easily be seen in the protection of the rights of local communities to take decisions that affect them especially. Thus, to

allow that there may be a special say for certain groups in matters of public policy may be compatible with the principle of equality, provided that the acknowledgement of the possibility is not unique to a particular group.

Despite these challenges, therefore, the principle of equality still remains as a central principle of modern political thought. But its very centrality means that continuing discussion over its exact interpretation and implications is likely for a long time to come.

ALBERT WEALE

References and further reading

Aristotle (*c*. mid 4th century BC) *The Politics*, trans. H. Rackham, Cambridge, MA: Harvard University Press, London: Heinemann, 1932.(Book 5 offers Aristotle’s thoughts on equality and inequality as the source of political revolution.)

Barry, B.M. (1989) *Theories of Justice*, London: Harvester Wheatsheaf.(An extended discussion of the implications for equality and social justice of the equality of strength argument and its weaknesses.)

Dworkin, R. (1981a) ‘What is Equality? Part 1: Equality of Welfare’, *Philosophy and Public Affairs* 10 (3): 185-246.(One of two articles arguing that equality should govern resources, not welfare.)

Dworkin, R. (1981b) ’What is Equality? Part 2: Equality of Resources’, *Philosophy and Public Affairs* 10 (4): 283-345.(The second of Dworkin’s articles arguing that equality should govern resources, not welfare.)

Goodin, R.E. (1985) *Protecting the Vulnerable*, Chicago, IL, and London: University of Chicago Press.(An analysis of the intuition that inequalities of strength should not be exploited.)

Hart H.L.A. (1962) *The Concept of Law*, Oxford: Clarendon Press.(Chapter 8 contains the best modern statement of the implications of the equal strength view discussed in §2.)

Hobbes, T. (1651) *Leviathan*, ed. M. Oakeshott, Oxford: Blackwell.(States the argument that humans are approximately equal in strength and draws conclusions.)

Hume, D. (1742) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press, 3rd edn, 1975. (Makes equality of strength one of the circumstances of justice.)

Kymlicka, W. (1990) *Contemporary Political Philosophy*, Oxford: Clarendon Press.(Argues that leading contemporary accounts of justice are all egalitarian in the sense that they conceive of persons as equals.)

Nagel, T. (1979) ’Equality’, *Mortal Questions*, Cambridge: Cambridge University Press, 106-27.(Argues that equality is tied to social decision procedures that seek to reflect the view of all affected parties.)

Nagel, T. (1991) *Equality and Partiality*, Oxford: Oxford University Press.(The source of the argument in §3 that equality may conflict with the subjective sense of personal attachments.)

Nozick, R. (1974) *Anarchy, State and Utopia*, Oxford: Blackwell.(Part 2 contains a statement of the libertarian case against equality of income and wealth, but on the basis of a Kantian understanding of persons.)

Nussbaum, M. and Sen, A. (1993) *The Quality of Life*, Oxford: Clarendon Press.(A useful collection reviewing the issue of what the metric of redistribution should be.)

Rawls, J. (1972) *A Theory of Justice*, Oxford: Clarendon Press.(A classic of twentieth-century political philosophy which is the source of the argument in §3 that benefit to the least well-off might justify certain inequalities.)

Vlastos, G. (1962) ’Justice and Equality’, in R.B. Brandt (ed.) *Social Justice*, Englewood Cliffs, NJ: Prentice Hall. (A beautifully lucid statement of the equal worth view.)

Walzer, M. (1983) *Spheres of Justice*, Oxford: Martin Robertson.(A statement of the view that it is cumulative inequalities, rather than inequality itself, that is morally objectionable.)

Williams, B. (1962) ’The Idea of Equality’, in P. Laslett and W.G. Runciman (eds) *Philosophy, Politics and Society*, second series, Oxford: Blackwell.(A statement of the equal worth view, with some interesting claims about ‘irrelevant’ bases of treatment.)

Young, I.M. (1991) *Justice and the Politics of Difference*, Princeton, NJ: Princeton University Press.(A statement of the view that difference calls into question universalistic assumptions about equality of citizenship.)

Erasmus, Desiderius (*c.*1466-1536)

Although Erasmus was not a systematic philosopher, he gave a philosophical cast to many of his writings. He believed in the human capacity for self-improvement through education and in the relative preponderance of nurture over nature. Ideally, education promoted docta pietas*, a combination of piety and learning. Erasmus' political thought is dominated by his vision of universal peace and the notions of consensus and consent, which he sees as the basis of the state. At the same time he upholds the ideal of the patriarchal prince, a godlike figure to his people, but accountable to God in turn. Erasmus' epistemology is characterized by scepticism. He advocates collating arguments on both sides of a question but suspending judgment. His scepticism does not extend to articles of faith, however. He believes in absolute knowledge through revelation and reserves calculations of probability for cases that are not settled by the authority of Scripture or the doctrinal pronouncements of the Church, the conduit of divine revelation. Erasmus' pioneering efforts as a textual critic of the Bible and his call for a reformation of the Church in its head and members brought him into conflict with conservative Catholic theologians. His support for the Reformation movement was equivocal, however. He refused to endorse the radical methods of the reformers and engaged in a polemic with Luther over the question of free will. On the whole, Erasmus was more interested in the moral and spiritual than in the doctrinal aspects of the Reformation. He promoted inner piety over the observance of rites, and disparaged scholastic speculations in favour of the* philosophia Christi *taught in the gospel. The term 'Christian humanism' best describes Erasmus' philosophy, which successfully combined Christian thought with the classical tradition revived by Renaissance humanists.*

1 Life and works

Erasmus was born at Rotterdam, the illegitimate son of a priest. Orphaned at an early age, he was persuaded by his guardians to enter an Augustinian monastery. He became a canon regular and was ordained priest in 1492. In 1495 he was sent to Paris to study theology, but soon developed a strong distaste for the scholastic method taught there. When the financial help promised by his bishop did not materialize, Erasmus was obliged to support himself by tutoring well-to-do young men. In 1499 he accompanied one of his pupils to England. The visit led to important connections with future patrons (among them William Warham, Archbishop of Canterbury) and to lifelong friendships with Thomas More (see Utopianism §1) and John Colet. Over the next two decades Erasmus travelled extensively. He returned to France, visited the Low Countries, made a second journey to England, and then spent three years (1506-9) in Italy. While in Turin he obtained a Doctorate of Theology *per saltum*, that is, without fulfilling the usual academic and residence requirements. For a year he lived in Venice, doing editorial work and research for the Aldine press. On the accession of Henry VIII in 1509 he returned to England and taught Greek at Cambridge, but his hopes for royal patronage were only modestly rewarded.

By this time Erasmus had made a name for himself as the author of the homiletic *Enchiridion Militis Christiani (Handbook of the Christian Soldier)* (1503) and the witty *In Praise of Folly* (1511), both calling for spiritual renewal of church and society. In 1515 he was appointed counsellor to Prince Charles of Burgundy (later Emperor Charles V) and through the generosity of his patrons in England and the Low Countries acquired sufficient wealth to live the life of an independent scholar. He took up residence at Louvain in 1517 and with short interruptions remained there until 1521. In 1516 he published his *magnum opus*, an edition of the New Testament containing a revised text of the Vulgate faced by the Greek original (the first printed Greek text of the New Testament to reach the market). Erasmus' pioneering work as textual critic was immediately attacked by the Louvain theologians, and his move to Basle in 1521 was partly motivated by a desire to escape their hostilities. However, geographical distance did not deter Erasmus' opponents in the Low Countries. Soon opposition to his work spread to France and Spain, involving Erasmus in bitter controversy and culminating in a formal condemnation by the faculty of theology at Paris in 1531. His critics accused Erasmus of attacking the principle of divine inspiration of Scripture, undermining the traditions of the Church, and providing the underpinning for the Lutheran 'heresy'. Erasmus denied these charges. He noted that his revision of the Vulgate text was not aimed at changing the words of the inspired evangelist, but at correcting the mistakes of an incompetent translator or a careless scribe. Although he occasionally challenged the traditional interpretation of a scriptural passage, he was willing in all instances to submit his views to the verdict of the Church. While he was no doubt sympathetic to the goals of the Reformation, he disapproved of the radical methods adopted by the reformers. Trying to maintain a scholarly detachment was

difficult, however. Erasmus soon found himself attacked by both parties. Conservative theologians perceived him as Luther's inspiration; followers of Luther resented his refusal to endorse the reformer unequivocally and accused him of hypocrisy. Challenged to clarify his position, Erasmus reluctantly entered the arena against Luther with a *Discourse on Free Will* (1524). The ensuing polemic led to a breach between the two men, but failed to convince Catholic critics of the orthodoxy of Erasmus' views. The controversy over Erasmus' commitment to the Catholic faith continued after his death in 1536. In the wake of the Council of Trent which imposed strict doctrinal norms, Erasmus' works were placed on the Index of Prohibited Books.

2 Educational philosophy

Erasmus' pedagogical theories are based on the anthropological optimism that characterizes Renaissance humanism. His philosophy of education is formulated in a number of treatises, dealing with curriculum, methods and objectives. Curriculum is discussed in *De ratione studii (On the Method of Study)* (1512) and *Ratio verae theologiae (The Method of True Theology)* (1518). Erasmus rejected the traditional emphasis on logic and speculative sciences in favour of language studies, history and moral philosophy. Such a curriculum was meant not only to develop skills and train the intellect, but also to provide moral instruction and promote civic virtue. *De pueris instituendis (On the Education of Children)* (1529) is a comprehensive statement of Erasmus' views on the goals of education and the respective roles played by parent, teacher and pupil in the process of education.

Erasmus believed in the human capacity for self-improvement through education and in the relative preponderance of nurture over nature. Indeed he believed that individuals are 'made human' through education. Although Erasmus placed the customary emphasis on imitation as the principal method of learning, he defined it in a wider sense as comprising the thorough understanding of a subject. Copying models was only a first step in the learning process. This must be followed by the judicious application of skills acquired by imitation, which includes choosing elements suitable to the subject, the occasion and the parties involved. Thus Erasmus acknowledged that mastering a subject required analysis, judgment, and a process of creative appropriation. Erasmus abhorred corporal punishment and insisted that positive reinforcement and healthy competition were the most effective means of motivating the learner. In the educational process Erasmus attached a disproportionate importance to the teacher, whom he held largely responsible for the success or failure of a pupil. Apparently he considered a pupil's negative attitude to learning or inability to benefit from instruction the direct result of inadequate teaching methods. While Erasmus asserted that everyone was capable of self-improvement, he grew up with the belief that females were less likely to benefit from education than males. He tells us that he revised his views when confronted with evidence to the contrary. All of Erasmus' pronouncements on education have a Christian subtext. He believed that the student must be transformed, not only into a learned person, but a learned Christian. In his estimate, St Jerome was the perfect scholar, embodying *docta pietas*, that is, combining erudition with piety.

3 Political thought

The main source for Erasmus' political thought is his *Institutio Principis Christiani (The Education of a Christian Prince)*, published in 1516 and dedicated to the 16-year-old Prince Charles. In this essay the ruler appears as a father figure, who looks after the welfare of his people, dispenses justice and provides moral leadership. He is the representative of God and as such is owed unquestioning obedience by his people. The ruler in turn must render an account of his stewardship to God. 'He does not have the same rights over men as over cattle,' Erasmus notes. 'Government depends to a large extent on the consent of the people, which was what created kings in the first place' (*Collected Works*: vol. 27, p. 284). Consent and consultation are prominent themes in Erasmus' political writings. He saw consensus-building as an important process in the government of both state and church. Accordingly, he proposed to settle territorial conflicts among states by calling on a group of referees, and to determine dogmatic questions within the church by summoning a general council. His emphasis on consensus as a means of ensuring a correct decision also of course has epistemological implications (see §4).

The maintenance of peace and concord is another central theme of Erasmus' political writings. It is the subject of two essays, 'Dulce Bellum Inexpertis' (War is Sweet to the Inexperienced) (1515) and *Querela Pacis (The Complaint of Peace)* (1517). Erasmus' pacifism is informed by the classical ideal of *humanitas* and the Christian ideal of a universal fellowship. In *De bello Turcico (On War Against the Turks)* (1530) Erasmus accepts the concept of a 'just war' against the Ottoman Empire but disparages a military solution. Instead he promotes the idea of the Turks as a scourge of God, which calls for spiritual weapons rather than battle gear. A transfer from the

political to the theological realm is apparent also in Erasmus' image of the Christian commonwealth with Christ surrounded by clergy, princes and common people in three concentric circles. This iconography is reminiscent of the traditional division of the polity into king and estates. It assigns to Christ a position analogous to that of the king and dispenser of power. In Christ's case, this is the power to purify those on earth from moral corruption. As heavenly king he delegates this power to the lower orders, enabling them to carry out his work of purification according to their own capacities (see Sovereignty).

4 Scepticism

Erasmus' moral integrity was questioned by his opponents when he declared that, although it was not permissible to 'go against' the truth, it might be expedient in some circumstances to conceal it. He also considered careful judgment and a fine sense of timing to be of prime importance in the deployment of truth. The notion that the truth needs husbanding is best understood as an aspect of Erasmus' scepticism. He believed that human beings were unable to grasp the absolute truth by their own intellectual efforts. Through discussing arguments on both sides of a question they could arrive at an approximation of the truth, but the investigation must necessarily end in *epoche*, the suspension of judgment. Erasmus' polemic with Luther over free will illustrates this method of investigation. His *Discourse on Free Will* consists of a comparison of biblical and patristic passages for and against the concept of free will. He concludes that the human will is exceedingly weak but able, with the help of divine grace, to choose the path of salvation. Erasmus does not assert this view, however, but presents it as the more likely to be true, emphasizing that his role was not to judge matters but to discuss them. It is important to note, however, that Erasmus set limits to scepticism in the religious sphere. He believed in absolute knowledge through revelation and resorted to calculations of probability only in cases that could not be settled by the authority of Scripture and had not been determined by the doctrinal pronouncements of the Church (see Scepticism, Renaissance §3).

5 Philosophy of language

As the author of several textbooks on style and composition, Erasmus took a substantial interest in language. However, he was concerned primarily with the practicalities of teaching language skills and does not seem to have a coherent philosophy concerning the nature, origin and function of language. Pronouncements such as 'In principle, knowledge as a whole seems to be of two kinds, of things and of words. Knowledge of words comes earlier, but that of things is the more important' (the opening sentence of *De ratione studii*) are isolated statements and do not lead to a fuller discussion of the underlying issues. They suggest the metaphysical questions raised and discussed in Plato's *Cratylus*, Aristotle's *Metaphysics*, or Augustine's *On Christian Doctrine* (see Plato §15; Aristotle §11; Augustine §8), but are merely echoes of classical commonplaces. Indeed, Erasmus is concerned, not with philosophy, but with rhetoric, and more specifically, with effective methods of teaching rhetoric.

De ratione studii, a curriculum outline, continues appropriately with remarks of a pedagogical rather than a philosophical nature: 'But some people, while they hurry on to learn about things, neglect a concern for language.' Elsewhere, while discussing the relation between words and things, Erasmus claims that people who are unskilled in the use of language are 'short-sighted', 'deluded' and unable to form proper judgments about things. Once again, the philosophical implications are left unexplained. The statement therefore amounts to little more than a comment on the importance of effective communication. The subject of the dichotomy between words and things recurs in *De copia (On the Abundance of Style)* (1512). There it serves as an organizational principle (Book I deals with 'words', that is, style; Book II with 'things', that is, subject matter). As in *De ratione*, references to the dichotomy are linked with remarks on the effectiveness of communication: subject matter is obscured by the failure to organize words and expressions correctly. Indeed Erasmus seems to disavow any philosophical purpose. Considering distinctions between words and things to be a matter of theory rather than practice, he explained that the arrangement in his book had no philosophical basis, but was a 'teaching procedure'.

More significant in the context of a philosophy of language is a statement found in the prolegomena of Erasmus' edition of the New Testament. Defending his revision of the Vulgate translation, he argues that speech is made up of two things: words (which he relates to the body) and meaning (which is related to the soul). Similarly in *De copia*, a textbook on style, he writes of style being to thought what clothes are to the body. Like his remarks on the dichotomy of words and things, however, these statements operate on a rhetorical rather than a philosophical level and concern method rather than logic. He goes on, in one case, to discuss techniques of translation, in the other, to suggest techniques for varying expression. The overwhelming consensus of modern scholars is that it is futile 'to

impute a philosophy of language to Erasmus' (O'Rourke Boyle 1977) on the basis of these and other isolated remarks, or at any rate, that his thinking on the subject is ill-disciplined and 'very imperfectly worked out' (Woodward [1904] 1964). Erasmus' comments on language bear out his propensity to focus on practical application rather than on theory, a propensity also found in his religious writings with their pronounced emphasis on Christian living rather than on doctrinal definitions (see Language, Renaissance philosophy of).

6 *Philosophia Christi*

Erasmus called his *Enchiridion Militis Christiani* a 'summary guide' to Christian living. In this work he developed the concept of a *philosophia Christi*, a term which is taken from patristic writings and develops Pauline notions: the dichotomy of spirit and flesh and the inferiority of visible to invisible things. Erasmus describes the human being as having a divine soul but the body of 'a brute beast'. He therefore places emphasis on the inner man rather than on the observance of rites, on faith rather than works. He does not reject the ceremonial aspect of religion out of hand, but reduces its role to that of a crutch aiding the weak in their progression towards a more perfect faith. Erasmus also takes a stand against scholastic theologians who place inordinate trust in Aristotelian dialectics (see Major, J.). The philosophy of Christ advocated by him does not require scholarship, but manifests itself in the believer's heart and through the conduct of their life. The body of theological work published by Erasmus - editions and translations of patristic writings; an annotated edition of the New Testament; paraphrases and commentaries on Scripture - indicates the focal points of his religious thought. In Erasmus' view, the essence of the Christian faith could be discovered, not from the disputations of scholastic theologians, but from the life of Christ related by the evangelists and the interpretations of the divine word by the early Church Fathers.

Erasmus was not a systematic philosopher, but a keen observer of society, commenting on the questions that agitated his contemporaries and at the same time shaping the opinions of his own and future generations. His call for reform of church and society merited him the title of a 'forerunner' of the Reformation, and his pedagogical works had a pervasive influence on curriculum in his time. His criticism of the scholastic method and his scepticism in particular commended him to readers in the Age of Enlightenment. This induced a modest Erasmian renaissance after a decline in popularity during the seventeenth century, especially in areas where the Counter-Reformation had led to a strict enforcement of censorship laws. We find, however, that the positions taken by Erasmus on individual questions are not integrated into a whole: patriarchal government is advocated along with the egalitarian ideal of consensus and consent; scepticism is juxtaposed with devout and unquestioning belief; anthropological optimism with the reduction of the human will to a spark kindled by divine grace. Erasmus' ideas are clearly modified by the tenets of the Catholic Church. His belief in the human potential is tempered by the doctrine of God's omnipotence, his rationalism by the tenet of God's inscrutability, his intellectual curiosity is circumscribed by articles of faith, and his scholarly findings subjected to the verdict of the Church. The term 'Christian humanism' best describes Erasmus' brand of philosophy. It acknowledges that Erasmus' thought is firmly rooted in the Christian tradition and that he was a man of letters rather than a professional philosopher.

See also: Education, history of philosophy of; Free will; Humanism, Renaissance; Luther, M.; Scepticism; Scepticism, Renaissance

ERIKA RUMMEL

List of works

Erasmus, D. (1503-36) *Desiderii Erasmi Roterodami Opera Omnia (Complete Works of Desiderius Erasmus of Rotterdam)*, ed. J. Leclerc, Leiden: Vander Aa, 1703-6; repr. Gregg Press, 1962.(The Gregg Press edition is a facsimile reprint.)

Erasmus, D. (1503-36) *Erasmi Opuscula: A supplement to the Opera Omnia*, ed. W. Ferguson, The Hague: Martinus Nijhoff, 1933.(Collection of Erasmus' short works.)

Erasmus, D. (1503-36) *Opus epistolarum Des. Erasmi Roterodami*, ed. P.S. Allen, Oxford: Clarendon Press, 1906-47.(Collection of Erasmus' letters.)

Erasmus, D. (1503-36) *Des. Erasmi Opera Omnia (Complete Works of Desiderius Erasmus)*, Amsterdam: North Holland, 1969-.(New Critical Latin edition, in progress. It does not include Erasmus' correspondence.)

Erasmus, D. (1503-36) *The Collected Works of Erasmus*, Toronto, Ont.: University of Toronto Press, 1974-. (English edition in progress. *Ratio verae theologiae* and *De bello Turcico* are forthcoming; the other works mentioned in this article have appeared.)

Erasmus, D. (1503) *Enchiridion Militis Christiani (The Handbook of the Christian Soldier)*, in *The Collected Works of Erasmus*, vol. 66, Toronto, Ont.: University of Toronto Press, 1988.(Develops Erasmus' concept of the *philosophia Christi*.)

Erasmus, D. (1511) *In Praise of Folly*, in R.M. Adams (ed.) *The Praise of Folly and Other Writings*, New York: W.W. Norton, 1990.(Includes *Paraclesis*, *Querela Pacis*, the foreword to the New Testament, a selection of colloquies and letters, and critical essays by modern scholars.)

Erasmus, D. (1512) *De copia (On the Abundance of Style)*, in *The Collected Works of Erasmus*, vol. 24, Toronto, Ont.: University of Toronto Press, 1978.(One of Erasmus' many textbooks on style. Book 1 deals with style, Book 2 with subject matter.)

Erasmus, D. (1512) *De ratione studii (On the Method of Study)*, in *The Collected Works of Erasmus*, vol. 24, Toronto, Ont.: University of Toronto Press, 1978.(A curriculum outline and approach to a philosophy of language from pedagogical standpoint.)

Erasmus, D. (1515) 'Dulce Bellum Inexpertis' (War is Sweet to the Inexperienced), in M. Mann Phillips (trans.) *Erasmus on his Times*, Cambridge: Cambridge University Press, 1967.(Outlines Erasmus' pacifism in the light of classical notions of humanity and the idea of Christian fellowship.)

Erasmus, D. (1516) *Institutio Principis Christiani (The Education of a Christian Prince)*, in *The Collected Works of Erasmus*, vol. 27, Toronto, Ont.: University of Toronto Press, 1986.(Presents Erasmus' ideas on sovereignty and good government.)

Erasmus, D. (1517) *Querela Pacis (Complaint of Peace)*, in *The Collected Works of Erasmus*, vol. 27, Toronto, Ont.: University of Toronto Press, 1986.(Concerned with the maintenance of peace and concord, appeals to Christian virtues in an argument against the necessity of war.)

Erasmus, D. (1518) *Ratio verae theologiae (The Method of True Theology)*, in H. Holborn (ed.) *Desiderius Erasmus Roterdamus: Ausgewählte Werke*, Munich: C.H. Beck, 1933.(On the aims and methods of theological studies.)

Erasmus, D. (1524) *Erasmus-Luther: Discourse on Free Will*, trans. E. Winter, New York: Frederick Ungar, 1961.(Erasmus' investigation of the concept of free will, using biblical and patristic sources; part of a major dispute with Luther.)

Erasmus, D. (1529) *De pueris instituendis (On the Education of Children)*, in *The Collected Works of Erasmus*, vol. 26, Toronto, Ont.: University of Toronto Press, 1985.(A comprehensive statement of Erasmus' views on the goals of education and the respective roles played by parent, teacher and pupil in the process of education.)

Erasmus, D. (1530) *De bello Turcico (On War Against the Turks)*, in *Des. Erasmi Opera Omnia*, vol. 3, Amsterdam: North Holland, 1986.(Accepts the concept of a 'just war' against the Ottoman Empire but recasts it as a primarily spiritual conflict. *De bello Turcico* is partially translated into English in *The Erasmus Reader*.)

Erasmus, D. (1990) *The Erasmus Reader*, ed. E. Rummel, Toronto, Ont.: University of Toronto Press.(This collection includes a partial translation of *De bello Turcico*.)

References and further reading

Augustijn, C. (1991) *Erasmus: His Life, Works, and Influence*, Toronto, Ont.: University of Toronto Press. (Extensive bibliography.)

Hoffmann, M. (1994) *Rhetoric and Theology: The Hermeneutics of Erasmus*, Toronto, Ont.: University of Toronto Press.(Provides the best analysis of Erasmus' *Ecclesiastes (The Preacher)* available in English.)

Jardine, L. (1993) *Erasmus, Man of Letters: The Construction of Charisma in Print*, Princeton, NJ: Princeton University Press.(Documents Erasmus' skill as a publicist; serves as a comment on his philosophy of language.)

Mansfield, B. (1979, 1992) *Phoenix of his Age: Interpretations of Erasmus, 1550-1750*, vol. 1; *Interpretations of Erasmus, c.1750-1920: Man on His Own*, vol. 2, Toronto, Ont.: University of Toronto Press.(Two-volume work on Erasmus' influence.)

McConica, J.K. (1991) *Erasmus*, Oxford: Oxford University Press.(A concise, insightful examination of Erasmus' thought.)

O'Malley, J. (1988) 'Introduction', *The Collected Works of Erasmus*, vol. 66, Toronto, Ont.: University of Toronto Press.(This is the best discussion in English of Erasmus' *philosophia Christi* and contains further bibliographical references.)

O'Rourke Boyle, M. (1977) *Erasmus on Language and Method in Theology*, Toronto, Ont.: University of Toronto Press.(Referred to in §5. An investigation of Erasmus' linguistic principles.)

O'Rourke Boyle, M. (1983) *Rhetoric and Reform: Erasmus' Civil Dispute with Luther*, Cambridge, MA: Harvard University Press.(Includes a discussion of Erasmus' scepticism.)

Popkin, R. (1964) *The History of Scepticism from Erasmus to Descartes*, New York: Humanities Press.(Chapter 1 deals with Erasmus' scepticism.)

Rummel, E. (1989) *Erasmus and His Catholic Critics*, Nieuwkoop: de Graaf.(Examines Erasmus' biblical scholarship and Christian humanism in the context of his polemics.)

Rummel, E. (1995) *The Humanist-Scholastic Debate in the Renaissance and Reformation*, Cambridge, MA: Harvard University Press.(Discusses Erasmus' role in this curricular/philosophical debate passim; chapter 6 explores his position as humanist and reformer.)

Tracy, J. (1970) *The Politics of Erasmus: A Pacifist Intellectual and His Political Milieu*, Toronto, Ont.: University of Toronto Press.(On Erasmus' political thought.)

Tracy, J. (1972) *Erasmus: The Growth of a Mind*, Geneva: Droz.(An incisive study of Erasmus' thought.)

Tracy, J. (1996) *Erasmus of the Low Countries*, Berkeley, CA: University of California Press.(An intellectual biography situating Erasmus in his political and cultural milieu; extensive bibliography.)

Woodward, W.H. (1904) *Desiderius Erasmus: Concerning the Aim and Method of Education*; repr. New York: Columbia University Press, 1964.(Most recent reprint of a work first published in 1904; this remains the most comprehensive study of Erasmus' educational philosophy. Referred to in §5.)

Eriugena, Johannes Scottus (*c*.800-*c*.877)

Johannes Scottus Eriugena is the most important philosopher writing in Latin between Boethius and Anselm. A Christian Neoplatonist, he developed a unique synthesis between the Neoplatonic traditions of Pseudo-Dionysius and Augustine. Eriugena knew Greek, which was highly unusual in the West at that time, and his translations of Dionysius and other Greek authors provided access to a theological tradition hitherto unknown in the Latin West. From these sources, Eriugena produced an original cosmology with Nature as the first principle. Nature, the totality of all things that are and are not, includes both God and creation, and has four divisions: nature which creates and is not created, nature which creates and is created, nature which is created and does not create, and nature which is neither created nor creates. These divisions participate in the cosmic procession of creatures from God and in their return to God. As everything takes place within Nature, God is present in all four divisions. Eriugena influenced twelfth-century Neoplatonists but was condemned in the thirteenth century for teaching the identity of God and creation.

1 Life and writings

Little is known about Eriugena's life. His exact place and date of birth are unknown, but contemporary sources refer to him as *scottus*, meaning 'Irish', and he signed one of his manuscripts with the pleonasm 'Eriugena' (Irish born). No evidence of his education survives. Johannes joined the court of Charles the Bald, possibly in the 840s, and achieved recognition as a liberal arts master. Two partial commentaries on *De nuptiis Philologiae et Mercurii (The Marriage of Philology and Mercury)*, the liberal arts handbook of the late Latin author Martianus Capella, survive from Eriugena's period as royal master (see Encyclopedists, medieval §4).

Eriugena was commissioned to refute a treatise by a Saxon monk, Gottschalk, who interpreted Augustine as teaching a twofold predestination: of the elect and of the damned. Eriugena's reponse, *De divina praedestinatione (On Divine Predestination)*, denied predestination by defending God's transcendence and goodness. God, being perfectly good, wants all humans to be saved. Humans, however, damn themselves through their own free choices. God, who is outside time, cannot be said to '*fore*-know' or to '*pre*-destine', terms which involve temporal predicates. Hence God does not predestine souls to damnation (see Eternity). Due to a perceived overemphasis on human free-will in the salvific process, Eriugena was accused of Pelagianism (the heresy that human beings can be saved through their own resources rather than by divine grace) (see Pelagianism), and *De divina praedestinatione* was condemned in France at the councils of Valence (855) and Langres (859). While purporting merely to interpret Augustinian texts, this early treatise is philosophically significant for its rationalistic, dialectical analysis of key theological concepts.

Despite the condemnations of *De divina praedestinatione*, Eriugena continued to have the patronage of Charles the Bald, who invited him to translate the writings of Dionysius the Areopagite, a mysterious Christian Neoplatonist, possibly of the sixth century, who imitated Proclus and purported to be Dionysius, the first convert of St Paul at Athens (see Pseudo-Dionysius). This author had been wrongly identified with St Denis, patron saint of France. For Dionysius, it is more true to say 'God is not' than 'God is', since God is 'above all the things that are and are not'. Eriugena adopted Dionysius' affirmative and negative theology, according to which denials concerning God are 'more true', 'better', 'more apt', than affirmations.

Eriugena's major dialogue, *Periphyseon* or *De divisione naturae (On the Division of Nature)* was completed around 867. He also wrote a commentary on Dionysius' *Celestial Hierarchy (Expositiones in hierarchiam coelestem)* and on Maximus' *Ambigua*. A fragmentary commentary on the Gospel of St. John and a sermon, *Homilia in Johannem (Homily on the Prologue to St John's Gospel)*, were also written probably in the late 860s or 870s. A number of poems also survive. It is probable that Eriugena died sometime around 877; an apocryphal tale, dating from the twelfth century, records that he was stabbed to death by his students.

2 Sources

Eriugena had no direct knowledge of Plotinus, Porphyry or Proclus. He appears to have known Plato's *Timaeus* in Calcidius' translation (see Plato). He knew the Pseudo-Augustinian paraphrase of Aristotle's *Categories*, Rufinus's Latin translation of Origen's *De principiis (On First Principles)*, and Boethius' trinitarian tracts and *De*

consolatione philosophiae (Consolation of Philosophy) (see Boethius, A.M.S.; Origen). His chief authorities were Augustine, Ambrose, Hilary of Poitiers and Jerome, among the Latin Fathers, but he more often expresses a preference for Eastern Church Fathers, in particular Dionysius and the Cappadocians, Basil, Gregory of Nazianzus and Gregory of Nyssa, whose *De hominis opificio (The Creation of Man)* he translated under the title of *De imagine (On the Image of Good)*. He also translated Maximus Confessor's *Ambigua (Difficult Questions of Interpretation [in Pseudo-Dionysius])* and *Quaestiones ad Thalassium (Questions to Thalassius)* and he may have translated Epiphanius's *Anchoratus (De fide) (The Hermit State (On Faith))* (see Patristic philosophy).

Given the available resources, Eriugena was exceptionally learned, but his genius lay in being able to bring all these 'authorities' together in a new cosmological framework. Eriugena enthusiastically incorporated many Greek Christian theological concepts: God, the One, creates by self-emanation; creation is a timeless event and hence ongoing and always contemporary; human nature is originally a Platonic Idea in the mind of God. Humans fail to understand their true nature as image of God in God because they are distracted by created, temporal images (*phantasiai*), leading to a fall into the spatio-temporal realm of sense. However, through intellectual contemplation (*theoria*) and divine illumination (which is the divine self-manifestation, *theophania*), humans may achieve unification (*henosis*) with God, deification (*theosis*). God became man so that humans can become God (see Platonism, early and middle).

3 The divisions of nature

The *Periphyseon*, or *De divisione naturae (On the Division of Nature)*, a long dialogue in five books between an anonymous teacher and his student, is Eriugena's major philosophical treatise. It opens with an all-encompassing definition of its subject matter. Nature is the general term 'for all things that are and all things that are not' (CXXII I.441a), including both God and creation. The first principle of nature is God, 'the cause of all things that are and that are not' (I.442b). Nature is divided into four 'species', echoing similar divisions in Augustine and Marius Victorinus: first, that which creates and is not created (that is, God); second, that which creates and is created (that is, Primary Causes or Ideas); third, that which is created and does not create (that is, Temporal Effects, created things); and fourth, that which is neither created nor creates (that is, non-being or nothingness).

Eriugena proceeds to list five ways (*quinque modi*) in which things may be said to be and not to be. According to the first mode, things accessible to the senses and the intellect are said to be, whereas those which - through their excellence - transcend our faculties are said not to be. According to this classification, God is said not to be; he is 'nothingness through excellence' (*nihil per excellentiam*). The second mode of being and non-being is based on the 'orders and differences of created natures', whereby if one level of nature is said to be, those orders above or below it are said not to be. According to this mode, to affirm 'man' is to negate 'angel' and vice versa. This mode invokes the medieval hierarchy of being and at the same time subverts it by relativising the notion of being: only one level of the hierarchy can at any time be said to be. The third mode asserts that *actual* things are, whereas *potential* things still caught up 'in the secret recesses of nature' are not. The fourth mode says that things which the intellect contemplates truly are, whereas material things caught up in generation do not truly exist. The fifth mode says that humans sanctified by grace are, whereas sinful humans are not.

Eriugena's discussion of being and non-being is subject to a dialectic of affirmation and negation. Being and non-being, then, are correlative categories; something may be said to be under one mode and not to be under another mode. These modes further must be taken into account when assessing Eriugena's metaphysical statements. Thus when Eriugena calls God 'nothing,' he means that God transcends all created being. Matter is called 'nothing through privation' (*nihil per privationem*), and created things are called 'nothing' because they do not contain in themselves their principles of subsistence (see Being).

Periphyseon Book I examines the first division, God. The Aristotelian categories do not properly apply to God, who is not literally (*proprie*) substance or essence, does not possess quantity or quality, has no relations and is not in place and time. God transcends all. His 'being' is beyond being. God is a 'nothingness' whose real essence is unknown to all created beings, including the angels. Indeed, God's nature is unknown even to himself, since he is the 'infinity of infinities' and hence beyond all comprehension and circumscription. God does not know evil, and, in a real sense, God may be said not to know anything (see God, concepts of).

However, for Eriugena, God also creates himself. He manifests himself in theophanies. This self-creation is God's

self-manifestation and also the expression of the Word and the creation of all other things, since all things are contained in the Word. The Word enfolds in itself the Ideas or Primary Causes of all things. God is 'all in all'. All things are in God: '… the Creative nature permits nothing outside itself because outside it nothing can be, yet everything which it has created and creates it contains within itself, but in such a way that it itself is other, because it is superessential, than what it creates within itself' (III.675c). God's transcendent otherness above creatures is precisely that which allows creatures to be within God and yet other than God. Eriugena stresses both the divine transcendence and immanence in creation, which is at the same time the immanence of creatures within God (see Creation and conservation, religious doctrine of).

God and the creature are one and the same: 'It follows that we ought not to understand God and the creature as two things distinct from one another, but as one and the same. For both the creature, by subsisting, is in God; and God, by manifesting himself, in a marvellous and ineffable manner creates himself in the creature' (III.678c). Eriugena's assertion that God is the 'form of all things' (*forma omnium*) led to the accusation of heresy, an accusation which fixates on one side only of Eriugena's dialectical reasoning. Although Eriugena asserts the identity of God and creation, he explicitly rejects the view that God is the 'genus' or 'whole' of which the creatures are 'species' or 'parts'. Only metaphorically (*metaforice*, *translative*) can it be said that God is a 'genus' or a 'whole'. The immanence of God in creation is balanced by God's transcendence above all things. God is both form of all things and without form.

Periphyseon Book II discusses the Primary Causes located in the mind of God. This doctrine combines the Platonic theory of Forms, Dionysius's discussion of the divine names and the Stoic-Augustinian notion of eternal reasons. The number of causes is infinite and none has priority over the other; for example, Being is not prior to Goodness, or vice versa. Each is a divine theophany. The very nature of these Causes is to flow out from themselves, bringing about their Effects. This outflowing (*exitus*) creates the whole universe from the highest genus to the lowest species. In this causal procession, like produces like; incorporeal causes produce incorporeal effects (see Causation). All created things are essentially incorporeal, immaterial, intellectual and eternal. Place and time are definitions which locate things, and since definitions are in the mind, then place and time are also in the mind. The sensible, corporeal spatio-temporal appearance of things is produced by the qualities or 'circumstances' of place, time, position and so on, which surround the incorporeal essence. The whole spatio-temporal world and our corporeal bodies are a consequence of the Fall, an emanation of the mind.

Book III discusses the nature of creation from nothing (*ex nihilo*). 'Nothing' has two meanings. The lowest rung in the hierarchy of being, unformed matter, is 'almost nothing' (*prope nihil*), 'nothingness through privation' (*nihil per privationem*). In contrast, God is non-being through excellence. Creation from nothing cannot mean creation from a principle outside God, since there is nothing outside God. *Ex nihilo* creation then means 'out of God's superabundant nothingness'. God creates out of himself and all creation remains within him.

Books IV and V discuss the return (*reditus*) of all things to God. Corporeal things will return to their incorporeal causes, the temporal to the eternal. The human mind will achieve reunification with the divine, and then the corporeal, temporal and material world will become essentially incorporeal, timeless and intellectual. Human nature will return to its Idea in the mind of God. This perfect human nature is paradise. There is a general return of all things to God. Humans who refuse to let go of the 'circumstances' remain trapped in their own phantasies, and this constitutes hell. The elect achieve a special deification whereby they will merge with God completely, as lights blend into the one light. Neither hell or heaven are localized; for Eriugena, they are intellectual states.

Eriugena's anthropology has been the focus of much philosophical interest in the twentieth century. For Eriugena, 'rational animal' does not adequately define human nature, nor is humanity a microcosm: humanity is an Idea eternally made in the mind of God. Perfect human nature is omniscient and omnipotent. Human nature is open to infinite possibility and perfectibility (see Perfectionism).

4 Influence and significance

Eriugena had immediate influence in France, notably at the schools of Laon, Auxerre and Corbie. His translations of Dionysius were widely used into the thirteenth century. His *Vox spiritualis (The Spiritual Voice)* or *Homilia in Johannem* (often attributed to Origen) was widely read in the Middle Ages. The *Periphyseon* was popular in the twelfth century especially in the paraphrase of Honorius Augustodunensis; Hugh of St Victor, Alanus of Lille,

Suger of St Denis and William of Malmesbury were all influenced by Eriugena. In the thirteenth century, the *Periphyseon* was associated with the writings of David of Dinant and Amaury of Bène, two theologians at the University of Paris, and was condemned with them in 1210 and 1225. According to Thomas Aquinas, Amaury of Bène was condemned for saying that God is the formal principle of all things, an accusation of pantheism, which recalled Eriugena's statement that God is the 'form of all things' (*forma omnium*). In the later Middle Ages both Meister Eckhart of Hochheim and Nicholas of Cusa were sympathetic to Eriugena and familiar with his *Periphyseon*. Interest in Eriugena was revived by Thomas Gale's edition of *De divisione naturae* (1681). Hegel and his followers saw Eriugena as the father of German idealism (see German Idealism). Critical editions of his major works are only now being produced.

Eriugena is an original philosopher who articulates the relation between God and creation in a manner which preserves both divine transcendence and omnipresence. His theory of human nature is rationalist and intellectualist. His theory of place and time as defining structures of the mind anticipates Kant, and his dialectical reasoning prefigures Hegel. Eriugena is a mystic who emphasises the unity of human nature with God.

See also: Augustinianism; Carolingian Renaissance; Platonism, medieval

DERMOT MORAN

List of works

Eriugena, Johannes Scottus (*c*.800-*c*.877) *Complete Works*, ed. H.-J. Floss, *Johannis Scoti Opera quae supersunt Omnia*, Patrologia Latina 122, Paris: 1853.(The complete edition of Eriugena's works in Latin.)

Eriugena, Johannes Scottus (*c*.840-60) *Iohannis Scotti Annotationes in Marcianum, (Commentary on Martianus Capella's The Marriage of Philology and Mercury),* ed. C. Lutz, Cambridge, MA: Medieval Academy of America, 1939.(Edition of the Latin text of one of Eriugena's commentaries on Martianus Capella's *The Marriage of Philology and Mercury*.)

Eriugena, Johannes Scottus (*c*.851) *De divina praedestinatione (On Divine Predestination)*, ed. G. Madec, *Iohannis Scotti de divina praedestinatione*, Corpus Christianorum, Series Latina 50, Turnhout: Brepols, 1978. (Critical edition of the Latin text of Eriugena's *On Divine Predestination*.)

Eriugena, Johannes Scottus (*c*.865-70) *Expositiones in Ierarchiam coelestem (Commentaries on Pseudo-Dionysius' Celestial Hierarchy)*, ed. J. Barbet, *Iohannis Scoti Eriugenae Expositiones in Ierarchiam coelestem*, Turnhout: Brepols, 1975.(The critical Latin edition of Eriugena's Commentary on the *De coelesti hierarchia* of Pseudo-Dionysius.)

Eriugena, Johannes Scottus (*c*.865-70) *Commentarius in Evangelium Iohannis (Commentary on the Gospel of John)*, ed. E. Jeauneau, *Jean Scot: Commentaire sur l'Evangile de Jean*, Sources Chrétiennes 180, Paris: Éditions du Cerf, 1972.(Critical edition of the Latin text with French translation of Eriugena's incomplete Commentary on the Gospel of John.)

Eriugena, Johannes Scottus (*c*.865-70) *Homilia in Johannem (Homily on the Prologue to the Gospel of St John)*, ed. E. Jeauneau, *Jean Scot: L'Homélie sur le Prologue de Jean*, Sources Chrétiennes 151, Paris: Éditions du Cerf, 1969.(Critical edition of the Latin text with French translation of Eriugena's homily on the Prologue to St. John's Gospel.)

Eriugena, Johannes Scottus (*c*.860-64) *Maximi Confessoris Ambigua ad Iohannem iuxta Iohannis Scotti Eriugenae latinam interpretationem (Commentary on Maximus' Ambigua)*, ed. E. Jeauneau, Turnout: Brepols, Louvain: Leuven University Press, 1988.(Critical edition of the Latin text of Eriugena's translation of Maximus.)

Eriugena, Johannes Scottus (*c*.867) *Periphyseon, or De divisione naturae (On the Division of Nature)*, ed. I.-P. Sheldon-Williams, *Johannis Scotti Eriugenae Periphyseon*, Dublin: Institute for Advanced Studies, 1968-95, 4 vols.(Critical edition of the Latin text of the first four books of Eriugena's *De divisione naturae*, together with English translation and notes. It should be noted that vol. 4 is edited by E. Jeauneau, and a fifth volume is in preparation also under the editorship of E. Jeauneau. The indispensable English translation of the complete text can also be found in I.-P. Sheldon-Williams and J.J. O'Meara, *Eriugena: Periphyseon (The Division of Nature),* Montreal and Paris: Bellarmin, 1987. A very useful abridged English translation of major sections of the work is M. Uhlfelder and J. Potter (1976) *John the Scot. Periphyseon. On the Division of Nature*, Indianapolis: Bobbs-Merrill.)

References and further reading

Brennan, M. (1977) 'A Bibliography of Publications in the Field of Eriugena Studies 1800-1975', *Studi Medievali*

ser. 3a, 28: 401-47.(Lists all books and articles which treat aspects of Eriugena's work.)

Brennan, M (1986) 'Materials for the Biography of Johannes Scottus Eriugena', *Studi Medievali* ser. 3a, 27: 413-60.(Latin extracts and English translations of all the texts which give biographical information on Eriugena.)

Brennan, M (1989) *A Guide to Eriugenian Studies: A Survey of Publications 1930-87*, Paris: Éditions du Cerf.(Contains English-language summaries of all the major secondary literature on Eriugena published in the twentieth century.)

Cappuyns, M. (1933) *Jean Scot Erigène: sa vie, son oeuvre, sa pensèe (Johannes Scottus Eriugena: His Life, Works and Thought)*, Louvain: Abbaye de Mont Cesar.(The standard twentieth-century study on Eriugena.)

Gersh, S. (1978) *From Iamblichus to Eriugena*, Leiden: Brill.(A study of Eriugena in the context of late Greek Neoplatonism.)

Marenbon, J. (1981) *From the Circle of Alcuin to the School of Auxerre: Logic, Theology and Philosophy in the Early Middle Ages*, Cambridge: Cambridge University Press.(A study of Eriugena in the context of Latin philosophy of the early Middle Ages.)

Moran, D. (1989) *The Philosophy of John Scottus Eriugena: A Study of Idealism in the Middle Ages*, Cambridge: Cambridge University Press.(Philosophical interpretation of Eriugena which focuses on his negative dialectics and anthropology as a key to understanding his idealist metaphysics.)

Moran, D. (1990) 'Pantheism from John Scottus Eriugena to Nicholas of Cusa', *American Catholic Philosophical Quarterly* 64 (1): 131-52.(A study of the accusation of pantheism levelled against Eriugena and Nicholas of Cusa.)

Moran, D. (1992a) 'Time, Space and Matter in the *Periphyseon*: an Examination of Eriugena's Understanding of the Physical World', in F. O'Rourke (ed.) *At the Heart of the Real*, Dublin: Irish Academic Press.(A study of Eriugena's account of the spatio-temporal world.)

Moran, D. (1992b) 'Origen and Eriugena: Aspects of Christian Gnosis', in T. Finan and V. Twomey (eds) *The Relationship Between Neoplatonism and Christianity*, Dublin: Four Courts Press.(Shows Eriugena's debt to Origen's Platonism in the *On First Principles*.)

O'Meara, J.J. (1988) *Eriugena*, Oxford: Clarendon Press.(Contains a very useful summary of Eriugena's *On the Division of Nature* together with a translation of some poems.)

Roques, R. (1975) *Libres sentiers vers l'érigénisme (Free-Wheeling Paths Towards Eriugenian Thought)*, Rome: Edizioni dell'Ateneo.(A collection of Rene Roques's scholarly articles in French on Eriugena.)

Sheldon-Williams, I.-P. (1970) 'The Greek Platonist Tradition from the Cappadocians to Maximus and Eriugena', in A.H. Armstrong (ed.) *The Cambridge History of Later Greek and Early Medieval Thought*, Cambridge: Cambridge University Press. (Contains Sheldon-Williams' masterly account of the relation between Eriugena and the Greek Christian Platonist tradition.)

Erotic art

Erotic art is art with a sexual content, which may be more or less overt. The presence of sexual content, however, is not sufficient for a work of art to be considered erotic. Although there is more than one sense in which a work can be said to be erotic, an erotic work of art must aim at and to some extent succeed in evoking sexual thoughts, feeling or desires in the spectator, in virtue of the nature of the sexual scene it represents and the manner in which it represents it. This aim, definitive of erotic art, may be a work's principal aim, but need not be. Erotic art often tends to express the artist's interest in and attitude towards sexuality; and whether or not it does, seeing it as expressing the artist's sexuality is likely to contribute towards the spectator's sexual arousal. An erotic work of art has an intended audience of a more or less specific kind, most frequently men. Erotic art is distinguished from pornography in at least two ways. First, pornography lacks any artistic intent. Second, its main aim is not only to stimulate the spectator sexually but to degrade, dominate and depersonalize its subject, usually women. This article is restricted in scope in at least two ways. First, it concerns exclusively the visual arts. Second, its focus is Western art, and primarily art from the Renaissance onwards.

1 Main questions

The chief philosophical questions in the aesthetics of erotic art are: (1) What is the distinction, within art, between erotic and non-erotic art, and how sharp are the boundaries of this category? (2) What are the ideological implications, if any, of the different forms and manners of erotic art, and what, in particular, is the distinction between erotic art and pornography? (3) How can erotic art in fact be art, or something properly eliciting an aesthetic response that is traditionally characterized as disinterested, when it is also aimed at provoking sexual desire, the very paradigm of an interested reaction? (4) In what ways might the criteria for assessing erotic art differ from those appropriate to other sorts of art, and how does the degree of eroticism of erotic art connect, if at all, to its value or worth as art? This entry will be devoted almost exclusively to (1) and (2).

2 The concept of erotic art

A good proportion of the work of many great visual artists - Rubens, Ingres, Delacroix, Degas, Rodin, Gauguin, Matisse, Magritte, Munch, Klimt, Picasso, Modigliani - is unquestionably erotic. But what is it, precisely, for art to be erotic? It seems that, at a minimum, it must have sexual content. Though sexual content may be either overt or covert, let us first consider art with overt sexual content. Typically, this takes the form of depictions of unclothed or semi-clothed human beings, alone or accompanied, at rest or performing actions of a sexual nature. Yet for art to be accounted erotic, it must do more than represent the naked human body or otherwise make reference to sexual matters: not all art concerned in some way with sexuality is automatically erotic. Anatomical sketches of genitalia, a realistic study of a gynaecologist's examining room, or a modern comic strip featuring pneumatic bimbos, are not erotic, despite the sexual content they include. Rather, erotic art is art that treats its sexual content in a particular way or projects a certain attitude towards it. Erotic art is art aimed at arousing sexual interest, at evoking sexual thoughts, feelings or desires in viewers, in virtue of what it depicts and how it is depicted, and which achieves some measure of success in doing so. The intent to awaken and reward sexual interest through what is depicted can be taken as criterial of at least central cases of erotic art. The erotic work of art does more than refer to or acknowledge human sexuality; rather, it expresses an involved attitude towards it, whether of fascination, obsession or delectation, and in addition invites the viewer to engage their imagination along similar lines.

Erotic art typically aims not only to activate the viewer's sexuality but to reflect that of the artist. That is to say, erotic works usually embody a perspective on what is depicted that suggests sexual interest on the maker's part. Furthermore, the sense of sharing in what at least appears to have been sexually stimulating to the artist often plays a causal role in the viewer's own stimulation. It is worth emphasizing that the sexual response occasioned by erotic art occurs largely on the plane of imagination, consisting primarily of thoughts, images and feelings, and rarely goes as far as full physiological arousal; the upshot of engagement with erotic art is imagined desire as often as it is real desire.

As suggested above, the term 'erotic art', in its primary usage, covers art that is aimed at stimulating sexual thoughts and feelings in its target audience, and which at least minimally succeeds. But this gives way to two

secondary usages, according to which fulfilling either the intentional condition or the success condition (somewhat modified) independently qualifies a work as erotic. On the first such usage, a work counts as erotic if it is ostensibly aimed at stimulating sexual thoughts and feelings, even when it does not succeed in doing so; on the second, a work counts as erotic if it succeeds noticeably in stimulating viewers sexually, even when it is not intended or even apparently intended to do so. In addition, for works of art that are erotic in the central sense, it is natural to employ 'erotic' as a comparative term as well as a classificatory one. There are works one describes as mildly erotic and those one describes as highly erotic, depending on the degree of sexual involvement they tend to sustain.

Finally, perhaps some works of art reasonably accounted erotic neither aim at nor achieve viewer arousal (in other words, sexual thoughts, feelings or sensations directed towards what is depicted), but are instead erotic merely in virtue of the fact that they facilitate the imagining of erotic states of others, without erotic involvement as such on the viewer's part - without the viewer identifying with or entering into those states, either in reality or in imagination.

3 Instrumentally erotic art and anti-erotic art

With some erotic art, the evocation of erotic feelings is a secondary aim, employed or manipulated by the artist in order to achieve some further end, such as wry humour (for example, Tom Wesslemann's caricatures of pulchritude, in his 'Great American Nudes' series, or Mel Ramos's exaggeratedly voluptuous pin-ups), social commentary (Degas's monotypes of brothel scenes), or psychological disorientation (Magritte's or Dali's recombinant sexual imagery). We might label such art 'instrumentally' erotic art. As a result of these secondary aims, the excitatory tendency of such works may be weakened or even wholly neutralized. In some limiting cases works are in effect *about* erotic art - they comment on or satirically appropriate the conventions and mechanisms of normal erotic art - but without being erotic in the primary sense, that is, without being ultimately aimed at sexually engaging the viewer. Such art may be accounted erotic in virtue of the fact that it leads the viewer to question the presuppositions and consequences, social and otherwise, of erotic responses, without inviting or even permitting viewers to have such responses.

Some other cases of works representing sexual matters without appearing clearly erotic serve to illuminate further the boundaries of the category:

(1) Lysippus' *Aphrodite*, Botticelli's *Birth of Venus*, or Cranach's images of Eve occasion some hesitation if classified as erotic. Probably this is because we take the primary intent of the artist to have been to embody ideals of the human form, male and female, and not to prompt imaginative erotic engagement on the part of viewers of either sex. But this may be ingenuous; at any rate, such a line could not plausibly be extended to exclude from the erotic Donatello's sensuous, almost coquettish *David*.
(2) Picasso's *Demoiselles d'Avignon* occasions hesitation of a different sort. Though the painting presents women who are not only nude but in fact prostitutes, they are depicted in a highly nonrealistic mode, which shortcircuits erotic involvement, as well as drawing attention primarily to the painting's formal and expressive dimension.
(3) Judy Chicago's *The Dinner Party*, an elaborate sculptural installation, uses female genital imagery in a celebratory though arguably not erotic way; its sexual content is of a sort that is purely symbolic, rather than sexually involving.
(4) Though displaying some of the hallmarks of erotic art, Lucian Freud's paintings of naked subjects are not obviously erotic, being more evocative of the *boucherie* than the *boudoir* - an observation even truer of the images of nudes in Francis Bacon's paintings. Philip Pearlstein's superrealist figure paintings or, in another vein, Dubuffet's quasi-paleolithic images of squashed and splayed humanity belong here as well.

These latter artists - Bacon, Freud, Pearlstein, Dubuffet - are not aptly described merely as non-erotic - as is, say, a Corot landscape or Chardin still life - but rather as anti-erotic. Thus in a broader sense they, unlike Chardin or Corot, are erotic after all, that is to say, concerned with sexuality in a way that reflects the sexual interests of the maker and engages those of the viewer, if not in a positive manner. De Kooning's raw and primitivist images of women also come naturally to mind in this connection as well, though the case can be made that those images project a more ambivalent attitude to human sexuality than those of, say, Bacon - an admixture of terror, awe and admiration.

4 Covertly erotic art

It is relatively easy to give plausible examples of erotic works of art that contain no explicit depictions of sexuality or nakedness: Georgia O'Keeffe's landscapes and still lifes, with their oblique evocation of female anatomy; Caravaggio's paintings of Bacchus or St John the Baptist, with their coded references to homosexual experience; or Bernini's marble of St Teresa in spiritual ecstasy, a state readily transposed by the viewer into its profane cousin. The criterion of covert sexual content, however, remains unclear. Depiction of objects recognized as sexually symbolic - umbrellas or fruit, for example - especially when juxtaposed with human subjects, may be a typical indication of such content, but can hardly serve as a general mark. According to some writers, virtually all art has covert sexual content in virtue of being the expression of unconscious wishes or fantasies of a sexual sort. Thus, for Wollheim (1987), Ingres' history paintings, Bellotto's landscapes with buildings, and Poussin's landscapes with water are as substantially imbued with sexuality as Goya's *Naked Maja* or Titian's *Venus of Urbino*.

Even so, it seems that not all covertly sexual art is usefully considered erotic, but only that which is plausibly aimed, if unconsciously, at exciting sexual thoughts or feelings in target viewers, and which succeeds in doing so. In putative cases of covert sexual content, the arousal of the viewer of intended orientation and appropriate background may be just what signals the presence of such content and justifies its ascription.

5 The relationality of erotic art

If a painting is erotic, this is in virtue of its being aimed at and to some extent its eliciting an erotic response *from a certain class of viewer* - the work's intended or target audience. Such classes may be delimited not only by the requirements of sensitivity and background knowledge but also by less acquirable ones of physiological make-up or sexual orientation. Thus, a painting may be erotic in virtue of its being designed to produce an erotic reaction in heterosexual males, elderly homosexual males, young heterosexual girls, homosexual women, or bisexuals of either sex. There is a fact of the matter, albeit a hazy one, about whether a given painting is erotic, but it is an inherently relational one, whose nature is only fully evident when the group targeted for response is identified. Indeed, according to Nochlin (1988), '…. the very term "erotic art" is understood to imply the specification "erotic-for-men".' Still, once such implicit indexing has been made explicit, it may then be cancelled, so as to recognize art that is erotic relative to other target groups.

6 Social and political aspects of erotic art

Recent writers on erotic art stress the way in which entrenched genres and conventions of representation embody dominant ideas and assumptions about the nature of men and women and their proper relationship. Paintings such as Delacroix's *Death of Sardanapalus*, Gerome's *Oriental Slave Market*, Ingres' *The Turkish Bath* or *Jupiter and Thetis* lend themselves readily to such analysis. For example, Nochlin (1988) speaks of 'the power relations obtaining between men and women inscribed in visual representation' as a focus of her investigations.

Equally frequently noted is the element of voyeurism in erotic art. It is said that the spectator is a voyeur (at least fictionally), and that the work of art often reinforces or echoes this by depicting a spectator who, together with the viewer, regards the erotic object. Furthermore, the implicit or explicit voyeurism of erotic art is sometimes held to reflect the necessary impotence of the artist in respect of the imaginary and thus unattainable individuals depicted within their art.

Finally, the relationship of erotic art and pornography has been much debated. They may be distinguished, arguably, in at least two ways. First, pornography has, perhaps by definition, no significant artistic aspect. That is to say, pornography makes no credible appeal to viewers to consider the mode and means of depiction, as opposed merely to what is depicted; pornography, unlike art of any kind, is wholly transparent in both aim and effect. Second, pornography has, as a central intent and characteristic result, not only the stimulation of sexual feelings or fantasies in viewers, but the degradation, domination and depersonalization of what it depicts, usually women. Courbet's *Sleep*, which shows two beautiful nude women in the arms of Morpheus and each other, or Schiele's 1917 *Reclining Woman*, which presents its subject provocatively spread-legged and scarlet-nippled, perhaps court dismissal as pornography by some of these criteria, but on reflection they remain at a safe distance from it. Though the images in question are starkly arousing, even exploitative, the technique of their construction, the style in which they are rendered, the preceding art history they encapsulate, and the entrée they afford into

their makers' psyches, are at least as absorbing as what they flatly represent, and conspire to redeem them as art.

See also: Aesthetic attitude; Aesthetics and ethics; Art and morality; Emotion in response to art §§1, 2, 6; Feminist aesthetics §3; Freedom of speech; Pornography

JERROLD LEVINSON

References and further reading

Berger, J. (1972) *Ways of Seeing*, London: Penguin.(A politically sensitive tour of the history of pictorial representation.)

Clark, K. (1956, 1972) *The Nude: A Study in Ideal Form*, Princeton, NJ: Princeton University Press.(The classic study of its subject, which highlighted the distinction between the naked and the nude.)

Devereaux, M. (1990) 'Oppressive Texts, Resisting Readers and the Gendered Spectator: The ''New'' Aesthetics', *Journal of Aesthetics and Art Criticism* 48: 337-47.(Examines issues of representation in art from a feminist perspective.)

Hess, T. and Nochlin, L. (eds) (1972) *Woman as Sex Object: Studies in Erotic Art, 1730-1970*, New York: Newsweek, Inc.(A collection of essays accompanying an exhibition.)

Kronhausen, P. (1968) *Erotic Art*, Copenhagen: Uniprint A/S.(Introduction to catalogue, briefly addressing the problem of definition or delimitation of the sphere of erotic art.)

Lucie-Smith, E. (1972, 1991) *Sexuality in Western Art*, London: Thames & Hudson.(The standard survey of eroticism in the visual art of the West, focusing on painting.)

Mulvey, L. (1989) *Visual and Other Pleasures*, Bloomington, IN: Indiana University Press. (Includes 'Visual Pleasure and Narrative Cinema', an influential essay in feminist psychoanalytical film theory.)

Nead, L. (1992) *The Female Nude: Art, Obscenity and Sexuality*, London: Routledge. (Politically up-to-date successor to Clark's 1956 study.)

Nochlin, L. (1988) *Woman, Art, and Power and Other Essays*, New York: Harper & Row. (A collection of essays by the best known American feminist art historian.)

Pollock, G. (1988) *Vision and Difference: Femininity, Feminism and the Histories of Art*, London: Routledge.(A collection of essays by the best known British feminist art historian.)

Rose, J. (1986) *Sexuality in the Field of Vision*, London: Verso.(A study of sexuality in visual art by a prominent art critic.)

Scruton, R. (1986) *Sexual Desire: A Moral Philosophy of the Erotic*, New York: Free Press. (A wide-ranging treatment of the issues, with a valuable appendix on intentionality in general.)

Sircello, G. (1990) *Love and Beauty*, Princeton, NJ: Princeton University Press.(A philosophical study of its subjects, focusing on sexual love.)

Steinberg, L. (1972) 'Picasso's Sleepwatchers', in *Other Criteria*, London: Oxford University Press.(A brilliant essay on the erotic and other content of a particular slice of Picasso's oeuvre.)

Wollheim, R. (1987) *Painting as an Art*, Princeton, NJ: Princeton University Press.(A masterly survey, with a number of elaborate case studies ascribing sexual content to the work of painters in whom such content is not obvious.)

Error and illusion, Indian conceptions of

From the earliest Indian speculation, illusion has been key in the exposition of Indian mysticism. In classical philosophy proper, all schools take positions concerning illusion (sometimes called error) early in their histories, and in some schools successive refinements are achieved over the centuries.

There is a wealth of reflection on illusion from different angles - for example, psychological treatments that differentiate seven, eight or ten types of illusion, identifying causal factors for each variety. Illusion is taken to have ontological and epistemological ramifications brought out in elaborations of one or another metaphysical system.

The sections of classical philosophical texts devoted specifically to illusion generally presuppose or smuggle in criteria of veridical awareness in the midst of causal analyses or systematic explanations of such stock examples as a snake appearing as a rope, a piece of shell appearing as silver, two moons, a 'red' crystal with a red flower behind it, a mirage, the 'blue' of the sky, and dreams. Arguments in support of a criterion, or set of criteria, of veridicality do appear, however. How we understand (1) what counts as a veridical and a nonveridical experience and (2) why individual cases of illusion occur are distinct issues, but positions taken on the former determine important parts of conceptions about why illusions occur. A peculiar stylistic feature of the later and more refined philosophic treatments is a polemical ordering, where a first view's inadequacy is shown to lead to a second view whose inadequacy, demonstrated in turn, leads on to a better theory, and so on, until we reach the right view, which is thus established by a sequence of arguments.

1 Illusion in Indian metaphysical systems

In proclaiming the supreme value of a mystical experience, both Upaniṣadic (*c.*800 BC and on) and Buddhist texts (*c.*300 BC and on) put illusion to use as an analogy. This is most dramatically exemplified in the epithet *buddha*, which in Sanskrit is the past passive participle of the common verb for 'to awaken': thus 'the Awakened', implying that those who do not have the *nirvāṇa* experience need to awaken from their all too unpleasant dreams to a blissful reality (see Nirvāṇa). Similarly, in the Advaita Vedānta philosophy, which is one of two or three major classical systematizations of the thought of the early Upaniṣads, an illusion analogy is central to a teaching of an experience of Brahman, the Absolute and One, as supremely valuable (see Brahman).

It is worth dwelling on illusion within the Advaita system. The use of illusion as a metaphor for mysticism is generalized to a metaphysical position: as the illusory snake is to the real rope, so the manifest universe, in its diversity, is to Brahman, the supreme reality, who is One and Nondual. Later Advaitins bring out the tension between diverse appearances and Brahman's unity. The great eighth-century philosopher Śaōkara (§2) is more concerned with defending an understanding of mystical awareness of Brahman: like the experience of a rope sublating a snake appearance, Brahman-awareness sublates all other modes of experience.

Śaōkara uses the term *adhyāsa*, commonly translated 'superimposition', to indicate the relation of the world to Brahman. But that is from our point of view. Strictly speaking, there is no relation of Brahman *to* the world (no two-way street), since (as maintained by the early interpreter Padmapāda), the snake simply does not exist. On our side, the world, like the snake, does of course appear. Consonantly, Advaitins say (again from 'our side', from the perspective of spiritual ignorance) that the world has a problematic ontic status 'impossible to explain' (*anirvacanīya*). Matching such a conception is the view (see §3) that the content of an illusory awareness cannot be explained (*anirvacanīya*) as real or not real. In applying to the world the category 'being impossible to explain' (*anirvacanīyatva*), Advaitins generate an entire metaphysics. The world is not not-real because it appears (unlike the horn of a rabbit), but it should not be said to be real because it is (or is to be) sublated.

Illusion likewise plays a central role in Buddhist philosophy, both with the Mādhyamika of Nāgārjuna (*c.*150-200 CE) and his followers, and the Yogācāra of Vasubandhu (*c.*400), Dignāga (b. *circa* 480), Dharmakīrti (*c.*600-60) and company (see Buddhism, Mādhyamika: India and Tibet; Buddhism, Yogācāra school of). Nāgārjuna says that his seeming to maintain positive philosophical positions in spite of his proclamation that everything is empty is to be understood as analogous to the unreal words of an imaginary, magical man. 'Everything is empty' is no exception to itself, since that statement too is empty, like statements in an illusion.

Nāgārjuna's appeal to this possibility seems integral to his entire outlook, but just what that outlook is (must a sceptic have an outlook?) is controversial among modern interpreters, as indeed it was in classical debates. A more straightforwardly theoretical approach to illusion is made by Vasubandhu and his school. Though we may suspect peculiarly Buddhist motivations having to do with defending the possibility of *nirvāṇa* experience, the great Yogācārin launches a project known in the West from Cartesian epistemology - or from what is called the argument from illusion for phenomenalism, and a perceptual core common to illusion and veridical experience (see Perception, epistemic issues in §5). Vasubandhu embraces a phenomenalism of atomic awareness-events, events of colour, sound, feeling, and so on, arguing that these are what we directly experience, as is proved by illusion. There must be a phenomenal core common to both veridical and illusory awareness since illusions do occur that are indistinguishable from veridical experiences, at least at the time of their occurrence.

In subsequent philosophic discussion both within Buddhist traditions and outside, the term 'error' (Sanskrit, *bhrama*, *bhrānti*) is used to refer to perceptual illusion in a neutral, non-question-begging fashion. Usually the term is not used to refer to other types of cognitive failure, such as false claims or faulty logic. However, on some views - notably taken by some Yogācārins as well as proponents of an influential Mīmāṃsaka stance (to be explained) - the error in what is ordinarily taken to be perceptual illusion is attributed to nonexperiential factors. There is no nonveridical experience according to these views. Thus the explananda targeted by these theories are patterns of everyday speech and behaviour where we commonly take cases of (experiential) error, or illusion, to occur.

Later Yogācārins view illusion pragmatically, but Dignāga, an important theorist standing about midway in the history of the school, sees error as a matter of incorrect interpretation, with no illusion on the level of experience itself. Although he claims that his epistemological positions are independent of such considerations as whether things known are viewed as 'external' or (as Buddhists are wont to say) 'Void', it seems plausible that the metaphysics of atomic awareness-events contributed to his interpretationalist stance - for how with a phenomenal core seen as reality could there be illusion in experience?

Philosophers belonging to Indian realist schools are called upon to defend their views against objections based on the occurrence of illusions. The great names from the Mīmāṃsā and the Nyāya-Vaiśeṣika traditions forge positions on illusion in consonance with a realism about the objects revealed by veridical experiences - and, surprisingly, the objects of nonveridical experiences.

2 Criteria of veridicality

The question of the nature of veridicality - what veridicality is - connects with the question of how it is known, that is, how an awareness is known to be veridical. A centrepiece of Nyāya-Vaiśeṣika realism (henceforth, for convenience, simply 'Nyāya') is what may be called a correspondence view: facts are what they are independently of consciousness (about which there are other facts), and what it is for an awareness to be veridical is for it to reveal an object, or, more precisely, an objective complex. The simplest veridical awareness makes known an objective complex: '*a* is *F*' just in case *F* (a qualifier) qualifies *a* (the qualificandum) - or, to suggest more explicitly the tie to Nyāya's ontology, just in case *a* has *F*-hood. In other words, an awareness' propositional content (*Fa*) must reflect what is in fact (*a*'s being *F*) if the awareness is veridical, according to Nyāya. An awareness with the content *Fa* is erroneous just in case *a* is not qualified by *F*. The world contains both *a* and *F* even, as will be explained, in cases of error. Finally, to be accurate, the fourteenth century philosopher Gaṅgeśa (§2), a leader of the New Nyāya school, and others as well, criticize characterizations of veridicality as simply correspondence with fact (*yāthārthya*), demanding more precise formulations. But a correspondence view may be said to be the spirit of Nyāya throughout its two-millennia history (see Nyāya-Vaiśeṣika §6).

The question of the marks whereby such a match is itself discernible is not an easy one for the realist, the Naiyāyika or any other. Before reviewing Nyāya's proposals, let us look at two other views about what veridicality is, in Mīmāṃsā - specifically, the Prābhākara branch - and in Yogācāra.

Contrasting with the Nyāya position is Mīmāṃsā, a realist school that assumes an unexpected stance (see Mīmāṃsā §2). Veridicality is identified with awareness itself: all awarenesses are veridical. The Buddhist argument from illusion to find a phenomenal core is met head on by the (Prābhākara) Mīmāṃsaka: the nature of all awareness is to reveal objects in a person's immediate environment. The question of how an awareness is known

to be veridical is understood as asking how an awareness, our own awareness, is known to have occurred. (The Prābhākara answer is, like that of the Yogācāra Buddhist, that every awareness is self-aware, while also - a departure from the Buddhist position - making the self known: thus something about the world, the awareness itself and the cognizer are all made known by a single awareness.) Why take it as given, the Prābhākara asks, that any awareness misrepresents the world? It is natural for awareness to present things as they are, and so another source of error should be assumed. Common speech and other behaviour do of course show that error occurs; this is admitted by all disputants. But no criteria for differentiating the veridical and the nonveridical are needed. Nor need we say what veridicality is, apart from specifying what awareness is and how it is different from other realities.

Coherence, or agreement with other awarenesses, is both the nature of veridicality and the criterion whereby it is discerned, according to the later Yogācāra of Dharmakīrti and others. To be sure, Dharmakīrti proposes that there are pragmatic criteria - whether action based on an (interpreted or proposition-laden) awareness leads to successful results - but since the results are themselves (proposition-laden) awarenesses, on his view, coherence seems to remain the criterion, as it clearly is with earlier members of his school.

Finally, as early as Vātsyāyana (*c*.350), whose commentary on the *Nyāyasūtra* is the oldest extant - indeed in the very opening passage of Vātsyāyana's commentary and throughout the history of the school - a pragmatic criterion is proffered as the test of veridicality. Veridical awarenesses are proved so by their fruits; nonveridical awarenesses lead to error in action, preventing us from getting what we want and avoiding things to which we are averse.

Again, what it is for an awareness to be veridical is for it to reveal a fact; but we do not know from the occurrence of an awareness *A* itself that *A* is veridical (except in the case of an apperception, where what is known is a preceding awareness). We infer *A*'s veridicality or nonveridicality - if circumstances call it into question - from whether it has proved workable in helping us get what we want and avoid our aversions.

The Nyāya pragmatism does not amount to a hidden coherentism, although to judge from the accounts given of particular types of illusion, coherence considerations are, as one might expect, appealed to. Veridical awarenesses are the results of reliable sources, and are themselves the foundations of warranted statements. If there is reasonable doubt about a statement's truth, then one turns to reliable sources of knowledge, such as perception, to resolve the doubt. According to Nyāya, it may be impossible to tell at the time of occurrence whether an ordinarily presumed veridical perception is in fact veridical, and if there is reasonable doubt about this it is here that the pragmatic criterion comes into play. Thus the epistemological default is a fallibilistic presumption of veridicality, and we are presumed not ordinarily to be called to wonder about error in our awarenesses.

3 A polemical survey of conceptions

An illuminating attempt to refute non-Naiyāyika views of error is made by the modern Nyāya philosopher Ramanuja Tatacharya in his *Pratyakṣatattvacintāmaṇi Vimarśa (Consideration of the Tattvacintāmani's [Treatment of] Perception*; 1992), which is based on the thought of Gaógeśa. The polemical ordering of his argument - whereby successive refinements or corrections of a view are shown to be inadequate until the correct one is arrived at - is typical of later classical philosophy.

Tatacharya starts by characterizing the Yogācāra view as *ātma-khyāti* ('illusion is "self-" or "internal" manifestation': *khyāti* means 'perception' or 'manifestation' and is used in a neutral way with respect to all theories considered; *ātma* means 'self', and since there is no self on the Buddhist view the term should be taken as 'internal'). With the error 'This is silver', when the object in front is instead a piece of shell, the Yogācārin holds, says Tatacharya, that the silver form that appears is an internal manifestation and that the correcting cognition 'This is a piece of shell' negates only the 'thisness' of the error. The silver form is uncorrected by the succeeding moment of consciousness (it has simply passed away, though also helping to give rise to the consciousness of a piece of shell, according to Yogācāra psychology). Illusion shows that the silver form has to be internal; 'thisness' with respect to silver cannot be external. Therefore, illusion is an internal manifestation (due, like all occurrences of consciousness, to dispositions, some common to other persons, some not).

However, if this view were correct, Tatacharya continues, the error would be appropriately verbalized as 'I am silver', not 'This is silver'. Moreover, since it is held that the silver form is no illusion, we should be able to use it

in the market. Clearly, the silver, which is the object of the illusion, does not exist, as the correcting cognition, ‘Not (this-is-silver)’, shows. The right view would then be *asat-khyāti*, ‘manifestation of the nonexistent’, a view attributable to the Mādhyamika school of Nāgārjuna. How the nonexistent can be the object of a cognition should not be asked, for cognitive objecthood is not to be understood causally, but in its own way; the existent and nonexistent are both within its range. Error is a cognition with a nonexistent object.

However, with reference to the error ‘This is silver’, the Nyāya philosopher asks whether what is nonexistent is just the silver or also the ‘this’. It cannot be the latter, for the deluded person desiring silver reaches out to pick up the piece of shell (the ‘this’), and no such action would be taken with respect to something nonexistent. And with regard to the silver, is it nonexistent everywhere or is it simply not at this locus or place? It cannot be the former, for the universally nonexistent does not appear (whereas silver is appearing in the illusion). The right view is that cognition reflects reality (*yāthārthya*); only the existent appears (*sat-khyāti*). Error is not a single cognition, but two cognitions, a perceiving and a remembering, with their distinctness unrecognized; this is *akhyāti*, ‘nonmanifestation (of the distinctness between a perception and a memory)’, a view attributable to the Prābhākara Mīmāṃsaka. What gets verbalized as ‘this’ is reflected by a (veridical) perception of an object in the perceiver’s immediate environment; what gets verbalized as ‘silver’ is reflected by a (veridical) memory acquired from previous experience of silver. Due to a causal irregularity - various causal factors influence cognition - the class character of the piece of shell (namely, ‘piece-of-shell-hood’) is not apprehended, but its similarity to silverhood, along with the causal irregularity, triggers a silverhood memory. (Memories are stored associatively.) The two cognitions get fused, with no awareness of their distinctness. The deluded person thus reaches out to pick up what is presented experientially. The correcting cognition works with respect to the failure to grasp the distinctness. When that failure ceases, the activity based on it ceases. Error is just a matter of activity (including speech-behaviour) that can come to be corrected, not of a cognition’s object or content. Neither perceiving nor remembering manifests anything except the real. All cognition reflects reality. Thus is the right view established.

However, from the Nyāya perspective, it is a person looking to gain silver who moves to pick up the ‘this’ as indicated by the error ‘This is silver’, not a person who does not see it *as* silver. That someone simply fails to be aware - from not grasping a distinction between a memory and a perception - that the thing in front is not silver is insufficient to prompt the act of reaching out to pick it up. It must be cognized *as* silver. Therefore, the erroneous cognition is single but, like all verbalizable cognition, complex: what is verbalized as ‘this’ is the cognition’s indicating a qualificandum, a bearer of properties, and what is verbalized as ‘silver’ is the cognition’s predication content, a qualifier that the qualificandum is presented as having. This analysis gives us, finally, the truly right view, *anyathā-khyāti*, ‘manifestation (of a qualifier) elsewhere (than where in fact it obtains)’. A causal story is indeed to be told of why a particular illusion occurs, and of how such-and-such a cognition, whose predication content does not qualify the qualificandum cognized, can be generated. Cognition is a complicated process, and the possibilities for something going wrong are numerous, depending on what is cognized, the sense organs, internal dispositions, and other factors.

See also: Epistemology, Indian schools of; Knowledge, Indian views of; Sense perception, Indian views of

STEPHEN H. PHILLIPS

References and further reading

Gaōgeśa (first half of 14th century) *Tattvacintāmaṇi*, vol. 1, *pratyakṣa-khaṇḍa*, with the *Prakāśa* commentary by Rucidatta Miśra and a subcommentary by Rāmakṛṣṇādhvarin, ed. N.S.R. Tatacharya, Kendriya Sanskrit Vidypeetha Series 20, Tirupati: Kendriya Sanskrit Vidypeetha, 1972.(Contains a section on *anyathā-khyāti-vāda*, elaborated in Tatacharya 1992.)

Matilal, B.K. (1986) *Perception: An Essay on Classical Indian Theories of Knowledge*, Oxford: Oxford University Press.(The best philosophical study to appear on issues surrounding perception as understood in classical Indian philosophies; highly readable and strongly recommended.)

Mohanty, J.N. (1966) *Gaōgeśa’s Theory of Truth*, Delhi: Motilal Banarsidass, 1989. (Contains an annotated translation of Gaōgeśa’s reflections on how the veridicality of an awareness is cognized.)

Phillips, S.H. (1995) *Classical Indian Metaphysics: Refutations of Realism and the Emergence of ‘New Logic’*, La Salle, IL: Open Court.(Contains an annotated translation of Gaōgeśa’s reflections on how veridicality should be characterized or defined.)

Sinha, J. (1958) *Indian Psychology: Cognition*, Calcutta: Sinha Publishing.(Contains a wealth of information about classical conceptions of illusion; requires in places an understanding of Sanskrit philosophical terminology.)

Tatacharya, N.S.R. (1992) *Pratyakṣatattvacintāmaṇi Vimarśa: A Comprehensive and Evaluative Study of the Pratyakṣakhaṇḍa of the Tattvacintāmaṇi*, Tirupati: Rastriya Samskrita Vidyapitha.(In Sanskrit; the principal source for §3, which takes one about halfway through Tatacharya's defence of the Nyāya position.)

Eschatology

Eschatology is the study of or doctrine about the end of history or the last things. Eschatology is a branch of Christian theology, and the term still finds its primary home in that context, but it is also used broadly to cover any theory about the end of human life or of the world.

There are many types of eschatological theory. Some of the most important are those of Plato, Vedāntic Hinduism, Karl Marx and Christianity. The contemporary philosopher of religion who makes most use of eschatology in his thinking is doubtless John Hick. There are several issues that are of interest to philosophers in the area of eschatology. Among them are such questions as whether there is good reason to believe that human life and/or history are moving towards a final end; whether personal identity problems are solvable in the eschaton *(the end-state); whether eschatological considerations can help philosophers address other philosophical problems (for example, the problem of evil); whether the very notion of disembodied survival of death is coherent; and how (in Christian theology especially) immortality of the soul and bodily resurrection are related.*

1 Types and varieties of eschatological theory

An eschatological theory can be either religious (as in Christianity) or secular (as in the classless society anticipated by Marx). But it is important to note that eschatology always presupposes a linear view of history. The common assumption of all eschatology is that human life and/or history are moving in a certain direction, towards a certain end or goal (see History, philosophy of §§1-2). Accordingly, circular views of history, or even linear views of time with no envisaged afterlife or end-state, have no eschatology.

What we might call a theory of *individual eschatology* presupposes that there is life after death for human beings and deals with their postmortem fate. The three main options for those who affirm an individual eschatology are reincarnation, immortality of the soul, and bodily resurrection (see Reincarnation; Soul, nature and immortality of the; Resurrection). What we might call a theory of *cosmic eschatology* deals with the larger issue of the goal or end towards which (so it is claimed) history is moving and with what happens at the end of history or of the world.

Plato (§13) is an example of a philosopher who argued for an individual eschatology but no cosmic eschatology. It is exceedingly difficult to combine into one coherent theory everything that Plato said in his dialogues about the soul and its life after death. Nevertheless, in both the *Phaedo* and the *Republic* he argued that the human soul, which he took to be the essence of the person, is indestructible and thus immortal. It survives death and faces judgment for its earthly deeds. Since Plato was convinced of the moral governance of the universe, he expected that this judgment would be just, with punishment for evil people and reward for the righteous. He did not dogmatize about the details of the afterlife, offering only ‘possible’ or mythic explanations. This much can certainly be said about Plato’s individual eschatology. However, there seems to be no notion in his writings of history moving towards a goal or end.

Theories of reincarnation, at least those that posit a possible final escape from the cycle of life, death and rebirth, also typically suggest an individual eschatology, but no cosmic eschatology. The Vedāntic school of Hinduism affords an example. Ultimate reality is Brahman; it is pure being or pure consciousness and is beyond all qualities. The phenomenal world is *māyā*, the realm of illusion. *Māyā* is the product of the creative power of Brahman, not through conscious acts of creation, but through Brahman’s emitting successive waves or emanations of reality. *Māyā* is unreal in the sense of not being ultimately real, of being temporary and contingent, unlike Brahman. One aspect of *māyā* is the existence of *jīvas*, individual souls that are incarnate as empirical selves. Limitless numbers of *jīvas* have always existed and are born in various bodies, both human and nonhuman. After the death of one body, they are reborn in another, a *jīva*’s station in life being a function of its deeds in previous lives.

Karma is an impersonal law to the effect that *jīvas* reap in one life what they have sowed in previous lives. The karmic imprint is carried from one incarnation to the next by what is called the subtle body. Western philosophers would call it mental; it is a body in the sense that it is part of the realm of *māyā*. It registers and transmits the moral and spiritual influence of previous lives. Through a series of lives, it is possible for *jīvas* to be purged of ignorance (*avidyā*) and evil, and to attain true self-consciousness, which is consciousness of their own essential oneness with Brahman and of the illusoriness of all differentiation (reality for Vedāntists is non-dual). If *ātman* is the

non-empirical self, the eternal and immutable spiritual reality of each *jīva*, then liberation (*mokṣa*) and escape from rebirth are achieved when one realizes that '*ātman* is Brahman'. In other words, an empirical self is an expression of a *jīva*, which is itself ultimately one with *ātman*, which is itself ultimately one with Brahman.

Karl Marx (§12) is an example of a philosopher who argued for a cosmic eschatology but no individual eschatology. Marx believed neither in God nor survival of death, but he did posit an end-state towards which history is irresistibly moving, namely increasing suffering of the workers, increasing contradictions in capitalism, world revolution, the dictatorship of the proletariat, and a final ideal society. In this society there will be common ownership of property and of the means of production; productivity will increase and prosperity will prevail; and there will be no wages, government or class structure.

Like most historical religions, Christianity posits both an individual and a cosmic eschatology. Christian eschatological notions are not easy to summarize, and eschatology is an area of great disagreement among theologians. The idea of personal survival of death arose late in the Hebrew Bible, but the hope of national and spiritual revival on the 'Day of the Lord', when God will judge the wicked, vindicate the righteous and restore the nation of Israel, was ancient. In the first century AD, the Pharisaical party accepted the idea of life after death while the Sadducees did not.

Jesus announced the end of the present age and the arrival of the kingdom of God, and the New Testament writers interpreted his life, death and resurrection as the defeat of all the forces that oppose God's rule. They also held that Jesus' resurrection from the dead is a promise of a general resurrection and of eternal life for those who believe in him. In one sense the kingdom of God arrived with the advent of Jesus; in another sense, the sense in which it arrives publicly and with power, it is postponed until the end-time. Thus Christian eschatology points to the end-time as the period when Jesus will return and the divine redemptive purposes for creation will be fulfilled. There will be a new heaven and a new earth. People will be raised from the dead and judged; the damned will live forever apart from the presence of God in hell, and the blessed will 'see God' in heaven (see Hell; Heaven).

2 John Hick on eschatology

Among contemporary philosophers of religion, John Hick has made the most sustained use of eschatology. In his earlier writings he introduced the notion of 'eschatological verification'. In response to critics such as Antony Flew who argued that theological statements like 'God exists' and 'Jesus is the Son of God' are unfalsifiable and thus without cognitive content, Hick argued that the mere possibility that such statements might be verified in some future *eschaton* (one day we might have, for example, an unambiguous experience of 'the reign of Christ in the kingdom of God') renders those statements cognitively meaningful here and now (1957: ch. 7).

Following Origen (§4) and other theologians, Hick also posits an eschatological universalism in which every human person will eventually achieve full consciousness of God and thus salvation or liberation. There is no hell, at least no permanent hell. In *Death and Eternal Life* (1976), Hick argues for a possible afterlife in which this might occur, one involving various lives in different worlds. In the end-state, Hick says, personality is retained but egoicity is transcended. Hick also uses universalism as an important element in his soul-making theodicy. Unless everyone will eventually enter a limitlessly good end-state, he says, the problem of evil cannot be solved. Hick thus holds that there can be no successful theodicy apart from eschatology.

3 Philosophical issues in eschatology

There are many issues of interest to philosophers in eschatology. Four will be briefly noted, and a fifth discussed.

(1) Perhaps the most important issue is the question of whether there is any good reason to believe that human beings will survive death and/or whether history is moving towards a final end. This question obviously cannot be answered apart from larger questions about what sort of world we live in, and whether it can be fully described in entirely naturalistic terms or whether some sort of God or divine reality exists (see God, arguments for the existence of). As far as individual eschatology is concerned, most philosophers seem to have lost interest in the classic 'proofs' for survival of death, such as those of Plato or Kant (§11). Some look to such empirical or psychological phenomena as near-death experiences, psychical research or purported memories of previous lives. But in general the project of offering philosophical arguments for or against survival of death is not nearly as popular as it once was.

(2) One of the most interesting philosophical issues related to eschatology is that of personal identity in the afterlife, whether the denizens of the afterlife will be identical to the ante-mortem persons they will apparently claim to be. Will the traditional criteria of personal identity (see Personal identity) suffice to establish that identity? If not, will 'closest continuer' or other notions less robust than identity satisfy the concerns of the religious systems in which most survival theories are embedded?

(3) Another important question is whether eschatological considerations can or should play any role in solving the problem of evil, as Hick insists that they can (see Evil, problem of §4). For Christian theodicists at least, Hick seems right. As the apostle Paul said, 'I consider that the sufferings of this present time are not worth comparing to the glory about to be revealed to us' (Romans 8: 18). Critics will respond that it is question-begging or in some other way unconvincing to 'solve' the problem by appealing to a future outcome that is hoped for but in no way assured. Theodicists like Hick can reply that it is rationally permissible and even essential that they respond to the problem of evil from the perspective of their own beliefs or those of their religious community, which may well include belief in an afterlife and eschatology.

(4) In an often-reprinted article (1953), H.H. Price argued for the coherence of the notion of a purely mental world, thus supporting the possibility of disembodied survival of death. Price posited a world of mental images (visual, auditory, telepathic, and so on) in which souls are aware of each other, live in a coherent 'world' (consisting of mental images rather than physical objects), communicate with each other telepathically, and have dreamlike (rather than bodily) perceptions of their world. According to Price, such a world may have different causal laws from our present world (for example, wish-fulfilment will be powerfully operative), but it will seem to its denizens just as real and even as 'solid' a world as ours does to us.

Some have detected problems in Price's argument. Hick, for example, notes that it is hard to combine Price's insistence on the public, non-solipsistic nature of the world he describes with his notion that the character of the world will be in large part a function of the wishes of its inhabitants (1976: 265-77). If it is a public world, sustained by telepathic contact between different individuals, how can wish-fulfilment be as potent a force in shaping it as Price suggests? Hick may well be correct here in detecting a tension in Price's account, but the problem seems solvable. And if it is solvable, then Price's main point - that disembodied survival of death is logically possible - is correct.

(5) A question relating specifically to Christian eschatology is that of the relationship between immortality of the soul and bodily resurrection. Traditional Christian theories, such as those of Aquinas, combine the two, arguing that after death there is an interim period during which we exist merely as disembodied souls, and that only later, at the end-time, will our bodies be miraculously raised by God and reunited with our souls, thus making us whole and complete persons again.

This picture now faces three quite different sorts of challenge. First, the two theories have been radically separated by the New Testament scholar Oscar Cullmann (1973). He argues that only resurrection is a genuinely Christian notion; immortality is an alien theory imported from Greek philosophy. It played no role in primitive Christian conceptions of the afterlife and ought not to be part of Christian belief today. Second, several contemporary philosophers (such as Antony Flew and Terence Penelhum) argue that the very concept of disembodied existence is either incoherent or deeply problematic. There are, however, other philosophers (such as John Hick and Richard Swinburne) who strongly defend the coherence of the concept. Third, many contemporary liberal Christian theologians (such as Hans Küng) are inclined to dismiss bodily resurrection in favour of a theory (usually only vaguely spelled out) called 'spiritual resurrection'. As with virtually all Christian theologians, their eschatological views presuppose certain views (in this case revisionist ones) of Jesus and of the reliability of the biblical witness to Jesus. They claim that bodily resurrection amounts to a superstitious and unchristian theory of resuscitation; what was raised was not Jesus' body but his person or self. On the other hand, there are contemporary philosophers and theologians who defend more traditional Christian notions of resurrection, both the resurrection of Jesus (for example, William Craig) and the general resurrection (for example, Stephen Davis).

Returning to Cullmann, he is entirely correct in differentiating immortality from resurrection. Immortal souls naturally survive death; for resurrected persons, on the other hand, survival of death occurs only because of a miraculous intervention by God. On the view of souls as naturally immortal, the human body is the intrinsically evil prison-house of the soul; on the view that bodies are resurrected, both body and soul were created good by

God, are now equally corrupt due to sin, and are thus equally in need of redemption. On the first theory, death is a friend (at least, as Plato says, to the philosopher) precisely because it means escape from the body; on the second, death is a fearsome enemy that must be defeated.

Still, Cullmann seems too hasty in rejecting any notion of immortality as unchristian. The traditional or Thomistic notion of the bodiless interim state can easily be purged of the problematic aspects that Cullmann correctly sees in Plato's theory. The survival of death of even the soul can be attributed to a gracious and miraculous act of God; it can be affirmed that both body and soul are sinful and must be redeemed; and death can still be seen as an enemy. Furthermore, the traditional notion seems to offer a way of resolving a possible tension in the New Testament concerning the time of entry into eternal life. Pauline thought especially seems committed to the idea that the general resurrection occurs at the end-time (2 Thessalonians 2: 1-15); yet Jesus said to the 'good thief' on the cross, 'Today, you will be with me in paradise' (Luke 23: 43). The solution, then, is that the thief's soul went immediately to paradise, only much later in the end-time to be reunited with his body.

See also: Feminist theology §2

STEPHEN T. DAVIS

References and further reading

Aquinas, T. (1259-64) *Summa contra gentiles*, Notre Dame, IN: Notre Dame University Press, 1975, Book IV.(A classic statement of the traditional Christian view of personal eschatology.)

Badham, P. and Badham, L. (1982) *Immortality or Extinction?*, London: Macmillan.(A clear discussion of various options in personal eschatology and an assessment of the relevant evidence for and against survival of death.)

Craig, W.L. (1989) *Assessing the New Testament Evidence for the Historicity of the Resurrection of Jesus*, Lewiston, NY: Edwin Mellen Press.(An energetic defence of the resurrection of Jesus by a philosopher who is also trained in theology.)

Cullmann, O. (1973) 'Immortality of the Soul or Resurrection of the Dead?', in T. Penelhum (ed.) *Immortality*, Belmont, CA: Wadsworth Publishing Co.(An influential article in twentieth-century Christian eschatology.)

Davis, S.T. (1993) *Risen Indeed: Making Sense of the Resurrection*, Grand Rapids, MI: Eerdmans.(A defence of a fairly traditional notion of the resurrection of Jesus and of the general resurrection.)

Flew, A. (1987) *The Logic of Mortality*, Oxford: Blackwell.(Criticisms of survival-of-death theories.)

Hick, J. (1957) *Faith and Knowledge*, London: Macmillan, 2nd edn, 1988. (Eschatological verification is explained in Chapter 7.)

Hick, J. (1966) *Evil and the God of Love*, New York: Harper & Row.(Although an early work, this book is the fullest statement of Hick's 'soul-making' theodicy. Later writings do not change the argument appreciably.)

Hick, J. (1976) *Death and Eternal Life*, New York: Harper & Row.(Part 5 is a clear statement of Hick's eschatological views in the 1970s. Some of them are updated in his later writings.)

Küng, H. (1976) *On Being A Christian*, New York: Doubleday.(A summary of Küng's liberal Catholic theology. Part A, chapter 5 discusses resurrection.)

Marx, K. and Engels, F. (1848) *The Communist Manifesto*, in L.S. Feuer (ed.) *Marx and Engels: Basic Writings on Politics and Philosophy*, Garden City, NY: Doubleday, 1959. (A brief statement of Marxism and of Marx's eschatology.)

Penelhum, T. (1970) *Survival and Disembodied Existence*, New York: Humanities Press.(The author no longer appears to agree with everything that he said in this book, but it is a clear and well-argued critique of the notion of disembodied existence.)

Phillips, D.Z. (1970) *Death and Immortality*, London: Macmillan.(A Wittgensteinian critique of traditional conceptions of immortality.)

Plato (*c*.428-347 BC) *Phaedo*, in E. Hamilton and H. Cairns (eds) *The Collected Dialogues of Plato*, New York: Pantheon Books, 1961.(Concerned with immortality and the rewards of justice.)

Plato (*c*.428-347 BC) *The Republic*, in E. Hamilton and H. Cairns (eds) *The Collected Dialogues of Plato*, New York: Pantheon Books, 1961.(Book 10 deals with immortality and the rewards of justice.)

Price, H.H. (1953) 'Survival and the Idea of "Another World"', *Proceedings of the Society for Psychical Research* 50 (182): 1-25.(A classic argument for the coherence of the notion of a purely mental postmortem world.)

Reichenbach, B. (1983) *Is Man The Phoenix? A Study of Immortality*, Washington, DC: University Press of America.(A helpful discussion of immortality and its relation to resurrection.)

Reichenbach, B. (1990) *The Law of Karma*, London: Macmillan.(A Western philosopher analyses and criticizes

aspects of karma and reincarnation in Hinduism and Buddhism.)
Swinburne, R. (1986) *The Evolution of the Soul*, Oxford: Oxford University Press.(A defence of what the author calls ‘soft dualism’.)

Essentialism

Essentialists maintain that an object's properties are not all on an equal footing: some are 'essential' to it and the rest only 'accidental'. The hard part is to explain what 'essential' means.

The essential properties of a thing are the ones it needs to possess to be the thing it is. But this can be taken in several ways. Traditionally it was held that F is essential to x if and only if to be F is part of 'what x is', as elucidated in the definition of x. Since the 1950s, however, this definitional conception of essence has been losing ground to the modal conception: x is essentially F if and only if necessarily whatever is x has the property F; equivalently, x must be F to exist at all. A further approach conceives the essential properties of x as those which underlie and account for the bulk of its other properties. This entry emphasizes the modal conception of essentiality.

Acceptance of some form of the essential/accidental distinction appears to be implicit in the very practice of metaphysics. For what interests the metaphysician is not just any old feature of a thing, but the properties that make it the thing it is. The essential/accidental distinction helps in other words to demarcate the subject matter of metaphysics. But it also constitutes a part of that subject matter. If objects have certain of their properties in a specially fundamental way, then this is a phenomenon of great metaphysical significance.

1 Anti-essentialism

Essentialists have two basic commitments: to the essential/accidental distinction as such; and to the existence of properties of both types. Accordingly there are two main schools of anti-essentialism. 'Sceptical anti-essentialists' reject the very idea of essential versus accidental, while 'trivializing anti-essentialists' insist that all or nearly all properties fall on the same side of the line.

Sceptics typically argue as follows. Whether *x* is essentially *F* is supposed to turn on whether it is necessary that *x* be *F*. But this leads to contradictions. Is nine essentially greater than seven? Yes, because nine is seven plus two, and it is necessary that seven plus two exceeds seven. Yet also no, for nine is the number of planets, and it need not have been that the numbers of planets exceeded seven. The only way out is to admit that nothing is essentially *F* as such but only as described in a certain way. So-called 'essential' properties are really just properties entailed by some currently salient description.

What ought to make us suspicious is that similar worries can be raised about intuitively quite innocent distinctions, such as that between the constant properties of a thing - those that it always possesses - and its temporary ones. While it is always the case that seven plus two exceeds seven, for example, the number of planets was (let us imagine) once six. Described one way, then, nine is constantly greater than seven, while described another it is only temporarily greater than seven.

Here the fallacy seems clear. For the purposes of assessing constancy, 'it is always the case that the *D* is *F*' must be read *de re*: 'concerning the object which is in fact the *D*, it is always *F*' (see De re/de dicto). Read *de re*, the objector's claim that at one time the number of planets was six is simply false because nine was never six. Might not a similar response be available to the essentialist? It certainly might, unless the sceptic can convince us that *de re* modal talk is less intelligible than its temporal analogue.

Sceptics have tried, complaining that there is nothing on the modal side to match our well-developed criteria of identity over time (see Possible worlds §1). But since temporal criteria have clear trans-world implications (that a thing can evolve in such-and-such ways shows it to be capable of such-and-such otherworldly careers), the latter cannot be *too* shapeless without dragging the former into similar disrepute. Anyway, few essentialists would grant the sceptic's assumption that without trans-world identity criteria, *de re* modal discourse becomes emptied of content. If anything, content flows in the other direction; to call a counterfactual object identical to *x* is just to say that its properties are *ipso facto* properties *x* could have possessed.

Now turn to the trivializer's claim that all, or nearly all, of the properties of a thing are accidental; or else that all, or nearly all, are essential. The second idea seems to be present in Leibniz, who holds that Adam would not have existed had Peter not gone on to deny Christ some thousands of years after his death. (Even today one encounters

it in remarks such as ‘If the pistol had been equipped with a silencer, a death would have resulted, but not the same death’.) This ‘superessentialism’ is often backed by a strikingly unconvincing line of argument, namely that the identity of indiscernibles (see Identity) rules out a possible *x* differing from our actual *x* in any way. (G.E. Moore’s ‘External and Internal Relations’ (1919) contains a devastating critique.) The first idea of trivialization is far more common.

Not every property can be accidental, because it is sometimes necessary just as a matter of logic that *x* be *F*. For every *x* whatsoever, logic tells us that *x* is round if round. Since what logic tells us is necessary, ‘being round if round’ is essential to *x*. But perhaps the trivializer will try to draw the line at these ‘logically essential’ properties, counting all other properties accidental.

Such an approach will seem more trivializing than it is if one supposes that a property logically essential to one thing is thereby logically essential to everything. ‘Identity with *x*’ is the obvious counterexample; it is logically necessary that $x = x$ but not that $y = x$, so ‘identity with *x*’ comes out essential to *x* alone.

Now, though, it becomes hard not to allow additional properties as essential. Without accidental identity, for example, how can there be accidental distinctness? Apparently, then, nine, in addition to being essentially identical to nine, should be reckoned essentially distinct from seven. But to be identical to seven, another number after all, would seem to be less contrary to the nature of nine than to be an entirely different kind of thing, such as a painting or person. Once we recognize non-membership in these other kinds as essential to nine, there seems little point in refusing to count it essentially a member of its actual kind, that is, essentially a number. By this point, we have shed our former colours and taken on a modest and unassuming essentialism.

2 Grades of essential involvement

Nearly all essentialists regard at least kind-properties as essential: Aristotle is essentially human, the set of horses is essentially a set, ‘Cow’s Skull’ is essentially a painting, and so on. But it is common to go further and conceive a thing’s kind as the key to its essential properties more generally. The simplest version of this ‘priority of kinds’ doctrine maintains that (ignoring identity-properties and other such trivia) the essential properties of a *K* are, first, to be a *K*, and, second, whatever being a *K* entails. Thus while *K*s have different essences than things of other kinds, between themselves all *K*s are essentially alike.

Yet there might be reasons for allowing essential differences within a kind. Does not each set, for instance, have its specific membership essentially, and is not each painting essentially due to its actual painter? Such a view may seem at odds with the priority of kinds doctrine; but the conflict is only superficial, for we can understand a thing’s kind to dictate not the essentiality of this or that specific property, but the essentiality of its properties of such-and-such types, whatever those properties may in fact be. So, Kripke proposes that a person essentially derives from whatever gametes they in fact derive from: ‘How could a person originating from different parents, from a totally different sperm and egg, be *this very woman*?’ (1980: 113; original emphasis). Again, ‘could *this table* have been made from a completely *different* block of wood, or even of water cleverly hardened into ice…?’ (1980: 113; original emphasis). It could not, so the table essentially originated in this block of wood or one sufficiently like it.

Now that we are countenancing essential differences among conspecifics, we might wonder how deep these differences run; indeed, whether each object emerges with a uniquely identifying modal profile. By an ‘individual essence’ of *x*, let us mean a collection of properties such that

(1) necessarily, whatever has these properties is *x*, and
(2) necessarily, whatever is *x* has these properties.

The second clause asks for properties necessitated by ‘identity with *x*’, that is, properties essential to *x*. But where shall we look for properties with the further feature, demanded by the first clause, that to possess them is *sufficient* for identity with *x*? Of course, the property of ‘identity with *x*’ handles the job easily. But a property whose identity itself depends on that of *x* seems ill-suited to the task of singling *x* out. Thus our problem becomes one of finding individual essences untainted by identity-properties and similar trivia. Essentialists of an Aristotelian bent call this impossible: ‘to make clear which thing a thing is, it is not enough (*pace* the friends of logically particularized essence) to say however lengthily that it is *such*, or *so and so*’ (Wiggins 1980: 104; original

emphasis).

Opposing this pessimism are, first, the Leibnizian strategy of specifying which thing *x* is through a world-by-world catalogue of its properties and, second, Forbes' idea of identifying *x* as the unique entity of its kind originating in a certain way from certain other entities: its members if *x* is a set, its original matter if *x* is an artefact, its gametes if *x* is an organism, and so on (Forbes 1985: ch. 7).

Note that the second approach does not offer to identify objects in purely qualitative terms, but only in terms of prior objects. This is not objectionable in itself but it does leave a puzzle about items to which nothing is prior, say, coexistent eternal particles of the same kind. How will their essences differ? (Faced with a similar problem, Aquinas decided that each angel was a species unto itself.) Proponents of the first approach can say that, for each particle, there is a world where it exists all by itself. But since these worlds are not themselves distinguishable except in terms of their solitary inhabitants, this leaves us not much further ahead. Ultimately, then, the Aristotelian may have a point; at least some identity facts will have to be taken as primitive and unexplained.

3 Essential epistemology

Kant famously remarked that experience tells us how a thing is, but not that the thing could not have been otherwise. If this is true, how do we discover essential properties?

At least the outlines of an answer are set forth in Kripke's *Naming and Necessity* (1980). Suppose that the essential properties of *x* flow from its kind *K* in the following sense: *x* is essentially *F* if and only if *x* is *F* and *F* is a type of property that *K*s possess essentially if at all. Then to know *x*'s essential properties, it would be enough to know (1) its kind, (2) which types of properties things of that kind possess essentially, and (3) which properties of those types *x* actually possesses.

Can we claim access to this information? For at least many objects, it seems arguable that (2) is knowable a priori and (3) a posteriori. One knows a priori, for example, that paintings essentially derive from their actual painters, and a posteriori that Georgia O'Keeffe painted 'Cow's Skull'. As for (1) - in this case, the knowledge that 'Cow's Skull' is a painting, this involves a priori and a posteriori factors working together. Experience reveals that 'Cow's Skull' has a certain history, internal make-up and so on, and it is a priori that these features suffice to make it a painting.

4 Applications of essentialism

Apart from uses already mentioned, what is the essential/accidental distinction good for? From Descartes on, many have seen its potential as a counter to identity theses. The argument is simple. Identicals are indiscernible, so if *x* has a feature essentially that is at best accidental to *y*, then *x* and *y* are distinct. Allowing that my body is essentially a thing that takes up space and I am not, it follows that I am distinct from my body. If the tree in the quad can exist without my idea of it, then since the same cannot be said of my idea, my idea is not the tree.

Yet the argument might appear to prove too much. Imagine a statue that is always composed of the same hunk of clay, while the clay, for its part, always composes the statue. Given their overwhelming similarity, the statue and the clay might well seem identical. True, the statue is essentially humanoid in form while the clay would survive reshaping into a ball. But are we really to conclude, on the basis of such a subtle difference, that the statue and its clay are two distinct objects?

So-called 'one-thingers' suspect a fallacy of equivocation. Substituting 'the clay' for 'the statue' in 'The statue is essentially humanoid' reverses its truth-value, all right, but is this due to a change in subject matter or a change in what is said about that subject matter? Perhaps the substitution works to deflect attention from the property of having humanoid statue-counterparts to that of having humanoid clay-counterparts. That two separate properties are involved removes the need for a distinction between the objects (see Lewis 1971).

'Two-thingers' urge us to accept the distinction while rethinking its significance. If objects as similar as a statue and its clay can fail to be identical, then non-identity is not *per se* a very powerful conclusion. This makes life harder for philosophers promoting substantive forms of dualism; they must now explain what beyond mere non-identity they intend, and what they can offer as evidence beyond mere differences in essential properties. But life becomes easier for those struggling to understand the various intimate identity-like relations so important in

recent metaphysics; for instance, the relations between material objects and their constituent matter, between actions and their associated bodily movements, between mental states and the physical states that realize them, and between fine-grained events occupying the same spatiotemporal region. The last example will be developed further since it sets up one final application of essentialism, to the problem of causation.

What is the relation between the Titanic's sinking so swiftly and its sinking at all? Both events were swift but only the first, arguably, had to be so. This comes out in the fact that the Titanic's sinking might have stretched out over days or weeks (suppose that certain hatches had held) whereas its *swiftly* sinking could not have been that prolonged. Of course it is not just in this respect that the events differ. That the hatches broke was crucial to the ship's swiftly sinking, but no factor at all in its sinking as such; and we can imagine that it was the ship's sinking as such (not its sinking so swiftly) that led to the navigator's dismissal. Is this only happenstance or can we find a theory of causation capable of 'predicting' the causal differences between the events from their essential differences? A counterfactual theory looks promising since the conditions under which an event would not have occurred are visibly sensitive to its essential properties.

5 Conceptions of the essential

So far we have been understanding an essential property of *x* as a property that *x* cannot exist without. But although this is the prevailing conception of essentiality, even its advocates admit that it suffers from certain anomalies. The first and best-known concerns the property of existence. Since it is impossible to exist without existing, the modal conception extends to absolutely everything a compliment normally reserved for God, namely essential existence. The problem arises because of the way we conditionalize on existence in the definition of an essential property. Suppose then that we drop the existence condition and define the essential properties of *x* simply as the ones it must possess, regardless of whether it exists or not. This has the desired effect of eliminating essential existence for contingent beings, but at a cost: no property presupposing existence can be essential to such beings either. So, since to be human one must exist, you and I are not essentially human; hence, perhaps, not human at all.

Now to a second and deeper problem. A thing's essential properties are supposed to be the properties that make it the thing it is. But the modal conception has no way of distinguishing the properties that make *x* the thing it is from the ones it has as a necessary result of being that thing; it cannot distinguish the *conditions* of *x*'s identity from the *consequences* of its identity. This is clearest in the case of universally necessary properties such as that of 'being an element if gold', or 'being such that 2 + 3 = 5'. Neither helps make Aristotle the thing he is, but since nothing can be Aristotle without them, the modal conception reckons them essential. Now consider Aristotle's not-universally-necessary property of being distinct from the Eiffel Tower. This is not a factor in Aristotle's identity either; or, to explain what Aristotle was, we would have to mention every other object, past, present and future. Nevertheless, the modal conception calls it essential to Aristotle to be distinct from the Eiffel Tower. A final example seems decisive. To go by the modal conception, Aristotle's membership of {Aristotle} (the set whose only member is Aristotle) is essential to man and set alike. But the truth is surely different: although it lies in the nature of the set to contain the man, the man's membership of the set is not a condition of his identity but a consequence of it. No purely modal account can deliver this result; the case presents no modal asymmetries whatever, hence none for a modal account to exploit.

Troubled by these anomalies, Fine has urged a revival of the definitional conception of essence. With each object *x*, he associates a proposition $D(x)$ to function as the 'real definition' of *x*. The properties essential to *x* are those that can be assigned to it just on the basis of $D(x)$, with no help from any other source. Assuming suitable definitions, this approach allows him to resist the unwelcome essential attributions of the last few paragraphs. Aristotle's definition makes no mention of the Eiffel Tower, so it cannot pronounce on the two objects' relations. Since {Aristotle}'s definition describes it as containing Aristotle, but not the other way around, their relationship will be essential to the set only (Fine 1994).

Not every modally essential property will be definitionally essential, but there is room for debate about the converse. Suppose I make a statue out of the one hunk of clay in my studio. Then in defining the statue - in explaining what it is - I will say that it was created out of this hunk of clay. I will say this despite the fact that a distinct but sufficiently overlapping hunk would have resulted in the very same statue. (In explaining what the statue *is*, why would I mention the various objects it *could* have been fashioned from?) So, 'originating in this

hunk of clay' looks like an example of a definitionally essential property that is not modally essential. Yet it could equally be argued that since the statue did not need to originate in the given clay, originating in that clay is not a condition of identity with the statue; hence it should not be considered essential even on the definitional approach.

Disputes like this force a closer look at the phrase 'conditions of identity with *x*'. Due to a familiar ambiguity in 'condition', this can mean either the necessary *prerequisites* of identity with *x* (the properties a thing would have to have in order to be *x*) or the factors actually *constituting* its identity (the properties that *x* actually possesses by which it succeeds in being *x*). Deriving from this particular hunk of clay may not be required for identity with the statue, but it seems still to be essential in the constitutive sense. To have derived from the given clay is part of what it is, even if not part of what it had to be, to be that statue.

So we end up with three conceptions of essential property, depending on whether *x*'s essential properties are understood as (1) the necessary prerequisites of identity with *x*, (2) the factors actually constituting *x*'s identity or (3) the necessary consequences of being *x*. (3) expresses the modal conception of essentiality. (2) is a fully de-modalized version of the definitional conception. (1) lies somewhere between; it is the definitional conception but with a modal twist. Each of the three seems worthy of further study.

See also: Aristotle §8; Definition; Locke, J. §5; Identity of indiscernibles §2; Logical and mathematical terms, glossary of;Natural kinds §3; Substance §1

STEPHEN YABLO

References and further reading

Adams, R. (1979) 'Primitive Thisness and Primitive Identity', *Journal of Philosophy* 76: 5-26.(Defends primitive identity against a Leibnizian position on individual essence; see §2.)

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press, 2nd edn, 1956.(Attempts to explain *de re* necessity in terms of meaning relations.)

Fine, K. (1994) 'Essence and Modality', *Philosophical Perspectives* 8: 1-16.(Basis for much of §§1, 5. Criticizes modal conception of essence and revives definitional conception.)

Forbes, G. (1985) *The Metaphysics of Modality*, Oxford: Clarendon Press.(Cited in §2 as defending individual essences. Excellent overall treatise on *de re* modality.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.(Cited in §§2-3. Extremely important. Argues for essentiality of kinds and origins, and sketches a *de re* modal epistemology.)

Lewis, D.K. (1971) 'Counterparts of Persons and Their Bodies', *Journal of Philosophy* 68: 203-11.(Defends 'one-thingism' using counterpart theory. See §4.)

Moore, G.E. (1919) 'External and Internal Relations', repr. in *Philosophical Studies*, London: Routledge & Kegan Paul, 1922.(Cited in §2. Early statement of the modal conception of essence.)

Quine, W.V. (1954) 'Reference and Modality', repr. in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1980.(Classic presentation of sceptical anti-essentialism. See §1.)

Sidelle, A. (1989) *Necessity, Essence, and Individuation*, Ithaca, NY: Cornell University Press.(Defends a sophisticated conventionalism about essential properties.)

Wiggins, D. (1980) *Sameness and Substance*, Cambridge, MA: Harvard University Press.(Cited in §2. Stresses essentiality of kinds and connections with trans-temporal identity.)

Yablo, S. (1987) 'Identity, Essence, & Indiscernibility', *Journal of Philosophy* 84: 293-314.(Expands on §§2, 4.)

Eternity

The distinctive, philosophically interesting concept of eternity arose very early in the history of philosophy as the concept of a mode of existence that was not only beginningless and endless but also essentially different from time. It was introduced into early Greek philosophy as the mode of existence required for fundamental reality (being) contrasted with ordinary appearance (becoming). But the concept was given its classic formulation by Boethius, who thought of eternity as God's mode of existence and defined God's eternality as 'the complete possession all at once of illimitable life'. As defined by Boethius the concept was important in medieval philosophy. The elements of the Boethian definition are life, illimitability (and hence duration), and absence of succession (or timelessness). Defined in this way, eternality is proper to an entity identifiable as a mind or a person (and in just that sense living) but existing beginninglessly, endlessly and timelessly.

Such a concept raises obvious difficulties. Some philosophers think the difficulties can be resolved, but others think that in the light of such difficulties the concept must be modified or simply rejected as incoherent. The most obvious difficulty has to do with the combination of atemporality and duration.

Special objections have arisen in connection with ascribing eternality to God. Some people have thought that an eternal being could not do anything at all, especially not in the temporal world. But the notion of an atemporal person's acting is not incoherent. Such acts as knowing necessary truths or willing that a world exist for a certain length of time are acts that themselves take no time and require no temporal location. An eternal God could engage in acts of cognition and of volition and could even do things that might seem to require a temporal location, such as answering a prayer.

The concept of God's eternality is relevant to several issues in philosophy of religion, including the apparent irreconcilability of divine omniscience with divine immutability and with human freedom.

1 The concept

Eternality - the condition of having eternity as one's mode of existence - has been understood in more than one way. Sometimes 'eternal' has been associated with endless temporal existence (as in 'eternal life') or with beginningless temporal existence (as in the medieval debate over whether the world was eternal). In these senses the concept of eternality presents no distinctive philosophical difficulties. But there is another sense of 'eternality' in which the concept has been an issue in philosophy from the Greeks to the present. The concept was given its classic formulation in this sense by Boethius (§5), who defined it as 'the complete possession all at once of illimitable life' (*The Consolation of Philosophy*, bk V, pr. 6).

Although the interpretation of the definition becomes controversial in its details, it pretty clearly identifies four elements of eternality. First, anything eternal in the Boethian sense has life. In this sense, then, eternality could not characterize numbers, truth or the world. Next, the life of whatever is eternal is illimitable - necessarily beginningless and endless. Sometimes this element has been interpreted as attributing to whatever is eternal a mode of existence that is illimitable in virtue of being absolutely unextended (like a single instant) and only in that way without a beginning or an end. But a more natural reading and one more consonant with other things Boethius and his successors say about eternity is that it is illimitable in virtue of its infinite duration. Duration is thus the (implicit) third element of eternality in Boethius' formulation. The fourth and last is conveyed by the phrase 'complete possession all at once'. Although living temporal persons may be said to possess their life, they do not possess it completely all at once because they live out their life successively. Past parts of their life they possess no longer, future parts not yet. Consequently, whatever is eternal is also not in time. Eternality thus combines atemporality and duration. (This apparent incoherence is discussed in §2 below.)

Eternity, then, is a real, atemporal mode of existence characterized by both the absence of succession and limitless duration. Nothing in that concept denies the reality of time or implies that temporal experiences are illusory. Boethius and others who use the concept suppose that reality includes time and eternity as two distinct modes of real existence, neither of them reducible to or incompatible with the other.

Temporal events are instructively ordered in terms of the A-series - past, present and future - and the B-series - earlier than, simultaneous with, later than (see McTaggart, J.M.E.). Because an eternal entity is atemporal, its life

cannot be ordered successively in either of those series. Moreover, no temporal entity or event can be past or future with respect to, or earlier or later than, the whole life of an eternal entity, because otherwise the eternal entity would itself be part of a temporal series. But nothing in eternality's absence of successiveness entails that it cannot be characterized by *presentness*, or that an eternal entity's cognitive or causal relationship with temporal entities and events cannot be a kind of *simultaneity*. Of course, the presentness and simultaneity associated with an eternal entity could not be temporal presentness or temporal simultaneity. Taking the concept of eternality seriously involves recognizing that it introduces technical senses for several familiar words, including 'now', 'present', and 'simultaneous with', as well as the present-tense forms of many verbs.

In order to allow for real relationships between what is eternal and what is temporal it is particularly important to establish a special sense of 'simultaneous'. A relationship that can be recognized as a kind of simultaneity will of course be symmetric; but, since its relata have relevantly distinct modes of existence, it will be neither reflexive nor transitive. In this sense of 'simultaneous', each of two temporal events can be simultaneous - co-occurrent - with one and the same eternal event without being in any sense simultaneous with each other. This special sort of simultaneity has been called 'ET-simultaneity' (for 'simultaneity between what is eternal and what is temporal'). From a temporal standpoint, the temporal present is ET-simultaneous with the whole infinite extent of an eternal entity's life. From an eternal standpoint, every time is present to or co-occurrent with the whole infinite atemporal duration; that is, each instant of time as it is actually present temporally is ET-simultaneous with the one enduring present of an eternal entity, so that for an eternal entity all of time is present at once.

2 Difficulties associated with the concept

Many of the difficulties in the concept of eternality have been discussed in twentieth-century philosophical literature. The most obvious difficulty arises from the combination of atemporality with duration - the heart of the concept. Ordinarily, 'duration' means persistence through time, and the incoherence of atemporally persisting through time needn't be argued. But the philosophers who developed the concept of eternality were using ordinary terms in extraordinary ways in order to express their theoretical notion of an illimitable life possessed completely all at once. Of course, language is strained when it is stretched to accommodate things utterly outside the ordinary experience language is founded on, as in the black holes and the Big Bang of twentieth-century cosmology (see Cosmology §3). Serious attempts to show that eternity really is an incoherent concept require showing that the apparent incoherence persists when the technical interpretations of its terms are fully taken into account.

One attempt at doing just that involves taking atemporal duration as a species of extension and then arguing that any extension must be divisible and so cannot be all at once, or atemporal. The problem with such an attempt is that it uses an inductive survey of temporal and spatial extensions to reach the generalization that all extensions are divisible. But since eternity is neither temporal nor spatial, it will not be surprising if the properties attributed to extension as a result of such an induction fail to apply to what is eternal.

Still, this way of avoiding the ascription of divisibility to what is eternal seems to run into an old problem: insisting that terms cannot be used univocally of temporal and of eternal things looks like introducing equivocation into the description of eternality. Not only does 'duration' not have its ordinary sense when used of the eternal, but it is also difficult to say precisely what its extraordinary sense is. This sort of problem has become familiar in connection with discourse about God. It has often been pointed out that to use ordinary terms univocally of God and creatures is to deny the transcendence of God, but to use them equivocally masks a radical agnosticism about God's nature and activity. So discourse about God can use neither univocal nor equivocal predication (see Religious language §4).

Analogical predication is the traditionally recognized solution to this dilemma, and it is also what is needed for interpreting the description of the eternal. Atemporal duration is *analogous* to temporal duration, enough like temporal duration to make using the term 'duration' helpful, but enough unlike it to mean that the definition of '(temporal) duration' will not apply. Eternal duration is fully actualized duration, none of which is already lost or not yet gained: beginningless, endless, non-successive existence possessed completely, all at once, present entirely to its possessor - a mode of existence consisting entirely in a present that is infinite rather than instantaneous. Timeless duration might well be thought of as Plato thought of it, as the genuine duration of which temporal duration is only the moving image. Not all critics of the concept of eternity find such a response adequate, and the most persistent objections to the concept concentrate on the difficulties of ascribing duration to what is atemporal.

Because eternality in the sense at issue here is taken primarily to characterize God's mode of existence, other objections to the concept stem from combining it with traditional concepts of God. In this vein philosophers have objected that a God who is eternal could not act at all, and especially not in time. But this objection is based on a confusion. Of course, there are things an atemporal God could not do - such as remembering, or planning ahead. But not all cognitive and volitional acts require temporal location, and God could engage in those that do not. Furthermore, an atemporal God could not change the past or foreknow the future. Such actions, if possible at all, would require a temporal location, without which there can be neither past nor future. Still, eternal God, present at once to each temporal instant in its temporal presentness, could in the eternal present directly affect events that are past with respect to us and be directly aware of events future with respect to us. He could also act in time. He could, for example, will timelessly that something occur or come into existence at a particular time. By the same token, he could also do things that might appear to require a particular temporal location, such as answering a particular prayer. Because both the time of the prayer and the time at which the answer to it occurs are ET-simultaneous with the whole of eternity, an eternal God could be aware, timelessly, of a prayer prayed at one time, willing (ET-simultaneously) that the answer to that prayer occur at a later time.

Finally, some critics suppose that if God is eternal and creatures are temporal, then God could not be directly aware of creatures or interact with them directly and immediately as he is traditionally said to do. Such criticisms must presuppose that for one being to interact directly with another, the two must share a mode of existence. But traditional theists are already committed to rejecting this presupposition as regards space. God is traditionally described as non-spatial and thus as not sharing with creatures a spatial mode of existence, and yet that difference in modes of existence is generally thought to be no obstacle to God's being directly aware of or directly interacting with his creatures. If the presupposition is false as regards space, however, it is hard to see why it should be accepted as regards time.

3 The history of the concept

The earliest indisputable appearance of the concept of eternity is in Plato's *Timaeus*. Parmenides' description of the mode of existence of Being, or the One, in his *Way of Truth* is much older, but scholars disagree over whether Parmenides intended to ascribe atemporality to Being (see Parmenides §3). Whatever Parmenides meant, what Plato says about eternity is in several respects just what Parmenides says about Being's mode of existence, and to that extent at least Parmenides may be thought of as the inventor or discoverer of the concept of eternity.

Many scholars believe that Aristotle rejected Plato's notion of eternity, although there is also some textual evidence suggesting that, on the contrary, he accepted and made use of it in describing the life of the Prime Mover. Whether or not Aristotle himself accepted the concept of eternity, it is indisputable that the concept came into medieval philosophy through the Platonic rather than the Aristotelian tradition. Plotinus (§4), for instance, has a well-developed concept of eternity, and in *Enneads* III 7 he stresses the importance of duration in the concept. Boethius seems to have based his definition of eternity on the one Plotinus develops.

Augustine (§8), who was even more clearly in the Platonic tradition than Boethius was, understood and accepted the concept of eternity, which plays a significant part in two of his most important works, the *Confessions* (bk XI, ch. 11) and *The City of God* (bk XI, ch. 21). Like Boethius' formulations, Augustine's views of eternity were an important influence on later medieval philosophy.

In the Middle Ages, the concept of eternity was widely used and discussed. It can be found, for example, in Anselm's *Monologion* (ch. 24) and *Proslogion* (ch. 13), where it seems taken for granted, as a standard part of traditional theology. But it received its most sophisticated development in the work of Aquinas, who discussed and employed it in several of his works (for example, *Summa theologiae* Ia, q.10). After Aquinas, although many medieval philosophers and theologians continued to hold that God is eternal, they did not always mean by 'eternal' what Boethius (and Aquinas) had meant by it. Duns Scotus, for example, seems to have held that God's eternality is not co-occurrent with all of time.

In the modern period, with the rejection of the medieval synthesis in theology, the notion of eternity, in the special sense at issue here, was largely abandoned. Hobbes is still aware of it in the Boethian sense '*as a permanent now*' (1680: 435), but Locke, for example, takes eternity to be just an infinity of temporal duration: 'By being able to repeat any such *Idea* of any length of time… and add them one to another, without ever coming to the end of such

addition… we come by the *Idea* of *Eternity*' (1689, bk II, ch. 14).

4 Applications of the concept to issues in the philosophy of religion

Applying the concept of eternality makes a significant difference in considering various issues in philosophy of religion. Here we will concentrate on just two of the most important: omniscience and immutability, and foreknowledge and free will.

It has been argued that omniscience and immutability, two traditional divine attributes, are not compossible. An omniscient knower always knows what time it is (or precisely what is going on) now, and any knower who always knows what time it is now is a knower whose knowledge is always changing. Consequently, a knower could be omniscient and mutable, or immutable and not omniscient, but there could not be an omniscient, immutable knower (see Immutability).

This argument presupposes that knowers are temporal. If omniscient God is eternal, however, the argument becomes more complicated. For example, in the claim that an omniscient knower always knows what time it is now, 'now' and the present tense of 'knows' can each be read as indicating either the temporal or the eternal present, thus allowing for four different interpretations of the claim. Not all of those interpretations make sense; for example, 'An omniscient knower always knows in the eternal present what time it is in the eternal present' incoherently attributes time to the eternal present. The most reasonable interpretation of the claim, on the supposition that the omniscient knower in question is eternal, is that the knower always knows in the eternal present what time it is in the temporal present. But on that interpretation it is much more difficult to show that such a knower could not be immutable.

Even if we suppose that the indexical 'now' is ineliminable, that there is an absolute temporal present as distinct from a present that is merely relative to some particular temporal entity, it is not clear that an eternal God could not know what time it is without constantly changing. On the view that the whole of eternity is ET-simultaneous with each temporal event as it is actually happening, an eternal omniscient knower will know all the events actually occurring at a particular time as well as the temporal location of that time and its being experienced as present by temporal entities at that time. Such a knower will also know that from the standpoint of eternity every temporal event is actually happening. There is nothing further for an eternal entity to know about what time it is now, for either the eternal or the temporal now; and nothing in what it does know requires constant change or change of any sort. Thus, while the argument may show that no temporal knower can be both omniscient and immutable, it does not make its case if the omniscient, immutable knower is eternal.

Arguments that knowledge of future contingent events is irreconcilable with human freedom depend on the notion of *fore*knowledge, on someone's knowing ahead of time what someone else will 'freely' decide to do (see Omniscience §§3-4). On the face of it, then, the concept of eternity provides a solution to the problem of foreknowledge and free will, as Boethius maintained in introducing the concept. An eternal omniscient knower will be eternally aware of *all* contingent events as they are occurring, including those that occur in the temporal future, but he will not *fore*know them, since nothing eternal can be earlier than anything else. Consequently, arguments purporting to show that foreknowledge and free will are incompatible will not apply to eternal omniscient knowledge, which is evidently compatible with human free will.

But some philosophers have thought that eternality nonetheless fails to provide a solution to the problem of divine knowledge and human freedom, because the fixity and infallibility of divine knowledge seem enough by themselves to make God's knowledge of future contingents incompatible with free will. 'God knows in the eternal present that Paula mows her lawn in 2095' entails that Paula mows her lawn in 2095, and so God's eternal awareness of a future event seems to have the result that the event is inevitable *now*, before the event occurs, in a way incompatible with Paula's freedom of action. Consequently, the concept of eternity seems after all unhelpful for resolving the apparent incompatibility between divine knowledge and human freedom.

The idea of this line of argument is that a proposition such as 'God eternally knows that *p*' (where *p* is of the form 'Paula mows her lawn (at some date future with respect to us in the present)') entails 'It is now the case that *p*'. That is why eternal knowledge is supposed to have the result that the future is somehow fixed and inevitable now. But is there such an entailment? 'God eternally knows that *p*' does entail *p*, and *p* does seem equivalent to 'It is now the case that *p*'. But is it equivalent in a context involving eternality? In that context, God knows that *p* in

virtue of being ET-simultaneous with the future events he is eternally aware of. In that context, furthermore, it is also true that God's knowledge that *p* is ET-simultaneous with the temporal now. That is why we can appropriately say such things as 'It is now (in the temporal present) true that God eternally knows that *p*'. But, as has already been pointed out, ET-simultaneity is not transitive. From the facts that the state of affairs that *p* is ET-simultaneous with eternity and that eternity is ET-simultaneous with the temporal present, it does not follow that it is *now* the case that *p*. Therefore, while it is true that God's knowing that *p* entails that *p*, the relationship between time and eternity is such that God's knowing that *p* does not entail that it is *now* the case that *p*. Hence, it seems that the concept of eternity can constitute the basis for an adequate solution to the problem of foreknowledge and free will.

See also: God, concepts of; Necessary being; Simplicity, divine

ELEONORE STUMP
NORMAN KRETZMANN

References and further reading

Adams, M. (1987) *William Ockham*, Notre Dame, IN: University of Notre Dame Press, 1115-48.(Contains a helpful explanation of Duns Scotus' and Ockham's views on God's knowledge of future contingents and of how their views are related to Aquinas'.)

Anselm (1076) *Monologion*, in J. Hopkins and H. Richardson (trans and eds) *Anselm of Canterbury*, vol. 1, Toronto, Ont. and New York: Edwin Mellen Press, 1974.(A classic of natural theology.)

Anselm (1077-8) *Proslogion*, in J. Hopkins and H. Richardson (trans and eds) *Anselm of Canterbury*, vol. 1, Toronto, Ont. and New York: Edwin Mellen Press, 1974.(Anselm supposes that God's eternality can be proven from the basic idea of what God is.)

Aquinas, T. (1266-73) *Summa theologiae* Ia, q.10, in P. Caramello (ed.) *S.Thomae Aquinatis Summa theologiae*, Turin: Marietti, 1952.(The classic scholastic formulation of the concept of eternity.)

Augustine (397-401) *Confessions*, trans. W. Watts, Cambridge, MA: Harvard University Press, and London: Heinemann, 1968-70, vol. 2, bk XI, ch. 11. (Augustine's story of his conversion to Christianity, containing much important work in metaphysics and ethics.)

Augustine (413-27) *The City of God*, trans. D.S. Wiesen, Cambridge, MA: Harvard University Press, and London: Heinemann, 1968, vol. 3, bk XI, ch. 21.(Contains presentations and applications of the concept that were influential in later medieval philosophy.)

Boethius (521-4) *De trinitate*, in H.F. Stewart, E.K. Rand and S.J. Tester (eds) *Boethius: The Theological Tractates and The Consolation of Philosophy*, London: Heinemann, and Cambridge, MA: Harvard University Press, 1973.(A classic attempt to resolve philosophical problems in the concept of the Trinity.)

Boethius (525-6) *The Consolation of Philosophy*, in H.F. Stewart, E.K. Rand and S.J. Tester (eds) *Boethius: The Theological Tractates and The Consolation of Philosophy*, London: Heinemann, and Cambridge, MA: Harvard University Press, 1973.(Boethius' presentation and applications of the concept became the *locus classicus* for later medieval discussion.)

Fitzgerald, P. (1985) 'Stump and Kretzmann on Time and Eternity', *Journal of Philosophy* 82: 260-9.(Argues that the notion of atemporal duration is incoherent.)

Hasker, W. (1989) *God, Time, and Knowledge*, Ithaca, NY, and London: Cornell University Press.(Contains a careful examination of the concept of eternity, finally rejecting it.)

Hobbes, T. (1680) *Considerations upon the Reputation, Loyalty, Manners and Religion of Thomas Hobbes of Malmsbury, written by himself by way of Letter to a Learned Person*, London: Crooke.(Referred to in §3.)

Leftow, B. (1991) *Time and Eternity*, Ithaca, NY, and London: Cornell University Press.(An excellent discussion of all aspects of the concept of eternity, with some philosophically interesting modifications of it.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P. Nidditch, Oxford: Clarendon Press, 1975. (Referred to in §3.)

Nelson, H. (1987) 'Time(s), Eternity, and Duration', *International Journal for Philosophy of Religion* 22: 3-19. (An argument that duration cannot be part of a coherent concept of eternity.)

Pike, N. (1970) *God and Timelessness*, London: Routledge & Kegan Paul.(A helpful overview of eternality in philosophy of religion, raising many of the classic objections.)

Plantinga, A. (1986) 'On Ockham's Way Out', *Faith and Philosophy* 3: 235-69; reprinted in T. Morris (ed.) *The Concept of God*, Oxford: Oxford University Press, 1987, 171-200.(Argues that the concept of eternity does not constitute an adequate solution to the problem of foreknowledge and free will.)

Plato (*c*.366-360 BC) *Timaeus*, trans. R.G. Bury, Cambridge, MA: Harvard University Press, and London: Heinemann, 1961, 37D-38C.(Plato's classic exposition of the concept of eternity.)

Plotinus (*c.*250-66) *Enneads*, trans. A.H. Armstrong, London: Heinemann, and Cambridge, MA: Harvard University Press, 1967, vol. 3, III 7.(Plotinus' definition of the concept of eternity and his argument that duration is part of the concept.)

Stump, E. and Kretzmann, N. (1981) 'Eternity', *Journal of Philosophy* 78: 429-57; repr. in T. Morris (ed.) *The Concept of God*, Oxford: Oxford University Press, 1987, 219-52.(An exposition and defence of Boethius' concept of eternity.)

Stump, E. and Kretzmann, N. (1991) 'Prophecy, Past Truth, and Eternity', in J.E. Tomberlin (ed.) *Philosophical Perspectives (5): Philosophy of Religion*, Atascadero, CA: Ridgeview Publishing Company.(Argues that the concept of eternity provides an adequate solution to the problem of foreknowledge and free will.)

Stump, E. and Kretzmann, N. (1992) 'Eternity, Awareness, and Action', *Faith and Philosophy* 9: 463-82. (Argues that duration is a coherent part of the concept of eternity.)

Wierenga, E. (1989) *The Nature of God: An Inquiry into Divine Attributes*, Ithaca, NY, and London: Cornell University Press.(Includes detailed arguments against the concept of eternity.)

Zagzebski, L. (1991) *The Dilemma of Freedom and Foreknowledge*, New York: Oxford University Press. (Includes a careful examination of the concept of eternity and of objections to it.)

Eternity of the world, medieval views of

The problem of the eternity of the world was much debated in Western philosophy from the twelfth through the fourteenth centuries, but its history goes back as far as Philo of Alexandria and the Church Fathers. The principal topic of controversy was the possibility of a beginningless and yet created world. The arguments that fashioned the medieval discussion rested upon assumptions concerning the concepts of eternity and creation. In addition, the issue of eternity intertwined with discussions of the relationship of God to creation, with proofs of the existence of God, with the nature of the material universe and with the nature of infinity. Some of the most ingenious ideas in these debates were obtained from pagan Greek, Islamic and Jewish traditions.

1 Early stages

According to Judaeo-Christian tradition, based in particular upon the opening words of Genesis ('In the beginning God created the heaven and the earth'), the universe had a beginning. From the fourth century onwards, however, Christian thinkers had to take into consideration accounts from antiquity, transmitted by authors such as Augustine and Boethius, according to which the existence of the material universe (*mundus*) was 'from eternity' (*ab aeterno*), that is, beginningless.

In its early stages, the medieval discussion of the eternity of the world was preoccupied with two types of questions, one asking whether the world had existed from eternity and the other examining the concept of eternity and its relation with time. In general, 'time' (*tempus*) was understood to imply having both a beginning and an end, and perpetuity (*aevum*) involved a beginning but not an end, whereas 'eternity' (*aeternitas*) had neither beginning nor end. 'Eternity' in this sense was understood as a temporal notion, meaning 'infinite temporal extension'. However as Boethius pointed out in *De consolatione philosophiae (The Consolation of Philosophy)*, 'eternity' could also be taken as atemporal. This conception of eternity, sometimes called 'eternity proper' by later authors, introduced the notion of timelessness within the notions of 'all at once' (*tota simul*) and of 'life', and referred to God's mode of existence.

Medieval thinkers all agreed that the universe was not coeternal with God. They were divided, however, over the question whether the universe had always existed in a temporal sense, and over how this question should be understood in the first place. Using a distinction derived from Augustine in *De civitate Dei (The City of God)* and *Confessiones (Confessions)*, many medieval thinkers would observe that the world was created *together with* time, and in this sense had *always* existed: there was no past at the time of creation.

In the early thirteenth century, the discussion of the eternity of the world was raised to a higher level. Three events contributed to this. First, the Fourth Lateran Council (1215) declared the temporal beginning of the world to be an article of faith. Second, around that same time the *Sentences* by Peter Lombard, which included a statement (Book 2, distinction 1) about the world's eternity, became an official textbook in the faculty of theology. Third, the translation into Latin of Aristotle's books on natural philosophy (*Libri naturales*), especially *Physics* VIII, brought medieval thinkers in the arts faculty into routine contact with arguments proving the eternity of motion and generation (see Aristotle). Aristotle's views on eternity, especially the question whether he had intended to prove the beginninglesness of the world, became an issue of debate that culminated in the condemnations of 1270 and 1277 by Bishop Etienne Tempier (see Aristotelianism, medieval). Aristotle's views were perceived through the interpretations offered by Averroes' *Commentary on Aristotle's Physics* and Maimonides' *Guide to the Perplexed*, works that had also been translated recently into Latin (see Averroism; Ibn Rushd; Maimonides, M.).

The most important result of these events was that the issue to be debated came into sharper focus: could the world have been eternal, if God had so willed? In other words, must the beginning of the world be accepted as an article of faith, or can it also be proved by (demonstrative) arguments? In this way, the issue of the eternity of the world helped to determine the position philosophers held with regard to the relation between faith and reason.

2 The two sides

The various conceivable ways of responding to the question of the eternity of the world can be reduced to the following two positions. One group of thinkers, of whom Bonaventure is the archetype, claimed that it could be demonstrated that the world began to exist. Other adherents of this position were Matthew of Aquasparta, Henry of

Ghent, John Pecham, Richard of Middleton, Peter Aureol, William of Alnwick, Henry Totting of Oyta and Marsilius of Inghen. The demonstrations hinged on two basic assumptions concerning creation and eternity.

The Christian view had come to be that God had created the world where before there had been nothing: in other words, creation from nothing. In an argument that goes back to Richard of St Victor, it was maintained that since creation from nothing (*ex nihilo*) is a transition from non-being (*nihil*) to being, non-being necessarily precedes being. Hence, creation necessarily implies a beginning in time. In sum, creation 'from nothing' was understood as creation 'after nothing' (*post nihil*).

Furthermore, an eternal world was conceived to imply the existence of an infinite series of past events. The contradictions that allegedly arose from this assumption were considered reasons why an eternal world is impossible. Most of the arguments were drawn from Aristotle's widely accepted theory of the infinite, but were here employed in a new context to substantiate the un-Aristotelian conclusion that the world is not eternal but had a beginning. In particular, the following generally accepted Aristotelian rules were considered to be violated by the idea of a beginningless world: first, that it is impossible to add to the infinite; second, that it is impossible to traverse what is infinite; third, that it is impossible for the infinite to be grasped by a finite power; and fourth, that it is impossible that there be simultaneously an infinite number of things. In addition, the theory of a possible eternal world seemed to clash with self-evident principles such as that the whole is greater than the part (John Pecham) or that, since the infinite is *infinite*, one infinity cannot be greater than another, or that there is no order in the infinite because there is no first element (Bonaventure). Some of these alleged contradictions had already been pointed out by John Philoponus in his *De aeternitate mundi contra Proclum (On the Eternity of the World Against Proclus)* and had circulated among Arabic authors such as Algazel in his *Metaphysica* before they were transmitted to the Latin West (see al-Ghazali; Islamic philosophy: transmission into Western Europe).

Another group of thinkers, including Siger of Brabant, Boethius of Dacia, Thomas Aquinas, Giles of Rome, Godfrey of Fontaines, William of Ockham, Henry of Harclay, Thomas of Wylton and Thomas of Strasbourg, argued that the universe could have existed from eternity. In part, their argumentation rested on the rebuttal of those arguments that the proponents of a demonstrable beginning had invoked to refute the possibility of an eternal world. They concluded that since the beginning of the world could not be *demonstratively* proved, the universe could have existed without beginning to exist.

Proponents of the possibility of a beginningless universe interpreted creation out of nothing as creation not out of anything, that is, not out of any independently existing matter. Creation was understood as a relation of causal dependence of creatures upon God, and in this interpretation the status of being a creature was not necessarily inconsistent with being beginningless. The argument that the universe depends for its existence upon a superior principle that is not prior in time but prior in the order of things can be found in Avicenna's *Metaphysica*, and was at the heart of Aquinas' rebuttal of Bonaventure's interpretation of creation from nothing (see Aquinas, T. §9; Ibn Sina §4).

The infinity arguments in favour of a beginning were considered off the mark by the proponents of the possibility of an eternal world. With regard to the traversal argument, for instance (that is, the argument that if the world had always been, an infinite past time would have been traversed), they emphasized that the traversed infinite time was a successive and not a simultaneous infinity, and hence, that the contradictions that seemed to follow from the premise of an eternal world were not pertinent. There will always be only finitely many past days between this and any past day. A beginningless world does not imply that any past day was infinitely remote from this day.

In general, the adherents of a possible beginningless world held the same Aristotelian views on the infinite as their opponents. They disagreed only over the kind of infinity that was involved in a possible eternal world. Henry of Harclay was an interesting exception. He agreed with his opponents that a possible eternal world would entail an *actual* infinity, but he denied that the conclusions that followed from this stance were contradictory. Harclay argued that the infinite can be traversed, that it can be exceeded, and that not all infinites are equal (see Infinity).

See also: Aristotelianism, medieval; Creation and conservation, religious doctrine of; Eternity; Natural philosophy, medieval; Richard Rufus of Cornwall; Time

J.M.M.H. THIJSSEN

References and further reading

Boethius of Dacia (*c*.1275) *De aeternitate mundi (On the Eternity of the World)*, trans J.F. Wippel, *On the Supreme Good, On the Eternity of the World, On Dreams*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1987.(An important text from the faculty of arts in the early 1270s.)

Brown, S.F. (1991) 'The Eternity of the World Discussion at Early Oxford', in A. Zimmermann and A. Speer (eds) *Mensch und Natur im Mittelalter (Man and Nature in the Middle Ages),* Berlin: de Gruyter, vol. 1, 259-80.(An examination of how the conception of theology influenced the discussion of the eternity of the world, with an edition of Richard Rufus of Cornwall II *Sent*. d.1 q.1.)

Bukowski, T. (1991) 'Understanding St. Thomas on the Eternity of the World: Help from Giles of Rome?', *Recherches de théologie ancienne et médiévale* 58: 113-25.(A contribution to the controversy over the interpretation of Thomas Aquinas' views, which at the same time clarifies some intricacies of the medieval debate concerning the eternity of the world in general.)

Dales, R.C. (1990) *Medieval Discussions of the Eternity of the World*, Leiden: Brill.(Assembles and discusses a wealth of source material and provides an extensive bibliographical guide.)

Dales, R.C. and Argerami, O. (1991) *Medieval Latin Texts on the Eternity of the World*, Leiden: Brill.(A dossier of medieval Latin texts extending from the 1220s to the second quarter of the fourteenth century, accompanied by brief introductions.)

Davidson, H.A. (1987) *Proofs for Eternity, Creation and the Existence of God in Medieval Islamic and Jewish Philosophy*, Oxford: Oxford University Press.(A brilliant typology of the medieval Islamic and Jewish arguments for eternity, many of which recur in the Latin West.)

Hoenen, M.J.F.M. (1992) 'The Eternity of the World according to Marsilius of Inghen with an Edition of the Dubium in II *Sent*. q.1 a.2', in H.A.G. Braakhuis and M.J.F.M. Hoenen (eds) *Marsilius of Inghen*, Acts of the International Marsilius of Inghen Symposium, Nijmegen: Ingenium, 117-43.(A presentation of the views of the late-fourteenth-century authors Thomas of Strasbourg, Henry Totting of Oyta, Marsilius of Inghen and Conrad of Soltau, whose discussions of the eternity of the world rest upon assumptions concerning the nature of God as an immeasurable spiritual quantity.)

John Pecham (1269-73) *Quaestiones de aeternitate mundi (Questions on the Eternity of the World)*, trans. V.G. Potter, *Questions concerning the Eternity of the World*, New York: Fordham University Press, 1993. (Translation of a key text, reflecting the state of the debate in the 1270s.)

Kretzmann, N. (1985) 'Ockham and the Creation of the Beginningless World', *Franciscan Studies* 45: 1-31.(A philosophical analysis of Ockham's views on the eternity of the world and an extremely useful general introduction to the entire debate.)

MacIntosh, J.J. (1994) 'St Thomas and the Traversal of the Infinite', *American Catholic Philosophical Quarterly* 68: 157-77.(A consideration of the so-called traversal-of-the-infinite argument in its ancient, medieval and early modern versions, and a contemporary assessment of Aquinas' response to it.)

Sorabji, R. (1983) *Time, Creation and the Continuum: Theories in Antiquity and the Early Middle Ages*, London: Duckworth.(A philosophical analysis of discussions of eternity and creation in Greek philosophy. Very important for understanding the medieval debates.)

Stump, E. and Kretzmann, N. (1981) 'Eternity', *Journal of Philosophy* 78: 429-70.(The most profound analysis to date of Boethius' atemporal notion of eternity.)

Vollert, C., Kendzierski, L.H. and Byrne, P.M. (eds) (1964) *On the Eternity of the World*, Milwaukee, WI: Marquette University Press.(A translation (by the editors) of three important texts under this title by Thomas Aquinas, Siger of Brabant and Bonaventure, which provide a useful catalogue of the arguments that played a role in the eternity debate.)

Wissink, J.B.M. (1990) *The Eternity of the World in the Thought of Thomas Aquinas and His Contemporaries*, Leiden: Brill.(Contains contributions by M.J.F.M. Hoenen about the discussion of the world's eternity in the *Correctoria* literature of the Dominican and Franciscan schools, and by J.M.M.H. Thijssen about the reception of Aquinas' views concerning infinity in the works of the fourteenth-century thinkers Harclay, Alnwick and Wylton.)

Ethical systems, African

Ethical thought in sub-Saharan Africa grows largely out of traditions that are communalistic, not based in individual consent, anti-universalizing, naturalistic, and humanist. Within such thought, the general vocabulary of evaluation, like such English words as 'good' and 'bad', does not strongly differentiate between narrow moral assessments, on the one hand, and technical or aesthetic evaluation on the other. This is true of places where Islam has been present for many centuries. The substantial exposure, in the colonial and postcolonial periods, to European moral ideas (both through various forms of Christian missionary evangelism and as a result of contact with secular moral and political traditions from elsewhere), along with the changes produced by the modern economy, have produced a wide range of ethical ideas. However, the residue of precolonial ethical theory remains the most distinctive, if not always the most important, component and is the main topic of this entry.

1 Defining the subject matter

Sub-Saharan Africa has a wide range of varied cultures. A standard audit (Greenberg 1970: 25) shows that more than one thousand languages are spoken on the continent. It is made up of more than fifty modern states which include peoples descended from hundreds of earlier societies, ranging, by degrees of political organization, from the hunter-gathering King Kung of central southern Africa, who were without a ruler, through to the hierarchical states of Buganda in the east and the Yoruba and Asante in the west, to the Islamic kingdoms, emirates and sultanates of the Sahel and northwest Africa.

Most contemporary sub-Saharan African cultures have grown out of interactions between the so-called 'traditional' systems of thought of these earlier societies, on the one hand, and Christianity, Islam and the many influences of the extensive colonial administrations of modern European and US cultures, on the other. Such colonial rule, which spanned more than a century, continued until the 1970s. Ethical theory in the Maghreb (the largely Arabic-speaking region of north Africa, which has been part of the literate Islamic world for more than a millennium), deserves separate consideration in the broader context (see Ethics in Islamic philosophy; Islamic philosophy). For similar reasons, Ethiopia's literate philosophical traditions are best considered separately (see Ethiopia, philosophy in). Contemporary moral philosophy in universities is largely an extension of ethics in the colonial metropolitan cultures, more particularly of the UK and France (see African philosophy, Anglophone; African philosophy, Francophone). The ethical traditions of white settler populations in southern Africa are continuous with those of their European ancestors. This entry deals generally with those sub-Saharan African traditions that remain.

2 Traditional cultures: general features

To begin with traditional cultures, sociological and anthropological accounts suggest a number of generalizations about the ethical dimensions (in the broadest sense of practical reason) of their modes of thought and behaviour. A number of formal features related more to the kinds of moral systems there are, and less to the particular ends or values they endorse, must be considered.

First, it is widely agreed (Danquah 1944; Mbiti 1969) that these earlier societies were essentially *communitarian* or *communalistic* in their ethical ideas. This means that they held (like many pre-modern societies elsewhere) that prerogatives of rights to land use and property inheritance, for example, inhere in various corporate groups, such as families, lineages, villages and societies, rather than in individuals. These societies also maintained that the flourishing of corporate interests were superior to the projects of individuals. Thus, property rights - the claim to exclusive use of an area of land for farming for a period of time - were assigned by chiefs or kings in many cultures to lineages. The head and senior members of each such group would allocate responsibilities and a share of the crop to members of the group and would manage profits from sales or exchanges to cover the needs of individual members. Marriage and the attendant obligation to raise and support children would be a relationship between families in which control of children and their obedience often belonged to the family of one spouse. This was symbolically acknowledged by dowry payments for a future bride.

A second significant aspect of traditional ethical thought relates to the extended family or clan. This form of moral thought permits someone, the agent, to treat someone else, the patient, in a certain manner because the two are

related, not because the patient has particular qualities or needs, for example, I might give shelter to someone 'because she is my kinswoman'. This produces a structure of thought that is, in a sense, *anti-universalizing*. The reasoning behind it is historical as well as indexical and relates to the agent's conviction that the patient can be treated in a certain way: 'because she's my kinswoman', although this is not consonant with some understandings of the Kantian demand for universalizability (see Kantian ethics). (However, it is possible to defend partiality towards, for example, one's close kin as a universalizable principle. It is a familiar idea in European societies that obligations to a parent or sibling should not depend on their intrinsic attributes: Europeans and Americans do not normally suppose that the reason we should care more for our siblings than for strangers is that they have general features that make them abstractly more deserving.) Such reasons for anti-universalizability are independent of the character and needs of the patient, and do not arise out of either obligations created through consent whose model is the contract in liberal political theory (for example, Hobbes, Locke and Rawls) or reciprocation (which might account for obligation) to those who bore and brought one up. This range of obligations, rooted in unchosen relationships to people from whom one has received nothing, is much more extensive in traditional Africa than in Europe or the USA.

In matrilineal societies, for example, the Akan of Ghana and the Ivory Coast, children 'belong' to their mothers' families, which are headed ideally by a maternal uncle or great-uncle. In patrilineal traditions, for example, the Yoruba, the child 'belongs' to their fathers' lineage, which is headed ideally by a father, paternal uncle or grandfather, although these are not the only forms of affiliation known in Africa. These differing ideas of kinship have profound consequences because they determine the range of familial obligation and define who retains custody of the children after divorce. At this point, relationships enter the anti-universalizing range of the indexical reason mentioned earlier: 'because they are my kin'.

A third aspect of traditional ethical thought deals with traditional religious ideas. These were *naturalistic*, conceiving of gods, spirits and ancestors as continuous with and continually operative within the natural world. These ideas did not make a sharp ontological distinction, of the kind central to Christian and Platonist thinking, between a realm of spirit and a realm of matter. Spiritual beings were supposed to be concerned with the flourishing of specific groups of people of certain places or lineages. Because of the centrality of the prescriptions of various spiritual forces in shaping behaviour, the separation of moral principles from principles of prudence is not always easy in such a setting. (Plato's discussion in the *Euthyphro* (*c*.386-*c*.380) of whether an act is right because the gods love it or they love it because it is right, raises an unanswerable question (see Plato §2).) On the one hand, it seems clear that traditional peoples felt obligations to and wished to help others for reasons other than the spiritually-underwritten sanctions associated with not doing so (in the same way that most people do not commit murder simply because they know they would be punished if they did). On the other hand, certain proscriptions, of incest for example, that might be thought of as moral in some European societies, seem to have derived from beliefs about the consequences, as when the Azande understood leprosy to be caused by incest.

Many areas of life, such as food preparation, ritual practice in association with cults for spirits and ancestors, and everyday contact with others, generally involved the observance of taboos, for which the sanctions were often misfortunes, like hunting accidents, disease, or a failed crop visited upon an individual or the group to which they belonged. The avoidance of harm of this type was apparent in practical reasoning. Rarely was there an articulated notion of doing right that was distinct from doing what was appropriate and conformed to custom and prudence so that no taboos were violated. Indeed, one of the highest terms of commendation for conduct in many African societies was also used to describe both physical beauty and the aesthetic appeal in artefacts. Thus, in Twi, an Akan language, *fé* refers to physical beauty in people, to a well-made artefact and to commend good behaviour. A common criticism of a child's behaviour would be that it is not *fé* (see Akan philosophical psychology).

Fourth, there is often no clear distinction between moral vocabulary and aesthetic or technical commendation. This reflects the low degree of linguistic division of labour within most pre-literate communities.

A further consequence of regular involvement with spirits of lineage and place is that, in the majority of cultures that recognize a high god associated with the heavens, this god is a *deus remotus* whose interest in peoples' affairs is minimal. The notion that one's conduct should be aimed at affecting an afterlife is unusual and concern for affairs after one's death focuses on questions of one's descendants performing cultish acts for their ancestors to conserve their memory. Such systems of thought are thus *anthropocentric* or *humanistic*, centring on the

contemporary concerns of human beings, including ancestors (who are always with their kin), rather than on otherworldly considerations.

3 Traditional cultures: ethical education

The transmission of ethical ideas from one generation to the next in pre-literate societies was accomplished by counsel and correction, as well as in other important ways. In many cultures young people underwent periods of education removed from the community prior to the transition to adulthood. In these periods ideas about behaviour appropriate to one's gender and status, notions of what kind of person is worthy of respect, and conceptions of one's duties to others (including ancestors and other nonhuman spirits) were often inculcated.

It was common to recount the history (rather than the myth, as in contemporary usage myth connotes fictional status and these narratives are usually presented as true), of one's own people in terms that underwrote existing social arrangements. Ritual enactment of these histories provided a reinforcement for the broadly ethical understandings they underwrote. Another major source of moral instruction was through the various forms of *orature*, the verbal art that corresponds to the literature of literate societies. Among the most important of these are the folktale, a story told as fiction often with an Aesopian moral, and the proverb, a shorthand reference to history and the folktale. Proverbs have been seen as evidence of the moral notions of earlier African societies. Many of those who have written on this subject, including J.B. Danquah (1944), have relied on their interpretation to deduce traditional moral, and sometimes metaphysical, views. This process has its dangers: the interpretation of proverbs is a matter of context. In English, for example, it is easy to come up with two proverbs that might sound contradictory: 'more haste, less speed' and 'a stitch in time saves nine'. Which one is appropriate depends on whether the context refers to being timely or being careful. To infer from the first of these proverbs that UK culture values care over punctuality would be an error. Analogous mistakes have been made by observers interpreting African proverbs.

In African traditions the interpretation of proverbs is a difficult matter. Indeed, one Akan proverb says: 'if we tell a fool a proverb, we explain its meaning to him'. Many proverbs are less obvious in their application than the few with which most Europeans and Americans are familiar, because the language is densely figurative. A typical Akan proverb says: 'The drongo (bird) says: if he had known that the palm nuts were going to ripen, then he would not have married the raffia palm with a twisted leg'. Any contextual application of this proverb requires knowledge of the two plants in question: they are both palms, but one produces a valuable oil, while the other produces less valuable fibres. The knowledge that, before a certain stage of development, it is not obvious which is which, is also required. With this knowledge, it might be grasped that, in a particular context, the proverb was being used to express regret at having made the wrong choice in a situation of insufficient knowledge. In this context, the proverb could be said to be analagous to: 'more haste, less speed'. However, it would be wrong to infer that Akan culture values caution over efficacy. In fact, more can be learned about cultural values by exploring a proverb's presuppositions, than by the message it communicates.

There is an Akan proverb that says: 'when your child cheats your co-wife's child, it is not good; but when your co-wife's child cheats your child it is also not good'. It is often used in circumstances where the UK proverb, 'do as you would be done by', might be invoked. This proverb is used to make a point unconnected with marriage or family life and presupposes that men may have more than one wife at a time, that there is likely competition among co-wives and that this is something to be avoided. These presuppositions are communicated to young people as they grow up through the contextual and didactic use of such proverbs.

Proverbs are, therefore, an important source of ethical ideas in many cultures in Africa. As a result, the careful analysis of proverbs, especially where the analyst deals with their presuppositions, can reveal a culture's values.

4 Comparisons with 'modern' societies

In terms of content, there is less similarity among African traditions than at the level of the abstract formal features catalogued in §§2 and 3. The communitarian and anti-universalist character of ethical thought means that in most traditional African systems of ideas, public questions of practical reason, for example, in courts or in family decision making, rarely focused on the enforcement of individual rights. Instead, they concentrated on individual needs and a concern for corporate harmony. These tendencies remain in the areas where the influences of

modernity (in the form of contemporary critiques by African intellectuals, of world religions and of Euro-American ideas), have not percolated. This is in contrast to formal juridical practice in most modern societies (including the modern sectors of most African states).

Another major contrast is with the central questions of propriety, reputation and status in practical reasoning. This is striking because in Euro-American culture, at least since the Enlightenment, such questions are related to the surface of action and a concern with them is seen as superficial in ways that contrast with genuine moral concerns.

In Akan cultures the notion of *animuonyam*, which comes close to the meaning of 'respect' or 'honour', plays a central role. The possibility of losing respect is very troubling for those who have it. (Well-known Akan proverbs such as the following show that respect is not something that belong to everybody: 'father soul and father slave Kyereme, neither of them has any respect'. This means that whatever you call him or her, a slave is still a slave.) But *animuonyam* is not just related to being thought well of: it is a matter of meriting a high opinion and being seen to merit it. One way of earning respect is by providing generously for one's fellow citizens through acts of munificence or serving one's country well in warfare.

A familiarity with European or Asian moral philosophy begs the question of how public ethical discourse relates to the inner language of private decision. The difficulty in understanding this relationship stems from the fact that access to such information in those cultures is a consequence of literacy. In pre-literate societies there is no record of ethical views except through the collective sphere of oral traditions, neither are there generic conventions, such as the confession (on the model of St Augustine), or the protestant diary (as the record of an individual conscience) to provide models of organized reflection on questions outside the public sphere. More than this, central genres such as the novel provide us with models of interiority in moral life that do not occur in the major genres of oral literature. To the extent that a nuanced language of private moral reflection, distinct from the discourse of collective decision making, is the product of the kind of privacy made possible through the conventions that constitute genres of writing such as autobiography and the novel, it can be said that there is no such private morality in the pre-literate cultures of precolonial Africa.

All of these latter facts lead anthropologists to speak of cultures of (external) shame as opposed to (internal) guilt. In using such a classification, it is important to remember that shame, like guilt, requires a person to share the values that they have failed to live up to, rather than simply incurring embarrassment. Therefore, it would be misleading to think of African cultures without developed discourses of individualized (or private) morality being governed by the notion that it is only what is known publicly that is of any real ethical significance. In Asante, it is said that the man whose wife has had intercourse with another must be paid adultery fees even if intercourse occurred in a dream. This is because, in the Akante culture it is understood that dreams reflect the real experiences of one's *sunsum* and the idea that the dream is private does not exist. Thus, if a man dreams of intercourse with a woman she is assumed to have dreamed of intercourse with him. It is difficult to see how such a rule would come into being in a culture where people thought that what others did not know was of no ethical significance (see Akan philosophical psychology).

The distinction between what I know and what others know is less crucial in cultures where everyone is supposed to be the subject of continuous surveillance by spirits of various sorts and where forms of divination and spiritual vision may allow others to discover what I have done by means other than witnessing or hearing about it.

A final difference, which is as a result of the lack of models of self-reflection provided in the major modern literary genres, is that certain moral ideas, such as the modern notion of authenticity, do not appear to have been present in most African moral traditions. There are notions of behaviour appropriate to one's status or role, and notions of honesty that come close to the idea of sincerity. However, the idea that a person can truly be defined as one thing or another beneath the masks that convention and ascribed status impose, is foreign. Instead, there are notions of destiny or fate. For example, in many west African traditions, along the whole of the Guinea coast, there is an idea that people are offered or given a fate by a high god before entering the world, and that this prenatal message shapes their life on earth. It is worthy of note that, the contrasts between African ethical ideas and those of modern Europe and the USA, do not necessarily distinguish African ethics from those of other traditional peoples who are outside the areas of influence of the literate ethical culture of the monotheistic world religions.

5 Contemporary trends

Against this background, it is possible to understand the ethical crises associated, on the one hand, with the growth of literacy, the development of the African novel, and access to Western literary forms, and on the other, with urbanization and the growth of commodity production and wage labour. Each of these changes contributes to undermining the compatibility of traditional ethical ideas of kinship and community and contemporary practical life. The problem of political corruption in many African states has arisen in part from the fact that state officials in the modern sector are paid salaries intended to support a family unit closer to a Euro-American nuclear model. These officials inherit obligations to corporate groups, such as lineages and places of origin, that they cannot sustain. Similarly, the question of nepotism or tribalism in the state bureaucracy should be seen in the context of a conflict between formal rights-based notions of the role of state agents and traditional corporate obligations.

As far as sub-Saharan ethical discourse is concerned, there is evidence of the influences of Islamic and Christian theory, as well as Marxist and liberal political and moral ideas introduced through European colonial education. These latter secular traditions are not too different from their Euro-American counterparts, although Marxist revolutionaries in Africa, such as Amílcar Cabral have shared with Mao Zedong the problem of reshaping the ideas of Marx and Engels for predominantly nonindustrial societies.

With regard to major religious traditions, the contrast between traditional and contemporary ideas is less striking in Islamic areas. This is because, in many parts of sub-Saharan Africa, Islam was well-established before the period of European colonialism and existed for a long period as a literate tradition alongside the popular culture of a largely nonliterate populace. As a consequence, there was a mutual accommodation of Islamic ideas and the ethical thought of the pre-Islamic cultures into which they were introduced. It is said, therefore, that African Islam is highly syncretic, melding with local traditions in the various places it has been established. Just as Islam in other parts of the contemporary Muslim world has developed new secularizing traditions and new forms of fundamentalism, so has Africa. Secularization has tended to be the response of a literate intelligentsia with substantial familiarity with European culture, while fundamentalism has more of a popular, although literate, base. A consequence of the increasing flow of information through broadcast media and literacy, is that public ethical discourse in sub-Saharan Islamic Africa has grown closer to that of the rest of the Muslim world than it has been since the early period of African Islam.

The effect of Christianity has been to challenge the content of traditional ethical thought in many areas, such as objections to polygamy and caste-like hereditary status systems, and to appeal to ancestors and other spirits. (This latter conflict is less distinct because of the strong tradition of reference to saints, angels, devils and other ‘principalities and powers’ in European and north African Christian traditions whose spirits resemble African ones.) Furthermore, Christian otherworldliness and individualistic notions of merit and responsibility (deriving from contemporary economic relations) have penetrated many societies. It is reasonable to assume that the ethical thought of most contemporary African Christians is closer than it may appear to that of the precolonial period, especially in its communalism. To the extent that African Christians participate in formal literate ethical discourse, they do so within contemporary traditions that belong to European Christianity, with the exception of Coptic churches in Egypt and Ethiopia.

See also: Aesthetics, African; Confucian philosophy, Chinese; Family, ethics and the; Honour; Practical reason and ethics; African traditional religions

K. ANTHONY APPIAH

References and further reading

Ackah, C.A. (1988) *Akan ethics: a study of the moral ideas and the moral behaviour of the Akan tribes of Ghana*, Accra: Ghana University Press.(Discussion of Akan ethics in theory and practice.)

Beidelman, T.O. (1993) *Moral imagination in Kaguru modes of thought*, Washington, DC: Smithsonian Institution Press.(Exploration of the ethnography of Kaguru moral life.)

Danquah, J.B. (1944) *The Akan Doctrine of God: A Fragment of Gold Coast Ethics and Religion*, London: Frank Cass, 1968.(A good source for the claim of the communitarian character of African traditions.)

Evans-Pritchard, E.E (1976) *Witchcraft, Oracles, and Magic Among the Azande*, abridged with introduction by E. Gillies, Oxford: Clarendon Press.(A classical discussion of how practical reasoning is affected by traditional religious beliefs.)

Fortes, M. (1983) *Oedipus and Job in West African religion*, Cambridge: Cambridge University Press.(Discusses

the role of ideas about fate in the thought of a number of west African cultures.)

Greenberg, J.H. (1970) *The Languages of Africa*, Bloomington, IN: Indiana University Press, 3rd edn.(The best source for material on African languages.)

Howard, R. and Donnelly, J. (1996) 'Human Dignity, Human Rights and Political Regimes', *American Political Science Review* 80: 801-17.(Makes the point that notions of dignity are more central than notions of rights in many African traditions.)

Mbiti, J. (1969) *African Religions and Philosophy*, London: Heinemann, 2nd edn, 1990.(Covers a range of philosophical and religious ideas across Africa.)

Rattray, R.S. (1916) *Ashanti Proverbs*, Oxford: Clarendon Press.(An early collection of African proverbs which indicates their range and ethical significance.)

Ethics

1 Ethics and meta-ethics

What is ethics? First, the systems of value and custom instantiated in the lives of particular groups of human beings are described as the ethics of these groups. Philosophers may concern themselves with articulating these systems, but this is usually seen as the task of anthropology.

Second, the term is used to refer to one in particular of these systems, 'morality', which involves notions such as rightness and wrongness, guilt and shame, and so on (see Rectification and remainders). A central question here is how best to characterize this system. Is a moral system one with a certain function, such as to enable cooperation among individuals, or must it involve certain sentiments, such as those concerned with blame (see Morality and ethics; Moral sentiments; Praise and blame; Reciprocity)?

Third, 'ethics' can, within this system of morality itself, refer to actual moral principles: 'Why did you return the book?' 'It was the only ethical thing to do in the circumstances.'

Finally, ethics is that area of philosophy concerned with the study of ethics in its other senses (see Ethics in Islamic philosophy). It is important to remember that philosophical ethics is not independent of other areas of philosophy. The answers to many ethical questions depend on answers to questions in metaphysics and other areas (see Aesthetics and ethics; Metaphysics; Pragmatism in ethics). Furthermore, philosophers have been concerned to establish links between the ethical sphere of life itself and other spheres (see Art and morality; Law and morality). Some philosophers have, for philosophical reasons, had doubts about whether philosophy provides anyway the best approach to ethics (see Theory and practice; Wittgensteinian ethics). And even those who believe philosophy has a contribution to make may suggest that ethical justification must refer outside philosophy to common sense beliefs or real-life examples (see Examples in ethics; Moral justification).

A central task of philosophical ethics is to articulate what constitutes ethics or morality. This project is that of meta-ethics. What is it that especially constitutes the moral point of view as opposed to others? Some argue that what is morally required is equivalent to what is required by reason overall, whereas others see morality as providing just one source of reasons (see Practical reason and ethics). Yet others have suggested that all reasons are self-interested, and that concern for others is ultimately irrational (see Egoism and altruism). This has not been seen to be inimical in itself to the notion of morality, however, since a moral system can be seen to benefit its participants (see Contractarianism; Decision and game theory).

The moral point of view itself is often spelled out as grounded on a conception of equal respect (see Equality; Respect for persons). But there is some debate about how impartial morality requires us to be (see Impartiality).

Another set of issues concerns what it is that gives a being moral status, either as an object of moral concern or as an actual moral agent (see Moral agents; Moral standing; Responsibility). And how do our understandings of human nature impinge on our conception of morality and moral agency (see Morality and identity)?

Once we have some grip on what ethics is, we can begin to ask questions about moral principles themselves. Moral principles have often been put in terms of what is required by duty, but there has been something of a reaction against this notion (see Duty). Some have seen it as outdated, depending on a conception of divine law with little relevance to the modern world (see Anscombe, G.E.M.; Schopenhauer, A.); while others have reacted against it as a result of a masculine overemphasis on rules at the cost of empathy and care (see Feminist ethics; Wollstonecraft, M.).

These doubts are related to general concerns about the role principles should play in ethical thought. Situation ethicists suggest that circumstances can lead to the abandonment of any moral principle, particularists arguing that this is because it cannot be assumed that a reason that applies in one case will apply in others (see Moral particularism; Situation ethics). The casuistical tradition has employed moral principles, but on the understanding that there is no 'super-principle' to decide conflicts of principles (see Casuistry). At the other end of the spectrum, some philosophers have sought to understand morality as itself constituted by a single principle, such as that not to lie (see Wollaston, W.).

Duties have been seen also as constituting only a part of morality, allowing for the possibility of heroically going beyond the call of duty (see Supererogation). This is a matter of the scope of the notion of duty within morality. There are also issues concerning the scope of moral principles more generally. Does a given moral principle apply everywhere, and at all times, or is morality somehow bounded by space or time (see Moral relativism; Universalism in ethics)? This question is related to that concerning what is going on when someone allows morality to guide them, or asserts a moral principle (see Epistemology and ethics; Moral judgment; Moral knowledge). How is the capacity of moral judgment acquired (see Moral education)? The view that humans possess a special moral sense or capacity for intuition, often identified with conscience, is still found among contemporary intuitionists (see Common-sense ethics; Conscience; Cudworth, R.; Hutcheson, F.; Intuitionism in ethics; Moral sense theories; Moore, G.E.; Ross, W.D.; Shaftesbury). Scepticism about the claims of morality, however, remains a common view (see Moral scepticism; Nietzsche, F.).

In recent centuries, a dichotomy has opened up between those who believe that morality is based solely on reason, and those who suggest that some nonrational component such as desire or emotion is also involved (see Hume, D.; Morality and emotions; Rationalism). Denial of pure rationalism need not lead to the giving up of morality. Much work in the twentieth century was devoted to the question whether moral judgments were best understood as beliefs (and so candidates for truth and falsity), or as disguised expressions of emotions or commands (see Analytic ethics; Emotivism; Hare, R.M.; Logic of ethical discourse; Prescriptivism; Stevenson, C.L.). Can there be moral experts, or is each person entirely responsible for developing their own morality (see Existentialist ethics; Moral expertise)? These questions have been seen as closely tied to issues concerning moral motivation itself (see Moral motivation). Moral judgments seem to motivate people, so it is tempting to think that they crucially involve a desire.

Moral principles can be understood to rest on moral values, and debate continues about how to characterize these values and about how many evaluative assumptions are required to ground ethical claims (see Axiology; Constructivism in ethics; Moral pluralism; Values). Against the emotivists and others, moral realists have asserted the existence of values, some identifying moral properties with those properties postulated in a fully scientific worldview (see Fact/value distinction; Moral realism; Naturalism in ethics; Value, ontological status of).

2 Ethical concepts and ethical theories

Some philosophical ethics is broad and general, seeking to find general principles or explanations of morality. Much, however, focuses on analysis of notions central to ethics itself. One such notion which has been the focus of much discussion in recent years is that of autonomy (see Autonomy, ethical). The interest in self-governance sits alongside other issues concerning the self, its moral nature and its ethical relation to others (see Akrasia; Determinism and indeterminism; Evolution and ethics; Free will; Self-deception, ethics of; Self-respect; Will, the); and the relations of these selves in a social context (see Recognition; Solidarity; Vulnerability and finitude). Other topics discussed include the nature of moral ideals, and the notions of desert and moral responsibility (see Ideals; Desert and merit; Moral luck).

The question of what makes for a human life that is good for the person living it has been at the heart of ethics since the Greek philosophers enquired into *eudaimonia* (‘happiness’) (see Aristotle; Eudaimonia; Happiness; Life, meaning of; Plato; Socrates). Once again, a philosopher’s theory of the good will almost always be closely bound up with their views on other central matters (see Good, theories of the). For example, some of those who put weight on sense experience in our understanding of the world have been tempted by the view that the good consists entirely in a particular kind of experience, pleasure (see Empiricism; Pleasure). Others have claimed that there is more to life than mere pleasure, and that the good life consists in fulfilling our complex human nature (see Perfectionism; Self-realization). Nor have philosophers forgotten ‘the bad’ (see Evil; Suffering; Suffering, Buddhist views of origination of).

Moral philosophy, or ethics, has long been at least partly concerned with the advocacy of particular ways of living or acting. Some traditions have now declined (see Asceticism; MacIntyre, A.); but there is still a large range of views on how we should live. One central modern tradition is that of consequentialism (see Consequentialism). On this view, as it is usually understood, we are required by morality to bring about the greatest good overall (see Teleological ethics). The nature of any particular consequentialist view, therefore, depends on its view of the good.

The most influential theory has been that the only good is the welfare or happiness of individual human and other animals, which, when combined with consequentialism, is utilitarianism (see Bentham, J.; Mill, J.S.; Utilitarianism).

It is commonly said that consequentialist views are based on the good, rather than on the right (see Right and good; Rights). Theories based on the right may be described as deontological (see Deontological ethics). The towering figure in the deontological tradition has been the eighteenth-century German philosopher, Immanuel Kant (see Kant, I.; Kantian ethics). Such theories will claim, for example, that we should keep a promise even if more good overall would come from breaking it, or that there are restrictions on what we can intentionally do in pursuit of the good (see Double effect, principle of; Promising).

In the second half of the twentieth century there was a reaction against some of the perceived excesses of consequentialist and deontological ethics, and a return to the ancient notion of the virtues (see Aretē; Theological virtues; Virtue ethics; Virtues and vices). Work in this area has consisted partly in attacks on modern ethics, but also in further elaborations and analyses of the virtues and related concepts (see Charity; Forgiveness and mercy; Help and beneficence; Honour; Hope; Innocence; Love; Prudence; Self-control; Trust; Truthfulness).

3 Applied ethics

Philosophical ethics has always been to some degree applied to real-life. Aristotle, for example, believed that there was no point in studying ethics unless it would have some beneficial effect on the way one lived one's life. But, since the 1960s, there has been a renewed interest in detailed discussion of particular issues of contemporary practical concern (see Applied ethics).

One area in which ethics has always played an important role is medicine, in particular in issues involving life and death (see Bioethics; Bioethics, Jewish; Life and death; Medical ethics; Suicide, ethics of). Recently, partly as a result of advances in science and technology, new areas of enquiry have been explored (see Genetics and ethics; Reproduction and ethics). In addition, certain parts of medical practice which previously lacked their own distinctive ethics have begun to develop them (see Nursing ethics).

This development is part of a wider movement involving research into the ethical requirements on those with particular occupations. Some of this research is again related to scientific advance and its implications for public policy (see Information technology and ethics; Responsibilities of scientists and intellectuals; Risk; Technology and ethics). But, again, attention has also been given to occupations not in the past subjected to much philosophical ethical analysis (see Business ethics; Journalism, ethics of; Professional ethics; Sport and ethics).

The planet, and those who live and will live on it, have in recent times become the focus of much political concern, and this has had its effect on philosophy (see Agricultural ethics; Animals and ethics; Development ethics; Ecological philosophy; Environmental ethics; Future generations, obligations to; Population and ethics). But just as the scope of ethical enquiry has broadened, so there has been renewed interest in the specific details of human relationships, whether personal or between society, state and individual (see Economics and ethics; Market, ethics of the; Family, ethics and the; Friendship; Paternalism; Political philosophy; Pornography; Sexuality, philosophy of).

See also: Aretē; Aristotle §§22-26; Augustine §§12-14; Callicles; Cheng; Confucian philosophy, Chinese; Cyrenaics §§3-4; Daoist philosophy; Duty and virtue, Indian conceptions of; Epictetus; Ethics in Islamic philosophy; Jia Yi; Li; Mencius; Panaetius; Self-cultivation in Chinese philosophy; Watsuji Tetsurō; Xin (trustworthiness); Zhu Xi

ROGER CRISP

References and further reading

Rachels, J. (1986) *The Elements of Moral Philosophy*, New York: Random House.(A helpful introduction to metaethics and ethical theory, with good use made of real-life examples. Contains suggestions for further reading.)

Singer, P. (ed.) (1991) *A Companion to Ethics*, Oxford: Blackwell.(Contains short, pithy articles on central topics in ethics, including metaethics, ethical theory, and applied ethics. The articles include useful bibliographies.)

Ethics in Islamic philosophy

*The study of Islamic ethics, whether philosophical or theological, grew out of early discussions of the questions of predetermination (*qadar*), obligation (*taklif*) and the injustices of temporal rulers, particularly the caliphs. Early writers on ethics from the Mu'tazila school were probably influenced by Greek philosophy. By the third century AH (ninth century AD) a clearly discernible current of philosophical ethics began to take shape, with strong influences from Greek ethics including Stoicism, Platonism and Aristotelianism.*

Al-Kindi, the first genuine philosopher of Islam, appears from his extant ethical writings to have been particularly influenced by Socrates and Diogenes the Cynic. Other classical influences can be seen in the work of Platonists such as Abu Bakr al-Razi, who followed Plato's division of the parts of the souls, and Neoplatonists such as al-Farabi, while Aristotelian influences can be seen in al-Farabi, who also discussed the problem of evil, Ibn Sina and Ibn Rushd. Ibn Sina developed a theory of the conjunction of the soul with the active intellect; with this conjunction is bound up the ultimate perfection of the soul which has attained the highest degree of wisdom and virtue.

Neoplatonism again surfaces in the work of Ibn Miskawayh and his followers, to whom we owe the groundwork of a whole ethical tradition which flourished in Persia well into the twelfth century AH (eighteenth century AD) and beyond. Onto Plato's threefold division of the soul, Ibn Miskawayh grafts a threefold division of virtue into wisdom, courage and temperance. His views were elaborated upon by al-Tusi and al-Dawwani, among others. A blend of philosophical and religious ethics is characteristic of the work of some later writers such as al-Ghazali and Fakhr al-Din al-Razi, in which the road to moral and spiritual perfection has mystical overtones.

1 Theological preludes

The earliest ethical discussions in the seventh and eighth centuries appear to have centred on the question of *qadar*, which could equally mean 'capacity', as predicated of man, or 'predetermination', as predicated of God. The early so-called Qadarites of Damascus raised the question of *qadar* in the context of the moral responsibilities of the Umayyad caliphs, who justified their most oppressive policies on the ground that they were part of the divine decree (*qada' wa qadar*). Subsequently, the Mu'tazilite theologians of Basra and Baghdad refined upon the speculation of their Qadarite predecessors and attempted for the first time to give an adequate definition of right and wrong and its bearing on God's justice and his decrees in the world. This definition was expressed in essentially rationalist and deontological terms and was received with disapproval by their traditionalist and conservative rivals, who adhered to a voluntarist thesis according to which right is by definition what God commands and wrong is what he prohibits (see Ash'ariyya and Mu'tazila §5).

Other instances of Mu'tazilite rationalism include the espousal of the absolute identity of God's essence and his attributes, the irreversibility of his decrees and the freedom of the will as a precondition of moral responsibility. Mu'tazilite theologians also stressed God's wisdom and goodness and exonerated him of the responsibility for evil in the world, which was 'created' by humankind. The degree to which Mu'tazilite theological ethics was influenced by Greek philosophy cannot be fully determined from our present knowledge of the early sources, since the translation of philosophical texts had not been started by the time the founder of the school, Wasil ibn 'Ata, launched this radical theological movement in the second century AH (eighth century AD). However, contact with Christian theologians, such as John of Damascus and his disciple Theodore Abu Qurrah, was definitely a factor in initiating the Qadarite theologians, the forerunners of the Mu'tazila, into the scholastic methods of discourse that Syriac-speaking Christian scholars had been applying to theological questions prior to the Arab conquest of Syria, Egypt and Iraq.

2 The rise of philosophical ethics: early Socratic and Stoic trends

Al-Kindi, the first genuine philosopher of Islam, was also the first writer on philosophical ethics, and it is significant that he was in sympathy with Mu'tazilite theology during the heyday of that movement. However, unlike his Mu'tazilite contemporaries, whose starting-point was the Qur'an and the Traditions (*hadith*) of Muhammad, al-Kindi's starting point was Greek philosophy. He is reported by the classical bibliographers to have written a number of ethical treatises reflecting a profound interest in Socratic thought. Thus, in addition to a

treatise on *Ethics*, he is credited with a work on *Paving the Way to Virtue*, as well as an extant tract, *Fi al-hila li-daf‘ al-ahzan (On the Art of Dispelling Sorrows)*. Of his Socratic writings, a tract on the *Excellence of Socrates*, *A Dialogue between Socrates and Aschines* and a short collection, *Alfaz Sugrat (Socratic Utterances)*, which has survived, are mentioned in the classical sources.

It is noteworthy, however, that in the last-mentioned collection the personalities of Socrates and Diogenes the Cynic (see Diogenes of Sinope) are fused and both emerge as paragons of virtue and asceticism. However, in the more discursive *Fi al-hila li-daf‘ al-ahzan* the Stoic ideal of *apatheia* (freedom from passion) and the consequent indifference to the vicissitudes of fortune are set out in eloquent terms. The antidote of sorrow, which al-Kindi argues in Stoic fashion is inseparable from humanity’s ephemeral condition in this world of generation and corruption, is to consider that sorrow results either from our actions or from the actions of others. In the first case, it is our duty as rational agents to refrain from doing that which is the cause of sorrow. In the second case, averting the sorrow which results from the actions of others is either in our power or it is not. If it is in our power, we ought certainly to avert it; if it is not in our power, we should not grieve at the prospect of injury in the hope that it might somehow be turned away. However, should we nevertheless be afflicted by sorrow resulting from actions over which we have no control, it is our duty as rational beings to bear this with fortitude. This view, and the exhortation to shun material possessions as temporary acquisitions of which we are mere borrowers and not real owners, reflect clearly the influences of the great second-century Stoic teacher Epictetus and appear to derive from his famous *Enchiridion*.

However, it was Socrates and his disciple Plato who were at the centre of the moral-ascetic speculation of the early Muslim ethical philosophers. Abu Bakr al-Razi, in his ethical treatise *al-Tibb al-ruhani (The Spiritual Physic)*, refers to Plato as ‘the master of the Philosophers and their leader’, and to his teacher Socrates as ‘the ascetic and godly’ sage. He speaks of the three Platonic parts of the soul as the rational or divine, the irascible or animal and the concupiscent or vegetative souls, as these parts are designated by the Alexandrian physician-philosopher Galen in a treatise on ethics (which has survived in Arabic translation only). He then proceeds to summarize Plato’s teaching on the manner in which each of the parts or ‘souls’ should be ‘managed’ through reasoning and demonstration, a process which he labels ‘spiritual physic’ or therapy.

Another Socratic-Platonic theme which recurs in al-Razi’s writings is the folly of the hedonistic life which turns man into a slave or a beast. Because so many of our pleasures are either ephemeral or unattainable, we are assailed by anxiety or grief. But the true philosopher will not succumb to grief, because he understands that nothing in this world of generation and corruption is ever permanent and that whatever cannot be turned away should be ignored or disregarded, since it is often the product of passion and not of reason: ‘For reason summons us only to what is susceptible of bringing about profit sooner or later; grief does not bring any advantage…. That is why the perfectly rational man will only follow the summons of reason…and will never follow the summons of passion or allow himself to be led by it or get close to it’ (*Rasa’il al-Razi al-falsafiya*: 69).

Like Socrates and Plato, al-Razi believed that the soul on leaving the body will return to its original abode in the intelligible world, after passing through an endless cycle of purifications. That, he argues, is why the fear of death is irrational, and like al-Kindi and other moral philosophers he admonishes the truly reasonable man to resign himself to the prospect of death as a logical consequence of our being human. As al-Kindi had written: ‘Since the definition of man is that he is a living, dying, rational being, then if there was no death, there would be no man’ (*al-Hila li-daf‘ al-ahzan*: 45), death being an essential part of the very definition of man.

However, al-Razi adds to this argument another argument which appears to derive from Epicurus, namely that death is the privation of sensation (see Epicureanism §13); in death, man is stripped of the sensations of pleasure or pain and thus is in a better condition than the living who are constantly subjected to pain, let alone the fact that pleasure sought by our concupiscent nature is really nothing but ‘relief from pain’ or return to the natural condition (as Plato has taught in the *Philebus*). Therefore, ‘according to the judgment of reason the condition of death is better than the condition of life’ (*Rasa’il al-Razi al-falsafiya*: 93).

3 The advent of Aristotle

The first systematic writer on philosophical questions in Islam was al-Farabi, who had also contributed to ethical discussions. He is reported in the classical sources to have written a commentary on ‘parts’ of the *Nicomachean*

Ethics, translated into Arabic by Ishaq ibn Hunayn. This commentary is lost, but judging from an extant collection of *Fusul muntaza'ah (Excerpts on Ethics)*,which he is established as having written, he appears to have followed Aristotle's lead in dividing the virtues into moral (practical) and intellectual (see Aristotle §§22-25). The former, he says, are the virtues or perfections of the concupiscent part or faculty of the soul; they include temperance, courage, liberality and justice. The latter are the perfections of the intellectual part and include practical reasoning, good judgment, sagacity and sound understanding.

In his discussion of justice, al-Farabi also follows Aristotle's lead, arguing that justice consists in the equitable distribution of 'common goods' in the city or the state. These goods include security, wealth, dignity and public office, of which every member of the city or state is entitled to a share. Another more general meaning of justice is given as 'man's exercise of virtuous actions in himself and in relation to others, whatever such a virtue might be'.

An interesting feature of al-Farabi's ethics which has no Aristotelian parallel is the discussion of evil. He starts in Neoplatonic fashion by asserting that 'evil has no existence as such in anything found in these worlds; that is generally in whatever does not exist through human volition. Everything therein is good' (*Fusul muntaza'a*: 81). Evil, then, is a predicate of human action, not of physical occurrences. However, al-Farabi disagrees with traditional Neoplatonists who identified being with the good and not-being with evil pure and simple, on the ground that 'being is good only when it is in conformity with justice (or merit); not-being is evil when it is not in conformity with justice' (*Fusul muntaza'a*: 81). This appears to reflect Heraclitus' concept of *dike* as chaos (see Heraclitus §3).

Al-Farabi's successor and spiritual disciple Ibn Sina (Avicenna) is the author of a very short tract on ethics which follows closely the Platonic model in psychology. Ibn Sina divides the soul into the rational, irascible and concupiscent, to which correspond the virtues of wisdom, courage and temperance respectively, and with justice being the 'summation' of all three. To ensure the enforcement of justice within the state, argues Ibn Sina, the existence of the caliph, as conceived of by the Shi'ites, is necessary as the sovereign of the world and God's vicegerent on earth.

More explicitly than al-Farabi, Ibn Sina develops in his psychological writings a theory of conjunction (*ittisal*) of the soul with the active intellect, that supermundane agency which according to the Muslim Neoplatonists governs the sublunary world (see Neoplatonism in Islamic philosophy §2). With this conjunction, he argues, is bound up the ultimate perfection of the soul which has attained the highest degree of wisdom and virtue, becoming thereby a replica or a mirror of the higher intelligible world. Therein lies man's ultimate happiness while his soul is still in the body.

Ibn Rushd, the great Aristotelian philosopher and commentator, is known from the bibliographical sources to have written a paraphrase and a middle commentary on the *Nicomachean Ethics*, which have survived only in Hebrew and Latin, together with a paraphrase of Plato's *Republic* which is also relevant to his ethical theory. The principal virtues, according to Ibn Rushd, correspond to the perfection of the three parts of the soul, the rational, the irascible and the concupiscent. Justice is then described along Platonic lines as the 'harmony' of the three corresponding virtues of wisdom, courage and temperance; but it has, as Aristotle stated in the *Nicomachean Ethics*, two subdivisions which Ibn Rushd calls common or universal, corresponding to 'perfect virtue', and particular, whose further subdivisions are distributive and rectificatory. Contrary to expectations, however, Ibn Rushd does not identify happiness with the contemplative life, as Aristotle (§26) had done, but rather with conjunction (*ittisal*) with the active intellect, which the Muslim Neoplatonists - with whom he was at loggerheads - had regarded as man's ultimate goal.

4 Ibn Miskawayh and the Persian writers on ethics

The most important writer on ethics in Islam, however, was an eclectic who inclined to Platonism, Ahmad ibn Muhammad Ibn Miskawayh. He laid down in his *Tahdhib al-akhlaq (The Cultivation of Morals)* and other ethical writings the groundwork for a whole tradition of Persian ethical writing, the chief representatives of which were Nasir-i Khusraw (d. AH 467/AD 1074), Nasir al-Din al-Tusi, Jalal al-Din al-Dawani, Hussain Kashifi (d. AH 910/AD 1504), Mulla Ahmad Nuraqi (d. AH 1244/AD 1828) among others.

The psychological basis of Ibn Miskawayh's ethics is clearly Platonic, as was generally the case in Islamic philosophical circles. Onto Plato's threefold division of the soul (see Plato §14), as modified by Galen, he grafts a

threefold division of virtue into wisdom, corresponding to the rational part of the soul; courage, corresponding to the irascible part; and temperance, corresponding to the concupiscent part. Justice, which Ibn Miskawayh describes as a form of moderation (*i'tidal*) or proportion (*nisba*), arises when the three powers or parts of the soul are in harmony. This virtue is not part of virtue, but virtue entire, as Aristotle (§22) had argued. Its subdivisions, according to Aristotle as interpreted probably by Porphyry, are then given as three: our duties to God, to our superiors or equals, and finally to our ancestors. Ibn Miskawayh, however, refers to the genuine Aristotelian twofold subdivisions of justice into distributive and rectificatory and predicates the exercise of this supreme virtue on submission to the holy law (*shari'a*), which emanates from God. He then assigns to the 'just imam' or caliph the function of warding off the different forms of violating justice. Ibn Miskawayh even attributes to Aristotle, quoting an apocryphal Aristotelian source in Arabic translation, the view that justice stipulates rendering God the kind of worship due to him as our beneficent creator.

However, the Neoplatonic element in Ibn Miskawayh's ethics is nowhere more pronounced than in his analysis of happiness. Its two subdivisions, according to him, are practical and theoretical. The latter consists in 'conjunction' with the active intellect, whereby man is able to join the 'higher intellectual' realm. However, Ibn Miskawayh recognizes beyond this intellectual perfection a 'divine' or supernatural condition whereby man partakes of divine perfection or achieves a condition of self-divinization which goes far beyond his worldly conditions. This 'divine condition' is also alleged to derive from an Aristotelian fragment *On the Virtues of the Soul*, which Ibn Miskawayh quotes in Arabic translation, but which is clearly different from the apocryphal tract of the Aristotelian corpus known as *De virtutibus et vitiis*.

Ibn Miskawayh's two best-known Persian followers are Nasir al-Din al-Tusi, author of *Akhlaq-i Nasiri (Nasirean Ethics)*, and Jalal al-Din al-Dawani, author of *Lawami al-ishraq fi makarim al-akhlaq (Flashes of Illumination on the Nobility of Character)*, known also as *Akhlaq-i Jalali*. Both al-Tusi and al-Dawani follow closely the lead of Ibn Miskawayh in the *Tahdhib al-akhlaq*. A fundamental difference between Ibn Miskawayh and the latter two authors is the addition of 'household management' and politics to the purely ethical part of their work by both al-Tusi and al-Dawwani. This may be viewed as a broadening, in Aristotelian fashion, of the scope of practical philosophy, which Ibn Miskawayh had tended to confine to ethical discourse only.

In the political section, inspired chiefly by al-Farabi, al-Tusi argues that orderly association is an essential precondition of the good life. Of the three forms of government, the monarchical, the tyrannical and the democratic (which he attributes to Aristotle), he favours the monarchical, identified like Plato's with the 'rule of the virtuous' or aristocrats. However, the true monarch is assisted by divine inspiration but is subordinate to the imam, who according to Shi'ite doctrine is in 'temporary concealment'. This monarch acts accordingly in a vicarious or interim capacity to ensure the administration of justice in the absence of the true head of the community or 'hidden imam'.

Al-Dawani's ethical treatise follows essentially al-Tusi's lead, but in genuine Shi'ite fashion he stresses more than his predecessor the position of humans as God's vicegerent (*khalifa*) on earth (Surah 2: 30). In mystical fashion, he then goes on to argue that people reflect in their capacity as God's vicegerent the dual character of the divine nature, the outer and the inner, the spiritual and the corporeal, and more than any other creatures, including the angels, can be described as the 'image' of God. The foremost duty of the ruler, he argues, is to preserve the ordinances of the divine law (*shari'a*) and to conduct the affairs of state in accordance with universal principles and the requirements of the times. The ruler is for that reason God's 'shadow' and the vicar of the Prophet.

5 Philosophical and religious ethics

A specific blend of philosophical and religious ethics is characteristic of the writings of some late authors, including al-Raghib al-Isfahani (d. AH 502/AD 1108), Abu Hamid al-Ghazali, Fakhr al-Din al-Razi and others. Al-Ghazali is the foremost representative of this group, who in both his ethical treatise *Mizan al-amal (The Balance of Action)* and his religious summa, *al-Ihya' 'ulum al-din (The Revival of the Religious Sciences)*, has developed an ethical theory in which Platonic psychology serves as the groundwork of an essentially Islamic and mystical worldview. In this theory, the table of the four cardinal virtues accords with the Platonic virtues but admits of a series of subdivisions or ramifications analogous to those of his predecessors. A good example of the combination of religious and philosophical ideas in al-Ghazali is the manner in which happiness can be achieved. Happiness, as the chief good, admits of two subdivisions, the worldly and the otherworldly. Otherworldly

happiness, which is our ultimate goal, cannot be achieved without certain worldly goods. These include the four cardinal virtues of wisdom, courage, temperance and justice, the bodily virtues of health, strength, good fortune and a long life, the external virtues of wealth, kin, social position and noble birth, and finally the 'divine virtues' of guidance, good counsel, direction and divine support. Those virtues are referred to in the Qur'an and the *hadith*, al-Ghazali says, and the final virtue, 'divine support', is identified with the Holy Spirit (Surah 2: 87, 253) (see Virtues and vices).

The road to moral and spiritual perfection is described as the 'quest for God'. The seekers after God must satisfy two conditions: their actions must be governed by the prescriptions or ordinances of the 'divine law' (*al-shar'*), and they must ensure that God is constantly present in their hearts. By this presence al-Ghazali means genuine contrition, adoration and submission, born of the seeker's awareness of the beauty and majesty of God which al-Ghazali, like other Muslim mystics or Sufis, regards as analogous to human passion or love (*'ishq*) (see Mystical philosophy in Islam).

See also: Aristotelianism in Islamic philosophy; Ethics; al-Ghazali; Greek philosophy: impact on Islamic philosophy;Ibn Miskawayh; Logic in Islamic philosophy; Platonism in Islamic philosophy

MAJID FAKHRY

References and further reading

Alon, I. (1991) *Socrates in Mediaeval Arabic Literature*, Leiden: Brill; Jerusalem: Magnes Press, The Hebrew University.(A reliable review of Socratic literature in Arabic.)

Arkoun, M. (1970) *Contribution à l'étude de l'humanisme arabe au IV/Xe siècle (Contribution to the Study of Arab Humanism in the 4th-10th Centuries)*, Paris: Vrin. (A detailed account of Ibn Miskawayh's ethics and his philosophical impact.)

Donaldson, D.M. (1953) *Studies in Muslim Ethics*, London: SPCK.(Contains useful information on Arabic and Persian ethics, but is now somewhat out of date.)

Fakhry, M. (1994) *Ethical Theories in Islam*, Leiden: Brill, 2nd enlarged edn.(A systematic analysis of philosophical and religious ethical theories in Islam.)

al-Farabi (*c*.870-950) *Fusul muntaza'ah (Excerpt on Ethics)*, ed. F. Najjar, Beirut, 1971.(Al-Farabi's work on ethics, showing the influence of Aristotle.)

Hourani, G. (1985) *Reason and Tradition in Islamic Ethics*, Cambridge: Cambridge University Press.(An important collection of ethical studies by an eminent writer on ethical questions in Islam.)

Ibn Miskawayh (before 1030) *Tahdhib al-akhlaq (The Cultivation of Morals)*, trans. C.K. Zurayk, *The Refinement of Character*, Beirut: American University of Beirut, 1966.(Reliable and annotated translation.)

Khadduri, M. (1984) *The Islamic Conception of Justice*, Baltimore, MD: Johns Hopkins University Press.(A discursive and informative account of this central ethical concept in Islamic thought.)

al-Kindi (before 873) *Rasa'il al-Kindi al-falsafiyah (Al-Kindi's Philosophical Writings)*, ed. A.H. Abu Rida, Cairo, 1950.(Collected philosophical writings of al-Kindi.)

al-Kindi (before 873) *al-Hila fi-daf' al-ahzan (On the Art of Dispelling Sorrows)*, ed. H. Ritter and R. Walzer, *Uno scritto morale de al-Kindi*, Rome, 1938.(Contains al-Kindi's account of Stoic ideals.)

Leaman, O. (1995) 'Christian Ethics in the Light of Muslim Ethics', in C. Rodd (ed.) *New Occasions Teach New Duties?*, Edinburgh: T. & T. Clark, 219-31.(Comparative study of the differing roles played by ethics in Islam and Christianity.)

al-Razi (before 925) *Rasa'il al-Razi al-falsafiyah (Al-Razi's Philosophical Writings)*, ed. P. Kraus, Cairo, 1939. (Collected philosophical writings of Abu Bakr al-Razi.)

al-Tusi (1201-74) *Akhlaq-i Nasiri (The Nasirean Ethics)*, trans. G.M. Wickens, London: Allen & Unwin, 1964. (Translation of al-Tusi's *Akhlaq-i Nasiri* from Persian. Reliable on the whole.)

Ethiopia, philosophy in

Ethiopia is a unique phenomenon in Africa for three reasons. First, because of its historical continuity and political independence; second because of its written language and third because of its written philosophy. Ethiopian philosophy in the broadest sense is expressed in both oral and written language. In this entry only the written documents are dealt with. This methodological approach limits the investigation to linguistic and cultural phenomena as it deals with the ancient Semitic language known as Ethiopic and the Christian zones of influence on the high plateaus of Ethiopia.

There are five basic texts of Ethiopian philosophical literature: the Physiologue (the Fisalgwos) (c. *fifth century* AD*), The Book of the Wise Philosophers (1510/22), The Life and Maxims of Skendes* (c. *eleventh century* AD*), The Treatise of Zar'a Ya'ecob (1667) and The Treatise of Wäldä Heywåt* (c. *eighteenth century). The first three are adaptations of works transmitted from Greek sources through Arabic; the latter two (appearing in modern publications in a combined form) are original works of a rationalist flavour.*

1 The basic texts

The Ethiopic *Physiologue* (*c*. fifth century AD), a naturalistic-theological book of Christian symbolism, is a translation and adaptation from the Greek original. The Greek *Physiologue* seems to have been written between the end of the second century AD and the beginning of the third. There is near unanimity among scholars that Alexandria, Egypt was the place where it was revised. The Ethiopian translation was probably adapted around the beginning or middle of the fifth century. It was completed in Egypt by an Ethiopian living in a monastery of Skete where he had access to the Greek original.

The sixteenth century is particularly rich in the history of Ethiopic literature with a philosophical slant. At this time, Abba Mikael, an Egyptian, translated and adapted into Ethiopic *The Book of the Wise Philosophers* (1510/22). After its adaptation, an Ethiopian transcribed it into on parchment in Ethiopic. The Arabic text is thought to have been written by Hunain ibn Ishâq (*c*.809-?), the son of a Nestorian Christian. The Arabic version is based on the lost Greek original. This original belonged to the later literature of apophthegm in the form in which it flourished during the Byzantine period (seventh to early ninth century). The Greek influence is evident throughout *The Book of the Wise Philosophers* in its allusion to Presocratic (sixth and fifth centuries BC), Socratic (469-399 BC), Aristotelian (384-322 BC), and especially Platonic (427-347 BC) and Neoplatonic (from third to sixth or seventh centuries AD) dialogues. *The Book of the Wise Philosophers* is a collection of sayings, or a *liber sententiarum*. It presents the quintessence of what various philosophers have said on a certain number of topics, the greater number of which are ethical.

The Life and Maxims of Skendes (*c*. eleventh century AD) was also translated and adapted from the Arabic into Ethiopic at this time, although the identity of the author is unknown. The Arabic story seems to have been based on a Greek original of the second century AD. A careful examination of the Ethiopic and the Arabic texts shows that the Ethiopian translator of this version of the Oedipus story is clearly a deeply thinking person with sensitive powers of perception. The Skendes story, as it is conveyed in Ethiopic, is the most perfect and morally chastened of all preserved accounts of this type.

A close comparison of the Arabic, Ethiopic and Greek texts (there are parallels of the lost Greek original in other Greek works) reveals that the Ethiopians never translate literally except for the Holy Scriptures. They adapt, modify, add and subtract. A translation therefore bears a typically Ethiopian stamp. Although the nucleus of what is translated is foreign, the way it is assimilated and transformed is typically Ethiopian. This phenomenon, which is not limited to philosophical works but also applies to other cultural areas like iconography, is known as 'creative incorporation'.

The Treatise of Zar'a Ya'ecob (1667) is original in many ways. The author begins the treatise with the story of his life. It is the only autobiography in Ethiopic literature. Zar'a Ya'ecob (1599-1692) was born in the environs of Aksum, northwest Ethiopia. He attended traditional Ethiopian schools and attained the highest learning levels. Afterwards he taught Holy Scriptures for ten years. Denounced before the king, he was compelled to flee for his life. On his way to Shoa in the south (Addis Ababa is the modern capital of the province of Shoa), he found an

uninhabited cave at the foot of a valley, where he lived for two years. In peace and solitude he developed his philosophy. His philosophy is presented as the fruit of his own personal reflection and not as a translation-adaptation from foreign sources as is the case with most Ethiopic literature. It is clearly rationalistic in a religious (not scientific) sense. Rationalism is understood as the view which recognizes as true only that content of faith which can be made to appeal to reason. In Ethiopia, traditional philosophy in its written form is closely linked with Christianity in general and monasticism in particular. Zar'a Ya'ecob's rationalism, in contrast, insists precisely on the absolute and exclusive sufficiency of human reason and denies all dogmatic assertion that reason is unable to establish by its own means.

Through the application of his method, the light of reason, whose immediacy enlightened his investigation, Zar'a Ya'ecob discovered a basic principle: the goodness of created nature. From this foundation he moved towards theodicy, ethics and psychology. In his ethics, because he accepted the fundamental goodness of creation, he endorsed married life and the enjoyment of food, and rejected the time-honoured traditions of celibacy and fasting in Ethiopian monastic life. In his psychology he emphasized the freedom of the human and their superiority over the rest of creation.

As a contemporary of René Descartes, Zar'a Ya'ecob resembled him in many ways, although Ya'ecob did not make use of the method of universal doubt. In both these philosophers one finds a method, occasion for a critical inquiry and the discovery of a criterion that leads to the establishment of a basic principle, which is then applied in both authors to theodicy, ethics and psychology. Also in both, the method of inquiry is revolutionary, although its roots are deeply theological. The historical circumstances from which the rationalism of both philosophers originated are equally similar. They are based in the collapse of scholasticism, the harmony between faith and reason in Europe, and in Ethiopia, the confrontation with Western culture and violent religious discord (see Descartes, R. §1).

Zar'a Ya'ecob, unlike Descartes, was not a mathematical genius. His philosophy does not develop deductively as a linear growth from a first principle or original idea, but rather, like the unfolding of the sun's rays, it emanates from a single centre and safeguards the complexity and richness of reality.

After his death, Zar'a Ya'ecob's disciple Wäldä Heywåt wrote *The Treatise of Wäldä Heywåt* (*c*. eighteenth century) in which his master's last years are related and his thought is presented in a more pedagogical and accessible fashion.

2 From wisdom to rationalism

The five basic texts that have been presented span more than twelve centuries of literary production. They show a double evolution: from translation, the *Physiologue* (*c*. fifth century AD), with its frequent misunderstandings, *The Book of the Wise Philosophers* (1510/22), with its intuitions, *The Life and Maxims of Skendes* (*c*. eleventh century AD), with its sensitive, reflective perceptions) to originality, *The Treatise of Zar'a Ya'ecob* (1667), *The Treatise of Wäldä Heywåt* (*c*. eighteenth century); from wisdom to rationalism. The link with Christianity evolves from the Christian symbolism of the *Physiologue* to the theologically-oriented anthropocentrism of *The Book of the Wise Philosophers*, to the theistic pantheism of the first series of maxims in *The Life and Maxims of Skendes* to the anti-Christian, though mystical radicalism of Zar'a Ya'ecob and his disciple.

However, it is the very notion of philosophy, in both a broad and narrow sense, which shows the greatest evolution. In the widest sense it is synonymous with wisdom understood as involving knowledge and the ability, inclination and steady purpose of putting knowledge to good use. Thus evidence of wisdom in a person involves an end or purpose to be attained, an understanding of this purpose and an effort to achieve it in the best possible manner. Objectively, wisdom is the sum total of things worth knowing and ends worth working for.

The Book of the Wise Philosophers and *The Life and Maxims of Skendes* attempt, each in their own way, to express wisdom in the broad sense of philosophy. *The Book of the Wise Philosophers* presents the simplest and most common form of wisdom literature: the *angarä*, or saying. *The Life and Maxims of Skendes*, on the other hand, is a philosophical novel. The second and third sections, within the question and answer frame, cover the spectrum of forms of wisdom literature, including maxims, numbers, exhortations and the biographical portrait.

The Treatise of Zar'a Ya'ecob and *The Treatise of Wäldä Heywåt* are philosophy in a stricter sense. Each is the

result of the reflection of one person, and each has a clearly defined method, characterized by criticism, inquiry (the name of the treatise in Ethiopic is *hatätä*, or inquiry), and the use of the criterion of the light of reason as a basis upon which the whole edifice is founded. But, within the same rationalistic approach, a clear distinction appears between the master and his disciple. Zar'a Ya'ecob is the only one of the two who presents in a complete, systematic and original way, the methodology of his philosophy. Although Wäldä Heywåt shares the ideas of his master, the starting point and the dialectical movement from this origin are not found in his treatise. In this sense, he is in line with Ethiopian traditional thought (see Rationalism).

Since the *Physiologue* is not a philosophical work, one may ask why it has been placed at the beginning of a series of texts on Ethiopian philosophical literature. The following quotation might indicate why:

> Seen from the vantage point of later developments in Ethiopian philosophy, especially the XVI, XVII and XVIII centuries, the Fisalgʷos (the Physiologue) affords a first glimpse into its rudimentary beginning. It is like observing a great river in its initial streamlets. Nearly all the main characteristics of Ethiopian philosophy that will flourish in Mäshafä fälasfa (The Book of the Wise Philosophers), The Life and Maxims of Skendes, and The Treatise of Zar'a Ya'ecob and of Wäldä Heywåt are already present in the Fisalgʷos, although in an inchoate way: its thought patterns, its archetypal images and symbolism, its world view…, its anthropology, its social approach and its set of ethical values. Without the later developments of Ethiopian philosophy, the Fisalgʷos is deprived of its significance; without the Fisalgʷos, Ethiopian philosophy is deprived of its roots.
>
> (Sumner 1985: 8-9)

3 Dialectical pattern

Like ancient and medieval philosophies in Europe, classical Ethiopian philosophical literature has a definite beginning and end. The history of Ethiopian philosophy appears as a dialectical process representing a curve rather than a straight line. This curve is also apparent within *The Life and Maxims of Skendes*, as it moves from the life of Skendes through the first series of maxims and into the second.

The thesis of the dialectical pattern is represented by Ethiopian philosophy of Greek origin. Just as Aristotelian thought was introduced into the Latin medieval world through translations made from the original Greek into Arabic and later into Latin, Greek Platonic and Aristotelian thought was introduced into Ethiopia during the sixteenth century by translations made from the original Greek into Arabic and later into Ethiopic. This stage can be seen in *The Book of the Wise Philosophers* and *The Life and Maxims of Skendes*. The Ethiopian response to this alien influence did not represent a slavish adherence to imported forms, but rather, a creative incorporation. These two works offer adaptations which are so free and original that they are practically the equivalent of an authentic Ethiopian work. The link with Christianity is also apparent in both works.

Zar'a Ya'ecob's work breaks the intimate link with Christianity in general and monasticism in particular, even though his thought remains profoundly theistic. Therefore, he represents an antithetic position to the works of his predecessors. The departure is even greater if one considers that his work is original, the product of personal reflection and not a translation or adaptation from foreign sources.

The treatise written by Wäldä Heywåt after his master's death attempted to synthesize Ya'ecob's ideas with the traditional wisdom literature represented by *The Book of the Wise Philosophers*. His treatise is interspersed with illustrations taken from this sixteenth-century work, although they are more dramatic and picturesque in Wäldä Heywåt than in their original models. The result of this synthesis is no less deep than his master's investigation, but does establish a rapport between the autobiographical insights of Ya'ecob and other practical and educational considerations.

Wäldä Heywåt represents a major contribution to Ethiopian philosophy with respect to ordinary human life. His treatise stresses the value of work, develops a social philosophy with its different practical applications, affirms the equality of all human beings whatever their beliefs, and insists on the beauty of marriage and family life. Yet looked at from the perspective of originality, critical rigour and methodological structure, the synthesis represented by Wäldä Heywåt's treatise is not the summit of the dialectical movement. The line that began with wisdom literature and ascended in an original arc curves back on itself.

CLAUDE SUMNER

References and further reading

Arefayne Hagos (1987) ‘Where is the Spring of African Philosophy?’, *The Ethiopian Herald* 48 (165): 3-6.(The precise meaning of the title of Sumner’s book: *The Source of African Philosophy: The Ethiopian Philosophy of Man*, 1986.)

Littmann, E. (1909) ‘Geschichte der Ethiopischen Literatur’ (History of Ethiopian Literature), in C. Brockelmann (ed.) *Geschichte der christlichen Literaturen des Orients (History of Christian Literature in the East)*, Leipzig: C.F. Zmelangs.(This book on oriental Christian literature contains a chapter on Ethiopian literary history.)

Sumner, C. (1510/22) *Ethiopian Philosophy*, vol. 1, *The Book of the Wise Philosophers*, Addis Ababa: Ministry of Culture and Sports, 1974.(Large sections of text in English translation; analysis in English.)

Sumner, C. (1667; *c*.18th century) *Ethiopian Philosophy*, vol. 2, *The Treatise of Zar’a Ya‘ecob and of Wäldä Heywåt*, Addis Ababa: Addis Ababa University, 1976.(Complete English translation of text and discussion of scholarship.)

Sumner, C. (1667; *c*.18th century) *Ethiopian Philosophy*, vol. 3, *The Treatise of Zar’a Ya‘ecob and of Wäldä Heywåt: an Analysis*, Addis Ababa: Addis Ababa University, 1978.(Thorough analysis of preceding text.)

Sumner, C. (*c*.11th century AD) *Ethiopian Philosophy*, vol. 4, *The Life and Maxims of Skendes*, Addis Ababa: Addis Ababa University, 1978.(Analysis and complete English translation of text.)

Sumner, C. (*c*.5th century AD) *Ethiopian Philosophy*, vol. 5, *the Fisalgʷos*, Addis Ababa: Addis Ababa University, 1982.(Analysis and complete English translation of text.)

Sumner, C. (1985) *Classical Ethiopian Philosophy*, Addis Ababa: Alliance Ethio-Française.(Complete English translation of texts contained in vols 1, 2, 3, 4 and 5. General introduction on Ethiopia and its culture and specific introductions for each of the five basic texts. Recommended as introductory reading.)

Sumner, C. (1986) *The Source of African Philosophy: the Ethiopian Philosophy of Man*, Stuttgart: Franz Steiner Verlag Wiesbaden.(Analysis of the five basic texts. Recommended as a companion to *Classical Ethiopian Philosophy*, 1985.)

Sumner, C. (1988) ‘The Ethiopian Philosophy of Greek Origin’, in S. Uhlig and Bairu Tafla (eds) *Collectanea Aethiopica*, Stuttgart: Franz Steiner Verlag Wiesbaden, 1988.(Comparative study of Greek, Arabic and Ethiopic texts.)

Ethnophilosophy, African

Ethnophilosophy refers to bodies of belief and knowledge that have philosophical relevance and which can be redescribed in terms drawn from academic philosophy, but which have not been consciously formulated as philosophy by philosophers. These bodies of belief and knowledge are manifested in the thoughts and actions of people who share a common culture.

Most of the literature on ethnophilosophy is written about African cultures. Ethnophilosophy's most immediate African antecedents include Leopold Senghor's philosophy of négritude *and the writings of the Belgian missionary to the Congo (later Zaire), Placide Tempels. Ethnophilosophy examines the systems of thought of existing and precolonial African communities in order to determine what can be the ideal forms of 'authentic' African philosophy and praxis in the emerging postcolonial situation. In addition to the pioneering work of Senghor and Tempels, this school is represented in the writings of philosopher Alexis Kagamé (1956) and theologian John Mbiti (1969) among others, many of whom were regarded as Tempels's disciples.*

The central themes of the work of these disciples include assertions that there is a unified 'Bantu philosophy' and that its fundamental categories are manifested in features of language such as grammar, or features of culture such as cosmology and ritual. According to many of these authors writing about Bantu philosophy, the boundaries between self and other are not as rigid as in Western philosophy. Also, interdependence rather than competition is a primary social value and the human and nonhuman world is animated by a 'vital force', which underlies the perception of reality.

1 Defining ethnophilosophy

The literature defining African socialism can be included in the ethnophilosophical approach. This is a body of materials combining Marxist social and economic theories with *négritude*'s politics of difference. Two works by Nyerere ((1968a, 1968b) reveal that these political accounts share with other works of ethnophilosophy the thesis that the central values of Africa are communal rather than individual (see Ethical systems, African). The definition of ethnophilosophy is complex and ambiguous because it arises out of a basic ambiguity in the definition of philosophy itself. The idea of philosophy refers to two related but separable orders of discourse. The first order is located in the forms of talk and action found among the ordinary members of society and elaborated by them in the course of their everyday life and activities. In common parlance, a person, group, historical period or culture could be referred to as having a philosophy when the systematic nature of their moral judgments, the ontological notions that underlie their view of the world, the ways in which they argue and the criteria of evidence and truth that are exhibited in their arguments are under discussion. Usually the philosophy attributed to a person or social entity is not consciously held or espoused, but is part of the tacit set of conventions that make discourse intelligible.

The second order of discourse is dependent on the first order and uses first-order discourse to develop a self-consciously critical order of discourse that is philosophical in a narrower sense and is perceived as having the capacity to correct the flaws of first-order discourse. This is the critical aspect of second-order discourse, but it can also have a speculative side that has not always been acknowledged by the analytic movement which has dominated so much of the history of modern philosophy.

The second problem in defining philosophy as second-order discourse returns to the basic question of defining ethnophilosophy. In a sense all philosophy is culturally shaped and socially determined. It is doubtful that any form of discourse is independent of its cultural and social contexts. If that is so, then all thought, even second-order thought, is shaped by such aspects of the lives of thinkers as their identity, social relations and the culture they share in common with their audiences and interlocutors. From a sociological and anthropological point of view all philosophy is ethnophilosophy. The distinction between first- and second-order thinking, while useful as a means of distinguishing modes of thought used within a single setting, does not help a great deal in specifying the defining features of philosophy believed to transcend the cultural and historical contexts in which they were elaborated.

These observations suggest a second definition for ethnophilosophy which arises more from the practices of philosophers than from the definitions they themselves have offered: ethnophilosophy is a term used to describe

the bodies of belief and knowledge defined by colonized people whose institutions cannot easily be assimilated into models derived from the Western experience in general and the Enlightenment in particular, notably those historical contexts out of which the modern discipline of (Western) philosophy emerged and from which it draws its specificity. From this point of view the ambiguity that created ethnophilosophy has little to do with the people to whom the term is applied and much to do with the social contexts from which a specialism such as philosophy is derived.

2 African ethnophilosophy

Both African socialism and more strictly philosophical works of ethnophilosophy celebrate the subordination of the individual to the community that they argue is central to African culture and philosophy. On the surface this position appears to be radically non-Western and opposed to the emphasis on the adversarial individual in Western thought and culture. Actually it accepts a distinct opposition between individual and community that is not only Western but profoundly rooted in nineteenth-century utilitarian thought. The primary difference between utilitarian thinkers and ethnophilosophers on the subject of the individual and community is that the direction of causation is reversed. In the philosophy of the social sciences, for example, the utilitarian position known as ‘methodological individualism’ asserts that collective and communal forms are reducible to calculating individuals. In ethnophilosophy, by contrast, individuals are reducible to (or explicable in terms of) communities (see Holism and individualism in history and social science).

A number of contradictions are exhibited in ethnophilosophical writings. First and foremost, they remain profoundly descriptive and nonjudgmental, that is noncritical, about African traditions and customs. But they are generally offered in the service of a discourse that is powerfully critical of colonial rule and culture. This is the basis for the reproach aimed at ethnophilosophy by the African philosophers Paulin Hontoundji (1983) and Kwasi Wiredu 1980), that critical discourse cannot exempt one side from the criticism it levels at the other side.

What is even more important, however, is the criticism that ethnophilosophy is oppositional without being radical. On the surface ethnophilosophy is robustly anti-colonial, yet it still accepts the basic categories in terms of which colonial culture defines other cultures and peoples. It attempts to re-evaluate them instead of seeking to criticize the grounds out of which colonial discourse emerges, such as the distinction between culture and civilization, or the primitive or ‘traditional’ and the ‘modern’.

Thus, it appears that ethnophilosophy has two contradictory aspects. It is a critical discourse that defines itself in opposition to colonialism, yet it begins by accepting the colonial categories of traditional and modern. The most significant difference between the original colonial categories and their use in ethnophilosophy is that instead of treating them as diametrically opposed in the colonial fashion, ethnophilosophy tries to merge them by re-appraising indigenous values as worthy of attention and then accordingly discovering the traditional in the modern.

Ethnophilosophers tend to be scholars trained in the West who work on materials derived from outside the cultural contexts in which they were trained. The context and structure of their work engages them in an activity that is culturally hybridized. Since many of them are Africans, the hybridity also works in the other direction, from Africa to the West. Ethnophilosophers have occasionally used the hybrid nature of their cultural productions to create knowledge that is itself hybrid by arguing that characteristics of Western intellectual history can be found in African traditional thought. For example Alexis Kagamé((1956) argues that African thought utilizes Aristotelian and Thomistic elements. Placide Tempels(1959), arguably the father of ethnohistory, even founded a Catholic religious movement in the Congo (later Zaire) based on the convergence of Bantu philosophy and Christianity.

The literature on ethnophilosophy does not usually combine discussions of ethnophilosophy proper with accounts of the writings on African socialism and Afrocentricity. The notable exception is Anthony Appiah’s *In My Father’s House* (1982) which examines nationalist themes and assertions about the person found in pan-Africanism, the counterhegemonic discourses on race and the literature on African philosophy (see Pan-Africanism). Appiah’s work represents a rare attempt to take a stance which acknowledges the ways in which philosophical discourse emerges from cultural and historical conditions and is manifested in a range of materials that include but are not limited to philosophy. By avoiding the increasingly sterile debate about whether African discourse has a second-order dimension, Appiah is able to discuss philosophical aspects of a broad range of

materials in a way that has escaped the grasp of philosophers more concerned with the boundaries of their discipline.

Perhaps the most notable characteristic of the ethnophilosophy school was its characterization of philosophy as a kind of collective narrative. Ethnophilosophers treated African philosophy as a narrative whose content is revealed through various codes, such as myth, symbolic systems and religious and ordinary language. In this sense, although African philosophy was represented as an innate form of indigenous expression, it could only be recovered and re-evaluated from a hybridized postcolonial present. This is another example of ethnophilosophers' ambivalent mediation between two points, the 'traditional' and the 'modern'.

3 Ethnophilosophy and the politics of culture

The political implications of ethnophilosophy are also displayed in the idea that traditional knowledge is or was collectively produced and appropriated. This central theme of ethnophilosophy suggests that individuals cannot be free; that in a society where knowledge is a collective product, cultural criticism is not possible. This in turn raises fundamental questions about personhood, agency and the possibility of change.

The idea that African thought is collective and unchanging has been accepted by some of the critics of ethnophilosophy. Hountondji 1983) and Wiredu (1980) argue that the capacity for change depends on three elements of Western Enlightenment inheritance lacking in traditional societies: individual freedom, abstract theory and openness to alternative theories and interpretation. In their view traditional philosophy appears to reject these features of thought. Wiredu (1980) identifies these with science and sees hope for this redemption in the application of analytic practice as people seek new methods and solutions to old problems. Hountondji (1983) proposes that this redemption takes as its foundation Althusserian neo-Marxist notions which he believes govern the mobility of knowledge through history. Both regard the individual as the agent of change through social and cultural criticism. In other words, Wiredu and Hountondji defend the colonial and postcolonial as the new spatial and temporal realities not to be ignored by Africa.

This countercritique ranks the West above the traditional as they separate the past from the present. This is because the influence of the past on the present should be minimized to avoid anachronism. In effect, it offers an alternative representation of the hybrid postcolonial social and cultural condition.

The critique of ethnophilosophy began around the 1960s after the first works of ethnophilosophy were published. Among the first critics of ethnophilosophy was Franz Crahay (1965). In his paper 'Le Décollage Conceptuel: Conditions d'une Philosophie bantoue' (Conceptual Launch: the Terms of a Bantu Philosophy) he argued that the colonial distinction between the traditional and the modern is analogous to the distinction between 'constructing myth' and 'practising philosophy'. This distinction is in turn similar to the metaphysical distinction between form and matter. Although form and matter make complementary contributions to the identity of things, they none the less remain conceptually distinct in nature and function. According to this perspective philosophy like form is of the mind. It deals with those elements of thought in which experience as event and practice is presented in and to our minds. It deals with ideas, related to and distinct from the particulars from which they derive. Though it is a human practice, philosophy diverges from other human practices such as customary ways of living and traditional (or any) group behavioural patterns.

Like matter, myth, tradition, custom and mores are glued to the sensual mode of experience. Their language remains unabstracted from the metaphors and experiences of everyday life; they are sociologically immediate and concrete. In this sense Crahay (1965) argues that while philosophy frees itself from its conceptually limiting fixation with sociological conditions by 'taking-off' to a free, or universal conceptual level, ethnophilosophy remains trapped in the closed confines of sociological structures and relativism like anthropology. In anthropology the argument that African systems of thought have been closed was developed by Robin Horton (1967). He argues that the nature of African social structures prevents African thought treating society and nature as abstract rather than personalized. In his view social relations are the idiom for thinking about natural elements, not vice versa.

The difficulty with the critics of ethnophilosophy, as well as with the alternatives developed by scholars such as Horton, is that by reserving criticism for themselves, the authors of this literature also reserve the role of criticism for the Western trained philosopher. This position implicitly supposes that traditional African society is still collective and so cannot regenerate itself by means of social and cultural criticism. From a political point of view

the strategy of constructing a subject who can only be represented by an other is characteristic of nationalist thought and assertion, where some political actor or writer sets themselves up as the voice and representative of an emerging collectivity which can not speak for itself. The parallel between ethnophilosophy and nationalist discourse is not accidental. The guiding assumption of ethnophilosophy, that philosophical notions can be held unconsciously, requires a conscious agent to represent and lead the mass of ordinary Africans to a different world and a better life. This political project shared by ethnophilosophy and many of its critics is unmistakably nationalist and modernist. It is organized by ideas about progress and divides the social world into two classes, citizens who are critically self-aware agents and subjects who have yet to acquire the cultural apparatus that would enable them to become citizens. The debate about whether Africans and others possess second-order modes of thought also exhibits modernist premises about critical self-awareness, but is more concerned with the relationship of philosophy (as an emergent profession) with society than with how cultures should be classified. Ethnophilosophy brings together two usually separate cultural arenas: professional discourse and cultural discourse. The result sheds light on unresolved issues in the definition of philosophy, as well as problems of how cultures are classified and ranked.

See also: African philosophy, Anglophone; African philosophy, Francophone §5; Akan philosophical psychology; Yoruba epistemology

IVAN KARP
D.A. MASOLO

References and further reading

Abraham, W.E. (1962) *The Mind of Africa*, Chicago, IL: University of Chicago Press.(The first book on African philosophy, containing a sophisticated account of ethnophilosophy and a critique of colonialism.)

Appiah, A.K. (1982) *In My Father's House: Africa in the Philosophy of Culture*, London: Methuen; Oxford: Oxford University Press, 1992.(A collection of essays on African culture, philosophy and pan-Africanism which uses analytical philosophy to examine changing definitions of race, problems of ethnophilosophy and the ideology of culture in Africa.)

Asante, K.M. and Kariamu, W.A. (1990) *African Culture: The Rhythms of Unity*, Trenton, NJ: New World Press.(One of many books on Afrocentricity. It is a broad-ranging exposition of Asante, a major figure, whose thinking exhibits many of the features of the literature in ethnophilosophy.)

Crahay, F. (1965) 'Le Décollage Conceptuel: Conditions d'une Philosophie bantoue' (Conceptual Launch: the Terms of a Bantu Philosophy), in *Diogène* 52: 61-84.(The classic critique of ethnophilosophy discussed in the text in §3)

Fløistad, G. (1987) *Contemporary Philosophy*, vol. 5: *African Philosophy*, Boston, MA: Kluwer.(A collection of essays by major figures writing about African philosophy. Perhaps the best text in which to survey the range of work on the subject.)

Horton, R. (1967) 'African Traditional Thought and Western Science', *Africa* 37 (1-2): 50-71, 155-87.(The most important modern essay by an anthropologist who compares the structure of African traditional thought to the structure of scientific thinking.)

Horton, R. (1982) 'Tradition and Modernity Revisited', in M. Hollis and S. Lukes (eds) *Rationality and Relativism*, Oxford: Basil Blackwell.(Influential articles by an anthropologist who examines African worldviews as forms of intellectual activity and compares them with the worldview of science. He finds many parallels and some significant differences, notably in the attitude to alternative explanations of the world found in each system.)

Hountondji, P.J. (1983) *African Philosophy: Myth and Reality*, Bloomington, IN: Indiana University Press; 2nd edn, 1996.(A major work by an African philosopher trained in France. Hountondji writes in the tradition of Bachelard and Althusser and seeks to find the space in which to situate an 'epistemological break' in African worldviews.)

Kagamé, A. (1956) *La philosophie bantu-rwandaise de l'être (The Bantu-Rwandan Philosophy of Being)*, Brussels: Académie Royale des Sciences Coloniales.(Perhaps the most detailed and analytical study of an African system of thought from an ethnophilosophical point of view.)

Masolo, D.A. (1994) *African Philosophy in Search of Identity*, Bloomington, IN: Indiana University Press.(A thorough study of ethnophilosophy and its African critics. The book takes account of relevant anthropological literature.)

Mbiti, J. (1969) *African Religions and Philosophy*, London: Heinemann, 2nd edn, 1990.(Ethnophilosophy is

combined with Christian theology in this influential book.)
Mudimbe, V.Y. (1988) *The Invention of Africa: Gnosis, Philosophy, and the Order of Knowledge*, Bloomington, IN: Indiana University Press.(A major survey of what Mudimbe calls the ‘colonial archive’ of images about Africa. Special attention is paid to the literature on African philosophy.)
Nyerere, J.K. (1968a) *Ujamaa: Essays on Socialism*, New York: Oxford University Press.(The classic exposition of African socialism, which examines African society to find socialist priciples.)
Nyerere, J.K. (1968b) *Ujamaa: The basis of African Socialism*, Dar es Salaam: Oxford University Press.(The most influential writer on African socialism.)
Pearce, C. (1992) ‘African Philosophy and the Sociological Thesis’, in *Philosophy of the Social Sciences* 22 (4): 440-60.(An important review of African philosophy.)
Senghor, L.S. (1964) *Liberté 1: négritude et humanisme (Liberty 1: négritude and humanism)*, Paris: Editions de Seuil.(An explanation of *négritude* by Senegal’s president and distinguished poet.)
Tempels, P. (1959) *Bantu Philosophy*, Paris: Présence Africaine.(The founding text of ethnophilosophy. Argues that Bantu philosophy is expressed in collective forms and that it perceives the world as animated by a vital force.)
Wiredu, K. (1980) *Philosophy and An African Culture*, Cambridge: Cambridge University Press.(A rigorous examination of the ethnophilosophical point of view from the twin perspectives of analytical philosophy and pragmatism.)

Eudaimonia

The literal sense of the Greek word eudaimonia *is 'having a good guardian spirit': that is, the state of having an objectively desirable life, universally agreed by ancient philosophical theory and popular thought to be the supreme human good. This objective character distinguishes it from the modern concept of happiness: a subjectively satisfactory life. Much ancient theory concerns the question of what constitutes the good life: for example, whether virtue is sufficient for it, as Socrates and the Stoics held, or whether external goods are also necessary, as Aristotle maintained. Immoralists such as Thrasymachus (in Plato's* Republic*) sought to discredit morality by arguing that it prevents the achievement of* eudaimonia, *while its defenders (including Plato) argued that it is necessary and/or sufficient for* eudaimonia. *The primacy of* eudaimonia *does not, however, imply either egoism (since altruism may itself be a constituent of the good life), or consequentialism (since the good life need not be specifiable independently of the moral life). The gulf between 'eudaimonistic' and 'Kantian' theories is therefore narrower than is generally thought.*

1 Conventional morality and Aristotle

The word *eudaimonia* and the corresponding adjective *eudaimōn* are derived from *eu* 'good, well' and *daimōn* 'spirit', giving the literal sense 'having a good guardian spirit', hence being blessed, having the life of one who enjoys divine favour. To possess *eudaimonia* is to have a life which is objectively desirable, and thereby to have achieved the most worthwhile of conditions available to humans. It thus overlaps substantially with the modern concept of happiness (see Happiness), where that is conceived, not as a (possibly short-lived) episode of feeling, but as a condition characterizing a whole life, or a substantial part of it, as when one wishes a newly married couple happiness in their life together. But even when thus conceived, happiness in the modern sense is determined primarily by one's subjective attitude to one's life; broadly, to be happy is to be content with one's life, whereas contentment with one's life, although arguably necessary for *eudaimonia*, is not sufficient for it. According to Aristotle (§21) (*Nicomachean Ethics*, 1095a15-22) everyone agrees (1) that *eudaimonia* is the highest human good and (2) that the word means 'doing and living well', but there are substantive disagreements on what kind of life counts as doing and living well. Some kinds, such as the life devoted to bodily pleasure, are disqualified not on the ground that people who live like that are always dissatisfied, but because they fail to satisfy objective requirements for a worthwhile life, in particular that it should be an excellent realization of specifically human potentiality. Again, Aristotle asserts (1100a1-2) that a child is not *eudaimōn*, meaning not that no one has a happy childhood, but that the life of a child is not one in which human potential is realized to its highest bent.

In popular usage the term has strong connotations of material prosperity, traditionally regarded as one of the signal marks of divine favour. This conception is in tension with a more specifically ethical one, according to which the condition of the agent's personality or character is at least the primary determinant of living well. The latter conception, found already in Democritus §§15-16 (fr. 171), finds its strongest expression in the Socratic thesis, later adopted as the cardinal tenet of Stoicism, that virtue is the only intrinsic human good and wickedness the only intrinsic evil (see Stoicism §15-16). Aristotle's treatment (*Nicomachean Ethics* I 7-10) is a characteristic exercise in reconciliation of these standpoints. On the one hand *eudaimonia* must be comprehensive and self-sufficient, requirements which seem to imply that it includes the gifts of fortune; on the other the achievement of the good life must be in our power. Aristotle's middle way is the thesis that excellence or virtue (see Aretē) of character and intellect is (1) indispensable for the good life, (2) although not totally invulnerable, less subject to the vicissitudes of fortune than anything else. Hence the achievement of excellence, although not strictly sufficient for *eudaimonia*, as Socrates arguably and the Stoics certainly claimed, comes as close to guaranteeing it as the limitations of human life permit.

2 Attacks on conventional morality

Since *eudaimonia* was universally agreed to be the good, critics of conventional morality such as Antiphon, Callicles (in Plato's *Gorgias*) and Thrasymachus (in Plato's *Republic*) sought to devalue morality by arguing that it hinders or prevents the achievement of *eudaimonia*, while defenders of morality sought to show that it is necessary and/or sufficient for *eudaimonia*. The attacks assume that *eudaimonia* is to be achieved through the gratification of desire, whereas morality requires resistance to or suppression of desire; the defences take different

forms. One line of defence, seen in Democritus, Protagoras' myth (in Plato's *Protagoras*) and Glaucon's social contract account of morality in *Republic* II, stresses the benefits which the individual derives from social institutions, for the existence of which morality is necessary. In the *Republic* (see Plato §14) Plato takes the more radical step of internalizing social morality; the 'individual' is in fact a community in miniature, and the requirements of morality are therefore reduced to the proper organization of that community: that is, to the harmonious organization of the 'individual' personality. Plato, and Greek philosophers generally, take the primacy of individual *eudaimonia* pretty well for granted; the possibility that morality might require the sacrifice of the agent's *eudaimonia* for the good of others, or of society, is hardly canvassed.

This raises the difficult problem of the extent to which Greek ethical theory allowed for altruism. Certainly, all the main writers and schools - Plato, Aristotle, Epicurus and the Stoics - stress the importance of one form or another of social life, whether that of the city, the Epicurean community or the cosmos as a whole, insisting that interpersonal ties of one or other kind are themselves essential constituents of individual *eudaimonia* (see Aristotle §§27-8; Epicureanism §11; Stoicism §14). If the most worthwhile life is the most excellent one, and the most excellent thing to do is to sacrifice one's life for one's friend, as Aristotle recommends, or to take one's turn ruling instead of devoting oneself full-time to metaphysics, as Plato requires of his guardians in the *Republic*, then in sacrificing oneself one does not sacrifice, but perfects, one's *eudaimonia*. The primacy of *eudaimonia* is thus preserved, but at the cost of moralizing the concept of *eudaimonia* itself; the good life includes the moral (whether the just or the altruistic) life as an essential component. This in turn threatens the project of defending morality by showing it to be necessary and/or sufficient for the independently acknowledged good of *eudaimonia*.

3 Eudaimonistic and Kantian ethics

The primacy of *eudaimonia* in Greek ethical thought standardly leads to its being characterized as 'teleological', in opposition to 'deontological' theories of a Kantian type (see Teleological ethics; Deontological ethics), in which the fundamental concept is that of obligation binding on all rational beings (see Kant §9). This contrast is superficial. As we have seen, 'eudaimonistic' theories need not be consequentialist, since the good life need not be specifiable independently of the moral life (see Consequentialism). Furthermore, the tendency of much Greek thought, notably that of Plato, Aristotle and the Stoics, to identify the real self with the rational self tends towards the identification of the achievement of *eudaimonia* with conformity to demands of reason. The Stoic community of all rational nature is the forerunner of the Kantian kingdom of ends.

See also: Cyrenaics §§3-4; Wisdom

C.C.W. TAYLOR

References and further reading

Annas, J. (1993) *The Morality of Happiness*, Oxford and New York: Oxford University Press.(Aristotle and later Greek philosophers on happiness and the good.)

Aristotle (*c.* mid 4th century BC) *Nicomachean Ethics*, trans. T. Irwin, Indianapolis, IN: Hackett, 1985.(Includes notes.)

Irwin, T. (1986) 'Stoic and Aristotelian Conceptions of Happiness', in M. Schofield and G. Striker (eds) *The Norms of Nature*, Cambridge: Cambridge University Press, 205-44.(An illuminating comparison of the two conceptions.)

Irwin, T. (1995) *Plato's Ethics*, Oxford and New York: Oxford University Press. (Comprehensive treatment of the ethical theories of the early and middle dialogues, with brief discussion of the later dialogues.)

Kenny, A. (1992) *Aristotle on the Perfect Life*, Oxford: Clarendon Press.(Aristotle's treatments of virtue and the good life.)

Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, vol. 1, Cambridge: Cambridge University Press.(Translates the principal sources for Stoicism and Epicureanism; includes commentary.)

Plato (*c.*390s-380s BC) *Gorgias*, trans. T. Irwin, Oxford: Clarendon Press, 1979.(Includes notes.)

Long, A.A. and Sedley, D.N. (*c.*380s BC) *Protagoras*, trans. C.C.W. Taylor, Oxford: Clarendon Press, 2nd edn, 1991.(Includes notes.)

Long, A.A. and Sedley, D.N. (*c.*370s BC) *Republic*, trans. R. Waterfield, Oxford and New York: Oxford University Press, 1994.(See especially Books I, IV, VIII and IX.)

White, S.A. (1992) *Sovereign Virtue*, Stanford, CA: Stanford University Press.(Aristotle on the relation between happiness and prosperity.)

Eudoxus (*c*.390-*c*.340 BC)

Eudoxus of Cnidos was a Greek mathematician with wide-ranging philosophical and scientific interests. He was known in antiquity for his mathematical astronomy and his philosophical hedonism, and placed, if loosely, within the Pythagorean tradition. More recently, debate about the correspondence between mathematical models produced by astronomers and the physical motions of the astronomical bodies themselves has focused especially on Eudoxus, since he is generally regarded as the first to have used a mathematical model in astronomy.

Very little about Eudoxus' life and work is known with any certainty. Diogenes Laertius (VIII 86-91), in his biography of Eudoxus, reports that he was called 'Endoxus' ('illustrious') because of his brilliant reputation. None of his works survives, only accounts and fragments in other ancient authors. Even the long-standing belief that he was a leading member of Plato's Academy (see Academy) is now in doubt. However, there is evidence that he founded a school at Cyzicus, on the Hellespont, which may have survived into the third century BC. Tantalizingly little is known about his relationships to such important contemporaries as his supposed teacher Archytas, Plato and Aristotle. Anecdotes relate his invention of a type of sundial, and his introduction of the practice of arranging furniture in a semi-circle to accommodate more people.

Eudoxus is mentioned, together with Anaxagoras, by Aristotle in his arguments against the theory of Platonic Forms (see Aristotle §15). The reference is short and ambiguous; consequently scholars have debated whether Eudoxus actually had a 'theory' of Forms. Alexander of Aphrodisias, in his commentary on Aristotle, provides a lengthier discussion of what he understands as Eudoxus' views, including the idea that forms are mixed together with and are 'homoeomerous' (of like parts) to those things they inform. Eudoxus can be understood as suggesting that forms are immanent (rather than transcendent), for they are physical ingredients in entities (Dancy 1991).

Eudoxus is particularly celebrated for his geometric model of planetary motion. According to the ancient story, he was the first to answer Plato's challenge to account for the apparently irregular motions of the celestial bodies by using only uniform, circular and regular motions in order to describe the phenomena or 'save the appearances'. He devised a separate system of nested homocentric spheres to describe the apparent motions of each of the heavenly bodies (reported by Aristotle in *Metaphysics* XII 8). Writing much later, Simplicius explained that according to Eudoxus the path of a planet describes a figure called a 'hippopede' or horse-fetter. Eudoxus was able to account for various phenomena and 'save the appearances' in an approximate way, although not all phenomena could be explained, including the changing brightness of planets. Aristotle mentions that Callipus (of Cyzicus) adopted Eudoxus' approach but added to the number of spheres, in order to explain the phenomena better. Aristotle objected to the number of spheres of both Eudoxus and Callipus, and offered his own improvements to the system. Although the ancient descriptions of Eudoxus' astronomy are brief, several nineteenth century scholars reconstructed in detail the geometry thought to underlie his solution.

Questions regarding the application of mathematical methods to the explanation of natural phenomena are of great interest to historians and philosophers of science, and Eudoxus' work often figures importantly in their discussions. Eudoxus played a key role in the history of science, according to Pierre Duhem, who was himself a proponent of an instrumentalist view of science and, as a physicist, philosopher and historian of science, was interested in the relationship between mathematics and physics (Duhem 1908) (see Scientific realism and antirealism §1; Ptolemy). According to instrumentalists, theories permit predictions which correspond to observations, but do not necessarily describe physical reality. Many, including Duhem, have assumed that Eudoxus did not believe that his geometric model had any physical reality. This would qualify Eudoxus as the first 'instrumentalist' in the history of science. However, the paucity of our knowledge of Eudoxus' work makes any statement regarding the intention underlying his astronomical solution purely speculative.

Within the broader context of early Greek astronomy, it has been suggested that Eudoxus' originality lay in providing the basis of a geometric explanation useful in accounts of problems of practical interest, namely the risings and settings of stars, in geographical studies, and in justifying a more mathematically sophisticated sundial (Goldstein and Bowen 1983). On this view, Eudoxus' work represents a synthesis of two existing activities, calendar-making and cosmological theorizing, and his introduction of a geometric model to explain celestial motions inaugurated a new phase in astronomy.

It is tempting to look for some connection between Eudoxus' astronomical work and his ethical philosophy. Both the science of music and cosmological speculation may have had an important influence on him: the Pythagoreans and Plato were already pointing to a moral order manifested in the circular motions of the heavenly bodies (see Pythagoreanism §§2-3). Furthermore, there is evidence that theological concerns motivated the followers of Eudoxus in his school at Cyzicus (Sedley 1976).

As a hedonist, Eudoxus thought that pleasure was the good and the goal of life (see Hedonism). Aristotle (*Nicomachean Ethics* X 2) explains Eudoxus' argument, which was based on the observation that all things aim towards pleasure. Because that which is the object of choice is excellent and that which is most the object of choice is most excellent, the fact that all things move towards the same goal indicates that for all things it is the good. Aristotle observes that Eudoxus' views regarding pleasure were taken seriously because of his good character, rather than on their own merits. Because he was regarded as remarkably self-controlled, it was thought that Eudoxus presented his views 'not simply as a friend of pleasure, but because the facts really were so'.

LIBA TAUB

References and further reading

Aristotle (*c*. mid 4th century BC) *Metaphysics*, trans. H. Tredennick, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933, 2 vols.(Parallel Greek text and English translation; Eudoxus' astronomy is covered in XII 8; his theory of Forms is described in I 9.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. H. Rackham, Cambridge, MA.: Loeb Classical Library, Harvard University Press and London: Heinemann, 1982.(Parallel Greek text and English translation; X 2 discusses Eudoxus' hedonism.)

Dancy, R.M. (1991) *Two Studies in the Early Academy*, Albany, NY: State University of New York.(Considers Eudoxus' theory of Forms.)

Diogenes Laertius (*c*. 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(A life of Eudoxus can be found at VIII 86-91.)

Dreyer, J.L.E. (1906) *History of the Planetary Systems from Thales to Kepler*, Cambridge: Cambridge University Press; 2nd edn, *A History of Astronomy from Thales to Kepler*, New York: Dover, 1953.(Gives a detailed account of Schiaparelli's reconstruction of Exodus' possible geometry.)

Duhem, P. (1908) *ΣΩΖΕΙΝ ΤΑ ΦΑΙΝΟΜΕΝΑ, Essai sur la notion de théorie physique, de Platon à Galilée*, Paris: A. Hermann; repr. Paris: Vrin, 1990; trans. E. Dolan and C. Maschier *To Save the Phenomena: An Essay on the Idea of Physical Theory from Plato to Galileo*, Chicago, IL: University of Chicago Press, 1969.(Influential interpretation of the relationship between astronomy and physics.)

Goldstein, B. and Bowen, A.C. (1983) 'A New View of Early Greek Astronomy', *Isis* 74: 330-40.(A reappraisal of the role of Eudoxus within the Greek astronomical tradition.)

Gosling, J.C.B. and Taylor, C.C.W. (1982) *The Greeks on Pleasure*, Oxford: Clarendon Press.(Chapter 14 discusses Eudoxus' hedonism.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols. (Volume 5 pages 447-57 covers Eudoxus.)

Lasserre, F. (1966) *Die Fragmente des Eudoxos von Knidos*, Berlin: de Gruyter.(Presents the evidence regarding Eudoxus.)

Lloyd, G.E.R. (1978) 'Saving the Appearances', *Classical Quarterly* 28: 202-22; repr. in *Methods and Problems in Greek Science*, Cambridge: Cambridge University Press, 1991.(Expresses important reservations about Duhem's interpretation.)

North, J. (1994) *The Fontana History of Astronomy and Cosmology*, London: Fontana, 1994. (A useful description of Eudoxus' geometric model.)

Sedley, D. (1976) 'Epicurus and the Mathematicians of Cyzicus', *Cronache Ercolanesi* 6: 23-54.(Shows that Eudoxus' students built physical models of his system.)

Simplicius, (6th century AD) *In Aristotelis de caelo commentaris*, ed. J.L. Heiberg, *Commentaria in Aristotelem Graeca VII*, Berlin: Reimer, 1894, 493-507; trans. G.V. Sciaparelli, *Scritti sulla storia della astronomia antica*, Bologna: Zanichelli, 3 vols, 1925-7, II, 95-112; trans. of major sections, T. Heath, *Aristarchus of Samos: The Ancient Copernicus*, New York: Dover, 1981, 201-2, 213, 221-3.(A late ancient account of Eudoxus' astronomy.)

Eurasian movement

The Eurasian movement was a creation of émigré Russian intellectuals following the First World War and the October Revolution. The ideology of Eurasianism was formally proclaimed in 1921. It obtained considerable development, prominence, distinction and notoriety in the two following decades, essentially running its course by the time of the Second World War. Eurasianism attracts attention because of the novelty and originality of its central argument, proposing that the Russian Empire and later the Soviet Union constituted an independent organic entity separate from both Europe and Asia, called Eurasia (it was the complete separation from Europe that represented an explosive novelty), and because of the intellectual variety and abilities, at times brilliance, of its proponents. Also, as an extraordinary phenomenon in Russian intellectual history, it demands explanation.

1 Eurasianism

Eurasianism emerged quite explicitly in 1921 when four young Russian intellectuals published a collective volume, *Iskhod k vostoku (Exodus to the East)*. The four were Prince Nikolai Sergeevich Trubetskoi (1890-1938), later to earn fame as a linguistic scholar, Pëtr Nikolaevich Savitskii (1894-1968), an economist-geographer and specialist in many subjects, Pëtr Petrovich Suvchinskii (1892-1985), a gifted music critic and many-sided intellectual, and Georgii Vasilievich Florovskii (1893-1979), a theologian, intellectual historian, and a man of numerous interests with a remarkable breadth of knowledge. The anthology consisted of an introduction and ten essays.

The introduction spoke of a world cataclysm, of a catastrophic change of scenes, of a new age, of the dying of the West and of the imminent rise of the East. It concluded: ‘Russians and those who belong to the peoples of "the Russian world" are neither Europeans nor Asiatics. Merging with the native element of culture and life which surrounds us, we are not ashamed to declare ourselves *Eurasians*’ (1921: 7).

In the essays which followed Florovskii and Suvchinskii concentrated on a fundamental and sweeping criticism of the hopelessly rationalistic, war-devastated, and moribund West, as well as on the catastrophe of Russian history and society, which eventuated in the unparalleled disaster of communist revolution; Savitskii and Trubetskoi, however, paid major attention to the new redeeming concept of Eurasia. In an essay on ’The Migration of Culture’ Savitskii outlined how in world history cultural centres and leadership moved with each millennium to colder and colder climates; if the migration continued, the next step will transfer the world leadership to Canada and ’northern Minnesota’ on the one hand, and on the other to ’parts of northern and central Russia together with all of eastern Russia, both eastern European Russia and Siberia’. And in another essay, ’Continent-Ocean (Russia and the World Market)’, the economist and geographer insisted that because of its landlocked position and huge bulk Russia had to minimize recourse to oceanic trade routes and devote itself instead to the entirely feasible rich and varied development of its own continental economy. Trubetskoi’s essay, ’The Upper and Lower Layers of Russian Culture (The Ethnic Base of Russian Culture)’, sketched in effect cultural Eurasia: in contrast to the superficially and unsuccessfully Westernized upper classes, Russian masses and their culture were in no sense simply European, because proto-Slav dialects (and hence the Russian language) occupied a middle geographic position between the proto-Indo-Iranian and proto-Western-Indo-European ones, closer to the former than to the latter in essential characteristics; because Great Russian folk songs utilized the ’five-tone’ or ’Indo-Chinese’ scale; because Russian folk dance, in contrast to Western dance, was not based on a couple, a man and a woman, dancing together; because Russian material culture indicated the closest possible association with Ugrofinnic peoples; because Russians had Ugrofinnic and Turkic as well as Slav blood; because they were inclined to oriental contemplation and devotion to ritual, but also to *udal’*, extravagant daring or audacity, ’a purely steppe virtue, which the Turkic peoples understand, but neither the Romanogermans nor the Slavs can’, and so on. The two Eurasias, Savitskii’s geopolitical and Trubetskoi’s cultural, became partners but also independent streams in the new ideology.

The first Eurasian symposium was followed by others, the seventh and last coming out in 1931. Also, twelve volumes of the *Evraziiskaia khronika (Eurasian Chronicle)* appeared between 1925 and 1937, and they were joined by numerous individual publications by participants in the movement. Indeed in the 1920s and 1930s Eurasianism acquired some prominence and attracted much attention among Russian exiles in Europe. Eurasian

publications were supplemented by ’seminars’, public lectures, formal debates and private disputations. Although Florovskii, one of the founders, rather quickly abandoned Eurasianism and in 1928 published a crushing critique of it, other émigré intellectuals, a few of them of great prominence, joined the movement. Thus Roman Osipovich Jakobson (1896-1982), another linguist of the first rank, followed Trubetskoi’s lead to make some remarkable discoveries about languages in Eurasia, while Georgii Vladimirovich Vernadsky (1887-1973) became the prolific Eurasian historian, and, because he taught Russian history at Yale and published many books in English, the greatest propagandist of Eurasianism to the world outside Russian émigré circles. The movement also displayed a certain political bent, even aspiring to replace communism, regarded as a rival and undesirable ideocracy, in the government of Russia-Eurasia. Still, always confined to a small minority and essentially to one particular generation of exiles, the Eurasian movement ended with the new perturbations of the Second World War and the decline and death of the generation in question.

2 The historical position of Eurasianism

Eurasianism is both easy and difficult to understand and explain. It certainly belongs with other European doctrines of bitterness, suffering and despair, as well as at times apocalyptic hopes, which proliferated after the First World War and the disasters which attended and followed it. In the Russian case the October Revolution, civil war and famine could be considered even more horrible than the great War itself, and they all found rich response in Eurasianism. Still another major historical development, described by phrases like the rise of colonial peoples, the decline of imperialism, or the gradual loss by the so-called White race of its dominant global position, also found powerful response. In fact, Trubetskoi had published already in 1920, the year preceding the inauguration of Eurasianism, a brief book *Europe and Mankind*, which was a relentless indictment of imperialism and a declaration of essential Russian solidarity with colonial peoples.

The difficulty resides in the positive part of the Eurasian teaching, in the concept of Eurasia itself. More exactly, of the two main versions of that concept, Savitskii’s geopolitical Eurasia found considerable understanding and some approval. But Trubetskoi’s cultural Eurasia, the alleged Eurasian identity of the Russians, exploded among the émigrés like a bombshell. No Russian intellectual prior to the Eurasian movement had denied a fundamental Russian affiliation with Europe. Such severe critics of the West as the Slavophiles and Dostoevskii emphasized the very point that the West was guilty of not treating Russians like brothers. Nikolai Danilevskii, it is true, rejected the West, but his entire allegiance went to Slavdom and the coming Slav age in Europe (see Pan-Slavism §3). The new identity was so revolutionary and so baffling that most critics totally failed to understand, let alone accept it. Where did this identity come from? A few disparate suggestions appear possible. As already indicated, the Eurasians reacted to the new phenomenon of the rise of the colonial peoples. Moreover, without doubting the sincerity of Eurasians, it seems clear that if Eurasia, organic and glorious, were also real, the Russian or the Soviet Empire, rebaptized Eurasia, would have perdured in the age when other empires crumbled. The Russian intellectual climate also made its contributions. Russian scholarship was in the process of discovering ever more about the character and the role of non-Slav peoples, whether Finnic tribes of the north or steppe inhabitants to the southeast, in Russian history and culture. Beyond that, on the eve of the appearance of Eurasianism a few poets, most notably Aleksandr Blok in a poem entitled ’The Scythians’ written 30 January 1918, began to identify the Russians themselves as such an intermediate people between Europe and Asia. Finally, the personal circumstances of the Eurasians deserve consideration. They were gifted young intellectuals ready to participate in the cultural leadership of Russia, when Russia disappeared (and they found themselves permanently in exile). Eurasianism can be regarded as a desperate bid to compensate for that loss.

Essentially fantastic and certainly abortive, Eurasianism deserves attention because of its intellectual variety and richness and also as a sign of the times. Understandably it has attracted new interest with the collapse of the Soviet Union and the Russian search for identity.

NICHOLAS V. RIASANOVSKY

List of works

Florovskii, G., Savitskii, P., Suvchinskii, P and Trubetskoi, N. (1921) *Iskhod k vostoku: Predchustviia i sversheniia. Utverzhdenie evraziitsev*, *Sofia: Tipografiia ’Balkan’*; translated in I. Vinkovetsky, N.V. Riasanovsky and C. Schlacks, Jr (eds), *Exodus to the East: Foreboding and Events: An Affirmation of the Eurasians*, Marina del Rey, CA: Charles Schlacks, Jr, in preparation.(This and other Eurasian publications are

rare and usually exist only in Russian. At last, a scholarly translation is scheduled to appear.)
Trubetskoi, N.S. [Trubetzkoy] (1991) *The Legacy of Genghiz Khan and Other Essays on Russian Identity*, ed. and with postscript by A. Liberman, preface by V.V. Ivanov, Ann Arbor, MI: Michigan Slavic Publication.
Vernadsky, G. (1929) *A History of Russia*, New Haven, CT: Yale University Press.
Vernadsky, G. (1953) *The Mongols and Russia*, New Haven, CT: Yale University Press.

References and further reading

Böss, O. (1961) *Die Lehre der Eurasier: Ein Beitrag zur Russischen Ideengeschichte des 20. Jahrhunderts (The Teaching of the Eurasians: A Contribution to the Russian History of Ideas of the Twentieth Century),* Wiesbaden: Otto Harrassowitz.(For a bibliography see 125-30.)
Riasanovsky, N.V. (1964) 'Prince N.S. Trubetskoi's "Europe and Mankind"', in *Jahrbücher für Geschichte Osteuropas, Neue Folge* 12 (2): 207-20.
Riasanovsky, N.V. (1967) 'The Emergence of Eurasianism' in *California Slavic Studies*, vol. 4, eds N.V. Riasanovsky and G. Struve, 39-72.
Riasanovsky, N.V. (1972) 'Asia through Russian Eyes' in *Russia and Asia: Essays on the Influence of Russia on the Asian Peoples*, ed. W. Vucinich, Stanford, CA: Hoover Institution Press, 3-29, 369-75.

Eusebius (*c.* AD 264-*c.*339)

Eusebius, Bishop of Caesarea in Palestine from c.*314, was the foremost Christian scholar of his age and wrote extensively on history, geography, chronology, apologetics and philosophical and biblical theology. He is best known for his pioneering History of the Church, but philosophers may prefer to consult his Preparation of the Gospel, which argues that basic Christian doctrines had been foreshadowed by well-known philosophers and preserves valuable extracts from writers whose works have otherwise been lost. In his own philosophical ideas, Eusebius was strongly influenced by classical philosophy, especially Platonism, and sought to reconcile this with Christian theology.*

1 Life, works and philosophical sources

Eusebius was born *c.* AD 264 and became Bishop of Caesarea in Palestine *c.*314. The library at Caesarea contained books brought from Alexandria by Origen with additions by Eusebius' mentor Pamphilus. Eusebius became an enormously industrious and wide-ranging scholar with a good general knowledge of philosophy. His *Historia ecclesiastica* (History of the Church) appeared in successive editions from *c.*300 onwards. His *Praeparatio evangelica* (Preparation of the Gospel), in fifteen books, was written *c.*312-18; it argues that Christian doctrines were foreshadowed by the philosophers, and is an invaluable source-book, especially for Middle Platonist authors whose works have otherwise been lost. A companion work, the *Demonstratio evangelica* (Proof of the Gospel) appeared about the same time.

Eusebius was a conservative theologian, anxious to maintain clear distinctions between the three divine Persons. He had some sympathy with Arius, signed the Creed of Nicaea (325) with reluctance, and led a partially successful reaction against its extreme adherents. Yet he warmly approved of the Emperor's reforms and wrote an eloquent address in praise of Constantine (*Laus Constantini*). He died *c.*339.

Eusebius quotes Plato very freely from original texts; but his knowledge of the Presocratics and of Aristotle, Epicurus and the Stoics is plainly derivative. He uses Philo of Alexandria and Plutarch, including some works not otherwise preserved, and is an important source for Atticus and Numenius. He quotes two books of Plotinus, now known as *Enneads* IV 7 and V 1, but not from Porphyry's edition. He uses Porphyry's *Philosophy from Oracles* and other writings, despite Porphyry's anti-Christian stance, but ignores his logical works.

2 Logic and ethics

Eusebius accepts the conventional division of philosophy into ethics, physics and logic (see Stoicism §2). But his treatment of logic is perfunctory; he ignores Plato's dialectic and does not use Albinus/Alcinous or Clement of Alexandria's *Stromateis* VIII, let alone Aristotle's or Chrysippus' logic. The 'logical skill' that he claims for Hebrew writers is identified with 'comprehension of the truth itself', which he finds in the Book of Proverbs; he also argues that Hebrew names correspond with realities, appealing to Genesis 2: 19 and adducing Plato's *Cratylus*.

His treatment of ethics, again, hardly goes beyond discussion of the *summum bonum*. He adopts the conventional criticisms of Epicurus as making pleasure the sole good, and of Aristotle as teaching a 'triad of goods' which values health and prosperity as much as virtue. He does, however, cite Plato's *Philebus* 67a, which assigns a positive though limited value to pleasure. Stoic ethics pass almost unnoticed. Eusebius praises Plato for locating the good in the intelligible world and assimilates Hebrew piety to Platonism, describing God as the source of all good things, so that the 'God-loving' person alone lives a happy and blessed life. Eusebius does not discuss the Jewish law in detail, though his own moral axioms are shown indirectly in his repeated condemnation of polytheism.

3 The physical world

For Eusebius, physics includes the study of the intelligible, incorporeal world as well as perceptible things, a Platonic doctrine which he claims the Hebrews accepted. Plato agreed with Moses in regarding the world as created; and Eusebius repeatedly criticizes the doctrines of the 'physicists' - some of them were atheists, while others deified the material elements, or at least regarded matter as the ultimate reality. Anaxagoras revered a divine

creative mind, but failed to explain its operations. Parmenides, however, rightly believed in unchanging reality, and Heraclitus in a divine *logos*. Aristotle's doctrine of the eternity of the cosmos is rejected in favour of the literalist interpretation of the *Timaeus* offered by Atticus; Eusebius ignores the common Platonist view that the myth of a divine creative act really signifies eternal dependence. Epicurus rejects any providential order, and his doctrine of atoms denies the unity of the cosmos, though Eusebius apparently ignores his view that the cosmos originated in an uncaused 'swerve' of free-falling atoms. The Stoics are blamed for explaining the universe in material terms, though their concept of God as 'intelligent fire' is defended, following Clement of Alexandria, as compatible with Scripture.

Apart from accepting Empedocles' doctrine of four elements, Eusebius says little about astronomy, geography and physics in its modern sense; he is content to reproduce a doxography from Pseudo-Plutarch. His claim of a biblical 'physiology' relies on Solomon's reputed expertise (1 Kings 4: 29-34), and is not further pursued.

4 The world of Ideas, and theology

Eusebius agrees with the 'dogmatic' Platonists in the importance he attaches to the Ideas, or intelligible reality; but like most Platonists before Plotinus he fails to include them in a consistent theory. He shows no serious interest in Plato's attempts to classify the Ideas by genera and species. He presents them as eternal and unchanging and adopts the prevalent view of them as thoughts in God's mind which function as prototypes for God's works in creation. But this view is presented mainly in relation to the species of things in the world, whereas a more consistent Platonist would begin the creative process with the highest genera: possibly the Idea of mutable being, or that of the living creature as such. Eusebius takes from Philo of Alexandria and Clement a Stoicized theory which gives first place to the cosmos and its elements. Yet he also presents the Ideas as intelligent beings, which so far from being changeless are capable of 'fall and divergence' from their original position as archangels, thus becoming demons. Eusebius quotes Plato's *Laws* in support; but an assimilation of Ideas to souls, based perhaps on *Phaedo* 79 and *Sophist* 248e-249a, was assumed by Philo and many Platonists. Eusebius of course accepts the immortality of the soul, and approves Plato's description of a judgment after death, but shows without difficulty that Plato offers inconsistent pictures of the soul's continued existence.

Eusebius also draws on Platonism for his description of God's being, presenting him not only as transcending the universe but also as incomprehensible. He claims that God is 'above all intelligence and beyond every concept and consideration', which recalls the Platonic phrase 'beyond mind and being' (*epekeina nou kai ousias*), based on Plato's *Republic* 509b (compare Aristotle fr. 46). He uses the human mind and its utterance (*logos*) as an analogy for God's being and self-disclosure, but (unlike Numenius) he seldom actually designates God as mind. He thinks of the divine Logos as a distinct personality, though in a later work he notes the unpopularity of the phrase 'a second God'. Opposing any theology which blurred the distinction between the divine Persons or envisaged their unity in material terms, he gave qualified support to the strongly subordinationist view of Arius; he accepted the Creed of Nicaea (325) only with reservations.

5 Providence and human freedom

Besides creation, Eusebius upheld the doctrine of divine providence against Epicurus and others, but was also concerned to establish the freedom of the human will, as presupposed by the concepts of moral effort and of praise and blame. He therefore attacked the determinism presupposed by the astrologers and pagan oracles, and condemned as inadequate the Stoics' attempt to reconcile determinism with moral freedom. In Eusebius' view, Chrysippus offered us a 'semi-slavery'. Eusebius appealed to our consciousness of freedom and correctly pointed out that future events are differently determined. Some events are bound to happen; others result unpredictably from a coincidence of natural causes. Nevertheless, the human will is a true cause, determining some events that would otherwise be undetermined or otherwise determined, and is not itself necessarily controlled either by physical causes or by social pressures. God's providence disposes the general laws of the universe, but also cooperates with our free wills to encourage us to virtue. Eusebius presents evil in terms of human wrongdoing, no doubt presupposing the Stoic doctrine that natural disasters are not evils unless feared or resented.

All this is intelligently argued; but Eusebius does not eliminate the problems presented by the predictive element in *Christian* prophecy and by the well-established doctrine of God's total foreknowledge. Nevertheless, his philosophical ability deserved more recognition than it received. The relatively open-minded approach that he

inherited from Origen rapidly became unpopular in the Eastern Church. The Nicene Creed became the standard of orthodoxy, and Eusebius' coolness towards it was remembered; he was never canonized, and his reputation rested upon his much valued and repeatedly imitated *History of the Church*.

See also: Platonism, Early and Middle

CHRISTOPHER STEAD

List of works

A complete list is given by D.S. Wallace-Hadrill (1960 - see below).

Eusebius (*c*.300-24) *Historia ecclesiastica (History of the Church)*, ed. E. Schwartz, *Eusebius Werke 2: Die Kirchengeschichte*, Leipzig: Hinrichs, 1903-8.(Massive edition containing all evidence; other useful editions include: Lake, K. (trans. and ed.) (1926, 1932) *Eusebius: The Ecclesiastical History*, London: Loeb Classical Library, Heinemann - Greek text with English translation; Lawlor, H.J. and Oulton, J.E.L. (trans. and eds) (1927-8) *Eusebius: The Ecclesiastical History*, London: SPCK - English translation with valuable notes; and Williamson, G.A. (trans. and ed.) (1965) *The History of the Church from Christ to Constantine*, Harmondsworth: Penguin - translation with introduction.)

Eusebius (*c*.312-18) *Praeparatio evangelica (Preparation of the Gospel)*, trans. and ed. E.H. Gifford, *Eusebii praeparatio evangelica*, Oxford: Oxford University Press, 5 vols, 1903.(Outstandingly good edition, with Greek text, English translation and valuable notes; see also Mras, K. (ed.) (1954, 1956) *Eusebius Werke 8: Die Praeparatio Evangelica*, Berlin: Akademie Verlag, 2 vols; 2nd edn ed. E. des Places, 1982 - authoritative Greek text.)

Eusebius (*c*.312-18) *Demonstratio evangelica (Proof of the Gospel)*, ed. I.A. Heikel, *Eusebius Werke 6: Die Demonstratio Evangelica*, Leipzig: Hinrichs, 1913.(Critical Greek text; see also Ferrar, W.J. (trans.) (1920) *Eusebius: The Proof of the Gospel*, London: SPCK, 2 vols - an English translation.)

Eusebius (335-6) *Laus Constantini (In Praise of Constantine)*, ed. I.A. Heikel, *Eusebius Werke I*, Leipzig: Hinrichs, 1902, 193-259.(Critical Greek text; see also Drake, H.A. (1975) *In Praise of Constantine*, Berkeley, CA: University of California Press - includes introduction, translation and notes.)

References and further reading

Des Places, E. (1982) 'Eusèbe de Césarée commentateur', *Théologie Historique* 63. (Catalogues Eusebius' use of Platonic, Platonist and scriptural texts.)

Dillon, J. (1977) *The Middle Platonists*, London: Duckworth.(Essential information on most philosophers immediately prior to Eusebius.)

Lyman, J.R. (1993) *Christology and Cosmology*, Oxford: Clarendon Press.(Well-written, simple account of basic themes in philosophical theology; see pages 82-123 for Eusebius.)

Stead, G.C. (1994) *Philosophy in Christian Antiquity*, Cambridge: Cambridge University Press.(Outline of Greek philosophy, and study of its use by Christians.)

Wallace-Hadrill, D.S. (1960) *Eusebius of Caesarea*, London: Mowbray.(Reliable introduction to Eusebius' life and works, a complete list of which can be found on pages 57-8.)

Wallace-Hadrill, D.S. (1982) 'Eusebius von Caesarea', in G. Krause *et al.* (eds) *Theologische Realenzyklopädie*, Berlin: de Gruyter, vol. 10, 537-43.(Succinct and authoritative, with full bibliography.)

Evans, Gareth (1946-80)

Frege's notion of sense is a conception of content whose application is controlled by the idea of rationality. A Fregean outlook is often taken to imply that singular reference is always mediated by descriptions, and hence to be inconsistent with the insights that have motivated proponents of 'direct reference'. But in his major work, The Varieties of Reference (1982), Gareth Evans showed how a treatment of singular reference on Fregean lines can accommodate those insights. This means that the semantics of singular reference need not be distanced from the philosophy of mind, in the way that proponents of 'direct reference' typically suppose. Within the framework provided by this synthesis, Evans gave detailed treatments of the different ways in which thought and speech are directed at particular objects. Particularly notable are his discussions of demonstrative thinking, which exploits the perceptible presence to the thinker of the object it concerns; of first-personal thinking; and of singular statements of nonexistence.

1 Life

Gareth Evans was a Fellow of University College, Oxford; in 1979 he became Wilde Reader in Mental Philosophy at the University of Oxford. His articles, on issues in metaphysics, philosophy of language and philosophy of mind, are assembled in his *Collected Papers* (1985), but his major contribution to philosophy is the posthumously published *The Varieties of Reference* (1982).

2 Historical background

In the Theory of Descriptions, Russell gave a precise account of one way in which the content of an episode of thinking or speaking can be targeted on an individual, in the sense that its truth-value depends on how things are in respect of that individual: the individual is singled out by uniquely fitting a specification that figures in the content - say, 'presently reigning over France' (see Russell, B. §9; Descriptions §§2-3). It is central to Russell's point that contents of this shape are available to be entertained in an episode of thinking, whether or not anything conforms to the specification. If nothing does, the thought is simply false, not deprived of content.

Russell contrasts this kind of content with contents that are more intimately related to objects on whose properties their truth depends. A content of this second kind can be entertained only by a thinker 'acquainted with' the objects concerned, and the very existence of such a content depends on the existence of these objects. Contents of this second kind are expressed by forms of words containing what Russell calls 'logically proper names'. Compared with the first kind of content, contents of this second kind seem to be genuinely singular; and compared with definite descriptions, 'logically proper names' seem to be genuinely referring expressions.

However, Russell thinks there are hardly any 'logically proper names'. In his view one cannot entertain or express one of these genuinely singular contents if there is so much as a possibility, by the lights of hyperbolical Cartesian doubt, that one's conviction that a suitable object exists is an illusion. Thus almost all expressions that would ordinarily be classified as referring expressions (for instance ordinary proper names) must be interpreted as 'disguised descriptions' - that is, abbreviations for expressions of the specifications that figure in the other kind of content. This 'disguised description' picture became something of an orthodoxy, supposedly with referring as its subject matter, even though in Russell's own thinking the idea of referring seems more at home in connection with 'logically proper names', which Russell himself sets in contrast with 'disguised descriptions'.

Frege had argued that singular terms must be credited with senses as well as referents (see Frege, G. §3; Sense and reference §1). The sense of a sentence is the thought expressible by it, and it is the sense rather than the referent of a singular term that contributes to determining the sense of a sentence in which the term figures. Frege's point here is to delineate a notion of content, thinkable or expressible, whose employment is controlled by the idea of rationality, in a way that can be illustrated as follows. Suppose two singular terms have the same referent; take a pair of sentences which are alike except that one contains one of those terms, where the other contains the other. If a rational subject can understand these sentences and take rationally opposed attitudes to them (for instance, endorsing one and rejecting the other), then the thoughts they express must differ; and this difference in the senses of the sentences can only trace back to a difference in the senses of the singular terms.

In the orthodoxy, the Fregean idea that singular terms have senses became assimilated to the neo-Russellian idea

that they are 'disguised descriptions'; the assimilation set up a supposedly unified doctrine, against which Saul Kripke and others recoiled. Their alternative yields a liberalized analogue to Russell's conception of 'logically proper names'. In the new picture, various contextual relations between thinkers or speakers and objects play something like the role in which Russell cast 'acquaintance', and genuinely referring expressions are not as rare as Russell thought. With Frege assimilated into the orthodoxy, this new picture is conceived as rejecting the Fregean apparatus. If Fregean sense figures at all in the new picture, it is not as an aspect of the semantics of singular terms, but in a supposedly separate role, as an instrument for describing the configurations of thinkers' minds (see Proper names §1).

3 Evans on Fregean sense

Evans offers a reading of Frege that undermines this assimilation. There are three main claims. First, the assimilation is not required by the motivation for Frege's notion of sense. In the case of what singular terms contribute to thoughts, Frege's point is that if attributing propositional attitudes is to be controlled by the idea of rationality, then individuating the contents of the attitudes by the identity of the objects that figure in them is too coarse-grained. Evans insists that singular thoughts can be individuated more finely, as Frege required, but still be 'Russellian' - that is, depend for their very existence on the existence of the objects they are about.

Second, Frege himself evidently feels the attraction of such a conception. He maintains that a putative singular utterance with no referents for its putative singular terms has no truth-value. We can best make sense of that as an implication of the idea that there is no thought (of the appropriate kind) for such an utterance to express - since the thoughts that singular utterances express, when they do express thoughts, are 'Russellian'. Frege's own principles make it difficult, perhaps impossible, to make sense of the alternative idea that such an utterance does express a thought, but one with no truth-value.

Third, although Frege often seems to say that a singular term's possession of a sense does not depend on whether it has a referent, this can be explained away. Such claims belong in the context of Frege's thesis that in cases of reference-failure one lapses into fiction. In fiction, he says, one expresses apparent thoughts. This suggests that when he says that an apparent singular term can have a sense even though it lacks a referent, what he really means is that such a term gives the appearance of having a sense, not that it genuinely has a sense.

It is not as important to settle what Frege himself held as to see that his basic principles allow for expressions whose semantics - possession of senses and all - qualify them to count, by quasi-Russellian lights, as genuine singular terms, rather than something on the lines of the 'disguised descriptions' that Russell contrasts with them. This means that a broadly Fregean conception of thought and its expression can accommodate the insights of the recoil from the 'disguised description' picture. We can acknowledge cases in which reference is not mediated by descriptions, but partly constituted by contextual relations between thinkers and objects, without needing to separate such reference from the characteristically Fregean topic, the rationally organized configurations of thinkers' minds.

4 Varieties of reference

Evans gives a general account of how 'informational' relations between thinkers and objects can enter into determining referential relations between thoughts and objects, even while the thoughts are conceived in a fundamentally Fregean way, as individuated by their positions in a space of possibilities open to a rational subject. He recommends his picture by contrasting it with 'the Photograph Model of mental representation', according to which relations of 'informational' type suffice for a mental state or episode to be about an object, without any need for a Fregean framework. In a transition from philosophy of mind to philosophy of language, he discusses at a general level how 'information-based' thoughts figure in the understanding that is required for some communication.

The general picture of 'information-based' thoughts is applied in detailed accounts of thoughts of three interdependent kinds: (1) thoughts in which objects are singled out demonstratively in a way that exploits their perceptible presence to the thinker; (2) thoughts about places or regions of space expressible by the use of words such as 'here'; and (3) first-personal thoughts. A central aim is to reconcile the fact that these thoughts single out their objects egocentrically with the claim that they are targeted on elements in the objective order. The capacities

to think these thoughts belong to a subject who is experientially and practically placed in the objective world. Such a subject enjoys a course of experience understood as jointly determined by the way things are anyway and by the subject's trajectory through objective reality, and thereby knows what it is for egocentric identifications of objects (including places) to be directed at items that are also identifiable in more objective ways (for instance by map references). Evans thus displays first-personal thoughts, in particular, as being in their own way about particular objects (the human beings - animals of a certain distinctive kind - who self-consciously think them). This stands in radical contrast to the Cartesian position, in which at least some such thoughts involve reference to immaterial entities problematically related to certain animals: it also defuses a temptation to avoid the Cartesian position by denying that those thoughts involve reference to anything at all.

Evans' general account of 'informational' relations might lead one to expect discussions of thoughts linked to their objects by relations of testimony and memory, corresponding to his treatment of perceptually demonstrative thoughts. The book contains nothing answering to the former, however, and where one might expect the latter, a treatment of thoughts whose objects are singled out by the thinker's ability to recognize them. This is insightful about recognitional capacities, but arguably not well integrated into the overall structure of the book.

A problem for 'Russellian' conceptions of reference is posed by negative existential statements that appear to have singular terms in subject position (for example, 'The planet Vulcan does not exist'). The 'disguised description' picture has no difficulty here; such a statement simply says that the associated specification is not satisfied. Metalinguistic treatments are not satisfying; such a statement seems somehow to use, rather than talk about, the singular terms it contains. One would like to say: 'Such a statement says, of some specified thing (for example, the planet Vulcan), that it does not exist.' But that would describe the statement in a way that undermines what it tries to say. Evans gives an ingenious and original account of such statements, which allows us to get as close as possible to that desideratum without falling into nonsense. One can go through the motions of referring within the scope of a game of make-believe (as when one 'talks about' fictional characters); the idea is that a statement such as 'Vulcan does not exist' exploits such a game of make-believe for the serious purpose of declaring that that is all it is. The book concludes with a suggestive discussion of proper names, centring on different degrees of involvement in name-using practices.

See also: Reference §§4, 8

JOHN McDOWELL

List of works

Evans, G. (1982) *The Varieties of Reference*, ed. J. McDowell, Oxford: Clarendon Press. (Evans' major work; it places Frege and Russell with respect to more recent reflection about singular reference, and discusses the various ways in which thought and speech are targeted on particular objects. Rough in places, and occasionally technical, but well worth the effort.)

Evans, G. (1985) *Collected Papers*, Oxford: Clarendon Press.(Evans' shorter writings, ranging from technical philosophy of language to general metaphysics and philosophy of mind. Perhaps especially noteworthy is the brilliant paper 'Things without the Mind'.)

References and further reading

Frege, G. (1892) 'On Sense and Meaning', in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984.(Frege's introduction of the notion of sense. A classic.)

Kripke, S. (1980) *Naming and Necessity*, Oxford: Blackwell.(A brilliant and trenchant attack on the 'disguised description' picture of how proper names refer.)

McDowell, J. (1986) 'Singular Thought and the Extent of Inner Space', in P. Pettit and J. McDowell (eds) *Subject, Thought and Context*, Oxford: Clarendon Press, 137-68. (Discusses the significance, for philosophy of mind and metaphysics, of the way Evans synthesizes a basically Fregean outlook with rejecting the 'disguised description' picture.)

Russell, B. (1905) 'On Denoting', in *Logic and Knowledge*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 39-56.(Another classic: Russell's announcement of the Theory of Descriptions. Not difficult, apart from a famously opaque section on Frege.)

Russell, B. (1918) 'The Philosophy of Logical Atomism', in *Logic and Knowledge*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 177-281.(Lectures for nonspecialists: contains a lucid exposition of Russell's central

doctrines about reference.)

Events

Events are entities like collisions and speeches, as opposed to things like planets and people. Many are changes, for example things being first hot and then cold. All lack a thing's full identity over time: either they are instantaneous, or they have temporal parts, like a speech's words, which stop them being wholly present at an instant; whereas things, which lack temporal parts, are wholly present throughout their lives.

Events may be identified with two types of entity: facts, like the fact that David Hume dies, corresponding to truths like 'Hume dies'; or particulars which, like things, correspond to names, for example 'Hume's death'. Which one they are taken to be affects the content of many metaphysical theories: such as that all particulars are things; that times, or causes and effects, or actions, are events; or that mental events are physical.

1 Events and things

Many kinds of entity, from any cause or effect to everything a space-time region contains, have been called 'events'. But events usually so-called - deaths, collisions, speeches - form an apparently distinct kind, different from things like people, planets and books (see Continuants). What is the difference? Many events are changes, for example, human bodies being first alive and then dead. But this may not define events. For first, we may need events to distinguish intrinsic changes, like dying, from some relational ones, like being orphaned; the latter being mere entailments of the former, which are real events, with contiguous causes and effects (see Change). Second, events that begin or end things, like the Big Bang and other explosions, cannot be changes in them and may not, if nothing precedes or survives them, be changes in anything else.

The difference between things and events, whether changes or not, may be that things keep a full identity over time, which events lack. First, some events may be instantaneous and lack any identity over time. Second, temporally extended events are deprived of full identity over time by their temporal parts, like a speech's spoken words, which stop them ever being wholly present at an instant; whereas people and other things have no temporal parts and are wholly present at every instant of their lives. This full identity over time will then distinguish one thing changing from successive things having different properties, thus explaining why only things can change and why changes, being events, are not things (Mellor 1981).

This difference may be denied by giving things temporal parts by definition, such as Hume-in-1739. But these are mere logical constructions from things and times, not independent events like the words in a speech. Some apparent things might indeed be mere strings of contiguous and causally related events (it has been suggested, for example, that we are strings of experiences). But not all: unchanging elementary particles involve no independent events. Moreover, since contiguity and causation can always link one event or thing to two successors, as when a cell divides, they cannot entail a thing's identity over time (see Personal identity). So, equating us to strings of experiences implies not that things can be strings of events but that we are not things. Events and things remain distinct types of entity.

2 Events and facts

Assuming there are things, are there also events? That may depend on whether events are facts, corresponding to truths like 'Hume dies in 1776', or particulars corresponding to names or descriptions like 'Hume's death' (see Facts; Particulars). Changes look like facts, for example the fact that a thing is first hot and then cold; and that things start and cease to exist at certain times are also facts. Thus events of both types mentioned above may be facts, and what many authors call 'events' certainly are: Kim (1976), for whom events are things having properties at times, like Hume being alive in 1775, is an example. Events in this sense are real entities if and only if facts are.

This being so, 'event' is best reserved, as by Davidson, for particulars like Hume's death. Their reality is independent of that of facts, but equally contentious: after all, the only particular apparently referred to in 'Hume dies' is Hume. Yet Davidson argues that 'Hume dies' also entails that an event exists which is a death of Hume. For first, this shows how 'Hume dies slowly' entails 'Hume dies', since a slow death must be a death. Second, identifying actions with particular events satisfying different descriptions dissolves puzzles about their identity: for example my bid can be a purchase even though many bids are not purchases (see Action). Similarly, if mental

events are particulars, they can also be physical brain events satisfying neurophysiological descriptions (see Identity theory of mind). This explains, without invoking non-physical causes or effects, how events satisfying mental descriptions - such as ‘is a decision to bid’ - can have physical causes and effects, like hand movements (see Anomalous monism). This explanation assumes moreover that causes and effects are particulars, not facts, and this requires particular events. For only if Hume’s death exists can the effect of whatever caused Hume to die be a particular; otherwise the effect can only be the fact that Hume dies.

All these arguments have been disputed. Events remain more contentious than things, despite having identity criteria: for example, that a headache is a certain brain event if it has the same causes and effects as that event. Yet since such criteria may only relate particular events to each other, we may still need people and other things to identify some events to start with, as in ‘Hume’s death’ (see Strawson, P.F.). This may explain scepticism about events, but cannot make them less real than things, or less able to be particulars. For to be a particular is just to be of a kind we make true first-order generalizations about (see Quine, W.V.), many of which, like Newton’s ‘to every action there is always opposed an equal reaction’, are about events as well as things.

See also: Adverbs; Causation; Logical form; Momentariness, Buddhist doctrine of; Mujo; Ontology; Reichenbach, H.; Time; Whitehead, A.N.

D.H. MELLOR

References and further reading

Bennett, J. (1988) *Events and Their Names*, Oxford: Clarendon Press.(An analytic assessment of rival views of events, concluding that they are facts.)

Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Clarendon Press.(A collection of articles arguing for and applying the view that events are particulars.)

Kim, J. (1976) ‘Events as property exemplifications’, in M. Brand and D. Walton (eds) *Action Theory*, Dordrecht: Reidel, 159-77.(Referred to in §2 above. Equates events with things having properties at times.)

Lombard, L.B. (1986) *Events: A Metaphysical Study*, London: Routledge & Kegan Paul. (Argues that events are changes.)

Mellor, D.H. (1981) *Real Time*, Cambridge: Cambridge University Press.(Referred to in §1 above. Chapter 7 argues that events are particulars which differ from things in having temporal parts.)

Evil

Evil is serious unjustified harm inflicted on sentient beings. Two types of evil can be distinguished: 'natural evil', which is the product of nonhuman agency, and 'moral evil', which is the product of human agency. Moral thinking tends to focus on moral evil, and three main interpretations of it have been made. One, initiated by Socrates, holds moral evil to be deviation from the good; another, favoured by Stoic-Spinozists, views it as illusory; the third, made originally by Leibniz, sees it as a contrast necessary for the existence of the good. A realistic account must face the fact that moral evil does exist, and much of it is due to common human vices, which coexist with virtues in human character. It is primarily the proportion of the mixture, not the knowledge and intentions of agents, that determines how much evil will be caused by specific individuals in specific contexts.

1 The nature of evil

Evil is the most severe condemnation our moral vocabulary allows. Murder, torture, enslavement and prolonged humiliation are some examples of it. Evil must involve harm, and it must be serious enough to damage its victims' capacity to function normally (see Suffering). Furthermore, the harm must be unjustified, since not even serious harm is in itself necessarily evil, as it may be just punishment for crimes committed or the only means of preventing even greater harm (see Crime and punishment §2). What harm is justified is one of the fundamental questions of moral philosophy. The competing answers to it, however, share the key idea of a moral equilibrium. In general terms, harms that tend to maintain the moral equilibrium are justified, while those that tend to produce a disequilibrium are unjustified. The generality of this explanation allows for disagreements about what count specifically as harms, and about how the moral equilibrium can be best maintained.

Evil may be the product of human or nonhuman agency. Inclement weather that causes crop failure and widespread starvation is an example of the latter, and it is usually described as natural evil. Evil caused by human beings, such as torturing an innocent person, is moral. This traditional distinction between natural and moral evil is useful, but it should not be drawn too sharply because human beings may be natural agents, as carriers of a disease, for instance, and evil caused by natural agency may warrant moral opprobrium, if it was preventable and those responsible for doing so failed. Moral thinking nevertheless tends to focus on moral evil, since it is much more likely to be within human control than natural evil.

The primary subjects to which moral evil (simply 'evil' from now on) may be ascribed are human actions. Intentions, agents, and institutions may also be evil, but only in a derivative sense. For intentions are evil if they lead to evil actions; agents are evil if the preponderance of their actions are evil; and institutions are evil if they regularly prompt agents representing them to perform evil actions. In its primary sense, therefore, evil is connected essentially with causing serious unjustified harm to sentient beings, and since the means by which this is done are human actions, an account of evil should begin by concentrating on them.

It is clear that evil actions are widespread, and that they are responsible for much suffering. The obvious explanation of this fact is that human beings are motivated by greed, cruelty, envy, rage, hatred, and so forth, and evil actions are the manifestations of these vices. But this is unilluminating, unless it is combined with an explanation of why human beings possess and act on vices. To attribute vices to choice is a poor explanation, since many vices are the unchosen consequences of genetic predispositions and corrupting circumstances, and even if vices are the results of choices, the question of why vices rather than virtues are chosen remains (see Virtues and vices §4).

2 Evil as deviation from the good

The philosophically most influential explanation of evil is embedded in the Socratic view that no one does evil knowingly (see Socrates §6). The thought behind the apparently obvious falsehood of this claim is that human agents are normally guided in their actions by what seems to be good to them. The explanation of evil actions must therefore be either that the agents are ignorant of the good, and perform evil actions in the mistaken belief that they are good, or, while they know what the good is, they do evil unintentionally, through accident, coercion, or some incapacity (see Moral knowledge §1; Akrasia). The remedy for evil, consequently, is moral education that imparts genuine knowledge of the good and strengthens the intention to act on it.

This Socratic view, however, is driven to rely on a metaphysical assumption about the nature of reality and its effect on human aspirations. For, since human experience of the world testifies that thorough knowledge of the good and good intentions are compatible with evil actions, it must be supposed that human experiences disclose only appearances, not reality. The metaphysical assumption that needs to be made, therefore, is: first, that beyond human experiences of the world that appears to contain chaos and evil, there is a suprasensible true reality, in which a moral order prevails; and second, that good lives for human beings depend on learning to live in conformity to this order, rather than being led astray by deceptive appearances. Plato's Socrates explains evil, therefore, as a deviation from the good due to a human defect in cognition or intention that leads to mistaking appearance for reality.

This metaphysical assumption and the explanation of evil implied by it has passed from Greek thought to Christian theology chiefly through the works of Augustine and Aquinas. Christianity attributes to an all-knowing, all-powerful, all-good God the creation of the moral order that permeates reality, and it explains the prevalence of evil by the corrupting influence of original sin, which leads human beings to choose evil over the good, and thereby wilfully or weakly pit themselves against God's moral order (see Sin §2). Although Christian thinking about evil has dominated Western thought between the Greeks' time and ours, it nevertheless must be seen, if we ignore some twists and turns of theological sophistication, as a particular adaptation and elaboration of the metaphysical assumption and explanation of evil first advanced by Plato's Socrates.

This metaphysical assumption, however, cannot be reasonably maintained in the light of well-known objections, which can only be stated here without elaboration. First, any evidence that may be cited in favour of the supposed existence of a moral order in a supposed suprasensible reality beyond the world as it appears to normal human observers must be derived from the world as it appears to normal human observers, since, as a matter of logic, there is no other possible source of evidence. The evidence derived from appearances, however, cannot reasonably be taken to point to any suprasensible order in reality because the most such evidence can imply is that human knowledge of the world as it appears is limited and fallible. It is logically impossible for evidence to support inferences about what may lie beyond all possible evidence. Second, if, undeterred by this logical obstacle, defenders of the metaphysical assumption pursue their speculations about the implications of the evidence, they must recognize that evidentially unsupported implications can be derived both in favour of and against their assumption. If the existence of a moral order in suprasensible reality is inferred from observed instances of apparent goodness, then the existence of an evil order in suprasensible reality must be analogously inferrable from observed instances of evil. There is, consequently, no more reason to think of evil as deviation from the good than there is to think of the good as deviation from evil. Third, even if it is assumed for the sake of argument that the metaphysical assumption is defensible, it accounts only for moral evil, caused by human failure, and not for natural evil, whose occurrence cannot be attributed to human agency.

3 Evil as illusory

The Stoic-Spinozistic view attempts to avoid these objections by denying the reality of evil. It acknowledges that evil appears to exist, but its appearance is said to be an illusion to which human beings succumb through misdirected desires (see Spinoza 1677). Such desires cannot be satisfied because they are contrary to the moral order of reality, which need not be supposed to be suprasensible. If the misdirection of these desires is recognized, then the unavoidable experience of frustration they cause will be seen as emotional detritus that has been misidentified as evil. The advantage of the Stoic-Spinozistic view is that if evil were indeed illusory rather than real, then the objection to the Socratic view that it is incapable of explaining the reality of evil would be met (see Spinoza, B. de §10; Stoicism §19).

The Stoic-Spinozistic view is undoubtedly right in that what appears to be evil may not be and that greater self-knowledge and self-control make it possible to avoid much unnecessary suffering caused by mistaking for evil the frustration of misdirected desires. This view, however, is advanced not merely as a proposal for alleviating some instances of evil, but as an explanation of all evil. And as such, it is a failure for several reasons.

First, it cannot reasonably be held that all desires are misdirected, for human nature requires having and satisfying some desires. Rightly directed desires, however, are often frustrated, their frustration often results in serious unjustified harm, and that is real evil which cannot be alleviated by growth in self-knowledge and self-control. Second, the distinction between real and illusory evil rests on moral beliefs, which may be true or false. This view

is committed to holding that beliefs about the occurrence of evil are always false. For if some were true, not all evil could be illusory. If, on the other hand, beliefs about the occurrence of evil were always false, then the belief that torturing innocent people is evil, for instance, would also be false. It is, however, a precondition of morality, and indeed of civilized life, that such basic moral beliefs are regarded as true. The view that all evil is illusory, therefore, is incompatible with human nature, morality, and civilized life.

4 Evil as a contrast necessary for the good

Another attempt to account for evil is the Leibnizian view that recognizes its reality, but argues that the evil that exists is the minimum necessary for the existence of the good, which far outweighs the amount of evil there is (see Leibniz 1710). Evil is thus seen as the cost of the great benefits the good provides (see Leibniz, G.W. §3). The assumption behind this view is that the good could exist only in contrast with evil. But whatever is true of phenomena requiring contrasting aspects, it is not true of good and evil. It is absurd to suppose that there can be kindness only if there is cruelty, or freedom only if there is tyranny. Defenders of this view therefore tend towards an epistemological sense of the contrast: evil is said to be required so that the good could be appreciated as good. The difficulty with this is that even if a contrast were necessary for appreciation, drawing it would not require the existence of evil. The good could be properly appreciated even in contrast with imaginative depictions of evil. It is, for instance, unnecessary to have people actually drawn and quartered in order to maintain a lively appreciation of one’s good health. Nor is it required for the appreciation of the good that it be contrasted with evil, since the neutral or the indifferent would serve as a contrast just as well. People’s dying in their sleep, without being tortured to death, is sufficient to enhance one’s appreciation of the good of being alive.

5 Facing evil

It will perhaps be apparent that the various attempts to account for evil are not among the highest achievements of Western philosophy. All the surveyed accounts begin with the assumption that the good is primary and then vainly struggle to explain the prevalence of evil. It is hard to avoid the conclusion that the history of this subject is riddled with bad arguments and sentimentalism. A reasonable account of evil must acknowledge the reality and prevalence of evil. It must recognize that much evil that prevents the wellbeing of humanity is caused by human beings who are not moral monsters but ordinary people going about their ordinary lives. The character of such people includes virtues and vices, and a struggle between them (seeVirtues and vices §5). Depending on the hardships they face, the traditions and institutions that guide their conduct, and their capacity, opportunity, and motivation for moral reflection, sometimes their virtues and sometimes their vices prevail. Human nature is mixed; it is neither simply good, nor simply evil. It is primarily the proportion of the mixture, not the knowledge and intentions of agents, that determines how much evil will be caused by specific individuals in specific contexts. The search for a metaphysical explanation for this banal fact is a diversion from the morally necessary task of decreasing evil by improving the conditions and character of individual moral agents.

See also: Evil, Problem of; Holocaust, the

JOHN KEKES

References and further reading

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, trans. English Dominican Fathers, London: Dent, 1934, Ia.47-9.(A classic of the Christian approach to evil.)

Augustine (397-401) *Confessionum libri tredecim (Confessions)*, trans. E.B. Pusey, London: Dent, 1907, book VII, chaps 3-5.(Another classic of the Christian approach to evil.)

Kant, I. (1793) *Die Religion innerhalb der Grenzen der blossen Vernunft*, trans. T.M. Greene and H.H. Hudson, *Religion Within the Limits of Reason Alone*, New York: Harper & Row, 1960.(An attempt to combine the Christian approach with a secular understanding of evil.)

Kekes, J. (1990) *Facing Evil*, Princeton, NJ: Princeton University Press.(A treatment of evil from a secular point of view.)

Leibniz, G.W. (1710) *Essai de Theodicée sur la bonté de Dieu, la liberté de l’homme, et l’origine du mal*, trans. E.M. Huggard, *Theodicy: Essays on the Goodness of God, the Freedom of Man, and the Origin of Evil*, London: Routledge & Kegan Paul, 1952.(A classic statement of the view that evil is required by the good.)

Madden, E.H. and Hare, P.H. (1968) *Evil and the Concept of God*, Springfield, IL: Charles C. Thomas.(A criticism of the main Christian approaches.)

McCord Adams, M. and Adams, R.M. (eds) (1991) *The Problem of Evil*, New York: Oxford University Press.(A collection of articles and a bibliography representing recent work on the Christian problem of evil.)
Plato (*c*.386-380 BC) *Meno*, trans. W.K.C. Guthrie, in *The Collected Dialogues*, Princeton, NJ: Princeton University Press, 1961.(One source of the Socratic view.)
Spinoza, B. de (1677) *Ethica Ordine Geometrico Demonstrata (Ethics Demonstrated in a Geometrical Manner)*, trans. E. Curley, *Ethics*, Princeton, NJ: Princeton University Press, 1985, parts IV-V.(A classic statement of the view that evil is due to misdirected desires.)

Evil, problem of

In this context, 'evil' is given the widest possible scope to signify all of life's minuses. Within this range, philosophers and theologians distinguish 'moral evils' such as war, betrayal and cruelty from 'natural evils' such as earthquakes, floods and disease. Usually the inescapability of death is numbered among the greatest natural evils. The existence of broad-sense evils is obvious and spawns a variety of problems, most prominently the practical one of how to cope with life and the existential one of what sort of meaning human life can have.

Philosophical discussion has focused on two theoretical difficulties posed for biblical theism. First, does the existence of evils show biblical theism to be logically inconsistent? Is it logically possible for an omnipotent, omniscient and perfectly good God to create a world containing evil? One classical response to this, following Leibniz, is to argue that such a God would create the best of all possible worlds, but that such a world may contain evil as an indispensable element. Alternatively, evil may be an unavoidable consequence of the boon of free will, or it may be part of a divine plan to ensure that all souls attain perfection.

The second difficulty for biblical theism is, even if we grant logical consistency, does evil (in the form, for instance, of apparently pointless suffering) nevertheless count as evidence against the existence of the Bible's God? One frequent theistic response here is to argue that the apparent pointlessness of evil may be merely a result of our limited cognitive powers; things would appear the same to us whether or not there were a point, so it is not legitimate to argue from the evidence.

1 Problems of evil

The so-called 'logical' problem of evil rests on the contention that the following two claims of biblical theism:

(I) God exists, and is essentially omnipotent, omniscient and perfectly good; and
(II) evil exists,

combine with the following plausible attribute analyses:

(P1) a perfectly good being would always eliminate evil so far as it could;
(P2) an omniscient being would know all about evils; and
(P3) there are no limits to what an omnipotent being can do,

to form an inconsistent quintet, so that the conjunction of any four entails the denial of the fifth; most notably the conjunction of (P1)-(P3) with either of (I) or (II) entails that the other is false.

Such an argument can be taken aporetically, as a challenge to propose more subtle alternatives to (P1)-(P3), but it has usually (in analytical philosophy of religion since the 1950s) been advanced 'atheologically' as an argument against the existence of God (see Atheism §4; Natural theology §5). Likewise important is the distinction between the abstract problem, which takes 'evil' in (II) to refer generally to some evil or other (say the pain of a single hangnail), and the concrete problem, which construes (II) as shorthand for the existence of evils in the amounts and of the kinds and with the distribution found in the actual world. While the abstract problem raises a question of conceptual interest, it is the concrete version that gives the issue its bite.

Bold responses deny (P3), maintaining variously that God cannot overcome certain natural necessities (like Plato's Demiurge), that he cannot conquer his evil twin (as in Manichean dualism), or even that he lacks the power to compel at all (see Process theism §2). Some reject (P2), observing that many evils arise from free choice, while future contingents are in principle unknowable (see Omniscience §4). (P1) is the most obviously vulnerable because it is contrary to the common intuition that ignorance and weakness excuse, and is best replaced with:

(P4) it is logically impossible for an omniscient, omnipotent being to have a reason compatible with perfect goodness for permitting (bringing about) evils.

Rebuttals seek to counterexemplify (P4) by identifying logically possible reasons available even to an omniscient, omnipotent God.

2 Logically necessary connections with greater goods

Since omnipotence is not bound by causally necessary connections, it is natural to look for reasons among the logically necessary connections of evils with greater goods. Because the piecemeal approach of correlating distinctive sorts of good with different kinds of evil (for example, courage with danger, forgiveness with injury) threatens to be endless, it seems advantageous to identify a single comprehensive good that logically integrates all ills. One promising strategy takes its inspiration from Leibniz and develops his 'best of all possible worlds' ('BPW') theodicy in terms of contemporary possible-worlds semantics (see Leibniz, G.W. §3; Modal logic, philosophical issues in §1). If a possible world is a maximal consistent state of affairs, each of infinitely many constitutive details is essential to the possible world of which it is a part. Assuming (P5) that possible worlds as wholes have values (P6) that can be ranked relative to one another and (P7) that the value scale has a maximum (P8) occupied by one and only one world, one can interpret divine creation in terms of actualizing a possible world and reason (P9) that necessarily an essentially omniscient, omnipotent and perfectly good God would actualize the best. Given the further controversial claim that:

(P10) the BPW contains instances of evil as logically indispensable components,

it follows that the desire to create the BPW is a reason compossible with perfect goodness for God not to prevent or eliminate all instances of evil.

(P10) contradicts our *prima facie* intuition that the BPW should be homogeneously good. Defenders of BPW approaches (see Chisholm 1968) distinguish two ways in which value-parts may be related to value-wholes. The one presupposed by the critics is simply additive: negatively and positively valued parts simply 'balance off' one another and the inclusion of any 'minuses' inevitably lowers the value total. By contrast, parts may be integrated into wholes by relations of organic unity, in such a way that the positive value of the whole may defeat the negative value of the part (for example, the way the beauty of Monet's design defeats the ugliness of some colour patches). (P10) envisages the defeat of evil within the context of the possible world as a whole.

Leibniz thought he could prove the necessity of (P10) on the basis of his a priori arguments for the necessity of (I) and (P9); he believed that (P10) followed from the fact that God had actualized this world. Yet (P10) seems to fall into that class of propositions that are logically possible if and only if logically necessary. Those who recognize no sound demonstrations of (I) are left to claim that (P10) is epistemically possible. Since the atheologian is in the same epistemic predicament with respect to (P10), this epistemic defence would be sufficient *ceteris paribus* to discharge the burden of proof imposed on the theist by the argument from evil.

This BPW approach makes several other debatable value-theory assumptions. Augustine's notion (*contra* P8) that many alternative worlds have maximum value imposes no damage. Aquinas' insistence (*contra* P7) that for every collection of creatures there is a better one would not be crippling if every possible world above a certain value-level included evil. The rejection of (P5) and (P6), however, would be fatal for BPW approaches. Some question whether our comparative evaluations of small-scale states of affairs (for example, Jones' enjoying a symphony as better than his experiencing excruciating pain) is good evidence that the values of maximal states of affairs form a hierarchy. More fundamentally, some have argued (*contra* P5) that states of affairs are not intrinsically good or bad, although they can be good or bad for certain persons or projects and can ground different moral evaluations by particular agents. Anti-consequentialists in ethics also challenge whether (P9) follows from (P5)-(P8) (see Consequentialism). Deontologists would let justice to individuals trump putative increases in the value of states of affairs (see Deontological ethics). Could creating the BPW be a reason compossible with perfect goodness for permitting suffering and degradation for the relatively innocent? Even if such value-maximizing were compatible with perfect goodness, it is not obviously required. For example, divine goodness is often interpreted as grace, a disposition to show favour independently of merit.

Finally, this modified Leibnizian approach entails divine determinism, because in choosing which of infinitely many fully determinate possible worlds to actualize, God is deciding on each and every detail. Some find this theologically objectionable, either because it seems incongruous for God to hold created persons responsible to himself for actions he determined, or because it fails to put enough distance between evil and divine aims.

3 Free-will defences

The last-mentioned worries are well accommodated by the other main traditional theme - that (some or all) evil originates in the wrong or evil choices of free creatures. Free-will approaches contend that:

(A1) created free will is a very great good, whether intrinsically or as a necessary means to God's central purposes in creation;
(A2) God cannot fulfil his purposes for and with free creatures without accepting the possibility that some will misuse their freedom, thereby introducing evil into the world.

In classical developments of this defence, (A1) is supposed to be a reason compossible with perfect goodness for making free creatures, while (A2) is compatible with the claim that evil is not necessary to the perfection of the universe or any other divine purpose. Some or all evil is not something God causes or does, but something he allows, a (perhaps) known but unintended side effect of his aims. The introduction of evil into the world is explained by the doctrine of 'the Fall', according to which God made angelic and human free agents in naturally optimal condition and placed them in utopian environments. God wanted them freely to choose what is right or good, but some angels and the primordial humans Adam and Eve chose what is wrong, thereby actualizing the possibility of evil.

Contemporary attention (beginning with Plantinga) has turned away from free-will defences based on the principles of double effect and doing-allowing - the principle that agents are not as responsible for the known but unintended side effects of their actions as they are for their chosen means and ends; that they are not as responsible for what they allow as for what they do - to others that reconnect with possible-worlds semantics (see Double effect, principle of). Once again, God creates by actualizing a possible world, but freedom is now taken to be incompatible with determinism, with the consequence that God and free creatures collaborate in determining which possible world becomes actual. Created freedom does not so much 'distance' God from evils as limit which worlds God can create. As with BPW approaches, God evaluates possible worlds as to their global features - (P5) and (P6) are assumed true, although not necessarily (P7) and (P8) - but this time he evaluates those that are a function of created incompatibilist-free choice: for example, a very good world with the optimal balance of created moral goodness over moral evil.

In defence of (A2), both classical and possible-worlds approaches appeal first to the notion that not even God can cause someone else's incompatibilist-free choices. To the objection that God should use his foreknowledge to actualize only incompatibilist-free creatures who will never sin, free-will defenders reply that such foreknowledge is not prior in the order of explanation to God's decision to create. To the suggestion that God should use his middle knowledge of what free creatures would do in particular circumstances, some (notably Plantinga 1974) grant that such counterfactuals of freedom can be true, but argue that it is logically possible that all incompatibilist-free creatures be 'transworld depraved' - that is, that no matter which combinations of individuals and circumstances God actualized, each would go wrong at least once - and logically possible that any world containing as much moral goodness as the actual world would also include at least as much moral evil as the actual world contains. Thus, it is logically possible that God could not create a world with a better balance of moral good over moral evil - which would be a reason compossible with perfect goodness for his not doing so.

This ingenious argument is controverted both by those who agree and those who deny that counterfactuals of freedom can be true. Among the former, Suárez (§1) defends middle knowledge but arguably finds transworld depravity impossible because of God's necessary resourcefulness, which he takes to have the following implication: necessarily, for any possible person and any situation in which they can exist, there are some helps of grace that would (should God supply them) win the creature over without compromising its incompatibilist freedom. Others (notably R.M. Adams 1977) wonder what could make such counterfactuals true about creatures considered as merely possible. Incompatibilist freedom rules out divine choices or any native features of the creative will. To appeal to a contingent condition (*habitudo*, or primitive property) independent of both is too close for comfort to the ancient doctrine of fate that falls alike on the gods and their creatures, and contradicts traditional Christian views of divine providence (see Providence §1). To maintain that counterfactuals of freedom are true although there is nothing to make them true violates a correspondence theory of truth (see Truth, correspondence theory of). Denying truth to such counterfactuals of freedom does not automatically put (A2) clear of the objection from omniscience, however, if God could know about merely possible creatures what they probably would do in any given circumstance. But the meaning and ground of such probability assessments is at least as problematic as

that of the original counterfactuals (see R.M. Adams 1977, 1985).

Even if (A2) were unproblematic, it could still be asked whether (A1) necessarily constitutes a reason compossible with perfect goodness for allowing evils. Two dimensions of divine goodness may be distinguished: ‘global’ goodness and goodness to individual created persons. The possible-worlds approaches cite global features - ‘the best of all possible worlds’, ‘a world a more perfect than which is impossible’, ‘a world exhibiting a perfect balance of retributive justice’, ‘a world with as favourable a balance as God can get of created moral good over moral evil’ - by way of producing some generic and comprehensive reason for allowing evil. But worlds with evils in the amounts and of the kind and with the distribution found in the actual world contain horrendous evils - evils the participation in (the doing or suffering of) which gives one *prima facie* reason to doubt whether one’s life could (given their inclusion in it) be a great good to one on the whole - unevenly distributed among humans and uncorrelated with variations in desert. Even if horrors thus apportioned were epistemically compatible with global perfections, these defences of divine goodness as a producer of global perfection would not so much guarantee as raise doubts about God’s goodness to individual participants in horrors. Divine goodness to them would require God to defeat the disvalue of horrors not only within the context of the world as a whole, but also within the framework of the individual participant’s life. Nor will precise individual retribution fit this bill where the perpetrators of horrors are concerned. ‘Balancing’ horror with horror only deepens the difficulty. Some Christians bite this bullet, insisting that decisive defeat of evil is promised only to the obedient, while the wicked can expect the reverse, a decisive defeat of positive meaning in their lives in the form of eternal damnation. Others insist that the doctrine of hell only makes matters worse by giving rise to a specialized version of the problem of evil (see Hell).

4 Divine goodness to creatures

Soul-making theodicies try to fill the explanatory gap regarding divine goodness to individual created persons by adding further hypotheses as to what they might get out of existence in an environment in which they are so vulnerable to sin, suffering and horrors. Some versions stipulate:

(A3) God’s purpose in creation culminates in a process of spiritual development in which autonomous created persons with their own free participation are perfected, and transformed from self-centred to other-centred, God-centred, Christlike or otherwise virtuous souls; and
(A4) environmental evils are permitted because they create an environment favourable to soul-making.

(A3) is compatible both with the notion that humans are initially created with mature unobstructed agency and so are fully responsible for their choices, and with the alternative idea (retrieved from Irenaeus by Hick (1966)) that human agency began immature, so that sin was to be expected in the course of the ‘growing-up’ process. The idea is that life in a world with evils such as this is, or with created cooperation can be, ‘good for the soul’.

Establishing (A4) is difficult thrice over because: (i) the task shatters into piecemeal cataloguing, with separate demonstrations for each sub-type of environmental evil; (ii) relevant necessary connections with the soul-making environment can be hard to show; and (iii) experience makes it *prima facie* implausible that a world with evils such as ours is a good classroom for the soul. In response to (ii), some (notably Hick 1966) ingeniously contend that ‘dysteleological’ evils lend an air of mystery which is itself favourable to soul-making. Others (for example, Stump 1985) modify (A4) to acknowledge that some environmental evils are consequences of sin.

Where God’s soul-making purpose succeeds, it is easy to see how the painful journey is worth the individual’s while. What about where it fails? Some (especially Swinburne 1983) reply that the dignity of self-determination is enough, whatever the outcome. The credibility of this contention varies with one’s estimate of the robustness of human nature as well as one’s conception of the natural or punitive consequences of repeated bad choices. Pessimists argue that ante-mortem participation in horrors makes a mockery of human self-determination; *a fortiori*, so does decisive personal ruin in hell.

Others (notably Hick 1966) embrace a doctrine of universal salvation: if ante-mortem horrors remain undefeated between birth and the grave, education will continue after death, probably in a series of careers, until the soul is perfected and brought into intimacy with God. Thus, God does guarantee each created person an overall existence that is a great good to them on the whole, one in which participation in horrors is balanced by the incommensurate

goodness of intimacy with God. Are such horrors likewise defeated within the context of the individual's existence? The stout of heart might say 'yes', because participation in horrors that remain undefeated within the individual's ante-mortem career contributes to the sense of mystery that makes a positive contribution to the soul-making of others. Since one is at least the agent-cause of the willy-nilly sacrifice of one's ante-mortem good, participation in horrors would constitute some sort of shift from self- to other- or God-centredness after all. Even if this putative positive dimension of participation in horrors is swamped by its negative aspect when considered within the framework of the individual's ante-mortem career, it provides a means for participation in horrors to be integrated into the overall development that gives positive meaning to the individual's life and thus defeated within the context of the individual's existence as a whole.

Some (notably M.M. Adams) contend, on the contrary, that divine goodness to created persons would do more to lend positive meaning to any careers in which they participate in horrors. The sacrifice involved in participation in horrors is pedagogically inept as a first lesson because it can damage the person so much as to make much further ante-mortem progress from self- to other- or God-centredness virtually impossible. This combines with the delay in gratification to another or perhaps many lives later to de-emphasize the importance of this life, leaving the impression that it would have been better skipped by those whose spiritual development was significantly set back through participation in horrors. To give this life, or any career involving participation in horrors, positive significance, some parameter of positive meaning other than contribution to soul-making must be found. Given two further assumptions - that divine metaphysical goodness is infinite, and that intimacy with God is incommensurably good for created persons - the mystical literature suggests several ways for participation in horrors to be integrated into the created person's relationship with God, ranging from divine gratitude for one's earthly career to various types of mystical identification between God and creatures in the midst of horrors. Because the identification occurs in this life and divine gratitude is for this life, they add positive significance to this life even where the creature has no ante-mortem but only postmortem recognition of these facts.

5 Methodological notes

Much contemporary discussion of BPW and free-will defences has addressed itself to the logical problem of evil because we seem epistemically in a better position to assess the compossibility of logically possible reasons with various conceptions of perfect goodness than to pronounce on what God's actual reasons are. In identifying logically possible defeaters, many of the earlier discussions (including those by Pike (1963) and Plantinga (1974)) confine themselves to a religion-neutral value theory, the better to answer the atheologian on their own turf. By contrast, soul-making, mystical and other explanatory theodicies draw on the resources of revelation for their speculations about God's actual reasons for the evils of this world and usually address their remarks in the first instance to the believing community. The distinction between these approaches blurs when attention is riveted on the concrete logical problem of evil - that is, on the logical compossibility of God with evils in the amounts and of the kinds and with the distribution found in the actual world. In so far as the consistency of actual religious belief is at stake, it becomes highly relevant to test the reasons supplied by revelation for logical compossibility with the existence of evils and the goodness of God. Where they pass, they can be advanced as solving the concrete logical problem of evil, whether or not their truth can be proved to the atheologian.

Once the wider resources of the religions under attack are allowed to interpret (I) and (II), it becomes clear that explanatory reasons come in two broad types: reasons why God causes or permits evils, and does not prevent or eliminate them; and explanations as to how God could be good to created persons despite their participation in evil. Reasons-why identify some great-enough good with which evils are necessarily connected, while reasons-how specify ways God could defeat evils in which the created person has participated and thus give that person a life that is a great good to them on the whole. Much philosophical discussion (Swinburne is particularly insistent on this point) presupposes that the problem cannot be solved without sufficient reasons-why. The criticized religions arguably take a mixed approach. Assuming that what perfect goodness can permit or cause is a function of what it can defeat, they combine partial reasons-why with elaborate scenarios by which God defeats even the worst horrors.

6 The evidential problem of evil

Recently many philosophers (notably Rowe, Alston, van Inwagen and Wykstra) have concluded that the most serious version of the problem of evil concerns not the logical but the evidential relation between (I) and (II). The

mere logical possibility that a student has broken all four limbs and been hospitalized for a heart attack will win them no extension of essay deadlines if the tutor can see that the student is in fact physically sound. Likewise, the evidential argument contends, many actual evils - such as the slow, painful death of a fawn severely burned in a forest fire started by lightning - appear pointless, in the sense that our composite empirical evidence constitutes strong reason to believe they have no point. But an omniscient, omnipotent being could have prevented some of them, while a perfectly good being would not allow or cause any of them it could avoid. Therefore, (II) concretely construed constitutes decisive evidence against (I).

Once again, replies could take the piecemeal approach, trying to show for each type of very intense suffering that it has a discernible point after all. It would not be necessary to complete the process to undermine the evidential argument. Success with some important cases would increase the probability that defeating goods are also present in other cases where we have not discovered any (see Suffering §4).

The favourite response (for example, by Wykstra, Alston and van Inwagen) attacks the argument at its epistemological foundations. The contention is that our composite empirical evidence could constitute strong reason to believe some actual evils pointless only if our cognitive powers would afford access to any point such evils might have were they to have one. If things would seem roughly the same to us (that is, if our evidence would be roughly the same) whether or not such evils had a point, the fact that we detect no point is not good evidence that there is no point. In particular, we are in no position to see that many instances of intense suffering are not explained by some of the reasons appealed to in traditional theodicies.

Defenders of the evidential argument (notably Rowe) grant the appeal of the underlying evidential principle, but relocate the disagreement in the richness of the theological hypothesis on which one draws. They argue that if one restricts oneself to a straightforward philosophical reading of (I), then it is likely that the situation with regard to intense suffering would be different in ways discernible by us. Expanded theism might import assumptions about the hiddenness of divine providence, mystical identification with suffering creatures, etc., but deploying these resources in the evidential debate carries a cost, because the prior probability of expanded theism is lower than that of (I).

This last point holds only if the richer theological theory is advanced as true. If instead it is used, as with the logical problem, to generate possible - this time not merely logically but epistemically possible - explanations, then no dilution in prior probabilities need be accepted. And once again, the more epistemically possible explanations there are, the greater the probability that the suffering in question is not pointless.

See also: Evil; God, arguments for the existence of; Goodness, perfect; Holocaust, the; Liberation theology; Religion and epistemology

MARILYN McCORD ADAMS

References and further reading

Adams, M.M. (1989) ‘Horrendous Evils and the Goodness of God’, *Proceedings of the Aristotelian Society*, supplementary vol. 63: 297-310.(Argues that in view of horrendous evils, global and generic hypotheses do nothing to defend divine goodness to created persons.)

Adams, R.M. (1972) ‘Must God Create the Best?’, *Philosophical Review* 81: 317-32. (Elaborates the argument referred to in §2 that a perfectly good God would not be obliged to create the best.)

Adams, R.M. (1977) ‘Middle Knowledge and the Problem of Evil’, *American Philosophical Quarterly* 14: 109-17.(Rejects middle knowledge and argues that counterfactuals of freedom cannot be true, as discussed in §3.)

Adams, R.M. (1985) ‘Plantinga on the Problem of Evil’, in J.E. Tomberlin and P. van Inwagen (eds) *Profiles: Alvin Plantinga*, Dordrecht: Reidel, 225-55.(Criticizes Plantinga’s treatment of the problem of evil, especially issues about counterfactuals and probability.)

Alston, W.P. (1991) ‘The Inductive Argument from Evil and the Human Cognitive Condition’, *Philosophical Perspectives* 5: 26-67.(Responds to evidential arguments by arguing that our composite empirical evidence does not constitute good reason to think none of the traditional theodicies are true.)

Chisholm, R. (1968) ‘The Defeat of Good and Evil’, *Proceedings of the American Philosophical Association* 42: 21-38.(Distinguishes balancing-off from defeat and assesses its relevance to the problem of evil.)

Griffin, D.R. (1981) ‘Creation out of Chaos and the Problem of Evil’, in S. Davis (ed.) *Encountering Evil: Live*

Options in Theodicy, Atlanta, GA: John Knox Press, 101-23, 125-36.(Draws on process philosophy to extend a free-will defence to natural evils.)

Hick, J. (1966) *Evil and the God of Love*, San Francisco, CA: Harper & Row, 2nd edn, 1978.(Criticizes 'Augustinian' theodicies that rely on the doctrine of the Fall, and develops a soul-making theodicy in the spirit of Irenaeus.)

Mackie, J.L. (1955) 'Evil and Omnipotence', *Mind* 64: 200-12.(The most influential formulation of the atheological argument from evil, slightly modified in §1.)

Pike, N. (1963) 'Hume on Evil', *Philosophical Review* 72: 180-97.(Articulates an epistemic defence using a BPW approach discussed in §2.)

Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Clarendon Press, ch. 9, §10, 191-3.(Develops the possible-worlds version of the free-will defence discussed in §3 of the present entry, with considerable attention to counterfactuals of freedom and the hypothesis of transworld depravity.)

Rowe, W.L. (1979) 'The Problem of Evil and Some Varieties of Atheism', *American Philosophical Quarterly* 16: 335-41.(Formulates the most discussed version of the evidential problem.)

Rowe, W.L. (1984) 'Evil and the Theistic Hypothesis: A Response to Wykstra', *International Journal for the Philosophy of Religion* 16: 95-100.(Replies to Wykstra's objection.)

Rowe, W.L. (1991) 'Ruminations about Evil', *Philosophical Perspectives*, 69-88.(Further reflections on his version of the evidential argument.)

Stump, E. (1985) 'The Problem of Evil', with comment by M. Smith and Stump's reply, *Faith and Philosophy* 2: 392-435.(Develops a theodicy based on the doctrine of the Fall.)

Swinburne, R. (1983) 'A Theodicy of Heaven and Hell', in A. Freddoso (ed.) *The Existence and Nature of God*, Notre Dame, IN: University of Notre Dame Press, 37-54. (Offers a free-will defence of annihilation or benevolent alternative environments for those who die with characters unfit for heaven.)

Swinburne, R. (1987) 'Knowledge from Experience and the Problem of Evil', in W.J. Abraham and S.W. Holtzer (eds) *The Rationality of Religious Belief: Essays in Honour of Basil Mitchell*, Oxford: Clarendon Press, 141-67. (Argues that natural evils contribute to a good environment for humans' self-determined choices of their own destinies.)

Van Inwagen, P. (1991) 'The Problem of Evil, the Problem of Air, and the Problem of Silence', *Philosophical Perspectives* 5: 135-65.(Develops a response to the evidential argument.)

Wykstra, S.J. (1984) 'The Humean Obstacle to Evidential Arguments from Suffering: On Avoiding the Evils of "Appearance"', *International Journal for the Philosophy of Religion* 16: 73-94.(Formulates a principle of epistemic access and uses it to attack Rowe's evidential argument.)

Evolution and ethics

The fact that human beings are a product of biological evolution has been thought to impinge on the study of ethics in two quite different ways. First, evolutionary ideas may help account for why people have the ethical thoughts and feelings they do. Second, evolutionary ideas may help illuminate which normative ethical claims, if any, are true or right or correct. These twin tasks - explanation and justification - may each be subdivided. Evolutionary considerations may be relevant to explaining elements of morality that are culturally universal; they also may help explain why individuals or societies differ in the ethics they espouse. With respect to the question of justification, evolutionary considerations have sometimes been cited to show that ethics is an elaborate illusion - that is, to defend versions of ethical subjectivism and emotivism. However, evolutionary considerations also have been invoked to justify ethical norms. Although there is no conflict between using evolution both to explain traits that are universal and to explain traits that vary, it is not consistent to claim both that evolution unmasks ethics and *justifies particular ethical norms.*

1 Explanation

Before we can assess whether evolution helps explain why we have the ethical beliefs we possess, we need to be clear on what the phenotype in question is. What part of human culture is delimited by the term ‘ethics’?

Some of the evaluations an organism may place on an action or a person or a motive may count as ethical in character. What makes an evaluation ethical? The question here is not the normative question, ‘What standards *should* an individual use?’ Rather, it is descriptive; its answer helps define a line of inquiry in the social and natural sciences. If we visited an alien society - either that of human beings or of some other species - how would we tell whether that society has an ethics?

Even if the individuals concerned have altruistic motives, this does not show that they possess an ethics. Parents may care for the welfare of their children, but feeling love and concern are not the same as possesing a morality. Closer to the mark is the idea that an ethics is a set of principles that possesses a certain sort of generality (see Morality and ethics).

In his *Methods of Ethics* (1874), Sidgwick suggested that it is a self-evident truth of common-sense morality that if one person has a right to do something, then so does anyone else who is similarly situated. This normative claim has a descriptive analogue: an organism has a morality only if it possesses a set of principles that dictate what anyone fitting a certain qualitative description is entitled or required to do. These principles need not provide an algorithm for deciding which actions are morally right; they simply may specify a set of ethically relevant considerations. Nor does this idea require that a morality treat all human beings as having equal ethical standing. Tribal moralities, no less than universal moralities, state what individuals must be like to be entitled to perform various actions. Nor does the Kantian pedigree of the idea mean that ethics focuses exclusively on disinterested reason and has nothing to do with engaged emotion (see Kant, I. §§9-10); an ethics of caring about personal relations may still find expression in a set of general principles.

Although generality is arguably a necessary feature of a morality, it is not sufficient. The hypothetical imperatives that specify rules for different aspects of cultural practice - of etiquette, of sports, and so on - may also be general. Ethics may be special because it dictates what individuals ought to care about, no matter what else they happen to want.

Even after we single out the aspect of a culture that falls under the rubric of ‘ethics’, we must take care to describe this phenotype accurately. A person’s or a society’s ethics may fail to coincide with the slogans that figure in public discourse. Aliens visiting Earth from another galaxy would make an enormous mistake if they thought that the people who pay homage to the Ten Commandments believe that the acts listed are always prohibited or always required. An ethics may be compared to a grammar; just as people speak a language without being able to articulate the grammatical principles their language obeys, so their ethical intuitions correspond to a set of principles that they often are unable to articulate. Descriptive ethics poses problems that are no less subtle than those addressed in theoretical linguistics.

Regardless of how exactly a morality is defined, the fact that moralities are so common in human societies raises a

fundamental evolutionary puzzle. An individual may possess a mixture of selfish and altruistic motives without having a morality. Why, then, do moralities exist? It is worth considering the possibility that moralities began life and continue to exist because they perform a certain social function. The idea that there are obligations, prohibitions and entitlements that people must respect regardless of whatever else they happen to want obviously contributes to a well-functioning society. Perhaps everyone benefits from a society in which a morality is present; alternatively, it may be that some benefit at the expense of others. Each of these patterns can be accommodated in an evolutionary explanation of why morality emerged as a social form in ancestral populations.

In addition to asking why moralities exist, evolutionary theory can be asked to explain why moralities have the features they do. But here we must be careful. To ask whether evolution can explain ‘our ethics’ is to begin with a poorly formulated problem. There are many features of the ethical norms endorsed by a person or a society. Some of these may have important connections with evolutionary considerations; others may not.

In virtually every human society, killing a member of one’s own society is viewed as a morally more weighty act than killing a chicken. It is hard to believe that evolution is irrelevant to why this is so. On the other hand, the fact that many societies have changed their views about capital punishment in the last century may be explicable by cultural, rather than evolutionary, factors.

This is not to deny that having views about capital punishment requires an organism to have a big brain, and that having a big brain is a product of evolution. The point remains, a big brain is something that people of today share with their ancestors of a hundred years ago; thus it does not explain why our views on capital punishment differ from those held a few generations ago.

The present example about the ethics of killing should not be taken to imply that evolution can help explain only those features of ethics that are culturally universal. Evolution has endowed human beings with enormous behavioural plasticity; however, sometimes the details of an organism’s plasticity are themselves dictated by evolution. For example, polar bears grow thick fur when it is cold and thinner fur when it is warm; natural selection has made the organism’s phenotype respond to environmental changes in ways that maximize fitness. So the mere fact that human beings espouse different ethical systems in different cultures does not imply that evolution is irrelevant to understanding that variation. The question is whether the pattern of variation can be understood as a phenotypically plastic adaptation.

2 Genetic arguments

If morality emerged and persisted because it conferred some evolutionary benefit, then we may speculate that morality would have had this utility even if there really were no such things as the obligations and rights that people objectively possess. It is worth remembering that ethics and religion were closely connected in the histories of many societies. Although Western philosophy since the Greeks prominently featured the idea that ethical principles have a status independent of divine decree, this declaration of independence is hardly a cultural universal. If theism emerged and persisted for reasons having nothing to do with its objective correctness, one may ask whether the same is true of moral convictions.

Here we find the beginning of a genetic argument for ethical subjectivism. The idea is that the causes of a conviction - the factors that generated it - provide evidence about whether the conviction is true. Arguments to this effect are called genetic because they refer to an idea’s genesis; they need have nothing to do with chromosomes. The question we need to consider is whether the fact that ethical beliefs can be explained - either by evolution or by details concerning human culture - show that these beliefs are never true or right or correct.

We must be careful not to dismiss this suggestion simply by invoking the so-called ‘genetic fallacy’. To be sure, one cannot deduce that a proposition is false just from information concerning why someone came to believe it. Even if Kekulé first thought of the benzene ring while he was hallucinating, this does not prove that benzene is not a ring.

Even so, facts about the genesis of a belief can provide non-deductive evidence about whether the proposition believed is true. If you form your belief concerning how many planets there are by drawing a number at random from an urn, then your belief is probably mistaken. The *process* of belief formation provides an indication of whether its *product* is likely to be correct.

If evolution produced some of the ethical convictions we have, what does that show about whether those convictions are likely to be true? Is this fact about their genesis evidence that the beliefs are false? It is easy to see how a genetic argument could show that a specific ethical belief is probably mistaken. If your belief about the number of planets can be cast in doubt by a genetic argument, there is no reason why your ethics should be immune from this challenge.

What is more puzzling is how the whole status of ethics as a subject can be undermined by genetic considerations. In the case of your belief about the number of planets, we assume that there is a fact of the matter concerning how many planets there are. We further assume that the process of drawing a number at random from an urn is not connected in the right way to that fact. This set of background assumptions explains why we are entitled to be confident that your belief about the number of planets is probably untrue.

For a genetic argument to support ethical subjectivism, it would have to be shown that the process of evolution is completely unrelated to the factors that make an ethical proposition true or false. But even if this could be demonstrated, it would not follow that no normative statements are true, only that the ones we happen to espouse are probably mistaken. How could a genetic argument support the conclusion that ethics, in its entirety, not just in the details we happen to believe now, is an illusion? Subjectivism and emotivism, if they are to be defended, must be defended by some other route (see Moral Scepticism).

3 The is/ought distinction

Hume (§4) famously observed that a normative conclusion cannot be deduced validly from purely descriptive premises. One cannot derive an *ought* from an *is*. This claim is not undermined by the fact that terms in natural languages sometimes combine normative and descriptive elements. To say that someone is cruel is both to describe and criticize the person in question. However, that does not affect the Humean thesis.

It follows from this Humean doctrine that descriptive facts about evolution cannot be sufficient to deduce normative conclusions. The fact that human beings in ancestral populations were omnivores does not provide an ethical justification for eating meat. It is ‘natural’ for human beings to eat meat, in the sense that this is a trait that is found in nature. Nothing normative follows from this point.

The Humean thesis concerns deduction. Even if an ought-statement cannot be deduced from purely is-premises, the question remains of whether descriptive premises provide non-deductive support for normative conclusions. Here the Humean thesis should be extended to the domain of non-deductive inference; purely is-premises do not, by themselves, provide evidence for an ought-conclusion.

It does not follow that evolutionary information is always irrelevant to ethical problems. Evolutionary facts may be relevant, even if they do not provide the whole story. Whether this is so depends on the specific ethical system one considers and on the details of the ethical problem at hand. In just the same way, human psychology is relevant to ethics only in so far as some ethical theory makes this so. For example, hedonistic utilitarianism says that an action’s moral standing is determined by its propensity to cause pleasure and pain. If psychology tells us about the pains and pleasures that an action is apt to cause, this information is ethically relevant because utilitarianism (or some other ethical theory of interest) says that it is.

One ethical idea that has been thought to establish the relevance of evolution to ethics is the ‘ought-implies-can’ principle. This principle says that obligations encompass only what is possible; we are not obliged to perform actions that are impossible for us to perform. If evolutionary considerations establish that certain types of behaviour are outside the human repertoire, then the idea of requiring them would not only be hopelessly utopian - such reforms also would fail to be ethically imperative.

This concept sometimes surfaces in discussion of the idea that natural selection will lead females to evolve the inclination to be more choosy about their sexual partners than males; females may have more to lose in fitness terms by making a poor choice. If evolution has hard-wired a sexual double standard into our ethical intuitions, then it will be false that we ought to change the way we think. However, a moment’s reflection shows that a double standard does not strike all people everywhere as obviously correct. For many species, natural selection has forged the details of behaviour; but for human beings, it has caused a big brain to evolve. This big brain allows us to reflect on what would be for many other species an automatic reflex. Paradoxically enough, natural selection has

caused a trait to evolve in human beings that limits the hegemony of natural selection in determining human behaviour. The human brain is a monkey wrench of the first order; apparently, it frequently prevents the ought-implies-can principle from demonstrating that evolutionary theory has a great deal to say about what our obligations are.

Still, ethics since Aristotle has repeatedly found attractive the thought that reflection on the biological nature of human beings must yield important information concerning what human flourishing amounts to, and a conception of human flourishing must be an important tool in forging theories of the right and the good. The more appealing such theories become, the more relevant to ethics will evolutionary ideas appear. There is no way to settle, once and for all, what evolution can teach us about ethics. As ethical theory evolves, the relevance of evolution will also experience a continuing metamorphosis.

See also: Human nature; Spencer, H.; Sociobiology

ELLIOTT SOBER

References and further reading

Alexander, R.D. (1987) *The Biology of Moral Systems*, New York: de Gruyter.(A systematic argument for the fruitfulness of evolutionary explanations of moral practices.)

Huxley, T.H. (1894) *Evolution and Ethics*, Princeton, NJ: Princeton University Press, 1989. (Argues that evolution often produces morally repulsive outcomes, which human culture should attempt to modify.)

Kitcher, P. (1994) 'Four Ways of Biologicizing Ethics', in E. Sober (ed.) *Conceptual Issues in Evolutionary Biology*, Cambridge, MA: MIT Press, 2nd edn, 439-50.(A criticism of Ruse and Wilson (1986).)

Ruse, M. (1986) *Taking Darwin Seriously*, Oxford: Blackwell.(A systematic defence of the relevance of evolution to ethics and epistemology.)

Ruse, M. and Wilson, E. (1986) 'Moral Philosophy as Applied Science', *Philosophy* 61: 173-92; repr. in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*, Cambridge, MA: MIT Press, 2nd edn, 1994, 421-38.(A defence of the relevance of evolution to both meta-ethics and normative ethics.)

Sidgwick, H. (1874) *Methods of Ethics*, London: Macmillan.(A classic text, which examines intutionism, egoism and utilitarianism.)

Sober, E. (1994) 'Prospects for Evolutionary Ethics', in *From a Biological Point of View*, Cambridge: Cambridge University Press.(Assesses how evolutionary theory bears on the tasks of explanation and justification in ethics.)

Evolution, theory of

The biological theory of evolution advances the view that the variety and forms of life on earth are the result of descent with modification from the earliest forms of life. Evolutionary theory does not attempt to explain the origin of life itself, that is, how the earliest forms of life came to exist, nor does it apply to the history of changes of the non-biological parts of the universe, which are also often described as 'evolutionary'. The mechanisms of natural selection, mutation and speciation are used in evolutionary theory to explain the relations and characteristics of all life forms. Modern evolutionary theory explains a wide range of natural phenomena, including the deep resemblances among organisms, the diversity of life forms, organisms' possession of vestigial organs and the good fit or 'adaptedness' between organisms and their environment.

Often summarized as 'survival of the fittest', the mechanism of natural selection *actually includes several distinct processes. There must be variation in traits among the members of a population; these traits must be passed on from parents to offspring; and the different traits must confer differential advantage for reproducing successfully in that environment. Because evidence for each of these processes can be gathered independently of the evolutionary claim, natural selection scenarios are robustly testable. When a trait in a population has arisen because it was directly selected in this fashion, it is called an* adaptation*.*

Genetic mutation is the originating source of variation, and selection processes shape that variation into adaptive forms; random genetic drift and various levels and forms of selection dynamic developed by geneticists have been integrated into a general theory of evolutionary change that encompasses natural selection and genetic mutation as complementary processes. Detailed ecological studies are used to provide evidence for selection scenarios involving the evolution of species in the wild.

Evolutionary theory is supported by an unusually wide range of scientific evidence, gaining its support from fields as diverse as geology, embryology, molecular genetics, palaeontology, climatology and functional morphology. Because of tensions between an evolutionary view of homo sapiens *and some religious beliefs, evolutionary theory has remained controversial in the public sphere far longer than no less well-supported scientific theories from other sciences.*

1 Basics of evolutionary theory

According to the biological theory of evolution the variety and forms of life on earth are the result of descent with modification from the earliest forms of life. The theory does not attempt to explain the origin of life itself (see Life, origin of), nor does it apply to the history of changes of the non-biological parts of the universe, which are also often described as 'evolutionary'. A number of European thinkers in the seventeenth and eighteenth centuries advocated an evolutionary view of life, but it is Charles Darwin and Alfred Russel Wallace who offered the mechanism, *natural selection*, which is accepted as a primary process by which evolutionary change is effected.

At the time of the proposals of evolutionary theory presented in Darwin's book, *On the Origin of Species* (1859), and Wallace's article 'On the tendency of varieties to depart indefinitely from the original type' (1858), there was no adequate naturalistic (non-miraculous) explanation for the diversity of life forms, their adaptations to their environments, their possession of vestigial organs, the resemblances among them, or their history. Evolutionary theory gives unified and naturalistic explanations for these features of life and for a wide range of other biological phenomena.

The problem of mechanism in evolutionary biology is to explain how the originating group of organisms has led, through descent with modification, to other groups of organisms with very different characteristics from those of the ancestral group. The mechanism of natural selection requires a particular type of set-up: in the standard case, the organisms in a particular population must vary among themselves in some trait that is important to surviving and reproducing in their environment; the differences in that trait must be inheritable to some degree; and the different versions of the trait must actually make a difference in that environment to the success of reproducing and raising offspring to reproductively successful adulthood. Given this set-up, and given the continual importance of that feature of the environment which advantages some versions of the trait over others, the result is change in the traits of the population over time, that is, evolutionary change.

A simple illustration: take a population of brown bears living in the Arctic. They vary in the darkness of their fur, and their offspring tend to resemble their parents in fur colour. Most importantly, they are hunters living in an environment with little cover; prey find it more difficult to escape those with lighter fur, as they see these predators less well against the arctic landscape. Hence, those bears with lighter fur will on average be more successful in hunting, less likely to starve, more likely to reproduce and provide for their young than are those with darker fur. Over tens of thousands of generations of this cyclic process, this population of bears will become lighter and lighter in colour. Note that this explanation assumes a number of facts: heritable variations in fur colour continue to arise; the bears choose to stay in an arctic environment; the arctic environment itself remains snowy and icy for a significant portion of the year; and the prey populations continue to respond differently to lighter and darker predators.

The basic selection process incorporates several distinct biological processes, including heredity, reproduction and the production of variation, as well as the ecological and behavioural processes involved in differential survival and reproduction. Each of these components of evolutionary theory has been investigated, developed and empirically confirmed. Perhaps the most significant successes have been in the science of genetics, which gives accounts of mutation and genetic recombination as causes of the variation on which selection acts, as well as offering insight into the mechanisms and measurements of genetic drift. The most serious deficiency of Darwin's own view of evolutionary change was its lack of a well-confirmed theory of heredity and variation. Modern genetics has provided that theory, and has become a thriving science since the discovery of the structure of the hereditary material, DNA, by Francis Crick and James Watson in 1953. The discovery and investigation of the basic biochemistry of heredity, involving DNA and RNA, has confirmed the similarities among living things at the most fundamental levels, and has also led to huge bodies of evidence that can be used to compare and differentiate lineages (see Genetics; Molecular biology).

The modern view is not that every detail of evolutionary change can be attributed to natural selection within species; genetic drift, in which forms fluctuate or persist by chance, also leads to substantial genetic variation within and among species. Evolutionary phenomena, then, are understood as consequences of the actions of drift and selection on various levels of biological entities, including genes, organisms and groups of organisms.

Moreover, the existence of mutation and the operation of natural selection do not, by themselves, suggest that new species could arise, nor do they account for the increased diversification of species that we see in the history of life. Given only the facts mentioned in the scenario, our white bears, considered above, will not necessarily comprise a new species of bear. Thus, some attention to the nature of speciation is needed.

Darwin suggested that speciation occurred when there was relatively rapid change in the environment; these different environmental pressures, operating on variations that arose over the generations, could eventually result in a divergence of the descendent group from the ancestral one that was substantial enough that they could or would no longer interbreed. Thus, reproductive isolation is considered a primary cause and result of speciation in sexually reproducing species.

Ernst Mayr, a leading evolutionary biologist, has emphasized a crucial conceptual change (1982) that was involved in changing to the modern evolutionary view of life, namely, the switch from '*typological*' thinking - or thinking of organisms and species in terms of their essences or basic types - to '*population*' thinking, in which species are irreducible populations of unique individual organisms, and in which the *variability* among individuals is of the most fundamental importance. The difference between the evolutionary view of species and the essentialist view is thus not primarily that evolutionary species change over time, while essentialist species do not. Individual organisms in modern evolutionary theory are not members of their species in virtue of their sharing some basic group of properties or traits, or because they belong to a particular type or have certain features. On an evolutionary view of species, each organism is unique in its non-trivial properties, and there is no average type or typical specimen of any species. This is not to say that particular individuals cannot exhibit some calculated average height, colour or weight; the point is rather that all such averages over individuals must be interpreted as abstractions, and not even as temporary essential features of that species (see Species).

Developments in evolutionary genetics helped to clarify the importance of thinking in terms of populations. Mathematical population genetics, first developed by R.A. Fisher, J.B.S. Haldane and Sewall Wright, explored mathematically the issue of which evolutionary changes are possible under what conditions and how quickly

genetic change could occur. They examined the relative effectiveness and roles of mutation and selection in evolutionary change and, generally, developed models of evolution under a wide variety of conditions and assumptions.

In the late 1930s and early 1940s a general approach called ’the evolutionary synthesis’ expressed the newly united views of population geneticists, systematists and palaeontologists. Genetic mutation was seen as the originating source of variation, and selection as the shaper of that variation into adaptive forms; random genetic drift and various levels and forms of selection dynamic developed by the geneticists were integrated into a general theory of evolutionary change that encompassed both natural selection and genetic mutation as complementary processes. Detailed ecological studies were used to provide evidence for selection scenarios involving the evolution of species in the wild. Subsequently, synthesis of the fields of population genetics and ecology created the field of population biology, which studies such phenomena as population dynamics and ecological diversity in their effects on the genetic characteristics of populations.

In sum, contemporary evolutionary biology addresses a broad range of issues, all involving processes that produce key aspects of living beings, including their history, forms, functions, distribution, and their ways of earning a living and reproducing. Evolutionary explanations involve an interlacing of theories and evidence from an array of scientific subfields, including molecular and population genetics, anatomy and morphology (the study of life forms), palaeontology and ecology.

2 The nature of the supporting evidence

Darwin’s evolutionary claims involve two central theses: that present and past forms of life are all descended from one or few primitive life forms (descent with modification); and, second, that a primary mechanism of evolutionary change is the mechanism of natural selection. The major kinds of evidence supporting each of these claims are reviewed in this section.

Descent with modification. The field of palaeontology provides some of the most important evidence for descent with modification. In Darwin’s time, geologists recognized that there are patterns of geological formations which are parallel the world over, each containing distinctive sets of fossil species; these patterns helped establish the progression of the history of life. Two features of the fossil record were especially important: the fact that each animal and plant form appeared at certain times and certain places only, showing a pattern of increased diversification of life forms; and the evidence that extinction occurred. Moreover, the more ancient a form was, the more markedly it differed from living forms, and fossils from adjacent formations resemble one another far more than those from formations remote in time. The geographical distribution of fossils was also significant; extinct forms on a continent are closely related to living forms there. All of these facts support the view that the extinct forms of life are the ancestors of the living ones, and that modifications in form resulted in the diversification of life.

Morphology, the study of plant and animal form, also provides vital support for descent with modification. All mammals share a common form - as do all birds, all fish and all insects. These basic anatomical similarities within groups are evidence that they descended from a common ancestor. Admittedly, there are variations within the basic form: a bat’s arms now appear as wings, while a dog’s forelegs and a chimpanzee’s arms - defined as ’homologous’ because they are derived from the same character of the mammalian ancestor - serve for running and climbing respectively. The fact that a wide variety of forms of forelimb are variations on a common pattern is evidence that a common ancestral population evolved and diversified. Microstructures of organisms can also serve as evidence for evolution. The fact that cells share many basic structural properties across all plants and animals provides evidence for genealogical affinities among organisms. Perhaps most powerfully, the structure of some macromolecules, for example, specific proteins, has been useful in determining the relations among different groupings of organisms; resemblances among proteins has been used to help reconstruct the history of the evolution and diversification of different forms of life, that is, phylogenetic trees.

Embryological morphology, especially, provided important evidence for common ancestors and a continuous process of branching. Darwin emphasized the differences between the embryonic and adult forms of many animals as well as the very close resemblances of embryos within the same class, arguing that ’ community in embryonic structure reveals community of descent’ (1859: 449). The fact that land-dwelling vertebrates all go through a

gill-arch stage in embryo, for example, is explained by their having descended from fish. Vestigial organs similarly provided strong support for common descent. The presence of the remnants of a pelvis and hind limbs in whales, for example, was explicable through their descent from land mammals.

Finally, the facts of geographic distribution provided particularly important evidence for common ancestors. Darwin emphasized certain features of oceanic islands and their differences from continental islands; oceanic islands have many fewer species, and are typically missing certain types, such as terrestrial mammals and freshwater fishes. Oceanic island species also resemble most of the species from their nearest continent. The inhabitants of the Galapagos archipelago, for example, closely resemble species from the nearby tropical Americas, rather than those from similar temperate climates.

Natural selection. The details of the evidential support for the mechanism of natural selection need to be understood in terms of the structure of natural selection explanations. One remarkable feature of the theory of evolution by natural selection is its explanatory power; it seems that the basic selection framework can be used, when applied with imagination, to account for the emergence, transformation or elimination of *any* trait or population. This flexibility is, in fact, the source of the theory's capacity to underwrite so many hypotheses, and is a primary source of its fruitfulness as a scientific theory.

However, the *proposal* of a selectionist explanation of a trait or evolutionary change is never enough for its scientific acceptance; the hypothesis must be tested. Unfortunately, the traditional emphasis placed on prediction by philosophers of science has contributed to a widespread misperception of the scientific status of these selection hypotheses (Hempel 1965). Under a classical hypothetico-deductive view of theory testing, the theory is used to generate an *outcome* or prediction about the natural system being modelled. If the outcome turns out to hold in nature, the theory is confirmed (see Confirmation theory). In the case of natural selection explanations, however, the outcome - the frequency or nature of the evolved traits - is standardly already known, and the substance of the selection hypothesis involves an account of how the population or trait *got* that way. It may seem, therefore, that all selection explanations could be successfully confirmed as predictions, whether they are true or not.

The logic of testing selection hypotheses was clarified by Darwin: the various empirical claims that appear in any selection explanation - about the existence of variation of reproductive success with the possession of different traits, about the existence and character of environmental stresses, about the inheritance of the relevant variety of traits - must be *independently* confirmed empirically. This independent confirmation of the detailed empirical claims made in selection explanations comprises the chief form of empirical testing for specific evolutionary explanations (Lloyd 1994).

In evolutionary explanations, the abstract framework of selection is filled in with specific details of the particular population, traits and environments at issue; the *empirical content* of any explanation by selection consists of particular claims about particular organisms and their environments. These particular empirical claims are subject to robust evidential testing, independently of any selectionist explanation of which they are a part. There remains, however, a question about the evidential status of the selection framework itself: is it proper to consider the abstract framework well-confirmed, on the basis of its independently confirmed successes in accounting for particular cases of evolution by selection? We tend to regard general theories as well-confirmed if, taken in conjunction with particular facts, they are good at accounting for other particular facts (see §4).

We are now in a position to appreciate the range of evidence relevant to Darwin's theory of natural selection. Important empirical support for natural selection explanations comes from several of the fields that provide evidence for descent with modification. The investigation of homologies, from morphology for example, has provided significant instances of the evolution of special adaptations through natural selection. The bat forelimb - homologous to both the human arm and the canine foreleg - presents an intriguing problem for the theory: how and why did bat ancestors evolve from squirrel-like creatures without wings? The occurrence of random mutation in forelimb form, combined with strong selection on the resulting variation, is used to explain the divergence of bat species from other mammalian species. Similarly, special adaptations to local environments are found to explain a variety of specialized mammalian forelimb forms and functions. The evolution of hundreds of species of bat, all descended from a single ancestral population, is an example of adaptive radiation; specializations in particular prey and particular environments proved advantageous to the survival and reproduction of past bat groups. Even vestigial traits can support the effectiveness of natural selection. The mammalian ancestors of whales, who entered

the water around 55 million years ago, experienced little or no selection pressure to maintain working hindlimbs: mutations and variations in which the hindlimbs were reduced or nonfunctional may have been advantageous, or may have been linked to traits that were advantageous in the new aquatic environments, and subsequently became fixed in the population.

Finally, evolutionary systematics has provided abundant evidence on evolutionary trends and patterns of adaptations, as well as detailing the types of evolutionary transformation that organisms' features undergo, and the patterns of variation within species (Futuyma 1979, ch. 1).

3 Scientifically contested issues

There are currently several areas of controversy within evolutionary biology. One of the most vigorous involves issues of macro-evolution, that is, changes in diversity, the distribution of traits across lineages, extinctions of major taxa, and trends in evolution, such as increase in average size of organisms. The key theoretical questions revolve around the tempo and mode of evolutionary change: when exactly does most evolutionary change occur? Do forms of life evolve mostly gradually, or is evolutionary change concentrated in speciation events? Under what is called the gradualist view, patterns of macro-evolutionary change are explained completely by patterns of gradual selection within species. An important alternative set of mechanisms, advocated since the 1970s by Stephen Jay Gould (1989), Niles Eldredge (1985), Steven Stanley, and other palaeontologists, involves phenomena higher than the species' level, for example, rates of speciation and extinction; under this approach, large-scale events of evolutionary history are explained as consequences of drift of species and species selection (seen as analogous to drift and selection of alleles within populations.)

Another debate about mechanism and the explanation of patterns involves the role of biological development and the ultimate sources of organismal structure. Developmental biologists have been concerned to explore how mutations in genes affect changes in form; some forms are much more likely to arise than others, due to the biochemistry of how organisms grow and change. Selection processes always require the presence of *variation* in traits that matter to the survival and reproduction of genes in an environment. Investigations into which variations are likely and which are even possible can provide information crucial to selection explanations. In addition, the fact that some traits vary along with other ones simply in virtue of how the organism grows makes it vital to incorporate information about the complexities of development into evolutionary explanations.

A very visible and important set of debates has revolved around the levels and units of selection. Generally, disagreement concerns the relative importance of different mechanisms of selection that cause evolutionary change.

Finally, there are robust debates in phylogenetic systematics, some of which concern the detailed evolutionary history of particular species or groups of organisms, and some of which involve the proper role of evolution in phylogenetic systematics. Regarding the latter, Darwin made it clear that he thought that evolution could help systematics by making the classification of life fundamentally a genealogy (Mayr 1982) (see Taxonomy).

4 Philosophical worries - tautology

There is a philosophical concern about modern evolutionary biology which seems to be quite serious: is evolutionary theory based on a tautology? A tautology is a statement that is necessarily true because of the meaning of its terms. Briefly: the theory of natural selection is based on 'the survival of the fittest'; but which organisms are the fittest, according to evolutionary theory? Answer: those that survive. This is the logical truism that is frequently taken to represent the core of evolutionary theory.

Although this objection about tautology has been raised and answered in different ways since the nineteenth century, it continues to cause confusion in each new generation of thinkers facing evolutionary theory. There are two general strategies in answering it. The first is to highlight that, as a matter of fact, the 'fittest' are not defined in terms of survival by evolutionists (Hull 1974). The term 'fitness' has a potentially confusing multiplicity of meanings in evolutionary theory: it can refer to the expected rate of survival and reproduction of a certain type (usually defined in terms of genes); it can also refer to the 'adaptedness' of a certain type of trait to the pressures of its local environment - here, adaptedness is analogous to being well-suited or well-engineered for a specific task. When using this latter meaning of 'fitness', an evolutionist appeals to the suitability of a certain character for

coping with the challenges presented to its owner in the expected environments. (For example, the ’adaptedness’ of being insulated by blubber in arctic environments depends on its effectiveness at maintaining body temperature in mammals compared to other available options, for example, faster metabolism.) Using this substantive notion of the ’fitness’ of an organism possessing a particular trait in a particular environment thus involves no hint of tautology; not only is such a claim substantive, it may be independently testable. Hence, explanations of evolution by natural selection are not tautological.

The second approach to answering the tautology objection is to point out that, strictly speaking, many accepted scientific theories are in fact empirically empty, especially those expressed in mathematical form. A model of a system in Newtonian mechanics is, for example, purely mathematical, as are all predictions arising from the model. This does not prevent the theory from being scientific and empirical, because a proper understanding of the theory entails that the model *represents* some real system which it is used to explain. Similarly, models from mathematical population genetics may show that, as a matter of mathematical deduction, those organisms with higher fitness parameters will necessarily have a higher probability of surviving and contributing to future generations; the related empirical application of the model, however, is not guaranteed to be true. In other words, many mathematically expressed theories - from physics and from natural selection theory - are self-contained or tautologous when taken alone; when used in their scientific context, however, they have precise and testable empirical content (Sober 1993).

The tautology objection can be answered by calling attention to the role of the principle of natural selection - the survival of the fittest - within the rest of the theoretical and explanatory parts of evolutionary theory. When this is done, the objection appears to be based on a misunderstanding of how either the specific theory, or scientific theories generally, really do their work (see Theories, scientific).

5 Philosophical worries - prediction

One of the most common misperceptions of evolutionary theory is that it makes no predictions, concerning, as it does, the past history of life. But being about past events does not prevent a science from being predictive; there are numerous predictions from evolutionary biology which concern the fossil record. For example, whatever will be discovered in the fossil record in the future, mammals will not appear at a date prior to the evolution of bony fishes. Nor will hominids be found to coincide in time with the large dinosaurs. Nor will fossil birds predate the earliest amphibia. Evolutionary theory also underwrites many predictions about genetic findings; closely related species are predicted to share a larger percentage of their basic genetic material than less closely related species and groups. These genetic predictions, made on the basis of evidence preceding the discovery of DNA, have recently gathered abundant empirical confirmation.

There is a distinct concern about evolutionary explanations, which is that they depend so strongly on uncontrollable and, in most cases, unpredictable contingencies - such as environmental changes, chance mutations, or complex interactions among members of a species - that it is impossible for evolutionists to predict the direction that evolutionary processes will take a given population in the future. Evolutionists faced with the question, ‘Which way is humankind evolving?’, admit that evolutionary theory does not enable this type of prediction.

The mathematical models of modern population genetics do allow precise predictions to be made regarding, for example, the future frequencies of specific forms of genes (called ’alleles’) in a population of laboratory fruitflies. These successes are very significant in their support of the efficacy of natural selection, random genetic drift or mutation rates, as causes of evolutionary change. However, population genetics does not presently have the capacity to predict changes over the periods of time involved in major evolutionary events such as speciation, which take place on the order of tens of thousands of years.

More generally (as reviewed in §2) the overemphasis on prediction in understanding scientific evidence has contributed to a failure to understand the evidential relations on the basis of which evolutionary claims are scientifically evaluated.

6 Philosophical worries - design

Design seems to be everywhere in nature; this fact supported the efforts of eighteenth and nineteenth century churchmen, such as William Paley, to encourage piety through the study of nature (see Paley, W. §2). The

existence of order, complexity and adaptedness in living things is still the most common reason given by those who believe in direct supernatural creation of the myriad forms of life on earth. As reviewed in §1, the presentation, development and broad, detailed evidential support of the theory of evolution have provided a scientific and empirically tested alternative to the hypotheses of direct supernatural design. The empirical successes of the modern theory, its compatibility and interdependence with other natural sciences, and its fruitfulness in facilitating scientific investigations into life, have led to its scientific acceptance as an adequate non-supernatural account of the very features of life that were formerly taken to be explicable only by the actions of a supernatural creator (Dawkins 1986) (see God, arguments for the existence of §§4-5; Religion and science §§4-5).

The distinction between proximate and ultimate causes in biological explanations is useful in understanding goal-like states and the appearance of design. *Proximate* causes concern the functions of an organism and its parts, as well as its development. *Ultimate* or evolutionary causes concern why an organism is the way it is. Any biological phenomenon has two independent sets of causes, according to Ernst Mayr, ‘ because organisms have a genetic programme. Proximate causes have to do with the decoding of the programme of a given individual: evolutionary causes have to do with the changes of genetic programmes through time, and with the reasons for these changes’ (1982: 68). Thus, for example, the proximate causes of sexual differences between male and female fruitflies may be hormonal, while the selective advantage of differential utilization of the food niche may be the ultimate cause of these differences. The design of the system is explained through evolution by natural selection over time, while its local functioning is explained through local conditions.

7 Progress and goals

The mechanisms of evolution seem especially mysterious in their capacity to generate what appears to be *progress*. Reviewing the increased cranial capacity of the hominid lines, for instance, it can appear that there was some goal towards which the hominid group was progressing. However, the appearance of ’upward progress’ is, on the modern theory, misleading; there is no natural goal or state of perfection towards which any lineage has evolved (Gould 1989). The series of improvements manifested in any lineage is solely the produce of evolutionary processes operating locally. As Futuyma notes, ’ future conditions cannot affect present survival. The enduring variations may increase the organism’s complexity or behavioural repertoire, or they may decrease it. They may increase the likelihood of survival through subsequent environmental changes, or they may increase the likelihood of subsequent extinction’ (1979: 7).

See also: al-Afghani §2

ELISABETH A. LLOYD

References and further reading

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*, London: John Murray. (Revolutionary argument for both the evolution of varied forms of life from earlier ones, and the theory of evolution by natural selection. Accessible to general readers.)

Dawkins, R. (1986) *The Blind Watchmaker*, London: Longman.(Accessible defence of the powers of natural selection to create adaptations.)

Eldredge, N. (1985) *The Unfinished Synthesis*, New York: Oxford University Press. (Discussion and defence of operation of selection at levels above the organism.)

Futuyma, D.J. (1979) *Evolutionary Biology*, Sunderland, MA: Sinauer.(Outstanding and accessible text, with abundant examples and careful discussion of conceptual issues.)

Gould, S.J. (1989) *Wonderful Life*, New York: W.W. Norton.(Accessible discussion of the role of processes above the species level and of chance events in the evolution of life. Highly recommended.)

Hempel, C.G. (1965) *The Structure of Scientific Explanations*, New York: Free Press.(Classic presentation of scientific explanation and scientific confirmation using a logical positivist framework. For upper-level philosophy students.)

Hull, D.L. (1974) *Philosophy of Biological Science*, Englewood Cliffs, NJ: Prentice Hall.(Classic presentation and discussion of key philosophical issues in evolutionary theory; especially notable discussion of goal-orientated processes.)

Hull, D.L. (1980) ‘Individuality and Selection’, *Annual Review of Ecology and Systematics* 11: 311-32.(Extremely influential rethinking of conceptual issues in levels of selection debates. Intermediate level.)

Keller, E.F. and Lloyd, E.A. (eds) (1992) *Keywords in Evolutionary Biology*, Cambridge, MA: Harvard University Press.(Essays by biologists, historians, philosophers about evolutionary terms with multiple meanings. Suitable for novice to expert readers.)

Lloyd, E.A. (1994) *The Structure and Confirmation of Evolutionary Theory*, Princeton, NJ: Princeton University Press.(Fairly advanced philosophical analysis of modern evolutionary theory.)

Maynard Smith, J. (1989) *Evolutionary Genetics*, Oxford: Oxford University Press. (Widely-used text, with interesting examples.)

Mayr, E. (1982) *The Growth of Biological Thought: Diversity, Evolution and Inheritance*, Cambridge, MA: Harvard University Press.(Readable and comprehensive history of evolutionary biology; especially useful discussions of proximate and ultimate explanations, typological thinking, and a defence of evolutionary taxonomy.)

Ruse, M. (1979) *The Darwinian Revolution: Science Red in Tooth and Claw*, Chicago, IL: Chicago University Press.(Discussion of chief conceptual issues involved in the nineteenth-century context of Darwinism; excellent review of Darwin's evidence. Suitable for upper-level students.)

Sober, E. (1993) *Philosophy of Biology*, Boulder, CO: Westview.(Sophisticated discussion of key philosophical issues in evolutionary biology and systematics. Suitable for upper-level students.)

Wallace, A.R. (1858) 'On the Tendency of Varieties to Depart Indefinitely from the Original Type', *Journal of the Proceedings of the Linnaean Society (Zoology)* 3: 53-62. (One of the original statements of the theory of evolution by natural selection.)

Evolutionary theory and social science

Ideas from evolutionary theory impinge on the social sciences in two ways. First, there is the research programme of sociobiology, which attempts to demonstrate the impact of biological evolution on important features of human mind and culture. Second, there is the idea that biological evolution provides a suggestive analogy for the processes that drive cultural change. Both research programmes have tended to focus on the idea of natural selection, even though the theory of biological evolution considers processes besides selection. Sociobiology attempts to show that the following conditional helps explain psychological traits just as it applies to traits of morphology and physiology: if a trait varies in a population, makes a difference for the survival and reproduction of individuals, and is influenced by genetic factors, then natural selection will lead the trait to change its frequency in the population. Models of cultural evolution are built on an analogous conditional: if a set of alternative ideas are found in a culture, and people tend to find some of these ideas more attractive than others, then the mix of ideas in the culture will change. Sociobiology and the understanding of cultural change as an evolutionary process are approaches that have a history and both will continue to be explored in the future. Each is a flexible instrument, which may be better suited to some tasks than to others, and may be handled well by some practitioners and poorly by others. As a consequence, neither can be said to be 'verified' and 'falsified' by their track records to date.

1 What is biological evolution?

The term 'evolution' has a misleading etymology. It comes from the Latin *evolvere*, which means to unfold. In the eighteenth century, the term was used to name the process of ontogeny in which a single organism develops from a fertilized egg into an adult. Later, in the aftermath of Darwin's theory, the term itself underwent an evolution, shifting its application from ontogeny to phylogeny (see Darwin, C.). It now no longer describes the development of an individual organism, but applies to the process of 'descent with modification', whereby new species emerge from ancestral forms.

This transposition in subject matter did not leave the meaning of the term intact. It is natural to think that important aspects of the development of an organism are 'preprogrammed'. Although the environment obviously influences the organism's phenotype in many ways, a great deal of the developmental sequence is relatively invariant over environmental changes. Organisms in the same species often share a fixed sequence of developmental stages, each marked by a special constellation of distinctive traits.

Darwin's theory of evolution entirely rejects this view when it is applied to phylogeny. Although organisms may have a preprogrammed ontogeny, species do not have a preordained phylogeny. There is no such thing as the developmental stages through which a species must pass. Rather, the Darwinian idea is that evolution by natural selection is 'opportunistic'; the trajectory of change is an adventitious consequence of the environmental conditions that the species experiences and the variation that the species happens to contain. There is no inherent tendency for species to evolve in the direction of greater complexity, specialization, or body size. Evolution often moves in those directions, but often it does just the reverse.

Contemporary evolutionary theory retains the twin Darwinian ideas of descent with modification and natural selection as an important cause of diversity. The former idea is usually formulated by saying that all living things on earth, both now and in the past, are part of a single genealogical tree. Current species may be closely related to each other or related only distantly, but all are related. The status of the latter idea is the focus of the debate about adaptationism: how important is natural selection as a cause of the similarities and differences we see within and among species?

When biologists apply this two-part theory, they use the term 'evolution' with a restricted and technical meaning. It does not mean mere change; planets change, but this is not evolution in the biologist's sense. And even when the change occurs in the traits of a population of organisms, the term may still fail to apply. If the organisms in one generation are taller than their parents, this may simply reflect a change in the nutrition available. If the population has not changed its genetic composition, this will be change without evolution. Biologists do not restrict the term 'evolution' to genetic changes; traits of morphology, physiology, and behaviour evolve in the strict sense. However, these phenotypic changes must reflect genetic changes, if they are to be genuinely evolutionary.

2 Two approaches

Biology recognizes other possible causes of evolution besides natural selection. Mutation, migration, systems of mating, recombination, and random genetic drift are examples. However, since natural selection is the factor that has been of greatest interest to the social sciences, it is well to explore its logic a bit further.

If natural selection is to produce biological evolution in a population of organisms, there must be variation in fitness and heredity. Let us explore a simple example to see what these twin requirements mean. Consider a population of zebras in which faster running speed evolves. If this is due to natural selection, then the population must contain variation. So let us suppose that some zebras run faster, others slower, and that this makes a difference in the organisms' fitness - their capacity to survive and reproduce. In particular, we may imagine that running speed affects fitness because faster zebras are better able to evade predators than are slower ones. In addition to these facts about fitness, there also must be a process of heredity: offspring tend to resemble their parents in running speed, and this resemblance, at least in part, is due to genetic factors. These assumptions permit the phenotypic trait to evolve in the population; we expect the average running speed of the organisms in the population to increase.

When the concept of natural selection is applied to the subject matter of the social sciences, these two elements are interpreted either literally or metaphorically. The literal application gives rise to the research programme called sociobiology; the metaphorical application leads to the idea that cultural change involves competition among ideas in which the 'best' ideas survive. A standard example of sociobiological explanation concerns the existence of an incest taboo. Human beings seldom reproduce with close genetic relatives. In addition to this regularity in behaviour, there is also the fact that societies often have norms of prohibition. Sociobiologists propose to explain both the behaviour and the taboo by describing the fitness consequences of inbreeding and outbreeding. Reproduction with close relatives increases the chance that the offspring will possess two copies of deleterious recessive genes. As a result, there has been selection for avoiding such matings. The result is not just that human beings now avoid inbreeding, but that they have a genetic inclination to do so. The dual ideas of fitness and heredity are said to apply to incest avoidance in human beings just as literally as they apply to running speed in zebras.

A good example of the metaphorical application of the biological idea of natural selection is the economic theory of the firm (reviewed in Hirshliefer (1977). The problem is to explain why businesses are usually efficient profit maximizers. One hypothesis is that managers are rational and well-informed; they determine which strategies work best and then they adopt them. The theory of the firm offers an alternative explanation. Firms that are efficient stay in business; those that are not, go bankrupt. The economic market induces a selection process in which only the fit survive. Firms do not rationally adjust their behaviour; rather, the market adjusts the composition of the population of firms.

Notice how the concepts of fitness and heredity are interpreted in this theory. The idea is not that people who run efficient businesses live longer or have more children. The fitness of firms has nothing to do with the biology of survival and reproduction. In addition, the reason that firms in later years tend to resemble their former selves (that is, are relatively stable in the strategies they deploy) has nothing to do with genetic transmission. Although sociobiologists often suggest that incest avoidance is 'in the genes', theorists of the firm do not claim that economic behaviour is genetically encoded. The theory of the firm posits a selection process, but one that is related only metaphorically to the process of biological selection (see Sociobiology).

3 The need for a piecemeal assessment

Proponents and critics of sociobiology often treat that subject as a monolithic entity. Proponents defend sociobiology by emphasizing that human beings are the product of evolution; critics argue that human behaviour is enormously flexible and responsive to cultural context - so much so, that the human mind is said to liberate us from the 'tyranny of the genes'.

Although the interest of reaching a general assessment of the promise of sociobiology is undeniable, it is important to consider its merits and demerits on a case by case basis (Kitcher 1985). Perhaps some elements of human behaviour are adequately explained by the theory of natural selection while others are not. Even if sociobiologists

are right about incest avoidance, it is a separate question whether other behavioural regularities can be explained similarly. We should not accept sociobiology because it is sometimes correct; nor should we reject it because it is sometimes mistaken.

Parallel remarks apply to metaphorical applications of the biological concept of natural selection. It is not inevitable that the theory of the firm is correct. But even if it is, the question remains of which other social regularities can be explained within an evolutionary framework. For example, it is often true that unemployment reduces inflation. The usual supply-and-demand explanation of this fact does not advert to different individuals deploying different strategies and, as a result, prospering to different degrees. Rather, the usual approach holds that all individuals are rational or approximately so. All individuals try to pay the lowest price they can. The social regularity is explained by the hypothesis that people are, in the relevant respect, always the same. This is not an explanation in which behavioural variation is sifted and winnowed.

4 The explanation of variation in sociobiology

Well posed problems about the evolution of behaviour concern the explanation of variation. It is all too easy to explain why human beings ‘tend’ to avoid incest by saying that this behaviour maximizes fitness. The word ‘tend’ is vague; how much incest would we have to observe to cast doubt on the evolutionary explanation? A better formulated problem would begin by measuring the amount of inbreeding that occurs in a range of species, our own included. Accounting for this quantitative data gives more structure to the problem.

Although sociobiologists often are interested in explaining features of human behaviour that they believe are cultural universals, they also are interested in explaining behavioural variation among human beings. Perhaps the most familiar context for this type of problem concerns differences between the sexes. Behavioural differences between men and women, they argue, can often be explained by the hypothesis that men and women have evolved in response to different selective pressures. A standard example is their claim that men tend to be more sexually promiscuous than women because selection placed a larger premium on being discriminating about sexual partners in females than it did in males. The claim is not that all men are more promiscuous than all women; rather, selection is said to be an important influence on the behaviour, the result of which is different averages for the two sexes.

Sociobiologists are often criticized for assuming that ‘genetic determinism’ is true - that human behaviour is rigidly controlled by genes and cannot be changed by environmental factors. There are two defects in this general criticism, correct though it may be in some instances. First, even when sociobiologists claim that genetic differences are a cause of the phenotypic variation under study, they are not committed to saying that this is the only cause. For example, even if there is a genetic basis for some behavioural differences between the sexes, this does not exclude the possibility that environmental causes also play a role.

The second defect in the charge of genetic determinism is that some sociobiological explanations of behavioural variation are purely environmental. Consider, for example, Alexander’s 1979) proposed explanation of the kinship system known as the avunculate. In some societies, men provide more care for their sister’s children than for the children of their own wives. Alexander argues that this behaviour will be selectively advantageous when there is considerable female promiscuity. The avunculate maximizes a man’s inclusive fitness if he is more likely to have genes in common with his nephews and nieces than he is with his wife’s children.

The point of interest here is not whether this explanation is empirically correct, but the role it assigns to genetic factors. Alexander is not saying that societies that deploy the avunculate are genetically different from societies that do not. Rather, his idea is that human beings have basically the same genetic endowment, one that leads them to behave in ways that maximize their fitness. A behaviour that maximizes fitness in one circumstance may not do so in another. According to Alexander’s model, the within-species variation in behaviour is explained environmentally, not genetically.

5 The meaning of fitness in evolutionary epistemology

John Stuart Mill defended the value of freedom of thought and expression by describing a ‘marketplace of ideas’ in which ideas compete for acceptance (see Mill, J.S.). Consumers judge which ideas are best, and so the quality of ideas found in the market gradually improves. Evolutionary epistemologists, such as Karl Popper, Donald

Campbell, and David Hull, argue that this idea may be applied fruitfully to the scientific process itself (see Bradie 1994). Scientists are the consumers; they propose theories that compete with each other in an ideational environment. Successful theories persist; unsuccessful theories suffer extinction.

Literal applications of the concept of biological evolution by natural selection not only posit fitness differences among traits, but explain why the traits have the fitness consequences they do. In the mundane example of running speed in zebras, the claim was not just that fast zebras are fitter than slow ones, but that this is so because running speed enhances a zebra's ability to evade predators. It is important not to forget this further element when evolutionary epistemologists propose to use natural selection as a metaphor for scientific change.

It is harmless to say that theories differ in their (cultural) fitness. But what makes for a fitter theory? That is, which properties of a theory bear on its chances of surviving and spreading in the market-place of ideas? There are many possibilities, as the ongoing dispute between internalists and externalists in the history of science attests. An extreme internalist might say that observational support is the only important factor influencing a theory's viability. An extreme externalist might say that ideological palatability is all that matters. In between these extremes are pluralistic theories which say that a variety of factors influence whether a theory will be accepted. The point is that in a model of ideational selection, everything depends on the substantive account of fitness. The model risks vacuity if fitness is left unexplained (see Internalism and externalism in epistemology §4).

In developing a substantive model, the theorist must decide whether the goal is descriptive, or normative, or both. Is the point to describe how scientists actually evaluate theories or how they should do so? These two goals will coincide to the extent that the theorist believes that scientists are rational.

6 Cultural evolution without selection

The concept of natural selection is so striking an ingredient in evolutionary theory that it is easy to forget the theory's other resources. As noted earlier, a fundamental ingredient in current evolutionary theory is the idea of phylogenetic relationship. Current species have common ancestors, and a great deal of work in evolutionary biology is devoted to determining what the family trees are for various taxonomic groups. The detailed working out of ancestry is a quite separate task from the problem of identifying the adaptive significance, if any, of the traits that various species possess.

Social scientists have explored the idea that elements of culture are sometimes related to each other genealogically. For example, historical linguists attempt to determine which languages are closely related to each other and which are related only distantly. This enterprise leads them to postulate unobserved languages, such as proto-Indo European, which are said to be the ancestors of current languages. In similar fashion, philologists study the provenance of manuscripts and try to reconstruct an *Urtext* that lies at the root of a branching tree, the tips of which represent the variants of the manuscript they observe.

There is an important isomorphism between the biologist's problem of phylogenetic inference and the historian's problem of inferring the genealogy of cultural objects. What is observed, in the first instance, are species or artefacts that exist now. From the pattern of similarity and difference they exhibit, one postulates a branching tree, whose internal nodes and root represent species or artefacts that are not observed. The epistemological problem is how to judge this postulated tree in the light of available evidence. What makes one postulated tree more plausible than another? This common methodological problem is, and will continue to be, an arena for mutual influence between biology and the social sciences.

7 Philosophical connections

In addition to the relationships just sketched between evolutionary theory and the social sciences, there is considerable overlap in the philosophical questions pertinent to each.

The issue of reductionism arises in several forms. One involves sociobiology's attempt to 'reduce' certain social and psychological phenomena to biology. A second concerns whether biology and the social sciences are each 'reducible in principle' to physics. A third is occasioned by the fact that the relationship of wholes and parts is an important subject in both sciences; the dispute between methodological holism and individualism arises in both.

A second domain of philosophical overlap concerns the role of teleological concepts. Just as a biologist might say

that the function of the heart is to pump blood, so a social scientist might say that the function of ideology is to preserve existing power relations. It is worth asking what such functional claims mean and what ramifications they have for the aims and methods of the sciences in which they occur.

Finally, both the social sciences and biology deal with systems on which a large number of different causes impinge. This makes especially important the project of understanding how real causes may be disentangled from spurious ones and what it means for one cause to be more important than another.

See also: Evolution and ethics; Evolution, theory of

ELLIOTT SOBER

References and further reading

Alexander, R. (1979) *Darwinism and Human Affairs*, Seattle, WA: University of Washington Press.(An energetic defence of sociobiology.)

Barlow, J., Cosmides, L. and Tooby, J. (1992) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, Oxford: Oxford University Press.(A collection of essays that aims to apply Darwinian ideas to cognitive structures, rather than directly to behaviour.)

Boyd, R. and Richerson, P. (1985) *Culture and the Evolutionary Process*, Chicago, IL: University of Chicago Press.(Discusses both the origin of culture and the changes in evolutionary processes occasioned by the emergence of culture.)

Bradie, M. (1994) 'Epistemology from an Evolutionary Point of View', in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*, 2nd edn, Cambridge, MA: MIT Press, 453-76.(A useful survey article.)

Campbell, D. (1974) 'Evolutionary Epistemology', in P. Schilpp (ed.), *The Philosophy of Karl Popper*, Peru, IL: Open Court.(Elaborates a selective-retention model of scientific change and of problem solving generally.)

Cavalli-Sforza, L. and Feldman, M. (1981) *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton, NJ: Princeton University Press.(Develops models that represent the simultaneous effects of biological and cultural evolution.)

Hirshliefer, J. (1977) 'Economics from a Biological Viewpoint', *Journal of Law and Economics* 1: 1-52.(A useful review of attempts to develop evolutionary models in economics.)

Hull, D. (1988) *Science as a Process*, Chicago, IL: University of Chicago Press. (Argues that many of the group beneficial features of the scientific process are side effects of individuals seeking to maximize self interest.)

Kitcher, P. (1985) *Vaulting Ambition: Sociobiology and the Quest for Human Nature*, Cambridge, MA: MIT Press.(A philosophical critique of sociobiology that leaves open the possibility that human sociobiology may evolve into a viable research programme.)

Popper, K. (1973) *Objective Knowledge*, Oxford: Oxford University Press.(Argues that the epistemology of conjectures and refutations can be thought of as a Darwinian selection process.)

Sober, E. (1993) *Philosophy of Biology*, Oxford: Oxford University Press.(An introduction to philosophy of biology that discusses adaptationism, sociobiology, models of cultural evolution, and the issue of phylogenetic reconstruction.)

Examples in ethics

Philosophers often employ examples to illustrate how their favoured principles are to be applied to concrete cases, and sometimes even to show that principles are of no help in decision-making. Examples are also used to convince readers of the existence of moral dilemmas - unresolvable conflicts between moral obligations. But a variety of different philosophical questions concerning the role, status, and nature of examples used in ethics have also been raised. One such question concerns the role that examples should play in our moral experience: should this be a rhetorical, pedagogic role of persuading us to do what is right, as determined by pre-existing principle; or a stronger, logical role of helping to determine what is morally right? Another query relates to moral teaching: is exposure to and reflection on stories, tales, narratives and exemplars sufficient for moral education, or is there a further need for exposure to principles and theories of ethics? Third, in terms of the kinds of examples employed in moral philosophy and reflection, should such examples be culled from great literature or sacred texts? Alternatively, should they be actual case studies drawn from real life, or hypothetical but realistic examples constructed by theorists? Or should they be imaginary, highly improbable cases designed to test our intuitions? A fourth question asks how examples are best identified and described, and to what extent the examples used in ethics are themselves theory-laden or even theory-constituted.

1 A rhetorical or logical role?

There is widespread agreement among moral philosophers that relevant examples play an important role in persuading us to do what is morally appropriate and in encouraging us to try to become morally virtuous agents. The right kind of example, skilfully employed and presented, can bring home the moral point to us in ways that the abstractions of theories, principles, and rules often cannot. Vivid examples frequently possess a motivating power that principles lack.

But according to this familiar view, the standards of moral rightness (or courage, generosity, and so on) are always established antecedently by principle and theory. Examples, while important and necessary, particularly in the early stages of moral education, in bringing the abstractions of theory down to earth, thus remain the handmaiden of principle and theory. Kant's warning in *Grundlegung zur Metaphysik der Sitten (Groundwork of the Metaphysics of Morals)* ([1785] 1903: 409) that cogent examples, while serving to 'make perceptible what the practical law expresses more generally', must nevertheless always be judged by 'their true original, which resides in reason' is a classic expression of this position.

However, some philosophers advocate a stronger role for examples, one in which the examples themselves are held to play a logical role in determining what is morally appropriate. On this latter view, examples function not just as applications of rules or illustrations of principles but rather as challenges for moral judgment to respond to particular situations in a manner which goes beyond mere rule-following. Aristotle's doctrine of practical wisdom or *phronēsis*, which holds that the correct decision 'depends on the particular facts and on perception [of the particular facts]' (*Nicomachean Ethics* 1126b4, 1109b20-3) is the classic representative of this latter view (see Aristotle §23; Virtue ethics §6). The medieval tradition of moral casuistry, particularly as elaborated by Aquinas, according to which individual circumstances may 'touch' or alter cases in important ways, represents a critical historical development of this Aristotelian perspective towards examples (see Casuistry).

2 Are examples enough?

Some philosophers argue that moral principles simply do not exist. According to ethical particularists, moral decisions should always be made case by case, without the false support or illusory guidance of moral principles (see Universalism in ethics §3; Moral realism §5; Logic of ethical discourse). Others hold that moral rules and principles, while not illusory, are unhelpful and often harmful in determining what to do and how to live. Antitheorists in ethics attack the whole idea of a moral 'theory' which seeks to evaluate particular situations by reference to general principles. The intellectual ancestry of particularism and antitheory in ethics is diverse; such views have been imputed variously to Aristotle, Wittgenstein, and postmodernists such as Foucault and Lyotard (see Virtue ethics §6; Wittgensteinian ethics §§2-3; Postmodernism). But the upshot of both views is simply that examples are enough. The best approach to moral education and reflection, for adults as well as for children, is one that emphasizes not impersonal principles and abstract arguments but personal exemplars and concrete cases.

Opponents argue that the worthiness of any alleged moral exemplar or example must always be judged according to antecedently established and defended general norms. Here again, Kant is the classic advocate: 'Every example of… [morality] presented to me must first itself be judged by moral principles in order to decide if it is fit to serve as… a model: it can in no way supply the prime source for the concept of morality' ([1785] 1903: 408). Additionally, defenders of theory hold that a primary goal of moral education must always be critical, autonomous reflection concerning norms. Autonomous moral agents do not assent blindly to others' teachings, but strive continually to make the truths of moral exemplars their own by reflecting critically on the legitimacy and justifiability of these teachings.

3 What kind of examples?

Disagreement also exists concerning what kind of examples to use in ethics. Some prefer literary examples, on the ground that they characterize our moral situation more profoundly and precisely than the less engaging, schematic examples constructed by theorists. On this view, personal experience is, in comparison with fiction, too confined and too parochial. Well-crafted literary examples give readers data that is deeper, sharper, and more precise than much of what occurs in real life. Antitheory tendencies often prevail among strong defenders of literary examples as well, in so far as they are committed to the view that the process of understanding our moral situation is a task not for theory but for nonphilosophical genres.

Others hold that case studies from real life are preferable, for the simple reason that moral theories stand a better chance of guiding moral practices effectively if they draw their data from reality rather than fiction. Additionally, it is often held that such examples are less 'cooked' and have fewer hidden theoretical strings attached to them (see §4). Granted, the theorist's own life experience may well be too confined and too parochial to serve as a satisfactory resource, but examples drawn from, for instance, medical practice, environmental policy, or international relations present theorists with more than ample opportunities. Much recent work in applied and professional ethics is inspired by this desire to bring theory into closer contact with actual practice via real-life examples (see Applied ethics §§2, 6). The concern of many virtue ethicists to articulate a more realistic moral psychology by examining the personalities of real-life moral exemplars represents an analogous development.

A third, more traditional camp claims that hypothetical examples constructed by the theorists themselves are the best way to go, on the ground that a better fit between principle and example is thereby achieved. Theorists here are the creators of their examples, and can embellish them to suit their own purposes. Because they have more control over the examples, a clearer, less *ad hoc* picture of how moral choices are to be made can also supposedly be gained. Within this third category one also finds a wide range of hypothetical example-types, extending from bare, highly schematic accounts at one end to intensely detailed and finely nuanced accounts at the other.

Finally, a fourth kind of example which also has its philosophical defenders is what might be called the imaginary case. This type too is of course hypothetical, but more extremely or fantastically so: it frequently involves counterfactual scenarios quite removed from everyday life. Sometimes one senses a frustrated science fiction writer behind the philosopher who concocts such cases, but when skilfully constructed and presented they can serve to flesh out our moral intuitions in ways that less fantastic stories cannot. The idea is then to transfer these intuitions to relevant (more earthly) cases where our intuitions are confused.

4 How to identify and describe examples

Some authors have also pressed sceptical questions concerning the ways in which specific cases are conceptualized and singled out by moral theorists. Postpositivist philosophers of science are fond of slogans such as 'All facts are theory-laden' (see Observation §§3-4), but it would appear that the theory-ladenness of examples used in ethics is radically deeper. What counts as an important moral problem, and the ways and reasons why it counts, are strongly affected by the theory one advocates or presupposes, as well as the time and place in which one lives. Applied and professional ethicists who work closely with practitioners also run the danger of allowing their perspective to be unduly influenced by the concerns of various interest groups and bodies of professional opinion. When unchecked, the result is not a critical reflection on moral practices but merely a reiteration of them. Given the evident theory-ladenness and conventionalism of many examples in ethics, it has been argued that philosophers who employ examples in ethical discussion need to pay much closer attention to the ways in which their examples are chosen and described. Additionally, stronger justificatory strategies are needed to defend such choices and

descriptions against possible alternative conceptualizations.

5 Concluding remarks

The kinds of question philosophers have raised about examples in ethics are diverse, and the above discussion does not claim to be exhaustive. At the same time, some commonality can be detected within this diversity. Simplifying somewhat, one can say that a unifying theme running through at least parts of §§1, 2 and 4 is an ethical variation on traditional metaphysical and epistemological themes concerning the relationship between universals and particulars (see Nominalism; Universals). Within philosophical discussions of ethics, should we grant pride of place to universals or to particulars? Should moral reflection and theory construction proceed in a 'top-down' application of principle to problem, or in a 'bottom-up' manner which insists that all rules and principles must submit to the jurisdiction of specific examples and not the other way around? (Here too, however, one needs to remember to allow room for adamant 'bottom-dwellers' - particularists and antitheorists who dismiss principles altogether.)

Philosophers as a group are undoubtedly more constitutionally principle-prone and theory-inclined than are novelists, dramatists, story tellers, and others who provide moralists with their richest stores of examples. Nevertheless, when philosophers decide to philosophize about ethics they need to keep in mind that they are ultimately philosophizing not about timeless entities but rather about how agents should live and act - agents, that is, who live in a world not only of space, time, and causality, but of history, politics, culture, and assorted other contingencies as well. The philosopher's natural impulse towards abstraction needs to be continually checked in ethics if we are not to lose sight of our subject matter. At the same time, while ethical theory cannot realistically aspire to the same level of generality as, say, astrophysical theory, the perpetual need to reflect on, understand, and evaluate moral examples eventually necessitates the resources of normative theory.

See also: Conscience; Casuistry; Moral education; Moral knowledge; Situation ethics; Theory and practice

ROBERT B. LOUDEN

References and further reading

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, London: Burnes, Oates, 1924, IaIIae.18.10-11. (Outlines the maxim that circumstances alter cases.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. W.D. Ross, revised by J.O. Urmson, Oxford: Oxford University Press, 1975, book VI 8-13.(Contains an extended discussion of *phronēsis*.)

Clarke, S.G. and Simpson, E. (1989) *Anti-Theory in Ethics and Moral Conservatism*, Albany, NY: State University of New York Press.(Good source book on antitheory movement. See especially contributions by A. Baier, B. Williams, J. McDowell, and bibliographic essay by editors.)

Dancy, J. (1983) 'Ethical Particularism and Morally Relevant Properties', *Mind* 92: 530-47.(Strong defence of ethical particularism.)

Flanagan, O. (1991) *Varieties of Moral Personality: Ethics and Psychological Realism*, Cambridge, MA: Harvard University Press.(Develops exemplar issue within virtue ethics movement.)

Jonsen, A.R. and Toulmin, S. (1988) *The Abuse of Casuistry: A History of Moral Reasoning*, London: University of California Press.(Excellent history and defence of moral casuistry tradition.)

Kant, I. (1780-81) *Lectures on Ethics*, trans. L. Infield, New York: Harper & Row, 1963, esp. 109-10.(Argues that examples properly 'serve for our encouragement and emulation', but that we should not copy them in our own behaviour.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, in *Kants gesammelte Schriften*, ed. Königlichen Preußischen Akademie der Wissenschaften, Berlin: Reimer, vol. 4, 1903; trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(References made to this work in the entry give the page number from the 1903 Berlin Akademie volume; these page numbers are included in the Paton translation. The work expounds the view that any alleged moral exemplar or example must always be judged according to antecedently established principles.)

Kant, I. (1788) *Kritik der praktischen Vernunft*, trans. L.W. Beck, *Critique of Practical Reason*, New York: Macmillan, Library of Liberal Arts, 3rd edn, 1993, esp. 157-68.(Advises moral educators to search through 'biographies of ancient and modern times' to find relevant examples that will help strengthen moral judgment of students.)

Kant, I. (1793) *Die Religion innerhalb der Grenzen der blossen Vernunft*, trans. T.M. Greene and H.H. Hudson and J.R. Silber, *Religion Within the Limits of Reason Alone*, New York: Harper & Row, 1960, esp. 41-49, 54-60.(Argues that human beings always need to resort to analogies to natural beings in order to make moral concepts comprehensible to themselves.)

Kant, I. (1797) *Die Metaphysik der Sitten*, trans M.J. Gregor, *The Metaphysics of Morals*, Cambridge: Cambridge University Press, 1991, esp. 266-72.(Argues that 'good example on the part of the teacher' plays a key role in the cultivation of human virtue.)

Kant, I. (1803) *Education*, trans. A. Churton, Ann Arbor, MI: University of Michigan Press, 1960.(Discusses the role of examples in the moral education of children.)

Larmore, C.E. (1987) *Patterns of Moral Complexity*, Cambridge: Cambridge University Press, ch. 1.(Presents the distinction between a rhetorical and a logical role for examples, and defends the logical role.)

Louden, R.B. (1992) 'Go-carts of Judgment: Exemplars in Kantian Moral Education', *Archiv für Geschichte der Philosophie* 74: 303-22.(A close textual analysis of Kant's views on both the value and the danger of teaching morality by exemplification.)

Louden, R.B. (1992) *Morality and Moral Theory: A Reappraisal and Reaffirmation*, Oxford: Oxford University Press.(A response to antitheorists and particularists which defends a more empirically informed model of ethical theory. Also argues that the moral theories of Aristotle and Kant have far more in common than is usually thought.)

Nussbaum, M. (1990) *Love's Knowledge: Essays on Philosophy and Literature*, Oxford: Oxford University Press. (Strong defence of literary examples in ethics.)

O'Neill, O. (1986) 'The Power of Example', in *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge: Cambridge University Press, 1989.(A Kantian critique of various example-oriented approaches to ethics.)

O'Neill, O. (1988) 'How Can We Individuate Moral Problems?', in D.M. Rosenthal and F. Shehadi (eds) *Applied Ethics and Ethical Theory*, Salt Lake City, UT: University of Utah Press.(Argues that the examples and cases discussed by contemporary ethicists are themselves heavily theory-laden.)

Existence

Philosophical problems concerning existence fall under two main headings: 'What is existence?' and 'What things exist?'. Although these questions cannot be entirely separated, this entry will concentrate on the first.

The question 'What is existence?' has produced a surprising variety of answers. Some hold that existence is a property that every individual has, others that it is a characteristic that some individuals have but other (for example, imaginary) individuals lack, while proponents of the thesis that 'existence is not a predicate' hold that existence is not a property or characteristic of individuals at all.

Other philosophical issues concerning existence include: disputes about whether there are abstract objects (for example, numbers, universals) as well as concrete ones, immaterial souls as well as bodies, possible objects as well as actual ones, and so on; and questions about which entities (if any) are the fundamental constituents of reality.

1 Objects and existence

To the question 'What is existence?' it is tempting to respond: 'A property that belongs to everything'. However, this answer appears to have paradoxical consequences.

An ancient puzzle, traceable to Parmenides, concerns denials of existence. If existence is a property of everything, one cannot talk about anything that does not exist. But if one cannot talk about what does not exist, how can one truly deny the existence of anything? Yet it is obvious that there are true denials of existence (true negative existential statements): Santa Claus does not exist, nor do Henry VIII's seventh wife or talking tigers.

A second puzzle concerns assertions of existence (positive existential statements). If everything that can be talked about exists, how can a true assertion that something exists be other than trivial? Yet it is obvious that there can be true but non-trivial assertions of existence: 'Egg-laying mammals exist', for example, is not trivial; and 'Homer existed' would be informative if true.

One response to these puzzles consists in holding that as well as objects that exist, there are objects that do not exist. Theories of this type are called 'Meinongian' (see Meinong 1904), after the most famous exponent of such a theory (see Meinong, A.). According to a Meinongian theory, the possibility of true negative existential statements is explained by the fact that Santa Claus, the seventh wife of Henry VIII and so on are objects which can be talked about and can possess various properties (such as driving reindeer and being married), but which lack existence. As for the second puzzle: if some objects are nonexistent, it can clearly be informative to be told, of an object, that it exists.

Such a theory may appear to offer an attractive account of discourse about fiction; one that allows us to take at face value the fact that both 'Pegasus has wings' and 'Pegasus does not exist' appear to be true. However, there are features of this Meinongian theory that many find unacceptable. For example, if we are to use the theory to explain all true negative existential statements, we must follow Meinong in admitting that there are impossible objects such as the round square (since 'The object that is both round and square does not exist' is evidently true). Also, the Meinongian theory implies that Pegasus, although nonexistent, is a horse, yet it is debatable whether the property of being a horse can belong to something nonexistent. Even the fundamental tenet of the theory - that we can distinguish being an object from being an existent object - is open to question. If we admit that Pegasus is an object, a possible subject of discourse and a bearer of properties, does it make sense to add that Pegasus does not exist? If the answer is that existence requires location in space and time, although being an object does not, we may ask why we should accept this distinction.

2 The quantifier account of existence

An influential rival to the Meinongian account holds that the puzzles described in §1 above arise from the mistaken assumption that the term 'exist(s)' attributes a property to individuals. This solution employs an account of existence given by Frege and elaborated by Russell (see Frege, G. §2; Russell, B. §9). The account is associated with the slogan 'existence is not a predicate', a dictum derived from the writings of Kant, who anticipated the Frege-Russell theory in claiming that the ontological argument for God's existence depends on the illegitimate

assumption that existence is a property (see God, arguments for the existence of §§2-3).

According to the Frege-Russell theory, the superficial grammatical form of the sentence ‘Egg-laying mammals exist’ (according to which it appears to attribute a property of existence to egg-laying mammals) is misleading. Its true logical form is revealed by reading it as ‘There are egg-laying mammals’, which can be analysed using logical symbolism as ‘ $(\exists x)x$ is an egg-laying mammal’, where ‘ $\exists x$’ is the existential quantifier (to be read as ‘There is some *x* such that…’). According to this ‘quantifier’ account of existence (or of the word ‘exist(s)’) the grammatical predicate ‘exist(s)’ disappears when the sentence is subjected to logical analysis, being replaced by the existential quantifier. As for ‘Talking tigers do not exist’, this is to be treated as equivalent to ‘There are no talking tigers’, which can be analysed as ‘It is not the case that: $(\exists x)x$ is a talking tiger’. Frege and Russell combined this quantifier account of existence with a claim to the effect that although ‘exist(s)’ is not a first-order predicate (a predicate of individuals), it is a higher-order predicate: a predicate of properties. For example, although in ‘Egg-laying mammals exist’ the word ‘exist’ does not ascribe the property of existence to egg-laying mammals, it does ascribe a property - that of being instantiated (‘having an instance’) - to the property of being an egg-laying mammal.

Using his celebrated theory of descriptions, Russell (1918) argued that the quantifier account can be extended to existential sentences involving definite descriptions and proper names, thus providing an ingenious general solution to the puzzles described above (see Descriptions). According to Russell, ‘Henry VIII’s seventh wife does not exist’ and ‘Santa Claus does not exist’ should be construed not as saying, of certain objects, that they lack existence, but rather as saying that certain properties (for example, being a seventh wife of Henry VIII or being a unique white-bearded, reindeer-driving, gift-delivering inhabitant of the North Pole) are not instantiated; that nothing has these properties. Similarly, ‘God does not exist’ is to be interpreted as the claim that a certain property - perhaps that of being a uniquely powerful, wise and benevolent creator of the universe - is not possessed by anything. Finally, if, following Russell, we treat the proper name ‘Homer’ as equivalent to the description ‘the author of the *Iliad* and the *Odyssey*’, we can construe the assertion that Homer existed as the claim that one person wrote both of these poems, an item of information that would be far from trivial.

The quantifier account is widely accepted as an analysis of general existential statements. However, many philosophers have been reluctant to follow Russell in holding that all meaningful existential statements are to be analysed in terms of the existential quantifier. (A notable exception is Quine (1953).) One difficulty is that there appear to be meaningful statements asserting the existence of particular things that resist the quantifier treatment. Even if we accept Russell’s (controversial) analysis of definite descriptions, and accept his descriptive analysis of proper names such as ‘Santa Claus’ and ‘God’, there are powerful reasons for thinking that most proper names are not equivalent to descriptions (see Proper names §§1-3). Hence it is doubtful whether ‘Bertrand Russell existed’ is equivalent to any claim that some set of properties is or was instantiated. Even more recalcitrant are existential statements involving demonstratives, such as ‘This exists’ or ‘That elephant exists’. Unless we say, implausibly, that all such sentences are false or meaningless, we appear to be committed to the conclusion that there are, after all, some true sentences in which ‘exist(s)’ ascribes a property of existence to particular objects. Further, it has been argued that some statements saying that things have existed, will exist, or might not have existed cannot be given a satisfactory analysis without invoking a property of existence.

However, if existence is a genuine property, an account must be given of the relation between this property and the existential quantifier, since it is undeniable that some general existential sentences are closely related to, if not equivalent to, existential quantifications. Moreover, if ‘exist(s)’ can express a property of individuals, does it play this role in any of the negative and positive existential statements discussed in §1? If so, how - assuming that the Meinongian account is rejected - are the puzzles to be solved? And what, if anything, can be said about the alleged property of existence beyond the claim that it belongs to everything? There is at present no consensus about these issues.

See also: Being; Fiction, semantics of; Free logics; Ontology; Ontological commitment; Logical and mathematical terms, glossary of

PENELOPE MACKIE

References and further reading

Barnes, J. (1972) *The Ontological Argument*, London: Macmillan, ch. 3.(A succinct critical discussion of the thesis that existence is not a predicate, including many references to other works.)

Evans, G. (1982) *The Varieties of Reference*, Oxford: Clarendon Press, ch. 10. (Defends the thesis that existence is a property of everything, offering a sophisticated solution to some of the problems mentioned in §2.)

Meinong, A. (1904) 'Über Gegenstandstheorie' (The Theory of Objects), in R.M. Chisholm (ed.) *Realism and the Background of Phenomenology*, New York: Free Press, 1960, 76-117.(A difficult work. Parsons' *Nonexistent Objects* is an easier introduction to Meinongian views.)

Parsons, T. (1980) *Nonexistent Objects*, New Haven, CT: Yale University Press.(A modern and thorough development of Meinong's views. Quite technical in places.)

Quine, W.V. (1953) 'On What There Is', in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1980.(A classic piece of advocacy for the quantifier account of existence, introducing Quine's famous 'criterion of ontological commitment'.)

Russell, B.A.W. (1918) 'The Philosophy of Logical Atomism', Lectures 5 and 6, in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992. (An informal presentation by Russell of his version of the quantifier account.)

Williams, C.J.F. (1981) *What is Existence?*, Oxford: Clarendon Press.(A wide-ranging discussion, including a defence of a version of the quantifier account against numerous objections.)

Existentialism

The term ‘existentialism’ is sometimes reserved for the works of Jean-Paul Sartre, who used it to refer to his own philosophy in the 1940s. But it is more often used as a general name for a number of thinkers in the nineteenth and twentieth centuries who made the concrete individual central to their thought. Existentialism in this broader sense arose as a backlash against philosophical and scientific systems that treat all particulars, including humans, as members of a genus or instances of universal laws. It claims that our own existence as unique individuals in concrete situations cannot be grasped adequately in such theories, and that systems of this sort conceal from us the highly personal task of trying to achieve self-fulfilment in our lives. Existentialists therefore start out with a detailed description of the self as an ‘existing individual’, understood as an agent involved in a specific social and historical world. One of their chief aims is to understand how the individual can achieve the richest and most fulfilling life in the modern world.

Existentialists hold widely differing views about human existence, but there are a number of recurring themes in their writings. First, existentialists hold that humans have no pregiven purpose or essence laid out for them by God or by nature; it is up to each one of us to decide who and what we are through our own actions. This is the point of Sartre’s definition of existentialism as the view that, for humans, ‘existence precedes essence’. What this means is that we first simply exist *- find ourselves born into a world not of our own choosing - and it is then up to each of us to define our own identity or essential characteristics in the course of what we do in living out our lives. Thus, our essence (our set of defining traits) is chosen, not given.*

Second, existentialists hold that people decide their own fates and are responsible for what they make of their lives. Humans have free will in the sense that, no matter what social and biological factors influence their decisions, they can reflect on those conditions, decide what they mean, and then make their own choices as to how to handle those factors in acting in the world. Because we are self-creating or self-fashioning beings in this sense, we have full responsibility for what we make of our lives.

Finally, existentialists are concerned with identifying the most authentic and fulfilling way of life possible for individuals. In their view, most of us tend to conform to the ways of living of the ‘herd’: we feel we are doing well if we do what ‘one’ does in familiar social situations. In this respect, our lives are said to be ‘inauthentic’, not really our own. To become authentic, according to this view, an individual must take over their own existence with clarity and intensity. Such a transformation is made possible by such profound emotional experiences as anxiety or the experience of existential guilt. When we face up to what is revealed in such experiences, existentialists claim, we will have a clearer grasp of what is at stake in life, and we will be able to become more committed and integrated individuals.

1 Historical development

Although such earlier thinkers as Augustine, Montaigne, Shakespeare and Pascal have been called existentialists, the term should be reserved for a loosely connected group of thinkers in recent times who were responding to certain views that became widespread in the nineteenth century. These views include, first, the scientific picture of reality as a meaningless, value-free collection of material objects in causal interactions, and second, the modern sense of society as an artificial construct that is inevitably in conflict with the aspirations of the individual. German Idealism had attempted to counteract the implications of these new ideas, but idealism had collapsed by the 1840s, and the result was a growing feeling that the individual is ultimately alone and unsupported in a cold and meaningless universe (see Idealism).

Existentialism appeared in the nineteenth century alongside romanticism, but it was different from romanticism in important respects. For one thing, where romanticism tried to evoke a sense of the individual’s participation in the larger context of nature, the first great existentialist, Søren Kierkegaard, held that humans are at the most basic level solitary, ‘existing individuals’ with no real connections to anything in this world. Instead of suggesting that we are at home in this world, Kierkegaard tried to intensify the individual’s feeling of anxiety and despair in order to bring about a ‘leap of faith’ that would bring the person into a defining relationship to the God-man (Christ).

The next figure usually included in the pantheon of existentialists, Friedrich Nietzsche, began from the assumption

that the development of science and critical thinking in Western history has led to the result that people have lost the ability to believe in a transcendent basis for values and belief. When Nietzsche said that ‘God is dead’, he meant that all the things people previously thought of as absolutes - the cosmic order, Platonic Forms, divine will, Reason, History - have been shown to be human constructions, with no ultimate authority in telling us how to live our lives. In the face of the growing ‘nihilism’ that results from the death of God, Nietzsche tried to formulate a vision of a healthy form of life people can achieve once they have given up all belief in absolutes (see Nihilism).

The translation of Kierkegaard’s works and the discovery of Nietzsche’s writings had an immense impact on German thought after the First World War. The psychiatrist and philosopher, Karl Jaspers, drew on these two figures to develop what he called a ‘philosophy of existence’. Martin Heidegger, influenced by Kierkegaard as well as by the movement called ‘philosophy of life’ (then associated with the names of Nietzsche, Wilhelm Dilthey and Henri Bergson), began his major work, *Being and Time* (1927), with an ‘existential analytic’ aimed at describing life from the standpoint of concrete, everyday being-in-the-world (see Lebensphilosophie). Heidegger’s thought was also influenced by Edmund Husserl’s phenomenology, an approach to philosophy that emphasizes description of our experience as it is prior to reflection and theorizing (see Phenomenological movement).

Working independently in France, Gabriel Marcel was building on Bergson’s philosophy to develop an alternative to the dominant idealist philosophy taught in the universities. Basing his reflections on his own experience of life, Marcel claimed that a human being must be understood as an embodied existence bound up with a concrete situation. Because the body and the situation can never be completely comprehended by the intellect, Marcel sees them as part of what he calls the ‘mystery’. Maurice Merleau-Ponty took over Marcel’s notion of embodied being-in-a-situation as basic for his own existential phenomenology. Jean-Paul Sartre also drew on Marcel’s thought, but he was especially influenced by the thought of Husserl and Heidegger. It seems that the term ‘existentialism’ was first used by critics of Sartre, but it came to be accepted in the 1940s by Sartre and Simone de Beauvoir as they replied to their critics. Merleau-Ponty and Albert Camus were initially associated with the movement called existentialism during its heyday after the Second World War, but both eventually rejected the term as they came to distance themselves from Sartre due to political differences.

There have been important developments outside Germany and France. The Spanish philosopher, José Ortega y Gasset, influenced by Dilthey’s philosophy of life, developed a number of ideas that closely parallel those of Heidegger and other existentialist thinkers. The novels and short stories of the Russian writer, Fëdor Dostoevskii, were influential not only for Russian existentialists like Nikolai Berdiaev, but for Heidegger, Sartre and Camus as well. Existentialism has also had a profound impact on other fields. The movement of existential theology remains influential today (see Existentialist theology), and existential psychoanalysis (especially Ludwig Binswanger, Medard Boss and Rollo May) continues to be of interest in psychotherapy. Though existentialism is no longer a central movement in philosophy, many of its principal exponents continue to be important in current philosophical discussions.

2 The human condition

Existentialists start out from the assumption that it is no longer possible to believe that there is some transcendent justification or underlying ground for our existence. If God is dead, then we find ourselves ‘abandoned’, ‘forlorn’, ‘thrown’ into a world, with no pregiven direction or legitimation. Though we seek some overarching meaning and purpose for our lives, we have to face the fact that there is no ‘proper function of humans’ or ‘plan in God’s mind’ that tells us the right way to be human.

This picture of our predicament leads to a particular view of human existence that is accepted by many existentialists. In contrast to traditional theories, which think of a human as a thing or object of some sort (whether a mind or a body or some combination of the two), existentialists characterize human existence as involving a deep tension or conflict between two different aspects of our being. On the one hand, we are organisms among other living beings, creatures with specific needs and drives, who operate at the level of sensation and desire in dealing with the present. At this level, we are not much different from other animals. On the other hand, there is a crucial respect in which we differ from other organisms. One way to describe this difference is to say that, because we are capable of self-awareness, we are able to reflect on our own desires and evaluate ourselves in terms of some larger vision of what our lives are adding up to. In this sense we transcend our own being as mere things. What is

characteristic of our being as humans is that we *care* about the kinds of beings we are, and we therefore take a stand on our basic desires. According to the existentialists, humans are unique among entities in that they form second-order desires about their first-order desires, and they therefore have aspirations that go beyond the immediacy of their sensual lives.

Heidegger and Sartre try to capture this reflexive dimension of human existence by saying that what is unique about humans is that their own being is ‘in question’ or ‘at issue’ for them. What kind of person I am matters to me, and because I am concerned about what I am and will be, I take some concrete *stand* on my life by assuming roles and developing a specific character through my actions. But this means that my existence is characterized by a fundamental tension or clash between my immediate sensations and desires on the one hand, and my long-range aims and projects on the other. As Sartre puts it, a ‘rift’ or a ‘gap’ - a ‘nothingness’ - is introduced into the fullness of being in the universe by human existence. Because consciousness makes us more than what we are as creatures with immediate sensations and desires, Sartre says that human reality ‘is not what it is and is what it is not’.

The conception of human existence as a tension also appears in Kierkegaard’s description of the self. For Kierkegaard, humans are both finite and infinite, temporal and eternal, contingent and free. What defines our identity as selves is the concrete way we relate ourselves to this tension. In a similar way, Nietzsche holds that we are both creatures and creators, and we have to embrace both these dimensions of ourselves in order to be fully human. Heidegger and Sartre refer to the two aspects of the self as ‘facticity’ (our mere givenness) and ‘transcendence’ (our ability to surpass our givenness through our interpretations and aspirations). In their view, life is a continuous tension between these elements, a tension resolved only in death. Finally, Jaspers seems to have a similar conception of humans in mind when he points out the polarity between our being as an empirical consciousness-as-such and our desire to grasp the general and realize our freedom as *Existenz*.

If we regard the self as a tension or struggle, it is natural to think of human existence not as a thing or object of some sort, but as an unfolding event or happening - the story of how the tension is dealt with. What defines my existence, according to this view, is not some set of properties that remain the same through time, but the ‘event of becoming’ through which I carry out the struggle to resolve the tension that defines my condition in the world. As an ongoing happening, I *am* what I make of myself throughout the course of my life as a whole. In Ortega’s words, a human ‘does not have a nature, but rather a history’ (see Ortega Y Gasset, J.). What defines my existence as an individual is the ongoing story of what I accomplish throughout my life.

To think of a human as an unfolding story suggests that human existence has a specific sort of temporal structure. We are not like rocks and cauliflowers which continue to exist through an endless sequence of ‘nows’. Instead, human temporality has a kind of cumulativeness and future-directedness that is different from the enduring presence of physical things. First, our existence is directed toward the future to the extent that we are striving to realize something for our lives. Heidegger calls this element of ‘futurity’ our ‘being-*towards*-death,’ understood as a movement toward realizing our own being by achieving certain things throughout our active lives. Second, the past shows up for us as something retained and carried forward for the purposes of our future. Depending on our projects at any given time, our past actions show up for us as assets or as liabilities in relation to what we are doing. Finally, our present appears as a point of intersection between our future projects and our past accomplishments. Because we are time-binding beings whose lives always reach out into the future and hold on to the past, we can never achieve the kind of direct presence of self to self that Descartes thought he had found in the *Cogito* (‘I think’).

To say that human temporality is cumulative is to say that everything we do is contributing to creating our ‘being’ as a totality. In this sense, we *are* what we *do* in living out our lives: we define our own identity through the choices we make in dealing with the world. Because there is no fixed essential nature which we have in advance, our ‘essence’ as individuals is defined and realized through our concrete existence in the world. Whatever capacities and traits I am born with, it is up to me to take them over and make something of them in what I do. Thus, whether aware of it or not, I am creating my own identity in my actions.

3 Being-in-the-world

Existentialists are deeply suspicious of the high-level, abstract theorizing about humans found in traditional philosophy and in the sciences. In their view, the concern with subsuming all particulars under concepts and

building systems tends to conceal crucial features of our lives as individuals. For this reason, existentialists generally start out from a description of ourselves as agents in everyday contexts, prior to reflection and theorizing. These descriptions reveal that it is part of our 'facticity' that we are generally caught up in the midst of things, involved with others in trying to accomplish specific goals, and affected by moods and commitments that influence our perception and thoughts. Furthermore, we are embodied beings who encounter the world only from the standpoint of a particular bodily orientation that gives us a set on things: 'I am my body' Marcel wrote, and this theme of embodiment became central to the thought of Beauvoir and Merleau-Ponty. We are also bound up with contexts of equipment in practical situations where we are trying to accomplish certain goals. Finally, as social beings, we always find ourselves embedded in a particular cultural and historical milieu that conditions our outlook and determines our basic orientation toward the world. To say that we are 'factical' beings is to say that we are always 'being-in-a-situation', where our being as selves is inseparable from a shared, meaningful life-world.

If we are always embedded in a situation, then all inquiry must start out from an 'insider's perspective' on things, that is, from a description of the world *as it appears to us* - to beings who are participants in our forms of life, with our unique sort of bodily set, feelings and modes of perception. We have no choice but to begin from where we stand in the thick of our actual lives, with our local attachments and particular cares and concerns. But this means that there is no way to achieve the sort of global 'God's-eye view' on ourselves and our world philosophers have sought ever since Plato. Existentialists are critical of the philosophical ideal of achieving a totally detached, disinterested, disengaged 'view from nowhere' that will provide us with completely objective knowledge. The attempt to step back from our ordinary concerns in order to achieve a totally detached and dispassionate standpoint - the stance Marcel calls 'desertion' and Merleau-Ponty calls 'high-altitude thinking' - will always give us a distorted view of the world, because it bleaches out our normal sense of the significance and worth of the things we encounter around us. In order to be able to gain an insight into the way reality presents itself to us at the most basic level, then, we need to start from a description of what Heidegger calls our 'average everydayness', our ordinary, familiar ways of being absorbed in practical affairs.

The idea that our being-in-a-situation or being-in-the-world is fundamental and inescapable gives the existentialists a way of criticizing the idea, central to philosophy since Descartes, that we are at the most basic level *minds* receiving sensory inputs and processing information. Sartre, for example, rejects the idea that the self can be thought of as a 'thinking substance' or self-encapsulated 'field of consciousness' distinct from the world. In my pre-reflective activities, Sartre says, I encounter myself not as a bundle of beliefs and desires in a mental container, but as being 'out there' *with* the things I am concerned about. When I am chasing a bus, I encounter my *self* as a 'running-toward-the-bus'. My being is found not in my head, but with the bus. Sartre thinks that this follows from Husserl's view that consciousness is always 'intentional,' always directed *towards* entities in the world (see Intentionality). If Husserl is right, according to Sartre, then the 'I' is not an object, not a 'something', but is instead sheer intentional activity directed towards things in the world. The totality of my intentional acts defines me; there is no residue of 'substantial thinghood' distinct from my acts.

The existentialist conception of our irreducible being-in-a-situation calls into question some of the dualisms that have dominated so much of Western thought. First, existentialists deny the romantic distinction between an outer self - what we *do* in the world - and an inner, 'true' self that embodies our genuine nature. If we just *are* what we *do*, as existentialists contend, then there is no basis for positing a substantive 'real me' distinct from the parts I play and the things I do. Second, the account of the primacy of being-in-the-world tends to undermine the traditional subject-object model of our epistemological situation. Existentialists suggest that the assumption that humans are, at the most basic level, subjects of experience trying to formulate beliefs about objects on the basis of their inner representations, distorts our situation. If it is true that we are initially and most basically already out there with things in the world, then there must be something wrong with traditional epistemological puzzles about how a knowing subject can 'transcend' its veil of ideas to gain knowledge of objects in the external world. Finally, the existentialist picture of our basic situation as always bound up with a practical life-world seems to raise questions about the traditional fact-value distinction. Existentialists hold that we always encounter the everyday life-world as a context of equipment bound up with our aims as agents in the world. If the things we encounter are initially and most basically functional entities tied up with our purposive activities, however, then it is an illusion to think that what is given 'at first' is a collection of brute objects we subsequently invest with subjective values.

In our everyday lives, fact and value are inseparable.

In general, existentialists hold that traditional dualisms arise only when we try to adopt a cool, detached, theoretical stance towards things. But since such a stance is derivative from and parasitic on a more basic way of being in which we are inseparably bound up with things in practical contexts, it cannot be regarded as providing us with a privileged insight into the way things really are.

4 Freedom and responsibility

As being-in-the-world, we are already engaged in a shared life-world that gives us a prior sense of what is possible, and we find ourselves with choices in our past that carry weight in determining how we can act in the future. This is our 'facticity', and it makes up what is just 'given' in our lives. However, existentialists regard facticity as only one aspect of human existence, for they hold that humans always have the ability to transcend their given situation by taking a stand on their own lives. As 'transcendence' we are always taking over our situations and making something of them through our choices. This ability to transcend our facticity means that we have free will. Our choices are free in the sense that (1) no outside factors determine our will, (2) in any particular case we could have acted otherwise than we did, and (3) we are therefore responsible for our choices in a way that justifies moral praise and blame (see Free will). (Nietzsche is inclined to reject the third sense because of its role in imposing feelings of guilt on people, but in other respects he seems to be committed to believing in human freedom.)

The existentialist belief in human freedom is based on a phenomenological description of our everyday lives. In confronting situations where I must make a choice, I find myself facing an open range of possible courses of action where nothing compels me to choose one course of action over the others. Even in cases where I am not aware of making choices, a moment's reflection shows that I am in fact deciding my own life. Suppose that I show up for work faithfully each day, and I believe that I am compelled to do this because I need to make money to support my family. Does this mean that I am forced to do what I do? An existentialist like Sartre would say that it is self-deception to think I am *compelled* to be a conscientious worker, for I *could* always walk away from it all and join a monastery or turn to a life of crime. If I am choosing to let considerations of duty or money be deciding factors for me in this way, then this is my choice. What this suggests is that even in my habitual and seemingly 'automatic' actions I am actually *assuming* a particular identity for myself through my own free choices, and am therefore responsible for what I do.

Sartre tries to capture this idea by saying that humans are 'condemned to be free'. Because our being is 'in question' for us, we are always taking it over and giving it some concrete shape through our actions. And this means that, whether we are aware of it or not, in continuing to act in familiar ways we are constantly renewing our decisions at every moment, for we could always change our ways of living through some radical self-transformation. Moreover, since all criteria or standards for evaluating our actions are also freely chosen, in our actions we are also deciding what sorts of reasons are going to guide our actions. With no higher tribunal for evaluating reasons for acting, we are entirely responsible for what we do: we have 'no excuses behind us nor justifications before us'.

Existentialists generally hold that we are not only responsible for the direction our own lives take, but also for the way the world around us appears. This idea has its roots in Kant's view that the reality we experience is partly shaped by the constituting activity of our own minds, though existentialists differ from Kant in holding that our construction of reality depends on our own choices (see Kant, I. §5). Kierkegaard, for example, contends that one's sense of reality is determined by the 'sphere of existence' in which one lives, so that the person who lives the life of a pleasure-loving aesthete will experience a world that is quite different from that of the duty-bound follower of the ethical. Similarly, Nietzsche holds that reality is accessible to us only through some 'perspective' or other, that there is no way to get in touch with reality as it is in itself, independent of any point of view or framework of interpretation.

Sartre works out an especially strong version of this Kantian outlook by developing the theory of constitution in Husserl's phenomenology. Husserl held that the world we experience is constituted through the meaning-giving activity of consciousness. Sartre takes this account of constitution to mean that, because I shape the world around me through my meaning-giving activity, I am ultimately responsible for the way the world presents itself to me in

my experience. Thus if I have had some painful experiences as a child, it is up to me to decide what these mean to me. I can use them as an excuse for going through life feeling cheated, or regard them as challenges that will make me stronger. Sartre's point is not that there are no constraints on the ways I interpret my situation, but that constraints and obstacles gain their meaning from me, and since there are indefinitely many possible meanings any situation can have, there is no way to identify any supposedly 'hard' facts that could be said to compel me to see things one way rather than another. But this means, according to Sartre, that in choosing my interpretation of myself, I simultaneously choose the world. It is our own freely chosen projects that determine how reality is to be carved up and how things are to count. Sartre even goes so far as to say that, if a war breaks out around me, then I am responsible for that war, because it is up to me to decide what the war is going to mean to me in my life.

Other existentialists have tried to formulate a more tempered conception of freedom. Kierkegaard argues that, because being human involves both necessity and possibility, the extreme sort of 'anything-goes' freedom (such as that later envisaged by Sartre) would lead to the 'despair of lack of necessity'. Both Heidegger and Merleau-Ponty work towards a notion of 'situated freedom' according to which choice is always embedded in and dependent upon the meaningful choices disclosed by a specific social and historical situation. Beauvoir tries to show how institutions and social practices can cut off the choices open to women and oppressed groups. Finally, Nietzsche calls attention to the way biological and historical factors operate 'behind our backs,' influencing our decisions without our awareness. But even when such limitations are recognized, the belief that we can rise above our situations to be 'creators' remains fundamental to existentialist thought.

5 Everyday existence, anxiety and guilt

Though existentialists agree that people are free to choose their own fates, they also hold that most people are quite unaware of their freedom. This obliviousness results not from ignorance or oversight, but from the fact that we usually try to avoid facing up to our responsibility for our lives. For the most part we are 'fleeing' from ourselves, throwing ourselves into mundane concerns and drifting into standardized public ways of acting. Existentialists are generally quite critical of everyday social existence. As they see it, there is a strong temptation to let ourselves get swallowed up by the 'public', the 'they', the 'herd' or the 'masses'. We try to do what 'one does' in familiar situations, and we assume that our lives are justified so long as we are following the norms and conventions laid out in our social context. In throwing ourselves into the kinds of busy-ness characteristic of contemporary society, we become more and more effective at finding means to achieving socially accepted ends, but at the same time we lose the ability to understand what is at stake in existing. Life then becomes a disjointed series of episodes with no real coherence or direction, and we end up dispersed and distracted, lacking any basis for meaningful action.

Existentialists give similar accounts of how social existence undermines our ability to realize ourselves as individuals. Kierkegaard describes the way that being a well adjusted member of 'the public' levels everything down to the lowest common denominator, with the result that nothing can really *count* or *matter* to people anymore. Similarly, Nietzsche describes the way that our being as 'herd' animals domesticates us and deadens our creativity, and Heidegger points out the 'tranquilization' and 'alienation from ourselves' that results from our absorption in the familiar social world. Sartre presents an especially harsh picture of social relations. Since, in his view, people can only see each other as objects and not as free beings, the Look of the Other always objectifies me and pressures me into thinking I am just a brute thing. As each individual struggles to affirm their being as a free 'transcendence' against the objectifying Look of others, the result is inevitable conflict: in the words of a character in one of Sartre's plays, 'Hell is the others'.

But many existentialists also see a positive side to social life. Though Heidegger criticizes the temptation to self-loss in our participation in the 'they', he also holds that all our possible ways of interpreting ourselves ultimately come from the social context in which we find ourselves. For this reason, becoming authentic is not a matter of escaping society, but of embracing our social existence in the right way. Marcel's attitude toward social existence shows how different he is from Sartre. He criticizes the 'technocratic attitude' of mass society not because it leads to conformism, but because it breeds an 'atomic individualism' that robs us of our deep sense of connection and obligation to others. And Jaspers and Buber both emphasize the importance of 'I-Thou' relations in realizing a full and meaningful life.

Although existentialists differ in their assessment of social existence, they agree in thinking that our ordinary, day-to-day existence is shot through with concealment and self-deception. What can free us from this distorted

sense of things is not rational reflection, but a profound affective experience. This emphasis on the role of emotions or moods in giving us access to the truth about ourselves is one of the most characteristic features of existentialist thought. Kierkegaard and Heidegger, for example, focus on the disclosive role of anxiety in leading us to confront the fact that we exist as finite beings who must decide the content of our own lives. Jaspers' concept of 'limit-situations' refers to situations in which our ordinary ways of handling our lives 'founder' as we encounter certain inescapable 'antinomies' of life. For Sartre, the feeling of *nausea* shows us that it is up to us to impart a meaning to things, and *anguish* reveals our 'terrible freedom' to decide our own fates. Finally, Marcel refers to the experience of *mystery* in which we encounter that which defies our ability to gain intellectual mastery through our problem-solving skills.

Some existentialists also talk about an experience of the absurd that can come over us in our rationalistic age. Sartre claims that there are no ultimate grounds that validate our choices, so that any fundamental project we adopt must be absurd in the sense that it is ultimately unjustified. Camus' conception of the absurd is perhaps the best known of all, though it is not really representative of existentialist thought. In *The Myth of Sisyphus* (1942a), he describes the feeling of futility we can experience when we become aware of the repetitiveness and pointlessness of our everyday routines and rituals. For Camus, this feeling of the absurdity of existence, a feeling in which suicide begins to seem like a real possibility, is the most fundamental experience philosophy must confront.

Finally, many existentialists point to the experience of guilt as providing an insight into our own being. Existential guilt refers to something broader than the feeling we sometimes have when we have done something wrong. In its broadest significance, existential guilt refers to the fact that there is no pre-given legitimation or justification for our existence. Though we are creatures who feel the need for some 'reason for existing', we find ourselves thrown into a world where there is no higher court of appeals that could validate our lives. We are ultimately answerable only to ourselves. In a somewhat narrower sense, existential guilt can refer to the fact that, because we are always engaged in acting in concrete situations, we are implicated in whatever happens in the world, and so we always have 'dirty hands'.

6 Authenticity

Experiences like anxiety and existential guilt are important, according to existentialists, because they reveal basic truths about our own condition as humans. Everyday life is characterized by 'inauthenticity', and in our ordinary busy-ness and social conformism we are refusing to take responsibility for our own lives. In throwing ourselves into socially approved activities and roles, we disown ourselves and spin a web of self-deception in trying to avoid facing up to the truth about what we are. This picture of inauthentic existence is contrasted with a vision of a way of living that does not slide into self-loss and self-deception. Such a life is (using the term found in Heidegger and Sartre) 'authentic'. Authenticity suggests the idea of being true to yourself - of owning up to who you really are. However, it is important to see that authenticity has nothing to do with the romantic ideal of getting in touch with an 'inner self' that contains one's true nature, for existentialists hold that we have no pregiven 'nature' or 'essence' distinct from what we do in the world.

If authenticity is not a matter of being true to some core of traits definitive of the 'real me', what is it? For most existentialists, becoming authentic is first of all a matter of lucidly grasping the seriousness of your own existence as an individual - the raw fact of the 'I exist' - and facing up to the task of making something of your own life. Kierkegaard, for example, holds that the only way to succeed in becoming a 'self' (understood as an 'existing individual') is by living in such a way that you have 'infinite passion' in your life. This kind of intensity is possible, he thinks, only through a total, life-defining commitment to something that gives your life an ultimate content and meaning. Nietzsche is also concerned with getting us to take hold of our own lives in a more intense and clear-sighted way. To free people from the attempt to find some overarching meaning for their lives, he proposes the idea of eternal recurrence: the idea that everything that happens in your life has happened before in exactly the same way, and will happen again and again, an endless number of times. If we accept this, Nietzsche suggests, we will be able to embrace our lives as they are, on their own terms, without regrets or dreams about how things could be different. Heidegger suggests that, in the experience of anxiety, one confronts one's own 'naked' existence as 'individualized, pure and thrown'. As we become aware of our 'being-towards-death' in this experience, we will grasp the weightiness of our own finite lives, and we will then be able to seize on our own existence with integrity, steadiness and self-constancy.

Many existentialists agree that owning up to one's own existence requires a defining commitment that gives one's life a focus and sense of direction. For Kierkegaard, a religious thinker, self-fulfilment is possible only for the 'knight of faith', the person who has a world-defining relation to a particular being which has infinite importance (the eternal being who has existed in time, the God-man). For Heidegger, authenticity requires 'resoluteness', a commitment to some specific range of possibilities opened up by one's historical 'heritage'. The fact that the ideal of commitment or engagement appears in such widely different existentialist works raises a question about the distinction, first made by Sartre, between 'religious' and 'atheist' existentialists. Kierkegaard, Marcel and Jaspers are often grouped together as religious existentialists, yet there are profound differences in their views of the nature of religious commitment. Where Kierkegaard emphasizes the importance of relating oneself to a concrete particular, Marcel and Jaspers speak of a relation to the 'mystery' or to 'transcendence' (respectively). At the same time, so-called 'atheist' existentialists like Heidegger and Sartre tend to agree with Kierkegaard's view that being 'engaged' or having a 'fundamental project' is necessary to achieving a focused, intense, coherent life. The distinction between atheist and religious existentialists becomes harder to maintain when we realize that what is important for religious thinkers is not so much the factual properties of the object of commitment as the inner condition of faith of the committed individual. Thus, Kierkegaard says that what is crucial to faith is not the 'objective truth' about what one believes, but rather the intensity of one's commitment (the 'subjective truth').

The idea that intensity and commitment are central to being authentic is shared by all types of existentialists. Another characteristic attributed to an authentic life by most existentialists is a lucid awareness of one's own responsibility for one's choices in shaping one's life. For Sartre, authenticity involves the awareness that, because we are always free to transform our lives through our decisions, if we maintain a particular identity through time, this is because we are choosing that identity at each moment. Similarly, Kierkegaard and Heidegger talk about the need to sustain our identity at each moment through a 'repetition' of our choice of who we are. In recognizing our freedom to determine our own lives, we also come to accept our responsibility for who we are.

The notion of authenticity is supposed to give us a picture of the most fulfilling life possible for us after the 'death of God'. It calls on us to assume our own identities by embracing our lives and making something of them in our own way. It presupposes lucidity, honesty, courage, intensity, openness to the realities of one's situation and a firm awareness of one's own responsibility for one's life. But it would be wrong to think of authenticity as an *ethical* ideal as this is normally interpreted. First, becoming authentic does not imply that one adopts any particular moral code or follows any particular path: an authentic individual might be a liberal or a conservative, a duty-bound citizen or a wild-eyed revolutionary. In this respect, authenticity pertains not to *what* specific kinds of things you do, but *how* you live - it is a matter of the style of your life rather than of its concrete content. Second, in formulating their different conceptions of authenticity, many existentialists describe the ideal of authenticity in terms that suggest that it can be opposed to ethics as ordinarily understood. Kierkegaard, for example, says that it is possible that the knight of faith might have to 'transcend the ethical', and Nietzsche holds that authentic individuals will live 'beyond good and evil'. Thus, authenticity seems to have more to do with what is called the 'art of self-cultivation' than it does with ethics as traditionally understood.

7 Criticisms and prospects

Existentialism has been criticized from a number of different angles. One line of criticism holds that the emphasis on individual freedom and the rejection of absolutes in existentialism tends to undermine ethics; by suggesting that everyday life is 'absurd' and by denying the existence of fixed, binding principles for evaluating our actions, existentialists promote an 'anything-goes' view of freedom that exacerbates the nihilism already present in contemporary life. Camus' novel *The Stranger* (1942b), for example, has come under attack for glorifying immoral 'gratuitous acts' as a way of affirming one's own absolute freedom. In reply, supporters of existentialism have noted that the stance portrayed in the work is not at all typical of existentialist views, and that existentialism's ideal of freedom and its sense of the need for human solidarity after the 'death of God', far from undermining ethics, might provide a very good basis for a moral point of view in the modern world (see Existentialist ethics).

Other critics have tried to show that the basic picture of reality presupposed by existentialism necessarily leads to nihilism. Hans Jonas (1966) argues that existentialism, despite its avowed goal of overcoming Cartesianism, tends to introduce a new kind of dualism with its sharp distinction between humans (who are thought of as absolutely free centres of choice and action), and an inert, meaningless 'being' that is on hand for humans to interpret and

transform as they please. Not only does this extreme opposition exclude animals from the realm of beings with intrinsic worth, its view of humans as thrown into an indifferent universe seems to give us freedom only at the cost of making nothing really worthy of choice.

This line of criticism is closely connected to the claim, formulated by various postmodern theorists, that existentialism is still trapped within the assumptions of Humanism, a view now supposed to have been discredited. Humanism in this context means the view, central to modern philosophers from Descartes to Kant, that the human subject is immediately present to itself as a centre of thought and action, and that the rest of the universe should be thought of as a collection of things on hand to be represented and manipulated by the subject. Postmodern theorists claim that a number of intellectual developments in the last two centuries have made it impossible to accept this picture of the centrality of the subject. The semiotic theories of Saussure, for example, have shown how language tends to work behind our backs, controlling our capacities for thought and speech, and Freudian theory has shown how unconscious drives and desires lie behind many of our conscious thoughts and actions. Given these developments, it is claimed, we can no longer accept the idea that humans are capable of the sorts of self-transparency and self-determination that seem to be presupposed by existentialists like Sartre (see Postmodernism).

In reply to this objection, one might point out that most existentialists have been very critical of the Cartesian belief in the transparency of consciousness to itself. Such themes as being-in-a-situation, 'thrownness', embodiment and mystery show the extent to which many existentialists think of humans as embedded in a wider context they can never totally master or comprehend. Moreover, the existentialist description of humans as temporal beings whose 'present' is always mediated by what is projected into the future and retained from the past undermines any Cartesian conception of the immediate presence of self to self in self-awareness. Finally, as Sonia Kruks (1990) argues, postmodern theorists seem to have run up against a wall in their attempts to 'de-centre the subject'. Having identified the pervasive background structures that influence the thoughts and actions of subjects, these theorists now find it difficult to give an account of the kind of critical thinking they see as central to the postmodern stance. In Kruks' view, existentialists have much to offer postmodern theory in formulating the conception of a 'situated subjectivity' that will fill this gap.

It is not clear what the future holds in store for existentialism understood as a philosophical movement. Many of the ideas that sounded so exciting in Paris in the 1940s now seem terribly old-fashioned. Many of the more viable themes in existentialism have been absorbed into new philosophical movements, especially into hermeneutics with its emphasis on humans as self-interpreting beings (see Hermeneutics). While some existentialist writers have faded from the scene, others have become more and more influential (though not always *as* existentialists). There has been an explosion of interest in Heidegger and Nietzsche recently, and the works of Kierkegaard, Sartre and Beauvoir are widely discussed.

Whether or not existentialism as such will continue to thrive, it seems that there will always be a place for the style of critique of society and the concern with the concrete realities of life that are central to existentialist thought. As a reactive movement, existentialism challenged the uncritical assumptions of mainstream philosophy as well as the complacency of everyday social existence. In its more positive side, it attempted to counteract the tendency to self-loss in contemporary life by formulating a vision of the kind of coherent, focused way of living that would provide a basis for meaningful action. These certainly seem to be valuable aims, and it is likely that existentialist writers will always have important contributions to make toward realizing them.

See also: Existentialist thought in Latin America

CHARLES B. GUIGNON

References and further reading

Blackham, H.J. (1952) *Six Existentialist Thinkers*, New York: Harper Torchbooks, 2nd edn, 1959.(Dependable introductions to Kierkegaard, Nietzsche, Jaspers, Marcel, Heidegger and Sartre.)

Camus, A. (1942a) *Le Mythe de Sisyphe*, trans. J. O'Brien in *The Myth of Sisyphus and Other Essays*, New York: Knopf, 1955.(A classic account of the 'absurdity' of everyday life.)

Camus, A. (1942b) *L'Etranger* (*The Stranger*), trans. S. Gilbert, New York: Knopf, 1973. (Description of the kind of 'gratuitous act' discussed in §7 above.)

Cooper, D.E. (1990) *Existentialism: A Reconstruction*, Oxford: Blackwell.(An up-to-date and thorough survey of

existentialist thought. Helpful bibliography.)
Dilman, İ. (1993) *Existentialist Critiques of Cartesianism*, Lanham, MD: Barnes & Noble Books.(Clear and interesting account of how existentialism criticizes ideas inherited from Descartes; focuses primarily on Sartre and Marcel.)
Golomb, J. (1995) *In Search of Authenticity: From Kierkegaard to Camus*, London: Routledge.(Clearly written survey of existentialist conceptions of authenticity.)
Guignon, C. and Pereboom, D. (eds) (1995) *Existentialism: Basic Writings*, Indianapolis, IN: Hackett Publishing Company.(Core texts by Kierkegaard, Nietzsche, Heidegger and Sartre, with extensive introductions by the editors.)
Heidegger, M. (1927) *Sein und Zeit* (*Being and Time*), trans. J. Macquarrie and E. Robinson, New York: Harper & Row, 1962.(A seminal work: the first systematic presentation of such existentialist themes as death, guilt, conscience, authenticity and being-in-the-world.)
Jonas, H. (1966) 'Gnosticism, Existentialism and Nihilism', in *The Phenomenon of Life*, New York: Harper & Row.(Interesting criticisms of existentialism and especially of Heidegger.)
Kierkegaard, S. (1843a) *Enten-eller* (*Either/Or*), trans. H.V. Kong and E.H. Kong, Princeton, NJ: Princeton University Presss, 1987.(Develops the idea of different 'spheres of existence' in which humans live.)
Kierkegaard, S. (1843b) *Frygt og Baeven*, trans W. Lowrie in *Fear and Trembling and The Sickness Unto Death*, Garden City, NY: Doubleday, 1954.(Kierkegaard's poetic evocation of the 'Knight of Faith'.)
Kruks, S. (1990) *Situation and Human Existence: Freedom, Subjectivity and Society*, London: Unwin Hyman. (Excellent study of the relevance of French existentialist thought to current postmodern developments. Covers Marcel, Sartre, Beauvoir and Merleau-Ponty.)
Macquarrie, J. (1972) *Existentialism*, Philadelphia, PA: Westminster.(A dependable and clear survey of existentialist thought. Extensive bibliography listing older works on existentialism.)
McBride, W.L. (ed.) (1997) *Sartre and Existentialism: Philosophy, Politics, Ethics, The Psyche, Literature, and Aesthetics*, New York: Garland, 8 vols.(Collections of classic essays in English on existentialism, with volumes on its background and development.)
May, R., Angel, E. and Ellenberger, H.F. (eds) (1958) *Existence: A New Dimension in Psychiatry and Psychology*, New York: Basic Books.(Collection of classic readings in existential psychoanalysis.)
Nietzsche, F. (1892) *Also Sprach Zarathustra* (*Thus Spoke Zarathustra*), trans. W. Kaufman in *The Portable Nietzsche*, New York: Viking, 1954.(Nietzsche's vision of the *Übermensch* or 'superman'.)
Sartre, J.-P. (1943) *L'Être et le Néant: Essai d'ontologie phénoménologique* (*Being and Nothingness: An Essay on Phenomenological Ontology*), trans. H.E. Barnes, New York: Philosophical Library, 1956.(The most representative work of the existentialist tradition.)
Sartre, J.-P. (1946) *L'Existentialisme est un humanisme*, trans. B. Frechtman as *Existentialism*, New York: Philosophical Library, 1946.(Influential, brief statement of Sartre's existentialist views.)
Schrag, C.O. (1961) *Existence and Freedom: Towards an Ontology of Human Finitude*, Evanston, IL: Northwestern University Press.(A classic synthesis of different existentialist philosophers and themes.)
Solomon, R.C. (1972) *From Rationalism to Existentialism: The Existentialists and Their Nineteenth-Century Backgrounds*, New York: Harper & Row.(The best account of existentialism in its historical context.)
Wahl, J. (1969) *Philosophies of Existence: An Introduction to the Basic Thought of Kierkegaard, Heidegger, Jaspers, Marcel, and Sartre*, London: Routledge & Kegan Paul. (Difficult but trustworthy introduction by a respected French scholar.)
Warnock, M. (1970) *Existentialism*, Oxford: Oxford University Press.(Clear exposition of the basic concepts.)
Yalom, I.D. (1980) *Existential Psychotherapy*, New York: Basic Books.(Well written and insightful study of existentialist themes and their relevance to psychotherapy.)

Existentialist ethics

Central to existentialism is a radical doctrine of individual freedom and responsibility. On the basis of this, writers such as Sartre have offered an account of the nature of morality and also advanced proposals for moral conduct. Important in that account are the claims that (a) moral values are 'created' rather than 'discovered', (b) moral responsibility is more extensive than usually assumed, and (c) moral life should not be a matter of following rules. Existentialist proposals for conduct derive from the notion of authenticity*, understood as 'facing up' to one's responsibility and not 'fleeing' it in 'bad faith'. Authenticity, many argue, entails treating other people so as to encourage a sense of freedom on their part, although there is disagreement as to the primary forms such treatment should take. Some have argued that we promote a sense of freedom through commitment to certain causes; others that this is best achieved through personal relationships.*

1 Existential freedom

Existentialism has been described as a 'philosophy for living', an engagement with the realities of the human predicament. But some have argued that an existentialist ethics is impossible, on the grounds that the existentialist understanding of moral choice and value precludes advocating any particular choices and values. This charge is sometimes supported by citing (a) the variety of personal commitments - Christian, communist, even fascist - among writers in this tradition, and (b) the reluctance of some of these writers to express their commitments in moral terms (Kierkegaard sometimes seems to commend the person-of-faith's 'suspension of the ethical', and Nietzsche the 'higher' person's going 'beyond good and evil'). To assess these competing views, we need to understand (a) the existentialist emphasis on freedom and responsibility, (b) the resulting account of moral choice, values and rules, and (c) the ideal of authenticity.

Existentialism is not a sharply defined tradition. Here, we are concerned primarily with the views of Jean-Paul Sartre, but also with those of predecessors and contemporaries - including Kierkegaard, Heidegger and Gabriel Marcel - to the extent that these have affinities with Sartre's position. Central to the tradition is the idea of radical individual freedom and responsibility to which, as human beings, we are 'condemned'. This is not freedom in the sense of absence of constraints, for this varies from society to society. Nor is it simply the freedom to will which actions to perform. Rather, it is a person's freedom to 'make himself… since nothing foreign has decided what we are' (Sartre 1943: 554). We are responsible, therefore, for our characters and emotional make-up, and even for the situations in which we 'find ourselves', since circumstances only constitute a *situation* when interpreted by us in a certain way. Someone who interprets the mountain as a challenge is in a different situation from one who sees it merely as an obstacle. Such interpretations of the world are embedded in purposes for which we are responsible. Neither human beings nor the world, therefore, have a given nature or 'essence': it is through purposive activities that they become what they are.

Sartre spoke of our 'choosing' to be what we are, but others identify existential freedom with a capacity to stand back from our lives and change them. I may not have chosen my social situation, but I could reflect upon it, 'refuse' it and 'begin something else' (Merleau-Ponty 1945: 452). Although 'thrown' into an already interpreted world, we can, with 'resoluteness', become 'released' from the prevailing interpretations (Heidegger 1927: 163).

2 Values, responsibility and rules

From such accounts of freedom, existentialists have drawn conclusions concerning moral values, responsibility, and the place of rules in moral life.

(a) Values are, in a sense, 'subjective' functions of our interpretative activity rather than features of an independent reality. They are not, wrote Sartre, 'strewn in our path', but 'created' or 'invented'. This does not mean, however, that existentialists insist on a 'gap' between facts and values. Nietzsche (1886) denied that there are moral facts, arguing there are no facts at all, independent of human 'perspectives'. Values are *no more* 'subjective' than other features of the world, for even 'factual' descriptions are inseparable from evaluation. Values, as Heidegger stressed (1927), are not 'stuck on' to a world first articulated in a value-free manner, since all articulation reflects human 'projects' which are shaped by what matters to us, what we find worth pursuing.

It follows from this that the existentialist view of values is 'anti-foundationalist' (though see §4). There can be no

final justification for a particular value, since all justification presupposes an interpretative stance which is itself imbued with values. Some writers dramatize this 'predicament' by referring to the 'absurdity' which thereby infects our lives. It is not moral values themselves which are absurd, but the discrepancy between the serious commitment they require and the recognition that they lack any final foundation. (This 'absurd' discrepancy is distinct from the one, discussed by Albert Camus (1943), between our aspirations to knowledge and the world's 'refusal' to supply it - see Camus, A.).

(b) The existentialist view of moral responsibility is predictably demanding. First, for some writers, notably Kierkegaard, adoption of the moral point of view is itself an individual decision that cannot be dictated by reasons, since any such reasons would presuppose a commitment to that point of view. Even after it has been adopted, there remains a choice as to the *kind* of morality to follow. Sartre's wartime student, who must decide between joining the Free French and looking after his mother, is in effect having to choose between the 'collective' good and a 'morality of… personal devotion' (1946: 36).

Second, our responsibility allows for much less mitigation than is usually thought. We are, according to Sartre, 'without excuses' (1946: 34). The 'villains' in existentialist fiction often excuse their behaviour by citing their 'character' or the exigencies of their job, failing to admit that their 'character' and the demands incumbent on them are of their own making. Nor is my responsibility confined to what I *do*, for by *failing* to oppose a war, say, I have 'dirty hands' and share responsibility with those actually waging the war. Neither is it to the results of what *I* do that my responsibility is limited. By acting in a certain way, I give a green light to others to act in the same way, and hence bear responsibility for the results of their doing so.

It might be thought that for 'religious existentialists' our responsibility is reduced by the obligation to obey God's commands. But the tendency of these writers is to emphasize that faith itself is a free commitment and one, moreover, which reflects, rather than constrains, a person's moral values. My perception that it is God's command that I am obeying is therefore one for which I am accountable.

(c) Existentialists generally reject the familiar picture of morality as a system of rules or principles. This is not due to a predilection for motiveless *actes gratuits*, but for two other reasons. First, the familiar picture encourages people to go along with conventions and to respond to situations mechanically. One can 'do anything "on principle" and avoid all personal responsibility', Kierkegaard complained (1846: 85). Adherence to rules, moreover, detracts from the task of identifying what these situations really are, something that itself calls for moral deliberation since situations are functions of how circumstances are interpreted and evaluated (see §1). An obsession with following rules, then, implies abdication from responsible appraisal of our situations and deeds.

Second, the feasibility of basing decisions upon rules is much exaggerated. Moral rules are typically vague and, in many contexts, speak ambiguously. Sartre's student (see §2 (b)) can easily cite rules which, suitably construed, would endorse either of the courses of action contemplated. By their very nature, moreover, moral rules are tailored to recurrent, everyday situations, and are useless, therefore, in what Jaspers calls 'boundary situations'. How I should confront my death, for instance, is an issue so individual to me that the precedents set by other people can have no automatic bearing. Even much less dramatic circumstances can have a complexity and uniqueness which render the application of rules hazardous. As one of Simone de Beauvoir's characters puts it, 'I can't tell you [what to have done]… because I wasn't in your place' (1954: 646).

3 Authenticity

These claims about the nature of morality do not themselves provide practical guidance to the values and goals people should adopt. 'Anti-foundationalism' and hostility to rules may even seem to preclude this. If existentialism is to furnish a practical ethics, this will be on the basis of a notion, already visible in Kierkegaard's and Nietzsche's sketches of the 'subjective existing individual' and the 'overman' respectively, which later writers call 'authenticity'.

The authentic person is one who lives in clear, honest recognition of existential freedom. This is best understood in contrast with the various ways of being inauthentic that Sartre calls 'bad faith'. These are strategies for denying or disguising one's freedom and responsibility in order to minimize the 'anxiety' which full appreciation of these would induce. Inauthentic life, as Heidegger puts it, 'tranquillizes'. One such strategy is to regard oneself as being in the grip of unconscious drives; another is illustrated by the woman, in Sartre's example, who pretends that her

seduction by a man is something that is merely ‘happening’ to her (1943: 55-6).

The most common and insidious strategy, however, is that of identifying too closely with one’s ‘being-for-others’, with the images and expectations that other people have of one. Simone de Beauvoir (1949) complained of women’s tendency to match themselves with the image men have of them and then behave as women are ‘supposed’ to do (see Beauvoir, S. de). More generally, I surrender my ‘ownmost being’, as Heidegger calls it, when I do what ‘one does’, when I tailor my values and attitudes to those expected of me. I have surrendered to the anonymous ‘Them’, to Kierkegaard’s ‘Public’ or Nietzsche’s ‘Herd’. ‘Authentic’ is the translation of the German *eigentlich*, from a word meaning ‘own’. The authentic individual ‘owns’ and ‘owns to’ their life.

4 Reciprocal freedom

When Sartre says that an authentic decision cannot be ‘judged’, this seems to confirm the charge that authenticity is at best a recipe for how to ‘think and feel’, not for what ‘we ought to *do*’ (Warnock 1970: 125; original emphasis). But he also holds, qualifying his ‘anti-foundationalism’ apparently, that freedom is ‘the foundation of all values’ (1946: 51). This refers us to a theme stated succinctly by Jaspers, that ‘man becomes free only in so far as the other becomes free’ (1957: 85). This theme is inherited from German Idealists, like Johann Fichte, rather than from Kierkegaard and Nietzsche, whose heroes engage in solitary self-cultivation to the relative exclusion of concern for ‘the other’.

In contrast with his earlier view that people assert their own freedom at the expense of that of others, Sartre came to hold that ‘my freedom impl[ies] mutual recognition of others’ freedom’ (1983: 487). Although he writes that, in oppressing others, ‘the oppressor oppresses himself’, the point is not primarily a political one. The idea is that unless I recognize the freedom of others, I cannot properly appreciate my own and so live authentically. This is because my conception of myself is indelibly coloured by how I take other people to view me. Now I cannot suppose that they regard me as free unless I regard *them* as free. Hence my obligation to cultivate an authentic sense of myself as a free person commits me to cultivating in others a similar sense of their freedom. In Heidegger’s language, people exist authentically when each ‘frees the other in his freedom for himself’ (1927: 159).

But how is this cultivation of other people’s sense of their own freedom to proceed? For Sartre, it is primarily through commitment to large causes, such as resistance to fascism. But for others, it is in personal relationships of love and friendship that, in Marcel’s words, I ‘apprehend’ another person as free, ‘help him to be freed, [and] collaborate with his freedom’ (1935: 107).

The ideal of authenticity, it seems, does have broad implications for the treatment of other people, in which case the charge that existentialism cannot furnish a practical ethics is misplaced. That existentialists differ as to the precise implications of the ideal may reflect a human dilemma rather than confusion on their part. It is, in effect, the dilemma confronting Sartre’s student, for either of the actions he considers could be seen as honouring the freedom of others (in a way that his going off to peddle heroin to children could not). Existentialists, like the student himself, have no short answer to the question of which kind of concern - one for people at large, or one for those to whom one is intimately related - is the better expression of an authentic life. That is hardly cause for concluding that existentialism cannot make ‘any contribution to moral philosophy at all’ (Warnock 1970: 125).

See also: Autonomy, ethical; Existentialism; Existentialist theology

DAVID E. COOPER

References and further reading

Baldwin, T. (1986) ‘Sartre, Existentialism and Humanism’, in G. Vesey (ed.) *Philosophers Past and Present*, Cambridge: Cambridge University Press.(Acute analysis of Sartre’s lecture, with useful discussion of the later *Notebooks* (1983).)

Beauvoir, S. de (1949) *Le deuxième sexe, tome I, Les faits et les mythes, tome II, L’expérience vécue*, Paris: Gallimard; trans. H.M. Parshley, *The Second Sex*, London: New English Library, 1962.(Classic feminist text, which emphasizes women’s own responsibility for accommodating themselves to men’s image of them.)

Beauvoir, S. de (1954) *Mandarins*, Paris: Gallimard, 2 vols; trans. L.B. Friedman, London: Fontana, 1982. (Massive novel of human relationships and betrayals. A fictional vehicle for its author’s views on ethics.)

Camus, A. (1943) *Le Mythe de Sisyphe*, Paris: Gallimard; trans. J. O’Brien, *The Myth of Sisyphus*,

Harmondsworth: Penguin, 1986.(Famous discussion of absurdity, different from Sartre’s.)
Cooper, D.E. (1990) *Existentialism: A Reconstruction*, Oxford: Blackwell. (Revisionary account, but suitable as an introduction. Defends prospects for an existentialist ethics.)
Heidegger, M. (1927) *Sein und Zeit*, trans. J. Macquarrie and E. Robinson, *Being and Time*, Oxford: Blackwell, 1962.(Difficult but seminal twentieth-century work. Essential reading on authenticity and anxiety.)
Jaspers, K. (1957) ‘Philosophical autobiography’, in P. Schilpp (ed.) *The Philosophy of Karl Jaspers*, La Salle, IL: Open Court.(A leading and usually verbose existentialist’s briefest statement of his position.)
Kierkegaard, S. (1846) *The Present Age*, trans. A. Dru, London: Fontana, 1962.(A bitter attack on the ‘Public’s suppression of individuality.)
Macquarrie, J. (1977) *Existentialism: An Introduction, Guide and Assessment*, Harmondsworth: Penguin. (Focuses on ‘religious existentialism’.)
Marcel, G. (1935) *Être et Avoir*, trans. K. Farrer, *Being and Having*, London: Dacre, 1949. (Sensitive exploration of personal relationships by a neglected thinker.)
Merleau-Ponty, M. (1945) *Phénoménologie de la perception*, Paris: Gallimard; trans. C. Smith, *The Phenomenology of Perception*, London: Routledge & Kegan Paul, 1962.(Classic work of existential phenomenology; presents the most plausible account of existential freedom.)
Nietzsche, F. (1886) *Jenseits von Gut und Böse*, trans. W. Kaufmann, *Beyond Good and Evil*, New York: Vintage, 1966.(One of Nietzsche’s most important and closely structured works. Contains essential chapters on religion, the history of morals, ‘free spirits’ and nobility.)
Olafson, F. (1967) *Principles and Persons: An Ethical Interpretation of Existentialism*, Baltimore, MD: Johns Hopkins University Press.(Full and penetrating discussion of existentialist ethics. Usable as a sophisticated introduction.)
Sartre, J-P. (1943) *L’Être et le Néant: Essai d’ontologie phénoménologique*, Paris: Gallimard; trans. H. Barnes, *Being and Nothingness: An Essay on Phenomenological Ontology*, London: Methuen, 1957.(*The* classic existentialist tome, as difficult as its title suggests; includes discussion of ‘bad faith’, which is essential reading.)
Sartre, J-P. (1946) *L’Existentialisme est un humanisme*, Paris: Nagel; trans. P. Mairet, *Existentialism and Humanism*, London: Methuen, 1966.(Punchy, readable lecture, but not always a reliable expression of Sartre’s views. Contains the example of the student torn between joining the Free French and looking after his mother.)
Sartre, J-P. (1983) *Cahiers pour une morale*, Paris: Gallimard; trans. D. Pellauer, *Notebooks for an Ethics*, Chicago, IL: University of Chicago Press, 1992.(These posthumously published notes from the 1940s for a book which was never to be give Sartre’s fullest account of reciprocal freedom.)
Warnock, M. (1970) *Existentialism*, Oxford: Oxford University Press.(Introductory and not very sympathetic account of main existentialist figures. Rejects possibility of existentialist ethics.)

Existentialist theology

Existentialist theology is a term used to describe the work of a number of theologians, chiefly from the twentieth century, whose writings were strongly influenced by the literary and philosophical movement known as existentialism. Because of the diversity of the movement, it is difficult to say much that is illuminating about existentialist theology as a whole. In general, however, these theologians attempt to understand God in relation to the situation of the concretely existing human individual. Their analysis of human existence is one that emphasizes the freedom of individuals to shape their own identities through choices, and the paradoxical, ambiguous or even absurd character of the reality that humans encounter. Religious faith is seen as closely related to feelings of alienation and despair; faith may grow out of such emotions or it may provide the key to overcoming them, or both these relations may be present at once.

Though the designation of any particular theologian as 'existentialist' is a controversial matter, Karl Barth, Rudolf Bultmann and Paul Tillich are among the more important thinkers whose work exhibits existentialist themes. The entire movement has been strongly influenced, directly or indirectly, by the nineteenth-century Danish philosopher-theologian Søren Kierkegaard, while the works of the Russian novelist Fëdor Dostoevskii and the German philosopher Friedrich Nietzsche, both from the late nineteenth century, have also been important.

1 Kierkegaard's critique of Enlightenment-inspired theologies

Though Kierkegaard lived from 1813 to 1855, his writings were not widely known outside Denmark until the end of the nineteenth century. As these works were translated, Kierkegaard became increasingly influential, both directly and indirectly through his impact on people such as Heidegger. Much of his influence stemmed from devastating criticisms of nineteenth-century liberal Protestant thought.

Kierkegaard was a Christian writer who saw himself as a kind of missionary, whose task was to 'reintroduce Christianity into Christendom' (1967, vol. 1: 160). He believed that the established churches of Europe had departed from genuine Christianity, which required a risky commitment to and willingness to suffer with the crucified Christ. These churches had instead developed a complacent religion that identified Christian faith with conventional ethical virtue. Kierkegaard felt that what was needed was the abolition of 'Christendom', the easy assumption that all born in Christian countries become Christians as a matter of course; Christianity requires a decisive commitment on the part of the individual. Kierkegaard believed that the rationalistic theologies that had come out of the Enlightenment provided a justification for 'Christendom' because of their underlying optimistic view of human nature. Such theologies assumed that humans had a natural capacity to gain religious truth on their own, and so the necessity of historical revelation as the foundation of faith was eliminated. They also de-emphasized human sinfulness by a tendency to reduce the religious life to a form of morality.

Kant (§11), for example, had attempted to develop a 'religion within the limits of reason alone', in which the essential tenets of Christian faith, a pure moral religion, could be derived from practical reason. Hegel (§8) had constructed a philosophical system in which the content of Christianity was supposedly justified by being philosophically developed and articulated. Schleiermacher (§7) had developed a theology in which the general religious experience of dependence became the foundation of faith. Though all of these thinkers attempted to find some continued role for historical revelation, the truth of any such revelation was for them something to be vindicated by showing that the revelation was in agreement with the conclusions of critical reflection. Revelation was thus subordinated to reason. In contrast, Kierkegaard attempted to make the categories of faith and revelation once again central for religious understanding.

2 Insights of Kierkegaard and other nineteenth-century thinkers

In *Fear and Trembling* (1843), Kierkegaard tries to show that the life of faith, exemplified for both Jews and Christians by Abraham, cannot be understood solely in ethical terms, since Abraham's faith was shown by an action that could not be ethically justified, namely his willingness to obey God by sacrificing his son Isaac. In *Philosophical Fragments* (1844), he goes on to argue that the 'Socratic' assumption that humans can achieve religious truth on their own is incompatible with the Christian view that humans are sinners who must be transformed and given the truth by God himself. This perspective is developed further in *Concluding Unscientific*

Postscript (1846), in which Kierkegaard describes Christian faith as involving a particular kind of inward passion, or 'subjectivity', that cannot be the result of philosophical reflection. The *Postscript* contains a thoroughgoing critique of the Hegelian claim to have achieved a systematic understanding of the whole of reality; it emphasizes the unfinished, historical character of human existence and the failure of thinking to reach objective certainty on existential concerns. The Hegelian system has difficulties both at the beginning and at the end: it cannot really begin, as it claims, with no presuppositions, because of the situated, interested character of human existence; and it cannot finally be completed, which means that it cannot be a system. A system requires a completeness and finality that human existence does not allow.

Kierkegaard does not deny objective truth; he affirms that reality is a system for God (that is, from God's point of view, reality is a system). When it comes to what he calls essential truth, however, he affirms that this truth can only be achieved by the existing thinker through passionate commitment. The person whose life is qualified by the right kind of 'subjectivity' may be said to be living truly even if that person believes things that are not objectively true. The truth of Christianity, which assumes that humans cannot achieve this kind of subjective truth on their own, is grasped in the passion of faith that is created by God within the person who encounters God in the historical figure of Jesus of Nazareth.

The centre of Christian faith for Kierkegaard is the incarnation. Far from wishing to prove historically that Jesus of Nazareth was God incarnate, Kierkegaard insists that faith does not require objective certainty; in fact, it thrives on the risk generated by objective uncertainty (see Faith §5). The idea of God becoming incarnate is, to us humans, the absolute paradox; as such it cannot be an object of proof. We come to believe in Christ only when we receive at first hand from him an awareness of our sinfulness and need of revelation. Since the object of faith is paradoxical, it necessarily appears 'absurd' and constitutes a 'possibility of offence' to the sinful self, which insists on holding on to its own autonomy. The believer, however, is not offended. To believers, Christ as the God-man is the truth, the truth that is at the same time 'the way and the life'. Thus, the true Christian does not seek speculatively to understand the incarnation, but to live as a follower of Jesus, who is for the believer both the redeemer and 'the pattern'. The believer must be willing to suffer with Christ and break with the established patterns of the world.

Though Kierkegaard provides the most important nineteenth-century background for existentialist theology, the figures of Dostoevskii and Nietzsche also deserve mention. Though independent of Kierkegaard, both emphasize the idea that human thinking is passionate and interested rather than completely objective and disinterested, and both emphasize the ways in which the universe does not appear to make sense to existing human beings. In *Notes from Underground* (1864), Dostoevskii highlights the way in which choice, even irrational choice, preserves our individuality and humanity, and in *The Brothers Karamazov* (1879-80), he powerfully presents the ways in which human suffering and evil seem to make no sense in a world God has created. Yet in such an ambiguous world our freedom is preserved, a freedom which God honours by approaching us in the humble figure of Christ.

Nietzsche not only emphasizes the interested character of human thought, but provides a challenge for theology by reinterpreting Christian morality as a morality of weakness, based on resentment and envy. He challenges the Christian to examine the unconscious motives that may lie beneath what is presented as Christian love, and poses the question as to whether God is a threat to human freedom and autonomy.

3 Twentieth-century existentialist theologians

Karl Barth (1886-1968) is both the earliest and the most influential of twentieth-century theologians indebted to existential thought. Barth's first major work, a commentary on Romans (1919), was forged in the disillusionment created by the First World War and clearly shows the influence of Kierkegaard and Dostoevskii. In this work, Barth challenged the optimism of liberal Protestant thought by attacking its assumptions that human beings could develop a natural knowledge of God and that Christianity could be vindicated as a higher evolutionary stage in the natural development of such knowledge.

Barth stressed instead the 'infinite qualitative difference' between a God who is 'wholly other' and human beings. The knowledge of God cannot be achieved through human religious experience, but is completely dependent on God's self-revelation. This revelation can be given no philosophical foundation, but is understood as true in the existential situation in which human beings encounter God. The Church must listen to the word of the living God,

before whom human religious systems and institutions are judged and found wanting.

Barth's emphasis on the transcendence of God won great respect because of his opposition to the Nazi attempt to give a theological justification of their racist and nationalist policies. Barth clearly saw the idolatrous character of National Socialism and became a leading figure in the protest of the Confessing Church. He largely authored the 'Barmen Declaration' (1934), which argued strongly that the lordship of Christ over the state must be acknowledged and that Nazi ideology must be judged in the light of God's transcendent revelation.

In his later writings, notably the massive *Church Dogmatics* (1939-67), which occupied most of the rest of Barth's life, there is a movement away from any overt reliance on Kierkegaard or other existentialist thinkers. In his later work, Barth emphasizes not only the 'no' that God pronounces against autonomous human religious and political striving, but the 'yes' that God graciously grants in Christ. All of the classic Christian doctrines are reinterpreted in a Christocentric manner, since the eternal Word of God has become incarnate in Jesus of Nazareth. That Word is still seen as a call to decision and commitment.

Rudolf Bultmann (1884-1976) was linked to Barth early on, but the two quickly distanced themselves from each other. Bultmann is equally renowned as a theologian and New Testament scholar, and his theological work reflects his insights as an inventor of form criticism, a discipline that studies the New Testament in the light of the situation of the early Church. Bultmann agreed with Barth that the 'quest for the historical Jesus' carried on by liberal Protestantism had failed to provide a foundation for faith, and that it was imperative to hear again the message of God as proclaimed by the Church. Unlike Barth, however, he saw the need for a radical translation of the message of the New Testament if it was to be heard by twentieth-century people.

As Bultmann saw it, the New Testament message is presented in mythical language that reflects the prescientific cosmology of the day. That message must be demythologized if it is to be meaningful to contemporary persons, and the key to this reinterpretation is provided by the analysis of human existence found in the early writings of Martin Heidegger. When the New Testament is read in this way, the offer of salvation is understood as an offer of a new form of authentic existence. Human persons must face the limits of human existence, especially their own death, without fleeing to some form of inauthenticity in which their being is identified with some objective entity in the world. Instead, humans must recognize that their authentic being lies in possibility, in their potentiality-to-be. In a way that is not completely clear, even to many of Bultmann's followers, this form of authentic existence is offered to humans in the historical figure of Jesus of Nazareth. History must here be understood as meaningful narrative (*Geschichte*), rather than as 'what actually happened', the object of scholarly investigation (*Historie*). What is important about Jesus is the 'historic' significance of the narrative, its meaning for the individual facing a decision, not the 'historical' (factual) accuracy of the narrative, regarding whichBultmann takes a critical, sceptical view.

Paul Tillich (1886-1965) differs from both Barth and Bultmann in being less tied to biblical exegesis and more overtly engaged in philosophical reflection. Tillich practised what he termed a 'method of correlation', in which the questions posed by human existence philosophically interpreted and the answers provided by religious revelation (also philosophically interpreted) were to be understood as mutually illuminating. In his analysis, God is understood as 'being-itself' or 'the ground of being' rather than as a supernatural being. The modern religious task in the West is to discover the power of being-itself, to find the 'God above the God of theism'. It is the 'God above the God of theism' who appears when the traditional God, understood as a distinct personal being, has died.

Tillich reflects the concerns of Nietzsche, Sartre and other existentialists that the traditional God of theism would limit human freedom and autonomy. The human task is therefore to confront a world of meaninglessness and discover 'the courage to be', the power that allows us to accept ourselves even when there is no personal God who accepts us.

For Tillich, the power of being-itself is mediated through 'symbols' that are said to participate in the power of being. For Christians, the story of Jesus can be such a symbol, but Tillich does not claim that the Christian symbols are unique or universal, and their truth does not lie in depicting the way things are objectively. The power of a symbol depends partly on the ability of a community for which that symbol functions to perceive that power.

4 Critical problems and lasting contributions

The diversity of views found in existentialist theology demonstrates that it cannot be defined in terms of particular conclusions. In relation to Christianity, it is useful to distinguish between thinkers such as Kierkegaard and Barth, who reformulated Christian themes to show their relevance to the human predicament, from those such as Bultmann and Tillich, who use Christian language to present 'existentialist' views radically different from traditional religious faith. Kierkegaard and Barth use new thought-forms creatively to express and make sense of classical religious convictions, while Bultmann and Tillich use traditional language to express very nontraditional views.

While some would argue that such theologies simply reflect the cultural and political situation of twentieth-century Europe, the attempt to relate religious concerns to the human existential predicament cannot be dismissed so easily. Existentialist theology reflects the perennial encounter between philosophical reflection and religious commitment and illustrates the power of each to shape the other. The existentialist theologian challenges the philosopher by highlighting the 'interested' character of human thinking, while the philosopher challenges the theologian to make sense of religious commitments in the light of all that humans know and experience.

See also: Berdiaev, N.A.; Brunner, E.; Buber, M.; Existentialism; Existentialist ethics

C. STEPHEN EVANS

References and further reading

Barth, K. (1919) *The Epistle to the Romans*, trans. E.C. Hoskyn, Oxford: Oxford University Press, 1933.(A translation of the sixth edition. This work shattered classical liberal theology and initiated existential or 'dialectical' theology in twentieth-century Europe.)

Barth, K. (1934) 'The Barmen Theological Declaration', trans. D.S. Bax, in E. Jungel (ed.) *Christ, Justice and Peace*, Edinburgh: T. & T. Clark, 1992.(Issued by the German Evangelical Church, this strongly repudiated the Nazi attempt to justify their programme on theological grounds.)

Barth, K. (1939-67) *Church Dogmatics*, trans. and ed. G.W. Bromiley and T.F. Torrance, Edinburgh: T. & T. Clark, 1956-69.(In four volumes, with each volume published in several books. This is Barth's *magnum opus*, arguably the most significant work of systematic theology of the twentieth century.)

Brunner, E. (1937) *Man in Revolt*, trans. O. Wyon, London: Lutterworth, 1939.(Provides a theology of the human person heavily shaped by existential philosophy.)

Bultmann, R. (1957) *History and Eschatology*, Edinburgh: Edinburgh University Press.(Develops the distinction between what is historic as existentially meaningful and what is merely historical in the sense of being the object of historical science.)

Bultmann, R. (1984) *New Testament and Mythology*, trans. and ed. S. Ogden, Philadelphia, PA: Fortress Press.(A collection of some of Bultmann's most important writings.)

Dostoevskii, F. (1864) *Notes from Underground*, trans. A.R. MacAndrew, New York: New American Library, 1961.(This volume also contains *White Nights*, *The Dream of a Ridiculous Man* and selections from *The House of the Dead*.)

Dostoevskii, F. (1879-80) *The Brothers Karamazov*, trans. D. Magarshack, London: Folio Society, 1964.(A rich and complex novel that resists any brief summation, but deals with the conflict between faith and unbelief, the anxieties linked to freedom, and whether it is true that 'if God does not exist, then everything is permitted'.)

Kierkegaard, S. (1843) *Fear and Trembling*, trans. H.V. Hong and E.H. Hong, Princeton, NJ: Princeton University Press, 1983.(Kierkegaard's argument that faith is not reducible to morality, developed through reflection on the story of Abraham and Isaac. This edition also includes his *Repetition* of the same year.)

Kierkegaard, S. (1844) *Philosophical Fragments*, trans. H.V. Hong and E.H. Hong, Princeton, NJ: Princeton University Press, 1985.(Develops the idea of the incarnation as the absolute paradox that must be accepted through faith received directly from God.)

Kierkegaard, S. (1846) *Concluding Unscientific Postscript*, trans. H.V. Hong and E.H. Hong, Princeton, NJ: Princeton University Press, 1992.(Kierkegaard's philosophical critique of Hegel, with the development of his own analysis of human existence.)

Kierkegaard, S. (1967) *Journals and Papers*, trans. and ed. H.V. Hong and E.H. Hong, Bloomington, IN: Indiana University Press.(Entry 388 contains the famous phrase about reintroducing Christianity into Christendom, although the idea is present in many places in Kierkegaard's writings.)

Tillich, P. (1951-63) *Systematic Theology*, Chicago, IL: University of Chicago Press.(A three-volume comprehensive statement of Tillich's theological views.)

Tillich, P. (1952) *The Courage to Be*, New Haven, CT: Yale University Press.(Connects religious faith to a form

of courage that struggles with meaninglessness and despair.)
Tillich, P. (1957) *The Dynamics of Faith*, New York: Harper & Rowe.(Looks at faith from the viewpoint of existential philosophy and psychology.)

Existentialist thought in Latin America

In Latin America the thought and teaching of José Ortega y Gasset have been very influential. Their influence leaves an important mark on the substance of existentialism. The most effective aspect of Ortega y Gasset's philosophical conception was his thesis that humans do not have a nature, only a history. It is this concept that encouraged Latin American thinkers to create their own original thought as a product of their concrete historical circumstances. This entry will deal with Latin American existence, unique in its historical concreteness, and from this general view will attempt to construct a metaphysical theory of Latin America's historical endeavour.

Given that all historicist conception involves values that are objective in nature, it is not surprising that Latin American existentialism was profoundly influenced by Max Scheler and Nicolai Hartmann, together with the existential analysis of Martin Heidegger. Consequently, in opposition to the phenomenologists with a Husserlian orientation, implicit in Latin American existentialism, there is a phenomenological methodology in the interpretation of history as culture in accordance with the analysis of the dialectics of the structure of history. This opposes all possible perceptions of pure essences which might precede existence. The essence of existence is seen as progressive, constructing itself as it is bypassed by historical events.

Both in terms of the search for an original philosophy, which could be reduced to a philosophy of history (for example in the Orteguian philosophy of life) and in terms of a Heideggerian approach, Latin American philosophy applies a phenomenological method in its analysis. This would explain the fusion of phenomenology and existentialism in the works of Latin American philosophers. All Latin American phenomenological-existentialist philosophical effort is a struggle between the analysis and interpretation of the European currents and their search for the historical realization of the autonomous Latin American being.

1 Existentialism and literature

Existentialist thought centres its philosophical speculation around concrete human existence, focusing on lived experiences, or *Erlebnisse*, and emphasizing the freedom of choice and the subsequent responsibility of the human being as self, or *être-en-soi*. The statement given by Jean-Paul Sartre that man is condemned to be free implies the necessity of 'engagement' with one's own life project. Because of these principles, many existentialist thinkers chose literature as a vehicle through which they could better express the search for the authentic self. This choice often caused some confusion between philosophy and literature, for example in the work of Sartre and Miguel de Unamuno, the latter undoubtedly the precursor of the trend of thought that can be identified as strictly existentialist. In fact, novelist and thinker, de Unamuno, coined the definition 'El hombre de carne y hueso' (the man of flesh and bone).

In Latin America, the existentialist search for authenticity was similar to the quest for an authentic national identity. Specific examples can be found in Puerto Rico, with its special political situation and the continuous intellectual struggle to defend Puerto Rican and Latin American culture in its development as Hispanic culture, set against North American culture, which is held to be alien and imposed on Puerto Rico by colonial imperialism. Hence, the existentialist vision of both humanity and life has been influential for writers, such as Edelmira González Maldonado (1923-) and Emilio Díaz Valcárcel (1929-).

González Maldonado is a well-known and well-respected short story writer, although her literary work is not very extensive. Her most important short story collections include, *Crisis* (1973), *Soledumbre (Solitude)* (1976) and *Alucinaciones (Hallucinations)* (1981). Her literary work has focused on the themes of humankind's loneliness and the limitations and ambiguities of language and the problem of interpretation between speaker and listener.

Crisis (1973) was the first attempt to express the concept and feeling of existentialist anguish. Argentine thinker Angel Casares who wrote the prologue for the book states: 'This book is, then, a piece of life, a life'. The title of *Soledumbre* is the deliberate use of an ancient word meaning both solitude and remoteness. *Soledumbre* is clearly an existentialist writing as it focuses on the concrete *hic et nunc*, the space and time of the literary work of art. Vital space and time are set against radical loneliness representing the horizon of the self. One of González Maldonado's short stories (1976), whose title is a quote from Martin Heidegger's *Unterwegs zur Sprache (On the Way to Language)* (1959), *Die Sprache spracht* (language speaks), shows with its play on words, the impossibility

of communication between human beings and establishes the dark dimension of existentialism. This dimension is the alienation which occurs with the loss of self and identity and the way in which the quality of being unique and authentic is lost within humdrum existence. In *Alucinaciones (Hallucinations)* (1981) the writer searches for the solution to metaphysical anguish in the realm of dazzling fantasy.

Emilio Díaz Valcárcel felt a deep commitment to his country and its problematic future. Díaz Valcárcel believed that the solution must begin with the construction of the self as a Puerto Rican Latin American, although he considered that Puerto Rican writers such as himself are trapped in the gap between cultural and national identity and political identity. His most important novel with the striking title of *Figuraciones en el mes de marzo (Representations in the Month of March)* (1972), is considered to rank alongside the works of García Márquez, Fuentes and Vargas Llosa. In this novel the protagonist, a mask for the author himself, struggles with two of the most significant existentialist concepts: freedom and authenticity. He is incapable of accomplishing the task of the construction of the self and is dominated by feelings of solitude and impotence in the face of his existence. He escapes by entering a realm of strange and often frightening fantasy in which he is persecuted by something monstrous, bizarre and terrible. He tries to maintain the ties that bind him to his country and his family, but the conflict between dreams, patriotism, love and the necessity of writing drive him towards a fantasmagoric fantasy, the parody of literature, and worse, solipsism. At this point, his writing falls into the existentialist black abyss of nothingness.

2 Relevant thinkers: Argentina

Carlos Astrada (1894-1970) used the ontology of Heidegger as a starting point and addressed many Marxist issues. He centred his thought on the idea of the risk implied by all metaphysical speculation. The game unleashes its own process within an existential scope. Astrada (1942) concerned himself with certain structures which were not thoroughly developed in the work of Heidegger, such as, the concrete composition of the ‘Dasein’ (being-here) and the relation of the ‘Dasein’ to things, before moving towards the Marxist dialectic in the latter stages of writing (1970).

Jorge Casares was professor at the University of Buenos Aires from 1953 to 1964 and after moved to the University of Puerto Rico. Mostly concerned with Heidegger, he emphasized three fundamental points in Heideggerian metaphysics: the loss of being as domination of thought, being in its totality and the loss by and for humans of their essence. These three points, tied in with other existentialist themes, are thoroughly explored in his work *Sobre la esencia del hombre (On the Essence of Man)* (1979).

Vicente Fatone (1903-62) was professor of cosmology and metaphysics at the National University of Litoral and of the history of religion at the National University of La Plata. He mostly concerned himself with the mystical experience and philosophical thought in India. He focused on the theme of liberty, which he interpreted from different existentialist viewpoints (1952; 1972). Fatone maintained: ‘I am nothing but my freedom; but I am not my freedom because I am not: I have to be’ (1948: 25). For Fatone, humanity is not free, but creates its own freedom, because liberty is not a property of the human being. It is a condition by virtue of which man becomes man and woman becomes woman. Towards the end of his life, he took up the theme of mysticism again.

Eugenio Pucciarelli’s existentialist roots lay in his interest in ontological, historical and aesthetic topics. A student of Alejandro Korn and Francisco Romero he was preoccupied with the sciences of the spirit. He was more distinguished for his teaching than for his written work. He maintained that metaphysical testimony - perception of the a priori conditions that make experience and morality feasible - is perceived through experience and morality.

Ismael Quiles (1958) was professor of philosophy at the University of Buenos Aires. A Jesuit priest, he edited the journal *Ciencia y fe (Science and Faith)* (1944-64). His thought is known as the philosophy of insistence. He was not a strict existentialist, but took advantage of certain Heideggerian concepts featured in *Being and Time* (1927) in his metaphysics of interiority in which humans are fulfilled through the knowledge of the world (be-in), of others (be-with) and of God as the absolute transcendence of all finitude.

Anibal Sánchez Reulet (1910-) wrote extensively on Husserl and Nicolai Hartmann. He followed, in part, Romero’s direction and was influenced by Ortega y Gasset. He was interested in the aspiration to transcendence in human existence as an affirmation of freedom through objective value judgments. For Sánchez Reulet this transcendence, in the end, becomes a question of transcending the historical conditions which determine the

finitude of existence.

Miguel Angel Virasoro (1900-66) was professor at the University of Buenos Aires and at the Universities of Córdoba, Bahía Blanca and Mendoza. Virasoro (1942) was influenced by Hegel and later, Heidegger. His work was thus an existentialist synthesis of Hegelian dialectics (1957). His starting point was the intuition of the finitude in the concreteness of existence (see Argentina, philosophy in).

3 Relevant thinkers: Brazil

Gerd A. Bornheim is not very well known internationally. He distinguished himself among scholars of existentialism in Brazil with *Sartre: metafisica e existencialismo (Sartre: Metaphysics and Existentialism)* (1971). He dealt with problems of an ontological nature and, to a certain degree, reaffirmed Hegelian ontology. He showed deep insight in unravelling the metaphysical aspect of Sartre's thought as he tried to expose the metaphysical underpinnings of the phenomenon of being.

Euryalo Cannabrava (1908-91), like many other Latin American thinkers, was attracted by the phenomenological-existentialist movement. He drifted towards logical neopositivism as he was interested in the philosophy of language and in the attempt to reduce philosophy to a method. His most important existential writings were *Descartes e Bergson (Descartes and Bergson)* (1938) and *Seis temas de espírito moderno (Six Themes of Modern Spirit)* (1941). In *Descartes e Bergson*, he attempted to surpass rationalism and vitalism. In 1956 he published *Elementos de metodología filosófica (Elements of Philosophical Methodology)* in which he revealed his increasing distance from Heidegger.

Vicente Ferreira da Silva's philosophical endeavours raised consciousness about the need for philosophizing in Brazil. He was one of the most important and enthusiastic followers of the existentialist movement. He wrote *Ensaios filosóficos (Philosophical Essays)* (1964) (see Brazil, philosophy in).

4 Relevant thinkers: Mexico

Adolfo Menéndez y Samara (1908-54) was deeply influenced by Heidegger and his question of being and its constant emergence from nothingness. In his youth he wrote *Dos ensayos sobre Heidegger (Two Essays on Heidegger)* (1939), which reveals a preoccupation with ontology. His early death did not allow the metaphysical-existentialist aspect of his work to develop fully.

José Romano Muñoz (1953) was one of the first disseminators of Orteguian thought in Hispanic America. His principal preoccupation was with finding answers to the questions of those who are more interested in wisdom than in science, such as: Who are we? What role do we play in the world? What is our fate? These questions brought him to the field of ethics which became his major preoccupation under the influence of Max Scheler and Nicolai Hartmann. His book, *Hacia una filosofía existencial: al margen de la nada, de la muerte y de la náusea metafísica (Towards an Existential Philosophy: at the Edge of Nothingness, Death and Metaphysical Nausea)* (1953) had a strictly existentialist perspective.

José Sánchez Villaseñor (1950) was a Catholic thinker, who worked on the thought of Sartre from a critical perspective, particularly with respect to the assertion of the absurd nature of existence. He was also a critic of the thought of Ortega y Gasset (1949) (see Mexico, philosophy in).

5 Relevant thinkers: Peru

Luis Felipe Alarco was a student of Alejandro Deústua (1849-1945), professor of philosophy at the University of San Marcos, Lima. Deústua (1967) defended a philosophy of creative freedom in which liberty is the basis of all social and moral values. Alarco is considered to be the initiator and catalyst of the philosophical movement in Peru. He was principally interested in the thought of Nicolai Hartmann. His inclination towards the philosophy of Hartmann was owing to his interest in overcoming the propaedeutic character of Heideggerian phenomenological analysis. His three most representative writings are *Nicolai Hartmann y la idea de la metafísica (Nicolai Hartmann and the Idea of Metaphysics)* (1943), *Lecciones de metafísica (Lessons in Metaphysics)* (1947) and *Ensayos de filosofía prima (Essays on Prime Philosophy)* (1951).

Alberto Wagner de Reyna (1939) studied in Berlin with Nicolai Hartmann and in Freiburg with Heidegger. His

first work focused on the thought of Heidegger. Later, in an approximation to Gabriel Marcel (1889-1973), he became interested in several existentialist themes, including the priority of existence over essence and commitment and engagement in one's own historical reality, that were useful to Catholic thought. He was one of the first to propagate the thought of Heidegger in Hispanic America. He focused his attention on themes that are typically Heideggerian such as death and care. He attempted to demonstrate that existential analysis meant to clarify beings as such and that temporality is the ontological condition of human existence.

6 Relevant thinkers: Venezuela

Ernesto Mayz Vallenilla (1925-) is a phenomenologist who deals with existentialist thought (1966), influenced by Dilthey and Heidegger. One of the themes that has interested him most is technology as a means of overcoming the finitude imposed by nothingness from a temporal perspective. For Mayz Vallenilla, technology has an ontological foundation in its utility, one of the ingredients of existence according to Heidegger, but also in the will to dominate.

See also: Existentialism; Heidegger, M.; Husserl, E.; Literature, Philosophy in Latin American; Ortega y Gasset, J.; Phenomenology in Latin America

MARÍA TERESA BERTELLONI

References and further reading

Aftalión, E.R. (1954) *Peligrosidad y existencialismo (Danger and Existentialism)*, Buenos Aires: Librería Jurídica de V. Abeledo.

Alarco, L.-F. (1943) *Nicolai Hartmann y la idea de la metafísica (Nicolai Hartmann and the Idea of Metaphysics)*, Lima: Librería e Imprenta D. Miranda.(Alarco tries to overcome the propaedeutic character of Heideggerian phenomenology.)

Alarco, L.-F. (1947) *Lecciones de metafísica (Lessons in Metaphysics)*, Lima: Universidad Nacional Mayor de San Marcos, 4th edn, 1965.(A description of Alarco's view of metaphysics.)

(1951) *Ensayos de filosofía prima (Essays on Prime Philosophy)*, Lima: Imprenta Santa María.(An overview of philosophy by the the initiator of the philosophical movement in Peru.)

Astrada, C. (1942) *El juego metafísico (para una filosofía de la finitud) (The Metaphysical Game (for a Philosophy of Finitude))*, Buenos Aires: El Ateneo.(Astrada concerns himself with certain structures which were not thoroughly developed in the work of Heidegger.)

Astrada, C. (1943) *Temporalidad (Temporality)*, Buenos Aires: Ediciones Cultura viva.

Astrada, C. (1952a) *Existencialismo y crisis de la filosofía (Existentialism and the Crisis of Philosophy)*, Buenos Aires: Editorial Devenir, 2nd edn, 1963.

Astrada, C. (1952b) *La revolución existencialista: hacia un humanismo de la libertad (The Existentialist Revolution: towards a humanism of liberty)*, Buenos Aires: Editiones Nuevo Destino.

Astrada, C. (1970) *Martin Heidegger: de la analítica ontologíca a la dimensión dialéctica (Martin Heidegger: From Analytical Ontology to the Dialectical Dimension)*, Buenos Aires: Juárez Editor.(Includes bibliographical references to all the works of Heidegger.)

Augustín Martines, A. (1948) *Información sobre el existencialismo (Information About Existentialism)*, Santiago: El Esfuerzo.

Bornheim, G.A. (1971) *Sartre: metafísica e existencialismo (Sartre: Metaphysics and Existentialism)*, São Paulo: Editora Perspectiva.(A distinguished work among scholars of existentialism in Brazil.)

Canessa Prando, A. (1961) *Ensayos sobre existencialismo (Essa on Existentialism)*, Montevideo: Barreiro.

Cannabrava, E. (1938) *Descartes e Bergson (Descartes and Bergson)*, São Paulo: Amigos de Livro.(His most important existential work.)

Cannabrava, E. (1941) *Seis temas do espírito moderno (Six Themes of Modern Spirit)*, São Paulo: S.E. Panorama. (Another important work in the area of existentialism.)

Cannabrava, E. (1956) *Elementos de metodología filosófica (Elements of Philosophical Methodology)*, São Paulo: Companhia Editora Nacional.(Cannabrava establishes his distance from Heidegger.)

Casares, J. (1979) *Sobre la esencia del hombre (On the Essence of Man)*, Rio Piedras: Editorial de la Universidad de Puerto Rico.(Concerned with Heideggerian metaphysics.)

Castillo González, C. (1962) *Existencialismo y relativismo moral (Existential and Moral Relativism)*, Bogotá: Kelly.

Cortes Guimaraes, A. (1979) *Farias Brito e as origens do existencia no Brasil*, Rio de Janeiro: Tempo Brasileiro.

Deústua, A. (1967) *Ensayos escogidos de filosofía y educación nacional (Selected Essays on Philosophy and National Education)*, Huancayo: Concejo Provincial De Huancayo, Inspección de Cultura.(Defends a philosophy of creative freedom.)

Díaz Valcárcel, E. (1972) *Figuraciones en el mes de marzo (Representations in the Month of March)*, Barcelona: Seix Barral.(A novel which ranks alongside the works of García Márquez.)

Fatone, V. (1948) *El existencialismo y la libertad creadora: una crítica al existencialismo de Jean-Paul Sartre (Existentialism and Creative Freedom: a Critique of the Existentialism of Jean-Paul Sartre)*, Buenos Aires: Argos.(A discussion opposing the views of Sartre.)

Fatone, V. (1952) *Introducción al existencialismo (Introduction to Existentialism)*, Buenos Aires: Columba.(A general introduction to his own views of existentialism.)

Fatone, V. (1972) La existencia humana y sus filósofos (Human Existence and its Philosophers), in *Obras completas (Complete Works)*, Buenos Aires: Editorial Sudamericana.(A discussion of the view that humanity is not free, but creates its own freedom.)

Fernándes, J. (1986) *O existencialismo na ficcao brasileira*, Brazil: Universidad Federal de Goias.

Ferreira da Silva, V. (1964) *Ensaios filosóficos (Philosophical Essays)*, in *Obras Completas (Complete Works)*, São Paulo: Instituto Brasileiro de Filosofía.(Ferreira da Silva attempts to raise consciousness about the need for philosophizing in Brazil.)

González Maldonado (1973) *Crisis*, Río Piedros: Edil.(Expresses the concept and feeling of existentialist anguish.)

González Maldonado (1976) *Soledumbre (Solitude)*, San Juan: Instituto de Cultura Puertorriqueña.(A novel which focuses on the space and time of the literary work of art.)

González Maldonado (1981) *Alucinaciones (Hallucinations)*, San Juan: Instituto de Cultura Puertorriqueña.(A novel in which the writer searches for the solution to metaphysical anguish.)

Heidegger, M. (1927) 'Sein und Zeit', *Jahrbuch für Philosophie und phänomenologische Forschung* 8: 1-438; *Sein und Zeit*, Halle an der Salle: Max Niemeyer; trans. J. Macquarrie and E. Robinson, *Being and Time*, New York: Harper & Row, 1962; trans. J. Stambaugh *Being and Time*, Albany, NY: State University of New York Press, 1996.(Heidegger's most famous work. The unpublished second half of the work was to have shown that the meaning of being is time.)

Laborde, H. (1949) *El existencialismo, filosofía reaccionaria (Existentialism, Reactionary Philosophy)*, Mexico: Ediciones Noviembre.

Laman, M. (1967) *Existencialismo y literatura (Existentialism and Literature)*, Buenos Aires: Centro editor de America Latina.

Larroyo, F. (1951) *El existencialismo, sus fuentes y direcciones*, Mexico: Stylo.

Lertora, A.C. (1969) *Existencialismo y materialism dialéctico (Existentialism and Materialism)*, Buenes Aires: Ediciones Sílaba.

Letizia, F. (1986) *El existencialismo de Gabriel Marcel (The Existentialism of Gabriel Marcel)*, Mendoza: Universidad Nacional de Cuyo.

Lima, A.A. (1956) *O existencialismo e outros mitos de nosso tempo*, Rio de Janeiro: Agir.

Mayz Vallenilla, E. (1966) *Del hombre y su aliencación (Of Man and his Alienation)*, Venezuela: Instituto Nacional de Cultura y Bellas Artes.(Existentialism and alienation are discussed.)

Menéndez y Samara, A. (1939) *Dos ensayos sobre Heidegger (Two Essays on Heidegger)*, Mexico: Letras de México.(Reveals a preoccupation with ontology.)

Morales Dávila, M. (1963) *El existencialismo y la creación de la personalidad (Existentialism and the Creation of a Personality)*, Bolivia: Canela.

Penha, J. da (1982) *O que e existencialismo? (What is Existentialism?)*, São Paulo: Brasilense.

Quiles, I. (1958) *Más allá del existencialismo (filosofía in-sistencial: una filosofía del ser y dignidad del hombre) (Beyond Existentialism (In-sistential Philosophy: a Philosophy of the Being and the Dignity of Man))*, Barcelona: Luis Miracle.(Influenced by Heideggerian concepts.)

Romano Muñoz, J. (1953) *Hacia una filosofía existencial: al margen de la nada, de la muerte y de la náusea metafísica (Towards an Existential Philosophy: at the Edge of Nothingness, Death and Metaphysical Nausea)*, Mexico: Imprenta Universitaria.(A strictly existentialist perspective.)

Salas Subirat, J. (1954) *Carta abierta sobre el existencialismo*, Buenes Aires: Rueda.

Sánchez-Reulet, A. (ed.) (1954) *Contemporary Latin American Philosophy: a Selection, with an Introduction and Notes*, trans. W.R. Trask, Albuquerque: University of New Mexico Press.(An anthology of Latin American existentialist thought.)

Sánchez Villaseñor, J. (1949) *Ortega y Gasset, existencialista: un estudio crítico de su pensamiento y sus fuentes*

(Ortega y Gasset, Existentialist: A Critical Study of his Thought and its Sources), Chicago, IL: H. Regnery Co. (An extensive study of Ortega y Gasset.)

Sánchez Villaseñor, J. (1950) *Introducción al pensamiento de Jean-Paul Sartre (Introduction to the Thought of Jean-Paul Sartre)*, Mexico: Editorial Jus.(Wrote about Sartre from a critical perspective.)

Virasoro, M.A. (1942) *La libertad, la existencia y el ser (Freedom, Existence and Being)*, Buenos Aires: Imprenta López.(His starting point was the intuition of the finitude in the concreteness of existence.)

Virasoro, M.A. (1957) *Existencialismo y moral: Heidegger-Sartre (Existentialism and Morality: Heidegger-Sartre)*, Sante Fe, NM: Librería y Editorial Castellvi.(An existentialist synthesis of Hegelian dialectics.)

Vives, E.F. (1948) *Introducción al existencialismo (An Introduction to Existentialism)*, Chile: del Pacífico.

Wagner De Reyna, A. (1939) *La ontología fundamental de Heidegger: su motivo y significación (The Fundamental Ontology of Heidegger: its Motivation and Meaning)*, Buenos Aires: Editorial Losada, repr. 1945.(The author’s first work.)

Experiment

Experiment, as a specific category of scientific activity, did not emerge until the Scientific Revolution of the seventeenth century. Seen primarily as an arbiter in theory choice, there was little, if any, analysis of experimental techniques or, the ways in which data become transformed into established facts. Philosophical analysis of experiment was typically simplistic, focusing on the role of observation alone as the foundation for experimental facts. This was challenged by Thomas Kuhn who stressed the importance of background theory and beliefs in all perception, including (its role in) scientific experiment. This interconnection between theory and experiment severely undermined the idea that experiment could stand as an independent and objective criterion for judging the merits of one theory over another.

In the 1980s new philosophical analyses of experiment began to emerge, emphasizing the ways in which experiment could be seen to have a life of its own embodying activities that could supposedly be understood without recourse to theory. Factors important in the evaluation of experimental results as well as the ways in which laboratory science differs from its theoretical counterpart became the focus for a new history and philosophy of experiment. Consequently, further debates arose regarding the relationship of experiment to theory, and whether it is possible to provide a methodological framework within which experimental practice can be evaluated.

1 Historical overview

Although the use of trial and error methods to establish scientific results can be traced back to the Greeks, experiment as a specific category of scientific activity did not emerge until the Scientific Revolution of the seventeenth century. This occurred partly in response to the development of precision instruments like the telescope, microscope and pendulum clock, but also as a result of the growth of scientific academies that provided laboratory environments. Galileo's work on hydrostatics was perhaps the first instance of what we now think of as the experimental method. He showed how experiments on bodies floating in water could be used to confirm Archimedes' principles and refute those of Aristotle. Francis Bacon was the first to provide a classification for various kinds of experiment including the concept of a crucial experiment, which came to be thought of as a decisive test of one theory over its rival (see Crucial experiments).

Even though experimental demonstration was an important component of the Scientific Revolution its primary function was thought to be that of an arbiter in theory choice. There was little, if any, analysis of experimental techniques. Issues about the epistemic status of experiment were typically assimilated to debates about the nature of experience in general. Emphasis on experiment was associated with empiricism; consequently, not everyone saw experiment as the foundation for scientific knowledge. Descartes, for example, relied heavily on experiment in his scientific work; yet in his philosophical writings deduction from first principles or innate ideas took precedence over empirical inquiry as the ultimate ground for knowledge.

2 Experiment as distinct but subordinate

Although experiment occupied a role distinct from theory, at least in historical and philosophical writing on science, theory continued to hold the dominant position. Even the most respected physicists, such as James Clerk Maxwell and Albert Einstein, devoted little if any attention to experiment in their writings about science (even though Maxwell was a gifted experimenter and Einstein emphasized the role of thought experiments in the development of relativity theory (see Thought experiments). Philosophical and historical accounts focused either on crucial experiments, such as Fresnel's diffraction experiments which decided in favour of the wave over the corpuscular theory of light (see Optics §1), or experiments that were associated with important discoveries, for instance those that led Roentgen to the discovery of X-rays, or Millikan's measurements of the charge on the electron.

Although these kinds of experiment were discussed in articles and textbooks, much of the intricate detail that represents such an important part of experimental work, detail that is extracted from notebooks and investigations of laboratory practices, was excluded. As a result, the experiments were characterized in a highly idealized way, embodying none of the difficulties often encountered by working scientists in establishing specific results.

Philosophical analysis of experiment also presented a simplistic view of its function and its relationship to theory. Twentieth-century positivism (or logical empiricism) emphasized the importance of observation as the secure foundation for science, but provided no analysis of experiment itself. Nor did it question the authority of observational reports. For the positivists, the meaning of a particular term was its method of verification, and many considered simple observation reports entered in a log book or protocol to be the ultimate basis of science (see Meaning and verification; Theories, scientific §§1-2; Logical positivism §4). Because all of science had to be understood in terms of empirical concepts, experimental procedures were sometimes appealed to as providing meanings for theoretical terms. One way of doing this was to identify terms like ‘electric force’, which is not empirically observable, specifically the operations necessary to measure the force, with the instrument or pointer readings that indicated its presence. This particular method of assigning meanings or values to scientific concepts was known as operationalism (see Operationalism). Despite the importance of testability and observation reports for establishing meaning there was no analysis of how experimental results and procedures became stabilized and validated.

The emphasis on definitive observational reports was challenged most influentially by Thomas Kuhn in *The Structure of Scientific Revolutions* (1962) (see also Hanson, N.R.; Popper, K.R.). He argued that not only scientific experiments but also the interpretation of all experience was heavily dependent on background theoretical knowledge (see Observation §§3-4). Kuhn’s thesis, known as the theory-ladenness of observation, entails that two individuals, such as Kepler and Tycho Brahe, who hold different theories about the sun literally see different things when they observe it. One of them sees a stationary body around which the earth moves while the other sees a moving body rotating around the earth. As a result neither simple experience nor experiments (including crucial experiments) can provide a neutral ground for theory choice. The analysis of experiment that Kuhn provides is really an account of how psychological factors influence perception. Consequently, theory becomes the focal point from which the interpretation and relevance of all experimental data is assessed.

In response to theory-ladeness, sociologists of science denied that appeals to nature in the form of experiment were ever decisive in deciding between competing theories. Such decisions, and even the interpretation of data, were made simply on the basis of social and political factors that often stood outside the laboratory domain. Philosophers responded by accepting that observation was theory-laden while maintaining that those theories nevertheless corresponded to a determinate world, which, if characterized correctly, would reveal its true nature.

3 The rise of experimentalism

The emphasis on experiment as a distinct field of philosophical inquiry began with the work of Ian Hacking (1983). He urged philosophers to focus on activities like the building and operation of instruments, calculation and measurement, the creation of phenomena and manipulation of entities in the laboratory, all of which are part of an experimental culture that is distinct from the theoretician’s world and from questions that have traditionally dominated philosophy. As a result, a new set of philosophical issues arose: for example, how should we characterize the relationship of experiment to theory in light of Hacking’s claim that ‘ experiment has a life of its own’; is there a methodology of experiment that can furnish validating procedures and constraints for evaluating results; and how do the credibility of particular instruments and the theories implicit in their operation bear on the integrity of laboratory events?

Hacking stressed the independence from theory of various aspects of experimental practice, such as viewing and manipulating microscopic images and specimens, the manipulation of entities such as electrons using electron guns, and the creation of particles in accelerators. He claimed that most of these images and entities are robust and survive even radical theory change. Knowledge of them derives from their low-level causal properties (for example, mass, charge, and so on) that are involved in the design and operation of devices used to measure, observe and create them. However, it is not simply the fact that these entities can be experimented on that attests to their reality; they must be capable of being manipulated and used as tools in the investigation of other aspects of nature, not simply observed or detected in the laboratory. Hence, the goal is not just understanding the nature of these entities, but using them to transform nature. As Hacking (1983) remarked about electrons - ‘If you can spray them then they are real’.

Despite the importance of isolating experimental practice in this way, questions have arisen concerning the interplay of experiment and theory and whether Hacking’s reliance on manipulation can underwrite the realistic

status of experimental entities. A much greater theory than that of low-level causal generalization is needed to differentiate, for example, between specific kinds of particle interactions that take place, even in a straightforward case like the identification of a particle in a cloud chamber. In this context one must appeal to a scaffolding of theoretical knowledge that comes from electrodynamics, mechanics and so on, in order to distinguish an electron from a negatively charged pion. Similarly, when a beam is produced using a simple ruby laser, the apparatus utilizes causal properties such as the mass and charge of the electron, but in order to understand simulated emission, the very basis of laser action, we need the quantum theory of the electron. A great deal of theory, over and above the theory about how the apparatus produces the beam, is required before one can meaningfully say that one is producing laser light by the manipulation of electrons.

In addition to simply describing the activity of manipulating particles, interpretive controversies often arise in cases where different groups evaluate experimental results and practices in different ways. Experiments done in the late 1970s to investigate the existence of charmed quarks involved photon, neutrino and hadron beams. One group of physicists thought that quarks were the constituents of hadrons, and that, in that sense, quarks were being manipulated in order to investigate charm. Another group, who thought quarks were merely fictional entities that provided a mathematical description of hadron collisions, saw the manipulation of the beams as having nothing to do with quarks; still another group with a different account of quarks also had a different interpretation of the practice of manipulating hadron beams.

What the examples show, and indeed what an emphasis on experiment has clarified, is the degree of complexity involved in the relationships among experiments, phenomena, instruments and theories. The idea that an experiment can serve as an independent judge in a theoretical dispute is not necessarily vindicated by showing how experiment has a life distinct from theory. Very often an analysis of experimental work exposes subtle theoretical influences not obviously present in textbook histories. When Heinrich Hertz successfully produced electromagnetic waves in his laboratory in 1888, it was against the backdrop of competing views about electromagnetism from Maxwell and Helmholtz. The experiment acquired a different meaning and had different implications for British scientists (who saw it as definitive confirmation of Maxwell's electrodynamics) than it did for Hertz. There was no ambiguity that what were produced were electromagnetic waves, but the velocities of the waves contradicted a fundamental component of Maxwell's theory, a result ignored by the British, but which caused Hertz great concern. It was not as though the British could explain away the experimental anomaly, or that Hertz had reasons unknown to the British causing him to distrust his results. Hertz saw the experiments as proving that electrical action is propagated in space rather than at a distance, but did not sanction any particular account of how that process took place. Although Hertz's anomalous results were never explained, similar experiments done by others established the velocities predicted by Maxwell's theory, to Hertz's satisfaction. As we can see, the transition from laboratory event to experimental fact is one that involves more than 'reading the book of nature'.

Establishing an experimental fact also involves evaluating the interpretation of an experiment; this involves determining the legitimacy of a particular result, together with assessing the validity of background assumptions that are brought to bear in understanding the meaning and significance of experimental outcomes. There are competing philosophical views about the justification of an experimental result and the role it plays in the broader theoretical context. The evidence model emphasizes the connection between a particular result as the manifestation of a determinate world and a theory that attempts to describe that world. Thus Allan Franklin (1986) suggests that there is a set of strategies explicated in terms of Bayesian confirmation theory that determines the validity of an experimental outcome and the role it plays in deciding between competing theories (see Confirmation theory; Inductive inference §3). On this view, it is simply a fact whether or not a particular outcome occurs and whether it is relevant in a theoretical context. In contrast, the social model advocated by Andrew Pickering (1984) and others proposes that the way a piece of evidence is interpreted and judged relevant depends on sociological and political factors including the interests of specific groups of scientists who wish to promote particular theoretical views, rather than on its reflecting a determinate way the world is. Consequently there is no unique characterization of experimental evidence. Others like Peter Galison (1987) recognize a role for both natural and social constraints. By showing how experiment involves not just the work of a single scientist or group, but a complex community of subgroups, each with its own goals and standards, Galison shows the diversity that constitutes experimental work and the various levels at which the interplay between natural and social concerns emerges.

The philosophy of experiment recognizes that experimental practice is an autonomous part of a broader scientific

culture with distinct problems and issues. The literature has focused primarily on two kinds of concern: first, the 'how to' of experiment, which emphasizes laboratory practices and techniques that differentiate it from theoretical work; and, second, the epistemological debate about the proper characterization of experiment and its implications for scientific realism and theory confirmation. The latter presents two options. One involves attempts to provide a framework within which to understand and codify that practice - a 'theory' of experiment. The other focuses on the local nature of experimental practice by presenting it in a way that defies a unitary view or method describing how specific results are reached and legitimated. Emphasis on experiment has thus produced a new theoretical debate about what a proper philosophy of experiment and science ought to be.

See also: Experiments in social science; Scientific realism and antirealism

MARGARET C. MORRISON

References and further reading

Ackerman, R. (1985) *Data, Instruments and Theory*, Princeton, NJ: Princeton University Press.(A discussion of the role of instruments in providing stable data. Ackerman argues that experimental practice provides continuity in the face of theoretical change.)

Bogen, J. and Woodward, J. (1983) 'Saving the Phenomena', *Philosophical Review* 97: 302-52.(Introduction of distinction between data and phenomena that brings the complexities of experimentalism into distinctions of explanation, testing and theory structure, which should be understood as encompassing phenomena only.)

Franklin, A. (1986) *The Neglect of Experiment*, Cambridge: Cambridge University Press.(See Franklin 1990.)

Franklin, A. (1990) *Experiment, Right or Wrong*, Cambridge: Cambridge University Press. (Both books discuss the role experiments and their results play in scientific practice, the difficulties that arise in interpreting results and the methods involved in their justification. Some of the case studies are fairly technical. Both books expand on the arguments mentioned in §3.)

Galison, P. (1987) *How Experiments End*, Chicago, IL: University of Chicago Press.(An account of how historical, philosophical and sociological factors play a role in laboratory science. Technical in parts, but with a helpful introduction.)

Gooding, D., Pinch, T. and Schaffer, S. (1989) *The Uses of Experiment*, Cambridge: Cambridge University Press.(A collection of essays on a variety of topics in the history, philosophy and sociology of experiment. Some involve intricate detail but most are not technically difficult.)

Hacking, I. (1983) *Representing and Intervening*, Cambridge: Cambridge University Press.(The first systematic treatment of the role of experiment in philosophy of science. The history of theory-dominated philosophy is discussed in a clear and comprehensive way. Also expands on material in §3.)

Latour, B. and Woolgar, S. (1986) *Laboratory Life*, Princeton, NJ: Princeton University Press.(The book presents a sociological account of how scientific facts are established and is based on field work done at the Salk Institute.)

Morrison, M. (1990) 'Theory, Intervention and Realism', *Synthèse* 82: 1-22.(A critical discussion of some of the issues in philosophy of experiment, particularly Hacking's view that experiment and theory can be successfully separated.)

Pickering, A. (1984) *Constructing Quarks*, Chicago, IL: University of Chicago Press.(A detailed account of sociological factors in the practice of elementary particle physics, particularly the acceptance of the quark model and interpretation of important experiments. Technical in parts.)

Woodward, J. (1989) 'Date and Phenomena', *Synthèse* 79: 393-472.(Like the article with Bogen but in a more detailed discussion, the distinction between data and phenomena cautions against confused philosophical understanding of data as directly relevant to explanation and theory testing, which really concerns phenomena. Data, with their experimental dimension, play the role of observation but only in witness for claims about phenomena.)

Experiments in social science

Within social science the experiment has an ambiguous place. With the possible exception of social psychology, there are few examples of strictly experimental studies. The classic study still often cited is the Hawthorne experiments, which began in 1927, and is used mainly to illustrate what became known as the 'Hawthorne Effect', that is, the unintended influence of the research itself on the results of the study. Yet, experimental design is often taken within social research as the embodiment of the scientific method which, if the social sciences are to reach the maturity of the natural sciences, social research should seek to emulate. Meeting this challenge meant trying to devise ways of applying the logic of the experiment to 'non-experimental' situations where it was not possible directly to manipulate the experimental conditions. Criticisms have come from two main sources: first, from researchers who claim that the techniques used to control factors within non-experimental situations are unrealizable with current statistical methods and, second, those who reject the very idea of hypothesis-testing as an ambition for social research.

1 The logic of the experiment

The intention of the experiment is to organize inquiry so that it bears decisively on hypotheses and the theories from which they are deduced. The classic means of achieving this is to set up a controlled comparison between two groups which are as alike as possible at the outset of the experiment. Only one of the groups is subjected to the experimental treatment. Thus, if there is a difference in the experimental outcome, then this can only be due to the experimental treatment. Control is typically achieved by matching subjects and randomizing their allocation to the experimental or to the control group so that possible confounding factors are excluded as possible explanations of the outcome.

John Stuart Mill's formalization of the general logic of the experiment (Mill (1843) was used as a framework for Durkheim's efforts to formulate sociology as a discipline with a properly scientific method. In his *The Rules of Sociological Method* (Durkheim 1895), he proposed the 'method of concomitant variation' as the means by which the causes of 'social facts' can be identified (see Mill, J.S. §2). In his later study, *Suicide* (Durkheim 1897), applying this logic to relevant official statistics, he identified variations in the suicide rate among various subgroups and, by eliminating alternative explanations, arrived at what he regarded as the social causes of different types of suicide.

Distinctive about Durkheim's work was his attempt to exploit the logic of the experiment in non-experimental situations; that is, situations in which it is impossible directly to manipulate causal and possible confounding factors. His own procedure, using official statistics, was to identify variations in the suicide rate among ever finer delineated subgroups, and a precursor of what became known as 'variable analysis' which is now one of the standard social research methods (see Durkheim, É.).

2 Variable analysis

However, the refinement of what became known as variable analysis owes more to the work of Lazarsfeld and his colleagues than directly to Durkheim (Lazarsfeld and Rosenberg 1955; Campbell and Stanley 1963; Stouffer 1950). Immediately after the Second World War, mainly in the USA, considerable effort was devoted to developing research methods which, though used in non-experimental situations, would provide a means of approximating to experimental designs and so make empirical research more decisive for social theory. This period led to the refinement of social survey and questionnaire design, measurement techniques and statistical multivariate methods for the analysis of social survey data.

Unike the classic experimental design which directly allocates subjects to control and experimental groups, in variable analysis, typically using survey data, the grouping of respondents takes place after the data have been collected. In place of matching and randomization, relationships of interest to the researcher are examined by holding constant other factors which may influence such relationships. For example, suppose a researcher is interested in the effects of ethnicity on voting choice, then the task is to isolate the effects of ethnicity given that other factors, such as occupation, gender, age, income, level of education, and more, may also have an effect on voting. Grouping the respondents into such other categories, and then determining the strength of the association

between ethnicity and voting within each of them, is seen as holding constant any effects on voting choice which may arise from factors other than ethnicity. However, there is a limit to the number of factors which can be handled in this way before the statistical measures used to determine the relationship become unstable and, of course, only those factors on which data have been collected can enter into the analysis.

3 Problems with the experimental model

More seriously, the procedures of variable analysis depend heavily upon the random selection of respondents in order to neutralize the effects of possible confounding factors, whereas in the classic experimental design control is done through matching and the random allocation to control and experimental groups. In non-experimental research, the random selection of cases is an approximation to experimental randomization if it can be assumed that the population from which the sample is drawn is homogeneous; a reasonable assumption when dealing with a population of seeds, plants or germs, but less secure when dealing with populations of human beings. Indeed, one of the founding presuppositions of social science is that individuals are not isolated but related in a myriad of ways. It becomes harder to make the assumption that samples of individuals - which is the typical method employed in survey research - selected for their distinctiveness on one variable will vary randomly on others. Indeed, Lieberson makes the point that although most data in social research are non-experimental in origin, they are ‘treated as if they were truly experimental data…sliced, chopped, beaten, molded, baked, and finally artificially coloured until the researcher is able to serve us proudly with a plateful of mock experiment’ (Lieberson 1985: 4). The burden of this critique of the use of the experimental model in social research, is that controlling fails to deal adequately with selectivity in the control variables. That is, social research deals with situations in which people, apart from their birth, have not been randomly assigned to their respective social worlds. Social processes of selectivity operate throughout life among subgroups, with varying effects and with different magnitudes, with the result that controlling in social research is not a ‘benign procedure’ since it takes little account of these selectivity processes.

Lieberson’s arguments direct attention to what it is matching and randomization are intended to provide, namely, a basis for sustaining the *ceteris paribus* conditions which govern any theoretical hypothesis. Selectivity processes, under current practices of social research, cannot be controlled which means that such *ceteris paribus* assumptions cannot be sustained. Similar considerations, though from a very different perspective, emerge in consideration of even laboratory experiments using human subjects. It was noted earlier that one of the claims to fame of the Hawthorne experiments was the discovery of ‘reactive effects’, that is, effects of the experiment itself on its results. Although this was not a standard laboratory experiment, none the less, it did provoke serious considerations of the ways in which the very organization of experimental situations could itself violate *ceteris paribus* assumptions by introducing uncontrolled ‘reactive effects’ (Rosenthal and Rosnow 1969, especially the paper by Boring who suggests that such ‘reactive effects’ were originally the rationale underlying the idea of the control group). Among such ‘reactive effects’ are the meanings and understandings that subjects themselves bring to the experimental situation. As Cicourel 1964) points out, experimenters have to make assumptions about basic social processes in order to select subjects, organize their experimental activities, give them directions, and so on, but which are not explicated, let alone controlled. Rarely is the question of what is commonly perceived by subjects and experimenter as invariant about the social scene of the experimental situation addressed. As Milgram’s (1974) experiments on obedience show, subjects themselves give meanings to experimental situations, for example, as normal, as involving legitimate activities and trusting in the experimenters that they will come to no harm, and so on. So much so that they are ‘willing’ to engage in actions, such as supposedly inflicting intense pain on others, out of obedience to what is perceived as the normative structure of the experimental situation.

A broader issue of debate relevant to the experiment in social science is the nature of the social sciences themselves; a debate which, of course, enters into epistemology. The experimental model is very much a hypothesis-testing procedure and, in this respect, intended as a means of building general theories. However, as discussed earlier, in social science research the direct manipulation of the factors and the experimental conditions is not possible. There are approaches which question whether this model is appropriate given that the experiment presumes that a great deal is known about the phenomena in question in order that the method may be brought to bear decisively on hypotheses. Approaches which draw their inspiration from less positivistic theories of knowledge, such as symbolic interactionism and ethnomethodology, emphasize the importance of studying the social world as a naturally occurring world without the intervention of experimental or quasi-experimental

methods (Hughes 1990).

See also: Experiment; Statistics and social science

JOHN A. HUGHES

References and further reading

Campbell, D.T. and Stanley, J.C. (1963) *Experimental Designs and Quasi-experimental Designs for Research*, Chicago, IL: Rand McNally.(A valuable discussion of the efforts to incorporate the logic of the experiment into social research.)

Cicourel, A.V. (1964) *Method and Measurement in Social Research*, New York: Free Press.(A serious critique of the foundations of orthodox social research methods, including the experiment.)

Durkheim, É. (1895) *The Rules of Sociological Method*, ed. G.E.G. Catlin, trans. S.A. Solovay and J.H. Mueller, New York: Free Press, 1938.(A classic effort to formulate the methodological framework for a scientific sociology using Mill's specification of the experimental method. First published in French.)

Durkheim, É. (1897) *Suicide: A Study in Sociology*, trans. J.A. Spaulding and G. Simpson, London: Routledge & Kegan Paul, 1952.(An early and classic investigation using multivariate analysis. First published in French.)

Hughes, J.A. (1990) *The Philosophy of Social Research*, London: Longman, 2nd revised and enlarged edn.(A discussion of some of the philosophical problems arising from social research methods, including the experiment.)

Lazarsfeld, P.F. and Rosenberg, M. (eds) (1955) *The Language of Social Research*, New York: Free Press.(A pioneering collection of readings setting out the procedures of variable analysis.)

Lieberson, S. (1985) *Making it Count: The Improvement of Social Research and Theory*, Berkeley, CA: University of California Press.(An 'insider's' critique of the assumptions of experimental thinking in social research.)

Milgram, S. (1974) *Obedience to Authority: An Experimental View*, London: Tavistock.(A discussion of Milgram's much misunderstood experiments.)

Mill, J.S. (1843) *System of Logic*, London: Longman, 1970.(The classic formulation of the logic of the experiment.)

Roethlisberger, F.J. and Dickson, W.J. (1939) *Management and the Worker*, Cambridge, MA: Harvard University Press.(A presentation of the famous Hawthorne studies.)

Rosenthal, R. and Rosnow, R.L. (eds) (1969) *Artifact in Behavioural Research*, New York: Academic Press.(A collection of papers dealing with 'reactive effects' in social research.)

Stouffer, S. (1950) 'Some Observations on Study Designs', *American Journal of Sociology* 53: 355-61.(A classic statement on the need for experimental and quasi-experimental design in social research.)

Explanation

Philosophical reflections about explanation are common in the history of philosophy, and important proposals were made by Aristotle, Hume, Kant and Mill. But the subject came of age in the twentieth century with the provision of detailed models of scientific explanation, prominently the covering-law *model, which takes explanations to be arguments in which a law of nature plays an essential role among the premises. In the heyday of logical empiricism, philosophers achieved a consensus on the covering law model, but, during the 1960s and 1970s, that consensus was challenged through the recognition of four major kinds of difficulty: first, a problem about the relation between idealized arguments and the actual practice of explaining; second, the difficulty of characterizing the underlying notion of a law of nature; third, troubles in accounting for the asymmetries of explanation; and, four, recalcitrant problems in treating statistical explanations.*

Appreciation of these difficulties has led to the widespread abandonment of the covering-law model, and currently there is no consensus on how to understand explanation. The main contemporary view seeks to characterize explanation in terms of causation, that is, explanations are accounts that trace the causes of the events (states, conditions) explained. Other philosophers believe that there is no general account of explanation, and offer pragmatic theories. A third option sees explanation as consisting in the unification of the phenomena. All of these approaches have associated successes, and face particular anomalies.

Although the general character of explanation is now a subject for philosophical debate, some particular kinds of explanation seem to be relatively well understood. In particular, functional explanations in biology, which logical empiricists found puzzling, now appear to be treated quite naturally by supposing them to make tacit reference to natural selection.

1 Early history

Thinking about explanation goes back at least to Aristotle, whose discussion of four kinds of causation in the *Posterior Analytics* can properly be viewed as distinguishing modes of scientific explanation (see Aristotle §9). In the modern period, the writings of Hume, Kant and Mill offer many insights on causation, laws and regularities in nature that, sometimes explicitly, sometimes implicitly, propose doctrines about the character of scientific explanation (see Hume, D. §2; Kant, I. §§4-7; Mill, J.S. §5). However questions about scientific explanation became sharply focused in the mid-twentieth century, with the emergence of an orthodoxy about scientific explanation, which, despite its later demise, stands as one of the most significant achievements of the movement known as ‘logical empiricism’ (see Logical positivism §4). The writings of Karl Popper (1959), R.B. Braithwaite (1953), Ernest Nagel (1961) and especially, C.G. Hempel (1965), articulate an influential conception. Namely, scientific explanations are viewed as arguments in which a statement describing the fact (or regularity) to be explained is derived from premises, at least one of which is a law of nature (see Hempel, C.G. §3). The underlying idea is that scientific explanations provide understanding by showing that the phenomena to be explained should be expected as a consequence of the general laws of nature.

2 The covering-law model

One important and much-discussed species of scientific explanation according with this general conception is *deductive-nomological* explanation (D-N explanation). In cases of this type, the argument is deductive, so that the statement describing the phenomenon to be explained (the *explanandum*) is a deductive consequence of the premises advanced in giving the explanation (the *explanans*). D-N explanations may be provided for explananda that describe particular facts or for explananda that announce general regularities. In the former case, there is a simple schema which exhibits the form of the explanation

$$\frac{\begin{array}{c} C_1, C_2, \ldots, C_n \\ L_1, L_2, \ldots, L_m \end{array}}{E}$$

where the statement *E* is the explanandum, describing the fact to be explained, the statements $L_1, L_2, \ldots, L_m$ are laws of nature, and the statements $C_1, C_2, \ldots, C_n$ describe particular facts (such as initial conditions). It is not hard to construct arguments that accord with this schema and which seem to explain their conclusions: derivations

in classical Newtonian dynamics that deduce the trajectories of bodies from force laws and initial conditions supply many examples.

Not all explanatory arguments are deductive. The logical empiricist orthodoxy admitted *inductive-statistical* explanations (I-S explanations) as well as D-N explanations. In an I-S explanation, the explanandum is inferred inductively from premises at least one of which is a probabilistic law, for example a statement that assigns a value to the probability with which a particular trait is found among members of a specific class. Thus, to cite a famous example of Hempel's, we may explain why a child, Henrietta, contracted measles, by noting that she has been in contact with another child, Henry, who has measles, and that a large percentage of children who come into contact with measles' patients (say 99 per cent) subsequently come down with measles. Imitating the schema for D-N explanation, we can present this modest derivation as follows:

Henrietta has been in contact with Henry, and Henry has measles.
The frequency with which children in contact with measles' patients subsequently acquire measles is 99 per cent.

$$\frac{\qquad\qquad (0.99) \qquad\qquad}{\text{Henrietta has measles}}$$

Here, the rule indicates that the inference from premises to conclusion is inductively strong, rather than deductively valid; the figure in brackets (0.99) reveals the strength of the inference. I-S explanations have to meet several requirements. First, the numerical strength of the inductive inference must be high (close to 1). Second, the explanans must meet a requirement of maximal specificity: there must not be known further premises which, if added to the explanans, would change the strength of the inference - as, for example, the inductive reasoning would be modified if we knew that Henrietta had received a measles' shot, and that children given such shots have a very low probability of acquiring measles.

Plainly, the explanations that scientists and others actually put forward do not look much like these stripped-down arguments. The logical empiricists claimed only that the everyday provision of explanations could be reconstructed by identifying arguments of D-N or I-S form, and that these reconstructions brought into the open what it was about the explanations that enabled them to fulfil their function. In the 1940s and 1950s, many scholars were happy to concede that the covering-law model of explanation, which assimilated explanations to arguments with laws among their premises, worked well as a reconstruction of explanations in the natural sciences - especially in physics and chemistry - but there were important debates about the application of the model to the social sciences and to explanation in everyday life. Controversy focused in particular on the activity of historical explanation. Historians construct detailed narratives that appear to explain particular events - the outbreak of the American Civil War or Henry VIII's dissolution of the monasteries. If the covering-law model is correct, then a proper reconstruction of these accounts must expose general laws. Are there indeed general laws in history? Or are the general laws that underlie historical explanation simply psychological laws that connect the motivations of historical actors with their actions (see Explanation in history and social science)?

3 Four kinds of difficulty

Troubles with history aside, the covering-law model appeared remarkably successful, a rare example of a convincing solution to a philosophical problem. Yet, in the 1960s, it came under sustained attack, and, by the end of the decade, it had been almost entirely abandoned. Four separate kinds of consideration contributed to this swift reversal of fortune.

First was a complaint, articulated by Michael Scriven (1962), that perfectly satisfactory explanations can be given, and understood, by people who are quite ignorant of the covering-laws that are essential to the supposed reconstruction. It is easy to explain to a friend why there is a mess on the floor by pointing out that your arm knocked the open ink bottle off the desk at which you were writing. Perhaps it would be possible for a knowing philosopher of science to cite the general laws that govern the behaviour of the bottle and the spilled ink, but this knowledge seems entirely irrelevant to the episode in which the chagrined mess-maker explains what has occurred. At the heart of Scriven's complaint lay the recognition that the covering-law model had failed to show how the idealized derivations that supposedly highlight how the explanatory work is done are adapted, in specific local situations, to transmit understanding from one person to another. Without a *pragmatics* of explanation, an account of how the ideal arguments that fit particular logical forms relate to what people actually do in giving explanations, it was possible to challenge the claim that the structures exposed by the logical empiricists reveal the crucial

features that make the explanation successful.

A second difficulty resulted from continued inability to provide a satisfactory account of natural laws. From the earliest formulations of the covering-law model, its champions had insisted that not every generalization counts as a law. So-called *accidental* generalizations cannot discharge any explanatory function: it may be a timeless truth about the universe that all ball games played by a red-haired left-hander who forgoes lunch are won by the opposite team, but that accidental generalization sheds no light on the outcome of any particular game (see Laws, natural §2). Prior to Nelson Goodman's formulation of a cluster of difficulties surrounding counterfactuals, induction and laws (1956), the problem of distinguishing laws from accidental generalizations appeared an interesting challenge to the logical empiricist project (see Goodman, N. §3). Once the depth of Goodman's 'new riddle of induction' had become apparent, it seemed impossible to find a solution within the constraints that empiricists allowed themselves (see Induction, epistemic issues in).

A third trouble emerged from the recognition that, even if the distinction between laws and accidental generalizations could be drawn, the covering-law model would still be too liberal. Introducing an example that was to become famous, Sylvain Bromberger (1966) pointed out that the model is blind to certain asymmetries in explanation. We can explain the length of the shadow cast by a flagpole by deriving a statement ascribing the pertinent numerical value from premises identifying the height of the pole and the angle of elevation of the sun, together with the law of the rectilinear propagation of light. This derivation fits the D-N schema beautifully. The trouble, however, is that we can produce a modified argument, according equally well with the D-N schema, by interchanging the premise that identifies the height of the pole with the conclusion: the height of the flagpole is deducible from the length of the shadow, the elevation of the sun and the law of rectilinear propagation of light. This new derivation does not seem explanatory, for it appears wrong to explain the heights of poles (or, more generally, the sizes of physical objects) in terms of the measurements of the shadows they cast. Scrupulous about appealing to causation, logical empiricism had tried to construct an account of explanation without invoking causal notions that would offend Humean sensibilities (hence, in part, the difficulty of characterizing natural laws). Bromberger's critique suggests that the omission of causal concepts assimilates cases that are importantly different: after all, it is tempting to characterize the difference between the two derivations by pointing out that the height of the flagpole *causes* the shadow to have the length it does, but that the length of the shadow does not cause the flagpole to have the height it does.

Perhaps the most influential difficulty, was the fourth, which focused on the failure of the account of statistical explanation. Alberto Coffa (1974) probed the conditions required of I-S explanation, revealing that they involved an essential reference to the state of knowledge, which made it impossible to develop a concept of a true inductive explanation. Coffa's critique complemented the work of Richard Jeffrey (1969), who had earlier argued that it is possible to explain individual events that do not have high probability in the light of background conditions, and thus that the high probability requirement was also defective. At the same time, Wesley Salmon (1970) worked out, in considerable detail, an account of statistical explanation that, like Jeffrey's, rejected the thesis that explanations are arguments. Central to Salmon's account was the idea that we explain by citing probabilistically relevant information. In the early versions of his model of statistical explanation, Salmon proposed that probabilistic explanations gain their force from the recognition that the probability that an individual has a property has been raised. Schematically, the information that a is F helps explain why a is G when the probability of something's being G is increased if that thing is F (more exactly: $\mathrm{P}(G/F) > \mathrm{P}(G)$). In this way, Salmon was able to respond to a difficulty noted earlier by Scriven - we may explain the fact that the mayor has paresis by noting that he previously had untreated syphilis, even though the frequency with which untreated syphilitics contract paresis is small (around 15 per cent). On Salmon's account, noting that the mayor had untreated syphilis gives an enormous boost to the probability of his having paresis, raising it from the baseline figure of close to 0 to about 15 per cent.

4 Picking up the pieces

Salmon's account of explanation was deliberately motivated by the felt need to allow for explanations in the indeterministic contexts of contemporary physics. His approach dovetailed neatly with attempts, like those of Patrick Suppes (1970), to fashion a conception of causality that would no longer be restricted to deterministic situations (see Causation §4; Determinism and indeterminism). From 1970 to the present, one important strand in

contemporary theories of explanation has taken explanation to consist in delineating the causes of events, and has tried to honour Humean concerns about the invocation of causality by providing a theory of causation that will define causal relations in statistical terms. The simplest account of probabilistic causality would propose that *A* is causally relevant to *B* just in case $\mathrm{P}(B/A) \neq \mathrm{P}(B)$. Unfortunately, this account is too simple. As Hans Reichenbach pointed out in the 1950s the inequality will obtain when *A* and *B* are both effects of a common cause (Reichenbach 1956). Thus further conditions must be imposed to identify the statistical relations constitutive of probabilistic causation.

Since the 1970s a number of different proposals have competed to inherit the position of orthodoxy once occupied by the covering-law model. Most popular have been causal approaches to explanation, and, initially, proposals to ground explanation in a detailed conception of probabilistic causality promised to answer (or sidestep) the four principal difficulties outlined above. However, it has proved remarkably difficult to work out a satisfactory account of explanation along these lines, and a number of critiques, most notably that by Nancy Cartwright (1983), have cast doubt on the viability of the enterprise. Faced with powerful objections, champions of causal approaches to explanation have pursued one of two options. One is to continue to honour Humean concerns about the causal relation, and to seek an analysis of causation that will not make use of metaphysical notions that empiricists consider dubious. The most thorough attempt to carry out this programme has been undertaken by Wesley Salmon (1984), who has attempted to develop Reichenbach’s account of causation in terms of the fundamental notion of mark transmission. The alternative approach is to declare victory by taking some causal notion as an unanalysed primitive, resisting Humean scruples about how we might know how to apply this notion as misguided (perhaps the proposals of Humphreys (1992) and Cartwright (1989) should be viewed as embodying this approach).

One evident attraction of the causal programme is that it provides an immediate response to the problem posed by the asymmetries of explanation. However, not all current theories view explanation as a matter of tracing causes. In recent years, Bas van Fraassen (1980), Peter Achinstein (1983) and Peter Railton (1981) have all made important contributions to the pragmatics of explanation, and the first two authors have defended the view that the enterprise of seeking substantive necessary conditions that apply across all contexts in which people seek and give explanations is misguided. The danger is that such pragmatic theories of explanation reduce the enterprise to triviality. For any explanation-seeking question and any proposition we choose, it seems that we can construct a context in which that proposition is licensed as an adequate (or even a perfect) explanatory answer to that question (see Kitcher and Salmon 1987, and, for a response, Lloyd and Anderson 1994).

A third cluster of positions stays close to the covering-law model’s conception of explanations as arguments, proposing that explanatory arguments are not distinguished singly but emerge from the best way of systematizing our body of knowledge. Michael Friedman (1974) and Philip Kitcher (1981) have developed (different) accounts of explanation that take arguments to be explanatory if they belong to a system of arguments that best unifies our beliefs. One virtue of this approach is its ready provision of an analysis of theoretical explanation; its principal difficulties lie in formulating appropriate criteria for unification and for addressing the asymmetries of explanation.

The present debate echoes themes from earlier chapters in the history of philosophy. Hume’s scruples about causation loom behind some efforts to articulate causal theories of explanation, those who oppose Hume on causation (like Cartwright) sometimes seem to harken back to Aristotle, and the unification approach has made an explicit connection with Kant. Perhaps these affinities suggest that contemporary debates about scientific explanation turn on larger metaphysical questions that need to be confronted directly (see Unity of science).

5 Functional explanation: a recent success story

Ironically, after the fragmentation of the consensus on the covering-law model of explanation, considerable progress has been made on studying a species of explanation that, despite careful studies by Hempel and Nagel, was always somewhat difficult for logical empiricist orthodoxy. Biologists often appear to explain the presence of a trait or structure by identifying its function, and it is not clear how such explanations should be assimilated to a D-N (or an I-S) schema. Thanks to pioneering work by Larry Wright (1976), contemporary philosophers of biology are largely agreed on a central idea: functional explanations are abridged versions of explanations in terms of natural selection. The identification of the function is thus seen as picking out the kind of selection pressure that causes the trait (or structure) to become originally established (or, maybe, to be maintained). The details of this idea are worked out in different ways by different authors, but the selectionist (or etiological) account of functional

explanation appears to provide a philosophically satisfactory reconstruction of parts of biological practice (see Evolution, theory of; Functional explanation).

A possible moral of the comparative success in studying functional explanation is that philosophers may be too ambitious in seeking general theories of explanation. Pragmatists, like van Fraassen, sometimes suggest that there are many different kinds of successful explanation, and that there may be no interesting general conditions that all must meet. Perhaps the most fundamental issue confronting the theory of explanation today is whether it is reasonable to seek a theory of explanation across all contexts and all epochs or whether the study of scientific explanation should be more local, concentrating on specific types of explanation.

See also: Meyerson, É.

PHILIP KITCHER

References and further reading

Achinstein, P. (1983) *The Nature of Explanation*, New York: Oxford University Press.(See §4. A sophisticated exploration of the practices of asking for and giving explanations.)

Braithwaite, R.B. (1953) *Scientific Explanation*, Cambridge: Cambridge University Press.(See §1. A major work from the heyday of logical empiricism.)

Bromberger, S. (1966) 'Why-Questions', in R. Colodny (ed.) *Mind and Cosmos*, Pittsburgh, PA: University of Pittsburgh Press.(See §3. An important critique of covering-law models.)

Cartwright, N. (1983) *How the Laws of Physics Lie*, New York: Oxford University Press.(See §4. A sustained attack on traditional conceptions of law, cause and explanation.)

Cartwright, N. (1989) *Nature's Capacities and their Measurement*, New York: Oxford University Press.(See §4. A defence of approaching causation and explanation in non-Humean terms.)

Coffa, J.A. (1974) 'Hempel's Ambiguity', *Synthèse* 28: 141-63.(See §3. A penetrating critique of Hempel's views about statistical explanation.)

Fraassen, B. van (1980) *The Scientific Image*, Oxford: Oxford University Press.(In particular, chapter 5. An important work on theory and explanation in science. See §4.)

Friedman, M. (1974) 'Explanation and Scientific Understanding', *Journal of Philosophy* 71: 5-19.(See §4. An important presentation of the unificationist approach to explanation.)

Goodman, N. (1956) *Fact, Fiction, and Forecast*, Indianapolis, IN: Bobbs-Merrill. (See §3. The classic source of the difficulties faced by logical empiricists in explicating the notion of a law of nature.)

Hempel, C.G. (1965) *Aspects of Scientific Explanation and Other Essays*, New York: Free Press.(Contains the *locus classicus* of the covering-law model. See §§1- 2.)

Humphreys, P. (1992) *The Chances of Explanation*, Princeton, NJ: Princeton University Press.(See §4. An articulation of a causal approach to explanation.)

Jeffrey, R. (1969) 'Statistical Explanation vs Statistical Inference', in N. Rescher (ed.) *Essays in Honor of Carl G. Hempel*, Dordrecht: Reidel, 104-13.(See §3. An important critique of covering-law models of explanation.)

Kitcher, P. (1981) 'Explanatory Unification', *Philosophy of Science* 48: 507-31. (See §4. An account of explanation in terms of unification.)

Kitcher, P. and Salmon, W. (1987) 'Van Fraassen on Explanation', *Journal of Philosophy* 84: 315-30.(See §4. A critique of van Fraassen's pragmatic theory of explanation.)

Lloyd, E. and Anderson, C. (1994) 'Empiricism, Objectivity, and Explanation', *Midwest Studies in Philosophy* 18: 121-31.(See §4. A defence of van Fraassen's pragmatic approach to explanation.)

Nagel, E. (1961) *The Structure of Science*, New York: Harcourt Brace.(See §1. A classical logical empiricist text that takes explanation to be central to scientific practice.)

Pitt, J. (1989) *Theories of Explanation*, New York: Oxford University Press.(A collection containing many important contemporary articles.)

Popper, K. (1959) *The Logic of Scientific Discovery*, New York: Basic Books.(See §1. A seminal logical empiricist work, containing an account of explanation as subsumption under law.)

Railton, P. (1981) 'Probability, Explanation, and Information', *Synthèse* 48: 233-56. (See §4. An important contribution to the pragmatics of scientific explanation.)

Reichenbach, H. (1956) *The Direction of Time*, Berkeley, CA: University of California Press.(See §4. A work containing a seminal discussion of probabilistic causation.)

Rubin, D.-H. (1994) *Explanation*, Oxford: Oxford University Press.(A collection explicitly designed to complement the Pitt volume.)

Salmon, W. (1970) ‘Statistical Explanation’, in R. Colodny (ed.) *The Nature and Function of Scientific Theories*, Pittsburgh, PA: University of Pittsburgh Press, 173-231.(See §3. A thorough scrutiny of Hempel’s views about statistical explanation, together with an important rival theory.)

Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*, Princeton, NJ: Princeton University Press.(See §4. The *locus classicus* for one of the leading contemporary accounts of scientific explanation.)

Scriven, M. (1962) ‘Explanations, Predictions and Laws’, in H. Feigl and G. Maxwell (eds) *Minnesota Studies in the Philosophy of Science*, vol. III, Minneapolis, MN: University of Minnesota Press.(See §3. An important early critique of covering-law models.)

Suppes, P. (1970) *A Probabilistic Theory of Causality*, Amsterdam: North Holland. (See §4. A pioneering study of probabilistic causation.)

Wright, L. (1976) *Teleological Explanation*, Berkeley, CA: University of California Press.(See §5. The major source of the most widely accepted current account of functional explanation.)

Explanation in history and social science

Historians and social scientists explain at least two sorts of things: (a) those individual human actions that have historical or social significance, such as Stalin's decision to hold the show trials, Diocletian's division of the Roman Empire, and the Lord Chief Justice's attempt to reform the English judicial system; (b) historical and social events and structures ('large-scale' social phenomena), such as wars, economic depressions, social customs, the class system, the family, the state, and the crime rate. Philosophical questions arise about explanations of both kinds (a) and (b).

Concerning (b), perhaps the most pressing question is whether explanations of this sort can, ultimately, be understood as merely explanations of a large number of individual human actions, that is, as a complex set of explanations of the first kind, (a).

A causal explanation is an explanation of something in terms of its event-cause(s). Some explanations under (b) appear not to be causal explanations in this sense. There are two ways in which this appears to happen. First, we sometimes seem to explain a social structure or event by giving its function or purpose. This seems to be an explanation in terms of its effects rather than by its causes. For example, it might be claimed that the explanation for a certain social custom in a tribal society is the way in which it contributes to social stability or group solidarity. An explanation of a thing in terms of its effects cannot be a causal explanation of that thing. Second, we sometimes seem to cite social structure as the explanation of something. Whatever a social structure is, it is not itself an event, and since only (it is often said) events can be causes, such a 'structural' explanation does not seem to be a causal explanation.

A second question, then, about explanations of kind (b) is whether some of them, at any rate, are genuinely non-causal explanations, or whether functional and/or structural explanations of this sort can be seen as special sorts of causal explanation.

Explanations of kind (a) are a proper subset of explanations of human actions generally. Although some of the discussion of these issues began life as a distinct literature within the philosophy of history, it has now been absorbed into philosophical action theory more generally. Even so, a question that remains is just which proper subset of human actions are the ones of interest to the historical and social sciences: how can we discriminate within the class of human actions between those in which historians or social scientists have a legitimate interest and those outside their purview?

1 Holistic explanation

Many explanations in history and social science are explanations of large-scale events or structures such as wars, invasions, economic depressions, social customs, the class system, the family, the state, and the crime rate. A plausible thought is that these large-scale phenomena are merely patterns, or complexes, of individual actions, that the former are, in principle, ontologically reducible to the latter. This idea is sometimes captured by using the terminology of wholes and parts (see Mereology). Large-scale social events or structures are wholes, whose parts are the individual actions which constitute them. If this is so, then explanations of such large-scale social and historical phenomena are nothing more than complexes of explanations of those constituent actions of individuals which are the former's parts.

The term 'methodological individualism' is ambiguous (see Ruben 1985). Sometimes it is said to be the view that social wholes are nothing more than the sums of their parts. In this sense, it is not a thesis about language, that is, about the translatability without remainder of the language of social science into the language of individual psychology, although some writers have expressed it in this form. Nor is it a thesis about the methodology of social scientific practice, as its name might misleadingly suggest. It is a metaphysical thesis about what social phenomena really are, namely that they are just, or are wholly constituted by, individual persons with beliefs, desires, and other mental states, who stand in various relations with one another. This states the metaphysical thesis both in terms of identity and constitution. It is controversial whether the numerical identity and constitution relations are the same relation, and no statement of methodological individualism carefully delineates these two alternatives.

At other times, methodological individualism is described as a view about explanation: that the explanation of social facts is ultimately in terms of individual psychological facts (perhaps with purely physical facts added to the psychological ones). Stated in this way, one needs an account of social fact, in order to clarify the scope of the claim. What makes a fact a *social* fact? Presumably, social facts are not only facts about large-scale social phenomena: the fact that LaGuardia was a mayor is a social fact, because a social property, being a mayor, was true of that person. LaGuardia was himself not a large-scale social event or structure, and being a mayor is a property, neither an event nor a structure, large-scale or otherwise. Presumably, the fact that he was a mayor is a social fact, because being a mayor is a social property. So, in the end, methodological individualism as a view about explanation will need a perspicuous account of social property.

The important point, though, is that the metaphysical doctrine of methodological individualism does not, by itself, warrant any conclusion about the explanation of social facts. No such inference, from a metaphysical premise to an epistemological conclusion, could be valid without the addition of further premises. The methodological individualist thesis about the explanation of social facts in terms of facts about individuals would require some additional argument independent of the metaphysical thesis, in order to claim credibility.

So even if it were true that social wholes were nothing more than the ‘sums’ of individual actions, nothing immediately follows about the explanation of the former in terms of the latter. This point is further strengthened by the reminder that, whereas full constitution (and identity, if this is a distinct relation) is an extensional relation, explanation is non-extensional. That is, if {a and b} fully constitute w, and if a = d, b = e, and w = x, then {d and e} fully constitute x. How we describe constituents and what they constitute is not relevant to the truth of constitution (or identity) claims.

But this is not the case for explanation. With explanation, description matters: something can be explanatory or get explained when described in one way but not in another (see Explanation). Even though red is the colour of ripe strawberries, one might be able to explain why bulls charge at certain flags waved in front of them by the fact that the flags are red, but not by the fact that the flags are the colour of ripe strawberries. So it may well be that even if a social whole, say a war or an economic slump or the state, is nothing more than, is wholly constituted by, the actions of individuals, no explanation of the war or the slump or the state, could be adequate, even in principle, if expressed in the language of individual action. The relationship between the metaphysical truth of methodological individualism (if it were one) and the parallel methodological individualist thesis about ultimate explanation in the social sciences seems far more complicated than writers on these issues have typically assumed (see Methodological individualism §2, 3).

2 Functional and structural explanation

There is a ‘strong’ view that all explanations of large-scale social phenomena are functional explanations. Even if this view is false, it seems undeniable that there are at least some functional explanations in history and the social sciences (see Functionalism in social science §1). Certain customs, practices or institutions in a society might be explained by the role they play: religion; etiquette; bureaucracy; education or the state might be explained by their function for a society or for a specific group within the society in question. There seem to be functional explanations in biology as well, but the parallel between biology and social science on this issue is controversial, and it is better, therefore, to consider functional explanation in social science and history on its own, without considering the issue in biology and attempting to establish a general account of functional explanation, equally applicable to both fields.

Functional explanation in social science should be sharply distinguished from purposive (or intentional) explanation of individual human action. In the latter case, explanation is by way of the beliefs and desires or intentions of the agent. It has sometimes been alleged that functional explanation in social science also requires reference to mental states such as belief and desire, but mental states attributable to a group as a whole rather than to individual persons. Such an implausible view, with its notorious idea of a group mind, seems unnecessary, if we refuse to assimilate functional and purposive explanation. The former does not require reference to any mental states whatever, whether of an individual or of a group.

Are functional explanations a type of causal explanation? It might seem not, since such explanations seem to be in terms of the beneficial effects of a thing rather than in terms of its cause(s). But one plausible suggestion is that,

for an adequate functional explanation, there must be some causal 'feedback' loop: if education is to be explained by the way in which it helps a certain group retain power, then it is insufficient that education merely leads to the group's retaining power. If that were all there were to it, we would have an ordinary causal explanation of the group's power retention by the existence of the system of education.

On the contrary, what we had hoped for was the explanation of the education system in terms of the group's retaining power. On one view of the matter, to have the latter explanation, the group retaining its power must itself lead, by a causal loop, to further expansion or entrenchment of education. As such, we can see how the fact that education has the result of a group's power-retention is explanatory of the education itself, since education having this effect is itself a cause of more such education. Perhaps the stronger the group gets, the more it reinforces and develops the educational system. Education's effect further strengthens the system of education itself. We can also see, on this account of functional explanation, that functional explanation is merely a more complicated kind of causal explanation, involving a feedback loop in the causal chain (Elster 1983).

What about explanations in which social structure of some sort is cited as what does the explaining? How could these be causal explanations? For example, the kinship structure in a specific society might be said to explain why certain persons do and do not marry other persons to whom they are related in certain ways. One way in which to understand these cases might be to distinguish more carefully just what it is that one is explaining. Does the social structure explain some token event, like Tom marrying Rita, or does it explain some more general structural fact, like the fact that men must marry their dead brother's male-childless wife? If the former, then presumably the kinship system by itself cannot explain the specific marriage of Tom to Rita: the beliefs and wants of Tom and Rita, at the very least, will have to be included. Social structure cannot have an effect on the agents' behaviour directly; its effect must be mediated through, or combined with, the psychology of the agents. These mental states must figure in the explanation, and they will ensure that the explanation of the specific marriage is an ordinary causal explanation (see Structuralism in social science §1).

On the other hand, if what we are explaining is the structural fact about levirate marriage, then that feature of the kinship structure might follow from certain other features of the kinship structure, in the sense that the fact can be derived, or follows, from them. Writers in this tradition of explanation, basing themselves on themes and modes of explanation in the philosophies of Hegel and Marx, often speak of the underlying 'logic' of the social structure working itself out, thereby signifying that such explanations are akin to logical deductions (Zeleny 1980). In something like this sense, we could 'explain' some theorem of Euclidean geometry by deriving it from axioms and definitions. In the case of structural explanations in the social sciences, social structures are sometimes said to have their own dynamic, based on what they essentially are, and to develop according to their own inner laws or tendencies. If the concept of explanation were to be extended in this way, then there may well be structural explanations in social science which are not causal explanations. Even so, the thought that all explanations of events are causal explanations remains undisturbed; whatever such structural explanations are explanations of, they do not seem to be events in any sense.

3 Individual action explanation

Even if history and social science sometimes explain structures or events which are irreducible to individual actions and events, it is undeniable that they sometimes at least do engage in explanation of the actions of specific individual persons. What kind of explanations are these action explanations?

There are two 'grand' traditions in the philosophy of history and social science that supply an answer to this question, one represented (*inter alia*) by John Stuart Mill and Carl Hempel, the other by Wilhelm Dilthey, William Dray, and Peter Winch. The traditions' conventional names conjure up too many other views, which bear no logical connection to their views on action (see Action), so one can call the first tradition simply 'the causal tradition', and the second, 'the non-causal tradition'. By the former, the thesis intended is that every human action is to be explained by its causes, which are to be found, at least in part, in the mental states of the agents; by the latter, one understands any thesis which gives some other account of action explanation. This latter tradition is something of a catch-all. It is sometimes called 'hermeneutic', because many non-causal theorists liken explaining human action to interpreting a text. It is also sometimes called the 'interpretative' view, since it tends to stress the way in which action needs to be interpreted in order to be rendered intelligible. A specific version of this view is called 'contextualism', since it focuses on the societal norms or rules in the social setting in which the agent acts,

with which we can interpret what he is doing as an action of one kind or another. For example, it is only because there exist certain religious rules, or rules about banking, that we can say of someone that what they are doing is celebrating mass or cashing a cheque.

There are at least three distinctions required, in order to make any progress in deciding between these views. The first is between a mere bodily movement and an action. A person's hand may move, or their finger bend, without them moving their hand or bending their finger. The former may not be actions, things they do, and their explanation may not include any of their mental states at all. They might, for example, be merely reflex 'actions'. Even if mental states are included in their explanation, it may still not be actions which are being explained. Fear may cause my pupils to dilate; St Thomas' example is that of lust causing a man's erection. Neither the dilation nor the erection are the actions of a person. These latter explanations are not within the scope of social science at all, but rather of biology.

The second and third distinctions both have to do with what is sometimes called 'explaining under a description'. Explanation is, as claimed above, non-extensional; how something is described matters to whether it explains or is explained by something else.

The second distinction is between the intentional and non-intentional descriptions of action. On one plausible account of action individuation, what we explain are not just actions, but actions under some specific description or other. For example, every action will have an indefinitely large number of unintended effects, and the action can always be redescribed as the action having such-and-such an unintended effect. Theories of action explanation focus on explanation of actions under their intended descriptions. Once an explanation of action under its intended description is found, an explanation under its unintended description will be quickly forthcoming: it is just that the action as intended happened to have such a causal consequence.

The third distinction is between the action under its basic and its non-basic descriptions. (Non-basic descriptions can be either intentional or unintentional.) A man flexes his finger, pulls the trigger, shoots the gun, kills Smith, and executes a criminal. He executes the criminal by killing Smith; he kills Smith by shooting the gun; he shoots the gun by pulling the trigger; he pulls the trigger by flexing his finger. 'Flexing his finger' is the basic description of the action, if there is no other description of the action such that he flexes his finger by doing the action described under that other description.

The causal tradition, deriving from Hobbes and Hume, is the view that every action is to be explained, *inter alia*, by the mental states that cause it. The mental states may be only part of the full explanation of the action, but they are a necessary part. Metaphysically, this view about explanation is consistent with, but does not entail, physicalism: such rationalizing mental states may or may not be physical states (see Materialism §1).

On the causal view, the item that is explained is an action only if the mental causes not only cause the movement but also rationalize it, make it the rational thing to have done in the circumstances, relative to the agent's beliefs and desires. Rationalizing and causing are two distinct relations and provide two distinct requirements. On the causal view, the mental states that explain an action must both cause and rationalize it. On the traditional causal view, such a mental state must include a belief and a desire. This is the old Humean view of rationality as instrumental: an action is rationalized if it can be 'justified' in terms of what the agent believes they need to do in order to satisfy their desires (see Hume, D. §3). On other accounts, such a mental state may be an intention rather than a belief and desire pair. An agent's doing A can be explained by their intention to do so, if such an intention both causes their doing that action and makes it rational for them to do it (Ayer 1967; Gardiner 1978).

Since the non-causal tradition is, as has been described, something of a ragbag, let us select one strand from many possible ones. Often, non-causal theories focus on explanation of action occurring in an alien culture. How can we understand, or make sense of, what someone from such a society is doing, if it makes no obvious sense in our own terms? The answer surely is that we must know the rules and norms of that society, in order to interpret what they are doing, to endow their movements with sense and significance. We must, as it were, go native. If we are seeking to explain action under its intentional description, then we must know what the intentions of the agent are, and one can know the latter only by knowing the rules and norms of the society which give rise to, or at least make possible, the having of such intentions. Movements of one's body can only be given meaning in terms of such rules; one's hand movements can only constitute a writing of a cheque, in the light of regulations about banking.

And what is true of alien societies is true of our own; our own behaviour only acquires social meaning or significance, in the light of our own social rules and conventions (Dilthey 1976; Winch 1958).

The non-causal view seems plausible as a view about at least some non-basic action. So-called social actions are just actions under non-basic social descriptions. A person moves their hand (if they are acting, then it is not just that their hand moves). To understand their moving of their hand as their writing a cheque, one must cite, *inter alia*, the rules and regulations of banking. In order to do that, the explainer will need to cite, or suppose, the existence of a social institution, that is, banking. Social rules and institutions may only be relevant to action explanation at the level of non-basic descriptions of action. The non-causal view explains the possibility of giving social descriptions of action.

The causal theory, on the other hand, seems to be doing at least two different things. First, it accounts for the differences between mere bodily movement and intentional action. Second, given that the action performed was of a certain kind, it offers an explanation of why the agent acted in that way.

4 In virtue of what is an action of interest to the historian or social scientist?

Some actions will be of no possible interest to the social scientist. On 17 July 1995, I went into my garden and picked some sour cherries. Such an action certainly has an explanation: I desired some sour cherries and I believed that by picking some fruit from a specific tree in my garden, I could obtain some. Given the circumstances, it is hard to see how an explanation of that action could be of any interest to the social scientist. Other cherry pickings might be, in different circumstances. But that picking was not. Why should that be so?

It is important to see that this question is not the same as another question, with which it is sometimes conflated: which actions are social actions? Whether or not an action is a social action (as so described) can be determined a priori. Some action descriptions are social: writing a cheque, celebrating a marriage, witnessing a will. That is, it is logically or conceptually necessary for there to be a society, and certain rules or conventions, in order for persons to engage in actions of that description. Some philosophical disputes can be understood as disputes about whether certain action descriptions are social, for example, ‘speaking a language’.

We need here another distinction: between action types and singular action tokens. An action type is a universal, which has many possible instances. Stalin moving his left thumb is an action type, since there could have been many such movings by him. An action token is an individual instance; it is an unrepeatable, dated item. Stalin moving his left thumb at a specific time is an action token.

Do social scientists explain action types or action tokens? One might suppose that to explain why Hitler annexed Austria in 1938 is to explain an action token; to explain why investors withdrew their bank deposits in 1929 is to explain why some action type had many instances. However, another way to consider this might be to say that social scientists only ever explain types, but that sometimes they explain why an action type had at least one instance (why there was at least one annexing of Austria by Hitler in 1938), and at other times why an action type had many instances (why there were many withdrawals of bank deposits by investors in 1929).

Now, not all social action, whether type or token, is of any interest to the social scientist. Not only will they be uninterested in explaining why I picked those sour cherries, they might equally be uninterested in explaining the social action token, my writing a cheque at a particular time for sour cherries. Further, they may have no interest in explaining why the social action type, writing a cheque for sour cherries, has many instances. They might be interested in explaining the origins and development of the system of capitalist agriculture, but that is not the same as an explanation of any action, type or token, one or many, made possible by that institution (unless, perhaps, methodological individualism is true).

So only some social actions and only some non-social actions (whether types or tokens) are of interest to the social scientist. Which ones are they? Those, presumably, which are thought to have consequences of social significance. But what kinds of consequences are those? I do not think that one can ultimately make sense of the idea of ‘consequences of social significance’ except via an account of ‘consequences for the wellbeing or ill-being of a large number of individuals’. That latter concept has itself nothing irreducibly social about it. This is not to deny that the social scientist will also be interested in giving explanations of historical and social events and structures, nor is it to argue that such explanations must be in principle reducible to explanations of individual actions of

significance. It is merely to say that there is no way in which to circumscribe which sorts of actions the social scientist will want to explain other than by using an ultimately ethical notion of the wellbeing of individuals. It is, of course, consistent with this that the wellbeing of an individual will itself include the satisfaction of their need for relationships with other persons.

See also: Functional explanation; Holism and individualism in history and social science

DAVID-HILLEL RUBEN

References and further reading

Ayer, A.J. (1967) ‘Man as a subject for science’, in P. Laslett and W.G. Runciman (eds), *Philosophy, Politics and Society*, Oxford: Blackwell, 3rd series.(A good discussion of the compatibility of the interpretative and causal approaches to the explanation of human action.)

Dilthey, W. (1976) *Selected Writings*, ed. by H.P. Rickman, Cambridge: Cambridge University Press.(A major contribution to the hermeneutic tradition of action explanation.)

Elster, J. (1983) *Explaining Technical Change*, Cambridge: Cambridge University Press.(A presentation of the causal feedback loop account of functional explanation.)

Gardiner, P. (ed.) (1978) *The Philosophy of History*, Oxford: Oxford University Press.(A useful collection of the literature in the philosophy of history on the explanation of action. Essays by William Dray, Carl Hempel, and others are included.)

Hempel, C. (1942) ‘The function of general laws in history’, repr. in *Aspects of Scientific Explanation*, New York: Free Press, 1965.(A major contribution to the causal tradition of action explanation, although Hempel himself eschews explicit use of the concept of causation.)

Hempel, C. (1959) ‘The logic of functional analysis’, repr. in *Aspects of Scientific Explanation*, New York: Free Press, 1965.(An alternative to the causal feedback loop account of functional explanation.)

Mill. J.S. (1843) *System of Logic: Ratiocinative and Inductive*, London: Longman, 1970.(A classical statement of the causal position.)

Ruben, D.-H. (1985) *The Metaphysics of the Social World*, London: Routledge.(A discussion of the varieties of methodological individualism and an assessment of their plausibility.)

Ruben, D.-H. (1990) ‘Singular explanation and the social sciences’, in P. French, T. Uehling, Jr., and H. Wettstein (eds), *Midwest Studies in Philosophy: The Philosophy of the Human Sciences*, Notre Dame, IN: University of Notre Dame Press, vol XV.(The volume as a whole is well worth browsing.)

Ryan, A. (1973) *The Philosophy of Social Explanation*, Oxford: Oxford University Press.(A somewhat dated but still useful collection of articles on a range of issues about explanation in the social sciences.)

Winch, P. (1958) *The Idea of a Social Science*, London: Routledge.(A well-known statement of the non-causal view of action explanation.)

Wright, G.H. von (1971) *Explanation and Understanding*, London: Routledge & Kegan Paul. (An unrivalled, insightful discussion of the contrasting merits of the two great traditions of explanation.)

Zeleny, J. (1980) *The Logic of Marx*, Oxford: Blackwell.(A defence and elaboration of the method of structural explanation.)

Fa

Fa *is a technical term in a variety of Chinese philosophical traditions. As a noun it means 'standard' or 'norm', and, by extension, 'law'. As a verb it means 'to be in accord with'. The disagreements among the various indigenous schools tended to focus on the origin and nature of* fa, *whether it is an order or pattern within things or an external pattern to which things conform. When Buddhism entered China, translators often used* fa *for rendering the Sanskrit term* dharma*. Depending on context, the Buddhist term could refer to true teachings, moral duty, phenomenal reality or rule.*

Like other philosophical terms, *fa* originally had an ordinary use in the Chinese language. As such, *fa* appears as a noun or a verb, suggesting either a standard to be met or the action of meeting that standard. Chinese philosophers differed over the external or internal nature of *fa*. Some maintained it is internal to things, part of their natural expression. Others held that *fa* is an external standard to which things normatively conform. Daoist philosophy tended to favour the internalist view; for example, Laozi used *fa* as a verb in maintaining that the *dao* itself 'accorded with' things as they naturally are (see Daodejing). Most other classical Chinese philosophers, however, used the term primarily as a noun to express a norm or standard.

Because the social, moral and political dimensions of human experience were central to Chinese systems of thought, philosophers often analysed norms relevant to the harmony of the state. Generally, the Chinese thinkers assumed the ruler to be responsible for setting and enforcing these standards. In such contexts, the term *fa* assumed a meaning close to that of the English term 'law' or, more specifically, 'penal law' (see Law, philosophy of; Law and ritual in Chinese philosophy). The Confucians and Legalists were especially inclined to view *fa* in such a manner.

For Confucius, social order arises from the practice of 'ritual' or 'propriety' (*li*) rather than law. Social harmony derives from each person's following the appropriate forms of basic human relations. If the conditions are right and the ruler is a person of sufficient charismatic virtue, Confucius maintained, then the society will organize itself harmoniously without legal sanctions. However, the Confucians also recognized that such ideal conditions do not necessarily prevail. When harmony breaks down, the ruler must enforce certain minimal standards of social behaviour through penal law (*fa*). In short, *fa* compensates for the inadequate practice of ritualized human interaction (see Confucian Philosophy, Chinese).

The Legalists, as represented by Han Feizi for example, characterized the role of *fa* differently. For them, *fa* is not a last resort, but instead the very foundation of the harmonious state. According to the Legalists, the state evolves out of the authority of the ruler and that ruler's capacity to establish a system of laws enforced by sanctions. For the Legalists, therefore, *fa* is neither part of the natural order (as the Daoists generally maintained) nor a back-up system for the virtuous charisma of the ruler (as the Confucians held).

The Legalist philosophy was perhaps the first systematic attempt in China to deal with the state independently of discourse about virtue or morality (see Legalist philosophy, Chinese). It stressed rule by law rather than rule by virtuous people. The Legalists argued for a universal legal system applying even to the ruler who originally established the laws. In a pragmatic vein, the Legalists maintained that government office should be assigned on the grounds of the candidates' practical talents and skills rather than education or virtue. This was, however, a strict totalitarian system of penal law in which offenders were not re-educated but were immediately (and by modern standards severely) punished. Although the strict Legalist philosophy was generally discounted as too harsh and inflexible, its ideal of *fa* endured as a central philosophical category within Chinese political theories and practices.

Buddhism brought a new layer of meanings to the term. When the Chinese began to translate Indian Buddhist texts two thousand years ago, they used the Chinese word *fa* to render the Sanskrit term *dharma*. The Buddhist term itself has several facets of meaning, including 'real (that is, non-delusory) phenomenon', 'true teachings', 'moral duty' and 'rule'. The match between the two terms was not perfect. At the time of its introduction, Buddhism was understood by the Chinese to be a philosophy akin to Daoism (see Buddhist philosophy, Chinese). It is helpful, therefore, to consider how the Daoist notion of *fa* relates to the Buddhist *dharma*.

Both Daoism and Buddhism aim to understand phenomena for what they are in themselves, without their being coloured by human or personal ideas of 'usefulness'. There are vast differences between Daoist and Buddhist practice, however. In Daoism, the focus is on observing the natural pattern or way (*dao*) of things, of seeing human beings as part of a larger natural process (see Dao; De). In Buddhism, on the other hand, the focus is on introspection and the eradication of the ego-centred projection of delusory contexts that lead one to interpret phenomena as 'mine', as 'permanent' and as 'desirable.' The Daoists and Buddhists share, however, the idea that if one sets right one's relation with the world, social morality will follow spontaneously. Specifically, for the Chinese Buddhists, if one eradicates the ego, one can experience the phenomena (*fa*) as they are in their own contexts without distortion. This experience will confirm the true teachings (*fa*) of Buddhism. Without ego, one will enact a spontaneous morality (*fa*). To achieve the proper attitude for the introspection leading to the eradication of ego, the monastic community sets social rules or laws (*fa*) that nurture detachment. When viewed in this light, we see how the different meanings of *dharma* in Buddhism are related and how the Buddhist term *dharma* could be construed as being consistent with aspects of the term fa, especially in its Daoist sense.

See also: Buddhist philosophy, Chinese; Legalist philosophy, Chinese; Confucian philosophy, Chinese; Daoist philosophy; Duty and virtue, Indian conceptions of; Law and ritual in Chinese philosophy; Law, philosophy of

THOMAS P. KASULIS

References and further reading

Ames, R.T. (1983) *The Art of Rulership: A Study in Ancient Chinese Political Thought*, Honolulu, HI: University of Hawaii Press.(Chapter 4, pages 108-41, is a thorough discussion of pre-Buddhist uses of the term *fa* in Chinese philosophy.)

Chan Wing-tsit (1963) *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press.(A useful, inclusive definition of *fa* is found in the appendix, page 786; this index is generally useful in finding key uses of the term in various Chinese philosophies by following references to either 'law' or '*dharma*'.)

Hansen, C. (1994) '*Fa* (standards: laws) and meaning changes in Chinese philosophy', *Philosophy East and West* 44 (3): 433-88.(Discussion of the meaning of *fa* within the context of a broader discussion of the philosophical issue of what it means for a word to change its meaning.)

Wu, J.C.H. (1967) 'The Status of the Individual in the Political and Legal Traditions of Old and New China', in C.A. Moore (ed.) *The Chinese Mind: Essentials of Chinese Philosophy and Culture*, Honolulu, HI: University of Hawaii Press, 213-37. (Helpful overview of key philosophical motifs in traditional Chinese conceptions of law.)

Fackenheim, Emil Ludwig (1916-)

Fackenheim is best known for his account of authentic philosophical and Jewish responses to the Nazi Holocaust. Fackenheim's thought, indebted to German philosophy, always had a practical, existential purpose. He aimed, consistently, to show how philosophical and theological thought could pay attention to lived experience in order to gain the understanding it sought. In this way, he exposed the varied ways in which thought and life, ideas and history, are interdependent. Philosophical and religious thought are grounded in historical experience and give direction to that experience; life is the ground of reflection and derives its meaning from it.

Fackenheim was born in Halle, Germany. Although aware of the dangers of Nazi persecution, he left Halle for Berlin, where in 1935 he began rabbinic studies at the Hochschule für die Wissenschaft des Judentums. After *Kristalnacht*, Fackenheim fled to Scotland and finally to Toronto, where he completed his doctorate with a dissertation on the doctrine of substance in the thought of the medieval Islamic authors known as the Sincere Brethren of Basra (see Ikhwan Al-Safa'). Four years later, having served as a congregational rabbi in Hamilton, Ontario, he began to teach philosophy at Toronto. He spent the post-war years studying German Idealism. This period culminated with the publication of a book on Hegel's treatment of religion. In this early period he defended faith and revelation. Understanding revelation and the faith it engendered as a divine-human encounter, Fackenheim took that encounter to be the ground of genuine religious life. After 1966-7, Fackenheim turned his attention to the memory of the Nazi death camps. He came to see the radical evil that erupted there as a threat both to reason and to faith. His defence of reason and faith became increasingly historical as he interpreted the continuity of religious faith as part of an authentic philosophical and religious response to the Nazi atrocities. Fackenheim remained in Toronto until his retirement and subsequent emigration to Israel in 1982.

Fackenheim's work and thought are united by several interrelated themes. The first of these is the existential relevance of philosophical and theological thought. This was a central attraction of existential philosophy: thinking is not detached from life; it arises in a concrete situation and is a committed response to that situation. Fackenheim shunned antiquarianism as he immersed himself in the great philosophers of the past, from Plato and Aristotle and Plotinus to Kant, Fichte, Schelling, Hegel and Heidegger. His aim was to grasp and grapple with the central issues of the present.

One of those issues is the conflict between reason and revelation. Stimulated by the thinking of Soren Kierkegaard, Martin Buber, and later Franz Rosenzweig, Fackenheim struggled with the challenges made by modernity and modern philosophy to the very possibility of faith and revelation. First, there is modern secular philosophy and thought, the challenge of figures like Feuerbach and Freud. Then there is the philosophical overcoming of revelation in Schelling and ultimately in Hegel, where a genuine divine-human encounter is transformed into an expressive monism and religion is superseded by philosophy. Finally, there is the existential, historical challenge of an evil unprecedented and radical, the evil of the Holocaust and the Nazi death camps. Throughout his career, in the face of each wave of attack, Fackenheim held firm to the banner of revelation and defended it philosophically - early on as the possibility of a fundamental commitment that is no more an act of sheer faith than are the secularism or naturalism that denied it, later as a possible and even necessary response to the deepest vector of Nazi evil and hence as an expression of hope.

The final theme that pervades Fackenheim's career is a life-long reflection on the complex relationships between thought and life. It is this issue that leads him to Hegel. If any modern philosopher succeeded in articulating how reason and thought, while doing justice to all of history and life, still tame them, it was Hegel. If Hegel failed, perhaps the project cannot succeed. Perhaps, ultimately, life is intractable to thought; perhaps thought must unavoidably go to school with life.

As Fackenheim, in 1966 and after, makes the important, even heroic decision to confront the Nazi Holocaust in all its darkness and its horror, the interdependence of thought and life emerges as central to his thinking (see Holocaust, the). In the essays of 1967 and 1968 Fackenheim's analysis proceeds this way. The thinker, whether philosopher, theologian, historian, or social theorist, confronts Naziism and the death camps with the desire to understand. This desire is quite natural, an expression of aspiration to transcendence that is characteristic of all thought. But probity requires a sense of unease and even humility. The cognitive capacity is paralyzed by the

enormity of the event; what can thought do? It can recognize that meaning and purpose are one thing, response another. The elicitation of meaning here may be impossible. But the response is unavoidable. Hence the thinker, seeking guidance, comes to the realization that real, living Jews have all along been responding in their own ways. In order to understand how one ought to respond and why, the thinker turns to actual response and engages in a 'transcendental' analysis of what makes such response necessary and possible. It is this analysis that leads to Fackenheim's momentous and famous conclusion - that for Jews, whether religious or non-religious, the Divine Commanding Voice was present at Auschwitz and that the imperative which it 'spoke' (and still speaks) was a commandment to resist Nazi purposes, not to give Hitler posthumous victories. This command becomes a means of identifying the ground and content of the necessity for contemporary Jewish existence.

But if authentic post-Holocaust existence - here, Jewish existence - is necessary, is it possible? In the essays of 1967-8, Fackenheim had relied on the Kantian conviction that ought implies can. Paralleling this is the neo-orthodox faith that a God who commands at the same time graciously gives the freedom to respond. By the late 1970s and early 1980s, if not before, these strategies had come to seem inadequate. Indeed, they seemed to belittle the heroism of those who had resisted, as well as the grim courage of those who had not. In the central sections of *To Mend the World* (1982) Fackenheim attempts to rectify the fault. Thought, he acknowledges, tries to comprehend the evil, the pain, the horror, the agency, but as it moves from victim to victim and from agent to agent it acknowledges its own inadequacy. As it tries to grasp why the perpetrators of genocide committed their many crimes, thought fails and fails again until, in a shocking resolution, it grasps the death camps as a whole that is greater than its parts, a whole of horror, a world of evil, and thought is surprised. But the surprise is not neutral; it is recognition, amazement, and horrified recoil, all at once, and while it contains a kind of cognitive grasp, it registers as well a compulsion to oppose. Fackenheim finds this conjunction, this attitude, realized in the actions of some who lived that recognition and that opposition, victims who at the moment of truth resisted and tell us how their resistance was born out of a recognition of the evil, its intent, and the sense of obligation that they felt. In this way, an ontological possibility is grounded in an ontic reality, and in this way too theory and praxis, realized in the event itself as a synthetic whole, lead to action. The theory is incomplete without the praxis, the praxis impossible without the theory.

See also: Holocaust, the; Jewish philosophy, contemporary

MICHAEL L. MORGAN

List of works

Fackenheim, E.L. (1961) *Metaphysics and Historicity*, Milwaukee, WI: Marquette University Press.(Aquinas Lecture on the universality of philosophical thought and a defence of existential, situated selfhood against the charge of historicism.)

Fackenheim, E.L. (1967) *The Religious Dimension in Hegel's Thought*, Bloomington, IN: Indiana University Press.(Systematic examination of the role of religion in Hegel's thought, from the youthful works and the Phenomenology of Spirit through the *Encyclopedia of the Philosophical Sciences*.)

Fackenheim, E.L. (1970) *God's Presence in History*, New York: New York University Press.(Deems Lectures on Jewish theology in midrashic literature, the challenges of modernity and the Enlightenment, and exposure to the Holocaust.)

Fackenheim, E.L. (1973) *Encounters Between Judaism and Modern Philosophy*, New York: Basic Books. (Philosophical confrontations between Judaism and modern empirical and ananlytical philosophy, Kantian thought, Hegel and existential philosophy.)

Fackenheim, E.L. (1982) *To Mend the World*, New York: Schocken Books; 3rd edn, 1994.(Systematic recovery of Jewish thought through encounters with Spinoza, Rosenzweig, Hegel and Heidegger; justification for the necessity and possibility of post-Holocaust response by Judaism, Christianity, philosophers and others.)

Fackenheim, E.L. (1996) *The God Within: Kant, Schelling and Historicity, Essays by E.L. Fackenheim*, ed. J. Burbidge, Toronto: University of Toronto Press.(Collection of Fackenheim's essays on Kant, Schelling, Hegel and German philosophy.)

References and further reading

Greenspan, L. and Nicholson, G. (eds) (1992) *Fackenheim: German Philosophy & Jewish Thought*, Toronto: University of Toronto Press.(Essays on philosophical and theological themes in Fackenheim's work; includes a complete bibliography.)

Morgan, M.L. (ed.) (1987) *The Jewish Thought of Emil Fackenheim*, Detroit: Wayne State University Press. (Comprehensive anthology with introductions and complete bibliography.)

Morgan, M.L. (1996) *Jewish Philosophers and Jewish Philosophy: Essays by Emil Fackenheim*, Bloomington, IN: Indiana University Press.(Essays on Jewish philosophers and post-Holocaust Jewish philosophy.)

Morgan, M.L. (1996) 'The Central Problem of Fackenheim's *To Mend the World*', *The Journal of Jewish Thought and Philosophy* 5: 297-312.(Analysis of Fackenheim's argument for the ground of the possibility of post-Holocaust response.)

Facts

The existence and nature of facts is disputed. In ordinary language we often speak of facts ('that's a fact') but it is hard to take such talk seriously since it can be paraphrased away. It is better to argue for the existence of facts on the basis of three connected theoretical roles for facts. First, facts as the referents of true sentences: 'the cat sat on the mat', if true, refers to the fact that the cat sat on the mat. Second, facts as the truth-makers of true sentences: the fact that the cat sat on the mat is what makes 'the cat sat on the mat' true. Third, facts as causal relata, related in such sentences as 'Caesar died because Brutus stabbed him'. The so-called 'slingshot' argument aims to show that these roles are misconceived.

1 Roles for facts

We often speak of facts in ordinary discourse. For example, we append 'the fact' to an indirect sentential clause in sentences such as 'Smith was surprised at the fact that Brown's wife left him'. It is controversial whether this talk of facts is to be taken at face value since the apparent commitment to facts disappears under analysis, becoming 'Smith heard with surprise that Brown's wife left him. Brown's wife left him', in which no mention is made of any fact. Other ordinary talk of facts is also subject to paraphrase: 'It is a fact that *p*' is equivalent to 'It is true that *p*', 'That is a fact' is equivalent to 'That is true'. This last style of paraphrase has been taken in two ways: either as showing that facts do not exist, or that they exist but can be identified with the true bearers of truth-values. Following the second course, the bearers of truth-values are commonly taken to be propositions, that is, the meanings of sentences, and facts are then identified with true propositions.

Rather than squeeze ontological juice from ordinary language, we can argue for the existence of facts on the basis of theoretical demands. First, there is a role for facts within semantic theory, as the worldly referents of true sentences. So a true sentence such as 'The cat sat on the mat' refers to the fact that the cat sat on the mat, sometimes called the situation or state of affairs of the cat sitting on the mat.

A related second role for facts lies within a correspondence theory of truth (see Truth, correspondence theory of). The unworkable correspondence theory of logical atomism, in which correspondence consists in a structural isomorphism between a true atomic sentence and the fact to which it corresponds, is replaced with a truth-maker principle of the form 'every true sentence has a truth-maker' (see Logical atomism). This principle retains the doctrine that the world is independent of linguistic description and must be a certain way for a given sentence to be true of it. The intuitive idea is that a truth-maker grounds the truth of a sentence, where the grounding can be spelled out in a modal fashion: entity *E* is a truth-maker for the sentence *S* just in case necessarily, if *E* exists, then *S* is true. Facts are introduced as the required truth-makers. Neither the particular cat nor the property of sitting on the mat makes 'The cat sat on the mat' true; each might exist even if the sentence is false. But the fact that the cat sat on the mat, consisting of the cat instantiating the property of sitting on the mat, will do the job. Facts as worldly referents or truth-makers of true sentences are best conceived as composite entities, having particulars and properties or relations as constituents, and as themselves parts of the actual world, unlike true propositions which belong to the abstract and other-worldly realm of sense.

Composite worldly facts may play a third role as causal relata (see Causation §6). One way of reporting singular causation uses the sentential connective 'because' (as in 'Caesar died because Brutus stabbed him'), which is true only if the flanking sentences are true. And if the truth-maker principle is right, these sentences need facts as truth-makers, and so the original sentence can be understood as reporting a causal relation between these facts. This account of causal relata has been criticized by Davidson (1967) using the so-called slingshot argument.

2 The slingshot argument

The aim of the slingshot argument is to show that a causal 'because' would have to be truth-functional: that is, the truth-value of a sentence containing the causal 'because' is preserved if one of the sentences flanking the 'because' is replaced by another which shares its truth-value. But it cannot be truth-functional, so the assumption that there is a causal 'because' must be rejected (see Davidson, D.).

In more detail, suppose we are given a true sentence featuring the purported causal 'because': for example, 'Caesar died because Brutus stabbed Caesar'. The idea of the slingshot is to show that this sentence entails a sentence such

as ‘Caesar died because Caesar had a big nose’, which differs from the first only in that a true sentence is substituted for a true sentence flanking the ‘because’. So we have reached the absurd result that the causal ‘because’ is truth-functional.

The steps of the argument are licensed by two premises: that a necessarily equivalent sentence can be substituted for a sentence flanking the ‘because’ without change of truth-value; and that a co-referential singular term can be substituted for a singular term occurring in one of the flanking sentences without change of truth-value. Here is one way to construct a slingshot argument.

(1) Caesar died because Brutus stabbed Caesar.
(2) Caesar died because { x: x is a natural number and Brutus stabbed Caesar}= { x: x is a natural number}.
(3) Caesar died because { x: x is a natural number and Caesar had a big nose}= { x: x is a natural number}.
(4) Caesar died because Caesar had a big nose.

The move from (1) to (2) is justified by the first premise. The singular term ‘{ x: x is a natural number and Brutus stabbed Caesar}’ names the set which contains every object to which the predicate ‘… is a natural number and Brutus stabbed Caesar’ applies; similarly for ‘{ x: x is a natural number}’. So ‘Brutus stabbed Caesar’ is necessarily equivalent to ‘{ x: x is a natural number and Brutus stabbed Caesar}= { x: x is a natural number}’.

The move from (2) to (3) is justified by the second premise. We replace the singular term ‘{ x: x is a natural number and Brutus stabbed Caesar}’ with the co-referential ‘{ x: x is a natural number and Caesar had a big nose}’; these are guaranteed to be co-referential given that ‘Brutus stabbed Caesar’ and ‘Caesar had a big nose’ have the same truth-value. The move from (3) to (4) is justified by the first premise and the fact that ‘{ x: x is a natural number and Caesar had a big nose}= { x: x is a natural number}’ and ‘Caesar had a big nose’ are necessarily equivalent.

We have thus shown that a causal ‘because’ would have to be truth-functional. But it cannot be, so there is no such ‘because’. The slingshot argument is valid, so at least one of the premises must be rejected if facts are to be causal relata. Indeed the slingshot argument also threatens facts in their roles as referents of true sentences and truth-makers for true sentences. For the argument can be used to establish the truth-functionality of ‘sentence *S* refers to the fact that *P*’ and ‘sentence *S* is made true by the fact that *P*’, by operating on the sentence following ‘the fact that’. This yields the undesirable result that a true sentence refers to and is made true by any fact. The slingshot argument cannot be dismissed as logical trickery, for it exposes a critical lacuna in the theory of composite facts. Any motivated criticism of the argument must rely on an account of the identity conditions of facts, and these remain undecided.

See also: Events; Negative facts in classical Indian philosophy; Ontological commitment

ALEX OLIVER

References and further reading

Armstrong, D.M. (1989) *Universals: An Opinionated Introduction*, Boulder, CO: Westview Press.(An introductory guide to the theory of properties, with facts - or states of affairs as Armstrong calls them - as worldly composites in their role as truth-makers.)

Barwise, J. and Perry, J. (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press. (Advocates of a formal semantic theory, situation semantics, based on the idea that true sentences refer to worldly facts or ‘situations’.)

Davidson, D. (1967) ‘Causal Relations’, in *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, 149-62. (Contains the slingshot argument against facts as causal relata.)

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.(Chapters 2 and 3 contain further uses of the slingshot argument, undermining the theoretical roles of worldly facts by showing that every true sentence refers to the same fact and that a given true sentence corresponds to every fact.)

Geach, P.T. (1963) ‘Aristotle on Conjunctive Propositions’, in *Logic Matters*, Oxford: Blackwell, 1972.(Facts as ‘a dream of our language’. §1.2, 21-3 is especially relevant.)

Mellor, D.H. (1987) ‘The Singularly Affecting Facts of Causation’, in *Matters of Metaphysics*, Cambridge: Cambridge University Press, 1991, 201-24.(A friend of facts as truth-makers and causal relata; criticises Davidson’s slingshot argument by denying its second premise.)

Olson, K.R. (1987) *An Essay on Facts*, CSLI Lecture Notes Number 6, Stanford, CA: Center for the Study of

Language and Information, Stanford University.(A detailed appraisal of rival accounts of facts and a history of the slingshot argument.)

Russell, B.A.W. (1918) 'The Philosophy of Logical Atomism', in *Logic and Knowledge*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 177-281.(Facts as constituents of the world corresponding to true *and* false sentences within the framework of logical atomism.)

Fact/value distinction

According to proponents of the fact/value distinction, no states of affairs in the world can be said to be values, and evaluative judgments are best understood not to be pure statements of fact. The distinction was important in twentieth-century ethics, and debate continues about the metaphysical status of value, the epistemology of value, and the best characterization of value-judgments.

A fact is an actual state of affairs (see Facts). A value is either something good (pleasure, for example), or a belief that something is good (to say that pleasure is one of my values is to say that I believe pleasure to be good) (see Values). The fact/value distinction was of great importance in twentieth-century moral philosophy, a distinction being drawn between actual states of affairs and values in both senses (each sense not always being clearly distinguished from the other).

On one version of the fact/value distinction, there are no values 'in the world'. John Mackie (1977), for example, argued that such items were too peculiar to be fitted into any decent metaphysics or epistemology, and that the nonexistence of values was the best way to explain disagreements in evaluations. According to existentialist ethics, the non-factuality of value leaves us in a position of radical freedom to choose (see Existentialist ethics §§1-2).

The distinction understood as concerning evaluations suggests evaluations are not pure attempts to state facts. One famous and influential version of this view is that of Hume (1739/40), who claimed that 'ought' conclusions do not follow logically from 'is' statements (see Hume. D. §4; Logic of ethical discourse §§1-3). So if you claim successfully that something ought to be done (this may be one of your values) on the basis of an argument apparently referring only to facts, it must be the case that one of your 'factual' statements involves a covert 'ought'.

This version of the fact/value distinction, allied to a narrow conception of what can count as a factual statement, has been of great importance. If facts are, for example, restricted to purely neutral descriptions, such as those found in natural science, moral judgments may be seen as other than fact-stating. (It has been argued by some that science itself is an evaluative enterprise, so that the fact/value distinction is spurious - see Analytic ethics §2.) Words such as 'good' or 'right' can then be claimed to have special, non-descriptive roles. According to emotivism, to claim that X is good is to express a pro-attitude towards it, and perhaps to encourage others to adopt that attitude; according to prescriptivism, the claim is to be understood as imperatival (see Emotivism; Prescriptivism). On such views, certain words, such as 'courageous' for example, may have some factual content; but this can always be distinguished, at least conceptually, from the evaluative content.

Those who have claimed that values are part of the world and that evaluations are fact-stating include advocates of moral realism (see Moral realism). This comes in at least two varieties. Ethical naturalism states that values are natural facts, 'natural' meaning that such facts are to be identified with, or seen as constituted by, facts open to investigation by natural science. Ethical non-naturalism sees values as facts *sui generis*, any attempt to identify them with natural facts being to commit what G.E. Moore (1903) called the 'naturalistic fallacy' (see Naturalism in ethics; Moore, G.E. §1).

See also: Logic of ethical discourse

ROGER CRISP

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz; 2nd edn, 1946, ch. 6.(An early and influential statement of emotivism.)

Foot, P.F. (1978) *Virtues and Vices*, Oxford: Blackwell, esp. ch. 8.(Collection of essays by an influential ethical naturalist.)

Hare, R.M. (1952) *The Language of Morals*, Oxford: Oxford University Press.(Key exposition of prescriptivism.)

Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978, book 3, part I, section 1. (Contains 'Hume's Law' concerning 'is' and 'ought'.)

Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin. (Claims that values are not

facts, but that moral language attempts to state facts.)
Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press. (Critique of naturalism and defence of non-naturalism.)

Faith

Faith became a topic of discussion in the Western philosophical tradition on account of its prominence in the New Testament, where the having or taking up of faith is often urged by writers. The New Testament itself echoes both Hellenistic concepts of faith and older biblical traditions, specifically that of Abraham in the Book of Genesis.

The subsequent attention of philosophers has been focused primarily on three topics: the nature of faith, the connection between God's goodness and human responsibility, and the relation of faith to reason. Discussions on the nature of faith, from Aquinas to Tillich, have tried to examine the subject in terms of whether it is a particular form of knowledge, virtue, trust and so on. Regarding divine goodness, the argument has primarily focused on the relationship between faith and free will, and whether lack of faith is the responsibility of the individual or of God. Concerning the relation between faith and reason, there are two quite separate issues: the relation of faith to theorizing, and the rationality of faith. Aquinas in particular argued that faith is a necessary prerequisite for reasoning and intellectual activity, while later, John Locke explored the relationship between faith, reason and rationality, and concluded that faith can be reached through reason. This latter viewpoint was later heavily criticized by Wittgenstein and his followers.

1 The New Testament background

In the course of everyday life we speak of persons as having faith in one thing and another - in themselves, in other persons, in causes, in the ability of themselves or others to carry out some project, and so forth. One can imagine some philosopher undertaking, in disinterested fashion, to uncover the nature and role in human life of this phenomenon of having faith; and others then taking up the topic in response to his discussion. There have in fact been a few such discussions. But that is not how faith became a prominent topic of discussion in Western philosophy. It became that because of its prominence in the New Testament; more precisely, because of the prominence the New Testament gives to faith in its description of the mode of life that it advocates. The context is, of course, that the New Testament belongs to the canon of the Christian Church and that Christianity has been powerfully formative of philosophy in the West.

Various writers of the New Testament record Jesus as having urged his hearers to take up the stance of faith. Likewise, those and other New Testament writers urge their own readers to take up this stance. Indeed, they urge their readers to take it up towards Jesus himself. It is clear that when the writers urged this on their readers, and when Jesus urged it on his hearers, they and he saw themselves as doing so on behalf of God.

Sometimes when the New Testament writers represent God as requiring faith of us, they are thinking of faith as one among other things that God requires of us. Faith is then one among other 'virtues'. Paul, in the famous thirteenth chapter of his first letter to the Christians in Corinth, urges on his readers hope and charity along with faith; he concludes that of the three, charity is greatest - on the ground that whereas in the New Age there will no longer be any need for faith and hope, the relevance of charity will abide forever. In other passages, however, the New Testament writers appear to use 'faith' to refer to that total stance towards God to which, on behalf of God, they call us. Then faith in things not seen, hope and charity are three of the requisite *manifestations* of faith. As Rudolf Bultmann remarks, 'In Primitive Christianity, *pistis* became the leading term for the relation of man to God' (1968: 205).

Bultmann observes that in both classical and Hellenistic Greek, the root meaning of *pistis* and its grammatical variants is 'trust' (reliance, belief in, confidence). The objects of *pisteuein* are characteristically such things as contracts and oaths, laws, armaments and persons. Likewise the words of a person can be the object of trust, in which case the sense of *pisteuein* shades towards that of our 'believe'. And sometimes, not surprisingly, *pisteuein* has the nuance of 'to obey'.

In the New Testament as well, the root meaning of the term, when used for that comprehensive stance towards God to which God calls us, is 'trust'. That is especially clear in the famous discourse on faith in the eleventh chapter of Hebrews. The chapter opens with a crisp definition of faith: 'faith is the *hypostasis* of things hoped for, the *elenchos* of things not seen'. In older translations, *hypostasis* and *elenchos* were translated, literally and straightforwardly, as 'substance' and 'evidence' respectively. The authors of the Revised Standard Version, in

recognition of the fact that both words are here being used eccentrically, translate the passage as 'the assurance of things hoped for, the conviction of things not seen'. After giving his definition of faith, and indicating its importance by adding that 'by it the men of old received divine approval', the writer cites examples. His first example is a case of 'conviction of things not seen', not of 'assurance of things hoped for': 'By faith we understand that the world was created by the word of God, so that what is seen was made out of things which do not appear'. All the other examples are cases of both. And in each case, the ground of the assured hope is trust in God's promise to save and bless. It was on account of hope thus grounded that men and women of old 'received divine approval'.

Not only has the Western intellectual tradition been influenced by this general picture of faith and its proper role in human life, but some of the specifics of the New Testament discussion have been influential as well. That is especially true for the discussion of faith in Paul's letter to the Roman Christians. The topic in hand is how to understand theologically the relation of the Christian Church, with its mixture of Jews and Gentiles, to the people of Israel, given that in Genesis it says that God's covenant was with Abraham and all his descendants. The issue was of intense existential concern to Paul, since he was both ethnically Jewish and a Christian. A corollary of the theological issue was the practical issue of whether Christians should be required to follow the ritual practices of Judaism.

The first move in Paul's solution is to refer to Genesis 15: 3, where it says, using Paul's near-paraphrase, that 'Abraham believed God, and it was reckoned to him as righteousness' (Romans 4: 3). From this, along with certain ancillary considerations, Paul draws two conclusions. First, that the divine promise (covenant) is not with those who are descendants of Abraham 'after the flesh' but with those who are descendants of Abraham 'according to the Spirit' - that is to say, not with those who happen to be ethnically Jewish but with those who share Abraham's faith in God and God's promises. Though Paul never offers a definition of faith in the manner of the writer of the book of Hebrews, it is clear that for him, too, trusting in the promises of God is the core, if not the whole, of what he has in mind. The second conclusion Paul draws is that what makes persons pleasing to God is not their performance of some array of ethical and ritual 'works', but simply faith. In the language Paul uses, we are 'justified by faith'. To suppose that one's performance of certain ethical and ritual 'works' would make one just before God would be to suppose that one could achieve that status on one's own — that one could earn it. But we cannot earn it, since we all inevitably in one way or another fail to perform the full panoply of 'works' prescribed in the manner prescribed. Yet Abraham was declared to be righteous before God - that is, just (justified). It follows that having the status of being just before God is a gift on God's part, a status God accords one - on the ground of one's faith. If we are not to fall into contradictions, we must also then understand faith itself as a gift of God, rather than as a 'work' which earns the status of being righteous (justified) before God.

2 The nature of faith

The philosophical discussions of faith almost all fall under one or the other of three headings. In the first place, there is a long history of discussions on the nature of faith. Of what genus is faith a species? Is it a species of believing propositions on say-so? Is it a species of loyalty to some person or cause? Is it a species of trusting someone? Is it a species of believing what someone has promised? Is it a species of 'concern'? Is it a virtue of a certain sort? Is it a species of knowledge? If so, of what sort: *agnitio* (recognition, acknowledgement), *cognitio*, or what? An exceptionally fine example of a philosophical discussion concerning the nature of faith is that by Thomas Aquinas (§14) in *Summa theologiae* IIaIIae, qq.1-7.

The disputes which have arisen among those who propound different analyses of faith have no doubt often had a large component of purely verbal disagreement; if one looks beneath the surface, however, usually one will find a great deal more at stake than just how the word 'faith' should be used. What is customarily at stake is how to understand the fundamental character of a person's proper relation to God; what is often also at stake is the nature of God and of the human being. To give an example: the twentieth-century German-American theologian Paul Tillich (§4) argued that faith is 'ultimate concern'. That proposal was in good measure motivated by Tillich's traditional view that faith is the core of a person's proper relation to God, combined with his very nontraditional conviction that God is the impersonal 'Ground of Being'. The latter conviction made it impossible for him to think of faith as trusting in God's promises - an impersonal entity cannot make promises. Another example: there was a wide-ranging and highly diffuse dispute in the twentieth century as to whether faith has propositional content. The

quick way with the dispute would have been to say that on one concept of faith it does, and on another it does not. But once again, what was actually under dispute, in the context of the shared conviction that faith constitutes the fundamental character of the Christian life, was the nature of divine revelation. Those who held out for faith as having propositional content were of the view that God reveals truths, and that faith consists in part in accepting those truths. Those who held that faith does not have propositional content were of the view that divine revelation consists simply in God manifesting God's self, and that propositions enter the picture only when we human beings interpret the divine self-manifestation. Many other examples could be cited to make the same point: for example, the dispute between the Reformers, who insisted that faith at bottom is *fiducia* (belief *in*), and their medieval opponents, who customarily defined faith as *fides* (belief *that*).

3 Faith, volition and responsibility

Paul's theology of faith in his letter to the Roman Christians has spawned an enormous quantity of theological reflections and controversies. These - the second major type of discussion on faith - have repeatedly spilled over into philosophy. The reason is obvious. Paul interpreted Judaism as a religion of 'works righteousness'. That is to say, he interpreted Judaism as holding that the satisfactory performance of prescribed ethical and ritual 'works' would earn one the status of being righteous (justified) before God. To forestall Christianity becoming Judaism under another guise, with the 'work' that earns justification now being faith, he went on to argue that faith itself is not a 'work' of the person, in which that person can 'glory', but a 'work' of God. Some of the questions this picture poses are subtle; others are obvious. If faith is entirely a gift of God and not a 'work' of the human being, how can anyone other than God rightly be held accountable for a given human being's lack of faith? How can it be just of God to punish a person for lacking faith? Indeed, how can it be just of God to grant it to some and not to others - since presumably the gift of faith itself can no more be earned than can the status of being justified before God? If faith is a good, how can it be good of God to withhold it from certain human beings? To say that the Pauline view posed profound issues of responsibility, free agency, and justice would be to understate the matter. It can scarcely be doubted that the depth of the discussion of these issues in the philosophical tradition of the West is in good measure due to the Pauline inheritance.

4 Faith and reason: the relation of faith to theorizing

The bulk of the remaining discussions of faith, in the philosophical tradition of the West, have been concerned with faith and reason, to use the traditional nomenclature. The nomenclature in fact covers issues of two quite different sorts: the role of faith in theorizing, and the rationality of faith itself.

Issues surrounding the role of faith in theorizing arose out of the confluence between the Bible on the one hand and the texts of classical Greek philosophy on the other. Both Plato and Aristotle held that the ideal human life is the life devoted to theorizing. Though their understanding of theorizing was different in various ways, both regarded it as the acquisition of knowledge, *episteme*, this being understood as awareness or insight; and Plato especially was emphatic in his insistence that the acquisition of *episteme* requires turning away from opinion or belief (*doxa*). Almost everybody in the Christian tradition of late antiquity and the Middle Ages agreed with the classical Greeks that rational insight is a human ideal. Yet none of them thought of faith as *episteme* - or, to use the Latin, *scientia*; all were vividly aware of Paul's statement, in 1 Corinthians 13, that 'now we see in a mirror dimly, but then face to face. Now I know in part; then I shall understand fully'. Hence the problem: as Christians they could not forswear faith, nor regard it as unimportant. Yet, as inheritors of the Greek tradition, they found themselves incapable of renouncing the ideal of rational insight into reality. The 'faith and reason' problem of late antiquity and the Middle Ages was the problem of how to resolve this tension.

Rather than offering at this point a rapid taxonomy of the various attempts at resolution, the main contours of a representative and influential formulation - that of Thomas Aquinas (§11) - will be sketched. Since the deepest happiness of a rational being is found in, and to be derived from, rational activity of the noblest sort, and since God is the one who, above all, is worth our being in rational contact with, our deepest happiness is to be achieved by gaining rational, intellectual contact with God. In general, it is not just the *activity* of exercising reason that gives us happiness, but the *love* of the object known, which is the intrinsic accompaniment of achieving the goal of the activity - namely, coming to know.

One way of gaining intellectual contact with God is by the practice of Aristotelian *scientia*: one starts from

propositions which some rational being has seen to be true, and then proceeds to deduce consequences. In the ideal case, one's premises are propositions that one has oneself seen to be true. Given our human limitations, however, that is often not possible. *Scientia* in its actuality is a communitarian enterprise in which one practitioner often takes it on the say-so of another that they, the latter, have constructed a demonstration of *P* starting from propositions which they personally saw to be true, and in which the first practitioner then takes *P* as a premise for their own work, without trying to prove it independently for themselves. It was Aquinas' conviction that it is possible to prove certain things about God in the ideal fashion; it is possible to construct a natural theology - a *scientia* of God (see Natural philosophy, medieval §§3, 7).

This *scientia* has its limitations, however, especially when assessed by reference to the happiness it contributes to human life. Prominent among the deficiencies, though by no means exhaustive of them, is the fact that, relative to the fullness of God's reality, there is not very much that one can prove about God. Natural theology, as Aquinas saw and practised it, was as subtle and complex in its argumentation and as rich in its results as any other natural scientific enterprise; yet, when one surveys the whole, one notices that nowhere does it yield knowledge of the nature (essence) of God. All that is provable is various relations of God to other things, various consequences of God's activity, and various things that God is not.

Now suppose we understand faith as belonging to the genus of 'believing something on someone's say-so' - specifically, on God's say-so. Faith thus understood also puts us in rational, intellectual contact with God. And when compared to natural theology, it proves advantageous in a number of ways. Prominent among them is the fact that the content of faith extends, at various points, beyond the reach of natural theology; belief that God is trinitarian in nature is an example. On the other hand, there is one serious and undeniable deficiency in faith when compared to natural theology. Whereas the practitioner of natural theology comes to 'see' various things to be true of God, the believer is confined to the inferior mental state of believing things about God on say-so. It is indeed God's say-so; and that makes faith more than mere opinion. Though like opinion in being belief, faith is unlike opinion in being utterly reliable belief, since God is the one believed. None the less, there is no gainsaying the fact that the mental act of 'seeing' a proposition to be true is distinctly better than that of believing it on say-so. It follows that even for a believer there is an advantage in the pursuit of natural theology - given that the believer in question has the time, the intelligence and the training necessary. By the practice of natural theology, the believer 'transmutes' that mode of intellectual contact which they have with regard to various propositions about God, from believing them on say-so to 'seeing' them to be true.

In this precise way, Aquinas would be happy to affirm the Augustinian formula, 'faith seeking understanding' (*fides quaerens intellectum*). He would, in his own way, also be willing to affirm Augustine's other formula, bequeathed to him by Clement of Alexandria, 'I believe in order that I may know' (*credo ut intelligam*). Augustine (§11) not only held that believers should strive to understand those things about God which they already believe, but also, conversely, that a condition of their coming to understand those things is that first they believe them. Augustine's thought was that faith is intertwined with love; and that there is no hope of attaining genuine understanding of God until one loves God. The theme is picked up and expressed in plaintive language by Anselm in the opening pages of his *Proslogion*, where he says that 'the smoke of our wrongdoing' will prohibit us from knowledge of God until that smoke is dissipated by faith. And it is picked up by Aquinas in his teaching that faith and charity are prerequisite for the achievement of such intellectual states as *scientia* and *sapientia* (wisdom).

5 Faith and reason: the rationality of faith

What strikes us in reading Aquinas and other medieval thinkers is that virtually no attention is paid to the question whether faith is itself rational. That question - the other 'faith and reason' topic - first became a matter of concern in the early Enlightenment. Again, rather than attempting a rapid taxonomy of positions held, it will be useful to develop the thought of a representative and influential figure - John Locke (§7).

Locke was living at a time when the traditional medieval view, that the great bulk of the texts bequeathed us together contain a unified body of highly articulate wisdom on moral, religious and philosophical matters, no longer seemed plausible. The textual tradition was now seen as fractured into incompatible components. Furthermore, though the Aristotelian concept of *scientia* remained part of Locke's conceptual repertoire, he himself was of the view that human knowledge in general, and *scientia* in particular, comes to very little. Thus it was that the main question to which Locke addressed himself was this: how should opinion (belief, *doxa*) be

regulated? His proposal was that on those matters where we are obliged to do our best to find out the truth of the matter, we ought to employ the following practice: first, collect satisfactory evidence concerning the truth or falsity of the proposition in question, this evidence consisting of things of which one is certain because one ‘sees’ them to be true; then determine the probability of the proposition in question on that evidence; and then believe or disbelieve the proposition with a firmness proportioned to its probability on that evidence. Locke believed that every normal adult is obliged to do their best on matters of religion - hence, to employ his proposed method on such matters. And since it was his view that no substantive propositions about God can be ‘seen’ to be true, it followed straightforwardly that he was an ‘evidentialist’ concerning theistic belief. One is entitled to believe things about God only if one believes them on propositional evidence. Evidentialism concerning theistic belief has been a prominent position in the West ever since Locke.

Locke’s own concept of faith was exactly the same as Aquinas’: faith is believing things on God’s say-so. Locke did not, however, follow Aquinas in holding that faith, thus understood, is a thing between knowledge and opinion; faith is just a species of opinion. Even more importantly, Locke differed from Aquinas in his emphatic insistence that it is irresponsible to believe something on the ground that God has said or revealed it unless one has good reason, of the indicated sort, for the assumption that God has indeed said or revealed it. Locke himself was of the view that only the occurrence of miracles provides good evidence for divine speech or revelation. In summary: believing something on God’s say-so is responsible only if one has first established in the specified way that there indeed exists a God who is capable of speaking or revealing, and then established, secondly, that there is good evidence of the specified sort for the conclusion that God has said that particular thing.

The way in which Locke connected reason, responsibility and faith has proved enormously influential in the modern world. A great many believers have attempted to meet the Lockean requirement for responsible faith by a combination of collecting evidence of the requisite sort and narrowing the scope of their faith until it contains nothing that the evidence does not justify. A great many nonbelievers have tried to show that the attempts all fail - and that faith is accordingly irresponsible. And a helpful way to look at the project of the Oxford philosopher Richard Swinburne is to look at it as a rigorous and far-ranging attempt to resuscitate the Lockean project in the twentieth century (see Swinburne 1984).

In recent years, however, Locke’s epistemology of religious belief has itself come under persistent attack. Wittgensteinian philosophers of religion, represented especially by D.Z. Phillips (1971, 1976), have argued that religion has its own unique criteria for acceptable belief, and that the flaw in Locke’s argument is that he attempts to impose criteria of acceptability on religious belief which are alien thereto. By contrast, the so-called ‘Reformed epistemologists’ - Alston, Plantinga and Wolterstorff being prominent representatives - while vigorously affirming the propriety of formulating a general epistemology and working at doing so, have gone on to contend that Locke’s foundationalist epistemology as a whole is seriously flawed; and that, in particular, it is often the case that a person who believes some theistic proposition immediately (or basically - that is, not on the basis of propositional evidence) is entirely within their rights in so doing. It is Locke’s universalizing insistence on the need for propositional evidence concerning religious beliefs that is misguided (see Religion and epistemology).

It should be noted that neither the Wittgensteinians nor the Reformed epistemologists espouse the view that the rationality of faith is irrelevant to whether or not it is acceptable to embrace it. That position has traditionally gone under the name of ‘fideism’, and such thinkers as Tertullian, William of Ockham (§9), Pierre Bayle (§1), Søren Kierkegaard (§§4-5) and Leo Shestov (§§3-4) have regularly been classified as fideists. However, when one digs into the thought of the five thinkers mentioned on the topic of faith and reason, one not only finds it to be in each case subtle, and the various positions taken diverse, one also becomes less and less convinced of the propriety of describing any of them as a defender of irrationality in matters of faith. The difficulty lies in good measure in the protean character of our word ‘rational’. If one accepts the Lockean paradigm of rationality, then the Wittgensteinians and the Reformed epistemologists will also be regarded as defenders of irrationality in matters of faith - a characterization which both parties themselves, however, would firmly reject. In short, interwoven with different positions on the rationality of faith are different positions on the nature of rationality - as, indeed, on the nature of faith.

See also: Justification, religious; Natural theology; Negative theology; Religious experience; Theological virtues

NICHOLAS P. WOLTERSTORFF

References and further reading

Alston, W.P. (1991) *Perceiving God*, Ithaca, NY: Cornell University Press.(An extended 'Reformed epistemological' treatment of religious experience.)

Aquinas, T. (1259-65) *Summa contra gentiles* I, trans. A.C. Pegis, Notre Dame, IN: University of Notre Dame Press, 1975.(The best introduction to Aquinas' understanding of the *scientia* of theology and its relation to faith; though see also the opening of his later *Summa theologiae*.)

Aquinas, T. (1266-73) *Summa theologica* IIaIIae, qq. 1-7, trans. Dominican Brothers, New York: Benziger Brothers, 1947.(Aquinas' fullest and most mature discussion of faith.)

Bultmann, R. (1968) '*pisteuo*', in G. Kittel (ed.) *Theological Word Book of the New Testament*, Grand Rapids, MI: Eerdmans, vol. 6, 174-228.(Though Bultmann has his own theological axe to grind, this remains a classic discussion of the various understandings of faith operative in the Bible.)

Kenny, A. (1992) *What Is Faith?*, Oxford: Oxford University Press.(This, along with Swinburne's book below, is a good entry into contemporary philosophical discussions concerning faith.)

Locke, J. (1689) *Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1975. (Locke's account of the relation between faith and reason is developed in Book IV.)

Matczak, S.A. (1967) 'Fideism', in W.J. McDonald (ed.) *New Catholic Encyclopedia*, New York: McGraw-Hill. (An authoritative Catholic treatment of fideism.)

Phillips, D.Z. (1971) *Faith and Philosophical Inquiry*, New York: Schocken Books. (This and the following work are the best examples of the Wittgensteinian philosophy of religious belief.)

Phillips, D.Z. (1976) *Religion without Explanation*, Oxford: Blackwell.(See previous annotation.)

Plantinga, A. (1993) *Warrant*, Oxford: Oxford University Press, 2 vols.(A detailed treatment of the general epistemological theory underlying Plantinga's 'Reformed epistemology'.)

Plantinga, A. and Wolterstorff, N. (eds) (1983) *Faith and Rationality*, Notre Dame, IN: Notre Dame University Press.(The articles by Alston, Plantinga and Wolterstorff are seminal examples of 'Reformed epistemology'.)

Popkin, R.H. (1972) 'Fideism', in Edwards, P. (ed.) *Encyclopedia of Philosophy*, New York: Macmillan.(A good, traditional philosophical treatment of fideism.)

Swinburne, R. (1984) *Faith and Reason*, Oxford: Oxford University Press. (Swinburne's corpus as a whole can be seen as an attempt to revive the Lockean vision in the twentieth century; this is perhaps the best point of entry.)

Wolterstorff, N. (1995) *Divine Discourse: Philosophical Reflections on the Claim that God Speaks*, Cambridge: Cambridge University Press.(A treatment of biblical interpretation by a 'Reformed epistemologist'.)

Fallacies

Fallacies are common types of arguments that have a strong tendency to go badly wrong, or to be used as deceptive tricks when two parties reason together. In some instances they are simply careless errors in thinking, while in other cases they are techniques of argumentation used by one arguer more or less deliberately to try to fool another into accepting a false conclusion. Fallacies have been described in logic textbooks since the time of Aristotle; their study, long neglected, has begun to be revived in recent years, as its practical importance in natural language reasoning has become apparent.

Many fallacies have been recognized, but some major ones that are unquestionably very common and influential in familiar thinking on all subjects have been emphasized in logic textbooks and manuals since the time of Aristotle - see Hamblin (1970) for a survey of these and their historical background. Fallacies were first systematically studied by Aristotle in his book *De Sophisticis Elenchis (On Sophistical Refutations)* and many of the textbook treatments have basically followed Aristotle's, except that more fallacies have been added along the way. Below, some of the traditionally most important fallacies are briefly characterized.

The '*ad hominem* fallacy' is the use of personal attack to try to undermine or refute a person's argument. Broadly, this takes two forms. The 'abusive *ad hominem*' consists of a direct attack on the arguer's character, and in particular, honesty. The 'circumstantial *ad hominem*' alleges some form of conflict between the arguer's personal circumstances and the argument. For example, it might be the charge that the arguer 'does not practise what he preaches', or that the arguer is biased because of having something to gain financially by advocating the argument in question.

The '*ad verecundiam* fallacy' is the use of appeal to expert opinion to support one's argument in such a way as to take advantage of one's opponent's respect for authority, allowing no room to question the pronouncement of an expert. Appeal to expert opinion is a common, and in principle legitimate, type of argumentation, but one that is easily abused by being pressed too heavily in an argumentative exchange. Experts can be (spectacularly) wrong in some cases, and their views can be misquoted, or otherwise distorted, especially when they use technical language.

The '*ad ignorantiam* fallacy' is the exploitation of ignorance (lack of knowledge) by one party in a dialogue to try to shift the burden of proof unjustifiably onto the other party. This type of argument takes two forms: (1) It is not known that proposition *A* is true; therefore *A* is false; (2) it is not known that *A* is false; therefore *A* is true. An example would be: 'This supposed theorem has never in fact been proved to be true, therefore it must be false'. In such a case, it could be that the conjecture has just been very difficult to prove, and no one has succeeded so far, but that somebody will prove it in the future. The fallaciousness of such an argument is indicated by the fact that there is such a thing as an incompleteness proof in logic; a proof that some propositions cannot be proved (see Gödel's Theorems).

The '*ad baculum* fallacy' is the use of threat or, more broadly, scaremongering, to try to exploit the insecurity of one's opponent to get them to accept some proposition, instead of justifying the proposition by good evidence. Instances of the *ad baculum* fallacy often take the form of covert threats, for example, 'I wouldn't say that if I were you - the last person who said it wound up on the bottom of the river wearing cement boots!'

The '*ad misericordiam* fallacy' is the blowing up out of proportion of an appeal to pity to cover up for lack of evidence to support a claim; or to brush aside other considerations that should be given more importance in arriving at a conclusion. There are many instances of quite reasonable arguments based on appeals to compassion or sympathy, but it is easy to blow such emotional appeals out of proportion to the slender weight of evidence that should be attached to them (within a larger evidential picture).

The '*ad populum* fallacy' is committed when an arguer suggests that someone's argument is wrong or unacceptable in so far as it deviates from popular opinion, or the conventionally accepted view. When pursued with rhetorical enthusiasm, this type of argumentation is called 'mob appeal'.

'Begging the question' is a kind of fallacious circular reasoning in which the conclusion is itself required as a premise to support the argument being advanced to justify the conclusion. Many cases of this type of fallacy are

disguised by being quite complex arguments - see Walton (1991) - but a simple case would be the following argument: someone asked to prove that God is benevolent replies, ‘God has all the virtues, therefore God is benevolent.’ Since benevolence is presumably a virtue, the premise requires the prior assumption that God is benevolent.

‘Equivocation’ is the fallacy whereby an argument contains an ambiguous word or phrase that means one thing in one premise, but something different in another premise or in the conclusion. As a consequence of this shift of meaning, one may be fooled into thinking one has been presented with a single argument, which looks pretty good.

The ‘straw man fallacy’ is committed when one arguer misrepresents another’s position to make it appear less plausible than it really is, in order more easily to criticize or refute it. For example, an environmentalist position may be exaggerated by saying, ‘They are trying to make this country a paradise on earth, by barring all industrial development. They will cause all of us to become unemployed.’

The ‘fallacy of affirming the consequent’ appears to be valid because of its superficial similarity to the valid form of argument *modus ponens*. An example of this valid form would be: ‘If it is raining, the streets are wet. It is raining. Therefore, the streets are wet.’ The comparable argument in the form of affirming the consequent may have the superficial appearance of also being valid: ‘If it is raining, the streets are wet. The streets are wet. Therefore, it is raining.’ (But maybe the streets are wet because a water main has burst.) It is perhaps because of our tendency to reverse the first premise in some instances that we can be deceived by this kind of fallacious reasoning.

See also: Logic, ancient §1; Logical and mathematical terms, glossary of;Formal and informal logic

DOUGLAS WALTON

References and further reading

Aristotle (*c*.330 BC) *De Sophisticis Elenchis*, trans. E.S. Forster, *On Sophistical Refutations*, Loeb Classical Library Edition, Cambridge, MA: Harvard University Press, 1938.(Historical.)

Hamblin, C.L. (1970) *Fallacies*, London: Methuen; repr. Newport News, VA: Vale Press, 1993.(Still extremely useful, especially on historical aspects.)

Walton, D.N. (1991) *Begging the Question*, New York: Greenwood Press.(On circular arguments.)

Walton, D.N. (1995) *A Pragmatic Theory of Fallacy*, Tuscaloosa, AL: University of Alabama Press.(Up to date and comprehensive.)

Fallibilism

Fallibilism is a philosophical doctrine regarding natural science, most closely associated with Charles Sanders Peirce, which maintains that our scientific knowledge claims are invariably vulnerable and may turn out to be false. Scientific theories cannot be asserted as true categorically, but only as having some probability of being true. Fallibilists insist on our inability to attain the final and definitive truth regarding the theoretical concerns of natural science - in particular at the level of theoretical physics. At any rate, at this level of generality and precision each of our accepted beliefs may turn out to be false, and many of them will. Fallibilism does not insist on the falsity of our scientific claims but rather on their tentativity as inevitable estimates: it does not hold that knowledge is unavailable here, but rather that it is always provisional.

1 The fallibilist perspective

Fallibilism is the view that we have no assurance that our scientific theories or systems are definitely true; they are simply the best we can do here and now. New theoretical knowledge does not just supplement but generally upsets our knowledge-in-hand. A scientific theory or system - like any product of human contrivance be it a house or a knowledge-claim - is ultimately bound to encounter conditions that its constructors did not and could not anticipate and which render its ultimate failure likely. Changed social conditions destabilize social systems; changed physical conditions destabilize physical structures; changed experiential (that is, experimental and observational) conditions - changed scientific technology, if you will - destabilize scientific theories. Rational inquiry links the products our understanding to the experienced conditions of a world in which chance and chaos plays an ineliminable role, so that there will always be new conditions and circumstances that ultimately threaten our rational contrivances. Thomas Kuhn's picture of scientific progress, which envisages a repetitive cycle of periods of normal science punctuated by revolutionary revisions, lends itself naturally to such a construction (see Kuhn, T. §§2-3).

Peirce, the most prominent exponent of fallibilism, has not been alone. Other recent thinkers who have espoused scientific fallibilism include Gaston Bachelard, Rudolf Carnap, Pierre Duhem, and Karl Popper. In antiquity, the mitigated sceptics of the middle academy were the precursors of this position, since they taught that we can never achieve certain knowledge (*epistēmē*) in matters regarding the world's ways, but have to make do with what is merely probable or plausible (*to pithanon*) (see Carneades §4).

2 Grounds for fallibilism

Why should we see all of our claims to knowledge of the world - even our best state-of-the-art scientific theories - as vulnerable? One of the most basic lessons of modern epistemology is that the available observational data underdetermine theories. For observations are always discrete, finite in number, and episodic. They deal in specific information-yielding episodes that occur at particular times and places. Theories, on the other hand, are general and non-finite: they deal in how certain features of a generic sort characterize situations of a certain kind always and everywhere. Accordingly, theories transcend the data and reach beyond them, so that there is always an evidential gap between the claims of a theory and the particular facts that are its supporting data. (The fitting of a continuous curve to discrete data points provides a somewhat oversimplified analogy.) Theories - as Pierre Duhem emphasized - lie beyond the possibility of definitive confirmation by data.

This data-transcendence of all theoretical claims means that there will always be various alternative ways of rounding the data off into generalized theories. The available data themselves can never constrain unique theoretical resolutions. Observed data are finite, possible hypotheses infinite, and the finite cannot constrain the infinite. The concrete realities we experience observationally are always to some extent ambiguous. They always admit, in principle, of alternative theoretical rationalizations. Theorizing without data is futile. But even theorizing on the basis of data involves interpretation, and thus carries in its wake the prospect of ambiguity, diversity and discord. The experiential realities determine and canalize the work of theorizing reason, but they cannot impose altogether unique resolutions upon it. In sum, theories are always informatively underdetermined by data (see Inference to the best explanation; Underdetermination).

There is much in this picture of the cognitive situation that rings true. The fact is that the equilibrium achieved by natural science at *any* given stage of its development is always an unstable one. The subject's history indicates that

scientific theories have a finite life span; they come to be modified or replaced under various innovative pressures, in particular the enhancement of observational and experimental evidence (through improved techniques of experimentation, more powerful means of observation and detection, superior procedures for data-processing, and so on). As C.S. Peirce and, subsequently, Karl Popper have insisted, we must acknowledge an inability to attain the final and definitive truth in the theoretical concerns of natural science - in particular at the level of theoretical physics. Our present science cannot plausibly claim to deliver a definitive picture of physical reality. We would like to think of our science as something safe, solid and reliable. Unfortunately, however, history militates against this comfortable view. In science, new knowledge does not just supplement but generally upsets our knowledge-in-hand.

We learn by empirical inquiry about empirical inquiry, and one of the key things we learn is that at no actual stage does science yield a final and unchanging result. We have no responsible alternative to supposing the imperfection of what we take ourselves to know. (And there is no reason to see the posture of our successors as fundamentally different from our own in this respect.) We occupy the predicament of the 'Preface Paradox' exemplified by the author who apologizes in his preface for those errors that have doubtless made their way into his work, and yet blithely remains committed to all those assertions he makes in the body of the work itself. We know or must presume that (at the synoptic level) there are errors, though we certainly cannot say where and how they arise (see Paradoxes, epistemic §1).

3 Approximation and its problems

To protect scientific realism in the face of the prospect of unending scientific change, it is tempting to adopt Peirce's stratagem of a resort to *convergent approximation*. This calls for envisaging a situation where, with the passage of time, the theories we successively arrive at grow increasingly concordant and their claims less and less differentiated. In the face of such a course of successive changes of ever-diminishing significance, we could proceed to maintain that the world really is not as *present* science claims it to be, but rather is as the ever-more clearly emerging science-in-the-limit claims it to be. The reality of ongoing change is now irrelevant, as with the passage of time the changes matter less and less. We increasingly approximate an essentially stable picture. This prospect is certainly a theoretically possible one. But neither historical experience nor considerations of general principle provide reason to think that it is actually possible.

Any theory of convergence in science, however carefully crafted, will shatter against the *conceptual innovation* that continually brings entirely new, radically different scientific concepts to the fore, carrying in its wake an ongoing wholesale revision of 'established fact'. Investigators of an earlier era not only did not *know* what the half-life of californium was, but they would not have *understood* it even if this fact had been explained to them. The questions we can pose are limited by our conceptual horizons. It did not appear to be a realistic prospect to all those late nineteenth-century physicists who investigated the properties of the luminiferous ether that no such medium for the transmission of light and electromagnetism might exist at all. Ongoing scientific progress is not simply a matter of increasing accuracy by extending the numbers in our otherwise stable descriptions of nature to a few more decimal places. Significant scientific progress is genuinely revolutionary in involving a fundamental change of mind about how things happen in the world. Progress of this calibre is generally a matter not of adding further facts, like filling in a crossword puzzle, but of changing the framework itself. And this blocks the theory of convergence.

Accordingly, we should temper our claims to scientific knowledge with a Cognitive Copernicanism. The original Copernican revolution made the point that there is nothing *ontologically* privileged about our own position in space. The fallibilistic doctrine now at issue effectively holds that there is nothing *cognitively* privileged about our own position in time. A kind of intellectual humility is in order - a diffidence that abstains from the hubris of pretensions to cognitive finality or centrality. Such a position calls for the humbling view that just as we think our predecessors of a hundred years had a fundamentally inadequate grasp on the furniture of the world, so our successors of a hundred years hence will almost certainly take the same view of *our* purported knowledge of things. The facts of epistemic life require us to recognize that, as concerns our scientific understanding of the world, our most secure knowledge is very likely no more than presently acceptable error. Such a view of scientific progress enjoins an even more modest position than Peircean convergentism.

In any convergent process, later is lesser. But since scientific progress on matters of fundamental importance is

generally a matter of replacement rather than mere supplementation, there is no reason to see the later issues of science as lesser issues in the significance of their bearing upon science as a cognitive enterprise - to think that nature will be cooperative in always yielding its most important secrets early on and reserving nothing but the relatively insignificant for later on. The fact is that a very small-scale effect - even one that lies very far out along the extremes of a 'range exploration' in terms of temperature, pressure, velocity, or the like - can force a far-reaching revolution and have a profound impact by way of major theoretical revisions. (Think of special relativity in relation to etherdrift experimentation, or general relativity in relation to the perihelion of Mercury.) Given that natural science progresses mainly by substitutions and replacements that go back to first principles and lead to comprehensive overall revisions of our picture of the phenomena at issue, it seems sensible to say that the shifts across successive scientific 'revolutions' maintain the same level of overall significance - it is neither a convergent nor a divergent process. Thus if we are to be realistic fallibilists we must hold a position more radical than that of Peirce. His approximationism cannot offset the sceptical impetus of fallibilism. There are, however, mitigating circumstances.

4 Scientific versus everyday knowledge

Fallibilists are prepared to concede that the claims of descriptive astronomy or of human anatomy are true enough. But they insist that the frontiers of theorizing are very different. With any sort of estimate, there is always a characteristic trade-off relationship between the evidential *security* of the estimate on the one hand (as determinable on the basis of its probability or degree of acceptability), and its contentual *definiteness* (exactness, detail, precision, and so on) on the other. But technical science forswears the looseness of vague generality: its law claims are strict and unshaded. In making the assertion 'The melting point of lead is $327.5°$,' we mean to assert that *all* (not most) pieces of (pure) lead will unfailingly melt at *exactly* (not somewhere around) this temperature. In technical science we always aim at the maximum of universality, precision, accuracy and exactness. And this commitment to generality and detailed precision renders the claims of science highly vulnerable. We realize that none of the hard claims of present-day frontier natural science will move untouched down the corridors of time.

The interesting fact is that fallibilism is a more plausible doctrine with respect to scientific knowledge than with respect to the less demanding 'knowledge' of everyday life. Ordinary-life generalizations are such as to allow one who asserts (say) that peaches are delicious, to be asserting something like, 'Most people will find the eating of suitably grown and duly matured peaches a rather pleasurable experience'. Such statements have all sorts of implicit safeguards, that can be cashed in through such guarding locutions as 'more or less', 'in ordinary circumstances', 'by and large' and so on. They are thus so well hedged that it is eminently implausible that contentions such as these should be overthrown.

In science, however, we willingly accept greater cognitive risks because we ask much more of the project. Here the objectives are primarily theoretical and governed by the aims of disinterested inquiry. Accordingly, the claims of informativeness - of generality, exactness, and precision - are paramount. We deliberately court risk by aiming at maximal definiteness and thus at maximal informativeness and testability. Aristotle's view that terrestrial science deals with what happens ordinarily and in the normal course of things has long ago been left by the wayside. The theories of modern natural science have little interest in what happens generally or by and large; they seek to transact their explanatory business in terms of strict universality - in terms of what happens always and everywhere and in all kinds of circumstances. We therefore have little choice but to acknowledge the vulnerability of our scientific statements, subject to the operation of the security-definiteness trade-off. Ironically, then, the 'common sense' of everyday life is securer than the more sophisticated claims of frontier science.

This distinction between the rougher but more secure level of 'schoolbook science', as it were, and the more demanding (but destabilizing) level of technical science at the frontiers of research, enables one to hold the realistic view that science does actually describe the real world, while accepting the implications of ongoing change in our scientific view of the world. The point is not that our pursuit of truth in science is futile, but rather that the information we obtain is to be seen as no more (but also no less) than the best available *estimate* of the truth that is available to us in the circumstances. Fallibilism is something very different from nihilistic scepticism.

5 Ethical ramifications of fallibilism

Some philosophers (Peirce included) see fallibilism as having ethical implications. They project an ethics of belief,

according to which we have no right to claim definitive truth for our current scientific claims, a view which they combine with a purported duty for the community of thinkers to pursue inquiry to the greatest extent possible in the circumstances. Accordingly, they insist that the fallibilism of our cognitive endeavours must emphatically *not* be construed as an open invitation to a sceptical abandonment of the scientific enterprise. Instead, it is an incentive to do the very best we can. In human inquiry, the cognitive ideal is correlative with the striving for definitive systematization. And this is an ideal that, like other ideals, is worthy of pursuit despite the fact that we must recognize that its full attainment lies beyond our grasp.

See also: Commonsensism; Epistemology and ethics; Scepticism

NICHOLAS RESCHER

References and further reading

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(Classic statement of scientific revisionism which stresses the dynamism of scientific change.)

Peirce, C.S. (1931) 'Principles of Philosophy', in *Collected Papers of C. S. Peirce*, Vol. 1, ed. C. Hartshorne and P. Weiss Cambridge, MA: Harvard University Press. (Classic statement of fallibilism by the father of the doctrine. Section 1.120 'The Uncertainty of Scientific Results' is especially relevant.)

Popper, K.R. (1959) *The Logic of Scientific Discovery*, New York: Basic Books.(Forceful attack on claims to definitive finality in science. Depicts Popper's 'conjection and refutation' theory of cognitive progress.)

Rescher, N. (1984) *The Limits of Science*, Berkeley and Los Angeles, CA: University of California Press.(A synoptic examination of scientific fallibilism. Extensive bibliographic information.)

Family, ethics and the

Do obligations to children take priority over filial and other family obligations? Do blood kin have stronger moral claims than relatives acquired through marriage? Whatever their origin, do family obligations take precedence over obligations to friends, neighbours, fellow citizens? Do family moral ties presuppose specific family feeling, love, or loyalty? Is the traditional family of a married, heterosexual couple with biological offspring morally preferable to families formed by adoptive, single, remarried or same-sex parents, or with the help of gamete donors or gestational ('maternal') surrogates? On what grounds may friends, neighbours or government agencies intrude upon 'family privacy'?

To simplify the complexity and diversity of family life, reasoned answers to such questions may stress a single dimension. A metaphysical *approach draws on the commands of a deity or the needs of a nation. A* biological *approach appeals to physical resemblance, blood or genes. An* economic *approach focuses on family property, income, division of work and resources, and inheritance. A related* political *approach attends to power, subordination, and rights within a family, as well as to their regulation by the state. A* psychological *approach takes affection, identification, intimacy, and emotional needs as morally decisive. A* narrative *approach makes recalling and revision of family stories the basis of moral education and the definition of family ties.*

Although mutually compatible, these approaches do each tend to favour particular moral theories.

1 Metaphysical conceptions of families

As in the major religions of Middle Eastern origin, parents may be regarded as recipients of God's gifts of newly created life. As such they are deputized guardians, not creators with powers of life and death over their creatures. Or, parents and children are seen as dutiful representatives of generations of long dead, but still attentive and demanding ancestors. In some versions or secular analogues, it is the nation or race that married couples must preserve and serve through the children they bear and rear. Accordingly, decisions about suitable marriage partners, or family size and comportment are decided by authorities speaking in the name of the transcendental agent or group, not by romantic couples or family-planning individuals.

Each of these theories prompts questions about the existence and moral authority of the metaphysical agent or collective invoked, such as: How can long dead ancestors continue to make demands of service or obedience on the living? How can a racial or national identity require those to whom it is ascribed to marry and procreate within its socially constructed limits? Answers to such questions often suppose that individuals are defined by particular relational properties (for instance, children of God the Father, sons of Erin, descendants of slaves). By contrast, Kantian and Utilitarian moral theories presuppose a metaphysical individualism of persons, each defined by nonrelational capacities for reason, good will, interests, or preferences (see Kantian ethics; Utilitarianism). Accordingly, children are not conceived as belonging to their parents, clan, or God, but as dependent on others for gradual development of their autonomy.

2 Biological conceptions of families

On a biological view, family ties are defined primarily through reproductive and genetic links. We distinguish between 'real' or 'true' mothers and adoptive, 'social' or step-mothers, full brothers and half-brothers, first cousins and distant cousins. Such traditional distinctions are accompanied by different levels of priority within families: the saying that 'blood is thicker than water' (and than diluted blood) indicates that blood kin have greater obligations to one another than to other relatives, but presumably find those greater demands less onerous in virtue of a natural sympathy that sociobiologists attribute to gene-preserving strategies (see Sociobiology).

However enshrined in legal and moral sentiment, this biological-based approach has problems of gradation, derivation and scope. Whatever genetic links or tendencies, do we owe half-sisters less loyal support? Do adoptive parents have more excuse for failing in their parental duties of care and comfort? Do adopted children have less obligation to 'honour' their parents, or more? If the moral gradations set out in such questions are indefensible, it would seem that an appeal to more than Darwinian notions would be needed to bridge Hume's gap between *is* and *ought* (see Logic of ethical discourse §§1-3).

But even if such prescriptive norms for family composition and conduct can be drawn from evolutionary or any other biological theory, there is an issue of scope. Whatever the value ascribed to biological connections, they are only a fraction of the ties by which families are formed and maintained. Biological considerations may govern marital and reproductive choices but have little or no bearing on many other family obligations, such as those of parents to their in-laws, or to one another, or to their children once they have become adult (grandchildren aside). Non-biological bases must be found for these other familial ties.

3 Economic conceptions of families

For many social scientists, a family is a household, a group of cohabitants who stand in various kinds of economic relationships, actual or potential. Kant and other family traditionalists have claimed male heads of family to have property rights in their wives and children (see Property). In many cultures, parents regard early support as an investment that children repay by working in or outside the home, by suitable marriages, and eventual care in old age. Adult children often regard their grandparents and parents as stewards of family assets to be used and distributed with due regard for the next generation (see Future generations, obligations to). So viewed, a family produces, consumes, accumulates and transmits economic goods.

This approach focuses on what for most families are principal sources of distress and dispute, namely, distribution of labour, goods, and economic responsibilities. It also helps expose unexamined divisions of family labour, for example, the unpaid labour of women who maintain the home and care for the young, ill and elderly family members who otherwise would need public services. In some areas of family life, economic assessment of women's traditional work is explicit, for example, in the hiring of substitute home childcare workers or gestational ('maternal') surrogates - practices often criticized for having a financial aspect that subverts the natural affection that ideally motivates the bearing and rearing of children.

Economic analyses are most compatible with utilitarian moral theories: both share the ideal of a quasi-calculus (of costs and benefits, or positive and negative utilities) for reaching principled decisions on actions or policies, for individuals or for groups with disparate interests or preferences (see Utilitarianism; Rational choice theory). In general, however, both approaches seem to ignore or discount a variety of morally relevant factors (fairness, needs, altruistic and selfish motives) that seem important in family matters. Moreover, both seem to presuppose individuals, with clearly distinguishable preferences and interests. But those families that do satisfy this requirement are not likely to have the solidarity, or intimacy, or communal history that other accounts of the family take to be definitive or primary (see Solidarity). Families composed of rational, self-interested, negotiating utility-maximizers seem like temporary roommates, not people who share a life over their lifetimes. This criticism does not, however, count against economic analyses less reliant on a calculus of individual interests.

4 Political conceptions of families

As in society generally, economic relationships reflect and produce within the family differences of power and subordination. Families are political in at least two senses: they are the 'basic units of society' in which social values are inculcated, enforced, or subverted; and they are also mini-realms in which power is distributed and exercised among and over its members. In some political theories, families have been the model for the larger polity, rulers cast as fathers and their subjects as their children.

John Locke (1689) attacked such patriarchy, both political and domestic, limiting the scope and duration of parents' authority over their children. Contemporary egalitarian critics view the traditional family as exploiting gender and age inequalities, as well as fostering unjust disparities of wealth. Some family theorists have found a remedy in Kantian notions of autonomy or in Rawls' 'difference principle', applied to families taken to be part of the basic structure of a society (see Autonomy, ethical; Rawls, J. §1). This principle allows differences in wealth, opportunities, and other 'primary social goods' only to the extent that they make the worst off members better than they would otherwise be.

Some feminist critics think that this principle still allows for traditional gender divisions of family labour (see Feminist political philosophy §5); communitarian critics find it too individualistic to do justice to the shared interests that sustain family life and obligation (see Community and communitarianism). Both types of critic tend to see the notion of individual rights as essentially adversarial, needed only when family feeling and harmony are

failing. On the other hand, family harmony and feeling may depend on the weaker members' confidence that conflicts of interest will be resolved equitably, a confidence that requires explicit recognition and respect of their individual rights of liberty and reciprocity.

5 Psychological conception of families

Neither the political nor the other approaches take much thought for the complex emotions that give families their peculiar fraught unity. In most modern cultures families are supposed to begin in love, or at least the hope of love, and be sustained by the love that grows among its members. Maternal love especially - whether connected with gestation and birth - is taken to be a deep, abiding base for self-sacrificial care and forgiveness (see Feminist ethics §2); brotherly love is supposed to be a model for relationships outside the family. But these emotional stereotypes ignore the variety and complexity of maternal or sibling sentiments.

To correct such sentimental simplicities, there is a growing philosophical literature on various intimate relationships and their emotions from which to draw. And, of course, any philosophically adequate phenomenology of family feeling must draw freely and widely from the world of plays, poetry, and novels in which familial emotional ties, both preserving and abusive, are central.

6 Narrative conceptions of families

Literature aside, families tend to create and recreate their own stories that serve to inspire loyalty, pride and identification with one another and with selected ancestors. By such stories of courageous grandparents, generous uncles, devoted sisters, good and bad in-laws, notions of family honour are defined and family trees are selectively pruned and shaped, thereby setting the moral parameters within which family members are expected to move, even if at odds with wider social mores.

This narrative approach is far from the Kantian conception of morality as based on general principles, applied impersonally across family and cultural boundaries (see Universalism in ethics). For any moral philosopher who insists that moral principles or lists of virtues and vices be derived from or constitute a human nature, to give moral weight to such family stories might seem to endorse a kind of idiosyncratic, asocial moralism without sufficient concern for the role of families in a civic culture. There is, however, nothing inherently asocial in a moral education that appeals to ancestral deeds and family tradition. Family heroes and villains are often remembered for their public as well as their private acts, virtues, and vices. Told and retold with revisions, these family litanies are akin to the common law, or much religious moral pedagogy (see Moral education). Philosophically, this narrative approach to family morality can find a framework in the existentialist ethics of thinkers such as Sartre, de Beauvoir, and Marcel (see Existentialist ethics).

Such literary approaches would favour an ethics of precedent rather than an ethics of principles, not unlike the common law and the teachings of the major Middle Eastern religions.

7 Conclusion

Each approach may yield reasoned answers to some of the various questions a comprehensive ethics of the family would pose. Whether we should expect any single approach to answer all such questions is debatable, especially in cultures for whom 'the family' no longer refers to a single, well-defined and widely-accepted norm. The diversity of families, as well as each family's complexity, may require a plurality of approaches, perhaps even greater than the six sketched here. What we should hope for, perhaps, is not a systematic ethics of the family, but a conceptually enriched capacity for moral reflection on family matters as they arise in context at different stages of our lives.

See also: Friendship; Genetics and ethics; Impartiality; Love; Reproduction and ethics; Sexuality, philosophy of

WILLIAM RUDDICK

References and further reading

Blustein, J. (1982) *Parents and Children: The Ethics of the Family*, New York: Oxford University Press.(An analysis that draws on Locke and Rawls for an analysis of parental obligations to meet a child's needs for 'primary goods' and to foster a child's autonomy.)

Bubeck, D. (1995) *Care, Gender, and Justice*, Oxford: Clarendon Press.(An economic and political analysis of women's traditional work of care in the family and elsewhere.)

Locke, J. (1689) *Two Treatises of Civil Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1963, esp. *Second Treatise*, ch. 6.(A political analysis of the limited scope and duration of parental authority and children's duty of obedience.)

Okin, S.M. (1989) *Justice, Gender, and the Family*. New York: Basic Books.(A liberal theory of justice meant to address women's inequalities within the traditional family more adequately than other theories of justice.)

O'Neill, O. (1988) 'Children's Rights and Children's Lives', *Ethics* 98: 445-63; repr. in *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge: Cambridge University Press, 1989.(A shift from political focus on children's rights to parental obligations of care, kindness, and other contributions to 'the genial play of life'.)

Ross, J.J. (1994) *The Virtues of the Family*, New York: Free Press.(A biologically-based defence of the traditional nuclear family.)

Sartre, J-P. (1964) *Les Mots*, trans. B. Frechtman, *The Words*, New York: Random House, 1981. (Autobiographical reflections on early relationships to his parents and grandparents, informed by his existentialist theories of authenticity and bad faith. An example of a narrative approach.)

Schoeman, F. (1980) 'Rights of Children, Rights of Parents, and the Moral Basis of the Family', *Ethics* 91: 6-19. (A psychological approach to family commitments in terms of intimacy.)Anthologies of original and reprinted philosophical essays, as well as useful bibliographies:

Aiken, W. and LaFollette, H. (eds) (1980) *Whose Child? Children's Rights, Parental Authority, and State Power*, Totowa, NJ: Rowman & Littlefield.(A collection of journalists' and philosophers' essays on child abuse and children's moral and legal rights.)

Archard, D. (1993) *Children: Rights and Childhood*, London: Routledge.(A concise critique of the 'liberal standard' for parental privacy, child abuse and state intervention, with a 'modest collectivist proposal' for children's sexual and voting rights. A compendious bibliographic essay.)

Ekman, R. (ed.) (1996) *Children's Rights Re-Visioned: Philosophical Readings*, Belmont, CA: Wadsworth.(A collection of philosophical essays examining the distinct interests and justified claims children have in regard to parental and state control and resources.)

Houlgate, L. (ed.) (1998) *Family Values: Issues in Ethics, Society and the Family*, Boulder, CO: Westview.(A collection of classic and current philosophic commentaries on ethical problems in marriage and family relations.)

Ladd, R.E. (ed.) (1996) *Children's Rights Re-Visioned: Philosophical Readings*, Belmont, CA: Wadsworth.(A collection of philosophical essays examining the distinct interests and justified claims children have in regard to parental and state control and resources.)

Meyers, D.T., Kipnis, K. and Murphy, C., Jr, (eds) (1993) *Kindred Matters: Rethinking the Philosophy of the Family*, Ithaca, NY: Cornell University Press. (Papers by philosophers and lawyers drawn mostly from the 1988 AMINTAPHIL conference on the family in moral and legal theory.)

Nelson, H. (1996) *Feminism and Families*, New York: Routledge.(A collection of critiques of, and various alternatives to, the 'traditional family'.)

O'Neill, O. and Ruddick, W. (eds) (1979) *Having Children: Philosophical and Legal Reflections on Parenthood*, New York: Oxford University Press.(A collection of philosophic essays and US legal decisions on the bearing and rearing of children.)

Scarre, G. (1989) *Children, Parents, and Politics*, Cambridge: Cambridge University Press.(A collection of philosophical and other essays on various conceptions of childhood and their bearing on parental and state rights and responsibilities in liberal democracies.)

Fanon, Frantz (1925-61)

Fanon's views (and often various misinterpretations of them) on the nature of colonialism, racism and the role of violence in Third-World revolutions were enormously influential. The main themes of all his writing are the critique of ethnopsychiatry and the Eurocentrism of psychoanalysis, the critique of négritude *and the development of a political philosophy for Third-World liberation.*

Frantz Fanon was born in the French Antilles on the island of Martinique and was educated there and in France. He served in the Free French army during the Second World War, both in north Africa and in Europe. He went on to study medicine and psychiatry at the University of Lyons between 1947 and 1951. In 1953 he was appointed chief of service of the psychiatry department of a hospital in Algeria (which was then still a French territory). He joined the Algerian liberation movement in 1954 and began to work for its underground newspaper El Moudjahid a few years later. His political activities caused him to leave his job, after which he moved to Tunisia where he practised psychiatry from 1957 to 1959. In 1961 he was appointed ambassador to Ghana by the Algerian provisional government. He died of leukaemia in 1961.

1 *Black Skin, White Masks*

Fanon's work is profoundly shaped both by his (largely psychoanalytic) training as a psychiatrist and his response to the work of European ethnopsychiatrists trying to understand the psychology of non-European peoples. But, like all African and Afro-Caribbean intellectuals in the Francophone world in the 1950s, he was also shaped (if only in reaction) by the ideas of the *négritude* movement (see African philosophy, Francophone §4). In his first book *Black Skin, White Masks* (1952), Fanon asserted that 'the black soul is a white man's artefact'. He claimed that the purportedly essential qualities of the negro spirit celebrated by the writers of *négritude* were a European fantasy. Fanon also argued against *négritude* that its assumption of a natural solidarity of all black people in the Caribbean and Africa was a political error; that far from needing to return to an African past, black intellectuals needed to adapt to modern European culture and that the movement had done nothing to change the everyday life of ordinary black people.

Yet even in this work (and more so later) he conceded that *négritude* could play an important role in freeing native intellectuals of dependence on metropolitan culture. This position which combined a critique of the theoretical claims of *négritude* with an account of it as a historically progressive movement, has affinities with Sartre's account of *négritude* in his essay 'Orphée noir' (1948) (see Sartre, J.-P.). Fanon's vehement attack on Sartre is surprising, therefore, in which he says that Sartre has undermined black fervour in his account of *négritude* as the 'minor term' of a dialectical progression from racism to humanism.

In *Black Skin, White Masks* Fanon developed an account of the psychological effects of racism. However, he did not acknowledge that he was referring to a racism that was based on the experiences of the black middle class in the French Caribbean. The dominant colonial culture, he argued, identifies the black skin of the negro with impurity. The Antilleans accept this association and have come to despise themselves. For example, colonial women exhibit their identification with whiteness by attempting neurotically to avoid black men and to get close to (and ultimately cohabit with) white men. This process Fanon dubbed 'lactification'.

This self-contempt manifests itself in other ways, for example, as anxiety in the presence of whites about revealing one's 'natural' negro inferiority; in a pathological hypersensitivity that Fanon dubbed 'affective erethism'; in existential dread and in a neurotic refusal to face up to the fact of one's own blackness. ('Erethism' is a medical word for pathologically irritable tissue: 'affective erethism', by analogy, is pathological irritability of the feelings.) Black children raised within the colonial system where racist cultural assumptions were prevalent can partially resolve the tension between contempt for blackness and their own dark skins by coming to think of themselves, in some sense, as white (hence the 'white masks' of the title). This fantasy can only be maintained away from direct exposure to Europeans: once the Antillean reaches metropolitan France, Fanon says, this illusion is destroyed with consequent damage to the ego.

2 Other works

Fanon's approach in *Black Skin, White Masks* focuses on the problems of identity created for the colonial subject

by colonial racism. In the course of the book he not only develops this account, but also criticizes the alternative accounts of the 'native mind' in the ethnopsychiatry of Octave Mannoni, proposing that his *Prospero and Caliban* (1964) is an apology for colonialism. He argues, too, that psychoanalytic theory (in the Adlerian and Jungian forms he draws on most heavily) requires substantial revision if it is to give an adequate account of the psychological development of non-European peoples.

In *Toward the African Revolution* (1964), Fanon developed an account of the ways in which the colonial system and its ideology over time adapted themselves to the changed economic realities. He argued that colonial racism begins as 'vulgar' racism. The aim of this racism is to prove the inferiority of colonial subjects by appealing to anatomy or physiology and especially to notions of the inferiority of their nervous systems. A later form of 'cultural' racism proposes that it is the culture of the native that is inferior. Fanon argued that the shift from vulgar to cultural racism reflected changes in the economic character of colonialism. As the colonial system developed it created a new class of colonial subjects trained in European methods and who, in collaborating with the colonizer, profited from the colonial system.

The changes associated with the creation of this new colonial class are characterized in *The Wretched of the Earth* (1961) as a transition from 'extractive' to 'consumer' colonialism. Whereas his earlier work had focused on a primary distinction between colonizer and the colonized, Fanon (1961) introduced, for the first time, a class analysis of colonialism which offered a more detailed account of the internal divisions of colonial societies.

In the extractive phase of colonialism there exists a peasantry, a traditional ruling group and a small number of native traders. In the transition from extractive to consumer colonialism a native petty bourgeoisie, a landed bourgeoisie, a proletariat and a lumpen-proletariat develop. However, Fanon rejects a conventional Marxist account of the colonial situation because he believed that the association of whiteness with money and power meant that the ruling class was identified by race, not by its relation to the means of production. Therefore, conventional class analysis and accounts of the role of ideology fail, power relations are crudely fixed and no ideology conceals the reality of the situation (see Marx, K.).

Instead, Fanon offers an account in which the peasantry are the revolutionary class who have escaped the total devastation of the traditional culture that has overtaken the bourgeoisie and the proletariat. Their revolutionary urge derives from their bid for land. Such a bid leads them to reject the colonizers' expropriation of the colonial territory. While Fanon expressed some ambivalence towards the revolutionary ardour of the peasantry, what was seen as his celebration of their revolutionary potential, rooted in their resistance to colonial cultural domination, was taken up by many Third-World intellectuals in the 1960s.

As Fanon died young and his last book was written under great pressure of time, there is no final attempt to bring consistency to the many strands of his work. For example, the psychological analysis of questions of identity begun in *Black Skin, White Masks* is never fully integrated with the increasingly political analysis of later essays and books.

See also: African philosophy, Francophone §7; Cultural identity; Marginality

K. ANTHONY APPIAH

List of works

Fanon, F. (1952) *Black Skin, White Masks*, London: Pluto Press, 1986.(His first major work.)

Fanon, F. (1959) *Studies in a Dying Colonialism*, London: Earthscan, 1989.(Essays mostly on the psychological harm done in Algeria by colonialism.)

Fanon, F. (1961) *The Wretched of the Earth*, London: Penguin, 1983.(His final book written over a ten-week period with the knowledge of his impending death from leukaemia.)

Fanon, F. (1964) *Toward the African Revolution*, London: Writers and Readers, 1980.(A collection of essays written between 1952 and 1959. The essay 'Racism and Culture' of 1956 is of special interest.)

References and further reading

Caute, D. (1970) *Fanon*, London: Fontana.(A brisk intellectual biography.)

Mannoni, O. (1964) *Prospero and Caliban*, trans. P. Powesland, New York, NY: Praeger.(A classic of ethnopsychiatry to which Fanon responded in *Black Skin, White Masks*.)

McCulloch, J. (1983) *Black Soul, White Artifact*, Cambridge: Cambridge University Press.(A careful discussion of Fanon's intellectual development that focuses on the connection between the clinical psychology and the political theory.)

Sartre, J.-P. (1948) 'Orphée noir', *Situations III*, Paris: Gallimard; trans. 'Black Orpheus', *Présence Africaine*, 1963.(Sartre's influential essay on the *négritude* movement.)

Fardella, Michelangelo (*c*.1646-1718)

Fardella was one of the first and most famous Italian Cartesians. Influenced by Malebranche and Leibniz, he rejected materialism in metaphysics, and endorsed a strongly Augustinian form of Cartesian dualism in philosophy of mind. In mathematics, he popularized Descartes' analytic method. An insufficiently radical thinker, the cultural significance of his work rests on his defence of the Cartesian method against scholasticism.

The Italian Michelangelo Fardella was educated by the Franciscans. Introduced by Giovanni Alfonso Borelli to mathematics and the Galilean method, he soon acquired an everlasting dislike for scholasticism. In Paris (1678-80), he became acquainted with the Cartesian movement and possibly Malebranche, whose writings affected his Augustinianism. He taught in Rome, Modena and Padua (1694-1710), where he was an influential advocate of Cartesianism. Between 1690 and 1714 he corresponded with Leibniz, and published his three major works: *Universae philosophiae systema* (1691), a methodological treatise presenting a critical combination of themes from the Logic of Port-Royal and Descartes; *Universae usualis mathematicae theoria* (1691), a presentation of the various branches of mathematics based on Descartes' method; and *Animae humanae natura ab Augustino detecta* (1698), a long, erudite commentary on Augustine's works, meant to demonstrate that the mind or soul is incorporeal and eternal (see Augustine §§5-6). In 1710 he became theologian and mathematician to Charles III in Barcelona. He died leaving most of his ambitious projects uncompleted.

Fardella celebrates Descartes as the first philosopher to have provided a proper method of philosophical investigation: as in mathematics, truth must be pursued by following only clear and certain inferences. Not only the rationality of Christianity, but also Descartes' metaphysics itself could profit from a firm application of the method. For, on its basis, it is possible to argue against the existence of the material world, as Fardella argues in the second appendix to *Universae philosophiae systema*. The mind is acquainted only with its ideas which are more like conceptual dispositions than images, and can be innate (God and mathematical principles) or acquired. Ideas are not different from their references, but since they could be exactly the same even if their ontological source were only another mind - Fardella denies any distinction in this respect between primary and secondary qualities - there is no reason to postulate an unlimited domain of independent, material bodies on the basis of a limited perception of appearances. This does not mean that only minds are real. The notion of a material substance is epistemologically redundant, but God still guarantees the existence of physical bodies, a conclusion that Berkeley explicitly criticized in his *Notebooks*. It does imply, however, that we should refrain from endorsing the Cartesian equation between *res extensa* and matter, and the subsequent mechanical physics of vortex.

Fardella interprets mathematics like Galileo, as the art whereby God created the universe, and infers that the world is intrinsically mathematical and that a science of nature must be quantitative. But then, instead of material substance we should speak of geometrical points, unextended atoms somewhat similar to Leibniz's monads, acting as the source of extension. The mind itself, whose actions Fardella places in the brain, is a mathematical point, immaterial and unextended. How this ontology might be possible, Fardella never made completely clear. Leibniz and Bayle objected that such geometrical points could not be material, or they would divisible, but could not be purely mathematical either, otherwise they would not give rise to extension.

Fardella's critique of our belief in the necessary existence of an external world remains nevertheless original. It differs both from Malebranche's position (Fardella does not provide probabilistic arguments for an external world), and from that of Berkeley (Fardella does not deny the existence of a mind-independent world). Unlike Descartes, Fardella maintains that God leaves us free to withhold our assent to any impression, and that our belief in the existence of material bodies is justified only by our faith in the revelation of the Holy Scriptures. Fardella's ontology is based on a weak understanding of the principle of sufficient reason, resembling the one endorsed by Samuel Clarke in his correspondence with Leibniz, according to which God's will is a sufficient *ratio essendi* of the external world.

See also: Augustinianism

LUCIANO FLORIDI

List of works

The first three works were meant to be part of a larger and very ambitious system of ‘philosophia architectonica’ which Fardella never completed.

Fardella, M. (1691) *Universae philosophiae systema*, Venice.(A textbook of philosophical logic in five sections: on the origin of error, the theory of perception and judgment, the principles of reasoning and the method of investigation. The second of two appendices contains Fardella’s anti-materialist argument.)

Fardella, M. (1691) *Universae usualis mathematicae theoria*, Venice.(The first comprehensive presentation of the various branches of mathematics to be written by an Italian according to the Cartesian method.)

Fardella, M. (1698) *Animae humanae natura ab Augustino detecta in libris de Animae Quantitate, decimo de Trinitate, et de Animae Immortalitate*, Venice; partial Italian trans. R. Venza (1991-2) as ‘L’Opera Filosofica di Michelangelo Fardella’, pts 1 and 2, *La Fardelliana* 10: 107-49 and 11: 107-64.(Meant to show how Cartesian rationalism combined easily with Augustinian spiritualism. The mind or soul is represented in Leibnizian terms, as a metaphysical point. Epicurus and Lucretius are the main polemical targets of Fardella’s anti-materialism.)

Fardella, M. (1691-1709) *Lettere ad Antonio Magliabechi*, Cassino: Editrice Garigliano, 1978.(Sixty-nine letters on disparate topics written by Fardella to the famous Florentine librarian.)

Fardella, M. (1698) *Pensieri Scientifici e Lettera Antiscolastica*, Naples: Bibliopolis, 1987. (A critical edition of two previously unpublished manuscripts. The *Lettera Antiscolastica* is a powerful and detailed attack against scholasticism in all fields of knowledge, from logic to physics and metaphysics.)

References and further reading

Barbata, A. and Corso, C. (1993) ‘Fra’ Michelangelo Fardella - Trapanese ‘‘in fuga’’ tra XVII e XVIII sec. nella Repubblica delle Lettere d’Europa’, *La Fardelliana* 12: 65-159.(Detailed reconstruction of Fardella’s life and intellectual biography; contains further bibliography.)

Berkeley, G. (1707-8) *Philosophical Commentaries*, in *The Works of George Berkeley, Bishop of Cloyne*, vol. 1, ed. A.A. Luce and T.E. Jessop, Edinburgh: Thomas Nelson.(Berkeley’s notebooks, first published in 1871. On page 15 Berkeley criticizes Fardella for being an immaterialist.)

Femiano, S. (1979) *Ricerca su Michelangelo Fardella, Filosofo e Matematico, in appendice due editi non conosciuti*, Cassino: Tipografia ‘S. Benedetto’.(A historical reconstruction of Fardella’s intellectual development and scientific production, which however does not discuss his philosophy.)

Garin, E. (1933) ‘Michelangelo Fardella’, *Giornale Critico della Filosofia Italiana* 14: 394-408.(A philosophical analysis of Fardella’s thought and his intellectual position within Italian Cartesianism.)

Lauria, D. (1974) *Agostinismo e Cartesianesimo in Michelangelo Fardella*, Catania: Niccolò Gannotta Editore.(A brief introduction to Fardella’s Augustinian Cartesianism including correspondence between Fardella and Matteo Giorgi about Descartes’ theory of material substance).

McCracken, C.J. (1986) ‘Stages on a Cartesian Road to Immaterialism’, *Journal of the History of Philosophy* 24: 19-40.(Places Fardella’s immaterialism in relation to the theories of Malebranche, Norris, Collier, Berkeley and Bayle.)

Meschini F.A. (1994) ‘Fardella, Michelangelo’, in *Dizionario Biografico degli Italiani*, vol. 44, Rome: Istituto dell’Enciclopedia Italiana, 776-81.(A good synthesis of Fardella’s life and works, with an extended bibliography.)

Pepe, L. (1982) ‘Notes on the dissemination of Descartes’ Geometrie in Italy in the 17th century’, *Bollettino di Storia delle Scienze Matematiche* 2: 249-88.(An interesting evaluation of Fardella’s work in mathematics.)

Predaval Magrini, M.V. (1986) ‘Problemi connessi all’edizione di un epistolario sei-settecentesco. La corrispondenza Leibniz-Fardella’, in G. Canziani and G. Paganini (eds) *Le Edizioni dei Testi Filosofici e Scientifici del ’500 e ’600*, Milan: Angeli, 183-91. (A well-documented overview of the correspondence between Leibniz and Fardella.)

Sasso, R. (1980) ‘Fardella et la Philosophie comme discours systématique’, *Recherches sur le XVIIème Siècle* 4: 115-25.(Focuses on Fardella’s conception of philosophy as an encyclopedic system of knowledge.)

Farrer, Austin Marsden (1904-68)

Austin Farrer is widely regarded as the Church of England's most brilliant philosophical theologian of the twentieth century. His elegant, sometimes difficult, writings aim more at synthesis and the elaboration of images than at analytical rigour; they characteristically take the form of a running debate between opposing positions. He focuses on the relations between divine and creaturely action in the spheres of nature, revelation and grace.

In the foreword to *The Freedom of the Will*, Austin Farrer writes:

> A host of able men, their wits sharpened by logical training, have written ingenious papers lately in the philosophical journals, each discussing voluntary freedom within the area of a carefully limited question. I could not hope to vie with their detailed thoroughness, or formal elegance. If I was to contribute anything, it would have to be by aiming at greater completeness, and at a synthesis of topics. So that is what I have done, well or ill.
>
> (1958: v)

This statement could aptly serve as a description of Farrer's entire philosophical output. He had an acute awareness of the metaphorical character of much apparently literal thinking, and an unrivalled deftness in penetrating and extending those metaphors, as his sermons and exegetical writings bear witness. And he was loath to take up any issue piecemeal. These features of Farrer's thought perhaps explain why his writing seems to lack rigour in comparison with those 'ingenious papers… in the philosophical journals'. They certainly make impossible the task of summarizing his thought in a short space. (Readers who want a quick introduction to Farrer's mature thinking on natural theology and on human freedom can do no better than to read his summaries at the end of *Faith and Speculation* and *The Freedom of the Will*.) So this entry will focus on an issue that was central to Farrer's thought: the relation between divine and human action.

The bare assertion of the existence of God, Farrer says, is 'of metaphysical interest only'. Practical importance attaches instead to assertions about God's particular actions. Since religion is precisely the voluntary association of one's own action with the divine, there can be no religion without some conception of how God acts in the world. And so a central problem for Farrer is how to conceive the relation between divine and creaturely action. Surprisingly, though, he insists that the 'causal joint' between infinite and finite action cannot be of any practical concern to us. The Christian holds that the divine action embraces whatever happens. But every event in our world is an immediate act of natural agents - otherwise it would be no part of our world. So:

> God's agency must actually be such as to work omnipotently on, in, or through creaturely agencies without either forcing them or competing with them. But as soon as we try to conceive it in action, we degrade it to the creaturely level and place it in the field of interacting causalities. The result can only be (if we take it literally) monstrosity and confusion.
>
> (1967: 62)

Given this position, the problem of human freedom and divine predestination in the achievement of salvation disappears. 'We know that the action of a man can be the action of God in him; our religious existence is an experimenting with this relation' (1967: 66). But since we cannot conceive the manner in which God acts, we cannot pose the question of the relation between his action and ours. Similarly, there is no room for a special concept of grace. 'Grace is an action of the Creator in the creature. He acts in the creature everywhere; when he acts in the rational creature he is pleased to act in that creature's mental and voluntary life, bringing them into his own' (1967: 67).

Farrer's understanding of revelation also depends on this conception of divine and human action. When we try to understand the experience of receiving a divine communication, we are often preoccupied with trying to locate the 'point of punctuation' between the divine and the human. But no such question should arise. God's will is always taking effect in the actions of his creatures. The events that make up revelation are simply those actions of God for the eternal salvation of his rational creatures 'in which the relation of human instrument to divine act becomes transparent' (1967: 103). 'God reveals himself effectively through fallible minds and takes care that their

imperfections shall not frustrate his purpose' (1967: 102).

See also: Grace; Revelation

THOMAS WILLIAMS

List of works

Farrer, A.M. (1943) *Finite and Infinite*, London: Dacre Press.(An early work of natural theology, sketching a voluntarist metaphysics.)

Farrer, A.M. (1948) *The Glass of Vision*, London: Dacre Press.(An investigation of the nature of revelation, with particular attention to the role of images in metaphysics, poetry and Scripture.)

Farrer, A.M. (1958) *The Freedom of the Will*, London: Adam & Charles Black; 2nd edn, New York: Charles Scribner's Sons, 1960.(The second edition includes a 'Summary of the Argument'. Farrer defends an incompatibilist account of human freedom and relates that account to empirical psychology, action theory and natural theology.)

Farrer, A.M. (1961) *Love Almighty and Ills Unlimited*, London: Collins.(An essay in theodicy, strictly philosophical in its discussion of physical disasters and animal suffering, but explicitly appealing to Christian revelation for the explanation of human suffering.)

Farrer, A.M. (1967) *Faith and Speculation*, London: Adam & Charles Black.(Farrer's mature thinking in natural theology, synthesizing his work on a great range of topics.)

Farrer, A.M. (1972) *Reflective Faith: Essays in Philosophical Theology*, ed. C.C. Conti, London: Society for Promoting Christian Knowledge.(A collection of essays spanning Farrer's whole career.)

References and further reading

Conti, C. (1995) *Metaphysical Personalism: An Analysis of Austin Farrer's Metaphysics of Theism*, Oxford: Clarendon Press.(Conti emphasizes the influence of process theology and personalism in the development of Farrer's thought and highlights the central role played by his notion of divine action.)

Eaton, J.C. and Loades, A. (eds) (1983) *For God and Clarity: New Essays in Honor of Austin Farrer*, Allison Park, PA: Pickwick Publications.(Chiefly valuable for its bibliography of works by and about Farrer.)

Hebblethwaite, B. and Henderson, E. (eds) (1990) *Divine Action: Studies Inspired by the Philosophical Theology of Austin Farrer*, Edinburgh: T. & T. Clark.(A collection of essays on divine action, most of which explore and extend Farrer's work. Contains an index to several of Farrer's books.)

Fascism

'Fascism' is a term referring both to a political ideology and to a concrete set of political movements and regimes. Its most prominent examples were the Italian and German regimes in the interwar period. Fascist ideology is sometimes portrayed as merely a mantle for political movements in search of power, but in reality it set forth a new vision of society, drawing on both left- and right-wing ideas. Fascists stressed the need for social cohesion and for strong leadership. They were more concerned to revitalize nations by cultural change than to propose institutional changes, but they saw themselves as offering a third way between capitalism and communism. There was no fascist philosophy as such, but fascist ideology drew inspiration from earlier philosophers, most notably Nietzsche and Sorel, and was supported by several contemporary philosophers, including Heidegger, Gentile and Schmitt.

There has always been a small number of philosophers and historians willing to concede that fascism has a serious intellectual pedigree. Some have traced its origins to Rousseau's concept of the 'general will' or Hegel's defence of the state. More commonly, the focus has been on the new nationalism which became prominent during the late nineteenth century, and which is best epitomized by the writings of Barrès or the German *völkisch* (blood nationalist) theorists such as de Lagarde. Whereas the nationalism associated with the French revolutionary doctrine of 'sovereignty of the people' tended to be universalistic and humanistic, the new strand was more emotive, stressing the bonds which linked the holistic community (see Nations and nationalism §2). This in turn was often linked to the rise of a more systematic form of racist thinking, inspired both by developments in science - especially Darwinism - and by a reading of history which saw nation and race as the main engine of change and conflict (see Race, theories of).

Other major sources of influence which have often been highlighted are the writings of Nietzsche, with his emphasis on the 'will to power', leadership and the total decadence of European bourgeois society, or Sorel's belief in the importance of myths as a means of motivating the masses (see Sorel, G.). These philosophical developments were reinforced by new ideas in the social sciences emerging at the turn of the twentieth century. Of particular importance was the work of elite theorists like Michels and Pareto, who argued that democracy was an impossibility, and the theories of psychologists like Freud and Le Bon, who pointed to the essential irrationality of much human behaviour.

These precursors were not, however, fascists in the full sense. Nietzsche, for instance, was opposed to anti-Semitism and emotive German nationalism - yet these were to be central elements in Nazism. Sorel was a revisionist socialist concerned with the problem of how to achieve a revolutionary working class, whereas the *völkisch* thinkers were nothing if not conservative. The First World War provided the impetus for fascism to develop as a more coherent ideology, synthesizing a variety of elements in pre-fascist thought.

Fascism took its name from the *fasci di combattimento*, a movement founded in 1919 by Mussolini, an ex-socialist attracted to nationalism as a means of uniting the people and promoting economic development which could fund radical social reform. Similar movements emerged elsewhere, sharing a set of ideological dispositions, although some, like the Nazis, did not normally refer to themselves as fascist.

The starting point was a view of human nature: fascists like Valois reversed Rousseau in the sense that they believed that 'man' was a savage until given a purpose by strong leaders. The need for a new fascist man was linked to a cyclical view of history, which saw conflict as inherent in the world: Hitler argued that it was necessary to ensure that the nation was sufficiently revitalized to withstand challenges from outside. Social cohesion required the destruction of deleterious humanistic or individualistic influences, like religion and liberalism, whilst the need to forge social unity made fascist theorists like Gentile and Schmitt hostile to the class analysis preached by much of the left (see Gentile, G. §3). In general, fascists paid little attention to specific institutional forms, preferring to talk vaguely about a Third Way (neither left nor right, rather lying between capitalism and communism), or the need to find a new sense of being - a major reason for Heidegger's conversion to fascism (see Heidegger, M. §1). The main focus was on changing culture and on the propaganda that was seen as necessary to do this, for fascism held that people were essentially swayed by myths and symbols rather than by rational argument - although some fascists, such as De Man and Mosley, were capable of developing serious economic and political programmes

based on a form of authoritarian and statist corporatism.

There were important differences between specific national manifestations of fascism. Nazism, for instance, saw the basis of the community as lying in ‘blood’, whereas Italian Fascism saw nationalism more in terms of culture. Thus, early fascism had no serious anti-Semitic side, or developed racial theory although it had an ethnocentric commitment to colonial expansion). Most Nazis also placed far less emphasis on the role of the state than Italian Fascists, seeing the state as a repository for reactionary elements and a source of division between the people and its leader. Moreover, some fascist intellectuals, like Drieu La Rochelle, were more concerned with the threat (from both internal decadence and more virile external enemies) to European rather than specifically national culture.

Linking all true examples of fascism, however, was a broad desire to revitalize society by forging a *holistic-national-radical-Third Way*. This characterization is flexible enough to embrace diverse examples of fascism like Nazism and Italian Fascism, while excluding other alleged cases such as General Franco’s Spain as too conservative. It is also capable of encompassing the more subtle form of post-war intellectual fascism, such as the writings of the key *nouvelle droite* theorist, de Benoist.

See also: Anti-Semitism; Totalitarianism

ROGER EATWELL

References and further reading

Aschheim. S. (1993) *The Nietzsche Legacy in Germany, 1890-1990*, Berkeley, CA: University of California Press. (Shows how Nietzsche’s legacy was manipulated, but also clearly illustrates the affinities between aspects of Nietzsche’s thought and fascism.)

Bendersky, J.W. (1983) *Carl Schmitt. Theorist for the Reich*, Princeton, NJ: Princeton University Press.(Shows that Schmitt was largely an authoritarian conservative opportunist - albeit one capable of producing a brilliant defence of statist fascism.)

Eatwell, R. (1996) ’On Defining the "Fascist Minimum": The Centrality of Ideology’, *Journal of Political Ideologies* 1 (3): 303-19.(Focusing especially on a critique of the notable interpretations of Griffin, Payne and Sternhell, this article moves on to offer an expanded version of the definition offered above.)

Gregor, A.J. (1979) *Young Mussolini and the Intellectual Origins of Fascism*, Berkeley, CA: University of California Press.(Views Mussolini as a relatively serious thinker, and traces fascism to left-wing roots; sees the resulting fascist ideology as being as coherent as liberalism or Marxism.)

Griffin, R. (1995) *Fascism: A Reader*, Oxford: Oxford University Press.(Excellent set of texts - coming up to the 1990s - which seek to illustrate the author’s thesis that fascism is a form a ‘palingenetic, populist, ultranationalism’, but cover other interpretations as well.)

Mosse, G. (1966) *The Crisis of German Ideology*, New York: Grosset & Dunlap.(Important study of the origins of Nazism in Romantic and *völkisch* thought; plays down more left-wing influences on Nazism seeing the Nazi revolution as threatening ‘none of the vested economic interests of the middle class’.)

Mosse, G. (1978) *Toward the Final Solution*, New York: Howard Fertig.(History of ideas approach to the rise of racism, showing its centrality to late nineteenth- and early twentieth-century European thought.)

Payne, S. (1980) *Fascism: Comparison and Definition*, Madison, WI: University of Wisconsin Press.(Important comparative typological book, written by a historian who defines fascism in terms of: its style, such as mass rallies; its negations, such as anti-communism; and its positive points, such as nationalism and authoritarianism.)

Popper, K. (1945) *The Open Society and Its Enemies*, London: Routledge & Kegan Paul, 2 vols.(Classic work which seeks to trace the roots of fascism - and communism - to important developments in Western thought. In the case of fascism, Hegel is seen as the main challenge to the liberal Western tradition of pluralism and tolerance.)

Rockmore, T. (1992) *On Heidegger’s Nazism and Philosophy*, London: Harvester.(One of several books to appear during the 1980s and 1990s which accepts that Heidegger can reasonably be considered a fascist.)

Sternhell, Z. (1986) *Neither Right Nor Left*, Berkeley, CA: University of California Press. (Argues both that fascism was a serious ideology and that French fascism, by not having to deal with the problems of office, produced the purest form of fascist theory; part of a trilogy, beginning in the late nineteenth century with Barrès.)

Sternhell, Z., Sznajder, M. and Asheri, M. (1994) *The Birth of Fascist Ideology*, Princeton, NJ: Princeton University Press.(Sternhell sees fascism as based on nationalism and socialism. Covers both France and Italy,

and pays particular emphasis to the impact of Sorel and anti-Marxist socialism. Sternhell does not consider Nazism to be fascist on account of its biological racist core).

Fatalism

'Fatalism' is sometimes used to mean the acceptance of determinism, along with a readiness to accept the consequence that there is no such thing as human freedom. The word is also often used in connection with a theological question: whether God's supposed foreknowledge means that the future is already fixed. But it is sometimes explained very differently, as the view that human choice and action have no influence on future events, which will be as they will be whatever we think or do. On the face of it this is barely coherent, and invites the assessment that fatalism is simply an expression of resigned acceptance.

Taken as meaning exactly what it says, the dictum that human choice and action have no influence on future events is absurd, since any action must make some aspects of some future events different from how they would have been had that action not been performed. If I leave the house, then *something* happens in the future which would not have happened had I not left it, even if that just turns out to be my rapid return to where I would have been had I stayed at home. Likewise, unless we are to deny that how we decide to behave ever has any effect on how we actually do behave, we must agree that our choices of action frequently affect future events. So if fatalism is not to fall into immediate incoherence, it must be understood rather differently, perhaps as saying only that there are *certain* things destined to happen irrespective of what we do. (Usually these will be events of particular significance: the soothsayer's prediction will come true, regardless of any attempts we may make to ensure that it does not.)

It follows that fatalism, in this form, cannot be supported by any argument which, if it worked at all, would apply to all future events. Any argument for thoroughgoing causal determinism would fall into this category. So would the notorious argument from the sea-battle: if there will be a sea-battle tomorrow, it is true now that there will be; if there will not, it is true now that there will not be. So since one of these is true now, there is nothing we can do to influence whether there will be a battle or not. If this reasoning works, it works for every future event. But in any case, although this argument and variants of it are found, it faces a severe difficulty. For it will have to address the crucial question about the direction of dependence: does the occurrence of the battle depend on the truth of today's statement, or the truth of today's statement on tomorrow's events? The fatalist conclusion calls for the former, but the latter is far more plausible and the basic argument does nothing to refute it.

Fatalism can appear more coherent if seen against a certain kind of background. Whether it appears more plausible depends on the reader's attitude to the background, which many will find rather too superstitious for their taste. We are to think of powers, watching over human life, who have decreed that certain things shall happen to certain people (that a particular individual will die young, that Oedipus will kill his own father), and are bent on manipulating the world, and perhaps also the fated individual's state of knowledge, so as to bring their decrees about. Imagine a rat in a maze, from which you have decided that it will never find its way out. If it turns left (which leads to the exit), you take out a piece from somewhere else in the maze and slot it in round the next corner, turning the chosen route into a dead end. If it turns right, you do nothing, since the right turn leads to a dead end anyway.

It should be noted that fatalism, understood in this way, has very little to do with problems about the freedom of the will. In choosing which way to turn, our rat can be free in the strongest sense that any libertarian has ever dreamed of; the fates make their move after it has made its choice. Oedipus may be equally free; the fates make sure that he does not know who the obstructive old man at the crossroads really is. Had he known that, no doubt he would have made his free choice differently, leaving the fates to outwit him in another way on another occasion.

See also: Determinism and indeterminism; Free will; Łukasiewicz, J. §3; Many-valued logics, philosophical issues in §1; Predestination; Stoicism §§20-1

EDWARD CRAIG

References and further reading

Aristotle (*c.* mid 3rd century BC) *De Interpretatione*, in *The Complete Works of Aristotle* ed. J Barnes, Princeton, NJ: Princeton University Press, 1984, vol. 1, 25-38.(Chapter 9 introduces the notorious sea-battle example. Aristotle appears to be arguing that to avoid fatalism we must give up the idea that the law of excluded middle

applies to statements about the future. However, the passage has given rise to a great deal of conflicting interpretation.)

Cicero (44 BC) *De Fato (On Fate)*, with trans. and commentary by R.W. Sharples, Warminster: Aris & Phillips, 1991.(Sections 28-30 offer an account of Stoic views on fatalism, which they appear to have understood both as determinism and in the sense emphasized in this entry.)

Taylor, R. (1963) *Metaphysics*, Englewood Cliffs, NJ: Prentice Hall.(Chapter 5 is on fatalism. It presents a version of the sea-battle argument, with the usual concealed defect.)

Fatalism, Indian

*Indian speculation about the vicissitudes of human life has a long and complex history. Life in the early Vedic period was considered to be largely hostage to the 'fate' of natural and psychic forces controlled by various gods (*devas*). Fate was what proceeded 'from the gods' (*daiva*), who were considered to be the guardians of the cosmic order and the ultimate source of prosperity. Sacrifice and prayer were the principal means to win their favour.*

Later the idea arose that one's present lot is due, not to the whim of some god, but to karma, the effect of one's own actions performed in this or previous lives. On this view, humans do have some scope or 'freedom' to change themselves and the environment in which they live. This more individual potential is called puruṣakāra, *which, to varying degrees, may modify* daiva. *The literal meaning of this term is 'human action' (from the Sanskrit for 'man' and a verbal root meaning 'to act'). With the increasing popularity of the karma theory,* daiva *tended to become equated with the effects of past behaviour.*

Finally, in the context of the spiritual ascent towards a unifying vision of existence, the status of human agency itself became an issue. As long as the seeker remains blinded by false notions of 'I', the ego must experience a sense of agency and a modicum of freedom to chart its course of life. However, from the perspective of enlightenment, or mokṣa, *all is 'fate' in the hands of a personal God or a Supreme Self.*

1 Background

The concept of 'fate' (*daiva*) is already well known in the *Ṛg Veda* (completed before 800 BC), the earliest of the sacred texts of modern Hinduism. Here fate is simply the favours obtained from the gods in return for praise or sacrifice. Sacrifices are performed to ensure the good order of the world, to secure a boon or to remove impurity associated with 'sin' (for example, disease). This seeming dependence on the whims of the gods changes rather dramatically, however, at the somewhat later period of the Brāhmaṇa texts (no later than 600 BC), when the priests take control of the gods themselves through their knowledge of the sacrifice. Power is understood to reside in the ritual itself rather than in the gods (who are virtually forced to act). The belief also arises that the cosmic powers can be harnessed by means of ascetic practices known as *tapas*, and by the appropriate use of incantations or mantras. This new magical tradition received its orthodox stamp of approval in the *Atharva Veda*.

There is still little at this period to suggest the existence of a 'person' in the modern sense of an autonomous, self-directing centre of willing and doing. The human agent remains at the mercy, as it were, of external forces. Plans to shape one's own future 'destiny' (in the form of prosperity, progeny, and so on) rely less on an inner sense of agency than on the power of gods, ritual or magic to exert control over the general order of the world. Furthermore, the esoteric knowledge on which this control is based is not in the hands of the individual, but remains a closely guarded secret of the priestly caste.

A shift towards the assertion of more distinctly human powers is signalled by the emergence of a doctrine of moral causality, subsequently known as the karma doctrine. First presented in the Bṛhadāraṇyaka Upaniṣad, this doctrine came to represent the hard facts of human life not as fate, but as the inevitable result of past behaviour. The results of all deeds, good or bad as the case may be, must ultimately bear fruit in one form or another. There is no such thing as chance or luck, and even the intervention of God himself becomes problematic in certain traditions. We are punished not *for* our sins but *by* them.

The doctrine drew its logic from the companion belief in reincarnation. Eventually, life itself, including character, the circumstances of birth, the span of life, and even the daily alternation of pleasure and pain, came to be regarded as the maturation of attitudes and actions in this and previous existences. In sum, the cumulative effects of the past are seen progressively to take the place of fate as the source of human character and circumstances.

2 Fate in the *Mahābhārata*

The *Mahābhārata* (probably compiled between 200 BC and AD 300-400) is a great repository of ideas on fate in the Indian context, moving from those reminiscent of the early Vedas to concepts of divine grace and the mature doctrines of karma. A humanization of divine powers is implicit in the circumstances surrounding the central subject of this epic, a great fratricidal war whose roots are anchored in a conflict of gods and demons for control of

heaven. For example, the difference between gods and humans is blurred by the fact that the battle is being waged on the earth, where human 'sons' of the divine prototypes battle for control of a dynastic succession. More importantly, this divine infusion has evidently sired in humans a sense of an inherent power of their own.

The most common Sanskrit term for this human capacity for self-effort is *puruṣakāra* (literally 'human action'). Only what is preordained by universal consensus of the heavenly gods is now sure to prevail on the Earth. Humans are otherwise at liberty to challenge the gods, although they do so at their peril since these gods may be jealous, or even fearful, of human aspirations. Indra, the king of the gods himself, can feel threatened by the fierce austerities of some power-hungry sage. The best-laid human plans may thus be blunted or bent by a *daiva* that confronts the individual either from within, as an eruption of doubt, inner conflict or irrational behaviour, or from without, in the form of some untoward reversal of fortune. The respective spheres and strengths of these two forces is a popular topic of learned discussion in the epic.

From the macrocosmic perspective, the fate of the cosmos is in the hands of Viṣṇu-Nārāyaṇa in his supreme aspect. It is he who controls the Wheel of Time (*kālacakra*), the immense temporal cycle of creation (activity) and destruction (quiescence) of the seven worlds (*lokas*). It is he too who must provide the superordinate backing to settle the difficult or contentious rulings of Brahmā, the epic arbiter of heavenly disputes. At the human level, the fate of society is governed by a 'morality' of world ages (*yugas*) that repeat themselves with chronological precision as long as the universe is in being. The crisis that arises at each low point of moral decay can only be resolved by the eschatological destruction of the old order of society by the Divinity in his lower, human, aspect (*avatāra*). The *Mahābhārata* is essentially the (quasi-) history of one such intervention, which follows upon a prearranged heavenly accord between the old Vedic gods and Viṣṇu-Nārāyaṇa, incarnated as Kṛṣṇa. He does no actual fighting, but it is his strategic guidance that leads the human 'sons' of these gods to eventual victory.

As events escalate towards their momentously regenerative climax, the fate of individuals becomes ever more hopelessly caught up in the chain of events. Not surprisingly, the protagonists tend to rail against fate, as their own ambitions encounter inexplicable reversals in the world around them. This is particularly true of blind King Dhṛtarāṣṭra. As Kaurava losses mount, he becomes more and more convinced that he is the victim of a fate opposed to reason and human volition. *Puruṣakāra* appears increasingly ephemeral, inconsequential or futile as he is dragged from one disaster to another by his own emotional attachment to his son. These feelings are expressed in an endless litany of complaints against cosmic justice.

The more widely held belief, however, is that human planning can succeed through *puruṣakāra*, provided the undertaking is in accord with the prescribed order of *dharma* (see Duty and virtue, Indian conceptions of §§1-2). *Puruṣakāra* is considered a particularly vital quality for a king, whose *dharma* is to protect the all-important Brahmans and his subjects as a whole. The important watchword 'Victory is where the *dharma* [or Kṛṣṇa] is', is not only a harbinger of victory on the battlefield; it also has important spiritual overtones. Whatever the challenges of the world, it is believed, these can only be successfully met if individuals are able to change themselves for the better through self-control and spiritual insight. The epic expresses this notion in the well-known imagery of the chariot: one must learn to control the chariot of the body by taming the wild horses of the senses with the reins of the intellect.

This mastery of the unruly senses through the intellect (*buddhi*) underlines the epic, and subsequent Hindu, idea that volition is not a separate act in itself, but follows automatically upon an act of spiritual insight. However, while *puruṣakāra* is an important element in this process it must not be confused with 'free will' or the modern concept of freedom. What is regarded as free is not the executive function of an empirical self (*jīvātman*) but the *puruṣa* or *ātman*, the self-realized transcendental Self (see Self, Indian theories of §1). The great problem of human life is precisely the bondage caused by the ego-centred self and its 'will'. Even full control of the chariot is not true freedom, since the very need for control implies that the unruly senses are still drawn by various desires to their worldly objects. The empirical self may identify with particular motives and will them to the extent that it casts its lot in with them. But in doing so, it effectively squanders its freedom, since it is precisely the identification of the *ātman* with these various dispositions and impulses that constitutes what the *Bhagavad Gītā* regards as the false notion of 'I' (*ahaṃkāra*), with its sense of agency.

These ideas can only be fully understood in the broader context of the ascent of life through many lives and life forms (exemplified in an epic story of a lowly worm that finally becomes Brahmā). Humanity is regarded as a sort

of bridge between the animals and the gods, the point at which instinct gives way to the need for self-control, and where consciousness becomes the possibility of knowledge of the Self. On the one hand, a human birth is the envy of the lower orders of being. On the other hand, the soul now finds itself in a difficult predicament. It no longer automatically follows the *dharma* of its species (like the animals), but it is also not yet able consciously to follow the imperatives of *dharma* as if they were its own (like God himself). Instead, it faces the arduous task of battling the atavistic forces of desire, aversion and anger that constantly seek to usurp the throne of the higher values and ideals that point the way to its own transcendence. The fratricidal war itself is a fitting metaphor for this human struggle on the 'field of *dharma*' between a lower nature experienced as the 'fate' of the gods or past karma as the case may be and a higher nature acting as a proxy for the spirit (Kṛṣṇa) that takes no active part in the conflict.

The protagonists face further complications over the many branches of *dharma* itself. Conflicts of values cause confusion and heartache, making it even more difficult to know what course to take. King Yudhiṣṭhira, the son of the god Dharma no less, is notably torn between the *dharma* of Brahmanical world-renunciation (*nivṛttidharma*) and the active *dharma* (*pravṛttidharma*) of the warrior king. Arjuna too, the main hero of the epic, is moved by caste and family values to avoid his moral obligation as a warrior chieftain to 'get up and fight' this war. It takes Kṛṣṇa himself to resolve this issue by transforming the traditional Brahmanical practice of renunciation into its inner form. Any undertaking, says Kṛṣṇa, including the most inhuman violence against one's own kith and kin, can be 'right' for the world and for the individual (that is, can be *dharma*), provided it is performed with no thought of personal gain (*karmayoga*), and in a spirit of devotion to Kṛṣṇa (*bhaktiyoga*). This suggests that *puruṣakāra* is no longer judged for its outward function or form, or even for the results to which it leads, but for the spirit that animates it from within.

3 Subsequent elaborations

The later commentaries generally treat the problem of fate in the context of discussions about *mokṣa* or spiritual freedom, the final goal of this spiritual journey. This is due to the fact that, psychologically, *mokṣa* involves a quantum shift in self-identity, in which the human ego, together with its sense of agency, is transcended in favour of a larger system of identity described in the *Bhagavad Gītā* as 'the self of the self of all beings'. Since *puruṣakāra* is based on ideas of 'I' and 'mine', it is fated to dissolve with the ego into the larger 'fate' of the will of God. This interplay between the individual and God raises the issue of the identity of the real agent of action. The consensus in the principal philosophical schools is that, ultimately, it is the Supreme Divinity (*puruṣottama*) who is the agent. Śaōkara and Rāmānuja both approach the issue in the context of their respective theories. Arguing from the relative standpoint in *Brahmasūtrabhāṣya* 2.3.42, Śaōkara reverts to the traditional mode of thinking that the soul (*jīva*) is driven by a god, in this case the personalized form of the Supreme Reality (*īśvara*) acting in accord with the previous efforts of the soul itself. In *Śrībhāṣya* 1.3.41, Rāmānuja too argues that the Lord is the source of all agency. However, being more conscious of preserving the integrity of the Vedic injunctions, he manages to salvage some responsibility for the individual by falling back on his theory of qualified difference. Since the embodied soul (*jīvātman*) is independent of the Supreme Self it can act on its own. However, the latter is still given the final word in 'granting permission' for the soul to act.

Later views on fate thus develop along the lines of subsequent religious beliefs. Ascetic traditions such as Vedānta accept self-effort as a practical fiction that dissolves with the advent of true knowledge, at which point one would no doubt view all of creation as moved by fate. However, in devotional sects for whom the permission, assistance or 'grace' of a personal God is called for, the devotee in some sense remains separate from God. In this case, the soul surrenders to the higher Divine Agency and itself becomes an instrument of fate. An example of this would be Kṛṣṇa's exhortation to Arjuna to be his chosen instrument of cosmic destruction.

In the last analysis, therefore, fate is all. In terms of the tasks and responsibilities of everyday life, humans may be regarded as enjoying a modicum of freedom to chart their own course of life, thereby modifying their individual fates. However, from the perspective of the higher freedom of *mokṣa*, this self-concept is inherently flawed, the sense of 'action' and freedom being nothing but fate in disguise.

See also: Karma and rebirth, Indian conceptions of; Predestination

JULIAN F. WOODS

References and further reading

No comprehensive analysis of Indian approaches to the problem of fate has been published to date. The issue is often raised in discussions concerning karma and related themes, or in more general works related to Indian mythology, the Dharmaśāstras or the epic literature; see, for example, Hopkins 1915 and O'Flaherty 1980.

Bedekar, V.M. (1961) 'The Doctrines of *Svabhāva* and *Kāla* in the *Mahābhārata* and other Old Sanskrit Works', *Journal of the University of Poona (Humanities)* 13: 17-28.(Discusses fate in the form of 'own nature' and time (*kāla*).)

Bhagavad Gītā (*c.*200 BC - AD 300-400), trans. W. Sargeant, ed. C. Chapple, Albany, NY: State University of New York Press, revised edn, 1984.(This text contains the Sanskrit with both literal and more readable English translations, plus grammatical notes on the Sanskrit vocabulary.)

Bharadwaj, S. (1992) *The Concept of 'Daiva' in the Mahābhārata*, Delhi: Nag Publishers.(A compendium of the concepts and the Sanskrit words used to convey the notion of fate in the *Mahābhārata*.)

Bhattacharyya, K. (1967) 'The Status of the Individual in Indian Metaphysics', in C.A. Moore with the assistance of A.V. Morris (ed.) *The Indian Mind: Essentials of Indian Philosophy and Culture*, Honolulu, HI: East-West Center Press and University of Hawaii Press, 299-319.(Indian views of the human being as the subject of fate.)

Bhattacharyya, K. (1971) 'The Indian Concept of Freedom', *Bulletin of the Ramakrishna Mission, Institute of Culture* 22 (9): 348-60.(Argues the case for determinism in Indian thought, with reference to Western and later Indian philosophical literature.)

De Smet, R.V. (1972) 'Early Trends in the Indian Understanding of Man', *Philosophy East and West* 22 (3): 259-68.(A brief but informative survey of Indian self-understanding from the early Vedic period to the rise of *bhakti* devotionalism.)

Divanji, P.C. (1946) '*Puruṣārtha*, *Daiva* and *Niyati*', parts 1-2, *Annals of the Bhandarkar Oriental Research Institute* 26: 142-51.(A discussion of the factors that influence the human capacity to pursue the traditional Indian goals of life.)

Hiriyanna, M. (1952) 'Karma and Free Will', in *Popular Essays in Indian Philosophy*, Mysore: Kavyalaya Publishers.(Discusses the relationship of karma to fate and free will.)

Hopkins, E.W. (1915) 'The Gods Collectively as Fate', in *Epic Mythology*, Strasbourg: Verlag von Karl J. Trübner.(A review of the various Sanskrit references to fate in the two major Indian epics, the *Mahābhārata* and the somewhat older *Rāmāyaṇa*.)

Hopkins, E.W. (1906-7) 'Modifications of the Karma Doctrine', *The Journal of The Royal Asiatic Society* 38: 581-93; 39: 397-401, 665-72. (Hopkins was one of the first to note the plethora of seemingly conflicting notions about fate, chance, luck, karma, grace, and so on in Indian traditions.)

Mahābhārata (*c.*200 BC - AD 300-400), trans. K.M. Ganguli, *The Mahābhārata of Krishna Dwaipayana Vyasa*, Calcutta: Oriental Publications Co., 2nd edn, 1952-62, 12 vols; Books 1-5 trans. J.A.B. van Buitenen, *The Mahābhārata*, Chicago, IL: University of Chicago Press, 1973-8, 3 vols.(Ganguli's translation, originally published in eighteen volumes in 1887-96, is the best complete English translation to convey the flavour of this Hindu epic; however, Buitenen's is the most faithful English rendition of the first five books.)

O'Flaherty, W.D. (ed.) (1980) *Karma and Rebirth in Classical Indian Traditions*, Berkeley, CA: University of California Press.(A useful collection of articles by well-known authorities on the subject of karma in Hinduism, Buddhism, Jainism and various Indian philosophical traditions.)

Rāmānuja (11th century) *Śrībhāṣya*, ed. and trans. R.D. Karmarkar, *Śrībhāṣya of Rāmānuja: Edited with a Complete English Translation, Introduction, Notes and Appendices*, Poona: University of Poona Sanskrit and Prakrit Series, 1964, 3 vols.(An English-language translation of Rāmānuja's commentary on the *Brahmasūtra* of Bādarāyaṇa.)

Śaōkara (early 8th century) *Brahmasūtrabhāṣya*, trans. Swami Gambhirananda, *Brahmasūtrabhāṣya of ŚrīŚaṃkarācārya*, Delhi: Advaita Ashrama, 3rd edn, 1977. (An English-language translation of the philosopher Śaōkara's commentary on the Brahmasūtra of Bādarāyaṇa.)

Sharma, A. (1980) 'Fate and Free Will in the *Bhagavadgītā*', *Religious Studies* 15: 531-7.(An attempt to reconcile the conflicting statements about fate and 'free will' in the *Bhagavad Gītā*.)

Walli, K. (1977) *Theory of Karma in Indian Thought*, Varanasi: Bharata Manisha.(A wide-ranging attempt to trace the development of the karma doctrine in Hindi, Buddhist and Jaina thought, based on the author's doctoral thesis at the University of Allahabad.)

Woods, J.F. (1988-9) 'The Doctrine of Karma in the *Bhagavadgītā*', *The Journal of Studies in the Bhagavadgītā* 8-9: 47-81.(An examination of the views of Śaōkara and Rāmānuja with respect to action (*karma, karmayoga*) in their respective commentaries on the *Bhagavad Gītā*.)

Fazang (643-712)

The monk-scholar Fazang is one of China's great Buddhist thinkers. Drawing on Buddhist scriptural literature and exegetical and systematic works of his predecessors, he fashioned a highly elaborate philosophy that served to provide a rational explanation of his vision of the way things really are. Huayan philosophy is an attempt to show rationally and systematically how the many phenomena that make up existence abide harmoniously in a double relationship of identity and interpenetration. Fazang was credited by his successors with being the third patriarch of the Huayan school.

Fazang was born in Changan (modern Xian), at that time the capital of China. He left home in his mid-teens to study Buddhism, and after several years met Zhiyan, who became the second Huayan patriarch. With Zhiyan he became acquainted with the *Huayan Sūtra* and its depiction of the cosmic identity and interpenetration of the many individual entities that constitute existence. Fazang became a monk in his mid-twenties and settled in a monastery in Changan as a scholar. He assisted Diksānanda in the latter's translation work, where he became intimately acquainted with the Mahāyāna Buddhist texts and doctrines that later would be cited in support of his portrayal of reality. Along with his translation work and his exegetical and commentarial writings, he composed several treatises that attempt to communicate the Huayan world view.

Fazang's more important works include the *Huayan wujiaozhang (Huayan Treatise on the Five Doctrines)*. Ostensibly a work on doctrinal classification, this text is the most important source of the complete Huayan system. He also composed the *Huayanjing tanxuanji (Notes on Searching for the Profundities of the Huayan Sūtra)*, an enormous phrase-by-phrase commentary on the sixty-volume version of the Huayan Sūtra. His *Huayan jinshizizhang (Huayan Essay on the Golden Lion)* is a demonstration of identity and interpenetration using the image of a golden lion as a device. The *Banluoboluomituo xinjing liushu (Brief Commentary on the Sūtra of the Heart of Prajñāpāramitā)* is a phrase-by-phrase commentary on the extremely influential Sūtra on the Heart of the Perfection of Wisdom. In the *Qixinlun yiji (Notes on the Meaning of the Treatise on the Awakening of Mahāyāna Faith)*, he provides a commentary on the enormously influential text mentioned in the title.

The Huayan philosophy is the result of Fazang's attempt to provide a reasoned, coherent description of what he believed to be the enlightened vision of the Buddha. This grasp of reality reveals a cosmos in which each individual particular is identical with all other particulars and at the same time interpenetrates with them. No thing exists in and of itself but rather is both the conditioned result of the many and a condition for all others. Consequently, the real cosmos is a place where all things of whatever nature are bound together intimately in a vast net of relationships.

This description of the true state of things is presented as the highest, final teaching of Mahāyāna Buddhism, and since Fazang's time, most Mahāyāna schools have accepted it as the high water mark of Buddhist thought. It was assembled from several parts that were central and foundational to Mahāyāna Buddhist thought and practice. One element is the doctrine of emptiness (*śūnyatā*) which, as a critique of things, is the denial that individuals exist autonomously by virtue of substances or essences. Primarily a critique of language but also secondarily a critique of the material world, 'emptiness' means that anything exists solely as the result of a vast array of conditions that generate and maintain the individual (see Buddhist concept of emptiness).

Another source of Huayan is the doctrine of the 'womb (or matrix) of Buddha' (*tathāgatagarbha*). In its simplest form, this is the doctrine that human beings are like wombs in which the seed or embryo of Buddhahood grows. In its most radical form, it teaches that everything, animate and inanimate, contains the Buddha and is thus both an ordinary, mundane individual and the body of the Buddha. Fazang also drew heavily on his predecessors in China, such as the Dilun School and the Shelun School, and on influential texts such as the *Awakening of Mahāyāna Faith* (see Buddhist philosophy, Chinese; Awakening of Faith in Mahāyāna).

Huayan thought should be seen as Fazang's attempt to unpack and develop the basic teaching of emptiness. If things are empty of independent existence, then if we examine any individual *x* in its relationship with the total environment, it will be seen that *x* arises out of a vast array of conditions and remains dependent on them. However, by the same token, because all things are simultaneously empty and existent, *x* in its capacity of being an

actual entity acts as a condition for the others. In this way, each item of existence is at once an actual fact of existence and empty. As conditions for others, all things retain their unique characteristics and qualities; as empty results of others, all things are identical. As a concrete individual with its own peculiar characteristics, the conditioning entity is simply one phenomenal item among myriads; as an instantiation of emptiness, it is the noumenous body of the Buddha Vairocana. The deluded see only the ordinary and phenomenal; the enlightened see the Buddha in each thing.

The Huayan cosmos is fundamentally and essentially a dynamic and creative place. Each individual is the focal point of vast number of causes and conditions that impinge on it and modify it from moment to moment. The ever-changing individual simultaneously acts creatively on its environment, for to be an existent being is necessarily to possess conditioning power. Such a situation has great implications for how the human individual grasps his or her own status and function in the world and for how one should act in a world where the fundamental fact is codependence.

See also: Buddhist philosophy, Chinese; Buddhist concept of emptiness

FRANCIS H. COOK

List of works

Fazang (643-712) *Huayan wujiaozhang (Huayan Treatise on the Five Doctrines)*, trans. F.H. Cook, 'Fa-tsang's Treatise on the Five Doctrines: An Annotated Translation', unpublished Ph.D. dissertation, University of Wisconsin, 1970; text in *Taishō Shinshū Daizōkyō* 45, no. 1866. (Fazang's most complete discussion of the intricacies of Huayan philosophy; he was the chief architect of Huayan thought.)

Fazang (643-712) *Huayanjing tanxuanji (Notes on Searching for the Profundities of the Huayan Sūtra)*, text in *Taishō Shinshū Daizōkyōō* 35, no. 1733. (An exhaustive phrase-by-phrase commentary on the *Huayan Sūtra* (Sanskrit title, *Avatamsaka Sūtra*, translated into Chinese in sixty volumes by Buddhabhadra). The *sūtra* or scripture is an important source for Huayan thought.)

Fazang (643-712) *Huayan jinshizizhang (Huayan Essay on the Golden Lion)*, text in *Taishō Shinsh Daizōkyō* 45, no. 1880. (Fazang's attempt to demonstrate the relationship of absolute and relative, and individual and totality, by the use of a small golden statue of a lion.)

Fazang (643-712) *Banluoboluomituo xinjing liushu (Brief Commentary on the Sūtra of the Heart of Prajñāpāramitā)*, text in *Taishō Shinshū Daizōyō* 33, no. 1712. (Fazang's Huayan interpretation of the *Prajñāāramitā Heart Sūtra*, a basic source for the Huayan teaching of emptiness as interdependence.)

Fazang (643-712) *Qixinlun yiji (Notes on the Meaning of the Treatise on the Awakening of Mahāyāna Faith)*, text in *Taishō Shinshū Daizōkyō* 44, no. 1846. (A discussion of the teachings of the *Awakening of Mahāyāna Faith*, a fertile source of ideas for Chinese innovators such as Fazang.)

References and further reading

Cook, F.H. (1977) *Hua-Yen Buddhism: The Jewel Net of Indra*, University Park: Pennsylvania State University Press. (An accessible interpretive study of key philosophical ideas.)

Liu Ming-wood (1979) 'The Teaching of Fa-tsang: An Examination of Buddhist Metaphysics', unpublished Ph.D. dissertation, University of California, Los Angeles. (The only Western-language critical study of Huayan metaphysics; in the course of commenting critically on the Huayan system, the author provides the reader with a thorough understanding of difficult Huayan concepts.)

Fechner, Gustav Theodor (1801-87)

Fechner was a pioneer in experimental psychology and the founder of psychophysics, the speciality within psychology devoted to quantitative studies of perception. In his foundational Elemente der Psychophysik (Elements of Psychophysics) (1860), he defined the mission of the new science to be the development of an 'exact theory of the functionally dependent relations of... the physical and the psychological worlds'. It is in this work that Fechner developed the law of sensation-magnitudes (Fechner's Law): the strength of a sensation is proportional to the logarithmic value of the intensity of the stimulus. Among his contemporaries he was well known not only for basic research in the field of electricity, but also as the author of a number of satirical works under the name 'Dr Mises'.

1 Life

Fechner was born in Saxony, in the village of Grossärchen, son and grandson of Protestant clergymen. Seventy years of his life were spent in Leipzig, many of them at the University, first as a medical student, then as an instructor, and finally as a professor.

Although trained in medicine, Fechner first distinguished himself in physics. While still a student he published translations of Biot's *Textbook of Physics* and Thénard's *Textbook of Theoretical and Practical Chemistry*. Under the pseudonym 'Dr Mises' he also published satirical and poetical works, the first appearing in 1821 as a *Proof That the Moon Is Made of Iodine*. It was his research on electrical currents that earned him a salaried ('Ordinarius') professorship when he was 33. But later experiments on visual after-images resulted in temporary blindness which, coupled with a harrowing nervous exhaustion, resulted in his resignation from the Leipzig faculty in 1840. A period of exile and depression followed for several years, whereupon recovered vision and restored strength permitted him to resume a productive life of research and writing.

2 Fechner's law

Elemente der Psychophysik (*Elements of Psychophysics*) (1860) is at once a defence of scientific approaches to the study of mind and a detailed treatise on the proper methods of experimental research in perception. Moreover, it contains the mathematical rationale for the derivation of what might be regarded as the first general law of perception, 'Fechner's Law', which declares the magnitude of sensations to grow as a logarithmic function of the intensity of physical stimuli.

The starting point for this derivation had been provided by the research of E.H. Weber, under whom Fechner had studied physiology and mathematics at Leipzig. Weber had shown that a constant ratio obtained between a stimulus and the amount by which it must be increased or diminished for the change to be just perceptible. Thus, a standard weight of, say, ten units might have to be increased by two units to be judged as 'just heavier'. In this case, on average, a weight of twenty units would have to be increased by four; one of thirty units by six. The general law, which Fechner named 'Weber's law', is expressed, '$\Delta S/S = C$', where *S* refers to the magnitude of the standard stimulus, ΔS to the difference or change in the magnitude of the standard stimulus sufficient to be just perceptible, and *C* to the constant fraction discovered for a given class of stimuli. It should be noted that this constant ratio obtains where the standard and the comparison stimuli are just noticeably different. Thus, for any value of the standard stimulus, the amount by which a comparison must differ to result in a just-noticeable difference (the 'jnd') is a constant fraction of the magnitude of the standard stimulus.

Weber's law is a law of discrimination, not a law relating the physical intensity of stimuli to the magnitude of the resulting sensations. To derive the latter from the former Fechner adopted a number of assumptions, chief among them that a given sensation represents (or may be taken as representing) the summation of an indefinitely large number of just-noticeable differences. He reasoned that the law of sensation-magnitude could be found by mathematical integration. Thus, by integrating the Weber ratio over all values of *S*, Fechner derived the general expression relating sensation-magnitude (*R*) to stimulus-magnitude (*S*): '$R = k \log_{\mathrm{e}} S$'. Transforming this to the base-10 system through the appropriate constant then yields '$R = k \log S$', the familiar form of Fechner's law.

Fechner developed the central ideas of his *Elements of Psychophysics* in a number of subsequent publications and extended the general perspective to the domain of aesthetics in his *Vorschule der Aesthetik (Propaedeutic to*

Aesthetics) (1876). His defence of experimental approaches to the study of mental processes and his development of methods useful for the purpose place him among the most influential figures in the history of psychology.

3 Panpsychism

Although the psychophysical research and theory were rich in mathematical and experimental content, their underlying rationale was drawn from cosmological and metaphysical conceptions somewhat misleadingly referred to asFechner's 'panpsychism'. In a number of books and articles Fechner defended the thesis that the cosmos itself has a life - an inner life not unlike human consciousness and not to be confused with its mere externals. The parallel is clear between this view and his distinction between 'inner' and 'outer' psychophysics. Inner psychophysics refers to the inner dynamics of the relations between the psychic and the physiological; outer psychophysics to the relations between stimulation and immediate experience. So too is the balance of nature impelled by animating principles not discernible in the mere outward appearances of things. A relentlessly materialistic science is doomed to ignore such principles and thus to arrive at a conception of nature that is literally meaningless (see Panpsychism).

In *Über die Seelenfrage* (*On the Soul*) (1861), Fechner criticized the conventional scientific perspective that divorced the earth from the diverse lives that arise from and return to it, rather than recognizing all living things as aspects of the earth's own inward development. A decade earlier in *Zend-Avesta* he had argued for the view that consciousness permeates all, but in a manner not accessible to prosaic modes of experience. It was only when he could once again look without pain at the flowers in his garden that he conceived of *Nanna, oder Über das seelen-leben der pflanzen* (*Nanna, The Soul Life of Plants*) (1848) in which he attributed the beauty of nature to inner principles of an irreducibly psychic sort. As he said in the conclusion of this work, 'Indeed, one will hardly believe how new and vivid is the nature which meets the man who himself comes to meet it with new eyes'.

See also: Psychology, theories of §§1-2

DANIEL N. ROBINSON

List of works

Fechner, G.T. (1836) *Das Büchlein vom Leben nach dem Tode*, Leipzig: Insel-verlag; trans. M. Wadsworth, *Life after Death*, New York: Pantheon, 1943.(Fechner's little book of eschatological theories emphasizes the indestructibility of the psychic.)

Fechner, G.T. (1848) *Nanna, oder Über das seelen-leben der pflanzen*, Leipzig: Voss; trans. W. Lowrie, *Nanna, the Soul-Life of Plants*, in *Religion of a Scientist: Selections from Gustav Th. Fechner*, New York: Pantheon, 1946.(Fechner again emphasizes the life-principle of creation, absorbing both plant and animal forms into an organic and creative cosmos.)

Fechner, G.T. (1860) *Elemente der Psychophysik*, Leipzig: Breitkopf & Hartel; trans. H. Adler, *Elements of Psychophysics*, New York: Holt, Rinehart & Winston, 1960.(Fechner's major contribution to experimental psychology in which a precise methodology is developed and its rationale defended productively and with rigour. It is one of the classical works in the history of psychological thought. Chapter 7 contains Fechner's mathematical rationale for his law. A list of 175 of Fechner's publications, including those under the name 'Dr Mises', was compiled by R. Müller and furnished by Wilhelm Wundt in the third German edition.)

Fechner, G.T. (1861) *Über die Seelenfrage*, Leipzig: C.F. Amelang; trans. W. Lowrie, *On the Soul*, in *Religion of a Scientist: Selections from Gustav Th. Fechner*, New York: Pantheon, 1946.(The tension between science and religion is acknowledged but Fechner insists that it is only through a consideration of spirit that the material transactions of the physical world become intelligible.)

Fechner, G.T. (1863) *Die drei Motive und Gründe des Glaubens*, Leipzig: Breitkopf & Hartel; trans. W. Lowrie, *The Three Motives and Grounds of Faith*, in *Religion of a Scientist: Selections from Gustav Th. Fechner*, New York: Pantheon, 1946.(More on 'the religion of a scientist'.)

Fechner, G.T. (1876) *Vorschule der Aesthetik (Propaedeutic to Aesthetics)*, Leipzig: Breitkopf & Hartel. (Fechner's place in the history of experimental aesthetics begins with this work, but as the speciality itself has progressed little beyond where Fechner left it, this remains something of a promissory history.)

Fechner, G.T. (1906) *Zend-Avesta, oder Über die Dinge des Himmels und des Jenseits (Zend-Avesta, or that Concerning Matters of Heaven and the Future Life)*, collected in 2 vols by K. Lasswitz, Leipzig: L. Voss, 3rd edn.(The collection in these volumes offers Fechner's views on religion and philosophy, on immortality and the divine, with acknowledged debts to Eastern religion and thought.)

References and further reading

Boring, E.G. (1950) *A History of Experimental Psychology*, New York: Appleton, Century, 2nd edn.(Boring's chapter on psychophysics and the debt of experimental psychology to Fechner is nontechnical and historically accurate.)

James, W. (1909) *A Pluralistic Universe*, New York: Longmans, Green.(Chapter 7 of this work is a retrospective appreciation of Fechner.)

Federalism and confederalism

Federative arrangements involve two or more governments ruling over the same territory and population. They have been of interest to political philosophers because they challenge, or at least complicate, some fundamental political concepts like authority, sovereignty, democracy and citizenship. Like citizens in actual federations, philosophers do not treat the terms of federation as a merely technocratic matter: they believe that there are morally legitimate and illegitimate ways of, among other things, dividing powers between governments, determining the representation of the subunits (for example, provinces) within federal institutions and amending the constitution. Philosophers also see in federalism a means of securing a degree of self-determination for ethnic minorities who cannot realistically expect to have their own homogeneous nation-states.

Most writers call a political arrangement federal if it has: (1) two or more levels or orders of government ruling over the same territory and citizens; (2) a division of legislative or administrative powers between these orders of government, with at least some areas of exclusive responsibility at each level; and (3) a constitutional guarantee that this division of powers cannot be changed without the agreement of governments of all orders. Contemporary federal theorists usually characterize the implications of these three conditions in a way that most philosophers prior to the twentieth century would have thought logically impossible. The combination of (1) and (2) implies that citizens of federal states are members of more than one political community and hence have a special kind of multiple citizenship; (2) and (3) imply that sovereignty in federations is divided.

An arrangement is likely to be called confederal rather than federal if effective sovereignty rests with the constituent units. In practice this is the case if the governments of the member states can unilaterally alter the terms of the arrangement, including the division of powers, or if member states can unilaterally secede from the union.

The development of federal and confederal theories in the modern era has always been closely linked to actual experiences with these sorts of political arrangements. Just before the predominance of the unitary nation-state there was widespread experimentation with leagues and confederations, including the Holy Roman Empire (800-1806), the Hanseatic League in Germany (1158-1669), the Swiss Confederation (1291-1848) and the United Provinces in Holland (1567-1798). Within these territories theorists attempted to work out a political and legal philosophy of confederation. So parallel to what was to become the mainstream tradition of political philosophy for the unitary nation-state, which begins with Bodin and Hobbes, is another philosophy which is much more receptive to federative arrangements and to what we would now call cultural and political pluralism. It starts with Althusius and Pufendorf in the seventeenth century and runs through the Abbé de Saint Pierre, Montesquieu, Rousseau, Kant, Madison and Hamilton, Tocqueville, Proudhon and Carl Schmitt, among others.

From Althusius to Kant, confederal arrangements were advocated primarily as a means of achieving international order and peace. What they contemplated in various forms was a permanent association of independent sovereign states that would delegate limited powers to a common council. Philosophical debates generally focused on three issues. The first concerned whether confederations violated the supposed indivisibility of sovereignty (see Sovereignty). Following Pufendorf, most theorists employed a distinction between sovereignty-in-itself, which remained indivisible, and the 'marks' of sovereignty - such as judicial and legislative power and powers of peace and war - which could be divided and shared within confederations. Second, there were debates about which forms of international cooperation were necessary and sufficient to realize the idea of international law. Kant's own views on this question evolved considerably in the 1780s and 1790s: as he became more optimistic about the possibility of international law for states sending delegates to a peace congress, he came to see confederations, and in particular a worldwide confederal state, as less essential for the preservation of peace. Third, there were disputes about what kinds of states could enter into the same confederation. Early writers found no objections to unions involving a mixture of feudal and republican states. But from Montesquieu and Rousseau onwards, most republicans insisted on the need for common forms of government. They saw confederation as a necessary feature of republicanism, since to be just and effective republics must be small; yet being small, they have to band together to defend themselves against larger enemies (see Republicanism). An analogous argument for the desirability of smaller political units within a larger economic union is still popular today.

The idea of federalism, as outlined above, was not seriously developed by anyone, even in theory, until the

American Confederation began to founder shortly after the thirteen colonies' seceded from Britain. Its constitution was replaced in 1789 by one setting up the first modern federation. In their brilliant defence of the new constitution, Madison, Jay and Hamilton (Hamilton *et al.* 1787-8) followed continental republicans in stressing the possibility of achieving within a federation the advantages enjoyed by both large states (security and economic strength) and small states (liberty and active democracy), without the usual disadvantages of either. Unlike Montesquieu and Rousseau, however, they paid careful attention to institutional design and argued for a strong central government responsible to a nation of equal citizens as well as to a union of states. They persuaded Tocqueville and J.S. Mill as well as the Swiss, who reconstructed their 500-year-old confederation on federal lines in 1848. Successful constitutional designs in the dozens of federations born since the late nineteenth century have all been guided by the idea of finding a division of powers and a system of federal representation that will protect diversity while encouraging a panfederal patriotism and stability (see Constitutionalism).

Concern for stability in a federation tends to introduce a number of normative considerations because in democracies political arrangements will be unstable if perceived to be unfair or illegitimate. Regional groups find an arrangement unjust if, for example, it imposes disproportionate economic burdens on them; denies them adequate representation in central institutions, such as the parliament or the supreme court; does not give them sufficient control over matters of cultural importance; or fails to recognize in an appropriate way their existence as distinct national communities. Many of these concerns involve group rights and identity politics - issues largely absent from federal debates before the twentieth century, and largely neglected by political philosophers until recently.

See also: Multiculturalism

WAYNE NORMAN

References and further reading

Carney, F.S. (ed.) (1964) *The Politics of Johannes Althusius*, Boston, MA: Beacon Press.(Includes a translation of the third edition of Althusius's *Politica* from 1614.)

Forsyth, M. (1981) *Unions of States: The Theory and Practice of Confederation*, Leicester: Leicester University Press.(An excellent and accessible introduction to the history of federal and confederal thought.)

Friedrich, C.J. (1968) *Trends of Federalism in Theory and Practice*, New York: Praeger.(Includes a short, useful survey of the history of federal and confederal thought.)

Hamilton, A., Madison, J. and Jay, J. (1787-8) *The Federalist Papers*, New York: New American Library, 1961. (See papers 3-5, 9, 14 and 39 for some of their more explicit discussions of the classic republican case for federalism over confederalism.)

Mill, J.S. (1863) *Considerations on Representative Government*, New York: Dutton, 1904.(See chapter 17, on federal government.)

Montesquieu, C.-L. de S. (1748) *The Spirit of Laws*, trans. A. Cohler, B.C. Miller and H.S. Stone, Cambridge: Cambridge University Press, 1989.(Discusses republican confederation in book IX, chapters 1-3.)

Norman, W. (1994) 'Towards a Philosophy of Federalism' in J. Baker (ed.) *Group Rights*, Toronto, Ont.: University of Toronto Press.(An exploration of criteria relevant to the design and evaluation of federal constitutions.)

Pufendorf, S. (1672) *The Law of Nature and Nations*, trans. C.H. Oldfather and W.A. Oldfather, Oxford: Clarendon Press, 1934.(See books V1 and VII for discussion of confederal structures and divisions of powers, as well as the challenges these raise for the concept of sovereignty.)

Reiss, H. (ed.) (1991) *Kant's Political Writings*, Cambridge: Cambridge University Press, 2nd edn. (Kant discusses confederal arrangements and international law in *Perpetual Peace, Idea of a Universal History*, the essay on *Theory and Practice* and in *Metaphysical Elements of Justice*.)

Riker, W.H. (1964) *Federalism: Origin, Operation, Significance*, Boston, MA: Little, Brown & Company.(The definition of 'federalism', above, is adapted from this philosophically literate, if somewhat dated, introduction to the idea and practice of federalism.)

Rousseau, J.-J. (1761) *Abstract of the Abbé de Saint Pierre's Project for Perpetual Peace*, in M. Forsyth *et al.* (eds) *The Theory of International Relations*, London: Allen, 1970.(Rousseau's most extensive discussion of confederation, apart from a manuscript on the subject which was apparently destroyed during the French Revolution.)

Tocqueville, A. de (1835-40) *Democracy in America*, trans. H. Reeve, New York: Knopf, 1966.(See especially volume I, chapter 8.)

Vernon, R. (1987) 'The Federal Citizen', in R.D. Olling and M. Westmacott (eds) *Perspectives on Canadian Federalism*, Scarborough, Ont.: Prentice Hall.(A provocative analysis of the way federalism complicates the notion of citizenship.)

Feminism

Feminism is grounded on the belief that women are oppressed or disadvantaged by comparison with men, and that their oppression is in some way illegitimate or unjustified. Under the umbrella of this general characterization there are, however, many interpretations of women and their oppression, so that it is a mistake to think of feminism as a single philosophical doctrine, or as implying an agreed political programme. Just as there are diverse images of liberation, so there are a number of feminist philosophies, yoked together not so much by their particular claims or prescriptions as by their interest in a common theme.

In the earlier phases of feminism, advocates focused largely on the reform of women's social position, arguing that they should have access to education, work or civil rights. During the latter half of the twentieth century, however, feminists have become increasingly interested in the variety of social practices (including theoretical ones) through which our understandings of femininity and masculinity are created and maintained. As a result, the scope of feminist enquiry has broadened to include, for example, jurisprudence, epistemology and psychoanalysis.

Despite its diversity, this work characteristically draws on and grapples with a set of deeply-rooted historical attempts to explain the domination of women. Aristotle's claim that they are mutilated males, together with the biblical account of the sin of Eve, gave rise to an authoritative tradition in which the weakness, irrationality and ineducability of women, the inconstancy, inability to control their emotions and lack of moral virtue, were all regularly cited and assumed as grounds for controlling them and excluding them from the public realm.

1 Feminism and feminisms

While there have been, throughout the history of philosophy, writers who challenge the sexual stereotype and offer different pictures of women, their works do not contribute to a single story. It can therefore be misleading to assimilate them too quickly to the philosophical literature and political campaigns which initiated later feminist movements, or to contemporary feminist positions. Only at the end of the eighteenth century did a stream of philosophical arguments aimed at the emancipation of women begin to gather force. Only at the end of the nineteenth century did the term *la féminisme* appear, put into circulation after the fact in France during the 1890s, and rapidly taken up in the rest of Europe and then in America. The label 'feminist' thus arose out of, and was in many ways continuous with, the sequence of diverse campaigns for female emancipation fought throughout the nineteenth century - campaigns for the vote, for access to education and the professions, for the right of married women to own property and have custody of their children, for the abolition of laws about female prostitution which assumed the double standard. While the character and success of these movements varied from country to country (for example, women's suffrage was introduced in New Zealand in 1893, Finland in 1906 and Britain in 1918) they all drew upon, and generated, arguments about the nature and capacities of women and the character of their oppression, and entertained, explicitly or implicitly, images of what a better condition would be like. Many of the most influential philosophical defences of women's emancipation dating from this period were in fact written by people involved in political work - to name only two, John Stuart Mill, the author of *The Subjection of Women*, proposed to the British parliament in 1867 an amendment to the Reform Bill designed to give votes to women, while Emily Davies, author of *The Higher Education of Women*, was the foundress of Girton College, Cambridge, the first women's college of higher education in England.

We have no difficulty in retrospectively classifying works such as these as feminist, although this is not a description their authors would have used, because they contain analyses of women's oppression and proposals for overcoming it which mesh easily with analyses and proposals later regarded as central to the feminist cause. However, there are also significant divergences between feminist writers, past as well as present. Different interpretations of the disadvantages to which women are subject, allied to different conceptions of what would constitute an improvement, gave rise to distinctive and sometimes irreconcilable feminisms. Compare, for example, the broadly liberal view that the oppression of women consists in their lack of political equality with men and can be alleviated by giving women and men the same political rights, with the separatist view that women's oppression lies principally in their sexual subordination to men and can only be overcome in societies that are, as far as possible, exclusively female. As these examples indicate, there are many feminisms, each with a history of its own.

Historians are bound to select their material in the light of the kind or kinds of feminism that concern them, and to work with interpretations that are used to distinguish texts and movements that qualify as feminist from those that are merely about women. To pursue the examples already discussed, historians whose interest in feminism focuses on the quest for equality between the sexes may identify certain writers as feminists *avant la lettre*. For example, they may include in their canon Poulain de la Barre, author of *De L'Égalité des deux sexes* (1673), or Mary Wollstonecraft. By contrast, a history of feminism understood as the quest for a separate society of women is more likely to pick out Mary Astell's proposal that ladies should retire from the society of men who debar them from realizing the natural desire to advance and perfect their being, or Charlotte Perkins Gilman's utopia *Herland* about an isolated society of women who are able to have children without male assistance. As long as there is more than one interpretation of feminism, feminism will not have a unified history.

2 Renaissance and early-modern forerunners

Although female inferiority is the dominant note that sounds through the Western philosophical tradition, its character was never a matter of consensus. Long drawn out theological debates about whether woman is a human being, whether she is made in the image of God, whether she is a perfect creation of God or an imperfect version of man, and whether men and women are equal before God, all appeal to classical authorities, to the Bible and to the Church Fathers, and rumble on through the Middle Ages and into the Renaissance. Complementing them are a series of more secular discussions, of which one of the most consistent focuses on women's learning. In her *Livre de la cité des dames* (1405) Christine de Pizan extols the advantages of educating women, a theme subsequently taken up by Renaissance writers for whom it played a part in the so-called *Querelle des Femmes* - a sequence of philosophically repetitive, interwoven disputes about whether fidelity in marriage should be demanded of both sexes, whether and to what extent women should be educated, and whether women were entitled to the respect and gratitude of men for the services they rendered them. On one side of these debates, women's inferiority was reasserted by appeal to example, authority and reason. On the other, their superiority was defended in a variety of genres. Some writers - for example Cornelius Agrippa - employed the rhetorical device of the paradoxical encomium, attempting to entertain and impress by ingeniously reversing conventional evaluations of men and women (see Agrippa von Nettesheim). Others drew on a well-tried stock of cases to illustrate women's superior virtue, intelligence or judiciousness. The choice of these genres contributes to the impression that, while these champions of women sometimes propose certain social reforms, they are on the whole anxious not to unsettle the status quo. Their aim is to entertain - to tease men and flatter women, and perhaps in doing so to make both reconsider their roles - rather than to foment social change.

Traces of this style endured well into the seventeenth century and are visible even in writers who in other ways broke with the terms of the *querelle*. A particularly striking change is the move away from debates about the relative inferiority or superiority of women, to works purporting to show that the sexes are equal. Marie de Gournay, who claimed that she was the first to take this view, published her *Égalité des hommes et des femmes* in 1622, and the same theme was taken up with a new determination later in the century. In France, Poulain de la Barre adopted a fresh approach when he appealed to Cartesian scientific method: a clear and distinct understanding of the issue can be arrived at, he insists, by rational demonstration. Although his *De l'égalité des deux sexes* (1673) sometimes lapses into the older style of argument - women are more decorous and discreet than men, women's work is more valuable than that of men, and so on - Poulain is remarkable for the forthright manner with which he asserts that the relations between mind and body and the capacities of the mind are the same in both sexes, and even more for the consequences he draws from this claim. There is no reason, in his view, why women should not occupy all the public roles currently held by men. Since they are capable of equalling men in understanding all the sciences (including both civil and canon law) they could, if educated, teach in the universities, be legislators, rulers, generals of armies, judges and - most radical of all - preachers and ministers of the Church.

Poulain's willingness to contemplate such dramatic social change is unusual, but less so is his emphasis on intellectual equality and downplaying of the significance of the bodily differences between men and women. This is shared by a number of women writing in the second half of the seventeenth century, who criticize men for depriving them of learning and education, and imply that women are quite capable of ruling themselves, and indeed men. In the Netherlands, Anna Maria von Schurman writes in favour of the education of girls. In England, women such as Margaret Cavendish and Mary Astell are by turns bitter and witty in their undermining of women's subjection to men.

3 Claims of right

As early as 1673, Poulain de la Barre argued that women and men possess an equal right to knowledge, conferred on them by nature. All humans pursue happiness; no one can achieve happiness without knowledge; so everyone needs knowledge. To ensure that people are able to pursue their proper end, nature has supplied the necessary means in the form of a right. We find here the beginnings of an appeal to rights which became progressively more central until, a century or so later, it dominated debate. In the immediate wake of the French Revolution, Olymphe de Gouges presented the French Assembly with a Declaration of the Rights of Women (which it declined to ratify). Women, she argued, should have rights to employment, legal rights within the family, a right to free speech and a separate assembly in which they could represent themselves. The same theme was taken up in England by Mary Wollstonecraft, who in 1792 published *A Vindication of the Rights of Women*. Challenging Rousseau, Wollstonecraft argued that the education and emancipation of women are conditions of a truly civilized society. God has endowed all humans with reason so that they can use it to govern their passions and attain knowledge and virtue. To deprive women of the opportunity to perfect their nature and increase their capacity for happiness is to treat them as less than human and render them ‘gentle, domestic brutes’. It is to trample on their rights and keep them in a state of subjection which damages both them and their male captors.

Far from being natural, Wollstonecraft explains echoing the arguments of Mary Astell, the presumed inferiority of women stems primarily from their lack of education. Cut off from learning and encouraged to care only for love and fashion, they are unable to cultivate any solid virtues, and do indeed display the flightiness and stupidity for which they are criticized. However, as well as damaging themselves, women in this condition diminish others. First, they damage men. To treat a fellow human despotically shows a lack of virtue, and just as kings are corrupted by their excessive power, so men are corrupted by the tyranny they exercise over their sisters, daughters and wives. Second, ignorant and powerless women are unfitted to instil virtue into their children. ‘To be a good mother **-** a woman must have sense and that independence of mind which few women possess who are taught to depend entirely on their husbands’ (Wollstonecraft [1792] 1995: 243).

Although Wollstonecraft’s argument hinges on her claim that women are as rational as men, she has no sympathy for what she calls ‘masculine women’. The aim of educating women is, in her view, to make them into virtuous wives and mothers who, by fulfilling these natural duties, will become useful members of society. Freed from male subjection, educated women would not usurp the roles of men but would freely and virtuously pursue their domestic lives to the benefit of society as a whole. The claim that men and women are intellectual equals is here allied to the view that there are natural differences between them which fit them for distinct ways of life: rational women will see that their place is in the home.

This easy division of labour was put under increasing pressure during the nineteenth century, as feminist thinking became less concerned with women’s overarching moral right to liberty and focused instead on particular legal entitlements such as the right to own property, to enter the professions, and above all to vote. Nevertheless, arguments which appeal simultaneously to the equality and difference of the sexes, and sustain the view that women excel in certain domestic virtues, remained common. The US suffragist, Elizabeth Cady Stanton, demanded the vote for women from the New York legislature during the 1850s on the grounds that ‘the rights of every human being are the same and identical’. But she also argued that, if women were able to represent themselves by voting, they would make a distinctive contribution which would balance that of men.

The wish to reconcile the demands of equality and difference is also evident in John Stuart Mill’s *The Subjection of Women* (1869). Mill argues that women are entitled to the same rights as men and should be able to hold public office, to work, to own property and to vote. He also argues that married women should not be required to obey their husbands and should have custody over their children. His primary ground for these conclusions is that women and men are equal, but he supplements this argument with further claims about the benefits that the freedom of women would bring. Like Wollstonecraft, he claims that the power of men over women ‘perverts the whole manner of existence of the man, both as an individual and a social being’, and reiterates her view that there can be no true affection between spouses who have nothing in common. It is only once women are educated that there can be the solid, enduring friendship between the sexes that heralds the moral regeneration of mankind. However, two further lines of thought appeal to assumed differences. Mill first argues that women possess a distinctive aversion to war and addiction to philanthropy, of which they would make better use if they were better

informed. In addition, although women should have the right to work, Mill takes it that when they marry they make ‘a choice of the management of a household, and the bringing up of a family’ as the first call on their exertions. Older women who have completed this task may decide to direct their energies to public life, for instance by standing for parliament. But the first place of married women is, once again, in the home.

4 Sexual oppression and emancipation

The view that the oppression of women could be overcome once they had the same rights as men was therefore compatible with a conventional understanding of the division of male and female labour. But doubt was cast on this whole approach to emancipation by the fact that, once the vote was won, women did not on the whole use their new-found political power to press for further reform. Many suffragists were keenly disappointed, and feminists of more radical political persuasions were strengthened in their conviction that the source of women’s oppression did not lie in their lack of political rights. Reforms such as the married women’s property act and the right to higher education, they pointed out, benefited middle-class more than working women. More important still, the root of women’s subordination lay not in their civic but in their private lives - in their roles as wives and mothers.

This latter view was partly derived from Engels’ *Origins of the Family, Private Property and the State* (1884) in which he argued that women’s oppression is primarily sexual. There is nothing natural about the patriarchal family. Rather, this institution came into existence at a particular point in history together with private property. To be able to hand down their property to their sons, men needed complete sexual possession of the mothers of their children, and to this end they reduced women to servitude. In capitalist society, women’s subjection consists not in their lack of legal rights, but in their weak position in the labour market which in turn forces them into marriage. Women face a choice between lives of near-destitution as workers or lives of slavery as wives and mothers, or in the case of working class women, both exploitation at work and subjection in marriage. Only once capitalism is overthrown will they escape this plight and be freed from dependence.

In Russia, the predicament diagnosed by Engels was confronted by the revolutionary Alexandra Kollontai (1872-1952), who insisted in *The Social Basis of the Woman Question* (1909) that proletarian women must refuse to cooperate with the bourgeois feminist movement and attack capitalism, the source of their oppression. As Commissar of Social Welfare in the Russian Revolutionary Government of 1917, Kollontai oversaw the drafting of far-reaching legal reforms designed to revolutionize the family and sexual relations between men and women, and to relieve women of the ‘triple load’ of wage worker, housekeeper and mother. These reforms were organized around a distinction between productive and non-productive labour, and were based on the view that women should be relieved of the burden of non-productive domestic labour (cleaning, cooking, washing, caring for clothes, and many aspects of child-rearing) to engage in productive labour alongside men. In this way they would achieve economic independence. At the same time, women’s work was to take account of their productive childbearing role. The work of carrying and bringing up children was no longer to be seen as the responsibility of individual families but as a task for the state, since it was in the interest of the workers’ collective that children should be born and that they should grow up to be able-bodied and good revolutionaries.

In the early years of the Bolshevik government, Kollontai began to implement a series of radical though short-lived reforms. Women were to have full civil rights; civil marriage and divorce laws were introduced; legitimate and illegitimate children were to have the same legal rights; and in 1920 abortion was legalized. As far as labour was concerned, women’s work was to take account of childbearing. They were not to do heavy work which might damage their health or work long hours or night shifts. They were to have paid maternity leave and health care during pregnancy. As for their children, once out of infancy they were to be cared for in crèches, kindergartens and schools which would also provide meals and clothing.

According to Kollontai, the dictatorship of the proletariat will abolish the family and with it bourgeois sexual morality. For though the state must, in her view, concern itself with children, it did not have any more extended interest in the relations between their parents. Conventional notions of romantic love must not undermine comradeship; yet Kollontai stresses that solidarity can only exist between those who are capable of love and sympathy, and envisages a society in which people are emotionally educated to feel many forms of love for different people.

In the USA, Engels’ view that women’s oppression is rooted in the family was used by the anarchist, Emma

Goldman (1869-1940), to ground a different set of conclusions. Access to education and work, for which emancipationists had fought so hard, produced women who were ‘professional automatons’ and lacked ‘the essence that enriches the soul’. By entering the public sphere, women had joined an impure State which prevents both women and men from developing the inner qualities that spring from sexual intimacy and constitute freedom, but is particularly distorting for women, for whom love is even more important than it is for men. The question of how to become free is therefore a question about how to foster sexual self-expression, and Goldman is adamant that this can only happen once women cease to be the sexual possessions of their husbands. As well as eschewing the public sphere, women must reject the private institution of marriage in which, driven by economic need, they purchase financial security at the price of their independence. They must learn instead to recognize and follow what Goldman calls their instinct.

Goldman and Kollontai share with some of their liberal forebears and contemporaries the premise that an institution of marriage in which women are sexually dominated by, and economically dependent on, their husbands, makes them unfree. More radically, both claim that these evils can only be overcome by sweeping away conventional notions of marriage and family. Beyond this, however, they diverge sharply. For Kollontai, liberty consists in productive labour in which both women and men must engage if they are to be equal and equally free. In the case of women, however, productive labour can take the distinctive form of bearing children. Motherhood (women’s difference), is subordinated to an overall conception of equality according to which men and women are not treated in the same way, but make the same kind of contribution by working productively. Goldman, by contrast, conceives freedom as a state of individual exploration and self-expression which needs to be pursued outside the impurity and corruption of the State and has little to do with work. Both men and women need love in order to become free, but for women, sexual intimacy plays a particularly important part in this process. While Kollontai separates reproductive sex from other erotic relations, Goldman tips the balance away from motherhood. Unconstrained love, which may or may not be the love of mothers for their children, is what enables women to fulfil themselves and become free.

5 The pervasiveness of male domination

It has become customary to distinguish a first wave of feminism, dating from the mid nineteenth century to the 1930s, from a second wave, breaking in the 1970s. This chronology is designed to highlight the absence of specifically feminist political campaigns in the intervening period, but is misleading if applied more generally, since one of the most influential works of modern feminist philosophy, Simone de Beauvoir’s *The Second Sex*, was published in 1949 (see Beauvoir, S. de). Dissatisfied with existing accounts of women’s subordination to men, Beauvoir confronted the question ‘What is Woman?’ by exploring the limited answers offered by historical materialism and psychoanalysis. Both these theories, she claims, beg the question. In *The Origin of the Family*, Engels asserts that the institution of private property results in the enslavement of women, but offers no means of explaining why this should have been so. Equally, Freud’s account of sexual differentiation fails to say what previous evaluation of virility makes boys proud of their penises and makes girls attribute special significance to their lack of this bodily part. To explain women’s oppression, in which women themselves are complicitous, it is not enough to appeal merely to economic categories or patterns of psychological development already imbued with the evaluations that constitute male power. What is needed is a theory grounded on dynamics of consciousness running deeper than physiological, psychological or economic forces, capable of doing justice to the vast variety of practices that contribute to women’s subordination.

Beauvoir derives her analysis of woman from the existentialist view of consciousness articulated by Sartre in *Being and Nothingness* (see Sartre, J.-P.). Each consciousness faces the world alone, and must create itself through its own choices by responding to the things around it, whether these are passive natural objects or other consciousnesses. Adapting Hegel’s account of the relations between master and slave, Sartre portrays the meeting of one consciousness with another as profoundly disturbing (see Hegel, G.F.W. §5). In the gaze of the Other, a consciousness recognizes a point of view which is different from its own and unattainable, a mark of its own incompleteness. At the same time, the gaze of the Other threatens to destroy it by turning it into an object. In response, the consciousness can choose to retaliate - to objectify the Other. But in doing so it destroys an external view of itself and must resign itself to the incompleteness of its self-understanding. The consciousness is therefore caught: it can dominate the Other, or live with the threat it poses.

It is through woman, Beauvoir argues, that male consciousness alleviates this conflict. Like men, women are conscious beings capable of returning the male gaze, and yet they allow themselves to be dominated. By possessing them, men are able to control the Other without destroying it, to withstand a gaze which is not unbearably threatening. Why do women occupy this position? Why do they not try to dominate men? Although Beauvoir suggests that the comparative passivity of women originates in childbearing, she is more interested in analysing the multitude of social practices which conspire to keep women in the position of the Other and prevent them from seeking their own transcendence. These practices, she argues, are sustained both by men, who encourage and reward female passivity, and by women, who cooperate in their own domination. The latter, however, is an example of bad faith. To allow oneself to be treated as an object is to fail to realize one's being by making ones own choices, and to shirk the painful project of becoming free. How, then, are women to liberate themselves? To avoid becoming Other, Beauvoir suggests, women must abandon the roles of wives and mothers in which they are most easily objectified and compete with men through work. Once they begin to exercise the assertiveness and courage essential to freedom, conceptions of what it is to be a woman will alter, and women and men will find ways to treat one another as equals.

One of Beauvoir's most profound contributions to feminist philosophy lay in her insistence that women are dominated in all aspects of their lives. Their comparative lack of freedom does not consist merely in lack of civic rights, or in particular institutions of motherhood and marriage, although these are contributory factors. Rather, they are kept in their inferior place by 'the whole of civilization' - by a multitude of evaluations and social practices (tellingly described in chapters on childhood, the young girl, sexual initiation and so on) which shape our understandings of male and female, masculine and feminine. As she indicates in her celebrated remark, 'One is not born, but rather becomes, a women', Beauvoir holds that it is through social practices that bodies come to be understood as sexually differentiated, and through these same practices that the differences between them are invested with evaluative significance. Becoming a woman is a cultural and historical process which is never completed. Although Beauvoir allows that there will always be differences between women and men deriving from their bodily distinctions and the effect these have on their sensuality, there is no one thing that women intrinsically or naturally are. Correspondingly, there is no discernible limit to what they may become.

6 Second- and third-wave feminism

Many of the critical and constructive themes discussed by Beauvoir were taken up again in the late 1960s and 1970s (though often without much reference to *The Second Sex*) by a generation of women who struggled, in the light of their personal experience, to revise the social and psychological theories around which academic debate revolved. On a critical plane, they enlarged Beauvoir's objections to Marxism and psychoanalysis and added criticisms of other sociological approaches such as functionalism, sometimes engendering debates which remained lively throughout the next twenty or so years. For feminists concerned with Marxism, the key issues were whether women could be satisfactorily accommodated within the class structure of society, and whether women's oppression could be adequately explained in terms of their place in the relations of production and the ideologies to which these gave rise. Studies of domestic labour and of women's sexual subordination suggested that, while Marxist analyses of women in capitalist societies remained valuable, the answer to these questions was negative. Turning their attention to psychoanalysis, a number of writers launched an influential attack on what they saw as Freud's construction of femininity as a passive, masochistic, narcissistic and intellectually limited condition. These readings of Freud and his successors gave feminists pause, and initiated a series of fruitful reinterpretations and modifications within psychoanalytic theory (see Feminism and psychoanalysis; Irigaray, L.; Kristeva, J.).

This critical work was also the vehicle for a number of important innovations in feminist thinking which raised fresh questions and consolidated novel approaches. Writers such as Kate Millett and Shulamith Firestone argued in the early 1970s that the forms of domination isolated by feminists are all relatively superficial in comparison with patriarchy - the sexual power that men exercise over women, primarily within the family, but also in social, economic and political institutions. In a wide range of societies, it was pointed out, men's sexuality is the source and justification of their power, the purportedly natural characteristic that gives them the right to rule women. The workings of patriarchy are evident not just in erotic relations between the sexes, but in the manifold means by which men and women are socialized as to temperament, role and status, men being taught to regard themselves as potent and active, women to perceive themselves as subordinate and sexually impure.

Patriarchy, then, relies not so much on the biological differences between men and women as on deep-seated cultural interpretations that give them value and significance. In the early 1970s this distinction came to be regarded as crucial, and writers such as Millett and Ann Oakley took over the terms ‘sex’ and ‘gender’ to mark it: sex refers to the biological traits that make a person male or female, gender to culturally variable conceptions of masculinity and femininity. Taken together, the notions of patriarchy, sex and gender provided an Anglo-American articulation of many of the themes announced by Beauvoir, and gave rise to a series of theoretical debates, some of which are still going on. Is sex really separable from gender, or is our experience and theorizing so mediated by culture that the idea of the simply biological ceases to make sense? Is patriarchy a useful analytical category, or is it either unduly general, or unduly reductionist? How, in any case, is patriarchal power related to other forms of political and economic power? And is it really as strong and pervasive as its exponents claim?

Regardless of the fate of these questions, the belief that men’s domination of women may be sustained by all sorts of practices had a vast impact on the Academy, as feminists began to take a fresh look at the texts and theories they studied professionally. This approach proved fruitful when applied to literary texts - Simone de Beauvoir had included a study of ‘The Myth of Woman in Five Authors’ in *The Second Sex*, and Millett’s *Sexual Politics* opens with striking readings of Henry Miller, Norman Mailer and Jean Genet (see Feminist literary criticism). It was soon adopted by philosophers, who started to analyse the conceptions of gender embedded in the great works of the philosophical tradition. Genevieve Lloyd’s *The Man of Reason* (1984) and Carole Pateman’s articles on contractarian political theory are notable early examples of this kind of work, and were rapidly followed by critical scrutinies both of the various areas of philosophy and particular positions within them (see Feminist aesthetics; Feminist epistemology; Feminist ethics; Feminist jurisprudence; Feminist political philosophy; Feminism and social science; Feminist theology).

While the results of this academic flowering have been extremely diverse, two themes stand out. First, some of the most impressive contributions to this recent work have shown how philosophical standards and doctrines that have claimed for themselves an objective and universal status reflect particular interests, values and priorities attuned to broader conceptions of masculinity. In this way, philosophy contributes to the cultural constructions of gender that play a part in legitimating and maintaining men’s power over women. Second, the third-wave feminism of the 1980s and 1990s has turned its critical techniques back on feminism’s own long-standing habit of making claims on behalf of ‘women’. These purportedly universal pronouncements, it is now pointed out, fail to take account of the differences between women of diverse races, sexual orientations, nationalities or classes. Moreover, if gender is not a natural category, there is nothing to be said about women as such, and we must become more sensitive to the many conceptions of femininity found in different societies. This anti-essentialism has profound implications for feminism, both as an academic preoccupation and as a political movement, and marks an important shift away from its own origins towards new themes and questions.

See also: Gender and science; Language and gender

SUSAN JAMES

References and further reading

Aristotle (*c.* mid 4th century BC) *Generation of Animals*, in J. Barnes (ed.) *The Complete Works of Aristotle*, vol. 1, Princeton, NJ: Princeton University Press, 1984, 1111-218.(Aristotle’s discussion of male and female roles in procreation.)

Beauvoir, S. de (1949) *The Second Sex* , trans. and ed. H.M. Parshley, Harmondsworth: Penguin, 1972.(The classic existentialist interpretation of masculinity and femininity and a path-breaking analysis of the social construction of gender.)

Davies, E. (1866) *The Higher Education of Women*, London: Hambledon, 1988.(The case for women’s education put by a pioneer.)

Engels, F. (1884) *The Origin of the Family, Private Property and the State*, intro. M. Barrett, Harmondsworth: Penguin, 1985.(A vastly influential discussion of the family by one of the founders of Marxism.)

Firestone, S. (1970) *The Dialectic of Sex*, London: The Women’s Press, 1979.(Firestone argues that women’s oppression lies in their reproductive role and that the goal of feminist revolution is the elimination of sexual difference.)

Goldman, E. (1972) *Red Emma Speaks. Selected Speeches and Writings of the anarchist and feminist Emma Goldman*, ed. A. Kates Shulman, New York: Random House.(Includes selections from Goldman’s writing on

the individual and the state, woman's emancipation, prostitution and marriage.)

Gouges, O. de (1782) *The Rights of Women*, trans. V. Stevenson, London: Pythia, 1989.(An impressive early statement of women's rights.)

Gournay, M. Le Jars de (1622) *Égalité des Hommes et des Femmes*, Paris: Côté-Femme, 1989.(A brief defence of the equality of the sexes by a woman whom Montaigne described as his adopted daughter. This edition includes the *Grief des Dames*, a defence of women of letters.)

Griffith, E. (1984) *In her own Right. The life of Elizabeth Cady Stanton*, New York: Oxford University Press.(A biography of one of the best-known US suffragists.)

Kollontai, A. (1977) *Selected Writings*, trans. and intro. A. Holt, London: Allison & Busby. (Includes selections from Kollontai's writing on women and work, work and motherhood, communism, the family and marriage.)

Lloyd, G. (1984) *The Man of Reason*, London: Methuen.(Innovative analysis of the prevalence of the association of women and irrationality within the philosophical tradition.)

Maclean, I. (1977) *Woman Triumphant. Feminism in French literature 1610-1652*, Oxford: Clarendon Press.(A study of works by and about women in seventeenth-century France. Includes a chapter on the *Querelle des femmes*.)

Mill, J.S. (1869) *On the Subjection of Women*, in S. Collini (ed.) *On Liberty and other Writings*, Cambridge: Cambridge University Press, 1989.(A classic, nineteenth-century account of the advantages of giving women the same access to education and work, and the same civil rights, as men.)

Millett, K. (1969) *Sexual Politics*, London: Virago, 1977.(A formative analysis of the construction of femininity in twentieth-century literature and in social theory.)

Oakley, A. (1972) *Sex, Gender and Society*, London: Temple Smith.(An influential discussion of the biological and social aspects of sexuality.)

Pateman, C. (1989) *The Disorder of Women*, Cambridge: Polity Press.(Includes seminal essays on the place of women in contractarian political theories.)

Perkins Gilman, C. (1919) *Herland*, London: Women's Press, 1979.(Utopian novel by a leading American feminist about a community of women who live without men.)

Pizan, C. de (1405) *The Book of the City of Ladies*, trans. E.J. Richards, New York: Persea Books, 1982.(An exchange between Christine and the figures of Reason, Rectitude and Justice, who reassure her that women are not evil and instruct her to build a city for ladies.)

Poulain de la Barre, F. (1673) *The Equality of the Sexes*, trans. D.M. Clarke, Manchester: Manchester University Press, 1990.(An early defence, along Cartesian lines, of the equality of the sexes.)

Sartre, J.-P. (1943) *Being and Nothingness*, trans. H. Barnes, London: Methuen, 1958. (Sartre's classic work, mentioned in §5 above.)

Wollstonecraft, M. (1792) *A Vindication of the Rights of Women*, in S. Tomaselli (ed.) *A Vindication of the Rights of Man and A Vindication of the Rights of Women*, Cambridge: Cambridge University Press, 1995.(A classic defence of the view that women and men have the same rights though different social roles.)

Feminism and social science

Feminists have two sorts of interest in the social sciences. With the advent of the second-wave women's movement, they developed wide-ranging critiques of gender bias in the conceptual framework and methodology, as well as in the goals, institutions and practice of virtually all the social sciences; they argue that the social sciences both reflect and contribute to the sexism of the larger societies in which they are embedded. Alongside these critiques feminist practitioners have established constructive programmes of research that are intended to rectify the inadequacies of existing traditions of research and to address questions of concern to women. In this they are committed both to improving the disciplines in which they participate and to establishing a sound empirical and theoretical basis for feminist activism. This engagement of feminists with social science, as commentators and practitioners, raises a number of philosophical issues that have been addressed by feminist social scientists and philosophers. These include questions about ideals of objectivity and the role of contextual values in social scientific inquiry, the goals of feminist research, the forms of practice appropriate to these goals, and the responsibilities of feminist researchers to the subjects of inquiry and to those who may otherwise be affected by its conduct or results.

1 Feminist critiques of the social sciences

Feminist critics of the social sciences make the case, usually with respect to specific examples, that often even the most rigorous and conscientious research in the social sciences has been androcentric or sexist or both: it embodies a male-centred view of its subject matter, and sometimes explicit assumptions of male superiority. These critiques began to appear regularly in most fields in the early 1970s and quickly gained wide currency. They frequently coincided with the founding of committees on the status of women which undertook to document and to counteract inequities in the recruitment, training, employment and professional recognition of women within the social sciences. Although there are important differences in the form that feminist critiques take and in the impact they have had in various fields and subfields of the social sciences, they follow similar lines of development and converge on four broad areas of concern (Wylie *et al.* 1989).

Often the first issues raised by feminist critics have to do with the goals of the social sciences. They object that the questions central to much mainstream research reflect androcentric interests in the social world; women's activities, roles and contributions have been largely ignored, as have issues of particular concern to women (for example, violence against women, the history of domestic technologies, gender segregation in the workplace). At the very least this line of criticism suggests that the scope of social scientific inquiry should be expanded. In addition, however, feminists ask how adequate our understanding can be, even of the subjects that are typically considered, if substantial dimensions of social life have been left out of account.

A second family of objections, content critiques, draws attention to ways in which social research is compromised when its subjects are conceptualized in terms that reflect sexist or androcentric presuppositions. These include implicit, 'taken-for-granted' assumptions about gender categories and relations - about the typical (or 'natural') roles and capabilities of women and men. Male experience, behaviour, interests and attributes may be treated as normative, with the result that gender differences are ignored. In some cases the whole subject domain is characterized in male-specific terms or, when women are considered, it is exclusively in relation to men (Eichler describes these as errors of overgeneralization and gender insensitivity). Some widely cited cases include ethnographic studies in which 'hunter-gatherer societies' are characterized in terms of the hunting activities of men (Rosaldo and Lamphere 1974), and the 'man the hunter' theories of evolution informed by these accounts (Dahlberg 1981; Hager 1997), historical periodization schemes which reflect a preoccupation with the fortunes of (elite) men (Kelly-Gadol 1976), and theories of human psychological development based exclusively on male samples (Gilligan 1982; see Moral development). Gender stereotypes may also be reflected in the converse assumption that men and women are categorically different. In this case, similar and overlapping traits are either ignored or are characterized in different terms when associated with men rather than women (in Eichler's terms these are errors that arise from imposing a 'double standard' and from 'sexual dichotomism', 1988); sex difference research abounds with examples of this practice (Fausto-Sterling 1985). The central point feminist critics make is that the background assumptions underlying such 'distortions' of content should be made explicit and subjected to critical empirical scrutiny.

These critiques both generate and depend upon ’remedial’ research; the study of ‘women worthies’, ’women’s contributions’ and ’women victims’ (Harding 1986) has been crucial in establishing, in concrete terms, what androcentric social science has left out of its account. But wherever such compensatory strategies have been adopted, autocritiques have appeared which made it clear that they are not sufficient on their own. To take one example among many that could be cited, through the 1970s and early 1980s feminist anthropologists, historians and political scientists argued (often independently) that a central limitation of their respective fields of study is a preoccupation with a male-identified ’public’ domain (for example, public figures, public exercises of power, politics in a juridical sense); they urged that feminists turn their attention to the neglected, private, domestic spheres in which women are to be found. Critical reflection shows, however, that simply inverting the emphasis of previous research leaves intact precisely the gender categories and stereotypes that feminists had insisted should be called into question (Rosaldo 1980; Tomm and Hamilton 1988). To rectify the androcentrism operating at this deeper level it is necessary, not just to add considerations of (for example) the work women do, the symbolic universes they construct and the forms of power they exercise in domestic and other contexts, but to more fundamentally reconceptualize what counts as work or as power, and as the ’domestic’ or private sphere both in its own terms and in relation to ’public’ life.

By the mid-1980s it was widely noted that the effects of androcentric goals and presuppositions were pervasive in two senses. Deepening critiques of content suggested that all aspects of research may be affected, not just the formation of questions or the interpretation of results, but also domain-defining assumptions that determine the design of research, the selection of methodologies and the formulation of categories of description and analysis. Moreover, androcentric bias of one kind or another had been identified in virtually all the fields of social research in which feminists had undertaken critical analyses. A third type of critique, of research methodologies, arises when feminists ask how it is possible that androcentric bias in the goals and content of inquiry could be as deeply rooted and persistent as their collective analyses suggest.

The point of departure for methodological critiques is a suspicion that the forms of inquiry typical in the social sciences may be responsible, in some part, for the production and persistence of androcentric bias. Some feminists maintain that established methods for evaluating knowledge claims are adequate as they stand, but that they have simply not been applied as widely or as systematically as they should have been; this constitutes a form of feminist empiricism (Harding 1986). Others insist that we should reassess our faith in the capacity of these methods to expose systematic bias. They may be powerful tools for assessing the empirical adequacy, reliability and scope of specific knowledge claims within a frame of reference, but will often simply reproduce error if it is rooted in the fundamental presuppositions of inquiry. Still others suggest that some methods may actually generate androcentric and sexist bias. For example, in many contexts a central theme in feminist critiques has been a rejection of the ’naturalist’ conviction that social research must conform to models of natural scientific practice (Mies 1983; Nielsen 1990; Stanley and Wise 1983) (see Naturalism in social science). Especially when formulated in terms of positivist theories of science, this enforces a preoccupation with quasi-experimental and quantitative methodologies which may systematically obscure devalued subordinate perspectives (for example, of women) that do not conform to the conventions of dominant ideology. Although few endorse this thesis in its strongest form - few maintain that research methods are intrinsically sexist or androcentric - it is clear that, in order to recover dimensions of social life ignored by mainstream research, feminist practitioners frequently and increasingly turn to forms of evidence considered ephemeral (for example, diaries, private papers, material culture, as opposed to archives of public record) and rely on methodologies that are nonstandard or marginal by the conventions of their fields (for example, open-ended qualitative methodologies in fields dominated by naturalism).

A fourth and final type of feminist critique is also inspired by reflection on the ways in which research is typically conducted in the social sciences, especially when informed by objectivist and positivist ideals (Fonow and Cook 1991; Nebraska Collective 1983; see the summary discussions in Reinharz 1992). A recurrent theme is concern that human subjects, especially those who are disempowered by established social hierarchies, may be demeaned and exploited by research practices which require that they be ‘objectified’ or treated like objects (Stanley and Wise 1983) (see Objectivity). In this feminists join many other social scientists who have insisted on the need to take responsibility for the effect of their research practice on subjects, and to scrutinize the ways in which they, and their research enterprises, function as agents of social manipulation and control (Smith 1987).

2 Constructive initiatives

These critiques by no means exhaust feminist interests in the social sciences. As insiders to the social sciences (Westcott 1979; Collins in Fonow and Cook 1991), feminist practitioners actively work against the sexism and androcentrism they identify in extant traditions of research; in this they are concerned to improve the research enterprises in which they are engaged. In addition, as feminists, many have a strong commitment to change conditions of life that are oppressive for women and they recognize that, if they are to be effective in realizing these goals, it is necessary to understand with accuracy, subtlety and explanatory precision the nature and sources of the inequities that disadvantage women. Whatever their limitations, social-scientific modes of inquiry are among the most powerful tools available for doing this. Feminists also have an interest, then, in developing programmes of research which address questions relevant to women and promise a secure empirical and conceptual foundation for feminist activism.

In taking up these constructive projects, feminists immediately confront the questions of how research should be conducted if it is to avoid the pitfalls of androcentric practice, and whether there is any distinctively 'feminist' mode of inquiry. Consensus emerged by the late 1980s that the quest for a 'feminist method' or 'feminist science' is misguided; it bespeaks a faith in methodological solutions that feminists have long cautioned against, and presupposes an implausible unity of feminist interests. It is more fruitful to ask what it means to 'do [social] science as a feminist' (Harding 1987; Longino 1987, 1994). In this spirit feminist practitioners have typically formulated quite general guidelines for ensuring that research practice is nonsexist, and is otherwise consistent with feminist ideals (Eichler 1988; Mies 1983; Reinharz 1992). These may be characterized in terms of responses to the four types of criticism described above.

First, feminists have insisted that, at the very least, the goals of social scientific inquiry should include consideration of women and gender. In addition, many feminist social scientists advocate research that has (potential) practical import: research that addresses questions of concern to women, and that promises an understanding relevant for improving their lives.

Second, a persistent, if controversial, theme is the recommendation that feminist research should be 'grounded in' women's experience (Mies 1983; Stanley and Wise 1983). This principle is interpreted in widely divergent ways. In some cases it is understood to mean that feminist research should take women's lives and experience as a primary subject of inquiry; it should redress the silencing and devaluation of women's perspectives that is reproduced by male-normative thinking in the social sciences. Many consider this too restrictive a mandate for feminist researchers, and urge that the experience and understanding of women should serve as a basis for identifying new questions, formulating new categories of description and analysis, and expanding our interpretive repertoire; it represents a standpoint from which the limitations of androcentric thinking may be discernible in all areas, not just those in which women predominate (see Harding 1991 on 'thinking from women's lives').

Third, feminist practitioners frequently insist on thoroughgoing 'reflexivity': they hold that the theoretical and methodological presuppositions of their research should be subject to the same sort of critical scrutiny as they apply to mainstream research (Longino's 'provisionalism', 1994).

Finally, by extension of these three principles, feminist social scientists are actively concerned with issues of accountability to research subjects and to others who may be affected by the practice or results of their research. They insist that, as a minimum requirement, research practices should not, themselves, exploit or oppress women. Often they urge that research be designed, as far as possible, as interventions that may benefit research subjects and they recommend collaborative models of inquiry.

Consistent with the requirement for reflexivity, there is active ongoing debate about all these recommendations and their implications for practice.

3 Philosophical implications

These constructive guidelines for doing research as feminists, and the critiques from which they arise, pose a number of philosophical problems, foundational and epistemic. In the process of interrogating entrenched assumptions about women and gender, feminist social scientists have contributed to wider debates that substantially destabilize the assumption that these constitute coherent categories which can be presupposed as the

basis either for feminist activism or feminist scholarship. As critics of the inherent elitism, racism, and ethnocentrism of second-wave feminism point out, women's experience is so diverse and gender categories and roles so deeply structured by other systems of social differentiation, it is untenable to treat 'gender' as an autonomous category for analysis or to presume a unitary 'women's' standpoint (Anderson and Collins 1995; Mohanty, Russo and Torres 1991). This reinforces the commitment to reflexivity; it is a thoroughly contingent, empirical question how gender should be conceptualized and whether it is a relevant category for critical or constructive analysis. The interests of feminists in the social sciences thus converge on those of critics working from innumerable other subordinate standpoints.

In addition to these foundational issues, feminist critiques of and contributions to the social sciences raise fundamental epistemic questions about entrenched ideals of objectivity and the relationship of 'external' contextual values and interests to the 'internal' epistemic values generally assumed to be constitutive of scientific inquiry (Longino 1990). The extent and depth of androcentric bias identified by feminist critics makes it increasingly implausible that the difficulty lies only with instances of 'bad science'; it suggests that good science, 'science as usual', even our best science, is more deeply infused with the values and interests of its practitioners than has generally been acknowledged (Harding 1986). Moreover, where feminists working from an explicitly interested standpoint have made important contributions to the social sciences, even judged by quite traditional standards, it seems that contextual values should not always be regarded as a contaminant of inquiry (Alcoff 1987). Standpoint theorists argue, on the basis of such observations, that 'contextual' values play an irreducible role in all research practice, and postmodern critics are frequently identified as drawing strong relativist conclusions on this basis (see Harding 1986 for a summary of these positions). In practice, however, few feminist practitioners are prepared to embrace such conclusions (di Leonardo 1991; Hawkesworth 1989). This is not just a strategic stance motivated by concern that feminists' insights should be taken seriously as offering an improved rather than just an alternative understanding of the social world. It reflects, as well, an appreciation, often born of activist engagement, that it is possible to be (disastrously) wrong even about subjects as enigmatic as the cultural, intentional systems studied by social scientists. As is often observed, only the most powerful could imagine that the world is as they construct it; for feminists much depends on establishing an empirically and explanatorily accurate understanding of the world in which they hope to intervene.

The epistemological challenge posed by feminist research is to develop an account of epistemic values and ideals, such as empirical adequacy, explanatory power, objectivity, which makes sense of the nuanced judgments social scientists (including feminists) routinely make about the relative credibility of knowledge claims without requiring untenable conditions of value-neutrality or empirical foundationalism (see Foundationalism). In taking up these questions, feminists engage issues which are central to postpositivist philosophy of science (Alcoff and Potter 1993; Lloyd 1995).

See also: Feminism; Feminist epistemology; Feminist ethics; Feminist jurisprudence; Feminist political philosophy; Gender and science; Logical positivism; Scientific method; Social sciences, philosophy of

ALISON WYLIE

References and further reading

Alcoff, L. (1987) 'Justifying Feminist Social Science', *Hypatia*, special issue on 'Feminism and Science', 2: 107-27.(An excellent survey of epistemological issues raised by feminist work in the social sciences.)

Alcoff, L. and Potter, E. (eds) (1993) *Feminist Epistemologies*, New York: Routledge.(Provides a good introduction to the issues in feminist epistemology and philosophy of science discussed in §3.)

Anderson, M.L. and Collins, P.H. (eds) (1995) *Race, Class, and Gender*, Belmont, CA: Wadsworth.(An excellent anthology on the intersections between dimensions of oppression, cited in §3.)

Dahlberg, F. (ed.) (1981) *Woman the Gatherer*, New Haven, CT: Yale University Press.(Includes examples of feminist critiques of androcentrism in theories of evolution and ethnographies of foraging societies, as cited in §1.)

Eichler, M. (1988) *Nonsexist Research Methods*, Boston, MA: Allen & Unwin.(A comprehensive summary of the feminist critiques discussed in §1.)

Fausto-Sterling, A. (1985) *Myths of Gender*, New York: Basic Books.(Includes a review of feminist critiques of sex-difference research, as cited in §1.)

Fonow, M.M. and Cook, J.A. (eds) (1991) *Beyond Methodology: Feminist Scholarship as Lived Research*,

Bloomington, IN: Indiana University Press.(A useful collection of essays on feminist contributions to the social sciences with an excellent introductory summary of the principles that inform feminist research.)
Gilligan, C. (1982) *In a Different Voice: Psychological Theories and Women's Development*, Cambridge, MA: Harvard University Press.(A critique of Kholberg's theory of moral development, cited in §1.)
Hager, L. (1997) *Women in Human Evolution*, London: Routledge.(A selection of essays on feminist research on evolution, cited in §1.)
Harding, S. (1986) *The Science Question in Feminism*, Ithaca, NY: Cornell University Press.(Includes an influential discussion of feminist empiricism, standpoint theory and postmodernism.)
Harding, S. (ed.) (1987) *Feminism and Methodology: Social Science Issues*, Bloomington, IN: Indiana University Press.(Includes influential examples of feminist research in the social sciences and a useful introduction to the epistemological issues they raise.)
Harding, S. (1991) *Whose Science? Whose Knowledge?*, Ithaca, NY: Cornell University Press.(Includes discussion of concepts of objectivity relevant to §3.)
Hawkesworth, M.E. (1989) 'Knowers, Knowing, Known: Feminist Theory and Claims of Truth', *Signs* 14: 533-57.(A useful summary of feminist epistemologies, including a critique of feminist postmodernism. Cited in §3.)
Kelly-Gadol, J. (1976) 'The Social Relations of the Sexes: Methodological Implications of Women's History', *Signs* 1: 809-23.(Cited in §1, a feminist critique of underlying assumptions that structure historical inquiry.)
Leonardo, M. di (1991) *Gender at the Crossroads of Knowledge: Feminist Anthropology in the Postmodern Era*, Berkeley, CA: University of California Press.(A valuable selection of feminist research in anthropology with an excellent introductory essay, cited in §3.)
Lloyd, E.A. (1995) 'Objectivity and the Double Standard for Feminist Epistemologies', *Synthèse* 104: 351-81.(An account of the relationship between feminist critiques of objectivity and central themes in postpositivist philosophy of science. Cited in §3.)
Longino, H. (1987) 'Can there Be a Feminist Science?', *Hypatia* 2: 51-64.(An excellent summary of and response to the debates about 'feminist method'. Cited in §2.)
Longino, H. (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*, Princeton, NY: Princeton University Press.(Provides a sophisticated introduction to the philosophical issues discussed in §3.)
Longino, H. (1994) 'In Search of Feminist Epistemology', *The Monist* 77: 472-85.(An overview of values that inform feminist research, cited in §§2 and 3.)
Mies, M. (1983) 'Toward a Methodology of Feminist Research', in G. Bowles and R.D. Klein (eds) *Theories of Women's Studies*, New York: Routledge & Kegan Paul.(An influential and controversial proposal for feminist research that focuses on epistemological issues. Cited in §§1 and 2. See also Mies' contribution to Fonow and Cook 1991.)
Mohanty, C.T., Russo, A. and Torres, L. (eds) (1991) *Third World Women and the Politics of Feminism*, Bloomington, IN: Indiana University Press.(An anthology of critiques of the elitism of much first world feminist scholarship and activism with an excellent introductory analysis by Mohanty. Cited in §3.)
Nebraska Feminist Collective (1983) 'A Feminist Ethic for Social Science Research', *Women's Studies International Forum* 6: 535-45.(A classic statement of ethical issues addressed by feminist social scientists. Cited in §§2 and 3.)
Nielsen, J.M. (ed.) (1990) *Feminist Research Methods: Exemplary Readings in the Social Sciences*, Boulder, CO: Westview Press.(A collection of essays that both describe and exemplify principles of feminist research with a useful introduction to these principles and the epistemological issues they raise. Cited in §§2 and 3.)
Reinharz, S. (1992) *Feminist Methods in Social Research*, Oxford: Oxford University Press.(The most comprehensive account available of feminist research practice in the social sciences.)
Reinharz, S., Bombyk, M. and Wright, J. (1983) 'Methodological Issues in Feminist Research: A Bibliography of Literature in Women's Studies, Sociology and Psychology', *Women's Studies International Forum* 6: 437-54.(An excellent guide to the first decade of feminist research in the social sciences; includes references to much of the literature cited in §1.)
Roberts, H. (ed.) (1981) *Doing Feminist Research*, London: Routledge & Kegan Paul.(Includes important critiques of traditional research methodologies, as described in §1.)
Rosaldo, M.Z. (1980) 'The Use and Abuse of Anthropology: Reflections on Feminism and Cross-Cultural Understanding', *Signs* 5: 389-417.(An example of internal feminist critiques that call into question 'remedial' research, as described in §§1 and 2.)
Rosaldo, M.Z. and Lamphere, L. (eds) (1974) *Women, Culture and Society*, Stanford, CA: Stanford University Press.(An anthology of early and widely influential feminist critiques of anthropology of the sort cited in

§1.)
Smith, D. (1987) *The Everyday World as Problematic: A Feminist Sociology*, Toronto, Ont.: University of Toronto Press.(An influential feminist critique of sociological technologies of control, informed by Marxist and ethnomethodological critiques of sociology.)
Stanley, L. and Wise, S. (1983) *Breaking Out: Feminist Consciousness and Feminist Research*, London: Routledge & Kegan Paul.(One of the most radical and influential epistemological critiques of traditional research, cited in §1.)
Tomm, W. and Hamilton, G. (eds) (1988) *Gender Bias in Scholarship: The Pervasive Prejudice*, Waterloo, Ont.: Wilfrid Laurier University Press.(A useful collection of essays summarizing feminist critiques of and contributions to the social sciences.)
Westkott, M. (1979) 'Feminist Criticism of the Social Sciences', *Harvard Educational Review* 49: 422-30. (Provides a review of feminist critiques discussed in §1.)
Wylie, A., Okruhlik, K., Thielen-Wilson, L. and Morton, S. (1989) 'Feminist Critiques of Science: The Epistemological and Methodological Literature', *Women's Studies International Forum* 12: 379-88.(A general guide to feminist critiques of science which includes sections on the social sciences. Provides references to much of the literature considered in §1.)

Feminist aesthetics

Feminist perspectives in aesthetics and philosophy of art have emerged not only from the discipline of philosophy but also from cognate fields such as literary theory, film studies and art history. Like other feminist philosophy, feminist aesthetics is founded upon critiques of fundamental assumptions that have traditionally governed this area of study. Such staple concepts as aesthetic value, disinterested attention, aesthetic perception and fine art have been analysed for biased perspectives that explicitly or covertly favour masculine gender.

Gender bias has been located, for example, in theories of aesthetic attention and appreciation. Feminist analyses have speculated that the traditional, ideal, 'disinterested' aesthetic perceiver covertly stands in a position of masculine privilege encouraging desire and control of the object of contemplation. Calling attention to the masculine position of the perceiver goes hand in hand with analyses of visual arts and literature that focus on their representations of gender - formalist interpretive methods that ignore the portrayal of women and sexuality have been especially criticized. The paucity of female artists on lists of acknowledged geniuses and the relative absence of their work from canons of art has occasioned speculation that the concept of art itself is biased in so far as it excludes the creations of most women.

Feminist analyses stress the social contexts within which theories develop. Not only do concepts basic to aesthetics manifest ideas about gender that derive from formative traditions, but notions of beauty and art are themselves influential components of culture and contribute to the shape and perpetuation of gender roles.

1 Historical critiques

Feminist challenges to academic disciplines begin with critiques of the basic assumptions of a field and often consist of examinations of canonical texts for any gender-related implications. In philosophy, this involves demonstrating that theories that purport to be universalist, that is, to pertain generically to all human beings, are in fact models of ideal masculine nature. The gender biases of Western philosophy are often related in some way to the fundamental reliance of philosophical views on the centrality of reason. Reason is the standard faculty of knowledge and of moral responsibility, and it operates in most theories as the chief essential characteristic of being human. Reason raises the human creature above nature, making possible creation beyond what nature has given. But rationality is also a trait that is employed to contrast male with female dispositions, for the nonrational domains of emotion, sense and the body are widely associated with femininity. Thus many theories that at first glance seem to be general views about all people may be discovered to refer standardly to men, and sometimes actually to exclude women. In the history of aesthetics, reason is often not the chief faculty governing artistic creativity or aesthetic experience. However, still operative in aesthetic theory are the binary correlations that link reason to masculinity and the transcendence of nature, and that associate emotion, the body, femininity and rootedness in nature. Two examples that illustrate this phenomenon are provided by the notion of artistic genius and the concepts of the beautiful and the sublime.

Though the idea of genius has taken a variety of forms throughout its long history, none has countenanced a female genius. The sensitivity, intuition and even madness sometimes associated with genius contravene rationality; nonetheless genius is preserved as a masculine domain by its bold originality and by the ability of the creative artist to transcend what is given by nature or culture. In the Romantic period - arguably the time when genius was of the greatest importance in aesthetic theory - the exclusion of women from the company of geniuses was especially marked; Kant, Nietzsche and Schopenhauer are among those who discuss genius in exclusively masculine terms. (Kant notoriously remarked that a woman who aspired to higher learning, let alone genius, might as well wear a beard.) The presumption of the masculinity of genius effectively screens out women from the ranks of the most important producers of art; however skilful their creations, lacking genius, they can presumably produce little more than derivative versions of the works that define and vivify a culture.

In the formative period of eighteenth-century European theory, the category of the aesthetic itself was becoming established, and this period framed views of perception, beauty and criticism upon which contemporary theories rest. At this time, beauty and sublimity were important theoretical categories, and the language of analysis directed to these topics is consistently gendered. The experience of the sublime - unbounded, vast, awesome - was cast in terms that not only stress the adventurous masculinity of the experience but also suggest that its grasp exceeds the

capacity of the female mind. Beauty was conceived as a pleasurable, bounded, controllable object of enjoyment and described by terms such as ‘graceful’ or ‘delicate’. In fact, the bodies of women figured as examples of objects of beauty more often than females were invoked as aesthetic perceivers. Implicitly, the ideal critic was male (indeed, a heterosexual male), and judgments of beauty and taste remained masculine activities.

2 Creativity and art

Since the greatest geniuses were conceived to be masculine, it was correspondingly difficult to make room for the works of female artists in the lists of important art. The degree to which women have been hindered from practising art varies according to genre: arts such as painting and music have been highly exclusionary; literature and dance less so. Philosophically, the most important issue regarding the absence of women from the canon is the effect of this on the concept of art itself. Here conceptual analysis merges with social history: the idea of fine art includes the production of objects that are conceived not only to have enduring value, but to be public objects that are products of individual talent. Throughout the history of Western art, few women have had the access to education and studio training that men have enjoyed, and furthermore many of them have chiefly been assigned domestic roles. What they created in such an ambience were not the sorts of objects that could easily join the category of fine art, but were rather objects of domestic crafts, designed as much for practical function as for beauty or expressivity. Moreover, products such as quilts are often the work of many hands, making the creative talents of individuals hard to discern. Thus the limitations of education and social roles are not the sole explanation for the scant presence of women in the history of Western painting, music and sculpture. The reigning concept of art and its attendant notion of creative genius is blind to the objects that most women in fact have created during much of recorded history. In recognition of this fact, many feminist scholars of the arts are expanding the scope of their enquiries to include hitherto neglected genres, such as domestic arts and crafts or private diaries and letters, in order to bring the art of women of the past into full view.

3 Perception and the male gaze

Above it was noted that formative eighteenth-century theories of the beautiful frequently mention women as objects of beauty and aesthetic attention. This fact needs to be placed in context with one of the most influential ideas that has governed the idea of the aesthetic in Euro-American theory: that of disinterested attention. (Several terms are employed to express this notion, including distanced appreciation and aesthetic contemplation.) For two centuries this idea has formed the heart of theories of the aesthetic; indeed, it has virtually defined the notion. The aesthetic is classically defined as the realm of pleasure or enjoyment that is contrasted to practical, moral, religious or political interests. The desires of the individual perceiver are set aside in aesthetic enjoyment, which is a receptive state of appreciation of presentational qualities of an object of art or nature.

Such a description of a mental state makes it seem generic, that is, to be the same whoever is looking; gender difference would not be seem to play any role in the concept of aesthetic attention. However, feminist critics have juxtaposed this ideal mental state with standard examples of works of art: the history of art, sculpture and literature is rife with seductions and rapes, nude figures of odalisques and prostitutes. Sceptical about the figurative distance that the notion of the aesthetic puts between the viewer and the object, feminist theorists have speculated that the disinterested viewer actually assumes a position of power over the object viewed. Especially when the object is a (real or portrayed) female character or figure, the aesthetic attitude is one that assumes a kind of prerogative on the part of a viewer and passivity on the part of the object, whose chief purpose is to be posed for the enjoyment of the viewer. In place of the notion of disinterested aesthetic attention, feminists have introduced the notion of the ‘male gaze’.

While the general point is made for all art forms, the critique of aesthetic perception has been especially strong in discussion of the visual arts of film and painting. (The work of film-maker and theorist Laura Mulvey has been particularly influential.) The term ‘male gaze’ sums up an analysis of the privileged art connoisseur, who embodies a perspective of social dominance and particularly of traditional, patriarchal authority. Psychoanalysis provides a theoretical scaffolding congenial to many theorists of the gaze, because it fosters an understanding of how perception and pleasure become gendered. It is important to note that the viewer is in an imaginative masculine position, but need not actually be male. Male or female, socially powerful or marginal, the viewer assumes this position in order to appreciate art according to accepted norms. Nevertheless, identifying the ideal aesthetic position as masculine and patriarchal also raises questions about the critical, appreciative and interpretive act, and

about the uniformity of aesthetic responses.

4 Style

Given the above analysis of the history of aesthetic theory as gender-biased, and the suspicion that many of the central concepts and terms with which theory is conducted retain the mark of gender, what positive remedies are available to feminist theorists? Are there distinctions between the art works of females and those of males that might be developed into a 'feminine' aesthetic theory that offers fruitful alternatives to customary aesthetic values? Is there a discernible female creative tradition? Does feminist analysis suggest gender differences in looking, reading, listening or appreciating?

There are proponents of both affirmative and negative positions on all of these issues. A particular problem that must be addressed in answering any of these questions affirmatively is the problem of essentialism, by which is meant the view that all women at all times share certain common aesthetic responses or stylistic traits - identifiably feminine qualities that are manifest in spite of their social differences.

As feminist theory developed in the 1970s, some theorists argued that there are certain common themes and styles in women's art works. Since women regularly share experiences that are different from those of men, the case might be made that their common social histories yield similar artistic themes, such as childbirth and domesticity. Some tied thematic focus to speculations about style, surmising that fluidity and softness are more characteristic of feminine style than are rigid boundaries and divisions. Perhaps the strongest and certainly the most sophisticated suggestions regarding the feminine style have come from francophone theorists such as Julia Kristeva, Luce Irigaray and Hélène Cixous. Pursuing suggestions made by Lacan and Derrida that language represents the feminine only in terms of absence, these thinkers (in their different ways) introduce the possibility of writing that disrupts patriarchal, symbolic order. *L'écriture féminine* - one of the concepts that has issued from this work - is posited as women's writing that rejects patriarchal discourse and inscribes a distinctively feminine voice. This type of theory is both a deep critique of language that inscribes a patriarchal order in which women have no subject position of their own, and a visionary prescription for women's language once these fetters are recognized and broken.

Against these views lie the objections that such categorization merely mimics stereotyped notions of femininity; that women's social circumstances often vary; and that historical periods, positions of social class, race, ethnicity, nationality and religion are all factors that distinguish women from one another. According to this view, to look for a common 'feminine style' is to assume wrongly that there is some essential femininity to be discovered. An emphasis on variety in gender identities and fluidity of social roles undermines the search for distinctively feminine voices and styles. However, the search for common feminine aesthetic qualities needs to be distinguished from the empirical search for the work that women have created outside traditionally recognized art forms. Such a search presumes no female essence, but acknowledges that if one is going to study the work of women, one must look in the right places.

Not only have feminist perspectives in aesthetics been developed to re-examine assumptions and basic concepts and to envision alternatives, but it is also the case that aesthetics contributes to projects of feminist theory. Art, language and representation are powerful components of the formation of subjectivities and self-concepts, as well as of public culture; they perpetuate cultural norms and provide imaginative space for violating or transcending these norms. Thus the aesthetic and its pleasures, art, style and representation are locations for the investigation of gender and the boundaries of masculinity and femininity.

See also: Beauty; Derrida, J.; Feminism; Film, aesthetics of §1; Lacan, J.; Sublime, the

CAROLYN KORSMEYER

References and further reading

Battersby, C. (1989) *Gender and Genius: Towards a Feminist Aesthetic*, London: The Women's Press.(An analysis of the concept of genius from classical antiquity to the present.)

Brand, P.Z. and Korsmeyer, C. (eds) (1994) *Feminism and Tradition in Aesthetics*, University Park, PA: Penn State Press.(This anthology investigates philosophical issues in art from the perspective of feminism.)

Feski, R. (1989) *Beyond Feminist Aesthetics*, Cambridge, MA: Harvard University Press.(This work reviews

feminist theories of the arts, especially literature, arguing that there is no property common to the work of women.)
Hein, H. and Korsmeyer, C. (eds) (1993) *Aesthetics in Feminist Perspective*, Bloomington, IN: Indiana University Press.(Collection of articles exploring aesthetics in the light of feminist theory.)
Korsmeyer, C. (1996) 'Perceptions, Pleasures, Arts: Considering Aesthetics', inJ. Kouraney (ed.) *Philosophy in a Feminist Voice*, Princeton, NJ: Princeton University Press.(Expansion of the material of this entry.)
Marks, E. and de Courtivron, I. (eds) (1980) *New French Feminisms*, Amherst, MA: University of Massachussetts Press.(This collection contains selections from works of various francophone theorists.)
Mulvey, L. (1988) *Visual and Other Pleasures*, Basingstoke: Macmillan.(Essays analysing film and exploring psychoanalysis, gender and perpetual pleasures.)
Pollock, G. (1988) *Vision and Difference*, London: Routledge.(An art historian analyses painting and the male gaze.)
Wolff, J. (1990) *Feminine Sentences: Essays on Women and Culture*, Cambridge: Polity Press.(This work by a sociologist places problems of aesthetics in historical and social context.)

Feminist epistemology

The impact of feminism on epistemology has been to move the question 'Whose knowledge are we talking about?' to a central place in epistemological inquiry. Hence feminist epistemologists are producing conceptions of knowledge that are quite specifically contextualized and situated, and of socially responsible epistemic agency. They have elaborated genealogical/interpretive methods, have advocated reconstructions of empiricism, have articulated standpoint positions and have demonstrated the potential of psychosocial and post-structural analyses to counter the hegemony of epistemological master narratives. In these reconfigured epistemologies, feminists have argued that the cognitive status and circumstances of the knower(s) are central among conditions for the possibility of knowledge. They have demonstrated the salience, in evaluating any epistemic event, of the social arrangements of power and privilege by which it is legitimated or discredited.

Feminists are engaged at once in critical projects of demonstrating the privilege-sustaining, androcentric character of 'the epistemological project' in most of its received forms, and in transformative projects of reconstructing methodologies and justificatory procedures so as to eradicate their oppressive, exclusionary effects. They have shown that, in late-twentieth-century western philosophy, the circumstances of mature white men continue to generate prevailing ideals and norms of 'human nature', while the ideals of reason, objectivity and value-neutrality around which most mainstream theories of knowledge are constructed, like the knowledge they legitimate, tacitly validate affluent male experiences and values. Scientific knowledge, which is still an overwhelmingly male preserve, stands as the regulative model of objective epistemic authority; and the experiences and values of non-male, non-white and otherwise differently placed knowers typically have to accommodate themselves, Procrustean-style, to an idealized scientific and implicitly masculine norm, or risk dismissal as inconsequential, aberrant, mere opinion.

In engaging with these issues, most feminists - like many other participants in 'successor epistemology' projects - retain a realist commitment to empirical evidence, while denying that facts or experiences 'speak for themselves' and maintaining that most truths are as artefactual as they are factual. Questions of cognitive authority and answerability thus figure as prominently as issues of epistemic warrant in these projects, where feminists are concentrating less on formal, universal conditions for making and justifying knowledge 'in general' than on the specificities of knowledge construction. Hence these inquiries are often interdisciplinary, producing detailed analyses of everyday knowledge-making and of scientific or social scientific inquiry; drawing out their gendered and other locational implications. In these projects feminists are showing that avowedly engaged, politically committed investigations can yield well-warranted conclusions.

1 Epistemologies of privilege

Feminist epistemologies, like many other post-Enlightenment, post-colonial epistemological projects, have grown out of critical interrogations of the universalistic presumptions of the theories of knowledge of the Western philosophical tradition. Sceptical about the very possibility of developing a theory of knowledge 'in general', whose claims to universal validity are premised on its abstraction from the specificities of human circumstances, feminist epistemologists have insisted on the constitutive part that epistemic location plays in the making and evaluating of knowledge claims. Many of the best-established post-positivist and neo-rationalist theories of knowledge - together with the conceptions of reason, epistemic agency, objectivity, experience and knowledge itself that comprise their core - tacitly draw their conceptual and theoretical apparatus from an idealized view of the knowledge produced and validated by the (male) occupants of the dominant social, political and economic positions in white Western societies. The consequent distribution of epistemic authority perpetuates a hierarchical order in which women and other 'others' occupy the least authoritative positions. Eschewing traditional a priori approaches with their normative goal of determining what an ideal knower ought to do, feminists are producing critical and self-critical analyses of what variously embodied, historically and materially 'situated' knowers (in Donna Haraway's (1991) phrase) actually do, deriving normative conclusions from critical-constructive readings of epistemic practice.

The dominant epistemologies of modernity, as they have developed out of the Enlightenment with a later infusion of positivist-empiricist principles, have defined themselves around ideals of objectivity and value-neutrality. Ideal

objectivity, in post-positivist times, has come to mean a detached, neutral and disinterested approach to a subject matter that exists in a publicly observable space, separate from knowers/observers and making no personal claims on them. Value-neutrality elaborates this disinterested aspect to insist that knowers must have no vested interest in the object of knowledge and are responsible only to the evidence. Such ideals are best suited to govern evaluations of the knowledge claims of persons whose situations allow them to assume that theirs is a ‘view from nowhere’, that through the autonomous exercise of their reason they can transcend particularity and contingency and the accidents of gendered embodiment. Such persons have usually tended, within the social arrangements of affluent Western societies, to be white and male, though the specificity of their identity and circumstances are usually effaced in their self-presentation as ‘representative’ human subjects.

Ideals of Reason, both in theories of knowledge, and in their trickle-down effects in everyday discourse, have been consistent even across centuries of historical variation in yielding a regulative conception of rationality in which, as Genevieve Lloyd ([1984] 1993) has shown, traits, values and activities commonly associated with ‘the feminine’ are systematically suppressed. Analogously, feminists such as Evelyn Fox Keller (1985) and Susan Bordo (1987) have shown that the psychosocial characteristics that affluent, white, male children in such societies have commonly been nurtured to embody have been just those that prepare them for a lifetime of detached, objective control in everyday and scientific knowledge-seeking, and in a public world of work and deliberation. Instrumental reason and a science-derived ideal of knowledge have shaped conceptions of good epistemic practice, with consequences that are often exclusionary not just of women’s knowledge, but of the knowledgeable activities of other marginalized people. The point is not that women and other ‘others’ cannot emulate the ideal, but rather (as Lloyd has shown) that the symbolisms out of which the ideal acquires its (historically mutable) content work simultaneously to validate the knowledge and epistemic status of would-be knowers whose socio-culturally fostered character traits coincide with the ideal, and to suppress those whose subjectivities are differently produced. In short, feminist genealogies and deconstructions of these ideals have shown that, despite their proclaimed neutrality, they derive from, naturalize and normalize the experiences and social positions of privileged European men and their (male) descendants.

No simple taxonomy of feminist epistemologies classified as distinct or self-contained categories could present an accurate, state-of-the-art picture of these ongoing projects. Yet certain strands run through them, sometimes separately but usually intertwined. Here I elaborate the principal features of an empiricist strand in feminist theories of knowledge, of standpoint positions, and of genealogical and interpretive practices (see Feminism and social science; Gender and science §§1-2, 6).

2 Feminism and empiricism

The relationship between feminism and empiricism, both classical and post-positivist, has been uneasy. Feminists have to work with empirical evidence if they are to move knowledgeably about the physical world and engage effectively with the social, political and ‘natural’ realities that sustain hierarchical social structures. Yet, classical and many contemporary versions of empiricism are constructed around assumptions that are inimical to feminist emancipatory projects (see Empiricism). They are as androcentered as the liberal moral-political theories that they inform (see Liberalism). A woman can claim space within them only in so far as she is prepared and able to be ‘more like a man’ - that is to say, a privileged, able-bodied white man.

It is with the abstract individualism of empiricist orthodoxy, and its residues within everyday conceptions of knowledge in Anglo-American societies, that many feminists take issue. Individualist assumptions yield a picture of knowers as interchangeable in their autonomous epistemic projects, where they are at once sceptical of testimony and unswayed by emotion. Such knowers are ‘individuals’ in their place-holder status as units of analysis, yet they are never individuated. Correlatively, knowledge is objective, universally valid, available impartially and indiscriminately to everyone in identical observation conditions; and observation often reduces to simple, atomic givens, reportable in discrete propositions of the form ‘*S* knows that *p*’ (‘Sara knows that it is raining’). The apolitical character of such utterances, together with their paradigm (and often foundational) status in theories of knowledge, generates an assumption that all knowledge claims worthy of the title will be equivalently apolitical; and that propositional knowledge alone merits the (honorific) title ‘knowledge’.

Yet many classical versions of empiricism and their latter-day analogues come apart around a paradox: for all their alleged grounding in experience, that experience is itself an abstraction. Orthodox empiricists are not equipped,

conceptually or theoretically, to deal with real, idiosyncratic, specifically located experiences. Historical, gendered, locational differences reduce to bias, aberration, errors, to be eradicated and thence discounted in justificatory procedures. When empiricist claims are upheld within social-political structures that deploy a rhetoric of formal equality, yet depend upon structural inequalities to maintain themselves (as do liberal, capitalist societies), the elusiveness of their democratic egalitarianism is apparent. None the less, committed to eradicating uneven distributions of epistemic power and privilege - where gendered unevenness always intersects with the unevenness of race, class, ethnicity and other disprivileged locations - feminists have drawn, albeit critically and selectively, on the resources of many of these same theories to enlist them in transformative and emancipatory projects.

For feminist empiricists, then, the goal of inquiry, both secular and scientific, is to produce knowledge that is neither androcentric nor marked by sexist, racist, classist or other biases. They contend that an unabashedly value-laden yet rigorous empiricism can yield more adequate knowledge than methods whose practitioners are ignorant of their specificity, and of their complicity in a sex/gender system, can yield. Objectivity, in these feminist projects, is disconnected from the universalist assumptions on which analyses of knowledge ‘in general’ depend, and reconfigured around a requirement to take the subjectivity of the inquirers as seriously into account as the objects of inquiry. Thus for feminist empiricists, investigators cannot function as anonymous, isolated and silent spectators. They become answerable for their interventions and epistemic negotiations, as much to the epistemic community as to the evidence; and details about an inquirer’s epistemic location and interests are likewise subject to empirical scrutiny. The central idea is that politically-informed inquiry fosters a better empiricism, and what Sandra Harding (1991) calls a ‘strong objectivity’ that is more objective than an objectivity that defines itself by bypassing the conditions of its own possibility.

According to Lynn Nelson (1990), Quinean empiricism demands neither the stark individualism nor the theory-neutrality of the classical theories; hence it lends itself to feminist reconstructions in which communities, not individuals, are knowers and knowledge claims are entangled in and shaped by webs of belief, testable always against (communal) experience (see Quine, W.V. §2). For Nelson, communities are knowers in a strong sense, for which ‘individual’ experience and knowledge are possible only within a community. The point recalls Wittgenstein’s private language argument: it amounts to a contention that there could be no knowledge, no appropriately justified beliefs, without communal standards of justification and critique (see Private language argument). In a radical rereading of Quinean ‘naturalized’ epistemology, Nelson finds a rich resource for feminist successor epistemology projects (see Naturalized epistemology). Departing from preoccupations with determining whether knowledge is possible, naturalized epistemologies start from an assumption that people can and do have knowledge. Appealing to scientific psychology, they abandon transcendence to examine how people actually know, individually and socially. Yet because they adhere to principles of empirical objectivity and retain a realist commitment to ‘the evidence’, they are effective in producing reliable knowledge of the physical and social world. For Nelson, the value of these inquiries for feminist epistemologies depends on their being ‘socialized’ to focus on questions about how knowledge is made and adjudicated in gender-inflected social groups. Lorraine Code argues that it also depends on their taking an appropriately critical stance towards residues of a positivistic orthodoxy that informs much of present-day cognitive psychology, with the individualistic, tacitly masculine conception of ‘human nature’ it often presupposes. Jane Duran (1991) contests that conception by advocating a justificatory approach that is both naturalized and gynocentric, appealing to ‘essentially feminine’ principles. Object relations theory, psychoanalysis read through French feminist theory, and cognitive science will inform its studies of how gendered knowers are psychologically produced and reproduced, she maintains.

In the contextual empiricism that Helen Longino (1990) elaborates, evidential reasoning is context-dependent, and data count as evidence only in relation to background assumptions and hypotheses. Knowledge construction is a thoroughly social practice; hence acknowledging the constitutive role of values and ideology in inquiry does not require an indiscriminate tolerance of individual subjective preferences. Objectivity is ensured by high standards of social criticism, which all epistemic products must satisfy. Such criticism can unmask androcentricity and other ‘centricities’ even in ‘good’ science and inquiry, even from an admittedly interested position that is at once open to scrutiny and rigorously committed to working within the limits and multiple possibilities of empirical evidence.

Lorraine Code’s position is residually empiricist in its realism. Yet it departs from canonical versions of empiricism in its conception of socially constructed and interactive subjectivities, located and enacted within

uneven structures of power and privilege; and in situating issues of responsibility firmly within epistemological inquiry. Here the monologic individualism of post-positivist theories gives way to a picture of situated, socially embedded knowers conducting epistemic negotiations across the multiply configured rhetorical spaces of the social-political world. Code proposes that knowing other people is as exemplary an epistemic activity as knowing medium-sized physical objects, and that testimony is as crucial a source of knowledge as perception and memory. She advocates an ecologically modelled epistemology that draws on narrative analyses to position human knowing within interconnected systems of social, natural and other environmental relations.

3 Standpoints, interpretations, genealogies

Feminist standpoint theorists such as Patricia Hill Collins, Sandra Harding, Nancy Hartsock and Hilary Rose contend that neither orthodox nor feminist empiricists can adequately address the historical and material conditions that produce both epistemic agency and knowledge itself. Because the authoritative, standard-setting knowledge in Western societies is derived from and tested against the social experiences and material circumstances of white, middle-class, educated men, women (like the proletariat of Marxist theory) are oppressed in marginal, underclass epistemic positions. Science as practice has created an esoteric discourse to which few women and non-white men gain ready access. It explains their limited success with ‘scientifically proven facts’ about their natural intellectual inferiority: facts that have been established by a methodology not explicitly designed to oppress, but whose oppressive consequences are manifold. Yet it is possible to transform oppression into epistemic advantage. Just as Marxist inquiry started from within the lives of the proletariat to produce historically-materially located analyses that offered a sharply focused lens through which to inspect the social system as a whole, shorn of its naturalistic ideology, so starting from the diverse and often contradictory lives of women casts a critical frame around taken-for-granted epistemic hierarchies and practices of disempowerment. Collins (1990) claims, for example, that a Black feminist standpoint rooted in the everyday experiences of African-American women resonates with the epistemologies of subordinate groups in a multitude of disparate locations. It ‘makes strange’ the taken-for-granted practices of rejecting concrete experiences and testimony in favour of standardized abstractions, and it values wisdom over knowledge more conventionally (propositionally) conceived. And Rose (1983, 1994) maintains that an epistemology that bypasses the hands-on, affectively engaged labour that attests to a profound knowledge of the nature of things cannot hope to be adequate to the knowledge that enables people to work and live well.

A standpoint is more than merely a perspective, one among many. It is an achieved intellectual and embodied political position, forged out of painstaking analyses of the systems that legitimate oppressive practices, and firmly located *in media res*. Standpoints are neither guaranteed innocence, nor are they immune from criticism: their affinities with the consciousness-raising practices of the 1970s work to preserve a critical interpretive stance for which even first-person experiential claims have often to be interrogated, albeit responsibly, in mutually respectful debate.

Some critics claim that because there is no single, unified feminist position, standpoint theory obliterates differences and hence fails by its own feminist standards. Others argue that its ‘locatedness’ merely produces a perspective that is as limited as any other. Yet few self-identified standpoint theorists would assume that there can be a single privileged speaking position or one unified epistemic voice. Because they work from within the specificities of women’s lives and are resistant to the reductivism that obliterates lived, practical differences, they take limitation and partial perspective as facts within which epistemologies must be produced, not as obstacles they must seek to overcome. Points of commonality from one material, embodied location to another produce sites of feminist solidarity and opportunities for strategic coalition-building; points of difference confirm the necessity of ongoing critical negotiation.

In the interpretive and genealogical projects of postmodernity, feminists such as Linda Alcoff and Susan Hekman have found critical-constructive tools that work sometimes in concert with empiricist and standpoint theories, sometimes in tension with them. Interpretive strategies, with their origins in Gadamerian hermeneutics, contest any claim to the effect that experience, evidence or texts speak for themselves (see Gadamer, H.-G.). It is philosophy’s task to interpret the cultural-historical prejudgments out of which knowledge necessarily comes into being: prejudgments which are neither pernicious nor escapable, but which have to be confronted if the knowledge that they inform is to be adequate. Hekman (1990), for example, argues that all knowledge is interpretive; and Alcoff (1996) proposes a coherentist, ‘immanent epistemology’ that contrasts with transcendent, observational

theories of knowledge in revealing connections between power-infused political issues and epistemic justification.

Genealogical inquiry, with its Nietzschean origins and later Foucauldian elaborations, situates knowledge-production within historically changing structures of power, maintaining the radical contingency of currently hegemonic modes of understanding, legitimating, and establishing knowledge claims (see Genealogy). Some theorists, such as Kathy Ferguson, see a tension between interpretation with its presumed quest for one true original meaning, and genealogy with its commitment to unearthing a multiplicity of meanings which gain hegemonic status through complex power structures and strategies. Again, some self-identified feminist empiricists and/or standpoint theorists resist making common cause with these postmodern projects because they fear that such resistance to totalizing, stabilizing theory initiates a slide into a pernicious form of relativism. Others, such as Haraway, Code and Alcoff, see interpretation and genealogy as mutually informative; they draw on these strategies from within feminist inquiries that resist exclusive alignment with any single category or strand.

In their concentration on the real-world effects of knowledge-making within a community of inquirers, many of these feminist positions demonstrate affinities with the epistemological preoccupations that Charlene Siegfried discerns in the thought of the American pragmatists. Siegfried (1996) shows how feminism and pragmatism might make common cause as social movements inspired by the emancipatory potential of the everyday experiences of real, embodied and active knowers.

4 Implications

If feminist critiques are taken seriously, then epistemologists cannot assume that ‘reason is alike in all men’; nor can they represent knowers as mere place-holders in an infinitely replicable process, whose minds convert information, mechanically and indifferently, into knowledge. Many of the dichotomies around which western philosophy has been built are pulled apart under feminist scrutiny. Feminists have shown that such hierarchically ordered pairs as reason/emotion, mind/body, abstract/concrete, objective/subjective, theory/praxis, universal/particular - and of course male/female - with their persistent veneration of the first and devaluation of the second term, have sustained conceptions of knowledge and subjectivity that are neither dictated by natural necessity, nor exhaustive of the available options. Knowledge is as much body-dependent as it is mind-dependent; fostered as much by well-schooled emotions as by a well-honed reason; at once subjective and objective; its universal and abstract claims are only as good as their concrete and particular manifestations allow. Working across the territory opened out by the revealed instability of these dichotomies, feminists have resituated knowledge-constructing practices within human lives. They have reclaimed testimony as a source of knowledge as valuable (and as fallible) as perception and memory (see Testimony); and have challenged the divide that is often thought to separate ‘knowing how’, ‘folk wisdom’, and narrative knowledge from ‘knowledge properly so called’. Feminists have demonstrated the effectiveness of dialogic, negotiated epistemic deliberation and the integrity of local knowledge, contrasted with monologic or reductive pronouncements and projects designed to affirm the unity of science. Feminists are not alone in articulating these challenges: neo-pragmatists, deconstructionists and hermeneuticists, naturalistic epistemologists, Marxists, narrative theorists and reassessors of relativism, to mention only a few, often make good scholarly allies (see Cognitive pluralism; Epistemic relativism). But few of their projects are explicitly gender-sensitive to the extent that feminist inquiry has to be: hence feminist voices cannot simply speak in chorus with these other voices. Theirs are distinctive, and often necessarily dissident, speaking parts.

None the less, most late-twentieth-century feminists resist affirming simplistic, across-the-board alignments, say of men with objectivist, distanced, positivistic, scientific methods and women with subjectivist, connected, interpretive, non-scientific methods. Nor do they opt for an essentialism that would identify quantitative methodologies as male, qualitative ones as female; positivism as male, interpretation as female. Few would endorse a wholesale science-bashing that smacks more of ideological excess than of a genuine quest for knowledge. Even feminists who are wary of the oppressive effects of scientific and other authoritative knowledge rely upon its (intermittent) successes. Yet feminists are also convinced, on good evidence across a range of disciplines and areas of inquiry, that feminist research makes a difference. Hence questions about ‘method’, which are central to feminist epistemological debates become, in effect, questions about what difference feminism makes, and how. Its answers come from as many directions as there are epistemic practices. They are not reducible to a single, closed set of ideals and principles, for the era of theoretical and methodological monotheism has passed.

But they provide nodal points for ongoing deliberation and active engagement with social and material issues.

See also: Feminism and psychoanalysis; Feminist political philosophy

LORRAINE CODE

References and further reading

Alcoff, L. (1996) *Real Knowing: New Versions of the Coherence Theory*, Ithaca, NY: Cornell University Press. (Breaks down the analytic-continental divide in a normative epistemology that takes power and desire as epistemically salient.)

Alcoff, L. and Potter, E. (eds) (1993) *Feminist Epistemologies*, New York: Routledge.(Elaborations and refinements of most of the principal approaches to feminist epistemology in the early 1990s.)

Antony, L. and Witt, C. (eds) (1993) *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, Boulder, CO: Westview Press.(Critical responses to contentions that the domain of reason is masculine.)

Bordo, S. (1987) *The Flight to Objectivity: Essays on Cartesianism and Culture*, Albany, NY: State University of New York Press.(A psycho-historical reading of the development of the role of Cartesian doubt in modern philosophy.)

Code, L. (1987) *Epistemic Responsibility*, Hanover, NH: University Press of New England.(Discussion of responsibility issues in epistemic communities.)

Code, L. (1991) *What Can She Know? Feminist Theory and the Construction of Knowledge*, Ithaca, NY: Cornell University Press.(Analysis of the androcentric character of traditional epistemologies, an examination of the politics of knowledge, and a sketch of new directions for theory of knowledge.)

Code, L. (1995) *Rhetorical Spaces: Essays on (Gendered) Locations*, New York: Routledge.(Addresses questions about testimony, empathy, knowing other people, and epistemic authority in specific, power-infused situations.)

Collins, P.H. (1990) *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*, London: Harper Collins.(Analysis of the social construction of Black feminist thought and of Black feminist standpoint epistemology.)

Duran, J. (1991) *Toward a Feminist Epistemology*, Savage, MD: Rowman & Littlefield.(A discussion of the resources of 'naturalized' epistemology and a proposal for a gynocentric epistemics.)

Ferguson, K. (1993) *The Man Question: Visions of Subjectivity in Feminist Theory*, Berkeley, CA: University of California Press.(Critical reading of interpretive and genealogical strategies.)

Haraway, D. (1991) 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective', in *Simians, Cyborgs and Women: The Reinvention of Nature*, New York: Routledge.(Argues for an embodied, specifically situated, critical feminist approach to knowledge questions.)

Harding, S. (1986) *The Science Question in Feminism*, Ithaca, NY: Cornell University Press.(Classifies feminist epistemologies as empiricist, standpoint, and postmodern.)

Harding, S. (1991) *Whose Science? Whose Knowledge? Thinking from Women's Lives*, Ithaca, NY: Cornell University Press.(Elaborates a conception of strong objectivity, and refines the 1986 account of standpoint theory.)

Hartsock, N. (1983) *Money, Sex, and Power: Toward a Feminist Historical Materialism*, Boston, MA: Northeastern University Press.(Develops a standpoint epistemology from a Marxist starting point.)

Harvey, E. and Okruhlik, K. (eds) (1992) *Women and Reason*, Ann Arbor, MI: University of Michigan Press. (Critical feminist rereadings of Western ideals of reason and rationality.)

Hekman, S. (1990) *Gender and Knowledge: Elements of a Postmodern Feminism*, Boston, MA: Northeastern University Press.(Advocates a postmodern, deconstructive and genealogical epistemological approach.)

Keller, E.F. (1985) *Reflections on Gender and Science*, New Haven, CT: Yale University Press.(Exposes the masculine assumptions in the history of western science.)

Lennon, K. and Whitford, M. (eds) (1994) *Knowing the Difference*, London: Routledge.(Postmodern essays on feminism and epistemological questions.)

Lloyd, G. (1984) *The Man of Reason: 'Male' and 'Female' in Western Philosophy*, Minneapolis, MN: University of Minnesota Press; 2nd edn, 1993.(Traces the symbolisms that construct dominant conceptions of reason and masculinity throughout the history of western philosophy.)

Longino, H. (1990) *Science As Social Knowledge*, Princeton, NJ: Princeton University Press.(Shows that background assumptions shape scientific knowledge, and that communities are the principal knowers.)

Nelson, L.H. (1990) *Who Knows: From Quine to a Feminist Empiricism*, Philadephia, PA: Temple University Press.(Argues that Quinean empiricism is a valuable resource for a feminist-informed empiricism in which

communities are the primary knowers.)
Rose, H. (1983) ‘Hand, Brain and Heart: A Feminist Epistemology for the Natural Sciences’, *Signs: Journal of Women in Culture and Society* 9 (1): 73-90.(Argues for a praxis-based feminist standpoint approach to knowledge issues.)
Rose, H. (1994) *Love, Power and Knowledge: Towards a Feminist Transformation of The Sciences*, Cambridge: Polity Press.(Expands and elaborates standpoint theory.)
Siegfried, C.H. (1996) *Pragmatism and Feminism: Reweaving the Social Fabric*, Chicago, IL: University of Chicago Press.(Sets up a dialogue between American pragmatism and feminist epistemology.)

Feminist ethics

Critics greet feminist ethics with suspicion, alleging that it is biased towards the interests of women. Feminist ethicists reply that it is traditional ethics which is biased. As they see it, for centuries traditional ethicists claimed to speak for all of humanity, when they were speaking only or primarily for men, and the most privileged of men at that. In contrast, although feminist ethicists openly admit that they proceed from the perspective of women's experience, their paramount goal is simply to reconstruct traditional ethics so that it becomes more universal and objective by including women's as well as men's moral voices.

Far from being monolithic, feminist ethics encompasses a wide variety of woman-centred approaches to the moral life. Feminine *approaches to ethics, with their stress on personal relationships and an ethics of care, put a premium on the value of human connection.* Maternal *approaches focus on one relationship in particular, that between mothers and children, as the paradigm for moral interaction.* Lesbian *approaches stress choice rather than duty, aiming to define the conditions in which lesbian women can flourish. Finally, specifically* feminist *approaches to ethics emphasize the political task of eliminating systems and structures of male domination and female subordination in both the public and the private domains.*

1 Feminine ethics

Woman-centred approaches to ethics, as well as debates about gender and morality, are hardly new. Thinkers like Mary Wollstonecraft, John Stuart Mill, Harriet Taylor, Catherine Beecher, Charlotte Perkins Gilman, Elizabeth Cady Stanton and others considered the 'woman question' in the eighteenth and nineteenth centuries. They asked whether morality is gendered, and whether moral and psychological traits come from culture or biology. Such thinkers paved the way for twentieth-century debate about whether women's biology and maternal experience predispose them to espouse an ethics of care, while men's biology and public experiences predispose them to espouse an ethics of justice.

For centuries, men wrote most of the great books of ethics, speaking, as might be expected, from a male perspective. The response of those developing feminine approaches to ethics has been to offer the traditionally 'feminine' characteristics of care, compassion, benevolence, nurturing and kindness as no less morally significant than the traditionally 'masculine' characteristics of rationality, justice and independence. Among others, psychologist Carol Gilligan (1982) has claimed that women, for whatever reasons, tend to speak in a voice of care that focuses on concrete relationships, rather than in a voice of justice that focuses on abstract rules (see Virtue ethics §6). In the personal narratives of women contemplating abortion, for example, Gilligan heard more concerns expressed about the personal and professional needs of particular people than about a foetus' right to life or a woman's right to choose. Although Gilligan denies that she regards women's and men's respective moralities as decidedly separate, many of her critics still view her approach to morality not only as a deeply gendered ethics, but also as a highly femininized ethics meant to counter the supposedly 'masculinist' morality of psychologist Lawrence Kohlberg.

Kohlberg, who mentored Gilligan early in her career, constructed a six-stage scale of moral development: while men tend to ascend to at least stage five on this ladder, 'the social contract legalistic orientation', women tend to stop at stage three, 'the interpersonal concordance of "good boy-nice girl" orientation' (1971: 164-5). Doubting that women are in any way men's moral inferiors, Gilligan hypothesized that women did poorly on Kohlberg's scale simply because it was calibrated to gauge male, rather than human moral development. As such, she proclaimed it an inappropriate scale for judging female moral development. Whereas men typically interpret their greater independence and reliance on an ethics of justice as signs of their moral growth, Gilligan speculated that women tend to regard their greater interdependence and reliance on an ethics of care as signs of their moral growth (see Moral development; Moral education §1).

Nell Noddings' relational ethics is another example of a 'feminine' approach to ethics. In any good human relationship, says Noddings (1984), there is the 'one-caring' and the 'cared-for'. The one-caring should engross themselves in the cared-for, focusing on their needs, actions and thoughts. In return, the cared-for should welcome the one-caring's attention and willingly share their own needs, hopes and accomplishments with the one-caring. Noddings believes that caring consists more in specific obligations to particular people than in general duties to

human kind. Concrete actions, far more than abstract intentions, define real caring.

There is reason to suspect, however, that Noddings - like Gilligan - makes caring the special province of women. Although Noddings claims that caring should be the responsibility of all humans, in most of her examples the carers are women, some of whom push their acts of caring to the point where their own identity, needs, and even existence are sacrificed. Thus, many critics fault Noddings as well as Gilligan for over-stressing women's role as carers. As Sheila Mullett points out (1988), since equality of care between men and women is unlikely in a still patriarchal society, caring remains dangerous for women. Until full social, political and psychological equality between the sexes exists, men will be inclined to let women take care of them without reciprocating the favour.

2 Maternal ethics

Critics also express reservations about maternal ethics, another feminine approach to ethics. Virginia Held (1993), Caroline Whitbeck (1974/5) and Sara Ruddick (1980) all offer a good mother-child relationship as a promising paradigm for any relationship. They find the traditional rational-contractor model inadequate (see Contractarianism). After all, most human connection and interaction does not take place between equally autonomous, powerful and informed adults (typically men). Instead it occurs between unequals, such as parent-child, student-professor, professional-client, doctor-patient, and employer-worker. Therefore, a caring, although unequal, mother-child relationship provides a normative standard for the assessment of relationships that are inevitably imbalanced. Presumably, mothers have their children's best interests at heart. They try to protect and teach them, and make them socially acceptable and successful. In the process, they teach both themselves and their children to be sensitive, responsible people, aware of others' needs (Ruddick 1984) (see Family, ethics and the §5).

Both feminists and non-feminists challenge maternal approaches to ethics, however. Non-feminist critics object to using one relationship in particular as a moral paradigm for all relationships in general, especially a relationship that seems to exclude all men and non-reproducing women. Feminist critics also challenge the adequacy of the mother-child relationship. As they see it, in a patriarchal society, mothers are already laden with guilt, expectations and demands. They doubt that anyone, but especially women, should seek to emulate the all-sacrificing mother who denies her own needs to serve those of her child. Instead people should pattern their relationships on a good friendship relationship (see Friendship). Like the mother-child paradigm, a friendship paradigm rings truer with our moral experience than the abstract rational-contractor relationship; unlike the mother-child paradigm, however, it comes closer to signalling the kind of give-and-take only equals can achieve.

3 Lesbian ethics

The women-centred approach to ethics most specifically tailored for women is clearly the lesbian one. Lesbian ethicists focus unapologetically not simply on women, but on women-loving women. Lesbian approaches are varied and complex, and generalizations are difficult, but most lesbian approaches seem to replace the traditional moral question, 'Is this action good?' with, 'Does this action help my quest for freedom and self-identity?' Although lesbian ethicists stress choice rather than duty, they are not relativists (see Duty; Moral relativism). As Sarah Lucia Hoagland asserts (1988), lesbian ethics concerns community: when a lesbian chooses for herself, she chooses for other lesbians, and those lesbians in turn will choose for her. Ethics, on this view, involves not an individual lesbian doing what she believes is absolutely right, but a community of lesbians weaving moral values together. Moral value and meaning is born and sustained in the spaces between oneself and others. Lesbian ethicists seek to become persons 'who are not accustomed to participating in relationships of domination and subordination' (1988: 241), but who are 'playful' beings with 'the ability to travel in and out of each other's world' (Lugones 1987: 13). To the degree that lesbians learn to move among and delight in a variety of worldviews, they lose interest in using ethics as what Hoagland has termed a 'tool of control' (1988: 246).

4 Feminist ethics

What makes specifically feminist approaches to ethics different? The answer is their strong emphasis on politics. Susan Sherwin aptly comments that a feminist approach to ethics 'applies a specifically political perspective and offers suggestions for how ethics must be revised if it is to get at the patterns of dominance and oppression as they affect women' (1992: 42). Feminist ethicists focus not so much on goodness as on power. They emphasize the ways in which traditional approaches to ethics maintain a status quo oppressive to women.

Many feminists discuss ways in which traditional ethics support women's subordination. Alison Jaggar, for example, discusses five specific ways in which this occurs, in her essay, *Feminist Ethics* (1992). First, traditional ethics focuses on men's rights and interests at the expense of women's. Indeed, traditional ethics has even encouraged women to develop supposed 'virtues' like patience and self-sacrifice that tend to further women's subordination. Second, traditional ethics justifies its neglect of women's interests by claiming that important moral questions rarely arise in women's world, the domestic sphere. Thus, women's double-workday, reproductive burdens, and sexual vulnerabilities, for example, are often dismissed as 'private' matters of little genuine moral significance. Third, traditional ethics often adds insult to injury by alleging that women are men's moral inferiors. Since the time of Aristotle and Plato, to Freud and the present day, (male) philosophers have described women as capable of only limited moral thought and action. Fourth, traditional ethics tends to overrate traditionally masculine traits such as 'independence, autonomy, intellect, will, wariness, hierarchy, domination, culture, transcendence, product, asceticism, war, death', and to undervalue traditionally feminine traits, such as 'interdependence, community, connection, sharing, emotion, body, trust, absence of hierarchy, nature, immanence, process, joy, peace and life' (Jaggar 1992: 364). Finally, traditional ethics tends to employ a 'masculine' way of thinking characterized by rules, abstraction, and impartiality, eschewing 'feminine' ways of thinking as 'emotional', rooted as they are in relationships, particularity, and partiality.

As a remedy for the sexism in traditional ethics, Jaggar suggests a feminist approach to ethics that does at least the following: (1) begins by acknowledging that men and women have different experiences and situations in life; (2) provides guides to action 'that will tend to subvert rather than reinforce the present systematic subordination of women'; (3) offers action guides and approaches to handle issues in both the public and the private realms; and (4) 'takes the moral experience of all women seriously, though not, of course, uncritically' (Jaggar 1992: 366). Thus, feminist approaches to ethics - unlike feminine and maternal approaches - do not call upon women to become exemplary carers. Nor do they, like lesbian approaches, ask women to develop meaning and value only for women-loving women. Instead, feminist approaches call upon women *and* men to overcome gender inequality and oppression. Therefore, they offer something of value to all human beings - for care, justice, and moral value can flourish only in a truly equal and balanced world.

See also: Feminism; Feminist epistemology; Feminist jurisprudence; Feminist political philosophy; Nursing ethics; Sexuality, philosophy of

ROSEMARIE TONG

References and further reading

Gilligan, C. (1982) *In a Different Voice: Psychological Theory and Women's Development*, Cambridge, MA: Harvard University Press, 76-92.(Claims that, on the average, and for a variety of cultural reasons, women tend to espouse an ethics of care that stresses relationships and responsibilities, whereas men tend to espouse an ethics of justice that stresses rules and rights.)

Held, V. (1993) *Feminist Morality: Transforming Culture, Society, and Politics*, Chicago, IL: University of Chicago Press.(Argues that society would be more moral if the ideal of the 'autonomous man' were replaced with the ideal of the 'relational woman'.)

Hoagland, S.L. (1988) *Lesbian Ethics*. Palo Alto, CA: Institute of Lesbian Studies. (Argues that choice, not duty, is the basis of lesbian ethics.)

Jaggar, A.M. (1992) 'Feminist Ethics', in L. Becker and C. Becker (eds) *Encyclopedia of Ethics*, Lawrence, KS: University Press of Kansas.(Identifies the necessary conditions for a feminist ethics.)

Kohlberg, L. (1971) 'From Is to Ought: How to Commit the Naturalistic Fallacy and Get Away With It in the Study of Moral Development', in T. Mischel (ed.) *Cognitive Development and Epistemology*, New York: Academic Press, 164-5.(Proposes a six-stage interpretation of moral development, which is taken to be universal, invariant and hierarchical.)

Lugones, M. (1987) 'Playfulness, "World"-Travelling, and Loving Perception', *Hypatia* 2: 13.(Argues that we cannot be moral unless we are willing to see reality through eyes different from our own.)

Mullett, S. (1988) 'Shifting Perspectives: A New Approach to Ethics', in L. Code, S. Mullett, C. Overall (eds) *Feminist Perspectives: Philosophical Essays on Method and Morals*, Toronto: University of Toronto, esp. 119. (Presents the moral agent as constructing a moral perspective within the context of a collective endeavour to transform an oppressive *status quo*.)

Noddings, N. (1984) *Caring: A Feminine Approach to Ethics and Moral Education*, Berkeley, CA: University of

California Press, esp. 9.(Argues that ethics concerns caring relationships between particular individuals.)
Ruddick, S. (1980) ‘Maternal Thinking’, *Feminist Studies* 6 (2); shortened version repr. in J. Trebilcott (ed.) *Mothering: Essays in Feminist Theory*, Totowa, NJ: Rowman & Allanheld, 1983, 213-30. (Describes the three fundamental goals of maternal practice as being the preservation, growth, and making acceptable of children.)
Sherwin, S. (1992) *No Longer Patient: Feminist Ethics and Health Care*, Philadelphia, PA: Temple University Press.(Shows how raising the ‘woman question’ makes a moral difference in the world of medicine.)
Tong, R. (1993) *Feminine and Feminist Ethics*, Belmont, CA: Wadsworth.(Explains the difference between a ‘feminine’ ethics of care and a ‘feminist’ ethic of power.)
Whitbeck, C. (1974/5) ‘The Maternal Instinct’, *Philosophical Forum* 6 (2, 3); repr. in J. Trebilcott (ed.) *Mothering: Essays in Feminist Theory*, Totowa, NJ: Rowman & Allanheld, 1983, 185-98.(Suggests that men’s and women’s different biological experiences typically affect the intensity of their respective attachments to their children.)

Feminist jurisprudence

The diversity of feminist philosophy and theory is represented in feminist jurisprudence, but two models of feminist jurisprudence predominate: the parity model, according to which women should be given legal parity with men; and the transformative model, which proposes the transformation of male legal categories and concepts to address women's experiences. The parity model has also been identified by the terms 'the male monopoly of law', or 'law as male bias'. The transformative model has sometimes been equated with feminist jurisprudence per se, *sometimes more specifically with US feminist jurisprudence. The two models differ primarily in their response to the claim of liberal jurisprudence - a claim made by law itself - that law is a neutral, rational and fair institution which defends individual liberty and treats people equally. The models also differ over the analysis of rights, and over the place of feminist jurisprudence in the legal curriculum. Both models have been subjected to a subversionist critique of any form of feminist jurisprudence.*

The parity model supports the values of liberal jurisprudence as imputed to law, but identifies a discrepancy between those liberal values and legal practice, such that women are not accorded parity with men. It follows either that law must be persuaded to apply these standards more rigorously in the case of women or that liberal values must be revised to recognize gender as a source of social injustice. The objective is to give women genuine, as opposed to nominal, equal rights or, where their special social situation demands it, special rights. On this model, courses in feminist jurisprudence comprise what have come to be known as 'women and law' studies which generally promote the visibility of women in jurisprudence. These studies may include documentation of law's discrimination against women, analyses of law's male bias against women, and reviews of all liberal jurisprudence which omits reference to gender.

The transformative model also notes the discrepancy between the liberal values imputed to law and law's treatment of women but recognizes the limitations of attempting to close the gap between liberal jurisprudence and legal practice either by making law apply liberal principles more scrupulously in the area of gender or by revising liberal principles. Instead, feminists working with this model argue that liberal jurisprudence can make no impact on law's treatment of women so long as legal categories, such as crime or family law, and legal concepts, such as provocation or marriage, embody male norms and accordingly fail to address women's experiences. It follows that such legal categories and concepts must be transformed to address women's social position and experiences. In so far as rights discourse embodies male norms, it too must be transformed. On this view, courses in feminist jurisprudence comprise the transformation of broad legal categories and specific legal concepts so that they engage with the reality of women's lives.

The subversionist critique seeks to undermine both the parity model and the transformative model. This critique questions the value of feminist jurisprudence for feminist politics. The reason given is that to work within the paradigm of jurisprudence is to legitimate the strategy of recourse to law as the proper means of solving social problems, the very strategy which both the parity model and the transformative model have exposed as inadequate. The subversionist critique recommends instead that feminists subvert the paradigm of jurisprudence, if necessary by abstaining from engagement with it. The use of rights discourse becomes a tactical calculation, and the inclusion of feminist jurisprudence in the law curriculum is a dubious strategy for feminists. The subversionist critique has been criticized in its turn for undervaluing the achievements of liberal legal systems and liberal jurisprudence.

1 The parity model

Central to the parity model of feminist jurisprudence is the theory of law as an institution for the defence of individual liberty and the resolution of disputes between individuals which upholds the values of neutrality, rationality and fairness, all individuals being equal before the law. This doctrine is fundamental to law's conception of itself, as expressed in the ideology of the rule of law, and it provides the rationale for the introduction of equality legislation, such as anti-discrimination and equal pay legislation. It is also typical of contemporary jurisprudence dedicated to the development of liberal theories of justice. For example, Rawls (1972) is a major exponent of the liberal jurisprudential doctrine of justice as fairness. Similarly, Dworkin (1978) is concerned to develop a jurisprudence according to which all people have the fundamental right to be treated as

equals (see Dworkin, R.; Rawls, J. §1).

The parity model of feminist jurisprudence subscribes to such liberal doctrines, but identifies a discrepancy between liberal principles and legal practice such that, in failing to uphold the values of neutrality, rationality, fairness, and liberty and equality in their treatment of women, neither the law nor contemporary liberal jurisprudence ascribes parity to women (see Law and morality §2). The parity model offers two main types of argument to support that claim: the socio-legal and the philosophical.

First, Sachs and Wilson (1978) have amassed a wealth of socio-legal materials to show that the law has systematically operated as a male monopoly. They examine the 'persons' cases in the UK and the 'monopoly cases' in the USA, which focus on the law's resistance to extending the category of the legal person to include women. They argue that, through the biased use of judicial discretion, the judiciary has pursued male interests in maintaining obstacles to women's entry into public spheres of citizenship and the professions, notably medicine and law.

The work of Sachs and Wilson has provided the impetus for many similar socio-legal studies revealing the discrepancy between anti-discrimination legislation, such as the principle of equal pay for work of equal value, and the actual benefits experienced by women (see Discrimination). Feminists have identified sex stereotyping as a key mechanism for maintaining this discrepancy. For example, Atkins and Hoggett (1984) argue that the law has consistently worked on the assumption that women are primarily dependent on men, and that their lives are defined by their domestic and sexual roles. These socio-legal materials are the basis for the parity model's claim that, in its failure to uphold its own proclaimed values of equality and neutrality, the law continually fails to accord parity to women.

Here, the parity model urges ever more vigilance in the pursuit of genuine equality for women and their unbiased treatment in legal practice. The strategies include general vigilance for discrepancy between law's stated values and intentions and its actual practice in relation to women, political pressure for improved anti-discrimination legislation, the introduction of gender issues into legal education, and reform of the judiciary and parliament so that women's issues are given greater recognition.

Second, feminist philosophers have challenged liberal jurisprudence for its articulation of the values of liberty and equality in ways which neglect the inferior social status of women. For example, Pateman (1988) has criticized Rawls' theory of the social contract, in particular his use of the 'original position'. She argues that Rawls' attempt to justify our intuitions about contemporary society incorporates implicit but unjustifiable assumptions about gender and sexual relations, thereby obscuring 'the sexual contract' which provides the foundation of modern patriarchy and the political right of men over women. Similarly, Okin (1989) has pointed to the omission of family justice and gendered family structures from contemporary theories of social justice. This omission reinforces the notion of justice as a public virtue, and, in perpetuating women's exclusion from social justice, impedes progress towards a just and democratic society (see Justice §6).

The parity model of feminist jurisprudence here requires revision of liberal jurisprudence so that the social significance of gender inequalities is given due recognition. Of particular importance is the need to extend the scope of liberal jurisprudence so that it is not restricted to the public sphere of social justice but extends also to the private sphere. By pointing to the limitations of liberal jurisprudence in this regard, feminists hope to undermine the assumption that the attribution of formal equality to women is sufficient to ensure that women enjoy material equality with men. Working within the parity model, feminists have identified the discrepancy between women's formal equality, for example in labour legislation, and their material inequality, for example through financial disadvantage incurred by their assumption of domestic and child-rearing responsibilities.

This concern with the neglect of the private sphere is typically expressed in terms of rights. For while the parity model accepts the claim of liberal jurisprudence that rights are the inalienable moral possessions of equal individuals, feminists have pointed out that many rights have not been fully enjoyed by women. While liberal jurisprudence articulates concepts of rights situated in the public sphere, such as civil rights, the parity model emphasizes rights which should be enjoyed in the private sphere, such as the right to protection against domestic violence (see Privacy §4).

A corollary of this view is scepticism about campaigns for women's rights expressed in terms of equal rights when

these are inappropriate to women's social situation. Here supporters of the parity model argue that women need special rights in recognition of their different social position from men's, particularly in the sphere of reproductive rights and in the institution of the family. Yet there are difficulties in determining when special rights are more appropriate than equal rights. Accordingly, some feminists within the parity model have become disillusioned with the potentially divisive nature of rights and argue that rights discourse must be transcended in order to achieve the goals of a communitarian society in which persons can live as genuine equals with respect for their different talents (see Rights).

On the parity model, courses in feminist jurisprudence comprise what has come to be known as 'women and law' work: cataloguing the failures of anti-discrimination law and equality legislation, the operation of male bias and sexist stereotypes (for example, in family law), the shortcomings of legislation and legal practice in the areas of abortion, rape and domestic violence, and exposure of the legal profession's exclusion of women. The parity model of feminist jurisprudence also includes the revision or adaptation of liberal jurisprudence to produce analyses which draw attention to gendered social institutions, such as the family, as major sources of social injustice.

The parity model has been subject to the criticism that the discrepancy between liberal principles and legal practice cannot be made good either by exhorting law to uphold them more scrupulously or by revising liberal principles. The reason given is that, by concentrating on the discrepancy between liberal principles and legal practice, and by giving prominence to 'women's issues', the parity model leaves intact existing categories of law, such as crime and family law, and existing concepts of law, such as provocation and marriage. These criticisms are more fully developed by feminists working within the framework of the transformative model of feminist jurisprudence.

2 The transformative model

The transformative model follows the parity model in noting the discrepancy between the liberal values imputed to law and law's actual treatment of women. But feminists working with the transformative model argue that the attempt to make good that discrepancy either by reforming legal practice or by amending liberal jurisprudence is doomed. The attempt is doomed because it leaves undisturbed broad legal categories and legal concepts which embody male norms and which consequently fail to address the reality of women's social position and experiences. It follows that these legal categories and concepts must be transformed so as to address the concrete reality of women's lives. There are two main versions of the transformative model of feminist jurisprudence: the systematic and the thematic.

The systematic version is associated almost exclusively with the work of MacKinnon (1989). She argues that the legal manifestation of the liberal state is institutionalized male power based on male sexual dominance of women. She identifies two main ways in which law effects male domination. The first is by the eroticization of sexual abuse, so that legislation on rape, prostitution and obscenity is proposed ostensibly for the protection of women but works in practice to protect men's sexual power over women. The second is by law's claim to the standard of neutrality. Here MacKinnon argues that it is futile and dangerous for feminists to demand that law apply this standard. For this legal standard is itself male. Furthermore, the greater the *prima facie* neutrality of law - for example, in equality legislation governing paid work - the more effectively 'neutrality' works as a key mechanism for masking the law's defence of male domination, for example, by requiring women to fit into an economic system which denies them substantive equality.

For these reasons, MacKinnon believes that feminists have wrongly made their demands in terms of rights, whether equal or special, because the pursuit of rights requires women to aspire - in vain - to men's rights. Rights are little more than the expression of institutionalized male power. Consequently, MacKinnon advocates the development of a feminist jurisprudence, namely the systematic inversion of the complete and comprehensive institutionalization of male power and ideology, all derived from male sexuality. This feminist jurisprudence would have a central place in legal education.

Other feminists working with the transformative model, however, have reservations about the project of an exclusively feminist jurisprudence. They suspect that it is a form of essentialism, namely the view that masculine law and feminine law are discrete and unitary spheres. On this view, MacKinnon's feminist jurisprudence is essentialist because it theorizes women as the permanent victims of male sexual power. Instead, these feminists

have advocated the development of thematic feminist jurisprudence through the strategic transformations of key areas of law. This exercise is characterized by two linked approaches: first, the transformation of legal categories; and second, the transformation of legal concepts.

Key exponents of the transformation of legal categories are Graycar and Morgan (1990). They show how it is not possible to consult a traditional legal textbook for the collected legal rules and regulatory practices concerning, for example, domestic violence. Relevant material will be found in textbooks on criminal law, family law, housing and social security rules, tort law, and in some jurisdictions constitutional law. These classifications, however, are meaningless for women who experience domestic violence. They and their lawyers typically need to identify an immediate practical remedy, one which can be rapidly selected from a compendium of legal doctrines and rulings related specifically to domestic violence. Graycar and Morgan accordingly engage in a fundamental recategorization of law according to themes such as women’s sources of financial income, legal responses to motherhood, and gendered injuries to women, such as rape.

The transformation of legal categories may well involve the transformation of specific legal concepts. For example, feminist scrutiny of criminal law from the point of view of women’s experiences has focused on provocation as a partial defence open to women and men charged with murder. For the charge to be reduced to one of manslaughter, the accused needs to prove that the act of killing followed a sudden and temporary loss of control, such as might be experienced by a person goaded beyond endurance. Feminists have argued that this concept of provocation embodies male norms of physical strength and anger. One feminist response, articulated by the pressure group Rights of Women (1992), has been to recommend the introduction of a new partial defence to murder, namely self-preservation. This defence would be open to persons who kill someone who has subjected them to continuing sexual and/or physical abuse and intimidation, to the extent that they honestly believe they have reached a point where there is no protection or safety from the abuse and that they will die if the aggressor remains alive.

A further example of the transformation of a specific legal concept to reflect women’s lives is Howe’s use of the term ’social injury’ (1991). This concept has had an established use in criminology, where it was used to broaden the definition of crime to include injury to the state, thereby bringing white-collar crime within the scope of criminology. Howe appropriates this concept for feminist jurisprudence with the identification of social injuries which are gender-specific, or social injuries by gender stereotype. The transformed concept can then be used to theorize not only offences such as rape, sexual assault and domestic violence but also superficially gender-neutral offences under anti-discrimination legislation and divorce law reform.

The scrutiny of legal categories and concepts for the extent to which they embody male norms will also include the investigation of rights discourse in law-oriented feminist campaigns. Unlike the systematic transformative model, however, the thematic transformative model makes no a priori judgment about whether feminists should continue to use appeals to rights. Rather, each instance of rights discourse requires evaluation. Sometimes the appeal to a right has great popular appeal for a specific cause; none the less it may have damaging implications. For example, justifying abortion as a woman’s right to choose has been a powerful slogan for law reform. At the same time, it invites the assertion of counter-rights, such as fathers’ rights and foetal rights, which are potentially damaging to women. Instead, feminists have argued, rights discourse in the context of abortion needs to be transformed into ‘safe strategies’ which emphasize not the individual choice of a pregnant woman but social responsibility for women in a state of emotional distress and economic need.

The principle that feminist jurisprudence should transform legal categories and concepts to reflect the reality of women’s lives is carried into the legal curriculum. At Dahl’s law school in Oslo, there are specialist courses in women’s law (1987). These organize materials from existing law into collections which follow the typical pattern of women’s experience, such as money law, law surrounding birth, and housewives’ law. Further, all mandatory courses in the law school are required to include feminist legal analyses. Where such radical measures are not possible, however, feminists have also seen the merits of a process of ‘infecting’ the traditional legal curriculum with feminist perspectives.

The main criticism of the transformative model, whether in its systematic or thematic form, is that feminist jurisprudence is of questionable value for feminist politics. This is because, however much feminists seek to transform legal categories and concepts, feminist jurisprudence remains a form of jurisprudence. As such it will

retain the notion, common to all forms of jurisprudence, that law is the proper site for the resolution of social conflicts. Yet feminist analysis has shown that this strategy has been unreliable for the achievement of feminist goals. This criticism constitutes the subversionist critique.

3 The subversionist critique

This critique does not comprise a third model of feminist jurisprudence; rather it constitutes a warning to feminists about the impressive resilience of jurisprudence in the face of critical analyses in and from other disciplines. Of special note is its capacity to absorb and redefine interventions issuing from feminist analyses.

On this view, every engagement with law, either as an object of theoretical analysis or as a political reality, needs to be reprimanded by the reminder that, by entering on to the terrain of law, and by implication the field of jurisprudence, the endeavour may be, and probably is, conceding territory for law’s hegemony. To the extent that it is incorporated within the paradigm of jurisprudence, feminist jurisprudence will reproduce the assumption that law is the proper terrain for feminist politics. Feminist jurisprudence is therefore in danger of collusion with jurisprudence. For example, according to the subversionist critique, the feminist recommendation that criminal law introduce self-preservation as a new partial defence to murder is helpful to women only when they are already embroiled in certain criminal law procedures; it leaves unanswered the question of whether the courts are the appropriate context for solving problems of domestic violence.

A further example of the subversionist critique draws attention to feminist engagement with the policy debate on how women should be supported financially after divorce. Smart (1984) shows that the policy debate is constructed so that the policy options are reduced to only two alternatives: support by former husbands or support by the state. Smart has argued that neither option permits a satisfactory answer, because feminists have argued both for women’s financial independence from men and for recognition of the value to their husbands of women’s unpaid domestic labour in the fixing of maintenance payments.

It follows easily from this example of the subversionist critique that formulating feminist politics in terms of equal rights, or indeed in terms of special rights, is unlikely to produce justice for women. For rights are law’s preferred language, and they are tainted with law’s claim to sole authority to adjudicate disputes between individuals who advance competing rights claims. Similarly, the inclusion of feminist jurisprudence into legal education carries all the risks of assimilation - the depoliticization of feminism in the academic sphere.

The subversionist critique has itself been criticized for its tendency to undervalue the genuine achievements of liberal legal systems and of liberal jurisprudence. For example, it undervalues relatively liberal abortion legislation, anti-discrimination legislation, and greater awareness on the part of the judiciary of the need to strengthen the position of vulnerable female cohabitants. Similarly, the subversionist critique misguidedly undermines the success of feminists working with the transformative model of feminist jurisprudence in forcing liberal jurisprudence to acknowledge gender as a significant social division impeding progress to greater social justice.

See also: Critical Legal Studies §5; Feminist political philosophy; Law, philosophy of

E.F. KINGDOM

References and further reading

Atkins, S. and Hoggett, B. (1984) *Women and the Law*, Oxford: Blackwell.(The classic text of the parity model of feminist jurisprudence.)

Dahl, T.S. (1987) *Women’s Law: an Introduction to Feminist Jurisprudence*, Oslo: Norwegian University Press. (Pioneering exponent of the thematic transformation model of feminist jurisprudence.)

Dworkin, R. (1978) *Taking Rights Seriously*, London: Duckworth.(Criticized by Okin for neglect of family justice.)

Graycar, R. and Morgan, J. (1990) *The Hidden Gender of Law*, Annandale, New South Wales: The Federation Press, also available through The Blackstone Press. (The most extensive collection of legal materials exemplifying the thematic transformation model of feminist jurisprudence.)

Howe, A. (1991) ‘The Problem of Privatised Injuries: Feminist Strategies for Litigation’, in M.A. Fineman and N.S. Thomadsen (eds), *At the Boundaries of Law: Feminism and Legal Theory*, London: Routledge.(A key example of the feminist transformation of a legal concept.)

Kingdom, E.F. (1991) *What's Wrong with Rights? Problems for Feminist Politics of Law*, Edinburgh: Edinburgh University Press.(Text documenting limitations of rights discourse for feminist politics.)

MacKinnon, C.A. (1989) *Toward a Feminist Theory of the State*, Cambridge, MA: Harvard University Press.(The definitive statement of the systematic transformation model of feminist jurisprudence.)

Naffine, N. (1990) *Law and the Sexes: Explorations in Feminist Jurisprudence*, London: Allen & Unwin.(A most accessible account of feminist discourses and law, focusing on gender and class bias in legal constructions of the individual.)

Okin, S.M. (1989) *Justice, Gender and the Family*, New York: Basic Books. (Thoroughgoing criticism of the failure of liberal jurisprudence to address gendered relations in the family.)

Pateman, C. (1988) *The Sexual Contract*, Cambridge: Polity Press.(Influential text revealing the gendered assumptions of Rawls.)

Rawls, J. (1972) *A Theory of Justice*, Oxford: Oxford University Press.(Criticized by Pateman for omission of gender in his account of social justice.)

Rights of Women (1992) *Proposals for Amending the 1957 Homicide Act: Including a New Defence of Self Preservation*, submission to the Royal Commission on Criminal Justice, London: Rights of Women.(A key example of the development of a feminist transformation of legal defences to murder.)

Sachs, A. and Wilson, J.H. (1978) *Sexism and the Law: A Study of Male Beliefs and Legal Bias in Britain and the United States*, Oxford: Martin Robertson.(A landmark text of the parity model of feminist jurisprudence.)

Smart, C. (1984) *The Ties that Bind: Law, Marriage and the Reproduction of Patriarchal Relations*, London: Routledge & Kegan Paul.(An early statement of the subversionist critique.)

Smart, C. (1989) *Feminism and the Power of Law*, London: Routledge.(The most influential account of feminist discourses and law. A critique of feminist jurisprudence in chapter 4 and a summary of feminist theories of rights.)

Smart, C. (1991) 'Feminist Jurisprudence', in P. Fitzpatrick (ed.) *Dangerous Supplements*, London: Pluto Press. (The best exposition of the subversionist critique.)

Feminist literary criticism

Feminist literary criticism looks at literature assuming its production from a male-dominated perspective. It re-examines canonical works to show how gender stereotypes are involved in their functioning. It examines (and often rediscovers) works by women for a possible alternative voice. A study of the social suppression and minimalization of women's literature becomes necessary. These questions emerge: What is sexual difference and how has it been represented? How has the representation of woman relied on a presupposition of inequality between the sexes? Is there a feminine essence, biological or otherwise, that produces 'women's writing'? Feminists who believe that a 'woman' is culturally or socially constructed look for evidence of that process in literature. The socio-cultural and politico-economic construction of sexual difference is 'gender.' A study of the difference between sexual and gender difference can establish alliances with gay and lesbian studies. Feminist criticism sometimes relates to psychoanalysis and/or Marxism, criticizing their masculinism and using their resources. It expands into film/video as well as social-scientific or philosophical texts. Feminists sensitive to racism and imperialism demonstrate the culture-specificity of all of the above.

1 History

Although the remote antecedents of feminist literary criticism may be located in many ways, the source-texts of modern feminist criticism are Virginia Woolf in Britain and Simone de Beauvoir in France. The generations that came of age during the Vietnam War found evidence of discrimination against women within the New Left Anti-War movement in the USA. In Britain, a comparable dissatisfaction with Marxism is to be found in works such as Sheila Rowbotham, Lynne Segal and Hilary Wainwright's *Beyond the Fragments* (1979). Women of the 'baby-boom' generation (born at the end of the Second World War), though not necessarily involved in political struggle, became increasingly dissatisfied with the treatment of women at home and in the workplace. This combination of protest and dissatisfaction worked to produce the first wave of feminist literary criticism in the USA and Britain. The first important texts are Elaine Showalter's A *Literature of Their Own* (1977) - echoing Woolf's title *A Room of Her Own* - and Sandra Gilbert and Susan Gubar's *The Madwoman in the Attic* (1979) - establishing Charlotte Brontë's *Jane Eyre* as the model feminist text (see Feminist political theory §2).

British feminist criticism was, at its inception, more marked by a sociological and psychoanalytic stance (see Feminism and social science; Feminism and psychoanalysis). Two path-breaking texts in this mode are *Women Take Issue* (1978) and Juliet Mitchell's *Psychoanalysis and Feminism* (1974). In the field of literary criticism proper, British feminists were reading US books such as Mary Ellmann's *Thinking About Women* (1968).

Although Simone de Beauvoir's *The Second Sex* (1949) was internationally influential as a source-text of feminist consciousness, it had little acknowledged influence on French feminist literary criticism. In France, feminist literary criticism was less disciplinary, more imbricated in the general social field than in the USA, where the professionalization of the academy is more systematic, and the apparent possibility of resistance more academically tolerated. The bringing together of Marx and Freud, or the feminization of philosophy, was the agenda of interventionist intellectual French feminism. The events of May 1968, when workers and students came together and almost toppled the State, were a motive force of French feminism. The groups *Psych et po* (psychoanalysis and politics) and *Cheveux longues idées courtes* (long hair short ideas) convey an idea of these characteristics of the French scene.

By the 1990s, Julia Kristeva, with her new readings of Freudian psychoanalysis, and Helène Cixous, with her writerly assimilation of the philosophy of deconstruction, had come to dominate the French feminist literary-critical scene. (It should be noted that the word 'feminism' denotes a particular group-politics in France, and therefore would not necessarily be acknowledged as a description by either of these women.)

In the USA and Britain, the impact of French feminist theory has divided the feminist literary-critical scene into two broad camps. (The work of the philosopher Luce Irigaray must also be counted as a French influence on literary criticism written in English.) This division is charted in Toril Moi's *Sexual/Textual Politics* (1985). In the field of feminist literary, film and cultural criticism the work of bell hooks (Gloria Watkins) has been a sustained combative voice. Black feminist criticism is in many ways at the cutting edge of British and US feminist thought.

As mentioned earlier, *Women Take Issue* was one of the early texts of British feminist cultural and literary intervention. Emerging as it did from the Birmingham School of Cultural Studies, it was a statement from postcolonial Black British women. (Hazel Carby, one of the chief spokeswomen of that group, has since moved to the USA and works both in the area of restoring the history of literature by African-American women as well as laying down ground-rules for a cultural critique of the representation of Black women, using some of the categories established by the philosopher Louis Althusser and the sociologist Stuart Hall.) In the USA, where the history of slavery, 'settlement' of indigenous peoples and immigration is complex and multi-layered, 'multiculturalist' feminist literary criticism began to appear in the 1980s, although the label is of later provenance. Cherrié Moraga and Gloria Anzaldua's *This Bridge Called My Back* (1983), with its focus on Chicana feminism, was an important early text of cultural criticism. As a result of a change in US Immigration Law in 1965, the influx of Asian immigration increased 500 per cent. It was among second-generation women of this group that a specifically 'postcolonial' multiculturalist feminism took root (see Multiculturalism §3). A not uncritical but fruitful exchange between the former British and French colonies and the metropolitan countries has been the result. In the British and French context might be mentioned Buci Emecheta from sub-Saharan Africa, Assia Djebar and Nawal el-Saadawi from North Africa, and Meenakshi Mukherjee from India. Feminist literary criticism from Japan and the Philippines is closely interactive with feminist criticism in the USA.

The relationship between feminist literary and cultural critique and the women's movement in the various countries mentioned here is a subject that must be taken into consideration in any extended study. Luce Irigaray (and Michel Foucault) in France, cultural workers such as Pratibha Parmar in Britain, and theorists such as Judith Butler in the United States have shown how an apparently empirical positioning outside heterosexuality can provide new presuppositions for general ethical, social, legal and cultural speculation. Their work is important for those feminists who feel that the alliance with gay and lesbian studies means a fuller critique of the assumption that the historico-legally sanctioned 'male' perspective is objective and just.

In colloquial academic parlance, and taking mainstream Britain and the USA as our examples, 'sixties feminism' was marked by protest and 'consciousness-raising' (collective self-education through sharing personal experience), 'seventies feminism' was Eurocentric and essentialist (emphasizing benevolent middle-class women as women as such); 'eighties feminism' was obliged to take race-class-gender and 'theory' into account; and 'nineties feminism' was marked by a post-feminist and/or multiculturalist impatience by the generation that grew up on, or on the other side of, the gains of 'white' feminist struggle. Feminist literary criticism relates to this general movement. A vade mecum for the twenty-first century is yet to be established.

2 Theory

If the feminine is ungeneralizable as the human, what is the validity of the human sciences? This is the first question of feminism. All methods that question the validity of the apparently reasonable can therefore potentially be its ally. This is how Marxism, psychoanalysis, the general critique of 'humanism', multiculturalism and multi-ethnicity (questioning 'Europe's claim to the culture of Reason), and gay theory (questioning the heterosexual male as the model of man), have contributed their methodological resources to feminism. At the same time, their theory and practice have also been questioned for participating in masculinism. Mainstream feminism, which builds on the strengths of liberal humanism, adopts Marxism via the Engelsian narrative of the origin of the family (Friedrich Engels, *The Origin of the Family and Private Property*; Gayle Rubin, 'The Traffic in Women') or the theory of the modes of production (Christine Delphy, *Close to Home*); psychoanalysis via its Oedipal storyline (Nancy Chodorow, *The Reproduction of Mothering*), and multiculturalism as liberal pluralism with an enlightened civil society at the centre. Feminist literary and cultural criticism are related to these moves.

Lively controversies within feminist literary and cultural criticism are a sign of its strength. The main line of debate is between 'essentialism' and 'anti-essentialism,' although sometimes the rubrics used are different. Is woman a special kind of human being by nature (or essence)? Or is what is understood by the word corresponding to 'woman' in English, a cultural and social construction? These seem to be the broad lines of the controversy.

There is a smaller group of more radical 'anti-essentialist' feminists who presume that the necessary focusing or centring of the thinking and acting subject may not be accessible to direct analysis. Whether such feminist literary and cultural critique can be 'politically engaged' is a question that is not without consequence.

If the non-'European' feminine is ungeneralizable as woman, socio-culturally constructed or otherwise, what is the validity of Eurocentric feminism? (The 'European' here is of Northwest Europe or its legacy: Euro-US, Euro-Australia, Euro-Canada.) The poignancy of this question is that it can only be asked for global attention by women marked by the modern university - an institution developed within the European tradition - or by migrancy - a Eurocentric phenomenon. Testimonies obtained from non-academic Third-World women are sometimes used in an effort to bypass this problem.

See also: Feminism; Feminist aesthetics

GAYATRI CHAKRAVORTY SPIVAK

References and further reading

Beauvoir, S. de (1949) *Le deuxième sexe* (*The Second Sex*), trans. H.M. Parshley, New York: Alfred A. Knopf, 1953.(Womanhood considered from a number of theoretical angles, followed by a contemporary account of woman's life. Source-text of Western feminism.)

Brontë, C. (1969) *Jane Eyre*, ed. J. Jack and M. Smith, Oxford: Oxford University Press. (The classic feminist story of the social and psychological development of a gifted orphan girl.)

Butler, J. (1990) *Gender Trouble: Feminism and the Subversion of Identity*, New York: Routledge.(Source book of lesbian feminist theory.)

Chodorow, N. (1978) *The Reproduction of Mothering: Psychoanalysis and the Sociology of Gender*, Berkeley, CA: University of California Press.(Source book of US-style feminist psychoanalytic theory.)

Delphy, C. (1984) *Close to Home: A Marxist Analysis of Women's Oppression*, trans. D. Leonard, London: Hutchinson.(Considers the economic dimensions of patriarchy.)

Djébar, A. (1980) 'Regard interdit, son coupé' ('Forbidden Gaze, Severed Sound') in *Femmes d'Alger dans leur appartement* (*Women of Algiers in their Apartment*), trans. M. de Jaegar, Charlottesville, VA: University of Virginia Press, 1992, 133-51.(Combines psychoanalytic and postcolonial feminist theory.)

Ellmann, M. (1968) *Thinking About Women*, New York: Harcourt, Brace & World. (Outstanding analysis of common-sense sexism, written in the style of the familiar essay.)

Emecheta, B. (1983) *The Rape of Shavi*, London and Ibuza: Ogwugu Afor.(Fictional allegory of the relationship between cultural construction of African and British women.)

Engels, F. (1884) *Der Ursprung der Familie, des Privateigentums und des Staats* (*The Origin of the Family, Private Property and the State*), ed. E.B. Leacock, New York: International Publishers.(Powerful communist classic of a broad spectrum of influence connecting economic history to gender through nineteenth-century anthropology.)

Gates, H.L. (ed.) (1990) *Reading Black, Reading Feminist: A Critical Anthology*, New York: Meridian Books. (Primary and secondary texts of and on black women writing, suitable for introductory as well as advanced work, for students as well as teachers.)

Gilbert, S.M. and Gubar, S. (1979) *The Madwoman in the Attic: The Woman Writer and the Nineteenth-Century Literary Imagination*, New Haven, CT: Yale University Press.(Path-breaking discussing of *Jane Eyre* as feminist allegory.)

hooks, b. (1992) *Black Looks: Race and Representation*, Boston, MA: South End Press. (African-American feminist cultural theory with the greatest international currency.)

Landry, D. and McLean, G. (1993) *Materialist Feminisms*, Cambridge, MA: Blackwell. (Broad analytic summary of Western feminism.)

Mitchell, J. (1974) *Psychoanalysis and Feminism*, London: Allen Lane.(Unpolemical account of Freud, attempting to make him useful for feminism.)

Moi, T. (1985) *Sexual/Textual Politics: Feminist Literary Theory*, London: Methuen.(A representative collection of essays from the 1950s to the 1980s.)

Moi, T. (1987) *French Feminist Thought: A Reader*, Oxford: Blackwell.(Introduces the important anti-essentialist debate within feminist literary theory.)

Morraga, C. and Anzaldua, G. (1983) *This Bridge Called My Back: Writings by Radical Women of Color*, New York: Kitchen Table.(Mixed-form classic feminist text by women of colour in the USA, antedating today's diasporic feminism.)

Mukhejee, M. (1993) 'Story, History and Her Story', *Studies in History* 9 (1): 72-85. (Variations in the construction of woman as hero in Indian nationalist historiography, literature and autobiography.)

Mulvey, L. (1973) 'Visual Pleasure and Narrative Cinema', in *Visual and Other Pleasures*, London: Macmillan, 14-26.(Lays out principles of feminist psychoanalytic film criticism.)

Rowbotham, S., Segal, L. and Wainwright, H. (1979) *Beyond the Fragments: Feminism and the Making of Socialism*, London: Merlin Press, 2nd edn.(Account of British feminism's disaffection from Marxism in a compelling activist style.)

Rubin, G. (1975) 'The Traffic in Women: Notes on the Political Economy of Sex', in R. Reiter (ed.) *Toward An Anthropology of Women*, New York: Monthly Review Press, 157-210.(Source-text of gender theory.)

Saadawi, N. (1980) *The Hidden Face of Eve: Women in the Arab World*, London: Zed Press. (Hortatory and descriptive text for and about the past, present and future of women of the eastern Arab world. Extensive documentation.)

Showalter, E. (1977) *A Literature of Their Own: British Women Novelists from Brontë to Lessing*, Princeton, NJ: Princeton University Press.(Introduction to feminist literature, establishing the difference between 'female', 'feminine' and 'feminist'.)

Spivak, G.C. (1987) *In Other Worlds: Essays in Cultural Politics*, New York: Methuen. (Source book of postcolonial feminist theory.)

Women's Studies Group, Centre for Contemporary Cultural Studies, University of Birmingham (1978) *Women Take Issue: Aspects of Women's Subordination*, London: Hutchinson.(Early and practical intervention by gays and women of colour in the field of Womens Studies. Excellent example of British Althusserian feminism.)

Feminist political philosophy

In all its forms, feminism asserts that social and political structures in society discriminate against women. Feminist political philosophy aims to show how traditional political philosophy is implicated in that discrimination and how the resources of political philosophy may nevertheless be employed in the service of women. Sometimes, feminist political philosophy extends the arguments of traditional political philosophy to indicate that women are unjustly treated and to propose ways in which that injustice might be removed. This is clearest in liberal feminism, where it is argued that since women are essentially the same as men in being rational creatures, they are entitled to the same legal and political rights as men: arguments which defend the rights of man also support the rights of women. Similarly, Marxist and socialist feminism extend the insights of Marxism and socialism in an attempt to expose and remove the oppression of women: Marxist emphasis on the exploitation of labour under capital is supplemented by Marxist feminist emphasis on the exploitation of women under patriarchy.

However, there are also forms of feminist political philosophy which are more critical of traditional political philosophy and which question the very distinctions upon which it is premised. Thus, radical feminist philosophers question the scope of the term 'political' as it is usually used by political philosophers, and argue that by excluding domestic concerns, traditional political philosophy excludes many of the things which are most important to women. The aim here is not to extend the insights of political philosophy, but rather to highlight the ways in which political philosophy itself shows a distinct gender bias.

Yet more radically, the postmodernists have been critical of philosophy's emphasis on truth and objectivity, and some feminists have extended their arguments to suggest that the very language of philosophy, and by extension of political philosophy, is 'man-made'.

Feminist political philosophy is therefore not one thing but many, and feminist political philosophers are deeply divided as to whether traditional political philosophy may be modified so as to include women's interests, or whether it is itself one of the ways in which women's politically disadvantaged position is legitimized and perpetuated.

1 The scope of feminist political philosophy

Political philosophy is primarily concerned with the justification of the state. It raises questions about what the state may demand of citizens and what citizens, in turn, owe to the state. Feminist political philosophy addresses those same concerns as they affect women. It asks whether and why there are differences between the ways in which men and women are treated by the state. Thus, in the nineteenth century, feminist political philosophers questioned the laws which dictated that men but not women were allowed to vote, and similarly they questioned the laws which prevented married women from retaining rights over their property. Here, feminist political philosophy emphasized the equality of men and women and demanded a justification for their different political and legal treatment.

More recently, feminist political philosophers have noted that even where the laws are gender neutral, women remain at a disadvantage: in most Western liberal democracies there are no longer laws which prevent women from being politically active, but nevertheless there are far fewer women than men in positions of political power or influence. This suggests that attaining full, as distinct from formal, political equality may require something more than laws which are gender neutral. As a result, feminist political philosophers have been concerned to argue for an understanding of the concept of equality which goes beyond the formal equality of gender-neutral laws (see Equality).

These two schools of thought exemplify distinct feminist approaches to political philosophy. In the first case, political philosophy's concern for equality is shown to be incompatible with laws which give women fewer political rights than men. In the second case, political philosophy's formalistic interpretation of equality is questioned. The two examples show how different forms of feminist political philosophy relate to traditional political philosophy: feminist political philosophers sometimes argue for the consistent application of the central concepts of political philosophy; sometimes for a reconsideration of those concepts.

The latter strategy manifests the more subversive character of some feminist political philosophy: traditional concern for justice and equality is often premised on a distinction between public and private life, and concern about the justice of political institutions and practices does not, in general, extend to the private world. But if political philosophy takes as its central concern the justice of political institutions, it needs to explain which institutions count as political and why. In particular, it needs to explain the grounds on which domestic affairs are deemed to be private and outside the domain of the political, outside the scope of theories of justice: for feminists, the decision to differentiate between public and private, between an area which is and an area which is not the legitimate concern of the state, is itself a political decision of some consequence (see Privacy §4). Hence the feminist claim that ‘the personal is political’, for what is meant by this is precisely that the decision as to what will and what will not count as the proper business of the state is a decision which itself has a political dimension.

Feminists who argue that the personal is political may take a further and yet more radical step, questioning the privileged position accorded to the concepts of justice and equality in traditional political philosophy. In private life, concern and compassion may be more important than justice, and if the distinction between the private world and the political world is undermined, then it seems to follow that concern and compassion should be more prominent, and justice and equality less prominent, than they usually are in political philosophy.

The views outlined above indicate that the term ‘feminist political philosophy’ covers a wide range of arguments. Some feminists argue for the consistent application of terms like ‘justice’ to women; others argue for the revision of those concepts; while yet others demand their replacement. At its most radical, feminist political philosophy is highly subversive of political philosophy, since it denies the existence of a separate, political realm or insists that the concept of the political is itself ‘male’. Nevertheless, in all its forms, feminist political philosophy seeks to identify the causes of women’s political disadvantage and to suggest remedies for it.

2 The history of feminist political philosophy

Feminist political philosophy may plausibly be said to begin in the late eighteenth century, with the publication in 1792 of Mary Wollstonecraft’s *A Vindication of the Rights of Woman* (see Wollstonecraft, M.). Wollstonecraft’s arguments owe much to Enlightenment philosophy and she champions vigorously the extension of the rights of man to women. One of her most important and controversial claims is that both men and women are essentially rational beings. Since they are alike in this respect, and since it is this which justifies the ascription of political rights, women should be accorded the same political rights as men. In its time, Wollstonecraft’s assertion that women are rational was highly controversial and had been denied by many of the great philosophers. Most importantly for her, it was denied by Rousseau, J.-J., her contemporary and one of the foremost philosophers of the Enlightenment. On Rousseau’s account, men were rational, whereas women were intuitive and emotional. *A Vindication* is in large part an attack on Rousseau’s political philosophy, and especially on its assertion that women lack the capacity for rationality.

Although it is somewhat anachronistic to refer to Mary Wollstonecraft as a feminist (the term was not used in England until 1894), *A Vindication* does nevertheless display both the critical and the constructive aims of feminist political philosophy: it exposes the ways in which traditional political philosophy excludes women and it presents a vigorous argument in favour of the extension of political rights to women.

However, Wollstonecraft also insists that, although men and women are equally rational and therefore equally entitled to be citizens of the state, they will express their citizenship in different ways. While men will occupy public positions, women will remain primarily within the private or domestic realm, and will express their citizenship in that context. Opinions differ as to the worth of Wollstonecraft’s thought in this area: some see her insistence on women’s role in the home as inappropriately conservative, while others argue that she is in fact a radical before her time, acknowledging the political importance of the domestic sphere. Whatever the correct judgment, we can see in her writings several of the most important features of feminist political philosophy: Wollstonecraft shows a keen awareness of the legal and political disadvantages which women suffered in the later eighteenth century, she identifies the cause of those disadvantages as being the belief that women and men have different natures, and she argues for the extension of political rights to women. Perhaps also she goes further, questioning the division between political and private on which political philosophy is premised, but even if she does not, her insistence on the common nature of men and women is still a central feature of feminist political philosophy and was taken up by other philosophers in the nineteenth century.

Chief among these were J.S. Mill and Harriet Taylor. Like Wollstonecraft, Mill presented an argument which was both critical and constructive. In *The Subjection of Women* (1869), he attacked those, like Rousseau, who argued that woman's political exclusion was justified by her different nature, insisting instead that women had been so repressed by social and political discrimination that it was impossible to tell what their true nature was. In so far as woman's nature appeared different from man's this was most likely to be a result of an education and upbringing which was distorting and repressive.

However, in his positive proposals, Mill appears less radical than either Mary Wollstonecraft or Harriet Taylor, who also wrote extensively on the subject of women's political position, most notably in her essay *The Enfranchisement of Women* (1851). For while Mill is eager to emphasize the fact that woman's nature has been deformed by her poor treatment, he also predicts that when women are granted equal legal and political rights with men, they will nevertheless not choose to exercise those rights in the public arena, but will prefer to remain within the home, pursuing the occupations traditionally reserved for women. And, contrary to his official position, he sometimes implies that this is because women are more emotional and intuitive than men and therefore better suited to domestic than to political occupations.

We have already seen that Mary Wollstonecraft favoured equal legal and political rights for women, yet also argued that men and women would express their citizenship differently. J.S. Mill also inveighs against women's inequality in legal and political contexts, but he sometimes implies that men and women are by nature the same, sometimes that they are different. His uncertainty about whether the differences between men and women are of political significance, and whether they spring from nature or from nurture, recur in the disputes which characterize much twentieth-century feminist political philosophy.

What is overwhelmingly clear, however, is that Wollstonecraft, Mill and Taylor set the agenda of modern feminist political philosophy. At a time when the rights of man were being declared in France and in the USA, Wollstonecraft raised the question of why those legal and political rights had not been extended to women. She argues in terms of justice and equality. Similarly, Mill and Taylor emphasize justice and equality in their discussion of women's subjection. In so doing, they provide the framework for modern liberal feminism, but it is arguable that they do more, and in some ways also anticipate the more radical forms of feminism which have become prominent in the late twentieth century.

3 Liberal feminism

The difficulties inherent in defining liberal feminism arise in part from the difficulties in defining liberalism itself, but in general we can identify the salient features of liberalism as being its individualism, its emphasis on justice and equality, and its understanding of the state ideally as a neutral arena within which people should be left to pursue their own interests in their own way (see Liberalism; Neutrality, political). For liberals, the most important function of the state is to protect individuals from interference by others. Liberalism takes as its premise the freedom and equality of all human beings and construes the state as justified in so far as it preserves that freedom and equality. Accepting these central premises, liberal feminism also emphasizes the rights of women as individuals; specifically, their right to equal treatment under the law and the responsibility of the state to ensure their freedom to develop as autonomous individuals.

Liberal feminism is primarily concerned with the legal and political rights of women and with the justice of political arrangements. Thus, members of the suffrage movement emphasized the injustice of denying the vote to women, and nineteenth-century campaigns for married women to be allowed to own property were also based on liberal premises: they appealed to the equality of men and women and to the consequent injustice of denying women the rights accorded to men.

More recently, liberal feminism has been concerned to argue for women's rights in the workplace: rights to equal pay and to equal opportunities more generally. These issues, however, raise doubts about the scope of liberalism in general, and liberal feminism in particular. The first doubt arises in connection with the individualism of liberal political philosophy: as we have seen, early liberal feminists argued for the essential sameness of men and women as rational beings. For liberal political philosophy, human beings are essentially rational creatures, and physical differences between people should not be used to justify differential treatment. Obviously, this argument was important in, for example, the Civil Rights movement in the USA. In similar vein, modern liberal feminists argue

that since women are, like men, essentially rational beings, they should be accorded all the legal and political rights of man. But in the case of women, it is far from clear that physical differences can be so easily disregarded, much less that it is just to disregard them. In this respect, it has been argued, liberal political philosophy is gendered, since it assumes the centrality of rationality and thereby ignores the important differences which spring from woman’s role as child-bearer and child-rearer. Connectedly, liberal feminism has been criticized for being altogether too willing to comply with the denial of difference and the characterization of human beings as essentially rational.

These reflections on woman’s different physical and biological nature can either inform or undermine liberal feminism and liberalism generally. Some have argued that differences can be incorporated within the liberal understanding of justice, while others have insisted that the differences cast doubt upon the very concepts of justice and equality themselves and their centrality to traditional political philosophy. Thus, Susan Okin (1990) argues that John Rawls’ theory of justice may be extended in such a way that the justice of a society is partly determined by the justice of domestic arrangements. Others have argued that the extension of justice to the domestic realm is inconsistent with central liberal principles, while yet others have cast doubt upon the importance which liberalism places on justice (see Justice §5).

4 Marxist and socialist feminism

Where liberal feminism emphasizes justice and equality and locates the source of women’s oppression in their unjustified exclusion from the political and legal rights accorded to men, Marxist and socialist feminism see women’s oppression as springing from their socio-economic condition, arguing that it will end only when social and economic circumstances change. The terms ‘Marxist feminism’ and ‘socialist feminism’ require elucidation because they are not used uniformly in the literature: sometimes they are used interchangeably, while at other times Marxist feminism is used to refer strictly to those who base their analysis on Marxist theory, and socialist feminism to refer to the more general belief that the liberation of women requires the transformation of the socio-economic system. Rather differently, socialist feminism is often used to refer to the synthesis of Marxist feminism and radical feminism. In general, however, we may take both Marxist and socialist feminism to be characterized by their belief that the causes of women’s oppression are rooted in the socio-economic system and by their proposals for remedies for that oppression - namely, the transformation of the system.

Orthodox Marxist feminism aims to extend the concepts and insights of Marx’s writings to the situation of women (see Marx, K.; Marxism, Western). In the past, this involved an understanding of women’s oppression as a special case of the quite general exploitation of workers within capitalist societies, and the remedy was to treat women’s work simply as one form of necessary labour under capitalism. This certainly seems to have been Engels’ position (see Engels). However, the analysis raises a number of problems: it is unclear whether housework should remain a ‘private’ matter, but shared equally between men and women or whether it should itself become a form of labour which attracts recompense. For many Marxist feminists, housework, or domestic labour, should not be seen as a personal service to an individual man, but as a more general presupposition of capitalism, one which is necessary if the workforce is to be reproduced. Additionally, and crucially, feminists have argued that housework is an area in which women are not merely exploited by men as capitalists, but by men as husbands. Put bluntly, Marxist feminism must account for the fact that it is not only capitalists who benefit from the oppression of women, but men generally. This argument undermines the ability of Marxist feminism to reduce all exploitation to class exploitation. For women, there is, or appears to be, a double exploitation - one rooted in class, the other rooted in gender. In consequence, there is reason to believe that women’s oppression will survive the demise of capitalism: there is no simple Marxist solution to the problem of patriarchy unless the Marxist understanding of production is extended to include reproduction.

These problems have generated a more general ‘socialist’ feminism which acknowledges the separate and distinct forms which oppression may take and which is reluctant to see all oppression as rooted in class. Here it is argued that women experience the dual oppressions of class and gender differently. Potential unities of class are disrupted by conflicts of gender (see Socialism §4).

But if this point holds, it signals not only the demise of Marxist feminism, narrowly construed, but also the possibility that different kinds of oppression will require different, and perhaps conflicting, remedies. This possibility, implicit in the denial of a single source of oppression, is problematic for feminist political philosophy

because it undermines the attempt to identify a single source of women's oppression, suggesting instead that for different women there will be multiple and conflicting sources, not easily reconcilable one with another.

5 Radical feminism

Where liberal feminism traces the oppression of women to unjust laws, and Marxist and socialist feminism traces it to socio-economic structures, radical feminism identifies the cause of women's oppression as men. It is to radical feminism that we owe the central feminist term, 'patriarchy', and the feminist slogan, 'the personal is political'. As the name implies, it is also in radical feminism that we find the most vigorous attacks on the aims and preoccupations of traditional political philosophy. For whereas liberal feminism, Marxist feminism and socialist feminism all subscribe to political philosophies which they believe can be modified to include women's interests, radical feminism is inherently critical of political philosophy itself, seeing it as one of the many ways in which male power is legitimized. On this account, socialism, Marxism and liberalism all misidentify the cause of women's oppression by interpreting it as a specific case of a more general problem. But for radical feminists, the oppression of women is a distinctive form of oppression in itself: it transcends other forms of oppression and in order to combat it women must unite as women against men as the oppressors.

From the perspective of political philosophy, the importance of radical feminism lies in its reconstruction of the concept of the political. Where liberal feminism emphasizes the injustice of existing laws, and urges the extension of the rights of man to woman, radical feminists construe the injustice of existing laws as part of an all-pervasive structure of male domination which begins in the family and spreads outward to political institutions. Hence the claim that 'the personal is political'. On this account, changes to existing laws governing divorce, property ownership and suffrage are merely cosmetic devices which conceal, and sometimes even perpetuate the injustice which exists in the structure of the family itself, and about which liberalism and liberal feminism, are largely silent. Similarly, the socialist and Marxist emphasis on the economic basis of women's oppression overlooks the important respects in which oppression may take non-economic, specifically sexual, forms. Thus, the ways in which women are disadvantaged in the workplace, and the problems associated with the status of domestic work, are but symptoms of a wider problem - the problem of women's sexual subordination to men.

But if the causes of women's oppression are as identified by radical feminists, what are the remedies? Here, radical feminists disagree profoundly: some claim that the demise of the family itself, and of women's role as child-bearers is what is required; others argue that what is needed is the re-evaluation of those roles and institutions. Expressed more generally, this is a disagreement about whether woman's biological nature is an inconvenient encumbrance obstructing her liberation, or whether it is the source of values which are currently unrecognized in politics but which, if cultivated, could provide a richer and more humane understanding of the political.

An example of the former strategy is to be found in the work of Shulamith Firestone, who argues in *The Dialectic of Sex* (1970) that the source of woman's oppression lies in her child-bearing capacity, in her 'fundamentally oppressive biological condition'. However, reproductive technology may enable woman to break free of the burden of her biology. If, therefore, women can control reproductive technology, they will, for the first time, be able to overcome oppression. Adapting Marx's arguments, Firestone argues that, like workers under capitalism, women must seize the means of (re)production with the ultimate aim of eliminating sex power and the sex distinction itself. By contrast, other radical feminists have asserted the value of reproduction and child-rearing, and have argued that women should not aspire to free themselves of those roles, but should rather emphasize the importance of the qualities which are developed through child-bearing and rearing. Thus, Sara Ruddick (1990) argues that distinctive ways of thinking may grow out of the work that mothers do. These ways of thinking are not, however, the unique preserve of women: they are developed via 'woman's work', but they are not rooted in woman's nature. They are, moreover, qualities which could have a political use in, for example, our dealings with conflict, but political philosophy's emphasis on justice and equality tends to disguise their significance.

In radical feminism, therefore, we find not so much a political philosophy as a critique of political philosophy. For this reason, radical feminism's lack of a theory of the state is not an indication that it is apolitical. On the contrary, its emphases on sexuality and child-bearing are themselves a political statement. For radical feminists, the identification of a separate, political realm is a mechanism for disguising the vast operations of power within the private world of the family. Until these latter are explained and removed, attempts to secure the liberation of

women via the 'political' methods advocated by liberal feminists, or the social and economic methods advocated by socialist and Marxist feminists, will be doomed to failure.

6 Postmodernism and feminism

In recent years, feminist political philosophy has been much influenced by the postmodernists. The term 'postmodern' covers a very wide range of thinkers, including the poststructuralists, deconstructionists, French theorists and psychoanalytic thinkers (see Postmodernism and political philosophy). However, from a feminist perspective, the important feature of all these thinkers is, as the name implies, their rejection of the modernism of Enlightenment philosophy. In the Enlightenment tradition, philosophy aims at the rational pursuit of truth, certainty and objectivity. We have already seen how some feminist philosophers reject the Enlightenment emphasis on reason and subvert traditional political philosophy by drawing attention to the 'maleness' of its central concepts. Thus, in moral and political philosophy, some radical feminists have argued that care and compassion should supplement, or even replace, justice and equality as core terms (see Feminist ethics). However, postmodernism goes further: its rejection of the Enlightenment search for truth is part of a general denial that truth or objectivity can be attained. Postmodernism is therefore at odds with both liberal feminism and Marxist feminism, since both, in their different ways, are committed to the possibility of certainty and objectivity. For postmodernists, all such commitment is itself a form of domination and therefore cannot form any part of a feminist project.

In addition, for postmodernists, the denial that there can be objective truth or certainty is often linked with an emphasis on the importance of language as a mediator between individuals and the world. Here, feminists sympathetic to postmodernism have emphasized the ways in which language itself is 'man-made' and thus constitutes a form of oppression for women. For many 'French feminists' such as Cixous and Irigary, the removal of patriarchy requires that we think as women, but thinking as a woman requires a recognition of the extent to which the language in which we think is itself man-made. The question which now arises is whether there can be such a thing as 'thinking as a woman' or whether there are simply different subjectivities and no unified category of woman.

Although postmodernism has found enthusiastic support among many feminists, it is not without difficulties. For one thing, there is an air of paradox about its confident assertion that there can be no such thing as certainty. But even ignoring that problem, postmodernism threatens to leave feminism as a rebel without a cause, for if there is no certainty or objectivity, what is the status of women's oppression? Is it also simply a feature of different subjectivities and not an objective fact? By denying the possibility of objectivity, postmodernist feminists appear to be depriving feminism of its central tenet - its claim that women as a group suffer systematic political disadvantage - and reducing it to little more than an analysis of different subjectivities. And this threatens to undermine both the possibility of, and the motivation for, feminist political action.

7 Recent feminist political philosophy

The categories of feminist political philosophy are, in part, artificial. The distinctions between socialist, Marxist and radical feminism are often blurred and controversial. Moreover, many writers resist categorization in this way - unsurprisingly, since one of the main aims of much feminist political philosophy is to indicate the ways in which the traditional branches of political philosophy themselves contribute to the oppression of women. In recent years, therefore, attention has shifted away from the distinct categories of feminist political philosophy towards the ways in which the insights of feminism, however categorized, may inform our understanding of political arrangements.

One important area here is feminist discussion of democracy and of the way in which democracy may involve a recognition of differences between people (see Democracy). In the history of political philosophy, and in the history of feminist political philosophy, individual rights have usually been asserted on the basis of the similarities between men and women. Mary Wollstonecraft, J.S. Mill and Harriet Taylor all emphasized the essential sameness of men and women as rational beings, and concluded that since women were the same as men in this respect, they should be accorded equal rights. However, the argument from sameness has its limitations. There are many important respects in which women are not the same as men, and these differences affect women's ability to make full use of their political rights, even when such rights are granted. For this reason, feminist political philosophers have turned their attention to the ways in which equality and justice between men and women may be guaranteed

through a recognition of difference, rather than through its denial or removal. One of the main proponents of this argument is the US writer, Iris Marion Young, who argues for the recognition of differences between different groups of people and for the differential treatment of those groups, in the name of fairness. Justice, she maintains, must not be interpreted as requiring abstraction from the particular circumstances of one's life (Young 1990). On the contrary, it requires a recognition and accommodation of those circumstances.

Young's argument against abstraction and universalism can, however, appear to license a special pleading at odds with an understanding of citizenship as essentially a relationship between equals. Thus, Mary Dietz argues that citizenship is about getting beyond one's immediate sphere, and dealing as a citizen with other citizens (Dietz 1985). In this respect she appears to insist upon precisely the abstraction and universalism of which Young is deeply suspicious. But the differences between them are less important than the similarities, since both insist on the importance of difference and on the need to find ways of acknowledging difference in political contexts (see Citizenship §3).

From Wollstonecraft to Dietz, feminist political philosophy has been concerned with the political significance of differences between men and women, and the ways in which those differences may be rendered compatible with political justice and equality. Often, that project has been pursued through a critique of the central concepts of political philosophy, and sometimes through a critique of political philosophy itself. In these respects, feminist political philosophy serves not only as a mechanism for drawing attention to the injustices of political arrangements as they affect women, but also as a means of revising our understanding of what political philosophy is.

See also: Feminism; Feminist Ethics; Feminist jurisprudence

SUSAN MENDUS

References and further reading

Bryson, V. (1992) *Feminist Political Theory*, London: Macmillan.(An excellent introductory text, clear and comprehensive, containing an extensive bibliography.)

Dietz, M. (1985) 'Citizenship with a Feminist Face. The Problem with Maternal Thinking', *Political Theory* 13 (1): 19-37.(A critique of the political importance of maternal thinking; referred to in §7 of this entry.)

Firestone, S. (1970) *The Dialectic of Sex*, New York: William Morrow.(A highly controversial and provocative book; discussed in §5 of this entry.)

Mill, J.S. (1869) *The Subjection of Women*, P, 1983.(The most important nineteenth-century argument for the enfranchisement of women; discussed in §2 of this entry.)

Okin, S. (1990) *Justice, Gender and the Family*, New York: Basic Books.(An incisive discussion of the gendered nature of modern political philosophy; referred to in §3 of this entry.)

Pateman, C. (1988) *The Sexual Contract*, Oxford and Cambridge: Polity Press.(A highly influential critique of contract theory, covering much of the history of political philosophy. A difficult but rewarding book.)

Phillips, A. (1991) *Engendering Democracy*, Oxford and Cambridge: Polity Press.(A clearly written and accessible account of the ways in which democratic theory is gendered.)

Ruddick, S. (1990) *Maternal Thinking. Towards a Politics of Peace*, London: Women's Press.(A clearly written and influential book emphasizing the importance of ways of thinking developed through 'women's work'; discussed in §5 of this entry.)

Taylor, H. (1851) *The Enfranchisement of Women*, London: Virago, 1983.(Contemporary with J.S. Mill's *Subjection of Women*, the text arguably more radical but less influential than Mill's work; discussed in §2 of this entry.)

Wollstonecraft, M. (1792) *A Vindication of the Rights of Woman*, Harmondsworth: Penguin, 1978.(The first significant philosophical argument for the rights of women; discussed in §2 of this entry.)

Young, I.M. (1990) *Justice and the Politics of Difference*, Oxford: Princeton University Press.(A very influential discussion of the shortcomings of traditional democratic theory. Young argues for the affirmation rather than the suppression of social group difference. Discussed in §7 of this entry.)

Feminist theology

Feminist theology began as a reaction to the exclusion of women and women's concerns from traditional Christian theology, but it soon incorporated constructive as well as critical elements. Originating 'from the margins' of women's exclusion, it now is a major force within Christian theological thought. The issue it raised initially was the cultural and social suppositions that inform all theological thinking and that enter into theologies as 'universals'. In response to such universals, a feminist 'hermeneutic of suspicion' seeks out the hidden norms and biases within religious texts.

Feminist theology is diverse, but it is characterized by pervasive themes. Immanence is valued over transcendence, relation over substance, change over immutability, liberation over salvation, and ecological concerns over traditional Christian eschatological concerns. As could be expected from its hermeneutic of suspicion, feminist theology is also characterized by its insistence on the social location of all thinking. Feminist theologians uniformly use gender-inclusive language not only in reference to humanity, but also in reference to God. Finally, all feminist theologians manifest a concern for liberation from every type of oppression, environmental as well as social, and not just liberation from the oppression women have experienced.

1 A brief history of contemporary feminist theology

Feminist theology began modestly in April 1960 with the appearance of Valerie Saiving's article 'The Human Situation: A Feminine View'. Saiving raised a critical question concerning the work of the prominent theologian Reinhold Niebuhr. In Niebuhr's monumental *The Nature and Destiny of Man* (1941-3), he analysed sin as pride and sensuousness, with pride clearly being the greater sin. Saiving suggested that perhaps pride was the distinctive sin of men; perhaps women's experience suggested a different understanding of sin as triviality, shallowness, and a deficiency of pride, which hindered women's self-development. Although it was eight years before the first major feminist work appeared (Mary Daly's *The Church and the Second Sex* (1968)), distinctive themes of feminist theology are already apparent in Saiving's article. These are: the particularity and social location of all theology; the marginality of women within male theology; and the social framework of patriarchy as the mediator of values, and the negative effect of these values upon women. These three themes formed the matrix of feminist theology in its early years. Feminist theology was a largely critical theology, aimed at exposing the pretensions to universality and the inherent sexism of all Western Christian thought.

Feminist theology moved towards constructive thought with Mary Daly's second book, *Beyond God the Father* (1973). Daly redefined the Christian doctrines of God, humanity, sin, salvation, Christology and ecclesiology by casting them within the framework of feminist values. She rejected the traditional notion of God as Being, and developed 'God the Verb' (see God, concepts of §§1-3). Humanity was depicted as distorted by patriarchy, so that any ideal of humanity could only be developed in separation from patriarchy. Women unfolding their potential in response to the call of God the Verb were presented as the model for anthropology. Daly identified the primal sin as 'stolen energy', or men's systemic victimization of women together with the attribution of blame for this victimization to women themselves. Salvation is not through a Christ figure (for no male can save women), but through women reclaiming their energy and answering death-oriented male values with life-oriented female values. With regard to ecclesiology, Daly called for a new sisterhood of humanity governed by the vision of a community of women living in harmony with the world of animals and nature, beyond the reaches of the evil patriarchy. The work had a catalytic effect within the feminist religious community.

Two other major feminist theologians in the 1970s were Rosemary Radford Ruether and Letty Russell. Ruether's first major feminist work was *Liberation Theology* (1973), which was quickly followed by *New Woman, New Earth* (1975). In the first text, Ruether related feminist theology to the emergent liberation theologies of African-Americans and South Americans (see Liberation theology); in the second work, she developed what was to become a dominant feminist theme concerning the connections between women and the earth in patriarchal theology. The earth, like woman, becomes object to man's subject; the earth, like woman, is exploitable for man's purposes. Thus the revaluation of woman entails a revaluation of earth as well. The redemption of women and men from patriarchy will result not only in new possibilities for women, but new possibilities for the earth.

Letty Russell's work focuses upon the implications of feminist theology for the redevelopment of the Church.

Unlike Daly (and like Ruether), Russell argues that Christianity contains the seeds of its own redemption. The prophetic strain of Christianity always offers a critique of every present state of affairs, and the role of women in the present is to call Christianity to a more faithful expression of itself in communities of justice. A dominant motif for Russell is 'partnership', which signifies women and men together transforming the patriarchal Church into a newly redemptive community of wellbeing.

Daly, Ruether and Russell continued to have a major influence on feminist theology through the 1980s and into the 1990s. Following *Beyond God the Father*, however, Daly abandoned Christianity entirely and devoted her subsequent works to feminist philosophy rather than theology. Ruether's major subsequent works are *Sexism and God-talk* (1983) and *Gaia and God* (1992). In both works Ruether continues to develop her earlier themes, exploring reconceptualization of Christian doctrines, further analysing the intertwined history of sexism and the disregard of nature, and proposing a transforming synthesis of Christian notions of God with ecological notions of the earth as Gaia. Ruether maintains that only through such a reconceptualization can we address the almost irreversible destruction of the earth's biosphere that has been brought about by patriarchy.

Other major feminist voices are Rebecca Chopp, author of *The Power to Speak* (1989), and Elizabeth Johnson, author of *She Who Is* (1992). Both focus on the doctrine of God in relation to communities of justice and wellbeing, with Chopp developing the traditional language of God as 'Word' into the notion of a 'perfectly open sign' that cannot be restricted within one image. She contrasts 'Word' as perfectly open sign with the restrictive metaphor of 'father', arguing that because of the latter's patriarchal limitations, it is no longer an adequate portrayal of the God whose very being is directed towards emancipation for the oppressed.

Johnson develops a new understanding of the Trinitarian God, both drawing from the tradition and transforming it through explicitly putting feminist theology and classical theology in dialogue with one another. In the process, she shows that feminist sensitivities can illuminate much in the ancient concept of God - particularly through revision of concepts such as 'omnipotence' and 'impassivity'. She also argues that feminist theology is given increased depth through incorporating aspects of the tradition.

In the decades since Saiving's initial article, it has become apparent that feminist theology can no more be reduced to a single voice than traditional theology. Not only is feminist theology itself diverse, but it has led to such alternative modes of theology as Womanist Theology, Mujerista Theology, Asian Women's Theology and Goddess Thealogy. The first three criticize the implicit and sometimes explicit racism of feminist theologians, who are seen to speak only for white middle- or upper-class women. These modes of theology go beyond critique, however, and create constructive theology by building upon insights developed through the distinctive ethnicities of (respectively) African-American women, Hispanic/Latina women and women from the various countries in Asia. Goddess Thealogy criticizes the orientation towards God and/or Christianity of feminist theologians, claiming that Christianity, in large part by virtue of its notion of God, is the primary perpetrator of sexism and should therefore be abandoned. Despite its diversity, however, feminist theology tends to be characterized by the themes described below.

2 Feminist themes

Immanence over transcendence. Traditional theologies claim both of these categories, but the emphasis is clearly on transcendence. Feminists either retain both, but reverse the valuation, or eliminate or ignore transcendence in favour of immanence. Carter Heyward, in *The Redemption of God* (1982), falls into the latter category. For Heyward, God is wholly immanent, and is defined as the power of relationships between persons. Whether or not there is a transcendent locus to this power is irrelevant; the issue is our ability to relate to one another in constructive ways that value not only the individual but also the community. The power that generates this form of relation is also generated by this form of relation, so that, in a sense, the God who empowers relation is also strengthened by the dynamism of the resultant relations.

While Heyward comes close to collapsing the notion of God into the notion of community, other feminists, such as Mary Daly, retain the sense of transcendence in balance with immanence. Daly's naming of God moves successively from 'God the Verb' in *Beyond God the Father* to 'Goddess the Verb', the 'Unfolding Verb', 'Be-ing' and 'Meta-being' in her later works. The transcendent aspect of this 'Be-ing' functions as a power of the future. Patriarchal theologians such as Wolfhart Pannenberg also named God the power of the future, but in Daly's

treatment the deity's transcendence acts as a lure to women who are freed from, or seek to be freed from, the powers of patriarchy. This dynamic 'Be-ing' calls women into a new future; hence this divine reality transcends the present moment. As a new future, this transcendent reality is also a transformative reality, calling women into a mode of being that has never before been achieved. Thus, in its transformative mode, 'Be-ing' is immanent as well as transcendent. 'Be-ing' becomes the other side of woman's self as a kind of meta-self, or, in Daly's capitalized version, the Self. It is the transcendent 'Be-ing' of the future become immanent, empowering woman's transformation. Thus in feminist theology, transcendence becomes a temporal rather than a spatial term, and immanence becomes an inward 'otherness' that is more than the self, yet intimately connected to it.

Relation over substance. The major category of Western theology is arguably 'substance' (see Substance). Substance is that which requires nothing other than itself in order to exist, with the understanding, of course, that only God conforms perfectly to this ideal. Feminists roundly repudiate the notion of an independent substance, capable of existing apart from any relation, and argue instead that existence is itself profoundly relational. To exist is to be in relation; thus substance as that which requires nothing other than itself in order to exist is an absurdity.

The feminist value of relation over substance is not unique in late-twentieth-century theology; process theologians, who follow Alfred North Whitehead (§4), have long argued that process is reality, and that process is a dynamic movement in which every existent reality comes into being through integrating the influence of its predecessors into its own becoming (see Process theism). Feminist theologians have generally not developed the metaphysical basis of relationships; rather, they tend to assume the primacy of relationships. In many respects, the appeal is simply to women's experience. Who one is develops in and through the relationships in which one participates. Relationships can be negative when one is subsumed into subservience to another, as in the sexist attempt to 'own' another person, or relationships can be positive, as in the richness of mutual relation. But that existence is formed in and through relationships tends to function as a given in feminist theology.

Change over immutability. Traditional theology tended to attribute immutability to God and mutability to the world, with mutability being the clearly subordinate and dependent state (see Immutability). Given the feminist assertion of the primacy of relation over substance, feminists likewise value mutability over immutability. The latter is regarded as an abstraction, merely posited as the antithesis of what is considered to be universal movement, process or change.

A major implication of the feminist rejection of immutability is the redefinition of perfection. Classical theology related perfection to immutability, holding that if a subject could change, it was either better or worse than the state from which it changed; hence perfection required immutability. The feminist rejection of immutability entails a change in the understanding of perfection. In feminist hands, perfection is a dynamic condition that requires change in order continuously to manifest itself. Far from being a subordinate category, change is essential to existence.

Liberation over salvation. Traditional theology maintained that there were both historical and everlasting aspects to salvation, but feminist theologians focus almost entirely on the historical. For them, salvation is not a change of state before God, an imputed righteousness covering one's sins, or an individual affair. But since 'salvation' connotes all these things, feminists eschew the word and speak instead of 'liberation'. Far from being individualistically oriented, liberation bespeaks a transformation of the social/symbolic system of one's culture from patriarchal exclusiveness to communities of inclusive wellbeing. In *The Power to Speak*, Rebecca Chopp describes such communities as existing through 'emancipatory transformation'.

Many feminist theologians envisage such communities of liberation as already existing through para-Church organizations such as 'women-church' (Ruether 1985). These small group communities are both inside and outside the Church. Inside, they function to bring about change within congregations; outside, they function as small worship and support groups for like-minded feminists. Other feminists, such as Russell (1993), focus more clearly on the existing Church as a community of liberation, finding congregations within the Church that are willing to experiment with new modes of inclusive wellbeing. But all feminists call for transformation not only within the Church, but within society, with the aim of achieving more just and sustainable conditions of life. This is liberation, and like the symbol of 'salvation' before it, it is both the goal and the process of Christian life.

Ecology over eschatology. As implied in the preceding section, feminist concerns focus on the wellbeing of this world, both the many-cultured world of peoples and the many-splendoured world of nature. As early as the 1970s, feminists pointed out the symbolic connections between women and nature, and argued that the liberation of one entails the liberation of the other. For example, Ruether in *Gaia and God* contrasts the feminist ecological concern with Christian eschatology (see Eschatology §1). While the latter dealt in temporal as well as eternal salvation, it tended to include an end of the world, along with other-worldly redemption and spiritualized bodies. Feminist adaptation of this eschatology also speaks about an end of the world, but through human wastefulness rather than through divine judgment. The apocalypse often associated with eschatology becomes the natural disasters resulting from deforestation, ozone depletion, air, ground and water pollution, and the rapid extinction of species. This feminist adaptation of eschatology into ecology decries eschatology's depiction of the loss of this world in favour of another, calling instead for social and individual conversion in order to realize responsible, just and sustainable communities of wellbeing.

The above five themes characterize feminist theology. Emphasis and development vary, depending upon the perspective of the particular theologian. But the impact of feminist theology as a new genre within theological discourse is raising awareness of these issues within all modes of contemporary theology.

See also: Feminism; Feminist ethics

MARJORIE SUCHOCKI

References and further reading

Chopp, R. (1989) *The Power to Speak: Feminism, Language, God*, New York: Crossroad Publishing.(Replaces the notion of 'God' with 'Word' as the 'perfectly open sign', open for its own and for society's transformation.)

Chopp, R. and Taylor, M.L. (eds) (1994) *Reconstructing Christian Theology*, Minneapolis, MN: Fortress Press. (Collection of essays juxtaposing traditional Christian doctrines with contemporary challenges and reconstructing the doctrines accordingly.)

Daly, M. (1968) *The Church and the Second Sex*, Boston, MA: Beacon Press. (Uncovers the history of misogyny in the Church.)

Daly, M. (1973) *Beyond God the Father*, Boston, MA: Beacon Press.(Redevelops the major doctrines of Christianity according to feminist criteria.)

Heyward, C. (1982) *The Redemption of God: A Theology of Mutual Relation*, Lanham, MD: University Press of America.(A theology of immanence, stressing the power of God within and for human relationships.)

Johnson, E. (1992) *She Who Is: The Mystery of God in Feminist Theological Discourse*, New York: Crossroad Publishing.(Redevelops the notion of the trinity through the centrality of Sophia-Spirit, with implications developed for human communities involving just and liberating relationships.)

McFague, S. (1993) *The Body of God: An Ecological Theology*, Minneapolis, MD: Fortress Press.(A revision of theology developed around the central metaphor of the world as God's body.)

Niebuhr, R. (1941) *The Nature and Destiny of Man*, vol. 1, *Human Nature*, New York: Charles Scribner's Sons. (An investigation of what it is to be human, including consideration of finitude, freedom and sin.)

Ruether, R.R. (1973) *Liberation Theology*, New York: Paulist Press.(An exploration of the issues raised in Black theology and in the emerging feminist theology of the 1960s and early 1970s in America.)

Ruether, R.R. (1975) *New Woman, New Earth*, New York: The Seabury Press.(An analysis of the formation of images of women in religion through study of the ideologies and structures of sexism, racism and capitalism.)

Ruether, R.R. (1983) *Sexism and God-Talk: Toward a Feminist Theology*, Boston, MA: Beacon Press.(Argues that Christianity has a prophetic stream within it that works towards its own transformation, and redevelops major Christian doctrines accordingly.)

Ruether, R.R. (1985) *Women-Church: Theology and Practice of Feminist Liturgical Communities*, New York: Harper & Row.(Practical analysis of the historical and theological understanding of the Church as a community of liberation. Offers guidelines for developing a feminist renewal movement within the Church.)

Ruether, R.R. (1992) *Gaia and God: An Ecofeminist Theology of Earth Healing*, San Francisco, CA: Harper. (Outlines the history of human destruction of the earth, relating this destruction to social symbol systems. Proposes a new symbolism to catalyse healing of the earth and its communities.)

Russell, L. (1993) *The Church in the Round: Feminist Interpretation of the Church*, Louisville, KY: Westminster/John Knox Press.(Uses a paradigm of partnership to develop an ecclesiology that lessens distinctions between clergy and laity, and that interprets the Church in terms of its mission of liberation both

within the Church and in society.)

Saiving, V. (1960) ‘The Human Situation: A Feminine View’, *Journal of Religion* 40: 100-12.(A consideration from the point of view of women of Niebuhr’s analysis of sin as pride and sensuousness. Argues that Niebuhr concentrates overly on the sin of pride, which peculiarly defines male experience, and neglects the issues problematic to women, such as triviality and a failure to have enough pride in oneself and one’s work.)

Thistlethwaite, S.B. and Potter, M.E. (eds) (1990) *Lift Every Voice: Constructing Christian Theologies from the Underside*, New York: Crossroad Publishing.(A collection of essays treating major Christian doctrines from the point of view of oppression and liberation.)

Feminist thought in Latin America

Any analysis of feminist thought in Latin America is burdened by the task of combatting the frequent assumption that feminism is an ideology imported from the USA or Europe. One could begin by arguing that in certain senses autocthonous feminist thought has existed in Latin America for centuries. The thought of Sor Juana Inés de la Cruz, a seventeenth-century Mexican writer and nun, had certain qualities, themes and perspectives that can be called feminist. Her autobiographical essay, 'Reply to Sor Filotea de la Cruz', is a brilliant defence of a woman's right to engage in intellectual pursuits and includes many feminist strategies and dimensions.

Among the Latin American feminists through the centuries, exemplary passages are easily found, such as the following from the Peruvian Flora Tristán, who said 'Without liberation of woman, there will be no liberation of man' (1843). Revisionist reappraisals of underappreciated women thinkers have grown more common. Venezuelan Teresa de la Parra's important writings which have been re-examined as texts by women continue to be rescued from relative obscurity. Rosario Castellanos, Rigoberta Menchú and Domitila Barrios de Chungara count among those whose work has enjoyed increased critical attention. Concurrently, the voices of traditionally voiceless women are being heard through expanding oral history projects, such as those of domestic workers in Bogotá: 'It is not enough to have rights. We must raise consciousness and organize ourselves to defend those rights'.

In recent years, particularly after the international year of the woman in 1975 and the decade of the woman sponsored by the United Nations from 1976 to 1985, women and men have continued to develop feminist philosophies appropriate for Latin American contexts.

Professional feminist philosophy has been practised in Latin America since the early part of the twentieth century. Perhaps surprisingly, a theoretically sophisticated feminist philosophy was practised in Uruguay at this time by male social philosopher, Carlos Vaz Ferreira (1871-1958). His work had significant impact on women's rights in Latin America. Vaz Ferreira was a pioneer in feminist theory.

1 The early twentieth century: Carlos Vaz Ferreira

The ideas of this seminal Latin American social thinker and his thought-provoking study of gender and family have stood the test of time since they were first delivered in Uruguay in 1914. *Sobre feminismo (On Feminism)* (1933) was written between 1914 and 1922 in the form of lectures which were delivered at the University of Montevideo where Vaz Ferreira was an academic. The book is concerned primarily with examining 'factual' differences between the sexes and 'normative' issues, such as the political, civil rights, their social life and the organization of the family within society. Vaz Ferreira analyses the disproportion between the ideas and faculties of women and the scope which society allows to their activity. He advocates the right of women to participate in all that makes life valuable to the individual human being. The ideas he expresses in *Sobre feminismo (On Feminism)* are poignant, relevant and innovative in the light of contemporary social debates throughout the Americas.

Sobre feminismo (On Feminism) is the Latin American counterpart of John Stuart Mill's *The Subjection of Women* (1869) in terms of content and tone, and more importantly, its impact on elite thinking. *Sobre feminismo (On Feminism)*, however, reflects the social changes that could be expected more than fifty years later. Vaz Ferreira's early influence on the suffrage movement was significant and his contributions are particularly noteworthy given that in general Hispanic men of his culture and age were not renowned for their progressive attitudes towards women. In Latin American intellectual circles in the early part of the twentieth century, an effect of the pervasive *machismo* and its complementary femininity was to marginalize women so thoroughly that thought about gender and family roles only had widespread impact when expressed by powerful men.

As one of Latin America's most influential social philosophers in the early twentieth century, Vaz Ferreira's complete works were collected in twenty-five volumes (1957-63). *Sobre feminismo (On Feminism)*, set within the Latin American experience, presents cogent arguments against the marginalization of women, the infringement of their political rights and their second-class status in marriage. Vaz Ferreira outlined a theory of cooperation between men and women which privileged monogamy, the family and the equitable division of household tasks. He studied the ways in which pregnancy can be a disadvantage for women and suggested remedies to compensate

for what he viewed as biological inequity. Ahead of his time, Vaz Ferreira reflected on divorce, artificial insemination and abortion. He was a painstaking, self-consciously philosophical craftsman who clearly grappled with what evidence he could muster to support what was essentially cultural and social criticism of the intimate dealings of men and women. He was avid in avoiding contradiction so as to insist on a certain philosophical probity which was not expected in Latin American discourses on women's roles at that time. Because of this insistence, he was considered a model of anthropologically-sensitive social philosophy.

The first print-run of *Sobre feminismo (On Feminism)* (1933) coincided with the year in which Uruguayan women's suffrage was enacted. Since Uruguay was one of the first Latin American countries in which women's suffrage was achieved, the lengthy gestation period and public lectures which led to its publication reveal both its timeliness within Uruguayan society and Vaz Ferreira's role as an influential public thinker.

These dates are historically significant because they show that serious feminist thought has been going on in Latin America since the early part of the twentieth century. However, the theories of Vaz Ferreira have not received the critical attention they deserve. *Sobre feminismo (On Feminism)* is still a relatively unknown work, while European and North American analyses of feminism remain privileged and widely disseminated. One distinguishing feature of Vaz Ferreira's work is that it includes analysis of justice for women within the context of the family while many contemporary theories of justice omit consideration of women in families (see Family, ethics and the; Justice).

Sobre feminismo (On Feminism) is an analysis of the social situation of the woman 'of flesh and bone' in the context of 'feminism' and 'antifeminism'. In Vaz Ferreira's words, 'those terms "feminism" and "antifeminism", "feminist" and ''antifeminist", in reality, do more harm than good, and they complicate the many and sometimes enormous difficulties of the problems: they complicate them further with questions of words and with confusions derived from the words' (1933: 14). He believed that a false polarization is produced by the terms 'feminism' and 'antifeminism' because there are those who believe 'we are the true feminists because we want to preserve the distinctive traits of the female sex. You want to make men of women; your true name ought to be ''hominists" and not "feminists"' (1933: 17). In Vaz Ferreira's view, the issues are neither so polarized nor so simple and this kind of resentment impedes serious analysis of the situation in which modern woman finds herself. Nevertheless, considering the various connotations implicit in the term 'feminism', contemporaries run the risk of being misunderstood when calling themselves 'feminist'. Vaz Ferreira recognized the importance of clarifying language usage and tried to extract concrete meanings when discussing feminism.

He felt the best strategy for confronting the problem of the social situation of women is to examine, on the one hand, the possible questions relating to the similarities and differences between the two sexes and, on the other, examine normative problems. Vaz Ferreira distinguished factual questions from normative ones in *Lógica viva (Living Logic)* (1910). Such factual questions relate to knowledge and verification and the normative questions relate to action, preference and choice. The latter are most relevant in terms of the female condition.

Among the questions of similarities and differences between the sexes, Vaz Ferreira maintains that there are debatable and undebatable data. The important undebatable fact, radical for his time is: 'From the union between a man and a woman, the woman can become pregnant; nothing happens to the man'. He argued that, 'finding this fact to be satisfactory is to be ''antifeminist"' (1933: 25). The factual data are of three types: biological, physiological and psychological. A recurring debate is that of 'the comparative intelligence of the two sexes, a special case in the category of comparative psychology' (1933: 12). However, in his treatment of the intelligence and mental aptitudes of women, Vaz Ferreira takes seriously the possibility that it could be verifiable that women might be less intelligent than men as he ponders why there have been no female Beethovens or Darwins, for example. This is the weakest point in his argument and one of the rare occasions when he fails to take socialization into account when trying to explain differences between social groups and their roles.

More convincing is his treatment of normative problems in *Sobre feminismo (On Feminism)* (1933). The normative problems for Vaz Ferreira are women's political rights, their activity in society, their access to public office, careers, professions and education; civil rights, the relations between the sexes and the organization of the family.

The basic idea in his analysis of these problems is to maintain the difference between 'feminism of equality' and 'feminism of compensation'. The former, according to Vaz Ferreira, is based on the idea that 'jobs and careers should be open to women as they are to men; that women should have the same civil capacity as men, the same

level of education; that, in general, the sexes should be equalized by diminishing the difference between them and by placing women in the same situation as men, making them more like men' (1933: 16). For him, 'feminism of equality' does not merit much attention because of the mere fact that women are biologically mistreated by the likelihood of pregnancy in their unions with men and therefore it is not possible to speak of 'equalization'. The only acceptable feminism to Vaz Ferreira is that of 'compensation', based on the idea that it is necessary to compensate physiological injustice given that it will never be possible to equalize it and that it would be counterproductive to attempt to do so. Thus, 'Antifeminism takes as its guide that fact (women's biological disadvantage). Bad feminism does not even take it into account. Good feminism strives to correct it and compensate for it' (1933: 38).

With respect to the civil and political rights and social participation of women, Vaz Ferreira had a decisive impact in favour of women in the Uruguayan legislature. Suffrage in Uruguay was enacted in 1933, following the USA (1920) and Ecuador (1929). Vaz Ferreira also proposed a bill, the law of unilateral divorce, which passed into law in the form he had conceived it. The law gave women the power to obtain a divorce at will 'without giving cause, while men have to show just cause' (1933: 83). This law is consistent with his theory that the situations of men and women are fundamentally different. When the law was passed, 'opponents of divorce did not like it because of their need to preserve the family as the basis of society. Proponents of the right to a divorce did not like it either because they framed the question as one of "equality"' (1933: 83). Vaz Ferreira's position can be criticized as a case of reverse sexism in which men do not have the same rights as women. It can also be placed in the context of his theory of 'feminism of compensation' in the way he seems to propose replacing patriarchy with matriarchy to correct historical inequities. To some extent Vaz Ferreira also believed that matriarchy deserved a turn in beginning the long process of compensation.

The normative problem which most concerned Vaz Ferreira was that of relations between the sexes and the organization of the family. His analysis of marriage and divorce was a curious mix of obsolete and progressive ideas. He asserted that roles outside the home are for men and those inside the home are for women; that it is possible that women are less intelligent than men because most great cultural figures are men and that 'free love' is a destructive social force.

At the same time, Vaz Ferreira was a pioneer of feminist ideas which became widespread much later. For example, although he believed that relationships are ideally constituted as monogamous marriages, he also indentified marriage as an institution which unfairly regulates and limits the role of women both in professions and as part of a workforce, and therefore needed modification. Vaz Ferreira wrote, 'A woman's capacity to live for herself, which has to do with power, ability, and opportunity, should not depend wholly on marriage, as it seems to in mainstream society, which is one of the saddest and most unpleasant aspects of traditional society' (1933). He also criticized the arguments of opponents of divorce who 'reason as if those who support the right to a divorce maintained that divorce is a good' (1933: 81). He believed that it was unfair that women have been expected to change their names when they marry while men do not modify theirs. He asked: 'Isn't this a relic of antiquated social structures in which the man owned the woman, or she was subordinate to him?' (1933: 141).

2 Recent trends

Women's suffrage did not immediately translate into social change. After 1933, although suffrage spread throughout Latin America, with Paraguay being the last country to enact it in 1961, philosophy continued to be practised almost exclusively by men. However, feminist concerns sometimes arose tangentially in other philosophical discussions. The Mexican philosopher Leopoldo Zea (1912-) focused on women, indigenous peoples and children as early as 1952 in his analyses of marginality (see Marginality; Mexico, philosophy in). Similarly, theories of liberation often selfconsciously advocated the rights of women (see Liberation philosophy).

Since the 1980s Latin American women have entered the academy in greater numbers, many of whom have become actively engaged in feminist philosophy in addition to other philosophical areas of expertise. Scholars of literature and the social sciences have increasingly drawn philosophers into their debates about feminist theory. There has been an increase in the number of feminist conferences, books and journals, most notably in Mexico and Argentina, as well as throughout Latin America. *Fem* (1975-) and *Debate feminista* (1988-) are journals of feminist thought published in Mexico. *Feminaria* has been published in Argentina since 1988.

One of the key points of debate among feminist thinkers has centred on the primordial nature of gender and class. While some have maintained a traditional socialist idea that women must work towards justice with men rather than emphasize differences between men and women, others have argued that patriarchy must be subverted before there can exist a society unconstrained by divisions of gender. Thus, the former believe that class is primordial, while the latter privilege gender-based social order in their analyses.

Feminist thought in Latin America can be differentiated from many of its North American and European counterparts by an almost ever-present concern for the family and certain forms of Latin social life and relationships. Although alternative lifestyles do exist among women in Latin America, feminist philosophy has commonly attempted to end discrimination against women while simultaneously accepting the family as the fundamental social unit. Many translations of North American and European feminist thought are published in Latin America and read widely, despite the fact that their emphasis on the individual rather than the family is not easily adapted to Latin American contexts and is often seized on as evidence of fundamental cultural difference.

Feminist thought in Latin America continues to focus on the pressing concerns of Latin American societies first, without necessarily concerning itself with questions of its applicability to other regions of the world. Rather than being a separatist movement, feminist thought is often well-served by connections forged with other areas of contemporary thought in Latin America, such as the development of democracy, equality and socialism, religious and secular ethics and liberation philosophy.

See also: Feminism; Feminist political philosophy

AMY A. OLIVER

References and further reading

Barrios de Chungara, D. (1978) *Let Me Speak! Testimony of Domitila, a Woman of the Bolivian Mines*, trans. V. Ortiz, New York: Monthly Review Press.(Written with M. Viezzer.)

Castellanos, R. (1973) *Mujer que sabe latín (A Woman Who Knows Latin)*, Mexico: Secretaría de Educación Pública.(Counts among some of the feminists whose work has been rescued from relative obscurity.)

Cruz, J. Inés de la (*c*.1700) *A Woman of Genius: the Intellectual Autobiography of Sor Juana Inés de la Cruz*, trans. and intro. M. Sayers Peden, Salisbury, CT: Lime Rock Press, 1982.(Contains her autobiographical essay, 'Reply to Sor Filotea de la Cruz'.)

Hierro, G. (1985) *La naturaleza feminina: tercer coloquio nacional de filosof ía (Feminine Nature: Third National Convention on Philosophy)*, Mexico: Universidad Nacional Autónoma de México.(Anthology of articles by feminist philosophers in Mexico, many of which centre on the question of the existence of feminine nature.)

Hierro, G. (1990) *Ética y feminismo (Ethics and Feminism)*, Mexico: Universidad Nacional Autónoma de México.(Analysis of the ethics of 'being for others'.)

Menchú, R. (1984) *I, Rigoberta Menchú, an Indian Woman in Guatemala*, ed. with intro. E. Burgos-Debray, trans. A. Wright, London and New York: Verso.(An account of being a woman in a country where women traditionally have been voiceless.)

Patai, D. (1988) *Brazilian Women Speak: Contemporary Life Stories*, New Brunswick, NJ: Rutgers University Press.(Tales of women in Brazil.)

Schutte, O. (1993) *Cultural Identity and Social Liberation in Latin American Thought*, Albany, NY: State University of New York Press.(The chapter on cultural identity, liberation and feminist theory provides an excellent overview of this topic.)

Tristán, F. (1843) *La emancipación de la mujer (The Emancipation of Woman)*, trans. M.E. Mur de Lara, Lima: Editorial P.T.C.M., 1948.(A posthumous work completed from her notes by A. Constant.)

Vaz Ferreira, C. (1910) Lógica viva (Living Logic), in *Obras: Homenaje*, 25 vols, Montevideo: Cámara de Representantes de la República Oriental de Uruguay, 1957-63.(Vaz Ferreira distinguishes factual questions relating to knowledge from normative ones.)

Vaz Ferreira, C. (1933) *Sobre feminismo (On Feminism)*, Buenos Aires: Editorial Losada, 2nd edn. 1957, 3rd edn. 1963.(A collection of lectures on feminism that constitute the foundations of feminist theory in twentieth-century Latin America.)

Vaz Ferreira, C. (1957-63) *Obras: Homenaje (Complete Works: Homage)*, 25 vols, Montevideo: Cámara de Representantes de la República Oriental de Uruguay.(All of his original Spanish works are included.)

Zea, L. (1952) *Conciencia y posibilidad del mexicano (Consciousness and Possibilities of the Mexican Citizen*,

Mexico: Porrúa.(Analysis of the Mexican condition of marginality which includes discussion of women's treatment.)

Ferguson, Adam (1723-1815)

Rarely mentioned by philosophers except as companion of David Hume and Adam Smith, Ferguson contributed a political consciousness to the moral philosophy of eighteenth-century Scotland. In An Essay on the History of Civil Society (1767), Ferguson used a comparative method to reflect on a commercial society distinguished by refined division of labour and to caution against its political dangers. With his intentionally elevated rhetoric he sought to counter his philosophical contemporaries' analytical aloofness from the negative effects of the civility, commerce, security and critical philosophy they prized. Ferguson's textbooks and Roman history deserve philosophical attention for their help with interpreting his distinctive social diagnosis of the liberal political constitution.

Ferguson's Highlands birth and Gaelic made him unique among the figures prominent in the 'Scottish Enlightenment' (see Enlightenment, Scottish), but he was unexceptionably Whig in politics and Moderate in Church affairs. Born 1723 in Logierait near Perth, Ferguson moved to Edinburgh for divinity studies after an MA from the then provincial St Andrews. An intimate of the circle around David Hume, he early abandoned a clerical career. Ferguson served five years as professor of natural philosophy in the University of Edinburgh, thanks to political patrons of the literary set, and then, in 1765, began a popular twenty years as professor of pneumatics and moral philosophy. The professorial position was more didactic than scholarly, as was clear from Ferguson's effectiveness in the science chair; but Ferguson's *Essay on the History of Civil Society* (1767) gained him recognition beyond the lecture room and salon. There were French and German translations, and attention from philosophers on the Continent. In addition to four pamphlets on political controversies, Ferguson published two textbooks for his course, the second a retrospective work, in which he systematized the changes he made during his years of teaching and a long-standard, but largely derivative, *History of the Progress and Termination of the Roman Republic* (1785). He died in 1815. His books were staples of instruction in American colleges well into the nineteenth century.

Neither Hume nor Adam Smith was impressed by Ferguson as a theorist. Philosophers have left his work to on the one hand intellectual historians, who find there the epitome of republican misgivings - perhaps conditioned by a pre-Romantic Highlands sensibility - about the Scottish race to modernity after the Jacobite Rising of 1745, and on the other hand to sociologists, who honour Ferguson as a science-minded progenitor of sociological theories of modernity. With his invocation of classical models and his alarms against corruption in nations dedicated to individual gain rather than heroic action, Ferguson amplifies the discourse of 'civic humanism' or 'Machiavellian moralism', as many historians emphasize; but he also fits in with a wider class of didactic rhetoricians who sought to bolster a religiously latitudinarian moral community in eighteenth-century Scotland. Sociologists' claims to Ferguson are no less justified, first and especially by Ferguson's insistence that the study of human nature is an empirical study of humans in civil societies differing in design from one developmental stage to another, and second by his conjectural reconstruction of these designs to recognize not only political, economic and ideological differences, in the manner of Montesquieu, but also sociological mechanisms such as integration through conflict, stratification, and diverse forms of sociation. Karl Marx's incidental commendation of Ferguson as a materialist thinker with original insights into the harm done to workers by a refined division of labour (see Marx, K.) stimulated a tradition of commentary that emphasizes - and exaggerates - the extent to which Ferguson first identifies each historical period with a distinctive mode of subsistence, second, rationalizes the respective schemes of stratification, property, mentality, political formation by reference to this feature, and, third, identifies a potentially explosive tension in the latest stage - a commercial society whose political order is progressively undermined by its own socioeconomic arrangements.

Although modest about the originality and rigour of his efforts, Ferguson counted himself a participant in the philosophical debates of his time and, indeed, as Adam Smith's competitor in addressing what George Davie calls 'the central problem of Scottish philosophy', namely, reconciling material advance with traditional moral standards in a 'teachable philosophy of civilization' (Davie 1976). Viewed from this perspective, Ferguson's writings appear as exercises in complementarity, in which their author is seeking a reasonable strategy for co-ordinating two distinct modes of reflection. On one side is an analytical model of the progressive development of the human species through a sequence of stages in the history of civil society, with each stage conceived as a configuration of diversely formed productive, legal, moral, intellectual and hierarchical relations, some more

systematic than others. On the other side is the evocation of a practical capacity to act in the diverse historical contexts, with the quality of actions assessed by standards of active virtue, especially manifest in political practice.

Ferguson submits the first level of study to Newtonian rules of method, in principle; and his work often varies themes from Hume and Smith. The second type of knowledge he distinguishes from the first as the knowledge appropriate to actors, not spectators; and he attacks Smith for neglecting this dimension and for allowing moral psychology or economic laws to stand in for responsible decisions. Moral conduct is opaque without a science of the facts of human existence, but incoherent without practical understanding. Historical stages constitute scenes for actions subject to principles derived from a practical philosophy of elevated happiness. Political arrangements express modes of action designed to counter the negative effects of adventitious social circumstances. The essay form bypasses philosophical difficulties in interrelating these levels, but Ferguson's experiment gains in philosophical seriousness from his astute reflection on literary forms like the essay.

In his moral philosophy class, as documented in the *Institutes of Moral Philosophy* (1768) and the retrospective *Principles of Moral and Political Science* (1792), Ferguson grounded the 'history of the species' in a psychological theory of the individual, and elaborated a distinction between explanatory and normative theories. Although his decisive construct, of 'man's progressive nature', has the wobble of improvisation, the effort to constitute a structured connection between the distinct domains marks an advance in the literature of Common Sense responses (see Common Sense School) to Hume. Put in the context of his view of constitutions as imperfect syntheses of conflicts, Ferguson's designs are suggestive for a period of philosophical experimentation.

See also: Human nature, science of in the 18th century; Social science, history of philosophy of

DAVID KETTLER

List of works

Ferguson, A. (1767) *An Essay on the History of Civil Society*, ed. and intro by D. Forbes, Edinburgh: Edinburgh University Press, 1966; ed. D. Forbes with intro. L. Schneider, New Brunswick, NJ, and London: Transaction, 1980.(Forbes 1966 edition is a reprint of the first edition, with a collation of the variants in the seventh edition of 1814. Ferguson's best-known work, presenting the moral and political issues of the time as a function of the 'advancement of civil and commercial arts' and emphasizing the threat of civic corruption in a 'polished and commercial' society.)

Ferguson, A. (1768) *Institutes of Moral Philosophy*, New York: Garland, 1978.(1773 edition, 1785 edition is expanded; contemporary translations in French, German, and Russian. A textbook in which Ferguson elaborates the claim of a natural congruence betwen the optimal workings of human impulses to pursue happiness, as established by a science of human nature avowedly Baconian in method, and Ciceronian moral teachings.)

Ferguson, A. (1785) *The History of the Progress and Termination of the Roman Republic*; new edition, Edinburgh: Bell & Bradfute, 1825.(Enlarged and revised in 1793; several nineteenth-century American editions. A narrative history, covering the period from the beginning of the First Punic Wars to the end of the reign of Augustus. Ferguson punctutates the text with moralizing set-pieces on Caesar, Brutus and others, but he also makes a contribution to the problem of' 'republican' historiography, interweaving the actions of leading actors with the operation of social forces, moral climates and political institutions.)

Ferguson, A. (1792) *Principles of Moral and Political Science*, preface L. Castiglione, New York: AMS Press, 1973.(A philosophically more ambitious reaction of Ferguson's textbook teachings, now divided into a separate volume on 'the facts' and another on the 'excellences or defects' of human nature. The unifying theme is a conception of man's 'progressive' nature.)

Philip, W. (1986) *The Unpublished Essays of Adam Ferguson*, ed. and commentary W.M. Philip, Kilberry, Argyll: W.M. Philip.(Ferguson's brief essays, written in old age, range from mildly pious meditations to astute political analyses. A pugnacious essay challenges Adam Smith's conception of moral theory.)

Ferguson, A. (1995) *The Correspondence of Adam Ferguson*, ed. V. Merolle, Brookfield, VT: William Pickering. (Ferguson's correspondence is of interest mainly from the standpoint of biography and political history.)

References and further reading

Allen, D. (1993) *Virtue, Learning and the Scottish Enlightenment*, Edinburgh: Edinburgh University Press.(A discussion of Ferguson and other Enlightenment historians in the context of a century of 'self-consciously

edificatory discourse'.)

Brewer, J. (1986) 'Adam Ferguson and the Theme of Exploitation', *British Journal of Sociology* 37: 461-87. (Ferguson is viewed as a precursor of post-Marxist sociology, realistically insightful about class relations, without reducing the theory of exploitation to a mechanical economic determinism. Author rejects the emphasis on 'virtue' in the standard interpretations.)

Broadie, A. (1990) *The Tradition of Scottish Philosophy: A New Perspective on the Enlightenment*, Edinburgh: Polygon.(Useful re-analysis of issues between Common Sense philosophers and Hume.)

Davie, G. (1976) 'Berkeley, Hume, and the Central Problem of Scottish Philosophy', in D.F. Norton *et al.* (eds) *McGill Hume Studies*, San Diego, CA: Austin Press.(The eighteenth-century problem constellation is defined by the conflict between the insights of Mandeville and Hutcheson, and this sets Hume and his circle the task of a theory of civilization that recognizes the realism of the one without abandoning the social idealism of the other.)

Davie, G. (1990) *The Scottish Enlightenment and Other Essays*, Edinburgh: Polygon, 1991.(The philosophical problem constellation.)

Emerson, R.L. (1990) 'Science and Moral Philosophy in the Scottish Enlightenment', in M.A. Stewart (ed.) *Studies in the Philosophy of the Enlightenment*, Oxford: Clarendon Press, 11-36.(Representative study by prolific intellectual historian, here cautioning against over-emphasis on political discourse in interpreting Ferguson, among others.)

Hamowy, R. (1987) *The Scottish Enlightenment and the Theory of Spontaneous Order*, Carbondale, IL: Southern Illinois University Press.(A short monograph on an analytical figure developed by F. Hayek.)

Hont, I. and Ignatieff, M. (1983) *Wealth and Virtue*, Cambridge: Cambridge University Press.(Contains several statements of the 'civic humanist' thesis.)

Jack, M. (1989) *Corruption and Progress: The Eighteenth-Century Debate*, New York: AMS Press.(Introductory, comparing Ferguson with Mandeville and Rousseau.)

Kettler, D. (1965) *The Social and Political Thought of Adam Ferguson*, Columbus OH: Ohio State University Press.(A sociology of knowledge approach, with special emphasis on Ferguson as a representative 'intellectual'; as well as university teacher and popular moralist.)

Kettler, D. (1977) 'History and Theory in Ferguson's *Essay on the History of Civil Society*: A Reconsideration', *Political Theory* 5 (4): 437-60.(A challenge to the interpretation that Ferguson's theory is derived from his conception of historical stages. Instead, a theory of human action is seen as complementary to the proto-sociological theory of a developmental process, corresponding to Ferguson's distinction between the point of view of actors and spectators.)

Kettler, D. (1978) 'Ferguson's Principles: Constitution in Permanence', *Studies in Burke and His Time* 19 (3): 208-22.(An application of Kenneth Burke's concept of 'constitution' to Ferguson's integration of the complementary elements in his social and political thought.)

Lehman, W. (1930) *Adam Ferguson and the Beginnings of Modern Sociology*, New York: Columbia University Press.(The classic of the sociological thesis.)

Meek, R. (1976) *Social Science and the Ignoble Savage*, Cambridge: Cambridge University Press.(The four stages and history of the Americas.)

Sher, R. (1990) 'Professors of Virtue: The Social History of the Edinburgh Moral Philosophy Chair in the Eighteenth Century', in M.A. Stewart (ed.) *Studies in the Philosophy of the Enlightenment*, Oxford: Clarendon Press.(Highlights didactic functions of professorship.)

Ferrier, James Frederick (1806-64)

Ferrier represents the transition within nineteenth-century Scottish philosophy from the tradition of common-sense realism begun by Thomas Reid, the last major exponent of which was Ferrier's mentor, Sir William Hamilton, to versions of idealism influenced by German philosophers, especially Hegel. Although he is largely forgotten, Ferrier merits study for at least two reasons. First, he had a role in importing Hegelian ideas into British thought; and second there are parallels between his arguments and those advanced by antirealist philosophers in the analytical tradition.

Ferrier was born in Edinburgh into a Scottish legal family. He studied at Edinburgh University (1825-7) and at Magdalen College, Oxford (1828-32), then began a short-lived legal career before turning to writing. In 1834 he toured Germany and in Berlin was taken with the ideas and cultural influence of Hegel who had died there three years previously. In 1837 Ferrier married and around that time began writing and publishing philosophical work. His first and most accessible publication was 'An Introduction to the Philosophy of Consciousness' which appeared in seven parts in *Blackwood's Magazine* during 1838-9. Subsequent works included 'Berkeley and Idealism' (1842); his *magnum opus*, the *Institutes of Metaphysics* (1854); and 'Scottish Philosophy, the Old and the New' (1856). Further material was published posthumously, such as the 'Lectures on Greek Philosophy' (1875) and 'Criticism of Adam Smith's Ethical System' (1986). Ferrier began his academic career as Professor of Civil History at Edinburgh (1842-6) but failed in his efforts to succeed Sir William Hamilton (1791-1856) in the Chair of Logic and Metaphysics there. In 1845, however, he was appointed Professor of Moral Philosophy at St Andrews, where he remained until his death.

Although much indebted to the teaching and writings of William Hamilton, who had edited the works of Thomas Reid, the 'father' of the Scottish 'common sense' tradition, Ferrier rejected the doctrines and methods of the 'Scottish Philosophy', as that tradition had come to be titled (see Common Sense School). In an effort to refute Humean scepticism,Reid and his followers had argued that in cognition the mind is directly engaged with reality and not with mental intermediaries such as 'impressions' or 'ideas'. This cognitive power was taken to be a constitutive feature of the human mind, and the study of it and of other mental faculties had come to be a major focus of interest for Reid's realist contemporaries and successors, who sought thereby to develop a 'Science of the Human Mind'.

As this study became more empirical a new form of scepticism began to emerge, since it was realized that all the method could reveal is how things appear to consciousness. Given a different empirical constitution, however, it might be that a subject's experiences would be quite unlike those taken to be common among human beings as they actually are. The resulting conception of the 'Relativity of Knowledge', as it came to be called, also drew upon Kantian themes which suggested the possibility that we are not even acquainted with the world and the self as they are *in themselves*, but only with subjectively conditioned *appearances*. Ironically, then, by the time of Hamilton, the anti-sceptical realist tradition initiated by Reid had come to the point of finding our knowledge of self and of reality problematic.

Ferrier's rejection of this tradition was due in part to its tendency to generate scepticism, which he judged to be ultimately incoherent, but also to what he saw as its disastrous substitution of empirical psychology for a priori reasoning. Following in the tradition of the Rationalists, in particular Descartes, Ferrier sought to show that philosophy is able to determine the necessary limits and structure of reality, showing what can and must be so and not restricting itself to what contingently exists. In the *Institutes of Metaphysics* he even goes as far as to claim that his entire philosophical system, comprised of 'epistemology' (a term of which Ferrier was the author), 'agniology' and 'ontology' (the theories of knowledge, ignorance and being, respectively) is derivable by deduction from the law of identity: 'that a thing must be what it is. A is A'.

Ferrier's charge against his predecessors is related to the objection made by later writers to the psychologizing of logical reasoning. One aspect of this is his observation that empirical study is incapable of refuting philosophical arguments such as those taken to imply scepticism. His own response to the sceptic is unmistakably aprioristic, involving the claim that epistemological and metaphysical relations between the subject of cognition, the act of cognitive consciousness and the object of cognition are internal or non-contingent. All 'subjects' and all

‘objects’ are, as such, necessarily terms of actual or possible cognitive relations. Even ignorance ensures the possibility of knowledge inasmuch as one can only fail to know that which can be known. Thus, anticipating arguments of later phenomenologists and antirealists, he concluded that the idea that reality might in principle transcend our capacity to know it is an incoherent one: the real is the knowable.

Among his contemporaries, other than those within the sphere of his immediate influence in Edinburgh and St Andrews, Ferrier’s idealist system was not well-received. Those who knew his work recognized his ability but were unpersuaded by his doctrines. In part that was probably due to the fact that the temper of the period was increasingly one of empiricism and naturalism, approaches opposed to that of a priori metaphysics. Set against this, however, there are interesting parallels between aspects of Ferrier’s philosophical system and ideas developed within analytical philosophy by writers such asMichael Dummett andHilary Putnam, concerning the extent to which ‘reality’ is a function of our ways of thinking, (see Realism and antirealism), and there is certainly scope for useful scholarship exploring these points of similarity. Indeed, the time is overdue for a proper assessment of Ferrier’s philosophical contribution.

See also: Rationalism; Scepticism

JOHN J. HALDANE

List of works

Ferrier, J.F. (1854) *Institutes of Metaphysics*, Edinburgh: Blackwood.(Ferrier’s definitive presentation of his version of Absolute Idealism. It is markedly rationalist in method, attempting to derive important claims about knowledge and reality from the ‘law of identity’ - that a thing must be what it is.)

Ferrier, J.F. (1875) *Philosophical Works of the late James Frederick Ferrier*, ed. A. Grant and E.L. Lushington, Edinburgh: Blackwood, 3 vols.(Comprises the *Institutes* (vol. 1), *Lectures on Greek Philosophy* (vol. 2) and *Philosophical Remains* (vol. 3), the last containing ‘Introduction to the Philosophy of Consciousness’. This edition of Ferrier’s works has been reprinted by Garland Press.)

Ferrier, J.F. (1986) ‘The 1849-1850 Lectures of J.F. Ferrier: Criticism of Adam Smith’s Ethical System’, ed. A. Thomson, *Edinburgh Review* 74: 100-7.(Two short essays in which Ferrier expounds Smith’s account of moral judgment as resting upon natural sympathy and then argues that such sympathy is acquired, not innate.)

References and further reading

Davie, G. (1967) ‘J. Ferrier’, in *Encyclopedia of Philosophy*, ed. P. Edwards, New York: Macmillan.(A brief account of Ferrier’s life and work emphasizing the interest of Ferrier’s earlier ‘phenomenological’ writings.)

Haldane, E. (1899) *James Frederick Ferrier*, Edinburgh: Oliphant.(Notwithstanding a more recent study by Thomson (see below), this remains the most comprehensive and reliable introduction to Ferrier’s life and work.)

Mayo, B. (1969) *The Moral and Physical Order: A Reappraisal of James Frederick Ferrier*, St Andrews: University of St Andrews.(A short but interesting discussion of aspects of Ferrier’s epistemology and moral psychology which argues that they anticipate ideas associated with twentieth-century analytical philosophy).

Thomson, A. (1964) ‘The Philosophy of J.F. Ferrier’, *Philosophy* 39: 46-62.(An account of Ferrier’s thought, as presented in the ‘Introduction to the Philosophy of Consciousness’, which holds that among Scottish philosophers he is second only to Hume in the originality and value of his ideas.)

Thomson, A. (1985) *Ferrier of St Andrews: An Academic Tragedy*, Edinburgh: Scottish Academic Press.(Although much more detailed in biographical aspects thanHaldane (1899), like Thompson’s *Philosophy* article (1964) this exaggerates Ferrier’s philosophical importance. It also makes controversial claims concerning his private life.)

Feuerbach, Ludwig Andreas (1804-72)

Ludwig Feuerbach, one of the critical Young Hegelian intellectuals of the nineteenth century, has become famous for his radical critique of religious belief. In Das Wesen des Christentums (Essence of Christianity) (1841) he develops the idea that God does not exist in reality but as a human projection only, and that the Christian principles of love and solidarity should be applied directly to fellow humans rather than being regarded as an indirect reflection of God's love. In religion, the believer 'projects his being into objectivity, and then again makes himself an object of an object, another being than himself'. Religious orientation is an illusion and is unhealthy, as it deprives and alienates the believer from true autonomy, virtue and community, 'for even love, in itself the deepest, truest emotion, becomes by means of religiousness merely ostensible, illusory, since religious love gives itself to man only for God's sake, so that it is given only in appearance to man, but in reality to God' (Feuerbach 1841: 44, 48). In Grundsätze der Philosophie der Zukunft (Principles of the Philosophy of the Future) (1843) he extends his criticism to all forms of metaphysics and religion: 'True Dialectics is not the Monologue of the sole Thinker, rather the Dialogue between I and Thou', he writes in paragraph 62 (1846-66 II: 345), criticizing in particular his former teacher Hegel. The philosophy of the future has to be both sensual and communal, equally based on theory and practice and among individuals. In an anonymous encyclopedia article (1847) he defines his position: 'the principle from which Feuerbach derives everything and towards which he targets everything is "the human being on the ground and foundation of nature"', a principle which 'bases truth on sensuous experience and thus replaces previous particular and abstract philosophical and religious principles' (1964- III: 331). Feuerbach's sensualism and communalism had great influence on the young Karl Marx's development of an anthropological humanism, and on his contemporaries in providing a cultural and moral system of reference for humanism outside of religious orientation and rationalistic psychology. In the twentieth century, Feuerbach influenced existential theology (Martin Buber, Karl Barth) as well as existentialist and phenomenological thought.

1 Philosophy of the I and the Thou

Born in Landshut, son of the legal scholar and Bavarian diplomat Anselm von Feuerbach, Ludwig Andreas Feuerbach studied theology with Karl Daub in Heidelberg and philosophy with Hegel in Berlin. He became lecturer in philosophy at the Bavarian university of Erlangen in 1829, but retired into private life after his marriage in 1837. His later years were without much communication and social activities; he died a poor and sick man in Rechenberg, near Nuremberg.

The interaction between the individual and the worlds of facts and thought, and the interrelatedness of individuals, were the central topics in Feuerbach's thinking. His dissertation *De infinitate, unitate atque communitate rationis (On the Infinity, Unity and Universality of Reason)* (1828) concerns the relationship between the individual and the general, the subjective and the objective; here his answer is that panlogic rationality is the universal individual as well as the general mode of essence and existence. Two years later he published his fiercest critique of Christian belief in immortality and eternal life after death in *Gedanken über Tod und Unsterblichkeit (Thoughts on Death and Immortality)* (1830). Here he holds that the mystical experience of sensual love and interpersonal community has infinity and eternity in itself, and that the neglect of this world and its riches in favour of the anticipated eternal world to come causes unhappiness and great discontent. Both studies mark two different attempts at a non-religious, humanistic transformation of traditional mystical insights concerning the unity of object and subject into, first, panlogical rationalism and, second, sensualistic realism. The publication of the critique of immortality, even though it appeared anonymously, made it impossible for Feuerbach to get a permanent salaried position as professor of philosophy in Christian Bavaria; he quit his teaching post after nine years of only moderately successful academic teaching at the University of Erlangen.

Feuerbach's existentialist concept of philosophy is explicated in *Abelard and Heloise* (1834) and *Pierre Bayle* (1838), as well as in his selected studies on Spinoza and Leibniz. In *Grundsätze der Philosophie der Zukunft* (*Principles of the Philosophy of Future*) (1843) he asserts that the philosophy of the future is 'the complete, coherent, and absolute resolution of theology into anthropology', into the 'unity of the I and the Thou': 'The essence of man is contained only in community, in the unity of man with man - a unity which however is founded only in the reality of the differences between I and Thou'. Thus, the new philosophy 'replaces religion… contains the essence of religion… truly is religion'. Epistemologically, reality resides in human rational and sensuous

perception, in the rational, sensuous and passionate nature of man; 'Truth, reality, and sensation are identical. Only the sensuous being is a true and real being' (Feuerbach 1846-66 II: 344, 346). In critically reconstructing the history of religious systems of reference, Feuerbach finds in references to God the believers' complex reference to themselves. Thus his method of critical reconstruction of previous systems of reference and self-reference reveals, as Wartofsky (1977) observes, a mode of self-revelation and self-articulation which represents the essential pattern of human consciousness and self-understanding, a consequent application of the existentialist interpretation of Hegel's *Phenomenology of Spirit* in Feuerbach's anthropological interpretation: 'the true critique lies in the development itself'.

This unorthodox approach to traditional metaphysics and anthropology influenced twentieth-century theologians and philosophers such as Martin Buber, Karl Barth, Karl Loewith, Henri Arvon and Louis Althusser. Feuerbach's influence on the young Marx also played an important role in the post-Leninist interpretation of the humanism of the young Marx as an alternative to Marxism-Leninism and Stalinist interpretation of political economy.

In his last period in *Über Spiritualismus und Materialismus (On Spiritualism and Materialism)* (1866) Feuerbach developed an eudaimonistic and utilitarian ethics on the basis of the I-Thou anthropology, and a critique of perverted forms of seeking happiness and community in religion: 'Good is the acceptance, bad the rejection of the drive to happiness. Happiness, but not reduced into one single person, rather disseminated among different persons, I and Thou integrating, therefore not one-sided but dual-sided and all-sided, is the principle of morality'. Sexual happiness is his paradigm for 'the human drive to happiness, which can only be fulfilled in and through the happiness of the other' (Feuerbach 1846-66 X: 62). The individual moral conscience thus works on the basis of reciprocity in assessing moral norms and attitudes: 'The conscience is nothing else than the I which looks at things from the perspective of the vulnerable Thou, nothing else than the proxy for the happiness of the other on the basis and on the calling of one's own happiness' (Feuerbach 1846-66 X: 65).

2 Critique of religion

Feuerbach's humanist approach prevented his philosophy of religion from merely rejecting religion as unrealistic and unfounded. On the contrary, he believed that religion contains the dreams and visions of individuals and cultures in an indirect way, a human product of self-transcending and visualizing oneself and humanity in ideal terms, not a divine inspiration: religion is the 'knowledge of the infinite, it is therefore and can be nothing else than the consciousness which man has of his own - not finite and limited - but infinite nature' (Feuerbach 1846-66 VII: 26). Feuerbach therefore rejects the notion that he is a promoter of atheism; he calls himself 'a natural philosopher in the domain of the mind', an 'anthropocentric' thinker: 'the mystery of theology is anthropology, that of the divine being the human being' (Feuerbach 1846-66 V: xiv-xv). This was exactly the interpretation of religion used by Hegel to interpret Greek and Roman mythology and religion as centred around the human world and simply transposing human characters, virtues and vices into the gods and goddesses. While Hegel's main interest in the philosophy of religion was dialectically and speculatively to harmonize Christian tradition with speculative idealism, for Feuerbach all religious references are models of 'alienation' of man from himself, as God is nothing else but the artificially created outside nature of man's own inside nature (see Alienation §2). Feuerbach rejects religion to the extent that it separates God and humans, and deprives humans of their best by making them a part of the divine; but he accepts religion to the extent that it recognizes divine powers and challenges which are genuinely human.

Feuerbach originally intended to entitle *Das Wesen des Christentums* (*Essence of Christianity*) (1841) as 'Critique of Pure Unreason' in order to set it alongside Kant's *Critique of Pure Reason*; and indeed the composition of his book mirrors that of Kant. Its first part characterizes religion as a human product and its positive content (that is, the 'true and anthropological essence of religion' which understands God as 'love' in the New Testament or as 'a moral being or law' in the Old Testament), just as Kant in the first part of the *Critique of Pure Reason* develops the positive capacities of epistemology in his transcendental analytics and aesthetics. The second part deconstructs religious speculation and metaphysics as being unrealistic and inhuman (that is, 'the false or theological essence of religion'), just as Kant destroys the arguments for metaphysics and the existence of God independently of human projection and existence (see Kant, I. §8). Thus, in making Christian monotheism his predominant object of critique, he presents an anthropological reduction of what in Hegel's interpretation still had

the form of speculative transcendentalism. The human being is not a bearer of reason, subordinated to as well as participating in the 'idea', but a natural being existing in relationship with others. The revised second edition (1843) moves away from the dominance of the 'species being' model of human religious projection and alienation, which was still strongly influenced by Hegel, towards a more naturalistic sensualistic approach under the influence of reading Luther.

Later, in a short essay on the nature of religion, *Das Wesen der Religion (Essence of Religion)* (1845), Feuerbach additionally identifies angst and fear of unknown natural elements as the prime sources of religious projection. This generalizes his earlier theory into the explanation of religion as a response to social as well as natural environmental challenges based in feelings of dependence of and exposure to nature: 'man's original dependence and the feelings of dependency relates to nothing else than nature'; therefore, '"spirit" is nothing other than a general term to name essences, things, utensils which man is confronted with and which are different to his own products, he also uses the collective term "nature"' (Feuerbach 1964- X: 4).

In his widely read *Vorlesungen über das Wesen der Religion (Lectures on the Essence of Religion)* (1851), originally delivered publicly in the revolutionary year 1848 on the invitation of students and citizens of Heidelberg, Feuerbach describes the cultural and social implications and consequences of religion and belief. Here, as in the *Essence of Religion*, the true object of religion and religious self-articulation or alienation is not primarily the individual and the social environment, but nature all-encompassing, the sum of powers and forces on which the individual feels absolutely dependent while simultaneously seeking freedom and self-determination.

His final exercise in promoting the thesis that 'man created God' is the voluminous book *Der Ursprung der Götter nach den Quellen des classischen, hebräischen und christlichen Altertums (Theogony following the Sources of the Classical, Hebrew, and Christian Past)*, written between 1852 and 1857, with an even greater arsenal of quotes and references from a broad variety of religious literature. Different forms of religion demonstrate different attitudes of humans towards themselves and towards nature. While polytheism expresses the subordination of man under a multitude of external powers in nature, monotheism suppresses those many forces and powers in nature under one single rule or ruler, indirectly under man, who has created this divine supreme power after his own human image.

It is this analytical and educational drive to 'transform friends of God into friends of man, believers into thinkers, devotees of prayer into devotees of work, candidates for the hereafter into students of this world' and the I-and-Thou which became instrumental in shaping modern existentialist theology, both Jewish (Martin Buber) and Protestant (Karl Barth). Despite being religious believers, Buber and Barth were nevertheless struck by Feuerbach's suggestion that all previous forms of theology and dogmatics, including Hegel's philosophy of religion, were nothing other than 'esoteric psychology', and they themselves became critics of religious fundamentalism. Most of Feuerbach's theological critics agree that the importance of religion as a mode of reference to oneself, to others and to the social and natural world rarely has been grasped as comprehensively as it was by Feuerbach.

3 Individual emancipation versus political revolution

Feuerbach refused to accompany Marx, Ruge and other political revolutionaries and liberals into the course of direct political argumentation and action. He held that the time was not yet right for a political revolution because individual emancipation and the development of a culture of self-determination had not run its course in Germany. Therefore he chose the emancipatory and educative approach, analysing and criticizing traditional models of orientation in religion, philosophy and ideology in order to promote political change by promoting self-understanding and self-determination based on the principles of enlightenment. Contrasting his method to that of socialist political action, he compares his philosophical methodology to the methods of medical diagnosis, prognosis and therapy: 'I have, determined by internal and external factors, decided to assess the diseases of the head and of the heart of mankind'.

Feuerbach's belief that the grounds for a true revolution lay in the liberation of human sensuousness and feeling in interpersonal relations, not primarily in the externalities of political life, had an enormous influence on the more radical former disciples of Hegel, the Young Hegelians Bruno Bauer, Karl Marx, Friedrich Engels, and Max Stirner. Instead of becoming a member of the 1848 revolutionary parliament in Frankfurt, he chose to lecture

publicly in Heidelberg where he declared religion to be closely related to politics, emphasizing that his lectures were focused on practical rather than theoretical politics. He maintained that 'individualism in practical consequence is socialism, but not in the sense of French individualism'; for him 'the dissolution of theology into anthropology in the realm of thinking is the dissolution of the monarchy into the republic in the realm of practice, of life…. Imagination is the power of religion and imagination is the power of the monarchy. Only so long will monarchs rule over humans as humans will be ruled by imagination. Despots rule only where phantasies rule' (Feuerbach 1846-66 VIII: 460). It has been stressed in the literature (Arvon 1964; Cabada 1975; Kamenka 1970; Sass 1972) that Feuerbach is not merely a transitional figure towards Marx and Marxism-Leninism, but an alternative to the concept of political revolution and party elites.

Thus Feuerbach gave his own meaning to the term 'communist', in direct opposition to that employed by Marx and the socialist movement, when answering Max Stirner's question 'where does Feuerbach stand?': 'neither a materialist, nor an idealist, nor an identity philosopher is Feuerbach. So, what is he? He is in thought what he is in his deeds, in spirit what he is in flesh, in essence what he is in his senses - human being [mensch]; for Feuerbach only sees the essence of man in community: communal being, communist' (Feuerbach 1846-66 V: 359).

In his *Grundsätze der Philosophie der Zukunft* (*Principles of the Philosophy of the Future*) (1843) Feuerbach uses the image of the ellipse having two focal points, as compared to the circle having only one centre, to compare the purely intellectual philosophy of Hegel and other schools of rationalism and metaphysics with his own approach of 'head and heart', rationality and sensuality, as the two commanding principles for theory and practice (Feuerbach 1846-66 II: 344). Marx took this image of the two centres of the ellipse from Feuerbach in 1844 in order to identify philosophy and proletariat as 'head and heart' of the emancipatorial process of history, this time in terms of political revolution, not educational and critical evolutionary emancipation as Feuerbach had done.

See also: Hegelianism

HANS-MARTIN SASS

List of works

Feuerbach, L. (1846-66) *Ludwig Feuerbach's Sämmtliche Werke*, Leipzig: Otto Wigand.(This is the edition of collected works published by Feuerbach himself, with revised versions of earlier articles and new works.)

Feuerbach, L. (1903-11) *Ludwig Feuerbach Sämtliche Werke*, ed. W. Bolin and F. Jodl, Stuttgart, 10 vols; reprint 1959-64; 3 suppl. vols, ed. H.M. Sass, 1962-4.(Collected works, with a bibliography in volume 11.)

Feuerbach, L. (1964-) *Historical-critical edition: Ludwig Feuerbach Gesammelte Werke*, ed. W. Schuffenhauer, Berlin: Akademie Verlag, vols 1-12; 17; 19 published.(This edition is also available in electronic form on disk or CD-ROM, eds M. Neumann and S.T. Stoler, Georgetown University Washington DC: Center for Text and Technology, 1994.)

Feuerbach, L. (1828) *De infinitate, unitate atque communitate rationis (On the infinity, unity and universality of reason)*, University Archives Erlangen.(Dissertation manuscript, first published in 1962 in volume 11 of the 1846-66 *Sämtliche Werke*, pages 11-62.)

Feuerbach, L. (1830) *Gedanken über Tod und Unsterblichkeit* (*Thoughts on Death and Immortality*), ed. and trans. J.A. Massey, Berkeley, CA: University of California Press, 1980.(Originally published anonymously in Nürnberg.)

Feuerbach, L. (1834) *Abelard und Heloise*, Ansbach: Brügel.(Outlines Feuerbach's existentialist concept of philosophy.)

Feuerbach, L. (1838) *Pierre Bayle*, Ansbach: Brügel.(Another explication of Feuerbach's existentialist concept of philosophy.)

Feuerbach, L. (1839) 'Zur Kritik der Hegelschen Philosophie'; trans. Z. Hanfi, 'Towards a Critique of Hegelian Philosophy', in Z. Hanfi (ed.) *The Fiery Brook: Selected Writings of Ludwig Feuerbach*, New York: Anchor Books.(This work established Feuerbach's position as a spokesman for the Hegelian Left.)

Feuerbach, L. (1841) *Das Wesen des Christentums*, trans. M. Evans (George Eliot) as *Essence of Christianity*, 1854; new edn, intro. K. Barth, foreword H.R. Niebuhr, New York: Harper & Row, 1957.(Feuerbach's most influential work.)

Feuerbach, L. (1842) *Vorläufige Thesen zur Reformation der Philosophie (Provisional Theses Towards the Reform of Philosophy)*.(Develops Feuerbach's humanistic materialism.)

Feuerbach, L. (1843) *Grundsätze der Philosophie der Zukunft*, trans. and intr. M.H. Vogel as *Principles of the Philosophy of Future*, Indianapolis, IN and New York: Hackett Publishing Company, 1966.(Sketches the new

philosophy of ‘I and thou’.)
Feuerbach, L. (1844) *Das Wesen des Glaubens im Sinne Luthers, ein Beitrag zum Wesen des Christentums*, trans. M. Cherno as *The Essence of Faith according to Luther*, New York: Harper & Row, 1967.
Feuerbach, L. (1845) *Das Wesen der Religion (Essence of Religion),* in *Die Epigonen I*, Leipzig: Wigand, 117-78. (An essay on the nature of religion)
Feuerbach, L. (1851) *Vorlesungen über das Wesen der Religion, nebst Zusätzen und Anmerkungen*, trans. R. Manheim as *Lectures on the Essence of Religion*, New York: Harper & Row, 1967.(Originally delivered at the invitation of students in Heidelberg during the revolutionary year of 1848.)
Feuerbach, L. (1857) *Der Ursprung der Götter nach den Quellen des classischen, hebräischen und christlichen Altertums (Theogony following the Sources of the Classical, Hebrew, and Christian Past)*, in *Ludwig Feuerbach’s Sämmtliche Werke* IX, Leipzig: Otto Wigand, 1846-66.(A most extensive scholarly analysis of modes of religious thought.)
Feuerbach, L. (1866) *Über Spiritualismus und Materialismus*, in *Ludwig Feuerbach’s Sämmtliche Werke* X, Leipzig: Otto Wigand, 1846-66.(Develops a eudaimonistic and utilitarian ethics on the basis of I-Thou anthropology.)
Feuerbach, L. (1972) *The Fiery Brook. Selected Writings of Ludwig Feuerbach*, trans. Z. Hanfi, New York: Anchor Press.(Contains, among other selections, translations of ‘Zur Kritik der Hegelschen Philosophie’ (1839, 1657-), *Vorläufige Thesen zur Reform der Philosophie* (1843, vol. 2, 62-), and *Grundsätze der Philosophie der Zukunft* (1843).)

References and further reading

Amengual, G. (1980) *Critica de la religion y antropologia en Ludwig Feuerbach (Ludwig Feuerbach’s critique of religion and anthropology)*, Barcelona: Laia.(Comprehensive interpretation of Feuerbach’s anthropology and critique of religion.)
Arvon, H. (1964) *Ludwig Feuerbach ou la transformation du sacre (Ludwig Feuerbach, on the transformation of the sacred)*, Paris: PUF.(An existentialist interpretation of Feuerbach as one of the sources of modern existentialism.)
Barata-Moura, J. and Marques, V.S. (eds) (1993) *Pensar Feuerbach*, Lisbon: Edicoes Colibri.(Collection of Portuguese essays from 1991, commemorating the 150th anniversary of Feuerbach’s *Essence of Christianity*.)
Braun, H.J., Sass, H.-M., Schuffenhauer, W. and Tomasoni, F. (eds) (1990) *Ludwig Feuerbach und die Philosophie der Zukunft (Ludwig Feuerbach and the Philosophy of the future)*, Berlin: Akademie Verlag. (Collection of essays originally contributed to the Second International Feuerbach Conference in 1989.)
Braun, H.J. (ed.) (1994) *Solidarität oder Egoismus. Studien zu einer Ethik bei und nach Feuerbach (Solidarity or Egoism: Studies toward an ethics with and after Feuerbach)*, Berlin: Akademie Verlag.(Collection of conference papers from 1992.)
Cabada Castro, M. (1975) *El Humanismos Premarxista de Ludwig Feuerbach (The pre-Marxist humanism of Ludwig Feuerbach)*, Madrid: Bibliotheka des Autores Christianos.(Detailed interpretation of Feuerbach’s humanism and his influence on Marx.)
Cesa, C. (1972) *Studi sulla sinistra hegeliana (Studies on the left Hegelianism)*, Urbino: Argalia Editore(Interpretation of Feuerbach’s historiography (167-84) and philosophy of religion (185-248) within the context of Young Hegelian thought.)
Harvey, Van A. (1995) *Feuerbach and the Interpretation of Religion*, Cambridge: Cambridge University Press. (Classic study of Feuerbach’s epistemology and interpretation of religion in the context of the philosophical and theological disputes on the validity of Feuerbach’s critique of religion.)
Jaeschke, W. (ed.) (1992) *Sinnlichkeit und Rationalität. Der Umbruch in der Philosophie des 19. Jahrhunderts (Sensibility and rationality: the revolution in nineteenth-century philosophers)*, Berlin: Akademie Verlag. (Collection of essays, with an English-language bibliography, 1873-1991, compiled by Scott Stebelman.)
Kamenka, E. (1970) *The Philosophy of Ludwig Feuerbach*, New York: Praeger.(Focuses mainly on the interpretation of Feuerbach’s humanist I-Thou philosophy.)
Loewith, K. (1964) *From Hegel to Nietzsche. The Revolution in Nineteenth Century Thought*, trans. D. Green, New York: Holt, Rinehart & Winston.(Classic study of Young Hegelians and Feuerbach’s role in promoting the radicalization of critical theory.)
Luebbe, H. and Sass, H.-M. (1975) *Atheismus in der Diskussion. Kontroversen um Ludwig Feuerbach (Atheism in debate: Controversies concerning Ludwig Feuerbach)*, Munich: Kaiser.(Collection of essays, originally contributions to the First International Feuerbach Conference in 1973.)
Rawidowitz, S. (1964) *Ludwig Feuerbach*, Berlin: de Gruyter.(An elaborate scholarly interpretation (in German)

of Feuerbach and his influence in the history of ideas.)
Sass, H.-M. (1972) *Ludwig Feuerbach*, Reinbek bei Hamburg: Rowohlt, 4th edn, 1994.(Illustrated intellectual biography (in German) using as yet unpublished material; Korean translation by Moon-Gil Chung, 1986.)
Sass, H.-M. (1983) ‘The Transition from Feuerbach to Marx. A Reinterpretation’, *Studies of Soviet Thought* 26: 123-42.(Analyses the conceptual differences between Feuerbach and Marx from a post-Leninist perspective.)
Sass, H.-M. and Wartofsky, M.W. (eds) (1978) *Feuerbach, Marx and the Left Hegelians*, special issue of *The Philosophical Forum*.(Contains English translations of minor works by Feuerbach and of some correspondence.)
Tomasoni, F. (1986) *Ludwig Feuerbach e la natura non umana (Ludwig Feuerbach and the non-human nature)*, Florence: La Nuova Italia Editrice.(An interpretation of Feuerbach’s philosophy of nature (in Italian), based on previously unpublished text edited by Tomasoni.)
Wartofsky, M.W. (1977) *Ludwig Feuerbach*, Cambridge: Cambridge University Press.(Classic epistemological interpretation of Feuerbach’s ideas.)

Feyerabend, Paul Karl (1924-94)

Feyerabend was an Austrian philosopher of science who spent most of his academic career in the USA. He was an early, persistent and influential critic of the positivist interpretation of science. Though his views have some affinities with those of Thomas Kuhn, they are in important ways more radical. Not only did Feyerabend become famous (or notorious) for advocating 'epistemological anarchism' - the position that there is no such thing as scientific method, so that in advancing scientific research 'anything goes' - he also argued that the scientific outlook is itself just one approach to dealing with the world, an approach that is not self-evidently superior in all respects to other approaches. This radicalism led to his being widely attacked as an irrationalist though perhaps he might better be seen as a sceptic in the humane and tolerant tradition of Sextus Empiricus and Montaigne.

Paul Feyerabend was born and educated in Austria. His studies were interrupted by the Second World War, in which he served as an infantry officer in the German army and was awarded the Iron Cross. His wartime service ended when a bullet in the spine left him with injuries from which he never fully recovered. After the war, he resumed his education, eventually switching from physics to philosophy of science. His academic career, which began in England and included a year working with Karl Popper, was spent mostly in the USA.

Feyerabend was an early and persistent critic of the once-dominant logical positivist interpretation of science. Like other such critics - Thomas Kuhn and N.R. Hanson for example - Feyerabend insisted on paying close attention to the history of science and to contemporary scientific debates. Again like them, he thought that such attention showed standard philosophical accounts of science to be wildly at variance with actual scientific practice.

Feyerabend is perhaps best remembered for two views. First, he was an early advocate of the view that competing theories are often 'incommensurable', meaning that there is no common standard by which their respective merits can be judged. Positivist philosophers of science sought such a standard in the 'basic' or 'protocol' statements of a purely observational language, uncontaminated by theoretical preconceptions (see Logical positivism §4). But according to Feyerabend - and this aspect of his thought goes all the way back to his doctoral dissertation - there is no such language. The meanings of the observation-terms that figure in statements relevant to the confirmation and disconfirmation of scientific theories are affected by the theories in which they are embedded. This can be disguised by the retention of the same words, as when both classical mechanics and special relativity talk about 'length'. But classical and relativistic length are different quantities, so that the measurement of 'length' is a different operation, depending on which theory we are working with (see Incommensurability; Observation).

Feyerabend argued further that, even if there were a pure observation language, this would not guarantee commensurability, since competing theories typically address themselves to different ranges of data. It is rarely, if ever, the case that a new theory explains everything an old theory explains and more besides. In consequence, choosing a new theory over an old one will typically involve explanatory loss as well as gain. For example, after Newton, theories of the heavens no longer attempt to explain the number of the planets, though it was previously thought essential to do so.

The second idea for which Feyerabend became famous - or notorious - is his 'epistemological anarchism'. Feyerabend denied that there is such a thing as 'scientific method': there are no universally valid methodological precepts by which scientific practice either is or should be governed. If we insist that there must be some such precept, the only defensible candidate is 'anything goes' (see Pluralism).

One of Feyerabend's reasons for taking this position was that methodology can no more be abstracted from substantive, theoretical ideas than can observation. (The quest for a permanent, theory-neutral, observation language is just one way of trying to force science into a methodological straightjacket.) But, even if it could, strict adherence to abstract methodological rules (such as Popper's critical rationalism, with its emphasis on attempts to falsify theories via severe tests) would spell the death of scientific progress. The history of science shows that all theories, even the most successful, swim in a sea of anomalies; but this is reason to develop them, even with the aid of *ad hoc* hypotheses, not a reason for going back to the drawing board. Furthermore, old, well-developed theories typically do better than new theories by just about any methodological standard one cares to name. In part, this reflects the 'law of uneven development': a persuasive defence of a new theory in one area may depend on

theoretical advances in various ancillary domains, and there is no guarantee that these will appear on cue. New ideas therefore need breathing space if we are to appreciate their full potential. But granting them this breathing space can involve violating any and every canon of ’rationality’. Feyerabend did not deny that, if a new idea turns out to be a good one, it will be possible - after the fact - to demonstrate its superiority according to one’s favourite methodological prescriptions. His point, however, was that such prescriptions provide no guidance whatsoever to the practice of research. Whatever story one tells about the merits of an idea that worked out well, there will be alternatives that, in their undeveloped state, look worse but which, in their turn, could prove to be better. One simply never knows, which is why there are no rules to guide our efforts (see Scientific method §2).

Feyerabend’s views have clear affinities with Kuhn’s. Nevertheless, while he admired Kuhn, he criticized him for remaining in thrall to the idea of method. According to Feyerabend, Kuhn’s distinction between ’normal’ and ’revolutionary’ science applies to different *aspects* of scientific research but not, contrary to Kuhn himself, to separable *phases*. However, Feyerabend did not only see Kuhn’s model as failing to correspond to the actual workings of science: he also regarded it as a disastrous normative ideal. This shows that, his professed anarchism notwithstanding, Feyerabend did have methodological advice to offer. Where Kuhn commends working within a dominant theoretical tradition, until it collapses from internal strains, Feyerabend held that progress depends on the constant proliferation of theoretical ideas, hence continuous ‘philosophical’ criticism of entrenched views. He saw the desirability of proliferation as a consequence of the fact that observation-terms and methodological preferences do not float free of background theoretical ideas, so that the experiments that supposedly *confirm* such ideas are often better seen as *illustrating* them. To be fully understood, hence seriously tested, a theory needs to be criticized from the outside, not just the inside. And since there is no neutral standpoint, beyond all theoretical preconceptions, this means criticized from the perspective of some alternative theory. Feyerabend differed from Popper in many ways, but his epistemological views, with their scepticism about justification and stress on the importance of critical argument, retained a distinctly Popperian coloration.

At the time of their publication, Feyerabend’s writings, especially his book *Against Method* (1975), enjoyed a considerable *succes de scandale*. They have come to seem less outrageous. While some of his views may strike some philosophers as overstated, their general spirit has some claim to be seen as today’s conventional wisdom. However, there is a way in which Feyerabend differed from most academic philosophers of science of his generation and which sets him apart even now. This is that he did not recommend proliferation merely as the best way of proceeding within science. On the contrary, he held that modern science is itself just one approach to dealing with the world and should not simply be assumed to be superior to alternative approaches.

By being willing to argue that, say, witchcraft could usefully be studied with the degree of seriousness currently reserved for science, Feyerabend acquired a reputation for irrationalism. This is hardly surprising, for when Feyerabend recommended studying witchcraft, he did not mean studying it in a scientific manner, as a detached observer: he meant immersing oneself in its practices, with a view to becoming a witch. In fact, it is not easy to gauge how seriously Feyerabend himself took the nonscientific outlooks he so often commended. He enjoyed playing the *enfant terrible* and, as he remarked in response to criticism of his supposed beliefs, ’argument is not confession’. But the main point is that what he opposed was not science as such but - our according any outlook, the scientific outlook included - authoritative status. His epistemological aim was to sharpen our critical sense by making us more keenly aware of the ways in which all outlooks have their weaknesses as well as their strengths, their costs as well as their benefits. Not surprisingly, then, he rejected the ’irrationalist’ label. What he objected to, he claimed, was all petrified conceptions of rationality, together with the orthodoxies they support. He can thus usefully be seen as a late representative of the classical sceptical tradition - the urbane and tolerant tradition of Sextus Empiricus and M. de Montaigne - in which the sceptic explores and counterposes all manner of competing ideas without regarding any as definitely established. The sceptic of this stamp need not deny his own *preferences*, including theoretical preferences, but he will not see them as singled out for approval by some extra-human authority: reason, for example. His scepticism thus grows out of the acuteness of his critical faculties. To the sceptic, his dogmatist opponents are the real irrationalists.

Feyerabend’s distrust of authority led him to argue that a truly humane society would make room for the greatest possible variety of what Mill called ’experiments in living’ (see Mill, J.S. §12). But although his philosophy took an increasingly political turn, he never gave serious thought to the question of how this conception of society might be given institutional expression under the conditions of modern life. In consequence, the political

dimension of his thought amounts only to a vague and romantic utopianism.

See also: Incommensurability

MICHAEL WILLIAMS

List of works

Feyerabend, P. (1975) *Against Method*, London: New Left Books.(Feyerabend's best known book, in which he defends the position that the only generally valid methodological maxim for science is 'anything goes'.)

Feyerabend, P. (1978) *Science in a Free Society*, London, New Left Books.(Collection of essays containing responses to several reviews of *Against Method*.)

Feyerabend, P. (1981) *Philosophical Papers*, vols 1 and 2, Cambridge: Cambridge University Press. (Feyerabend's earlier and more standardly 'philosophical' essays on scientific method and the philosophy of physics.)

Feyerabend, P. (1991) *Three Dialogues on Knowledge*, Oxford: Blackwell.(Late defence of epistemological anarchism.)

Feyerabend, P. (1995) *Killing Time*, Chicago, IL: University of Chicago Press.(Feyerabend's autobiography, completed just before his death.)

References and further reading

Anderssen, G. (1994) *Criticism and the History of Science: Kuhn's, Lakatos' and Feyerabend's Criticisms of Critical Rationalism*, Leiden, and New York: Brill. (Critical evaluation of Feyerabend *et al*. from a Popperian standpoint.)

Couvalis, G. (1989) *Feyerabend's Critique of Foundationalism,* Aldershot: Avebury. (Sympathetic account of Feyerabend's views drawing mainly on his earlier (pre-*Against Method*) essays.)

Munevar, G. (ed.) (1991) *Beyond Reason: Essays on the Philosophy of Paul Feyerabend*, Dordrecht, and Boston, MA: Kluwer.(Essays on Feyerabend by various contemporary philosophers of science.)

Stove, D. (1982) *Popper and After: Four Modern Irrationalists*, Oxford, Pergamon Press.(Traces the alleged irrationalism of Feyerabend and others to the mistaken 'deductivism' the author claims to be in Hume's original argument for inductive scepticism.)

Fichte, Johann Gottlieb (1762-1814)

Fichte developed Kant's Critical philosophy into a system of his own, which he named 'Theory of Science' or Wissenschaftslehre. *Though Fichte continued to revise this system until the end of his life, almost all of his best known and most influential philosophical works were written in first portion of his career, when he was a professor at the University of Jena.*

The task of philosophy, as understood by Fichte, is to provide a transcendental explanation of ordinary consciousness and of everyday experience, from the standpoint of which philosophy must therefore abstract. Such an explanation can start either with the concept of free subjectivity ('the I') or with that of pure objectivity (the 'thing in itself'), the former being the principle of idealism and the latter that of what Fichte called 'dogmatism' (or transcendental realism). Though neither of these first principles can be theoretically demonstrated, the principle of freedom possesses the advantage of being practically or morally certain. Moreover, according to Fichte, only transcendental idealism, which begins with the principle of subjective freedom and then proceeds to derive objectivity and limitation as conditions for the possibility of any selfhood whatsoever, can actually accomplish the task of philosophy.

One of the distinctive features of Fichte's Jena system is its thoroughgoing integration of theoretical and practical reason, that is, its demonstration that there can be no (theoretical) cognition without (practical) striving, and vice versa. Another important feature is Fichte's demonstration of the necessary finitude of all actual selfhood. The 'absolute I' with which the system seems to begin turns out to be only a practical ideal of total self-determination, an ideal toward which the finite I continuously strives but can never achieve. Also emphasized in Fichte's Jena writings is the social or intersubjective character of all selfhood: an I is an I only in relationship to other finite rational subjects. This insight provides the basis for Fichte's political philosophy or 'theory of right', which is one of the more original portions of the overall system of the Wissenschaftslehre, *a system that also includes a foundational portion (or 'first philosophy'), a philosophy of nature, an ethics and a philosophy of religion.*

1 Life and works I: Jena

Born in the village of Rammenau in the Oberlausitz area of Saxony, Fichte was the eldest son in a family of poor and pious ribbon weavers. At the age of 9 his extraordinary intellectual talents brought him to the attention of a local baron, who sponsored his education, first at the Pforta school and then at the Universities of Jena and Leipzig. With the death of his patron, Fichte was forced to discontinue his studies and earn his livelihood as a private tutor, a profession the proud young man quickly came to detest.

After several years of such employment, including a long sojourn in Zurich, where he met his future wife, Johanna Rahn, Fichte returned to Leipzig to pursue a literary career. All his projects failed, however, and he was again forced to survive as a tutor. This time he was asked by a university student to give lessons in the Kantian philosophy (see KANT, I.). This encounter with Kant's writings in the summer of 1790 proved to be decisive, and Fichte became an instant convert to a philosophy he described as having 'occasioned a revolution in my way of thinking'. Eventually he made his way to Königsberg, where he lived for a few months. After his first disappointing interview with Kant, he resolved to demonstrate his mastery of the new philosophy by writing a treatise on a theme as yet unaddressed by Kant: namely, the question of the compatibility of the former with any concept of divine revelation. In just a few weeks Fichte composed a remarkable manuscript in which he concluded that the only revelation consistent with the Critical philosophy is the moral law itself. Suitably impressed, Kant helped to arrange the publication of Fichte's manuscript, which was published by Kant's own publisher in 1792 under the title *Versuch eine Kritik aller Offenbarung (Attempt at a Critique of All Revelation)*. For reasons that have never been satisfactorily explained, the first edition appeared without the author's name and preface. Not surprisingly, the book was hailed as a product of Kant's own pen and was widely praised as such. When the true identity of its author was revealed, Fichte was immediately catapulted from total obscurity to philosophical celebrity.

At the time that this stroke of fortune occurred Fichte was again employed as a private tutor, this time in Poland, where he was also working on several political tracts, including the provocatively entitled *Reclamation of the Freedom of Thought from the Princes of Europe, who have Hitherto Suppressed it* (1793). In the summer of 1793,

however, he returned to Zurich where he married his fiancée and oversaw the publication of the first two instalments of his spirited *Contribution to the Rectification of the Public's Judgment of the French Revolution* (1793 and 1794), a work in which he outlined his own democratic view of legitimate state authority and insisted passionately on the right of revolution. Even though both of these political tracts were published anonymously, Fichte was widely known to be their author and thus acquired an early reputation as a political - as well as a religious and philosophical - radical.

In Zurich, Fichte's time was increasingly occupied by his efforts to construct for himself a systematic revision of what he still took to be Kant's Critical philosophy. It was while he was so engaged that he received an invitation to assume the recently vacated chair of Critical Philosophy at the University of Jena, which was rapidly emerging as the capital of the new philosophy and, together with nearby Weimar, as one of the liveliest intellectual centres in the German-speaking world.

Fichte arrived in Jena in May of 1794, and enjoyed tremendous success there for the next six years, during which time he laid the foundations and developed the first systematic articulations of his new system (the *Wissenschaftslehre* or 'Theory of Science'). Even as he was engaged in this tremendous theoretical labour, he also tried to address a larger, popular audience and also threw himself into various practical efforts to reform university life. As one bemused colleague accurately remarked, 'His is a restless spirit; he thirsts for some opportunity to act in the world. Fichte wants to employ his philosophy to guide the spirit of his age'. This passionate striving to address the pressing needs of his own time is plainly evident in the text of the enormously popular public lectures that he began to deliver immediately upon his arrival in Jena and published under the title *Einige Vorlesungen über die Bestimmung des Gelehrten (Some Lectures concerning the Scholar's Vocation)* (1794c).

The first hint of Fichte's new philosophical project and strategy came in his 1794a review of G.E. Schulze's *Aenesidemus*, which was soon followed by a brief but important manifesto for the new system, *Über den Begriff der Wissenschaftslehre (Concerning the Concept of the Wissenschaftslehre)* (1794b). The first full - albeit still provisional - presentation of the first or 'foundational' portion of the new system was presented in the so-called private lectures that Fichte delivered during the 1794-5 academic year, lectures he arranged to have printed in fascicles for the convenience of his students, to whom these printed sheets were distributed in instalments over the course of the year. Eventually, these same pages were bound together and offered to the public under the title *Grundlage der gesamten Wissenschaftslehre Wissenschaftslehre (Foundations of the Entire Wissenschaftslehre)* (Parts I and II, 1794; Part III, 1795), supplemented by the *Grunriß des Eigenthümlichen der Wissenschaftslehre in Rüchsicht auf das theoretishce Vermögen (Outline of the Distinctive Character of the Wissenschaftslehre with Respect to the Theoretical Faculty)* (1795), though each of these publications still bore on its title page the words 'a manuscript for the use of his listeners'.

Dissatisfied with many features of this initial presentation and surprised and shocked by the virtually universal misunderstanding of his published *Foundations*, Fichte immediately set to work on an entirely new exposition of the same, which he repeated three times in his private lectures on 'The Foundations of Transcendental Philosophy (*Wissenschaftslehre*) *nova methodo*' (1796-7, 1797-8, 1798-9). Though he intended to revise these lectures for serial publication under the title *An Attempt at a New Presentation of the Wissenschaftslehre in the Philosophisches Journal einer Gesellschaft Teutscher Gelehrten*, of which he himself was by then co-editor, the only portions of this projected *New Presentation* that ever appeared were the 'First and Second Introductions to the *Wissenschaftslehre*' (1797) and Chapter One of the same (1798).

At the same time that Fichte was revising the presentation of the foundational portion of his system he was also elaborating its various subdivisions or branches - first in private lectures and then in published treatises on *Foundations of Natural Right Based on the Wissenschaftslehre* (1796-7) and *Das System der Sittenlehre nach den Principien der Wissenschaftslehre) (System of Ethical Theory Based on the Wissenschaftslehre)* (1798a). Fichte's intention was then to turn to the development of the remaining subdivision of his system and to expound a philosophy of religion 'in accordance with the principles of the Wissenschaftslehre', but before he had a chance to do this he published a brief essay *Üeber den Grund unsers Glaubens an eine göttliche Weltregierung (On the Basis of Our Belief in a Divine Governance of the World)* (1798b), in which he attempted to sketch some of his preliminary ideas on this topic. This essay, along with another by K.L. Forberg, that was contained in the same issue of the *Philosophisches Journal*, provoked an anonymous author to publish a pamphlet charging Fichte with

atheism, which eventually led to the official suppression of the offending issue of the journal and to public threats by various German princes to prevent their students from enrolling at the University of Jena. The crisis produced by these actions and the growing number of publications for and against Fichte - which included an intemperate *Appeal to the Public* by Fichte himself (1799) - quickly assumed a life and even a name of its own: 'the Atheism Controversy' or *Atheismusstreit*. In the end, Fichte badly miscalculated his own position and was finally forced to resign from Jena and flee to Berlin, where he arrived in the summer of 1799 and where the final phase of his career largely unfolded.

2 Life and works II: Berlin

In Berlin, which was not yet home to a university, Fichte supported himself by giving private lectures on his philosophy and by a new flurry of literary production, increasingly aimed at a large, popular audience. The first of these new writings was a brilliant popular presentation of some of the characteristic features and conclusions of Fichte's system, with a strong emphasis upon the moral and religious character of the same. This work, *Die Bestimmung des Menschen (The Vocation of Man)* (1800b), which is perhaps Fichte's greatest literary achievements, was published in 1800, the same year that he also published a typically bold foray into political economy, *The Closed Commercial State*, in which he propounds a curious blend of socialist political ideas and protectionist economic principles. Defending his own philosophy against misunderstanding remained a primary concern for Fichte, however, as is attested by the publication in 1801 of the poignantly titled *Sonnenklarer Bericht an das größere Publikum über das eigentliche Wesen der neuesten Philosophie. Ein Versuch, die Leser zum verstehen zu zwingen (Sun-Clear Report to the Public at Large on the Actual Character of the latest Philosophy: An Attempt to Force the Reader to Understand).*

At the same time that he was addressing the public, Fichte was deeply engrossed in yet another effort to rethink and to restructure the foundations of his system. Hardly a year went by that he did not produce, either for his own use, or else for use in conjunction with the private lectures he was delivering in Berlin, yet another completely new version of the *Wissenschaftslehre*, each of which differed more and more dramatically from the familiar published version of 1794-5. (By 1804, for example, most references to 'the I' had been replaced with references to 'the absolute' and its various 'appearances'.) Having learned a lesson from the public reception of his 1794-5 presentation, Fichte resolved never to publish any of these new versions of his system (of which more than a dozen survive in manuscript), explaining in an official pro memoria of 3 January 1804 that the author of the *Wissenschaftslehre* 'wishes to confine himself to oral communication, so that misunderstanding can thereby be detected and eliminated on the spot'. Consequently, the sole public hint of Fichte's new understanding of his system was a brief and enigmatic *Die Wissenschaftslehre, in ihrem allgemeinen Umrisse dargestellt (General Outline of the Wissenschaftslehre)* published in 1810, as well as whatever hints in this direction readers might glean from the popular works published during this period.

In 1805 Fichte spent a semester lecturing at the University of Erlangen, but returned in the autumn to Berlin, where, in 1806, he published in rapid succession three popular and well-received books, all of which were based upon earlier lectures: *On the Essence of the Scholar* (a reworking of some of the same themes first addressed in the similarly titled lectures of 1794); *Der Grundzüge des gegewärtigen Zeitalters (Characteristics of the Present Age)* (1806a) (an attempt to show the implications of his 'system of freedom' for a speculative philosophy of history); and *Guide to the Blessed Life* (an eloquent and rather mystical treatise on philosophy and religion). With the entry of the French army into Berlin in 1806, Fichte joined the Prussian government in exile in Königsberg, where he held yet another course of lectures on the *Wissenschaftslehre* and wrote an important short work on *Machiavelli as Author* (1807), in which he defends a form of Realpolitik that appears to contrast starkly with the liberalism and political idealism of his earlier writings. Eventually, however, he returned to Berlin, where, under the eyes of the occupying forces, he delivered his celebrated *Reden an die deutsche Nation (Addresses to the German Nation)* (1808). Though these lectures later obtained a place of dubious honour as founding documents in the history of German nationalism, they are mainly concerned with the issue of national identity - and hence with the question of national education (which is the main topic of the work) - as a means to a larger, cosmopolitan end.

Fichte had always had a lively interest in pedagogical issues and hence took a leading role in planning the new Prussian university to be established in Berlin. Appropriately, when the new university finally opened in 1810, Fichte himself served as the first head of the philosophical faculty as well as the first elected rector of the entire

university. His final years saw no diminution in the pace either of his public activity or of his philosophical efforts. He continued to produce new lectures on the foundations and first principles of his system, as well as new introductory lectures on philosophy in general (‘Logic and Philosophy’ (1812) and ‘The Facts of Consciousness’ (1813)), political philosophy (‘System of the Theory of Right’ (1812) and ‘Theory of the State’ (1813)) and ethics (‘System of Ethical Theory’ (1812)). As presaged by the book on Machiavelli, these late forays into the domain of practical philosophy paint a far darker picture of human nature and defend a much more authoritarian view of the state than anything to be found in the earlier, published writings on these subject.

In 1813 Fichte cancelled his lectures so that his students could enlist in the ‘War of Liberation’ against Napoleon, of which Fichte himself proved to be an indirect casualty. Through his wife, who was serving as a volunteer nurse in a Berlin military hospital, he contracted an infection of which he died on 29 January 1814.

3 The project of a *Wissenschaftslehre*

The primary task of Fichte’s system of philosophy is to reconcile freedom with necessity, or, more specifically, to explain how freely willing, morally responsible agents can at the same time be considered part of a world of causally conditioned material objects in space and time. Fichte’s strategy for answering this question was to begin simply with the ungrounded assertion of the subjective spontaneity and freedom (infinity) of the I and then to proceed to a transcendental derivation of objective necessity and limitation (finitude) as a condition necessary for the possibility of the former. Hence, in his ‘First Introduction to the *Wissenschaftslehre*’ (1797), he describes philosophy’s task as that of ‘displaying the foundation of experience’, that is, ‘explaining the basis of the system of representations accompanied by a feeling of necessity’. Fichte owed this conception of the task and strategy of philosophy entirely to his reading of Kant, and no matter how far his own presentation of his system seemed to diverge from ‘the letter’ of Kant’s own presentation, he always maintained that it remained true to ‘the spirit’ of the same - which, for Fichte, lay entirely in its uncompromising insistence upon the practical indubitability of human freedom and its thoroughgoing dedication to the project of giving a transcendental account of ordinary experience that could explain the objectivity and necessity of theoretical reason (cognition) in a manner consistent with the affirmation of human liberty. For this reason Fichte sometimes referred to his own presentation of this same philosophy as ‘the system of human freedom’.

In accordance with this conception of philosophy’s task, Fichte insisted firmly upon the sharp distinction between the ‘standpoint’ of natural consciousness (which it is the task of philosophy to somehow ‘explain’) and that of transcendental reflection (from which alone such a philosophical explanation can be conducted - if at all). Thus there is no possible conflict between transcendental idealism and everyday realism; on the contrary, the whole point of the former is to demonstrate the necessity and unavoidability of the latter.

However ‘Kantian’ in spirit Fichte’s enterprise might have been, he was at the same time all too keenly aware of certain glaring weaknesses and inadequacies in Kant’s own execution of this project. Taking to heart the criticisms of such contemporaries as F. JACOBI, Salomon Maimon and G.E. Schulze, Fichte realized, first, that the doctrine of the thing in itself, understood as an external cause of sensations, was indefensible on Critical grounds, and second, that Kant’s denial of the possibility of ‘intellectual intuition’, though certainly justified as a denial of the possibility of any non-sensory awareness of objects, was nevertheless difficult to reconcile with certain other Kantian doctrines regarding the subject’s presence to itself both as a (theoretically) cognizing subject (the doctrine of the transcendental apperception) and as a (practically) striving moral agent (the doctrine of the categorical imperative).

Fichte was also persuaded by his reading of the works of K. L. REINHOLD that the systematic unity of the Critical philosophy - specifically, the unity of theoretical and practical reason, of the First and Second *Critiques* - was insufficiently apparent in Kant’s own presentation of his philosophy. His study of Reinhold also convinced him that the most promising way to display the unity in question would be to provide both theoretical and practical philosophy with a common foundation. The task, then, was to discover a single, self-evident starting point or first principle from which one could then somehow deduce or ‘derive’ both theoretical and practical philosophy. Not only would such a method guarantee the systematic unity of philosophy itself, but, more importantly, it would also display what Kant hinted at but never demonstrated: namely, the underlying unity of reason itself.

Since it is a central task of philosophy, so construed, to establish the very possibility of any knowledge or science

(*Wissenschaft*) whatsoever, Fichte proposed to replace the disputed term 'philosophy' (or 'love of wisdom') with the new term *Wissenschaftslehre* or 'Theory of Science' - a name that clearly designates the distinctively 'second order' character of philosophical reflection. Though Fichte's proposal never caught on as a general name for what was once called 'philosophy', it did become the universally acknowledged name for his own distinctive version of transcendental idealism. It is, however, important to note that *Wissenschaftslehre* is not the name of any particular Fichtean treatise, but is instead the general name for his entire system or project - a larger system that consists of a number of interrelated parts or systematic subdisciplines and an overarching project that could be expounded in a series of radically different presentations and a bewildering variety of systematic vocabularies. In fact, Fichte spent his entire life demonstrating this last point: that is, expounding and re-expounding what he himself took to be the same philosophy ('the *Wissenschaftslehre*') in a number of radically different forms and guises (*The Foundations of the Entire Wissenschaftslehre* of 1794-5, 'Wissenschaftslehre nova methodo', 'the *Wissenschaftslehre* of 1801-2', 'the 1804 *Wissenschaftslehre*', 'the *Wissenschaftslehre* of 1805', and so on) - more than a dozen in all between 1794 and 1814.

4 Idealism versus dogmatism

In order to construct any genuine philosophy of freedom, thought Fichte, the reality of freedom itself must simply be presupposed and thus treated as an incontrovertible 'fact of reason' in the Kantian sense. This, of course, is not to deny the possibility of raising theoretical objections to such claims; on the contrary, it was the very impossibility of any theoretically satisfactory refutation of scepticism concerning the reality of freedom that led Fichte to the recognition of the inescapable 'primacy of the practical' with respect to the selection of any philosophical starting point.

In so far as any proposed first principle of philosophy is really supposed to be the 'first principle' of all knowledge and hence of all argument, it clearly cannot be derived from any higher principle and hence cannot be established by any sort of argument. Furthermore, Fichte maintained that there are two and only two possible starting points for the philosophical 'explanation' of experience sketched above: namely, the concept of pure selfhood (freedom) and that of pure thinghood (necessity) - neither of which can be warranted, *qua* philosophical starting point, by a direct appeal to experience, and both of which can be arrived at only by a selfconscious act of philosophical abstraction from ordinary experience (within which freedom and necessity, subject and object, are invariably joined).

The two rival philosophical strategies made possible by these opposed starting points are unforgettably limned by Fichte his two 1797 *Introductions to the Wissenschaftslehre*, where he characterizes the sort of philosophy that begins with the pure I as 'idealism' and that which begins with the thing in itself as 'dogmatism'. Since, according to Fichte's earlier argument in *Concerning the Concept of the Wissenschaftslehre*, a unified system of philosophy can have one and only one first principle, and since there are two and only two possible first principles, then it follows that no 'mixed' system of idealism/dogmatism is possible. Moreover, since dogmatism, as understood by Fichte, unavoidably implies a strict form of determinism or 'fatalism', whereas idealism is, from the start, committed to the reality of human freedom, it is also practically impossible to reach any sort of 'compromise' between two such radically opposed systems.

Though Fichte conceded that neither dogmatism nor idealism could directly refute its opposite and also recognized that the choice between philosophical starting points could not be resolved philosophically, he nevertheless denied that any dogmatic system - any system that started with the concept of sheer objectivity - could ever succeed in accomplishing what was required of all philosophy. Dogmatism, he argued, could never provide a transcendental deduction of ordinary consciousness, for, in order to accomplish this, it would have to make an illicit leap from the realm of things to the realm of mental ideas or 'representations' (*Vorstellungen*). Idealism, in contrast, at least when correctly understood as the sort of Critical idealism that recognizes that the intellect itself most operate in accordance with certain necessary laws, could - at least in principle - accomplish the prescribed task and explain our experience of objects ('representations accompanied by a feeling of necessity') in terms of the necessary operations of the intellect itself, without having to make an illicit appeal to things in themselves. To be sure, one cannot decide in advance whether or not any such deduction of experience from the mere concept of free selfconsciousness is *in fact* possible. This is something that can be decided only after the construction of the system in question. Until then, it remains a mere hypothesis that the principle of human freedom, for all of its

practical certainty, is also the proper starting point for a transcendental account of objective experience.

Returning to the question of the starting point, however, it still must be granted that the truth of the latter cannot be established by any philosophical means, including its utility as a philosophical first principle. On the contrary, one must be convinced, on wholly extraphilosophical grounds, of the reality of one's own freedom before one can enter into the chain of deductions and arguments that constitute the system of transcendental idealism. This is the meaning of Fichte's oft-cited assertion that 'the kind of philosophy one chooses depends upon the kind of person one is'.

In the end, the only reason why transcendental idealists come to a stop with - and thus begins their philosophy with - the proposition 'the I freely posits itself' is not because they are unable to entertain theoretical doubts on this point nor because they cannot continue the process of reflective abstraction. Instead, as Fichte explains in his essay *On the Basis of Our Belief in a Divine Governance of the World*, 'I cannot go beyond this standpoint because I am not permitted to do so'. Of course I could think of myself as determined by things, but still 'I ought to begin my thinking with the thought of the pure I, and I ought to think of this pure I as acting with absolute spontaneity - not as determined by things, but rather, as determining them'. It is precisely because the categorical imperative is in this way invoked to secure the first principle of his entire system that Fichte can make the rather startling claim that the '*Wissenschaftslehre* is the only kind of philosophical thinking that accords with duty'.

5 Foundations of the Jena *Wissenschaftslehre*

The published presentation of the first principles of the Jena *Wissenschaftslehre* commences with the proposition 'the I posits itself'; more specifically, 'the I posits itself as an I - that is, as positing itself'. To be sure, this starting point is somewhat obscured in the *Foundations of the Entire Wissenschaftslehre* by a difficult and somewhat forced attempt to connect this starting point to the logical law of identity, as well as by the introduction of two additional 'first principles', corresponding to the logical laws of non-contradiction and sufficient reason. (Significantly, this distraction is eliminated entirely in the 1796-9 lectures on *Wissenschaftslehre nova methodo*.)

'To posit' (*setzen*) means simply 'to be aware of', 'to reflect upon', or 'to be conscious of', and thus the principle in question simply states that the essence of selfhood lies in the assertion of ones own self-identity. Such immediate self-identify, however, cannot be understood as a 'fact', no matter how privileged, nor as an 'accident' of some previously existing substance or being. Instead, it must be understood as an *activity*, albeit an activity of a most extraordinary, auto-productive type: in Fichte's own language, it is a *Tathandlung* or 'fact/act', a unity that is presupposed by and contained within every fact and every act of empirical consciousness, though it never appears as such therein.

This same 'identity in difference' of self-consciousness might also be described as an 'intellectual intuition', since it involves the immediate presence of the I to itself, prior to and independently of any sensory content. Understood in this way, this original intellectual intuition is merely the reflectively realized presupposition for the possibility of actual consciousness and does not designate any act or state of actual, empirical consciousness - including the consciousness of the philosopher. Fichte himself, however, confuses matters by sometimes using the term 'intellectual intuition' to designate the act of philosophical reflection through which the philosopher arrives at the former concept or, on still other occasions, to indicate our direct, practical awareness within everyday life of our moral obligations (categorical imperative *qua* 'real intellectual intuition'). Given the subsequent abuse of this term by SCHELLING and the romantics, it is important to recognize its systematic ambiguity in Fichte's own writings. Significantly, the term 'intellectual intuition' does not even occur in the 1794-5 *Foundations*.

A fundamental corollary of Fichte's understanding of selfhood as a kind of act is his denial that the I is originally any sort of 'substance'. Instead, the I is simply what it posits itself to be, and thus its 'being' is, so to speak, a consequence of its self-positing. The first principle of the Jena *Wissenschaftslehre* is thus equally 'practical' and 'theoretical', insofar as the act described by this principle is a 'doing' as well as a 'knowing', a deed as well as a cognition. Thus the problematic unity of theoretical and practical reason is guaranteed from the start, inasmuch as this very unity is a condition for the possibility of selfconsciousness.

After establishing the first principle and conceiving the act expressed therein, the philosophical task is then to discover what other acts must occur as conditions for the possibility of the original, 'simply posited', first act and then to do the same for each of these successively discovered acts (or the theorems in which they are formulated).

By continuing in this manner, one will finally arrive at a complete deduction of the a priori structure of ordinary experience - or, what amount to the same thing, a complete inventory of the ‘original acts of the mind’. This is precisely the task of the first or ‘foundational’ portion of the Jena system.

Just as we are never directly aware of the original act of self-positing with which the system commences, so we are also unaware - except, of course, from the artificial standpoint of philosophical reflection - of each of these additional ‘necessary but unconscious’, acts that are derived as conditions necessary for the possibility of this first act. Furthermore, though we must, due to the discursive character of reflection itself, distinguish each of these acts from the others that it is conditioned by and that are, in turn, conditioned by it, none of these individual acts can actually occur in isolation from all of the others. Thus what transcendental philosophy actually does is to *analyse* what is in fact the single, *synthetic* act through which the I posits for itself both itself and its world, thereby becoming aware in a single moment of both its freedom and its limitations, its infinity and its finitude. The result of such an analysis is a clear realization on the part of the transcendental philosopher that, although ‘the I simply posits itself’, its freedom is never ‘absolute’ or ‘unlimited’; instead, freedom proves to be conceivable - and hence the I itself proves to be ‘possible’ - only as limited and finite. Despite widespread misunderstanding of this point, the *Wissenschaftslehre* is not a philosophy of the absolute I. Instead, the conclusion of both the *Foundations of the Entire Wissenschaftslehre* and of the *Wissenschaftslehre nova methodo* is that the ‘absolute I’ is a mere abstraction and that the only sort of I that can actually exist or act is a finite, empirical, embodied, individual self.

An I must posit itself in order to be an I at all; but it can posit itself only insofar as it posits itself as limited (and hence divided against itself). Moreover, it cannot even posit for itself its own limitations, in the sense of producing or creating these limits. Instead, according to Fichte’s analysis, if the I is to posit itself at all, it must simply discover itself to be limited, a discovery that Fichte characterizes as a ‘check’ or *Anstoß* to the free, practical activity of the I. Such an original limitation of the I itself is first posited as a mere ‘feeling’, then as a ‘sensation’, then as an ‘intuition’ of a thing, and finally as a ‘concept’. The *Anstoß* thus provides the essential occasion that first sets in motion the entire complex train of activities that finally result in our conscious experience both of ourselves as empirical individuals and of a world of spatio-temporal material objects.

Though this doctrine of the *Anstoß* may seem to play a role in Fichte’s philosophy not unlike that which has sometimes been assigned to the thing in itself in the Kantian system, the fundamental difference is this: the *Anstoß* is not something foreign to the I. Instead, this term signifies the I’s original encounter with its own finitude. Rather than claim that the Not-I is the cause or ground of the *Anstoß*, Fichte argues that the former is posited by the I in order to explain to itself the latter. Though the *Wissenschaftslehre* demonstrates that such an *Anstoß* must occur if selfconsciousness is to be actual, philosophy itself is quite unable to ‘deduce’ or to ‘explain’ the actual occurrence of such an *Anstoß* - except as a condition for the possibility of consciousness. Accordingly, there are strict limits to what can be expected from any a priori deduction of experience. Though transcendental philosophy can explain, for example, why the world has a spatio-temporal character and a causal structure, it can never explain why objects have the particular sensible properties they happen to have. This is something that the I simply has to ‘discover’ at the same time that it discovers its own freedom, and indeed, as a condition for the latter.

But there remains much that can be demonstrated within the foundational portion of the *Wissenschaftslehre*. For example, it can be shown that the I could not become conscious of its own limits in the manner required for the possibility of any self-consciousness unless it also possessed an original and spontaneous ability to synthesize the finite and the infinite. In this sense, the *Wissenschaftslehre* deduces the power of productive imagination as an original power of the mind. Similarly, it can be shown that the I could not be ‘checked’ in the manner required for the possibility of consciousness unless it possessed, in addition to its original ‘theoretical’ power of productive imagination, an equally original ‘practical’ power of sheer willing, which, once ‘checked’, is immediately converted into a capacity for endless striving. The *Wissenschaftslehre* thus also includes a deduction of the categorical imperative and of the practical power of the self. For Fichte, therefore the Kantian principle of ‘the primacy of practical reason’ means not simply that philosophy must recognize a certain autonomous sphere within which practical reason is efficacious and practical considerations are appropriate; instead, it implies something much stronger: namely, the recognition that ‘the practical power is the innermost root of the I’ and thus that ‘our freedom itself is a theoretical determining principle of our world’. The *Wissenschaftslehre* demonstrates from the very start that reason could not be theoretical if it were not also practical - at the same time, to be sure, that it also demonstrates the converse of this proposition, that reason could not be practical if were not also theoretical.

Freedom, according to Fichte's argument, is possible and actual only within the context of natural necessity, and thus it is never 'absolute', but always limited and finite. On the other hand, just as surely as a free subject must posit its freedom 'absolutely' - that is to say, 'purely and simply' (*schlechthin*) or 'for no reason' - so must it never identify itself with any determinate or limited state whatsoever. Instead, a finite free self must constantly strive to transform both the natural and the human worlds in accordance with its own freely-posited goals. The sheer unity of the self posited in the starting point of the *Foundations* is thereby transformed into an idea of reason in the Kantian sense: the actual I is always finite and divided against itself, and hence always striving for a sheer self-determinacy that it never achieves. Between the original abstraction of pure selfhood as pure *Tathandlung* and the concluding ideal of a self that is only what it determines itself to be, in which 'is' and 'ought' wholly coincide, lies the entire realm of actual consciousness and experience.

6 Philosophy of nature and ethics

Having established the foundations of his new system, Fichte then turned to the task of erecting a fully-articulated transcendental system according to a plan that is perhaps most clearly outlined in 'The Deduction of the Subdivisions of the *Wissenschaftslehre*', with which the lectures on '*Wissenschaftslehre nova methodo*' conclude. At least as conceived during the Jena years, the entire *Wissenschaftslehre* consists of four parts: (1) first philosophy, which is to say, the 'foundational' portion of the system; (2) 'theoretical philosophy' or 'philosophy of nature', the latter understood in a sense corresponding to Kant's *Metaphysical First Principles of Nature*; (3) 'practical philosophy' or ethics; and (4) 'philosophy of the postulates', which includes the subdisciplines of political philosophy or 'theory of right' and philosophy of religion.

The closest Fichte himself ever came to developing a philosophy of nature according to transcendental principles is the compressed account of the same presented in the *Outline of the Distinctive Character of the Wissenschaftslehre with Respect to the Theoretical Faculty* (1795), in which, however, the domain of specifically 'theoretical' philosophy is not rigorously distinguished from the 'theoretical' portion of the *Foundations*, which also includes a discussion of the a priori structure of the objective world. In fact, the kind of theoretical philosophy of nature made possible by the *Wissenschaftslehre* turns out to be even more modest than Kant's and more closely resembles what later came to be called the philosophy of (natural) science than it does the speculative *Naturphilosophie* of Schelling and HEGEL.

In contrast to Fichte's rather cursory treatment of purely theoretical philosophy, ethics or 'practical philosophy', which analyses the determinate ways in which willing and acting are determinable by principles of pure reason, constitutes a major portion of the Jena system, and the *System of Ethical Theory* (1798a) is Fichte's longest single book. Whereas theoretical philosophy 'explains how the world is, and the result is the same as pure experience', practical philosophy 'explains how the world ought to be constructed by rational beings, and its result is something ideal'. Ethics thus considers the object of consciousness not as something given or even constructed by necessary laws of consciousness, but rather as something to be produced by a freely acting subject, consciously striving to establish and to accomplish its own goals. The specific task of Fichte's ethics is therefore to deduce from the general obligation to determine oneself freely the particular obligations of every finite rational being.

Viewed from the perspective of practical philosophy, the world really is nothing more than what Fichte once described as 'the material of our duty made sensible', which is precisely the viewpoint adopted by the morally engaged, practically striving subject. On the other hand, this is not the only way the world can be viewed, and, more specifically, it is not the only way in which it is construed by transcendental philosophy. For this reason it is somewhat misleading to characterize the *Wissenschaftslehre* as a whole as a system of 'ethical idealism'. As noted above, Fichte certainly does succeed in constructing an account of consciousness that fully integrates the imperatives and activities of practical reason into the very structure of the latter, but this integration is always balanced by a recognition, first, of the constitutive role of theoretical reason, and secondly, of the sheer 'giveness' of the I's original determinacy (doctrine of the *Anstoß*).

7 Political philosophy and philosophy of religion

The final portion of the Jena system is devoted to 'the philosophy of the postulates', a discipline conceived as occupying the middle ground between purely theoretical and purely practical philosophy. In this portion of the system the world is considered neither as it simply is nor as it simply ought to be; instead, the moral world is itself

considered from the perspective of the natural world (that is, one considers the postulates that theoretical reason addresses to the practical realm) or else, alternatively, the natural world is considered from the perspective of the moral world order (that is, one considers the postulates that practical reason addresses to the realm of theory). The first of these perspectives is that of political philosophy, or what Fichte calls ‘theory of right’ or sometimes the domain of ‘natural right’; the latter is that of the philosophy of religion.

A transcendental theory of right proceeds from the general principle that one must limit one’s own freedom in accordance with an a priori concept of the other free beings with whom one comes into contact, and goes on to consider the precise conditions under which such a postulated society of free and equal individuals is in fact possible. Ultimately, this leads Fichte to the formulation of his own version of the contract theory of political legitimacy.

If one is to posit one’s own freedom, then one must posit the freedom of others and limit one’s own freedom accordingly. It follows that a just political order is a demand of reason itself, since ‘the concept of justice or right is a condition of selfconsciousness’. This passage, from the *Foundations of Natural Right*, points to what is perhaps the most original feature of Fichte’s political philosophy: its attempt to demonstrate the intrinsically social character of reason itself and thus to deduce the concept of right from the ‘reciprocal concept’ of an individual free being. The strategy of Fichte’s deduction of intersubjectivity is to establish, first, that an I can posit itself only as an individual and, second, that it can posit itself as an individual only insofar as it recognizes other free individuals - a recognition facilitated by the presence, within selfconsciousness itself, of an immediate awareness of oneself as ‘summoned’ by the other to limit oneself freely in recognition of the latter’s freedom. This a priori deduction of intersubjectivity is so central to the conception of selfhood developed in the Jena *Wissenschaftslehre* that Fichte, in his lectures on *Wissenschaftslehre nova methodo*, incorporated it into his revised presentation of the very foundations of his system.

The theory of natural right developed by Fichte during his Jena period is clearly quite different from the kind of theories that have traditionally gone by this name. The task of Fichte’s theory of right is to consider the specific ways in which the freedom of each individual must be restricted so that several individuals can live together with the maximum amount of mutual freedom. Furthermore, it derives its a priori concepts of the laws of social interaction entirely from the sheer concept of an individual I, as conditions for the possibility of the latter. The concept of right obtains its binding force not from the ethical law, but rather from the general laws of thinking and from enlightened self-interest, and the force of such considerations is hypothetical rather than categorical. The theory of right examines how the freedom of each must be externally limited if a free society of equals is to be possible.

Unlike Kant, Fichte does not treat political philosophy merely as a subdivision of moral theory, but as an independent philosophical discipline with a topic and laws of its own. Whereas ethics analyses the concept of what is *demanded* of a freely willing subject, the theory of right describes what such a subject is *permitted* to do; and whereas ethics is concerned with the inner world of conscience, the theory of right is concerned only with the external, public realm, though only insofar as the latter can be viewed as an embodiment of freedom. If the *Wissenschaftslehre* as a whole can be described, in Fichte’s words, as a ‘system of freedom’, then the spirit of such a system is perhaps best reflected in the distinctively liberal political philosophy based thereupon.

In addition to the postulates addressed by theoretical to practical reason, there are also those addressed by practical reason to nature itself. The latter is the domain of the transcendental philosophy of religion, which is concerned solely with the question of the extent to which the realm of nature can be said to accommodate itself to the aims of morality. The questions dealt with within such a philosophy of religion are those concerning the nature, limits, and legitimacy of our belief in divine providence.

As it happened, Fichte never had a chance to develop this final subdivision of his Jena system and, somewhat ironically, he was prevented from doing so by a controversy that erupted in the wake of the publication, in 1798, of a tentative foray into this very domain in his essay *On the Basis of Our Belief in a Divine Governance of the World*. In this essay Fichte seems to contend that, so far as philosophy is concerned, the realm of the divine is wholly coextensive with that of the moral law itself and that no further inference to a ‘moral lawgiver’ is theoretically or practically required or warranted. In this same essay Fichte also sought to draw a sharp distinction between religion and philosophy (a distinction parallel to the crucial distinction between the ‘ordinary’ and

'transcendental' standpoints) and to defend philosophy's right to postulate something like a 'moral world order'. Philosophy of religion would thus include a deduction of the postulate that our moral actions really do make some difference in the world. The argument of the essay *On the Basis of Our Belief* is, nevertheless, primarily a negative one, designed to deny that any postulate of the existence of a God independent of the moral law is justifiable on philosophical grounds. In the wake of the atheism controversy, Fichte returned to this subject and, in his *From a Private Letter* (1800) and in Part III of *The Vocation of Man* (1800b), attempted to restate his position in a manner that at least appeared to be more compatible with the claims of theism.

See also: Enlightenment, Continental; Idealism; KANTIAN ETHICS

DANIEL BREAZEALE

List of works

Fichte, I. H. (ed.) (1834-5) *Johann Gottlieb Fichtes nachgelassene Werke*, Bonn: Adolph-Marcus, 3 vols.

Fichte, I. H. (1845-6) *Johann Gottlieb Fichtes sämmtliche Werke*, Berlin: Veit, 8 vols.(Taken together, these eleven volumes, edited by Fichte's son, constitute the first attempt at a complete edition of his works and are still widely cited and reprinted.)

Fichte, J.G. (1792) *Versuch einer Kritik aller Offenbarung (Attempt at a Critique of All Revelation)*, trans. G. Green, New York: Cambridge University Press, 1978.

Fichte, J.G. (1794a) '[Rezension:] *Aenesidemus*' (Aenesidemus Review), in D. Breazeale (trans. and ed.) *Early Philosophical Writings*, Ithaca, NY: Cornell University Press, 1988.

Fichte, J.G. (1794b) 'Ueber den Begriff der Wissenschaftslehre' (Concerning the Concept of the Wissenschaftslehre), in D. Breazeale (trans. and ed.) *Early Philosophical Writings*, Ithaca, NY: Cornell University Press, 1988.

Fichte, J.G. (1794c) 'Einige Vorlesungen über die Bestimmung des Gelehrten' (Some Lectures Concerning the Scholar's Vocation), in D. Breazeale (trans. and ed.) *Early Philosophical Writings*, Ithaca, NY: Cornell University Press, 1988.

Fichte, J.G. (1794/5) 'Grundlage der gesamten Wissenschaftslehre' (Foundations of the Entire Science of Knowledge), in P. Heath and J. Lachs (eds) *Science of Knowledge (Wissenschaftslehre)*, trans. P. Heath, New York: Appleton-Century-Crofts, 1970; 2nd edn., Cambridge: Cambridge University Press, 1982.(An adequate translation of Fichte's most famous work, despite the retention of the traditional and misleading title 'Science of Knowledge'.)

Fichte, J.G. (1795) 'Grundriß des Eigenthümlichen der Wissenschaftslehre in Rücksicht auf das theoretische Vermögen' (Outline of the Distinctive Character of the Wissenschaftslehre with respect to the Theoretical Faculty), in D. Breazeale (trans. and ed.) *Early Philosophical Writings*, Ithaca, NY: Cornell University Press, 1988.

Fichte, J.G. (1796-99) 'Wissenschaftslehre nova methodo (student lecture transcripts)', (Foundations of Transcendental Philosophy (Wissenschaftslehre) nova methodo), trans. and ed. D. Breazeale, Ithaca, NY: Cornell University Press, 1992.

Fichte, J.G. (1796/7) *Grundlage des Naturrechts nach Principien der Wissenschaftslehre (The Science of Rights)*, trans. A.E. Kroeger, Philadelphia, PA: Lippincott, 1869.(Very unreliable translation.)

Fichte, J.G. (1797/8) 'Versuch einer neuen Darstellung der Wissenschaftslehre ('Erste' und 'Zweite Einleitung,' 1797; Erste Capitel,' 1798)' (Attempt at a New Presentation of the Wissenschaftslehre), in D. Breazeale (trans. and ed.) *Introductions to the Wissenschaftslehre and Other Writings (1797 - 1800)*, Indianapolis, IN: Hackett, 1994.

Fichte, J.G. (1798a) *Das System der Sittenlehre nach den Principien der Wissenschaftslehre (The Science of Ethics as Based on the Science of Knowledge)*, trans. A.E. Kroeger, London: Kegan Paul, 1897.(Very unreliable translation.)

Fichte, J.G. (1798b) 'Über den Grund unsers Glaubens an eine göttliche Weltregierung' (On the Basis of our Belief in a Divine Governance of the World), in D. Breazeale (trans. and ed.) *Introductions to the Wissenschaftslehre and Other Writings (1797 - 1800)*, Indianapolis, IN: Hackett, 1994.

Fichte, J.G. (1800a) 'Aus einem Privatschreiben' (From a Private Letter), in D. Breazeale (trans. and ed.) *Introductions to the Wissenschaftslehre and Other Writings (1797 - 1800)*, Indianapolis, IN: Hackett, 1994.

Fichte, J.G. (1800b) *Die Bestimmung des Menschen (The Vocation of Man)*, trans. P. Preuss, Indianapolis, IN: Hackett, 1987.

Fichte, J.G. (1801) 'Sonnenklarer Bericht an das größere Publikum über das eigentliche Wesen der neuesten Philosophie. Ein Versuch, die Leser zum Verstehen zu zwingen' (A Crystal Clear Report to the General Public

Concerning the Actual Essence of the Newest Philosophy: An Attempt to Force the Reader to Understand), in E. Behler (ed.) *Philosophy of German Idealism*, trans. J. Botterman and W. Rash, New York: Continuum, 1987.
Fichte, J.G. (1806a) 'Der Grundzüge des gegewärtigen Zeitalters' (The Characteristics of the Present Age), in *The Popular Works of Johann Gottlieb Fichte*, trans. W. Smith, London: Chapman, 2 vols, 1848/9; 4th edn, 1889.
Fichte, J.G. (1806b) 'Über des Wesen des Gelehrten, und seine Erscheinungen im Gebiete der Freiheit' (On The Nature of the Scholar and Its Manifestations), in *The Popular Works of Johann Gottlieb Fichte*, trans. W. Smith, London: Chapman, 2 vols, 1848/9; 4th edn, 1889.
Fichte, J.G. (1806c) 'Die Anweisung zum seeligen Leben, oder auch die Religionslehre' (The Way Towards the Blessed Life; or, the Doctrine of Religion), in *The Popular Works of Johann Gottlieb Fichte*, trans. W. Smith, London: Chapman, 2 vols, 1848/9; 4th edn, 1889.
Fichte, J.G. (1808) *Reden an die deutsche Nation (Addresses to the German Nation)*, trans. R. F. Jones and G. H. Turnbull, eds G. Armstrong Kelly, New York: Harper & Row, 1968.
Fichte, J.G. (1810) 'Die Wissenschaftslehre, in ihrem allgemeinen Umrisse dargestellt' (The Science of Knowledge in its General Outline), trans. W.E. Wright, *Idealistic Studies* 6: 106-17, 1976.
Lauth, R., Jacobs, H. and Gliwitzsky, H. (eds) (1964 -) *J.G. Fichte: Gesamtausgabe der Bayerischen Akademie der Wissenschaften*, Stuttgart-Bad Cannstatt: Frommann, 25 vols to date.(Organized into four separate series - writings published by Fichte, unpublished writings, correspondence and student lecture transcripts - this monumental critical edition, though still incomplete, supersedes all earlier editions.)

References and further reading

The journal *Fichte-Studien* (1990- Amsterdam and Atlanta, GA: Editions Rodopi) appears roughly once a year and publishes papers, most of them in German, on every aspect of Fichte's life and thought.

Adamson, R. (1881) *Fichte*, Edinburgh: Blackwood.(Though seriously out of date, this remains the only full-scale treatment of Fichte in English.)
Baumanns, P. (1974) *Fichtes Wissenschaftslehre. Probleme ihres Anfangs*, Bonn: Bouvier.(Useful exposition of various ways of interpreting the starting point of the first *Wissenschaftslehre*.)
Baumanns, P. (1990) *J.G. Fichte: Kritische Gesamtdarstellung seiner Philosophie*, Freiburg: Alber.(A useful and critical overview of Fichte's philosophical development.)
Baumgartner, M. and Jacobs, W.G. (1968) *J. G. Fichte: Bibliographie*, Stuttgart-Bad Cannstatt: Frommann. (A complete bibliography. Supplemented by Doyé's bibliography.)
Breazeale, D. and Rockmore, T. (eds) (1993) *Fichte: Historical Context/Contemporary Controversies*, Atlantic Highlands, NJ: Humanities Press.(A collection of essays on various aspects of Fichte's philosophy. Includes a complete bibliography of works in English by and about Fichte.)
Doyé, S. (ed.) (1993) *J. G. Fichte-Bibliographie (1969-1991)*, Amsterdam and Atlanta, GA: Editions Rodopi.(An essential addition to the bibliography by Baumgartner and Jacobs.)
Gueroult, M. (1930) *L'evolution et la structure de la doctrine de la science chez Fichte*, Paris: Société de l'édition, 2 vols.(A pioneering developmental study of the *Wissenschaftslehre*.)
Everett, C.C. (1884) *Fichte's Science of Knowledge: A Critical Exposition*, Chicago, IL: Griggs.(The only monograph in English devoted entirely to the Foundations of the Entire *Wissenschaftslehre*.)
'Fichte and Contemporary Philosophy' (1988) special issue of *Philosophical Forum* 19 (2 and 3).
Fuchs, E. (ed.) (1978-92) *J.G. Fichte im Gespräch: Berichte der Zeitgenossen*, Stuttgart-Bad Cannstatt: Frommann, 6 vols.(An encyclopedic collection of contemporary reports on Fichte and his writings. An invaluable research tool.)
Henrich, D. (1967) *Fichtes ursprüngliche Einsicht*, Frankfurt: Klostermann; trans. D. Lachterman, 'Fichte's Original Insight', *Contemporary German Philosophy* 1: 15-52, 1982.(An influential reading of Fichte's alleged movement beyond a 'reflective theory of consciousness'.)
Janke, W. (1970) *Fichte: Sein und Reflexion - Grundlagen der kristischen Vernunft*, Berlin: de Gruyter.(A hermeneutic reading of Fichte, with an emphasis upon the later writings.)
Lauth, R. (1984) *Die transzendentale Naturlehre Fichtes nach den Prinzipien der Wissenschaftslehre*, Hamburg: Meiner.(A masterly exposition of Fichte's transcendental approach to the philosophy of nature.)
Lauth, R. (1987) *Hegel vor der Wissenschaftslehre*, Stuttgart: Steiner-Verlag.(A vigorous and convincing defence of Fichte against the familiar criticisms of Hegel.)
Lauth, R. (1989) *Transzendentale Entwicklungslinien von Descartes bis zu Marx und Dostojewski*, Hamburg: Meiner.(A collection of essays, most of them on Fichte, by the leading Fichte scholar of the age.)
Léon, X. (1922, 1927) *Fichte et son temps*, Paris: Armand Colin, 3 vols.(Though outdated, this remains the best study of Fichte's 'life and thought'.)

Neuhouser, F. (1990) *Fichte's Theory of Subjectivity*, Cambridge: Cambridge University Press.(A good example of a contemporary appropriation of Fichte's thought and of an analytically sensitive exposition of the same.)

'New Studies in the Philosophy of Fichte' (1979) special issue of *Idealistic Studies* 6 (2).(A collection of essays on Fichte in English.)

Pareyson, L. (1976) *Fichte. Il sistema della liberté*, Milan: Mursia, 2nd edn.(Along with Philonenko, Pareyson's exposition of the early system as a 'system of freedom' is one of the most influential works on Fichte of the post-war period.)

Philonenko, A. (1966) *La liberté humaine dans la philosophie de Fichte*, Paris: Vrin; 2nd edn, 1980.(A tremendously original and influential study of Fichte's early philosophy, interpreted as a 'philosophy of human finitude'. Essential.)

Radrizzani, I. (1993) *Vers la fondation de l'intersubjectivité chez Fichte: des Prinzipes à la Nova Methodo*, Paris: Vrin.(Argues for the continuity of Fichte's development within the Jena period and for the centrality therein of the *Wissenschaftslehre* nova methodo.)

Renaut, A. (1986) *Le système du droit: Philosophie et droit dans la pensée de Fichte*, Paris: Presses Universitaires de France.(A powerful reading of the *Foundations of Natural Right*. Argues that political philosophy is the keystone of the Jena *Wissenschaftslehre*.)

Rockmore, T. (1980) *Fichte, Marx, and the German Philosophical Tradition*, Carbondale, IL: Southern Illinois University Press.(The first successful effort in English to liberate Fichte's philosophy from the shadow of Hegel's.)

Seidel, G.J. (1993) *Fichte's Wissenschaftslehre of 1794: A Commentary on Part I*, Lafayette, IN: Purdue University Press.(An elementary introduction to the early system. Written with the beginning student in mind.)

Widmann, J. (1977) *Der Grundstrukture des transzendentalen Wissens nach J. G. Fichtes Wissenschaftslehre 1804*, Hamburg: Meiner. (A elaborate formal analysis of what is widely considered to be the most successful and complete presentation of the later Wissenschaftslehre.)

Wundt, M. (1929) *Fichte-Forschungen*, Stuttgart: Frommann.(Another pioneering study of the development of Fichte's thought, with an emphasis upon the 'spirits' of the various Wissenschaftslehren.)

Ficino, Marsilio (1433-99)

With Giovanni Pico della Mirandola, Marsilio Ficino was the most important philosopher working under the patronage of Lorenzo de'Medici, 'Il Magnifico', in the Florence of the High Renaissance. Ficino's main contribution was as a translator of Platonic philosophy from Greek into Latin: he produced the first complete Latin version of the works of Plato (1484) and Plotinus (1492) as well as renderings of a number of minor Platonists. He supplied many of his translations with philosophical commentaries, and these came to exercise great influence on the interpretation of Platonic philosophy in the Renaissance and early modern periods. Ficino's most important philosophical work, the Theologia platonica de immortalitate animae (Platonic Theology, On the Immortality of the Soul) (1474) aimed to use Platonic arguments to combat the Averroists, 'impious' scholastic philosophers who denied that the immortality of the soul could be proven by reason. The most famous concept associated with his name is that of 'Platonic love'.

1 Life and mission

Marsilio Ficino was born in 1433 in Figline Valdarno, a small town in the territory of Florence; his father, Dietifeci, was a doctor who worked in the city of Florence and served as private physician to Cosimo de'Medici, then the *de facto* ruler of Florence. Dietifeci intended that his son should follow him in the medical profession, and the young Marsilio was sent to the University of Florence to hear the lectures on Aristotelian logic and natural philosophy that would fit him for his medical studies. Though Ficino retained a lifelong interest in medicine, he eventually left the university without taking his degree. He had discovered that his true bent was for philosophy and theology, especially the Platonic and Augustinian theology he met in his confraternity, led by the priest and theologian Lorenzo Pisano. Thanks to the indulgence of Cosimo de'Medici, Ficino was able to study Greek so as to have direct access to Plato's works, only a few of which were available in Latin at the time (see Platonism, Renaissance). Ficino's desire to return to the original Greek text was also a symptom of his ties with the humanist movement; he had enjoyed something of a classical education under two minor humanist schoolmasters, and was surrounded in later life by friends, patrons and pupils who were deeply involved in the movement for 'good letters' (see Humanism, Renaissance).

In addition to supporting his Greek studies, Cosimo de'Medici had lent the young Ficino a precious manuscript containing the complete works of Plato, and Ficino's first large translation project was a Latin version of ten dialogues of Plato, dedicated to Cosimo de'Medici a few weeks before the latter's death in 1464. Before 1469 Ficino had drafted a translation of all thirty-six dialogues in the Thrasyllan canon, though the final version was not printed until 1484. All of these dialogues were provided with short 'arguments', philosophical in character, and several were also supplied with full-scale philosophical commentaries. Ficino eventually wrote commentaries on the *Timaeus*, *Symposium*, *Philebus*, *Phaedrus*, *Parmenides*, *Sophist*, and on the mysterious passage in Book VIII of the *Republic* on the 'Platonic number'; these were published together in 1496. In addition to his work on Plato, Ficino also produced an exceptionally fine translation and commentary on Plotinus' *Enneads* (published 1492), a version of Pseudo-Dionysius (dedicated 1492), and versions of texts by the Platonists Iamblichus, Proclus, Porphyry, Synesius, Michael Psellus, Priscianus Lydus, Albinus, pseudo-Pythagoras, Speusippus and pseudo-Xenocrates (collection published 1499). Ficino's work of translation and commentary makes him by far the most important conduit for the transmission of the ancient Platonic tradition to Renaissance and early modern Europe.

Ficino was able to devote himself to his scholarly and philosophical labours thanks largely to the support of the Medici family over four generations as well as of other wealthy Florentine aristocrats. Cosimo de' Medici had given Ficino a house in Florence and a small farm near his own villa at Careggi; Cosimo's son Piero gave Ficino a brief appointment as a lecturer at the Florentine *studio*; Piero's son, the great Lorenzo, arranged for Ficino to have two small ecclesiastical benefices after the philosopher was ordained a priest in 1473 and in 1487 had him made a canon of the Cathedral of Florence. Before receiving this major benefice Ficino supplemented his income by tutoring wealthy pupils in rhetoric, philosophy and other subjects. He referred to his pupils collectively as his *gymnasium* or *academy* and, since most of his students concurrently attended the University of Florence (also called a gymnasium or academy by contemporary humanists), the confusion of terminology led later scholars to elaborate the rich fable of Ficino's 'Platonic Academy' (an expression not documented before 1638). In fact,

Ficino's closest, though informal, institutional link was with the University of Florence.

Though Ficino was not the leader of a Platonic Academy at the centre of Florentine intellectual life, his influence went far beyond what might be expected of a private scholar and teacher. Thanks to his position as a protégé of the Medicis, his own wide circle of pupils and friends, his numerous publications and an extensive, Europe-wide correspondence, Ficino became the leader of a philosophical counterculture that challenged in a small but significant way the dominance of Aristotelian school philosophy (see Aristotelianism, Renaissance). Ficino saw himself as an educational and cultural reformer. A sincere Christian, he believed that religious belief in his day was under threat from the growing estrangement of piety from philosophy. On the one hand were priests and other religious authorities, too ignorant of philosophy to defend Christianity; on the other hand were impious university philosophers, chiefly those whom Ficino labelled the 'Averroists' and 'Alexandrists' (that is, followers of Averroes' and Alexander of Aphrodisias' interpretations of Aristotle's psychology), who denied the ability of philosophy to prove central tenets of the faith like the immortality of the soul. In Ficino's view the great medieval project to integrate Christianity and Aristotelian philosophy had ended in intellectual disaster. But Ficino believed the best way to combat the growing autonomy of philosophy from religion in the schools was neither to retreat to fideism nor to seek a more successful Christian interpretation of Aristotle, but to replace Aristotle with Plato as the primary philosophical authority of Christendom. It was Ficino's belief that Plato, who had believed in creation and in the immortality of the soul, would provide a better foundation for Christian belief than Aristotle, who had believed in the eternity of the world and had been fatally unclear on the immortality question. In this Ficino was supported by the authority of Augustine, who in a much-quoted passage of *De civitate Dei (The City of God)* had declared Plato to be the pagan philosopher closest to Christianity. Indeed, Ficino (who also thought himself to have prophetic powers) declared that he had personally been appointed by divine providence to bring about a rebirth of Platonic theology in order to save the Christian faith from impious intellectuals.

Ficino's other sharp difference from contemporary scholasticism was his interest in the therapeutic side of ancient philosophy. This interest grew out of his early medical training, which helped him recognize the close connection between medicine and philosophy in Plato, Aristotle and the Hellenistic philosophers. Psychic health for Ficino involved both the cultivation of moral virtue and the achievement of a proper equilibrium between states of the body and states of the soul. The former theme is mostly explored in Ficino's letters; the latter in his major work on magic, the *De vita (On Life)*. In this work Ficino tries to find magical remedies for scholar's melancholy, in particular psychosomatic remedies that work on the *spiritus* - the *tertium quid* linking soul and body. It was once believed on the authority of Frances Yates that the sources of Ficino's magic were to be found in Hermes Trismegistus, but this has recently been shown to be an error; Ficino's actual sources were the same late ancient Platonic texts that guided his interpretation of Plato (see Hermetism).

2 The 'ancient theology'

Ficino's real interest in 'Thrice-Great Hermes' was theological rather than magical. Ficino believed that the texts going under the names of Hermes Trismegistus and Orpheus, as well as the Chaldaean Oracles which he (following Pletho) attributed to Zoroaster, were theological writings of extreme antiquity, belonging roughly to the time of Moses. (In fact they were late ancient forgeries.) For him, they represented a tradition of gentile theology, parallel to the Hebrew Old Testament, that culminated in the writings of Plato. Like the Old Testament, they were inspired but obscure writings whose full meaning only became clear with the advent of Christ. Following Eusebius and Lactantius, Ficino held that the writings of the ancient theologians had prepared the gentiles for Christianity as the Old Testament writings had prepared the Jews. In Christian times the writings of the pagan philosophers and theologians in the Platonic tradition retained value for a variety of reasons. They provided a model for the kind of *pia philosophia* or *docta religio*, combining religious belief and philosophical wisdom, that Ficino wished to see revived in his own time. If Platonism could be successfully reconciled with Christian faith, the latter's credibility would be much enhanced among people educated by humanists for whom ancient authority was paramount. Platonism could also act as an introductory credo preparing the philosophically minded for the higher truths of Christianity, as it had in the case of Origen and Augustine. It could serve as an apologetic weapon against sceptical intellectuals: Ficino said as much in a letter praising Pico della Mirandola as a 'fisher of men' for having used Plato as a net to draw 'Averroists and Epicureans' back to Christianity.

It may be asked what could have led Ficino and his circle to promote so unusual a view of the role of ancient

theology and Plato within Christianity. It must be remembered that the efforts of those in the Renaissance movement to revalue all things classical led to a certain cognitive dissonance between classical and Christian values. While most Christian writers in antiquity were anxious to distinguish Christianity from the surrounding pagan culture, in the Renaissance many humanist writers wished to emphasize the common ground between Christian doctrine and the best pagan thought. Ficino's interpretation of the historical role of Platonism integrated the greatest of ancient philosophers into the Christian tradition, and strengthened the humanist hope that Christendom could renew itself from ancient sources without alienating itself wholly from pagan culture. The 'ancient theology' also presupposed Ficino's wider belief in natural religion: that religion was natural to all men, and that all manifestations of religion were in some degree good as expressions of the natural desire of man for God. Though Ficino did not declare all religions equal - he held that each species of belief should be arranged hierarchically with Christianity holding the first place in the genus - his attitude to religions other than Christianity stands in marked contrast with older medieval attitudes. While traditional church teachings declared Islam, for example, to be a demonic invention or a divinely-ordained scourge of sinful Christians, Ficino (like Nicholas of Cusa) regards Islam as a sincere though misguided form of worship. Ficino's conception of natural religion and the 'ancient theology' continued to retain a certain appeal throughout the later Renaissance; it has been shown that as late as the seventeenth century the Jesuits in China were attempting to convince their catechumens that the 'ancient theology' of the Chinese was a foreshadowing of the truths of Christianity.

3 The immortality of the soul

The question of the soul's immortality was perhaps the most hotly debated philosophical issue of the later fifteenth and early sixteenth century. After a long period in which 'the common doctrine of the philosophers' (that is, Averroism) had reigned almost unchecked in Italian universities (see Averroism), the 1470s saw a revival, in the religious orders and among secular thinkers like Ficino and Apollinaris Offredi, of attempts to establish rational proofs for the immortality of the soul. The campaign culminated with the Fifth Lateran Council (1512-7), summoned by Ficino's former pupil Leo X, which denounced 'those who assert that the intellective soul is mortal or is one for all mankind'. Ficino, like Thomas Aquinas and many others who argued for immortality, held that those who denied personal immortality were undermining the traditional belief in rewards and punishments in the afterlife, and thus removing a key sanction against immoral behaviour. In his argument to the tenth book of Plato's *Laws* Ficino went so far as to say that the immortality of the soul was the main foundation for religion.

This historical background does not however account fully for Ficino's preoccupation with human immortality. Ficino was the first philosopher to give the immortality of the soul the central place in his thought, a prominence indicated by the full title of his principal work, *Theologia platonica de immortalitate animae (Platonic Theology, On the Immortality of the Soul)* (1474). In his classic study of Ficino's thought (1943), P.O. Kristeller maintained that immortality was central to Ficino's work because of its close connection with contemplative experience. For Ficino, the experience of God which may be attained in contemplation constitutes the highest human act of consciousness, and it therefore both defines the essence of humanity and establishes the end of human existence. Since the contemplative experience of God in this life is only partial and transitory, the soul must be capable of some form of separate incorporeal existence in which its natural appetite for the knowledge and enjoyment of God may be fulfilled. Ficino does not seem to realize that this is at best an argument for the soul's survival after death and not necessarily for eternal existence, since he admits elsewhere that souls are created in time. But in fact, many of the dozens of immortality arguments that fill the last thirteen books of the *Theologia platonica* begin from similar subjective analyses of acts, habits, faculties, appetites and affinities of soul. The arguments from affinity, drawn in the first instance from Plato's *Phaedo*, are of particular importance, but Ficino also employs the argument from self-motion found in the *Phaedrus*, while the arguments from *appetitus naturalis* have Augustinian and Thomistic antecedents.

The immortal soul is also central to Ficino's philosophy in a broader metaphysical context. The early books of the *Theologia platonica* lay out an ontological hierarchy of five substances, God, angel, soul, quality and matter. This hierarchy was probably taken from Proclus, though the hypostasis of quality seems to be original with Ficino. Within the hierarchy soul functions as the central link: what is above soul is eternal, immaterial, unchanging, intelligible; what is below it is temporal, material, mutable and sensible. It is soul that binds the two spheres together and makes them a unity. Paradoxically, the unifying functions of soul within total metaphysical reality mean that the soul is radically divided within itself. It contains within itself two separate and opposing impulses. It

has a natural desire for God which drives it to cut itself off from the body and empirical reality, to turn within and upwards to the source of its being through rational activity. In addition to this contemplative nisus, obviously inspired by Plotinus, Ficino posits another which leads the soul to ‘care for’ lower things such as the body, for which it has a natural affection. To these two natural tendencies or affections there correspond the higher parts of the soul, pre-eminently the reasoning and noetic powers, and the lower parts of the soul such as sensation and vegetation which are responsible for its empirical activity.

Ficino’s definition of the soul, partly in consequence of this radical division within the soul, is not entirely satisfactory. In some passages he asserts, following Aristotle and Aquinas, that it has a hylomorphic relationship to the body, that is, it is the substantial form of the body. He rejects the Plotinian formulation of the relation of soul to body, in which the soul, while remaining separate from body, controls it by means of a *physis* or reflection of itself which it projects into the body. He also will not endorse the implausible solution of some medieval Augustinians, which calls for the soul to be a distinct substance, though composed of form and ‘spiritual matter’. Yet in most contexts Ficino does speak of the soul in Platonic terms as though it were an independent substance and not subject to material potencies. He infers from the fact that we form simple concepts and can conceive of pure simplicity that the soul itself is simple; he uses many of the old Platonic analogies, such as that the soul is imprisoned in the body, that the soul is to the body as the person at the helm is to the ship, and so forth.

Plato’s doctrine of the pre-existence and transmigration of souls presented Ficino with a major interpretative challenge, since it threatened the Florentine’s larger project of demonstrating the compatibility of Platonism and Christianity. The doctrine had been confidently attributed to Plato by Plotinus, Augustine, the standard commentary on the Bible (where King Herod is also said to have believed in it) and Thomas Aquinas. Several of Ficino’s contemporary opponents, such as the Dominicans Savonarola and Dominic of Flanders, used it to discredit Ficino’s Platonic revival and the doctrine was, significantly, condemned by the Inquisition in articles published at the University of Pisa in 1490. Ficino’s response was to deny that Plato had ever really held the doctrine. Metempsychosis as referred to in Plato and the other ancient theologians was to be understood as a mystery; to understand it in a literal sense was a vulgar error first put about in the late Academy. In reality the doctrine of transmigration must be taken as an allegory of the return of the soul to the One in contemplative experience, or typologically as a prophecy of the resurrection of the body, or as an obscure, proto-Christian premonition of the doctrine of purgatory. Ficino also employed the not wholly consistent argument that when Plato discussed transmigration in his works he was simply retailing a Pythagorean doctrine to which he was not himself necessarily committed. To the metaphysical problem of explaining how an eternal substance could have been created in particular souls at particular moments of time, Ficino argues that creation in this instance is to be understood as the ontological dependence of the temporal manifestations of soul on their eternal source, not as the production *ex nihilo* of new substances in time.

4 Platonic love

If the *Theologia platonica* was Ficino’s most substantial independent work of philosophy, his most influential work was his commentary on Plato’s *Symposium*, known as *De amore (On Love)* (1469). Unlike Ficino’s other commentaries, the *De amore* was cast in the form of a literary dialogue, set in the villa of Lorenzo de’Medici, with Ficino’s friends and patrons as the interlocutors. Each interlocutor gives a speech which is effectively a Neoplatonic reading of the several speeches in Plato’s *Symposium*. It was this work, and translations of it into Italian, French and German, that popularized the concept of ‘Platonic love’, a concept that became a popular poetical conceit in the later Renaissance and is still used in a debased sense in modern colloquial speech (see Love).

Ficino is usually credited as the inventor of this concept, and the expression *amor platonicus* actually occurs in one of his letters (though not in the *De amore*). Ficino’s account of love, however, is closely based on Plotinus (in particular *Enneads* 3.5), and a similar Plotinian reading of the *Symposium* can be found in Cardinal Bessarion’s *In Calumniatorem Platonis (Against Plato’s Calumniator)*, printed within a few months of the publication of Ficino’s *De amore* in 1469 (see Platonism, Renaissance §3). Nevertheless, it was certainly Ficino who gave the concept its currency in Renaissance Europe and invested it with the philosophical richness that gave it its wide appeal. As Kristeller (1943) points out, the concept combines a Plotinian reading of love with the will of St Augustine, the charity of St Paul, and ideas on friendship found in Aristotle, the Stoics, and Cicero’s *Laelius*.

Ficino's basic move is to interpret our experience of love in terms of the spiritual dynamics of the Neoplatonic cosmos. A true experience of love awakens one to the natural desire of the soul for union with God. It may begin with a sensual element but that is a mere preparation for genuine love, which is the love of God. The instantiations of beauty or goodness that kindle mutual desire between human beings are to be understood as reflections of the divine beauty and goodness. What we love in others rightly belongs to God; to give love to another without at the same time giving love to God, as Ficino says in a striking formulation, is 'nothing but robbery'. Yet the true basis of active love is not the unconscious dependence of attributes on their divine source, but a conscious striving of souls together towards God in contemplative experience. It is the active search for truth in the philosophical life which is the true basis of love and forms a genuine union between lovers. Real, divine love is thus independent of the sex of the lovers and can exist between members of the same or opposite genders. Ficino's concept of love in this way subsumes the Pauline and Augustinian concept of charity; it also absorbs the classical pagan concept of friendship. In the *Nicomachean Ethics*, for example, Aristotle argues that true friendship is necessarily between equals and has as its subject common pursuits. Ficino understands this to mean that souls united in contemplation of the same object are *eo ipso* equal. Such souls are always present to each other even when separated, and love even has the power to transform the lover into the image of the beloved.

Ficino insists that the biblical command to love one another has a metaphysical basis; he argues that unrequited love is, so to speak, spiritually defective. Mutual love which constitutes a concrete communion between persons is the only true and perfect form of love; it is not only a moral obligation but a cosmic necessity. If love is based on a similarity or equality between the souls of lovers and has the same divine source, then it must exist in both souls equally. A failure of charity is a denial of one's essential nature. It has recently been suggested that Ficino's emphasis on Platonic love has a political meaning, that it was meant as a cure for the endemic divisions in Florentine civil society. It has also been argued that Ficino feared criticism of Plato on account of the numerous scenes of homosexual gallantry depicted in the dialogues, and that the doctrine of Platonic love was in part intended as an exegetical device to protect Plato from this charge. Both statements may well be true. But the doctrine of Platonic love is one that grows naturally from the central themes of Ficino's philosophy, especially his emphasis on the special dignity of contemplative noesis among human cognitive powers and his belief in the unitive functions of soul within creation.

See also: Hermetism; Plato; Platonism, Renaissance

JAMES HANKINS

List of works

Ficino, M. (1433-99) *Opera omnia (Complete Works)*, Basle: Heinrich Petri, 1576; repr. Turin: Bottega d'Erasmo, 1959.(The only widely available edition for many of Ficino's Latin works; most of those not included here are given in Kristeller's *Supplementum Ficinianum*.)

Ficino, M. (1469) *De amore*, ed. and trans. R. Marcel, *Commentaire sur le Banquet de Platon*, Paris: Les Belles Lettres, 1956; trans. S. Jayne, *Commentary on Plato's Symposium On Love*, Dallas, TX: Spring Books, 1985. (Referred to in §4. Marcel's edition gives the Latin text with French translation; Jayne has a useful introduction describing the influence of *De amore* on Renaissance literature.)

Ficino, M. (1474) *Theologia platonica de immortalitate animae (Platonic Theology, On the Immortality of the Soul)*, ed. and trans. R. Marcel, *Théologie platonicienne de l'immortalité des âmes*, Paris: Les Belles Lettres, 1964-79, 3 vols.(Referred to in §3. Edition of the Latin text with a French translation; the third volume also contains most of Book 2 of Ficino's *Letters*.)

Ficino, M. (1489) *De vita libri tres*, ed. and trans. C.V. Kaske and J.R. Clark, *Three Books On Life*, Binghamton, NY: Center for Medieval and Early Renaissance Studies, 1989. (Referred to in §1. Critical edition of the Latin text of the *De vita* with English translation and notes.)

Ficino, M. (1496) *Commentaria in Platonem (Commentaries on Plato)*, Florence: Lorenzo di Francesco da Venezia; ed. M.J.B. Allen, *The Philebus Commentary*, Berkeley, CA: University of California Press, 1975; *Marsilio Ficino and the Phaedran Charioteer*, Berkeley, CA: University of California Press, 1981; *Icastes: Marsilio Ficino's Interpretation of Plato's Sophist*, Berkeley, CA: University of California Press, 1989; *Nuptial Arithmetic: Marsilio Ficino's Commentary on the Fatal Number in Book VIII of the Republic*, Berkeley, CA: University of California Press, 1994. (The 1496 edition is the final Latin text of the commentaries; now largely superseded by M.J.B. Allen's critical editions of the Latin text, which also provide English translations, monographs and notes.)

Ficino, M. (1937) *Supplementum Ficinianum: Marsilii Ficini… opuscula inedita et dispersa*, ed. P.O. Kristeller, Florence: Olschki.(Contains most of the Latin works not included in Ficino's *Opera* of 1576; gives textual improvements to, and the textual history of, many of the works included in the *Opera*.)

Ficino, M. (1975-88) *The Letters of Marsilio Ficino*, trans. by Members of the Language Department of the School of Economic Science, London: Shepheard-Walwyn, 4 vols to date. (Referred to in §1. English translation of the first five of Ficino's twelve books of philosophical letters.)

Ficino, M. (1990) *Lettere. I. Epistolarum familiarium liber I (First Book of Familiar Letters)*, ed. S. Gentile, Florence: Olschki.(Referred to in §1. Critical edition of the Latin text of Book I of Ficino's *Epistolarum familiarium*, with important biographical material in the introduction.)

References and further reading

Allen, M.J.B. (1984) *The Platonism of Marsilio Ficino: A Study of His Phaedrus Commentary, Its Sources and Genesis*, Berkeley, CA: University of California Press.(Supplements Kristeller 1943 for a number of themes in Ficino's philosophy.)

Hankins, J. (1990) *Plato in the Italian Renaissance*, Leiden: Brill.(A study of Ficino as a translator and interpreter of Plato.)

Hankins, J. (1991) 'The Myth of the Platonic Academy of Florence', *Renaissance Quarterly* 44: 429-75. (Expansion of material in §1; disproves the myth of Ficino's 'Platonic Academy'.)

Kristeller, P.O. (1943) *The Philosophy of Marsilio Ficino*, New York: Columbia University Press; revised edn, *Il pensiero filosofico di Marsilio Ficino*, Florence: Le Lettere, 1988.(Referred to in §§3-4. The classic study of Ficino's philosophical thought. Revised edition was published in Italian with an updated bibliography.)

Kristeller, P.O. (1987) *Marsilio Ficino and His Work after Five Hundred Years*, Istituto Nazionale di Studi sul Rinascimento, *Quaderni di 'Rinascimento'* 7, Florence: Olschki.(Contains an exhaustive bibliography of secondary literature on Ficino as well as additions and corrections to Kristeller's *Supplementum Ficinianum* (1937).)

Marcel, R. (1958) *Marsile Ficin*, Paris: Les Belles Lettres.(The standard biography, not very critical.)

Fiction, semantics of

Taken at face value, 'Anna Karenina is a woman' seems true. By using Tolstoi's name 'Anna Karenina' and the predicate 'is a woman' we appear to refer to the character Anna and to attribute to her a property which she has. Yet how can this be? There is no actual woman to whom the name refers. Such problems of reference, predication and truth also arise in connection with representational art and with beliefs and other attitudes.

Meinong distinguishes the 'being' of objects (including fictional objects) from the 'existence' of actual objects such as Napoleon. 'Anna Karenina' refers to a concrete, particular, nonexistent object that has the property of womanhood. However, Meinong's distinction seems to many ontologically suspect. Perhaps, then, being is existence and 'Anna Karenina is a woman' is actually false because 'Anna Karenina' has no referent. Russell in 'On Denoting' (1905) agrees. But how can we explain the apparent contrast in truth between this sentence and the unquestionably erroneous 'Anna Karenina is from Moscow'? Or is it that being is existence but 'Anna Karenina' refers to an abstract, not a concrete, thing - an existent, abstract thing that does not have the property of being a woman but has merely the property of being said, by Tolstoi's novel, to be a woman? Then, however, the meaning of our sentence about Anna no longer parallels that of 'Emily Dickinson is a woman'. Perhaps, as many argue, we only pretend that 'Anna Karenina' refers and that the sentence is true. This position may not adequately explain the intuitions that support Anna Karenina as a genuine object of reference and predication, however.

1 Meinongian and abstract-object views

Meinong distinguishes the 'being' of objects (including fictional objects) from the 'existence' of actual objects such as Napoleon. Although Meinong's position at first blush respects our face-value understanding of sentences about fiction, it encounters difficulties (see Meinong, A.). Fictional characters normally are incomplete: it is not true, according to the novel, that Anna is right-handed; and it is not true, according to the novel, that Anna is not right-handed. Yet how, without violating the law of excluded middle, can there be such a thing? Again, some fictions are inconsistent: in a story, a wood sprite squares the circle. Does the realm of being then include an object that infringes the law of contradiction?

Contemporary Meinongians suggest answers to such difficulties. Thus Parsons (1980) distinguishes sentence negation ('*x* is not *P*') from predicate negation ('*x* is non-*P*'). He correlates existent objects one to one with the sets of ordinary properties that they have and extends the correlation by introducing nonexistent objects as entities that are correlated one to one with the sets of ordinary properties that are not correlated with existent objects. No object is contradictory, that is, no object satisfies a formula of the form '*x* is *F* and not (*x* is *F*)'. But objects may have impossible properties - for example, being *P*-and-non-*P* - without implying the truth of any contradictions. (Such objects are impossible objects and cannot exist.) Objects also may be incomplete without violating excluded middle: the lack both of being-*P* and of being-non-*P* does not contradict the fact that an object either is *P* or is not *P*.

Such theories show that Meinongian views can be defended with far more plausibility than has been thought by philosophers heavily influenced by Russell's and others' identification of being and existence. Nevertheless, these views do not easily capture all aspects of sentences relating to fiction. For one thing, there are problems specific to particular Meinongian approaches. Thus in Parsons' theory, as in non-Meinongian theories that regard characters as collections of properties (for example, Wolterstorff 1980), it is difficult to allow that a novel may introduce two distinct characters assigned the same properties. Again, our epistemic situation and that of Sherlock Holmes are relevantly similar: Holmes senses and reasons about objects in the world, and so do we. How then can we know (as we do) that we exist, given that he - a nonexistent thing - cannot know that he exists? For another thing, the 'face-value' understanding just sketched of sentences relating to fiction - in which in saying 'Anna Karenina is a woman' we state a truth, refer to the character Anna, and attribute to her the property of womanhood - is over simple.

Thus although 'Anna Karenina is a woman' may be true *in the story*, so also is 'Anna Karenina exists'; and if we grant that the same character may occur in different stories and be described contradictorily by them, then it may be true that the character both is *P* (in one story) and is not *P* (in another story). On pain of contradiction, Meinongians cannot understand such claims as together implying that objects both have and do not have the

relevant properties. Complex Meinongian treatments of these claims can be devised. But it seems simpler to allow story-relative properties and then take fictional names to designate abstract, but existent, things that have properties such as being-*P*-according-to-story-*s* and being-not-*P*-according-to-story-*t* without having the non-relative properties of being-*P* and of being-not-*P*. We can thus proceed without having to introduce mysterious Meinongian nonexistent, particular things at all. As discussed in §§2-3 below, there are still further alternatives to considering claims relating to fiction as simple, literal truths in which objects of reference are assigned properties. Many writers would say that unless all these alternatives fail, it is unreasonable to commit ourselves to the perplexities of the Meinongian approach.

The one alternative that is clearly unsatisfactory is the unsupplemented Russellian view that a claim such as ‘Anna Karenina is a woman’ is false (or lacks truth-value, as Fregeans hold). Perhaps such a claim, taken on its own, is indeed false (or expresses no proposition at all, given direct-reference theories of proper names). But such a view leaves unexplained the contrast between that claim and ‘Anna Karenina is from Moscow’. Nor does it explain our inclination to ascribe truth to further claims such as ‘Sherlock Holmes is a fictional detective’, ‘Odysseus is the same as Ulysses’ and ‘Sherlock Holmes is far more famous than any real detective’ (Howell 1979; Parsons 1980; van Inwagen 1977). Unsupplemented Russellian views suggest nothing about the semantic or pragmatic factors which may enter into the understanding and acceptance of these further claims. (With qualifications, a similar point also holds for proposals such as that made by Woods (1974) to understand the truth of claims about fiction by appeal to the specific words or sentences that actually occur in the fictions themselves.)

The abstract-existent-object position allows for the further claims just noted, and that fact together with its non-Meinongian way of accepting genuine fictional entities makes it attractive. In van Inwagen’s version of this position, characters of fiction are theoretical entities of literary criticism. Sentences in a work of fiction that describe a character must be distinguished from our own sentences about that character. As part of the novel, Tolstoi’s sentence ‘Anna took the tray’ is not about anything, expresses no proposition, is not used by Tolstoi as a vehicle of an assertion. But my description of the novel, using the same sentence, expresses a true proposition about Tolstoi’s character: the proposition that the property of taking a tray is ascribed to that character in *Anna Karenina*. Like all entities, that character exists and obeys the laws of excluded middle and of contradiction. The evidence for the existence of characters is provided by sentences such as ‘Anna is the principal character of *Anna Karenina*’ or ‘There is a character in a nineteenth-century novel who is presented in greater physical detail than are any characters in eighteenth-century novels’. We accept these sentences as true, and it seems impossible to paraphrase them in ways that say the same thing yet involve no quantification over characters. Yet these and other sentences can be straightforwardly represented using van Inwagen’s apparatus: ‘Anna Karenina has the property of being the principal character in *Anna Karenina*’; or (roughly) ‘There is a character *c* and a nineteenth-century novel *n* such that there are properties ascribed to *c* in *n* and some of those properties imply more physical detail than any properties ascribed to any characters in any eighteenth-century novels’.

The abstract-existent-object theory provides the most plausible account so far that recognizes genuine fictional objects without accepting Meinongianism. But the theory has never been presented in full detail. Moreover, as used by us the name ‘Anna Karenina’ (unlike ‘Emily Dickinson’) is not the usual proper name that, according to direct-reference theories of proper names, refers directly to its bearer through an initial act of baptism. Rather, we refer to the theoretical entity via the name ascribed to it in the novel. As noted, we also attribute to this entity the property of having the property of womanhood ascribed to it in the novel, rather than the simple property of womanhood. Yet the theory by itself offers no particularly plausible explanation why, when we are really treating Tolstoi’s character in this way, we none the less use language appropriate to the reference and predication that concern ordinary concrete particulars (‘Anna Karenina is a woman’).

2 Pretence and make-believe

Instead of directly attacking questions about fictional reference, predication, truth and ontology, perhaps we should focus on the fictional use of language. Clearly (to ignore issues about fictional names), sentences in fiction or about characters in fiction are not distinguished from ordinary literal sentences by special lexical meanings or grammatical constructions. Like ordinary literal sentences, such sentences also may be true (as are various of Tolstoi’s claims about Napoleon in *War and Peace*) and may be vehicles of assertion (as, again, are various of Tolstoi’s claims). Perhaps, however, as some speech-act theorists have suggested, fictional claims are specially

marked out by their use in communication, for example, by possessing a special illocutionary force (see Speech acts §1).

In one version of this suggestion, in producing fiction the author performs no new illocutionary act analogous to ordinary acts of asserting, questioning or requesting. Instead, authors pretend, in a non-deceptive way, to perform such acts; for example, pretend to be making assertions about events of which they have knowledge. This suggestion fails, for not all pretences to assert produce fiction (for example, putting on an illustrative performance of making a silly claim), and not all productions of fictional sentences are pretences to assert (typing out one's story and mailing it to a publisher).

In another version of the suggestion, the author performs a special new sort of illocutionary act. For example (as Currie (1990) proposes), the author produces sentences with the intention that the audience will *make believe* the content of those sentences (that there is someone named 'Anna Karenina' who is a woman, and so on) through the audience's recognition of the author's intention. Like other pretence and make-believe approaches (including Walton's, described below, which has influenced Currie), this suggestion explains why both Tolstoi and we produce the sentence 'Anna Karenina is a woman' even while we recognize that 'Anna Karenina' denotes no actual person. We do so because that sentence gives the content that we are to make believe is true. (Because the sentence mimics claims describing real persons, we can also easily use it in making believe that there is a real person whom it truly describes.) One might object that the notion of fiction includes nonlinguistic items such as representational paintings and sculptures as well as sentences. However, its defenders argue that the preceding suggestion can be generalized to cover such items.

A deeper objection derives from Walton's nonlinguistic theory of fiction. For Walton (1990), individuals and groups play games of make-believe, the rules of which require the participants to imagine various matters to be true. Props generate fictional truths in such games. (Thus in a game in which the rule is that stumps are to be imagined to be bears, a stump by its presence generates the fictional truth that a bear is present.) Fictions - including linguistic fictions, representational paintings and sculptures - serve as props in games of make-believe. A text constitutes a fiction when, roughly, there is a rule in force that we are to make believe that there are objects such that the words of that text refer to and describe those objects. (Thus we make believe that 'Anna Karenina' is a genuine proper name that directly refers to a Russian woman of whom the story is telling us, and so on.) In Walton's account, the fact that words constitute a fiction depends on their thus being a prop in a game of make-believe, not on the fact that an author intends the audience to make believe that the propositional content expressed by the words is true. For Walton, fiction has nothing special to do with communicative acts of intentional agents; naturally occurring cracks in a rock can spell out a story.

These last points are controversial. (Perhaps the cracks are only treated by us as fiction, without really being so, as Currie argues.) But the basic idea, developed in detail by Walton, that our claims about fiction are to be understood in terms of make-believe provides a deep, powerful explanation of much that concerns those claims. Quite possibly, and as both Walton's and Currie's views suggest, we only make believe that through Tolstoi's sentences about Anna Karenina we get genuine reference, predication and truth.

3 Further questions

Nevertheless, questions remain: about the content of our make-believe, about non-make-believe claims concerning fiction, and about ontology. If we make believe that 'Anna Karenina is a woman' is true, what exact proposition do we make believe is true? If in a piece of literary criticism I assert that Anna is from St Petersburg - or if I compare her with real persons or with other characters ('Anna Karenina is a stronger, more developed personality than is Theodor Fontane's Effi Briest') - then I seem to be stating a sober, literal truth rather than simply making believe that something is true. If you and I independently remark that Tolstoi's character Anna is from St Petersburg, our remarks seemingly concern the selfsame character. But how can they, if, without having any particular object in mind, we each independently merely make believe that 'Anna Karenina' is a genuine name that refers to some real entity?

Writers such as Walton and Currie suggest somewhat different answers to such questions. Common to their answers, however, is the view that the most we can make believe - in propositional content - is the existential general claim that there is a woman who is named 'Anna Karenina', and so on. We cannot literally make believe *of*

a concrete, particular fictional character that *it* is so-and-so because, actually, there are no such characters. This last point can be supported by an argument of Kripke's in *Naming and Necessity* (1980). Our make-believe is not *of* any actual, concrete, particular entity genuinely referred to by 'Anna Karenina'. Nor (to ignore inconsistent fictions) can it be *of* any purely possible thing that exists in all the possible worlds compatible with what is said in the novel and is picked out from each such world by the properties that the novel assigns to Anna. There is no way to decide which of the many possible entities that have all those properties is the unique, particular referent of the name.

Again, writers in the make-believe tradition may understand our non-make-believe claims to indicate or allude to practices of make-believe. Thus, as Walton suggests, my sober, literal assertion that Anna Karenina is from St Petersburg may in effect claim that to say such a thing in the appropriate game of make-believe is fictionally to speak truly. Finally, when both you and I independently talk of Anna Karenina, or when I compare Anna Karenina to Effi Briest, then, Walton urges, we engage in a shared game of make-believe, or I simply combine the games individually appropriate to *Anna Karenina* and *Effi Briest*. (About such literal assertions and comparisons, Currie offers somewhat different treatments, involving what he holds are further, 'metafictive' and 'transfictive' uses of fictional names.)

Some of these suggestions may succeed. But in the end they deny that there can be a single fictional character - a single entity created by the author - that in actual fact two people independently refer to and describe. They also turn out to deny that it is literally and actually true that one and the same character - one single entity - can occur in different works of fiction (or can be the common focus of different games of make-believe; or of different dreams, illusions and beliefs). If we take seriously the idea that there actually are such characters of fiction, then we are forced back into accepting characters as objects of genuine reference and true predication. If so, then unless some Meinongian or similar view is adopted, it seems that we should combine the abstract-existent-object theory with the make-believe approach. Or perhaps (swallowing popular objections to Lewis's modal realism) we should accept a version of Lewis's counterpart theory (see Lewis, D.K.). We might hold, for example, that 'Anna Karenina' refers to various concrete, particular, non-actual individuals that exist in the possible worlds in which Tolstoi's novel is told as known fact. Those would be the individuals that have all the properties the novel attributes to Anna and that (while not strictly identical to one another) are enough alike to function as counterparts (see Modal logic, philosophical issues in §4). It remains to be seen, however, whether these proposals, or any other ideas for countenancing genuine reference to and predication of characters of fiction, can be successfully defended.

See also: Fictional entities; Proper names; Reference

ROBERT HOWELL

References and further reading

Adams, F., Fuller, G. and Stecker, R. (forthcoming) 'The Semantics of Fictional Names', *Pacific Philosophical Quarterly*.(Proposes forgoing fictional characters in favour of talk of fictional names, using the notion of a partial proposition.)

Crittenden, C. (1991) *Unreality*, Ithaca, NY, and London: Cornell University Press. (Defends a Meinongian view against competing theories.)

Currie, G. (1990) *The Nature of Fiction*, Cambridge: Cambridge University Press.(A comprehensive, well-argued make-believe account of the semantics of fiction and related issues. Many valuable references to the earlier literature. Technical in places. Discussed in §§2-3.)

Howell, R. (1979) 'Fictional Objects: How They Are and How They Aren't', *Poetics* 8: 129-77.(Presents difficulties for views rejecting fictional objects. Technical in places.)

Inwagen, P. van (1977) 'Creatures of Fiction', *American Philosophical Quarterly* 14: 299-308.(A forceful defence of an abstract-existent-object view. Discussed in §1.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.(A centrally important discussion of reference in general with application to fictional terms such as 'unicorn' and 'Sherlock Holmes'.)

Kroon, F. (1994) 'Make-Believe and Fictional Reference', *Journal of Aesthetics and Art Criticism* 52: 207-14. (Presents problems for make-believe accounts of fictional names; suggests a counterpart theory.)

Lewis, D.K. (1978) 'Truth in Fiction', in *Philosophical Papers*, New York and Oxford: Oxford University Press, 1983, vol. 1, 261-76, with 'Postscripts', 277-80.(A classic pretence account of truth in fiction, with analyses of

'in fiction *f*, so-and-so is the case' claims.)

Parsons, T. (1980) *Nonexistent Objects*, New Haven, CT, and London: Yale University Press.(A careful defence of a Meinongian theory. Technical in parts but with excellent nontechnical discussions. Discussed in §1.)

Routley, R. (1980) *Exploring Meinong's Jungle and Beyond*, Canberra: Research School of Social Sciences, Australian National University.(A broad discussion of Meinongian theories and their applications. Technical in places.)

Russell, B. (1905) 'On Denoting', *Mind* 14: 479-93; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992.(A classic treatment of reference and description.)

Walton, K. (1990) *Mimesis as Make-Believe*, Cambridge, MA: Harvard University Press.(A rich, highly original development of a make-believe theory of all aspects of verbal and nonverbal fictions. Excellent references to the earlier literature. Offers a subtle account of how fictional truths are generated - implicit truths such as 'Anna Karenina has a cerebral cortex' as well as explicit ones. Discussed in §§2-3.)

Wolterstorff, N. (1980) *Works and Worlds of Art*, Oxford: Clarendon Press.(Treats characters as collections of properties. Regards works of fiction as vehicles of special illocutionary acts.)

Woods, J. (1974) *The Logic of Fiction*, The Hague: Mouton.(Treats fictional claims via the sentences occurring in fiction, utilizing substitutional quantification. Technical.)

Fictional entities

By 'fictional entities', philosophers principally mean those entities originating in and defined by myths, legends, fairy tales, novels, dramas and other works of fiction. In this sense unicorns, centaurs, Pegasus, the Time Machine and Sherlock Holmes are all fictional entities.

A somewhat different category of fictional entities is associated with empiricist philosophy. It includes entities apparently assumed by common discourse but which admit of no direct empirical experience. Thus Jeremy Bentham classified as 'fictitious entities' motion, relation, power and matter, as well as, notoriously, rights, obligations and duties. David Hume called substance, the self, even space and time 'fictions' andBertrand Russell thought ordinary things, such as Piccadilly or Socrates, were fictions, on the grounds that they are 'constructed' out of simpler, more immediate objects of acquaintance.

Philosophical interest in fictional entities thus covers a surprisingly wide area of the subject, including ontology and metaphysics, epistemology, logic, philosophy of language, and aesthetics. The first question that arises is how the distinction should be drawn between fictional and nonfictional entities. As the examples from Bentham, Hume andRussell show, this is by no means a straightforward matter. The next question concerns what to do with fictional entities once they have been identified. Here the primary philosophical task has been to try to accommodate two powerful yet apparently conflicting intuitions: on the one hand, the intuition that there are no such things as fictional entities, so that any seeming reference to them must be explained away; on the other hand, the intuition that because 'things' like Sherlock Holmes and Anna Karenina are so vividly drawn, so seemingly 'real', objects of thoughts and emotions, they must after all have some kind of reality. Broadly speaking, we can discern two kinds of philosophical approach: those which incline towards the latter intuition, being in some way hospitable to fictional entities; and the less hospitable kind, which incline towards the former and seek only to show how fictional entities can be eliminated altogether in the strict regime of rational discourse.

1 Distinguishing the fictional from the nonfictional

It might be supposed that what determines whether an entity is fictional or not is whether it exists. Perhaps fictional just means nonexistent. Certainly it is commonplace to contrast what is fictional with what is real. The crucial difference between Socrates and Sherlock Holmes is that the one existed and the other did not. The trouble comes when we find serious philosophers, such as Russell, describing even beings like Socrates as 'fictions'. That invites the thought that there are different kinds of fiction and even different kinds of nonexistence. If that is right it compromises any simple identification of the fictional with the nonexistent. Also philosophers in the twentieth century have become increasingly wary of unqualified talk about what exists and what does not. Among logicians and philosophers of science there has been a pronounced shift away from asking what entities exist (or are real) towards asking what entities particular theories are committed to. Underlying this shift and acting as a constraint on theory construction is the heuristic principle called Ockham's razor (after the medieval logician William of Ockham), namely 'Do not multiply entities beyond necessity'.

In seeking to refine the association of the fictional with the nonexistent it is helpful to distinguish the role of fictional entities in logic, epistemology and literary narrative. There is no compulsion to suppose that some uniform account of fictional entities must span these three applications. Indeed it becomes evident that different criteria for distinguishing the fictional from the nonfictional operate in each case.

First of all, fictions in logic are most commonly associated with eliminability by logical paraphrase. The idea is roughly this: that a fictional entity is the purported referent of an eliminable singular term. It relies on a distinction between apparent reference and genuine reference; the object of a genuine reference must exist, or be assumed to exist by a theory, while apparent but non-genuine reference is that which is eliminable by paraphrase without loss of content. The conception probably originated with Bentham, who thought that rights, for example, were fictitious entities, precisely because sentences containing 'rights' as a noun (such as 'The people demanded their rights') can be paraphrased into sentences in which no such noun or singular term appears (such as 'The people demanded fair treatment under the law'). As long as the paraphrase captures the significant content of the original it shows that the apparent reference in the original (to a specific class of entities) is not a genuine reference: thus are the entities 'fictional'.

Bentham's theory was a clear forerunner of the school of twentieth-century logical analysis epitomized by such philosophers asBertrand Russell, Rudolf Carnap and W.V. Quine. Russell's Theory of Descriptions eloquently showed that singular descriptive phrases, such as 'the highest prime' or 'the golden mountain', do not need to function logically as naming expressions (that is, having, in our earlier terminology, 'genuine reference') in order to be meaningful (see Russell, B. §9). Quine extended the thesis to all singular terms, including proper names, arguing that each is eliminable by logical paraphrase, in favour of quantifiers and predicates. Only bound variables, for Quine, are the true bearers of referential commitments (see Quine, W.V. §5). We will return to Russell's and Quine's theories in the next section, for they have been highly influential in deflationary accounts of fictional entities.

While fictions in logic connect with reference and paraphrase, fictions in epistemology rest on the idea of 'construction'. Ordinary objects can be thought of as fictions, in epistemological theories, just to the extent that they are thought of as 'constructed' (by the human mind) out of more basic, perhaps more real, elements. A characteristic kind of empiricism equates what is real with what is knowable and what is knowable with what is given in experience. For those empiricists who hold that the only immediate objects of experience are subjective entities like 'impressions' (Hume) or 'sense-data' (Russell), it is a short and natural step to the idea that enduring objects in space and time are mere 'constructs' or posits of the mind that round out the flux of experience. Quine, albeit rejecting any privileged foundation of knowledge, has famously compared physical objects with the gods of Homer and spoken of both as 'myths'.

No doubt for those who think of fictional characters in legend or literature as paradigmatic fictional entities, it will seem tendentious, even paradoxical, to describe physical objects as 'fictions'. However, etymology associates fiction more with 'making' or 'feigning' than with nonexistence and empiricist epistemologists can be taken to be emphasizing the different kinds of 'making', or imaginative invention, involved in human knowledge.Nelson Goodman has described knowledge itself as a species of 'worldmaking'. None of these philosophers wants to collapse the distinction between a fictional character, in the familiar literary-critical sense, and a real person.

So what sets apart the characters of literary narratives from the 'fictions' in logic and epistemology? One common idea is that fictional entities such as the Time Machine and Sherlock Holmes are logically rooted in the narratives in which they are first introduced; their very conception derives from the descriptive content of those narratives. To find out about Holmes one must read the Holmes stories. This already hints at a fundamental difference from nonfictional objects, for the latter do not derive their existence and nature from any narrative; to find out about them we must investigate the world, not merely a source text. Fictional entities of the literary type are 'made' or 'made up' precisely by being described in an act of storytelling, under the appropriate conventions. The fact that they do not exist in the real world can be seen to be as much a consequence of their origin (real objects are not made up by storytellers) as a defining feature of their fictionality.

2 Deflationary theories

The intuition that there are no such things as fictional entities, so that any apparent ontological commitment to them must be removed, is powerful among philosophers. The logical analysis of Russell and Quine provides a striking example of how such a commitment can be avoided. The eliminative strategy applied to logical fictions can be applied to more familiar literary fictions as well. When we describe the exploits of Sherlock Holmes or the properties of Pegasus we appear to be referring to, or speaking about, entities which have some kind of existence. But this is an illusion, according to deflationary theories. In fact we are not referring to any such things.

Russell's Theory of Descriptions sets the standard for deflationary theories. Suppose, regarding a character in a novel, we want to assert, 'The invisible man could see but not be seen.' For Russell, a logical analysis of our assertion would be somewhat as follows: 'There is one and only one thing that is a man, is invisible, and can see without being seen'. The latter sentence removes the apparent naming expression 'the invisible man' and asserts only that something or other satisfies a collection of predicates. The sentence turns out to be false (for there is no such thing) but is perfectly meaningful.

Quine showed that the analysis could be generalized to eliminate all singular terms, proper names included. He suggested that, from a logical point of view, names can be turned into predicates purely formally. Thus he analyses the seemingly true sentence 'Pegasus does not exist' as 'Nothing pegasizes'. By losing the troublesome singular

term ‘Pegasus’ and speaking only of the instantiation (or lack of it) of a complex, albeit artificial, predicate, the analysis neatly shows how it is possible to deny the existence of a fictional entity without incurring any ontological commitment to that entity. It is a lesson, for example, that atheists can usefully apply in their denials of the existence of God (for it might seem that even to use the name ‘God’ is to admit that there is such a being).

To the extent that the logical problem of fictional entities is that of how to discriminate apparent reference from genuine reference the Russell/Quine deflationary theory is highly effective. But it also has serious drawbacks in accounting for other intuitions concerning ordinary talk ‘about fictional characters’. For one thing it has the consequence that all sentences containing singular terms for fictional characters turn out to be false (because they involve false existence claims). Yet is there not a sense in which the assertion ‘Sherlock Holmes is a detective’ is true, in contrast, say, to ‘Sherlock Holmes is a ballet dancer’, which is clearly false? On the Russell/Quine view they are both false. What needs to be captured is the fact that the claims are *about the Holmes stories*, not about persons in the real world. What is meant is something like: ‘Within the Holmes stories, as told by Conan Doyle, Holmes is a detective (and not a ballet dancer)’. It is a further question how assertions of that kind are to be analysed. In some deflationary theories they are said to refer only to novels and their component sentences. That has the advantage that it too removes unwelcome ontological commitments, but is surprisingly difficult to sustain in practice (not least because truths about the content of fictional stories extend beyond the explicit content of the stories’ sentences: for example, it is surely true that Holmes did not travel in a rocket, though that is never made explicit by Conan Doyle).

Another drawback of the Russell/Quine view is that it fails to distinguish storytelling discourse from discourse about stories. If the sentence ‘Holmes returned to London’ occurs in a Holmes story it seems inappropriate to analyse it as a false assertion about the real world. Conan Doyle, the author, is not asserting, or attempting to assert, facts about the world; he is making up a story. If on the other hand readers report the content of the story by using the same sentence they, in contrast to the author, are making an assertion (true or false), but one about the story, and again not about the world.

There are other deflationary theories, which attempt to remove commitment to fictional entities. Two deserve particular mention. The first is that of Nelson Goodman (§2), who defends a strict nominalist ontology. Given that there are no unicorns and no centaurs, the predicates ‘is a unicorn’ and ‘is a centaur’ have the same extension, according to Goodman, namely, a null extension: the set of unicorns is identical to the set of centaurs because it is the empty set. Yet surely unicorns and centaurs are different? Goodman seeks to explain the difference, not by invoking fictional entities, but by appeal to further predicates that do differ in extension. For example, a unicorn-picture, as he puts it, is different from a centaur-picture and can be recognized as such. While normally ‘X is a picture of Y’ is a two-place predicate, relating a picture and an object (for example, a portrait and the Duke of Wellington), in ‘fictional’ cases, such as ‘This is a picture of Sherlock Holmes’, there is only a one-place predicate involved, namely ‘is a Holmes-picture’. Pictures can be sorted into ‘Holmes-pictures’, ‘unicorn-pictures’ and so on, without commitment to a separate class of ‘fictional entities’. Pictures, after all, are unproblematically real.

Another deflationary strategy comes fromKendall Walton, for whom works of fiction are ‘props in games of make-believe’. Readers of *Don Quixote* seem to be introduced to a dreamy knight, who tilts at windmills, crusades against Evil, and so forth. Furthermore, they seem to think about him, refer to him and even feel sympathy for his predicament. However, for Walton it is only make-believe that they do such things. It is all pretence; they are playing a game with the novel as prop. In Walton’s alternative idiom, ‘it is fictional that’ they describe him, refer to him and respond emotionally to him. There is not even any such proposition that Don Quixote tilted at windmills.Walton offers a systematic theory which locates all ‘relations with fictional characters’ in imaginative games; to speak ‘about characters’ is just an elliptical way of speaking about the relevant games.

3 Hospitable theories

A presupposition behind deflationary theories is that if there is no such thing as X then X cannot (literally) be referred to or spoken about. The focus in all cases is in removing the appearance of reference. Some philosophers, notably Richard Rorty, find this association between reference and ontological commitment unjustified. Referring, forRorty, is just ‘talking about’ and he claims we can talk about Holmes, the number three, the beauty of a landscape or moral values, for instance, without engaging any deep issues in ontology. Rorty views the very idea of ‘ontological commitment’ as pointless.

A more widely held view is that 'referring to fictional entities' does have ontological significance but that different kinds of *being* are involved. Fictional entities are, after all, entities. If Russell's Theory of Descriptions is the starting point for (modern) deflationary theories, then Alexius Meinong's Theory of Objects (*Gegenstandstheorie*) is the starting point for (modern) hospitable theories (see Meinong, A. §§2-4). Meinong postulated a realm of objects of which only a tiny subset are existent objects. He seems at least partially to have worked with a syntactic criterion for objecthood: any expression functioning as a singular term in a well-formed sentence designates an object. Notoriously, Meinong held that even 'the round square' denotes an object (albeit one that does not, and could not, exist in reality). Although Meinong's theory was strongly attacked byRussell, it has proved remarkably resilient, with defences by, for example, Terence Parsons and Charles Crittenden. The point, again, is not to collapse the distinction between Socrates and Sherlock Holmes but only to claim they are both objects, the one existent, the other nonexistent. An added difference, for Parsons, is that Holmes, unlike Socrates, is an 'incomplete' object, in the sense that for any given property it is not always determinate whether Holmes has that property or not. But what is to be gained by saying that *there are* nonexistent objects? Principally, it makes sense out of common ways of speaking, without the need for paraphrase: we refer to Holmes, we distinguish him from Dr Watson, we ascribe properties to him. What could be more natural than saying there is an *object* here?

Other theorists sympathetic to fictional entities attribute to them existence as *abstract* objects. Nicholas Wolterstorff has argued that fictional characters are kinds (the character Holmes is a person-kind, though not a kind of person), Peter van Inwagen views them as 'theoretical entities of literary criticism', comparable in status to plots, metres and rhyme schemes. A consequence for Wolterstorff is that characters not only exist (as kinds) but do so eternally; a writer 'selects' but does not strictly create characters, for the properties ('being a detective', 'solving mysteries', etc.) that constitute the relevant kinds are not themselves created. If fictional characters exist as abstract objects, what properties do they possess? Holmes, it seems, is a detective, yet no abstract object could be a detective. Van Inwagen responds by distinguishing those properties 'ascribed' to characters (such as 'being a man', 'smoking a pipe') that they do not literally possess and those which characters 'exemplify' (such as 'being introduced in Chapter 29', 'being wittily conceived'). The intuition that fictional characters are in some such way real is further supported by other common modes of speech, as when we say, for example, that Dostoevsky's characters are more realistic than those of Cervantes. It is difficult for deflationary theories to paraphrase away such references.

The debate about fictional entities goes to the heart of methodological issues in philosophy. It highlights the status and role of logical analysis, the nature of ontological commitment, and above all the remarkable indifference of ordinary language to philosophical worries about what does and does not exist.

See also: Carnap, R.; Emotion in response to art §3; Fiction, semantics of; Ontological commitments; Reference; Rights; Semantics

PETER LAMARQUE

References and further reading

Crittenden, C. (1991) *Unreality: The Metaphysics of Fictional Objects*, Ithaca, NY, and London: Cornell University Press.(Mentioned in §3. Good general introduction to the subject, with a point of view sympathetic to that of Meinong.)

Currie, G. (1990) *The Nature of Fiction*, Cambridge: Cambridge University Press.(Presents a detailed logical theory of all aspects of fictionality. Quite technical.)

Goodman, N. (1968) *The Languages of Art*, New York: Bobbs-Merrill.(Referred to in §2. A strict eliminative theory of fictional entities applied to all art forms.)

Howell, R. (1979) 'Fictional Objects: How They Are and How They Aren't', *Poetics* 8: 129- 77.(Careful, critical exposition of the main logical approaches.)

Inwagen, P. van (1977) 'Creatures of Fiction', *American Philosophical Quarterly* 14: 299- 308.(Referred to in §3. Attributes some reality to fictional entities.)

Lamarque, P.V. and Olsen, S.H. (1994) *Truth, Fiction and Literature: A Philosophical Perspective*, Oxford: Clarendon Press.(Useful introduction to logical, epistemological and literary conceptions of fiction, as outlined in §1.)

Meinong, A. (1960) 'Theory of Objects', in R.M. Chisholm (ed.) *Realism and the Background of Phenomenology*, Glencoe, IL: Free Press.(Referred to in §3. Classic statement of the case for nonexistent objects as well as

existent objects.)

Ogden, C.K. (ed.) (1932) *Bentham's Theory of Fictions*, London: Kegan Paul, Trench, Trubner & Co., Ltd. (Referred to in §2. Bentham's theory of fictions and paraphrase was a strong influence on twentieth-century logical analysis.)

Parsons, T. (1980) *Nonexistent Objects*, New Haven, CT: Yale University Press.(Referred to in §3. A sophisticated defence and development ofMeinong's theory. Technical in parts.)

Quine, W.V. (1953) 'On What There Is', in *From a Logical Point of View*, Cambridge, MA: Harvard University Press.(Referred to in §2. Classic statement of the logical criterion of ontological commitment.)

Rorty, R. (1982) 'Is There a Problem About Fictional Discourse?', in *Consequences of Pragmatism*, Brighton: Harvester Wheatsheaf.(Referred to in §3. Dismisses the whole approach of analytic philosophy to fictional entities.)

Russell, B. (1956) *Logic and Knowledge*, ed. R.C. Marsh, London: Allen & Unwin. (Includes 'On Denoting' and 'The Philosophy of Logical Atomism', where Russell develops and applies his Theory of Descriptions and conception of logical fictions. See §2.)

Walton, K.L. (1990) *Mimesis as Make-Believe: On the Foundations of the Representational Arts*, Cambridge, MA: Harvard University Press.(Referred to in §2. Presents an eliminative theory of fictional entities, in terms of 'games of make-believe'.)

Wolterstorff, N. (1980) *Works and Worlds of Art*, Oxford: Clarendon Press.(Referred to in §3. At times difficult; but a powerful case for fictional characters as eternal kinds.)

Fictionalism

'Fictionalism' generally refers to a pragmatic, antirealist position in the debate over scientific realism. The use of a theory or concept can be reliable without the theory being true and without the entities mentioned actually existing. When truth (or existence) is lacking we are dealing with a fiction. Thus fictionalism is a corollary of instrumentalism, the view that what matters about a theory is its reliability in practice, adding to it the claim that science often employs useful fictions. Perhaps the fullest expression of fictionalism occurs in Vaihinger's once popular philosophy of 'as if'.

Fictionalism is allied to instrumentalism, the brand of pragmatism associated with Dewey's 'Chicago School of Thought' (see Dewey, J.; Pragmatism). There is also fictionalism in the philosophy of logic and of mathematics (see Antirealism in the philosophy of mathematics). Dewey coined the term ' instrumentalism' (sometimes also 'experimentalism') to describe his pragmatic treatment of 'how thought functions in the experimental determination of future consequences' (Dewey 1943: 463). According to instrumentalism what we look for in all the various contexts of inquiry, whether around the house or in the laboratory, is instrumental reliability; that is, we want our theories or concepts to be useful in all the practical and theoretical endeavours for which we try them out. Such usefulness or reliability does not imply that the theory is true nor that the 'entities' involved actually exist. In such a case the 'entities' mentioned by the theory are in fact useful *fictions*.

Perhaps the fullest expression of fictionalism occurs in Vaihinger's once popular philosophy of 'as if' (see Vaihinger, H.). According to Vaihinger (1911), thinking by means of fictions is a fundamental cognitive process, as basic and pervasive as deductive or inductive thought themselves. Vaihinger makes a primary distinction between 'scientific' and the 'unscientific' fictions: that is, between those that actually prove useful (scientific) and those that do not. He cites atoms as an example of a scientific fiction, believing that they rest on the concept of point masses and the vacuum, neither of which he regards as real. Newton's laws would be another scientific fiction (see Mechanics, classical). By contrast the infamous 'dormitive virtue' proposed by the doctors in Molière's *Le Misanthrope* would be an example of an unscientific fiction. Vaihinger goes on to classify many different types of fiction and to document their occurrence in virtually every scientific discipline. The fundamental distinction for Vaihinger's account, as for any version of fictionalism, is that between a fictional proposition and a scientific hypothesis. For Vaihinger, scientific hypotheses are verifiable by observation and are chosen over rivals according to the probability that they are true. Fictions on the other hand are justifiable only in terms of their utility (not truth or probability) and are chosen over rivals as being the most expedient means to the ends we have in mind.

Vaihinger's work, and fictionalism in general, enters the debate over scientific realism by challenging the inference from something's being useful to its being real, from utility to reality. A fiction is precisely a useful construct that is not real. In terms of recent debates (see Scientific realism and antirealism) fictionalism challenges the explanationist argument for scientific realism, which holds that only the truth of a scientific theory and the actual existence of the entities occurring in the theory could account for the theory's instrumental success. Fictionalism's challenge to realism is to point out that we can account for success quite well by supposing nothing more than this, that at the observable level it is just *as if* the scientific theory were actually true. This may not seem satisfactory to the realist who will accept no explanation short of the literal truth of science and will therefore press the question as to why our theories possess this dispositional property of instrumental reliability (the 'as if'). Pragmatists, however, point out that requests for explanation have to stop somewhere and that stopping here, with the fundamental reliability of our theories, is the right place (indeed the only place) if we want to align our beliefs with our epistemological practice, since that practice all along is pitched to reliability. Whatever the final verdict, it does seem that we can respond to the realist and explain scientific success by treating successful scientific stories as, roughly speaking, scientific fictions.

Idealizations and approximations play an essential role in science (see Idealizations). These are paradigm examples of scientific fictions. Moreover if we think of models as maps, useful tools for navigating our way around but always and necessarily false to reality, then the idea of a scientific fiction encompasses models as well. In the present era of computer simulations and virtual reality, a dominant picture is the image of science as the builder of

useful models. Thus one productive way to appreciate fictionalism is to see it as concerned with how models are articulated and work in scientific thought. In that light it may have more to offer than just another twist in the realism debate. Rather we might think of fictionalism as the beginning of a more comprehensive philosophical treatment of modelling (see Models).

See also: Bentham, J.; Maxwell, J.C.

ARTHUR FINE

References and further reading

Cohen, M.R. (1923) 'On the Logic of Fictions', *Journal of Philosophy* 20: 477-88. (Questions Vaihinger's concept of a genuine fiction.)

Dewey, J. (1916) *Essays in Experimental Logic*, Chicago, IL: University of Chicago Press.(Introduces 'instrumentalism'.)

Dewey, J. (1943) 'The Development of American Pragmatism', in D.D. Runes (ed.) *Twentieth Century Philosophy*, New York: Philosophical Library, 451-67.(Definitive account of the Chicago School.)

Fine, A. (1993) 'Fictionalism', *Midwest Studies in Philosophy* 18: 1-18.(Expansion of the material of this essay that traces the historical connections between fictionalism and logical positivism.)

Frank, J. (1970) *Law and the Modern Mind*, Gloucester, MA: Peter Smith.(Applies fictionalism to the law.)

Schaper, E. (1966) 'The Kantian Thing-In-Itself as a Philosophical Fiction', *Philosophical Quarterly* 16: 233-43. (Traces the Kantian roots of fictionalism.)

Scheffler, I. (1963) *The Anatomy of Inquiry*, New York: Alfred A. Knopf.(Treats fictionalism as a thesis about the language of a regimented science.)

Vaihinger, H. (1911) *Die Philosophie des Als Ob*, Leipzig: Felix Meiner Verlag; trans. C.K. Ogden, *The Philosophy of 'As If'*, London: Routledge & Kegan Paul, 1924.(First publication of Vaihinger's philosophy. Contains significant introduction for the English translation.)

Field theory, classical

A physical quantity (such as mass, temperature or electrical strength) appears as a field *if it is distributed continuously and variably throughout a region. In distinction to a 'lumped' quantity, whose condition at any time can be specified by a finite list of numbers, a complete description of a field requires infinitely many bits of data (it is said to 'possess infinite degrees of freedom'). A field is* classical *if it fits consistently within the general framework of classical mechanics. By the start of the twentieth century, orthodox mechanics had evolved to a state of ontological* dualism, *incorporating a worldview where massive matter appears as 'lumped' points which communicate electrical and magnetic influences to one another through a continuous intervening medium called the electromagnetic field. The problem of consistently describing how matter and fields function together has yet to be fully resolved.*

Initially, most of the fields considered in mechanics were *material*: they possess mass and can be acted upon by forces. Simple examples are flexible strings, membranes, fluids (which include gases) and elastic spheres (as long as these objects are treated as continuously distributed and not merely as swarms of point masses). Derivative kinds of field can depend upon these material underpinnings - for example, the distribution of a temperature field. The laws for such material 'continua' can be rather tricky to formulate (see Mechanics, classical).

By the end of the nineteenth century, massless fields independent of particulate matter began to be considered in classical mechanics. When J.C. Maxwell originally proposed his equations for electromagnetism, he believed that this field was composed of an 'ether' that was similar in many respects to a common elastic solid (see Maxwell, J.C. §2). After the Michelson-Morley experiment of 1887, this material interpretation was abandoned and the electromagnetic field became regarded as a physical entity *sui generis*. It was now a substance without mass that could store and transmit energy as ably as matter.

In this respect, electromagnetism differs from what is loosely dubbed 'the classical gravitational field'. On the Newtonian account, if a massive particle is wiggled, any distant particle will instantly feel a gravitational tug due to an 'action-at-a-distance' force acting directly between the matter (see Newton, I. §§3-4). No intervening medium is needed to transmit the force between the particles or to store the tug's energy while in passage. If the gravitational attraction of a massive body is approximated as an 'external force', a so-called 'field of the gravitational potential' can be introduced which, mathematically, proves a great convenience, but this 'field' does not qualify as an independent entity because it transmits all effects instantaneously and its state is always entirely conditioned by the state of the originating particles. In electromagnetism, by contrast, one could know the 'initial condition' of all matter in the universe, yet be unable to determine how the universe will behave. The condition of the field apart from the matter must also be specified, for significant amounts of energy might be traversing the field in the form of electromagnetic waves.

A basic problem tends to plague all forms of particle-field duality. If the two entities subsist independently, the field created by a bit of matter ought to affect that bit of the matter itself according to the same laws as govern the coupling between the field and any other body. The reaction of matter to a field that it itself initiates is called the back action of the field. If the matter is concentrated within an isolated point, as often happens in classical theories, it will usually create too strong a field within its own locality, leading to infinite amounts of back action against itself.

A variety of strategies have been proposed to rectify this problem: regard the 'field' as merely a mathematical convenience - retain a 'monism' of particles only. A careful look at how gravitational 'fields' are introduced into 'point mass' mechanics reveals that no overall gravitational field is ever defined by standard arguments, but merely a set of so-called 'other fields' that codify the forces exerted upon a given mass by the other bodies in the universe. 'Other fields', by their very definition, exert no back action and they should not be regarded as truly independent of the matter. Another strategy would be to smear out the mass, charge, and so on, of the particle over a finite volume, so that the total back action of its component points never becomes infinite. 'Other fields' are thus avoided, although this kind of field will still transmit all effects instantaneously and cannot sustain gravitational waves. In the 1890s, H.A. Lorentz tried to resolve the difficulties of electromagnetic dualism in a 'smeared out' manner also. A mechanism for creating a resistant stress becomes necessary to prevent the finite back action from

quickly collapsing or inflating the particle. Suitable candidates for this ’stress’ are hard to find. A third strategy could be to remove the dualism by converting the ’particle’ into a singularity within the field - this supplies a ’monism’ of fields. Examples of the approach are (William Thomson) Lord Kelvin’s proposal that ordinary atoms represent the centres of vortices twisting in the ether and William Clifford’s suggestion that particles are geometrical singularities within a non-Euclidean space.

In electromagnetism, none of these alternatives works very well. Because of field energy storage, one cannot easily retreat to the simple ’monism’ permitted in the case of gravitation. An interesting effort at electromagnetic particle monism was proposed by Feynmann and Wheeler, based upon the work of Gustav Mie. Philosophers sometimes claim that this theory, despite its distinct ontology, is ’observationally equivalent’ to standard electromagnetism. In the author’s opinion, such claims do not survive scrutiny. In any case, since a measurable electromagnetic back action is experimentally detectable, the simple ’Lorentz law’ treatment of field-particle coupling described in most textbooks is too simple to be correct. The more refined classical treatment articulated by P.A.M. Dirac in the 1930s turns out to have very bizarre consequences with respect to causality and the like.

Related problems of back action continue to trouble quantum versions of field theory as well.

See also: Electrodynamics; Field theory, quantum

MARK WILSON

References and further reading

Hesse, M. (1965) *Forces and Fields*, Totowa, NJ: Littlefield, Adams.(Readable survey of the historical struggle between field and action-at-a-distance perspectives.)

Maxwell, J.C. (1873) *Treatise on Electricity and Magnetism*, Oxford: Clarendon Press, 2 vols.(The classic work.)

Parrott, S. (1987) *Relativistic Electrodynamics and Differential Geometry*, New York: Springer.(Detailed exploration of the problems of dualism in electromagnetism. Requires university level mathematics.)

Whittaker, E. (1951) *A History of the Theories of Aether and Electricity*, London: Nelson.(Despite its grudging attitude towards Einstein, still the best general history.)

Field theory, quantum

Quantum field theory extends the basic ideas of quantum mechanics for a fixed, finite number of particles to systems comprising fields and an unlimited, indefinite number of particles, providing a coherent blend of field-like and particle-like concepts. One can start from either field- or particle-like concepts, apply the methods of quantum mechanics, and arrive at the same theory. The result inherits all the puzzles of conventional quantum mechanics, such as measurement, superposition and quantum correlations; and it adds a new roster of conceptual difficulties. To mention three: the vacuum seems not really to be empty; the particle concept clashes with classical intuitions; and a method called 'renormalization' gets the best predictions in physics, apparently by dropping infinite terms.

1 Approaches to quantum field theory

There are three conceptually distinct ways of introducing quantum field theory. We grasp the theory best by taking note of all three.

Classical objects have degrees of freedom, independent ways they can move - a locomotive confined to its tracks has one degree of freedom, a baseball has three (ignoring spin). We start with the classical picture of a system of discrete particles, each with three degrees of freedom, each degree of freedom in turn described with a pair of (conjugate) variables, typically position and momentum (see Mechanics, classical). Conventional quantum mechanics 'quantizes' such a system by replacing each pair of variables with a pair of operators (satisfying a commutation relation) that transform vectors in a vector space. These vectors describe the quantum states, and a simple formalism involving the operators and a state gives the values for quantities in the state. Unlike a classical system, the states do not in general give exact values for quantities, but only probabilities for what values will be found if a measurement is made. At most, one of a pair of quantities, such as position and momentum, can take on an exact value in a given state (see Quantum mechanics, interpretation of).

The first approach to quantum field theory applies this procedure to a classical field conceived of as a system with an infinite number of degrees of freedom (see Field theory, classical). In classical field theory one associates a field value, such as a value of the electric field, with each point in space. Think of such a field value as a 'position' in an abstract space. A standard (Lagrangian) formalism associates a second 'momentum' variable with each of these 'position' variables. So the field has been described as a system with infinitely many degrees of freedom, one at each spatial point, and each described with a 'position' and a 'momentum' variable. To quantize, one replaces each pair of variables with a pair of operators, all acting on a giant vector space. As before, vectors represent states, in general giving only probabilities on measurement for observed field values. The classical field equation, describing the relation between classical field values at neighbouring places and times, generalizes to an operator equation constraining the probabilities for field values at neighbouring points.

So far we see the theory's connection with fields but not with particles: a second approach allows us to see the latter connection. Quantum descriptions of 'particles' depart from classical descriptions in never assigning exact spacetime trajectories and in recognizing no 'individuality' distinct from qualitative properties. To mark the differences we will talk about *quanta* instead of particles. To reveal the theory's commitment to quanta, start with a classically described field, but redescribe it as a sum of simple waves - called 'harmonic modes' - of various frequencies. Each mode has the form of a harmonic oscillator - something behaving like a pendulum or oscillating spring - so that the field has been redescribed formally as a collection of independent harmonic oscillators. Now apply the quantization procedure to each of the (formal) harmonic oscillators.

A quantized harmonic oscillator has evenly spaced, exact energy states, beginning with a lowest, inviting us to interpret them as a state with zero quanta, a state with one quantum, a state with two quanta, and so on, each quantum with the same momentum. Different momenta go with different oscillators so that any 'excitation' state of the collection of oscillators can be reinterpreted as a collection of various number of quanta with various momenta. So far the account describes only Bosons, quanta of which there can be any whole number in a given state; but with an adjustment in the operators used in quantizing the oscillators (changing commutators to anticommutators) one gets a description of Fermions, quanta of which there can be only zero or one quantum in a given state.

In this formulation the collection of possible quantum states, called 'Fock space', is described in terms of basis

states, written ‘ $| n_1, n_2, n_3, \ldots n_k\rangle$’, in which, for each n_i, there are n_i quanta described by some maximal list of properties, and each n_i a non-negative integer for Bosons, or zero or one for Fermions. Fock space includes weighted sums (superpositions) of these basis vectors, introducing an element not found in conventional quantum mechanics, states with an indefinite number of quanta and temporal change from states with one number of quanta into states with another number. Manipulation of the formalism is facilitated by special (non-Hermitian) *raising* and *lowering* operators which transform a Fock space state into one describing, respectively, one more or one less quantum. Often these are called ‘creation’ and ’annihilation’ operators, but since arguably they do not describe processes of creation and annihilation the alternative ‘raising’ and ‘lowering’ operator terminology is recommmended.

Approaching the theory via quantized harmonic oscillators yields a formulation which seems to talk about quanta. But a simple redescription (a transformation very like the conventional Fourier transform from the momentum to the position basis) restores a field-theoretic appearance. After redescription one has raising and lowering operators, $\Psi^\dagger(x)$ and $\Psi(x)$, indexed by a spatial variable. (To facilitate relativistic description practitioners generally use a spacetime variable.) Often one informally describes $\Psi^\dagger(x)$ and $\Psi(x)$ as ‘creating and annihilating quanta at *x*’. But the raising and lowering operators do not describe creation and annihilation events, and the quanta associated with such operators are concentrated around a point, *x*, but not strictly localized at it.

Given the operators indexed by spatial (or spacetime) variables, people often talk of an ‘operator valued field’, as if the theory specified values of a physical quantity to spatial points by providing operators indexed with a spatial variable. This is misleading: an operator does not correspond to a value of a quantity, but to the whole spectrum of values which the quantity can assume. One sees the field-theoretical aspect more accurately in the assignment of probabilities of values to spatial points or, better yet, in the connection with classical field theory through the first form of field quantization.

The two methods described above are called *field quantization*. An extension of the technique starts with a spacetime description of the state from conventional or *first quantized* quantum mechanics and applies the quantization procedure to this state function, or *second quantizes* it, treating it as if it were itself a classical field. Field quantization of the classical electromagnetic field results in a theory of photons. Second quantization of a first quantized theory of particles produces a theory of quanta such as electrons and protons.

A third approach to quantum field theory starts with the quanta themselves. One assumes a state to be completely described by specifying the number of quanta described by each possible maximal list of properties: one cannot get a new state description by switching around ‘which’ quanta have which properties. But these are just the basis states of the Fock space from the last approach. Taking the full Fock space to include the weighted sums (superpositions) of these basis states gives a theory including indefinite and variable number of quanta; and by generalizing on the assumption that the quanta individually obey conventional quantum mechanics, one gets the same theories generated by field quantization and its second quantization analogue.

2 Puzzles in quantum field theory

In the *vacuum*, the state with no quanta, one would expect all quantities to have a null value. But many have a positive expected average on measurement (expectation value). A second problem concerns any detector designed to register ‘no quanta’ when it is non-accelerating in the vacuum state. If one accelerates this detector through the vacuum it will react as if it were moving through a thermal bath of so-called *Rindler quanta*! So is the vacuum a state with no quanta or is it not?

Both these problems turn out to have the same formal basis as the fact from conventional quantum mechanics that a state which has an exact value for either position or momentum will have no exact value for the other quantity. The vacuum is exact for the quantity, *number of quanta*, and inexact for (complementary) field quantities. Consequently the vacuum definitely has zero quanta and has no actual values for field quantities, though this state has positive *probability* for finding non-null values for field quantities on measurement. Then field quantities such as energy, which have only positive values, can still have a positive average (expectation value) for the results of measurement. Formally analogous considerations apply to Rindler quanta. Interpreters disagree as to whether these considerations resolve the puzzle. There are further puzzles about the free vacuum in a relativistic theory, involving distant correlations which suggest that any quantal concept must, strictly speaking, be a global concept,

not a local one.

The presentation to this point has described only free quantum field theory, a theory of non-interacting quanta. But different kinds of quanta interact with one another: free quantum field theory is a severe idealization since, as physicists say, ‘we cannot turn off the interaction’. One transforms the free theory into interacting quantum field theory by adding the effects of interaction into the equations of motion, resulting in a theory vastly more complex and laden with difficulties. Calculation requires approximation methods which, if not to be interpreted instrumentalistically, introduce new interpretive problems.

A common method of approximation (perturbation expansion) results in divergent integrals arising from quanta interacting with themselves. An ingenious scheme, known as *renormalization*, allows the offending terms to be grouped systematically with terms describing the quanta’s mass and strength of interaction. By substituting the observed values of the mass and ‘coupling constants’ for these ‘renormalized’ terms the approximation method produces startlingly good predictions. This procedure has shocked many who gloss it as ‘discarding infinities’.

Physics texts make clear that renormalization involves no mathematical sleight of hand. The theory must break down at very high momenta, if only because not all interactions have been taken into account. In effect one substitutes observed values for masses and coupling constants where the theory breaks down, ensuring mathematical honesty by making the substitutions before taking the limits on the offending integrals. It is a little like using an observed value for Hooke’s spring constant where one could not calculate the constant from first principles. (While most endorse the foregoing, two of the subject’s giants, Dirac and Feynman, thought renormalization could not be so easily resolved.)

Many likewise accept resolution of puzzles surrounding *virtual quanta*. The terms of an approximation expansion can be presented by a graphic technique called *Feynman diagrams*. These diagrams depict a network of creation and annihilation of ‘virtual’ quanta thought of as ‘mediating’ the interactions, tempting people to think that the virtual quanta actually exist. But they exist only as elements of a gigantic sum (superposition), so that they exist as do component simple waves whose sum is a complicated actually occurring wave pattern.

Quantum field theory as used by working physicists is not a formally exact theory. There is a version, axiomatic quantum field theory, which achieves rigour at the expense of most practical application. Within axiomatic quantum field theory one can prove (Haag’s theorem) that in the interacting theory observed states of quanta are not related by intermediate states in the way usually assumed by the working theory, again suggesting that the theory recognizes only a global concept of quanta. Most working physicists simply dismiss these and related problems: after all (precise) axiomatic quantum field theory and (imprecise) working quantum field theory are not the same theory. But some maintain that Haag’s theorem signals important formal and interpretive problems.

Description of electrons involves an arbitrary element (the absolute phase) superficially similar to an arbitrary choice of unit of length. Setting this element arbitrarily at each point in space seems to force the assumption of a ‘compensating’ field, interpretable as the photon field, which describes electromagnetic forces. Generalization of this idea of *gauge theories* extends quantum field theory: internal degrees of freedom of quanta involve similar ‘arbitrary’ elements again appearing to require assumption of further quantized fields, interpretable as the agents of the weak and strong forces. The resulting theory, known as the *standard model*, provides our most detailed picture of the internal structure of matter with the electro-weak theory and quantum chromodynamics. But the quarks of chromodynamics apparently cannot exist in a free state. Is that a problem? The fact that deep inelastic scattering experiments are interpretable as extremely indirect detection of quarks probably means that such quark confinement should not worry us. But at the time of writing the nature of the argument for the existence of gauge fields stands to be better understood.

See also: Bell’s theorem; Quantum measurement problem

PAUL TELLER

References and further reading

Each of the following include extensive further bibliography.

Auyang, S.Y. (1995) *How is Quantum Field Theory Possible?*, New York: Oxford University Press.(A Kantian interpretation of contemporary field theories.)

Redhead, M.L.G. (1983) ‘Quantum Field Theory for Philosophers’, in P.D. Asquith and T. Nickles (eds) *PSA 1982*, East Lansing, MI: Philosophy of Science Association, vol. 2, 57-99.(A relatively accessible introduction to some of the basic ideas and interpretive problems of quantum field theory.)

Redhead, M.L.G. (1988) ‘A Philosopher Looks at Quantum Field Theory’, in H. Brown and R. Harre (eds) *Philosophical Foundations of Quantum Field Theory*, Oxford: Clarendon Press.(Review and elaboration on material in the 1983 article.)

Teller, P. (1995) *An Interpretive Introduction to Quantum Field Theory*, Princeton, NJ: Princeton University Press.(A general introduction to the elementary formalism and interpretive problems of quantum field theory, accessible to anyone who knows a little quantum mechanics.)

Film, aesthetics of

Film aesthetics has been dominated by issues of realism. Three kinds of realism attributable to film may be distinguished: (1) the realism inherent in film because of its use of the photographic method (realism of method); (2) realism as a style which approximates the normal conditions of perception (realism of style); (3) realism as the capacity of film to engender in the viewer an illusion of the reality and presentness of fictional characters and events (realism of effect). Some theorists have argued that realism of method requires us to avoid realist style, others that it requires us to adopt realist style. Most have agreed that realist style makes for realism of effect; they disagree about whether this is a desirable goal. It is argued here that these realisms are independent of one another, that realism of style does not entail any kind of metaphysical realism, and that realism of effect is irrelevant to understanding the normal experience of cinema. Realism of style suggests a way of making precise the claim that cinema is an art of time and of space, because this kind of realism is partially explicated in terms of the representation of time by time and of space by space. Psychological theorizing about the cinema has been strongly connected with realism of effect, and with the idea that an illusion of the film's reality is created by the identification of the viewer's position with that of the camera. Another version of illusionism has it that the experience of film-watching is significantly similar to that of dreaming. Such doctrines are undermined when we acknowledge that realism of effect is an insignificant phenomenon.

1 History

The first period of serious writing on film aesthetics (roughly from the 1920s to the 1950s) was devoted to the problems raised by the photographic method, and to the accusation that film was merely the automatic recording of reality. The second period, still under way, has been dominated by attempts to elaborate a theory about the role of the camera as spectator, participant and the object or vehicle of the viewer's identification. But issues of realism have continued to be influential.

Two opposing tendencies were evident in the first period. One sought to show that what was distinctive about cinema, and therefore to be encouraged, was the means, notably editing, by which the product deviated from a mere recording of the real world (realism of method). This view was often combined with an opposition to the use of integrated sound, on the grounds that its use resulted in a hybridized and weakened medium. In film practice this tendency was evident in the montage style of the early Soviet cinema, which emphasized the juxtaposition of objects and events by quick cuts, close-ups and striking camera angles. Thus realism of method was made the grounds for preferring an unrealistic style. The contrary and somewhat later tendency was to celebrate the dependence of film on the process of automatic recording, to argue that the medium of film is reality itself, and that in consequence developments like sound and colour are to be seen as fulfilling the historical destiny of cinema because they are additions to the realism of film. Central to this view was an endorsement of so-called 'long-take, deep-focus' style, thought to provide a visual experience approximating to our visual experience of the real world, and enabling different actions to take place within the same frame - a capacity later enhanced by the introduction of wide screen. Thus realism of style was justified by an appeal to realism of method.

Both tendencies exhibit a high degree of concern for prescriptive issues about film-making. The writing that initiated the second period during the 1960s rejected this approach and sought a new, ostensibly more scientific attitude to cinema through the articulation of a supposed language of film, which would enable us to analyse film technique and the viewer's strategies for understanding film. But this approach soon abandoned its structuralist and quasi-scientific impetus in favour of a psychologically oriented approach that sought to analyse the way that film, especially the mimetic cinema of Hollywood, engages the viewer, and to characterize the experience of engagement as a species of entrapment. Thus the second period has generally been hostile to the realist tendency of the first period. Film theory in this second period has had a distinctly political agenda, drawing on Marxism and psychoanalysis to explore the role of film in reinforcing the viewer's subjective identity. Feminist theorists have argued that conventional film-making is geared to the satisfaction of male voyeuristic desire (Mulvay 1976).

2 Realism of method and of style

The idea behind realism of method seems to be that the photographic medium enables us to see things themselves rather than representations of them; in this respect photographs are said to be akin to lenses and mirrors, in contrast

to paintings (and, presumably, animated cartoons) which give us mere representations of things see Photography, aesthetics of). Even assuming that realism of method is correct, no theorist has adequately explained how this purely descriptive claim entails the evaluative claim that long-take style is preferable to montage style. And those who sought to justify realist style in terms of realism of method overlooked a crucial distinction between what the camera literally records - actors, props and sets - and the fictional objects and events it presents. Realism of method applies only to the first of these, whereas realism of style is primarily a vehicle for the presentation of the second. Further, if realism of method unequivocally favours realist style, we must endorse what are generally agreed to be rather problematic developments: 3D, smellorama, and the ultra-long-take method of Hitchcock's *Rope*.

Some writers have denied that long-take, deep-focus style is realistic. In fact, a defence of the realism of that style is possible. Let us say that a mode of representation is realistic when, or to the degree that, we employ the same capacities in recognizing its representational content that we employ in recognizing the (kind of) objects its represents. A good-quality, well focused, middle-distance photograph of a horse is realistic in this sense: we employ our capacity to recognize horses visually so as to determine that this is, indeed, a photograph of a horse. Many paintings would count as realistic by the same criterion. A linguistic description of a horse, by contrast, is not realistic, for the capacity to recognize horses visually is not sufficient to enable you to recognize this as a representation of a horse; that requires a knowledge of the conventions of language. (There may be other senses in which the description is realistic.) Realism of style is a matter of degree; some aspects of the content of a representation may be recognized by deploying the capacity for object-recognition, while others are not.

By this criterion, long-take, deep-focus style is (relatively) realistic. Watching a film in this style, we judge the spatial and temporal relations between the objects (and their parts) and the events that the image represents by using the capacity visually to judge spatial and temporal relations between real things. We judge the spatial relations between objects visible in the same frame by seeing that they are spatially related thus and so within the visual field; we judge the temporal properties of and relations between events within the take by noting that this event took (roughly) so long to observe, while this one was experienced as occurring later than or earlier than that one. That is just how we perceive the spatial and temporal properties of things and events in the real world. With montage style, by contrast, where there is quick cutting between very distinct spatial (and sometimes temporal) perspectives, these properties and relations have to be judged, with greater frequency, by means of inference from the overall dramatic structure of the film.

It has been said that deep-focus style is unrealistic in that it presents us with an image in which objects at considerably different distances from the camera are simultaneously in sharp focus, whereas objects at comparable distances from the eye could not be seen in focus together (Ogle 1972). This does not seriously detract from the realism of deep focus. Deep focus, particularly when used in conjunction with wide screen, enables us to concentrate our attention on one object, and then to shift our attention at will to another object, just as we are able to do when perceiving the real world. Since we are usually not very conscious of refocusing our eyes, the similarities between viewing deep-focus style and perceiving the real world are more striking than the differences. With montage style, on the other hand, we are severely limited, by shot length and depth of field, in our capacity to shift our attention from one object to another at will - though this feature is not entirely absent in montage style.

Explicating the idea of realism of style in this way helps us to avoid an error that has dogged theorizing about the cinema: that realism in film can be attacked on metaphysical grounds because it postulates a real, observer-independent world, an idea that some theorists then further associate with a politically conservative agenda of submission to prevailing conditions. But realism of style as I have explicated it here appeals to no such postulate of an observer-independent world (though one might argue that such a postulate is both philosophically respectable and politically neutral). The claim of stylistic realism is not the claim that cinema presents objects and events isomorphic to those that exist in an observer-independent world. It is the claim that, in crucial respects, the experience of film-watching is similar to our ordinary perceptual experience of the world, irrespective of whether and to what extent that world is independent of our experience of it.

3 Space, time and film

We can now see an important sense in which film is both a spatial and a temporal medium. Film represents space

by means of space, and time by means of time. It is the spatial (temporal) properties of the cinematic representation that we observe and rely upon in order to figure out what spatial (temporal) properties of the fictional characters and events are portrayed. It is correctly said that painting and still photography are capable of representing the temporal: by inference, by juxtaposition of distinct static images, by transforming temporal properties into spatial ones (where, say, being further to the right represents being later in time), and by special techniques such as blurring and multiple exposure. But these possibilities do not constitute grounds for calling painting and still photography arts of time in the way that cinema is, for they do not represent time by means of time.

So far we have considered untensed temporal properties of duration and precedence. What about the representation of tensed temporal properties: pastness, presentness and futurity? Theorists have often argued that film represents fictional events as occurring now, in the sense that viewers are to think of those events going on in the present as they watch. This is one consequence of the doctrine of realism of effect, according to which film typically creates an illusion in viewers' minds that the fictional events represented on screen are real, and that they, the viewers, are present (spatially and temporally) at their occurrence. Realism of effect will be discussed in the next section; here we concentrate on the claim of temporal presentness. One problem with the view that film represents fictional events as occurring in the viewer's present is that it then becomes difficult to make sense of the idea of 'anachrony' in film - the flashback or flashforward. If viewers are to think of the image they see as representing something occurring now, an anachronous sequence would require them to imagine travelling in time, viewing what is objectively past or future in their subjective present. But this is neither plausible psychologically, since viewers of anachronous material do not seem to imagine themselves travellers in time, nor helpful in maintaining the integrity of the action, since time travel introduces an element of fantasy that would be unwelcome in our experience of many naturalistic films that none the less display anachrony in their narratives.

If, on the other hand, we say that film does not represent fictional events as present to the viewer, but merely as standing in untensed relations of precedence and simultaneity to one another, how shall we explain anachrony? After all, anachronous representations seem to be ones that represent the past (as in flashback) and the future (as in flashforward). In fact, we can explain anachrony in untensed terms as follows: we have anachrony when an event occurring earlier in fictional time than another occurs later in the time of narrative exposition; here we appeal only to untensed notions of precedence. Such a tenseless approach implies a certain relativity and even arbitrariness in the application of the notions of flashback and flashforward; a film's narrative structure, or part of it, could sometimes equally well be described as employing the one as the other. If, on the other hand, we could appeal to tense and identify non-anachronous material as temporally present, we would have an absolute distinction between flashback (past) and flashforward (future). But this relativity in the tenseless account of anachrony is arguably no drawback. The fact that one of the two labels (flashforward, flashback) often seems more natural than the other may be accounted for by appeal to the fact that one of them allows a simpler, and therefore preferable, description of the film's temporal structure. It may also be that one or other description accords better with the dramatic structure of the film at that point, which may involve an episode of memory (in which case the description in terms of flashback will seem preferable) or premonition (which favours flashforward). There will, on this view, be occasions when there is nothing to choose between the two descriptions; and indeed there are films (*Last Year at Marienbad*, for example) of which it seems right to say that there is no uniquely correct description of their temporal structure. In that case, while anachrony itself is a notion definable in temporal but tenseless terms, the direction of anachrony needs to be defined in pragmatic and dramatic terms.

4 The camera, the eye and realism of effect

Many theorists have argued that the mechanism or 'apparatus' of cinema encourages and perhaps requires the viewer to think of cinematic images as corresponding to the perceptual states of an observer whose eye is the lens of the camera itself (Aumont 1989: 2). This idea is reinforced by the doctrine of realism of effect, in the following way. If viewers are victims of an illusion created by the film, then presumably they must think that the reality supposedly seen is seen from their point of view, by means of their visual organs. Since film manifestly works by presenting events from another, independent position - that of the camera - illusion-dominated viewers must come to think of themselves as occupying that position.

There are two objections to the idea that film induces the illusion that fictional events are real and that the viewer

is directly witnessing them. The first is that there is little evidence that film typically creates, even temporarily or in part, the false beliefs necessary to sustain such an illusion; film watchers do not behave like people who believe, or even suspect, that they are in the presence of axe murderers, world-destroying monsters or nuclear explosions. Second, this theory is at odds with much of the experience of film-watching; identification with the camera would frequently require us to imagine ourselves in peculiar or impossible locations, undertaking movements out of keeping with the natural limitations of our bodies, and peculiarly invisible to the characters. None of this seems to be part of the ordinary experience of film-watching. In the attempt to identify the camera with some observer within the world of the action with whom the viewer can in turn identify, theorists have exaggerated the extent to which shots within a film can be thought of as point-of-view shots, and have sometimes postulated, quite ad hoc, an invisible narrator from whose position the action is displayed. It would be better to acknowledge that cinematic shots are only rarely from a psychological point of view, and abandon the thesis that the viewer identifies with an intelligence whose point of view is the camera.

A variant of illusionism says that the situation of the film watcher approximates to that of a dreamer, in that both situations present us with a strong impression of the reality of that which is actually unreal (Metz 1970). In that case the camera would correspond to a supposed ‘inner eye’ by means of which we perceive the images of dreams. This analogy has been a powerful stimulus to the development of psychoanalytic theories of film and film experience. In fact the analogy with dreaming fails to compare like with like. Dreamers, like film watchers, are usually physically passive while watching or dreaming. But the experience of dreaming is usually one that involves action - sometimes ineffectual - on the dreamer’s part, while the experience of film-watching, our reflex responses aside, rarely involves physical action. And it is the experience of film-watching and the experience of dreaming that are claimed by the advocates of the dream/film analogy to be alike. In dreams, our own actions and sufferings are of central concern to us; the experience of film-watching makes us largely forgetful of ourselves while we concentrate on the fate of the characters.

See also: Semiotics

GREGORY CURRIE

References and further reading

Arnheim, R. (1958) *Film as Art*, London: Faber & Faber.(A translation of essays written in German between 1933 and 1938. An antirealist statement from the first period.)

Aumont, J. (1989) ‘The Point of View’, trans. A. Denner, *Quarterly Review of Film and Video* 11: 1-22.(On the supposed subjectivity of the camera.)

Bazin, A. (1967) *What is Cinema?*, vol. 1, trans. H. Gray, Berkeley, CA, and Los Angeles, CA: University of California Press. A translation of selected essays from *Qu-est-ce que le cinéma?*, 4 vols, Paris: Éditions du Cerf, 1958-65.(The classic statement of realism.)

Bordwell, D. and Carroll, N. (eds) (1995) *Post-Theory: Reconstructing Film Studies*, Madison, WI: University of Wisconsin Press.(A collection of writings by philosophers and theorists opposed to the dominant models of the second period.)

Carroll, N. (1988) *Philosophical Problems of Classical Film Theory*, Princeton, NJ: Princeton University Press.(A useful discussion of expressionist and realist tendencies.)

(1990) *Mystifying Movies*, Ithaca, NY: Cornell University Press.(An attack on the film theory of the second period from the perspective of contemporary Anglo-American philosophy.)

Cavell, S. (1971) *The World Viewed*, New York: Viking Press.(A statement of the realist view.)

Currie, G. (1995) *Image and Mind. Film, Philosophy, and Cognitive Science*, New York: Cambridge University Press.(A theory of film which rejects semiotic and psychoanalytic models in favour of Anglo-American philosophy of mind and cognitive science.)

Metz, C. (1977) *Le signifiant imaginaire. Psychoanalyse et cinéma*, Paris: Union Générale d’Éditions; trans. C. Britton *et al.*, *The Imaginary Signifier*, Bloomington, IN: Indiana University Press, 1982.(Represents a semiotician’s turn to psychoanalytic and especially Lacanian ideas.)

Mulvay, L. (1976) ‘Visual Pleasure and Narrative Cinema’, *Screen* 16, reprinted in G. Mast, M. Cohen, and L. Braudy (eds).(An account of the viewer’s relation to the screen image, influenced byLacan.)

Nichols, B. (ed.) (1976, 1985) *Movies and Methods*, 2 vols, Berkeley and Los Angeles, CA: University of California Press.(An anthology in which Marxist, semiotic and feminist perspectives are well represented.)

Ogle, P. (1972) ‘Technological and Aesthetic Influences on the Development of Deep-Focus Cinematography in

the United States', *Screen* 13, reprinted in Nichols (ed.), vol. 2.(A historical account of deep-focus style.)
Wilson, G. (1986) *Narration in Light*, Baltimore, MD, and London: Johns Hopkins University Press.(This work by a philosopher develops an account of narration in film through close analysis of five films.)

Filmer, Sir Robert (1588-1653)

Filmer was one of the most important political thinkers in seventeenth-century England, and the author of Patriarcha. Locke replied to this and other works by Filmer in the Two Treatises of Government - perhaps the most famous of all works of liberal political theory. Filmer argued that notions of mixed or limited government were false and pernicious, and that the powers of all legitimate rulers were derived not from the people but directly from God, to whom alone rulers were accountable. Filmer's contemporaries commonly held that the authority of a father and husband over his family stemmed not from the consent of his wife and children but from the natural and divinely appointed order of things. Filmer harnessed such ideas to the cause of royal absolutism by arguing that the state and the family were essentially the same institution.

1 Life and works

Filmer was born in 1588 into a large and wealthy Kentish family. He was an eldest son, and upheld political beliefs which took it for granted that primogeniture in the male line was the proper means of succession in families and states. After attending Trinity College, Cambridge, and Lincoln's Inn, in 1613 he was called to the bar. He married a bishop's daughter and the couple settled in lodgings at Westminster Abbey. In 1619 Robert was knighted. A decade later he inherited his father's estate in Kent.

Sir Robert had connections with the royal court (where a brother held office) and with the upper ranks of the church (through his wife and through his friend Peter Heylin, the chaplain and biographer of William Laud, Archbishop of Canterbury). The ideological stance which Filmer adopted in his writings was strongly influenced by attitudes common among courtiers and high-ranking clerics. It is unclear when he began to write, but not long before 8 February 1632 he brought to the relevant royal official 'a Discourse to bee licensed for printing, written of Government and in praise of Royaltie and the supreme authority thereof'. The official asked the king himself whether it would be expedient to publish the book, and on 8 February Charles I examined the work and decided to forbid publication. The book was almost certainly *Patriarcha. The Naturall Power of Kinges Defended Against the Unnatural Liberty of the People*: the text of one manuscript of *Patriarcha* is dateable on internal evidence to between 1628 and 1631, and it is plausible that parts of the work are from much earlier still. Charles refused to license Filmer's treatise for the press, probably because he feared that it would foment unnecessary strife as its political doctrines were so uncompromising and so trenchantly expressed. Filmer later revised the book and drew on it in his published pamphlets, but *Patriarcha* itself was printed only in 1680.

In the Civil War between Charles and parliament (1642-6), Filmer's eldest son left Kent and sided actively with the king, but the ageing Sir Robert stayed at home. Kent quickly fell into the hands of parliament, and for a while Filmer was imprisoned. In 1648 the leading royalist publisher Richard Royston brought out two pamphlets by Filmer, *The Free-holders Grand Inquest Touching Our Soveraigne Lord the King and His Parliament* and *The Anarchy of a Limited or Mixed Monarchy* (Filmer's authorship of *The Free-holders* has been challenged but is generally accepted). Royston also published another work, *The Necessity of the Absolute Power of All Kings: and in particular, of the King of England*, compiled by Filmer from the writings of Bodin (§§2-3). Internal evidence suggests that the two former texts were written around 1644. Until Charles I's final defeat in 1646, most royalist pamphlets attempted to win support for the king by stressing the moderation of his cause, but Filmer's writings were outspokenly absolutist; this may explain why they (like other trenchantly absolutist writings such as Hobbes' *Elements of Law*) were not published until later (see Absolutism). Two final political writings by Filmer appeared in 1652, the year before his death. One, *Observations Concerning the Originall of Government*, criticized the ideas of Hobbes (§§6-8), Milton and Grotius. The other, *Observations upon Aristotles Politiques, Touching Forms of Government*, included an appendix, 'Directions for Obedience to Governours in Dangerous and Doubtfull Times', in which Filmer argued that limited obedience to the usurping government of the Rump parliament was justifiable, but that no usurpation - however lengthy - could extinguish the right of the true rulers (contradicting his earlier views on usurpation). Some other brief political writings have been attributed to Filmer, but on insufficient grounds. Filmer did, however, write short published works on witchcraft and usury, and a manuscript essay 'In Praise of the Vertuous Wife'.

2 Political theory

Filmer's political writings were intended to refute ideas of legitimate resistance and limited or mixed monarchy which circulated among critics of royal policy in the early seventeenth century (these included the Catholics Suarez (§4) and Bellarmine, and a number of parliamentarian pamphleteers). The cardinal error of most political theorists, Filmer argued, was their supposition that people were born free and equal. From this false premise they drew the erroneous conclusions that states had originally arisen by the consent of their members, that political power had at first grown out of the community as a whole (for if all were equal, no one had any more right to exercise such power than anyone else), and that the powers of rulers were therefore derived from the community, which could discipline and perhaps even depose them if they failed to abide by the conditions upon which authority had been transferred to them.

Filmer claimed that this whole line of reasoning was faulty since people were in fact born into subjection to their fathers. He argued that Adam (the first father of all) had held authority over his children and their descendants, and over all property in the world. Adam's authority arose from his position as the ultimate ancestor of the group. Filmer claimed that Adam's power was fully political, and that it included the right to execute his subjects. By nature, he held, children remained subject to their fathers throughout their lives, although a father could decide to free them from subjection. If a son were freed in this way, he would then be able to wield full fatherly/political power over his own descendants. The ordinary means of succession to a father's power and property was by primogeniture in the male line, but a ruler could alter this arrangement (Noah, for example, divided the earth among his sons). Sovereigns could also hand over authority to someone else, and God might providentially intervene in human affairs by changing the ruling family or even the form of government (although monarchy was the best form). Filmer did not argue that the whole world should be ruled by Adam's heir (whoever that might be), for he recognized that Adam's original empire had been split into a large number of states. His central contention was that the rulers in all of these states exercised the same powers that Adam had - powers which were sovereign and independent of the consent of the people.

Filmer frequently quoted scripture (particularly the book of Genesis), the Civil Law, Aristotle and Bodin. His patriarchal theory of the origins and nature of government is, however, not easy to find in these sources. Bodin did indeed assert that by nature fathers have the power of life and death over their wives and children, but the Adamite elements of Filmerian patriarchalism are not present in his writings. They are present in the works of some English authors, notably Hadrian Saravia, who put forward views which are virtually indistinguishable from Filmer's in the 1593 Latin text, *De imperandi authoritate*. Like Filmer, Saravia argued that the family and the state were not merely analogous but identical institutions. Filmer, throughout his political writings, stressed that sovereignty in every state must be unlimited (except by the laws of God and nature) and indivisible, ideas which he took unaltered from Bodin (see Sovereignty §1).

Arguably Filmer was at his most effective and most original in his criticisms of theories of original popular sovereignty and the contractual origins of political power. He argued that contract theory was unconvincing on both historical and philosophical grounds and claimed that contract theorists themselves inconsistently adopted patriarchalist principles - for example, to argue that women and children were bound by political arrangements to which they had not consented (see Contractarianism).

3 Posthumous debate on his ideas

In 1680 *Patriarcha* was published as a contribution to the debate over the nature and extent of royal power which accompanied the Exclusion Crisis. Filmer's other political works had been reissued in the previous year. His writings soon attracted responses from James Tyrrell, Algernon Sidney and John Locke. For many years it was almost universally accepted that Locke had refuted Filmer's theory, successfully exposing his opponent's absurdities (Locke 1690). More recently, some feminist scholars have argued that this verdict is only partially justified and that Locke (and other liberal theorists) did not adequately meet Filmer's criticisms of contract theory, since they tacitly and inconsistently accepted patriarchalist assumptions.

JOHANN P. SOMMERVILLE

List of works

Filmer, Sir R. (1648a) *The Free-holders Grand Inquest Touching Our Soveraigne Lord the King and His Parliament*, London: Royston.(A long historical account of English constitutional development, arguing that

parliament is subordinate to the king.)
— (1648b) *The Anarchy of a Limited or Mixed Monarchy*, London: Royston.(Attacks the claim of the parliamentarian pamphleteer, Philip Hunton, and others that England is a mixed and limited monarchy by asserting that such a monarchy is impossible.)
— (1648c) *The Necessity of the Absolute Power of All Kings: and in particular, of the King of England*, London: Royston; compiled from J. Bodin, *The Six Bookes of a Commonweale*, trans. R Knolles, London: Bishop, 1606. (A collection of extracts from Bodin on absolute power and sovereignty.)
— (1652a) *Observations Concerning the Originall of Government*, London: Royston.(Attacks Hobbes, Grotius and Milton. Agrees with Hobbes on the nature of sovereign power, but denies that it is rooted in popular consent.)
— (1652b) *Observations upon Aristotles Politiques, Touching Forms of Government*, London: Royston.(Argues that Aristotle's *Politics* supports absolute monarchy.)
— (*c*.1628-31) *Patriarcha. The Naturall Power of Kinges Defended Against the Unnatural Liberty of the People*, London: Davis, 1680.(Presents the fullest version of Filmer's theory. First published a s Tory propaganda long after his death.)
— (1991) *Patriarcha and Other Writings*, ed. J.P. Sommerville, Cambridge: Cambridge University Press.(Includes the writings listed above, edited from the manuscripts and original editions.)

References and further reading

Daly, J. (1979) *Sir Robert Filmer and English Political Thought*, Toronto, Ont.: Toronto University Press.(The fullest account, but overestimates Filmer's originality.)
Saravia, H. (1593) *De imperandi authorite, et Christiana obedientia (On the Authority to Command, and Christian Obedience)*, London: Barker.(A vigorously absolutist work on the origins and nature of political power.)
Laslett, P. (1949) *Patriarcha and Other Political Works of Sir Robert Filmer*, Oxford: Blackwell.(Informative but a little dated.)
Locke, J. (1690) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1960. (Replies in detail to Filmer's theory, especially in the first treatise.)
Pateman, C. (1988) *The Sexual Contract*, Stanford, CT: Stanford University Press.(Includes discussion of the theories of Filmer and his critics from a feminist perspective.)
Schochet, G.J. (1975) *Patriarchalism and Political Thought*, Oxford: Blackwell.(Contains much important background information and two useful chapters on Filmer.)
Sommerville, J.P. (1986) *Politics and Ideology in England, 1603-1640*, Harlow: Longman.(Discusses the political and ideological context of *Patriarcha*. Chapter 1 is especially relevant.)

Florenskii, Pavel Aleksandrovich (1882-1937)

A figure of genius in the history of twentieth-century Russian religious philosophy, Florenskii did much to influence the directions of subsequent Russian thought, both within the Soviet Union and abroad in the Russian diaspora. Florenskii's originality is most noticeable in his chief philosophical work, Stolp i utverzhdenie istiny (The Pillar and Foundation of Truth) (1914), a somewhat eclectic and romantic work in which he sets forth his basic tenets in epistemology and sophiology. The work, which was his doctoral dissertation, represents a decisive rejection of rationalist and Western-orientated religious philosophy and theology in favour of a more concrete and experiential methodology.

1 Life

Florenskii was born in 1882 in Yevlakh, Azerbaijan, to a Russian father and an Armenian mother, both of a liberal secular orientation indifferent to religion. His own personal journey to devout religious practice, which culminated in theological studies (1904-8) and ordination to the Orthodox priesthood (1911), came only gradually and began, as he states in his memoirs, with an early fascination with the aesthetic form of nature that matured in a lifelong interest in natural science and mathematics. Declining an offer to pursue advanced studies in mathematics upon graduation from Moscow University (1904), where he studied under the noted mathematician N.V. Bugaev, Florenskii, to the dismay of family and friends, chose to enrol at the Moscow Theological Academy whose faculty he himself was to join (1908).

Bolshevik repression after the Russian Revolution brought about a change of direction in both his professional activities and his philosophical thought. With the forced closure of the Moscow Theological Academy, Florenskii was, owing to his scientific talents, still able to collaborate in various Soviet undertakings, including the Commission for the Electrification of Soviet Russia. His 1927 invention of a noncoagulating machine oil was even called 'dekanite' by the Soviets, in commemoration of the tenth anniversary of the Bolshevik Revolution. It was during this time that Florenskii shifted his philosophical interest to aesthetics and the philosophy of language. Refusing to renounce the priesthood, he was exiled in 1933 to Siberia and then in 1934 to the Solovki Island concentration camp in the White Sea, where he was executed by firing squad in 1937, contrary to official Soviet data which for decades listed the year of his death as 1943.

2 Theodicy: epistemology and sophiology

The first period of Florenskii's philosophical development reaches its climax in his already noted master-work, which, even according to his critic G. Florovskii, is the most characteristic work of the religious renaissance during Russia's short-lived Silver Age at the beginning of this century (see Russian Religious-Philosophical Renaissance). Florenskii subtitled his study 'The Experience of Orthodox Theodicy in Twelve Letters', indicating the root motivation inspiring the development of his reflections. One might say that *Stolp i utverzhdenie istiny* constitutes a form of intellectual autobiography in which Florenskii's own inner struggle for truth is articulated and justified. Indeed, the justification of the mind's claims to know truth lies at the core of Florenskii's own unique understanding of theodicy. After an initial discussion of his point of departure - concrete experience - he considers the divisions characteristic of a fallen world (along with the partial truths about the world) as facts crying out for Ultimate Truth that can undergird reality, imparting integrality to it.

Postulating 'living religious experience as the sole legitimate method for understanding dogma', Florenskii seeks to rejuvenate intellectual thought beyond confessional Orthodox lines towards a true existential appreciation of the whole of reality. Orthodox theologians as diverse as S.N. Bulgakov, J. Meyendorff and A. Schmemann have all followed his lead on this point. His stress on the immediateness of concrete experience also influenced other Russian philosophers like A.F. Losev, whose own work has similarities to Husserlian phenomenology.

Florenskii's aim is to foster dialectical thinking, which he conceives as living, unmediated thought unlike the 'schooled', rationalistic variety that prescinds from concrete experience. His emphases in cognition are an offshoot of the Slavophile theory of integral knowledge, first developed by I.V. Kireevskii and A.S. Khomiakov, then continued by V.S. Solov'ëv (see Slavophilism §1; Solov'ëv, V.S. §1). The whole thrust of Florenskii's epistemological query is to shed light on the experience of truth as gained in lived contact with reality. He

approaches the topic from various angles, finally defining truth as an ‘intuition-discursion’ in which the immediate givenness of truth is not blunt or blind, but is instead infused with intelligibility and *ratio*. Florenskii in this fashion offers a dynamic understanding of the principle of identity, making ’otherness’ constitutive of identity through a process of adoption (*usvoenie*) and assimilation (*upodobenie sebe*), thus grounding it in the principle of sufficient reason. Comparative analysis reveals similar (although subsequent and independent) insights on these themes in the work of such diverse and seminal thinkers as Heidegger and Whitehead.

Florenskii’s epistemological study of the experience of truth sets the stage for his metaphysical worldview, which, borrowing a term from the history of patristic theology, he fashions as *homoousian* philosophy or a philosophy of consubstantiality in opposition to a *homoiousian* one of mere similarity. The whole thrust of Florenskii’s metaphysics is to lay out the fundamental truth of pan-unity (*vseedinstvo*) at the root of all being. All created beings are not atomized units in isolation, but are consubstantial with one another. Cognition thereby enjoys an ontological moment in which knower and known are truly united in opposition to all representationalist theories of knowledge.

Florenskii transcribes his metaphysics into a sophiological key, holding all being to be nothing but a symbol of Sophia, which, for him, is an all-embracing reality linking both Creator and creature together and is considered variously as the ‘great root of the total creature’, the ‘guardian angel of creation’ and the ‘eternal spouse of the Word of God’, among other designations. As an attempt to address the ontological problem of the relations obtaining between Creator and creature, Florenskii’s sophiology, however, raises a host of problems on its own. In particular, critics (for instance N.O. Lossky and J. Meyendorff) have charged him with both Gnosticism and pantheism, since his conception of Sophia as a ‘fourth hypostatic element’ seems to compromise God’s freedom in creation as well as efface the ontological difference between God and creation (see Gnosticism; Pantheism). Others (for instance Slesinski) more sympathetic to Florenskii suggest that a more careful analysis of love as a central category of his metaphysics of consubstantiality can obviate these difficulties.

3 Anthropodicy: aesthetics and philosophy of language

If the major concern of *Stolp i utverzhdenie istiny* is the justification of the claims of truth, Florenskii makes the specific truth about the human person the focus of his later works, which he dedicates to concrete metaphysics or ’anthropodicy’. The centrality of art and language for the meaning of the human person is the main theme of this endeavour, the human person and the world being expressions of the spiritual in the sensual or empirical order of being. Being for Florenskii is no longer Sophia, but rather ‘symbol’ or ‘icon’, ontologically conceived as a dynamic interchange between the knower and symbolized reality, visually in the case of art and verbally in that of language, wherein the symbol is not a sign of an absent reality, but its very presence. Florenskii defines symbol as ‘being which is greater than itself’, and thus understands it as a manifestation or energy of being without exhausting its essence.

An heir to the linguistic school of W. von Humboldt, Florenskii centred his studies on the dynamic aspect of language as a pining of the spirit to express itself and, accordingly, was sympathetic to avant-garde poetic theory. From his university years, he enjoyed a close friendship with Russian Symbolist poets, A. Belyi in particular. There were also theological roots to Florenskii’s philosophy of language. The controversy (1912) over the ‘glorification of the name’ (*imiaslavie*) among Athonite monks regarding the immediate presence of God in his name had a decisive bearing on the development of Florenskii’s thought. Theological concerns also motivated his studies in aesthetics and the philosophy of cult in which the meaning of iconography and the significance of liturgical action are primary themes.

The chief works in which he explores these ideas are *U vodorazdelov mysli (At the Watersheds of Thought)* (1990), *Ikonostas (Iconostasis)* (1972), and a series of lectures on the philosophy of cult now published under the heading of *Iz bogoslovskogo nalsediia sviashchennika Pavla Florenskogo (From the Theological Heritage of the Priest Pavel Florenskii)*. Not published during Florenskii’s lifetime because of the political climate, these studies largely lay dormant even after his posthumous rehabilitation (1956). *Glasnost’* and the fall of communism reversed this situation. A complete publication of all his works remains to be achieved.

ROBERT SLESINSKI

List of works

Florenskii, P.A. (1914) *Stolp i utverzhdenie istiny (The Pillar and Foundation of Truth),* Moscow: Izdatel'stvo 'Pravda', 1990, 2 vols; trans. B. Jakim, *The Pillar and Ground of the Truth*, Princeton, NJ: Princeton University Press, 1997.(Florenskii's best-known work, detailing his epistemology and sophiology.)

Florenskii, P.A. (1915) *Smysl idealizma (The Meaning of Idealism)*, Sergiev Posad. (Provides Florenskii's original reading of Plato.)

Florenskii, P.A. (1922) *Mnimosti v geometrii (The Imaginary in Geometry),* Moscow: Pomor'e; repr. Munich: Otto Sagner, 1985.(Florenskii offers a defence of the Ptolemaic cosmological worldview of Dante's *Divine Comedy* in the last section of this work.)

Florenskii, P.A. (1972) *Ikonostas (Iconostasis)*, in *Bogoslovskie trudy (Theological Studies)* 9: 80-148; integral trans., Crestwood, NY: St Vladimir's Seminary Press, 1996.(A key work in the aesthetics of iconography. The previously unpublished manuscript dates to 1922.)

Florenskii, P.A. (1977) *Iz bogoslovskogo naslediia sviashchennika Pavla Florenskogo (From the Theological Heritage of the Priest Pavel Florenskii)* in *Bogoslovskie trudy (Theological studies)* 17: 85-248.(Written from 1918 to 1922, this work treats the philosophy of cult.)

Florenskii, P.A. (1990) *U vodorazdelov mysli (At the Watersheds of Thought),* Moscow: Izdatel'stvo 'Pravda'. (The original manuscript, itself incomplete, was written over 1917-22, and comprises various studies on the philosophy of language. It includes his noted monograph on inverse perspective. Another volume with the same title was published by YMCA-Press, Paris, 1985, but contains only articles on art, including his monographs on inverse perspective and the *Ikonostas*.)

Florenskii, P.A. (1992) *Detiam moim: vospominan'ia proshlykh dnei (To My Children: Reminiscences of Past Days)*, Moscow: Moskovskii rabochii, 1992.(The original manuscript was written intermittently between 1916 and 1925. Appended to this autobiography are Florenskii's own genealogical investigations and his letters to his family written from imprisonment.)

Florenskii, P.A. (1993) *Analiz prostranstvennosti i vremeni v khudozhestvenno-izobrazitel'nykh proizvedeniiakh (An Analysis of Space and Time in Works of Fine Art)*, Moscow: Izdatel'skaia gruppa 'Progress'. (This contains his lectures at the Higher Artistic-Technical Studios [VKHUTEMAS], 1924-5.)

Florenskii, P.A. (1994) *Sochineniia v chetyrekh tomakh (Compositions in Four Volumes)*, vol. 1 Moscow: Izdatel'stvo 'Mysl''.(A collection of his articles from 1903-9 along with Florenskii's résumé (*c*.1925-6) of his own thought. Volume 2 is to include subsequent articles from 1903-33 with volumes 3 and 4 containing *Stolp i utverzhdenie istiny* and *U vodorazdelov mysli* respectively.)

References and further reading

Bychkov, V. (1990) *Ėsteticheskii lik bytiia: Umozreniia Pavla Florenskogo*, Moscow: Znanie; trans. R. Pevear and L. Volokhonsky, *The Aesthetic Face of Being: Art in the Theology of Pavel Florensky*, Crestwood, NY: St Vladimir's Seminary Press, 1993.(An overview of Florenskii's aesthetics and his understanding of classical iconography in particular.)

Florovskii, G. (1937) *Puti russkogo bogosloviia (The Ways of Russian Theology)*, Paris: n.p.; 2nd edn, Paris: YMCA-Press, 1981.(For his critical commentary on Florenskii, see 493-8.)

Isupov, K.G. (ed.) (1996) *P.A. Florenskii: Pro et Contra*, St Petersburg: Izdatel'stvo Russkogo Khristianskogo Gumanitarnogo Instituta.(An anthology of critical reviews of Florenskii written by leading past and contemporary Russian philosophers.)

Lossky, N. (1951) *History of Russian Philosophy*, New York: International Universities Press, 176-91.(A critical synopsis of Florenskii's metaphysics and sophiology.)

Slesinski, R. (1984) *Pavel Florensky: A Metaphysics of Love*, Crestwood, NY: St Vladimir's Seminary Press.(The most extensive commentary on Florenskii in English.)

Trubachev, A. (1989) *'K 100-letiiu so dnia rozhdeniia sviashchennika Pavla Florenskogo. Ukazatel' pechatnykh trudov' (In Commemoration of the Centenary of the Birth of the Priest Pavel Florenskii. Index of Published Works)* in *Bogoslovskie trudy (Theological Studies)* 23: 264-309.(Lists all Florenskii's works published in Russia and the Soviet Union from 1901-82; 338 entries.)

Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, vol. 2, Paris: YMCA-Press, 413-30; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, London: Routledge & Kegan Paul and New York: Columbia University Press, 2 vols, 1953.(A digest of the metaphysics of pan-unity, along with an outline of Florenskii's epistemology and metaphysics.)

Fludd, Robert (1574-1637)

Fludd is a marginal figure in the mainstream development of philosophy in seventeenth-century England, but of some importance in the development of an esoteric and mystical philosophy which took the form of an 'underground' movement, Rosicrucianism. Rejecting the various types of scholastic philosophy that were still dominant in his day, Fludd drew on Neoplatonist and Renaissance sources in an effort to redirect learning in what he took to be a Christian direction. His writings centre around a search for hidden connections between a purely intelligible realm and the realm of sensation.

Robert Fludd was born in Kent in the south of England. After a period at St John's College, Oxford, he studied medicine, chemistry and the occult sciences in continental Europe, where he was employed as a tutor to noble families. He returned to England in 1604 and studied medicine at Christ Church, Oxford. Despite his rejection of the current Galenist orthodoxies, he was admitted a Fellow of the Royal College of Physicians, and pursued a form of holistic medicine in which magnetism and psychic healing played a significant role. From 1619 onwards he was engaged in various controversies with Gassendi, Kepler, Mersenne and others.

Fludd assumed a continuity or even an identity between ancient and Christian thought, to the extent that the whole of ancient thought must be shown to anticipate Christianity, even down to points of detail such as the doctrine of the Trinity. In setting out this account he relies very extensively on Hermetic texts, which he maintains date from remote antiquity, despite the fact that Isaac Casaubon had shown in 1614 that they were actually composed well into the Christian era. Much of his work takes its bearings from commentary on the book of Genesis, having as its aim the reconciliation of Scripture with natural philosophy. Fludd pursued this project in a way which had two distinctive features. First, he read Genesis as showing that light and dark are the two basic principles from which everything else follows. God created the universe by contracting into himself, thereby creating darkness, then expanding outwards again in the form of light. Fludd relied heavily on the metaphorical connotations of light and dark, associating them directly with God and Satan. The second distinctive feature of his approach is that it depends as much, if not more, on pictorial representation as on verbal description. The latter is insufficient if we are to capture the deep, hidden relations between the world of sensation and the intelligible reality that underlies it, and pictorial representation is able to capture such mysteries as the Trinity in a way that verbal description is not. Here Fludd believed he had discovered the key to knowledge that Kabbalists, Rosicrucians and various Hermetist sects had been seeking (see Hermetism; Kabbalah).

With this key, Fludd proceeded to unlock hidden antipathies and sympathies, using pairings of terms that mirrored the fundamental dichotomy between light and darkness in his exploration of various harmonic and magnetic phenomena. All natural phenomena are to be construed ultimately as manifestations of the one light-dark contrast, but Fludd interpreted this claim sometimes reductively (for example, in terms of a contrast between heat and cold, the universal effects of which were displayed by means of a primitive thermometer), and sometimes metaphorically and analogically, seeking parallels and similarities rather than simply trying to explain one thing in terms of another.

See also: Alchemy; Neoplatonism; Paracelsus; Theosophy

STEPHEN GAUKROGER

List of works

Fludd, R. (1617) *Tractatus Theologico-Philosophicus in Libros tres (Theological-Political Treatise in three books)*, Oppenheim: Johann Theodore de Bry.(Fludd's most extensive theological tract, written from a Hermetic and Neoplatonic perspective.)

Fludd, R. (1617/21) *Utriusque Cosmi Maioris scilicet et Minoris Metaphysica, Physica atque Technica Historia (History of the Microcosm and Macrocosm)*, Oppenheim/Frankfurt: Johann Theodore de Bry, 2 vols.(Fludd's account of the two worlds, the microcosm - the human being - and the macrocosm - God and the rest of his creation.)

Fludd, R. (1621) *Veritatis Proscenium (First Reply to Kepler)*, Frankfurt: Johann Theodore de Bry.(The first reply to Kepler sets out Fludd's defence of his application of the principles of harmony to the study of the heavens.)

Fludd, R. (1622) *Monochordum mundi symphoniacum (The second reply to Kepler)*, Frankfurt: Johann Theodore

de Bry.(Continues the theme of the ’First Reply’.)
Fludd, R. (1623) *Anatomiae amphitheatrum (The Amphitheatre of Anatomy)*, Frankfurt: Johann Theodore de Bry. (An account of anatomy and physiology on the basis of dissections, alchemical experiments and mystical principles.)
Fludd, R. (1633) *Clavis Philosophiae et alchymiae (The key to Philosophy and Alchemy)*, Frankfurt: Wilhelm Fitzer.(Fludd’s defence of alchemy and the principal response to Gassendi’s criticisms of alchemy from the point of view of atomism.)
Fludd, R. (1640) *Philosophia Moysaica*, Gouda: Petrus Rammazenius; trans. as *Mosaicall Philosophy: Grounded upon the Essentiall Truth or Eternal Sapience*, London: Humphrey Moseley, 1659.(Summary of Fludd’s cosmology, setting out a theory of universal attraction and repulsion.)

References and further reading

Debus, A.C. (1966) *The English Paracelsians*, New York: Franklin Watts, 105-27. (Good general account of Fludd’s work, paying special attention to his work in alchemy and medicine.)
Hutin, S. (1971) *Robert Fludd (1574-1637): alchimiste et philosophe rosicrucian (Robert Fludd: alchemist and rosicrucian philosopher)*, Paris: Alchimie et Alchimistes, vol. 8.(The only book-length treatment of Fludd that can be recommended without major qualification: available only in French.)

Fodor, Jerry Alan (1935-)

Jerry Fodor has been one of the most influential figures in the philosophy of mind, the philosophy of psychology, and 'cognitive science' through the latter part of the twentieth century. His primary concern has been to argue (vigorously) for a certain view of the nature of thought. According to this view, thinking is information processing within 'the language of thought'. The mind can be understood as a computer, which directs action with the aid of internal representations of the world.

1 Fodor's view of the mind

Like several other philosophers, Fodor in the 1960s and 1970s defended a functionalist view of the mind (Fodor 1968; see Functionalism). Functionalism makes possible a physicalist worldview but does not seek simply to reduce sciences such as psychology to physics. The computer model of the mind has been central to functionalism. Fodor, however, developed a more literal application of this model than other functionalists.

For Fodor, the value of the computer model lies in the light it sheds on reasoning, belief, planning and other intelligent thought-processes. (He does not see functionalism as a way of explaining the first-person 'feel' of experience.) Fodor claims that thinking is performing computational operations on mental representations. These inner representations form a system with many of the basic properties of a language, so the system can be called 'the language of thought' (Fodor 1975). This inner language is not identical to any public language, such as English. Rather, it is used in learning public languages. Thus the language of thought is innate (see Language of thought).

Fodor holds that to believe that limes contain vitamin C, for example, is to have in one's head a certain sentence-like formula in the language of thought. A formula is a belief that limes contain vitamin C in virtue of both its internal causal role, which makes it a belief rather than a hope or desire, and also in virtue of connections to the external world, which determine its content. These formulas are made up of mental 'terms' with their own properties of meaning and reference. The terms combine to generate the truth-condition of the whole. There are also 'implicit' beliefs which follow trivially from the explicitly represented ones, but only explicitly represented beliefs contribute to thought processes.

Though the inner formulas have semantic properties, their causal contributions to thought depend only on their formal or 'syntactic' properties; these are the properties relevant in computational processes. Indeed, these semantic properties depend partly on the nature of the thinker's environment, and what is outside the head cannot directly affect the production of behaviour. None the less, Fodor insists that there are many psychological generalizations that are naturally and perhaps necessarily expressed in terms of the semantic content of thoughts.

For some time Fodor posited an inner code with representational properties, but did not have a theory of how such representation was possible. He has tried to solve this problem with his 'asymmetric dependence' theory of meaning (Fodor 1990). This is a variety of informational or indicator semantics, based on causal and law-like connections between thoughts and their objects (see Semantics, informational). We cannot simply say that whatever can cause a representational state is represented by that state, since then error would be impossible. Fodor's proposal is that an inner symbol 'horse' represents horses if occurrences of this symbol are reliably caused by horses, and the only other things that can cause occurrences of this symbol do so *because* of the connection between the 'horse' and horses.

These are the central elements of Fodor's view, but he has contributed to many other topics. He has defended the 'automomy' of higher-level sciences against reductionism (Fodor 1975). He has outlined a 'modular' view of mind, in which specialized and partially autonomous mental faculties are responsible for perception and some other features of cognition (Fodor 1983). Recently he has also attacked the holism influential in much twentieth-century epistemology and semantics (Fodor and LePore 1992; see Holism: mental and semantic; Modularity of mind).

2 Debates

Fodor's programme is so ambitious that nearly every aspect of his view has been controversial. Some have found

the idea of an inner language to be either incoherent (as languages are essentially public) or empirically unmotivated. For others more sympathetic to the representationalist approach, Fodor has never resolved a tension between his stress on the 'syntactic' nature of mental processes, and his insistence on the importance of semantic properties in psychological explanation.

This problem is magnified by his acceptance of the view that the semantic properties of thought depend on factors external to the agent's body. For some years Fodor and others tried to develop a purified 'narrow' way of ascribing thoughts which would not distinguish physically identical agents who happen to inhabit different environments. More recently he has argued that there is no real problem here after all; intentional laws are 'implemented' in computational processes, and these are two distinct levels of description. For some, however, Fodor has never resolved this problem (Loewer and Rey 1991; see Content: wide and narrow).

Another problem, which Fodor himself has explored, concerns learning (Fodor 1981). Psychologists have often seen various types of learning as hypothesis testing. Fodor adopts a realist interpretation of this view: such theories must posit internal representations of hypotheses and the evidence used to choose between them. But how then can new *concepts* be learned? To test hypotheses about the concept 'cigar' we must already be able to represent internally the various possibilities for its meaning. Consequently we cannot acquire genuinely new representational capacities from experience; all we can do is rearrange and rename concepts we already have. So the representational approach, in Fodor's hands, leads to a view in which much of what appears to be learning is actually nothing of the sort (see Nativism §3; Concepts).

Fodor's 'asymmetric dependence' theory of meaning has not been widely accepted. (Loewer and Rey 1991). Some have raised problem cases, while others (like the present author) are more generally sceptical about the counterfactuals that Fodor's theory posits. Fodor says that the inner symbol 'horse' represents horses if it is reliably caused by the presence of horses, and other things only produce occurrences of the symbol because of the connection between the symbol and horses. This dependence is present-tense rather than historical. But why is it the connection between 'horse' and horses that explains the other connections? As a consequence of physical facts about the wiring connecting the symbol to the senses, some horses can cause occurrences of the symbol. These same facts explain why other objects, such as cows, sometimes cause occurrences of the symbol. The connection which explains the others is a connection between the symbol and a certain sensory appearence. If so, the connection between the symbol and horses is not fundamental in the way that Fodor claims. These debates are complex, however, and no naturalistic theory of meaning has generated wide acceptance.

For many years Fodor claimed that, with the demise of behaviourism, the computational, symbolic approach was the only genuine research programme we had for studying cognitive processes. Hence it was 'the only game in town' as a theory of the general nature of thought. The rise of connectionism during the 1980s partially undermined this claim. The connectionist model of thought is based upon parallel and simultaneous interactions, of a very simple nature, between neuron-like elements. Connectionism has closer links to the brain sciences than Fodor's more abstract approach, and it does not posit language-like inner symbols. Fodor has been an outspoken critic of connectionism (Fodor and Pylyshyn 1988), but there are now other games in town, and in this new context there is ongoing debate over the merits of the symbolic approach (see Connectionism).

PETER GODFREY-SMITH

List of works

Fodor, J.A. (1968) *Psychological Explanation*, New York: Random House.(Presents a functionalist view of mind and defends it against reductionism and behaviourism.)

Fodor, J.A. (1975) *The Language of Thought*, New York: Thomas Crowell.(Probably Fodor's most influential book. Outlines and defends the hypothesis of a language of thought, utilizing both philosophical and psychological evidence with great ingenuity.)

Fodor, J.A. (1981) *Representations*, Cambridge, MA: MIT Press.(Mostly a collection of reprinted articles, including 'Methodological Solipsism…', which influenced many discussions of 'narrow psychology'. Also includes Fodor's most detailed discussion of innate concepts.)

Fodor, J.A. (1983) *The Modularity of Mind*, Cambridge, MA: MIT Press.(An interdisciplinary work positing a range of specialized and partly independent mental faculties or 'modules' for certain cognitive tasks. Expresses pessimism about our chances of understanding mental capacities which do not have modules.)

Fodor, J.A. (1984) ‘Semantics, Wisconsin Style’, *Synthèse* 58: 231-50.(An influential article outlining the problem that the asymmetric dependence theory was intended to solve.)
Fodor, J.A. (1987) *Psychosemantics*, Cambridge, MA: MIT Press.(A systematic statement of Fodor’s view of the mind. Defends common-sense ‘folk psychology’, introduces the asymmetric dependence theory, criticizes meaning-holism, and defends the language of thought.)
Fodor, J.A. and Pylyshyn, Z. (1988) ‘Connectionism and Cognitive Architecture’, *Cognition* 28: 3-71.(An attack on connectionism.)
Fodor, J.A. (1990) *A Theory of Content, and Other Essays*, Cambridge, MA: MIT Press.(Contains his most detailed discussion of the asymmetric dependence theory of meaning, also ‘Semantics, Wisconsin Style’ and various other articles.)
Fodor, J.A. and LePore, E. (1992) *Holism: A Shopper’s Guide*, Cambridge, MA: MIT Press.(A criticism of holist positions in philosophy of mind, semantics and epistemology.)
Fodor, J.A. (1994) *The Elm and the Expert*, Cambridge, MA: MIT Press.(Attempts to reorient the debates about individualism and ‘narrow psychology’.)

References and further reading

Loewer, B. and Rey, G. (1991) *Meaning in Mind: Fodor and his Critics*, Oxford: Blackwell.(A collection of essays on Fodor’s work, with replies by Fodor.)
Dennett, D.C. (1978) ‘A Cure for the Common Code’, in *Brainstorms*, Cambridge, MA: Bradford Books.(A critical notice of *The Language of Thought*.)

Fonseca, Pedro (1528-99)

Called in his own time 'the Portuguese Aristotle', Pedro da Fonseca was a sixteenth-century Jesuit philosopher and theologian. Schooled as a Thomist, Fonseca was a master of the Greek, Arabic and scholastic traditions, which enabled him to pursue his own independent line on various issues dealt with by Aquinas and Aristotle. As reflected in his publications, his chief accomplishments were in logic and metaphysics. He authored two very important and widely used works: a clear, comprehensive and systematic textbook in logic (Institutionum dialecticarum) and an edition of Aristotle's Metaphysics with translation plus explanation and commentary. A third shorter work of introduction to logic (Isagoge philosophica) was also influential.

1 Life

Pedro da Fonseca (Petrus Fonseca) was born at Cortiçada (now Proença-a-Nova) in Portugal. He entered the Society of Jesus in 1548. After a period of noviciate, in 1551 he enrolled at the newly founded University of Évora, where between 1552 and 1555 he studied theology and also taught philosophy in 1552-3. From 1555 to 1561, he taught philosophy in the Jesuit-directed College of Arts at the University of Coimbra. During this last period, Fonseca promoted the idea of a *Cursus Conimbricensis* which later became a reality through the efforts of fellow Jesuits at Coimbra (see Collegium Conimbricense). From 1561 to 1564 he served the Jesuits in various administrative roles. In 1570 he received his doctorate in theology at Évora and became chancellor there. From 1572 to 1582 he served in Rome as general assistant for the Jesuit Province of Portugal. While in Rome, he worked with others on a 'Plan of Studies' (*Ratio studiorum*) which was later adopted by the Society of Jesus. Returning to Portugal in 1582, he became a Jesuit superior in Lisbon and then a visitor for the province. In 1592, he was again in Rome, a delegate to the fifth General Congregation of the Society, which among other things legislated that the Jesuits should follow Aristotle in philosophy. He died in Lisbon.

2 Logic

Fonseca's contribution to philosophy is in two main parts, logical and metaphysical. In logic, he authored two important works. The first was *Institutionum dialecticarum libri octo (Eight Books of Dialectical Instructions)*. Published at Lisbon in 1564 and re-edited fifty-two more times by 1625, it was adopted as a textbook, especially by the Jesuits, throughout Europe, America and the Far East. The second work was a much shorter *Isagoge philosophica (Philosophical Introduction)*, which was published initially in 1591 and re-edited eighteen times up to 1623. More than a simple commentary on the *Isagōgē* of Porphyry (§§2, 5), Fonseca's book was a new introduction to the Organon of Aristotle (§4). After a brief preface and a proem, he devoted six chapters to a general treatment of universals, particulars and the abstraction of one from the other. Then over five chapters he treated the five universals (genus, species, difference, property and accident) of Porphyry. Finally, in Chapter 12 he treated 'certain other species of universals [connected with the logic of the Trinity and the Incarnation] which pagan philosophers did not know' (*Isagoge philosophica*: 61).

The *Institutionum dialecticarum* is a systematic presentation of Aristotelian formal logic. Book 1 consists of thirty-two chapters in which Fonseca treats successively of the necessity, the names and the nature of logic. He distinguishes it from dialectic as Aristotle speaks of this in his *Topics* and then goes on to discuss nouns, verbs and signs (formal and instrumental as well as natural and conventional). After discussion of concepts as signs, he treats terms as equivocal (including analogical) and univocal, concrete and abstract, connotative and absolute, common and singular, transcendent and supertranscendent, positive and negative, contradictory and non-contradictory, as well as of first and second intention, and so on. Book 2 concerns the universal and covers much of the same matter (for example, genus, species and so on) treated in his *Isagoge*, plus the ten categories of Aristotle. Book 3 deals with various types of proposition. Book 4 covers division and Book 5 treats of definition. Book 6 deals with consequence, argumentation, invention and judgment, syllogisms, enthymemes, and induction. Book 7 is concerned with demonstration, dialectical reasoning, 'places' (or seats of argument), teaching procedure and the art of disputation. Book 8 mainly concerns fallacies, but Fonseca also treats supposition, ampliation, restriction and appellation in detail. As for the Aristotelian character of the *Institutionum dialecticarum*, Ferreira Gomes ([1564] 1964: xlvi-xlvii) notes that in them Aristotle is cited 600 times and he quotes a 1597 editor to the effect that Fonseca's work covers the Aristotelian logic so well that it makes Aristotle's own work, apart from its

historical value, almost useless (see Logic, Renaissance).

3 Metaphysics

Comprising four quarto volumes, Fonseca's *In libros Metaphysicorum Aristotelis Stagiritae (Commentary on the Books of Aristotle's Metaphysics)* contains a critical Greek text which he himself established from the best available manuscripts and printed editions. Through the four volumes, in a right-hand column matching the Greek to the left, he has given a fine Latin translation. An explanation of the text follows each chapter and then commentary 'by way of question' (*per modum quaestionis*) on most of the chapters through the first nine books of the *Metaphysics*. Published posthumously, Books 10, 11, and 12 continue to give the Greek and Latin plus the explanation, while Books 13 and 14 give only the text in the two languages. The questions, which contain Fonseca's own thought on subjects metaphysical, were developed in scholastic fashion. After asking a particular question, Fonseca, through a series of 'sections', presents objections and opinions, clarifies terms and concepts, gives and proves his own answer, and then returns to answer the objections raised in support of other views. Extraordinarily well versed in the earlier Greek, Arabic and scholastic traditions, Fonseca wherever possible follows the lead of Aristotle and Thomas Aquinas, but with a decided independence.

After rejecting opinions which hold that the subject of metaphysics is God, Aristotelian 'separate substances', or being in the categories, Fonseca says that the first and adequate subject of metaphysics is being in so far as it is common to God and creatures (*In libros Metaphysicorum* IV c.1 q.1 s.3). Understood in this way, being is analogous, although as said of species within one genus or of individuals within one species it is univocal. Between God and creatures, between created substance and accidents, between different classes of accident, and between real being and being of reason, being is analogous by analogies both of proportion and of attribution. As God is related to his being, so in proportion a created substance is related to its being. Likewise, as created substance and its being are related, so in proportion is an accident related to its being. Again, as one kind of accident is disposed to its existence so is each other kind of accident to its existence. And as real beings are disposed to their being, so beings of reason are to theirs (*Metaphysicorum* IV c.2 q.1 s.5, 7). An analogy of attribution obtains among accidents as an analogy of two things to a third (that is, created substance), while between accidents and substance it is analogy of one to the other. The same is true of beings of reason among themselves and then in comparison with real being; for beings of reason do not depend less upon real beings than do accidents upon substance. Again, a creature is being only by attribution or reference to God. Pursuing this, Fonseca distinguishes between formal and objective concepts. A formal concept is an 'actual likeness' (*actualis similitudo*) of a thing that is understood, produced by the intellect in order to express that thing. An objective concept is that thing which is understood in so far as it is conceived through the formal concept. Both the formal and the objective concept of being are one, but not perfectly so for the reason that they do not prescind perfectly from the concepts of the members which divide being. Being as such is transcendent as are also the concepts of thing, something, one, true and good (*Metaphysicorum* IV c.2 q.2 s.1, 4-5; q.5 s.2) (see Being).

In God alone there is a perfect identity of essence and existence. In every creature, essence is distinct from existence, but not as one thing from another. Rather, says Fonseca, a created essence is as distinct from its existence as a thing from its ultimate intrinsic mode. In this opinion, he tells us, he is following Alexander of Hales and Duns Scotus (§12) (*Metaphysicorum* IV c.2 q.3 s.4). It is possible that here Fonseca has also to some extent anticipated the Suárezian doctrine of modes.

Excluded from the subject of metaphysics are accidental beings (*entia per accidens*) and beings of reason. An accidental being, in the sense excluded, is a juxtaposition of two or more beings which lack any (intrinsic) relation to one another (*Metaphysicorum* IV c.1 q.1 s.3). Beings of reason are those which exist only inasmuch as they are objects of understanding. Within such beings of reason, as they stand in contrast with mind-independent real beings, Fonseca distinguishes a proper being of reason from one which is fictitious. Properly taken, a being of reason is one whose being depends upon the understanding in such way that it can still be said of real beings, for example, the concepts of genus, species, and the like. A fictitious being as such is a being whose essence depends upon the understanding in such way that it cannot be said of any real being, for example, a chimera, a goat-stag, or the like (*Metaphysicorum* IV c.7 q.6 s.5).

See also: Aristotelianism, Renaissance; Language, Renaissance philosophy of; Logic, Renaissance

JOHN P. DOYLE

List of works

Fonseca, P. da (1564) *Institutionum dialecticarum libri octo (Eight Books of Dialectical Instructions)*, Coimbra, 1575; ed. and trans. J. Ferreira Gomes, *Pedro da Fonseca: Instituiçiones dialécticas*, Coimbra: Universidade de Coimbra, 1964, 2 vols. (This edition of Fonseca's main logical work includes introduction, critical Latin text, notes and Portuguese translation. Ferreira Gomes' introduction lists the editions of this work and also those of the *Isagoge* and commentary on Aristotle's *Metaphysics*.)

Fonseca, P. da (1591) *Isagoge philosophica (Philosophical Introduction)*, Lisbon; ed. and trans. J. Ferreira Gomes, *Pedro da Fonseca. Isagoge filosófica*, Coimbra: Universidade de Coimbra, 1965.(Fonseca's introduction to Aristotelian logic; with introduction and Portuguese translation.)

Fonseca, P. da (1615-29) *In libros Metaphysicorum Aristotelis Stagiritae. Tomi quatuor (Four Books on Aristotle's Metaphysics)*, Coloniae; repr. Hildesheim: Olms, 1964.(Fonseca's Greek text of the *Metaphysics* with Latin translation, explanation and commentary. The complete text was published posthumously.)

References and further reading

Ashworth, E.J. (1974) *Language and Logic in the Post-Medieval Period*, Dordrecht and Boston, MA: Reidel.(A general study which includes discussion of Fonseca's logic.)

Ceñal, R. (1943) 'Pedro da Fonseca (1528-1599). Su crítica del Texto de la "Metaphysica" de Aristóteles' (His Textual Criticism of Aristotle's *Metaphysics*), *Revista de Filosofia* 2 (4): 124-46.(Treats Fonseca as a textual critic of Aristotle's *Metaphysics*.)

Giacon, C. (1944) *La seconda scolastica (The Second Scholasticism)*, Milan: Bocca, vol. 2, 31-66. (Fonseca treated, along with Toletus and Pereira, within the sixteenth-century revival of scholasticism.)

Lohr, C.H. (1987) *Latin Aristotle Commentaries. II Renaissance Authors*, Florence: Olschki, 150-1.(Contains bibliographical information on Fonseca.)

Risse, W. (1964) *Die Logik der Neuzeit (The Logic of the Modern Period)*, Stuttgart and Bad Cannstatt: Frommann, vol. 1, 361-72.(Fonseca's logic treated within its historical context.)

Solana, M. (1940) *Historia de la filosofia española. Época del renacimiento (siglo XVI) (History of Spanish Philosophy. The Time of the Renaissance (16th Century))*, Madrid: Asociación Española para el Progreso de las Ciencias, vol. 3, 339-66.(A good introductory overview of Fonseca.)

Fontenelle, Bernard de (1657-1757)

Despite his considerable historical importance and vast output of literary, critical and philosophical works, Fontenelle did not make original contributions to philosophy. He popularized a modern view of nature, and raised doubts about institutionalized religions and unexamined theistic beliefs. As a champion of science and secularization, Fontenelle was extremely influential; his Entretiens (Conversations) of 1686 were quickly translated into many languages and became one of the basic texts of the early Enlightenment.

1 Life and work

Bernard Le Bovier ('Boyer') de Fontenelle was born in Rouen, France, and received a Jesuit education at the Collège de Bourbon. Early literary fame and the support of Pierre and Thomas Corneille, his uncles, gained him access to the intellectual and social circles of Paris. He was elected to the Academie Française in 1687 and became its perpetual secretary ten years later. He died in Paris shortly before his one-hundredth birthday.

Fontenelle wrote poems, tragedies, novels, comedies and libretti for operas (set to music by Jean-Baptiste Lully). The pieces written for the stage met with only moderate success. Apart from satires, dialogues and critical essays, he composed so-called 'lettres galantes' and maintained a far-reaching correspondence with the leading thinkers of his time.

As a historian of science, Fontenelle distinguished himself through his exhaustive report of the Academy's work since its inception, the *Histoire de l'Académie royale des sciences depuis 1666 jusqu'en 1699* (1720-33). Several multivolume sequels followed at regular intervals. Parallel to these chronicles, Fontenelle drafted almost seventy eulogies on deceased members, the *Éloges des académiciens* (1708, 1722 and following), reviewing and summarizing their achievements.

Interested in the infinitesimal calculus, Fontenelle wrote a preface to a work by d'Hôpital and an account of recent mathematical research, the *Éléments de la géometrie de l'infini* (1727). In his philosophy of nature, Fontenelle defended theories developed by others instead of developing new ideas; in his philosophy of religion, he was a sceptic inspired by de Huet, Thomassin and Spinoza (see Huet, P.-D. §2; Spinoza, B. de §14). But as a popularizer and advocate, Fontenelle became an influential champion of science and secularization, and the leading pioneer of the early French Enlightenment.

2 Philosophy of nature

Fontenelle's most famous work is the 1686 *Entretiens sur la pluralité des mondes habités (Conversations on the Plurality of the Inhabited Worlds)*. This popular account of astronomy and natural philosophy consists of five dialogues, to which a sixth dialogue was added in a revised edition of 1687. The conversations begin with the assertion that our planetary system is not the only one in the universe: there are others whose suns are the fixed stars. Earth and the planets revolve around the sun; the celestial motions are described by the Copernican model and explained by Cartesian vortices. The second dialogue touches on the corpuscular theory of light, an explanation of solar eclipses and the possibility of space flight, and turns into a discussion about the moon. An examination of lunar features reveals the moon to resemble Earth in significant respects. Considering this, Fontenelle speculates that Earth's satellite is inhabited. This claim is expanded in the third dialogue. The microscope reveals a compelling ubiquity of life even in the tiniest water droplet and on the smallest stone in our world. Since Earth and other heavenly bodies are alike in their appearances, we can treat them alike in what is not apparent as well, as long as evidence to the contrary is lacking. Operating with this strong principle of analogical reasoning, Fontenelle argues for the possibility of life on other worlds. The fourth and fifth dialogues are imaginative summaries of the current knowledge about planets and stars. The supplementary sixth dialogue contains an update of recent astronomical and geological discoveries and a reflection on probability and analogical reasoning.

What makes this work interesting is Fontenelle's persuasive presentation of an entirely modern worldview. Anthropocentrism is obsolete; humans do not reside at the centre of a small enclosed universe, but are just one among many possible civilizations in a boundless cosmic ocean in which worlds are scattered like so many islands. It is doubtful that humans are the pinnacle of creation; far more advanced beings may exist. Anthropomorphism

fails to fathom how such extraterrestrials might look; the universe is too fertile and diverse for even the boldest imagination. But despite its diversity, the cosmos is a well-ordered system: a principle of order generates a maximum of effects with a minimum of means. These means are mechanical and law-governed, and modern science can account for them. The *Entretiens* proved to be extremely influential in the European thought of the eighteenth century; Fontenelle's conception of nature was picked up by subsequent *philosophes* and even influenced Kant's cosmology.

Throughout his life, from the *Entretiens* to the *Lettre à Basnage de Beauval* (1699) to his last work, the *Théorie des tourbillons cartésiens* (1752), Fontenelle defended Cartesian vortices against the rise of Newton's universal gravitation. Since whirls of matter presuppose a medium, Fontenelle argued for the existence of a cosmic ether in which heavenly bodies move. Apart from that, however, Fontenelle did not feel much allegiance to Cartesianism. Whereas Descartes excluded teleology, Fontenelle emphasized purposive explanations of nature. Fontenelle rejected Malebranche's occasionalist interpretation of Descartes in the *Doutes sur le système physique des causes occasionelles* (1686b), favouring a physical influx model of causality (see Occasionalism).

3 Philosophy of religion

On the surface, God seems to be an indispensable part of Fontenelle's universe, but it remains unclear what God is. Like Spinoza, Fontenelle subscribed to the idea that there is a fundamental conformity between the order of reason and the order of nature (see Spinoza, B. de §§5-6). This conformity permits humans to know nature but does not lead to knowledge of God. Like the Deists, in his correspondence with Leibniz (1701-4) Fontenelle took the principle of order to be immanent to nature: matter, not God, determines the laws of motion (see Deism). The nature of God's existence cannot be known; at the very least, the Cartesian proofs do not work (*Sur l'existence de Dieu* 1724). Fontenelle remained evasive about these issues through fear of censorship and persecution. At any rate, the utopian society envisaged in the clandestine novel *Histoire des Ajaoiens* (*c*.1680, attributed to Fontenelle) is devoid of institutionalized monotheism, and instead embraces pantheism.

Continuing themes of Fontenelle's work were his impatience with superstition and his criticisms of religious credulity. He derided pagan superstition in the *Dialogues des morts* (1683), and in *La Comète* (1681), a comedy written after the appearance of Halley's Comet in 1680, ridiculed the widespread belief that passing comets have negative effects. According to *De l'origine des fables* (1679, published 1724 as part of *Sur l'histoire*) and the *Histoire des oracles* (1686c), religions stem from an anthropological and a social source. Human imagination triggered by an ignorance of natural phenomena, and successful imposters exploiting human gullibility were the first causes of religious beliefs.

In the essay *Traité de la liberté de l'âme* (written 1700, seized and burnt on parliamentary order, republished anonymously in the 1743 collection *Nouvelles libertés de penser*), Fontenelle raised even more serious questions. Not only is God unknowable, not only do religions derive from human imagination, but it also appears that divine foreknowledge and human freedom are incompatible. Although Fontenelle makes an ostensible attempt at their reconciliation, the actual result of his examination is different: divine foreknowledge implies fatalism; human freedom is possible only without a prescient God.

See also: Clandestine literature

MARTIN SCHÖNFELD

List of works

Fontenelle, B. de (1742-6) *Oeuvres*, Paris: Brunet; repr. 1758-66, Amsterdam, 12 vols. (The first edition of collected works, published during Fontenelle's lifetime, originally in eleven volumes. The subsequent Amsterdam edition was augmented by an additional volume.)

Fontenelle, B. de (1818) *Oeuvres complètes*, ed. G.-B. Depping, Paris; repr. Geneva: Slatkine, 1968, 3 vols.(The second edition of Fontenelle's works; not complete, despite its title. Arranged by categories: science and mathematics in volume 1, philosophy and history in volume 2, poetry and fictional works in volume 3.)

Fontenelle, B. de (1989-) *Oeuvres complètes*, ed. A. Niderst, Paris: Arthème-Fayard, 7 vols.(The supposedly definitive edition. Fontenelle's works are arranged chronologically as well as topically.)

Fontenelle, B. de (1970) *The Achievement of Bernard Le Bovier de Fontenelle*, ed. L. Marsak, New York and London: Johnson Reprint Corporation.(The only English anthology of Fontenelle's works. Contains

photographic reproductions of the first English editions of Fontenelle's *Plurality of Worlds* (London 1688) and *The History of Oracles* (London 1688). It also contains a new translation of the *Discourse Concerning the Ancients and the Moderns* and short selections from other works.)

Fontenelle, B. de (1679) *De l'origine des fables (On the origin of myths)*, in *Oeuvres complètes* ed. G.-B. Depping, vol. 2; ed. A. Niderst, vol. 3.(An attack of religious superstitions in the form of a brief essay.)

Fontenelle, B. de (*c*.1680) *Histoire des Ajaoiens (History of the Ajaoiens)*.(Novel with antireligious overtones, the fictional account of a benevolent and atheist society.)

Fontenelle, B. de (1681) *La Cométe (The comet)*, in *Oeuvres complètes* ed. G.-B. Depping, vol. 3; ed. A. Niderst, vol. 4.(Comedy written after the appearance of Halley's Comet in 1680.)

Fontenelle, B. de (1683) *Dialogues des morts anciens et modernes (Dialogues of the dead in antiquity and modernity)*, in *Oeuvres complètes*, ed. G.-B. Depping, vol. 2.(Thirty fictional dialogues, some among ancient figures, such as Anacreon and Aristotle, others between ancients and moderns, such as Socrates and Montaigne, and again others among moderns, such as Cortez and Montezuma.)

Fontenelle, B. de (1686a) *Entretiens sur la pluralité des mondes habités*, in *Oeuvres complètes*, ed. G.-B. Depping, vol. 2; trans. H. Hargreaves, *Conversations on the Plurality of Worlds*, Berkeley and Los Angeles, CA: University of California Press, 1990.(Fontenelle's main work, which has appeared in almost one hundred editions since its first publication. The 1990 translation contains an introduction by N.R. Gelbart.)

Fontenelle, B. de (1686b) *Doutes sur le système physique des causes occasionelles (Doubts on the physical system of the occasional causes)*, in *Oeuvres complètes*, ed. G.-B. Depping, vol. 1; ed. A. Niderst, vol. 1. (Philosophical treatise; a defence of Cartesianism and a critique of Malebranche and Arnauld.)

Fontenelle, B. de (1686c) *Histoire des oracles (History of oracles)*, in *Oeuvres complètes*, ed. G.-B. Depping, vol. 2; ed. A. Niderst, vol. 2.(One of Fontenelle's main works in the philosophy of religion. Consists of two parts, 'That Oracles were not at all given by demons', and 'That Oracles did not cease to exist at the coming of Jesus Christ'. A characteristic work of the early French Enlightenment and a sequel to *De l'origine des fables*.)

Fontenelle, B. de (1699) *Lettre de M. Fontenelle à M. Basnage de Beauval (Letter from Fontenelle to Basnage de Beauval)*, in *Oeuvres complètes*, ed. A. Niderst, vol. 3.(Fontenelle defends himself against the critic of Entretiens, and insists that the sun rotates around its own axis. Both observational and theoretical evidence support the hypothesis of solar rotation: the movement of the sun spots and the fact that the vortical centre of a revolving cosmic liquid cannot remain at rest.)

Fontenelle, B. de (1700) *Traité de la liberté de l'âme (Treatise on the freedom of the soul)*, in *Oeuvres complètes*, ed. G.-B. Depping, vol. 2; ed. A. Niderst, vol. 3.(Short philosophical essay, published anonymously, involving implicit objections to Church doctrine.)

Fontenelle, B. de (1701-4) 'Correspondence with Leibniz', in *Oeuvres complètes*, ed. A. Niderst, vol. 3.(An exchange of eleven letters touching on astronomy, chemistry, the composition of fire, mathematics, Leibniz's *Essai de Science Numerique* and Fontenelle's own works.)

Fontenelle, B. de (1708, 1722-) *Éloges des académiciens (Eulogies of the Members of the Academy)*, in *Oeuvres complètes*, ed. G.-B. Depping, vol. 1.(Fontenelle eulogized sixty-nine Academicians, among them Bernoulli, Boerhaave, d'Hôpital, Leibniz, Malebranche, Newton and Tschirnhaus.)

Fontenelle, B. de (1720-33) *Histoire de l'Academie royale des sciences depuis 1666 jusqu'en 1699 (History of the Royal Academy of Sciences from 1666 to 1699)*.(A multivolume work published in Paris 1720-33. Fontenelle explains the research and discoveries of the new generation of scientists and natural philosophers in a non-technical language. The preface to the *Histoire* is reprinted in Depping, volume 1.)

Fontenelle, B. de (1724) 'Sur l'existence de Dieu' (On the existence of God), in *Oeuvres complètes*, ed. A. Niderst, vol. 3.(A short philosophical essay. It is unlikely that matter caused life and thus, the existence of life suggests that God exists. None the less, metaphysical proofs of God's existence aspiring to certitude, such as the argument from design, are problematic.)

Fontenelle, B. de (1727) *Éléments de la géometrie de l'infini (Elements of the geometry of the infinite)*, in *Oeuvres complètes*, ed. G.-B. Depping, vol. 1; repr. Paris: Editions Klincksieck, 1995.(Book-length treatise on the philosophy of mathematics that contains a systematic theory of the infinite. The 1995 edition is a photographic reproduction of the 1727 text, with Fontenelle's responses to objections; also has new introduction by M. Blay and A. Niderst.)

Fontenelle, B. de (1752) *Théorie des tourbillons cartésiens (A Cartesian Vortex Theory)*, in *Oeuvres complètes*, ed. G.-B. Depping, vol. 1.(A longer philosophical treatise. Fontenelle's defence of the Cartesian vortices involves general reflections on physics, an attempt at coming to terms with Descartes' theory, and arguments to support a solar vortex and the ethereal atmosphere enveloping celestial objects.)

References and further reading

Carré, J.-R. (1932) *La Philosophie de Fontenelle ou le sourire de la raison*, Paris: F. Alcan; repr. Geneva: Slatkine, 1970.(A general survey of Fontenelle's views.)

Ekstein, N. (1996) 'Appropriation and Gender: the Case of Catherine Bernard and Bernard de Fontenelle', *Eighteenth Century Studies* 30: 59-81.(A historical-feminist study: no philosophical content. The author argues that Fontenelle's tragedy *Brutus*, as well as certain other plays, was not written by the author alone, but in collaboration with Catherine Bernard.)

Fayol, A. (1961) *Fontenelle*, Paris: Debresse.(A readable but superficial essay.)

Grégoire, F. (1947) *Fontenelle, une 'philosophie' désabusée*, Paris: J. Vrin.(A study emphasizing Fontenelle's philosophy of nature.)

Howells, R. (1992) 'The principle of mobility in Fontanelle's 'Entretiens sur [*sic*] la Pluralité des Mondes'', *French Studies* 46: 129-44.(A literary-interpretative study; no philosophical content. The author investigates how the multiplicity of points of view contained in the *Entretiens* can be subsumed under the heading of mobility.)

James, ED. (1990) 'Fontenelle's *Entretiens sur la pluralité des mondes* and their Intellectual Context', in *Actes de Columbus Racine; Fontenelle: 'Entretiens sur la pluralité des mondes', histoire et littérature: actes du XXI Colloquie de la North American Society for Seventeenth-Century French Literature*, Paris: Papers on French Seventeenth Century Literature, 133-47.(A careful account of the background of the *Entretiens*.)

Maigron, L. (1906) *Fontenelle, l'homme, l'oeuvre, l'influence*, Paris: Librairie Plon.(The earliest and standard monograph on Fontenelle.)

Niderst, A. (1972) *Fontenelle. A la recherche de lui-même (1657-1702)*, Paris: Nizet.(A well-researched study of the early Fontenelle and his intellectual growth.)

Niderst, A. (1989) *Fontenelle. Actes du colloque tenu à Rouen du 6 au 10 octobre 1987*, Paris: Presses universitaires.(A useful compendium of the results of recent research on Fontenelle's philosophy.)

Forcing

The method of forcing was introduced by Paul J. Cohen in order to prove the independence of the axiom of choice (AC) from the basic (ZF) axioms of set theory, and of the continuum hypothesis (CH) from the accepted axioms (ZFC = ZF + AC) of set theory (see set theory, axiom of choice, continuum hypothesis). Given a model M of ZF and a certain $P \in M$*, it produces a 'generic'* $G \subseteq P$ *and a model N of ZF with* $M \subseteq N$ *and* $G \in N$*. By suitably choosing P, N can be 'forced' to be or not be a model of various hypotheses, which are thus shown to be consistent with or independent of the axioms. This method of proving undecidability has been very widely applied. The method has also motivated the proposal of new so-called forcing axioms to decide what is otherwise undecidable, the most important being that called Martin's axiom (MA).*

1 The method: three basic lemmas

Let *M* be a countable transitive model of ZFC (see Set theory §7, terminology and notation from which will be used throughout). Let $P \in M$. Then $p, q \in P$ are called *compatible* if there is an $r \in P$ with $p \subseteq r$ and $q \subseteq r$. $D \subseteq P$ is *dense* if it is nonempty and for every $p \in P$ there exists a $q \in D$ with $p \subseteq q$. If *F* is a set of dense sets, *G* is *generic* for *F* if: (1) $q \in G$ whenever $q \subseteq p$ and $p \in P$, (2) any $p, q \in G$ are compatible, and (3) for any $D \in F$, $G \cap D \neq \emptyset$. *G* is *generic* for *M* if *G* is generic for $F = \{D \subseteq P \mid D \text{ dense and } D \in M\}$ (which appears uncountable to *M* but is really countable). If $F = \{D_0, D_1, D_2, \ldots\}$ is countable, choosing $p_0 \subseteq p_1 \subseteq p_2 \subseteq \ldots$ with $p_k \in D_k$, as is possible since each D_k is dense, $G = \{q \in P \mid q \subseteq p_k \text{ for some } k \in \omega\}$ is generic. A generic *G* generally cannot be an element of *M*.

For example, let *P* be the set of all functions *p* with domain a finite subset of ω and range a subset of 2. (Recall that an ordinal is identical with the set lesser ordinals, so $2 = \{0, 1\}, \omega = \{0, 1, 2, \ldots\}$.) Then *p*,*q* are compatible unless they 'disagree' in the sense that there is an $n \in \mathrm{dom}(p) \cap \mathrm{dom}(q)$ with $p(n) \neq q(n)$. For each $n \in \omega$, $D_n = \{p \in P \mid n \in \mathrm{dom}(p)\}$ is dense, since any function *p* not already defined on *n* can be extended to a function *q* that is. Likewise, for any function $f \in M$ from ω to 2, $D_f = \{p \in P \mid p(n) \neq f(n)\}$ is dense, since any function *p* with a finite domain that does not already disagree with *f* can be extended to a function that *q* does. In the example, if all the $D_n, D_f \in F$ for all *n* and *f* as above, then $g = \cup G$ will be a function *g* from ω to 2 disagreeing with each $f \in M$. Hence $g \notin M$.

Cohen's method associates to any given countable transitive model *M* of ZFC, and any given $P \in M$, and any given $G \subseteq P$ generic for *M* another countable transitive model $N = M[G]$ of ZFC, with $M \subseteq M[G]$ and $G \in M[G]$, in such a way that three basic lemmas stated below hold. For various choices of *P* the three basic lemmas can be used to show that $M[G]$ will be a model of various hypotheses Δ. Before stating the three basic lemmas it should be remarked that the best way to gain an intuitive understanding of their rather technical content is to see how they are used in applications (as in §§2,3).

Minimality lemma The $b \subseteq M$ with $b \in M[G]$ will be precisely those $b \subseteq M$ of the form $\delta_G(c) = \{x \mid (p, x) \in c \text{ for some } p \in G\}$ for some $c \in M$. Note that such a $\delta(c)$ would certainly have to belong to *any* countable transitive model *N* with $M \subseteq N$ and $G \subseteq N$. It is sometimes said that while *M* cannot 'know' about which *p* will get into *G* or about which $b \subseteq M$ will get into $M[G] - M$ (since *G* and such *b* simply do not exist in *M*), *M* does contain 'codes' *c* for such *b*, which can immediately be 'decoded' by δ_G once *G* is given. Especially important are the *canonical* code $x^{\#} = P \otimes x$ for any $x \in M$, and the *canonical* code $\Gamma = \{(p, p) \mid p \in P\}$ for the generic set. It is fairly easily checked that for any generic *G*, $\delta_G(x^{\#}) = x$ and $\delta_G(\Gamma) = G$.

Truth lemma If $\Phi(\delta_G(a), \delta_G(b), \ldots)$ holds inside $M[G]$, then there is some $p \in G$ such that $\Phi(\delta_H(a), \delta_H(b), \ldots)$ holds inside $M[H]$ for *any H* generic for *M* with $p \in H$. Such a *p* is said to *force* $\Phi(a, b, \ldots)$ over *M*.

Definability lemma To every formula Φ there may be associated a formula $\Phi^{\dagger}$ such that given any $p \in P$ and any $a, b, \ldots \in M$, then $\Phi^{\dagger}(p, a, b, \ldots)$ holds in *M* if and only if *p* forces $\Phi(a, b, \ldots)$ over *M*. Note that it follows, by the axiom of separation, that for any set *A* that is an element of *M*, the set $\{(p, a, b, \ldots) \mid u, v, \ldots \in A \text{ and } p \text{ forces } \Phi(a, b, \ldots)\}$ is an element of *M*. It is sometimes said that though *M*

cannot 'know' which *p* will get into *G* or whether a given Φ will come to hold in $M[G]$, at least Φ will come to hold in $M[G]$ only if it is forced to hold by some *p* that gets into *G*, and *M* 'knows' which *p* will, if they get into *G*, force Φ to hold in $M[G]$.

2 Application to the continuum hypothesis

To treat CH, let now ξ, η be the ordinals inside *M* that appear to *M* to be ω_1, ω_1 (really, they are, like *M* itself, countable), and vary the example of §1 by letting *P* be the set of all functions *p* with domain a finite subset of $\eta \otimes \omega$ and range a subset of 2. If *G* is generic, $g = \cup G$ will be a function from $\eta \otimes \omega$ to 2, and there will be functions g_ρ from ω to 2 for $\rho < \eta$, given by $g_\rho(n) = g(\rho, n)$. Since for distinct $\rho, \sigma < \eta$, $D_{\rho,\sigma} = \{p \in P \mid (\rho, n), (\sigma, n) \in \mathrm{dom}(p) \text{ and } p(\rho n) \neq p(\sigma n)\}$ is dense, these functions will be distinct not only from any function in *M*, but from each other. So ~CH will hold inside $M[G]$ provided ξ, η still appear to $M[G]$ to be ω_1, ω_2.

The proviso means there is no function *e* or *f* in $M[G]$ with $\mathrm{dom}(e) = \omega$, $\xi \subseteq \mathrm{ran}(e)$ or $\mathrm{dom}(f) = \xi$, $\eta \subseteq \mathrm{ran}(f)$. Cohen showed that for this it suffices that *P* should appear to *M* to have a certain property, the *countable chain condition* (CCC). Here an *antichain* is an $A \subseteq P$ such that p, q are incompatible for any two $p, q \in A$, and CCC is the property that any antichain is countable. (It ought logically to be called the '*no* countable *anti*chain condition', but the illogical terminology is well established.) To sketch why this suffices to rule out the existence of an *e* as above (the case of an *f* as above being similar), suppose $e \in M[G]$ is a function with $\mathrm{dom}(e) \subseteq \omega$.

By the minimality lemma, $e = \delta_G(c) \in M$ for some $c \in M$. By the truth lemma there is a $p_0 \in G$ forcing '*c* is a function with $\mathrm{dom}(c) = \omega^\#$ and $\xi^\# \subseteq \mathrm{ran}(c)$', and by restricting attention to those $p \in P$ such that $p_0 \subseteq P$, we may suppose that every $p \in P$ forces this. By the definability lemma $E = \{(p, n, \rho) \in P \otimes \omega \otimes \xi \mid p \text{ forces 'the value of function } c \text{ for argument } n^\# \text{ is } \rho^\# \text{'}\} \in M$.

Now consider how things appear inside *M*, where *P* appears CCC. For each $n \in \omega$ let $C_n = \{\rho < \xi \mid (p, n, \rho) \in E \text{ for some } p \in P\}$, and let $C = \cup\{C_0, C_1, C_2, \ldots\}$. Intuitively, C_n is the set of 'candidates' to become the output of the function for input *n*, while *C* is the set of 'candidates' to belong to the range of the function. For each $n \in \omega$ and $\rho \in C_n$ choose a $p(n, \rho)$ with $(p(n, \rho), n, \rho) \in E$, and let $A_n = \{p(n, \rho) \mid \rho \in C_n\}$. Note that if (p, n, ρ), $(q, n, \sigma) \in E$ and $\rho \neq \sigma$, then p, q are incompatible (since they force things which cannot both hold in $M[G]$ for any *G*). Hence any two elements of any A_n are incompatible. By CCC, each A_n, hence each C_n and hence *C* are all countable. Since $\xi = \omega_1$ (or so it appears inside *M*), *C* is bounded by some $\sigma < \xi$. Back outside *M* this means that $e = \delta_G(c)$ will be a function with $\mathrm{ran}(e) \subseteq \sigma < \xi$, contrary to assumption that $\mathrm{ran}(e) = \xi$.

To show that the set of functions *p* with domain a finite subset of $\omega_2 \otimes \omega$ and range a subset of 2 (which set is what *P* appears inside *M* to be) is CCC, the *delta system theorem* from combinatorial set theory is used (see Set theory §4): given an uncountable set *A* of incompatible elements of *P*, there must be a finite *b* and an uncountable $B \subseteq A$ such that $\mathrm{dom}(p) \cap \mathrm{dom}(q) = b$ for all $p, q \in B$. Thus the restrictions of the $p \in B$ to *b* give uncountably many distinct functions from *b* to 2. But if *k* be the number of elements in *b*, the number of such functions is just 2^k, which is finite, a contradiction.

3 Other applications

Among many other uses of the method of forcing, the following may be cited: If CH fails, the number of cardinals between $\aleph_0$ and c can be almost anything from one (hypothesis CH') to c (hypothesis CH#). Moreover, as regards the function $\exp\alpha = 2^\alpha$, all kinds of combinations are possible: One can have $\exp\aleph_0 = \exp\aleph_1 = \aleph_2$, or one can have $\aleph_0 = \aleph_2$ and $\exp\aleph_1 = \aleph_3$, and so on.

Cohen himself proved the independence of AC from ZF, combining forcing with an earlier construction used to prove independence of AC from a variant of ZF allowing 'individuals'. Indeed, Cohen's proof established the independence of the weaker axiom of dependent choice DC. The independence of AC from ZF+DC can similarly be established. Indeed, a whole range of implications of AC have been studied and questions as to which imply which and which are independent of which have been settled.

Many undecidability results have also been obtained by forcing in combinatorial set (see Set theory §4). For instance, any combination of CH or ~CH with SH or ~SH is consistent, where SH is Suslin's hypothesis.

Many undecidability results have also been obtained by forcing in the related areas of descriptive set theory and large cardinal theory (see Set theory §§5,7). In particular, adding the hypotheses IC, MC, … of the existence of inaccessible, measurable, … cardinals to ZFC does not decide CH (or SH or any other such combinatorial hypothesis).

As for the hypothesis AM (resp. PM) that all (resp. all projective) sets of reals are Lebesgue measurable, in a noteworthy early extension of Cohen's work, Robert Solovay established consistency relative to ZF+IC for DC+AM and AC+PM. (Also projective uniformization PU fails in one of the models involved.) For the related hypothesis RM (resp. RSM) that there is a measure with all the properties of Lebesgue measure except invariance (resp. with all the properties of Lebesgue measure except invariance, plus superadditivity), Solovay proved the consistency of RSM (which is known to imply CH#) relative to ZFC+MC, and of MC relative to ZFC+RM (which is known to imply ~CH and ~CH').

In many applications, it is convenient to work with a generalization of the notions of §1: one starts with an arbitrary set *P* and an arbitrary partial order $\leq$ on it, and defines p, q to be compatible if there is an $r \in P$ with $p \leq r$ and $q \leq r$, $D \subseteq P$ to be dense if it is nonempty and for every $p \in P$ there exists a $q \in D$ with $p \leq q$, and so on. In §1 only the special case where $\leq$ is the subset relation $\subseteq$ was considered, but the general case easily reduces to this special case, since setting $[p] = \{r \in P \mid r \leq p\}$ one has $[p] \subseteq [q]$ if and only if $p \leq q$.

4 Martin's axiom

D.A. Martin and Solovay, by a method involving reduction of several successive applications of forcing by $P \in M$, $Q \in M[G]$, $R \in M[G][H] \ldots$ to a single application, showed the consistency with ~CH (indeed with CH') of a consequence of CH called *Martin's axiom* (MA). MA is the most basic example of a *forcing axiom*, asserting that anything of a certain kind that could be forced to hold already does hold. MA asserts that for any *P* that is CCC and any set *F* of fewer than continuum many (c) dense subsets of *P*, there exists a *G* generic for *F*. (Assuming CH, c amounts to $\leq \aleph_0$, and the existence of *G* in this case was essentially proved in §1.)

To give at least one example, consider the usual order on the reals R (as in Set theory, end of §4, beginning of §5): $S \subseteq \mathbf{R}$ is *open dense* if for every rational a, c with $a < c$ there are rational d, e with $a \leq d < e < c$ such that $b \in S$ for every real *b* with $d \leq b \leq e$. It is easily proved that the intersection of finitely many open dense sets is open dense and hence not empty. It is an important theorem of modern analysis and topology that the intersection of countably many open dense sets is not empty. It is an interesting consequence of MA that an intersection of fewer than continuum many open dense sets is not empty.

To derive this consequence, let *X* be a set of fewer than continuum many open dense sets, to prove $\cap X \neq \emptyset$. Consider the set *P* of triples (a, c, Y) where $Y \subseteq X$ is finite, where a, c are rational with $a < c$, and where $b \in \cap Y$ for every real *b* with $a \leq b \leq c$. Consider the partial order on *P* given by $(a, c, Y) \leq (d, e, Z)$ if and only if $Y \subseteq Z$ and $a \leq d < e \leq c$. Then (a, c, Y) and (a', c', Y') are compatible if and only if $\max(a, a') < \min(c, c')$. For then letting $Z = Y \cup Y'$, using the fact that $\cap Z$ is open dense, there are rational d, e with $\max(a, a') \leq d < e \leq \min(c, c')$ such $b \in \cap Z$ for every real *b* with $d \leq b \leq e$, and so $(a, c, Y), (a', c', Y') \leq (d, e, Z)$.

In particular, (a, c, Y) and (a', c', Y') are always compatible if $d = d'$ and $e = e'$, and since there are only countably many possibilities for d, e, any set of pairwise incompatible elements of *P* must be countable, and the partial order is CCC. Also, for any $(a, c, Y) \in P$ and any $S \in X$, letting $Z = Y \cup \{S\}$ and using the fact that $\cap Z$ is open dense, there are rational d, e with $a \leq d < e \leq c$ such that $b \in \cap Z$ for all real *b* with $d \leq b \leq e$, and so $(a, c, Y) \subseteq (d, e, Z)$ and $S \in Z$. Thus for each $S \in X$ the set $Q_S = \{(d, e, Z) \mid S \in Z\}$ is dense in *P*. A similar argument shows that for each $n = 1, 2, 3, \ldots$, the set $R_n = \{(d, e, Z) \mid d - e < 1/n\}$ is dense in *P*. Applying MA, there is a $G \subseteq P$ generic for $\{Q_S \mid S \in X\} \cup \{R_n \mid n = 1, 2, 3, \ldots\}$.

Let α be the least upper bound of $\{a \mid (a, c, Y) \in G \text{ for some } c, Y\}$ and let γ be the greatest lower bound of $\{c \mid (a, c, Y) \in G \text{ for some } a, Y\}$. Since any elements $(a, c, Y), (a', c', Y') \in G$ are compatible, and so have $\max(a, a') < \min(c, c')$, one has $\alpha \leq \gamma$. For any *n* there is an $(a, c, Y) \in G \cap R_n$, whence

$(\alpha - \gamma) \leq (a - c) < 1/n$, and since this holds for all $n = 1, 2, 3, \ldots$, one has $\alpha = \gamma$. For any $S \in X$ there is an $(a, c, Y) \in G \cap Q_s$, and since $a \leq \alpha \leq \gamma$ one has $\alpha \in S$, and since this holds for all $S \in X$, one has $\alpha \in \cap X$, completing the proof.

Of the most interesting consequences of CH, about half follow from MA alone, about half are refutable assuming MA+~CH, which also has other interesting consequences ($\sum_2^1$-measurability, SH). By contrast, ~CH by itself has few interesting consequences. MA+~CH thus provides a rather attractive alternative to CH, though there is no generalization of MA+~CH to higher cardinals comparable to the generalized continuum hypothesis (GCH) generalizing CH. Saharon Shelah has introduced forcing axioms considerably strengthening MA+~CH and implying CH'.

5 On the proofs of the three basic lemmas

Only the briefest sketch of the highly technical proofs of the three main lemmas may be given here. It cannot be emphasized strongly enough that the best way to gain an intuitive understanding of the three main lemmas is to see how they are used in applications (as in §§2,3), which hardly ever refer back to details of their proofs.

First, while the notion of coding given in §1 is that generally needed in applications, for the proofs a different notion of coding is needed, one providing codes for *all* elements of M[G] and not just for those that are subsets of *M*. Correspondingly, a different notion of decoding is needed: $\delta_G(c) = \{\delta_G(b) \mid (p, b) \in c \text{ for some } p \in G\}$. This definition appears circular but is really an instance of a legitimate method of definition of operations on sets by $\in$-recursion, a generalization of the method of definition of operations on ordinals by $<$-recursion (or transfinite recursion, as in Set theory §4): As an operation on ordinals may be defined by defining its value for ρ in terms of its values for $\sigma < \rho$, so an operation on sets may be defined by defining its value for *c* in terms of its values for $b \in c$. With this alternate definition of decoding, M[G] is simply defined to be $\{\delta_G(c) \mid c \in M\}$, and to establish the minimality lemma what remains to be proved is that M[G] so defined is a model of ZFC. To prove this, one makes heavy use of various necessary and sufficient conditions for being a model of this or that axiom of set theory (see Set theory §7). For instance, to handle the axiom of pairing, one must show that M[G] is closed under the operation of pairing: One must show that if $a, b \in M$, then there is a $c \in M$ such that $\delta_G(c) = \{\delta_G(a), \delta_G(b)\}$. In this case it is easily checked that the required *c* is provided by $P \otimes \{a, b\}$.

Second, while the definition of forcing given in §1 is that generally needed in applications, for the proofs a different definition is needed initially, though the two definitions will be proved equivalent *eventually*. This alternate definition of *p* forces Φ is by induction on the complexity of the formula Φ: (i) *p* forces 'not Φ' if and only if no *q* with $p \subseteq q$ forces Φ; (ii) *p* forces 'Φ and Ψ' if and only if *p* forces Φ and *p* forces Ψ; (iii) *p* forces 'for all *x*, $\Phi(x)$' if and only if for all *c*, *p* forces $\Phi(c)$; and so on. What is most difficult is to define *p* forces Φ for the *simplest* formulas '$c = d$' and '$c \in d$': The definition for these simplest formulas requires a highly technical $\in$-recursion. With the alternate definition of forcing, the definability lemma is almost immediate.

Towards establishing the truth lemma, and with it the equivalence of the alternate definition of forcing to the definition given in §1, one notes that it follows from clause (i) of the alternate definition of forcing that the set $D = \{p \mid p \text{ forces } \Phi \text{ or } p \text{ forces } \ \Phi\}$ is dense, and it follows from the definability lemma that $D \in M$. Hence for any generic *G* there will be a $p \in G \cap D$. What remains to be proved is that if $p \in G$ and *p* forces $\Phi(a, b, \ldots)$, then $\Phi(\delta_G(a), \delta_G(b), \ldots)$ holds in M[G]. This is proved by induction on the complexity of the formula Φ.

See also: Logical and mathematical terms, glossary of

JOHN P. BURGESS

References and further reading

Burgess, J. (1977) 'Forcing', in K.J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 403-52.(A survey in a standard reference, derived from lectures of Solovay and emphasizing consistency proofs for combinatorial principles.)

Cohen, P.J. (1966) *Set Theory and the Continuum Hypothesis*, New York: Benjamin.(The original source for forcing, now mainly of historical interest, having been superseded in technical aspects by later, streamlined treatments.)

Jech, T. (1978) *Set Theory*, New York: Academic Press.(A comprehensive treatise, covering the applications

mentioned and many more besides, with attributions to their original authors and references to the original technical literature.)

Jech, T. (1986) *Multiple Forcing*, Cambridge: Cambridge University Press.(A semi-popular account of the forbiddingly technical work of Shelah.)

Martin, D.A. and Solovay, R. (1970) ‘Internal Cohen Extensions’, *Annals of Mathematical Logic* 2: 143-78. (The original, and a highly readable, source for Martin’s axiom.)

Forgiveness and mercy

Forgiveness and mercy are regarded as virtues in many moral and religious traditions, although different traditions will emphasize different aspects. The Christian tradition, for example, tends to emphasize purity of heart as the core of the virtue of forgiveness, whereas the Judaic tradition gives priority to the social dimension of reintegration into the covenanted community. Forgiveness involves the overcoming of anger and resentment, and mercy involves the withholding of harsh treatment that one has a right to inflict. Both allow for healing, but some critics would say that this healing may come at too high a price. Forgiveness, if carried to extremes, can lapse into servility, entailing a loss of self-respect. There are similar paradoxes associated with mercy, particularly in the context of punishment; too strong an emphasis on mercy can lead to a departure from justice. Clearly, though both forgiveness and mercy are obvious virtues, there are difficulties in putting them into practice in the complex situations that make up everyday reality.

1 Forgiveness and mercy analysed

Forgiveness and mercy are typically regarded as virtues. As virtues, they represent moral demands that may not be satisfied by actions alone, but require certain motives or dispositions of character. Obligations are different; the obligation to keep a promise has been discharged when the promise is kept, whatever the motive. The virtue of mercy is not manifested merely in an act of leniency, however, but only if that act flows from a character motivated by compassion or some other appropriate motive (love of God, perhaps). The forgiving or merciful person will typically respond to others in characteristic ways, of course, and the person receiving the response might well claim to have been forgiven or shown mercy. Ordinary language is not always precise here. What has been done exemplifies a virtue, however, only if it is the product of a merciful or forgiving heart.

With respect to forgiveness, the disposition of heart may be all that is required. Bishop Butler, one of the most insightful of all philosophical writers on forgiveness, characterized the forgiving person simply as one who has, on moral or religious grounds, forsworn resentment - forsworn the very personal response of anger or sometimes even hatred that one naturally (and perhaps properly) initially feels when one believes that one has been wronged by another. The person who no longer resents will obviously avoid certain actions - namely, those without any basis other than resentment. Since grounds other than resentment may be the basis for harsh treatment of a wrongdoer, however, the demand for harsh treatment may well survive even a sincere and complete act of forgiveness. It would be consistent, for example, for a rape victim to forgive the person who has assaulted her - to harbour no resentment at all - but still demand that he be legally punished. She might make this demand not because of any residual anger or hatred that she personally feels, but rather because she believes that punishment (for reasons of crime control, perhaps) is a socially necessary response to such serious rights violations. The person truly manifesting the virtue of forgiveness is not to be confused with the unvirtuous person who is indifferent to or complicitous in wrongdoing, and who is willing to cooperate in the further corruption of the wrongdoer or of the life of the community.

Although a good case can be made that forgiveness may be understood solely in terms of such internal dispositions as the forswearing of resentment, mercy seems different in at least two important ways: first, the person in a position to bestow mercy (for example, a judge) need not have been personally wronged and thus need not have any feelings of resentment to overcome through forgiveness; and second, although mercy as a virtue necessarily involves dispositions of character (such as compassion), it also requires - in a way that forgiveness does not - a public manifestation. One can forgive another in one’s heart of hearts, but one cannot show mercy to another in one’s heart of hearts. This is because mercy has its primary life in an institutional context (typically, but not necessarily, legal) where one has the right or duty to inflict a certain level of harsh treatment on another, but refuses to do so, imposing instead a less harsh treatment or no harsh treatment at all.

Because mercy necessarily involves a mitigation or abandonment of hard treatment, it may easily be confused with mere leniency. When the fallen knight begs for mercy, he knows that the rules of chivalric combat give the victor the right to kill the vanquished and he is asking that the victor waive that right. If the victor does waive it, and spares the life of the vanquished, we may say - speaking casually - that mercy was shown, because ordinary language often allows us to confuse mercy with leniency. Surely, however, the victor reveals the *virtue* of mercy

only if he acts from such an appropriate motive as compassion or love of God. Or consider, in a legal context, an example from Shakespeare's *Measure for Measure*: when Isabella begs Angelo to show mercy to her brother Claudio, Angelo's false promise to free Claudio in exchange for sexual intercourse with Isabella is an insincere offer of leniency, but no kind of an offer of mercy in any sense in which mercy may be regarded as a virtue.

But are forgiveness and mercy properly regarded as virtues? They are both responsive, in a moral or legal context, to genuine cases of wrongdoing for which the wrongdoer is responsible. In this they are unlike both excuse and justification, concepts with which they are sometimes confused. The person with a valid excuse (insanity, perhaps) is not responsible for wrongdoing. The person with a valid justification (self-defence, perhaps) establishes that, initial appearances to the contrary, there was (all things considered) no ultimate wrongdoing. In these cases there is nothing to resent and nothing for which to demand punishment or other harsh treatment, and thus the avoidance of resentment and demands for harsh treatment is obviously required by elementary rationality and justice. Forgiveness and mercy, however, get their main life in cases where there *is* wrongdoing for which the wrongdoer is responsible, and thus where either resentment or the demand for punishment might well seem justified. But this is what makes both forgiveness and mercy problematic candidates for moral virtues: they demand that we avoid psychological responses and actions that seem not only understandable, but perhaps actually justified by a proper respect for rights and justice.

2 Forgiveness and self-respect

There are, of course, many things to be said in favour of forgiveness: it allows us to show our genuine commitment to love and compassion (central virtues in the Christian tradition) in really hard test cases, cases where we have been personally wronged or injured; it allows us to check excessive self-importance or self-love, tendencies that incline us to overdramatize some of the wrongs done to us (a point well made by Simone Weil and others); it allows us to avoid the self-poisoning effects of resentment itself (a poison graphically described by Nietzsche (§§8-9)); it may confer considerable social benefits, such as the avoidance of feuds and vigilante activities; and it allows us to maintain our most valued personal relationships. This last benefit arises from the fact that, since each of us will sometimes wrong the people that mean the most to us, there will be times when we will want to be forgiven by those whom we have wronged. Seeing this, no rational person would desire to live in a world where forgiveness was not seen as a healing virtue.

In spite of all these splendid fruits that may be born of forgiveness, it is hard to regard it as an unqualified virtue. Although one of course wants to guard against excessive self-importance or self-love, are not certain assertions of the value of the self - the assertion of one's moral rights, for example - morally permissible, perhaps even morally mandatory? If one thinks that self-respect, in contrast to servility, is a virtue, is it not plausible to argue that one important way in which this self-respect may sometimes be shown is through resenting those persons who maliciously inflict wrongs upon us? We show our moral respect for others in feeling indignation when they are wronged. May we not show moral respect for ourselves by feeling resentment when we are wronged? And if we do not, what does this show? That we are manifesting the virtues of love and forgiveness, or simply that we are servile persons with low self-esteem, unable to demand the respect that we deserve?

Forgiveness thus creates a problem: can it be conceptualized as a virtue in such a way that it does not immediately collapse into the vice of servility? One way out of this tension, perhaps, is to make forgiveness conditional on a change in the wrongdoer: namely, repentance. When we are wronged, one of the things most threatening to our self-respect is the symbolic message that the wrongdoer seems to be conveying about us: the message that we matter less than the wrongdoer and thus may simply be used, like an object, for the wrongdoer's own purposes. This is a message that one's own self-respect prompts one to resist, and resentment of the wrongdoer is our primary emotional way of expressing such resistance. As much as we might like to follow Augustine's counsel that we should 'hate the sin and not the sinner', it is difficult to follow this counsel when the wrongdoer remains attached to the wrongdoing. But what if the wrongdoer seeks to break this attachment - as shown by sincere repentance? Then there is no longer an endorsement of the insulting and humiliating message contained in the wrongful conduct, and the wronged person can now forgive and join with the wrongdoer in condemning the act from which the wrongdoer is now severed. In this case forgiveness and self-respect seem obviously compatible. For this reason it is not surprising that repentance, which opens the door to forgiveness by allowing the wronged person a self-respecting retreat from resentment, has loomed large in the moral literature on forgiveness. It is also

not surprising that repentance (and the purity of heart at which it aims) has been regarded within some (but not all) religious traditions as a precondition of God's forgiveness of sinners.

Of course, not everyone would agree that resentment of wrongdoers is essentially tied to human self-respect. (The whole conceptual framework of resentment and self-respect may not even apply to God, of course, and thus divine forgiveness may have to be understood in totally different terms.) A person who grounds their self-respect in something other than the regard of other people will perhaps see resentment as a natural response but also as a dangerous temptation - a temptation to rely on something that is an improper basis for self-respect. Such a person may well see forgiveness - even in the absence of repentance on the part of the wrongdoer - as a way of restoring full and properly founded self-respect. Persons who do not derive their self-respect from treatment by other people may, of course, rely on a variety of different sources and worldviews to support a sense of their own worth - for example, the moral elitism of a Socrates ('Why should I care if the ignorant seek to harm me?') or the belief of Christians that their self-respect is grounded in the fact that God loves them.

3 Mercy and justice

Mercy is also a concept that is capable of generating tension and paradox; this is particularly so in the context of punishment, where it initially seems most at home. If a certain punishment is required by justice, representing what the wrongdoer in justice deserves, then it would seem to be an injustice to depart from that punishment. We are fond of using the slogan that 'justice should be tempered with mercy', yet it is unclear that such a slogan represents morally wise counsel. If it is simply a somewhat misleading way of demanding proper individuation to guarantee that justice is finely tuned and thus truly done, then one can hardly have any deep quarrels with it. If it counsels an actual departure from justice (an injustice), however, then what it counsels seems manifestly unvirtuous. It is for this reason that Kant feared the role of compassion in the criminal law and argued that the sovereign's power of pardon (a dramatic kind of legal mercy) is one of the most dangerous of all sovereign powers because it might tempt the sovereign, out of compassion, to avoid a genuine duty: namely, the doing of justice. It is manifestly illegitimate, Kant argues, for the sovereign or judge to pursue personal and private commitments to compassion when such individuals have explicitly taken on the public responsibility of administering justice.

The tension - perhaps even contradiction - between justice and mercy produces paradoxes even with respect to the idea of divine mercy. Anselm, for example, found it paradoxical that God, who is committed to justice, could through his mercy spare the wicked from the eternal punishment that they deserved.

Perhaps the way out of this paradox, at both the human and the divine level, is to distinguish between two different models of mercy: the criminal law model and the private law model. On the criminal law model, a judge has an obligation (a role responsibility) to enforce just laws by imposing the just sanctions specified. The responsibility of the office precludes appeals to personal compassion, since the very essence of the job is defined in terms of certain specified duties. This criminal law model would, of course, have serious problems as a model for divine justice. God is presumably not like a judge in a criminal case, having the duty to enforce rules that are independent of his will, but is in some sense the author of the rules that are enforced. God is thus not accountable to a constituency in the way a human judge is, and thus does not experience anything that might plausibly be regarded as role responsibility or role conflict. Anselm, however, may have been tempted to see God occupying such a role; his paradox seems to depend upon seeing God caught in the very kind of dilemma that might plague a human judge in a criminal case.

A private law model of mercy, however, might allow us to avoid some of the tensions between mercy and justice at both the human and the divine level. According to this model, mercy is best illustrated not by a judge in a criminal case, but by a private party in a private lawsuit. Shakespeare's *The Merchant of Venice* nicely illustrates this private law model in its portrayal of Portia pleading with Shylock to show mercy to Antonio, a man who has defaulted on a contract and who now, according to the terms of the contract, owes Shylock 'a pound of flesh'. Unlike a judge, Shylock has no duty to enforce any legal rules. Rather, the legal rules of contract give him certain rights, rights that he is free to waive if he chooses. And it is such a waiver, motivated by compassion, that Portia is asking him to consider - a waiver that, since he has only rights and no duties in this situation, could not open him up to a valid charge of injustice or dereliction of duty.

So too, one might say, for God. If God is properly conceived as a being who mainly has rights over his creatures

rather than external (non-self-imposed) duties to them, then perhaps he may show mercy - waiving his right to consign a person to eternal damnation - without being validly charged with injustice. We may now begin to see our way out of at least one paradox about mercy. Others, however, may still be present - for example, what might be called the 'equal protection' paradox raised by Anselm when he wondered how God, a rational being, can consistently show mercy to some who are unworthy while punishing others who are equally unworthy.

Both forgiveness and mercy, because of the tensions they generate, will no doubt remain problematic concepts that will continue to generate rich philosophical and theological discussion. It is an obvious truth, of course, that the virtuous person is one lacking in neither charity (the primary foundation of mercy and forgiveness) nor justice (the primary foundation for resentment and punishment). What is by no means obvious, however, is how these virtues are to be brought into harmony in all the actual complex cases that arise in the moral lives of individuals and communities.

See also: Atonement; Grace; Justification, religious; Rectification and remainders; Salvation; Self-respect; Sin; Virtue ethics; Virtues and vices

JEFFRIE G. MURPHY

References and further reading

Adams, M.M. (1991) 'Forgiveness: A Christian Model', *Faith and Philosophy* 8 (3): 277-304.(Criticizes the work of Jeffrie Murphy and others who have argued that forgiveness may be incompatible with self-respect.)

Butler, J. (1722) *Sermons*, Oxford: Clarendon Press, 1897.(Sermon VIII, 'Upon Resentment', and Sermon IX, 'Upon Forgiveness of Injuries', present the essence of Butler's account of forgiveness.)

Card, C. (1972) 'On Mercy', *Philosophical Review* 81: 182-207.(Argues that mercy is a part of justice (on a sophisticated theory of justice) and not an autonomous moral virtue.)

Hill, T.E., Jr (1973) 'Servility and Self-respect', *The Monist* 57: 87-104.(Argues, on Kantian grounds of duty to self, that servility is a moral vice.)

Kant, I. (1797) *The Metaphysics of Morals*, trans. M. Gregor, Cambridge: Cambridge University Press, 1991, 140-45, 168-9.(Develops Kant's theory of justice and employs it to defend retributive punishment and to oppose pardons for those convicted of crime.)

Kolnai, A. (1973-4) 'Forgiveness', *Proceedings of the Aristotelian Society* 74: 91-106.(Argues that forgiveness can be a virtue only if it does not involve complicity in wrongdoing.)

Moore, K.D. (1989) *Pardons: Justice, Mercy, and the Public Interest*, Oxford: Oxford University Press.(Presents, in a generally Kantian framework, a theory of when legal pardons may be justified.)

Murphy, J.G. (1997) 'Repentance, Punishment and Mercy', in A. Brien (ed.) *The Quality of Mercy*, Value Inquiry Book Series, Amsterdam: Rodopi; also in A. Etzioni (ed.) *Repentance*, Lanham, MD: Rowman & Littlefield. (Presents a detailed analysis of the concept of repentance and argues that repentance can be relevant not simply to mercy and the reduction of punishment, but to the justification of punishment itself.)

Murphy, J.G. and Hampton, J. (1988) *Forgiveness and Mercy*, Cambridge: Cambridge University Press. (Hampton's two chapters seek to develop, with philosophical rigour, a Christian defence of forgiveness and mercy, while Murphy's three chapters explore the sceptical case against the claim that forgiveness and mercy are virtues.)

Murphy, J.G. and Morris, H. (1988) 'Exchange: Forgiveness and Mercy', *Criminal Justice Ethics* 7 (2): 3-22. (Murphy summarizes the view of forgiveness and mercy he develops in *Forgiveness and Mercy* and Morris builds his case against this from remarks on unjustified self-importance made by Simone Weil in her book *Gravity and Grace*.)

Newman, L.E. (1987) 'The Quality of Mercy: On the Duty to Forgive in the Judaic Tradition', *Journal of Religious Ethics* 15: 155-72.(Stresses the idea that, in the Judaic tradition, forgiveness is more a matter of community reintegration than purity of heart.)

Nietzsche, F. (1887) *On The Genealogy of Morals*, trans. W. Kaufmann, New York: Random House, 1967. (Explores, among many other things, the destructive role that resentment (*ressentiment*) plays in the moral life. See, for example, pages 36-9.)

Nussbaum, M. (1993) 'Equity and Mercy', *Philosophy and Public Affairs* 22 (2): 83-125.(Builds a case for mercy and empathetic individuation from Seneca's essays 'On Anger' and 'On Clemency'.)

Quinn, P.L. (1978) *Divine Commands and Moral Requirements*, Oxford: Oxford University Press.(Contains, in the last chapter, a fine discussion of Saint Anselm's paradoxes of divine mercy.)

Twambley, P. (1976) 'Mercy and Forgiveness', *Analysis* 36: 84-90.(Introduces the distinction between the

criminal law and private law models of mercy.)

Formal and informal logic

Formal logic abstracts the form of an argument from an instance of it that may be encountered, and then evaluates the form as being valid or invalid. The form is the important thing, rather than the concrete instance of the form. Informal logic, on the other hand, evaluates how an argument is used in a given context of conversation. This more practical, real-world orientation requires more judgment in interpreting what a given text of discourse should be taken to argue.

Formal logic abstracts from the context of use of an argument by seeing an argument as an instance of a form of argument, composed of constants and variables. The constants are defined in a precise way that makes any reference to context unnecessary (see Logical constants). The variables are blanks that are filled in by the component parts of the argument under consideration. Take the argument 'Jane and Dick will not both be elected; Jane will be elected; therefore Dick will not be elected'. In the modern formal logic of propositions, 'and' and 'not' are defined as constants, and 'Jane will be elected' and 'Dick will be elected' are represented using propositional variables, *A* and *B*. The constants 'and' and 'not' are defined as follows:

'*A* and *B*' is true only if both *A* and *B* are true; otherwise it is false.
'not *A*' is true if *A* is false; and false if *A* is true.

Symbolized thus, the argument above about Jane and Dick has the following form:

not(*A* and *B*)
A
Therefore not *B*.

This form of argument is 'valid', meaning that every single instance of it is such that if both premises are true, the conclusion must also be true. Thus formal logic guarantees that a valid argument will never take you from true premises to a false conclusion. This is simply a matter of the form of the argument and the way the constants have been defined. Hence formal logic is an abstract, mathematical discipline - a study of forms.

Formal logic of this mathematical sort, using constants and variables, is a relatively modern development, but the roots of it can be traced back to Aristotle's theory of the syllogism (see Aristotle §5; Logic, ancient §§1-3). A 'syllogism' is an argument made up of three propositions put together in a particular form, of which the following is an example: 'All bears climb trees; all grizzlies are bears; therefore, all grizzlies climb trees'. This kind of argument was ruled valid in Aristotle's theory: *if* the premises are true, then the conclusion must be true too. This does not imply that the premises *are* true. Validity, you could say, is a conditional concept in Aristotle's theory, and the same can be said in modern logic.

Modern formal logic has been centrally concerned with two kinds of arguments. Propositional logic has to do with arguments that are valid or not by virtue of certain constants, such as 'and', 'not' and 'if…, then…', which function as connectives between propositions. Quantifier logic is a sophisticated and powerful system of formal logic that not only encompasses the limited varieties of arguments treated in Aristotle's theory of the syllogism, but also many other more complex arguments depending on 'all' and 'some' (see Quantifiers; Predicate calculus).

Informal logic was also founded as a systematic subject by Aristotle, principally in his *Topics* and *Sophistical Refutations* (see *Topica et Sophistici Elenchi*), but never received the same consistent and more serious attention that formal logic did in the subsequent history of logic. Consequently, it was never developed as a scientific subject in the way that formal logic was. However, it survived in a practical form, in logic textbooks and manuals used for instruction in universities throughout the Middle Ages and into the modern period. Only recently, with the advent of argumentation theory, has there come to be a relatively serious and sustained scholarly attempt to develop informal logic as a subject of research. Yet even at the end of the twentieth century, informal logic is a more practical subject which addresses contextual factors of how arguments are actually used in everyday conversation in a given discourse. Unlike formal logic, informal logic cannot abstract from the context of use of an argument, and must evaluate it as subject to an interpretation of the speaker's discourse as a part of a conversational exchange.

Both formal and informal logic are concerned primarily with the task of evaluating arguments as correct or incorrect, according to standards. Formal logic, as noted above, works by identifying forms and proving them valid or invalid by means of some method of calculation independent of the context in which the argument was used. Applying the formal test to a real argument does involve some informal determining of what kind of argument was meant by the speaker, what its most precise form is, and so forth. But in informal logic such questions of interpretation are a much more central part of the task of evaluation.

Informal logic has to do with the use of an argument in a conversational setting where participants are supposed to contribute to the conversation by making moves appropriate to the current stage of the conversation (see Grice 1975; Pragmatics; Implicature). Different types of conversation can be recognized. For example, van Eemeren and Grootendorst (1984) cite one very important type where the purpose is to resolve a conflict of opinions by verbal means, using arguments. Generally an argument is correct if it contributes to the goal of the conversation, and incorrect if it fails to do so. Some particularly deceptive or misleading types of arguments can even actively interfere with the goal of a conversation, and the most common and dangerous of these are called fallacies (see Fallacies).

See also: Logical and mathematical terms, glossary of

DOUGLAS WALTON

References and further reading

Aristotle (*c*.330 BC) *Topica et Sophistici Elenchi (Topics and Sophistical Refutations)*, trans. W.A. Pickard-Cambridge, ed. W.D. Ross, New York: Oxford University Press, 1958.(Historical.)

Eemeren, F.H. van and Grootendorst, R. (1984) *Speech Acts in Argumentative Discussions*, Dordrecht: Foris. (Account of argumentation theory.)

Grice, H.P. (1975) ‘Logic and Conversation’, in D. Davidson and G. Harman (eds) *The Logic of Grammar*, Encino, CA: Dickenson.(Account of implicature.)

Jeffrey, R.C. (1967) *Formal Logic: Its Scope and Limits*, New York: McGraw-Hill, 2nd edn, 1981.(General introduction to formal logic.)

Walton, D. (1989) *Informal Logic*, Cambridge: Cambridge University Press.(General introduction to informal logic.)

Formal languages and systems

Formal languages and systems are concerned with symbolic structures considered under the aspect of generation by formal (syntactic) rules, that is, irrespective of their or their components' meaning(s). In the most general sense, a formal language is a set of expressions. The most important way of describing this set is by means of grammars. Formal systems are formal languages equipped with a consequence operation yielding a deductive system. If one further specifies the means by which expressions are built up (connectives, quantifiers) and the rules from which inferences are generated, one obtains logical calculi of various sorts, especially Frege-Hilbert-style and Gentzen-style systems.

1 Expressions and grammars

A formal language is a set $\mathcal{L}$ of expressions. Expressions are built up from a finite set Σ of atomic symbols or atoms (the 'alphabet' of the language) by means of certain constructors. Normally one confines oneself to linear association (concatenation) '$\circ$' as the only constructor, yielding 'strings' of atoms, also called 'words' (over Σ), starting with the empty word ϵ. Mathematically, the structure obtained is the semi-group freely generated by the alphabet Σ. When $a_1 \circ a_2 \circ \ldots \circ a_n$ with atoms $a_1, a_2, \ldots, a_n$ is written in the usual way as $a_1 a_2 \ldots a_n$, it is assumed that the decomposition of this expression into atoms is unique.

It should be noted that languages based on more than one constructor may well be considered. An example would be Frege's two-dimensional *Begriffsschrift*, which may be viewed as a formal language based on constructors for horizontal and vertical alignment (see Frege 1879). However, since any constructor can be written linearly as a function symbol applied to arguments, the notion of formal languages as based on linear concatenation is sufficiently general. The atomic symbols which form the basis of a formal language are themselves abstract entities, in contrast to the tokens which constitute their individual realizations.

A formal language $\mathcal{L}$ can be investigated from various points of view. One possibility is to specify a device $\mathcal{A}$ that recognizes exactly those words over Σ which belong to $\mathcal{L}$, that is, which enters a certain state if and only if a word of $\mathcal{L}$ is given as an input. When this occurs we say that $\mathcal{A}$ *accepts* $\mathcal{L}$. Such a device is called an 'automaton' (in the mathematical, not the physical sense). The relationship between languages and automata accepting them is studied in automata theory.

A formal grammar $\mathcal{G}$ for $\mathcal{L}$ describes a way to generate the words of $\mathcal{L}$ by means of formal rules, beginning with a special start symbol *S*. A rule (also called a 'production') has the form $u \rightarrow v$, which expresses that in any word of the form $w_1 u w_2$ the part *u* may be replaced by *v*, yielding $w_1 v w_2$. In addition to the atoms of the alphabet Σ under consideration (which are called 'terminals' in formal language theory), rules may contain variables (called 'non-terminals'), among which *S* is distinguished. For example, the grammar

$$S \rightarrow ASA$$

$$S \rightarrow a$$

$$A \rightarrow b$$

generates all words $b \ldots bab \ldots b$ over the alphabet $\{a, b\}$ with equally many instances of *b* on each side of *a*.

Formal languages may be classified by the type of grammar generating them. For example, a language $\mathcal{L}$ is called 'context-free' if it can be generated by a context-free grammar, which is one whose productions always have just a single variable on the left side (as in the example just given). Context-free grammars play a distinguished role in the construction of artificial languages (such as programming languages). However, for the study of (fragments of) natural languages, more complicated ('context-sensitive') types of grammar must be considered. Formal grammars have the same expressive power as computable functions: the computation by a Turing machine can be described by means of a formal grammar; and the generation of words in a formal grammar can be simulated by a Turing machine.

2 Deductive systems

A formal system is based on a formal language $\mathcal{L}$, endowing it with a 'consequence operation' $\mathcal{C}$. This operation

can be specified at different levels of abstraction. In the most general sense, $\mathcal{C}$ is just an arbitrary function transforming subsets of $\mathcal{L}$ into subsets of $\mathcal{L}$. $\mathcal{C}$ is said to be an 'inference operation' if the set of consequences of a set *X* comprises at least *X*.

$$X \subseteq \mathcal{C}(X)(\text{inclusion})$$

The pair $\langle \mathcal{L}, \mathcal{C} \rangle$ is then called an 'inference system'. It is called a 'closure system' if furthermore

$$\mathcal{C}(\mathcal{C}(X)) \subseteq \mathcal{C}(X)(\text{idempotence})$$

and

$$X \subseteq Y \Rightarrow \mathcal{C}(X) \subseteq \mathcal{C}(Y)(\text{monotonicity})$$

are fulfilled. It is called a 'deductive system' if the consequences of a set *X* can be obtained from the finite subsets of *X*, that is, if in addition to the three conditions mentioned,

$$\mathcal{C}(X) \subseteq \bigcup\{\mathcal{C}(Y) \colon Y \subseteq X, Y \text{ finite}\}(\text{compactness})$$

holds. Equivalently, formal systems can be described by a consequence relation $X \vdash A$ between subsets of $\mathcal{L}$ and expressions of $\mathcal{L}$. The four conditions mentioned then become

$$X \cup \{A\} \vdash A$$
$$X \vdash A \Rightarrow X \cup Y \vdash A$$
$$(X \vdash A \text{ for all } A \in Y \text{ and } Y \cup Z \vdash B) \Rightarrow X \cup Z \vdash B$$
$$X \vdash A \Rightarrow Y \vdash A \text{ for some finite } Y \subseteq X.$$

If we confine ourselves to consequences of finite sets *X*, *Y*, *Z*, which in the case of deductive systems is appropriate, these conditions are equivalent to

to $\{A\} \vdash A(\text{identity}) \quad X \vdash A \Rightarrow X \cup \{B\} \vdash A(\text{thinning}) \quad (X \vdash A \text{ and } \{A\} \cup Z \vdash B) \Rightarrow X \cup Z \vdash B(\text{cut})$

which are the basic structural (that is, logic-free) principles of Gentzen-style sequent systems (see Gentzen 1935).

Deductive systems can be given by means of a set Δ of inference rules *R*, where an inference rule is an $(n+1)$-ary relation $R \subseteq \mathcal{L}^n \times \mathcal{L}$. Any $(A_1, \ldots, A_n, B) \in R$ (also written as $A_1, \ldots, A_n \Rightarrow B$) is called an instance of *R*, and $A_1, \ldots, A_n$ are said to be the premises and *B* the conclusion of that instance. If $n = 0$ then the rule $\Rightarrow A$ is called an axiom. If we define $\mathcal{C}_\Delta(X)$ to be the smallest set of expressions in $\mathcal{L}$ containing *X* and closed under the rules in Δ (that is, if $\{A_1, \ldots, A_n\} \subseteq \mathcal{C}_\Delta(X)$, then $B \in \mathcal{C}_\Delta(X)$ for any instance $A_1, \ldots, A_n \Rightarrow B$ of a rule in Δ), then $\langle \mathcal{L}, \mathcal{C}_\Delta \rangle$ is a deductive system, called the rule-based system with respect to Δ.

Conversely, the consequence relation of any deductive system defines a set of rules with respect to which this system can be viewed as a rule-based system. Like grammars, rule-based systems are as powerful as Turing machines: they can be used to express any algorithm.

3 Logical calculi

In logic one considers deductive systems over a language $\mathcal{L}$ whose expressions, called formulas, are built up from certain basic expressions by means of functions, predicates and logical operators. In propositional logic, which for reasons of simplicity is considered here, formulas are built up from a subset $\mathcal{P}$ of $\mathcal{L}$ (the set of propositional letters or propositional variables) and a finite set $\mathcal{S}$ of symbols of the alphabet Σ of $\mathcal{L}$ (the set of propositional connectives). Prominent connectives are 'not' ($\sim$), 'and' ($\wedge$), 'or' ($\vee$) and 'implies' ($\supset$). The two principal ways of specifying a consequence operation $\mathcal{C}$ for such a system are (1) by presenting it as a rule-based system and (2) by formally describing its consequence relation $\vdash$. Other approaches modify or extend the notion of consequence by considering (3) multiple consequences, (4) consequence relations with restricted structural principles ('substructural logics') or (5) the fine structure of proofs.

(1) For a *rule-based system* a set of axioms and a set of inference rules are given. Axioms for a fragment of propositional logic based on conjunction and implication (the so-called 'positive implication-conjunction logic') are, for example, all formulas of the form

$A \supset (B \supset A)$
$(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$
$(A \supset (B \supset C)) \supset ((A \wedge B) \supset C)$
$((A \wedge B) \supset C) \supset (A \supset (B \supset C))$.

Inference rules are all instances of the following schema.

$$A, A \supset B \Rightarrow B(modus\ ponens)$$

Rule-based systems of this kind are often called Frege-Hilbert-style systems. Typically, they include a relatively long list of axioms and just a few inference rules (in ordinary propositional logic, just *modus ponens*). The consequence relation ‘ $X \vdash_{\mathcal{C}} A$’ corresponding to the consequence operator $\mathcal{C}$ is called ‘derivability from assumptions’.

(2) In contradistinction to this approach, Gentzen-style sequent systems formulate the desired properties of a consequence relation ‘ $X \vdash_{\mathcal{C}} A$’ as formal rules of inference. In a sequent system of propositional logic based on the connectives ‘ $\wedge$’ and ‘ $\supset$’ the above rules of identity, thinning and cut are typically postulated as logic-free structural rules, in addition to the following logical rules.

to $X \cup \{A\} \vdash B \Rightarrow X \vdash A \supset B(\supset\text{-introduction right})$ $X \vdash A,\ X \cup \{B\} \vdash C \Rightarrow X \cup \{A \supset B\} \vdash C(\supset\text{-introduction left})$ $X \vdash A,\ X \vdash B \Rightarrow X \vdash A \wedge B(\wedge\text{-introduction right})$ $X \cup \{A, B\} \vdash C \Rightarrow X \cup \{A \wedge B\} \vdash C(\wedge\text{-introduction left})$

Formally, this is just a specification of a rule-based system with respect to a more complicated language. Rather than a set of formulas $\mathcal{L}$ over $\mathcal{P}$ and $\mathcal{S}$ one considers a language $\mathcal{L}'$, whose expressions are ‘sequents’ of the form ‘ $X \vdash A$’, where $A \in \mathcal{L}$ with *X* finite. The deductive system $\mathcal{C}'$ over $\mathcal{L}'$ is then given by the rules stated. (Actually, to capture the idea of finite *sets X* in *expressions* ‘ $X \vdash A$’, one has to consider lists instead and to add special inference rules.) Intuitively, $\langle \mathcal{L}', \mathcal{C}' \rangle$ states features of the relation $\vdash_{\mathcal{C}}$ which are intended to describe properties of the connectives of $\mathcal{S}$. So a Gentzen-style sequent system is a rule-based system with a consequence relation $\vdash_{\mathcal{C}'}$ that can be viewed as a description of a deductive system with a consequence relation $\vdash_{\mathcal{C}}$ such that $X \vdash_{\mathcal{C}} A$ if and only if $\emptyset \vdash_{\mathcal{C}'} (X \vdash A)$.

(3) In the Gentzen tradition, more complicated types of consequence relations ‘ $\vdash$’ have been investigated that do not completely fit the pattern described. One such approach is to consider consequence relations with a set of formulas appearing on the right side of ‘ $\vdash$’ (‘multiple-conclusion logics’ or, sometimes, ‘Scott consequence relations’; see Multiple-conclusion logic). The intended reading is disjunctive: ‘ $X \vdash Y$’ means that under the assumptions *X* at least one element of *Y* holds (although that cannot be formally derived!). (Gentzen (1935) has proposed this idea to formulate sequent calculi for classical logic in which each connective obtains a ‘natural’ meaning via introduction rules for it on the right or left side of the turnstile without the need to consider special axioms or rules to guarantee the validity of, for example, $A \vee {\sim}A$ or ${\sim}{\sim}A \supset A$. Negation and disjunction rules in this system can be formulated as follows.

(Note the duality between these $\vee$-rules and the $\wedge$-rules formulated above, the right and left sides of the sequents being interchanged.) The *tertium non datur* (excluded middle) $\{\vdash A \vee {\sim}A\}$ is then obtained from $\{A\} \vdash \{A\}$ via $\vdash \{A, {\sim}A\}$. It is obvious that this notion of consequence has strong elements of symmetry which are characteristic of classical logic.

(4) Another approach is to restrict the structural principles underlying deductive systems. If in $X \vdash A$ we consider *X* no longer as a set but as a list of formulas, we may discuss whether in all cases we want to assume that elements of *X* are ‘permutable’ (that is, whether $Y, A, B, Z \vdash C \Rightarrow Y, B, A, Z \vdash C$ holds), or whether two identical formulas in *X* may be ‘contracted’ to a single one (that is, whether $Y, A, A, Z \vdash B \Rightarrow Y, A, Z \vdash B$ holds), or whether we want to assume or reject the principle of thinning (see above).

Such considerations, which apply analogously to the multiple-conclusion case, lead to so-called ‘substructural’ logics, which have become important in computer science and linguistics (but not only there); in particular, relevance logics (logics without thinning), contraction-free logics (in which the multiplicity of formulas counts), linear logics (without thinning and contraction) and Lambek logics without thinning, contraction and permutation). If at the level of general consequence relations we consider logics without monotonicity, we enter the general field of non-monotonic logics, which is of special interest in artificial intelligence, since it helps to

frame many common forms of reasoning (see Non-monotonic logic).

(5) In proof theory (see Proof theory) one is often interested in the structure of the derivation that leads from a set of assumptions *X* to a conclusion *A* and in criteria for when two proofs of the same 'derivability fact' $X \vdash A$ are to be seen as identical - rather than in whether $X \vdash A$ holds or not. Significant cases are calculi of natural deduction (see Natural deduction, tableau and sequent systems). With respect to derivability, they can be described as Gentzen sequent systems in the sense above. However, crucial features of the fine structure of derivations, such as the fact that different occurrences of the same formula as an assumption may be treated differently, cannot be captured in that way.

One should be aware that in cases (3)-(5), as well as in other cases, in which extensions or modifications of the usual concept of a deductive system are considered with respect to some 'object logic', the system in which this object logic is described has itself the structure of a deductive system in the ordinary sense. Sequent calculi for a multiple-conclusion consequence relation '$\vdash$', perhaps with substructural features, are themselves rule-based systems. Even the structure of natural deduction proofs can be characterized by a rule-based system, namely a system of type assignment for λ-terms. Therefore, it is not so much the specific type of consequence relation studied which makes logical systems formal systems, but rather the fact that the variety of logical systems can be described in terms of formal systems in the very specific sense of a rule-based system.

See also: Gentzen, G.K.E.; Logical and mathematical terms, glossary of; Turing machines

HEINRICH HERRE
PETER SCHROEDER-HEISTER

References and further reading

Došen, K. and Schroeder-Heister, P. (eds) (1993) *Substructural Logics*, Oxford: Clarendon Press. (Overview of logics with restricted structural rules.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. J. van Heijenoort, '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967. (The first formulation of a complete axiom system for first-order logic.)

Gentzen, G. (1935) 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* 39: 176-210, 405-565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969; trans. 'Investigations into Logical Deduction', in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969.(Gentzen's doctoral thesis, in which he created the notions of natural deduction and sequent calculus.)

Harrison, M.A. (1978) *Introduction to Formal Language Theory*, Reading, MA: Addison-Wesley. (Like Salomaa (1973), an advanced textbook of formal language theory.)

Hopcroft, J.E. and Ullman, J.D. (1979) *Introduction to Automata Theory, Languages and Computation*, Reading, MA: Addison-Wesley.(Elementary textbook of formal language theory.)

Kleene, S.C. (1952) *Introduction to Metamathematics*, Amsterdam: North Holland, 1974. (Classic textbook of mathematical logic with emphasis on logical calculi.)

Salomaa, A. (1973) *Formal Languages*, New York: Academic Press.(Like Harrison (1978), an advanced textbook of formal language theory.)

Takeuti, G. (1975) *Proof Theory*, Amsterdam: North Holland.(Textbook of Gentzen-style proof theory.)

Tarski, A. (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press; repr. Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(Includes Tarski's basic logical papers, particularly 'On Some Fundamental Concepts of Metamathematics' (1930), in which he defined the abstract notion of consequence for the first time.)

Troelstra, A.S. and Schwichtenberg, H. (1996) *Basic Proof Theory*, Cambridge: Cambridge University Press. (Textbook of Gentzen-style natural deduction and sequent systems.)

Wójcicki, R. (1988) *Theory of Logical Calculi: Basic Theory of Consequence Operations*, Dordrecht: Kluwer.(A comprehensive development of the theory of consequence relations.)

Formalism in art

Formalism in art is the doctrine that the artistic value of a work of art is determined solely by the work's form. The concept of artistic form is multiply ambiguous, however, and the precise meaning of formalism depends upon which sense of form it operates with. There are two main possibilities. The first understands form as the structure of a work's elements, the second as the manner in which it renders its 'content'. If form is understood as structure, formalism is still ambiguous: understood one way, it has never been denied; understood another way, it is untenable. If form is understood as manner, formalism is false.

1 Formalism and artistic value

Formalism as an artistic doctrine is a claim about the value of a work of art *as* a work of art. Let us call this 'artistic value' (see Art, value of §2). Formalism is the thesis that the artistic value of a work of literature, music, architecture or any other art is determined by the work's form, and by nothing else. This thesis presupposes that any work of art has two or more aspects, a formal aspect and one or a number of non-formal aspects, and implies that artistic appreciation should be focused exclusively on the form of the work. Note that it does not follow from the thought that a work's artistic value is determined solely by its form that all other aspects of the work are irrelevant to its value. For some or all of these other aspects may determine or contribute to form. If so, formalism does not require the aesthetic spectator to ignore them; rather, it requires the spectator to attend to them only in so far as it is necessary to perceive the work's form - only for this reason, and not for their own sakes.

But what is the form of a work of art? Is there a unitary notion of form, or does the idea of form change from art to art, or even from genre to genre?

2 Form as structure

'Form' often signifies the way in which the elements of an object are related to one another, that is to say, the structure of the elements. If 'form' is understood in this sense, formalism maintains the aesthetic irrelevance of the nature of the elements of a work of art, locating a work's value only in how it structures these elements: it does not matter what the elements of a work are, except in so far as only those elements, or elements of a similar kind, can be combined in the precise structure of the work. This version of formalism requires that a work of art is a composition of elements. But what are supposed to be the elements that compose works of art in the various art forms? The most obvious candidates for the elements of poems and musical works are, respectively, the words, and the notes, chords and rests that compose them. But it is not so clear what the elements of a play, a dance, a picture or a building might be supposed to be.

Let us shelve this difficulty, however, and concentrate on poetry and music. Given that a poem is a composition of words, and a piece of music a composition of notes, chords and rests, the crucial question is whether a work's form is separable from its elements. If a work's form is *its elements as concatenated in the work*, its form is not separable from its elements. Understood in this way, it is undeniable that a work's artistic value is determined not by its elements considered in themselves, but by its elements as related to one another in the work, that is, by its form. This variety of formalism is harmless. If, on the other hand, the form of a work of art can be shared with another work which has different elements, it is uncertain what the form of a work is, and there is no plausible candidate that would make formalism acceptable. To see this, consider what the form common to two poems or two symphonies might be. Whatever answer you give, it should be clear that the artistic value of a poem or symphony is not determined solely by its form.

3 Form as manner

The idea of artistic form as a structure of elements is not, however, the sole or even principal conception of form at work in the doctrine of formalism. Usually formalists allege the aesthetic irrelevance of one particular aspect of a work of art, which aspect changes from art form to art form. In architecture, form is principally opposed to function; in music, to the expression of emotion and other affective states; in the pictorial arts, to pictorial content (what is depicted); in poetry, to meaning or what is said. Accordingly, formalism in architecture maintains the artistic irrelevance of a building's function, except in so far as it contributes to the building's form; in music, of the expressive aspects of a work; in the pictorial arts, of what a picture depicts; in poetry, of what is said, as opposed

to how it is said. So it is the manner in which a function is realized, a subject depicted, an emotion or thought expressed - not the function, subject, emotion or thought - that determines the value of a work. Or so formalists claim.

What might lead one to embrace formalism? Let us call the non-formal aspect of a work of art that formalism dismisses F. A train of thought that often underlies formalism is this. A fine work of art can lack F. A poor one can possess F. Therefore, the value of a work of art that possesses F cannot be in any way due to its possessing F. For example: there are good paintings that do not depict significant subjects; there are poor paintings that do; hence, the significance of a painting's subject is irrelevant to its artistic value. Again: there are fine buildings that are not churches; there are poor buildings that are; hence, a building's being a church is irrelevant to its value as a work of art. But this line of thought is fallacious, as can be most easily seen by considering a parallel argument: there are good soccer players who lack a powerful kick; there are poor soccer players who do not; hence, the possession of a powerful kick is irrelevant to being a good soccer player. On the contrary, someone who possesses a powerful kick might be a good player partly because of that ability - it might be one of his strengths. The formalist argument mistakenly concludes from the fact that a certain feature - which it usually calls 'content' - is neither necessary nor sufficient for artistic value, that it is irrelevant to artistic value.

Formalists are motivated by a concern to preserve the autonomy of artistic value, but mistakenly believe that the only way to achieve this is by insulating artistic value from any aspect of a work - function, subject, emotion, thought - that has an extra-aesthetic reference. It is true that any aspect of a work of art contributes to the work's artistic value only as it is realized in the work. For example, the depicted content of a picture does not itself determine the picture's artistic value, which can easily be seen from the fact that two pictures can depict the same state of affairs but have unequal artistic values. But it does not follow from this that it is only how a picture depicts its content, not what that content is, that is relevant to its value as a picture. The consideration licences no stronger claim than that it is a picture's content *as depicted* that determines its value as art. Likewise, the fact that two buildings of unequal value can have the same function, two musical passages express the same emotion or mood, or two poems say the same thing shows no more than that a work's artistic value is determined not by its non-formal aspect, considered in itself, but only as realized in the building's structure, as expressed in the music's notes or in the arrangement of words that constitutes the poem. The truth of the matter is that the value of Pope's poetry is inseparable from the quality of the thought it expresses; the value of Poussin's *Echo and Narcissus* is inseparable from its being a depiction of that mythological story; and so on.

See also: Hanslick, E.; Kant, I. §12

MALCOLM BUDD

References and further reading

Bell, C. (1961) *Art*, London: Arrow Books.(The best-known statement of formalism with respect to visual art.)

Bradley, A.C. (1909) 'Poetry for Poetry's Sake', in *Oxford Lectures on Poetry*, London: Macmillan.(A sophisticated attempt to undermine the debate about formalism as applied to poetry.)

Hanslick, E. (1885) *Vom Musikalisch-Schönen*, 7th edn, trans. G. Cohen, *The Beautiful in Music*, London and New York: Novello, Ewer & Co, 1891.(The classic statement of a formalist view of music.)

Isenberg, A. (1973) 'Formalism', in *Aesthetics and the Theory of Criticism*, Chicago, IL, and London: University of Chicago Press.(An incisive critique of formalism.)

Kant, I. (1790) *The Critique of Judgment*, trans. J.C. Meredith, Oxford: Clarendon Press, 1957.(Part I containsKant's celebrated account of aesthetic judgment, which emphasizes the form of the object of the judgment. It is a contentious matter how formalisticKant's treatment of art is.)

Foucault, Michel (1926-84)

Michel Foucault was a French philosopher and historian of thought. Although his earliest writings developed within the frameworks of Marxism and existential phenomenology, he soon moved beyond these influences and developed his own distinctive approaches. There is no overall methodological or theoretical unity to Foucault's thought, but his writings do fall into several main groups, each characterized by distinctive problems and methods. In his early studies of psychiatry, clinical medicine and the social sciences, Foucault developed an 'archaeology of knowledge' that treated systems of thought as 'discursive formations' independent of the beliefs and intentions of individuals. Foucault's archaeology displaced the human subject from the central role it played in the humanism which had been dominant since Kant. While archaeology provided no account of transitions from one system to another, Foucault later introduced a 'genealogical' approach, which seeks to explain changes in systems of discourse by connecting them to changes in the non-discursive practices of social power structures. Like Nietzsche's, Foucault's genealogies refused all comprehensive explanatory schemes, such as those of Marx or Freud. Instead he viewed systems of thought as contingent products of many small, unrelated causes. Foucault's genealogical studies also emphasize the essential connection between knowledge and power. Bodies of knowledge are not autonomous intellectual structures that happen to be employed as Baconian instruments of power. Rather, they are essentially tied to systems of social control. Foucault first used his genealogical approach to study the relations between modern prisons and the psychological and sociological knowledge on which they are based. He next proposed a similar analysis of modern practices and 'sciences' of sexuality, but eventually decided that such a study had to begin with an understanding of ancient Greek and Roman conceptions of the ethical self. This study was published in two volumes that appeared just before his death.

1 Biography

Foucault was born on 15 June 1926 in Poitiers, where his father was a prominent physician. In 1946, after preparatory studies with Jean Hyppolite at the Lycée Henri IV, he entered the École Normale Supérieure. He completed advanced degrees in both philosophy and psychology, working with, among others, Maurice Merleau-Ponty (see Merleau-Ponty, M. §2). Dissatisfied with French culture and society, Foucault held various academic posts in Sweden, Poland and Germany from 1955 to 1960, while he also completed his thesis (on madness in the Classical Age) for the *doctorat ès lettres*, which he published in 1961. During the 1960s, Foucault held a series of positions in French universities, culminating in 1969 with his election to the Collège de France, where he was Professor of the History of Systems of Thought until he died. Throughout the 1970s and until his death, Foucault was active politically, helping to found the Groupe d'Information sur les Prisions and supporting protests on behalf of homosexuals and other marginalized groups. He also frequently lectured outside France, particularly in the USA, and in 1983 had agreed to teach annually at the University of California at Berkeley. One of the first victims of AIDS, Foucault died in Paris on 25 June 1984.

Contrary to common views of authorship as self-expression, Foucault said that he wrote to escape from himself, to become other than he was. Correspondingly, there is no methodological or theoretical unity of Foucault's thought that will support any single comprehensive interpretation. His writings instead fall into several main groups, each characterized by a distinctive problematic and method of approach. It is fruitful to follow certain themes through some or all of these groups, but the core of his effort at any point is defined by what is specific to the problems then engaging him.

2 The history of madness

Foucault's earliest publications dealt with psychology and mental illness. His initial approach, developed in a long introduction to the French translation of Ludwig Binswanger's *Traum und Existenz* (1954), was through existential phenomenology, particularly that of the early Heidegger. His 1954 book, *Maladie mentale et personnalité (Mental Illness and Psychology)*, combined this approach with a Marxist analysis (which, however, Foucault soon decisively rejected). His first major work, *Folie et déraison: histoire de la folie à l'âge classique (Madness and Civilization)* (1961), tried to combine the experiential emphasis of his earlier phenomenological discussions with an essentially historical approach.

Folie et déraison is a challenge to the modern use of the terms 'mad' and 'mentally ill' as synonyms. Beginning in

the early nineteenth century, doctors and other therapists rejected such traditional conceptions of madness as divine ecstasy or diabolical possession in favour of the 'enlightened' view that madness is mental illness. Standard histories of psychiatry canonize this view, telling the story of how brave and compassionate men such as Tuke and Pinel replaced superstitious cruelty with scientific treatment of the mad. Foucault's rejection of this view is based on a detailed analysis of the 'experience' of madness that prevailed in the Classical Age (roughly, 1650 to 1800). Then, he maintains, madness was regarded not as mental illness but as a fundamental choice in favour of unreason (*déraison*), where unreason is any basic rejection of the norms of rationality constituting the boundaries of bourgeois social life. Among the various forms of unreason (including sexual promiscuity and deviancy, irreligion and idleness) madness was distinguished by its embracing the animal aspect of human nature at the expense of all higher aspects. The mad were those who had stripped themselves of everything distinctive of their humanity and had chosen to live like beasts. Since, on the classical view, madness was defined by its rejection of reason, the only rational reaction to it was rejection and exclusion. (Foucault regards Descartes' dismissal of his possible madness as grounds for doubt as a paradigm of this point.) Since the Classical Age had no coherent way of giving the mad a place in society, the only alternative was to exclude them from rational society, an exclusion epitomized by the Great Confinement of 1656.

The implication of Foucault's analysis is that there was, even in the relatively recent past of our own culture, a view of madness which was radically different from our own and no less defensible. This alone, he suggests, should begin to undermine our idea that there is something inevitable about our conception of madness. Foucault drives home his point through an analysis of the development of the modern (post-French-Revolution) 'experience' of madness as mental illness. This experience restores a social locus to madness, seeing it as a deviation from norms (an illness), not a rejection of the entire framework of rationality that defines these norms. He takes particular pains to show that, in spite of its veneer of scientific objectivity, the modern view is based more on a moral disapproval of the values implicit in madness than on any objective scientific truth. Similarly, he argues that the modern treatment of the mad (in asylums) was not so much a matter of medical compassion as a concerted effort to bring the mad back beneath the yoke of bourgeois morality. Initiating a theme further developed in various literary essays during the 1960s, *Folie et déraison* continually evokes the lives and works of artists haunted by madness (Van Gogh, Roussel, Artaud, Nietzsche) as precious expressions of a truth suppressed by both classical and modern experiences of madness.

3 The archaeological method

The second major division of Foucault's work begins with his history of the origins of modern medicine, *Naissance de la clinique: une archéologie du régard médical (The Birth of the Clinic)* (1963). Its first pages suggest that it is an extension of Foucault's ethical critique of the concept of mental illness to that of physical illness. But very soon the study becomes an analysis of the linguistic and conceptual structures underlying the modern practice of medicine; or, in the phrase of its subtitle, 'an archaeology of medical perception'.

Foucault's development of his 'archaeological method' synthesized three fundamental lines of influence on his thought: the history and philosophy of science of Gaston Bachelard and Georges Canguilhem (see Bachelard, G. §2; French philosophy of science §§1-6); the modernist literature of (especially) Raymond Roussel, George Bataille and Maurice Blanchot; and the historiography of Fernand Braudel and his *Annales* School. The point of convergence of these influences was the elimination of the subject as the centre of historical and philosophical analysis. Bachelard and Canguilhem challenged what Foucault called the 'transcendental narcissism' of existential phenomenology through a philosophy of objective concepts opposed to the existentialists' philosophy of subjective experience. Bachelard worked primarily on the physical sciences and Canguilhem on the biological and medical sciences, but Foucault extended their viewpoint to the strongholds of the modern conception of subjectivity: the 'human sciences'. Modernist writing excited Foucault by its potential for, in Bataille's terminology, 'transgressing' the limits of standard knowledge and experience. As illustrated in his essay, 'What Is an Author?' (1969a), Foucault was particularly impressed with the modernists' decentring of the author and their constitution of language itself as the essence of literature. Braudel and his school had obtained extremely interesting results by varying the historiographical perspective; that is, by writing history not in terms of individuals' experience but from the broader standpoint of long-term factors such as geography, climate and natural resources. Foucault's archaeology did not take over any of Braudel's specific results or methods but tried to effect a parallel change of perspective in the history of thought: a move away from the individual thinker and

towards more fundamental categories and structures.

Foucault’s fullest deployment of his archaeological method was in *Les mots et les choses: une archéologie des sciences humaines (The Order of Things)* (1966), where he analysed the linguistic systems (‘epistemes’) characteristic of certain periods of thought. In particular, Foucault delineated the linguistic systems underlying the classical disciplines of general grammar, natural history and analysis of wealth, as well as those of the modern disciplines of philosophy, biology and economics that replaced them. He argued that there were strong structural similarities among the three classical disciplines and among the three modern disciplines, but a sharp break between classical and modern modes of thought taken as wholes. On this basis he rejected, for example, the common view that the work of nineteenth-century biologists such as Darwin was a continuous development of the work of eighteenth-century natural historians such as Lamarck. Specifically, he maintained that there is no hint of the Darwinian concept of evolution in Lamarck or any other classical thinker. He took such results as illustrative of the superiority of his archaeological approach to standard history of ideas, which focused on the specific concepts and theories of particular thinkers and not on the linguistic structures underlying them.

L’archéologie du savoir (The Archaeology of Knowledge) (1969b) systematically articulated the methodology Foucault had gradually forged in his preceding historical studies. It did this through an account of discourse, based on his notion of the statement (*l’énoncé*), which described a level of linguistic structure prior to and determining the range of objects, concepts, methodological resources and theoretical formulations available to individuals who speak and write. This account provided a theoretical elucidation of the decentring of the subject effected by Foucault’s histories.

4 Genealogy

Foucault’s writings during the 1970s constitute a third major division of his work. Although archaeological method is not abandoned in this period, it is subordinated to a new style of analysis that Foucault, with a bow to Nietzsche, dubs ‘genealogical’. A genealogical analysis explains changes in systems of discourse by connecting them to changes in the non-discursive practices of social power structures. Foucault recognizes the standard economic, social and political causes of such changes, but he rejects the efforts of many historians to fit these causes into unitary, teleological schemata, such as the rise of the bourgeoisie or Napoleonic ambition. Rather, he sees changes in non-discursive practices as due to a vast number of minute and unconnected facts, the sorts of ‘petty causes’ that Nietzsche evoked in his genealogies (see Genealogy).

Foucault’s genealogical studies emphasize the essential connection between knowledge and power. Bodies of knowledge are not autonomous intellectual structures that happen to be employed as instruments of power. Rather, precisely as bodies of knowledge, they are tied (but not reducible) to systems of social control. This essential connection of power and knowledge reflects Foucault’s view that power is not merely repressive but a creative, if always dangerous, source of positive values. Although systems of knowledge may express objective truth in their own right, they are none the less always tied to current regimes of power. Conversely, regimes of power necessarily give rise to bodies of knowledge about the objects they control; but this knowledge may - in its objectivity - go beyond and even ultimately threaten the project of domination from which it arises.

Surveiller et punir: naissance de la prison (Discipline and Punish) (1975) is the best example of Foucault’s genealogical approach. Here Foucault applies his conception of knowledge/power to the connection between modern disciplinary practices and modern social scientific disciplines. His primary example is the relation of the practice of imprisonment to such disciplines as criminology and social psychology. But imprisonment quickly becomes a model for the entire range of modern disciplinary practices, as employed in schools, factories, the military and so on. *Discipline and Punish* is a genealogical study in the precise sense that it shows how fundamental changes in thought (the emergence of new social scientific disciplines) were causally connected with changes in non-discursive practices (characteristically modern means of controlling the body).

Foucault’s *Histoire de la sexualité (History of Sexuality)* (1976-84) was initially conceived as a straightforward extension of the genealogical approach to sexuality. His idea was that modern bodies of knowledge about sexuality (the ‘sciences of sexuality’, including psychoanalysis) have an intimate association with the power structures of modern society. Volume 1, published in 1976, was intended as the introduction to a series of studies on particular aspects of modern sexuality (children, women, perverts and population), outlining the basic viewpoint and

methods of the project. A central contention was that the history of sexuality is distorted by our acceptance of the ‘repressive hypothesis’: the proposition that the primary attitude towards sex during the last three centuries was one of opposition, silencing and, as far as possible, elimination. Foucault argues that in fact this period produced a ‘discursive explosion’ regarding sex, beginning with the rules of the Counter-Reformation governing sacramental confession. These rules emphasized the need for penitents to examine themselves and articulate not just all their sinful sexual actions, but all the thoughts, desires and inclinations behind these actions. The distinctive modern turn is the secularization (in, for example, psychoanalysis) of this concern for knowing and expressing the truth about sex.

Foucault emphasizes the similarities in our views of sex and crime. Both are objects of allegedly scientific disciplines, which simultaneously offer knowledge and domination of their objects. In the case of sexuality, however, control is exercised not only through others’ knowledge of individuals but also through individuals’ knowledge of themselves. We internalize the norms laid down by the sciences of sexuality and monitor our own conformity to these norms. We are controlled not only as *objects* of disciplines but also as self-scrutinizing and self-forming *subjects*. Foucault thus sees our apparently liberating focus on our sexuality as just a reinforcement of the mechanisms of social control. The self-scrutiny that overcomes psychic repression to reveal our deep sexual nature is merely a subtle means of shaping us to the norms of modern society.

5 Sexuality and ethics

Foucault planned the second volume of his history of sexuality as a study of the origins of the modern notion of the subject in the practices of Christian confession. He wrote such a study, ‘Les aveux de la chair’ (The Confessions of the Flesh), but did not publish it because he decided that a proper understanding of the Christian development required a comparison with ancient conceptions of the ethical self. This led to two volumes on Greek and Roman sexuality: *L’usage des plaisirs (The Use of Pleasure)* (1984a) and *Le souci de soi (The Care of the Self)* (1984b). These two volumes mark the fourth and final period of Foucault’s work, a period most striking for its emphasis on the individual self: the ‘problematization’ of its world and actions and the ‘aesthetics of existence’ whereby it makes its life a work of art. It might seem that Foucault has finally rejected the derivative and ephemeral status of the individual. But this would be doubly mistaken. On the one hand, he still sees our history as strongly structured by discursive and non-discursive practices operating at much deeper levels than that of human consciousness. On the other hand, every stage of Foucault’s work was directed towards overcoming the limitations of individuals (himself and others). Previously, his effort was the negative one of dissolving the apparently necessary constraints of society and its discourses. In this final turn to what he calls ‘ethics’, he began to explore the positive possibilities of self-creation.

In *The Use of Pleasure* and *The Care of the Self*, Foucault compares ancient pagan and Christian ethics through studies of the test case of sexuality. He notes that the moral codes of pagans and Christians were similar, but maintains that there were fundamental differences in the ways in which individuals were subordinated to the codes (in the ‘forms of subjectification’). The Greeks of the fourth and fifth centuries BC, unlike the early Christians, did not regard the domain of sexual acts (*ta aphrodisia*) as evil in its own right, but as natural and necessary. The Greeks did see sexual acts as objects of moral concern because of their animality and their great intensity. What was dangerous, however, was not sex in itself but its excesses. Therefore, the Greek mode of subjection to the code of sexual ethics was a matter of the proper use (*chresis*) of pleasures. Unlike the Christians, the Greeks allowed the full range of sexual activities (heterosexual, homosexual, in marriage, out of marriage) with proper moderation. Properly used, sex was a major part of an aesthetics of the self: the self’s creation of a beautiful and enjoyable existence.

The Use of Pleasure analyses a variety of primary texts (for instance, those of Plato and Xenophon) in order to understand the classical Greek conception of an aesthetics of existence. *The Care of the Self* continues with studies (of Galen, Artemidorus and Plutarch, for example) showing how later antiquity gradually moved away from this aesthetics towards a hermeneutics of the self. The latter, fully developed only by Christianity, replaced the ideal of aesthetic self-creation with that of a deep understanding of a hidden ‘real self’. Foucault regards this Christian conception as the root of our domination by the sciences of sexuality he discussed in *The History of Sexuality I*. Although insisting that there can be no question of ‘going back to the Greeks’, he suggests that reflection on the aesthetics of existence may help us devise liberating alternatives to the traps of modern sexuality (see Sexuality,

philosophy of §3).

6 Conclusion

It is impossible to understand Foucault's work in the typical manner of histories of philosophy. There is not only no system, but no sustained vision, message or project (which we find even in such mavericks as Nietzsche, Kierkegaard and Wittgenstein). For Foucault, philosophy is always just a means of overcoming some specific set of historical limits. It has no final goal, no specific truth or effect, of its own. It is merely a set of intellectual techniques, tied to a consciousness of the historical enterprise that has been known as philosophy. If philosophy ever transforms its self-conception along the lines of Foucault's practice, then he will be recognized as a great philosopher (or, more likely, as someone who played a major role in eliminating philosophy as it had been understood since Plato). Otherwise, he will in all likelihood remain a minor figure, interesting for his odd historical perspectives and his quirky social criticism.

GARY GUTTING

List of works

Foucault, M. (1954) *Maladie mentale et personnalité*, Paris: Presses Universitaires de France; revised as *Maladie mentale et psychologie*, Paris: Presses Universitaires de France, 1962; trans. A. Sheridan, *Mental Illness and Psychology*, Berkeley, CA: University of California Press, 1987.(Foucault's first book, showing the strong influence of existentialist phenomenology and of Marxism. The 1962 revision (the version used in the translation) eliminates most of the Marxist discussions and develops a view much closer to that of *Folie et déraison*.)

Foucault, M. (1961) *Folie et déraison: Histoire de la folie à l'âge classique*, Paris: Plon; 2nd edn, with new preface and appendices, *Histoire de la folie à l'âge classique*, Paris: Gallimard, 1972; trans. R. Howard, *Madness and Civilization*, New York: Pantheon, 1965.(A critical history of the origins of modern psychiatry. The English translation is greatly abridged, amounting to less than half of the original.)

Foucault, M. (1963) *Naissance de la clinique: une archéologie du régard médical*, Paris: Presses Universitaires de France; trans. A. Sheridan, *The Birth of the Clinic*, New York: Vintage, 1973.(An analysis of the linguistic and conceptual structures underlying the modern practice of medicine.)

Foucault, M. (1966) *Les mots et les choses: une archéologie des sciences humaines*, Paris: Gallimard; trans. A. Sheridan, *The Order of Things*, New York: Random House, 1970. (A critical history of the origins of the modern social sciences.)

Foucault, M. (1969a) 'Qu'est-ce qu'un auteur?', *Bulletin de la Societé française de Philosophie* 63: 73-104; trans. 'What is an Author?', in D. Bouchard and S. Simon (eds) *Language, Countermemory, and Practice: Selected Essays and Interviews [of M. Foucault]*, Ithaca, NY: Cornell University Press, 1977.(Foucault's influential essay on the 'death of the author'.)

Foucault, M. (1969b) *L'archéologie du savoir*, Paris: Gallimard; trans. A. Sheridan, *The Archaeology of Knowledge*, New York: Pantheon, 1972.(Reflections on Foucault's method of archaeological analysis.)

Foucault, M. (1975) *Surveiller et punir: naissance de la prison*, Paris: Gallimard; trans. A. Sheridan, *Discipline and Punish*, New York: Pantheon, 1977.(A genealogical study of the modern prison and, more generally, of the disciplinary practices of modern society. The work introduces Foucault's influential views on knowledge and power.)

Foucault, M. (1976) *Histoire de la sexualité I: la volonté de savoir*, Paris: Gallimard; trans. R. Hurley, *The History of Sexuality, Volume I: An Introduction*, New York: Pantheon, 1978.(An introduction to a projected, but never completed, series of studies on aspects of modern sexuality.)

Foucault, M. (1984a) *L'usage des plaisirs: histoire de la sexualité, tome 2*, Paris: Gallimard; trans. R. Hurley, *The Use of Pleasure*, New York: Pantheon, 1985.(A study of texts by Plato, Xenophon and others, with a view to understanding the Classical Greek conception of an aesthetics of existence.)

Foucault, M. (1984b) *Le souci de soi: histoire de la sexualité, tome 3*, Paris: Gallimard; trans. R. Hurley, *The Care of the Self*, New York: Pantheon, 1985.(Studies of texts from late antiquity, tracing the move from an aesthetics of existence to a hermeneutics of the self.)

Foucault, M. (1994) *Dits et écrits, 1954-1988, tomes 1-4*, ed. D. Defert and F. Ewald with J. Lagrange, Paris: Gallimard.(A comprehensive collection of Foucault's essays, lectures and interviews.)

References and further reading

Armstrong, T.J. (ed. and trans.) (1991) *Michel Foucault, Philosopher*, New York: Routledge.(A collection of papers presented at a major international conference on Foucault.)

Binswanger, L. (1954) *Le rêve et l'existence*, trans. from the German by J. Verdeaux, Bruges: Desclée de Brouwer.(Foucault's long introductory essay shows the early influence of Sartre and Heidegger.)

Deleuze, G. (1988) *Michel Foucault*, trans. S. Hand, Minneapolis, MN: University of Minnesota Press.(An analysis of Foucault's work by a friend and major French philosopher.)

Dreyfus, H. and Rabinow, P. (1983) *Michel Foucault: Beyond Structuralism and Hermeneutics*, Chicago, IL: University of Chicago Press, 2nd edn.(A very influential early interpretation of Foucault's work.)

Gutting, G. (1989) *Michel Foucault's Archaeology of Scientific Reason*, Cambridge: Cambridge University Press. (A detailed study of Foucault's writings from the early works to the *Archaeology of Knowledge*.)

Gutting, G. (ed.) (1994) *The Companion to Foucault*, Cambridge: Cambridge University Press.(Essays covering most of the main areas of Foucault's thought.)

Hoy, D. (ed.) (1986) *Foucault: A Critical Reader*, Oxford: Blackwell.(A collection of some of the best essays on Foucault.)

McNay, L. (1994) *Foucault: A Critical Introduction*, New York: Continuum.(An excellent introductory survey of Foucault's work.)

Macey, D. (1993) *The Lives of Michel Foucault: A Biography*, New York: Pantheon. (The best of the several biographies of Foucault.)

Rajchman, J. (1985) *Michel Foucault: The Freedom of Philosophy*, New York: Columbia University Press.(A perceptive discussion of Foucault's 'scepticism of our modernity'.)

Foucher, Simon (1644-96)

Simon Foucher, Canon of Dijon, was a sceptical thinker, active in intellectual circles in Paris. His main philosophical project was the revival of Academic scepticism, and he emerged as an important and influential critic of Cartesian philosophy, questioning the consistency of the Cartesians' commitment both to mind-body dualism and to the claims that ideas in the mind represent and make known external bodies, and that mind and body interact. He was generally concerned to undermine the Cartesian pretension to know the real essences of mind and body. Foucher was also an early constructive critic of Leibniz's system of pre-established harmony.

1 Life and works

Foucher was born on 1 March 1644, in Dijon, France. He was ordained to the Roman Catholic priesthood while still quite young, and made an honorary canon of the Sainte Chapel of Dijon. Although he received a bachelor's degree in theology from the Sorbonne, and earned a small salary as the chaplain of a religious house in Paris, his real interests were philosophical. He spent the rest of his life in Paris, and quickly became known as a formidable critic of Cartesian philosophy. His first serious contacts with Cartesianism probably came in 1667-8 through the weekly lectures on physics given by Jacques Rohault. (There is an apocryphal story that Rohault asked Foucher to deliver a funeral oration when Descartes' remains were returned to Paris for burial in 1667). It was probably Rohault's lectures (and subsequent arguments between the two) that stimulated Foucher's interest in experimental physics, and in 1672 he published a letter describing various instruments for measuring air humidity. By that time, he had also developed a passion for Academic scepticism, not just as a means for criticizing what he perceived as Cartesian dogmatism, but also as a useful and important propaedeutic for Christian belief. In 1673, he published the first of three dissertations on the Academic philosophy, *Dissertation sur la recherche de la vérité, ou sur la logique des academiciens*. It was Foucher's belief that Nicolas Malebranche's *De la recherche de la vérité*, was a response to this work that led him to compose his *Critique de La recherche de la vérité* (1675), in which he criticizes Malebranche in particular and Cartesian metaphysics and epistemology generally. This was only the opening salvo in what would be an extended (and acrimonious) public polemic between Foucher and the Cartesians, especially Malebranche and Dom Robert Desgabets (who rallied to Malebranche's defence). Foucher also found himself engaged in a fruitful, twenty-year philosophical correspondence with Leibniz (whom he had met when Leibniz resided in Paris from 1672 to 1676). Foucher's early critical remarks on Leibniz's metaphysics forced Leibniz to clarify and develop his views, which Foucher helped get published. Foucher died (some said of overwork) on 27 April 1696.

2 Academic scepticism

In his assault on the Cartesian system, Foucher employed the weapons of classical scepticism. He looked, in particular, to the thinkers of the New Academy, such as Antilochus and Philo, although he also saw Academic scepticism as a perfectly Augustinian tool that promotes Christian faith (see Academy; Platonism, Early and Middle). Foucher's scepticism was of the moderate variety. He sought to steer a middle course between dogmatism and extreme Pyrrhonism (see Pyrrhonism). While there are, to be sure, many things that cannot be known, there are also some that can be known with certainty. The sincere sceptic does not doubt all things. Rather, the goal of the Academic sceptic is precisely to distinguish that which can be known from that which cannot. Philosophy just is the search for those *veritez evidentes* and *connoissances certaines* that lead to new knowledge.

According to Foucher, the criterion of truth is absolutely certain evidence, complete indubitability, particularly as detected by application of the law of non-contradiction. One should thus admit as true and as known only that which is evidently true, or that which cannot be denied without contradiction. On this criterion, it turns out that the most evident truths are mathematical and conceptual, along with introspective claims concerning states of mind. This class of self-evident truths gives us our first principles, from which we can then seek out further knowledge, particularly concerning the soul, God and the existence of the world. As long as we start from incontestable first principles and draw from them only necessary consequences, we will not be led into error and falsehood. Foucher thus sums up the Academic method in five simple principles: (1) in the matter of science, proceed only by demonstration; (2) do not concern oneself with problems that clearly cannot be resolved; (3) admit that one does not know those things that one effectively does not know; (4) distinguish those things that can be known from

those that cannot be known; and (5) always seek new knowledge.

While the evidence of introspection does provide knowledge of our internal states - states of the soul, such as pleasure, pain, heat and cold sensations, and other sensible qualities as sensed - and we can infer the *existence* of external bodies in the world from the testimony of sensation, we cannot have real knowledge of the natures or essences of external things, how they are 'in themselves'. The senses, he insists, can tell us nothing about the truth of things; they provide only phenomenal appearances, and we can never know whether those appearances represent to us things are they really are. Immediately-known appearances are only 'ways of being' (*façons d'être*) or 'modifications' of the soul, which we can never compare with the things themselves, to which we have no direct access. Similar considerations apply to our knowledge of the soul. We can be certain that the soul exists, and we have immediate knowledge of the modifications or properties of the soul - that it is a thinking thing - but we cannot have certain knowledge of what the soul is in its being or nature.

In the case of the particular sciences, which are based on sense experience, Foucher insists that one can attain many probabilities, as long as one follows proper (that is, academic) method. Speculation about the causes of our sensations can lead to a kind of knowledge about the world, and will even produce many useful and practical results. But one should be properly modest (and not dogmatic) about the status of the results thereby obtained.

3 Critique of Cartesianism

Academic scepticism gave Foucher the means to criticize the Cartesians. When Descartes and his followers claim to know the essence of matter and the essence of mind, and to know therefore that these are two substances that are radically different in nature, they go beyond what is evident and certain. While Foucher, in his dissertations on *les Academiciens*, contrasts in a general way the Academic philosophy with the Cartesian philosophy, the particular occasion for him to compose a detailed critical investigation into the principles of Cartesianism was the publication in 1674 of the first volume (the first three books) of Malebranche's *Recherche*. Foucher took this to be a complete work in its own right, and this led him to misunderstand some important features of Malebranche's system. (Malebranche himself testily replied that 'when one criticizes a book, it seems one ought at least to have read it'). But although Foucher may occasionally miss the mark with respect to Malebranche, he none the less scores some important points, later to prove influential, against a more orthodox version of the Cartesian system. Foucher, in fact, takes himself to be examining not just Malebranche's work, but at the same time *une partie des principes de M. Descartes*.

Foucher composed his *Critique* in 1674, and it was published in early 1675. His real object of attack, in this and subsequent critical works, is the confidence of Cartesians that they know the essences of mind (thought) and matter (extension). Two themes running throughout the *Critique* are, first, that these are matters for investigation, and are not to be assumed at the outset of one's inquiries; and, second, for all we know there may be something material about the mind. Perhaps, for example, our mental operations consist in certain brain activities. Moreover, if mind-body dualism is true, Foucher argues, then there are some serious internal and sceptical problems facing Cartesian epistemology and metaphysics.

Cartesians say (and Foucher agrees) that what the mind always immediately perceives is a representative idea: external bodies are perceived only by means of ideas. But then Cartesians also want to draw a distinction between quantitative ideas (of extension and motion) that truly represent features of external bodies and qualitative sensory ideas (colours, light, heat and so on) which do not. Foucher claims that they are not entitled to this distinction. For the Cartesian, he insists, *all* ideas are only modes of being of the soul and all are caused in the mind by bodies in the same way (here he misreads Malebranche, for whom ideas properly speaking are not modes of the soul but are in God (see Malebranche, N. §§2-3). On what grounds, therefore, can one say that some mental modes represent features of external bodies and others do not? Either both kinds of ideas represent, or neither does.

In the course of his arguments in the *Critique* and other works (in effect, arguments against the difference in epistemological value granted by philosophers to 'primary quality ideas' and 'secondary quality ideas'), Foucher introduced considerations that would prove to be influential in the history of sceptical objections to the so-called 'way of ideas'. Cartesians say that we know colours, light and other such qualities through the senses; that is, that they just are sensations and properties of the mind. But if this is an argument that there is nothing like colour in external bodies, it follows that it is equally an argument that there is nothing like extension in bodies, since we also

know extension through the senses. Thus, extension and figure would have to be no less 'in us' than are light and colours and heat. Moreover, if one grants that sensory qualities such as colour are in us, then it is difficult not to admit that the mathematical qualities are in us as well. First, the sceptical arguments that appeal to variations in sensory appearances to show that such qualities as heat and cold are in us and not in bodies can be used to demonstrate the same thing about all the other qualities, such as shape and size. Second, how could one claim that colours are in us without also recognizing that extension must be in us as well, since the extension of a colour, its shape, is obviously and undeniably where the colour is? Thus, once again, either both kinds of ideas represent, or neither does.

Foucher, however, raises an even deeper question for the Cartesian theory of ideas. How can *any* mental being such as an idea represent and thus make known something material? Since, as the Cartesians insist, 'the soul has nothing at all in it that is like matter and extended beings', a mode of the soul cannot be like a material body. And if there is no likeness between an idea and a body, then an idea cannot represent a body. *None* of our ideas, for the Cartesian, should be capable of giving us knowledge of objects outside us. To save their system, Foucher believes, Cartesians deny that representation requires resemblance; words, for example, represent objects without resembling them. Foucher replies that the notion of non-resembling representative ideas is incomprehensible. On that assumption, any idea could arbitrarily represent any object and all of our ideas could represent one and the same object. Furthermore, representation (at least in its primary signification) essentially involves resemblance. Words can represent objects they do not resemble only because they excite in the mind ideas that do represent things in a primary (that is, a resembling) manner. Foucher's conclusion is that since representation requires resemblance, to posit representational ideas in the mind is to admit that mind and matter cannot be as different as the Cartesians say they are. Either our ideas do not represent material things at all, or they do so because they are like those material things - in which case the Cartesians do not really know the essences of mind and matter.

The same general sceptical conclusion is drawn from a consideration of the question of mind-body interaction. Foucher claims that Cartesians are committed to a causal principle according to which there must be an essential or substantial likeness between a cause and its effect. But for Cartesians there is no respect whatsoever in which mind and body are alike. So they must, by their own principles, deny that mind and body causally interact. But, Foucher continues, mind and body obviously *do* interact; and the causal principle is self-evident. It follows that Cartesians cannot really know the essences of mind and matter.

4 Correspondence with Leibniz

In 1676, Foucher began a correspondence with Leibniz that lasted until the end of his life. Its interest lies both in Leibniz's own comments on his constantly developing metaphysics and in Foucher's insightful constructive criticisms. The extracts published in the *Journal des sçavans* between 1692 and 1696 constitute Leibniz's first public presentation of the principles of his *nouveau système*. There is much common ground between the correspondents. They agree, for example, on the value of Academic scepticism for philosophical inquiry, although Leibniz is the more insistent that the philosopher should seek to establish metaphysical and physical truths by building up a system from first principles. What Foucher seems to find most attractive about Leibniz's system is the critique of Cartesianism and the alternative to a strict dualism which it offers. He agrees with Leibniz, for example, that the essence of matter does not consist in extension, as the Cartesians had asserted.

Part of Leibniz's project in the 1680s and 1690s was to explain 'the communication of substances and the union of mind and body' without postulating either real causal interaction between them, or the continuous, quasi-miraculous agency of God (as Malebranche and, perhaps, if more equivocally, Descartes had done). Every individual substance, Leibniz claimed, spontaneously generates (from an inner principle of action) its own sequence of states, and God has so set up things from the start that there is, in the unfolding of the states of things over time, a 'grand concommitance' (what Leibniz later called the 'pre-established harmony') (see Leibniz, G.W. §6).

Foucher grants that God could adjust things in such a way that the body could produce by itself all the movements that the soul joined to the body would produce, without giving to the soul any power to effect motion in the body; and, conversely, that the thoughts and modifications of the soul that correspond to these bodily motions could arise successively at the precise moments when the body undergoes its movements. But what, he asks, is gained by such a 'grand artifice'? As much as Malebranche's occasionalism, it has God go to a great deal of trouble to make it

seem as though substances causally interact when in truth they do not (see Occasionalism). God might as well dispense with material bodies altogether and simply produce thoughts and modifications in the mind directly. Bodies, on Leibniz's (and the occasionalist) view, are useless anyway, since the mind can neither move nor know them. What Foucher really wants, however, is an explanation of *how* mind and body do interact, and not the implausible claim that the appearance of interaction is misleading.

Foucher is surprised by Leibniz's account. It is clear to him that Cartesians are forced into their metaphysical contortions by their commitment to dualism: since mind and body are so unlike each other, real interaction is not possible; thus, the recourse to a *deus ex machina*. But Leibniz rejects the conception of two utterly different kinds of substance, and he puts active force in things. So what, Foucher wonders, prevents mind and body from really interacting? But perhaps, he suggests in fine sceptical manner, these are questions that we simply cannot yet resolve, at least not until all philosophers agree on the infallible criterion of truth.

5 Influence

Unlike Descartes, Malebranche and Leibniz, Foucher was not a system-builder. His philosophical talents were mainly critical and analytical in nature. Yet he is, for that reason, important and influential in the history of seventeenth and early eighteenth century philosophy, particularly through the epistemological criticisms he directed at the Cartesian theory of ideas. His sceptical arguments against the distinction between ideas which represent features of external bodies and sensations which do not, were picked up by such contemporary and later thinkers as Bayle (*Dictionnaire historique et critique*, article on Pyrrho, remark B; article on Zeno, remarks G and H), Berkeley (*Principles of Human Knowledge*, 8-15, and *Three Dialogues* I) and Hume (*A Treatise on Human Nature* I.4.ii).

See also: Descartes, R.; Dualism; Pyrrho

STEVEN NADLER

List of works

Foucher, S. (1673) *Dissertation sur la recherche de la vérité, ou sur la logique des academiciens (Dissertation on the search after truth or on the logic of the academics)*, Dijon.(A presentation of the principles of Academic scepticism, including a comparison with Cartesian philosophy.)

Foucher, S. (1675) *Critique de La recherche de la vérité*, Paris; trans. R. A. Watson and M. Greene in *Malebranche's First and Last Critics*, Carbondale, IL: Southern Illinois University Press, 1995.(A critique of Malebranche's *Recherche de la vérité* in particular, and of Cartesian philosophy in general.)

Foucher, S. (1676a) *Réponse à la Critique de la Critique de La Recherche de la vérité (Response to the critique of the critique of the search after truth)*, Paris.(Foucher's response to Desgabets, who had rallied to Malebranche's defence.)

Foucher, S. (1676b) *Réponse pour la Critique à la Preface du second volume de La recherche de la vérité (Response to critique in the preface of the second volume of the search after truth)*, Paris.(Foucher's direct response to Malebranche's response to Foucher's original *Critique*.)

Foucher, S. (1693) *Dissertations sur la recherche de la vérité, contenant l'histoire et les principes de la philosophie des academiciens. Avec plusiers réflexions sur les sentimens de M. Descartes (Dissertation on the search after truth containing the history and the principles of the philosophy of the academics with several reflections on M. Descartes' sentiments)*, Paris.(A compilation of the material presented in Foucher's three treatises on the Academic philosophy.)

Foucher, S. (1693) 'Extrait d'une lettre de M. Foucher, chanoine de Dijon, pour repondre à M. de Leibniz sur quelques axiomes de philosophie' (Extract from a letter of M. Foucher Canon of Dijon to respond to M. Leibniz on some axioms of philosophy), *Journal des sçavans* 21 (16 March): 365-9.(Some remarks on philosophical method.)

Foucher, S. (1695) 'Reponse de M. S. F. à L. B. Z. sur son nouveau sisteme de la communication des substances, proposé dans les journaux du 27 juin & 4 juillet 1695' (Response of M.S.F. à L.B.Z. on his new system of the communication of substances proposed in the journals of 27 June and 4 July 1695), *Journal des sçavans* 23 (12 September): 639-45.(Further remarks on Leibniz's metaphysics.)

References and further reading

Bayle, P. (1696) *Dictionnaire historique et critique*, ed. and trans. R. Popkin, *Pierre Bayle, Historical and Critical*

Dictionary, Indianapolis IN: Hackett, 1995.(Bayle's most important philosophical work, influenced by Foucher.)

Berkeley, G. (1710/1713) *A Treatise concerning the Principles of Human Knowledge* and *Three Dialogues between Hylas and Philonous*, in *The Works of George Berkeley, Bishop of Cloyne*, ed. A.A. Luce and T.E. Jessop, vol. 2, Edinburgh: Thomas Nelson, 1948-57.(Both works bear Foucher's influence.)

Gouhier, H. (1927) 'La Première Polémique de Malebranche' (Malebranche's first polemic), *Revue d'histoire de la philosophie* 1: 23-48, 168-91.(Examines Foucher's criticisms of Malebranche and Malebranche's responses.)

Hume, D. (1740) *A Treatise on Human Nature*, ed. P.H. Nidditch, Oxford: Clarendon, 1978.(Foucher's influence is evident here at I.4.ii.)

Malebranche, N. (1674-5) *De la recherche de la vérité*, trans. T. Lennon and P.J. Olscamp as *The Search After Truth/Elucidations of the Search AfterTruth*, Columbus, OH: Ohio State University Press, 1980. (Malebranche's major philosophical work.)

Popkin, R. (1957) 'L'Abbé Foucher et le problème des qualités premières' (The Abbé Foucher and the problem of primary qualities), *XVIIe Siècle* 33: 633-47.(Discusses Foucher's arguments against distinguishing between ideas that represent the quantitative features of things, and ideas of sensible qualities that do not.)

Popkin, R. (1980) *The High Road to Skepticism*, San Diego, CA: Austin Hill Press; repr. as *The High Road to Pyrrhonism*, Indianapolis IN: Hackett Publishing, 1993.(A collection of essays, the first of which places Foucher's scepticism in its historical and philosophical context.)

Rabbe, F. (1867) *Étude philosophique. L'Abbé Simon Foucher chanoine de la Sainte Chapelle de Dijon (Philosophical study: Abbé Simon Foucher Canon of the holy church of Dijon)*, Paris: Didier.(Foucher's life and work.)

Watson, R.A. (1987) *The Breakdown of Cartesian Metaphysics*, Atlantic Highlands, NJ: Humanities Press. (Contains a reissue of *The Downfall of Cartesianism*, which analyses Foucher's critique of Cartesian philosophy and Cartesian responses to that critique.)

Watson, R.A. (1991) 'Foucher's Mistake and Malebranche's Break: Ideas, Intelligible Extension and the End of Ontology', in S. Brown (ed.) *Nicolas Malebranche: His Philosophical Critics and Successors*, Van Assen: Gorcum.(Examines, in the context of the Foucher-Malebranche debate, the issue of whether epistemological questions about ideas could or should have ontological answers.)

Foundationalism

Some foundationalists are rationalists who rely on intuition and deduction. Others are empiricists, in a broad sense, and accept observation and induction or abduction or yet other ways to support beliefs by means of other beliefs. What they have in common is that they are all willing to hazard a positive view about what in general makes a belief epistemically justified in the way required for it to be a case of knowledge; and they all propose something of the following general form: belief b is justified if and only if either b is foundationally justified through a psychological process of direct apprehension p (such as rational intuition, observation, introspection, and so on) or else b is inferentially justified through a psychological process of reasoning r (such as deduction, induction, abduction, and so on) ultimately from beliefs all of which are acquired or sustained through p. If one rejects all forms of such foundationalism, then a question remains as to what distinguishes in general the cases where a belief is epistemically justified from the cases in which it is not. Can anything general and illuminating be said about what confers epistemic justification on a belief, and what gives a belief the epistemic status required for it to constitute knowledge (provided it is true)?

1 Formal foundationalism versus substantive foundationalism

The foundationalism controversy in the late twentieth century can be clarified by a distinction between ‘formal foundationalism’ and ‘substantive foundationalism’. Substantive foundationalism is opposed to coherentism; formal foundationalism is opposed not to coherentism but to what is here called (epistemic) ‘pessimism’.

Formal foundationalism may be held with respect to the study of normative or evaluative principles of any sort. For instance, it may be held in ethics as well as in epistemology. Formal foundationalism in ethics tries to fix the goodness of events or states, or the rightness of actions, perhaps recursively. Thus a simple utilitarian theory might say that:

(1) every event of someone undergoing pleasure is good,
(2) every event that causes a good event is good, and
(3) every event that is good is so in virtue of (1) or (2).

(This is of course absurdly simple but it will serve as an example of a foundationalist ethical theory.)

Analogously, formal foundationalism in epistemology would say that:

(1) every belief with a certain non-epistemic property *g* is justified,
(2) if a belief bears relation *r* to a set of justified beliefs then it is itself justified, and
(3) every belief that is justified is so in virtue of (1) or (2)

There are various familiar candidates for the role of property *g* or of relation *r*. Thus property *g* may be indubitability or infallibility of belief, and relation *r* may hold between a belief *b* and a set α of further beliefs when *b* is deductively based on α. But these are only some examples.

Is there an alternative to foundationalism thus understood? If there is a class of justified beliefs at all, must it be specifiable in some illuminating way, at least recursively, or can it just be the class of justified beliefs and that is that? There is no obvious reason to think that the class of blue things must be recursively specifiable in terms of a basis and a generator that are colour-neutral. Why should we think that the class of justified beliefs must be recursively specifiable in terms of a basis and a generator that are epistemically neutral?

The main reason in favour of formal foundationalism - antecedent to an actual, compelling foundationalist theory - is the apparently supervenient or consequential character of the evaluative generally and of epistemic justification in particular. For example, an apple may be a good apple in virtue of certain non-evaluative properties: in virtue, say, of being sweet, juicy, large and so on. If so, then its evaluative property of being a good apple ‘supervenes’ upon its complex of non-evaluative properties that include being sweet, juicy, large and so on. And this means that any sweet, juicy, large, apple would also be a good apple. This example introduces the concept of the ‘supervenient’ to be defined in what follows. According to the doctrine of supervenience, evaluative and normative properties always supervene upon or derive from non-evaluative, non-normative properties, in the way the

goodness of an apple might supervene upon or derive from its juiciness, sweetness and so on (see Supervenience).

An acceptable formal foundationalist theory would specify a particular non-epistemic basis and generator, which would give more precise content to the doctrine of supervenience or consequentialism, and would fulfil its promise. Such a formal foundationalism would assure us that for every case of a justified belief, its being justified is supervenient on a set of non-epistemic facts involving only the basis property of the recursion and its generating relation. (The very same defence can be given for foundationalism in ethics, where, of course, utilitarianism is only one example of such foundationalism.)

It should be noted that formal foundationalism does not entail a doctrine of objectively unique foundations for empirical knowledge. For there might be several alternative recursive specifications of the class of justified beliefs, making use of different bases and generators, without any evident criterion for selecting one as objectively correct. If so, there might be different bases determining different foundations, none objectively prior or superior to the others.

That being so, if (for the sake of example) we presuppose a definition of the justified using indubitability as the basis and deduction as the generator, we may then think of the set of indubitable beliefs as the foundation of empirical knowledge. If so, we will be right only relative to our definition. For if the foundation is what is picked out by the basis property, then the indubitable is the foundation only relative to the definition that uses indubitability as the basis property. And so long as other definitions, using other bases, are equally possible and on a par with ours, the relativity is ineliminable.

2 Epistemic supervenience and formal foundationalism

Pessimism is to be distinguished from the more radical position that simply rejects supervenience. The ‘doctrine of supervenience’ for an evaluative property *f* is simply that, for every *x*, if *x* has ϕ then there is a non-evaluative property (perhaps a relational property) ψ such that:

(1) *x* has ψ, and
(2) necessarily, whatever has ψ has ϕ.

The denial of this for an evaluative property ϕ is a doctrine of the *autonomy* of ϕ. It holds that ϕ can be exemplified even though it does not supervene on any non-evaluative properties of what exemplifies it.

Formal foundationalism entails the doctrine of supervenience for epistemic justification, but is also considerably stronger. It requires in addition a certain faith in our intellectual powers, or a certain confidence in the manageable simplicity of the sphere of the relevant values (that is, epistemic if the foundationalism is epistemic, ethical if it is ethical, aesthetic if it is aesthetic, and so on). After all, formal foundationalism requires not just supervenience but *explicable, comprehensible* supervenience: supervenience at least surveyable by the human mind.

Accordingly, since pessimism is the denial of formal foundationalism and *not* of the weaker doctrine of supervenience, pessimism is a weaker claim than the doctrine of autonomy. Pessimism is compatible with supervenience and requires only a certain scepticism about our ability to comprehend the principles that underlie such supervenience, perhaps because they are infinite in number or degree of complexity.

Far from being pessimist, the coherence theory in epistemology is a kind of formal foundationalism, for it does try to provide principles that specify the conditions within which beliefs are justified. Thus a coherentist might choose coherence within a set of a certain sort as basis and deduction as generator. For example, a coherentist may hold that a belief is justified if and only if *either* it coheres within a large and diverse set of beliefs held by the subject *or* it is deduced by the subject from a set of such beliefs. In fact the coherentist usually has an all-encompassing basis that absorbs all generators, but this is quite compatible with formal foundationalism, though it is a limiting case (see Knowledge and justification, coherence theory of).

Coherentism is opposed not to formal foundationalism but at most to substantive foundationalism. The conflict here is over what basis to choose in the recursive definition of justification. Obviously, there are grades of coherentism and of foundationalism. Radical coherentism holds that *only* coherence can serve as a basis. Radical foundationalism holds that coherence *never* serves as a basis, that the basis property which gives a belief *B* access to the foundation never makes reference to *other* beliefs of the subject, except of course such beliefs as *B* itself

may refer to. And various intermediate positions are clearly possible.

An infinite regress of justification is incompatible neither with formal foundationalism nor even with radical foundationalism. Consider the infinite sequence:

(P1) That there is at least one real number in the interval (0-1)
(P2) That there are at least two real numbers in the interval (0-1)
⋮

I can think of no compelling reason why there could not be a sequence of justified dispositional beliefs in (P1), (P2), … such that each member of the sequence is justified in the following sense (J1) by its successor: (J1) *p justifies (would justify) q if and only if that p is justified is sufficient for q to be justified*. This sense is not ruled out by formal foundationalism, nor does formal foundationalism rule out the possibility of a sequence of dispositional beliefs in (P1), (P2), …. What formal foundationalism would require of such a sequence, however, is that for each of its members there be a possible finite epistemic account or explanation of how its justification supervenes on the non-epistemic (such an explanation to be carried out perhaps by means of a non-epistemic basis *g* and a non-epistemic generator *r*).

It seems, therefore, that foundationalists misplace their objection when they focus on the infinite regress of justification. What they really oppose is pessimism. Their fundamental thesis is formal foundationalism, which we have found to be a form of the doctrine of supervenience.

G.E. Moore was optimistic enough about the possibility of a normative ethics (that is, of an ethical theory or system) to write as follows in *Principia Ethica*:

> When *A* asks *B* what school he ought to send his son to, *B*'s answer will certainly be an ethical judgment. And similarly all distribution of praise or blame to any personage or thing that has existed, now exists, or will exist, does give some answer to the question 'What is good?'…. But this is not the sense in which a scientific Ethics asks the question. Not one, of all the many million answers of this kind, which must be true, can form a part of an ethical system; although *that science must contain reason and principles sufficient for deciding on the truth of all of them*.
>
> (Moore 1903: 3; emphasis added)

For Moore, a scientific ethics must have a kind of completeness which, given that it is comprehensible, it can have only if pessimism is false and formal foundationalism true.

By way of contrast, compare W.D. Ross' opposing view (1930: 41): 'For the estimation of the comparative stringency of… *prima facie* obligations no general rules can, so far as I can see, be laid down.' Ross' view is pessimist, but it may well be true, not only for ethics but also for epistemology. It seems unrealistic to suppose that either subject admits of the kind of completeness required by Moore. Only an evaluative monist, perhaps Mill, could reasonably believe otherwise. Mill himself was well aware of this, and made it the basis of scorn for pluralist 'intuitionism'. The pluralist must abandon all hope of attaining Moorean completeness, for there is no *general* answer accessible to our limited intellects to the question of how to resolve value conflicts or moral dilemmas.

Here we must distinguish between a general *method* for resolving such conflicts or dilemmas and its *application*. The value monist can have a method, even if it may be difficult to apply. Thus it may be difficult to tell whether a certain action would lead to more pleasure than any alternative, but at least the monist radical hedonist can tell us that it is right if and only if it does in fact do so.

Even if it turns out that one must in the end yield to pessimism, moreover, one can still reject autonomism. Suppose one resigns oneself to the fact that given epistemic pluralism (several different *basic* sources of justification) there is no possibility of a complete epistemology, such that every correct attribution of justification to a belief would follow logically from the principles of our complete epistemology and certain non-epistemic facts about the belief. Even so, one might still reasonably reject autonomism and accept supervenience by holding that for every justified belief there must be a property ψ (perhaps a very complex relational property) such that:

(1) that belief has ψ,
(2) ψ is not a normative epistemic property, and

(3) necessarily, whatever belief has ψ is a justified belief.

Here again the foundationalist and the coherentist could turn out to be allies. For each could surely carry on despite pessimism in an attempt to specify *as completely as possible* the conditions within which beliefs are justified. Each could retain formal foundationalism as an ideal which we might approach but probably could never reach.

Formal foundationalism leaves open the viability of coherentism even in its most radical forms, therefore, which means that *substantive* foundationalism requires additional support beyond that provided by *formal* foundationalism. We need not rehearse in detail the familiar arguments against radical coherentism, since we are not assessing the merits of that doctrine. But, these objections to coherentism should be detached from the alleged impossibility of an infinite regress of justification.

3 An argument for classical foundationalism

We turn now to the controversy between classical foundationalism (a sort of substantive foundationalism) and epistemic coherentism. Often when we are adequately justified in believing something, *p*, we are so justified on the basis of some inference or argument or adduced reasons. In such cases our belief is *mediately* justified. The foundationalistic thesis might then be put as follows:

(F) If a subject *S* has (epistemically) justified beliefs, not all of the beliefs held by *S* can be only mediately justified: *S* must (sometime) have beliefs that are immediately justified.

What is it for *S* to hold a belief *b* at least partly on the basis of another belief b'? What is it for the ‘basing relation’ *r* to hold thus from *b* to $brsquo;$? One thing seems clearly involved: *S* must hold *b* at least partly because it holds $brsquo;$: the latter must contribute causally to the former. More is presumably involved, but that much is enough to start a foundationalist argument as follows:

(1) There is at least one justified belief held by *S*.
(2) *r*, the basing relation among beliefs, is irreflexive.
(3) *r* is transitive.
(4) There is no infinite sequence of justified beliefs of *S*’s each member of which has a predecessor that bears *r* to it (on which it is based).
(5) Therefore, at least one justified belief held by *S* is based on no belief held by *S*.

The foundationalist concludes that there must be some way for a belief to be justified that does not involve its being based on some *other* already justified belief. Such ‘immediately justified’ beliefs would hence constitute a foundation on which one could erect one’s system of beliefs. One could base *other* beliefs on those that are foundationally, immediately justified.

That argument for foundationalism depends crucially on the premise that the basing relation, *r*, is irreflexive (premise 2). That premise is questionable, however, as follows. The relatively clear core of the basing relation is constituted by the causal contribution relation *c*, and *c* provides a good approximation to *r*. Consider now the ‘ancestral’ of that relation, c^* (roughly the relation that *x* bears to *y* if and only if either *x* bears *c* to *y* or *x* bears *c* to x_1 which bears *c* to x_2 which… which bears *c* to x_n which bears *c* to *y*). So long as we allow not only immediate or direct but, *also*, mediate and indirect basing, c^* provides an even better approximation to the basing relation than does *c*. Is c^* irreflexive? Consider two rigid cards c_1 and c_2 that support each other by each standing on end on a flat surface and leaning on the other. And suppose God creates these, already arranged thus, at a time *t*. In that case, is it not true that, for every *e*, c_1’s being positioned as it is during the interval [*t* to ($t+\varepsilon$)] *partly results* from c_2’s being positioned as it is during that interval, *and conversely*? Take then the relation of causal contribution *c* (the converse of the relation of partial resultance). From the fact that, as we have seen, *c* is not asymmetrical, it follows that c^* is not asymmetrical. Obviously, moreover, c^* is transitive. And it follows that c^* is *not* irreflexive. In our particular example what we are forced to say is that c_1’s being positioned as it is during the interval [*t* to ($t+\varepsilon$)] *results in part* from its being so positioned during that very interval. For this contributes to supporting c_2 which, being thus supported, can return the favour coincidentally.

The case for the irreflexivity of the belief-basing relation presumably rests at least in part on the supposed

irreflexivity of the causal contribution relation (or its ancestral c^*). In any case, once the latter is abandoned this is bound to weaken our intuitive support, such as it is, for accepting the former. And that intuitive support can also be attacked directly in terms of examples such as one in terms of the following five propositions:

(P1) It is drizzling.
(P2) Expanding circles are constantly forming in the puddles.
(P3) There is a pitter-patter on the window panes.
(P4) There was a forecast of drizzle or rain by now.
(P5) Car wipers are on.

If, at a given time, *t*, I believe all five of these, what is wrong with supposing each to be based in part (perhaps in *small* part) on the others?

4 Foundationalism versus coherentism

The beliefs supporting a belief p_1 held by a subject *S* might be thought to form one of at least three possible structures: (1) first, a ‘tree’ that branches infinitely upwards from the ‘root’ node constituted by *S*’s belief p_1; (2) second, a pyramid with foundational beliefs on which rest level after level of beliefs supported by reasoning and leading ultimately to *S*’s belief p_1 at the tip of the pyramid; (3) third, a raft, one of whose planks is *S*’s belief p_1, where each of the many planks is held in place by its ties of coherence with the other planks.

So far the foundationalist has ruled the tree out of the question for limited human minds, and argued against the raft by supposing that such belief structures must be built by means of some causal belief-basing relation that must be both irreflexive and transitive. Given these assumptions, it follows that any justified set of beliefs held by a limited subject must take the form of a foundational pyramid. As we have seen, however, it is questionable whether the relevant causal-contribution and belief-basing relations must be indeed irreflexive. And doubt on the irreflexivity assumption undermines the foregoing line of reasoning for pyramidal foundations of knowledge. We do well to consider another approach.

A belief may be both true and justified without being knowledge. This is shown by any Gettier example in which someone is justified in believing some falsehood *f* from which in turn they deduce a truth, *t*. Their belief that *t* is then both true and justified without being knowledge (see Gettier problems). A particularly striking Gettieresque case is that of the victim of an evil demon who is allowed a *true* belief that he faces a fire, along with a set of necessarily associated truths - one, however, that is minuscule by comparison with the falsehood that massively surrounds it in our victim’s belief set. Concerning the experiences, memories, introspections and reasonings of the victim, there is not the slightest flaw: the victim is, internally, as cognitively worthy as the best of us (see Internalism and externalism in epistemology). In that sense, therefore, the victim is perfectly well justified in believing that they face a nearby fire. And there is indeed such a fire there. But clearly our victim has no knowledge of that fact, despite the fact that the belief is true and well justified, not if their supporting web of beliefs, and their broader belief system, is almost entirely a tissue of falsehoods; so that, for example, their visual, auditory and other experiences as of a fire before them are entirely unaffected by the fire there, and so on.

What is involved in such internal justification? Classical coherentism and foundationalism are best understood, it seems, as accounts of such justification. According to coherentism, one’s belief *b* is thus justified if and only if *b* coheres well enough with one’s system of beliefs. Let us explore this briefly before turning to foundationalism.

Coherence involves the logical, explanatory, and probabilistic relations among one’s beliefs. However, it would not do to attain a tightly interrelated system by merely lopping off whatever beliefs may refuse to fit. Theoretically, it would be possible (perhaps with the aid of advanced cognitive technology) to perform such drastic surgery on one’s belief system. But that would not necessarily yield epistemic justification. Thus take arbitrary beliefs *p* (that the moon is made of cheese) and *q* (that clouds are cotton wool) and surround those beliefs with the likes of $(p \supset q)$, $(q \supset < eixi = "P021p7.22" >< bixi = "P021p7.23" > (p \supset q))$, $(q \supset p)$, $(p \supset (q \supset p))$, and so on. (The ‘ $\supset$’ stands for material implication; so an expression of the form ‘ $x \supset y$’ is shorthand for ‘either not-*x* or *y*’.) The result will be a set that is not only tightly coherent but also as large as one may like. Yet it will obviously lack a desirable sort of scope or comprehensiveness none the less. Such comprehensiveness is hence not just a matter of numerosity; scope of subject matter also counts.

Such comprehensiveness, even when tightly coherent, still will not suffice for justification, however. Take one's I/now perspective. Suppose one replaces one's concepts of (1) oneself, and (2) the present, wherever they may figure in one's vast system of object level beliefs, by corresponding concepts of (1′) the holder of passport number *n*, and (2′) 18 May 1998. This system will nearly match one's original system in true comprehensiveness and in interlocking coherence. Yet there is little semblance of justification in it. One will be wildly unjustified in attributing to the holder of passport number *n*, as of 18 May 1998, the vast set of things that one attributes to oneself now. The coherentist needs to require interlocking comprehensiveness not only with regard to the object level of beliefs, but also in a way that reaches up to meta-levels, where one takes note, at least implicitly, of the sources of one's beliefs and of how reliable these are, and so on. Once this is required, the transformation from the I/now system to the passport number *n*/18 May 1998 system is blocked.

Even that seems insufficient, however, if it is conceivable that such a comprehensively coherent system of beliefs could still fail to mesh properly with the subject's sensory experience. This failure of mesh might occur in either or both of two ways. Someone might experience as if *p* and might have no reason to question or resist such prompting to believe that *p*, but might none the less believe something, that *q*, incompatible with the proposition that *p*, and might even believe that $\sim p$ for good measure. In addition they might have a conspicuous experience *e* - a splitting headache, say - and might none the less believe that they do *not* have such an experience. If such failure of mesh is indeed possible, then it could preclude even a highly coherent and comprehensive system of beliefs from rendering its member beliefs justified.

Coherentism seems well advised to adopt such requirements of (1) comprehensiveness, (2) perspectival content, and (3) mesh with experience. With such needful qualifications and improvements, coherentism is on a rapprochement course with classical foundationalism as accounts of (internal) justification. This seems especially clear in light of the fact that foundationalism, for its part, seems forced to withdraw its objections to mutually supportive beliefs, and must willingly allow that the appropriate internal coherence of a body of beliefs can indeed account for much of the justification that resides in the member beliefs. The potential for a meeting of the minds is evident if we put it this way: the coherence required for epistemic justification in a system of beliefs requires that the system of beliefs be appropriately comprehensive. And this in turn requires that the system of beliefs should include an epistemic (meta) perspective, *and* a suitable complement of *foundational* beliefs! Once it is put thus, the game is obviously up. Both sides score significant points. Coherentism scores for its emphasis on appropriately comprehensive (and perspectival) coherence, an indispensable component of our concluding view. And foundationalism scores for its emphasis on the extra-belief components that are clearly needed in an appropriately justified system: especially the need for appropriate mesh with experience.

See also: Certainty; Empiricism; Justification, epistemic; Knowledge, concept of; Rationalism; Reasons for belief

ERNEST SOSA

References and further reading

Alston, W. (1990) *Epistemic Justification: Essays in the Theory of Knowledge*, Ithaca, NY: Cornell University Press.(Influential essays by a distinguished epistemologist.)

Audi, R. (1993) *The Structure of Justification*, Cambridge: Cambridge University Press.(Collected essays by an important defender of a moderate foundationalism.)

Chisholm, R.M. (1982) *The Foundations of Knowing*, Minneapolis, MN: University of Minnesota Press. (Influential defence of foundationalism, especially in chapters 1 and 6.)

Kuhn, T. (1977) *The Essential Tension*, Chicago, IL: University of Chicago Press.(See especially 'Objectivity, Value Judgment, and Theory Choice' for a case in favour of the reality and importance of epistemic dilemmas.)

Lewis, C.I. (1946) *An Analysis of Knowledge and its Valuation*, La Salle, IL: Open Court.(Thorough defence of a radical form of foundationalism by a major figure.)

Mill, J.S. (1861) *Utilitarianism*, London: Dent, 1971.(Classic account of right action unified under its single overarching principle of utility. Many editions are available.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press, 1930.(Main work in ethics by a major founding-figure of analytic philosophy.)

Moser, P. (1989) *Knowledge and Evidence*, Cambridge: Cambridge University Press. (Sophisticated book-length development of strong foundationalism, the most thorough and unified.)

Reid, T. (1785) *Essays on the Intellectual Powers of Man*, ed. B. Brody, Cambridge, MA: MIT Press, 1969.

(Classic work whose importance is growing steadily.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Clarendon Press.(Important work that defends radical pluralism.)

Sosa, E. (1991) *Knowledge in Perspective*, Cambridge: Cambridge University Press. (Detailed discussion of the foundationalism-coherentism controversy, especially in parts II and III.)

Francis of Meyronnes (d. after 1325)

Francis of Meyronnes, the doctor illuminatus *(Enlightened Doctor), was called the 'Prince of the Scotists' for his work in systematizing and propagating the philosophy of Duns Scotus in the fourteenth century. His work in metaphysics and theology, while heavily dependent on Scotus, shows originality and independence of mind, and is characterized by his dedication to finding rational defences of Catholic doctrine. His discussion of Ideas includes a critique of Aristotelian metaphysics, and he argues instead for a position based on his conception of Platonism.*

Born in Provence, Francis joined the Franciscan order and studied theology at the University of Paris, where he was probably a pupil of John Duns Scotus some time between 1304 and 1307. In 1323 he was awarded the degree of Master of Theology. Prior to this, in the academic year 1320-21, he represented the Scotist school in a spirited debate with the Thomist Peter Roger (later Pope Clement VI) over Trinitarian theology. However, Francis was never merely Scotus' mouthpiece, but was an original and creative thinker. He often addressed philosophical issues in ways Scotus did not, and criticized Scotus when he found the latter lacking. Francis's written work, which shows the influence also of Augustine, Pseudo-Dionysius and Avicenna (see Ibn Sina), includes sermons, commentaries on the Bible, studies of theology, metaphysics and moral philosophy, and commentaries on the works of Aristotle, Pseudo-Dionysius and Augustine. His writings became very popular in the later Middle Ages, in particular his *Conflatus*, a revision of his commentary on Book I of Peter Lombard's *Sentences* (see Lombard, P.).

Francis argues that theologians cannot demonstrate the doctrine of the Trinity, but they can defend it against objections. According to this doctrine, as Francis understands it, God is a Trinity of three persons - Father, Son and Holy Spirit - three things, each of which is the same as the divine essence (see Trinity). Francis recognizes that this doctrine is vulnerable to criticism. For instance, the Father begets the Son, and the Son is begotten by the Father. However, the divine essence, which is the same as the Father, does not beget; nor is it begotten, although it is the same as the Son. Either the principle of non-contradiction does not apply to God, or each divine person must be distinct in some way from the divine essence.

Accepting the first alternative would, of course, put an end to all rational inquiry about God. Francis holds instead that each person of the Trinity is distinct in some way from the divine essence. These are not real distinctions, for real distinctions hold between things, and the doctrine of the Trinity maintains that God is three things, not four. Nor, he argues, are the distinctions purely conceptual: mental constructs that do not reflect any distinction in the Trinity itself. The sort of distinction that holds between a person and the divine essence lies between the real and the purely conceptual. Following Scotus, Francis calls this the *formal* distinction.

Scotus makes the formal distinction a cornerstone of his philosophy, and Francis builds on that cornerstone. To say of *x* and *y* that they are formally distinct is to say that they are really the same - they are not two things - but that *x* can be characterized differently than *y*, and the different characterizations accurately capture facts about *x* and *y* and are not simply mental constructs. By appealing to the formal distinction in order to explain how each divine person is rightly characterized in a way different from the essence, Francis attempts both to defend the orthodox doctrine of the Trinity and to preserve the possibility of rational theology. Furthermore, Francis argues, this solution carries the weight of authority. Examining the history of Trinitarian theology, Francis finds that orthodox theologians such as Bonaventure and Aquinas had been working toward this solution, though they had not fully articulated it.

Francis's originality displays itself in the *Conflatus*, in a series of articles on Ideas. He begins by asking whether there are Ideas in the mind of God. Ideas conceived in this way - the way theologians conceive of them - are the eternal, immutable archetypes of created things. Francis, however, does not see any necessity for concluding that there are such Ideas in God's mind. He explores Augustine's reasons for maintaining that there are Ideas in God, but finds them inconclusive. Nevertheless, he finds the view that there are Ideas in God's mind a plausible one, and he is willing to concede this point on the authority of Augustine.

According to Francis, when metaphysicians discuss Ideas, they are referring not to divine, immutable archetypes, but to the quiddities, or defining characteristics, of things. Are these Ideas, as Plato argues, separated from

particulars? Francis recounts the commonly held view that the quiddities are conjoined to individuating conditions in particular things, but the intellect can abstract a quiddity from its individuating condition. According to this view, there are indeed separated Ideas, but only in intellects. Francis rejects this view, offering instead what he takes to be Plato's position: there are separated Ideas prior to any activity of the intellect, and in fact it is the Idea's separation that grounds its abstractability by the intellect. Ideas are separated not only from the individuating conditions, but also from potentiality and actuality, existence, time and place. The Idea *equinity*, for example, is not actual or potential, existent or nonexistent, located in any place or at any time. It is, as Avicenna maintains, just equinity.

The reason so few philosophers have subscribed to Plato's view, Francis suggests, is that Aristotle misrepresented it to posterity. According to Francis - who had almost no direct access to Plato's works - Aristotle ascribes to Plato the ridiculous view that the Ideas are bizarre particulars which exist in the air. Francis thinks that Aristotle, motivated by jealousy, tried to sabotage Plato's reputation. Otherwise we would have to say that although Aristotle was the best natural philosopher, he was the worst metaphysician, because he did not understand abstraction. Plato, Francis supposes, never held that Ideas are separated from particulars locally, but only formally. On Francis' view, the formal distinction is a cornerstone not just of Scotism, but of Platonism as well (see Nominalism; Realism and antirealism).

See also: Platonism, medieval

JEFFREY HAUSE

List of works

Francis of Meyronnes (1320-1) *In libros sententiarum (Commentary on the Sentences)*, first printed edition Venice, 1520; repr. Frankfurt: Minerva, 1966.(An extensive discussion of a vast array of philosophical and theological issues raised in or suggested by Peter Lombard's *Sentences*.)

Francis of Meyronnes (1323-25?) *Quodlibeta (Quodlibetal Questions)*, first printed edition Venice, 1520; repr. Frankfurt: Minerva, 1966.(Records Francis' answers to a series of questions originally posed during disputations by members of the academic community. The questions concern diverse issues ranging from Trinitarian theology to the obedience owed by secular rulers to the church.)

References and further reading

Maurer, A. (1990) *Being and Knowing: Studies in Thomas Aquinas and Later Medieval Philosophers*, Toronto, Ont.: Pontifical Institute of Medieval Studies.(Contains two earlier articles by Maurer, one on Francis' epistemology and his disagreements with Ockham and Aureol (311-31) and the second on Francis' account of infinity as the primary mode of being in God (333-59).)

Roth, B. (1936) *Franz von Mayronis, O.F.M. Sein Leben, seine Werke, seine Lehre vom Formalunterschied in Gott (Francis of Meyronnes, OFM: His Life, Works and Teaching on the Formal Distinction in God)*, Werl-in-Westfalen: Franziskus-Druckerei.(The most complete study of Francis's life and works.)

Vignaux, P. (1962) 'L'être comme perfection selon François de Meyronnes' (Being as Perfection according to Francis of Meyrones), *Études d'histoire littéraire et doctrinale*, Publications de l'Institut d'Études Médiévales, Paris: Vrin, and Montreal: Institut d'Études Médiévales, 259-318.(An excellent but technical essay focusing on Francis's differences from Scotus, in particular his Platonism.)

Frank, Jerome (1889-1957)

Jerome Frank was a significant contributor to the 'realist' movement in US legal theory. He is most closely associated with 'fact scepticism', the view that legal processes, especially court processes, are afflicted with pervasive uncertainty and unpredictability because of the difficulties of finding out with certainty or even strongly justified confidence what happened in the past.

Frank himself distinguished 'fact scepticism' from rule scepticism. By contrast with doubt arising from uncertainty in interpreting rules, 'fact scepticism' concerns the difficulties of finding out about past events, difficulties inevitably exacerbated by the adversarial character of legal processes conducted under common law, most notably where a jury is involved. Since 'facts' in the legal sense are both uncertain and malleable, the personal predilections and intuitions of the trier of fact come to be primary determinants of the practical outcomes of trials and litigations (see Legal evidence and inference §1; Legal reasoning and interpretation §§1,2).

Frank's theoretical views about law arose from his engagement in legal practice and in New Deal politics, and finally from experience as a Federal appellate judge in the USA. Although he taught occasionally at Yale in his later years, he was never a full-time academic lawyer. His philosophical contributions are the reflections of a legal practitioner, not of a philosopher of law. In fact, he came to legal practice with considerable reluctance and under pressure from his family. This may account for the fact that quite early in his career he underwent psychoanalysis, becoming in consequence an advocate of the relevance of Freud's theories to legal understanding. Humans' delusory belief in and search for certainty in law he likened to the search in life for a father-figure as a repository of unquestionable authority. The process of achieving adulthood was the process of abandoning the myths of certainty and acknowledging the irreducible indeterminacy and negotiability of practical life. This case was put vigorously in his *Law and the Modern Mind* (1930), which achieved some standing as an account of realist ideas accessible to the general reader. His later work, culminating in *Courts on Trial* (1949), carried forward his critique of adversarial processes and over-reliance on juries and other characteristic institutions of the common law.

Despite his scepticism concerning the determinability of past facts, he retained an ideal of a legal system that could function more objectively and predictably than the one in which he worked and to which he contributed a great deal both intellectually and judicially. As a judge of the Courts of Appeals for the Second Circuit (a position to which he was appointed by President Roosevelt in 1941), he showed himself to be a friend of civil liberties and a proponent of judicial activism especially in favour of individuals under threat from encroachment by the state. While many commentators have found his doctrines somewhat overstated, and indeed they do contain elements easy to caricature, there is no doubt that Frank left an important mark upon jurisprudence. After him, it has been impossible for serious thought about the judicial process to ignore the crucial elements of fact-finding, evidence and proof, and the special problems attendant on these subjects.

See also: Law, philosophy of; Legal realism §§2-3

NEIL DUXBURY
NEIL MacCORMICK

List of works

Frank, J. (1930) *Law and the Modern Mind*, Birmingham, AL: Legal Classics Library, 1985.(This is a reprint of the original edition, including a useful introduction by Julian Mack.)

Frank, J. (1949) *Courts on Trial*, Princeton, NJ: Princeton University Press.(This collection of papers expresses Frank's mature thought on the wide range of subjects about administering justice that his life's work focused on.)

References and further reading

Chase, A. (1979) 'Jerome Frank and American Psychoanalytic Jurisprudence', *International Journal of Law and Psychiatry* 2: 329-48.(A useful account of the influence of psychoanalytic thought on Frank, and of his application of it to law.)

Duxbury, N. (1991) 'Jerome Frank and the Legacy of Legal Realism', *Journal of Law and Society* 18: 175-205.(A particularly helpful and balanced summary of Frank's thought and influence.)

Glennon, R.J. (1985) *The Iconoclast as Reformer: Jerome Frank's Impact on American Law*, Ithaca, NY: Cornell University Press.(Valuable, constructively critical, biography and account of Frank's work and influence.)

Frank, Semën Liudvigovich (1877-1950)

The philosophy of S.L. Frank was one product of the renewed interest in epistemology, speculative metaphysics and religion among educated Russians in the quarter-century preceding the Revolution of 1917. Frank published the first volume of his philosophical system in 1915, but most of his major works were written after the Revolution, in European exile.

Influenced by tendencies in turn-of-the-century European thought that criticized the exaggerated pretensions of scientific reason, Frank formed the conviction that abstract, conceptual thought was inherently incapable of mastering ultimate reality. A valid metaphysics was nevertheless possible, founded on our capacity for direct, intuitive apprehension of reality in its living concreteness.

In intuitive knowledge, reality discloses itself as a 'total-unity' - an all-embracing unity in which the dualities with which conceptual thought wrestles are overcome without being dissolved. Ultimate reality is itself grounded in, and embraced by, a principle Frank termed 'Divinity', one that manifests itself in religious experience as the personal God of Christian faith. The rootedness of the human person in this divine principle is the condition of possibility of all spiritual creativity - of art, science, morality and law, and religion.

1 Life; sources of his thought

Semën Liudvigovich Frank was born in Moscow, of Jewish parentage. Converted to Marxism at the age of 16, he chose to devote his studies at Moscow University (1894-8) to political economy. In 1899, he was briefly detained by the tsarist police for his involvement in revolutionary activities. He subsequently attended lectures on economics and philosophy at the universities of Berlin and Heidelberg. By 1901, he had ceased to consider himself a Marxist, but remained a proponent of social as well as political democracy for Russia. In 1903, he began his long, intimate journalistic collaboration with P.B. Struve, a pioneer of the philosophical revival and a leading figure in the emerging Russian liberal movement. As a supporter of constitutional government, Frank initially hailed the Russian Revolution of 1905 with enthusiasm; its ultimate failure precipitated major changes in his views. After 1906, Frank was an unsparing critic of socialism; of the many articles he wrote criticizing the beliefs and mentality of the Russian left, the most famous is his contribution to the 1909 *Vekhi* (*Signposts*) symposium (see Signposts Movement §1-3). At this time, he also abandoned the Fichtean idealism professed in his early philosophical writings in favour of the spiritualist metaphysics characteristic of his mature works.

Although Frank now became the advocate of a religious outlook on life, he never returned to Judaism. In 1908, he married a Gentile; in 1912, he was baptized into the Russian Orthodox Church.

Soon after his marriage, Frank resolved upon a career as a professional philosopher. After passing his master's examination at the University of St Petersburg, he was appointed, in 1912, as an unsalaried lecturer there. He subsequently held professorships at the universities of Saratov and Moscow.

Frank was one of many prominent anti-Marxist academics arrested in 1922 and deported from Soviet Russia. He spent the first fifteen years of his exile in Berlin, where he taught at various émigré educational institutions, and, briefly, at the University. His Jewish ancestry led to his dismissal from his posts after Hitler came to power, but he was not able to escape from Germany until 1937. The final dozen years of his life, during which he wrote his last three books, were passed in France and Britain.

In exile, Frank styled himself a 'Christian Platonist', claiming Plotinus and Nicholas of Cusa as his 'masters' in philosophy. But his mature thought was actually shaped by a great variety of philosophical influences. The Kantian tradition, with which he associated himself in his earliest philosophical writings, impressed a permanent stamp on his thinking about epistemology (see Neo-Kantianism). Other, countervailing influences on his mature epistemological views were the writings of Goethe, Bergson, N.O. Lossky and Husserl. Frank's thought on logic was indebted to both Hegel and Hermann Cohen. While his mature metaphysical system bears a broad resemblance to that of Vladimir Solov'ëv (who popularized the term 'total-unity' in Russia), the extent of Solov'ëv's actual impact on Frank has arguably been overstated by historians of Russian philosophy.

2 First philosophy (epistemology, logic, metaphysics)

In his first major work, *Predmet znaniia* (*The Object of Knowledge*) (1915), Frank maintained that epistemology, logic and metaphysics constitute a single philosophical discipline, to which (following Aristotle) he gave the name ‘first philosophy’. Neither epistemology nor logic could claim the rank of an autonomous branch of philosophy, since a correct solution to their fundamental problems could only be attained through an inquiry which brought the problem of knowledge into relation with the problem of being. Metaphysics was not an independent discipline either, since an understanding of the true character of ultimate reality could only be reached by way of an analysis of human knowledge.

Frank distinguished three different modes of knowledge: ‘abstract’ or ‘conceptual’ knowledge; ‘intuition’; and ‘living knowledge’. Abstract knowledge, which relates particular concepts to each other in judgments, is generally thought to be the only kind of knowledge there is: both our everyday understanding of the world and organized science are systems of conceptual judgments. But Frank contended that we cannot explain how abstract knowledge is possible unless we recognize that it is dependent upon a prior form of cognitive access to reality: a direct intuition of being as an integral whole, unmediated by concepts.

The concepts with which abstract knowledge operates are fixed ‘determinations’, subject to the logical laws of identity, contradiction and the excluded middle. To account for our ability to affirm, in the judgment ‘A is B’, a necessary connection between two discrete, mutually exclusive conceptual contents, the philosopher must concede that we already ‘possess’ both A and B in a form in which they are not discrete determinations, but moments of a ‘metalogical unity’ to which these logical laws do not apply. Valid conceptual knowledge becomes possible only after we have intuited the necessary connection between A and B as nondiscrete elements in a continuous whole. This continuous whole is the ‘total-unity’, the all-encompassing unity outside of which nothing is conceivable. All particular knowledge is partial knowledge of this whole; abstract knowledge can be systematic only because the elements of the total-unity constitute a system.

We can possess the total-unity prior to the cognitive acts that engender conceptual knowledge because the total-unity transcends the opposition between being and knowledge; it is not simply ‘being’ but also ‘thought’; it is not objective being (being for a subject) but absolute being, a reality that reveals itself to itself. As existent entities, we humans belong to being; our being is the being of the total-unity. Accordingly, we participate, albeit imperfectly, in the knowledge reality has of itself. This ontological communion we enjoy with the reality we seek to know is the supreme condition of possibility of all knowledge.

In abstract knowledge, and even in intuition, our ontological communion with the object of knowledge is imperfect: the knowing subject remains enclosed in itself and merely sends out ‘rays’ which illuminate its object. Living knowledge, the third mode of knowing distinguished by Frank, involves a much deeper kind of communion with the object: a mutual penetration, as a result of which the knower merges with the thing known and, as it were, experiences it from inside. Living knowledge was not, for Frank, a rare or esoteric phenomenon; it is instanced in every genuine encounter with another ‘I’; in the uncanny insights of the real expert dealing with the objects of their passionate concern; and in all aesthetic, moral and religious experience.

Philosophy itself, as Frank conceived it, is ‘transcendental thinking’: living knowledge of absolute being as the ground of rational, conceptual thought. Since absolute being transcends the oppositions between concepts which are ultimate for abstract knowledge, philosophy can find verbal expression only in the form of ‘antinomian knowledge’. In speaking of the total-unity, the philosopher must couple every affirmative judgment with a negative judgment that logically contradicts it, but in uttering these judgments does not affirm any contradiction within the total-unity itself.

3 The metaphysics of human nature; the doctrine of Godmanhood

Frank’s metaphysics of total-unity provided him with a philosophical framework within which to examine a wide range of questions relating to our place, as human beings, in the universe and to the meaning and purposes of human life. His second book, *Dusha cheloveka* (*Man’s Soul*) (1917) shows that most of the principles which guided his mature thinking on these subjects had been formulated before the Revolution; but his ‘metaphysics of human nature’ and philosophy of religion were fully elaborated only in his writings of the 1930s and 1940s.

The foundation of all Frank’s thinking about human life was a conception of the ‘person’ as a ‘divine-human’ entity. Frank labelled this conception - a complex amalgam of Neoplatonic, Romantic and Christian ideas - the

doctrine of ‘Godmanhood’.

According to Frank, when we contemplate our own inner life - prescinding from the varied interactions with the external world which usually monopolize our conscious attention - we encounter an elusive mode of being which he terms ‘psychic life’ or ‘immediate self-being’. ‘Pure’ psychic life may be characterized as a restless chaos of formless potentiality; but we almost always experience psychic being as ‘formed’ by active forces which penetrate it, organize it and direct its flow into particular channels. Some of the forces which shape psychic life have their origin in this sphere of being itself. But the most exalted of them - those which direct the soul in its highest creative activities - emanate from a sphere of being that lies beyond the threshold of the individual self, the sphere of supra-individual ‘spiritual life’. The true ‘centre’ of the *person* lies outside the individual human self, in the realm of spirit, and through spirit is rooted in the divine principle which grounds all things. To conceive the person as an isolated individuality is the great error of modern humanism, an error which takes its revenge upon humanism in fascist and communist ideology and practice.

True philosophy, for Frank, must therefore be ‘religious’ philosophy. He believed that the eternal truths of religion had found their most perfect historical embodiment in the Christian faith. But a philosophy which is religious, while harmonizing with the essential teachings of Christianity, will rest on universal truths revealed in the philosopher’s own religious and metaphysical experience; it need not derive its premises from the positive revelation preserved by the Church.

4 Social philosophy

Frank outlined the fundamental theses of his social philosophy in *Dukhovnye osnovy obshchestva* (*The Spiritual Foundations of Society*) (1930); in *Svet vo t’me* (*The Light Shineth in Darkness*) (1949), he discussed the dilemmas posed for a Christian social ethics by the power of evil in human affairs.

On the empirical plane, society presents itself to us as an aggregate of competing individuals, who accept a measure of subordination to a guiding social will in the form of personal ‘power’ or impersonal ‘law’. The error of liberal social theory is to suppose that the inner essence of society coincides with this, its external manifestation. In its essence, apparent only to living knowledge, society is a primary ontological reality, no less ‘real’ than the individuals who constitute it, a spiritual communion founded on the shared obligation of service to divine truth. The impossibility of fully realizing the essential nature of human social life in a world subject to sin expresses itself in the unresolvable tension between the demands of inner moral life and the imperatives of law and outward social morality.

Since the late 1980s, many of Frank’s works have been reprinted in Russia. Until then, his influence was largely confined to Russian émigré circles. There, his achievement was much esteemed; V.V. Zenkovsky ranked him as the most eminent of all Russian philosophers. Frank wrote a philosophical prose remarkable for its clarity and concision.

See also: Russian Religious-Philosophical Renaissance §4

PHILIP J. SWOBODA

List of works

Frank, S. (1909) ‘Ėtika nigilizma’, in *Vekhi: Sbornik statei o russkoi intelligentsii*, Moscow: M. Sablin; trans. M. Shatz and J. Zimmerman, ‘The Ethic of Nihilism: A Characterization of the Russian Intelligentsia’s Moral Outlook’, *Vekhi (Landmarks): A Collection of Articles about the Russian Intelligentsia*, Armonk, NY: M.E. Sharpe, 1994, 131-55.(Frank’s extremely influential critique of the Russian revolutionary left’s incoherent ethical outlook; the English translation supplies historical background helpful in understanding the essay.)

Frank, S. (1915) *Predmet znaniia: Ob osnovakh i predelakh otvlechënnogo znaniia (The Object of Knowledge: On the Grounds and Limits of Abstract Knowledge)*, Petrograd: R.G. Shrëder; abridged French trans. Kaffi, Oldenbourg and Fedotoff, *La connaissance et l’être*, Paris: Aubier, 1937.(Frank’s fundamental work: a dense, technical and closely argued exposition of his ‘first philosophy’.)

Frank, S. (1917) *Dusha cheloveka: Opyt vvedeniia v filosofskuiu psikhologiiu*, Petrograd: Sakharov & Leman; trans. B. Jakim, *Man’s Soul: An Introductory Essay in Philosophical Psychology*, Athens, OH: Ohio University Press, 1993.(A subtle and complex analysis of human psychic life and its relation to other modes of being. ‘Foreword’ by P. Swoboda sets the work in its historical and philosophical context.)

Frank, S. (1930) *Dukhovnye osnovy obshchestva: Vvedenie v sotsial'nuiu filosofiiu*, Paris: YMCA-Press; trans. B. Jakim, *The Spiritual Foundations of Society: An Introduction to Social Philosophy*, Athens, OH: Ohio University Press, 1987. (A short, fairly accessible treatise.)

Frank, S. (1939) *Nepostizhimoe: Ontologicheskoe vvedenie v filosofiiu religii*, Paris: Dom Knigi & Sovremennye zapiski; trans. B. Jakim, *The Unknowable: An Ontological Introduction to the Philosophy of Religion*, Athens, OH: Ohio University Press, 1983.(A tour of the entire system of reality, often regarded as Frank's most important book; skilfully written and challenging to read.)

Frank, S. (1946) *God With Us: Three Meditations*, trans. N. Duddington, London: Jonathan Cape; *S nami Bog: Tri razmyshleniia*, Paris: YMCA-Press, 1964.(A presentation of Frank's philosophy of religion, first published in translation, intended for the general reader.)

Frank, S. (1949) *Svet vo t'me: Opyt khristianskoi ėtiki i sotsial'noi filosofii*, Paris: YMCA-Press; trans. B. Jakim, *The Light Shineth in Darkness: An Essay in Christian Ethics and Social Philosophy*, Athens, OH: Ohio University Press, 1989.(An eloquent philosophical meditation on the struggle with evil in the world; very accessible.)

Frank, S. (1956) *Real'nost' i chelovek: Metafizika chelovecheskogo bytiia*, Paris: YMCA-Press; trans. N. Duddington, *Reality and Man: An Essay on the Metaphysics of Human Nature*, New York: Taplinger, 1965. (Covers much the same ground as *Nepostizhimoe* on a more popular level.)

References and further reading

Boobbyer, P. (1995) *S.L. Frank: The Life and Work of a Russian Philosopher, 1877-1950*, Athens, OH: Ohio University Press.(The first full-length biography of Frank; some discussion of his philosophy; good bibliography.)

Copleston, F. (1986) *Philosophy in Russia: From Herzen to Lenin and Berdyaev*, Tunbridge Wells: Search Press. (Chapter 13 offers a very brief but eminently readable account of Frank's philosophy.)

Edie, J., Scanlan, J., Zeldin, M. and Kline, G. (1976) *Russian Philosophy*, Knoxville, TN: University of Tennessee Press, vol. 3, 277-304.(Contains a brief sketch of Frank's life and thought, and two well-chosen excerpts from his late writings.)

Frank, V. (1980) *Bibliographie des oeuvres de Simon Frank*, Paris: Institut d'Études Slaves.(Helpful, although not altogether complete.)

Lossky, N. (1951) *History of Russian Philosophy*, New York: International Universities Press, 266-92.(A detailed and generally critical summary of the contents of *Predmet znaniia* and *Nepostizhimoe*; fairly difficult.)

Swoboda, P. (1995) 'Windelband's Influence on S.L. Frank', *Studies in East European Thought* 47: 259-90.(A discussion of the enduring influence of Neo-Kantianism on Frank's thinking; fairly difficult.)

Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, 2 vols; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, vol. 2, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, 852-72.(The standard history; both admiring and critical in its treatment of Frank; ignores the Western influences on his thought; somewhat challenging for nonspecialists.)

Frankfurt School

The origins of the circle of philosophers and social scientists now known as the Frankfurt School lie in the 1920s when a number of critics and intellectuals were attempting to adapt Marxism to the theoretical and political needs of the time. The distinguishing feature of the approach adopted by the Frankfurt School lies less in its theoretical orientation than in its explicit intention to include each of the disciplines of the social sciences in the project of a critical theory of society. The objectives of this theoretical innovation vis-à-vis all the traditional Marxist approaches were established by Max Horkheimer in various articles written in the 1920s and 1930s. His critique of neo-idealist philosophy and contemporary empiricism sought to develop a philosophy of history which would comprehend the evolution of human reason; in so doing, he drew on empirical research. Thus the Institute of Social Research, conceived as a way of realizing this plan, was founded in 1929. Its work drew on economics, psychology and cultural theory, seeking to analyse, from a historical perspective, how a rational organization of society might be achieved.

However, after the National Socialists came to power and drove the Institute into exile, historical/philosophical optimism gave way to cultural/critical pessimism. Horkheimer and Adorno now saw it as the function of a critical theory of society to try, by returning to the history of civilization, to establish the reasons for the emergence of Fascism and Stalinism. Their Dialectic of Enlightenment, which bears some resemblance to Heidegger, impressively testifies to this change of orientation: it asks why totalitarianism came into being and it identifies a cognitive and practical perspective on the world which, because of its concern with the technical control of objects and persons, only allows for an instrumental rationality.

But there was some opposition to this critique of reason which tended to view totalitarianism as a consequence of an inescapable cycle of instrumental reason and social control. The concept of total reification was called into question by some of the more marginal members of the Institute working under Adorno and Horkheimer. These were far more interested in asking whether, even under totalitarian conditions, they could determine the remains of a desire for communicative solidarity. The work of philosopher Walter Benjamin constitutes an analysis of the interrelation of power and the imagination; Franz Neumann and Otto Kirchheimer inquired into legal consensus culture and social control; while Erich Fromm conducted a psychoanalytic investigation of communicative needs and their potential for resistance.

After the core members of the School had returned from exile, the Institute resumed its work in Frankfurt and embarked on large-scale empirical projects. From the very beginning, however, a considerable gap existed between the empirical investigations which focused on the industrial workplace and the philosophical radicalization of negativity on which Adorno and Horkheimer worked, albeit with differing emphasis. This gap was bridged only when Habermas began to challenge the systematic bases of critical theory, causing the basic philosophical concepts and the intentions of empirical social research once again to correspond. The central idea, with which Habermas introduced a new phase in the history of the Frankfurt School, was his understanding of a form of rationality which would describe the communicative agreement between subjects rather than the instrumental control of things. The concept of communicative rationality which emerged from this idea has since formed the basis for the moral grounds and democratic application of critical theory.

1 Origins and aims of critical theory: Horkheimer and Marcuse

Critical theory stands out from the various attempts to develop a productive model of Marxism which took place between the World Wars. It differed from comparable approaches primarily in its methodological objectives rather than its theoretical principles. These objectives derived from an unquestioned and programmatic acknowledgement of the sciences. Critical theory’s fundamental aim was to include all the disciplines of the social sciences in the project of a materialist theory of society. It thereby overcame the theoretical purism which had long persisted in historical materialism, and made space for the fruitful integration of academic social science and Marxist theory. This methodological vision was most ably represented by Max Horkheimer in whose hands the project of a more broadly interdisciplinary Marxism developed during the 1920s.

When Horkheimer succeeded Grünberg as Director of the Institute for Social Research in 1930, an opportunity presented itself for the realization of this project. He revealed, for the first time, the theoretical programme of a

critical theory of society in his inaugural speech (Horkheimer 1972). In *Zeitschrift für Sozialforschung (Journal for Social Research)*, founded in 1932 and which, until 1941, represented the intellectual centre of the Institute's work, he attempted to develop this approach in collaboration with Herbert Marcuse. The idea of overcoming the scientific and historical gulf which had arisen between factual research and philosophy formed the methodological framework. Following a Hegelian model of history, both branches of knowledge were again to be fused in a single form of reflection to such a degree that the empirical analysis of reality could coincide with the philosophical determination of reason. To achieve this, one needed a theory of history which would be able to determine the powers of reason as they operated within the historical process. Although their interpretative approaches differed, both Horkheimer and Marcuse inherited these basic philosophical-historical conceptions from the tradition of Marxist thought.

In the 1930s, Horkheimer and Marcuse continued to represent the classical Marxist theory of history. According to this, the central mechanism of social progress lies in the development of the forces of production, and as the domination of nature moves through increasingly technical stages, these determine new stages in the relations of production (Horkheimer 1932). Thus critical theory was to enter into this historical development not merely as a cognitive example of the work process, as was the case in the empirical sciences, but as a critical example of society's self-knowledge. It is by critical theory, according to Marcuse and Horkheimer, that we are made conscious of the possibilities which are already developing within the historical situation itself (Marcuse 1937). They no longer believed, however, that rationality, as embodied in the contemporary forces of production, simultaneously finds expression in the revolutionary awareness of the proletariat. From the very beginning they were conscious that, because of the increasing integration of the working classes into the social system of late capitalism, a strictly Marxist theory had lost its social relevance. Hence the empirical analyses, which are regarded as an immanent part of the theory of society which was evolving, should also aim to explain precisely those social and psychic mechanisms by which all potential social resistance has become integrated.

Horkheimer's and Marcuse's preliminary philosophical reflections, not unlike those of their contemporaries Lukács and Korsch, continued along the general lines of a productivist philosophy of history, and the openness of their methodology to empirical social research enabled them to formulate something new. The idea of including all the social sciences in the development of a critical theory of society, as well as the subjects of their empirical research, led them far beyond other contemporary attempts to revive Marxism. As a point of reference for the entire research of the Institute, Horkheimer sought the origin of the psychic mechanisms that prevent the outbreak of conflict between social classes which would otherwise result from the tension due to their economic differences (Horkheimer 1932). This question represents a soberly empirical change in the problem of the revolutionary subject which until this time had been fraught with philosophical and historical difficulty. The focus of the Institute's interdisciplinary investigations throughout the 1930s was now clear, and the projects to which each member contributed their specialist knowledge were to elucidate the specific form of social integration which capitalism had systematically constructed in its post-liberal phase.

To Horkheimer belongs the credit of having defined the form and content of a critical theory of society. He used his position as director to acquire specialists for the wide-ranging tasks of the interdisciplinary research projects, and the Institute's programme emerged from the interconnection of three disciplines. Horkheimer continued to regard the research as being fundamentally based on the economic analysis of the post-liberal phase of capitalism. He entrusted his friend Friedrich Pollock with the task of implementing this analysis and answering the question of whether a different principle of capitalist organization was beginning to emerge from the new, planned economy. The psychological study of the individual's integration into society represented the next stage of inquiry. Erich Fromm, who at this time still spoke for the Freudian Left and the fusion of historical materialism and psychoanalysis, was to implement this project. He started with the proposition that the integration of individuals into the capitalist system of social domination took place alongside the formation and socialization of their psycho-sexual characteristics. Hence he made it his task to investigate the consequences of capitalism's structural change for socialization within the family and individual personality formation. Cultural theory and its analysis of the operations of mass culture completed the Institute's programme of inquiry. Horkheimer's definitions of the object of this cultural analysis were ambivalent, even contradictory (Honneth 1991: 1.1), but within a short time it was focusing on the ways in which adaptations in behaviour were brought about by the media of the new culture industry. Theodor W. Adorno and Leo Löwenthal were entrusted with the study of this pivotal cultural dimension

of the capitalist process of integration.

Horkheimer and his colleagues were able to bring a theoretical unity to this range of subjects by employing a very simple functionalism. This guaranteed a conceptual fusing of the individual studies only because every social process was examined according to its function in the reproduction and development of work within the context of society (see Functionalism in social science). The reasons for this reduction to functionalism were the very historical-philosophical premises which consistently underwrite the thoughts of Marcuse, Adorno and Horkheimer. Because no other type of social activity was allowed other than work, even on the level of social theory only the instrumental forms of social praxis could be considered. But in the face of this functionalism one necessarily loses sight of that very dimension of everyday praxis by which the socialized subjects communicatively reproduce, and creatively learn to direct, their common action (see Honneth 1987).

2 The turn to a negative philosophy of history: Theodor W. Adorno

Within the Institute itself, however, the concept of interdisciplinary social research only existed in a vital and constructive form until the beginning of the 1940s. The articles which Horkheimer contributed to the last issue of the *Zeitschrift für Sozialforschung* in 1941 announced a general change of orientation affecting both the philosophical-historical premises of critical theory and its political conception of itself. While the Institute’s sense of resignation in the face of Fascism increased throughout the 1930s, its work was still sustained by hopes of progress theoretically founded on the Marxist conception of history. In spite of scepticism about the established workers’ parties, Horkheimer still regarded the research of the Institute as an intellectual reflection of the workers’ movement. At the end of the 1930s, however, the world represented by the Institute finally collapsed. The practical-political experience of the closed circle of Fascism, Stalinism and capitalist mass culture, which seemed to have become a totalitarian whole, led Horkheimer and his colleagues to abandon all Marxist conceptions of social progress. It was the change from a positive to a negative concept of work which reshaped critical theory and introduced a new phase in the history of the Institute. Thus a critique of reason and progress, so radical that it could not but cast general doubt on the potential for political revolution within social relations, replaced the conception of progress founded on production.

The outstanding representative of this new conception of critical theory is not Horkheimer but Theodor W. Adorno. His thinking, more than that of anyone else, is marked by the historical experience of Fascism as the fate of civilization. This made him sceptical of the Institute’s original historical-materialist ideas of progress. Moreover, the interests of his intellectual development were so predominantly artistic that he was bound quite naturally to doubt the limited rationalism of the Marxist theoretical tradition. Under the influence of Walter Benjamin he had already started to develop effective methods of aesthetic interpretation for a materialist philosophy of history. In his collaboration with Horkheimer (which intensified during the late 1930s), whose earlier reading of Schopenhauer had prepared him for critical reflections on civilization, Adorno’s scepticism about progress came to the fore. The *Dialectic of Enlightenment* (1947), which they wrote together in the early 1940s, articulated this new theme within a negative philosophy of history. In its philosophical-historical approach this work undoubtedly surpasses the Institute’s earlier efforts. The totalitarian condition which the rise of Fascist systems was bringing to the world was no longer to be explained as a consequence of the conflict between forces and relations of production, but as a consequence of the inner dynamic of the development of human consciousness. Horkheimer and Adorno abandoned the theoretical framework of capitalism and made the process of civilization as a whole their frame of reference. Hence Fascism came to be seen as the final historical stage in a logic of disintegration which is intrinsic to the original form of survival of the species itself. The prehistorical efforts of instrumental thought, by which mankind learns to survive and suppress the diversity of nature, gradually reproduce themselves in the disciplining of mankind’s natural impulses, in the impoverishment of its sensory faculties, and in the development of social power relations. Thus an increasing reification, set in motion by the first acts of domination of nature and reaching its logical conclusion in Fascism, shapes the process of civilization.

Even this philosophical-historical thesis, which implicitly rests on a series of anthropological and ethnological studies and their more recent interpretations, relies on Marxist premises. The prehistory of the human species is reduced to the single dimension of nature’s manipulation by mankind, and thus the development of civilization is interpreted according to a pattern of the domination of nature. But the motifs of Romanticism and *Lebensphilosophie*, which Horkheimer and Adorno take up again in their research, reveal the process of the

adaptation of nature in a completely different light (see Romanticism, German; Lebensphilosophie). If humankind's freedom is seen to lie in its ability to devote itself directly to natural life, every act of instrumental control over nature must be interpreted as one step towards the self-alienation of the species.

Two different points arise from this negative turn in the Marxist philosophy of history. On the one hand the consideration of the process of civilization so categorically excludes the whole sphere of everyday communicative praxis that the social advances which may have occurred can no longer be accommodated. Consequently, the *Dialectic of Enlightenment* is compelled to ignore another dimension of civilization's progress, namely one which is not expressed in the intensification of the forces of production, but in the extension of legal freedoms and the scope for individual action. The second consequence is of a political nature and of no less importance for the development of critical theory: every kind of political praxis is interpreted as a form of coercive action and is therefore excluded from the range of available positive alternatives. Thus Horkheimer and Adorno ultimately deny themselves the possibility of mobilizing their own research activity in the real political sphere. While this conclusion might not lead to complete political self-destruction, it necessitates a conception of revolution which no longer sees even the slightest possibility of political emancipation through the radical overthrow of social relations - instead, it looks toward messianic intervention in the process of civilization. Hence in conclusions such as these, drawn in fact by Horkheimer himself in his essay the 'Authoritarian State' (1942), the *Dialectic of Enlightenment* corresponds with the ideas explored in Benjamin's 'Theses on the Philosophy of History' in which he was attempting to criticize the model of social democratic progress.

3 Neglected margins: Benjamin, Kirchheimer, Neumann, Fromm

The only alternative which could have emerged within the Institute to counter such negativistic theories would have been found within the kind of research practised by Franz Neumann and Otto Kirchheimer. They formed a group, together with Walter Benjamin and later Erich Fromm, who both only briefly or indirectly contributed to the Institute's research. Independently, the work of each of these authors stands for an anti-functionalist impulse which could have called into question the image of a totally administrated society, an image which would inevitably have led to the dual impasse of political self-destruction or revolutionary messianism. But the dominating philosophical position of Adorno, Horkheimer and Marcuse never allowed such alternative theoretical approaches to be properly represented within the Institute (Honneth 1987).

Neumann and Kirchheimer, who were both trained in jurisprudence and had come of political age in German social democracy, contributed legal and constitutional investigations to the work of the Institute from their New York exile. Their scholarly and political background convinced them from the outset that the Law is a central medium of control in bourgeois society. Within the rules of constitutional law they recognized the social substance of a political compromise which the different classes, variously empowered by the conditions of private capitalism, all agreed upon. This social-theoretical premise provided the background for the analyses in which Neumann and Kirchheimer examined the changes in constitutional law which accompanied the structural change of the economy under capitalism (Neumann 1978; Kirchheimer 1976). Thus they are far better able to account for the political and legal mediations of capitalist rule than Adorno and Horkheimer who, following Pollock's state capitalism thesis, had been bound to the model of a centrally controlled society.

What also makes Neumann and Kirchheimer's conception superior are the social-theoretical ideas which tacitly influence their investigations. From the outset they perceived the social order from a different perspective than that which prevailed in Horkheimer's circle. For them, social integration represents a process which does not just take place over and above the always unconscious fulfilment of the functional imperatives of society, but via political communication between the social groups. Dealing with questions concerning the constitutional state had brought Neumann and Kirchheimer up against the phenomenon of political legitimacy; thence arose the view that the constitutional order of society is always the expression of a general compromise or consensus between the political powers. Active interest in class disputes, characteristic of the Weimar Republic, led to a realistic assessment of the power relations of social interests; the power which arose from private-capitalist control of the means of production could not be overestimated. The compromise character of the social order as a whole was finally revealed to them through their work on Austro-Marxism: the institutional structures of a society must be understood as nothing other than momentary determinations of social contracts in which the various interest groups are implicated according to their particular power potential.

Within Neumann's and Kirchheimer's thinking this all makes for a concept of society at the centre of which stands the overarching process of communication between social groups. If this concept had been used to direct the Institute's theoretical inquiries, it would not only have prevented the uncritical adoption of ideas of the total integration of society, but could also have opened up new modes of political orientation for its own research.

4 Continuation and revision of critical theory: Jürgen Habermas

When the Institute for Social Research re-opened in Frankfurt in 1950, it resumed its research activities without direct reference to its socio-philosophical identity of the previous two decades. During the post-war years, the unifying theoretical line which could have linked the empirical research and philosophical reflection was broken. The unity and coherence of the Frankfurt School no longer existed.

While scarcely any common denominator could be found for the empirical research projects of the Institute, the idea of a 'totally administrated world' came to represent a standard (if temporary) point of reference for the socio-philosophical work. This idea runs through the cultural-critical writing of Horkheimer, Adorno and Marcuse like a leitmotif; the central premise of the state-capitalism theory became the general framework for an analysis of post-war capitalism. Hence the theoretical perspective on totalitarianism, which had already shaped the image of society in the *Dialectic of Enlightenment*, also determined the current investigations. Because administrative social control and the individual's readiness to adapt were meshed seamlessly together, the life of society seemed to be integrated in a stable and now unassailable system of coercion. The three authors drew quite different conclusions for the project of a critical theory of society from their common diagnoses of the present. Horkheimer's thinking was affected by a deepening Schopenhauerian pessimism until it turned to negative theology (Horkheimer 1985). Adorno continued to pursue a self-critique of conceptual thought in which the idea of a mimetic rationality, vicariously preserved in the work of art, remained the normative point of reference (Adorno 1973). Marcuse alone reacted to the pessimistic diagnosis of the present situation with an impulse to save the lost idea of revolution, thereby repositioning reason below the social threshold, in the sphere of the natural libidinal needs of mankind (Marcuse 1955).

In spite of their different objectives, the common background to these three approaches remained a philosophy of history in which historical development is interpreted as a process of technological rationalization which reaches its conclusion in the closed system of rule of contemporary society. Only one theory, at first scarcely recognizable as a point of departure within critical theory, relinquished the philosophical premises of this diagnosis. Jürgen Habermas had emerged from the Institute of Social Research as Adorno's assistant. Yet his theoretical origins and orientation had never had much in common with the philosophical tradition of critical theory; more significant for the development of his thought were the theoretical movements of philosophical anthropology, hermeneutics, pragmatism and linguistic analysis, to which the older generation remained hostile. Nevertheless, from the work of Habermas a theory gradually took shape which was so clearly motivated by the original aims of critical theory that it counts today as the only serious attempt at theoretical renewal.

The insight into the linguistic intersubjectivity of social action forms the basis of this theory. Habermas arrived at his central premise through studying hermeneutic philosophy and the linguistic analysis of Wittgenstein, from which he learned that human subjects are always already linked through the medium of language. The human form of life is distinguished by an intersubjectivity embodied in the structures of language; linguistic communication between subjects constitutes an absolutely fundamental prerequisite for the reproduction of social life. Habermas gives this thesis socio-theoretical weight by making it the starting point of a debate with the socio-philosophical and sociological tradition. He is critical of the tendency of modern social philosophy gradually to reduce all practical intersubjective interests to decisions based on technical criteria (Habermas 1968). Against sociological functionalism he asserts that a society's tasks of reproduction are always fixed by the normative self-conceptions of the socialized communicative subjects. Consequently the necessary functions of life as such are nowhere to be found in the relations of human life (Habermas 1963). He is therefore also led to a critique of Marxism which results in an expanded conception of the theory of action within history. If the human form of life is distinguished by the medium of linguistic communication, then social reproduction cannot be reduced, as it was in the theoretical writings of Marx, to the single dimension of work. Rather, the praxis of linguistically-mediated interaction must be regarded as no less fundamental a dimension of the history of mankind than the activity of manipulating nature (Habermas 1971).

But the decisive step taken by Habermas towards his own theory of society, and hence a revision of critical theory, came about through his investing both concepts of action (work and interaction) with different categories of rationality. Within the subsystem of rationalized action in which the social tasks of work and political administration are organized, the species develops through the accumulation of technical and strategic knowledge. But within the institutional frame, in which the integrating norms of society are reproduced, it continues to develop by means of the liberation of communicative constraints (Habermas 1970). During the 1970s, Habermas' ramifications of his theory follow this concept of society, whereby rationalized systems are differentiated from the sphere of everyday communicative praxis, and separate forms of rationalization are postulated for each social sphere. A universal pragmatism further elucidates the linguistic infrastructure of communicative action; a theory of social evolution is to help resolve the logic of the development of social knowledge and thus the dual process of rationalization; and finally, the further development of systematic theoretical conceptions is intended to determine the mechanisms through which areas of social activity evolve into rationalized systems.

Although these theoretical endeavours reach into the most diverse fields of knowledge, they all aim to establish a critique of society with a basis in the theory of communication. Thus in seeking to prove that the rationality of communicative action is so fundamental a condition of social development, Habermas' work constitutes a critique of Adorno's and Horkheimer's biased thesis, revealing their diagnoses of the tendencies of instrumental reification to be forms of social rationalization wholly defined by functionalism. In the *Theory of Communicative Action* (1987) this programme is systematized for the first time. The fruits of the various studies are brought together to reconstruct the rationality of communicative action within the context of a speech-act theory, to develop it into the basis of a theory of society by going through the history of sociological theories, and finally, to make it the point of reference for a critical diagnosis of the present (see Communicative rationality).

One of Habermas' most recent studies can be understood as an answer to the pessimistic thesis in which claims are made for a tendency towards the colonization of the living world. In *Between Facts and Norms* (1992) he reconstructs, in the form of a critical legal philosophy, the legal and cultural conditions under which the project of a further democratization of our society can be continued today.

See also: Critical theory

Translated from the German by Bridget Thomson
Bibliographical annotations by Michael Bull
AXEL HONNETH

References and further reading

Adorno, T.W. (1973) *Negative Dialectics*, London: Verso.(Adorno's final statement on the nature of philosophical analysis in which he attacks identity thinking, replacing it with the notion of the dialectics of non-identity.)

Adorno, T.W. (1986) *Aesthetic Theory*, trans. C. Lenhardt, London and New York: Routledge & Kegan Paul. (Unfinished and published posthumously, this links Adorno's critique of human domination over nature with a critique of the administered society. As such it represents a continuation of the earlier dialectics of Enlightenment.)

Arato, A. and Gebhardt, E. (eds) (1978) *The Essential Frankfurt School Reader*, Oxford: Blackwell.(Excellent selection of articles on the central concerns of critical theory by members of The Frankfurt School, with additional explanatory commentaries by the editors.)

Benhabib, S. (1986) *Critique, Norm and Utopia*, New York: Columbia University Press.(A major work which analyses the notion of critique within the Frankfurt School up to Habermas.)

Benjamin, W. (1968) 'Theses on the Philosophy of History', in *Illuminations*, ed. H. Arendt, trans. H. Cohn, New York: Harcourt, Brace & World.(Essential for an understanding of Adorno's thought on the nature of history as developed in *The Dialectic of Enlightenment* and after.)

Benjamin, W. (1978) *Reflections. Selected Works: Essays, Aphorisms, Autobiographical Writings*, trans. E. Jenncott, New York: Harcourt Brace Jovanovitch.(Good selection of Benjamin's works, with extracts from his reflections on cities and his autobiographical writings.)

Fromm, E. (1992) *The Revision of Psychoanalysis*, ed. R. Funk, Boulder, CO: Westview Press.(Important essays by early and influential Frankfurt-School thinker, linking the works of Marx and Freud.)

Habermas, J. (1963) *Theorie und Praxis*, Darmstadt: Hermann Luchterand Verlag; trans. J. Viertal as *Theory and Practice*, Boston, MA: Beacon Press, 1973.(Contains important essays on modern political theory, including

several on Hegel.)
Habermas, J. (1968) *Erkenntnis und Interesse*, Frankfurt: Suhrkamp Verlag; trans. J.J. Shapiro as *Knowledge and Human Interests*, Boston, MA: Beacon Press, 1971. (An important interpretation of philosophy from Kant to the pragmatists, arguing that epistemology must now be pursued as social theory.)
Habermas, J. (1970) *Zur Logik der Sozialwissenschaft*, Frankfurt: Suhrkamp Verlag; trans. S. Nicholson and J. Stark as *On the Logic of the Social Sciences*, Cambridge MA: MIT Press, 1988.(Early sketch of Habermas' interpretive approach and his critique of Gadamer's philosophical hermeneutics.)
Habermas, J. (1971) 'Technology and Science as Ideology', trans. J.J. Shapiro in *Toward a Rational Society: Student Protest, Science and Politics*, Boston, MA: Beacon Press, 1971.(Important essay which formulates a method for studying the relationship between technology, science and ideology.)
Habermas, J. (1987) *Theory of Communicative Action*, Cambridge: Polity Press, 2 vols. (Habermas' central work which attempts to provide a normative basis and conceptual framework for a new critical theory.)
Habermas, J. (1992) *Faktizät und Geltung. Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaates (Between Facts and norms: Contributions to a Discourse Theory of Law and Democracy)*, trans. W. Rehg, Cambridge: Polity Press, 1996.(Works out the legal and political implications contained in the earlier *Theory of Communicative Action* and presents a sociologically informed conceptualization of law and rights, together with a normative account of the rule of law and the constitutional state.)
Held, D. (1980) *Introduction to Critical Theory*, London: Hutchinson.(Detailed and judicial treatment which clearly explores the central and substantive concerns of the first generation of Critical Theorists.)
Honneth, A. (1987) 'Critical Theory', in A. Gidden and J. Turner (eds) *Social Theory Today*, Cambridge: Polity Press.(Clear exposition of Critical Theory in the light of Habermas' theory of communicative action, highlighting possible future directions.)
Honneth, A. (1987) *Critique of Power: Stages of Reflection of a Critical Theory of Society*, trans. K. Baynes, Cambridge, MA: MIT Press.(Offers an interpretation of the history of Critical Theory focusing largely on Adorno and Horkheimer and debating their work on the nature of power within the context of the later debate between Foucault and Habermas.)
Honneth, A. and Joas, H. (eds) (1991) *Communicative Action*, Cambridge: Polity Press.(Collection of articles debating the strengths and limitations of Habermas' theory of communicative action.)
Horkheimer, M. (1932) 'Geschichte und Psychologie' (History and Psychology), *Zeitschrift für Sozialforschung* 1: 125-44.(Horkheimer's first discussion of the role of psychic mechanisms within the historical and social realm, later implemented by Fromm in his empirical analysis of the political and psychic attitudes of German workers in the 1930s.)
Horkheimer, M. (1941a) 'The End of Reason', *Zeitschrift für Sozialforschung* 9: 36-304.(This early essay introduces key distinctive concepts which distinguished Critical Theory from orthodox Marxism - the domination of nature, the decline of the individual and the role of technology in Western culture.)
Horkheimer, M. (1941b) 'Art and Mass Culture', *Zeitschrift für Sozialforschung* 9: 290-304. (Early articulation of what was to become the orthodox view of Critical Theorists on the potential critical function of art in relation to 'mass culture'.)
Horkheimer, M. (1942) 'The Authoritarian State', in *The Essential Frankfurt School Reader*, ed. A. Arato and E. Gerbhardt, trans. People's Translation Service, Oxford and New York: Blackwell, 1982.(Radical statement linking working-class organizations and orthodox Marxism to the notion of an 'authoritarian state'; puts into political context Horkheimer's analysis of 'The End of Reason'.)
Horkheimer, M. (1968) *Kritische Theorie (Critical Theory)*, ed. A Schmidt, Frankfurt: Fischer, 2 vols.(Seminal essays from the 1930s.)
Horkheimer, M. (1972) 'Die gegenwärtige Lage der Sozialphilosophie und die Aufgaben eines Instituts für Sozialforschung' (The Current Situation in social Philosophy and the Tasks of an Institute for Social Research), in W. Brede (ed.) *Sozialphilosophische Studien (Social-philosophical Studies)*, Frankfurt: Fischer, 33-46. (Inaugural speech of 1930, laying out the theoretical programme to be followed by The Frankfurt School in subsequent years.)
Horkheimer, M. (1985) *Gesammelte Schriften (Collected Writings)*, ed. A. Schmidt and G. Schmid Noerr, Frankfurt: Fischer.(Collected writings of Horkheimer in fourteen volumes covering the period from 1914 to 1972.)
Horkheimer, M. and Adorno, T.W. (1947) *Dialektik der Aufklärung. Philosophiche Fragmente (The Dialectic of Enlightenment)*, trans. J. Cummings, London: Verso, 1979.(Seminal work in which a new Benjamin-inspired philosophy of history is introduced, moving Critical Theory away from an orthodox Marxian concept of history to one that places power and the domination of nature at the centre.)

Jay, M. (1973) *The dialectical Imagination: A History of the Frankfurt School and the Institute of Social Research, 1923-50*, London: Heinemann.(A model of intellectual clarity and accuracy, perhaps the best and most accessible introduction to the early work of the Frankfurt School available in English.)

Kirchheimer, O. (1976) *Von der Weimarer Republik zum Faschismus: Die Auflösung der demokratischen Rechtsordnung (From the Weimar Republic to Fascism: The Dissolution of the Democratic Rule of Law)*, ed. W. Luthard, Frankfurt: Suhrkamp.(Account of the transformation of the German legal system with the rise of fascism. Demonstrates the ideological nature of Nazi political theory in its claim that it had brought together the spheres of law and morality.)

Marcuse, H. (1937) 'Philosophie und kritische Theorie' (Philosophy and Critical Theory), *Zeitschrift für Sozialforschung* 6: 632-47.(Early, important essay by Marcuse which maps out the relationship of Critical Theory to traditional forms of bourgeois philosophy.)

Marcuse, H. (1955) *Eros and Civilisation. A Philosophical Inquiry into Freud*, Boston, MA: Beacon Press. (Radical re-reading of Freud which produces a dialectically inspired account of Freudian instinct theory. Combines an analysis of a historicized and liberatory set of instincts with a Marxist inspired theory of historical development.)

Neumann, F. (1978) *Wirtschaft, Staat, Demokratie. Aufsätze 1930-1954 (Economy, State, Democracy. Essays 1930-1954)*, ed. A. Sollner, Frankfurt: Suhrkamp. (Collected essays of Neumann covering his significant contribution to Critical Theory's work on political economy during the fascist period of German history.)

Wiggerschaus, R. (1994) *The Frankfurt School: Its History, Theories and Political Significance*, trans. M. Robertson, Cambridge, MA: MIT Press.(The most comprehensive history available, providing a thorough analysis and overview of the intellectual development of The Frankfurt School, a wealth of biographical detail and an analysis of internal politics and conflicts within the Institute.)

Franklin, Benjamin (1709-70)

Benjamin Franklin was a candlemaker's son who became a successful businessman, politician, diplomat, scientist, philosopher, writer, and social reformer. He played a major role in winning American independence and in American intellectual history. In philosophy Franklin was a deist, and struggled much with problems of morality and determinism.

Benjamin Franklin was born on 17 January 1706 in Boston, Massachusetts. He was apprenticed to his brother James to become a printer. James Franklin was a freethinker, under whose influence Benjamin became a deist (see Deism). After his brother's arrest, Benjamin went to Philadelphia in 1723. In 1725 he went to England, and while there published *A Dissertation on Liberty and Necessity, Pleasure and Pain* (1725). In this work, Franklin adopted an extreme deism based on Newtonian mechanics (see Newton, I.); he asserted a total determinism, which, together with the claim that God is perfect, led to the conclusion that whatever is is right, and to a complete denial of a self-determining human will. He viewed people as machines governed by pain and pleasure, the latter defined as the absence of pain, and claimed that since all pain ends at death, the amounts of pain and pleasure in life are always equal. He denied the existence of an afterlife or any difference in merit or demerit among creatures; all are equal since all do what God ordains.

Returning to Philadelphia in 1726, Franklin became a successful printer, and revised his earlier extreme views. By 1732 the belief that God commands prayers, which, according to the necessitarian view he had espoused, would be pointless, led Franklin to conclude that God has endowed us with free will, and does directly intervene in human affairs. The admission of free will introduced virtue and vice, and, influenced by the benevolence morality of the time, Franklin concluded that virtue consists in doing good to others. He believed this could best be done by helping others to become able to provide for themselves. This he thought was possible through the universal ownership of property; thus property had a moral function for him. Such ownership would make people free to engage in disinterested benevolence, serving the public good. Franklin was thus a believer in the doctrine of civic virtue, which he practiced through social reform and politics, particularly after he retired from business in 1748.

In 1747 Franklin began working on electricity. He was the foremost American scientist of his century. His one-fluid theory of electricity, his law of the conservation of electric charge, and his demonstration of the electrical character of lightning, made him an international celebrity and won him honorary degrees from Oxford and Cambridge. Franklin capitalized on his fame to serve first Pennsylvania and then the Colonies as their representative in England and subsequently in France. He was a member of the committee that drafted the Declaration of Independence in 1776. The treaty with France, which he negotiated in 1778, gave the colonies the French support necessary to achieve independence. He was a negotiator of the peace treaty with England, and served in the convention that drafted the Constitution of the United States.

See also: American philosophy in the 18th and 19th centuries; Jefferson, T.

MURRAY G. MURPHEY

List of works

Franklin, B. (1840) *The Works of Benjamin Franklin*, ed. J. Sparks, Boston, MA: Tappan, Whittemore & Mason, 10 vols.(An old edition, but covering the span of Franklin's life.)

(1709-70) *The Papers of Benjamin Franklin*, ed. L.W. Labaree *et al*., New Haven, CT: Yale University Press, 1960-, 30 vols; vols 1-14, vols 15-26, ed. W. Willcox, vol. 27 ed. C. Lopez, vols. 28-30 ed. B. Oberg.(The best modern scholarly edition of Franklin's writings in chronological order, but still in progress; the last volume issued is for 1779).

(1725) *A Dissertation on Liberty and Necessity, Pleasure and Pain*, London: Samuel Palmer's Printing House. (Addressed to his friend James Ralph, the pamphlet advances an extreme form of mechanisitic Deism designed to prove that 'all is right'.)

(1790) *Autobiography*, ed. D. Aaron, New York, Vintage, 1990. ed. J.A. Leo Lemay and P.M. Zoll, *The Autobiography of Benjamin Franklin: A Genetic Text*, Knoxville, TN: University of Tennessee Press, 1981. (The textual and publication history of the *Autobiography* is very complex, with portions of it appearing as early as 1790. Franklin's best statement of his views on virtue.)

References and further reading

Cohen, I. B. (1956) *Franklin and Newton*, Philadelphia, PA: American Philosophical Society.(Excellent study of Franklin's scientific work.)

Flower, E. and Murphey, M.G. (1977) *A History of Philosophy in America*, New York: Putnam, vol. 1, ch. 2. (Brief treatment of Franklin's philosophical views.)

Free logics

We often need to reason about things that do not - or may not - exist. We might, for example, want to prove that there is no highest prime number by assuming its existence and deriving a contradiction. Our ordinary formal logic, however (that is, anything including standard quantification theory), automatically assumes that every singular term used has a denotation: if you can use the term 'God' - if that term is part of your language - automatically there is a denotation for it, that is, God exists. Some logicians have thought that this assumption prejudges too many important issues, and that it is best to get rid of it. So they have constructed logics free of this assumption, called 'free logics'.

A 'free logic' is a system of quantification theory, with or without identity, that allows for non-denoting singular terms. In other words, some of the expressions that may be considered singular terms - individual constants, free variables, definite or indefinite descriptions - are not assigned any object in some of the models of the system. In a system of *standard* quantification theory, schemata such as $\exists x(x = a)$ are logical truths; not so in free logics. Free logics reject the 'principle of particularization', $A(t/x) \supset \exists xA$, and all equivalent principles or rules, among them the 'principle of specification', $\forall xA \supset A(t/x)$, and the rules of existential generalization and universal instantiation.

Historically, contemporary logicians since Frege have tended to dispense with non-denoting singular terms, adopting either Frege's device of assigning them an arbitrary denotation or Russell's of denying them singular term status (via his theory of descriptions; see Descriptions). It was Henry Leonard (1956) who first proposed that standard quantification theory be revised to allow for non-denoting singular terms. Subsequently, Hailperin, Leblanc, Hintikka, Lambert and Smiley proposed various formal systems to this effect (between 1959 and 1967 - see Lambert 1991). But, until the mid-1960s, no semantic interpretation of these systems was provided, for good reason.

Standard semantics is based on the 'correspondence theory of truth': a statement such as 'Socrates is wise' is true if and only if the object corresponding to 'Socrates' has the property corresponding to 'wise'. But this scheme has no natural application to statements such as 'Pegasus is white', relating to nonexistent objects. So a semantics for a free logic (a 'free semantics') should provide for some alternative (or extension) to the correspondence theory, to account for those cases in which a singular term 'corresponds to' nothing - a problem for which it is difficult to find an uncontroversial solution.

There are currently three main kinds of free semantics. An 'outer domain' semantics adds to the ordinary domain of quantification a second ('outer') domain and allows singular terms to be interpreted on either domain. Intuitively, the outer domain is constituted by nonexistent objects and the reason why 'Pegasus is white', say, is true is that the nonexistent object corresponding to 'Pegasus' has the property corresponding to 'white'. A 'conventional' free semantics assigns a single truth-value (usually False) to all atomic formulas containing non-denoting singular terms, and then evaluates complex formulas accordingly. Both conventional and outer domain semantics are bivalent, and strong soundness and completeness theorems can be proved for them; that is, not only the set of logically true formulas but also the set of valid arguments can be axiomatized. Not so with the third kind of free semantics, 'supervaluational' semantics, due principally to Bas van Fraassen. This assigns truth-values to formulas containing non-denoting singular terms (relative to a model *M*), either by convention or by extending the domain, and then defines a 'supervaluation' (on *M*) that makes a formula true if it is true under all such assignments, false if it is false under all such assignments, and truth-valueless otherwise.

For illustration, suppose *a* is non-denoting in the model M, and consider the following formulas:

(1) Pa
(2) $\sim Pa$
(3) $Pa \vee \sim Pa$
(4) $Pa \;\&\; \sim Pa$.

There will be conventions on M (or extensions of its domain) that make (1) true, and others that make it false, and similarly for (2); so the supervaluation on M will leave both (1) and (2) truth-valueless. But *all* these conventions

(or extensions) will make (3) true and (4) false; so the supervaluation on M will agree on assigning them these truth-values. Intuitively, supervaluational semantics makes a statement containing non-denoting singular terms true (false) if it would be true (false) were its atomic components containing non-denoting singular terms to have *any* combination of truth-values (or its non-denoting singular terms to have *any* combination of denotations) - a ‘counterfactual’ theory of truth, as it is sometimes called.

Free logics have been applied in other parts of logic and to the formalization of scientific theories. Most typical of an application of the first kind is modal quantification theory, where a free logic base is quite natural (for consider: in the presence of the rule of ‘necessitation’, the theorem $\exists x(x = a)$ of standard quantification theory can be strengthened to $\Box\exists x(x = a)$ - that is, whatever we can name not only exists, but exists necessarily). Most typical of an application of the second kind is set theory, where a free logic base will let us use the *name* of, say, the Russellian set $\{x\colon x \notin x\}$ without being committed to its *existence* - indeed, while being able to prove that it does not exist (it cannot, by Russell’s paradox).

Finally, free logics must be distinguished from ‘inclusive’ logics, that is, logics that allow for an empty domain of quantification. Though it is natural to require inclusiveness of a free logic, and though indeed most free logics are also inclusive, the definitions of the two are distinct, and independent of one another.

See also: Existence; Free logics, philosophical issues in; Logical and mathematical terms, glossary of

ERMANNO BENCIVENGA

References and further reading

Bencivenga, E. (1986) ‘Free Logics’, in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 3, 373-426.(An ample survey of the subject.)

Lambert, K. (ed.) (1991) *Philosophical Applications of Free Logic*, New York: Oxford University Press.(A collection of articles illustrating various uses of free logic.)

Leonard, H. (1956) ‘The Logic of Existence’, *Philosophical Studies* 7: 49-64.(The first proposal that standard quantification theory be revised to allow for non-denoting singular terms.)

Free logics, philosophical issues in

The expression 'free logic' is a contraction of the more cumbersome 'logic free of existence assumptions with respect to both its general terms (predicates) and its singular terms'. Its most distinctive feature is the rejection of the principle of universal specification, a principle of classical predicate logic which licenses the logical truth of statements such as 'If everything rotates then (the planet) Mars rotates'. If a free logic contains the general term 'exists', this principle is replaced by a restricted version, one which licenses the logical truth only of statements such as 'If everything rotates then Mars rotates, provided that Mars exists'. If the free logic does not contain the general term 'exists', but contains the term 'is the same as', the principle is replaced by a version which licenses only statements such as 'If everything rotates then Mars rotates, provided that there is an object the same as Mars'.

Most free logicians regard the restricted version of universal specification as simply making explicit an implicit assumption, namely, that Mars exists. Indeed, free logic is the culmination of a long historical trend to rid logic of existence assumptions with respect to its terms. Just as classical predicate logic purports to be free of the hidden existence assumptions which pervaded the medieval theory of inference with respect to its general terms, so free logic rids classical predicate logic of hidden existence assumptions with respect to its singular terms.

There are various kinds of free logic, with many interesting and novel philosophical applications. These cover a wide range of issues from the philosophy of mathematics to the philosophy of religion. In addition to the issue of how to analyse singular existence statements, of the form '3 + 7 exists' and 'That than which nothing greater can be conceived exists', of special importance are issues in the theory of definite descriptions, set theory, the theory of reference, modal logic and the theory of complex general terms.

1 Kinds of free logic

Though there are various historical antecedents of free logic, it is a relatively recent development in the logic of terms - that is, in the development of predicate logic with or without identity. In fact, serious technical and philosophical study of this alternative to classical predicate logic originated only in the second half of the twentieth century with Henry Leonard's 1956 paper, 'Logic of Existence'. The technical rudiments of free logic were then worked out in the subsequent decade and a half, the expression 'free logic' being first coined in 1960 (by Karel Lambert). Since then many studies discussing or employing free logic in general philosophy, philosophical logic and in the philosophy of science have appeared.

'Free logic' is short for 'logic free of existence assumptions with respect to both its general terms (predicates) and its singular terms'. It restricts the principle of universal specification (which licenses the logical truth of statements such as 'If everything rotates then (the planet) Mars rotates') to a version which licenses only statements such as 'If everything rotates then Mars rotates, provided that Mars exists' (if the logic contains the general term 'exists') or such as 'If everything rotates then Mars rotates, provided that there is an object the same as Mars' (if the free logic does not contain the general term 'exists', but contains the term 'is the same as').

When a free logic also entertains the empty universe of discourse, it is called a universally free logic. In such comprehensive free logics, statements (perhaps without singular terms) such as 'There is an object which is tall or is not tall' and 'There is an object which is self-identical' - statements which Russell (§§9, 11) regarded as impure logical truths because of the evidently factual implication that there exists at least one thing - are also excluded from the class of logical truths.

In any free logic, universal or otherwise, the quantifier context 'There is an object…' has existential force just as it does in classical predicate logic. So in any free logic containing the general term 'is the same as' (the identity symbol), a statement such as 'Mars exists' can be taken as shorthand for the statement 'There is an object the same as Mars'. Indeed, statements of the latter form can be false in free logic with identity, and hence, in contrast to classical predicate logic with identity, are not logically true. An example is the false statement 'There is an object the same as (the planet) Vulcan'.

Free logics may be divided into three classes depending on the treatment of atomic (or simple) statements containing at least one singular term *t* such that the statement '*t* exists' is false. For example, the statement 'Vulcan

rotates' is an atomic statement containing a singular term, 'Vulcan', such that 'Vulcan exists' is false. Those which count all such statements as false are called 'negative', those which regard some such statements as true are called 'positive', and those which assign no truth-value at all to such statements - except perhaps statements of the form 'Mars exists'(or 'Vulcan exists') - are called 'neutral'.

In virtue of these different views about the truth-value of atomic statements with singular terms that refer to no existent object there is a significant inferential difference between the different kinds of free logic. For only in negative free logic does a version of the classical principle of particularization hold. In a negative free logic, the principle of particularization licenses the logical truth of the conditional statement 'If Vulcan rotates then there is an object which rotates' because the antecedent, 'Vulcan rotates', is an atomic statement. This does not alter the falsity of the conditional 'If Vulcan is not an object then there is an object which is not an object' because its (true) antecedent, 'Vulcan is not an object', is not atomic.

The kind of free logic one adopts has epistemological significance. For instance, when Descartes' declaration 'I think, therefore I am [exist]' is construed as an argument and formulated in a positive free logic it is invalid. For when the singular term 'I' is replaced by the singular term 'Sherlock Holmes', and the general term 'think' is replaced by the general term 'is self-identical', the resulting premise 'Sherlock Holmes is self-identical' is true in virtually any positive free logic but the resulting conclusion 'Sherlock Holmes exists' is false. On the other hand, if formulated in a negative or neutral free logic, Descartes' argument is valid because the premise is either false or without truth-value when 'I' is replaced by a singular term such as 'Sherlock Holmes', no matter what general term is substituted for 'thinks'.

2 Philosophical semantics

It is a common misunderstanding that in free logics some singular terms need not refer. There do exist, however, semantic developments of free logics which entertain disjoint (and possibly empty) universes of discourse. Usually (though not always) these are philosophically interpreted as sets of existent objects and sets of nonexistent objects, after Alexius Meinong. In such developments, a singular term *always* refers, but not always to an existent object - the singular term, 'Sherlock Holmes', for example, may be taken to refer to a nonexistent (fictional) object. On the other hand, there also exist semantic treatments of free logic with a single (and possibly empty) universe of discourse. This is virtually always interpreted philosophically as the set of existent objects, after Russell. In such developments, a singular term can fail to refer to an existent object, and hence to any thing at all. Since free logics can be based on either Meinongian or Russellian ontologies, it is only correct to say that in such logics a singular term may fail to refer to an *existent* object. In free logics, successful or failed reference is explained as follows. Let '*X*' be a place-holder for the name of a singular term, and let '' be a place-holder for the singular term named. Then a statement of the form '*X* refers (to an existent object)' is true if and only if a statement of the form ' exists' is true - hence, when the language contains identity, if and only if a statement of the form 'There exists an object the same as ' is true. For example, the statement 'The satellite of the Earth refers (to an existent object)' is true if and only if the statement 'The satellite of the Earth exists' is true (or, if the language contains identity, if and only if the statement 'There is an object the same as the satellite of the Earth' is true).

The quantifiers of free logic can be interpreted either objectually or substitutionally (see Quantifiers, substitutional and objectual); if interpreted in the latter way, then it is necessary to say in the clause for a quantifier context of the form 'There is an object such that it ' that it is true just in case a sentence of the form 'it ' becomes true when 'it' is replaced by some name '*n*' such that '*n* exists' is true. This condition is needed to give the quantifier context in question the requisite existential force. In current parlance, quantifier contexts of the form 'There is an object such that it ' in free logic are 'actualist'. Treatments seeking to preserve classical predicate logic by interpreting the quantifier context in question substitutionally while nevertheless admitting singular terms that refer to no existent object are not free logics because that quantifier context has no existential force; it is, in other words, 'non-actualist' in such treatments.

3 Implications and applications

Free logics are not committed to the philosophical doctrine that existence is a 'predicate' (see Existence) - indeed, there are free logics without identity or an existence symbol. Neither are they committed to the philosophical doctrine that proper names are truncated definite descriptions. They are not even committed to the doctrine that the

reference of most grammatically proper names is determined by the reference of some description, a view often attributed to Frege and Russell (see Proper names §1). But an important result in the proof theory of positive free logic does bear on the traditional problem of defining (and, hence, explaining away) existence. Since it can be shown that, in the otherwise classical formal idiom minus identity, statements of the form 'Mars exists' (or 'Vulcan exists') are not definable in positive free logics, then, if one chooses to couch one's logic of an existence predicate in positive free logic, that predicate cannot be eliminated save by means of the complex general term (predicate) 'is the same as something' or its equivalents.

Applications of free logic to important philosophical concerns are numerous. They range from the philosophy of mathematics to the philosophy of religion. Of special philosophical importance are the topics of definite descriptions, modality, the notion of reference, set theory and the theory of complex predicates.

In contrast to the Fregean tradition, which assigns an existent, sometimes arbitrarily, as referent to 'unfulfilled' definite descriptions such as 'the planet Vulcan', in contrast to the Russellian tradition which treats all definite descriptions as grammatical but not as genuine singular terms, and in contrast to the Hilbert-Bernays tradition which regards unfulfilled definite descriptions as formally ungrammatical (ill-formed), free theories of definite descriptions hold all definite descriptions to be genuine singular terms, but do not regard unfulfilled definite descriptions as referring to existents if they refer at all (see Descriptions). More precisely, in free theories of definite descriptions, if the noun or noun phrase following the word 'the' in a definite description is not true of exactly one existent thing, then it refers to no existent thing if it refers at all.

Free theories of definite descriptions are all based on the basic principle that

Everything is such that *it* is the same as *the* if and only if *it* and *it only* is .

This principle is known as Lambert's Law in the literature. Free theories of definite descriptions form a continuum. At one end there is a Russell-like theory which counts all atomic statements having at least one unfulfilled definite description as false, and contains scope distinctions *à la* Russell. At the other end there is a Frege-like theory which counts all atomic identity statements having only unfulfilled definite descriptions as true, and collapses all scope distinctions held to be dependent on definite descriptions. There are also a multitude of intermediate cases in the literature which, in contrast to the other traditions, is dramatic evidence of the fecundity of free logic in the logical treatment of definite descriptions.

In many, if not most, treatments of the logic of singular and general terms, supplemented by the logical modalities 'it is necessary that' and 'it is possible that', the purely quantificational fragment is free. For in such developments, the conditional

C If every (existent) thing exists, then Mars exists,

an instance of the classical principle of universal specification, is rejected on the ground that it leads, via unimpeachable modal principles, to the conditional

C^* If necessarily every (existent) thing exists, then necessarily Mars exists.

But C^* is false because though 'Necessarily every (existent) thing exists' is true, 'Necessarily Mars exists' is false. In contrast, the principle of restricted universal specification yields only the innocent and trivial

C^+ If necessarily every (existent) thing exists then necessarily Mars exists, *provided that* necessarily Mars exists.

The various treatments of free logic can influence the relationship between the meta-logical notion of reference and the logical notion of identity, the relationship reflected in the traditional adequacy condition that, for example, the singular term 'the planet causing the perturbations in the orbit of Mercury' refers to Vulcan if and only if the planet causing the perturbations in the orbit of Mercury is the same as Vulcan. To clarify the point, suppose the singular term 'the planet causing the perturbations in the orbit of Mercury' is replaced throughout the adequacy condition by the singular term 'Vulcan'. Then one obtains the special case that 'Vulcan' refers to Vulcan if and only if Vulcan is the same as Vulcan. Clearly the traditional adequacy condition is sustained in negative free logics because both the reference claim and the identity claim are false since there exists no such object as Vulcan. Similarly, if the free logic is positive, but treats all singular terms as referring, in the spirit of Meinong, then the

traditional adequacy condition is also sustained because both the identity claim and the reference claim will be true. However, if the free logic is positive and it recognizes that some singular terms do not refer, in the spirit of Russell, the traditional adequacy condition may have to be altered; in such a version of free logic, it is false that ‘Vulcan’ refers to Vulcan if Vulcan does not exist, but usually it is true that Vulcan is identical with Vulcan. In such a treatment of positive free logic the relationship between reference and identity is expressed in the following way: ‘the planet causing the perturbations in the orbit of Mercury’ refers to Vulcan if and only if Vulcan exists and the planet causing the perturbations in the orbit of Mercury is the same as Vulcan.

Free logic also exerts an influence in set theory; in free set theory, for instance, the naïve axiom of set abstraction, from which Russell deduced his famous paradox, is assertible without condition (see Paradoxes of set and property §4). Because of the replacement of universal specification by the restricted version of that principle, the most that can be deduced from the statement ‘Everything is a member of the set of sets which are not members of themselves if and only if it is not a member of itself’, an instance of the naïve principle of set abstraction, is that the set of all sets which are not members of themselves does not exist.

A final application of free logic concerns the important topic of complex general terms. In certain developments of free logic, both positive and negative, devices exist for generating complex general terms out of ‘open’ sentences, which are expressions without truth-value of the form ‘it rotates’. For instance, the open sentence in question can be turned into the complex general term ‘object such that it rotates’ by prefixing to it the general term-forming operator ‘object such that’. Then applying that complex general term to the singular term ‘Vulcan’ yields an expression capable of having a truth-value, the statement ‘Vulcan is an object such that it rotates’. In some developments this statement is equivalent to the statement ‘Vulcan rotates’ if and only if it is true that Vulcan exists. The result is a free logic in which scope distinctions with respect to the connective ‘it is not the case that’ can be made over the whole class of singular terms and not just with respect to the class of definite descriptions. For instance, in the developments in question, ‘Vulcan is an object such that it does not rotate’ may be false, although ‘It is not the case that Vulcan is an object such that it rotates’ may be true. If, however, ‘Vulcan’ is replaced by ‘Mars’, the distinction between negation in the general term and negation over the whole statement collapses because the resulting statements have the same truth-value in virtue of the existence of Mars. Indeed, the means are available in these augmented free logics to make a distinction in general between *de re* and *de dicto* predicates or properties (see De re/de dicto). In such treatments, ‘*t* exists’ can now be defined even if identity is not present - as Arthur Prior once anticipated. The requisite definition is that ‘*t* exists’ means ‘For any general term *G*, it is *not* the case that *t* is an object such that it is *G* if and only if *t* is an object such that it is *not* the case that it is *G*’. More idiomatically, the definition says that an object exists just in case denying it is *G* is tantamount to asserting it is *non-G*, and vice versa, for any *G*.

See also: Free logics; Logical and mathematical terms, glossary of

KAREL LAMBERT

References and further reading

Bencivenga, E. (1989) ‘Why Free Logic?’ in *Looser Ends*, Minneapolis, MN: University of Minnesota Press. (A very readable and distinctive philosophical explanation of free logic.)

Evans, G. (1979) ‘Reference and Contingency’, *The Monist* 62: 161-89. (An argument that Kripke’s doctrine of contingent a priori statements presupposes free logic.)

Lambert, K. (1983) *Meinong and the Principle of Independence*, Cambridge: Cambridge University Press, esp. ch. 5. (An informal account of the varieties of free logic and its connection with issues in the theory of predication.)

Leonard, H. (1956) ‘The Logic of Existence’, *Philosophical Studies* 7: 49-64. (The essay which initiated the development of free logics.)

Freedom and liberty

There are at least two basic ideas in the conceptual complex we call 'freedom'; namely, rightful self-government (autonomy), and the overall ability to do, choose or achieve things, which can be called 'optionality' and defined as the possession of open options. To be autonomous is to be free in the sense of 'self-governing' and 'independent', in a manner analogous to that in which sovereign nation states are free. Optionality is when a person has an open option in respect to some possible action, x*, when nothing in the objective circumstances prevents them from doing* x *should they choose to do so, and nothing requires them to do* x *should they choose not to. One has freedom of action when one can do what one wills, but in order to have the full benefit of optionality, it must be supplemented by freedom of choice (free will), which consists in being able to will what one wants to will, free of internal psychological impediments. Autonomy and optionality can vary independently of one another. A great deal of one can coexist with very little of the other.*

Perhaps the most controversial philosophical question about the analysis of freedom concerns its relation to wants or desires. Some philosophers maintain that only the actual wants that a person has at a given time are relevant to their freedom, and that a person is free to the extent that they can do what they want, even if they can do very little else. Other philosophers, urging that the function of freedom is to provide 'breathing space', insist that freedom is a function of a person's ability to satisfy possible (hypothetical) as well as actual wants. A third group consists of those who hold a 'value-oriented' theory according to which freedom is not merely the power of doing what one wants or may come to want, rather it is the capacity of doing something 'worth doing or enjoying', something that is important or significant to the person said to be free, or to others.

1 Freedom and liberty

These two terms are often used interchangeably, but on those occasions when they are not taken to be synonyms, the basis of the distinction between the two is usually clear. 'Freedom', when applied to persons and their actions, refers to the *ability* of a person in a given set of circumstances to act in some particular way. 'Liberty' refers to authoritative *permission* to act in some particular way. The contrast is a basis for the grammatical distinction between 'can' and 'may', between the *de facto* and the *de jure* perspectives, or between (overall) ability and permission.

The concept of a liberty is an important part of juridical systems. A set of governing rules can impose *duties* on those who are subject to their authority. But when the rules remain silent about a given type of activity, *x*, then they are said to leave the subjects *at liberty* to *x* or not to *x*, or however they see fit. To be at liberty to *x* is simply to have no duty not to *x* (see Rights §2).

So conceived, freedom and liberty can vary independently. For example, when a statute is only sporadically enforced, if at all, it may leave a person's *de facto* ability to do what it prohibits virtually unimpaired. Thus a cohabiting unmarried couple in Arizona is perfectly capable of doing what the law prohibits only because the law in question is hardly ever enforced. Cohabiting couples have almost perfect *de facto* freedom to cohabit because police indifference makes the risk of detection and conviction minimal. Not only are we sometimes free (to some degree or other) to do what we are not at liberty to do; conversely, we may be at liberty to do what we are not free to do, as when circumstances other than enforced rules prevent us from doing what we are legally permitted to do.

2 Freedom as autonomy

Judgments of freedom are made not only about individual persons but also about their political communities. Many nation states at some point in their histories have had occasion to declare their independence or, what amounts to the same thing, their sovereignty or status as free states. At the same time, other countries may have lost their freedom (independence) and become mere colonies of stronger nation states. When a political state becomes unfree in this way, each citizen can think of themselves as also deprived of the same kind of freedom, or something analogous to it, for few of them remain entirely self-governed when their own country is governed from afar by masters who simply impose their directives by force. 'None of us are truly free so long as our nation is not', a local patriot might say, even while conceding that the colonial power that governs them has treated them decently, allowing them many freedoms. The question of who rightly governs them, as Sir Isaiah Berlin puts it, is

logically distinct from the question of how their governors - foreign or domestic, legitimate or not - protect their liberties (Berlin 1958). The first question concerns autonomy; the second concerns optionality.

Autonomy and optionality also can vary independently. National autonomy can produce its own tyranny, as when a newly independent nation, free of its colonial repressors, refuses to govern democratically and recognizes no civil rights in its subjects. In such a case the native populace may feel as badly mistreated, if not worse, by their own government as before by the colonizers.

Autonomy or self-government can consist of sharing with one's fellow citizens political independence, or can consist of self-direction by that element of the self authorized by nature to rule (often identified with reason). To be self-governed requires that we not be governed by illegitimate outsiders and equally not by alien forces from within.

Note the close analogy to slavery. One slave owner, *A*, is very severe with his only slave, *S1*, permitting him only minimal free movement, no choice in deciding what his off-duty conduct shall be, or what he shall read, how he dresses and so on. Another slave-owner, *B*, is very easy-going. He treats his only slave, *S2*, as if he were a valued friend, and allows him to do anything short of harming others or leaving the plantation. It is clearly understood that *S2* may do all these things only because *B* permits him to, not as a matter of right. The only rights in this situation are *B*'s property rights. The rules of property ownership permit *B* to be as tough with *S2* as he wishes, but he prefers to be kind. So there is in *S2*'s situation a predominance of heteronomy (government by others) conjoined with high optionality (*de facto* freedom). It seems clear then that one can have little or no autonomy, and yet live a contented life with a high standard of living, something resembling friendship and respect, and most important - options left open for one's own choice to exercise, though not as a matter of right.

Suppose that in respect to a certain choice there are two possibilities left open for *S2*. His master can restore his autonomy and turn him loose into an unfriendly world where other human beings, even though they lack authority over him, treat him badly, effectively closing many key options that would have been left open for him by his earlier beneficent master. Should he accept this offer at the expense of much *de facto* freedom? If the question is a difficult one, it shows that it is not clear to which of the two contending values, independence or optionality, he attaches the greater importance (see Autonomy, ethical).

3 Negative and positive freedom

Philosophical advocates of 'positive freedom' are often reacting to a tradition among English empiricists that extends from Hobbes to J.S. Mill and Russell. Hobbes intended his definition to apply to the most essential of the common elements in free action and free movement generally. He defined 'free' as the absence of external impediments, to apply equally well, for example, to free-flowing (undammed) streams of water as to the purposeful conduct of human beings. Subsequent empiricists, including Locke (§10) and Hume (§3), also held that all freedom is essentially something negative; namely, the absence of restraint or impediment to our actions.

The family of theories to which Berlin attached the label 'positive liberty' are those that identify freedom (or liberty) with personal autonomy or self-government. One version of that theory emphasizes the internal forum that is the agent's self and the legitimate claim of the rational self to rule over the self's lesser elements. The second of the autonomy theories is less individualistic and more political. It holds that no individual can live their life autonomously except as a member of a free political community, a state that is not only independent of other states, but one that is itself organized democratically so that all citizens can share in its governance, and in that sense, at least, be self-governing.

The nineteenth-century idealist philosopher T.H. Green (§3) summed up these requirements in his definition of 'freedom' as a 'positive power or capacity of doing or enjoying something worth doing or enjoying, and that too, something that we do or enjoy in common with others' (1888: 371). Green's definition also expresses a second conception of positive freedom as more than the mere absence of impediment to our desires, even more than the absence of impediment to our 'worthy' desires. In addition to the absence of constraint, genuine freedom must provide full opportunity beyond the mere non-interference of the police and other people. If a person desires above all things to own and enjoy a Mercedes, and there are no external impediments, legal or nonlegal, in the community to such ownership, then both the legal code and the neighbours leave them free to do as they desire. But if that person has no money, then that negative freedom is effectively useless. To have true freedom, say

supporters of positive freedom, one must have what is required for the satisfaction of worthwhile wants, and that will usually include at least minimal wealth, physical health, talent and knowledge, including the sorts of knowledge normally imparted by formal education. The more we are able to do the things worth doing, they insist, the freer we are.

The positive freedom theorists may go on to charge that the negative theorist cannot explain why the laudatory title of freedom should ever characterize the person who is paralyzed, insane, infantile, impoverished and ignorant, as generally free. It must be ironic, they claim, to say that such an unfortunate person is well off in the manner implied by the term ‘free’. Negative freedom theorists argue that to be free does not mean to be well off; one may be free but discontented, unhappy, ignorant, hungry or in pain. An individual may have freedom but find that in their circumstances it is not worth much. What this shows is that freedom is the kind of good whose worth fluctuates, or in the words of John Rawls (§§1-2): ‘the worth of liberty is not the same for everyone’ (1971: 204).

The pauper is *unable* to buy the Mercedes, but according to the negative freedom advocate, this is not through being *unfree* to buy one. Most writers within the negative freedom tradition deny that all inabilities are also ‘unfreedoms’. The inabilities that constitute unfreedoms, they insist, are those that can be traced directly or indirectly to the deliberate actions or policies of other human beings, in particular legislators and police offers, who can intervene directly and forcefully in other persons’ lives. Sometimes the relevant explanation of some other person’s incapacity (for example, to earn a decent living) can be linked indirectly to various social influences. The impoverished person might be so because of a lack of technical skill, and that lack, in turn, could be a product of a poor education traceable, however obscurely, to the inequities of a national system of racial segregation, which in turn was supported as deliberate policy by an apartheid government. In that case we could say not only that they are unable to do *x*, but also that they are unfree, given the circumstances, to do so.

The gap between the positive and negative theorists can be further decreased by a theory which has a wider conception of ‘restraint’ and ‘impediment.’ Such a theory would have a place both for negative constraints like lack of money, and internal constraints like intense headaches. Such a theory could support negative freedom (that all freedom consists of the absence of impediments, whether positive, negative, internal or external), yet also encompass the important point made by positive freedom (that there is a lot more to freedom than simple non-interference from police officers and other persons, important as that is).

A large number of philosophers now reject the view that there are two irreducibly distinct concepts of freedom, one positive and the other negative (MacCallum 1967; Feinberg 1973; Rawls 1971). These ‘single concept’ theorists do not contend that one of the pair of allegedly distinct concepts is ‘the only, the "truest", or the "most worthwhile"’ (MacCallum 1967: 312), but rather that it is a mistake to make the distinction between positive and negative concepts in the first place. According to MacCallum, there is only one concept of liberty and that is best understood as ‘always one and the same triadic relation’ between a person (subject or agent), an intended action (actual or possible) and what MacCallum calls a ‘preventing condition’ (barriers, compulsions and constraints) (1967: 312, 314). Freedom, in his view, is always *of* someone, *from* something, *to* do, have or be something. Disputes about the nature of freedom like that which divided adherents of positive freedom from adherents of negative freedom are, according to MacCallum, really disagreements over the proper range of one or more of the three variables in a single analytic model: what the term ‘person’ is to stand for, what is to count as an obstacle, or impediment, or forceful interference, and what is to count as a wanted or intended action.

4 Freedom as optionality

Among the controversies that still divide writers about freedom is the question of whether freedom (in the sense of optionality alone) should be conceived of as simply the absence of present frustration or whether it is best understood as the absence of wider opportunities to do more than one wants to do now. We can call the former concept the ‘actual-want satisfaction’ theory and the latter the ‘hypothetical-want’ or ‘dispositional’ theory. The former allows a person to be called free to the extent that they can satisfy their present wants, without hindrance or frustration. The dispositional concept, however, will not consider them free unless they can also do things that they do not want to do at that present moment, but could, for all they now know, come to want to do at some future time. Why, one might ask, would added dispositional freedom be of value to the person whose actual wants are always permitted their satisfaction? Why should a person miss merely hypothetical want-satisfaction when they can do everything they want without frustration? The usual answer to this question points out that the love of

freedom can be a love of breathing space, or room to manoeuvre, of frequent opportunities to change one's mind. The hypothetical account of optionality, therefore, can also be called the 'breathing space' theory.

The most influential spokesmen for the 'actual-wants' concept were the ancient Stoics (see Stoicism §17). According to Stoic teaching, there are two ways a person can increase the degree of their want-satisfaction. One is to leave their wants as they are and work for the means to satisfy them. The second is to avoid trying to change the world - that is the path to misery - but instead to develop the techniques for changing desires so that they always accord with what happens. The Stoic does not need any breathing space. Whatever happens will please them because their only desire is that God's will be done, and since Zeus is believed to be omnipotent, that desire cannot be frustrated.

Consider Dorothy Doe, who can choose among 1,000 things at time *t*, but is prevented from choosing, or actually doing, the one thing she wants most to do. Richard Roe, on the other hand, can only do one thing at time *t*, but it happens to be the one thing he wants most to do. Richard Roe, one should say, is not simply 'comparatively' or 'largely' unfree; rather, he is totally unfree, for to say that he can do only one thing is to say that he is forced or made to do that thing. And it is beyond controversy that one cannot be both free and compelled to do the same thing. It may be the case, however, that Richard Roe will be entirely content with the arrangement, and actually welcome the compulsion, which if so, shows again that one can sometimes find more contentment in being unfree than in being free. This at least is the message derived from the example by the hypothetical-wants theorist. The actual-wants theorist will deny that Roe is truly unfree; after all, Roe *can* do what he most wants to do.

Dorothy Doe, on the other hand, may have just lost her last chance to pursue a career as a medical researcher, or her last chance to marry the man she loves, or to find a cure for her child's disease. But she does 'enjoy' thousands more options than Richard Roe. She might seem to be freer in respect (say) to prospective marriage partners (one hundred philosophers are eager to marry her tomorrow, but she loathes them all). But her greater freedom is of no significant use. She will be unhappier but more free than Roe. This example may please the breathing space theorist more than the actual-wants theorist, but it may give still more support to the value-oriented philosopher of freedom considered next.

Is Dorothy Doe really more free simply by having more open options? Does not the superior desirability, in her judgment, of some of the options count as well as the sheer number of them? The proponent of the hypothetical-wants theory of freedom often baulks at permitting desirability into our determinations, partly because of the danger that philosophers will reduce the issue to a purely normative question to be settled by considering which definition of the word 'free' links its meaning to something that is worthy of the value we associate with the word.

The problem of counting options remains a difficulty for all the above theories. Berlin had earlier written that 'the method for counting [possibilities] can never be more than impressionistic. Possibilities of action are not discrete entities like apples, which can be exhaustively enumerated' (1958: 130, footnote 1) But individuating possible actions is not the only problem for the philosopher who would apply quantitative measures to freedom's many dimensions, including comprehensiveness, fecundity, and diversity. And one does not exhaust the relevant possibilities by dividing all action into 'possible' and 'impossible.' There are also the component categories - difficult and easy, possible at great cost and possible at small cost, statistically probable and statistically improbable, multiple choices and either/or choices.

It is hard enough to deal with these problems of measurement, but they are almost equally difficult to evade, especially if we continue to speak of one individual or one society having more freedom than another. Moreover, if a philosopher maintains that both the number and the significance, importance or value of open options determines how free one is (and probably most philosophers take such a combination view), then the difficulties begin all over again when they leave off option-counting and begins option-evaluating.

See also: Coercion; Free will; Freedom of speech; Law, limits of; Liberalism

JOEL FEINBERG

References and further reading

Arneson, R.J. (1985) 'Freedom and Desire', *Canadian Journal of Philosophy* 15 (3): 425-48. (Considers the

question whether amounts of freedom can be measured, and if so, what role the free person's desires play in the process.)

Benn, S.I. and Weinstein, W.L. (1971) 'Being Free to Act and Being a Free Man', *Mind* 80 (3): 194-211.(Argues that the concept of freedom presupposes a concept of autonomy; that is, the concept of the free person as chooser.)

Berlin, I. (1958) 'Two Concepts of Liberty', in *Four Essays on Liberty*, Oxford: Oxford University Press, 1969; repr. in D. Miller (ed.) *Liberty* Oxford: Oxford University Press, 1991.(Classic source of the distinction between positive and negative liberty.)

Dworkin, G. (1982) 'Is More Choice Better than Less?', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Midwest Studies in Philosophy, VII*, Minneapolis, MN: Minnesota University Press, 47-61.(Considers the question of whether more choices are always preferable to fewer for a rational individual, concluding that more is not always better.)

Feinberg, J. (1973) *Social Philosophy*, Englewood Cliffs, NJ: Prentice Hall.(This introduction to social and philosophy concentrates on the concept and limits of freedom, the distinction between legal and moral rights, and the concept of social justice.)

Feinberg, J. (1980) 'The Interest in Liberty on the Scales', and 'The Idea of a Free Man', in *Rights, Justice, and the Bounds of Liberty*, Princeton, NJ: Princeton University Press.(The former deals with the problem of weighing the interest in liberty against other interests in the effort to minimalize social harms; the latter probes for conceptual linkages between the idea of a free person and of a free society by considering how the word 'free' has come to apply to both.)

Gray, J.N. (1980) 'On Negative and Positive Liberty', *Political Studies* 28 (4): 507-26.(Important critique of the theory of Isaiah Berlin.)

Green, T.H. (1888) 'Liberal Legislation and Freedom of Contract', in *Works of Thomas Hill Green*, vol. 3, ed. R.L. Nettleship, London: Longmans Green; abridged by D. Miller (ed.) *Liberty*, Oxford: Oxford University Press, 1991.(The author defines freedom as 'a positive power or capacity of doing or enjoying something worth doing or enjoying, and that too, something that we do or enjoy in common with others'.)

MacCallum, G.C. (1967) 'Negative and Positive Freedom', *Philosophical Review* 76 (3): 312-34; repr. in D. Miller (ed.) *Liberty*, Oxford: Oxford University Press, 1991.(Argues that there is only one concept of liberty; namely, that expressed in the schema *x* (the actor) is free from *y* (constraining condition) to do or become *z*.)

Miller, D. (ed.) (1991) *Liberty*, Oxford: Oxford University Press.(The most useful bibliography of philosophical analyses of freedom. Contains, among other things, influential articles by T.H. Green, Isaiah Berlin, Gerald C. MacCallum and G.A. Cohen.)

Pelczynski, Z.A. and Gray, J.N. (eds) (1984) *Conceptions of Liberty in Political Philosophy*, London: Athlone Press.(Contemporary authors discuss the differing conceptions of freedom that appear throughout the history of philosophy, from the ancient Greeks to John Rawls. A major theme is the comparison of the various views with Berlin's distinction between positive and negative freedom.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, 201-10.(Claims that the usual debates about liberty are not properly conceived of as disputes about the proper definition of 'liberty'. Instead, the concern is with the relative values of several distinct liberties, which can conflict.)

Swanton, C. (1992) *Freedom: A Coherence Theory*, Cambridge: Hackett Publishing Company. (Argues that a properly conceived coherence model can provide a theory of freedom that reconciles a wide variety of views on freedom while retaining the strengths of each.)

Freedom of speech

Freedom of speech is one of the most widely accepted principles of modern political and social life. The three arguments most commonly offered in its defence are that it is essential for the pursuit of truth, that it is a fundamental constituent of democracy, and that it is a liberty crucial to human dignity and wellbeing. Its advocates also plead the dangers of allowing governments to control what may be said or heard. Yet there is also general agreement that speech should be subject to some limits. Most contemporary controversies about free speech concern those limits; some focus upon what should count as 'speech', others upon the harms that speech may cause.

1 Definitions

In discussions of free speech, 'speech' is typically given a meaning that is both more and less inclusive than in ordinary usage. It includes written as well as spoken words and it may include nonverbal forms of communication such as pictures, symbols and gestures. That is why the phrase 'freedom of expression' is sometimes preferred to 'freedom of speech'. On the other hand, certain forms of speech fall outside this privileged domain; despite involving speech, perjury and blackmail, for example, receive no privileged treatment.

The 'freedom' that is most commonly of concern is legal freedom - the absence of legal restraints upon speech. But other sorts of freedom may also be at issue, for example freedom from inhibiting social pressures or economic threats, such as loss of trade or employment. Questions of free speech can also arise in matters such as the editorial policy of an academic journal or the purchasing policy of a library. The reality of people's opportunities to air their views through the media is another important matter for free speech, particularly in contexts of economic inequality or government-controlled media and in relation to competitive speech such as political argument (see Freedom and liberty).

Freedom of speech is typically asserted as a 'right', which therefore has a privileged moral and political status, and it has become a standard ingredient of constitutional bills of rights and declarations of human rights. The right to unfettered speech is most commonly claimed for those who wish to *speak*, but it may also be claimed for those who wish to *hear* and the right to *hear* what others wish to say may be no less important.

Analysts sometimes try to give content to freedom of speech as a right by dissecting the meanings of 'speech' and 'freedom', but philosophically it is impossible to divorce the proper content of the right from its justification. Only when we have considered why freedom of speech matters are we in a position to identify what sort of speech the right should encompass, what sort of freedom it should imply, and what sort of status that freedom should enjoy.

2 Justifications

The pursuit of truth. Freedom of speech has been commonly defended as an instrument of truth. That was the claim central to Milton's *Areopagitica* (1644) and to Mill's *On Liberty* (1859) - the two most celebrated defences of free speech in the English language (see Mill, J.S. §12). If there is to be progress in knowledge and understanding, people must be free to present, criticize and discuss ideas and information. Only in such a free and open exchange will the truth have maximum opportunity to emerge. Once the truth has emerged, we might suppose that free speech has done its job and can be safely abandoned. But the fallibility of our beliefs has always figured strongly in this defence: we may believe that the opinion we suppress is false, but history shows how often truth has been mistaken for error and error for truth. Moreover, as Mill argued, even if an opinion is true, we can be confident of its truth only so long as it can withstand challenge, and even a true belief is liable to become a 'dead dogma' rather than a genuine conviction if immured from challenge.

Powerful though this defence is, it encounters a number of objections. First, it makes free speech valuable only in so far as truth is valuable. In general, we are unlikely to deny that truth is better than error and knowledge better than ignorance, but on occasion truth and knowledge may conflict with other goods such as public security and peace of mind. Some doubt whether truth should always be the overriding value.

Second, the argument supposes that free speech will actually foster truth and understanding. Sceptics find such a claim unduly optimistic and doubt the ability of ordinary people to distinguish truth from error and irrational

prejudice from sober reasoning. In assessing that objection, however, we must be alive to the alternatives. The relevant question is not whether truth will always emerge from the marketplace of ideas but whether it will be better served by open and unfettered discussion than by government censorship.

Third, some forms of expression, such as some types of art, are not concerned with truth and cannot therefore be defended as its instruments. On the other hand, in those cases there are often other goods, such as creativity, which can take the place of truth in a similarly structured consequentialist defence of free expression.

Democracy. The fundamental idea of democracy is that of a people governing itself. This cannot be done if the members of a people are unable to present their views to one another. Democratic decision making requires discussion and debate as well as voting, and a significant limit upon its deliberative process will significantly limit its democratic character. Moreover, democracy implies that no one is authorized to limit a people’s freedom of discussion, except perhaps the people itself. Democracy in ideal form also entails that each person who is subject to its decisions should be entitled to participate equally in its processes, and hence that each should enjoy equal rights of free expression (see Democracy).

Given the indirect character of modern democracy, freedom of speech is also important politically for ensuring that rulers are aware of the wishes and opinions of the ruled, for enabling the conduct of opposition and for preventing abuses of power. Indeed, as a check upon the abuse of power, the case for free speech extends beyond democratic forms of government.

Democracy is sometimes thought to afford an unsafe haven for free speech. Suppose a majority wishes to silence a minority; does democracy not then sanction the removal of that minority’s freedom of speech? The answer depends in part on what sort of speech is at issue. Since democracy requires that all members of a demos should enjoy equal political rights, democracy itself cannot justify a majority’s removing or reducing the democratic rights - including democratic rights of free expression - of some members of the demos. However, the speech that can be defended by appealing to democracy is limited to that which is essential to participation in a democratic process. That does not encompass every form of speech. For example, if we wish to defend the right of scientists to challenge the orthodoxies of their peers, the prerequisites of democratic participation do not provide the most convincing foundations for that right - although it is not possible to state in a simple a priori fashion what sorts of speech will be relevant to a democratic process and what will not.

Individual liberty. Free speech is often valued as a constituent of individual liberty rather than only as an instrument of social purposes or as a specifically political institution. Sometimes what is emphasized is the moral standing of persons. To prevent people from communicating their beliefs and opinions is to violate the respect and standing to which they are entitled as persons. Equally, not allowing them to hear the views of others and to reach their own conclusions is inconsistent with their moral status as persons. Sometimes the emphasis shifts from the ’rightness’ to the ’goodness’ of free speech. The ability to communicate freely with one’s fellows, it is claimed, is essential to human development and human fulfilment. Free speech is a crucial feature of a society whose institutions and ethos enable individuals to shape their lives, to exercise meaningful choices and to achieve wellbeing in its fullest human sense.

3 Limits

Nowadays, controversy over free speech centres mainly upon its limits. Debates about those limits focus on two matters: the scope of speech and the harms that might be wrought by speech.

What some seek to protect as speech, others deny is speech in any credible sense. For example, attempts to incorporate pornographic displays within the free speech principle have met with the objection that those displays have no communicative content and are not speech in any reasonable sense (see Pornography). In addition, some of what would ordinarily be termed speech may not qualify as speech of a privileged kind. Advertisers, for instance, are engaged in intentional communication but arguably justifications of free speech provide no case for including commercial advertising in the category of privileged speech. Equally, there are acts, such as flag-burning or wearing a symbolic armband, which would not ordinarily be described as speech but which may be encompassed by the free speech principle because of their expressive or communicative character. In some measure, then, the limits of free speech are set by what its justifications imply should count as speech.

Second, speech is often limited because it proves harmful. It is sometimes held that speech, because it is speech and not action, cannot harm; but that is clearly not so. Defamation damages a person's reputation and that harm is usually accepted as reason enough for proscribing libellous speech, provided that what is said is neither true nor 'fair comment'. Speech which invades privacy is also a candidate for proscription where no genuine public interest is at stake; here even 'truth' is no defence - there are some things that a public has no right to know. Other types of speech that are often outlawed because of their alleged harmfulness are obscenity and pornography, incitements to violence and other sorts of illegal conduct, and speech which incites hatred, particularly racial hatred.

Four types of controversy surround claims about the harmfulness of speech. One is the essentially empirical question of whether a particular form of speech actually causes the harm it is alleged to cause. Many people, for example, believe that pornographic material encourages sex crimes, while others claim it has no such effect and may even provide a catharsis which reduces criminality. A second issue is whether an alleged harm is really a harm. If a speaker incites others to overthrow a political system or to defy a law, is that harmful? Clearly, the answer will depend, in part, upon our estimate of the political system or the law at stake (see Law, limits of §4).

A third matter is whether an acknowledged harmful effect is sufficiently harmful to justify restricting speech. Should we, for instance, outlaw blasphemous statements so that people with religious convictions are spared offence? Given the weight of the free speech principle, a harm will have to be significant and substantial to justify curtailing speech. On the other hand, in making that assessment it is reasonable also to take into account the nature of the speech involved, even when it is speech whose subject falls within the privileged realm. For example, the case against prohibiting offensive religious statements will be less strong if those statements are merely scurrilous and aim only to cause distress than if they form part of a serious disquisition.

In most cases in which we restrict harmful speech, we are delimiting the right of free speech. For example, in accepting laws of libel we are accepting that people have no right to engage in defamatory speech. But there is a fourth area of debate concerning harm that is about overriding rather than defining the right of free speech. Suppose someone addresses a hostile audience; they say nothing they are not entitled to say but the audience starts to react violently; the authorities, anxious to prevent a riot and to protect life and limb, step in and silence the speaker. In these circumstances, we might accept that, all things considered, the authorities rightly silenced the speaker even though the speaker had committed no wrong. In such cases, we are not setting the boundaries of the right of free speech; rather we are accepting that there are circumstances in which the right may be justifiably overridden.

Not every wrong that is cited as a reason for limiting speech is readily described as a 'harm'. 'Indecency' and 'obscenity', for example, are sometimes condemned for reasons other than the injury they do to others. Generally, there is a reluctance nowadays to accept that statements should be prohibited merely because they are false. After all, fallibility and uncertainty about truth figure prominently in the case for free speech. One current and controversial exception is the Holocaust, denial of which is now outlawed in some countries. Even that measure, however, is often justified as a way of combating racism and of preventing affronts to victims of the Holocaust, rather than as a simple endeavour to uphold truth by law.

Not all rules which govern speech should be conceived as restricting it. For example, orderly discussion of issues in a legislature requires rules of debate and, taken in the round, those rules regulate rather than restrict speech. Similarly, there is a difference (although one that is often controversial) between regulating and restricting street demonstrations.

In all cases in which speech is limited, determining where precisely the boundary should fall can prove difficult. Legislators and judges need precise distinctions between the tolerable and intolerable, yet general philosophical arguments often fail to supply those sharp distinctions. In addition, debates about the moral limits of free speech need not be wholly identical with debates about its proper legal limits. Because there is reason to fear political power and the clumsiness of law, the legal right of free speech may be cast more generously than the moral right. Thus, morally, there may be things that we ought not to say, even though, legally, we are and ought to be free to say them.

PETER JONES

References and further reading

Baker, C.E. (1989) *Human Liberty and Freedom of Speech*, Oxford: Oxford University Press.(A comprehensive study which integrates philosophical and legal issues.)

Fish, S. (1994) *There is No Such Thing as Free Speech, and It's a Good Thing Too*, Oxford: Oxford University Press.(A sceptical appraisal of enthusiasm for free speech.)

MacKinnon, C.A. (1994) *Only Words*, London: Harper Collins.(An impassioned attack upon attempts to defend pornography and racism through free speech.)

Meiklejohn, A. (1965) *Political Freedom*, New York: Oxford University Press.(A democratic defence of free speech.)

Mill, J.S. (1859) *On Liberty*, London: Dent, 1972.(Probably the most influential defence of free speech ever written.)

Milton, J. (1644) Areopagitica, in K. Barton (ed.) *Prose Writings of Milton*, London: Dent, 1970.(A classic early defence of free speech.)

Scanlon, T. (1972) 'A Theory of Freedom of Expression', *Philosophy and Public Affairs* 1 (2): 204-26. (A defence of free expression grounded in autonomy and rationality.)

Schauer, F. (1982) *Free Speech: A Philosophical Enquiry*, Cambridge: Cambridge University Press.(A comprehensive philosophical study.)

Freedom, divine

In the theistic tradition, many thinkers have held that God is infinitely powerful, all-knowing, perfectly good and perfectly free. But since a perfectly good being would invariably follow the best course of action, what can be meant when it is said that God acts freely? Two different views of divine freedom have emerged. According to the first view, God acts freely provided nothing outside him determines him to act. So when we consider God's action of creating a world, it is clear that on the first view he acts freely since there is nothing outside him to determine him to do as he does. The difficulty with this view is that it neglects the possibility that God's own nature might require him to create one particular world rather than another or none at all. According to the second view, God is free in an action provided it was within his power not to perform that action. Unlike the first view, on this view God acts freely only if nothing beyond God's control necessitates his performing that action. The problem for this view is that since it is impossible for God, being perfectly good, not to choose to follow the best course of action, it is difficult to see how God could be free in such an action.

1 Two views of God's freedom

In traditional theism, God is understood to be an omnipotent, omniscient, perfectly good creator of all things other than himself. But what of divine freedom? Is the God of traditional theism also free with respect to creating things other than himself? If God were not free not to create what he does create then it seems we could make no sense of thanking him or praising him for creating what he did. He would not merit our gratitude and praise for the simple reason that he would create of necessity and not freely. So in addition to his other attributes, traditional theism has insisted on ascribing freedom to God (see God, concepts of §6).

How are we to understand divine freedom? Two different views have emerged. According to the first view, God is free in creating a world or in acting within the world he has created provided nothing outside him determines him to create the world he creates or determines him to act in a particular way in the world he has created. According to the second view, God is free in creating or acting within his creation provided it was in his power not to create what he did or not to act within his creation as he did.

How are these two views of divine freedom connected? If it was in God's power not to create the world he did create, then there was nothing outside God that absolutely determined him to create the world he did create. So God's having the power not to create what he did create implies the absence of something outside him that absolutely determined him to create what he did. But the reverse is not true. The absence of something outside a being that determines that being to act in a certain way does not imply that the being has the power not to act in that way. For something within that being may absolutely determine it to act in a certain way. And if something within a being necessitates that it act in a certain way, then, unless the presence or absence of that necessitating factor is within the being's control, that being lacks the power not to act in that way.

2 The first view and its chief difficulty

The first of these two views has the advantage of establishing beyond question that God possesses freedom from external forces with respect to his selection of a world to create. For given that he is omnipotent and the creator of all things other than himself, it is evident that nothing outside him determines him to create whatever he does create. And given that whatever he creates is within his control, it would seem that he is completely at liberty to act as he sees fit within the world he has created. So the fact that nothing outside God determines him to create or act as he does clearly shows that God is an autonomous agent; he is self-determining in the sense that his actions are the result of decisions determined only by his own nature. But is this sufficient to establish that God is genuinely free? We believe that a human being need not be free in performing a certain action even when it is clear that they were not determined to perform that action by external forces. Perhaps the person was in the grip of some internal passion or irresistible impulse that necessitated the performance of that action, overcoming their judgment that the action is wrong or unwise. With respect to human beings, the defender of the first view of divine freedom can agree that the mere absence of determining external agents or forces is not sufficient for an individual's action to be free. But in the case of God, as opposed to humans, the defender can argue that it is sufficient. For in God there is no possibility of his passions overcoming the judgment of reason. As Leibniz remarks:

> The Stoics said that only the wise man is free; and one's mind is indeed not free when it is possessed by a great passion, for then one cannot will as one should, i.e. with proper deliberation. It is in that way that God alone is perfectly free, and the created minds are free only in proportion as they are above passion.
>
> (1704: 175)

The chief objection to this view of divine freedom is that it does not sufficiently recognize the importance of agents having control over their free acts. An action that an agent performs freely is an action the agent was free to perform and free not to perform; it was up to the agent whether to perform or not perform that act. If some external force or internal passion was beyond the control of the agent, and the agent's action was inevitable given that external force or internal passion, then the agent did not act freely in performing that action. Since God is a purely rational being and not subject to uncontrollable passions, which sometimes compel human agents to act, it is tempting to conclude that God enjoys perfect freedom of action. But this will be so only if there are no other features in God that both necessitate his actions and are not within his control. Because human beings are generally thought to have the power to act against the counsel of reason, we regard their acts due to reason - as opposed to their acts due to irresistible impulses - as acts they perform freely. For we believe they were free to reject the counsel of reason and act otherwise. But what if God cannot reject the counsel of his reason as to what action to perform? A human being who is morally good and rational may yet have the power to refrain from acting as his goodness and reason direct. But can this be true of God? And if it cannot be, how can we then say that God acts freely? Thus, it has seemed to many philosophers and theologians that the first view of divine freedom fails to capture an essential element in the idea of freedom, whether human or divine.

3 The second view and its chief difficulty

The second view of divine freedom embraces the idea that agents, whether divine or human, act freely only if they have the power to refrain from so acting. Although the idea of 'power to refrain from so acting' is not particularly clear, the proponents of the second view are united in rejecting any conditional account such as 'power to refrain from so acting if the agent wills to refrain', for such an account does not require that it was in the agent's power to will to refrain. So, according to the second view, God was free in creating the world he did create only if it was unconditionally in his power either to create some other world instead or to refrain from creating any world at all.

The chief difficulty facing this second view of divine freedom arises when we consider some of God's other attributes and ask whether they leave God with any significant degree of freedom. It has already been noted that the theistic God is a being of infinite power and perfect goodness. These perfections are generally regarded as *essential* to God. They constitute part of his very nature. That being so, we can ask whether God is ever free to do an evil (morally wrong) act. The answer seems to be no. Of course, being perfectly good, omnipotent and omniscient, God will never in fact do an evil act. No being who knowingly and willingly performs an evil act is perfectly good. Since being free to do an evil act is consistent with never in fact doing an evil act, it may seem that God could be free to perform such an act. But if God is free to perform an evil act, then he has it in his power to perform that act. And if God has it in his power to perform an evil act, then he has it in his power to deprive himself of one of his essential attributes (perfect goodness), to change his nature. But no being has the power to deprive itself of one of its essential attributes. Therefore, it would seem that God does not have it in his power to perform an evil act (see Goodness, perfect §2).

The fact that God cannot do evil does not severely limit God's freedom so long as there remains a sufficient range of actions that are open to God. For example, it has long been held that God's action in creating the world was a free action, that God was free to create a world and free not to create a world. Moreover, it has been held that among worlds God can create, God is free to select the one he will create. Of course, since he cannot do what is evil he cannot create a bad world. But presumably there are many good worlds from which he is free to select the one he wishes to create. Even here, however, there are serious problems in reconciling God's freedom with his omnipotence, omniscience and perfect goodness.

Suppose among creatable worlds there is one that is better than all others. Is God free to create some world other than this best world? If God creates some world that is less than the best world he can create, it would seem to be possible for there to be a being better than God. For it is the nature of perfect goodness to create the best it can.

Therefore, since God is perfectly good, a being whose goodness is unsurpassable, it follows that if there is a unique best world among worlds he can create, it is impossible for him to create a world inferior to it. For then he would not be a being whose goodness is unsurpassable (see Quinn 1982).

What if there is no best world among creatable worlds? This could be true in two ways. First, it might be that for any creatable world there is a better world God could create. Second, it might be that there is no unique best world. Perhaps instead there are many equally best worlds among the worlds God can create. Let us consider these two cases in turn.

On the assumption that for any world God might create there is a better world he could create instead, it is clear that it is impossible for God to do the best that he can. Whatever he does, he could have done better. This being so, it seems only reasonable that God's perfect goodness would be fully satisfied should he create a very good world. Presumably there are a large number of such worlds from which he can choose one to create. So long as he creates one of these worlds it seems he will have satisfied the demands of his perfect nature. For the idea that he should create the best world he can is an idea that logically cannot be realized. Hence, on the assumption that there is no best world among the worlds God can create, it would seem that God's perfect goodness is fully compatible with his freely creating any one of a number of good worlds that lie in his power to create. To complain that God cannot then be perfect because he could have created a better world is to raise a complaint that no creative action God took would have enabled him to avoid. Thus it appears that if for any creatable world there is a better creatable world, God would be free to select among an infinite number of good worlds to create. (Thomas Aquinas, for example, thought that no matter what world God created he could have created a better world in the sense of a world with a greater amount of good in it.) But the difficulty with this position is that it conflicts with the principle that if a being creates a world when there is a better world that it could have created, then it is possible that there be a being better than it. For it would be possible that there be a being whose degree of goodness is such that it would choose to create a better world. So if for every creatable world there is a better creatable world, it is not clear that the traditional theistic God could be perfectly good and be the creator of a world (see Rowe 1993).

Let us now consider the second way in which it could be true that no creatable world is better than all others. Suppose that among the worlds God can create there are a number of worlds that are equally best. It is plausible to think that if a perfectly good being creates a world, he cannot create a world that is inferior to some other that he can create, at least provided that there is a world than which none is better. Therefore, if there are a number of equally best worlds among the worlds he can create, then if he creates at all he must create one of those worlds. But unlike the case when there is exactly one best world among the worlds he can create, here some degree of divine freedom can exist in harmony with God's perfect goodness. For God would seem to be free to create any one of the equally best worlds. In any case, God's absolute perfection imposes no requirement on his creation from among those equally best worlds.

Is God free not to create a world at all? Here we may limit our inquiry to the possibility that there is exactly one best world among the worlds creatable by an omnipotent being. As we have seen, in this case there is some reason to think that God is not free to select any other world to create. Our present question is whether in this situation God is free not to create at all.

What world would be actual if God exists but does not create at all? Apart from God and whatever necessarily existing entities there are, in a world God inhabits but does not create, no other being would exist. To answer our question of whether God is free not to create a world at all, we must compare the best world God can create with a world whose inhabitants are simply God and whatever necessarily existing entities there are. Assuming such a world would not be incommensurate with the best among worlds God can create, there is reason to think that God is not free with respect to whether he will not create at all. For either the world he inhabits but does not create is better than the best world he can create or it is not. If it is better, then he is not free to create a world, and thus necessarily refrains from creating. If it is worse, then he is not free not to create a world, and thus necessarily creates a world. For it is plausible to hold that if a being can but does not create a world that is better than the one it merely inhabits, then it is possible that there be a being better than it. Would the world he inhabits but does not create be just as good as the best world among those an omnipotent being can create? If so, then, as in the case where there are a number of equally best worlds omnipotence can create, we again have room for some degree of divine freedom. But in this situation, God's freedom would be restricted to creating the best world or not creating

at all.

4 Possible solutions to the difficulties

The difficulty with the first view of divine freedom is that it fails to satisfy our basic conviction that a free action requires the freedom to have done otherwise. Once we see that God's nature would necessitate his action of creating the best of all creatable worlds, the failure to satisfy this basic conviction becomes apparent. The proponent of the first view may try to solve this difficulty by arguing that the intention behind emphasizing the freedom to have done otherwise is not to exclude all necessitating factors, only those for which the agent bears no responsibility. If by freely engaging in strenuous moral effort Mahatma Gandhi developed his moral nature to such a degree that in certain situations he was not capable of acting unkindly towards others, we would still view his acts of kindness as free in a derivative sense since the virtuous nature now necessitating his acts of kindness was itself brought about, at least in part, by his own free moral striving. The difficulty with this suggestion, however, is that God's properties of omnipotence, perfect goodness, and omniscience are not properties he acquired by any kind of striving. In fact, he did not acquire them at all, but had them from all eternity. Thus it seems that although we may describe God as perfectly free in the sense that his actions are never determined by agents or factors outside him, this perfect freedom does not by itself ensure that his actions are free in the sense required if he is to be thanked and praised for performing those actions.

The chief difficulty confronting the second view of God's freedom is that his attributes of omnipotence, omniscience and perfect goodness appear to impose significant restrictions on his freedom in creating a world and in acting within the world he creates. As we saw, it seems that it is not in God's power to do evil or to create less than the best that he can create. Proponents of the second view have tried to solve this problem by distinguishing the necessity imposed on his actions by God's absolute perfections from the sort of necessity that deprives an agent of the power to act otherwise. While admitting that the inflamed passions may causally necessitate an agent's action, thus depriving the agent of the power to do otherwise, some proponents of the second view hold that rational motives are not causes at all. In their view, rational motives do not deprive the agent of the power to act against their counsel. On the other hand, God's essential attributes of omnipotence and perfect goodness make it truly impossible that God should fail to follow what he sees to be the best course of action. How, then, can God be free in carrying out what he sees to be the best course of action? How can the fact that he cannot do other than choose the best not be seen as a limit on his freedom? The answer traditionally given is that power extends only to what is logically possible. A being who is essentially omnipotent and perfectly good cannot possibly fail to choose the best course of action. But this implies no lack of power, for power is lacking only when an agent cannot do what is possible to be done. Thus Samuel Clarke (§1) concludes his discussion of divine freedom by remarking:

> It is no diminution of power not to be able to do things which are no object of power. And it is in like manner no diminution either of power or liberty to have such a perfect and unalterable rectitude of will as never possibly to choose to do anything inconsistent with that rectitude.
>
> ([1706] 1738: 122)

The solution proposed by Clarke and others is promising. That it is not within God's power to weaken himself does not mean that his power is less than infinite. For it is logically impossible that an essentially omnipotent being should weaken itself. Hence, since power extends only to what is possible, the fact that it is not within God's power to weaken himself does not imply any diminution of power. And perhaps the same can be said for it not being within God's power to refrain from following what he sees as the best course of action. This too may not imply any diminution of power. But for God to act freely in following the best course of action it must be within his power not so to act. The fact that it is not within his power not so to act may not imply any diminution of power, but it remains unclear why it does not imply a diminution of freedom.

See also: Creation and conservation, religious doctrine of; Free will; Leibniz, G.W. §§3, 7; Omnipotence §5; Simplicity, divine §2

WILLIAM L. ROWE

References and further reading

Adams, R.M. (1972) 'Must God Create the Best?', *Philosophical Review* 81 (3): 317-32.(An important discussion of God's obligations with respect to what he creates.)

Clarke, S. (1706; 9th edn 1738) *A Discourse Concerning the Being and Attributes of God, the Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation*, British Philosophers and Theologians of the 17th and 18th Centuries, New York: Garland, 1978.(Contains Clarke's account of how God's inability to do less than the best is no diminution of his perfect power or his perfect freedom; the Garland reprint is of the ninth (1738) edition.)

Garcia, L.L. (1992) 'Divine Freedom and Creation', *Philosophical Quarterly* 42 (167): 191-213.(Defends a Thomistic view of God's perfect freedom with respect to creation against some recent challenges to it.)

Kretzmann, N. (1990) 'A General Problem of Creation', in S. MacDonald (ed.) *Being and Goodness*, Ithaca, NY, and London: Cornell University Press.(Discusses Aquinas' theory of God's freedom with respect to creating a world or not creating at all.)

Kretzmann, N. (1990) 'A Particular Problem of Creation', in S. MacDonald (ed.) *Being and Goodness*, Ithaca, NY and London: Cornell University Press.(Discusses Aquinas' theory of God's freedom in selecting among worlds the one he will create.)

Leibniz, G.W. (1704) *New Essays on Human Understanding*, trans. and ed. P. Remnant and J. Bennett, Cambridge: Cambridge University Press, 1982.(Contains Leibniz's discussion of Locke's account of free will.)

Leibniz, G.W. (1710) *Theodicy*, trans. E.M. Huggard, ed. A. Farrer, La Salle, IL: Open Court, 1985.(Contains Leibniz's extended treatment of divine freedom and the problem of evil.)

Morris, T.V. (1993) 'Perfection and Creation', in E. Stump (ed.) *Reasoned Faith*, Ithaca, NY, and London: Cornell University Press.(Argues that if it is impossible for God to do the best he can, his perfection is not impaired by his creating a world when he could have created a better world.)

Morriston, W. (1985) 'Is God Significantly Free?', *Faith and Philosophy* 2 (3): 257-64.(Argues that God lacks the freedom to refrain from doing what is morally obligatory.)

Quinn, P. (1982) 'God, Moral Perfection, and Possible Worlds', in F. Sontag and M. Darrol Bryant (eds.) *God: The Contemporary Discussion*, New York: The Rose of Sharon Press.(Somewhat technical discussion of the principles governing God's freedom in relation to creation.)

Rowe, W.L. (1993) 'The Problem of Divine Perfection and Freedom', in E. Stump (ed.) *Reasoned Faith*, Ithaca, NY, and London: Cornell University Press.(Argues that God is not free to create a given world if there is a better world he can create.)

Frege, Gottlob (1848-1925)

A German philosopher-mathematician, Gottlob Frege was primarily interested in understanding both the nature of mathematical truths and the means whereby they are ultimately to be justified. In general, he held that what justifies mathematical statements is reason alone; their justification proceeds without the benefit or need of either perceptual information or the deliverances of any faculty of intuition.

To give this view substance, Frege had to articulate an experience- and intuition-independent conception of reason. In 1879, with extreme clarity, rigour and technical brilliance, he first presented his conception of rational justification. In effect, it constitutes perhaps the greatest single contribution to logic ever made and it was, in any event, the most important advance since Aristotle. For the first time, a deep analysis was possible of deductive inferences involving sentences containing multiply embedded expressions of generality (such as 'Everyone loves someone'). Furthermore, he presented a logical system within which such arguments could be perspicuously represented: this was the most significant development in our understanding of axiomatic systems since Euclid.

Frege's goal was to show that most of mathematics could be reduced to logic, in the sense that the full content of all mathematical truths could be expressed using only logical notions and that the truths so expressed could be deduced from logical first principles using only logical means of inference. In this task, Frege is widely thought to have failed, but the attempted execution of his project was not in vain: for Frege did show how the axioms of arithmetic can be derived, using only logical resources, from a single principle which some have argued is, if not a logical principle, still appropriately fundamental. In addition, Frege contributed importantly to the philosophy of mathematics through his trenchant critiques of alternative conceptions of mathematics, in particular those advanced by John Stuart Mill and Immanuel Kant, and through his sustained inquiry into the nature of number and, more generally, of abstract objects.

In the course of offering an analysis of deductive argument, Frege was led to probe beneath the surface form of sentences to an underlying structure by virtue of which the cogency of inferences obtains. As a consequence of his explorations, Frege came to offer the first non-trivial and remotely plausible account of the functioning of language. Many of his specific theses about language - for instance, that understanding a linguistic expression does not consist merely of knowing which object it refers to - are acknowledged as of fundamental importance even by those who reject them.

More generally, three features of Frege's approach to philosophical problems have shaped the concerns and methods of analytic philosophy, one of the twentieth century's dominant traditions. First, Frege translates central philosophical problems into problems about language: for example, faced with the epistemological question of how we are able to have knowledge of objects which we can neither observe nor intuit, such as numbers, Frege replaces it with the question of how we are able to talk about those objects using language and, once the question is so put, avenues of exploration previously invisible come to seem plausible and even natural. Second, Frege's focus on language is governed by the principle that it is the operation of sentences that is explanatorily primary: the explanation of the functioning of all parts of speech is to be in terms of their contribution to the meanings of full sentences in which they occur. Finally, Frege insists that we not confuse such explanations with psychological accounts of the mental states of speakers: inquiry into the nature of the link between language and the world, on the one hand, and language and thought, on the other, must not concern itself with unshareable aspects of individual experience.

These three guiding ideas - lingua-centrism, the primacy of the sentence, and anti-psychologism - exercised a commanding influence on early analytic philosophers, such as Wittgenstein, Russell and Carnap. Through them, these ideas have been spread far and wide, and they have come to create and shape analytic philosophy, with whose fathering Frege, more than anyone else, must be credited.

1 Life and work

Gottlob Frege, a German philosopher and mathematician, is the father of modern logic and one of the founding figures of analytic philosophy. Trained as an algebraic geometer, he spent his professional life at the University of Jena, where, because his views about logic, mathematics and language were generally at odds with the dominant

trends of the time, he laboured independently on his central philosophical project.

Frege's main work consists of his *Begriffsschrift (Conceptual Notation)* (1879), in which he first presents his logic; *Die Grundlagen der Arithmetik (The Foundations of Arithmetic)* (1884), in which he outlines the strategy he is going to employ in reducing arithmetic to logic and then goes on to provide the reduction with a philosophical rationale and justification; *Grundgesetze der Arithmetik (Basic Laws of Arithmetic)* (volumes 1, 1893, and 2, 1903), in which he seeks to carry out the programme in detail (a planned third volume was aborted following Bertrand Russell's communication to Frege in 1902 of his discovery of paradox in Frege's logic); and a series of philosophical essays on language, the most important of which are *Funktion und Begriff (*'Function and Concept'*)* (1891), 'Über Sinn und Bedeutung' ('On Sense and Reference') (1892a), 'Über Begriff und Gegenstand' ('On Concept and Object') (1892b) and 'Der Gedanke: eine logische Untersuchung' ('Thoughts') (1918).

In general, Frege was not philosophically intrigued by what was specific to the human condition; for instance, he sought neither to probe the nature and limits of human knowledge nor to understand how humans actually reason. Yet he pursued his non-parochial philosophical interests by attending carefully to natural language and to the way it serves to express our thoughts. Frege's approach, in conjunction with his powerful tool of linguistic analysis (logic) and the collection of subtle, innovative and interwoven theses about language which he elaborated, spurred in others not only an intense interest in language as an object of inquiry but also adherence to a distinctive methodology which has come to characterize much analytic philosophy past and present.

2 Language and ontology

Frege's short book *Begriffsschrift* revolutionized the study of deductive inference. In the course of explaining how his 'conceptual notation' relates to natural language, Frege illuminates important features of language's underlying structure. He recognized that traditional grammatical categories have no logical significance and he urged the consideration instead of the categories of 'singular terms' (which he calls 'proper names') and of 'predicates' (which he calls 'concept-words'). For Frege, a singular term is a complete expression, one which contains no gaps into which another expression may be placed; for example, 'Virginia Woolf', 'the third planet from the Sun' and 'the largest prime number' are all singular terms according to Frege.

By contrast, a predicate such as '( ) was written by Virginia Woolf' is something incomplete; it is a linguistic expression which contains a gap and which becomes a sentence once this gap is filled by a singular term. (The parentheses are not part of the predicate but are intended only to indicate the location of the gap.) Thus, if we fill the gap with 'the third planet from the Sun', we get a complete sentence. This example shows that the resulting sentence may be false or perhaps even nonsense; what is important is that a complete sentence does result. Other examples of predicates are 'Leonard Woolf married ( )', '( ) orbits Jupiter' and '( ) is an even prime number greater than two'; there are infinitely many predicates in any natural language.

Frege distinguishes more finely between predicates, depending on how many gaps they contain and on the types of linguistic expressions that can fill them. The predicates we have so far considered each have just one gap and are known as 'one-place' predicates. '( ) is the mother of ( )' is an example of a 'two-place' predicate, for it contains two gaps, each to be filled by a singular term. Predicates whose gaps are to be filled by one or more singular terms, as is the case with all those mentioned so far, are said to be 'first-level'. Predicates whose gaps are to be filled by first-level predicates are said to be 'second-level', and so on. For example, when properly analysed, 'All [ ] are mammals' is seen to be a second-level predicate: its structure is really 'Everything is such that if it [ ], then it is a mammal', which makes it plainer that the gap is to be filled by a predicate of first-level. (We use square brackets to distinguish such gaps, which must be filled by first-level predicates, from those occurring in first-level predicates, which must be filled by singular terms.) And, for Frege, 'There is at least one thing which [ ]' is likewise a second-level predicate: as we shall see in a moment, this corresponds to his view that existence is not a concept that applies to objects but rather to concepts. This finer classification will not be discussed further here. Nor will the issue of how precisely to understand the incompleteness of predicates, and the corresponding completeness of singular terms, a question that continues to be debated.

The following two points, however, are not in dispute: first, Frege discerns in the categories of reality counterparts to the linguistic categories of singular term and predicate; he calls these ontological categories 'object' and 'concept', respectively. Second, Frege understands concepts on the model of functions as they are commonly

encountered in mathematics.

A singular term refers to, or designates, an object. A predicate refers to, or designates, a concept. (We shall use ‘designate’ and ‘refer’ and their cognates interchangeably; in some discussions of Frege, ‘denote’ and ‘mean’ are also used.) Corresponding to the fact that a first-level predicate yields a complete sentence upon its gap being filled by a singular term, we have the fact that a first-level concept is true or false of an object - or, as Frege puts it, that an object ‘falls under’ or fails to fall under a concept. For this reason, Frege calls concepts ‘unsaturated’: unlike objects, they await completion, whereupon they yield one of the two truth-values, which Frege takes to be objects: the True and the False (1891: 6). For example, the concept designated by ‘( ) is an Oxonian’ yields the value the False when completed with the object designated by ‘Gottlob Frege’ and the value the True when completed with the object designated by ‘John Locke’.

Concepts are incomplete in the way that functions in mathematics are. For example, the function designated by ‘ 2+ ( )’ yields the value 8 when completed with the object 6, or, as we might more simply say, it yields that value for the argument 6. The function ‘ 2+ ( )’ is not an object, but yields one upon completion by an argument. On Frege’s view, concepts are a kind of function, namely those that take as their only values the True or the False.

In describing this congruence of linguistic and ontological categories, Frege is not confusing use and mention (see Use/mention distinction and quotation); indeed, Frege exhibited an understanding of the distinction between words and what they designate not witnessed again until well into the twentieth century. Though linguistic expressions are claimed to slot into a certain set of categories and reality into another, an intimate connection obviously exists between the two categorial schemes. Whether Frege takes the linguistic scheme or instead the ontological one as fundamental is a subject of debate and bound up with the question of whether and in what sense Frege sees philosophical reflection on language as the foundation of philosophy.

Regardless of this debate’s resolution, we know that Frege takes the categories of concept and object to be fundamental ontologically, ones not amenable to further analysis. In addition, basic structural features of language raise insurmountable obstacles to expressing certain truths about these categories. This can be demonstrated by considering the obvious claim (1) seeks to articulate:

(1) The concept designated by ‘( ) is a horse’ is a concept.

This seems patently correct. Yet Frege recognizes in ‘On Concept and Object’ (1892b) that we must judge (1) to be false. This is because the expression ‘The concept designated by "( ) is a horse"’ is a singular term and hence refers to an object, not to a concept. (The expression does not contain a gap, though it does mention one.) So (1) does not succeed in expressing what we intended. To do this, we need to fill the gap in ‘( ) is a concept’ by an expression that refers to a concept. But the straightforward way of doing this yields (2):

(2) ( ) is a horse is a concept,

which is not even a sentence. In trying to articulate our thought, we are led either to falsity or to something that makes no claim at all. This ‘awkwardness of language’, as Frege calls it, has been the subject of much controversy, but it seems clear that Frege draws the lesson that there are elemental facts about language and the world that perforce escape expression (1892b: 196).

3 Sense and reference

How do what words refer to (concepts and objects) relate to our understanding of language? In his seminal essay ‘On Sense and Reference’ (1892a), Frege considered whether the ‘sense’ of an expression - what it is that we know when we understand the expression - is simply identical to what it designates (the ‘reference’). Frege offers the following argument, as famous as it is simple, to show that our understanding of singular terms cannot consist just of knowing their reference:

(3) (a) If two singular terms *t* and *t*′ have the same sense and *C* is any (first-level) predicate, then *C*(*t*) has the same sense as *C*(*t*′).
(b) ‘The evening star = the morning star’ does not have the same sense as ‘The evening star = the evening star’.
(c) ‘The evening star’ does not have the same sense as ‘the morning star’.
[This follows from (a) and (b): let *C* in (a) be ‘The evening star = ( )’.]

(d) ‘The evening star’ refers to the same object as ‘the morning star’.
(e) The reference of ‘the evening star’ is not identical with its sense.
[This follows from (c) and (d).]

So the sense of an expression - that which must be known in order for a speaker to understand it - cannot be identified with its reference (see Sense and reference). (A terminological digression: in this essay of 1892, Frege used the words ‘*Sinn*’ and ‘*Bedeutung*’, respectively. There is, however, no consensus on how these should be rendered in English. More importantly, there is not yet agreement about how precisely to understand these notions and how they are related to more everyday notions, such as meaning.)

Frege’s justification for premise (a) relies on a ‘compositionality thesis’: the sense of a sentence is determined by the senses of its components (and by the way in which the sentence is constructed from them; see Compositionality). Premise (b) is justified by noting that an alert speaker would find the one sentence obvious, but not the other: ‘The evening star = the evening star’ is an uninformative statement of self-identity, whereas ‘The evening star = the morning star’ may impart hitherto unsuspected information. This difference in what Frege called ‘cognitive value’ suffices, according to him, to register distinct senses (1892a: 25). Finally, premise (d) follows from the observation that both singular terms designate the planet Venus, an astronomical discovery that was not made for some time.

If sense is not reference, then what is it? To this, Frege provides no clear answer. He writes that the sense of an expression is ‘the mode of presentation of that which is designated’ (1892a: 26), but he does not offer any elaboration regarding the nature of these modes (see Proper names §1). Frege does, however, advance a number of other theses concerning the relation between sense and reference. In the first place, the sense of an expression determines the identity of its reference, but not vice versa. For instance, the expression ‘the author of *Begriffsschrift*’ designates a particular individual, Frege, and any expression with the same sense designates the same individual; yet the expression ‘the author of *Die Grundlagen der Arithmetik*’, which also refers to Frege, has a different sense. Likewise, though ‘George Orwell’ and ‘Eric Blair’ both designate the same person (they have the same reference), the two singular terms have distinct senses.

Second, expressions can be formed which, while possessing a sense, lack a reference. For example, ‘Sherlock Holmes’ is a singular term that has a sense but lacks a reference, for Holmes does not exist. This is not in conflict with saying that sense determines reference, for what this means is that if two expressions have the same sense, then they have the same reference. One might wonder, though, whether this thesis is at odds with Frege’s description of sense as the way ‘that which is designated’ is presented: how can there be such a way when that which is designated, the reference, does not exist? (See Evans 1981.)

Many have found in the distinction between sense and reference relief from the philosophical distress that follows upon assuming, first, that we understand an expression by directly associating a reference with it and, second, that we understand expressions that fail to refer. Holding both of these assumptions has led philosophers to extravagant claims about the reality that is actually designated by our expressions, for instance that Holmes must exist in some fashion if we are to speak of him intelligibly (if only to deny his existence). Frege’s distinction between sense and reference, and his thesis that an expression can have a sense while lacking a reference, simply dissolves the problem by rejecting the first assumption (see Descriptions §1).

The argument (3) concerns singular terms, but Frege believed that other kinds of expressions have sense and reference. The distinction can also be drawn in the case of predicates:

(4) Something is a bottle of claret if and only if it is a bottle of claret.
(5) Something is a bottle of claret if and only if it is a bottle of Hume’s favourite wine.

(4) does not have the same sense as (5): no one would deny the first, but the second might come as something of a gastronomic discovery. Though ‘( ) is a bottle of claret’ and ‘( ) is a bottle of Hume’s favourite wine’ are predicates that refer to the same concept, they have different senses. (Frege identifies two concepts if an object falls under the one if and only if it falls under the other.)

What is the reference of a whole sentence? Frege answers this by observing what remains unchanged about a sentence when we substitute coreferring expressions (that is, expressions that have the same reference) in it.

Assuming a compositionality thesis for reference (that the reference of a sentence is determined by the reference of its components), we then have some reason to take whatever remains unchanged to be the sentence's reference. Consider (6) and (7):

(6) George Orwell wrote *1984*.
(7) Eric Blair wrote *1984*.

What remains unchanged? Not the 'thought' expressed by each sentence: someone might believe one sentence to be true, but not the other. Rather, it is the truth-value of the sentences that is constant: (6) and (7) are either both true or both false. This leads Frege to identify the reference of a sentence with its truth-value; a sentence refers to one of the two truth-values. Frege takes the two truth-values to be objects and he observes that on his view all true (false) sentences are really singular terms that refer to the same object, the True (the False).

Because the sciences are interested in what is true, we can see why Frege holds that reference 'is thus shown at every point to be the essential thing for science' ([1892-5] 1979: 123). Natural language permits the formation of expressions that lack reference, and so Frege judged it unsuitable as a tool of rational inquiry. For this purpose, it was inferior to his *Begriffsschrift*, a formal language he designed with the intention that no expression without reference could be constructed in it.

Frege notes that if his thesis that the reference of a sentence is its truth-value is correct, then we would predict (again on the basis of the compositionality thesis for reference) that if a subordinate sentence is replaced by a coreferential one (which will have the same truth-value), the reference of the entire containing sentence (that is, its truth-value) will remain unchanged. For instance, consider:

(8) Ronald Reagan was elected President in 1984 and George Orwell wrote *1984*.
(9) Ronald Reagan was elected President in 1984 and Eric Blair wrote *1984*.

(8) and (9) have the same truth-value, as predicted if the reference of a sentence is its truth-value. But now consider:

(10) Stimpson believes that George Orwell wrote *1984*.
(11) Stimpson believes that Eric Blair wrote *1984*.

If Stimpson does not realize that George Orwell and Eric Blair are one and the same person, (10) and (11) might have different truth-values, that is, they might differ in reference. And yet one is obtained from the other merely through substitution of what, on Frege's hypothesis, are the coreferring sentences (6) and (7).

Frege defends his hypothesis by claiming that in certain contexts expressions refer not to their ordinary reference but rather to an 'indirect' reference (1892a: 28). And, he adds, the indirect reference of an expression is just its ordinary sense. Because 'Stimpson believes that ( ) wrote 1984' is just such a context and because 'George Orwell' has a different sense from 'Eric Blair' (and so, in this context, a different reference), the compositionality thesis for reference no longer forces us to the conclusion that (10) and (11) have the same truth-value. Frege's response inaugurated a long, fruitful and continuing debate about the nature of such linguistic contexts. (For further discussion, see Indirect discourse; Sense and reference §5; Propositional attitude statements.)

And what about the sense of a complete sentence? Given the compositionality thesis for sense, it will be preserved upon substitution of one expression in a sentence by another with identical sense. Frege says that such a substitution preserves the thought expressed by the sentence and he consequently identifies this thought with the sentence's sense. So, because 'to lie' has the same sense as 'to express something one believes to be false with the intention of deceiving', Frege would predict that the following two sentences express the same thought:

(12) Everyone has lied.
(13) Everyone has expressed something he believes to be false with the intention of deceiving.

And this prediction does seem to be borne out. But what precisely is the thought expressed by a sentence?

4 Thought and thinking

When one thinks the thought that lemons are sour, many different kinds of psychological events may transpire; certain memories, images or sensations may be triggered. These events, according to Frege, belong to the psychological world of the subject and, as such, are not fully shareable with others: much as one may try, one cannot experience what another does. It is a serious error, Frege says, to confuse such private events as may accompany our grasping of a thought with the thought that is grasped. One commits the sin of ‘psychologism’ if one does not sharply distinguish between the psychological process of thinking and the thoughts that are, as a consequence of this private activity, apprehended.

Thoughts then, in contrast to what Frege calls ‘ideas’, are fully shareable. When you and I grasp the sense of ‘Lemons are sour’, we arrive at the very same thought: there are not two different, related thoughts (as we might, for example, have two different mental images of lemons), but just the one that, through perhaps idiosyncratic and private paths, we both succeed in apprehending. And so it is in general with senses: they are not of the mental world, but objective in that different individuals can grasp them and associate them with their words. To stray from this perspective, according to Frege, is simply to abandon the view that communication is possible; that two speakers can understand a linguistic expression in the same way.

To grasp a thought is not to hold it true. For though one grasps a thought in the course of asserting it, it is no less grasped in the acts of assuming it to be true, wishing that it were true, commanding that it be made true, questioning whether it is true, and so on. These acts correspond to the different kinds of ‘force’ that may be attached to a thought. This is not quite Frege’s position, but it is an influential one closely related to his views (1892a: 38-9). It has been attractive to many students of language because it divides the daunting project of giving an explanation of linguistic understanding and use into two potentially more tractable components. The first task, the articulation of a ‘theory of sense’ or ‘theory of meaning’, is to explain how the sense of a sentence is determined by the senses of its parts. The second task, the articulation of a ‘theory of force’ (sometimes also called ‘pragmatics’), is to explain, taking for granted an account of the thoughts expressed by sentences, the different speech acts into which they may enter. For example, it falls to the theory of meaning to describe the content of ‘Lemurs are native to London’ and to show how that thought is determined by the senses of the sentence’s constituent components. It falls to the theory of force to discover what, beyond the apprehension of that thought, is involved in asking ‘Are lemurs native to London?’, in wishing ‘Oh, would that lemurs were native to London!’, and so on. Should the theory of force need to make reference to the mental states of agents, then it is not merely helpful to divide an account of linguistic understanding and use into these two components but, Frege would insist, essential if we are to keep psychology from intruding into an account of sense (see Pragmatics; Speech acts §1).

5 Objectivity and privacy

In the previous section, we saw that Frege insists on the objectivity of thoughts (and senses generally), intending by this that different speakers can attach the very same thoughts to their sentences. There is a second way in which Frege takes thoughts to be objective. To say that thoughts are shareable is compatible with saying that their existence and properties are dependent upon human activity. Frege’s view seems to be that thoughts are also objective in that they exist independently of human activity. Thoughts are not created or shaped by the process of thinking; they exist regardless of whether we have apprehended them, regardless of whether we shall ever apprehend them. Thoughts await our grasp in somewhat the way that physical objects await our observation, though the latter are located in space and time whereas thoughts are not.

These two kinds of objectivity, shareability and independence, are taken by Frege to apply to truth as well, which he considers a property of thoughts. There are not different properties - say, ‘true-for-you’ and ‘true-for-me’ - which are private unto individuals. There is just one property, being true, which some thoughts have and others lack. Furthermore, whether a thought possesses this property is in no way dependent upon our capacity to recognize that it does. A thought’s being true must be sharply distinguished from our believing it to be true or our being justified in taking it to be true. On Frege’s view, the truth of a thought is not dependent upon our beliefs, not even upon our beliefs in some ideal epistemic situation. Truth is one thing, our recognition of truth something else entirely. This position regarding the independence of truth, present throughout Frege’s writings, is a robust realist motif in his thought, the evaluation of which has occupied centre-stage of much contemporary philosophy of language (see Realism and antirealism §4).

Since on Frege’s view a sentence is true if and only if it refers to the value the True, his realism amounts to saying

that a sentence refers to what it does independently of our recognition of this fact. But a sentence refers to an object only via its sense (the thought it expresses), which is what determines its reference. And its sense is not something that it has independently of speakers (after all, an expression - like 'chat', for example - can have one sense in one language and another in another), but is associated with it through human activity. Putting all this together, we see that, according to Frege, humans associate senses with linguistic expressions - this is what their understanding of language consists of - on account of which those expressions take on references whose identity may remain forever unknown.

That such a realism should follow from the senses attached to expressions forces our attention not only to Frege's notion of sense but, relatedly, to his conception of what it is to grasp a sense and associate it with an expression. Frege's anti-psychologism with regard to sense is not extended by him to an account of the grasping of sense, for in his few remarks about the subject he appears to avow a psychological picture of the process of apprehending a thought, of judging it to be true and so on (see [1897a] 1979: 145). Given his view of the privacy of mental events, it might seem that for Frege one cannot always determine which thought another has associated with a sentence: not that one cannot apprehend the same thought as another (for one feature of the objectivity of thoughts, that they can be grasped by all alike, guarantees that one can); but that one cannot always ascertain that one has.

That a speaker's linguistic understanding might remain private is a view that has troubled many, especially those influenced by the work of Ludwig Wittgenstein (see Private language argument). Michael Dummett, the most influential interpreter of Frege's philosophy (see Dummett 1973), argues that, while much remains of lasting value in Frege's views on language, they should be modified in such a way that not only must senses be graspable by all, something on which Frege already insists, but the attachment of senses to expressions must be accessible as well; that is, the nature of a speaker's linguistic understanding must be a thoroughly public affair (see Private states and language). Once these modifications are made, Dummett claims, important consequences will follow about which senses can be coherently grasped: if the only intelligible senses are those the grasp of which is subject to public inspection, then we must be sceptical of any analysis of linguistic understanding in terms of senses that underwrite realism and we should instead consider seriously analyses employing senses that do not countenance the possibility of sentences being true independently of all potential human knowledge. Both Dummett's argument and his assumptions have been disputed, and the controversy continues to be a lively one. This is not an issue about which Frege explicitly said much, but it is among the many deep debates on the nature of language and thought which his work has made possible.

6 Contributions to logic

Infamous for using a notation difficult to learn to read, Frege's *Begriffsschrift* is one of the greatest logical works ever written. (Its full title in English is *Conceptual Notation: a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought*.) It contains a number of major innovations, two of which are of foundational importance for contemporary logic: a satisfactory logical treatment of generality and the development of the first formal system. *Begriffsschrift* also introduces (what are essentially) truth tables, contains Frege's definition of the ancestral (§9) and sows the seeds of his philosophy of language (§§2-5).

Let us first discuss Frege's treatment of generality, that is, his logical analysis of sentences containing such words as 'everything', 'something', 'no one' and the like. The foundation for this is Frege's predicate-singular term analysis of simple sentences (§2). Sentences such as 'Tony is alive' contain a singular term, 'Tony', and a predicate, '( ) is alive'. We can extend this analysis to such sentences as 'Everything is alive' by drawing 'upon [the formula language of] arithmetic'. In arithmetic, a sentence containing a variable, '*x*', is taken to be true if and only if a true sentence results no matter what *x* is supposed to be. For example, '$x + 2 = 2 + x$' is true if and only if, no matter what x may be, $x + 2 = 2 + x$. So, if we allow the argument of '( ) is alive' to be a variable, the resulting sentence '*x* is alive' may be taken to express the generalization of 'Tony is alive'; it will be true if and only if, no matter what *x* may be, *x* is alive. Similarly, 'Everything is not alive' may be represented as '*x* is not alive'. For, given the convention just explained, this will be true if and only if, no matter what *x* may be, *x* is not alive. But, as a little experimentation will show (and as can be proven), it is impossible to represent such sentences as 'Not everything is alive', so long as we use variables in this way alone. Nor can one so express 'If everything is alive, then snow is black'. One could try to represent it as 'If *x* is alive, then snow is black'. But this actually represents 'Everything is such that, if it is alive, then snow is black' (1879: §11).

What is required here is some way of confining the generality expressed by the variable to a part of the sentence. In his informal discussions, Frege uses the phrase ‘no matter what *x* may be’ to do this (for example, 1879: §12). We may then represent ‘Not everything is alive’ as ‘It is not the case that, no matter what *x* may be, *x* is alive’; and ‘If everything is alive, then snow is black’ as ‘If, no matter what *x* may be, *x* is alive, then snow is black’. The phrase ‘no matter what *x* may be’, and its placement in the sentence, is said to delimit the ‘scope’ of the variable (see Scope). Frege’s most vital discovery is not that variables may be used to indicate generality, but that variables have scope; his most significant innovation, the development of a notation in which scope can be represented, that is, his introduction of the *quantifier*.

Frege’s second fundamental contribution was his construction of the first formal system. A formal system, as Frege conceives it, has three parts: first, a highly structured ‘language’ in which thoughts may be expressed; second, precisely specified ‘axioms’, or basic truths, about the subject matter in question; and third, ‘rules of inference’ governing how one sentence may be inferred from others already established. Frege believed that there were a number of advantages to carrying out proofs in such formal systems, for example, that giving a proof in a formal system would better one’s understanding of the proof, by revealing precisely what principles it employs. Suppose, for example, that one wants to show that a given theorem can be proven without using the axiom of choice: the obvious method would be to prove the theorem without using the axiom. But how can one be sure that the axiom has not tacitly been employed (as it was in Richard Dedekind’s proof of Theorem 159 in *Was sind und was sollen die Zahlen?* in 1888)? Proving the theorem in a formal system makes this possible: the axioms which may be used are clearly specified and steps in the proof may be taken only in accordance with certain rules. (Compare 1879: Preface; 1884: §2. Formal systems are important in contemporary logic for related reasons; see Model theory.)

The formal system of *Begriffsschrift* does not actually live up to the standards Frege imposes upon it: not all its rules of inference are explicitly stated. However, this complaint cannot be made about the system presented in Part I of his *Grundgesetze der Arithmetik* (1893). (The first-order fragment of this system is complete; the second-order fragment is a formulation of standard, second-order logic.) The rigour of Frege’s formulation was not equalled until Kurt Gödel’s work in the early 1930s, almost forty years later.

It is arguable that Frege goes even further and, in *Grundgesetze*, presents a semantics for his system; that is to say, he attempts to explain, rigorously, how the formal system is to be interpreted; how its symbols are to be understood. Using these explanations, he attempts to prove that the axioms of the system, so interpreted, are true and that the rules of the system, so interpreted, preserve truth (that is, in technical parlance, to prove the system’s soundness). If these controversial claims are correct, then Frege may also be credited with having anticipated, to a limited extent, the development of model theory.

7 *Die Grundlagen der Arithmetik*: three fundamental principles

In the Preface to *Begriffsschrift*, Frege announced his interest in determining whether the basic truths of arithmetic could be proven ‘by means of pure logic’. Kant’s answer had been negative. According to Kant, the truths of arithmetic are synthetic a priori: for example, knowledge of ‘ $7+5=12$’ requires appeal to intuition (see Kant, I. §§4-5). One of Frege’s main goals in his *Die Grundlagen der Arithmetik (The Foundations of Arithmetic)* was to refute this view by giving purely logical proofs of the basic laws of arithmetic, thereby showing that arithmetical truths can be known independently of any intuition. Frege conceived the formal system of *Begriffsschrift* as an important prerequisite for this project: without it, it would be impossible to determine whether the complex proofs required do indeed depend only upon axioms of ‘pure logic’.

There has been some controversy about what motivated Frege’s logicism - his view that the truths of arithmetic are truths of logic (see Logicism §1). Frege says that his project is inspired by both mathematical and philosophical concerns (1884: §§1-3). Some, notably Paul Benacerraf (1981), have defended the view that Frege was interested in philosophical problems only in so far as they were susceptible of mathematical resolution. However that may be, Mark Wilson (1992) and Jamie Tappenden (1995) have argued that there are important connections between Frege’s work on arithmetic and then recent developments in geometry to which he would have been exposed during his graduate career.

Grundlagen is important for a number of reasons. The philosophy of arithmetic developed in the book is of

continuing interest. Moreover, a large number of more specific theses propounded there have had a profound influence on later philosophers, including Ludwig Wittgenstein, W.V. Quine and Michael Dummett, to name but three. More generally, *Grundlagen* is arguably the first work of analytic philosophy. At crucial points in the book, Frege makes ‘the linguistic turn’: that is, he recasts an ontological or epistemological question as a question about language. Unlike some ‘linguistic’ philosophers, his purpose is not to dissolve the philosophical problem - to unmask it as a ‘pseudo-problem’ - but to reformulate it so that it can be solved.

According to Frege, his work in *Grundlagen* is guided by ‘three fundamental principles’ he states in the Preface:

> always to separate sharply the psychological from the logical, the subjective from the objective;
> never to ask for the meaning of a word in isolation, but only in the context of a proposition;
> never to lose sight of the distinction between concept and object.
>
> (1884: x)

All of these are sufficiently important to warrant separate discussion. The first, which announces Frege’s opposition to ‘psychologism’, has been discussed in §§4-5; the second, which is called ‘the context principle’, will be discussed in §8 below.

The third principle, which distinguishes the sorts of things singular terms denote - objects - from the sorts of things predicates denote - concepts - has already been discussed in §2. Of present interest is the application Frege makes of this distinction in *Grundlagen* §§45-54. Just prior to these sections, Frege has been concerned with what numbers are, his results having been almost entirely negative: numbers are neither physical objects; nor collections or properties of such; nor subjective ideas. Frege now suggests that progress may be made by asking to what, exactly, number is ascribed. The crucial observation is that different numbers seem to be assignable to the same thing: of a pack of cards, for example, one could say that it was one pack or fifty-two cards. Frege realizes that this might suggest that ascription of number is subjective; dependent upon our way of thinking about the object in question (1884: §§25-6). But what is different in our way of regarding the pack is, specifically, the ‘concept’ we choose to employ: that denoted by ‘( ) is a pack’ in the one case, or by ‘( ) is a card’ in the other. If, with Frege, we insist that concepts, and facts about them, can be just as objective as objects and facts about them (1884: §48; 1891), there is no need to regard number as subjective. Rather, we must acknowledge that number is ascribed not to objects, nor to collections thereof, but to concepts:

> If I say ‘Venus has 0 moons’, there simply does not exist any moon or agglomeration of moons for anything to be asserted of; but what happens is that a property is assigned to the *concept* ‘moon of Venus’, namely that of including nothing under it. If I say, ‘the King’s carriage is drawn by four horses’, then I assign the number four to the concept ‘horse that draws the King’s carriage’.
>
> (1884: §46; emphasis added)

(Note that we will use ‘"moon of Venus"’ and ‘( ) is a moon of Venus’ to denote the same concept.) As Frege famously puts the point in §55: ‘the content of a statement of number is an assertion about a concept’.

Observe here how Frege’s interest in what numbers are has led him to an interest in the nature of ascriptions of number and, in particular, to an investigation of the ‘logical form’ of such statements. According to Frege, the most fundamental way of referring to a number is by means of an expression of the form ‘the number belonging to the concept *F*’; for example, ‘the number belonging to the concept "moon of the earth"’ refers to the number one, for there is just one object that is a moon of the earth. This seemingly innocuous linguistic claim plays a crucial role in Frege’s account of what numbers are.

8 *Die Grundlagen der Arithmetik*: the context principle

Frege denied both that numbers are physical objects and that they are objects of intuition, in Kant’s sense. In *Grundlagen* §62, he therefore raises the question of how ‘numbers [are] given to us, if we cannot have any ideas or intuitions of them’. This question is plainly epistemological, concerning how we can have knowledge of the objects (and so the truths) of arithmetic. What is astonishing is how Frege sets about answering it: ‘Since it is only in the context of a proposition that words have any meaning, our problem becomes this: to define the sense of a proposition in which a number word occurs.’ As Michael Dummett has emphasized, Frege here makes ‘the

linguistic turn' in a profound way: what was plainly an epistemological problem is converted into one about language (Dummett 1991: 111-12). The question of how we can have knowledge about numbers becomes the question of how we refer to - that is, succeed in talking about - numbers.

Frege suggests that this question can be answered by examining whole sentences in which names of numbers occur. It is here that the second of Frege's fundamental principles, the context principle, is at work. Frege rejects any requirement that he should point out or, in some other way, display numbers for his audience. He has already asserted that this would be impossible, since numbers cannot be encountered in perception (we have no 'ideas' of them) or in intuition. Rather, Frege insists, one's ability to refer to numbers should be explained in terms of one's understanding of complete sentences in which names of numbers are employed. That is to say, Frege refuses to say what 'zero' refers to except to say what such sentences as 'Zero is the number belonging to the concept "moon of Venus"' mean. More precisely, he insists that to explain the meanings of such sentences is to say what 'zero' refers to.

Frege intends this view to be generalized: indeed, his own discussion of it does not concern numbers directly, but the analogous case of directions. So-called abstract objects pose serious philosophical problems, both ontological and epistemological. Frege's strategy for defending their existence and for defusing worries about our cognitive access to them is appealing and may well seem the only workable option. His general idea, that our capacity to refer to objects of a given kind may be explained only in terms of our understanding of sentences containing names of them, has been of continuing influence.

To return to our main thread, the immediate goal is to explain the meaning of sentences in which reference is made to numbers. Frege claims that, when we are concerned with names of *objects*, the most important such sentences are those asserting an identity. Because Frege takes numbers to be objects, he focuses upon such sentences as 'The number belonging to the concept "plate on the table" is the same as that belonging to the concept "guest at dinner"'. Frege observes that this sentence will be true if and only if there is a way of assigning plates to guests such that each guest gets exactly one plate and each plate, exactly one guest; that is, if and only if there is a one-to-one correlation between the plates and the guests. More generally, say that the concept F is equinumerous with the concept G if and only if there is a one-to-one correlation between the objects falling under F and those falling under G. Then the thought is this: the number belonging to the concept $F =$ the number belonging to the concept G if and only if the concept F is equinumerous with the concept G. Since Frege introduces it with a quotation from Hume, this is sometimes called 'Hume's Principle' (1884: §§55-63). The principle had been known for some time, but it was only in the work of Georg Cantor (1897) that its mathematical significance was fully realized (see Set theory §2).

Of course, if Hume's Principle is to play any role in Frege's attempt to prove the axioms of arithmetic from logical principles alone, the notion of equinumerosity must be definable in purely logical terms. But Frege shows that it is, if the general theory of relations is accepted as part of logic (1884: §§70-2).

For reasons which are not entirely clear, Frege rejects the claim that Hume's Principle suffices to explain numerical identities. His stated reason (1884: §66; compare §56) is that it fails to decide whether Julius Caesar is the number zero! But there is little agreement about the point of this complaint or about its force. Still, Hume's Principle continues to be important to Frege, since he insists that any correct explanation of numbers must have Hume's Principle as a (relatively immediate) consequence. He himself settles upon an explicit definition of names of numbers: the number belonging to the concept F is to be the extension of the second-level concept '[ ] is a concept equinumerous with the concept F'. (Roughly, the extension of a concept is the collection of things that fall under that concept.) Frege shows how Hume's Principle may be derived from this definition (1884: §73). To make the definition and proof precise, however, he has to appeal to some axiom concerning extensions. Frege's idea, developed in *Grundgesetze*, was that extensions could be characterized by means of a principle analogous to Hume's Principle, namely, his Basic Law V: the extension of the concept F is the same as that of the concept G if and only if the very same objects fall under the concepts F and G (1893: §§3, 20). Famously and unfortunately, however, Bertrand Russell showed Frege, in 1902, that the resulting theory of extensions is inconsistent, since Russell's paradox can be derived from Basic Law V in (standard) second-order logic (see Paradoxes of set and property §4).

9 Frege's formal theory of arithmetic

The story of Frege's work on arithmetic might well have ended there. In *Grundlagen*, Frege does sketch proofs of axioms for arithmetic (1884: §§70-83) and, in *Grundgesetze*, offers formal versions of them (1893: §§78-119). But little attention was paid to these proofs for almost a century, on the ground that Frege gave his proofs in an inconsistent theory, and anything can be proven in an inconsistent theory. Frege himself decided in about 1906 that no suitable reformulation of Basic Law V was forthcoming and, his wife having died in 1904, appears to have become deeply depressed. He published nothing at all between 1908 and 1917 and only three further articles after that.

Closer attention to the structure of Frege's proofs reveals something interesting, however. As already stated, Frege requires that his explicit definition of names of numbers imply Hume's Principle, and he shows that it does (given Basic Law V). But neither the explicit definition nor Basic Law V is used essentially in the proof of any other arithmetical theorem; these other theorems are proven using only second-order logic and Hume's Principle. Thus, Frege in fact proves that axioms for arithmetic can be derived in second-order logic from Hume's Principle alone. Frege himself knew as much, but, sadly, never appreciated the philosophical and mathematical significance of this result, now known as Frege's Theorem.

The details of Frege's proof of this result are beyond the scope of this discussion. But a few points are worth mentioning. First, one cannot actually prove each of the infinitely many truths of arithmetic from logical principles, or from anything else. So any attempt to show that all truths of arithmetic follow from logical principles, or from Hume's Principle, will depend upon the identification of some finite number of basic laws, or axioms, of arithmetic from which we are confident all other arithmetical truths will follow. The most famous such axioms are those due to Dedekind (though widely known as the Peano axioms; see Numbers §6). Frege employs his own axiomatization which, while similar, is importantly different and arguably more intuitive (1893: §§128-57).

Second, Frege's hardest task is to prove that there are infinitely many numbers, and his method of doing so is extremely elegant. The basic idea is as follows. Begin by noting that 0 is the number belonging to the concept 'object not the same as itself' and then that 1 is the number belonging to the concept 'identical with 0'. Further, 2 is the number belonging to the concept 'identical with 0 or 1', and so on. More generally, if *n* is finite, then the number of the concept 'natural number less than or equal to *n*' is always one more than *n*, which implies that every finite number has a successor and (in the presence of the other axioms) that there are infinitely many numbers (1884: §§82-3; 1893: §§114-19).

Third, Frege needs, for a variety of reasons, to define the notion of a finite, or natural, number. He also needs to prove the validity of proof by induction, since one of his (and Dedekind's) axioms is, essentially, that such proofs are legitimate. Induction is a way of proving that all natural numbers fall under a concept *F*; a proof by induction proceeds by showing (1) that 0 falls under *F* and (2) that, if a number *n* falls under *F*, $n+1$ must also fall under *F*. In essence, Frege *defines* the natural numbers as those objects for which induction works. According to Frege's definition, a number is a natural number if and only if it falls under every concept *F* which is a concept (1′) under which 0 falls and (2′) which is 'hereditary in the number series', that is, under which $n+1$ falls whenever *n* does. It then follows, immediately, that proof by induction is valid: if *F* is a concept satisfying (1) and (2), then, since *F* is then a concept satisfying (1′) and (2′), every natural number must fall under it, by definition.

It is possible to see, intuitively, that this is a good definition, that is, that what Frege calls 'the natural numbers', the things that fall under every concept satisfying (1′) and (2′), *really are* the natural numbers. Certainly, if *x* is a natural number, then it falls under every concept satisfying (1′) and (2′). Conversely, suppose that *x* falls under every concept satisfying (1′) and (2′). Well, the concept 'natural number' is such a concept: for (1′′) 0 certainly falls under it, and (2′′) whenever a number *n* falls under it, $n+1$ also falls under it. So, since *x* falls under *every* concept satisfying (1′) and (2′), it must fall under this one, that is, it must be a natural number. Thus, falling under every concept satisfying (1′) and (2′) is both necessary and sufficient for being a natural number. (Objections can be made to this argument, chiefly on the grounds of its 'impredicativity'; see Property theory §1.)

This method of definition may be generalized to furnish a definition of the 'ancestral' of any given relation. Frege's definition of the ancestral, introduced in *Begriffsschrift* (and independently discovered by Dedekind), is of quite general importance in mathematics.

Frege's formal work in *Grundgesetze* does not end with his proof of the arithmetical axioms. He goes on to develop purely logical definitions both of finitude and of infinitude; he proves the so-called categoricity theorem, that any two structures which satisfy his axioms for arithmetic are isomorphic; and he proves the validity of definition by induction. (These last two results were first proven by Dedekind.) There is also some reason to believe that Frege was interested in proving results now known to depend upon the axiom of choice and that his investigations led him to discover - though not to communicate - this axiom some years before Ernst Zermelo's formulation of it in 1904 (see Axiom of choice). Finally, in later parts of *Grundgesetze*, Frege proves a number of preliminary results required for a logicist development of the theory of real numbers, but a projected third volume, which would have completed that part of the project, was abandoned in the wake of Russell's paradox.

10 The fate of Frege's logicism

Frege proved Frege's Theorem, that is, that axioms for arithmetic are (second-order) logical consequences of Hume's Principle. Of course, this would be of little interest but for the fact that, unlike Basic Law V, Hume's Principle is consistent. If, therefore, Hume's Principle could be argued to be a truth of logic, then the truths of arithmetic, being logical consequences of a truth of logic, would all be truths of logic and logicism would be vindicated! Work on Frege's philosophy of arithmetic has therefore tended to abstract from Frege's ill-fated explicit definition and to concern itself with the view Frege himself rejected, namely that names of numbers may be defined, or explained, by means of Hume's Principle.

No one nowadays really thinks that Hume's Principle is a truth of logic. Still, one might think that Hume's Principle is suitable as some kind of definition, or philosophical explanation, of names of numbers. And, or so argues Crispin Wright (1983), if the truths of arithmetic turn out to be logical consequences of an explanation, is that not almost as good as if they were truths of logic? This view is appealing precisely because Hume's Principle does seem to capture something very fundamental about (cardinal) numbers: saying that the number belonging to the concept *F* is the same as that belonging to *G* if and only if the concepts *F* and *G* are equinumerous just does seem a good way of explaining what (cardinal) numbers are. Still, there are serious worries. Saying that the extension of *F* is the same as the extension of *G* if and only if the same objects fall under *F* and *G* also *seems* to be a good way of explaining what extensions are - until one realizes this explanation is inconsistent. Hume's Principle is not inconsistent, of course, but one might wonder whether it can rightly be viewed as explanatory of what numbers are if such a kindred principle cannot rightly be viewed as explanatory of what extensions are. It is in the debate over this, the 'bad company objection' raised by George Boolos (1987) and Michael Dummett (1991), that the fate of Frege's logicism will be decided.

See also: Logic in the 19th century; Logic in the early 20th century; Logical and mathematical terms, glossary of; Meaning and truth; Intuitionistic logic and antirealism

ALEXANDER GEORGE
RICHARD HECK

List of works

Frege, G. (1952) *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980.(The page numbers of the original publications appear in the text.)

Frege, G. (1984) *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell.(This includes translations of almost all of Frege's published papers; the page numbers of the original publications appear in the margins.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82.(Includes the first formulation of Frege's system of logic, and discussion of its importance for philosophical reflection on language and thought.)

Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, 2nd edn, 1980.(Includes criticism of then extant views about the nature of arithmetic, motivation for Frege's own proposal and informal proofs of the laws of arithmetic. One of Frege's most central texts.)

Frege, G. (1891) *Funktion und Begriff*, Jena: Pohle; trans. P.T. Geach, 'Function and Concept', in *Translations*

and *Collected Papers*.(Important paper in which Frege offers his analysis of concepts as functions whose values are restricted to the two truth-values.)

Frege, G. (1892a) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Meaning', in *Translations* and *Collected Papers*.(Frege's most famous and seminal paper in which he argues for the distinction between the sense and reference of expressions and presents a number of important theses regarding the two notions.)

Frege, G. (1892b) 'Über Begriff und Gegenstand', *Vierteljahrsschrift für wissenschaftliche Philosophie* 16: 192-205; trans. P.T. Geach, 'On Concept and Object', in *Translations* and *Collected Papers*.(Expands on what Frege takes to be the fundamental distinction between concepts and objects.)

Frege, G. (1892-5) 'Ausführungen über Sinn und Bedeutung', trans. P. Long and R. White, 'Comments on Sense and Meaning', in *Posthumous Writings*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Oxford: Blackwell, 1979, 118-25. (Expands some of the themes in Frege 1892a and discusses the importance of the sense/reference distinction in connection with predicates.)

Frege, G. (1893, 1903) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, Jena: Pohle, 2 vols; repr. as *Grundgesetze der Arithmetik*, Hildesheim: Olms, 1966; Part 1 of vol. 1 trans. M. Furth, *Basic Laws of Arithmetic: An Exposition of the System*, Berkeley, CA: University of California Press, 1964; no complete translation of vol. 2 exists, though fragments, trans. M. Black and P.T. Geach, appear as 'Frege on Definitions', 'Frege Against the Formalists' and 'Frege on Russell's Paradox', in *Translations*.(Gives a rigorous exposition of Frege's mature logic, with some philosophical discussion of its importance and an extended discussion of psychologism. Also includes formal proofs of the laws of arithmetic within this system, and the beginnings of a development of a theory of real numbers.)

Frege, G. (1894) Review of Edmund Husserl's *Philosophie der Arithmetik*, *Zeitschrift für Philosophie und philosophische Kritik* 103: 313-32; trans. H. Kaal, in *Collected Papers*.(One of the most important of Frege's discussions of 'psychologism'.)

Frege, G. (1897a) 'Logik', in *Nachgelassene Schriften und Wissenschaftlicher Briefwechsel (Posthumous Writings)*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Hamburg: Meiner, 1969-76.(A philosophical treatise on notions fundamental to logic, such as assertion, thought, sense and the like. A predecessor of Frege 1918.)

Frege, G. (1897b) 'Über die Begriffsschrift des Herrn Peano und meine eigene', *Berichte über die Verhandlungen der königlich Sächsischen Gesellschaften der Wissenschaften zu Leipzig, mathematisch-physische Klasse* 48: 362-8; trans. V. Dudman, 'On Mr. Peano's Conceptual Notation and My Own', in *Collected Papers*.(A comparison of Frege's logic with that of Peano, bringing out certain of the advantages of the former.)

Frege, G. (1903, 1906) 'Über die Grundlagen der Geometrie' (first and second series), *Jahresbericht der Deutschen Mathematiker Vereinigung* 12: 319-24, 368-75 and 15: 293-309, 377-403, 423-30; trans. E.-H. Kluge, 'The Foundations of Geometry: First Series' and 'The Foundations of Geometry: Second Series', in *Collected Papers*.(A critical discussion of David Hilbert's work on the foundations of geometry, frequently cited in connection with Frege's attitude towards early work in model theory.)

Frege, G. (1918) 'Der Gedanke: eine logische Untersuchung', *Beiträge zur Philosophie des deutschen Idealismus* 1: 58-77; trans. P.T. Geach and R. Stoothoff, 'Thoughts', in *Collected Papers*.(Late paper of Frege's in which he focuses on the nature of thoughts, that is, the senses of complete statements, and emphasizes their objectivity and their independence of psychology.)

Frege, G. (1969-76) *Nachgelassene Schriften und Wissenschaftlicher Briefwechsel*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Hamburg: Meiner; trans. *Posthumous Writings*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Oxford: Blackwell, 1979. (Translations of Frege's extant unpublished writings, most of which were destroyed during the bombing of Münster in the Second World War.)

References and further reading

Benacerraf, P. (1981) 'Frege: The Last Logicist', in P. French *et al*. (eds) *Midwest Studies in Philosophy* vol. 6, Minneapolis, MN: University of Minnesota Press, 17- 35; repr. in Demopoulos 1995, 41-67. (Mentioned in §7, this paper discusses the mathematical motivations for Frege's logicism and contrasts them with the epistemological motivations of the logical positivists.)

Boolos, G. (1987) 'The Consistency of Frege's Foundations of Arithmetic', in J.J. Thomson (ed.) *On Being and Saying: Essays in Honor of Richard Cartwright*, Cambridge, MA: MIT Press, 3-20; repr. in Demopoulos 1995, 211-33. (Raises the 'bad company objection' discussed in §10 and also includes a proof of the consistency of Hume's Principle.)

Cantor, G. (1897) *Beiträge zur Begründung der transfiniten Mengenlehre, 2er artikel*, Halle; trans. P. Jourdain, *Contributions to the Founding of the Theory of Transfinite Numbers*, London: Open Court, 1915; repr. New

York: Dover, 1955. (This is mentioned in §7.)

Dedekind, R. (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; trans. W. Beman, ‘The Nature and Meaning of Numbers’, in R. Dedekind, *Essays on the Theory of Numbers*, New York: Dover, 1963. (Referred to on a number of occasions, this is a classic in the philosophy of mathematics and a useful contrast to Frege’s approach.)

Demopoulos, W. (ed.) (1995) *Frege’s Philosophy of Mathematics*, Cambridge, MA: Harvard University Press. (This work collects most of the classic papers on the subject, not all of which are mentioned here.)

Dummett, M. (1973) *Frege: Philosophy of Language*, London: Duckworth, 2nd edn, 1992. (This is a long, difficult, but superb and seminal study of Frege’s views on language and logic which is relied upon heavily throughout the above.)

Dummett, M. (1978) ‘Frege’s Distinction Between Sense and Reference’, in *Truth and Other Enigmas*, London: Duckworth, 116-44. (This offers a detailed reconstruction of Frege’s arguments.)

Dummett, M. (1981) *The Interpretation of Frege’s Philosophy*, London: Duckworth. (This is an extended defence of Dummett’s exegesis of Frege together with a critical assessment of other interpretations.)

Dummett, M. (1991) *Frege: Philosophy of Mathematics*, London: Duckworth; repr. 1995. (Mentioned in §§8, 10, this raises the ‘bad company objection’; a sequel to *Frege: Philosophy of Language*, it is somewhat more accessible but perhaps not as definitive.)

Dummett, M. (1993) *The Seas of Language*, Oxford: Clarendon Press. (This is a collection of essays some of which first forcefully raised the issues discussed in §5.)

Evans, G. (1981) ‘Understanding Demonstratives’, in H. Parrett and J. Bouveresse (eds) *Meaning and Understanding*, Berlin: de Gruyter, 280-303; repr. in *Collected Papers*, Oxford: Clarendon Press, 1985, 291-321. (Puts forward an important interpretation of Frege’s views on sense and reference; mentioned in §3.)

Heck, R.G., Jr (1993) ‘The Development of Arithmetic in Frege’s *Grundgesetze der Arithmetik*’, *Journal of Symbolic Logic* 58: 579-601; repr. in Demopoulos 1995, 257-94. (This paper includes an extended, technical discussion of Frege’s formal work on arithmetic.)

Parsons, C. (1965) ‘Frege’s Theory of Number’, in M. Black (ed.) *Philosophy in America*, Ithaca, NY: Cornell University Press, 180-203; repr. in Demopoulos 1995, 182-210. (This is a classic, accessible paper on Frege’s philosophy of arithmetic, that raises many central issues.)

Ricketts, T. (1986) ‘Objectivity and Objecthood: Frege’s Metaphysics of Judgment’, in J. Hintikka and L. Haaparanta (eds) *Frege Synthesized*, Dordrecht: Reidel, 65-95. (This is a dense work that discusses the question, mentioned in §6, of whether Frege anticipated modern semantics.)

Tappenden, J. (1995) ‘Frege on Extending Knowledge and "Fruitful Concepts"’, *Nous* 29: 427-67. (Mentioned in §7, it discusses the relation between Frege’s work on arithmetic and developments in geometry with which he would have been familiar.)

Wilson, M. (1992) ‘Frege: The Royal Road from Geometry’, Nous 26: 149-80; repr. in Demopoulos 1995, 108-59. (Mentioned in §7, it presents an interpretation of Frege’s work relying heavily on connections with work in geometry.)

Wright, C. (1983) *Frege’s Conception of Numbers as Objects*, Aberdeen: Aberdeen University Press. (Mentioned in §9, it includes important discussions of logicism and the context principle and the first post-Fregean proof of Frege’s Theorem.)

Frei, Hans (1922-88)

Frei was concerned with two particular questions: why had the traditional Christian mode of biblical interpretation collapsed in the modern world, and how could it be recovered? He argued that, while the traditional interpretation of the Bible treated it as realistic narrative, in the modern period it had become accepted to treat the events described in it as myths and allegories. He argued for a more literal interpretation of the Gospels, which could lead to a more realistic interpretation of the person and nature of Jesus.

Hans Frei was an American theologian who fled his native Germany before the Second World War, and spent almost all of his teaching career at Yale University. He was of the view that, until the modern period, there had been near-consensus in the Church around the convictions that the Bible must be interpreted as one book, that the Gospels must be given priority in that unified interpretation, and that the Gospels as a whole must be interpreted literally. It was especially the collapse of consensus over the third element that drew Frei's attention. In the modern world, the Gospels are regularly interpreted as allegories, myths or moral tales.

Frei argued that there is something very strange about this. The Gospels are narratives. If one looks at the sort of narratives they are, it seems obvious that they are very different from paradigmatic examples of allegorical, mythical or moral narratives. The genre that fits them best is that of realistic narrative. Frei delineated various hallmarks of this genre, emphasizing especially that a realistic narrative, by tracing persons' actions and their reactions to features and events in their situation, renders the narrative identity of persons. It tells us who they are. The application of this point to the Gospels is that they render the narrative identity of Jesus.

One can summarize that identity in various ways: they narrate Jesus' enactment of his Messiahship, they narrate Jesus' enactment of the coming of God's kingdom, and so forth. What one cannot say, however, is that they are about such things as 'the coming of God's kingdom' or 'divine-human reconciliation'. To interpret them thus is to overlook their realistic narrative character and turn them into allegories or myths about some abstraction. The point - or, as Frei usually called it, the 'meaning' - of a realistic narrative is not something outside the narrative but the narrative itself, with its thematic core.

If, then, the literary genre of the Gospels is realistic narrative, why in the modern world have they been interpreted as allegories, as myths, as moral tales? Frei's answer is that interpreters confused 'meaning' with 'reference'. They assumed that the point of these narratives lies in something outside the narrative rather than in the narrative itself. They furthermore confused *history-likeness* in a narrative with the *historicity* of the narrative. Putting these points together: they assumed that the point of these narratives was to put us in touch with events in ancient history. But then scepticism set in as to whether the events really did occur. Yet most interpreters resisted concluding that these narratives lack 'meaning'. So they searched for aspects of reality other than ancient historical events which these narratives might be 'about'; and they construed the narratives as allegories of or myths about those aspects. Thus there occurred, to quote the title of Frei's best-known book, an 'eclipse of biblical narrative'.

Frei repudiated the assumption that meaning is reference in the case of the Gospels. The meaning of the Gospel narratives is what they say. The narrative interpretation that he called for finds literary criticism of far more use than historical studies. Frei's *The Eclipse of Biblical Narrative* (1974) is open to the antirealist interpretation that the important thing in the case of the Gospels is nothing but the story. When Frei's other major book, *The Identity of Jesus Christ* (1975), published almost simultaneously with *Eclipse*, is brought into the picture, it becomes clear that that was not his view. The Gospels - provided we interpret them as realistic narratives - present to us an actual historical person, Jesus of Nazareth, in his fundamental identity. They constitute the best access we have to that person; and, for religious purposes, an entirely adequate access. It is true that we do not have a proof of this. The situation is rather that some people, upon reading the Gospels, find themselves believing that they are in the 'presence' of Jesus of Nazareth, and are entitled to so believe.

It comes as no surprise to learn that after the publication of *Eclipse* and *Identity*, Frei began to emphasize the legitimacy and importance of what some have called 'Anselmian theology'; that is, theology that from the beginning is conducted as an exposition of Christian belief, in contrast to theology that develops natural or foundational theology before it sets about expounding Christian belief.

See also: Hermeneutics, biblical; Revelation §3

NICHOLAS P. WOLTERSTORFF

List of works

Frei, H. (1974) *The Eclipse of Biblical Narrative: A Study in Eighteenth and Nineteenth Century Hermeneutics*, New Haven, CT: Yale University Press.(Frei's classic study of biblical hermeneutics; densely written.)

Frei, H. (1975) *The Identity of Jesus Christ: The Hermeneutical Bases of Dogmatic Theology*, Philadelphia, PA: Fortress Press.(Considerably less difficult than *Eclipse*; corrects the impression of antirealism conveyed by *Eclipse*.)

Frei, H. (1992) *Types of Christian Theology*, New Haven, CT: Yale University Press. (Frei had been working for some years before his death on this typology of Christian theology.)

Frei, H. (1993) *Theology and Narrative*, ed. G. Hunsinger and W.C. Placher, New York: Oxford University Press. (A very helpful collection of Frei's most important articles, collected after his death by former students.)

References and further reading

Green, G. (ed.) (1987) *Scriptural Authority and Narrative Interpretation*, Philadelphia, PA: Fortress Press.(As yet, no full-length treatment of Frei's work has been published. However, a good many articles have appeared; this is a representative collection.)

French philosophy of science

A distinctively French tradition in the philosophy of science began with Descartes, continued through the Enlightenment in works such as D'Alembert's Discours préliminaire and the Encyclopédie, and flowered in the ninteenth and the early twentieth century with the work of Comte, Duhem, Meyerson and Poincaré. Throughout the twentieth century, the dominant fashions in French philosophy derived more and more from German influences, especially idealism and phenomenology (Hegel to Heidegger). But amidst these developments, there persisted an essentially autonomous tradition of French philosophy of science that offered an indigenous alternative to the Germanic imports. Here the key figure was Gaston Bachelard (1884-1962), for many years professor at the Sorbonne and director of the Institut d'Histoire des Sciences et des Techniques. His work was continued and modified by Georges Canguilhem (1904-95), his successor at the Institute, who himself was an important influence on philosophers such as Louis Althusser, Michel Foucault and Michel Serres. Jean Cavaillès' critique of Husserl's philosophy of mathematics and his effort to develop a neo-Hegelian alternative to it had deep affinities with Bachelard's work and was also an important influence on Canguilhem.

The most important general features of twentieth-century French philosophy of science appear if we contrast it with its two major rivals: existential phenomenology and logical positivism. Existential phenomenology is a 'philosophy of the subject', maintaining that ultimate truth resides in the immediacy of lived experience. Bachelard and his followers, by contrast, proposed a 'philosophy of the concept', for which experiential immediacy is subordinate to and corrected by concepts produced by rational reflection. This process of rational reflection is, moreover, embodied in science, which is not, as existential phenomenology maintains, a derivative and incomplete form of knowing, but the very paradigm of knowledge. In giving science a privileged epistemic position, the French philosophers of science are like the logical positivists. But, unlike the positivists, they treat science as essentially historical, irreducible in either method or content to the rigour of a formal system. They also opposed the positivists' effort to find the foundations of scientific knowledge in sense experience, maintaining that there are no simply given data and that all experience is informed by conceptual interpretation.

1 Bachelard's philosophy of science: epistemology and method

Bachelard accepts Descartes' view that knowledge derives from the methodical questioning of accepted beliefs (see Descartes, R. §4). But he is explicitly anti-Cartesian in rejecting the idea that, from this questioning, there emerge intuitive certainties that provide an unshakeable foundation for knowledge. Descartes' absolute certainty requires an entirely simple object (lest there be any hidden complexity misleading our judgment) and a subject that is entirely transparent to itself (lest we be deceived by unexamined prejudices or other distortions). But Bachelard maintains that such simplicity and transparency are not available; as the history of science amply shows, future generations always discover complexities and prejudices unnoticed by previous inquirers.

Scientific inquiry, therefore, never attains the entirely clear path Descartes sought but is always blocked, in some key respects, by *epistemological obstacles*. These are ways of thinking and perceiving, typically tied to common sense or past scientific theories, that prevent us from seeing truths essential for the further progress of science. Early chemistry, for example, had to overcome faulty intuitions based on the animism of primitive common sense; twentieth-century physics had to break through inadequate concepts of space, time and causality implicit in classical physics. Because of the need, at every stage, to overcome such obstacles, the development of science is not a continuous progress but a series of *epistemological breaks*, whereby scientists attain new conceptions of nature and of the methods required to study it.

In Bachelard's view the development of science is not continuous, but it is progressive. This is so in two respects: first, some scientific results are permanently valid and must be preserved in all subsequent formulations; second, successive scientific theories in a given domain provide a more general perspective from which the claims of previous theories can be assessed and, to the extent they are found valid, incorporated into current thinking.

Given his anti-Cartesian stance, the methodology of Bachelard's philosophy of science must be historical. The nature of scientific thinking cannot be determined from a priori principles but must be discerned from a close examination of actual scientific practice. However, within this historical approach, current science holds a privileged position as the best available account of nature. It alone is a basis for judging the degree of descriptive

and methodological success attained by previous scientific efforts. In Bachelard's terminology, contemporary science allows us to judge which parts of scientific history are valid (*l'histoire sanctionée*) and which outdated (*l'histoire périmée*). Thus, in contrast to ordinary political history, the history of science does not aim at a narration of objective facts that refrains from judging past events by contemporary standards. History of science essentially involves normative judgments of the past it studies.

As science develops, then, so too does the philosophy of science. (Bachelard in fact thinks that all domains of philosophy, not just the philosophy of science, develop in tandem with science.) He sees his own philosophy of science, in particular, as developing out the 'new scientific spirit' that gave rise to the relativistic and quantum revolutions at the beginning of the twentieth century. For example, Bachelard's notions of epistemological obstacle and epistemological break reflect the self-consciously revolutionary character of twentieth-century physics. Similarly, physicists' derivation of classical laws as limiting cases of the principles of relativity and quantum theory are the model for Bachelard's account of the way later theories contain the truth of earlier ones.

2 Bachelard's philosophy of science: metaphysics

The influence of current science on Bachelard's philosophy of science is particularly evident in his treatment of the reality of scientific entities. He vigorously opposes what he calls 'realism', by which he means the view that the middle-sized objects of ordinary sense experience are the fundamental realities of the world. According to Bachelard, the development of science has amply demonstrated that accounts based on such objects are entirely unequal to the task of explaining the facts of experience and must, therefore, be replaced by the more adequate ontology of theoretical science.

Although Bachelard thinks the explanatory power of science establishes the ontological priority of its theoretical postulations, he also maintains that a proper understanding of contemporary physics excludes metaphysical realism, which would see scientific entities as mind-independent substances. First, the metaphysics underlying the mathematical formalisms of the new physics are of mutually interdependent processes, not autonomous substances. More importantly, Bachelard maintains that the way in which contemporary physics constructs its objects reveals that they cannot be regarded as entirely independent of mental processes. Any particular scientific object (for example, Bohr's atom) is a construction of theoretical reason. Nonetheless, Bachelard does not accept the idealist claim that objects are simply the creations of mind. Any particular act of theoretical construction operates on an object pregiven to it.

Bachelard characterizes his position as an 'applied rationalism' that mediates between the metaphysical poles of realism and idealism. It is a 'rationalism' because it maintains the active role of the mind in the constitution of its objects and the priority of theoretical conceptualization over sense experience. The mind's activity, however, is not an autonomous creation but an application of its concepts to an object already given to it (through, of course, previous conceptualizations). Bachelard further separates himself from classical forms of idealism by giving scientific instruments an essential role in the constitution of scientific objects. Instruments are the material realization of theories, and such realization is necessary for the constitution of a scientific object. The mind constitutes objects only through the material intermediary of technical (instrumental) intervention in the world (see Scientific realism and antirealism §§3-4).

Bachelard's applied rationalism is very nicely illustrated by his account of how a modern scientist would undertake the investigation of the nature of wax suggested by Descartes in his *Second Meditation*. Descartes focuses on a piece of wax in its natural state, fresh from the hive, describes its appearance under various everyday circumstances (warmed by the sun, rolled between the fingers, and the like), and derives the essential nature of wax (as an extended body) from intellectual reflection on these descriptions. A modern scientist, by contrast, systematically transforms the 'natural' wax through experimental procedures that theory tells us will allow the observation of its characteristic features. It is purified, melted and methodically reshaped, X-rayed to obtain diffraction patterns, and so on. Its qualities appear not as simple givens to our intuition but as products of our rational experimental techniques (see Experiment; Observation).

Bachelard's philosophy of science obviously relies on a sharp distinction between the scientific and the nonscientific (see Demarcation problem). He sees the history of thought as a long struggle of scientific reason to overcome obstacles posed by nonscientific modes of thought and experience. For Bachelard, the fundamental

epistemological break is the one that separates science from the illusions of the senses and the imagination. Bachelard does not, however, deny all positive significance to the nonscientific realm of sense and imagination. He makes room for this realm in a series of studies on the 'poetics' of human thought. Here he introduces the project of a 'psychoanalysis' of reason that would uncover the unconscious significance of primitive images (such as those of earth, air, fire and water). He continues to deny that such images have any role in the objective description of what the world is like, resisting, as he says, 'the ontological temptation of beauty'. But he insists on the irreducible value of the subjective realm of poetic images as a necessary aesthetic complement to scientific knowing.

3 Canguilhem on concepts and theories

Georges Canguilhem succeeded Bachelard as director of the Institut d'Histoire des Sciences et des Techniques. Although he starts from an essentially Bachelardian view of science, the foci of his work are different from Bachelard's: philosophical history rather than historical philosophizing, the biological and medical sciences rather than physics and chemistry. Further, his results suggest a number of important modifications to Bachelard's position.

Canguilhem's most important methodological contribution is his distinction between concepts and theories. In much twentieth-century philosophy of science, concepts are functions of theories, deriving their meaning from the roles they play in theoretical accounts of phenomena (see Theories, scientific). Newtonian and Einsteinian mass, for example, are regarded as fundamentally different concepts because they are embedded in fundamentally different physical theories. This subordination of concept to theory derives from the view that the interpretation of phenomena (that is, their subsumption under a given set of concepts) is a matter of explaining them on the basis of a particular theoretical framework. For Canguilhem, by contrast, there is a crucial distinction between the interpretation of phenomena (via concepts) and their theoretical explanation. According to him, a given set of concepts provides the preliminary descriptions of a phenomenon that allow the formulation of questions about how to explain it. Different theories (all, however, formulated in terms of the same set of basic concepts) will provide competing answers to these questions. Galileo, for example, introduced a new conception of the motion of falling bodies to replace the Aristotelian concept. Galileo, Descartes and Newton all employed this new conception in their descriptions of the motion of falling bodies and in the theories they developed to explain this motion. Although the basic concept of motion was the same, the explanatory theories were very different. This shows, according to Canguilhem, the 'theoretical polyvalence' of concepts: their ability to function in the context of widely differing theories. His own historical studies (for example, of reflex movement) are typically histories of concepts that persist through a series of theoretical formulations.

Taken seriously, Canguilhem's emphasis on the history of concepts as opposed to the history of theories requires important modifications to Bachelard's view of science. Epistemological breaks, for example, must be construed as due to conceptual rather than theoretical innovation. Since successful conceptualizations tend to reappear in even quite diverse theories, epistemological breaks are, for Canguilhem, less frequent and, in many cases, less radical than Bachelard had suggested. The priority of concepts also requires us to rethink the notion of an epistemological obstacle. The same piece of scientific work may be an obstacle in terms of the theoretical context in which it is formulated and a creative breakthrough in terms of some of its conceptual content. Thus Black, even though working in the now outdated context of phlogiston theory, introduced the enduring concept of specific heat. The notion of an epistemological obstacle is more ambivalent than Bachelard suggests. Canguilhem makes particularly effective use of this ambivalence in his discussion of vitalism, so often abused as an enemy of progress in biology. Canguilhem admits that vitalistic theories have generally impeded the development of more adequate mechanistic accounts, but he maintains that the concept of vitalism, through its insistence on the uniqueness of biological phenomena, has served as a valuable protection against unfortunate reductionist tendencies of mechanistic theories (see Vitalism).

Canguilhem's refinement of the notions of epistemological breaks and obstacles also suggests a weakening of Bachelard's sharp distinction between science and nonscience. Science is what overcomes epistemological obstacles and effects epistemological breaks. To the extent that the notions of obstacle and break have become ambivalent, so has the notion of science. As a result, Canguilhem is reluctant to say more than that - in a given context - a given idea or approach is 'more scientific' than another (that is, more fully integrated into current experimental procedures). Further developing this line of thought, Canguilhem (influenced here by his students,

Althusser and Foucault) introduced the notion of *scientific ideology* as an intermediary between science and nonscience.

A scientific ideology (Herbert Spencer's philosophy of evolution is a good example) is scientific in the sense that it models itself on a successful scientific theory. It is ideological, however, because it makes claims about the world that go beyond what the science contemporary with it is able to establish; it has, in other words, pretensions that are not scientifically grounded. Such pretensions may very well function as obstacles to the development of science. But Canguilhem also sees a positive role for scientific ideologies: they provide an essential, if not entirely responsible, dimension of intellectual adventure, without which many scientific advances would not occur. Scientific ideologies are a prime example of the ambivalence of epistemological obstacles.

4 Rationality and norms

Like other antifoundational philosophers of science, Bachelard and Canguilhem must pay particular attention to the problem of rationality and objectivity. If there are no Cartesian certainties grounding science, if its development is a contingent historical process, what guarantee do we have that it is a reliable source of truths about the world? The historicist turn in Anglo-American philosophy of science, which is in many ways a later parallel to the French development, encountered these same questions.

Bachelard tried to ground the objectivity of science through social norms. Contrary to Descartes, he holds that objectivity is not found in the individual self's intuitions (which will always remain obstacles to scientific progress) but in a move to considerations that convince not just a given individual but all rational minds. This move, which Bachelard characterizes as from the isolated *cogito* (I think) to the social *cogitamus* (we think), takes us, he maintains, from the subjectivity of the merely psychological to the objectivity of the epistemological.

Canguilhem offers a much more extensive treatment of norms, rooted in his analysis of biological norms. He notes that, whereas modern physics has rejected any distinction between normal and pathological states of its entities, biological systems (organisms) require a distinction between states that enhance their functioning and those that impede it; in other words, a distinction between health and disease. However, Canguilhem maintains, we cannot define health as simply life in accord with the relevant biological norms. In any state, even one that is clearly pathological, there will be norms specifying the proper functioning of the organism in that state. (For example, a person who has lost a kidney is in a pathological state, even though the norms for proper functioning in this state are the same as for someone with both kidneys.) The pathological must, accordingly, be understood rather as a reduction in the range of circumstances in which an organism can function properly. Correspondingly, health is a state in which an organism is not only able to survive in its current circumstances but is capable of surviving in a significant range of varying circumstances.

Canguilhem emphasizes that, according to his account, biological norms are not objective in any scientific sense. Physiology can describe the states that we call 'normal' or 'pathological', but their normative status as such derives not from the physiological description but from the meaning of those states for the organism. Put another way, biological norms are subjective in the sense that they are constituted by the organism itself. On the other hand, this constitution is not a matter of individual idiosyncrasy but corresponds to the essential nature of the organism in question. Biological norms are not objective in the sense of being derived from value-neutral scientific inquiry, but they are rooted in the biological reality of the organisms that they regulate.

Turning to the question of social norms, such as those Bachelard sees as governing scientific practice, Canguilhem notes that there are important ways in which societies are similar to organisms and that social norms can have the same sort of necessary force that biological norms have. The biological analogy works, however, only for so-called 'traditional' societies, where there is a set of norms that defines, once and for all, the essential nature and purpose of the society. Modern societies have no such 'intrinsic finality', since the question of what should be their fundamental direction is contested in principle. A distinguishing feature of a modern society is dissent regarding basic norms. Canguilhem does not conclude, however, that a consensus, no matter how formed, would legitimately establish norms in a modern society. He criticizes, for example, Thomas Kuhn's account of scientific norms because, in his view, it derives them from a contingent, merely psychological, agreement that has no genuine regulative force. Canguilhem makes a similar criticism of Bachelard, who, he suggests, poses but does not solve the problem of finding a middle ground between grounding scientific norms in the illusion of Cartesian

foundations and reducing them to the merely descriptive realm of empirical psychology. Unfortunately, Canguilhem himself never provides a solution to this problem and, after all the subtlety of his analysis of norms, they are still left without a philosophical basis.

5 Michel Foucault: the influence of Bachelard and Canguilhem

It would be a serious mistake to read Michel Foucault's work as simply another stage of the historical approach to philosophy of science developed by Bachelard and Canguilhem. Foucault is too original a thinker to be confined to any school, and his work is informed by numerous currents far removed from and even opposed to Bachelard's applied rationalism orCanguilhem's history of concepts. But there was considerable influence on Foucault by both Bachelard and, especially, Canguilhem, who supervised his doctoral thesis on the history of madness. Because of this influence some aspects of Foucault's work must be understood in the context defined by these two predecessors.

Foucault accepted many of the key claims of Bachelard and Canguilhem's philosophy of science. These included the essentially historical character of science and the central role of epistemological breaks or discontinuities. On an even deeper level, Bachelard's notion of epistemological obstacle foreshadows Foucault's project of showing the historical, contingent nature of concepts and practices that present themselves as ahistorical necessities. Just as Bachelard spoke of a psychoanalysis of knowledge, Foucault characterized his own work as an effort to discover the unconscious of our knowledge.

Canguilhem's strongest influence on Foucault is through his history of concepts. Foucault employed this approach both as an alternative to what he saw as the excessive subjectivism ('transcendental narcissism') of phenomenology and as a way of de-emphasizing the role of individual 'great minds' in the history of science. *The Birth of the Clinic* (1963) and *The Order of Things* (1966) are particularly influenced by Canguilhem, offering, respectively, a history of the concept of disease in modern medicine and a history of the concept of man in the modern social sciences. The history of concepts is an essential tool of Foucault's famous archaeology of knowledge.

6 Michel Foucault: beyond Bachelard and Canguilhem

Foucault's major differences with Bachelard and Canguilhem can best be understood through two of his most important epistemological distinctions: that between *savoir* and *connaissance* and that between the threshold of scientificity and the threshold of epistemologization. By *connaissance* Foucault means, in accord with standard French usage, any given domain of knowledge, such as biology, economics or philology. *Savoir* he uses to refer to the underlying epistemic conditions that, in a given period, make various domains of *connaissance* (various scientific disciplines) possible. For example, eighteenth-century knowledge (*connaissance*) of natural history was, according to Foucault, based on a fundamental conception of natural entities as representable through tables of timeless genera and species. In the nineteenth century, he maintains, the new discipline (*connaissance*) of biology was, on the contrary, based on the fundamental concept of the organism as an essentially historical entity (see Taxonomy §1). Such fundamental concepts belong to what Foucault calls *savoir*. They correspond to the 'unconscious' of knowledge that his archaeological method is designed to excavate.

From Foucault's standpoint, Bachelard and Canguilhem treat science only at the level of *connaissance*. They take as givens the bodies of knowledge produced by scientific disciplines and remain unaware of the epistemic realm of *savoir* that underlies this knowledge. This prevents them from discerning the common ground between apparently diverse sciences (for example, the epistemic structure - in Foucault's terminology, *episteme* - shared by modern biology and economics). It also keeps hidden important connections between scientific and nonscientific forms of knowledge.

Foucault distinguishes (in *The Archaeology of Knowledge*, 1969) four thresholds in the development of a body of knowledge. At one extreme is the threshold of positivity (the initial emergence of an autonomous body of discourse) and at the other the threshold of formalization (formulation in axiomatic terms). Between these two are the thresholds of epistemologization and of scientificity. The threshold of epistemologization is crossed when a body of discourse (what Foucault calls a 'discursive formation') comes to be governed by epistemic norms that provide criteria (logical coherence, support from observed facts) for evaluating knowledge claims. The threshold

of scientificity is crossed when these norms are extended to include the specific principles of a scientific methodology. Such principles introduce a kind of rigour and precision far beyond that of ordinary discourse and define the special form taken by scientific objectivity.

Foucault's studies of science, in contrast to those of Bachelard and Canguilhem, deal with *savoir* rather than *connaissance* and operate on disciplines that have passed the threshold of epistemologization but not that of scientificity. The 'epistemological history' of Bachelard and Canguilhem begins with disciplines well-established as scientific and reflects only on the methodological and empirical content of their concepts and theories. As a result, Bachelard and Canguilhem simply accept the norms of current science as given and use them to evaluate the content of earlier science as well as the content of nonscientific discourses (for example, poetry or ideology) insofar as the latter implies a description of the natural world. By focusing on the *savoir* presupposed by scientific concepts and theories and on disciplines, such as psychiatry and the human sciences, that are not fully entrenched as scientific, Foucault is able to study scientific norms as the products of historical processes. Canguilhem (1967), reviewing *Les mots et les choses* (Foucault 1966), wrote that ' there is today no philosophy less normative than that of Foucault'. This is true in the sense that Foucault, unlike his two predecessors, writes the history of scientific (or would-be scientific) disciplines without presupposing the norms of these very disciplines. On the other hand, Foucault's archaeological analysis of science is fundamentally concerned with norms, in the sense of understanding their precise role in the historical development of scientific practices.

What might Foucault's approach contribute to the problem of the objectivity of scientific norms? His contribution is not so much a solution as a transformation of the problem. The issue, he suggests, is not how to provide a philosophical basis for objective norms. There is nothing outside of a given scientific community and its practices that can intelligibly underwrite its objectivity. A widely accepted, successful and enduring scientific enterprise has, of itself, *prima facie* status as a source of objective knowledge. But although scientific claims cannot, and need not, be grounded by philosophical argument, they can be undermined by the special sort of historico-philosophical inquiry that Foucault practised in his historical studies. Specifically, a careful untangling of the structure of the *savoir* underlying a given discipline and a tracing of the historical development of this structure can demonstrate the arbitrariness or contingency of principles that the discipline presents as objectively necessary. Foucault's studies of psychiatry (in *The History of Madness*), of criminology and related disciplines (in *Discipline and Punish*), and of the modern 'sciences' of sexuality (in the first volume of *The History of Sexuality*) are all attempts at such demonstrations.

Foucault's critiques are not intended as rejections of scientific objectivity as such. He is concerned with only specific disciplines and never, for example, subjects physics or chemistry to his critique. Moreover, his critiques never conclude that the discipline under scrutiny entirely lacks objective validity. Even the modern psychiatry that he denounces with such virulence in *The History of Madness* he acknowledges to have a certain validity. This eschewal of global scepticism is just the inverse of his rejection of foundationalism. Doubting everything and justifying everything are both totalizing epistemological enterprises entirely contrary to the spirit of Foucault's work (see Foundationalism; Objectivity).

7 Michel Serres

Michel Serres is even less a disciple of Bachelard and Canguilhem than Foucault. Nonetheless, there are important ways in which his work is a continuation and transformation of theirs. Like Bachelard, Serres emphasizes the dispersed, regional character of scientific work. Each domain is like a Leibnizian monad, with a life and intelligibility of its own. (The comparison to Leibniz is far from superficial. Serres wrote his thesis on the Leibnizian system, and his subsequent work can very fruitfully be read as a twentieth-century reformulation of Leibniz's philosophy.) But, like Foucault, Serres sees also a structural unity that connects independent scientific domains. He explicates this unity in terms of the concept of communication, which he employs as a formalized and secularized version of Leibniz's pre-established harmony (see Leibniz, G.W. §6; Unity of science). Serres also goes much further than Canguilhem and Foucault in questioning Bachelard's sharp demarcation of science from nonscience. He maintains that domains conventionally regarded as nonscientific, such as art and literature, share the structures of scientific disciplines and must be regarded as their epistemic peers. In this regard, Serres makes some particularly striking claims; for example, that Emile Zola discovered thermodynamics before it was explicitly formulated by physicists.

Serres's own writings are an often disconcerting - but also stimulating - mixture of the philosophical and the artistic. He tries to show the structural identity of Descartes' *Meditations* and La Fontaine's fables, argues that 'Turner translates Carnot', and presents Lucretius' *De Rerum Natura* as a contribution to twentieth-century physical theory. His style correspondingly combines close philosophical analysis with poetic evocation; casual readers may be in doubt as to whether they have picked up a philosophical treatise or a prose poem. Serres would, of course, maintain that such apparent incongruities simply reflect the limitations of conventional intellectual categories and that the pursuit of truth requires violating artificial disciplinary boundaries.

8 Conclusion

The history of French philosophy of science in the twentieth century is a history of the radicalization of Bachelard's insight that science and our philosophical reflection on it are historical phenomena. Bachelard's original formulations try to make this insight consistent with traditional views of the autonomy, progressiveness and normativity of scientific knowledge. Canguilhem, Foucault and Serres in turn raise fundamental questions about the viability of such views in the face of a full acceptance of the historicality of science. The issues emerging from this line of development parallel those familiar to Anglo-American philosophers of science in the years since Thomas Kuhn's *The Structure of Scientific Revolutions*. Like Bachelard, Kuhn (writing over twenty-five years later) tried to take seriously the historical nature of science while still maintaining traditional views about its epistemic status. Post-Kuhnian discussions, from Feyerabend through late twentieth-century sociology of science, have raised questions very similar to those posed by Bachelard's successors. Despite formidable obstacles, both rhetorical and conceptual, there is much that each of these traditions could learn from the other.

GARY GUTTING

References and further reading

Bachelard, G. (1934) *Le nouvel esprit scientifique*, Paris: Alcan; trans. A. Goldhammer, *The New Scientific Spirit*, Boston, MA: Beacon Press, 1984.(An important early statement of Bachelard's approach to the philosophy of science.)

Bachelard, G. (1938) *La psychoanalyse du feu*, Paris: Gallimard; trans. A.C.M. Ross, *The Psychoanalysis of Fire*, Boston, MA: Beacon Press, 1964.(A study of the poetic images at the origins of our thought about nature.)

Bachelard, G. (1949) *Le rationalisme appliqué*, Paris: Presses Universitaire de France.(An important late formulation of Bachelard's philosophy of 'applied rationalism'.)

Canguilhem, G. (1966) *Le normale et le pathologique*, Paris: PUF; trans. C. Fawcett, *The Normal and the Pathological*, New York: Zone Books, 1991.(A study of the concept of the normal in biology and medicine; the English translation also includes an important essay on Canguilhem by Michel Foucault.)

Canguilhem, G. (1967) 'Mort de l'homme ou épuisement du cogito?', *Critique* 24: 599-618; trans. C. Porter, 'The Death of Man, or Exhaustion of the Cogito?', in G. Gutting (ed.) *The Cambridge Companion to Foucault*, Cambridge: Cambridge University Press, 1994.(Canguilhem's review of Foucault's *Les mots et les choses (The Order of Things)*.)

Canguilhem, G. (1983) *Etudes d'histoire et de philosophie des sciences (Studies in History and Philosophy of Science)*, Paris: Vrin, 5th edn.(A collection of some of Canguilhem's most important historical and philosophical articles, including his interpretive essays on Bachelard.)

Foucault, M. (1963) *Naissance de la clinique*, Paris: Presses Universitaire de France; trans. A. Sheridan, *The Birth of the Clinic*, New York: Vintage, 1973.(A history of the concept of illness, written for a series edited by Canguilhem.)

Foucault, M. (1966) *Les mots et les choses*, Paris: Gallimard; trans. A. Sheridan, *The Order of Things*, New York: Random House, 1970.(A study of the origins of the fundamental concepts of the modern human sciences, from the Renaissance through the nineteenth century.)

Foucault, M. (1969) *L'archéologie du savoir*, Paris: Gallimard; trans. A. Sheridan, *The Archaeology of Knowledge*, New York: Pantheon, 1972.(A systematic reflection on the 'archaeological method' employed in Foucault's previous historical studies.)

Gutting, G. (1989) *Michel Foucault's Archaeology of Scientific Reason*, Cambridge: Cambridge University Press. (A study of Foucault's earlier work, with emphasis on his connections to Bachelard and Canguilhem; chapter one provides an overview of the work of Bachelard and Canguilhem.)

Lecourt, D. (1972) *Pour une critique de l'épistémologie (Bachelard, Canguilhem, Foucault)*, Paris: Maspero; trans. B. Brewster, *Marxism and Epistemology: Bachelard, Canguilhem, and Foucault*, New York: New Left

Books, 1975.(A critical discussion from a Marxist viewpoint.)

Saint-Sernin, B. (ed.) (1985) 'Canguilhem', *Revue de la métaphysique et de morale* 90.(A special issue of the journal with essays on various aspects of Canguilhem's work, along with a comprehensive bibliography.)

Serres, M. (1968) *Le système de Leibniz et ses modèles mathématiques (Leibniz's System and its Mathematical Models),* Paris: Presses Universitaire de France.(Serres's doctoral thesis on Leibniz.)

Serres, M. (1969-80) *Hermès I-V*, Paris: Editions de Minuit; trans. J.V. Harari and D.F. Bell, *Hermes: Literature, Science, Philosophy*, Baltimore, MD: Johns Hopkins University Press, 1982.(A series of collections of Serres's essays; the English translation selects some of the more important pieces from the five original volumes.)

Tiles, M. (1984) *Bachelard: Science and Objectivity*, Cambridge: Cambridge University Press.(An analysis of Bachelard's work, emphasizing its relevance to problems of Anglo-American philosophy of science.)

Freud, Sigmund (1856-1939)

Freud developed the theory and practice of psychoanalysis, one of the most influential schools of psychology and psychotherapy of the twentieth century. He established a relationship with his patients which maximized information relevant to the interpretation of their behaviour, and this enabled him to find explanations of dreams, symptoms and many other phenomena not previously related to desire. In consequence he was able radically to extend our common-sense psychology of motive.

On Freud's account everyday actions are determined by motives which are far more numerous and complex than people realize, or than common-sense understanding takes into account. The most basic and constant motives which influence our actions are unconscious, that is, difficult to acknowledge or avow. Such motives are residues of encounters with significant persons and situations from the past, often reaching back to early childhood; and they operate not to achieve realistic satisfaction, but rather to secure a form of pacification through representation. When we interpret what others say and do we apply these patterns of satisfaction and pacification to explain their behaviour; and in so far as we succeed in understanding others in this way we support the patterns as empirical generalization. While we recognize that pacification consequent on genuine satisfaction is deeper and more lasting than that effected by representation alone, we also know that human desire outruns opportunities for satisfaction to such an extent that pacification via imagination is common. This is a view which psychoanalysis radically extends.

This understanding of the mind enabled Freud to give psychological accounts of neurosis and psychosis, and to explicate how the past gives significance to the present in normal mental functioning. Past desires, even those of infancy, are not psychologically lost; rather they are continually re-articulated through symbolism, so as to direct action towards their representational pacification throughout life. In this Freud provides both a radically holistic account of the causation of action and a naturalistic description of the generation of meaning in life. New goals acquire significance as representatives of the unremembered objects of our earliest and most visceral passions; and the depth of satisfaction we feel in present accomplishments flows from their unacknowledged pacification of unknown desires from the distant past. Thus, paradoxically, significant desires can remain forever flexible, renewable and satisfiable in their expressions, precisely because they are immutable, frustrated and unrelenting at the root.

1 Early life and research

Freud was born in Freiburg, Moravia, the first son of the third wife of a travelling wool merchant. In 1861, the year after Austria abolished legal restrictions on Jews, the family settled in Vienna, where Freud remained until the Nazi occupation. As a student at the Medical School of the University of Vienna he attended lectures by the physiologist Ernst Bruke, which set out the Darwinian and physicalistic approach to nature associated with the school of Helmholtz. Shortly afterwards Freud began research in Bruke's laboratory, and within a year published the first of many articles on the nervous system.

On becoming engaged Freud found that he could not support a family through neurological research. He prepared for medical practice at the General Hospital of Vienna, working also in Meynert's Institute of Cerebral Anatomy. His publications on disorders of the nervous system led to an appointment as Lecturer in Neuropathology, and in 1885 he was awarded funds to study in Paris under Charcot, whose work on hysteria included the use of hypnosis for producing and removing symptoms. Here Freud made his first characteristic psychological observation, noting that the regions of the body liable to hysterical paralysis or anaesthesia did not correspond to real functional demarcations. Hysteria, it seemed, knew nothing of anatomy.

2 Collaboration with Breuer

Freud's psychoanalytic work was based on a discovery by his senior colleague Joseph Breuer. Breuer's patient Anna O, diagnosed as hysterical, was an exceptionally intelligent and articulate woman, who enquired with him into her symptoms in great detail. They found that each symptom had meaningful connections with significant but forgotten events, related to feelings which she had neither expressed nor mastered. When she recovered these her symptoms were eased. Thus one symptom was an aversion to drinking: despite 'tormenting thirst', she would

push away a glass of water 'like someone suffering from hydrophobia' (Freud 1953-74: vol. 2, 34). Under hypnosis she traced this to an episode in which she had remained silent while a companion let a dog, a 'horrid creature', drink water from a glass. After reliving this and expressing her disgust she drank without difficulty. Again, a range of symptoms - hallucinations, paralyses and disturbances of speech - were similarly related to experiences which had distressed her while she nursed her father in his terminal illness; when she went through these with Breuer the symptoms were relieved.

The research of Charcot and others had indicated that hysterical symptoms could be understood in psychological terms, and were sometimes related to emotional trauma. Although Breuer's patient relapsed before finally recovering, his observations suggested two further more specific hypotheses: first, that symptoms expressed memories and feelings associated with significant forgotten events; and second, that symptoms could be relieved by a cathartic therapy, which enabled patients to relive these events, thereby working through feelings connected with them. So Freud followed Breuer's example, and began questioning his patients about their lives and feelings in great detail. In this he abandoned hypnotism, and for a time substituted a technique, derived from Bernheim, of pressing patients to remember significant events. His experience corroborated Breuer's sufficiently for them to publish a series of case reports arguing that 'hysterics suffer mainly from [unconscious] reminiscences' (Freud 1953-74: vol. 2, 7).

3 Investigation of sexual abuse

As his research progressed, Freud found that an important range of the memories connected with symptoms were sexual, and went back in time. Under the pressure of his technique, indeed, a number of his patients recovered apparent memories of sexual abuse from early childhood. As he pressed further, in a series of female patients, the role of abuser was consistently assigned to the father. (This experience has been repeated by a number of therapists in the USA.) In considering these scenes of seduction, however, Freud finally concluded that while parental abuse could be a factor in causing neurosis, it was not as widespread as suggested by the readiness of patients to recollect it. For, as he noted, there seemed to be 'no indications of reality in the unconscious', so that 'one cannot distinguish between truth and [emotionally charged] fiction', where this could include 'sexual fantasy [which] seizes on the theme of the parents' (Freud 1953-74: vol. 1, 260; also see Masson 1985: 265). Hence, it seemed, the primary cause of symptoms might be found in activity of the imagination.

These events also led Freud to change his therapeutic technique. It had became apparent both that he had to guard against effects of suggestion, and also that the most relevant and reliable material tended to emerge not when patients were pressed for memories, but rather when they expressed their thoughts and feelings spontaneously. So he began asking his patients simply to communicate each idea or thought which occurred to them as fully as possible, and without regard to whether it seemed significant, sensible or morally acceptable. This kind of immediate and unconstrained self-description, which Freud called free association, led both to the topics previously found important by questioning, and to others not yet investigated. Since this procedure maximized the data relevant to understanding his patients, Freud made it the fundamental rule of psychoanalytic treatment (see Unconscious mental states).

4 Dreams and wish-fulfilment

During this period Freud observed that dreams, like symptoms, could be understood as related to memories and motives which emerged in free association. In investigating dreams, moreover, he could make use of his own case. He thus began his self-analysis, applying to himself the interpretive techniques which he applied to his patients. As this progressed, Freud saw that his and Breuer's findings about symptoms were better understood on the model that he was developing for dreams. In consequence he was able to frame an account of both dreams and symptoms which was relatively simple and unified, and which he published in *The Interpretation of Dreams* (1900). Further, as he soon saw, this account could be extended to other products of the mind, including slips and errors, jokes and works of art.

Freud's discussion of dreams thus constituted a paradigm by which he consolidated his first phase of psychoanalytic research. We can see some main features of this by considering the first dream he analysed, his own dream of Irma's injection. In this dream Freud met Irma, a family friend whose pains he had diagnosed as hysterical and treated by analysis. He told her that if she still felt pains, this was her own fault, for not accepting

his ‘solution’ to her problems. He became alarmed, however, that he had failed to diagnose an organic cause of her suffering, and this turned out to be so. A senior colleague, M, examined Irma, and found that she was organically ill. The cause of her illness also became manifest: another of Freud’s colleagues, his family doctor Otto, had given her a toxic injection. As the dream ended Freud censured Otto firmly, saying, ‘Injections of that kind ought not to be made so thoughtlessly’ and ‘probably the syringe had not been clean’.

On the surface this dream dealt with topics which were unpleasant to Freud, such as the continued suffering of a friend and patient, and the possibility that he had misdiagnosed an organic illness, which he described as ‘a constant anxiety’ to a doctor offering psychological therapy. Freud’s associations, however, made clear that the dream represented these things in a way which was actually in accord with various of his desires or wishes. For Freud remembered that on the day before the dream he had discussed Irma with Otto, who had recently visited Irma’s family. Otto had been called away to give an injection while at Irma’s, and had told Freud that Irma was looking ‘better, but not yet well’. Freud imagined that he detected a professional reproof in this remark, considered it thoughtless, and felt vaguely annoyed. That night, in order to justify his treatment of Irma, he had started to write up her case to show to M, who was respected by both himself and Otto, and who appeared in the dream as diagnosing Irma’s illness and realizing that it was Otto’s fault. (Also, as it happened, Freud had just had news indicating that another of his female patients had been given a careless injection by some other doctor, and had been contemplating how safe his own practice with injections was.)

In considering this dream Freud noted that his desire to justify himself in respect of Irma’s case - and in particular not to be responsible for her continued suffering - was apparent from the beginning, in which he told Irma that her pains were now her own fault. Also, he felt that his alarm at her illness in the dream was not entirely genuine. So, as he realized, it seemed that the dream-situation, in which Irma was organically ill, actually served to fulfil a wish on his account: for as he undertook to treat only psychological complaints, this would mean that despite what he had taken Otto thoughtlessly to imply, he and his mode of treatment could not be held responsible for Irma’s continued illness. This theme, moreover, seemed carried further in the rest of the dream, in which M found that Otto, not Freud, bore responsibility for Irma’s suffering. It thus emerged that the dream could be seen as a deeply wishful response to Otto’s remark. For as the dream represented matters, Freud bore no responsibility whatever for Irma’s condition. Rather, Otto was the sole cause of her suffering; and this was the result of Otto’s thoughtless practice with injections, a matter about which Freud himself, as he had been considering, was particularly conscientious and careful. The dream thus repaid what Freud felt to be Otto’s professional reproof with a charge of serious malpractice on Otto’s part. So Freud concluded that the dream was a wish-fulfilment, that is, that it was caused by certain of his desires (or irresponsible wishes derived from these desires) and expressed these by representing them as fulfilled.

We can bring out the nature of Freud’s proposal here by contrasting the way his desire to be clear of responsibility for Irma’s suffering operated in his dream as opposed to his intentional action. Schematically, the role of a desire that P in intentional action is to bring about (cause) a situation that P, which both satisfies the desire and pacifies it, that is, causes the desire to cease to operate. Action on a desire that P (that one be cleared of responsibility) should ideally satisfy the desire, that is, should bring it about that P (that one is cleared of responsibility). This, in turn, should cause one to experience and believe that P (that one has been cleared of responsibility); and this, perhaps acting together with the satisfying situation, should pacify the desire that P, so that it ceases to govern action. This is roughly the sequence of results which Freud was seeking to produce, in accord with standard medical practice, in writing up Irma’s case history on the night of the dream to discuss with M, his respected senior colleague. M would be able to offer an independent and authoritative opinion on Freud’s treatment of Irma; so his judgment could partly serve to put Freud, and his mode of therapy, in the clear.

In Freud’s dream the same motive was apparently also at work, but in a quite different way. There it produced no rational action, but rather gave rise to a (dreamt) representation of a situation in which Freud was cleared of responsibility, and by M. This representation, moreover, was extravagantly wishful - Irma was made physically ill, Freud was cleared in a great number of ways, Otto was elaborately blamed, and so on. So we can contrast the causal role of desire, as between intentional action and what Freud called wish-fulfilment, as follows. In intentional action a desire that P serves to bring about a situation that P, and this a (justified and true) belief that P, so that the desire is pacified. In Freudian wish-fulfilment, by contrast, the desire that P causes a wishful representation that P which, although perhaps exaggerated and unrealistic, is experience- or belief-*like*, so that it

serves to pacify the desire directly, at least temporarily. So in rational action we find both the real satisfaction and also the pacification of desire, with the latter a rational and causal consequence of the former; whereas in wish-fulfilment we find only imaginary satisfaction, and the pacification which is consequent on this.

It seems intelligible that the mind (or brain) should operate in accord with both these patterns, for in both desire is ultimately pacified by the representation of satisfaction. On reflection, indeed, the latter pattern seems almost as familiar as the former. We are aware that our response to a desire or lack is often simply to imagine it satisfied; and this process seems clearly to have other forms. Children, for example, often pacify desires which arise from their being small and immature by representing themselves as various sorts of admired or impressive figures in play. Also it seems clear that many books, films, video games and so on enable people to imagine gratifying desires which they could not (and often would not) otherwise satisfy. In such cases we speak of make-believe, suspension of disbelief and virtual reality, thereby indicating the way the kind of representation involved is comparable to experience or belief in serving to pacify desire (see Action; Desire).

5 Transference, childhood conflict and sexuality

Freud found that the motives expressed in dreams and symptoms could be retraced to the past. Also, he noted that in the course of analysis his patients were liable to experience towards himself versions of motives and feelings which they had felt towards earlier objects of love and hate, in particular their parents. Freud called this phenomenon ‘transference’. Thus a main symptom of his patient the Rat Man was involuntarily to imagine his father being punished, by being eaten into from behind by hungry rats. This caused him anxiety and depression, from which he protected himself by the rituals of an obsessional neurosis. He insisted that this symptom did not show hostility towards his father, with whom he had been the best of friends. He could not remember his father having punished him; but as the treatment progressed, he developed a fear that Freud would punish him - would beat him and throw him out because of the dreadful things he said in his free associations.

In discussing these matters with Freud the patient went down to the end of the room, saying that he did so from delicacy of feeling for Freud, to whom he was saying such dreadful things. Freud interpreted, however, that he had moved away because, as he had previously been saying, he was afraid that Freud would beat him. To this he responded like ‘a man in desperation and one who was trying to save himself from blows of desperate violence; he buried his head in his hands, rushed away, covered his face with his arms, etc.’ (Freud 1953-74: vol. 10, 284). He was reliving an episode from his childhood, in which, while lying between his parents in bed, he had urinated, and his father had beaten him and thrown him out. He thus recovered a buried image of his father as a terrifying, punishing figure in relation to whom he felt helpless and hostile. This in turn made it possible to understand his involuntary imaginings, in which his father suffered such cruel punishments, as wish-fulfilments fitting the same pattern as Freud’s dream above. Just as Freud’s dream could be seen as a wishful and extravagant reversal of the feelings of responsibility prompted by Otto’s remark, so this patient’s symptom of imagining his father cruelly punished could be seen as a wishful and extravagant reversal of his own feelings as a punished child. But these childhood feelings were re-experienced towards Freud in the transference before being consciously remembered as relating to his father.

The childhood motives revealed by analysis characteristically included sensual love for one parent combined with rivalry and jealous hatred for the other, a constellation Freud called the ‘Oedipus complex’. It now appeared that little children were liable to intense psychical conflict, as between desires to harm or displace each parent, envied and hated as a rival for the love of the other, and desires to preserve and protect that same parent, loved sensually and also as a caretaker, helper and model. Further, children apparently attached great emotional significance to their interactions with their parents in such basics of disciplined and cooperative activity as feeding and the expulsion and management of waste. These activities also involved the first use, and hence the first stimulation, of bodily organs or zones - particularly the mouth, genitals and anus - which would later figure in the emotionally significant activities of normal and abnormal sexuality. Analysis indicated that the feelings related to the early uses of these organs had significant continuity with those aroused by their later uses; and on this basis Freud framed an account which systematically linked normal and abnormal sexual phenomena in the development of the individual.

In normal development, weaker and less acceptable childhood feelings were subjected to a process of repression, which removed them from conscious thinking and planning, so that they could achieve pacification only indirectly. Apart from causing dreams or symptoms, such motives could undergo a process of sublimation, whereby they

provided symbolic significance for everyday activities, and could thus be pacified in the course of them. This achievement can be illustrated by a relatively successful teacher and writer, who had been surprised when one of his pupils - who had made a special effort to be taught by him, and was trying hard to master his ideas - had unexpectedly offered to suck his penis. He had not accepted this offer, but that night dreamed that a lamb had come to suck milk from his finger. On waking he realized that the lamb represented the pupil who had come to imbibe his ideas, and his milk-giving finger the penis his pupil had wanted to suck. So the dream could be seen as representing the fulfilment, in a more acceptable and symbolic form, of a sexual wish which had arisen on the day before.

The symbolism, however, went deeper, for the dreamer also represented himself as a mother nursing a child. In this he represented his finger/penis in the role of a feeding breast, and his writing and teaching as the production of milk as well as semen. He thus represented himself as enjoying a combination of feminine nurture and masculine potency which was impossible in real life, and his desires for which had been repressed in early childhood. Still, these same desires could to some degree be pacified in his adult work, owing to the symbolic significance which he attached to it. In writing or teaching he could with some justice see himself - to use more familiar metaphors - as potent and seminal, and at the same time as giving others food for thought.

6 Fantasy and primary process

Psychoanalysts now commonly describe the kind of imaginative representational activity which serves to pacify unconscious desire as fantasy. Fantasies can be shown not only in dreams or symptoms, but any form of activity which has representational significance; and fantasies characteristically realize the psychological mechanisms and processes described in psychoanalytic theory. Thus the fantasy of Freud’s patient above - that Freud was about to give him a beating - simultaneously implemented his transference onto Freud of emotions originally felt towards his father; the repression of his memory of an episode in which his father had punished and terrified him; and the projection into Freud of his own desire to punish, which was apparently related to the episode with his father, and otherwise expressed in his symptom of imagining his father punished. Again, persons form permanent life-shaping fantasies of themselves on the model of others, thereby establishing identifications with those persons, which process Freud took to be central to the constitution of the self.

Freud held that the pattern of wish-fulfilling representation, in which a desire that P produces a pacifying (and perhaps symbolic) representation that P, marks a primary process in the pacification of desire, whereas the pattern of intentional action, in which a desire that P gives rise to a real action or situation that P, marks a secondary process, to be seen both as developing from the first and occurring in the context of it. We can understand this as the claim that present action characteristically stands in a representational and pacificatory relation to desires from the past. Thus Freud’s patient above consciously desired to go to the end of the room partly in response to his unconscious fantasy that Freud, representing his father, was about to beat him; and the author who dreamed of a sucking lamb consciously desired to teach and write partly because these activities unconsciously realized fantasies of taking the role of a feeding mother. That is to say, a person acting in the present may thereby be representing himself as satisfying, and so may thereby now actually be pacifying, desires which have remained unsatisfied from early childhood.

7 Psychology and the brain

Freud’s initial research was on the nervous system, and in his *Project for a Scientific Psychology* (1895) he sketched a version of physicalism - an account of psychological processes as neural processes - which could accommodate his psychological findings. He took the brain to operate by transferring excitation through networks of neurons, so that information was stored in the form of facilitations or inhibitions of the interneural links. Hence, as he put it, ‘psychic acquisition generally’, including memory, would be ‘represented by the differences in the facilitations’ of neural connections (Freud 1953-74: vol. 1, 300). In this he anticipated the contemporary conception of the brain as a computational device whose knowledge is ‘in the connections’ among neuronal processing units, and also the associated view of mental processes as forms of neural activation, and mental states as dispositions to these, or structures determining them (see Glymour 1992; Connectionism).

In Freud’s model the signalling of a bodily need - say, for nutrition in an infant - causes a disequilibrium in neural excitation. This at first results in crying and uncoordinated bodily movements, which have at best a fleeting

tendency to stabilize it. Better and more lasting equilibration requires satisfaction, for instance, by feeding; and this causes the facilitation of the neural connections involved in the satisfying events. The brain thus constantly lays down neural records, or prototypes, of the sequences of perceptions, internal changes, bodily movements and so on, involved in the restoration of equilibrium by satisfaction. Then when disequilibrium again occurs - such as when the infant is again hungry - the input signals engage previously facilitated pathways, so that the records of the best past attempts at coping with comparable situations are naturally reactivated. This, Freud hypothesized, constitutes early wish-fulfilment.

Freud thus identified the wish-fulfilling pacification of infantile proto-desire with a form of neural prototype activation (see Churchland 1995). He took it that this served to stabilize and organize the infant's responses to need, by reproducing those previously associated with satisfaction. Then, as the infant continued to lay down prototype upon prototype, the original wishful stabilizations evolved towards a system of thought, while also coming to govern a growing range of behaviour, increasingly coordinated to the securing of satisfaction. This, however, required the brain to learn to delay the activation of prototypes of past satisfactions until present circumstances were perceptually appropriate - that is, to come into accord with what Freud called the reality principle. This in turn depended upon a tolerance of frustration, and the absence of a satisfying object, which permitted reality-testing, and hence the binding of the neural connections involved in the securing of satisfaction to perceptual information about the object, and later to rational thought. This benign development could, however, be blighted, if frustration (or intolerance of it) led to the overactivation of inappropriate prototypes, and this to greater frustration. Such a process could render the mind or brain increasingly vulnerable to disequilibrium and delusion, and hence increasingly reliant on wish-fulfilling modes of stabilization, in a vicious circle constitutive of mental illness.

8 Psychic structure and function

Freud allocated the task of fostering the sense of reality to a hypothetical neural structure, or functional part of the mind, which he called 'das Ich', or the *ego*. (The literal meaning of Freud's phrase is 'the I'; but the Latin pronoun used in the English translation has now acquired a life of its own.) In his final conceptualizations he linked this structure with two others, the 'super-ego', which judged or criticized the ego, and which included the ego-ideal, representing the ideals or standards by which the ego was judged; and the primitive 'it', or *id*, the natural matrix of basic and potentially conflicting instincts or drives - that is, structures which would yield basic emotions and motives for action - out of which these others developed. His late discussions of these notions are particularly difficult, partly because they combine differing modes of explanation.

Overall the ego, super-ego and id are neural systems described in a functional way, that is, in terms of the goals which their operation secures and the information upon which they operate. This kind of explanation has been refined in contemporary cognitive science, in which distinct functional units are often represented by boxes, in a flow chart which describes the contribution of each boxed unit to psychological functioning overall. In his later work Freud sought to combine this mode of explanation with the empirical claim that the main functional systems of the human mind are partly constituted by the mind's internalized representations of significant persons in the environment, particularly the parents. Hence the working of these systems is partly felt, and can partly be described, via the motives, feelings or actions of the imagos of persons which they embody (see Functionalism).

Freud took the drives constituting the id as divisible into two main categories: those which engender motives which are creative and constructive, such as affection, love and care, which he called the life instincts; and those which yield motives linked to aggression, such as envy and hate, which he called the destructive or death instincts. The sexual drives (or motives), together with those aimed at self-preservation, were among the life instincts generally; but owing to their great plasticity they were liable to be mixed with aggression, as in the case of sadism, masochism, devaluation of the object of love, and so on. His final view was thus that the primary conflicts in a person's life - those which necessitated repression and could become constitutive of mental illness - might involve sexuality, but were ultimately to be seen as holding between impulses to create or destroy.

On Freud's account the ego and super-ego develop out of the id, mainly through the child's formative identifications with others, particularly the parents. The child ordinarily begins to advance towards self-control by laying down prototypical images of the parents, in their role as regulators of socially significant primitive bodily activities, particularly, as noted, those involved in feeding and the elimination of waste. These 'earliest parental

imagos' (Freud 1953-74: vol. 22, 54) provide the basis of the super-ego. This self-critical faculty embodies the child's aggression in a projected form, and so tends to be far more punitive than the actual parents. Hence it can be a source of great anxiety or guilt, or even, in the extreme, of suicide. (Compare the material in §5 above, which might also be described as involving projection of the patient's punitive super-ego.)

The child further constitutes its ego by identifying with the parents as agents, that is, as desirers and satisfiers of desire. A main step in normal development is identification with the parent of the same sex, which has the consequence that sexual (and other) desires are rendered non-incestuous, heterosexual and reproductive. For this to occur, however, the child must renounce the goal of replacing the envied parent with that of becoming like him or her. Hence the final establishment of the ego and super-ego coincides with what Freud called the dissolution of the Oedipus complex.

9 Social psychology

Freud held that the cohesiveness of many groups results from their members identifying with one another by putting a common idealized figure (or cause or creed) in the place of the ego-ideal. Groups may likewise be identified by the projection of their bad aspects - and in particular their own hostile and destructive motives - into some common locus, which thereby becomes a focus of collective and legitimated hate. This may be a particular figure, such as the devil, or another family, race or nation, the leaders of which may also be denigrated or demonized.

People who identify themselves with the same idealized object, or who represent a common object as hateful, can thereby feel purified, unified and justified in cooperating in destructive activities validated by common ideals. Identification by the projection of good or bad aspects into a common locus thus serves systematically to organize persons into groups which represent themselves as unrealistically good and others as unrealistically bad; and this in turn allows aggressive and destructive motives to be entertained and satisfied with a minimum of indirectness and guilt. This pattern of good us/bad them is shared by most participants in the great variety of human conflicts which it encourages. Where groups are organized on this basis their rational disagreements tend to be underpinned and exaggerated by suspicions and hatreds which can be seen to be irrational, but which remain none the less intractable, in so far as they are integral to group cohesion.

See also: Jung, C.G.; Nietzsche, F.; Psychoanalysis, methodological issues in; Psychoanalysis, post-Freudian

JAMES HOPKINS

List of works

Freud, S. (1953-74) *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, trans. and ed. J. Strachey *et al*., London: Hogarth Press, 24 vols.(The standard English edition of Freud's works.)

Freud, S. (1979) *The Pelican Freud Library*, Harmondsworth and New York: Penguin.(A slightly abridged version, this omits Freud's original case notes on the Rat Man, from which the description of treatment in §5 is taken.)

Freud, S. (1895) *Project for a Scientific Psychology*, in Freud 1953-74: vol. 1, 283-.(Freud's physiological hypotheses about the mind, sketched but not published while he was formulating his account of dreams. He took mental processes to be realized by the passing of excitation through networks of neurons, which also altered neural connections so as to store information for later use. The mind/brain thus naturally laid down neural records of its activities relating to the satisfaction of desire, and these prototypes were automatically reactivated so as to guide action, and receive further modification, in comparable situations later. Freud expressed a number of his basic ideas in these terms, and his later discussions are consistent with them.)

Freud, S. (1900) *The Interpretation of Dreams*, in Freud 1953-74, vols 4-5.(Freud's account of dreams and symptoms as wish-fulfilments, that is, as serving to pacify desires by representing them as satisfied. Freud later applied this paradigm to other kinds of representation, and extended it to everyday intentional action through his account of the sublimation of infantile sexual and aggressive desires. These hypotheses explain present desires by mapping them to objects and situations desired in the past, and hence in accord with the role assigned past neural prototypes in the Project.)

References and further reading

Cavell, M. (1993) *The Psychoanalytic Mind*, Cambridge, MA: Harvard University Press.(Relates Freud's work to

Donald Davidson's philosophy of mind.)
Churchland, P. (1995) *The Engine of Reason, The Seat of the Soul*, Cambridge, MA, and London: MIT Press. (Enthusiastic account of the connectionist understanding of the brain foreshadowed by Freud's *Project* of 1896; does not mention Freud, hence usefully consulted together with Gill and Pribram 1976, and Glymour 1992.)
Clark, P. and Wright, C. (eds) (1988) *Mind, Psychoanalysis and Science*, Oxford: Blackwell.(Essays by philosophers and psychologists, including summary and discussion of Grunbaum 1984.)
Erwin, E. (1996) *A Final Accounting: Philosophical and Empirical Issues in Freudian Psychology*, Cambridge, MA, and London: MIT Press.(Critique which follows Grunbaum in arguing against the role assigned to clinical data in supporting psychoanalytic theory.)
Gardner, S. (1993) *Irrationality and the Philosophy of Psychoanalysis*, Cambridge: Cambridge University Press. (Discusses psychoanalysis and the explanation of irrationality.)
Gay, P. (1988) *Freud, A Life for Our Time*, London and Melbourne: J.M. Dent.(Biography which contains accounts of Freud's main ideas and a comprehensive Bibliographical Essay.)
Gill, M. and Pribram, K. (1976) *Freud's 'Project' Re-assessed*, London: Hutchinson.(Early appreciation of the contemporary relevance of Freud's understanding of the brain.)
Glymour, C. (1992) 'Freud's Androids', in J. Neu (ed.) *The Cambridge Companion to Freud*, Cambridge: Cambridge University Press.(Sketches relation of Freud to recent work in computational cognitive science.)
Grunbaum, A. (1984) *The Foundations of Psychoanalysis: A Philosophical Critique*, University of California Press.(Most influential of recent empiricist critiques of Freud's work, arguing against the role assigned by Freud and his successors to clinical data in support of psychoanalytic theory. Summarized and discussed, with criticisms, in Clark and Wright 1988; carried forward in Erwin 1996.)
Grunbaum, A. (1993) *Validation in the Clinical Theory of Psychoanalysis*, Madison, CT: International Universities Press.(Further discussions of themes raised in Grunbaum 1984.)
Hopkins, J. and Savile, A. (eds) (1992) *Psychoanalysis, Mind, and Art: Perspectives on Richard Wollheim*, Oxford: Blackwell.(Includes philosophical essays explicating and making use of psychoanalytic concepts.)
Kitcher, P. (1992) *Freud's Dream: A Complete Interdisciplinary Science of Mind*, Cambridge, MA, and London: MIT Press.(Discusses Freud's systematic psychology in comparison to contemporary cognitive science.)
Kline, P. (1984) *Psychology and Freudian Theory: An Introduction*, London: Methuen.(Clear introduction to the literature on empirical testing of Freudian concepts and theories.)
Laplanche, J. and Pontalis, J.B. (1973) *The Language of Psychoanalysis*, London: Hogarth Press.(Basic explanations of psychoanalytic concepts.)
Lear, J. (1990) *Love and Its Place in Nature*, New York: Farrar, Strauss & Giroux.(Philosophical discussion centred on the concept of love in psychoanalysis.)
Masson, J. (ed.) (1985) *The Complete Letters of Sigmund Freud to Wilhelm Fliess 1887-1904*, Boston, MA: Harvard University Press.(Personal letters in which Freud describes his developing ideas, including his self-analysis.)
MacDonald, C. and MacDonald, G. (eds) (1995) *Philosophy of Psychology: Debates on Psychological Explanation*, Oxford: Blackwell.(Discussions of psychoanalysis in the context of other debates in the philosophy of psychology.)
Neu, J. (ed.) (1992) *The Cambridge Companion to Freud*, Cambridge: Cambridge University Press.(Philosophical essays on Freud and psychoanalysis.)
Wollheim, R. (1984) *The Thread of Life*, Cambridge, MA: Harvard University Press.(Philosophical and psychoanalytic approach to the concept of the life of a person.)
Wollheim, R. (1991) *Freud*, London: Fontana, 2nd edn.(Lucid and philosophically informed account of Freud's work).
Wollheim, R. (1993) *The Mind and Its Depths*, Cambridge, MA: Harvard University Press.(Psychoanalytically informed philosophical essays on the mind.)
Wollheim, R. and Hopkins, J. (1982) *Philosophical Essays on Freud*, Cambridge: Cambridge University Press. (Essays discussing psychoanalytic concepts and modes of explanation.)

Friendship

Philosophical interest in friendship has revived after a long eclipse. This is due largely to a renewed interest in ancient moral philosophy, in the role of emotion in morality, and in the ethical dimensions of personal relations in general. Questions about friendship are concerned with issues such as whether it is only an instrumental value (a means to other values), or also an intrinsic value - a value in its own right; whether it is a mark of psychological and moral self-sufficiency, or rather of deficiency; and how friendship-love differs from the unconditional love of agapē*. Other issues at stake include how - if at all - friendship is related to justice; whether the particularist, partialist perspective of friendship can be reconciled with the universalist, impartialist perspective of morality; and whether friendship is morally neutral.*

1 Ancient philosophy

Ancient moral philosophy devoted considerable attention to understanding friendship (*philia*) and passionate love (*eros*) (see Love §§2-3). In its widest sense, *philia* covers both amicable relations among casual acquaintances, and intimate, loving relationships within and beyond the family. Plato's *Lysis* is the first serious philosophical treatment of friendship, and sets the stage for subsequent treatments. The dialogue is largely inconclusive, but some of the claims made in it re-emerge in Plato's later dialogues: we love people, like other objects of love, because of some lack in ourselves, and we love them only in so far as they are useful in pursuing our final end or *eudaimonia*, achieved with knowledge of Goodness (see Plato; Eudaimonia). Thus, the notion of friendship as an enduring and reciprocal love between equals, and as a mark of self-sufficiency rather than deficiency, remains absent from Plato's account.

These lacunae are made good in Aristotle's writings. A flourishing (*eudaimōn*) human life, a life that is complete and self-sufficient (lacking in nothing), includes friendship - reciprocal affection and goodwill - because human flourishing is relational. People can be loved for their pleasure- or utility-value, or for their good character; accordingly, there are pleasure-friendships, utility-friendships and character-friendships. Character-friendships include pleasure and utility, because good character is both pleasurable and useful. Here friends love (*philein*) each other for who they are, their character, and 'not coincidentally' (*Nicomachean Ethics* 1156b11-12); hence, they wish each other 'to be' and to flourish for their own sakes (1166a3-20). Character-friendship requires virtuous activity, and a choice (*prohairesis*) of 'another self' (1170b6-7). Such a choice presupposes genuine self-love, which only virtuous people possess; hence, only they can choose another self. Moreover, virtue is best exercised in pleasurable and beneficial activities with and towards friends, so it is virtuous people who most *want* friends.

A much-discussed problem in Aristotle's account is that he first makes non-instrumental concern - goodwill - a distinguishing feature of *all* friendships (1155b29-1156a5), but subsequently restricts it to character-friendships (1156b10-11, 1167a11-14). He does not attempt to reconcile these claims; nor does he explain *why* we cannot have goodwill towards utility- or pleasure-friends, given his belief that we *can* towards strangers. So perhaps we should take Aristotle to mean simply that active and *enduring* goodwill is found only in character-friendships.

Aristotle also discusses friendship in the wider sense of relations of mutual goodwill and justice among people, including strangers. The general claim is that friendship, justice and community are coextensive, and so all humans capable of community, including slaves, have a sort of friendship with each other (1159b25-30, 1161b6). This belies the widespread view that Aristotle's ethics is parochial. Also noteworthy is Aristotle's insistence that the claims of justice increase with the closeness of the friendship, a claim that runs contrary to many contemporary depictions of friendship as being 'beyond justice' (see Aristotle §25).

The Aristotelian idea that noninstrumental concern is essential to friendship is central to Hellenistic accounts of friendship, though sometimes uneasily so. Thus, Epicurus recognizes friendship as an intrinsic good which necessarily involves loving our friends as we love ourselves, yet insists that every act of friendship must aim at our *eudaimonia* (see Mitsis 1988). But this insistence, conjoined with his conception of *eudaimonia* as a static end-state of unhindered pleasure (*ataraxia*, tranquillity), rather than as pleasurable activity, makes friendship simply instrumental to *eudaimonia*, and so not an intrinsic good. Another problem arises from Epicurus' conception of *eudaimonia* as invulnerable and up to us (*par' hemas*). *Eudaimonia* is achieved through a rational control of the strength and range of one's desires, a control that makes it relatively immune to chance. But what if

the desire that one's friend live and flourish is frustrated? Epicurusmust either insist that we can and should eliminate such desires in the interests of *eudaimonia*, thereby rendering his conception of friendship rather shallow, or acknowledge that *eudaimonia* is not entirely *par' hemas* or invulnerable (see Epicureanism §11).

Laelius, De amicitia (Laelius: On Friendship) by Cicero (106-43 BC) was largely responsible for transmitting Aristotle's conception of friendship to Christian philosophers, notably St Augustine (354-430) and Aelred of Rievaulx (1109- 66), exemplifying his aim of using rhetoric to disseminate philosophical ideas (see Cicero §§4-5). Building on Aristotle's account, Cicero stressed the internal component of self-sufficiency - a sense of security and self-confidence - as a necessary condition of friendship, and made explicit the centrality of mutual respect in perfect friendship, a factor only implicit in Aristotle. In addition, he emphasized rationality by the admonition to base friendship on careful observation and occasional testing of a person's character, and continual self-scrutiny.

2 Medieval Christian philosophy

Friendship raises new questions and problems in the context of the Christian conception of perfect love, *agapē* or charity (see Charity; Love §2). For *agapē* is universal and unconditional, directed at saint and sinner alike, whereas friendship is preferential and conditional on the friend's goodness (see Universalism in ethics). Does not the exclusiveness of friendship, then, contradict the inclusiveness of *agapē*? Augustine's answer is 'no', so long as we recognize that friendship is a gift of divine *agapē*, and use it - like all earthly loves - as a means to perfect love for God, the only love to be enjoyed for its own sake. To love persons as ends is to seek infinite satisfaction in them, and this is idolatory, a violation of the commandment of total devotion to God. As in Plato, so in Augustine, personal love ultimately seeks, and finds satisfaction in, a good that transcends any earthly good (see Augustine §12).

Not all Christian philosophers, however, see personal love as a mere means to a divine end. Like Augustine, Aelred believed that only Christians can have true ('spiritual') friendship, because true friendship requires agreement on both human and divine affairs and a mutual wish for both temporal and eternal happiness. However, Aelred agreed with Cicero that friendship is its own 'fruition and reward' (*De spiritali amicitia* 1:45). For spiritual friendship combines the joys of intimate, trusting, mutual openness with justice, prudence, fortitude, and temperance, which are forms of charity. And, since God is charity, 'he that abides in friendship, abides in God, and God in him' (1:70).

Likewise, the Aristotelian Aquinas (*c*.1224-74) saw the true friend as one who is loved 'simply and for itself', not 'from any extrinsic cause' (*Summa theologiae* IaIIae.26.4, IaIIae.28.2). So exalted is friendship in Aquinas' eyes, that it even serves as a model for charity-love of God. Thus, charity involves communication with God the friend, a communication that enables us to achieve our 'ultimate end' of becoming more like God *Summa contra gentiles* III.1.3, c.19). What remains unexplained is how friendship can have value for God, given that his perfect self-sufficiency renders external goods superfluous (see Aquinas §13).

3 Modern philosophy

Medieval Christian thinkers see *agapē* as predicated on the good or God in everyone, even the sinner, and friendship as predicated on the friend's good character. Martin Luther, however, regarded all conditional-on-goodness love as instrumental, and reinterpreted *agapē* as a spontaneous, gratuitous love, unmotivated by goodness. This turn had a profound influence on later philosophers, both secular and Christian, notably, Immanuel Kant (1785, 1797) and Søren Kierkegaard (1847).

In Kant's discussion, the Lutheran influence was curiously intermingled with the influence of ancient accounts. Like Aristotle, Kant distinguished friendships based on a shared moral attitude - perfect friendships - from lesser sorts and, like Cicero, he emphasized equality and mutuality of love and respect in perfect friendship. Contrary to Aristotle, however, Kant denied emotions any intrinsic moral worth: emotions are subject to natural laws, whereas morality requires freedom or autonomy (self-governance) (see Kant, I. §9). Hence, intrinsic moral worth accrues only to respect and 'practical' love (*agapē*'s secular analogue), which can be willed, and their object, the friend's 'humanity' - the purely rational capacity for moral self-legislation that makes rational beings ends in themselves. Emotional love, and its object, the friend's good character, have, at best, instrumental moral worth. Furthermore, Kant regarded the desire for our friends' happiness as a mere means to our own. Hence, emotional love is

inevitably instrumental self-love. Finally, the view that emotions are involuntary implies that it is *morally* inconsequential whether we love virtuous or vicious people. These counter-intuitive consequences raise the question of whether Kant's view of emotion and emotional love - later echoed by Kierkegaard - can do justice to friendship.

This question, and the utilitarian worry about how to justify the differential concern of friendship if morality requires impartial concern for all, dominate twentieth-century discussions of friendship.

4 Twentieth-century philosophy

The Kantian view of emotion, and its contribution to philosophers' neglect of friendship, is discussed thoroughly by Lawrence Blum in *Friendship, Altruism, and Morality* (1980), the first philosophical book on friendship since Aelred's. In this work, Blum argues that love and concern require thought and choice, are crucial to moral perception and motivation, and can be non-instrumentally (though conditionally and preferentially) directed at friends' good (see Virtue ethics §4). Hence, friendship-love is morally significant.

The idea of the moral significance of friendship is also central to discussions of friendship inspired by Aristotle. Thus, several philosophers have emphasized the importance of friendship to self-knowledge and moral growth. Since friends share the same basic values, each gains greater self-awareness by observing the other's actions and responses, and since friends differ in important ways, each grows through emulation of the other's admirable qualities.

Elizabeth Telfer (1970/1) addresses the utilitarian worry about justifying differential concern (see Utilitarianism §3; Impartiality §2). Telfer's solution is that unequal concern is justified because we are more effective producers of utility (the good) when we concentrate our energies on the needs of the few we know best. However, if utilitarians must justify their friendships as instrumental means to utility, their moral commitments seem psychologically incompatible with friendship's commitments. The usual response to this is that utility-maximization is only a *standard* of right action, not its primary *motivation*. Indeed, utilitarians *should* cultivate the motivations of friendship, subject to the (background) proviso that if, overall, these motivations failed to maximize utility, they would seek to change them. But this is just to say that utilitarians' motivations are, ultimately, shaped by instrumental considerations. So, even if they are *psychologically* compatible with dispositions of friendship, they seem *logically* inconsistent with them.

The theme of preferential concern is salient in feminist and communitarian writings. Feminists emphasize the 'care perspective' as a moral perspective independent of justice, and friendship as a caring relationship that is particularly valuable because of its voluntariness, mutuality and equality. But like most contemporary philosophers on friendship, including communitarians, who oppose justice to friendship, most feminists also oppose the universalist, impartialist justice perspective to the particularist, partialist care perspective (see Feminist ethics; Community and communitarianism). Some, however, recognizing that there is no inherent tension between these perspectives, and deploring the frequent inequality of caring work in female-male friendships, have argued for the reconciliation of the two perspectives, and for the importance of justice in friendship.

But more needs saying. Our particularity cannot exist, or be understood, independently of our common humanity. Hence, caring for someone implies seeing their particularity as a distinctive expression of common aspirations and needs. Conversely, justice implies giving due weight to the particular expressions of people's commonalities. Sensitivity to these aspects requires both justice and care. Justice and care are thus mutually dependent, and equally important in friendship. Indeed, in so far as justice depends on perceptiveness and imaginative effort, and the rightful expectation of these is greater within friendship, friendship *intensifies* the claims of justice.

See also: Morality and emotions; Sexuality, philosophy of; Trust

NEERA K. BADHWAR

References and further reading

Aelred of Rievaulx (1148) *De spiritali amicitia*, trans. E. Laker, *Spiritual Friendship*, Kalamazoo, MI: Cistercian Publications, 1974.(The original title shows the influence of Cicero's *De amicitia* on Aelred. In a true friendship, which is necessarily spiritual, the friends are willing to die for each other, love each other unconditionally and never-endingly, and hold all possessions in common. The only alternative to true

friendship is sinful friendhip.)
Aquinas, T. (*c*.1259-65) *Summa contra gentiles (Synopsis [of Christian Doctrine] Directed Against Unbelievers)*, trans. V.J. Bourke, Notre Dame, IN: University of Notre Dame Press, 1975, III.1.3, c.19.(Expresses the view that our 'ultimate end' is to become more like God.)
Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, trans. Fathers of the English Dominican Province, Glencoe, MN: McGraw-Hill, 1947.(IaIIae.26-8 discuss love as an emotion; IaIIae.26.4 distinguishes friendship based on friendship-love (friendship proper) from friendships based on love of concupiscence or desire (friendships of pleasure and utility). IIaIIae.23-46 discuss charity as friendship with God.)
Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. W.D. Ross, revised by J.O. Urmson, in J. Barnes (ed.) *The Complete Works of Aristotle*, vol. 2, Princeton, NJ: Princeton University Press, 1984.(In addition to the issues discussed in this entry, books VIII-IX on friendship also discuss the justification of friendship, conflicts among friends, when friendships should be dissolved, civic friendship, and other topics.)
Augustine (397-401) *Confessionum libri tredecim*, trans. F.J. Sheed, *Confessions*, Indianapolis, IN: Hackett Publishing Company, 1993, IV, 4-12.(Describes Augustine's grief on the death of a friend, and his later reflection that he had mistakenly loved his friend as an end in himself.)
Augustine (396-426) *De doctrina christiana (On Christian Doctrine)*, trans. D.W. Robertson, Jr, New York: Liberal Arts Press, 1958.(See book I: section iv, passage 4 and xxii, 20-1, for the view that all earthly loves, including neighbour-love, are instrumental or utility loves, and love of God alone an end love to be enjoyed for its own sake.)
Badhwar, N.K. (ed.) (1993) *Friendship: A Philosophical Reader*, Ithaca, NY: Cornell University Press.(A collection of fifteen contemporary articles, or excerpts from books, on the psychological, moral and political significance of friendship. The editor's 36-page introduction expands on some of the material in this entry, discusses related issues, such as the object of love in friendship, and contains extensive references to historical and contemporary literature.)
Badhwar, N.K. (1985) 'Friendship, Justice and Supererogation', *American Philosophical Quarterly* 22 (2): 123-31.(Argues that justice and benevolence are mutually supportive and equally central, constitutive virtues of friendship.)
Blum, L. (1980) *Friendship, Altruism, and Morality*, London: Routledge & Kegan Paul.(The first book on friendship since Aelred's *De spiritali amicitia*, and probably the best-known single contemporary work on the topic. Argues that sympathy, compassion and concern are cognitive emotions that are intentionally diverted at others' good and since friendship is a locus of these emotions, it is an intrinsically moral phenomenon.)
Cicero, M.T. (mid-44 BC) *Laelius, De amicitia (Laelius: On Friendship)*, Latin text with trans. and notes by J.G.F. Powell, Warminster: Aris & Phillips, 1990.(A somewhat unsystematic, but historically influential, dialogue on friendship. Gaius Laelius (186 BC-?), who studied philosophy under the Stoics, is the main speaker.)
Cooper, J.M. (1977) 'Aristotle on the Forms of Friendship', *Review of Metaphysics* 30: 618-48; 'Friendship and the Good in Aristotle', *Philosophical Review* 86: 290-315; a revised and condensed version of these articles appears as 'Aristotle on Friendship', in A.O. Rorty (ed.) *Essays on Aristotle's Ethics*, Berkeley, CA: University of California Press, 1980, 301-40.(Very influential treatment of Aristotle's conception of friendship.)
Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. M.J. Gregor, *Groundwork of the Metaphysics of Morals*, in *Practical Philosophy*, Cambridge: Cambridge University Press, 1993, esp. 389-401, 401 note. (Presents Kant's sharp dichotomy between the moral-rational and the natural-emotional realms, and his view that emotional love is instrumental self-love.)
Kant, I. (1797) *Metaphysische Anfangsgründe der Tugendlehre,* trans. M.J. Gregor, *The Doctrine of Virtue*, in *Practical Philosophy*, Cambridge: Cambridge University Press, 1993, 469-73.(Contains the only extended discussion of friendship in Kant's developed philosophical works.)
Kierkegaard, S.A. (1847) *Kierlighedens Gjerninger*, trans. H. Hong and E. Hong, *Works of Love*, New York: Harper & Row, 1962.(Kierkegaard follows Luther in regarding friendship as essentially irreconcilable with *agapē*, and echoes Kant in declaring that friendship-love and *eros* 'contain no ethical task', both because they are forms of self-love, and because, as natural inclinations, they cannot be willed or, therefore, commanded by God's law.)
Mitsis, P. (1988) *Epicurus' Ethical Theory: The Pleasures of Invulnerability*, Ithaca, NY: Cornell University Press, ch. 3.(Expands on Epicurus' interpretation of friendship, particularly his conception of *eudamonia* as invulnerable.)
Pakaluk, M. (ed.) (1991) *Other Selves: Philosophers on Friendship*, Indianpolis, IN: Hackett Publishing Company.(A very useful introduction to historical writings on friendship by Plato, Aristotle, Cicero, Seneca, Aelred, Aquinas, Montaigne, Bacon, Kant, Emerson, Kierkegaard and Telfer.)

Plato (*c.*380-367 BC) *Lysis*, trans. S. Lombardo, in *Complete Works*, ed. J.M. Cooper, Indianapolis, IN: Hackett Publishing Company, 1997.(The only Platonic dialogue devoted entirely to friendship.)

Price, A.W. (1989) *Love and Friendship in Plato and Aristotle*, Oxford: Clarendon Press.(A detailed, scholarly, investigation of these topics.)

Railton, P. (1984) 'Alienation, Consequentialism, and the Demands of Morality', *Philosophy and Public Affairs* 13 (2): 134-71; repr. in N.K. Badhwar (ed.) *Friendship: A Philosophical Reader*, Ithaca, NY: Cornell University Press, 1993.(Expands on the idea that commitment to utilitarianism or consequentialism is psychologically compatible with personal commitments.)

Stocker, M. (1981) 'Values and Purposes: The Limits of Teleology and the Ends of Friendship', *Journal of Philosophy* 78: 747-65; repr. in N.K. Badhwar (ed.) *Friendship: A Philosophical Reader*, Ithaca, NY: Cornell University Press, 1993.(Argues that *no* theory that analyses or evaluates value or action entirely in terms of goals can fully understand friendship.)

Telfer, E. (1970/1) 'Friendship', *Proceedings of the Aristotelian Society*, supplementary volume: 223-41. (Examines the nature, duties and values of friendship, arguing that friendship is valuable because it is useful, pleasurable and life-enhancing. The article marks the revival of friendship as an important philosophical topic in this century.)

Fries, Jacob Friedrich (1773-1843)

Fries was a German post-Kantian philosopher, active chiefly in Jena and Heidelberg. He was a personal as well as a philosophical enemy of Hegel. Fries' version of Kantian philosophy opposed the speculative idealism of Fichte, Schelling and Hegel, developing an 'anthropological critique of reason'. Fries also emphasized subjectivity in ethics and religion. In politics he was a republican and a German nationalist. For his participation in the Wartburg Festival of 1817 (a gathering of radical student fraternities), Fries was removed from his professorship at Jena in 1820, but restored in 1824. He wrote both scholarly and popular treatises on metaphysics, logic, ethics and politics, as well as mathematics and natural science. His continuing influence early in this century was mediated chiefly by the Göttingen Neo-Kantian Leonard Nelson and by Rudolf Otto's theory of religious experience.

Fries accepted Kant's distinction between sensibility and understanding, the transcendental ideality of empirical objects and the consequent unknowability of things-in-themselves (see Kant, I. §5). Early in his career he defended a critical Kantian position against the emerging speculative idealism in the polemical essay *Reinhold, Fichte und Schelling* (1803). In his *Neue Kritik der Vernunft (New Critique of Reason)* (1807, revised in 1838 under the title *Anthropologische Kritik der Vernunft (Anthropological Critique of Reason)*). Fries was critical of Kant's strategy of arguing for a priori principles by transcendental deduction from the possibility of experience; Fries claimed that a priori principles must instead be known 'anthropologically'. Accordingly, he admitted that we cannot prove the objective validity of Kantian principles such as substance and causality, but only that such principles are indispensable for our own cognitive activities.Fries' approach is sometimes described as psychologism; it does not deny the Kantian a priori, however, but claims that it is knowable through the observation of our faculties.

In ethics, Fries placed primary emphasis on an agent's possession of a 'pure will', one which follows the dictates of conscience and acts solely from the motive of duty (see Kantian ethics). In his *Handbuch der praktischen Philosophie (Manual of Practical Philosophy)* (1818), Fries maintained that we know our duty through the exercise of moral sentiments leading to conscientious 'convictions'; but conscience is always 'educable', and moral judgment fallible. Erring moral judgment, however, does not imply that conscience itself is fallible. An agent whose convictions are objectively wrong still has a pure will, and is therefore to be esteemed rather than blamed. Hence, when judging the acts of others, the standard should always be their convictions, rather than objectively right convictions.

Like Kant, Fries based religion on a faith in divine providence which is rationally justifiable on moral grounds (see Kant, I. §11). But Fries' religious upbringing in Moravian pietism led him to emphasize far more than Kant both the specifically Christian and the experiential aspects of religion. Fries drew the categories of the beautiful and the sublime from Kantian aesthetic theory and - again like Kant - he interpreted these aesthetic experiences as having a moral-religious significance (see Kant, I. §12). For such experiences Fries coined a special term, with a deliberately archaic spelling: '*Ahndung*'; it might be translated 'presentiment', 'divination' or 'inkling'. Inklings do not count as empirical cognitions of any kind, but they do reveal, through feeling, the limitation of the sensible world by the supersensible, and thus count as a kind of experiential awareness of a higher spiritual reality. Fries' philosophy of religion was first expounded in *Wissen, Glauben und Ahndung (Knowledge, Faith and Inkling)* (1805), but later popularly presented in his dialogue *Julius und Evagoras: ein philosophischer Roman (Julius and Evagoras: A Philosophical Novel)* (1813, revised 1822).

Fries is sometimes described as a 'liberal' in opposition to the 'conservatism' of Hegel. This terminology certainly involves a grosser distortion of Hegel's thought than of Fries', but it is oversimplified and misleading in his case too. Fries was a defender of individual rights and liberty, and an egalitarian in economic as well as political matters; he hated the fact that modern society is unified by a cash nexus rather than by moral ties, and he was outraged by the exploitation of the poor by the rich. Fries' views on both points have a German nationalist as well as a communitarian bent, which the term 'liberal' is not well suited to capture. In *Von deutschem Bund und deutscher Staatsverfassung (German Federation and German Constitution)* (1816), Fries advocated the unification of all German states into a single federated constitutional republic - which entailed the abolition of the various monarchical forms of government prevailing in all of them at the time. His German cultural nationalism, however,

made him as hostile to the French Republic and the ideals of the French Revolution as any ultra-reactionary monarchist. While Fries was a social egalitarian, like Kant he favoured a representative republic with a voting franchise limited by property and occupational restrictions (as well as by those of age and gender). Hence he was hostile to the idea of a democratic government, which he equated with mob rule.

Fries was also an anti-Semite, as appears in his pamphlet: *Über die Gefährung des Wohlstandes und Characters der Deutschen durch die Juden* (On the Danger Posed by the Jews to German Well-Being and Character) (1816). He did not recognize Judaism as a religion at all, but saw Jewry as a politico-economic state within a state, one founded on ancient and barbarous principles and hostile alike to individual freedom, moral universality and human dignity. Blaming Jews for the rule of money in society, he opposed any toleration of Judaism, as well as the extension of political and civil rights to Jews. To qualify for such rights, Fries maintained, Jews must first declare their allegiance to the German state (requiring that they renounce Judaism). Judaism was to be 'extirpated root and branch' from German society.

In later life Fries wrote on mathematics and natural science. In general he took Kantianism in a moralistic and empiricist direction, in opposition to the speculative tendencies of German idealism.

See also: German idealism

ALLEN W. WOOD

List of works

Fries, J.F. (1967-76) *Sämtliche Schriften (Collected Writings)*, Aalen: Scientia Verlag.(The standard collected works, in six divisions, itemized below.)

Fries, J.F. (1967-71) *Schriften zur reinen Philosophie (Writings on Pure Philosophy), Sämtliche Schriften (Collected Writings)* Div. 1, Aalen: Scientia Verlag, 5 vols. (Contains the *Neue Kritik der Vernunft (New Critique of Reason)* (1807), revised in 1838 under the title *Anthropologische Kritik der Vernunft (Anthropological Critique of Reason)*.)

Fries, J.F. (1969) *Schriften zur Geschichte der Philosophie (Writings on the History of Philosophy), Sämtliche Schriften (Collected Writings)* Div. 4, Aalen: Scientia Verlag, 3 vols.(Contains Fries' works on the history of philosophy.)

Fries, J.F. (1970-1) *Schriften zur angewandten Philosophie 1 (Writings on Applied Philosophy 1), Sämtliche Schriften (Collected Writings)* Div. 2, Aalen: Scientia Verlag, 4 vols.(Contains *Wissen, Glauben und Ahndung (Knowledge, Faith and Inkling)* (1805).)

Fries, J.F. (1973a) *Schriften zur angewandten Philosophie 2 (Writings on Applied Philosophy 2), Sämtliche Schriften (Collected Writings)* Div. 3, Aalen: Scientia Verlag, 1 vol.(Contains *Handbuch der praktischen Philosophie (Manual of Practical Philosophy)* (1818), *Von deutschem Bund und deutscher Staatsverfassung (German Federation and German Constitution)* (1816) and *Julius und Evagoras: ein philosophischer Roman (Julius and Evagoras: A Philosophical Novel)* (1813, revised 1822).)

Fries, J.F. (1973b) *Populäre Schriften (Popular Writings), Sämtliche Schriften (Collected Writings)* Div. 5, Aalen: Scientia Verlag, 3 vols.(Contains occasional writings of a popular nature.)

Fries, J.F. (1976) *Polemische Schriften (Polemical Writings), Sämtliche Schriften (Collected Writings)* Div. 6, Aalen: Scientia Verlag, 1 vol.(Contains, among other writings, some of Fries' attacks onHegel.)

Fries, J.F. (1982) *Dialogues on Morality and Religion*, ed. D.Z. Phillips, trans. D. Walford, Totowa, NJ: Barnes & Noble.(The recent translation of Fries, an abridged version of *Julius and Evagoras*. The editor emphasizes the affinity of Fries' philosophy of religion with Wittgensteinian Fideism.)

References and further reading

Nelson, L. (1962) *Fortschritte und Rückschritte der Philosophie (Steps Forward and Backward in Philosophy)*, Frankfurt: Suhrkamp.(Contains a polemical defence of Fries by the leading twentieth-century German exponent of his philosophy.)

Otto, R. (1931) *The Philosophy of Religion Based on Kant and Fries*, trans. E.B. Dickler, London: Allen & Unwin.(The author is an influential twentieth-century philosopher of religion, whose classic study *The Idea of the Holy* was strongly influenced by Fries' theory of religious experience.)

Wood, A.W. (1990) *Hegel's Ethical Thought*, New York: Cambridge.(Chapter 10 focuses on a discussion of Fries' 'ethics of conviction' andHegel's critique of it.)

Fujiwara Seika (1561-1619)

In Tokugawa intellectual historiography, Fujiwara Seika has been traditionally deemed the founding father of the Zhu Xi school of neo-Confucianism in Japan. He emphasized seiza *(quiet-sitting) in order to perceive the ethical essence of human nature, and asserted the priority of principle, moving away from dualism towards a more rationalistic monism.*

Prior to the Tokugawa period (1600-1868), Zhu Xi's commentaries on the Four Books - the *Daxue (Great Learning)*, the *Analects*, the *Mengzi* and the *Zhongyong (Doctrine of the Mean)* - had long circulated among Rinzai Zen monks (see Zhu Xi; Daxue; Confucius; Mencius; Zhongyong). They were understood, however, as ingredients in a syncretic, religio-philosophical system combining Buddhism, neo-Confucianism, and Shintō. Only in the Tokugawa period did scholars who were keenly interested in neo-Confucianism break away from Rinzai temples to make neo-Confucianism the independent, anti-Buddhist philosophy that Zhu Xi had intended it to be (see Neo-Confucian philosophy; Confucian Philosophy, Japanese). Seika played a cardinal role in this process because he, though initially a Rinzai monk, quit Buddhism to study and teach neo-Confucianism as a purely secular scholar in Kyoto, the imperial capital.

One of Seika's students, Hayashi Razan (1583-1657), also severed his links with Buddhism to pursue a purer neo-Confucian teaching. Razan studied only briefly under Seika, who later recommended him to Tokugawa Ieyasu (1543-1616), the *samurai*-unifier of Japan as of 1600. The Hayashi family, having established itself in Edo (now Tokyo), the shogun's capital, became hereditary scholars serving the Tokugawa regime; they later praised Seika for initiating the neo-Confucian break with Buddhism in Japan, which they continued wholeheartedly. With the Hayashi family, neo-Confucians increasingly came to serve the Tokugawa in ways that Buddhist scholars had served other earlier military regimes (see Buddhist philosophy, Japanese).

Some historians, however, have questioned the extent to which Seika (1) broke with Buddhism, and (2) taught Razan. Some suggest that Razan, a largely self-educated scholar, was legitimizing his family's teachings by tracing them to Seika, a reclusive scholar of aristocratic lineage. It is true that Seika's ideas were developed most faithfully not in Edo, but instead within the confines of Kyoto, an aristocratic haven surrounded by the *samurai*-dominated world of the Tokugawa shogunate (see Bushi philosophy §3). Seika's longest standing, most respected disciple was not Razan but Matsunaga Sekigo (1592-1657), who remained in Kyoto and founded an academy which flourished after Seika's death, largely dedicated to propagating Seika's teachings.

Seika's thought was, unlike Razan's, rather Buddhistic. Seika emphasized *seiza* or 'quiet-sitting', a neo-Confucian meditative practice outwardly similar to *zazen*, or 'sitting-in-Zen meditation', but aimed at perceiving the ethical essence of human nature. Razan had no use for *seiza*. Metaphysically, Seika asserted the priority of principle, skewing Zhu Xi's dualism of principle and material force towards a more rationalistic monism (see Qi). Razan's thinking tended towards materialism and activism. Like Rinzai syncretism, Seika's teachings were eclectic, accommodating critics of Zhu Xi like Wang Yangming, and elements of Buddhism and Shintōas well; Razan disliked Seika's easy eclecticism. Seika and his disciples mostly refused to serve *samurai*, but Razan's family became scholar-servants *par excellence* of the *samurai* estate.

See also: Confucian philosophy, Japanese; Japanese philosophy; Neo-Confucian philosophy; Shintō; Zhu Xi

JOHN ALLEN TUCKER

List of works

Fujiwara Seika (1561-1619) *Fujiwara Seika shū(Complete Works of Fujiwara Seika)*, Tokyo: Kokumin seishin bunka kenkyūjo, 1938-9; ed. Ishida Ichirōand Kanaya Osamu, *Fujiwara Seika/Hayashi Razan*, Nihon shisōtaikei vol. 28, Tokyo: Iwanami shoten, 1975.(The Complete Works includes essays discussing Seika's works; analyses the *Kana seiri (Neo-Confucianism for Japanese Readers)*, traditionally attributed to Seika and part of his Complete Works, and explores questions about its authenticity. This Ishida and Kanaya edition includes Seika's *Suntetsu roku (Philosophical Apothegms)*, *Daigaku yōryaku (Essential Teachings of the Great Learning)* and selections from the *Fujiwara Seika bunshū*, as well as analytic essays by the editors. The *Kana seiri* is included, along with the similar, anonymous text, *Shingaku gorinsho (Ethical Teachings of the Learning*

of the Mind).)

References and further reading

de Bary, W.T. and Bloom, I. (eds) (1979) *Principle and Practicality: Essays in Neo-Confucianism and Practical Learning*, New York: Columbia University Press.(An anthology including sophisticated essays, many pertaining to Seika, by leading Japanese and American scholars.)

Boot, W.J. (1982) 'The Adoption and Adaptation of Neo-Confucianism in Japan: The Role of Fujiwara Seika and Hayashi Razan', unpublished Ph.D. dissertation, University of Leiden.(Traces Razan's ideas to pre-Tokugawa Japanese thinkers, scrutinizing the real impact of Razan's tutelage under Seika.)

Imanaka Kanshi (1973) *Kinsei Nihon seiji shisōno seiritsu: Seika gaku to Razan gaku (The Establishment of Early Modern Japanese Political Thought: The Learning of Fujiwara Seika and Hayashi Razan)*, Tokyo: Sōbunsha.(Questions traditional accounts regarding the Seika-Razan relationship.)

Ota Seikyū (1985) *Fujiwara Seika*, Tokyo: Yoshikawa kōbunkan.(The most recent biography of Seika in Japanese.)

Fuller, Lon Louvois (1902-78)

Lon Louvois Fuller was a leading US legal philosopher and contracts lawyer who in his controversies with H.L.A. Hart and with US 'legal realists' advanced a version of 'procedural natural law' deriving an 'inner morality of law' from the formal properties of law. At the same time, through his insistence that legal interpretation must always consider the essentially purposive character of legal activity, he forms an intellectual bridge between earlier pragmatist accounts of law and the late twentieth-century interpretivist approach associated particularly with Ronald Dworkin.

Soon after Fuller's birth in Hereford, Texas, his family moved to California, where Fuller received his education, partly at Berkeley, but finally, both in economics and law, at Stanford University. In 1926 he took up a post in the Law School of the University of Oregon, subsequently moving to Illinois, to Duke University, and finally to Harvard in 1940, retiring in 1972.

Anti-formalism is the leading theme in Fuller's theoretical work. It is expressed primarily in giving attention to the extra-statal forms of law, to what he calls 'implicit law', and to extrajudicial and extra-legislative forms of production of law, for instance contracts (his speciality) and industrial relations. Law was for him not just the outcome of decisions by authorities, but also and especially of agreements, compromises and mediations, if not indeed of spontaneous and in a certain sense unconscious harmonization of interests and wills. Fuller's anti-formalism was, moreover, expressed not just in the area of the theory of sources of law. The area dearest to him, to which he devoted special attention in his years of maturity, was that of the relation between law and ethics. Formalists were in his view all those who denied that relation, isolating the law in a sphere made either of forms of various types or of acts of authorities or mere external behaviour, or of meanings of linguistic propositions. It was not just legal positivist doctrine that in his view was guilty of this sort of formalism, but also the realism of such jurists as Karl Llewellyn and the pragmatism of philosophers like John Dewey.

He wrote, 'For its meaning any rule of law is tragically dependent upon an understanding of the purpose pursued and this in turn usually requires a feeling for the extra-legal community and its problems.' But legal positivists and 'realists' denied the need for this reference to purposes and to extra-legal problems in order to understand and explain legal norms. These purposes and problems are not included among the conditions for the validity or existence of a norm, whether by legal positivists like Hans Kelsen or Hart, or realists like Llewellyn or Jerome Frank (see Frank, J.; Legal realism §1; Legal positivism). For all these the jurist's ethos is broadly identical with the scientists' ethos, and is value-free. There is no 'role morality' that binds the jurist's activity and conditions the possible contents of their social function. But for Fuller the problem is just that without reference to values one cannot even identify the scope of what is legal.

The values that Fuller refers to in this connection are not those of good morality. They do not entirely and at every point coincide with ethical values. He wrote to his adversary, Llewellyn, 'You say that there are obviously bad rules which exist. I don't deny that.' Saying, then, that a legal norm exists does not mean preaching its morality (goodness). But in order for a legal norm to exist it has to pursue certain purposes, derived from law's external character, that are separate from the specific content of the norms. Thus Fuller's finalism proves to be a formalist approach, but a deliberately *sui generis* formalism, quite different from the imperativist and legal positivist variety. His formalism is essentially Kantian (see Kantian ethics; Legal idealism §2), since the specific purposes of law also define a procedural scope to be respected, without which the legal norm would be an arbitrary act or a mere instrumental measure. Fuller thus seems to anticipate the distinction recently made familiar by John Rawls between 'right' and 'good'. For him, the law is largely equal to just what Rawls today (along with Habermas) would call the sphere of 'right': a series of formal and procedural principles within which each subject can pursue their own purposes and develop their own conception of the good. Fuller identifies in this connection eight fundamental qualities of legal acts, understood as elements in a value-laden legal order: (1) the generality of the act; (2) its publicity; (3) its orientation towards future behaviour; (4) its comprehensibility; (5) the noncontradictoriness of the provision; (6) consistency and constancy in time; (7) not demanding the impossible; and finally (8) an interpretation and application that are sufficiently respectful of the text and of its ordinary meaning. Failing such requirements, a legal act would not be called 'nonexistent' or 'invalid' by Fuller; it would

however be denied the possibility of acting as a justificatory basis for any legally relevant conduct (a judicial decision or a contract between private individuals, for instance). These eight aspects define the 'inner morality of law' or 'procedural natural law' for which Fuller contended in his great debate with H.L.A. Hart and in his greatest book, *The Morality of Law* (1964).

See also: Law, philosophy of; Norms, legal; Rule of law/Rechsstaat §2

MASSIMO LA TORRE

List of works

Fuller, L.L. (1940) *The Law in Quest of Itself*, 2nd edn, Cambridge, MA: Harvard University Press, 1966.(This is Fuller's earliest work of legal theory, in which he stresses the significance of the purposive approach to law and its interpretation.)

Fuller, L.L. (1964) *The Morality of Law*, 2nd revised edn, New Haven, CT: Yale University Press, 1969.(The great work which establishes the thesis of the 'inner morality of law' in opposition to the version of legal positivism proposed by H.L.A. Hart.)

Fuller, L.L. (1967) *Legal Fictions*, Stanford, CA: Stanford University Press.(Three essays giving a lucid account of the development and significance of 'fictions' in law.)

Fuller, L.L. (1968) *Anatomy of the Law*, New York: Praeger.(A short overview of jurisprudence and legal system, cast in Fuller's characteristic line of thought.)

References and further reading

Summers, R.S. (1984) *Lon L. Fuller*, London: Edward Arnold.(A very thorough account of Fuller's work, with full bibliography, by an author who started as an opponent but became (with qualifications) a supporter of Fuller's doctrine.)

Functional explanation

Explanations appealing to the functions of items are common in everyday discourse and in science: we say that the heart pumps blood because that is its function, and that the car fails to start because the ignition is not functioning. Moreover, we distinguish the functions things perform from other things they do: the heart makes a noise, but that is not one of its functions. Philosophical discussions in this area attempt to specify conditions under which it is appropriate to ascribe functions to items and under which it is appropriate to appeal to those functions in explanations. Difficulties arise because functions are normative: there is some sense in which items ought to perform their functions; failure to perform is a kind of error. Philosophical discussions investigate whether and how this normativity can be understood in scientifically respectable terms. This is important, because biological entities are among the most characteristically functional items. This issue gives rise to differing views as to what it is that functional explanations explain. One view is that they explain how a containing system achieves some goal or effect. Another is that functional explanations explain causally why the functional item exists.

1 Historical context

Science and explanatory practice prior to the modern age was based on the Aristotelian view that explanation is by appeal to causes, and that the primary sense of ‘cause’ is final cause - *telos* or purpose (see Aristotle §9). Activities of things which are parts of wholes were understood as subserving the final causes of the wholes of which they are parts. Their activities, insofar as understood, are their functions. With the discovery of precise mathematically formulable laws of dynamics for general application to all types of inanimate motion, appeal to purposes falls not just out of service but into disrepute. Contemporary science generally conceives nature as the realm of law-like regularity, devoid of irreducible purpose and design. Explanation is still appeal to causes, but these are now conceived in terms of regularity and mechanism. The advent of genetics, the means by which Darwinian natural selection can account for the origins of complex animate behaviour, gives further weight to this conception (see Evolution, theory of).

Yet despite its apparent appeal to purposes and its lack of overt appeal to natural laws, functional discourse and explanation has persisted and been of great utility, especially in the biological sciences. Contemporary discussions seek to provide for these practices a basis which is naturalistic in the modern sense. Three interrelated problems beset such attempts. First, not all effects are functional (the heart has the mere effect, but not the function, of making sound); so functional must be distinguished from ‘mere’ effects. Second, items may have functions which they do not perform. A heart in fibrillation, and the eyes of a congenitally blind person, still have their functions. Third is the specification of what functional explanations explain.

2 Explaining the accomplishment of real goals

Some philosophers distinguish functional from mere effects by taking functions to be contributions to real goals of containing systems, whose accomplishment is partially explained by the resulting function ascription. Ernest Nagel and Morton Beckner, among others, have offered naturalistic criteria for such goals. Roughly, systems are goal-directed if and only if they maintain some feature (possibly a developmental tendency) in relative homeostasis through varying conditions, where that relative homeostasis is achieved by the combined operation of multiple components of the system, whose states are logically independent of one another but causally related. The goal of the system may then be understood as the maintenance of the homeostatic feature (or the state towards which the development tends).

Nagel himself recognized that such non-directed objects as elastic solids and chemical systems in thermodynamic equilibrium are not obviously ruled out by his proposed definition. Moreover, nothing in this approach helps to solve the second, ‘malfunction’, problem, for, unless more is said, a system which ceases to behave in a goal-directed manner ceases to have a goal, and its subparts cease to have a function. Efforts to solve this difficulty generally specify system types in terms of typical members. Since tokens of a type can fail to be typical, token systems can have typical goals they do not accomplish, and their components typical functions they do not perform. However, attempts to distinguish typical members have generally seemed either arbitrary or else to appeal, implicitly or explicitly, to their proper functioning, rendering the account circular for present purposes.

3 Explaining the presence of the functional item

Following claims of sociologists such as Malinowski and Merton that functional explanation is a distinctive method in sociology (see Functionalism in social science), positivist philosophers, notably Hempel and Nagel, attempted to assimilate functional explanation to the deductive-nomological model thought to characterize all explanation. On the deductive-nomological (D-N) model, explanation is a species of inference: from the relevant laws of nature and further facts and relevant background conditions, explananda can be deduced or, in the case of statistical laws, induced (see Explanation §2). Function ascriptions are here taken as equivalent to arguments in which the presence of the functionally characterized item is inferred from a containing system's having certain effects, and from the need for an item of that sort to produce those effects.

It is rarely if ever the case, however, that the only means by which a system can have a given effect is by means of a single specific item. For example, the pumping of blood may be achieved by means of an artificial device. For this reason, plausible function ascriptions will fail as D-N explanations, because the inference from the effect to the presence of the item in question will typically be unsound. This problem has been generally regarded as fatal to any attempt to accommodate function ascription to the D-N model.

The demand that explanations support valid or strong statistical inferences has been generally attacked as too strict. Understanding explanation more modestly, some philosophers continue to maintain that legitimate function ascription explains the presence of functionally characterized items. In a highly influential analysis, Larry Wright (1973) claims that 'the function of I is F' means

(a) I is there because it Fs;
(b) F is a result (or consequence) of I's being there.

The first, etiological requirement is designed to eliminate mere effects, but is clearly too weak to do the job. To cite a well-known example, suppose a pebble has the effect of propping up a rock in a stream. But for the presence of the rock, the pebble would be washed away. Hence the pebble is where it is because it has the effect of propping up the rock. On Wright's analysis, then, the pebble has the function of propping up the rock, which is surely a misapplication of the concept of function. The account also seems under-inclusive, as the second clause demands that the item actually have the functional effect, eliminating mal- and non-functioning items.

The most plausible efforts to refine Wright's basic etiological framework attempt to solve these two problems at once, holding that the etiologies which mark out functional effects are selectional, understood roughly on the model of selectional explanation in evolutionary biology. Importantly, selection always involves the relative replicative or reproductive success of objects with a given feature as compared with competing objects without that feature. The function of the feature is the effect it had in virtue of which objects possessing it were selected. The 'malfunction problem' is supposedly solved by the fact that the function of an item is not identified with any of its actual or even possible effects, but rather with selected-for effects of some predecessor items.

Whether selection history, even in the biological case, actually provides a genuine etiology for particular items is disputed. Robert Cummins (1975) and Elliott Sober (1984), for example, each deny that it does, holding that the etiology of a biological trait is exclusively a matter of the genetic plan inherited by the containing organism. Proponents of the selection view counter that selection accounts for the increased incidence of the item in the population, thereby explaining the fact that the token system in issue possesses it.

Appeal to original selection etiology seems in any event to give rise to further counterexamples, since the effects for which an item has been selected may intuitively be non-functional or positively disfunctional in subsequent generations of the containing systems, and items may have effects in systems that are plausibly regarded as functional, even where their etiologies are either non-selectional, or involve selection for some other effects. Ruth Millikan (1984), whose influential selectional theory of 'proper functions' is intended for use in a naturalized semantics (see Semantics, teleological), has attempted to deflect such counter-examples by claiming that her account is a theoretical definition, rather than an analysis of our ordinary concept of 'function'; failure of overlap with our intuitive function ascriptions is beside the point. However, actual selection histories are generally difficult if not impossible to determine, and the boundaries of the theoretical concept, from an instrumental or operational point of view, are no sharper than those of the intuitive. Moreover, Millikan's account is intended to capture not just biological functions, but others as well, and it is not clear that the notions of ancestry and replication it

requires are applicable outside of the biological context (or perhaps even within it).

Notice that the types of counterexample mentioned above involve the temporal dimension (items can lose functions they once had, and can gain functions after having been selected for other effects). Peter Godfrey-Smith (1994) and Paul Griffiths (1992, 1993), each concerned particularly with biological function, have suggested that these counterexamples can be avoided if only recent or proximal selection history is regarded as relevant to function ascription. The viability of such accounts depends upon whether the notions of ‘recent’ or ‘proximal’ history can be made precise. This is turn depends upon empirical assumptions about likely mutation rates and the incidence and nature of competitors, as well as the assumption that evolution is efficient in the sense that, if an item is not providing some fitness benefit to an organism, then the incidence of that item in the population will decline. Though not implausible, such assumptions are far from clearly warranted. Any attempt similarly to refine a selectional account of function for general application beyond the context of biology would involve a more dubious generalization of these assumptions. Finally, some biological items with selectional etiologies, notably, so-called ‘segregation distorters’, provide no benefit to containing systems, and are even positively harmful. Unless we wish, counter-intuitively, to regard these effects as functional, some non-arbitrary means of ruling them out of the analysis is needed.

4 Explaining interesting capacities

Because of difficulties of these kinds, some philosophers deny that the distinction between functional and mere effects is ultimately real and that function ascriptions do or can operate as explanations of the presence of the functional item. Robert Cummins has offered a well-known account along these lines, on which ascriptions of functional properties to items are always relative to an analytical account of some capacity of some containing system, that is, to an account which breaks down the net capacity into sub-processes which, taken together, are sufficient to explain the capacity. On this view, any capacity of a containing system which results from contributions of its parts is subject to functional analysis, and such parts could conceivably be ascribed innumerable distinct, even incompatible functions (though because of the relativization of function ascription to analysis, no inconsistency results). Cummins’ analysis is thoroughly instrumental, holding that our explanatory interests determine which effects we single out in the sciences for this kind of analysis, hence our ascriptions of functions to particular items, and also which tokens we take as typical, hence what we treat as malfunctional.

There is little doubt that Cummins’ functions are naturalistically respectable, so long as the particular analyses offered are themselves naturalistic, and that they capture much of the actual practice of functional explanation. However, function ascription along Cummins’ lines seems inadequate to account for those cases in which selectional or other etiological accounts adverting to function can be given, notably in the case of artifacts. Moreover, Cummins’ functions are plainly inadequate for service in broader philosophical projects, such as naturalized semantics and epistemology, which require a robustly realist conception of the normativity of function.

See also: Causation;Teleology; Technology, philosophy of

RICHARD N. MANNING

References and further reading

Beckner, M. (1959) *The Biological Way of Thought*, New York: Columbia University Press.(See chapters 6 and 7. Moderately technical accounts of functions as contributions to goals and of goal-directed systems, in the biological context.)

Bigelow, J. and Pargetter, R. (1987) ‘Functions’, *Journal of Philosophy* 84: 181-97. (Functions as contributions to goals).

Boorse, C. (1976) ‘Wright on Functions’, *Philosophical Review* 85: 70-86.(Counterexamples to the etiological theory.)

Cummins, R. (1975) ‘Functional Analysis’, *Journal of Philosophy* 72: 741-65. (Detailed critique of positivist and etiological theories of function; presentation of instrumental account.)

Davies, P.S. (1994) ‘Troubles for Direct Proper Functions’, *Nous* XXVII: 363-81.(Somewhat technical challenge to the applicability of Millikan’s theory of proper functions to biological cases.)

Godfrey-Smith, P. (1993) ‘Functions: Consensus Without Unity’, *Pacific Philosophical Quarterly* 74: 196-208. (Highly readable presentation of the etiological and instrumental accounts as distinct and equally important.)

Godfrey-Smith, P. (1994) ‘A Modern History Theory of Functions’, *Nous* 27: 344-62.(Attempts to accommodate

loss and gain of functions within the selection etiology framework, along with a discussion of segregation distortion.)

Griffiths, P. (1992) ‘Adaptive Explanation and the Concept of a Vestige’, in P. Griffiths (ed.) *Trees of Life*, Dordrecht: Kluwer, 111-31.(Somewhat technical attempt to accommodate loss and gain of function within the selection etiology framework.)

Griffiths, P. (1993) ‘Functional Analysis and Proper Functions’, *British Journal of Philosophy of Science* 44: 409-22.(As above, with a discussion of segregation distortion and an extension to artifacts.)

Hempel, C.G. (1965) ‘The Logic of Functional Analysis’, in *Aspects of Scientific Explanation*, New York: Free Press.(Classic discussion of functional explanation as a species of D-N explanation.)

Kitcher, P. (1988) ‘Function and Design’, *Midwest Studies in Philosphy* 18: 379-97.(Attempts to unify function under the concept of design; moderately difficult.)

Manning, R.N. (1997) ‘Biological Function, Selection, and Reduction’, *British Journal for the Philosophy of Science* 48: 69-82.(Rejecting reductive status of selectional accounts of function; moderate difficulty.)

Millikan, R.G. (1984) *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press. (Influential theory of ‘proper functions’ for use in naturalized semantics; technical but rewarding.)

Millikan, R.G. (1993) *White Queen Psychology and other Essay*, Cambridge, MA: MIT Press.(Further clarification, defence, and application of ‘proper function’.)

Nagel, E. (1953) *The Structure of Science*, New York: Harcourt Brace.(See chapter 12. A typical account, in a positivist framework, of goal-directedness for systems; somewhat technical.)

Neander, K. (1988) ‘What Does Natural Selection Explain? Correction to Sober’, *Philosophy of Science* 55: 422-6.(Nontechnical defence of the view that selection explains the presence of selected for items.)

Neander, K. (1991a) ‘The Teleological Notion of "Function"’, *Australasian Journal of Philosophy* 69: 454-68.(A selectional account of function; moderate difficulty.)

Neander, K. (1991b) ‘Functions as Selected Effects: The Conceptual Analyst’s Defence’, *Philosophy of Science* 58: 168-84.(Rejecting Millikan’s defence of ‘proper function’ as a theoretical definition.)

Sober, E. (1984) *The Nature of Selection*, Cambridge, MA: MIT Press.(See chapter 5. Argues that selection does not explain the presence of biological items; moderately difficult.)

Sober, E (1985) ‘Panglossian Functionalism and the Philosophy of Mind’, *Synthèse* 64: 165-93.(Nontechnical discussion of functionalism in biology, and its possible use in accounting for mental contents.)

Wright, L. (1973) ‘Functions’, *Philosophical Review* 82: 139-68.(An etiological theory of function; highly readable.)

Wright, L. (1976) *Teleological Explanations*, Berkeley, CA: University of California Press.(Booklength version of the above.)

Functionalism in social science

In the social sciences, functionalists are theorists who give an especially prominent role to functional explanations. One of the most influential self-defined functionalists, Malinowski (1926), summed up this view: the functionalist 'insists… upon the principle that in every type of civilisation, every custom, material object, idea and belief fulfils some vital function, has some task to accomplish, represents an indispensable part within a working whole'. As an example of a functionalist explanation, one might consider the hypothesis, as argued for instance by Evans-Pritchard in his work on the Azande in Africa, that belief in witches generally plays a role in maintaining social stability (1937).

In the last few decades of the twentieth century, postmodern, or post-structuralist, sociologists have largely disavowed the pursuit of functional explanations. The extremism of some functionalist theses has been matched by an equal extremism in postmodern antitheses. In denying that everything must be explained functionally, some go so far as to say that nothing should ever be explained functionally.

Yet there is liveable logical space between the modernist's 'There has always got to be a reason, the real reason, for everything', and the postmodernist's 'There is never any real reason for anything'. We do not have to be card-carrying functionalists to suspect that functional explanations might be found for at least some of the bewildering things that some people do in various parts of the world. New models of functional explanation are emerging from recent ferment in the biological sciences, and these new models may suggest new ways of approaching functional explanations in the social sciences.

1 Apparent limits of rational choice explanations

When European explorers made first contact with various different cultures, they were sometimes at a loss to understand many of the practices they witnessed. Potlatch for instance: this was a spectacular practice among the indigenous peoples of north-western North America of the giving and then almost immediate destruction of extremely impressive gifts.

After the initial shock of the exotic, some anthropologists then began to imagine how many of their own practices might appear to strangers. Then familiar things in their own lives began to cry out for explanation. In a similar way, Americans may have been prompted to wonder (in, say, the 1980s): why do businessmen wear ties, and business women shoulder pads, for instance?

There is often a common thread running through those things that seemed especially difficult to understand: it is initially difficult to see goals which you could believe the agents to have, and which you could believe that the agents were thinking they could achieve by the actions they were performing. You ask the agents themselves why they do these things, and either they say they do not know, or they do their best to tell you their motives but somehow you still feel that there is more to be explained.

Research in the social sciences is often driven by a sense of bewilderment at some of the things people do. When our common sense (individualistic, rational choice theoretic) explanatory strategies fail to get a grip, functional explanations naturally emerge to fill the explanatory vacuum.

Functional explanations are not the only sorts of explanations that present themselves as rivals to rational choice theoretic explanations. Functionalism emerged from a range of competing explanatory strategies, as for instance: causal/historical explanations, diffusionism, evolutionary theories of the sort typified by Frazer's *The Golden Bough* (1890), and the possibility of explaining some social phenomena by reference to instincts. But functional explanations have had a tendency to become peculiarly pervasive, as soon as we move beyond rational choice explanations.

If you wonder why a person holds a particular belief, or engages in a particular practice, then one answer that is worth pondering is this: perhaps they believe or act as they do just because that is what other people believe and do. But then we are slipping into a regress: why do those other people hold this belief and engage in this practice? You might speculate that this regress ends with a person who, once upon a time, introduced this belief and associated practices for reasons that were perfectly comprehensible at the time, easily explicable by common-sense

rational choice theory. Then this belief and associated practices have been replicated through the ages even though people have forgotten the initial rationale behind them, just because it is, in certain contexts, individually rational to conform to the beliefs and practices of others. In brief, one way to explain something through rational choice theory is by tracing it back to its origins.

This sort of explanation is problematic, however, for several reasons. Often, there are too few empirical constraints on speculations about the origins of a practice. And even if one were confident about the origin of a practice, there remains something to be explained in the fact that while other practices die out, this one persists. Malinowski said: ‘There are no survivors’ (1922). This may be an overstatement: there may be some practices which really are explained by there having been a rational choice by some originator, and then transmission by rational imitation from there on. Nevertheless, it is arguable that our default assumption should be in the spirit of Malinowski’s dictum; usually, there is something important left unexplained by the theory of rational origin followed by rational imitation.

When our default explanations of behaviour, by common-sense rational choice theory, fail to satisfy, we are at a loss for understanding, and functional explanations in the social sciences aim to fill this explanatory gap. This is the first step towards functionalism.

The second step towards functionalism is taken when social scientists begin to look more sceptically at cases that used to seem straightforwardly explicable in terms of common-sense rational choice theory. A suspicion begins to grow, that the veneer of rationality is always just rationalization or, to take a phrase from Marxism, false consciousness. An analogy can be drawn with psychoanalysis. Some sorts of bizarre behaviour seem difficult to explain by common-sense rational choice theory, so extraordinary modes of explanation are introduced, that posit unconscious beliefs and desires. Then the psychoanalyst looks back at superficially rational choices of everyday life and begins to suspect that our ‘real’ reasons are never the conscious reasons. Likewise, the extreme functionalist always looks for ‘latent’ functions behind whatever ‘manifest’ functions there may be, if any.

2 Functionalists and their influence

Many theorists, some less outspoken than Malinowski, have been collected under the label of functionalism because of family resemblances, acknowledged indebtednesses, and affiliations to others travelling under that label. Famous names in the family are predominantly anthropologists: Marcel Mauss, Émile Durkheim, Talcott Parsons, E.E. Evans-Pritchard, Arnold Radcliffe-Brown, Franz Boas, Ruth Benedict, Margaret Mead, Edward Sapir.

Some theorists, like Claude Lévi-Strauss, broke away from some (though not all) of the distinctive doctrines of the self-defined functionalists and travelled under the label of structuralism (see Lévi-Strauss, C.; Structuralism in social science). But the relationship between structuralism and functionalism is often far from clear, and it is not clear whether structuralism is a friend or a foe of functionalism. Structuralists aim to understand a social phenomenon by describing a network of interdependent phenomena within which the thing to be explained is located. The whole network is then seen to be a structure that, at a sufficiently high level of abstraction, is instantiated in every human society. The metaphors guiding structuralism are more often drawn from grammar than from biology; the natural bedfellow of this sort of theory is what is known as systems theory. The supposed universality of certain deep structures suggests that they are in some important sense innate. Structuralism may thus be seen as a sophisticated version of social explanation by reference to human instincts. And this is often seen as a rival to explanation by reference to functions.

Nevertheless, the structuralists’ explanatory strategy does still have a close kinship with functional explanations. Systems theory embodies much of what is distinctive of functional explanations, and is applied across a wide range of research in the social sciences: in politics, social psychology, agrarian studies, and in sociology generally. Much of structuralism can be absorbed into what is called structural-functionalism (this label attaches especially to Talcott Parsons (1952)), and much of systems theory can be construed as raw materials for a more complete functional explanation. A description of ‘how a system works’ can be construed as a description of ‘the functions’ which the elements are playing in the whole.

3 Critiques of functionalism

Yet, as with the posits of psychoanalysis, a social scientist's attributions of functions can be clouded by logical and epistemological miasmas. It is sometimes unclear just what it means to say that some practice has a certain function. And while it is often relatively easy to think up a story about what function some practice might have, it is often hard to know what sort of evidence would count for or against that hypothesis: verification or falsification are not so straightforward.

For instance, Cohen used the tools of the logical positivists, especially Hempel's model of scientific explanation (1965), to articulate the logic of functional explanations. He also drew heavily on what was a novel idea at that time: feedback loops. In *Karl Marx's Theory of History: A Defence* (1978), he brought all this to bear on political examples of special interest to Marxists. This brought the nature of functional explanation into sharp focus, and thereby helped to define some of the growing unease about these sorts of explanations. In an influential critique, in *Ulysses and the Sirens* (1979) and elsewhere, Jon Elster (another Marxist) argued that the required postulations of feedback loops are almost never empirically testable. Hence functional explanations are just as empirically unconstrained as fanciful speculations about origins, which they were supposed to supplant. Much can be said in defence of Cohen, but the kinds of worries expressed by Elster have led many people to retreat from the functionalist bandwagon.

4 Social functions compared with biological functions

To get a grip on the nature of functional explanation in general, it may be useful to consider some of the most successful functional explanations outside the social sciences. Think first of functional explanations of features of artefacts, as for instance when we are puzzled by one of the attachments on a Swiss army knife. However these functional explanations seem to draw us back towards the intentions of a designer, and this is not a model which helps us with the kinds of puzzle-cases which mostly cry out for functional explanation in the social sciences.

Functional explanation in the social sciences can be more profitably compared with functional explanation in the biological sciences. Just as an anthropologist is puzzled by a practice like potlatch, so too a biologist is puzzled by something like the peacock's tail. Adaptationism is the name given to the default assumption, for every feature of an animal which catches your interest, that it must be serving a function. The functionalism of Malinowski's dictum, that 'there are no survivors', is like an extreme adaptationism that would look for a current function not only for say the pancreas, but even for the human appendix.

In the biological sciences, it used to be thought that any references to functions and purposes would presuppose the intentions of a creator. There are analogous viewpoints that could be held in the social sciences. It might be thought that the attribution of social functions would presuppose the intentions of some agent: either the intentions of the human or deity who invented the practice, or the unconscious intentions of the people who engage in the practice, or else intentions embodied in some sort of group mind. For various reasons, the prospects for functional explanation in the social sciences would be slim if they carried that much metaphysical baggage with them.

Fortunately, however, it has become apparent that sense can be made of biological functions without presupposing any conscious processes of design or selection or intention. A biologist may undertake to find a purely biological explanation of the presence of the pancreas for instance, or of the peacock's tail. One way to do so is by finding that this thing has had a propensity to bring about certain characteristic effects, which in turn have contributed to the reproductive capacities of the animals which possessed it. One key concept in evolutionary thinking is that of 'inclusive fitness': this is taken to be the propensity of an organism to reproduce other organisms with this same propensity.

Thus, a biologist who wants to explain the presence of something biological can legitimately do so by reference to its functions, and can do so without thereby incurring a theological commitment. And this is a relief: it does seem obvious that one reasonable way of explaining the presence of, say, the pancreas is by identifying some of its functions, and that this is not tantamount to belief in a creator.

According to what is called the etiological theory of biological functions, for something to have a biological function is for it to have an evolutionary history that explains its current existence by reference to those of its past effects which have contributed to its reproductive fitness and subsequent survival and replication.

The application of the etiological theory of functions to the social sciences generates some interesting results. One

example is provided by Marvin Harris in *Cows, Pigs, Wars and Witches* (1974). Here, some of the puzzling activities of people in New Guinea are explained by attributing to these practices the function of population control.

It is not implied by Harris that the people engaging in various social practices in New Guinea are either consciously or unconsciously aiming at population control. It is not implied that any person or deity designed the practices with the intention that this would result in population control, or that there was any intention on behalf of something you might call a group mind. Yet Harris meant more than just that these puzzling practices resulted in population control. He meant that these practices existed in these social groups precisely because they have a propensity to result in population control. What he hypothesizes is that, over the ages, those social groups which engaged in such practices prospered. Furthermore, those groups prospered because these practices resulted in population control, which in turn ensured that the social group did not exceed the carrying capacity of their environment. If Harris were right about this sort of evolutionary past for the relevant social groups, then the etiological theory of functions would license the attribution of functions to those practices.

Functional explanations of this sort, however, will seldom be well supported by evidence. It is seldom that we will have good reason to believe that the practices we seek to explain have the kind of history which would be called for by the etiological theory of functions.

Consider an example. Some witch beliefs and practices are pandemic, and one might reasonably suppose them to have had enough evolutionary history to justify an etiologically grounded attribution of functions. But the case is different with the large scale European witch-hunts that began a little before 1492, and periodically flared up or smouldered on until the last of these horrific witch-hunts, in Salem in 1692. This is just the sort of case in which explanation by common-sense rational choice theory seems to be missing something of underlying importance, and hence in which we might hope to find some more satisfying functional explanations. Yet the European witch-hunts also constitute a case in which it seems doubtful that we could ever establish the kind of evolutionary history that Harris postulates for the New Guinea case.

There is a striking feature of early functionalism that can be seen to arise naturally out of an etiological conception of functions. There used to be a propensity of early functionalists to attribute to every practice the very same function: that of promoting social stability. There was a tendency to see every belief and every practice as cogs in a social machine for suppressing change. Later functionalists have tried to explain change as well as stability. But if we are using an etiological theory of functions, then functional explanations of change do not rest easily on those etiological foundations.

There is an instructive parallel in biology. On the etiologist’s conception, biological functions cannot contribute to an explanation of an increasing frequency of some brand new feature: with no evolutionary history there can be no etiological function. The etiologist’s functions can only explain the presence of a feature which has contributed to fitness over a significant period of time. So for the etiologist, functional explanations can only get a grip when there is stability rather than change.

Likewise in the social sciences, the etiologist’s functions cannot explain change but only the existence of ‘survivors’. This is a disappointment, because among the things that a functionalist was hoping to explain were, not just ‘where do we come from?’, but also ‘what is going on here?’ and ‘where are we going?’

5 New tools from the biological sciences

Some functional explanations appeal to an intelligent creator, others to Darwinian evolution; both of these kinds are of limited use in social theory. It is thus worth asking whether there may be other logical forms which functional explanations might take, and which might better serve the needs of the social sciences. Here are some models which have been emerging in the biological sciences.

Instead of wondering about why a characteristic is present in an animal, a biologist might take an interest in the fitness of the animal in some particular environment. Overall fitness may then be analysed as the product of several distinct components, somewhat as a physicist analyses a resultant force as deriving from several distinct component forces. Whenever some feature of an organism is currently contributing to the propensity for survival and replication, it is natural to speak of this feature as serving a function. This may be so, even if there is not yet

any history during which this propensity has actually contributed to survival and replication. Not yet having enough history, the relevant feature cannot 'have a biological function' according to the etiologist. Nevertheless, if something is currently contributing to overall fitness, then there is already a clear enough non-etiological sense in which it makes sense to speak of it as serving some biological function.

This construal of functions might be instructively applicable to various problem cases in the social sciences. It is more likely that a sociologist could establish that something is currently contributing to success, than that it has an etiology which explains its current existence in terms of past successes. In the case of the European witch-hunts, for instance, the spread of witch-hunts did not happen by magic; there must have been something about witch-hunts that explained their alarming propensity for replication. This sort of explanation would speak more directly to our bewilderment, than would the backward-looking etiologist's search for an evolutionary history behind the witch-hunts. What is especially puzzling about witch-hunts, after all, what is crying out for explanation, is the way they spread and kept recurring, their replicatory fitness so to speak. And what a non-etiological functional explanation can hope to offer is an identification of the component functions which contributed to that overall replicatory fitness.

In recent biology, there has been heated debate about the so-called units of selection. There have been interesting challenges to the narrowly Darwinian assumption that biologically functional features of an organism must always enhance the inclusive fitness of the individual organism which carries that feature. In opposition, it has been argued that in special circumstances it can sometimes happen that a gene will increase in frequency even if it lowers the survival and reproductive prospects for the organism which carries it. This is, for instance, the point of the Dawkins' 'selfish gene' theory (1976). Others have argued that patterns can evolve in groups of individuals, or even in 'developmental systems' or 'ecologies' that blur the boundaries between organisms and environment, and that these patterns can evolve in a way which is to some degree autonomous from any contributions that are being made to the fitness of the individuals involved. Whether or not the empirical evidence in biology backs up these less narrowly Darwinian biological theories, the logic of these debates over units of selection can teach lessons of value for functional explanation in the social sciences.

If we model social science functional explanations exclusively on narrowly Darwinian evolutionary theory, then the only things with social functions will be things that 'benefit' the particular societies in which they appear. Yet consider again the example of the European witch-hunts. On the face of things, they had a high propensity to replicate themselves even though it was far from obvious what 'benefits' if any they conferred on the social groups in which they occurred. Thus, it is desirable to at least try out the hypothesis that witch-hunts were analogous to an epidemic, which has a propensity to spread and persist even though it does nothing but harm to its victims. It is therefore of theoretical interest if we can find ways of articulating functional explanations of things like the witch-hunts without presupposing that whatever is functionally explainable has to be socially beneficial.

Functionalism in the social sciences has had a troubled history. It arose out of an urgent need for explanations of various bewildering things people do; and at least in some cases it is very tempting to explain these things by reference to social functions rather than to rational choices, instincts, or other explanatory candidates. Arguably, functionalists overplayed their hand in demanding functional explanations for everything. They also had difficulty getting a secure grip on the logic and epistemology of functional explanations. And they had trouble understanding how functional explanations could account for change as well as stability. However, similar difficulties also beset functional explanations in the biological sciences, and progress has been made in overcoming these difficulties. It is premature to conclude, as some postmodernists have done, that functional explanations can have no place in the social sciences (see Post-structuralism in the social sciences; Postmodernism).

See also: Evolutionary theory and social science; Explanation in history and social science; Functional explanation; Social science, methodology of

JOHN C. BIGELOW

References and further reading

Cohen, G.A. (1978) *Karl Marx's Theory of History: A Defence*, Princeton, NJ: Princeton University Press.(For its time, one of the most careful and penetrating analyses and applications of the concept of function as it occurs in social explanations, presenting a clearly defined target for Elster's (1979) influential epistemological critique.)

Dawkins, R. (1976) *The Selfish Gene*, Oxford: Oxford University Press.(An influential critique of individualistic

presuppositions of many Darwinians, proposing the controversial thesis that features of organisms serve the function of benefiting, not the individual, but the individual's genes.)

Durkheim, É. (1915) *The Elementary Forms of the Religious Life*, trans. J.W. Swain, London: Allen & Unwin, 1971.(A classic work in sociology, portraying societies as more like organisms than like mere aggregates of rational individuals.)

Elster, J. (1979) *Ulysses and the Sirens*, Cambridge: Cambridge University Press.(An influential Marxist critique of functional explanations, especially Cohen (1978).)

Evans-Pritchard, E.E. (1937) *Witchcraft, Oracles and Magic Among the Azande*, Oxford: Clarendon Press.(A functionalist classic in anthropology, seeking the social functions served by apparently irrational magical beliefs.)

Frazer, J.G. (1890) *The Golden Bough*, London: Macmillan, 3rd edn, 1925-36. (An 'evolutionary theory' in anthropology, laden with stories of strange beliefs and practices from all over the world.)

Harris, M. (1974) *Cows, Pigs, Wars and Witches*, New York: Random House.(The chapter on 'Primitive War' is especially interesting, pp. 61-80, proposing to explain puzzling practices in New Guinea as serving the function of population control.)

Hempel, C.G. (1965) *Aspects of Scientific Explanation*, New York: Free Press.(A classic in the broadly logical positivist tradition, including an influential analysis of scientific explanation in general, and functional explanations in particular.)

Malinowski, B. (1922) *Argonauts of the Western Pacific: An Account of Native Enterprises and Adventure in the Archipelagoes of Melanesian New Guinea*, London: Routledge.(A functionalist classic in anthropology, seeking a functional explanation for every social phenomenon.)

Malinowski, B. (1926) 'Anthropology' in *Encyclopaedia Britannica*, London: Encyclopaedia Britannica Inc, 13th edn, supplement I.(A manifesto for an extreme functionalist thesis: that there is a functional reason for everything social.)

Mauss, M. and Humbert, H. (1899) *Sacrifice: Its Nature and Function*, trans. W.D. Halls, London: Cohen & West, 1964.(A functionalist classic in anthropology, seeking the social function of rituals of sacrifice.)

Merton, R.K. (1957) *Social Theory and Social Structure*, Glencoe, IL: Free Press.(Contains an influential essay on 'Manifest and Latent Functions': functionalists are especially concerned with uncovering 'latent' functions.)

Midelfort, C.E. (1872) *Witch Hunting in Southwestern Germany 1562-1684*, Stanford, CA: Stanford University Press.(Excellent material, with critical comments on early functionalism, and reasons for thinking witch hunting in Europe did not function to maintain social stability.)

Parijs, P. van (1981) *Evolutionary Explanation in the Social Sciences*, Totoawa, NJ: Rowman & Littlefield.(Clear and comprehensive exposition and further references.)

Parsons, T. (1952) *The Social System*, London: Routledge & Kegan Paul.(A sociological classic in 'structural-functionalism', construing functions both as essential elements within the structures which constitute any human society ('structuralism'), and as processes which meet the needs of the societies in which they occur ('functionalism').)

Radcliffe-Brown, A.R. (1922) *The Andaman Islanders*, London: Cambridge University Press.(A functionalist classic in anthropology.)

Summers, M. (1971) *The Malleus Maleficarum of Heinrich Kramer and James Sprenger*, New York: Dover. (Source material on early witch-hunts in Europe, full of expressions of and exhortations to apparently irrational and socially dysfunctional beliefs and practices.)

Thomas, K. (1971) *Religion and the Decline of Magic*, London: Weidenfield and Nicolson.(A classic on a topic on the interface between history, sociology and anthropology, seeking explanations for things which are not easy to understand just by reference to rational choices of individuals.)

Weber, M. (1922) *Die protestantische Ethik und der Geist des Kapitalismus (The Protestant Ethic and the Spirit of Capitalism)*, trans. T. Parsons, London: Allen & Unwin, 1930.(A classic in sociology, seeking to explain various features of religion by reference, neither to reason not to revelation, but to the functions they serve in a capitalist society.)

Wilson, E.O. (1975) *Sociobiology*, Cambridge, MA: Harvard University Press.(An influential attack on the individualistic presuppositions of many Darwinists in biology, arguing like Dawkins (1976), that many things serve functions other than those of benefiting individuals.)

Future generations, obligations to

There are at least three different views concerning obligations to future generations. One is that morality does not apply here, future generations not being in any reciprocal relationship with us. Another is that, though we are not obliged to do anything for future generations, it would be praiseworthy to do so. A third view is that justice demands that we respect the interests of future generations.

Philosophers and others have discussed obligations in three main areas: the environment, and the damage inflicted upon it in pursuit of profit; savings and the accumulations of capital; and population policy.

Different theoretical approaches have been taken. According to utilitarianism, the interests of people count equally with those of present people, and all interests are to be satisfied maximally. This may have very demanding implications. Contractarianism rests morality on the agreement of all affected parties. But whose views will be considered in the case of future generations? Perhaps the most plausible approach is communitarianism, according to which obligations can rest on a sense of community which stretches into the future.

1 Areas of debate

The idea that one generation should regard itself as having duties to succeeding generations is not new. We find the following story, for instance, in the Talmud. A a young man once saw an old man planting a carob tree. Knowing that the tree would bear fruit only seventy years later, the young man asked him whether he was sure that he would live long enough to eat the fruit, only to be answered, 'When I came into the world I found it planted with carob trees. Just as my ancestors planted for me, so shall I plant for posterity.' Discussions of the issue in Christian and secular Western philosophy can also be found in the works of Bacon, Kant and others (Passmore 1974). And yet, only in the 1970s did a few scholars start to discuss the issue in depth (Rawls 1971; Sikora and Barry 1978; Parfit 1984).

Relations between distant generations involve important political and moral issues in three areas. The first is the environment. To what extent do contemporary people have the right to deplete resources, to leave 'time-bombs' such as radioactive waste for future persons, and to change the environment while seeking to improve their own material welfare? The treatment of such problems as the depletion of non-renewable natural resources, the pollution of the soil, the contamination of water, the production of toxic and radioactive waste, the destruction of rare species, conservation, and preservation represents the neglect of our duties not only to other contemporaries and, perhaps, to the environment itself, but also to the people of the future. Failure on our side to fulfil these duties incurs much economic cost and considerable difficulties in terms of health and quality of life for future people (thus the optimistic view that technology and science will solve all environmental problems is rejected) (see Environmental ethics.)

The second area in which inter-generational issues are at stake is savings and the multi-generational accumulation of capital. Should we save for future people, even though this prevents us from improving the welfare of the most disadvantaged among contemporaries? Should our governments be allowed to take very long-term loans, assuming that future people will repay the loan plus interest (without necessarily enjoying the fruits of these loans)? The third area is population policies, which also involves the issue of inter-generational distribution and the ever-growing demands for such goods as energy and food (see Population and ethics).

2 Utilitarianism

The question of the grounds for our obligations to future generations raises many theoretical difficulties. The most obvious is that this kind of relationship is different from the usual intra-generational relationship. For example, relevant information is absent. And so, although environmental policies may be seen in terms of inter-generational distribution of access to goods, such as clean air and beautiful landscape, and although we may assume that a certain act or policy will affect future generations, we do not know precisely how bad the consequences of our actions will be, what future generations will want or need, what will bring them happiness, and how many people will be affected (indeed, the very existence and identity of future generations depends on our actions today, and, at least theoretically, we can decide not to reproduce).

These are the difficulties faced by consequentialist theorists such as utilitarians when constructing a theory of inter-generational justice (see Consequentialism; Utilitarianism). It seems impossible to measure or to calculate accurately the personal utilities of the not-yet-born. Moreover, if the time at which people exist cannot affect the value of their happiness, how far should we consider the interests of posterity when they seem to conflict with those of existing persons? A utilitarian who is committed to temporal universalism (that is, regarding the utilities of all persons as equal, whatever the period in which they live) will find it impossible to achieve an equilibrium between our obligations to posterity and the necessity to improve the welfare of contemporaries. If we assume that an investment is productive and is used wisely by subsequent generations, the bigger the initial investment is, the larger the total utility will be. Thus, we may infer that we should reduce our standard of living to a minimum; but then, the next generation will find itself in the same situation, with the same moral demands. Are we then to assume that each generation should save and conserve as much as it can, consuming as little as possible, or almost nothing? While some radical Greens find this idea appealing, many people would find it unreasonable and perhaps impossible to put into practice (see Green political philosophy §4). In addition, there is a serious question about which generation should break the chain and enjoy the fruits of savings by past generations.

3 Contractarianism

Perhaps a contractarian theory could avoid some of these problems (see Contractarianism). The basic idea of contractarianism is that a policy is approved if all concerned people would have approved it had they been asked. Therefore, the first question which arises is who will be ‘invited’ to decide upon the principles of justice between generations: one’s contemporaries, all possible people, or all actual people? The difficulty is that our decisions and policies will also affect the identity and number of future people. Do we first decide upon the policies (that will determine the identity of future people), or upon the identity of the parties to the contract (who will decide upon the policies)?

There are other difficulties with the contractarian theory. It is widely agreed that a contract will be reached in ‘circumstances of justice’, that is, a rough equality in the powers and capacities of the parties concerned, a situation of moderate scarcity, and a conflict of interests, which is moderated by a sense of potential mutual benefit from social and economic cooperation. But it is not certain that, when applied to the inter-generational context, these circumstances still exist. The second and third conditions are a matter of information which is not now available; and it is certain that the first - equality of powers between generations - does not exist, both because the current generation has the capacity and power to decide not to reproduce, and also because it is we who decide about every matter which concerns our relationship with future generations. Even a sense of mutual advantage is lacking: it seems quite reasonable, from the point of view of self-interest, to suppose that, after benefiting from a contract with previous generations, a generation will change the terms of the contract, and hence of its obligations to future generations.

It has thus been suggested that an inter-generational contract should be based on the idea of impartiality rather than on mutual advantage (Barry 1989) (see Impartiality). Justice is an impartial identification with the interests of everyone, and thus in deciding upon principles of justice we should not be influenced by our knowledge of what makes us different from others (such as the fact that we are contemporaries).This contractarian theory is strongly related, however, to the idea of neutrality about the good, and the theorist must then explain, for example, why one should care about the distant future or the conservation of ecosystems rather than show greater concern for the welfare of disadvantaged contemporaries (is the environment, or nature, being unjustifiably assumed as part of the good?).

4 A communitarian theory

Perhaps our obligations to future generations can be derived from the sense of a community that stretches and extends over generations and into the future (see Community and communitarianism; Solidarity). However, unlike the backward-looking Burkeian notion of a trans-generational community that grounds an obligation to continue the heritage of previous generations, this notion of trans-generational community would be based on the argument that, just as many people regard the past as part of what constitutes their ‘selves’, so they regard the future as part of their ‘selves’ (see Burke, E.). These, then, would be the relationships that form the trans-generational community, which constitute the source of our obligations to future generations.

Often, the concept of community is related to face-to-face interaction. It goes without saying that such interaction does not exist in relations between distant generations. But community is arguably a level of relationships standing beyond contractual or legal relations. As such, the trans-generational community is defined not geographically but rather as a moral entity which does not have to rely on face-to-face interactions. Instead, it is constituted by a framework from which people draw their values and derive their cultural and moral identities. It is within this framework that the discourse about the moral and political ideas constituting social life and identity takes place. This discourse is carried on through critical discussion, mass-media, art, literature, academic research, and so on. But is there an interaction of this kind between distant - and not necessarily overlapping - generations?

The answer seems to be affirmative: the moral discourse extends beyond one’s lifetime. There is no reason why, if I want to do something because I think it is good, I should be indifferent to its value if it turns out that it could only be accomplished after my death (compare Rolston 1981; Barry 1977). When we conceive an idea or create something, we reflect on its relevance not only to the political and moral environment in which we live, but also to the political and moral environment of the future. This is illustrated in many spheres of human activity. For example, what is pure mathematics today may become applied mathematics within a few generations; several technological inventions that are utilized now will still be useful in a generation or more. The same applies to the humanities and arts, and also to political theory and reflections on the values a society holds. Thus, philosophers consider how the future will view their ideas, and politicians do certain things because they wish to be remembered in the future in a certain way. Just as contemporaries study and reflect on the ideas of Plato, Kant or Hume, so will future people discuss ideas which were conceived by our contemporaries. Distant generations do reflect on each other’s cultural and moral ideas, and these reflections are always directed towards the future. In that sense, even our notion of the present is future-oriented. This constructs the multi-generational community which extends into the future, and is the source of our obligations to other members of this community (future generations).

All these theories offer explanations of the moral grounds for our obligations to future generations. We are still left with questions such as how much we should save and what we should conserve. The answers to these questions depend partly on the value we attach to natural resources, rare species, and objects of conservation, and are thus related to Green theory and environmental philosophy.

See also: Agricultural ethics; Development ethics; Reciprocity; Technology and ethics

AVNER DE-SHALIT

References and further reading

Barry, B. (1977) ‘Justice Between Generations’, in P.M.S. Hacker and J. Raz (eds) *Law, Morality and Society*, Oxford: Clarendon Press.(An analysis of the contractarian approach to inter-generational justice.)

Barry, B. (1989) *Theories of Justice*, London: Harvester Wheatsheaf, ch. 5.(Uses the idea of inter-generational relations to illustrate the two different contractarian traditions and the advantage of the theory of impartiality.)

Busttil, S., Agius, E., *et al*. (1990) *Our Responsibilities Towards Future Generations*, Malta: Foundation for International Studies and UNESCO.(This report includes religious perspectives on the issue.)

de-Shalit, A. (1994) *Why Posterity Matters*, London: Routledge.(Expounds a communitarian theory of justice between generations.)

Gauthier, D. (1986) *Morals By Agreement*, Oxford: Clarendon Press, ch. 9.(Tries to turn the issue into a matter of making a contract between a chain of overlapping generations.)

Heyd, D. (1992) *Genethics*, *Berkeley, CA: University of California Press*.(Mainly discusses population policies, defending the person-regarding view.)

Laslett, P. and Fishkin, J. (eds) (1992) *Justice Between Age Groups and Generations*, New Haven, CT: Yale University Press.(A collection of essays which considers both familiar and newer theoretical questions as well as the application of the theories and their implementation.)

MacLean, D. and Brown, P. (eds) (1983) *Energy and the Future*, Totowa, NJ: Rowman and Littlefield.(A collection of essays which discuss theoretical and practical questions, mainly in the context of the energy crisis of the 1980s.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press.(An important and provocative book which attacks person-regarding morality, partly as applied to inter-generational relations.)

Partridge, E. (ed.) (1981) *Responsibilities to Future Generations*, Buffalo, NY: Prometheus Books.(A collection of essays which mainly discusses theoretical questions. Very good as an introduction.)

Passmore, J. (1974) *Man’s Responsibility for Nature*, London: Duckworth.(Focuses specifically on the theme of

responsibility for the natural world.)
Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(Discusses our obligations to future generations mainly with regard to savings and the accumulation of capital.)
Rolston, H., III (1981) ‘The River of Life: Past, Present, Future’, in E. Partridge (ed.) *Responsibilities to Future Generations*, Buffalo, NY: Prometheus Books.(The issue is treated by this environmental philosopher as another way - in addition to non-anthropocentric ethics - of justifying environmental policies.)
Sen, A. (1961) ‘On Optimising the Saving Rate’, *Economic Journal* 71: 479-97.(A pioneering work of an economist in this field.)
Sikora, R.I. and Barry, B. (eds) (1978) *Obligations to Future Generations*, Philadelphia, PA: Temple University Press.(An excellent collection of essays dealing both with population policies and with the question of inter-generational distribution.)

Fuzzy logic

The term 'fuzzy' refers to concepts without precise borders. Membership in a 'fuzzy' set - the set of things to which a 'fuzzy' concept (fuzzily) applies - is to be thought of as being a matter of degree. Hence, in order to specify a fuzzy set, one must specify for every item in the universe the extent to which the item is a member of the set. The engineer Lotfi Zadeh developed a theory of fuzzy sets and advocated their use in many areas of engineering and science. Zadeh and his zealous followers have attempted to develop fuzzy systems theory, fuzzy algorithms and even fuzzy arithmetic. The phrase 'fuzzy logic' has come to be applied rather imprecisely to any analysis that is not strictly binary. It does not refer to any particular formal logic, in the sense in which the term 'logic' is used by philosophers and mathematicians. ('Fuzzy logic' is sometimes used anachronistically to refer to any many-valued logic.)

1 Philosophy

Fuzzy theory has, paradoxically, tended to partition the intellectual community into true believers and unbelievers. Early fuzzy researchers sometimes ignored or failed to credit properly ideas and similar approaches in other areas, resulting in much duplication of work. Simplistic condemnations of bivalent thinking and everything based on it (virtually all of science) were made, along with broad, often unsupported, claims for the utility of a fuzzy approach. These factors together with the intellectual conservatism of the engineering community led to the unwarranted view in some quarters that fuzzy theory was a fringe development designed only to appeal to those unable to measure up to the rigours of higher mathematical techniques. As a result, fuzzy theory was often ridiculed, frequently ignored, and sometimes excluded from conferences and publications. This situation resulted in some circling of the intellectual wagons and the founding of several conferences and journals devoted solely to fuzzy theory.

Some advocates have attributed the schism to philosophical differences, suggesting that a fuzzy world view has roots in eastern philosophies while western thinking has been dominated by bivalence. The fuzzy world view may be summed up by saying that everything is a matter of degree; that nothing real has any property to the complete exclusion of its negation. Fuzzy advocates criticize bivalent thinking as too simplistic.

Advocates of the fuzzy world view claim that bivalent thinking leads to paradoxes, such as the sorites paradox (see Vagueness §2). (If we pluck one hair from the face of a bearded man, he is still bearded; applying this argument many times over, there is no real difference between a bearded man and one without a beard.) They claim that we can avoid such paradoxes by recognizing that most predicates apply only to some degree. (If we pluck one hair from a man's face, then he is simply slightly less bearded than before: any sharp line drawn between having a beard and not having a beard is completely artificial.)

Critics respond that progress in science is usually the result of abstracting or idealizing: maps are useful because they leave out a lot of detail; computation of pressure/temperature/volume relationships does not require, and would be hindered by, the details of molecular positions, masses and velocities. Further, (western) science has developed mathematical techniques for dealing with matters of degree that do not require a fundamental change of logic. With regard to the paradoxes, critics point out that the specification of precise numerical degrees for concepts such as 'bearded' is artificial and impossible to carry out in any meaningful way. Further, there is no reason to think that we can linearly order the universe in terms of degree of 'beardedness'; some individuals are just not comparable. And if fuzzy advocates avoid drawing artificial distinctions, the paradoxes will simply reappear. For example, let 'really-bearded' apply 100% to those individuals whose degree of beardedness is 0.5 or greater and 0% to those whose degree of beardedness is less than 0.5. Unless fuzzy theorists are going to draw some very artificial fine distinctions, then it seems that removing one facial hair should not change one from being 'really-bearded'. If fuzzy advocates draw fine distinctions to avoid the paradox, then they can hardly criticize bivalent logic for drawing such distinctions.

2 Vague concepts

Most of the concepts which we employ are vague rather than sharp. Vague concepts (see Vagueness) are extremely useful for getting around in the world and for communication. Suppose you have been told to look for a

tall, bald man with a beard, wearing a flowered shirt and walking with a limp. You do not need to have the criteria for the term 'tall' specified to within a millimetre; you do not need criteria for baldness spelled out in terms of percentage of coverage of scalp; and so on. If we tried to make these terms very precise, we would have more difficulty applying them. It seems that the more precise a term, the more difficult it is to determine whether it applies.

Classical logic makes the simplifying assumption of bivalence: every sentence is either true or false. But consider the concept of baldness. Certainly there are clear cases in which we would agree that '*X* is bald' is true, and others in which we would agree that '*X* is bald' is false. But there are also a lot of cases on the borderline. Instead of just 'true' or 'false', it is tempting to use locutions such as 'partially true', 'more true than not', and so on. Classical logic does not seem suited to reasoning with vague concepts. Such observations are cited as a motivation for many-valued logic (see Many-valued logics).

Max Black (1937) was the first to publish systematic suggestions for formalizing reasoning with vague concepts. Suppose our universe is human beings, and consider the predicate 'bald'. We may think of the predicate as standing for the set of objects which have the property. In classical logic we would have to decide all borderline cases arbitrarily. Classically, we may identify the set with its two-valued 'characteristic' function: $b(x) = 1$ if *x* is bald, and $b(x) = 0$ if *x* is not bald. Black suggested that instead of taking the characteristic function to be two-valued, we should allow it to take any value in the closed real interval $[0, 1]$. Ordering the universe in terms of degree of baldness, we could plot a 'baldness' curve as shown.

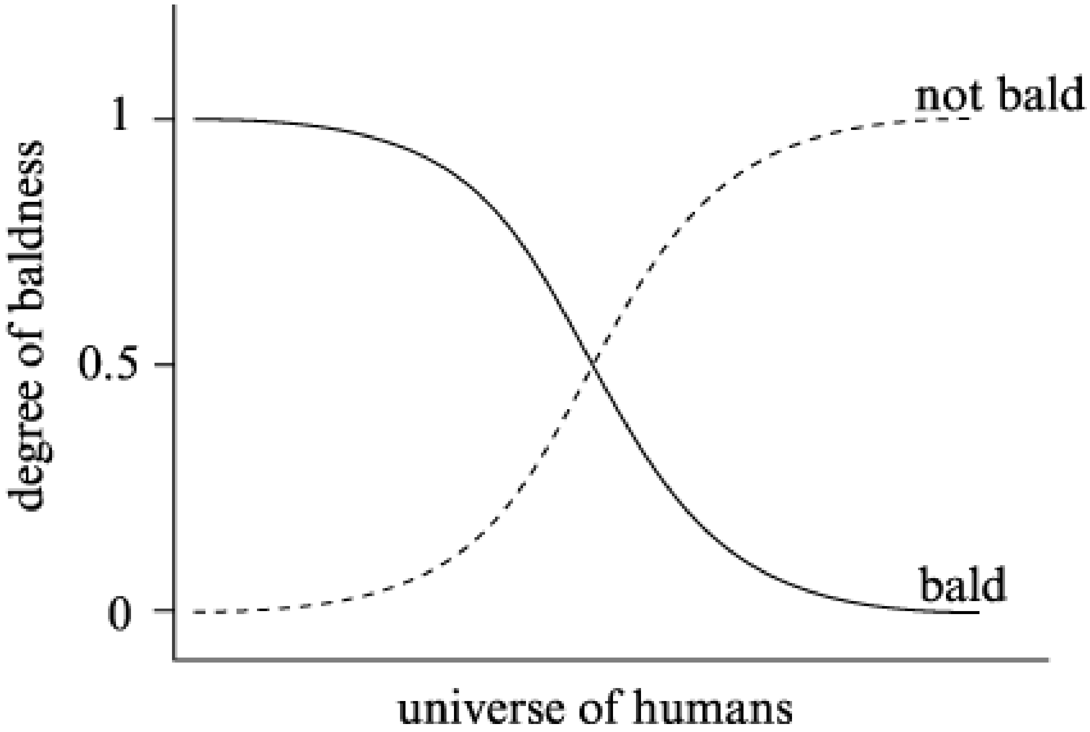

Black advocated treating the negation of a concept as the complement with respect to 1, that is, $\sim b(x) = 1 - b(x)$. He noted that the curve for a vague concept and its negation would cross at the 0.5 position. He further noted that for items in the grey area between 0 and 1, it made sense to say that the item both did and did not have the property. Black recognized the rather arbitrary nature of assigning precise numerical degrees to the membership relation. He suggested using statistical surveys and assigning $b(x)$ to be the percentage of the linguistic community that would agree that *x* has property *b*. Paradoxically, this procedure is based on the possibility of forcing vague concepts into a binary ideal that Black's analysis was intended to avoid.

3 Fuzzy theory

In name, fuzzy set theory was first developed by Lotfi Zadeh (1965, 1975), though the content is strikingly similar to Black's earlier analysis. However, Zadeh took the theory much further.

Assume some universe, *U*; identify each fuzzy set *f* with its characteristic function $f : U \rightarrow [0, 1]$. Zadeh suggested the following definitions.

union:	$(f \cup g)(x) = \max\{f(x), g(x)\}$
intersection:	$(f \cap g)(x) = \min\{f(x), g(x)\}$
complement:	$f'(x) = 1 - f(x)$
universal set:	$u(x) = 1$
empty set:	$\emptyset(x) = 0$

These definitions reduce to the classical notions if we restrict characteristic functions to being two-valued. Union and intersection so defined are commutative, associative, idempotent and mutually distributive, and De Morgan's laws are satisfied. (Zadeh's definitions are not the only ones to satisfy these properties.) However, if there is an *x* in *U* for which $0 < f(x) < 1$, then $f \cup f' \neq u$ and $f \cap f' \neq \emptyset$. These observations correspond well with our intuition that we cannot draw a sharp line between, for example, 'bald' and 'not bald'.

Difficulties arise when trying to treat fuzzy set theory as a basis for a logic of vagueness. There is a simple correspondence between classical set theory on a universal set and classical logic. We may think of the universe as the domain of a model, with sets corresponding to predicates; and union, intersection and complement corresponding to disjunction, conjunction and negation, respectively. The same sort of correspondence exists between Zadeh's fuzzy set theory and the infinite-valued logic of Łukasiewicz (§3). But the quantificational version of infinite-valued Łukasiewicz logic is not finitely axiomatizable. Further, trying to assign a precise numerical degree to the extent to which each person is bald seems even more artificial than simply assigning True or False to each attribution.

Finally, Łukasiewicz logic is not expressively strong enough to allow one to express in the object language the fundamental semantic concepts which served as the motivation. We want our formal language to contain 'autodescriptive' operators J_i for each truth-value *i*, such that, in any particular model, the sentence $J_i(E)$ takes the value 1 if *E* takes the value *i* in that model, and 0 otherwise. But if we include autodescriptive operators in the syntax, it is not possible to give a finitary proof theory for even the propositional logic. So, either the object language must be inadequate to express the notions fundamental to fuzzy theory, or there can be no finitary proof theory.

Zadeh's response to such criticisms was to fuzzify even the metatheory. He suggested that a fuzzy logic should be regarded as a complex association of some formal base logic (for example, that of Łukasiewicz) with fuzzy linguistic structures. Semantic attributions such as 'true', 'very true', 'somewhat true' and so on are to be treated as fuzzy sets. He claimed that all rules of inference are only approximate, criteria for good arguments can be given only semantically, and that the standard metatheoretical notions such as consistency and completeness are not relevant to fuzzy logic.

In spite of the difficulties of developing a formal logic based on fuzzy principles, a fuzzy analysis has led to some elegantly simple techniques for mathematical approximation of complex systems. As a result there have been some impressive systems developed in areas of expert systems and control engineering.

See also: Logical and mathematical terms, glossary of

CHARLES G. MORGAN

References and further reading

Black, M. (1937) 'Vagueness: An Exercise in Logical Analysis', *Philosophy of Science* 427-55. (Original analysis of vague concepts; referred to in §2.)

Dubois, D. and Prade, H. (1980) *Fuzzy Sets and Systems*, New York: Academic Press.(Excellent formal development and review of the literature of the theory of fuzzy sets and related matters.)

Kosko, B. (1992) *Neural Networks and Fuzzy Systems*, Englewood Cliffs, NJ: Prentice Hall.(Formal development of important results in fuzzy systems theory, along with partisan philosophical commentary on the fuzzy world view.)

Kosko, B. (1993) *Fuzzy Thinking*, New York: Hyperion.(A much less formal version of the previous item with some intellectual history.)

Morgan, C.G. and Pelletier, F.J. (1976) ‘Some Notes concerning Fuzzy Logics’, *Linguistics and Philosophy* 1: 79-97.(Problems finding proof theory for fuzzy logic.)
Scarpellini, B. (1962) ‘Die Nicht-Axiomatisierbarkeit des unendlichwertigen Prädikatenkalküls von Łukasiewicz’, *Journal of Symbolic Logic* 27: 159-70.(Quantified infinite-valued Łukasiewicz logic is not axiomatizable; see §3.)
Zadeh, L.A. (1965) ‘Fuzzy Sets’, *Information and Control* 8: 338-53.(Classic paper on fuzzy set theory; referred to in §3.)
Zadeh, L.A. (1975) ‘Fuzzy Logic and Approximate Reasoning’, *Synthèse* 30: 407-28.(Rejection of formal proof theory for fuzzy logic; referred to in §3.)

Fëdorov, Nikolai Fëdorovich (1829-1903)

Like many other major figures in the nineteenth-century Russian tradition of speculation, Fëdorov was not an academic philosopher, but an unsystematic religious thinker who sought working answers to the fundamental questions of life. Fëdorov's basic question was: 'Why do the living die?' His answer, in short, was that we die because we neglect our God-given duty to regulate nature. Fëdorov's life work was to formulate an activist approach to the problem of death, a 'common task' in which all people living on earth, all religions and all sciences would eventually be united in a universal project to resurrect all the dead.

Born in southern Russia near Tambov as an illegitimate son in the princely Gagarin family, Nikolai Fëdorovich Fëdorov always wrote from the viewpoint of the outsider looking in, the unlearned addressing the learned, raising for serious discussion the 'naïve' questions that philosophy had previously failed to answer: Why do we kill? Why do we hunger? Why are some people kin and others strangers? Why do we die?

With support from the Gagarin family, Fëdorov received a sound education, first at the Tambov gymnasium, and then at the respected Richelieu Lyceum in Odessa. In 1851, upon the death of the uncle who was his benefactor, Fëdorov left the Lyceum without a degree, ending his formal education. For the next several years, he wandered from village to village through central and southern Russia, serving as a teacher of elementary history and geography in such places as Lipetsk, Bogorodsk, Uglich and Podolsk. As a teacher, Fëdorov seems to have been loved by his pupils but viewed by headmasters as an overzealous nuisance. Narratives from people who knew him in those years depict a devoted, saintly, eccentric educator who engaged his pupils in unusual group research projects and who on one occasion even gave up his own teacher's uniform to pay for the burial of an indigent pupil's father.

In 1869, Fëdorov took a position as assistant librarian at the Rumiantsev Museum in Moscow, where he became a legend among writers and scholars over the next thirty years. Living alone as an ascetic vegetarian, sleeping on a humpback trunk in a tiny rented room, wearing the same overcoat winter and summer, giving away most of his meagre salary to the poor, always the first to arrive at work and the last to leave, he was said to know not only the location and title but also the contents of every book in the vast library. When a scholar would order books on an obscure research topic, it was said that Fëdorov would usually bring double that number, including titles that the scholar had been unaware of but that gave new depth or direction to the research.

His writing was done for the most part in collaboration with disciples, late at night, on work holidays, and in his last years after retirement from the library. His great work, published in two posthumous volumes under the title *Filosofiia obschago dela (The Philosophy of the Common Task)* (1907?-13), is essentially a 1,200-page miscellany of long and short essays, unfinished drafts, fragments and inspired jottings, all variations on a single theme. In the world as it is, ruled by nature, the universe of matter and man is disintegrating into isolated particles; in the world as it ought to be, regulated by human reason, eternal unity and harmony will prevail. Therefore the human task, our common duty, is to join a universal project to guide our own evolution to the point where we may exercise benign control over nature and complete the perfection of ourselves and our universe.

Knowledge, in Fëdorov, is neither subjective nor objective, but 'projective', action guided by reason towards the realization of an ideal. The 'project' is the bridge between the real and ideal, and between all other opposite poles in Western dualism.

By uniting all to overcome the only true enemy of all, namely death, Fëdorov believed that the project of resurrection would also solve the social, economic and other problems of his day. Energies and resources now directed towards war or commercial exploitation would be redirected towards resurrection. Historical enemies would find mutual assistance not only possible but necessary. Unbelievers, who in theory might find Christianity unacceptable, would, by resurrecting the dead, become in practice followers of Christ.

The first steps towards resurrection might consist of little more than the brief, temporary resuscitation of a person who had just died. But, as all scientific technology, sociopolitical organization - indeed, all human knowledge and action - gradually became directed towards the goal of resurrection, more than brief and temporary resuscitation would become possible. Eventually the synthesizing of bodies should be feasible, and ultimately, Fëdorov

believed, whole persons could be recreated from the least trace. To recover particles of disintegrated ancestors, Fëdorov imagined, research teams would have to travel to the moon, the planets, and to distant points throughout the universe. Eventually these outer points of the cosmos would be inhabited by the resurrected ancestors, whose bodies might be synthesized so as to live under conditions that could not now support human life as it is known.

Fëdorov published almost nothing during his lifetime, and his posthumous works were circulated haphazardly in tiny editions. Nevertheless, his ideas had a strong impact on both late nineteenth-, and early twentieth-century Russian intellectuals. In varying ways and to varying degrees, Dostoevskii, Tolstoi and Vladimir Solov'ëv all incorporated Fëdorov's ideas into their work. In the twentieth century, Fëdorov's influence is apparent in the works of the rocket scientist Tsiolkovskii, the writers Briusov, Belyi Maiakovskii, Khlebnikov, Platonov and Pasternak, the religious thinkers Bulgakov, Florenskii and Berdiaev, and the natural scientists Vernadskii, Chizhevskii, Kholodnyi, Kuprevich and Maneev.

Viewed strictly as a philosopher, Fëdorov was an amateur who contributed little to the history of the discipline, but as an imaginative thinker his bold, comprehensive project of resurrection represents a unique coalescence of several previously divergent tendencies in Russian thought, Slavophile and Westernist, scientific and religious, traditionalist and futurist, probably making him, as Berdiaev observed, 'the most Russian' of Russian thinkers.

GEORGE M. YOUNG

List of works

Fëdorov, N.F. (1982) *Sochineniia (Works)*, ed. S.G. Semenova, Moscow: Mysl'.(Consists mainly of selections from *Filosofiia obschago dela*. Includes bibliographical references and indexes.)

Fëdorov, N.F. (1907?-13) *Filosofiia obschago dela. Stat'i, mysli i pis'ma Nikolaia Fëdorovicha Fëdorova (The Philosophy of the Common Task. Essays, Thoughts and Letters of Nikolai Fëdorovich Fëdorov)*, ed. V.A. Kozhevnikov and N.P. Peterson, vol. l, Verny, vol. 2, Moscow, 1970; repr. Farnborough: Gregg International.

Fëdorov, N.F. (1990) *What Was Man Created For? The Philosophy of the Common Task. Selected Works*, trans. E. Koutaissoff and M. Minto, UK: Honeyglen and Switzerland: L'Age d'Homme.(The first English translation of extensive selections from Fëdorov's major essays.)

Edie, J., Scanlan, J., Zeldin, M. and Kline, G. (eds) (1965) *Russian Philosophy*, Chicago: Quadrangle Books, vol. 3, 16-54.(Translation of part of the opening essay in Fëdorov's collected works.)

Schmemann, A. (ed.) (1965) *Ultimate Questions: An Anthology of Modern Russian Religious Thought*, New York, 175-223.(Translation of selected passages from the collected works.)

References and further reading

Hagemeister, M. (1989) *Nikolaj Fëdorov: Studien zu Leben, Werk und Wirkung (Nikolai Fëdorov: A Study of his Life, Work and Influence)*, Munich: Otto Sagner.(Detailed and thorough, with special attention to Fëdorov's twentieth-century followers.)

Koehler, L. (1979) *N.F. Fëdorov: The Philosophy of Action*, Pittsburgh, PA: Institute for the Human Sciences. (Special attention to Fëdorov's relationship to twentieth-century Russian religious thinkers.)

Lukashevich, S. (1977) *N.F. Fedorov (1828-1903): A Study in Russian Eupsychian and Utopian Thought*, Newark, NJ: University of Delaware Press.(An interpretive study.)

Masing-Delic, I. (1992) *Abolishing Death: A Salvation Myth of Russian Twentieth-Century Literature*, Stanford, CA: Stanford University Press.(Gives special attention to Fëdorov's impact on twentieth-century Russian literature.)

Semenova, S.G. (1990) *Nikolai Fëdorov: Tvorchestvo zhizni*, Moscow: Sovetskii Pisatel'.(A thorough study of the life and works using information from archives available only in Russia.)

Young, G.M. (1979) *Nikolai Fedorov: An Introduction*, Belmont, MA: Nordland.(A general introduction to the life and thought.)

Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, 2 vols; 2nd edn, 1989; trans. G.L. Kline, *A History of Russian Philosophy*, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953.

Fénelon, François de Salignac de la Mothe (1651-1715)

Fénelon is best-known for his utopian political novel Aventures de Télémaque fils d'Ulysse (Telemachus, Son of Ulysses) (1699), which contrasts the rustic simplicity of Greek antiquity with the corrupt luxuriousness of Louis XIV's Versailles. The crucial philosophical works of Fénelon are the Refutation of Malebranche (c.1686-7) and the Maxims of the Saints (1697).

Fénelon came from a provincial aristocratic family (Périgord) which had long served the French crown. First noticed by Bishop Bossuet in the 1670s, Fénelon spent that decade ministering to the 'new Catholics' (ex-Huguenots) in Northern France. He was made tutor to the grandson of Louis XIV in 1689, then Archbishop of Cambrai in 1695.

Although Fénelon is best-known for his utopian political novel *Aventures de Télémaque fils d'Ulysse (Telemachus, Son of Ulysses)* (1699), which contrasts the rustic simplicity of Greek antiquity with the corrupt luxuriousness of Louis XIV's Versailles, the *Réfutation du système du Pére Malebranche* (*c*.1686-7) is the best evidence of his philosophical abilities. Malebranche had argued in his controversial *Traité de la nature et de la grace* (1680) that God governs the realms of nature and of grace equally by 'simple' and 'general' invariable laws; that these laws are the product of divine 'general will' or *volonté générale*; that nature is 'nothing but the general laws which God has established in order to construct or to preserve his work by the simplest means'. God, according to Malebranche, 'does not act at all' by *volontés particulières*, by lawless *ad hoc* volitions, as do 'limited intelligences'; and the generality and simplicity of the divine operation even explains the origin of evil in an imperfect world: if, Malebranche says, 'rain falls on certain lands, and if the sun roasts others… this is not at all because God wanted to produce those effects by *volontés particulières*; it is because he has established [general] laws… whose effects are necessary consequences'. Those who claim that God *ought*, through special *volontés particulières,* to suspend general laws if their operation will harm the innocent, or that he ought to confer grace particularly only on those who will be saved by it, fail to understand that it is unworthy of a wise being to abandon general rules. Those who want God to act through particular wills, simply 'imagine that God at every moment is performing miracles in their favour'. But this 'self-love', Malebranche insists, rests on 'ignorance' of the fact that God 'rarely' departs from generality and simplicity (unless 'order' permits the departure), that he acts 'not often' through anything as *particulier* as a miracle.

In his *Réfutation*, Fénelon challenged the whole Malebranchian notion of a 'general' and 'simple' universe, and thought that he had found a fatal flaw in Malebranche's admission that God acts only *usually*, but not invariably, through general wills and general laws - that God sometimes, though 'rarely', acts through *volontés particulières*. 'In what', Fénelon demands, 'consists that which the author calls "rarely"? These words signify nothing, unless they mean that there is a certain small number of *volontés particulières* which order permits to God outside the general laws, after which he can will nothing particularly'. But if order permits God a small number of particular wills, then 'it follows not only that these *volontés particulières* do not harm in the slightest the simplicity of God's ways, but even that it is more perfect of God to mix some *volontés particulières* in his general plan, than to limit himself absolutely to his *volontés générales*'. Fénelon imagines a hypothetical case in which 'order' has permitted God to have a hundred 'particular wills'; and asks the rhetorical question: 'What, then, is this "simplicity" which is able to accommodate a hundred [particular] wills, which even requires them, but which invincibly rejects the hundred and first?' Fénelon responds in an effective passage: 'If God did not have these hundred *volontés particulières,* he would cease to be God; for he would violate the order which requires them… [but] if he had the hundred and first *volonté*, he would also cease to be God, for he would destroy the simplicity of his ways'. Finally, Fénelon inquires, triumphantly and sarcastically, whether there is 'a fatal number of exceptions which God is obliged to use up, after which he can will nothing except according to general laws?' No one who takes divine omnipotence seriously, Fénelon insists, would 'dare' to say this.

In 1697, Fénelon wrote the book which caused his banishment by Louis XIV, the *Maximes des saints (Maxims of the Saints)* - the 'quietist' work arguing for the 'disinterested' or 'pure' love of God which got Fénelon into such trouble with the Church. In the *Maxims*, Fénelon argued for five degrees of 'purity' or 'disinterestedness' in human love of God. At the lowest end of the scale one finds 'purely servile love': the love of God, not for himself

but for 'the goods which depend on his power and which one hopes to obtain'. One small notch above this Fénelon places loving God, not for 'goods' which he can provide but as the 'instrument' of our salvation; even this 'higher' love, however, is still 'at the level of self-love'. At the third and fourth levels Fénelon finds a mixture of self-love and true love of God; but what really interests him is the fifth and highest degree, the 'pure love' of God that one finds only in 'saints': 'One can love God', Fénelon urges, 'from a love which is pure charity, and without the slightest mixture of self-interested motivation'. In such a love, neither the 'fear of punishment' nor the 'hope of reward' plays any part at all. To achieve this love, one must 'go out of oneself' (*sortir de soi*), even 'hate oneself' (*se haïr*). Malebranche, in his *Traité de l'amour de Dieu* (1697), argued that Fénelon's 'disinterested' love excluded all hope of salvation, as well as all fear of justified punishment, and thus subverted Christianity; and Fénelon's work was finally placed on the Index. Only Leibniz (among important contemporaries) continued to urge that Fénelon's version of 'quietism' was well-meant.

In 1713, near the end of his life, Fénelon wrote *De l'existence de Dieu*, which is not as original and controversial as it is lucid and graceful; its object was to turn back Jansenism as a form of demi-Calvinism. The sheer orthodoxy of this work left it unattacked - an uncommon event in Fénelon's life. But it is the works on Malebranche and on 'quietism' which are still taken seriously - together with *Aventures de Télémaque fils d'Ulysse*, which became the most read book in eighteenth-century France after the Bible.

See also: Pietism

PATRICK RILEY

List of works

Fénelon, F. de S. de la M. (1699) *Aventures de Télémaque fils d'Ulysse*, ed. and trans. P. Riley, *Telemachus, Son of Ulysses*, Cambridge: Cambridge University Press, 1994.(Fénelon's political utopian novel.)
Fénelon, F. de S. de la M. (1835) *Oeuvres de Fénelon*, Paris, 3 vols.(All the works of Fénelon mentioned are most conveniently to be found in this collection. Including *Réfutation du système du Pére Malebranche* (*c*.1686-7), *Maximes des saints* (1697), *Telemachus* (1699) and *De l'existence de Dieu* (1713).)

References and further reading

Goré, J.L. (1957) *L'Itinéraire de Fénélon: humanisme et spiritualité*, Paris.
Gouhier, H. (1977) *Fénelon philosophe*, Librairie Philosophique, Paris: Vrin.
Malebranche, N. (1680) *Traité de la nature et de la grace*, in *Oeuvres complètes de Malebranche (Complete Works of Malebranche)*, vol. 5, ed. A. Robinet, Paris: Vrin, 1958-65; trans. P. Riley as *Treatise on Nature and Grace*, Oxford: Oxford University Press, 1992.(Malebranche's theodicy and account of human freedom.)
Malebranche, N. (1697) *Traité de l'amour de Dieu*, in *Oeuvres complètes de Malebranche (Complete Works of Malebranche)*, vol. 14, ed. A. Robinet, Paris: Vrin, 1958-65.(Criticism of Fénelon's 'disinterested love'.)

Gadādhara (1604-1709)

Gadādhara Bhaṭṭācārya was a seventeenth-century Indian philosopher belonging to a school of thinkers, Navya-Nyāya, noted for its extreme realism and its contributions to philosophical methodology. Though Gadādhara's commentaries on the school's key texts are recognized as among the latest, most detailed and innovative, his greater claim to fame is due to his composition of a number of independent tracts on topics in philosophy of language, philosophy of mind, ethics and legal theory. He may be credited in particular with the discovery of a version of the pragmatic theory of pronominal anaphora. His work on case grammar and inferential fallacies is highly admired in India, while recent translations into English have begun to make him better known outside.

1 Life and work

Gadādhara Bhaṭṭācārya lived from 1604 to 1709. He was a native of Bengal, and became a student of the 'celebrated Naiyāyika' Harirāma Tarkavāgīśa at Navadvīpa, the famous centre for philosophical studies in the district of Nadia, northwest of Calcutta. The late medieval philosophical school to which he belonged, Navya-Nyāya, was especially noted for its contributions to logical theory (quantification, propositional laws, soundness of argumentative schema, negation), epistemology (perception, testimony, casual/reliabilist theory of knowledge), and for developing a quasi-formal technical language capable of great exactness.

Gadādhara is known first as a masterly Nyāya commentator, and second for having written a large number of celebrated tracts on particular philosophical concepts. His subcommentary on Raghunātha's *Tattvacintāmaṇidīdhiti (Ray in the Jewel of Knowledge)* soon became a standard and authoritative work, gaining wide popularity throughout India. Gadādhara's principal contributions to Indian logical theory are to be found within this commentary, notably his remarks on the proper analysis of the 'pervasion' (*vyāpti*) relation between an inferential sign and the property to be inferred (Ingalls 1951: 154-61), and his study of the inferential fallacies. Among his philosophical tracts, which are always illuminating and often brilliant, the ones which brought him greatest fame are those on the philosophy of language (the *Śaktivāda* and *Vyutpattivāda (Theory of the Analysis of Sentence Meaning)*), the philosophy of mind (the *Viṣayatāvāda*), and ethical and legal theory (the *Muktivāda*, *Vidhisvarūpavādārtha* and *Vivāhavādārtha*).

2 Philosophy of language

Gadādhara's contribution to the study of language in India is of particular importance. His influence here lies in his move away from the earlier Nyāya epistemic conception of language, of language as essentially an instrument for communication of knowledge, towards a conception more influenced by semantic theory and linguistics. When an expression is indexical its meaning, according to Gadādhara, is not to be identified with its reference, but rather with what is called the *śakyatāvacchedaka* or 'limiter-of-the-property-of-being-a-referent'. Gadādhara supposes that an ordinary nominal, such as '(the) cow', is an indexical referring expression, whose limiter is the property cowhood. Introducing an important semantic distinction related to Kripke's distinction between reference-fixing and meaning-giving (see Kripke, S.A. §2), he asks whether this property is an 'indicating' property (*upalakṣaṇa*) or a 'qualifying' property (*viśeṣaṇa*). According to Gadādhara's definition, cowhood is here a qualifying property just in case one correctly understands an utterance of the word 'cow' only if one thinks of the referent under the mode or sortal 'cowhood'. In the Nyāya technical language, cowhood is then the 'limiter-of-the-property-of-being-the-object-of-the-hearer's-thought' (*viṣayatāvacchedaka*). Gadādhara argues that the limiter of an ordinary noun-phrase must be a qualifying property, for otherwise two coextensive nominals would be synonymous, but that as there is no single conventionally fixed way of thinking about the reference of a proper name, the limiter of a name merely 'indicates' its reference.

Extending this approach to other types of expression, Gadādhara says that the meaning, that is, limiter of reference, of the first person pronoun 'I' is the property 'being the speaker', which here functions merely to 'indicate' the reference. In constructions such as

(1) The teacher said: 'I am not coming today',

Gadādhara distinguishes between 'dependent' and 'independent' occurrences of 'I', a dependent occurrence being defined as one in which the first person is embedded in a sentence or noun-phrase dominated by a verb-phrase. Such occurrences are governed by a coreference rule of the form: 'I' corefers with the principal noun-phrase. Another construction involving pronominal anaphora is examined at great length by Gadādhara:

(2) There is a pot next door - bring it!

Here, however, Gadādhara rejects a coreference rule in favour of a pragmatic account of anaphora: the reference of the pronoun is determined by the speaker's referential intentions.

Pāṇini, the fourth-century BC Indian grammarian, is renowned for his introduction of a *kāraka* or deep-level analysis of sentence structure, which permits one to say that

(3) Rāma cooks rice for Sitā, and
(4) Rice is cooked by Rāma for Sitā

have the same deep structure, the inflected noun 'Rāma' in both sentences designating the logical 'agent', even though it has nominative case inflexion in (3) but instrumental case inflexion in (4). In his *Vyutpattivāda* Gadādhara adapts this method of sentence analysis, interpreting the *kārakas* in a manner akin to the 'participant roles' of modern case grammar, and making them the basis for a compositional account of the relation between word-meaning and sentence-meaning.

3 Legal theory

Like many late Indian logicians, Gadādhara wrote also on philosophy of law, reflecting the political importance of Hindu legal theory in Islamic, and later in British, India. Raghunātha, for example, argued that the older definition of property (*svatva*) as the capacity of a thing to be used at pleasure hid a circularity, for even if an object is owned there are still restrictions on its use, and these restrictions cannot be explicated without appeal to the notion of property itself. Raghunātha (?1500-50) proposed instead that property is a *sui generis* ontological category. Some later Naiyāyikas (apparently including Gadādhara) rejected this 'category' view in favour of a subjectivist definition of property as a legally grounded dispositional judgment 'this is mine', a definition which was itself subjected to further refinement (see Derrett 1956).

See also: Anaphora; Language, Indian theories of; Nyāya-Vaiśeṣika; Proper names

JONARDON GANERI

List of works

Gadādhara Bhaṭṭācārya (1604-1709) *The Gādādharī*, ed. V.P. Dvivedi *et al*., Varanasi: Chowkhamba Sanskrit Series Office, 1970.(A commentary on the *Tattvacintāmaṇidīdhiti*.)

Gadādhara Bhaṭṭācārya (1604-1709) *Vādavāridhi*, ed. B. Mishra and D. Shastri, Varanasi: Chowkhamba Sanskrit Series Office, 1933.(The collected tracts.)

Gadādhara Bhaṭṭācārya (1604-1709) *Vyutpattivāda (Theory of the Analysis of Sentence Meaning)*, trans. V.P. Bhatta, Delhi: Eastern Book Linkers, 1990.(Gadādhara's account of sentential semantic structure).

Gadādhara Bhaṭṭācārya (1604-1709) *Viṣayatāvāda*, trans. S. Bhattacharya, *Gadādhara's Theory of Objectivity, Containing the Text of Gadādhara's Viṣayatāvāda with an English Translation and Explanatory Notes*, Part II, Delhi: Motilal Banarsidass, 1990.(A clear presentation of Gadādhara's work on mental content and intentionality.)

Gadādhara Bhaṭṭācārya (1604-1709) *Vivāhavādārtha*, trans. J.D.M. Derrett, 'The Discussion of Marriage by Gadādhara: A Preliminary Investigation', *The Adyar Library Bulletin* 27 (1): 171-99, 1963.(Includes a study of the text; see also Derrett 1956.)

References and further reading

Annaṃbhaṭṭa (18th century) *Tarkasaṃgrahadīpikāon Tarkasaṃgraha (Commentary on the Compendium of Logic)*, ed. and trans. G. Bhattacharya, Calcutta: Progressive Publishers, 1983.(An introduction to Navya-Nyāya, sometimes called the *Bālagādādharī* or 'Gadādhara for Beginners'.)

Bhattacharya, D.C. (1951) *Bāōgalīr Sarasvat Avadīn, pratham bhāg, Baōge Navyanyāyacarccā (History of Navya-Nyāya in Bengal)*, Calcutta, 182-3. (Biographical details; in Bengali.)

Derrett, J.D.M. (1956) 'An Indian Contribution to the Study of Property', *Bulletin of the School of Oriental and African Studies* 18 (3): 475-98.

Ganeri, J. (forthcoming) *Meaning, Realism & Testimony: Philosophy of Language in Classical India*, Oxford: Oxford University Press.(A study of Gadādhara's theory in the context of contemporary philosophy of language.)

Ingalls, D.H.H. (1951) *Materials for the Study of Navya-nyāya Logic,* Cambridge, MA: Harvard University Press. (A work which marked the beginnings of the modern study of Navya-Nyāya.)

Tarkatīrtha, V. (1992) 'The Nyāya on the Meaning of some Words', *Journal of Indian Philosophy* 20 (1): 41-88. (A summary of Gadādhara's views on meaning.)

Gadamer, Hans-Georg (1900-)

Hans-Georg Gadamer is best known for his philosophical hermeneutics. Gadamer studied with Martin Heidegger during his preparation of Being and Time (1927). Like Heidegger, Gadamer rejects the idea of hermeneutics as merely a method for the human and historical sciences comparable to the method of the natural sciences. Philosophical hermeneutics is instead about a process of human understanding that is inevitably circular because we come to understand the whole through the parts and the parts through the whole. Understanding in this sense is not an 'act' that can be secured methodically and verified objectively. It is an 'event' or 'experience' that we undergo. It occurs paradigmatically in our experience of works of art and literature. But it also takes place in our disciplined and scholarly study of the works of other human beings in the humanities and social sciences. In each case, understanding brings self-understanding.

Philosophical hermeneutics advocates a mediated approach to self-understanding on the model of a conversation with the texts and works of others. The concept of dialogue employed here is one of question and answer and is taken from Plato. Such understanding never becomes absolute knowledge. It is finite because we remain conditioned by our historical situation, and partial because we are interested in the truth that we come to understand. By grounding understanding in language and dialogue as opposed to subjectivity, Gadamer's philosophical hermeneutics avoids the danger of arbitrariness in interpreting the works of others.

Gadamer's most important publication is Wahrheit and Methode (Truth and Method) (1960). He has also published four volumes of short works, Kleine Schriften (1967-77), containing important hermeneutical studies of Plato, Hegel, and Paul Celan among others. His many books and essays are collected into ten volumes (Gesammelte Werke). Gadamer is widely known as a teacher who practises the dialogue which is at the core of his philosophical hermeneutics.

1 Life

Gadamer was born in Marburg and brought up in Breslau, and his academic life mirrors the most important philosophical movements in Germany in the twentieth century. He encountered Neo-Kantianism when he studied philosophy at Marburg from 1918 to 1922, completing his doctoral dissertation in 1922 under the direction of Nicolai Hartmann and Paul Natorp (see Neo-Kantianism §§3-5). In 1923 he transferred to Freiburg to study phenomenology with Edmund Husserl and Martin Heidegger, whom he followed to Marburg in 1924. Gadamer completed his habilitation under Heidegger's sponsorship in 1929, just two years after Heidegger had published his existential phenomenology in *Being and Time* (1927). Gadamer was a precocious student, obtaining his doctorate and his habilitation at the unusually early ages of 22 and 29 respectively. While at the University of Marburg, Gadamer also studied classical philology and completed his state examination in philology in 1927. Gadamer's doctoral dissertation and habilitation were both on Plato. He published the latter, *Platos dialektische ethik*, in 1931.

Gadamer began teaching while still at Marburg. He also taught at Leipzig (1938-47), Frankfurt (1947-9), and Heidelberg, where in 1949 he assumed the chair in philosophy previously occupied by Karl Jaspers. After his retirement in 1968, Gadamer continued to lecture at the University of Heidelberg and began teaching part of the year in the USA, notably at Boston College.

In 1952 Gadamer and Helmut Kuhn founded the *Philosophische Rundschau*, which quickly became one of Germany's most important philosophy journals. Gadamer published relatively little after 1931, until he produced *Wahrheit und Methode (Truth and Method)* in 1960. Like Kant's *Critique of Pure Reason*, which was also published late in the author's life (Kant was 57), Gadamer's book immediately established its author's reputation. *Truth and Method* and *Being and Time* were recognized as the most important contributions to hermeneutics in the twentieth century. In the 1960s Gadamer became engaged in a lively debate with Jürgen Habermas and the Frankfurt School about hermeneutics and critical theory. In the 1970s, and in connection with his second teaching career in the USA, Gadamer was able to expand on the contribution that *Truth and Method* had made to practical philosophy, a move that brought him into contact with Americans influenced by pragmatism, such as Richard Bernstein and Richard Rorty. Later, in the 1980s, Gadamer initiated a debate with Jacques Derrida on hermeneutics and deconstruction.

2 *Truth and Method* and Kant’s aesthetics

Even in the earliest stages of his philosophical journey, when he was writing on Plato, Gadamer remained convinced that the experience of art was connected to philosophy. But it was not until *Truth and Method*, his late and mature work, that Gadamer makes this connection philosophically explicit. Gadamer maintains that the experience of art is an experience of truth as an ‘event’ that ‘ happens to us over and above our wanting and doing’ ([1960] 1975: xvi). This event structure of truth is, Gadamer believes, a universal characteristic of human understanding. The phenomenon of understanding - *Verstehen* - that concerns Gadamer should not be confused with *Verstand*, which is also translated as ‘understanding’. *Verstand*, most notably in Kant’s *Critique of Pure Reason*, is objectifying and means the natural scientific understanding subjects have of something other than themselves, namely, spatiotemporal objects (see Kant, I. §6). The phenomenon of understanding that Gadamer takes up in *Truth and Method* is that introduced by Heidegger in §32 of *Being and Time* (see Heidegger, M. §3). Understanding as *Verstehen* is non-objectifying. It is ontological and concerns a subject’s very being in the world. Understanding is once again of something other than oneself, but in this case it is of other possibilities for oneself as a being in the world.

Truth and Method begins when Gadamer takes up the question of truth as it emerges from the experience of art. The target of Gadamer’s criticism is Kant’s aesthetics and the abstraction of the art work from its original world and from our own concrete historical situation when we experience a work aesthetically. Gadamer’s goal in the first part of *Truth and Method* is to transform aesthetics by transcending the limits placed upon what counts as aesthetic experience and opening up the truth-claims made on us in this experience by works of art and literature. Gadamer shows how Kant’s *Critique of Judgment* subjectivizes aesthetics by narrowing the concept of aesthetic experience to the experience (*Erlebnis*) of the subject’s pleasurable ‘state of mind’ (the free play of the mental faculties) (see Kant, I. §12). Gadamer contrasts this narrowing to the role played earlier by judgment and taste in connection with culture and the formation of a *sensus communis* (‘common sense’). Gadamer proposes that we think instead of aesthetic experience on the model of *Erfahrung*, Hegel’s concept of experience. *Erfahrung* gives us something to understand. It challenges us so that we must ‘change our minds’. The subjective universality or ‘disinterested interest’ that Kant claims for aesthetic experience and the judgments we make about it cuts us off, Gadamer concludes, from the interest we take in the truth of works of art and literature in so far as they challenge us to change our minds and therefore our lives.

The experience of art offers Gadamer a model for how truth happens in human experience generally. Gadamer models human experience not on the sense perception of what is immediately given but on the experience of reading and understanding a text which has become historically distant from us. He reminds us of the practice of hermeneutics or interpretation in the reading and understanding of legal and religious texts that are temporally distant. These texts remain ‘classical’ (like those literary texts, ‘the classics’, that admit of no final interpretation) so long as they continue to challenge present-day readers to produce new interpretations of them. According to Gadamer, we understand these texts when we understand them to be making a true claim about how we are to understand something that is an issue for us in our own world and historical age. Like the experience we undergo with art, the event of understanding these texts changes our minds. Despite the prominent role played in *Truth and Method* by the experience of art, this book turns out not to be a work in aesthetics. Instead, *Truth and Method* transforms Kantian aesthetics into a philosophical hermeneutics, a general theory of understanding and interpretation that is also an ontology of ourselves as historical beings (see Hermeneutics §5).

3 Basic concepts of philosophical hermeneutics

Four interrelated concepts define Gadamer’s philosophical hermeneutics. The first of these is ‘effective-historical consciousness’ (*wirkungsgeschichtliches Bewusstsein*), which means that consciousness is at once ‘affected by history’ and ‘open to the effects of history’. Gadamer’s philosophical hermeneutics preserves Heidegger’s central insight in *Being and Time* that the understanding of oneself and one’s possibilities as a being in the world is temporal and situated historically. Understanding is therefore constituted in a positive way by its own historicality instead of being threatened by it, as Enlightenment thinkers after Descartes had feared. To be affected by history means that there can be no unmediated starting-point of the kind Descartes sought. To be open to the effects of history means to allow one’s present horizon of understanding to be called into question by texts and works of art handed down to us by tradition.

Gadamer retains Heidegger's concept of the 'hermeneutical circle' to express the way understanding is always mediated. Understanding is always circular, but in a non-vicious sense. Anticipation of meaning, preconceptions or prejudgments ('prejudices') play a constitutive role as we use them to project wholes out of the parts of the text or art work that we are interpreting. Gadamer's claim that prejudgments enable understanding, that is, positively condition it, has been mistaken to mean that he affirms prejudice blindly. But Gadamer does not mean that we must remain uncritical of those prejudices that block or distort our understanding of a text or art work. Rather, since we cannot get out of the hermeneutical circle of understanding and assume an unprejudiced view from nowhere, we must aim instead at becoming more conscious of our particular hermeneutical situatedness so that new and more appropriate judgments can play the role of older and less appropriate prejudgments in conditioning our understanding of the text or art work.

The third concept is that of 'play' (*Spiel*), which Gadamer borrows and transforms from Kant's aesthetics (the free play of the mental faculties). 'Play' or 'game' is used at two crucial points in *Truth and Method* to describe how his philosophical hermeneutics transcends the subjectivism of aesthetics and of modern philosophy in general. Gadamer speaks first of the game we play with art to show how individual subjects most lose themselves to experience the truth claimed by the art work, just as an individual player must in order to become absorbed in the game. And when, in Part 3, Gadamer turns to focus upon language as the medium of hermeneutical experience, he grounds language not in the consciousness of an individual subject but in the language game we call dialogue or conversation.

Language in the form of conversation leads us to the fourth concept, the 'fusion of horizons'. The understanding of a text or art work involves a challenge in so far as the truth claimed by the text or art work is at variance with what the interpreter believes to be the truth about the matter at issue ('*die Sache*'). The interpreter's horizon of understanding first excludes the truth claimed by the text or art work. The interpreter opens up to the horizon of the other (the text or art work) by allowing it to question the interpreter's own prejudgments about the matter at issue. What ensues is a dialogue of question and answer where the interpreter not only questions the truth claimed by the work but also allows what the interpreter prejudges to be true to be put into question by the work. The fusion of horizons that ends the dialogue occurs when the interpreter understands differently. This may require altering the interpreter's prejudgments in line with what has been learned from the text or art work. But it may also mean reaffirming the original prejudgments for different reasons since they have survived the challenge by another way to judge the matter at issue.

4 *Truth and Method* and Heidegger's existential phenomenology

Truth and Method also takes up the existential phenomenology of Heidegger's *Being and Time* and transforms it into a philosophical hermeneutics, and here too the experience of art is pivotal. Gadamer has always acknowledged that Heidegger's *The Origin of the Work of Art* (1936) played a vital role in his first formulation of the idea of a philosophical hermeneutics in the 1930s. In this essay, Heidegger asserts that the truth which an art work discloses emerges out of the strife between 'earth' and 'world'. What strikes Gadamer as so insightful about Heidegger's characterization of the experience of art as the 'event of truth' is that 'earth' becomes a 'necessary determination of Being of the work of art' (1994: 100). 'World', Gadamer notes, is a familiar concept from *Being and Time*. 'Earth', however, introduces an altogether new conceptuality, one drawn from the poetry of Friedrich Hölderlin. 'Earth' (*Erde*) does not mean 'matter' (*Stoff*), but that which continues to conceal itself even though it is that out of which everything emerges and into which everything retires.

Although Gadamer underplays this, Heidegger concentrates almost exclusively on the poetry of Hölderlin in the Nazi period, because he believes it to be history-making, that is, capable of inaugurating a new epoch of European history, one identified from now on as German, just as poetry had inaugurated the Greek beginning of European history. Gadamer nowhere shares Heidegger's belief in the need to inaugurate a new beginning. On the contrary, what is necessary is a creative renewal of the truth-claims made by the tradition and history of Europe's past. Gadamer aims, therefore, both to retain what is insightful about Heidegger's turn to art, poetry and language in his works after *Being and Time,* and to redirect it to the past. Accordingly, Gadamer reinterprets 'earth', which in Heidegger is associated with the German country and land, so that it takes on the meaning of the European tradition in which Gadamer finds himself embedded. This tradition is like 'earth' in that it continues to conceal resources within itself. But for Gadamer these become resources for renewal instead of unmediated sources of the

altogether new.

Not surprisingly, we find that Hölderlin's poetry plays only a minor role in Gadamer's thinking, compared to the importance Gadamer continues to attach to the Plato's dialogues and Aristotle's practical philosophy. Plato's concept of dialogue, for example, lives on in Gadamer's concept of language in and as a conversation with the text. And Aristotle's concept of practical wisdom (*phronesis*) survives and flourishes in Gadamer's solution to the problem of the application of our understanding of a text to our particular situation (see Aristotle §23; Plato §4).

5 Criticism and self-criticism

Gadamer has continued to define his hermeneutical philosophy in response to questions raised about *Truth and Method*. Emilio Betti (1980) and E.D. Hirsch (1967) were the first to object to Gadamer's philosophical hermeneutics in the belief that it failed to provide a method for validating interpretation. Habermas questioned instead Gadamer's rehabilitation of prejudice and the authority of tradition (see Frankfurt School §4). Derrida has raised doubts about Gadamer's reliance on a hermeneutics of good will after Nietzsche's hermeneutics of suspicion (see Nietzsche, F. §7). In each case, Gadamer has used these objections as an occasion to expand upon the scope of his hermeneutical philosophy and clarify its potential.

In 1992, in reflecting on his philosophical journey during the better part of the twentieth century, Gadamer acknowledges the validity of the following criticisms of *Truth and Method*. He admits that his concept of scientific method did not reflect the nuanced and complex concept of method that had grown up in the natural sciences between the 1930s and 1960. He grants that *Truth and Method* did not sufficiently differentiate the historian's work of interpreting a historical text from the philologist's work of interpreting a literary and classical text. Finally, he recognizes a potential problem with his concept of play or game. In conversation, the game of language that mediates understanding, we transcend our subjectivity dialogically by changing our self-understanding along with changing our understanding of the truth claimed by the other. In the experience of art, however, we seem to transcend subjectivity in a more one-sided way. The game we engage in with art challenges us to change our minds and ourselves without apparently being able at the same time to reciprocate and challenge the work's claim to truth.

Gadamer's own self-criticism has led him to renew the question of the relation between philosophy and art or literature. Gadamer argues that there is indeed reciprocity in the game that we engage in with art. Just as the art work challenges us to change our minds, so too our interpretation of the art work challenges it to produce a response to our questions. By interpreting the art work, we do not merely reproduce the truth claimed by the art work but we produce a different understanding of its claim to truth.

Ironically, once play and gaming became associated with post-structuralist theories of interpretation, especially that of Derrida, Gadamer's attempt to rethink his earlier concept of the game we play with the art work risks being misunderstood as an affirmation of the very subjectivity in understanding and interpretation that *Truth and Method* sets out to deny. Gadamer attempted but failed to engage successfully in a dialogue with Derrida on the concept of the interpretation of a text that appears to divide the hermeneutical renewal of tradition from its deconstruction (Michelfelder and Palmer 1989). The confrontation of hermeneutics and deconstruction, both legacies of Heidegger's later philosophy, remains a task for the future.

KATHLEEN WRIGHT

List of works

Gadamer, H.-G. (1985-) *Gesammelte Werke*, Tübingen: J.C.B. Mohr (Paul Siebeck), 10 vols.(Gadamer's *Collected Works*: vols 1 and 2 are on hermeneutics; vols 3 and 4 on recent philosophy; vols 8 and 9 on aesthetics and poetics; vol. 10 on hermeneutics in retrospective.)

Gadamer, H.-G. (1931) *Platos dialektische Ethik: Phänomenologische Interpretationen zum Philebos*, in *Platos dialektische Ethik*, Hamburg: Felix Meiner, 1968; trans. R.M. Wallace, *Plato's Dialectical Ethics*, New Haven, CT: Yale University Press, 1991.(Gadamer's 1931 habilitation, a phenomenological interpretation of Plato's *Philebus*.)

Gadamer, H.-G. (1960) *Wahrheit und Methode*, Tübingen: J.C.B. Mohr (Paul Siebeck); trans. *Truth and Method*, New York: Sheed & Ward, 1975; 2nd edn, rev. trans. J. Weinsheimer and D.G. Marshall, 1989.(Gadamer's philosophical hermeneutics. The original translation was edited by G. Barden and J. Cumming in 1975. The

later, revised translation is preferred.)
Gadamer, H.-G. (1967a) *Hegels Dialektik*, Tübingen: J.C.B. Mohr; trans. P. Christopher Smith, *Hegel's Dialectic*, New Haven, CT: Yale University Press.(Five hermeneutical studies.)
Gadamer, H.-G. (1967b) *Kleine Schriften I: Philosophie, Hermeneutik*, Tübingen: J.C.B. Mohr (Paul Siebeck). (These essays address early formulations and further developments of Gadamer's philosophical hermeneutics.)
Gadamer, H.-G. (1976) *Philosophical Hermeneutics*, trans. D. Linge, Berkeley, CA: University of California Press.(Essays developing Gadamer's hermeneutics.)
Gadamer, H.-G. (1972) *Kleine Schriften III: Idee und Sprache*, Tübingen: J.C.B. Mohr (Paul Siebeck), 2nd edn. (These essays interpret the thought of Plato, Husserl, Heidegger and Hegel, among others.)
Gadamer, H.-G. (1972) *Kleine Schriften IV: Variationen*, Tübingen: J.C.B. Mohr (Paul Siebeck), 2nd edn.(These essays respond to further questions about the philosophical hermeneutics presented in *Truth and Method*.)
Gadamer, H.-G. (1979) *Kleine Schriften II: Interpretationen*, Tübingen: J.C.B. Mohr (Paul Siebeck), 2nd edn. (These essays exemplify the practice of hermeneutics in the domain of aesthetics.)
Gadamer, H.-G. (1980) *Dialogue and Dialectic*, trans. P. Christopher Smith, New Haven, CT: Yale University Press.(Eight hermeneutical studies of Plato.)
Gadamer, H.-G. (1981) *Reason in the Age of Science*, trans. F.G. Lawrence, Cambridge, MA: MIT Press.(Essays on practical philosophy.)
Gadamer, H.-G. (1986) *The Relevance of the Beautiful*, trans. N. Walker, ed. R. Bernasconi, Cambridge: Cambridge University Press.(Essays on art, literature and philosophy.)
Gadamer, H.-G. (1994) *Heidegger's Ways*, trans. J. Stanley, Albany, NY: State University of New York Press. (Essays on Heidegger.)

References and further reading

Betti, E. (1980) 'Hermeneutics as the General Methodology of the *Geisteswissenschaften*', in *Contemporary Hermeneutics*, ed. J. Bleicher, London: Routledge & Kegan Paul.(Criticism of Gadamer's hermeneutics.)
Grondin, J. (1982) *Hermeneutische Wahrheit?*, Königstein: Forum Academicum. (Gadamer's concept of truth.)
Habermas, J. (1977) 'A Review of Gadamer's *Truth and Method*', in *Understanding and Social Inquiry*, ed. F. Dallmayr and T. McCarthy, Notre Dame, IN: Notre Dame University Press.(Discussion from the perspective of critical theory.)
Heidegger, M. (1927) *Being and Time*, New York: Harper & Row, 1962.(The first major contribution to hermeneutics in the twentieth century.)
Heidegger, M. (1977) 'The Origin of the Work of Art', in *Basic Writing*, ed. D.F. Krell, New York: Harper & Row.(A major influence on Gadamer's hermeneutics.)
Hirsch, E.D. (1967) *Validity in Interpretation*, New Haven, CT: Yale University Press. (Criticism of Gadamer's hermeneutics.)
Hoy, D. (1978) *The Critical Circle*, Berkeley, CA: University of California Press.(On Gadamer and Hirsch, Habermas and Derrida.)
Michelfelder, D. and Palmer, R. (eds) (1989) *Dialogue and Deconstruction*, Albany, NY: State University of New York Press.(The 1981 encounter between Gadamer and Derrida.)
Palmer, R. (1969) *Hermeneutics*, Evanston, IL: Northwestern University Press.(The interpretation theory of Schleiermacher, Dilthey, Heidegger and Gadamer.)
Wachterhauser, B. (ed.) (1986) *Hermeneutics and Modern Philosophy*, Albany, NY: State University of New York Press.(Important essays on Gadamer.)
Warnke, G. (1987) *Gadamer: Hermeneutics, Tradition and Reason*, Stanford, CA: Stanford University Press. (Gadamer, Frankfurt School, and pragmatism.)
Weinsheimer, J. (1985) *Gadamer's Hermeneutics: A Reading of Truth and Method*, New Haven, CT: Yale University Press.(A useful commentary.)
Wright, K. (ed.) (1990) *Festivals of Interpretation*, Albany, NY: State University of New York Press.(Essays on concepts central to Gadamer's philosophical hermeneutics.)

Gaius (*c*.110-*c*.180)

*Gaius was a great but not a typical classical Roman jurist. His Institutes (*c.AD *161), an introductory textbook, is the first known attempt to see Roman law as a systematic whole. His scheme was used by Justinian and so played a central role in subsequent European thought on the classification of law.*

Little is known of the Roman jurist Gaius apart from his writings. An adherent of the Sabinian school at Rome, he seems to have lived in the provinces. Gaius was not typical of the jurists of the classical age (*c*. 50 BC-*c*.AD 250), who do not cite his opinions. Unlike them, he held no political office. Evidently a teacher of law rather than a practising lawyer, his writings show an interest in legal history, Greek philosophy and the classification of law. None of these was characteristic of the other jurists, who clung to a tradition of clever argument in individual cases, the casuistic approach.

Extracts from his many works in the *Digest* show that Gaius was also a master of practical legal reasoning in this sense. But his *Institutes*, a short introductory textbook written about AD 161, is unique. The only classical law book to have survived complete and unaltered by Justinian (see Justinian), a manuscript of it was found by the German historian Niebuhr in 1816. After some opening remarks on the sources of law, Gaius presents his famous tripartite division of private law into persons (personal status), things (property, succession and obligations) and actions (forms of action and procedure). There are many subdivisions until, for example, the level of individual contracts is reached. Out of a great heap of statutes, edicts, opinions and actions Gaius produced for his students a clear map of the law.

Was Gaius an original thinker? Some have argued that he borrowed his system from an earlier legal work, but no such book is known to exist. And when, for example, Gaius deals with the classification of obligations, he sounds like someone breaking new ground. His debt to Greek philosophy raises a different question. That he was familiar with some common philosophical ideas is suggested, for instance, by his many reference to what is ‘natural’ in relation to law. But even if he did sometimes adapt and apply classifications he found in Aristotle and elsewhere to the intricacies of actual Roman law, this in itself was an original achievement. Gaius deserves to be considered a great lawyer, an excellent teacher and a brilliant inventor.

Gaius’ merits were eventually recognized in the later Empire, when his *Institutes* became a standard work in the law schools. Justinian calls him ‘our Gaius’ and based his own *Institutes* squarely on the work of Gaius. In this way Gaius became the teacher of Europe. The institutional scheme was not perfect, but it was there. Through it, Gaius can claim to be the father of systematic legal thinking in the West.

See also: Law, philosophy of; Legal concepts §1; Roman law

GRANT McLEOD

List of works

Gaius (*c*.AD 161) *The Institutes of Gaius*, F. de Zulueta (ed.), Oxford: Oxford University Press, 1946, 1953; W.M. Gordon and P.F. Robinson (eds), London: Duckworth, 1988.(In the former part 1 is the text with critical notes and a translation, part 2 an extensive and useful commentary. The latter edition has the text with a readable, modern translation.)

References and further reading

Honoré, A.M. (1962) *Gaius*, Oxford: Oxford University Press.(Brilliant but speculative account of Gaius’ life. Technical in places.)

Jolowicz, H.F. and Nicholas, B. (1972) *Historical Introduction to the Study of Roman Law*, Cambridge: Cambridge University Press, esp. 374-94, 405-20.(An accessible account, with extensive references to the literature, of Gaius and the classical Roman jurists.)

Justinian (AD 533) *The Digest of Justinian*, T. Mommsen, P. Krueger and A.Watson (eds), Philadelphia, PA: University of Pennsylvania Press, 1985.(Contains excerpts from Gaius’ other works. The Latin text by Mommsen and Krueger with a readable translation by various writers edited by A. Watson.)

Justinian (AD 533) *Justinian’s Institutes*, P. Birks and G. McLeod (eds), London: Duckworth, 1987.(The Latin text

by Krueger with a good modern translation and useful introduction.)

Galen (AD 129-*c*.210)

Galen was the most influential doctor of late Greco-Roman antiquity. But he was also a notable philosopher, who desired to effect a synthesis of what was best in the work of his predecessors, not only in medicine but also in logic, epistemology, philosophical psychology and the philosophy of science and explanation. In logic he made use of both Aristotelian and Stoic material, but supplemented them with his own treatment of relational logic. His epistemology, while resolutely anti-sceptical on the grounds that nature could not have furnished us with systematically delusive sense-organs, was none the less sober and cautious: some philosophers' questions are simply unanswerable. He attacked the Stoics' unitary psychology, establishing by means of detailed experiments that the brain was the source of voluntary action. Finally, drawing on the philosophical and medical tradition, he crafted a theory of cause and explanation sophisticated enough to rebut the sceptical challenges to such notions, and rich enough to enable him to construct a comprehensive physiology and pathology on its basis.

1 Life and work

Galen is best known to history as a doctor, the systematizer of a Hippocratism (see Hippocratic medicine) which became the foundation of medicine for more than 1,500 years. But he was also a considerable philosopher in his own right, writing extensively on logic and language, and scattering philosophical remarks, arguments and controversies throughout his vast medical *oeuvre*. And this is no mere accidental collocation of interests, to be explained by the fact that he trained first in philosophy before turning to medicine. He wrote a short pamphlet *That the Best Doctor is Also a Philosopher*, which survives, stressing the intimate connections between the two disciplines, and never tired of pointing out the errors made by his medical colleagues caused by their insufficient attention to logical detail and their ethical shortcomings.

Galen was born into a well-to-do family in Greek Asia Minor, and received an extensive liberal education, learning philosophy from leading Platonists and Peripatetics, as well as Stoic logic. He had little time for atomism in any of its forms (see Atomism, ancient), and is scathing about the inability of the atomists, and doctors influenced by them, to give satisfying explanations of complex phenomena in purely material terms without recourse to teleological explanation. For Galen, teleology was an essential part of any rationally acceptable explanation of the world and its contents.

Galen travelled widely in pursuit of his education, absorbing both Empiricist and Rationalist views (see Hellenistic medical epistemology), which conspired to nurture his distinctively eclectic views in methodology, as well as a syncretism in philosophical outlook typical of the times (although original enough in its elaboration). He moved to Rome in 162, making (on his own account, in *On Prognosis*) an immediate impact upon the cultural and aristocratic elite for his competence as a doctor and for the soundness of his logical reasoning.

He wrote at a furious pace, dictating to relays of scribes, and his output of treatises medical and philosophical, in the form of original monographs or of commentaries on the masters was correspondingly enormous (see §2 below). Some of this *oeuvre* was destroyed by fire in his own lifetime, and much else has been lost since. But a great deal survives in Greek (Kühn's monumental edition runs to about 10,000 pages of text), and much else in other languages, notably Arabic, Latin and Hebrew, a fact ensured by his early acceptance in the Islamic world as the great repository of medical wisdom. His influence on the subsequent history of medicine was enormous, his views in physiology and anatomy not being superseded until the sixteenth century, his medicine itself surviving even longer.

2 Galen's intellectual inheritance

Galen's chief sources of inspiration were Hippocrates and Plato, his *On the Doctrines of Hippocrates and Plato* being a typically syncretistic attempt to demonstrate their basic agreement on all important issues. His theory of disease is fundamentally Hippocratic in conception: an illness consists in the impediment to or destruction of one of the crucial functions of the body, while treatment involves restoring that function to its proper state. These impediments can take the form of imbalances in the fundamental four humours, blood, phlegm, and yellow and black bile, which regulate the body; and each of the four humours is itself associated with one pair of the four basic qualities, hot, cold, wet and dry, which may by distempered either individually, or in their humoral pairs.

Furthermore, there are four types of disease of the organic parts, plus a general category he terms ‘loss of cohesion’ (this scheme forms the basis of *On the Therapeutic Method*). However, for Galen, Hippocrates was important not so much as a medical gospel, but rather for having exemplified a methodology of investigation, theorizing, and testing, which, if competently pursued, would in time reveal the whole of medical truth.

Galen’s greatest philosophical debt to Plato is his adoption and elaboration of the type of creationist natural teleology sketched in the latter’s *Timaeus* (see Plato §16; Teleology). For Galen, this time in agreement with the Stoics (see Stoicism §5), the idea that the world, in all its structural regularity, might be a spontaneous efflorescence of blind mechanical forces was simply anathema: no mechanical theory could possibly account for the complexity and organization of the world and its inhabitants. Moreover, the artifice of such a world demands an actual Creator: Aristotelian purely immanentist teleology is equally unsatisfactory. Galen calls his own great teleological account of biological structure and function, *On the Function of the Parts*, a ’hymn to Nature’ and the Creator. Those (such as Erasistratus; see Hellenistic medical epistemology), who have dared to say that Nature on occasion does something in vain have simply failed to appreciate the subtlety and complexity of her works, which are only apparent to the skilled and conscientious investigator.

Galen also seeks in *On the Doctrines of Hippocrates and Plato* to ground the Platonic hypothesis of a tripartite soul, with reason in the brain, emotion in the heart and desire in the liver, upon detailed anatomical investigation and experiment. He subjects the unitary psychology of the Stoics (see Stoicism §19) to a withering attack, accusing them of logical incompetence as well as blindness to the evidence. Yet Galen was equally powerfully influenced by the Peripatetic tradition in natural science: while the underlying pattern to his teleology is Platonic, the great model for his practical work in the investigation of function is Aristotle’s biology. And he admired Aristotle’s logic, sometimes troubling to cast his arguments in syllogistic form (but see §4 below).

3 Theory of causation

Equally, Galen happily adopts the basic model of Aristotle’s four causes (*Physics* II 3) and deploys it in his own analysis (see Aristotle §9). There is ample room for the final cause in nature given his commitment to teleology; efficient causes clearly have a role in explaining how things come to be what they are; while a just appreciation of the importance of the material cause will help to evade some of the stronger sceptical attacks (notably again those of Erasistratus) upon the coherence of the notion of causation in the first place. The reason why different things respond differently to the same influences (the prime empirical weapon in the sceptical armoury) is that they are differently constituted, and so differently susceptible (a view perfectly consistent with Hippocratism: the different constitutions, at least in the case of animals, will be humoral in structure).

Yet Galen’s use of Aristotle is selective and creative. He has very little use for the formal cause (or indeed for Platonic hypostasized Forms) as such, although he is perfectly well aware of the importance of form for function. He simply does not regard it as a properly separate explanatory category; instead he adopts the Middle Platonist category of the instrumental cause.

Equally, his treatment of efficient causes is unorthodox, since he happily embraces the Stoic (or at least Stoic-influenced) treatment of causal agency as involving categorially distinct stages. There are antecedent causes, external influences brought to bear on objects (or physical systems), which may, if they are suitably disposed, trigger in them a sequence of events leading to the establishment of some further condition (for example, disease). But antecedent causes are not on their own sufficient for their effects; they operate only on patients in a suitable condition, and even then their progress may, if caught early enough, be arrested. For they set in motion processes within the organism, a sequence of preceding causes, which then culminate (unless appropriate action has been taken) in a set of conditions both necessary and sufficient for the disease in question, as well as being contemporary with it: the so-called ’containing’, or sustaining, cause.

Thus Galen’s detailed account of causal efficiency involves a wholly un-Aristotelian concentration on the sequential nature of causation as a temporal phenomenon; and, in addition to being of sufficient complexity and sophistication to defuse sceptical objections, it allows him to build up a detailed theoretical account of the structure of human pathology.

4 Reason and experience

Galen's commitment to theory places him in the Rationalist camp (see Hellenistic medical epistemology); but he insisted with the Empiricists upon the importance of testing, and sought to effect a reconciliation: theory without experience was empty, but experience without theory, blind. He emphasizes the importance of practical engagement in physiology and medicine; of regular, repeated observation and experiment; but he also believes that such observation is hamstrung unless it can be placed within its appropriate theoretical context. Students of medicine must make repeated dissections and vivisections in order to learn the structure and arrangement of things in the body; but they must also learn to sytematize those observations by attending to logic, and to understand them in nature's teleological light.

The basic theoretical categories he adopts, the four qualities, while not directly equivalent to their phenomenal counterparts, are still susceptible of empirical appreciation. Chilled white wine is, surface indications notwithstanding, hot and dry, since its effect is to heat and desiccate. Empirical testing, the main tool in the Empiricist context of discovery, is equally essential for Galen, but rather in the context of justification. The theories elaborated by reason on the basis of observation must themselves answer at the tribunal of experience.

Galen never abandons his belief (stated in the opening books of *On the Therapeutic Method*) that medicine can be made into an Aristotelian demonstrative science, where necessarily true theorems flow ineluctably from indubitable axioms, which include definitions stating the essence of things as well as fundamental logical and metaphysical principles, such as the law of the excluded middle and the principle of causality. Yet for such a science to be of practical use, it must be applicable to contingent circumstances, and pass stiff empirical tests. Moreover, given the diversity of circumstances, practical medicine will never be infallible; but if it is probabilistic, it is so not because of any intrinsic uncertainty in nature itself, but only because of our own inabilities to see clearly or accurately enough into it.

Galen insisted on the importance of training in logic, for systematizing empirical evidence and exposing others' fallacies. He made use of Stoic as well as Aristotelian argument schemata (although not uncritically: he rejects the Stoic third 'indemonstrable' (see Stoicism §11) as incapable of delivering necessary conclusions); and he saw more clearly than either school that neither system could handle relations properly, a deficiency he sought, albeit rather naïvely, to remedy with an account of 'relational syllogisms'.

Galen was not overly sanguine about knowledge; indeed he held that many of the disputes of the theoretical philosophers and cosmologists, over the nature of the soul or the existence of an extra-mundane void, were simply undecidable, since reason alone cannot decide them and neither is there any relevant experience to be won regarding them. But where they can operate, the sense-organs are 'natural criteria'; and nature does nothing in vain. Thus animals must naturally be able to assimilate perceptible form, and no inductive problem of the sort which plagues empiricist accounts of concept-formation arises. Equally, scepticism is refuted by teleology. It is by analysis of these naturally arising conceptions that we may, if we are able and conscientious enough, arrive at the real, definitional principles of things upon which all genuine, scientific knowledge is based.

R.J. HANKINSON

List of works

Galen (*c.* AD 150-210) *Opera omnia*, ed. C.G. Kühn, Leipzig: Knoblauch, 1821-33.(Greek text with Latin translation; Galen is standardly cited by volume and page number from this edition.)

Galen (*c.* AD 150-210) *Selected Works*, trans. P.N. Singer, Oxford and New York: Oxford University Press, 1997. (A broad range of works.)

Galen (*c.* AD 170) *On the Function of the Parts*, trans. M.T. May, *Galen on the Usefulness of the Parts of the Body*, Baltimore, MD: Johns Hopkins University Press, 2 vols, 1967.(Translation with useful notes of Galen's defence of natural teleology.)

Galen (*c.* AD 175) *On the Therapeutic Method, Books I and II*, trans. R.J. Hankinson, Oxford: Oxford University Press, 1991.(Translation and commentary, with introduction and notes, of Galen's most extended discussion of scientific method.)

Galen (*c.* AD 175) *On Antecedent Causes*, trans. R.J. Hankinson, Cambridge: Cambridge University Press, 1998. (New edition, with translation, commentary and general introduction, of Galen's important short treatise on causation.)

Galen (*c.* AD 175) *On the Doctrines of Hippocrates and Plato*, trans. P.H. De Lacy, *Galeni de Placitis Hippocratis*

et Platonis, in *Corpus Medicorum Graecorum* V 4 2 1, Berlin: Akademie-Verlag, 3 vols, 1978-83.(Magisterial edition and translation, with short commentary, on Galen's great text.)

Galen (*c.* AD 178) *On Prognosis*, trans. V. Nutton, *Galeni de Praecognitione*, in *Corpus Medicorum Graecorum* V 4 2 1, Berlin: Akademie-Verlag, 1979.(Fine edition, with commentary, of an interesting text in which Galen recounts many of his early medical triumphs in Rome.)

Galen (*c.* AD 190, 195, 150) *Three Treatises on the Nature of Science: On the Sects for Beginners, An Outline of Empiricism, On Medical Experience*, trans. M. Frede, Indianapolis, IN: Hackett Publishing Company, 1985. (Clear translations of three key short works on medical methodology, with an excellent introduction.)

Galen (*c.* AD 195) *On Anatomical Procedures*, trans. C. Singer, Oxford: Oxford University Press, 1956; later books trans. W.H.L. Duckworth, Cambridge: Cambridge University Press, 1962.(Complete English version of Galen's major treatise on anatomy.)

References and further reading

Barnes, J. (1991) 'Galen on Logic and Therapy', in R. Durling and F. Kudlien (eds) *Galen on the Method of Healing*, Leiden: Brill, 50-102.(A magisterial account of the philosophical basis of Galen's medical theory.)

Frede, M. (1982) 'On Galen's Epistemology', in V. Nutton (ed.) *Galen: Problems and Propects*, London: Wellcome Institute, 65-86.(A brief, comprehensive overview.)

Hankinson, R.J. (1989) 'Galen and the Best of All Possible Worlds', *Classical Quarterly* 39: 206-27.(Discussion of Galen's creationist teleology; see §2.)

Hankinson, R.J. (1992) 'Galen's Philosophical Eclecticism', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36 5: 3,505-22. (Brief, general account of Galen's philosophy and its sources.)

Hankinson, R.J. (1994a) 'Galen's Anatomical Procedures', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 37 2: 1,834-55. (An account of the theoretical basis of Galen's anatomy, and its relation to his teleology.)

Hankinson, R.J. (1994b) 'Galen's Theory of Causation', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 37 2: 1,757-74. (Expanded version of §3.)

Temkin, O. (1973) *Galenism*, Ithaca, NY: Cornell University Press.(A good general account of Galen's medical views and their afterlife.)

Galilei, Galileo (1564-1642)

Galileo Galilei, one of the most colourful figures in the long history of the natural sciences, is remembered best today for two quite different sorts of reason. He has often been described as the 'father' of modern natural science because of his achievements in the fields of mechanics and astronomy, and for what today would be called his philosophy of science, his vision of how the practice of science should be carried on and what a completed piece of natural science should look like. While none of the elements of that philosophy was entirely new, the way in which he combined them was so effective that it did much to shape all that came after in the sciences. In the popular mind, however, as a continuing stream of biographies attest, it is his struggle with Church authority that remains the centre of attention, symbolic as it is of the often troubled, but always intriguing, relationship between science and religion.

1 Life and works

Born in Pisa, Galileo enrolled in 1581 in medicine at the University of Pisa. But his interests lay rather in mathematics and he left the university without completing a degree. He continued his work in mathematics, however, and obtained a teaching post in mathematics and natural philosophy at Pisa in 1589. To aid his teaching, he composed several commentaries on standard logical topics in the Aristotelian curriculum of the day, drawing heavily (as Wallace (1992a) has persuasively argued) on class-notes from lectures at the Jesuit Collegio Romano. Around this time, he also composed a treatise on motion which departed from Aristotle in some significant respects, but supported Aristotelian doctrine against its critics in the main, holding, for example, that the speed of natural fall, after a brief initial period, is constant and that the motion of a body on a frictionless plane will eventually cease.

In 1592, he gained a teaching position at the University of Padua, then perhaps the leading centre in natural philosophy. There he continued his work in mechanics along lines that diverged more and more from Aristotelian norms. Aided by experiments with an inclined plane, he concluded that free fall is uniformly accelerated. The mathematics of uniform acceleration had already been explored by the Merton school at Oxford (see Oxford Calculators); Galileo applied this mathematics to falling motion, with the crucial proviso that the motion be idealized by supposing the resistance of the medium to be removed. There is evidence, as Drake and others have shown, that he also discovered the parabolic shape of projectile motion by empirical means at this time.

In 1609, Galileo turned his newly constructed telescope to the skies and in the short space of a few years made a series of remarkable discoveries that altogether changed the direction of his researches and of his life generally. He announced the earth-like nature of the lunar surface, the existence of sunspots and the apparent axial rotation of the sun, the four moons of Jupiter (which he named the 'Medicean' planets in honour of the Medici ruler of Florence, Cosimo II) and the phases of Venus which showed it to revolve around the sun. His first book-length publication, the *Sidereus Nuncius (The Starry Messenger)* (1610), was an instant best-seller, provoking heated debate. His new-found fame led to the appointment he had so eagerly sought, as court mathematician and philosopher to Cosimo.

Galileo's discoveries definitively undermined the Aristotelian geocentric world-system (see Cosmology). Though he took this to affirm its Copernican rival, there was, in fact, another alternative, the Tychonic model which had the planets orbit the sun and the sun in turn revolve around an immobile earth. Galileo's intuitions led him to reject this model as dynamically absurd, but it was greeted warmly by those who wanted to maintain the stability of the earth either for physical or for theological reasons.

Galileo soon found himself opposed not only by his beleaguered Aristotelian colleagues but by the theologians whose aid the latter solicited. A number of biblical passages referred, directly or indirectly, to the motion of the sun or the stability of the earth. In a letter to his former student Benedetto Castelli, and in a much longer *Letter* (1615) addressed nominally to Cosimo's mother, the dowager Grand Duchess Christina, Galileo formulated a set of hermeneutic principles addressed to cases where science and Scripture appear to clash. All of the principles found a precedent, he emphasized, in St Augustine's *De Genesi ad litteram*. But his arguments were in vain. In 1616, the Roman Congregation of the Index banned Copernicus' *De revolutionibus*, until the passages implying the reality of the earth's motion and the sun's rest should be 'corrected', and asserted in passing that the

Copernican theses, taken literally, were ‘contrary to Scripture’. Galileo was summoned before Cardinal Bellarmine and warned to abandon the Copernican claims. Whether he also received the further injunction not to teach, defend or even discuss the Copernican theses, which was to be delivered only if he proved recalcitrant, has been much debated; the documentary evidence on this issue is clouded.

Burdened by recurrent illness, Galileo returned to the astronomical topics he had already addressed in his *Letters on Sunspots* (1613). He turned now to comets and engaged in an increasingly contentious dispute about their nature with the Jesuit natural philosopher Orazio Grassi, culminating in the publication of *The Assayer* (1623). It was here that Galileo broke entirely with Aristotelian natural philosophy, maintaining that the Book of Nature is written in the language of mathematics, that the qualitative properties of sensible objects are merely subjective, and that matter is composed of imperceptible corpuscles.

In 1623, Galileo’s friend, Maffeo Barberini, was elected Pope as Urban VIII. In a series of interviews with the new Pope, Galileo won an evidently reluctant permission to proceed with a treatise on the Copernican system, provided it be treated as a hypothesis. The traditional understanding of mathematical astronomy was that mathematical ’hypotheses’, like those of Ptolemy and Copernicus, were to be regarded as convenient predictive devices only. Furthermore, for theological reasons rooted in the earlier nominalist-voluntarist tradition, Urban himself maintained that the underlying causes of a physical phenomenon could never be established demonstratively, as the Aristotelian notion of science required, since God could always bring about the phenomenon by other hidden means. But Galileo evidently considered himself licensed by Urban to argue for the possible *truth* of the Copernican claims, understanding ’hypothesis’ now in a very different sense, a sense closer to the modern one. And so he began work on a treatise that would make the best case possible for Copernicanism, stopping short only of claiming it had been demonstrated. After a lengthy struggle with the ecclesiastical censors in Rome and in Florence, he was finally permitted to publish the *Dialogue Concerning the Two Chief World Systems* in 1632.

The book immediately encountered strenuous opposition from Roman theologians, most particularly from the Pope himself, principally on the grounds that it contravened the decree of 1616 declaring the Copernican doctrine to be contrary to Scripture. The Pope ordered Galileo to be brought to trial before the Holy Office in 1633. Found ’suspect of heresy’, he was ordered to recant. After his recantation he was sentenced to imprisonment, commuted to house arrest.

Deeply discouraged and forced to abandon his cherished project of showing that the Church had nothing to fear from Copernicanism, Galileo returned to the treatise on motion he had commenced long before in Padua. It still needed to be given a literary form, and a properly axiomatic structure had to be found for the two laws of motion. The completed manuscript was smuggled to Holland and the *Two New Sciences* was finally published in 1638. It announced ’a brand new science concerning a very old subject’, and set forth in elegant geometrical form his two laws of motion. It was his final testament, one that would ’open wide a gateway’ for the labours of others, as he prophesied that it would.

2 The practice of natural science

The most striking feature of Galilean science was its emphasis on *mathematization*. Not only did it treat quantity to the exclusion of quality, but quantity itself was assumed to be expressible in purely geometric terms. Time was represented as a line, and velocity, instead of being treated in the traditional way as an irreducible ratio between distance covered and time taken, was presented as a mathematical quantity in its own right. Galileo’s two laws were both purely kinematic; they simply described motion. Quantities like mass or force would not have lent themselves to the simple sort of geometrization that length and time allowed. (Descartes achieved a similar geometrical reduction by equating matter with extension.) Furthermore, the axiomatic mode of presentation customary for geometry encouraged Galileo (as it later did Newton) to adopt a similar mode for mechanics. Galileo’s success in dispensing with weight and resistance of the medium, the two non-geometrical factors determining velocity in older accounts of falling motion, may have led him to be over-optimistic about the prospects for a reductively geometric science of motion. But his resolute move to mathematically expressible quantities as the proper basis of physics was to prove its long-term worth.

What made both of his laws possible was his decision to *idealize*, that is, to extrapolate to the (at the time empirically unachievable) case of *in vacuo* fall. The medium invariably present in actually observed fall was to be

treated as an ‘impediment’, a complicating factor that should as far as possible be eliminated, in order to focus on the ’pure’ case of fall. Though conceptual idealization was not new in natural philosophy, finding a precedent especially in the work of Archimedes, the construction of deliberately simplified material situations was a distinctively Galilean contribution to the science of motion. Galileo was convinced that it would lead to the mathematically simple laws that lie hidden beneath the complexities of natural process. This aspect of his practice has often been described as ’Platonic’, but it runs contrary to the Platonic conviction that the sense-world, far from being ’written in the language of mathematics’ as Galileo’s famous slogan ran, lends itself only imperfectly to the intelligibility of mathematics.

Idealization of this sort, opposed even more to Aristotle’s way with nature, finds material expression in *experiment*. Galileo had to contrive situations that would reveal how selected physical factors would relate to one another in the absence of disturbing causal influences, such as friction. Though his inclined-plane experiment became a model for later emulation, it must be added that he could, on occasion, manifest a rather casual attitude towards experimental warrant. His allusions to experiments sometimes appear as rhetorical devices rather than as reports of actual performance. Koyré (1940) claimed, indeed, that Galileo rarely performed the experiments he describes, that these were no more than thought-experiments resting on his own powerful intuitions (see Thought experiments). Recent work on Galileo’s work-notes has shown this to be exaggerated. Galileo undoubtedly sought support for his new science in repeated observation.

This in turn required specially-designed *instruments* to supplement the testimony of the unaided senses. The new emphasis on mathematically expressible quantities meant that a degree of precision would be needed that the senses alone could never achieve. Furthermore, instruments such as the telescope and the microscope could extend the very notion of observation to realms the senses could not reach. Instruments could even be devised to measure quantities, like the ’strength’ of magnets, that escape the senses entirely. A significant part of Galileo’s success lay in his talent for improving instruments already in use, or for constructing instruments of entirely novel kinds. The new science would depend heavily on technological skills, a feature entirely alien to natural philosophy in the Greek tradition (see Experiment; Scientific method).

This novel complex of practices gave the new science an immediate degree of autonomy relative to the rest of philosophy. Thus began the long and often painful process of separation between two disciplines that had up to this point been considered as one. The ridicule that Galileo often directed to ’the philosophers’ was clearly meant for Aristotelian philosophers, not for philosophy itself. He had, after all, insisted on the title of ’Philosopher’ when he joined the Medici court, in part to ensure that his work would not be dismissed as mere ’mathematics’. How would he have regarded the redefinition of philosophy itself that his work did so much to set in train, had this been presented to him? Despite what some historians have claimed, he was not a positivist in the making. His conception of science as a structure of ‘necessary demonstrations’ leant as much to a past when the principles of natural philosophy carried their own intuitive self-evidence as to a future when scientific theories would be warranted primarily by their observational consequences.

3 The conception of science

Given the marked shift in scientific practice represented by Galileo’s mature work, one might have expected that his conception of what makes a knowledge-claim truly scientific would have altered correspondingly. But in this respect, he stayed close to the Aristotelian ideal absorbed in his youth. His new science departed in one important respect, of course, from the earlier conception: it was mathematical, not syllogistic, in form. But this made it *more* scientific in his eyes, not less. Indeed, it made it equal in certainty to God’s own knowledge. His opening promise was that his science ’ would prove demonstratively, and not just persuade by probable arguments’ (1638: 16). And the familiar Aristotelian phrase, ’necessary demonstration’, is dotted throughout his writings. He contrasts the probable mode of argumentation characteristic of the humanities with the demonstration proper to the natural sciences ‘ whose conclusions are true and necessary, and have nothing to do with human will’ (1632: 53).

But how could he make good on this claim? In demonstration, the inference proceeds deductively from the premises, which must be seen to be true and necessary in their own right; whereas in an observational science the order is reversed, since inferences proceed backwards, as it were, from observed particulars to (universal) premises. The warrant for the postulated premises lies in the consequences drawn from them, not in their own self-evidence. Such an inference cannot be deductive, it would seem, unless the observed particulars can somehow

prompt the mind to formulate a universal principle. This had been Aristotle's doctrine of *epagōgē*, extensively elaborated by Galileo's predecessors in Padua as a 'demonstrative regress' from observation to principle and back down to observation. Could an investigation relying on experiment fit this model? Galileo seems, at times at least, to have thought so.

Contrary to a once-popular presumption, Galileo was no inductivist. He would never have supposed that, for example, observation alone could establish as a *principle* that projectile motion follows a parabolic path. The material world, he remarks, is too full of potentially disturbing factors, and observation itself is always subject to correction. Instead, one would have to show either that the parabolic law is a principle in its own right or that it follows from other such principles. Primary among these latter, of course, would be the law of free fall, described as ' the simplest and most evident rule' (1638: 154). But how can one be sure that nature actually follows this rule? Galileo concedes that this has to be 'confirmed' by experiment. But if that is so, can the supposed 'principle' exhibit the necessity that the starting point of strict demonstration demands? Galileo was well-aware of this difficulty and hedged his claim in phrases like ' little short of necessary demonstration', 'worthy of being conceded as if demonstrated' (1638: 162, 164). The tension here between two very different conceptions of science, one demonstrative, the other hypothetical, is evident.

The axiomatic structure of his mechanics undoubtedly encouraged him, if encouragement was needed, in his retention of the language of demonstration. Regarded purely as mathematics, his mechanics could be called 'demonstrative', though admittedly in a weaker sense of that term, a sense that Galileo often exploited. In response to critics after the publication of *Two New Sciences*, he could claim to have 'demonstrated conclusively' what the properties of uniformly accelerated motion were, 'on the supposition' (*ex suppositione*) that falling motion actually *is* uniformly accelerated in nature. This allowed him to affirm that the reasoning would remain 'demonstrative' even if experiment were to show that downward motion does not quite conform to the supposition (Letters to Baliani and de Carcavi, *Opera*, XVIII: 12-13; XVII: 90). The retreat to an *ex suppositione* claim, however, was equivalently, to open the possibility that his mechanics, considered as *physics* rather than just as mathematics, could claim only a hypothetical warrant. Or else it was to assume that experimental discrepancies can be disregarded in such a case because of the inherent messiness of the causal order impeding the operation of the 'pure' physical principles. But *ought* one to rely on intuition in this way, especially in the face of counter-evidence? This was the challenge that defenders of the traditional regress model of proof had never really been able to handle.

The question of *why* bodies fall as they do was expressly laid aside; had it been acknowledged, even the appearance of demonstration could hardly have been maintained. But elsewhere in his inquiries, notably in his cosmology, the quest for agent-causes could not be avoided. What was the nature of the lunar surface, of comets, of sunspots? The only way to answer questions like these about natures that were not immediately accessible was to postulate in hypothetical fashion causes that would explain the observed effects. This sort of inference could never fit the mould of strict demonstration. The natures of these distinct objects could not become the subject of self-authenticating principle, and effect-to-cause reasoning inevitably leaves open the possibility of alternative explanations that might fit the data equally well. Galileo tried to limit this latter difficulty by proposing that for any given effect there should be one unique 'true cause'. But it was difficult to deny that in practice, in his detailed argument for the earth-like character of the lunar surface, for example, he was proposing a plausible hypothesis whose epistemic status depended on its being confirmed by the verification of consequences drawn from it (see Inference to the best explanation).

Galileo's attachment to the notion of demonstration was such that he never really came to terms with the notion of hypothesis. Unlike others of his day (Kepler, Descartes and Boyle immediately come to mind), he did not concern himself with the question, urgent in a science that argues from observed effects to unobserved agent-causes, of how the credentials of an explanatory hypothesis are to be assessed. This was to prove a critical weakness in his attempts to prove the reality of the earth's double motion, the key to the Copernican debate. His opponents, both the Aristotelian philosophers and the Roman theologians, insisted, for different reasons, on the standards of demonstration. And Galileo on his own account was not disposed to challenge these standards. But lacking demonstration of these motions, as he rather evidently did, what lesser status could he claim for the Copernican theses? The hesitant language in which he couches his conclusions in *Two Chief World Systems* betrays not only the theological dilemma in which he found himself, but also a lifelong reluctance to allow properly 'scientific'

status to anything less than demonstration.

4 Theology and science

In conclusion, something should be said about his theological dilemma. What the Church was at pains to defend throughout the Galileo affair was not so much the Aristotelian worldview, as has often been supposed, as the authority of Scripture and the Church's own authority as its interpreter. Had the handful of references in Scripture to the motion of the sun or the stability of the earth not been present, as they might well not have been, the Church would have been unlikely to become so forcefully involved in the Copernican debate. Furthermore, had the debate itself come a century earlier or a century later, a major confrontation might well have been avoided. But this was the age of the Counter-Reformation, when one of the principal issues in dispute between Protestants and Catholics involved hermeneutics: how was Scripture to be interpreted, and who had the authority to interpret it? And here was a layman affirming on his own authority (so it would have seemed in Rome) a nonstandard interpretation of biblical passages. No matter that he could advance good arguments in favour of his interpretation and even cite the authority of one of the greatest of the Church's own theologians on behalf of the principles of interpretation he employed, it was still not to be borne that such a challenge should go unanswered.

Galileo's *Letter to the Grand Duchess* has been described as one of the most effective pieces of theological writing of its century, an ironic assessment, given the lay status and the later fate of its author. Had its arguments been heeded, the fateful 1616 decree of the Congregation of the Index, effectively outlawing the Copernican theses as contrary to Scripture might never have been penned. Or would it? There was a hidden tension between the hermeneutic principles Galileo advanced, a tension to be found likewise in the Augustinian principles he appealed to. On the one hand, Scripture, was held to have no bearing on matters of natural science; these latter were said to lie outside its concerns entirely. On the other hand, where the literal reading of Scripture appears to conflict with a scientific finding, his principles implied that this latter finding has to have the status of demonstration in order to warrant the search for a different reading of Scripture.

The implications of these two directives were quite different for the Copernican debate. Were one to be guided by the first, the status of the scientific arguments for the Copernican theses would be irrelevant; the debated biblical passages mentioning sun and earth would have no scientific standing in the first place. Were one to be guided by the second, the Copernican theses would have to be demonstrated. How crucial the difference is needs no emphasis. Galileo's opponents, notably Cardinal Bellarmine in 1616, adhered to the second and more traditional directive, and demanded demonstration before they would consider a reinterpretation of the Scriptural passages along less literal lines. Galileo himself evidently favoured the first principle, which would have absolved him from the necessity of providing a demonstration of the earth's motion that he simply did not have. To defenders of the second approach (the probable arguments Galileo advanced) arguments that carried weight, though admittedly not conclusive weight with his fellow scientists, were of no avail. And Galileo's own hesitations about the epistemic status of such arguments complicated matters further.

The complexities of character and political circumstance that propelled the Galileo affair continue to fascinate modern readers. But it should also be said that this event, one of the greatest moments in human affairs, hinged to no small extent around an issue in philosophy of science.

See also: Aristotelianism, Renaissance; Explanation; Idealizations; Inductive inference; Mechanics, Aristotelian; Mechanics, classical; Platonism, Renaissance

ERNAN McMULLIN

List of works

Galilei, G. (1610) *The Starry Messenger*, ed. and trans. S. Drake, *Discoveries and Opinions of Galileo*, New York: Doubleday, 1957.(Includes a helpful narrative by the translator, situating this and the following two works that marked Galileo's early involvement in the Copernican debate.)

Galilei, G. (1613) *Letters on Sunspots*, ed. and trans. S. Drake, *Discoveries and Opinions of Galileo*, New York: Doubleday, abridged, 1957.

Galilei, G. (1615) *Letter to the Grand Duchess Christina*, ed. and trans. S. Drake, *Discoveries and Opinions of Galileo*, New York: Doubleday, 1957.

Galilei, G. (1623) *The Assayer*, trans. S. Drake, in C.D. O'Malley and S. Drake (eds), *The Controversy on the*

Comets of 1618, Philadelphia, PA: University of Pennsylvania Press, 1960.(Where Galileo lays out some of his most characteristic philosophic theses in the context of an increasingly bitter dispute about the nature of comets with the Jesuit natural philosopher Orazio Grassi.)

Galilei, G. (1632) *Dialogue Concerning the Two Chief World Systems*, trans. S. Drake, Berkeley, CA: University of California Press, 1953.(The classic work in which Galileo undermines the Aristotelian arguments against the Copernican proposal, refutes Aristotle's own concentric world-system, and advances a number of arguments in support of the Copernican alternative.)

Galilei, G. (1638) *Two New Sciences*, trans. S. Drake, Madison, WI: University of Wisconsin Press, 1974.(The foundational work in modern mechanics, proposing Galileo's two laws of motion. The 'new science' concerns the strength of materials.)

Galilei, G. (1890-1909) *Edizione Nazionale delle Opera di Galileo Galilei*, ed. A. Favaro, Florence: Barbera; repr. 1968.(Definitive 20-volume edition of Galileo's works and correspondence.)

Galilei, G. (1989) *The Galileo Affair*, ed. and trans. M. Finocchiaro, Berkeley, CA: University of California Press. (Useful collection of documents bearing on the 'Galileo affair'. Includes the *Letter to the Grand Duchess Christina*, selected correspondence, and Holy Office files.)

References and further reading

Biagioli, M. (1993) *Galileo, Courtier: The Practice of Science in the Culture of Absolutism*, Chicago, IL: University of Chicago Press.(Critical of the 'idealism' of the traditional approaches to Galileo's science in terms of arguments advanced, evidence presented and methods followed; suggests that when faced with controversy Galileo relied for legitimation to a significant extent on princely patronage. Employing the methods of cultural anthropology, the author focuses on the dynamics of princely patronage that made possible Galileo's 'self-fashioning' as a 'new philosopher' and contributed to his ultimate downfall.)

Butts, R.E. and Pitts, J. (eds) (1978) *New Perspectives on Galileo*, Dordrecht: Reidel. (Detailed essays dealing primarily with Galileo's scientific methods and his notion of what constitutes science.)

Drake, S. (1978) *Galileo at Work: His Scientific Biography*, Chicago, IL: University of Chicago Press.(Detailed chronicle, year by year, of Galileo's scientific work.)

Fantoli, A. (1994) *Galileo: For Copernicanism and for the Church*; trans. G. Coyne, Rome: Vatican City Publications, 2nd revised and enlarged edn, 1996.(With 180 pages of footnotes and extensive bibliography, an amply documented account of Galileo's protracted attempt to convince the church authorities in Rome of the merits of the Copernican system, the dramatic failure of his efforts, and the aftermath in Rome up to the mid-1990s.)

Finocchiaro, M.A. (1980) *Galileo and the Art of Reasoning: Rhetorical Foundations of Logic and Scientific Method*, Dordrecht: Reidel.(Immensely detailed analysis of *Dialogue Concerning the Two Chief World Systems*, dividing it into sixteen main arguments and 229 sub-arguments. The author claims this work as the book of choice for anyone who wants to see critical reason at its best.)

Koyré, A. (1940) *Etudes Galiléennes*, Paris: Hermann; trans. J. Mepham, *Galileo Studies*, Atlantic Highlands, NJ: Humanities Press, 1978.(Groundbreaking analyses of Galileo's science, situating it within the Archimedean tradition, and strongly contesting the nineteenth-century view of Galileo as the model empiricist.)

Machamer, P. (ed.) (1998) *Cambridge Companion to Galileo*, Cambridge: Cambridge University Press.(A set of essays from recent contributors to Galileo scholarship that effectively illustrates how diverse that scholarship's findings still remain.)

McMullin, E. (ed.) (1967) *Galileo: Man of Science*, Notre Dame, IN: University of Notre Dame Press.(A collection of twenty-three studies of Galileo's scientific achievements, including classical essays by notable scholars of an earlier generation: Koyré, Olschki, Tannery, Cassirer. Comprehensive bibliography covering 1940-64.)

McMullin, E. (1985) 'Galilean Idealization', *Studies in the History and Philosophy of Science*, 16: 247-73. (Analysis of six different modes of idealization employed by Galileo in his scientific work.)

Wallace, W.A. (1992a) *Galileo's Logical Treatises*, Dordrecht: Kluwer. (Translation of, and commentary on, two unpublished treatises Galileo evidently composed as aids in his early teaching duties *c*.1590. They deal with two themes in Aristotle's *Posterior Analytics*: demonstration and prior knowledge. Wallace has shown by dint of patient research that these treatises almost certainly derive from the 1588 class-notes of Paulus Vallius, a professor at the Jesuit Collegio Romano.)

Wallace, W.A. (1992b) *Galileo's Logic of Discovery and Proof*, Dordrecht: Kluwer. (Culmination of a series of works analysing the background and content of Galileo's early commentaries on Aristotelian logical themes. Argues cogently for the influence of these doctrines on the logical structure of Galileo's later writings on

motion.)

Gandhi, Mohandas Karamchand (1869-1948)

Gandhi was called Mahatma *(the Great Soul) by Rabindranath Tagore and many in the West, while Gandhi's followers often simply called him* Bapuji *(Father). His confrontation with racism in South Africa provided a challenging context for the development of his idea of* satyāgraha *(holding fast to the truth), a method of nonviolent, noncooperative resistance to the authorities. Influenced by several religious traditions, such as Hinduism (especially Vaishnava), Jainism, Islam and Christianity, Gandhi was both a religious thinker and practical reformer. While in jail on several occasions, he wrote prolifically. He was murdered on 30 January 1948 by a Hindu zealot.*

1 Life and career

Gandhi was born on 2 October 1869 in Porbandar, Kathiawad to a politician father, Karamchand Uttamchand and a religiously devout mother, Putalibai. At the age of thirteen an arranged marriage took place between Gandhi and Kasturbai, although he was later to oppose child marriage. Educated in India and London, Gandhi passed his bar examination with ease, enrolled at the High Court on 11 June 1891 and sailed for India the following day. After a brief and unsuccessful attempt at law in India he went to South Africa, initially to represent the interests of a Muslim firm. From 1893-1914 Gandhi struggled against racism in South Africa.

Gandhi returned to India in 1915 to engage in a struggle with the British for *swarāj* (self-rule). He lived a life of simplicity and renunciation and used the method of fasting to the brink of death as a way of persuading opponents.

The Amritsar massacre of 1919 was decisive in Gandhi's rejection of British colonial rule in any form. *Khādī*, white home-spun cotton clothes and *charkha*, the spinning wheel, became Gandhian symbols. *Charkhas* were difficult to find, as they were rare among the professional elite in Gandhi's time, and the making of *khādīs* had to be learned.

2 Truth, religion and politics

Gandhi made his first public speech on truth in business dealings in Pretoria, South Africa. He believed in both 'truth of statement' and the notion of 'truth of being'. He was preoccupied with truth and felt it must be expressed in a life of *ahiṃsā*, or non-injury (which entails vegetarianism). He always saw the divine in the less fortunate (for instance by opposing untouchability). Gandhi believed that a *satyāgrahī*, one 'steadfast in truth', ideally must exemplify celibacy (following the vow of *brahmacarya*). Only then could the *satyāgrahī* rise above all passionate attachments and generate enough internal 'heat' (*tapas*) to do battle with untruth. Gandhi's goal was one of serving the public rather than only achieving his individual salvation. He was motivated by the ideal of *sarvodaya*, the welfare of all (see Duty and virtue, Indian conceptions of).

Gandhi appropriated the term *sadāgraha* and changed it to *satyāgraha* (holding fast to the truth), with *satya* representing truth and *āgraha* firmness. In *An Autobiography, or the Story of My Experiments with Truth* (1927) Gandhi spoke of truth as an underground mine holding many opportunities for service. Proposing that 'there is no other God than truth', Gandhi understood divinity as immanently realizable through dedicating one's work to God. In Hindu tradition this idea is *karma yoga*. For Gandhi *karma yoga* meant a dedication to truth in each particular case, rather than in mystical contemplation of an abstract Godhead. He often appealed to 'the Voice' of conscience wherein God speaks. Accepting the principle of *karma* and the doctrine of transmigration, Gandhi hoped to be reborn an untouchable to be of further service.

In *An Autobiography, or the Story of My Experiments with Truth* (1927), Gandhi revealed that he was not well-versed in Sanskrit and the Hindu classics at the time when the the role of guru was thrust upon him by the public. His 'philosophy of religion' was confessional in form, not analytical. He thought at a metareligious level, believing that all religions participate in truth. The Sermon on the Mount, the *Bhagavad Gītā* (which Gandhi himself translated) and the writings of Tolstoi were sources of particular fascination for him. The Jain doctrine of *ahiṃsā* (non-injury) was one of his main tenets and provided the background to his vegetarianism (see Jaina philosophy). His close friend C.F. Andrews supplied a Christian stimulus in his life, but Gandhi found deeper inspiration in his native *Bhagavad Gītā*. Self-purification and dedication to God through service to humanity provided the basis for religiosity according to Gandhi. For Gandhi, religion must bind together humankind and

include politics. There were no particular religious images in Gandhi's *ashram*. With the knowledge that God has many names, several scriptures were recited in the daily routine. In this way Gandhi's practice exemplified the tolerance of Jain *syādvāda*, the view that all judgments of nonomniscient beings must be qualified by 'somehow' or 'perhaps' (see Manifoldness, Jaina theory of). Gandhi was influenced by Jainism from early on. Philosophically, his position expressed a reverence for life characteristic of both Jainism and Buddhism. He disputed the notion that Buddhism is atheistic, preferring to interpret Buddhist *dharma* in the sense of 'truth' as representing God. In his view, religion at its best is concerned with human community and the transcendence of cultural tribalism and parochialism.

Gandhi's goals were the achievement of *sarvodaya*, sovereignty and self-sufficiency in India, encompassing the idea of teaching in native languages. Although untouchability and child marriage were vigorously opposed as they oppressed the less fortunate and defenceless, Gandhi defended stratification in terms of *varṇa* in the limited sense of the division of labour. He was also opposed to individualistic capitalism. Education was important to Gandhi, and he thought that colleges should be linked to particular industries. Like Martin Luther King, he did not offer a theory of justice, but believed, as King was later to say: 'injustice anywhere is a threat to justice everywhere'.

See also: Duty and virtue, Indian conceptions of; Political philosophy, Indian; Tagore, R.

FRANK J. HOFFMAN

List of works

Gandhi, M.K. (1958-) *The Collected Works of Mahatma Gandhi*, New Delhi: Ministry of Information and Broadcasting, 100 volumes to date.(This voluminous work is the authoritative edition of Gandhi's collected works.)

Gandhi, M.K. (1927) *An Autobiography, or the Story of My Experiments with Truth*, trans. M. Desai, Boston, MA: Beacon Press, 1957.(This book is the single most important work by Gandhi for understanding his philosophy of life.)

References and further reading

Carter, A. (1996) *Mahatma Gandhi: A Selected Bibliography*, Bibliographies of World Leaders 2, Westport, CT: Greenwood Press.(This work is an up-to-date, indispensable bibliography.)

Chatterjee, M. (1983) *Gandhi's Religious Thought*, London: Macmillan.(This is the best single work on Gandhi from a theological viewpoint currently available.)

Herman, A.L. (1990) *The Ways of Philosophy*, Atlanta, GA: Scholars Press.(Chapters 11-12 describe the link between the ideas of Gandhi and Martin Luther King.)

Hick, J. and Hempel, L. (1989) *Gandhi's Significance for Today*, New York: St Martin's Press.(This volume constitutes an important collection of essays on Gandhi, his religion, ethics, politics and economics.)

Iyer Raghavan, N. (1986-7) *The Moral and Political Writings of Gandhi*, Oxford: Oxford University Press, 3 vols.(This volume is a major scholarly study of Gandhi's moral and political thought.)

Jack, H.A. (1956) *The Gandhi Reader*, Bloomington, IN: Indiana University Press.(This source book is ideal for classes and study groups on Gandhi.)

Juergensmeyer, M. (1984) *Fighting with Gandhi*, San Francisco, CA: Harper & Row.(This work is an engaged, thoughtful study of Gandhian *satyāgraha* that includes both case studies and confrontations with the intellectual currents of Marx, Freud and Niebuhr.)

Rani, A. (1981) *Gandhian Non-Violence and India's Freedom Struggle*, Delhi: Shree Publishing House.(This work offers a detailed insider's view of Gandhi's philosophy-in-practice, amply researched from Indian archives.)

Richards, G. (1982) *The Philosophy of Gandhi: A study of his Basic Ideas*, London: Curzon Press.(This is an excellent philosophical study of Gandhi's thought.)

Gaōgeśa (*fl. c.*1325)

Gaōgeśa launched and solidified advances in logic and epistemology within the classical Indian school of Logic, Nyāya. He is traditionally taken to have inaugurated the 'New' school, Navya-Nyāya. Nyāya, both Old and New, is a multidimensional system that belies the stereotype of Indian philosophy as idealist and mystical in orientation. Gaōgeśa worked with a realist ontology of objects spoken about and experienced every day. He articulated what may be called a reliabilist theory of knowledge: under specified conditions, sense-mediated and inferential cognitions (along with two other types) are reliable sources of information about reality.

Gaōgeśa was a pivotal figure in classical Indian philosophy; most later debate both within his school and outside it presupposed cognitive analyses that he standardized. These analyses focus on properties exhibited by things known, properties central to the processes whereby they are known. Properties relating the cognized to the cognizer are especially important. Though Gaōgeśa had a lot to say about the ontological status of these properties, others in his school found them problematic. Such controversy appears to have contributed to New Logic's success: proponents of rival views were able to utilize Gaōgeśa's formulas and definitions without abandoning their own positions on what is real.

1 Historical particulars

Gaōgeśa Upādhyāya lived in the first half of the fourteenth century, in Mithilā, in northern India, where he was a prominent teacher. Gaōgeśa's school of New Logic remained a Mithilāmonopoly for almost two hundred years; students travelled there to learn the system and returned with it to their own regions. Within another century, New Logic came to dominate late classical thought all over India, influencing jurisprudence and aesthetics as well as shaping the philosophical scene. But unfortunately we know Gaōgeśa the person almost exclusively as the author of a single work, the *Tattvacintāmaṇi (Jewel of Reflection on Reality)*.

The *Tattvacintāmaṇi* is New Logic's root text. Numerous commentaries were written on it, and all later New Logic writers presupposed its acquaintance. It is without question one of the premier works of the whole of classical Indian thought. The *Tattvacintāmaṇi* contains the distilled reflections of generations of earlier philosophers, mainly within the sister schools of Nyāya and Vaiśeṣika (see Nyāya-Vaiśeṣika), but also arguments of historic opponents of these two complementary realist views. Gaōgeśa's originality is in fact less than one might expect given the acclaim for his *Tattvacintāmaṇi*. He himself recognizes Udayana (*fl. c.*1000), and rightly so, as the innovator of positions characteristic of the New Logic movement. But Gaōgeśa does refine the Logic understanding of the cognitively veridical and nonveridical (among other things), and it is here that he has his greatest influence on subsequent Indian thought.

The *Tattvacintāmaṇi* comprises four chapters, each devoted to a separate source of right cognition: veridical perception, cogent inference, analogical vocabulary acquisition, and reliable testimony. Within each chapter there are clearly delineated sections on various subtopics, including reflections, mainly in the first chapter, concerning cognition and veridicality in general, considered independently of particular instruments. Gaōgeśa is lucid and precise; he argues forcefully, ingeniously leading the reader into positions by careful examination of alternatives.

2 Perception and inference

Cognition (*jñāna*, often translated 'awareness') is the most important item in Gaōgeśa's philosophy. Ontologically, cognitions are short-lived, episodic attributes or qualities of the self or soul, and, in the case of sensory awarenesses, are causally continuous with physical realities as a result of the relations between sense organ and object. A cognition may be understood as a mental event, but it is a product or state rather than an act. Cognitions are intentional; they are invariably *of* some object or objects, or, more precisely, objective complex(es).

Gaōgeśa posits sensory cognitions below the level of what we can consciously articulate; these are called indeterminate cognitions, by contrast with the determinate. Every determinate cognition is verbalizable, but indeterminate cognitions are not. Even the simplest verbalizable cognition has an object that is ontologically complex: that is, a qualificandum is cognized through one of its qualifiers in the case of a veridical awareness. It is to avoid an infinite regress of qualification that Gaōgeśa posits indeterminate cognition, where qualifiers are

directly known. Otherwise, there would be the problem of how a $qualifier_1$ of $qualificandum_1$ is known; that is, if the qualifier had to be itself a qualificandum, $qualificandum_2$, to be known through another qualifier, $qualifier_2$, for anything to be known, there would have to be an infinite series of qualifiers and qualificanda. Thus some qualifiers are cognized directly though indeterminately.

Indeterminate cognition is said to give rise to determinate cognition through a causal process. Although acquaintance with an individual is determinate, it should be thought of as causally direct: perception is a causal process spread over time, and indeterminate cognition may be viewed as a first phase. Verbalizable cognition of an individual is mediated by - or, better, fused with - what the individual appears as. Only determinate cognition is verbalizable, in that its content is expressible in propositional form as something, *a*, being an *F* (*Fa*).

All simple (non-doubly, triply, and so on) qualificative or determinate cognitions have a content that exhibits a common structure: qualificandum, qualificative relation, qualifier (*a*-*R*-*b*). Qualification is a relational property that obtains in the world, but it is used to talk about cognitions. The simplest veridical sensory cognition reflects a fact or state of affairs, an individual cognized (the qualificandum) as related to a qualifier. With the veridical awareness '(A) pot', for example, we would have the particular pot- inherence-potness as the objective complex (*a*-*R*-*b*) cognized as something appearing as a pot. The qualificandum is the bearer or possessor of properties; with the veridical awareness '(A) pot', the particular pot is the subject of qualification and possesses the property of potness. The awareness '(A) pot' has 'potness' as its predication content.

With this apparatus, Gaōgeśa is able to analyse nonveridical sensory awareness as a cognition whose predication content does not qualify the object cognized, that is, the object that stands in relation to the sense organ. A causal story is to be told why there is such a foul-up, why something appears as other than it is, a story that will vary with the circumstances of individual instances of error (see Error and illusion, Indian conceptions of §3).

Inferential knowledge is defined by Gaōgeśa as the cognition arising as the result of the perceptual cognition of something *a* as possessing a property *F* that is understood as pervaded by (or invariably concomitant with) another property *G* (*F* as invariably co-located with *G*). To quote Gaōgeśa: 'Inferential knowledge is cognition generated by cognition of a property-belonging-to-a-locus-and-qualified-by-a-pervasion' (*Tattvacintāmaṇi*, 1982: 5; the dashes in the translation are to indicate that the cognition has to have all this as content). Logicians throughout the history of their school are concerned with inference only in so far as it provides knowledge of the world. Thus any correct inference must have a perceptual premise, which is its starting point, so to speak: for example, 'There is smoke on yonder hill', as known perceptually. (Perceptual cognition may be said to be the premier knowledge source in another way as well: inferential cognitions that are *prima facie* correct can be overridden by sensory awareness.) But of course it is not just any perceptual awareness that leads to inferential knowledge: for example, the smoke, or smokiness, as a qualifier of a locus or qualificandum must be seen as invariably related to another property called technically the property to be proved; it must be qualified by pervasion, or co-locatedness, with the property of being-fiery. Throughout New Logic it is the nature of the relation called pervasion (*vyāpti*, often translated 'invariable concomitance') that is the main focus of reflection under the general rubric of inference. Gaōgeśa considers and rejects a total of twenty-one definitions of pervasion. He accepts a total of five.

Among other important aspects of Gaōgeśa's epistemology is fallibilism, which feeds his response to scepticism concerning how a pervasion in its full generality can by known (occurrence of *G wherever* there is occurrence of *F*: how can this be known?). The answer is wide experience, but future experience could prove us wrong.

Gaōgeśa shows a pragmatic orientation in his final response to scepticism on this score. The sceptic's behaviour gives the lie to the complaint: does the sceptic not, for example, in voicing the complaint, assume a *vyāpti* between speaking and others' comprehension?

3 Ontology

Like all Nyāya ontology, that of Gaōgeśa's *Tattvacintāmaṇi* is realistic in the sense of a commitment to entities whose existence is independent of consciousness. Realism led the philosophers of Gaōgeśa's Nyāya school to embrace fallibilism, because the transcendence, so to speak, of a physical object with respect to the instruments of knowledge means that the possibility of error cannot be ruled out. Gaōgeśa's definitional projects are attempts at philosophical characterization as opposed to linguistic analysis, because the various definienda (veridical

awareness, inference-grounding pervasion, and so on) are realities in the world, not linguistic items. Then regarding the backbone of Nyāya ontology - its categorial system - all individuals of whatever fundamental type (substances, qualities, motions, natural kinds, inherence, ultimate individuators, and absences) are viewed as reals.

Gaōgeśa's ontological labours are generally in the service of epistemology. Consonantly, his treatment of six 'operative relations in sense experience' is extensive. So also are his reflections on *vyāpti*, inference-underpinning pervasion. The material on *vyāpti* received an enormous measure of attention from later Logicians.

Depending on what is cognized - individual substance, quality, natural-kind character, an absence (for example, of my glasses on the table), a previous cognition (as in apperception), and so on - the relationality between the grasping and the grasped varies. Moreover, the different sense organs relate to objects differently. And there are special problems about the 'operative relation' in a perception of an absence, as well as with what in the West is known as the 'Bradley problem' of a possible infinite regress of relations centring on inherence as joining a quality to a substance (for example) - what then joins the inherence? The importance of the ontology of absence looms especially large because the acceptable characterizations of pervasion require mention of certain general types of absence. That characterization project also involves Gaōgeśa in explaining ontologically what 'co-locatedness' amounts to.

Gaōgeśa takes very seriously a principle of ontological parsimony (*lāghavatva*), and so his reasoning through all the complexities of a variety of desiderata and principles potentially in tension is at times tortuous. But it is invariably tight, and though later Logicians abandon some of his positions (Gaōgeśa sometimes appears too conservative, too intent on defending Udayana's ontological stances), his arguments form the problem space for later reflection.

See also: Epistemology, Indian schools of §2; Knowledge, Indian views of; Ontology in Indian philosophy

STEPHEN H. PHILLIPS

List of works

Gaōgeśa (first half of 14th century) *Tattvacintāmaṇi (Jewel of Reflection on Reality)*, with the *Māthurī* commentary of Mathurānātha (or others where that is not extant), ed. K. Tarkavagish, Calcutta: Asiatic Society, 5 vols, 1884-1901; repr. Delhi: Chowkhambha Sanskrit Pratisthan, 1990.(The reprint has made Gaōgeśa easily available to the Sanskrit-reading community in India, whose numbers are considerable.)

Gaōgeśa (first half of 14th century) *Tattvacintāmaṇi (Jewel of Reflection on Reality)*, vol. 1, *pratyakṣa-khaṇḍa*, with the *Prakāśa* commentary by Rucidatta Miśra and a subcommentary by Rāmakṛṣṇādhvarin, vol. 2, *anumāna-khaṇḍa*, with the *Prakāśa* commentary by Rucidatta Miśra and a subcommentary by Dharmarājādhvarin, ed. N.S. Ramanuja Tatacharya, Kendriya Sanskrit Vidypeetha Series 20 and 33, Tirupati, 1972, 1982.(The commentary and subcommentary, as well as the alternative readings and punctuation supplied by the editor, make these volumes invaluable to scholars.)

References and further reading

Bhattacharya, D.C. (1958) *History of Navya-nyāya in Mithilā*, Darbhanga: Mithila Institute.(Informative historical research.)

Bhattacharya, K. (trans.) (1977, 1978, 1980, 1982, 1984) 'Le *Siddhāntalakṣaṇaprakaraṇa* du *Tattvacintāmaṇi* de Gaōgeśa avec la *Dīdhiti* de Raghunātha Śiromaṇi et la *Ṭīkā* de Jagadīśa Tarkālaṃkāra', *Journal Asiatique*, 1977: 97-139, 1978: 97-124, 1980: 275-322, 1982: 401-13, 1984: 47-82.(Elegant translation of difficult materials; the commentaries are important in their own right.)

Bhattacharyya, S. (1993) *Gaōgeśa's Theory of Indeterminate Perception*, New Delhi: Indian Council of Philosophical Research.(Lucid explanation of key concepts of Gaōgeśa's epistemology and philosophy of perception.)

Goekoop, C. (1967) *The Logic of Invariable Concomitance in the Tattvacintāmaṇi*, Dordrecht: D. Reidel.(A translation and explanation of a portion of the *Tattvacintāmaṇi* devoted to defining 'pervasion', *vyāpti*.)

Ingalls, D.H.H. (1951) *Materials for the Study of Navya-Nyāya Logic*, Cambridge, MA: Harvard University Press.(Though now a little dated, this is a masterly exploration of the analytic weaponry of New Logic.)

Matilal, B.K. (1968) *The Navya-Nyāya Doctrine of Negation*, Cambridge, MA: Harvard University Press.(Matilal's long introduction to a translation of a *Tattvacintāmaṇi* section - the section on the ontology of absence - is an excellent overall introduction to the central concepts and claims of Gaōgeśa's system.)

Mohanty, J.N. (1966) *Gaōgeśa's Theory of Truth*, repr. Delhi: Motilal Banarsidass, 1989.(Though Mohanty renders *pramātva* as 'truth' instead of 'veridicality', his investigation of epistemological views belonging to Gaōgeśa's opponents as well as to the great Logician himself is excellent philosophy.)

Phillips, S.H. (1995) *Classical Indian Metaphysics: Refutations of Realism and the Emergence of 'New Logic'*, La Salle, IL: Open Court.(Contains elaboration of, in particular, what is discussed §2 of the present entry.)

Phillips, S.H. and Tatacharya N.S.R. (trans.) (1998) *Gaōgeśa's 'Jewel of Reflection on Reality', the Perception Chapter*, Delhi: Motilal Banarsidass.(Translation, interspersed with the translators' commentary, of the entire first chapter of Gaōgeśa's 'Jewel'.)

Potter, K.H. (ed.) (1977) *Encyclopedia of Indian Philosophies*, vol. 2, *Nyāya-Vaiśeṣika*, Princeton, NJ: Princeton University Press.(The best introduction to the philosophies and philosophers that most inform Gaōgeśa's reasonings.)

Potter, K.H. and Bhattacharyya, S. (eds) (1994) *Encyclopedia of Indian Philosophies*, vol. 6, *Nyāya-Vaiśeṣika from Gaōgeśa to Raghunātha Śiromaṇi*, Princeton, NJ: Princeton University Press.(Contains summaries of most of the *Tattvacintāmaṇi* written by a group of scholars and philosophers; should prove a landmark publication.)

Garrigou-Lagrange, Réginald (1887-1964)

Garrigou-Lagrange was a French Dominican who for decades adorned the Angelicum in Rome, where in his courses he commented closely on the Summa theologiae. The spiritual life was a principal interest of Garrigou-Lagrange, and many of his books are devoted to the theology and practice of mystical union with God. Impatient with theological novelty, Garrigou-Lagrange came to be caricatured by the champions of innovators. His own work, solid, careful, illuminating, is a monument to a golden period of the Thomistic revival.

Garrigou-Lagrange was an influential figure in the second phase of the revival of Thomism initiated by Pope Leo XIII's 1879 encyclical *Aeterni Patris*. He first studied medicine, then entered the Order of Preachers, where, after studying philosophy and theology, he began in 1905 a teaching career which was to last sixty years. He taught at Le Saulchoir, Belgium, until, in 1909, he was assigned to the Dominican university in Rome, the Angelicum.

Mentor of Jacques Maritain, *bête noire* of Étienne Gilson and Henri de Lubac, Garrigou-Lagrange was a Thomist of the strict observance. His writings, vast in number, fall into three main classes: philosophical, theological and spiritual. His theological work consists of commentaries on Aquinas, in which he follows the lead of the great commentators of the Dominican Order - Cajetan, John of St Thomas and Báñez - as well as original work in a Thomistic key. His spiritual writings deal with the progressive appropriation of the life of grace leading to mystical union with God. His philosophical works are of particular interest to the philosopher of religion.

In theory of knowledge, Garrigou-Lagrange was an ardent defender of the doctrine of Aquinas which, while taking due account of the contribution of our mind to the object of intellectual knowledge, sees that contribution as modal rather than substantive. The essences or quiddities that are the object of intellectual knowledge are those of really existing things. These essences are individuated in material substances and are universal and immaterial (modally) as known (see Aquinas, T. §11).

Garrigou-Lagrange's Thomism can best be sampled in the lengthy article he contributed to the *Dictionnaire de théologie catholique*, which became the basis for his book *La Synthèse thomiste (The Thomist Synthesis*, 1946). After a bibliographical introduction to the philosophical and theological works of Aquinas, Garrigou-Lagrange significantly discusses the Thomistic commentators. He saw his own efforts as a continuation of a tradition that had been broken and was to be renewed. The 'metaphysical synthesis of Thomism' concentrates on act and potency. The bulk of the work deals with Aquinas' explicitly theological doctrines, following the general plan of the *Summa theologiae*. It ends with a section devoted to the 'Twenty-Four Theses', a list of key tenets of Thomism which had been drawn up by several professors and approved by the Congregation of Sacred Studies as an adequate summary of Thomas' teaching, and a discussion of epistemological realism.

Garrigou-Lagrange's magisterial *Dieu, son existence et sa nature* (*God: His Existence and His Nature*, 1915) is a running debate with post-Kantian thought. Thus he defends the epistemological assumptions of proofs for the existence of God before giving a detailed analysis of Aquinas' Five Ways and discussing the divine attributes. Throughout, Garrigou-Lagrange indicates how Aquinas' views differ from those of major modern philosophers. The work ends with a discussion of the principle of inertia and conservation of energy, an indication of the author's concern to deal with apparent difficulties for the assumptions of the traditional proofs of God's existence. He also argues that agnosticism leads to atheistic evolutionism.

Other works of interest to philosophers of religion are *De de uno (The One God*, 1937), *La prédestination des saints et la grâce (The Predestination of the Saints and Grace*, 1936) and *Providence et la confiance en Dieu (Providence and Confidence in God*, 1935). *Le Sens commun (Common Sense*, 1921) is at once a defence of realism and an extended engagement with the thought of Bergson and with the Scottish Common-Sense school. Throughout, Garrigou-Lagrange recommends Thomistic positions as preferable to those of the rivals he examines. The work ends with criticism of the modernist interpretation of dogmatic formulas. *Le Réalisme du principe de finalité (The Realism of the Principle of Finality*, 1932) compares knowledge of natural and supernatural truths. The motto on the title page reads thus: 'Every being acts for an end, from the grain of sand to God. Our intellect knows its own finality: to judge in conformity with the nature and existence of things and to raise itself to their first cause and ultimate end.'

The style and approach of Garrigou-Lagrange have been deplored by some post-conciliar Catholics. This is in many ways unjust. He is an engaging writer, a thinker of great power, a Thomist who takes the thought of his master to be the answer to some of the twentieth century's more vexing philosophical divagations. But that, of course, was the message of *Aeterni Patris*.

See also: Thomism

RALPH McINERNY

List of works

Garrigou-Lagrange, R. (1915) *Dieu, son existence et sa nature*, trans. Dom B. Rose, *God: His Existence and His Nature*, St Louis, MO: Herder, 2 vols, 1934, 1935.(Proof of the demonstrability of God; treats the Five Ways of St Thomas, and various divine attributes and their interrelationship.)

Garrigou-Lagrange, R. (1921) *Le Sens commun (Common Sense)*, Paris: Nouvelle Librairie Nationale.(An examination of whether the immortality of the soul and the first principles are known by common sense, and the role of common sense in proving the existence of God.)

Garrigou-Lagrange, R. (1932) *Le Réalisme du principe de finalité (The Realism of the Principle of Finality)*, Paris: Desclée de Brouwer.(A discussion of being, change and finality, with reference to chance, and the application of of the principle of finality to epistemological realism.)

Garrigou-Lagrange, R. (1933) *Le Sauveur et son amour pour nous*, Paris: Éditions du Cedre; trans. A. Bouchard, *Our Savior and His Love for Us*, St Louis, MO: Herder, 1951.(A discussion based on *Summa theologiae* IIIa: the personality of Christ and his redemptive act.)

Garrigou-Lagrange, R. (1935) *Providence et la confiance en Dieu*, Paris: Desclée de Brouwer; trans. Dom B. Rose, *Providence*, St Louis, MO: Herder, 1937.(A practical rather than theoretical treatment of providence from the point of view of philosophy, Scripture and providence, and religious submission to providence.)

Garrigou-Lagrange, R. (1936) *La Prédestination des saints et la grâce*, Paris: Desclée de Brouwer; trans. Dom B. Rose, *Predestination*, St Louis, MO: Herder, 1953.(Comparison of Thomas' teaching on predestination and the later disputes between Jesuits and Dominicans that are summed up under the tag *de auxiliis*.)

Garrigou-Lagrange, R. (1937) *De de uno*, Paris: Desclée de Brouwer; trans. Dom B. Rose, *The One God*, St Louis, MO: Herder, 1951.(A commentary on questions 1-26 of *Summa theologiae* Ia.)

Garrigou-Lagrange, R. (1938) *Les Trois Âges de la vie intérieure*, Paris: Éditions du Cerf, 2 vols; trans. Sister M.T. Doyle, O.P., *The Three Ages of the Interior Life*, St Louis, MO: Herder, 2 vols, 1948, 1952.(The sources and end of the interior life; a statement of the classical conception of degrees of spirituality. This masterpiece in mysticism treats grace, the gifts of the Holy Spirit, the indwelling of the Trinity, purification and illumination.)

Garrigou-Lagrange, R. (1946) *La Synthèse thomiste*, Paris: Desclée de Brouwer; trans. P. Cummins, *Reality: A Synthesis of Thomistic Thought*, St Louis, MO: Herder, 1950.(An examination of the nature of theology, proofs of God's existence, his nature, Trinity, angels, humanity and morals.)

Garrigou-Lagrange, R. (1948) *De unione sacerdotis cum Christo sacerdote*, Turin: Marietti; trans. E. Hayden, *The Priesthood and Perfection*, Westminster, MD: Newman Press, 1955.(Treats the pursuit of perfection, the spirituality fitting for priests, mental prayer, cultivation of the virtues, and the Eucharist.)

References and further reading

Gilson, É. and Maritain, J. (1990) *Étienne Gilson, Jacques Maritain, Deux approches de l'Être: Correspondance, 1923-1971*, ed. G. Prouvost, Paris: J. Vrin.(The two leading Thomists correspond over the years; their estimates of the work of Garrigou-Lagrange differ fundamentally, Maritain defending it, Gilson criticizing it.)

Leo XIII (1879) *Aeterni Patris (Of the Eternal Father)*, in J. Maritain, *Le Docteur angélique*, Paris: Desclée de Brouwer, 1930; trans. J. Evans and P. O'Reilly, *St Thomas Aquinas*, New York: Meridian, 1958.(Includes a discussion of the encyclical, as well as its text.)

Lubac, H. de and Gilson, É. (1988) *Letters of Étienne Gilson to Henri de Lubac*, ed. H. de Lubac, San Francisco, CA: Ignatius Press.(Gilson's letters deftly used to win arguments with various departed foes of de Lubac. Garrigou-Lagrange subjected to much bantering and some serious criticism.)

Gassendi, Pierre (1592-1655)

Pierre Gassendi, a French Catholic priest, introduced the philosophy of the ancient atomist Epicurus into the mainstream of European thought. Like many of his contemporaries in the first half of the seventeenth century, he sought to articulate a new philosophy of nature to replace the Aristotelianism that had traditionally provided foundations for natural philosophy. Before European intellectuals could accept the philosophy of Epicurus, it had to be purged of various heterodox notions. Accordingly, Gassendi modified the philosophy of his ancient model to make it conform to the demands of Christian theology.

Like Epicurus, Gassendi claimed that the physical world consists of indivisible atoms moving in void space. Unlike the ancient atomist, Gassendi argued that there exists only a finite, though very large number of atoms, that these atoms were created by God, and that the resulting world is ruled by divine providence rather than blind chance. In contrast to Epicurus' materialism, Gassendi enriched his atomism by arguing for the existence of an immaterial, immortal soul. He also believed in the existence of angels and demons. His theology was voluntarist, emphasizing God's freedom to impose his will on the Creation.

Gassendi's empiricist theory of knowledge was an outgrowth of his response to scepticism. Accepting the sceptical critique of sensory knowledge, he denied that we can have certain knowledge of the real essences of things. Rather than falling into sceptical despair, however, he argued that we can acquire knowledge of the way things appear to us. This 'science of appearances' is based on sensory experience and can only attain probability. It can, none the less, provide knowledge useful for living in the world. Gassendi denied the existence of essences in either the Platonic or Aristotelian sense and numbered himself among the nominalists.

Adopting the hedonistic ethics of Epicurus, which sought to maximize pleasure and minimize pain, Gassendi reinterpreted the concept of pleasure in a distinctly Christian way. He believed that God endowed humans with free will and an innate desire for pleasure. Thus, by utilizing the calculus of pleasure and pain and by exercising their ability to make free choices, they participate in God's providential plans for the Creation. The greatest pleasure humans can attain is the beatific vision of God after death. Based on his hedonistic ethics, Gassendi's political philosophy was a theory of social contract, a view which influenced the writings of Hobbes and Locke.

Gassendi was an active participant in the philosophical and natural philosophical communities of his day. He corresponded with Hobbes and Descartes, and conducted experiments on various topics, wrote about astronomy, corresponded with important natural philosophers, and wrote a treatise defending Galileo's new science of motion. His philosophy was very influential, particularly on the development of British empiricism and liberalism.

1 Life and works

Pierre Gassendi was born in Champtercier, a village near Digne in Provence, France. He received his early education in Digne and Riez and was admitted to the clerical state in 1604. The remainder of his formal education was supervised by the Church, as it formed part of his preparation for the priesthood. From 1604 through 1611 he studied Aristotelian philosophy and Catholic theology at the college of Aix-en-Provence. Upon receiving his doctorate in theology at Avignon in 1614, he was appointed official diocesan teacher of theology and superintendant of theological education. In 1616 he assumed the chair of philosophy at Aix-en-Provence, where he taught Aristotelian philosophy for the next six years. He voiced his dissatisfaction with Aristotelianism in his first published work, *Exercitationes Paradoxicae Adversus Aristoteleos (Paradoxical Exercises against the Aristotelians)* (1624), in which he claimed to have taught his students the principles of Aristotelianism, only to refute them with sceptical arguments. When the college at Aix was turned over to the Jesuits in 1622, Gassendi and the rest of the faculty were forced to leave their positions. Since his student days, he had been a member canon of the Cathedral of Digne. He was appointed provost of the Cathedral in 1634, a position which he held for the rest of his life.

Apart from his administrative position at the Cathedral of Digne, Gassendi relied on patronage for his scholarly work. His first patron was the polymath and humanist Nicolas-Claude Fabri de Peiresc (1580-1637), councillor of the Parlement of Aix and patron of arts and letters. Gassendi and Peiresc conducted scientific experiments together and corresponded about astronomy and philosophy. During the years until Peiresc's death, Gassendi divided his

time among Aix, Digne, and Paris. During the winter of 1628, he travelled with his friend the *libertin érudit* François Lullier (*c.*1600-51) to Holland where he met a number of important intellectual figures, including the atomist Isaac Beeckman (1588-1637), who reinforced his resolve to restore the philosophy of Epicurus, an enterprise that became his life's work.

In 1630, Gassendi published two works not directly related to his Epicurean project. In the *Epistolica exercitatio, in qua principia philosophiae Roberti Fluddi, medici, reteguntur* he responded to Marin Mersenne's request for an evaluation of the natural philosophy of the Paracelsian Robert Fludd (see Mersenne). In the *Parhelia, sive soles quat[u]or qui circa verum apparuerunt Romae, die XX mensis martii anno 1629* he gave an account of a strange appearance of parhelia (or the appearance of multiple suns and patches of shimmering light around the sun). Two years later, in *Mercurius in sole visus et Venus invisa Parisiis anno 1631*, he confirmed Kepler's prediction of Mercury's transit of the sun.

Gassendi suffered a major loss when Peiresc died in 1637. Losing not only his patron but also a close and valued friend, he wrote little for the next four years. In 1641 he published a memorial to his friend, *Viri illustris Nicolai Fabricii de Peiresc Senatoris Aquisextiensis vita*. Having returned to Provence to perform his clerical duties, Gassendi was invited by Louis-Emmanuel de Valois, count of Alais, a man closely connected to the royal family, 'to wait upon' him, that is, to enter a client-patron relationship which continued until Valois' death in 1653.

Gassendi was deeply involved in the major developments in natural philosophy in his day. He published an exposition of Galileo's new science of motion, *De motu impresso a motore translato* (1642), which contains the first correct statement of the principle of inertia in print (see Galileo, G.). Prior to the publication of Descartes' *Meditations* in 1641, Mersenne invited Gassendi, among other philosophers and theologians, to comment on the manuscript of the work. These comments, first published as the fifth set of 'Objections', were enlarged in Gassendi's *Disquisitio metaphysica, seu dubitationes et instantiae adversus Renati Cartesii metaphysicam et responsa (Metaphysical disquisition, or doubts and instances against the metaphysics of René Descartes and responses)* (1644), which contains Gassendi's objections to the *Meditations*, Descartes' replies, and Gassendi's rejoinders (see Descartes, R.). He also published an account of the new astronomy in *Institutio astronomica juxta hypotheseis tam veteram quam Copernici et Tychonis Brahei* (1647).

In 1645, Gassendi received an appointment as professor of mathematics at the Collège Royal. In 1653, after Valois' death, Gassendi came under the protection of Henri-Louis Habert, lord of Montmor, a patron of natural philosophy, with whom he lived until his own death in 1655.

During the last decade of his life, Gassendi published the products of his extensive work on the life and philosophy of Epicurus. The three major Epicurean works include *De vita et moribus Epicuri libri octo* (1647), *Animadversiones in decimum librum Diogenis Laertii, qui est de vita, moribus, placitisque Epicuri (Observations on Book X of Diogenes Laertius, Which is about the Life, Morals, and Opinions of Epicurus)* (1649), and the *Syntagma Philosophicum*, published posthumously in 1658.

2 Gassendi's Epicurean project

The unifying theme of Gassendi's work was his project to restore the philosophy of Epicurus, making it acceptable to orthodox Christians. He envisaged this Christianized Epicureanism as a complete philosophy that would replace Aristotelianism, which had been dominant in the universities for centuries (see Aristotelianism in the 17th century; Epicurianism).

Gassendi's rejection of Aristotelianism is the theme of his 1624 *Exercitationes Paradoxicae Adversus Aristoteleos*. Drawing on the Pyrrhonian scepticism of Sextus Empiricus, Gassendi attacked the epistemological and metaphysical foundations of Aristotelianism (see Pyrronism). He rejected Aristotle's philosophy of matter and form, as well as his ideal of demonstrative certain knowledge as the epistemological goal of natural philosophy (see Aristotle). Rather than sinking into sceptical despair, however, Gassendi adopted what Popkin (1979) has happily called 'mitigated scepticism', advocating a science of appearances which can attain at best probable knowledge of things based on their appearances. He denied the possibility of knowing essences and explicitly allied himself with the nominalists (see Nominalism). An empiricist epistemology and a nominalist metaphysics were central themes in his philosophical writings. In this early work as well as in correspondence from the second half of the 1620s, he expressed a growing interest in Epicureanism as a substitute for Aristotelianism.

Although there is evidence that Gassendi was attracted to the philosophy of Epicurus from the mid-1620s, he did not defend Epicureanism in print until the 1640s. During the intervening years, his Epicurean project grew from the straightforward humanist task of translating Book X of Diogenes Laertius' *Lives of Eminent Philosophers*, one of the major sources for knowledge of Epicurus' writings, to a full-fledged rehabilitation of Epicureanism (see Diogenes Laertius). During the early 1630s, Gassendi produced a manuscript of his Epicurean commentary, which he shared chapter-by-chapter with Peiresc and some other friends. By 1634, he completed a draft of what would later be published as *De vita et moribus Epicuri* (1647). During 1641 and 1642, he wrote a series of letters containing a sketch of his Epicurean project to his patron, the new governor of Provence, Louis-Emmanuel de Valois. In *De vita et moribus Epicuri*, Gassendi defended Epicurus against the allegations of decadence and immorality that had dogged his reputation since antiquity. In 1649, he published *Animadversiones in decimum librum Diogenis Laertii*, a book still conceived in the humanist format as a commentary on an ancient text. The posthumous *Syntagma Philosophicum* (1658) was the culmination of Gassendi's Epicurean project, incorporating material from contemporary natural philosophy into an exposition of Epicureanism.

At every stage of his Epicurean project, Gassendi was concerned to modify certain of the ancient philosopher's doctrines in order to make his philosophy consistent with Christian orthodoxy. He rejected the theologically objectionable components of Epicureanism: polytheism, a corporeal conception of the divine nature, the negation of all providence, the denial of Creation *ex nihilo*, the infinitude and eternity of atoms and the universe, the plurality of worlds, the attribution of the cause of the world to chance, a materialistic cosmogony, the denial of all finality in biology, and the corporeality and mortality of the human soul. Gassendi replaced these doctrines with a Christianized atomism which asserted the creation of the world and its constituent atoms by a wise and all-powerful God who rules the world providentially, the existence of a large but finite number of atoms in a single world, the evidence of design throughout the Creation, a role for final causes in natural philosophy, and the immortality and immateriality of the human soul.

Gassendi's theology was thoroughly voluntarist. Voluntarism is an interpretation of God's relationship to the Creation which insists on God's omnipotence and his absolute freedom of will. Nothing exists independently of him, and nothing that he creates can bind or impede him. Emphasizing the contingency of the world on God's will, Gassendi believed that God's omnipotence is in no way constrained by the Creation, which contains no necessary relations that might limit God's power or will. 'There is nothing in the universe that God cannot destroy, nothing that he cannot produce, nothing that he cannot change, even into its opposite qualities' (1658 vol 1: 308). Consequently there are no universal or eternal essences of created things. Even the laws of nature lack necessity. 'He is free from the laws of nature, which he constituted by his own free will' (1658 vol 1: 381). The laws of nature have no existence apart from serving as descriptions of the regularities we observe in the operations of nature. In contrast to Descartes, Gassendi never identified any particular propositions as laws. Like everything else God created, he can negate them. He could have created an entirely different natural order if it had pleased him to do so. Similar to other voluntarists, for example Ockham, Gassendi believed that God's will was constrained only by the law of non-contradiction and that nothing God creates can prevent him from acting directly on the Creation. God does not directly produce all the motions of bodies, nor does he simply create bodies in the beginning, leaving them to act in accordance with their individual natures as bodies. He makes use of second causes - that is, natural causes that he has created. But he can always intervene and act directly if he wants. Nothing he creates constrains him in any way. 'God… is the most free; and he is not bound, as he can do whatever… he wishes' (1658 vol 1: 309). Nominalism was one important implication of Gassendi's voluntarism. The existence of universals, even universals created by God, would limit God's freedom of action. Gassendi's voluntarism and anti-essentialism played a central role in his debate with Descartes over the *Meditations*. These assumptions infused every part of his version of Epicureanism.

Gassendi employed humanist methods in philosophy. That is, he insisted on finding an ancient model for his philosophical views, and he wrote in a style marked by frequent allusions to and quotations from classical authors. Each section of his massive *Syntagma Philosophicum* begins with a summary of all existing views on the subject at hand. Only after having rehearsed all the traditional views did he embark on his own account of a subject. He articulated his own views in dialogue with his ancient model, Epicurus. In this respect, he was traditional, even though the content of his philosophy in many ways broke from traditional ideas. He claimed that he chose Epicurus as his model because his atomistic physics and hedonistic ethics could be more readily reconciled with

‘the Sacred Faith’ than any of the other ancient schools of philosophy. Presented as a complete philosophy to replace Aristotelianism, the *Syntagma Philosophicum* is divided into three large sections, entitled ‘Logic’, ‘Physics’, and ‘Ethics’.

3 Logic

The first part of the *Syntagma Philosophicum*, entitled ‘Logic’, is the culmination of Gassendi’s lifelong considerations of this subject. He proposed to substitute the canonic of Epicurus for Aristotelian dialectic which he judged to be useless for obtaining knowledge of nature. He considered logic to be the art of thinking well. He developed it as a theory of knowledge and a primitive psychology to explain how ideas get into the mind rather than simply a study of the forms of syllogism and the relationships among propositions, although he discussed these topics as well. He advocated an empirical approach to knowledge of the world, one modelled on the Epicurean canonic.

In his early *Exercitationes Paradoxicae Adversus Aristoteleos*, Gassendi attacked Aristotelian dialectic as overly complex and abstruse, useless as a method for making new discoveries. He used the classical sceptical arguments to question the reliability and validity of sensory experience. Since Aristotelian demonstration requires premises based on experience, Gassendi argued that the syllogisms involved cannot produce certain knowledge about the world. Moreover, since the conclusion of a syllogism contains no information not already present in the premises, syllogistic demonstration is incapable of producing new knowledge. Gassendi concluded that the entire method of Aristotelian demonstration is without foundation or utility. Not satisfied with the suspension of judgment advocated by the ancient sceptics, he opted for a middle way, ‘mitigated scepticism’. He based this approach on a series of rules or canons, which he drew from the Epicurean Canonic. These canons defined sensations, ideas, propositions and syllogisms. Gassendi then elaborated a criterion of truth, based on empiricist assumptions.

Adopting the fundamental empiricist doctrine that all ideas contained in the mind derive from the senses, Gassendi distinguished between two sorts of truth, what he called ‘truth of existence’ and ‘truth of judgment’. Truth of existence refers to the content of sensation itself. That it is what it is must be true. In this respect the senses are infallible. The sensation of gold is a sensation of gold, whether or not the object sensed is really gold or even exists. Truths of judgment, in contrast to truths of existence, are truths about the judgments we make about sensations. They are fallible, since they can assert statements about the world which might not be true. An example of such a judgment would be the proposition that this object is gold.

On the basis of this distinction, Gassendi argued that sensations, which are truths of existence that he called ‘appearances’, provide the basis for our knowledge of the world. This knowledge cannot penetrate to the inner natures of things precisely because it is knowledge of how they appear to us. On the basis of the appearances, however, it is possible to seek causal explanations, with the understanding that such reasoning is always conjectural and subject to revision in the face of further knowledge. This science of appearances can never achieve certainty, only probability, but such probability is adequate for our needs. In settling for probability rather than certainty as the epistemic goal of natural philosophy, Gassendi was rejecting the traditional Aristotelian and Scholastic conception of *scientia* or demonstrative knowledge. His redefinition of the goals of natural philosophy influenced the conceptions of later natural philosophers such as Robert Boyle John Locke and Isaac Newton.

4 Physics

The ‘Physics’ is by far the longest part of Gassendi’s *Syntagma Philosophicum*. Here he laid down the basic principles of his natural philosophy. He claimed that the fundamental components of the natural world are atoms and the void.

Gassendi began his account of the nature of things in general with a discussion of the void, since he thought that the entire universe is contained in empty space. In accordance with traditional discussions, Gassendi classified the void into three categories: the separate, extra-cosmic void; the interparticulate, interstitial, or disseminated void; and the *coacervatum*, produced by collecting a number of interstitial voids, usually by means of some kind of mechanical device. The question of the existence of void led him into a discussion of space and time more generally. Borrowing from the Renaissance Platonist Francesco Patrizi, Gassendi argued that space is neither substance nor accident but, rather, a kind of incorporeal extension. He thus avoided Aristotle’s error of

confounding dimensionality and corporeality. Space, according to Gassendi, continues to exist even when the matter contained in it moves away or ceases to exist.

The large extramundane void is the space in which God created the universe. It is boundless, incorporeal extension. Following Epicurus, Gassendi argued that the existence of interstitial void is a necessary condition for motion; for without void spaces, there would be no place into which particles of matter could move. Other Epicurean arguments included the saturation of water with salt, the dissemination of dyes through water, the penetration of air by light, heat and cold, all of which he assumed to be corpuscular. These phenomena seemed to require empty spaces between the particles composing material bodies. Gassendi also borrowed several arguments from Hero of Alexandria (*fl*. 62 AD), who had drawn an analogy between the matter composing bodies and a heap of grain or sand. Just as the individual grains are separated from each other by air or water, so the particles composing bodies are separated by small void spaces. The compressibility of air, which Hero had demonstrated by means of several inventions, including the pneumatic cannon and the aeolipile (a prototype of the steam engine), seemed to call for the existence of interstitial void between the material particles composing air.

Gassendi drew on contemporary natural philosophy to argue for the existence of the *coacervatum* void, especially the barometric experiments of Torricelli and Pascal. Gassendi accepted the mechanical explanation of the suspension of mercury in the barometer, as the result of atmospheric pressure, thereby rejecting the Aristotelian explanation which appealed to the paradigmatic occult quality, the *horror vacui*. He argued that the space in the tube above the mercury is void.

Gassendi's atoms are perfectly full, solid, hard, indivisible particles, so small that they fall below the threshold of sense. Following Lucretius, the Roman expositor of Epicureanism, Gassendi noted several commonly observed phenomena that lend support to the existence of such atoms. Wind is evidence that invisible matter can produce visible, physical effects. So is the fact that paving stones and ploughshares wear away because of constant rubbing, even though individual acts of rubbing produce no discernible change. The passage of odours through the air can be explained in terms of tiny particles travelling from the source of the odour to the nose.

Atoms possess only a few primary qualities: magnitude and figure, resistance or solidity, and heaviness. Gassendi appealed to observations using the new instrument, the microscope, to provide evidence for their small size. He also cited traditional observations such as the dispersion of pigment in water and the large quantity of smoke emanating from a smouldering log. He argued for the indivisibility of these atoms by appealing to Zeno's paradoxes, which he interpreted as implying the absurdity of the idea of the infinite divisibility of matter (see Zeno of Elea).

Matter, in the form of atoms, is the material principle in Gassendi's world. The efficient principle explains the causal structure of the world. The first cause is God, who created the world, including the atoms. Second causes, the natural causes operating in the physical world, are reduced to collisions among atoms moving in void space. In contrast to Epicurus who had claimed that an endless series of worlds is being produced by an eternal series of chance collisions among an infinite number of uncreated atoms, Gassendi argued that the world and its constituent atoms - a large, but finite number of them - had been created by God, who continues to rule the world providentially, with special providence for humankind. Rejecting the atomic swerves or *clinamen* that Epicurus had introduced to account for the collision of atoms that would otherwise only fall downward in parallel paths, Gassendi maintained that the motions of atoms had been caused by God at the beginning. Gassendi argued for God's providential relationship to the creation on the basis of an extended argument from design.

Having established the material and efficient principles of things, Gassendi proceeded to argue that all the qualities of bodies can be explained in terms of the motions and configurations of their constituent atoms. He gave mechanical, atomistic explanations of the whole range of qualities, including rarity and density, transparency and opacity, size and shape, smoothness and roughness, heaviness and lightness, fluidity and firmness, moistness and dryness, softness and hardness, flexibility and ductility, flavour and odour, sound, light, and colour. He concluded his account of qualities with a chapter on the so-called occult qualities, in which he argued that there is no action at a distance and that even apparently occult qualities such as magnetism and the sympathies and antipathies favoured by the Renaissance naturalists can be explained in mechanical terms.

In this first section of the 'Physics', Gassendi created the blueprint for his version of a mechanical philosophy of

nature, a Christianized version of Epicurean atomism, designed to replace Aristotle's *Physics* and consonant in spirit, if not detail, with that of his contemporary Descartes (see Descartes, R. §11). In the remaining sections of the 'Physics', Gassendi tried to explain all the phenomena of the world in these terms. His work paralleled Aristotle's treatises *De caelo*, *Meteorologica*, *De partibus animalium*, and *De anima*; but he took account of recent developments in natural philosophy. His intentions to mechanize notwithstanding, he wrote eclectically, often appealing to concepts and terms drawn from traditions such as Aristotelianism, Renaissance naturalism and even alchemy (see Alchemy).

Having considered the universal principles of physics, Gassendi proceeded to consider the created world, starting with celestial things. He discussed the following questions: the substance of the sky and stars; the variety, position, and magnitude of the stars; the motions of the stars; the light of the stars; comets and new stars; and the effects of the stars. In his youth, Gassendi had endorsed the new Copernican astronomy enthusiastically. The condemnation of Galileo in 1633 dampened his enthusiasm, at least in print, where he expressed sceptical doubts about being able to prove any of the three main world systems - Ptolemaic, Copernican, and Tychonic - conclusively. In the *Syntagma Philosophicum*, he proposed the system of Tycho Brahe as a compromise approved by the Church, but not before having stated that the Copernican theory was 'more probable and evident'.

As for the effects of the stars, Gassendi thought astrology 'inane and futile.' He rejected the possibility that the stars cause terrestrial and human events. Sidereal and planetary configurations may be signs of some events on earth, such as the seasons or the weather, but they are not their causes. Moreover, the ability to prognosticate the future is the prerogative of God alone. Gassendi found horoscopes based on the moment of nativity ridiculous. Why, he asked, should the heavenly bodies have more influence at the moment of birth than at any other moment in a person's life? He thought that the principles of astrology were based on insufficient evidence, and that astrologers often resort to deception.

Having discussed the heavens, Gassendi turned his attention to terrestrial phenomena, considering inanimate things first. He described the properties of Earth, the distribution of water and land, the tides, subterranean heat and the saltiness of the sea. He then turned to 'meteorological' phenomena, which included winds, rain, snow, ice, lightning and thunder, rainbows and parhelia, and the Aurora Borealis. Shifting his attention to smaller things, he wrote about stones and metals, paying particular attention to recent observations of the magnet and to the question of the transmutation of gold,which he considered possible. Finally, among inanimate things, he discussed plants which, he claimed, Epicurus had believed to be inanimate, that is, lacking soul. Gassendi discussed the varieties of plants and their parts, considering their various physiological processes, including grafting, nutrition, germination, growth, and death.

The final section of the 'Physics' was devoted to terrestrial living things, or animals. This section contains discussions of the varieties of animals, the parts of animals - which he described in explicitly finalistic terms, and various physiological topics including generation, nutrition, respiration and motion. Gassendi devoted about half of this lengthy section to the topics of sensation, perception and the immortality of the human soul.

Gassendi's argument for the immortality of the soul was a key factor in his Christianization of Epicureanism. Epicurus, in order to eliminate fear of death and anxiety about punishment and reward in the afterlife as major sources of human distress, claimed that the soul is material and mortal (see Epicurianism §13). In addressing the nature of the soul, Gassendi said that the soul is what distinguishes living things from inanimate things. Adopting the distinction between *anima* and *animus* directly from Lucretius, Gassendi argued that the *anima* or sentient soul is material and present throughout the body but that the *animus* or rational soul is incorporeal. Like animals, humans possess an *anima*, but the possession of an *animus* distinguishes them from animals.

The *anima* was thought to be composed of very fine and intensely active atoms, 'like the flower of matter'. It is the principle of organization and activity for the organism and the source of the animal's vital heat. It is also responsible for perception, forming the imagination or 'phantasy', a physical organ which forms images derived from perception. The *anima* is transmitted from one generation to the next at the moment of conception.

The *animus* or rational soul is an incorporeal substance, created by God, infused in the body and functioning like an informing form. Gassendi argued on several traditional grounds for the incorporeality of the rational soul. We know that it is distinct from the corporeal imagination or phantasy, because we can understand some things of

which we cannot form images, for example, that the sun is 160 times larger than the earth. Moreover, unlike corporeal things, the rational soul is capable of reflecting on itself. It is also able to reflect on the nature of universality *per se*, whereas animals, possessing only the corporeal *anima*, are limited to forming universal concepts without having the ability to reflect on them abstractly. Gassendi's claim that the rational soul, in contrast with the animal soul, is incorporeal established one of the boundaries of his mechanization of the world.

Having established the immateriality of the rational soul, Gassendi proceeded to argue for its immortality. He wrote of this topic as the 'crown of the treatise' and the 'last touch of universal physics'. Although he knew of the soul's immortality on the basis of 'the Sacred Faith', he supported this article of faith using philosophical and physical arguments, his response to the Fifth Lateran Council's call on philosophers in 1513 'to use all their powers, including natural reason, to defend the immortality of the soul'. Epistemologically similar to all reasoning in natural philosophy, the conclusions drawn from physics and philosophy could at best be highly probable. Nevertheless Gassendi was certain of the soul's immortality which was ultimately grounded in faith.

Gassendi argued on the basis of physics that the soul is immortal because it is immaterial. Lacking matter, an immaterial thing 'also lacks mass and parts into which it can be divided and analyzed.' This argument was similar to arguments used by Kenelm Digby, Henry More and others. Another approach, which he called 'moral', argued for immortality on the grounds that the afterlife is necessary in order to compensate for various injustices in this life. Gassendi went on to argue against many detractors of the soul's immortality, especially Epicurus against whose arguments he devoted an entire chapter of the *Syntagma Philosophicum*.

Gassendi's argument that the rational soul is immaterial and immortal provides evidence that he was not a materialist, despite arguments to the contrary by some scholars, notably O.R. Bloch. Bloch (1971) bases his claim that Gassendi defended the immateriality of the rational soul only in deference to the Church on two claims: that Gassendi ascribed many aspects of cognition to the material *anima*; and that he did not fully articulate his arguments for an immaterial *animus* until 1642. Bloch's interpretation of Gassendi as a clandestine materialist belies the fact that Gassendi's assertion of the existence of God, angels, demons and an immaterial immortal soul is to be found throughout his Epicurean writings. These topics appear as early as a manuscript outline of his Epicurean project which he sent to Peiresc in 1631 at the inception of his project, in his letters to Valois in the 1640s, and in the posthumous *Syntagma Philosophicum*.

Gassendi's Christianized Epicureanism had significant influence on natural philosophy in the latter half of the seventeenth century. Brought to England in Walter Charleton's translation in the 1650s, Gassendi's ideas influenced the thinking of Robert Boyle and Isaac Newton. His epistemology was also important for the thought of John Locke.

5 Ethics and political philosophy

The 'Ethics' comprises the third and final part of the *Syntagma Philosophicum*. Here, Gassendi fulfilled his youthful determination to Christianize all parts of Epicureanism. Epicurean ethics was founded on the principle that pleasure is the end of life (see Hedonism). Pleasure, according to Epicurus, consists of freedom from bodily pain and freedom from mental turmoil. The greatest pleasure is associated with the absence of both anxiety and physical pain, a state of tranquillity. The achievement of tranquillity gives individuals self-sufficiency, so that they are free to pursue their own pleasures. Epicurus recognized that not all pleasures are of equal value, and gave reason the role of calculating pleasure and pain. In this way, he could claim that long-term pleasure was of greater value than short-term pleasure that might lead to long-term pain. The freedom humans need to implement the calculus of pleasure and pain was insured by the random swerve of atoms, the *clinamen*, which added a dimension of indeterminism to the human soul which Epicurus believed to be composed of atoms. Epicurean hedonism did not receive a good press, either in antiquity or in the Christian Middle Ages. Dogged by Epicurus' reputation for atheism and moral decadence, this philosophy needed to be restored and Christianized in order to be acceptable to orthodox intellectuals in the seventeenth century. Gassendi addressed this task in the 'Ethics'.

Although Gassendi accepted the Epicurean principle that equated pleasure with the good, he reinterpreted the concepts of pleasure and human action, thereby creating a Christian hedonism which found a natural place in his providential worldview. Gassendi described four kinds of pleasure: the instinctive desire for pleasure that even irrational creatures possess; the calculated strategy of maximizing physical pleasure; the prudence of the wise who

understand that true pleasure consists of tranquillity and the absence of pain; and finally the sublime pleasure of the beatific vision of God. The prudence of the wise is based on understanding the vanity of most human desires. The wise person will employ the calculus of pleasure and pain to achieve the state of tranquillity. Gassendi united this hedonistic ethics with his providential worldview by claiming that God has instilled in humans a natural desire for pleasure and a natural aversion to pain. In this way, God guides human choices, without negating free will. The prudent pursuit of pleasure will ultimately lead to the greatest pleasure of all, presence with God in heaven.

Consistent with the voluntarist theology that informed all of Gassendi's philosophy, the 'Ethics' presumed both divine and human freedom. Human freedom is a necessary concomitant of voluntarism, for if human actions are inexorably determined, that determinism will limit God's freedom to intervene in their lives. Gassendi considered true freedom, *libertas*, to be the freedom of indifference, the ability of the mind to make judgments and take action without being determined in one direction or another. This kind of freedom gives reason a central role in moral deliberation. Gassendi contrasted *libertas* with *libentia* - spontaneity or willingness - which is characteristic of boys, brutes and stones, creatures that are impelled to move in certain ways, but not on the basis of judgments deriving from the freedom of indifference.

Gassendi's emphasis on freedom, both human and divine, led him to consider the question of predestination, which was the main context for discussions of freedom and determinism in the post-Reformation setting. How can human action be free if God has foreknowledge of who will be saved and who will be damned? Influenced by the Jesuit Luis de Molina's moderate stance on the question of predestination (see Molina, L. de; Predestination), Gassendi argued that God created people free to choose, even though he knows from his eternal viewpoint how they will choose. Gassendi, following Molina, claimed that divine foreknowledge does not interfere with human freedom. Similar concerns with free will led Gassendi to reject both Stoic fatalism and astrology.

Gassendi's hedonism led him to formulate a political philosophy based on the idea of *pactum* or contract (see Contractarianism). Starting from the idea of a hypothetical state of nature in which there was no secure ownership of property, a state which would inevitably degenerate into turmoil and conflict, Gassendi argued that individuals could secure greater happiness for themselves by forming societies. These societies are based on pacts or contracts in which both individual rights and property rights are defined and in which the weaker are protected from the stronger. The contracts establish rights, which Gassendi considered natural in the sense that they follow from the calculus of pleasure and pain. Civil society is thus a natural outcome of human nature. A system of justice comes into being to restore rights that have been violated and to prevent further violations. Gassendi favoured monarchy as simpler and more efficient than the other traditional forms of government. However, he argued that the power of the monarch remains answerable to the consent of the governed who first established the contract. He was therefore opposed to absolutism on the grounds that the absolute monarch had severed his relationship with the governed and was consequently answerable to no one. Gassendi developed his political philosophy in close contact with Thomas Hobbes, and his ideas had a profound influence on John Locke, who is usually named as the founder of the liberal tradition in political philosophy (see Liberalism).

See also: Atomism, ancient; Bernier, F.; Democritus; Freedom, divine; Scepticism, Renaissance; Will, the

MARGARET J. OSLER

List of works

Gassendi, P. (1658) *Opera Omnia (Collected works)*, Lyon; repr. Stuttgart-Bad Canstatt: Friedrich Frommann Verlag, 1964.(The most easily accessible edition of Gassendi's works.)

Gassendi, P. (1972) *The Selected Works of Pierre Gassendi*, trans. C. Brush, New York: Johnson Reprint.(English translations of selections from *Exercitationes paradoxicae adversus Aristoteleos*, *De motu impressu*, *Rebuttals to Descartes* from the *Disquisitio metaphysica*, and *Syntagma philosophicum*.)

Gassendi, P. (1624) *Exercitationes Paradoxicae Adversus Aristoteleos (Paradoxical Exercises against the Aristotelians)*; Books I and II ed. and trans. B. Rochot as *Dissertations en forme de paradoxes contre les Aristotéliciens*, Paris: Vrin, 1959.(French translation with facing pages in Latin of Gassendi's first published work.)

Gassendi, P. (1630a) *Epistolica exercitatio, in qua principia philosophiae Roberti Fluddi medici, reteguntur (An Epistolary Exercise, in which the principles of the philosophy of the physician Robert Fludd are brought to light)*, in *Opera omnia*, vol. 3.(Critical response to Robert Fludd's alchemical philosophy of nature.)

Gassendi, P. (1630b) *Parhelia, sive soles quat[u]or qui circa verum apparuerunt Romae, die XX mensis Martii*

anno 1629 (Parhelia, or the four suns which appeared real at Rome on 20 March 1629), in *Opera omnia*, vol. 3 as *Parhelia, sive soles Quatuor Spurii, qui circa verum apparverunt Romae Anno MDCXXIX, Die XX Martij (Parhelia or four spurious suns…)*.(Deals with optical illusions in astronomy.)

Gassendi, P. (1632) *Mercurius in sole visus et Venus invisa Parisiis anno 1631 (Mercury seen in the sun and Venus unseen at Paris in 1631)*, in *Opera omnia*, vol. 4.(On observations of Mercury's transit of the Sun.)

Gassendi, P. (1632-3) *Lettres Familières à François Lullier pendant l'hiver 1632-1633 (Familiar letters to François Luillier during the winter of 1632-1633)*, ed. B. Rochot, Paris: Vrin, 1944.(Discusses his reading and travels during this year.)

Gassendi, P. (1641) *Viri illustri Nicolai Fabricii de Peiresc Senatoris Aquisextiensis vita (The Life of the famous Senator of Aix, Nicholas Fabri de Peiresc)*, in *Opera omnia*, vol. 5; French trans. R. Lasalle and A. Bresson as *Peiresc, 1580-1637. Vie de l'illustre Nicolas-Claude Fabri de Peiresc*, Paris: Belin, 1992.(Biography of Gassendi's patron, the polymath Peiresc.)

Gassendi, P. (1642) *De motu impresso a motore translato (On motion impressed by the moved mover)*, in *Opera omnia*, vol. 3.(An exposition of Galileo's new science of motion and contains the first correct statement of the principle of inertia in print.)

Gassendi, P. (1644) *Disquisitio Metaphysica seu dubitationes et instantiae adversus Renati Cartesii Metaphysicam et responsa (Metaphysical disquisition, or doubts and instances against the metaphysics of René Descartes and responses)*, ed. and trans. B. Rochot, Paris: Vrin, 1962.(French translation with facing pages in Latin of Gassendi's objections to Descartes' *Meditations*, Descartes' replies, and Gassendi's rejoinders.)

Gassendi, P. (1647) *Institutio astronomica juxta hypotheseis tam veteram quam Copernici et Tychonis Brahei (Astronomical instruction according to the ancient hypotheses as well as those of Copernicus and Tycho Brahe)*, in *Opera omnia*, vol. 4.(An account of the astronomical theories of Ptolemy, Copernicus and Tycho Brahe.)

Gassendi, P. (1649) *Animadversiones in decimum librum Diogenis Laertii, qui est de vita, moribus, pacitisque Epicuri (Observations on Book X of Diogenes Laertius, Which is about the Life, Morals, and Opinions of Epicurus)*, repr. New York and London: Garland, 3 vols, 1987.(Translation of and commentary on Book X of Diogenes Laertius, *Lives of Eminent Philosophers*. Not in *Opera omnia*.)

Gassendi, P. (1658) *Syntagma philosophicum (Philosophical treatise)*, in *Opera omnia*, vols 1 and 2.(Final version of Gassendi's Epicurean project, first published in the *Opera omnia*.)

Jones, H. (1981) *Pierre Gassendi's Institutio Logica, 1658*, Assen: Van Gorcum. (English translation of Part IV of the 'Logic' of Gassendi's *Syntagma Philosophicum*, his major statement on method.)

Murr, S. (1993) 'Pierre Gassendi - Préliminaires à La Physique *Syntagma Philosophicum,' Dix-septième Siècle*', 45 (2): 353-85.(French translation of the introductory section of Gassendi's *Syntagma Philosophicum*.)

Peiresc, N.C.F. de (1888-98) *Lettres de Peiresc*, ed. P.T. de Larroque in *Documents inédits sur l'histoire de France*, 7 vols, Paris: Imprimerie Nationale. (Contains Gassendi's correspondence with Peiresc.)

References and further reading

Bloch, O.R. (1971) *La philosophie de Gassendi: Nominalisme, matérialisme, et métaphysique (Gassendi's philosophy: nominalism, materialism and metaphysics)*, The Hague: Martinus Nijhoff.(The major modern study of Gassendi's philosophy. Argues that Gassendi was a materialist.)

Brundell, B. (1987) *Pierre Gassendi: From Aristotelianism to a New Natural Philosophy*, Dordrecht: Reidel. (Argues that Gassendi sought to replace Aristotelianism with a modified Epicureanism, which would then serve as a complete philosophy of nature.)

Descartes, R. (1641) *Meditations on First Philosophy and Objections and Replies*, in *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff, D. Murdoch, and A. Kenny, Cambridge: Cambridge University Press, 1984.(Gassendi wrote the fifth set of *Objections*.)

Jones, H. (1981) *Pierre Gassendi 1592-1655. An Intellectual Biography*, Nieuwkoop: B. de Graaf.(A biography of Gassendi which focuses on his philosophy but pays little attention to his scientific interests.)

Joy, L.S. (1987) *Gassendi the Atomist: Advocate of History in an Age of Science*, Cambridge: Cambridge University Press.(Interprets Gassendi's atomism in light of his use of the history of philosophy.)

Michael, F. (1992) 'La place de Gassendi dans l'histoire de la logique' (Gassendi's place in the history of logic), *Corpus* (Corpus des oeuvres de philosophie en langue française) 20/21: 9-36.(Discusses the historical background to and influence of Gassendi's writings on logic.)

Osler, M.J. (1994) *Divine Will and the Mechanical Philosophy: Gassendi and Descartes on Contingency and Necessity in the Created World*, Cambridge: Cambridge University Press.(Examines the influence of voluntarist theology on Gassendi's articulation of the mechanical philosophy. Contains extensive

bibliography.)
Popkin, R.H. (1979) *The History of Scepticism from Erasmus to Spinoza*, Berkeley, CA: University of California Press.(Situates Gassendi’s epistemology in the context of the sceptical crisis of the early modern period.)
Rochot, B. (1944) *Les travaux de Gassendi sur Épicure et sur l’atomisme, 1619-1658 (Gassendi’s works on Epicurius and atomism)*, Paris: Vrin.(Thorough examination of the development of Gassendi’s project to Christianize Epicurean philosophy. A basic work.)
Sarasohn, L.T. (1996) *Gassendi’s Ethics: Freedom in a Mechanistic Universe*, Ithaca, NY: Cornell University Press.(Study of Gassendi’s moral and political philosophy in relation to his physics.)

Gauḍīya Vaiṣṇavism

The philosophical school encompassing the Bengali devotees of the god Viṣṇu is traditionally known as the Gauḍīya Vaiṣṇava school. Caitanya is considered to be the founder of this school in the sense that he led a revival of Kṛṣṇa devotionalism in Bengal during the early sixteenth century, inspiring a number of contemporary intellectuals to some original speculations of a metaphysical nature. Some of these were directly related to Vedānta, others were not. Since the eighteenth century the school has also become associated through Baladeva Vidyābhūṣaṇa with the Madhva school. However, the latter propounded a dualistic doctrine, while the Gauḍīyas are believers in the inconceivable simultaneous difference and oneness of the Supreme and his creation.

1 Kṛṣṇa Caitanya (1486-1534)

Born in Nabadwip, Bengal as Viśvambhara Miśra, Caitanya was first a precocious student and then worked as a teacher of Sanskrit grammar until becoming possessed by unusually strong devotional impulses at the age of twenty-three. He took the name Kṛṣṇa Caitanya when taking monastic vows in 1510. Despite entering a Śaōkarite order, Caitanya had a deep-rooted distaste for the conclusions of exclusive nondualism (see Śaōkara). He thus reversed Śaōkara's interpretation of the essential statements of the Vedānta, taking 'Thou art that' (*tat tvam asi*) as 'Thou art his (servant)'. After his renunciation, Caitanya travelled throughout India before retiring to Puri in Orissa where he remained fully absorbed in devotional acts until his death. Caitanya's influence was born primarily of his charisma and he himself wrote nothing more than a few devotional verses (*Śikṣāṣṭaka* (Eight verses of instruction)) which encapsulate his doctrine: by singing the names of God one can be liberated from material suffering and reunite with him.

2 The Gosvāmins of Vṛndān

The brothers Sanātana (d. 1556) and Rūpa Gosvāmin (d. 1566) and their nephew Jīva Gosvāmin (d. 1612) are considered to be the foremost thinkers of the school. Sanātana's commentary on the tenth book of the *Bhāgavatapurāṇa* and his *Bhāgavatamṛta* (Essence of the *Bhāgavata*) are primarily concerned with the descriptions of the avatar Kṛṣṇa and the distinctions made between the various personal manifestations of Viṣṇu (or Kṛṣṇa) and giving them a hierarchical classification. This same preoccupation is found in all the Gauḍīya writers. Rūpa Gosvāmin is notable for his utilization of the theories of Sanskrit poeticians, the writers on literature and rhetoric, in particular that of aesthetic experience (*rasa*). Viśvanātha Kavirāja (fourteenth century) had identified aesthetic experience as *brahma-sahodara*, the 'twin of spiritual experience' and Rūpa adapted the current taxonomy of *rasa* as given in Siōgabhūpāla's *Rasasudhākara* to the *līlā* of Rādhā and Kṛṣṇa. His justification for such a jump was as follows: of the three features of the Supreme, existence (*sat*), consciousness (*cit*) and joy (*ānanda*), joy is the highest. It is to realize this joy that the Supreme creates, as joy cannot exist in a nonmanifold state. The highest truth is thus revealed in emotional experience which is reflected in our personal human relations. Just as these experiences are distilled and experienced transcendentally in works of drama or poetry, so too is an even higher transcendental emotional relation experienced with the Supreme. Just as sexual love is the most intense and dominating of the varieties of human relationships, its spiritual counterpart is also the preeminent amongst those experienced with God. Because this relationship is found in its most relishable form exclusively in God's manifestation as Kṛṣṇa, Kṛṣṇa is the highest form of God. The monistic view of God, in which personal relationships are denied, is considered a partial and defective view of the Supreme. Its monotheistic counterpart - the concept of God as possessing only certain attributes of omniscience and omnipresence - is conducive to the mood of devotion known as the 'peaceful' (*śānta*) which is also considered to be inferior to the filial and servile (*dāsya*), the friendly (*sākhya*) or protective (*vātsalya*) moods.

Jīva attacked more basic questions of doctrine in his *Bhāgavatasandarbha* (Treatise on the *Bhāgavata*), a systematic treatment of the theology of the *Bhāgavatapurāṇa*. As with the other Vedāntists, Jīva accepts the three sources of knowledge (*pramāṇas*): direct perception, logical reason and revelation (see Knowledge, Indian views of). However, Jīva considers the *Bhāgavata* to be the most perfect source of revelation which gives the clearest explanation of the *Vedāntasūtras*. Many of his key concepts are derived from the *Viṣṇupurāṇa* and a great deal of his thought concurs with that of Rāmānuja. For Jīva the problem of unity and difference is resolved through the concept of *śakti* (power): all creation is a product of God's power or energy which is innate in him. God's energies

form a unity with him just as the heat, light and sparks of a fire form a unity with the fire. God's energies are divided into three categories: internal, external and marginal. The temporal creation is external to God, comparable to a shadow which, although devoid of light depends on light for its existence and is unconscious. The internal energy is the world of God's grace; it has three aspects according to the Vedāntic categories of eternity, consciousness and bliss. The marginal energy (*taṭasthaśakti*) is the living being, a spark of consciousness who is conditioned by association with matter. The conditioned soul possesses limited freedom to choose between the internal and external energies. This ability to choose is awakened by association with God's devotees who give him a taste of the joy of the perfections available in the internal realm. Through such association, an individual becomes aware that relations in the conditioned existence are partial and unsatisfying and he learns to seek the essential relationship that he possesses with the root of all existence, the Supreme soul. Through the purifying process of devotion in practice (*sādhanābhakti*), the specific character of this emotional relationship is awakened in him. Love of God (*premabhakti*) is the ultimate goal of perfection.

3 Later Gauḍīya Vaiṣṇavas

Prominent later members of the school include Viśvanātha Cakravartī (seventeenth century) and Baladeva Vidyābhūṣaṇa (early eighteenth century). Baladeva's main contribution is his commentary on the *Vedāntasūtra*, the *Govindabhāṣya*, in which he arranges the ideas of his predecessors according to the sūtras. He also introduced certain elements of Madhvācārya's thought into the tradition, despite the fact that Madhvācārya (fourteenth century) propounded a dualist (*dvaita*) doctrine, while the Gauḍīyas style themselves as believers in the inconceivable simultaneous difference and oneness of the Supreme and his creation (*acintyabhedābhedavādins*).

See also: God, Indian conceptions of; Madhva; Vedānta

JAN K. BRZEZINSKI

References and further reading

Bhaktivedanta Swami, A.C. (1970) *The Teachings of Lord Caitanya*, New York: Bhaktivedanta Book Trust.(A good summary of the essential teachings found in Caitanya's biographies.)

Bhaktivedanta Swami, A.C. and H. Goswami (1972-85) *Śrīmad-Bhāgavatam*, Los Angeles, CA: Bhaktivedanta Book Trust, 41 vols.(A complete multivolume work.)

Brahmachari, M.N.B. (1974) *Vaiṣṇava Vedānta: The Philosophy of Śrī Jīva Gosvāmī*, Calcutta: Das Gupta and Co.(A useful introduction to some epistemological questions.)

Chakravarty, R. (1985) *Bengal Vaiṣṇavism, 1486-1900*, Calcutta: Sanskrit Pustaka Bhandar.(An excellent history of the school.)

Chaudhuri, R. (1975) *Ten Schools of the Vedānta*, Calcutta: Rabindra Bharati University.(Vol. 3: 241-365 gives a detailed account of the philosophy of Baladeva Vidyābhūṣaṇa.)

De, S.K. (1943) *Early Vaishnava Faith and Movement in Bengal*, Calcutta: Firma K.L. Mukhopadhyay, 2nd edn, 1962.(A wide-ranging and thorough study.)

Delmonico, N. Gorton (1990) 'Sacred Rapture: A Study of the Religious Aesthetic of Rūpa Gosvāmin', unpublished doctoral dissertation, University of Chicago, IL.(The best study of Rūpa Gosvāmin's religious aesthetic.)

Eidlitz, W. (1968) *Kṛṣṇa-Caitanya: Sein Leben und Seine Lehre*, Stockholm: Almqvist & Wiksell.(One of the earliest scholarly works on the Caitanya movement, it still provides a good introduction of its thought and literature.)

Elkman, S. (1985) *The Tattvasandarbha of Jīva Gosvāmīn*, Delhi: Motilal Banarsidass.(Deals with the first volume of Jīva's oeuvre which sets the stage for accepting the *Bhāgavatapurāṇa* as the main source of knowledge and the analysis of the tripartite understanding of the Supreme as *Brahma*, *paramaātman*, *bhagavān*.)

Kapoor, O.B.L. (1983) *The Religion and Philosophy of Śrī Caitanya*, Delhi: Motilal Banarsidass.(Provides insight into major theological issues in Gauḍīya Vaiṣṇavism.)

Masson, J. and Patwardhan, M.V. (1970) *Aesthetic Rapture*, Poona: Deccan College, 2 vols.(A more general discussion of *rasa*.)

Gautama, Akṣapāda

‘Akṣapāda Gautama’ (or ‘Gotama’) stands for the legendary founder of the Nyāya (‘Logic’) school of Indian philosophy, who is reputed also to be the author of its basic text, the Nyāyasūtra. This compilation of roughly 500 mnemonic sentences reached its first defined form around AD 400. Its oldest core preserves a manual of philosophical debate supplemented by elements of an early philosophy of nature and a basic soteriology. The later parts of the text deal with the number and nature of the means of valid cognition; they further treat the objects of valid cognition and discuss basic questions of metaphysical content.

Pakṣilasvāmin Vātsyāyana (*circa* second half of the fifth century AD), the author of the first preserved commentary on the *Nyāyasūtra*, refers to the founder of Nyāya as the sage Akṣapāda. This name, meaning ‘the one with eyes on his feet’, appears repeatedly in the school’s older tradition. Later anecdotes current among Indian pandits connect it with the sage’s deep philosophical contemplation and the superiority of Nyāya over its critics from the more orthodox camp, who held the Vedic revelation to be the most authoritative means of valid cognition. From the tenth century onward, the personal name Gotama or clan name Gautama (‘descendant of Gotama’) is connected with the school’s basic teachings; both names become more current in the eleventh and following centuries, often appearing next to the older ‘Akṣapāda’. The identity of this Akṣapāda Gautama/Gotama - if there ever was a historical personage of that name - remains a matter of speculation, as does his possible role in the formation of the Nyāya school and its basic text.

The *Nyāyasūtra*, divided into five chapters, reflects in its textual history the origin and early development of Nyāya itself. The ancient core of the text is preserved in chapters one and five, and could stem in this or a similar form from the second century AD. It is actually a manual of debate (*vāda*) which enumerates and concisely defines sixteen categories (*padārtha*) relevant to the orderly and successful conduct of a public philosophical debate. These are: (1) means of valid cognition, (2) objects of valid cognition, (3) doubt, (4) purpose, (5) generally accepted fact, (6) tenets, (7) members of proof, (8) reasoning, (9) decision, (10) friendly debate, (11) wrangle, (12) dispute involving mere criticism by one party, (13) fallacious reasons, (14) distortions, (15) sophistries and (16) points of defeat. An early philosophy of nature and a rudimentary soteriology were integrated into the category of objects of valid cognition (*prameya*); these are: (1) self, (2) senses, (3) body, (4) sense objects, (5) cognition, (6) ‘internal instrument’, (7) linguistic, mental and physical activity, (8) emotional and intellectual faults that instigate activity, (9) rebirth, (10) the result of activity and faults - that is, the experience of pleasure and pain, (11) suffering and (12) release. The text’s initial sentence proclaims that adequate knowledge of all sixteen categories effects the attainment of the highest good - namely, release from suffering, or liberation. In the Indian context, the school’s teachings could not be presented as a complete philosophical doctrine without this soteriological component.

Owing to its origin in a tradition concerned with methodical and sound argumentation, logical and epistemological issues are prominent in Nyāya from early on. In chapter two of the *Nyāyasūtra*, the very possibility of means of valid cognition (*pramāṇa*) is defended against attacks by an opponent who professes a functional essentialism; related attacks are also found in works ascribed to Nāgārjuna (§2). Subsequently, the four means of valid cognition accepted in Nyāya are examined in detail and established as distinct; these are sensory perception, inference, analogy or comparison, and verbal testimony (see Knowledge, Indian views of). Further means accepted by other schools, such as traditional lore, circumstantial evidence or presumption, coincidence or inclusion, and absence or negation, are shown to be included in verbal testimony and inference. The examination of sensory perception entails an exposition and defence of the doctrine of the whole as distinct from its parts (contested by Buddhist opponents), which is taken up again in chapter four and expanded with a defence of the atom as the ultimate partless constituent of matter (see Matter, Indian conceptions of §2; Sense perception, Indian views of §§4-5). In this context, the realistic outlook of Nyāya, as opposed to Buddhist idealism, is maintained. The topic of verbal testimony, instruction by a trustworthy person which may have visible or invisible objects, occasions a defence of the authority of the Veda. This places Nyāya within the larger Brahmanical tradition, although it differs from the orthodox Mīmāṃsā school in assuming a personal authorship of these texts and denying their eternity (see Mīmāṃsā §2). A digression on the impermanence of articulated sound and a refutation of its eternity underscores the difference between their philosophies of language, as does the concluding section on the referent of the word.

Like chapter two, chapters three and four proceed in a dialogical-dialectical manner, treating mainly the objects of valid cognition. The existence of a self or soul (*ātman*) is proved by various arguments; arguments based on innate dispositions imply its previous existence in other bodies. Cognition, which is momentary, is shown to be a quality of the soul, not of the body or of the 'internal instrument' called the *manas* (often misleadingly rendered as 'mind'); the *manas* functions in non-sensory cognitive events, such as recollection, and regulates cognitions in general. The senses are argued to be material, of the nature of the five elements earth, water, fire, wind and 'ether', whose specific qualities they grasp - these are smell, taste, colour, temperature and sound respectively. Sight receives special treatment as consisting in fiery rays issuing from the eyes to make contact with objects. Finally, several sections address soteriological issues, the possibility of release and the knowledge eventually leading to it. This knowledge chiefly concerns the true nature of the sense objects and can be attained by meditative contemplation and yogic practices aided by self-restraint. Friendly debate with disciples, teachers, fellow students and others striving for salvation serves the same purpose. Disputations with opponents protect one's nascent ascertainment of the truth.

See also: Nyāya-Vaiśeṣika

ELI FRANCO
KARIN PREISENDANZ

List of works

Nyāyasūtra (*c*.400 AD), trans. G. Jha, *The Nyāya-Sūtras of Gautama with the Bhāṣya of Vātsyāyana and the Vārtika of Uddyotakara*, *Indian Thought* 4-11, 1912-19; repr. Delhi: Motilal Banarsidass, 4 vols, 1984. (Complete translation together with the first preserved commentary and its subcommentary; contains profuse notes from further commentaries. However, the translation of the *Sūtra* is heavily influenced by the commentaries.)

Nyāyasūtra (*c*.400 AD), trans. W. Ruben *Die Nyāyasūtra's. Text, Übersetzung, Erläuterung und Glossar (The Nyāyasūtras: Text, Translation, Commentary and Glossary)*, Abhandlungen für die Kunde des Morgenlandes 18, 2, Leipzig: Deutsche Morgenländische Gesellschaft, 1928.(Standard critical edition and still by far the best annotated translation of the complete text. Contains useful index and glossary.)

References and further reading

Frauwallner, E. (1984) 'Erkenntnistheorie und Logik der klassischen Zeit' (Epistemology and Logic of the Classical Period), in *Nachgelassene Werke*, vol. 1, *Aufsätze, Beiträge, Skizzen (Posthumous Works, vol. 1, Essays, Contributions, Outlines)*, ed. E. Steinkellner, Österreichische Akademie der Wissenschaften, phil.-hist. Klasse, Sitzungsberichte 438, Veröffentlichungen der Kommission für Sprachen und Kulturen Südasiens 19, Wien: Verlag der österreichischen Akademie der Wissenschaften, 66-87.(Traces the development from debate to epistemology and logic by demonstrating the refinement of the Nyāya categories from an early manual of debate that survives - explained for the needs of physicians - in a classical compilation on Indian medicine.)

Junankar, N.S. (1978) *Gautama: The Nyāya Philosophy*, Delhi: Motilal Banarsidass. (Detailed account and extensive discussion of the *Nyāyasūtra* and the two earliest commentaries, concentrating on the theory of cognition and the concept of proof and inference.)

Meuthrath, A. (1996) *Untersuchungen zur Kompositionsgeschichte der Nyāyasūtras (Investigations into the Compositional History of the Nyāyasūtras)*, Religionswissenschaftliche Studien 36, Würzburg: Echter Verlag, and Altenberg: Oros Verlag.(Discusses the entire *sūtra*-text with a view to its philological-historical stratification.)

Oetke, C. (1988) *'Ich' und das Ich. Analytische Untersuchungen zur buddhistisch-brahmanischen Ātmankontroverse ('I' and the Self. Analytical Investigations into the Controversy over Ātman between Buddhists and Brahmans)*, Alt- und Neu-Indische Studien 33, Stuttgart: Franz Steiner, 247-77, 288-98. (Penetrating analytical study of the arguments for the existence of a self or soul and for cognition being its quality in the *Nyāyasūtra* and its earliest commentary.)

Oetke, C. (1991) Zur Methode der Analyse philosophischer Sūtratexte. Die Pramāṇa Passagen der Nyāyasūtren (On the Method of Analysis of Philosophical Sūtra-Texts. The Passages on Pramāṇa in the *Nyāyasūtra*), Studien zur Indologie und Iranistik, Monographie 11, Reinbek: Dr Inge Wezler.(Excellent philological-philosophical analysis of the key passages on the possibility of means of valid cognition, also of consequence for the text's relation to early Madhyamaka works.)

Preisendanz, K. (1994) *Studien zu Nyāyasūtra III.1 mit dem Nyāyatattvāloka Vācaspati Miśras II.*, Alt- und

Neu-Indische Studien 46, 1-2, Stuttgart: Franz Steiner. (Combined diachronic and synchronic study of the *Nyāyasūtra* on the soul, senses, body and sense objects, together with a fifteenth-century commentary.)

Randle, H.N. (1930) *Indian Logic in the Early Schools*, Oxford: Oxford University Press; repr. Delhi: Oriental Books Reprint Corporation, 1976.(Study of the epistemology and logic of the *Nyāyasūtra* in its wider historical context; although partially outdated, it is still valuable for the understanding of these topics in the *Nyāyasūtra* itself.)

Saha, S. (1987) *Perspectives on Nyaya Logic and Epistemology*, Calcutta/New Delhi: K P Bagchi & Company. (Although the author makes some use of later materials, he concentrates on the *Nyāyasūtra* and follows the text's order in his clear presentation.)

Gender and science

Gender-based analyses of philosophies of science have arisen at the conjunction of two other movements. First, in the 1960s it became increasingly the accepted position that scientific claims failed to reflect only an external reality. Scientific processes are not transparent; they necessarily permit cultural and social values and interests to contribute to the descriptions and explanations of nature's order. Thus gender values and interests, too, could have shaped scientific practices and claims. Second, women's movements developed powerful gender analyses of all other aspects of social relations. What resources could such accounts provide to illuminate also the practices and cultures of the sciences?

Consequently, diverse analyses have appeared showing how sexist and androcentric values and interests have shaped scientific projects, and shaped them with unfortunate results not only for gender relations, but also for the advance of both sciences and philosophy. They also examine how other values and interests, ones that are gender-neutral and ones that draw on resources found in women's lives, can have beneficial effects. Concern has been raised in the following issues. Is correcting 'bad science' sufficient to eliminate sexist and androcentric results of research? How has the exclusion of women from those groups that select what will count as scientific problems resulted in narrow and distorted representations of nature's order? How has excessive reliance on gendered meanings of nature, the scientist, and scientific method - gender coding - shaped scientific claims? If the subject of science remains coded masculine, how can women claim positions as speakers/authors of scientific research - as scientific subjects? In what ways have research procedures (technologies) been gender-coded and how has their use directed the development of subsequent sexist and androcentric technologies and applications of science? Why have standards for maximizing objectivity and good method been too weak to prevent of knowledge-decreasing values and interests from shaping the results of research?

1 Origins and scope

By the late 1960s, the Enlightenment vision of science had been undermined by views that observations are theory-laden, our beliefs form a network such that none are in principle immune from revision, and our best theories are underdetermined by any possible collection of evidence (see Observation §§3-4; Underdetermination). Thus empirical evidence could no longer be regarded as foundational. Slack in the procedures of the scientific system of belief-sorting permits the play of local values and interests across the entire fabric of scientific practices and claims - values and interests that can advance as well as retard the growth of knowledge. Moreover, there are no longer good grounds for assuming the unity of science, since neither physics' formally expressed laws of nature, nor the logic supposed to ground them, can be regarded as foundational in the sense of being assumed to have escaped the play of values and interests (see Unity of science). Sciences are embedded in their surrounding cultures - or, as the point was later made, co-constituted with them.

These findings made it difficult to reserve a subject matter for the philosophy of science that was safe from the legitimate gaze of historians and, later, sociologists, anthropologists, political theorists and even literary critics. After all, what is the point of philosophies of science that cannot account for the successes of science? In this first stage of post-Enlightenment philosophy of science, feminists raised issues about how gender relations, too, had shaped the sciences and their philosophies.

A second stage focused on one or another specific science, since it was thought that science in general had turned out to be too complicated to permit any comprehensive philosophical claims. Participation in this stage included Evelyn Fox Keller's works on, and in, physics and biology and Helen Longino, Elisabeth Lloyd and others in, and on, biology. Ethnographies of scientific practices and culture meanwhile were produced by historians, sociologists and anthropologists (for example, Donna Haraway's on twentieth-century primatology (1989) and Sharon Traweek's on contemporary physics (1988)). Feminist political theorists' questions about philosophical issues in the social sciences were understood often to be relevant also to the natural sciences. By the late 1980s, a third stage had produced a revival of more general philosophic concerns. These analyses have been able to draw on both earlier stages, and on the resources of literary and other cultural analyses of representation, narrative, discourse and rhetoric.

Feminism does not introduce gender into science or the philosophy of science, for it is already there. Instead,

feminism introduces gender as an analytic category (parallel to economic or Christian categories) into the philosophy and social studies of science discourses. As Keller puts the point, feminism tries to locate the scientific subject in *his* (masculine) practices and *his* rhetoric of neutrality that deny *his* very existence (Fox Keller 1983: 16-). Determining where this occurs is an empirical project. Moreover, in these accounts, gender is understood as socially constructed, not as determined by sex differences. As such, gender is not only a property of individuals but, far more importantly, of institutions and of symbolic systems. Thus science and the military are more masculine than are nursing or social welfare in their gendered divisions of activities and social meanings.

Feminist science studies were initially focused on six distinct areas. One, where have women been located in sciences' institutions, and what have been the obstacles to their entrance, retention and advancement? What are the particular consequences for the content of knowledge in each scientific field of the exclusion or restriction of women? Two, in what ways are scientific applications and technologies sexist or androcentric? Are these only 'misuses and abuses', or are blueprints for them provided by research methods themselves? Three, where are there sexist or androcentric silences and distortions in the results of research in biology and the social sciences? How do these biases reflect back on 'logics' of research and explanation for the natural sciences? Four, how do gendered models and meanings of nature, the scientist and the research process distort scientific representations of nature and social relations? Five, how are central epistemological regulatory notions - objectivity, good method, rationality, the abstract individual knower - conceptualized to favour the values and interests of the dominant groups? And six, what is the role of science education in exacerbating these problems?

Feminist work in other disciplines has been self-transforming. Women of colour in the West, working-class women, third-world women, sexual minorities and others marginalized by the dominant discourses have all challenged the centring of concerns and perspectives characteristic primarily of privileged women in the West. Thinking from the margins has become a crucial theoretical and empirical strategy for many 'centre' women, too. This work has only begun for the philosophy and social studies of science. Nevertheless, outlines for such projects have already been sketched out in the work of a few thinkers such as Harawayand Harding, and Shiva (1989) (see Postcolonial philosophy of science). The serious attention to issues of self-reflexivity characteristic of feminist work more generally provides stimulus and resources for the philosophy and the social studies of science also.

Finally, feminist philosophies of science draw on diverse disciplinary and theoretical resources. There is not now, nor will there ever be, a single feminist philosophy of science. Philosophical issues have been raised within the conceptual frameworks of the special sciences studied, disciplinary and theoretical frameworks of the social studies of science, and diverse political philosophies (see Feminism and social science; Feminist political philosophy). This entry will focus on five issues that surface repeatedly in these diverse approaches.

2 Bad science or standard science?

Are sexist and androcentric silences and distortions in the results of research in biology and the social sciences the consequence merely of bad science or, worse, also of standard science? The line between the empiricists and the post-empiricists, as we can refer to those who prefer one or the other of these positions, is not always hard and fast. Nevertheless, it is useful to characterize in such a way the two poles of the continuum on which feminists think about this question. Those who are satisfied with the first choice believe that there is nothing wrong with existing principles of good research in each disciplinary area, or with an empiricist logic of explanation. The problem is that researchers have not carefully or thoughtfully enough followed them. Basing generalizations about humans only on data about men, assuming that women's biology normally functions in immature or pathological ways, or that only males have evolved - such claims are examples only of bad science. Philosophers in this group often go on to criticize conventional assumptions that knowers are abstract individuals and to stress the importance of accountability to an epistemic community. Whatever their concerns, this group is marked by its vigorous intention to keep its projects firmly located within the empiricist tradition. For them, feminism raises new issues that enlarge the scope of research and reveal cause for greater care in its conduct and philosophic formulations, but leaves fundamentally unchallenged an empiricist logic of research and explanation.

A diverse group think gender problems in the natural sciences cannot adequately be understood or resolved by empiricist approaches: standpoint theorists such as Hilary Rose (1983), Dorothy Smith (1987) and Nancy Hartsock, standpoint theorists who also draw on post-structuralist resources; Donna Haraway (1991) and Sandra Harding (1986), and others like Evelyn Fox Keller who do not give their distinctive analyses a special name.

These feminist post-empiricists differ among themselves on the whole range of issues that divide the rest of post-empiricists; philosophically some feminist post-empiricists are more conservative and others more radical. All share the suspicion that there is something wrong with the standard ‘logics’ of both research and explanation. Androcentric values and interests are not just contingently present in the results of research; they also play a role in the cognitive, technical core of science that, in the Enlightenment vision, was held to be immune to all social fingerprints. If such values and interests can permeate the cognitive core of science - its ’method’ and collection of formal claims - then it is not sufficient simply to more carefully follow existing methods, theories and norms of science; instead these features must be transformed if the results of research are to escape distortion. Nor are remedies that rely on the critical effectiveness of a community of experts - an epistemic or scientific community - sufficient, since it is precisely such communities that have been unable to detect their own shared androcentric assumptions. There are problems with standard science that post-empiricists think require more radical kinds of resolution (see Scientific method §§2-3).

3 Whose problems?

Androcentric biases enter sciences at every stage of the research process, including the selection of problems. Just which array of problems scientists will pursue at any moment is far less due to scientists’ individual choices to ’pursue the truth wherever it may lead’ than it is to priorities of their sponsors and culture. However, the prevailing view has been that the selection of problems has no effect on the cognitive, technical core of science, that there is no logic of discovery, and thus that any attempt to regulate science prior to the context of justification is not only futile, but an infringement on sciences’ creative potential (see Discovery, logic of).

There is something odd about this protection of problem selection from responsibility for distortion in the results of research. The post-Kuhnian histories and ethnographies of science have revealed in great detail exactly how cultural assumptions and politically interested decisions made prior to the context of justification, the ones that constitute scientific projects and select methods to test hypotheses, preclude the possibility of the results of this research achieving cultural neutrality. If the problem is conceptualized as how to identify the proximate biological causes of cancer, one can be sure of a very different picture of nature and social relations in the results of that research than if the problem were conceptualized as how to identify the environmental carcinogens that are known to be the most powerful of the many causes of cancer. Focusing on proximate causes supports the interests and values of medical researchers, the tobacco, junk food and meat industries, and industrial and military polluters; focusing on environmental causes supports the interests and values of their critics and environmental researchers. Neither approach can be interest-free.

Feminists have argued that our collective knowledge about the natural and social worlds disproportionately represents the kinds of question of interest to the dominant social institutions - those interested in increasing profit, maximizing social control, legitimating hierarchical authority and the *status quo* and thereby advancing the values and interests of men in the advantaged groups, frequently at the expense of the values and interests of the poor, women, racial and ethnic minorities, and ‘have nots’ in other parts of the world. The patterns and contents of what sciences tend to discover about nature’s order disproportionately reflect what the dominant groups both have and have not wanted to know.

Faced with this problem, all feminists have recommended that scientific communities become more inclusive for scientific reasons as well as ones of social justice. Since the victims tend to be more sensitive to less obvious expressions of discrimination and devaluing of their concerns, their presence and critical analyses would be likely to improve the ability of the sciences to arrive at less partial and distorted representations of nature. However, such strategies are of dubious effectiveness by themselves; the concerns of marginalized people tend to remain marginalized when individuals from such groups are simply inserted into dominant and often resisting institutions that make no other changes.

Some feminists have proposed an additional strategy that could systematically help eliminate distortions in sciences’ representations of nature that are produced by androcentrism, Eurocentrism, or other values and interests that are shared by research communities (and, often, their cultures). They propose a strategy of conceptualizing scientific projects in the first place from outside the conceptual frameworks of dominant social institutions and research disciplines, namely from the standpoint of the lives of those who bear a disproportionate share of the costs

of how such institutions and disciplines function. Standpoint epistemologies thus direct researchers to start research from women's lives ('lives' plural, since 'women' is as heterogeneous a category as is 'humans') to generate scientific problems that are ones for women too, rather than only problems for the dominant institutions and the disciplines that work up knowledge for them. This 'method' is a collective and political process - a social and scientific achievement rather than an automatic outcome of women's experience or of living a woman's life. In a sense, they are proposing a kind of logic of discovery for more accurate, because interest-balanced, representations of nature and social relations.

4 Gender coding

What should one make of gendered metaphors and models of nature, the scientist and the research process that have been prevalent throughout the history of modern science? Historians and philosophers have pointed to the 'nature, she; scientist, he' language so familiar from Bacon and Machiavelli's writings to those of contemporary science texts. Susan Bordo (1987) Evelyn Fox Keller (1984) and Carolyn Merchant (1980) show how invoking images of the hyper-masculinity of science provided solace for the loss of the organicist 'mother-world' in early modern science. Such metaphors and models have helped to constitute the cognitive content of scientific claims, they and others argue.

The interactionist theory of metaphors and models holds that such forms of socially-meaningful language and concepts are a necessary part of every scientific theory. 'Nature is a machine' is not only heuristically or pedagogically useful; it also tells scientists how to develop their theories into new domains and where to revise when observations prove recalcitrant. As machines become more familiar, more natural, so too nature is experienced as more mechanical. And as the metaphor becomes more familiar, its regulatory functions shift from explicit plans into tacit knowledge. Similarly, conceptualizing nature as a wild and unruly woman who must be tamed if man is to be able to control his fate, as did Machiavelli, simultaneously directs scientists to favour certain kinds of interaction with nature and naturalizes certain ways of treating women. Enlightenment sciences have been directed by the illusion that there is a 'bare nature' at which scientists can gaze if they strip away the veils under which she hides her true features. But this metaphor distorts our understanding of how nature-as-an-object-of-knowledge is always already socially constructed, and yet also not agreeable to just any old fears, fantasies or desires that cultures choose to project on to its order. Humans speak in diverse metaphors, models and conceptual frameworks, but nature does answer back with as much (and only as much) precision as human interactions with it demand.

Such gender coding appears in ancient dichotomizing practices evident from Parmenides to the present day that construct self-other schemas coded masculine-feminine: rational-irrational, logic-emotions/passions, culture-nature, dynamic-static, mind-body, adult-infant, abstract-concrete, autonomous-dependent/relational/contextual, and so forth. Reliance on these symbolic systems can lead to oppressive social relations and distorted understandings of nature's order. After all, the dichotomies are empirically false of men and women's natures. Our understanding of nature can benefit from a broader array of metaphors and scientific processes than containment within such androcentric metaphors permits.

A persistent feminist concern has been with the modern coding of neutrality as masculine. If maximizing objectivity and rationality are conceptualized as requiring neutrality, and masculinity is defined partially in terms of its neutrality (even to gender, since 'man' and 'he' are taken to represent the generically human), then it is a kind of contradiction in terms to imagine that women can be objective and rational. This remains so even when their achievements are indistinguishable from men's. There is no speaking position available to women as rational, objective thinkers since that position has been fully occupied by the masculine. Moreover, those areas of science that are thought least shaped by social values and interests thus are the most vigorously masculine. Precisely because they are thought to be the most neutral, physics, mathematics and logic are the most masculine intellectual areas.

It should be stressed that these arguments about the role of gender metaphors and models in science are no more about individuals' idiosyncratic motivations than are arguments about the role of mechanistic or organicist metaphors. Androcentric individuals are not the issue: cultural assumptions, fears, desires, values and interests are. The metaphors and models used by Plato, Machiavelli, Bacon and Descartes are interesting because of what they reveal of the cultures within which these individuals constructed their texts for imagined audiences.

What should be done about these distorting conceptual frameworks? One widely shared proposal is to recover and revalue the feminine in nature, scientific practice, and the history and philosophy of science in order to enlarge sciences' conceptual resources. If hierarchical models of nature are preferred by cultures seeking to legitimate gender or other social hierarchies, then we lose the chance to understand those 'democratic' aspects of nature's order that are often coded feminine. If masculine models of research are preferred, we lose the chance to understand those aspects of nature's order that could be revealed by methods associated with the feminine. If reason must be divorced from the emotions and objectivity from cultural location, we lose the chance to understand the positive role that emotions and cultures can play in the growth of knowledge.

Another proposal is to seek metaphors and models that do not emphasize the importance of gender difference, and point out the limitations of those that do. We would thus have the chance to grasp heretofore less visible aspects of nature's order as well as undermine the oppressive rule of hegemonic gender difference.

5 Technology

Like the science discussions, feminist approaches to technology have drawn upon recent constructivist tendencies to show how artefacts have politics: technological change is a site for gender struggles over the meanings of and access to technological literacy, the use of new technologies and control over their consequences (see Constructivism). Judy Wajcman's (1988) important analysis focuses on such issues as they occur in the organization of work, reproduction, the built environment, militarism, and computer culture. Research technologies, too, are gendered. Since modern science is often distinguished from other more speculative science traditions precisely by its reliance on experimental methods' technologies of intervention in nature, these studies show yet another way in which the social consequences of sciences are a direct result of features of their cognitive core. If one learns about genes or extra-uterine forms of reproduction by manipulating them, one creates blueprints for how to change them. Technological uses of sciences are partially directed by successful research methods.

These accounts are also valuable because they provide rich resources for understanding how Western claims to scientificity are used to impose technologically oppressive social relations on women in developing countries. Here too, feminist standpoint approaches reveal why philosophies and other social studies of technology should start off their projects from the lives of those who have suffered from Western sciences and technologies, in order to gain a less partial and distorted understanding of their strengths and weaknesses (see Technology and ethics).

6 Objectivity and method

Feminist reflections on the goal of objectivity, the neutrality ideal and methods to achieve these are part of a more general discomfort with the West's assumed authority about the nature of reality and desirable forms of social relations. Five analyses will suggest the range of proposals here.

The first of these was referred to in §4. If objectivity requires neutrality, and neutrality is coded masculine, how can a woman - or a feminism that speaks on behalf of women - ever claim objectivity for women's speech? And how can sciences' claims to objectivity avoid continuing to overvalue manliness? Many feminists have found these objections compelling and have recommended adopting some other language -for example, subjectivism or relativism - for discussing the reliability of our understandings of nature and social relations. However, others have sought to clarify or transform the concept.

Helen Longino (1990) sets out to show how even though science is socially constructed, it already has good procedures for achieving objectivity. The relevance of evidence to a theory is determined by the context of inquiry, by background beliefs. All too often, social or practical interests (contextual values), including gendered ones, in background beliefs have functioned as the cognitive values (constitutive ones) that determine what counts as a good scientific judgment. However, she argues that we can identify these background beliefs independently, and thus show how certain states of affairs have been rightly or wrongly taken as evidence for a hypothesis. Such processes are social and thus public, for they depend upon a scientific community sharing a common language with which to describe experience and intersubjective agreement that screens out idiosyncratic and subjective elements. Most importantly, since the scientific community rather than the individual scientist is the knower, enhanced attention to the role of criticism is crucial for overcoming power imbalances within such a community that can result in the devaluation of legitimate criticisms. Objectivity requires a transformative criticism such that

the community adjusts its assumptions and processes in light of criticism by all of its qualified members.

Evelyn Fox Keller draws on psychoanalytic theory to argue that objectivity is best reconceptualized as profiting from the use of subjective experience rather than as opposed to it. This ‘dynamic objectivity’ contrasts with the prevailing philosophical preference for ‘static objectivity’ through which the knower must take a basically adversarial relation to the objects of study. Such a science, with its rhetoric of domination and coercion, attracts individuals who find emotional and cognitive comfort in such ways of relating to nature and other people to the exclusion of other, at least equally revealing styles, methods and theories of such relations. ‘Good science’ gets defined in an excessively narrow way. Dynamic objectivity ‘ grants to the world around us its independent integrity but does so in a way that remains cognizant of, indeed relies on, our connectivity with that world… dynamic objectivity is not unlike empathy’ (Keller 1984). Dynamic objectivity’s long history can be detected in the thought of the many scientists who have seen their work primarily as an erotic rather than adversarial activity.

Sandra Harding has argued (1991) that the neutrality ideal can usefully be detached from the goal of objectivity; methods for maximizing ’strong objectivity’ can be substituted for it. Androcentrism, Eurocentrism and the like have been shared by virtually entire research communities; they are not the kind of subjective, idiosyncratic values that the research methods activated in the context of justification are capable of detecting. The neutrality ideal is an obstacle to detecting the conceptual practices of power since it categorizes as political only the critics of power, not those who claim the ‘view from nowhere’ within its protective normalizing assumptions and procedures. Standpoint epistemologies, that direct one to formulate questions in the first place from outside the dominant institutions, practices and culture, and to start instead from disadvantaged lives, can provide systematic resources - a ‘strong method’ - for philosophies of the natural and social sciences. Abjuring both epistemic foundationalism and relativism, and careful neither fully to identify with nor to appropriate the voice of the other, standpoint theories insist that only socially situated knowledge, only partial perspectives, are possible. Those issuing from disadvantaged social locations can provide less partial and distorted understandings of nature and social relations.

Donna Haraway’s rethinking of objectivity occurs in the context of her incisive questions about the relation between the production of science (‘technoscience’) and of culture. She develops the ethical and political implications of standpoint theory further, stressing the need for a notion of objectivity that initiates, instead of closing off, issues of responsibility for science’s images of nature and the rational knower. She argues that the objects of knowledge, too, must be reconceptualized:

> A corollary of the insistence that ethics and politics covertly or overtly provide the bases for objectivity in the sciences as a heterogeneous whole, and not just in the social sciences, is granting the status of agent/actor to the ‘objects’ of the world…. Accounts of a real world do not, then, depend on a logic of ‘discovery’, but on a power-charged social relation of ‘conversation’.
>
> (Haraway 1991)

Such an agent could be pictured as the Coyote or Trickster, she proposes, embodied in American southwest Indian accounts, for this figure ‘ suggests our situation when we give up mastery but keep searching for fidelity, knowing all the while we will be hoodwinked’ by nature.

In conclusion, it is striking how little attention has been paid to the philosophic effects of gender relations by even the most radical of mainstream philosophies and social studies of science. In effect they exclude such relations from the realm of the social that otherwise is a major concern. Feminist accounts are more objective in this respect, as well as in their attention to the conceptual effects of race/imperial relations and more rigorous understanding of self-reflexivity issues.

In their attention to relations between power and knowledge, feminist science analyses both reflect and promote the widespread critical re-examination of Enlightenment ideals and their political consequences. Their diverse attempts to envisage post-Enlightenment sciences and philosophies of science join widespread global projects to bring Western sciences under democratic control. And yet the feminist projects also are fully inside Western traditions, for they emerge from and advance one long-standing project there of criticizing the favoured beliefs of the powerful in order to block ‘might makes right’ in the domain of knowledge production.

See also: Feminism; Feminist epistemology; Objectivity

SANDRA G. HARDING

References and further reading

Bordo, S.R. (1987) *The Flight to Objectivity: Essays on Cartesianism and Culture*, Albany, NY: State University of New York Press.(Important analysis of issues in §4.)

Haraway, D. (1989) *Primate Visions: Gender, Race and Nature in the World of Modern Science*, New York: Routledge.(Primatology's role in the production of androcentric and Eurocentric culture, and vice versa. Early classic accounts of how gender and race shaped the choice and conceptual frameworks of scientific projects.)

Haraway, D. (1991) 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective', in *Simians, Cyborgs and Women: The Reinvention of Nature*, New York: Routledge.(Widely cited analysis of the non-absolutist and non-relativistic epistemology emerging from philosophies of science. Originated as a review of Harding 1986.)

Harding, S. (1986) *The Science Question in Feminism*, Ithaca, NY: Cornell University Press.(Expansion of the material of §§1 and 2. Mapped the feminist empiricist and standpoint philosophies of science that were being developed by practising scientists and science theorists in philosophy and the social sciences, and the challenges to them emerging from feminist poststructuralism.)

Harding, S. (1991) *Whose Science? Whose Knowledge?*, Ithaca, NY: Cornell University Press.(Expansion of the material of §§3 and 6. Develops programme for 'strong objectivity' to replace conventional neutrality requirement.)

Harding, S. (1998) *Is Science Multicultural? Postcolonialism, Feminism and Epistemology*, Bloomington, IN: Indiana University Press.(Shows how feminist philosophies of science and two decades of multicultural and postcolonial science studies can inform each other.)

Harding, S. and Hintikka, M. (eds) (1983) *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology and Philosophy of Science*, Dordrecht: Reidel.(Earliest collection of philosophically focused papers, including standpoint theory essay by Nancy Hartsock.)

Keller, E.F. (1983) 'Feminism as an Analytic Tool for the Study of Science', *Academe* September-October: 15-21. (A concise discussion of themes from her 1984 publication.)

Keller, E.F. (1984) *Reflections on Gender and Science*, New Haven, CT: Yale University Press.(Important arguments for §§4, 6. Classic early analyses of gender and method, objectivity, Plato's epistemology, Baconian science, contemporary physics and biology.)

Longino, H. (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*, Princeton, NJ: Princeton University Press.(Analysis of issues in §6. How gendered meanings infuse not only modern sciences' background beliefs, but also their constituting values.)

Merchant, C. (1980) *The Death of Nature: Women, Ecology and the Scientific Revolution*, New York: Harper & Row.(Important history of early modern science; arguments for §4. How modern sciences came to reflect and help to shape shifting gender relations of fifteenth- through seventeenth-century Europe.)

Nelson, L.H. and Nelson, J. (eds) (1996) *Feminism, Science, and the Philosophy of Science*, Dordrecht: Kluwer. (Thirteen philosophers - men and women - discuss a range of issues that feminist scholars have raised in philosophy of science. Diverse topics and philosophical approaches.)

Rose, H. (1983) 'Hand, Brain and Heart: A Feminist Epistemology for the Natural Sciences', *Signs: Journal of Women in Culture and Society* 9: 1.(Important development of standpoint theory; argument of §3.)

Shiva, V. (1989) *Staying Alive: Women, Ecology and Development*, Atlantic Highlands, NJ: Zed Books.(Critique of Western science, its patriarchal conceptions of nature, rationality and objectivity by an Indian physicist and environmental activist.)

Smith, D.E. (1987) *The Everyday World as Problematic: A Feminist Sociology*, Boston, MA: Northeastern University Press.(Important standpoint theory essays by the most eminent feminist philosopher-sociologist writing from the perspective of concerns about research methodologies.)

Traweek, S. (1988) *Beamtimes and Lifetimes: The World of High Energy Physics*, Cambridge, MA: Harvard University Press.(Ethnography of physics, including its gender relations.)

Wajcman, J. (1991) *Feminism Confronts Technology*, University Park, PA: Pennsylvania State University Press. (Analysis of issues in §5. A pathbreaking constructivist analysis of technology that enables development of a distinctive and influential philosophy of technology.)

Wylie, A., Okruhlik, K., Morton, S. and Thielen-Wilson, L. (1990) 'Philosophical Feminism: A Bibliographic Guide to Critiques of Science', *RFR/DRF New Feminist Research* 19 (2): 2-36.(The most comprehensive guide to the first two decades of the literature.)

Genealogy

'Genealogy' is an expression that has come into currency since the 1970s, a result of Michel Foucault's works Surveiller et punir (1975) (Discipline and Punish, 1977) and The History of Sexuality (1976, 1984a, 1984b). Foucault's use of the term continues Nietzsche's in his On the Genealogy of Morals ([1887] 1967). For both philosophers, genealogy is a form of historical critique, designed to overturn our norms by revealing their origins. Whereas Nietzsche's method relies on psychological explanations, and attacks modern conceptions of equality in favour of a perfectionist ethic, Foucault's relies on micro-sociological explanations, and attacks modern forms of domination in favour of radical politics.

1 Genealogy as philosophical method

While some claim that genealogy is a new method for investigating history, others have maintained that genealogy is simply historical investigation itself, and nothing new. Neither of these claims captures what is distinctive about Nietzsche's and Foucault's enterprises. Rather, genealogy is a new method of applying historical investigation to philosophical concerns.

In Book I of the *Republic*, Plato portrays Socrates as seeking substantive philosophical knowledge of justice by discussing proposed definitions of 'justice'. The *elenchus* is the method for testing proposed definitions. In the second essay of *On the Genealogy of Morals* ([1887] 1967, II: 13) Nietzsche claims that 'only that which has no history is definable'. If the expressions which concern us have a history, then Socrates' project fails. We can better clarify the sense behind such expressions as 'justice' by examining the history of the practices that have shaped our disposition to classify some acts as just. According to Plato, we apprehend an unchanging, independently real *eidos* of justice, and it is our capacity to intuit the presence of justice in an act that accounts for our linguistic dispositions. Consequently, the history of our linguistic dispositions can only be the history of our clear vision of what is always already there; such a history cannot possess much philosophical interest. On the genealogist's view, however, there is no such *eidos*. Therefore, it cannot be responsible for our semantic intuitions or our consensus about them. For the genealogist, history itself moulds and shapes our practices and intuitions over time; in an account of what moulds the history, we will find nothing but competing and cooperating forces, interpretations, interests. The genealogist's interest is thus directed at normative practices whose dominance requires that their participants not understand how the practice came to be.

As Nietzsche and Foucault view it, genealogy supersedes earlier forms of philosophy. The sweeping character of this methodological claim rests on the genealogist's confidence that traditional philosophical strategies fail on their own terms. (This point is crucial for escaping the common complaint that the genealogist commits the genetic fallacy.) Traditional strategies falsely presuppose that our intuitions are reliable. If, for example, our moral intuitions were broadly unreliable, traditional moral theory would collapse, given the moral theorist's reliance on these intuitions for data by which to judge the moral theories under discussion. If our conceptual or semantic intuitions are unreliable, similar considerations will apply to conceptual or linguistic analysis.

History, according to genealogists, is not teleological (as it is for Hegel). They cannot identify a *goal* of a historical process, and then go on to show how it gradually emerged from its embryonic beginnings. Rather, they chart the processes that, by contingent confluence, produce a contemporary result. Hence the metaphor: no individual is the goal of a family history. Rather, a family is a vast fabric of relationships, and any one individual represents only one among many confluences of past lines of descent.

To avoid the temptation of projecting current norms into the past, which would promote a teleological picture, genealogy adopts the German historicist view that each cross-section of history has its own autonomous structures, meanings, values, etc. which the historian must work to recover hermeneutically (see Historicism; Hermeneutics). However, unlike the hermeneutic historian, the genealogist allows that the alien culture's self-understanding is not the end, but the beginning of investigation. Since the function of genealogy is to undermine our own hermeneutical self-understanding, we cannot grant self-understanding final authority elsewhere either.

What determines the content of such a self-understanding is, ultimately, multiple conflicting interests and interested interpretations. Genealogy characterizes the underlying historical processes as a 'power struggle'. Just

as an account of a war will differ depending upon which combatant pens it, aspects of these historical processes will appear differently to different viewers. Attention to this plurality of 'perspectives' leads the genealogist to claim that there is no absolute fact as to the significance or value of a particular historical event or process. If we add the claim that experienced fact is largely constituted by the interpretations of those experiencing it, a question arises: does it make sense to speak of historical fact at all? This feature of genealogy raises familiar epistemological problems, and sits uncomfortably with its intended critical function. However that may be, the genealogist does *not* claim that one perspective is as good as another, either normatively or cognitively. Genealogy serves a critical function, undermining the myths and mystifications a particular contemporary perspective may have about itself. To the extent that the perspective depends upon mystification, participants' understanding of the processes that constructed the perspective will lead to its disintegration.

The critical nature of genealogy can be understood as follows. The genealogist is concerned to undermine ahistorical and inflationary interpretations of mundane facts about human life. These interpretations typically appeal to structures, norms, entities, etc. which stand above both nature and history. The genealogist is also apt to undermine the propensity to regard what is artificial (and therefore mutable) as naturally, rationally or metaphysically necessary (and therefore immutable). Norms may appear to possess a greater degree of legitimacy if they are associated with such necessities as opposed to artifice. If norms are artificial, produced by struggle and veiled by mystification, they will be an expression of someone's pragmatic interests. Inflationary interpretations conceal these interests and the ways that the practice in question promotes them.

Finally, genealogists must accept that they too are engaged in a practice (doing genealogy) which has a history, which expresses various pragmatic interests, which is perspectival, etc. Whether genealogy's reflexivity hoists the genealogists by their own petard has been much discussed since the 1970s.

Many broad themes of contemporary genealogical inquiry have resonances in earlier European intellectual history. The notion of the artificiality of norms, and their emergence from conflict is central to social contract theory since Hobbes (see Contractarianism). However, modern dissenters from social contract theory laid the groundwork for Nietzsche's novel interpretation of morality. Though himself a social contract theorist, in the *Discourse on Inequality* ([1755] 1988) Rousseau transforms the model *state of nature-contract-civil society* into an anthropological and historical, if speculative, narrative. Rousseau outlines how we might have first become social. Later, through mystified economic processes, we come to consent to political institutions which benefit the wealthy at the expense of the poor. By emphasizing the historical and artificial character of this process, Rousseau rejects the social contract arguments defending the economic and political institutions of his day.

2 Nietzsche

The image of Rousseau haunted Nietzsche throughout his career. Both the similar styles of argumentation, and the aristocratic values Nietzsche champions make 'A Discourse on Equality' an apt alternate title for his *Genealogy*. Nietzsche's thoughts on the development of morality can be seen in several texts prior to the *Genealogy*. In 'On the Utility and Liability of History for Life' ([1873b] 1995), Nietzsche deplores the rise of 'disinterested' historical research, which has replaced more edifying and useful portrayals of the past with pointless accumulations of fact. Creative living requires that we often forget the past. If we are to benefit from the study of the past, we must practise history with one or more of three aims in mind. 'Antiquarian history' seeks to revive the lost cultural treasure house of the past. 'Monumental history' seeks to memorialize ethically exceptional individuals and deeds, which we might wish to emulate. 'Critical history' uncovers the evils of the past that we might better renounce them in the present. All three forms appear woven together in several later works: 'On Truth and Lie in an Extra-moral Sense' ([1873a] 1990), 'On the History of the Moral Sensations' in the first volume of *Human, All Too Human* ([1878] 1986), 'Natural History of Morality' in *Beyond Good and Evil* ([1886] 1966) and the *Genealogy* itself.

In the first essay of the *Genealogy*, Nietzsche argues that 'English psychologists' err in regarding the original sense of 'good' as 'useful'. The sense of 'good' varied, depending upon which class perceived it. For the dominant class, or 'masters', the primary sense of what is good is the self and that which resembles the self. By contrast, the socially subordinated, the 'slaves' behave in very different ways, which the master designates as 'bad', ignoble or base. However, from the slave perspective, what the master calls 'good' causes the slave's suffering and subordination. Thus, he designates the master as 'evil' or wicked. The slave then invents the fiction of free will.

The master is free to not behave in a wicked way, and the slave is equally free to begin to do so. The slave then constructs a moral scheme according to which his own passivity makes him morally superior to the master and therefore 'good'. Here we see many characteristic themes of genealogy sounded: pragmatic interests, perspectivality, demystification. The relevance of Nietzsche's critique of Christianity is clear. Beyond that, he broadens his analysis by claiming that modern liberal institutions, norms and intuitions are so much secularized Christianity. Thus liberalism inherits much of Christianity's unappetizing 'slavishness' without the religious rationale which legitimated it.

In the second essay of the *Genealogy*, Nietzsche traces the psychological origin of guilt to the primitive experience of debt (at a time when enforcement of debt was more cruel than now). Anticipating Freud, he speculates that the inner mechanism mediating this experience is the product of introjected aggression. While we cannot dismantle the conscience, we can 'reprogramme' it with values and standards other than those derived from slave morality. In the third essay, Nietzsche asks why we regard asceticism so highly. He reveals its psychological roots in the sour grapes of the ascetic. Its persistence is due to the socially stabilizing function of the ascetic's ideology of sin and contrition. Furthermore, modern science is an outgrowth of the ascetic ideal, its self-hating contempt for pleasurable illusion expressing itself as a commitment to truth. These essays stand as early paradigms of genealogical critique.

3 Foucault

Foucault turned towards Nietzsche and genealogy late in his career. In his earlier, 'archaeological' works, Foucault analysed medicine, the human sciences (for example, economics, linguistics, biology) and philosophy. According to Foucault, structured codes organize scientists' perception, discourse and social practices at a level unknown to them. Such codes, or 'epistemes', play a quasi-Kantian role, as 'conditions of the possibility' of knowledge. Foucault's conception of episteme thus makes the work of scientists subject to a cultural force beyond anyone's conscious control. His conception of these epistemes and their historical succession is indebted to the later Heidegger's conception of a 'history of being' as well as to the work of the structuralists (for example, Althusser, Lacan, Lévi-Strauss) (see Heidegger, M.; Structuralism). His later turn to genealogy would inherit both the culturalism, and the epistemological interests, of his earlier work.

Foucault came to believe that he was reifying unduly in claiming that epistemes, which are abstract, governed (at least some domains of) social life. Instead, he substituted the notion of regimes of power, which coerce people into following specific norms. The mechanisms which maintain this coercion are micro-practices demanding specific kinds of behaviour in specific contexts (for example, the use of time-cards and clocks in factories). Although these micro-practices in effect subordinate some people to the benefit of others, they need not be the products of any deliberate design.

Foucault's analysis of society often seems to jump from detailed concrete descriptions of micro-practices, to more impressionistic claims about the common style of the micro-practices. This common style betrays a pervasive, invisible, impersonal force: power. Power structures behaviour; at the same time, it structures the perception, discourse and practices of scientists who are concerned to understand human behaviour scientifically. Foucault's account of the relationship between power and knowledge echoes Nietzsche's in *Genealogy* Essay III. However, where Nietzsche's analysis of the will to truth depends on *psychological* explanations, in Foucault's, micro-sociology and a broad cultural thematics replace psychology.

Foucault's genealogical method, which he sketches in 'Nietzsche, genealogie, histoire' (1971), does not merely echo Nietzsche's concern with power. Both philosophers' historical inquiries are directed at practices whose character bears on philosophical concerns (for example, the nature of personal and political freedom, the value of self-knowledge). Furthermore, Foucault's analyses, like Nietzsche's, portray historical processes with an emphasis on their contingent, non-teleological character. He is sensitive to how alien even something as recent as the Enlightenment can appear, if viewed through historicist lenses. However, what links him most decisively with Nietzsche is his commitment to demystifying accounts of the historical processes that most concern him. We easily ignore our 'disciplinary' practices in their institutional home, the prison, because our ideological commitments interpret such practices as just. Once we see the practices described in coolly non-ideological terms and in concrete detail, we clearly perceive their outrageousness. What is more, the presence of similar 'carceral' practices in other social domains can then be more clearly perceived. Foucault intends his readers to transpose their outrage from the

one setting to the other. His work thus shares Nietzsche's critical intent: Nietzsche wishes to use his critique as a tool to re-animate 'noble' values, whereas Foucault places his critique in the service of radical politics.

Foucault had intended to continue his genealogy of modernity. He meant to supplement his account in *Discipline and Punish* (1975) of the growing 'carceralization' of modernity with a parallel account of its growing 'sexualization' in the projected multi-volume *History of Sexuality*. However, after one promising introductory volume, Foucault turned his attention away from the historical processes that culminated in contemporary society. Instead, he turned toward the genealogy of Western ethical ideals, and their relationship to sexual self-mastery. This project begins with an analysis of Greek ethics in *The Use of Pleasure* (1984a), and continues with an account of Roman ethics in *The Care of the Self* (1984b). A final volume, which was to deal with medieval ethics and the 'hermeneutics of the self' has not yet appeared. In these final works, Foucault relegates his political passion, so reminiscent of Rousseau, to the background. Instead, we see the beginnings of work much closer in subject and spirit to Nietzsche's original concerns.

See also: Post-structuralism in the social sciences §3

R. KEVIN HILL

References and further reading

Dreyfus, H.L. and Rabinow, P. (1982) *Michel Foucault: Beyond Structuralism and Hermeneutics*, Chicago, IL: University of Chicago Press.(An accessible introduction to Foucault's thought, with extensive discussion of the methodological issues raised by genealogy.)

Foucault, M. (1971) 'Nietzsche, généalogie, histoire', in *Hommage à Jean Hyppolite*, Paris: Presses Universitaires de France; trans. D. Bouchard and S. Simon, 'Nietzsche, Genealogy, History', in *Language, Counter-memory, Practice*, ed. D. Bouchard, Ithaca, NY: Cornell University Press, 1977; reprinted in P. Rabinow (ed.) *The Foucault Reader*, New York: Pantheon Books, 1984.(This latter volume also includes extensive selections from *Discipline and Punish* and *History of Sexuality*, vols 1 and 2, and makes a good starting point for reading Foucault.)

Foucault, M. (1975) *Surveiller et punir: naissance de la prison*, Paris: Gallimard; trans. A. Sheridan, *Discipline and Punish: The Birth of the Prison*, New York: Pantheon Books, 1977.(Foucault's first genealogical exercise, focusing on the relationships between the history of punitive practices, criminology and the formation of the modern subject.)

Foucault, M. (1976) *La Volonté de savoir (The Will to Knowledge)*, trans. R. Hurley, *The History of Sexuality*, vol. 1, *An Introduction*, New York: Pantheon Books, 1978.(Foucault's second genealogical exercise, never completed, was to trace connections between sexology, psychiatry, the 'sexualization' of modern culture and the emergence of the hermeneutical, self-interpreting subject.)

Foucault, M. (1984a) *L'Usage des plaisirs*, Paris: Gallimard; trans. R. Hurley, *The History of Sexuality*, vol. 2, *The Use of Pleasure*, New York: Pantheon Books, 1985.(The first of two studies in Foucault's self-described 'genealogy of ethics' proposes that ancient Greek ethics was inseparable from personal projects of sexual self-mastery.)

Foucault, M. (1984b) *Le Souci de soi*, Paris: Gallimard; trans. R. Hurley, *The History of Sexuality*, vol. 3, *The Care of the Self*, New York: Pantheon Books, 1986.(In the second of two studies in ancient ethics. Foucault turns his attention to Rome.)

Nietzsche, F. (1873a) 'On Truth and Lie in an Extra-moral Sense', trans. D. Breazeale, in *Philosophy and Truth*, Atlantic Highlands, NJ and London: Humanities Press International, 1990.(This early text of Nietzsche's critically discusses the development of the 'drive for truth' in a manner that anticipates his later genealogies. It has been very influential in both philosophical and literary theory circles.)

Nietzsche, F. (1873b) 'Vom Nutzen und Nachtheil der Historie für das Leben', Leipzig: Fritzsch; trans. R.T. Gray, 'On the Utility and Liability of History for Life', in *The Complete Works of Friedrich Nietzsche*, vol. 2, ed. E. Behler, Stanford, CA: 1995. (Vol. 2 is the only volume of this twenty-volume 'Stanford Edition' to have appeared at the time of writing. This edition is based on the Colli-Montinari *Kritische Studienausgabe*, a large subset of the *Kritische Gesamteausgabe Werke*, which is regarded as the definitive German edition. 'On Truth and Lie in an Extra-moral Sense' will appear in vol. 1; *Human, All Too Human* will appear in vols 3 and 4; *Beyond Good and Evil* and *On the Genealogy of Morals* will appear in vol. 8.)

Nietzsche, F. (1878) 'Zur Geschichte der moralischen Empfindungen', in *Menschliches, Allzumenschliches*, Chemnitz: E Schmeitzner; trans. R.J. Hollingdale, 'On the History of the Moral Sensations', in *Human, All Too Human*, Cambridge: Cambridge University Press, 1986.(These aphorisms represent Nietzsche's first attempt at

a genealogy of morals.)

Nietzsche, F. (1886) ‘Zur Naturgeschichte der Moral’, in *Jenseits von Gut und Böse*, Chemnitz: E. Schmeitzner; trans. W. Kaufmann, ‘Natural History of Morals’, in *Beyond Good and Evil*, New York: Random House, 1966. (A precis of the late Nietzsche’s genealogy of morals.)

Nietzsche, F. (1887) *Zur Genealogie der Moral*, Chemnitz: E. Schmeitzner; trans. W. Kaufmann and R.J. Hollingdale, *On the Genealogy of Morals*, New York: Random House, 1967.(The classic text which introduced the concept of ‘genealogy’ into philosophical discussion.)

Plato (*c*.380s-370s BC) *Republic*, trans. G.M.A. Grube, revised C.D.C. Reeve, Indianapolis, IN: Hackett Publishing Company, 1992.(Reeve’s introduction is especially relevant.)

Rousseau, J.-J. (1755) *Discours sur l’origine de l’inégalité*, in *Oeuvres complètes*, vol. 3, Paris: Gallimard, 1975; trans. J. Conaway Bondanella, *Discourse on Inequality*, in *Rousseau’s Political Writings*, ed. A. Ritter and J. Conaway Bondanella, New York and London: W.W. Norton, 1988.(Rousseau’s critique of inegalitarian attitudes, practices and institutions rests on demystifying historical narrative in a fashion which anticipates Nietzsche, despite the latter’s opposing normative commitments.)

Rousseau, J.-J. (1762) *Contrat social*, in *Oeuvres complètes*, vol. 3, Paris: Gallimard, 1975; trans. J. Conaway Bonadella, *On the Social Contract*, in *Rousseau’s Political Writings*, ed. A. Ritter and J. Conaway Bondanella, New York and London: W.W. Norton, 1988.(Rousseau’s contribution to liberal political theory also emphasized the artificiality of the social.)

Schacht, R. (1983) *Nietzsche*, London and New York: Routledge, 2nd edn, 1992. (Systematic, thorough, moderately difficult. Issues associated with genealogy can be found throughout, but see especially chapters 2, 5, 6 and 7.)

Schacht, R. (ed.) (1994) *Nietzsche, Genealogy, Morality: Essays on Nietzsche’s Genealogy of Morals*, Berkeley, CA: University of California Press.(This collection is an ideal starting point for further reading about Nietzsche’s *Genealogy*.)

General relativity, philosophical responses to

Much of the early philosophical attention given Einstein's theory of gravitation was not uncontaminated by a grim post-war atmosphere conducive to public diversions, hysteria and national chauvinism. Most was ill-informed regarding the mathematical and physical content of the theory. Even amongst the scientifically literate, there was disagreement as to the theory's philosophical implications. In part, this lack of clarity is due to Einstein. In an endeavour to eliminate references to 'absolute space' as the earlier special (or, as it was then known, restricted) theory of relativity (STR) had eliminated reference to 'absolute time', Einstein had motivated his theory of gravitation as arising from an epistemologically mandated generalization of the relativity principle of STR, which governed only inertial motions, and he misleadingly baptized it a theory of 'general relativity', wherein all motions are regarded as relative to other bodies. This the theory does not show. Also, some incautiously expressed remarks on the formal requirement of general covariance were seized upon as evidence for antithetical epistemological and ontological attitudes. Amidst such confusions, it is not at all surprising that inherently antagonistic philosophical outlooks claimed vindication or confirmation by the general theory of relativity (GTR). In turn, the perceived failure of both Machian positivism and Neo-Kantianism to accommodate the revolutionary theory spurred the development of a new 'scientific philosophy', logical positivism. The subsequent course of philosophy of science in the twentieth century was indelibly marked by this development. Yet it would turn out that Einstein himself refused to be a cooperative exemplar for any of the major philosophical schools, positivism, Kantianism, or, to its embarrassment, logical positivism.

1 Positivism

Positivists generally welcomed the special theory of relativity (STR), viewing Einstein's operational analysis of the 'simultaneity' of separated events as a vindication of Mach's account of concepts as mere tools, economical shorthand for operations exhibiting the dependence of phenomena on one another (see Logical positivism §§2, 4; Mach, E.; Operationalism). Some of Einstein's epistemological remarks in the context of STR appear to underscore this alignment: for example, ' The concept does not exist for the physicist until he has the possibility of discovering whether or not it is fulfilled in an actual case' (Einstein 1917: 22). Such pronouncements served to cast Einstein's philosophical orientation in a positivist light and, naturally enough, promoted tendencies both to fashion similar interpretations of the general theory of relativity (GTR) and, in view of ensuing difficulties, to criticize GTR as a violation of Einstein's own methodological precepts (Bridgman 1949).

A single passage from section three of Einstein's 1916 exposition of GTR provided further grist for the mill of Machian positivism. There, in the course of a heuristic 'reflection' on the reason for the necessity of general covariance of the field equations of gravitation (that they remain unchanged under arbitrary transformation of the coordinates), Einstein proclaimed that this requirement, 'takes away from space and time the last remnant of physical objectivity' since 'all our space-time verifications invariably amount to a determination of space-time coincidences' and that consequently, 'all our physical experience can ultimately be reduced to [point-event] coincidences'. Following the interpretation given these remarks by the Machian physicist Philipp Frank, the philosopher Joseph Petzoldt stated that 'the theory of relativity rests, in the end, on the perception of the coincidence of sensations' and, in particular, that GTR 'is fully in accord with Mach's world-view, which is best characterized as relativistic positivism', his designation for Mach's philosophical direction (see Petzodt 1921: 516). Despite the obvious embarrassment of Mach's apparent disavowal of the theory of relativity in the 'Preface', dated 1913, to his posthumously published (1921) book on physical optics (Mach died in 1916), Einstein's illustrative reflection seemed an incontrovertible avowal of the centrepiece of Mach's phenomenalist epistemology with its attendant ontological parsimony.

Contemporary scholarship has shown that these remarks of Einstein were but elliptical references to an argument for a different and considerably more intricate conclusion, that the points of the spacetime manifold have no inherent individuality (inherited say, from the underlying topology) (see Spacetime), hence they have no reality independently of the presence of physical fields defined on the manifold (Stachel 1989). And while Einstein made no secret of his ambition in the context of GTR (in recognition of a 'general principle of relativity') to fully implement what he termed 'Mach's programme' against Newtonian 'absolute motions', he had already distanced himself from Mach's phenomenalist epistemology. As for the infamous 'Preface', a strong case has recently been

made that it was forged after Mach's death by his son Ludwig, under the influence of a rival guardian of Mach's legacy, Hugo Dingler.

2 Neo-Kantianism

Among the scientifically minded philosophical community of Wilhelmian and early Weimar Germany, the philosophy of Kant held pride of place, with that of Mach a distant second. Only a few Kantians adopted strategies to preserve Kantian orthodoxy intact in the face of empirical success of GTR, for example, claiming that Kant's views on space and time pertained solely to a psychological 'intuitive space' and were not touched by Einstein's empirical theory of spacetime (Sellien 1919). Others sought vindication on grounds which downplayed the doctrine of space and time, for example, holding that Einstein's finite but unbounded universe supported the dialectical argument of the first antinomy of pure reason (Schneider 1921). But since, according to GTR, the behaviour of light rays and measuring rods in gravitational fields no longer corresponded to the relations of Euclidean geometry, knowledgeable Kantians concluded that the necessarily Euclidean character of space claimed in the transcendental aesthetic (TA) was untenable (see Kant, I. §5). At least, that is, in finite regions, for it could be maintained that those (essentially gravitation-free) infinitesimal regions where the rigorous validity of STR was upheld within GTR did conform to Kantian a priori conditions of experience (Cassirer 1921: 436-7; Bollert 1921: 61). As a matter of mathematical fact, the metric in the immediate neighbourhood of any point in a Riemannian manifold (such as is employed in Einstein's theory of gravitation) is 'locally Euclidean' (in spacetime, pseudo-Euclidean, for example a Minkowski metric); in such regions the spacetime manifold is homogeneous, isotropic and flat. But it is not generally possible to extend this metrical characterization to cover the entire spacetime manifold since the generic spacetimes allowed in GTR are inhomogeneous and variably curved (see Spacetime §3). And these latter do fall outside the supposed necessity of conditions of possible experience.

Characteristically, the deliverances of the new physics were accepted as fact while resources broadly within Kantian orthodoxy were sought for updating 'critical idealism'. Bollert (1921, cited by Gödel 1946/9) 'clarifies' the latter as concerned with the identification of progressively deeper 'levels of objectivation' underlying physics; in this regard, GTR, with its variably curved spacetimes, is a further step below the space and time of both Newton and STR. More influential was Ernst Cassirer's interpretation of Einstein's comment that space and time have lost 'the last remnants of physical objectivity'. According to Cassirer, GTR confirms the pivotal Kantian lesson of the ideality of space and time, that space and time are not 'things' but only 'forms of appearance' or, better, 'ideal principles of order'; each is a schema applying to the phenomena of the external world as a necessary condition of their possible experience (see Space §3). GTR has thus given explicit formulation to what Kant tried, with partial success, to articulate in the doctrine of 'pure intuition', namely, space and time are methodological presuppositions of empirical knowledge in physics. What remains of 'pure intuition' are 'the serial form of the relations of coexistence and succession', a structure consisting of (broadly speaking) topological relations which is far more impoverished than a globally Euclidean metric. In the light of GTR, Cassirer recast Kant's doctrine of space and time as essentially a version of Leibnizian relationism (see Leibniz, G.W. §11).

A more radical revision of Kant is given in Winternitz (1924) for whom the essential Kantian conception is just the recognition that the world is *nicht gegeben, sondern aufgegeben* (not given but posed) to us as a task (see Kant, I., *Critique of Pure Reason*, A498/B526). Regarding the TA as undeniably asserting the necessarily Euclidean character of geometry, hence in direct contradiction with GTR, Winternitz accordingly jettisoned it as a confusing and unnecessary appendage to what is taken as the fundamental Kantian endeavour: to establish the a priori presuppositions of physical knowledge. These are spatiality and temporality as 'unintuitive schemata of order' in general (as distinct from any particular chronometrical relations), the law of causality and assumption of continuity of physical processes, the principle of sufficient reason, and the conservation laws. Remarkably, the *necessity* of each of these putative a priori elements was soon to be challenged by the new quantum mechanics (see Quantum mechanics, interpretations of).

3 Logical positivism

The last redoubt still held by Kantians concerned the existence of synthetic a priori elements of physical knowledge; their defence was also attempted in the earliest works of Rudolf Carnap and Hans Reichenbach. Carnap's first philosophical publication, his dissertation on space (1922), essentially seeks a reconciliation

between GTR and the Kantian notion of a form of intuition. Carefully distinguishing between intuitive, physical and mathematical conceptions of space, Carnap also draws upon the infinitesimal validity of Euclidean geometry in a Riemannian manifold to argue that this geometry is presented in the intuition of a limited region. But since the global extension of this structure depends upon conventions, Euclidean geometry is not constitutive of experience in Kant's sense. Instead, inspired by Einstein's remarks on point coincidences, Carnap conceives the empirical content of physics ('matters of fact') as entirely given by spacetime coincidences while, as a synthetic a priori condition of the possibility of experience, intutive space is assigned a structure of topological relations since these alone preserve empirical content for all possible metric stipulations. By the late 1920s, Carnap had come to view form solely in logical or analytic, not synthetic, terms, concluding that 'there is no synthetic a priori'. This declaration became an article of faith among logical empiricists (see Carnap, R.).

Of all the logical empiricists, Reichenbach wrote the most extensively about problems of space and time and the philosophical interpretation of GTR. His early work (1920) also attempts to reconcile GTR with the Kantian emphasis on the a priori, although a modification in the conception of the a priori must be made in view of inconsistencies resulting from a straightforward combination of Kantian precepts and empirically confirmed principles of GTR (for example, the equivalence of inertial and gravitational mass). Reichenbach suggested that of the two primary senses of the Kantian a priori, as valid for all time and as constitutive of the object of knowledge, only the latter was compatible with the new physics of GTR. The remnant of the a priori is placed in 'principles and axioms of coordination', linking the (analytic) mathematics of physical theory with objects in the physical world, in effect, 'constituting' the latter. The systematic task of the philosopher of science now becomes the provision of an 'axiomatic' analysis of domains of physical knowledge. The aim here is to locate a theory's empirical core and then sharply to separate this 'objective content' from 'the subjective contribution of reason', systems of equations and the respective principles of coordination (seeTheories, scientific). Reichenbach came to view the important role of coordinative principles in less and less Kantian terms, stressing instead their character as freely chosen conventions or definitions, not derived from experience but essential to the applicability of mathematical theory to physical reality. From this viewpoint, 'the philosophical significance' of GTR is that it shows how empiricist analyses of scientific knowledge must be tempered with a recognition of the fundamental role of conventional elements in the basic propositions of physical theory. Thus, in order to obtain empirical spacetime measurements, certain stipulations must first be made governing the behaviour of measuring rods and clocks. Reichenbach's analysis of the relation of mathematical theory to empirical objects in terms of 'coordinative definitions' became a central tenet of the standard 'received view' of scientific theories dominant in philosophy of science till the 1960s (see Reichenbach, H.).

The founder of the Vienna Circle, Moritz Schlick, was already an opponent of the Kantian conception of physical knowledge; he had not yet embraced neopositivism (see Schlick, F.A.M. §1). In one of the earliest lay monographs on GTR, Schlick argued (1917), in a concluding philosophical section applauded by Einstein, for a tempered realism about physical theories which nonetheless recognized that no physical theory fails to have at least some arbitrary elements. Subsequently (1921), he subjected Cassirer's very general defence of a priori elements as methodological presuppositions in physical knowledge to a withering critique, arguing that such vague formulations could have no physical significance and hence were best abandoned. Together with H. Weyl, who was fundamentally opposed to positivism, Schlick emphasized that the union of space, time, *and* matter wrought by the geometrization of gravity in GTR is a mind-independent reality. Of course, Kant does not allow, on pain of surrendering transcendental idealism, that there is a coherent conception of space except as a form of sensibility.

4 Einstein's response

Einstein's reaction to the various philosophical appropriations of his theory are perhaps most widely known through casual remarks made in other contexts, but they may be tracked more closely in several explicit encounters: a meeting at the Sorbonne in Paris on 6 April 1922, with French philosophers, physicists and mathematicians, and in several rare book reviews of philosophical works dealing with his theory. The significance of these encounters lies in how Einstein distinguishes his views from others, not in the presentation of a comprehensive alternative. At the Paris meeting, Einstein made short shrift of Machian philosophy, praising Mach as a 'good mechanician' (no doubt in reference to Mach's views of the relativity of motion) but as a 'deplorable philosopher'. In this venue Einstein also suggestively remarked that 'each philosopher has his own Kant', probably alluding to the plethora of Kantian attempts to assimilate GTR, but perhaps as well intentionally

including himself. In this regard, it is instructive to consider that his most fully articulated objection to claims of the necessity of a priori elements in physical knowledge, voiced in a review of an otherwise forgettable book by a student of Cassirer (Elsbach 1924), is considerably different from the verificationist objections of Schlick. It is not that such elements are so general as to be beyond empirical confirmation or disconfirmation but rather that it is entirely arbitrary *which* components of a theoretical system are to be considered a priori. Here also Einstein concedes the justice of the criticism, already implicitly made by Cassirer, of his remark (cited in §1 above) that concepts in physics have validity only if they are individually connected with experience; now an explicitly holist line is affirmed: ‘ in general, it is not to single concepts but only to the system as a whole that possible experiences must correspond’ (1924b: 1,692). Reviewing Winternitz’s book, Einstein expressly agrees with the Kantian point that the fundamental principles of a theoretical structure ‘do not themselves originate in experience’; the difference with Kantians arises only in Einstein’s view that such principles are ‘ *insgesamt* [collectively] mere conventions like the ordering principle of words in a dictionary’ (1924a: 22). Interestingly, late in life Einstein returns to the language of Winternitz’s formulation of the essential Kantian postulate, that the world is ‘not given but posed’ as comprising ‘the truly valuable’ in Kantian doctrine. As before, however, Einstein takes the so-called categories not as ‘unalterable’ but as ‘(in the logical sense) free conventions’ (1949: 680, 674). Finally, in a review of a work which emphasized the on-going development (then underway by Einstein and others) of GTR into a ‘unified field theory’ which would contain all laws of nature as logical consequences, Einstein appears to endorse Meyerson’s explicitly realist orientation, noting that ‘ at the basis of all physical science lies a philosophical realism’ (1928: 253) (see Meyerson, É.). Expressions of such ‘metaphysical’ articles of faith, made with increasing frequency in his later years, were anathema to logical positivism.

See also: Bridgman, P.W.; Einstein, A.; Geometry, philosophical issues in §3; Relativity theory, philosophical significance of

T.A. RYCKMAN

References and further reading

Bollert, K. (1921) *Einstein’s Relativitätstheorie und ihre Stellung im System der Gesamterfahrung (Einstein’s Theory of Relativity and its Position in the System of Experience as a Whole)*, Leipzig: Steinkopff.(Technically proficient Neo-Kantian interpretation.)

Bridgman, P.W. (1949) ‘Einstein’s Theories and the Operational Point of View’, in P.A. Schilpp (ed.) *Albert Einstein: Philosopher-Scientist*, Evanston, IL: Northwestern University Press, 335-54.(Indicts GTR as not following methodology of STR.)

Carnap, R. (1922) *Der Raum. Ein Beitrag zur Wissenschaftslehre (Space. A Contribution to the Theory of Science)*, Berlin: Reuther & Reichard.(Sophisticated discussion of different conceptions of space and their interrelations.)

Cassirer, E. (1921) *Zur Einstein’schen Relativitätstheorie*, Berlin: Bruno Cassirer; trans. W.C. Swabey and M.C. Swabey, *Substance and Function and Einstein’s Theory of Relativity*, Chicago, IL: Open Court, 1923.(Richly textured, historically-oriented Neo-Kantian interpretation.)

Einstein, A. (1917) *Ueber die spezielle und die allgemeine Relativitätstheorie*, Braunschweig: Vieweg; trans. R.W. Lawson, *Relativity: The Special and the General Theory*, New York: Crown, 1961.(His lay exposition of special and general relativity.)

Einstein, A. (1924a) ‘Review of *Relativitätstheorie und Erkenntnislehre* by J. Winternitz’, *Deutsche Literaturzeitung*, 20-2.(Conventionalism of a priori elements in physical knowledge.)

Einstein, A. (1924b) ‘Review of *Kant und Einstein* by A. Elsbach’, *Deutsche Literaturzeitung* 45: 1,685-92.(Early statement of holist views on the relation of theory to experience.)

Einstein, A. (1928) ‘À Propos de *La Déduction Relativiste* by É. Meyerson’, *Revue philosophique de la France et de l’étranger*, 105: 161-6; trans. in É. Meyerson, *The Relativistic Deduction*, Dordrecht: Reidel, 1985, 252-6. (Endorses author’s realist orientation.)

Einstein, A. (1949) ‘Replies to Criticisms’, in P.A. Schilpp (ed.) *Albert Einstein: Philosopher-Scientist*, Evanston, IL: Northwestern University Press, 665-88.(Einstein’s later philosophical views.)

Elsbach, A. (1924) *Kant und Einstein*, Berlin and Leipzig: de Gruyter.(Neo-Kantian work drawing heavily upon Cassirer.)

Gödel, K. (1946/9) ‘Some Observations about the Relationship between Theory of Relativity and Kantian Philosophy’, in S. Feferman *et al*. (eds) *Kurt Goedel Collected Works*, New York: Oxford University Press, 1995, vol. III, 230-59.(Two versions of an unpublished essay attempting to minimize Kantian difficulties with

general relativity.)
Hentschel, K. (1990) *Interpretationen und Fehlinterpretationen der speziellen und der allgemeinen Relativitätstheorie durch Zeitgenossen Albert Einstein* (Interpretations and Misinterpretations of the Special and the General Theory of Relativity by Einstein’s Contempories), Basel-Boston-Berlin: Birkhäuser. (Encyclopedic reference of all the various ‘schools’ of interpretation.)
Mach, E. (1921) *Die Prinzipien der physikalischen Optik. Historisch and erkenntnispsychologisch entwickelt*, Leipzig: J.A. Barth; trans. J. Anderson and A. Young, *The Principles of Physical Optics: An Historical and Philosophical Treatment*, London: Methuen, 1926; repr. New York: Dover, 1953.(Published posthumously, contains the preface of July 1913 refered to in §1.)
Meyerson, É. (1925) *La Déduction Relativiste*, Paris: Payot; trans. with supplementary material D.A. Sipfle and M.A. Sipfle, *The Relativistic Deduction*, Dordrecht: Reidel, 1985.(GTR carries out Descartes’ ‘geometrization of Nature’.)
Petzoldt, J. (1921) ‘Das Verhältnis der Machschen Gedankenwelt zur Relativitätstheorie’ (The Relation of Mach’s Philosophy to the Theory of Relativity), an appendix to E. Mach, *Die Mechanik in ihrer Entwicklung: Historisch-Kritisch Dargestellt*, 8th edn, Leipzig: Brockhaus, 490-517.(Machian assimilation of GTR.)
Reichenbach, H. (1920) *Relativitätstheorie und Erkenntnis A priori*, Berlin: Springer; trans. M. Reichenbach, *The Theory of Relativity and A Priori Knowledge*, Berkeley, CA: University of California Press, 1965.(Early attempt to modify Kant to accommodate GTR.)
Reichenbach, H. (1928) *Philosophie der Raum-Zeit-Lehre*, Berlin and Leipzig: de Gruyter; trans. M. Reichenbach and J. Freund, *The Philosophy of Space and Time*, New York: Dover, 1957.(GTR shows role of stipulations in knowledge of nature.)
Schlick, M. (1917) *Raum und Zeit in der gegenwärtigen Physik*, Berlin: Springer; trans. of 4th edn of 1922 by H.L. Brose and P. Heath, ‘Space and Time in Contemporary Physics’, *Moritz Schlick: Philosophical Papers*, Dordrecht: Reidel, 1979, vol. 1, 207-69.(Concluding philosophical chapter praised by Einstein.)
Schlick, M. (1921) ‘Kritizistische oder empiristische Deutung der neuen Physik?’, *Kant-Studien* 26, 96-111; trans. P. Heath, ‘Critical or Empiricist Interpretation of Modern Physics’, in *Moritz Schlick: Philosophical Papers*, Dordrecht: Reidel, 1979, vol. 1, 322-34.(Attack upon a priori principles as unverifiable.)
Schneider, I. (1921) *Das Raum-Zeit Problem bei Kant und Einstein (The Space-Time Problem in Kant and Einstein)*, Berlin: Springer.(Novel Neo-Kantian assimilation of Einstein’s cosmology.)
Sellien, E. (1919) *Die erkenntnistheoretische Bedeutung der Relativitätstheorie*, Berlin: Reuther & Reichard. (Argues Kant’s doctrine of space and time is unscathed by GTR.)
Stachel, J. (1989) ‘Einstein’s Search for General Covariance, 1912-1915’, in D. Howard and J. Stachel (eds) *Einstein and the History of General Relativity*, Boston-Basel: Birkhäuser, 48-100.(Uncovers elliptically expressed significance of talk of ‘point coincidences’.)
Winternitz, J. (1924) *Relativitätstheorie und Erkenntnislehre (Theory of Relativity and the Theory of Knowledge)*, Leipzig and Berlin: B.G. Teubner.(Neo-Kantian interpretation abandoning Transcendental Aesthetic.)

General will

The fundamental claim for general will is that the members of a political community, as members, share a public or general interest or good which is for the benefit of them all and which should be put before private interests. When the members put the general good first, they are willing the general will of their community. The claim was given special and influential shape by Rousseau. He produced a comprehensive theory of the legitimacy of the state and of government, revolving around the general will. Some contend this solves the central problem of political philosophy - how the individual can both be obliged to obey the state's laws, and be free. If laws are made by the general will, aimed at the common good and expressed by all the citizens, the laws must be in accordance with the public interest and therefore in the interest of each, and each is obliged by the law yet free because they are its author. Rousseau's formulation has been much criticized. But others have found it essentially true and have variously adapted it.

1 Rousseau and the general will

Rousseau's general will is the will of a people or state (see Rousseau, J.-J. §§2-3). A people is not a natural society as is the family, but conventional. Rousseau uses the framework of social contract theory: independent individuals (men only) freely, equally and unanimously bind themselves into a single association (see Contractarianism §6). Thereby private persons become a public person, the people, and a collective body, the state, with a single will. That will is necessarily general in its object, its source and its application. (1) It aims exclusively at the common good of the whole body, which is the good of every member, good being moral personhood, freedom, and security under the law, possible only in a state. (2) The will of the whole body can be expressed only through the wills of its parts, that is, by all the members voting together. It is impossible to transfer one's will, so no member can give his right to express the general will to a representative or government or leader. Rousseau thus ties the general will to direct popular sovereignty. (3) The general will, in the form of laws, applies equally to all members.

The laws of a society ought to conform to the general will, and its government ought to conform to the laws. Rousseau realizes this would happen only if certain conditions are met, and the individual associates together consider themselves a single body so that they have a single will. The members of the state, the citizens, must be socialized to be patriotic and public spirited. There must be no extremes of wealth and poverty to heighten conflicting private interests. Private associations, factionalizing society and encouraging citizens to substitute a partial interest for the general interest, must be prevented or controlled. The government, appointed by the people as its servant, and potentially the most dangerous partial interest, must be kept in its place.

These conditions met, the general will is obvious to every citizen and guides the state according to the common end for which it was instituted. The citizens remain equal; none can impose a burden on others without undertaking it himself. The laws are just, being in the common interest. Everyone is free, because each individual obeys laws which he has made himself and which require what everyone wants for himself.

2 Criticisms

Some common criticisms rest on misunderstanding. Rousseau is attacked for stating that the general will is 'always right', that if one is outvoted that shows one made a mistake, and that one can be 'forced to be free'. These claims startle, but can be understood as true by definition provided the conditions for the expression of the general will are met.

Other criticisms, of both the general will having a general object and of its being willed generally, are more serious. It can be denied that individuals in a society constitute, or can constitute themselves, by a contract, a kind of person with a single will. This objection raises controversial and unsettled issues about the nature of social wholes. Even if such a 'person' is possible, it can still be denied that there can be a general interest which is in the interest of all members of a society. It can be argued that on specific matters, there is no policy or action which is in everyone's interest because individuals tend to have different aims and desires. People may agree, but that agreement is contingent, temporary, subjective and possibly arbitrary. Rousseau's belief that all men have the same basic moral and material interests (see his *Discourse on the Origins of Inequality* (1755)), so that they share a permanent and objective interest, requires assumptions about human nature and men's true interests which are

disputable (see Public interest §2).

If the idea of a public interest and a general will aiming at it is accepted, it is still vulnerable to powerful criticism. It can only be expressed through the votes of all the citizens, unanimously or by a majority. Rousseau admits the people's deliberations are not always right: a people is often fooled. But he offers no clear way of determining which majority decisions express the general will and which do not.

Even when the public interest is known, Rousseau recognizes that men are inclined to give preference to their own personal interests (which may be opposed to each other and to the public interest), following their private, particular will rather than the general will. This could be prevented, but only under improbable conditions and at great cost. In a suitably small and socially homogeneous society, if it could be established, all the citizens might well be sufficiently close in their values and aims to agree on common interests and put them first, as Rousseau found so admirable in ancient Sparta. However this would involve a high degree of deliberate socialization. The society's institutions would be designed to make citizens uniform. Under such conditions, the emergence of a general will is more plausible; but the price would be a society in which individuality was not developed but discouraged. There is always a balance to be struck between the freedom of individual members of society and the restrictions necessary for its existence. Rousseau's general will seems to require minimizing the individual's freedom, in the public sphere at least.

3 Later versions

Attempts, not altogether successful, have been made to retain the essence of Rousseau's idea while abandoning elements which are subject to criticism. Kant (§10) holds that it is not necessary that the people actually express the general will: it is enough if the law is what the people ought to have willed if they had been asked. The general will becomes hypothetical, what rational men would will, and operates as a 'regulative ideal', a standard to guide conduct. This avoids the practical problems in Rousseau. However, an acceptable test of what laws it is rational to will is needed, and what Kant provides - a political version of his categorical imperative - is disputable.

Hegel (§8) dispenses with social contract theory, and its individualism, instead situating the general will in actual historical societies. The moral and cultural beliefs and ideals of a society, and the legal, moral, economic, social and political institutions and practices which embody them, are its general will as it has developed; each society has established its own view of the public good. Its individual members learn this as they grow up, and express it and contribute to its further development as they live their lives in and through the society and its beliefs, rules and institutions. The general will ceases to be an ideal, hard to realize practically, and is given specific concrete content. Responsibility for expressing it is given to the civil service and the government. The people does not legislate directly, but elects representatives. Private associations are no longer seen as factional but as important media for the education of citizens. This is the general will brought down to earth - perhaps too much. Hegel realizes the ideas and institutions of one's society are inadequate in various ways; however, the direction of reform is indicated only very generally.

Recently the general will has been considered in terms of social choice theory (see Social choice). This has led to some clarification, especially of how the general will might emerge through voting. However, to the extent that this approach aggregates individual preferences (taking individuals as isolable from their society), and sets no moral limit to the content of decisions, it departs significantly from Rousseau's and gives the term a different meaning. Rousseau's general will requires the existence of a social unity between individuals, who thereby have a collective interest as members of that union in addition to their private interests as individuals; their general interest is to have a certain kind of society, bringing everyone specific moral benefits; and the verdict on what the general will is in any particular case must be delivered by the citizens themselves. Hence Rousseau's ideas, although complex and problematic, remain relevant, for instance to debates about freedom, democracy, republicanism and communitarianism.

PETER P. NICHOLSON

References and further reading

Bosanquet, B. (1899) *The Philosophical Theory of the State*, London: Macmillan, chaps 4-7.(Reworks Rousseau on Hegelian lines.)

Emmet, D. (1994) *The Role of the Unrealisable: A Study of Regulative Ideals*, London: Macmillan, chaps 1-4.

(Regulative ideals, and the general will as one.)

Grofman, B. and Feld, S.L. (1988) 'Rousseau's General Will: A Condorcetian Perspective', *American Political Science Review* 82 (2): 567-76.(How the general will could emerge through voting.)

Hiley, D.R. (1990) 'The Individual and the General Will: Rousseau Reconsidered', *History of Philosophy Quarterly* 7 (2): 159-78.(Connects Rousseau with contemporary debates between liberals and communitarians.)

Levine, A. (1993) *The General Will: Rousseau, Marx, Communism*, Cambridge: Cambridge University Press.(The use of the idea of the general will in current political philosophy.)

Nicholson, P.P. (1990) *The Political Philosophy of the British Idealists: Selected Studies*, Cambridge: Cambridge University Press, study VI.(Rousseau, Kant and Hegel discussed briefly, Bosanquet fully.)

Riley, P. (1986) *The General Will Before Rousseau: The Transformation of the Divine into the Civic*, Princeton, NJ: Princeton University Press.(Traces the theological origins of the idea of the general will in writers such as Pascal and Malebranche.)

Rousseau, J.-J. (1762) *Social Contract, Discourse on the Virtue Most Necessary For a Hero, Political Fragments, and Geneva Manuscript*, trans. J.R. Bush, R.D. Masters, C. Kelly and T. Marshall in R.D. Masters and C. Kelly (eds) *The Collected Writings of Rousseau*, Hanover, NH, and London: University Press of New England, 1994, vol. 4.(Rousseau's fullest account of the general will.)

Rousseau, J.-J. (1755) *Discourse on the Origins of Inequality Among Men (Second Discourse) Polemics and Political Economy*, trans. J.R. Bush, R.D. Masters, C. Kelly and T. Marshall in R.D. Masters and C. Kelly (eds) *The Collected Writings of Rousseau*, Hanover, NH, and London: University Press of New England, 1992, vol. 3.(Rousseau's view of human nature.)

Runciman, W.G. and Sen, A.K. (1965) 'Games, Justice and the General Will', *Mind* 74: 554-62; repr. in W.G. Runciman, *Sociology in its Place and Other Essays*, Cambridge: Cambridge University Press, 1970, 224-32. (Claims a useful sense may be given to the general will by reference to the Prisoner's Dilemma.)

Genetics

Genetics studies the problem of heredity, namely why offspring resemble their parents. The field emerged in 1900 with the rediscovery of the 1865 work of Gregor Mendel. William Bateson called the new field 'genetics' in 1905, and W. Johannsen used the term 'gene' in 1909. By analysing data about patterns of inheritance of characters, such as yellow and green peas, Mendelian geneticists infer the number and type of hypothetical genes. The major components of the theory of the gene, which proposed the model of genes as beads on a string, were in place by the 1920s. In the 1930s, the field of population genetics emerged from the synthesis of results from Mendelian genetics with Darwinian natural selection. Population geneticists study the distribution of genes in the gene pool of a population and changes caused by selection and other factors. The 1940s and 1950s saw the development of molecular genetics, which investigates problems about gene reproduction, mutation and function at the molecular level.

Philosophical issues arise: the question about the evidence for the reality of hypothetical genes, and the status of Mendel's laws, given that they are not universal generalizations. Debates have occurred about the nature of the relation between Mendelian and molecular genetics. Population genetics provides the perspective of the gene as the unit of selection in evolutionary theory. Molecular genetics and its accompanying technologies raise ethical issues about humans' genetic information, such as the issue of privacy of information about one's genome and the morality of changing a person's genes. The nature-nurture debate involves the issue of genetic determinism, the extent to which genes control human traits and behaviour.

1 Historical development of Mendelian genetics

Mendelian genetics arose from three traditions in the nineteenth century: one, the hybridist tradition of Mendel; two, speculative theories of heredity, variation and development; and, three, studies of the hereditary variation important in evolutionary change. Genetics emerged when the problems of heredity and embryological development were separated: the transmission of hereditary characters through generations can be studied independently of the development of the characters in embryos. Thus, bifurcation rather than unification marked early progress in genetics. Only recently have efforts to reunite genetics and embryology begun to be fruitful.

Mendel and Darwin viewed organisms as being composed of separable unit-characters that could vary independently of one another. The focus on contrasting variations among individual organisms, rather than on the similarities on which taxonomists had focused, was a key conceptual shift that enabled the development of both Darwinian natural selection and Mendelian genetics (see Evolution, theory of §1).

Mendelian genetics is usually presented by following the details of an exemplary cross-breeding experiment. The technique used by the Mendelian geneticist is artificial breeding. For example, pure breeding plants that produce yellow peas are bred with others that produce only green peas. The resulting hybrid peas are all yellow. After the hybrids self-fertilize, the resulting peas occur in a ratio of three yellow to one green. The empirical generalization of $3{:}\,1$ ratios is explained by assuming the existence of paired genes. It is a matter of dispute how much of this explanation is found in Mendel's paper. The post-1900 account assumes that paired genes occur in the pure breeding parents: designate yellow, AA; green, aa. The hybrid has the 'phenotype' of the dominant parent, yellow, but its 'genotype' is Aa. During the formation of germ cells in the hybrid, the two genes segregate (separate) to form germ cells (pollen and eggs) that are either A or a. Segregation is the key theoretical assumption of Mendelian genetics. The two types of germ cell form in equal numbers and combine randomly during fertilization; symbolically: $(\mathrm{A}+\mathrm{a})(\mathrm{A}+\mathrm{a})=1\mathrm{AA}+2\mathrm{Aa}+1\mathrm{aa}$. The AA and Aa plants have yellow peas; the aa, green. These theoretical assumptions thus explain the empirical generalization of $3{:}\,1$ ratios. Numerous characters in plants and animals exhibit such ratios. The law of segregation is also called 'Mendel's first law'.

Although Mendel himself did not distinguish them, Mendelian genetics came to contain what are called 'Mendel's two laws'. Mendel's second law was formulated separately only after exceptions were found, and is called the 'law of independent assortment'. The law states that genes for different traits, such as pea colour and pea shape, are inherited independently. In 1910, Morgan explained anomalies for independent assortment by postulating that some genes are linked in inheritance. Thus, the overgeneralization that all genes assort independently was specialized to the claim that all genes in different linkage groups assort independently. Linkage is explained by the

chromosome theory, which is discussed in §2. Because of the significant changes that linkage introduced, the field is sometimes called ‘Mendelian-Morganist genetics’.

In 1926, Morgan stated what he called ‘the theory of the gene’:

> The theory [of the gene] states that the characters of the individual are referable to paired elements (genes) in the germinal material that are held together in a definite number of linkage groups; it states that the members of each pair of genes separate when their germ-cells mature in accordance with Mendel’s first law, and in consequence, each germ-cell comes to contain one set only; it states that the members belonging to different linkage groups assort independently in accordance with Mendel’s second law; it states that an orderly interchange - crossing over - also takes place, at times between the elements in corresponding linkage groups; and it states that the frequency of crossing-over furnishes evidence of the linear order of the elements in each linkage group and of the relative position of the elements with respect to each other.
>
> (Morgan 1926: 25)

The occasional crossing over between genes in corresponding linkage groups enables the construction of genetic maps, which show the relative positions of genes to each other. Numerous questions were left unanswered by the theory of the gene of the 1920s. What is the chemical nature of the gene? What is the size of the gene? How do genes mutate and then stably reproduce those mutations? How do genes function to produce characters? These questions were not answerable by the technique of artificial breeding of Mendelian genetics.

2 Reality of the genes

The genes were hypothetical, theoretical entities, whose numbers and types were inferred in order to explain the distribution of characters in breeding experiments. The first step in providing the physical basis of heredity was the proposal in 1903-4 of the chromosome theory, the claim that Mendelian genes are arranged linearly along the chromosomes, which are thread-like bodies in the nuclei of cells. The chromosome theory was an interfield theory which integrated findings from the fields of cytology and genetics. Information about chromosomes, obtained by fixing and staining parts of cells and viewing them in a microscope, was related to information about genes, inferred from patterns of inheritance of characters. Predictions were made about the behaviour of chromosomes, based on genetic information, and vice versa. Chromosomes are chemically composed of proteins and deoxyribonucleic acid (DNA). Early speculations were that the genes were proteins. That view was challenged in the 1940s and 1950s as evidence mounted that DNA is the genetic material (or occasionally ribonucleic acid - RNA). The field of molecular biology (also called ‘molecular genetics’) emerged with the discovery of the double helix structure of DNA by Watson and Crick in 1953. By 1970, numerous problems had been solved: gene reproduction (opening and copying of the double helix of DNA), the genetic code and the sequence hypothesis (three bases along the linear length of the DNA code for one amino acid in a linear protein), and genetic function (DNA is transcribed to messenger RNA which is translated to protein). From a molecular genetics perspective, genes are segments of DNA or RNA which code for a molecular product or regulate or otherwise influence other genes. After 1970, a new era of molecular genetics began. With the discovery of enzymes that precisely cut and ligate DNA, DNA from widely divergent species can be recombined. Techniques for sequencing DNA have made possible the human genome project in which all the DNA in a human will be sequenced (see Molecular biology; Scientific realism and antirealism).

3 Universality of genetic laws

Philosophers trained in logic are fond of universal generalizations. At least some laws in physics appear to be universal generalizations, admitting of no exceptions, applying to all objects in the universe and throughout all time (at least after the first three minutes following the Big Bang). Biology seems to lack such strong universal generalizations. Although all living things on earth share enough of their metabolic and genetic machinery to indicate a common origin for all life, exceptions to generalizations abound. Certainly Mendel’s two laws are not without exceptions. The $1:1$ segregation of genes is violated when one gene knocks out its competitor. Such segregation distortion is uncommon, but does constitute an exception to the law of segregation. The oversimplification that one gene is associated with one character was complicated when it was found that one gene can influence many characters and one character may be influenced by many genes. One character is not always dominant over another; the hybrid can show a blend of the two parental characters or even a character different

from those of its parents. Special-case exceptions also exist for the general claims of molecular genetics, such as the universality of the genetic code and the claim that DNA is the genetic material.

Given the perspective of evolutionary theory, namely that adapted variants survive in natural selection, such variation and lack of universality are not surprising. Furthermore, the genetic apparatus in sexually breeding organisms, which is the object of study in Mendelian genetics, is itself a product of evolution. Future evolutionary changes may result in segregation being the special case or even disappearing, if other mechanisms prove more advantageous.

Defenders of the claim that there are laws in biology take a more local or conditional perspective on the nature of laws. Within the time span of genetic experiments, Mendel’s laws support predictions about the ratios of characters expected in hybrid crosses (see Laws, natural §1).

4 Reduction and theory change in genetics

Philosophers have debated whether the relation between Mendelian and molecular genetics is one of theory reduction. Theory reduction, in the sense proposed by the logical empiricists, requires that theories be in axiomatic form, such that the reduced theory can be deductively derived from the reducing theory. Furthermore, any term in the reduced theory not also in the reducing theory must be connected to a term in the reducing theory by a ‘reduction function’. Severe complications arise when trying to connect the Mendelian gene, characterized in terms of effects upon one or many phenotypic characters, and the molecular gene, which may code for a single protein that has myriad effects upon numerous biosynthetic pathways. Furthermore, although some possibility exists for putting Mendel’s laws of segregation and independent assortment into an axiomatic form, no ‘laws of molecular biology’ seem identifiable. No deduction seems possible.

The debate about derivational theory reduction arose within the context of the logical empiricist approach to philosophy of science (see Logical positivism §4; Unity of science). Alternatives to reduction have arisen within the context of the study of theory change, rather than the analysis of logical relations between static, axiomatically characterized, atemporal theories. From this historical perspective, one view is that Mendelian genetics was replaced by molecular genetics. Another view is that molecular genetics is the later development of the research programme begun by Mendelian genetics. In this latter alternative, Mendelian segregation, for example, is viewed as being explained at the level of the chromosome. Pairs of chromosomes separate in the formation of germ cells and this explains why genes, carried by the chromosomes, segregate. However, the problems of gene reproduction and gene function were not solved at the chromosomal level. Their solution required the level of the macromolecule. Molecular genetics thus solved problems posed by, but not solvable by, Mendelian genetics and the chromosome theory (see Reduction, problems of).

5 The gene as the unit of selection

Philosophers of biology, as well as evolutionary biologists, have debated the appropriate unit for evolutionary theory. Dawkins (1982) proposed that ‘selfish’ genes construct extended phenotypes to increase their chances of survival in future generations. Extended phenotypes include not only the hereditary characters of typical Mendelian phenotypes, but also any behaviour under genetic control, such as the building of dams by beavers. The genes are the replicators and the important units of selection; the extended phenotypes are vehicles for carrying the genes. Others have objected that the phenotype is the unit that interacts with the environment and hence the phenotype (the interactor) is the more important unit in selection. Other critics of genic selection argue that attempting to reduce the theory of evolution to an account of changes in individual genes produces artifacts, because the individual genes are usually parts of gene complexes that produce phenotype characters in complex ways. Also, the selective advantage of a gene may be dependent on its relative frequency in a gene pool; thus, say critics of genic selection, the genetic context is relevant to the selective advantage of a given gene. Reducing evolutionary theory to the level of the single gene may be too low a level for an adequate theory. Defenders of genic selectionism disagree.

There is another problem with characterizing the units of selection theory as the replicator and the interactor - it neglects the essential role of genetic variation in producing the raw material on which selection acts. Selection type theories are characterized by genetic variation, then interaction, which causes a benefit to accrue to some of the

variants, a benefit which may include genetic replication (see Evolution, theory of §3).

6 Genetic determinism and sociobiology

The issue of the role of genes in determining human traits has occasioned much debate. Human sociobiology raises new questions for the old nature-nurture debate about the role of genes in determining human social behaviour and the past evolutionary advantage of such hypothetical genes (see Sociobiology). The fallacy of genetic fatalism is the erroneous assumption that possession of a gene infallibly leads to a single phenotypic expression of the genetic trait. Different phenotypic responses are possible, given the genetic context and environmental variables that play a role in the expression of a trait.

See also: Genetics and ethics; Species

LINDLEY DARDEN

References and further reading

Beatty, J. (1995) 'The Evolutionary Contingency Thesis', in J.G. Lennox and G. Wolters (eds) *Philosophy of Biology*, Konstanz: University of Konstanz Press and Pittsburgh, PA: University of Pittsburgh Press, 45-81. (Argues against the claim that there are biological laws. See §3.)

Brandon, R.N. (1990) *Adaptation and Environment*, Princeton, NJ: Princeton University Press.(Argues for a phenotypic perspective in evolutionary theory.)

Brandon, R.N. and Burian, R. (eds) (1984) *Genes, Organisms and Populations*, Cambridge, MA: MIT Press. (Collection of articles by both biologists and philosophers about units of selection.)

Carlson, E.A. (1966) *The Gene: A Critical History*, Philadelphia, PA: Saunders. (History of the gene concept.)

Darden, L. (1991) *Theory Change in Science: Strategies from Mendelian Genetics*, New York: Oxford University Press.(Discusses reasoning strategies in theory change by tracing the development of Morgan's theory of the gene.)

Darden, L. and Cain, J.A. (1989) 'Selection Type Theories', *Philosophy of Science* 56: 106-29.(Argues for the role of genetic variation, in addition to interaction and replication, in natural selection.)

Darden, L. and Maull, N. (1977) 'Interfield Theories', *Philosophy of Science* 44: 43- 64.(Discusses the chromosome theory mentioned in §2, as well as theories from molecular biology that bridge two fields.)

Dawkins, R. (1982) *The Extended Phenotype*, New York: Oxford University Press. (An extended argument for the selfish gene perspective.)

Dunn, L.C. (1965) *A Short History of Genetics*, New York: McGraw-Hill.(History of genetics from Mendel to 1939.)

Judson, H.F. (1979) *The Eighth Day of Creation*, New York: Simon & Schuster. (Account of the development of molecular biology, with extensive quotations from interviews with participants.)

Kevles, D.J. and Hood, L. (eds) (1992) *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, Cambridge, MA: Harvard University Press. (Useful collection of articles from diverse perspectives.)

Kitcher, P. (1984) '1953 and All That: A Tale of Two Sciences', *The Philosophical Review* 93: 335-73; repr. in Sober, E. (ed.) *Conceptual Issues in Evolutionary Biology*, Cambridge, MA: MIT Press, 1994, 2nd edn, 379-99. (Argues against derivational reduction; in favour of molecular genetics as an 'explanatory extension' of classical genetics and in favour of autonomous levels of biological explanation.)

Morgan, T.H., Sturtevant, A.H., Muller, H.J. and Bridges, C.B. (1915) *The Mechanism of Mendelian Heredity*, New York: Holt.(Seminal monograph on the chromosome theory of Mendelian heredity.)

Morgan, T.H. (1926) *The Theory of the Gene*, New Haven, CT: Yale University Press.(The classic statement of the theory by its principal architect.)

Olby, R. (1985) *Origins of Mendelism*, Chicago, IL: University of Chicago Press, 2nd edn.(Places Mendel in the older hybridist tradition rather than as the first Mendelian geneticist.)

Olby, R. (1974) *The Path to the Double Helix*, Seattle, WA: University of Washington Press.(History of developments from 1930s to the 1953 Watson-Crick discovery of the structure of DNA.)

Sober, E. and Lewontin, R. (1982) 'Artifact, Cause, and Genic Selection', *Philosophy of Science* 47: 157-80. (Argues against genic selectionism. See §5)

Sterelny, K. and Kitcher, P. (1988) 'The Return of the Gene', *Journal of Philosophy* 85: 339-61.(A defence of genic selectionism. See §5.)

Stern, C. and Sherwood, E. (eds) (1966) *The Origin of Genetics, A Mendel Source Book*, San Francisco, CA: W.H. Freeman.(Contains translations of Mendel's original papers, letters and the papers of his rediscovers,

Hugo de Vries and Carl Correns.)

Strickberger, M. (1985) *Genetics*, New York: Macmillan, 3rd edn.(A standard genetics textbook with more detail on Mendelian genetics, which was presented briefly in §1.)

Waters, C.K. (1990) 'Why the Anti-reductionist Consensus Won't Survive the Case of Classical Mendelian Genetics', in A. Fine, M. Forbes and L. Wessels (eds) *PSA 1990*, East Lansing, MI: Philosophy of Science Association, vol. 1, 125-39; repr. in E. Sober (ed.) *Conceptual Issues in Evolutionary Biology*, Cambridge, MA: MIT Press, 1994, 2nd edn, 401-17.(A defence of reduction and a source for references to the reduction debate discussed in §4.)

Watson, J. (1980) *The Double Helix, A Norton Critical Edition*, New York: W.W. Norton.(Watson's autobiographical account of the discovery of the DNA double helix, reprints of original articles, and other related material.)

Watson, J., Hopkins, N.H., Roberts, J.W., Steitz, J.A. and Weiner A.M. (1988) *Molecular Biology of the Gene*, Menlo Park, CA: Benjamin Cummings, 4th edn. (Developments in molecular biology can be traced by examining the changes through the four editions of Watson's textbook.)

Wimsatt, W. (1980) 'Reductionist Research Strategies and their Biases in the Units of Selection Controversy', in T. Nickles (ed.) *Scientific Discovery: Case Studies*, Dordrecht: Reidel, 213-59.(Relates the issues of reduction and the controversy about units of selection, discussed in §4 and §5.)

Genetics and ethics

The identification of human genes poses problems about the use of resources, and about ownership and use of genetic information, and could lead to overemphasis of the importance of genetic make-up. Genetic screening raises problems of consent, stigmatization, discrimination and public anxiety. Counselling will be required, but whether this can facilitate individual choice is unclear. It will also involve problems of confidentiality. On the other hand genetic knowledge will pave the way for genetic therapies for hereditary disease. This raises the question whether a therapy which alters an individual at the genetic level is different in kind from conventional medical treatment. Genetic alterations passed on to future generations raise problems regarding consent. Genetic intervention could also be used to make 'improvements' in human genetic potential, leading to anxieties about eugenic attempts to design the species. Transgenics, the introduction of foreign genes into a genome, raises questions about the integrity of species boundaries and the assessment of risk.

1 Genetic screening, determinism and discrimination

It is with regard to the medical implications of genetic research that most ethical issues arise. The first concerns genetic screening. There is a distinction between screening and testing, although similar ethical issues arise in both cases: whereas testing applies to the analysis of the genetic make-up of individuals, to establish whether they carry a certain gene, screening determines the prevalence of a gene in a particular population. Different kinds of screening can be distinguished. (1) Neonatal screening, such as that used for phenylketonuria, which is primarily for identifying those suffering from diseases that can be ameliorated by early treatment. (2) Screening or testing in childhood, which will facilitate both the detection of those predisposed to develop late-onset diseases, and the identification of carrier status prior to reproductive age. There are particular problems here with consent, possible stigmatization, and the disclosure of information to the child. (3) Pre-conception screening of adults, which may increase the possibilities of informed reproductive choice. (4) Prenatal screening, which will have particular relevance for assessing the status of the foetus, with the associated ethical dilemmas concerning abortion (see Reproduction and ethics).

Screening of adults will increasingly have implications beyond reproductive choice. Whereas in the past, screening tended to be targeted at individuals known to be at risk, greater genetic knowledge will facilitate the screening of whole populations, given available resources. It may become possible to do 'broad spectrum' screening of individuals for a wide range of conditions, with the associated likelihood of increased public anxiety arising from changes in self-image following testing, and the danger of stigmatization.

Three categories of ethical issue arise. What criteria should be satisfied before the introduction of a particular screening programme? In accordance with what principles should a screening programme be carried out? What are the criteria for success - how should it be evaluated afterwards? This may involve follow-up of individuals screened, for example, with problems of privacy and confidentiality.

Genetic screening could lead to new forms of discrimination on the basis of a person's genetic make-up, irrespective of symptoms. The implementation of sickle-cell screening in the USA led to stigmatization of carriers, despite the fact that they did not have sickle-cell disease. Further worry surrounds the use of screening and testing to identify people who have a genetic predisposition to develop a disease in later life which has a genetic component, but which may require some environmental component to trigger it. If it were possible to identify which individuals were at increased risk, because of their genetic make-up, of succumbing to cancer caused by a particular toxin in the workplace, such individuals might find themselves unemployable or uninsurable. There is a question concerning what information about genetic predispositions ought to be made available to employers and insurers, and a conceptual question about what counts as a predisposition: there is a danger that the distinction between 'predisposed' and 'predetermined' might become blurred.

Arguments against genetic discrimination, or 'geneticism', are analogous to those opposing racism and sexism. There are worries, however, that greater genetic knowledge may lead to less tolerance of genetic disease, which may come to be seen as avoidable. A distinction has to be drawn, however, between the evaluation of a particular genetic condition and the value of the individual who suffers from it. Hence there are issues about choice of terminology. The choice between 'handicap', 'defect', 'disorder' and 'variant' is not ethically neutral.

2 Genetic counselling

Genetic screening programmes will increase the need for counselling. Genetic counselling is commonly understood to include the following elements: estimating the risk of having or propagating a genetic disorder; advising the counsellee at risk about the medical facts underlying the disorder; and facilitating and supporting a decision by the counsellee.

The first ethical dilemma concerns autonomy (see Autonomy, ethical). There is a question whether the very existence of genetic screening and counselling encourages pressure on individuals to make particular reproductive decisions. In line with the importance commonly given to autonomy, the prevailing model in genetic counselling has been ‘nondirective’. Counselling, in facilitating counsellee decision-making, is thus distinguished from advice. It has been suggested, however, that the very structure of the counselling interview undermines the possibility of nondirectiveness, and the fact that, for example, termination is an option may somehow suggest to clients that this is a choice they ought to make.

Informed consent is particularly problematic in genetic counselling, when variation in perceptions of risk is taken into account. To a couple being counselled about a one in four chance of giving birth to a child suffering from a recessive genetic condition, the numerical probability may be less meaningful than the nature of the condition itself. Furthermore, individuals’ perception of risk can be affected by the way in which information is presented. The introduction of population broad-spectrum screening will further complicate the issues. Informed consent will be yet more difficult when a number of diseases are at issue at the same time (see Consent).

The applicability of principles of medical ethics to genetic counselling is also complicated by the fact that the ‘client’ in this context is normally a couple or family rather than an individual. This raises the issue of confidentiality, relevant to both screening and counselling. The discovery of nonpaternity is an example of a dilemma that might arise. For example, genetic testing of a child and its parents may reveal that the male partner is not the genetic father of the child. Does he have a right to this information, or does the mother have a right to confidentiality?

The issue of confidentiality is linked with that of ownership, and the sense in which an individual can be said to own their genes. The fact that genetic relatives share genes has been advanced as a reason for questioning the individual’s right to confidentiality in some cases, where there is information that might affect the reproductive decisions of others.

3 Gene therapy

Genetic knowledge will facilitate gene therapy, which is controversial because it involves alteration of an individual at the genetic level. An analogy is frequently drawn, however, between gene therapy and organ transplants. Fears that genetic alteration would somehow interfere with the identity of the person are unrealistic in the light of proper understanding of what the techniques involve, and appreciation of the extent to which individuals can be affected by environmental techniques. Regarding the view that gene therapy should be seen in the same way as other forms of therapy, and similarly subject to requirements of informed consent, there are two complicating factors, however. One is the difference between somatic and germline gene therapy; the other is the potential use of gene therapy techniques for enhancement purposes.

Somatic gene therapy is the treatment of the body cells of an individual; the genetic alteration will not be passed on to descendants. Germline therapy, however, will involve changes that will affect subsequent generations. Many commentators see a moral difference between these two, suggesting that somatic, but not germline, therapy is acceptable. The arguments for this position are, first, that there is at present insufficient knowledge to justify taking risks that may be irrevocable (see Risk). If, for example, action were taken to eliminate a particular gene from the gene pool, we may discover too late that it had some hidden advantage and lose valuable genetic diversity. Second, whereas the individual undergoing somatic therapy has the opportunity to consent to or refuse therapy, those affected by germline therapy will be as yet unborn generations, who cannot express a choice in the matter. This raises the general ethical question whether we have obligations to future generations (see Future generations, obligations to). Third, there is an argument that individuals have a right to an unmodified genetic inheritance. In a more plausible form the argument claims a right to a genetic inheritance that has not been

tampered with except to remove pathology. Here it slides into the therapy/enhancement distinction.

A further problem associated with germline therapy concerns the means of carrying it out. It must either target the germ cells of an adult or be carried out at the embryonic stage. Its development will therefore involve research on human embryos and the associated ethical dilemmas.

On the other hand, there are arguments in favour of a positive obligation to develop germline therapy. First, there is said to be a general obligation to relieve suffering where we can, in this case by seeking treatments for genetic disease. Second, there is an argument from scientific freedom, though this is largely connected with a view that classifies gene therapy as research. Third, there is an argument from reproductive autonomy. Germline therapy might be, for some couples, the only way they can achieve a ‘winning combination’ of genes. Even if this is not the case, if a woman can consent to somatic therapy on her child, why cannot she ensure that all her descendants should be free of a condition such as Huntington’s chorea?

The debate over the distinction between gene therapy and genetic enhancement raises the question of what the goals of medicine should be. If medical technology is not limited to therapeutic interventions elsewhere, it is difficult to see the argument for limiting it in the genetic sphere alone: particular uses would have to be considered on their individual merits. An example to illustrate the distinction is the use of growth hormone to treat dwarfism, on the one hand, and to produce talented basketball players, on the other.

4 Eugenics and transgenics

Underlying many of the objections to genetic screening, genetic counselling with the option of termination of affected foetuses, and gene therapy, is an anxiety about eugenics, the attempt to improve the human gene pool. This arises because of historical precedents which show the potential for abuse. It is also one factor in the drive towards the upholding of nondirectiveness in counselling. The question arises as to the extent to which it is possible to advocate steps to improve genetic public health without the associations of discredited genetic eugenic policies. If public health is a good that should be pursued, then, *a fortiori*, genetic public health is a good that should be pursued. A distinction has commonly been drawn between negative and positive eugenics, negative being the attempt to reduce the frequency of genetic disease and positive the attempt to improve such qualities as intelligence or moral sensitivity. Some take the view that all eugenics is unacceptable; some that negative eugenics is morally acceptable, but not positive; others that both, under certain conditions, are acceptable. There are conceptual problems about how the negative-positive distinction is to be drawn, and also problems about the arguments for a moral distinction. There are arguments about the greater moral urgency of relieving suffering, but also intuitive objections to ‘designer babies’, the basis of which is not always clear, but which rely partly on an appeal to human dignity, partly on an objection to ‘playing God’, and partly on an appeal to lack of sufficient knowledge.

Transgenics, the incorporation of foreign genes into the genome of an organism, is commonly discussed in connection with animals and plants. One issue concerns the rights of the host organism, particularly in cases where suffering is likely to ensue. A more fundamental question is whether there is anything wrong in principle in crossing species barriers. Supporters point out that cross-breeding is nothing new, and that the definition of species is itself problematic (see Species). Opponents may fall back on arguments about what is natural to a certain form of life, or point to the risks of releasing genetically engineered organisms into the environment. In addition to the interests of the modified organisms themselves, there are consequences for human health, for other species, and for the biosphere to be considered. There is a question, first, as to what constitutes ‘release’, and, second, as to whether risk-assessment exhausts the moral issues.

5 The human genome project

The attempt to map and sequence the entire human genome raises ethical questions about whether that attempt is worthwhile. Large parts of the genome, for example, are referred to as ‘junk DNA’ - are they worth sequencing? The project is defended on the grounds that knowledge is worthwhile in itself: junk DNA may have functions we are as yet unaware of, and the mapping of the human genome will give us insight into the nature of humanity. The latter point is attacked as an overly reductionist view of human beings which ignores spiritual and environmental factors (see Human nature §2).

A second problem concerns ownership and exploitation of results of human genome analysis: whether, for example, it is permissible to patent human genetic sequences, and how the benefits of the knowledge gained will be distributed, given that there is a sense in which the genome belongs to the whole of humanity (see Responsibilities of scientists and intellectuals §3). These issues have been given added urgency by the result of a recent genetic experiment: the succesful cloning of a sheep. Nicknamed ‘Dolly’, the cloned sheep was produced by isolating a segment of DNA from another sheep and replicating it in a laboratory. It has been admitted by the scientists who performed the experiment that the same process could be applied to humans.

Because of their revolutionary and far-reaching potential, and because the fact of genetic relatedness poses particular problems for an individualistic ethic, advances in genetics have provoked discussion about the appropriate ethical framework to apply: whether there is any categorical objection to genetic interventions or whether some form of consequentialist analysis is adequate (see Consequentialism; Deontological ethics); whether established principles of medical ethics are applicable or whether an ethics of care based on feminist thought is preferable (see Medical ethics; Feminist ethics).

See also: Applied ethics; Bioethics; Genetics; Technology and ethics

RUTH CHADWICK

References and further reading

Annas, G.J. and Elias, S. (eds) (1992) *Gene Mapping: Using Law and Ethics as Guides*, New York: Oxford University Press.(Focuses on the medical implications of genetic research.)

Chadwick, D. *et al*. (1990) *Human Genetic Information: Science, Law and Ethics*, Chichester: John Wiley.(Deals with issues raised by the attempt to map and sequence the human genome.)

Chadwick, R. (1988) ‘Genetic improvement’, in D. Braine and H. Lesser (eds) *Ethics, Technology and Medicine*, Aldershot: Avebury.(On the distinction which has been drawn between negative and positive eugenics, negative being the attempt to reduce the frequency of genetic disease and positive the attempt to improve such qualities as intelligence or moral sensitivity.)

Chadwick, R. (1992) *Ethics, Reproduction and Genetic Control*, London: Routledge. (Examines issues relating to eugenics.)

Clarke, A. (1991) ‘Is non-directive counselling possible?’, *Lancet* 338: 998-1001. (Considers issues in genetic counselling.)

Clothier, C.M. *et al*. (1992) *Report of the Committee on the Ethics of Gene Therapy*, London: HMSO. (Multi-authored report on the ethical issues raised by gene therapy.)

Macer, D.R.J. (1990) *Shaping Genes*, Christchurch, New Zealand: Eubios Ethics Institute.(Contains material on transgenics; has a good bibliography and survey of relevant literature.)

Pembrey, M.E. (1991) ‘Embryo therapy: is there a clinical need?’, in D.R. Bromham *et al*. (eds) *Ethics in Reproductive Medicine*, London: Springer-Verlag.(On germline therapy and the possibility of a winning combination of genes.)

Wertz, D.C. and Fletcher, J.C. (1991) ‘Privacy and disclosure in medical genetics examined in an ethics of care’, *Bioethics* 5: 212-32.(Considers privacy and disclosure as they arise in relation to issues in genetics.)

Gentile, Giovanni (1875-1944)

Best known as the self-styled philosopher of Fascism, Gentile, along with Benedetto Croce, was responsible for the ascendance of Hegelian idealism in Italy during the first half of the twentieth century. His 'actual' idealism or 'actualism' was a radical attempt to integrate our consciousness of experience with its creation in the 'pure act of thought', thereby abolishing the distinction between theory and practice. He held an extreme subjectivist version of idealism, and rejected both empirical and transcendental arguments as forms of 'realism' that posited the existence of a reality outside thought.

His thesis developed through a radicalization of Hegel's critique of Kant that drew on the work of the nineteenth-century Neapolitan Hegelian Bertrando Spaventa. He argued that it represented both the natural conclusion of the whole tradition of Western philosophy, and had a basis in the concrete experience of each individual. He illustrated these arguments in detailed writings on the history of Italian philosophy and the philosophy of education respectively. He joined the Fascist Party in 1923 and thereafter placed his philosophy at the service of the regime. He contended that Fascism was best understood in terms of his reworking of the Hegelian idea of the ethical state, a view that occasionally proved useful for ideological purposes but which had little practical influence.

1 Life and works

Giovanni Gentile was born on 30 May 1875 at Castelvetrano in Sicily. He studied philosophy at the University of Pisa under Donato Jaja, a pupil of Bertrando Spaventa. Spaventa's ideas were also to have a profound impact on Gentile's thought. He graduated in 1897 with a thesis on the nineteenth-century Italian philosophers Rosmini and Gioberti, through whom he explored Kant and Hegel respectively. He later claimed that the germs of his mature philosophy were already present in this early attempt to pursue the Hegelian critique of Kant. This study was published in 1898 and was followed the next year by a critical examination of the writings of Karl Marx, *La filosofia di Marx: Studi critici (The Philosophy of Marx: Critical Studies)*, which took this project further. In this work he offered one of the earliest Hegelian readings of Marxism. Although he maintained that Marx's materialism involved an unsuperseded dualism between mind and matter, he shared his belief in the unity of thought and practice. He translated Marx's *Theses on Feuerbach* into Italian for the first time, using them to assimilate the Marxian doctrine of praxis to Vico's notion that we can only know what we have made. His own doctrine largely grew out of a radicalization of this thesis: namely, that we make the world through thought.

Gentile's friendship with Croce began during this period (see Croce, B. §§1, 4). Their collaboration on the latter's journal *La critica* was to have a profound impact on Italian cultural life. Together they attacked the dominant positivist tradition and worked for a revival of the idealist tradition of the Neapolitan Hegelians. In pursuance of this goal, Gentile contributed a series of detailed historical studies of Italian philosophy in the eighteenth and nineteenth centuries to *La critica*, and republished the works of Spaventa.

Gentile first developed his own theory through writings on the philosophy of education, beginning with his essay 'Il concetto scientifico della pedagogia' (The Scientific Concept of Education) (1901) and culminating in his two-volume *Sommario di pedagogia come scienza filosofica (Summary of the Philosophy of Education)* (1913-14). He became a *Privatdozent* at Naples in 1903, but only secured a permanent post in 1906, when he became professor of philosophy at the university of Palermo. He began to work out his theory in earnest, publishing an important article in 1912, 'L'atto di pensare come atto puro' (The Act of Thought as Pure Act), and a series of essays in 1913, *La riforma della dialettica hegeliana (The Reform of the Hegelian Dialectic)*. Up until this time, his relationship with Croce had been one of friendly and creative disagreement. Indeed, he could claim to have had an important influence on Croce, moving him in a more markedly historicist direction. Croce, however, now accused him of advocating either emotivism or a mystical spiritualism, while Gentile countered that his friend remained at heart a naturalist. Their disagreements, hitherto confined to private correspondence, went public in 1913 in a famous debate in the journal *La voce*. It marked the first step to their eventual dramatic split.

In 1914, Gentile succeeded to Jaja's chair at Pisa, where he wrote his major work, the *Teoria generale della spirito come atto puro (The Theory of Mind as Pure Act)* (1916). In 1917 he moved to Rome, where he completed the systematic exposition of his thesis with his two-volume *Sistema di logica come teoria del conoscere (System of*

Logic as Theory of Knowing) (1917, 1923). In 1920 he started up the *Giornale di filosofia italiana (Journal of Italian Philosophy)* as an organ for his students, although he continued to write for *La critica*.

In 1922 Gentile entered Mussolini's first cabinet as Minister of Education, instituting *La riforma Gentile*, a comprehensive overhaul of the secondary education system. In 1923 he joined the Fascist Party and thereafter became the self-styled philosopher of Fascism - a move that precipitated the final break with Croce, who remained a confirmed liberal. From 1924 he was director of the National Fascist Institute of Culture. He also acted as editorial director of the *Enciclopedia italiana* from its inception in 1925 to its completion in 1937. The only books of philosophical note from this period were his *Filosofia dell'arte (The Philosophy of Art)* (1931), which criticizes Croce's aesthetics, and *Genesi e struttura della società (Genesis and Structure of Society)* (1946), which offers the fullest account of his social and political philosophy. This last was written following the outbreak of civil war in Italy and was published posthumously. Gentile had joined the last ditchers in the German puppet Fascist Social Republic of Salò, and was assassinated by communist partisans in Florence on 15 April 1944.

2 Actualism

Gentile sought to elaborate a theory of knowledge that rejected all presuppositions of either a transcendent or an empirical reality lying outside human consciousness. Following Hegel, Gentile believed that by positing a fixed noumenal world as the transcendental ground of phenomenal experience Kant denied the truly creative faculty of the Understanding (see Kant, I. §5). Hegel's criticism did not go far enough for Gentile, however. Hegel did not wish to replace transcendental with subjectivist idealism. The contribution of the thinking subject was vital, but this did not mean that all knowledge was relative to the knower. Hegel's epistemology was underwritten by an ontological argument, which regarded being and consciousness as logically related as part of the development of a metaphysical entity - Spirit or Mind - immanent to them both (see Hegel, G.W.F. §3). Gentile argued that this solution remained unsatisfactory because the Hegelian Spirit was as transcendent from the human viewpoint as the Kantian categories. He believed the Hegelian distinction between *Phenomenology*, or the development of individual consciousness, and *Logic*, which regarded reality as the product of Spirit, created an apparent dualism between Spirit as the essence of everything and human thought. Drawing on Spaventa, Gentile sought to overcome this alleged Hegelian dualism by identifying the individual mind of the *Phenomenology* with the Spirit of the *Logic*. Transcendence was finally overcome in a philosophy of absolute immanence, in which the activity of the thinking subject was the sole basis for human knowledge and hence of the objects that constituted human experience.

Gentile, via Hegel and Spaventa, concluded by transferring to thought *tout court* the faculties originally ascribed by Kant to the transcendental subject and by Hegel to Spirit. The method of immanence, he contended, rendered thought a 'pure act' that is presuppositionless, the absolute foundation of the human world of experience. Sensation, on this theory, is misconstrued as a response to an object external to the subject. Rather, it represents an act of self-awareness on the part of the thinking subject through which the external world is internally comprehended as part of a process of self-constitution (*autoctisi*).

There is a clear danger that Gentile's theory risks degenerating into solipsism. He attempted to avoid this pitfall through a theory of language that in a number of respects anticipates the arguments of Wittgenstein. Language, Gentile argued, is the embodiment of thought through which we organize experience. It is the form and not simply the vehicle of consciousness. However, there can be no purely personal or private language (see Private language argument §§1-2; Wittgenstein, L.J.J. §13). The very concept of language entails a spiritual universe comprised of a system of meanings that are essentially public and communicable. Consequently, all thought implied the existence of an 'other' with whom we entered into dialogue within our own consciousness. The search for self-knowledge, and in the process truth, involved the aspiration to membership of an ideal community speaking a shared universal language.

Gentile provided two proofs of his 'theory of spirit as pure act'. First, he claimed it was the logical outcome of the rationalist enterprise of modern philosophy since Descartes and, borrowing from Hegel, the essence of Christianity. He adopted Spaventa's theory of the 'circulation of European thought' to illustrate this thesis in order to show how the ideas of the German philosophers he admired were adopted, or independently conceived, by Italian thinkers as part of a single philosophical tradition. Second, he maintained that the theory reflected the phenomenological development of human consciousness within the individual. This argument provided the main

theme of his writings on education and influenced his practical proposals for changes in the school curriculum, which gave pride of place to languages and philosophy and treated history and the natural sciences as preliminary stages of human understanding.

3 Social and political philosophy

Gentile's 'actual idealism' entailed the unity of theory and practice. Thinking and acting formed part of the same activity of self-constitution, whereby experience involved thought simultaneously comprehending and transforming the world. Drawing on Marx and Vico, Gentile elaborated an idealist theory of praxis whereby thought creates the world of fact through the transformation of nature - a process he termed the 'humanization of labour'. Actualism was the only philosophy compatible with human autonomy. It brought moral freedom back to earth from the noumenal world where Kant placed it, to become part of our everyday economic and political activity for which we were totally responsible. As he put it in a famous speech in support of the Fascist *squadristi*, it is impossible

> 'to distinguish moral force from material force …. Every force is a moral force, for it is always an expression of will, and whatever method of argument it uses - from sermon to cudgel - its efficacy cannot be other than that of entreating the inner man and persuading him to agree.'
>
> (1925: 50-1)

Just as his subjectivist idealism risks solipsism at the level of theory, so it appears to endorse a Hobbes-like 'war of all against all', in which might is right, at the level of practice. Once again, Gentile's solution to this dilemma was to argue that action, like consciousness, presupposed interaction with a community of values. Drawing on both the Hegelian conception of the state as the rational expression of the ethical life of the nation and Giuseppe Mazzini's theory of nationalism as the spiritual expression of 'God and People', he argued that each individual came to appreciate how their self-realization was intimately bound up with those of their fellows through social institutions that embodied shared cultural traditions and were oriented towards the common good. His defence of Fascism was largely in these terms, although this involved a considerable blindness to the difference between official doctrine and the practical reality. He coined the term totalitarianism to signify his ideal of the ethical state as the unity of the whole nation's thought and action. However, the Fascist corporatist state he idealized as the mechanism for bringing this unification about never had more than a merely formal existence.

The central difficulty of Gentile's thought lies in his attempt to reconcile an extreme subjectivism with a universalist theory of meaning and a collectivist conception of the state. Arguably the clearest account of his position comes in Chapter 4 of the posthumously published *Genesis and Structure of Society* on 'Transcendental Society or Society *in interiore homine*'. Gentile insisted that our consciousness of society and the community of meaning comes not from our actual interaction with other people but is generated within ourselves and the reflection of individuals upon themselves. Society and even language originates from within the individual (*in interiore homine*), not between individuals (*inter homines*), a view Gentile condemned as atomistic. The temptation to slip from this position into the assumption of an undifferentiated identity among individuals is all too easy, however, as Gentile's connivance with a highly repressive and authoritarian regime amply testifies.

See also: Italy, philosophy in §3

RICHARD BELLAMY

List of works

Gentile, G. (1950-) *Opere complete*, Florence: Sansoni, 55 vols.(The definitive edition of Gentile's works.)

Gentile, G. (1991) *Opere filosofiche*, ed. E. Garin, Milan: Garzanti.(A collection of the principal philosophical works leading up to and including the *Teoria generale dello spirito come atto puro*, together with a useful editorial introduction.)

Gentile, G. (1899) *La filosofia di Marx: Studi critici (The Philosophy of Marx)*, Pisa: Spoerri.(This book contains Gentile's *abilitazione* thesis in the history of philosophy of 1897 on historical materialism and studies on the philosophy of praxis emerging out of his discussions with Sorel, Labriola and Croce, to whom it was dedicated.)

Gentile, G. (1901) 'Il concetto scientifico della pedagogia' (The Scientific Concept of Education), *Rendiconti Accademia Lincei* 9; repr. *Scritti pedagogia I: Educazione e scuola laica*, 4th edn, Milan: Treves, 1932, 1-47.

(Gentile's early work on the philosophy of education.)

Gentile, G. (1912) 'L'atto di pensare come atto puro' (The Act of Thought as Pure Act), *Annuario della Biblioteca filosofica di Palermo* 1; repr. in Gentile 1991, vol. 27, 138-95.(The first statement of Gentile's 'actualism'.)

Gentile, G. (1913) *La riforma della dialettica hegeliana (The Reform of the Hegelian Dialectic)*, Messina: Principato.(Collects Gentile's early essays outlining his Spaventian reinterpretation of Hegel and the 'actualism' he derived from it.)

Gentile, G. (1913-14) *Sommario di pedagogia come scienza filosofica (Summary of the Philosophy of Education)*, Bari: Laterza, 2 vols.(Gentile's main treatise on the philosophy of education.)

Gentile, G. (1916) *Teoria generale della spirito come atto puro*, Pisa: Mariotti; trans. H. Wildon Carr, *The Theory of Mind as Pure Act*, London: Macmillan, 1922.(Presents the first full account of Gentile's actualist doctrine.)

Gentile, G. (1917, 1923) *Sistema di logica come teoria del conoscere (System of Logic as Theory of Knowing)*, Pisa: Spoerri. (The final extended statement of Gentile's theory.)

Gentile, G. (1920) *La riforma dell'educazione*. *Discorsi ai maestri di Triestre*, Bari: Laterza; trans. D. Bigongiari, *The Reform of Education*, New York: Harcourt Brace, 1922.(A collection of lectures delivered to Italian school teachers which offers the most accessible introduction to Gentile's ideas. It includes an interesting preface by Croce.)

Gentile, G. (1925) *Che cosa è il fascismo? Discorsi e polemiche (What is Fascism?)*, Florence: Vallecchi.(A collection of Gentile's early articles and speeches supporting fascism.)

Gentile, G. (1931) *Filosofia dell'arte*, Milan: Treves; trans. G. Gullace, *The Philosophy of Art*, Ithaca, NY: Cornell University Press, 1972.(The fullest account of Gentile's aesthetic theory, which had a major influence on that of the British philosopher R.G. Collingwood.)

Gentile, G. (1946) *Genesi e struttura della società*, Florence: Sansoni; trans. H.S. Harris, *Genesis and Structure of Society*, Urbana, IL: University of Illinois Press, 1960.(The fullest exposition of Gentile's social and political thought. The introduction to the translation traces the influence of Gentile outside Italy, particularly on Collingwood.)

References and further reading

Bellamy, R.P. (1987) *Modern Italian Social Theory: Ideology and Politics from Pareto to the Present*, Cambridge: Polity Press, ch. 6.(Locates Gentile's ideas within the context of the Italian idealist tradition and the political concerns of the times.)

Del Noce, A. (1990) *Giovanni Gentile. Per una interpretazione filosofica della storia contemporania*, Bologna: Il Mulino.(A sympathetic and at times contentious defence of Gentile's ideas.)

Harris, H.S. (1960) *The Social Philosophy of Giovanni Gentile*, Urbana, IL: University of Illinois Press.(The best account of Gentile's philosophy, with particular reference to his social and political thought.)

Holmes, R.W. (1937) *The Idealism of Giovanni Gentile*, New York: Macmillan.(A useful, if somewhat dated, overview.)

Jacobelli, J. (1989) *Croce e Gentile: Da sodalizio al dramma*, Milan: Rizzoli.(An account of the intellectual relationship between Croce and Gentile.)

Turi, G. (1995) *Giovanni Gentile. Una biografia*, Florence: Giunti.(The most recent intellectual and political biography.)

Gentzen, Gerhard Karl Erich (1909-45)

The German mathematician and logician Gerhard Gentzen devoted his life to proving the consistency of arithmetic and analysis. His work should be seen as contributing to the post-Gödelian development of Hilbert's programme. In this connection he developed several logical calculi. The main device used in his proofs was a theorem in which he proved the eliminability of the inference known as 'cut' from a variety of different kinds of proofs. This 'cut-elimination theorem' yields the consistency of both classical and intuitionistic logic, and the consistency of arithmetic without complete induction. His later work was aimed at providing consistency proofs for less restricted systems of arithmetic and analysis.

Gerhard Gentzen studied mathematics at Greifswald, where he was born, Göttingen, Munich and Berlin. In 1933 he obtained his doctoral degree at Göttingen under the supervision of Hermann Weyl and then from 1935 to 1943 he carried out research as an assistant to David Hilbert. Gentzen was conscripted in 1939 and then released from military service in 1942 after having become seriously ill. On his release, he took up a post at the University of Göttingen, where he had completed his post-doctoral thesis in 1940, moving to the German university in Prague in 1943. Gentzen was imprisoned by the Czech authorities in May 1945 because of his membership of National Socialist organizations. Three months later he died of malnutrition.

Gentzen was one of the most important contributors to Hilbert's programme of founding classical mathematics via finitary consistency proofs (see Hilbert's programme and formalism). For this purpose he developed several logical calculi building on earlier work on logical deduction by the Göttingen physicist Paul Hertz (see Gentzen 1933). Gentzen's main idea, already present in his doctoral dissertation, 'Untersuchungen über das logische Schließen' ('Investigations into Logical Deduction'), was to replace Hilbert-style calculi, in which theorems are deduced from axioms, by 'calculi of natural deduction' which start not with axioms but with sets of formulas considered as assumptions (see Natural deduction, tableau and sequent systems). The general form of an inference figure in such calculi is

$$\frac{A_1 \ldots A_\nu}{B} \quad (\nu \geq 1),$$

where $A_1, \ldots, A_\nu, B$ are formulas, and where the 'lower formula' *B* is deduced from the 'upper formulas' $A_1, \ldots, A_\nu$. Propositional connectives and quantifiers are defined by rules governing their introduction and elimination in proofs, and theorems are derived tree-style beginning with certain initial formulas treated as (possibly hypothetical) 'assumptions'. The idea was that such a procedure would be more like our 'natural' way of reasoning in mathematics than are the procedures of a Hilbert-style system. The resulting intuitionist calculus (NJ) can be extended into a complete classical calculus (NK) by simply admitting additional assumption formulas, for example, in the case of propositional logic, sentences of the form $A \vee \neg A$.

In order to avoid assumption formulas, Gentzen suggested intuitionist and classical 'sequent calculi' (LJ and LK, respectively). A metalogical sequent of the type $A_1, \ldots, A_\nu \vdash B_1, \ldots, B_\mu$ can be read as an abbreviation of the logical formula $A_1 \wedge \ldots \wedge A_\nu \supset B_1 \vee \ldots \vee B_\mu$. Sequents can be arranged in derivations with initial sequents of the unconditional form $D \vdash D$. Besides the operational rules for introduction and elimination of particular operators, Gentzen introduced structural rules (so-called 'inference figure schemes'). They include 'thinning', which allows the introduction of extra formulas in antecedents, 'contraction' of formulas stated more than once, 'interchange' of formulas in antecedents and consequents, and the important 'cut' rule

$$\frac{\Sigma \vdash A \quad \Sigma, A \vdash B}{\Sigma \vdash B},$$

with Σ an arbitrary set of formulas. Gentzen's cut-elimination theorem ('*Hauptsatz*') says that every derivation in LJ or LK can be transformed into a derivation with the same end sequent in which the cut rule is not used. The *Hauptsatz* can be used to prove the consistency of classical and intuitionist predicate logic. Gentzen also used it to solve the decision problem for intuitionistic propositional logic and to give a new proof that the law of excluded middle cannot be derived in formalized intuitionistic logic.

Using a sharpened form of the cut-elimination theorem (commonly referred to as the Gentzen-Herbrand theorem) and the subformula property (stating that all formulas occurring in the course of a derivation in the sequent

calculus without cuts are subformulas of the formulas in the end sequent), Gentzen succeeded in proving the consistency of arithmetic without complete induction. However, he saw this as only a first step, since, as he noted, arithmetic without induction is of only little practical significance in number theory (1969: 115). In approaching the project of finding consistency proofs for less restricted systems of arithmetic and analysis Gentzen had to face Gödel’s result that the consistency of a formalized theory embracing arithmetic cannot be proved using only means formalizable in this theory. For Gentzen, this did not yield the impossibility of consistency proofs for such theories. He admitted a ‘restricted transfinite induction’ up to Cantor’s first epsilon number ε_0 (the limit of $\omega, \omega^{\omega}, \omega^{\omega^{\omega}}, \ldots$), which cannot be formalized in classical arithmetic, which he nevertheless claimed could be interpreted constructively (1969: 231). Hence he regarded a consistency proof using ε_0-induction as a contribution to the realization of a liberalized Hilbert’s programme although the form of induction used goes beyond the finitary methods of proof as described by Hilbert.

In ‘Die Widerspruchsfreiheit der reinen Zahlentheorie’ (1936a), Gentzen offered a first proof of the consistency of classical arithmetic by arranging the proofs of classical number theory into an ordering of type ε_0 such that the consistency of a given proof in the sequence follows from that of its predecessors. This done, an application of ε_0-induction secures the proof of the consistency of the full family of proofs of classical arithmetic.

Gentzen provided new insights into the interrelationship between the intuitionist and formalist approaches to the philosophy of mathematics. He aimed at reconciling both positions with the help of a consistency proof which saved all of classical analysis by constructive techniques. The techniques he developed are now standard in proof theory.

See also: Gödel’s theorems; Logical and mathematical terms, glossary of; Proof theory

VOLKER PECKHAUS

List of works

Gentzen, G.(1969) *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland.(English translations of most of Gentzen’s papers, including the first publication of some. See also Szabo’s ‘Biographical Sketch’ and ‘Introduction’.)

Gentzen, G.(1933) ‘Über die Existenz unabhängiger Axiomensysteme zu unendlichen Satzsystemen’, *Mathematische Annalen* 107: 329-50; trans. ‘On the Existence of Independent Axiom Systems for Infinite Sentence Systems’, in *The Collected Papers*, 1969, 29-52.(A discussion of Paul Hertz’s theory of logical deduction.)

Gentzen, G.(1935) ‘Untersuchungen über das logische Schließen’, *Mathematische Zeitschrift* 39: 176-210, 405-565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969; trans. ‘Investigations into Logical Deduction’, in *The Collected Papers*, 1969, 68-131.(Gentzen’s doctoral dissertation, including the calculi of natural deduction and sequents, and the consistency proof for arithmetic without complete induction.)

Gentzen, G.(1936a) ‘Die Widerspruchsfreiheit der reinen Zahlentheorie’, *Mathematische Annalen* 112: 493-565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1967; trans. ‘The Consistency of Elementary Number Theory’, in *The Collected Papers*, 1969, 132-213.(Includes a discussion of finitism and the introduction of transfinite induction.)

Gentzen, G.(1936b) ‘Die Widerspruchsfreiheit der Stufenlogik’, *Mathematische Zeitschrift* 41 (3): 357-66; trans. ‘The Consistency of the Simple Theory of Types’, in *The Collected Papers*, 1969, 214-22. (Discusses the consistency of the simple theory of types.)

Gentzen, G.(1938) *Die gegenwärtige Lage in der mathematischen Grundlagenforschung: neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie*, Leipzig: Hirzel; part 1 repr. in *Deutsche Mathematik* 3: 255-68, 1939; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969; trans. ‘The Present State of Research into the Foundations of Mathematics’, in *The Collected Papers*, 1969, 234-51.(The first part gives an excellent account of the state of proof theory in the 1930s.)

Gentzen, G.(1943) ‘Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie’, *Mathematische Annalen* 119 (1): 140-61; trans. ‘Provability and Nonprovability of Restricted Transfinite Induction in Elementary Number Theory’, in *The Collected Papers*, 1969, 287-308. (Gentzen’s post-doctoral thesis.)

References and further reading

Prawitz, D.(1965) *Natural Deduction: A Proof-Theoretical Study*, Stockholm Studies in Philosophy 3,

Stockholm, Göteborg and Uppsala: Almqvist & Wiksell.(A classic study of Gentzen's logic and its development up to 1965.)

Vihan, P.(1995) 'The Last Months of Gerhard Gentzen in Prague', *Collegium Logicum* 1: 1-7.(An account of Gentzen's imprisonment.)

Geology, philosophy of

In the mid-1960s, geology underwent a conceptual revolution. Prior to that time, most geologists believed that the continents and oceans were fixed and permanent, the basic features of the earth's crust. Subsequently they came to agree that the earth was covered by rigid plates, thin in relation to the earth's diameter, in which the continents were embedded like logs in icebergs. It was the creation, movement and destruction of these plates that were responsible for the mid-ocean ridges, the areas of mountain building and earthquake activity, and the deep ocean trenches.

This conceptual revolution also marks a shift in the philosophy of geology. From the early nineteenth century, the chief philosophical question posed by geology was whether a historical science encountered special epistemic problems, a question that was usually answered by invoking the principle of uniformitarianism. In its strict form this stated that the only kind and intensities of causes that could be used to explain past geological phenomena were those that could be directly observed. Many sloppier formulations were invoked under the same name. Since the revolution, philosophers have turned to geology chiefly to use the revolution to exemplify or challenge one or another theory of scientific change.

1 Geology as a historical science: uniformitarianism

Early in the nineteenth century, geologists, who had previously concerned themselves primarily with the location and chemistry of economically valuable ores, began instead to concentrate on unravelling the earth's history. For the next century-and-a-half, the major issue in philosophy of geology was what it was to be a historical science. Most of the hard thinking about this was done by geologists, not philosophers.

Often cited in this regard is James Hutton, the late-eighteenth-century Scottish geologist and natural philosopher, who coined the phrase that the processes of geology showed 'no vestige of a beginning, no prospect of an end'. But in fact the major analysis of geology as a historical science was advanced by Charles Lyell, the British geologist, in his three-volume *Principles of Geology* (1830-3). Although Lyell himself did not call this analysis uniformitarianism, the Victorian philosopher-scientist William Whewell dubbed it so in the *Quarterly Review* in 1832. What Lyell had suggested was: one, that the laws of nature have not changed over time (law uniformitarianism); two, that the kinds of cause operating have not changed over time (kind uniformitarianism); and three, that the intensity of those causes has not changed over time either (degree uniformitarianism).

Law uniformitarianism was an uncontroversial position in the early nineteenth century. Geologists agreed that the laws of nature had neither changed nor been suspended. With the possible exception of the introduction of new animal and plant species, miracles were no longer part of the geological repertoire. Even in the case of new species, their possibly miraculous creation did not signify a change of natural law but an exception to it. Lyell himself dealt with this by having new species created in a law-like manner.

Kind and degree uniformitarianism, by contrast, were controversial. It seems plausible to believe that Lyell had taken these positions as a result of adapting a well-known methodological principle to geology, the principle of true causes (or vera causa principle) laid down by Newton in the Rules of Reasoning in the *Principia mathematica* (1687). This stated that the scientist should 'admit no more causes of natural things than such as are both true and sufficient to explain their appearances'. Newton's prestige was such that, during the eighteenth century, many different glosses were placed on the method of true causes (see Scientific method §2). It seems likely that Lyell took his particular interpretation from the Scottish philosopher Thomas Reid (1710-96), who argued in his *Essays on the Intellectual Powers of Man* (1785) that the vera causa principle meant that: one, there must be direct observational evidence that causes of the kind assumed do exist, and two, there must be evidence that the causes were sufficient to produce the purported effects.

As Lyell saw it, the problem for the geologist who attempted to interpret the past was the lack of direct observational access to past causes. The only causes that could be readily observed were those acting at present. Therefore the geologist who attempted to follow the method of true causes was in general limited to causes of a kind that operate at present. Equally, in conjecturing about the unknown effects of some observed cause, geologists must confine themselves in a similar way. In short, kind and degree uniformitarianism can be seen as a

straightforward extension of the principle of true causes to situations in which cause and effect are widely separated in time.

Geologists' attempts to reconstruct the past, therefore, posed a special case of an epistemic problem central to the other sciences: how to go from the known to the unknown, from the observable to the unobservable. But whereas for physicists this jump was generally from the macro-world to the micro, for geologists it was from the present to the past. Even if geology were an historical science, its epistemic problems could be tackled by a suitable modification of methodological principles already established in the other sciences.

Lyell put his principle of uniformitarianism to good effect in constructing his system of geology. He was unhappy with the common belief that the earth had once been molten and that its temperature was declining. Such a theory did not meet his vera causa criterion for there was no direct observational access to the original state of the globe. There was, however, evidence (fossilized tropical plants, for example) that the temperature of the northern hemisphere had been lower in the past. To explain this, Lyell proposed an alternative. At present, the earth's temperature at a given spot depends not solely on latitude but also on the distribution of land and sea, continental climates being very different from oceanic ones. Distribution of land and sea is thus a vera causa of climate. If the distribution of continents and islands had been different in the past, then the temperature in the northern hemisphere could have been warmer even though the average temperature of the globe remained the same.

Lyellian uniformitarianism won certain influential adherents in the nineteenth century, among them the astronomer and philosopher of science John Herschel, the mathematician Charles Babbage, and the natural historian Charles Darwin. Indeed, all three, realizing that the success of Lyell's vera causa theory of climate depended on a prior vera causa theory of the elevation and depression of land masses, attempted to identify a 'true cause' for land elevation, something that Lyell had not succeeded in doing in the *Principles*.

Most geologists, though, were sceptical. To them, it appeared quite possible that certain causes now in operation (ice, for example) might not have existed in the past (if the globe had been warmer, for example). To almost all geologists, the suggestion that causes had never varied in intensity was little short of perverse. Consequently most subscribed to the rival position, termed catastrophism by Whewell, which admitted the possibility of past causes different in kind and in intensity from those currently operating. Catastrophism in this strict sense was not tantamount to asserting that all geological effects were to be explained by sudden and dramatic changes and certainly not to the assertion that the laws of nature had changed over time.

In the twentieth century, for reasons that are not entirely clear, it became geological orthodoxy to invoke uniformitarianism, though by now much of the force of Lyell's original thesis had been lost. The term was frequently used to mean 'actualism' - that the past should be interpreted in terms of present causes. Lyell himself had never simplistically equated the 'present' and the 'known'. For him, the core dichotomy was between the observable and the unobservable, not between the present and the past. While most observable causes were present causes, sometimes we had access to past causes from the written records of antiquity. For example, on the basis of historical records of volcanic activity, Lyell was quite prepared to argue that the places where volcanoes would erupt in the future would vary just as they had varied in the past.

Uniformitarianism was also confused with 'gradualism', the belief that all geological processes had acted slowly and gradually, again something that Lyell never claimed. For him, the question of the pace of geological processes was simply a matter of what was observed. If there was observational evidence of meteor impact or of large-scale floods, then such agencies could be invoked in the past.

In the 1980s and 1990s, catastrophism (understood as anti-gradualism) became newly fashionable as geologists came to understand the effect of sudden and dramatic changes on the earth's surface. The detailed investigation of mass extinctions and the debate about whether these could be explained by meteor impact contributed to this (see Evolution, theory of §3). (Lyell would have been quite happy to accept our current evidence of meteor impact as adequate to establish meteors as a vera causa of geological change.)

Meanwhile, throughout the first half of the twentieth century, professional philosophers remained silent about the problems of geology. When they did speak, it was to say that, since geology dealt with the past, its laws must perforce be supplied by physics and chemistry. Hence, once the philosophical problems of physics and chemistry had been solved, solutions to the problems of geology would follow without further ado.

2 Scientific change and the plate tectonic revolution

During the 1960s, geology underwent a sudden and dramatic theoretical revolution. According to the new plate tectonic theory, the earth is covered by a limited number of rigid plates, thin in relation to the diameter of the earth, created at mid-ocean ridges (actually large mountain chains), spreading out from those ridges to be consumed again in subduction zones. Other phenomena - mountain building, earthquake activity, gravity anomalies, and so on - are explained in relation to the movement and intersection of these plates.

The revolution changed the emphasis of geology once more; reconstructing the earth’s past declined in importance and understanding the causal processes going on at present became the major focus of attention. These causes, while in principle compatible with those invoked by physicists and chemists, were in practice of a different order, including creation of new land at mid-ocean ridges, mountain building where one plate overrode another, destruction of land in subduction zones, and possible convection currents in the earth’s interior. Nonetheless, neither geologists nor philosophers have worried overtly about the philosophical foundations of the entities and processes asserted in plate tectonic theory. Instead the attention has gone to the plate tectonic revolution as a case of scientific change.

The plate tectonic revolution disturbed geologists deeply: within a matter of a few years, their most basic ideas about the structure of the earth were overturned. Equally important, their implicit philosophy of science simply gave no warrant for such a change. Ultimately derived from positivism, geologists’ image of science suggested that it progressed by a simple confrontation of hypothesis and evidence so that gradually and cumulatively a body of well-founded conclusions was established (see Logical positivism §4).

Casting around for some interpretation of science that would make sense of this event, geologists involved in the revolution quickly lit upon Thomas Kuhn’s *Structure of Scientific Revolutions*, serendipitously published just a few years previously in 1962. This gave licence for the possibility of revolutionary change in the most basic postulates of a science (see Kuhn, T.S.).

Philosophers of science, while agreeing that the changes were indeed dramatic, have been less inclined to treat the revolution as specifically Kuhnian. They have denied both that pre-plate tectonic geology was immature and that it fitted Kuhn’s description of normal science. They have questioned whether geological methods and aims changed concurrently with theory. They have dismissed the idea that pre-plate tectonic geology and plate tectonics were incommensurable (see Incommensurability).

But while being unconvinced that the geological revolution could be characterized as Kuhnian, they have agreed with geologists that this dramatic theoretical change does pose a challenge to the methodologies of science proposed by positivists. There is now a substantial and growing literature that seeks to interpret these changes in light of different methodologies.

See also: Causation; Laws, natural; Observation; Science, 19th century, philosophy of; Scientific realism and antirealism

RACHEL LAUDAN

References and further reading

Frankel, H. (1987) ‘The Continental Drift Debate’, in H. Tristram Engelhardt Jr. and A.L. Caplan (eds) *Scientific Controversies: Case Studies in the Resolution and Closure of Disputes in Science and Technology*, Cambridge: Cambridge University Press.(A good short account of the history of plate tectonics, both historically and philosophically sophisticated.)

Kitts, D. (1977) *The Structure of Geology*, Dallas, TX: Southern Methodist University Press.(The sole significant attempt to apply positivist philosophy of science to geology.)

Laudan, R. (1987) *From Mineralogy to Geology: The Foundations of a Science, 1650-1830*, Chicago, IL: University of Chicago Press.(Uses history to explore the conceptual foundations of geology.)

LeGrand, H. (1988) *Drifting Continents and Shifting Theories*, Cambridge: Cambridge University Press. (Contrasts philosophical and sociological accounts of scientific change.)

Lyell, C. (1830-3) *Principles of Geology*, London: John Murray, 3 vols; repr. Chicago, IL: Chicago University Press, 1989.(The classic formulation of uniformitarianism aptly introduced by Martin Rudwick, a leading

historian of geology.)

Geometry, philosophical issues in

The least abstract form of mathematics, geometry has, from the earliest Hellenic times, been accorded a curious position straddling empirical and exact science. Its standing as an empirical and approximate science stems from the practical pursuits of land surveying and measuring, from the prominence of visual aids (figures and constructions) in geometric proofs and, in the twentieth century, from Einstein's General Theory of Relativity, which holds that the geometry of spacetime is dependent upon physical quantities. On the other hand, very early on, the symmetry and perfect regularity of certain geometric figures were taken as representative of a higher knowledge than that afforded by sense experience. And its concern with figures and constructions, rather than with number and calculation, rendered geometry amenable to axiomatic formulation and syllogistic deduction, establishing a paradigm of demonstrative knowledge which endured for two millennia. While the progress of mathematics has surmounted traditional distinctions between geometry and the mathematics of number, leaving only a heuristic role for geometric intuition, geometric thinking remains a vital component of mathematical cognition.

1 The idea of a demonstrative science

The earliest records of activity in geometry come from Babylon. What we know as Pythagoras' theorem appears in cuneiform texts of around 2600 BC, where it is given an empirical and approximate verification. As the name indicates, the subject of geometry originated in the practical pursuits of land measurement and surveying (though probably not first among the Egyptians, as Herodotus famously reports). Although the Greeks did not originate geometry, it is to them that we owe the conception of geometry as an exact demonstrative science as against an empirical and practical discipline. While anecdotal evidence identifies Thales of Miletus as the originator of the idea that geometric statements are to be proved, it is generally accepted that the Pythagoreans had transformed mathematics into a deductive science by around 500 BC. Pythagoras himself is said to have conferred upon geometry the standing of *theōria*, the manner of contemplative knowledge alone worthy of a free and not a slave people; and to have propounded the belief that its true subject matter pertained to intelligible objects rather than to sense experience. Later, in Plato, geometric figures - circles, triangles, the regular solids - became archetypal ideal forms: perfect, universal, absolute, eternal and harmonious; whereas their sensible instantiations in matter were seen as inexact, particular, relative, temporal and discordant (see Pythagoreanism §2).

Early Pythagorean methods of proof had probably not advanced much beyond primitive diagrammatic methods. By arranging pebbles in simple geometric arrays, a number of elementary theorems had been discovered by direct observation. Other visual methods included superposition, whereby the identity or congruence of two geometric figures was established by moving one figure into point-for-point coincidence with another. Some have conjectured that the first Greek proofs of the Pythagorean theorem were discovered in this manner. While such methods were suitable to the discovery of simple arithmetic and geometric facts, they were not as conclusive as the deductive methods canonically compiled in Euclid's *Elements of Geometry* (*c*.300 bc). Through its influence, the idea of a demonstrative science was established and geometry became the paradigm of systematic presentation of a body of knowledge in terms of logical deductions from axioms, whose truth was antecedently recognized.

At the basis of geometry, Euclid distinguished axioms, postulates and definitions. Euclid calls axioms 'common notions' which, as the starting points of demonstration, are held to be immediately evident, non-demonstrable truths or 'intelligible principles' deniable only on pain of absurdity. In this, he followed Aristotle rather than Plato, for whom a dialectical demonstration of the starting point of any particular systematic inquiry into the nature of things was required (*Republic* VII.533c). The five axioms all concern the concept of equality and might be considered as stipulating properties of identity applicable to magnitudes generally, for example, the third: 'Two things equal to a third thing are equal to each other.'

In addition to his axioms or 'common notions', Euclid also stated five postulates which are not asserted as true but rather as hypotheses to be tested through the conformity with experience of results deduced from them. The first three outline the permissible means by which geometric figures may be constructed: (1) by drawing a straight line from any point to any other point; (2) by producing a finite straight line continuously in a straight line; (3) by describing a circle with any given point as centre and any given distance as radius. These postulates essentially

delimit geometric construction as a procedure involving only straight edge and compass. This bounds what has been traditionally termed 'elementary geometry' and one may ask why such constructions were singled out, for at an early date Greek geometers had already recognized that there were many interesting problems in construction which did not appear to be solvable by these means alone. Three famous examples known in ancient times are: to construct a square having the area of a given circle ('squaring the circle'); to trisect a given angle; and to construct a cube having double the volume of a given cube. Plutarch (*c*.46-120) reports that Plato was incensed by the attempts by Eudoxus and Archytas to solve the problem of doubling the cube by employing mechanical instruments in trial and error fashion; for Plato, such methods accorded sense experience an unwarranted legitimacy and hence were a corruption of the 'pure excellence of geometry'. In fact, not until the nineteenth century were rigorous proofs given that these are impossible ruler and compass constructions, each additionally requiring some version of a principle of continuity.

Of the remaining two postulates, (4) All right angles are equal to each other, and (5) Given a line *l* and a point P not on it, there is a unique line parallel to *l* through P, (5) was often singled out as lacking the evidential force of the others. As a result, from Euclid's time up to the early part of the nineteenth century, many attempts were made in vain to derive it from Euclid's other assumptions. (Mathematical practice over the centuries had largely submerged Euclid's evidential distinction between axioms and postulates.) Reflecting on this dismal record in 1819, C.F. Gauss had to admit that 'In the theory of parallels, we are still no further than Euclid', a situation he called 'the shame of mathematics'. Resolution took quite an unexpected form. After unsuccessfully attempting to demonstrate the postulate, both J. Bolyai (in 1823) and N. Lobachevskii (in 1826) independently formulated a consistent non-Euclidean (hyperbolic) geometry in which the fifth postulate did not hold, the other Euclidean axioms and postulates remaining intact. Gauss himself had privately entertained such thoughts as early as 1799. Significantly, Gauss and Lobachevskii (but not Bolyai, who conceived of geometry as a purely logical system based upon axioms) seem to have thought that the truth as to the Euclidean or non-Euclidean character of geometry was an empirical matter to be resolved by certain observations in geodesy or astronomy (for example, stellar parallax). The mathematical and philosophical ramifications of these discoveries were slow to emerge; Bolyai and Lobachevskii had published obscurely and Gauss not at all. But by mid-century, a transformation was underway in the Euclidean conception of geometric knowledge as a unique body of truths concerning space whose certainty was secured by intuition.

2 Towards formal rigour

For nearly two millennia, Euclid's *Elements* was the exemplar of exact reasoning. It is instructive to consider the developments which resulted in overturning this perspective. The first took place between the end of the seventeenth and the end of the eighteenth centuries. For Hobbes, Spinoza and Descartes, reasoning *more geometrico* ('in the geometric manner') attained the deductive ideal, whereby the truth of single propositions inexorably followed from propositions either antecedently established or accepted as fundamental truths. According to Pascal (*c*.1657), the 'art of persuasion' is identified with the method whereby evident axioms are laid down, clear definitions of non-primitive terms are given, and demonstration proceeds in step-by-step fashion. Even Newton, the co-discoverer of the differential and integral calculus, wrote his *Mathematical Principles of Natural Philosophy* (1687) in the axiomatic style of a classical geometric treatise, employing geometric methods of proof where the new methods of the calculus would have been more appropriate.

However, the enormous progress of mathematics in the eighteenth century was almost entirely due to developments in algebra and the new science of analysis, the study of functions and variables pioneered, above all, by Euler. Correspondingly, the guiding ideal of rigour in mathematical proof shifted from the figures and constructions of geometry to the permitted manipulations and transpositions of symbols characteristic of algebra and analysis. Thus Lagrange would famously boast in the beginning of his *Mécanique analytique* (1788) that 'One will not find any figures in this work. The methods that I present require neither constructions, nor geometrical or mechanical reasoning, but only algebraic operations subjected to a regular and uniform development'. Another shift took place at the end of the nineteenth century in tandem with the programme of seeking greater rigour in the calculus, the so-called 'arithmetization' of analysis. Here, the analytic idea of rigour was carried into geometry itself in the demand that deduction not only be independent of diagrams and figures but also of the meaning of geometric concepts. In 1882, Pasch provided an initial account of Euclidean geometry as a hypothetico-deductive system in which the Euclidean definitions of 'point', 'line' and 'plane' were replaced by stipulated unanalysed

relations between these concepts. However, the classic exposition of this point of view is Hilbert's *Grundlagen der Geometrie* (1899) which famously posits 'three different systems of things…points…straight lines…and planes', entities defined only implicitly through relations designated by terms such as 'lie upon', 'between', 'parallel', 'congruent' and 'continuous' whose complete characterization, in turn, lies in those propositions derivable from the chosen axioms. In this abstract conception, there is no longer any trace of the traditional conception of geometry as a theory of *actual* space and its properties. By the early twentieth century, this trend towards axiomatic formalism had completely supplanted the Euclidean paradigm of an exact science.

3 The nature of geometry

According to a once-customary view of mathematics, the two primordial mathematical conceptions are the notions of whole number (as exhibited in counting or ordering) and continuous magnitude or mere extension (as determined in measurement); hence mathematics is ultimately rooted in the sciences of number (arithmetic, algebra, analysis) and the science of space (geometry). However, the conception of geometry as the science of space seems only to have emerged in the late Renaissance, as the modern notion of 'space' arose around 1600 AD (see Space), whereas the science of number has a pedigree reaching back to the speculations of the Pythagoreans. Indeed, the Pythagoreans initially did not clearly distinguish geometry from number. Number was the essential reality, though numbers appear to have been conceived as dots in particular geometric arrangements (thus, for example, the terms 'square number' and 'triangular number', which still survive). The Pythagorean discovery of incommensurable lines (for example, $\sqrt{2}$ which is the diagonal of a unit square) perhaps first effected a separation, for a 'horror of the irrational' prevented a generalization of the concept of number suitable for measurement as well as counting. Hence, irrational magnitudes could only be represented geometrically. As a result, the science of number remained paramount and, in the *Republic*, Plato places the study of geometry after the study of number on the grounds that, unlike the figures of geometry, numbers have no sensible and tangible bodies but are purely objects of thought accessible only to the soul (VII.526). A more compelling reason for the subordination is simply that geometry, in speaking of triangles, quadrilaterals and so on, presupposes number.

The ancient distinction of geometry and the science of number was blurred by the invention of analytic geometry by Fermat (1629) and Descartes (1637). The essential idea here - the introduction of numerical coordinates into the plane - establishes a one-to-one correspondence between plane curves and equations in two variables: to each distinct curve corresponds a definite equation $f(x, y) = 0$. This permits the study of the geometric properties of curves through algebraic or analytic relations connecting their equations, an innovation greatly expanding the variety of curves and figures which can be submitted to precise mathematical treatment. Moreover, coordinates provide a ready means of generalizing from two to three dimensions, and then, as Grassmann (1844) showed, to *n*-dimensional spaces. Thus it is only with the introduction of coordinates that geometry acquires the requisite generality of a 'science of space' (and indeed, more than the requisite generality); ironically, geometry obtains the status of 'science of space' through methods which enable a reduction of geometry to number.

However, the triumph of analytic methods in geometry was carefully circumscribed. Newton, for example, held that equations as 'Expressions of Arithmetical Computation…properly have no place in Geometry', and he celebrated 'the Ancients' over 'the Moderns', in that the former 'never introduced Arithmetical Terms into Geometry' whereas the latter 'by confounding both, have lost the simplicity in which all the Elegancy of Geometry consists' ([1707] 1728: 228; 1967: 120). Newton's intent, it seems, was to insulate only elementary geometry from algebraic methods but not higher geometry, where coordinate methods were actively employed and extended. Other dissenters from the 'triumph of analysis' soon emerged. Most influential were the early nineteenth century French school of projective geometers (L. Carnot, Chasles, Poncelet, among others). Expressing empiricist disdain for the use in calculus and algebra of such 'metaphysical' entities as infinitesimals and negative numbers, they championed instead coordinate-free 'synthetic' methods of proof whereby mathematical certainty was safeguarded by exclusive focus upon a constructed figure (see Poncelet 1822).

At issue here is the status of the diagrams employed in geometric proofs. Were constructed figures merely superfluous, *in concreto* visual aids to clarify bits of abstract reasoning in the text? Or were geometric truths somehow necessarily connected with the quasi-visual evidence provided by representation in 'the mind's eye' of intuition? For Kant, it was clearly the latter. Euclidean geometry characterizes that 'pure form of our sensible intuition' which is space, a condition imposed by the mind upon the experience of the outer sensible world and

thus a condition of the possibility of experience of this world. Furthermore, the ruler and compass constructions of Euclidean geometry are examples *par excellence* of how, for Kant, mathematical knowledge is attained solely through construction of concepts, that is, exhibition a priori of an intuition (for example, a figure such as a triangle) corresponding to its concept which, though particular, is none the less representative of all the universality contained in the concept. A geometric proof is thus 'a chain of inference guided throughout by intuition' (1781/1787: a717; b745).

Mathematical developments in the nineteenth century undermined Kant's conception. First, as noted above, the discovery of consistent (that is, relative to Euclidean geometry) non-Euclidean geometries undercut the prevalent belief in the necessary truth of Euclidean geometry as a characterization of space. Second, the development of projective geometry led to a vastly generalized conception of a geometry as the study of properties of figures invariant under a given group of transformations. Thus, there are as many geometries as there are groups of transformations acting on a space. Third, under the influence of Riemann (see below), physical space came to be seen as a special three-dimensional case of a more general geometric structure. From each of these developments, it followed that the structure of physical space was no longer geometry's sole subject matter. Thus different 'mathematical spaces' came to be distinguished from physical space, and both from a quasi-perceptual and psychological 'space of intuition'. Finally, the trend towards greater rigour in mathematical proof in the latter part of the century denied intuition any intrinsic evidential standing. This in turn promoted axiomatic formalism in geometry, where an explicit recognition that geometric primitives have no pre-axiomatic meaning replaced the traditional view of its primitive terms as indefinable but intuitively secure concepts.

4 Geometry of physical space

Once the conceptual possibility opened that the geometry of physical space is not necessarily Euclidean, the question concerning the structure of this space appeared to be empirically resolvable through measurement. Poincaré (1902) argued, however, that any empirical determination of the geometry of physical space must rest on certain prior conventions concerning metric concepts, such as whether the path of a light ray is considered 'straight'. Others saw a still more fundamental problem concerning what is presupposed by measurement at all. To this, Riemann (1854) and Helmholtz (1868) provided superficially similar but in fact quite different solutions. For both, measurement rests upon the possibility of free mobility of the standard of measurement. However, Riemann interpreted this to mean that magnitudes are independent of their position in what he termed an *n*-fold extended quantity, of which physical space is merely a particular three-dimensional instance (see Space §4). Riemann postulated an analytical expression, a generalization of the Pythagorean theorem to *n* dimensions, which is an invariant measure of the infinitesimal lengths between neighbouring points in such a structure, roughly the modern idea of a real *n*-dimensional differentiable manifold. Given this measure, one can define the arc length of a curve, the angle between intersecting curves, the volume of a region and other geometric concepts. This geometry is locally Euclidean in that it gives a Euclidean metric structure to the tangent space at each point P, hence the measure applies generally to flat and to curved manifolds and also to manifolds with variable (non-constant) curvature. However, in this latter type of manifold, Riemann went on to observe, the magnitude of a solid body is not independent of the position of the body.

Helmholtz, on the other hand, saw the postulate of free mobility as resting upon presumed facts about the behaviour of such bodies *in physical space*, in particular, that there are rigid motions. While acknowledging that the notion of a perfectly rigid body is an idealization to which nothing in the physical world corresponds, Helmholtz none the less saw this concept as a presupposition of measurement in general and hence of the geometry of physical space. As subsequently refined by Reichenbach (1928), who emphasized the impossibility of actually ascertaining whether a physical measuring rod undergoes alterations of length in transport, this view introduces an arbitrary element into measurement in that it makes determinations of the geometry of physical space rest on the adoption of conventions regarding the behaviour of rigid bodies. However, in the generic case of variably curved space-times permitted by Einstein's Theory of General Relativity (1915), the notion of a rigid body is suspect for just the reason suggested 61 years earlier by Riemann: there are no congruences corresponding to the supposed invariant length of a rigid body. In this situation the most consistent procedure, as Weyl pointed out, is to renounce the concept of a transportable standard length. Weyl then showed (1921) how metric properties of the space-time manifold could be mathematically derived from weaker conformal and affine properties having, as their respective empirical correlates, two metric-independent physical processes, the paths of light rays (so-called null geodesics)

and the inertial trajectories of point masses (particles 'too small' to be affected by forces of surrounding physical fields). The essentials of this procedure remain a matter of live interest among general relativists (see Ehlers 1988).

5 But what is geometry?

With the triumph of formal axiomatics, geometry had lost its two most distinctive characteristics: its standing as the theory of physical space and the intuitional trappings attending its traditional reliance on figures and constructions. What then distinguishes geometry from any other formally axiomatized mathematical theory? Surveying the subject in 1932, two eminent geometers wrote:

> The question then arises why the name geometry is given to some mathematical sciences and not to others. It is likely that there is no definite answer to this question, but that a branch of mathematics is called a geometry because the name seems good, on emotional and traditional grounds, to a sufficient number of competent people.
>
> (Veblen and Whitehead 1932: 17)

Yet, as Cartan (1936) quickly pointed out, even in mathematics such a consensus may be difficult to achieve. This is not to say that boundaries are not still drawn. A leading contemporary geometer recently suggested that 'a property is geometrical, if it does not deal directly with numbers or if it happens on a manifold, where the coordinates themselves have no meaning' (Chern 1990: 685). In any case, it must be admitted that geometry has not been completely assimilated into the rest of mathematics, in part, perhaps, because the physical world continues to suggest geometric analogies. Moreover, mathematicians continue to speak of 'geometric intuition' as a fertile source of mathematical invention, even while recognizing that distinctions between geometry and other areas of mathematics remain a matter of taste. And although some mathematicians doubt that 'the psychological aspects of true geometric intuition' will ever be 'cleared up' (Weil 1978), others speculate that the visual-intuitive mode of cognition so characteristic of geometry stems from a lateral specialization of the brain localized in the right hemisphere of the cerebral cortex (Yaglom 1988: 26).

See also: Conventionalism; Einstein, A.; Helmholtz, H. von; Hilbert's programme and formalism; Kant, I.; Logical and mathematical terms, glossary of; Mathematics, foundations of; Poincaré, J.H.; Relativity theory, philosophical significance of; Spacetime; Weyl, H.

T.A. RYCKMAN

References and further reading

Atiyah, M. (1982) 'What is Geometry?', *The Mathematical Gazette* 66: 179-84.(The suggestion, by an eminent geometer, that geometry is that part of mathematics where visual thought is dominant.)

Boyer, C. (1956) *History of Analytic Geometry*, New York: Scripta mathematica.(A comprehensive account of the innovations of Fermat and Descartes.)

Cartan, E. (1936) 'Le rôle de la théorie des groupes de lie dans l'évolution de la géométrie moderne', *Comptes rendus du congrès international des mathématiciens* 1: 92-103.(The case for continuous groups as the unifying concept in geometry.)

Chern, S.-S. (1990) 'What is Geometry?', *American Mathematical Monthly* 97: 679-86.(A survey from Euclid to fibre bundles.)

Coolidge, J. (1940) *A History of Geometrical Methods*, Oxford: Clarendon Press.(Compendium covering Babylonia to Weyl's generalization of Riemannian geometry.)

Daston, L.J. (1986) 'The Physicalist Tradition in Early Nineteenth Century French Geometry', *Studies in the History and Philosophy of Science* 17: 269-95.(An informative essay on the geometric empiricism of an influential school of projective geometers.)

Descartes, R. (1637) *La Géométrie*, Leiden: Maire; trans. D.E. Smith and M.L. Latham, *The Geometry of René Descartes*, with a facsimile of the 1st edn, New York: Dover, 1954.(The first published account of analytic geometry.)

Ehlers, J. (1988) 'Einführung der Raum-Zeit-Struktur mittels Lichtstrahlen und Teilchen' (The Introduction of Space-Time Structure through Light Rays and Particles), in J. Audretsch and K. Mainzer (eds) *Philosophie und Physik der Raum-Zeit*, Mannheim: Wissenschaftsverlag, 145-62.(A discussion of the problem of measurement in General Relativity.)

Euclid (*c*.300 BC) *The Thirteen Books of Euclid's Elements*, trans. T. Heath, Cambridge: Cambridge University

Press, 2nd edn, 1956, 3 vols.(First articulated conception of geometry as a deductive science, based upon axioms, postulates and definitions.)

Fermat, P. de (1629) 'Ad Locos Planos et Solidos Isagoge', in *Oeuvres*, ed. P. Tannery and C. Henry, Paris: Gauthiers-Villars, 1891-1922, vol. 1, 91-110 (Latin); vol. 3, 85-101 (French); partial English translation in D.E. Smith (ed.) *A Source Book in Mathematics*, New York: Dover, 1984, vol. 2, 389- 96.(Fermat's discovery of analytic geometry, published posthumously.)

Grassmann, H. (1844) *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik*, Leipzig: Otto Wigand, 2nd edn, 1862; trans. L.C. Kannenberg, 'The Linear Theory of Extension: A New Branch of Mathematics', in *A New Branch of Mathematics: The 'Ausdehnungslehre' of 1844, and Other Works*, La Salle, IL: Open Court, 1995.(Incipient general formulation of vector analysis in *n*-dimensional space.)

Helmholtz, H. von (1868) 'Über die Tatsachen, die der Geometrie zugrunde liegen', *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen* 9: 193-221; trans. M. Lowe, 'On the Facts Underlying Geometry', in *Epistemological Writings*, ed. P. Hertz and M. Schlick, Dordrecht: Reidel, 1977, 39-71.(The view that measurement presupposes the notion of a rigid body.)

Hilbert, D. (1899) *Grundlagen der Geometrie*, Leipzig: Teubner, 7th edn, 1930; 2nd edn trans. L. Unger and P. Bernays, *Foundations of Geometry*, La Salle, IL: Open Court, 1971.(Influential abstract axiomatization of Euclidean geometry.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1929.(The 'Transcendental Aesthetic' and 'Transcendental Doctrine of Method' sections give Kant's views on geometry.)

Klein, F. (1895) *Vorträge über ausgewahlte Fragen der Elementargeometrie*, Leipzig: Barth; trans. *Famous Problems of Elementary Geometry*, New York: Dover, 1956.(Classic examination of ruler and compass constructions involving no higher mathematics.)

Müller, I. (1981) *Philosophy of Mathematics and Deductive Structure in Euclid's Elements*, Cambridge, MA: MIT Press.(A modern perspective on axiomatic method in Euclid.)

Nagel, E. (1939) 'The Formation of Modern Conceptions of Formal Logic in the Development of Geometry', *Osiris* 7: 142-224; repr. in *Teleology Revisited and Other Essays in the Philosophy and History of Science*, New York: Columbia University Press, 1982, 195-259.(Classic study of the rise of the 'formalistic' conception of geometry.)

Newton, I. (1687) *The Mathematical Principles of Natural Philosophy*, trans. A. Motte (1729) and ed. F. Cajori, Berkeley, CA: University of California Press, 1962, 2 vols.(Its propositions are demonstrated synthetically, 'that the system of the heavens might be founded on good geometry'.)

Newton, I. (1707) *Arithmetica Universalis*, Cambridge: Whiston; trans. Ralphson and Cunn (1722), *Universal Arithmetik: Or, A Treatise of Arithmetical Composition and Resolution*, London: Sener, 2nd very much revised edn, 1728; repr. in *The Mathematical Works of Isaac Newton*, ed. D.T. Whiteside, London and New York: Johnson Reprint, 1967, vol. 2, 1-257.(Criticism of the Cartesian association of algebra and geometry.)

Pascal, B. (*c*.1657) *De l'esprit géométrique*, fragments published posthumously in *Oeuvres complètes*, Paris: Desclée de Brouwer, 1991, vol. 3, 390-428; trans. R. Schofield in *Scientific Treatises*, Chicago, IL, and London: Encyclopaedia Britannica, 1952, 430-46.(The view that 'geometry has all the perfection men are capable of'.)

Plato (*c*.380-367 BC) *Republic*, trans. P. Shorey (1930), London and Cambridge, MA: Loeb Classical Library, Heinemann and Harvard University Press, 1969, 2 vols.(Book VII locates the study of geometry as knowledge of 'the eternally existent' in the curriculum for future guardians of the ideal state.)

Plutarch (*c*.46-120) *Lives*, vol. 5, Cambridge, MA: Harvard University Press (Loeb Classical Library), 1961, 472-3.(Report of Plato's wrath at empirical methods in geometry.)

Poincaré, H. (1902) *La Science et l'hypothèse*, Paris: Flammarion; trans. W. Greenstreet, *Science and Hypothesis*, New York: Dover, 1952, ch. 5.(Classic statement of geometric conventionalism.)

Poncelet, J.V. (1822) *Traité des propriétés projectives des figures (Treatise on the Projective Properties of Figures)*, Paris: Bachelier; trans. selections entitled 'On Projective Geometry', in D.E. Smith (ed.) *A Source Book in Mathematics*, New York: Dover, 1984, vol. 2, 315-23.(Sketch of the significance of coordinate-free projective methods.)

Reichenbach, H. (1928) *Philosophie der Raum-Zeit Lehre*, Berlin: de Gruyter; trans. M. Reichenbach and J. Freund, *The Philosophy of Space and Time*, New York: Dover, 1982.(The case for conventionalism in the context of the General Theory of Relativity.)

Riemann, B. (1854) 'Über die Hypothesen, welche der Geometrie zu Grunde liegen', *Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen* 13: 133-52; trans. with mathematical commentary, 'On the Hypotheses which lie at the Foundations of Geometry', in M. Spivak, *A Comprehensive Introduction to*

Differential Geometry, Houston: Publish or Perish, 2nd edn, 1990, vol. 2, 135-53.(Densely compressed, the deep mathematical and philosophical content of this article repays study.)

Szabó, A. (1964) ‘The Transformation of Mathematics into Deductive Science and the Beginnings of its Foundation on Definitions and Axioms’, *Scripta Mathematica* 28: 27-48A, 113-39.(A scholarly yet speculative account of the Pythagorean origins of mathematics as a demonstrative science.)

Tarski, A. (1959) ‘What is Elementary Geometry?’, in L. Henkin, P. Suppes and A. Tarski (eds) *The Axiomatic Method*, Amsterdam: North Holland, 16-29.(A logical formulation of elementary geometry without a set-theoretic basis.)

Torretti, R. (1978) *Philosophy of Geometry from Riemann to Poincaré*, Dordrecht: Reidel.(A work of high mathematical and philosophical distinction.)

Veblen, O. and Whitehead, J.H.C. (1932) *The Foundations of Differential Geometry*, London: Cambridge University Press.(A concise account reflecting stimulus provided by the theory of General Relativity.)

Weil, A. (1978) ‘S.S. Chern as Geometer and Friend’, in S.-S. Chern, *Selected Papers*, New York and Berlin: Springer, 1989, ix-xii.(Assessment of the place of geometry within mathematics by a leading number theorist.)

Weyl, H. (1921) ‘Zur Infinitesimal Geometrie: Einordnung der projektiven und konformen Auffassung’, *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse*, 99-112; repr. in *Gesammelte Abhandlungen*, Berlin and New York: Springer, 1968, vol. 2, 195-207. (The theorem that conformal and affine properties of a Riemannian manifold determine the metric.)

Yaglom, I.M. (1988) *Felix Klein and Sophus Lie; Evolution of the Idea of Symmetry in the Nineteenth Century*, trans. S. Sossinsky, Boston, MA, and Basle: Birkhäuser.(A readable account of the development of the group concept in geometry.)

George of Trebizond (*c.*1396-*c.*1472)

George was a fifteenth-century humanist important for his work in rhetoric, his translations from the Greek, and his role in the Renaissance Plato-Aristotle controversy. In 1458, as a fierce opponent of Plato and supporter of Aristotle, he transformed what had previously been a quarrel among Byzantines into a major European controversy. He also wrote the first and, for a time, the most popular humanist manual of logic.

Born and brought up in Crete (his Greek name 'Trapezountios' is a patronymic, not a toponymic), George emigrated to Italy around 1416 when twenty years old. Mastering Latin with startling speed, he became a teacher of Latin eloquence in northern Italy. He confirmed his reputation with the publication at Venice in 1433-4 of his book on rhetoric, which was the first full-fledged work on the subject in the classical mode since antiquity. In the late 1430s, at Florence, he published an *Isagoge Dialectica (Introduction to Logic)*, which became a best-seller in the first half of the sixteenth century as humanists attempted to wrest the teaching of logic from the philosophers (see Logic, Renaissance). George himself had no polemical purpose. He merely wished to extract from Aristotle and the manuals of contemporary Aristotelians that small and simplified amount of logic needed by someone primarily interested in rhetoric and literature. Around this time he also joined the papal bureaucracy, becoming eventually a papal secretary. In Florence and (after 1442) in Rome, he continued to teach rhetoric, but he also began to translate from the Greek. His work as a translator peaked under Pope Nicholas V (1447-55), who commissioned him to translate into Latin all of Aristotle's zoological works, the pseudo-Aristotelian *Problemata*, Plato's *Laws*, Ptolemy's *Almagest* and patristic texts (including Eusebius of Caesaria's *Preparation of the Gospels*, which preserves a great many classical philosophical fragments). A quarrel with the pope over his commentary on Ptolemy's *Almagest* drove George to the Neapolitan court of King Alfonso of Aragon in 1452, but he returned to Rome in 1455. Thereafter, he lived in Rome until he died about 1472 (apart from a foray to Venice in 1460-1, where he propounded the theory that the Venetians had realized the ideal mixed constitution of monarchy, aristocracy and democracy).

In the mid-1450s George turned on the Greek Cardinal Bessarion. Though himself a pious Christian, Bessarion had studied with the neo-pagan Byzantine Platonist Pletho. George became convinced that Bessarion and his circle of Greek and Latin humanists were engaged in a Platonic conspiracy to paganize the Latin West. The first result was George's treatise of 1456, the *Protectio Aristotelis Problematum (Protection of Aristotle's Problemata)*, which warned of the attempt of the Bessarion circle (and especially Bessarion's client, the Greek scholar Theodore Gaza) to undermine Latin Aristotelianism and theology by disseminating perverse translations of Aristotle. In 1458 George published (in manuscript form) the first major Latin work in the Plato-Aristotle controversy, the *Comparatio Philosophorum Aristotelis et Platonis (Comparison of the Philosophers Aristotle and Plato)* (printed in 1523 as *Comparationes philosophorum Aristotelis et Platonis*).

The *Comparatio* is divided into three books. In the first, George compares the two philosophers as scholars, and seeks to demonstrate that while Aristotle is the foundation of scientific knowledge, Plato was a mere literary dabbler. In the second, he measures the doctrines of both Aristotle and Plato against the standard of Christian truth. Plato fails miserably because he believed in the transmigration of souls, creation from pre-existent matter, and a hierarchy of gods. Aristotle, on the other hand, taught the immortality of the soul, creation of the world from nothing, and monotheism. George even argues that Aristotle had an inkling of the Christian Trinity. To prove the latter contention, George resorted to the exemplarist doctrine of the medieval Platonic tradition, which found Trinitarian vestiges in all creation. George attributed to Aristotle a hylomorphic conception of the soul and argued that creation *ex nihilo* was a philosophically necessary doctrine. Ironically, both positions reflected the medieval Platonic tradition of the Franciscan-Augustinian school rather than the Christian Aristotelianism of Thomas Aquinas, whom George cited. In Book 3, George revealed his true purpose: to warn the West of the Platonic conspiracy. He showed that Platonists propounded and practised homosexuality. He warned that their doctrines had been the cause of Christian heresies, that Muhammad himself had been trained by a Platonist, and that more recently Pletho had tried to revive heathenism. Now, the Platonists (meaning Bessarion and his circle) stood poised to capture the church. Bessarion answered George in 1469 with his monumental *In calumniatorem Platonis (Against Plato's Calumniator)*, which rehabilitated Plato by making him a semi-Christian Neoplatonist, by arguing that Aristotle was a pagan, and by ignoring George's attack on Pletho's pagan agenda. This phase of the

Plato-Aristotle controversy ended with the deaths of Bessarion and George about 1472.

See also: Aristotelianism, Renaissance; Humanism, Renaissance; Platonism, Renaissance

JOHN MONFASANI

List of works

George of Trebizond [Georgius Trapezuntius] (*c*.1438) *Isagoge Dialectica (Introduction to Logic)*, Cologne: Eucharius, 1539; repr. Frankfurt am Main: Minerva, 1966.(A simplified Aristotelian logic, useful in rhetorical education. George's *Isagoge* went through many early modern editions; the Minerva edition is a photographic reprint of Eucharius.)

George of Trebizond [Georgius Trapezuntius] (1456) *Protectio Aristotelis Problematum (Protection of Aristotle's Problemata)*, in L. Mohler (ed.) *Kardinal Bessarion*, Paderborn: Schöningh, 1923-42; repr. Aalen: Scientia and Paderborn: Schöningh, 1967, vol. 3, 274-342.(A critique of Theodore Gaza's translation of pseudo-Aristotle's *Problemata* and an attack on those who would subvert Aristotelianism.)

George of Trebizond [Georgius Trapezuntius] (1458) *Comparationes Philosophorum Aristotelis et Platonis (Comparisons of the Philosophers Aristotle and Plato)*, Venice: Iacobus Pentius, 1523; repr. Frankfurt am Main: Minerva, 1965.(Condemns Plato as an incompetent philosopher whose latter-day followers subvert Christianity, but praises Aristotle for giving philosophical support to Christian truths and for having intimations of some Christian mysteries. The manuscript (published in 1458) gives the title of this work as *Comparatio philosophorum Aristotelis et Platonis*.)

References and further reading

Monfasani, J. (1976) *George of Trebizond: A Biography and a Study of His Rhetoric and Logic*, Leiden: Brill. (The Plato-Aristotle controversy is treated at length.)

Monfasani, J. (ed.) (1984) *Collectanea Trapezuntiana: Texts, Documents, and Bibliographies of George of Trebizond*, Binghamton, NY: Medieval & Renaissance Texts & Studies.(The Greek texts are accompanied by an English translation.)

Gerard of Cremona (1114-87)

Gerard of Cremona was the most important translator of philosophical works from Arabic to Latin in the twelfth century. During a career of about forty years, he translated at least seventy books. The most famous translations are those of works of Aristotle, including Posterior Analytics, Physics, On the Heavens, On Generation and Corruption and Meteorology 1-3. Gerard also translated a number of works as part of the Aristotelian corpus that were not at all Aristotelian; the most important of these is the so-called Liber de causis (Book of Causes). However, the Aristotelian translations were only a small part of his labour. He translated many more works that were medical, astronomical or mathematical, bringing into Latin several small libraries of fundamental natural science.

What is known of Gerard's life comes principally from three related documents: a list of his translations prepared by his assistants, a later biographical note and a verse eulogy. According to the biographical note, Gerard came to Toledo from Italy in search of Ptolemy's *Almagest*. He stayed to learn Arabic when he discovered just how much in that language was inaccessible to the Latins. So far as modern scholarship can determine, this would have been about 1140. His first translation seems indeed to have been of Ptolemy, since it is only for this work that the list of translations assigns a native collaborator. After the *Almagest*, Gerard is known to have translated at least seventy more works.

Gerard's assistants divided the translated works by subject. The richest subjects are medicine (24 works), mathematics, including optics, weight and dynamics (17 works), and logic or philosophy (14 works). The last group contains some of Gerard's most famous translations, including Aristotle's *Posterior Analytics* (revised from the version of James of Venice, with the commentary of Themistius), *Physics* (with the commentary of al-Farabi), *On the Heavens*, *On Generation and Corruption* and *Meteorology* 1-3. It should be noted that Gerard was concerned to provide commentaries alongside the Aristotelian texts. Some of these commentaries were originally Greek (as in the case of Themistius on the *Posterior Analytics*), though Gerard brought them in from Arabic. Other commentaries and expository treatises had been written in Arabic as part of the Arab peoples' own appropriation of Greek learning (as in the case of al-Farabi's commentary on the *Physics*).

Gerard also translated a number of works as part of the Aristotelian corpus that were not at all Aristotelian. The most important of these is the so-called *Liber de causis (Book of Causes)*, a compilation of material mostly from Proclus (see Liber de causis). This treatise on the cosmic participation of such transcendental features as goodness or unity was read for about a century after Gerard's translation as if it had been written by Aristotle, and guided interpretations of the whole Aristotelian corpus.

Gerard's translations are notable for their fidelity to the Arabic. They preserve its syntax and its lexicon as much possible. Even when Gerard had access to an earlier Latin translation based on the Greek, he would weigh its meaning against what he had gathered from the Arabic, often standardizing the technical terminology to reflect Arabic usage. When the original was in Arabic, Gerard would revise earlier Latin versions and restore passages that had been abridged or omitted. The resulting version does not always make for easy reading, but Gerard seems to have imagined that the texts would not function as isolated discourses so much as parts or stages of a complete curriculum. The influx of Arabic learning into the Latin West was an influx of new texts that mediated organized bodies of knowledge. It was not enough that the texts be put into a passable Latin; they had to be glossed, interpreted, reconciled and corrected in circumstances very different from the circumstances of their Arabic transmission. Gerard meant to help this labour by providing not just single texts, but groups of texts that would serve to introduce bodies of knowledge in a coherent and ordered way. Gerard was not just translating Arabic texts, he was importing Arabic curricula. It would be at least a century before these curricula would be appropriated in any active way by his Latin readers.

See also: Aristotelianism, medieval; Islamic philosophy: transmission into Western Europe; Translators

MARK D. JORDAN

List of works

Gerard of Cremona (before 1187) Revision of James of Venice's translation of Aristotle's *Posterior Analytics*,

ed. L. Minio-Paluello and B. Dod, Aristoteles Latinus 4.3, 2nd edn, Leiden: Brill, 1968.(This Aristotelian work was fundamental for the medieval reconceiving of the nature of knowledge.)

Gerard of Cremona (before 1187) Translation of Aristotle's *Meterology* 1-3, ed. A. Pattin and É. van de Vyver, *Études d'histoire littéraire sur la scolastique médiévale*, Philosophes Médiévaux 8, Louvain and Paris: Publications universitaires, 1964, 241-71.(This Aristotelian work provided medieval Latin readers with the bases for an account of elemental interactions.)

Gerard of Cremona (before 1187) Translation of Themistius' commentary on Aristotle's *Posterior Analytics*, ed. J.R. O'Donnell, *Mediaeval Studies* 20, 1958, 239-315.(Themistius' commentary provided medieval Latin readers with some access to late classical readings of Aristotle.)

Gerard of Cremona (before 1187) Translation of al-Kindi's *De quinque essentiis (On the Five Essences)*, ed. A. Nagy, *Die philosophischen Abhandlungen des Ja'qub ben Ishaq al-Kindi*, Beiträge zur Geschichte der Philosophie des Mittelalters 2.5, Münster: Aschendorff, 1897, 28-40.(This text provided medieval Latin readers with an introduction to Neoplatonic cosmology.)

Gerard of Cremona (before 1187) Revision of Gundissalinus' abridgement of al-Farabi's *Catalogue of Sciences*, ed. A. González Paléncia, Madrid and Granada: Instituto Miguel Asin, 1953.(The *Catalogue* describes a hierarchy of liberal arts and philosophical sciences.)

Gerard of Cremona (before 1187) Revision of an earlier translation of the *Liber de causis (Book of Causes)*, ed. A. Pattin, *Tijdschrift voor Filosofie* 28, 1966, 90-203.(The *Liber de causis*, believed still to be a work of Aristotle, provided a deeply Neoplatonic ontology.)

Anon. (*c*.1187) List of Gerard's Translations, ed. F. Wüstenfeld, 'Die Übersetzungen arabischer Werke in das lateinische seit dem XI. Jahrhunderts,' *Sitzungsberichte der K. Akademie der Wissenschaften in Wien*, Phil.-hist. Kl. 149 (4), 1904, and 151 (1), 1906; trans. M. McVaugh in *A Source Book in Medieval Science*, ed. E. Grant, Cambridge, MA: Harvard University Press, 1974, 35-8.(List of Gerard's translations compiled by his assistants.)

References and further reading

Lemay, R. (1978) 'Gerard of Cremona,' *Dictionary of Scientific Biography* 15 (Supplement 1): 173-92.(The most useful introduction in English; contains an exhaustive list of the translations, with bibliography.)

Gerard of Odo (*c.*1290-*c.*1349)

Gerard of Odo, a scholastic philosopher and theologian who wrote a long commentary on Aristotle's Nicomachean Ethics, is one of many scholastics who attempted to reconcile Aristotle's teachings with the views of Christian authorities. Gerard's work declares the subject of ethics to be the human being as free, makes the will's power of self-determination a necessary condition for moral responsibility, and in other respects reflects the voluntarism commonly found in Franciscan writings of the period.

Gerard of Odo has long been remembered by Church historians as the papal associate who became minister general of the Franciscan order in 1329. At the time, the Franciscans were badly divided about the kind of life required by their vow of poverty. Pope John XXII deposed the head of the order and nominated Gerard, his own candidate, as the new minister general. Gerard continued in that position until 1342, when he was named patriarch of Antioch and bishop of Catania.

Gerard's accomplishments as a bachelor and master of theology at Paris were better received than his performance as an administrator. His writings include a commentary on the *Sentences* of Peter Lombard, two treatises on logic - one on syllogisms, the other on supposition (*De suppositionibus*) - as well as a treatise on economics (*De contractibus*) and a long commentary on the *Nicomachean Ethics*. His discussion of continua, once thought to be a fragment of some lost work on Aristotle's *Physics*, comes from his *Sentences* commentary.

Gerard's most important contribution by far was his commentary on Aristotle's *Ethics*, which might explain how he came to be called 'the Moral Doctor.' The better-known *Ethics* commentary by the nominalist John Buridan borrows extensively from Gerard's work, which favours a more realist, Neoplatonist metaphysics. Despite the seemingly prohibitive metaphysical differences, Buridan plainly found the work of the Moral Doctor well worth his attention (Walsh 1975).

Gerard's commentary on the *Ethics*, probably the first such work by a Franciscan, signals its emphasis on freedom from the outset. The prologue declares the subject of ethics to be the human being as *free*. Though this opens the possibility that ethics extends to angels, Gerard argues that angelic freedom differs significantly from human freedom, and since we know little about angels in any case, we apply ethics only to mankind. Nevertheless, he adds, ethics could apply to other creatures if any possessed a freedom similar to our own.

Gerard sees the will's power of self-determination as the source of freedom and a necessary condition for moral responsibility. Among the powers of the soul, the will holds the position of king, the intellect only of king's counsellor. Because the will can act against the advice of the intellect, and the intellect cannot influence the emotions without the concurrence of the will, moral weakness arises from weakness of will. In a similar vein, Gerard opposes Aquinas in insisting that all moral virtues must belong to the will. Aquinas' commentary on the *Ethics* claims that temperance and courage have to do mainly with how someone is internally affected by the passions, whereas justice has mainly to do with how one acts externally; thus the former are virtues of the emotional part of the soul, while the latter is a virtue of the will. Quoting Aquinas with strong disapproval, Gerard argues that *all* moral virtues moderate passion but nonetheless are principally concerned with action. For him, all moral virtues are habits of choice: hence habits of the will, since choice is an act of the will (see Free will).

On these and many other points, Gerard's commentary reflects the voluntarism so often found in Franciscan writings of the late thirteenth and early fourteenth centuries. His work attempts to interpret Aristotle's *Ethics* in such a way that it supports, or at least does not contradict, teachings prevalent in his own order (Kent 1984).

Gerard's *Sentences* commentary placed him among the 'indivisibilists' in the scholastic debate about continuous magnitudes. He argues that a continuum is composed of extensionless points, distinct from mathematical points in admitting relative differences in time and space (Zoubov 1959). Here, as in the metaphysics of his *Ethics* commentary, Gerard diverges sharply from Aristotle's teachings. The arguments against indivisibles in Book VI of the *Physics* allow for no genuine doubt about Aristotle's views.

See also: Aristotelianism, medieval

BONNIE KENT

List of works

Gerard of Odo (*c.* 1320-9) *Expositio in Aristotelis Ethicam (Commentary on Aristotle's Nicomachean Ethics)*, Venice: De Lovere, 1500.(Gerard's best known work, which influenced John Buridan.)

Gerard of Odo (*c*.1320-9) *De suppositionibus (On Supposition)*, ed. S. Brown, 'Gerard Odon's 'De Suppositionibus'', *Franciscan Studies* 35, 1975, 5-44.(Logical treatise, on suppositions.)

Gerard of Odo (*c*.1326-8) *In quattuor libros Sententiarum*, MS Sarnano E. 98; MS Paris, BN lat. 3068.(No edition of this work exists. Other useful manuscripts can be found in the national library in Naples and the cathedral library in Valencia.)

Gerard of Odo (??) *De contractibus*, MS Siena Bcom. U.V.8.(Discussed in Langholm (1979), chapter 6, this work is quoted extensively and with approval by St Bernardino of Siena.)

References and further reading

Kent, B. (1984) *Aristotle and the Franciscans: Gerald Odonis' Commentary on the "Nicomachean Ethics"*, Ph.D. dissertation, Columbia University.(Places Gerard's *Ethics* commentary in the context of late thirteenth- and early fourteenth-century Franciscan psychology and ethics.)

Langholm, O. (1979) *Price and Value in the Aristotelian Tradition*, Bergen: Universitetsforlaget.(Discusses Gerard's *De contractibus* and other scholastic works on economics.)

Murdoch, J.E. (1982) 'Infinity and continuity,' in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(Explains fourteenth-century debate regarding indivisibles.)

Saarinen, R. (1994) *Weakness of the Will in Medieval Thought: From Augustine to Buridan*, Leiden: Brill. (Exegesis of teachings by Gerard and other medieval thinkers on the problem of *akrasia*.)

Walsh, J.J. (1975) 'Some Relationships Between Gerald Odo's and John Buridan's Commentaries on Aristotle's 'Ethics',' *Franciscan Studies* 35: 237-75.(Analysis of Buridan's borrowings from Gerard's commentary.)

Zoubov, V. (1959) 'Walter Catton, Gérard d'Odon et Nicolas Bonet,' *Physis* 1: 261-78. (Gerard's arguments regarding the composition of continua and his influence on other thinkers.)

Gerard, Alexander (1728-95)

Alexander Gerard was Professor of Moral Philosophy and Logic (1752) and Professor of Divinity (1759) at Marischal College, and Professor of Divinity (1773) at King's College, Aberdeen. A leading member of the Aberdeen Philosophical Society, he wrote a new plan of education for Marischal College as well as works on divinity. He is best known, however, for his Essay on Taste (1759). In 1774, he returned to the subject with An Essay on Genius. Gerard was associated with Thomas Reid (1710-96) in the Philosophical Society until Reid's transfer to Glasgow in 1764. The work of David Hume (1711-76) was a principal influence.

Alexander Gerard's aesthetic philosophy is a response to the theories of Francis Hutcheson (1694-1746) and David Hume (see Hume, D. §3). Like Hutcheson, Gerard made use of the concepts of an internal sense and of association to explain the subjectivity of taste. However, Gerard's use of an internal sense, which he called a reflex act, must be severely qualified. He distinguished two ways of considering taste:

> It may be considered either as a species of *sensation*, or as a species of *discernment*…. Taste considered in the former of these lights, in respect of what we may call its *direct exercise*, cannot properly admit any standard… But notwithstanding this, there may be a standard of taste in respect of its *reflex acts*: and it is only in respect of these, that a standard should be sought for.
>
> (1759: 214-6)

This distinction distances Gerard from Hume's more radical reliance on sentiment and Hutcheson's elaborate theory of an internal sense. Gerard depended on a species of induction (perhaps derived from another Aberdeen philosopher, George Turnbull) to correct immediate sensation.

Gerard's own theory rests on a view of the imagination as exercising the mind. Freed of the need for the actual presence of objective sources, which limit the external senses, the imagination combines reflective ideas supplied by fancy. The operation of the imagination exercises the mind, and when that exercise falls within a moderate range, it is experienced as pleasurable. If it is either too languid and easy or too excited and difficult, discomfort (or indifference) results. From the beginning, these were the controlling principles of Gerard's discussion. But the terms of the problem set for him (and in some sense his essay was a set piece for the Edinburgh Society for the Encouragement of Arts, Sciences, Manufactures and Agriculture) come from Hutcheson's internal sense theory and Hume's search for a standard of taste. So Gerard used 'taste' as a vehicle to explain the faculty of the imagination. He was not always consistent, and sometimes it seems that he forgot that the problem of taste arose because an internal sense would lack the checks and confirmations of an external sense.

Gerard tended to multiply senses because they were little more to him than aesthetic predicates. Association establishes the reference for such predicates. Thus, while continuing to use the language of taste and internal sense, he was moving decisively in the direction of Archibald Alison's more elaborate scheme of imaginative and associative aesthetics (see Alison, A.). At the same time, Gerard remains within the scope set by Reid and his school.

See also: Aberdeen Philosophical Society; Emotion in response to art; Enlightenment, Scottish; Artistic taste

DABNEY TOWNSEND

List of works

Gerard, A. (1759) *An Essay on Taste*, London: Miller; 3rd edn repr. Gainsville, FL: Scholars' Facsimiles and Reprints, 1963.(Gerard's prize essay for The Edinburgh Society for the Encouragement of the Arts, Sciences, Manufactures and Agriculture. The Essay on Taste sifts internal sense toward a theory of imagination and association. Includes a useful introduction by W.J. Hipple, Jr.)

(1774) *An Essay on Genius*, London: Strahan; repr. ed. B. Fabian, Munich: Fink.(Less widely read because it fails to break new ground. Continues Gerard's emphasis on imagination and association as the leading concepts of an aesthetic psychology.)

References and further reading

Dickie, G. (1996) *The Century of Taste*, Oxford: Oxford University Press.(A defence of Hume against the forms of association advocated by Alison and Gerard.)

Kivy, P. (1976) *The Seventh Sense*, New York: Burt Franklin.(Traces the concept of internal sense through eighteenth-century aesthetics.)

McCosh, J. (1875) *The Scottish Philosophy*, New York: Carter, 1911-2; repr. Bristol: Thoemmes, 1990.(Brief biographical sketch and commentary.)

Stulmitz, J. (1961) ‘Beauty: Some Stages in the History of Ideas’, *Journal of the History of Ideas* 22: 185-204. (Locates Gerard as the mediator between the internal sense theory of Hutcheson and the associationism of A. Alison.)

Townsend, D. (1987) ’From Shaftesbury to Kant’, *Journal of the History of Ideas* 48: 287-305.(Traces the development of the empiricist tradition, including Gerard. Argues that activity of the mind is more important than association or sense to Gerard.)

Gerbert of Aurillac (938-1003)

Gerbert is chiefly remembered as an educational reformer. He established a syllabus for the university course in logic, the logica vetus*, that remained in use until the mid-twelfth century. Most of his academic writings are instructional works on mathematics. In his single philosophical work, De rationali et ratione uti (On That Which is Rational and Using Reason), he uses Boethius' logical commentaries to develop a distinctly Platonic solution to a problem he derives from Porphyry's Isagōgē.*

Gerbert began his studies as a Cluniac monk at Aurillac. It is generally supposed, although documentation for this period is not good, that he was exposed to Islamic science and mathematics during a subsequent stay in Catalonia. He moved to Rheims around 972 to continue his studies, and later became director of schools there. After a brief and stormy tenure as Abbot of Bobbio, *circa* 980-3, he returned to Rheims, where he continued to engage in political controversy. In 991 he was made archbishop, and in 998 was transferred to Ravenna. He was elected as Pope Sylvester II in 999, and died in Rome in 1003.

During Gerbert's first stay at Rheims, he instigated educational reforms which were to influence the curriculum of cathedral schools and universities for at least the following two centuries. Gerbert is the first person documented to have taught the syllabus of the *logica vetus*, the basis for logic teaching until the mid-twelfth century. He used Porphyry's *Isagōgē* in the translations of Marius Victorinus and Boethius, Aristotle's *Categories* and *De interpretatione* in Boethius' versions, Cicero's *Topics*, Boethius' commentaries on these works and Boethius' own logical works (see Aristotle; Cicero; Porphyry). In addition to studying these works, his students were required to engage in 'controversies' on set questions (see also Language, medieval theories of; Logic, medieval).

In teaching astronomy, Gerbert used a spherical model to illustrate the movement of the planets and had his students make astronomical observations. Mathematics was an area of particular interest for Gerbert; several of his letters include mathematical discussions in response to questions from his students, and he also wrote longer mathematical treatises. He taught arithmetic using the arabic numerals from one to nine rather than Roman numerals (though he did not adopt the arabic zero), and instructed his students in the use of the abacus. His treatise on the use of the abacus later became a standard work. For geometry, he supplemented a fragmentary version of Euclid with his own treatise. Gerbert is also known to have reintroduced the study of classical authors including Virgil, Terence, Horace, Persius, Lucan, Seneca and Quintilian. Fulbert of Rheims, founder of the school of Chartres, was one of his students (see Chartres, School of).

Gerbert's one extant philosophical work, *De rationali et ratione uti (On That Which is Rational and Using Reason)* has been examined by Marenbon (1983). It shows Gerbert using Porphyry's *Isagōgē*, Aristotle's *De interpretatione* and Boethius' commentaries on the same works to explain a problem arising from Porphyry's discussion of differentia: why *ratione uti* can be predicated of *rationale*. As Marenbon explains, Gerbert uses the opposition between act and potency (discussed in both the *Isagōgē* and the *De interpretatione*), Aristotle's remarks on psychology at the beginning of the *De interpretatione* and a metaphysical hierarchy from the beginning of Boethius's first *Isagōgē* commentary in developing his solution. Gerbert claims that when *rationale* is considered just as a concept, apart from any instance in a human being, then its scope is no narrower than that of *ratione uti*. It is only when the concept *rationale* is attached to a corporeal human being that it is no longer always in act, but becomes a potentiality realized in act only when the person who is rational is indeed using reason. When *rationale* is attached to a corporeal being, then, it is acceptable to predicate *ratione uti* of it. When we say 'that which is rational uses reason', it is understood that we are referring to some particular rational being which is currently using its reason, not to 'that which is rational' taken in a general sense.

A philosophical debate between Gerbert and Otric held before Emperor Otto II is recounted by Gerbert's student Richer. While the word-for-word dialogue Richer includes is most likely of his own composition, the event may well have occurred. According to Richer, the debate ranged over the proper division of philosophy, the cause of the creation of the world, whether all causes can be expressed by the single word 'good', the cause of a shadow and whether 'mortal' or 'rational' is the wider concept. Whether or not Richer's account faithfully reports Gerbert's replies, it remains a testimony to his effective teaching.

See also: Chartres, School of; Logic, medieval; Natural philosophy, medieval (§1)

FIONA SOMERSET

List of works

Gerbert of Aurillac (*c*.972-99) *Opera (Works)*, ed. A. Duchesne in J.-P. Migne, *Patrologia Latina* 139: 201-64; revised edn, ed. A. Olleris, *Oeuvres de Gerbert, pape sous le nom de Sylvestre ii*, Clermont-Ferrand: Thibaud and Paris: Dumoulon, 1867.(Collected works of Gerbert.)

Gerbert of Aurillac (*c*.972-99) *Gerberti opera mathematica (Mathematical Works of Gerbert)*, ed. N. Bubnov, Berlin: Berloni (R. Friedlander & Sohn), 1889.(Includes the mathematical letters as well as the treatises.)

Gerbert of Aurillac (983-97) Letters, ed. J.V. Havet, *Lettres de Gerbert, 983-97*, Paris: A. Picard, 1889; re-edited and trans. H.P. Lattin, *The Letters of Gerbert with his Papal Privileges as Sylvester II*, Records of Civilization, Sources and Studies 60, New York: Columbia University Press, 1961. (Mathematical discussions and requests for books are included.)

Gerbert of Aurillac (*c*.972-99) *De rationale et ratione uti (On That Which is Rational and Using Reason)*, ed. A. Olleris, *Oeuvres de Gerbert, pape sous le nom de Sylvestre ii*, Clermont-Ferrand: Thibaud and Paris: Dumoulon, 1867, 297-310.(A short treatise on why ‘using reason’ can be predicated of ‘that which is rational’.)

References and further reading

Amann, E. (1941) ‘Silvestre II’, in *Dictionnaire de théologie catholique*, Paris: Letouzey et Ane, vol. 14 (2), cols 2075-83.(Still provides the most extensive bibliography of work on Gerbert.)

Marenbon, J. (1983) *Early Medieval Philosophy (480-1150)*, London: Routledge & Kegan Paul, 82-3.(A clear, accessible exposition of *De rationali et ratione uti*.)

Lindgren, U. (1976) *Gerbert von Aurillac und das Quadrivium (Gerbert of Aurillac and the Quadrivium)*, Wiesbaden: Steiner.(Discusses Gerbert’s mathematical and scientific interests and teaching.)

Richer (*c*.980-1000) *Histoire de France 888-995*, ed. R Latouche, Paris: H. Champion, 1937, vol. 2, 54-64.(This historian, a student of Gerbert’s, recounts Gerbert’s career and gives details of his teaching and of his debate with Otric before emperor Otto II.)

Gerdil, Giancinto Sigismondo (1718-1802)

A lifelong member of the Barnabite religious order, Gerdil became well-known as the most eminent Italian disciple of Malebranche (and critic of Locke); in 1764 he published a critique of Émile (1762) which Rousseau himself called the only attack worth reading in its entirety. Only extreme old age kept Cardinal Gerdil from being elected Pope at Venice in 1798.

Born in Savoy, Gerdil became a lifelong member of the Barnabite religious order and, as a Cardinal in 1798, was kept from being elected Pope at Venice only by extreme old age. He is best-known, however, as the most eminent Italian disciple of Malebranche (1638-1715) and a critic of Locke and Rousseau.

As a follower of Malebranche, Gerdil defended the principal doctrines of the celebrated Oratorian Father: that God governs the universe through consent, uniform *volontés générales* (general wills), not through *ad hoc* miraculous *volontés particulières* (particular wills); that human beings are only the ‘occasional’ causes of their own acts (while God is the true cause); that ‘we see all things in God’, not in Lockean sense-perception. Gerdil was also distressed by the Lockean notion that the natural immortality of the soul is only highly ‘probable’, that only revelation (but not reason) can offer certainty on this point. Here Gerdil’s key works are *The Immateriality of the Soul, Demonstrated against Locke* and *Defence of the Sentiment of Father Malebranche against this Philosopher* (both published 1747-8).

But it was as a critic of Rousseau that Gerdil gained a European reputation. In his *Réflexions* on Rousseau’s political, religious and pedagogical theory (1764), Gerdil urged that ‘the purpose of *The Social Contract* is the universal overturning of the civil order; the purpose of *Émile* (1762) is to prepare minds for this through a total revolution in ways of thinking’. Granting Rousseau’s eloquence and ‘brilliance of colouring’ as a writer, Gerdil insisted that Jean-Jacques’ clinging to (supposedly uncorrupted) ‘nature’ would produce ‘bad Christians and bad citizens’. Rousseau’s supposed ‘natural man’, for Gerdil, is ‘the most factitious [being] that has ever existed in the mind of any philosopher’: ‘no one today has seen a man detached from every social institution, no one can say what he is’. Despite the ‘persuasiveness’ of Rousseau’s ‘blinding’ eloquence, in Gerdil’s view, the citizen of Geneva’s thought amounts to:

> contempt for all revealed religion, and for Christianity in particular…; a revolt against all legitimate authority: … a false indulgence in not reprimanding the faults arising from children’s natural liberty; a false determination not to reason with them, and not to cultivate their minds by any of the studies suitable to their age - such are the fruits of this new plan for education.

(*Réflexions*)

If Rousseau alone is right, Gerdil adds tartly, then Socrates, Plato, Cicero, Seneca, Plutarch, Bacon, Bossuet and Fénelon knew nothing about human nature or suitable education. And he ends the *Réflexions* with a paragraph whose literary effectiveness Rousseau himself acknowledged:

> ‘M. Rousseau is opening a new career. Men will no longer be depraved by arbitrary institutions, they will no longer be terrified by the importunate threats of religion; they will no longer be fatigued by studies which are so remote from their nature. What chimeras! What visions!’

PATRICK RILEY

List of works

Gerdil, G.S. (1806-21) *Opere (Works)*, Rome, 20 vols.

(1747-8) *L’immatérialité de l’âme, démontrée contre Locke (The Immateriality of the Soul, Demonstrated against Locke)*, together with *Défense du sentiment du P. Malebranche, contre ce philosophe (Defence of the Sentiment of Father Malebranche against this Philosopher)*, Turin.

(1764) *Réflexions sur la théorie et la pratique de l’éducation, contre les principes de J.J. Rousseau (Reflections on the theory and practice of education, against the principles of J.-J. Rousseau)*, Paris.

References and further reading

Biographie universelle (1816) ‘Gerdil’, Paris, vol. 17.
Geymonat, L. (ed.) (1971) *Storia del pensiero filosofico e scientifico (History of Philosophical and Scientific Thought)*, vol. 3: 486-8, Milan.

German idealism

From the late eighteenth century until the middle of the nineteenth, German philosophy was dominated by the movement known as German idealism, which began as an attempt to complete Kant's revolutionary project: the derivation of the principles of knowledge and ethics from the spontaneity and autonomy of mind or spirit. However, German idealists produced systems whose relation to Kant is controversial, due to their emphasis on the absolute unity and historical development of reason.

As a movement to complete Kant's project, German idealism was punctuated by controversies about whether certain Kantian distinctions constitute *dualisms* - unbridgeable gaps between elements whose underlying unity must be demonstrated - and about how such dualisms can be overcome (see Kant, I. §3). One controversy concerned the distinction between the form of knowable objects - contributed by the mind, according to Kant - and the *matter* of sensation - contributed by mind-independent things-in-themselves. Jacobi objected that things-in-themselves lay beyond the boundaries of human knowledge, so Kant should profess 'transcendental ignorance' about the origin of sensible matter, leaving open the possibility that reality is mind-dependent. Another controversy concerned Kant's distinction between the spatio-temporal forms of sensibility and the categorial forms of understanding. Maimon (1790) argued that unless the underlying unity of this distinction were demonstrated, the applicability of categories to sensible objects could not be demonstrated against sceptics like Hume. Instead of defeating scepticism as he intended, some thought Kant had ensured its triumph by establishing unbridgeable dualisms between mind and reality, and between understanding and sensibility.

In the 1790s, some Kantians - notably Reinhold, Fichte and Schelling - sought to complete Kant's project through systematization. Troublesome dualisms would be overcome by positing distinct mental forms, as well as the distinction between mind and reality, as necessary conditions of the mind's free and unitary activity, and thus as necessary elements of a unified system (see Fichte, J.G. §3; Schelling, F.W.J. von §1). However, other Kantians - notably Buhle - accused the systematizers of undermining the distinctness of form and matter, and of attempting to generate matter from pure form. Meanwhile the systematizers, seeking to defeat scepticism, claimed metaphysical knowledge grounded in intellectual intuition of the mind's spontaneity (see Fichte, J.G. §5). But Kant had explicitly denied that humans could attain such knowledge.

Professing continued allegiance to Kant despite these apparent departures, some systematizers - notably Fichte - claimed that Kant's teaching was only intelligible from a special standpoint and that, having attained that standpoint, they were expressing Kant's spirit, if not his letter. However, in his 1799 *Open Letter on Fichte's Wissenschaftslehre*, Kant publicly repudiated all attempts to discern his philosophy's spirit from a special standpoint and rejected any endeavour to bridge the gap between form and matter. But those who were repudiated did not change their ways. Finding Kant unable to complete the revolution he had started, they henceforth constructed their systems more independently of Kant's writings. The influence of pre-Kantian philosophers, notably Spinoza, was explicitly acknowledged.

In order to overcome Kantian dualisms without ignoring his distinctions, Idealists produced a variety of developmental monisms (see Fichte, J.G. §3; Schelling, F.W.J. von §2; Hegel, G.W.F. §4). Such systems portray a single, developing principle expressing itself in dualisms whose unstable, conflictual nature necessitates further developments. Thus reality is a developing, organic whole whose principle can be grasped and whose unity can be articulated in a philosophical system. But the dualisms encountered in everyday experience are not illusory. Rather, they are necessary stages in reality's development towards its full realization. This conception of development is often called dialectic.

Developmental monism emphasized the sociality and historicity of reason. Fichte was the first to emphasize sociality, arguing that the development of individual self-consciousness required consciousness of another mind, and deriving a theory of justice from the idea of one individual *recognizing* another as such (see Fichte, J.G. §7). Hegel placed particular emphasis on historicity, portraying human history as a series of conflicts and resolutions culminating in a just society that would enable the reciprocal recognition of individuals, as well as the perfect self-recognition of reason at which philosophy had always aimed (see Hegel, G.W.F. §8). Thus history - especially the history of philosophy - acquired unprecedented significance as the narrative of the mind's ascent to

self-knowledge. And it was hoped that a philosophical account of society's historical development would correct the deficiencies of the French revolution, which was often called the political equivalent of Kant's revolution.

Idealists disagreed about whether Kant's distinction between mind and nature was another problematic dualism. Schelling and Hegel argued that a systematic philosophy must portray nature as the mind's preconscious development. But Fichte regarded their philosophy of nature as a betrayal of Idealism that explained the mind in nonmental terms and deprived the mind of its autonomy. By 1801 the disagreement was explicit (see Fichte, J.G. §6; Schelling, F.W.J. von §§1, 2; Hegel, G.W.F. §§3, 7).

Controversy about another putative dualism - between concept and intuition - ended the alliance between Schelling and Hegel. Without naming him, Hegel's *Phenomenology of Spirit* (1807) appeared to criticize Schelling's view that philosophy can only be understood by those innately able to intuit - to grasp non-discursively - the identity implicit in apparent dualisms. Hegel argued that the completed philosophical system must be conceptualized and rendered discursively intelligible to everyone. However, only those who transformed their accustomed ways of thinking could understand the system (see Hegel, G.W.F. §§5, 6). So Hegel undertook to guide his readers through a series of transformations of consciousness representing the history of human thought, as well as the education of the individual. The Napoleonic wars forced Hegel from university life, but after his return in 1814, and especially after his move to Berlin in 1818, his version of idealism - with its portrayal of reason developing in both nature and culture towards conceptual articulation - became dominant.

However, Hegel died in 1831 and Schelling raised influential criticisms of Hegel when he began teaching in Berlin in 1841. In an inaugural lecture before an audience including Engels and Kierkegaard, Schelling argued that Hegel's system was an inevitably failed attempt to overcome the dualism between conceptual thought and intuited existence. Schelling's criticism was seminal for Marxism and existentialism and was more influential than his alternative proposals, which he had been developing under the influence of theosophy since 1809 (see Existentialism; Marxism, Western; Schelling, F.W.J. von §§3, 4; Hegelianism §3). The relationship between thought and existence remains problematic for post-idealist philosophy, and German idealism remains both an object of criticism and a source of insight.

See also: Absolute, the; Idealism; Romanticism, German; Rosenzweig, F.; Buddhism, Yogācāra school of §2

PAUL FRANKS

List of works

Collected works, critical editions and English translations can be found in the entries on individual philosopher throughout the *Encylopedia*. Below are listed some texts of particular importance for the historical development of German idealism.

Behler, E. (ed.) (1985) *Fichte, Jacobi and Schelling: Philosophy of German Idealism*, New York: Continuum. (Contains translations of Fichte's Sun-Clear Report and Jacobi's Letter to Fichte, both relevant to the controversy about whether Idealism has atheistic and nihilistic implications, as well as important writings by Schelling on aesthetics, the philosophy of nature and the difficulties facing any philosophical account of the freedom to do evil.)

Giovanni, G. di and Harris, H. (eds) (1985) *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism*, Albany, NY: State University of New York Press.(Contains translated and annotated extracts from important texts by Reinhold, Aenesidemus, Maimon and Beck that were pivotal in the debate between sceptical and systematic interpretations of Kant, as well as articles from the *Critical Journal* of Hegel and Schelling, dating from their early alliance.)

Hegel, G.W.F. (1807) *Phänomenologie des Geistes (Phenomenology of Spirit)*, trans. A.V. Miller, Oxford: Clarendon Press, 1977.(Hegel's early attempt to complete Kant's revolution by leading the reader through a series of dualistic forms of consciousness, representing the history of human thought, until the reader reaches the standpoint of absolute knowledge by seeing the dialectical development of each form into the next, and the dialectical development of the entire series into knowledge free of dualism.)

Hegel, G.W.F., Schelling, F.W. J. von, Hölderlin, J.C.F. (1796) *Das älteste Systemprogram; Studien zur Frühgeschichte des deutschen Idealismus*, ed. R. Bubner, Bonn: Bouvier, 1973; trans. as *The Oldest System Programme of German Idealism*, in E. Behler (ed.) *Fichte, Jacobi and Schelling: Philosophy of German Idealism*, New York: Continuum, 1985.(Written in Hegel's hand but of disputed authorship, this provides a remarkable glimpse of the origins of German idealism; for interpretation and reception-history see F.-P.

Hansen's *'Das älteste Systemprogram': Rezeptionsgeschichte und Interpretation*, Berlin and New York: de Gruyter, 1989. The Bubner volume contains useful German articles on the content and authorship of the manuscript.)

Jacobi, F. (1785-1815) *The Main Philosophical Writings and the Novel Allwill*, trans. G. di Giovanni, Montreal and Kingston, London, Buffalo, NY: McGill-Queen's University Press, 1994.(Collection of important texts, with an introduction, notes and bibliography. Although Jacobi was not, strictly speaking, a German idealist, many of his ideas greatly influenced Idealism. For example, the Idealists should be understood as accepting his criticism of Kant's conception of the thing-in-itself, and as inspired by the Spinozism he brought to their attention, although Jacobi himself publicized Spinozism only in order to reject it.)

Kant, I. (1799) 'Erklärung in Beziehung auf Fichtes Wissenschaftslehre' (Open Letter on Fichte's Wissenschaftslehre), in *Kant: Philosophical Correspondence, 1759-99*, ed. and trans. A. Zweig, Chicago, IL: University of Chicago Press, 1967, 253-4.(Kant publicly repudiates Fichte and Beck for claiming that the correct interpretation of his writings would be understood only from a special standpoint, according to the spirit and not the letter. Implicitly repudiated are Schelling, Hegel and all others who distinguish, in their own ways, between the letter and the spirit of Kant's philosophy.)

Maimon, S. (1790) 'Versuch über die Transcendentalphilosophie mit einem Anhang über die symbolische Erkenntnis und Anmerkungen' (Essay on Transcendental Philosophy with an Appendix on Symbolic Cognition and Annotations), in S. Maimon, *Gesammelte Werke*, ed. V. Verra, Heidelberg: G. Olms, 1965-75, vol. 2.(An influential work that presses the case against various Kantian dualisms with great clarity, and proposes a monistic solution whereby the dualisms afflicting the finite, human mind represent an imperfect yet necessary stage on the way towards the realization of the infinite, divine mind.)

Simpson, D. (ed.) (1984) *German Aesthetic and Literary Criticism: Kant, Fichte, Schelling, Schopenhauer, Hegel*, Cambridge: Cambridge University Press.(Contains texts relevant to the philosophy of art and literature.)

References and further reading

Beiser, F. (1987) *The Fate of Reason: German Philosophy from Kant to Fichte*, Cambridge, MA: Harvard University Press.(Indispensable account of the immediate reaction to Kant's revolution.)

Beiser, F. (1992) *Enlightenment, Revolution and Romanticism: The Formation of Modern German Political Thought, 1792-1800*, Cambridge, MA: Harvard University Press. (Important account of the political dimension of early Idealism.)

Breazeale, D. (1981) 'Fichte's *Aenesidemus* Review and the Transformation of German Idealism', *Review of Metaphysics* 34: 545-68.(Clear, accessible account of the important role of sceptical challenges to Kant and Reinhold in the development of Fichte's thought.)

Breazeale, D. (1982) 'Between Kant and Fichte: Karl Leonhard Reinhold's "Elementary Philosophy"'', *Review of Metaphysics* 35: 785-822.(Clear, accessible account of this neglected philosopher's important contribution to early German idealism: the idea of completing Kant's revolution through systematization.)

Cassirer, E. (1917) *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit (The Problem of Knowledge in the Philosophy and Science of the Modern Age)*, vol. 2, Darmstadt: Wissenschaftliche Buchgesellschaft.(Classic study of the development from Kant to Hegel, challenging the traditional view of that development as leading inexorably to Hegel; a useful corrective to Kroner.)

Henrich, D. (1997) *The Course of Remembrance and Other Essays on Hölderlin*, ed E. Förster, Stanford, CA: Stanford University Press.(Important collection of essays by a leading German researcher, reporting groundbreaking work on the origins of German idealism at the University of Jena, and focusing on some neglected but influential figures.)

Kroner, R. (1921) *Von Kant bis Hegel (From Kant to Hegel)*, Tübingen: JCB Mohr, 3rd edn.(The classic account of German idealism as a dialectic culminating in Hegel.)

Nauern, F.-G. (1971) *Revolution, Idealism and Human Freedom: Schelling, Hölderlin and Hegel and the Crisis of Early German Idealism*, The Hague: Martinus Nijhoff. (Accessible account of the early thought of three important figures who studied together at the Tübingen theological seminary at the time of the French revolution.)

Royce, J. (1919) *Lectures on modern idealism*, foreword by J.E. Smith, New Haven, CT: Yale University Press, 1964.(A classic, if rather outdated, account of German idealism by a major US philosopher of the nineteenth and early twentieth centuries.)

Solomon, R. (1981) *Introducing the German idealists: mock interviews with Kant, Hegel, Fichte, Schelling, Reinhold, Jacobi, Schlegel, and a letter from Schopenhauer*, Indianapolis, IN: Hackett Publishing Company. (Accessible introduction to some important figures in German idealism.)

Solomon, R. and Higgins, K. (1993) *The Age of German Idealism*, London and New York: Routledge.(Accessible and useful collection of articles on individual thinkers, well suited to beginners.)
Taylor, C. (1979) *Hegel and Modern Society*, Cambridge: Cambridge University Press.(Illuminating and accessible account of the cultural and social background against which German idealism developed.)

Gerson, Jean (1363-1429)

Gerson was one of the leading theologians of the via moderna*, the 'modern way' of nominalism. A fervent critic of the 'formalists' of the* via antiqua*, Gerson stood in the Ockhamist tradition as a pastoral theologian opposed to strictly speculative questions. His overarching interests lay in the pastoral foundations of theology and opposed abstract and hence 'unedifying' metaphysical questions, as these dominated scholastic discourse in the theological faculty at Paris. He sought to mediate between increasingly polemical school disputes, arguing for the recovery of a 'biblical' theology that led away from speculative questions toward mystical encounter with God. Later known as* doctor christianissimus *(the most Christian doctor), Gerson exerted such a profound influence upon the subsequent theological horizon that one historian has aptly called the fifteenth century 'le siècle de Gerson'.*

Educated at the College of Navarre in Paris, Gerson attained his *licentia* (1381) and *magister* (1382) in arts before completing his *baccalaureus formatus* in 1390-2 and attaining the licentiate in theology in 1392. In 1395 he was appointed chancellor of Notre Dame and assumed administrative oversight for the University of Paris. Gerson's tenure of this office shaped in practical ways his approach to philosophical questions. His opposition to 'speculative' metaphysical questions that exceeded the 'revealed' truths of scripture must be set alongside his lifelong interest in 'mystical' theology (see for example his *De mystica theologia (On Mystical Theology)*), but he conceived of the latter not as an abstract or speculative form of cognition but in terms of its transformative power that led toward an experiential and affective understanding (*cognitio experimentalis et affectualis*). At the conclusion of the Council of Constance (1414-8), which Gerson attended as representative of the university and the province of Sens, he completed his important treatise *De consolatione theologiae (On the Consolation of Theology)*, a dialogue about ecclesial and theological questions which sought to extend rather than simply imitate the philosophical precedent of Boethius.

Gerson eventually returned to Lyons, where he took up residence at the Coelestine Priory and completed works of a wide topical range. Among these are several works devoted to epistemological questions, including a treatise exploring the relation of scholastic and mystical theology, *De elucidatione scholastica mysticae theologiae (On the Scholastic Elucidation of Mystical Theology)*, and two works of fifty propositions each, devoted to the epistemological debate between realists of the *via antiqua* and nominalists of the *via moderna*, *De modis significandi (On Modes of Signifying)* and *De concordia metaphysicae cum logica (On the Harmony of Metaphysics and Logic)*. Although he retained his appointment as university chancellor until his death, Gerson never returned to Paris after the council. He died on 12 July 1429, and was buried in the church of St Lawrence at Lyons.

The question of Gerson's philosophical commitments has been much disputed in recent studies. This derives in part from the moderating character of his thought, and in part from developments in his own thinking. The most convincing portrait of Gerson's mature philosophical position remains that offered by Gerhard Ritter (1922), who situates him as a philosophical theologian interested in finding a mediating position that would secure the foundation for a philosophical metaphysics on the basis of a nominalist epistemology. Gerson sought to maintain, according to Ritter, the nominalist distinction of signs and things signified, and identified 'natural reason' as limited to the former. At the same time, Gerson insisted that such signs were not devoid of true signification, and thus metaphysical arguments could hold real epistemological certitude. On the basis of Gerson's strong syncretistic tendencies, Ritter represents him as a conciliatory figure attempting to follow a middle way between the 'modern' heirs of William of Ockham and advocates of an earlier realist tradition.

Such later tendencies are already evident in his earlier program for curricular reform (see for example his letter to Pierre d'Ailly in 1401, in *Oeuvres Complètes* 2: 26-), as in his larger treatise *Contra curiositatem studentium (Against the Curiosity of Scholars)*. In these writings, Gerson avoided theoretical questions by evaluating the legitimate place of *philosophia* in terms of pastoral questions and concerns. Thus he insisted that the *articuli fidei* (articles of faith) were in no way contrary to 'natural reason'; the human pursuit of knowledge should respect certain limited boundaries within which philosophical speculation might legitimately proceed. Such 'natural reasoning' should not, he argued, apply itself through 'vain curiosity' to unedifying questions removed from the actual practice of Christian life, since this often transgressed the boundaries both of reason and of faith and

penetrated the hidden realm of the divine will (*secretum divinae voluntatis*). Rather, one should be guided by humility to accept as sufficient the knowledge gained by scaling the ladder of scripture (*scala scripturarum*), even though one should 'elucidate humbly [through the use of natural reason] the truth of holy scripture within the limits where this is possible' (*Contra curiositatem studentium*, in *Oeuvres Complètes* 3: 233). At stake for Gerson was how to apply reason properly to unfold the revealed truths of faith, in order to strive beyond such knowledge by mounting the 'other ladder' leading toward the 'higher [mystical] knowledge of God' (*Contra curiositatem studentium*: 233).

Later writings, particularly his *De consolatione theologiae*, further develop this argument. Here Gerson insists that the revealed truths of scripture (*articuli fidei*), while not immediately accessible to 'natural reason,' are not on this account irrational. Applying Thomist logic to this question, he suggests that 'just as grace exceeds [but does not destroy or oppose] nature…so does theology surpass philosophy - not by rejecting it but by taking it into servitude' (*De consolatione theologiae*, in *Oeuvres Complètes* 9: 188). Philosophy brings us to the 'summit' of natural reason and its consolations, only to yield to the higher reaches accessible only by theology. Thus *philosophia* for Gerson established a useful but limited prolegomenon for theology, finally directing us toward the 'higher' knowledge of faith revealed in and through the church. On the basis of such a mediating position, it is not surprising to note that the official controversy over nominalism that took place in the later fifteenth century found Parisian nominalists seeking to claim Gerson as an authority endorsing their position, while the official edict condemning this movement did not include him among the 'new doctors' condemned.

See also: Nominalism; William of Ockham

MARK S. BURROWS

List of works

Gerson, J. (1401-26) *Opera omnia*, ed. P. Glorieux, *Oeuvres Complètes*, 10 vols, Paris: Desclée, 1960-73.(This is the standard critical edition of Gerson's works. Gerson's extensive writings were already compiled and published in several editions during the later fifteenth century, the first by Johannes Koelhoff, Cologne, 1483/4, and the second by Geiler von Kaysersberg and Petrus Schott, Strassburg, 1488 (repr. 1489, 1494). An early modern edition was prepared by E. du Pin, *Opera omnia*, 4 vols, Antwerp, 1706.)

Gerson, J. (1402) *Contra curiositatem studentium (Against the Curiosity of Scholars)*, in P. Glorieux (ed.) *Oeuvres Complètes*, Paris: Desclée, 1960-73.(Treatise on educational reform.)

Gerson, J. (1402-3) *De mystica theologia (On Mystical Theology)*, in P. Glorieux (ed.) *Oeuvres Complètes*, Paris: Desclée, 1960-73.(Reflects Gerson's lifelong interest in mystical theology.)

Gerson, J. (1418) *De consolatione theologiae (On the Consolation of Theology)*, in P. Glorieux (ed.) *Oeuvres Complètes*, Paris: Desclée, 1960-73.(Influenced by Boethius, Gerson here argues that the revealed truths of scripture are not irrational simply because they are not immediately accessible to natural reason.)

Gerson, J. (1424) *De elucidatione scholastica mysticae theologiae (On the Scholastic Elucidation of Mystical Theology)*, in P. Glorieux (ed.) *Oeuvres Complètes*, Paris: Desclée, 1960-73.(Treatise exploring the relationship between scholasticism and mystical theology.)

Gerson, J. (1426) *De modis significandi (On the Mode of Signifying)*, in P. Glorieux (ed.) *Oeuvres Complètes*, Paris: Desclée, 1960-73.(Propositions exploring the debate between realism and nominalism.)

Gerson, J. (1426) *De concordia metaphysicae cum logica (On the Harmony of Metaphysics and Logic)*, in P. Glorieux (ed.) *Oeuvres Complètes*, Paris: Desclée, 1960-73.(Propositions exploring the debate between realism and nominalism.)

References and further reading

Burger, C. (1986) *Aedificatio, Fructus, Utilitas. Johannes Gerson als Professor der Theologie und Kanzler der Universität Paris (Edification, Fruition, Usefulness: Jean Gerson as Professor of Theology and Chancellor of the University of Paris)*, Tübingen: J.C.B. Mohr (Paul Siebeck).(Critical overview of Gerson's university writings to 1415, emphasizing the pastoral tone and ecclesial context of his thought.)

Burrows, M. (1991) *Jean Gerson and 'De Consolatione Theologiae' (1418): The Consolation of a Biblical and Reforming Theology for a Disordered Age*, Tübingen: J.C.B. Mohr (Paul Siebeck).(Oriented toward Gerson's writings from 1415, with opening discussion of current Gerson historiography, full bibliography and detailed consideration of how pastoral and ecclesiological concerns shaped philosophical and theological arguments.)

Combes, A. (1945-72) *Essai sur la critique de Ruysbroeck par Gerson (A Study of Gerson's Criticism of*

Ruysbroeck), Paris: Vrin, 3 vols. (Detailed analysis of Gerson's theological criticism of Ruysbroeck's union mysticism, with careful attention to theological and soteriological questions.)

Combes, A. (1963-4) *La théologie mystique de Gerson. Profil de son évolution (Gerson's Mystical Theology: Profile of its Develpment)*, Paris: Vrin, 2 vols.(Masterly survey of Gerson's mystical theology, early and late, with detailed discussion of how his thought related to that of his contemporaries.)

Delaruelle, E. (1964) *L'Eglise au temps du Grand Schisme et de la crise conciliaire (The Church from the Time of the Great Schism to the Conciliar Crisis)*, Brussels: Bloud & Gay, 1964.(Standard French study of the period, emphasizing the relationship of institutional and intellectual developments.)

Ritter, G. (1922) *Studien zur Spätscholastik (Studies of Late Scholasticism)*, vol. 2, *Via antiqua und via moderna auf den deutschen Universitäten des XV. Jahrhunderts (The Via Antiqua and Via Moderna in German Universities in the Fifteenth Century)*, Heidelberg: Carl Winter. (Careful, technically demanding study of late-scholastic philosophical debates between advocates of 'old' and 'new' schools; unsurpassed by subsequent studies.)

Vignaux, P. (1948) *Nominalisme au XIVe siècle (Nominalism to the Fourteenth Century)*, Paris: Vrin.(Broad and masterful orientation to late medieval nominalism, with subtle interpretation of both philosophical issues and their theological consequences.)

Gersonides (1288-1344)

Living all his life in southern France, Levi ben Gershom, known as Gersonides in Latin texts, was an accomplished astronomer and mathematician as well as a philosopher. A prolific and engaged exegete, Gersonides wrote biblical commentaries that are still studied today. His philosophical magnum opus, *Milhamot ha-Shem (The Wars of the Lord), reached original and often unorthodox conclusions regarding many of the great issues of medieval philosophical theology. It denied creation* ex nihilo, *preferring a modified version of the doctrine of* formatio mundi *traditionally ascribed to Plato (formation of the world from pre-existing matter). It qualified traditional doctrines of divine omniscience by denying God's determinate knowledge of future contingent events. And it confined personal immortality to the rational portion of the soul, that is, the intellect.*

1 Life and works

Known in traditional Hebrew sources by the acronym Ralbag (Rabbi Levi ben Gershom), Gersonides was born in Provence and lived there all his life. Besides his philosophical work, he made important contributions to astronomy, mathematics, and biblical exegesis. Both a theorist and an observer in astronomy, he held to the Ptolemic system but rejected the epicycles and other aspects of the medieval Ptolemaic worldview. His commentaries on the Pentateuch, Proverbs, Job, the Former Prophets and several of the smaller Hagiographia are full of philosophy, which he saw as the key to the true understanding of Scripture. He continued the Jewish tradition of writing supercommentaries on Averroes' commentaries on Aristotle, but his original philosophical work, *Milhamot ha-Shem*, written in Hebrew in six books, covers the full range of medieval psychology, metaphysics, philosophy of nature and cosmology. Gersonides was less interested in moral and political philosophy, but his biblical commentaries do reflect on these topics. This essay will consider three of the central issues treated in *Milhamot ha-Shem*: creation, divine knowledge, and immortality.

2 Creation

Creation was a persistent problem in medieval philosophy (see Creation and conservation, religious doctrine of). Both Aristotelians and the Neoplatonists regarded the universe as eternal; yet the Bible and the Qur'an teach creation. Since Gersonides believed that philosophy and the Torah teach the same truth, the apparent disagreement needed to be resolved. His admired philosophical predecessors Averroes (see Ibn Rushd) and Moses Maimonides had taught two contrary theses: Averroes, that the universe is eternal, Maimonides, that it was created *ex nihilo* but that neither the world's creation nor its eternity is demonstrable. Gersonides rejected both views in favour of *formatio mundi*: creation was not bringing something into existence from nothing but the organizing of an eternal, hitherto formless matter. This theory, Gersonides urged, is demonstrable. He devoted the longest part of *Milhamot ha-Shem* to formulating numerous arguments in support of some form of creation. These fall into three distinct groups.

(1) Clearly favouring teleology in metaphysics and cosmology, Gersonides makes the classic argument from design a key proof not just of the existence of God but of creation as he understands it. Anything with teleological properties, he argues, must be the work of a purposive agent - unless it is a chance phenomenon. Focusing on the heavenly bodies, eternal beings for Aristotle (§16), and certainly not chance phenomena in Aristotelian philosophy, Gersonides urges that these do exhibit teleological properties, most plainly in the case of the sun, whose beneficent functioning is obvious. The system and arrangement of the heavenly bodies in general similarly manifest order and design and so show them to be products of purpose and intent, thus, created (*Milhamot ha-Shem*: 6.1, 7 and 9).

(2) Gersonides reinforces this argument with the 'particularization argument' often favoured by Muslim theologians of the *kalam* (see Islamic theology). Aristotle's heavenly bodies all consist of a fifth element (the 'quintessence' of the scholastics). Yet they exhibit diverse properties. Mars, for example, emits a reddish light, whereas Venus has a bluish cast. Aristotle's attempt to explain such properties in terms of distances from the earth fails, since Saturn and Jupiter both emit a whitish light, although Saturn is much further from the earth than Jupiter. Since chance was excluded by Aristotle himself, the observed differences can be explained only by reference to a 'producing agent' (*Milhamot ha-Shem*: 6.1, 8).

(3) Relying on Aristotle's doctrine (*Physics* III 4-8) that an actual infinity is impossible, Gersonides argues, in the style of *kalam*, that the world's eternity would entail an actual infinity. Among his several variations on this long familiar line of argument one is especially interesting, since it is not invalidated by modern astronomy or mathematics: Gersonides argues from our intuition that the past is closed and the future open. Unlike the future, where it is often not determined whether p or ¬p, for all states of affairs in the past it is already the case that p or ¬p; that is, the past is completely determinate. Each fact about the past is real: it is a definite outcome of a preceding state of affairs and has causal consequences, many of which have lasting effects. In this sense the past is always with us. To Gersonides this means that unlike the future, which to a considerable extent is empty, the past is actual. If it were infinite, we would have an actual infinity (*Milhamot ha-Shem*: 6.1, 10-11) (Eternity of the world, medieval views of).

Concluding that the world must be created, Gersonides goes on to argue that its creation cannot be *ex nihilo*. Here he relies on Aristotle's physics and on the fact that familiar cases of generation are always from, or out of, something. He also appeals to the idea that in nature 'like produces like'. Anticipating the critics of Cartesian mind-body interaction, he finds it implausible to say that pure form could bring bodies into being from no pre-existing matter. God, an incorporeal being, could no more create something physical *ex nihilo* than a stallion and a mare could produce a bear.

Defenders of *ex nihilo* creation might not feel troubled by these objections, since creation for them is no ordinary natural event but the 'mother of miracles'. Why should we expect it to conform to the familiar natural laws? But Gersonides goes further. Creation *ex nihilo,* he argues, would entail a vacuum, both before and after the event. And a vacuum, it was widely assumed, was not just a natural but a logical impossibility. In the beginning, according to the adherents of absolute creation, there was only God; then a world was made. If this world was preceded by and subsequently located in empty space - since it is a finite body - then there would inevitably be a vacuum not only prior to the world's creation, but even afterwards, unless the world were infinite in size, a possibility that Aristotle was generally thought to have ruled out (*Physics* IV 7; *On the Heavens* III 2; see *Milhamot ha-Shem*: 6.1, 17-8).

In response to these difficulties, Gersonides adopted the account that medieval thinkers usually thought of as Plato's: the Creator fashioned the universe out of a pre-existing matter (see Plato §16). Although eternal, this matter was not divine. For it was unstable and utterly imperfect, having no intrinsic form or shape. 'Not preserving its shape', it is no competition to God, who is immutable and utterly perfect. Matter is not worthy of worship. It is simply the raw material from which our world was manufactured. It is also the source of all disorder and defect - including moral imperfection (*Perush 'al Sefer Iyyov [Commentary on Job]* - introduction).

Gersonides caps his cosmology by arguing that the universe is indestructible, although generated, contrary to Aristotle, who held that whatever is generated is destroyed (*On the Heavens* I 12). Both Plato and Maimonides had believed the world indestructible, but they ascribed its endurance to God's will (see Plato, *Timaeus* 41a-b; Maimonides, *Guide to the Perplexed* II 27). Gersonides went further, arguing that the cosmos is intrinsically indestructible, since it contains no inherent cause of corruption. God not only would not but could not destroy it; for God would have no reason to destroy what is as perfect as it can be. Once created, then, the world endures forever (*Milhamot ha-Shem*: 6.1, 16).

3 Divine knowledge

Gersonides deviates from familiar views on divine omniscience as well. Ever since Augustine, most medieval thinkers had opted for some kind of compatibilist solution to the venerable difficulties of reconciling divine foreknowledge with human freedom. Not so Gersonides. Embracing the position of the 'philosophers' (that is, Neoplatonic Aristotelians), he concludes that an eternal and incorporeal God cannot know spatio-temporal particulars as such, least of all, future contingent events. If God knew these as particulars, they would lose their contingency. In an argument that anticipates the thinking of recent philosophers, Gersonides preserves human freedom even at the cost of limiting God's knowledge: God knows the general laws governing human behaviour - how humans in general would react to a divine command to kill their only child, and that it is possible for someone to be willing to do so. But God does not know whether Abraham or Jephtha is that person. For humans have the power to act contrary to the biological-psychological generalities of human behaviour, and God knows that these patterns can be broken by human freedom (see Free will; Omniscience).

Gersonides did not regard his solution as a weakening of God’s omniscience, since an omniscient being knows only what is knowable. Future contingent events are not knowable in principle. If they were, they would forfeit their contingency (see Aristotle, *On Interpretation* IX). If it is objected that this means that God lacks knowledge of particulars that even a child might know, Gersonides is ready with an answer: ignorance of spatio-temporal events is not a genuine deficiency in a being who transcends the spatio-temporal order.

Thomas Aquinas admitted that God would not know future events as future but as ‘eternally present’ (*Summa theologiae* Ia.14.13). Gersonides goes further, holding that an eternal being does not know temporal particulars at all. For, as Aristotle claims, genuine knowledge is of the universal (*On the Soul* II 5).

How then can we make sense of the biblical stories in which God speaks to individuals such as Abraham and Moses at specific times? Here Gersonides calls on the aid of the highly regarded twelfth-century biblical exegete, Abraham ibn Ezra, who had reached the same conclusion about divine knowledge. Relying heavily on the traditional hermeneutical liberty to interpret biblical texts so as to reconcile them with philosophical knowledge (*Milhamot ha-Shem*: 3.6), he adapts his reading of Scripture to fit his philosophical conclusions; for example, Bilaam’s donkey didn’t really speak: that incident was just one episode in Bilaam’s prophetic dream. More literal readers of the Bible might find the outcome strained. But Gersonides’ boldness is noteworthy and admirable: he does not allow even the authority of text or tradition to subvert his philosophical conclusions but bends both to accommodate the truth as he believes argument to have shown it.

4 Immortality

Gersonides shows similar independence regarding immortality. By his time the question had become quite focused: what exactly in humans is immortal and how does it become immortal? Within Gersonides’ philosophical orbit the standard reply was that the human intellect becomes immortal through ‘conjunction’, or linkage with the ‘active intellect’. This active intellect was a transcendent, incorporeal cosmic intelligence responsible not only for human intellectual knowledge (including prophecy) but also for informing all things with their objective natures and thus for the governance of biological reproduction and development. The idea of the active intellect stems from a somewhat elusive paragraph in Aristotle’s *On the Soul* (III 5), which speaks of an active intellect that is a precondition of human understanding. In interpreting this passage, Gersonides follows the tradition laid down by Alexander of Aphrodisias (*fl*. *c*.AD 200), who regarded the active intellect as a transcendent and eternal power but the human intellect as a material disposition or capacity of the body, specifically, of the imagination. Other interpreters such as Themistius and Avicenna (see Ibn Sina) believed the human intellect to be an incorporeal substance, inherently capable of immortality. How could Gersonides, as a follower of Alexander’s tradition, account for immortality, if the human intellect was just a biological function and thus susceptible to corruption?

Unlike Alexander or Averroes (whose position is closer to that of Themistius), Gersonides rejected ‘linkage’ with the active intellect. He did not think that we ever reach an intellectual level that would allow so intimate a bond with the transcendent. Besides, the Averroist version of ‘conjunction’ implied the obliteration of individuality: all minds become one in the active intellect. Such loss of identity was objectionable both philosophically and religiously. What Gersonides was seeking was personal immortality.

If the human intellect is only a capacity of the imagination to acquire knowledge, by abstraction from and generalization about sense-data, albeit with the aid of the active intellect, what persists after physical death? Gersonides turns for help to Plato: although the senses are needed initially to acquire the data from which we derive our general concepts, these concepts are grounded not only in the natural world, in the types, or natural kinds, of things, but also in the active intellect itself, which exhibits the ‘rational order’ (a favourite phrase of Gersonides’) of the terrestrial world. Like Plato’s eternal paradigm (*Timaeus* 28-9) the rational order in the active intellect is an everlasting blueprint for the natural world (see Plato §16). In so far as the human mind apprehends, or ‘participates’, in this plan it becomes incorruptible and everlasting. This is the meaning of the expression ‘acquired intellect’ as used in Alexander’s psychology. That intellect in us is the repository of the intellectual capital our minds have accrued. Immortality is achieved not through conjunction with the active intellect but by our individual intellectual efforts and successes. Since what is immortal is our own knowledge, unique to each individual, the immortality we are assured is personal.

5 Conclusion

Although working within a well-defined philosophical tradition that continually tried to show its religious fidelity, Gersonides departed from some of the central doctrines of traditional medieval faith. His loyalty to biblical creation led him to reject Aristotle's notion of the eternity of the universe; but his commitment to Aristotle's physics and metaphysics forced him to reject *creatio ex nihilo* as well. In his theories of divine omniscience and human immortality, he sided with the 'philosophers' more than with the 'theologians'. Yet he did not reject divine omniscience and he worked creatively to sustain individual immortality. His views provoked much controversy, and the *Milhamot ha-Shem* received hostile criticism (notably from Crescas). But Gersonides' biblical commentaries, which often voice his philosophical conclusions, are still studied by traditional Jews. Like Maimonides, Gersonides could think radically and originally, even to the point of heterodoxy, but could win the acceptance of the religious community through a combination of philosophical acumen, biblical and rabbinic erudition, and a life conducted within the norms and traditions of practice of the community.

See also: Aristotelianism in Islamic philosophy; Aristotelianism, medieval; Eternity of the world, medieval views of; God, concepts of; Maimonides, M.; Natural philosophy, medieval; Platonism, medieval; Soul, nature and immortality of the

SEYMOUR FELDMAN

List of works

Levi ben Gershom (Gersonides) (1317-29) *Milhamot ha-Shem (The Wars of the Lord)*, Leipzig: Lorck, 1866; Books 1-4 trans. S. Feldman, *The Wars of the Lord*, Philadelphia, PA: Jewish Publication Society, 1984, 1987, 2 vols.(Covers the full range of medieval psychology, metaphysics, philosophy of nature and cosmology. The longest part of the work is devoted to formulating numerous arguments in support of some form of creation. Volume 3 of the Jewish Publication Society edition is due to be published, containing Books 5 and 6, a bibliography and indices.)

Levi ben Gershom (Gersonides) (1325) *Perush 'al Sefer Iyyov (Commentary on the Book of Job)*, trans. A. Lassen, *Commentary of Levi ben Gershom on the Book of Job*, New York: Bloch Publishing House, 1946.(In the introduction to his commentary, Gersonides makes the argument that matter is not worthy of worship. This argument suggests that matter is simply the raw material from which our world was manufactured, and that it is the source of all disorder and defect - including moral imperfection.)

Levi ben Gershom (Gersonides) (1319) *Sefer ha-Heqqesh ha-Yashar (Book of the Correct Syllogism)*, trans. C. Manekin, *The Logic of Gersonides: A Translation of the Sefer ha-Heqqesh ha-Yashar*, Dordrecht: Kluwer Academic Publishers, 1992.(Influenced by the works of Aristotle and Averroes, Gersonides wrote this work, criticizing several arguments of Aristotle.)

Levi ben Gershom (Gersonides) (1329-38) *Perush 'al ha-Torah (Commentary on the Torah)*, ed. J. Levy, *Perushei Ralbag 'al ha-Torah*, Jerusalem: Mossad Ha-Rav Kook, 1992, 1994, 2 vols.(The two volumes which comprise the Levy edition are the parts of the commentary on Genesis (1992) and Exodus (1994).)

Levi ben Gershom (Gersonides) (1328-40) *The Astronomy of Levi ben Gershom (1288-1344)*, trans. with commentary by B. Goldstein, New York: Springer, 1985.(This translation is of the first 20 chapters of the astronomical-mathematical portion of Book 5, part 1 of the *Milhamot ha-Shem*.)

References and further reading

Dahan, G. (ed.) (1991) *Gersonides en son temps*, Louvain and Paris: E. Peeters.(A collection of recent essays on Gersonides' life and thought, with a good bibliography.)

Feldman, S. (1967) 'Gersonides' Proofs for the Creation of the Universe', *Proceedings of the American Academy of Jewish Research* 35: 113-37.(Detailed analysis of Gersonides' various proofs for the creation of the universe.)

Feldman, S. (1975) 'Platonic Themes in Gersonides' Cosmology', *Salo W. Baron Jubilee Volume*, ed. S. Lieberman and A. Hyman (Jerusalem: American Academy for Jewish Research) l: 383-405.(Intensive discussion of Gersonides' critique of *ex nihilo* creation and defence of creation out of eternal formless matter; arguments for the indestructibility of a created universe.)

Feldman, S. (1978) 'Gersonides on the Possibility of Conjunction with the Agent Intellect', *Association for Jewish Studies Review* 3: 99-120.(Philosophical and historical analysis of Gersonides' critique of Averroes' theory of intellect and immortality.)

Goldstein, B. (1969) 'Preliminary Remarks on Levi ben Gerson's Contribution to Astronomy', *Proceedings of the Israel Academy of Sciences and Humanities* 3 (9): 239-254.(A good introduction to Gersonides' astronomy.)

Kellner, M. (1980) ‘Gersonides on the Problem of Volitional Creation’, *Hebrew Union College Annual* 51: 111-28.(A critical analysis of Gersonides’ doctrine of creation, especially his rejection of *ex nihilo* creation.)
Manekin, C. (1985) ‘Preliminary Observations on Gersonides’ Logical Writings’, *Proceedings of the American Academy for Jewish Research* 52: 85-113.(A good but technical treatment of some of the more significant issues in Gersonides’ logic.)
Rudavsky, T. (1983) ‘Divine Omniscience and Future Contingents in Gersonides’, *Journal of the History of Philosophy* 21: 513-36.(Good but difficult discussion of Gersonides’ doctrine of divine omniscience, focusing upon its philosophical adequacy and fidelity to Scripture.)
Samuelson, N. (1972) ‘Gersonides’ Account of God’s Knowledge’, *Journal of the History of Philosophy* 10: 399-416.(Comprehensive philosophical analysis of Gersonides’ doctrine of divine foreknowledge.)
Touati, C. (1973) *La Pensée philosophique et théologique de Gersonide*, Paris: Les Editions de Minuit.(The most comprehensive study of Gersonides’ philosophy.)
Wolfson, H. (1953) ‘Maimonides and Gersonides on Divine Attributes as Ambiguous Terms’, *Mordecai M. Kaplan Jubilee Volume* (ed. M. Davis, New York: Jewish Theological Seminary of America), 515-30.(Classic treatment of the philological-textual background of this problem.)

Gettier problems

The expression 'the Gettier problem' refers to one or another problem exposed by Edmund Gettier when discussing the relation between several examples that he constructed and analyses of knowing advanced by various philosophers, including Plato in the Theatetus. Gettier's examples appear to run counter to these 'standard' or 'traditional' analyses. A few philosophers take this appearance to be deceptive and regard the genuine problem revealed by Gettier to be: 'How can one show that Gettier's examples are not really counterexamples to the standard analyses?' But most philosophers take seriously the problem which is the central concern of this entry: 'How can such standard analyses be altered so that Gettier's cases do not constitute counterexamples to the modified analyses (and without opening the analyses to further objections?)'.

Gettier's short paper spawned many important, ongoing projects in contemporary epistemology - for instance, attempts to add a fourth condition of knowing to the traditional analyses, attempts to replace some conditions of those analyses, such as externalist accounts of knowing or justification (causal theories and reliability theories), and revived interest in scepticism, including an investigation of the deductive closure principle. Difficulties uncovered at each stage of this research have generated an ever more sophisticated set of accounts of knowing and justification, as well as a wealth of examples useful for testing proposed analyses. In spite of the vast literature that Gettier's brief paper elicited, there is still no widespread agreement as to whether the Gettier problem has been solved, nor as to what constitutes the most promising line of research.

1 Gettier-type cases

Clarifying the Gettier problem as most philosophers have seen it ('How can the standard analyses be altered so that Gettier-type cases do not constitute counterexamples to the modified analyses, and without opening the analyses to further objections?') involves clarifying the terms 'a standard analysis' and 'Gettier-type cases'. We may here sidestep concern with the former (but see Shope 1981) by rephrasing the problem as: 'What aspect of a correct analysis of one's knowing that *p* prevents Gettier-type cases from constituting counterexamples to it?' Philosophers who adopt the broadest perspective regard a Gettier-type case as any situation where justified, true believing falls short of knowing. Others are influenced by the manner in which they expect the central problem to be resolved. For example, those who think that the counterexamples in question can be avoided by clauses in a correct analysis ruling out the falsity of specific types of propositions to which the knower is related will characterize Gettier-type cases as situations where knowledge is absent because various propositions are false. Those who hope to avoid the counterexamples by strictures on the durability or degree of justification of one's believing that *p* may prefer to define 'Gettier-type case' by mentioning aspects of such justification. Alternatively, Alvin Plantinga (1993b) suggests that Gettier-type cases involve a slight mismatch between the proper functioning of one's cognitive equipment and one's circumstances.

Approaches to resolving the central problem have been too multifarious for a brief summary. So, setting aside differences over what constitutes an analysis (Shope 1981, and forthcoming), consider how variations on the following important Gettier-type case (Lehrer 1965) impact on a number of prominent attempts to analyse knowing:

Mr Nogot: Somebody in one's office, Mr Nogot, has given evidence, *e*, that completely justifies one in believing that *f*: 'Mr Nogot, who is in the office, owns a Ford'. Evidence *e* consists in such things as Nogot's having been reliable in dealings with one in the past, having just said to one that he owns a Ford, and having just shown legal documents confirming it. From the proposition that *f*, one deduces and thereby comes to believe that *p*: 'Somebody in the office owns a Ford'. Unsuspectedly, Nogot has been lying and it is someone else in the office who happens to own a Ford. One has true, justified belief that *p* but not knowledge that *p*.

This case differs slightly from one of Gettier's original (1963) examples, in which one notices that by picking at random a city name such as 'Barcelona', one can deduce from the unsuspectedly false proposition that *f′*, 'Jones owns a Ford', the conclusion that *p′*, 'Either Jones owns a Ford or Brown is in Barcelona', which one proceeds to believe on the grounds of the proposition that *f′*.

In Gettier's other original example, one justifiably believes the unsuspectedly false proposition that *f″*: 'Jones, who

is not myself, is the man who will get the job, and Jones has ten coins in his pocket', and one sees that it entails that *p*″: 'The man who will get the job has ten coins in his pocket', which one then believes on the grounds of the proposition that *f*″. Unsuspectedly, not only does one have ten coins in one's own pocket, but it is oneself who is going to get the job.

Some variants of the Nogot case do not involve one's having relevant false beliefs. (Indeed, Gettier only spoke of a *proposition* as the basis for one's inferring that *p*′ [that *p*″].) For instance, one may happen to be a clever and cautious reasoner not interested in who owns a Ford but only in whether it is true that *p*, and realizing that someone in the office other than Nogot might happen to own a Ford, one decides that it is safer merely to accept that *p* than to believe that *f* (see Lehrer 1974). Alternatively, Richard Feldman (1974) has noted that one may not arrive at a belief that *p* by relying upon considerations about the proposition that *f*, but instead by utilizing as a basis for believing that *p* the following true existential generalization of one's evidence: 'There is someone in the office who has provided me with evidence *e*'.

2 Defeasibility analyses

Defeasibility analyses of knowing attempt to exploit the insight that specific propositions are true (for example, 'Nogot owns no Ford'), whose hypothetical inclusion in the set of propositions that one believes would have a negative impact either upon the justified status of the belief or upon the justified status of the proposition that *p* (see Knowledge, defeasibility theory of). For instance, Peter Klein (1971) advanced the requirement: 'No proposition is true such that if one were to believe it, then the justification condition would not hold'. We might call such a truth a 'defeater' of *p*. The requirement shows that one fails to know that *p* in the following case:

Tom Grabit's Twin: One is well-acquainted with Tom Grabit, and clearly sees him steal a book from the library. But in consequently coming to believe that *p*, 'Tom stole the book', one does not realize that *t*, 'Tom has an identical twin and that the twin was in the library around the time of the theft'.

Similarly, if in the two variants of the Nogot case mentioned above one were also to have the true belief that not-*f*, then one would not be justified in believing that *p*.

But here we encounter the difficulty of avoiding Gettier-type cases without creating further problems, one such problem being the generation of non-Gettier-type counterexamples. Klein's defeasibility analysis faces this difficulty concerning a variant of the Tom Grabit case as it was originally described by Keith Lehrer and Thomas Paxson (1969) - one indeed knows that *p* but does not realize that *t*: 'Tom's mother has testified to police that *f*: "Tom has an identical twin and was miles away from the library at the time of the theft"', and that *t*′: 'The twin is not real and Mrs Grabit is lying'. If one were merely to come to believe the misleading evidence that *t* without also coming to believe the truth that *t*′, one would not be justified in believing that *p*.

In response, Klein (1981) deepened the defeasibility approach by requiring, roughly, that the combination of a defeater (such as *t*) with *S*'s evidence should not remove *S*'s justification for believing *p*, essentially through justifying *S* in believing a false proposition (such as *f*). But the resulting analysis of knowing is too weak to deal with a variant of the Nogot case in which (1) after coming to believe that *n*: 'Nogot owns a Ford', *S* sees, as luck would have it, a true sentence expressing a disjunction, *d*, which, unsuspected by *S*, is such that any defeater of *n* entails one or another disjunct of *d* that *S* is not presently justified in believing; (2) on the basis of believing that *n*, *S* comes to believe that *p*: '*n* or *d*'; (3) it is true that *n* because Nogot is lucky enough to win a Ford in a raffle while in the company of *S*. Defeaters of *n* fail to be defeaters of *p*, since they support its second disjunct. Yet there need be no other truths that defeat *p*. So Klein does not show why *S* fails to know that *p*.

Perhaps this case of Lucky *S* and Lucky Nogot could be handled by a sufficiently clarified defeasibility condition inspired by John Barker's (1976) concern with the fact that there may be other truths, which, when combined with a defeater of *p* and with *S*'s evidence for *p*, restore *S*'s original justification for believing *p*. (Yet Klein's 1981 work pointed out that restoring evidence may be too easy to come by, for example, exact specifications of nearby lookalikes whose nearness should deny knowledge - such as a real twin in a variant of the Grabit case; see also the case of the Back-Up Volcano in Shope (1981).)

In a sophisticated defeasibility account, Barker's perspective is extended by John Pollock (but in a fashion that may fail to handle the case of Lucky *S* and Lucky Nogot). Pollock (1986) focuses on the justification for the

statement that *p* which is provided by an argument whose premises are drawn from the reasons forming one's basis for believing that *p*. He defines the defeat of such an argument by a set, *S*, of propositions as a situation where: (1) one has the reasons in question and it is logically possible for one to become justified in believing that *p* on the basis of those reasons; and (2) one believes the members of set *S*, where *S* is consistent with the proposition that *p* and is such that it is not logically possible for one to become justified in believing that *p* on the basis of the combination of one's present reasons and the members of *S*. As a condition of one's knowing that *p*, Pollock proposes that the rational presumption in favour of believing that *p* created by reasons drawn from one's present reasons for believing that *p* is undefeated by the set of all truths. Pollock appears to be requiring that the following conditional proposition, *R*, must be true: 'If one were to retain one's present reasons for believing that *p* and were to believe the members of the set of all truths, then one would be justified in believing that *p* on that basis'. Thus, one lacks knowledge in the original Nogot case because of the inclusion of the proposition that Nogot owns no Ford in the set of all truths. In contrast, one succeeds in knowing in the case of the lying Mrs Grabit thanks to the inclusion in the set of all truths of the proposition that *t′* (Plantinga 1993a gives fuller exposition).

Some examples of introspective knowledge are a challenge for defeasibility theories and Pollock (1992) seeks to handle them by treating conditionals as true when they have logically or metaphysically impossible antecedents. For instance, when one knows that *p*: 'I believe that *f*', one's reasons for believing may include the presence of the introspected state, one's believing that *f*. Suppose that one does not realize that it is false that *f*. At least for some instantiations for '*f*', the antecedent of Pollock's conditional involves the impossible situation of retaining one's reason, that is continuing to believe that *f*, while also believing that not-*f* - since the proposition that not-*f* belongs to the set of all truths. (In contrast, Klein needs here to maintain the less-than-obvious claim that *S*'s believing the conjunction 'Not-*f* and *S* believes that *f*' justifies *S*'s believing that *S* believes that *f*.)

But Pollock's treatment of conditionals as being true when they have impossible antecedents opens his analysis to non-Gettier-type counterexamples concerning one's coming to believe that one is in a particular mental state through inference from false beliefs rather than through introspection. On the basis of a justified but imperfect psychological theory, I may justifiably believe the following two propositions that I do not realize are false - f_1: 'Everyone has memories of early childhood', and f_2: 'Everyone who has memories of early childhood believes everyone to have memories of early childhood'. On this basis I come justifiably to believe the true proposition that *b*: 'I believe that f_1'. Pollock's account incorrectly entails my knowing that *b*, since it will be impossible for me to retain my reasons for believing that *b* while adding the true beliefs that not-f_1 and that not-f_2.

3 Causal analyses and proper functioning

The preceding Gettier-type cases involve no causal relationship between one's believing that *p* and the state of affairs that one believes to obtain. So various analyses requiring such specified causal connections have been proposed (see Knowledge, causal theory of). Alternatively, sometimes conditionals have been included in a fourth condition of knowing, relating believing that *p* to what one believes to obtain. For instance, Robert Nozick (1981) initiated a new line of research by requiring that if it were the case that *p* then one would believe that *p*. As it stands, this suggestion faces a difficulty that also affects simple versions of causal theories: these views fail to rule out knowledge in a refinement of another variant of the Nogot case that was devised by Keith Lehrer (1979). Here he imagines that Nogot is manifesting a compulsion to trick people into justifiedly believing truths by concocting Gettierized evidence of type *e* for those truths. Nozick's conditional is satisfied if we add the details that: (1) Mr Nogot's neurosis is highly specific to the type of information contained in the proposition that *p*; (2) his compulsion is not easily changed; (3) while in the office he has no other easy trick of the relevant type to play on one; and (4) one arrives at one's belief that *p* not by reasoning through false beliefs but by basing the belief that *p* on a true existential generalization of one's evidence (Shope 1992).

Further variations on the case of tricky Mr Nogot challenge Alvin Plantinga's (1993b) proposal that one's knowing that *p* requires not only that one have a true, justified, sufficiently strong belief that *p*, but that one also meets the following requirements: (*1*) the cognitive faculties involved in the production of one's belief are functioning properly in an environment sufficiently similar to the one for which they were designed; (*2*) the portion of one's design plan covering formation of beliefs when in the latter circumstances specifies that such formation directly serves the function of forming true beliefs; (*3*) if those circumstances include additional beliefs or testimony, then

the latter are or express beliefs also satisfying (*2*) - and so on, backwards through any chain of input beliefs or testimony from one person to another; (*4*) there is a high statistical or objective probability that a belief produced in accordance with that portion of one's design plan in one's type of circumstances is true.

Plantinga notes that the original Nogot case involves credulity, that is believing for the most part what one's fellows tell one. When Nogot says that he owns a Ford, credulity helps to produce one's false belief that *f*, which serves as further input to one's belief-forming faculties so as to generate the true belief that *p*. Plantinga suggests that credulity was designed to function in circumstances where testimony from our fellows does not deceive us and is intended to produce true beliefs in us. Both of the latter circumstances are absent when Nogot brags about his Ford and one comes to believe that *f*.

But in the case of tricky Mr Nogot the informant does intend to produce a true belief in one. Of course, the case does not satisfy (*4*), since it is likely that one will not only believe that *p*, but also believe that *f*. So consider instead a variant of the case where one is in addition a clever and cautious reasoner of the type described earlier, highly likely to avoid believing that *f*, and suppose that Nogot realizes this. He then takes no significant risk of one's forming a false belief.

The example may violate (*3*) if the purpose specified in the design plan for the use of testimony is to induce or sustain belief on the part of another *in what the testifier says*. But Plantinga admits that Nogot provides a wide variety of evidence that supports both the proposition that *f* and the proposition that *p*. So the Nogot case can be altered by imagining that he provides a large quantity of evidence involving no false testimony. Then it is obscure how Plantinga can show that tricky Mr Nogot is providing evidence for a non-designed purpose, since evidence never supports only one proposition, and we rarely, if ever, come to believe everything for which we have evidence.

Suppose that Plantinga is defended by the suggestion that the circumstances in which one's relevant cognitive faculties were designed to operate exclude the type found when tricky Mr Nogot meets the clever reasoner. This response exposes a 'generality problem' in Plantinga's account: what level of generality is relevant to specifying the circumstances pertinent to conditions (*2*), (*3*) and (*4*)? Plantinga rejects the possibility of a naturalistic account of proper function, and focuses attention on what a theistic God would have in mind when designing us in his image, part of which would involve our powers as knowers being analogous to God's in so far as he has beliefs that are true and he, although not designed, fulfils a specification that is an ideal for cognitive design plans, namely, to believe that *p* if and only if it is true that *p*. So what does it matter that a neurosis is what motivates the way in which evidence is provided by tricky Mr Nogot, since true beliefs are ensured?

In any event, Plantinga's requirements are too weak to deal with the following variant of the original Nogot case: Nogot is entirely sincere but his Ford has, unknown to everyone in the office, been repossessed or destroyed by a meteorite since he parked it. Requirement (*3*) implies that when one's belief depends upon a chain of testimony, then at each stage the input propositions must have the right kind of credentials, namely, being a belief that has warrant for the testifier. But the sincere Nogot would have warrant for believing the truth that *f* in a variant of the above case where nothing untoward happens to his Ford. So it is difficult to see how the repossession or meteorite strike removes such warrant in the former case. (Moreover, the case of unlucky, sincere Mr Nogot could also be varied so as not to include any false testimonial evidence.)

4 The role of false propositions

Solutions concerning the role of false propositions continue to intrigue researchers. Roderick Chisholm (1966) requires that when one knows that *p*, something makes *p* evident for one without making any falsehood evident for one - unless the proposition that *p* is a conjunction (for example, 'Nogot is in the office and has presented *e*'), in which case Chisholm also requires that there is something that makes any given conjunct evident for one without making any falsehood evident for one. But the physicist Millikan, who mistakenly accepted the false statement that the charge of the electron was a certain value, did at least know that the value was very, very close to the one that he picked. Anything that made the latter proposition evident for Millikan also made the former falsehood evident for him (Plantinga 1993a provides other counterexamples).

Another solution of the Gettier problem that concerns the exclusion of falsehoods requires that one's believing that *p* be justified through one's grasping enough of a 'justification-explaining chain' - a specific type of structure of

justified propositions concerning what explains the justification of other members without including falsehoods at certain locations (see Shope 1981). The rationale for such a requirement is that fully satisfactory explanations avoid crucial reliance on falsehoods. We must acknowledge and analyse a broader category of knowing that *p* than the narrower, debate-oriented species of concern in a standard analysis, a genus that grants some knowledge to infants and perhaps even to some brutes. Then we may say that: (1) grasping a proposition within a justification-explaining chain, for example, the proposition that *h*, is having knowledge that *h* belonging to the broad category; and (2) the amount of the chain that one must grasp in this way in order to have knowledge that *p* belonging to the narrow category is an amount that suffices for having knowledge that *p* belonging to the broad category (see Shope forthcoming).

See also: Internalism and externalism in epistemology; Introspection, epistemology of; Reliabilism

ROBERT K. SHOPE

References and further reading

Barker, J. (1976) 'What You Don't Know Won't Hurt You?', *The American Philosophical Quarterly* 13 (4): 303-8.(Defeasibility approach mentioned in §2 above, which was the first to deal with misleading defeaters, such as *t*.)

Chisholm R. (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 3rd edn, 1989.(The third edition of this textbook refines earlier research by its author and presents a solution mentioned in §4 above concerning falsehoods.)

Feldman, R. (1974) 'An Alleged Defect in Gettier Counter-Examples', *The Australasian Journal of Philosophy* 52 (1): 68-9.(Describes an example mentioned above in §1 not involving falsely believing.)

Gettier, E. (1963) 'Is Justified True Belief Knowledge?', *Analysis* 23 (6): 121-3; repr. in M. Roth and L. Galis (eds) *Knowing: Essays in the Analysis of Knowledge*, New York: Random House, 1970, 35-8.(Gettier's presentation of his own two examples.)

Klein, P. (1971) 'A Proposed Definition of Propositional Knowledge', *The Journal of Philosophy* 68 (16): 471-82. (A defeasibility analysis presenting the Tom Grabit example mentioned above in §2 that has inspired important variants.)

Klein, P. (1981) *Certainty: A Refutation of Scepticism*, Minneapolis, MN: University of Minnesota Press.(Defends an improvement upon Klein (1971) mentioned in §2 above; relates defeasibility to issues about certainty and scepticism.)

Lehrer, K. (1965) 'Knowledge, Truth, and Evidence', *Analysis* 25 (5): 168-75; repr. in M. Roth and L. Galis (eds) *Knowing: Essays in the Analysis of Knowledge*, New York: Random House, 1970, 55-66.(A defeasibility analysis presenting the Mr Nogot example mentioned above in §2 that has inspired important variants.)

Lehrer, K. (1974) *Knowledge*, London: Oxford University Press.(Presents the clever reasoner example mentioned in §1 above, and analyses knowing so as to accommodate it.)

Lehrer, K. (1979) 'The Gettier Problem and the Analysis of Knowledge', in G. Pappas (ed.) *Justification and Knowledge*, Dordrecht, Boston, MA, and London: Reidel, 65-78.(Presents the tricky Mr Nogot example mentioned in §3 above.)

Lehrer, K. (1990) *Theory of Knowledge*, Boulder, CO and San Francisco, CA: Westview Press.(A textbook refining earlier research by its author.)

Lehrer, K. and Paxson, T., Jr. (1969) 'Knowledge: Undefeated Justified True Belief', *The Journal of Philosophy* 66 (8): 225-37; repr. in M. Roth and L. Galis (eds) *Knowing: Essays in the Analysis of Knowledge*, New York: Random House, 1970, 55-66.(Presents a defeasibility analysis and describes a variant of the Tom Grabit example mentioned above in §2.)

Moser, P. (1989) *Knowledge and Evidence*, Cambridge: Cambridge University Press. (Presents a defeasibility analysis related to Barker's; contains a critique of the solution to the Gettier problem proposed in Shope (1981).)

Nozick, R. (1981) *Philosophical Explanation*, Cambridge, MA: Harvard University Press.(Contains a section on epistemology analysing knowing in terms of 'tracking the truth,' which involves conditionals including the one mentioned in §3 above.)

Plantinga, A. (1993a) *Warrant: the Current Debate*, New York and Oxford: Oxford University Press.(A thorough survey and critique of contemporary attempts to develop a fourth condition of knowing, which includes counterexamples to Chisholm's account.)

Plantinga, A. (1993b) *Warrant and Proper Function*, New York and Oxford: Oxford University Press.(Presents the analysis of knowing discussed in §3 above.)

Pollock, J.L. (1986) *Contemporary Theories of Knowledge*, Totowa, NJ: Rowman & Littlefield.(A textbook including one of the most sophisticated defeasibility analyses, discussed above in §2.)

Pollock, J.L. (1992) 'Reply to Shope', *Philosophy and Phenomenological Research* 52 (2): 411-3.(Explains how his defeasibility theory relates to introspective knowledge; discussed in §2 above.)

Shope, R.K. (1981) *The Analysis of Knowing: a Decade of Research*, Princeton. NJ: Princeton University Press. (Surveys and assesses analyses of knowing presented after the period covered by Slaught (1977), especially in regard to the Gettier problem; offers a characterization of the problem and sketches a solution mentioned above in §4.)

Shope, R.K. (1992) 'Propositional Knowledge', in J. Dancy and E. Sosa (eds) *A Companion to Epistemology*, Oxford: Blackwell, 396-401.(A brief handbook survey of main types of analyses of knowing prior to that offered in Plantinga (1993b), with some consideration of their relation to the Gettier problem.)

Shope, R.K.(forthcoming) *Knowledge as Power*.(Completes the sketch in Shope (1981) of a solution to the Gettier problem in the fashion mentioned above in §4 and defends it against critics, including Moser (1989).)

Slaught, R.L. (1977) 'Is Justified True Belief Knowledge?: A Selective Critical Survey of Recent Work', *Philosophy Research Archives* 3: 1-135.(A thorough survey of early literature concerning the Gettier problem.)

Geulincx, Arnold (1624-69)

Rediscovered in the middle of the nineteenth century, for a long time it was only because of his relation with other more conspicuous philosophers, such as Spinoza, Leibniz and Kant, that interest in Geulincx arose. It has since become clear that he was an original thinker in his own right, who proposed an intriguing metaphysics and made interesting contributions to logic.

1 Life and works

Arnold Geulincx was born in Belgium at Antwerp. In 1641 he matriculated at Louvain, where he became a professor of philosophy in 1646. For reasons that are not clear but were probably of a religious nature (at Catholic Louvain there was much sensitivity over Jansenism), he was suspended from his duties and subsequently dismissed in 1657/8. He moved to Leiden and converted to Calvinism. After taking a degree in medicine he obtained permission to lecture on philosophy, but his position was regularized only in 1662 when he was appointed reader in logic. He became professor ‘extra ordinem’ in 1665, but died in 1669 from the plague. Most of his works, dealing with logic, moral philosophy, physics and metaphysics, were published posthumously.

2 Logic

Although the main merits of Geulincx’s works on logic seem to be their elegance and precision, Dürr and Nuchelmans have shown that he made some important steps towards a logic of propositions. According to him, words like *est* and *non* are signs (*notae*) by which to indicate the mental act performed with respect to a particular content. Every denial is the negation of an affirmative claim, and that means that an affirmation has been present to the mind (*affirmatio inclusa*). Accordingly, ‘Peter is not learned’ must be interpreted as ‘It is not the case that Peter is learned’, or also as ‘The sentence "Peter is learned" is false’. Speaking of compound conditional sentences, Geulincx defines an antecedent as a statement which says that the whole of that which some other statement (the consequent) says to be the case is indeed the case. Consequence is a form of containment (*continentia*): between two statements *A* and *B* there obtains a relation of consequence if *A* says the dictum of *B*. Both the theory of containment and the theorem that every *A* implies the statement ‘*A* is true’ are corollaries of the idea put forward by Geulincx that by making a statement one commits oneself to the truth of that statement and of everything entailed by that statement. For example, if one says ‘I am standing’, this must be taken as an affirmation of whatever is entailed by that statement, such as, for example, ‘I am capable of standing’. Accordingly ‘I am standing’ serves as the antecedent of any number of other statements to the truth of which we commit ourselves (see Conditionals; Propositional logic).

3 Metaphysics and moral philosophy

According to Geulincx, metaphysics is first philosophy or first science. It deals with the human subject, body and God, each of which are the object of a separate science: autologia, somatologia and theologia. The autologia consists in an exploration of the Cartesian *Cogito*, which, however, Geulincx does not see as the basic principle of his philosophy but, rather, as a way to gain access to the realm of necessary truths (see Descartes, R. §5). In fact, the more fundamental principle is the axiom that we can truly be said to make or do something only if we know how it is made or done (*quod nescis quomodo fiat id non facis*). This axiom allows Geulincx to claim that we are passive spectators of the world, our only activities being to will and think, albeit in a purely immanent way. Indeed, the world cannot be the cause of our seeing and perceiving: since it can neither think nor know anything, it cannot be active. The only true cause is God and the only truly causal relation is that between God and the world. In fact, all philosophy should begin with the concept of God; we have to start with the *Cogito* only because the Fall has obscured our faculties. The result of Geulincx’s analysis is that God is Being simpliciter as well as Mind simpliciter. This implies not only that all reality is ultimately mental but also that whatever is neither God nor part of God is nothing but an appearance. In fact, there are only two things that really exist: Mind, which is the creator, and Body, which is the created. Our minds are parts of the Divine mind. Our bodies are part of a phenomenal world. Particular three-dimensional bodies can be understood as limitations of the archetypal Extension that was produced in the act of creation. However, that there is a world extended in three dimensions can be known through the sensations God causes us to have. Finally, since contingent facts cannot be accounted for by principles of metaphysics (which explain only what is necessary), physics makes use of ‘hypotheses’. These must consist of

clear and distinct ideas which together with the principles of metaphysics must be sufficient to explain all phenomena (see Persons; God, concepts of).

Like his metaphysics, Geulincx's moral philosophy is based on a corollary of his fundamental axiom, namely, that where there is no possibility to act there can be no will either (*ubi nihil vales, ibi etiam nihil velis*). In whatever way we act, it is God that makes us act in that particular way. Accordingly, virtue is not to act in a particular way but to yield internally to God's will. Morality lies in the intention, not in the act. As a result, the cardinal virtues are dispositions: diligence, obedience, justice and, above all, humility. Passions on the other hand are like sense impressions. Although they belong to human nature, they are relevant only in so far as they prevent us from developing the right attitude towards God's will. The most dangerous passion in this respect is self-love. In any case, the reward of virtue is that, freed from self-love, we enjoy peace and tranquillity in this life.

4 Relation to other thinkers

Although much in Geulincx's philosophy goes back to Descartes, it would probably be wrong to call him a Cartesian. For not only are the various parts of his philosophy differently connected, his metaphysics is crowned by his moral philosophy, not by his physics, which is basically an attempt to provide a metaphysical account of Divine Creation. Accordingly, his philosophy has more affinity with Malebranche, with whom Geulincx shares a basically occasionalist interpretation of causality (see Malebranche, N. §4). There is also some similarity with Spinoza's philosophy - except, of course, that Spinoza rejects the idea of creation (see Spinoza, B. de §4). In fact, Geulincx's starting point is fundamentally different. Whereas Spinoza argues that on the basis of Cartesian metaphysics it is impossible to account for creation, Geulincx takes creation to be a fact and attempts to make sense of it in terms of Cartesian metaphysics. However, the only way to do this is, he believes, to assume that creation consists in producing a world of appearance. Geulincx's physics is also different from that of Descartes, not because it would involve different concepts but because they understand the status of concepts differently: concepts are 'hypotheses' which, even if clear and distinct, are not automatically and necessarily true. This doctrine, which involves an interplay of empirical and metaphysical principles, has often been associated with Kant's theory of judgment (Cassirer 1911, De Vleeschauwer), but this is somewhat far fetched (see Kant, I. §7). The best characterization seems to be that he is a Christian philosopher trying to find his way in the world of post-Aristotelian philosophy and availing himself of the language and concepts of his contemporaries to provide an intelligible account of the mysteries of faith.

THEO VERBEEK

List of works

Geulincx, A. (1891-3) *Opera philosophica*, ed. J.P.N. Land, The Hague: Martinus Nijhoff, 3 vols; repr. with some additions, *Sämtliche Schriften*, ed. H.J. De Vleeschauwer, Stuttgart Bad Canstatt: Frommann, 1965-8, 3 vols.

Geulincx, A. (1653) *Quaestiones quodlibeticae*, Antwerp; repr. Leiden, 1665; repr. *Opera philosophica/Sämtliche Schriften*, vol. 1, 67-147.

Geulincx, A. (1662) *Logica suis fundamentis restituta*, Leiden; repr. Amsterdam 1689; repr. *Opera philosophica/Sämtliche Schriften*, vol. 1, 165-454.

Geulincx, A. (1663) *Methodus inveniendi argumenta*, Leiden; repr. *Opera philosophica/Sämtliche Schriften*, vol. 2, 1-111.

Geulincx, A. (1667) *Van de Hooft-deuchden*, Leiden; repr. ed. J.P.N. Land, Antwerp, 1895; repr. ed. C. Verhoevan, Baarn: Ambo, 1986.(Dutch translation of Book 1 of the *Ethica*.)

Geulincx, A. (1675) *'Gnoti seauton' sive Ethica integra*, ed. 'Philaretus' (Cornelis Bontekoe), Leiden; 5th edn, 1709; repr. *Opera philosophica/Sämtliche Schriften*, vol. 3, 1-271.

Geulincx, A. (1688) 'Physica vera', in C. Bontekow, *Metaphysica*, Leiden; repr. *Opera philosophica/Sämtliche Schriften*, vol. 2, 368-446.

Geulincx, A. (1690) *Annotata praecurrentia ad R. Cartesii Principia*, Dordrecht.

Geulincx, A. (1691a) *Annotata majora in Principia philosophiae Renati Des Cartes*, Dordrecht; repr. *Opera philosophica/Sämtliche Schriften*, vol. 3, 361-521.

Geulincx, A. (1691b) *Metaphysica vera et ad mentem peripateticam*, Amsterdam; repr. *Opera philosophica/Sämtliche Schriften*, vol. 2, 137-265.

References and further reading

Cassirer, E. (1911) *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit*, Darmstadt: Wissenschaftliche Buchgesellschaft, 1971-3, vol. 1, 532- 43.
Cooney, B. (1972) *The Development of Cartesian metaphysics: Descartes, Malebranche, Geulincx*, Ph.D. dissertation, McGill University, Montreal.
Cooney, B. (1978) 'Arnold Geulincx: A Cartesian idealist', *Journal of the History of Philosophy* 16: 167-80.
de Lattre, A. (1966) *L'occasionalisme d'Arnold Geulincx*, Paris: éditions de Minuit.
De Vleeschauwer, H.J. (1957) *Three centuries of Geulincx research*, Communications of the University of South Africa, vol. 1, Pretoria.
De Vleeschauwer, H.J. (1961) *More seu ordine geometrico demonstratum*, Communications of the University of South Africa, vol. 48, Pretoria.
De Vleeschauwer, H.J. (1965) *Le problème du suicide dans la philosophie morale d'Arnold Geulincx*, Communications of the University of South Africa, vol. 54, Pretoria.
Dürr, K. (1939-40) 'Die mathematische Logik des Arnold Geulincx', *Erkenntnis* 8: 361-68.
Dürr, K. (1965) 'Arnold Geulincx und die klassische Logik des 17, Jahrhunderts', *Studium generale* 18: 520-41.
Land, J.P.N. (1895) *Arnold Geulincx und seine Philosophie*, The Hague: Martinus Nijoff.
Malbreil, G. (1974) 'L'occasionalisme d'Arnold Geulincx', *Archives de Philosophie* 37: 77-105.(Based on Alain de Lattre).
Monchamp, G. (1886) *Histoire du cartésianism en Belgique*, Brussels, 219-46.
Nuchelmans, G. (1983) *Judgment and proposition from Descartes to Kant*, Amsterdam, Oxford and New York: North Holland, 99-120.
Nuchelmans, G. (1984) *'Geulincx' karakterisering van relationele predikaten en van voegwoorden. Mededelingen van de Koninklijke Akademie voor Wetenschappen, Letteren en Schone Kunsten van Belgie. Klasse der Letteren*, Brussels, 73-89.
Nuchelmans, G. (1986) *Geulincx' containment theory of logic*, Amsterdam, Oxford and New York: North Holland.
Vander Haeghen, V. (1886) *Geulincx: étude sur sa vie, sa philosophie et ses ouvrages*, Gent.

Gilbert of Poitiers (*c*.1085-1154)

Gilbert's most important work is his commentary on the theological treatises of Boethius. His contemporaries valued him not only as a theologian but also as a philosopher, especially as a logician. Their estimation was well-founded. Although today we possess only theological writings from his own hand, these allow us to reconstruct a body of rich and independent philosophical thought.

The most salient characteristic of Gilbert's thought is the precise, analytical reflection that he brought to bear on the linguistic and conceptual means by which we think about whatever exists. In Gilbert's thought, two things go hand in hand: a philosophy of the concrete and the particular and an intellectual viewpoint whose conceptual resources are manifestly Platonist. In the history of philosophy, these two things are not usually found together.

1 Life

Gilbert of Poitiers, named after the city of his birth, studied in Chartres and Laon and then taught theology in Chartres and in Paris. In 1142 he was made Bishop of Poitiers. At two Councils he had to defend his doctrine of the Trinity against charges of heresy. In both cases his learning, power of argumentation and self-assured presence were convincing, and his defence succeeded. He died, highly respected, in 1154 (see Chartres, School of).

Gilbert's influence survived for some time after his death, thanks to the efforts of his students. The school he founded (named apparently after the Porretain quarter in Paris, in which it was located, rather than after Gilbert himself) survived well into the second half of the twelfth century, teaching grammar, logic and theology according to the philosophical methods and concepts that Gilbert had introduced.

2 The multiformity of being

At the foundation of Gilbert's whole ontology lies the distinction between 'that which is' (*quod est*) and 'that through which a particular thing is' (*quo est*). Corresponding to this distinction, there are different kinds of linguistic expression. 'That which is' is referred to and described by concrete terms (such as 'human being', 'bodily, body', 'colourful, the coloured'). Every concrete term in turn has a correlated abstract term (such as 'humanity', 'corporeity', 'colour'). The abstract term designates 'forms', or determinations of that which is. Thus every being can be referred to and described in many different ways. Gilbert construes this fact as meaning that one can see each being as a concretion of many formal determinations; any being is multiform, in that it is constituted by a plurality of formal determinations. Conversely, forms are real only through their concretion; they have no independent existence.

Among the forms that determine any being, there are some that determine a given being with respect to its essence. These make up 'the being of something with respect to what it is' (*esse aliquid in eo quod quid est*) (Commentary on Boethius' *De hebdomadibus* 1, nn. 46-7, 11.255-62). Other forms also make a being be something particular, but they determine the being only with respect to some non-essential property (*esse aliquid tantum*), not with respect to what it is. Gilbert calls a being, considered with regard to its essential determinations, 'a subsistent something' (*subsistens*), and he calls these formal determinations 'subsistences' (*subsistentiae*), belonging to the subsistent something. Every subsistent something has several subsistences, for everything determines a being essentially that belongs to its definition or to the further explanation of a determination referred to in the definition. For each being, it is the task of the natural sciences to discern correctly the boundaries between the determinations belonging to the being's definition and those having to do with its merely accidental properties. The criterion for making this decision between the ways that a form determines a being is the regularity with which a given being has certain properties (*consuetudo rerum*). It is generally possible that some formal determination that is a subsistence in relation to one being is accidental in relation to another.

Gilbert distinguishes sharply between designations or descriptions that apply to something in and of itself, and those that characterize it in comparison with something else. Aside from the subsistences of a thing, only its qualitative and quantitative determinations pertain to the thing's own being. Determinations of space, time and position, as well as determinations of how a thing acts or is affected and of what a thing possesses, always characterize a thing in relation to other things. Such determinations are relational, pertaining to the circumstances in which a thing exists.

3 The individual

Anything that subsists is a singular thing (*singulare*); and every subsistence, and every quality and quantity, is also single. Any single entity is different from every other. Gilbert adheres to this proposition closely. When one and the same subsistence, or one and the same property, belongs to different things, then these things are 'the same in form' (*conformes*) with regard to each subsistence or property: but the form, considered as the form of some one thing, is different from the form of some other thing even though it is the same in content. Every human being is a human being through its own humanity; every white thing is white through its own whiteness.

In Gilbert's view, the singularity of form is the basis of the singularity of that which is formed by it (Commentary on Boethius' *De trinitate* I, 5, n. 23, 1.162). Consequently, when Gilbert considers what individuality consists in, he takes forms, not what is formed, as his starting point. Formal determinations are either simple or complex (for example, humanity is complex compared with being rational since rationality is a conceptual element of humanity.) All simple formal determinations are suited to determine equally a number of subsistent things. Gilbert calls such forms 'divisible among many' (*dividua*). If two things *a* and *b* are the same in form with respect to the form *F*, then *F* is 'divisible' with respect to *a* and *b* (and probably with respect to other things as well). A complex formal determination could be represented by a conjunction of expressions standing for single forms (*F* and *G* and…) or, considered as the form of *a*, by the conjunction (*F(a)* and *G(a)* and…). Such a formal complex remains divisible as long as it can be completed by further formal elements and, while it is not complete, the complex taken as a whole could apply to other things. However, the complete formal determination, to which nothing further can be added, is not divisible. With respect to it, no two things are the same in form. Formal determination is thus the basis of individuality: an individual, determined in every respect by its complete form, may be similar to other individuals in many ways, but as a whole it is dissimilar from every other individual.

Finite reason can never grasp the individuality of an individual with regard to its content, because finite reason is discursive (that is, it proceeds by joining together partial determinations of things). Any content attainable by finite reason is always partial and therefore 'divisible'. When we want to refer unmistakably to precisely one individual, we have surrogates of complete understanding, such as dating and locating. But individuals are not individuated by their place in space and time; rather, they are individuated by their complete form. The complete form is inherently unrepeatable, and thus individual things are unrepeatable. This proposition is indemonstrable; it is a first principle of Gilbert's ontology.

4 Theory of knowledge

Gilbert distinguishes three branches of theoretical knowledge: physics, mathematics, and theology. This distinction is common in medieval philosophy. Its ultimate derivation is from Aristotle's *Metaphysics*, Book E, recalled by Boethius in the first of his treatises on the Trinity. However, in the twelfth century, at the time that Gilbert was commenting on Boethius, Aristotle's *Metaphysics* itself was not yet known in the West.

In his interpretation of the distinction, Gilbert arrives at a special understanding of the three sorts of theoretical knowledge. His first step is to take up objects of knowledge. We consider objects that are either concrete or abstract, and we consider what is abstract when we consider the first principles of all being in their simple being in themselves. Gilbert often speaks of unformed matter and of form which is not the form of some matter as principles. For Gilbert, however, there is ultimately just a single 'separated' first principle of being, God. Theology is the science whose object is the non-concrete.

Gilbert's second step is to take up different ways of knowing. That which is concrete can be considered either as it is, or in an abstract way (*abstractim*) when we attend only to forms, disregarding matter. The first way Gilbert calls 'natural consideration', the second he calls 'mathematical consideration' (or sometimes also *mathesis* or *disciplina*). It should be noted that *mathesis* concerns not quantities but abstracted forms.

The first way of knowing considers natural things in a 'natural' way and describes things as they present themselves in experience. What is concrete is complex and multiform; its formal aspects are understood as they are found in their concretion. This is reflected in the language of natural science. In its statements, which are descriptive, there are concrete terms in both subject position and predicate position, though a predicate term can be filled out by an abstract term in the ablative case (for example, 'Socrates is a human being *on account of*

humanity', or 'something is white *on account of whiteness*').

Mathesis does not have its own sphere of objects, separable with regard to content from the objects of natural science. *Mathesis* is special because it is a certain way of knowing, consisting in conceptual analysis. Its inquiry concerns the meaning of concepts such as corporeity or life. Thus concepts of forms are freed from the concrete state of affairs with which they are connected. *Mathesis* concerns itself with the question of what other concepts are implied by the range of meaning of a certain concept: with what other concepts is a given concept compatible or incompatible? It is mathematical rather than natural consideration of objects that leads to general concepts under which other concepts are ordered. Ultimately, such consideration leads to the Aristotelian categories. Gilbert emphasizes explicitly that the categories are referred to more adequately by abstract than by concrete terms (Commentary on Boethius' *De trinitate* I, 4, nn. 7-21).

Natural and mathematical consideration must not be mixed together. Nor must concrete and abstract terms be mixed together in statements, for instance, by putting a concrete term in subject position and an abstract term in predicate position or vice versa. The descriptive consideration *of things* and the consideration that develops theories *about the concepts of forms* are different and complementary.

5 The nature of theological discourse

Theological discourse is reflective and has no special language appropriate to it. We put together statements using concrete or abstract terms, but none of these statements conveys God's simplicity (see Simplicity, divine). This inappropriateness is something about which the theologian or metaphysician thinks, but he cannot make statements that are in fact appropriate to the object under consideration.

From Gilbert's writings, we can infer a methodology of speculative theology. This methodology consists in precisely regulated transpositions (*transsumptio*) of ways of using expressions and making statements. The ways of using expressions and making statements cannot themselves be altered, for we cannot speak in any other ways than those resulting from natural or mathematical consideration, using concrete and abstract terms. Gilbert's theory of transpositions is a theory about the sort of altered sense that statements of an ordinary form must be understood as having when they are used in theological discourse. One of the first transpositions is that of word meaning. The predicate 'is powerful' has a different meaning when applied to God than when applied to a king. When God is said to be just, this is not to be understood as a determination of a quality; nor, when God is said to be greatest, is this to be understood as the determination of a quantity (Commentary on Boethius' *De trinitate* I, 4, n. 28, 11.161-9).

More important is the transposition of the meaning of predication. Ordinarily, when we say of something that it is, or that it is powerful, or that it is wise, we are making very different statements. Recourse to the abstract concepts of existence, reality, power or wisdom serves to make clear the differences between these formal determinations. However, God must not be thought of as multiform. When it is said of God that he exists, or is powerful, or is wise, these statements are just various approximations - from the human standpoint - regarding the being who is simple. Every statement about God must be understood as expressing its object as a whole. Whereas ordinarily we must understand and evaluate the statements that we made about something separately (*divisive*), in theology we must understand all statements as conjoined and interrelated (*coniuncte et copulate*), in other words, in such a way that each statement is thought of as including all other statements (Commentary on Boethius' *De trinitate* I, 4, nn. 37-48).

When reflecting on the inappropriateness of our talk about God, Gilbert occasionally makes statements of the following form, which he otherwise rejects: 'God is not powerful but rather power itself'. It is a mistake to think that in statements of this form, one has discovered the special language appropriate to theological discourse. Such statements are only admissible in the development of theological inquiry, arising as our thought about God corrects itself, and should not be taken as fixed results of theological consideration. Abstract terms ordinarily are indications of a possible conformity among beings; but there is no conformity between God and anything else.

Gilbert sees theology as a reflection on that which is primordial, that is, as providing a theory of first principles. However, such a theory must also teach us to see that which has come into being (that which derives from first principles) in a new light. The path of human thought leads from what has come into being to what is primordial, and it also leads back again, connecting our consideration of the primordial to our concepts of what has come into

being. Although the primordial is incomprehensible, our thinking about it changes our understanding of what has come into being.

In Gilbert's view, there is not just one transposition; there is also another, complementary transposition leading from thought about God to a new understanding of the natural world. Gilbert demonstrates this reversal of the direction of transposition in the case of the predicate 'good'. He explains that the proposition that every being is good insofar as it is a being lies beyond any natural consideration, springing instead from theology. Because God is good insofar as he is, everything created by him is good. In its primary sense, therefore, 'is good' is predicated of God, and when we say of a creature that it is good, we are using a derivative designation (*denominatio*) (Commentary on Boethius' *De hebdomadibus* 2, nn. 150-6).

It is clear that construing the predicate 'is good' along these lines is not just an *ad hoc* idea, for Gilbert interprets each of the axioms that Boethius presents in his treatise *De hebdomadibus* in two ways (with the exception of the axiom concerning simplicity, which can only be understood theologically). Thus being (*id quod est*) should be treated in the following two ways: theologically speaking, God is the being of everything; ontologically speaking, the totality of its subsistences is the being of each existent thing. Theologically speaking, then, God alone is, in the full sense of the word 'is'. Each particular thing is just that particular thing, and is not something else that also is. Given the historical context, it is surprising that in his theological works Gilbert construes the Boethian axioms ontologically as well as theologically. Indeed, Gilbert apparently takes special interest in the ontological interpretation, one which is clearly not demanded by Boethius' text. From a philosophical perspective, the converse is equally noteworthy: in Gilbert's view, to speak of the being of natural beings is to speak of 'being' and 'beings' in respect of a certain derivative designation (*denominatio*).

See also: Aristotelianism, medieval; Being; Boethius, A.M.S.; Chartres, School of; Language, medieval theories of (§8); Natural theology

Translated from the German by Hannes Jarka-Sellers
KLAUS JACOBI

List of works

Gilbert of Poitiers (*c*.1085-1154) Commentaries on Boethius, ed. N.M. Häring, *The Commentaries on Boethius by Gilbert of Poitiers*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1966.(Includes Gilbert's commentaries on *De hebdomadibus* and *De trinitate*.)

Gilbert of Poitiers (*c*.1085-1154) *Compendium logicae (Compendium of Logic)*, ed. S. Ebbesen, K.M. Fredborg and L.O. Nielsen, *Compendium Logicae Porrentanum ex codice Oxoniensi collegii Corporis Christi 250: A Manual of Porretan Doctrine by a Pupil of Gilbert's*, Cahiers de l'Institut du Moyen-Âge Grec et Latin 57, 11-67, Copenhagen: Paludan, 1983.(Gilbert's work on logic.)

References and further reading

Elswijk, H.C. van (1966) *Gilbert Porreta, Sa Vie, son oeuvre, sa pensée (Gilbert of Poiters: His Life, Work and Thought)*, Louvain: Spicilegium sacrum Lovaniense.(Reliably informative about Gilbert's life and work.)

Gracia, J.J.E. (1984) *Introduction to the Problem of Individuation in the Early Middle Ages*, Munich: Philosophia Verlag, 155-78, 187-93.(Systematic investigation of the history of the problem, carried out on a high analytical level.)

Jacobi, K. (1995) 'Einzelnes-Individuum-Person. Gilbert von Poitier's Philosophie des Individuellen' (Particular-Individual-Person: Gilbert of Poitier's Philosophy of Individuals), in J. Aertsen and A. Speer (eds) *Individuum und Individualität im Mittelalter*, Miscellanea Mediaevalia 23, Berlin: de Gruyter.(An elaboration of the material in §2 above.)

Maioli, B. (1979) *Gilberto Porretano. Della grammatica speculativa alla metafisica del concreto (Gilbert of Poitiers: From Speculative Grammar to the Metaphysics of the Concrete)*, Biblioteca di cultura 173, Roma: Bulzoni Editore.(A good presentation of Gilbert's philosophy, taking into account the historical context of this thought.)

Nielsen, L.O. (1982) *Theology and Philosophy in the Twelfth Century: A Study of Gilbert Porreta's Thinking and the Theological Exposition of the Doctrine of the Incarnation during the Period 1130-1180*, Leiden: Brill.(Part I, *Gilbert Porreta's Life, Works and Thought* begins with an informative biography at the cutting edge of Gilbert scholarship. It is followed by a well thought-out presentation of Gilbert's metaphysics, theory of science, philosophy of language and theology.)

Rijk, L.M. de (1988-9) ‘Semantics and Metaphysics in Gilbert of Poitiers: A Chapter of Twelfth-Century Platonism’, *Vivarium* 26 (2): 73-112; 27 (1): 11-35.(Important excerpts from Gilbert’s commentaries are carefully selected, translated and commented on, in accordance with the central concepts and themes of Gilbert’s philosophy.)

Schmidt, M.A. (1956) *Gottheit und Trinität, nach dem Kommentar des Gilbert Porreta zu Boethius, De Trinitate (Godhead and Trinity in Gilbert of Poitiers’ Commentary on Boethius’ De trinitate)*, Studia Philosophica, Jahrbuch der Schweizer Philosophischen Gesellschaft, Suppl. 7, Basel: Verlag für Recht und Gesellschaft.(A study of Gilbert’s theology, full of sound philosophical observations.)

Giles of Rome (*c*.1243/7-1316)

Giles of Rome was one of the most eminent theologians and commentators on the works of Aristotle at the University of Paris in the second half of the thirteenth century. He was probably a pupil of Thomas Aquinas, who exerted a deep influence on Giles' metaphysical and theological thought. Giles' reception of Aquinas' positions, however, was often critical and original.

*For historians of medieval philosophy, Giles' name is mainly associated with the doctrine of 'the real distinction' between essence (*essentia*) and existence (*esse*). According to this doctrine, essence and existence are two completely distinct things (*res*) of which the ontological structure of every created being is composed. On the issue of the relationship between essence and existence Giles took a firm position against his contemporary Henry of Ghent, who maintained that existence is a mere relation of the essence of a created being to its creator. Giles was also involved in the debate over the unity of the substantial form in composite substances, another burning issue in the thirteenth century. As a commentator on Aristotle's works, Giles made original contributions to the tradition of Aristotelian natural philosophy, especially in his treatment of extension, place, time and motion in a vacuum.*

1 Life

Giles of Rome was born, probably in Rome, in the fifth decade of the thirteenth century. It seems that he entered the Augustinian Order at Rome. In order to study philosophy and theology he was sent to Paris, where he probably attended Thomas Aquinas' lectures during the years 1269-72. Giles started his academic carrier as a bachelor of theology at the university of Paris around 1270, and in the period between 1270 and 1277 he lectured on the four books of the *Sentences* by Peter Lombard and wrote most of his commentaries on Aristotle's works.

In 1277 he was one of those whose views were attacked in the condemnation of heterodox Aristotelianism initiated by the bishop of Paris, Etienne Tempier (see Aristotelianism, medieval). Three main reasons for Giles' condemnation have been detected: first, the similarities between some of his doctrinal positions and those of the masters of arts who were the principal targets of the condemnation (for example, Siger of Brabant); second, his critical attitude towards some masters of theology, especially Henry of Ghent; and third, his defence of some Thomist positions, such as those about the unity of the substantial form and the possibility of an eternal world (see Aquinas, T.; Eternity of the world, medieval views of). As a reaction against the condemnation, he wrote the sharply polemic treatise *Contra gradus et pluralitatem formarum (Against the Degrees and Plurality of Forms)*, in which he argued that the doctrine of the unity of substantial form is not only sound but also the only one in agreement with faith.

Little is known about Giles' activity in the years immediately after 1277. He probably left Paris and returned to Italy. During this period he wrote the *De regimine principum (On the Governance of Princes)*, dedicated to the future French king Philip the Fair, whose tutor he is said to have been, possibly in the period between 1277/8 and 1281.

In 1285 Giles returned to Paris, where he was appointed master of theology. His activity in this position is mainly represented by six *Quodlibeta* discussed between 1286 and 1291 and by several collections of *Quaestiones disputatae (Disputed Questions)*. The most heated phase of Giles' polemic against Henry of Ghent on the issue of the distinction between essence and existence also dates to this period. In 1287 his teachings became the official doctrine of the Augustinian Order, of which he was elected prior general in 1292.

In 1295, Pope Boniface VIII appointed Giles archbishop of Bourges. He took an active part in the ecclesiastical life of that period. On the occasion of the controversy over the legitimacy of Boniface VIII's election, Giles wrote the *De renuntiatione papae (On the Abdication of the Pope)*, in which he took the pope's side. In connection with the conflict between Boniface VIII and Philip the Fair, he wrote the *De ecclesiastica potestate (On the Ecclesiastical Power)*, a work which exerted a strong influence on the pope. Giles died in Avignon on 22 December 1316.

2 Metaphysics

Giles' philosophical and theological thought is much influenced by the writings of Aquinas, but his views do not faithfully express those of Aquinas. On the contrary, Giles often criticizes Aquinas' positions, although in most cases his criticisms aim at correcting and integrating Aquinas' doctrines rather than rejecting them completely. Furthermore, Giles' methodological approach to philosophical problems is not the same as that of Aquinas. Giles pays more attention to the logical structure of argumentation and linguistic analysis, and also tries to formulate more general solutions.

Giles' metaphysics is mainly characterized by two doctrines: the real distinction between essence and existence and the unity of the substantial form. As to the first doctrine, although Giles finds a distinction between essence and existence in Aquinas' writings, he is particularly original in his interpretation of this distinction. Indeed, Giles' view on this topic is much more clear-cut and radical than that of Aquinas, for he maintains that essence and existence are two completely distinct things (*res*), of which the ontological structure of any created being is composed. According to Giles, in order to account for the passage of an entity from potentially existing to actually existing, it is necessary to regard existence as a *res* which is added to the essence of this entity. On the basis of this account of coming into being, Giles strongly criticizes and rejects Henry of Ghent's notion of existence as a mere relation of the created essence to the creator. Despite Giles' refusal to identify existence with an accidental form, the essence-existence relationship in his treatment of it turns out to be very similar to the Aristotelian substance-accident relationship (see Aristotle). This implies in particular that essence is ontologically prior to existence. In this respect, Giles' position is closer to an Avicennian than to a Thomist position (see Existence; Ibn Sina).

Giles' position on the issue of the unity of the substantial form is not perfectly uniform. In the *Errores philosophorum (Errors of the Philosophers)* he criticizes Aristotle's theory of the unity of the substantial form in every composite substance (however, Giles' authorship is not certain). On the contrary, in the *Contra gradus et pluralitatem formarum* he denies the plurality of forms in any composite. In his later writings, however, he seems to be more concerned with the theological implications of the doctrine as applied to the human being, and adopts a more prudent view, holding that there is only one substantial form in any composite with the exception of the human being, and leaving the question open in this latter case.

The nature of individuation is another important topic in Giles' metaphysics. Following Aquinas, he maintains that prime matter is pure potentiality and that therefore a quantitative principle must be added to matter in order to account for the multiplication of the substantial form. This quantitative principle is described by Giles in terms of Averroes' notion of indefinite dimension (*dimensio indeterminata*), where the adjective 'indefinite' reflects the fact that this dimension exists in matter prior to the substantial form (see Ibn Rushd). In replying to some objections against the priority of extension over substance resulting from his account of individuation, Giles specifies that the principle of individuation is not quantity conceived of as a thing distinct from matter, but a quantitative mode (*modus quantitativus*) of being which indefinite dimensions confer upon matter (see Matter).

3 Theory of soul and of knowledge

Giles' psychology and epistemology reflect mainly Aristotelian and Thomistic principles, with no special inclinations towards Augustinian or Averroistic positions. As to the theory of soul, for instance, Giles follows Aquinas' view on the real distinction between the soul and its faculties. In theory of knowledge, Giles emphasizes the Aristotelian principle that any knowledge starts from sense experience. Accordingly, he rejects the Platonic theory of innate ideas as well as the Augustinian theory of illumination (see Augustinianism; Platonism, medieval). While the starting point of intellectual knowledge is the sensory cognition of sensible substances, its proper object is the immaterial quiddity of these substances. Therefore, intellectual knowledge also requires a process of abstraction from material conditions, in which, according to Giles, both the agent intellect (*intellectus agens*) and an intelligible species (*species intelligibilis*) play an essential role (see Knowledge, concept of).

As to the theory of truth, while in Aquinas both the Augustinian-Anselmian and the Aristotelian traditions are combined, Giles firmly adopts the latter. That is, he maintains that truth is not a property inhering in extra-mental things independently of any cognitive act, but a correspondence between these things and the propositions formed by the intellect (*adaequatio intellectus et rei*).

Regarding the issue of the unicity of the intellect, Giles strongly rejects both the doctrine of the unicity of the agent

intellect, which he ascribes to Avicenna (see Ibn Sina §3) and the Averroistic doctrine of the unity of the possible intellect (see Averroism) In Giles' view, individual knowledge and the ontological status of the intellect as an intermediate entity between material and immaterial forms can be accounted for only if rational soul is the form of the body and is multiplied by inhering in different bodies.

4 Natural philosophy

Giles' natural philosophy is still basically Aristotelian. His most systematic and extensive treatment of physical concepts is contained in his commentary on Aristotle's *Physics*, which was widely used by later commentators. In addition to a careful exegesis of Aristotle's text, Giles also presents original positions on some of the fundamental issues of Aristotelian natural philosophy, as for instance extension, place, time and the eternity of the world.

In order to account for the processes of condensation and rarefaction, Giles postulates the existence of two independent quantities inhering in a physical body: the corporeal dimensions, which vary when the body is condensed or rarefied, and the quantity of matter (*quantitas materiae*), which does not vary. This distinction has strong similarities to the modern distinction between volume and mass.

Giles also distinguishes between material place (*locus materialiter*) and formal place (*locus formaliter*). Material place is place in the Aristotelian sense, that is, the inner surface of the containing body. Formal place is an order (*ordo*) or a distance (*distantia*) between the located body and the fixed points of the universe. Following Aquinas, Giles introduces the notion of formal place in order to account for the immobility of place, but he departs from Aquinas regarding the relationship between material place and formal place. In Giles' view, formal place is not a property of material place, but something independent of it. His motivation for separating formal from material place is to be found in the distinction between two roles that the notion of place must play: first, a principle of delimitation of the extension of the located body; and second, a frame of reference for describing the body's motion and rest. The first requirement is adequately met by material place, the second by formal place.

Although Giles agrees with most of Aristotle's presuppositions about time, he also admits the simultaneous existence of many temporal durations and the existence of a discrete time, that is, of a succession of instants conceived of as temporal atoms rather than as mere cuts in a temporal continuum, whereas Aristotle holds that there is just one time and that it is a continuum. In particular, the notion of a discrete time plays an important role in Giles' treatment of motion in the void, since he argues that such a motion would not take place instantaneously, as Aristotle and Averroes claim, but in a succession of instants, that is, in a discrete time.

On the issue of the eternity of the world, in his early writings Giles follows a Thomistic position. On the one hand, he claims that Aristotle's and Averroes' arguments for an eternal world are not conclusive, since they falsely assume that any production involves motion; on the other hand, however, he admits that an eternal world is at least theoretically possible. After the condemnation of 1277, Giles takes a more moderate position, according to which it is theoretically possible to prove that the world had a temporal beginning, although adequate arguments for this conclusion have not yet been proposed (see Eternity of the world, medieval views of; Natural philosophy, medieval).

5 Theology

Giles' conception of God centres on the notion of divine unity, which he regards as the primary divine attribute. The priority of divine unity is particularly evident in his approach to trinitarian problems, in which he stresses the supremacy of the unity of the divine essence over the multiplicity of the divine relations. Giles' emphasis on the unity of God, however, does not imply that divine relations are weakened. On the contrary, his doctrine of divine relations is much more complex and articulated than that of Aquinas. Giles' doctrine is based on a distinction between three kinds of relations (opposite, disparate and similar), which turns out to be a very powerful conceptual tool for dealing with some traditional problems of trinitarian theology, such as those concerning the priority-posteriority relationship among divine relations, their real distinction and their multiplication.

Among divine attributes, infinity ranks second only to unity. The importance of the infinity of God appears especially in Giles' discussion of the problem of the object of theology. Contrary to Aquinas, he constantly remarks that the object of theology cannot be God as God, since our finite intellect cannot grasp the infinity of God. Instead, the object of theology is to be identified with God regarded as the 'principle of our restoration and

the completion of our glorification' (*principium nostrae restaurationis et consummatio nostrae glorificationis*), according to the formula which characterized Giles' position in a long lasting debate in which Henry of Ghent and Godfrey of Fontaines were also involved. Giles' emphasis on the wide gap between finite creatures and infinite God does not lead him to a sceptical conclusion about the possibility of rational knowledge of the divine nature. He shows, for instance, an optimistic attitude towards our knowledge of the Trinity, since he maintains that natural reason can prove that the arguments against the Trinity of the divine persons are not conclusive (see God, concepts of; Trinity).

Finally, Giles was also a preacher. His preaching activity, as attested to by a collection of seventy-six sermons written by Giles himself, is one of the most original in the thirteenth century, since it is characterized by strong doctrinal commitments and is closely related to his theological works. Indeed most of Giles' sermons can be considered as short theological treatises and represent an important source for a final assessment of his theological thought.

6 Politics

Giles' political thought centres on the theory of the supremacy of spiritual power over temporal power. In the *De regimine principum* he holds that the state originates from the natural tendency of human beings to live in a civil society, but in the *De ecclesiastica potestate* he also stresses that the state must be subordinated to spiritual power. In Giles' view, the distinction between temporal and spiritual power is based on the distinction in the human being between body and soul, but the subordination of temporal to spiritual power is based on the order in the universe, according to which only spiritual power derives directly from God, whereas temporal power is dependent upon the spiritual.

See also: Aquinas, T.; Aristotelianism, medieval; Augustinianism; Godfrey of Fontaines; Henry of Ghent §2; Natural philosophy, medieval

FRANCESCO DEL PUNTA
CECILIA TRIFOGLI

List of works

Giles' works can be chronologically divided into four chronological periods: before 1277/8, 1278-85, 1286-95 and 1296-1316. The main works belonging to each of these groups are listed below, after the complete works.

Giles of Rome [Aegidius Romanus] (before 1271-after 1309) *Aegidii Romani Opera Omnia (Complete Works of Giles of Rome)*. In addition to the volume containing the edition of Giles' *Apologia* and the one containing the *Repertorio dei sermoni*, noted above, the following volumes of *census* and description of the manuscripts have been published: *Aegidii Romani Opera Omnia*, Catalogo dei manoscritti: I, 1, 1 (Vatican), a cura di B. Faes de Mottoni and C. Luna, Florence: Olschki, 1987; I, 1, 2* (Florence, Padua, Venice), a cura di F. Del Punta and C. Luna, Florence: Olschki, 1989; I, 1, 3* (Dipartimenti), a cura di F. Del Punta and C. Luna, Florence: Olschki, 1987; I, 1, 3** (Paris), a cura di C. Luna, Florence: Olschki, 1988; I, 1, 5* (Munich), a cura di B. Faes de Mottoni, Florence: Olschki, 1990; I, 1, 11 (*De Regimine Principum*) (Vatican), a cura di F. Del Punta and C. Luna, Florence: Olschki, 1993.

1 Before 1277/8

Giles of Rome [Aegidius Romanus] (before 1271) *Super librum I Sententiarum, reportatio (On the First Book of the Sentences)*, ed. C. Luna (1990), 'Fragments d'une reportation du commentaire de Gilles de Rome sur le premier livre des *Sentences*. Les extraits des mss. Clm. 8005 et Paris, B. N., Lat. 15819', *Revue des sciences philosophiques et théologiques* 74: 234-54, 446-56.(Giles' commentary on the first book of the *Sentences* of Peter Lombard.)

Giles of Rome [Aegidius Romanus] (before 1271) *Super librum II Sententiarum, reportatio (On the Second Book of the Sentences)*.(An edition of this recently discovered work is in preparation.)

Giles of Rome [Aegidius Romanus] (before 1271) *Super librum III Sententiarum, reportatio (On the Third Book of the Sentences)*, ed. C. Luna (1990), 'La *Reportatio* della lettura di Egidio Romano sul libro III delle *Sentenze* e il problema dell'autenticità dell'*Ordinatio*', *Documenti e studi sulla tradizione filosofica medievale* 1 (1): 179-225.(Giles' commentary on the third book of the *Sentences* of Peter Lombard.)

Giles of Rome [Aegidius Romanus] (before 1271) *Super librum IV Sententiarum, reportatio (On the Fourth Book of the Sentences)*, ed. C. Luna (1990), 'La lecture de Gilles de Rome sur le quatrième livre des *Sentences*. Les

extraits du Clm. 8005', *Recherches de théologie ancienne et médiévale* 57: 216-54.(Giles' commentary on the fourth book of the *Sentences* of Peter Lombard.)

Giles of Rome [Aegidius Romanus] (*c*.1269-72/3) *Quaestiones metaphysicales, reportatio (Metaphysical Questions)*, Venice, 1501.(Edition of questions on metaphysics.)

Giles of Rome [Aegidius Romanus] (*c*.1271-3) *Super librum I Sententiarum, ordinatio (On the First Book of Sentences)*, Venice, 1521.(Further commenary on the *Sentences* of Peter Lombard.)

Giles of Rome [Aegidius Romanus] (*c*.1271-4) *Super libros Rhetoricorum (On Aristotle's Rhetoric)*, Venice, 1515.(Commentary on Aristotle.)

Giles of Rome [Aegidius Romanus] (*c*.1274) *Super De generatione et corruptione (On Aristotle's On Generation and Corruption)*, Venice, 1505.(Commentary on Aristotle.)

Giles of Rome [Aegidius Romanus] (*c*.1274) *Quaestiones super librum I De generatione et corruptione (Questions on the first book of Aristotle's On Generation and Corruption)*, Venice, 1505.(Questions on Aristotle.)

Giles of Rome [Aegidius Romanus] (*c*.1274) *Theoremata de Corpore Christi (Theorems about Christ's Body)*, Rome, 1554.(Metaphysical work.)

Giles of Rome [Aegidius Romanus] (*c*.1274-5) *Super libros Elenchorum (On Aristotle's Sophistical Refutations)*, Venice, 1496.(Commentary on Aristotle.)

Giles of Rome [Aegidius Romanus] (*c*.1274-5) *Super Physicam (On Aristotle's Physics)*, Venice, 1502. (Commentary on Aristotle.)

Giles of Rome [Aegidius Romanus] (after 1274) *De plurificatione intellectus possibilis (On the Multiplication of the Possible Intellect)*, Venice, 1500; ed. H. Bullotta, Rome: Barracco, 1957.(Giles on the intellect.)

Giles of Rome [Aegidius Romanus] (before 1275) *Errores philosophorum (Errors of the Philosophers)*, ed. J. Koch, Milwaukee, WI, 1944.(Criticizes Aristotle on the unity of the substantial form.)

Giles of Rome [Aegidius Romanus] (*c*.1276) *Super De anima (On Aristotle's On the Soul)*, Venice, 1500. (Commentary on Aristotle.)

Giles of Rome [Aegidius Romanus] (1277-8) *Contra gradus et pluralitatem formarum (Against the Degrees and Plurality of the forms)*, Venice, 1500.(Argues that the doctrine of the unity of substantial form is in agreement with faith.)

2 1278-85

Giles of Rome [Aegidius Romanus] (*c*.1280) *De regimine principum (On the Governance of Princes)*, Rome, 1607.(Gives the view that the state originates from the natural tendency of human beings to live in a civil society.)

Giles of Rome [Aegidius Romanus] (1281) *Quaestiones disputatae in capitulo generali Paduae (Questions Disputed at the General chapter of Padua)*, ed. G. Bruni (1939), *Analecta Augustiniana* 17: 125-50.(Edition of disputed questions.)

Giles of Rome [Aegidius Romanus] *Theoremata de esse et essentia (Theorems on Being and Essence)*, ed. E. Hocedez, Louvain: Musuem Lessianum, 1930.(Metaphysical theorems.)

3 1286-95

Giles of Rome [Aegidius Romanus] (*c*.1286-7) *Quaestiones de esse et essentia (Questions on Being and Essence)*, Venice, 1503.(Metaphysical questions.)

Giles of Rome [Aegidius Romanus] (before 1288) *De materia caeli contra Averroistas (On the Matter of the Heavens, Against Averroists)*, Venice, 1500.(Treatise against the Averroists.)

Giles of Rome [Aegidius Romanus] (*c*.1287-8) *Quaestiones de cognitione angelorum (Questions on the Angels' Knowledge)*, Venice, 1503.(Early edition; no modern edition exists at present.)

Giles of Rome [Aegidius Romanus] (*c*.1288-9) *Quaestiones de mensura angelorum (Questions on the Measure of the Angels)*, Venice, 1503.(Early edition; no modern edition exists at present.)

Giles of Rome [Aegidius Romanus] (*c*.1287-90) *Quaestiones de compositione angelorum (Questions on the Composition of the Angels)*.(No published edition exists. The work can be found in manuscript at Toulouse, Bibl. Mun., ms. 739, fols. 96ra-103vb.)

Giles of Rome [Aegidius Romanus] (1286-91) *Quodlibeta*, Louvain, 1646.(Edition of Giles' quodlibetal questions.)

Giles of Rome [Aegidius Romanus] (*c*.1289-91) *Super De causis (On the Book of the Causes)*, Venice, 1550. (Commentary on Aristotle.)

Giles of Rome [Aegidius Romanus] (after 1288) *Tractatus de subiecto theologiae (Treatise on the Object of Theology)*, Venice, 1504; ed. C. Luna (1991) 'Una nuova questione di Egidio Romano *De subiecto theologiae*', *Freiburger Zeitschrift für Philosophie und Theologie* 38: 129-58.(Theological treatise.)

Giles of Rome [Aegidius Romanus] (after 1288-9) *Quaestiones de motu angelorum (Questions on the Movement of the Angels)*, ed. G. Bruni (1939) *Analecta Augustiniana* 17: 22-66.(Metaphysical work.)

Giles of Rome [Aegidius Romanus] (after 1288-9) *Super Posteriora Analytica (On Aristotle's Posterior Analytics)*, Venice, 1488.(Commentary on Aristotle.)

Giles of Rome [Aegidius Romanus] (after 1288-9) *De formatione corporis humani in utero (On the Formation of the Human Body in the Uterus)*, Arimini, 1626.(Scientific work.)

Giles of Rome [Aegidius Romanus] (after 1288-9) *Quaestio de subiecto theologiae (Question on the Subject of Theology)*, ed. C. Luna (1990) 'Una nuova questione di Egidio Romano *De subiecto theologiae*', *Freiburger Zeitschrift für Philosophie und Theologie* 37: 429-39.(Edition of a theological question.)

4 1296-1316

Giles of Rome [Aegidius Romanus] (after May-June 1297) *De renuntiatione papae (On the Abdication of the Pope)*, Rome, 1554.(Sides with Pope Boniface VIII in the dispute over the latter's election.)

Giles of Rome [Aegidius Romanus] (finished before 1303) *Hexaemeron*, Rome, 1555.(No modern edition of this work exists at present.)

Giles of Rome [Aegidius Romanus] (1301-2) *De ecclesiastica potestate (On the Ecclesiastical Power)*, ed. R. Scholz, Weimar: Hermann Böhlaus Nachfolger, 1929.(On the conflict between Pope Boniface VIII and King Philip IV of France.)

Giles of Rome [Aegidius Romanus] (finished not before 1309) *Super librum II Sententiarum, ordinatio (On the Second Book of the Sentences)*, Venice, 1581.(Further work on the second book of Peter Lombard's *Sentences*.)

Giles of Rome [Aegidius Romanus] *Super librum III Sententiarum, ordinatio (On the Third Book of the Sentences)*, Rome, 1623.(Spurious. See C. Luna (1990-1) 'La *Reportatio* della lettura di Egidio Romano sul libro III delle *Sentenze* e il problema dell'autenticità dell'*Ordinatio*', *Documenti e studi sulla tradizione filosofica medievale* 1 (1): 113-78; 2 (1): 75-146.)

5 Undated and complete works

Giles of Rome [Aegidius Romanus] (n.d.) *Sermones (Sermons)*, partially published in C. Luna, *Aegidii Romani Opera Omnia*, I, 6, *Repertorio dei sermoni*, Florence: Olschki, 1990.(Edition of Giles' collected sermons.)

References and further reading

Bruni, G. (1933) 'Egidio Romano antiaverroista', *Sophia* 1: 208-19.(An assessment of Giles' anti-Averroistic position on the issue of the unity of the intellect.)

Conti, A.D. (1992) 'Conoscenza e verità in Egidio Romano', *Documenti e studi sulla tradizione filosofica medievale* 3 (1): 305-61.(A detailed account of Giles' theory of truth and of knowledge and of its relation to Aquinas' positions.)

Del Punta, F., Donati, S. and Luna, C. (1993) 'Egidio Romano', in *Dizionario biografico degli italiani* 42: 319-41.(The most up-to-date biography, with a full and comprehensive bibliography.)

Donati, S. (1988) 'La dottrina delle dimensioni indeterminate in Egidio Romano', *Medioevo* 14: 149-233.(An expansion of the material in §2 of this entry on individuation and in §4 on extension.)

Donati, S. (1990-1) 'Studi per una cronologia delle opere di Egidio Romano. I: Le opere prima del 1285. I commenti aristotelici', *Documenti e studi sulla tradizione filosofica medievale* 1 (1): 1-112; 2 (1): 1-74.(A systematic study of the relative and absolute chronology of Giles' commentaries on Aristotle's works before 1285.)

Lambertini, R. (1992) 'Tra etica e politica: la *prudentia* del principe nel *De regimine* di Egidio Romano', *Documenti e studi sulla tradizione filosofica medievale* 3 (1): 77-144. (An account of Giles' theory of practical wisdom in the *De regimine principum*.)

Luna, C. (1988) 'Essenza divina e relazioni trinitarie nella critica di Egidio Romano a Tommaso d'Aquino', *Medioevo* 14: 3-69.(An expansion of the material in §5 above on Giles' conception of God and of the trinitarian relations.)

Luna, C. (1991) 'Théologie trinitaire et prédication dans les sermons de Gilles de Rome', *Archives d'histoire doctrinale et littéraire du Moyen Age* 58: 99-135.(An analysis of the relationship between Giles' theological treatment of trinitarian problems and his preaching activity.)

Nash, P.W. (1950) 'Giles of Rome on Boethius' *Diversum est esse et id quod est*', *Mediaeval Studies* 12: 57-91.(A study on Giles' doctrine of the real distinction between essence and existence. An expansion of the material in §2 above on this topic.)

Trifogli, C. (1988) 'La dottrina del luogo in Egidio Romano', *Medioevo* 14: 235-90.(An expansion of the material in §4 above on place.)

Trifogli, C. (1990) 'La dottrina del tempo in Egidio Romano', *Documenti e studi sulla tradizione filosofica medievale* 1 (1): 247-76.(An expansion of the material in §4 above on time.)

Trifogli, C. (1992) 'Giles of Rome on Natural Motion in the Void', *Mediaeval studies* 54: 136-61.(An account of Giles' position in the debate on natural motion in the void.)

Trifogli, C. (1993) 'Giles of Rome on the Instant of Change', *Synthese* 96 (1): 93-114.(An account of Giles' doctrine of the change between contradictory states.)

Gioberti, Vincenzo (1801-52)

The work of Vincenzo Gioberti was a life-long attempt to reconnect philosophy and Christianity, and tradition and progress, within the political turmoil of early nineteenth-century Italy and the rise of new philosophies of history. His critique of subjectivism led him to propose a Neoplatonic scheme (epitomized in what he called the 'ideal formula'), which finds its root in an original intuition of being. From this intuition he deduced that the Being as the creator is God. But reflective judgment is not mere contemplation: as thinking and creating are the same in God (God is 'the first philosopher'), so thinking and acting are the same in man, as an image of God. History and civilization are the continuation of the creative process in which the return of existence to being leads duality to unity again, although it keeps the ontological gap of creatural relationship.

Gioberti was an outstanding intellectual of the Italian Risorgimento and the leader of the Neo-guelphic movement, whose aim was national unification under the Pope's leadership and national identity based on Catholicism. However, despite his involvement in active politics (he was both exiled and later became Prime Minister of Piedmont), his philosophical vocation was constantly pre-eminent.

Gioberti's philosophy took systematic shape in *Introduzione allo studio della filosofia (Introduction to the Study of Philosophy)* (1840). For him the main object of philosophy is the Idea, which cannot be considered as a mere mental representation or as a psychological construction. Gioberti's argument starts from a critique of Descartes' philosophy of the subject, which he calls 'psychologism'. This is 'the modern heterodoxy of philosophical sciences' which alters the legitimate notion of Idea, whose meaning expresses 'neither our concept, nor any other created thing or property, but the absolute and eternal truth, as it reveals itself to man's intuition' (1840 (2): 1). According to Gioberti, Descartes did in philosophy what Luther had done in theology. In both cases truth is derived from the state of the subject, from its intimate sense. The priority of the Idea is denied and what takes its place is a conceptual simulacrum, in which sensibilia and intelligibilia are mixed into a contradictory synthesis. Gioberti describes five possible developments of Descartes' position which correspond to five moments of modern philosophy: (1) Cartesianism itself; (2) Lockean empiricism; (3) Spinozism and German idealism; (4) Kantianism and French sensism; and (5) Humean scepticism. The order of classification finds its own consistency in the progressive shift of Descartes' subjective foundation towards the giving up of every ontological perspective until the Humean elision of the foundation itself. In this sequence, the third moment takes a central and recursive place. Between the permanence of an ontology which contradicts its own foundation, as in Cartesianism, and the negation of any ontology, as in Hume, this moment presents an attempt to construct a new ontology fully consistent with the new method and therefore totally different from traditional ontology.

Gioberti's argument is that the new method perverts the ontological outlook implied in all men by original intuition, and leads it to pantheism, that is, an ontology which confuses being and existence, through a reflection misled by subjectivism. As Spinozism was a necessary result of Cartesianism, so the necessary outcome of Kantian criticism was German idealism. Gioberti's notion of pantheism is very wide-ranging and covers several philosophical positions, for it summarizes the consequences of a deceptive approach to the relationship between intuition and reflection. Its real meaning is clarified by the polemic Gioberti elaborated in *Degli errori filosofici di Antonio Rosmini (On the Philosophical Mistakes of Antonio Rosmini)* (1841-43). Although Gioberti was close to Rosmini in his judgment of post-Cartesian philosophy, he could not accept his claim that being, which is the object of our intuition, is, first of all, ideal being. Rosmini's being is a psychological *primum* - a mental object and, therefore, a form of the human mind - and not an ontological *primum*: the latter should be founded on the former and their coincidence should be the aim of the reflection (see Rosmini-Serbati, A.). Against Rosmini, Gioberti asserts that the Idea itself is present as a true reality to intuition, whose passivity we have to acknowledge. Truth does not belong to a performance of mind, but to the Idea which becomes evident by penetrating the mind and making it partake of itself. In this way it is possible to escape psychologistic anthropocentrism: philosophy performs a reflective judgment which is really 'voluntary, subjective, human', although it finds legitimacy and objectivity because it is a repetition of intuitive judgment, which precedes, establishes and authorizes it.

This is the birth of the 'ideal formula' by which Gioberti expresses his own philosophical position called 'ontologism'. The formula consists of three statements connected to each other: (1) being is by necessity, (2) being

creates the existent and (3) the existent comes back to being. These statements identify two cycles. The first clarifies the relation between Creator and creatures; for existing-beings are literally so (as *ex-sistentes*) and their reason is outside themselves, in being as a creator. In the second cycle, man, as an image of being, finds his own creative dimension in thinking. The return to being takes the features of the human *methexis* of being's activity: it is man who makes himself the creator of history, institutions and culture, and it is man's creative progress which makes intelligible the sensible of nature by realizing its potential to be word.

The ideal formula formed a theoretical structure for a philosophical encyclopedia, whose first developments were the writings *Del bello (On Beauty)* (1841) and *Del buono (On Good)* (1843a), and to which are correlated *Primato morale e civile degli Italiani (Moral and Civil Primacy of Italians)* (1843b) and *Del rinnovamento civile in Italia (Of Italy's Civil Renewal)* (1851). These mark the birth of the Neo-guelphic project and its fall as a result of the First War of National Liberation (1848-9). In *Primato* Gioberti attributed to Italian national events a crucial role in the fate of Europe and at the same time wished to force the Catholic Church to become the guarantor of progress and of liberties. This explains his polemic against the Jesuits in *Il Gesuita moderno (The Modern Jesuit)* (1846-7). In *Del rinnovamento* a secular Papacy is no longer given a central role, and the defence of the internal *status quo* of states gives way to the acceptance of the ineluctable republican destiny of post-1848 Europe. Sovereignty is of the Idea, the choice between monarchy or republic only a matter of political opportunity: the Piedmontese way to Italian unity is supported, but only on this basis. What remains is the emphasis on the religious determination of progress: its catholicity.

Gioberti's legacy in Italian philosophy is linked to the participative moment of the second cycle of the ideal formula and especially to its version as it appears in the posthumously published *Protologia (First Science)* (1857). This work uses a great deal of Hegelian vocabulary and puts more stress on the dialectical determination of the formula. Man's thought in his return to being takes on the character of a con-creation, and Gioberti goes so far as to speak of a real 'palingenesis'. Some scholars see here a new phase in Gioberti's philosophy, where the opposition between transcendence and immanence loses its traditional sharpness. Supporting this interpretation, Bertrando Spaventa in his book *La filosofia di Gioberti (Gioberti's Philosophy)* (1863) recognized in Gioberti's work the inquiry of a principle of all things not in the absolute objectivity, material or ideal, but in the absolute mind. For him it represented a moment of the homecoming of Italian thought which had its origins inBruno andVico, and thence emigrated into German philosophy. In the late nineteenth century these aspects of Gioberti's work contributed to the theoretical roots of the actualism of Giovanni Gentile (§2).

See also: German idealism; Italy, philosophy in §2

MARIO PICCININI

List of works

Gioberti, V. (1838) *Teorica del sovrannaturale (Theory of the Supernatural)*, Padua: Cedam, 1972.(An early statement of Gioberti's philosophy.)

Gioberti, V. (1840) *Introduzione allo studio della filosofia (Introduction to the Study of Philosophy)*, Milan: Bocca, 2 vols, 1939-41.(The most organized exposition of Gioberti's philosophical system.)

Gioberti, V. (1841) *Del bello*, Milan: Bocca, 1939; trans. E. Thomas, *On Beauty*, London, Taylor, 1871.('Ideal formula' and aesthetics in Gioberti.)

Gioberti, V. (1841-4) *Degli errori filosofici di A. Rosmini (On the Philosophical Mistakes of Antonio Rosmini)*, Milan: Bocca, 1939.(Gioberti's critique of Rosmini's view of the Idea.)

Gioberti, V. (1843a) *Del buono (On Good)*, Milan: Bocca, 1940.(A reflection on ethics and politics strongly connected with the *Primato*.)

Gioberti, V. (1843b) *Primato morale e civile degli italiani (Moral and Civil Primacy of Italians)*, Milan: Bocca, 1939.(The most famous work by Gioberti and a major text of the Italian Risorgimento.)

Gioberti, V. (1847) *Il Gesuita Moderno (The Modern Jesuit)*, Milan: Bocca, 1940.(The polemic against Jesuits and their reading of the *Primato*.)

Gioberti, V. (1851) *Del Rinnovamento Civile d'Italia (Of Italy's Civil Renewal)*, Rome: ABETE, 1969. (Gioberti's view of the political condition of Europe and Italy after 1848.)

Gioberti, V. (1857) *Protologia (First Science)*, ed. G. Massari, Padua: Cedam, 1983-6.(A posthumous collection of fragments documenting the so-called second phase of Gioberti's thought.)Listed twentieth-century editions belong to the National Edition of *Opere edite e inedite Vincenzo Gioberti*, (originally edited by Enrico Castelli)

which began in 1939 and is still being printed by several publishers.

References and further reading

Gentile, G. (1898) *Rosmini e Gioberti. Saggio storico sulla filosofia italiana del Risorgimento (Rosmini and Gioberti. Historical Essay on Italian Philosophy of Risorgimento)*, Florence: Sansoni; 3rd edn, 1958.(The actualistic version of the idealistic interpretation of Gioberti.)

Derossi, G. (1970) *La teoria giobertiana del linguaggio come dono divino e il suo significato storico e speculativo (Gioberti's Theory of Language as a Divine Gift and Its Historical and Theoretical Meaning)*, Milan: Marzorati.(The linguistic side of Gioberti's philosophy.)

Padovani, U. (1927) *Gioberti e il cattolicesimo (Gioberti and Catholicism)*, Milan: Vita & Pensiero.(Represents the criticism of Gioberti by the Catholic orthodoxy.)

Portale, V. (1968) *Vincenzo Gioberti e l'ontologismo (Vincenzo Gioberti and Ontologism)*, Cosenza: Pellegrini.(A useful tool for a first approach.)

Spaventa, B. (1863) *La filosofia di Gioberti (Gioberti's Philosophy)*, Naples: Vitale. (The main text in the Italian idealistic reception of Gioberti.)

Stefanini, L. (1947) *Gioberti*, Milan: Bocca.(A wide study which stresses the elements of continuity in Gioberti's thought against the neo-idealistic interpretation.)

Glanvill, Joseph (1636-80)

Joseph Glanvill was an opponent of the scholastic philosophy which he had been taught in England, supporting instead the new learning associated with Francis Bacon and the Royal Society of which he became a Fellow in 1664. Although he called himself a sceptic and is often so classified, this may easily give a misleading picture of his philosophy. He was a sceptic in so far as he believed that human knowledge is very limited, opposing both the extravagant religious 'enthusiasm' for doctrines to the contrary which still retained adherents throughout his life, and the more general dogmatism which holds firmly to an opinion even though the evidence does not warrant it. Although he was not untouched by the revival of Platonic philosophies in Cambridge and was a great admirer of Henry More, in general he advocated an anti-dogmatic and generally empirical philosophy which in some ways anticipates the thought of Locke. In one area he might be accused of succumbing to the enthusiasm of his opponents, and that was in his espousal of a belief in witchcraft, though he claimed that in his explorations of the spirit world he was merely concerned to gather empirical evidence for a religious view. His analysis of causation has been held to anticipate that of Hume. Although it is true that he does have some of Hume's insights, the extent of that anticipation has sometimes been exaggerated.

1 The way to science

The tone of the English philosopher Glanvill's thought is almost always aimed at reconciling the new philosophy, associated with Francis Bacon, Descartes and the early members of the Royal Society, with traditional Christian doctrine as espoused by the Anglican Church. In doing so he laid great stress on the merits of the new philosophy and the limitations of human knowledge, itself substantially, but not wholly, the product of experience.

Glanvill's first and most important work, *The Vanity of Dogmatizing* (1661), was somewhat modified and improved as *Scepsis Scientifica* (1665). In them he sees 'confest ignorance as the way to science'. He argues that a proper recognition of our limited understanding of the world will provide an antidote to the dogmatic but unwarranted claims to knowledge all too often made, especially by the Aristotelians whom he regarded as the major opponents. Conceits and fancies too often stand in the way to true knowledge and we must recognize that, since the fall of man, ignorance has been the dominant human condition. The book develops examples of that ignorance: as the 'divine Plato' said, we have both mind and body. But how they are connected is a mystery. We can give no account of the nature of spirit. Even the 'ingenious Descartes' cannot help us. We are entirely dependent on our senses for our knowledge but, as we can see only the surface of things, our knowledge is scant and limited.

Despite these limitations, Glanvill regards Descartes' philosophy as providing the best alternative to Aristotelianism, although his understanding of Descartes' philosophy as providing merely hypothetical explanations of phenomena is undoubtedly one that Descartes would not have recognized. And although his empiricism was tempered with a commitment to some innate knowledge of the Platonic kind, he was not a rationalist in any serious sense. He follows Descartes in accepting a distinction between primary and secondary qualities and a generally mechanical philosophy.

Several chapters are a direct attack on Aristotelian philosophy, regarded as entirely devoid of merit. It is merely verbal, litigious, gives no account of the phenomena, makes no discoveries and is inconsistent with both divinity and itself.

2 Causation and contingency

Among the many other examples of our sources of error and ignorance given by Glanvill, the one that has drawn most attention to his writings over the years has been his account of causation, which in important ways anticipates the much more famous analysis of David Hume in the following century. The crucial section in Glanvill is very short:

> All knowledge of causes is *deductive*: for we know none by simple intuition; but through the mediation of their effects. So that we cannot conclude, anything to be the cause of another; but from its continual accompanying it: for the *causality* it self is *insensible*. But now to argue from a concomitancy to a causality, is not infallibly conclusive: yea in this way lies notorious delusion.

(Glanvill 1665: 142)

Glanvill, like Hume, recognizes the contingency of causal inference, its empirical basis in constant conjunction, and its fallibility. To claim knowledge of causes from our simple observations of constancies is merely a conceit, we never have more than conjectures.

In recognizing that we cannot observe the causal relationship but only the constant conjunction, Glanvill clearly went a long way towards Hume's position. Nor is this so surprising, given that both of them begin from the empiricist position that the source of knowledge is experience. And, whatever Hume's position might be, it is also clear that Glanvill does not deny the reality of causal relations in the world, only the certainty of our knowledge of them. His argument is entirely epistemic: we have no knowledge of necessary connection, but that does not entail that necessary connections do not exist. He does not explore the further question as to whether we have any reason at all to believe that there are any such connections. But this was in a sense already excluded by his realist commitment to the physical world, which remains an unexamined assumption in his philosophy. This emerges clearly in his account of our ignorance of the natural causes of phenomena in a later chapter. We cannot, he says, infer to the hidden causes of phenomena except by analogy, but this is notoriously unreliable because there is so often little resemblance between an effect and its cause. To claim that the causes must be as we conceive them is to set unwarranted bounds to omnipotency. We have to settle for the limited knowledge we have through our senses and refrain from the all-too-human tendency to make dogmatic claims which extend beyond our ability to know.

3 Glanvill and More

Glanvill can sound extremely modern and he does anticipate the later British empiricists, particularly Locke and Hume, in important ways. But we should also appreciate that he combined his empiricism with an early admiration for the philosophy of Henry More and Plato which fits less easily with that empirical tradition. In his early *Lux Orientalis* (1662) he followed More, Origen, Plato and the Cambridge Platonist George Rust in a commitment to pre-existence (see Cambridge Platonism). Yet, consistent with his position in *The Vanity of Dogmatizing*, he recognized that he had shown only that it is probably true and not that it must be.

Glanvill, once again following More, argues that, although spirit is primary, it is totally insensate without body and never acts 'but in some body or other', even though that body does not have to be the one we have now. He also sees the arguments for pre-existence as providing evidence for the independent existence of the soul and thus for the possibility of immortality. In subscribing to the Platonist theory he also commits himself to a theory of plastic nature which seems somewhat at odds with his cautious scepticism in other areas.

4 Witches and spirits as natural phenomena

Glanvill maintained a strong commitment to the existence of witches, which appeared in several of his publications. He claimed there was strong empirical evidence for their existence, and that they did exist itself argued for the spiritual world and our immortality. To deny their existence, as did Hobbes, was a short step away from atheism. Witchcraft and the world of spirits Glanvill regarded as further natural phenomena that deserved investigation entirely within the spirit of the Royal Society.

It is this last point that is perhaps the central feature of Glanvill's thought and which gives him a modest but real place in the history of philosophy. For in the early decades of the Royal Society he provided a philosophical defence of its methods and procedures against its assailants, in particular Henry Stubbe and Meric Casaubon. Central to much of Glanvill's argument was the belief that the study of nature in the way advocated by the Royal Society provided strong support for religion, rather than the opposite, particularly in exhibiting, as he expressed it in *Philosophia Pia* (1671: 17), 'the wonderful art, and contrivance that is in the contexture of the effects of nature'. In mounting this defence of a rational, cautious empiricism as a road to a deeper religious knowledge, Glanvill contributed substantially to the intellectual respectability of the new natural sciences without ever reaching a philosophical depth comparable to that achieved later in the century by John Locke.

See also: Aristotelianism in the 17th century

G.A.J. ROGERS

List of works

Glanvill, J. (1661) *The Vanity of Dogmatizing: or Confidence in Opinions. Manifested in a Discourse of the Shortness and Incertainty of our Knowledge, and its Causes, with some reflections on Peripateticism; and an Apology for Philosophy*, London.(With *Scepsis Scientifica*, Glanvill's most important work. It argues a cautious empiricist philosophy in keeping with the programme of the Royal Society.)

Glanvill, J. (1662) *Lux Orientalis, or An Enquiry into the Opinion of the Eastern Sages, Concerning the Præexistence of Souls. Being a key to unlock the Grand Mysteries of Providence, In relation to mans sin and misery*, London.(Argues for the Platonic doctrine of the plausibility of the pre-existence of souls. The argument is much indebted to Henry More.)

Glanvill, J. (1665) *Scepsis Scientifica: or Confest Ignorance, the way to Science; In an Essay of the Vanity of Dogmatizing, and Confident Opinion. With a Reply to the Learned Thomas Albius*, London.(Essentially a second edition of *The Vanity of Dogmatizing*.)

Glanvill, J. (1667) *Some Philosophical Considerations Touching the Being of Witches and Witchcraft*, London. (Argues that there is substantial empirical evidence for the existence of witches, which itself is evidence of a spiritual world.)

Glanvill, J. (1668a) *A Blow at Modern Sadducism in some Philosophical Considerations about Witchcraft*, London.(Further evidence for witches and therefore for spirits.)

Glanvill, J. (1668b) *Plus Ultra: or The Progress and Advancement of Knowledge since the Days of Aristotle*, London.(A championing of modern learning and a defence of the Royal Society against those who asserted the superiority of the ancients.)

Glanvill, J. (1671) *Philosophia Pia; or, A discourse of the Religious Temper, and Tendencies of the Experimental Philosophy, which is profest by the Royal Society*, London.(A further defence of the Royal Society and its goals, especially with regard to the religious implications of its method and findings which Glanvill claims strongly supported the religious position.)

Glanvill, J. (1676) *Essays on Several Important Subjects in Philosophy and Religion*, London.(Essays defending positions he had already advocated on reason, spirits, religion, scepticism and knowledge.)

Glanvill, J. (1681) *Sadducismus Triuphatus: or, Full and Plain Evidence concerning Witches and Apparitions. In two Parts. the First treating of their Possibility, the Second of their Real Existence*, London.(Yet more argument and evidence for the existence of witches.)

References and further reading

Cope, J.I. (1956) *Joseph Glanvill. Anglican Apologist*, St Louis, MO: Washington University Studies.(A scholarly and comprehensive intellectual biography.)

Leeuwan, H.G. van (1970) *The Problem of Certainty in English Thought 1630-1690*, The Hague: Martin Nijhoff. (Pages 71-89 are useful on Glanvill, and the remainder of the book provides the context of Glanvill's thought.)

Popkin, R.H. (1953) 'Joseph Glanvill: A Precursor of David Hume', *Journal of the History of Ideas* 14: 293-303; repr. in R.H. Popkin *The High Road to Pyrrhonism*, Indianapolis, IN: Hackett Publishing Company, 1993. (Argues that Glanvill anticipated Hume on the fallacy of induction.)

Gnosticism

Gnosticism comprises a loosely associated group of teachers, teachings and sects which professed to offer 'gnosis', saving knowledge or enlightenment, conveyed in various myths which sought to explain the origin of the world and of the human soul and the destiny of the latter. Everything originated from a transcendent spiritual power; but corruption set in and inferior powers emerged, resulting in the creation of the material world in which the human spirit is now imprisoned. Salvation is sought by cultivating the inner life while neglecting the body and social duties unconnected with the cult. The Gnostic movement emerged in the first and second centuries AD *and was seen as a rival to orthodox Christianity, though in fact some Gnostic sects were more closely linked with Judaism or with Iranian religion. By the fourth century its influence was waning, but it persisted with sporadic revivals into the Middle Ages.*

1 Basic doctrines

Gnosticism can best be understood in terms of family likeness. One can identify characteristic features, most of which are found in most Gnostic sects; but the attempts often made to define Gnosticism in terms of universally present common features can only approximate to the truth.

Characteristic tenets include:

1 A radical dualism, contrasting a transcendent realm of pure spirit with the world of gross matter. The human makeup likewise presents a sharp contrast of spirit and sensuality, with a corresponding distinction between the 'elect' or spiritual people and the rest of society, though some systems introduce an intermediate grade.
2 A creator presented as imperfect or evil, though commonly identified with the God of Judaism, and sharply contrasted with the supreme divinity, who is his ultimate source. His existence is explained by various myths depicting events prior to the creation and claiming to show how evil dispositions arose by accumulated lapses among the heavenly powers.
3 The human spirit originated in the higher realm, but is now imprisoned in the form of a soul within the material body. Many Gnostic sects taught that the same spirit can live many lives. But it is often seen as predestined to salvation or the reverse.
4 The Gnostics' aim was to liberate their spirits from all attachment to material things, and thereby return with the elect minority to ultimate happiness. Most Gnostic sects therefore adopted a puritan ethic, though some held that all physical actions are contemptible and approved licentious conduct as a sign of liberation.

2 Definitions, origins and dating

'Gnosticism' is a term coined by modern scholars. Ancient writers allude to 'gnosis', that is, knowledge, especially spiritual knowledge or enlightenment. St Paul speaks disparagingly of Christians who laid claim to it (1 Corinthians 8: 1), but it was nevertheless commended by Clement of Alexandria and others. Clement also uses 'Gnostic' to mean a devout and instructed Christian. The Gnostic sects themselves had no common self-designation parallel to 'Jew' or 'Christian', and were commonly named after their founders. Irenaeus (*c*.130-*c*.200), however, implies that the term was appropriated by several sects, and its use was soon extended to include all similar schools.

There has been much debate about the origins of Gnosticism, whether Greek, Jewish or Iranian. It now appears that the movement was too diversified for any single-source theory to be acceptable, and many forms of it clearly presuppose an amalgamation of different cultures. Its ablest exponents, including Valentinus and Marcion (*fl.* *c*.140-60), inherit the traditions of Hellenistic Judaism, incorporating Christian elements; the rest are unlikely to interest students of philosophy.

The problem of dating has been complicated by ill-defined terminology. It has been claimed that Gnosticism originated in Iran before the Christian movement emerged. But it now appears that the emergence of *systematic* Gnostic teaching is roughly contemporaneous with a parallel Christian development, though many of the ideas found in Gnosticism were current earlier. Scholars, especially in Germany, now tend to reserve the term 'Gnosticism' for the elaborate systems described by Irenaeus around 180 AD, for example, using 'Gnosis' as an

inclusive term for its constituent ideas.

Mandaeism, a small Gnostic sect unnoticed by Christian writers, has attracted some attention from scholars, as the sect still survives and has preserved sacred writings of great antiquity. Its claim to derive from John the Baptist is probably unfounded.

3 Sources

For many centuries Gnosticism was known only through the writings of its Christian opponents, notably Irenaeus, Tertullian and Clement, who did however embody quotations from the works they criticized. Some later Gnostic texts of dubious value emerged in the eighteenth century, supplemented by the important Berlin Codex 8502 (discovered 1901, fully edited 1955). But the situation was transformed by the discovery of forty-four books in codex form at Nag-Hammadi in Upper Egypt in 1944 (though once again publication was delayed). Most are Coptic translations of Greek originals, some of which probably date from the first century AD. Many of them introduce biblical characters, though strongly influenced by Gnostic assumptions. Three may be mentioned in particular: the Apocryphon of John, which abounds in fanciful mythology, but was apparently authoritative and survives in several copies; the Gospel of Thomas, a collection of sayings ascribed to Jesus, isolated from their settings and accompanying actions, but sometimes presenting variant forms of canonical Gospel texts; and the so-called Gospel of Truth (*Evangelium veritatis*), the one item in the collection which could without absurdity be annexed to the Christian scriptures; it has no marked heretical features and offers an original meditation on the passion of Christ.

The list of sources should be extended by a brief note on Manicheism, a Gnostic sect founded by Mani (Manes, Manichaeus) around 216-76 AD in Iran, and influential especially in the fourth century, when it briefly captured St Augustine (see Manicheism). Earlier patristic and Muslim sources can now be compared with Manichean documents found at Turfan from 1898 and in Egypt from 1930 onwards, including a biography of the founder.

4 Philosophical content and value

Early Christian writers, especially Hippolytus (*c*.170-*c*.236), argued that the Gnostics were influenced by Greek philosophy. In most cases this is unlikely, and where such influence existed it has been overlaid by mythology. Some philosophical schools, for example, Sceptics and Epicureans, can be discounted; Stoic influence was slight and indirect. One must also exclude the dominant Platonism, which taught the eternity of the world, taking the *Timaeus* to symbolize its eternal dependence on creative goodness. But some Gnostic sects echo the Jewish and Christian Platonism that assimilated the early chapters of Genesis with the *Timaeus* interpreted historically, as it was by Plutarch and Atticus. The Nag-Hammadi texts include an extract from Plato’s *Republic* (588a-589b) inaccurately reproduced in Coptic. Pythagorean influence appears in the significance assigned to numbers, already present in the Jewish practice of Gematria, where numbers, commonly expressed by letters, are regrouped to yield significant words; thus 666, ‘the number of the beast’ (Revelation 13: 18), can be split up, it is alleged, to yield the letters NERO CAESAR.

Platonic traits appear clearly in the Valentinian system, which was widely influential and is commonly taken as typical. Valentinus was a gifted man who hoped to be made Bishop of Rome, and so presumably restrained his speculative powers, as his scanty surviving fragments suggest. Yet the Valentinian system, as known only one generation later, presents a bewilderingly complex mythology, clearly unacceptable to mainstream Christians.

Note however the following features:

1 Dualism is modified to include an intermediate grade: three levels of being - spirit, soul and matter; and three classes of people - the Gnostic elect, the conventional church member, the unregenerate outsider - which recalls the three classes of citizens in Plato’s *Republic*.
2 Evil is traced to defective cognition.
3 The ultimate divinity expands to form a series of powers or ‘Aeons’. The first derivative is God’s self-knowledge (his ‘Ennoia’). But the process goes wrong: in the developed Valentinian myth, the primal fault is ascribed to the last in a series of thirty Aeons, who nevertheless bears the prestigious name of Sophia. This may point to an earlier conception in which it is the *first* derivative, Ennoia-Sophia, who fails. Conversely, Irenaeus describes a further development in which the erring Sophia herself is duplicated.

The myth serves to express a fundamental problem of theology. The ancients commonly thought of knowledge as a process of copying; thus we may be said to know someone when we can recall that person's features. Thus any knowledge of God must be a kind of replica. But it cannot be perfect, or it would amount to a second divinity. So any attempt to elucidate God must be presumptuous. The most thoughtful treatment of the problem appears in the Tripartite Tractate from Nag-Hammadi. Here God's nature is expressed in a series of powers which at first appear as impersonal attributes; but to replicate God's being each one must become a sovereign will; they thus incur a common failure, as each one fails to consider its own incompleteness and its need of the others. This account of the primary fault is clearly more persuasive than the official Valentinian theory, which fixes the blame exclusively on Sophia. But the majority of Gnostic teachers were catering for untrained minds, and sought to impress them with increasingly complex and pretentious mythology, a feature which today baffles and repels many philosophers.

CHRISTOPHER STEAD

References and further reading

Foerster, W. (ed. and trans.) (1972, 1974) *Gnosis*, Oxford: Clarendon Press.(An English translation of Gnostic texts in two volumes: 1, Patristic Evidence: 2, Coptic and Mandaean Sources.)

Grant, R.M. (ed. and trans.) (1961) *Gnosticism, An Anthology*, London: Collins. (Good introductory collection.)

Jonas, H. (1958) *The Gnostic Religion*, Boston, MA: Beacon Press.(Authoritative brief survey.)

Koester, H. (1990) *Ancient Christian Gospels, their History and Development*, London: SCM Press.(Compares the canonical Gospels with Gnostic parallels.)

Layton, B. (ed. and trans.) (1987) *The Gnostic Scriptures: A New Translation*, London: SCM.(Complete English version of the Nag-Hammadi texts.)

Pagels, E.H. (1975) *The Gnostic Paul*, Philadelphia, PA: Fortress Press.(Introduces Gnostic biblical scholarship.)

Pagels, E.H. (1979) *The Gnostic Gospels*, New York: Random House; repr. Harmondsworth: Penguin, 1982. (Introduces Gnostic biblical scholarship.)

Robinson, J.M. (ed. and trans.) (1988) *The Nag-Hammadi Library*, Leiden: Brill. (Complete translation of the collection.)

Rudolph, K. (1983) *Gnosis*, Edinburgh: T. & T. Clark.(Authoritative, comprehensive study.)

Stead, G.C. (1969) 'The Valentinian Myth of Sophia', *Journal of Theological Studies* 20: 75-104.(Examines the growth of a Gnostic myth).

Wilson, R.M. (1968) *Gnosis and the New Testament*, Oxford: Blackwell.(Fairly simple in treatment, but totally reliable.)

Yamauchi, E. (1973) *Pre-Christian Gnosticism*, London: Tyndale Press.(Discusses the question whether Gnosticism antedates Christianity.)

God, arguments for the existence of

Arguments for the existence of God go back at least to Aristotle, who argued that there must be a first mover, itself unmoved. All the great medieval philosophers (Arabic and Jewish as well as Christian) proposed and developed theistic arguments - for example, Augustine, al-Ghazali, Anselm, Moses Maimonides, Thomas Aquinas and Duns Scotus. Most of the great modern philosophers - in particular René Descartes, Gottfried Leibniz and Immanuel Kant - have also offered theistic arguments. They remain a subject of considerable contemporary concern; the twentieth century has seen important work on all the main varieties of these arguments.

These arguments come in several varieties. Since Kant, the traditional Big Three have been the cosmological, ontological and teleological arguments. The cosmological argument goes back to Aristotle, but gets its classic statement (at least for European philosophy) in the famous 'five ways' of Aquinas, in particular his arguments for a first uncaused cause, a first unmoved mover, and a necessary being. According to the first-mover argument (which is a special case of the first-cause argument), whatever is moved (that is, caused to move) is moved by something else. It is impossible, however, that there should be an infinite series of moved and moving beings; hence there must be a first unmoved mover. Aquinas goes on to argue that a first mover would have to be both a first cause and a necessary being; he then goes on in the next parts (Ia, qq.3-11) of the Summa theologiae to argue that such a being must have the attributes of God.

The perennially fascinating ontological argument, in Anselm's version, goes as follows: God is by definition the being than which none greater can be conceived. Now suppose God did not exist. It is greater to exist than not to exist; so if God did not exist, a being greater than God could be conceived. Since God is by definition the being than which none greater can be conceived, that is absurd. Therefore the supposition that God does not exist implies an absurdity and must be false. This argument has had many illustrious defenders and equally illustrious attackers from Anselm's time to ours; the twentieth century has seen the development of a new (modal) version of the argument.

Aquinas' fifth way is a version of the third kind of theistic argument, the teleological argument; but it was left to modern and contemporary philosophy to propose fuller and better-developed versions of it. Its basic idea is simple: the universe and many of its parts look as if they have been designed, and the only real candidate for the post of designer of the universe is God. Many take evolutionary theory to undercut this sort of argument by showing how all of this apparent design could have been the result of blind, mechanical forces. Supporters of the argument dispute this claim and retort that the enormously delicate 'fine tuning' of the cosmological constants required for the existence of life strongly suggests design.

In addition to the traditional Big Three, there are in fact many more theistic arguments. There are arguments from the nature of morality, from the nature of propositions, numbers and sets, from intentionality, from reference, simplicity, intuition and love, from colours and flavours, miracles, play and enjoyment, from beauty, and from the meaning of life; and there is even an argument from the existence of evil.

1 Cosmological arguments

Cosmological arguments start from some obvious and general but a posteriori fact about the universe: that there are contingent beings, for example, or that things move or change. We find first steps towards such an argument in Plato (*Laws* 10); Aristotle (§16) (*Metaphysics* 12; *Physics* 7, 8) gives it a fuller statement; the medieval Arabic (especially al-Ghazali) and Jewish philosophers (especially Maimonides) gave elaborate statements of the argument; but its *locus classicus* (for Westerners, anyway) is the first three of the famous 'five ways' of Aquinas' *Summa theologiae*. Following Aquinas, Duns Scotus (§§7-11) presented a subtle and powerful version of the argument, and in modern times the most influential versions of the argument are to be found in the works of Leibniz (§3) and Samuel Clarke (§1). (The most influential criticisms of the argument are given by Hume (§6) and Kant (§8).)

Following William Craig (1980), we may distinguish substantially three versions of the cosmological argument. First, the so-called *kalam* (Arabic, 'speculative theology') argument, developed by Arabic thinkers (for example, al-Kindi (§2) and al-Ghazali). Put most schematically, this argument goes as follows:

(1) Whatever begins to exist is caused to exist by something else.
(2) The universe began to exist.
(3) Therefore the universe was caused to exist; and the cause of its existence is God.

The second premise was supported by arguments for the conclusion that an 'actual infinite' is not possible: it is not possible, for example, that there have been infinitely many temporally non-overlapping beings each existing for at least a second; alternatively, it is not possible that an infinite number of seconds have elapsed. These arguments proceed by pointing out some of the paradoxes or peculiarities that an actual infinite involves (see Infinity). For example, suppose there were a hotel with infinitely many rooms ('Hilbert's Hotel'). The hotel is full; a new guest arrives; despite the fact that each room is already occupied, the proprietor accommodates the guest by putting them in room 1, moving the occupant of room 1 to room 2, of room 2 to room 3, and in general the occupant of *n* to $n + 1$. No problem! Indeed, when a large bus containing infinitely many new guests pulls up, they too can all be accommodated: for any odd-numbered room *n*, move its occupant into room 2*n* (moving the occupant of *that* room n^* into $2n^*$, and so on), thus freeing up the infinitely many odd-numbered rooms. In fact, if it is a busy weekend and an infinite fleet of buses pulls up, each with infinitely many new guests, they too can all be easily accommodated. And the question is: is it really possible, in the broadly logical sense, that such a hotel could actually exist? The friend of the *kalam* argument thinks not, and adds that no other actual infinite is possible either. If so, then the universe has not existed for an infinite stretch of time, but had a beginning. Contemporary cosmological theory in physics has seemed to some to provide scientific, empirical support for the claim that the universe had a beginning; according to 'Big Bang' cosmology, the universe came into being something like 15 billion years ago, give or take a few billion (see Cosmology §3).

Given that the universe has a beginning, the next step is to argue (by way of the first premise) that it must therefore have had a cause; it could not have popped into existence uncaused. And the final step is to argue that the cause of the universe would have to have certain important properties - properties of God.

The second kind of cosmological argument is the kind to be found in the first three of Aquinas' five ways. His second way, for example, goes like this:

(1) Many things in nature are caused.
(2) Nothing is a cause of itself.
(3) An infinite regress of essentially ordered efficient causes is impossible.
(4) Therefore, there is a first uncaused cause - 'to which', says Aquinas, 'everyone gives the name of God.'

There are two points of particular interest about this argument. First, Aquinas disagrees with a premise of the *kalam* argument, according to which it is impossible that there be an actual infinite. He argues that it cannot be proved that the universe had a beginning; he thinks it possible (though false) that the universe has existed for an infinite stretch of time. How then are we to understand premise (3)? Aquinas is here speaking of a certain kind of series, an 'essentially ordered' series, a series of causes in which any cause of an effect must be operating throughout the whole duration of the effect's operation. It is only such series, he says, that cannot proceed to infinity. (Aquinas gives the example of a stick moving a stone, a hand moving the stick, and so on.) So the upshot of the argument, if it is successful, is that there exists at least one thing which causes other things to exist, but is not itself caused to exist by anything else.

But could there not be many such things? And would each of them be God? This brings us to the second point of interest. Aquinas argues that there must be a first unmoved mover, a first uncaused cause, a necessary being, and the like; but his theistic argument is not finished there. In the next eight questions he argues that anything that was a first efficient cause would have to be immaterial, unchanging, eternal, simple and the possessor of all the perfections to be found in those things dependent upon it - in a word, God. It is therefore incorrect to follow the usual custom of criticizing Aquinas for hastily concluding that a first cause or unmoved mover or necessary being would have to be God.

The third sort of cosmological argument is associated especially with Leibniz and Samuel Clarke; according to this version of the argument, there must be a sufficient reason for the actuality of any contingent state of affairs. Therefore there must be a sufficient reason for the existence of any contingent being - but also, says Leibniz, for the whole series of contingent beings. This sufficient reason must be the activity of God.

2 Ontological arguments

Anselm’s ontological argument has excited enormous controversy (see Anselm of Canterbury §§3-4). Aquinas rejected it, Duns Scotus ‘coloured’ it a bit and then accepted it; Descartes (§6) and Malebranche endorsed it; like Duns Scotus, Leibniz thought it needed just a bit of work to be successful; Kant rejected it and delivered what many thought to be the final quietus (though others have found Kant’s criticisms both intrinsically obscure and of doubtful relevance to the argument); and Schopenhauer thought it a ‘charming joke’. Although in the twentieth century it was defended by (among others) Charles Hartshorne, Norman Malcolm and Alvin Plantinga, probably most contemporary philosophers reject the argument, thinking it a joke, but not particularly charming.

Anselm’s version goes as follows:

> Hence, even the fool is convinced that something exists in the understanding, at least, than which nothing greater can be conceived…. And assuredly that, than which nothing greater can be conceived, cannot exist in the understanding alone; for suppose it exists in the understanding alone; then it can be conceived to exist in reality; which is greater.
>
> Therefore, if that, than which nothing greater can be conceived, exists in the understanding alone, the very being, than which nothing greater can be conceived is one, than which a greater can be conceived. But obviously this is impossible. Hence, there is no doubt that there exists a being, than which nothing greater can be conceived, and it exists both in the understanding and in reality.
>
> (*Proslogion*, ch. 2)

This argument is a *reductio ad absurdum*: postulate the nonexistence of God, and show that this leads to an absurdity. Perhaps we can outline the argument as follows:

(1) A maximally great being (one than which nothing greater can be conceived) exists in the understanding (that is, is such that we can conceive of it).
(2) It is greater to exist in reality than to exist merely in the understanding.
(3) Therefore, if the maximally great being existed *only* in the understanding, it would be less than maximally great.

But it is impossible that the maximally great being be less than maximally great; hence this being exists in reality as well as in the understanding - that is, it exists. And clearly this maximally great being is God.

The earliest objection to this argument was proposed by Anselm’s contemporary and fellow monk Gaunilo in his *On Behalf of the Fool* (Psalm 14: ‘The fool has said in his heart "There is no God"’). According to Gaunilo, the argument must be defective, because we can use an argument of the very same form to demonstrate the existence of such absurdities as an island (or chocolate sundae, or hamster, for that matter) than which none greater can be conceived. (Says Gaunilo: ‘I know not which I ought to regard as the greater fool: myself, supposing that I should allow this proof; or him, if he should suppose that he had established with any certainty the existence of this island.’) But Anselm has a reply: the notion of a maximally great island, like that of a largest integer, does not make sense, cannot be exemplified. The reason is that the properties that make for greatness in an island - size, number of palm trees, quality of coconuts - do not have intrinsic maxima; for any island, no matter how large and no matter how many palm trees, it is possible that there be one even larger and with more palm trees. But the properties that make for greatness in a being - knowledge, power and goodness, for example - do have intrinsic maxima: omniscience, omnipotence and being perfectly good.

3 Ontological arguments (cont.)

The most celebrated criticism of the ontological argument comes from Immanuel Kant, who apparently argues in his *Critique of Pure Reason* (Transcendental Dialectic, bk II, ch. III, section 4) that if this argument were sound, the proposition ‘there is a being than which none greater can be conceived’ would have to be logically necessary; but there cannot be an existential proposition that is logically necessary. Sadly, his reason for making this declaration is itself maximally obscure. He adds that ‘existence is not a real predicate’, which is widely quoted as the principal objection to the argument. Unfortunately this dictum is of dubious relevance to Anselm’s argument and a dark saying in its own right. What might it mean to say that existence is not a real property or predicate?

And if it is not, how is that relevant to the argument? Why should Anselm care whether it is or not?

Perhaps we can understand Kant as follows. The argument as stated begins with the assertion that a maximally great being exists in the understanding; the idea is that this much is obvious, whether or not this being also exists in reality (that is, actually exists). Anselm then goes on to reason about this being, arguing that a being with the properties this one has - of being maximally great - cannot exist only in the understanding, but must exist in reality as well. So the argument depends upon the assumption that there *is* a maximally great being, and now the question is: does this being actually exist? Use the term ‘actualism’ for the view that there are not (and could not be) things that do not exist; the things that exist are all the things there are. Note that if this is true, then existence is a very special property: it is redundant, in that it is implied by every other property; anything that has *any* property (including the property of being maximally great) also has existence. But if actualism is true, the ontological argument as formulated above cannot work. For if it is not possible that there be things that do not exist, then in saying initially that there is a maximally great being, one that at any rate exists in the understanding, we are already saying that there exists a maximally great being, thus begging the question. If no maximally great being exists, then there simply is no such thing as a maximally great being, in which case we cannot (following Anselm) suppose initially that the maximally great being does not exist in reality and then argue that this being would be greater if it did exist in reality. If actualism is true, existence is a redundant property; but then to say that there is a maximally great being that exists in the understanding is already to say that there really exists a maximally great being. So perhaps Kant’s puzzling dictum should be seen as an early endorsement of actualism.

Of course Anselm might reply that the fault lies not with his argument, but with actualism; in any event, there are other versions that do not conflict with actualism. Charles Hartshorne (1941) claimed to detect two quite different versions of the argument in Anselm’s work; the second version is consistent with actualism and thus sidesteps Kant’s criticism. This version proceeds from the thought that a really great being would be one that would have been great even if things had been different; its greatness is stable across possible worlds, to put it in a misleading if picturesque way. So say that a being has *maximal excellence* in a given possible world *W* if and only if it is omnipotent, omniscient and wholly good in *W*; and say that a being has *maximal greatness* if it has maximal excellence in every possible world. Then the premise of the argument (thus restated) is simply:

Maximal greatness is possibly exemplified.

That is, it is possible that there be a being that has maximal greatness. But (given the widely accepted view that if a proposition is possibly necessary in the broadly logical sense, then it is necessary), it follows by ordinary modal logic that maximal greatness is not just possibly exemplified, but exemplified in fact. For maximal greatness is exemplified if and only if there is a being *B* such that the proposition

B is omnipotent and omniscient and wholly good (has maximal excellence)

is necessary. If maximal greatness is possibly exemplified, therefore, then some proposition of that sort is possibly necessary. By the above principle, whatever is possibly necessary is necessary; accordingly, that proposition is necessarily true and hence true.

So stated, the ontological argument breaches no laws of logic, commits no confusions and is entirely immune to Kant’s criticism. The only remaining question of interest is whether its premise, that maximal greatness is possibly exemplified, is indeed true. That certainly seems to be a rational claim; but it is not one that cannot rationally be denied. A remaining problem with the argument, perhaps, is that it might be thought that the epistemic distance between premise and conclusion is insufficiently great. Once you see how the argument works, you may think that asserting or believing the premise is tantamount to asserting or believing the conclusion; the canny atheist will say that he does not believe it is possible that there be a maximally great being. But would not a similar criticism hold of any valid argument? Take any valid argument: once you see how it works, you may think that asserting or believing the premise is tantamount to asserting or believing the conclusion. The ontological argument remains as intriguing as ever.

4 Teleological arguments

Teleological arguments start from contingent premises that involve more specific features of the universe, features which in one way or another suggest that the universe has been designed by a conscious and intelligent being.

These arguments have often been developed in close connection with modern science; they have been endorsed by many of the giants of modern science, including Isaac Newton. Here is a classic statement of the argument by William Paley (§2):

> In crossing a heath, suppose I pitched my foot against a *stone*, and were asked how the stone came to be there, I might possibly answer that for any thing I know to the contrary, it had lain there for ever: nor would it perhaps be very easy to show the absurdity of this answer. But suppose I had found a *watch* upon the ground and it should be inquired how the watch happened to be in that place, I should hardly think of the answer which I had before given, that, for any thing I know the watch might have always been there. Yet why should not this answer serve for the watch, as well as for the stone? For this reason and for no other: viz., that, when we come to inspect the watch, we perceive (what we could not discover in the stone) that its several parts are framed and put together for a purpose.
>
> ([1802] 1804: 1-2)

Paley then points out that the universe and some of its parts - for example, living things and their organs - resemble a watch, in that they give the appearance of having been designed in order to accomplish certain purposes. An eye, for example, looks like an extremely subtle and sophisticated mechanism designed to enable its owner to see. But the only serious candidate for the post of designer of the universe is God.

Kant, who had little but contempt for the cosmological and ontological arguments, was much less dismissive of this one. He still rejected the argument, however, pointing out that at most it shows that it is likely that there is a designer or architect of the universe; and it is a long way from a designer to the God of the theistic religions, an almighty, omniscient, wholly good creator of the world, by whose power the universe sprang into being. Of course, a cosmic architect - a being who has designed our entire universe, with its elements ranging across many orders of magnitude from gigantic galaxies to the minutest things we know - is no mean conclusion, and it seems churlish to dismiss it with an airy wave in order to point out that there is something even stronger that the teleological argument does not show.

5 Teleological arguments (cont.)

Many people, however, have rejected the teleological argument even taken as an argument for a designer. The eighteenth-century proponents of the argument invariably mentioned the apparent teleology in the biological world; but (so say the critics) Darwin changed all that. We now know that the apparent design in the world of living things is *merely* apparent. The enormous variety of flora and fauna, those enormously elaborate and articulate mechanisms, and finely detailed systems and organs such as the mammalian eye and the human brain give a powerful impression of design; but in fact they are the product of such blind mechanisms as random genetic mutation and natural selection. The idea is that there is a source of genetic variation which produces mutation in the structure and function of existing organisms. Most of these mutations are deleterious; a few are adaptive and their lucky owners will have an adaptive edge, eventually coming to predominate in a population. Given enough time, so the story goes, this process can produce all the splendid complexity and detail that characterize the contemporary living world (see Evolution, theory of).

Of course there is little real evidence that these processes can in fact achieve this much: naturally enough, we have not been able to follow their operation in such a way as to observe them produce, say, birds or mammals from reptiles, or even human beings from simian precursors. And even if we did observe the course of animate history (even if we had a detailed record on film), this would by no means show that blind mechanisms are in fact sufficient for this effect; for of course there would be nothing in the film record to show that those random genetic mutations were not in fact guided and orchestrated by God.

Still, the critic of the teleological argument claims not that in fact evolution has been accomplished just by these blind mechanisms, but that it could have been; if so, there is a real alternative to design. That these mechanisms really could have produced effects of this magnitude is far from clear; we have little real reason to suppose that there is a path through the space of possible animal design plans, a path leading from bacteria to human beings, and such that each new step is both adaptive and reachable from the previous step by mechanisms we understand. Still, the suggestion does perhaps damage the teleological argument by suggesting a naturalistic candidate for the post of producer of apparent design.

But organic evolution addresses only one of the areas of apparent design. There is also the origin of life; even the simplest unicellular creatures (prokaryotes such as bacteria and certain algae, for example) are enormously complex and upon close inspection look for all the world as if they have been designed; it is fair to say that no one, so far, has a decent idea as to how these creatures might have come into being just by way of the operation of the regularities of physics and chemistry. There are also the various considerations connected with the so-called 'fine tuning' of the universe. First, there is the 'flatness' problem. The mass density of the universe is at present very close to the density corresponding to the borderline between an open universe (one that goes on expanding for ever) and a closed universe (one that expands to a certain size and then collapses). The ratio between the forces making for expansion and those making for contraction is close to one. But then shortly after the Big Bang this value would have to have been inside a very narrow band indeed. Thus Stephen Hawking (1974): 'reduction of the rate of expansion by one part in 10^{12} at the time when the temperature of the Universe was 10^{10} K would have resulted in the Universe's starting to recollapse when its radius was only 1/3000 of the present value and the temperature was still 10,000 K' - much too warm for the development of life. On the other hand, if the rate of expansion had been even minutely greater, the universe would have expanded much too fast for the formation of stars and galaxies, required for the formation of the heavy elements necessary for the development of life.

Another kind of fine tuning was also necessary: of the fundamental physical constants. If any of the four fundamental forces (weak and strong nuclear forces, electromagnetic force, electron charge) had been even minutely different, the universe would not have supported life; they too must have been fine-tuned to an almost unbelievable accuracy. And the suggestion, again, is that given the infinite range of possible values for the fundamental constants, design is suggested by the fact that the actual values fall in that extremely narrow range of values that permits the development of intelligent life.

But there is a naturalistic riposte. Since the 1970s, several different sorts of 'inflationary' scenario have shown up. These postulate the formation (at very early times) of many different universes or subuniverses, with different rates of expansion, and different values for the fundamental constants. These inflationary models are motivated, in part, by a desire to avoid singularities and the accompanying appearance of design. If all possible values for the fundamental constants and the rate of expansion are actually exemplified in different subuniverses, then the fact that there is a subuniverse with the values ours displays no longer requires explanation or suggests a Designer. Many of these scenarios are wildly speculative and unencumbered by empirical evidence, but (if physically acceptable) they do tend to blunt the force of a design argument from fine tuning. (Of course, someone who already believed in God and saw no need to eliminate suggestions of design might be inclined to reject these suggestions as metaphysically extravagant.) But there are also counterarguments here; the discussion goes on. It is hard to see a verdict, at present, on the prospects of this form of the argument. The teleological argument seems to have enormous vitality; its epitaph is often read, but the argument regularly reappears in new forms. As for a final evaluation, the best perhaps comes from Kant, who said that this argument 'always deserves to be mentioned with respect. It is the oldest, the clearest and the most accordant with the common reason of mankind' (1787: 520).

6 Other theistic arguments

We have examined the Big Three among theistic arguments, but there are many more. First, there are moral arguments of at least two sorts. These are arguments that the very nature of morality - the unconditioned character of the moral law - requires a divine lawgiver. You might find yourself utterly convinced that:

Morality is objective, not dependent upon what human beings know or think or do.

You may also be convinced that:

The objective character of morality cannot be explained in terms of any 'natural' facts about human beings (or other things), so there could not be such a thing as objective moral law unless there were a being like God who legislates it.

Then you will have a theistic argument from the nature of morality. This argument can go in either of two directions: some people think we can simply *see* that moral obligation is impossible apart from a divine will and lawgiver, while others think that the dependence of moral obligation upon the will of God is the best explanation for its objectivity and special deontological force.

A second main type of moral argument is due to Kant (1788), who argues first that virtue deserves to be proportionally rewarded with happiness: the more virtuous you are, the more happiness you deserve. But nature by itself does not seem able to guarantee anything like this sort of coincidence. If morality is to make sense, however, it must be supposed that there is such a coincidence; practical reason, therefore, is entitled to postulate a supernatural being with enough knowledge, power and goodness to ensure that we receive the happiness we deserve as a reward for our virtue. So taken, the argument is for the rationality of making the assumption that there is a being of this sort; it is not really an argument for the actual existence of such a being. This argument receives criticism from several sides: some hold that we do not have to assume that there is proportionality between virtue and happiness in order to carry out the moral life; others (for example, many Christians) argue that both happiness and the ability to live a moral life are gifts of grace and that if we really got what we deserve, we should all be thoroughly miserable.

There are many other theistic arguments - arguments from the nature of proper function, from the nature of propositions, numbers and sets, from intentionality, from counterfactuals, from the confluence of epistemic reliability with epistemic justification, from reference, simplicity, intuition, love, colours and flavours, miracles, play and enjoyment, morality, beauty, the meaning of life, and even from the existence of evil. There is no space even to outline all these arguments, so we will look at just three.

First, the argument from intentionality (or aboutness). Consider propositions - the things that are true or false, that are capable of being believed, and that stand in logical relations to one another. Propositions have another property: *aboutness* or *intentionality*. They represent reality or some part of it as being thus and so, and it is by virtue of this property that propositions (as opposed, for example, to sets) are true or false. Most who have thought about the matter have found it incredible that propositions should exist apart from the activity of minds. How could they just *be* there, if never thought of? Further, representing things as being thus and so - being about something or other - seems to be a property or activity of minds or perhaps thoughts. It is therefore extremely plausible to think of propositions as ontologically dependent upon mental or intellectual activity in such a way that either they just *are* thoughts, or else at any rate could not exist if not thought of. But propositions cannot be human thoughts; there are far too many of them for that. (For each real number *r*, for example, there is the proposition that *r* is distinct from the Taj Mahal.) Hence the only viable possibility is that they are divine thoughts, God's thoughts (so that when we think, we literally think God's thoughts after him).

Second, there is the argument from sets or collections. Many think of sets as displaying the following characteristics: (1) no set is a member of itself; (2) sets (unlike properties) have their extensions essentially - hence many sets are contingent beings and no set could have existed if one of its members had not; (3) sets form an iterated structure - at the first level, there are sets whose members are nonsets, at the second, sets whose members are nonsets or first level sets, …, at the *n*th level, sets whose members are nonsets or sets of index less than *n*, …, and so on.

It is also natural to follow Georg Cantor, the father of modern set theory, in thinking of sets as collections - that is, as things whose existence depends upon a certain sort of intellectual activity, a collecting or 'thinking together' as Cantor put it. If sets were collections, that would explain their having the first three features. But of course there are far too many sets for them to be a product of human thinking together; there are many sets such that no human being has ever thought their members together and many such that no human being *could* think them together. That requires an infinite mind - one such as God's.

For a third example, consider the argument from appalling evil. Many philosophers offer antitheistic arguments from evil, and perhaps they have some force. But there is also a *theistic* argument from evil. The premise is that there is real and objectively horrifying evil in the world. Examples would be certain sorts of appalling evil characteristic of Nazi concentration camps: guards found pleasure in devising tortures, making mothers decide which of their children would go to the gas chamber and which be spared; small children were hanged, dying (because of their light weight) a slow and agonizing death; victims were taunted with the claim that no one would ever know of their fate and how they were treated. Of course, Nazi concentration camps have no monopoly on this sort of evil: there are also Stalin, Pol Pot and a thousand lesser villains. These states of affairs, one thinks, are objectively horrifying, in the sense that they would constitute enormous evil even if we and everyone else came perversely to approve of them.

Naturalism does not have the resources to accommodate or explain this fact about these states of affairs. From a naturalistic point of view, about all one can say is that we do indeed hate them; but this is far short of seeing them as intrinsically horrifying. How can we understand this intrinsically horrifying character? After all, as much misery and suffering can occur in a death from cancer as in a death caused by someone else's wickedness. What is the difference? The difference lies in the perpetrators and their intentions. Those who engage in this sort of evil are purposely and intentionally setting themselves to do these wicked things. But why is that objectively horrifying? A good answer (and one for which it is hard to think of an alternative) is that this evil consists in defying God, the source of all that is good and just, and the first being of the universe. What is horrifying here is not merely going contrary to God's will, but consciously choosing to invert the true scale of values, explicitly aiming at what is abhorrent to God. This is an offence and affront to God; it is defiance of God himself, and so is objectively horrifying. Appalling evil thus has a sort of cosmic significance. But of course there could be no evil of this sort if there were no such being as God.

See also: Agnosticism; Atheism; Deism; God, concepts of; Natural theology; Religion and epistemology; Religious experience

ALVIN PLANTINGA

References and further reading

Aquinas, T. (1266-73) *Summa theologiae*, London: Eyre & Spottiswoode, and New York: McGraw-Hill, for Blackfriars, 1964, 60 vols, Ia, qq.2-11.(The *locus classicus* of the cosmological argument; there is also a fuller statement of the first of the five ways in Aquinas' *Summa contra gentiles* I, 12-22.)

Anselm (1077-8) *Proslogion*, in *Opera omnia*, ed. F.S. Schmitt, Stuttgart: F. Frommann, 1984.(Contains the original statement of the ontological argument(s).)

Aristotle (*c*.350 BC) *Metaphysics* 12, in *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984.(An early source for the cosmological argument.)

Aristotle (*c*.350 BC) *Physics* 7, 8, in *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984.(A version of the cosmological argument can also be found here.)

Clarke, S. (1705, 1706) *A Demonstration of the Being and Attributes of God* and *A Discourse Concerning the Unchangeable Obligations of Natural Religion*, London: Faksimile-Neudruck der Londoner Ausgaben. (Clarke's Boyle Lectures, containing an elaborate statement of the 'sufficient reasons' version of the cosmological argument, the one criticized by Hume.)

Clifford, W.K. (1879) 'The Ethics of Belief', in *Lectures and Essays*, London: Macmillan. (Classic statement of the view that one needs evidence in order to be justified (within one's intellectual rights) in believing that there is such a person as God.)

Craig, W. (1980) *The Cosmological Argument from Plato to Leibniz*, London: Macmillan. (An excellent history of the cosmological argument, especially good on the versions given by the medieval Arabic and Jewish philosophers.)

Descartes, R. (1641) *Meditations*, New York: Dover, 1955.(Contains a classic statement of the ontological argument, the statement that (apparently) Kant had in mind in arguing that no version of the argument could be successful; contains also Descartes' own idiosyncratic version of the cosmological argument.)

Gaunilo (*c*.1078) *On Behalf of the Fool*, in Anselm, *Monologion and Proslogion: with the replies of Gaunilo and Anselm*, Indianapolis, IN, 1966.(A criticism of Anselm's ontological argument.)

Hartshorne, C. (1941) *Man's Vision of God*, New York: Harper & Row.(The first modal version of the ontological argument.)

Hawking, S. (1974) 'The Anisotropy of the Universe at Large Times', in M.S. Longair (ed.) *Confrontation of Cosmological Theories with Observational Data*, Dordrecht: Reidel, 72. (An early statement of and response to fine-tuning issues by one of the most eminent contemporary physicists.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. and ed. N. Kemp Smith, London: Macmillan, 1929.(Contains influential criticisms of the cosmological, teleological and ontological arguments; Kant holds that the first two are unsuccessful without the third, but the third is vitiated by the fact (as Kant sees it) that 'existence is not a predicate'.)

Kant, I. (1788) *Critique of Practical Reason*, trans. and ed. L. White Beck, New York: Macmillan, 1993.(Presents a moral argument: postulation of the existence of God is a necessary condition for morality.)

Leibniz, G.W. (1714) *Monadology*, in *The Monadology and other Philosophical Writings*, trans. R. Latta, Oxford: Oxford University Press, 1898.(Contains Leibniz's statement of the 'sufficient reasons' version of the cosmological argument, as well as one of his statements of the ontological argument; Leibniz accepted the

ontological argument, but thought it was necessary to prove and add as a preliminary premise the proposition that it is possible that there is such a person as God.)

Mavrodes, G. (1986) 'Religion and the Queerness of Morality', in R. Audi and W. Wainwright (eds) *Rationality, Religious Belief and Moral Commitment*, Ithaca, NY: Cornell University Press.(Mavrodes argues that if naturalism were true, there would be no such thing as morality; this is therefore a moral argument for the existence of God.)

Paley, W. (1802) *Natural Theology, or Evidences of the Existence and Attributes of the Deity collected from the Appearances of Nature*, London: R. Faulder & Son; Weybridge: S. Hamilton, 1804.(Classic modern statement of the teleological argument.)

Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Clarendon Press.(Contains a version of the ontological argument developed in terms of the metaphysics of modality (possible worlds) that is immune to Kant's criticism (whatever precisely it was).)

Plantinga, A. (1983) 'Reason and Belief in God', in A. Plantinga and N. Wolterstorff (eds) *Faith and Rationality*, Notre Dame, IN: University of Notre Dame Press.(Argues against the claim that propositional evidence (as in the theistic arguments) is needed for rational justification in believing in God; see also the other essays in the collection.)

Plato (*c*.360 BC) *Laws*, Chicago, IL: University of Chicago Press.(One of the earliest adumbrations of the cosmological argument.)

Rowe, W. (1975) *The Cosmological Argument*, Princeton, NJ: Princeton University Press.(A penetrating and thorough examination of the several forms of the cosmological argument.)

Swinburne, R. (1979) *The Existence of God*, Oxford: Oxford University Press.(Contains a very full and sophisticated version of the teleological argument, developed in terms of probability.)

God, concepts of

We think of God as an ultimate reality, the source or ground of all else, perfect and deserving of worship. Such a conception is common to both Eastern and Western religions. Some trace this to human psychology or sociology: Freud regarded God as a wish-fulfilling projection of a perfect, comforting father-figure; Marxists see belief in God as arising from the capitalist structure of society. Believers, however, trace their belief to religious experience, revealed or authoritative texts, and rational reflection.

Philosophers flesh out the concept of God by drawing inferences from God's relation to the universe ('first-cause theology') and from the claim that God is a perfect being. 'Perfect-being' theology is the more fundamental method. Its history stretches from Plato and Aristotle, through the Stoics, and into the Christian tradition as early as Augustine and Boethius; it plays an important role in underwriting such ontological arguments for God's existence as those of Anselm and Descartes. It draws on four root intuitions: that to be perfect is perfectly to be, that it includes being complete, that it includes being all-inclusive, and that it includes being personal. Variously balanced, these intuitions yield our varied concepts of God.

Criticisms of perfect-being theology have focused both on the possibility that the set of candidate divine perfections may not be consistent or unique, and doubts as to whether human judgment can be adequate for forming concepts of God. Another problem with the method is that different accounts of perfection will yield different accounts of God: Ibn Sina and Ibn Rushd, for instance, appear to have held that God would be the more perfect for lacking some knowledge, while most Christian writers hold that perfection requires omniscience.

Views of God's relation to the universe vary greatly. Pantheists say that God is *the universe. Panentheists assert that God includes the universe, or is related to it as soul to body. They ascribe to God the limitations associated with being a person - such as limited power and knowledge - but argue that being a person is nevertheless a state of perfection. Other philosophers, however, assert that God is wholly different from the universe.*

Some of these think that God created the universe ex nihilo, *that is, from no pre-existing material. Some add that God conserves the universe in being moment by moment, and is thus provident for his creatures. Still others think that God 'found' some pre-existing material and 'creates' by gradually improving this material - this view goes back to the myth of the Demiurge in Plato's Timaeus, and also entails that God is provident. By contrast, deists deny providence and think that once God made it, the universe ran on its own. Still others argue that God neither is nor has been involved in the world. The common thread lies in the concept of perfection: thinkers relate God to the universe in the way that their thoughts about God's perfection make most appropriate.*

1 The logic of 'God'

We use the term 'God' in two main ways. We use 'God' to address God, as in prayer. So used, 'God' seems like a proper name. We also use 'God' like a general predicate. For we can and do ask whether there is more than one God: the concept of God allows this question a 'yes' answer. The concept of Moses does not allow a 'yes' answer to 'was there more than one Moses?' One cannot use proper names this way, for there cannot have been two Moseses. If 'two' people are identical with Moses, they are one person, not two. 'Was there more than one Moses?' makes sense only taken as 'was there more than one man called Moses?' or 'did one person do the deeds in the Moses story, or does the story conflate the deeds of many men?'

The ambiguity between name and predicate suggests that 'God' is a title-term, like 'Pastor' or 'Bishop'. Many people can be bishops; in this way title-terms are like general predicates. But one can also address the office-holder by the title ('Dear Bishop…'); one can use the title as a name for the person who holds the office. Thus, the concept of God is a concept of an individual holding a special office.

Thinkers have disagreed about what this office is. Aquinas (§9) suggests that to be God is to have providence over all (*Summa theologiae* Ia, q.13, a.8). But without obviously contradicting themselves, some philosophers (such as Aristotle (§16) and Plotinus (§§3-5)) deny that God provides for creatures. Some say that to be God is to deserve worship. But arguably one's act cannot be an act of worship if one does not take its object to be truly divine, and something cannot deserve worship if it is not divine. If either of these claims is true, 'to be God is to deserve worship' amounts to something like 'to be God is to deserve to be treated as God' - which is true but not helpful.

The suggestion that this entry will explore is that the role or office of God is that of ultimate reality, than which no reality is more basic. Consider some claims on which Eastern and Western religions agree:

(1) Nothing made God; anything that was made is not God.
(2) God is the source or ground of all that is not God.

Eastern and Western religions also agree that God is the Supreme Being. They use 'supreme' to speak of God's perfection and authority and so agree that:

(3) God rules all that is not God.
(4) God is the most perfect being.

(1)-(4) are necessary truths *de dicto*; they state requirements for counting as God. All express facets of God's ultimacy. (1) states that nothing is causally more basic than God. On (1), no account of why things exist can go beyond God; if God explains anything, God is an ultimate explainer. On (2), all explanations of existence do in fact terminate in God. On (3), God has ultimate control over all things; on (4), God is ultimate in perfection. If (1)-(4) are near the core of the concept of God, so is ultimacy.

Quite diverse religious contexts tie ultimacy and deity. Revelation 21: 12 calls God 'the Alpha and the Omega, the beginning and the end'. To be last is to be ultimate. That is first to which nothing is prior. So being first is also a way to be ultimate, and being first is a prerequisite of being taken as God. John Chrysostom comments that 'men most honour the eldest of beings which was before all, and account this to be God' (*Homilies on St John* [1994: 7]). Pantheists claim not to twist the meaning of 'God' in calling the universe 'God' because, as they see it, the universe is the ultimate reality; there is nothing beyond it. Even believers in many gods sometimes distinguish mere gods from God on grounds of ultimacy. Thus, Hindus see their many *devas* (gods) as finally just creations and manifestations of the one ultimate being, Brahman (see God, Indian conceptions of); and Proclus (§4), a Neoplatonic polytheist, writes that:

> God and the One are the same because there is nothing greater than God and nothing greater than the One…[Plato's] Demiurge is a god, not God. The god that is the One is not a god, but God simply.
>
> (*Commentary on Plato's Parmenides* 641-3)

Proclus' One counts as God precisely because it is, in Proclus' view, the ultimate reality. Tillich (§§2-3) sums it up: 'Only [an] ultimate reality can…be our unconditional concern [that is, a genuine God]. Faith in anything which has only preliminary reality is idolatrous' (1955: 59). Thus, concepts of God are concepts of what the ultimate reality is and how it is related to the rest of reality.

2 Data and methods

Human beings draw their views of God from religious experience, revealed or authoritative texts, and rational reflection. The third is, in a way, basic. Not all seeming experience of God is genuine. Texts and philosophical thought help sort the true from the illusory. Then, too, the genuine experiences of God, once found, do not wear their meanings on their sleeves. They need interpretation, which texts and reason provide. Finally, authoritative texts themselves need interpretation, and reason provides this. So religious experience and authoritative texts are data, which philosophical and theological reasoning work into fully articulate theories of the nature of God.

Some trace views of God to still other sources. For example, Freud argued that belief in God is a form of wish-fulfilment, in which human beings give themselves the protected, loved feeling of an idealized childhood by projecting into reality a vision of a perfect, benevolent father-figure who will make everything right in the end. Other psychologists take a still dimmer view of God, seeing him as not a comforting dream but the projection into reality of neurotic self-loathing; they see belief in divine judgment (for example) as a form of self-persecution. Durkheim(§4) (1915) regarded the concept of God as a symbol expressing society's role in our lives. Marxists see this role as oppressive, at least under capitalism, and so see concepts of God as reflecting and reacting to social evils. On their view, belief in God is doomed to wither away when society progresses.

These are all speculations, and beliefs may be true even if one can explain our holding them in purely psychological or sociological terms. But still, theists claim that reason, experience and authority yield *warranted*

belief. On some views, beliefs are warranted only if their sources are reliably truth-producing ways of forming beliefs. Wish-fulfilment and neurosis-expression are not reliably truth-producing ways of forming beliefs. So, on these views of warrant, a persuasive psychological or sociological explanation of apparent religious experience or the intuitions that guide theistic reasoning could undermine theists' claims to warranted belief. Again, some think a belief warranted only if one holds it for the right kind of reason. 'To feel good' is not the right kind of reason.

Still, the concept of God gives as much grist to wish-fulfilment as to self-persecution or Marxist theories. Consequently, the two sorts may cancel out: the facets of God which count *for* the one count *against* the other. Belief in God cannot both comfort and torment us. It can at most comfort one part of us and torment another part. But the more elaborate the genetic account of theism's attractions, the less plausible the account becomes. Furthermore, one can give equally unflattering accounts of the appeal of atheism, Freudianism and Marxism. If so, the genetic argument again cancels out.

Western philosophy works up theories about God in two main ways. *First-cause theology* draws out implications of God's relation to the universe. *Perfect-being theology* reasons out consequences of the claim that God is a perfect being. Each method gets at part of God's ultimacy. First-cause theology explores God's causal ultimacy. Perfect-being theology explores God's ultimacy in value. Most religious thinkers use both methods. Aquinas, for instance, sets up his basic descriptions of God by giving five ways to argue for God's existence (*Summa theologiae* Ia, q.2, a.3). The first three ways argue the existence of a first cause, the fourth that of a perfect being, and the fifth that of an intelligent designer of nature. Aquinas then argues that a first cause must be purely actual and so perfect and uses both perfect-being and first-cause theology to work out his concept of God.

Perfect-being theology is the more basic theological method. In the West, for instance, first-cause theology often explicates and/or argues the claim that God created the universe *ex nihilo* (see Creation and conservation, religious doctrine of §§1-2). But while Western Scriptures say that God created the universe, they do not so clearly state that God made it *ex nihilo*. The most famous biblical passage on creation is Genesis 1: 1-2: '[1]In the beginning God created the heavens and the earth. [2]Now the earth was formless and empty, darkness was over the surface of the deep, and the Spirit of God was hovering over the waters.' Verse 1 could assert a creation *ex nihilo*, with verse 2 describing that act's initial result. But verse 1 could just be a topic statement for the overall creation story, which might begin in verse 2. In that case, in the story, God finds some initial chaotic state of things, and creates a universe by forming or moulding that chaos into the structured world we now inhabit. This second reading does not involve creation *ex nihilo*. It is perfect-being theology that leads Western thinkers to take creation as *ex nihilo*; by doing so, they magnify things' dependence on God and God's importance for and superiority to the universe. We ought, then, to focus on perfect-being theology.

3 Perfect-being theology

Perfect-being theology has a long history. Plato (§14) assumes that gods are 'the…best possible' to argue that they are immutable (*Republic* II, 381c). To work out the nature of God, Aristotle takes as a premise that God is 'the best substance' (*Metaphysics* XII, 9, 1074b). Cicero (§3) records the perfect-being arguments of the early Stoic Zeno of Citium:

> That which is rational is better than that which is not rational. But nothing is better than the cosmos. Therefore the cosmos is rational. One can prove in a similar manner that the cosmos is wise, happy and eternal, since all of these are better than things which lack them, and nothing is better than the cosmos.
>
> (*On the Nature of the Gods* II, 21)

Zeno plainly envisages many applications of this basic argument form:

(5) Whatever is *F* is more perfect than whatever is not *F*.
(6) Whatever is most perfect is more perfect than everything else.
(7) So whatever is most perfect is *F*.
(8) *X* is most perfect.
(9) So *X* is *F*.

The project of filling out a concept of God by successive applications of (5)-(9) is a version of perfect-being theology. To many, Zeno's taking 'the cosmos' as a value of *X* in (8) will look odd. Cicero's case for Zeno is that

‘since the cosmos includes everything and since there is nothing which is not in it, it is perfect in all respects’ (*On the Nature of the Gods*). Cicero’s thought is perhaps this: the cosmos is composed of other objects, its parts. So the cosmos’ perfection is the sum of its parts’ perfections, and thus it is greater than the perfection of every object it contains. But the cosmos contains every object other than itself. So the cosmos is the most perfect actual being.

Cicero seems even to think that the cosmos is the most perfect *possible* being:

> Nothing at all is better…than the cosmos. Not only *is* nothing better, but nothing can even be conceived of which is better. And if nothing is better than reason and wisdom, it is necessary that these be present in that which we have granted to be the best.
>
> (*On the Nature of the Gods* II, 18)

Reasoning out the nature of a greatest *possible* being is another form of perfect-being theology. Taking the cosmos as the greatest *possible* being, Cicero applies to ‘reason’ and ‘wisdom’ a form of argument which, like (5)-(9), yields a general theological programme. Why might Stoics call the cosmos the most perfect possible being? Suppose that to be a cosmos is to be all-inclusive, a sum of all other things. If so, then no matter what, if anything exists, a cosmos also exists: there is a sum of all things, and this sum is the most perfect actual thing. Suppose we now add that no matter what objects made it up, the cosmos which existed would be the same individual cosmos as the one which now exists. It then follows that the cosmos is the greatest possible being, in the sense that no matter what possible things existed, our cosmos would be greater than any of them. Stoics may have thought this way (see Stoicism §§3-5).

The Christian tradition invokes perfect-being theology at least as early as Augustine (§7), who wrote: ‘When we think of…God…thought takes the form of an attempt to conceive something than which nothing more excellent or sublime exists’ (*Christian Doctrine* I, 7, 7). Augustine tells us to construct a concept of God by seeking a concept of the most perfect being there is. He adds two concrete rules. One is to deny God whatever attributes we think to be imperfections, the other is to affirm of God in the highest degree any attributes we find to be perfections (*On the Trinity* V, 1, 2; XV, 4, 6).

Boethius (§5) stated an axiom stronger than Augustine’s, that God is such that nothing greater than him is even conceivable (*The Consolation of Philosophy* III, 10). Anselm of Canterbury (§§3-4) at first based his perfect-being theology on Augustine’s axiom, that God is the greatest actual being (*Monologion*, chaps 1-3). To fill out the concept of God, Anselm directed, ascribe to God all attributes *F* such that whatever is *F* is better than whatever is not *F* (*Monologion*, ch. 15). Thus Anselm’s *Monologion* recaps Zeno’s theological programme. In his *Proslogion*, Anselm switches to Boethius’ axiom, and suggests filling out the concept of God by many arguments of this form:

(10) Nothing greater than God is conceivable.
(11) If God is not *F*, something greater than God is conceivable.
(12) So God is *F*.

Replacing ‘is conceivable’ with ‘is possible’ yields a form of argument theistic philosophers still respect. Replacing ‘*F*’ with ‘existent’ yields Anselm’s ontological argument for God’s existence (see God, arguments for the existence of §2).

4 Limits of perfect-being theology

Duns Scotus’ gloss on Boethius’ axiom calls God the greatest being conceivable without contradiction (*On the First Principle*, 4.65) (see Duns Scotus, J. §§7-11). So perhaps Duns Scotus saw a question Leibniz (§3) was to raise explicitly: what if some values of *F* are incompatible? What if nothing can have all values of *F*? Perfect-being theology replies: if so, God has a set of compatible attributes *S* which does not include all values of *F*, such that nothing can be greater than a thing with all members of *S*. This raises another question. Suppose that there is such a set. There may also be a second set *T* such that nothing can be greater than a thing with all members of *T*, but either an *S*-God would be just as great as a *T*-God or an *S*-God would be neither greater than, less great than nor as great as a *T*-God. What then? Perfect-being theology can say that God has any attributes common to *S* and *T* and can try to decide between *S* and *T* on further grounds, but may have to settle for saying that God has either *S* or *T*. There is no guarantee that perfect-being theology alone can fill out the concept of God.

The metaphysics one brings to perfect-being theology affects one's conclusions. The Stoics agreed with Christians that being an agent is a perfection. But the Stoics thought that only material things can act. They inferred that God must be material and so must be the most perfect material thing, the cosmos. Being perfect is being maximally good, and so thinkers' intuitions about what is perfect can differ and be as hard to reconcile as their intuitions about what is good. But perfect-being theology yields divergent results when those who do it do not share intuitions about what it is to be perfect. Anselm, a Christian, thought that a perfect being must be all-knowing (*Proslogion*, ch. 6). But Alcinous, a second-century Platonist, held that all God's ideas are 'eternal and perfect' and therefore denied that God has ideas - or therefore knowledge - of disease, artefacts, dirt or individuals as such (*The Handbook of Platonism*, 9.2, 7; 11.27-31). Ibn Sina and Ibn Rushd too seem to have thought that a perfect being has some but not all knowledge - that there is knowledge that is not good, so that God is more perfect without it (see Ibn Rushd §2). It is not clear that perfect-being theology can adjudicate such differences. So while it may help one work out one's own concept of God, perfect-being theology may have less promise as an avenue of theological agreement. This might in turn cast doubt on perfect-being theology's ability to give theological beliefs *warrant*.

Some theists reject perfect-being theology. Plotinus held with Plato that goodness is 'superior to being' (*Republic* VI, 509b). He thus refused to use perfect-being theology to explicate his view of the truly ultimate God, the One or Good; Plotinus' 'perfect being' is his second, subordinate god, *Nous*, the divine mind. For Plotinus, then, being truly ultimate is incompatible with being the subject of perfect-being theology. Theists who hold that God is utterly ineffable or entirely beyond our conceptual grasp doubt that perfect-being theology's development of human concepts of perfection can really explicate God's nature. As one's concept of perfection reflects one's concept of goodness, theists in the tradition of Augustine and Calvin (§§2-3), who emphasize the corrupting effects of sin on the human power to judge what is good, find perfect-being theology radically unreliable. Freudian views of religious belief can also question perfect-being theology, for it can look like explicit, conscious wish-fulfilment: is it not a way of ascribing to God a 'wish list' of perfections? Finally, those who criticize the ontological argument for God's existence may also criticize perfect-being theology. For perfect-being theology underwrote the ontological argument in Anselm and Descartes (§6), and its connection with that argument is not accidental. Perfect-being theology claims in effect that in God's case, 'ought' implies 'is', and one thing God ought to be is existent.

5 Intuitions about perfection

Four intuitions about what it is to be perfect are found in the history of philosophical theology. One is that to be perfect is perfectly to be, and so draws on an account of what it is to *be*. We find this account and its consequences in Parmenides (§6) and in Plato's account of his Forms. Augustine puts it this way:

> Something which changes does not retain its own being. A thing which can change, even if it does not change, cannot be what it had been. So only what does not only not change but also cannot at all change falls most truly…under the category of being.
>
> (*On the Trinity* V, 2-3)

A God perfect in this way must be wholly discrete from the universe: the universe changes, God does not.

The Latin root of 'perfect', *perficio*, has the sense of 'complete'. Thus, a second perfection-intuition has been that what is perfect lacks nothing, is fully self-sufficient, and so is independent of all else. If persons by nature change, are finite or incomplete in various ways, or require others for their full flourishing, a God whose perfection reflects these first two intuitions may be personlike, but is not in the ordinary sense a person. A third intuition is that what is perfect, being complete, is all-inclusive. Classical theism takes God to be all-inclusive in perfection. Thus, Aquinas argues that if a being lacks some perfection, it is not a God (*Summa theologiae* Ia, q.11, a.3). The Stoics and Parmenides took being perfect as being all-inclusive and reasoned that as all-inclusive, a perfect being must include the universe. One way to develop this thought holds that God is the universe, and so sees God as impersonal. A second way to develop this view sees God as including, but also as being 'more than', the universe, just as persons include but are more than their bodies. The last intuition takes persons as the standard of perfection, and so insists that any perfect being must know, will and (according to some philosophers) feel.

Below is a survey of some of the main concepts of God that these intuitions have yielded.

6 Classical theism

Classical theism's ancestry includes Plato, Aristotle, Middle Platonism and Neoplatonism. It entered Judaism through Philo of Alexandria (§4), reaching its apogee there in Maimonides (§3). It entered Christianity as early as Irenaeus and Clement of Alexandria and became Christian orthodoxy as the Roman Empire wound down. Though more and more challenged after 1300, it remains orthodox. Classical theism filtered into Islam as early as al-Kindi (§§1-2). Al-Ghazali attacked it as the view of Islamic Aristotelians, and it suffered in Islamic orthodoxy's successful reaction against Aristotle.

Much of classical theism's concept of God unfolds from the claim that God is the ultimate reality. According to classical theism, God is:

A se - wholly independent of all else. God is absolutely the first being. He exists before there is anything else for him to depend on. So he must need or depend on nothing in any way other than himself. (Classical theism sits well with intuitions that, as perfect, God is self-sufficient.)

Simple - completely without parts. Whatever has parts depends on them for its existence and nature; bricks make a wall and make it what it is. So a God wholly *a se* has no parts (see Simplicity, divine).

Having no parts, God is:

Immaterial - whatever is made of matter has parts, for the matter of which it is made has parts. So a partless God cannot be made of matter, or include matter in any way which makes its parts his.

Not spatially extended - whatever extends through space has parts covering the parts of the space through which it extends (or so most think).

Without accidents - lacking non-relational contingent attributes. If God has attributes distinct from himself, he depends on them for his nature and existence. So if he is entirely *a se*, he has no such attributes: God's attributes = God. Nothing can be identical with a contingent attribute. For if one has an attribute contingently, one can only be contingently identical with it. But all genuine identities are necessary. So God has no accidents.

If so, God is:

Immutable - or unable to undergo real, intrinsic change. For real, intrinsic change is change in accident (see Immutability).

If *a se*, God is also:

Impassible - unable to be affected by beings other than himself. For if we affected God - if, say, our suffering made him sorrowful - his emotional state would depend on us, and so God would not be wholly *a se*.

According to classical theism, God is also:

Eternal - in the sense of timeless, that is, alive without past or future, living a life neither containing nor located in any series of earlier and later events. Much traditional perfect-being theology converges on this claim. Boethius, for instance, argued that a perfect being must be timeless because timeless existence is superior to temporal existence. Temporal beings lose their pasts and lack their futures, and so enjoy only an instant-thin slice of their existences at any one time. Timeless life has no past or future. A timeless being enjoys its entire life in one timeless present. Thus Boethius saw timeless life as 'the all-at-once and perfect possession of interminable life' (*The Consolation of Philosophy* V, 6), and so most appropriate to a perfect being (see Eternity).

Necessarily existent - perfect-being theology yields divine necessity. To exist contingently is to be able not to exist. A being is more perfect if wholly immune to nonexistence (see Necessary being).

Omnipresent - present in all space and time, though not contained by either (see Omnipresence). This follows via perfect-being theology: a God not in some way everywhere and everywhen would be more limited and less perfect than a God with these attributes. As creator and sustainer, God is present everywhere and everywhen in the sense that he sustains in being and knows immediately every place and time and their contents.

Classical theism thinks God personal enough to have intellect and will. Perfect-being theology backs this if

nothing incompatible with intellect and will adds more value to the concept of God. If God has intellect and will, perfect-being theology ascribes to him also the perfect versions of these.

Perfect intellect includes *perfect wisdom and rationality* and *perfect knowledge*. It thus includes omniscience, variously defined as knowing all truths and/or all facts, or all that is knowable, or having the greatest amount of knowledge possible for a single individual (see Omniscience). Classical theism typically holds that God's knowledge includes knowledge of free creaturely actions that, with respect to us, are future; this raises the question of how such knowledge is compatible with creaturely freedom. Perfection in knowledge also includes having knowledge in the best possible way (perhaps immediately rather than inferentially, or by direct intuition of fact rather than through grasping some representation such as a proposition) and on the best possible grounds. As perfectly wise or rational, God makes optimal use of his knowledge.

Perfect will includes *perfect power* and *perfect goodness*. Perfect power includes omnipotence, defined roughly and with some qualifications as the power to actualize any broadly logically possible state of affairs (see Omnipotence). Perfect goodness includes always acting as moral norms dictate, doing great supererogatory good and having perfect versions of at least some moral virtues (justice, mercy, altruistic love). The claim that God is perfect in knowledge, power and goodness sets up the various versions of the problem of evil (see Evil, problem of). The simplest version is this: if God knows that and when each evil will occur if he does not intervene, and God has the power to prevent each evil, and God is perfectly good and so does not approve of evil, how then can evil occur?

Classical theism constrains what we make of God's personal attributes. For instance, if wholly *a se*, God cannot know creatures by observing them, since states of observational knowledge causally depend on the objects observed. If I know that Fido is here by seeing Fido, then as Fido helps cause my seeing, Fido helps cause my knowing. Perfect-being theism also affects our view of God's personal attributes. It would be better to be necessarily omniscient, omnipotent and so on than merely contingently so. So perfect-being theism pushes us to say that God has necessarily his personal perfections. This can raise questions about, for example, God's goodness: some argue that if God necessarily does no wrong, he does not qualify as a morally responsible agent (see Freedom, divine).

Classical theism holds God wholly discrete from the world in substance. Classical theists differ over God's relation to the world. Aristotle held that God is merely a final cause, a lure who draws the universe's efficient causes into action (*Metaphysics* XII, 6-7). But most classical theists hold that God efficiently causes the world to exist by creating it from no pre-existing material; classical theists often compare this with our thinking up imaginary worlds just as we choose (without implying that the universe is a figment of God's imagination).

Most classical theists hold that God created freely and directly rather than via an intermediary. Other classical theists demur. Plotinus, for instance, argued that his ultimate reality, the One, creates necessarily (*Enneads* II, 9, 3), and has as its direct effect only a second deity, the divine mind, which creates all 'lower' beings. Both claims have had later partisans in classical theism - for example, such Arabic Aristotelians as Ibn Sina and Ibn Rushd.

Most classical theists who hold that God creates also hold that God conserves the world - that is, they hold that the universe depends on God for its being not merely in its first instant of existence (if any), but equally throughout its duration. For classical theists, God would not have to do anything positive to annihilate the universe. He would merely have to stop doing what he is always doing to keep it in existence. As a book falls to the floor once a supporting hand is withdrawn, so the world would fall back into nothingness were God's supporting power withdrawn (see Creation and conservation, religious doctrine of §5). If God conserves the world, God also is provident for his creatures, at least in keeping in being the things they need. Creation and conservation raise again the problem of evil: if all things stem from God and God actively preserves them, how can God avoid responsibility for the evils things bring?

Classical theism's God is infinite or unlimited in not depending on other things, and in perfection, power, knowledge, goodness and creative responsibility. Thus, to some, the God of classical theism is personlike but not a person (as some say of Brahman, conceived as being, consciousness and bliss). Some have denied classical theism because they see God as limited in some of these ways and more like the persons we know. Some deists denied divine providence, and perhaps conservation (see Deism §§1-2). Plato held that the Demiurge 'found' a disordered

cosmos, and 'created' only by bringing it into better order (*Timaeus*, 30a), thus freeing him from responsibility for evil (*Republic* II, 379c). Theists such as John Stuart Mill and William James (§4) tried to free God of this responsibility by supposing that God has only limited power and/or knowledge.

7 Pantheism

Pantheism consists of two theses: that (a) all things are parts of, appearances of or really identical with some one being, and (b) this being is divine. Claim (a) is compatible with the view that all items not identical with this one being are unreal, or ultimately unreal, but does not entail it. Nor does (a) entail (b), for one who holds that there is just one really real or all-inclusive thing need not treat it as divine. Still, if there is just one thing, or an all-inclusive thing, that thing is the ultimate thing, and has a claim to be the most perfect thing. So given the tie between ultimacy and deity and the tradition of perfect-being theism, those who hold (a) naturally tend towards (b).

Reverence for nature and mystical experiences of the unity of all things are a perennial source of pantheism, the former persuasive to nineteenth-century Romantics, the latter to Eastern thinkers. Once tilted towards this view of God, pantheists sometimes argue that their Unity has some attributes classical theism calls distinctively divine. Stoics argue that the universe is conscious and wise, Spinoza (§§2-4) that the one universal substance is simple, Hegel that the world-spirit exercises a kind of providence, and some contemporary philosophers that the universe viewed tenselessly exists independently and is eternal. The universe is certainly omnipresent in the sense that every place is part of it, though not as in classical theism (in which God is present *as a whole* in every place and time). In pantheism, God = universe, and so the claim that God and universe are causally related is problematic. Spinoza tries to preserve an efficient-causal relation by distinguishing an active aspect of the universe (*natura naturans*) from a passive (*natura naturata*). Hegel sees God as the universe's final cause, claiming that full realization of its divine nature is the goal towards which the universe's development moves.

The line between pantheism and panentheism is narrow. Pantheism holds that God is the sum of, the reality behind or the true identity of all things, and nothing beyond this. If a being with awareness and/or will is in some way more than its body, then a view can remain pure pantheism only by denying that God is a person or personlike, and many thinkers usually called pantheists (for example, Spinoza, the Stoics, some Hindus) may actually have been panentheists. Many find pantheism's implications distasteful. If everything is part of God, or an aspect of him, and so forth, then God includes or really is the most unsavoury items, and (for example) whoever eats, chews on God. Thus 'pantheist' often serves as a term of theological abuse, and not everyone given the label deserves it.

8 Panentheism

Panentheism can be traced back at least to the Hindu thinker Rāmānuja (§3); its chief twentieth-century friends have been 'process philosophers' (see Process theism). Panentheism seeks a middle ground between classical theism and pantheism, preserving the former's claim that God has intellect and will and the latter's sense of intimate connection between God and universe. In panentheism, God is a person who includes the universe, or a soul whose body is the universe. Thus, God is larger than but also like the persons we know: physically embodied in a finite material object, growing older through time and changing as his material parts change. Panentheists claim that even so, God is perfect, because persons are the acme of perfection. Panentheists deny that God is immaterial; most also affirm that he has spatial extension, and offer their own distinctive slants on other traditional divine attributes.

The panentheist's God includes other things, and so depends on them. Such a God is thus not *a se* in the classical sense. But such a God may still be more independent than anything else. He depends only on things he includes, in contrast to depending on things that include him or that lie beyond him. Some panentheists (such as Charles Hartshorne) see some sorts of dependence as perfections. Compassion and sympathy are (they argue) moral perfections that involve having feelings that depend on others' fortunes, and all knowledge depends on the objects known. The panentheist's God has the complexity of the universe and changes as it does. But some panentheists (such as Hartshorne) maintain that God's abstract nature is simple, while Rāmānuja sites all complexity in God's body, the universe, holding that God's inner self, the pure Brahman, is partless. Hartshorne and Rāmānuja claim that God is immutable in certain respects, Hartshorne that his basic character (perfect benevolence, and so forth) cannot change, Rāmānuja that all change occurs in God's body, the universe, not in Brahman proper. Panentheists

hold that God is eternal in the sense of everlasting in time; Hartshorne defends a version of the ontological argument, contending that God exists necessarily. Process panentheists usually claim that God cannot know the future, as it is indeterminate and so unknowable, and that God's power, while greater than any other being's, is not literally omnipotence. In this they follow Plato, whose Demiurge seeks to mould the world into a better place, but is limited by the recalcitrant material with which he works.

See also: Bible, Hebrew; Epicureanism §9; Feminist theology; God, arguments for the existence of; Goodness, perfect; Incarnation and Christology §1; Kabbalah §3; Monotheism; Pantheism; Postmodern theology; Theology, Rabbinic; Tian; Trinity

BRIAN LEFTOW

References and further reading

Alcinous (2nd century AD) *The Handbook of Platonism*, trans. J. Dillon, New York: Oxford University Press, 1993.(Synopsis of Middle Platonism.)

Alston, W. (1989) *Divine Nature and Human Language*, Ithaca, NY: Cornell University Press.(Important essays on divine action, knowledge and timelessness.)

Anselm (1076) *Monologion*, trans. J. Hopkins and H. Richardson, in J. Hopkins and H. Richardson (eds and trans) *Anselm of Canterbury*, Toronto, Ont.: Edwin Mellen Press, 1974, vol. 1.('Perfect-being' account of God's existence and nature.)

Anselm (1077-8) *Proslogion*, trans. J. Hopkins and H. Richardson, in J. Hopkins and H. Richardson (eds and trans) *Anselm of Canterbury*, Toronto, Ont.: Edwin Mellen Press, 1974, vol. 1.(*Locus classicus* of perfect-being theology.)

Aquinas, T. (1266-73) *Summa theologiae*, New York: Benziger Brothers, 1948. (*Locus classicus* of classical theism.)

Aristotle (mid 4th century BC) *Metaphysics*, trans. W.D. Ross, in R. McKeon (ed.) *The Basic Works of Aristotle*, New York: Random House, 1941.(Primary source for his philosophical theology.)

Augustine (397-426) *Christian Doctrine*, trans. F. Shaw, in P. Schaff (ed.) *Nicene and Post-Nicene Fathers*, series 1, Peabody, MA: Hendrickson Publishers, 1994, vol. 2.(On how to interpret Scripture.)

Augustine (398-425) *On the Trinity*, trans.A. Haddan and W. Shedd, in P. Schaff (ed.) *Nicene and Post-Nicene Fathers*, series 1, Peabody, MA: Hendrickson Publishers, 1994, vol. 3.(His central treatment of the nature of God.)

Augustine (388) *On the Morals of the Manichaeans*, trans. R. Stothert, in P. Schaff (ed.) *Nicene and Post-Nicene Fathers*, series 1, Peabody, MA: Hendrickson Publishers, 1994, vol. 4.(Polemical work, on ethics and metaphysics.)

Boethius, A.M.S. (*c*.522-4) *The Consolation of Philosophy*, trans. V.E. Watt, New York: Penguin, 1969.(On the nature of happiness, luck, freedom and divine foreknowledge.)

Chrysostom, J. (390-8) *Homilies on St. John*, trans. P. Schaff, in P. Schaff (ed.) *Nicene and Post-Nicene Fathers*, series 1, Peabody, MA: Hendrickson Publishers, 1994, vol. 14.(Series of occasionally philosophical sermons.)

Cicero (45 BC) *De natura deorum*, trans. H. Rackham, *On the Nature of the Gods*, London: Heinemann, 1933. (Rich source of Graeco-Roman philosophical theology.)

Davis, S.T. (1983) *Logic and the Nature of God*, Grand Rapids, MI: Eerdmans. (Readable account of central divine attributes; slightly harder than Morris (1991).)

Duns Scotus, J. (*c*.1306) *On the First Principle*, trans. A. Wolter, Chicago, IL: Franciscan Herald Press, 1966. (Fullest version of his argument for God's existence, compact outline of his view of God.)

Durkheim, É. (1915) *The Elementary Forms of the Religious Life*, London: Allen & Unwin.(Sociological account of religion.)

Freud, S. (1927) *The Future of an Illusion*, trans. J. Strachey, New York: W.W. Norton, 1961.(Psychological critique of religious belief.)

Hartshorne, C. (1948) *The Divine Relativity*, New Haven, CT: Yale University Press.(Exposition of 'process' panentheism.)

Hartshorne, C. and Reese, W. (1953) *Philosophers Speak of God*, Chicago, IL: University of Chicago Press. (Selective, idiosyncratic history, coloured by the authors' panentheist views.)

Levine, M. (1994) *Pantheism*, New York: Routledge.(Fullest presentation of this view to date.)

Morris, T.V. (ed.) (1987) *The Concept of God*, New York: Oxford University Press. (Collection of important recent articles. Rigorous but not technical.)

Morris, T.V. (1991) *Our Idea of God*, Notre Dame, IN: University of Notre Dame Press. (Good introduction to

issues in philosophical theology.)
Plato (*c*.380-367 BC) *Republic*, trans. G.M.A. Grube and C.D.C. Reeve, Indianapolis, IN: Hackett Publishing Company, 1992.(The nature of justice, in individuals and society.)
Plato (*c*.366-360 BC) *Timaeus*, trans. B. Jowett, in E. Hamilton and H. Cairns (eds) *Collected Dialogues of Plato*, Princeton, NJ: Princeton University Press, 1961.(His cosmology, with some attention to philosophical theology.)
Plotinus (*c*.250-66) *Enneads* II, trans. A.H. Armstrong, Cambridge, MA: Harvard University Press, 1979.(First Neoplatonist metaphysical system.)
Proclus (*c*.460-70) *Commentary on Plato's Parmenides*, trans. G. Morrow and J. Dillon, Princeton, NJ: Princeton University Press, 1987.(Really an exposition of Neoplatonist metaphysics and theology.)
Sontag, F. (1962) *Divine Perfection*, New York: Harper.(Impressionistic account of relevant intuitions.)
Swinburne, R. (1993) *The Coherence of Theism*, New York: Oxford University Press, revised edn.(Clear, careful treatment of central divine attributes.)
Tillich, P. (1955) *Biblical Religion and the Search for Ultimate Reality*, Chicago, IL: University of Chicago Press. (On religion's relation to philosophy.)

God, Indian conceptions of

*In the Ṛg Veda, the oldest text in India, many gods and goddesses are mentioned by name; most of them appear to be deifications of natural powers, such as fire, water, rivers, wind, the sun, dusk and dawn. The Mīmāṃsā school started by Jaimini (*c. AD *50) adopts a nominalistic interpretation of the Vedas. There are words like 'Indra', 'Varuṇa', and so on, which are names of gods, but there is no god over and above the names. God is the sacred word (mantra) which has the potency to produce magical results. The Yoga system of Patañjali (*c. AD *300) postulates God as a soul different from individual souls in that God does not have any blemishes and is eternally free. The ultimate aim of life is not to realize God, but to realize the nature of one's own soul. God-realization may help some individuals to attain self-realization, but it is not compulsory to believe in God to attain the* summum bonum *of human life. Śaōkara (*c. AD *780), who propounded the Advaita Vedānta school of Indian philosophy, agrees that God-realization is not the ultimate aim of human life. Plurality, and therefore this world, are mere appearances, and God, as the creator of the world, is himself relative to the concept of the world. Rāmānuja (traditionally 1016-1137), the propounder of the Viśiṣṭādvaita school, holds God to be ultimate reality, and God-realization to be the ultimate goal of human life. The way to realize God is through total self-surrender to God. Nyāya theory also postulates one God who is an infinite soul, a Person with omniscience and omnipresence as his attributes. God is the creator of language, the author of the sacred Vedas, and the first teacher of all the arts and crafts.*

1 Mysticism in the *Ṛg Veda*

The Vedic texts (*c.*2000-1500 BC) suggest that divinity permeates the whole universe, the fields, the crops, animals, human speech, and so on. This conception involves many gods and goddesses, but it is not polytheism, for each god or goddess is praised as the supreme deity in the verses addressed to them. Max Müller has coined the term 'henotheism' to denote systems where the position of supreme deity is occupied now by one god or goddess, then by another, and so on (see Chatterjee and Datta 1984). No more than one god is ever worshipped at a time. There is also an emphasis on monotheism, pantheism and mysticism; the different gods and goddesses are said to be different only in name.

It is often claimed that the *Ṛg Veda* (*c.*200 BC) represents nature worship, polytheism and a rather crude imagery of gods and goddesses. This claim is certainly unjustified, for all through the *Ṛg Veda* there is a sense of all-pervasive divine presence, immanent in humanity and nature, unifying the universe and illuminating it. To the seers (*ṛṣis*) of the *Ṛg Veda*, nature is a living divine presence with which they can have communion and be existentially united. They worship the sun, the moon, the dawn, the night, the firmament, fire, water, rivers, seasons, herbs, trees, forests, the grass, furrows, agriculture, mountains, stones, wind and so forth not as mere natural forces, not as anthropomorphic gods and goddesses, but as the various expressions of one divine foundational reality. 'The real is one, the learned call it by various names: Agni, Yama, Mātariśvan' (*Ṛg Veda* 1.164.46). The only mode of worship is performing sacrifices to gods and goddesses. It is not merely that the gods are one, but that the whole of the universe in its physical, material and natural aspects is permeated by one divine presence. Even in the description of the horse sacrifice (1.162), the entire sacrifice, the apparent victim, all the implements used and actions performed are united in the all-comprehending divine presence. The result of this sacrifice, too, is all-pervasive, mundane as well as spiritual.

The whole of *Ṛg Veda* 1.164 is permeated with a sense of divine unity in all creation. The human being has a body and a soul, which is the principle of consciousness. This consciousness has two levels. At the surface level, the human subject knows objects, performs actions and experiences pleasure and pain. Then there is the deeper level of consciousness, where the subject is simply the witness of the empirical self, unaffected by its pleasures and sufferings, by wants and needs, by motives and intentions. This deeper consciousness is the same as the divinity of the whole universe. This the predominant tradition in the *Ṛg Veda*, but there is a second tradition, which is found in one hymn of the tenth chapter (10.136). This is the tradition of the Munis, who are completely different from the members of mainstream society. Physically, the Munis have long hair, are dirty and wear yellow clothes. They are wanderers, not living in societies or performing any social function and not performing any sacrifices. The Munis claim that spiritual realization brings magical powers - a doctrine that is otherwise absent in the *Ṛg Veda*. According to Sāyana (after *c.*AD 780) the Munis 'become gods by the might of their penance'. This is also the first

time we find a reference to penance in the *Ṛg Veda*. The Munis have the power to transmit spiritual excellence and power to others: 'Exhilarated by the sanctity of the Munis, we have mounted upon the winds' (10.136.3). The plurality of gods and goddesses does not make any difference to the spiritual realization of the Muni, who is equally attached to all of them, 'the friend of each deity' (10.136.4). All persons also regard the Munis as their true friends, 'sweet and most delightful friends' (10.136.6).

2 The Pūrva-Mīmāṃsā conception

God is supposed to be known from the revealed scriptures (Vedas). Pūrva-Mīmāṃsā philosophers reject this theory. The revealed texts are only the source of knowing our duty, which is to perform sacrifices. There are names of gods and goddesses, but they do not have any reference beyond these names; there is no divinity beyond the names. There is nothing real beyond the texts. The revealed texts are eternal; they do not have any author. So God is not their author. The universe, too, is eternal and self-sustaining; God cannot, therefore, be postulated as the creator, sustainer and destroyer of the world. So gods and goddesses are just names that have magical powers to produce merit. When performing sacrifices one has to utter the sentences very carefully; as a result, one goes to heaven after death. Neither God-realization nor self-realization is the ultimate aim. In heaven, there are no gods or goddesses, for they are only names. To be in heaven is to enjoy bliss eternally; there can be no fall from heaven (see Mīmāṃsā).

3 The Vedānta conception

According to Śaōkara (*c.* AD 780), who propounded the Advaita Vedānta school of Indian philosophy, God-realization is not the ultimate aim of human life. God is one; he is the creator, sustainer and destroyer of the world, and is therefore omnipotent and omniscient, and the object of worship. But God is not the ultimate reality. For plurality, and therefore this world, are mere appearances (*māyā*), and God, as the creator of the world, is himself related to appearance. Ultimate reality (Brahman), on the other hand, is beyond all relations to objects, to humanity, to thought and to language. Ultimate reality is pure, impersonal, universal consciousness, devoid of any internal or external difference. This consciousness is identical with the self of the individual, which is different from other selves and material objects only at the surface level. At the deepest level, every person, as also every material object, is identical with Brahman. All reality is Brahman. One cannot worship Brahman, for worshipping presupposes a distinction between the subject and object of worship. Brahman is pure subject and can never become an object of worship or even of knowledge. One can only *be* Brahman, that is, realize that one is pure, impersonal, universal consciousness (see Brahman; Vedānta).

In the Vedānta system, God is conceived under two aspects: God pervades the world, and is also beyond it. God is both immanent and transcendent. The world of material objects and living beings contains a plurality of innumerable objects. The question which arises here is: how can the immanent unity be reconciled with the phenomenal plurality? The two major schools of Vedānta, the Advaita Vedānta of Śaōkara and the Viśiṣṭādvaita of Rāmānuja, answer this question differently. According to Śaōkara, manyness or plurality is inconsistent with unity, and as the inner reality is one, the surface reality, that is, the phenomenal world, is only apparent. This apparent world, the world-show, cannot be really related to the inner unity which is God. The relation between God and the world is like the relation between a piece of rope and the illusory snake for which it is mistaken. The piece of rope is, of course, necessary as the foundation of the illusory snake. So also God is the foundation of the illusory world. The cause of this illusory appearance is *māyā*, which has the dual function of covering reality and projecting illusory objects. The world projected by *māyā* is not real, but it is not wholly unreal. Thus the illusory object, the appearance, is neither real nor unreal. It is not real, for when one has spiritual illumination the world-show vanishes, just as, in good light, the illusory snake vanishes and the piece of rope is seen (see Error and illusion, Indian conceptions of §1). Nor is it wholly unreal, for a wholly unreal thing (such as a barren woman's son) cannot be perceived. The world is not unreal, for it is perceived to be real; the world is not real, for it is contradicted by spiritual realization of the universal unity. The bare identity of consciousness is the only reality; this consciousness is universal, impersonal, eternal. If the conception of God is necessarily that of a Person, then this universal consciousness cannot be called God; it is Brahman. So, according to Śaōkara, Brahman is not the creator of the world, for the world is not eternal, and whatever follows from Brahman must partake of its eternal reality. Brahman is the foundation of the world in the sense in which a real object is the foundation of an illusory object. There cannot be any illusion unless there is a real object which is mistaken for an illusory object. Thus illusion

differs from hallucination. Identity cannot tolerate any difference; so Brahman is pure identity devoid of all internal and external difference. Being aware of identity with Brahman, the illumination of the real self, is the ultimate aim of life.

Rāmānuja, on the other hand, argues that the empirical world of plurality cannot be illusory, for it is reality related to God, who is infinite. The world is, of course, not created by an antecedently existing God; God and the world are necessary for each other. Rāmānuja, in his Viśiṣṭādvaita philosophy, conceives the relation between God and the world on the analogy of the self and body of a human being. Just as a human being is an embodied consciousness, so also God is the self and the world is his body. The entire empirical universe is the body of God, who, as the indwelling spirit, is immanent in the universe. But even though both God and the world are real, still God as the soul of the world has supremacy over the world. Thus unity and plurality are both real; the ultimate reality - God - is unity in difference. To see God through self-surrender, not to be identical with him, is the ultimate aim of human life. Going to heaven cannot be the *summum bonum*, for one can stay in heaven only so long as one's merit is not exhausted. So there is necessarily a fall from heaven.

4 The Yoga conception

Patañjali, founder of the Yoga system, postulates God as a soul different from individual souls in that he does not have any blemishes and is eternally free. God is omniscient and omnipresent and is the creator of the world, which arises from ultimate matter (*prakṛti*) because of its association with individual souls due to nescience. All souls, individual as well as divine, are of the nature of consciousness not really affected by pain and suffering. The ultimate aim of life is not to realize God, but to realize the nature of one's own soul, which is wrongly believed to be associated with the body. But God-realization may help some individuals to attain self-realization. It is not compulsory to believe in God to attain the *summum bonum* of human life.

Patañjali did not feel the necessity of God for solving any theoretical problem of philosophy. For him, God has more a practical than a theoretical value. Devotion to God is considered to be of great practical value, inasmuch as it forms part of the practice of yoga and is one of the many means for the final attainment of concentration leading to yoga, defined as 'restraint of the mind'. The ultimate aim of human life is self-realization; God-realization is an optional means.

According to the Yoga system, God is the Supreme Person who is above all individual selves and free from all defects. He is the Perfect Being, eternal, all-pervading and omniscient. All individual selves are more or less subject to the afflictions (*kleśa*) of ignorance, egoism, desire, aversion and dread of death. They have to do various kinds of works (*karma*) - good, bad and indifferent - and reap the consequences (*vipāka*) thereof. They are also infected and influenced by the latent impressions of their past experiences (*āśaya*). Even if the liberated self is released from all these troubles, it cannot be said that it was always free from them. It is God and God alone who is eternally free from all defects. God is a perfect immortal spirit who remains ever untouched by afflictions and actions, and their effects and impressions. He possesses a perfect nature and also the fullest possible knowledge of all acts; he is, therefore, capable of maintaining the whole world by his mere wish or thought. He is the Supreme Ruler of the world and has infinite knowledge, unlimited power and the wisest desires, which distinguish him from all other selves.

5 The Nyāya conception

According to Nyāya, both old and new, God is a self-substance which has special qualities that distinguish God from individual selves. God as Supreme Self (*paramātman*) has eight qualities: number, magnitude, separateness, conjunction, disjunction, consciousness or knowledge, desire or will, and inner effort. God does not have pleasure, pain, aversion, memory traces, merit or demerit. This conception of God is fundamentally different from the Advaita Vedānta theory of Brahman, which identifies Brahman with existence, consciousness and bliss. The Nyāya conception of God, on the other hand, is of a substance which has existence, consciousness, will and inner effort, but not bliss, as special qualities. Moreover, God is one real object among many. There are several qualities that make God the Supreme Self. So far as omnipresence is concerned, God and individual selves have this quality in the same sense; both God and individual selves (or souls) are ubiquitous substances. It does not make sense to say that God is in the world or in humanity, or immanent in any sense. This is because ubiquitous substance cannot be in anything. On the other hand, all substances of finite magnitude are in contact with ubiquitous substances;

they are in these substances, but these substances are not in anything.

The difference between individual selves and God consists primarily in his not having any body. God is the Supreme Self without any body. The supremacy of the Supreme Self consists in his omniscience and omnipotence. But these qualities are conceived very differently in new Nyāya (Navya-Nyāya) than in other Indian philosophical systems. God's knowledge of all things and events, past, present and future, is one cognition and is eternal. Although God knows all objects, he does not know them in different cognitive states. He has only one cognition, which is eternally there. Moreover, this eternal knowledge can only be perceptual; it cannot be inferential, for any mediate cognition has to be produced by another cognition. But as God does not have a plurality of cognitions, and as his knowledge is eternal, his knowledge cannot be produced at all; for whatever is produced cannot be eternal.

The omnipotence of God is also conceived very differently in Navya-Nyāya than in other Indian philosophical systems. God creates everything through his will guided by his knowledge. Again, the will of God is one act of will and is eternal: God does not have to perform more than one act of will to create the world. Furthermore, God does not create everything that there is. According to Navya-Nyāya, there are eternal objects of various categories, such as the self, atoms, space, time, the mind (*manas*) and ether (*ākāśa*), all of which are substances. Also eternal are universals, the relation of inherence (*samavāya*) and the individuality (*viśeṣa*) that belongs to each eternal substance (see Nyāya-Vaiśeṣika §§4-5). These objects are all coeternal with God and cannot be created by him. Thus God's omnipresence and omnipotence are very much restricted in that he cannot be present in any ubiquitous substance such as space or time, nor in our souls. However, he does have unrestricted, eternal perception of all things past, present and future, as also of all eternal substances and objects.

See also: Gauḍīya Vaiṣṇavism; God, concepts of; Heaven, Indian conceptions of

SIBAJIBAN BHATTACHARYYA

References and further reading

Bhattacharyya, S. and Potter, K.H. (eds) (1993) *Encyclopedia of Indian Philosophies*, vol. 6, *Indian Philosophical Analysis: Nyāya-Vaiśeṣika from Gaōgeśa to Raghunātha Śiromaṇi*, Princeton, NJ: Princeton University Press.(See especially page 211 of Part Two, chapter one (on Gaōgeśa) for proofs of the existence of God in Nyāya.)

Chatterjee, S.C. and Datta, D.M. (1984) *An Introdution to Indian Philosophy*, Calcutta: University of Calcutta. (Includes an explanation of Müller's conception of henotheism.)

Frauwallner, E. (1973) *History of Indian Philosophy*, trans. V.M. Bedekar, Delhi: Motilal Banarsidass, 2 vols. (Contains original researches into many Indian systems.)

Radhakrishan, S. (1923) *Indian Philosophy*, Delhi: Oxford University Press, 1989, 2 vols.(Contains detailed information about all the Indian systems.)

Ṛg Veda (*c.*1200 BC), trans. H.H. Wilson, *Ṛg-Veda Samhita*, Delhi: Nag Publishers, 1990, 7 vols.(An English translation of the entire *Ṛg Veda*. The Indian tradition is that the Vedas are beginningless; Wilson is of the opinion that they are of human composition.)

Gödel, Kurt (1906-78)

The greatest logician of the twentieth century, Gödel is renowned for his advocacy of mathematical Platonism and for three fundamental theorems in logic: the completeness of first-order logic; the incompleteness of formalized arithmetic; and the consistency of the axiom of choice and the continuum hypothesis with the axioms of Zermelo-Fraenkel set theory.

A Sudeten German born in Brno, Moravia, Gödel received his doctorate in 1930 at the University of Vienna. His dissertation, extending earlier work by Skolem, demonstrated that first-order predicate calculus is semantically complete, that is, that every logical truth is provable. His incompleteness paper (1931, subsequently his *Habilitationsschrift*) showed that every ω-consistent recursively axiomatizable formalization of Peano arithmetic yields statements that are formally undecidable (neither provable nor refutable) within the theory. Rosser later showed that 'ω-consistent' could be weakened to 'consistent'. One of the undecidable sentences expresses the theory's own consistency. Hilbert's goal of reducing the consistency of arithmetic to that of weaker systems (see Hilbert's programme and formalism) thus cannot be realized.

While *Privatdozent* at the University of Vienna (1933-9) Gödel published several papers bearing on the decision problem (see Proof theory) and the relation between classical and intuitionistic logic (see Intuitionism). He also attended some meetings of the Vienna Circle, but disagreed with its philosophy (see Logical positivism).

Gödel spent the academic year 1933-4 at the Institute for Advanced Study in Princeton. Then, after recovering from an incapacitating bout of depression, he turned to set theory. In his address to the International Congress of Mathematicians in 1900, Hilbert had asked for a proof that every infinite set of real numbers is equinumerous either with the integers or the set of all reals (Cantor's continuum hypothesis). Gödel showed in 1938 that that assumption cannot be disproved within Zermelo-Fraenkel set theory, since it and the axiom of choice hold within a particular model thereof (see Set theory).

In 1940 Gödel emigrated to the United States. He returned to the Princeton Institute and remained there until his death from malnutrition - the result of further psychiatric disturbance - in 1978.

After 1942, frustrated by his failure to prove that the negation of the continuum hypothesis is also consistent with the axioms of set theory (a result finally established by Paul Cohen in 1963 - see Cohen 1966), Gödel worked primarily in philosophy and cosmology. Among the philosophical works published during his lifetime his commentary on 'Russell's Mathematical Logic' (1944) and his essay 'What is Cantor's Continuum Problem?' (1947, revised 1964) are notable for their mathematical Platonism. In the former, Gödel maintained that 'classes and concepts may…be conceived as real objects' whose existence is as tenable as that of physical bodies; in the latter that 'despite their remoteness from sense experience, we…have something like a perception also of the objects of set theory', whose 'axioms force themselves upon us as being true'. Posthumous philosophical works include a modal formalization of Leibniz's ontological proof.

See also: Cantor's theorem; Church's theorem and the decision problem; Computability theory; Continuum hypothesis; Gödel's theorems; Logical and mathematical terms, glossary of; Provability logic; Realism in the philosophy of mathematics; Time travel; Turing machines

JOHN W. DAWSON, JR

List of works

Gödel, K. (1986-95) *Kurt Gödel: Collected Works*, vol. 1, *Publications 1929-1936*, 1986; vol. 2, *Publications 1938-1974*, 1990; vol. 3, *Unpublished Essays and Lectures*, 1995, ed. S. Feferman *et al*., New York and Oxford: Oxford University Press.(Contains full texts of all Gödel's published works as well as some previously unpublished work, with commentaries and an extensive bibliography. Includes parallel English translations of all German items.)

(1930) 'Die Vollständigkeit der Axiome des logischen Funktionenkalküls', *Monatshefte für Mathematik und Physik* 37: 349-60; trans. 'The Completeness of the Axioms of the Functional Calculus of Logic', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 582-91.(Revised version of Gödel's doctoral dissertation.)

(1931) 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38: 173-98; trans. 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 592-617.(The incompleteness paper.)

(1940) *The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis with the Axioms of Set Theory*, Annals of Mathematics Studies, vol. 3, Princeton, NJ: Princeton University Press.(A monograph including full details of Gödel's consistency proofs.)

(1941) 'In What Sense is Intuitionistic Logic Constructive?', in *Collected Works*, vol. 3, 189- 201.(A particularly lucid discussion of the ideas underlying Gödel's consistency proof for arithmetic via functionals of finite type.)

(1944) 'Russell's Mathematical Logic', in P.A. Schilpp (ed.) *The Philosophy of Bertrand Russell*, Library of Living Philosophers, vol. 5, Evanston, IL: Northwestern University Press.(An intricate critique of Russell's contributions to logic.)

(1946-9) 'Some Observations About the Relationship Between Theory of Relativity and Kantian Philosophy' (two versions), in *Collected Works*, vol. 3, 230-60.(The accompanying introductory note, by Howard Stein, provides penetrating historical and technical commentary.)

(1947) 'What is Cantor's Continuum Problem?', *American Mathematical Monthly* 54: 515-25; errata, 55: 151.(A discussion of the problem of determining the cardinality of the continuum, written for a general mathematical audience.)

(1953-9) 'Is Mathematics Syntax of Language?' (two versions), in *Collected Works*, vol. 3, 334-63.(An essay opposing views of Rudolf Carnap.)

(1970) 'Ontological Proof', in *Collected Works*, vol. 3, 403-4.(Preceded by an illuminating note by Robert M. Adams.)

References and further reading

Boolos, G. (1979) *The Unprovability of Consistency: An Essay in Modal Logic*, Cambridge: Cambridge University Press.(The work of Gödel and Löb, viewed from a modal perspective.)

Boolos, G. (1993) *The Logic of Provability*, Cambridge: Cambridge University Press.(Revised and expanded sequel to Boolos (1979).)

Cohen, P. (1966) *Set Theory and the Continuum Hypothesis*, New York: Benjamin.(A self-contained exposition of Cohen's independence proofs.)

Dawson, J. (1997) *Logical Dilemmas: The Life and Work of Kurt Gödel*, Wellesley, MA: Peters.(A full-length scientific biography.)

Nagel, E. and Newman, J.R. (1958) *Gödel's Proof*, New York: New York University Press; repr. London: Routledge, 1971.(Still a classic exposition of Gödel's incompleteness results for the lay reader.)

Shanker, S. (ed.) (1988) *Gödel's Theorem in Focus*, London: Croom Helm.(An anthology of essays about Gödel's life and work, together with a translation of his incompleteness paper.)

Yourgrau, P. (1991) *The Disappearance of Time: Kurt Gödel and the Idealistic Tradition in Philosophy*, Cambridge: Cambridge University Press.(A study of the implications of Gödel's cosmological work for the philosophy of time.)

Gödel's theorems

Utilizing the formalization of mathematics and logic found in Whitehead and Russell's Principia Mathematica (1910), Hilbert and Ackermann (1928) gave precise formulations of a variety of foundational and methodological problems, among them the so-called 'completeness problem' for formal axiomatic theories - the problem of whether all truths or laws pertaining to their subjects are provable within them. Applied to a proposed system for first-order quantificational logic, the completeness problem is the problem of whether all logically valid formulas are provable in it.

In his doctoral dissertation (1929), Gödel gave a positive solution to the completeness problem for a system of quantificational logic based on the work of Whitehead and Russell. This is the first of the three theorems that we here refer to as 'Gödel's theorems'.

The other two theorems arose from Gödel's continued investigation of the completeness problem for more comprehensive formal systems - including, especially, systems comprehensive enough to encompass all known methods of mathematical proof. Here, however, the question was not whether all logically valid formulas are provable (they are), but whether all formulas true in the intended interpretations of the systems are.

For this to be the case, the systems would have to prove either S or the denial of S for each sentence S of their languages. In his first incompleteness theorem, Gödel showed that the systems investigated were not complete in this sense. Indeed, there are even sentences of a simple arithmetic type that the systems can neither prove nor refute, provided they are consistent. So even the class of simple arithmetic truths is not formally axiomatizable.

The idea behind Gödel's proof is basically as follows. Let a given system T satisfy the following conditions: (1) it is powerful enough to prove of each sentence in its language that if it proves it, then it proves that it proves it, and (2) it is capable of proving of a certain sentence G (Gödel's self-referential sentence) that it is equivalent to 'G is not provable in T'. Under these conditions, T cannot prove G, so long as T is consistent. For suppose T proved G. By (1) it would also prove 'G is provable in T', and by (2) it would prove 'G is not provable in T'. Hence, T would be inconsistent.

Under slightly stronger conditions - specifically, (2) and (1′) every sentence of the form 'X is provable in T' that T proves is true - it can be shown that a consistent T cannot prove 'not G' either. For if 'not G' were provable in T it would follow by (2) that 'G is provable in T' would also be provable in T. But then by (1′) G would be provable. Hence, T would be inconsistent.

The proof of Gödel's second incompleteness theorem essentially involves formalizing in T a proof of a formula expressing the proposition that if T is consistent, then G. The second incompleteness theorem (that is, the claim that if T is consistent it cannot prove its own consistency) then follows from this and the first part of the proof of the first incompleteness theorem.

The two incompleteness theorems have been applied to a wide variety of concerns in philosophy. The best known of these are critical applications to Hilbert's programme and logicism in the philosophy of mathematics and to mechanism in the philosophy of mind.

1 Background

Interest in questions of completeness dates from quite early on in the history of logic and the foundations of mathematics. Aristotle (*Posterior Analytics* 87b) described the goal of science as the complete identification of the laws pertaining to its various areas. Much later, Kant expressed concern with completeness when he described human reason as inevitably disposed to pose questions to itself that it cannot answer. Clearer still, perhaps, are various statements in letters by Gauss (1817, 1829), in which he asserted that it may in principle be impossible for us to give a complete specification of the laws of geometry a priori owing to the fact that its object, space, is at least partially 'external' to our minds. (See Kronecker ([1887] 1932: 265) for an expression of the same view.) Finally, we find something approaching modern formulations of the notion in Hilbert ([1900] 1930: 242), where it is said to be a duty to show that a set of axioms for geometry capture every geometric law, and in Huntington (1905: 17), where a set of postulates for real algebra is said to form a 'complete' axiomatic basis for the subject.

However, it was not until the appearance of Hilbert and Ackermann's *Grundzüge der theoretischen Logik* (1928: 42-3) that the completeness problem received its first fully explicit, precise and general formulation. Hilbert and Ackermann believed that the formal-symbolic approach of modern mathematics had contributed greatly to its remarkable development, and they therefore advocated taking the same approach to logic. Such an approach would, they believed, pave the way for a mathematically precise study of both logical and mathematical thinking, and would make it possible to settle in a mathematical way various of the methodological and foundational questions pertaining to them. Among these was the question of 'completeness' - that is, whether the full class of truths or laws properly pertaining to a given domain of thought can be proved in a given system purporting to formalize it. Whitehead and Russell (1910) presented a formalism (PM) for logic and mathematics, and this became the focus of the early investigations of the completeness problem.

The first results were obtained (independently) by Bernays, in his unpublished *Habilitationsschrift* of 1918 (see Bernays 1926), and Post, in his doctoral dissertation of 1920 (see Post 1921). They isolated the propositional part of Whitehead and Russell's PM and showed that it proved all valid sentences of the propositional variety. Post indeed proved the stronger result (known as 'Post-completeness') that no sentence of the propositional variety can be added to PM without creating an inconsistency.

Hilbert and Ackermann observed that the broader logical system known as the 'restricted', or 'first-order' predicate calculus is not Post-complete (1928: 92). With this as his point of departure, Gödel, in his doctoral dissertation (1929), then took up the ordinary completeness problem for this system. Using ideas introduced by Löwenheim (1915) and Skolem (1923), he obtained the basic positive result that every valid formula of the first-order quantificational variety is provable in PM. Indeed, he proved the somewhat stronger result that every first-order quantificational formula is either refutable in PM or is satisfiable in a denumerable domain. He then extended the basic result to first-order logic with identity and the stronger result from single formulas to denumerable systems of formulas. In this latter extension, he made use of a 'compactness' result (namely, that a denumerably infinite set of formulas is satisfiable if and only if each of its finite subsets is) that he proved by semantic means.

In the introduction to his dissertation (deleted from the version published in 1930), Gödel expressed awareness of the possibility that the completeness problem for other areas of mathematical thought might not result in the same positive type of result as the completeness problem for PM. It was therefore not surprising that in 1930 he turned his attention to the completeness problem for the architectonic systems of Whitehead and Russell and Zermelo and Fraenkel (see Gödel 1930, 1931). These systems interested Gödel precisely because of their comprehensiveness. Such comprehensiveness might, he noted, lead some to conjecture that they can decide all mathematical questions formulable in their languages. What Gödel showed was that this is mistaken. There are problems - indeed, relatively elementary problems of the arithmetic of the integers - that even the most comprehensive systems cannot decide. Gödel's investigation of the completeness problem thus ended with a negative result.

There is a relationship between Gödel's 1929/1930 and 1931 studies that seems worth noting, even though he himself did not note it. It stems from the fact that every formal system is composed of both a logical and a non-logical part, deficiencies in either of which can lead to incompleteness. A defect of the first sort would mean that though the non-logical part might itself be perfect in content, the logical part would not be able fully to extract that content.

Gödel ([1929] 1930) can be seen as showing that incompleteness in systems whose logic is rightly taken to be first-order is due to deficiencies in the non-logical part, and his 1931 paper can then be seen as showing that such deficiencies extend to a wide range of systems. The latter also points to *logical* deficiencies in another class of systems, however - namely, systems whose non-logical parts are composed of the second-order Peano postulates and whose logical part is some formalizable scheme of second-order logical principles. This is a result of the categoricity of the second-order Peano postulates (first established in Dedekind 1888). Being categorical, they are in a certain sense perfect in their content. Still, Gödel (1931) establishes that any formal system based on them must be incomplete. Hence, any formalizable scheme of second-order logical consequence must fail to capture some of the content latent in them. In other words, any formal system of second-order logic must fail to capture some valid second-order logical principle. The 1931 paper can thus be seen as rounding out the logical investigations of the 1930 paper by providing a negative solution to the completeness problem for second-order

logic.

2 The completeness theorem

Gödel's original completeness proof was based on a normalization of quantificational formulas developed by Löwenheim (1915) and Skolem (1923). According to this normalization, there is an effective way of finding, for each formula ϕ of a first-order quantificational language L, a formula $\sigma(\phi)$ such that ϕ is satisfiable iff $\sigma(\phi)$ is, and $\sigma(\phi)$ is of the form $U–E–QF$, where *U* is a (non-empty) string of universal quantifiers, *E* a (possibly empty) string of existential quantifiers and *QF* a quantifier-free formula. Gödel showed how to construct conjunctions $K_n = QF_1 \,\&\ldots\&\, QF_n$ of instances QF_i of *QF* such that either for every *n*, K_n is truth-functionally satisfiable, or, for some *n*, K_n is not. Like Skolem, he then showed that if the first alternative holds, $\sigma(\phi)$ is satisfiable in the domain of the natural numbers. Finally, he was able to show - and this was his distinctive contribution - that if the second alternative holds, $\sigma(\phi)$ is refutable (that is, $\vdash_{\mathrm{PM}} \neg\sigma(\phi)$). With this came the crucial link between satisfiability and derivability in PM needed to prove PM's completeness.

It follows that $\sigma(\phi)$ is either denumerably satisfiable or refutable, and from this and the fact (also used by Gödel) that $\sigma(\phi)$ is refutable only if ϕ is, it then follows that for any formula ϕ of the language of the pure predicate calculus either ϕ is (denumerably) satisfiable or ϕ is refutable.

Gödel's proof was given for a specific logical calculus having a countable language (that is, a language containing countably many individual constants). Since then, completeness proofs have been devised both for a wider variety of calculi having countable languages (see, for example, Henkin 1949; Hasenjäger 1953; Hintikka 1955) as well as for calculi with uncountable languages (see Maltsev 1936). From this work, the following general method for constructing models for consistent sets of sentences of infinite cardinality has emerged.

To construct a model of infinite cardinality α for a consistent set S of sentences of language L of cardinality not exceeding α, one begins by adding α-many constants to L to form the language L^*. One then extends S to a maximal consistent and term-complete set S^*. (A consistent set is 'maximal consistent' iff no set properly containing it is consistent. A set is 'term-complete' just in case for every formula ϕ and all terms $t_1, \ldots, t_i, \ldots, t_\alpha$ of L^*, if $\phi(t_1), \ldots, \phi(t_i), \ldots, \phi(t_\alpha)$ are all elements of the set, then so is $\forall x\phi(x)$.) Next one constructs a model for S^* whose domain is the class of constants of L^*. And finally, one restricts this model to L and automatically obtains a model for S.

The proof that this construction yields a model *M* proceeds roughly as follows. (1) For every closed term *c* that appears in S^*, let *M* assign *c* itself as value. In this way, every element of the domain of *M* becomes the referent of a term of L^*. (2) For each basic *n*-ary predicate P^n of L^*, let *M* assign the *n*-tuple $\langle t_1, \ldots, t_n\rangle$ to the extension of P^n iff $P^n(t_1, \ldots, t_n)$ is an element of S^*. In this way, all atomic sentences and all denials of atomic sentences that are elements of S^* become true in *M*. This is so because S^*, being consistent, does not contain any atomic sentence that is denied by another element of S^*.

This completes the basis clause. In the induction step one calls upon the fact that, owing to its maximal consistency (and, in the case of quantified sentences, its term-completeness), S^* contains, for each of its compound elements σ_c (of complexity greater than that of the denial of an atomic sentence), a set of less complex elements whose joint truth suffices for the truth of σ_c. These less complex sentences fall under the hypothesis of the induction. The truth of the sentences of the basis clause thus 'induces' the truth of all elements of S^* of the next higher degree of complexity, and this process continues through all degrees of complexity. The truth-in-*M* of the sentences covered in the basis step thus eventually 'oozes up' to each element of S^*.

Every consistent set of sentences thus has a model. From this, both 'weak' completeness (every valid sentence is derivable) and 'strong' completeness (every sentence validly implied by a set of sentences is derivable from it) follow for systems admitting the usual types of derivations.

3 The first incompleteness theorem

Gödel made use of a formalization of mathematics in his 1931 paper that was related to that given in Whitehead and Russell (1910). But he added a new device - the 'arithmetization' of formal systems. This was based upon a division of the objects of a formal language into three basic categories: (1) the basic symbols of the language; (2)

finite sequences of these symbols; and (3) finite sequences of finite sequences of them. In order to express properties and propositions concerning these objects as predicates and sentences of arithmetic, Gödel defined a correlation (now called a 'Gödel-numbering') between the objects of these three types and the natural numbers. He coded the individual symbols of the systems by distinct odd numbers. To each finite sequence $\sigma_1, \ldots, \sigma_n$ of symbols he then assigned the number

$$2^{g(\sigma_1)} \cdot 3^{g(\sigma_2)} \cdot \ldots \cdot p_n^{g(\sigma_n)},$$

where p_n is the *n*th prime number and $g(\sigma_i)$ $(1 \leq i \leq n)$ the Gödel number of σ_i. Finally, to each finite sequence of finite sequences of symbols $\Sigma_1, \ldots, \Sigma_k$, he assigned the number

$$2^{g(\Sigma_1)} \cdot 3^{g(\Sigma_2)} \cdot \ldots \cdot p_k^{g(\Sigma_k)},$$

where p_k is the *k*th prime number and $g(\Sigma_i)$ $(1 \leq i \leq k)$ the Gödel number of Σ_i.

Infinitely many others schemes would be possible. The essential features are that distinct expressions are assigned distinct numbers and that there are effective procedures for encoding expressions and for decoding numbers to find which expression (if any) they represent.

Under such an encoding, certain formulas express not only arithmetic statements about numbers but also metamathematical statements about such objects as formulas, axioms and proofs. Using this observation, Gödel exploited an analogy with the Richard and liar paradoxes (see Paradoxes of set and property §6). (In passing he also observed that any other 'epistemological antinomy' could be put to similar use. See Gödel (1931: 598).) The analogy runs as follows.

Suppose that a theory *T* proves only true sentences and that a sentence *G* 'says' of itself that it is not provable in *T*. Neither *G* nor $\neg G$ is then provable in *T*. For if *G* is true then it is not provable in *T*, and if *G* is false then it cannot be provable because *T* proves only true sentences. So *G* is not provable in *T* and hence true. Its denial, $\neg G$, is thus false and hence unprovable in *T* since *T* proves only true sentences.

Gödel's proof follows this reasoning with one exception that he himself emphasized (1931: 599): it replaces the assumption that *T* proves only true sentences with the weaker assumption of ω-consistency (see below). The core of the proof is the construction of the 'self-referential' sentence *G* in the so-called 'diagonal lemma'. In its general form, this lemma states that for every formula F of L_T (that is, the language of *T*) of one free variable, there is a sentence D of L_T such that $\vdash_T D \leftrightarrow F(\ulcorner D \urcorner)$ (the term '$\ulcorner D \urcorner$' stands for the numeral of L_T whose referent under the standard interpretation of L_T is the Gödel number of *D*). We need only the following specialization of the diagonal lemma to *G*:

(DL) For every theory *T* that contains a certain fragment of arithmetic, there is a sentence *G* of L_T such that $\vdash_T G \leftrightarrow \neg\,\mathrm{Prov}_T(\ulcorner G \urcorner)$.

This lemma is also referred to as Gödel's 'fixed point theorem', *A* being a 'fixed point' in *T* of a formula $\phi(x)$ just in case $\vdash_T A \leftrightarrow \phi(\ulcorner A \urcorner)$. The formula $\mathrm{Prov}_T(x)$ 'expresses' the notion of a formula's being provable in *T* in the limited sense that the set of *A* such that $\vdash_T \mathrm{Prov}_T(\ulcorner A \urcorner)$ is the set of theorems of *T*. In the standard terminology, $\mathrm{Prov}_T(x)$ 'weakly represents' the set of Gödel numbers of theorems of *T* in *T*; that is

(WR) For every sentence *A* of L_T, $\vdash_T A$ iff $\vdash_T \mathrm{Prov}_T(\ulcorner A \urcorner)$.

From DL and WR, the proof of Gödel's first incompleteness theorem (G1) proceeds in two parts.

Part 1. If *T* is a consistent theory and $\mathrm{Prov}_T(x)$ satisfies both WR in the 'only if' direction and DL, then $\nvdash_T G$.

Proof. Suppose that

(1) $\vdash_T G$.

From this and WR in the 'only if' direction, we get

(2) $\vdash_T \mathrm{Prov}_T(\ulcorner G \urcorner)$.

But by (1), DL and the assumption that the logic of *T* allows *modus ponens* inference, we get

$\vdash_T \neg\mathrm{Prov}_T(\ulcorner G \urcorner)$,

which together with (2) implies the inconsistency of *T*.

Part 2. If *T* is consistent and $\mathrm{Prov}_T(x)$ satisfies both WR in the 'if' direction and DL, then $\nvdash_T \neg G$.

Proof. Suppose that

(3) $\vdash_T \neg G$.

From (3), DL and the fact that the logic of *T* allows both *modus tollens* inference and inference from $\neg\neg A$ to *A*, it follows that

$\vdash_T \mathrm{Prov}_T(\ulcorner G \urcorner)$.

But this and WR in the 'if' direction imply that

$\vdash_T G$,

which together with (3) implies the inconsistency of *T*.

Together, parts 1 and 2 establish

(G1) If *T* is consistent and $\mathrm{Prov}_T(x)$ satisfies DL and WR, then $\nvdash_T G$ and $\nvdash_T \neg G$.

It should be noted that WR in the 'if' direction is a different condition from that of the consistency of *T*. It forces all theorems of *T* of the form $\mathrm{Prov}_T(\ulcorner A \urcorner)$ to be true. Equivalently, given the decidability in *T* of the arithmetized proof relation of *T*, it forces *T* to be ω-consistent (that is, it forces *T* not to prove, for any formula ϕ, both all instances of $\phi(n)$ and also $\exists x \neg\phi(x)$). ω-consistency, which was explicitly used as a condition in Gödel's original proof, thus figures in our proof as well. Using a more complicated fixed point formula than *G* and a more complicated proof than the one just given, Rosser (1936) proved a version of G1 that does not require ω-consistency.

Parts 1 and 2 reveal the basic structure of the proof for G1. The details that remain are proofs of DL and WR. These can be found in many standard logic texts (for example, Enderton 1972; Boolos and Jeffrey 1974). Since Gödel's proof in 1931, a mathematically more interesting case of a purely combinatorial sentence undecidable in first-order Peano arithmetic (PA) has been given, in Paris and Harrington (1977). In the early 1980s, Friedman gave a still more striking case (involving Kruskal's theorem) of an arithmetical sentence undecidable in fragments of second-order Peano arithmetic considerably stronger than PA.

4 The second incompleteness theorem

Gödel arrived at the idea for his second incompleteness theorem (G2) by noting that, using the provability expression $\mathrm{Prov}_T(x)$ of part 1 of the proof of G1, one can define a consistency formula Con_T for *T* and that the proof of part 1 can then be carried out in *T*. In other words, one can prove '$\mathrm{Con}_T \rightarrow G$' in *T*. Together with part 1, this then implies the unprovability in *T* of Con_T.

The proof of '$\mathrm{Con}_T \rightarrow G$', however, requires conditions on $\mathrm{Prov}_T(x)$ that are different from those (namely DL and WR) required for the proof of G1. What exactly these conditions are becomes clear from the proof of '$\mathrm{Con}_T \rightarrow G$' given below.

The details of this proof require that we choose a specific definition of the consistency expression, Con_T. The definition we choose is the basic one, the formula '$\forall x(\mathrm{Prov}_T(x) \rightarrow \neg\mathrm{Prov}_T(\mathrm{neg}(x)))$' (where $\mathrm{neg}(x)$ stands for the function that sends the Gödel number of a formula to the Gödel number of its negation). We could define Con_T differently without altering the conditions on $\mathrm{Prov}_T(x)$ that are required for the proof of '$\mathrm{Con}_T \rightarrow G$' in *T*. The present definition, however, has the virtues of being both familiar and basic. We therefore proceed with the proof of

$\vdash_T \text{Con}_T \rightarrow G.$(G2 lemma)

Its core element is the construction of a derivation in *T* of $\neg\text{Con}_T$ from $\neg G$. We begin by stating five schemata whose provability in *T* will be assumed in the proof:

(a) $\text{Prov}_T(\ulcorner A \urcorner) \rightarrow \text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner A \urcorner) \urcorner)$
(b) $\text{Prov}_T(\ulcorner A \leftrightarrow B \urcorner) \rightarrow \text{Prov}_T(\ulcorner \neg B \leftrightarrow \neg A \urcorner)$
(c) $\text{Prov}_T(\ulcorner A \leftrightarrow B \urcorner) \rightarrow \text{Prov}_T(\ulcorner A \rightarrow B \urcorner)$
(d) $\text{Prov}_T(\ulcorner A \rightarrow B \urcorner) \rightarrow (\text{Prov}_T(\ulcorner A \urcorner) \rightarrow \text{Prov}_T(\ulcorner B \urcorner))$
(e) $\text{Prov}_T(\ulcorner A \urcorner) \rightarrow \text{Prov}_T(\ulcorner \neg\neg A \urcorner)$.

The proof proper begins by taking $\neg G$ as a hypothesis in *T*. We then reason in *T* as follows. Applying DL (whose content '$G \leftrightarrow \neg\text{Prov}_T(\ulcorner G \urcorner)$' is available in *T*), we get

(4) $\text{Prov}_T(\ulcorner G \urcorner)$.

From (4) and an instance of (a), we get

(5) $\text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner G \urcorner) \urcorner)$.

We now add a *T*-theoretic 'formalization' of DL; that is,

$\text{Prov}_T(\ulcorner G \leftrightarrow \neg\text{Prov}_T(\ulcorner G \urcorner) \urcorner)$.

This together with an instance of schema (b) gives

$\text{Prov}_T(\ulcorner \neg\neg\text{Prov}_T(\ulcorner G \urcorner) \leftrightarrow \neg G \urcorner)$,

which together with an instance of (c) then yields

$\text{Prov}_T(\ulcorner \neg\neg\text{Prov}_T(\ulcorner G \urcorner) \rightarrow \neg G \urcorner)$.

From this and an instance of (d) we get

$\text{Prov}_T(\ulcorner \neg\neg\text{Prov}_T(\ulcorner G \urcorner) \urcorner) \rightarrow \text{Prov}_T(\ulcorner \neg G \urcorner)$,

which with an instance of (e) gives us

$\text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner G \urcorner) \urcorner) \rightarrow \text{Prov}_T(\ulcorner \neg G \urcorner)$.

This and (5) via *modus ponens* in *T* imply

(6) $\text{Prov}_T(\ulcorner \neg G \urcorner)$.

(6) together with (4) then gives us

$\neg(\text{Prov}_T(\ulcorner G \urcorner) \rightarrow \neg\text{Prov}_T(\ulcorner \neg G \urcorner))$.

This and the definition of Con_T imply $\neg\text{Con}_T$. This is what was to be proved in *T* from our hypothesis $\neg G$. Discharging this hypothesis, we then get

$\neg G \rightarrow \neg\text{Con}_T$,

and from this, by contraposition in *T*, we conclude

$\text{Con}_T \rightarrow G$.

Inspecting this proof we see that it makes various common assumptions concerning the logic of *T*; for example, that it allows *modus ponens* inference, classical contraposition and universal instantiation. It also utilizes various conditions on $\text{Prov}_T(x)$ - specifically, schemata (a)-(e) and the 'formalized' version of DL. All of these, we will now show, can be obtained from the following 'derivability conditions'.

(DC1) For every sentence *A* of L_T, $\vdash_T A$ only if $\vdash_T \mathrm{Prov}_T(\ulcorner A \urcorner)$.
(DC2) For all formulas *A* and *B* of L_T,
$\vdash_T \mathrm{Prov}_T(\ulcorner A \to B \urcorner) \to (\mathrm{Prov}_T(\ulcorner A \urcorner) \to \mathrm{Prov}_T(\ulcorner B \urcorner))$.
(DC3) For every sentence *A* of L_T, $\vdash_T \mathrm{Prov}_T(\ulcorner A \urcorner) \to \mathrm{Prov}_T(\ulcorner \mathrm{Prov}_T(\ulcorner A \urcorner) \urcorner)$.

The sufficiency of DC1-3 can be seen as follows. Schema (a) is simply DC3 and schema (d) is DC2. The 'formalized' version of DL used above is obtained immediately by applying DC1 to DL. Schema (b) is obtained by applying DC1 to the logical truths of the form $(A \leftrightarrow B) \to (\neg B \leftrightarrow \neg A)$ to get the several instances of $\mathrm{Prov}_T(\ulcorner (A \leftrightarrow B) \to (\neg B \leftrightarrow \neg A) \urcorner)$, and then applying DC2. The remaining schemata are similarly obtained. Schema (c) is obtained by applying DC1 to the logical principle $(A \leftrightarrow B) \to (A \to B)$ to get $\mathrm{Prov}_T(\ulcorner (A \leftrightarrow B) \to (A \to B) \urcorner)$, which then, by DC2, yields (c). Schema (e) is obtained by application of DC1 to $A \to \neg\neg A$ to get $\mathrm{Prov}_T(\ulcorner A \to \neg\neg A \urcorner)$, and then application of DC2.

We thus arrive at the following generalized version of Gödel's second incompleteness theorem.

(G2) If *T* is consistent, Con_T is defined as $\forall x(\mathrm{Prov}_T(x) \to \neg \mathrm{Prov}_T(\mathrm{neg}(x)))$ and $\mathrm{Prov}_T(x)$ satisfies DC1-3, then $\nvdash_T \mathrm{Con}_T$.

5 The derivability conditions

DC1-3 represent a streamlining of the original 'Ableitbarkeitsforderungen' used in Hilbert and Bernays (1939: 285-6) to give the first full proof of G2. The improvements are due mainly to Löb (1955), though his aim was to identify a set of conditions on $\mathrm{Prov}_T(x)$ sufficient to solve a problem of Leon Henkin's and not to provide a general set of conditions under which G2 holds.

Jeroslow (1973) produced a further streamlining of the derivability conditions by showing how to eliminate DC2. The key idea was a new fixed-point or diagonalization construction - referred to as a 'literal' fixed point. It states that

For every formula $\psi(x)$ of L_T of one free variable, one can effectively find a formula ϕ and a closed term τ of L_T such that ϕ is $\psi(\tau)$, and $\vdash_T \tau = \ulcorner \phi \urcorner$.

What Jeroslow proved was a version of G2 lemma with a literal Gödel sentence G_{Lit} - that is, a sentence of the form $\neg \mathrm{Prov}_T(\tau)$, where $\vdash_T \tau = \ulcorner G_{\mathrm{Lit}} \urcorner$, and not merely a sentence provably equivalent in *T* to the sentence $\neg \mathrm{Prov}_T(\ulcorner G_{\mathrm{Lit}} \urcorner)$ - in place of the usual Gödel sentence *G*. His proof relies only upon the literal diagonal lemma, DC1 and DC3. Likewise, the relevant part of Jeroslow's version of G1 - namely, that if *T* is consistent, then $\nvdash_T G_{\mathrm{Lit}}$ - relies only upon the literal diagonal lemma and DC1. Together these two results yield a version of G2 that requires only the literal diagonal lemma, DC1 and DC3. We thus see that G2 can hold even when the formula expressing provability does not satisfy DC2. Among other things, this appears to distinguish G2 from Löb's theorem (that is, the theorem that for any sentence *A* of L_T and any formula $\mathrm{Prov}_T(x)$ satisfying DC1-3, if $\vdash_T \mathrm{Prov}_T(\ulcorner A \urcorner) \to A$, then $\vdash_T A$), since the known proofs of the latter demand satisfaction of DC2 (see Kreisel and Takeuti 1974: 7-8).

The approaches of Hilbert and Bernays (1939), Löb (1955) and Jeroslow (1973) are what might be called 'derivability conditions' approaches since they all rely for their application upon the satisfaction of some set of abstract conditions on $\mathrm{Prov}_T(x)$. Feferman (1957, 1960) has criticized such approaches on the ground that, for a given theory *T* and a given formula $\mathrm{Prov}_T(x)$ weakly representing the theorems of *T*, it is often difficult to determine whether or not such a set of conditions holds. He therefore sought a condition on provability expressions whose satisfaction or non-satisfaction is easier to determine, but which is still sufficient for G2.

By requiring that provability expressions preserve the logical and inductive structures of the usual definitions of formal provability, Feferman was able to reduce his condition to one governing the choice of expressions for the axioms of a system. That requirement is that the axiomhood formula be RE, where, roughly, this means that it preserves the way in which the axioms are 'presented'. An RE ('recursively enumerable') formula is one which transparently defines a recursively enumerable set of axioms. An RE formula representing an infinite set of axioms (which is the difficult case) imitates the inductive character of the metamathematical definition of that set (see

Smullyan 1961; Kreisel 1965; Feferman 1982, 1989). The advantage of using the RE criterion for G2 is that the RE representations form a primitive recursive set. Because of this, the question of whether a given expression of provability is RE is generally easier to answer than the question of whether it satisfies a given set of derivability conditions.

Using his new condition, Feferman proved the following version of G2 (1960: theorem 5.6).

If *T* is a system containing a minimal amount of arithmetic, $Ax_T(x)$ is an RE formula representing the axioms of *T* in *T*, and Con_T is built up from $Ax_T(x)$ in a way that corresponds to any of the usual ways in which the definition of consistency is related to the definition of axiomhood for *T*, then $\nvdash_T \mathrm{Con}_T$.

He also showed (theorem 5.9) that the restriction to RE cases of $Ax_T(x)$ is important since there are theories *S* and non-RE representations of their axioms $Ax^*_S(x)$ such that consistency formulas Con^*S - built up from $Ax^*_S(x)$ in exactly the same way that Con_T is built up from $Ax_T(x)$ - *are* provable in *S*. See Feferman (1982, 1989) for further development of his search for a general, readily applicable characterization of the representations of metamathematical notions with respect to which G2 holds.

Derivability conditions approaches have generally been enhanced by the recent development of provability logics (see Provability logic). The choice between the two types of approaches is complicated by the fact that Jeroslow’s criterion for G2 appears to apply to theories to which Feferman’s does not (namely, various cut-free systems of analysis - see Kreisel and Takeuti 1974). Thus, though Feferman’s G2 criterion may generally be easier to apply than those of the derivability conditions approaches, it does not appear to be as broad in scope (see Feferman’s remark 1 in Gödel (1990: 282-7) for a related discussion).

6 Philosophical applications

There have been many attempts to apply Gödel’s incompleteness theorems to philosophical concerns. For the most part, these have to do with issues in the philosophy of mathematics and the philosophy of mind, though there have also been attempts in other areas of philosophy, and also in areas of intellectual life not lying wholly within philosophy (see Thomas 1995). Here we consider some of the better-known and/or better-developed applications.

In the philosophy of mathematics the chief application is to Hilbert’s programme which is widely believed to have been refuted by G2. This is so because G2 is seen as prohibiting the finitary consistency proof that Hilbert demanded for classical mathematics (see, for example, Gödel 1958; Kitcher 1976; Kreisel 1958; Prawitz 1981; Resnik 1974; Smorynski 1977, 1985; Simpson 1988; for a dissenting view see Detlefsen 1986, 1990). G1 has been presented as having a similar effect through its exhibition of a recognizedly true real sentence of classical mathematics (namely, *G*) that is not finitarily provable (see Smorynski 1977, 1985; Hilbert’s programme and formalism §4).

Gödel’s theorems have also been used to argue against logicism. One example of this is in Hellman (1981), which maintains that G2 implies that no finitely axiomatizable logicist system exists. For non-finitely axiomatizable systems, the claim is weaker: such systems may exist, but G2 prohibits our being able to know of any particular system that it is one of them. A major attraction of Hellman’s argument is that, unlike other arguments against logicism, it does not require that one know where to draw the line between logic and non-logic.

Gödel’s theorems have also figured in philosophical discussions concerning the nature of proof. Among the more interesting of these are Myhill (1960) and Reinhardt (1985, 1986). Myhill argues that G1 and G2 establish that for any correct system containing arithmetic there are ‘correct’ inferences - inferences to whose acceptability the user of the system is rationally committed - that cannot be captured in it. There is thus, he says, an absolute epistemic notion of provability (neither syntactic, nor semantic nor psychological in character), according to which Gödel’s undecidable sentences are provable. Hence, this notion of absolute provability is not formalizable. Reinhardt (1985, 1986) refines aspects of this argument and also argues that if ‘humanly provable’ means ‘formally provable’ there must be absolutely undecidable sentences of arithmetic. He urges the view that G2 is an epistemic phenomenon - one that arises from the epistemic properties of that type of belief that mathematical proof is supposed to sponsor. These properties, it turns out, lead to something tantamount to the derivability conditions.

In a related vein, Dummett (1963) has used G1 to argue that mathematical proof (including that which lies within a

well-circumscribed subject area) is an 'indefinitely extensible concept' and cannot therefore be identified with derivation in any formal system. Since, he believes, the meanings of mathematical statements are to be given in terms of a concept of proof, this means that the notion of proof that determines the meanings of mathematical statements must be an essentially vague and unformalizable one. In this he sounds a theme similar in certain respects to one associated with the intuitionist, Brouwer.

Chaitin (1974, 1975, 1982) offers an analysis of Gödel's theorems that links them to limits on the extent to which information in a given mathematical field can be 'compressed' into a formal system. They thus, according to Chaitin, point up the need for a continuing search for new axioms (see Computability and information). This view of the relationship between Gödel's theorems and algorithmic information theory is challenged in van Lambalgem (1989).

In the philosophy of mind and metaphysics, the main application of the theorems is to the question of mechanism ('Do human-like minds have only such capabilities as are simulable by computational devices?' - see Mind, computational theories of §5) and, more generally, to the question of materialism ('Are all objects, events and/or forces in the world reducible to physical matter and its physical properties?' - see Materialism; Materialism in the philosophy of mind). Gödel himself (1951) presented such an application, arguing that either the mind of the human mathematician cannot be codified by any formal system or that there exist absolutely unsolvable arithmetical problems of an elementary sort. On the former alternative, supposing Church's thesis, mechanism clearly fails. On the latter, Gödel argued, materialism fails. For if there are mathematical problems that are absolutely unsolvable, then mathematics is not our own creation; and if this is so, its objects must exist independently of us, in which case a materialist view of reality fails. See Wang (1993) for a useful discussion of Gödel's argument.

The most well-known anti-mechanist argument is that of Lucas (1961), modified by Benacerraf (1967) and repeated in Penrose (1989, 1994). In outline, the argument is as follows. (1) If human minds were mechanizable (or, in some versions, if they could be known to be identical to some particular machine), then, by G2, they could not know that their beliefs are consistent. But (2) human minds can know that their beliefs are consistent; therefore, (3) they are not mechanizable (or, as in Benacerraf's variant, they cannot be known to be identical to any particular machine).

Lucas' argument has been widely criticized (see, for example, Boolos 1968; Chihara 1972; Dennett 1978; Putnam 1960 (before the fact); Smart 1961; Wang 1974; Webb 1968, 1980). Its chief difficulty is perhaps one that is partially concealed by the unclarity of premise (2). It is true that humans are capable of examining various of their sets of beliefs and arriving at credible judgments regarding their consistency. It does not follow from this, however, that among the sets of beliefs they can judge to be consistent are some that contain that very consistency judgment itself. Indeed, if fixing a set of beliefs is a typical precondition for its evaluation, it would seem that consistency judgments do not typically belong to the sets to which they are applied; for at the time the beliefs are fixed such judgments would not yet have been made and so would not exist as elements of the set being evaluated. For consistency evaluations not requiring such prior fixing of beliefs, on the other hand, it is not clear that G2 applies. Either way, Lucas-type arguments face serious difficulties (see Detlefsen 1995).

See also: Church's theorem and the decision problem; Church's thesis; Computability theory; Logical and mathematical terms, glossary of; Ordinal logics; Proof theory

MICHAEL DETLEFSEN

References and further reading

Benacerraf, P. (1967) 'God, the Devil, and Gödel', *The Monist* 51: 9-32.(A revised and chastened version of Lucas' anti-mechanist argument from the incompleteness theorems.)

Bernays, P. (1926) 'Axiomatische Untersuchung des Aussagen-Kalküls der *Principia Mathematica*' (Axiomatic Investigation of the Propositional Calculus of *Principia Mathematica*), *Mathematische Zeitschrift* 25: 305-20. (Condensed version of the author's 1918 *Habilitationsschrift* in which the first proof of the completeness of the propositional calculus of Whitehead and Russell was given.)

Boolos, G. (1968) 'Review of ''Minds, Machines and Gödel'', by J.R. Lucas, and ''God, the Devil, and Gödel'', by P. Benacerraf', *Journal of Symbolic Logic* 33: 613-15.(Critical review of Lucas (1961) and Benacerraf (1967).)

Boolos, G. and Jeffrey, R. (1974) *Computability and Logic*, Cambridge: Cambridge University Press, 3rd edn, 1989.(Thorough and well-written textbook of mathematical logic that provides much of the technical material required for the proofs given here.)

Chaitin, G.J. (1974) 'Information-Theoretic Limitations of Formal Systems', *Journal of the Association for Computing Machinery* 21: 403-24.(Introduction to Chaitin's analysis of Gödel's incompleteness theorems; argues that they show a limit on the information that can be 'compressed' into a formal axiomatic system, and infers that the search for new axioms must be an enduring part of mathematical research.)

Chaitin, G.J. (1975) 'Randomness and Mathematical Proof', *Scientific American* 232: 47-52.(Less technical version of various of the ideas in Chaitin (1974).)

Chaitin, G.J. (1982) 'Gödel's Theorem and Information', *International Journal of Theoretical Physics* 21: 941-54.(Attempt to explain Gödelian incompleteness as a phenomenon in algorithmic information theory.)

Chihara, C. (1972) 'On Alleged Refutations of Mechanism Using Gödel's Incompleteness Results', *Journal of Philosophy* 69: 507-26.(Critical discussion of attempted anti-mechanist applications of Gödel's incompleteness theorems.)

Dedekind, R. (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; trans. W.W. Beman (1901), 'The Nature and Meaning of Numbers', in *Essays on the Theory of Numbers*, New York: Dover, 1963. (Includes Dedekind's proof of the categoricity of the second-order Peano postulates, chapter 10, theorem 132.)

Dennett, D.C. (1978) 'The Abilities of Men and Machines', in *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: MIT Press, 1981.(Criticism of Lucas (1961) and similar arguments.)

Detlefsen, M. (1986) *Hilbert's Program: An Essay on Mathematical Instrumentalism*, Boston, MA, and Dordrecht: Reidel.(Detailed critical examination of the arguments from Gödel's incompleteness theorems against Hilbert's programme; focuses primarily on the arguments from G2 and the question of what is required of formulas that are to express consistency.)

Detlefsen, M. (1990) 'On an Alleged Refutation of Hilbert's Program using Gödel's First Incompleteness Theorem', *Journal of Philosophical Logic* 19: 343-77.(Critical examination of the arguments against Hilbert's programme from G1.)

Detlefsen, M. (1995) 'Wright on the Non-Mechanizability of Intuitionist Reasoning', *Philosophia Mathematica*, series 3, 3: 103-19.(Critical examination of an argument by Crispin Wright that Gödel's theorems imply the non-mechanizability of intuitionist mathematics. Includes brief statement of what the author regards as the central mistake of all Lucas-type arguments.)

Dummett, M.A.E. (1963) 'The Philosophical Significance of Gödel's Theorem', in *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press, 1978.(Includes Dummett's argument that Gödel's incompleteness theorems show that the notion of mathematical proof is essentially unformalizable.)

Enderton, H.B. (1972) *A Mathematical Introduction to Logic*, New York: Academic Press, 1972. (Well-conceived and -executed basic textbook of mathematical logic; supplies definitions and proofs of the basic notions and results needed to prove G1. Treatment of G2 is sketchy.)

Feferman, S. (1957) 'Formal Consistency Proofs and Interpretability of Theories', in *Summaries of Talks Presented at the Summer Institute of Symbolic Logic in 1957 at Cornell University*, vol. 1, 71-7. Mimeographed.(A succinct and readable summary of the author's Berkeley doctoral dissertation.)

Feferman, S. (1960) 'Arithmetization of Metamathematics in a General Setting', *Fundamenta Mathematicae* 49: 3-92.(Thorough treatment of the differences between G1 and G2 as regards the conditions they require of arithmetical expressions of metamathematical notions. Develops a novel and interesting alternative to the traditional derivability conditions approach to the proof of G2.)

Feferman, S. (1982) 'Inductively Presented Systems and the Formalization of Metamathematics', in D. van Dalen, D. Lascar and T. Smiley (eds) *Logic Colloquium '80*, Studies in Logic and the Foundations of Mathematics, 108, Amsterdam: North Holland, 1982.(Outline of investigation designed to extend and refine the study begun in the author's dissertation and 1960 paper. The aim is a fully explicit and precise account of canonical representations of the various metamathematical notions pertaining to formal axiomatic systems.)

Feferman, S. (1989) 'Finitary Inductively Presented Logics', in R. Ferro, Bonotto, S. Valentini and A. Zanardo (eds) *Logic Colloquium '88*, Studies in Logic and the Foundations of Mathematics, 127, Amsterdam: North Holland, 1989.(Continuation of the author's long-standing study of what is required for the canonical representation of formal systems.)

Gauss, K. (1817, 1829) letter to Olbers (1817), in *Briefwechsel mit H.W.M. Olbers*, Hildesheim: Olms, 1976; letter to Bessel (1829), in *Werke*, Leipzig: Teubner, 1863-1903, vol. 8, 200.(Statement of Gauss' belief that though a human knower might be able to attain complete knowledge of the laws of arithmetic a priori, the same would not be true of the laws of geometry.)

Gödel, K. (1929) ‘Über die Vollständigkeit des Logikkalküls’, doctoral dissertation, University of Vienna; trans. ‘On the Completeness of the Logical Calculus’, in Gödel (1986), 44-101; slightly modified version, ‘Die Vollständigkeit der Axiome des logischen Funktionenkalküls’, *Monatshefte für Mathematik und Physik* 37: 349-60, 1930; trans. ‘The Completeness of the Axioms of the Functional Calculus of Logic’, in van Heijenoort (1967), 582-91; and in Gödel (1986), 102-23.(Gödel’s doctoral dissertation in which he first proved the completeness of the quantificational portion of the logic of Whitehead and Russell.)

Gödel, K. (1930) ‘Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit’, repr. and trans. ‘Some Metamathematical Results on Decidability and Consistency’, in Gödel (1986); trans. in van Heijenoort (1967).(Brief announcement of Gödel’s incompleteness theorems.)

Gödel, K. (1931) ‘Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I’, *Monatshefte für Mathematik und Physik* 38: 173-98; trans. ‘On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems’, in van Heijenoort (1967), 592-617; and in Gödel (1986), 144-95.(Paper in which were originally published Gödel’s proofs of his incompleteness theorems.)

Gödel, K. (1951) ‘Some Basic Theorems on the Foundations of Mathematics and Their Implications’, in Gödel (1995).(Edited text of Gödel’s 1951 Josiah Gibbs Lecture delivered at Brown University at a meeting of the American Mathematical Society. Gödel here formulated his argument concerning the anti-materialist implications of his incompleteness theorems; also includes an argument for a Platonist conception and an argument against a conventionalist conception of mathematics. Useful introductory essay by George Boolos.)

Gödel, K. (1958) ‘Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes’, repr. and trans. ‘On a Hitherto Unexploited Extension of the Finitary Standpoint’, in Gödel (1990).(Gödel’s attempt to find a legitimate epistemological extension of Hilbert’s finitary standpoint that is capable of providing the kinds of soundness proofs for classical mathematics that Hilbert called for. Includes Gödel’s statement that, unlike in his 1931 paper, he now accepted the idea that G2 essentially refutes Hilbert’s programme.)

Gödel, K. (1986-95) *Collected Works*, ed. S. Feferman *et al.*, vol. 1, *Publications 1929-1936*, 1986; vol. 2, *Publications 1938-1974*, 1990; vol. 3, *Unpublished Essays and Lectures*, 1995, New York and Oxford: Oxford University Press. German originals and English translations.(Accurate yet readable translations, useful introductory essays and bibliography.)

Hasenjäger, G. (1953) ‘Eine Bemerkung zu Henkins Beweis für die Vollständigkeit des Prädikatenkalküls der ersten Stufe’ (A Remark on Henkin’s Proof of the Completeness of the First-Order Predicate Calculus), *Journal of Symbolic Logic* 18: 42-8.(A simplification of Henkin’s completeness proof.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press.(Basic papers in mathematical logic and the foundations of mathematics, with useful forewords and bibliography.)

Hellman, G. (1981) ‘How to Gödel a Frege-Russell: Gödel’s Incompleteness Theorems and Logicism’, *Noûs* 25: 451-68.(Argues that G2 counts against logicism.)

Henkin, L. (1949) ‘The Completeness of the First-Order Functional Calculus’, *Journal of Symbolic Logic* 14: 159-66.(First presentation of the currently prevailing method for completeness proofs; introduced devices making the proof less dependent upon the details of the particular calculus involved and, hence, more general in scope.)

Henkin, L. (1952) ‘A Problem concerning Provability’, *Journal of Symbolic Logic* 17: 160.(Posed the question ‘Is a sentence expressing the statement that it is itself provable in *T*, a first-order formal system containing enough arithmetic to enable it to represent all recursive relations of natural numbers, provable in *T* or not?’ Solved in the affirmative by Löb (1955).)

Hilbert, D. (1900) ‘Über den Zahlbegriff’ (On the Number-Concept), *Jahresbericht der Deutschen Mathematiker-Vereinigung* 8; revised as appendix 6 of *Grundlagen der Geometrie*, Leipzig and Berlin: Teubner, 7th edn, 1930.(Early statement by Hilbert of the chief aims of foundational studies. Explicitly identified are a completeness problem, and the consistency problem, which is described as necessary for the demonstration of the existence of a concept.)

Hilbert, D. and Ackermann, W. (1928) *Grundzüge der theoretischen Logik*, Berlin: Springer, 2nd edn, 1938; 2nd edn trans. L.M. Hammond, G.G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, ed. R.E. Luce, New York: Chelsea, 1950.(Text in which the standard problems of metamathematics - consistency, completeness, independence and decidability - received their first precise formulations; readable and illuminating source for students today. These formulations guided Gödel and Church in their subsequent research.)

Hilbert, D. and Bernays, P. (1939) *Grundlagen der Mathematik (Foundations of Mathematics)*, vol. 2, Berlin: Springer; 2nd edn, 1970.(Classic work of mathematical logic; gives the first full proof of G2.)

Hintikka, J. (1955) ‘Form and Content in Quantification Theory’, *Acta Philosophica Fennica* 8: 7-55. (Description of a tableau method that supports a simple proof of completeness.)

Huntington, E.V. (1905) ‘A Set of Postulates for Real Algebra, Comprising Postulates for a One-Dimensional Continuum and for the Theory of Groups’, *Transactions of the American Mathematical Society* 6: 17-41. (Includes early explicit mention of completeness as a property of an axiom system (see page 17).)

Jeroslow, R. (1973) ‘Redundancies in the Hilbert-Bernays Derivability Conditions for Gödel’s Second Incompleteness Theorem’, *Journal of Symbolic Logic* 38: 359-67.(Shows how to eliminate DC2 as a condition on provability expressions whose associated consistency expressions support a version of G2.)

Kitcher, P. (1976) ‘Hilbert’s Epistemology’, *Philosophy of Science* 43: 99-115.(Useful discussion of the epistemic character of Hilbert’s so-called ‘finitary’ judgments and of the bearing of Gödel’s theorems on Hilbert’s programme.)

Kreisel, G. (1958) ‘Hilbert’s Programme’, *Dialectica* 12: 346-72; revised version in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (Essay on the plausibility of Hilbert’s programme in light of Gödel’s theorems and other developments; discusses the differences between G1 and G2 and between what Kreisel calls the broader and narrower versions of Hilbert’s programme; claims that G1 counts against the broader version.)

Kreisel, G. (1965) ‘Mathematical Logic’, in T.L. Saaty (ed.) *Lectures in Modern Mathematics*, vol. 3, New York: Wiley.(Survey of modern mathematical logic. Considers the conditions that must be satisfied by formulas expressing metamathematical notions, and introduces the notion of a ‘canonical’ representation.)

Kreisel, G. and Takeuti, G. (1974) ‘Formally Self-Referential Propositions for Cut-Free Classical Analysis and Related Systems’, *Dissertationes Mathematicae* 118: 4-50.(Useful study of the conditions on formulas expressing provability that are required for G2 and such related results as Löb’s theorem. Shows how the streamlining of the derivability conditions discovered by Jeroslow allows a proof of G2 for certain cut-free systems of analysis.)

Kronecker, L. (1887) ‘Über den Zahlbegriff’ (On the Concept of Number), *Journal für die reine und angewandte Mathematik* 101: 337-55; repr. in *Werke*, ed. K. Hensel, Leipzig: Teubner, 1932, vol. 3, part 1, 249-74. (Kronecker’s only philosophical essay, in which he expresses the view, also expressed earlier by Gauss, that although humans might be able to attain a complete knowledge of the laws of arithmetic a priori, the same cannot be expected of the laws of geometry.)

Lambalgem, M. van (1989) ‘Algorithmic Information Theory’, *Journal of Symbolic Logic* 54: 1,389-400.(Argues that Chaitin’s analysis of Gödel’s theorem in terms of algorithmic information theory does not in fact explain it.)

Löb, M.H. (1955) ‘Solution of a Problem of Leon Henkin’, *Journal of Symbolic Logic* 20: 115-18. (Positive solution of the problem posed by Henkin (1952). Streamlined the original conditions on provability expressions used in the first full proof of G2 in Hilbert and Bernays (1939).)

Löwenheim, L. (1915) ‘Über Möglichkeiten im Relativkalkül’, *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, ‘On Possibilities in the Calculus of Relatives’, in van Heijenoort (1967), 232-51.(Includes a proof of Löwenheim’s theorem that a formula true in all finite and denumerably infinite domains is true in all domains. Makes use of a normal form for first-order formulas that Gödel also made use of in his completeness proof.)

Lucas, J.R. (1961) ‘Minds, Machines, and Gödel’, *Philosophy* 36: 112-27; revised version appears in *Freedom of the Will*, Oxford: Clarendon Press, 1970.(First statement of a well-known and widely discussed argument that Gödel’s incompleteness theorems refute mechanism.)

Lucas, J.R. (1970) *Freedom of the Will*, Oxford: Clarendon Press.(Includes responses to some of the criticisms of the earlier version of the argument (1961) and gives a select bibliography of the anti-mechanist argument from Gödel’s theorems.)

Maltsev, A.I. (1936) ‘Untersuchungen aus dem Gebiete der mathematischen Logik’, *Matematicheskii Sbornik* 1: 323-6; trans. ‘Investigations in the Area of Mathematical Logic’, in B. Wells (ed.) *The Metamathematics of Algebraic Systems: Collected Papers, 1936-1967*, Amsterdam: North Holland, 1971.(Extends to uncountable sets of formulas the proof of compactness given in Gödel (1930) for countable sets of formulas.)

Myhill, J. (1960) ‘Some Remarks on the Notion of Proof’, *Journal of Philosophy* 62: 461-70.(Argues that Gödel’s incompleteness theorems show that for any correct formal system containing arithmetic there are inferences to whose acceptability the user of the system is rationally committed that cannot be captured in it.)

Paris, J. and Harrington, L. (1977) ‘A Mathematical Incompleteness in Peano Arithmetic’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland.(First proof of a reasonably natural and true statement of pure combinatorics that is not provable in first-order Peano arithmetic. Followed a few years later

by an even more striking discovery of Harvey Friedman's (see Simpson 1985).)
Penrose, R. (1989) *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, New York: Oxford University Press.(Argues against mechanism in a manner essentially that of Lucas (1961).)
Penrose, R. (1994) *Shadows of the Mind: A Search for the Missing Science of Consciousness*, New York: Oxford University Press.(Includes replies to critics of the anti-mechanist argument of Penrose (1989). Essential elements of the argument remain unchanged.)
Post, E. (1921) 'Introduction to a General Theory of Elementary Propositions', *American Journal of Mathematics* 43: 163-85; repr. in van Heijenoort (1967), 264-83.(First published proof of the completeness of the propositional calculus of Whitehead and Russell; also proved its Post-completeness.)
Prawitz, D. (1981) 'Philosophical Aspects of Proof Theory', in G. Fløistad (ed.) *Contemporary Philosophy: A New Survey*, vol. 1, The Hague: Martinus Nijhoff, 1981.(Useful presentation of the case against Hilbert's programme via Gödel's theorems, especially G1.)
Putnam, H. (1960) 'Minds and Machines', in S. Hook (ed.) *Dimensions of Mind: A Symposium*, New York: Macmillan.(Predates Lucas (1961), but none the less has the makings of an effective argument against it.)
Reinhardt, W. (1985) 'Absolute Versions of Incompleteness Theorems', *Noûs* 19: 317-46.(Argues that epistemic properties of an intuitive notion of provability give rise to G1 and G2 phenomena.)
Reinhardt, W. (1986) 'Epistemic Theories and the Interpretation of Gödel's Incompleteness Theorems', *Journal of Philosophical Logic* 15: 427-74.(Extends and refines the discussion of Reinhardt (1985).)
Resnik, M. (1974) 'The Philosophical Significance of Consistency Proofs', *Journal of Philosophical Logic* 3: 133-47.(Discussion of certain elements of Feferman (1960) and the question of the difference between normal and 'pathological' expressions of metamathematical notions.)
Rosser, J.B. (1936) 'Extensions of Some Theorems of Gödel and Church', *Journal of Symbolic Logic* 1: 87-91. (Proof of a version of G1 that does not rely upon a theory's being ω-consistent.)
Simpson, S. (1985) 'Nonprovability of certain combinatorial properties of finite trees', in L. Harrington *et al.* (eds) *Harvey Friedman's Research in the Foundations of Mathematics*, Amsterdam: North Holland, 1985.(First published presentation of Friedman's proof in the early 1980s that a certain proposition of finite combinatorics (a form of Kruskal's theorem on finite trees) is independent not only of first-order Peano arithmetic but also of two subsystems of second-order Peano arithmetic that go well beyond it.)
Simpson, S. (1988) 'Partial Realization of Hilbert's Program', *Journal of Symbolic Logic* 53: 349-63.(Useful survey and basic exposition of the work of the so-called 'reverse mathematics' programme begun by Harvey Friedman. Argues that though Hilbert's programme in its original form is refuted by G2, a significant portion of that programme can still be carried out.)
Skolem, T. (1923) 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre', in *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, Helsinki: Akademiska Bokhandeln, 217-32; trans. S. Bauer-Mengelberg, 'Some Remarks on Axiomatized Set Theory', in van Heijenoort (1967), 291-301. (Anticipated the chief result of Gödel (1930). Gödel was apparently unaware of it at the time he wrote his paper.)
Smart, J.J.C. (1961) 'Gödel's Theorem, Church's Theorem and Mechanism', *Synthese* 13: 105-10. (Critical discussion of Lucas-type arguments.)
Smorynski, C. (1977) 'The Incompleteness Theorems', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam, North Holland.(Exposition of Gödel's incompleteness theorems, their proofs and some of their applications.)
Smorynski, C. (1985) *Self-Reference and Modal Logic*, New York: Springer.(Good exposition of provability logic. Chapter 0 presents argument against Hilbert's programme via G1.)
Smullyan, R. (1961) *Theory of Formal Systems*, Princeton, NJ: Princeton University Press.(Important study of the nature of formal systems and what is required for the adequate representation of metamathematical notions via arithmetization.)
Thomas, D. (1995) 'Gödel's Theorem and Postmodern Theory', *Publications of the Modern Language Association* 110: 248-61.(Discussion of the significance of Gödel's incompleteness theorems for literary criticism.)
Wang, H. (1974) *From Mathematics to Philosophy*, New York: Humanities Press.(Includes worthwhile discussion of a variety of aspects of Gödel's theorems and of his philosophical ideas; also a critical discussion of Lucas-type arguments against mechanism.)
Wang, H. (1993) 'Can Machines Think?', *Philosophia Mathematica*, series 3, 1: 97-138.(Discussion of Gödel's ideas on mechanism. Reconstructs the many conversations between the author and Gödel on the subject.)
Webb, J. (1968) 'Metamathematics and the Philosophy of Mind', *Philosophy of Science* 35: 156-78. (Interesting

discussion of some of the difficulties involved in trying to link Gödel’s incompleteness theorems to anti-mechanist views of mind.)

Webb, J. (1980) *Mechanism, Mentalism, and Metamathematics: An Essay on Finitism*, Boston, MA: Reidel. (Thorough study of various themes in philosophy with which Gödel’s incompleteness theorems have sometimes been associated. Argues that, contrary to the opinion of some, these theorems actually confirm a type of formalism in the philosophy of mathematics.)

Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press; 2nd edn, 1925; repr. 1994.(The classical symbolic formalization of logic and mathematics; basis of much of the greatest work in mathematical logic and the foundations of mathematics in the twentieth century.)

Godfrey of Fontaines (*c*.1250-*c*.1306/9)

Godfrey of Fontaines studied philosophy and theology at the University of Paris and subsequently taught theology there. A theologian by profession, he developed a highly interesting philosophy, especially a metaphysics. For Godfrey, metaphysics studies being as being. Being itself is divided into cognitive being and real being, and real being is divided into being in act and being in potency. In finite beings, essence and existence are neither really distinct nor intentionally distinct; they are identical. Human reason can prove that God exists, and reach some imperfect knowledge concerning what God is, but cannot prove that the world began to be. For Godfrey, corporeal entities in this world are composed of matter and form, but heavenly bodies probably lack prime matter. On philosophical grounds, he favours the theory that there is only one substantial form in human beings - the intellective soul - but for theological reasons leaves this question open. His philosophy is somewhat more Aristotelian and less Neoplatonic than that of most of his contemporaries.

1 Life and works

Born in the principality of Liège in modern Belgium, probably shortly before 1250, Godfrey of Fontaines studied philosophy at the Faculty of Arts of the University of Paris, presumably before 1274, and then completed the lengthy programme in the theology faculty, eventually becoming regent master in theology in 1285. He continued to teach theology at Paris until about 1298-9, and then returned to the same position around 1303-4. He died around 1306, or perhaps in 1309.

His major contributions to philosophical and theological literature were his fifteen *Quodlibeta (Quodlibetal Questions)*. These originated from public disputations which were open to anyone who might wish to pose questions about any reasonable topic, and which were again disputed and 'determined' (resolved) by the presiding master on some following day. Godfrey also conducted other 'ordinary' disputations, and some of these have been preserved. His philosophy, especially his metaphysics, is highly interesting, and he often turned to it in deciding theological issues.

2 Theory of being

With Aristotle, Godfrey holds that metaphysics has being as being as its subject. Being is the most general notion the human mind can discover, and is affirmed primarily of substance and of everything else as in some way ordered to substance. Godfrey divides being into cognitive being (being in the mind of a knower, whether human or divine), and real being (being outside the mind). Real being is divided into being in potency and being in act. For a thing to have being in potency is for it to exist in one or more causes which have the capacity to bring it into actual being. A thing may have potential being by reason of an intrinsic cause, for example, pre-existing matter which would enter into that thing's structure when it receives actual being, or by reason of an extrinsic cause (or causes) which can bring it into actual being. For a thing to have actual being is for it to exist in its own right with its own complete form or nature (see Being).

3 Theory of essence and existence

Much debated during Godfrey's time was how essence is related to existence in created entities. It was generally conceded that they are identical in God. Godfrey strongly rejects the view that essence and existence are really distinct in creatures. He directs special attention to the way this theory was defended by Giles of Rome, who had referred on some occasions to this as a distinction between two 'things'. However, Godfrey also rejects, at least by implication, the more nuanced distinction of essence and act of being (*esse*) defended by Thomas Aquinas. He also rejects Henry of Ghent's claim that they are intentionally distinct, that is, by a distinction that is less than real but more than purely mental (see Henry of Ghent). For Godfrey, essence and existence are really identical, so strictly so that whatever is true of one is true of the other, and whatever is known of one is known of the other. He denies (against Thomas and Giles, but also against Henry) that one can know what something is (its essence) without knowing that it is (its existence). He tries to account for our awareness of nonexistent possible entities simply by appealing to the distinction between our knowing something as only potentially existing and our knowing it as actually existing. Neither real nor intentional distinction of essence and existence is needed to account for this (see Existence).

To account for the non-simple character of created entities, especially purely spiritual entities such as angels, Godfrey maintains that such beings fall short of God, the supreme or first ‘One’, by possessing an intermediary nature. In comparison with God they may be regarded as potential, but in comparison with less perfect created beings they may be regarded as actual. This is enough for him to defend their act-potency ‘composition’ and thereby distinguish them from God. This solution blends Aristotelianism (act-potency composition) and Neoplatonism (Proclus’ notion that what participates in the ‘One’ or God, is both one and not one) (see Aristotelianism, medieval; Platonism, medieval).

4 Philosophical knowledge of God

Godfrey accepts the prevailing view of his time that in this life human beings do not enjoy direct or intuitive knowledge of God. In addition to the knowledge of God afforded believers by divine revelation, Godfrey defends the human mind’s ability to reason from knowledge of effects to knowledge of God as their first cause. He assigns some role in this process to argumentation based on motion which concludes to a First Mover in accord with Aristotle’s procedure in the *Physics*, but reserves more precise knowledge of God to metaphysics.

Godfrey does not fully accept Aquinas’s view that by reasoning philosophically we can know that God is and what God is not (with which Godfrey agrees) but not what God is. Godfrey maintains that through appropriate metaphysical reasoning we can also know something about what God is, though such knowledge will always be incomplete and imperfect. In discussing God’s knowledge and production of created entities, Godfrey appeals to a theory of divine ideas. For him these are simply different ways in which God understands his essence as capable of being imitated by creatures. The divine ideas are not really distinct either from one another or from the divine essence. Even so, Godfrey does postulate a plurality of divine ideas to correspond to different specific classes of creatures but, against the prevailing view of his time and in agreement with Henry of Ghent, he denies that divine ideas are multiplied to correspond to different individuals within species (see God, concepts of).

5 Eternal creation

Sharply contested during Godfrey’s time was the issue concerning whether or not philosophical argumentation can demonstrate that the world began to be. All participants in this discussion acknowledged that, on the strength of Scripture, we can be certain that the world began to be, but some, such as Henry of Ghent, maintained that human argumentation can demonstrate this conclusion. Others, including Thomas Aquinas, Siger of Brabant and Boethius of Dacia denied that human reason can prove this. In addition some, including Aquinas near the end of his career in his *De aeternitate mundi (On the Eternity of the World)*, went so far as to maintain that human reason can establish the possibility of an eternally created world. Godfrey holds that human reason cannot prove that the world began to be, but, while he is heavily dependent on Aquinas’ *De aeternitate mundi*, he refuses to go so far as to claim that human reason can prove that an eternally created world is possible (see Eternity of the world, medieval views of).

6 Substance and accidents

Godfrey accepts the usual Aristotelian distinction between substance and accidents, but rigidly applies the equally Aristotelian theory of act and potency to their relationship. If accidents inhere in their appropriate substances as secondary actualities in an underlying potential subject, can a substance efficiently cause any of the accidents that inhere in it? For Godfrey this is impossible, for it would mean that a substance would be active (as efficiently causing its accidents) and passive (as receiving them) at one and the same time with respect to the same thing. He is especially critical of Henry of Ghent’s claim that a substance may be regarded as the proximate efficient cause of its proper accidents and operations. One particular point of disagreement had to do with the causal explanation of human activities, especially volition. In such cases Henry maintains that a human agent, because of the spiritual nature of the will, can reduce itself from what he calls virtual act to formal act, and that this happens in acts of choice (see Henry of Ghent). For Godfrey this would violate the Aristotelian view that whatever is moved is moved by something else and that nothing can reduce itself from potency to act. The human will is not the efficient cause of its acts of choice since they inhere in the will itself: their efficient cause is rather the object of choice insofar as this is presented to the will by the intellect. Not surprisingly, Godfrey’s highly intellectualist account of human choice was criticized by some as leading to determinism. Closely connected with Godfrey’s position concerning this is his defence of a real distinction between the soul and its powers, including the intellect and the

will.

7 Matter and form

Godfrey accepts the Aristotelian view that corporeal entities are composed of prime matter and substantial form, and that they unite as potency and act to constitute the essence of such entities. He strongly insists that prime matter is pure potency, and denies, especially against Henry of Ghent, that it enjoys some minimum degree of actuality in itself that would be sufficient for God to keep it in existence without any form. In arguing for the matter-form composition of material entities to account for the change of one material substance into another, Godfrey is ultimately inspired by Aristotle's procedure in *Physics* I. Against the view defended by some thirteenth-century thinkers, Godfrey rejects any kind of matter-form composition of purely spiritual entities such as human souls and angels. He also rejects the view proposed by some, such as Aquinas, that because of their incorruptible character, heavenly bodies are composed of a special kind of prime matter and their appropriate substantial forms. Because he also accepts the prevailing view that heavenly bodies are not subject to corruption, Godfrey is forced into a difficult position. Drawing upon Averroes, he proposes that it is better to say that there is no matter in the proper sense in the heavenly bodies; for if there were, they would be subject to corruption (see Ibn Rushd). Godfrey refers to this position as being 'more probable'.

Closely related to Godfrey's views on matter and form is his consideration of another contested issue: unity versus plurality of substantial forms in corporeal substances, especially in human beings. Aquinas had argued that in human beings the intellective soul is their single substantial form, but his position had been condemned after his death by two Archbishops of Canterbury, Robert Kilwardby in 1277 and John Pecham in 1284 and 1286. Godfrey's philosophical argumentation leads him to favour unicity of substantial form in all corporeal entities, including human beings. However, theological concerns, especially the need to explain how the body of Christ could continue to retain its identity as his body during the three days in the tomb, perplex him. Hence, while he insists upon the theologian's right to defend unicity of substantial form, Godfrey styles this theory as more probable but does not adopt it definitively.

Godfrey devotes considerable effort to determining the respective roles of form, prime matter, and quantity in the individuation of material entities. He concludes that a thing's substantial form is its formal principle of individuation, and that quantity plays the role of material dispositive cause by rendering matter divisible into numerically distinct parts which can therefore receive different substantial forms of the same kind. His philosophy is somewhat more Aristotelian and somewhat less Neoplatonic than that of most of his contemporaries.

See also: Aristotelianism, medieval; Henry of Ghent §2

JOHN F. WIPPEL

List of works

Godfrey of Fontaines (1285-1304) *Quodlibeta (Quodlibetal Questions)*, ed. M. De Wulf, J. Hoffmans and O. Lottin in Les Philosophes Belges II-V, XIV, Louvain: Institut supérieur de Philosophie de l'Université, 1904-37.(Quodlibets 1-4 are in vol. II, ed. M. De Wulf and J. Hoffmans, 1904; Quodlibets 5-7 are in vol. III, ed. M. De Wulf and J. Hoffmans, 1914; Quodlibets 8-10 are in vol, IV, ed. J. Hoffmans, 1924-31; Quodlibets 11-14 are in vol. V, ed. J. Hoffmans, 1932-35; Quodlibets 15 plus three ordinary questions are in vol. XIV, ed. O. Lottin, 1937. Wippel (1981) has proposed dates of composition for individual quodlibets.)

Godfrey of Fontaines (*c*.1285) Disputed Questions.(Edited questions can be found in B. Neumann, *Der Mensch und die himmlische Seligkeit nach der Lehre Gottfrieds von Fontaines*, Limburg: Lahn-Verlag, 1958; O. Lottin, *Psychologie et morale aux XII^e et XIII^e siècle*, Louvain: Abbaye du Mont César, and Gembloux, J. Duculot, 1954; and J.F. Wippel, 'Godfrey of Fontaines: Disputed Questions 9, 10 and 12', *Franciscan Studies* 33: 356-72. For a listing of those questions which have been edited either in whole or in part see Wippel 1981: xxxi-xxxii.)

References and further reading

De Wulf, M. (1904) *Un théologien-philosophe du XIIIe siècle. Etude sur la vie, les oeuvres et l'influence de Godefroid de Fontaines (A Theologian-Philosopher of the Thirteenth Century: Study on the Life, Works and Influence of Godfrey of Fontaines)*, Brussels: Hayez.(Still the most thorough account of Godfrey's life.)

Putallaz, F.-X. (1995) *Insolente liberté - controverses et condamnations au XIII^e siècle*, Fribourg: Éditions

Universitaires de Fribourg Suisse.(Chapters (here called ‘situations’) VI and VII are especially helpful on Godfrey’s place in late thirteenth-century discussions concerning the respective roles of the intellect and the will in accounting for human choice.)

Wippel, J.F. (1981) *The Metaphysical Thought of Godfrey of Fontaines: A Study in Late Thirteenth-Century Philosophy*, Washington, DC: Catholic University of America Press.(A general account of Godfrey’s metaphysical thought and his contributions to philosophy, with bibliography.)

Wippel, J.F. (1982) ‘Godfrey of Fontaines’ Disputed Questions 9 and 10 (Bruges 491): by Godfrey or by Giles of Rome?’, *Franciscan Studies* 42: 216-47.(On the authenticity of two Disputed Questions attributed to Godfrey and edited by Wippel in 1973.)

Wippel, J.F. (1984) ‘Possible Sources for Godfrey of Fontaines’ Views on the Act-Potency ‘Composition’ of Simple Creatures’, *Mediaeval Studies* 46: 222-44.(Proposes an anonymous set of questions on Aristotle’s *Posterior Analytics* as another source for Godfrey’s theory in addition to Siger and Proclus.)

Godwin, William (1756-1836)

William Godwin is considered the founder of philosophical anarchism. His An Enquiry Concerning Political Justice (1793) contends that although government is a corrupt force in society, perpetuating dependence and ignorance, it will increasingly be rendered impotent by the gradual spread of knowledge. Politics will be displaced by an enlarged morality as truth conquers error and mind subordinates matter. He predicts the end to cooperative activities (which restrain individual freedom), the abandonment of marriage and private property, and increasing longevity and ultimate immortality.

Godwin's moral theory is often described as utilitarian, but he understands pleasure to be inseparable from the development of truth and wisdom through the full and free exercise of private judgment and public discussion. As such, his position is better understood as perfectionist.

1 Life

William Godwin was born 3 March 1756, the son of a Dissenting minister, and was educated for the ministry at the Hoxton Dissenting Academy. When his vocation failed he turned to political journalism. Caught up in the wave of enthusiasm which greeted the French Revolution in 1789, he took an active interest in the controversy which followed Edmund Burke's attack (see Burke, E.) on the revolutionaries and their English supporters (notably Richard Price and the Dissent-dominated London Revolution Society). Godwin's *Enquiry Concerning Political Justice*, written between September 1791 and January 1793, was intended as a summary of recent developments in moral and political philosophy. It begins with evident debts to the *philosophes* (especially Baron d'Holbach, Claude-Adrien Helvetius and Jean-Jacques Rousseau), to Thomas Paine's *Common Sense* (1775) and *Rights of Man* (1791-2; see Paine, T.) and to Price's Platonism. But as the work develops, Godwin's own position emerges, the *philosophe* influence is displaced by a secular interpretation of several central themes of rational dissent: notably the importance of candour, the doctrine of philosophical necessity, the irresistible power of truth, and the sanctity of private judgment. From these materials Godwin forges a resounding attack on contemporary political institutions and a philosophical anarchism which rejects all government interference.

Political Justice brought Godwin widespread renown, which *Caleb Williams* (1794), his most successful novel, confirmed. By the end of the decade, however, he had fallen victim to the loyalist reaction - a fate he partly invited by his frank memoir of his deceased wife, Mary Wollstonecraft (1798). While he continued to write until his death, he never regained his reputation.

2 Ethics

Godwin's moral position as expressed in *Political Justice* is complex because he endorses two potentially conflicting moral principles. In the chapter 'Of Justice' (II.2), he uses his most famous moral example, 'The Famous Fire Cause'. He argues that when faced with a situation in which I can save either Fenelon (see Fenelon, F.), the illustrious Archbishop of Cambrai, or his chambermaid, I should save Fenelon, even if the chambermaid is my mother. This is justified not on grounds of desert, but because of the benefit which Fenelon will confer on humankind. This argument encapsulates Godwin's more general claim that we should act to maximize utility, motivated wholly by disinterested benevolence, and unaffected by familial or other attachments.

Godwin's argument involves a most demanding conception of personal rectitude (albeit softened by changes in later editions of *Political Justice* and elsewhere). His position rests on three elements: that we can all come to recognize the same or similar moral or ethical principles upon which to base our conduct; that we will act on the basis of these principles without incentives from self-interest or coercion; and that we will order every aspect of our lives, public and private, in accordance with these principles. The first premise derives from Godwin's background in rational dissent, coupled with a more general enlightenment optimism, which assumed convergence in matters of belief. The second rests on an understanding of necessitarianism derived initially from David Hartley, Jonathan Edwards and Joseph Priestley, which held that when feeling contends with moral truth the understanding experiences this as a contention of ideas, and will resolve in favour of the more general and powerful (a view also indebted to Price's Platonism). The third has its origins both in Godwin's classical reading, with its ideal of the virtuous citizen, and in his theological training, in which subordination to God is equally demanding of public and

private life.

Godwin's endorsement of the utility principle is qualified by two arguments: that the happiness of another can best be promoted indirectly, by enlightening the understanding; and that pleasures differ qualitatively, with wisdom and benevolence affording infinitely superior pleasure to sensual gratification. His utilitarianism is, then, indirect and ideal (see Utilitarianism).

Godwin's second basic moral principle is the right of private judgment. Morality requires that we act wholly upon the dictates of the understanding. We have a duty to justice, but our only route to justice is through the full and free exercise of private judgment. Private judgment is best seen as a right for Godwin, despite his rejection (most forcefully in the first edition of *Political Justice*) of the language of rights. It is not reducible to utility because its exercise is both an indefeasible duty and the highest form of existence we can attain. Clearly, there is utility to be gained from respecting private judgment, but we cannot abrogate an individual's judgment for the sake of utility - not least because there is no infallible standard for what utility demands. This suggests that Godwin's position is best characterized as perfectionist: the true end of moral action is the progressive improvement of our moral and intellectual faculties.

Godwin's position is distinctive because he combines an act-centred conception of moral duty - act so as to maximize overall benefit - with an agent-centred, perfectionist conception of moral agency and virtue - according to which action is only fully moral when intention and outcome are brought together in a fully rational will. Neither government nor individuals may legitimately transgress another's private judgment. We have a right to remonstrate with them, but we have a duty to respect their considered judgment. Behind this account lies a deeper theory, driven by the doctrine of necessity and by the belief that mind is potentially sovereign over desire and feeling, which sees the apex of human perfection as a life lived in the light of truth. It is not difficult to recognize here the residual effect of Godwin's Dissenting inheritance.

Godwin's philosophical position underwent a number of changes at the end of the 1790s and after: his rationalism ebbed, and he endorsed a sentiment-based theory of moral judgment and moral motivation. Yet, even while making the changes, Godwin insisted that they did not touch the central tenets of his position. This is not an easy claim to assess. The later editions of *Political Justice* (1796, 1798) advance a more tepid doctrine of progress than the first. In *The Enquirer: Reflections on Education, Manners and Literature* (1797) sympathy is clearly rehabilitated, as is feeling more generally. It is also possible to identify a more nuanced doctrine of moral motivation, in which the love of fame or distinction is accorded a more positive role than in *Political Justice*. The emphasis on stimulating the imagination, rather than on reason, and the suggestion that the imagination is the basis for moral sense can be recognized in both *The Enquirer* and the later *Letter of Advice* (1818). Moreover, *Thoughts on Man* (1831) contains arguments which seem wholly at odds with his earlier egalitarianism.

3 Anarchism and liberalism

There is, nonetheless, a consistent core to Godwin's work. His philosophical speculations are best understood as speculations, not foundations. What is foundational is his basic conviction that private judgment and public discussion are the essential means to the development of moral and political truth, and that we become fully human only by coming to live wholly in accord with the dictates of the understanding. To do less is to be less - less rational, less individual, less worthy and less human. Godwin's novels, and much of his other writing, explore how we come to be less than we might be; and it is in these accounts, and in the swingeing attack on monarchical and aristocratic regimes in *Political Justice*, that he takes to task the world of aristocratic power and privilege in which he lived. His religious writings perform much the same role with respect to the other central institution of imposture and fraud, the Christian Church. Indeed, the separation of his works into disciplinary categories misses the point that in nearly everything he wrote Godwin was concerned to further the moral and intellectual development of his fellow men and women, and to reveal to them the extent to which they have been misled and imposed upon by the political, social and religious institutions of their society.

Political Justice remains Godwin's most forceful expression of this position. Despite its eccentricities and occasional extremism, it is a work of considerable philosophical originality and power. Sixty years before John Stuart Mill's *On Liberty* (1859), Godwin synthesized an indirect and ideal utilitarianism with a conception of individual wellbeing and a rights-based doctrine of private judgment, which is in many respects more compelling

than Mill’s. Moreover, where Mill hedges liberty against the masses, Godwin shares the eighteenth-century view that the greatest threat is from government. His anarchism simply takes this distrust to its ultimate conclusion. Liberals have tended to dismiss Godwin as an extremist and utopian, yet by combining his commitment to utility as a principle of justice with an equal commitment to private judgment as an indefeasible adjunct principle of right, he offers us a radical vision set within impeccably liberal constraints.

See also: Anarchism

MARK PHILP

List of works

Godwin, W. (1993) *The Political and Philosophical Writings of William Godwin*, ed. M. Philp, London: Pickering & Chatto.(Volumes 1 and 2 contain Godwin’s political writings, from his first published work, a biography of Pitt the Elder, and a number of pamphlets from the 1780s, through to his defence of the radicals charged with Treason in 1794 and his replies to the critics - notably Samuel Parr and Thomas Malthus. Volume 3 reprints the 1793 edition of *An Enquiry Concerning Political Justice* (Godwin’s major philosophical work); while volume 4 contains variant material from the manuscript and from the second edition of 1796 and the third edition of 1798, together with Godwin’s manuscript comments on the development of his views. Volume 5 contains Godwin’s essays and educational writings, notably *The Enquirer* which develops less formally many of the themes of his *Political Justice*; volume 6 his *Thoughts on Man*, his last major philosophical work, and volume 7 his religious writings which range from Sermons first delivered in the early 1780s through to a substantial but uncompleted manuscript denouncing Christianity which was published posthumously. (The extended introduction to volume 1 of this contemporary series provides an overview of Godwin’s central contribution.))

Godwin, W. (1992) *The Collected Novels and Memoirs of William Godwin*, ed. M. Philp, 8 vols, London: Pickering & Chatto.(The extended introduction to volume 1 of this contemporary series provides an overview of Godwin’s central contribution.)

Godwin, W. (1793) *An Enquiry Concerning Political Justice*, London: Robinson; revised 1796, 1798.(Godwin’s major philosophical treatise in which he develops a philosophical anarchism which looks forward to the increasing perfection of the human race through the development of mind.)

Godwin, W. (1794) *Things as They Are: Or the Adventures of Caleb Williams*, London: Robinson(Godwin’s most successful novel. A tale of domestic despotism and of flight and pursuit which makes original use of a flawed first person narrator.)

Godwin, W. (1797) *The Enquirer: Reflections on Education, Manners and Literature*, London: Robinson.(A collection of diverse essays in which some of the themes of his *Enquiry Concerning Political Justice* are analysed less formally, resulting in some modification of his philosophical position with respect to equality.)

Godwin, W. (1798) *Memoirs of the Author of a Vindication of the Rights of Woman*, London: J. Johnson; in W. Goodwin (1992), vol. 1.(Godwin’s affectionate and candid memoir of his wife, Mary Wollstonecraft.)

Godwin, W. (1818) *Letter of Advice to a Young American*, London: M.J. Godwin; in W. Goodwin (1993), vol. 5. (A synopsis of Godwin’s beliefs about the appropriate philosophical education.)

Godwin, W. (1831) *Thoughts on Man*, London: Effingham Wilson; in W. Goodwin (1992), vol.6.(Godwin’s last major philosophical work, which largely confirms his continuing commitment to the central doctrines of his *Enquiry Concerning Political Justice*.)

References and further reading

Barry, B. (1955) *Justice as Impartiality*, Oxford: Clarendon Press.(A defence of impartialist ethics which discusses Godwin’s relevance to our understanding of the character of impartiality in chapter 9.)

Clark, J.P. (1977) *The Philosophical Anarchism of William Godwin*, Princeton, NJ: Princeton University Press.(A sophisticated reading of Godwin’s political philosophy, which regards him as a utilitarian.)

Locke, D. (1980) *A Fantasy of Reason: The Life and Thought of William Godwin*, London: Routledge & Kegan Paul.(A major biography with substantial philosophical commentary.)

Mill, J.S. (1859) *On Liberty*, London: J.W. Parker and Son.(A seminal defence of liberty in all its forms in which the only justification for constraint is the avoidance of harm to others.)

Philp, M. (1986) *Godwin’s Political Justice*, London: Duckworth.(A contextual study of Godwin’s work 1789-1800 which advances a perfectionist interpretation of his moral philosophy.)

Goethe, Johann Wolfgang von (1749-1832)

*Goethe was a statesman, scientist, amateur artist, theatrical impresario, dramatist, novelist and Germany's supreme lyric poet; indeed he provided the Romantic generation which followed him with their conception of what a poet should be. His works, diaries and about 12,000 letters run to nearly 150 volumes. His drama Faust (1790-1832) is the greatest long poem in modern European literature and made the legend of Dr Faust a modern myth. He knew most of the significant figures in the philosophical movement of German Idealism (though he never met Kant), but he was not himself a philosopher. His literary works certainly addressed contemporary philosophical concerns: Iphigenie auf Tauris (Iphigenia in Tauris) (1779-86) seems a prophetic dramatization of the ethical and religious autonomy Kant was to proclaim from 1785; in his novel Die Wahlverwandtschaften (The Elective Affinities) (1809) a mysterious natural or supernatural world of chemistry, magnetism or Fate, such as '*Naturphilosophie*' envisaged, seems to underlie and perhaps determine a human story of spiritual adultery; in Faust, particularly Part Two, the tale of a pact or wager with the Devil seems to develop into a survey of world cultural history, which has been held to have overtones of Schelling,Hegel or even Marx. But whatever their conceptual materials, Goethe's literary works require literary rather than philosophical analysis. There are, however, certain discrete concepts prominent in his scientific work, or in the expressions of his 'wisdom' - maxims, essays, autobiographies, letters and conversations - with which Goethe's name is particularly associated and which are capable of being separately discussed. Notable among these are: Nature and metamorphosis (*Bildung*), polarity and 'intensification' (*Steigerung*), the 'primal phenomena' (*Urphänomene*), 'the daemonic' (*das Dämonische*) and renunciation (*Entsagung*).*

1 Life and works

Johann Wolfgang Goethe was born on 28 August 1749 in Frankfurt am Main, then a self-governing free city within the Holy Roman Empire. His father lived off his capital and his mother was the daughter of the city's principal official. He studied law in Leipzig, where he met J.C. Gottsched, the literary reformer and interpreter of Leibniz, and then in Strasbourg, where he met J.G. Herder, who opened his eyes to the merits of folk poetry andShakespeare. Goethe's first great literary success was the Sentimentalist novel in letters, *Die Leiden des jungen Werthers (The Sorrows of Young Werther)* (1774). Its story of a young intellectual who falls in love with a married woman and commits suicide made him a European celebrity. The mid-1770s were a period of great fertility for Goethe, and in about 1773 he began *Faust*, the work which was to occupy him throughout the rest of his life. In 1775 he accepted an invitation from the young Duke of Weimar, a sovereign prince, to visit his court and then to become a Privy Councillor of the duchy. Goethe did not return to Frankfurt and remained permanently in the Weimar administration. At first he had a heavy load of official work and his responsibility for the (unproductive) Weimar silver mines led him to an interest in geology and other branches of natural history. In 1782 he was ennobled, becoming 'von Goethe', but his literary powers seemed to be languishing. However, after he had spent a sabbatical period from 1786-8 in Italy (where he met the aesthetician K.P. Moritz), the Duke allowed him to concentrate on writing and on managing the Weimar theatre. A fragmentary version of *Faust* was published in the first authorized edition of his collected writings (1787-90), which also contained some other noteworthy plays, such as *Iphigenie auf Tauris (Iphigenia in Tauris)*, *Egmont* and *Torquato Tasso* (believed to be the first tragedy with a poet as its hero), and the first collection of his lyric poems. In 1790 he published a brief general theory of botany, *Versuch, die Metamorphose der Pflanzen zu erklären (Essay in Elucidation of the Metamorphosis of Plants)*.

Already in 1776 Goethe had arranged for Herder to come to Weimar as the head of the Lutheran Church in the duchy, and in 1789 he recommended Friedrich Schiller for a chair in history at the local university of Jena. After his return from Italy, Goethe became increasingly involved in the affairs of the university, paying particular attention to the appointments in the natural sciences and in philosophy, where he followed a policy of encouraging young talent and especially the new Kantian school (already established in Jena by Reinhold). He had some part in the successive appointments of Fichte, Schelling and Hegel, and after he had established a close friendship in 1794 with Schiller, who was just completing the major aesthetic essays of his Kantian period and with whom he undertook various joint editorial ventures, Goethe spent as much time in the university world of Jena and in the circle of Romantic writers centred on F. Schlegel as in the court world of Weimar. During this period he published

a long novel, *Wilhelm Meisters Lehrjahre (Wilhelm Meister's Years of Apprenticeship)*, which had been in the making for twenty years, and his interests in natural history were displaced by a passion for the theory of colour, of which he had become convinced Newtonian optics gave a perversely false account (see Colour and qualia). His reactions to the French Revolution, which threatened to overturn the world he had grown up in, were expressed in the play *Die Natürliche Tochter (The Natural Daughter)* (1803). From 1799 Schiller's mature plays provided the Weimar theatre with the centrepiece of a programme for raising theatrical and literary standards throughout Germany, but the German Romantic movement was already well under way and Goethe's Italianate and Hellenizing tastes already seemed old-fashioned. In 1805 Schiller died, and the following year Napoleon put an end to the old German order and the Holy Roman Empire itself at the battle of Jena. The university of Jena had already begun to decline with the controversial departure of Fichte and then Schelling, and after a period of closure could not recover its position as a centre of philosophical innovation.

Goethe marked the new epoch in his life that began in 1806 by marrying Christiane Vulpius, with whom he had lived since 1788 and who had borne him several children, of whom only one survived infancy. (Having ceased to be a communicant Christian at the age of 21, Goethe gave his refusal to undergo a church ceremony as his reason for postponing marriage so long.) 1806 also saw the completion of *Faust. Part One*, though publication was delayed until 1808. *Die Wahlverwandtschaften (The Elective Affinities)* (1809) was a kind of reckoning with a decade of Romantic culture, and in 1810, after a long gestation, *Zur Farbenlehre (On the Theory of Colour)* appeared. Much of the next ten years Goethe devoted to autobiographical writings, notably *Dichtung und Wahrheit (Poetry and Truth)* and *Italienische Reise (Italian Journey)*, but from 1814 to 1817 there came a remarkable resurgence of lyric poetry, in imitation of the medieval Persian poet Hafiz, *West-östlicher Divan (The West-Eastern Divan)*. A more sober atmosphere prevailed after the death of Christiane in 1816, and Goethe began to publish, rewrite and reinterpret his earlier studies in natural science. He remained hostile to the Christianizing tendencies in Romantic literature and art, resisted Schelling's proposed election to a chair in philosophy in Jena in 1816, since Schelling wished to combine it with a chair in theology, and favoured instead the appointment of Fries. Schopenhauer and his mother, who had settled in Weimar, were both close friends at this time.

For twenty years Goethe was a regular visitor to the Bohemian spas. At the age of 74, when in Marienbad, he proposed marriage to Ulrike von Levetzow, then 19 years old, and on being refused by her family gave up travelling and spent the last nine years of his life in Weimar preparing the final collected edition of his works. A stream of visitors - including Victor Cousin and Hegel - kept him in touch with European culture, and his list of correspondents - Byron and Carlyle among them - became immense. His last works, in which the theme of renunciation is increasingly prominent, show a concern with problems of world-historical dimensions: the nature of education and culture, the relation of America and Europe, the impending industrial revolution, 'world-literature' (*Weltliteratur*, a term he invented) and the preservation of humane values in a changing world. In addition to rounding off his autobiographical works he wrote a sequel to his major novel, *Wilhelm Meisters Wanderjahre (Wilhelm Meister's Years of Wandering)*, and in his last five years wrote most of *Faust. Part Two*, which was not published until after his death. At his own suggestion, many aphoristic remarks scattered among his papers were collected together as his *Maximen und Reflexionen (Maxims and Reflections)* and published posthumously by his amanuensis J.P. Eckermann, whose *Gespräche mit Goethe in den letzten Jahren seines Lebens (Conversations with Goethe in the Last Years of his Life)* were described by Nietzsche as 'the best German book there is'.

2 Goethe and the philosophers

The philosophical orthodoxy prevailing at all German universities in Goethe's youth was the Wolffian version of Leibnizianism, and throughout his life it is possible to find in his literary and scientific works characteristically Leibnizian traits: in a conversation of 1813 he expressed a preference for 'the Leibnizian term' monad as a description of 'the ultimate primal components of all beings' (see Leibniz, G.W. §§4,5). None the less it was long fashionable to regard Spinoza as the main philosophical influence onGoethe, despite Goethe's own admission thatSpinoza - whom he found 'abstruse' - had an effect on him like Shakespeare and Linnaeus, that is, he showed him the way he could not himself follow. A fragmentary essay of 1784-5, dubbed by its editors 'Studie nach Spinoza' ('Spinozan Study') has little in it that is certainly Spinozan and much that is probably Leibnizian.

Asked by Eckermann in 1827 who was pre-eminent among the modern philosophers, Goethe replied

'Kant…without any doubt'. He first sought help from Reinhold in studying the new system in 1789, and made an attempt at reading the first *Critique*. In 1790, immediately after its publication, he made a careful study of the *Critique of Judgment*, which he ever after regarded as Kant's most important work, particularly the 'Critique of Teleological Judgment' (see Kant, I. §12). In 1794, when his closer acquaintance with Schiller was only beginning, he read Fichte's first systematic writings and could summarize them to their author's entire satisfaction (see Fichte, J.G.). Goethe was aware of, and to a great extent accepted, the Copernican revolution which founded the Idealist period of German philosophy - no one in the world of learning, he wrote in 1805, could ignore it with impunity.

From 1797, after Schiller had turned back to drama, Schelling and F.I. Niethammer were Goethe's main informants about contemporary philosophical developments, and for a while it seemed to Goethe that Schelling might be the man to heal the division between mind and nature which the new philosophy had opened up. Despite some admiration for J.W. Ritter, however, Goethe's attitude to '*Naturphilosophie*' was consistently sceptical. He had less personal contact with Hegel than with Schelling, but had a high opinion of his character and abilities. Given the rapidly changing nature of scientific understanding, though, he doubted whether it was prudent for Hegel to incorporate so much empirical material into his philosophy, and objected to Hegel's continual recourse to the categories of the Christian religion (see Hegel, G.W.F. §8). Goethe thought world history 'the most absurd thing there is', but his relations with Hegel in the last years of his life were genial.

3 Some Goethean ideas

A poem Goethe wrote in 1826 ends:

> What greater gain can a human being have from life than that God-or-Nature should reveal herself to him? - how she dissolves firm material into spirit; how she firmly preserves what spirit has produced.
>
> (1988: 1, 367)

Goethe early became familiar with the Spinozan tag '*Deus sive natura*', and throughout his life 'Nature' was to him an expression of what is most perfect and most worthy of veneration. The content of the concept changed considerably: at first, in Rousseauist manner, it was an antithesis to the 'artificial' and was referred to an almost personal, creative and arbitrary power; later, as Goethe's scientific studies began, Nature was distinguished from the human world by being 'unfeeling', an impersonal repository of passionless order, contrasted with the turmoil of human emotions; later still, once Goethe had absorbed the implications of the Kantian revolution, he regarded the works of the human mind, which to his contemporaries seemed to transcend Nature, as, rather, her supreme product. In so far as this is the position of what Fichte called a 'realist', Goethe could be said never to have deviated from a Kantian belief in the independent reality of things in themselves, though, as he remarks in the essay 'Einwirkung der neueren Philosophie' ('Influence of Modern Philosophy'), his views appeared to Kant's disciples to be 'an analogy of Kantian conceptions, but a strange one'.

The most important single concept in Goethe's understanding of natural processes was that of 'metamorphosis'. (The term 'formation', '*Bildung*', probably borrowed from J.F. Blumenbach, is largely synonymous, but more indeterminate and multivalent.) Early in his botanical studies Goethe gave up the notion that all plants could be related to an ideal type, or 'primal plant' ('*Urpflanze*'), and substituted for it the notion that all vegetable shapes could be understood as the consequences of an ordered but unending series of transformations of a single fundamental organ, called 'the leaf'. An attempt to transfer this concept to vertebrate zoology was abandoned in 1795, under the influence of Kantian criticism, but Goethe's name remained associated with 'metamorphosis' - seen as an early form of evolutionary theory - in nineteenth-century biological textbooks. In Goethe's later years this (Leibnizian) conception of transformation in accordance with an innate rule was associated with two concepts for which he was particularly, though not wholly, indebted to Schelling: 'polarity' (*Polarität*) and 'intensification' (*Steigerung*). Not only is polar opposition held to be fundamental to the natural order (in colour theory, for example, the opposition of blue and yellow), but through 'intensification' opposites have a tendency to transform themselves jointly into a third term on a higher level (in colour theory: red).

However, at least from the time when he began reading Kant, Goethe thought of science not simply as speculation about the order of things, but as the attempt to understand human experience. Natural science was not concerned with an unobservable, mathematically constructed substratum to things (as Newton's optics seemed to imply), but

with the material offered to us in and by our senses. The theories of natural science simply sought to make explicit an order already implicitly present in our observations. Experiments therefore were not devices for confirming or disproving hypotheses, but were the very material of science, and should consist of a large series of observations arranged so as to manifest all aspects of the phenomenon concerned. The ultimate constituents of our knowledge (never completely isolated in practice) are therefore 'primal phenomena' (*Urphänomene*), which are utterly simple manifestations to our senses of what is, and of which no explanation is possible (for any explanation would take us away from our senses into a merely imagined world). An example of the approach to such a 'pure' or 'primal' phenomenon would be the revelation of colour when a prism is held up against a boundary between light and dark.

Writing to Hegel, Goethe called the 'primal phenomena' 'daemonic' (*dämonisch*), which he later glossed as meaning 'which cannot be resolved by the understanding or the reason'. In the Kantian terms of this definition, the 'daemonic' is identical, therefore, with things as they are in themselves. The word, however, which has been the subject of much scholarly controversy, is also applied to personal individuality, especially when strongly marked, and to the operations of chance, of an unfathomable providence (benign or malign) and of some animals (notably, bats and parrots). Although therefore Goethe might at first sight seem to share Hegel's principle that the noumenon is present to us in our experience, he also clearly refuses to identify the noumenon with our own spirit and insists that it is something *given*: human experience - and so human history - is not to be interpreted simply as the realm of the rational. The achievements of Hellenic civilization, for example, are extraordinary gifts of good fortune, not part of some inevitable process - not least because that might imply that their loss was inevitable too. But for Goethe the combination of ethical and sensual perfection in the Greek world can always in principle be recovered in uncovenanted moments of fulfilment, of which works of art are the principal monuments. Our normal lot, of course, is to do without such fulfilment, and to suffer our deprivation, and the absurd disorder of most of what is called world-history, in a spirit of renunciation (*Entsagung*). But that is the mirror-image of our tacit but never abandoned hope that the perfect order will one day be restored.

See also: German idealism; Nature, aesthetic appreciation of; Naturphilosophie; Poetry; Romanticism, German

NICHOLAS BOYLE

List of works

Goethe's main works have been mentioned in the text, and most, it will be plain, are at least indirectly relevant to philosophy. The following is a selection of non- or semi-fictional works of more directly philosophical interest (the titles are often editorial, and may be misleading).

Goethe, J.W. von (1887-1919) *Goethes Werke, herausgegeben im Auftrage der Großherzogin Sophie von Sachsen (Weimarer Ausgabe)*, Weimar: Hermann Böhlaus Nachfolger.(Other, more fully annotated, editions exist, or are in progress, but this is, and is likely to remain, the only historical-critical edition. It is available on CD-ROM as *Goethes Werke auf CD-ROM*, Cambridge: Chadwyck-Healey, 1995, supplemented by Goethe's conversations and the letters not included in the edition of 1887-1919.)

Goethe, J.W. von (1988) *Johann Wolfgang von Goethe. Werke. Hamburger Ausgabe in 14 Bänden*, Munich: C.H. Beck, 1988.(A popular students' edition, with commentary. Volume 14 has a comprehensive bibliography; for Goethe and philosophy, see 618-19. Four volumes of selected letters, and two of letters to Goethe, are uniform with the works.)

Goethe, J.W. von (1983-9) *Goethe: Collected Works*, Cambridge, MA: Suhrkamp Publishers Inc., Suhrkamp Edition, 12 vols.(An extensive collection of modern translations, with some notes, and including two volumes of essays, maxims and scientific writing.)

Goethe, J.W. von (1772) 'Brief des Pastors zu *** an den neuen Pastor zu Čast;Čast;Čast;' ('Letter from the pastor of *** to the new pastor of Čast;Čast;Čast;'), in Goethe (1988), vol. 12, 228-39.(Theology of toleration.)

Goethe, J.W. von (1784) 'Über den Granit' ('On Granite') in Goethe (1983-9), vol. 12, 131-4.(Goethe's geological study.)

Goethe, J.W. von (1784-5) 'Studie nach Spinoza' ('A Study Based on Spinoza'), in Goethe (1983-9), vol. 12, 8-10.(Goethe's theory of knowledge.)

Goethe, J.W. von (1790a) *Versuch, die Metamorphose der Pflanzen zu erklären (Essay in Elucidation of the Metamorphosis of Plants)*, in Goethe (1983-89), vol. 12, 76-97. (Goethe's principal work on botany.)

Goethe, J.W. von (1790b) 'Versuch einer allgemeinen Vergleichungslehre' ('Toward a General Comparative Theory'), in Goethe (1983-9), vol.12, 53-6.(His first attempt at a general morphology.)

Goethe, J.W. von (1792) 'Der Versuch als Vermittler von Objekt und Subjekt' ('The Experiment as Mediator

between Object and Subject'), in Goethe (1983-9), vol. 12, 11-17.(Titled in 1823; despite the title the thinking is almost pre-Kantian)

Goethe, J.W. von (1795) 'Plato als Mitgenosse einer christlichen Offenbarung' ('Plato as Party to a Christian Revelation'), in Goethe (1983-9), vol. 3, 200-3.(Against Christian exclusivism.)

Goethe, J.W. von (1798a) 'Einleitung in die Propyläen' ('Introduction to the *Propylaea*'), in Goethe (1983-9), vol. 3, 78-90.(Art theory in a historical context.)

Goethe, J.W. von (1798b) 'Erfahrung und Wissenschaft' ('Empirical Observation and Science'), in Goethe (1983-9), vol. 12, 24-5.(His theory of 'empirical' and 'pure' phenomena.)

Goethe, J.W. von (1799) 'Der Sammler und die Seinigen' (The Collector and his Circle), in Goethe (1983-9), vol. 3, 121-59.(Aesthetic theory in a fictionalized form.)

Goethe, J.W. von (1805) 'Letzte Kunstausstellung. 1805' ('Final Art Exhibition, 1805') in Goethe (1887-1919), part 1, vol. 36, 265-7.(Retrospect on Weimar art competitions.)

Goethe, J.W. von (1817) 'Geistesepochen' ('Stages of Man's Mind'), in Goethe (1983-9), vol. 3, 203-4.(Goethe's philosophy of history.)

Goethe, J.W. von (1817-20) 'Einwirkung der neueren Philosophie' ('The Influence of Modern Philosophy'), in Goethe (1983-9), vol. 12, 28-30.(Goethe's assessment of the influence on him of contemporary philosophers.)

Goethe, J.W. von (1827) 'Nachlese zu Aristoteles' Poetik' ('On interpreting Aristotle's *Poetics*'), in Goethe (1983-9), vol. 3, 197-9.(Goethe's interpretation of 'catharsis'.)

Goethe, J.W. von (1828) 'Erläuterung zu dem aphoristischen Aufsatz "Die Natur"' ('A Commentary on the Aphoristic Essay "Nature"'), in Goethe (1983-9), vol. 12, 6-7. (General theory of Nature - the essay 'Die Natur' ('Nature') is not by Goethe but by G.C. Tobler.)

References and further reading

Boyle, N. (1991) *Goethe: The Poet and the Age. Volume One: The Poetry of Desire (1749-1790)*, Oxford: Oxford University Press.(Includes the historical and philosophical background.)

Bruford, W.H. (1962) *Culture and Society in Classical Weimar: 1775-1806*, Cambridge: Cambridge University Press.(Good accounts ofHerder, Schiller andFichte, as well asGoethe.)

Goethe-Jahrbuch (1994) Weimar: Hermann Böhlaus Nachfolger.(Papers read in 1993 at the 73rd Congress of the International Goethe Society, which together constitute the most comprehensive study of Goethe's thinking about history.)

Heller, E. (1952) *The Disinherited Mind*, Cambridge: Bowes & Bowes.(Classic essays on Goethe's place in the modern German tradition.)

Molnár, G. von (1993) *Goethes Kantstudien*, Weimar: Hermann Böhlaus Nachfolger. (Facsimiles and analysis of all Goethe's marginalia to the First and Third *Critiques*.)

Nisbet, H.B. (1972) *Goethe and the Scientific Tradition*, London: Institute of Germanic Studies.(Particularly useful on Goethe's view of Bacon.)

Rabel, G. (1927) *Goethe und Kant*, Vienna: privately printed, 2 vols.(A very thorough, if slightly obsessional, compilation, regrettably under-used by later writers.)

Reed, T.J. (1984) *Goethe*, Oxford: Oxford University Press.(Excellent summary account of life and works.)

Wells, G.A. (1978) *Goethe and the Development of Science: 1750-1900*, Alphen aan den Rijn: Sijthoff & Noordhoff.(A Darwinian-rationalist critique, which is extremely well-informed, both about the science and about Goethe's writings.)

Good, theories of the

'Good' is the most general term of positive evaluation, used to recommend or express approval in a wide range of contexts. It indicates that a thing is desirable or worthy of choice, so that normally, if you have reason to want a certain kind of thing, you also have reason to prefer a good *thing of that kind.*

A theory of the good may consist in a general account of the good, which is meant to apply to all good things; or in a definition of 'good', an account of how the term functions in the language. Theories of the good have metaphysical implications about the relations of fact and value. Many ancient and medieval philosophers believed in the ultimate identity of the real and the good. Modern philosophers reject this identification, and have held a range of positions: realists, for example, hold that the good is part of reality, while certain moral sense theorists hold that when we call something good we are projecting human interests onto reality; and emotivists hold that we use the term 'good' only to signify subjective approval.

Theorists of the good also categorize different kinds of goodness and explain how they are related. Good things are standardly classified as ends, which are valued for their own sakes, or means, valued for the sake of the ends they promote. Some philosophers also divide them into intrinsic goods, which have their value in themselves, and extrinsic goods, which get their value from their relation to something else. Various theories have been held about the relation between these two distinctions - about whether an end must be something with intrinsic value. Philosophers also distinguish subjective goods - things which are good for someone in particular - from objective goods, which are good from everyone's point of view. Views about how these kinds of goodness are related have important implications for moral philosophy.

Usually, a theory of the good is constructed in the hope of shedding light on more substantive questions, such as what makes a person, an action, or a human life good. These questions raise issues about the relation between ethical and other values. For example, we may ask whether moral virtue is a special sort of goodness, or just the ordinary sort applied to persons. Or, since actions are valued as 'right' or 'wrong', we may ask how these values are related to the action's goodness or badness. We may also pose the question of whether a life that is good in the sense of being happy *must also be a* morally good *or virtuous life. This last question has occupied the attention of philosophers ever since Plato.*

1 History and metaphysics of the good

Almost anything may be assessed as 'good' or 'bad'. This ubiquity of 'good' and its cognates in other languages has suggested nearly opposite conclusions about the metaphysics of goodness to different philosophers. At one extreme we find Plato's view that the good is the fundamental principle of reality. Through works such as the *Republic*, Plato expounded his view that the reality of an object consists in its 'participation' in a 'Form'. A Form is both an archetype or pattern, and an ideal, a perfect version, of the things of which it is the Form. Each Form, because it is perfect, in turn participates in the Form of the Good. A thing is real, then, to the extent that it participates in the Form of the Good.

In *Nicomachean Ethics*, Aristotle criticizes Plato's account for not telling us anything about particular kinds of goodness (see Aristotle §§21-6). Yet Aristotle's own metaphysics retains a version of the equation of goodness and reality. The essential nature of each thing, according to Aristotle, consists in its characteristic activity or function, and it is both most perfect and most real when it performs that function well (see Perfectionism).

These metaphysical views may seem remote from our everyday employment of the idea of goodness. Yet most philosophers agree that the basic insight behind Aristotle's account throws important light on many uses of 'good'. To say that a thing is good is to say that it is a well-functioning thing of its kind, and its well-functioning is related to its reality: a good heart is one that pumps blood well, and a heart that ceases to pump blood altogether ceases to be a heart. This functional account of goodness applies most clearly to things that have purposes - instruments and tools, biological organs, parts of machines, crafts and professions - and it has the advantage of making it clear why we care about having things that are good. But when we say that happiness, or beauty, or freedom, is good, we do not seem to be talking about the performance of a function. Efforts have therefore been made to extend the basic idea of the functional account to things which are not clearly purposive, like people and lives. In *A Theory of*

Justice (1971), Rawls proposes that ‘a good x’ means ‘an x that has the properties it is rational to want in an x’. A good life, for example, has the properties it is rational to want in a life. The goodness of pleasure, freedom, or beauty may then be interpreted in terms of the role these things play in good lives.

The relinquishment of the view that the good coincides with the real marks the transition from the ancient and medieval to the modern world. Modern thinkers confront a value-neutral world, the world of matter and motion described by physics. Modern ‘realists’, although they reject the general equation of the real and the good, still believe that goodness is an objective property of certain objects (see Moral realism). Other modern philosophers believe that goodness is not a property that exists independently of the human mind, but rather some sort of projection or construction out of the needs, desires, and interests of human or sensate beings.

Philosophers of the seventeenth and eighteenth centuries distinguished between the ‘naturally good’ and the ‘morally good’. The naturally good is what is pleasant or desirable or makes us happy. Some early modern philosophers believed that the morally good is constructed out of the naturally good. For example, natural law theorists like Hobbes and Pufendorf thought that naturally good actions become morally obligatory when we are commanded to perform them by God or a sovereign (see Natural law), while sentimentalists, such as Hutcheson and Hume, held that naturally good dispositions - pleasant or useful character traits - are rendered moral virtues by the fact that we approve of them (see Moral sense theories). Rationalists such as Clarke and Price, on the other hand, believed that certain actions have a special kind of moral value which is independent of natural goodness, namely rightness.

By the late eighteenth century the distinction between natural and moral goodness began to blur, but from two opposed directions. Kant (1788) argued that moral goodness is the necessary condition of natural goodness. Happiness purchased by immoral action, for example, is not good at all (see Kantian ethics; Kant, I§§9-11). The utilitarians, by contrast, made natural goodness the source of all value, arguing that morally right actions are simply those that produce the maximum amount of the natural good, happiness or pleasure (see Utilitarianism).

The claim that happiness or pleasure *just is* the good puzzled Moore, who, in the early years of the twentieth century, pointed out that ‘good’ certainly does not *mean* ‘pleasant’ (see Moore, G.E. §1). Moore argued (1903) that any attempt to identify ‘good’ with a natural property is an instance of the ‘naturalistic fallacy’, and that we must therefore suppose that ‘good’ is a ‘non-natural’ property (see Naturalism in ethics §3). This attempt to establish value-realism on linguistic grounds set off a discussion of what the word ‘good’ means, or how it is used. Does it describe some property of objects, or is it used just to prescribe or recommend? The most extreme view is that of the emotivists, who stand in diametric opposition to Plato. Emotivists believe not only that goodness is not the fundamental principle of reality, but that strictly speaking the word ‘good’ does not refer to anything real at all. Its ubiquity, they think, can be explained only by the supposition that ‘good’ is used merely to express the speaker’s subjective approval, like a squeal of delight (see Emotivism).

2 Distinctions in goodness

The most obvious distinction in goodness is that between things which are valued as means, or instrumental goods, and things which are valued as ends, or final goods (see Values). This distinction is often confused with the distinction between intrinsic and extrinsic value. To say that something is intrinsically good is to say that it is good in virtue of what it is - in virtue of its own intrinsic nature; while to say that something is extrinsically good is to say that it is good in virtue of the relations in which it stands to things outside of itself. Means, for instance, are obviously extrinsically valuable, because their goodness springs from the fact that they promote other good things. Ends, by contrast, are often characterized as intrinsically good. In fact, however, it is an open question whether in order to be an end - to be valued or valuable for its own sake - an object must have intrinsic value.

Early twentieth-century moral philosophers debated this question. Empiricists argued that for something to be a final good is just for it to be desired for its own sake; philosophers in the idealist tradition, in contrast, believed that to be a final good is to be the object of a rational will. On both accounts, final goods are extrinsically valuable, deriving their value from the desires or volitions of human beings. But Moore argued that because ‘good’ does not mean ‘desired’ or ‘willed’, it is always an open question whether something desired or willed is good. Final goods, he thought, must therefore be intrinsically valuable. Moore suggested that in order to ascertain whether something is intrinsically good we should use a test of isolation: we consider whether the object has value apart from its

relations to other things. He claimed that when we used this method, we would discover that value belongs to complex states of affairs which he called ‘organic unities’, such as a person’s contemplating a beautiful object, or two friends enjoying each other’s company.

One of Moore’s aims in advancing this theory was to oppose hedonism, the view that pleasure is the good (see Hedonism). Yet the theory that final goods must be intrinsically valuable may push one towards hedonism, as we can see by applying Moore’s isolation test. A beautiful painting must surely be a good thing, yet if we imagine it existing in isolation from all viewers who might enjoy it, it seems to be without value. So we may decide that the value belongs instead to an organic unity consisting of someone’s enjoying the beautiful painting. This line of thought led Moore to conclude that human experiences, especially pleasant ones, are an element in most intrinsically valuable organic unities. But the same idea - that a thing’s goodness cannot be completely independent of its relation to human or sensate experiences and concerns - led utilitarians to suppose that pleasure must be the good.

The categories of value which Kant adopted suggest a different way of thinking about the relation between value and human concerns. Kant distinguished unconditional and conditional value, a distinction that resembles the intrinsic/extrinsic distinction. According to Kant (1785), a thing has unconditional goodness if we value it under any and all conditions, whereas it is conditionally good if the value we accord it depends on circumstances. The only thing we value unconditionally, he argued, is a good will, and human beings as the possessors of the capacity for good will. Yet the things people desire and care about do have conditional value - they are valuable because they matter to people, who have value. In this way, all values are related to human concerns.

If value must be related to human concerns, then we might think that anything that is good must be good *for* someone. Goodness for someone is sometimes characterized as ‘subjective’ or ‘agent-relative’ goodness, as opposed to ‘objective’ or ‘agent-neutral’ goodness, which pertains to everyone. If something is subjectively good for me, then I have reason to promote it and care about it; while if something is objectively good, then everyone has reason to promote it and care about it.

Twentieth-century philosophers have debated the question of the relation between these two kinds of goodness. Some think that goodness is inherently subjective, and people have reason to pursue common objects only when their interests happen to coincide. Others think that subjective values always give rise to objective ones, so that if it is (subjectively) good *for* me to have something, then it is objectively good *that* I should have it. Still others think that some subjective values - say, the ones associated with needs - give rise to objective ones, while others do not. And finally, at the other extreme, there are philosophers who think that subjective values are derived *from* objective ones. According to these philosophers, I cannot claim that something is good for me ‘because it makes me happy’, unless I consider my happiness to be, independently of my personal interest in it, an objectively good thing.

3 The goodness of people

When we call a person ‘good’ are we using ‘good’ in the ordinary sense? Sometimes it seems clear that we are. A person may be good *at* things - talented at sports or crafts, or master of an intellectual discipline. A person may also be good in certain roles - a good mother or teacher, for instance. Aristotle’s functional account of goodness, or something like it, seems to apply to these cases: a good teacher is good at carrying out the functions of a teacher, or has the qualities that it is rational to want in a teacher. But what about when we say that a person is good, just as a person?

In fact there are two different ways in which people are said, just as people, to have value. Ordinarily, when we say of a particular person that they are ‘a good person’, we mean that they are morally good or virtuous. Aristotle applied his functional account of goodness to moral virtue in a straightforward way. He identified reason as the human function, meaning that what is distinctive of human beings is the use of reason to govern our activities. Virtues are qualities that foster the good performance of this function. Both Plato and Aristotle compared moral virtue to health: it is a way of being in good psychological condition. The functional account has also been used in a more broadly social way to explain moral goodness. On this view you are morally good if you are good at the performance of all of your various social roles, or if you have the qualities it is rational for your friends and fellow citizens to want in a friend and fellow citizen; or, as in certain sentimentalist theories, if you are an object of moral

approval because you have these qualities (see Virtues and vices §§2-3).

Those accounts identify moral goodness with the possession of certain dispositions - character traits - that influence a person's conduct. Philosophers who favour such accounts usually suppose that your actions issue directly from your character. But other philosophers claim that human beings have a power distinct from and more directly related to actions than character traits are, namely the will (see Will, the). The will enables you to act freely, even to the point of doing what is 'out of character'; and it makes you responsible for your actions, even when they seem to be determined inevitably by your character. Kant, accordingly, recognized a form of moral goodness distinct from virtue or the goodness of character traits, namely goodness of the will or *moral worth*, which pertains to the well-functioning of the will itself, either as it is exercised in particular actions, or as a standing disposition. Kant famously claimed that you may achieve moral worth even if you have temperamental qualities which make it hard for you to do what duty demands (see Self-control). This claim raises important questions about how the will and character are related.

The second way in which people are considered to be valuable just as people, enshrined in many religious and philosophical systems, involves the thought that every human being as such has a fundamental value which it is wrong to deny or overlook. The religious view that we are all God's children, the political view that all human beings are created equal, and the Kantian moral view that every human being is an end-in-itself are all expressions of this idea. There are various views about what makes people valuable in this way - freedom of the will, rationality, consciousness, the possession of identifiable interests, the capacity for pleasure and pain, or simply life itself - and so what treatment is called for. Some of these options raise the question whether other living things should also be accorded such value (see Moral standing §§1-3). This kind of value is different from that which we attribute to particular people when we say they are morally good, for one need not be an especially good person to lay claim to the political rights or the moral respect due to every human being. But the two ideas are sometimes related by the thought that it is the capacity for moral goodness, or the capacity that makes us capable of moral goodness, that gives us this fundamental value (see Respect for persons).

4 The right and the good

Good is a ubiquitous term, applying to almost any sort of thing, but actions, policies and laws are also praised as being 'right' or 'just'. Right actions are those which are required by morality or, more extensively, those not forbidden by morality, not 'wrong'. So a question arises about the relation between these two kinds of value, the right and the good. Consequentialists think that the relation is simple: right actions are those which tend to maximize good results, so that rightness is actually a form of instrumental goodness (see Consequentialism). Yet we sometimes seem to care about doing the right thing independently of, or even in the teeth of, the consequences that it produces. We may decide that we will uphold someone's rights, or obey the law, or keep a promise, though we know the results will be bad, because, as we say, 'it is the principle of the thing'. Consequentialists think this attitude is either an acknowledgement of the especially important consequences that result from the observance of certain rules, or else a misguided form of rule-worship. But 'deontologists' believe there is something valuable about doing the right thing apart from the good results it may or may not produce (see Deontological ethics).

Since deontologists deny that rightness is merely an instrumental value, we might be tempted to say they think of rightness as either a final good or a special kind of intrinsic value, characteristic of actions. But this does not completely capture the deontological intuition. Final and intrinsic values, as they are often conceived, may be weighed in with other values, so on this view we might sometimes endure wrongdoing, as we sometimes endure pain, for the sake of the larger benefits it brings. But deontologists deny that such values as freedom, justice, or fidelity may be traded off for other goods. They also believe that the way in which right actions serve these values is not by producing or causing them, but instead is direct or constitutive. For example, a deontologist may think that it is right to keep a promise, not because fidelity to promises is thereby efficiently produced (for we can imagine circumstances in which one person's keeping a promise will induce others to break theirs), but because it is an act of fidelity. And we respect human rights, not because this produces freedom, but because this is what freedom consists in - living in a world in which human rights are respected.

The question of the relation between the right and the good therefore gives rise to deep questions about the relations between actions and the values they serve. But these questions belong more properly to a discussion of the right (see Right and good).

5 The good life

One of the oldest questions of moral philosophy is what the best life is for a human being. A standard view in ancient Greek philosophy was that there are three types of life: a contemplative or philosophical life; a life of virtuous political activity; and a hedonistic or money-making life. Plato and Aristotle agreed that the contemplative life is best and the political life second best; in their view, only those who do not know the true pleasures of contemplation and virtuous action resort to hedonistic pursuits (see Eudaimonia; Hedonism).

The idea that so specific a life can be identified as best may seem paradoxical. If there is a best life for human beings, does that mean it is the best life for any human being, regardless of personal endowments or natural tastes? On this view, lives are like looks: there may be a best way for each person to look, but the best way for you to look may not, unfortunately, be the best way to look. Some philosophers argue that this makes no sense: how can a certain way of life be better for you, if there is no way in which you could enjoy, appreciate, or be interested in such a life and still be *yourself*?

What sorts of arguments might be used to show that one type of life is best? Plato favoured a test of experience: we should take as authoritative the preferences of those people who have experienced the kinds of activities central to all three types of life. John Stuart Mill (1861) suggested that this test could be used to identify those pleasures whose ‘quality’ is so high as to outweigh considerations of ‘quantity,’ and so which belong in the best life (see Mill, J.S. §9). Aristotle appealed to his own idea of function: if human beings have a function, the person who performs the human function well must lead the best life. But Aristotle also thought we could identify certain criteria which any good life must meet, and rate lives by the extent to which they fulfil these criteria: the pursuits central to a good life must be active, pleasant, self-sufficient, and done for their own sake alone. Moore’s theory suggests a simpler account: the good life is one that consists of intrinsically valuable states and activities, such as appreciating beauty, having friends, and seeking knowledge.

Perhaps the most common strategy is to appeal to human psychology, to what people actually care about. Classical utilitarians argued that human beings care about only two things: getting pleasure and avoiding pain. The best life must therefore be the one with the greatest balance of pleasure over pain. Others claim that the goodness of your life is a function of how many of your desires are satisfied and the strength of those desires. But the content of your desires and their strength may be determined in unfortunate ways by the limitations of your knowledge or imagination, and this has led some philosophers to adopt an idealized version of this account: the good life for you is the one you *would* choose under conditions of perfect knowledge and imaginative reflection.

But we must also ask whether the *good* life and the *happy* life are the same. The early modern philosophers’ distinction between moral and natural goodness brought this question sharply into focus. They thought of happiness as a natural good and many of them believed that reason demands the pursuit of happiness just as obviously as it demands the practice of virtue. Yet virtue does not always bring happiness. Are human beings then subject to conflicting demands of reason? Ancient Greek philosophers had raised a parallel question, whether being virtuous is a good thing for the virtuous person. But ancient and modern solutions are different in an important way. For the Greeks, the answer lay in demonstrating that the qualities we ordinarily regard as moral virtues really are qualities that make us good at the performance of the human function - qualities without which we would be incapable of choosing and acting well. Once this is established, it is evident that a virtuous person will necessarily have a better life. Modern philosophers, however, are more inclined to believe that happiness and virtue are independent. Many modern philosophers have therefore tried to produce what Rawls calls ‘congruence’ arguments: arguments that show that the pursuit of virtue will also bring happiness, and so that the two kinds of good, although independent, come together in practice. But others have drawn a more austere conclusion, namely that a life which is both morally good and happy is open to us only in favourable circumstances - circumstances which must be secured by divine arrangements or, more optimistically, by political action.

See also: Evil; Happiness; Practical reason and ethics; Right and good; Xunzi

CHRISTINE M. KORSGAARD

References and further reading

Anderson, E. (1993) *Value in Ethics and Economics*, Cambridge, MA: Harvard University Press.(Categorizes different ways in which we value things and examines the moral and other consequences of these categorizations; especially concerned to show that not every form of value may be handled properly by the mechanisms of the market.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, in *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984, esp. books I 6-7, X.(Criticizes Plato's lack of guidance on particular kinds of goodness, gives a functional account of goodness, and discusses the three kinds of life: contemplative, political and hedonistic.)

Aristotle (*c*. mid 4th century BC) *Metaphysics*, in *The Complete Works of Aristotle*. ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984, esp. books VII-X, XII. (Outlines the connections between goodness, function and reality.)

Ayer, A.J. (1936) *Language, Truth, and Logic*, London: Gollancz; 2nd edn, 1946, ch. VI.(The founding statement of emotivism.)

Butler, J. (1726) *Fifteen Sermons Preached at the Rolls Chapel*, Sermons I, II, III, XI, XII; repr. in S. Darwall (ed.) *Five Sermons Preached at the Rolls Chapel and A Dissertation Upon the Nature of Virtue*, Indianapolis, IN: Hackett Publishing Company, 1983, Sermons XI, XII.(The most influential 'congruence' argument of the eighteenth century, to the effect that self-love favours the practice of virtue; later adopted by David Hume and others.)

Foot, P. (1961) 'Goodness and Choice', *Proceedings of the Aristotelian Society*, supplementary vol.; repr. in *Virtues and Vices and Other Essays in Moral Philosophy*, Berkeley, CA: University of California Press, 1978. (An influential twentieth-century presentation of the functional account of goodness.)

Griffin, J. (1986) *Well-Being*, Oxford: Clarendon Press.(One of the most complete presentations of the view that the good is what we would desire under ideal conditions of knowledge and reflection.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. and ed. M. Gregor, *Groundwork of the Metaphysics of Morals*, Cambridge: Cambridge University Press, 1997.(Introduces the ideas of unconditional value and of goodness of will; argues for the unconditional value of humanity and the good will.)

Kant, I. (1788) *Critik der practischen Vernunft*, trans. L.W. Beck, *Critique of Practical Reason*, Upper Saddle River, NJ: Prentice Hall, 1993.(Part 1, book I, ch. 2 deals with the relation between moral and natural goodness; part 1, book II deals with the relation between virtue and happiness.)

Korsgaard, C.M. (1996) *Creating the Kingdom of Ends*, New York: Cambridge University Press.(Chaps 4, 8 and 9 concern the concept of unconditional value; ch. 10 discusses various theories of the relation between subjective and objective value.)

Lewis, C.I. (1946) *An Analysis of Knowledge and Valuation*, La Salle, IL: Open Court.(Critical exploration of the concept of intrinsic value; introduces a new category, inherent value, in response to problems with the former category.)

Mill, J.S. (1861) *Utilitarianism*, ed. G. Sher, Indianapolis, IN: Hackett Publishing Company, 1979, chaps II, IV. (Uses Plato's test of experience to identify the 'higher' pleasures, and provides a 'proof' that pleasure and the avoidance of pain are the only desirable things.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press, 1971.(Argues that goodness must be an intrinsic non-natural property, and explains the idea of 'organic unities'.)

Moore, G.E. (1912) *Ethics*, Oxford: Oxford University Press, chaps IV, VII.(Presents Moore's theory of the good in the context of a theory about right and wrong.)

Moore, G.E. (1922) 'The Conception of Intrinsic Value', in *Philosophical Studies*, London: Kegan Paul.(Argues that final goods must be intrinsically valuable.)

Nagel, T. (1970) *The Possibility of Altruism*, Oxford: Clarendon Press, reprinted Princeton, NJ: Princeton University Press, 1978.(The most important examination of the relation between subjective and objective value written in the twentieth century, in which it is argued that all subjective values have objective correlates.)

Nagel, T. (1986) *The View From Nowhere*, Oxford: Oxford University Press, chaps VIII, IX.(Modifying his earlier position, this work argues that only some subjective values have objective correlates.)

Perry, R.B. (1926) *General Theory of Value: Its Meaning and Basic Principles Construed in Terms of Interest*, Cambridge, MA: Harvard University Press.(An influential work in twentieth-century American value theory, proposing that to be valuable means to be the object of any interest.)

Plato (*c*.380-367 BC) *Republic*, trans. G.M.A. Grube, revised by C.D.C. Reeve, Indianapolis, IN: Hackett Publishing Company, 1992.(Presents the Theory of Forms and an account of the fundamental role of the Form of the Good.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, part III.(Ch. 7 argues that for a

thing to be good is to have the properties it is rational to want in a thing of that kind; ch. 9 presents an argument for the ‘congruence’ of happiness and justice.)
Stevenson, C.L. (1944) *Ethics and Language*, New Haven, CT: Yale University Press.(The most complete and systematic development of the emotivist theory of evaluative language.)

Goodman, Nelson (1906-)

Nelson Goodman is an American philosopher who has written important works in metaphysics, aesthetics and epistemology. Throughout his work runs a concern with the ways that the symbols we construct inform the facts that we find and structure our understanding of them. Different symbol systems yield irreconcilable structures. So there is no one way things really are. There are, he concludes, many worlds if any. Moreover, worlds are made rather than found, for the categories we construct fix the criteria of identity for the individuals and kinds we recognize. Thus they determine what objects and kinds constitute a world.

Goodman argues that the arts as well as the sciences make and reveal worlds. Aesthetics as he construes it is a branch of epistemology. He analyses a variety of modes of symbolization, literal and metaphorical, and shows how they contribute in the arts and elsewhere to the advancement of understanding.

Goodman's 'new riddle of *induction' reveals that the problem of induction runs deeper than philosophers had thought. He defines the predicate 'grue' as 'examined before future time* t *and found to be green or not so examined and blue.' All emeralds examined to date have been both green and grue. What justifies our expecting future emeralds to be green rather than grue? Inductive validity, the new riddle shows, turns not only on the constitution of an evidence class, but also on its characterization. The question then is what favours one characterization over its rivals. The fact that 'green' has been used far more often than 'grue' in induction, Goodman contends, provides the answer - not because it increases our odds of being right, but because of its pragmatic advantages.*

1 Metaphysics

Nelson Goodman has made groundbreaking contributions to aesthetics, epistemology, metaphysics and the philosophy of science. In his youth he ran an art gallery. Throughout his life he has been an avid and eclectic collector of art. He founded the Harvard Summer Dance programme and Project Zero, an ongoing research programme in arts education at the Harvard Graduate School of Education. As graduate students at Harvard, he and Henry Leonard developed a version of mereology that they dubbed 'the calculus of individuals'. Elaborated in *The Structure of Appearance*, it forms the basis for Goodman's nominalism. As Goodman construes it, the difference between mereology and set theory lies in the sorts of constructions they permit. Set theory admits infinitely many distinct entities - sets of sets of sets of sets … - all composed of the same basic elements. Mereology holds that the same basic elements are parts of but a single whole. Goodman's nominalism consists in a refusal to recognize more than one entity comprised of the same basic elements. It says nothing about the metaphysical constitution of the elements. The decision of whether to countenance abstract or concrete, immaterial or material, mental or physical entities requires more than a commitment to nominalism.

The constructionalism Goodman espouses in *The Structure of Appearance* is a methodology for systematizing a body of pretheoretical beliefs. Such beliefs tend to be inchoate, vague, even inconsistent. By devising an interpreted formal system that derives them from or explicates them in terms of a suitable base of primitives, we can bring them into logical contact with each other, eliminate inconsistencies and disclose unanticipated logical and theoretical connections. Were such a system required to reflect all our relevant pretheoretical beliefs exactly, it would simply replicate received errors and confusions. But, Goodman argues, the regimentation constructionalism countenances involves judicious correction, refinement and even repudiation of presystematic convictions in the interests of simplicity, coherence, theoretical tractability and the like.

Multiple, divergent systems do justice to the same range of pretheoretical beliefs. One system might identify a geometrical point with the intersection of two non-parallel lines. Another might identify it with the limit of a sequence of nested spheres. Neither invalidates the other, for each provides a geometrically acceptable definition of a point. Here lies the root of Goodman's relativism. Relative to each acceptable system, the constitution of a point is determinate. But absolutely and independently of the systems we construct, it is indeterminate.

In *Ways of Worldmaking*, Goodman provides a less formal treatment of the same themes. Worlds and the objects that comprise them are made, he contends, not found. The members of any group are alike in some respects, different in others. So mere inspection cannot settle whether two manifestations are of the same thing or of the

same kind. To answer such questions we require a category scheme or system of classification that distinguishes differences that matter from differences that do not. Such schemes are not dictated by nature, but are human constructs. We draw the lines. Lines can be drawn in various places, resulting in divergent but equally viable world versions. One person might count a newspaper with a new publisher and a radically revised editorial stance as the same newspaper it always was. Another might count it as a different newspaper. One might classify a black hole as a star, another as the residue of an extinguished star. Relative to its own world version, each is right. Relative to its rival's, each is wrong. Neither is absolutely right or wrong.

If all overlapping world versions were reducible to or supervenient on a single base, such divergences would be insignificant. But, Goodman insists, such is not the case. In *The Structure of Appearance* he develops a phenomenalist constructional system whose primitives are qualia - phenomenal individuals out of which enduring perceptibles are constructed. He does not claim that this is the only viable form of phenomenalism and he recognizes that it neither underwrites nor reduces to a physicalist system. But it is none the worse for that. Rather, he urges that physicalist and phenomenalist systems are distinct, valid constructions of independent interest and importance. Neither is parasitic on the other.

If the proliferation derives from the availability of clashing category schemes, why not take it to show simply that there are multiple conceptualizations? Then we could retain our pretheoretical conviction that there exists exactly one world underlying them. Were the analytic/synthetic distinction tenable, such a strategy might work. But as Goodman, W.V. Quine, and Morton White demonstrated, it is not. Although statements depend on both meaning and fact, the dependence on meaning cannot be separated from the dependence on fact. There is thus no way to make sense of the claim that the difference between clashing world versions derives from different ways of conceptualizing the same facts. Category schemes dictate the criteria of identity of their objects, so mutually irreducible schemes do not treat of the same things. Since a world is the totality of things that comprise it, irreducible schemes define distinct worlds. There are, Goodman concludes, many worlds if any.

Still, it is not the case that just anything goes. Goodman describes his position as relativism under rigorous restraints. Consistency, coherence, suitability for a purpose, as well as accord with past practice and antecedent convictions are among the restraints he recognizes. Fitting and working are the marks of a successful version. A world version must consist of components that fit together; it must fit reasonably well with our considered judgments about the subject at hand, and must work to further our cognitive objectives. A version that is internally incoherent, is inconsonant with our antecedent considered judgments or impedes the advancement of understanding is unacceptable. Worldmaking need not be deliberate. *Ways of Worldmaking* discusses how, with only sparse cues, the visual system constructs the apparent motion it discerns. Nor is worldmaking exclusively the province of science - the arts as well as the sciences make worlds.

2 Aesthetics

Works of art, Goodman contends, are symbols. As such, they require interpretation. The aesthetic attitude, then, is not one of passive contemplation, but of active cognitive engagement; its main goal is understanding, not pleasure. Aesthetics, as Goodman conceives it, is a branch of epistemology.

Languages of Art characterizes a range of symbol systems used in the arts and elsewhere, and identifies the syntactic and semantic structures that give rise to their powers and limitations. The basic modes of reference are denotation and exemplification. Denotation is the relation of a name to its bearer, a predicate to the members of its extension, a picture to its subject. 'Charles I', 'the father of Charles II' and a Van Dyck portrait all denote Charles I. Fictional symbols lack denotata. Their significance, Goodman maintains, derives from symbols that denote them. Although the term 'unicorn' fails to denote, the terms 'unicorn-description' and 'unicorn-picture' denote a variety of symbols that collectively constitute the meaning of the term 'unicorn'. Many works of art - abstract art, most instrumental music and much dance - do not even purport to denote. They refer, Goodman maintains, by means of exemplification. In exemplification, a symbol points up and hence refers to features it serves as a sample or example of. Thus, a commercial paint sample exemplifies its colour and sheen; a late Mondrian painting exemplifies squareness. Exemplification is widespread not only in the arts but also in science, commerce and pedagogy - indeed wherever samples and examples are used.

Denotation and exemplification are not mutually exclusive. Denoting symbols in the arts typically also exemplify.

A portrait by Whistler denotes his mother and exemplifies a seemingly infinite spectrum of shades of grey. Seventeenth-century Dutch still lifes denote opulent arrangements of flowers and fruit and exemplify ambivalence about worldly success. Symbols, particularly aesthetic ones, often perform a variety of interanimating referential functions.

Symbols normally do not operate in isolation. They belong to schemes that collectively sort the objects in a realm. In metaphor a scheme that normally sorts one realm is imported to effect a re-sorting of another. New groupings emerge as items that belong to disjoint literal categories are classed together under a single metaphorical label and affinities between literal and metaphorical referents of a term are brought to light. In calling a plumber a virtuoso, for example, we import a scheme that literally sorts musicians to effect a sorting of craftsmen. New patterns and distinctions emerge as we recognize the delicacy, dexterity and skill displayed by some few members of the several crafts. Because they draw their lines where no literal label does, metaphors resist paraphrase. No literal label quite captures what all and only the virtuosi in the building trades have in common with one another or with literal virtuosi.

Figurative reference is real reference, Goodman maintains, and figurative truth is real truth. ‘Feline cunning’ genuinely if metaphorically denotes some scheming politicians. ‘The walls of the Alhambra are made of lace’ is genuinely if metaphorically true, and Michelangelo’s *Pietà* genuinely if metaphorically exemplifies incalculable sorrow.

Expression, he contends, is metaphorical exemplification by a work of art functioning as such. The *Pietà* then expresses the incalculable sorrow it metaphorically exemplifies. But expression is not limited to emotions. A work of art, functioning as such, expresses any feature it metaphorically exemplifies. Thus, for example, Brancusi’s *Bird in Space* expresses fluidity and flight. There is, evidently, no limit on the range of features that works of art can express.

Reference is not always direct. In indirect reference, one symbol refers to another by means of a chain consisting of denotational and exemplificational links. Rembrandt’s *Night Watch* alludes to the history of the militia it depicts by portraying its subjects wearing costumes that exemplify important periods of the militia’s history. In *Reconceptions* Goodman construes variation as a form of indirect reference, where a variation must in some respects resemble its theme and must differ from it in others. But every two passages do that, and not every passage is a variation on every other. A symbol functions as a variation on a theme when it uses the resemblances and differences as routes of reference to the theme. It exemplifies features it shares with the theme, contrastively exemplifies features it does not share with the theme, and effects reference back to the theme via both exemplification and contrastive exemplification. On Goodman’s analysis, variation is not restricted to music. Nor need a variation occur in the same work as the theme on which it is a variation.

Scientific symbols tend to be attenuated, symbolizing along comparatively few dimensions, whereas aesthetic symbols are relatively replete. For example, the same configuration might serve as an electrocardiogram or a line drawing. In an electrocardiogram, only its shape is significant, whereas in a line drawing, the exact colour and thickness of the line, the precise shade of the background, the size and shape of the paper, the position of the line on it and even the texture of the paper may be significant. The electrocardiogram is referentially austere denoting a pattern of heartbeats and perhaps exemplifying a certain symptomatology. The drawing, however, is apt to perform a variety of interanimating referential functions.

Whether the symbol is an electrocardiogram or a drawing depends on its function. It counts as a work of art as long as it functions as an aesthetic symbol. And it may do so intermittently. The crucial question for Goodman is not ‘What is art?’ but ‘When is art?’, that is, ‘Under what circumstances does an object function as an aesthetic symbol?’ He gives no criterion, but identifies five symptoms of the aesthetic: syntactic density, semantic density, relative repleteness, exemplification and complex and indirect reference. A symbol system is syntactically dense when the finest differences between signs constitute a difference between symbols, that is, when it can mark the finest differences between items in its domain. Symptoms, Goodman acknowledges, are neither necessary nor sufficient conditions. But they are good, if defeasible, indications of the presence of a condition.

Interpretation requires discovering what symbols constitute a work, how they function, what they refer to and what they achieve. The richness and complexity of aesthetic symbols means that the task may be interminable and that

multiple, irreconcilable interpretations may be correct. But not every interpretation is correct. Only those that make maximally good sense of the work's symbolic functions are to be accepted. Goodman's pluralism consists in his recognition that more than one interpretation can often do so.

3 Induction

Symbol systems are artefacts. Their syntactic and semantic features are not dictated by the domain but result from decisions about how the domain is to be organized, which can be done in an enormous number of divergent ways. Acceptability of a symbol system depends on its suitability for the purposes at hand.

In empirical science and elsewhere the purpose is often inductive - a matter of projecting from limited evidence to a wider class of cases. The traditional problem of induction is to say when and to what extent a limited evidence-class warrants such an inference. How many emeralds need be examined and from what sources need they be drawn before we are justified in inferring that all emeralds are green? In *Fact, Fiction, and Forecast*, Goodman shows that the problem runs deeper. It is a matter not just of the composition of the evidence class, but also of its characterization. He defines novel predicates 'grue' and 'bleen' as follows:

x is 'grue' = *x* is examined before future time *t* and is found to be green, or *x* is not so examined and is blue.
x is 'bleen' = *x* is examined before future time *t* and is found to be blue, or *x* is not so examined and is green.

All emeralds examined to date have been grue as well as green, for the extensions of the two predicates do not yet diverge. Yet we confidently expect emeralds found after *t* to be green, not grue. What, if anything, is our justification?

We can't dismiss 'grue' as derivative from 'green', for 'green' can be defined in terms of 'grue' and 'bleen', just as 'grue' is defined in terms of 'green' and 'blue':

x is 'green' = *x* is examined before future time *t* and found to be grue, or *x* is not so examined and is bleen.

Which predicate is basic and which is derivative depends entirely on where we start. Nor can we dismiss 'grue' on the grounds that it makes essential reference to a specific time *t*. For whether 'grue' or 'green' makes reference to *t* again depends on which is taken as primitive. One might argue that 'green' marks a more natural kind than 'grue' does. Then it is, in some sense, essentially more primitive. But in the absence of an acceptable standard of naturalness that does not presuppose the very differences in projectibility we are trying to account for, this claim rings hollow. For we neither know what it means, nor how to tell whether one predicate is more natural than another.

The solution, Goodman maintains, lies in entrenchment. What favours 'green' over 'grue' is the brute fact that 'green' and its cognates have been successfully projected far more often than 'grue'. The fact that up to now 'grue' would have worked as well is irrelevant. The decision favours the predicates that were actually successfully used.

Induction provides no guarantees. Goodman recognizes that we currently have no way of knowing whether future emeralds will be grue, green or something else entirely. The problem as he sees it is how to proceed in the absence of such knowledge. He argues that entrenched predicates are to be preferred, not because they have any lien on the future, but because they have served us well so far, and their continued use enables us to make efficient use of available cognitive resources and habits of thought. But the presumption in favour of entrenched predicates evaporates as soon as counterexamples emerge. When the first non-green emerald is found, 'All emeralds are green' loses its claim on our epistemic allegiance.

The emphasis on entrenchment does not preclude innovation. Novel predicates can be projected when entrenched hypotheses are violated. Thus, for example, the Michelson-Morley experiment, by violating Newtonian generalizations, opened the way for the projection of novel, relativistic predicates. New predicates can also be introduced at interstices where no entrenched predicate prevails. A term like 'quark' can be introduced to denote phenomena that previously lacked a label. Such terms, Goodman maintains, derive their projectibility from related terms such as 'subatomic particle'. Novel predicates thus become projectible by fitting into working inductive systems or into replacements for systems that do not work.

See also: Analyticity; Fictional entities §2; Induction, epistemic issues; Nominalism; Reference; Relativism

CATHERINE Z. ELGIN

List of works

Goodman, N. (1951) *The Structure of Appearance*, Boston, MA: Reidel, 3rd edn, 1977.(This work uses devices of formal logic to construct systems that solve or dissolve perennial problems in epistemology. It argues that the availability of multiple, divergent constructional systems is a virtue rather than a flaw.)
(1954) *Fact, Fiction, and Forecast*, Cambridge, MA: Harvard University Press, 4th edn, 1984.(This work poses the new riddle of induction, and proposes a solution to it. It argues that only by looking at past inductive practice do we have the resouces for precluding the induction of ‘All emeralds are grue’.)
(1968) *Languages of Art*, Indianapolis, IN: Hackett Publishing Company.(A development of a general theory of symbols and an application to the arts.)
(1972) *Problems and Projects*, Indianapolis, IN: Hackett Publishing Company.(A collection of papers on a variety of issues including art, meaning, induction and simplicity.)
(1978) *Ways of Worldmaking*, Indianapolis, IN: Hackett Publishing Company.(This work argues that worlds are made through the construction and deployment of symbol systems, and that there are many worlds if any.)
(1984) *Of Mind and Other Matters*, Cambridge, MA: Harvard University Press.(A collection of papers on metaphysics, aesthetics and the theory of reference.)
Goodman, N. and Elgin, C.Z. (1988) *Reconceptions in Philosophy and Other Arts and Sciences*, Indianapolis, IN: Hackett Publishing Company.(This book reconceives the ends and means of philosophy, taking understanding and rightness of symbols rather than knowledge and truth as central notions.)
Goodman, N. (1990) *A Study of Qualities*, New York: Garland Publishing Company. (Goodman’s Ph.D. dissertation. A precursor to the position developed more fully in *The Structure of Appearance*.)

References and further reading

Elgin, C.Z. (1983) *With Reference to Reference*, Indianapolis, IN: Hackett Publishing Company.(An explication and elaboration of Goodman’s theory of symbols, which also extends it to the philosophy of language.)
Hellman, G. (1977) ‘Introduction’ to N. Goodman *The Structure of Appearance*, 3rd edn, Boston, MA: Reidel. (An excellent, accessible introduction to the main themes in Goodman’s technical work.)
Schwartz, R. (1986) ‘I’m Going to Make You a Star’, *Midwest Studies in Philosophy* 11: 427-39.(A discussion of the metaphysical position Goodman adopts in *Ways of Worldmaking*.)
Stalker, D. (ed.) (1994) *Grue*, Chicago, IL: Open Court.(An anthology of papers on the new riddle of induction. Contains an excellent annotated bibliography.)

Goodness, perfect

The concept of perfect goodness had a central place in ancient Greek and medieval philosophy, and is still frequently discussed in contemporary natural theology. Medieval philosophers adopted the idea from the philosophies of Plato and Aristotle, with the difference that they identified perfect goodness with a personal God. In ancient and medieval philosophy the concept is primarily a metaphysical one, since goodness was thought to be extensionally equivalent to being, but it is secondarily a moral concept referring to the distinctive sort of goodness appropriate to those beings that have wills. Thus it is fundamental to a long tradition on the metaphysical basis of value which lasted from Plato until at least the sixteenth century.

In Plato, perfect goodness is the Form of the Good, upon which everything that has being is ontologically and causally dependent. In Aristotle, the good is identified with the end or purpose of a natural being. The good is that towards which all things move for the fulfilment of their natures. By the time of Aquinas, medieval philosophers had identified the good in both the Platonic and Aristotelian senses with the Christian God and had argued that God is both the perfectly good creative source and the perfectly good end of all beings other than himself.

The concept of a perfectly good being in Christian philosophical theology faces two major kinds of difficulty. One is the problem that perfect goodness appears to be incompatible with the divine attributes of omnipotence and freedom of the divine will. And if a perfectly good being does not have a will that is free in a morally significant sense, that being seems to lack goodness in the moral sense of goodness. The second kind of problem is that the existence of a being who is both omnipotent and perfectly good seems to be incompatible with the existence of evil. In spite of these problems, there is a strong attraction to the idea of a perfectly good God in contemporary philosophical theology. The category of perfect goodness is therefore one of the most persistent of the concepts in the Platonic legacy.

1 Perfect goodness in Greek and medieval philosophy

In Plato’s metaphysics it is the Form of the Good, not the Form of Being, which is at the pinnacle of the realm of the Forms (see Plato §14). Since any object of kind *K* is real to the extent that it approximates perfect *K*-ness, the better an object is as a *K*, the more real it is. The source of the being of any individual object of kind *K* is the Form *K*, and the individual Forms themselves derive their being from the Form of the Good. In Plato, then, perfect goodness is ultimately the source of the being of everything else. In the *Republic*, Plato asserts the dependency of everything on the Form of the Good by the relation of participation. In the *Timaeus*, Plato develops the Theory of Forms into a cosmological theory according to which all things in the physical world are caused to be by emanation from the Form of the Good, and ultimately return to it.

Aristotle (§§9, 16-17) introduced the connection between the idea of the good and the notion of an end. Each thing of kind *K* has an end or purpose (*telos*) which is the fulfilment of the nature of things of kind *K*, and an individual is good to the extent to which its nature is actualized. Since desires are related to ends, the notion of an end gives desire a role in achieving an end. Aristotle’s teleological theory of nature therefore includes a teleological view of human psychology. Aristotelian natural teleology had a significant impact on the treatment of the concept of the good in later philosophy.

One of the most important features of the concept of goodness arising from Greek philosophy was the identification of goodness with being. The idea that everything is good in so far as it exists leads to the position taken explicitly by Augustine (§§7, 9) in the *Confessions* that evil is only the privation of being. This means that good and evil are not ontologically on a par and, with certain other assumptions deriving from Greek metaphysics, it led to the view that perfect goodness can and, in fact, must of necessity be actualized, whereas perfect evil cannot be. Furthermore, on the view that everything is good in so far as it has being and evil is only a privation of being, it seems to follow that evil is metaphysically necessary if anything exists besides a perfect being. The doctrine that evil is only a privation of being, therefore, suggests a way out of the problem of evil (see §3 below).

Boethius (§2) defended the thesis that everything is good in so far as it exists in the first of his five theological treatises, which became known as *De hebdomadibus*. There he explains how substances can be good in virtue of the fact that they have being even though they are not good in virtue of their substance or nature. Goodness does

not constitute the substance of created substances, but constitutes the substance only of God, from whom all things proceed. Every created substance is therefore good in that it flows from the first and perfect good.

Thomas Aquinas (§§2, 9) synthesized a number of elements of the pre-existing tradition on perfect goodness. Like Plato, he believed that creatures are good by virtue of participation in divine perfect goodness; the created universe is an outpouring of the goodness of God. As Aquinas reads Aristotle, in every genus there is something highest that is the cause of the other members of the genus. Aquinas employs this view in his ‘Fourth Way’ of proving the existence of God:

> That which is said most of all to be such and such in some genus is the cause of all the things that belong to that genus. In this way fire, which is hot to the highest degree, is the cause of all hot things, as is said in *Metaphysics*, Bk II. Therefore there exists something that is the cause of being and goodness and any perfection in all things; and this we call God.
>
> (*Summa theologiae* Ia, q.2, a.3)

In another work he says, ‘Therefore, the highest good [*summum bonum*], which is God, is the cause of goodness in all things’ (*Summa contra gentiles* III, 17: 2). Aquinas follows Aristotle in identifying the good of a thing with its complete being and its end as the achieving of its complete being. And he follows Aristotle’s view of the will as naturally directed towards the proper end of the willer, which is the good for the willer. He follows Augustine and Boethius in identifying good with being and perfect goodness with pure being, in which there is no potentiality unactualized.

After its high point in late medieval philosophy, there was a decline of interest in the concept of perfect goodness. But in the later twentieth century there has been both a resurgence of interest in medieval philosophy and a renaissance in the philosophy of religion. Part of the work in these areas has focused on an approach to theology in which the concept of a perfectly good being occupies the centre of attention. Some of this work aims at reworking medieval philosophy into the contemporary milieu. Other work aims at defending the concept of perfect goodness from the kinds of objection many contemporary philosophers find compelling. Some of the puzzles most commonly addressed will be the subject of the next two sections.

2 The possible incompatibility of perfect goodness with other divine attributes

The idea of a God who is the personification of perfect goodness creates problems when combined with some of the other traditional attributes of the deity. In particular, three problems of compatibility have been discussed in the recent literature, and some of them have a long history.

First, there is the alleged incompatibility of perfect goodness and omnipotence. A perfectly good being was traditionally understood to be one whose will is so fixed in goodness that he is actually incapable of willing anything other than good, a property sometimes called ‘impeccability’. An omnipotent being is one who has maximal power, and while there are many different accounts of what maximal power entails, it has often been understood to involve the power to do anything possible (see Omnipotence §§1, 5). But since sinning or doing evil is a possible thing to do, if a perfectly good being lacks the power to do evil, such a being lacks the power to do something possible and hence is not omnipotent. This puzzle has recently been articulated by Pike (1969), but it was also known to Aquinas and other medieval philosophers.

Aquinas’ way out of this problem was to deny that the power to do evil is a power in the true sense (*Summa theologiae* Ia, q.25, a.3, ad.2), and many philosophers subsequently have taken some version of this approach. To be successful, this approach must elucidate a plausible concept of power which includes what is intuitively desired in the powers of the deity but excludes what is not desired. A different approach is to argue that God *can* do evil, but simply does not. Reichenbach (1982) maintains a position of this kind. Clearly, this move uses a different notion of the nature of perfect goodness from one in which perfect goodness is entailed by God’s very nature. Taking this approach involves weakening the metaphysical connection between God’s nature and being, on the one hand, and his goodness, on the other. A third solution to the problem is to claim that God has the power to do anything possible, including the things which are, in fact, evil, but if he were to do such things, they would not be evil. This approach utilizes the ‘divine command’ theory, according to which God’s will and God’s nature are prior to morality, although perhaps not prior to good in the metaphysical sense of good as pure being (see Religion

and morality §1; Voluntarism). Divine command theories claim that what is morally right (or good) is made to be right (or good) by God’s will. A frequently mentioned problem with this approach is that it makes God’s will appear to be arbitrary. This is because God cannot have a moral reason for willing as he does, since it is his will that determines what counts as a moral reason. In addition, a divine command theory seems to make the concept of perfect goodness vacuous. That is, there is no content to the property of goodness other than this: goodness is whatever it is that God is, and rightness is whatever it is that God wills.

The reasoning behind the alleged incompatibility of perfect goodness and omnipotence leads to a second difficulty. Under the assumption that a perfectly good being is incapable of doing wrong or willing anything but good, the will of such a being does not appear to be free in any morally significant sense (see Freedom, divine). On a common interpretation of the conditions for moral praise and blame in the human case, persons are morally praised because they choose good when they could have chosen evil instead, and they are morally blamed because they choose evil when they could have chosen good instead. The understanding of the conditions of moral praise and blame as entailing the ability to do otherwise was not stressed prior to the modern era, but it is also one which many philosophers think important (see Free will; Praise and blame §4). But then, on the understanding of perfect goodness as involving the inability to will evil, a perfectly good being is not free in a morally significant sense. And it then seems to follow from that that a perfectly good being cannot be praised in the moral sense of praise because he lacks morally significant freedom, and so cannot be good in the moral sense of good. This leads to a third and even more serious problem. If the concept of perfect goodness is intended to include moral goodness and yet the concept of perfect goodness is inconsistent with the concept of moral goodness, as allegedly demonstrated by the foregoing argument, it seems to follow that the concept of perfect goodness is self-inconsistent.

An interesting attempt has been made to use the Thomistic doctrine of divine simplicity (see Simplicity, divine) to resolve the above puzzles about the compatibility of perfect goodness with other divine attributes. According to Aquinas’ concept of simplicity, there is no composition in God at all, so God does not even have attributes, even though it is necessary for us to think of God as having attributes because of the limitations of our finite minds. As this doctrine is described by Stump and Kretzmann (1985), the actual referent of such terms as ‘divine omnipotence’, ‘divine goodness’ and ‘divine being’ is the (simple) divine nature, and so there can be no conflict in reality between any pair of attributes.

3 Perfect goodness and the problem of evil

Even if the problem of the alleged inconsistency of perfect goodness and omnipotence can be resolved, there is still another difficulty. The combination of these two attributes leads to one of the most intractable of all theological problems, the problem of evil (see Evil, problem of). This problem has a number of forms, but it most directly threatens the concept of perfect goodness by way of the following argument: (1) a perfectly good being would be motivated to eliminate all evil; (2) an omnipotent being would be able to eliminate all evil; (3) if a being is both motivated to eliminate evil and is able to do so, he would do so; (4) so if there was a perfectly good and omnipotent being, evil would not exist; (5) but evil does exist; therefore, (6) there is no being who is both perfectly good and omnipotent.

Defences of the compatibility of evil with divine omnipotence and perfect goodness generally focus on modifying either premise (2) or premise (3) of the above argument, or on interpreting them in such a way that (4) does not follow. It is virtually always taken for granted that a perfectly good being is motivated to prevent or to eliminate evil, and so premise (1) is usually accepted by all sides of the discussion. Theists looking for ways out of the dilemma then typically look for reasons why a perfectly good God would permit evil for the sake of some good - often the good of free will or, in general, the good of the world and its creatures. The idea is that a certain kind and amount of evil is a logically necessary prerequisite for obtaining such good.

But there are ways of understanding the concept of perfect goodness which are both faithful to the tradition of reflection on the concept and which lead to the denial of (1) in the sense intended in this dilemma. One is suggested by the divine command theory mentioned in §2. Since divine command theories maintain that what is right or good is determined to be right or good by the will of God, they maintain (1) only in the trivial sense that since good just is what God wills, and since it is trivially true that God is motivated to will what he wills and is not motivated to will what he does not will, then God is motivated to will good and not to will evil. Such an interpretation of (1) does not generate the problem of evil. A modification of this approach is to attempt to ground

God's perfect goodness in those conditions of his nature out of which his will arises, but without making his will primary. One way this might be done is to make motivations the primary good. If God's motivations are an intrinsic feature of his nature and hence good in the primary sense, it follows that good is not something against which the motivations of a good being are to be measured. Instead, the dependency goes the other way. What is good is good because a perfectly good being is motivated to bring it about. It is not the case that a being is good in part because his motivations are to do or to bring about good. On this approach there is no independent conception of good for the sake of which a perfectly good being derives its motivations and for the sake of which he permits some independent conception of evil. This approach could be called a 'divine motivation' theory rather than a divine command theory.

Another way out of the problem of evil raises one of the deepest questions on the metaphysics of value. This is the view that evil does not exist at all. This approach has been favoured at various times in the history of philosophy, but the present era is not one of them. Both inside and outside professional philosophy, many persons believe that the twentieth century has been especially evil and that it is no longer tenable to maintain that evil is a subjective reaction to the undesirable rather than an active and virulent force in history and human life. This attitude threatens the viability of the concept of perfect goodness to the extent that on the traditional conception, a perfectly good being does not *compete* with evil or evil beings. Unlike some of the religions of the East, Western philosophy and religion have never understood good and evil to be on the same level, either ontologically or practically (see Evil). For this reason the concept of perfect goodness has been able to exert a hold on the philosophical imagination far greater than it would have done if evil had been supposed to be an equal combatant.

See also: God, concepts of §§1-6

LINDA ZAGZEBSKI

References and further reading

Aquinas, T. (1266-73) *Summa theologica*, trans. Fathers of the English Dominican Province, Westminster, MD, 1981, Ia, q.5, a.1; q.25, a.3.(Aquinas' treatment of the connection between being and goodness appears in many places in his work. These passages are among the more important. Other references appear in Stump and Kretzmann (1991).)

Aquinas, T. (1259-65) *Summa contra gentiles*, trans. A.C. Pegis, J.F. Anderson, V.J. Bourke and C.J. O'Neil, Notre Dame, IN: University of Notre Dame Press, 1975.(Divided into four books, of which the first three deal with natural theology and the last with revealed theology. The *Summa contra gentiles* has a more philosophical purpose than the *Summa theologiae* since it is directed towards non-believers.)

Augustine (397-401) *Confessions*, trans. with commentary by H. Chadwick, Oxford and New York: Oxford University Press, 1991, bk VII.(An important book in the history of the Christian concept of the relation between being and goodness, frequently cited throughout the medieval period.)

Boethius (early 6th century) *De hebdomadibus*, trans. S. MacDonald, in S. MacDonald (ed.) *Being and Goodness*, Ithaca, NY: Cornell University Press, 1991.(The third of Boethius' *Opuscula sacra (Theological Tractates)*, this aims to discover 'how substances are good in that they have being, yet are not substantial goods'.)

Kretzmann, N. (1991) 'A General Problem of Creation: Why Would God Create Anything at All?' and 'A Particular Problem of Creation: Why Would God Create This World?', in S. MacDonald (ed.) *Being and Goodness*, Ithaca, NY: Cornell University Press.(In this pair of papers, Kretzmann poses a serious puzzle for the idea that God is free in choosing to create the world. He traces two different lines - a Platonistic necessitarian one and a libertarian one - running through the theology of the medieval period, and finds Aquinas' position inconsistent.)

MacDonald, S. (ed.) (1991) *Being and Goodness*, Ithaca, NY: Cornell University Press.(This collection provides an up-to-date treatment of the metaphysics of goodness as it appears in medieval philosophy broadly construed. It includes a translation of Boethius' *De hebdomadibus* in the appendix.)

Morris, T.V. (1987) 'Duty and Divine Goodness', 'The Necessity of God's Goodness' and 'Perfection and Power', in *Anselmian Explorations*, Notre Dame, IN: University of Notre Dame Press.(Good examples of contemporary work attempting to resolve puzzles about the connection between divine goodness and certain other attributes.)

Pike, N. (1969) 'Omnipotence and God's Ability to Sin', *American Philosophical Quarterly* 6 (3): 208-16.(The most important contemporary article on this puzzle.)

Plato (*c*.380-367 BC) *Republic*, trans. G.M.A. Grube, Indianapolis: Hackett Publishing Company, 1974.(This is a classic of political philosophy as well ethics and metaphysics, and its influence on theology has been

considerable.)

Plato (*c*.366-360 BC) *Timaeus*, trans. B. Jowett, in E. Hamilton and H. Cairns (eds) *The Collected Dialogues of Plato*, New York: Pantheon Books, 1961.(This work exerted a profound influence on the medieval mind. It is the account in mythical form of Plato's cosmology.)

Reichenbach, B.R. (1982) *Evil and a Good God*, New York: Fordham University Press.(An example of the idea that God has the ability to do evil, but simply does not.)

Stump, E. and Kretzmann, N. (1985) 'Absolute Simplicity', *Faith and Philosophy* 2 (3): 353-82.(This paper explores the Thomistic concept of divine simplicity as a way of resolving a number of puzzles about the divine attributes.)

Stump, E. and Kretzmann, N. (1991) 'Being and Goodness', in S. MacDonald (ed.) *Being and Goodness*, Ithaca, NY: Cornell University Press, 98-128.(This paper explicates Aquinas' conception of the relation between being and goodness and argues that this position can be used to solve some problems discussed in contemporary ethics.)

Gorgias (late 5th century BC)

The most important of the fifth-century BC Greek Sophists after Protagoras, Gorgias was a famous rhetorician, a major influence on the development of artistic prose and a gifted dabbler in philosophy. His display speeches, Encomium of Helen of Troy and Defence of Palamedes, are masterpieces of the art of making a weak case seem strong, and brilliant exercises in symmetrical and antithetical sentence structure. Of philosophical importance is his treatise On Not-Being, or On the Nature of Things, an elaborate reversal of the metaphysical argument of Parmenides, showing: (1) that nothing exists; (2) that if anything exists, it cannot be known; and (3) if anything can be known, it cannot be communicated. This nihilistic tour de force *is probably a caricature rather than a serious statement of a philosophical position. Gorgias is a master of the persuasive use of* logos *(discourse), understood both as eloquence and as argumentative skill.*

1 Career

Born in Leontini in Sicily, Gorgias came under the influence of the Sicilian philosopher Empedocles as well as the local school of rhetoric associated with the names of Corax and Tisias. In 427 BC Gorgias led a delegation from Leontini to Athens, where his new oratorical style proved to be a great success. His lessons were very highly paid, and as a teacher he travelled throughout Greece. He gave public speeches at Olympia (where he urged the Greeks to unite against the Persians), at Delphi and elsewhere. His written works, such as the epideictic or model speeches *Encomium of Helen of Troy* and *refid="A052bibent4">Defence of Palamedes*, created a new, elaborate prose style which profoundly influenced the development not only of oratory but of literary prose generally. His pupil Isocrates founded the first school of rhetoric and the first institution of higher learning in Athens.

Gorgias lived to be over one hundred years old, and he is said to have read with amusement Plato's dialogue named *Gorgias* (written *c*.395-387 BC) and remarked: 'How well Plato knows how to make fun of people!'.

2 Teaching

Unlike Protagoras (§2) and many later Sophists (see Sophists), Gorgias made no claim to train students in moral and political virtue, but only to make them skilful public speakers. He developed both the theory and practice of persuasive speech (*logos*) as a technique of power (see Logos §1). Comparing the effect of speech on the soul to the action of drugs on the body, Gorgias describes how 'some speeches cause grief, some cause delight, some produce fear in the hearers while others produce confidence, and some by an evil persuasion drug and bewitch the soul' (*Helen* 14, fr.11). Hence in legal and political competitions 'a single *logos* delights and persuades a great crowd, because it is written with art, not because it is spoken with truth' (*Helen* 13, fr.11). In the dialogue *Gorgias*, Plato calls attention both to the power and to the lack of moral responsibility involved in Gorgianic training in rhetoric. And by presenting as followers of Gorgias the subsequent speakers Polus and Callicles, who systematically attack traditional Greek notions of morality, Plato's dialogue clearly implies that 'Gorgias' teaching is the seed of which the Calliclean life is the poisonous fruit' (Dodds 1959: 15).

Gorgias' own conception of 'writing with art' produced a highly ornamental prose style, where balanced clauses and euphonious word choice dazzle the ear and titillate the mind in sentence after sentence. We can still appreciate this effect in the extant *Encomium of Helen* and *Defence of Palamedes*, two display pieces that were apparently written as models for his students to memorize and imitate. A later taste finds this style artificial and exaggerated, but as a sheer technical exploit Gorgias' compositions made such an impression on his contemporaries that the art of writing in prose was permanently changed. Agathon's speech in Plato's *Symposium* (194e-197e) is one of the more elegant examples of direct Gorgianic influence. Gorgias' most distinguished pupil, Isocrates, established the classical style in Greek oratory with milder, less mechanical forms of parallelism and antithesis in sentence structure, which Cicero was later to reproduce in Latin.

3 Philosophical work

It is not clear that Gorgias made any major contribution to the development of Greek philosophy. Whether or not he was personally a pupil of Empedocles, as an ancient tradition reports, he was certainly familiar with the new natural philosophy which had come from Ionia to Sicily by the middle of the fifth century BC, and he made use of it in his teaching. Thus Meno, who is represented by Plato as a pupil of Gorgias, is fond of physical explanations

of sensory perception in the style of Empedocles (*Meno* 76c-e) (see Empedocles §6). To this extent, as a popularizer of the new science Gorgias was (like the other Sophists) a typical representative of the fifth-century Enlightenment. His own, more personal, achievement is represented by a treatise *On Not-Being or On the Nature of Things*, which has come down to us in two later summaries, one in the Aristotelian corpus (the pseudo-Aristotelian text *On Melissus, Xenophanes and Gorgias*, 979a-980b) and another version given by Sextus Empiricus (*Against the Professors* VII 65-87).

It would seem that the title of this treatise is already a joke, since it equates a discussion of nonentity or nothingness (the Parmenidean notion of not-being) with the Ionian investigation of the nature of things (*peri physeōs*) (compare Parmenides §2). Even if the intention of the author is not entirely serious, the text is of considerable historical interest, on several counts. It contains the longest example of continuous argumentation that has reached us from the fifth century BC. The closest precedent is provided by the poem of Parmenides and the paradoxes of Zeno of Elea). Gorgias' argument is in fact a direct inversion of Parmenides' reasoning in favour of being, which, according to Parmenides' argument, is the only subject that can be known and described in rational language. By contrast, Gorgias argues: (1) that nothing exists, or that there is no being at all; (2) that if anything exists, it cannot be known; and (3) if anything can be known, it cannot be communicated in language. Gorgias' reversal of Parmenides' reasoning follows the original so carefully that it has permitted modern scholars to reconstruct some features that are badly preserved in Parmenides' text. Gorgias' reasoning is indirect in form, like Parmenides', and it has benefited from the additional subtlety introduced by Zeno's paradoxes. Gorgias' work thus served to popularize not only natural philosophy but also Eleatic logic.

This treatise is, like the *Encomium of Helen*, a *paignion* or plaything. But what Gorgias is playing with is Parmenidean ontology and Zenonian dialectic. Consider the argument for the first thesis, 'that nothing exists' or 'that there is nothing'. (Translation is difficult, since in Greek the verb *esti*, 'is', and the participle *on*, 'being', cover both existence and predication; the notion of being represents a fusion of the two.) In the version of the argument given by Sextus, the proof of this thesis begins with a trilemma, echoing the three Paths of Parmenides' discourse on Truth. 'If something exists, then either Being (*to on*) exists or Not-being (*to mēon*) exists or both Being and Not-being exist'. The three options are then eliminated *seriatim*. First of all, 'Not-being is not (does not exist). For if Not-being exists, it will at the same time be and not be. Understood as Not-being, it will not be. But insofar as Not-being exists it will also be. But it is utterly absurd for something at the same time to be and not to be. Therefore Not-being does not exist'. The proof that being also does not exist begins again with a trilemma: 'If Being exists, it is either eternal or generated or both'. All three possibilities are then refuted.

The arguments to show (2), that, if anything exists, it is unknowable, and (3), that, if anything is known, it cannot be communicated, are of philosophical interest inasmuch as they emphasize the gap between thought and reality on the one hand, and between thought and language on the other. But the preserved arguments seem grossly fallacious. They are good illustrations of the later notion of sophistry as making use of specious argument. On the other hand, the general form of the reasoning is reminiscent of the dialectical second part of Plato's *Parmenides*. Since we do not have Gorgias' own wording, it is difficult to know whether Plato is imitating Gorgias or whether our text has been reshaped under the influence of later, more sophisticated argumentation.

Besides displaying the virtuoso skills of its author, Gorgias' treatise seems designed to discredit ambitious philosophical reasoning of the sort initiated by Parmenides and Zeno. This is in any case the conclusion drawn by Isocrates (Gorgias fr. 1): if these theses can be logically argued for, why should we take any such arguments seriously? In a more light-hearted way, but with greater technical sophistication, Gorgias thus joins the attack on Parmenidean dogmatism that was launched by Protagoras with his 'Man-the-measure' thesis.

CHARLES H. KAHN

References and further reading

Anon. (date unknown) *On Melissus, Xenophanes and Gorgias*, trans. W. S. Hett, *Aristotle, Minor Works*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1936.(Although spuriously attributed to Aristotle, and of uncertain date, it is an important source on Gorgias' *On Not-Being*, omitted from Diels and Kranz (1954); includes Greek text with English translation.)

Cole, T. (1991) *The Origins of Rhetoric in Ancient Greece*, Baltimore, MD, and London: Johns Hopkins University Press.(Strikingly original view of Gorgias' place in the development of rhetoric and written prose.)

Dodds, E.R. (1959) *Plato, 'Gorgias', A Revised Text with Introduction and Commentary*, Oxford: Oxford University Press.(Masterful discussion of the relation between Plato's dialogue and the historical figure of Gorgias.)

Gorgias (late 5th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 7th edn, 1954, vol. 2, 271-307.(The standard collection of the ancient sources; includes Greek texts with translations in German.)

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, vol. 3, Cambridge: Cambridge University Press.(A full, scholarly account.)

Kerferd, G.B. (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press.(A briefer, more personal interpretation than that of Guthrie (1969).)

Plato (*c*.395-387 BC) *Gorgias*, trans. T. Irwin, Oxford: Oxford University Press, 1979. (Includes commentary.)

Plato (*c*.386-380 BC) *Meno*, trans. W.K.C. Guthrie, Harmondsworth: Penguin, 1956.(Presents Meno as a pupil of Gorgias.)

Plato (*c*. 386-380 BC) *Symposium*, trans. A. Nehamas and P. Woodruff, Indianapolis, IN and Cambridge: Hackett Publishing Company, 1989.(Contains display speech in Gorgianic style.)

Sextus Empiricus (*c*.200 AD) *Against the Professors*, trans. R.G. Bury, *Against the Logicians*, vol. 1, *Against the Physicists, Against the Ethicists*, vol. 2 and *Against the Professors*, vol. 3, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 3 vols, 1935, 1936, 1949.(Contains summary of *On Not Being*.)

Sprague, R.K. (ed.) (1972) *The Older Sophists*, Columbia, SC: University of South Carolina Press, 30-67.(Full translation of the fragments.)

Grace

Grace is a gift of personal relationship with God surpassing the powers of nature. Such relationship presupposes the relation every creature has with God as immediately dependent on him for its very existence and continuance, but goes beyond it in its distinctive personal character and the resulting more intimate dependence.

Such personal relationship necessarily depends upon some will or expression of will. The development of the relationship must be free in relation to what has gone before - it must in no way be necessitated. Therefore, if the relationship arises within some established natural system involving some degree of conformity to law at some levels of description (physical or otherwise), such a system and laws must leave some degree of indeterminacy. This allows the possibility that free actions, physically and psychologically contingent, should be among the ultimate determinants of what unfolds. For the workings of grace must always be free or contingent in relation to such a system of nature, concordant with and fulfilling nature, but working according to distinct principles - grace and nature are both equally God's gift, but always distinct.

The notion of grace has no place outside some kind of theism. Implicit in Judaism, its explicit development has been in Christian theology, which sees human beings as radically dependent upon God not only for redemption from personal sin, but for any personal or 'supernatural' relationship to God at all. This emphasis on grace is peculiar to Christianity, although the general conception seems implicit in most theistic religions; it can perhaps be found in the more theistic forms of Hindu devotion, although it is absent from Buddhism and the stricter forms of Hinduism.

1 Grace and nature

In considering grace, there are questions which spring from a need for logical clarification and questions arising from real theological disagreement. Differences in philosophical framework often make applying this distinction difficult. The same formulation may disguise real differences, while opposed forms of words sometimes conceal agreement.

The first distinction to be made is between what belongs to God's unconstrained will simply by his being and nature (for example, his existence and his delight in it, and, in Christian understanding, the relationships of Father, Son and Holy Spirit, none of which are by free choice) and what arises from his free will. Creation and grace both belong to his free overflow of love, entirely without necessity.

Grace is a gift of God constituting or expressing a supernatural relationship (analogous to friendship) or relational disposition (personal goodwill directed towards such friendship): many of the words for 'grace' in different languages signify 'being favoured', either as a steady state or as a possibly temporary disposition. As such, it arises in connection with created persons, not just as created beings but precisely as persons. In the human case, these persons begin in the womb and mature in stages. In Christian understanding, their relationship is to God as three persons, Father, Son and Holy Spirit, united in one single act of being. The word 'supernatural' in this context means something essentially geared to such a relationship or possibility of relationship, freely given and entered into. As such, it is entirely distinct from the merely 'preternatural', a word which refers to what is coeval with nature even if interfering with it, such as poltergeists and other supposed paranormal phenomena, and any supposed activities of spirits, devils or disembodied souls not directed to assisting or resisting divine activity.

However, God's special favours towards individuals have sometimes been conceived solely in terms of his providing some special created happiness (such as envisaged in portrayals of paradise as an earthly garden), or special help in dealings in this life (such as described in some of the Psalms), or special prophetic knowledge. This is compatible with a way of thinking which regards virtue pleasing to God, and even perseverance so as to attain heaven (identified with paradise), as within human power, perhaps assisted by the imitation of moral examples such as Jesus. This view is represented in the fifth-century British monk Pelagius as well as in modern exemplarist theories of the atonement, such as that of Hastings Rashdall (see Pelagianism §1).

Such approaches led Western Christian theology, following Augustine (§13), to give primary emphasis to the necessity of grace for any state, act or virtue pleasing to God, for any repentance and any perseverance in virtue. Augustine conceived mankind as having forfeited all right to friendship with God by a primeval choice preferring

lesser goods (for, since God had created mankind to be a family in solidarity with one another, the sin of one affected all) (see Sin §2). Hence, the first role of grace was to undo the effects of this 'original sin'. Accordingly, the primary role of Jesus' redeeming action tended to be seen as the undoing of this negative effect so as to reopen the possibility of supernatural friendship with God.

But such a negative conception of grace, as necessary simply to undo the effects of sin, obscures the absolutely basic Christian presumption that any act of love towards God appropriate to the kind of friendship God intends is beyond human power alone. In this conception, such acts of love would have been beyond human power even if there had been no sin - indeed they are equally beyond angelic power, so that even the angels depend on grace. Such love requires the coaction of God within our act, and it is this which makes the kind of friendship concerned essentially 'supernatural'. This principle was plain in John's Gospel, Augustine and many other Christian 'Fathers' long before being given Thomas Aquinas' careful articulation, according to which any act meriting salvation has to be moved by charity (*Summa theologiae* IaIIae, q.109; q.114, aa.2, 4), which is explained as a God-empowered mutual friendship involving God's co-working in us (IIaIIae, q.23, aa.1-2; compare a.7; on co-working, IaIIae, q.111, a.2).

Many other theists share this general conception, but Christian teaching gives it special emphasis. Christianity envisages a relationship which enables us to pray 'Abba, Father' ('Abba' being an intimate word like 'papa' or 'daddy'). Such prayer expresses some intimate sharing in the life of the Godhead alongside Jesus, a sharing that gives hope of the vision of God. Such intimacy and such vision are not within natural human power.

2 Grace essentially related to loving personal relationship with God

Historically we find two different accents in the treatment of grace. First, there is an accent on grace as God himself, under the aspect of giving himself in some kind of divine indwelling, and under this aspect referred to as 'uncreated grace'. Second, there is an accent on created grace, grace considered as involving alteration in the created person and their intellect and will. Eastern Christian tradition has always emphasized uncreated grace, that is, God himself, sometimes speaking of this as an uncreated divine energy, and regarding this as primary. By contrast, Western tradition has commonly treated the real alteration of created persons, their transformation into new creatures, as most important. This created grace has been conceived as going to the root of the person's being, but also as co-working within their faith and love and their particular acts of faith and love.

The insistence on grace as something real in the creature (on created grace), has been especially evident in Roman Catholic teaching; for example, it is found in Aquinas' rejection of the idea that grace consists only in an attitude of God (of favour') towards someone, rather than in a gift of God to that person making them pleasing to God (equally, Aquinas rejects the idea that the remission of sins involves only imputing justice to people rather than effecting a real change in them, rendering them actually just and pleasing; see *Summa theologiae* IaIIae, q.110, a.1, q.113, a.2). This has sometimes generated the picture of grace as being like a liquid poured upon us or into us, it then being mistakenly conceived of as something which, once given to us by God, we possess neither more nor less independently of him than we possess his natural created gifts; this picture is often associated with the (mistaken) belief that medieval thinkers supposed the sacraments to communicate grace by created human power (contrary to the Catholic insistence that God alone causes grace and is the principal agent in the sacraments). Luther revolted against such notions, insisting on speaking of justification as something imputed rather than imparted, and rejecting the supposed conception of the sacraments as human works causing grace. In such a context, the Reformers unsurprisingly preferred such translations as 'highly favoured' to 'full of grace' as descriptions of the Virgin Mary.

Hans Küng (1964) interprets Luther's texts in a way seemingly compatible with regarding the grace of being justified as internal to the person justified, it being wrong to suppose that what God regards as just he could so regard if he did not in the same act make it actually just. Whether or not Küng's own explanations are satisfactory, granted the obscurities of his phenomenological/existentialist idioms, clearly many other points of dispute between Catholic and Protestant conceptions are terminological rather than real. Standard logical distinctions have to be taken into account: for example, whether or not someone is courageous is a matter of fact, not a matter of degree, whereas how courageous someone is is a matter of degree. Thus, in the standard Protestant idiom, 'the grace of justification' refers to the matter of fact and 'sanctification' to what has degrees - whereas the standard Catholic idioms for the same matter of fact are 'being in a state of sanctifying grace' or 'being in communion with God',

and ‘the grace of justification’ and ‘sanctification’ are treated as the same reality, distinct only in the aspect described, both equally having degrees.

John Oman (1917) indicates how the underlying problem is to be resolved by recovering the full implications of conceiving grace relationally. His explanation is reminiscent of Aquinas’ treatment of the supernatural (‘theological’) virtue of charity as involving essential mutuality, mutuality deeper than in normal human friendship. Aquinas’ explanation of this depth is that this friendship is only established by the love of God first giving us the requisite faith and goodwill towards him, and that the acts belonging to it essentially involve, on the human side, God’s co-working in us. The general problem is how to give an adequate explanation of the particular way in which grace in the creature has no existence independently of relationship with God. It is the peculiar way in which personal relationship is internal to grace that makes it intrinsically supernatural.

Karl Rahner (§5) uses ideas from Aquinas to explain how God dwelling in us (uncreated grace) is specially internal to the transformation of the creature (created grace), noting that Aquinas regarded the sanctifying grace required for charity towards God in this life as having the same ontological presuppositions as the immediate vision of God (Rahner 1961). Only God (Jesus, the Word of the Father) dwelling in the soul could be the principle of this beatific vision - God is related to it as the impression or concept shaping our seeing or understanding of something is related to that seeing or understanding. God the Father and Son, with the Spirit breathed by them, make themselves, as it were, the internal principles of the workings of our own heart and mind. It is this presence which makes us ‘new creatures’, adapted to God’s co-working in us in this love and vision, things whose possibility is known only through revelation.

In Catholic and Orthodox thinking, grace always perfects nature, never abolishes it. It has to be the very same human being, previously sinful, who is now in love with God: continuity in individual identity depends on some continuity of nature. Nature is not of itself capable of grace, nor is the finite of itself capable of the infinite, but created nature can be enhanced by grace so as to be able, with God’s co-working, to reach the love and vision spoken of. However, this still involves not just an addition to or strengthening of the powers of the human soul or person, but a change at a logically deeper level. For, as the difference between a human being and an amoeba is not just that additional powers are added to a nature which is the same, but that the possession of these additional powers reflects a richer nature of which these powers are only an expression, so the difference between a person supposed to be without grace and a person in a state of grace involves an enlargement of a person’s nature. A person’s power to know and love God in the way appropriate to the friendship opened out by faith presupposes such enlargement. Therefore, grace has the same logical relation to the powers and activities which it makes possible as human nature has to natural human powers and activities, and it is in this sense that it belongs to the essence of the soul rather than being a mere accident external to it - that is, it constitutes a real transformation at the root of the human being, making a person, in St Paul’s words, ‘a new creature’.

3 Grace within the human framework

The phrase ‘the economy of grace’ traditionally signified the household provision for the distribution of grace among the human race as a historically ramified family (subject to no restrictions of scarcity such as the modern use of the word ‘economy’ suggests). In the traditional Christian conception, every giving of grace in God’s providence is preceded by some preparedness for the receiving of the grace concerned (some ‘prevenient’ grace). Thus, in each individual’s life, some remnants of the original good in human nature give them some openness to truth and desire for relationship in their heart with God. More particularly, before giving any special gift, God tends first to raise up the desire and then even the will and prayer for it. In this conception, the whole law of God’s working towards humanity is not that he imposes his will, but that he waits on human readiness, consent and cooperation, each as appropriate to the degree of maturity of the human beings concerned.

In Jewish as well as Aristotelian thinking, human beings are essentially social by nature - an inseparable consequence of their bodiliness and biological nature (in scholastic thinking, putting human beings in contrast with the angels, whose sociality is supposedly differently rooted). This essential sociality is the key to Mersch’s explanation of ‘original sin’ (Mersch 1951): God allowed the sin of one to affect all, preserving human solidarity, because he was intent on human beings coming to salvation as one community, so that in Jesus’ becoming man the same solidarity should become the means of opening up a new way to salvation, the obedient love of one being effective for all. It also explains how the prayer or consent of one person or group, or the Church, can avail for

others, so that none need be too immature, too weak or even too proud (since pride can be broken) to be outside the scope of grace - and none too ignorant by reason of their historical or geographical situation.

In common Christian thinking, grace restoring human beings to fellowship with God ('freeing them from original sin') has been made available to every human being throughout time. Thus, before the coming of Christ, independently of historically received knowledge of his redemptive suffering and death on the cross (his 'passion'), human beings received grace in anticipation of this redemptive work. All this is conceived as geared to a plan of God for the whole of mankind; so, for example, the fact that Old Testament prophets, and, in Catholic and Orthodox belief, the Virgin Mary in an especially complete way, should have received grace and grown in friendship with God from early in their lives was part of God's wider plan, each stage of the unfolding of which should wait upon appropriate human preparation, consent and cooperation.

It accords with this conception that, within this plan, the immature (for example, the infants massacred in Bethlehem) with a place in the prayer of Israel or of other God-inspired religious company (in the case of infant baptism by the prayer of the Church), and all young or old who suffer 'for the sanctification of the Name' (Jews, notably in the Holocaust, and Christians and others in other martyrdom), are in reach of salvation - and, along with these, everybody who, although mature, lacks effective explicit knowledge of God's desire and plan (whether through living in times and places too early or remote to have heard the gospel, or through accidents of culture affecting response to the gospel). All these stand as people waiting, waiting to enter salvation in one company with those to whom Christ's coming, the Church, Scripture and history make a more fully human response possible in this life. 'Fully human' means fully relational, personal and communal, fully intellectual with appropriate bodily and social expression by sign and voice. In this conception, such waiting is absolutely required if the salvation into which all who are willing are to enter is to be as fully humanly unified and complete as the work of Christ makes possible: bodiliness and historicality necessarily involve such a structured communal entry into salvation. In this conception, a person can be excluded from such salvation only by their own decision of heart and will. Tradition and different theologians have indicated so many different ways God might prepare human beings so as to be able to receive grace that it has become increasingly obscure whether a doctrine of limbo is necessary at all - so long as grace is only in anticipation of Christ's work and salvation is completed only in being joined in community with Christ and waiting 'to reach their perfection only in company with us' (Hebrews 11: 40) (see Limbo).

St Paul tells us that God desires the salvation of all human beings. Not all have interpreted Paul in the way sketched above (Aquinas outlines the three traditional alternatives in *Summa theologiae* Ia, q.19, a.6, ad.1). For some, grace seemed inaccessible to the unbaptized and even to all the immature, as well as to virtuous heathen ancient or modern. Others made the human will seem almost irrelevant by considering all salvation to depend upon God's determination in giving the graces of conversion and perseverance to an effective extent, either predestining only some (Calvin's tendency) or by predestining all ('universalism'). Modern theological history is littered with controversies in this area, with views associated with Calvin (§§3-4) and Jansen (tending in one direction), and Molina (§§2-3) and Arminius (tending in the other) all seeming one-sided or unacceptable to some, while Dominicans and Jesuits were restrained from controversy even in the supposed middle ground. In short, the problems of conceding a genuine reality to time and a real coauthorship in history between God and human beings in the giving and receiving of grace have proved more difficult than those concerning God's foreknowledge.

There has been much confusion between foreordination to salvation and knowledge of it (called 'assurance'). Aquinas argued that none have certain knowledge of their damnation (such knowledge would make them worse) and that few have certain knowledge of their salvation (knowledge tending to generate arrogance), regarding this as a special gift reserved for a few (*Disputed Questions on Truth*, q.6, a.5); many Protestants, however, regard it as the norm, and this has been presented as one of the defining features of Methodism.

4 Grace and creation

One problem with conceiving grace negatively as simply undoing the effects of sin was discussed in §1 above. A second defect in this conception is that it obscures the logical point that the good gained in redemption may be greater than that lost through sin. Christian theology requires not just that this may be so, but is so, for, as Augustine (§9) says, God only permits evil because through it he may bring greater good: Adam consorted with God in the first paradise but, after the death and ascension of Jesus the Son, the whole Godhead, Father, Son and Holy Spirit, are to come together to live in those who abide in the Son's love - God not only with us but within us.

The Roman Catholic Easter liturgy even refers to Adam's sin as 'happy fault, necessary sin, to gain for us so great a Redeemer'.

Medieval theologians asked whether Jesus, God the Son, became incarnate only to undo the effects of sin, further benefits being accidental, or whether he would have adopted this solidarity with mankind anyway, independently of sin. Aquinas, regarding Scripture as making sin the only certain reason for the Incarnation, left it open that God might still have willed the Incarnation even if there had been no sin. By contrast, Duns Scotus, relying on the principle that what is fitting must be true (a principle Aquinas did not regard as giving proof in revealed theology), envisaged the willing of the Incarnation as unconditional, a way followed by many modern theologians, especially those influenced by Karl Barth; it is echoed by de Lubac (1984) and Rahner (1961), who both represent 'nature' as a mere abstraction, historically never free of the supernatural, but always already laden with grace. They refuse to envisage the plan of creation as established first, with grace arising only as a special response to a contingency within it; instead they regard the whole history of creation as set within the context of the order of grace rather than vice versa, so that it is essentially, not just incidentally, a *Heilsgeschichte* (conventionally translated as 'Salvation History'). Within such a perspective, the greater good arising in response to a foreknown Fall is conceived as foreordained.

See also: Justification, religious; Predestination; Providence; Reprobation; Sacraments; Sanctification

DAVID BRAINE

References and further reading

Aquinas, T. (1266-73) *Summa theologiae*, London: Blackfriars, 1961-70, Ia, q.19, a.6; IaIIae, qq.109-14; IIaIIae, q.23, aa.1-3.(These are the key questions respectively on the distribution of grace; grace and merit; and on charity.)

Aquinas, T. (1256-9) *Disputed Questions on Truth*, Chicago, IL: Regnery, 1952.(Also commonly referred to as *QD. de veritate*; on predestination and the relation of foreknowledge and foreordination, see q.2, a.12; q.5, a.5; and q.6, aa.3, 5.)

Küng, H. (1964) *Justification: The Doctrine of Karl Barth and a Catholic Reflection*, London: Burns & Oates. (The then Catholic theologian's attempt to reconcile Protestant and Catholic views of justification.)

Lubac, H. de (1984) *A Brief Catechism on Nature and Grace*, San Francisco, CA: Ignatius.(Presents the idea that human nature's orientation towards grace is a datum of human existence; see also his *The Mystery of the Supernatural*, London: Chapman, 1967.)

Mersch, É. (1951) *The Theology of the Mystical Body*, St Louis, MO: Herder. (Chapter 7, on original sin, explains the notion in terms of our understanding of salvation, rather than vice versa.)

Oman, J. (1917) *Grace and Personality*, Cambridge: Cambridge University Press. (Presents the idea of explaining 'grace' in terms of relationship.)

Rahner, K. (1961) *Theological Investigations*, London: Darton, Longman & Todd, vol. 1, chaps 9-10.(Chapter 9, 'Concerning the Relationship between Nature and Grace', presents 'nature' as an abstraction rather than an empirical reality, the possibility of grace being the context of any human existence; chapter 10, 'Some Implications of the Scholastic Concept of Uncreated Grace', shows how the concept of created grace appears to be ultimately dependent on the concept of uncreated grace, that is, on special ways of God's being personally present.)

Gramsci, Antonio (1891-1937)

An Italian Marxist theorist and activist, Gramsci's main contribution lies in his critique of dialectical materialism. This school treated both the power of the bourgeois state and the prospects for its revolutionary overthrow and replacement by a stateless communist society as necessary consequences of the autonomous development of the economic forces of production. In contrast, Gramsci emphasized the relatively independent role played by politics and culture in upholding the authority of the state and in organizing popular resistance to it. A canonical figure of Western Marxism, he is credited with formulating a strategy applicable to all communist parties operating in the democratic states of advanced industrial societies. However, the posthumously published Quaderni del carcere (Prison Notebooks) elaborate on theoretical questions that had preoccupied Gramsci throughout most of his political career and which reflect peculiarities of the Italian political situation and traditions.

1 Life

Born in Sardinia, Gramsci won a scholarship to study at the university of Turin in 1911. He became involved in the Socialist Party and eventually gave up his studies to become a full-time journalist for various Party newspapers. The journal *L'Ordine Nuovo (New Order)*, which he edited with the future communist leader Palmiro Togliatti, provided the main theoretical inspiration for the occupation of the factories in the Turin region in September 1920. Dissatisfaction with the Socialist Party and union support for the movement led him to join the secession to form the Communist Party of Italy (PCd'I) in 1921. However, his unorthodox Marxism and his advocacy of workers' democracy and the formation of an alliance between proletariat and peasants were initially just as out of favour here. Mussolini's seizure of power in 1922 led to a change of Party policy. In 1924 Gramsci was elected to Parliament. He returned from Moscow, where he had acted as the PCd'I's representative and met his wife Julia Schuct, and was appointed General Secretary of the Party. In 1926, however, he was arrested, and in 1928 sentenced to twenty years imprisonment. In spite of the Fascist prosecutor's desire 'to stop this brain functioning', enforced withdrawal from politics enabled him to reflect upon it: between 1929 and 1935, when chronic illness sapped even his intellectual willpower, he filled thirty-three exercise books with some 2,848 pages of notes. He had suffered from ill health all his life which was exacerbated by prison conditions and died a week after his commuted term ended.

2 Historical materialism and Crocean historicism

The distinctiveness of Gramsci's thought stems from his critique of the positivist Marxism of his day. He began by attacking the materialist and associationist psychologies employed by certain versions of historical materialism. While matter might be the essence of things, such theories ignored humanity's capacity to control natural processes and mould them to its will. He criticized the crude empiricism of certain positivist Marxists on similar grounds: facts do not speak for themselves but only make sense within some theory that provides criteria for their selection and significance. He then built on these criticisms to dispute the claim that the economic base determined in a mechanical manner the ideological and political superstructure (see Marx, K. §7). As before, such a statement ignored the independent role played by human consciousness. Gramsci pointed out that institutions and belief systems had internal dynamics of their own, which had no direct connection with economic developments. More important, the influence of the base was mediated through particular superstructures, and we engage with the structure and partially constitute it through our activity and ideologies. In one of his most famous early articles, he contended that the Russian Revolution could only be understood in this way. The collapse of the Tsarist regime and Lenin's ability to seize his opportunity had been to a large extent superstructural. The Soviet Union's task was now to build the economic conditions from which a true communist society could emerge. As such it was a 'Revolution against *Capital*', at least as economistic Marxists interpreted this work.

Gramsci's critique drew on the ideas of the idealist philosophers Croce (§3) and Gentile (§3), both of whom had written important studies of Marxism. Particularly influential were Croce's denial of a priori principles and his identification of philosophy with historical judgment, and Gentile's thesis of self-constitution through the unity of theory and practice. However, Gramsci was keen to rebut the charges of idealism and voluntarism that had been levelled against him, and a major section of the *Quaderni del carcere (Prison Notebooks)* is devoted to pointing out Croce's errors. Gramsci argued that just as materialists treated matter as a metaphysical entity that determined

all human action, so Croce had a tendency to attribute all human thought to the unfolding of a mystical ideal entity - spirit. He charged Croce with treating history as the product of the internal dynamics of thought, and ignoring the interplay of theory and action in the practical task of engaging with and changing the world. Gentile’s actualism might seem more promising in this respect, but Gramsci believed he committed the even graver error of emotivism, which culminated in the conclusion that might is right, as his allegiance to Fascism demonstrated only too clearly.

Thus, Gramsci contended that in different ways both materialism and idealism resulted in passivity. Materialism encouraged the masses to regard revolution as an inevitable product of changes to the economic base to which they need not contribute. However, the existence of the objective conditions for revolution was not in itself sufficient: ‘One must "know them" and how to use them. And want to use them’ (*Quaderni*: 1338). Idealism led to an uncritical acceptance of prevailing ideas and a failure to relate them to economic and social circumstances. Going back to Marx’s 1857 ‘Preface’ and the *Theses on Feuerbach*, Gramsci argued that the true Marxist message was that the economic base constrained rather than determined the superstructure. The superiority of Marxism over other social philosophies lay in its claim to be able to make the most of existing circumstances so as to benefit humanity at large - a goal Gramsci largely associated with the achievement of the maximum productivity possible with the available techniques. However, this claim had to be sustained by engaging with existing ideologies and persuading the masses to act on the basis of Marxist principles.

3 Hegemony

Gramsci’s emphasis on the role of will and consciousness fed into what most commentators regard as his chief contribution to philosophy - his concept of hegemony. Gramsci employed this term in two related ways. First, he used it to denote the consensual and ideological, as opposed to the coercive, basis of a political system. The state consisted of two elements. On the one hand, there was the coercive apparatus comprising the police, army and judiciary who could uphold the authority of the ruling class through force. On the other hand, there were the various institutions of civil society, such as the media, schools, churches, clubs, parties and trade unions. These organizations were the instruments of hegemony, the means whereby the dominant class obtained the spontaneous adherence of the rest of the population to its rule. Hegemony would allow a ruling group to hang on to power long after it had ceased to be the leading economic class. This point was vital to understanding the resilience of liberal democracies during the economic crisis following the First World War. Gramsci claimed that civil society was infinitely more developed within advanced industrial societies. In contrast to Russia, which was comparatively backward, in these countries revolution could not take the form of a direct military assault on the forces of coercion until the battle for civil society had been won and the masses convinced that their true interests lay in overthrowing the old order. In Gramsci’s terminology, it was necessary to fight a ‘war of position’ to capture or supplant the institutions of hegemonic power before turning to a ‘war of manoeuvre’ and violent revolutionary action.

This leads to Gramsci’s second usage of ‘hegemony’; this time referring to the cultural and educative task of the Party in the formation of a coherent moral awareness and political will among the proletariat. The building of this new hegemony went through numerous phases. At the most basic level, it involved the political organization of the proletariat and making them conscious of their class and economic interests. However, to be truly hegemonic it was necessary to go beyond this stage and persuade other groups, particular the peasantry and the petit bourgeoisie, to identify their interests with those of the proletariat. Gramsci believed that the bourgeoisie in the UK and France had successfully universalized their point of view in this manner, convincing workers of the validity of capitalist economics and getting them to adopt bourgeois aspirations.

While Gramsci believed that this strategy had to have some foundation in economic reality to be plausible, it was an essentially political task. The Party had to constitute itself as an alternative civil society, offering schools, clubs and a focus for social life. A particularly important role would be played by intellectuals, who had the job of mediating between the Party leadership and ordinary people. Gramsci distinguished between ‘traditional’ and ‘organic’ intellectuals. The first were self-styled ‘detached’ scholars such as Croce. The second, in contrast, did not see themselves as a class apart. They acknowledged their membership of a particular social group and sought to lead its members to greater self-awareness. Gramsci did not think that the masses could ever be totally deluded by bourgeois ideology. The seeds of dissent from hegemonic views always existed in embryonic form if these failed to correspond to real interests. Organic leadership simply consisted of making the masses aware of this fact

and getting them to relate their dissatisfaction to those of others. The result was to be the formation of a collective will for change.

Gramsci rejected Lenin's view that the Party cadres had to form an elitist vanguard and dictate to the people what they ought to believe and do by referring to scientific laws of history (see Lenin, V.). He opposed 'democratic centralism', that brought together leaders and led 'organically' to 'bureaucratic centralism' that operated 'mechanically' on the basis of directives sent from above. A Party organized in this fashion would always be a 'policing organ' rather than 'a deliberative body' (*Quaderni*: 1691-2). However, his theory was not without a sinister side of its own. The moral and intellectual revolution within people's consciences was to be so complete that the Party was to take the place of 'the Divinity or the categorical imperative' in their minds, with all acts judged good or bad to the extent that they benefited the Party (*Quaderni*: 1561).

4 The New Order

Gramsci believed that a cultural consensus around a shared moral vision would make the coercive force of the state redundant under communism. Drawing on Lenin's somewhat idealized account of the soviets, his own experience of worker's self-management during the occupation of the Turin factories, and his reading of accounts of the introduction of Frederick Taylor's management techniques by Henry Ford, he argued that factory production provided the economic and social preconditions for the autonomous organization of workers. He believed that the industrial system was potentially self-regulating. All that was required to ensure that production was related to the needs of the producers and that supply exactly matched demand, was to raise each individual worker's consciousness of their role within the productive process and their relationship to the rest of the global economy. This goal could be achieved through a worldwide network of factory and peasants councils that was hierarchically organized according to trade and region, and which culminated in an international administration that supervised the world's resources in the interests of all humanity. This network would serve as a vast information system through which workers both channelled their requirements and modified them to take into account those of others. In this way, a collective will gradually came into being that was linked to the maximal employment of the productive system. In so arguing, however, Gramsci risked justifying as total a subordination of the individual to the demands of technological development as he accused deterministic Marxists of proposing. His theory has no place for the traditional Marxist concern with alienation, concentrating rather on the removal of anomie through the assimilation of norms most suited to the regulation of production (see Alienation §§3-5).

5 An Italian theorist

Gramsci's ideas grew out of his reflections on the relatively backward nature of Italian capitalism and liberal democracy, and his attempts to account for the failure of revolutionary action and the triumph of fascism in the aftermath of the First World War. In addressing these problems, Gramsci was influenced as much by certain themes of the Italian political tradition as he was by his reading of Marx and Lenin. The machiavellian view that politics involves the mobilization of consent as well as the judicious employment of force, which was taken up in very different ways by contemporary Italian political sociologists such as Mosca and Pareto, and idealist philosophers such as Croce and Gentile, was particularly important in this regard. Indeed, Gramsci significantly referred to the Party as the Modern Prince (see Machiavelli, N. §6).

In common with a long tradition of nineteenth- and twentieth-century Italian writers, Gramsci believed that the unification of Italy had been incomplete. As a Sardinian, he was all too aware that the Piedmontese state had succeeded in imposing political unity upon the region but had failed to assert its hegemony so as to 'make Italians'. Gramsci thought Fascism had arisen in Italy in part because the capitalist class had not become hegemonic and so had had to rely on more directly coercive methods. A successful and enduring social revolution would have to make good this failure. Gramsci's emphasis on consent and cultural consciousness, therefore, resulted in part from reading Marx from the perspective of a characteristic Italian concern with moral unity in order to construct a *via italiana* to communism.

See also: Marxism, Wester

RICHARD BELLAMY

List of works

Gramsci, A. (1913-17) *Cronache torinese (Turin Diary)*, ed. S. Caprioglio, Turin: Einaudi, 1980.(Contains Gramsci's earliest journalism.)

Gramsci, A. (1917-18) *La città futura (The Future City)*, ed. S. Caprioglio, Turin: Einaudi, 1982. (Contains Gramsci's reaction to the Russian Revolution and the First World War.)

Gramsci, A. (1919-20a) *Il nostro Marx (Our Marx)*, ed. S. Caprioglio, Turin: Einaudi, 1984. (Contains his shift towards a redefinition of Marxism in more cultural terms.)

Gramsci, A. (1919-20b) *L'Ordine Nuovo (The New Order)*, ed. V. Gerratana and A.A. Santucci, Turin: Einaudi, 1987.(Contains the articles written during the occupation of the factories.)

Gramsci, A. (1921-22) *Socialismo e fascismo: L'Ordine Nuovo, 1921-22 (Socialism and Fascism)*, Turin: Einaudi, 1966.(Contains Gramsci's analysis of fascism and his work on the split of the Communists from the Italian Socialist Party.)

Gramsci, A. (1923-6) *La costruzione del Partito communista (Building the Communist Party)*, ed. S. Caprioglio, Turin: Einaudi, 1971.(Contains articles written as leader of the Communist Party on the strategy they should adopt.)

Gramsci, A. (1926-37) *Lettere dal carcere (Prison Letters)*, ed. S. Caprioglio and E. Furbini, Turin: Einaudi, 1965; trans. R. Rosenthal and ed. F. Rosengarten, *Prison Letters*, New York: Columbia University Press. (Written while imprisoned by Fascist authorities, these letters are an invaluable source for the evolution of Gramsci's thought at this time; definitive English translation.)

Gramsci, A. (1948-51) *Quaderni del carcere (Prison Notebooks)*, ed. V. Gerratana, Turin: Einaudi, 4 vols, 1975; trans. J. Buttigieg and A. Callari, *Prison Notebooks*, New York: Columbia University Press, 1992-. (The former is the standard Italian edition of the *Prison Notebooks*; the latter is a translation of the Gerratana edition, of which vol. 1 has so far appeared.)

Gramsci, A. (1992) *Lettere, 1908-26 (Letters)*, ed. A.A. Santucci, Turin: Einaudi.(Collects together the relatively few surviving letters from Gramsci's early years.)

Gramsci, A. (1993) *Pre-Prison Writings*, trans. V. Cox, ed. R. Bellamy, Cambridge: Cambridge University Press. (Contains a wide-ranging selection of Gramsci's journalism, Party memoranda and letters, taken from the seven volumes of pre-prison writings and letters listed above.)

References and further reading

Adamson, W.L. (1980) *Hegemony and Revolution*, Berkeley, CA: University of California Press.(A good piece of intellectual history.)

Bellamy, R. and Schecter, D. (1993) *Gramsci and the Italian State*, Manchester: Manchester University Press. (Situates Gramsci's thought in the contemporary debate between various shades of liberals, fascists and socialists about the nature and future of the Italian state.)

Clark, M. (1977) *Antonio Gramsci and the Revolution that Failed*, New Haven, CT, and London: Yale University Press.(An excellent study of the occupation of the factories.)

Femia, J. (1981) *Gramsci's Political Thought*, Oxford: Clarendon Press.(Best analysis of the concept of hegemony outlined in Gramsci's *Prison Notebooks*.)

Fiori, G. (1977) *Antonio Gramsci: Life of a Revolutionary*, trans. T. Nairn, London: Verso. (The standard biography.)

Greek philosophy: impact on Islamic philosophy

During the Hellenistic period (323-43 BC*), classical Greek philosophy underwent a radical transformation. From being an essentially Greek product, it developed into a cosmopolitan and eclectic cultural movement in which Greek, Egyptian, Phoenician and other Near Eastern religious and ethical elements coalesced. This transformation is best symbolized by the role Alexandria played as the hub of diverse currents of thought making up the new philosophy.*

When the Abbasid Caliphate was founded in Baghdad in 750 AD*, the centre of learning gradually moved to the Abbasid capital, which became in due course the heir of Athens and Alexandria as the new cultural metropolis of the medieval world. About two centuries later Cordoba, capital of Muslim Spain, began to vie with Baghdad as the centre of 'ancient learning'. From Cordoba, Greek-Arabic philosophy and science were transmitted across the Pyrenees to Paris, Bologna and Oxford in the twelfth and the thirteenth centuries.*

The initial reception of Greek-Hellenistic philosophy in the Islamic world was mixed. It was frowned upon at first as being suspiciously foreign or pagan, and was dismissed by conservative theologians, legal scholars and grammarians as pernicious or superfluous. By the middle of the eighth century AD *the picture had changed somewhat, with the appearance of the rationalist theologians of Islam known as the Mu'tazilites, who were thoroughly influenced by the methods of discourse or dialectic favoured by the Muslim philosophers. Of those philosophers, the two outstanding figures of the ninth and tenth centuries were al-Kindi and al-Razi, who hailed Greek philosophy as a form of liberation from the shackles of dogma or blind imitation (*taqlid*). For al-Kindi, the goals of philosophy are perfectly compatible with those of religion, and, for al-Razi, philosophy was the highest expression of man's intellectual ambitions and the noblest achievement of that noble people, the Greeks, who were unsurpassed in their quest for wisdom (*hikma*).*

1 The rise of Neoplatonism

Neoplatonism has been described as the final summation or synthesis of the major currents in Greek philosophy, Pythagoreanism, Stoicism, Platonism and Aristotelianism, into which an oriental religious and mystical spirit was infused (see Neoplatonism). Its founder, Plotinus, was born in Lycopolis in Egypt, studied at Alexandria and lectured in Rome. He studied with Ammonius Saccas and was a classmate of Origen, who became a Christian. Plotinus was so profoundly interested in oriental religions that he joined the abortive expedition of the Roman Emperor Gordian to Persia, as we are told by Porphyry, Plotinus' disciple and biographer, in search for 'the ways and beliefs of the Persians and Indians'. When the Emperor died on the way and the expedition came to grief, Plotinus sailed to Rome in AD 244, where he achieved great success as a teacher.

The other great representative of Neoplatonism was Diadochus Proclus, whose metaphysical outlook, like that of Plotinus, marks the final phase in the struggle of Greek paganism against Christianity at Athens, where he taught, and Alexandria, where he studied under Olympiodorus. It also marked that version of Greek philosophy which exerted a particular fascination upon Muslim minds. Other representatives of Neoplatonism during the Byzantine period include Syrianus, the teacher of Proclus, Damascius, Proclus' pupil, Simplicius and Philoponus. All those philosophers or commentators were known to the Arabs and some of their writings, sometimes lost in Greek, were translated into Arabic.

In AD 529 the Byzantine Emperor Justinian ordered the school of Athens, the last bastion of Greek paganism, to be closed. Seven of its teachers, including Syrianus and Damascius, emigrated to Persia and were well received by the Emperor Chosroes I, who was an admirer of Greek learning and the founder, in AD 555, of the School of Jundishapur which was destined to become a major centre of Greek medical and scientific studies (see Neoplatonism in Islamic philosophy §1). Subsequently, due to its proximity to Baghdad and the close political links between the Abbasids and the Persians, Jundishapur served as a staging station in the process of transmitting Greek medical and scientific learning into the Islamic world.

However, the first phase in the process of transmitting Greek learning into the Near East was the translation of theological treatises, such as Eusebius' *Ecclesiastical History* and Clement's *Recognitiones*, into Syriac. As a prelude, a series of logical texts were also translated into Syriac, including the *Isagōgē* of Porphyry, Aristotle's

Categories, *De interpretatione* and *Prior Analytics*. Beyond these, for theological reasons, the Syriac translators were not allowed to proceed.

2 Arabic translations of Greek philosophical texts

The Arab conquest of Syria and Iraq in the seventh century did not, on the whole, interfere with the academic pursuits of Syriac scholars at Edessa, Nisibis, Qinnesrin and other centres of Syriac-Greek learning. To these Christian centres should also be added Harran in Northern Syria, home of a sect of star-worshippers known in the Arabic sources as the Sabaeans and alleged to have been founded by Alexander the Great. Throughout the eighth and ninth centuries (second and third centuries AH), a new impetus was given to the translation movement thanks to the enlightened patronage of three of the early Abbasid caliphs at Baghdad, al-Mansur, Harun and his son al-Ma'mun, who founded the House of Wisdom in Baghdad to serve as a library and institute of translation. It was during the reign of al-Ma'mun that the translation of medical, scientific and philosophical texts, chiefly from Greek or Syriac, was placed on an official footing. The major translators who flourished during al-Ma'mun's reign include Yahya ibn al-Bitriq, credited with translating into Arabic Plato's *Timaeus*, Aristotle's *On the Soul*, *On the Heavens* and *Prior Analytics* as well as the *Secret of Secrets*, an apocryphal political treatise of unknown authorship attributed to Aristotle.

However, the shining star of al-Ma'mun's reign was the Nestorian Hunayn ibn Ishaq (d. AH 264/AD 873), who hailed from al-Hirah in Iraq and, jointly with his son Ishaq (d. AH 299/AD 911), his nephew Hubaysh and other associates, placed the translation of Greek medieval and philosophical texts on a sound scientific footing. The chief interests of Hunayn himself were medical, and we owe to him the translation of the complete medical corpus of Hippocrates and Galen, but Hunayn and his associates were also responsible for translating Galen's treatises on logic, his *Ethics* (the Greek original of which is lost) and his epitomes of Plato's *Sophist*, *Parmenides*, *Cratylus*, *Euthydemus*, *Timaeus*, *Statesman*, *Republic* and *Laws*. (Only the epitomes of the *Timaeus* and the *Laws* have survived in Arabic.)

The interest of Hunayn and his school in Galen, the outstanding Alexandrian physician and Platonist, is noteworthy and this philosopher-physician is a major figure in the history of the transmission of Greek learning into Arabic. Not only his sixteen books on medicine but a series of his logical and ethical writings were translated and played an important role in the development of Arabic thought. Apart from the epitomes of Plato's *Dialogues* already mentioned, his *Pinax* (list of his own writings), *That the Virtuous can Profit from Knowing Their Enemies*, *That One Should Know His Own Faults* and especially his *Ethics* have influenced moral philosophers from Abu Bakr al-Razi to Ibn Miskawayh and beyond.

Of the works of Aristotle, Hunayn's son Ishaq is responsible for translating the *Categories*, *De interpretatione*, *On Generation and Corruption*, the *Physics*, *On the Soul*, the *Nicomachean Ethics* and the spurious *De Plantis*, written by the Peripatetic philosopher Nicolaus of Damascus (first century BC). By far the most important Aristotelian treatise to be translated into Arabic during this period is the *Metaphysics*, known in the Arabic sources as the *Book of Letters* or the *Theologica* (*al-Ilahiyat*). According to reliable authorities, a little-known translator named Astat (Eustathius) translated the twelve books (excluding M and N) for al-Kindi, as did Yahya ibn 'Adi a century later. However, Ishaq, Abu Bishr Matta and others are also credited with translating some parts of the *Metaphysics*.

Equally important is the translation by Ibn Na'imah al-Himsi (d. AH 220/AD 835) of a treatise allegedly written by Aristotle and referred to in the Arabic sources at *Uthulugia* or *Theologia Aristotelis*. This treatise, which consists of a paraphrase of Plotinus' *Enneads* IV-VI, made by an anonymous Greek author (who could very well be Porphyry of Tyre), together with Proclus' *Elements of Theology* (known as the *Pure Good* or *Liber de causis*), thoroughly conditioned the whole development of Arab-Islamic Neoplatonism (see Neoplatonism in Islamic philosophy). Al-Kindi is said to have commented on the *Theologia Aristotelis* as did Ibn Sina and others, and al-Farabi refers to it as an undoubted work of Aristotle. A series of other pseudo-Aristotelian works also found their way into Arabic, including the already mentioned *Secret of Secrets*, *De Plantis*, *Economica* and the *Book of Minerals*.

Among other translators of Greek philosophical texts, we should mention Qusta ibn Luqa (d. AH 300/AD 912), Abu 'Uthman al-Dimashqi (d. AH 298/AD 910), Ibn Zur'ah (d. AH 398/AD 1008) and Ibn al-Khammar (d. AH 408/AD

1017), as well as the already-mentioned Abu Bishr Matta (d. AH 328/AD 940) and his disciple Yahya Ibn ‘Adi. None of those translators made any significant or original contribution to Arabic philosophical literature, although they laid the groundwork for subsequent developments and served as the chief purveyors of Greek philosophy and science into the Islamic world. However, there were noteworthy exceptions: Abu Bishr Matta was a skilled logician, and the Jacobite Yahya ibn ‘Adi stands out as the best-known writer on Christian theological questions and on ethics in Arabic. The Harranean Thabit ibn Qurra (d. AH 289/AD 901) was an outstanding mathematician and astronomer as well as a translator.

3 Eclecticism and the systematization of philosophical ideas

Al-Kindi, already mentioned in connection with the translation of Aristotle's *Metaphysics* and the apocryphal *Theologia Aristotalis*, should be regarded as the first genuine writer on philosophical subjects in Islam. He appears from bibliographical sources to have been a truly encyclopedic writer on every philosophical or scientific subject; his works total around three hundred, of which only a small number have survived in Arabic or Latin. He was profoundly interested in Greek thought as well as Indian religious ideas, and was a professional astrologer. His extant works, however, give the impression of eclectic and hasty composition, reflecting the influence of Socrates, Plato and Aristotle in an undefined manner. We owe to him however the first treatise on philosophical terms and definitions, modelled on Aristotle's Book Delta of the *Metaphysics*, which became in due course the model of almost all subsequent parallel treatises (see Aristotelianism in Islamic philosophy).

The first genuine system-builder in Islam, however, was al-Farabi. He was the first outstanding logician of Islam, who commented on or paraphrased the six books of Aristotle's Organon, together with the *Rhetoric* and the *Poetics*, which formed part of the Organon in the Syriac-Arabic tradition and to which the *Isagōgē* of Porphyry, also paraphrased by al-Farabi, was added. He also wrote several original treatises on the analysis of logical terms, which had no parallels until modern times. He defended Aristotelian logic against the Arabic grammarians who regarded logic as a foreign importation, doubly superfluous and pernicious (see Logic in Islamic philosophy). He also laid down the foundations of Arab-Islamic Neoplatonism in a series of writings, the best-known of which is *al-Madina al-fadila (Opinions of the Inhabitants of the Virtuous City)*. This treatise is inspired by the same utopian ideal as Plato's *Republic* (see Plato §14), but is essentially an exposition of the emanationist world-view of Plotinus to which a political dimension has been added. In that latter respect, it had hardly any impact on political developments in Islam, but it did inspire subsequent writers on political philosophy such as Ibn Bajja. Another great champion of the emanationist world-view was Abu ‘Ali al-Husayn Ibn Sina (Avicenna), who was a confessed spiritual disciple of al-Farabi.

The only great Platonist of Islam was Abu Bakr al-Razi, the greatest medical author and practitioner of the third and fourth centuries AH (ninth and tenth centuries AD). Both in ethics and in metaphysics, al-Razi exhibits a profound veneration for Plato, ‘the master of the philosophers and their leader’, as well as to the great Alexandrian doctor and philosopher Galen, whose epitomes of Plato's *Dialogues*, as we have mentioned, formed the basis of the Arabic translations of Plato's works in the third century AH (ninth century AD). Although a Manichean or Harranean influence is discernible in his thought, the ‘five eternal principles’ which form the substance of al-Razi's metaphysics and cosmology, namely the Creator (*al-Bari’*), the soul, prime matter (*al-hayula*, *hyle*) space and time, can be shown to derive from Plato's *Timaeus* and his other dialogues. The other noteworthy Platonist, especially in the realms of ethics and psychology, was a fellow Persian philosopher, Ibn Miskawayh, the best known ethical philosopher of Islam (see Ethics in Islamic philosophy §4; Platonism in Islamic philosophy).

4 The legacy of Aristotle

As already mentioned, all the works of Aristotle were translated into Arabic with the exception of the *Politics*, which for some obscure reason remained unknown to the Arabs. In addition a large number of apocryphal writings, including the *Secret of Secrets*, the *Economica*, *De plantis*, *De mundo*, the *Theologia* and the *Liber de causis* were also attributed to Aristotle. Of these works, the Organon, *On the Soul*, the *Nicomachean Ethics*, the *Physics*, the *Metaphysics*, *On Generation and Corruption* and *On the Heavens* played a decisive role in the development of logical and metaphysical ideas in Islam. However, chiefly due to the influence of the apocryphal *Theologia Aristotalis*, the tendency of early Muslim philosophers was to interpret Aristotle in Neoplatonic terms; a basic premise of this interpretation was the total agreement of Plato and Aristotle on all major issues alleged to separate them. (A famous treatise of al-Farabi entitled *al-Jam‘ bayn ra’yay al-hakimayn (The Reconciliation of*

Plato and Aristotle), probably modelled on Porphyry's lost work of the same title, illustrates this point.)

The picture radically changed with the appearance on the philosophical scene of the greatest Arab Aristotelian, Ibn Rushd of Cordoba, known in Latin as Averroes. Ibn Rushd continued the tradition of commenting on Aristotle's works initiated in Arab Spain by Ibn Bajja (Avempace) and in the East by al-Farabi. Ibn Rushd, however, produced the most extensive commentaries on all the works of Aristotle with the exception of the *Politics*, for which he substituted the *Republic* of Plato. These commentaries, which have survived in Arabic, Hebrew or Latin, earned him in the Middle Ages the title of the Commentator, or as Dante put it in *Inferno* V. 144, 'che'l gran commento feo' (he who wrote the grand commentary). Ibn Rushd actually wrote three types of commentaries, known as the large, middle and short commentaries, on the major Aristotelian treatises, notably the *Physics*, the *Metaphysics*, the *Posterior Analytics*, *On the Soul* and *On the Heavens*. In addition, he defended Aristotle against the onslaughts of al-Ghazali, the famous Ash'arite theologian, in a great work of philosophical debate entitled the *Tahafut al-tahafut (Incoherence of the Incoherence)*, a rebuttal of al-Ghazali's *Tahafut al-falasifa (Incoherence of the Philosophers)*.

5 Presocratic and post-Aristotelian philosophers

The records of the early Greek philosophers in Arabic, such as al-Sijistani's *Siwan al-hikma (Vessel of Wisdom)* and al-Shahrastani's *al-Milal wa'l nihal*, usually begin with the Seven Sages, followed by a list of the Presocratics, including Thales, Anaxamines, Anaximander, Democritus, Pythagoras and Heraclitus, with very brief accounts of their views, which are invariably given a religious and mystical twist. This is particularly true of Empedocles and Pythagoras, who are said to have received 'wisdom' (*hikma*) from Semitic sources, notably Luqman and Solomon, and to have asserted as genuine Muslims *avant la lettre* the unity of God, the creation of the world and the resurrection.

Of the post-Aristotelian philosophers, Chrysippus is almost the only Stoic philosopher mentioned; Zeno of Citium is almost unknown, despite the significant influence of Stoicism both in logic and in ethics on the early philosophers and theologians. Diogenes the Cynic (see Diogenes of Sinope), often confused with Socrates, is represented as a key moral and ascetic figure, but of the latter Stoic philosophers such as Cicero, Epictetus and Marcus Aurelius, no mention is made in the classical Arabic sources. Of the materialists, only the names of Democritus and Epicurus are mentioned. The impression one gains from reading these sources is that Greek philosophy had, according to the authors, reached its zenith with Aristotle, who superseded his predecessors in such a way as to render them superfluous. Ibn Bajja, for instance, in his paraphrase of the *Physics* of Aristotle, justifies his total omission of the views of the Presocratics on the grounds that Aristotle had refuted their 'dialectical' views and therefore it was unnecessary to dwell on them, a position which al-Ghazali had also adopted in his *Tahafut al-falasifa*. As his target, al-Ghazali chose Aristotle to the exclusion of all others, on the ground that 'he (Aristotle) had refuted all his predecessors, including his teacher, whom they nickname Plato the Divine. Then he excused himself for disagreeing with his teacher, saying: "Plato is a friend and the truth is a friend, but truth is a greater friend than Plato"' (an obvious paraphrase of *Nicomachean Ethics* I). Of Aristotle's Greek commentators, the Arab historians of ideas and philosophers refer frequently to, or quote from, the writings of Alexander of Aphrodisias, Themistius, Olympiodorus, Simplicius, Philoponus and Nicolaus of Damascus (first century BC). The earliest commentators, such as Theophrastus (d. 288 BC) and Strato (d. *c*.269 BC), were virtually unknown to the Arabs.

The chief historical significance of the Muslim-Spanish phase in the rise and development of Muslim philosophy is that it served as a major link in the transmission of Greek philosophy to Western Europe. The Muslims had been the custodians of that philosophy, which had been almost completely forgotten in Western Europe since the time of Boethius, who was responsible chiefly for translating the Aristotelian logic into Latin. By the end of the twelfth century, the translation of Arabic philosophical, scientific and medical works into Hebrew or Latin by such eminent scholars as Gerard of Cremona, Michael Scot and Herman the German (see Translators) had wrought a genuine intellectual revolution in learned circles. The most influential Muslim philosopher to leave a lasting impression on Western thought was Ibn Rushd (Averroes). During the thirteenth century, philosophers and theologians split into two rival groups, the Latin Averroists with Siger of Brabant and Boethius of Dacia at their head, and the anti-Averroist group led by Albert the Great and Thomas Aquinas (see Averroism). The confrontation between these two groups became so violent that in the second half of the thirteenth century the

Bishop of Paris, Étienne Tempier, had to intervene and in 1270 and 1277 issued ecclesiastical condemnations of a total of 219 propositions which were of Aristotelian or Averroist inspiration (see Aristotelianism, medieval). It should be noted, however, that it was thanks to the Latin translations of Ibn Rushd's commentaries that the rediscovery of Aristotle in Western Europe and the concurrent emergence of Latin Scholasticism, one of the glories of late medieval thought, were made possible.

See also: Aristotelianism in Islamic philosophy; Hellenistic philosophy; Islam, concept of philosophy in; Islamic philosophy: transmission into Western Europe; Neoplatonism in Islamic philosophy; Platonism in Islamic philosophy

MAJID FAKHRY

References and further reading

Badawi, A.-R. (1987) *La transmission de la philosophie grecque au monde arabe (The Transmission of Greek Philosophy into the Arab World)*, Paris: Vrin.(Contains useful information on manuscripts and Greek-Arabic texts or translations.)

Duval, R. (1970) *La littérature Syriaque (Syriac Literature)*, Amsterdam: Philo Press. (First published in 1899, this is a detailed account of the Syriac translations of philosophical texts.)

Madkour, I. (1969) *L'Organon d'Aristote au monde arabe (The Organon of Aristotle in the Arab World)*, Paris: Vrin.(First published in 1934, an important review of the stages through which Aristotelian logic passed in the Arab world.)

Peters, F. (1996) 'The Greek and Syriac Background', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 3, 40-51.(Discussion of some of the important features of Greek and Syriac culture as sources of Islamic philosophy.)

Pines, S. (1986) *Studies in Arabic Versions of Greek Texts*, Jerusalem: Hebrew University and Brill.(The first part consists of a series of important Greek texts in Arabic translation, with valuable comments and information.)

Shayegan, Y. (1996) 'The Transmission of Greek Philosophy into the Islamic World', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 6, 98-104.(Detailed account of how the transmission took place, paying particular attention to the Persian background.)

Steinschneider, M. (1960) *Die Arabischen Übersetzungen aus dem Griechischen (The Arabic Translations of Greek Writings)*, Graz: Akademische Druck U. Verlagsanstalt. (First published 1889, this is the earliest authoritative review in a European language of the Arabic translation of Greek texts.)

Vacherot, E. (1846-51) *Histoire critique de l'École d'Alexandrie (Critical History of the Alexandrian School)*, Paris.(The most comprehensive history of philosophical developments during the Alexandrian period.)

Walzer, R. (1962) *Greek into Arabic*, Cambridge, MA: Harvard University Press.(A collection of studies dealing with the transmission of Greek philosophy to the Arabs, by an eminent classical scholar and Arabist.)

Green political philosophy

All the major political philosophies have been born of crisis. Green political philosophy is no exception to this general rule. It has emerged from that interconnected series of crises that is often termed 'the environmental crisis'. As we enter the third millennium and the twenty-first century it seems quite clear that the level and degree of environmental degradation and destruction cannot be sustained over the longer term without dire consequences for human and other animal species, and the ecosystems on which all depend. A veritable explosion in the human population, the pollution of air and water, the over-fishing of the oceans, the destruction of tropical and temperate rain forests, the extinction of entire species, the depletion of the ozone layer, the build-up of greenhouse gases, global warming, desertification, wind and water erosion of precious topsoil, the disappearance of valuable farmland and wilderness for 'development' - these and many other interrelated phenomena provide the backdrop and justification for the 'greening' of much of modern political thinking.

The task of outlining and summarizing the state of green political philosophy is made more difficult because there is as yet no agreement among 'green' political thinkers. Indeed there is, at present, no definitive 'green political philosophy' as such. The environmental or green movement is diverse and disparate, and appears in different shades of green. These range from 'light green' conservationists to 'dark green' deep ecologists, from ecofeminists to social ecologists, from the militant ecoteurs of Earth First! to the low-keyed gradualists of the Sierra Club and the Nature Conservancy. These groups differ not only over strategy and tactics, but also over fundamental philosophy.

While there is no single, systematically articulated and agreed-upon green political philosophy, however, there are none the less recurring topics, themes, categories and concepts that are surely central to such a political philosophy. These include the idea that humans are part of nature and members of a larger and more inclusive 'biotic community' to which they have obligations or duties. This community includes both human and non-human animals, and the conditions conducive to their survival and flourishing. Such a community consists, moreover, not only of members who are alive but those who are as yet unborn. A green political philosophy values both biological and cultural diversity, and views sustainability as a standard by which to judge the justness of human actions and practices. Exactly how these themes might fit together to form some larger, systematic and coherent whole is still being worked out.

1 Political philosophy and green political philosophy

Green political philosophy is the attempt to think clearly and systematically about the political and ethical aspects of environmental issues and problems. Just as political philosophy is a sub-species of philosophy, so green political philosophy is a sub-species of political philosophy. Political philosophy is the attempt to think systematically about the ethical and political problems that arise in the course of living together (see Political philosophy, nature of). Green political philosophy is the attempt to think systematically about those problems and prospects as they are reflected and refracted through the lens of an 'ecological' or 'environmental' perspective.

An environmental perspective places the natural environment at the centre of attention and concern. It sees human beings as one of many species, and views all as interdependent and dependent on the integrity, stability and carrying capacity of the ecosystems that sustain human and non-human animals alike. Humans are members of a community whose boundaries are both wider than and different from those of conventional political communities. As Aldo Leopold puts the point in his now-classic *A Sand County Almanac* (1949: 240): 'a land ethic changes the role of *Homo sapiens* from conqueror of the land-community to plain member and citizen of it. It implies respect for his fellow-members, and also respect for the community as such'. An environmental perspective also recognizes that there are natural limits to human activity and endeavour. Thus one may not, for example, extract more energy from an ecosystem than is required to sustain it, that is, without that ecosystem ceasing to function well, or at all. 'Sustainability' thus becomes a standard for assessing the rightness or wrongness of human activities, practices and interventions in natural processes. As Leopold observes: 'A thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise' (1949: 260). (Whether this is to commit the 'naturalistic fallacy' of deriving 'ought' from 'is', is discussed in Callicott (1989).)

An environmental or green political philosophy questions a number of conventional philosophical views. It is especially critical of the anthropocentric or 'humanist' bias of much of Western philosophy, that puts human beings at the apex of creation. Whether it is because we humans are created in God's image, or are alone capable of rational thought and speech, or belong (as Kant put it) to the kingdom of ends, we have set ourselves apart from, and above, nature and its myriad species. On the conventional view, nature and non-human animals are means to human ends and endeavours. Many Western (and some non-Western) thinkers have held humans to be superior to 'lesser' creatures and have seen the natural environment as a cornucopia of 'resources' for our use and enjoyment.

If this is the conventional view, then green thinking is, by implication, unconventional. It offers alternative ways of thinking about nature and our species' place in - and responsibility for - the natural world. Much of modern green philosophizing takes the form of a critique of conventional ways of thinking about nature, as a prelude to thinking anew about our species' place in the natural order (see Ecological philosophy §3). Let us consider several specific ways in which we are asked to think anew about these matters.

2 A green theory of value

The green theory of value holds that the worth of some things does not derive solely from human assessments of their utility or beauty, and still less from their price or market value (see §5 below). Some things have intrinsic value; that is, they are valuable in and of themselves, quite apart from any human estimate of their worth or any value they might have as means to some other end. This is especially true, some greens argue (see for example Goodin 1992: 30-41), of certain natural objects or entities. For example, wilderness *per se* has no instrumental value; indeed, it is often called 'wasteland'. Locke put the point well in the second of his *Two Treatises of Government* (1690: para 42): 'land that is left wholly to nature, that hath no improvement or pasturage, tillage, or planting, is called, as indeed it is, waste; and we shall find the benefit of it amount to little more than nothing'. Nor of course is Locke alone in taking this view; he merely articulates a pervasive - and decidedly anthropocentric - view of the value of cultivated versus uncultivated land. Against Locke and other like-minded thinkers, environmentally minded philosophers maintain that wilderness has value in itself, and for the non-human creatures whose haven it is. Likewise, many of these species - the northern spotted owl, for example - have no instrumental or market value; they are not a means to any human end, nor are they sold or traded in any market. And yet all are valuable and worth protecting, for all have a place and a function in their respective ecosystems.

Greens tend to reject, with widely varying degrees of rigour, the anthropocentric view that human needs and wants supply the only standard of value or worth. They opt instead for one or another version of the biocentric view that the health and wellbeing of the biotic community - for example, an ecosystem and the myriad species it sustains - takes precedence over any of its individual members. This amounts to a conception of value that is both naturalistic and holistic; that is, it takes nature - and not one of its creatures, namely man - as the source and measure of value; and it views all creatures as part of a larger, life-sustaining whole. Some dark green thinkers, especially those calling themselves 'deep ecologists', say that this new way of thinking requires a radical shift in perspectives - roughly, from a hierarchical pyramid with humans at the apex, to an interdependent web in which humans are but one species amongst many (Devall and Sessions 1985) (see Environmental ethics).

Not all green thinkers agree, however. Indeed, some - including 'social ecologists' such as Murray Bookchin (1990) - are highly critical of any attempt to make *Homo sapiens* merely one species among many. Such a view would, they contend, diminish the value of human beings more than it would re-value nature. And such a view would fail to recognize that, like it or not, human beings, because of their knowledge and technology, have a disproportionate power over nature and its creatures, and therefore disproportionate responsibility for the health and wellbeing of both. To see the human species as co-equal with other species is to be blind to this exceedingly important - and indeed unique - aspect of human existence.

This in turn introduces another 'green' theme: the use and abuse of scientific knowledge and technology. Much of Western thought since the scientific revolution of the seventeenth century has tended to celebrate the increasing power of the human species over nature. Sir Francis Bacon and other seventeenth-century philosophers saw science and technology as means of mastering or dominating nature for human ends (Leiss 1972). Variations on this view can be found among later thinkers. Karl Marx, for one, looked forward to the pacification or 'humanization' of nature which modern science made possible. The productive forces - roughly, natural resources and the technology used to turn them into objects useful to humans - that were developed under capitalism have

transformed nature beyond all recognition, and that, Marx thought, was a commendably progressive development (see Marx, K.).

Right and left, capitalist, communist, conservative or liberal, Western political philosophers have for the most part embraced and celebrated the 'conquest' or 'pacification' of nature for human purposes. Green political philosophers have, by contrast, been highly critical of any philosophy that views nature only as a 'resource base' or a means to human ends and is, in consequence, heedless of the conditions conducive to the health and wellbeing of nature's myriad species.

In different ways and with different emphases, green political philosophers generally subscribe to what one might call a 'systemic' or 'ecological' view of value. Some things, as noted already, have intrinsic value as ends in themselves. Others, by contrast, have value by virtue of the contribution they make to a larger whole. The value of such a thing is determined by its place in, and contribution to, some larger functional whole - an ecosystem or (in Leopold's phrase) a biotic community. Thus, for example, certain predator species - for example wolves - have value not only in themselves but because of their function within the ecosystem of which they are an integral part. Wolves cull sick, lame or deformed deer; by preventing the weaker members of the deer population from reproducing, wolves actually benefit that species. And by controlling the deer population wolves protect the larger ecosystem which both share with other species. This green view of value is directly contrary to the conventional view - long established in law and public policy in North America - that wolves and other predator species should be trapped or shot (and bounties paid to trappers and hunters) because predation is cruel, wanton, wasteful and without value to human beings.

But what, exactly, is 'political' about predators and prey, about anthropocentric versus biocentric conceptions of value, and the like? The most obvious answer, of course, is that laws, rules, regulations and government policies on environmental matters must be made on the basis of our beliefs about nature and its creatures. These include laws and public policies concerning mining and forestry, tourism and economic development, land use and property rights, motorway construction, parks and recreation, re-forestation, the preservation of wilderness, the protection of endangered species, and many other matters. Some of the most heated and hard-fought political battles of the modern age have been concerned with these and other broadly environmental issues.

These matters are also 'political' in a second, and much older, sense. For politics is concerned (as Aristotle famously put it) with the good life and the conditions conducive to it. But while Aristotle was concerned with the good life as lived by (some) human beings, modern greens cast their net more widely, to include non-human creatures and the biotic communities they share with human beings. And still, in our day as in Aristotle's, what constitutes or counts as the good life is hotly disputed amongst philosophers.

Contemporary political debates and divisions over environmental issues are, as often as not, based on philosophical differences over value (anthropocentric versus biocentric; instrumental versus intrinsic), over what counts as the good life, over the proper place and role of human beings in the natural order, over obligations to non-human creatures and future humans, and so on. A green political philosophy is simply the attempt to articulate and justify a biocentric view of the good life for human beings and for other creatures with whom they share a common planet.

What will count as the good life for fish will of course differ from what counts as the good life for monkeys. But what all creatures share is an interest in a healthy habitat. Fish and frogs have an interest in clean or unpolluted water, monkeys in intact tree canopies, whales in plankton, koala bears in eucalyptus trees, and so on. They need not know or be consciously cognizant that they have an interest in these things in order to actually have an interest, since being aware that one has a need for or interest in X is not a necessary condition for having an interest in X (see Needs and interests §1).

Humans are of course able to know, as non-human animals are not, what is required to sustain non-human species. This confers upon humans the 'epistemic responsibility' that comes with knowing what members of other species do not or cannot know. Our knowing what conditions are conducive to some species' survival and flourishing confers upon our species an added measure of responsibility. More specifically, it requires that we recognize, and not wantonly disregard, the interest such a species has in the conditions conducive to its survival and flourishing (Johnson 1991, ch. 6). Our ever-expanding knowledge of the natural world brings with it an expanded

responsibility to recognize the interests (some greens go further, and say ‘rights’) of other creatures. And while we may not always be able to promote these interests, we must, as moral and political agents, at least accord them serious consideration (see Animals and ethics).

A green political philosophy expands the circle of value and moral concerns not only spatially, so to speak - to include non-humans - but also temporally, to include generations of human beings who are as yet unborn. A green political philosophy - and a green theory of value in particular - places posterity on a par with people now living. Our distant descendants are our moral equals, their happiness and wellbeing as valuable as our own (see Future generations, obligations to).

3 Green political agency

If the green theory of value sets the ends, the green theory of political agency offers an account of the means to achieve those ends. A theory of agency operates at two levels. The first, and more fundamental, level is individual agency, which specifies the characteristics of individual agents; the second, and derivative, level is that of collective agency, which describes the main features of organizations and institutions within which individual agents work. Taken together, these two features of a theory of agency supply an account of the kinds of actors and institutions that would be required to achieve the ends stipulated by a green theory of value (Goodin 1992, ch. 4). Two questions are particularly pertinent. The first is: what kinds of agents - that is, political actors or citizens, consumers and so on - are best suited for realizing green ends? The second is: what kinds of political organizations or institutions, what form of government, and what strategies and tactics are most likely to implement green values?

First, at the individual level: green agents or political actors must be motivated by a love of and respect for the natural world, of which they are a small but important and morally responsible part; their satisfactions and pleasures will not, in the main, be materialistic; their wants will be satisfiable in sustainable ways; they will act in nonviolent ways; their time horizon will typically extend further than their own and one or two adjoining generations.

There is, as some have noted, nothing all that new or novel in this picture of green political agency. It is a picture painted, with many minor variations, by philosophers from Plato onward, of the good man and the good citizen motivated by a vision of the good life (O’Neill 1993, ch. 1). The good life for humans consists of appreciating the superiority of spiritual and intellectual satisfactions over material ones, of getting outside ourselves, as it were, to care about something larger and longer-lived than our own mortal selves. And how better to do this, greens ask, than recognizing our place in, and responsibility to care for, the natural world? (Passmore 1980).

A second set of questions concerns collective agency. What kinds of political organizations, institutions and strategies might best achieve green goals? Here again green thinkers differ amongst themselves. Most greens tend to believe that the most desirable and effective institutions are broadly democratic, decentralized and participatory (see for example Porritt 1984; Dobson 1990). Others are more sceptical, suggesting that solutions to environmental problems require the kind of coordination that can come only through the modern state, its allied agencies and cooperation between states (for example, Ostrom 1991; Goodin 1992). Others are more sceptical and despairing still, arguing that environmental crises are apt to be so numerous, so pervasive and so severe as to require the harsh interventions of an authoritarian, hierarchical, and not necessarily democratic, state (for example Heilbroner 1975; Ophuls 1977; Catton 1980). Such dire predictions and prescriptions, in combination with the radical beliefs of some militant environmentalists, have led some critics to suggest that there is a historical, if not a logical, link between green political thought and fascist or Nazi-like political practice (Bramwell 1989; Pois 1986).

This last criticism relies for its force on a single historical case. Certainly it is true that some Nazis did indeed embrace green themes, values, beliefs and attitudes not unlike those held by some amongst the more militant and mystical greens (nature goddess worship, a marked hostility toward the natural sciences and so on). And this fact should give self-critical and reflective greens some pause. But it does not follow, of course, that anyone who subscribes to green values is a Nazi. To claim otherwise is analogous to arguing that: ‘Nazis loved dogs; therefore dog-lovers are Nazis.’

A more serious set of questions concerns the shape and structure of green political institutions. For example, are the interest-group politics of Western liberal democracy well-suited to achieving environmental ends? Ought

greens to organize themselves into parties and pressure groups? If so, should they nominate candidates for election to public office and lobby on behalf of their green agenda? Or should greens remain a broad-based movement, aloof from party politics and pressure-group tactics?

Such questions admit of no easy answers and are hotly debated amongst greens. In two-party systems such as the USA, third parties have little hope of success, and the best bet might be to try to influence the platforms and policies of the major parties. In multi-party parliamentary systems, by contrast, it might be more rational to organize green political parties - as has been done in the UK and Germany, for example, albeit without notable electoral success.

Some of the more militant greens hold that conventional party politics will not suffice and that direct action - civil disobedience, protest marches and demonstrations, even ‘ecotage’ or ‘monkey-wrenching’ - is required (see Civil disobedience §2). Such action is politically and morally justifiable, its defenders say, if it is nonviolent. Thus, for example, motorway construction and the destruction of wildlife habitat may be slowed or stopped by nonviolent civil disobedience. By drawing public attention to such issues, protestors hope to educate or ‘raise the consciousness’ of their fellow citizens so as to put pressure on their political representatives, planners and policy makers. Some more militant green activists advocate such tactics as monkey-wrenching (the disabling of machinery), the destruction of surveyors’ stakes, the ‘spiking’ of old-growth trees to prevent their being cut. The arguments for and against such measures have been a lively topic of political and philosophical debate amongst environmentalists (Goodin 1992: 133-5).

Arguments about strategies, tactics, and institutions are arguments about agency, both individual and collective, and constitute an important aspect of green political philosophy.

4 Green economic thought

A green view of economic relations is an integral part of a green political philosophy, its theory of value, and its theory of individual and collective agency. In the main, green economic thinking represents a critique of, and a challenge to, much of conventional market-based economic thought - especially in so far as such thinking forms the basis and justification for present-day policies regarding energy and other broadly environmental issues.

What is wrong with modern economic thought, green critics contend, is that it knows the price of everything and the value of nothing. Or, perhaps more accurately, it is mistaken in assuming without argument that everything has a price, even if some things are not actually traded in any market. Three features of green economic thought merit special mention. The first is its critique of, and alternative to, the practice of assigning ‘shadow prices’ to goods not traded in markets. The second concerns the practice of ‘discounting’ the welfare of future generations via a ‘social discount rate’. And the third is the green questioning of the use of cost-benefit analysis in making and justifying political decisions and policies.

Conventional economic thinking translates the question, ‘What is the value of X?’ into another question, ‘What is the price of X?’ That is, what would X be worth in monetary terms, if it were to be traded in a market? Now clearly, since some things are not bought and sold, some way must be found for determining their price. Some economists contend that so-called ‘shadow prices’ can be assigned to such goods as clean air, scenic beauty, the preservation of a particular species, and so on. The hypothetical price is determined by asking people what they would be ‘willing to pay’ to (for example) preserve the Grand Canyon or to prevent the extinction of an entire species of plant or animal. By this means it should be possible to ascertain the economic value of these and other ‘environmental’ goods.

The problem with the practice of shadow pricing, say critics (see for example Sagoff 1988), is that it cheapens things that are beyond price, that is to say, literally priceless. To ask what someone would be willing to pay for being treated with respect would thereby demean the very idea of kindness and respect. Some things actually lose value (and, arguably, all meaning) when they are bought and sold. Such is the case, critics argue, when a price, even a hypothetical one, is put on species, ecosystems and other natural entities.

A second, and scarcely less controversial, practice is that of discounting the wellbeing of future generations by means of the social rate of discount. Roughly, the idea is this: just as individuals discount their own future, so too does an entire society at some time discount its future members’ welfare at all later times. And, just as it is rational

for individuals to discount their own future wellbeing, so it is rational for one generation to discount the welfare of future generations.

The green critique of social discounting is easily summarized. It is one thing to discount one's own future wellbeing; it is a morally much more questionable matter to discount other people's wellbeing. I am not entitled, rationally or morally, to discount your future wellbeing at my personal rate of discount. And yet that is precisely what defenders of social discounting attempt to do. One's moral worth does not vary according to one's place in the temporal order of succession. I am not entitled to discount your wellbeing, whether you are my contemporary or my very distant descendant. The practice of social discounting, its critics claim, works to the distinct disadvantage of future generations and is clearly unjust (Cowen and Parfit 1992; Cowen 1992; O'Neill 1993, ch. 4).

The practice of social discounting, when combined with other economic tools or techniques, such as cost-benefit analysis, further disadvantages future people (O'Neill 1993, chs 4 and 5). Consider, by way of example, the claim that nuclear power is preferable to other alternatives because it has a higher benefit-to-cost ratio. In practice, however, the benefits - including access to cheap and plentiful electricity - accrue to those now living, whilst the costs will be borne by future people, in the form of increased risk of radiation exposure. To measure the benefits over the short term and the costs over the longer term is systematically to disadvantage future people.

It is important to note that this is not an argument against cost-benefit analysis *per se* or in principle, but against taking a too-constricted time horizon over which to measure costs and benefits. If the wellbeing of future people is not discounted, and the benefits not enjoyed exclusively by one generation while the costs are borne by another, then cost-benefit analysis can be a useful tool of analysis, even for environmentally minded policy analysts, legislators and concerned citizens.

Greens tend to be critical of conventional economic thinking, and particularly of the view that market allocations of valued goods are always fair or necessarily just. This does not mean, however, that all greens are therefore socialists of one or another stripe (although some certainly are). Green economic thought is at present more critical than constructive - more articulate about what is wrong with modern market economics than what alternative system might be devised (see Market, Ethics of the).

5 Conclusion

The task of a green political philosophy is twofold. It is critical in that it aims to expose and criticize the flaws, fallacies and contradictions that lurk in conventional ways of thinking about human beings and their proper place in the natural world. And it is constructive in that it aims to sketch the contours and outline the institutions of a society that values the natural world, practises sustainability and concerns itself with the longer-term consequences of present-day policies and practices. Much of the critical groundwork has now been laid, but the task of constructing a coherent, systematic and persuasive political philosophy has only just begun, and remains a task for the twenty-first century.

See also: Ecology; Environmental ethics

TERENCE BALL

References and further reading

Berry, W. (1981) *The Gift of Good Land*, San Francisco, CA: North Point Press.(A thoughtful, provocative and beautifully written book about agriculture, conservation, and human community and responsibility.)

Bookchin, M. (1990) *Remaking Society: Pathways to a Green Future*, Boston, MA: South End Press.(A philosophical and programmatic statement by a leading social ecologist.)

Bramwell, A. (1989) *Ecology in the 20th Century: A History*, New Haven, CT: Yale University Press.(A very critical history.)

Callicott, J.B. (1989) *In Defense of the Land Ethic*, Albany, NY: State University of New York Press.(A collection of essays on issues and topics deriving from Leopold's 'land ethic'. Includes an account of American Indian environmental ethics.)

Catton, W.R. (1980) *Overshoot: The Ecological Basis of Revolutionary Change*, Urbana, IL: University of Illinois Press.(Argues that overpopulation threatens to exceed or 'overshoot' the earth's carrying capacity.)

Cowen, T. (1992) 'Consequentialism Implies a Zero rate of Intergenerational Discount', in P. Laslett and J.S.

Fishkin (eds), *Justice Between Age Groups and Generations*, 'Philosophy, Politics, and Society', series 6, New Haven, CT, and London: Yale University Press, ch. 8.(Argues on consequentialist grounds against applying any social rate of discount.)

Cowen, T. and Parfit, D. (1992) 'Against the Social Discount Rate', in P. Laslett and J.S. Fishkin (eds), *Justice Between Age Groups and Generations*, 'Philosophy, Politics, and Society', series 6, New Haven, CT, and London: Yale University Press, ch. 7.(A critique of the very idea of a social rate of discount.)

de-Shalit, A. (1995) *Why Posterity Matters: Environmental Policies and Future Generations*, London: Routledge.(An intelligent discussion of a notoriously vexed topic.)

Devall, B. and Sessions, G. (1985) *Deep Ecology*, Salt Lake City, UT: Peregrine Smith Books.(A wide-ranging critique of anthropocentrism and a defence of a 'deep ecology' ethic.)

Dobson, A. (1990) *Green Political Thought*, London: Unwin Hyman.(A useful introduction.)

Dryzek, J. (1987) *Rational Ecology*, Oxford: Blackwell.(An attempt to apply rational choice theory to environmental decision making.)

Foreman, D. (1993) *Confessions of an Eco-Warrior*, New York: Crown.(A passionate and sometimes amusing analysis of a number of issues by a leading radical environmentalist.)

Goodin, R.E. (1992) *Green Political Theory*, Cambridge: Polity Press.(A clear and systematic account of what a rationally reconstructed green political philosophy might look like.)

Heilbroner, R.L. (1975) *An Inquiry into the Human Prospect*, New York: W.W. Norton; revised 1975.(An economist's sombre inquiry into the trends and tendencies that are likely to lead to environmental disaster and political repression.)

Johnson, L.E. (1991) *A Morally Deep World: An Essay on Moral Significance and Environmental Ethics*, Cambridge: Cambridge University Press.(Argues that animal species and ecosystems have morally considerable interest that humans must take into account.)

Laslett, P, and Fishkin, J.S. (eds) (1992) *Justice Between Age Groups and Generations*, Philosophy, Politics, and Society, series 6, New Haven, CT, and London: Yale University Press.(A collection of essays about intergenerational issues.)

Leiss, W. (1972) *The Domination of Nature*, New York: George Braziller.(Explores the theme of the domination or 'conquest' of nature in Western thought.)

Leopold, A. (1949) *A Sand County Almanac*, Oxford and New York: Oxford University Press; repr. New York: Ballantine Books, 1970.(The concluding section - 'The Land Ethic' - is widely considered to be a modern classic of environmental thinking.)

Locke, J. (1690) *Two Treatises of Government*, repr. Cambridge: Cambridge University Press, 1992.(Propounds a theory of property according to which ownership and value are determined by the labour expended in obtaining and transforming land and other natural goods.)

Nash, R.F. (1989) *The Rights of Nature: A History of Environmental Ethics*, Madison, WI: University of Wisconsin Press.(Holds that the history of environmental ethics is the story of an ever more extensive widening of the circle of moral considerability - roughly, from the ascription of 'natural rights' to humans to the 'rights' of nature itself.)

O'Neill, J. (1993) *Ecology, Policy and Politics: Human Well-being and the Natural World*, London: Routledge.(A sophisticated and wide-ranging analysis of a variety of issues ranging from green value theory to green economics.)

Ophuls, W. (1977) *Ecology and the Politics of Scarcity: Prologue to a Theory of the Steady State*, San Francisco, CA: W.H. Freeman.(A defence of no-growth or 'steady-state' economy and society.)

Ostrom, E. (1991) *Governing the Commons*, Cambridge: Cambridge University Press.(Analyses strategies for bringing those resources we have in common under rational and responsible collective control.)

Partridge, E. (ed.) (1981) *Responsibilities to Future Generations*, Buffalo, NY: Prometheus Books.(An excellent anthology.)

Passmore, J. (1980) *Man's Responsibility for Nature*, London: Duckworth, 2nd edn.(Defends the view that humans have a special species-specific responsibility to protect and preserve the natural environment.)

Pois, R.A. (1986) *National Socialism and the Religion of Nature*, London: Croom Helm.(A historical exploration of 'environmental' themes in Nazi theory and practice.)

Porritt, J. (1984) *Seeing Green: The Politics of Ecology Explained*, Oxford: Blackwell.(A useful primer on the political theory and practice of the green movement.)

Sagoff, M. (1988) *The Economy of the Earth: Philosophy, Law, and the Environment*, Cambridge: Cambridge University Press.(A wide-ranging critique of conventional 'economic' thinking about environmental issues.)

Worster, D. (1994) *Nature's Economy: A History of Ecological Ideas*, Cambridge: Cambridge University Press,

2nd edn.(A useful survey of key themes and thinkers.)
Sikora, R.I. and Barry, B. (eds) (1978) *Obligations to Future Generations*, Philadelphia, PA: Temple University Press.(A collection of essays about intergenerational justice and related issues.)
Taylor, P.W. (1986) *Respect For Nature: A Theory of Environmental Ethics*, Princeton, NJ: Princeton University Press.(A philosophical inquiry into the foundations of environmental ethics and human responsibility.)
Wenz, P. (1988) *Environmental Justice*, Albany, NY: State University of New York Press.(Attempts to formulate a theory of justice that takes animals and the ecosystem into account.)

Green, Thomas Hill (1836-82)

Green was a prominent Oxford idealist philosopher, who criticized both the epistemological and ethical implications of the dominant empiricist and utilitarian theories of the time. He contended that experience could not be explained merely as the product of sensations acting on the human mind. Like Kant, Green argued that knowledge presupposes certain a priori categories, such as substance, causation, space and time, which enable us to structure our understanding of empirical reality. Physical objects and even the most simple feelings are only intelligible as relations of ideas constituted by human consciousness. However, unlike Kant, he did not draw the conclusion that things in themselves are consequently unknowable. Rather, he argued that reality itself is ultimately spiritual, the product of an eternal consciousness operating within both the world and human reason. Green adopted a similarly anti-naturalist and holistic position in ethics, in which desires are seen as orientated towards the realization of the good - both within the individual and in society at large. In politics, this led him to criticize the laissez-faire *individualist liberalism of Herbert Spencer and, to a lesser extent, of J.S. Mill, and to advocate a more collectivist liberalism in which the state seeks to promote the positive liberty of its members.*

1 Life and works

Green was born in Birkin, Yorkshire, on 7 April 1836, the fourth child of the Reverend Valentine Green. His evangelical religious views played a profound part in his later philosophy, alienating the orthodox but having a great appeal for those seeking a rational basis for Christianity that was compatible both with Darwinism and the new historical scholarship, which were then calling traditional beliefs into question, and which issued in a broadly humanist social ethic. He was educated first by his father and then, from the age of 14, at Rugby School, where the future Cambridge philosopher Henry Sidgwick was also a pupil. He went up to Balliol in 1855, coming under the influence of Benjamin Jowett, who introduced him to German as well as Greek philosophy. He took a First in Greats in 1859, and was elected to a college fellowship the following year. In 1863 he refused the editorship of the *Times of India* and in 1864 was an unsuccessful candidate for the chair in philosophy at the University of St Andrews, largely due to his religious opinions. He served as an assistant to the Schools Inquiry Commission in 1865 and 1866, and was later involved in the founding of Oxford School for Boys. From 1866, he became heavily involved in college business and was the first Tutor not to be enrolled in holy orders. Together with Jowett, he played a prominent role in the university reforms that were to orientate the curriculum towards the needs of modern society and to professionalize academia. Green has been described as the first professional philosopher in the modern sense, and he influenced the emergence of a new kind of hard-working undergraduate, who viewed a degree as a qualification for a career, with Greats in particular becoming a training for politics or the imperial civil service.

The Philosophical Works of David Hume, edited with T.H. Gross, appeared in 1874-5. Green’s long introduction contained his critique of empiricism and naturalism in Locke and Hume, an argument that was extended to J.S. Mill,Spencer andLewes in subsequent articles. He had failed to get elected as Waynflete Professor in 1867, but in 1878 became Whyte’s Professor of Moral Philosophy. He took his lecturing duties seriously and his best-known books, the *Prolegomena to Ethics* (1883) and the *Lectures on the Principles of Political Obligation* (1886), were both published posthumously on the basis of his lecture notes. Many of his other essays, including the important ’On the Different Senses of "Freedom" as Applied to Will and to the Moral Progress of Man’, similarly originated as lectures and likewise only appeared with the publication of his collected works after his death.

Green played an active part in radical liberal politics, supporting John Bright and the cause of the North in the American Civil War. He was a prominent member of the temperance movement from 1872, and in 1876 was the first Oxford academic to be elected to the city council as a representative of the town rather than the University. Together with his wife Charlotte Symonds, whom he married in 1871, he participated in the University extension movement and in attempts to open up higher education to women. His lecture *Liberal Legislation and Freedom of Contract* (1881) has often been taken as providing the philosophic basis for the shift from *laissez-faire* to social liberalism at the turn of the century. He died prematurely on 26 March 1882.

2 Critique of empiricism and naturalism

Green first presented his own theory in terms of a critique of the native empiricist tradition of Locke,Hume and

their nineteenth-century followers. For Green, the heart of the empiricist project lies in the contention that ideas arise from sensation. According to this thesis, at least as Green understood it, a distinction is made between those ’simple’ ideas that the mind passively receives through the sensory mechanisms, and ’complex’ ideas that result from reflection upon these basic notions and the combinations which the mind performs between them. The crux of Green’s criticism consisted in denying the validity of this distinction. Drawing on Kant, he claimed that what he calls ‘formal conceptions’, such as subject and object, or cause and effect, cannot be derived from simpler ideas stemming directly from sensation because even our most basic notions presuppose these more complex concepts. For example, knowledge of cause and effect cannot be acquired (as Green held that Locke maintained) through reflection on the circumstances under which things come to exist, because the very idea of an external thing can only be derived from sensations if we already conceive of them as being caused by outer objects. A conception of causation is therefore intrinsic to our knowledge of those very objects from which we are supposed to obtain it. Green found additional evidence of the circularity of the empiricist argument in the tendency to equivocate over whether simple ideas are the result of mere feelings or feelings that already have certain properties built into them, such as notions of space or blueness. If the latter, then the empiricist has simply assumed what needs to be explained. If the former, then there remains the difficulty that the simple ideas from which we allegedly develop our more complex ideas seem to require such notions for us to be able to conceive of them in the first place.

From this critique, Green drew the conclusion that conscious experience could not be accounted for in terms of a succession of passively received feelings that act upon us in much the same way as certain climatic or chemical changes act on a plant and cause it to pass through a number of different physical states. Mind or consciousness had to be ascribed an active role from the beginning. On this view, feelings or sensations can only be experienced in the relevant sense of constituting knowledge if they are present to a self-conscious mind that is distinct from them and is capable of relating them to each other. Green insisted that one could not even conceive of a present and immediate sensation without relating it to other general notions of immediacy, presentness and sensation that denied its particular character.

This emphasis on relations is vital to Green’s argument and was later taken up by F.H. Bradley (see Bradley §5). For Green, ’formal conceptions’ or any kind of general notion are essentially relations. However, this does not imply that there are ideas or objects that exist independently of relations. Thought does not simply relate things that are already there. His thesis is essentially holistic, in which a single self-consciousness mind is regarded as constituting the world as part of a unified system of mutually related and entailed elements. Thus, particulars always stand in relation to something else, and are real only in so far as they form part of those relations. Likewise, general categories exist only in so far as they comprise relations. This inter-relatedness of universal and particular is generally known as the doctrine of the concrete universal.

Much of Green’s argument can be characterized as broadly Kantian. Indeed, Green identified his principle of consciousness with Kant’s synthetic unity of apperception. However, he went beyond Kant in denying that there are ’things-in-themselves’ that mind plays no role in constituting (see Kant, I. §5). Green believed that in this respect Kant had failed to take his critique of Humean empiricism far enough. He held that Kant regarded the noumenal world as totally inaccessible to thought, and charged this position with incoherence on the grounds that such a world would be unknowable and so it would be impossible to say whether it existed and hence was antithetical to thought in the first place. Denial of the noumenal world, however, raises the problem of solipsism. If mind constitutes reality, does that mean that there are as many worlds as there are individual minds? Or do our individual minds somehow collaborate together to create the phenomenal world? Green rejected both these suggestions to argue that reality itself is the product of a single, eternal self-conscious mind which manifests itself, albeit incompletely, in individual human minds. Green’s religious beliefs play a major role in his argument at this point - in particular, the view that the essence of Christianity is that human beings partake in some measure in the divine spirit. As we saw, Green contended that this single eternal Mind unifies the whole of reality, relating every part to every other part. No single element exists or can be known, therefore, except in relation to everything else. As a consequence, human knowledge will always be incomplete, although for the divine consciousness, which is omniscient, all relations are known and hence, Green believed, necessary.

3 Ethics and politics

Green’s ethics follows on from his critique of naturalism. Here his target is utilitarianism, which he regarded as

intimately bound up with empiricist epistemology and a hedonist psychology. A utilitarian morality of this kind holds that human beings are motivated by the pursuit of pleasure and the avoidance of pain. However, Green argued that even those forms of happiness that originated in purely animal wants or impulses could only become objects of conscious human desire if they are related to an ideal of self-perfection describing the sort of person one ought to be and a corresponding conception of the good. We cannot derive our view of morality from pleasure, any more than we can acquire other forms of knowledge from different kinds of sensations. The experience of pleasure presupposes notions of good and bad in exactly the same way as different sensations are bound up with other ‘formal conceptions’.

Green maintained that we can only conceive of and actively pursue self-perfection within a society, for our sense of personal identity arises out of our relations with others, for instance in the various professional, familial and other social roles which we perform or to which we aspire. As a result, the goals which we come to pursue will be similarly tied up with how we relate to others. The only good that can foster and preserve the relations between individuals is a common good which is non-competitive in nature and the attainment of which by each promotes that of everyone else. Such a good cannot be conceived in material terms but is rather a state of mind or character. Toleration provides an example of what Green had in mind, since tolerance is a good that can only be understood as a social relation between two or more persons, and each individual’s aiming to be tolerant fosters the ability of everyone else to be so.

The origins of this argument are firmly located in his metaphysics and his holistic account of reality as a system of necessary relations. It is important to stress that Green never treats society as a collective subject. He regarded any suggestion that society and the common good might be distinct from or superior to its individual members and their particular goods as being as meaningless as attempts to establish the converse.

Green’s conception of positive liberty, his main contribution to political philosophy, is closely tied to his ethics. Green associated a purely negative conception of freedom, as the absence of external physical constraints, with a naturalistic view of human beings as driven by innate impulses and appetites which they sought to satisfy in the pursuit of various material goods. As we have seen, he disputed this conception of human motivation. He also believed that it issued in an atomistic and ultimately incoherent view of society, since it failed to offer a satisfactory account of social relations. Negative liberty was the freedom of Hobbes’ state of nature, of the war of all against all. Positive freedom, in contrast, was the attempt to secure the conditions for the self-realization of all. It was orientated towards a common, moral good, rather than purely individual, material goods. Green noted that both negative liberties and certain material conditions formed preconditions for the attainment of this moral freedom. A moral action had to issue from a free will and be voluntarily performed on Green’s somewhat Kantian account, so that individuals needed to be protected from undue coercion by others. However, he also claimed that some obstacles were insurmountable by individual action alone, so that public education and welfare legislation were also necessary.

In this way, Green’s ethical theory fed into a progressive liberal politics, and he defended measures such as the Factory Acts, regulating conditions at work, on these grounds. He argued that properly understood freedom of contract entailed a mutual recognition of positive rights guaranteeing certain standards of safety, remuneration and limiting hours rather than just the negative right to sell and hire labour freely. He rejected both the minimal state of *laissez-faire* liberals such as Spencer and the conception of the state as the monopoly holder of force, which he associated with Hobbes and J. Austin (see Liberalism §1). Rather, he argued that the state’s role consists in coordinating the complex relations between the various institutions and individuals that comprise it. Our obligation to the state is moral and stems from a consciousness of its function in promoting certain common ends. Rights are neither natural properties of individuals nor positive enactments of the sovereign power, but grow out of a mutual awareness of those common goods necessary for the fullest degree of self-realization by all. He held that our understanding of the nature and purposes of the state has steadily broadened, as our consciousness of the common good and the resulting rights stemming from the relations between individuals has developed. Some governments may move more slowly than others in responding to this greater awareness among the people, and in such cases civil disobedience or even revolution might be justified. Green used these arguments to support the North in the American Civil War, Giuseppe Mazzini’s movement for national liberation in Italy, and the campaign for suffrage reform in Britain.

Critics, such as Isaiah Berlin, have complained of two major problems with Green's advocacy of positive liberty. First, they have argued that there are a number of different moral goals that it may be rational to pursue and these are not always compatible or commensurable with each other. Second, they have noted that the welfare goods advocated by Green are capable of being objects of competition. Education, for example, can give someone a positional advantage when it comes to getting a given job. Both these objections suggest that social relations are liable to be less harmonious than Green supposes and that an attempt to coordinate all forms of self-realization will either prove incoherent or totalitarian. Green disputed the validity of these sorts of criticism at both a metaphysical and a practical level. While he asserted that ultimately all forms of genuine human self-perfection prove mutually supportive, he insisted that full consciousness of these relations was only accessible to God. However, he did believe that human beings were capable of making progressive steps towards such a perfected state. Nevertheless, moves in this direction always had to come from below, since morality could not be imposed.

It is undeniable that an optimistic teleology underlies his thesis, in which the widening area of state action and social cooperation results from the gradual diffusion of reason and self-consciousness among the general public. The ambivalences of Green's position come out clearly when the people failed to move in the direction he desired. One prime example is temperance reform, for which Green ultimately abandoned schemes for voluntary abstinence and advocated the compulsory prohibition of alcohol.

Green's political writings had an important impact at the time, influencing a whole generation of British liberal politicians and theorists, such as L.T. Hobhouse and J. Hobson. His arguments, however, are deeply embedded in a metaphysics the religious assumptions of which came to seem increasingly dated and questionable.

See also: Empiricism; Freedom and liberty §3; Hegelianism §5; Naturalism in ethics

RICHARD BELLAMY

List of works

Green, T.H. (1883) *Prolegomena to Ethics*, ed. A.C. Bradley, Oxford: Clarendon Press.(This book is not contained in the complete works.)

Green, T.H. (1997) *Works*, ed. P. Nicholson, 5 vols, Bristol: Thoemmes Press.(Volumes 1-3 reproduce the Nettleship edition of the Works, volume 4 the 1883 edition of the *Prolegomena to Ethics*, which was not included in Nettleship's *Works*, and volume 5 reprints all previously uncollected published material, including letters and selections from Green's papers, plus a new bibliography and an introduction by the editor.)

Green, T.H. (1885-8) *The Works of Thomas Hill Green*, ed. R.L. Nettleship, London: Longman's, Green, 3 vols. (Volume 3 contains a valuable 'Memoir' by the editor.)

Green, T.H. (1986) *Lectures on the Principles of Political Obligation and Other Writings*, ed. P. Harris and J. Morrow, Cambridge: Cambridge University Press.(The definitive edition of the *Lectures*, corrected against the original manuscript, together with extracts from the *Prolegomena* that have been similarly checked, and some previously unpublished essays and other important political articles.)

References and further reading

Bellamy, R.P. (1990) 'T.H. Green and the Morality of Victorian Liberalism', in R.P. Bellamy (ed.) *Victorian Liberalism: Nineteenth-Century Political Thought and Practice*, London: Routledge.(Locates Green's thought in the context of his times, arguing that he was not so innovative or radical as is often thought.)

Bellamy, R.P. (1992) 'T.H. Green, J.S. Mill, and Isaiah Berlin on the Nature of Liberty and Liberalism', in H. Gross and R. Harrison (eds) *Jurisprudence: Cambridge Essays*, Oxford: Clarendon Press.(Defends Green against the criticisms of Berlin.)

Berlin, I. (1969) 'Two Concepts of Liberty', in *Four Essays on Liberty*, Oxford: Oxford University Press.(Explicit criticism of Green is found on 133 and 150, but he associates him with the errors of the concept of positive liberty in general.)

Hylton, P. (1990) *Russell, Idealism and the Emergence of Analytical Philosophy*, Oxford: Clarendon Press, ch. 1. (A fine account of Green's critique of naturalism, which forms the basis of §2 above.)

Nicholson, P.P. (1990) *The Political Philosophy of the British Idealists*, Cambridge: Cambridge University Press, Studies 2-5.(A critical defence of various aspects of Green's political thought.)

Pucelle, J. (1961, 1965) *La Nature et l'esprit dans la philosophie de T.H. Green (Nature and Mind in the Philosophy of T.H. Green)*, Paris and Louvain: Beatrice Nauwellaerts Editions, 2 vols.(The most extensive study of Green's epistemology.)

Richter, M. (1964) *The Politics of Conscience: T.H. Green and his Times*, London: Weidenfeld & Nicolson.(A classic work of intellectual history, particularly good at tracing the origins of Green's views to his evangelical background.)

Thomas, G. (1987) *The Moral Philosophy of T.H. Green*, Oxford: Clarendon Press.(A recent analysis and defence of Green's ethics.)

Vincent, A. (ed.) (1986) *The Philosophy of T.H. Green*, Aldershot: Gower.(Contains articles covering all aspects of Green's activity and thought, together with a very useful bibliography.)

Gregory of Rimini (*c*.1300-58)

Gregory of Rimini was for a long time known primarily for his doctrine of predestination and for his notion of 'the complexly signifiable' in the semantics of propositions. However, he also provides an interesting alternative to William of Ockham among medieval nominalists. His chief work was his Lectura super primum et secundum Sententiarum (Lectures on Books I and II of Peter Lombard's Sentences).

Born in Rimini in Italy, Gregory began studies in theology at Paris in 1323 and attained the rank of lector there in 1329. He then taught at Augustinian houses of study in Bologna, Padua and Perugia, before returning to Paris in 1341 or 1342 to prepare for his lectures on the *Sentences* of Peter Lombard as a baccalaureus from 1342-3 or 1343-4. Gregory returned to his native Rimini as regent of the Augustinian *studium* (house of studies) in 1351 and taught there until 1357. He replaced Thomas of Strasbourg as General Prior of the Augustinians in 1357, but died a year later.

Gregory's chief work was his *Lectura super primum et secundum Sententiarum (Lectures on Books I and II of Peter Lombard's Sentences)*. Through it he helped introduce to Paris many new English philosophers, including William of Ockham, Walter Chatton, Adam Wodeham, Richard Fitzralph and, to a lesser degree, Thomas Bradwardine, Richard Kilvington, William Heytesbury, Thomas Buckingham and Robert of Halifax. He had encountered the works of some of these authors during his many years of teaching in Italy, and of others during his preparatory year in Paris. Gregory's own doctrinal positions were carried back to the Augustinian house of studies at Oxford in the 1350s by John Klenkok. His philosophical works, *Tractatus de intensione et remissione formarum corporalium (Treatise on the Intension and Remission of Corporeal Forms)* and *De quattuor virtutibus cardinalium (On the Four Cardinal Virtues)* are complemented by a number of scriptural commentaries, theological treatises and his administrative records.

Although Gregory was portrayed by Joannes Aventinus (John Thurmayr) in 1517 as 'the standard-bearer of the nominalists', some historians have considered him to be a leader of the anti-nominalists and promoter of the condemnations touching John Mirecourt in 1347. To evaluate the tag of nominalism is a complex affair, especially in the fourteenth century, when it swelled from a denial of real entities corresponding to our universal concepts to include a dozen other points. The most recent editors of his *Lectura* label him 'a nominalistic alternative to William of Ockham'.

Gregory is known for his staunch allegiance to Augustine: he criticizes Peter Aureol in particular for imprecise Augustinian quotations and argues against Thomas Aquinas and for the intellectual knowledge of singulars by citing chapter after chapter of Augustine's *De libero arbitrio (On Free Choice)*. Experience plays a very strong role in Gregory's philosophy as it did in Augustine's, since there are many things that we experience both within and outside ourselves that we cannot deny even though we also cannot explain them in terms of universal scientific knowledge. The intellect, according to Gregory, knows singulars directly before it knows universals: 'a universal is not some thing outside the mind but is rather a concept created (*fictus*) or formed by the soul that is common to many things' (*Lectura*, I, 396). This *fictum* theory of the concept was defended earlier by Henry of Harclay and William of Ockham, but it was challenged by Walter Chatton and seemingly abandoned by Ockham in his later works.

Gregory argues against Ockham that the object of science cannot be identified as the conclusion of a demonstration. Yet he does not agree with Walter Chatton's claim that things outside the mind are the objects of science:

> since if this were the case, many sciences would be about contingent things that could be different than they are, whereas for strict science the object must be eternal and necessary. Every being, however, besides God is contingent and not necessary. If things outside the mind were the objects of the sciences, then many sciences, physical and geometrical, and many others, would be about things other than God, and therefore about contingent things.
>
> (*Lectura* I, 6)

Following the lead of Adam Wodeham, Gregory found a middle path between the positions of Ockham and

Chatton and located the eternal and necessary knowledge of science in the total over-all meaning or significate (the *complexe significabile*) of the conclusion of a syllogism. Thus, the total significate of the proposition 'God is eternal' is neither the proposition 'God is eternal', nor God, nor eternity, but the *dictum* (which might be expressed as God's-being-eternal). The total significate of 'God is eternal' or the state of affairs expressible as God's-being-eternal is the object of our scientific knowledge (see Logic, medieval; Wodeham, A.).

Gregory accepts with little alteration many claims of Ockham's natural philosophy. Gregory, like Ockham, employs a razor to establish that motion, time and sudden change are not distinct and definable entities in themselves. 'Sudden change', for example, does not signify some thing over and above the permanent things involved in the change, that is, over and above the subject which is changed and the form gained which the subject did not have previously, or the form lost which it previously had. Gregory stresses the contingency of the natural world. Since God is the only necessary being, all creatures and thus the whole created universe are contingent. The laws of nature have been freely chosen by God and have no absolute necessity of their own (see Natural philosophy, medieval).

If Gregory welcomes a great deal of Ockham's logic, scientific theory and physics, especially as adjusted by the latter's pupil, Adam Wodeham, his anthropology is quite critical of these Englishmen and he accuses both of being modern Pelagians in their appreciation of man (see Pelagianism). He stresses the weakness of fallen human nature both in our ability to know what we should choose and what we should shun and in our ability to carry out our will in the proper direction even if we had the right knowledge.

In his scientific methodology, in his interest in questions of logic and language and in his natural philosophy, Gregory brings a new form of Augustinianism to Paris and to the Augustinian Order elsewhere. The form of Augustinianism initiated in the late thirteenth century by Giles of Rome was not long-lasting and was without great difficulty replaced by Gregory's stronger, English-influenced approach to philosophical and theological questions.

See also: Augustinianism; Aureol, P.; Chatton, W.; William of Ockham; Wodeham, A.

STEPHEN F. BROWN

List of works

Gregory of Rimini [Gregorius Ariminensis] *Lectura super Primum et Secundum Sententiarum (Lectures of Books I and II of Peter Lombard's Sentences)*, ed. A. Damasus Trapp *et al*., Berlin: de Gruyter, 1979-87.(These volumes replace the Venice edition of 1522, reprinted St Bonaventure, NY: The Franciscan Institute, 1952. The introduction to Volume I provides a description of the manuscripts for Book I, the principles of the edition and the distinction between the final text and the additions, which paradoxically are pre-Parisian elements that are subtracted in the Parisian *Lectura*. The introduction to Volume IV provides a description of the manuscripts for Book II, a comment on marginal notes and sources and doctrinal reflections called 'Themes and Problems'.)

References and further reading

Courtenay, W.J. (1978) *Adam Wodeham: An Introduction to his Life and Writings*, Leiden: Brill, 123-33. (Provides the *curriculum vitae* of Gregory.)

dal Pra, M. (1956) 'La Teoria del 'significato totale' della proposizione nel pensiero di Gregorio da Rimini' (The Theory of the 'Complete Significate' of the Proposition in the Thought of Gregory of Rimini), *Rivista critica di storia della filosofia* XI: 287-311.(On the significate of a proposition.)

de Muralt, A. (1991) *L'enjeu de la philosophie médiévale (What is at Stake in Medieval Philosophy)*, Studien und Texte zur Geistesgeschichte des Mittelalters XXIV, Leiden: Brill.(Gregory's theory concerning the object of knowledge.)

Eckermann, W. (1978) *Wort und Wirklichkeit: Das Sprachverständnis in der Theologie Gregors von Rimini und sein Weiterwirken in der Augustinerschule, Cassiacum (Word and Reality: The Understanding of Language in the Theology of Gregory of Rimini and its Effects in Augustinianism)*, Würzburg: Augustinus-Verlag.(On the theory of language and its relation to reality.)

Gál, G. (1977) 'Adam Wodeham's Question on the *complexe significabile* as the Immediate Object of Scientific Knowledge', *Franciscan Studies* 37: 66-102.(Shows the dependence of Gregory on Adam Wodeham regarding the significate of a proposition.)

Kölmel, W. (1955) 'Von Ockham zur Gabriel Biel. Zur Naturrechtslehre des 14. und 15. Jahrhunderts' (From

Ockham to Gabriel Biel: Natural Law Theory in the Fourteenth and Fifteenth Centuries), *Franziskanische Studien* 37: 767-87.(On the influence of Gregory on Gabriel Biel and Suárez concerning natural law.)

Kretzmann, N. (1970) 'Medieval Logicians on the Meaning of the *Propositio*', *The Journal of Philosophy* 67: 767-87.(Discusses the medieval debate on propositions.)

Kretzmann, N., Kenny, A. and Pinborg, J. (eds) (1982) *Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(Passages of interest include G. Nuchelmans on the significate of propositions, 203-6; C. Normore on future contingents, 376-80; and J.E. Murdoch on infinites, 572-3.)

Marmo, C. (1990) 'Gregory of Rimini: *notitia intuitiva*, *species* and Semiotics of Images', in S. Knuuttila, R. Työrinoja and S. Ebbesen (eds) *Knowledge and the Sciences in Medieval Philosophy (Proceedings of the Eighth International Congress of Medieval Philosophy, SIEPM*, Helsinki: Publications of Luther-Agricola Society B 19, II: 257-64.(Conference paper on Gregory's logic and language.)

Leff, G. (1961) *Gregory of Rimini: Tradition and Innovation in Fourteenth Century Thought*, Manchester: Manchester University Press.(Full survey of Gregory's life and works.)

Miccolo, L. (1987) 'L'oggetto della conoscenza scientifica nel Prologo del Commento alle Sentenze di Gregorio da Rimini' (The Object of Scientific Knowledge in the Prologue to Gregory of Rimini's Commentary on the *Sentences*), *Journal Philosophique* 13: 86-109.(Analysis of the opening question of Gregory's prologue to Book I, concerning the object of scientific knowledge.)

Oberman, H.A. (ed.) (1981) *Gregor von Rimini. Werk und Wirkung bis zur Reformation (Gregory of Rimini: His Work and its Influence on the Reformation)*, Spätmittelalter und Reformation Texte und Untersuchungen 20, Berlin: de Gruyter, 1-126, 127-94, 241-300.(Passages of interest include M. Schulze, an overview of Gregory's accomplishments, 1-126; V. Marcolino on the so-called 'additions' to Gregory's commentary, 127-94; and V. Wendland on the theory of scientific knowledge, 241-300.)

Paqué, R. (1970) *Das Pariser Nominalistenstatut: Zur Entstehung des Realitätsbegriffs der neuzeitlichen Naturwissenschaft (The Parisian Statute Regarding Nominalism: the Origin of the Concept of Reality in Modern Natural Science)*, Berlin: de Gruyter. (Discusses the 1347 Paris condemnations of John Mirecourt, in which Gregory reputedly played a major role.)

Tachau, K.H. (1988) *Vision and Certitude in the Age of Ockham*, Leiden: Brill. (Treats Gregory in passing, especially on pages 357-71 and 373-7.)

Thijssen, J. (1985) 'Het continuum-debat bij Gregorius van Rimini (1300-1358): Wijsbegeerte en theologia in de 14de eeuw' (Gregory of Rimini (1300-58) the Debate over the Continuum: Science and Theology in the Fourteenth Century), *Algemeen Nederlands Tijdschrift voor Wijsbegeerte* 77: 109-19.(A detailed study of Gregory's discussion concerning the nature of a continuum.)

Trapp, A.D. (1956) 'Augustinian Theology of the 14th Century', *Augustiniana* 6: 146-274.(Gregory is treated in great detail on pages 182-213.)

Worek, J. (1964) 'Agustinismo y aristotelismo tomista en la doctrina gnoseológica de Gregorio Ariminense' (Augustinianism and Thomistic Aristotelianism in the Epistemological Theory of Gregory of Rimini), *La Ciudad de Dios* 177: 435-68, 635-82.(A treatment of the more traditional elements in Gregory's philosophy and theology.)

Grice, Herbert Paul (1913-88)

Grice was a leading member of the post-war Oxford group of analytic philosophers. His small body of published work, together with an oral tradition, has been deeply influential among both philosophers and theoretical linguists. His outline of general rules of conversation began a new era in pragmatics. Grice's analysis of speaker's meaning explicates semantic notions in terms of the psychological concepts of intention and belief. His theory of conversation is based on the nature of language as a rational, cooperative activity. His account of conversational rules gave him a tool that he applied to a wide class of philosophical problems. Although Grice is most famous for his work on language and meaning, his interests cover a full range of philosophical topics, including ethics, moral psychology and philosophical psychology.

1 Life

Paul Grice was born on 15 March 1913 in Birmingham, England. A fellow and tutor at St John's College, Oxford, from 1938 to 1967, except when serving in naval intelligence during the Second World War, he was a member of a group of young Oxford philosophers who met under the leadership of J.L. Austin (see Austin, J.L. §2). He became Professor of Philosophy at the University of California at Berkeley in 1967, and died in Berkeley in 1988. Since he published reluctantly, much of his considerable influence came from the seminars and lectures he gave as a visitor to numerous universities and from his collaborative work. Grice was interested in a wide range of philosophical problems; more than most of his contemporaries of his rank he studied the great philosophers of the past, especially admiring Aristotle and Kant. He was exacting in presenting a problem, and lucid in exposing underlying issues and questions of methodology. Although his texts are often difficult to read, they are filled with wit, occasional irreverence and an exuberance which comes from his love of ideas.

Grice is most famous as a philosopher of language. His views on ethics, moral psychology and philosophical psychology, to some extent presented in public lectures, are mainly unpublished. The lectures, along with notebooks, are available in the Archives of the University of California.

2 Language and meaning

Meaning, Grice argued, does not result from conventions but is a function of what users do - with sounds and gestures as well as sentences and words. It is to be analysed in terms of intentions, on the part of an utterer, to produce a particular response in an audience. Utterer's meaning provides the basis for an analysis of utterance-type meaning, the meaning of words or sentences. The analysis explicates the stable truth-functional content of meaning - what, in Grice's terms, is 'said' by a speaker. But utterers may use language to mean more than what they say. Grice's theory of conversational implicature, propounded in 1967 in his William James lectures (1989), explains how this is possible (see Language, philosophy of).

3 Utterer's meaning

The meaning in general of a sign is to be explained in terms of what users of the sign mean by it on particular occasions. Grice's core idea is simple; the difficulties lie in the details. He begins by suggesting that '*x* meant something' would be true if *x* was intended by its utterer to induce a belief in some audience. To say what the belief was would be to say what *U* meant. But there is more to meaning that Jones is a murderer than acting so as to get someone to believe Jones to be a murderer: you could do that by leaving his handkerchief at the scene of the crime. Grice added nested intentions: you do something, meaning by it that *p*, if you do it intending to get someone to think that *p*, and intending to produce this result partly through the person's realizing that that is what you are up to ([1957] 1989). The idea seems obvious: meaning something by an utterance is distinguished from causing someone to believe something (or have some other propositional attitude) by the utterers' using their audience's *recognition* of their intention. The task of unpacking the idea is quite complex. The 'openness' of an utterer's intentions would later become an issue of considerable debate.

In his essay 'Utterer's Meaning and Intentions' ([1968] 1989), Grice proceeded to analyse utterer's meaning in two stages: (1) what it is for an individual to mean something; (2) what it is to mean that *p*. The interesting case, (2), is explained in terms of (1).

(1) ‘By uttering *x*, *U* meant something’ is true if and only if, for some audience *A*, *U* uttered *x* intending:
(i) *A* to produce some particular response *r*,
(ii) *A* to recognize that *U* intends (i), and
(iii) *A*s recognition that *U* intends (i) to function, in part, as a reason for (i).

To specify what *U* meant is to specify the nature of the intended response: where *x* is an indicative utterance, the response is *A*’s believing something, and where *x* is an imperative type the response is *A*’s intending to do something.

(2) ‘By uttering *x*, *U* meant that *p*’ is true if and only if for some audience *A*, *U* uttered *x* intending:
(i) *A* to believe that *p*,
(ii) and (iii) above.

This type of complex intention is called an ‘*M*-intention’. In a series of articles Grice revised this structure of nested intentions using more formal apparatus. He directed his attention to the meaning of imperative-type utterances, to a large range of speech-act types, and to meeting various counterexamples that he and others had proposed.

One series of issues concerns the ‘openness’ of meaning. A threatening infinity of intentions emerged in response to a case in which the three conditions for (2) - ‘By uttering *x*, *U* meant that *p*’ - are fulfilled. I do something (i) intending to get you to think that *p*, and (ii and iii) intending this to come about through your realizing that this is what I am up to, but *not* (iv) intending you to realize that (ii and iii) is what I am up to. For example, a man wants to ingratiate himself with his boss by letting him win at bridge. He does not wish to be blatant, but he does wish his boss, who likes such self-effacement, to know he wants him to win. To get this effect, he produces a smile when he has a good hand, intending this to be recognized as a simulation of a spontaneous response. Conditions i-iii of (2) are then fulfilled, but as Grice uses the idea of meaning, the employee did not mean by smiling that he had a good hand.

The problem, in the words of P.F. Strawson (1964), is that the employee’s intentions are not *wholly overt*; indeed, given that the employee does not wish to be crude, partial concealment is required. Utterer’s meaning requires openness; how is this to be conveyed in the analysans? Strawson proposed that Grice add a further condition, that the utterer *U* should utter *x* not only, as already provided, with the intention that *U* intends to obtain a certain response from *A*, but also with the intention that *A* should think that *U* has the intention just mentioned. This excludes the bridge-playing employee who wished his boss to take the smile as a spontaneous giveaway. With enough ingenuity, however, the same sort of counterexample can still be generated, and then a fifth clause will be needed, then a sixth, and so on.

Grice proposed to block an infinite regress by adding a condition which prevents ‘sneaky intentions’. Although there is continued debate about the extent of transparency required, as well as the form of its expression in the analysans, Grice himself felt that ‘the sneaky intention’ condition is *ad hoc*. Returning much later to the topic, he discussed what general conditions might legitimate our ‘deeming’ a speaker, given the absence of a sneaky intention, to have an infinity of intentions, and so to mean *p*. He worried about the legitimacy of ‘deeming’ someone to have an unrealizable intention.

The considerations which intrigued Grice, sketched in ‘Meaning Revisited’ ([1983] 1989) and in his ‘Reply to Richards’ (Grandy and Warner 1986), include his views on the nature of rational capacities, their connection to pre-rational states, and what he called value-oriented terms, that is, terms referring to a capacity or property such that attribution depends on proper or good specifications of that property. No completed account exists and much of the material is unfortunately unpublished. Some of his work on reasoning was however presented in the Locke/Kant lectures (1977, 1979).

4 Conversational implicature

Conversational implicature is not just a theory about language, but a tool to be applied to a wide class of philosophical problems. The theory simplifies semantics both by providing a single sense for the many uses of a word and by permitting a truth-functional analysis of ordinary language. Grice’s argument against the trenchant

criticisms of P.F. Strawson would return us, in some ways, to Bertrand Russell's conception of the relation between logic and language.

A person communicates more than what is said. Much of what is communicated is implied in one way or another. In the William James lectures Grice sketched the main types of implication, roughly characterized them, and proposed rules of conversation or communication which were specifications of a most general 'cooperative' principle. These rules govern a central class of implicatures, but they are not semantic in nature. They guide our conversations because they are rational, efficient ways to achieve common ends. Grice's work inaugurated a new era in what is now called pragmatics (see Pragmatics §§1, 3).

Grice's most general rule, the Cooperative Principle, requires that an individual's conversational contribution be such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange. More specific maxims (and their submaxims) articulate this rule: they enjoin truthfulness, informativeness, relevance and clarity.

Speakers conversationally imply that which they must be assumed to think if we are to suppose them to observe the conversational maxims, or at least the Cooperative Principle. Conversational implicatures are brought within Grice's general account of meaning in terms of speaker's intentions by the further condition that it is known (and known to be known) by both audience and speaker that the audience is capable of working out the assumption required.

Grice illustrates how implicature works. One example, in which no maxim is violated, is of such a recognizably common nature that we can also see the pervasiveness of conversational implicatures. A man stands by an obviously immobilized car and says to an approaching stranger 'I am out of petrol'. The stranger replies, 'There is a garage around the corner'. He implies that he thinks the garage is open, has petrol to sell and so on, for otherwise the relevance maxim would be infringed.

Grice's most arresting examples concern speakers who deliberately flout a subordinate maxim while adhering to the overarching Cooperative Principle. These are used to show how tautologies and figures of speech can be meaningful. In general, speakers may flout a subordinate maxim in order to introduce into the exchange something they do not want to say explicitly. Here is one of Grice's examples: a tutor, writing a testimonial for a pupil who is a candidate for a philosophy job, says only that the candidate's command of English is excellent, and his attendance at tutorials regular. Given his knowledge of his pupil, the tutor must be wishing to impart information that he is reluctant to write down, namely that his pupil is no good at philosophy.

The William James lectures ([1967] 1989) display the sensitivity to nuances of ordinary language that made Grice such a notable practitioner of 'Ordinary Language' philosophy. He did think that what he called linguistic botanizing, as practised in that tradition, was essential. He also regarded himself as part of that philosophical tradition that wished to defend common sense. None the less, he criticized both the practices of common-sense philosophers and the lack of theoretical justification in much ordinary language philosophy. In his analysis of perception and when treating the logical stucture of ordinary English, Grice protested against those who would block a philosophical analysis whenever its key terms were used in ways that were out of line with common speech. Oddness was to be distinguished from the truth or falsity of an utterance. Grice used his general theory of conversational implicature to explain the oddness, and in this way defended the unfashionable causal theory of perception. Similarly, in discussing Strawson's thesis that certain expressions, such as 'not', 'and', 'or', 'if', 'all', 'some' and, famously, the definite description operator 'the', diverge in meaning from their counterparts in formal logic, Grice held that both the advocates of truth-functional analyses and their opponents fail to attend to the conditions governing conversation (see Ordinary language philosophy, school of §3).

5 Value

Grice's views on value are subtle. Value is constructed by us, but objective. His lectures, 'The Conception of Value' ([1983] 1991) attempt to provide a metaphysical basis for value. Grice depicts an Aristotelian world in which objects and creatures are characterized in terms of what they are supposed to do, thus enabling us to evaluate by reference to function and finality. But he argues that the legitimacy of these evaluative judgments rests on an argument for absolute value. He makes the same claim for explanations in terms of purposiveness. Vitalistic phenomena can also be explained by mechanistic, cybernetic notions; these threaten the irreducibility of finality

features unless there is a domain of objects in which attribution of finality has no substitute. Rational beings, or persons, are said to furnish such a domain.

Grice's attempt to secure our evaluations takes as its first task understanding the concept of value, whose hybrid nature - describing and motivating - needs to be made intelligible. He begins by utilizing a device called 'Humean Projection'. Hume attributed to the mind a tendency 'to spread itself upon objects'; values start off as human attitudes, features of our minds or psychological states, but are projected onto the world. Motivation comes via our own attitudes, description from their projection. What served Hume as a diagnosis of one source of illusion becomes, under conditions Grice carefully specified, a legitimating procedure.

Humean projection was first used in an address, 'Method in Philosophical Psychology (from the Banal to the Bizarre)' ([1975] 1991), to justify the attribution to creatures of concepts which were unavailable to their predecessors. It offers a partial explanation of what it is to be a more developed psychological creature, by describing greater internalization or representational capacity in a creature. Internalized expressions include connectives, quantifiers, temporal modifiers, mood indicators, modal operators and names of psychological states, such as 'judge' and 'will'. The specific benefit of the procedure when applied to value is that judgments of value are not translatable back into preferences or imperatives or any other approving or volitional attitudes from which they are being constructed. Non-translatability does not, however, guarantee objectivity for value judgments. Grice thought that objectivity required the provision of 'something like' truth conditions for their expression. The lectures on value present a story, which Grice intended as the analogue of truth-conditions.

Objective value is attributed to the end-products of Humean projection if those who do the projecting are themselves of absolute value and can thus 'transmit' it - an adaptation of Aristotle's 'what seems good to the good man *is* good'. Two moves are thought to legitimate this attribution of value. First, Grice argues for the propriety of conceiving of human beings as persons. He introduces an operation flauntingly named Metaphysical Transubstantiation, a procedure for the redistribution but not the invention of properties. What was accidental becomes essential, and rationality, which attaches only accidentally to humans, becomes, via this construction routine, an essential property of persons. As persons, Grice claims, we will have a 'rational demand for absolute value'. Grice's second move is to invoke a 'metaphysical principle of supply and demand' which allows us to believe what can in a 'trouble-free' manner satisfy a rational demand. Despite his persuasive defence of these ideas, Grice recognized that, rather than offering a systematic account, he was sketching a story of value 'bristling with problems'.

See also: Analytical philosophy §3; Communication and intention; Value, ontological status of

JUDITH BAKER

List of works

Grice, H.P. (1941) 'Personal Identity', *Mind* 50: 330-50.(Grice states and defends a form of Logical Construction theory, offering a modification of Locke's memory-based theory of personal identity.)

Grice, H.P. and Strawson, P.F. (1956) 'In Defence of a Dogma', *Philosophical Review* 65: 141-58.(Grice defends the analytic-synthetic distinction, which had been criticized by W.V. Quine in 'Two Dogmas of Empiricism'.)

Grice, H.P. (1969) 'Vacuous Names', in D. Davidson and J. Hintikka (eds) *Words and Objections*, Dordrecht: Reidel, 118-45.(Grice presents an account of referential usage and defends Russell's Theory of Descriptions; he argues against the charge that definite descriptions are ambiguous between Russellian and referential readings.)

Grice, H.P.(1970) 'Lectures on Language and Reality', delivered at University of Illinois, Urbana, Illinois, unpublished: Archives, University of California, Berkeley, CA.(Lecture IV, 'Presupposition and Conversational Implicature', published in *Studies in the Way of Words*, presents Grice's defence of Russell and his attack on the notion of presupposition.)

Grice, H.P. (1971) 'Intention and Uncertainty', *Proceedings of the British Academy* 57: 263- 79.(Grice rejects his earlier (unpublished) view, which understood intention as a form of belief. He defends the thesis that there is a special attitude of willing distinct from belief and desire. Intentional action results from this special form of willing, which can occur in the absence of intention, but is a component of intention, where intention is present.)

Grice, H.P. (1977) 'The Immanuel Kant Lectures', delivered at Stanford University, revised and presented as the

John Locke lectures at the University of Oxford as ‘Aspects of Reason’ (1979), unpublished: Archives, University of California, Berkeley, CA.(These lectures on reasons and reasoning also contain material which bears importantly on a problem raised in connection with meaning. Grice’s analysis of meaning features utterers and audience guided implicitly by procedures. The difficulty is in understanding of what it is to be *implicitly* guided by procedures.)

Grice, H.P. and Baker, J. (1985) ‘Davidson on Weakness of the Will’, in B. Vermazen and M. Hintikka (eds) *Essays on Davidson: Actions and Events*, Oxford: Oxford University Press, 27-50.(Grice analyses and criticizes the account of weakness of will offered by Donald Davidson.)

Grice, H.P. (1986) ‘Reply to Richards’, in R.E. Grandy and R. Warner (eds) *Philosophical Grounds of Rationality*, Oxford: Clarendon Press.(Grice offers a general reply to the authors of the essays in his honour and a statement of his philosophical views.)

Grice, H.P. (1986) ‘Actions and Events’, *Pacific Philosophical Quarterly* 67: 1-35.(Grice’s discussion focuses on, but is not confined to, Davidson’s views on this topic. Grice offers an alternative account of actions and events.)

Grice, H.P. (1988) ‘Aristotle on the Multiplicity of Being’, *Pacific Philosophical Quarterly* 69: 175-200.(Grice discusses a well-known cluster of Aristotelian theses about the connection of semantic multiplicity and the notion of being. He discusses the different ways in which semantic multiplicity becomes unified - recursive, focal and analogical.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, MA, and London: Harvard University Press. (Collected essays, some previously published in journals but not listed above, written 1946-88. Includes ‘Meaning’ (1957), ‘The Causal Theory of Perception’ (1967), the first complete publication of The William James Lectures (1967) and ‘Utterer’s Meaning and Intentions’ (1969).)

Grice, H.P. (1991) *The Conception of Value*, Oxford: Clarendon Press.(Includes ‘The Carus Lectures on the Conception of Value’ (1983), which attempts to provide a metaphysical basis for value, ‘Method in Philosophical Psychology (From the Banal to the Bizarre)’ (1975) and ‘Reply to Richards’ (Final Section) (1986).)

References and further reading

Baker, J. (1991) ‘Introduction’, in H.P. Grice, *The Conception of Value*, Oxford: Clarendon Press.(An expansion of the material of §5 of this entry.)

Bennett, J. (1976) *Linguistic Behaviour*, Cambridge: Cambridge University Press.(Bennett’s project is influenced by the Gricean programme. He develops a view of language in general as a matter of systematic communicative behaviour and offers a rationale for the primacy of utterer’s meaning.)

Grandy, R. and Warner, R. (eds) (1986) *Philosophical Grounds of Rationality*, Oxford: Clarendon Press.(This is a Festschrift honoring Paul Grice and includes an overview of his work by Grandy and Warner, as well as essays by P. Suppes, ‘The Primacy of Utterer’s Meaning’, and A. Kemmerling, ‘Utterer’s Meaning Revisited’, defending Grice’s account. Kemmerling’s article is technical. See ‘On Defining Relevance’ by D. Wilson and B. Sperber for discussion of the conversational maxims.)

Levinson, S. (1983) *Pragmatics*, Cambridge. MA: Harvard University Press.(A good introduction to pragmatics and to Grice’s contribution to the field.)

Loar, B. (1981) *Mind and Meaning*, Cambridge: Cambridge University Press.(Influenced by Grice’s account of meaning, the chief concerns of the book are in the foundations of the theory of meaning and are directed towards understanding the contents of belief and linguistic meaning. Somewhat technical.)

Neale, S. (1992) ‘Paul Grice and the Philosophy of Language’, *Linguistics and Philosophy* 15: 509-59.(A comprehensive, critical commentary on the development of Grice’s work on language and meaning as displayed in Grice 1989. Excellent bibliography.)

Schiffer, S. (1972) *Meaning*, Oxford: Clarendon Press.(Schiffer presents objections to Grice’s account of speaker meaning and discusses what must be added for a satisfactory analysis. Schiffer attempts to cut off the threat of an infinite regress of intentions by appealing to mutual knowledge. Somewhat technical.)

Strawson, P.F. (1964) ‘Intention and Convention in Speech Acts’, *Philosophical Review* 73: 439-60.(Strawson proposes a modification to Grice’s account of utterer’s meaning in order to resolve a problem he attributes to the openness of utterer’s intentions.)

Grosseteste, Robert (*c.*1170-1253)

Grosseteste's thought is representative of the conflicting currents in the intellectual climate of Europe in the late twelfth and early thirteenth centuries. On the one hand, his commitment to acquiring, understanding and making accessible to his Latin contemporaries the texts and ideas of newly discovered Arabic and Greek intellectual traditions places him in the vanguard of a sweeping movement transforming European thought during his lifetime. His work in science and natural philosophy, for example, is inspired by material newly translated from Arabic sources and by the new Aristotelian natural philosophy, especially the Physics, On the Heavens and Posterior Analytics (Aristotle's treatise on the nature of scientific knowledge). Similarly, in his work in metaphysics, ethics and theology Grosseteste turns to ancient sources previously unknown (or incompletely known) to Western thinkers, prominent among which are Aristotle's Ethics and the writings of Pseudo-Dionysius. His work as a translator of and commentator on Aristotle and Pseudo-Dionysius places Grosseteste among the pioneers in the assimilation of these important strands of the Greek intellectual heritage into the mainstream of European thought.

On the other hand, Grosseteste's views are in significant respects conservative. His greatest debt is to Augustine, and his most original ideas - such as his view that light is a fundamental constituent of all corporeal reality - are extensions of recognizably Augustinian themes. Moreover, although his work on Aristotle is groundbreaking, his approach is judicious and measured, lacking any hint of the crusader's zeal that marks the work of the later radical Aristotelians. In general his practice conforms to the traditional Neoplatonist line, viewing Aristotle as a guide to logic and natural philosophy while turning to Platonism - in Grosseteste's case, Augustinian and Pseudo-Dionysian Platonism - for the correct account of the loftier matters of metaphysics and theology.

1 Life and works

Grosseteste was born probably shortly before 1170 in Suffolk, England. There is little evidence of his activities until the mid-1220s, when he would have been already well over fifty years old. In 1225 he was lecturing in theology at Oxford, where he also served for a time as chancellor of the university. In 1230 he left his post in Oxford's secular schools to become the first lecturer at the newly established house of studies for the Oxford Franciscans. He remained there until 1235 when he was elected bishop of Lincoln, England's largest diocese, which then included Oxford. He served in that position until his death in 1253.

We know little about Grosseteste's formal training or career prior to his appearance at Oxford in 1225. He may have begun his studies at Oxford or Cambridge in the 1180s, receiving the degree of Master of Arts by the end of that decade. He seems to have had an administrative career in the household of the Bishop of Hereford until perhaps 1198. It may have been in Hereford, an early centre of interest in scientific speculation, that Grosseteste developed his deep interests in natural philosophy, astronomy and mathematics. His earliest works, dating from about 1200-20, deal with subject matter of this sort: they include *De cometis (On Comets)*, *De artibus liberalibus (On the Liberal Arts)*, *De generatione sonorum (On the Generation of Sounds)*, *De sphaera (On the Sphere)* and *De impressionibus aeris (On the Influences of Air)*. Shortly after 1220 Grosseteste seems to have begun serious reflection on Aristotle. His *Commentarius in Posteriorum analyticorum libros (Commentary on Aristotle's Posterior Analytics)*, perhaps the first complete medieval commentary on that text, probably dates to the early 1220s. The earliest of his notes on Aristotle's *Physics* (which have been gathered together as the *Commentarius in VIII libros Physicorum*) may date from the same period. It is unclear whether Grosseteste ever studied at Paris or whether he spent time in Paris during the suspension of studies at Oxford in 1209-14, as many English academics did.

During the 1220s at Oxford, the study and teaching of theology moved to the center of Grosseteste's attention. As part of his study of the Bible and the Church Fathers, Grosseteste began learning Greek, an undertaking that would yield remarkable fruit in his late-blooming career as a philosophical theologian. The writings in philosophical theology that belong to the period from 1225-30 include *De veritate (On Truth)*, *De veritate propositionis (On the Truth of the Proposition)*, *De libero arbitrio (On Free Choice)* in two recensions, *De scientia Dei (On God's Knowledge) De statu causarum (On the Finitude of Causal Series)*, *De intelligentiis (On Intelligences)* and *De unica forma (On the Single Form)*.

Grosseteste's growing theological occupations did not displace his work in natural philosophy, however, and in the period from 1225 to about 1240 he continued to write on natural philosophy, weaving together his scientific and mathematical interests with his theological work. He continued to lecture on Aristotle's *Physics* into the early 1230s, and composed his treatise on the six days of creation, *Hexaemeron*, in 1230-5 as well as several treatises on broadly scientific matters, including *De impressionibus elementorum (On the Influences of the Elements)*, *De motu supercaelestium (On the Motion of What is Above the Heavens)*, *De motu corporali et luce (On Corporeal Motion and Light)*, *De lineis (On Lines)*, *De natura locorum (On the Nature of Places)*, *De iride (On the Rainbow)*, *De colore (On Colour)*, *De calore solis (On the Heat of the Sun)*, *De operationibus solis (On the Activities of the Sun, an exposition of Ecclesiasticus 43: 1-5)*, *De finitate motus et temporis (On the Finitude of Motion and Time)* and *De differentiis localibus (On the Differentiae Associated with Place)*. Central to Grosseteste's scientific reflections in this period is his developing account of the nature of light and its fundamental role in both natural and divine causality. That account reaches its fullest development in *De luce (On Light)*.

The crowning scholarly achievements from the years in which Grosseteste was Bishop of Lincoln result from monumental projects involving translation of and commentary on Greek philosophical and theological texts. In the mid-1230s he revised a twelfth-century translation of John Damascene's *De fide orthodoxa (On the Orthodox Faith)* (see John of Damascus) and produced the first Latin translations of the remainder of Damascene's corpus, *De logica (On Logic)*, *De centum heresibus (On the Hundred Heresies)*, *Elementarium dogmatum (Introduction to Doctrine)* and *De hymno Trisagion (On the Hymn 'Holy, Holy, Holy')*. Shortly thereafter he began the largest of his scholarly endeavours, the production of translations of and commentaries on each of Pseudo-Dionysius's four major treatises: *De caelesti hierarchia (On the Celestial Hierarchy)*, *De ecclesiastica hierarchica (On the Ecclesiastical Hierarchy)*, *De divinis nominibus (On the Divine Names)* and *De mystica theologia (On Mystical Theology)* (see Pseudo-Dionysius). By 1243 this massive undertaking was complete, and Grosseteste turned next to Aristotle's *Ethics* (see Aristotle). By the late 1240s he had finished the first Latin translation of the complete text of the *Ethics*. To accompany his translation Grosseteste produced translations of the ancient Greek commentators on the *Ethics* and his own glosses on the text. Apparently in the last years of his life, at close to eighty years of age, Grosseteste began studying Hebrew.

2 Augustinian psychology and epistemology

Following Augustine, Grosseteste appeals to the nature and behaviour of light to explain the fundamental nature of reality and human cognition of it. God is the first and highest light, and all creatures depend on God in the way rays of light depend on the light source from which they radiate. The eternal ideas in God act as principles in creation (*creatrices*) insofar as they are the formal exemplar causes of created things. Creatures have their being, and hence are 'true' (in the sense of 'true friend') to the extent to which they conform or are 'adequated' to the eternal ideas.

Grosseteste holds that since the eternal ideas in God are the exemplars of created things, knowledge of creatures depends on illumination from God:

> Since the truth of each thing is its conformity to the idea of it in the eternal Word, it is clear that every created truth is seen only in the light of the highest truth.…All created truth, therefore, is clear insofar as the light of its eternal idea is present to the observer, as Augustine testifies. Nor can anything be seen to be true in its created truth alone, just as a body cannot be seen to be coloured solely in its colour apart from the illumination of an extrinsic light.
>
> (*De veritate*)

Grosseteste believes that our cognitive dependence on God's illumination is a function of our weakness. Just as the weak eyes of the body cannot look at the sun itself, despite their depending on sunlight for seeing coloured bodies, the weak eyes of the mind cannot look on the highest truth itself, though the mind depends for its vision of truth on the light streaming from it. By contrast, pure intellects - the divine intellect, the separate intelligences and purified human intellects (in heaven) - have knowledge by virtue of their direct awareness of the eternal ideas and of the highest light itself (*De veritate*; *Commentarius in Posteriorum analyticorum libros* I.7, I.14).

In Grosseteste's view, our cognitive weakness is explained in part by the soul's embodiment:

> If the highest part of the human soul - the so-called intellective part which is not the actuality of any body and

> needs no corporeal instrument for its proper activity - were not clouded and weighed down by the weight of the corrupt body, it would have complete knowledge without the aid of sense-perception through an irradiation received from a higher light, just as it will have when the soul has cast aside the body.… But because the purity of the eye of the soul is clouded and weighed down by the weight of the body, all the powers of the rational soul are oppressed from birth by the weight of the body so that they cannot act, and so are in a certain way drowsy.
>
> (*Commentarius in Posteriorum analyticorum libros* I.14)

Our embodiment also accounts for the indispensable role sense-perception plays in our acquisition of knowledge: ‘In the course of time, as sense-perception acts through its many encounters with sensible things, reason (which is intermingled with the senses and is, as it were, ferried to sensible things by the senses as in a boat) is awakened’ (*Commentarius in Posteriorum analyticorum libros* I.14). Once awakened, reason begins to function, drawing distinctions and abstracting until it arrives at cognition first of universals, then of necessary truths and finally of the sorts of demonstrations that provide the strictest kind of knowledge.

Human cognitive weakness results not only from the soul’s embodiment but also from the soul’s misdirected love:

> The cause of the soul’s sight’s being clouded through the weight of the corrupt body is that the soul’s affection and vision (*affectus et aspectus*) are not distinct, and it attains its vision only by that by which it attains its affection or its love. Therefore since the soul’s love and affection are directed toward the body and bodily enticements, it necessarily drags the soul’s vision behind it and directs it away from its light.
>
> (*Commentarius in Posteriorum analyticorum libros* I.14)

In our present state, turned away from the light, we must make our way toward truth starting from vestiges of light discovered in the external senses; but to the extent that our love is directed away from corruptible things, our cognitive gaze will be directed toward its own light until, purified of bodily distractions, it will look on the light itself.

3 Light as the first corporeal form

Some of Grosseteste’s most original ideas result from his extension of the Augustinian metaphor of illumination to issues in natural philosophy. He is captivated by a particular characteristic of light, namely, its essential ability to multiply and diffuse itself instantaneously (as he thought) in all directions. He saw in this feature the mechanism of an ambitious account of the generation of the physical universe (see Illumination).

The problem, as Grosseteste sees it, is to account for the three-dimensional extension of bodies and of the corporeal universe in general. This is because the ultimate principles of corporeal things are prime matter and bodily form, each of which is in itself simple and utterly dimensionless. Prime matter must be dimensionless because its possessing dimensions would entail its being informed in some way; and bodily form, considered just as form, must be wholly immaterial and cannot therefore be spatially extended. Hence the combination of simple, unformed matter with simple form cannot give rise to quantitative extension. Moreover, no finite multiplication of a simple form in matter can give rise to quantitative extension because no aggregation of a finite number of extensionless entities can constitute anything extended. Grosseteste’s solution to this problem is to identify light with the primary corporeal form:

> I judge that the first corporeal form, which some call corporeity, is light. For light, of itself, diffuses itself in every direction so that from a point of light, a sphere of light as great as you please is generated instantaneously.… But corporeity is that from which the extension of matter into three dimensions necessarily follows.… But it was impossible for form, which in itself is simple and dimensionless, to introduce dimensions into matter in every direction - since matter is likewise simple and dimensionless - except by multiplying and diffusing itself instantaneously in every direction and by extending matter, in its diffusion. For form itself could not leave matter behind, since it is not separable.
>
> (*De luce*)

In its instantaneous infinite multiplication light, as it were, stretches the matter it informs into a three-dimensional quantity. Grosseteste holds that this generation of extended matter occurs instantaneously at the beginning of time

when God creates the first corporeal form in prime matter, which God also creates simultaneously. Moreover, he holds that since the infinite multiplication of something simple yields something finite (because what is produced by the infinite multiplication of something infinitely exceeds that by the multiplication of which it was produced), light's instantaneous infinite multiplication of itself yields finite corporeal extension in every direction. The full extent of that finite extension defines a sphere coextensive with the whole of the corporeal universe (*machina mundi*).

Grosseteste's reflections on the generation of the corporeal universe lead him to two particularly interesting corollaries. First, he reasons that if in general the infinite multiplication of something simple is required to produce a thing of finite quantity, then it must be that there are infinities of different magnitudes, the difference in the magnitude of the infinities accounting for the quantitative difference in objects of different sizes:

> Now, it is possible that an infinite aggregation of numbers is related to another infinite aggregation in any numeric ratio (and even in any non-numeric ratio), and [so] there are infinities that have more [elements] than other infinities and infinities that have fewer than others. (The aggregation of all the numbers, both even and odd, is infinite, and it is thus greater than the aggregation of all the even numbers, which is nevertheless infinite.) …It is clear therefore that by its infinite multiplication, light extends matter into smaller finite dimensions and into larger finite dimensions in any ratio whatever.

(*De luce*)

Thus, the first corporeal form - light - explains not only how there can be a three-dimensional corporeal world but also how there can be bodies in it of different sizes.

Second, Grosseteste supposes that since light is the first corporeal form and as such is intrinsic and fundamental to all corporeal reality, understanding the behaviour of light is fundamental to understanding all natural phenomena. Mathematics, therefore, and in particular geometry and optics (the science of refraction), are indispensable tools for natural philosophy: 'The consideration of lines, angles, and figures is especially useful since it is impossible to have knowledge of natural philosophy without them' (*De lineis*). Grosseteste's lifelong commitment to these disciplines provides the foundation for an enduring tradition of mathematical and scientific speculation at Oxford in the later Middle Ages (see Natural philosophy, medieval; Oxford calculators).

See also: Aristotelianism, medieval; Augustine; Augustinianism; Bradwardine, T.; Platonism, medieval; Pseudo-Grosseteste; Translators

SCOTT MacDONALD

List of works

Grosseteste, R. (*c*.1170-1253) *Opera Roberti Grosseteste Lincolniensis*, ed. J. McEvoy, Corpus Christianorum Continuatio Medievalis, Turnhout: Brepols.(A number of volumes are projected, including volumes devoted to Grosseteste's work on the Pseudo-Dionysian corpus. To date, only the first volume has appeared, containing Grosseteste's Pauline commentaries and his *Tabula*.)

Grosseteste, R. (*c*.1200) *De cometis (On Comets)*, ed. S. Thomson, 'The Text of Grosseteste's *De cometis*', *Isis* 19, 1933: 19-25.

Grosseteste, R. (1200-10) *De artibus liberalibus (On the Liberal Arts)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 1-7.

Grosseteste, R. (1200-10) *De generatione sonorum (On the Generation of Sounds)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 7-10.

Grosseteste, R. (*c*.1215) *De sphaera (On the Sphere)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 10-32.

Grosseteste, R. (1215-20) *De impressionibus aeris (On Atmospheric Pressures)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 41-51.

Grosseteste, R. (1220-32) *Commentarius in VIII libros Physicorum Aristotelis (Commentary on Aristotle's Physics)*, ed. R. Dales, Boulder, CO: University of Colorado Press, 1963.

Grosseteste, R. (early 1220s) *Commentarius in Posteriorum analyticorum libros (Commentary on Aristotle's Posterior Analytics)*, ed. P. Rossi, Florence: Olschki, 1981.(Translation by S. Macdonald, *Commentary on the Posterior Analytics*, New Haven, CT: Yale University Press, forthcoming.)

Grosseteste, R. (1225-30) *De veritate (On Truth)*, ed. L. Baur in *Die Philosophischen Werke des Robert*

Grosseteste, Münster: Aschendorff, 1912, 130-43; trans. R. McKeon in *Selections from Medieval Philosophers*, New York: Charles Scribner's Sons, 1929, vol. 1, 263-81.

Grosseteste, R. (1225-30) *De veritate propositionis (On the Truth of a Proposition)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 143-5; trans. R. McKeon in *Selections from Medieval Philosophers*, New York: Charles Scribner's Sons, 1929, vol. 1, 282-4; trans. N. Lewis in 'Time and Modality in Robert Grosseteste', Ph.D. dissertation, University of Pittsburgh, 1988, 427-9. (Lewis' dissertation also contains a corrected version of Baur's edition on pages 380-4.)

Grosseteste, R. (1225-30) *De libero arbitrio (On Free Will)*, ed. N. Lewis, 'The First Recension of Robert Grosseteste's "De libero arbitrio"', *Mediaeval Studies* 53, 1991: 1-88; trans. N. Lewis in 'Time and Modality in Robert Grosseteste', Ph.D. dissertation, University of Pittsburgh, 1988, 391-424.

Grosseteste, R. (1225-30) *De scientia Dei (On the Knowledge of God)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 145-7; trans. R. McKeon in *Selections from Medieval Philosophers*, New York: Charles Scribner's Sons, 1929, vol. 1, 285-7; trans. N. Lewis in 'Time and Modality in Robert Grosseteste', Ph.D. dissertation, University of Pittsburgh, 1988, 425-6.(Lewis' dissertation also contains a corrected version of Baur's edition on pages 377-9.)

Grosseteste, R. (1225-30) *De statu causarum (On the State of Causes)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 120-6.

Grosseteste, R. (1225-30) *De intelligentiis (On Intelligence)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 112-20. (This treatise is the second part of a single letter composed by Grosseteste; *De unica forma* is the first part.)

Grosseteste, R. (1225-30) *De unica forma (On the Unity of Forms)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 106-12.(This treatise is the first part of a single letter composed by Grosseteste; *De intelligentiis* is the first part.)

Grosseteste, R. (1230-5) *Hexaemeron*, ed. R. Dales and S. Gieben, London: Oxford University Press, 1982; trans. C. Martin, *Robert Grosseteste: On the Six Days of Creation*, London: Oxford University Press, 1996.

Grosseteste, R. (1230-5) *De impressionibus elementorum (The Impressions of the Elements)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 87-90.

(1230-5) *De motu supercaelestium (On the Motion of What is Above the Heavens)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 92-101.

Grosseteste, R. (1230-5) *De motu corporali et luce (On Corporeal Motion and Light)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 90-2.

Grosseteste, R. (1230-5) *De lineis (On Lines)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 59-65.(This treatise and *De natura locorum* are two parts of a single treatise.)

Grosseteste, R. (1230-5) *De natura locorum (On the Nature of Places)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 65-72.(This treatise and *De lineis* are two parts of a single treatise.)

Grosseteste, R. (1230-5) *De iride (On the Rainbow)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 72-8.

Grosseteste, R. (1230-5) *De colore (On Colour)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 78-9.

Grosseteste, R. (1230-5) *De calore solis (On the Heat of the Sun)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 79-84; trans. A.C. Crombie in D.A. Callus (ed.) *Robert Grosseteste: Scholar and Bishop*, Oxford: Clarendon Press, 1955.

Grosseteste, R. (1230-5) *De operationibus solis (On the Activities of the Sun)*, ed. J. McEvoy, 'The Sun as "Res" and "Signum": Grosseteste's Commentary on Ecclesiasticus ch. 43, vv. 1-5', *Recherches de théologie ancienne et médiévale* 41, 1974: 38-91.

Grosseteste, R. (1230-5) *De finitate motus et temporis (On the Finitude of Motion and Time)*, ed. R. Dales, 'Robert Grosseteste's Treatise "De finitate motus et temporis"', *Traditio* 19, 1963: 245-66.

Grosseteste, R. (1230-5) *De differentiis localibus (On the Differentiae Associated with Place)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 84-7.

Grosseteste, R. (1235-40) *De luce (On Light)*, ed. L. Baur in *Die Philosophischen Werke des Robert Grosseteste*, Münster: Aschendorff, 1912, 51-9; trans. C. Riedl, *Robert Grosseteste on Light*, Milwaukee, WI: Marquette University Press, 1942.

Grosseteste, R. (1238-43) *Notulae in De caelesti hierarchia (Glosses on Pseudo-Dionysius' On the Celestial Hierarchy)*, chs 1-9 ed. and trans. J. McQuade in 'Robert Grosseteste's Commentary on the "Celestial Hierarchy" of Pseudo-Dionysius the Areopagite', Ph.D. dissertation, Queen's University of Belfast, 1961; chs

10-15 ed. and trans. J. McEvoy, ‘Robert Grosseteste on the *Celestial Hierarchy* of Pseudo-Dionysius’, MA dissertation, Queen’s University of Belfast, 1967. (These two editions are scheduled to appear in *Opera Roberti Grosseteste Lincolniensis.*)

Grosseteste, R. (1238-43) *Notulae in De ecclesiastica hierarchia (Glosses on Pseudo-Dionysius’ On the Ecclesiastical Hierarchy)*, ed. C. Taylor Hogan, ‘Robert Grosseteste, Pseudo-Dionysius, and Hierarchy: A Medieval Trinity’, Ph.D. dissertation, Cornell University, 1991.(Editions of Grosseteste’s work on the Pseudo-Dionysian corpus are scheduled to appear in *Opera Roberti Grosseteste Lincolniensis*.)

Grosseteste, R. (1238-43) *Notulae in De divinis nominibus (Glosses on Pseudo-Dionysius’ On the Divine Names)*, chs 1 and 4 ed. F. Ruello, ‘La *Divinum Nominim Reseratio* selon Robert Grosseteste et Albert le Grand’, *Archives d’histoire doctrinale et littéraire du moyen âge* 34, 1959: 99-197.(Editions of Grosseteste’s work on the Pseudo-Dionysian corpus are scheduled to appear in *Opera Roberti Grosseteste Lincolniensis*.)

Grosseteste, R. (1238-43) *Notulae in De mystica theologia (Glosses on Pseudo-Dionysius’ On Mystical Theology)*, ed. U. Gamba, *Il Commento di Roberto Grossatesta al ‘De Mystica Theologia’ del Pseudo-Dionigi Areopagita*, Milan, 1942.(Editions of Grosseteste’s work on the Pseudo-Dionysian corpus are scheduled to appear in *Opera Roberti Grosseteste Lincolniensis*.)

Grosseteste, R. (1240-7) *Notulae in Ethicam Nicomacheam (Glosses on Aristotle’s Nicomachean Ethics)*, ed. H. Merken, *The Greek Commentaries on the Nicomachean Ethics of Aristotle in the Latin Translation of Robert Grosseteste*, Leiden: Brill, 1973-, 5 vols.(Vol. I containing Grosseteste’s translations of the Greek commentators on Books I-IV of the *Ethics* appeared in 1973; vol. III containing the translations of the Greek commentators on Books VII-X appeared in 1991; vol. II containing the translation of the commentators on Books V-VI, vol. IV containing indices for vols 1-III, and vol. V containing an edition of Grosseteste’s glosses have not to date been published.)

References and further reading

Callus, D.A. (ed.) (1955) *Robert Grosseteste: Scholar and Bishop*, Oxford: Clarendon Press.(Six seminal essays on Grosseteste’s life and thought.)

Crombie, A.C. (1953) *Robert Grosseteste and the Origins of Experimental Science 1100-1700*, Oxford: Clarendon Press; 2nd edn, 1962.(An ambitious argument, in some respects now qualified by more recent scholarship, for Grosseteste’s fundamental significance for the development of the experimental method in Western science.)

Gieben, S. (1969) ‘Bibliographia Universa Roberti Grosseteste ab an. 1473 ad an. 1969’, *Collectanea Franciscana* 39: 362-418.(Bibliography supplementing and extending Thompson (1940).)

Gieben, S. (1995) ‘Robertus Grosseteste: Bibliographia 1970-91’, in J. McEvoy (ed.) *Robert Grosseteste: New Perspectives on His Thought and Scholarship*, Steenbrugge: Brepols.(The continuation of Gieben (1969).)

McEvoy, J. (1982) *The Philosophy of Robert Grosseteste*, Oxford: Clarendon Press.(A compendious study of Grosseteste’s philosophical ideas with valuable biographical and bibliographical resources. The appendix to this masterful work updates the bibliographical work of Thomson (1940).)

McEvoy, J. (ed.) (1995) *Robert Grosseteste: New Perspectives on His Thought and Scholarship*, Steenbrugge: Brepols.(A collection of nineteen essays on a wide range of issues, presenting a synoptic view of the current state of Grosseteste scholarship.)

Southern, R.W. (1986) *Robert Grosseteste: The Growth of an English Mind in Medieval Europe*, Oxford: Clarendon Press; 2nd edn, 1992.(A provocative reassessment of Grosseteste’s intellectual development and historical significance. The biographical material presented in this article for the most part follows Southern’s account.)

Thomson, S.H. (1940) *The Writings of Robert Grosseteste, Bishop of Lincoln 1235-1253*, Cambridge: Cambridge University Press.(A pioneering catalogue and scholarly assessment of the manuscripts and published editions of Grosseteste’s works identified prior to 1940.)

Grote, John (1813-66)

From 1855 Grote was Knightbridge Professor at Cambridge. His literary legacy was largely posthumous. Often seen as unsystematic, he was in fact a penetrating thinker who forcefully criticized utilitarianism and positivism and ably argued that 'all that we call existence is for us a thought *of ours'; his was a seminal British idealism, stressing both the gulf between philosophical inquiry and the sciences and the difficulty of distinguishing the necessary from the contingent.*

Grote's views derived from Plato, Aristotle, Cicero, Christianity, Kant and the Cambridge Moralists. *An Examination of the Utilitarian Philosophy* exposes the gap between J.S. Mill's abstract utilitarian principle and concrete guidance, and shows how Mill failed to justify utilitarian benevolence over egoism or the demands of special relations - 'family, order, class, friends, country' (see Mill, J.S.). Though the ideal of happiness needs an intuitive ground, utilitarianism is partly right. What it misses is the more basic viewpoint of human agency: what we ought to do. Virtue and duty, requiring benevolent and distributivist action, are distinct demands, better addressing the basic ethical question: 'How different is moralized human nature to be from human nature unmoralized and as it is a subject of simple observation?' *A Treatise on the Moral Ideals* develops this system of 'Eudaemonics and Aretaics'; the former engages our sentient nature, that pain is undesirable to suffer, the latter our active nature, that pain ought not to be inflicted, though their common source is the claim that existence is driven by 'want' - a want to be, know, do and realize ideals. Positivism and utilitarianism miss the ideality or 'aspiringness' of thought, or how imagination constitutes our world. Ultimately, Eudaemonics is subordinate to Aretaics, which includes Deontics, and is differentiable into a plurality of other ideals with independent value - for example, justice, generosity, perfection, courage. The theistic faith that all these ideals cohere in an orderly universe is one 'we instinctively and intuitively have'.

Exploratio Philosophica develops Grote's original epistemology, critically engaging J.F. Ferrier but intricately demolishing a host of others for not keeping philosophical inquiry into truth separate from scientific inquiry into causes. For Grote, knowledge involves both propositions and judgments about what is 'given' immediately in experience, and progresses by 'distinction, not aggregation', discerning parts of a 'previously conceived whole'. He introduces the distinction between knowledge by acquaintance and by description, but for him, immediacy and reflection are continuous, complementary aspects of experience, whereby mind, via philosophy, traces itself in the order and purpose there. It is a basic truth that 'we know' and are free; we 'cannot superadd the idea of knowledge to that of existence and *then* analyse the knowledge' without disintegrating the idea of existence presupposed, just as the 'idea of freedom or choice of action will not superadd itself to that of a state of things independent of it, without pulling this latter to pieces'. Moreover, as 'one word of a language implies the whole system of it, so does one particular of knowledge involve thought of the, as yet unparticularized, universe'. 'On Glossology' (1872, 1874) also seeks the origins of knowledge in language, as speech, looking not to etymology but to what we 'wish to express, by the words' and 'what the person whom we address will understand by them'.

J.B. Mayor rightly noted not only Grote's 'freshness of thought', but also his 'fearlessness in the use of neologisms' - notably 'personalism', 'hedonics', and 'relativism'. Yet, however knotty his style, Grote foreshadowed much that was to come in idealistic, analytic and linguistic philosophy.

See also: Common Sense School; Epistemic relativism; Good, theories of the; Hedonism; Intuitionism in ethics; Moral knowledge

JOHN GIBBINS
BART SCHULTZ

List of works

The following list is limited to Grote's philosophical works, and does not include various other works. The best archival source is the Grote/Mayor Papers in the Wren Library, Trinity College, Cambridge.

Grote, J. (1865) *Exploratio Philosophica, Part I*, Cambridge: Deighton, Bell. (Published under his supervision, this work gives the substance of Grote's own idealist epistemology and ontology, which are elaborated through a conversational engagement with the work of such contemporaries as Ferrier, Whewell, Hamilton and J.S. Mill.)

Grote, J. (1867) ‘What is Materialism?’, *Macmillan’s Magazine* 15: 371-81.(Forming chapters 1-3 of *Exploratio Philosophica*, *Part II*, this work criticizes the development of psychology by a comparative analysis of the logical coherence of materialist physiology, associational psychology, philosophical epistemology and idealism.)

Grote, J. (1870) *An Examination of the Utilitarian Philosophy*, ed. J.B. Mayor, Cambridge: Deighton, Bell.(A brilliant assessment of utilitarianism, arguing that J.S. Mill’s revision of the doctrine is at odds with the earlier tradition and incoherent in itself.)

Grote, J. (1871) ’Thought versus Learning’, *Good Words* 12: 818-23.(Argues that the best strategy in adult education is to encourage thinking via engagement in conversation, rather than to impose learning via the digestion of books.)

Grote, J. (1871) ’On a Future State’, *Contemporary Review* 18: 133-40.(Attacks the appeal to supposed knowledge of the afterlife and the threat of damnation as a means of inculcating religious observance.)

Grote, J. (1872) ‘On Glossology’, *Journal of Philology* 4: 55-66, 157-81; also in *Journal of Philology* 1874 (5): 153-82.(Grote’s linguistic prolegomena to philosophy and ethics, which in opposition to the philologists argues that the meaning of words is best established by the study of everyday usage.)

Grote, J. (1876) *A Treatise on the Moral Ideals*,ed. J.B. Mayor, Cambridge: Deighton, Bell.(Being the positive statement of Grote’s moral philosophy, this text unites what is valid in both consequentialism and deontology in the systems of Eudaemonics and Aretaics, and grounds both of these in the experience of ‘want’ while also elaborating the plurality of moral ideals.)

Grote, J. (1877) ’Pascal and Montaigne’, *Contemporary Review* 30: 285-96.(A careful comparison of Pascal and Montaigne on the human condition.)

Grote, J. and Sidgwick, H. (1889) ’A Discussion Between Professor Henry Sidgwick and the Late Professor John Grote, on the Utilitarian Basis of Plato’s *Republic*’, *The Classical Review* 3: 98-102.(A reconstruction of an earlier debate on ethics between Grote and his younger colleague, cast in the form of a dialogue involving Thrasymachus, Adeimantes, Glaucon, Socrates and George Grote.)

Grote, J. (1900) *Exploratio Philosophica, Part II*, ed. J.B. Mayor, Cambridge: Cambridge University Press. (Elaborates and further illustrates Grote’s epistemology, nicely articulating the idealist theory of the development of knowledge and the controversy between idealism and positivism.)

References and further reading

Bosanquet, B. (1902) ’Review: *Exploratio Philosophica*’, *Archiv für Systematische Philosophie* 8: 128-30.(A vital source on how a major Oxford idealist recognized Grote as an objective idealist.)

Cunningham, G.W. (1933) *The Idealistic Argument in Recent British and American Philosophy*, New York: The Century Co.(Good chapter on Grote’s *Exploratio*, stressing the Hegelian elements in his idealism.)

Gibbins, J. (1997) *John Grote, Cambridge University and the Development of Victorian Ideas*, Edinburgh: Edinburgh University Press.(This book locates Grote in the context of his day, analyses his original theories of history, politics and language, and provides photographs and an up-to-date bibliography of manuscripts and references to Grote. Argues that while developing an idealist position, Grote also set a new direction for Cambridge philosophy and brings out his influence on or affinities with such later figures asJ. Ward, C.D. Broad, M. Oakeshott,J.A. Smith and H.W.B. Joseph.)

Lasson, G. (1878) ‘Review: *Treatise on the Moral Ideals*’, *Zeitschrift für Philosophie* u.k.r. 84: 149-54.(As the editor ofHegel’s *Works*, Lasson provides a review that indicates the German and idealist response to Grote.)

Macdonald, L. (1966) *John Grote, A Critical Estimate of His Writings*, The Hague: Martinus Nijhoff.(Good bibliography, but seems to have been unable to track down various writings. Densely written, it interprets Grote as a personal idealist and is good on the originality of Grote’s distinction between knowledge by description and by acquaintance, but contains less on his moral philosophy and very little on his views on history, politics and language.)

Mayor, J.B. (1870, 1876, 1900) Editorial introductions to *An Examination*, *Moral Ideals*, and *Exploratio Philosophica, Part II*.(These introductions are essential for understanding Grote’s development and significance, and the last noted work, which contains the index for the whole, explains how H. Sidgwick andH. Joseph were instrumental in helping Mayor publish it at last.)

Quinton, A. (1973) *Utilitarian Ethics*, New York: St Martin’s Press.(One of the few relatively recent introductory surveys of utilitarianism to pay Grote his due, with a short section summarizing his objections to Mill.)

Schneewind, J.B. (1977) *Sidgwick’s Ethics and Victorian Moral Philosophy*, Oxford: Clarendon Press.(A splendid overview of philosophical developments in the half-century leading up toSidgwick’s *Methods of Ethics*. Relates Grote more to the earlier conflict between intuitionism and utilitarianism than to the growth of

idealism - a key source of controversy in Grote studies.)

Seth, J. (1912) *English Philosophy and Schools of Philosophy*, London: J.M. Dent.(Detailed examination, identifying Grote, along with T.H. Green, as a typically English philosopher in both his procedures and his concern with the experiential basis of knowledge.)

Sidgwick, H. (1877) 'Review: *A Treatise on the Moral Ideals*', *Mind* 2: 239-44. (Sidgwick's various reviews of Grote tend towards the ungenerous in accusing Grote of being unsystematic and, consequently, unphilosophical.)

Whitmore, C.E. (1927) 'The Significance of John Grote', *Philosophical Review* 36: 307-37.(A very helpful overview.)

Grotius, Hugo (1583-1645)

Scholar, lawyer and statesman, Grotius contributed to a number of different disciplines. His reputation as the founder both of a new international order and of a new moral science rests largely on his De iure belli ac pacis (The Law of War and Peace) (1625). Though the tendency today is to regard Grotius as one figure among others in the development of the concept of international law, he is increasingly regarded as one of the most original moral philosophers of the seventeenth century, in particular as having laid the foundations for the post-sceptical doctrine of natural law that flourished during the Enlightenment.

Of several striking epithets applied to Grotius, the earliest was that of Henry IV, who called him ‘the miracle of Holland’. When only 15 years old Grotius had been taken on a diplomatic mission from the United Provinces to France, where he astonished the King with his prodigious learning. As a classicist, poet, historian, theologian and jurist he continued to impress his contemporaries throughout his life. He also continued to participate in international diplomacy during a period of almost constant conflict. For some twenty years he served the United Provinces until, on the fall of his patron Jan van Oldenbarnevelt in 1619, he was imprisoned for life. After a dramatic escape two years later he once again left The Netherlands for France, this time permanently. From 1634 until his death he acted as ambassador to France for Queen Christina of Sweden.

His interest in international relations earned Grotius another remarkable epithet, ‘the father of international law’. As early as 1605 he wrote a treatise in defence of Dutch incursions in the East Indies, of which one chapter was published as *Mare liberum* (The Freedom of the Seas) in 1609. The treatise as a whole did not appear in print until 1864 as *De iure praedae* (The Law of Plunder), but in 1625 Grotius provided a fuller and somewhat revised statement of his ideas in the work on which his reputation chiefly rests, *De iure belli ac pacis*. In the first of its three books Grotius examined the notion of a just war, before moving on in the second to review the causes of just wars and in the third to consider how wars were justly waged.

The *De iure belli* did much to promote the idea of a society of states bound together by the common recognition of a set of rules governing relations between them. None the less, Grotius’ reputation as the founder of an international legal order has not survived intact, partly because he wrote on only some aspects of international law, partly because his views were largely derivative. By contrast, his reputation as the founder of a new moral science has been rehabilitated. Although he wrote in the apparently familiar language of the natural law tradition, Grotius is believed to have established a new basis for ethical inquiry by tracing everything back to the one tenet of natural law that even sceptics would not deny, the idea that men were entitled to preserve themselves. From this foundation he - and to a greater degree his successors - reconstructed an elaborate body of natural law doctrine on a mathematical model (see Natural law §§4-5).

See also: Law, philosophy of; Pufendorf, S.; Rights; Roman law §3; War and peace, philosophy of

J.D. FORD

List of works

Grotius, H. (1609) *Mare liberum (The Freedom of the Seas)*, trans. R. Magoffin, Oxford: Oxford University Press, 1916.(Translation printed along with a facsimile of the 1633 edition.)

Grotius, H. (1625) *De iure belli ac pacis (The Law of War and Peace)*, trans. F.W. Kelsey, Oxford: Oxford University Press, 1925.(Translation printed along with a facsimile of the 1646 edition.)

Grotius, H. (1864) *De iure praedae (The Law of Plunder)*, trans. G.L. Williams, Oxford: Oxford University Press, 1950.(Translation printed along with a facsimile of Grotius’ original manuscript, replete with interesting alterations.)

References and further reading

Bull, H., Kingsbury, B. and Roberts, A. (eds) (1990) *Hugo Grotius and International Relations*, Oxford: Clarendon Press.(Contains essays reappraising Grotius’ role in the history of international relations, with a useful bibliography.)

Grotiana (1928-47) and *Grotiana*, new series (1980-) Assen: Grotiana Foundation. (Devoted to Grotian studies, with bibliography updated in the latter.)

Knight, W.S.M. (1925) *The Life and Works of Hugo Grotius*, London: Sweet and Maxwell.(The most detailed biography.)

Onuma, Y. (ed.) (1993) *A Normative Approach to War*, Oxford: Clarendon Press. (Provides reliable accounts of Grotius' central concepts.)

Tuck, R. (1993) *Philosophy and Government, 1572-1651*, Cambridge: Cambridge University Press.(Assigns Grotius a pivotal position in the development of seventeenth-century moral and political thought.)

Guanzi

The Guanzi, or 'Book of Master Guan', is an eclectic work including textual materials dating from the fourth century BC to the first century AD, drawing on themes from Daoism, the science of government, yin-yang *and the Five Phases, Legalism, Confucianism, the art of war, economics and the Huang-Lao movement. Its pragmatic outlook is founded on a kind of objective realism that is the signature of the famous fourth-century BC Jixia academy in the principality of Qi. It was this academy that consolidated several different schools of thought.*

Historians have for the most part regarded the *Guanzi* as a treatise on political economy, and have not appreciated its philosophical side. This is because it has come down to us in an incomplete, corrupt and reworked form. The current version was reputedly compiled by Liu Xiang (76-6 BC) and includes various kinds of texts, from short maxims and popular sayings to philosophical discourses and historical moralizings. Some chapters apparently were originally separate works, while others look like commentaries. The two main strains in the work seem to be the thought of Guan Zhong (the prime minister and architect of the economic dominance of the principality of Qi in the seventh century BC) and that of his anonymous disciples of the Han imperial epoch (second century BC to the second century AD).

Some of the ideas found in the *Guanzi* are reminiscent of legalism, including government by impersonal and invariable law, the concepts of 'profit' and 'propitious dates', the supreme position of the prince, the notion that one must adapt to circumstances with 'non-action' and that one must follow the will of Heaven as an impersonal, predetermined natural order (see Legalist philosophy, Chinese). Then there are other, more Confucian concepts: a respect for rituals that ensure social order by assigning a place to each person reflective of the inevitable differences in social status (the knowledge of how to allocate the places equitably being the mark of the Sage); a respect for unchanging, cosmic moral rules, the 'just way' (*dao*) and the cultivation of a 'sense of shame', along with a concern for the welfare of the people as necessary to the maintenance of good order (see Confucian philosophy, Chinese). In the *Guanzi*, Legalism is tinted by Confucian ideas while at the same time, Confucianism is influenced by Legalism and marked by a centralizing and authoritarian tendency closer to Xunzi than to Confucius. The two tendencies merge, respecting both the authority of the prince and the desires of the people.

The role of the prince is primary since he determines each person's rank and function, establishes laws and moral rules and hands out rewards and punishments. Being the central authority, he must exercise power by himself, maintaining rituals and dispensing law and punishments while also enforcing respect for political institutions. The prince must be a moral example, inspiring his people. His 'love' for the people is not purely humanitarian; it derives from the fact that the people, in their complete obedience, serve him and ensure his continued power without having any of their own. This contrasts with Confucius' insistence on the people's right and duty to give counsel, and Mencius' conception of the people's right to revolt against tyranny (see Mencius). In the *Guanzi*, the people need not be considered when their wants are individual and short-term; only the prince can decide how to evaluate the collective desires of his subjects and determine how to satisfy their deeper, long-term wishes.

The economic ideas in the *Guanzi*, with its insistence on the need to pay attention to prices and to seek a mean between poverty and excessive wealth, favour productivity and suggest some real commercial influence (see Economics, philosophy of). One famous chapter on 'the light and the heavy' is the first exposition of the need to watch over the circulation and amount of goods. However, the work also favours agriculture as a basic activity, ranking it higher than the occupations of craftsmen and merchants. The land itself constitutes the basis of government and its tax system. Emphasis is placed on the fair distribution of land, with attention to territorial organization and precise administrative boundaries, as well as to the nature of the soil and propitious dates. In this context there are several calendars, and a chapter on maps.

Alongside the notion of a Confucian-style *dao* or moral path and some Daoist-like ideas which had been rejected by the Daoists themselves, we also find chapters redolent of psycho-physiological Daoism, with ideas of non-action and the interior void that constitutes purity and receptiveness (see Daoist philosophy). In this vein there is also the idea of a metaphysical *dao*, ineffable and impossible to place, omnipresent as an all-encompassing void and as such, the source of all life. Paradoxically, a Confucian-like discussion on 'correcting names' is linked with a treatise on the Daoistic silence that the Sage must observe, a silence which allows for the establishment of an

indispensable, exact terminology.

See also: Confucian philosophy, Chinese; Dao; Daoist philosophy; Huainanzi; Legalist philosophy, Chinese; Political philosophy, history of; Smith, A.

ISABELLE ROBINET

References and further reading

Guanzi (4th century BC-1st century AD), trans. W. Allyn Rickett, *Kuan-tzu, a Repository of Early Chinese Thought*, Hong Kong: University of Hong Kong Press, 1965; trans. W. Allyn Rickett, *Guanzi: Political, Economic, and Philosophical Essays from Early China*, Princeton, NJ: Princeton University Press, 1985.(Two translations of the *Guanzi*.)

Haloun, G. (1951) ‘Legalist Fragments: Part I; *Kuan-tsï* 55 and Related Texts’, *Asia Major*, new series 2 (1): 85-120.(A translation of Chapter XVIII, 55 of the *Guanzi* with a study of the transmission of the text and its various editions.)

Gurney, Edmund (1847-88)

Edmund Gurney was an English psychologist and musician. His major work, The Power of Sound, is a vast treatise on musical aesthetics, ranging from issues in the physiology of hearing to the question of the relation of music to morality, but is mostly devoted to central questions of form, expression and value in music. It is the most significant work of its kind in the latter half of the nineteenth century.

Commentators often couple Gurney with Hanslick as a supporter of musical formalism, but his views on the expressive dimension of music are neither as restrictive nor as doctrinaire as Hanslick's. Hanslick insisted on denying specific emotional content to music, allowing it only to convey dynamic features, which emotions, among other things, might exhibit. Gurney, on the other hand, grants that some music possesses fairly definite emotional expression, and discusses at length the grounds of such expression; he is primarily concerned to deny that musical impressiveness, or beauty, is either the same as or depends on musical expressiveness.

Gurney maintains that overall form in music is not of primary relevance to the appreciation of music. This is because the central feature of musical comprehension is the grasping of individual parts as they occur, and the grasping of connections to immediately neighbouring parts, whatever the overarching form of a piece might be. The value of a piece is directly a function of the pleasurableness of its individual parts and the cogency of sequence exhibited at the transitions between them, not a function of its global architecture.

1 Form, understanding and value in music

A cornerstone of Gurney's view of music is the contrast between the temporal art of music and the static visual art of two dimensions. He is struck by the difference between the apprehension of music and the apprehension of a linear arabesque or an architectural façade. While one can have a single sweeping perception of the whole of an arabesque or façade, a musical work allows one only a series of perceptions of its parts as it unfolds in time, never a single perception of the work in its entirety. The experience of music is fundamentally a matter of individual, overlapping impressions of short duration.

Despite the successiveness of our apprehension of them, however, moments of music often group themselves into extended units that we seem to grasp as a whole, though we cannot literally perceive them at one time. We find that some sequences of minute, heard events cohere together strongly when sequentially perceived, so that we hear them as single, unified movements. This is the phenomenon of melody, a string of successive auditory impressions somehow yielding a single impression comprising them all. When a sequence of notes constitutes a melody, the notes together display a certain minimum rightness. This rightness can be understood counterfactually, in terms of the effects, usually deleterious, of changing a single note. The special quality or character of a given melody is dependent on every note that comprises it; change any one and invariably the result is either no melody or another melody. The general outline or contour of a melody gives practically no sense of it at all; its essence lies in the specific notes that make it up, and not in anything more general that may be abstracted from them. No pitch or rhythm sequence, furthermore, possesses any inherent goodness in itself, and there are in fact no necessary and sufficient conditions discoverable for what constitutes a good melody.

Gurney later supplements this counterfactual criterion of melody with one that is more directly phenomenological. It is a mark of a melody that one is familiar with, that when one is strictly only hearing one part of it, it is *as if* one is also hearing the whole thing; one is in some way aware of the remainder while hearing any one part. 'When a melody is familiar to us we realize it by a gradual process of advance along it, while yet the whole process is in some real manner present to us at each of the successive instants at which only a minute part of it is actually engaging our ears' (1880: 165). It is this that Gurney famously labels 'Ideal Motion', noting that it is as much a characteristic of the process by which we perceive melody as of melody itself. The peculiarity of audible as opposed to visible form, of which melody is the paradigm, is that it is both formed, implying completeness, and at the same time moving or developing, implying incompleteness. Though Ideal Motion bears some analogy to physical motion, and often induces the latter in sympathetic listeners, it is a distinct phenomenon, neither reducible to nor proportional to the other. Gurney also notes a related feature of familiar melodies, namely that in listening to them one seems to construct them, with their characteristic tensions and overall flavour, in the very act of listening.

The emergence of melody is thus the chief way in which the moment-to-moment nature of musical apprehension is tempered and, to an extent, transcended. Gurney does, however, admit the existence of larger unities (musical paragraphs, for example) provided the components are bound together in a certain way. Although such paragraphs, unlike melodies, are not experienced as wholes, they can be internally unified in such a fashion that their successive parts do not strike a listener as unconnected to each other. What is required for this Gurney names ‘organicity’ or ‘cogency of sequence’ - the one and only adequate criterion of effective musical form, the *sine qua non* of well-formedness at any level. It is possessed by the succession of unprepossessing two- and three-note fragments that make up a melody, and by the succession of melodic and transitional passages that make up a unified musical paragraph. In the first case, one has parts, the individual tones, which are not themselves impressive, combined into a unit that is; in the second, the parts are already independently impressive, though they yield a whole that is impressive as well. In discussing what it is for a musical paragraph to constitute a form in the strict sense, Gurney offers a characterization of cogency of sequence in which its additional application to melody is plain: ‘any paragraph which is to be musically valuable must satisfy the test that each bit shall necessitate, as it were, and so enter into organic union with, the one next to it’ (1880: 204).

A sequence is cogent, then, when each part - whether motive, phrase, melody or paragraph - leads convincingly to the next, each consequent appearing, upon familiarity, to be the natural and inevitable continuation of its antecedent. Where there is cogency of sequence, there is resistance to any imagined substitution for any part in the musical chain, the test of cogency being precisely ‘the feeling of resentment…with which the ear, after sufficient acquaintance, would receive changes or omissions’ (1880: 212). Cogency of sequence admits of degrees, however. It reaches its apex within single melodies, tying together component motives or phrases, and can run fairly high in a paragraph composed of several melodic or quasi-melodic passages. But it is usually on the wane within extended movements made up of several sections, and generally approaches zero between the movements of large-scale works.

For Gurney, cogency of sequence - including unity of melody, which is a species of it - is a purely *intuitive* matter, judged by an autonomous musical faculty. It cannot be assessed or demonstrated by rational means; no formulable rules of repetition, contrast or balance can guarantee the presence of cogency or organicity. The musical faculty, in accord with the experiential criteria that have been mentioned, is the ultimate arbiter of musical form and effectiveness.

Gurney insists that the essential sequentiality of musical apprehension, and our inability to take in an extended portion of music as a whole at one stroke, entails the absolute primacy of the *parts* of a musical composition with regard to musical form, musical enjoyment and musical worth. The real form of a piece of music is wholly exhausted by the constitution of the smallest independent units out of formless elements, and the specific manner in which each independent unit leads on to the next. There is in no important sense an overall form to an extended piece of music; there is only form within and between parts that are successively apprehended. Musical enjoyment is likewise grounded entirely in the grasp of individual moments of various lengths and the transitions between them. ‘The whole is a combination of parts successively enjoyed, and can only be impressive so far as the parts are impressive’ (1880: 97). The parts, even taken out of context, are often of exceptional value; the whole comprising all the parts in order, on the other hand, carries no additional value of its own. As for relations which a given part may bear to other parts distant from it, these may affect one’s experience of the given part, but attention remains fixed on the part being heard, and not on any skein of relations encompassing the piece as a whole in which the part abstractly figures. Furthermore, though awareness of such relations may enhance one’s pleasure in what is then being heard, there is no distinct pleasure to be had in the relations themselves.

This view, which emphasizes the concrete detail of musical surface and its quality and connectedness from moment to moment, is clearly inimical to that to which music theorists often implicitly subscribe - namely, that elucidation and awareness of large-scale structural relations is of primary relevance to the understanding and evaluation of musical works, and that the whole is indeed as important as the parts, if not more so. Nevertheless, it seems that Gurney is fundamentally right in this respect. The simple insight at the root of his thought, and one that promoters of overall form tend to forget, is that music consists of a series of successive events which can never be apprehended simultaneously in a single perceptual act. The parts of a façade can be taken in more or less in one sweep, unlike the parts of a symphony. Of course, earlier portions of a musical work can be abstractly reviewed in memory as listening proceeds, or later portions entertained in anticipation, but that is not the same as the

perceptual experience of a whole. What is crucial, according to Gurney, is involvement in local movement from phrase to phrase, and not architectonic vistas beyond aural experience.

Gurney distinguishes finally between two modes of listening to music, the ‘definite’ and the ‘indefinite’. His conclusions concerning musical form, value and enjoyment all presuppose listening that occurs in the definite mode. Someone listening definitely attends to the specific features of the melodic and harmonic motion of the music as it passes, registers the individuality of what they hear, and typically acquires a recognitional capacity for those parts of form as a result. Definite listening is active, alert, participatory and particularizing. Someone listening indefinitely, on the other hand, is mainly aware of a succession of sounds, and reacts to and registers only gross changes in the overall sound-image.

2 Beauty and expression in music

Gurney is a musical formalist in the following respects. First, he posits an autonomous musical faculty, whose charge is the grasping and assessing of musical forms. This faculty derives from the apprehension of ‘impressive’ forms, that is, ones that can be synthesized as wholes in the short span, and which thus exhibit Ideal Motion, a distinct and *sui generis* pleasure. Second, he maintains that the impressiveness of musical forms - their purely musical beauty - is not different from but entirely independent of whatever ‘expressiveness’ they possess. But Gurney, unlike Hanslick, clearly accepts the phenomenon of musical expressiveness and devotes considerable attention to it (see Hanslick, E.). He begins, in anticipation of Goodman’s claim that expressive properties are only metaphorically possessed, by observing that a thing does not express its central or defining attributes, but rather its incidental (or ‘occasional’) ones, for only the latter display sufficient separability from the thing to allow them to be expressed. Gurney’s specific analysis of musical expressiveness is thus: ‘when a particular feeling in ourselves is identified with a particular character in a particular bit of music, then we say without hesitation that such a particular bit *expresses* the quality or feeling’ (1880: 313). This formulation can evidently be construed in both an arousalist and a cognitivist way, depending on how the phrase ‘a particular feeling in ourselves’ is understood; indeed Gurney seems to waver between the two in subsequent discussion. On the first construal, what is expressed by music is necessarily some feeling evoked in us as we listen. On the second construal, what is expressed is a feeling that we have experienced and are familiar with, and which we are disposed to relate to the character of the part being heard, but this is not necessarily something we are then experiencing; it is a feeling the music suggests to us, rather than awakens in us.

An additional complication in Gurney’s account is that although the ‘feeling in ourselves’ by which the expressiveness of the music is gauged is in most cases the same as the quality we ascribe to the music, it is in some cases rather the feeling that *contemplation* of the quality ascribed is apt to induce in us. Thus, while the special feelings corresponding to melancholic or joyful music are melancholy and joy, those corresponding to humorous and terrible music, that is, music expressive of humour and terror, are amusement and fear, not wittiness or menace. Unfortunately, Gurney gives no principled criterion for when the relationship between expressed feeling and feeling in listeners conforms to the direct, as opposed to the indirect, model.

Gurney holds that the grounds of musical expressiveness are to be found in the resemblances between music and human behaviour that is expressive of emotion. Musical motion, in its pace, rhythm and shape, echoes that exhibited in the gestures, movements and vocal inflections that outwardly manifest human feeling. Gurney holds that a certain asymmetry obtains when accounting for expressiveness, which partly parallels the absence of criteria for impressiveness. When a moment or section of music expresses an emotion, it is always possible to find structural features of the music that, by underpinning a resemblance to standard expressive behaviour, are largely responsible for the expressiveness, and thus explain it; but we cannot reverse this procedure and dependably predict, on the basis of its structural features, the emotion that a moment or section of music will express.

Gurney’s arguments concerning impressiveness and expressiveness have a variety of targets and are of varying soundness. Some are designed merely to underline the difference between impressiveness and expressiveness, some are aimed at denying any dependence of impressiveness on expressiveness, and some are concerned with the different bearing of the two on musical worth.

Gurney notes that in much impressive music ‘definable expression’ is absent and that much music with ‘definable expression’ fails to be impressive. Thus - on the assumption that definable expression and

expressiveness are the same - expressiveness is distinct from, and neither necessary nor sufficient for, impressiveness. And reinforcing for Gurney the nonidentity of expressiveness and impressiveness is the fact that assessments of expressiveness often vary widely for different listeners, while judgments of impressiveness roughly coincide. But Gurney goes on to assert that musical features that are intended to bring about the definite expression of some emotion will only do so if the music is impressive as well - suggesting that impressiveness, far from depending on expressiveness, may even be a necessary condition of it.

Gurney notes that pieces with equivalent expressiveness are often musically valuable in different ways, so that we would not be content to keep one and not the other. He concludes that the musical value of such pieces must be independent of the expressiveness they have in common, and observes that emotions are expressed by human behaviour even more directly and unequivocally than by music, yet the former is not seen as a source of aesthetic satisfaction; again, the aesthetic worth of music cannot rest merely on its capacity to express emotion.

Reviewing the above, it is evident that Gurney does not sufficiently consider the possibility that *part* of a piece's value might reside in its expressiveness, even if its expressiveness is shared with other pieces, and even if a concomitant impressiveness is required for that value to be realized; nor does he consider that value might reside in the particular *way* in which music, as opposed to ordinary behaviour, is expressive of emotion; nor that the *interplay* of expressiveness and impressiveness in a given piece might be a focus of interest.

It is sometimes ventured that music, through its almost infinite variety of forms, might have the ability to convey correspondingly fine nuances of feeling, and thus might outstrip what word or gesture can accomplish in expressive capacity. However, if one holds with Gurney that the only way in which music is expressive is in its reflection of the normal, behavioural expression of emotion, then the idea that music will be able to express finer shades of emotion than are already differentiated in the full range of human behaviour necessarily comes to grief. On this hypothesis, the emotional suggestions of musical form and content can be no more finely-grained than pre-existing correlations of behaviour with emotion, by reference to which such suggestions are accomplished at all.

Finally, Gurney examines with devastating results the thesis that expressive music expresses the emotions of its creator. First, there is the gross implausibility of the claim that an extended piece might track the regularly fluctuating feelings of the composer; second, music expressive of a feeling is clearly composable by one who is not experiencing that feeling and has perhaps never experienced it; and third, observation gives no support even to any rough rules connecting the expressive character of a piece and a composer's contemporaneous state. A more plausible thesis, according to Gurney, proposes only that the dominant character of a composer's music, as evidenced in their entire *oeuvre* or a significant portion thereof, reliably reflects the composer's characteristic psychological condition.

See also: Art, understanding of; Emotion in response to art; Artistic expression; Formalism in art; Music, aesthetics of §§6, 10

JERROLD LEVINSON

List of works

Gurney, E. (1880) *The Power of Sound*, New York: Basic Books, 1966.(His major work, an important contribution to aesthetics of music in the nineteenth century.)

Gurney, E. (1887) *Tertium Quid*, London.(A collection of essays.)

References and further reading

Budd, M. (1985) *Music and the Emotions*, London: Routledge, ch. 4.(Discusses Gurney's views on the roots of musical affect and the nature of musical expressiveness.)

Cone, E. (1966) 'The Power of *The Power of Sound*', intro. to *The Power of Sound*, New York: Basic Books, 1966.(Contains biographical and anecdotal material, and contrasts Gurney's view of musical form with those advanced by other theorists of music.)

Hospers, J. (1982) *Understanding the Arts*, Englewood Cliffs, NJ: Prentice Hall, ch. 4. (Discusses Gurney's arguments concerning musical expressiveness.)

Levinson, J. (1993) 'Edmund Gurney and the Appreciation of Music', *IYYUN, The Jerusalem Philosophical Quarterly* 42: 181-205.(An elaboration of the material of the first part of this article, on musical

understanding.)
Levinson, J. (1997) *Music in the Moment*, Ithaca, NY: Cornell University Press.(A monograph that defends an account of musical comprehension based on Gurney's ideas about music listening.)

Ha'am, Ahad (1856-1927)

Ahad Ha'am (Asher Hirsch Ginzberg) was one of the most remarkable Jewish thinkers and Zionist ideologists of his time. Born in the province of Kiev in the Ukraine, he moved in 1884 to Odessa, an important centre of Hebrew literary activity. In 1907 he moved on to London, and in 1922 settled in the young city of Tel Aviv. He attended the universities of Vienna, Berlin and Breslau but did not pursue any regular course of study and was primarily an autodidact. Never a systematic philosopher, Ginzberg, who wrote in Hebrew and adopted the pen name Ahad Ha'am, 'one of the people', became a first-rate and widely read essayist and polemicist. He engaged in controversies over the practical problems of the early Jewish settlements in Palestine, his opposition to Theodore Herzl's drive to create a Jewish state, and numerous problems of Hebrew culture, tradition and literature. No single principle or theme stands out as the guiding idea of his thought. Indeed, his ideas are sometimes inconsistent. But his writings preserve the flavour of his values and commitments. Although his outlook never became the main road of Zionist ideology, its impact on Zionist thought was powerful, especially after the establishment of the State of Israel.

1 Jewish nationhood

Ahad Ha'am's thinking was strongly influenced by nineteenth-century evolutionism and the philosophy of Herbert Spencer. Unlike most other Jewish thinkers of his time, he was more under the spell of English and French than German philosophers, but he applied his eclectic thinking most concertedly to the complex of difficulties known in his time as 'the Jewish problem'. Jewish nationality, like any other, he argued, is acquired naturally - unlike the moral idea of humanity, which is derived from abstract reasoning. It is grounded in a spontaneous sentiment of national belonging, akin to family ties and needs no theoretical justification. For the sense of national belonging is prior to consciousness itself; it springs from a kind of biological impulse and is felt as an emotion, rather than reasoned like a universal moral truth.

A people, like any organism, Ahad Ha'am reasoned, strives for self-preservation. This national 'will to live' is the resultant of every individual's will to live. But its demands take precedence over individual survival, since the nation will endure long after the individual's death. Indeed, the national 'urge to survive' is not merely a Darwinian biological concept but a cultural and social imperative. Following up on a Hegelian idea elaborated by the Jewish philosopher Nachman Krochmal in the 1840s, Ahad Ha'am maintained that every nation has its own distinctive spiritual character. This spiritual factor, and no mere political fact like statehood, was the true expression of peoplehood and the true basis of national survival. Diverging from his positivist and evolutionary bent, Ginzberg turned here towardss ideas influenced by German idealism and romanticism. Every national spirit, he observed, has its sources in the past. In Judaism the national idea is inspired by the moral teaching of the Hebrew prophets. Thus the core of the Jewish spirit is not ultimately a matter of religion but of ethics (see Cohen, H.). Unlike many previous thinkers, Ahad Ha'am found the core message of Jewish ethics in turn not at the individual but at the social level, in the prophet's uncompromising insistence on objective justice, not mere love or compassion. What matters is the act, not the agent or the intention. He drew this theme from his reading of the Jewish sources, but he did not restrict it to Judaism. He conceived of the quest for justice as the universal human quest and thus found a world historical mission in the particularities of Jewish history. But the enduring message and the enduring bond were not those of religion. Monotheism and messianism, to be sure, had played their parts historically in upholding the national will to existence. But with the rise of secularism the Jewish religion could no longer unify the Jewish people. The historico-religious bond had crumbled. Ahad Ha'am himself had become a secular Jew, and he looked forward to the emergence of a new Judaism that would take over the unifying role once played by religion. Modern Judaism will respect the Jewish religion as the most significant creation of the Jewish people; but it can no longer accept that religion's imperatives. For these are not divine commands but historical expressions of peoplehood. These ideas of Ahad Ha'am's were later taken up by the American Jewish philosopher Mordecai Kaplan.

2 The spiritual centre

Despite the Hegelian and romantic essentialism at the root of his ideas of the Jewish spirit, it was in a Spencerian vein that Ahad Ha'am addressed the Jewish urge for survival, arguing, in effect, that there is no Jewish national

problem, any more than there is a French or English problem. The peculiarity of the Jewish question derives only from the complexity of the Jewish situation. Assimilation, long identified as a critical issue for the Jews of the diaspora, is the most portentous of the problems faced in that situation. In his famous essay 'Hikkuj ve-Hitbollelut' (Imitation and Assimilation) (1893), Ahad Ha'am defines assimilation as the disappearance of the natural will for national survival. This contretemps is seen by him as a danger to Jews only as individuals. Surprisingly, it need not threaten Jewish peoplehood. For Judaism, he reasoned, could be saved by the establishment of a spiritual centre in Palestine. The message of Judaism emanating from this centre to the Jewish periphery - the message oriented by the moral exhortations of the Hebrew prophets - will strengthen national consciousness in the diaspora and endow spiritual life with a true national content. It will thereby constitute an effective barrier against assimilation, without need of a state, as called for by the political Zionists of Theodore Herzl's movement (see Zionism).

The question remained, of course, whether the Jews to be influenced by the centre would willingly follow those who hoped to influence them. Clearly Ahad Ha'am's schema of spokes and a wheel more adequately reflects present-day relations between Israel and world Jewry than did the maximalist ideology that often characterized political Zionism. Yet, contrary to his hopes, even the State of Israel and the activity of its emissaries in the diaspora, not to mention the many cultural and educational institutions it shelters, have only partially arrested the trend towards assimilation among Jews in their countries of residence. The Jews of those lands have not come to see Israel as their capital; and Israel itself, in an ironic commentary on Ahad Ha'am's vision, is fully subject to the same forces that drive much of the assimilation of diaspora Jewry.

Ahad Ha'am's cultural Zionism was his alternative to assimilation on the one hand and political Zionism on the other. Unlike Pinsker and Herzl, whose Zionism was triggered by the collision of Jews with their social and political environment, and especially by the growth of anti-Semitism, Ha'am believed the Jewish predicament to consist in growing to feel too much at home in that environment and becoming estranged from the Jewish national heritage. His sensitivity to the problem of assimilation indeed tended to blind him to the threat of tragedy that hung over the diaspora Jewry in the form of anti-Semitism.

See also: Nation and Nationalism; Zionism

ZE'EV LEVY

List of works

Ahad Ha'am (1946) *Ahad Ha'am: Essays, Letters, Memoirs*, ed. and trans. L. Simon, Oxford: East and West Library.(Broad translated collection of Ahad Ha'am's writings in various forms.)

Ahad Ha'am (1950) *Kol Kit've Ahad Ha'am (Complete Works of Ahad Ha'am)*, Tel Aviv: Dvir.(The essays deal with Judaism, Zionism and the issues of the Zionist movement and colonization of Palstine.)

Ahad Ha'am (1970) *Selected Essays by Ahad Ha'am*, ed. and trans. L. Simon, New York: Atheneum.(A presentation of Ahad Ha'am's life, public activity in the Zionist movement, his controversy with Herzlian Zionism, and his literary and publicist work.)

Ahad Ha'am (1893) 'Hikkuj ve-Hitbollelut' Imitation and Assimilation, trans. and ed. L. Simon, in *Ahad Ha'am: Essays, Letters, Memoirs*, Oxford: East and West Library, 1946, 71-5.(Points out the positive aspects of cultural assimilation and the negative aspects of abandoning the Jewish heritage for its sake.)

References and further reading

Fraenkel, J. (1963) *Dubnow, Herzl and Ahad Ha'am*, London: Ararat.(Three monographs on a great non-Zionist Jewish scholar and two Zionist thinkers with divergent views on the aim of Zionism.)

Simon, L. (1960) *Ahad Ha'am Asher Ginzberg: A Biography*, New York: Herzl Press.(Biography of Ahad Ha'am.)

Habermas, Jürgen (1929-)

Jürgen Habermas, German philosopher and social theorist, is perhaps best known for his wide-ranging defence of the modern public sphere and its related ideals of publicity and free public reason, but he has also made important contributions to theories of communication and informal argumentation, ethics, and the foundations and methodology of the social sciences. He studied in Göttingen, Zurich and Bonn, completing a dissertation on Schelling's philosophy in 1954. After working for a short time as Theodor Adorno's research assistant at the Institute for Social Research in Frankfurt he held professorships in Heidelberg and Frankfurt and, from 1971 to 1981, was co-director of the Max Planck Institute in Starnberg. With the publication of Knowledge and Human Interests (1968) he became widely recognized as the leading intellectual heir to the Frankfurt School of Critical Theory, a variant of Western Marxism that included such figures as Adorno, Max Horkheimer and Herbert Marcuse. His two-volume The Theory of Communicative Action (1981) is a major contribution to social theory, in which he locates the origins of the various political, economic and cultural crises confronting modern society in a one-sided process of rationalization steered more by the media of money and administrative power than by forms of collective decision-making based on consensually grounded norms and values.

1 Early work

In his first book, *The Structural Transformation of the Public Sphere* (1962), Habermas traces the emergence of a bourgeois public sphere which, at least for a time, offered the prospect of an arena that would mediate between state and society. Rooted in the social and economic conditions of liberal capitalism, the 'bourgeois public sphere' refers to those sociocultural institutions that arose in the eighteenth century in opposition to the absolutist powers of the state - private clubs and coffeehouses, learned societies and literary associations, publishing houses, journals, and newspapers. Taken as a whole these institutions constituted a 'public realm of reasoning private persons' that was partially secured through the enactment of various constitutional rights and liberties. In theory and (to a more limited extent) in practice, this public sphere was distinct both from the private sphere of the market and the family and from the political authority of the state. It designated a sphere that comes about whenever private persons reason collectively about their common interests, and its function was both to restrain and legitimate the political power exercised by the administrative state.

Habermas' thesis, however, is that this idea of the public sphere had a very limited lifespan. At the level of theory, the idea of a 'reasoning public' was treated with suspicion by conservatives and progressives alike and, in liberal theory, the defence of a public sphere was often limited to the formal guarantee of a narrow set of 'natural' civil and political rights. As an historical phenomenon, the bourgeois public sphere suffered an equally unfortunate fate. In connection with what he refers to as the 'refeudalization' of civil society during the latter part of the nineteenth century, Habermas traces the commercialization of civil society, the bureaucratization of political and non-political authority, and the growth of a manipulative or propagandistic mass-media. A 'repoliticized social sphere' erodes the real distinction between state and society that is a necessary social condition for the bourgeois public sphere, and a society oriented to consumption and a politics based on the competition and bargaining between interest groups emerges in the place of a public sphere formed by an enlightened citizenry. Although Habermas never abandons the normative claims expressed in the bourgeois ideal, the conclusion of *The Structural Transformation of the Public Sphere* is extremely cautious about the possibility for a renewed public sphere under the altered conditions of late capitalist society. Only thirty years later in *Between Facts and Norms* (1992) does Habermas return, on a more optimistic note, to the prospects for democracy in complex, pluralist societies.

In the 1960s and 1970s, Habermas pursued a number of related issues. He sharply criticized the 'scientization of politics' and increase in 'technocratic consciousness' he discerned in contemporary societies. For Habermas the growth of technocracy was not inevitable but the result of a failure to preserve (albeit by other means) the classical distinction between theory and practice, and between practical wisdom (*phronēsis*) and technical skill (*technē*). In a series of influential essays, many of which are collected in *Theory and Practice* (1963), he traced the loss of these distinctions in modern political theory (from Hobbes to Hegel) as well as in Marx whose own concept of praxis blurred a related and equally important distinction between labour (*Arbeit*) and modes of social interaction based on shared interpretations of the world. Thus, contra Marx, Habermas argues that the end of alienated labour does not alone ensure social emancipation.

During this same period, Habermas also pursued a more systematic critique of positivism, found in his contributions to the so-called ‘positivist dispute’ in German sociology and in his popular study, *Knowledge and Human Interests* (1968). The latter work traces the ‘dissolution of epistemology’ from its origin as a critical enterprise with Kant to its status as a relatively uncritical ‘theory of science’ in the first half of this century. Positivism (or scientism), for Habermas, is the insistence that only the sciences constitute genuine knowledge together with the belief that science does not need any further critical analysis or justification. It refers less to the practice of the sciences than to their ‘scientistic self-misunderstanding’. Habermas challenges this view (which was still popular at the time) and attempts to secure an independent basis for critique by arguing that all forms of knowledge are rooted in fundamental human interests. He identifies three ‘quasi-transcendental’ or ‘anthropologically deep-seated’ cognitive interests with reference to which distinct forms of knowledge can be delineated: the natural sciences correspond to a technical interest; the historical-hermeneutic sciences, to a practical interest; and the critical sciences (for example, Marxism and psychoanalysis when each is freed from its own ‘scientistic self-misunderstanding’), to an emancipatory interest. Thus, through a kind of continuation of epistemology by social theory, Habermas sought to complete a critique of positivism and provide a ‘prolegomenon’ for a critical social theory. Though he quickly became dissatisfied with the anthropological underpinnings of this initial attempt, there are already hints of the ‘linguistic turn’ in his later theory, evident, for example, in his remark that ‘the human interest in autonomy and responsibility is not mere fancy, for it can be apprehended a priori. What raises us out of nature is the only thing whose nature we can know: language’ (1968: 314) (see Positivism in the social sciences).

2 *The Theory of Communicative Action*

The Theory of Communicative Action (1981), Habermas’ magnum opus, is a major contribution to social theory which in its aim and structure resembles Max Weber’s *Economy and Society* and Talcott Parson’s *The Structure of Social Action*. Like those works, it presents metatheoretic reflections on the basic concepts of social theory along with observations on the methodology of the social sciences and quasi-empirical hypotheses about modernization as a process of societal rationalization. However, in contrast particularly to Weber, Habermas does not regard rationalization as a process that inevitably culminates in the loss of meaning and freedom in the world, but as an ambivalent process that also opens up a potential for societal learning and new levels of human emancipation.

Like Weber and Parsons, Habermas begins with a set of metatheoretic reflections that yields a typology of action. His basic distinction is between ‘consent-oriented’ (or communicative) and ‘success-oriented’ (or purposive-rational) actions; within the latter he distinguishes further between strategic and instrumental action. Instrumental actions are goal-oriented interventions in the physical world. They can be appraised from the standpoint of efficiency and described as the following of technical rules. Strategic action, by contrast, is action which aims at influencing others for the purpose of achieving some end. It too can be appraised in terms of its efficiency and described with the tools of game theory and theories of rational choice (see Rational choice theory). Many instrumental actions can also be strategic, and some forms of strategic action may be instrumental. Communicative action, however, constitutes an independent and distinct type of social action. The goal or ‘telos’ of communicative action is not expressed or realized in an attempt to influence others, but in the attempt to reach an agreement or mutual understanding (*Verständigung*) about something in the world. Thus, while all action is teleological or goal-oriented in a broad sense, in the case of communicative action any further ends the agent may have are subordinated to the goal of achieving a mutually shared definition of the agent’s lifeworldly situation through a cooperative process of interpretation. In acting communicatively, individuals more or less naïvely accept as valid the various claims raised with their utterance or action and mutually suppose that each is prepared to provide reasons for them should their validity be questioned. In a slightly more technical (and controversial) sense, and one tied more specifically to modern structures of rationality, Habermas also holds that individuals who act communicatively aim at reaching understanding about something in the world by relating their interpretations to three general types of validity claims that are constitutive for three basic types of Speech acts: a claim to truth raised in constative speech acts, a claim to normative rightness raised in regulative speech acts, and a claim to truthfulness raised in expressive speech acts (1981, vol. I: 319-).

At a methodological level, this analysis of social action underlines the need for an interpretive or *Verstehende* approach in the social sciences. The requirement of *Verstehen* arises because the objects that the social sciences study - actions and their products - are embedded in ‘complexes of meaning’ (*Sinnzusammenhänge*) that can be

understood by the social inquirer only as he or she relates them to his or her own pre-theoretical knowledge as a member of the lifeworld. This, in turn, gives rise to the 'disquieting thesis' that the interpretation of action cannot be separated from the interpreter's taking a position on the validity of the claims explicitly or implicitly connected with the action. The process of identifying the reasons for an action unavoidably draws one into the process of assessment in which the inquirer must adopt the perspective of an (at least virtual) participant. Understanding the reasons for action requires taking a position on the validity of those reasons according to our own lights, and that means (at least initially) setting aside an external or 'third person' perspective in favour of an internal or 'first person' perspective in which both agent and interpreter belong to the same 'universe of discourse'. It is in this way that Habermas is able to connect the notion of rationality generally with his more specific claims that reason is a capacity to test validity claims.

However, despite his emphasis on social (or communicative) rationality with its ties to criticizable validity claims, Habermas does not think that society can be viewed as a sort of large-scale debating club. On the one hand, agents' interpretations are generally taken for granted and form part of an implicit background of knowledge and practices that constitute what he calls (following Edmund Husserl and Alfred Schutz) the 'lifeworld'. On the other hand, social integration can also be achieved in large measure by relegating tasks of action coordination to social institutions that operate, so to speak, 'behind the backs' of social agents. In fact, Habermas' distinction between society as lifeworld and society as system reflects a trend toward differentiation that is yet another feature of modern societies.

Beyond these metatheoretic and methodological concerns, *The Theory of Communicative Action* presents an interpretation of modern society as the outcome of a process of rationalization and societal differentiation. Rationalization, as in Weber, refers most generally to the extension of the criteria of rational choice and purposive-rational action to ever more aspects of social life (for example, science, law, business, and even religion and the arts). Differentiation in modern societies is reflected, first, in a (functional) separation of the economic and political subsystems from society as a whole and, second, within the lifeworld, in the division among institutional complexes devoted, respectively, to the tasks of the transmission of knowledge and interpretive patterns (culture), social integration ('society' in the narrower sense of normative orders), and socialization (personality). Finally, within culture one can also trace a differentiation among the three values spheres of science and technology, law and morality, and art and aesthetic criticism as each becomes independent of the other and develops its own internal standards of critique and evaluation. This interpretation of modern society as a process of rationalization is, of course, not new. However, in contrast to the classical theorists who generally viewed it as a threat to social life - Weber's 'iron cage', Marx's thesis of reification, Durkheim's analysis of anomie - Habermas emphasizes the potential for emancipation also made available. Thus, social pathologies are not an inevitable consequence of rationalization *per se* but result rather from a one-sided process in which the market and administrative state invade the lifeworld, displacing modes of integration based on communicative reason with their own form of functional rationality. Habermas dramatically describes this as the 'colonization of the lifeworld'. The primary task of a critical theory is to draw attention to this process of colonization and indicate the ways in which various social movements are a response to it (see Social sciences, philosophy of).

3 Discourse ethics

Habermas' development of a discourse (or communicative) ethics is an important corollary of his theory of communicative action and represents his formulation of a post-conventional moral theory as an alternative to utilitarianism and Kantian theories. Its basic idea is expressed in a principle of universalizability, analogous to the categorical imperative of Kant, which is intended to function as a rule of argumentation for testing the legitimacy of contested norms. The principle reads:

> Every valid norm must satisfy the condition that all affected can accept the consequences and the side effects its general observance can be anticipated to have for the satisfaction of everyone's interests (and these consequences are preferred to those of known alternative possibilities).
>
> (Habermas 1983: 65)

Habermas' strong claim is that this principle can be derived from the general pragmatic presuppositions of communication and argumentation. His strategy runs roughly as follows: In making utterances, speakers at least

implicitly raise different types of validity-claims - for example, claims to truth, normative rightness, and sincerity or truthfulness. These validity-claims, in turn, point to the notion of an ideal speech situation freed from all external constraints and in which nothing but the force of the better argument prevails. The principle of universalizability represents an attempt to formulate this counterfactual ideal as a constitutive rule of argument for moral-practical discourses: norms or maxims of action are only morally legitimate if, when contested, they could be justified in a moral-practical discourse (see Communicative rationality).

Discourse ethics shares several features with other broadly Kantian moral theories. It is a formalistic ethic in the sense that it does not presuppose substantive moral content (beyond the idea of practical reason) but rather specifies a formal procedure which any norm must satisfy if it is to be morally acceptable. It is also a cognitivist ethic, not in the sense that it supposes an independent order of moral facts, but in its insistence that there exists a sufficient analogy between moral discourse and scientific discourse to make it possible to speak, for example, of progress in learning or of a comparable notion of 'good reason' or argument in both. Finally, discourse ethics is a deontological moral theory in two senses. It assumes the priority of the right over the good. The basic moral principle must be specified in a way that does not presuppose a specific conception of the good life since that would violate the liberal commitment to a plurality of conceptions of the good. In a further sense the distinction between deontological and teleological theories is closely related to Kant's distinction between categorical and hypothetical imperatives, a distinction strongly contested in contemporary analytic moral philosophy. Habermas sides with those who argue that morality consists of categorical imperatives (imperatives that do not require non-reason-based interests or desires), but agrees with critics that Kant's defence of such imperatives was not successful. In his own theory, Habermas accounts for the obligatory character of moral norms in terms of their relation to communicative action: valid norms are morally binding because of their intimate connection with processes of social interaction and communication out of which one cannot easily (or even rationally) choose to step (1983: 109).

At the same time, discourse ethics differs from Kant's moral theory in some important respects. It breaks with Kant's two-world metaphysics (phenomenal/noumenal) and rejects his exclusively monological interpretation of the categorical imperative in favour of an intersubjective or communicative version of the principle of universalizability. Of course, Kant's categorical imperative - especially in its 'Kingdom of Ends' formulation - already has an intersubjective dimension to it. However, it is only because of his two-world metaphysics that Kant is able to equate what one person can consistently and rationally will with what everyone could consistently and rationally agree to. Only because interests, desires and inclinations are set over against reason and purged from the Kingdom of Ends can Kant assume a harmony between the individual and the collective rational will. In communicative ethics, by contrast, such simulated thought experiments should ideally be replaced by practical discourses. Even though any actual discourse will always be limited by constraints of space and time and will thus fall short of their ideal, virtual dialogues carried out by a few on behalf of others are not an adequate substitute.

It is perhaps still too soon to tell whether discourse ethics represents a distinct alternative in contemporary moral theory. On the one hand, it must respond to some of the same challenges that confront other impartialist theories and it has to make good its claim that a notion of rational consensus or rational acceptability (distinct from the Hobbesian sort) provides the best general account of more ordinary moral intuitions. On the other hand, discourse ethics does seem to capture central insights that have made utilitarian and Kantian theories attractive: It regards respect for individual integrity (or 'autonomy') and concern for the welfare of others (or 'solidarity') as two aspects of the abstract notion of equal respect implicit in the idea of communicative action (see Kantian ethics; Utilitarianism).

4 Politics and law

In *Between Facts and Norms* (1992) Habermas returns to the political questions that guided his early studies and outlines in greater detail some of the implications of *The Theory of Communicative Action* for democratic theory and practice. A central claim of the work is that the common contrast between democracy (or popular sovereignty) and constitutionalism (with its emphasis on individual rights) is overdrawn since each is conceptually dependent on the other. Popular sovereignty and basic rights are co-original (*gleichursprünglich*) in that each derives from an ideal implicit in the notion of communicative reason: the right not to be bound to norms other than those to which one could give uncoerced rational consent. Moreover, by stressing the procedural dimensions of communicative

reason Habermas offers a model of 'deliberative politics' that steers between the alternatives of liberalism and republicanism (or communitarianism). In particular, with the republican model, it rejects the vision of the political process as primarily a process of competition and aggregation of private preferences. However, more in keeping with the liberal model, it regards the republican vision of a citizenry united and actively motivated by a shared conception of the good life as unrealistic in modern, pluralist societies. By contrast, as he puts it, 'the success of deliberative politics depends not on a collectively acting citizenry but on the institutionalization of the corresponding procedures and conditions of communication, as well as on the interplay of institutionalized deliberative processes with informally constituted public opinions' (1992: 298). Central to this model is not only the institutionalization of different forms and levels of discourse that, taken together, allow for (a reasonable expectation of) a rational public opinion and will-formation, but, as the quotation indicates, a 'two-track' process in which there is a division of labour between 'weak publics' - the informally organized public sphere ranging from private associations to the mass media located in 'civil society' - and 'strong publics' - parliamentary bodies and other formally organized institutions of the political system. In this division of labour, 'weak publics' assume a central responsibility for identifying, interpreting and addressing social problems. Decision-making responsibility, however, as well as the further 'filtering' of reasons via more formal parliamentary procedures, remains the task of a strong public (for example, the formally organized political system).

Within legal theory Habermas also resists the sharp contrast between positivism ('law as the will of the sovereign' (see Legal positivism)) and natural law theories (see Natural law). Legitimate law must in some meaningful sense be construed as the will of the people or demos, yet a democratic process itself requires legal regulation. 'The idea of the rule of law [or 'constitutional state' (*Rechtsstaat*)] sets in motion a spiralling self-application of law' (1992: 39). Habermas' response to this paradox is twofold. At an institutional level he pursues a strategy of differentiation: legal discourses are but one (albeit important) form of discourse that must simultaneously constrain yet be pervious to other forms of practical discourse. This leads to a reformulation of the classical separation of powers in discourse-theoretic terms. At a normative level, he advocates a 'public reasons' approach similar to the recent proposal by John Rawls: ordinary law-making is subject to a higher law of 'constitutional essentials' (Rawls 1993: 227), but these essentials are themselves ultimately justified in terms of their ability to realize effectively citizens' basic moral powers - Habermas speaks of their 'public' and 'private' autonomy - and give expression to an ideal of free public reason (see Liberalism; Republicanism).

5 Habermas as polemicist

Throughout his career Habermas has been a frequent participant in cultural and political debates as well as more narrowly academic discussions. In fact, his reputation, at least in Germany, may be due as much to these interventions in public life as to his academic writings. From the 1950s when he challenged Heidegger's willingness to publish his 1935 lectures, Introduction to Metaphysics, without any apology for its support of fascism, through his spirited debates with radical students in the 1960s, his response to neoconservatism and the 'historian's debate' in the 1980s (which concerned the public use of revisionist historical interpretations of the 'final solution' for politcal purposes), and his reflections on nationalism and European identity in the 1990s, Habermas has exemplified in practice the ideals of the public sphere he has defended in his academic writings.

One of the more widely discussed of these debates is his engagement with postmodernism. In *The Philosophical Discourse of Modernity* (1985) Habermas argues that much of postmodern thought (from its roots in Nietzsche and Heidegger through to its representatives in Foucault and Derrida) rests ironically on a failure to break sufficiently with the modern philosophy of consciousness (or *Bewußtseinsphilosophie*) against which so much of its criticism is directed. Within this tradition, the (knowing and acting) subject is regarded either as one object among others in the world (and thus, too, as something to be dominated) or as a 'transcendental' subject whose conscious activity, while always presupposed, can itself never be consciously apprehended. Most modern and contemporary philosophy - including postmodernism or deconstructionism - oscillates between these alternative positions, thereby rendering incoherent their own normative claims and critical positions. Habermas argues (as he did in *The Theory of Communicative Action*) that one can escape the bottlenecks of the philosophy of consciousness only by making a 'paradigm shift' to a model of communicative action in which subjectivity is first approached through the intersubjective relations individuals establish with one another in the medium of language. Autonomy, for example, is then less a property that the individual agent obscurely possesses than it is a set of competencies that actors mutually ascribe to one another in specific forms of social interaction. Habermas' aim, once more, is to

show that the positive achievements of modernity can be salvaged from their shaky metaphysical foundations without, however, requiring that one relinquish a rational (or reasonable) grounding for normative criticism. Modernity thus remains for Habermas an 'unfulfilled project', though one that should not be abandoned (see Postmodernism).

See also: Frankfurt School; Modernism; Adorno; Marcuse

KENNETH BAYNES

List of works

Habermas, J. (1962) *Strukturwandel der Öffentlichkeit*, Darmstadt: Hermann Luchterland Verlag; trans. T. Burger, *The Structural Transformation of the Public Sphere*, Cambridge, MA: MIT Press, 1989.(Traces the history and ideology of the 'bourgeois public sphere' from its emergence in the eighteenth century to its decline in the twentieth century.)

Habermas, J. (1963) *Theorie und Praxis*, Darmstadt: Hermann Luchterhand Verlag; trans. J. Viertel, *Theory and Practice*, Boston, MA: Beacon Press, 1973.(Contains important essays on modern political theory, including several onHegel.)

Habermas, J. (1968) *Erkenntnis und Interesse*, Frankfurt: Suhrkamp; trans. J. Shapiro, *Knowledge and Human Interests*, Boston, MA: Beacon Press, 1971; Oxford: Polity Press, 1987.(An important interpretation of philosophy from Kant to the pragmatists arguing that epistemology must now be pursued as social theory. Part 1 contrasts Hegel's critique of knowledge with that of Kant and Marx.)

Habermas, J. (1970) *Zur Logik der Sozialwissenschaften*, Frankfurt: Suhrkamp; trans. S. Nicholsen and J. Stark, *On the Logic of the Social Sciences*, Cambridge, MA: MIT Press, 1988.(An early sketch of Habermas' interpretive approach and his critique of Gadamer's philosophical hermeneutics.)

Habermas, J. (1971) *Theorie und Praxis*, Frankfurt: Suhrkamp, 4th edn; trans. J. Viertel, *Theory and Practice*, Oxford: Polity Press, 1973.(Contains several influential essays on Hegel.)

Habermas, J. (1973) *Legitimationsprobleme im Spätkapitalismus*, Frankfurt: Suhrkamp; trans. T. McCarthy, *Legitimation Crisis*, Boston, MA: Beacon Press, 1975.(An important and influential early formulation of Habermas' critical analysis of modern society.)

Habermas, J. (1976) *Zur Rekonstruktion des historischen Materialismus*, Frankfurt: Suhrkamp; trans. T. McCarthy, *Communication and the Evolution of Society*, Boston, MA: Beacon Press, 1979.(The abridged English translation contains essays on Habermas' theory of social evolution and an important essay on the linguistic grounding of normative critique, 'What is Universal Pragmatics?'.)

Habermas, J. (1981) *Theorie des kommunikativen Handelns*, Frankfurt: Suhrkamp, 2 vols; trans. T. McCarthy, *The Theory of Communicative Action*, Boston, MA: Beacon Press, 1984/1987.(His two-volume magnum opus, a systematic treatise on the foundations of social theory and one-sided process of rationalization in modern societies.)

Habermas, J. (1983) *Moralbewußtsein und kommunikatives Handeln*, Frankfurt: Suhrkamp; trans. C. Lenhardt and S. Nicholsen, *Moral Consciousness and Communicative Action*, Cambridge, MA: MIT Press, 1990.(A collection of five essays, including a central essay on 'discourse ethics'.)

Habermas, J. (1985) *Der philosophische Diskurs der Moderne*, Frankfurt: Suhrkamp; trans. F. Lawrence, *The Philosophical Discourse of Modernity*, Cambridge, MA: MIT Press, 1987.(Series of polemical lectures, ranging from Hegel to Derrida and Foucault, in which Habermas challenges the postmodern turn in philosophy.)

Habermas, J. (1985/1987) *Kleine Politische Schriften*, vols. V and VI, Frankfurt: Suhrkamp; trans. S. Nicholsen, *The New Conservatism: Cultural Criticism and the Historians' Debate*, Cambridge, MA: MIT Press, 1989.(A collection of essays on politics and culture, including the 'historians' debate' on the memory of the Holocaust, Heidegger's politics, and an essay on Henrich Heine and the role of the intellectual in Germany.)

Habermas, J. (1988) *Nachmetaphysisches Denken*, Frankfurt: Suhrkamp; trans. W. Hohengarten, *Postmetaphysical Thinking*, Cambridge, MA: MIT Press, 1992. (Collection of philosophical essays defending his 'postmetaphysical' position from both traditional philosophy of consciousness and its postmodern alternative.)

Habermas, J. (1992) *Faktizität und Geltung*, Frankfurt: Suhrkamp; trans. W. Rehg, *Between Facts and Norms*, Cambridge, MA: MIT Press, 1996.(A discourse-theoretic interpretation of law and democracy, and their interconnection.)

References and further reading

Benhabib, S. and Dallmayr, F. (eds) (1990) *The Communicative Ethics Controversy*, Cambridge, MA: MIT Press.(Contains translations from the German discussion and a lengthy ‘Afterword’ by Benhabib.)
Bernstein, R. (ed.) (1985) *Habermas and Modernity*, Cambridge, MA: MIT Press.(Contains essays on Habermas’ later work and his debate with postmodernism.)
Calhoun, C. (ed.) (1992) *Habermas and the Public Sphere*, Cambridge, MA: MIT Press.(An excellent collection devoted to *The Structural Transformation of the Public Sphere*.)
Dews, P. (ed.) (1986) *Habermas: Autonomy and Solidarity - Interviews with Jürgen Habermas*, London: Verso.(An important collection of interviews with Habermas, including the ‘Dialectics of Rationalization’ with Axel Honneth.)
Honneth, A. and Joas, H. (eds.) (1991) *Communicative Action*, Cambridge, MA: MIT Press.(A valuable collection of essays on *The Theory of Communicative Action*.)
Ingram, D. (1987) *Habermas and the Dialectic of Reason*, New Haven: Yale University Press.(Full-length study focused primarily on *The Theory of Communicative Action*.)
McCarthy, T. (1978) *The Critical Theory of Jürgen Habermas*, Cambridge, MA: MIT Press.(A thorough study of Habermas’ development prior to *The Theory of Communicative Action*.)
McCarthy, T. and Hoy, D. (1994) *Critical Theory*, Cambridge, MA: Blackwell.(An exchange between a leading exponent of Habermas’ theory and a ‘postmodern’ critic.)
Parsons, T. (1937) *The Structure of Social Action*, New York: McGraw-Hill.(Classic survey of foundations of empirical social science.)
Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(Includes Rawl’s articulation of the role of a just constitution within a pluralist society.)
Rehg, W. (1994) *Insight and Solidarity: The Discourse Ethics of Jürgen Habermas*, Berkeley, CA: University of California Press.(A full-length study of Habermas’ discourse ethics.)
Thompson, J. and Held, D. (1982) *Habermas: Critical Debates*, Cambridge, MA: MIT Press.(An early but important collection of critical essays exploring various aspects of Habermas’ work, including a ‘Reply’ by Habermas.)
Weber, M. (1925) *Wirtschaft und Gesellschaft*, Tübingen: J.C.B. Mohr; revised 4th edn, J. Winckelmann, 1956; trans. as *Economy and Science*, ed. G. Roth and C. Wittich, New York: Bedminster Press, 1968.(Classic theory of social and economic structures.)
White, S. (1995) *The Cambridge Companion to Habermas*, New York: Cambridge University Press.(A volume dealing primarily with Habermas’ political theory and postmodernism.)

Haeckel, Ernst Heinrich (1834-1919)

Haeckel was the leading German Darwinist. His evolutionary philosophy of monism differed substantially from the views of Darwin or of British evolutionary philosophers such as Herbert Spencer or the dualist T.H. Huxley. Haeckel's monism asserted the unity of physical and organic nature, and included mental processes and social phenomena. Its initial form was mechanistic, seeking to reduce vital processes to physicochemical laws and substances. However, his efforts to construct the history of life meant that Haeckel became preoccupied with historical processes. In its final form, his monism was pantheistic. AlthoughHaeckel has been regarded as a forerunner of national socialism, a contextual reading of his works does not support this interpretation.

Haeckel studied medicine and zoology. His approach to biology should be contrasted with the reductionist biophysics programme of the physiologists Hermann Helmholtz, Emil Du Bois-Reymond and Rudolf Virchow. During the 1850s Virchow (for whom Haeckel was assistant) moved away from reductionism to a belief in distinct organizational properties of life - notably in the cell. Haeckel was influenced byVirchow's theories of cellular organization and by the vitalist comparative anatomist and physiologist, Johannes Müller. More distant influences included Goethe's concept of morphology (see Goethe, J.W.) and the *Naturphilosophen* Lorenz Oken and Friedrich Schelling.

Darwin's *Origin of Species* inspiredHaeckel to advance Darwinism scientifically and as a popular form of materialist ethics (see Darwin, C.). This coincided with Max Schultze's protoplasmic theory of the cell, which gave Haeckel the idea that the cell substance or 'Plasma' was equivalent to the first-formed life. All organisms were plasmatic bodies differing only in degrees of organization. In linking cell theory to evolutionary theory, Haeckel was not a conventional Darwinist. Natural selection assumed only a limited role, being circumscribed by environmental and historical forces: Haeckel endorsed Lamarckian factors, such as the direct effects of the environment.

In reconstructing the history of life, Haeckel adopted a causal and mechanistic approach: in the 1870s he developed the 'biogenetic law' that ontogeny (the development of an individual organism) recapitulated phylogeny (the evolution of the species). The Gastraea Theory of 1874 deduced an extinct primitive organism (the Gastraea) that explained the early stage of embryological development, which Haeckel termed 'gastrulation'. Past evolutionary history provided a causal explanation of development.

Haeckel's monism was first developed in the *Generelle Morphologie (General Morphology)* of 1866. It implied that there was no essential difference between organic and inorganic nature, life differing from inorganic nature only in its higher organization. Atoms had immanent properties of irritability. Substance united matter and spirit. Haeckel liked to evoke the image of 'crystal minds' (*Kristalseelen*) to convey the link between matter and mind. There were two fundamental laws of substance: that of energy and material as constant throughout the world, and that of evolution from the unformed to the fully formed. Similarly there was no fundamental difference between animals and the human species. All mental powers were derived from sensitivity and movement. Ethics was based on the natural sciences, particularly the evidence of the social instincts of animals. The advance of science had the consequence of making mankind more aware of the rationale of the new morality. Certain aspects of modern civilization could be condemned as degenerative.

Haeckel drew on non-Darwinian concepts of division of labour in order to demonstrate that a centralized nervous system was an essential feature of the evolution of higher organisms. Following Spencer, he argued that the more an organism was differentiated the greater its need of centralized coordination and control. He appliedVirchow's egalitarian concept of a 'cell state' in a hierarchical fashion to argue that higher organisms manifested greater centralization, just as empires were higher than republics.

Haeckel condemned Christianity, and especially Roman Catholicism, as a primitive superstition, accusing Christianity of despising nature. Monism should replace religion as the basis of education in the newly united nation. At the same time he denied liberal notions of free will, which led to controversy with Virchow in 1877. Virchow viewed Darwinism only as an hypothesis, whereas forHaeckel it was a proven law. This was the first of a number of disputes over the processes of development and heredity that leftHaeckel scientifically isolated. His

increasing popularity compensated for this, as his ideas contributed to the development of psychology, sociology and psychoanalysis. In 1906 Haeckel was a founder of the Monist League, in which he encountered opposition to his pantheism, particularly from the physicist Wilhelm Ostwald.

It has been suggested thatHaeckel’s monism supplied the scientific roots of National Socialism (Gasman 1971). He certainly endorsed eugenics, and from 1905 he was an honorary member of the German Society for Racial Hygiene. Although on occasion personally anti-Semitic, Haeckel’s monistic philosophy did not support anti-Semitism. However, his anthropology was based on progression from ’primitive races’ to Europeans. While for primitive races the survival of the fittest had been a historical factor, higher civilized races were immune from its workings. In 1872 he argued that the death penalty was necessary for irredeemable criminals, and ascribed ’less value’ to weak and sick persons.Haeckel noted that euthanasia of the malformed and crippled new-born children had served to maintain the military prowess of the Spartans, but that it had no place in modern, civilized society. This would suggest that certain Social Darwinist tenets were redundant in higher societies. While the Nazis veneratedHaeckel, it was on the basis of a very selective reading of his ideas. On the other hand, Haeckel was much appreciated for furthering materialism, by Lenin and by German socialists such as Walther Ulbricht. Although this was also a distortion, it gaveHaeckel an honoured place under communism. His evolutionary ethics interested a variety of thinkers, including Sigmund Freud, Magnus Hirschfeld (the campaigner for homosexual rights) and many authors and artists.

See also: Evolution and ethics; Evolution, theory of

PAUL WEINDLING

List of works

Haeckel, E.H. (1866) *Generelle Morphologie der Organismen (General Morphology)*, Berlin: G. Reimer. (Integrates Darwinism with cell theory and embryology.)

Haeckel, E.H. (1868) *Natürliche Schöpfungsgeschichte*, Berlin: G. Reimer; trans. E.R. Lankester, *The History of Creation*, London: Routledge, 1876.(Applies evolutionary theories to explain the development of life, culminating in human evolution.)

Haeckel, E.H. (1899) *Die Welträthsel*, Bonn: E. Strauss; trans. J. McCabe, *The Riddle of the Universe*, London: Watts, 1929.(A popular evolutionary cosmology.)

References and further reading

Gasman, D. (1971) *The Scientific Origins of National Socialism. Social Darwinism in Ernst Haeckel and the Monist League*, London: MacDonald.(Links monism and Nazism; now superseded.)

Holt, N.R. (1971) ’Ernst Haeckel’s Monistic Religion’, *Journal of the History of Ideas* 32: 265-80.(On Haeckel’s later pantheism.)

Sandmann, J. (1990) *Der Bruch mit der humanitären Tradition* (The Break With Humanitarian Tradition), Stuttgart: Gustav Fischer.(A careful exposition of Haeckel’s ideas, although it over-emphasizes selectionism.)

Weindling, P.J. (1989) ‘Ernst Haeckel, Darwinismus and the secularisation of nature’, inMoore, J.R. (ed.) *History, Humanity and Evolution*, Cambridge: Cambridge University Press, 311-29.(A contextualized exposition of Haeckel’s changing views.)

(1989) *Health, Race and German Politics between National Unification and Nazism*, Cambridge: Cambridge University Press.(A contextualized analysis of Social Darwinism and eugenics.)

Hägerström, Axel Anders Theodor (1868-1939)

Hägerström was professor of philosophy at Uppsala University, Sweden, from 1911 until 1933, and together with his pupil Adolf Phalén founded the Uppsala school of conceptual analysis. He first came to the attention in 1902 with a study of Kant's ethics, but his main claim to fame rests upon the anti-metaphysical and emotivist positions which he developed during the years 1905-39. Thus the two fundamental theses of his moral philosophy are that moral valuations are neither true nor false, and that the proper task of moral philosophy itself is only descriptive and not normative. In this connection he argued that the philosophical ideas of objective moral values and of absolute rights have created fanaticism and sharpened conflicts in human history. He is also well known for his work in jurisprudence as the founder of Scandinavian legal realism. Here again his thought is predominantly antimetaphysical, and he criticized many legal concepts for their metaphysical and magical elements.

1 Kant research

Hägerström's *Kants Ethik* (1902) is in part an attempt to find a coherent line of thought in Kant's ethics. But the first part of the book is a penetrating analysis of Kant's epistemology in general. Hägerström sharply criticizes the prevalent psychologistic interpretation of Kant (such as that of Vaihinger) according to which the central line of thought in *The Critique of Pure Reason* is a psychological theory of the causes of human knowledge; according to Hägerström's logistic interpretation it should be understood instead as a logical analysis of the reasons of knowledge (see Kant, I.; Neo-Kantianism).

2 Ontology and epistemology

The motto of Hägerström's *Selbstdarstellung (Self-description)* (1929) is '*Praeterea censeo metaphysicam esse delendam*' (Besides, it is my opinion that metaphysics ought to be destroyed). According to Hägerström metaphysics rests on a deep misunderstanding of the concept of reality: in metaphysics, reality (the universal concept) is conceived as itself something real. This error is a result of a confusion that is natural to the human mind; it is implied by the common-sense conception of temporal and spatial facts; it is connected, Hägerström argues, with the idea that the past is unreal, an idea which belongs to common sense but which, like many other common-sense ideas, is absurd.

By an abstract argument Hägerström tries to show that the real world must be identical with the spatiotemporal continuum of experience. But the concept of reality is not derived from experience; on the contrary, it is a presupposition of all experience (and a presupposition of every judgment). This is one of the rationalistic ingredients of Hägerström's epistemology. Another important ingredient is the criticism of subjectivism, that is, of the idea that the only thing immediately given for a subject is that subject itself and its perceptions. Hägerström's picture of reality is strictly deterministic, the principle of causality being a logical condition of all empirical knowledge. He rejects the conventionalist opinion that the principle of causality is just a postulate stipulated by science; nor does he share Kant's opinion that it is a synthetic a priori judgment. He argues instead that it is deducible from the logical principle of the self-identity of reality.

3 Moral philosophy and theory of value

The main thesis of Hägerström's famous inaugural lecture of 1911, *Om moraliska föreställningars sanning (On the Truth of Moral Ideas)*, is that moral valuations are neither true nor false. Even such moral valuations as are usually expressed by declarative sentences lack the character of judgment; the value predicate is an expression of emotion, and in every valuation an emotion is a constituent element. Hägerström was one of the first philosophers to hold a no-truth-value theory about moral valuations, and one of the first to propose an emotive theory that did not take a moral valuation as a judgment about an emotion, but as itself constituted by an emotion. Emotive theories must explain the existence of abstract value sentences (such as 'Honesty is something good'), since emotions are evoked by particular phenomena such as individual actions or persons, not by universal concepts. Similarly, it is also a problem for an emotive theory to explain the existence of inferences having a value sentence as a premise, since emotions cannot enter into an inference. As early as the 1910s Hägerström was conscious of these and similar problems, and gave interesting solutions.

The moral valuations are the primary elements of morals, but in addition there are, in common types of morality,

important elements which, according to Hägerström, are false judgments. The idea of free will, which is absurd according to Hägerström's determinism, plays an enormous role in the history of morals, and so does the idea of objective values, which again is false. Several metaphysical ideas are false ingredients of many kinds of morals: for example, a metaphysical concept of person is presupposed by many deontological norms and in the doctrine of human dignity; a metaphysical assumption is also presupposed by the idea of absolute values. Hägerström gave a complex analysis of the idea of duty and found certain absurd conceptions involved in it. Thus Hägerström combined his emotive theory of the primary elements in morality with an error theory about other elements. Hägerström was influenced in certain ways by Meinong, Westermarck and Nietzsche, but on the whole his moral theory is an original achievement (see Emotivism; Morality and emotions).

4 Legal philosophy

According to Hägerström several of our fundamental legal concepts are metaphysical and magical, and hence absurd. An example is the concept of ownership: a person's ownership is an intermediary link between its legal ground (for instance, a gift or a purchase) and its legal consequences. This link cannot be identified with any empirical fact; it is a metaphysical entity, or, more precisely, a metaphysical force or power. Thus, the transfer of an ownership between persons is a transfer of a metaphysical force, that is, a magical procedure. (Hägerström's strong interest in the social and cultural role of magic was stimulated by Frazer's celebrated *The Golden Bough* (1911-15). Hägerström's view on the essence of magic was not identical with Frazer's, but he was much influenced by his scheme of classifications and by certain details in Frazer's work.)

Hägerström made a thorough and comprehensive inquiry into certain central themes of Roman law, paying great attention to its magical elements. He analysed the connection between Roman law and the doctrine of Natural law. He rejected this doctrine in all its forms, and urged that even many modern authors who explicitly disavow the principles of Natural law have nevertheless resorted to such ideas. He also rejected the will theory and Kelsen's pure theory of law (see Kelsen, H.). He was the founder of the school of Scandinavian legal realism, according to which it is a main task of legal science to give a description of legal phenomena in strictly empirical terms, without any metaphysical or ethical ingredients (see Legal realism).

5 Philosophy of religion

In his lectures of the 1920s, Hägerström developed a theory of religion which resembles, in an important respect, his theory of morality. The primary element in religion, as in morality, is something which is constituted by emotions, and which is neither true nor false. With this element, however, are combined false ideas of different kinds, and in religious dogmatics such false ideas are of essential importance (for example, the idea of God as an objective reality and the common conceptions of the effects of prayer or sacramental acts); an important factor in the formation of these ideas is the objectification of the contents of religious emotions.

Hägerström made thoroughgoing psychological investigations of these false religious ideas, of their relations to the emotive element, to morals and to magic. His psychological views about emotions and about self-consciousness are very complex, and this is especially evident in his philosophy of religion. He analysed both positive religion and metaphysical religiosity. In his treatment of positive religion, Hägerström utilized his great classical learning by taking examples from Greek and Roman religion, but he also referred to the religions of India, to primitive religion and, most often, to Christianity. An important detail is his investigation of the relation between magic and religion in the Lutheran doctrine of justification by faith; another, his analysis of different forms of the consciousness of sin. In his treatment of metaphysical religiosity he dealt with Plotinus, Eckhart, Spinoza and others, analysing how mysticism passes over into scholasticism and scholasticism into mysticism (see Mysticism, nature of).

See also: Scandinavia, philosophy in

THORILD DAHLQUIST
ANN-MARI HENSCHEN-DAHLQUIST

List of works

Hägerström, A. (1902) *Kants Ethik im Verhältnis zu seinen erkenntnistheoretischen Grundgedanken systematisch dargestellt (Kant's Ethics Systematically Expounded with Reference to the Basic Ideas of his Theory of*

Knowledge), Uppsala: Almqvist & Wiksell.(An accurate and detailed interpretation (not a critical examination) of Kant's ethics and parts of Kant's epistemology.)
- (1908) *Das Prinzip der Wissenschaft. Eine logischerkenntnistheoretische Untersuchung. I. Die Realität (The Principle of Science: A Logical and Epistemological Inquiry, vol. 1, The Reality)*, Uppsala: Humanistiska Vetenskapssamfundet.(Hägerström called this his most important work.)
- (1911) *Om moraliska föreställningars sanning*, Stockhom: Bonniers; trans. T. Mautner, *'On the Truth of Moral Ideas'*, Canberra: The Australian National University, 1971. (Hägerström's inaugural lecture; also included in Hägerström 1964.)
- (1927-41) *Der römische Obligationsbegriff im Lichte der allgemeinen römischen Rechtsanschauung I-II (The Roman Concept of Obligation in the Light of General Roman Jurisprudence)*, Uppsala: Humanistiska Vetenskapssamfundet; trans. R.T. Sandin, in *Philosophy and Religion*, London: Allen & Unwin, 1964.
- (1929) 'Axel Hägerström', in R. Schmidt (ed.) *Die Philosophie der Gegenwart in Selbstdarstellungen (The Philosophy of the Present in Self-Descriptions)*, Leipzig: Felix Meiner, vol. 7; trans. R.T. Sandin, in *Philosophy and Religion*, London: Allen & Unwin, 1964.(The major part of this Selbstdarstellungen is mainly an account of Hägerström's ontology and his criticism of metaphysics.)
- (1939) *Socialfilosofiska uppsatser (Studies in Social Philosophy)*, Stockholm, 1966.(Eight papers on moral and political philosophy and on philosophy of law.)
- (1948) 'Lectures on So-called Spiritual Religion', trans. C.D. Broad, *Theoria* 14: 28-67.(A central text in Hägerström's philosophy of religion.)
- (1953) *Inquiries into the Nature of Law and Morals*, ed. K. Olivecrona, trans. C.D. Broad, Uppsala: Humanistiska Vetenskapssamfundet.(Contains a selection from Hägerström's philosophy of law.)
- (1964) *Philosophy and Religion*, trans. R.T. Sandin, London: Allen & Unwin. (Selected writings on moral philosophy and philosophy of religion.)
- (1987) *Moralfilosofins grundläggning (The Foundation of Moral Philosophy)*, ed. T. Mautner, Uppsala: Humanistiska Vetenskapssamfundet.(Contains a summary in English.)

References and further reading

Broad, C.D. (1951) 'Hägerström's Account of Sense of Duty and Certain Allied Experiences', *Philosophy* 26.(A discussion of Hägerström's moral philosophy.)

Broad, C.D. (1964) 'Memoir of Axel Hägerström', in A. Hägerström, *Philosophy and Religion*, trans. R.T. Sandin, London: Allen & Unwin.(A biographical sketch, mainly based on a biography published in Swedish in 1961 by Hägerström's daughter Margit Waller.)

Dahlquist, T. (1976) 'Some Pages from the History of Philosophy in Uppsala', in *Uppsala University: 500 Years*, Uppsala, vol. 5.(Discusses Hägerström's role in founding the Uppsala school of philosophy.)

Ehnmark, E. (1956) 'Religion and Magic - Frazer, Söderblom, and Hägerström', *Ethnos* 21.(An account of Hägerström's use of Frazer.)

Frazer, J.G. (1911-15) *The Golden Bough: A Study in Magic and Religion*, ed. R. Fraser, Oxford: Oxford Paperbacks, 1994.(Frazer's twenty-five year study of primitive customs, magic and superstitions around the world.)

Hedenius, I. (1941) *Om rätt och moral (On Law and Morals)*, Stockholm.(A book of central importance for the Scandinavian debate on Hägerström's moral and legal philosophy.)

Marc-Wogau, K. (1968) *Studier till Axel Hägerström s filosofi (Studies in the Philosophy of Axel Hägerström)*, Falköping.(An important critical discussion by one of Hägerström's former students.)

Marc-Wogau, K. (1972) 'Axel Hägerström's Ontology', in *Contemporary Philosophy in Scandinavia*, ed. R.E. Olson and A.M. Paul, Baltimore, MD, and London.(Discusses Hägerström's attack on metaphysics.)

Mautner, T. (1994) *'Vägledning till Hägerström studiet' (Guide to Hägerström Research)* , Uppsala: Humanistiska Vetenskapssamfundet.(Includes a complete bibliography and a list of extant manuscripts.)

Passmore, J. (1961) 'Hägerström's Philosophy of Law', *Philosophy* 36: 143-60. (Lucid exposition and critical discussion.)

Petersson, B. (1973) *Axel Hägerström värdeteori (The Value Theory of Axel Hägerström)*, Uppsala: Uppsala University Philosophy Department.(A lucid analysis of Hägerström's moral philosophy and value theory. Contains a summary in English.)

Sandin, R.T. (1962) 'The Founding of the Uppsala School', *Journal of the History of Ideas* 23: 496-512.(A largely historical account.)

Halakhah

The central ideal of rabbinic Judaism is that of living by the Torah, that is, God's teachings. These teachings are mediated by a detailed normative system called halakhah, *which might be translated as 'the Way'. The term 'rabbinic law' captures the form of* halakhic *discourse, but not its range. Appropriate sections of* halakhah *have indeed served as the law of Jewish communities for two millennia. But other sections relate to individual conscience and religious observance and are enforceable only by a 'heavenly court'.*

Although grounded in Scripture, halakhah*'s frame of reference is the 'oral Torah', a tradition of interpretation and argument culminating in the twenty volumes of the Talmud. God's authority is the foundational norm, but it is only invoked occasionally as superseding human understanding. Indeed, the rabbis disallowed divine interference in their deliberations, asserting, in keeping with Scripture, that Torah is 'not in heaven' (Bava Metzia 59b, citing Deuteronomy 30: 12).*

Given the lack of binding dogma in Judaism, halakhic *practice has often been regarded as the common denominator that unites the Jewish community. The enterprise of furnishing 'reasons of the commandments' (*ta'amei ha-mitzvot*), central to many thinkers in Judaism, accordingly reveals a great diversity of orientations. These range, in medieval Judaism, from esoteric mystical doctrines to Maimonides' rational and historical explanations; and among modern writers, from moral positivism to existentialism.*

1 Interpreting biblical law

Of the twenty-four books canonized in the Hebrew Bible, *halakhah* accords strictly binding status only to the norms recognized as having been promulgated in the five books of Moses, the Pentateuch. Revered as the document of God's revelation, these constitute Torah in its narrow sense, and are regarded as the core of Torah (instruction) in the broad sense, which encompasses all valid teachings in Judaism through to the present (see Bible, Hebrew).

The Pentateuch contains not only laws, but also their setting: the story of creation and of God's relationship with humankind in general and with the Israelite people in particular. This relationship is consummated in the Sinai Covenant, whose human partners become committed to God's commandments (*mitzvot*). Much of the Torah is indeed composed of laws, typically introduced by 'The Lord spoke to Moses'.

This divine legislation differs from familiar legal codes in both scope and form. Its scope extends to issues which in the modern West belong not to law but more typically to the realms of religious practice or personal morality. This broad scope is characteristic of *halakhah* throughout its history. Some contemporary scholars have therefore proposed to treat Jewish law as a subcategory of *halakhah* relating to those issues generally included in the realm of law. As to form, the legal passages in the Pentateuch hardly read as a coherent system of law. They contain many duplications, gaps and inconsistencies; and their texts are traced by Bible scholars to a variety of circles and periods between the times of Moses (thirteenth century BC) and the first exile (sixth century BC). While some of these biblical teachings undoubtedly embody ancient Israelite legal practices, others seem to enunciate utopian legislation, a prescriptive parallel to the social criticism of the post-Mosaic prophets.

Since all scriptural verses are, in the *halakhic* system, equally holy, the biblical heritage, from the outset, called for extensive interpretation and harmonization. For example, according to Exodus 21: 1-7, a fellow-Israelite bought as a slave must be set free at the seventh year; but if he formally refuses this option he remains a slave 'forever'. Leviticus 25: 39-43 restricts slavery in a rather different way: it makes no mention of a seventh-year release, but prohibits gruelling labour, and proclaims freedom for all Israelite slaves in the Jubilee year. Jeremiah 34: 8-16 suggests that neither legal scheme had the impact that the imperatives seem to expect. But the *halakhah* accepted both, offering a liberal harmonization: the slave could not be forced into gruelling labour, *and* was to be offered his freedom on the seventh year. Even if he chose servitude, 'forever' was interpreted as 'until the Jubilee'.

The Second Commonwealth (fifth century BC to first century AD) saw a great increase in reinterpretations and enhancements of Torah for diverse reasons, in particular among the 'Pharisees'. By contrast with their rivals, the Sadducees (the priestly party, who put primary emphasis upon the temple rituals), the Pharisees gave greater weight to personal spirituality. Inspired by biblical passages like Leviticus 19: 2 ('You shall be holy, for I the Lord

thy God am holy'), they sought to direct everyday life and behaviour towards an ideal of pervasive holiness. After the destruction of the Second Temple (70 AD), the Pharisees became, perforce, the preservers of Jewish norms and ideals. They and their heirs became known as 'the rabbis' (masters or teachers), and their work defines classical 'rabbinic Judaism', the socio-cultural context of *halakhah*.

2 Rabbinic Judaism: *halakhah*'s classical era

In the definitive work of early rabbinic Judaism, the Mishnah (compiled by Rabbi Judah the Prince in the Galilee, *c*.200 AD), some sections closely follow biblical law; others boldly modify it, not through outright rejection or explicit amendment, but through a mode of qualifications and reinterpretations, called midrash (exposition or explication) (see Midrash). The rabbis recognized the great disparity between rabbinic and biblical law, and emphasized that the 'written Torah' must be accompanied and complemented by the 'oral Torah'. Indeed, they came to describe the oral Torah as a historical concomitant of the written Torah and even claimed normative primacy for the oral law: 'God's covenant with Israel was in regard to the oral *torah*' (Gittin 60b) (see Theology, Rabbinic §5).

A fine example of rabbinic boldness is the Mishnah's treatment of capital punishment. Detailed codification of the many capital offences prescribed in biblical law is combined with a procedural policy explicitly aimed at making executions virtually impossible. The Mishnah's discussion of capital punishment concludes by citing Rabbi Akiba, widely considered the greatest of the rabbinic sages: 'Were [I] in the Sanhedrin [the High Court], no person would ever be put to death' (*Mishnah Makkot* 1: 7). The rabbinic policy expressed here reflects a pervasive religious humanism, harking back to the dictum of Genesis that humanity, and so each individual human being, is created in God's image.

The Mishnah faithfully preserves the record of disagreements among the rabbinic sages, but it typically omits mention of their argumentation and much more. Its condensed legal clauses became focal points for elaboration, explanation and enhancement. These afford much of the bulk of the Gemara, by which the Mishnah is expanded to form the Talmud. The Talmud seeks to leave no issue unexplained; indeed, its redactors seem committed to overexplanation, sometimes advancing three or four alternative answers to each of the questions it raises, or offering strings of arguments and counterarguments to render plausible each side of its myriad of debates. The Talmud's interest in reasons and ideas, rather than in mere legal pronouncements, is further reflected in the fact that perhaps a third of it consists of *aggadah*, that is, narrative material - stories and parables, expositions of the non-legal parts of the Bible, theological and moral homilies, and the like.

3 Post-Talmudic Judaism: decentralized *halakhah*

After the redaction of the Talmud (in Mesopotamia, *c*.500 AD - hence the name 'Babylonian Talmud', as distinct from the earlier and less influential 'Jerusalem Talmud'), Judaism saw a shift away from a single cultural centre. The commitment of Jewish communities throughout the diaspora to the *halakhah* canonized the Talmud with its complex dialectic form, where hardly any debate ends with an explicitly stated definitive conclusion. Practical rulings in moral, legal, and ritual matters of *halakhah* came to rest in the hands of local courts or individual rabbinic scholars, masters in the art of Talmudic interpretation. Jewish communities had a considerable measure of legal autonomy as tolerated minorities, and through their various internal processes endowed such courts and scholars with *halakhic* authority.

This reality fit well with the classical rabbinic application of the dictum of Deuteronomy that the Torah is 'not in heaven': God's will is to be determined not by prophecy or oracles but by study and interpretation. Thus, despite the system's ultimate grounding in revelation, post-Talmudic Halakhists rarely appeal to Scripture - let alone to revelation - in arguing or deciding points of *halakhah*. Instead, they focus on the oral Torah, God's word as mediated by tradition, precedent and practical reason (*sevara*). When Moses Maimonides sought to offer, in his monumental *Mishneh Torah*, a univocal codification of Talmudic law, his code was widely respected but not accepted as final or binding. One central criticism was that he had written 'like a prophet', failing to support his rulings or to quote and engage alternative opinions and interpretations. Despite the recognized usefulness of Maimonides' code, and others as well, then, the most important repository of *halakhic* sources from medieval and early modern times was and remains the Responsa literature: wide ranging and unsystematic collections of reasoned rulings, resembling court opinions, issued by eminent scholars in response to written queries.

4 Contemporary *halakhah*

Halakhic authority reached a watershed with the political changes of the eighteenth and nineteenth centuries, which greatly undermined the jurisdiction of *halakhic* courts. For Orthodox Jews, *halakhah* continues to function as a system of religious law: through contemporary Responsa, essays and monographs, it addresses a broad spectrum of issues in modern life. Conservative, Reform, Reconstructionist, and many other Jews relate in their own ways to *halakhah* as a less binding framework, which nevertheless informs their normative thinking. In the state of Israel, *halakhic* courts have limited jurisdiction, primarily in matters of personal and family law. The contemporary Jewish scene thus reflects an unprecedented diversity of perspectives regarding *halakhic* concepts and norms.

For example, an issue of major contention concerns defining the Jewish community itself. In rabbinic Judaism, joining the community from without was perceived as a transformation of religious identity. The national and religious communities were one and the same, and the convert joined the Jewish people through a ritual defined by religious law. In the modern setting, however, many Jews (both in Israel and in the diaspora) see themselves as members in a national or ethnic collectivity rather than as adherents of a religion. While the Orthodox continue to require ritual immersion for those who wish to become Jewish, and (for males) circumcision, accompanied by (at least an overt) commitment to *halakhah*, and overseen by a rabbinic court of three men, other circles, drawing on Talmudic sources, have adopted alternative criteria and procedures.

5 Philosophies of *halakhah*

Philosophical discourse on the *halakhah* has traditionally focused on two main problems: rationality and coherence.

The rationality of *halakhah* is problematic in light of the numerous details, mainly in ritual law, which may seem alien to the dictates of reason. The rabbis were prepared to grant, on occasion, that particular commandments were ‘divine decrees’, unfathomable demands of the divine Sovereign. But to medieval religious philosophers, who had come to share Aristotelian traditions, a much greater segment of *halakhah* appeared hard to justify. Starting with Saadiah Gaon al-Fayyumi, some offered a distinction between ‘commandments of reason’ and ‘commandments of obedience’; even the latter, however, were in principle justifiable - for example, by the value (intrinsic or instrumental) of obedience itself.

The second issue in the philosophy of *halakhah* relates to the system’s formal coherence, rather than to the substance and sense of its particular elements. The rabbinic idea of norms grounded in God’s revealed words, but subject to extensive midrashic remoulding, seems to pose a fundamental paradox: If the autonomy of human reasoning is upheld against the divine author’s intention, has not the commitment to revelation become insincere? This challenge was posed most poignantly by the Karaites, a powerful movement in medieval Judaism radically opposed to midrash and to the entire project of an oral law (see Karaism).

Rabbinic responses in this case diverge into two different tracks. Some play down the scope of innovation in *halakhah*, insisting that the entire oral law was revealed at Sinai. Others seek to distinguish between an unchanging core of divine law and a supple, growing body of human teachings.

Some views of Judaism emphasize elements outside of *halakhah*; they may contain a philosophy of *halakhah* but do not make it their main subject. This is true of Maimonides’ philosophy, as presented in his *Guide to the Perplexed*, which views the entirety of the divine law as a means to an end, the attainment of human perfection through the realization of the inner human propensity to become Godlike. It is equally true of various systems of Jewish mysticism (see Kabbalah). In their different ways, both rationalists of the Maimonidean persuasion and mystics of Kabbalistic inclination have seen intimate knowledge of God as the focus of religious life and have gauged the import of *halakhah* from that standpoint. The work of Franz Rosenzweig represents a modern counterpart to such approaches.

See also: Bible, Hebrew; Law, philosophy of; Leibowitz, Y.; Maimonides, M.; Theology, Rabbinic

NOAM J. ZOHAR

References and further reading

Berkovits, E. (1983) *Not in Heaven*, New York: Ktav.(Restatement of the traditional doctrine regarding the primacy of the oral Torah and the role of human reason in *halakhah*.)

Elon, M. (1994) *Jewish Law: History, Sources, Principles*, trans. B. Auerbach and M.J. Sykes, Philadelphia, PA: Jewish Publication Society.(A comprehensive description and analysis of the *halakhic* legal system: its constitution, textual sources, and evolution.)

Jacob, W. and Zemer, M. (eds) (1991) *Progressive Halakhah: Essence and Application*, Tel Aviv and Pittsburgh, PA: Freehof Institute of Progressive Halakha.(Essays on the problems of 'Liberal' (Reform) Jewish Law.)

Roth, J. (1986) *The Halakhic Process: A Systemic Analysis*, New York: Jewish Theological Seminary. (Well-documented, systematic presentation of modes of *halakhic* change.)

Halevi, Judah (before 1075-1141)

Physician, philosopher and perhaps the greatest Hebrew poet since the Psalms, Judah Halevi studied the Neoplatonic Aristotelianism widespread in Islamic Spain, but his loyalty to Judaic traditions, love of Israel and poetic empathy for the sufferings and aspirations of his people made him a powerful critic of that philosophical tradition. His philosophical masterpiece, the Kuzari, is a fictional dialogue set at the court of the king of the Khazars, a people of the Volga basin whose leaders had converted to Judaism in the early ninth century. Reports of the Khazar realm sparked Halevi's imagination and gave him the backdrop for this effort to celebrate and shape his ancestral faith.

Beyond heartening his fellow Jews in times of upheaval, Halevi confronted philosophical questions that conventional thinkers often begged or ignored. He found the erudite Neoplatonism of his day too confining to God, too speculative and a priori. Tellingly, he condemns Neoplatonism for cultural vacuity, moral sterility and spiritual escapism: while Christians and Muslims, with the highest spiritual intentions, earnestly set about one another's murder, the philosophers fail to differentiate one faith from another. What is needed, Halevi reasons, is not a still more spiritual intellectualism but a historically and geographically rooted tradition concretely directed by God's love.

Halevi did not, as romantics often suppose, simply turn his back on reason, or on philosophy generically. Rather, he used his own philosophical gifts and poetic tact to retune philosophy to the ground notes of Jewish experience. He retained but structurally adapted the Neoplatonic linkage of God to the world via emanation, replacing the elaborate hierarchy of star souls with the simple manifestation of God's word, the 'Amr. *Like Philo's* Logos, *Halevi's* 'Amr *was at once an attribute of God, his wisdom and a manifestation of God immanent in nature. Since the* 'Amr *is an imperative, it connotes power, volition and command, not just logical entailment or necessitation. Since it is immanent, it allows fuller appropriation than was possible for many philosophers and many of the pietists and mystics in their wake, of the material side of nature, including human nature: language, material culture - including agriculture and other economic activities - law and politics belong to realm of God's expression. Particularity is not isolated from God. Poetry and works of imagination can be expressions of the divine, not just stepchildren mediating the ever more abstract and abtruse flights of the intellect. Zion could be acknowledged as the land where the divine afflatus was most clearly articulated as a way of life. Longing for Zion need no longer be sublimated in prayer; rather, Israel's songs of longing for the robust life of the land of God's grace would voice a spiritual imperative that demanded practical expression and historical realization.*

1 Life

Born in Toledo and broadly educated in Arabic and Hebrew letters and sciences, Judah Halevi became a physician and poet in the Spain of the early Reconquista period. His poems speak of love, wine, gardens, friendship and the death of friends, of spiritual quest, devotion and the love of God, his Hebrew reluctantly attuned to Arabizing rhymes and cadences; but the ruined campsite that once prompted elegies to lost loves in pre-Islamic odes, is now the ruined Temple mount. Alienation gives edge to Halevi's poetry and thought. Philosophy, as he finds it, shows deep rifts between ideals and practice. Medicine is wanting both technically and spiritually. The Christian rulers of Northern Spain seem inhuman taskmasters, with the Jews their ministering slaves: 'we heal Babel, but it will not be healed' (letter to David Narboni, in *Diwan*, Brody 1974: 224).

Leaving Castile and settling in Cordova, Halevi saw the streams of Jewish refugees who fled the Almoravids and the Christian plundering and destruction of Jewish towns, and he learned of the terrors of the First Crusade. 'How can I savour what I eat', he wrote, 'when Zion is in Christian chains, and I in the shackle of Islam?' ('Libbi be-Mizrah', in *Diwan*, Brody 1974: 2). The smouldering scenes of a medieval Guernica, etched in his verses, made Halevi a proto-Zionist. Moving between Christian and Muslim Spain, protected by Jewish courtiers who prized his poetic and medical skills, Halevi saw no respite for his people.

Restless and troubled, he longed for Zion: 'My heart is in the East, but I am in the farthest West' ('Libbi be-Mizrah', in *Diwan*, Brody 1974: 2). The East here was more than geographical. The very direction held a spiritual significance for Halevi, conveyed in ideas of light, life and enlightenment. But could the East even in that sense exclude the earthly land of Israel, the land where the people of Israel had lived their earliest history and

fullest life as a nation? Could even the spiritual East be less than Zion? How could one mourn the lost Jerusalem and pray for its restoration, yet remain in exile? Friends might urge complaisance, but as Halevi's historic vision deepened and darkened, his friends seemed drunk. Casting them in the stock role of the 'reproacher' of Arabic love lyrics, he turned on them: 'Is it good for a pure and honest man to be led about like a captive bird in the hands of children?' Halevi expanded on such questions in his *Kuzari*, an Arabic philosophical dialogue (the full title of the work is *Kitab al-radd wa 'l-dalil fi 'l-din al-dhalil (A Defence and an Argument on behalf of the Abased Religion)*.)

2 The *Kuzari*: the basis of commitment

Written between 1130 and 1140, the *Kuzari* reflects a striking episode of Jewish history. In the early ninth century, Bulan, King of the Khazars, a people of the Volga basin, along with four thousand of his nobles, had adopted Judaism. Under Byzantine and Muslim pressure, the king sought his own monotheistic faith. Granted religious freedom, most Khazars probably never became Jews; but until the Khazars were swept away in the Tatar invasion of 1237, Judaism was the state religion. Hasdai ibn Shaprut, *wazir* of 'Abdu 'l-Rahman III of Cordova, thrilled to hear of a powerful Jewish state in the East, wrote to the Khazar monarch around 960. Replying, King Joseph told of the Khazar conversion and described his kingdom.

The conversion had followed a debate among Christian, Jewish and Muslim spokesmen, which Halevi fictively reconstructed. The king has dreamt that his intentions, but not his practices, are pleasing to God. He consults a philosopher but hears only a disappointingly intellectualist Neoplatonic Aristotelianism. 'If philosophers say that God created you, that is metaphorical, of course. For He is the Cause of all the causes that conspire in the creation of all things' (*Kuzari* 1.1).

The philosopher argues smoothly, with many therefores and an almost equally smooth stream of disembodied intellects through which God's act reaches the world: 'The world is eternal. Human beings have always arisen one from another, their forms compounded and their characters formed by those of their fathers and mothers, and by their environment - airs, lands, foods and waters - along with the influences of the spheres, constellations and signs of the Zodiac' (*Kuzari* 1.1). Man's goal is to purify the soul, for the perfect may reunite with the nearest hypostasis, the 'active intellect' - which the perfect man, ignoring mere limbs and organs, already is. Religion is a valuable moral conditioner, especially for the masses. Once its function is grasped, however, it may be moulded at will.

The king finds the philosopher's argument 'impressive', but unhelpful. He still seeks a way of life pleasing to God. Christians and Muslims divide the world; with the highest intentions, they earnestly pursue one another's murder. Philosophy seems to wish to stand above the fray, deeming all monotheists to be truth seekers. But how can these enemies all be right? The philosopher's affirmation, 'In the faith of the Philosophers there is no such killing, since we foster the mind' (*Kuzari* 1.1), betrays a stunning innocence. The highest intentions, it seems, do not differentiate acts of heroism from obscene atrocities. Ethos is grotesquely under-determined by ideology. *Khalas* (sincerity) must be given a material, not merely formal sense; devotion must extend to action, not just in cultivating personal virtues but communally. Neoplatonists, pietists and ascetics still must address the embodiment of their ideals and the delicate, dangerous nexus between ethos and ethnicity. The philosophers have paid too little mind to history.

The king grows curious about Christians and Muslims: 'Surely one of these two ways of life is the pleasing one' (*Kuzari* 1.1). The obvious abasement of the Jews seems to exclude theirs. A Christian and a Muslim speak, appealing, respectively, to the divinity of Christ and the inimitable language of the Qur'an. The king's responses vividly signal Halevi's outlook: a little philosophy would not hurt the Christian case. 'Not having grown up in these beliefs', the king feels no need to seek credibility for them. Logic will accommodate experience, but that is because experience takes root in the heart (*Kuzari* 1.5).

Like Muhammad, the Muslim spokesman avoids appeal to miracles, except the Qur'an, whose every verse Muslims call a portent. Again the Khazar answers existentially:

> If one hopes for guidance from God's Word and hopes to be convinced, against his own scepticism, that God does speak to mortals, things ought to be manifest and incontrovertible. Even then one would hardly credit that God spoke to a man. If your book is miraculous, being written in Arabic its uniqueness and inimitability are

> indiscernible to a non-Arab like me. When read to me, it sounds like any other Arabic book.
>
> (*Kuzari* 1.6)

Since both Christian and Muslim cite Jewish history, the Khazar summons a rabbi, called a *Haver* or Fellow of a Talmudical academy. The rabbi opens not with a cosmological credo but an invocation of the God of Abraham, Isaac and Jacob. He avoids proofs like the design argument:

> 'If you were told that the ruler of India was a virtuous man whom you should hold in awe and whose name you should revere, but his works were described to you in reports of the justice, good character and fair ways of the people of his land, would that bind you to him?'
>
> 'How could it, when the question remains whether the people of India act justly of their own accord and have no king at all, whether they do so on account of their king, or whether both are true?'
>
> (*Kuzari* 1.19-20)

Mere argument does not show whether the order and design of nature are the work of God, intrinsic to nature, or Neoplatonically imparted by God through the natures of things:

> 'But if a messenger came to you from that king with Indic gifts, that you were certain could be had only in India, and only in the palace of a king, and he brought you a written attestation that these came from the king, and enclosed medicines to treat your illnesses and preserve your health… would this not bind you to his allegiance?'
>
> (*Kuzari* 1.21)

Similarly, God confronts Pharaoh (Exodus 5: 1) not as the creator but as the ancestral help of Israel; at Sinai (Exodus 20: 2), he is the God who saved Israel from Egypt. What mattered then was not what God had done for the universe, but what he had done for them. Only Israel, the Rabbi argues, has a true and continuous tradition regarding the divine. India may be ancient, but its people have no coherent system of ideas, and they are polytheists. Greek philosophical originality is a *tour de force*, but it shows the unsteady gait of solecism. Aristotle has no tradition to keep him from going overboard, as when he ascribes intelligence to nature at large.

The religion of Israel, the *Haver* argues, uniquely combines the intimacy of God's unique historical relationship with Israel and the publicity of the entire nation's experience of God's act, receipt of his gifts and their written attestation, passed down undisrupted through the generations, preserving the certitude of God's ancient self-revelation. The true religion did not evolve like man-made faiths. Like creation, it was complete in a moment, when 600,000 Israelites experienced their redemption, and after wandering in the desert, somehow heard God's words, each individual directly and personally inspired.

Reacting to the palpable chauvinism (*ta'assub*) of the *Haver*'s claims, the king asks if the sin of the Golden Calf does not diminish the rabbi's all but insufferable pride. Every nation, the *Haver* replies, was full of idolaters at the time. Any philosophers who could prove that God was one would still have rationalized idolatry as symbolism mediating the divine to the masses. Israel's backsliding was grievous principally because the sin was theirs. The people sinned, but they were forgiven. What matters is that they were chosen. They preserved the perfection in which Adam was created. The perfect climate of the God-given land would prepare them to live by God's word. Prophecy will never leave them.

Israel's great gift is not the specious reward of a sensuous afterlife, or even a spiritual afterlife that no one really wants, but the abiding presence of the divine in this life. Philosophers imagine that only pure intellects are immortal; Muslims and Christians compound such exclusivism with superstitious notions that lipservice somehow confers immortality. But Jews believe that God rewards the righteous of all nations. What distinguishes Israel lies in this world. Jews do not spend their lives begging their king for safe conduct on the parlous journey that every mortal must undertake; they live confident that God, by whose word they live, will not forsake them.

Convinced that here must be the way of life his dream foretold, king and *wazir* embrace Judaism and gradually win over many of their nation. They study the Torah and gain great worldly success, honoring the Israelites among them as the first and most fully Jewish of their countrymen. Only after extensive study of the Torah does the king begin to inquire speculatively into its theology; existential commitment comes first.

3 The *Kuzari*: idea and embodiment

Through the dialogue, Halevi proposes that God is described in three ways: (1) negative predicates, indicative of perfection, as when we say ‘the living God,’ to distinguish him from the dead, false gods of idolaters; (2) relative predicates, expressing human attitudes toward God, as when we call him blessed and exalted; and (3) predicates of creativity, which speak of God’s acts through some natural medium or agency, as when we say, ‘making poor and rich’ God’s agency in nature is his will, the motive force behind all natural and supernatural events, and the source of the created glory that manifests God’s grace on Israel in their land, the favoured place for its appearance. When Israel dwells on its soil in peace and justice, prophecy becomes possible among the pious. All true prophecy was either in or for the land of Israel, the centre of the globe and reference point of day and night, east and west. Its very air imparts wisdom. Thus the sages were not misled in saying that one who walks four yards there is assured of happiness in the world to come: such a person already tastes transcendence.

Surely then, the king responds, the rabbi is remiss in not returning to that land. For even if the Shechinah, God’s immanence, is no longer present there, one should seek to purify the soul in such a holy place, if only because it once was there. The rabbi accepts the reproach, answering only that Israel’s return has always depended on the people’s willingness, for ‘God’s Word grants a man no more than he is capable of receiving’ (*Kuzari* 2.24). Israel today is no longer a body but only dry bones. Yet these bones preserve a trace of life, which can return to them if the Temple, which animated them, and made them vulnerable, is restored.

Adapting a thought that Plato had applied to philosophers, the Rabbi says that Israel is the heart among the nations, at once the most vital and the most vulnerable. Prophecy will return, but prophets and saints are not the same as hermits and ascetics. Mere renunciation does not achieve the intimacy with God that makes a nation the true seedbed of prophecy. Justice, not humility or spirituality, is the natural, rational and necessary foundation of a nation’s life: ‘The divine law cannot be fulfilled until the civil and rational laws are perfected.’ Israel cannot survive and fulfil God’s commandments as a soul without a body. Not withdrawal and asceticism are demanded but the full life of an economy and a state. It is as much a divine commandment to labour and cultivate the soil as it is to keep the Sabbath, for both celebrate God’s act of creation and the liberation of Israel from Egypt.

Piety is not best shown by upturned eyes, meditative postures and gestures, or fine words that intend no action, but by genuine commitment and sincere intentions, manifested in demanding actions performed with dedication. What Halevi calls for here is not simply a return to Zion; not mere spiritual longing, but the reconstitution of the full, robust life of Israel in its land, under its laws. The members of the Sanhedrin, he argues, were responsible for knowledge of every science: the authentic sciences, to fulfill the intentions of the Law and preserve the health and welfare of the people; the conventional sciences, to perfect their use of language; and even the specious sciences, evidently to understand ambient superstitions. These sciences, whose relics still distinguish Jews, must be restored, along with the Hebrew language, which has fallen into decline since the days of the psalmists and has become the toy of lackeys and misfits.

Israel’s aim is not otherworldliness. We love life and all its goods. True, a man who reaches moral perfection, like Enoch or Elijah, will grow uncomfortable in the world and will feel no isolation in solitude. Yet today, when there is no clear vision, the good man must be the guardian of his country. He must give all his powers their due, preparing them to serve when called on. The king is surprised at so political an idea of personal goodness, but the rabbi replies that human goodness is political for Plato’s reason, that it rests on command over one’s powers: ‘The good man is the prince, obeyed by his senses, and by his spiritual and physical powers…. It is he who is fit to rule. For if he led a state, he would apply the same justice in it as he does in governing his own body and soul’ (*Kuzari* 3.5).

Human rationality regulates the good man’s life. However, God adds further refinements to the life of Israel, rendering specific the generic obligations of reason, and instituting the visible symbolisms that save the idea of a covenant with God from reduction to a mere abstraction. The good Israelite lives in the thought that God is ever-present; the world is not finished or abandoned but an ongoing creation, in which our own words and the songs that spring from our mouths at God’s behest issue without our least knowledge of how the God-given powers of the body and creativity of the mind spring to his service.

Obedience, not zeal, is God’s desire. Moderation, not excess, anchors God’s plan. Just as only God, and no mere

alchemist, knows the proportions needed to compound a living body, so no mere tinkerer can fashion a law of life. No individual can possibly replace the careful, systemic modulation of the good life. Halevi's target here is the Karaites, whose rejection of oral or Talmudic Law seemed to leave each individual to comprehend the Torah in isolation (see Karaism). Without an oral tradition, we would not even know how to vocalize or parse the text, let alone govern by it. True, the Karaites show originality, the product of their isolation. They are like foot soldiers in no man's land; the Rabbanites, secure in their traditions, seem as confident and lax as the populace of a walled city. But on Karaite principles, there would be as many codes as opinions.

The Oral Law is stricter than the Law of Moses, making a margin (*seyag*) around the Torah, but that allows the Rabbis to qualify and mitigate their rulings, guided by God's still present Word. We are not confined to the flickering light of reason but know God through our intercourse with him and the long history of our growing awareness, traceable in a tradition recorded by Scripture all the way back to Adam.

God's will is executed in nature without intermediaries, the *Haver* argues, with a sidelong glance at the Neoplatonic cosmos, clogged with disembodied intelligences. Like other critics, including al-Ghazali, Halevi finds the idea that emanation is the truth behind scriptural creation reductionistic. It treats God's creativity too much as a mechanism or a necessity of logic, placing God at a remove and setting the ideas of things between God and his creatures. Knowing that his battle is with intellectual authority, Halevi turns satirical, asking how it is that thought in disembodied intellects generates a sphere. When Aristotle reflected, no such sphere arose!

Dispensing with the entire complex disembodied apparatus, Halevi ascribes God's knowledge and governance of the world directly to his Word, preserving the double edged efficacy of the ancient Philonic Logos: immanent in nature yet in no way separate from God, since it is His will. The *'Amr* or divine word of command is still, in a way, an emanation. For it does convey the divine idea and impress it upon the world, the Land and the prophets, who are recipients of inspiration. But the emergence of the Word from God in Halevi, like the initial differentiation of the First Essence in Ibn Gabirol, is now volitional. The work of emanation is no longer conceived through a mystification of logic that makes entailment somehow a source or vehicle of creation. The impact is to immanentize divine volition. This approach has a long afterlife: in Maimonides' theory of angels as forms and forces, in Kabbalistic developments pioneered by Nahmanides, in Spinoza's *conatus*, Bergson's *élan vital*, Whitehead's creativity and beyond.

In dismissing intellectualist emanation, Halevi has not rejected logic or philosophy; he has rejected the specific product that prominent philosophers ascribed to logic. Only tradition can explain their reliance on arguments as suppositious as those behind the Neoplatonic ontology. However, Halevi faults the philosophers not for their logic but for their want of logic. The naturalism to which Halevi appeals in rejecting the idea that mere self-reflection or contemplation of the divine can entail spheres or angels into being is philosophic, although it is corrosive to the intellectualist assumptions of the prevalent philosophical school of his day.

Philosophers may say that love of God follows from knowledge of his omnipotence, but such inferences are too abstract to command the heart. Similarly, even the most impressive arguments of the philosophers find no following among the common people, not because the people are too crude, as the philosophers suppose, but because the philosophers are too far removed from life. Humans need language, images and rituals. What we need to know of theology is that the world is created, that it has a cause in God, who is eternal and unconditioned, incorporeal, omniscient and omnipotent, living and willing eternally, and that the human will, like God's, is free. Volition is delegated to human beings, just as natural dispositions are imparted to all things. For if an external determinism were true, 'a man's speaking would be compulsory, like his pulse' (*Kuzari* 5.20).

The *Haver* draws his teaching to a close with new thoughts of Zion, which has never been far from the aim of his argument. The life of Israel in her land and the will of God are of a piece. The rabbi dismisses the pious notion that Israel's sins debar return. The liturgical confession 'and for our sins were we exiled from our land' is hortatory and admonitory, not explanatory or normative. The Psalmist's lines (102: 14-15) 'Thou wilt arise and take pity on Zion, for the time to favour her is here, the time is come - since Thy servants delight in her stones and cherish her dust', mean that Jerusalem will be rebuilt when Israel so yearns for it as to cherish her very stones and dust. The gloss sums up the hearty rootedness that Halevi opposes to Neoplatonic intellectualism and pietist asceticism. Responding, the king offers a courtly opportunity for his teacher to take his leave: 'If this be so, it would be culpable to detain you' (*Kuzari* 5.28).

Acting on the same conclusion, Halevi left Spain in 1140, travelling to Egypt and on to Palestine. In legend, he lived to kiss the ground outside Jerusalem, where, as he spoke the words of his Ode to Zion, he was killed by an Arab horseman. But the meaning of his journey is best captured in the direction he took, encoded not in the poet's life but in his work and above all in the *Kuzari*, with its intellectually serious call for the reintegration of Israel, body and spirit, land, language and Logos, the freely imparted and freely chosen direction of God's eternal Idea.

See also: God, concepts of; Zionism

L.E. GOODMAN

List of works

Halevi, Judah (*c*.1130-40) *Kitab al-radd wa- 'l-dalil fi 'l-din al-dhalil (Kuzari)(A Defence and an Argument on behalf of the Abased Religion)*, ed. D.H. Baneth, Jerusalem: Magnes Press; trans. H. Hirschfeld, London: Routledge, 1905; repr. New York: Schocken, 1974.(The Hirschfeld translation is misleadingly imprecise. A new translation is in preparation as part of the Yale Judaica series. Passages quoted here are translated by L.E. Goodman.)

Halevi, Judah (before 1141) *Diwan*, ed. H. Brody, *Jehuda Ha Levi, Die Sch'nen Versmasse*, Berlin, 1894-1930, 4 vols; ed. H. Brody, *Selected Poems of Jehudah Halevi*, Philadelphia, PA: Jewish Publication Society, 1974. (The Berlin edition is the classic four-volume edition of Halevi's poetry, which includes his secular poetry in vols 1 and 2 and his religious poetry in vols 3 and 4; selections from this edition are reprinted, with English translation by S. Bernstein in Brody's *Selected Poems*, which was first published in 1924.

Halevi, Judah (before 1141) 'Letter to Rabbi Haviv' (in Hebrew), ed. H. Ratzhaby, 'A Letter from R. Judah Halevi to R. Haviv', *Gilyonot* 28, 1953: 268-72.(Useful for biographical detail on Halevi.)

References and further reading

Allony, N. (1951) *Torat ha-Mishkalim - The Scansion of Medieval Hebrew Poetry: Dunash, Jehuda Halevi, and Abraham ibn Ezra*, Jerusalem.(In Hebrew, this work discusses Halevi's theories of prosody and his attitudes towards meter in Arabic and Hebrew.)

Baer, Y. (1971) *A History of the Jews of Christian Spain*, trans. L. Schoffman, Philadelphia, PA: Jewish Publication Society.(Valuable exposition of the historical backgrounds.)

Baljon, J.M.S. (1958) 'The *amr* of God in the Koran', *Acta Orientalia* 23: 7-18. (Explores the Islamic *logos* theory that influenced Halevi's conception.)

Baneth, D.H. (1981) 'Judah Halevi and al-Ghazali', trans. G. Hirschler in A. Jospe (ed.) *Studies in Jewish Thought: An Anthology of German Jewish Scholarship*, Detroit, MI: Wayne State University Press, 181-99. (First published in German in 1924; important comparative study.)

Baron, S. (1960) *A Social and Religious History of the Jews*, Philadelphia, PA: Jewish Publication Society. (Includes detailed discussions of the poet and his times.)

Brann, R. (1991) *The Compunctious Poet: Cultural Ambiguity and Hebrew Poetry in Muslim Spain*, Baltimore, MD: Johns Hopkins University Press.(A study of the ambivalent reception of Arabic themes and methods by Hebrew poets.)

Davidson, H. (1972) 'The Active Intellect in the Cuzari and Hallevi's Theory of Causality', *Revue des Études Juives* 131: 351-96.(Explores Halevi's ideas on the metaphysics of emanation.)

Dunlop, D.M. (1954) *The History of the Jewish Khazars*, Princeton, NJ: Princeton University Press.(A scholarly review of the evidence avoiding the extravagant speculations of some popularizers.)

Goitein, S.D. (1967) *A Mediterranean Society*, Berkeley, CA: University of California Press.(Magisterial study of the Cairo Geniza materials, which preserve primary evidence for the life and times of Halevi.)

Hamori, A. (1985) 'Lights in the Heart of the Sea: Some Images of Judah Halevi's', *Journal of Semitic Studies* 30: 75-83.(Examines Halevi's use of Neoplatonic imagery in his poetry.)

Motzkin, A. (1980) 'On Halevi's *Kuzari* as a Platonic Dialogue', *Interpretation* 9: 111-24.(A Straussian reading.)

Neumark, D. (1929) 'Jehuda Hallevi's Philosophy', in S. Cohon (ed.) *Essays in Jewish Philosophy*, Cincinnati, OH: Central Conference of American Rabbis; repr. Amsterdam: Philo Press, 1971, 219-300.(Draws the connection between Halevi and Ibn Gabirol.)

Pines, S. (1960) *Encyclopedia of Islam*, London: Luzac, 1.29-30, s.v. *'amr*.(This article connects Halevi's use of the term *'amr* with the Islamic background and usage.)

Pines, S. (1980) 'Shiite Terms and Conceptions in Judah Halevi's *Kuzari*', *Jerusalem Studies in Arabic and Islam* 2: 165-219.(Further clarification of the Islamic background for Halevi's philosophical theology.)

Scheindlin, R. (1986) *Wine, Women, and Death: Medieval Hebrew Poems on the Good Life*, Philadelphia, PA: Jewish Publication Society.(Excellent survey (with translations) of the thematics of Hebrew poetry of Islamic Spain.)

Scheindlin, R. (1991) *The Gazelle: Medieval Hebrew Poems on God, Israel and the Soul*, Philadelphia, PA: Jewish Publication Society.(Sequel to Scheindlin 1986.)

Silman, Y. (1995) *Philosopher and Prophet: Judah Halevi, the Kuzari and the Evolution of his Thought*, trans. L.J. Schramm, Albany, NY: State University of New York Press.(Analyses the *Kuzari* from a developmental perspective, showing how it embeds elements of the poet's early attachment to the Neoplatonic philosophical tradition and how the work grows beyond its original aim, as a riposte to Karaism.)

Spiegel, S. (1976) 'On Medieval Hebrew Poetry', repr. in J. Goldin (ed.) *The Jewish Expression*, New Haven, CN: Yale University Press.(A compact but wide-ranging survey of the medieval Hebrew poets, their themes and their relations with Biblical and post-Biblical poetry, liturgy and ideas.)

Strauss, L. (1952) 'The Law of Reason in the *Kuzari*', in *Persecution and the Art of Writing*, New York: Free Press, 95-141.(Classic study from the volume in which Strauss established the methodology of a still influential school of thought focused on the idea of indirection in the texts of philosophical past masters.)

Hamann, Johann Georg (1730-88)

Hamann was one of the most important critics of the German Enlightenment or Aufklärung. *He attacked the* Aufklärung *chiefly because it gave reason undue authority over faith. It misunderstood faith, which consists in an immediate personal experience, inaccessible to reason. The main fallacy of the* Aufklärung *was hypostasis, the reification of ideas, the artificial abstraction of reason from its social and historical context. Hamann stressed the social and historical dimension of reason, that it must be embodied in society, history and language. He also emphasized the pivotal role of language in the development of reason. The instrument and criterion of reason was language, whose only sanction was tradition and use.*

Hamann was a sharp critic of Kant, whose philosophy exemplified all the sins of the Aufklärung. *Hamann attacked the critical philosophy for its purification of reason from experience, language and tradition. He also strongly objected to all its dualisms, which seemed arbitrary and artificial. The task of philosophy was to unify all the various functions of the mind, seeing reason, will and feeling as an indivisible whole.*

Although he was original and unorthodox, Hamann's critique of reason should be placed within the tradition of Protestant nominalism. Hamann saw himself as a defender of Luther, whose reputation was on the wane in late eighteenth-century Germany.

Hamann was also a founder of the Sturm und Drang*, the late eighteenth-century literary movement which celebrated personal freedom and revolt. His aesthetics defended creative genius and the metaphysical powers of art. It marked a sharp break with the rationalism of the classical tradition and the empiricism of late eighteenth-century aesthetics.*

Hamann was a seminal influence upon Herder, Goethe, Jacobi, Friedrich Schlegel and Kierkegaard.

1 Critique of reason

Born into a middle class family, Hamann spent almost all of his life in Königsberg, Prussia. He formulated many of his ideas in exchanges with his friend and neighbour, Immanuel Kant. From 1762 to 1764 he was the tutor of Johann Herder, who was profoundly influenced by him. For most of his life Hamann was a minor official in the Prussian bureaucracy, a position procured for him by Kant. Because of his mysticism and sibylline style he earned the sobriquet 'The Magus of the North'.

The starting point of Hamann's philosophy was his mystical experience and conversion, which took place in London in March 1758. Hamann went through a deep personal crisis that shattered his earlier allegiance to the *Aufklärung*. The solution to his crisis was a mystical experience, a rebirth through faith in Christ. During his experience Hamann had a vision of the divine omnipresence, of God dwelling within him and speaking to him through all the events of nature and history. The creation was the secret language of God, the symbolism by which he communicated his message of redemption to human beings. All of nature consisted in hieroglyphs, secret symbols and puzzles, which could be interpreted through Scriptures alone. Hamann wrote down his experiences and reflections upon them in several early manuscripts, *Biblische Betrachtungen (Scriptural Meditations), Brocken (Fragments), Gedanken ueber meinen Lebenslauf (Thoughts on My Life's Path)*, and *Gedanken ueber Kirchenliedern (Thoughts on Hymns)*, which were all composed from March to May, 1758. Although it was deeply personal, Hamann's vision had its roots in a commonplace of the pietist tradition: that all the events portrayed in the Bible are a spiritual allegory, a metaphor for the soul's struggle for salvation (see Pietism).

Hamann's reaction to the *Aufklärung* began with his defence of his new faith, which he set forth in his short tract *Sokratische Denkwürdigkeiten (Socratic Memorabilia)* (1759). This work was dedicated to Christoph Berens, Hamann's employer, and Immanuel Kant, who attempted to reconvert him to the cause of the *Aufklärung*. Its central thesis is that reason does not have the power to criticize faith. The jurisdiction of reason, Hamann argues, is limited to the assessment of the truth-value of propositions. Its task is to determine whether we have sufficient evidence for some statement. Faith therefore falls outside the sphere of reason, because faith consists in a living experience, which cannot be expressed in any proposition or statement. Just as reason cannot criticize our ordinary sensations, whose qualities are just given and indescribable, so it cannot detract from a mystical experience, whose qualities are no less present and ineffable.

In his defence of faith Hamann appealed to the scepticism of David Hume. According to Hamann, the Scottish philosopher had shown that reason cannot demonstrate or refute the existence of anything, and that we need faith to sustain us even in our ordinary life. Contrary to Hume's intentions, though, Hamann used his scepticism to *defend* religious belief itself. If reason cannot prove or disprove the existence of ordinary things, then *a fortiori* it cannot prove or disprove the existence of God. If we need faith in ordinary life, then why cannot we have it in religion too? In general, Hamann proved to be an important transmitter of Hume's ideas in Germany. His July 1759 letter to Kant refers to Hume, providing Kant with his first knowledge of the philosopher who would later awaken him from his 'dogmatic slumber'.

In many of his later writings - especially his *Kreuzzüge des Philologen (Crusades of the Philologist)* (1762) *Philologische Einfälle und Zweifel (Philological Whimsys and Doubt)* (1772) and *Metakritik ueber den Purismus der Vernunft (Metacritique of the Purism of Reason)* (1781) - Hamann deepened and broadened his attack upon the faith in reason of the *Aufklärung*. The chief target of Hamann's critique was the tendency of the *Aufklärung* to abstract reason from its social, historical and linguistic context. The *Aufklärer* treated reason as if it were a selfsufficient, autonomous faculty, operating independently of political interests, cultural traditions or subconscious desires. If, however, we are to avoid hypostasis, then we must raise the questions 'Where is reason?', 'In what particular things does it exist?' We can answer these questions, Hamann argues, only by identifying the embodiment or manifestation of reason in language, custom and action. Reason then proves to be not a special kind of faculty existing in some Platonic or noumenal realm but only a specific manner of speaking, writing and acting in concrete cultural circumstances. Accordingly, Hamann stressed the social and cultural dimension of reason, which had been much neglected in the eighteenth century. In this regard his teaching was influential upon Herder and anticipates the historicism of the nineteenth century.

When Kant's *Critique of Pure Reason* appeared in 1781, Hamann attacked it as a perfect example of the *Aufklärung*'s hypostasis of reason. The main fallacy of the *Critique*, as he described it, is 'the purism of pure reason', the abstraction of reason from its embodiment in language, tradition and experience. According to Hamann, Kant hypostasized reason by postulating a selfsufficient noumenal realm that exists apart from the phenomenal realms of language, history and experience. Kant commits a threefold hypostasis or purification of reason. He abstracts it from sense perception, tradition or custom, and - worst of all - language. What remains after all this abstraction is nothing but the purely formal transcendental subject =X, 'the talisman and rosary of a transcendental superstition which believes in all *ens rationis*'. In general, Hamann deplored Kant's dualisms in all their forms -noumena/phenomena, understanding/sensibility, concept/intuition - since they were nothing more than arbitrary and artificial abstractions. The central task of the philosophy of mind, in his view, is not to divide but to unite the various powers, showing how they stem from a single source. This single source, Hamann suggested, lay in words, the embodiment of thought in sensible marks and sounds. The central task for the critique of reason therefore lay in the examination and criticism of language itself.

Although Hamann's critique of reason made him a rebel against the prevailing ideology of the *Aufklärung*, it was not entirely new or original. To say that Hamann is without any known debt to anyone else is to ignore his broader intellectual context. Hamann's critique of reason should be placed firmly within the Protestant tradition, which had always insisted upon a sharp distinction between realms of faith and reason. It was Hamann's aim to defend this distinction, the integrity of the sphere of faith, against the incursions of the *Aufklärung*. His critique of hypostasis, his emphasis upon concrete experience, his insistence upon the linguistic embodiment of reason, and his claim that language consists only in convention and custom, all show his debt to the nominalism so characteristic of the Protestant tradition. It is indeed important that Hamann saw himself as the defender of Luther, whose reputation had waned at the close of the eighteenth century. Although Hamann's critique of reason moves within the broad contours of the Protestant tradition, it would be a mistake to regard him as an orthodox Lutheran or typical pietist. In the famous dispute between Lessing and Goeze, Hamann refused to side with the orthodox pastor, who had defended the literal truth of the Bible. Hamann's ultimate authority was mystical experience more than the Bible, which required a more metaphoric and spiritual interpretation. It is more accurate to place Hamann within the spiritualist tradition of Protestantism, among such radical reformers as Gottfried Arnold, Valentin Weigel, Jakob Boehme, Hans Denck and Sebastian Franck. Like Hamann, these thinkers stressed the importance of personal freedom and mystical experience.

It would be a mistake, too, to describe Hamann's critique of reason as a form of 'irrationalism', as if he were

strongly prejudiced against reason and the sciences. Hamann frequently insisted that his critique was directed against only the *illegitimate extension* of reason beyond experience, and that reason had a perfectly valid use within the boundaries of ordinary experience. If we use the term 'irrationalism' strictly and accurately, then it applies to the acceptance of beliefs that are contrary to evidence, the attitude typified by Tertullian's famous maxim *credo quia absurdam* (I believe because it is absurd). But Hamann rejected any such 'leap of faith'. In this regard he was only keeping with the traditional Lutheran distinction between the heavenly and earthly realms, where the supernatural heavenly realm was the object of faith, the empirical earthly realm the domain of reason.

2 Philosophy of language

The most important aspect of Hamann's critique of reason was his philosophy of language, which he first sketched in his *Kreuzzüge* and then developed in many writings throughout his life. Hamann is one of the first thinkers in the modern tradition to stress the importance of language for thought. He denies the prevalent eighteenth-century view - common to both the empiricist and rationalist traditions - that there are clear and distinct ideas apart from their embodiment in language. Words are not just arbitrary signs for already formed ideas but the very medium by which ideas come into existence. Thinking is nothing more than the use of symbols. To talk about ideas or concepts apart from words is again to lapse into that old fallacy of the *Aufklärung*: hypostasis.

Hamann takes a very radical stand on the connection between ideas and signs, thoughts and words. It is not only that signs are the *medium of existence* of reason; they are also the *criterion of its truth*. True to the nominalism of the Protestant tradition, Hamann denies that a proposition is true because it corresponds to some special kind of entity, such as facts, universals or states of affairs. All that exists are particulars, which can be explained and classified in all kinds of ways, depending upon our interests. The main criterion for truth is only the correct use of language, which is determined by nothing more than convention and tradition. In making language the criterion of reason and use the sanction of language, Hamann flirted with relativism, though he never explicitly avowed it.

In 1772 Hamann became involved in the famous eighteenth-century controversy regarding the origin of language (see Herder, J. §3). This debate, launched by Condillac in 1746, concerned whether language had a merely human or a divine origin. Hamann criticized his erstwhile pupil, Herder, for his view that the origin of language could be explained purely naturalistically, as a product of human need and skill. Although Hamann is generally regarded as a defender of the divine origin view, his theory is much more complex and obscure. On at least one reading, Hamann does not deny that the origin of language can be explained naturalistically. Rather, he holds that language has *both* a divine and human, a supernatural *and* natural, source. Although language is created through natural means, by the use of native human powers, God is also coactive in the use of these powers. Since God acts through man, what man creates through his natural capacities is also what God creates through him. Sometimes, however, Hamann resorts to mystical and metaphorical expressions, as if God directly taught Adam how to use language after the creation.

3 Aesthetics

Hamann's historical significance lies in his aesthetics no less than his critique of reason. His main work in aesthetics is his *Aesthetica in nuce*, which appeared as part of his *Kreuzzüge des Philologen*. Its defence of artistic creativity and the metaphysical significance of art had a great influence upon the *Sturm und Drang* and ultimately Romanticism itself (see Romanticism, German).

One of the main aims of Hamann's text is to liberate the artist from the shackles of conventions and norms. *Aesthetica in nuce* is a manifesto on behalf of artistic creativity, a paean to the genius who dares to break all the rules. Rather than aspiring to portray archetypes or models, the artist should dare to express his passions and reveal his personal vision. What makes a work of art beautiful, Hamann suggests, is its expression, the revelation of the personality of the artist.

Another objective of the *Aesthetica* is to reaffirm the metaphysical significance of art, the classical equation of truth and beauty. This significance had been denied or depreciated by most eighteenth-century aestheticians, who usually regarded aesthetic experience as little more than a pleasant sensation or an amusing illusion. According to Hamann, however, art is nothing less than the purest medium of truth. Our knowledge of life and reality comes through immediate experience, and only a non-discursive medium such as art reproduces such sensations and

feelings. While philosophers must resign themselves to concepts, which are only artificial and arbitrary abstractions, artists deal with images, which capture all the vividness and richness of experience.

In defending the equation of truth and beauty, Hamann appears to return to the classical aesthetics of the seventeenth century. Indeed, he even retains the doctrine of imitation, so important for that tradition. But Hamann retained this doctrine in his own manner and for his own purposes. He reformulated it in accord with his empiricist and mystical epistemology, and so at odds with the more rationalist epistemology of classicism. He broke with classicism with regard to both the object and manner of imitation. What the artist should imitate, Hamann thinks, is the secret language of God, not the eternal archetypes of things, as in classicism; and he imitates this language not by following classical rules or norms but by creating images and symbols.

The *Aesthetica* appears to consist in two conflicting doctrines: an extreme subjectivism, which encourages artists to express their personal feelings and visions; and an extreme objectivism, which demands that they imitate nature and reveal the presence of God. What is central to, and characteristic of, Hamann's aesthetics, however, is precisely the synthesis of these doctrines. They come together in Hamann's mystical vision of the omnipresence of God. Since God is inside man, dwelling in man's inner heart, artists have only to reveal their innermost feelings to reveal God himself. Their personal symbols and images then become nothing less than the language of God.

4 Politics

Although not primarily a political thinker, Hamann made some important criticisms of the political doctrines of the *Aufklärung*. In *Golgotha und Scheblimini* (1784), a critique of Moses Mendelssohn's *Jerusalem*, he criticized the prevalent natural law doctrine (see Mendelssohn, M.). Since reason becomes determinate only in a specific cultural context, it is a false abstraction, Hamann argues, to seek some universal and eternal norm true for any culture. Natural law doctrine also assumes that people are rational outside society; yet it is only through society that they learn the use of language, restrain their appetites and conduct themselves according to rules.

In *Kreuzzüge des Philologen* Hamann attacked the individualism prevalent in the modern political tradition. He explicitly appealed to Aristotle's conception of man as a political animal to refute the notion that the individual is selfsufficient and born with natural needs. Freedom and reason are not properties inherent in each individual apart from society, Hamann contended, but only their manner of acting and speaking within it.

Because of his critique of the *Aufklärung* and the liberal values associated with it, Hamann has sometimes been portrayed as a reactionary, indeed as a founding father of the modern alliance of irrationalism and conservatism. But this interpretation is anachronistic, failing to consider the context in which Hamann lived and wrote. All his mature life Hamann was a passionate critic of the reigning Prussian monarch, Friedrich II, 'the philosopher king'. But his critique of Friedrich and the *Aufklärer* in Berlin is not a rejection of their liberal principles - toleration, freedom of press, equality before the law - but of the paternalism and authoritarianism of the absolutist Prussian state. Hamann too embraced these principles, and his critique of the *Aufklärung* is indeed motivated by them. He argued, however, that the *Aufklärer* had betrayed them through their intolerance toward revealed religion and through their alliance with the autocratic Prussian monarch. In his critical review of Kant's essay on enlightenment, Hamann attacked Kant not because he advocated emancipation - the right to think for oneself - but because he restricted it to the public sphere. By denying the people a right to think for themselves in their official duties, Kant had virtually sanctioned Friedrich's despotism, the old maxim 'Say what you like but obey'.

See also: Enlightenment, Continental; Faith; Jacobi, F.; Schlegel, F. von.

FREDERICK BEISER

List of works

Nadler, J. (ed.) (1949-57) *Sämtliche Werke*, Vienna: Herder, 6 vols.Includes the following:
Hamann, J.G. (1759) *Sokratische Denkwürdigkeiten (Socratic Memorabilia)*.
Hamann, J.G. (1762) *Briefwechsel*, Ziesemer, W. and Henkel, A. (eds) (1955-79), Wiesbaden and Frankfurt: Insel, 6 vols.
Hamann, J.G. *Hamann's Socratic Memorabilia: A Translation and Commentary*, trans. O'Flaherty, J.C. (1967), Baltimore, MD: Johns Hopkins University Press.
Hamann, J.G. *Golgotha and Scheblimini*, trans. Dunning, S.N. (1979), in *The Tongues of Men: Hegel and Hamann on Religious Language and History*, Missoula, MT: Scholars Press.

Hamann, J.G. (1985) *Aesthetica in nuce ‘Aesthetics in a Nutshell’: A Rhapsody in Cabbalistic Prose*, trans. Crick, J.P., in H.B. Nisbet (ed.) *German Aesthetic and Literary Criticism: Winckelmann, Lessing, Hamann, Herder, Schiller, Goethe*, Cambridge: Cambridge University Press.

References and further reading

Alexander, W.M. (1966) *Johann Georg Hamann: Philosophy and Faith*, The Hague: Martinus Nijhoff.
Berlin, I. (1993) *The Magus of the North: J.G. Hamann and the Origins of Modern Irrationalism*, London: John Murray.
German, T. (1981) *Hamann on Language and Religion*, Oxford: Oxford University Press.
O’Flaherty, J.C. (1952) *Johann Georg Hamann*, Chapel Hill, NC: University of North Carolina.
Smith, R.G. (1960) *J.G. Hamann 1730-1788 A Study in Christian Existence, with Selections from his Writings*, London: Collins.

Hamilton, William (1788-1856)

Sir William Hamilton was a leading exponent of the Scottish philosophy of 'common sense'. This philosophy had its origin in the works ofThomas Reid, but it was throughHamilton that it achieved its most subtle form and exerted its greatest influence.

'Common-sense' philosophy, on a superficial view, may seem to hold that philosophical problems should be settled by appealing to the commonly accepted opinions of ordinary people. But that is not *what it holds. The 'common sense' to which it refers are certain powers and beliefs natural to the mind and therefore common alike to the learned and vulgar.Hamilton holds that these powers and beliefs can neither be doubted nor justified. They carry their own authority. This view derives its significance from a point which has often been overlooked. When we doubt or justify a belief, we stand outside that belief and compare it with the world. But the power to compare a belief with the world itself presupposes beliefs about the world. We cannot step outside all our beliefs. That is why, according to Hamilton, certain powers and beliefs must carry authority.*

Hamilton was educated in Glasgow and Oxford. After graduation he entered the legal profession, passing as advocate at the Scottish Bar in 1813. His chief interest, however, was in philosophy. Between 1829 and 1836 he published in the *Edinburgh Review* a series of articles which established his reputation, both in Britain and on the Continent. In 1826 he was appointed to a chair in philosophy at the University of Edinburgh, which he held until his death.

Hamilton wrote very little and was not a gifted writer. In this respect, he suffers in comparison with his predecessor Thomas Reid and his associateHenry Mansel, both of whom were masters of English prose. But he was a deep thinker and a powerful teacher, greatly respected by his pupils. In his time, he exerted an influence on philosophy which extended throughout the English-speaking countries. In the colleges of America, for example, his was the predominant philosophy throughout the greater part of the nineteenth century.

His influence waned after the publication ofJ.S. Mill's *Examination of Sir William Hamilton's Philosophy* (1865). Mill tookHamilton to be the leading opponent of his own empiricism and therefore subjected his views to a severe criticism, which at the time was generally thought to be successful (see Mill, J.S.).

In logic,Hamilton is best known for his view that in the proposition 'All *A* is *B*', one may distinguish between 'All *A* is all *B*' and 'All *A* is some *B*'. This doctrine - known as the quantification of the predicate - served to reduce propositions to equations, thereby preparing the way for mathematical logic. Thus if the letters *X* and *Y* stand for classes, and are regarded as universally quantified, then for 'All *X*s are all *Y*s' one may substitute '$X = Y$'.

Hamilton's more general philosophy may be taken in two parts. The first is a reworking of views he shared with Reid; the second a development of these into a subtle form of realism. The views he shared with Reid may be contrasted with idealism on the one hand and hypothetical realism on the other. The idealist criticism of realism is that it leads to scepticism. The realist holds that the world exists independently of our beliefs. How then can we ever be sure that our beliefs correspond to that world? The hypothetical realist argues that our beliefs may be justified by an argument to the best explanation: that there is an independent world corresponding to our beliefs is the best explanation for our having them. Both these views, however, have an assumption in common. The assumption is that we are not entitled to our beliefs unless they can be supported by reason. It is precisely this assumption, according to Hamilton, which leads to scepticism. If every belief has to be justified, every justification has itself to be justified. The regress is vicious and scepticism inevitable. Against this,Hamilton argued (anticipating the later Wittgenstein) that belief is primary: 'belief is the primary condition of reason and not reason the ultimate ground of belief'.

Hamilton's development of realism is made distinctive by his insistence that knowledge is relative. The essence of realism is that objects exist independently of being known. It follows, on the realist view, that they can be known only in so far as they enter into relations with the mind. But the *existence* of an object does not depend on the relations by which it is known. It is only knowledge which is thus relative, not the existence of the object itself. For example, unless an object reflects light, it cannot be seen; but it exists, whether or not it reflects light. Consequently we have no reason to suppose that what we know of the object, through its reflecting light, is

exhaustive of the object itself. Moreover it is not simply our knowledge of the world which is thus relative; our conception of the world, so far as it is positive, must also be relative. For we have no such conception which does not depend on our knowledge. It follows that we have no positive conception of the world in its ultimate nature, but only of the world as it is in its relations to the mind. Moreover, Hamilton believed that this will always be so; the idea that we may obtain an absolute conception of the world is an aspiration without content and it forms no part of a coherent realism (see Common Sense School).

Modern readers who considerMill's criticism of Hamilton may be inclined to suppose that its success was due not so much to its validity as to the prestige ofMill himself. They may note also that empiricism is no longer as dominant as it was in Mill's day and that there has been a revival of interest in views, such as those of Wittgenstein or the pragmatists, which are clearly related to Hamilton's.

See also: Commonsensism; Logic in the 19 century; Realism and antirealism

H.O. MOUNCE

List of works

Hamilton, W. (1853) *Discussions on Philosophy and Literature, Education and University Reform*, London: Longman.(Hamilton's major work.)

Hamilton, W. (1858-60) *Lectures on Metaphysics and Logic*, Edinburgh: Blackwood, 4 vols. (Lectures to his students; they do not contain his fully developed views.)

References and further reading

Grave, S.A. (1960) *The Scottish Philosophy of Common Sense*, Oxford: Clarendon Press.(A modern survey of the Scottish school.)

Mansel, H. (1866) *The Philosophy of the Conditioned*, London and Edinburgh: Strachan.(A defence ofHamilton by his most famous associate.)

Mill, J.S. (1865) *An Examination of Sir William Hamilton's Philosophy*, London: Longman and Green, 2 vols. (The most famous criticism of Hamilton.)

Veitch, J. (1888) *Hamilton*, Edinburgh and London: Blackwood.(Contains a defence of Hamilton against Mill.)

Han Feizi (*c.*280-233 BC)

Han Feizi was the pre-eminent Legalist philosopher. The work attributed to him, the Han Feizi, is his conscious response to the general breakdown of civil order and the interminable inter-state struggle for survival and conquest during the Warring States period (463-222 BC). However, Han Feizi's work transcends the particular circumstances that gave rise to it and addresses perennial philosophical issues that continue to remain relevant.

In the *Han Feizi*, the works 'Gufen' (Solitary Indignation), 'Wudu' (Five Vermin) and 'Shuonan' (Difficulties of the Way of Persuasion) constitute Han Feizi's response to the political pathology of his time. The central and compelling concern was political survival in the vortex of transformation precipitated by the systemic disintegration of what was left of the Zhou feudal order. This overriding political reality in large measure explains the way Han Feizi identified the outstanding political problems of his time: (1) 'learned celebrities' deluding the ruler with irresponsible eloquence so as to undermine the law; (2) itinerant speakers mouthing deceptive theories to serve their own interests; (3) private swordsmen violating the interdicts of government; (4) well-connected courtiers engaging in all manner of corruption; and (5) nonproductive tradesmen and craftsmen collecting useless luxuries and exploiting peasants. These are the 'Five Vermin' that undermine and ruin the state.

Han Feizi identified the root causes underlying these problems and called for a new prescriptive model of sociopolitical organization to remedy them. First, he called for the establishment of objective and impartial standards of human conduct, impersonal standards patterned after *dao* (the Way the natural world works) and free of subjective preferences (see Dao). The term *fa* means both prescriptive standards and penal law; *fa* is designed to remould human behaviour so that individual interest will 'naturally' dovetail with public interest (see Fa). Moreover, it is *fa* alone, not moral suasion, that will serve as the guiding principle of sociopolitical behaviour. Second, authoritative power (*shi*) is 'well-ordered' when it is securely based on clearly established and objective *fa*. Thus, *shi* is the impersonal, institutionalized position of the rulership, rather than a charismatic power to inspire awe and obedience among the masses. Only when *shi* is 'well-ordered' will penal law function properly. Furthermore, *shi* cannot govern effectively without the organizational power of bureaucracy under the centralized control of the ruler. Thus, Han Feizi advocated a rational paradigm of bureaucratic accountability and merit-based functional specificity. The ruler needs to master the 'technique' (*shu*) of controlling bureaucracy by comparing 'word' (proposals) and actual 'performance'. If they tally, reward is in order; if not, then punishment is required. The ruler holds bureaucracy accountable not only by means of empirical verification but also by means of 'The Two Handles' of power over life and death.

Han Feizi elaborated on and synthesized the three interdependent pillars of the Legalist School: the *fa* of Shang Yang (390-338 BC), the *shi* of Shen Dao (*fl.* 310 BC) and the *shu* of Shen Buhai (d. 337 BC). In doing so, he produced a coherent and theoretically sophisticated Legalist philosophy.

See also: Legalist philosophy, Chinese; Fa

LEO S. CHANG

List of works

Han Feizi (*c.*280-233 BC) *Han Feizi*, ed. Wang Xianshen, *Han Feizi jijie (The Complete Works of Han Feizi with Collected Commentaries)*, 1896; ed. Chen Chiyu, *Han Feizi jishi (Collected Explanations)*, Beijing: Zhonghua Book Co., 1958; ed. Chen Chitian, *Zengding Han Feizi jiaoshi (Further Revisions of Collations and Explanations of Han Feizi)*, Taipei: Taiwan Commercial Press, 1972; trans. W.K. Liao, *The Complete Works of Han Fei Tzu*, London: Arthur Probsthian, vol. 1, 1938; vol. 2, 1959.(The work was in all probability written by Han Feizi himself with the exceptions of 'Interview with the King of Ch'in: A Memorial' and 'Having Regulations'. Jong Tsau-chu's *Textual Criticisms of Han Fei Tzu's Works* (Shanghai: Shanghai Commercial Press, 1936) notwithstanding, there is no convincing evidence that Han Feizi did not write 'Interpreting Laozi', 'Illustrating Laozi' and other writings with Daoist contents: see Appendix I in H.P. Wang and L. Chang, *The Philosophical Foundations of Han Fei's Political Theory*, Honolulu, HI: University of Hawaii Press, 1986, 87-109. Of the editions given here, Wang (1896), Chen (1958) and Chen (1972) are extensively annotated. Liao remains the only complete translation of the work in English, although twelve chapters are translated by B. Watson in *Han Fei Tzu: Basic Writings*, New York: Columbia University Press, 1964.)

References and further reading

de Bary, W.T. (1960) *Sources of Chinese Tradition*, New York: Columbia University Press, vol. 1, 122-49.(A brief introduction to Han Feizi and Li Si for students.)

Fung Yu-lan [Feng Youlan] (1952) *A History of Chinese Philosophy*, trans. D. Bodde, Princeton, NJ: Princeton University Press, vol. 1, ch. 13.(A clear exposition of Han Feizi and Legalist philosophy with many excerpts from the original texts; for upper level undergraduates.)

Graham, A.C. (1989) *Disputers of Tao*, La Salle, IL: Open Court.(Erudite and evenhanded treatment of Legalist philosophy for advanced students.)

Schwartz, B.I. (1985) *The World of Thought in Ancient China: Philosophical Argument in Ancient China*, Cambridge, MA: Harvard University Press.(Sophisticated and insightful exposition of Legalist philosophy that transcends the stereotypical views.)

Wang, H.P. (1991) *Hsien Ch'in fa-chia ssu-hsiang shih lun (Discourses on the Pre-Qin Legalist Thought)*, Taipei: Linking (Lien Ching) Publishing Co.(The most comprehensive and clear account of Legalist philosophy; see especially pages 267- 319.)

Wang, H. P. and Chang, L. (1986) *The Philosophical Foundations of Han Fei's Political Theory*, Honolulu, HI: University of Hawaii Press.(For graduate students.)

Han Wônjin (1682-1751)

Han Wônjin was a major thinker of the Korean neo-Confucian tradition. One of the leading scholars of his time, he is especially remembered as a protagonist in the Horak controversy, which he ignited with the observation that ki *(in Chinese,* qi*) or 'material force' is present even when the mind is in a meditative, quiescent state.*

Neo-Confucian metaphysics is a dualistic monism in which the single *dao* or pattern running through all things is concretized and individualized in real beings only through *ki* (see Qi). The patterning element, known as *i* and often translated as 'principle', also constitutes the innermost nature of all things; its presence as the substance of the human heart-and-mind accounts for our innate potential to respond appropriately to all things and situations (see Xin).

Han's observation that *ki* is always present is consistent with a metaphysics where *i* and *ki* are strictly interdependent and complimentary in function: any real existent always involves both. But in addition to individuation, imperfection or turbidity in *ki* also is a distorting element that accounts for our often imperfect response to things. A cornerstone of neo-Confucian self-cultivation was quiet-sitting, a meditative technique of mental quiescence understood as putting oneself in union with the perfection of *i*, the inner nature and substance of the heart-and-mind. By asserting that even in quiescence the imperfection of *ki* could not be considered absent Han aroused a storm of controversy. His monistic emphasis on absolute interdependence was metaphysically sound, but was problematic in the context of an ascetical system which was naturally more dualistic in emphasis in order to account for perfection and imperfection.

As the controversy unfolded it soon involved issues related to understanding the unitary *i* of the universe as the nature of individual things. His opponent upheld the more conventional bifurcation: considered in itself, *i* is unitary, but considered as concretized and differentiated by *ki* it is the nature of single, individual beings. Han fastened onto the need for a differentiated normative nature: the norm for a cow is not the same as the norm for a human. The problem is that the normative nature conventionally is *i* in itself, unlimited and undistorted by imperfect *ki* - and hence evidently undifferentiated as well. Han introduced a threefold schema: between the unitary nature and the completely individuated one he proposed a unique level, '*i* as based on *ki*'. On this level, *i* would be differentiated as the nature of various species, but not concretely limited to imperfect individuals.

Han's novel doctrine was roundly criticized by opponents as being unintelligible, for the new level seemed neither to consider *i* in itself nor actually join it with *ki*. The controversy crystallized as the question whether the normative 'original' nature was the same or different for all species, and continued for generations without resolution (see Xing). It calls attention to a deep seam in neo-Confucian thought, where the metaphysics of *i* and *ki* synthesize and mask quite different meanings of 'nature' from the Daoist and Buddhist traditions.

See also: Confucian philosophy, Korean; Neo-Confucian philosophy; Yi Kan

MICHAEL C. KALTON

List of works

Han Wônjin (1682-1751) *Namdang Chip (The Collected Writings of Han Wonjin)*, Seoul: Ch'ae Insik, 1976. (Photo reprint.)

References and further readings

There is at present no further information on Han in Western languages.

Han Yu (768-824)

Among the most important figures in the history of medieval Confucianism, Han Yu helped to redefine and adapt earlier Confucian teachings to the needs of his contemporary society. He strove to make Confucianism a more popular and comprehensive doctrine by devising a basic metaphysics and epistemology to complement the earlier Confucian emphasis on ethics.

Han Yu was a government official, writer, and scholar of the Tang dynasty (618-907) who played a central role in the medieval redefinition of the Confucian tradition. He was the major figure in the transition between the first state-supported formulation of Confucian ideology in the Han dynasty (206 BC-AD 220) and the later transformation of that tradition, known today as neo-Confucianism, in the Song dynasty (960-1279).

The centrepiece of this transformation was a new concept of the Confucian Sage. Whereas early Confucian teachings stressed the ethical dimension of the Sage, Han Yu and his Song dynasty successors worked to broaden this earlier concept of the Sage to include a metaphysical and epistemological dimension. According to most understandings of Chinese intellectual history, their effort resulted from the challenge and popularity of Buddhist and Daoist teachings in the Six Dynasties and early Tang dynasty (see Buddhist philosophy, Chinese; Daoist philosophy).

For Han Yu, the Sage was a person who had achieved a perfect integration in his own daily life between theory and action. In this process, the Sage becomes one with a metaphysical Absolute. This unity is described as *cheng*, translated as ‘integrity, being true to oneself’ (see Cheng). In forging this union, Han Yu sought to discover new epistemological values in the writings of pre-Han Confucian antiquity. In his writings, the phrase *gudao*, or ‘way of antiquity’, assumes an almost philosophical dimension as the source of positive human values.

To extract these values from the text of the Confucian canon, Han Yu espoused an interpretative method known as ‘overall meaning’, which rejected established commentary in favour of an intuitive understanding by direct appeal to the reader’s ‘mind’. The process attempted to make the old texts relevant to contemporary issues and resulted in a concentration on sections of the old canon that best lent themselves to such reinterpretation. One may see in the philosophical writings of Han Yu a concern with those texts that were later to become known as the Four Books: the *Analects, Daxue (Great Learning), Zhongyong (Doctrine of the Mean)* and *Mengzi* (see Confucius; Daxue; Zhongyong; Mencius).

Han Yu also constructed for his rejuvenated Confucianism a lineage or progression of orthodox teachers from antiquity down to his own time. The contemporary Chan Buddhist system of master-disciple lineages probably stimulated Han Yu’s thinking in this area, and the formation of such lineages became an important concern of later neo-Confucianism. Han Yu was known both in his own day and in later times primarily as a man of letters and creator of *guwen*, the ‘literature of antiquity’, a style of writing that was intended as a literary reflection of the ‘way of antiquity’.

See also: Cheng; Confucian philosophy, Chinese; Neo-Confucian philosophy

CHARLES HARTMAN

List of works

Han Yu (768-824) *Han Changli shi xinian jish (The Poetry of Han Yu Arranged Chronologically with Collected Annotations)*, ed. Qian Zhonglian, Shanghai: Gudian wenxue chubanshe, 1957.(Standard modern edition of Han Yu’s poetry.)

Han Yu (768-824) *Han Changli wen ji jiaozhu (The Prose Writings of Han Yu, Collated and Annotated)*, ed. Ma Chichang, Shanghai: Gudian wenxue chubanshe, 1957; partial trans. Liu Shih Shun, *Chinese Classical Prose: The Eight Masters of the T’ang-Sung Period*, Hong Kong: The Chinese University Press, 1979.(The Ma edition is the standard modern edition of Han Yu’s prose writings; Liu has English translations of Han Yu prose texts. For translations of key Han Yu philosophical texts, see W.T. de Bary (ed.) *Sources of Chinese Tradition*, New York: Columbia University Press, 1964, 426-37.)

References and further reading

Hartman, C. (1986) *Han Yü and the T'ang Search for Unity*, Princeton, NJ: Princeton University Press.(The only attempt at a comprehensive study of Han Yu in a Western language; Chapter 3 is devoted to Han Yu's philosophical thought.)

Hanslick, Eduard (1825-1904)

Eduard Hanslick, a music critic for the popular Viennese press, is principally known as the author of Vom Musikalisch-Schönen (1854). This is probably the most widely read work in the aesthetics of music for both philosophers and musicians, and remains the starting point for any discussion either of the place of emotion in music, or of the doctrine usually referred to as 'musical purism'. On the former, Hanslick maintained what he calls the negative thesis, which 'first and foremost opposes the widespread view that music is supposed to represent the feelings'; on purism, he proposed the positive thesis or antithesis, 'that the beauty of a piece of music is specifically musical, that is, is inherent in the tonal relationships without reference to an extraneous, extra-musical context'.

1 Major arguments

What was Hanslick denying? As becomes clear in the progress of his argument, the negative thesis denies that it is any part of the office of music, *qua* music, to arouse or represent 'feelings' of the kind I shall call 'ordinary', by which I mean feelings or emotions such as love, fear, hatred, anger and the like. Two further possibilities remain open: that music might arouse some other kind of feeling as part of its essential nature, and that it might arouse the ordinary emotions, but in an aesthetically irrelevant context. As we shall presently see, Hanslick affirmed both.

Hanslick presents three basic arguments in support of the negative thesis, all of which take the following general form. Music demonstrably does not (or cannot) arouse or represent the ordinary emotions in any aesthetically relevant way; therefore it cannot be any part of the aesthetic purpose of music to do so.

The first of these arguments, and certainly the most subtle, is this: 'The representation of a specific feeling or emotional state is not at all among the characteristic powers of music' because such feelings or emotional states 'depend upon ideas, judgments, and (in brief) the whole range of intelligible and rational thought to which some people readily oppose feeling' (1854 (1986): 9). In other words, emotions cannot either exist or be 'represented' in the absence of what contemporary philosophers, following Franz Brentano (§3), call 'intentional objects'. And music cannot in any way make those intentional objects available to the listener. Particularly in the latter form, the argument seems generally persuasive, although a possible response to each form of the argument might at least serve to qualify it. For it might be argued both that there are other ways that music can make up for lack of intentional objects in representing the emotions, and at least peripheral ways in which some emotions are perhaps aroused without them.

The second argument, which might be called the 'argument from disagreement', is directed specifically against the possibility that music can represent the ordinary emotions (although, if good, it would apply with equal force to their arousal). It is simply that, according to Hanslick, since there is no consensus at all in any individual case about which emotion a passage of music may represent, it cannot make sense to say that it can represent any at all. But anyone who has ever played passages with distinct emotive character to a typical listener or indeed has read any critical literature (including Hanslick's), will find Hanslick's premise of chaotic disagreement utterly false, and emotive consensus, at least within reasonable limits, fairly widespread. Nor, given the historical continuity of our musical materials, could this consensus come as a surprise to anyone who had not begged the question against it from the outset.

Hanslick's final argument is based on the growing awareness in the nineteenth century that composers of earlier times recycled their music, and in particular that they sometimes used the same music for different texts. It charges that if the same music can seem emotively appropriate to two texts of very different expression, one cannot say that the music has any particular emotive character at all. However, this is clearly a case of drawing an extreme and false conclusion from an undoubtedly true premise. And Hanslick's prime example, Handel's well-known use of music from his erotic love duets for sacred choruses in *Messiah*, simply shows that music expressive of melancholy or joy can seem appropriate both to the joy and melancholy of erotic love and to the joy and melancholy of the Christian religion, which is by no means surprising, and hardly proof that the music is not expressive both of joy and melancholy.

Of the positive thesis, little need (or can) be said here because, apart from some suggestive and, as it turned out,

fruitful metaphors, Hanslick was able to give his concept of musical beauty little substantive content. The metaphor which is most remembered, and in the event proved most fruitful, is: ‘The content of music is tonally moving forms’ (1854 (1986): 29). Hanslick combined this recognition of absolute music (that is, instrumental music) as a formal structure in apparent motion with the equally important ‘intuition’ that the formal structure is not merely a kaleidoscopic pattern (another of his metaphors) but a quasi-syntactical one, bearing, in that respect, an analogy to language. Beyond such intriguing, though fragmentary, remarks Hanslick was not able to put forward a systematic musical purism.

2 Music and emotion

The question now remains of what role, if any, human feeling plays in Hanslick’s musical philosophy. As regards the ordinary emotions, Hanslick does not deny them a role, but explicitly affirms that ’music can nevertheless excite such feelings as melancholy, gaiety and the like (can, not must) (1854 (1986): 9-10). How is this consistent with the negative thesis? It is consistent because what the negative thesis asserts is that it is no part of the *aesthetic* significance of music to arouse the ordinary emotions; it indeed *cannot* arouse them in an aesthetically meaningful way. What it can do is arouse such emotions in those who, by virtue of particular personal associations or a ’pathological’ nervous condition, may be abnormally susceptible.

Are we to conclude, then, that Hanslick views a proper experience of music as completely ’cold’ and ’unemotional’? Such a view would be hard to credit, but there is no need to impute it to him, for nothing he says implies it, and one intriguing passage, in the Foreword to the Eighth Edition (1891), suggests just the right response to the question. Hanslick writes: ’I share completely the view that the ultimate worth of the beautiful is always based on the immediate manifestation of feeling’ (1854 (1986): 22). Can we not read this as an assertion of the undoubted truth that the beautiful always deeply moves us upon its recognition? And what is true of beauty *sans phrase* must also be true of musical beauty. Beautiful music can no more leave those who perceive it unmoved, unfeeling and emotionally untouched than beauty in any instance.

The weakness in Hanslick’s position lies, then, not in his inability to accommodate an emotionally involved listener; rather, it involves his denial that music can sensibly be characterized in terms of the ordinary emotions. Apparently seeing emotive arousal or emotive representation as the only possible avenues to specific emotive description, and rejecting both for at least one of the right reasons, he was left with the implausible conclusion that calling music ‘sad’ or ‘happy’, ‘angry’ or ‘yearning’, is either figurative or foolish. But other ways in which music might ‘embody’ the ordinary emotions have been explored since Hanslick’s time; and there is a growing consensus that this aspect of his negative thesis is in need of serious revision. Nevertheless, *Vom Musikalisch-Schönen* remains the starting point of the philosophical aesthetics of music and its first modern text, and suffices to keep Eduard Hanslick’s name alive.

See also: Emotion in response to art §4; Emotions, philosophy of; Artistic expression; Formalism in art; Music, aesthetics of

PETER KIVY

List of works

Hanslick, E. (1854) *Vom Musikalisch-Schönen*, Wiesbaden: Breitkopf & Hartel, 1980; trans. G. Cohen, ed. M. Weitz, *The Beautiful in Music*, New York: The Liberal Arts Press, 1957; and trans. G. Payzant, *On the Musically Beautiful*, Indianapolis, IN: Hackett Publishing Company, 1986.(The twentieth edition of the German text, in paperback. The 1957 edition uses the original, but still useful, English translation (1891), with an insightful philosophical introduction by Morris Weitz, while the 1986 translation is new and philosophically superior.)

(1846-99) *Music Criticisms 1846-99*, ed. and trans. H. Pleasants, Baltimore, MD: Penguin, 1950.(A valuable volume allowing a comparison of Hanslick’s theory with practice.)

References and further reading

Budd, M. (1980) ’The Repudiation of Emotion: Hanslick on Music’, *British Journal of Aesthetics* 20: 29-43.(A philosophically astute account of Hanslick’s position.)

Kivy, P. (1990) ’What was Hanslick Denying?’, *Journal of Musicology* 8: 3-18.(A careful philosophical analysis of the negative thesis.)

Wilkinson, R. (1992) 'Art, Emotion and Expression', in O. Hanfling (ed.) *Philosophical Aesthetics, an Introduction*, Oxford: Blackwell, 207-20.(A clear and accurate discussion of Hanslick's arguments.)

Hanson, Norwood Russell (1924-67)

Hanson was a philosopher of science who introduced novel ways of relating logical, historical and linguistic analyses. His best-known book, Patterns of Discovery, stressed the theory-ladeness of observational reports and argued that causality is a feature of inference systems, rather than of nature as such. He pioneered in combining historical and analytic analyses of significant breakthroughs in science. Though he clarified patterns of discovery he never succeeded in the project of developing a logic of discovery, or an account of the inferences leading from problematic situations to novel explanatory hypotheses. A man of many talents, he also made contributions to the history of science, aerodynamics and epistemology.

Hanson was born in west New York, made his undergraduate studies at the University of Chicago, served as a marine fighter pilot in the Second World War, did further studies at both Oxford and Cambridge, where he became the first university lecturer in the philosophy of science. In 1957 he returned to the USA, set up the department of the history and logic of science at Indiana University, and vociferously opposed McCarthyism. In 1962 he accepted a position as philosophy professor at Yale. He died when the beloved Bearcat he was flying crashed into a mountain near Cortland, New York.

Hanson was a baseball player, a Golden Gloves heavyweight boxer, a proficient classical trumpet player, an artist especially adept in illustrating Homeric heroes, a decorated fighter pilot (Air Medal, Distinguished Flying Cross) shot down over Japan, a gadget designer, a daredevil pilot who set speed records for propeller driven airplanes, and a prolific writer in both the philosophy and history of science. The tension and conflict that stimulated him were reflected in his way of doing philosophy. He introduced into logical positivism analytic methods developed by the later Wittgenstein (see Logical positivism §3). Though, with the positivists, he professed a belief that the history of science is irrelevant to the philosophy of science, he pioneered the use of critical historical reconstructions as a testing ground for philosophical accounts of science. His first, and most influential book, *Patterns of Discovery* (1958), treated the issues that were explored in his later writings.

Hanson popularized the term 'theory-laden observation' (see Observation §§3-4), which he intended in a rather Wittgensteinian sense. One cannot *see* something *as* an X unless one already has the concept of an X. Such 'seeing as' depends more on acquired skills than on formal theories. His emphasis played a historically significant role in undercutting the positivist ideal of reporting observations in a theory-neutral language. Hanson's treatment of facts and causes illustrates his fusion of positivism and analysis. A fact is something reported in a that-clause: the fact that…. This requires a classification of propositional types into necessary and contingent, a priori and a posteriori, general and particular. This positivistic technique was supplemented by an analytic insistence that the meaning of terms used to report facts depends on their use. Scientific breakthroughs occur when old presumptions are discarded in favour of new ways of observing and reporting. Hanson supplemented this analysis by detailed historical studies of how Kepler used ellipses as a tool to approximate distorted circular orbits and finally saw that the ellipses fit the data; and of how a protracted muddling through involving Galileo, Descartes and Beeckman, finally led to a correct account of free fall. He explained causality as a feature of our explanations of nature, rather than of nature itself. We introduce causes to explain phenomena. In experiments we try to develop causal chains, not because these chains really describe nature, but because they simplify inferences.

The most novel feature of Hanson's philosophy of science was his protracted attempt to develop a logic of discovery. Detailed analysis of historical examples led to a general pattern of discovery. A curious iconoclast challenges traditional ways of reporting and explaining and comes to see accepted facts as involving problematic relationships. A tentative hypothesis, introduced to explain the newly perceived problem, must meet various criteria: initial plausibility, coherence with accepted background assumptions, logical and empirical constraints on testing hypothesis. A logic of discovery concerns inferring novel hypotheses from problematic situations.

This project encountered two chief difficulties. The first was the Newtonian precedent of rejecting any reliance on hypotheses. This Hanson tackled through his analytic approach to history. After distinguishing different senses of hypotheses he showed that Newton rejected the scholastic type of subjunctive hypothesis and emphatically rejected any Cartesian-style metaphysical hypotheses. However, Newton repeatedly and fruitfully introduced his own type of hypothesis, a categorical proposition that, if true, would both explain phenomena and also lead to

testable consequences. The second difficulty concerned the suitability of the term ’logic’. Here Hanson eventually distinguished between the precise hypothesis tested and inference from a problematic situation to a suitable type of hypothesis. This, he insisted, was a rational process although he could not formulate logical rules governing it.

Hanson championed the Copenhagen interpretation of quantum mechanics (see Quantum mechanics, interpretation of §3). He supplemented his flying activities with studies of the history and physics of aerodynamics. He became proficient in celestial mechanics and used detailed analyses of historical examples to illustrate subtle differences between philosophical ideals of explanation and the types of explanation scientists are willing to accept. He defended atheism on the grounds that there is no good reason for believing in the existence of a personal God, and attacked agnosticism as a timid compromise.

Hanson readily provoked controversy and polarized conferences into his adversaries and his supporters. He never produced a school of disciples nor developed a general system. His talent was for attacking, rather than formulating, systems.

See also: Discovery, logic of; Galilei, G.; Kepler, J.; Scientific method; Theories, scientific §4

EDWARD MacKINNON

List of works

Hanson, N.R. (1958) *Patterns of Discovery*, Cambridge: Cambridge University Press.(His basic work, reprinted in paperback in 1961.)

Hanson, N.R. (1963) *The Concept of the Positron: A Philosophical Analysis*, Cambridge: Cambridge University Press.(A defence and analysis of the Copenhagen interpretation of quantum mechanics and the discovery of the positron.)

Hanson, N.R. (1969) *Perception and Discovery: An Introduction to Scientific Inquiry*, ed. W.C. Humphreys, San Francisco, CA: Freeman, Cooper & Co.(Text constructed from Hanson’s lecture notes.)

Hanson, N.R. (1971) *Observation and Explanation: A Guide to the Philosophy of Science*, New York: Harper & Row.(An introductory work posthumously edited by Stephen Toulmin.)

Hanson, N.R. (1972) *What I Do Not Believe, and Other Essays*, eds S. Toulmin and H. Woolf, Dordrecht: Reidel. (A 2-volume collection of Hanson’s essays on philosophy of science, theology and the history of flight.)

Hanson, N.R. (1973) *Constellations and Conjectures*, ed. W.C. Humphreys, Dordrecht: Reidel.(A copiously illustrated history of planetary theory from Plato to Kepler emphasizing the gap between explanation and prediction.)

References and further reading

Cohen, R.S. and Marx M.W. (eds) (1968) *Boston Studies in the Philosophy of Science*, vol. III, Dordrecht: Kluwer.(Contains tributes to Hanson and a list of his publications.)

Duhem, P. (1906) *La théorie physique. Son objet et sa structure*, Paris: Chevalier et Rivière; 2nd edn, 1914; trans. P.P. Wiener, *The Aim and Structure of Physical Theory*, Princeton, NJ: Princeton University Press, 1954.(The English version influenced Hanson’s *Patterns of Discovery*.)

Feyerabend, P.K. (1960) ‘Patterns of Discovery’, *Philosophical Review* 59, 247-52.

Feyerabend, P.K. (1961) ‘Explanation, Reduction, and Empiricism’, in H. Feigl and G. Maxwell (eds) *Minnesota Studies in the Philosophy of Science*, vol. III, Minneapolis, MN: University of Minnesota Press.(An early summary of the position of Hanson’s primary opponent.)

Humphreys, W.C. (1968) *Anomalies and Scientific Theories*, Sudbury, MA: Jones & Bartlett.(An expanded version of a dissertation directed by Hanson.)

Wittgenstein, L. (1953) *Philosophical Investigations*, eds G.E.M. Anscombe and R. Rhees, trans. G.E.M. Anscombe, New York: Macmillan, and Oxford: Blackwell. (Hanson was one of the earliest philosophers of science to adopt Wittgenstein’s later ideas.)

Happiness

In ordinary use, the word 'happiness' has to do with one's situation (one is fortunate) or with one's state of mind (one is glad, cheerful) or, typically, with both. These two elements appear in different proportions on different occasions. If one is concerned with a long stretch of time (as in 'a happy life'), one is likely to focus more on situation than on state of mind. If a short period of time, it is not uncommon to focus on states of mind.

By and large philosophers are more interested in long-term cases. One's life is happy if one is content that life has brought one much of what one regards as important. There is a pull in these lifetime assessments towards a person's objective situation and away from the person's subjective responses. The important notion for ethics is 'wellbeing' - that is, a notion of what makes an individual life go well. 'Happiness' is important because many philosophers have thought that happiness is the only thing that contributes to wellbeing, or because they have used 'happiness' to mean the same as 'wellbeing'.

What, then, makes a life go well? Some have thought that it was the presence of a positive feeling tone. Others have thought that it was having one's desires fulfilled - either actual desires (as some would say) or informed desires (as others would say). It is unclear how stringent the requirement of 'informed' must be; if it is fairly stringent it can, in effect, require abandoning desire explanations and adopting instead an explanation in terms of a list of good-making features in human life.

1 The ordinary notion

'Happiness' is a central term - for some philosophers *the* central term - in ethics. It is also a term in everyday speech, and one should be aware of how the philosophical and the ordinary uses of the word are related.

'Happy' comes from the noun 'hap': what just happens, chance, luck. It came to mean having good hap, fortunate, lucky, a sense that it still retains (death can be a happy release). Etymology is not meaning, but there is a large etymological residue in our current use of the word. In a very common use now, to be 'happy' is to be satisfied or contented with having a good measure of what one regards as important in life. In this use of the word, 'happy' has to do with one's situation; one is fortunate. It also has to do with one's state of mind; one is glad or cheerful. It typically has to do with both situation and state of mind (one has the latter because of the former), but the two elements can appear in very different proportions in different cases. At one extreme, a martyr can go happy to the stake, merely secure in the conviction of right. At another extreme, a person can be happy (cheerful) for a few moments before realizing how unfavourable the situation actually is. There are other current uses of the word as well; for example, 'happy' can mean 'productive of favourable results' (a happy intervention). There is no definition of 'happiness', in the sense of a list of essential properties. Few words in a natural language, especially words covering as much ground as 'happiness', allow definition in that form. We can use these words correctly; hence we know their meaning. But we know it by catching on to the use of the words, not by catching on to a set of defining properties.

'Happiness' does not mean the same as 'pleasure', despite J.S. Mill's definition (1861) (see Mill, J.S. §8 Pleasure). He defined 'happiness' as 'pleasure, and the absence of pain'. But the words mark different features of life; the martyr who goes happy to the stake is unlikely to do so with pleasure. Nor does 'happiness' mean the same as 'wellbeing', if the latter term is used, as it often is, of one's actual situation. Because of the psychological element in the word, one may be 'happy' when things are going badly for one, if one is unaware that they are.

There are, as we have noted, two important strands to happiness, one's situation and one's state of mind. If it is a long stretch of time that one is interested in - say, 'a happy life' - one is likely to focus more on situation than on a flow of psychological states. If it is a short period, it is not uncommon to focus on psychological states. By and large philosophers are more interested in the long term use. Certain tensions can emerge in speaking of 'happy lives'. One's life is happy if one is content that life has brought one much of what one regards as important. The standard case is one in which one regards certain things as important because they *are* important and one thinks that life has brought them because it *has* brought them. But what happens to the use of the word 'happiness' if one or other of the standard conditions fails? Think, for instance, of a society in which women's expectations are very low. Aspirations are largely relative to expectations. A particular woman in that society might be pathetically

content with a small improvement in her generally miserable lot. Would we say that she has a ‘happy life’ merely because she is content? There would be strain in saying so because there is some pull in these lifetime assessments towards a person’s objective situation, and away from the person’s subjective responses. People may have various views about what good fortune in life is, but when they explain their conception of a happy life they will describe what they regard as good fortune, not what they believe, justified or not, will produce states of contentment or cheerfulness. I have been speaking of a tension in our use of the word ‘happiness’. There is also a troubling tension in our attempts to be happy. One route to happiness is to strive to achieve more of what is important in life. Another route is to be content with what one has already got. It is not easy, though just possible, to take both routes.

2 The notion in philosophy

‘Happiness’ is not the important notion in ethics; ‘wellbeing’ is - that is, a notion of good fortune, of what makes a life go well. ‘Happiness’ is important only if one thinks, as many philosophers of course have thought, that happiness is the only thing that contributes to wellbeing (a substantive claim), or if one uses ‘happiness’, as some philosophers have used it, to mean the same thing as ‘wellbeing’. What we want, and what philosophers have generally been in search of, is an account of what it is for a life to go well.

Aristotle spoke of ‘eudaimonia’ (literally, ‘good divine power’ or ‘good fortune’), normally translated in English as ‘happiness’ (see Eudaimonia). For him, it is the central term in ethics. ‘What is the supreme good attainable in our actions? Well, so far as the name goes, there is pretty general agreement. "It is happiness," say both intellectuals and the unsophisticated, meaning by "happiness" living well and faring well’ (*Nicomachean Ethics* 1095a). Aristotle’s notion therefore is what I have called ‘wellbeing’ - what actually makes a life go well - and it is often thought that the prominent psychological element in the English word ‘happiness’ makes it an inappropriate translation for ‘eudaimonia’. But even if ‘eudaimonia’ and ‘happiness’ do not mean quite the same, it does not follow that Aristotle, on the one hand, and anglophone philosophers who have used the term ‘happiness’ on the other, are talking about different things. For the reason given earlier, they are often both talking about wellbeing and, if they say different things about it, they may well be making contrary claims (see Aristotle §§21-2).

It is a plausible thought that faring well in some way involves acting well. How are virtue and happiness related? Some, such as Socrates, Plato, and Aristotle on some interpretations, have held that virtuous action is necessary and sufficient for happiness (see Plato §14; Socrates §4). On this view, all that matters for happiness is internal to one’s will: happiness *is* being virtuous and nothing else. Others, including Aristotle on other interpretations, have held that, important though virtue is to happiness, it is not all there is to it; external conditions such as health, wealth and avoidance of disasters matter too. Still others have thought that virtue is generally inimical to happiness (Thrasymachus as depicted by Plato in the *Republic* and Thomas Hobbes on some interpretations). It is somewhat easier to argue that virtue is sufficient for happiness than that it is necessary. Some elements of a good life (say, accomplishing things such as creating great art or making major scientific discoveries) do not seem to lose their value just because the agent concerned is not virtuous. If we could resolve these issues in favour of the first option, we should have a powerful answer to the vexed question, ‘Why be moral?’ Because, we could then say, it is a (or the only) way to be happy.

The most prominent use of the term ‘happiness’ in modern philosophy is to be found in the work of utilitarians (see Utilitarianism). They hold that acts are right in virtue of the value of their consequences, and what makes consequences valuable, according to classical utilitarians, such as Jeremy Bentham, J.S. Mill and Henry Sidgwick, is the presence of pleasure or happiness (see Consequentialism). This appeal to pleasure or happiness, although historically important in utilitarianism, is not essential to it. A utilitarian may say, and many modern utilitarians have said, that several irreducibly different things, perhaps including happiness, are valuable. The classical utilitarians put the term ‘happiness’ to various theoretical uses. Some of them used it in an empirical theory of action (psychological hedonism), which claims that pleasure or happiness is the only end which in fact we desire or at which we aim. They also used it in a normative theory of the ends of life (ethical hedonism), which claims that pleasure or happiness is the only thing worth acting for (see Hedonism §2). Once ‘happiness’ is given these theoretical roles, it comes under pressure to grow to fill them. If happiness is what in fact we aim at, then the term ‘happiness’ must encompass all that in fact we aim at (for example, saving one’s children at the cost of one’s own life). If it is the only thing worth aiming at, then whatever is worth aiming at (such as saving one’s children at the

cost of one's own life) must be fitted under the term 'happiness'. It is not that the term cannot be stretched to include them; it is rather that once it includes them it may have become a technical term, and we should then need to know its technical sense. The natural elasticity of the word 'happiness' can easily lead to confusion. It is easy to be deceived by shifts of sense such as this: 'happiness is what makes a life good' (a substantive claim: happiness is the one and only end in life); 'happiness is what makes life good' (a tautology: 'happiness' means 'what makes a life good').

The classical utilitarians paid insufficient attention to the meaning of their central term. Jeremy Bentham seems to have thought of 'pleasure' or 'happiness' primarily as a positive feeling tone. But when he defined 'utility', he described it as 'that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness (all this in the present case comes to the same thing)' (1789: 12), thus running together what improves one's situation and what produces certain mental states. J.S. Mill's approach was more complex. What links valuable states for him is that they are pleasurable, but this is not a case of there being any common positive feeling tone running through them all. In *Utilitarianism* (1861), he famously claims that pleasures differ in quality as well as quantity: states of higher quality are those that we prefer when sufficiently informed. Then, returning to differences in quantity, he observes that pleasures are not homogeneous, and that pleasure is always heterogeneous with pain. Informed preferences, he says, enter too in determining which is the greater of two pleasures or the worse of two pains. He may mean merely that the preference of the informed judge is a way for us to *learn* which pleasure is greater, but he sometimes seems to suggest that being the object of an informed preference is all that we can *mean* by a pleasure's being greater. Henry Sidgwick also adopted a preference account. He thought that the only things that we wanted for their own sakes were mental states, but that there was no single feeling tone running through all of them that gave them their unity. What unified them was that they were desired. The ultimate good, he concluded, was 'desirable consciousness' (1874: 397).

3 'Wellbeing', 'welfare', 'utility' and 'quality of life'

It is because 'wellbeing' (or 'welfare', 'utility', 'quality of life') is the important notion in ethics, not 'happiness', that most attention now goes to it (see Welfare). What, then, makes a life go well? How one answers that question will depend on how many of the following arguments one finds persuasive.

Wellbeing is not simply a positive feeling tone. There is no one tone running through all the things that make a life go well. Wellbeing, one might instead say, taking a lead from Sidgwick, is the fulfilment of desires. However, despite what some economists say, it certainly could not be the fulfilment of one's actual desires; they can be fulfilled and one be worse off. It would have to be, as Sidgwick himself says, the fulfilment of informed desires. But 'informed' in what sense? Suppose we say that a desire is informed if it exists when I am aware of all relevant facts and I commit no logical error (Brandt 1979: 10). But an irrational desire might well survive criticism by facts and logic, and its mere survival is less than it takes to make one better off. For instance, a man might have a crazy aim in life - say, counting the blades of grass in various lawns (Rawls 1971: 432-3). He accepts, let us say, that no one is interested, that the information is of no use, and he makes no logical error. Still, it is unlikely that we would regard the fulfilment of this obsessive desire as, in itself, enhancing his life - apart, that is, from preventing anxieties or tensions that might be set up by frustrating the desire, which are not the point. Cases like this suggest that our standard of 'informed' is not stiff enough yet. To make it stiffer, though, we should have to make desires 'informed' in some such strong sense as 'formed in proper appreciation of the nature of their object'. But this makes the mere occurrence of desire much less important and the nature of the object of desire much more important.

We might say therefore, as many philosophers do, that there are many different things that enhance life: happiness, seen as a state of mind, might be one, but perhaps also accomplishing something in the course of one's life, knowledge of certain basic metaphysical and moral matters, deep personal relations, and so on. Wellbeing, we could say, consists in having good things (a list of which we could provide).

The list-account of wellbeing has the potential of being much broader than the happiness-account of classical utilitarianism. Philosophers now debate just how broad it can be. Sidgwick thought that nothing enhanced a life unless it entered consciousness or experience. One might call this the 'experience requirement'. But some think that the requirement is too restrictive. They say that we sometimes want things other than states of consciousness, and these things seem to make our lives better. For instance we may desire a good reputation among people we

shall never know about, or posthumous fame, or to accomplish something with our lives. One way to clarify one's thoughts on this issue is to ask: supposing there were a foolproof machine that would give one any experience one wanted, would one plug in? What could matter except how life feels from the inside? Many would answer that what they want is to accomplish something with their lives (such as write a good novel, or discover a cure for AIDS), not to have the impression that they are. Many would say that they also want simply to be in touch with reality, even at a cost in desirable consciousness.

Some philosophers think that if, for these reasons, we drop the experience requirement, the list of things that make a life better will grow counter-intuitively large. Without the requirement, we seem to have to fall back on the view that wellbeing is the fulfilment of desires formed in proper appreciation of the nature of their object. I want our twenty-fifth century successors to flourish (any moderately decent person would), but surely the fulfilment of that desire centuries from now will not retroactively make my life better. The fulfilment of such desires has to be excluded, but the experience requirement is not the only way of doing it. The list-account of wellbeing will do it too. The list is composed by identifying what enhances life. Nothing enhances life but instances of items on that list: happiness, knowledge, accomplishment, personal relations, and so on. The fulfilment of my desire for our twenty-fifth century successors to flourish may be excluded simply because it does not fit under any of the headings on the list that we should eventually compile.

J.P. GRIFFIN

References and further reading

Annas, J. (1993) *The Morality of Happiness*, New York: Oxford University Press.(A history of ancient ethics, focused on the notions of virtue and happiness, from Aristotle onwards.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985, bks I, X.(Contains the core of his account of happiness.)

Austin, J. (1968) 'Pleasure and Happiness', *Philosophy* 43.(Examines the distinction.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, revised F. Rosen, Oxford: Clarendon Press, 1996.(Classic expression of the view that happiness consists in pleasure.)

Brandt, R.B. (1979) *A Theory of the Good and the Right*, Oxford: Clarendon Press.(Argues that rational persons would choose a utilitarian moral code, and provides important discussions of such issues as happiness, welfare and desire-satisfaction.)

Crisp, R. (1997) *Mill on Utilitarianism*, London: Routledge, chaps 2-3.(Discusses Mill's view of happiness.)

Den Uyl, D. and Machan, T.R. (1983) 'Recent Work on the Concept of Happiness', *American Philosophical Quarterly* 20.(A survey of contemporary discussions.)

Glover, J. (ed.) (1990) *Utilitarianism and its Critics*, New York: Macmillan, pt II. (Good selection of writings on happiness, with special relevance to utilitarianism.)

Griffin, J. (1986) *Well-Being*, Oxford: Clarendon Press, pt I.(Examines different accounts of wellbeing in modern ethics.)

Hobbes, T. (1651) *Leviathan*, ed. C.B. Macpherson, Harmondsworth: Penguin, 1968.(Can be interpreted as arguing that virtue and happiness are in opposition.)

Kraut, R. (1979) 'Two Conceptions of Happiness', *Philosophical Review* 80.(Explores the similarities and differences in two conceptions of happiness, Aristotle's and our own.)

Mill, J.S. (1861) *Utilitarianism*, ed. R. Crisp, Oxford: Clarendon Press, 1998, ch. 2.(Argues that happiness consists in both higher and lower pleasures.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press, appendix I. (Identifies different positions about what makes someone's life go well.)

Plato (*c*.380-367 BC) *Republic*, trans. G.M. Grube, revised by C. Reeve, Indianapolis, IN: Hackett Publishing Company, 1992.(Famous ancient argument for the identification of virtue and happiness.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (In the course of developing a non-utilitarian theory of justice, presents important criticisms of utilitarianism.)

Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan; 7th edn, 1907. (The fullest and most closely argued version of classical utilitarianism.)

Sumner, L.W. (1996) *Welfare, Happiness, and Ethics*, Oxford: Clarendon Press.(Rejects hedonism and desire accounts, and defends a close connection of wellbeing to happiness.)

Hare, Richard Mervyn (1919-)

R.M. Hare is the creator of the ethical theory called 'prescriptivism'. This holds that moral statements differ from purely factual ones in prescribing conduct; they differ from simple imperatives in invoking universal principles that apply to all similar cases. The theory has three aspects: prescriptivity and universalizability as formal features of moral statements; appeal to the Golden Rule (that we should do to others as we wish them to do to us) for selecting moral principles; two levels of practical thinking, critical and intuitive.

1 Prescriptivity and universalizability

Richard Hare is a British philosopher, educated at Oxford, who has held chairs in moral philosophy at Oxford and Gainesville, Florida. His best-known work, *The Language of Morals* (1952), introduces a distinction between prescriptive and descriptive meaning. Prescriptive meaning is defined in relation to imperatives: a statement is prescriptive if it entails, if necessary in conjunction with purely factual statements, at least one imperative; and to assent to an imperative is to prescribe action. Descriptive meaning is defined in relation to truth-conditions: a statement is descriptive to the extent that factual conditions for its correct application define its meaning. In this usage, the factual is that which is only contingently motivating: desire is no part of sincere assent to a purely factual statement. The meaning of a moral statement is prescriptive, but may also be partly descriptive. Thus '*A* [a person] ought to Čphis;' entails the imperative 'Let *A*Čphis;', so that to assent to it sincerely is to have an overriding desire (which in application to oneself will amount, if its satisfaction is practicable, to an intention) that *A*. If there are agreed reasons for -ing within a linguistic community, for example that it is enjoyable, '*A* ought to Čphis;' may take on the descriptive implication 'Čphis;-ing is enjoyable'. '*X* [a person, object or whatever] is a good *F* [a kind of thing]' prescribes choice within a certain range; it takes on a descriptive connotation if there are agreed standards for assessing *F* 's. Cases of failing to try to do what one admits one ought to do may involve psychological incapacity, or an off-colour use of 'ought' whereby it retains descriptive meaning but loses its prescriptive meaning. What the modal '*A* ought to Čphis;' adds to the simple 'Let *A*Čphis;' is universalizability: one who assents to the former is implicitly accepting a universal principle that applies equally to anyone else whose condition and situation are identical in kind (see Prescriptivism).

2 The Golden Rule

Freedom and Reason (1963) extracts from these features a mode of reasoning, the so-called 'Golden-Rule Argument', for selecting moral principles; this was subsequently clarified and fortified in writings leading up to *Moral Thinking* (1981). In wondering whether to assent to the statement '*A* ought to Čphis;', I have to reflect whether I can prescribe that everyone should act in the same way in every possible world, whatever my role within that world. (This talk of possible worlds is useful, though Hare denies it to be essential.) 'I' connotes no essence (for example, human): I occupy every possible role in some possible world. But there is a prescriptive aspect to its meaning: to take a role within some possible but non-actual world as *my* role is to give weight to the desires of the occupant of that role as if they were actually my own. Hence, I can rationally assent to a particular 'ought'-statement only if it is derivable from some universal principle that I shall accept if I give impartial and positive weight to all preferences whose fulfilment would be affected by its fulfilment. Thus moral reflection generates a universalized prudence. Possibly it should also respect 'external' preferences, that is, the preferences of those (say the dead) who will *experience* no satisfaction; *Moral Thinking* leaves this as 'unfinished business'. Moral ideals register within this framework simply as universal preferences; to allow one's own ideals to override the stronger or more prevalent desires and ideals of others is a kind of egoism, and so excluded. The upshot is a variant of utilitarianism that aims at the maximization not of happiness, but of the satisfaction of preferences (see Utilitarianism §2). Human decision remains free, however rational and informed, because anyone can avoid the constraints of morality by declining to moralize; for this reason, 'is' still fails to entail 'ought'.

3 Critical and intuitive thinking

Reasoning according to the Golden Rule will ascribe maximum observance-utility to highly specific universal principles. However, a greater acceptance-utility may attach to more general principles that it is easier to apply without error or self-deception. *Moral Thinking* elaborates a distinction between a 'critical' level of thinking, conducted by 'archangels' with the use of the Golden-Rule Argument, and an 'intuitive' level, conducted by

'proles' with the use of simple principles (often articulating emotional responses) whose acceptance can be justified at the critical level. These two levels define not two social castes, but two roles between which each of us learns to alternate as appropriate. Intuitive objections to utilitarianism (such as that it neglects rights) can often be accommodated at the intuitive level, and then cease to be objections (see Intuitionism in ethics §3).

Hare further argues that prescriptivism guarantees objectivity, while descriptivism collapses into relativism. How I can rationally apply the terms 'right' and 'wrong' is determined by what I can will universally, which rides free of the contingencies of my own tastes and intuitions. But if the correct application of such terms was determined by their descriptive meaning within some linguistic community, it would be relative to the culture and ideology of that community. Only a prescriptivist ethics can achieve moral universality. The theory has received much critical attention, and while few may accept it as a whole, the simplicity and fertility of its main ideas make it a paradigm of practical philosophy.

A.W. PRICE

List of works

Hare, R.M. (1952) *The Language of Morals*, Oxford: Clarendon Press.(Discusses imperative inference, explains 'good' as a term of commendation, and relates 'ought' to imperatives.)

Hare, R.M. (1963) *Freedom and Reason*, Oxford: Clarendon Press.(Derives from prescriptive universalizability a Golden-Rule Argument for testing moral maxims.)

Hare, R.M. (1971a) *Essays on Philosophical Method*, London: Macmillan.(Essays, partly historical, exploring the proper goals and methods of philosophy, especially of moral philosophy.)

Hare, R.M. (1971b) *Practical Inferences*, London: Macmillan.(An important set of essays on imperative logic, practical reasoning, and meaning and speech acts.)

Hare, R.M. (1972a) *Applications of Moral Philosophy*, London: Macmillan.(Essays on practical issues originally intended for a non-professional audience. 'Nothing Matters' is a lucid and engaging statement of Hare's scepticism about the objective/subjective debate in ethics.)

Hare, R.M. (1972b) *Essays on the Moral Concepts*, London: Macmillan.(Partly polemical essays exploring the relation between moral and descriptive concepts.)

Hare, R.M. (1981) *Moral Thinking: Its Levels, Method and Point*, Oxford: Clarendon Press. (Distinguishes two levels of moral thinking as 'critical' and 'intuitive'. Importantly discusses the Golden-Rule Argument, alleging a prescriptive use of 'I' and 'my'.)

Hare, R.M. (1982) *Plato*, Oxford: Oxford University Press.(Introduces Plato's philosophy, especially as a response to the moral uncertainties of his time.)

Hare, R.M. (1989a) *Essays in Ethical Theory*, Oxford: Clarendon Press.(A valuable collection of essays, some polemical, others constructive, preparatory or accessory to *Moral Thinking*.)

Hare, R.M. (1989b) *Essays on Political Morality*, Oxford: Clarendon Press.(Essays on legal obligation, war and terrorism, rights, justice and the environment.)

Hare, R.M. (1992) *Essays on Religion and Education*, Oxford: Clarendon Press.(Essays on the relation of religion to morals, and on the nature of moral education from the viewpoint of a prescriptivist meta-ethics.)

Hare, R.M. (1993) *Essays on Bioethics*, Oxford: Clarendon Press.(Essays in medical ethics and related topics. Particularly important are discussions, in relation to abortion, of decisions about possible people.)

Hare, R.M. (1997) *Sorting out Ethics*, Oxford: Clarendon Press.(Recommends prescriptivism in the context of a mapping of the possible alternatives within meta-ethics.)

References and further reading

Fehige, C. and Meggle, G. (eds) (1995) *Zum moralischen Denken (Towards Moral Thinking)*, Frankfurt: Suhrkamp.(Twenty-three essays, with replies by Hare, in German.)

Seanor, D. and Fotion, N. (eds) (1988) *Hare and Critics: Essays on Moral Thinking*, Oxford: Oxford University Press.(Contains concise comments by Hare; careful and sophisticated.)

Harrington, James (1611-77)

Harrington was the premier English republican political theorist. His The Commonwealth of Oceana (1656), published soon after the Civil War, analysed the collapse of monarchy and recommended institutions for a perfect commonwealth. He argued that forms of government were shaped by modes of land tenure; the decline of the feudal aristocracy and rise of the gentry rendered monarchy inviable. His proposed republic entailed regular elections for all public offices and secret ballots among a citizenry of independent gentlemen. Harrington influenced English, American and French radicals throughout the eighteenth century. Today he tends to be a talisman for those who would inject an aspect of 'civic republicanism' or 'public virtue' into contemporary politics, by contrast with the liberal rights theories of the natural jurisprudence tradition.

1 Life and works

Born to a gentry family in Lincolnshire, Harrington studied at Oxford and the Middle Temple. In the 1630s he served as a volunteer in the Thirty Years War in Germany and undertook the Grand Tour, visiting France and Italy. He served King Charles I and accompanied him during his captivity: his republicanism was not grounded in personal animosity. His publishing career was brief. His chief work, *The Commonwealth of Oceana*, appeared in 1656, and was followed by a series of vindications. His notoriety was greatest in 1659-60 when the Rota Club met in a London coffeehouse to debate models of government. Harrington fell silent after the restoration of monarchy in 1660, and was briefly jailed. He died apparently insane.

Oceana is dedicated to Oliver Cromwell; its Legislator, called Olphaus Megaletor, is plainly an Oliverian figure. Like other republicans, Harrington saw Cromwell's Protectorate as a betrayal of the 'Good Old Cause', a return to quasi-monarchic rule, yet he appealed to Cromwell to act heroically as midwife to a utopian commonwealth. Sir Henry Vane's *A Healing Question* and Marchamont Needham's *The Excellency of a Free State* were published in the same year (1656). The Harringtonians were pervasively civic humanist in their language and committed to constructing republican institutions - unlike John Milton, whose 'republicanism' amounted to a defence of king-killing, couched in the language of natural jurisprudence (see Republicanism).

Harrington's mentors were Aristotle, who provided the concept of the self-sufficient gentlemanly household; Polybius, who celebrated the mixed and balanced polity; Machiavelli, who suggested the emphasis on arms and citizen militias; and John Selden, who documented English feudalism. Harrington admired the republics of ancient Rome and modern Venice, as well as Israel in the age of the Judges. His hero-lawgivers were those of the Renaissance humanists: Solon, Lycurgus, Numa, Romulus and (echoing Machiavelli) Moses. Harrington was transfixed by 'the archives of ancient prudence' and by Machiavelli, 'the only politician of later ages', whom he read as the republican idealist of the *Discourses*.

There is also a distinctive Platonist voice. Harrington's only essay in metaphysics was his *Mechanics of Nature* (written around 1662 and published posthumously in 1700). It shows him to have adopted a hylozoist and pantheist position. Behind nature and human polities lies the *anima mundi* or world-soul. More broadly, Harrington inclined towards Socinianism. Grace and charism have little place: the divine is fulfilled through natural virtue, and the Kingdom of God is no other than the perfect commonwealth.

2 The argument of *Oceana*

Oceana is divided into two parts, a diagnosis of contemporary England in 'The Preliminaries', followed by a description of the imagined commonwealth. 'The Preliminaries' are known for their doctrine that forms of government tend to follow the 'balance of property', and hence that the art of politics consists of 'the skill of raising such superstructures of government as are natural to the known foundations' (1656 (1977): 202). There are 'natural' (stable) and 'violent' (unstable) governments, those which accord with, or violate, their tenurial foundations. Harrington argues that the two centuries prior to the Civil War saw a drastic change in the balance of property. Crown and aristocratic lands were gradually distributed to the gentry. Without the support of a dominant nobility, or a standing army, kingship became unsustainable. Thus, the feudal or 'Gothic' polity, in which land was the reward for military service to the crown, was gone forever. Harrington's editor, John Toland, remarked, 'that empire [government] follows the balance of property, whether lodged in one, in a few, or in many hands, he was

the first that ever made out; and is a noble discovery…the foundation of all politics' (1700a: preface).

The institutions of Oceana are grounded in the new circumstances of a gentry commonwealth. Oceana has a senate composed of 300 of the wealthier landholders, who have the power of debate, and an assembly of 1,050 smaller freeholders, who have the power of resolution. There is a three-year term of office, one-third of each house coming up for election annually. The citizen body - the electors - must be economically independent (not servants or poor), be over 30 years of age, and have served in the militia. There are four executive councils - of state, war, religion and trade - and members must stand down after two years in office. Olphaus Megaletor is the presidential figure. All offices, including religious and military, are elective, with regular rotation: 'They, who do command today, shall learn again tomorrow to obey'. The basic geographical unit remains the parish (rationalized into 10,000 in number), revitalized as the locus of civic life. There are counties and hundreds, but no boroughs: this is a resolutely rural commonwealth. Each year, one in five citizens compulsorily serves in the militia.

A crucial stipulation is the Agrarian Law, which, through the abolition of primogeniture and a ceiling of £2,000 worth of property, aimed to prevent the emergence of an overweening oligarchy. The Agrarian Law renders Oceana 'an immortal commonwealth', free from the instability of the Polybian cycles of rise and decline. It also inclines Oceana more towards the Athenian model of democracy than was customary among the Roman-inspired Renaissance republicans, although the presidential and senatorial elements sufficiently exemplify the traditional one-few-many formula.

The whole fabric is designed to ensure 'an empire of laws and not of men', a regime of reason and not of mere will. It is a commonwealth which nurtures civic virtue and service to the public. The landed gentleman governs himself and so is fit to govern the polity. The obverse of virtue is corruption: tyranny is that which serves private or factional interests rather than the public good (see Corruption §§1, 3).

3 Interpretations

Modern interpretations of Harrington fall into three camps. One group, in the thrall of positivist political science, celebrates Harrington's empiricism and secularism, his grasp of the mechanics of power and the calculus of interest. A second group, Marxian-inspired (notably R.H. Tawney, Christopher Hill and C.B. Macpherson), finds in Harrington an economic determinist, the first theorist to offer a structural interpretation of the English Revolution, and a spokesman of the bourgeois gentry. The third group (J.G.A. Pocock and his followers) emphasizes the civic humanist character of Harrington's thought. Against the political scientists, the Pocockians point to his Platonic and occasionally millenarian language. Against the Marxians, they argue that Harrington was no determinist, and that his grasp of economics went no further than land tenure. His account of the transition from feudalism rested on the contingent political changes by which Henry VII and Henry VIII curtailed aristocratic estates and retinues, and redistributed monastic lands. The Agrarian Law manifestly limits a free market in land.

Some attention has been paid to Harrington's 'civil religion'. Clerics quickly attacked *Oceana*, and Harrington retorted with extensive discussion of ecclesiastical power. He was vehemently anticlerical, praising Hobbes (from whom he otherwise diverged) for his critique of the pretensions of papalists, prelatists and presbyterians. Harrington coined the word 'priestcraft'. *Oceana* is secular in the sense that it rejects divine right clericalism and embraces religious toleration within a broad church. Yet Oceana is a Christian commonwealth, indeed one in which the Church finds its apotheosis in the commonwealth itself. Harrington argues that it was so in Israel, and he explains that the Greek '*ecclesia*' or 'congregation' meant only the civic assembly of the citizens. The ancient manner of choosing clergy by popular vote - *chirotonia* - was perverted by usurping ecclesiastics into the practice of clerical 'laying on of hands' - *chirothesia*. Harrington's millenarianism is most visible in his recommendation that Ireland be colonized by the Jews.

4 Influence

Harrington's republicanism was profoundly influential in the later seventeenth and eighteenth centuries. *Oceana* inspired a body of 'neo-Harringtonians', the True, Old, or Real Whigs, who provided a radical critique of England's monarchical polity, and later attacked the executive tyranny of prime ministerial and cabinet government. Harrington helped fuel assaults on standing armies, 'placemen' and irregular elections. Early examples are *Plato redivivus* (Plato Revived) (1680) by Harrington's friend Henry Neville, who also published an

edition of Machiavelli’s works; and *An Essay upon the Roman Government* (1698) by Walter Moyle. Toland published Harrington’s collected *Works* in 1700. Eighteenth-century political analysts, such as Bolingbroke and Hume, deployed the idea of ‘the balance of property’, although they rejected the republican utopianism. Harrington’s mark is also visible in some of the early constitutions of the American colonies, and among the rebels of 1776. It was said of General Warner, the hero of Bunker Hill, that ‘like Harrington he wrote, like Cicero he spoke, like Hampden he lived, like Wolfe he died’. In 1779 it was proposed, albeit ironically, that Massachusetts be renamed Oceana. In 1795 *Oceana* was republished in revolutionary France and the Abbé Sieyès drew up a Harringtonian constitutional plan. In the nineteenth century George Grote found inspiration in Harrington for his campaign for the secret ballot.

MARK GOLDIE

List of works

(1977) *The Political Works of James Harrington*, ed. J.G.A. Pocock, Cambridge: Cambridge University Press; abridged edn, *James Harrington: The Commonwealth of Oceana and A System of Politics*, ed. J.G.A. Pocock, Cambridge: Cambridge University Press, 1992.(Contains all of Harrington’s political writings, the chief of which are listed separately.)

(1656) *Oceana*, in *The Political Works of James Harrington*, ed. J.G.A. Pocock, Cambridge: Cambridge University Press, 155-360.(England’s premier republican text. Part 1 analyses the historical reasons for the collapse of the monarchy; part 2 sketches the spirit and institutions of an ideal agrarian republic.)

(1658) *The Prerogative of Popular Government*, in *The Political Works of James Harrington*, ed. J.G.A. Pocock, Cambridge: Cambridge University Press, 389-498.(Book 1 is a reply to Matthew Wren’s *Considerations upon Mr Harrington’s Commonwealth of Oceana*; book 2 is an anticlerical history of priestly ordination, directed against Presbyterians and Anglicans.)

(1659a) *The Art of Lawgiving*, in *The Political Works of James Harrington*, ed. J.G.A. Pocock, Cambridge: Cambridge University Press, 599-704.(Book 1 recapitulates arguments from *Oceana* about the ’foundations and superstructures of government’; book 2 discusses the commonwealth of ancient Israel; book 3 elaborates on the institutions needed in a republic.)

(1659b) *Aphorisms Political*, in *The Political Works of James Harrington*, ed. J.G.A. Pocock, Cambridge: Cambridge University Press, 761-80.(A short work consisting of 120 apothogems on government.)

(1700a) *The Mechanics of Nature*, in *Works*, ed. J. Toland, London.(A metaphysical essay of pantheist or neoplatonist leanings. The divine is embodied in the world soul and in natural virtue.)

Harrington, J. (1700b) *Works*, ed. J. Toland, London.(The neo-republican Toland provided the standard edition until the twentieth century and offered ammunition for the radical critique of English government during the eighteenth century.)

Harrington, J. (1700c) *A System of Politics*, in J. Toland (ed.), *Works*; repr. in *The Political Works of James Harrington*, ed. J.G.A. Pocock, Cambridge: Cambridge University Press, 833-54.(Written around 1661 and published posthumously, in it Harrington analyses the art of government, reducing it to several hundred aphorisms.)

References and further reading

Goldie, M. (1987) ‘The Civil Religion of James Harrington’, in A. Pagden (ed.) *The Languages of Political Theory in Early Modern Europe*, Cambridge: Cambridge University Press.(Explores Harrington’s religious thinking.)

Moyle, W. (1698) *An Essay upon the Roman Government*, in *Works*, London, 1726.(Moyle used a discussion of the virues and institutions, and the corruption and decline, of Rome as a warning for contemporary England.)

Needham, M. (1656) *The Excellency of a Free State*, London.(After Harrington’s *Oceana*, the principal republican work published during the interregnum. Comprises newspaper editorials published in 1651 and 1652 in *Mercurius Politicus*.)

Neville, H. (1680) *Plato redivivius*, London.(An adaptation of Harrington’s republican theories for the circumstances of the restored monarchy. Neville proposed to strip the crown of most of its executive powers.)

Pocock, J.G.A. (1975) *The Machiavellian Moment*, Princeton, NJ: Princeton University Press. (Places Harrington and the neo-Harringtonians in the context of civic humanism.)

Russell Smith, H.F. (1914) *Harrington and His Oceana*, Cambridge: Cambridge University Press. (Discusses Harrington’s republican constitutionalism and its legacy.)

Tawney, R.H. (1942), *Harrington’s Interpretation of his Age*, Oxford: Oxford University Press.(A Marxian

emphasis on Harrington as a theorist of economy and polity.)
Vane, Sir H. (1656) *A Healing Question*, London.(A work of 'godly' republicanism. The armed 'saints' have a duty to create a perfect commonwealth, protected by a righteous army.)

Hart, Herbert Lionel Adolphus (1907-93)

H.L.A. Hart, Professor of Jurisprudence at Oxford University, 1952-1968, is an outstanding representative of the analytical approach in jurisprudence and philosophy of law. He restated 'legal positivism' in the tradition of Jeremy Bentham and John Austin, differentiating between law's existence and its moral qualities. But he rejected the Benthamite identification of law with a sovereign's commands, advancing instead a theory of law as comprising a special, systematically organized, kind of social rules. He did this in a linguistic-analytical style, showing how attention to our way of speaking and thinking about rules can yield new insights into their nature.

Hart aimed to establish that legal obligation is intelligible in itself and yet conceptually distinct from moral obligation (see Law and morality §3). This depends on differentiating different types of rules in the legal setting, showing how some impose requirements upon conduct, while others confer power over rules of the former, 'primary' type. Legal obligations are imposed by 'primary rules', but since these can be varied through application of 'secondary' power-conferring rules, they can and do differ in content from moral obligations. He also argued for the possibility of a moral critique of law as it is, aimed at reforming it in favour of liberty and tolerance. His work here had important practical influence. He offered a distinctive account of rights, including human rights, and argued that a 'minimum content of natural law' is present in all viable forms of human social order (see Natural law §5).

The Concept of Law (1961), Hart's outstanding work, explains law as a system of social rules distinct from the rules and principles of morality, whether conventional or ideal. Legal rules divide into 'primary rules' laying down duties to act (for example, in pursuance of the terms of a contract) or to abstain (for example, to avoid harming others or attacking them in body or in reputation, or defrauding them) and 'secondary rules' relating in a systemic way to primary rules. Some secondary rules ('rules of adjudication') regulate the ways in which primary rules can be implemented through courts and other agencies. Others are 'rules of change', either enabling legislatures to secure large-scale legal changes, or enabling private persons to exercise powers under, for example, the law of contract, or property law, or trust law, so as to make changes in individual legal relationships.

Finally, there is the 'rule of recognition' containing the criteria for validity of primary rules and other secondary rules. A legal system thus comprises a rule of recognition and all the other rules valid by reference to it. The existence of a rule of recognition, and thus of the legal system, depends on human practice, especially the practices of officials. Acceptance of a common public standard under which all persons have an obligation to respect and apply laws that are valid by that standard is constitutive of the existence of a rule of recognition (see Legal concepts §3).

This framework also accounts for rights, the legal position held by those who are able to act in their own interest to activate or enforce duties laid upon others by law. Related kinds of right, also characterized by the law's deference to individual choice, are the liberties enjoyed by those who are not duty-bound to act in a certain way, or the powers conferred by 'rules of change', particularly those conferring private powers. This leaves open the question whether there are any 'natural' or 'human' rights. Hart's answer was hypothetical: if there are any rights that are a morally required part of any social order, the first and central of them must be the equal right of all humans to be free (see Rights §4). As for 'natural law', human nature makes restraints on violence and deception, and provision of some sort about possession of and access to material things, essential to the existence of any viable society. But this is not a detailed blueprint for either law or conventional morality, and hence does not justify a restricted definition of law that excludes from it any grossly unjust rules.

Law, Liberty and Morality (1963) was his own statement on the limits of law. He demanded that law be used only to prevent harm to others or for some restricted forms of 'paternalism' restraining people from severely self-harming activity (see Law, limits of §4). 'Positive morality', the conventional moral opinion of a community, is firmly differentiated from 'critical morality', rational and enlightened standards for criticizing or upholding norms of law or of positive morality. Hart rejected the use of law simply to enforce positive morality. *Punishment and Responsibility* (1968) was a collection of essays in which he developed a middle way between utilitarian and retributive principles of punishment, arguing that the latter ought to prevail in respect of distribution of punishments (only to those who deserve it, and in proportion to desert), while general deterrence remained the

justifying aim of the system (see Crime and punishment §2).

Causation in the Law (jointly with Tony Honoré, 1958) is an outstanding example of an effective combination of philosophical analysis and legal exposition, giving an account of the legally important idea of causality poised between common sense, science and law (see Causation). Hart's collected essays were published in two volumes (*Essays on Bentham*, 1982; *Essays in Jurisprudence and Philosophy*, 1983) one blue and one brown. If there was here an allusion to Wittgenstein's 'blue book' and 'brown book' it was not inapt; for Hart above all brought to jurisprudence the linguistic-philosophical insights of Wittgenstein, J.L. Austin and Ryle (see Ordinary language philosophy).

Hart's elucidation of rules involves a hermeneutic approach examining conduct 'from the internal point of view' of a group's members (see Legal hermeneutics). This owes much to Wittgenstein, and something also to Max Weber, but remains an original contribution to rule-analysis (see Norms, legal §2). Some critics have argued that one cannot start down this road without ending up in a fully 'natural law' approach. Others have argued that analysis of law purely in terms of conventional rules omits its essentially principled character, or that its sociological presuppositions are untenable. Hart's legal positivism, albeit highly persuasive, remains open to doubt.

See also: Austin, J.; Bentham, J.; Law, philosophy of

NEIL MacCORMICK

List of works

Hart, H.L.A. with Honoré, T. (1958) *Causation in the Law*, 2nd edn, Oxford: Clarendon Press, 1985.(A detailed study of causation, pursued first through general philosophical accounts and second, at considerable length, through analysis of case law dealing with causal problems. Accessible to the general reader.)

Hart, H.L.A. (1961) *The Concept of Law*, Oxford: Clarendon Press.(Hart's most celebrated work, originally intended as a high-level student's introduction to jurisprudence. A gracefully written book, deceptively easy to read.)

Hart, H.L.A. (1963) *Law, Liberty and Morality*, Oxford: Clarendon Press.(An accessible, well-argued account of the case for a liberal and anti-moralistic approach to legislation and law.)

Hart, H.L.A. (1968) *Punishment and Responsibility*, Oxford: Clarendon Press.(A collection of essays in which aspects of utilitarianism and of retributivism are reconciled.)

Hart, H.L.A. (1982) *Essays on Bentham*, Oxford: Clarendon Press.(A collection of papers mainly directed at positivism and utilitarianism, containing important revisions of Hart's earlier approach to jurisprudence.)

Hart, H.L.A. (1983) *Essays in Jurisprudence and Philosophy*, Oxford: Clarendon Press.(A collection of Hart's main contributions to discussions of basic jurisprudential issues, including critiques of his main rivals.)

References and further reading

Bayles, M. (1992) *Hart's Legal Philosophy: an Examination*, Dordrecht/Boston/London: Kluwer.(An extended account of Hart's theories and of objections to them; some telling criticisms carry the debate beyond Martin and MacCormick. Contains complete Hart bibliography.)

Cotterrell, R. (1991) *The Politics of Jurisprudence*, London: Butterworths.(Takes up a critical position on the style of philosophical jurisprudence developed by Hart and continued by various successors; a critique of Hartian (and other) legal philosophy from a critical sociological standpoint.)

MacCormick, N. (1981) *H.L.A. Hart*, London: Edward Arnold.(A relatively short intellectual biography and sympathetically critical account of Hart's main tenets; an essentially Hartian view of Hart.)

Martin, M. (1987) *The Legal Philosophy of H.L.A. Hart*, Philadelphia, PA: Temple University Press.(Longer and more detailed than MacCormick, and more distanced from Hart's own position.)

Hartley, David (1705-1757)

David Hartley commands a distinctive place in Enlightenment thinking for his attempt to establish an empiricist epistemology upon a foundation of ontological materialism - in other words, a philosophy of mind that incorporates a physiology of the brain. He also set forth an optimistic vision of human progress which was nonetheless cast within the framework of a transcendental theology. Though his views might seem to be a singular fusion of disparate strands, they nevertheless epitomized much liberal and advanced English thinking of the time, and exercised considerable influence upon the philosophical radicalism of subsequent generations.

1 Life and contacts

David Hartley was born the son of a poor Anglican clergyman in Armley, Yorkshire, probably on 30 August 1705. Both his father and mother died while he was still a child. He attended Bradford Grammar School before matriculating at Jesus College, Cambridge, as a sizar at the age of seventeen, at precisely the moment when a thoroughly modern synthesis of Newtonian natural philosophjy and Lockean philosophy was becoming the staple of the undergraduate curriculum. He graduated BA in 1726 and MA three years later, and held a fellowship at the college from 1727 until he married in 1730.

As his writings show, Hartley was a devout Christian; yet scruples against signing the Thirty-Nine Articles precluded him from taking orders, and he chose medicine as a career. Apparently without any medical degree or licence, he began to practise in Newark, moving on to Bury St Edmunds. After his wife's death, he remarried in 1735, and his second wife's wealth enabled him to settle just off London's fashionable Leicester Square (then the physicians' quarter); subsequently her ill-health induced the couple to move to Bath, where Hartley built up a successful medical practice.

A Fellow of the Royal Society, Hartley was able to move in the best intellectual and scientific circles. His friends included the physician Sir Hans Sloane (who became president of the Royal Society), the Revd Stephen Hales, famous for his pioneering physiological experiments, the evangelical William Law, and Joseph Butler, the leading rational Anglican theologian. Hartley had a part to play in major intellectual and philanthropic causes of the day, championing smallpox inoculation and John Byrom's shorthand system, and writing pamphlets to secure a parliamentary subvention for Mrs Joanna Stephens' lithontriptic nostrum against kidney and bladder stones (he had suffered from the disorder while still a young man). He died in Bath on 28 August 1757.

2 Physiological philosophy

Hartley is best remembered for his *Observations on Man, His Frame, His Duty, and His Expectations*, published in 1749 - a two-volume philosophical *tour de force* that systematized views set out in his earlier *The Progress of Happiness Deduced From Reason* (1734) and *Conjecturae quaedam de sensu, motu, et idearum generatione* (1746). The *Observations on Man* offered a comprehensive vision of the individual considered both as an earthly being and in regard to a future state. Emphasizing the view that all knowledge derives from experience, it drew heavily upon the empiricist theory of mental operations explicated in John Locke's *Essay Concerning Human Understanding* (1690; see Locke, J.). It also absorbed the innovative associationist utilitarianism of the Revd John Gay's *Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality* (1731), which set out a pleasure and pain psychology as the key to the formation of opinions and to the philosophy of action. Like Locke and Gay, Hartley sought to refute nativist theories of cognition and morality (see Nativism), insisting that complex ideas were built up from simple imputs by repeated combinations of what Hartley called the 'sensations of the soul'. Hartley thus grounded thoughts and values upon the Lockean principle of the association of ideas.

Unlike Locke and Gay, however, Hartley aimed to set these epistemological and psychological observations upon concrete physical foundations - the anatomy of the nervous system and the physiology of 'motions excited in the brain'. For this he drew upon the theory of sensation suggested in the 'Queries' to Isaac Newton's *Opticks* (1704; see Newton, I.). Newton had shown how light vibrated in a medium; such vibrations had an impact upon the retina; having impinged upon the eye, Hartley argued, these corpuscular motions set off further vibrating waves that passed along the nerves to the brain. The Lockean notion of the association of ideas was thus visualized and made material by Hartley in terms of reiterated vibrations in the white medullary matter of the brain and spinal cord,

which resulted in lasting traces or vestiges that served as the physical substrate of complex ideas, memory and dispositions. Unlike the French *philosophes* like La Mettrie, Hartley framed his materialist physiological psychology in terms of a Christian natural theology. In his view, materialism was not the slippery slope to atheism, precisely because it was the Christian God who, in His Wisdom, had endowed matter with all its powers and potentialities. The necessitarianism entailed by materialism was, in Hartley's opinion, the finest guarantee of the universal operation of cause and effect, hence of the uniformity of nature, and so of the boundless power of God.

As befitted a medical man, the first volume of *Observations on Man* explored major aspects of neurophysiology, discussing the human mind and appetites in terms of the evolution of complex ideas and habits from elementary sensations. Hartley demonstrated the formation of mental associations on the basis of the vibrations of particles in the nervous system which persisted in the form of the more minute 'vibratiuncles', which in turn provided the physical basis for memory, regarded in something of a rather Hobbesian way as decaying sense (see Hobbes, T.). Individual chapters explored the mental physics of feeling, taste, smell, seeing and hearing. The second volume extended the system to account for morality and the individual's prospects in a future state.

Unlike Stephen Hales, Hartley was less an experimenter than a systematizer, offering a comprehensive framework for interpreting the phenomena of life and mind. Among his followers, the *Observations on Man* came to be seen as the fountainhead of key biological, psychological and social doctrines. Hartley's tenets provided the framework for the associationist heritage in psychology - in particular, learning theory. His conjectures concerning the physiology of the nervous system offered suggestions for sensory-motor theories later influential in neurophysiology, and for the experimental localization of brain functions. Hartleian notions are the distant ancestors of Pavlovian notions of conditioned reflexes.

3 Influence

Hartley's particular neurophysiological theories proved erroneous, yet his work was of cardinal significance in the development of attempts to apply scientific concepts to the study of the individual as a social being and as a progressive creature. His was the first methodical elaboration of the explanatory principle (psychophysical parallelism) that came to play a role in the human sciences analogous to that played by the concept of gravity in the physical sciences. Though Hartley was personally devout, his unification of sensation, motion, association and volition within a mechanistic theory of consciousness and action created a framework of thinking that later supported more secular readings of the concepts of utility. His agenda was widely taken up in late Enlightenment and nineteenth-century doctrines as a means of accounting for cumulative ordered change through experience, changing responses being explained in terms of adaptation to pleasurable and painful consequences. Hartley put learning theory and moral judgment on a scientific basis.

Hartley had an influence on Samuel Taylor Coleridge, whose brief flirtation with his notions led to his first-born being baptised Hartley Coleridge. Joseph Priestley stressed Hartley's determinism, but omitted his materialist neurology in his edition of the *Observations on Man*, retitled [Hartley's] *Theory of the Human Mind* (1775). Priestley put his modified version of Hartley's theory at the service of a Unitarian philosophy of nature. Erasmus Darwin used Hartley's neurological mechanisms as the basis for his system of medical classification in *Zoonomia* (1794-6) and for his theory of evolution in the *Temple of Nature* (1803). In sociopolitical theory, the arguments advanced in William Godwin's *Political Justice* (1793) for inevitable human progress towards perfection were based on inferences from Hartley. The psychological, social, and political theories of James and John Stuart Mill and other English utilitarians were also based on Hartleian psychology and generalizations from it. In the nineteenth century, Hartley's fusion of corpuscular physics with empiricist epistemology and sensationalist psychology became reworked in more modern evolutionary terms to provide the foundations for theories in biology, neurophysiology, psychology, psychiatry, sociopolitical theory and to endorse a general faith in progress. Hartley's psychophysiological theory of learning underpins modern, evolutionary human science.

See also: Human nature, science of, in the 18th century; Utilitarianism

ROY PORTER

List of works

Hartley, D. (1730) *Conjecturae quaedam de sensu, motu, et idearum generatione*, London; 2nd edn, Bath, 1746; repr. in S. Parr (ed.) *Metaphysical Tracts*, London, 1837. (The first statement of Hartley's interests in the

physiological basis of perception and ideation.)
Hartley, D. (1749) *Observations on Man, His Frame, His Duty, and His Expectations*, London: Leake & Frederick, 2 vols.(Hartley's *opus magnum*, subsequently edited in 'bowdlerized' form by Joseph Priestley, sets out a materialist theory of the activity of mind, understood in terms of Newtonian corpuscular theory and a Providential reading of Christianity.)

References and further reading

Darwin, E. (1794-6) *Zoonomia*, London: J. Johnson, 2 vols; repr. New York: A.M.S. Press, 1974.
Darwin, E. (1803) *Temple of Nature*, London: J. Johnson; facsimile Menston: Scolar Press, 1973.
Gay, J. (1731) *A Dissertation concerning the Fundamental Principle and Immediate Criterion of Virtue*, London: W. Thurlborn.
Godwin, W. (1793) *An Enquiry Concerning Political Justice*, London: Robinson; revised 1796, 1798.
Halévy, E. (1928) *The Growth of Philosophic Radicalism*, trans. M. Morris, London: Faber & Faber; rev. edn, London: Faber & Faber, 1934.(By far the best survey of the radical intellectual tradition of which Hartley formed a part.)
Oberg, B.B. (1976) 'David Hartley and the Association of Ideas', *Journal of the History of Ideas* 37: 441-54.(The best brief account of Hartley's psychology.)
Smith, C.U.M. (1987) 'David Hartley's Newtonian Neuropsychology', *Journal of the History of the Behavioral Sciences*, 23: 123-36.(This offers a lucid exegesis of the largely Newtonian physical theories underpinning Hartley's science.)
Webb, M.E. (1988) 'A New History of Hartley's Observations on Man', *Journal of the History of the Behavioral Sciences* 24: 202-11.(Insists upon the importance of the iatrophysical medical theories of the early eighteenth century as a major context for Hartley's doctrines.)
Webb, M.E. (1989) 'The Early Medical Studies and Practice of Dr David Hartley', *Bulletin of the History of Medicine* 63: 618-36.(Informative on Hartley's medical career.)
Young, R.M. (1972) 'David Hartley', *Dictionary of Scientific Biography* 6, New York: Scribner, 138-40.(An exposition of Hartley's views which stresses the debts of later theorists to him.)
Young, R.M. (1973) 'Association of Ideas', *Dictionary of the History of Ideas*, ed. P.P. Wiener, New York: Scribner, 111-18.(A fine account of the post-Lockean tradition of associationism which stresses the major part played by Hartley.)

Hartmann, Karl Robert Eduard von (1842-1906)

Eduard von Hartmann was born in Berlin and lived there for most of his life. He was a prolific writer of both scholarly and popular works on a wide variety of topics, including aesthetics, ethics, religion and politics. He forecast a gloomy future for Germany, insisting that the only way to rectify the problems of modernity was with a strong nationalistic Germany ruled by an educated elite.

After distinguishing himself as a student Hartmann joined the military, intending to become an officer in the Prussian army. However, a severe knee injury frustrated these plans. During his convalescence he began a two-year study of music and painting. Believing he lacked the ’creative genius’ necessary for art he began an independent study of philosophy. In 1864 he began writing his best known work *Philosophie des Unbewußten (Philosophy of the Unconscious)*. When this three-volume work appeared in 1869, it made Hartmann an overnight success. Some ten thousand copies were sold between 1869 and 1875. The eleventh edition was published posthumously.Hartmann’s widow compiled a chronological overview of his writings with 400 entries. Hartmann died in Berlin.

Hartmann is perhaps best known as a philosophical pessimist, and is thus usually regarded as a disciple of Schopenhauer. Indeed Hartmann acknowledged that Schopenhauer exerted an early influence on his philosophical development. However, he insisted that he was a follower of Kant, whom he regarded as the true father of pessimism. Hartmann offered his own version of pessimism in his many writings, but perhaps the best statement of his position is found in his *Philosophie des Unbewußten*. Most of the first volume deals with Hartmann’s scientific and mathematical justifications for his belief that not only humans but all animals are prompted to actions by various unconscious motives; the ‘Will’ is the cause of everything that comes to be. He creditsKant with having discovered unconscious motives, but faults him for not having sufficiently appreciated the unconscious workings of the mind. In Hartmann’s view, Kant shared this deficiency with most of the figures in the history of philosophy.Schelling and Hegel barely improved uponKant’s philosophy. It was Schopenhauer who began to grasp the fundamental importance of the unconscious, what he called the ‘Will’. Unfortunately,Schopenhauer was too blinded by Eastern influences to appreciate sufficiently how the unconscious worked in the human realm. (ForHartmann’s view of his predecessors see the foreword to the tenth edition.)

There are, Hartmann maintained, three stages of the illusion regarding the state of happiness. In the first stage we believe that happiness is already attained in this life, yet there is far more pain than pleasure. In the second we believe that we will attain happiness in an afterlife. But happiness must be the concern of this life, and thus the second stage is a necessary transitional stage to the third stage, in which we believe in the progress of mankind. However, this concern with a future paradise on earth is also based on an illusion. Philosophical maturity comes in the realization that happiness is illusory.

Hartmann’s pessimism did not lead him to agree with Schopenhauer that this was the worst of all possible worlds; instead he thought thatLeibniz was actually closer to being correct in the belief that it was the best. However, he strongly rejectedLeibniz’s belief that we should consider evil to be a privation. Instead evil is manifest in this world. But it is part of the world’s teleological nature, the striving towards the final completion of the world, which means its utter and total annihilation. Again, despite some similarities with Schopenhauer,Hartmann believes that this outcome is not tragic. Rather, it is the culmination of the strivings of the unconscious through the entire human race - the end will be the highest expression of the unconscious in human nature.

Hartmann’s main debt is to Kant, whom he credits with emphasizing the subjective workings of the mind. According to Hartmann,Kant’s metaphysics and epistemology are primarily a doctrine of categories (see Kant, I. §6; Categories §3).Kant’s successors may claim to have continuedKant’s work, but Hartmann insists that his own philosophy represents the first major improvement onKant’s epistemology. In his *Kategorienlehre (Doctrine of Categories)* (1896a), Hartmann attempted to show that categories apply not only to the mind and nature but also essentially to what he called the ’metaphysical sphere’, that is, to the realm of the unconscious. He argued that there is not only intra-individual causality but also trans-individual causality, a kind of universal causality that organisms share at the biological level.

Hartmann sought to develop his theory of the unconscious and to apply it in many areas. Besides his attempt to reworkKant's categories, he also tried to build an ethics on his philosophy of the unconscious. He was also extremely interested in science, frequently using tables, graphs and calculations to illustrate his point. Yet he objected to the mechanical science of the day, occasionally singling out Darwin for special criticism.

He was, among other things, a historian of philosophy, a preoccupation he shared with many of his fellow nineteenth-century German philosophers. In his two-volume *Geschichte der Metaphysik (History of Metaphysics)* (1899-1900) he traced the role of metaphysics from the pre-Socratics to Nietzsche. He emphasized the place of metaphysics in the German philosophical tradition.

Hartmann's popularity waned immediately after his death, and today he is generally ignored. When he is remembered, it is primarily as one of the last of the nineteenth-century German speculative philosophers who sought to extend Kant's subjective philosophy by showing that our minds work not only at the conscious level but also at the unconscious level, thus prefiguring some aspects of modern psychology.

See also: German idealism

CHRISTOPHER ADAIR-TOTEFF

List of works

Hartmann, K.R.E. von (1869) *Philosophie des Unbewußten. Versuch einer Weltanschauung (Philosophy of the Unconscious: An Attempt at a World View)*, Berlin: Duncker, 11 editions; trans. W.C. Coupland, London: Macmillan, 1884; reprinted New York, Greenwood, 1972.(Hartmann's central work, and the best statement of his philosophical pessimism.)

Hartmann, K.R.E. von (1876) *Gesammelte Studien and Aufsätze (Collected Studies and Essays)*, Berlin: Duncker.(An early collection of essays.)

Hartmann, K.R.E. von (1876) *Phänomenologie des sittlichen Bewußtens. Prolegomena zu jeder künftigen Ethik (Phenomenology of Moral Consciousness. Prolegomena to any Future Ethic)*, Berlin: Duncker.(A copy of Kant's work, but intended to show how pessimism allows for ethics.)

Hartmann, K.R.E. von (1887) *Philosophie des Schönen. Zweiter systematischer Teil der Aesthetik (Philosophy of Beauty. Second Systematical Part of Aesthetic)*, Leipzig: Friedrich. (Hartmann's weak attempt at a theory of beauty and aesthetics.)

Hartmann, K.R.E. von (1888) *Lotze's Philosophie*, Leipzig: Friedrich; reprinted Ann Arbor, MI: Books on Demand, 1990.(An investigation of an influential thinker.)

Hartmann, K.R.E. von (1896a) *Kategorienlehre (Doctrine of Categories)*, Leipzig: Haake.(An attempt to show how we unconsciously apply categories.)

Hartmann, K.R.E. von (1896b) *Tagesfragen (Questions of the Day)*, Leipzig: Haake.(Discussion of contemporary issues - of little interest today.)

Hartmann, K.R.E. von (1899) *Ethische Studien (Ethical Studies)*, Leipzig: Haake.(Connects with his attempt to show that pessimism 'makes' ethics possible.)

Hartmann, K.R.E. von (1899-1900) *Geschichte der Metaphysik (History of Metaphysics)*, Leipzig: Haake; reprinted Darmstadt, Wissenschaftliche Buchgesellschaft, 1969.(A difficult and rather biased history of metaphysics.)

Hartmann, K.R.E. von (1885-1901) *Ausgewählte Werke (Selected Works)*, Leipzig: various publishers, 13 vols. (A huge collection, reviewing all aspects of Hartmann's writings.)

References and further reading

Bavarian Academy of Sciences Historical Commission (1966) *Neue Deutsche Biographie (New German Biography)*, Berlin: Duncker & Humblot, vol. 7: 738-40. (Contains biographical material.)

Caldwell, W. (1893) 'The Epistemology of Ed. V. Hartmann', *Mind* n.s. 11: 188-207. (Reviews and criticizesHartmann's thought.)

Caldwell, W. (1899) 'Moral and Social Philosophy. 1: The Positive Ethic', *Philosophical Review* 8 (5): 465-83. (Further criticism of Hartmann, focusing on his ethical theory.)

Caldwell, W. (1899) 'Moral and Social Philosophy. 2: The Metaphysic', *Philosophical Review* 8 (6): 589-603. (Caldwell's final review of Hartmann's central ideas, dealing with his writings on metaphysics.)

Copleston, F. (1962) *A History of Philosophy; Modern Philosophy: Schopenhauer to Nietzsche,* Garden City, NY: Image Books, vol. 7, 2: 57-9.(Gives a brief, readable account of Hartmann's thought.)

Hall, G.S. (1912) *Founders of Modern Psychology*, New York and London: D. Appleton; reprinted Ann Arbor: University Microfilms, 1978, 178-242.(A readable introduction by one of Hartmann's acquaintances.)

Hartmann, A. von (1912) 'Chronological Overview of the Works of Eduard von Hartmann', *Kant Studien* 17: 501-20.(The best bibliography of Hartmann's writings.)

Loemker, L. (1967) 'Hartmann, Eduard von', in *Encyclopedia of Philosophy*, ed. P. Edwards, New York: Macmillan, vol. 3, 419-21.(A broad account of the main themes ofHartmann's philosophy.)

Ringer, F.K. (1969) *The Decline of the German Mandarins. The German Academic Community, 1890-1933*, Cambridge, MA: Harvard University Press, 128-30.(A very interesting account of the German professorial community. Ringer chronicles the decline of their importance.)

Hartmann, Nicolai (1882-1950)

Nicolai Hartmann's intellectual trajectory was similar to that of his contemporary, Heidegger. He abandoned his early Neo-Kantian concern with knowledge and its foundations in favour of 'ontology', a study of the being of entities. Unlike Heidegger he assigned no ontological priority to human beings. Human beings are the highest level of entities, perched precariously above the physical, organic and animal levels, but conferring meaning and value on an otherwise meaningless and Godless universe.

Hartmann was born in Riga and, after a period as a medical student, studied history, classics and philosophy at the university of Saint Petersburg. In 1905 he moved to Marburg, and gained a professorship there in 1920. He later held chairs successively at Cologne, Berlin and Göttingen, where he died. For some years he remained under the spell of the Neo-Kantians, Cohen and Natorp, but by 1921, he had awoken from his 'critical slumber'. He rejected transcendental idealism, the transcendental ego, the primacy of practical reason and many other Kantian doctrines (see Kant, I.; Neo-Kantianism). Knowledge does not produce or alter what is known. Epistemology is not, as some Neo-Kantians believed, the sole preoccupation of philosophy. We must also consider being, the being of the objects known and of the knower.

Hartmann did not reject Kantianism *in toto*. Philosophers have constructed systems and they have also tackled problems, often reaching conclusions at odds with their chosen system. Systems must be rejected; they belong to the past. But problems are eternal, and philosophers have made enduring contributions to their solution. One such contribution was Kant's belief that our experience involves categories (see Kant, I. §6). Kant inferred that categories are subjective, supplied by us to our experience and inapplicable to things in themselves. Hartmann rejects this inference. Categories are involved both in our cognition and in things in themselves. On this basis, Hartmann constructed an ontology.

Entities form a hierarchy of levels. At the lowest level are physical entities, involving such categories as extension and causality. Above these are plants, which involve organic categories. Then, various forms of animal life, requiring such categories as consciousness and purpose. Finally there are human being with their social and cultural products, which Hartmann, following Hegel, calls 'objectivized spirit'. These levels or regions of beings are related in such ways as these. Lower entities need not constitute higher entities. There are, for example, purely material entities forming no part of a plant, animal or human being. Conversely, a higher entity necessarily involves the lower levels and the corresponding categories. A human being consists of matter, is a self-maintaining organism in the way that a plant is, and shares the basic features of animals. A higher entity is not determined solely by the laws governing the lower entities on whose presence it depends; it has considerable free-play. An organism consists of matter, but not the same material particles throughout its life; it organizes the matter of which it successively consists according to biological laws that do not govern purely material entities. Humans have free will: their decisions are not undetermined, but determined by factors peculiar to their spiritual life, not by the laws governing matter, plants or animals.

There are values, but they do not (as Kant supposed) depend on the legislation of the rational will nor are they all encompassed by the moral 'ought'. They constitute a realm of objective essences, as do the truths of logic and mathematics, and can similarly be discovered a priori. Values form a complex hierarchical system, and the realization of higher values depends on the realization of lower values, that is, nonmoral values and rudimentary moral values. One cannot properly aspire to sainthood unless one first fulfils one's familial and civic duties. Values can conflict: appropriate action in specific situations requires a combination of different and often conflicting virtues. Hartmann explores the relations between the ancient and the Christian conceptions of virtue.

Values are not realized independently of human activity nor is there any providential power to guarantee their ultimate realization. The nonhuman world is devoid of value and 'sense'. If it were not, human freedom would be restricted. We could not realize values that are already realized or frustrate the realization of values the eventual triumph of which is assured by God. Efficient causality does not threaten freedom, since a causal system or sequence can be diverted by external intervention. But final or teleological causality undermines freedom. Hence Hartmann rejects attempts, theological or otherwise, to find value, meaning and purpose in the nonhuman world. The human being is free, and alone actualizes values, not only moral values, but also aesthetic. A work of art, like

any other entity, is primarily a material thing. In it the artist realizes aesthetic value as a sort of superstructure. Hartmann explored the ontology of the work of art, as well as several other 'regional ontologies'. The priority of persons in the realm of value does not correspond to their ontological position. We are latecomers, resting on more fundamental levels of being.

Hartmann's work has much in common with that of other twentieth-century German philosophers. He believed, like Scheler, in objective, non-formal values, and, like Heidegger, that being is prior to knowledge and that philosophy goes hand in hand with history of philosophy. There are, however, differences. Scheler and Heidegger saw themselves as revolutionaries, changing the direction of philosophy. For Hartmann philosophy steadily advances in the solution of perennial problems. Philosophy is to be pursued for its own sake, not - as Heidegger sometimes suggests - as a prelude to life or *Existenz*. The 'subject' has no such ontological priority for Hartmann as 'Dasein' does for Heidegger. Hence Hartmann proceeds from lower ontological levels to higher, not, as Heidegger does, from the higher to the lower. For example, 'real time' is the uniformly flowing time in which physical objects and events occur; 'experience-time', the time of human awareness and agency, is located within real time but does not determine its nature. Hartmann lacks Heidegger's interest in *being as such*. This coheres with his other views. Our conception of being needs no radical overhaul; he largely accepts traditional categories, such as reality and value. He gives no priority to the human being, who (on Heidegger's view) needs an overall understanding of being in order to make possible the encountering of particular beings. Nor is he tempted to ask 'Why is there anything at all?' or 'Why are we in a world?' It is enough to describe systematically the wealth of beings we encounter.

Hartmann lacked sympathy with existentialism, and with such of its precursors as Nietzsche and Kierkegaard - though he had a close affinity to Hegel. Hartmann is overshadowed by his competitors, but has significant merits: an encyclopedic range, dispassionate rigour of argument, clarity of thought and style, and single-minded devotion to philosophy.

MICHAEL INWOOD

List of works

Hartmann, N. (1909) *Platos Logik des Seins (Plato's Logic of Being)*, Giessen: Töpelmann.(The first of many works on ancient philosophy.)

Hartmann, N. (1912) *Philosophische Grundfragen der Biologie (Basic Philosophical Questions of Biology)*, Göttingen: Vandenhoeck & Ruprecht.(The first of many works on the philosophy of nature.)

Hartmann, N. (1921) *Grundzüge der Metaphysik der Erkenntnis (The Basis of the Metaphysics of Knowledge)*, Berlin: de Gruyter, 2nd edn. 1925.(This work marks Hartmann's break with Neo-Kantianism.)

Hartmann, N. (1923) 'Aristoteles und Hegel', *Beiträge zur Philosophie des deutschen Idealismus* 3: 1-36.(An influential comparison of Aristotle and Hegel.)

Hartmann, N. (1923, 1929) *Die Philosophie des deutschen Idealismus (The Philosophy of German Idealism)*, vol.1 *Fichte, Schelling und die Romantik*, vol. 2 *Hegel*, Berlin: de Gruyter.(One of the best works on the subject, combining scholarship with critical acumen.)

Hartmann, N. (1924) 'Diesseits von Idealismus und Realismus' (This Side of Idealism and Realism), *Kantstudien* 29: 160-206.(Hartmann's critique of Kant.)

Hartmann, N. (1926) *Ethik*, Berlin: de Gruyter; trans as *Ethics*, London: Allen & Unwin, 1932, 3 vols. (Hartmann's main work on value was reviewed by J.S. Mackenzie in *Mind* 42 (1933-4):217-37.)

Hartmann, N. (1933) *Das Problem des geistigen Seins (The Problem of Spiritual Being)*, Berlin: de Gruyter.(This deals with objectivized spirit.)

Hartmann, N. (1934) 'Sinngebung und Sinnerfüllung' (Sense-giving and Sense-fulfilment), *Blatter fur deutsche Philosophie*, 8: 1-38.(Explains how we confer meaning on our own lives and on the world.)

Hartmann, N. (1935) *Zur Grundlegung der Ontologie (On the Foundations of Ontology)*, Berlin: de Gruyter. (Hartmann's first major work on general ontology.)

Hartmann, N. (1936) 'Der philosophische Gedanke und seine Geschichte' (Philosophical Thought and its History), *Abhandlungen der Preussischen Akademie der Wissenschaften: Phil.-hist. Klasse* 5: 975-81.(Explains the distinction between systems and problems in the history of philosophy.)

Hartmann, N. (1938) *Möglichkeit und Wirklichkeit (Possibility and Actuality)*, Berlin: de Gruyter.(An influential work.)

Hartmann, N. (1938/39) 'Zeitlichkeit und Substantialität' (Temporality and Substantiality), *Blätter fur deutsche*

Philosophie 12: 1-38.(Presents a non-Heideggerian account of time.)

Hartmann, N. (1940) *Der Aufbau der realen Welt (The Structure of the Real World)*, Berlin: de Gruyter.(Presents the general theory of categories.)

Hartmann, N. (1942) 'Neue Wege der Ontologie', in *Systematische Philosophie*, ed. N. Hartmann, Stuttgart and Berlin: W. Kohlhammer, 199-311; trans. as *New Ways of Ontology*, Chicago, IL: Regnery, 1953.(A clear, brief account of Hartmann's ontology.)

Hartmann, N. (1950) *Philosophie der Natur (Philosophy of Nature)*, Berlin: de Gruyter. (Hartmann's last work on the philosophy of nature.)

Hartmann, N. (1953) *Ästhetik*, Berlin: de Gruyter.(A posthumous work on aesthetics.)

Hartmann, N. (1955-7) *Kleinere Schriften (Shorter Writings)*, Berlin: de Gruyter, 2 vols. (These volumes contain Hartmann's diverse journal articles.)

References and further reading

Feyerabend, P.K. (1963) 'Professor Hartmann's Philosophy of Nature', *Ratio* V (1), 91-106. (An interesting piece by a well-known philosopher of science.)

Heimsoeth, H. and Heiss, H. (1952) *Nicolai Hartmann: Der Denker und sein Werk (Nicolai Hartmann: The Thinker and his Work)*, Göttingen: Vandenhoeck & Ruprecht. (Fifteen essays by German philosophers, with a bibliography of works by and about Hartmann.)

Scheler, M. (1927/8) 'Idealismus - Realismus', *Philosophischer Anzeiger* 2: 255-324; repr. with additions in *Späte Schriften (Late Works)*, Bern: Francke, 1976, 183-340; trans. as 'Idealism and Realism' in M. Scheler, *Selected Philosophical Essays*, Evanston, IL: Northwestern University Press, 1973, 288-356.(Criticizes Hartmann's metaphysics of knowledge, and compares him with Heidegger.)

Scheler, M. (1929) *Philosophische Weltanschauung (Philosophical Worldview)*, Bonn: Cohen.(Considers the anti-teleological atheism of Hartmann's *Ethics* together with other 'philosophical worldviews'.)

Stegmuller, W. (1970) *Main Currents in Contemporary German, British and American Philosophy*, Bloomington, IN: Indiana University Press.(Contains a good, clear account of Hartmann's thought.)

Hasidism

Its name literally meaning pietism, Hasidism is a mystical renewal movement that originated in eastern Europe in the mid-eighteenth century. It has become one of the most important spiritual and social developments of Orthodox Judaism and has exerted an influence as well on non-Jews and Jews who are not Orthodox. Early Hasidic leaders claimed their spiritual authority on the basis of heavenly revelations and mystical awakenings. But they generally differed from the more esoterically minded Kabbalists, from whom they drew their earliest following, in seeking to present the fruits of mystical inspiration to the community. Hasidic teachings fostered specific spiritual and ritual innovations, which gave outward expression to the profound nexus that the Hasidic masters saw between mundane existence and the inner, mystical meaning of God's law. According to Hasidic thinking, the divine and the human formed a single, all-encompassing unity, and it was on this basis that the Hasidic rabbis found in acts of Jewish piety means of linking divine experience with human responsiveness. Notable for its vitality and continuity in diversity, Hasidism continues its influence on religious Jewry and beyond to the present day.

1 Hasidic thought

Hasidism arises out of the tendency of Kabbalistic leaders to claim charismatic authority and mystical inspiration (see Kabbalah). The early Hasidic figures used such claims to set out original ideas about God and worship. They established new patterns of leadership based on a sense of immediate connection with the divine presence and the central relevance of that presence to all aspects of life. The founder of Hasidism, Rabbi Israel Ben Eliezer (1698-1760), known as the Ba'al Shem Tov or by the acronym BESHT, drew his following from the eastern European circles devoted to the Lurianic Kabbalah, a tradition of esoteric mysticism, ascetic seclusion, spiritual devotion and reliance on the efficacy of ecstatic prayer. The spiritual life organized by these charismatic leaders was routinized in a social structure grounded upon mystical leadership but capable of profound impact on broad circles of followers. The Ba'al Shem Tov sought to bring a similar structure to a much wider public.

Hasidic thought rests on the assumption that all things in this world are imbued with divine vitality. A hidden core of divine essence sustains all outward existence. God is omnipresent and manifest in all dimensions of existence, permeating every human act, every thought, and every material object (see Omnipresence). Physical objects, seen in the light of the infinite, are illusory mantles, mere vessels for the divine presence (*shekhinah*). Reality, then, is a spiritual core enveloped by a physical exterior, a divine unity disguised within the multiplicity of corporeality. The linked but opposing visages of the internal and external - the hidden divine substance and its material manifestation - condition one another. For the divine essence cannot be revealed except in its corporeal manifestation, but material existence has no actuality without the sustaining divine essence.

The matrix from which Hasidic worship springs is this assumption that the divine essence (*'ayin* - literally, 'nothingness') preconditions every mundane and spiritual phenomenon (*yesh* - literally, 'being') and that beyond perceived reality (*yesh*) abides a hidden truth (*'ayin*). Reality, then, is a dialectical whole composed from being and nothingness, in which all things are united with their opposites and combined with their inversions. The opposing elements annihilate, garb, sustain, transform and change one another from being into nothingness and from nothingness into being. The fullest expression of each spiritual element resides within its manifestation through inversion. The truth of reality is the struggle to achieve this reversal, abstraction and restoration of each thing to its source.

Hasidism strives to establish a connection both with the hidden divine element which vitalizes manifest concrete reality and with the divine presence beyond the material garment. Thus the paramount demand of Hasidic worship is for constant realization of the overwhelming presence of the divine and the acknowledgement of the inherent unity of opposites. Hasidic literature reflects this interest and is replete with expressions indicating the contemplative consciousness to which the *hasid* (devotee) is to aspire: *devekut* (literally, 'cleaving') devotion in study and contemplation of the incongruity between the divine essentiality and its physical manifestation is sought through the perception of the spiritual inwardness that lines material reality. Such awareness can lead to ecstatic union with the divine presence. *Bittul hayesh* - nullification of existence - is contemplation of divine reality through its corporeal garment, stripping away corporeality. *Avoda be-gashmiut* - service in corporeality - is the daily obligation to illuminate the mundane with its hidden divine essence. *Ha'alat nizozot* - literally, 'raising

sparks' - is elevation of divine elements in one's consciousness in order to assess the true meaning of *yesh* and *'ayin*. All of these notions are adapted from the Kabbalah, but now qualified to accommodate the new theology.

2 The Hasidic ethos and worship

Hasidic efforts to achieve contemplative consciousness by abolishing sensory experience are called the subordination of the eyes of the flesh to the eyes of the intellect. The goal is realization of the ultimate unreality of material existence. Only the all-pervasive divine presence is ultimately real. Existence and nothingness are no longer determined by empirical criteria but through contemplation. The transformation of human consciousness implies an obligation to lay bare the divine in all things and discover the unity behind the specious multifariousness of appearances. The separate existence of all things in thought must be nullified. The ultimate goal is proximity to God.

To remove all encumbrances to awareness, Hasidism proposes an ethos based on equanimity, worldly abnegation, and 'the nullification of existence'. Human thought is at the forefront of the spiritual struggle, since its creative and transformative powers and its capabilities of unifying opposites mark it clearly as the divine element in man. Although amorphous in itself, thought has the divine power to penetrate to the truth of things and illuminate the consciousness concealed and imprisoned within tangible externalities, to remove the barriers separating the divine presence from human contact, and so, ultimately, to achieve union with God.

The spiritual impetus that Hasidism founded upon transformative thought brought about a renewed appraisal of the means and ends of spiritual activity. Spiritual ascension, for example, had been pursued by Kabbalistic practitioners through penitential mortification and the cultivation of lofty spiritual and moral virtues arranged on a mystical ladder that led to God. This scheme was now discarded in favour of direct contact with a divine omnipresence that was freely accessible to all. Hasidic theology demanded only that human consciousness be directed towards devotion to God and proximity to his presence through contemplation of the dual visage of existence.

Hasidic dialectical assumptions made the mystic's focus on the inner, spiritual dimensions of religious life into a matter of public, communal concern. As one popular Hasidic saying has it, 'Because God is present in every place and human thought is present in every place about which one thinks, every act, every time and every place can serve as the point of departure for every man's contemplation of the divine inwardness of reality that lies beyond its physical garments'. In the Hasidic idiom this claim is finely honed by the radical pietist slogan 'In all your ways, know Him'.

The most conspicuous conclusion that Hasidism draws from this thought is an expansion of worship to include the most mundane and secular actions, by virtue of the thought and intentionality that illuminate and accompany them. 'Worship through corporeality' means that any mundane act may become divine service, if performed with the proper intention. The sanctification of the mundane and the conversion of all aspects of life to divine service become core features of Hasidism. But these reflect the centrality of thought and intention in Hasidism. The claim for divine immanence and omnipresence, reiterated throughout Hasidic literature, entails an unfettered but also unbounded human obligation to seek and attain this immanence. Herein lies the deep conflict with the basic worldview of the Kabbalah. For not only is the Kabbalistic hierarchy denied but the very boundaries between heavenly and earthly existence are obliterated. The transcendental Godhead has been displaced by an immanent and accessible divinity, and the wall between esoteric and exoteric has been removed, allowing the attainment of divinity, in principle, by anyone (see God, concepts of).

These changes had far-reaching social ramifications. Worship now extends far beyond the realm of accepted convention and common Kabbalistic interpretation. The often conventional reliance on acts alone as a standard of religious commitment is displaced as intention and consciousness move to the fore. Further, the assumption that the divine presence is the core within all fosters a substantial expansion of the social circles for whom a spiritual life on the highest order is accessible. All is God, and God is everywhere. Everyone, without exception, should approach the divine in all possible ways. The Hasidic mystical leader and his community become counterparts, envisaged through the polarity of *yesh ve-'ayin*. Detailed guidance in the worship of God and a sense of guardianship for the individual and the whole Jewish people led to the formation of a new spiritual agenda and new social affinities.

Opposition to the Hasidic movement was not slow to appear, focused on principled reservations to the doctrine of immanence but also aware of the broader significances of that idea. Controversy, infighting, schism, social conflict, denunciations to the ruling authorities of Russia and Poland, and a mountain of polemic marked the spread of Hasidism. The controversy, like the Hasidic ideas themselves, can still be felt within the Orthodox Jewish community today. On the other hand, Hasidism proved to be a source of inspiration to such thinkers as Martin Buber.

The spiritual self perception of Hasidism is best represented in its voluminous homiletic literature, known as *Sifrut ha-Drush*, emerging in 1780 and continuing to the present. This literature shows the diversity and creativity of Hasidic mysticism and sheds light on its social ramifications. The major innovations of Hasidism are also represented in short tracts from the end of the eighteenth century, known as *Sifrut ha-Hanhagot*, which are abridgements of the homiletic literature. The controversial position of Hasidism is reflected in the polemical literature known as *Sifrut Hasidim ve-Mitnaggdim* and *Sifrut ha-Haskalah*, representing the controversies with Orthodox ‘Opponents’ of Hasidism on the one hand and with the exponents of the Jewish Enlightenment of the other (see Enlightenment, Jewish). The unique Hasidic social structure and the major role of the *Zaddik* (Hasidic mystical leader) are reflected in the Hasidic tale, a literary genre known as *Sifrut ha-Sipurim ha-Hasidit*.

See also: Buddhist concept of emptiness; God, concepts of; Kabbalah; Mysticism, nature of

RACHEL ELIOR

References and further reading

Elior, R. (1988) ‘Between Yesh and Ayin: The Doctrine of the Zaddik in the Works of Jacob Issac, the Seer of Lublin’, in A. Rapoport-Albert and S. Zipperstein (eds) *Jewish History: Essays in Honour of Chimen Abramsky*, London: Peter Halban, 393-455.(A study of the mystical foundations of Hasidic social innovation.)

Elior, R. (1993) *The Paradoxical Ascent to God: The Kabbalistic Theosophy of Habad Hasidism*, Albany, NY: State University of New York Press.(A study of Lubavitch Hasidism, its mystical ideas and religious practice, in the late eighteenth and early nineteenth centuries.)

Elior, R. (1996) ‘The Paradigms of Yesh and Ayin in Hasidic Thought’, in A. Rapoport-Albert (ed.) *Hasidism Reappraised*, London: Littman Library.(An introductory essay on the foundations of Hasidic mysticism.)

Elior, R. (1997) *Israel Ba’al Shem Tov: Between Magic and Mysticism*, Jerusalem: Am Oved.(Comprehensive study of the BESHT and the innovative aspects of his mystical teaching.)

Etkes, I. (1988) ‘Hasidism as a Movement: The First Stage’, in B. Safran (ed.) *Hasidism, Continuity or Innovation*, Cambridge, MA: Harvard University Press.(Concise historical reappraisal of the emergence of Hasidism.)

Green, A. (1981) ‘Hasidism: Discovery and Retreat’, in P.L. Berger (ed.) *The Other Side of God*, Princeton, NJ: Princeton University Press.(A concise discussion of the major spiritual issues in Hasidism.)

Heschel, A.J. (1989) *The Circle of the Ba‘al Shem Tov*, Chicago. IL: University of Chicago Press.(A study of the group that initiated the Hasidic movement in the first half of the eighteenth century.)

Loewenthal N. (1990) *Communicating the Infinite*, Chicago, IL: University of Chicago Press.(A study of the methods of propagating mystical ideas to broad circles in Lubavitch Hasidism in the nineteenth century.)

Rapoport-Albert, A. (1979) ‘God and the Zaddik as Two Focal Points of Hasidic Worship’, *History of Religion* 18 (4): 296-325.(A study on the Hasidic Tzaddik and his mystical and social role.)

Schatz-Ufffenheimer, R. (1968) *Hasidism as Mysticism, Quietistic Elements in Eighteenth Century Thought*, Princeton, NJ: Princeton University Press.(Comprehensive discussion of the mystical elements in Hasidic thought.)

Scholem, G. (1949-50) ‘Devekut and Communion with God’, in *Review of Religion* 14: 115-39.(A study of a major Hasidic mystical concept.)

Weiss, J. (1985) *Studies in Eastern European Jewish Mysticism*, Oxford: Oxford University Press.(Essays on various aspects of Hasidism, integrating phenomenological and historical approaches.)

Hayek, Friedrich August von (1899-1992)

An Austrian-born British economist who turned political philosopher, Hayek was best known for his critique of socialism and the modern welfare state. Writing as an avowed classical liberal (he repudiated the label 'conservative'), he attempted to develop his account of the market as a mechanism facilitating economic coordination into a more general theory of law and politics.

Hayek's liberalism was first formulated as a response to totalitarianism, which he regarded as a tendency manifest in the regimes of Nazi Germany and Stalinist Russia, and inherent in proposals for central planning in society. This was the basis of his early opposition to socialism and his theory of limited constitutional government under the rule of law. The development of his political thought, however, saw him become increasingly critical of government and its interventions in the spontaneous evolution of society. Society was a 'spontaneous order' and not the product of human design. The threat to this order or civilization came from mankind's mistaken confidence in reason's capacity to take control of social processes to shape society in accordance with particular ideals. Socialism, as well as proposals for social justice, he regarded as variants of this tendency, which he labelled 'constructivist rationalism'. This social philosophy was underpinned by a philosophy of science which emphasized the subjective character of the data of the social sciences.

1 Life

Coming from a distinguished Austrian family of scientists and academics (he was a cousin of Ludwig Wittgenstein), Hayek entered the University of Vienna after serving in the First World War. Graduating with doctorates in law (1921) and political science (1923), he became a legal consultant in the civil service and, later, director of the Austrian Institute for Business Cycle Research. After a period as *Privatdozent* in Political Economy at the University of Vienna, Hayek accepted an invitation to lecture at the London School of Economics in 1931, and subsequently was appointed Tooke Professor of Economic Science and Statistics.

In the 1930s Hayek engaged in several important controversies in monetary economics, capital theory and business cycle research, taking issue with John Maynard Keynes (§3) and Pierro Sraffa. He also played a leading role in debates over the possibility of economic calculation under socialism. With Ludwig von Mises, Hayek argued against economists such as Oskar Lange that, once prices in a market economy were recognized as carriers of information, it had to be conceded that central planning was impossible: command economies had no mechanisms for coordinating economic knowledge.

When he became a British citizen in 1938, Hayek's reputation as an economist rivalled that of Keynes. Yet it was the publication in 1944 of *The Road to Serfdom* that brought him to the attention of a wider public. Warning that economic planning would subvert the institutions of a free society, the book become a controversial bestseller and established Hayek's reputation as a critic of government economic intervention at a time when arguments for central planning were in the ascendant.

In 1950 Hayek moved to the University of Chicago as Professor of Social and Moral Sciences on the Committee on Social Thought. Most of the following decade was devoted to work on his political treatise, *The Constitution of Liberty* (1960), although he also published an important collection of the J.S.Mill-Harriet Taylor correspondence, and edited a collection of papers on *Capitalism and the Historians* (1954). In 1962 he retired from Chicago to the University of Freiburg and then, in 1967, to the University of Salzburg. In the 1970s he continued his work on the philosophy of law in his trilogy, *Law, Legislation and Liberty* (1973-9). Interest in his work revived following his award of the 1974 Nobel Prize for Economics. In the 1980s Hayek set out upon another ambitious project: a three-volume refutation of socialism. A single volume, *The Fatal Conceit: The Intellectual Error of Socialism*, appeared in 1989, three years before his death.

2 Economics and social philosophy

Hayek belonged to the Austrian school of economics. He studied under Friedrich von Wieser, but his work also bears the influence of Carl Menger and Ludwig von Mises. Hayek's economic theory offers a general account of economics as a coordination problem. The economic problem of society is not how to allocate given resources; it is, rather, 'how to secure the best use of resources known to any of the members of society, for ends whose relative

importance only these individuals know'. Hayek's theory of capital, and his work on trade cycles and on money, attempt to account for failures of economic coordination, particularly over time. He rejected Keynes' economics which, he thought, ignored the temporal structure of production and so did not properly comprehend how market processes facilitated economic coordination over time. More generally, he objected to neoclassical equilibrium analysis which, he held, conflated individual choice at one point in time with market processes over time. Such models made assumptions about the knowledge possessed by economic agents which it was the task of economics to uncover.

Methodologically, these views have their origin in the Austrian school's subjective theory of value, which holds that value, far from being an inherent property of any object, is conferred upon goods by the subjective preferences of agents. Austrians such as Mises thus concluded that economics was an a priori or apodictic science: economic laws were deductions from a few basic axioms about human action. Hayek, however, rejected the radical subjectivism implicit in this conception of economic theory. In 'Economics and Knowledge' (1937) he articulated an empirical conception of the nature of economics as a body of testable theories; yet at the same time he rejected general equilibrium models of analysis.

The problem with such models, he thought, was that they were unable to explain how equilibrium (or a state of coordination of different individual purposes) was achieved. This was because the theory of perfect competition at their core ignored the adjustment process required to reach equilibrium. The standard assumptions of economic theory accepted as given: (1) complete knowledge on the part of economic agents; (2) the absence of costs or restraints upon the movement of goods and prices; and (3) a homogeneous commodity being supplied and demanded by many sellers and buyers. These assumptions were untenable because the economic world was a world in motion - one in which the process of coordinating economic plans took place despite the ignorance of economic agents, the existence of transaction costs and the subjective character of the individual valuation of commodities. Hayek thus rejected Lionel Robbins' influential interpretation of economics as the science that studied the allocation of scarce means among competing ends. This view presumed something false: that knowledge of those means and ends was given.

At the heart of Hayek's economics is a view of the nature of knowledge. This has two important aspects. First, Hayek distinguished two kinds of knowledge: theoretical or articulated knowledge, and tacit or unarticulated knowledge. Theoretical knowledge exists in the form of explicit statements and is the product of conscious reflection. Tacit knowledge, however, is unarticulated; it exists in the skills, instincts and dispositions of individuals and is also embodied in institutions and practices. (This distinction encompasses Gilbert Ryle's distinction between 'knowing that' and 'knowing how'; see Ryle G..) Most knowledge, Hayek thought, was of the latter sort; and its existence was also, therefore, dependent upon context or local circumstances. Second, Hayek argued that much of human knowledge is not only dispersed but also, because it is unarticulated, is incapable of being collected.

This view had an important bearing upon Hayek's economic argument, particularly in the socialist calculation debates. Here Hayek extended Mises' analysis of the impossibility of rational calculation under socialism. Mises had criticized socialist economics for not recognizing that economic value was subjective. Valuation could only take place in terms of units, yet it was impossible that there should ever be a unit of subjective use-value for goods. Judgments of value did not measure but established grades or scales, and, in the exchange economy, the 'objective exchange-value' of commodities became the unit of economic calcuation. In a monetary exchange economy *money* was the good used as the unit in which exchange-values are defined. While the value of money might fluctuate as its value relative to other goods changed, monetary calculation enabled us to judge the relative values of all goods and so to make production plans over time.

Hayek accepted Mises' argument that, without a pricing mechanism (given the defects of calculation in terms of labour rather than money), there could be no economic calculation. He pressed this critique of socialist economics further, however, by arguing that the fundamental problem was not calculational but epistemological. In 'The Use of Knowledge in Society' (1945) he argued that the problem to be solved when seeking to construct a 'rational economic order' is not a logical problem to be overcome using all relevant information concerning preferences and the factors of production. The data from which economic calculus begins are never given complete to a single mind which could work out their implications. The nature of the problem of a rational economic order is:

'determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess' (Hayek 1945) (see Socialism §2).

This division of knowledge Hayek thought to be the central problem of economics as a social science. However, investigating this problem meant moving away from the study of 'the economy' as a fixed domain marked by clear boundaries. Economies, in the strict sense in which households or enterprises can be called economies, consist of a complex of activities by which a given set of means is allocated in accordance with a unitary plan among the competing ends according to their relative importance. But the domain of human exchange encompasses not a single economy. What is commonly called a social or national economy is in fact a network of many economies. While such an order shares with an economy proper some formal characteristics, it does not share the most important one: its activities are not governed by a single hierarchy of ends. A better term for the extended order of human cooperation, Hayek suggests, is 'catallaxy': an order brought about by the mutual adjustment of many individual economies in a market. The task of economics is to investigate this order; but this also means that economics, ultimately, involves a more general investigation of social institutions.

Hayek's economics is therefore part of a larger social philosophy describing the extended order of society. The key to this account is his concept of 'spontaneous order'. For Hayek, an order is: 'a state of affairs in which a multiplicity of elements of various kinds are so related to each other that we may learn from our acquaintance with some spatial or temporal part of the whole to form correct expectations concerning the rest'. Order can be created by forces outside the system or it may be created from within as an equilibrium is generated by the interaction of elements whose natures impel them towards stable formations. This distinction is between an order that is 'made' and one that is formed spontaneously given the existence of particular elements in an environment. In the social world, the mind, language, law, and indeed society itself, should be understood as spontaneous orders.

In viewing society as a spontaneous order Hayek emphasizes that it is not the product of human design. While it is the result of human intention, it evolves without anyone controlling its development. This is partly because the knowledge possessed by any individual is so limited that no one could understand more than a fraction of the social environment. But it is also because not only society but also the individual is in a state of continuous evolution. Human nature - and mind and reason themselves - are not fixed but evolve with society (see Economics, philosophy of §3).

The important conclusion Hayek draws from this social philosophy is that individual reason is incapable of comprehending the social order in its entirety. Therefore, reason should not overreach itself by attempting to control or redirect the development of society. Such attempts would not only fail in their own terms but also threaten the progress of civilization, and the future evolution of reason itself. This concern about the 'abuse of reason' is the basis of Hayek's political philosophy.

3 Political philosophy

In his politics, Hayek defined himself as a classical liberal. He rejected socialism, but also denied that he was any sort of a conservative.

Hayek's liberalism stems from his social philosophy. His objection to socialism was that its aims, while noble, were unattainable. Its hope to bring about a planned society, organized in accordance with some common purpose, in which social forces were brought under human control, was in Hayek's view incapable of being fulfilled. The cost of the attempt, he argued, would be not only material impoverishment but also the gradual loss of individual liberty. The greatest danger to modern civilization came from the alliance of socialism with nationalism, which threatened to destroy the 'extended order of human cooperation' by drawing boundaries closing off societies in separate, controlled states.

Liberalism, for Hayek was a philosophy opposed to socialism and to nationalism. Its starting point was an understanding of society as a 'spontaneous' order. Any plausible answer to the question of what are the best social and political arrangements for human beings must be based upon this understanding. Hayek's answer is that human relations should be governed by arrangements which preserve liberty, with liberty understood as 'independence of the arbitrary will of another'. His political philosophy thus argues that a liberal society is one governed by the rule of law, and that justice is served only if law delimits the scope of individual freedom. In

short, liberalism favours a 'free' society in which individual conduct is regulated by rules of justice, so that all may pursue their own purposes in peace (see Liberalism §2).

The ideal of equality has a place in this scheme only in so far as Hayek concedes that the aim of the struggle for liberty has been equality before the law. Individual differences do not justify different treatment. However, this will inevitably lead to inequality in the actual positions people occupy and to material inequality. In Hayek's view, the state is not justified in using coercion to try to reduce or eliminate such inequalities. His objection is not to equality as such but to any attempt to impose a chosen pattern of distribution.

For the same reason, he objects to attempts to distribute goods according to 'merit' (see Desert and merit). If the principle of reward according to merit were institutionalized, we would end up with attempts to control remuneration which would in turn require even more controls upon human activity. For Hayek, this would not make for a free society since authority would decide what individuals were to do and how they were to do it.

The fear of this outcome is also the basis of his rejection of demands for equal distribution based upon membership of a particular community or nation. Membership of a national community does not, in Hayek's liberalism, confer entitlements to any share of national wealth. The result of recognizing any such entitlement would be that, rather than admit people to the advantages that living in their country offers, a nation will prefer to keep them out altogether.

In his early political writings, which include *The Road to Serfdom* and *The Constitution of Liberty*, Hayek argued that liberalism required a constitutional regime of limited government (see Constitutionalism §2). Government's essential function is to enforce the abstract rules of the society applying equally to all, although it is also charged with the task of preserving the market order. Beyond these, government also has a legitimate role to play in the administration of the resources of society to render some services to citizens, and also to provide a welfare safety net for the poor. It is not, however, any part of government's role to secure a particular distribution of income or wealth. This argument was developed at greater length in *The Mirage of Social Justice*, volume 2 of *Law, Legislation and Liberty*, which rejected the idea of social justice as not only unattainable but also 'meaningless'.

Yet while *The Constitution of Liberty* conceded a substantial - if limited - role for government in a constitutional regime, Hayek became progressively more critical of government as his political thinking developed. In *Law, Legislation and Liberty*, he began to criticize government's extensive legislative powers. This was an aspect of the development of his philosophy of law, which became increasingly critical of all forms of legal positivism (see Legal positivism). It also ushered in some interesting speculations about constitutional reform, although these sat uncomfortably with his growing scepticism about any kind of intervention in the spontaneous order of society by government.

This was the source of the fundamental difficulties in Hayek's political thought. As an intellectual who was genuinely concerned for the plight of post-war civilization, Hayek was a tireless advocate, arguing about the appropriate role of government and the conduct of policy. Yet his philosophical starting point was an emphasis upon human ignorance and the limits of reason, and the impossibility of humans controlling their social development. When, and for what reason, government could legitimately intervene in the processes of social evolution was thus difficult to establish.

Hayek's thought none the less is of interest as a liberal political philosophy. It is the most ambitious attempt in the twentieth century to develop a comprehensive liberal system of ideas. In its emphasis upon the danger of central control of knowledge, Hayek's thought also offers a liberalism whose core is strikingly postmodern.

See also: Economics, philosophy of; Liberalism

CHANDRAN KUKATHAS

List of works

Hayek, F.A. von (1937) 'Economics and Knowledge', *Economica* 4 (new series): 33-54.(Hayek's first articulation of his economic theory.)

Hayek, F.A. von (1944) *The Road to Serfdom*, London: Routledge & Kegan Paul.(Hayek's polemical first political work, arguing against central planning.)

Hayek, F.A. von (1945) 'The Use of Knowledge in Society', *American Economic Review* 35 (4): 519-30.(On how

to create a rational economic order in society.)
Hayek, F.A. von (1948) *Individualism and Economic Order*, Chicago, IL: University of Chicago Press.(Contains important economic essays, as well as his contributions to the socialist calculation debates.)
Hayek, F.A. von (1952) *The Counter-Revolution of Science*, Glencoe, IL: Free Press.(Develops Hayek's philosophy of science.)
Hayek, F.A. von (ed.) (1954) *Capitalism and the Historians*, Chicago, IL: University of Chicago Press.(A collection of revisionist essays defending capitalism, introduced by Hayek)
Hayek, F.A. von (1960) *The Constitution of Liberty*, London: Routledge.(A statement of Hayek's liberalism containing his theory of the rule of law.)
Hayek, F.A. von (1973-9) *Law, Legislation and Liberty*, London: Routledge, 3 vols.(Presents Hayek's later legal philosophy and also his arguments against social justice.)
Hayek, F.A. von (1989) *The Fatal Conceit: The Intellectual Error of Socialism*, London: Routledge. (The last statement of Hayek's critique of socialism.)

References and further reading

Barry, N. (1979) *Hayek's Social and Economic Philosophy*, London: Macmillan.(An expository account of Hayek's social and economic thought.)
Gray, J. (1986) *Hayek on Liberty*, Oxford: Blackwell.(A comprehensive review of Hayek's system of ideas with an extensive bibliography.)
Kukathas, C. (1989) *Hayek and Modern Liberalism*, Oxford: Oxford University Press.(A critical examination of Hayek's liberalism.)

Heaven

In Christian theology, heaven is both the dwelling place of God and the angels, and the place where all who are saved ultimately go after death and judgment to receive their eternal reward. The doctrine of the resurrection of the body requires that heaven be a place because it must contain the glorified bodies of the redeemed, but heaven is more theologically important as a state than as a place. This state is traditionally described as involving the most intimate union with God without the elimination of the individual human personality (the beatific vision); it is a state of perfect bliss beyond anything possible on earth. In high medieval theology, the happiness of heaven is understood to be so great that it is even beyond the capability of human nature to enjoy without divine aid. There are varying views on the nature of heavenly society, however, with some theologians (Augustine, Aquinas, Bonaventure) arguing that perfect happiness will be derived from the love of God alone, while others (for example, Giles of Rome) stress the joy that will be derived from the company of the elect. More recently, interest in the nature of heaven has declined, and Christian theology has tended to play down its importance.

1 Heaven in the Bible

The Hebrew word for heaven or the heavens (*samayim*) means ‘that which is above the earth’; in particular, it refers to that which is above the firmament (*raqia*). This, in the book of Genesis, is a solid vault in the visible sky holding back the waters beyond (Genesis 1: 8; Psalms 148: 4-6); it contains windows that open to let rain fall on the earth (Genesis 7: 11; 2 Kings 7: 2, 19), and the stars are suspended from it (Genesis 1: 14). The heavens are God’s handiwork, but are also the special domain of God himself. While it is vicious pride that leads human beings to attempt to reach heaven, as in the tower of Babel (Genesis 11: 1-9), the Hebrew tradition stresses the idea that while God is transcendent, he is also active in the history of his people and comes down repeatedly to intervene in human affairs; he is, in fact, omnipresent (Psalm 139: 8-10). In the New Testament, the ministry of Jesus begins with the heavens opening (Matthew 3: 16) in order that salvation might descend to earth. Christ returns to his Father in heaven (John 6: 62; 13: 1) and prepares a place for his followers (John 14: 3). The Ascension of Jesus into heaven (Mark 16: 19; Luke 24: 51; Acts 1: 9) led Christians to await the return of Christ, when both heaven and earth will be transformed (Isaiah 65: 17; Romans 8: 19-23; 2 Peter 3: 13; Revelation 21: 1). The Book of Revelation contains vivid imagery of resurrection, judgment and the City of God after the Second Coming.

2 Heaven in salvation history

In Hebrew thought, the belief that human beings were raised to heaven after death was not widespread. It was distinctive of Christian faith to interpret the crucifixion of Christ as a victory over death as well as over sin. Belief in a victory over death included not only belief in eternal life beyond the grave, but also belief in a resurrection of the body. This doctrine was predicated on the resurrection of Jesus from the tomb and was connected with the early Christian concern with the end of history and the expectation of the Second Coming (the *Parousia*); at this time, Christ would return in glory, the dead would be raised, all would be judged, and those who were saved would be taken to heaven to enjoy eternal glory. Heaven is therefore the culmination of salvation history. But it has always been a part of Christian belief that even before the general resurrection at least some of the redeemed will be with Christ, although in traditional Roman Catholic teaching some are detained for a time in purgatory. It is possible, then, to enjoy the glory of heaven without the resurrected body, so it must be a place/state for which bodies are not necessary. Still, Aquinas (§10) maintained that the enjoyment of heaven after the resurrection of the body will be greater than in the disembodied state. He argued that this must be the case because the operation of the soul united to a body is more perfect than the operation of the separated soul. Since happiness consists in an operation, he reasoned, the soul’s happiness after bodily reunion will be greater than before (*Summa theologiae* IIIa, q.93, a.1, corpus). Heavenly bliss is therefore enhanced by the union of the human soul with the resurrected body (see Eschatology §3).

Glorified bodies are said by Aquinas to have the qualities of impassibility, subtlety, agility and clarity (*Summa theologiae* IIIa, qq.82-5). Impassibility means that the glorified body does not suffer and does not need to preserve itself from wear or harm from within or without. Aquinas identifies subtlety with the power to penetrate. This was sometimes interpreted as the ability to pass through other bodies, but Aquinas identifies it with the perfect subjection of the body to the soul (*Summa theologiae* IIIa, q.83, a.1, corpus). Agility allows the glorified body to

move about unimpeded and with great rapidity. Clarity makes the glorified body luminous and brilliant to behold even to the non-glorified eye.

In Roman Catholic theology, there is no official position on the spatial characteristics of heaven or its relationship to the physical universe, in part because of the problem of explaining how it can be the abode of non-physical beings. However, some fundamentalist Christians interpret the Bible as referring to heaven as a physical place. The Latter Day Saints (Mormons) believe that the Second Coming will usher in a thousand-year reign of Jesus on earth which will involve the physical transformation of the planet into a Garden of Eden. This belief in a thousand-year reign after the Second Coming (millenarianism) has appeared from time to time in Christian history, in such groups as the Gnostics, the Montanists, the Anabaptists and the Pietists. Millenarianism never dominated the Christian tradition, but many theologians, including Aquinas, maintained that after the general resurrection and last judgment the earth will be transformed in a way analogous to the glorification of the human body (*Summa contra gentiles* IV, ch. 96). Since it was generally thought that the renewed earth would become part of heaven, it followed that heaven, or some part of heaven, has at least quasi-spatial features. On the other hand, Aquinas thought that the transformed physical universe will have no temporal features, but will be permanent and changeless. At the end of time, the earth and all heavenly bodies will stand still (*Summa theologiae* IIIa, q.91, a.2). Heaven itself is a static state to which nothing can be added and nothing taken away.

3 The beatific vision

The essential element in the state of heavenly glory is the beatific vision, which was defined by Aquinas as the intuitive vision of God in his essence. The idea is inspired by Paul: ‘We see now through a glass darkly, but then face to face’ (1 Corinthians 13: 12). This doctrine is theologically important because it indicates that the enjoyment of God in heaven is not only a historical end, or the aim of the divine plan of human salvation, but constitutes the fulfilment of the nature of each individual human person. Adopting Aristotle’s teleological view of nature, Aquinas describes human beings as incomplete and in a state of movement towards an end that fulfils that for which they were designed. Like Aristotle (§21), Aquinas identifies the end of human life as happiness; happiness is what human beings by nature desire. But unlike Aristotle, Aquinas maintains that true happiness is not possible in this life (*Summa theologiae* IaIIae, q.5, a.3). He interprets the desire for happiness as a natural but obscure awareness in each human person that their goal is union with the supreme good. Since the supreme good is God, it follows that there is a natural but obscure desire for union with God, and this union is the supernatural end for which every human being was created. Aquinas thus combines Aristotelian metaphysics with the Christian theology of Augustine (§11), whose famous prayer ‘Our hearts are restless until they rest in thee, oh Lord’ expresses the failure of the fulfilment of human desires in this life.

The doctrine of the beatific vision raises interesting questions about the nature and limits of human knowledge. Aquinas was clearly concerned with the problem of whether a finite created being can actually know an uncreated infinite deity. He decided that it is beyond the nature of created beings to do so, but supernatural help enhances the human and angelic intellects to make this possible. Aquinas calls this aid ‘illumination’. In the beatific vision, Aquinas says, God himself is in direct contact with the human mind; this contact is not via images or concepts, which are the only ways human beings can have knowledge otherwise (*Summa theologiae* IIIa, q.92, a.1). As infinite truth, God alone is able to satisfy fully the human and angelic intellects, which are made for the possession of truth. As infinitely desirable, God alone is able to satisfy the desires of the human heart or the angelic will. The end for which human beings were created is thus a gratuitous gift of God, one which is fully satisfying, but which human beings would be incapable of enjoying without supernatural aid. The blessed in heaven will not all enjoy the beatific vision in equal measure, however. There will be degrees of happiness and union with God corresponding to the degrees of merit of persons at death (*Summa theologiae* IIIa, q.93, a.2-3).

The secondary object of the beatific vision is our continuing knowledge of the beings and events of creation. These will be known through the vision of the divine essence in whom all things are known. This raises a series of questions about the importance of contingent relationships to human happiness. Since human beings in heaven are self-conscious egos, it is reasonable to think that they do not lose their memories of the past or their attachment to particular persons. But what if certain persons loved while on earth are absent from heaven? Aquinas says that no unhappiness would result from this, since all delight is in God and in God there can be no attachment to evil or to evil persons. But as long as it is admitted that there is enjoyment in the company of the elect, it is reasonable to

think that there would have been greater delight had certain loved ones been saved. Furthermore, it is difficult to believe that the blessed in heaven would want to observe the sufferings of the damned. Yet Aquinas not only maintains that they do, but that this enhances their bliss:

> Everything is known the more for being compared with its contrary…. Therefore in order that the happiness of the saints may be more delightful to them and that they may render more copious thanks to God for it, they are allowed to see perfectly the sufferings of the damned.
>
> (*Summa theologiae* IIIa, q.94, a.1, corpus)

4 The society of heaven

McDannell and Lang (1988) raise the problem of how happiness deriving from the particularity of human relationships is related to happiness deriving from the vision of God. They claim that there are two strands in the history of the idea of heaven in Christian thought, one theocentric and one anthropocentric, and that the two are in tension with each other. The dominant theocentric strand either ignores or at least downgrades social joys in heaven. Bonaventure, like Augustine, rejected any idea of individualized friendships in heaven. Likewise, Aquinas says that there will be no sexual love or enjoyment of food in heaven (*Summa contra gentiles*, ch. 83) - a view contrasting with the Islamic heaven - and that friendship is not necessary for perfect happiness: ‘But if we speak of perfect Happiness which will be in our heavenly Fatherland, the fellowship of friends is not essential to Happiness, since man has the entire fullness of his perfection in God’ (*Summa theologiae* IaIIae, q.4, a.8). Aquinas implies, though, that while the love of friends is unnecessary, it will still exist as derived from the love of God: ‘And so if there were but one soul enjoying God, it would be happy, though having no neighbour to love. But supposing one neighbour to be there, love of him results from love of God’ (*Summa theologiae* IaIIae, q.4, a.8, ad.3).

The other strand identified by McDannell and Lang emphasizes the joy of the company of the elect. This strand has appeared from time to time throughout the Christian era. While Aquinas was somewhat reticent about the extent of social life in heaven, his student Giles of Rome was not, and he speculated that the saints form a heavenly *societas perfecta*. During the nineteenth century, a conception of heaven as full of earthly delights, including all the pleasures of social relations, replaced the theocentric conception in the popular imagination and represented the apex of the anthropocentric view of heaven. Although this conception of heaven has largely been discarded in theology (with the exception of Mormon theology), it has not usually been replaced by a well-developed theocentric theology. With a few exceptions (for example, Hick 1976), heaven is not a significant topic in contemporary theology. Occasionally, however, some interesting new problems about heaven have been raised. Examples are the problem as to whether the unchangeability of heaven would make it boring, and the problem of there seeming to be a significant amount of luck involved in getting there. This is because part of what gets a person to heaven is what one does during life, but what one does during life is partly determined by such things as the kind of upbringing one has, the circumstances in which one makes choices, and even the timing of one’s death (see Zagzebski 1994). There is also, of course, the traditional dispute over predestination (see Predestination).

5 The decline of heaven in contemporary theology

There has been a decline of interest in Christian theology regarding what were traditionally called the ‘last things’: death, judgment, the resurrection of the body, and heaven and hell. In the Roman Catholic Church, this may be partly due to embarrassment over the centuries of teaching the faithful to accept their lot in life, no matter how poor, on the grounds that the important thing was the next life. Heaven is mentioned only briefly in the documents of Vatican II, and the influential Jesuit theologian Karl Rahner said late in life that we have but one life and no continued existence after death (1975). Liberal Protestant theologians have little to say about heaven, and when it is mentioned, it is usually described in minimalist terms as a state which concerns our contact with God in the afterlife. There has been more speculation on the existence of hell; and since the traditional theology of heaven is the counterpart to the theology of hell (the former being associated with reward and the latter with punishment) changes in the conception of the latter will require some changes in the former.

See also: Faith; Heaven, Indian conceptions of; Hell; Justification, religious; Limbo; Purgatory; Resurrection; Salvation; Soul, nature and immortality of the; Tian

LINDA ZAGZEBSKI

References and further reading

Aquinas, T. (1266-73) *Summa theologica*, trans. Fathers of the English Dominican Province, Westminster, MD: Christian Classics, 1981, IaIIae, qq.1-5; sup. qq.91-6.(A very influential account of heaven.)

Aquinas, T. (1259-65) *Summa contra gentiles*, trans. A.C. Pegis, J.F. Anderson, V.J. Bourke and C.J. O'Neil, Notre Dame, IN: University of Notre Dame Press, 1975, chaps 82-7, 92, 95-6.(One of the fullest and most influential treatments of the doctrine of heaven in Christian theology. The work of Aquinas is particularly important for those who wish to investigate the connection between the idea of heaven and philosophically important medieval views in ethics and metaphysics.)

Arendzen, J.P. (1951) *Purgatory and Heaven*, New York: Sheed & Ward.(A short and simple summary of traditional Roman Catholic teaching on judgment, the resurrection of the body, heaven and purgatory.)

Garrigou-Lagrange, R. (1952) *Life Everlasting*, trans. P. Cummins, St Louis and London: B. Herder Book Co.(A good explanation of Roman Catholic theology on the traditional last things - death, judgment, heaven, hell and purgatory - by an important theologian of the era preceding Vatican II.)

Hick, J. (1976) *Death and Eternal Life*, New York: Harper & Row.(An influential book by an important Christian theologian. A discussion of both Western and Eastern eschatologies is included.)

McDannell, C. and Lang, B. (1988) *Heaven: A History*, New Haven, CT: Yale University Press.(A useful and interesting history of the idea of heaven in Western religion, with discussions of its treatment in imaginative literature and art, as well as theology and philosophy; not theologically technical.)

Rahner, K. (1975) *Theological Investigations*, trans. D. Bourke, London: Darton, Longman & Todd.(Rahner was a leading twentieth-century Catholic theologian who attempted to work out a new teaching on heaven that did not include the idea of a continued personal life.)

Zagzebski, L. (1994) 'Religious Luck', *Faith and Philosophy* 11 (3): 397-413. (Argues that certain Christian doctrines make the problem of moral luck, which has received a lot of attention in contemporary ethics, particularly critical for Christian moral theology. The focus of the problems is the doctrine of an eternal heaven and hell, particularly the latter.)

Heaven, Indian conceptions of

*Heaven is an important part of Indian religious cosmology and also figures strongly in Indian philosophical discourse. In the cosmologies of the early period of Indian thought, from the Ṛg Veda to the advent of the so-called heterodox schools of Jainism and Buddhism (*c.*1500-500 BC), heaven was conceived of in relatively simple terms, as a happy and permanent abode for both the deceased (particularly those who performed Vedic sacrifices) and the gods. In all three of the major religious systems of classical India, namely Hinduism, Buddhism and Jainism, heavenly realms expanded in number, and were usually depicted in terms of the deities who inhabited them and the prevailing life-situations of the deceased. Geographically, the heavens of classical India were envisaged as either parallel, occupying separate, bounded, horizontal space, or vertical, existing on separate tiers of upper-level space. Philosophically, the existence and soteriological value of heaven were much debated, from the earliest Upaniṣads (*c.*500 BC) and philosophical* sūtras *of Vedānta and Pūrva Mīmāṃsā (the Brahmasūtra of Bādarāyaṇa and the Mīmāṃsāsūtra of Jaimini respectively) to the later logicians and systematizers of the standard philosophical systems (*c.*1500 and beyond). Religion and philosophy were always intimately linked in India, indeed were often barely distinguishable; thus the soteriology of a religious form would be logically supported and fully explicated by allied philosophical schools. Broadly, two points of view were represented: first, that heavenly realms did exist and were achievable for one's ultimate benefit, or, second, that they existed but were phenomena subject to decay, rather like ordinary reality, or (as the Buddhists thought) were mental constructs. In the second case, heaven was regarded as inferior to variously conceived states of enlightenment or liberation.*

1 Heaven in the Vedas

In the religious and philosophical schools of India, notions and descriptions of heaven follow one of two paths: heavens are either permanent abodes of bliss and happiness based on earthly models, or they are (along with various hells) temporary states where good (or bad) karma (Pāli, *kamma*) is consumed during an all-but-eternal cycle of births and deaths (*saṃsāra*). In the Vedas and most of the religious and philosophical traditions that follow in their wake, the word that comes closest to 'heaven' is *svarga*, from *s(u)var* ('light', 'the sun', 'the realm of celestial light') and *ga*, from the Sanskrit verbal root *gam* ('to go'). Thus the etymological meaning of *svarga* is 'going to the realm of celestial light' (see Gonda 1966: 73-106). Other words are used occasionally in post-Vedic literature, such as *devaloka* or *devālaya*, 'the realm (or abode) of the gods'. *Svarga*, more explicitly, was an abode of religious merit identified with the celestial counterpart of the Vedic sacrificial ritual, located in the highest part of the vault which comprised the tripartite Vedic universe: the earth, mid-region and celestium. *Svarga* was a ritually replicated embodiment of 'Vedic man's desire to live in a pure, strong and perfect "sphere" or "world"' (Gonda 1966: 73). It was connected to this world not just ideologically but physically. The *Ṛg Veda* (10.90.3-4) states that three feet of the archetypal cosmic being (*puruṣa*) obtain immortality in the divine celestium (*amṛtaṃ divi*), while the fourth foot 'is all beings'. This inspired the common Indian trope of the inverted tree, with its many roots in the sky and its leaves on earth, an image which envisages a connection of the world of the gods and the deceased with that of humanity (see Coomaraswamy 1938-9).

In order to grasp fully the significance of the idea of heaven in ancient and classical India, it is important to understand at least the broad outlines of the metaphorical uses of *svar* and *svarga* in the earliest Vedic texts. In the *Ṛg Veda* itself (7.90.6), the patrons of the sacrifice bring *svar*, 'celestial light', to the poets, presumed to take the form of worldly payment for services rendered. Gonda concludes: 'The ideas of sun, sunlight, celestial light are inextricably mixed up with those of well-being, good fortune, happiness, glory' (Gonda 1966: 78; see also Kirfel 1967: 22-).

Svarga is often thought of as the 'yonder world' to which the ritually perfected Vedic sacrificer travels after death to live in a state of happiness or bliss. It is the ultimate abode of the one 'who knows thus', the one who has acquired the esoteric meaning or connections of ritual performance. It is important to understand that this is not just armchair speculation or meditative cognition: ritual action and intent were considered essential to realizing and entering into the celestial region. This 'yonder world' was often identified as both the realm of the gods (*devalokaḥ*) and the realm of the deceased ancestors (*pitṛlokaḥ*), though these were just as often distinguished. *Svarga* was variously described as infinite (*ananta*), immeasurable (*aparimita*), undefined (*anirukta*), whole or complete (*sarva*, hence also identified with the year), immortal (*amṛta*), and as the sacrifice itself (*svargo vai loko*

yajñaḥ, Kauṣītaki Brāhmaṇa 14.1). It was thus every bit as much an abstraction or symbol, an ideal of earthly life, as a paradisiacal abode of the dead. Indeed, the Vedas do not contain florid descriptions of *svarga*. It is merely described as full of pleasures and resplendent. Occasionally it is said to be in the east or northeast; 'up from the earth', according to two Vedic texts (Aitareya Brāhmaṇa 7.5, Śatapatha Brāhmaṇa 3.5.1.9). It is difficult to attain. One mounts or ascends to *svarga* as if on a chariot or ship. 'Becoming a hawk one flies to it' (Taittirīya Saṃhitā (5.4.11.1). The sacrifice, as a ship, was oriented towards the east, for when the priest 'walks towards the east, he steers that [ship] eastward toward the heavenly realm [*svargaṃ lokam*]. By this he attains the heavenly realm' (Śatapatha Brāhmaṇa 2.3.3.16). Similarly, the eastern sacrificial fire (*āhavanīya*) is regarded as the world of heaven, so the sacrificer who is anointed near the *āhavanīya* attains this realm (Śatapatha Brāhmaṇa 9.3.4.11-13). Often *svarga* is associated with Brahman, which in the middle Vedic literature indicates the power of expansion that expresses all things superior in the Vedic universe. Thus this heavenly realm has both geographical and existential status: it is both 'a soteriological "world of heaven" (*svarga loka*) and a cosmological "world of the sky" (*svar*, *dyaus*)' (Smith 1994: 127).

2 Heaven in classical Indian religious systems

Heavenly or celestial realms multiply dramatically - they expand vertically as well as horizontally, which is to say they are multi-tiered and parallel - and receive much more detailed treatment in Jainism and Buddhism than in the Vedic literature, as well as in the Purāṇas and the Hindu *bhakti* traditions that grew out of it and which continue to dominate the Indian religious landscape. Much of the discourse on heaven is intimately connected with notions of karma and retribution. The general rule is that if one performs good actions, one attains a pleasant celestial abode, pleasant in a measure corresponding to the merit accruing from the good actions; but the abode can equally be destructive or frightening as a result of an overflow of bad actions. For example, the Buddhist cosmologies speak of pleasant heavenly realms of sensual pleasures, called *kāmadhātu*; realms in which gods are liberated from sensuality, called *rūpadhātu*; immaterial realms in which one may continue spiritual work, called *arūpyadhātu*; the abode of the 'satisfied' Bodhisattvas, called *tuṣita*; and finally the 'happy universe' or 'pure land', *sukhāvatī*, where blissful creatures are born from lotuses in the presence of the Buddha Amitābha, Avalokiteśvara and certain Bodhisattvas (La Vallée Poussin 1926; see also Hirakawa 1990: 170-; Kirfel 1967: 190-).

The Hindu Purāṇas have, in theory at least, five characteristic features (*pañcalakṣaṇa*), one of which is describing the knowledge of the cosmos (*sarga*, 'the creation'). Included in this cosmology are various heavenly realms (usually seven), which are usually ideal and paradigmatic, and mappable according to both physical, social and mythic notions, and ideas of psychological symmetry, order and justice. One of the common images (not always satisfactorily mapped) is the descent of the Ganges (and other rivers) from heaven. The Ganges, an artery of celestial water, connects the world of humans and the comparatively proximate world of the dead (*pitṛloka*) with the heavenly fountainhead of the gods. Thus, immersion of physical remains in rivers, especially the Ganges, connects the deceased (and the consequent offerings to them by the living) with the higher realms (see Witzel 1984; von Stietencron 1972; Lüders 1951: 138-61, 269-193; Kirfel 1967: 59-, 109-). Heavenly worlds in Jainism are every bit as elaborate as their Hindu and Buddhist counterparts, are equally dependent on notions of karma produced, and are populated by several levels of gods and exalted beings (see Kirfel 1967: 285-; Caillat and Kumar 1981).

Despite their splendour, these higher or heavenly worlds are, in post-Vedic India, almost invariably regarded as impermanent, subject to decay. Indeed, perhaps the most prominent feature of heavenly worlds, at least for Indian philosophical schools, is their ambiguity. The virtue which leads to residence in them and the rewards resulting from such residence are decidedly transitory, yet they represent progress towards emancipation. But it is a peculiar kind of progress, in that it that can be fully realized only when the denizens of heaven, from ordinary mortal souls to the highest of the gods, take rebirth once again in human form. On the whole, the philosophical schools recognize heaven as a reward of properly intended and performed ritual action, action performed for the explicit purpose of otherworldly reward. But as the merit and the corresponding participation in such worlds become exhausted, heavenly beings begin to lose their lustre in preparation for a fall. But the fall is deceptive: although it is a fall from certainty and near immortality in pleasant realms into the uncertainty and suffering of this mortal world, it is necessary and in the interest of the highest good because final liberating knowledge and experience can only be gained from decisively confronting the visage of mortality.

3 Heaven in philosophical thought

In India, philosophy was almost invariably a handmaiden of religious doctrine. Thus, heaven, which was intimately associated with specific religious doctrines, became a point of contention in sectarian and inter-religious debate. In general, the classical schools of Indian philosophy argued against the ultimate value of the Vedic notion of heaven, a place to which deserving humans or even sacrificed animals ascended after death, regarding it as limiting and therefore inferior to the 'higher' goal of spiritual emancipation. Regardless, philosophers regularly appealed in establishing the truth of their respective positions to the authority or verbal testimony (*śabdapramāṇa*) of *Śruti*, which indicated either the Vedic corpus vaguely or the Upaniṣads specifically. The first question considered by the philosophical schools was the very existence of heaven. This question was taken up because heaven is invisible and must therefore be proved to exist (the Cārvāka, or atheist, school declared that it is invisible and thus nonexistent). To this end the philosophers resorted to the authority of scripture. For example, Uddyotakara (a sixth-century logician) says in his *Nyāyavārttika* that, based on descriptions in Vedic texts, heaven and the gods are perceptible, have specific location and exist for the sake of others, but are not eternal (see Potter 1977: 311). He declares that the Veda aims at heaven through ritual means, while self-knowledge can be attained only through understanding of the ontological categories presented in Nyāya-Vaiśeṣika, including, among others, the nature of substances, wholes, parts and proper reasoning (see Nyāya-Vaiśeṣika §§4-5).

The philosophical system that most naturally defends the Vedic notion of heaven is Pūrva Mīmāṃsā (see Mīmāṃsā). Śabara (*c.* second century AD) states in his commentary on Jaimini's *Mīmāṃsāsūtra* (6.1.1, *c.* second-third century BC) that *svarga* is a form of happiness, because that is the way people describe it. Only in a 'secondary figurative sense', he says, is it 'applied to the thing or substance that causes happiness' (Jha 1942: 270). It is a condition in which one is devoid of pain and in which the self alone is the object of consciousness. It is also, according to Pūrva Mīmāṃsā, the result of a desire that is fulfilled by ritual action. The connection between sacrificial action and later residence in heaven is established by the later philosophers of Pūrva Mīmāṃsā (beginning with Kumārila, seventh century) by the assumption of *apūrva*, 'a potency produced by the sacrifice which makes it possible that its fruits be reaped at a later time; it is a bridge between the actions and their promised results' (Halbfass 1980: 275).

The notion of desire, however, which most of the philosophical traditions associate with popular religion, is generally rejected as a causal factor in the acquisition of the highest state of existence. Most of the philosophical schools regard as naïve both the Vedic notion of heaven and the correlative notion of desire as a means to attain it, and replace them with more systematic paradigms of liberation and proper understanding of reality, usually accompanied with programmatic spiritual practice. Indeed, desire and sacrifice come to be regarded as antithetical to the 'higher' goal of enlightenment (*mokṣa*, *mukti*). The major (but partial) exceptions to this are some of the devotionally based schools of Vedānta, including, for example, the Śuddhādvaita of Vallabhācārya (1479-1531), in which passionate attachment to Kṛṣṇa, to be sure a form of desire, can lead to final residence and near partnership in the archetypal realm of Kṛṣṇa's creative matrix. Vallabhācārya regards this blissful condition as superior to the state of enlightenment in which the individual self is absorbed into the abstract absolute (see Vedānta §3).

One example of the classical Advaita Vedānta position is found in Śaōkara's (early eighth century) commentary on Bṛhadāraṇyaka Upaniṣad 3.2-3, which discusses the question 'What is this "death" that the liberated man is freed from?' Śaōkara concludes, against the views of his (rather unfairly represented) Pūrva Mīmāṃsā (opponent, that ritual action cannot lead to liberation but only to heaven, that residence in heaven is not possible after the body has returned to its constituent elements after death, that dissolution cannot result in abidance in any 'state'. Indeed, liberation is characterized by destruction of ignorance (*avidyā*), the underlying cause of the world, and this is only possible through proper understanding of the relationship between the self and the absolute. And, he adds, it does not require dissolution of the physical body (see Śaōkara).

One final position may be mentioned here which allows for both the Pūrva Mīmāṃsā (and the Vedānta views. This is the view of the grammarians (*vaiyākaraṇāḥ*). Nāgeśa (seventeenth century) states in his commentary on Kaiyaṭa's *Pradīpa* on the *Mahābhāṣya* of Patañjali (second century BC) that the study of grammar is useful for the attainment of both *svarga* and *mokṣa*; it results in correct understanding of Vedic texts, which enables one to perform rituals properly, leading to *svarga*, while it is also useful for understanding the Upaniṣads, which lead to *mokṣa*.

See also: Cosmology and cosmogony, Indian theories of; Heaven; Hindu philosophy; Jaina philosophy §1; Karma and rebirth, Indian conceptions of

FREDERICK M. SMITH

References and further reading

Caillat, C. and Kumar, R. (1981) *The Jain Cosmology*, New York: Harmony Books.(A 'coffee-table' book with a good introduction and excellent colour reproductions of Jaina paintings and manuscript pages.)

Coomaraswamy, A.K. (1938-9) 'The Inverted Tree', *Quarterly Journal of the Mythic Society* (New Series) 29: 111-49; in R. Lipsey (ed.) *Coomaraswamy*, vol. 1, *Selected Papers: Traditional Art and Symbolism*, Bollingen Series 89, Princeton, NJ: Princeton University Press, 1977, 376-404.(An accessible and intriguing account of this important image.)

Gonda, J. (1966) *Loka: World and Heaven in the Veda*, Amsterdam: North Holland Publishing Company.(A thorough, indispensible and highly philological account of heaven as it is represented in the early and middle Vedic literature.)

Halbfass, W. (1980) 'Karma, *Apūrva*, and 'Natural' Causes: Observations on the Growth and Limits of the Theory of *Saṃsāra*', in W. Doniger O'Flaherty (ed.) *Karma and Rebirth in Classical Indian Traditions*, Berkeley, CA: University of California Press, 268-302.(An accessible article on a difficult topic.)

Hirakawa, Akira (1990) *A History of Indian Buddhism from Śākyamuni to Early Mahāyāna*, ed. and trans. P. Groner, Honolulu, HI: University of Hawaii Press.(The section on Buddhist heavens is the most complete among general textbooks on early Buddhism.)

Jha, G. (1942) *Pūrva-Mīmāṃsā in its Sources*, Varanasi: Banaras Hindu University.(This remains unsurpassed as a guidebook to early Pūrva Mīmāṃsā.)

Kirfel, W. (1967) *Die Cosmographie der Inder*, Hildesheim: Georg Olms.(This indispensible book includes many important texts on 'heaven' from post-Vedic India, including Jaina, Buddhist and Purāṇic.)

La Vallée Poussin, L. de (1926) 'Abode of the Blest (Buddhist)', in J. Hastings (ed.) *Encyclopedia of Religion and Ethics*, Edinburgh: T. & T. Clark, vol. 2, 687-9.(A good, but very brief, account of the subject.)

Lüders, H. (1951) *Varuṇa*, vol 1, *Varuṇa und die Wasser (Varuṇa and Water)*, Göttingen: Vandenhoeck & Ruprecht.(An excellent sourcebook, with much on the subject, as Varuṇa is a 'heavenly' deity.)

Potter, K.H. (ed.) (1977) *Encyclopedia of Indian Philosophies*, vol. 2, *Indian Metaphysics and Epistemology: The Tradition of Nyāya-Vaiśeṣika up to Gaōgeśa*, Princeton, NJ: Princeton University Press, and New Delhi: Motilal Banarsidass.(The editor's introduction includes a succinct account of heaven according to this system of Indian philosophy. The summaries of the texts also provide information on the subject.)

Smith, B.K. (1994) *Classifying the Universe: The Ancient Indian Varṇa System and the Origins of Caste*, New York and Oxford: Oxford University Press.(This book provides considerable material on Vedic notions of heaven and their correspondences with the physical world.)

Stietencron, H. von (1972) *Gaōgā und Yamunā: zur symbolischen Bedeutung der Flussgöttinnen an indischen Tempeln (Gaōgā and Yamunā: On the Symbolic Meaning of the River Goddesses on Indian Temples)*, Wiesbaden: Harrassowitz.(In India, rivers are said to originate in celestial or heavenly realms. This excellent book recounts both the mythology and the artistic representations of this notion.)

Witzel, M. (1984) 'Sur le chemin du ciel' (On the Way to Heaven), *Bulletin d'Études Indiennes* 2: 213-77.(This innovative and philologically thorough article makes the case that the Vedic Sarasvatī is in fact the Milky Way.)

Hedonism

Hedonism is the doctrine that pleasure is the good. It was important in ancient discussions, and many positions were taken, from the view that pleasure is to be avoided to the view that immediate bodily pleasure is to be sought. More elevated views of pleasure were also taken, and have been revived in modern times. There are three varieties of hedonism. Psychological hedonists hold that we can pursue only pleasure; evaluative hedonists that pleasure is what we ought to pursue; reflective hedonists that it is what on reflection gives value to any pursuit. Arguments for psychological hedonism suggest that an agent's actions are a function of what they think will maximize their pleasure overall. Explaining altruism can lead such theories into truism. Similar arguments are used for reflective hedonism, and the same problem arises. The difficulty for evaluative hedonism lies in deciding how we can establish certain ends as desirable. The claim that pleasure is to be maximized seems immoral to many. Hedonism also faces problems with the measurement of pleasure.

1 History and varieties of hedonism

Hedonism is the doctrine that pleasure is the good (see Pleasure; Good, theories of the). It has usually been viewed as a doctrine of self-indulgence, and so morally suspect. In classical Greece the pre-philosophical picture of pleasure is given in the Myth of Prodicus, where Pleasure vies with Virtue in tempting Hercules. Among philosophers, some advocated pursuit of immediate bodily pleasures (see Cyrenaics); some, hoping to show disreputable pleasures to be detrimental to overall pleasure, took the maximization of pleasure over one's life as the good (their hope is helped if non-bodily pleasures are included) (see Democritus; Protagoras); some (such as Plato in *Republic* and Aristotle in *Nicomachean Ethics*) held that at least the good life is the pleasantest, and perhaps that, if you understood about pleasure, arguing from hedonistic premises would lead you to the morally right conclusion (see Plato; Aristotle); some held that pleasure is to be avoided (see Cynics). After Aristotle, Epicurus and his followers advocated a life of physical pleasure unadulterated by pain, and of unanxious memory and anticipation of such pleasures. For a number of centuries they were an influential school. The Stoics, by contrast, considered pleasure to be a false belief in the agent's wellbeing (see Stoicism).

In the Middle Ages the topic did not arouse such excitement, although it was treated in commentaries on Aristotle. In post-Cartesian philosophy Claude-Adrien Helvetius was a proponent, and many of the British empiricists adopted some form of hedonism. This was developed into utilitarianism by Jeremy Bentham, James Mill and John Stuart Mill (see Utilitarianism). J.S. Mill (1861) distinguished higher from lower pleasures, declaring the former, more admirable, pleasures pleasanter. Hedonism received judicious consideration from Henry Sidgwick and was vigorously attacked by G.E. Moore, but has benefited from sophisticated development at the hands of Richard Brandt (1979).

Psychological hedonists hold that we can pursue only pleasure; evaluative hedonists that pleasure is what we ought to pursue; reflective hedonists that it is what on reflection gives value to any pursuit. Some have tried to be all three at once, usually through failure to notice the differences. The varieties of hedonism in the first paragraph above will match with types in these three to produce different forms of hedonism. Further variations may be introduced as follows: (a) while the pleasure in question is usually the agent's own, it may be that of any human or sentient being (see Animals and ethics); (b) different views of the nature of pleasure will yield different pictures of the goal in life (see Eudaimonia). For example, if pleasure is the realization of one's own good functioning, we get one picture (see Perfectionism); if it is the satisfaction of desire, another (see Desire).

2 Psychological, evaluative and reflective hedonism

Whether this distinction holds depends on conclusions about the distinction between fact and value (see Fact/value distinction). In classical Greek philosophy it was standard practice to argue from the supposed fact that something is the only thing wanted (or really wanted) by an individual or species, to the conclusion that it is the good of that individual or species. To show that everyone wants only pleasure was thought necessary and sufficient to show that pleasure is the human good, and so what humans ought to pursue. If it is invalid to argue from facts about what people do want to conclusions about what they ought to want, then there will be a clear difference between holding that we all do want pleasure and holding that we ought to. Indeed, it will be a waste of time telling people that they ought to want pleasure if pleasure is the only thing they are capable of wanting. Proponents of evaluative

hedonism seem to suppose the falsity of the psychological variety.

Reflective hedonism is a halfway house. The thesis is that when we reflect on what is valuable in life, then only considerations of pleasure weigh; but in many of our pursuits we lose sight of this. Thus I may become overconscientious and in all my actions be ruled by considerations of duty; when I stand back and review my life, it is its dreary lack of pleasure which convinces me that it is all worthless. This is a psychological thesis about our reflective valuing which allows us to make sense of raising questions about the worth of what we actually pursue.

3 Arguments for hedonism

(a) Psychological hedonism. Arguments may be to the effect that all of every agent’s actions are a function of what they consider will maximize their pleasure overall, or will be the pleasantest thing in itself of the available alternatives. It will depend on how the end of pleasure is conceived. Apparent exceptions, such as altruistic acts, will have to be shown nevertheless to be a function of the goal as specified. Two points are important here. First, suggesting a way in which an agent might see an act as maximizing pleasure is not showing that they do: hypotheses have to be substantiated. Second, showing that an act is a function of what the agent thinks the most titillating in prospect of the available alternatives is not showing it to be a function of what they think will maximize pleasure overall. In other words, one has to ensure that one has evidence for the hypothesis, and to be sufficiently precise about its formulation to be sure that the evidence tells in favour of it and not some other view.

Faced with apparent counterexamples, such as altruism, there is a temptation to claim that if the agent does not envisage some pleasure from their altruism, or view it with pleasure, then they do not really want to act as they do. This saves the thesis but at the cost of making it truistic; pleasure can no longer illuminate why we desire anything: if pleasure is absent we do not count as desiring.

In these arguments, not only is it important to beware the slide into making the thesis truistic; one must also take care about the relations between various explanations in terms of pleasure. If I am depressed, I may recoil from the prospect of going to a party which I know I should enjoy if I made the effort. Evidence about what I view with pleasure does not seem relevant to the question of what I think will produce pleasure, or conversely. It is necessary to argue the connection.

(b) Evaluative hedonism. The issue here depends on the possibilities for establishing the desirability of ends. Is it intuitively obvious that in some sense pleasure is (the) good? Or by some account of rationality is it the only rational thing to pursue? Or does pleasure alone make sense of our moral concerns? The argument is not now about actual pursuits, nor about what counts as wanting, but about how one might argue that pleasure ought to be pursued or desired.

(c) Reflective hedonism. Points paralleling those on psychological hedonism hold. First, there are arguments as to whether all our reflective valuing does in the last resort rely on appeal to hedonistic considerations; second, there is the possibility of arguing that we count as evaluating only if we refer to hedonistic grounds. There is the same risk of lapse into truism.

4 Problems for hedonism

Hedonism may be aimed at supplying either a rationale for morality, or an alternative to it. Since most forms of hedonism judge the worth of courses of action by their consequences in terms of pleasure, they encounter the problems, as rationales of morality, of consequentialism (see Consequentialism): moral judgments of worth seem often partly or wholly independent of consequences, certainly of pleasure, however long-term (see Deontological ethics; Intuitionism in ethics). For egoistic forms of hedonism, there is the problem of explaining the apparently anti-egoistic bias of morality (see Egoism and altruism).

Whether considered as a rationale of morality or an alternative, hedonism has problems with the measurement of pleasure. It is not clear what to make of the question whether a given episode of skating gave as much pleasure as the supping of a glass of wine. They seem incommensurable. If these problems can be surmounted, there remain problems, for non-egoistic hedonists, of comparing different people’s pleasures.

See also: Asceticism; Economics and ethics; Eudoxus; Happiness; Moral motivation; Rational Choice Theory; Suffering

JUSTIN GOSLING

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985.(Contains two important discussions of pleasure, 1152b-1154b and 1172a16-1176a29, which attempt to reconcile the hedonist and anti-hedonist insights of predecessors.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, revised F. Rosen, Oxford: Clarendon Press, 1996. (Central modern account, which gives a qualified hedonistic basis for utilitarianism.)

Brandt, R.B. (1979) *A Theory of the Good and the Right*, Oxford: Clarendon Press. (Sophisticated contemporary discussion, which defends hedonism against modern criticisms.)

Gosling, J.C.B. (1969) *Pleasure and Desire*, Oxford: Clarendon Press.(Discusses the issues introduced in this entry, in §§2-4.)

Gosling, J.C.B. and Taylor, C.C.W (1982) *The Greeks on Pleasure*, Oxford: Clarendon Press.(Comprehensive discussion of the disputes in classical Greece up to Epicurus and the early Stoics.)

Mill, J.S. (1861) *Utilitarianism*, in J. Gray (ed.) *On Liberty and Other Essays*, Oxford: Oxford University Press, 1991.(Classic modern discussion of hedonistic utilitarianism.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press. (Contains spirited attack on hedonism, especially that of J.S. Mill.)

Plamenatz, J. (1958) *The English Utilitarians*, Oxford: Blackwell.(Treatment of some of the main figures in the British empiricist tradition.)

Plato (*c*.380-367 BC) *Republic*, trans. G.M. Grube, revised by C. Reeve, Indianapolis, IN: Hackett Publishing Company, 1992.(Seeks to justify life of reason as most pleasant.)

Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan; 7th edn, 1907. (Contains well-balanced and careful assessment of hedonism.)

Hegel, Georg Wilhelm Friedrich (1770-1831)

Hegel was the last of the main representatives of a philosophical movement known as German Idealism, which developed towards the end of the eighteenth century primarily as a reaction against the philosophy of Kant, and whose main proponents, aside from Hegel, include Fichte and Schelling. The movement played an important role in the philosophical life of Germany until the fourth decade of the nineteenth century. Like the other German Idealists, Hegel was convinced that the philosophy of Kant did not represent the final word in philosophical matters, because it was not possible to conceive a unified theory of reality by means of Kantian principles alone. For Hegel and his two idealistic predecessors, a unified theory of reality is one which can systematically explain all forms of reality, starting from a single principle or a single subject. For Hegel, these forms of reality included not only solar systems, physical bodies and the various guises assumed by organic life, for example, plants, animals and human beings, but also psychic phenomena, social and political forms of organization as well as artistic creations and cultural achievements such as religion and philosophy. Hegel believed that one of the essential tasks of philosophy was the systematic explanation of all these various forms starting from one single principle, in other words, in the establishment of a unified theory of reality. He believed this because only a theory of this nature could permit knowledge to take the place of faith. Hegel's goal here, namely the conquest of faith, places his philosophical programme, like that of the other German Idealists, within the wider context of the philosophy of the German Enlightenment.

For Hegel, the fundamental principle which explains all reality is reason. Reason, as Hegel understands it, is not some quality which is attributed to some human subject; it is, by contrast, the sum of all reality. In accordance with this belief, Hegel claims that reason and reality are strictly identical: only reason is real and only reality is reasonable. The considerations which moved Hegel to identify reason with reality are various. On the one hand, certain motives rooted in Hegel's theological convictions play a role. According to these convictions, one must be able to give a philosophical interpretation of the whole of reality which can simultaneously act as a justification of the basic assumptions of Christianity. On the other hand, epistemological convictions also have to be identified to support Hegel's claim that reason and reality are one and the same. Among these convictions belong the assumptions (1) that knowledge of reality is only possible if reality is reasonable, because it would not otherwise be accessible to cognition, and (2) that we can only know that which is real.

According to Hegel, although reason is regarded as the sum total of reality, it must not be interpreted along the lines of Spinoza's model of substance. Reason is rather to be thought of as a process which has as its goal the recognition of reason through itself. Since reason is the whole of reality, this goal will be achieved when reason recognizes itself as total reality. It is the task of philosophy to give a coherent account of this process which leads to self-knowledge of reason. Hegel conceived this process by analogy with the model of organic development which takes place on various levels. The basic presupposition governing the conception of this process is that reason has to be interpreted in accordance with the paradigm of a living organism. Hegel thought of a living organism as an entity which represents the successful realization of a plan in which all individual characteristics of this entity are contained. He called this plan the concept of an entity, and conceived its successful realization as a developmental process, in the course of which each of the individual characteristics acquires reality. In accordance with these assumptions, Hegel distinguished the concept of reason from the process of the realization of this concept. He undertook the exposition of the concept of reason in that section of his philosophical system which he calls the Wissenschaft der Logik (Science of Logic). In this first part of his system, the various elements of the concept of reason are discussed and placed into a systematic context. He presented the process of the realization of this concept in the other two parts of his system, the Philosophie der Natur (Philosophy of Nature) and the Philosophie des Geistes (Philosophy of Spirit). Apart from their systematic function, which consists in demonstrating reason in the Hegelian sense as total reality, both parts have a specifically material function in each case. In the Philosophy of Nature, Hegel aims to describe comprehensively all aspects of natural phenomena as a system of increasingly complex facts. This system begins with the simple concepts of space, time and matter and ends with the theory of the animal organism. The Philosophy of Spirit treats of various psychological, social and cultural forms of reality. It is characterized by the assumption of the existence of something like genuine, spiritual facts, which cannot be described as subjective states of individual persons possessing consciousness, but which have an independent, objective existence. For Hegel, examples of such facts are the state, art, religion and

history.

In spite of the relatively abstract metaphysical background of his philosophy, which is difficult to reconcile with common sense, Hegel's insights in his analysis of concrete facts have guaranteed him a permanent place in the history of philosophy. None the less, for contemporary readers these insights are interesting hypotheses, rather than commonly accepted truths. Of lesser importance among these insights should be counted Hegel's results in the realm of natural philosophy, which soon suffered considerable criticism from practising natural scientists. The important insights apply more specifically to the spheres of the theory of knowledge as well as the philosophy of right, and social and cultural philosophy. Hegel is thus regarded as an astute and original representative of the thesis that our conception of objectivity is largely determined by social factors which also play a significant role in constituting the subject of cognition and knowledge. His criticisms of the seventeenth- and eighteenth-century concepts of natural law and his thoughts on the genesis and significance of right in the modern world have had a demonstrable influence on the theory of right in juridical contexts. Hegel's analysis of the relationship and interplay between social and political institutions became a constituent element in very influential social theories, in particular that of Marx. The same applies to his central theses on the theory of art and the philosophy of religion and history. Hegel's thoughts on the history of philosophy made that topic a philosophical discipline in its own right. Thus Hegel was a very influential philosopher. That his philosophy has none the less remained deeply contentious is due in part to the fact that his uncompromising struggle against traditional habits of thought and his attempt to establish a conceptual perspective on reality in contrast with the philosophical tradition of the time remains characterized by a large measure of obscurity and vagueness. Unfortunately these characteristics also infect every summary of his philosophy.

1 Life and works

Hegel was born on 27 August 1770 in Stuttgart, son of a Württemberg official. In the autumn of 1788, after attending the local grammar school, he began a course of study at the Protestant Seminary in Tübingen in preparation for a career as a Protestant clergyman. Two of his fellow students and friends were F.W.J. Schelling and F. Hölderlin. In autumn 1793, after successfully completing this period of study, Hegel became a private tutor in Berne, Switzerland, and remained there until 1796. From January 1797 until the end of 1800 he was a private tutor in Frankfurt am Main, where he again came into contact with Hölderlin, who played an important role in the formation of Hegel's early philosophical convictions. Thanks to a legacy, Hegel was able to abandon his position as a tutor and pursue his academic ambitions. Early in 1801 he went to Jena. His student friend Schelling had become Fichte's successor and was lecturing in philosophy at the university there. With Schelling's energetic support Hegel qualified as a *Privatdozent* in the autumn of 1801 with a thesis on natural philosophy. Initially, Schelling and Hegel worked closely together, a fact which is documented by a philosophical periodical which they published jointly from 1802 (although it ceased publication following Schelling's departure from Jena in 1803). In 1805 Hegel was appointed Extraordinary Professor, but financial difficulties forced him to abandon his activities at the University of Jena in the autumn of 1806. A friend's intervention enabled him to take over as editor of a daily newspaper in Bamberg in March 1807. In November 1808 the same friend then ensured that Hegel was nominated rector and professor at a grammar school in Nuremberg. After a few years in this capacity, Hegel was able to return to university life. In 1816 he was called to the University of Heidelberg, which he left again in 1818 to take a chair at the University of Berlin, as Fichte's successor. There he revealed a considerable talent for academic teaching and succeeded in assuring a dominant position in contemporary discussions for his philosophical doctrines. Hegel died in Berlin during a cholera epidemic on 14 November 1831, at the height of his fame.

Hegel's works can be divided into three groups: (1) texts written by Hegel and published during his lifetime; (2) texts written by him, but not published during his lifetime; and (3) texts neither written by him nor published during his lifetime. Two texts from his early years in Frankfurt do not fit into this scheme. The first is the translation of a pamphlet by Cart, a Berne lawyer, on the political situation in the Canton of Vaud, which was translated and annotated by Hegel, and which he published anonymously in 1798. This is the first printed text by Hegel; the second is a fragment dating from the same period and known as the *Systemprogramm des deutschen Idealismus (System-Programme of German Idealism)*. The text has survived in Hegel's handwriting, but his authorship remains controversial.

The earliest writings in the first group date from the beginning of Hegel's time in Jena. His first philosophical

work is entitled *Differenz des Fichte'schen und Schelling'schen Systems der Philosophie* (*The Difference between Fichte's and Schelling's System of Philosophy*) (1801b). This was followed later during the same year by the essay which he had to submit in order to qualify as *Privatdozent*, *De Orbitis Planetarum (On the Orbits of the Planets)*. In 1802-3 Hegel published various philosophical works in the periodical which he edited with Schelling, the *Kritisches Journal der Philosophie (Critical Journal of Philosophy)*. The most important among these were *Glauben und Wissen (Faith and Knowledge)*, *Verhältnis des Skeptizismus zur Philosophie (The Relationship of Scepticism to Philosophy)* and *über die wissenschaftlichen Behandlungsarten des Naturrechts (On the Scientific Ways of Dealing with Natural Law)*. Immediately after his period as a university teacher in Jena and at the beginning of his period in Bamberg, Hegel published his first great philosophical work, the *Phänomenologie des Geistes* (*Phenomenology of Spirit*) (1807). During the eight years in which he taught at the grammar school in Nuremberg, Hegel published his three-volume *Wissenschaft der Logik* (*Science of Logic*) (1812, 1813, 1816). While in Heidelberg, the complete presentation of his system appeared for the first time, in his *Enzyklopädie der philosophischen Wissenschaften in Grundrisse* (*Encyclopedia of the Philosophical Sciences in Outline*) (1817), which was reprinted twice during his Berlin period in two completely revised editions (1827, 1830). Also during this period he published *Naturrecht und Staatswissenschaft im Grundrisse. Grundlinien der Philosophie des Rechts (Natural Law and Politics in Outline. The Principles of the Philosophy of Right)* (1821). Apart from these, Hegel published only minor writings during his lifetime. These were written partly in response to events at the time, although most articles were for the *Jahrbücher für wissenschaftliche Kritik (Yearbooks of Scientific Criticism)*, which he co-edited from 1827. Among these is his final published work, *Über die englische Reform-Bill (On the English Reform Bill)* (1831).

The second group of texts includes those works which were written by Hegel but not published by him. Almost all these texts first became accessible in a more or less authentic form during the twentieth century. They can again be divided into three groups. The first group consists of the manuscripts which Hegel wrote between the end of his time as a student and the end of his time in Jena. Among the most important are the so-called *Theologische Jugendschriften* (*Early Theological Writings*), which were published in 1907 at the instigation of Wilhelm Dilthey by his pupil, H. Nohl. Today they are known as Hegel's *Frühschriften (Early Writings)*. Further important texts from this period are the three *Jenaer Systementwürfe* (*Jena Drafts of a Philosophical System*), written between 1803 and 1806, partly for publication and partly as lecture notes. The second group of writings not published by Hegel consists of works produced during his period in Nuremberg. Hegel's first biographer, K. Rosenkranz, presented excerpts from these writings as the *Philosophische Propädeutik (Philosophical Propaedeutic)* (1840). In this text Hegel attempted to present his philosophical views in a form suitable for use within the framework of his grammar-school teaching courses. The third group of texts comprises manuscripts and notes which he wrote in connection with his lectures in Heidelberg and Berlin. They are partly contained in the editions in which his pupils and friends published his works after his death.

The third major group of texts covers those works which were neither written nor published by Hegel. They form almost half the texts contained in the first complete edition of Hegel's works. Among them one finds Hegel's extremely influential lectures on aesthetics, the philosophy of history, the history of philosophy and the philosophy of religion. In the form in which they have become influential, these texts are the product of students, in most cases representing the result of notes compiled during Hegel's lectures. Insufficient attention has been paid to this remarkable fact, that is, that some of Hegel's most influential texts actually have the status of secondhand sources.

The first complete edition of Hegel's work, published during the years 1832-45, proved to be influential but highly unreliable both from a historical and a critical point of view. Since the beginning of the twentieth century several attempts have been made to produce a new edition. To date, none has reached a successful conclusion. Since 1968 a new historical and critical edition of Hegel's complete works has been in preparation. By the end of 1996 fifteen volumes had been published.

2 The development of the system: the early writings

The early years of Hegel's intellectual career were characterized less by philosophical ambitions than by interests in public enlightenment and public education. In contrast to his student friends Hölderlin and Schelling, whose activities were directly based on internal philosophical discussions, Hegel aimed in his early works to find ways 'to influence men's lives' (as he wrote to Schelling). He regarded as an appropriate starting point for these

attempts the analysis of the role and consequences which must be attributed to religion, especially Christianity, for the individual and for the social context of a nation. In this early approach two different interests are at work. On the one hand, Hegel aims to show how religion had developed into a power hostile to life, which produces its effect through fear and demands submission. On the other hand, however, he would like to understand the conditions under which it can prosper as a productive element in the life of the individual and society. Hegel's investigations of religion under these two aspects were strongly influenced during the early years (1793-1800) by the cultural criticism and social theories of Rousseau as well as the religious philosophy of Kant (§§11, 13), and by his critical assessment of the theological positions of his academic theology teachers in Tübingen (G.C. Storr and J.F. Flatt). The most important works during this period are represented by the texts which have been preserved as fragments, and which have become known under the titles *Die Positivität der christlichen Religion (The Positivity of the Christian Religion)* (1795-6) and *Der Geist des Christentums und sein Schicksal (The Spirit of Christianity and its Fate)*, (1798-9).

Hegel's religious criticism centres on the concept of 'positive religion'. For Hegel, a positive religion is one whose fundamental content and principles cannot be made comprehensible to human reason. They thus appear unnatural and supernatural, and are seen to be based on authority and to demand obedience. For Hegel, the Jewish religion represents the paradigm of a positive religion. Hegel also considers that the Christian religion has been transformed into a positive religion during the course of its history, in other words into a religion which alienates human beings from themselves and from their fellow creatures (see Alienation §1). He tries to identify cultural and social developments as an explanation for this transformation. In direct opposition to positive religion, Hegel conceives what he calls 'natural religion', which he defines as one whose doctrines correspond with human nature: one which permits or even encourages people to live not only in harmony with their own needs, inclinations and well-considered convictions, but also without being alienated from other people. Hegel's belief in the value for mankind of harmony with oneself (and others), which is strongly influenced by the Stoic ethic (which also via Rousseau had an impact on Kant's practical philosophy), is grounded in a quasi-metaphysical conception of love and life. It owes a considerable debt to the philosophical approach of Hölderlin, with whom Hegel again associated closely during his Frankfurt period. According to this conception, there is a sort of moral emotion of love, which rises above all separations and conflicts, in which persons might be involved in relation to themselves and to others. It is this emotion of love which makes people aware of their unity with others and with themselves. It cannot be adequately thematized by philosophy, which is based on reflection and (conceptual) distinction. It demonstrates vividly, however - and here metaphysics enters - the true constitution of reality, which consists of a state of unity forming the basis for all separations and conflicts and making these possible. This reality, which has to be thought of as unity, Hegel calls 'Life' (*Leben*) and also 'Being' (*Sein*). Hegel's efforts at the end of his Frankfurt period are directed towards thinking of reality in these terms in a sufficiently differentiated manner. In doing so he pursues above all the goal of conceiving of life as a process which generates as well as reconciles oppositions, a dynamic unity of generation and reconciliation. To explain this complex structure, which he conceives what he calls 'life' to be, Hegel devised in the so-called *Systemfragment von 1800* the formula 'Life is the connection of connection and non-connection'. This formula and the concept of life on which it is based already point clearly towards Hegel's later organicist metaphysics.

3 The development of the system: the Jena writings

The work of Hegel's Jena period (1801-6) can be divided into critical and systematic writings. Among the critical writings are his first philosophical publication, *The Difference between Fichte's and Schelling's Systems of Philosophy*, and most of the essays which he published during the years 1802-3 in the *Critical Journal of Philosophy*. In these essays, Hegel reveals himself as a critic of the philosophy of his age, especially of the positions of Kant, Jacobi and Fichte whom he accuses of practising a 'reflective philosophy of subjectivity' as he calls it in the sub-title of his essay *Faith and Knowledge* (1802). For Hegel, reflective philosophy is initially an expression of an age or historical situation. Such an age is subject to the dichotomies of culture (*Bildung*), which are the products of the understanding and whose activities are regarded as divisive and isolatory. Being subject to those dichotomies, it is impossible for such an age to overcome them and restore the harmony which the understanding has destroyed. A philosophy committed to such an age shares its fate, being also unable to remove, at least in theory, the conflicts which appear as the concrete forms of dichotomy. For even when philosophy strives to overcome these conflicts - according to Hegel, 'the only interest of reason' - and thus makes reference to a

particular idea of unity or harmony, even then it remains committed to the conditions of its age and will achieve nothing except newer and even more acute conflicts. According to Hegel, we can characterize the general form underlying the various conflicts as the conflict between subjectivity and objectivity. The attempts of reflective philosophy to overcome them fail, in Hegel's view, because they are largely abstract: that is, they fail to take into account either the subjective or the objective component of the conflict, and declare it to be resolved by neglecting or abstracting from either of these components. In abstracting from subjectivity, objectivity (in Hegel's terminology) is posited as absolute, which leads to the subordination of subjectivity. This way of reconciling the conflict between subjectivity and objectivity is characteristic of all religions describable as positive by Hegel's definition. If, on the other hand, abstraction is made from objectivity, and subjectivity is thus posited as absolute, then objectivity is regarded as being dependent on subjectivity. This one-sided absolutization of subjectivity is Hegel's objection to the philosophies of Kant, Jacobi and Fichte and the reason he describes their theories as forms of a reflective philosophy of subjectivity.

During his early years in Jena, in contrast to the philosophical attitudes which he criticized, Hegel assumes along with Schelling that the described conflict between subjectivity and objectivity can only be overcome by a philosophy of identity. A philosophy of identity is characterized by the preconditions (1) that for each opposition there is a unity which must be regarded as a unity of the opposing factors, and (2) that the opposing factors are nothing more than their unity under the description or *in the form* of the opposing factors. These preconditions suggest that one should understand the overcoming of the opposition between subjectivity and objectivity as a single process which *reconstructs* the unity underlying the opposing factors and makes them possible in the first place. Following the conceptual assumptions favoured by Hegel at the time, the unity to be reconstructed in a philosophy of identity is defined as the 'subject-object', and the subject and object themselves are characterized as 'subjective subject-object' or 'objective subject-object' respectively. The process of reconstruction of the subject-object by means of the assumptions of the philosophy of identity consists of recognizing the subjective and objective subject-object in their specific one-sidedness or opposition to each other, and thus gaining an insight into the internal structure of the subject-object as the unity which underlies the two conflicting factors and makes them possible in the first place. Although Hegel did not persist in using this terminology for long, for most of his time in Jena he nevertheless remained faithful to the project of the development of a unity which he considered to be comprehensive and which consists of its internal opposing elements. The various attempts at a formal description of a process which was aimed at a unity led Hegel to various system models. All of them contained - albeit with variations of terminology and detail - a discipline initially defined by Hegel as 'logic and metaphysics' as well as a so-called 'real philosophy' (*Real-Philosophie*), in other words a 'philosophy of nature' as well as what he later called a 'philosophy of spirit'.

The systematic works of the Jena period, apart from the *Phenomenology of Spirit*, principally include the three *Jena Drafts of a Philosophical System*. Of these (in some cases) comprehensive fragments, mainly the sections dealing with the philosophy of nature and of spirit are extant. As regards the philosophy of nature, in all the Jena versions of this part of Hegel's system the description of all natural phenomena, the analysis of their processes and their interrelationships is achieved by recourse to two essential factors, which Hegel calls 'Ether' and 'Matter' (*Materie*). 'Ether' describes something like a materialized absolute, which expresses and develops itself within the realm of space and time. This entity is now introduced by Hegel in connection with the development of the determinations of nature as absolute matter or alternatively as absolute being, and the task of philosophy of nature lies in interpreting the various natural phenomena - from the solar system and the laws governing its movements to illness and death of animal organisms - as different manifestations of this absolute matter. Hegel is concerned not merely to show that any particular natural phenomenon is in its peculiar way a specific expression of absolute matter. Above all, he is concerned to prove that nature is a unity ordered in a particular manner. As a specific expression of absolute matter, each natural phenomenon represents an element in the ordered succession of natural phenomena. The position of a natural phenomenon in the order of nature is laid down by the specific way in which absolute matter is expressed in it. A consequence of this approach is that here the natural order is understood as determined by certain postulates which result from the structural conditions of the absolute matter and the methodological maxims of the complete description of these conditions. Differences between the Jena versions of Hegel's philosophy of nature mainly result from the inclusion of new facts made available by current science; but they leave his basic assumptions untouched.

Things are different in the case of the Jena writings on the second part of real philosophy - the philosophy of spirit - initially still described by Hegel as the ‘philosophy of ethical life’ (*Philosophie der Sittlichkeit*). They reveal many changes, all linked to modifications of his conception of spirit. Initially, he presents his philosophy of spirit as a theory of ethical life, which he then transforms into a theory of consciousness. For reasons linked to a renewed preoccupation with Fichte and certain new insights into the logical structure of self-consciousness, towards the end of his Jena period Hegel found himself obliged to present an approach which had occupied him since at least 1804-5. This approach enabled him to liberate the philosophy of spirit from its narrow systematic links to a conception of ethical life based on assumptions incompatible with his new conception of spirit. It assumes that only the formal structure of self-consciousness, which consists in its being a unity of generality and singularity, can provide the framework within which the logical-metaphysical determinations, the natural world and psychosocial phenomena unite to form a meaningful systematic context. For the philosophy of spirit this means in particular that as far as method is concerned it is better equipped for the implementation of its systematic task of being the representation of the processes of self-realization of what Hegel calls ‘reason’. This insight into the formal structure of self-consciousness is the final achievement of his Jena period, and one which he never subsequently abandoned.

4 The system: metaphysical foundations

Hegel’s systematic philosophy attempts to comprehend reality in all its manifestations as a self-representation of reason (*Vernunft*). His conception of what he calls ‘reason’ combines various specifically Hegelian connotations, both ontological and epistemological. For him, ‘reason’ is not merely the name for a human faculty which contributes in a specific manner to our gaining knowledge; he also uses ‘reason’ to describe that which is ultimately and eminently real. This is the ontological connotation. Reason is reality, and that alone is truly real which is reasonable. This programmatic credo, which has become famous from the foreword to Hegel’s *Philosophy of Right*, is the basic precept determining the entire approach to his system.

At least three different convictions make up this basic precept of the ontological dignity of reason. The first is that the totality of that which in one sense or another is real must be considered as the differentiation and partial realization of a primary structure which forms the basis for all facts which are real in that specific sense. Hegel calls this primary structure ‘the absolute’ or ‘reason’. He shares this conviction of the necessity of assuming a primary structure called ‘reason’ (interpreted ontologically) with Fichte, Schelling, Hölderlin and other members of the post-Kantian idealistic movement whose monistic approach is, indeed, defined by this assumption. For Hegel, therefore, this conviction does not require detailed philosophical justification. For him, it is justified because it alone offers a basis for systematic philosophical considerations, following the failure of all previous philosophical attempts to conceive of a unified and complete representation of the world.

This first conviction, which forms a part of Hegel’s ontological conception of reason, is still too imprecise to provide a clue as to why exactly the concept ‘reason’ can be used to characterize the primary structure. Hegel’s second important conviction, however, makes this clearer. It relates to the internal constitution of the structure which he characterizes as reason. He understands this structure to be a complex unity of thinking and being. The relevant motives for this conviction can be summarized as follows: the only philosophical approach which can organize the whole of reality into a unified and coherent picture accessible to knowledge is one which insists that everything taken to be real is only real inasmuch as it can be comprehended as the actualization of some specific structural elements of reason. This assertion of the essential reasonableness of all being, together with the first conviction of the necessity of assuming a primary structure, leads directly to the concept of this primary structure as a unity of thinking and being, understood in the very radical sense that thinking and being are one and the same, or that only thinking has being. If we now call this unity of thinking and being ‘reason’, and if, like Hegel, we are convinced that the requisite primary structure must be thought of as this unity of thinking and being, then reason will be declared on the one hand to represent what in the final analysis is ultimately real, and on the other that which alone is real. Since a monistic position is one in which a single entity is maintained as the ultimate and sole reality, Hegel’s philosophical conception has rightly been called a ‘Monism of Reason’ (see Monism).

The third conviction which enters into Hegel’s basic assumption of reason as the primary structure constituting reality and thus being ultimately and only real is that this structure constitutes reality and thus its own objectivity in a teleological process which must be understood as a process of knowledge. It is this conviction which leads to

the characteristically Hegelian dogma that there can be no adequate theory of reality without a dynamic or process-oriented ontology (see Processes). The formula which Hegel uses to characterize this process from his early Jena works onwards shows very clearly the dominant role which he assigns to what he defines as ‘reason’ in the systematic approach designed to elaborate his third conviction. This process is described as ‘self-knowledge of reason’ (*Selbsterkenntnis der Vernunft*). Hegel tries to integrate within this formula various aspects of his conception of reason. The first aspect is that it is necessary to take reason, understood as the primary structure, as something which is essentially dynamic. By this he means that the element of self-realization forms part of the moments which determine the primary structure. It is difficult to understand the way in which Hegel links this element of self-realization into his idea of reason as the unity of thinking and being. In order to get a rather over-simplified idea of the background for Hegel’s claim, it might help to rely metaphorically on the theory of organism: just as an organism can be described as an entity whose development is linked to the concept or the structural plan of itself in such a way that the (more or less) successful realization of this concept or structural plan belongs to its being real, so we should think of Hegelian reason, understood as the ontologically relevant primary structure, as realizing in a quasi-organic developmental process the unity of thinking and being which characterizes its concept, thereby representing itself as real or as reality.

The second aspect Hegel has in mind when he speaks of ‘self-knowledge of reason’, describing a process which must indeed be understood as that of the self-realization of reason, is that this process represents a process of *recognition* for reason. It is apparently not sufficient for Hegel to embed his idea of reason as the ontological primary structure in a conception of realization based on the paradigm of the organism. Such a grounding seems to be too unspecific for him, because it does not show how to describe a process which is typical of *all* organisms in such a way that we understand more precisely and in detail what it means for the process to be one of self-realization of *reason*. The specific way in which reason realizes itself is to be characterized first of all as a process of recognition, because only this characterization takes into account the fact that that which is being realized, namely reason, must be thought of strictly as nothing more than thinking *qua* recognition. But even this way of conceiving the realization of reason is still too imprecise, unless one includes in the concept of realization the thesis that reason is the ultimately and only real ontological primary structure. The inclusion of this thesis then leads directly to the teleologically conceived description of the process of the realization of reason as a process of self-recognition. For if only reason - by which is meant the unity of thinking and being - is real, and if an integral part of this concept of reason is the conception of its realization in the form of a process of recognition, then this process can only be directed towards the recognition of reason itself, because nothing else exists. Since this process aims to make reason aware that it alone is real, the presentation of this process, in Hegel’s view, must take on the form of a system in which each manifestation of reality documents its reasonable nature. His philosophy aims to elaborate this system.

The project of exhibiting reason not only as the basis for all reality, but also as the whole of reality itself, was Hegel’s sole, lifelong philosophical goal. It took him some time to be able to formulate this project explicitly. This is linked to his intellectual development (see §2 above). He also considered various approaches to the realization and development of this project (see §3 above), but he never felt any need to question the project itself.

5 The system: *Phenomenology of Spirit*

The *Phenomenology of Spirit* (1807) is Hegel’s most influential work. It serves as an introduction to his philosophical system by means of a history of the experience of consciousness. The *Phenomenology of Spirit* represents only one of a number of introductory attempts he made. In the Jena writings and system drafts, a discipline which Hegel calls ‘logic’ assumes this function. This logic is intended to fulfil its introductory function by raising our ‘normal’ thinking, which is characterized by its confinement to irreconcilable oppositions, to the level of ‘speculation’ - Hegel’s term for philosophical thinking. Speculative thought is characterized by the knowledge of the reconcilability of oppositions and of the mechanisms of their coming about. That thinking, which by its insistence on oppositions simultaneously maintains their basic irresolvability, Hegel referred to at this time as ‘reflection’. He regarded the elevation of this thinking to the position of speculation as a destruction of the structures which characterize reflection, and which together constitute the finiteness of reflection. ‘Finiteness’ of reflection or (used by Hegel as a synonym) of the understanding (*Verstand*) is initially a way of saying that thinking whose oppositions are irreconcilable moves within limits and must thus be regarded as finite. According to Hegel, it is now the task of logic to carry out the destruction of the finiteness of reflection or of the thinking of

the understanding, thereby simultaneously leading to the standpoint of speculation or of the thinking of reason. Hegel sees the problem of a logic, which he understands as an introduction to philosophy, to be to carry out this destruction in such a way that not only the limitations of the thinking of the understanding and its preconditions are presented as mistakes and absurdities, but also that during this destructive process those structures become clear which guarantee a reasonable (that is, an intrinsic speculative) insight into the basic structures of reality.

Towards the end of his Jena period, Hegel abandoned the project of developing a system of logic as an introduction to philosophy, and in its place presented a new discipline which he called the 'Science of the Experience of Consciousness' or 'Phenomenology of Spirit'. The declared goal of this discipline is twofold: on the one hand, it should destroy our supposedly natural picture of the world, and thus also our understanding of ourselves as the more or less consistent holders or subjects of this view, by demonstrating the contradictions which arise in our normal, complex view of the world. And second, it should thereby demonstrate that our view of the world as something which is both alien and different from us is not tenable, but that we and the world represent a structural unity with the essential characteristic of self-knowledge and thus being conscious of itself.

Hegel pursues this dual goal in a complex and ambitious thought-process, which attempts to combine and position within a comprehensive context a wide range of themes - historical, epistemological, psychological, meta-scientifical, ideological-critical, ethical, aesthetical and religio-philosophical. This whole thought-process is based on two convictions which govern Hegel's entire construction: (1) It is possible to conceive of all epistemic attitudes of a consciousness towards a material world as relations between a subject termed 'cognition' (*Wissen*) and an object termed 'truth' (*Wahrheit*). That which is presented as cognition or truth is in each case determined by the description which the consciousness is able to furnish of its epistemic situation and its object corresponding to this situation. (2) 'Knowledge' (*Erkenntnis*) can only be taken to be that epistemic relation between cognition (subject) and truth (object) in which cognition and truth correspond with each other, which for Hegel is only the case if they are identical. A necessary, though not sufficient, condition for claiming this relationship of identity between cognition and truth is that what is regarded as cognition or as truth respectively is not formulated in a self-contradictory or inconsistent manner. For the Hegel of the *Phenomenology of Spirit*, and of the writings which were to follow, knowledge in the strict sense is thus really self-knowledge.

In characterizing the various epistemic attitudes of a consciousness to the world in the *Phenomenology of Spirit*, Hegel takes as his starting point something which he calls 'sense certainty'. He uses this term to describe an attitude which assumes that in order to know the true nature of reality we must rely on that which is sensually immediately present to us as a spatiotemporally given single object. Hegel demonstrates the untenability of this attitude by attempting to prove that in such an immediate reference to objects nothing true can be claimed of them. Moreover this immediate approach shows that any attempt to gain knowledge of what an object really is implies at the outset a different attitude towards objects. This attitude is determined by the assumption that what we are really dealing with if we refer to objects in order to know them is not the immediately given object, but the object of perception, which is characterized by Hegel as an entity defined through its qualities. According to Hegel, however, even this attitude is not tenable. Neither the perceiving consciousness nor the object perceived nor the relationship which is believed to exist between the two can be accepted in the manner in which they appear in this constellation: the subject, which aims to perceive the object of perception as that which it really is, can neither formulate a consistent concept of this object nor describe itself in unequivocal terms. The consciousness is thus led to a concept of an object which differentiates between what the object is in itself and what it appears to be. In order to differentiate in this manner, the consciousness must define itself as understanding, to which the inner constitution of the object in itself is disclosed as being constituted by its own laws, that is, by the laws of the understanding. Although, according to Hegel, this interpretation of the objective world through the cognizing subject also produces neither a truthful concept of the cognizing consciousness nor of the object in question, it none the less leads to the enforcement of an attitude according to which consciousness, when referring to an object, is referring to something which it is itself. The realization of this insight - that consciousness, when referring to objects, in reality relates to itself - converts consciousness into self-consciousness.

The various ways in which consciousness deals with itself and the objective manifestations corresponding with these ways as reason and spirit are comprehensively discussed by Hegel in the remainder of his *Phenomenology of Spirit*. It is in this context that he presents some of his most famous analyses, such as the account of the master-servant relationship, his critique of the Enlightenment and the French Revolution, his diagnosis of the

strengths and weaknesses of the ancients' ideas of morality and ethical life and his theory of religion. The conclusion of the *Phenomenology of Spirit* forms what Hegel calls 'absolute knowledge'. Hegel characterizes this knowledge also as 'comprehending knowledge' (*begreifendes Wissen*), aiming thereby to highlight two ideas: (1) that this knowledge is only present when the subject of the knowledge knows itself to be identical *under every description* with the object of that knowledge. Comprehending knowledge therefore only occurs when the self knows itself to be 'in its otherness with itself', as Hegel puts it at the end of the *Phenomenology of Spirit*. He also aims to point out (2) that this type of identity of a subject with an object is that which constitutes the essence of that which he calls the 'Concept' (*Begriff*) of reason. The task of the *Science of Logic* is to develop this 'Concept' of reason in all its logical qualities. The goal of the *Phenomenology of Spirit*, the discipline which is to provide an introduction to logic, is achieved when it becomes evident to the consciousness that (Hegelian) truth only belongs to the (Hegelian) Concept.

But the *Phenomenology of Spirit* is not just an introduction to the system. From another point of view, Hegel describes the phenomenological process as 'self-fulfilling scepticism'. By means of this metaphor he attempts to establish a link to a subject closely connected with his critical assessment of his cultural and political environment, namely that of dichotomy (*Entzweiung*). For Hegel, the modern age is characterized by the fact that unity has disappeared from people's lives. The all-embracing unity of life can no longer be experienced, as people are no longer in a position to integrate the various aspects of their understanding of the world in a conflict-free context. So, for example, their moral convictions will force upon them a view of the world in which something like freedom and consequently something like the belief in the possibility to cause events based on free decisions occupies an irrefutable position. This view, based on moral convictions, stands in a conflicting and, finally, aporetic relationship to their scientific view of the world, which commits them to an understanding of the world in which there are no first causes or unconditioned facts, because each cause must itself be interpreted anew as an effect, whose cause can only be seen as determined by previous circumstances. In this view of the world there is apparently no place for freedom. The conflict between various aspects of the understanding of the world, cited here by way of example, is for Hegel by no means singular; instead, it runs like a leitmotif through the modern conceptualizations of all spheres of life. He initially interprets it wholly in the spirit of Rousseau as a product of culture and civilization. This conflict is what separates a human being from itself, so that it continues to be denied a coherent image of the world. As a creature living in dichotomies, the modern person is an example of what Hegel calls the 'unhappy consciousness'.

Modern consciousness now attempts to solve this conflict by making each in turn of the conflicting views into the dominant attitude of its entire interpretation of the world. In this way, however, it can only achieve a one-sided interpretation of reality, which is just as incapable of doing justice to the true nature of reality as to the need of human consciousness to integrate all aspects of reality into its understanding of the world as a coherent unity. According to Hegel, it is in this situation that the need for philosophy arises. It is philosophy's task to destroy these one-sided total interpretations of consciousness and in this destruction to lay the foundation for the true complete interpretation of reality. The *Phenomenology of Spirit* describes this process of destruction and foundation-laying. The consciousness experiences it as a process of permanent destabilization of all the convictions on which it has always based its one-sided interpretations of the world. In this sceptical approach it is forced to doubt everything and to abandon all its supposed certainties. While the phenomenological process thus concedes a philosophical value to scepticism, in Hegel's understanding it simultaneously overcomes this scepticism by claiming a truth-revealing function for it. It is also Hegel's intention that the *Phenomenology of Spirit* should in this respect be understood as a treatise on the cathartic effect of philosophical scepticism.

Two questions have often been raised in connection with Hegel's conception of the *Phenomenology of Spirit* as an introduction to his 'System of Science', especially his logic. The first is whether Hegel does not assume in advance certain central theses of the discipline to which the *Phenomenology of Spirit* is intended to provide an introduction. This question draws attention to a methodological problem which originates from Hegel's assertion that the process of consciousness described in the *Phenomenology of Spirit* is not guided by any preconditions external to this process. This assertion seems difficult to square with certain manoeuvres which Hegel makes during the course of the *Phenomenology of Spirit*. The second question is of a more intrinsic nature and concerns the categorical apparatus employed by Hegel in the *Phenomenology of Spirit*. In this context, in particular his phenomenological conception of negation and identity as well as his concept of knowledge aroused critical interest

from the very beginning.

It is difficult to determine exactly how Hegel himself later assessed the success of the *Phenomenology of Spirit* as an introduction to the point of view which is assumed at the beginning of his *Science of Logic*. On the one hand, he seems to have allotted it a certain value throughout his entire life, not only as a history of consciousness but also as an introduction. This is shown not only by the fact that he made arrangements for the publication of a second edition of the work immediately before his death, but also by later statements in the various editions of the *Science of Logic* and the *Encyclopedia of the Philosophical Sciences*. However, it is in precisely these statements that one finds Hegel expressing an increasingly critical attitude towards his project of a phenomenological introduction to the system. In this context it should also be recorded that by 1827 at the latest (that is, from the second edition of the *Encyclopedia of the Philosophical Sciences*) Hegel no longer has recourse to a version of phenomenology as ‘a more detailed introduction, in order to explain and lead to the meaning and the point of view which is here allotted to logic’, but for this purpose uses instead a discussion which deals with three different ‘attitudes of thought to objectivity’.

6 The system: *Science of Logic*

The real centre of the Hegelian system is the discipline he described as ‘Logic’. It contains his doctrine of the categories, to use traditional terminology (see Categories §1). Hegel dedicated his most comprehensive and complex work to this discipline, the *Science of Logic* (1812-16), later adding a much shorter version within the framework of the *Encyclopedia of the Philosophical Sciences*.

The starting point of the *Logic* is the insight, justified in Hegel’s view by the result of the *Phenomenology of Spirit* (1) that all true knowledge is knowledge of oneself, and (2) that the subject of this knowledge, that is to say, that which knows about itself, is reason. Because Hegel - following Schelling - only considers that to be real which can also be known, he concludes from the results of the *Phenomenology of Spirit* that only reason is real. He thinks of this reason as, internally, an extremely complex entity. Hegel now distinguishes between the ‘Concept’ of reason and the process of its realization. The object of the *Science of Logic* is the conceptual, that is to say, for Hegel, the logical development of this Concept. Since this Concept is the Concept of that which alone is real, Hegel can maintain that his *Science of Logic* takes the place of traditional metaphysics, which concerned itself with the elucidation of the basic ways in which we can think of reality.

Since the object whose Concept is to be logically discussed is reason, understood to be the sum of reality, the Concept of reason must include not only those aspects which account for reason’s character of reality or of being, but also those aspects which do justice to the peculiar character of reason as thinking. Hegel calls these aspects ‘Determinations of the Concept’ (*Begriffsbestimmungen*). Those aspects of the Concept of reason which take into account its character of being are developed by Hegel in the section ‘Objective Logic’ in the *Science of Logic*. He presents those aspects which are intended to do justice to its thinking character in the section called ‘Subjective Logic’. He further subdivides ‘Objective Logic’ into ‘Logic of Being’ and ‘Logic of Essence’.

In his ‘Objective Logic’, Hegel tries to show how it is possible to generate from very simple, so-called ‘immediate’ determinations such as ‘Being’, ‘Nothing’ and ‘Becoming’ other categories of quality and quantity as well as relational and modal determinations, such as ‘Cause-Effect’, ‘Substance-Accidence’ and ‘Existence’, ‘Necessity’ and the like. As in the ‘Subjective Logic’, the basic strategy here for the creation of categories or determinations of the Concept, assumes that (1) for every category there is an opposing one which upon closer analysis reveals itself to be its true meaning, and that (2) for every two categories opposing each other in this manner there is a third category whose meaning is determined by that which makes the opposing categories compatible. Hegel considers these two assumptions justified because only they can lead to what in his eyes is a complete and non-contingent system of categories. Hegel himself, however, did very little to make their exact meaning clear, although he uses them with great skill. Immediately after his death this led to a confused and still inconclusive discussion regarding their interpretation. Many judgments concerning the worth or worthlessness of Hegel’s philosophy are linked to this discussion, which has taken its place in the annals of Hegel research as a discussion concerning the meaning, significance and value of the so-called ‘dialectical method’. (Hegel himself preferred ‘speculative method’.)

In particular, Hegel’s claims about the truth-generating function of contradiction have played a major role in the

discussion of the ‘dialectical method’ described in the *Science of Logic*. Though highly praised by Hegel himself, his doctrine of the nature and methodological merits of contradiction has proved to be inaccessible and obscure. This may have been caused in part by Hegel’s extremely concise and provocative formulations of this methodical maxim. The reader is reminded in this context not only of the succinct formulation which he chose to defend on the occasion of his Jena Habilitation - ‘contradiction is the rule of truth, non-contradiction the rule of falsehood’ - but also of his provocative version of the principle of contradiction, according to which ‘everything is inherently contradictory’. The difficulties associated with the comprehension of the Hegelian conception of contradiction have perforce a link with his particular unconventional concept of contradiction. Two points are particularly important, in that they differentiate his concept from the classical concept of contradiction of traditional logic: (1) A contradiction between two propositions cannot be confirmed solely on the basis of their ascription to a single subject of two contradictory predicates; it is also necessary to take into account the meaning of the subject of these propositions. If the contradictory predicates cannot meaningfully be attributed to the subject, then no contradiction arises. ‘Legible’ and ‘illegible’ are predicates which will only lead to contradiction if attributed to texts, but not, for example, to bananas. For Hegel, this means among other things that the relation of contradiction is dependent on the context. (2) Hegel thinks of contradictions as analogous to positive and negative determinations, which neutralize each other but without making that whose neutralizing determinations they are into a contradictory concept which has absolutely no meaning, which therefore means nothing (the Kantian ‘Nihil negativum’). Rather, the way in which positive and negative determinations neutralize each other tells us something informative about the object to which the neutralizing determinations apply. For example, possession of DM100 neutralizes a debt of DM100, without thereby making the concept of property a contradictory concept. Instead, the way in which this neutralization takes place makes clear that the concept ‘property’ means something which must be thought of as of a quantifiable size. For Hegel this is a consequence of ‘the logical principle that what is self-contradictory does not dissolve itself into a nullity, into abstract nothingness, but essentially only into the negation of its *particular* content’. Whether these two convictions are sufficient to justify Hegel’s thesis that contradictions play a ‘positive’ role in cognition procedures is rightly controversial.

Hegel’s ‘Subjective Logic’, the second part of the *Science of Logic*, contains not only his so-called ‘speculative’ interpretation of the objects of traditional logic, that is, his own doctrine of concepts, judgments and syllogisms, but above all his theory of the ‘Concept’. This theory is deeply rooted in Hegel’s critique of traditional metaphysics, and is thus most easily comprehended when placed in that context. He presents this critique most tellingly in the third edition of the *Encyclopedia of the Philosophical Sciences*. What matters in philosophy, and what philosophy aims at, is the ‘scientific recognition of truth’. This means among other things that philosophy is concerned with the recognition of ‘what objects really are’. According to Hegel, the question of what objects really are has been approached in philosophy from a variety of angles, but all the different modes in which the question has been answered to date are unacceptable because they are based on false premises. Traditional metaphysics is one of the ways of approaching the question of what objects really are. Hegel characterizes this approach as ‘the *impartial* method’, which is motivated by the assumption that ‘through *thinking the truth becomes known*, what objects really are is brought to one’s consciousness’. In contrast to other philosophical approaches that deal with the question posed, metaphysics, according to Hegel, is in principle certainly capable of contributing to the cognition of what things in truth are, because it starts from the correct assumption that ‘the determinations of thought’ are to be seen as ‘the fundamental determinations of things’. But traditional metaphysics has not made a significant contribution to the cognition of truth, because it was only able to transform its correct initial assumption in a systematically erroneous manner.

According to Hegel, the crucial weakness of traditional metaphysics is to use the form of judgment *in an unreflective manner*, and this shows itself in various ways. First it shows in the unfounded assumption of traditional metaphysics that judgments provide a particular and direct insight into the constitution of reality or that which really exists. In Hegel’s view this unfounded assumption has two consequences: the first is that it tends without reason to favour a particular ontological model of reality because it regards the subject-predicate form as the standard form of the judgment; the second consequence, in Hegel’s view incomparably more problematic, consists of the unfounded tendency of traditional metaphysics to conclude from the unquestioned and assumed correspondence between the form of judgment and constitution of reality that one can express by means of judgments what objects really are. Hegel does not find problematic the assumption contained in this conviction that one can make judgments concerning objects. He believes that the problem lies rather in the fact that one can

assume *without examination* that 'the form of the judgment could be the form of truth'. In Hegel's view, however, such an examination is essential because the traditional understanding of subject and predicate does not justify the assertion that a subject-predicate judgment actually contributes something to the determination of a real object. An additional problem lies in the fact that the unconsidered use of the form of judgment has led traditional metaphysics erroneously to use a 'natural' interpretation of the concepts of subject and predicate. The consequence of this interpretation is that judgments of the subject-predicate form can lay no claim to 'truth'.

Hegel's chief criticism of traditional metaphysics therefore lies in the lack of clarity associated with its interpretation of the form of the judgment. In particular he rejects its tendency to interpret the judgment 'naturally', which for him means to encourage a subjectivist interpretation of the judgment built on the concept of representation. Such a subjectivist interpretation cannot show how to guarantee for the judgment some sort of claim to truth or recognition of what something really is. Therefore the subjectivist metaphysical interpretation of the judgment is problematic at the very outset. Moreover, it becomes downright dangerous when one considers its ontological implications, for it leads erroneously to the assumption that the objects corresponding to the subjects of the judgment are to be thought of as substances to which are attributed the characteristics described by the predicate-concepts. The unreflective subjectivist interpretation implies or at least suggests what may be called a substance-ontology, according to which substances which are independent of each other are taken as the fundamental entities of reality, determined predicatively by accidental characteristics which are applicable or not applicable to them. It is this commitment to a substance-ontology which Hegel critically imputes to traditional metaphysics. From this criticism, he deduces that it is first necessary to reach an agreement as to what the object really is before one can adequately assess the function and the achievement of the judgment in the context of knowledge. To reach this agreement is the task of the Logic of the Concept.

The starting point of Hegel's theory of the Concept is the assumption which he imputes to traditional metaphysics as an insight which is in principle correct. This was the insight that only through thinking can one recognize what something really is. Since in Hegel's view thinking is concerned not with intuitions or representations, but with concepts, he identifies that which something in truth or really is with its Concept. Because of this identification, talk of the Concept acquires an ontological connotation. Hegelian Concepts must not be confused with the so-called general concepts of traditional logic. They are difficult to understand precisely and are characterized by the fact that they are (1) non-sensible - which means that they are a particular type of thought-object - and that they are (2) something objective as opposed to subjective. Regarded as these objective thoughts, these Concepts are determined in the sense that in them different relations of determinations of the Concept are to be encountered which occur as determinations of thinking or of thought (*Denkbestimmungen*). These determinations of thought can themselves be regarded as a kind of predicative characteristic. They make up the multitude of all those determinations on the basis of which the Concept of an object can be seen as completely fixed.

Now, in Hegel's view, not everything has a Concept which in one sense or another is ordinarily thought of as an object. A (Hegelian) Concept is only allotted to objects which can be thought of on the model of an organism. Hegel thus maintains that one can only regard those objects as real or as existing in truth for which there is a Concept which can be interpreted on the organic model. If, then, the 'scientific recognition of truth' consists in recognizing the Concept of something, and if a Concept is always a Concept of a organic-type object, then the question arises how one should conceive of such a Concept. For Hegel it is clear that in his concept of a Concept, he must include everything needed to describe an organism. This includes first of all what Hegel calls the subjective Concept, which one can best regard as the sum of all characteristics whose realization represents an organic-type object. For Hegel, in the case of the concept of reason, whose Concept the *Science of Logic* elaborates, these characteristics are exclusively logical data which can be presented in the form of determinations of concepts, judgments and syllogisms. Furthermore, Hegel's Concept must include the element of objectivity. 'Objectivity' here means more or less the same as reality or the state of being an object and suggests the fact that it is part of the Concept of an organism to realize itself. Since, however, Hegel holds that there is ultimately only one object which really exists, namely reason, the Concept of this object must include a characteristic which is exclusively applicable to itself. This characteristic must permit the justification of the claim that in reality there is only one Concept and therefore also only one object. Hegel calls this characteristic 'subjectivity'.

Although it is easy to see that the term 'subjectivity' describes a central element of Hegel's logical theory, it is very difficult to shed light upon its meaning and function therein. It is relatively obvious only that Hegel attributes

the characteristic of subjectivity not just to his Concept, but also to entities such as 'I', 'self-consciousness' and 'spirit'. We are therefore on safe ground if we assume that the subjectivity which is to be attributed to the Concept is precisely that which is also attributed to the I, self-consciousness or spirit and which distinguishes them from other types of organism. The ground becomes more dangerous when it is a matter of stating what subjectivity actually means. This is not merely because Hegel distinguishes between different types of subjectivity, but also because the subjectivity which is constitutive of the Concept is tied to conditions which are difficult to state with any precision. In general it seems to be correct to say that subjectivity occurs when something recognizes itself as being identical with something else. If we follow the *Science of Logic*, then this relationship of identity known as 'subjectivity' can only be established between entities which themselves can be thought of as being particular complexes of relations of similar elements or moments. Subjectivity in this sense is thus intended to describe a certain form of self-reference or self-relationship. According to Hegel, there should be only one entity to which the term 'subjectivity' can be attributed as a characteristic in the sense which has just been explained - the Hegelian 'Idea'. He says of it, 'The unity of the Idea is subjectivity'. This Idea now forms the end of the *Science of Logic*, because through it the Concept of reason has been completely explicated. He also describes this Idea as the absolute method, for it is not only the result, that is, the Concept which comprehends all his moments, but also the complete and systematically generated series of these moments.

The results of the Logic of the Concept represent the justification for Hegel's belief that, apart from a system of logic, a complete system of philosophy must include a so-called 'real philosophy', which is divided into a philosophy of nature and a philosophy of spirit. Hegel undertakes this justification within the framework of the exposition of what characterizes the fully developed (Hegelian) Concept. This exposition only becomes comprehensible if one remembers that Hegel is a supporter of the organological paradigm in metaphysics, according to which that which really is must be regarded as a particular type of organism. Hegel describes the type of organism relevant to his metaphysics as an object which has realized or objectivized its Concept in such a way that it comprehends itself as the objectivization of this Concept of itself. On the basis of this conception, Hegel now develops the following consideration: the (Hegelian) Concept is something which is to be regarded as a unity of (in some ways incompatible) determinations of the Concept. Among these determinations also belong, as Hegel believes he can show, that of objectivity. By this he means that it is a part of the nature of a Concept to become objective, to manifest itself as an object. Now, the only object which is an adequate realization of the Concept is the one to which what Hegel calls 'subjectivity' can be attributed. 'Subjectivity' is the name of a relational characteristic which is present when something *knows* itself to be identical with something else. For Hegel, it follows from these stipulations that subjectivity can only be attributed to the object which knows itself to be identical with its Concept. To produce this knowledge is therefore a demand inherent in the nature of the Concept. Since it is the sole task of the *Science of Logic* to exhibit the Concept of reason, and since this Concept contains the demand for the production of a form of knowledge which can only be acquired when (1) the Concept objectivizes itself, that is, becomes an object, and (2) this object comprehends itself as being identical with its Concept, then it is already a demand inherent in the Concept of reason that reason should be discussed (1) from under the point of view of its objectivity or as an object, and (2) under the aspect of its known identity with its Concept. The first of these topics is the subject of a philosophy of nature; the second that of a philosophy of spirit.

7 The system: philosophy of nature

Hegel's philosophy of nature is an attempt to explain how it is possible that we can recognize nature as a complex whole standing under a set of laws. He thereby takes up the question, important in particular to both Kant and Schelling, of which epistemological and ontological preconditions underlie our conviction that nature can be known. Although Hegel had thought about the problems of a philosophy of nature since his time in Frankfurt, and although he produced several versions of a philosophy of nature during his Jena period, he only published this part of his system once, quite late, in his *Encyclopedia of the Philosophical Sciences*. Hegel's philosophy of nature is of interest mainly in three respects. The first concerns the way in which he transforms his logical theory into an interpretation of natural phenomena. The second relates to the question of how far Hegel's conceptions in the field of the philosophy of nature take into account the scientific theories current at the time. Finally, the third leads to the question of what we should make of Hegel's approach to a philosophy of nature within the framework of present-day philosophy of science. Since the philosophy of nature is that part of the Hegelian system which is

traditionally regarded with the greatest suspicion and which for this reason has received the least scholarly attention, the assessment of the second and third aspects of this philosophy of nature has so far produced very few uncontroversial results.

As far as the construction of a philosophy of nature according to the requirements of the logical theory of the Concept is concerned, Hegel assumes in accordance with his organological conception of reason that we should think of nature '*in itself* as a living whole'. This living whole has to be conceived of primarily under three different determinations, which reproduce to a certain extent the central characteristics of the Concept of reason as developed by the logical process. According to the first of these determinations, nature is to be considered as a whole defined by space, time, matter and movement. This way of looking at nature makes it the object of what Hegel calls 'mechanics'. According to Hegelian mechanics, space, time, matter and movement, their characteristics and the laws of nature which describe their relationship are generated by the formal structural moments of the (Hegelian) Concept. The utilization of such a 'conceptual' procedure to gain and secure scientific results was never seen by Hegel himself as a direct alternative procedure to empirical scientific research. On the contrary, he was of the opinion that (for instance, by means of his philosophical mechanics) he only makes explicit the conceptual elements and secures for them a rational foundation which is implicitly contained in every scientific mechanical theory that acquires its data 'from experience and then applies the mathematical treatment'.

According to Hegel, his philosophical mechanics leads to the insight that we must think of the whole of physical nature as 'qualified matter', that is, as a totality of bodies with physical characteristics. This provides the second main determination by which nature is to be comprehended. Hegel ascribes to this way of comprehending nature a discipline which he calls 'physics', including under this heading everything which can in any way be linked with the material status of a body. Accordingly, from phenomena like specific weight, through those like sound, warmth, shape, electricity and magnetism, to the chemical reactions of substances, everything is described as being a consequence of and following from the constitution of the (Hegelian) Concept. He also adds theses concerning the nature of light and a doctrine concerning the elements earth, fire, water and air. It was this part of Hegel's philosophy of nature in particular which drove Hans Scholz, among others, to the following crushing judgment: 'Hegel's philosophy of nature is an experiment which set the philosophy of nature back several centuries instead of furthering its cause, returning it to the stage it had reached at about the time of Paracelsus'. Whether this dictum has substance, however, depends very much on what conception of nature and science one favours.

The third part of Hegel's philosophy of nature consists of the so-called 'organic physics' or 'organics'. In this section the characteristic of subjectivity, familiar from his logical theory, is the determination under which nature is to be regarded. Since in the context of the philosophy of nature, Hegel interprets subjectivity as an essential characteristic of organic life, this section of his philosophy of nature is concerned with nature as a hierarchy of organisms or as an 'organic system'. He distinguishes between three forms of organic life which are exemplified in three types of organism: the general form, which is represented by the geological organism, the particular, which is realized in vegetation, and the individual, which finds its expression in animal organisms. He regards these forms as hierarchically ordered by increasing degree of complexity. In some ways Hegel thematizes relations and conditions of dependence: just as vegetable life-forms presuppose geological structures and processes, so animal organisms presuppose a fully developed plant world. Hegel links this last part of his philosophy of nature to his philosophy of spirit by means of an analysis of the phenomenon of the death of an individual natural being. Here the leading idea is that although through death all natural determinations of the individual are removed, so that we can speak of the 'death of the natural element', none the less death does not annihilate the principle of life, that which is responsible for the essential unity of animal organization and which Hegel calls the 'soul'. Since Hegel interprets the soul as a form of spirit, and since according to his conception the soul is not destroyed by death, he can now postulate the reality of spirit independently of natural determinations as the result of his philosophy of nature, and investigate this reality in its various forms within the framework of a philosophy of spirit.

The question whether Hegel's philosophy of nature integrates in a relatively informed manner the state of science during his lifetime has provoked a number of fairly controversial answers, as has the question whether his approach can still provide any promising perspectives which are relevant today. During the nineteenth century, Hegel's philosophy of nature was broadly considered scandalous by the majority of scientists, an attitude which contributed in no small measure to the discrediting of his philosophy as a whole. This assessment also meant that Hegel's philosophy of nature has never really been taken seriously again. Since 1970, however, the situation has

changed somewhat. Starting from and relying on recent investigations in the history of science regarding the development and state of the sciences during the early nineteenth century, increasing numbers of scholars are inclining towards the view that Hegel was indeed much more familiar with the science of his time and its problems than was generally believed during the nineteenth and most of the twentieth century. It seems advisable at present to refrain from passing final judgment on this matter. The same cannot, however, be said with regard to the present relevance. Here one cannot ignore the fact that Hegel's theses concerning philosophy of nature are, quite simply, meaningless for present-day scientific theory.

8 The system: philosophy of spirit

Hegel's philosophy of spirit is divided into a theory of subjective, objective and absolute spirit. The philosophy of subjective spirit contains Hegel's philosophical psychology; his philosophy of objective spirit is devoted to his theory of law and politics and his conception of world history; and his philosophy of absolute spirit presents his theory of art, religion and philosophy. Hegel presented his philosophy of subjective spirit and in particular his philosophy of absolute spirit to a wider public only in outline in a few paragraphs of the *Encyclopedia of the Philosophical Sciences*. He presented his philosophy of objective spirit not only in the *Encyclopedia*, but also in detail in a work which was already highly regarded during his lifetime, *Natural Law and Politics in Outline: The Principles of the Philosophy of Right* (1821). In this part of his system Hegel again relies on the principle developed in his logical theory that something - here the entity called 'spirit' - must experience a process of realization in order to be able to recognize its truth, or what it is.

The philosophy of subjective spirit contains an anthropology, a phenomenology of spirit and a psychology. In these sections Hegel describes and analyses all the phenomena that influence the somatic, psychophysical and mental characteristics, conditions, processes and activities of the individual. The gamut of subjects he covers runs from the natural qualities of the individual, expressed in temperament, character and physiognomy, via sensibility, feeling, awareness and desire, to self-awareness, intuition, representation, thinking and wanting. Here one finds Hegel's theory of language acquisition, of practical feeling, of the achievements and function of imagination, his defence of the life-preserving power of habit, his solution of the mind-body problem, his understanding of the origin and treatment of mental illnesses and many other subjects. In these analyses Hegel's aim is to replace the 'ordinary approach' of empirical psychology with a 'philosophical perspective' towards psychological phenomena. The dominant characteristic of this philosophical attitude, it is claimed, is that it permits an interpretation of the subject of psychic processes as the product of psychic activity and not as an object to be thought of as a substance possessing certain powers and capacities which are its characteristics.

While the philosophy of subjective spirit really only attracted attention up to the middle of the nineteenth century, Hegel's philosophy of objective spirit, in other words his theory of law and politics, received a great deal of attention during the nineteenth and especially the twentieth century. This was not only because of the theory's great importance for the Marxist and other anti-liberalistic social theories (see Marx, K.; Western Marxism). It has also repeatedly been the object of violent controversy, especially because of its political implications. In all its versions, Hegel's political philosophy rests on three main convictions which he cherished from his early years and held for the rest of his life. The first is that every modern philosophy of law and politics must incorporate the conception of freedom which was central to the European Enlightenment, and in particular to that of Germany (see Enlightenment, Continental). The second is that, especially in the case of modern political philosophy, the insight that the whole takes priority over its separate parts, an insight formulated by Aristotle in his *Politics*, must be maintained and brought up to date. Finally, the third conviction consists in an application of the principle which shapes Hegel's whole philosophical enterprise, namely, that political philosophy must play its part in the confirmation of the thesis that only reason is real. Hegel attempts to do justice to these three convictions within the framework of his theory of objective spirit by (1) introducing an extravagant conception of freedom, (2) identifying the whole of Aristotle with the phenomenon which he calls 'ethical life' and (3) declaring this phenomenon called 'ethical life' to be the 'reality of reason'.

Hegel fulfils his self-imposed demand for the integration of freedom by making the conception of free will the fundamental concept of his philosophy of the objective spirit; this is where his characteristic conception of freedom comes into play. According to Hegel, a will is free not because it can choose its ends from a virtually limitless number of objective alternatives; the truly free will is the will which only determines itself. For Hegel,

self-determination means to refer willingly to oneself, that is, to will oneself. Thus he thinks of freedom as a case of self-reference and in this way assimilates it into his concept of cognition, which is also based on the idea of self-reference (see §5 above). This assimilation is utterly intentional on Hegel's part, because it gives him the opportunity to interpret the process of the systematic unfolding of the various determinations of the will not only as different ways of the realization of free will but also as a process of cognition (see Freedom and liberty; Free will).

Against this background, Hegel first develops his theories of law and morality, which derive all legal relationships and the obligatory character of moral acts from the concept of free will. In his theory of law, Hegel makes his contribution to the discussion of the philosophical foundations of civil and criminal law. His basic thesis is that property, the acquisition and use of which is a presupposition for being able to act freely, is the necessary condition of law in all its different variations. In his theory of morality, Hegel discusses the moral behaviour of autonomous subjects under the aspect of the gaining of moral standpoints for the purpose of judging actions and of the conversion of moral goals into actions. According to Hegel, however, legal relationships and moral standards are *founded* in social institutions. He thinks of these institutions as forms of what he calls 'ethical life' (*Sittlichkeit*). In Hegel's language, ethical life as the basis for the possibility of law and morality is the truth of free will, that which free will really is. Since it is a characteristic of the truth of free will to be real, it follows that, for Hegel, ethical life is also the reality of free will. This reality is thus the 'presupposed whole', without reference to which the discussion of law and morality makes no sense at all. This thesis of the function of real ethical life as the basis for law and morality is intended to account for the Aristotelian maxim of the primacy of the whole in political philosophy (see Legal idealism §§1-2).

For Hegel, ethical life appears in three institutional forms: family, bourgeois society and the state. The theory of the family contains his thoughts on the ethical function of marriage, his justification of monogamy, his views on family property and the laws of inheritance and his maxims for bringing up children. The theory of bourgeois society became well-known and influential, above all because of Hegel's diagnosis of the difficulties which will arise within a society based solely on economic interests and elementary needs of its individual members. This diagnosis is grounded in Hegel's analyses of a society founded solely on economic relationships. They owe much to the works on political economy by Adam Smith, J.P. Say and Ricardo, to whom Hegel often explicitly refers. According to Hegel, a bourgeois society considered as an economic community is defined by the fact that in it people can satisfy their needs through labour. The manifold nature of these needs means that they can only be satisfied by division of labour within the society. This leads economic subjects to join together into estates (*Stände*) and corporations whose members each undertake specific tasks with regard to the socially organized satisfaction of their needs. Hegel recognizes three estates: the peasant estate, which he calls the 'substantial estate'; the tradesmen's estate, among which he includes craftsmen, manufacturers and traders; and what he calls the 'general estate', whose members fulfil judicial and policing functions. Corporations are formed mainly in the tradesmen's estate. Although this entire realm of bourgeois society organized along these lines does involve legal restrictions, and is regulated by a civil and criminal legal code, it none the less cannot remain indefinitely stable. For it is not possible to prevent the polarization of the poor majority and the rich minority which leads to overpopulation, so that eventually the entire social wealth will not suffice to satisfy even the most elementary needs of all. The consequences are colonization and the formation of the 'proletariat'. Both will tend to destroy this bourgeois society.

If one follows Hegel's arguments, bourgeois society can only avoid this fate if its members act not according to their own particular interests and needs, but recognize the state as their 'general purpose', and direct all their activities to maintaining it (see Civil society §1; State). Hegel thinks of the state as a constitutional monarchy with division of power. For Hegel, the constitution of a state is in no sense the product of some constitution-creating institution or the work of individual persons. It is 'absolutely essential that the constitution, although the product of past history, *should not be seen as a finished entity*'. A constitution is rather the manifestation of the spirit of a people, created during the course of history through their customs and traditions. This view permits Hegel to maintain on the one hand that each people has the constitution 'which is appropriate to it and fits it', and on the other to insist that there is not much leeway for the modification of constitutions. The constitutional form of a reasonably organized state must be a monarchy because its characteristic individuality can only be appropriately represented by a concrete individual to whom as a person the sovereign acts of the state can be attributed. Hegel

also favours a hereditary monarchy, since he sees the process of determining a person as monarch by virtue of its origin as the method which is least dependent upon arbitrary decisions. Hegel's theory of the powers of the state (*Staatsgewalten*) recognizes, in addition to the princely power (*fürstliche Gewalt*) which represents the instance of ultimate decision-making within the constitutional framework, the governmental power (*Regierungsgewalt*) and the legislative power (*gesetzgebende Gewalt*). It is the task of the governmental power, which for Hegel also includes the judicial power, to pursue the general interests of the state, ensure the maintenance of right and enforce the laws. The legislative power is responsible for the 'further determination' of the constitution and laws. It is executed by an assembly of the estates which is divided into two chambers. The first chamber consists of a certain group of powerful landowners chosen by virtue of their birth; the second chamber comprises representatives of the corporate associations of the bourgeois society, who are sent to the assembly by their various corporations. Thus in Hegel's model state, both chambers are constituted without the direct political involvement of the population. Hegel's theory of the state provoked considerable controversy, particularly during his own time, because of its resolute defence of the hereditary monarchy and its strongly anti-democratic characteristics in all questions concerning the political representation of the citizens of the state. It was this section of his political philosophy in particular which, as early as the mid-nineteenth century, gave rise to the statement that Hegel was the philosopher of the Prussian state.

Hegel forges the link to his theory of the spirit, which contains his political philosophy, by interpreting what he calls 'ethical life' as the 'spirit of a people'. This allows him to elaborate his conception of history on the one hand and on the other to introduce his theory of the absolute spirit. The philosophy of history is introduced by the idea that ethical life as the reality of free will takes on different forms for different peoples. These forms differ from each other in the degree to which the different institutions of ethical life are actually developed. Now, Hegel believes that this development has taken place during the course of a historical process which he calls 'world history' (see History, philosophy of). This process of world history, which he sees as 'progress in the consciousness of freedom', can be divided into four distinct epochs, which correspond to four 'empires of world history'. Hegel describes this process of world history as beginning with the 'Oriental Empire', which is followed by the 'Greek' and then by the 'Roman' empires. The process is brought to a conclusion by the 'Germanic Empire', in which the 'Germanic peoples are given the task of accomplishing the principle of the unity of divine and human nature, of reconciling…bjective truth and freedom'. Hegel now interprets this reconciliation as the conclusion of the process of the self-recognition of reason. The result of this process consists of the insight that reason *knows itself* to be the whole of reality. Thus Hegel links the theory of the objective spirit with his metaphysics of reason and can now concentrate on the various aspects of this self-knowledge of reason as a theory of absolute spirit.

Hegel's philosophy of absolute spirit contains his philosophy of art, his philosophy of religion and his theory of philosophy. Although from the very first all these subjects had a fixed place in Hegel's attempts at a system, and although his philosophies of art and religion were to become very influential (the one in the history of art and the theory of aesthetics and the other in theology), none the less these sections of Hegel's philosophy are relatively little elaborated in the works published by Hegel himself. Apart from a few sketch-like hints in his first work, *Difference between the Systems of Fichte and Schelling* (see §3 above), and the two final chapters of the *Phenomenology of Spirit*, Hegel devoted only a few paragraphs to these themes at the end of the *Encyclopedia of Philosophical Sciences*. We can gather from these paragraphs that there are three different ways in which reason, which knows itself, relates to itself; these are manifested in art, religion and philosophy. They differ from each other in the way in which in each of these ways reason knows itself. In art, reason relates to itself *intuitively* or, as Hegel says, *knows itself immediately*, while in religion this knowing relationship with itself realizes itself in the form of *representation*, which is linked with the sublation of the immediacy of knowledge. In philosophy, the self-reference of reason is accounted for in the mode of *cognition*. The theory of epistemic modes which underpins this functional analysis of art, religion and philosophy, though obviously relying on the results of the Hegelian theory of the subjective spirit, none the less contains a number of difficulties which are hard to unravel.

Against this background of different forms of knowledge, Hegel first reveals his theory of art in the form of a theory of styles of art (*Kunstformen*) and of individual arts (*Kunstarten*). He recognizes three different styles of art, which he calls symbolic, classical and romantic. They differ from each other in their various means of expressing the distinguishing characteristics of the spiritual, which belong to the sensible and therefore to the intuitive

manifestations which reason gives itself. These styles themselves are characterized by the ways in which a spiritual content presents itself as the meaning of a sensible object. The symbolic style of art is thus the one in which the relationship between meaning and sensible appearance is relatively contingent, since it only arises through a randomly chosen attribute. By way of example Hegel takes the lion, which symbolizes strength. In the classical style of art the sensible appearance expresses adequately what it is intended to signify. For Hegel, the human figure serves as a paradigm for this adequate symbolization of the spiritual, especially in the way in which it is represented in sculpture and painting. Finally, the romantic style of art takes as its subject the representation of the ‘self-conscious inwardness’ of the spirit. In it, the emotional world of the subject is expressed by reference to sensible characteristics. Hegel interprets the various individual arts as realizations of styles of art in various materials. Although each individual art can present itself in each style of art, there is for each individual art an ideal style, which he calls its basic type. The first individual art which Hegel discusses is architecture. Its task is to deal with nonorganic nature in an artful manner. Its basic type is the symbolic style of art. The second individual art is sculpture, the basic type of which is the classical style. Sculpture aims to transform nonorganic nature into the physical form of the human body. The remaining individual arts are painting, music and poetry, whose basic type is represented by the romantic style of art. Painting marks the beginning of the separation of the direct processing of natural materials and thus a certain intellectualization of matter, which makes it capable of representing feelings, emotions, etc. Music is the romantic style of art *par excellence*. Its material is sound, which is matter only in a figurative sense and is therefore particularly suitable for the representation of even the most fleeting affects. Finally poetry, the last of the romantic arts, has as its material only signs, which here play no part as material entities, but as bearers of meaning. These meanings refer to the realm of imagination and other spiritual content, so that in poetry a spiritual content can be presented in a manner appropriate to its spirituality. Hegel could not resist the temptation to use his theory of individual arts and styles of art as a model for the interpretation of the history of the development of art. His historicizing of individual arts and styles of art played a significant role in making the concept of an epoch an important tool in the history of art.

In the philosophy of religion Hegel holds that only in Christianity are the conditions fulfilled which are characteristic of the representational self-knowledge of reason. Philosophy of religion has as its subject not only God, but also religion itself, and for Hegel that means the way in which God is present in the religious consciousness. By this characterization he aims to distinguish philosophy of religion from the traditional *theologia naturalis*. On the basis of the two components which make up its nature, the philosophy of religion attempts in the first instance to characterize more closely the concept of God and the various kinds of religious consciousness which Hegel takes to be feeling, intuition and representation. This will be found in the first part of the philosophy of religion, which thematizes the ‘concept of religion’. The second part of the philosophy of religion discusses what Hegel calls ‘determinate religion’. Here, he is concerned with something resembling a phenomenology of religions, the exposition of their various forms of appearance and objectivizations. This exposition starts with so-called natural religion, which according to Hegel assumes three forms: the religion of magic, the religion of substantiality and the religion of abstract subjectivity. The specific characteristic of natural religion is that it thinks of God in direct unity with nature. Natural religion finds its historical concept in the Oriental religions. Hegel regards the ‘religions of spiritual individuality’ as a second stage; these assume the forms of the religion of sublimity, the religion of beauty and the religion of teleology. At this stage, God is regarded as the primary spiritual being, which is not only nature but which also rules over and determines nature. Hegel puts the Jewish, Greek and Roman religions in this category. Finally, the third stage represents the ‘perfect religion’, to the discussion of which he devotes the third section of his philosophy of religion. In it, God is presented as He in reality is, namely the ‘infinite, absolute end in itself’. To the religious consciousness, the God of the perfect religion appears in the trinitarian form as the unity of the Father, Son and Holy Ghost. According to Hegel, this idea of religion was first realized adequately in Christianity. Hegel’s philosophy of religion greatly influenced theological discussions and points of view. None the less, it was not without its critics, for whom it represented a theory which, as, for example, R. Haym claimed in the last century, contributed to the dissolution of the Godly in reason and of Piety in knowledge.

As far as philosophy is concerned, Hegel maintains that its distinguishing mode of knowledge, namely cognition, is present when something is seen to be necessary. Since reason within the sphere of the absolute spirit relates only to itself, the achievement of the cognitive reference of reason to itself lies in the fact that it understands the progress of its realization in logic, nature and spirit as a necessary process. Philosophy is the representation of this

process in its necessity. This philosophical process also has its appearance in time in the form of the history of philosophy. For Hegel, the history of philosophy presents itself as a historical succession of philosophical positions in which in each case one of the essential characteristics of (Hegelian) reason is made the principle of a philosophical interpretation of the world in a one-sided and distorted way that is characteristic of its time. He sees the existence of political freedom as a necessary precondition for a philosophical interpretation of the world. Only in societies in which free constitutions exist can philosophical thought develop. Since, he claims, the concepts of freedom and constitution only arose as the products of Greek (that is, occidental) thought, philosophical discourse is really a specifically Western achievement. He therefore absolutely refuses to ascribe any philosophically relevant intellectual achievements to the Oriental world, the principle proponents of which are in his view China and India. All the doctrines of wisdom of the Orient can at most be accepted as codifications of religious ideas. If, for a Westerner, some of these doctrines none the less seem to express a philosophical thought, this is because they confuse the abstract generality of Oriental religious ideas with the generality which is applicable to the thoughts of reason engaged in thinking itself. Hegel divides Western philosophy into two main periods: Greek and Germanic philosophy. Up to a certain point, Greek philosophy also includes Roman, and Germanic philosophy includes not only German philosophy but that of other European peoples as well, since these peoples have ‘in their totality a Germanic culture’. The difference between Greek and Germanic philosophy lies in the fact that Greek philosophy was not yet in a position to comprehend the conception of spirit in all its profundity. This only became possible through Christianity and its acceptance throughout the Germanic world. For only in this historical context was it possible for the insight to establish itself that the essence of spirit is subjectivity and hence knowledge of itself. Hegel regards it as a great merit of his philosophy that it adequately explains this, and thus reconciles reason with reality in thought. In the last analysis, his message consists of a single proposition: Reason is and knows itself to be the ultimate reality. His system is brought to a conclusion in what is, in his view, a successful justification of that proposition. Even during the nineteenth century, the optimism of reason underlying Hegel’s system aroused criticism, for example, from Nietzsche and the representatives of Neo-Kantianism. It seems doubtful whether, at the end of the twentieth century, Hegel’s indomitable faith in reason can continue to convince.

See also Absolute, the; German idealism; Hegelianism; Neo-Kantianism

Translated from the German by Jane Michael-Rushmer
ROLF-PETER HORSTMANN

List of works

German editions of major works

Hegel, G.W.F. (1968-) *Gesammelte Werke*, ed. von der Rheinisch-westfälischen Akademie der Wissenschaften, Hamburg: Meiner.(This is the historical-critical edition in progress which aims at completeness. Up to the end of 1996 fifteen volumes have appeared.)

Hegel, G.W.F. (1832-45) *Werke. Vollständige Ausgabe durch einen Verein von Freunden des Verewigten*, 18 vols in 21, Berlin: Duncker & Humblot.(This is the first edition - inaugurated by an association of friends of Hegel - which contains a large amount of material from Hegel’s lectures. Of particular importance are vol. 9 *Lectures on the Philosophy of History*, vols 10, 1-3 *Lectures on Aesthetics*, vols 11-12 *Lectures on the Philosophy of Religion*, vols 13-15 *Lectures on the History of Philosophy*.)

Hegel, G.W.F. (1970) *Werke in 20 Bänden (Theorie-Werkausgabe)*, ed. E. Moldenhauer and K.M. Michel, Frankfurt: Suhrkamp.(This is a modified and slightly enlarged reprint of the 1832-45 edition.)

Hegel, G.W.F. (*c*.1793-1800) *Theologische Jugendschriften*, ed. H. Nohl, Tübingen: J.C.B. Mohr, 1907, reprinted 1968.(Collection of manuscripts and fragments written before 1801.)

Hegel, G.W.F. (1801a) *De Orbitis Planetarum*, Jena: Seidler.(Hegel’s Habilitation Essay.)

Hegel, G.W.F. (1801b) *Differenz des Fichte’schen und Schelling’schen Systems der Philosophie*, Jena: Seidler. (Hegel’s first philosophical publication.)

Hegel, G.W.F. (1802) *Glauben und Wissen oder die Reflexionsphilosophie der Subjektivität*, in *Kritisches Journal der Philosophie*, ed. F.W.J. Schelling and G.W.F. Hegel, Tübingen: Cotta.(Contains his early criticism of Kant, Jacobi and Fichte.)

Hegel, G.W.F. (*c*.1803-4) *Jenaer Systementwürfe I*, newly ed. K. Düsing and H. Kimmerle, Hamburg: Meiner, 1986.(Contains fragments of the first system draft.)

Hegel, G.W.F. (*c*.1804-5) *Jenaer Systementwürfe II*, newly ed. R.P. Horstmann, Hamburg: Meiner, 1982. (Documents his early conception of logic and metaphysics and contains a fragmentary version of his philosophy of nature.)

Hegel, G.W.F. (*c*.1805-6) *Jenaer Systementwürfe III*, newly ed. R.P. Horstmann, Hamburg: Meiner, 1987. (Incomplete versions of his philosophy of nature and of spirit.)

Hegel, G.W.F. (1807) *System der Wissenschaft. Erster Theil, die Phänomenologie des Geistes*, Bamberg & Würzburg: Goebhardt.(The famous introduction to the system.)

Hegel, G.W.F. (1812-16) *Wissenschaft der Logik*, Nürnberg: Schrag, 3 vols.(The most extensive elaboration of his metaphysical logic.)

Hegel, G.W.F. (1821) *Naturrecht und Staatswissenschaft im Grundrisse. Grundlinien der Philosophie des Rechts*, Berlin: Nicolai.(Contains his theory of law and morality as well as his social and political philosophy.)

Hegel, G.W.F. (1830) *Enzyklopädie der philosophischen Wissenschaften im Grundrisse. Dritte Ausgabe*, Heidelberg: Winter.(The last and most comprehensive version of his system.)

Hegel, G.W.F. (1952-81) *Briefe von und an Hegel*, ed. J. Hoffmeister and F. Nicolin, Hamburg: Meiner, 4 vols.(A richly commented edition of his correspondence.)

Hegel, G.W.F. (1983-) *Vorlesungen. Ausgewählte Nachschriften und Manuskripte*, Hamburg: Meiner.(Contains a large number of manuscripts which are based on student notes of Hegel’s lecture courses. Many of these manuscripts have been discovered only recently.)

1 English translations

Hegel, G.W.F. (1892-6) *Lectures on the History of Philosophy*, trans. E.S. Haldane and F.H. Simson, London: Kegan, Paul, Trench, Trübner, 3 vols; reprinted, Atlantic Highlands, NJ: Humanities Press, 1983.(Translation of vols 13-15 of Hegel (1832-45).)

Hegel, G.W.F. (1948) *Early Theological Writings*, trans. T.M. Knox, Chicago, IL: University of Chicago Press. (Translation of Hegel (*c*.1793-1800).)

Hegel, G.W.F. (1969) *Science of Logic*, trans. A.V. Miller, London: Allen & Unwin; repr. Atlantic Highlands, NJ: Humanities Press, 1993.(Translation of Hegel (1812-16).)

Hegel, G.W.F. (1970a) *Phenomenology of Spirit*, trans. A.V. Miller, Oxford: Clarendon Press. (Translation of Hegel (1807).)

Hegel, G.W.F. (1970b) *Philosophy of Nature*, trans. and ed. M.J. Petry, London: Allen & Unwin, 3 vols. (Translation of part 2 of the *Enzyklopädie der philosophischen Wissenschaften im Grundrisse* in the version presented as vol. 7, 1 in Hegel (1832-45). With informative introduction and excellent explanatory notes. Also translated by A.V. Miller, Oxford: Clarendon Press, 1970.)

Hegel, G.W.F. (1975) *Aesthetics*, trans. T.M. Knox, Oxford: Clarendon Press, 2 vols. (Translation of vols 10, 1-3 of Hegel (1832-45).)

Hegel, G.W.F. (1977a) *Faith and Knowledge*, trans. and ed. W. Cerf and H.S. Harris, Albany, NY: State University of New York Press.(Translation of Hegel (1802).)

Hegel, G.W.F. (1977b) *The Difference between Fichte’s and Schelling’s System of Philosophy*, trans. and ed. H.S. Harris and W. Cerf, Albany, NY: State University of New York Press.(Translation of Hegel (1801b).)

Hegel, G.W.F. (1978) *Philosophy of Subjective Spirit*, trans. and ed. M.J. Petry, Dordrecht and Boston, MA: Reidel, 3 vols.(Contains a translation of the first section of part 3 of the *Enzyklopädie der philosophischen Wissenschaften im Grundrisse* in the version printed as vol. 7, 2 in Hegel (1832-45). With introduction and explanatory notes.)

Hegel, G.W.F. (1984) *Letters*, trans. C. Butler and C. Seiler, Bloomington, IN: Indiana University Press.(Contains a selection from Hegel (1952-81).)

Hegel, G.W.F. (1984-7) *Lectures on the Philosophy of Religion*, trans. C.P. Hodgson and R.F. Brown, Los Angeles, CA: University of California Press, 3 vols.(Translation of vols 3-5 of Hegel (1983-).)

Hegel, G.W.F. (1986) *The Jena System of 1804-5: Logic and Metaphysics*, trans. J. Burbidge *et al.*, Kingston and Montreal: McGill-Queen’s Press.(Contains part of Hegel (*c*.1804-05).)

Hegel, G.W.F. (1988) *Introduction to the Philosophy of History*, trans. L. Rauch, Indianapolis, IN: Hackett Publishing Company.(Contains material from vol. 9 of Hegel (1832-45).)

Hegel, G.W.F. (1990) *Encyclopedia of the Philosophical Sciences in Outline, and Critical Writings*, ed. E. Behler, New York: Continuum.(Contains a translation of the first edition of the *Encyclopedia*.)

Hegel, G.W.F. (1991a) *The Encyclopaedia Logic*, trans. T.F. Geraets, W.A. Suchting and H.S. Harris, Indianapolis: Hackett Publishing Company.(Translation of part 1 of *Enzyklopädie der philosophischen Wissenschaften im Grundrisse* as printed in vol. 6 of Hegel (1832-45).)

Hegel, G.W.F. (1991b) *Elements of the Philosophy of Right*, trans. H.B. Nisbet, ed. A. Wood, Cambridge: Cambridge University Press.(Translation of Hegel (1821).)

References and further reading

Beiser, F.C. (ed.) (1993) *The Cambridge Companion to Hegel*, Cambridge: Cambridge University Press. (Collection of essays on various topics of Hegel's philosophy.)

Dickey, L. (1987) *Hegel: Religion, Economics and Politics of the Spirit, 1770-1807*, Cambridge: Cambridge University Press.(Social, religious and political background of Hegel's early philosophy.)

Düsing, K. (1983) *Hegel und die Geschichte der Philosophie (Hegel and the History of Philosophy)*, Darmstadt: Wissenschaftliche Buchgesellschaft.(Focuses on Hegel's interpretation of important positions in ancient and modern philosophy.)

Fulda, H.F. (1975) *Das Problem einer Einleitung in Hegels Wissenschaft der Logik (The Problem of an Introduction to Hegel's Science of Logic)*, Frankfurt: Klostermann, 2nd edn.(An essay on the role of the *Phenomenology of Spirit* in Hegel's system.)

Hardimon, M. (1994) *The Project of Reconciliation: Hegel's Social Philosophy*, Cambridge: Cambridge University Press.(Hegel's views on the relation between the individual and the modern social world.)

Harris, H.S. (1972-83) *Hegel's Development*, vol. 1 *Toward the Sunlight 1770-1801*, vol. 2 *Night Thoughts. Jena 1801-1806*, Oxford: Oxford University Press.(Hegel's intellectual development in the light of his early writings.)

Haym, R. (1857) *Hegel und seine Zeit (Hegel and his Age)*, Berlin: Gärtner; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1962.(Very important critical study of Hegel's system and its development.)

Henrich, D. (1971) *Hegel im Kontext (Hegel in Context)*, Frankfurt: Suhrkamp. (Influential essays on logical themes.)

Horstmann, R.P. (1990) *Wahrheit aus dem Begriff (Truth from the Concept)*, Frankfurt: Anton Hain.(Deals with metaphysical assumptions of Hegel's philosophy.)

Inwood, M. (1983) *Hegel*, London: Routledge & Kegan Paul.(Critical discussion of central topics.)

McTaggart, J. (1910) *A Commentary on Hegel's Logic*, Cambridge: Cambridge University Press.(Still very informative.)

McTaggart, J. (1922) *Studies in the Hegelian Dialectic*, Cambridge: Cambridge University Press.(A well-informed contribution to Hegel's methodology.)

Pinkard, T. (1994) *Hegel's Phenomenology. The Sociality of Reason*, Cambridge: Cambridge University Press.(A penetrating interpretation of the aim and the structure of Hegel's most influential book.)

Pippin, R.B. (1989) *Hegel's Idealism: The Satisfactions of Self-Consciousness*, Cambridge: Cambridge University Press.(Concentrates on Hegel's epistemology and its Kantian sources.)

Rosen, M. (1982) *Hegel's Dialectic and its Criticism*, Cambridge: Cambridge University Press.(Methodological problems of Hegel's metaphysics.)

Rosenkranz, K. (1844) *Hegel's Leben (Hegel's Life)*, Berlin: Duncker & Humblot. (The first biography of Hegel.)

Scholz, H. (1921) *Die Bedeutung der Hegelschen Philosophie für das philosophische Denken der Gegenwart (The Relevance of Hegel's Philosophy for Contemporary Philosophical Thought)*, Berlin: de Gruyter.(Contains a harsh criticism of Hegel's endeavour.)

Stern, R. (ed.) (1993) *G.W.F. Hegel: Critical Assessments*, London: Routledge & Kegan Paul, 4 vols.(A voluminous collection of essays on Hegel.)

Taylor, C. (1975) *Hegel*, Cambridge: Cambridge University Press.(The best and most comprehensive introduction to Hegel for the English speaking reader.)

Theunissen, M. (1978) *Sein und Schein. Die kritische Funktion der Hegelschen Logik (Being and Appearance. The critical function of Hegel's Logic)*, Frankfurt: Suhrkamp. (A very detailed account of central ideas of Hegel's Logic.)

Wood, A. (1990) *Hegel's Ethical Thought*, Cambridge: Cambridge University Press. (A modern interpretation of Hegel's moral and social philosophy.)

Hegelianism

As an intellectual tradition, the history of Hegelianism is the history of the reception and influence of the thought of G.W.F. Hegel. This tradition is notoriously complex and many-sided, because while some Hegelians have seen themselves as merely defending and developing his ideas along what they took to be orthodox lines, others have sought to 'reform' his system, or to appropriate individual aspects and overturn others, or to offer consciously revisionary readings of his work. This makes it very hard to identify any body of doctrine common to members of this tradition, and a wide range of divergent philosophical views can be found among those who (despite this) can none the less claim to be Hegelians.

There are both 'internal' and 'external' reasons for this: on one hand, Hegel's position itself brings together many different tendencies (idealism and objectivism, historicism and absolutism, rationalism and empiricism, Christianity and humanism, classicism and modernism, a liberal view of civil society with an organicist view of the state); any balance between them is hermeneutically very unstable, enabling existing readings to be challenged and old orthodoxies to be overturned. On the other hand, the critical response to Hegel's thought and the many attempts to undermine it have meant that Hegelians have continually needed to reconstruct his ideas and even to turn Hegel against himself, while each new intellectual development, such as Marxism, pragmatism, phenomenology or existential philosophy, has brought about some reassessment of his position. This feature of the Hegelian tradition has been heightened by the fact that Hegel's work has had an impact at different times over a long period and in a wide range of countries, so that divergent intellectual, social and historical pressures have influenced its distinct appropriations. At the hermeneutic level, these appropriations have contributed greatly to keeping the philosophical understanding of Hegel alive and open-ended, so that our present-day conception of his thought cannot properly be separated from them. Moreover, because questions of Hegel interpretation have so often revolved around the main philosophical, political and religious issues of the nineteenth and twentieth centuries, Hegelianism has also had a significant impact on the development of modern Western thought in its own right.

As a result of its complex evolution, Hegelianism is best understood historically, by showing how the changing representation of Hegel's ideas have come about, shaped by the different critical concerns, sociopolitical conditions and intellectual movements that dominated his reception in different countries at different times. Initially, Hegel's influence was naturally most strongly felt in Germany as a comprehensive, integrative philosophy that seemed to do justice to all realms of experience and promised to preserve the Christian heritage in a modern and progressive form within a speculative framework. However, this position was quickly challenged, both from other philosophical standpoints (such as F.W.J. Schelling's 'positive philosophy' and F.A. Trendelenburg's neo-Aristotelian empiricism), and by the celebrated generation of younger thinkers (the so-called 'Young' or 'Left' Hegelians, such as Ludwig Feuerbach, David Strauss, Bruno Bauer, Arnold Ruge and the early Karl Marx), who insisted that to discover what made Hegel a truly significant thinker (his dialectical method, his view of alienation, his 'sublation' of Christianity), this orthodoxy must be overturned. None the less, both among these radicals and in academic circles, Hegel's influence was considerably weakened in Germany by the 1860s and 1870s, while by this time developments in Hegelian thought had begun to take place elsewhere.

Hegel's work was known outside Germany from the 1820s onwards, and Hegelian schools developed in northern Europe, Italy, France, Eastern Europe, America and (somewhat later) Britain, each with their own distinctive line of interpretation, but all fairly uncritical in their attempts to assimilate his ideas. However, in each of these countries challenges to the Hegelian position were quick to arise, partly because the influence of Hegel's German critics soon spread abroad, and partly because of the growing impact of other philosophical positions (such as Neo-Kantianism, materialism and pragmatism). Nevertheless, Hegelianism outside Germany proved more durable in the face of these attacks, as new readings and approaches emerged to counter them, and ways were found to reinterpret Hegel's work to show that it could accommodate these other positions, once the earlier accounts of Hegel's metaphysics, political philosophy and philosophy of religion (in particular) were rejected as too crude.

This pattern has continued into the twentieth century, as many of the movements that began by defining themselves against *Hegel (such as Neo-Kantianism, Marxism, existentialism, pragmatism, post-structuralism and even 'analytic' philosophy) have then come to find unexpected common ground, giving a new impetus and depth to*

Hegelianism as it began to be assimilated within and influenced by these diverse approaches. Such efforts at rapprochement began in the early part of the century with Wilhelm Dilthey's attempt to link Hegel with his own historicism, and although they were more ambivalent, this connection was reinforced in Italy by Benedetto Croce and Giovanni Gentile. The realignment continued in France in the 1930s, as Jean Wahl brought out the more existentialist themes in Hegel's thought, followed in the 1940s by Alexander Kojève's influential Marxist readings. Hegelianism has also had an impact on Western Marxism through the writings of the Hungarian Georg Lukács, and this influence has continued in the critical reinterpretations offered by members of the Frankfurt School, particularly Theodor W. Adorno, Max Horkheimer, Herbert Marcuse, Jürgen Habermas and others. More recently, most of the major schools of philosophical thought (from French post-structuralism to Anglo-American 'analytic' philosophy) have emphasized the need to take account of Hegel, and as a result Hegelian thought (both exegetical and constructive) is continually finding new directions.

1 The Hegelian School in Germany 1816-40

Initially, Hegel's influence was naturally most strongly felt in Germany, and can be seen in the relatively rapid formation within the philosopher's lifetime of something like a 'Hegelian school'. The representatives of this school procured a considerable influence for themselves not only through the personal prestige of Hegel, but also through the foundation of important journals more or less expressly designed to propagate and disseminate the philosophical principles of Hegel himself and apply them to central theoretical and practical issues of the day. But the very comprehensiveness and richness of Hegel's systematic synthesis placed his more original students in an ambiguous and paradoxical position. Eduard Gans wrote 'Hegel has left behind a number of gifted students but no successor. For philosophy has now for the first time completed the cycle of its existence; further advance can only be expected as the further intelligent penetration of the material of knowledge.'

One of the earliest explicit champions of Hegel's thought was Georg Andreas Gabler (1786-1853), a student from Hegel's Jena period 1801-07, who later succeeded to Hegel's chair in Berlin (1835) and was one of the few students to write intensively on (part of) the *Phenomenology*, with the*Kritik des Bewußtseins (Critique of Consciousness)* (1827). When Hegel moved to take up his first chair in Heidelberg in 1816 he also found an ardent supporter in the theologian Karl Daub (1765-1836), who expounded a thoroughly Hegelian approach to religious questions with *Die dogmatische Theologie jetziger Zeit (The Dogmatic Theology of Our Times)* (1833). But it was essentially during his final Berlin period (1819-31) that Hegel began to develop a proper 'School' around him, with the founding of a *Gesellschaft für wissenschaftliche Kritik* (Society for Scientific Criticism) in 1825 and the consequent launching of the journal *Jahrbücher der wissenschaftlichen Kritik (Yearbook of Scientific Criticism)* under the editorship of Hegel and his more prominent students. The journal explicitly began to disseminate a Hegelian line on contemporary philosophical and cultural issues and was soon dubbed the 'Hegel newspaper' by its opponents.

Other followers at this time who produced Hegelian interpretations in the fields of ethics, history of philosophy, speculative theology, law and political thought were Leopold von Henning (1791-1866) with his *Prinzipien der Ethik in historischer Entwicklung (Principles of Ethics in Historical Development)* (1824), Karl Ludwig Michelet (1801-93) with the *Geschichte der letzten Systeme der Philosophie in Deutschland (History of the Most Recent Systems of Philosophy in Germany)* (1837-38), Philipp Karl Marheinecke (1780-1846) with *Die Grundlehren der christlichen Dogmatik als Wissenschaft (The Fundamental Doctrines of Christian Dogmatics as Science)* (1827) and, one of the most interesting and original, Eduard Gans (1798-1839). Gans had become a friend of Hegel's in Heidelberg and strongly under his influence produced his major work *Das Erbrecht in weltgeschichtlicher Entwicklung (The Law of Inheritance Considered in its World-Historical Development)* (1824-35) which forcefully pursued Hegel's own criticism of the 'Historical School' of jurisprudence defended by Karl von Savigny. Gans also lectured on the philosophy of world history from a liberal-progressive Hegelian perspective as well as upon law and may well have been a powerful influence upon the young Karl Marx who heard him lecture in Berlin in the mid-1830s (see Marx, K. §2). These early protagonists of Hegel's thought are sometimes described as the 'Old Hegelians' because they represented the first generation of the 'School', by contrast with the later so-called 'Young Hegelians' of the 1840s, but the label is often quite uninformative about the substance of their teachings or their political and religious persuasions.

Karl Rosenkranz (1805-1879) was another of these early disciples who remained perhaps most faithful to the

original Hegelian vision but also showed himself an independent thinker in his wide-ranging oeuvre. Rosenkranz consciously strove to defend and re-articulate Hegel's position in all its dialectical complexity and, unlike most of Hegel's followers, laid particular stress upon Hegel's fundamental debt to Kant and aspects of the Enlightenment heritage. Rosenkranz expressed his faith in the Hegelian 'middle' in declaring that 'only all of his students taken together are the equal of Hegel; each one on his own account merely represents a one-sided moment of Hegel' (Rosenkranz 1840a: xxxv).

2 The Critique of Hegelian idealism 1840-70

Rosenkranz's preface to his biography of Hegel, *Hegels Leben (Hegel's Life)* (1844), reveals something of the fervent ideological climate of the early 1840s and reflects the various splits within the Hegelian school which had developed in the previous decade, not to mention the counter-reaction to Hegel's influence in the later work of Schelling, Hegel's former friend and collaborator (see Schelling, F.W.J. §4). For it was during the 1830s that the apparent solidity and impressive unity of Hegel's achievement gradually began to fissure and the potentially centrifugal tendencies of the system revealed themselves under the pressure of significant new social and cultural developments.

These divisions first appeared in theology and the philosophy of religion as Hegel's successors attempted to clarify the contemporary implications of Hegel's famous philosophical appropriation of Christianity as the 'consummate' religion corresponding to the 'absolute' perspective of the speculative system. Nevertheless, it was far from clear how much of what many of Hegel's contemporaries still took to be the essence of Christianity really was preserved and adequately reformulated in Hegel, especially traditional dogmatic beliefs concerning individual immortality and the afterlife, the personal and transcendent God of theism, the uniqueness of the incarnation and the entire eschatological dimension.

The figure who brought the interpretation of Hegel's philosophy of religion to a head under all these aspects was David Friedrich Strauss, whose *Das Leben Jesu, kritisch bearbeitet* (*The Life of Jesus, Critically Examined*) (1835-6) represents a watershed in nineteenth-century religious Protestant thought (see Strauss, D.F. §1). Hegel himself had spoken of religious language in terms of pictorial representation, symbolism and on occasion myth, but it was Strauss who fearlessly subjected the received Gospel accounts to a 'demythologizing' technique and attempted to reveal the intelligible ethical and spiritual truths misleadingly couched in archaic symbolic form in the original texts of the tradition. He not only expressed doubts about the historical verisimilitude of the stories and discounted the miraculous and supernatural elements, but also reinterpreted the idea of special revelation in terms of an unfolding historical revelation and rejected traditional accounts of Christ's uniquely divine status. Thus Strauss brought latent tensions in Hegel's legacy into the open and considerably sharpened the ensuing debate. It is in this theological context that Strauss himself first made the distinction in his *Streitschriften zur Vertheidigung meiner Schrift (Polemical Writings in Defence of My Work)* (1837) between 'right', 'centre' and 'left' positions in the spectrum of Hegelian philosophy: the right held to orthodox tradition in emphasizing divine transcendence, personal deity and the doctrine of immortality; the left dissolved the radical uniqueness and sometimes even the historicity of Christ and adopted a progressive humanistic domestication of Christianity as a social creed not so far removed from the 'religion of humanity' of Auguste Comte, (see §6); while the centre attempted the most difficult task of all, upholding the complexity of the original Hegelian 'middle' and avoiding alike the extremes of traditional theism, romantic pantheism or humanist reduction.

Some of those who attempted to negotiate this path in a sensitive and interesting way, apart from Rosenkranz, fell into neglect once the poles of the ensuing debate had ossified into fixed positions. Thus Ferdinand Christian Baur (1792-1860), although he never considered himself a strict adherent of the 'School' in any of its forms, developed in his *Die christliche Gnosis oder die christliche Religionsphilosophie in ihrer geschichtlichen Entwicklung (Christian Gnosis, or the Christian Philosophy of Religion in its Historical Development)* (1835) a kind of speculative hermeneutic of biblical texts and traditional dogmas that remained closer in certain important respects to Hegel's spirit than the investigations of Baur's pupil Strauss. And Strauss' friend Wilhelm Vatke (1806-82) brought a Hegelian perspective to the study of Judaic thought, a neglected subject at the time, with *Die Religion des alten Testaments (The Religion of the Old Testament)* (1835), and produced detailed work on central religio-philosophical questions with *Die menschliche Freiheit in ihrem Verhältniss zur Sünde und Gnade (Human Freedom in its Relation to Sin and Divine Forgiveness)* (1841). Alois Emanuel Biedermann (1819-1885) was

another thinker who engaged with the theological debates on the left and later continued to exploit Hegelian ideas in the quest for a responsible modern Christology which would avoid the pitfalls of anthropological reduction and antiquated supranaturalism in his *Christliche Dogmatik (Christian Dogmatics)* (1868).

The traditional division between 'right' and 'left', with the 'centre' being largely ignored, is an extremely inadequate intellectual shorthand that threatens to obscure rather than illuminate the complexity of the central issues, especially in the 1830s. For it is really only with the development of a radically secular and increasingly naturalistic worldview in the next couple of decades that the earlier Hegelian positions could globally be labelled as 'right-Hegelian', and it is historically anachronistic to regard thinkers such as Gans and most of Hegel's earlier students as politically 'conservative'. In fact many representatives of the 'School' supported liberal-progressive causes and were not initially disappointed by the revolutionary events of 1848.

From the end of the 1830s and throughout the 1840s the ideological fronts sharpened radically in the context of social and political thought. Thus the continuing concern with 'saving' historical Christianity through philosophy on the part of nearly all the original Hegelians came increasingly to seem an antiquated and regressive debate with the growing importance of radical humanistic political thought as the primary site of opposition to entrenched and anti-liberal state social policies in the period up to 1848. It was symptomatic of this trend when the Polish Count August von Cieszkowski reinterpreted Hegel's philosophy of religion in terms of a secularized eschatological philosophy of history with practical intent in his *Prolegomena zur Historiosophie (Prolegomena to the Wisdom of History)* of 1838. He had concluded that the ultimate logic of Hegel's thought demanded not a contemplative or predominantly theoretical relation to reality but rather a 'philosophy of action' ('praxis'). If, as Hegel had claimed, the future could not be predicted, it could nevertheless be shaped with will and consciousness: the task therefore was no longer to recognize the supposed actuality of reason, but actively to procure a place for the emerging rationality of the future. In emphasizing the open and dynamic element of Hegel's thought, stressing the immanent negativity of the dialectical 'method' at the expense of the apparently static 'system', and in elevating the active will over purely retrospective thought, Cieszkowski epitomized the Young Hegelian approach to Hegel's philosophical legacy. A similar position was adopted by Moses Hess who also preached the transformation of traditional religious ideas into an ethical programme for the future with *Die heilige Geschichte der Menschheit (The Sacred History of Humanity)* (1837).

The remarkable intellectual career of Bruno Bauer vividly illustrates these developments since he began as a protagonist of the theological Hegelian right and subsequently progressed through the centre towards a radically atheistic stance: in his *Die Posaune des jüngsten Gerichts über Hegel den Atheisten und Antichristen* (*The Trumpet of the Last Judgement upon Hegel the Atheist and Antichrist*) (1841), Bauer ventriloquized strategically from an apparently orthodox theological perspective precisely in order to reveal the ultimately heterodox and destructive implications of Hegelian philosophy for traditional Christian belief. These radical developments within the Hegelian school were most clearly registered in the journal founded by Arnold Ruge and T. Echtermeyer in 1838, the *Hallische Jahrbücher für deutsche Wissenschaft und Kunst (Halle Yearbook for German Science and Art)*. Although initially representatives of the whole spectrum of the school published articles in the journal, the general tenor of the contributions soon began to reflect the most advanced position of the left. In this respect the article 'Zur Kritik der Hegelschen Philosophie'(Towards a Critique of Hegelian Philosophy) (1839) by Ludwig proved symptomatic. Indeed it was Feuerbach's influential book *Das Wesen des Christentums* (*The Essence of Christianity*) (1841) which seemed in the eyes of many to draw the ultimate conclusions from Hegel's philosophy of religion and Strauss's development of it by 'unmasking' all theological discourse as an alienated and 'inverted' projection of human imagination and desire. He proposed to reveal through his 'transformational method' that the ultimate truth of theology is anthropology (in the sense that chemistry is the truth of alchemy). This interpretation of religion generally as a compensating 'ideology' has proved enormously influential in modern thought (see Feuerbach, L.A. §2).

Feuerbach also turned his critique of religion against Hegel's philosophy itself, and in particular against his idealism, accusing speculative philosophy of making the same mistake as theology: it prioritizes the infinite over the finite, thought over sense, the abstract over the concrete, and so ends up as a panlogistic idealism which sets essence above existence. This nominalistic attack on Hegel exerted a great influence, and marks the beginning of a turn away from idealism towards a new materialist metaphysics, as the dominant philosophical outlook ceased to be speculative and became anthropological and naturalistic.

3 The critique of Hegelian idealism 1840-70 (cont.)

Under the influence of this critique of Hegel’s idealism, those who succeeded Feuerbach among the so-called ‘Young Hegelians’ (such as Ruge, Friedrich Engels, Hess and the early Marx) extended it to include Hegel’s political thought, while at the same time this turn towards naturalism was treated as a key to the reinterpretation and radicalization of some of Hegel’s fundamental doctrines. Thus, in the first place, Ruge objected that Hegel’s ‘metaphysics of politics’ lacks a proper critical standpoint because it ‘would offer us the passing realities of history as eternal figures’, and is thereby rendered ‘impotent’: ‘Hegel undertook to present the hereditary monarch, the majority, the bicameral system, etc, as *logical necessities*, whereas it had to be a matter of establishing all these as products of history and of explaining and criticizing them as *historical existences*’ (Ruge (1842: 763) 1983: 228). In a similar vein, Marx accused Hegel of ‘logical, pantheistic mysticism’, of attempting ‘to provide the political constitution with a relationship to the abstract Idea, and to establish it as a link in the life-history of the Idea - an obvious mystification’ (Marx 1975: 69-70). It is evident, therefore, how the turn against Hegel’s idealism decisively influenced the Young Hegelians in their attitude to his *Philosophy of Right* and its place in the speculative system.

In the second place, the Young Hegelians saw the need (in Marx’s famous phrase) to locate properly the ‘rational kernel within the mystical shell’ of Hegel’s philosophy: to rescue what is valuable in Hegel from his idealistic metaphysics. So, for example, Engels argued that Hegel’s dialectical procedure, while apparently based on an abstract logic of concepts, is (as Marx’s work showed) nothing more than a *historical* method, ‘which ultimately amounts to the discovery of the general laws of motion which assert themselves as the ruling ones in the history of human society’ (Engels (1886) 1968: 612). Likewise, Marx himself took Hegel’s analysis of the estrangement between man and nature, based on his conception of nature as the ‘otherness of the idea’, and interpreted this in anthropological terms, as the separation of man from the human process of productive activity. By approaching Hegel in this heterodox manner, the Young Hegelians hoped to recover the radical historicism, humanism and social critique that lay obscured in the empty abstractions of his metaphysical idealism.

If most of the Young Hegelians of radical political persuasion tended to substitute the idea of a new collective humanity or an appropriately transformed ‘species being’ for the spiritual teleology of Hegel’s thought, it was left to Max Stirner (pseudonym of Johann Kaspar Schmidt) to develop the other individualistic extreme of the Hegelian mediation with *Der Einzige und sein Eigentum* (*The Ego and Its Own*) (1845), exalting the sovereign negativity of the singular ego in an almost proto-Nietzschean sense to create and recreate its own value systems and emancipate itself from all heteronomous givenness through tradition and previous history. In drawing the ultimate conclusions from the modern liberal emphasis upon subjective freedom Stirner’s philosophy of the liberated ‘self’ represents the extreme counter-position to Feuerbach’s and Marx’s conception of the ‘social individual’.

Alongside this revolt against idealism brought about by the turn towards naturalism and materialism by the Young Hegelians, Hegel’s alleged panlogicism also came under attack from F.W.J. Schelling and his ‘positive philosophy’, which he adopted from around 1827 until his death in 1854. This position was explicitly conceived in contrast to the ‘negative philosophy’ Schelling claimed to find in Hegel, which is confined to concepts and essences, but neglects being or existence; as a result, it overlooks the fact that it cannot answer the fundamental question ‘Why does anything exist at all? Why is there not nothing?’, and so cannot make the transition from the Idea to nature. Schelling therefore insists that Hegel fails to surmount the ‘nasty broad ditch’ between the first and second parts of the *Encyclopedia*, because concepts are mere abstractions from the empirical world, and so cannot be treated as ideal forms from which the latter can be deduced; on the contrary, the limits of Hegel’s rationalistic metaphysics are shown by the fact that existence must be taken (by us) to be an inexplicable *prius*. In attacking Hegel’s idealism in this way, Schelling began an antirationalistic revolt against his panlogicism which has become one of the fundamental critical reactions to his thought.

Another significant strand in this broadly existentialist critique of Hegel’s idealism which emerged in the 1840s lies in the assertion that Hegel is unable to grasp the reality of becoming, finitude and temporality, despite his talk of movement in his dialectical treatment of the categories. The claim (made, for example, by F.A. Trendelenburg (1802-72) and echoed by Kierkegaard) is that like all idealists, Hegel posits a world of abstract essences behind the world of time and transience, and so fails to give due weight to the reality of finite existence; where Hegel is

deceptive, however, is in the way in which he attributes a dynamic interrelation to the categories, and talks in terms of 'transition', 'development' and 'movement'. Hegel's critics insisted, however, that this talk of movement can only be figurative, and that in fact it is senseless to talk of real change and development in connection with Hegel's Logic. Hegel's followers tried to respond to this wave of anti-idealist criticism: Rosenkranz, for example, insisted in vain that Hegel was not a Platonist, to be 'reproached with offering up the world of blooming life to the idea as to a desolate Hades' (Rosenkranz (1870: 125) 1993 I: 283-4); on the contrary, he argued, Hegel saw universals as more like souls that must be embodied in concrete particulars.

None the less, the effect of this materialist and existentialist critique meant that from around 1860 only the more moderate epistemological idealism of the Neo-Kantians was taken seriously as a systematic philosophy; among those self-confessed Hegelians who remained academically active, the scope of their operations was considerably narrowed, so that Johann Erdmann (1805-92), Eduard Zeller (1814-1908) and Kuno Fischer (1824-1907) are principally known as historians of philosophy. Another figure whose considerable output reflects something of the vicissitudes of the Hegelian tradition in Germany throughout this period is the prolific writer and critic Friedrich Theodor Vischer (1807-87). His earlier works, such as *Über das Erhabene und Komische (On the Sublime and the Comic)* (1837) and the monumental *Ästhetik oder Wissenschaft des Schönen (Aesthetics or the Science of Beauty)* (1845-57), express more or less total commitment to Hegel's philosophy as a whole; but his later contributions represent a progressive abandonment of all ambitious metaphysical claims for art and religion in the modern world in favour of an increasingly sceptical and critical relationship to social reality and to the classical Hegelian project of reconciliation as he had earlier understood it.

4 Hegelianism outside Germany in the nineteenth century: France, Northern Europe and Italy

While Hegelianism in Germany was gradually eclipsed, in several other countries it continued to have an impact into the second half of the nineteenth century. Although in its earlier stages, the reception of Hegel in these countries broke little new ground, none the less an inevitable diversification occurred as Hegel's ideas were taken up in different climates of thought, while Hegel was later read both as part of the broader development of German Idealism, and as closer in outlook to some of his critics. This process has continued into the twentieth century, and has yielded some profound reassessments of his ideas.

France. Although French Hegelianism is best known for its influence on European thought in the 1930s onwards (see §6 below), France was also one of the first countries outside Germany to feel the impact of Hegel's ideas in the nineteenth century, largely due to the efforts of Victor Cousin. Having met Hegel in Heidelberg in 1817, Cousin became an enthusiastic admirer, returning several times to Germany thereafter. He helped give currency to Hegel's ideas through his lectures of 1828-9 at the École Normale in Paris, and with the advent of the July Monarchy in 1830, he was able to acknowledge Hegel's influence explicitly. In his later work, however, he was more guarded, partly due to his growing support for Schelling, and partly due to his increasingly conservative and conformist position. None the less, it was through Cousin that many in France came to know of Hegel's work (such as Pierre-Joseph Proudhon), while he also encouraged others, such as the Italian Augusto Vera (1813-85), who later translated several of Hegel's works into French.

With the advent of the Second Empire in 1852, Cousin lost his official posts, while the growing influence of Auguste Comte meant that the outlook of many thinkers in France became increasingly positivisitic. As a result, Hegel came to be viewed in a new light, as attempts were made to find a fruitful synthesis of both positions, particularly by Ernest Renan (1823-92) and Hippolyte Taine, both of whom had discovered Hegel in the 1840s. Renan sought to develop a less secularized positivism, using Hegel's conception of progress as bringing a divine consciousness into existence through the realization of reason. Taine was likewise attracted to Hegel's idea of a temporal development of reason, and tried to use it to give a historical dimension to the static metaphysics of Spinoza, while fusing the rationalism of the latter with a positivistic recognition of empirical knowledge and apparent contigency.

By the 1850s and 1860s there was a growing awareness of the critical debate surrounding Hegel that had developed in Germany, while Vera attempted to win disciples for the Hegelian cause in France with his *Introduction à la Philosophie de Hegel (Introduction to the Philosophy of Hegel)* (1855), though with little obvious success. Publications by Vera, Rosenkranz and Hegel's critic Rudolf Haym, were reviewed by Edmond Scherer in 1861, who commented that 'Hegel cannot begin to be known, and his philosophy assessed, since there

are no longer any Hegelians' (Scherer 1861: 813). He himself offered an influential assessment of what was valuable in Hegel's thought, emphasizing broadly Left Hegelian themes (such as Hegel's notion of contradiction and historical change), and analysing his *Philosophy of Right* and philosophy of religion (which he considered in relation to D.F. Strauss' *Leben Jesu*). However, positive discussion and dissemination of Hegel's ideas was brought to a halt by the Prussian invasion of France in the 1870s, as (not for the first time) he was blamed for fostering the expansionist nationalism of his country.

The credit for subsequently rehabilitating Hegel in France is usually given to Lucien Herr (1864-1926), who wrote an article on him for the *Grande Encyclopédie* (1893-4). Moreoever, as librarian at the École Normale Superieure from 1886, Herr was able to introduce a large number of philosophy students to Hegel's ideas during this period. In his article, Herr emphasized and appreciated Hegel's systematic ambitions, and placed considerable emphasis on the *Logic*. Though he did not try to resolve any of the philosophical cruxes of his thought, Herr did none the less present a reasonably clear and appealing synopsis of Hegel's views. An equally sympathetic but more partisan view of Hegel, intended as a rebuttal of positivism and Neo-Kantianism, was offered by Georges Noël in his study of Hegel's *Logic*. It is significant, too, that by this time Hegel was being recognized as an important precursor of Marxist thought, and that this rapprochement led to a less panlogistic and quietistic reading of his work (as can be seen in René Bertholet's address to the French Philosophy Society of 1907).

Northern Europe. Around the middle of the nineteenth century, Hegelian ideas had an important impact on the intellectual life of several northern European countries.

In Denmark, the person most responsible for introducing these ideas was the dramatist and man of letters Johan Ludwig Heiberg (1791-1860). Heiberg met Hegel in Berlin in 1824, and in the same year he brought out his *Om den menneskelige Frihed (On Human Freedom)*, in which he makes several references to Hegel, while using distinctly Hegelian ideas and terminology in his dispute with F.G. Howitz over this issue. Heiberg subsequently produced other works that established him as a spokesman for Hegelianism, and in June 1837 he began publication of *Perseus, Journal for den speculative Idee*. At the same time, Hegel's ideas were also being critically discussed by Poul Martin Moller (1794-1838) and Frederich Christian Sibbern (1785-1872). The former left Denmark to occupy a chair at Oslo University from 1826-31, and introduced the study of Hegel into Norway.

Among a slightly younger generation, Hegelian ideas were enthusiastically taken up by Hans Martensen (1808-84) and Rasmus Nielsen (1809-84). Martensen saw Hegel in much the same way as he presented himself - as attempting to bring modern philosophy to its highest standpoint by overcoming all previous one-sided approaches, and as therefore forming the culminating point of philosophical development. Martensen also argued that Christian orthodoxy had nothing to fear from Hegel, whom he followed in seeking to reconcile philosophy and theology by making the latter speculative, and applying the methods of philosophy to the received dogmas of the church. Nielsen also lectured and wrote extensively on Hegel, and his main work *Grundidéernes Logik (The Logic of Fundamental Ideas)* of 1864-6 gave a full account of his Hegelian views in this area. None the less, he came under the influence of Kierkegaard's attack on Hegel's treatment of religion, and so criticized Martensen's position as being too complacent in this regard. In 1860, Nielsen was joined at Copenhagen as a professor by Hans Bockner (1820-75), who also thought and wrote as a Hegelian, principally on the history of philosophy.

As well as having an influence on Hegel's reception in Denmark, Kierkegaard is clearly the most philosophically significant thinker to have responded to his work in this country. While Kierkegaard attacked Hegel from a theistic perspective, his own form of Christianity was sufficiently radical to set him apart from any standard Right Hegelian approach; his critique can rather be seen as undermining Hegel's entire project, which was apparently to provide a systematic, rational and complete conception of the world, of the sort traditionally associated with a divine understanding. In rejecting this ambition as 'comic' and 'absurd', Kierkegaard therefore gave a very special twist to some of the themes found earlier in Schelling, and so deepened this existentialist reaction to Hegel's work. Kierkegaard came to this position out of a desire to save the religious outlook from the claim (made by Martensen, for example) that this Hegelian standpoint could give Christianity a rational basis. Kierkegaard argued that this was impossible, as philosophical speculation could never assimilate both the metaphysical and ethical paradoxes of true Christian faith: that God has become man, that religious knowledge can be based on subjective feeling, and that the religiously inspired individual (such as Abraham) may act out of a purely individual sense of the will of God. Kierkegaard therefore sets Christianity *against* the Hegelian conception of philosophy and philosophical

reason, in order to demonstrate the limitations of the latter (see Kierkegaard, S.A. §2).

In Holland, Hegel's earliest follower was P.G. van Ghert (1782-1852), who was a student of his in Jena and later became his friend. A more significant spokesman for Hegelianism was G.J.P.J. Bolland (1854-1922), who, as professor at Leiden (from 1896), established a kind of Hegelian sect which later infiltrated all parts of Dutch intellectual life (J. Hessing (1874-1944), J.G. Wattjes (1879-1944) and Esther Vas Nunes (1866-1929) being his most important pupils). However, this school lost its influence after the Second World War, due to the anti-Semitic views of Bolland himself, and the extreme right-wing affiliations of his pupils.

Italy. While Gioberti and Rosmini drew in a general way on aspects of post-Kantian German Idealist metaphysics, Hegel's ideas were more explicitly introduced into mainstream Italian culture through the efforts of Augusto Vera and Bertrando Spaventa (1817-82), who founded an influential Hegelian school in Naples and expounded Hegel's social and political thought with his *Studi sull'etica hegeliana (Studies on Hegelian Ethics)* (1869). Hegelian ideas were also represented by Francesco de Sanctis (1817-83), whose classic literary history, *La storia della letteraria italiana* (*History of Italian Literature*) (1870-1) is much influenced by Hegel's aesthetics, and by Raffaele Mariano (1840-1912) and Pasquale d'Ercole (1831-1917). One of Spaventa's pupils was Antonio Labriola, who later proved to be an independent Marxist thinker who appreciated the importance of Hegel for the evolution of historical materialism. He avoided the reductive positivist interpretation of Marxism which was currently being codified as a system of 'dialectical materialism' and was not inhibited from drawing freely on his Hegelian teachers and predecessors. For him as for them the living heritage of Hegel lay in his profoundly historical conception of social and political life, not in his metaphysical ambitions. In regarding Hegel as pre-eminently a great philosopher of culture Labriola anticipated much of the later Italian reception of Hegelian thought, by Marxists and non-Marxists alike. Labriola's expressly non-positivist interpretation of Marxism as essentially a 'philosophy of praxis' rather than a supposedly scientific and comprehensive worldview was a significant precursor of the Hegelian-Marxist approach that would emerge in Germany in the 1920s.

5 Hegelianism outside Germany in the nineteenth century: America and Britain

America. In the second half of the nineteenth century, Hegel's ideas came to play an important part in the intellectual life of the USA, where two centres of Hegelian thought began to develop. The first was a loosely associated group of friends and acquaintances based at this time in Cincinnati, Ohio, the most important of whom were John Bernard Stallo (1823-1900), August Willich (1810-78) and Moncure Conway (1832-1907). Broadly speaking, the Cincinnati Hegelians offered a left-wing interpretation of his views, which stressed his conception of a cosmos 'full of life and reason' (as Conway put it), in which scientific and social progress were possible, leading to a more liberal and rational political and religious order.

A similar outlook can be found in the second centre of Hegelian ideas at this time, in St Louis. The leading figures here were Henry Conrad Brokmeyer (1826-1906) and William Torrey Harris (1835-1909). After the Civil War, members of the Kant Club in St Louis formed the Philosophical Society, inaugurated in 1866 with Brokmeyer as president, Harris as secretary, and Denton Snider (1841-1925), G.H. Howison (1834-1916), A.E. Kroeger (1837-82) and Thomas Davidson (1840-1900) among its leading members. All were to contribute articles and translations to *The Journal of Speculative Philosophy*, which Harris edited from 1867 to 1893. The *Journal* had considerable influence in making Hegelian ideas part of the mainstream philosophical discussion in America, while Harris's own large output made a major contribution to the study of Hegel's works. Many of the St Louis Hegelians (including Brokmeyer and Harris) also had important institutional positions, in which they tried to apply his ideas in the fields of government and education.

Of this group, Harris was perhaps the most successful in developing a general philosophical outlook that is clearly Hegelian in character. He argued that in its highest stage, knowledge reveals 'independence and self-relation underlying all dependence and relativity' (Easton 1966: 481); and he used this structure, as Hegel had done, to develop a dialectical conception of 'identity-in-difference' that provided the basis for his account of the universe, God's relation to the world and the place of the individual in society.

By the end of the nineteenth century, many of the major academic posts in America were occupied by self-styled idealists, who accepted Hegel's central place in this tradition of thought. At this time, Hegel in particular and idealism in general were used to come to terms with the growing impact of Darwinian ideas on theology and

philosophy, in part by using the notion of the dialectic to find reason in the process of evolution itself.

From the 1880s onwards, however, the claim by Hegel's earlier American disciples that he represented the highest point of German thought began to be challenged, as pragmatism started to make its mark in academic philosophical circles (see Pragmatism). This drew on a much broader range of idealist thinkers than just Hegel, who was no longer viewed as the culminating point of the tradition. Thus, for an influential figure such as William James, the less rationalistic and metaphysical idealism of Kant, Schopenhauer and Lotze was more congenial to his pragmatic outlook. The central target of James' attack was Hegel's 'vicious intellectualism', to which he opposed his own radical empiricism (James 1909: 105). James argued that the concrete world of experience has a different structure from the world of thought, and that the particularity of things can never be adequately conceptualized. He criticized intellectualism for substituting 'a pallid outline for the real world's richness', and (like Kierkegaard) claimed that it sought to transcend becoming and temporality by abandoning the human point of view. In voicing these misgivings about Hegel's alleged essentialism, James was developing a familiar line of criticism, but in a way that was new in the American reception of Hegel's work.

The effect of this critique can be seen in the writings of James' Harvard colleague and contemporary, Josiah Royce. Unlike James, Royce was prepared to follow through the developments in idealism that led to Hegel, and so became his most sympathetic and sophisticated interpreter in this period, basing his conception of the Absolute on Hegel's account of the concrete universal as an organic unity of individual minds. In his posthumously published lectures on 'Aspects of Post-Kantian Idealism', delivered in 1906, Royce broke new ground by laying greater stress on the *Phenomenology of Spirit* than the *Logic*, emphasizing the voluntaristic aspects of the former, as showing that 'for Hegel, thought is inseparable from will' (1919: 145). By adopting this approach, Royce hoped to show that Hegel's real intention was to portray a 'logic of passion', and of the conflicts of the will, and not a system of abstract thought; this would demonstrate the continuity of Hegel's ideas with the outlook of pragmatism.

The other leading American pragmatists, C.S. Peirce and John Dewey were also influenced by their encounter with Hegel. While Peirce was quick to distance himself from American Hegelianism as a school (entering into a sharp critical exchange with W.T. Harris in 1868), he none the less acknowledged the affinities that existed between Hegel's outlook and his own, while more broadly he may be seen as a Neo-Kantian. The greatest convergence comes in Peirce's phenomenological deduction of the categories of Firstness, Secondness and Thirdness, and his demonstration that our immediate perceptual judgments (Firstness) and our relational judgments (Secondness) require mediation by reference to generalities (Thirdness); as Peirce admits in his *Lectures on Pragmatism* (1903), this deduction echoes Hegel's opening arguments in the *Phenomenology* (see Peirce, C.S. §7). None the less, Peirce complains that Hegel appears to reduce Firstness and Secondness to Thirdness, instead of recognizing that all three categories must be present in any coherent conception of the world. In Dewey, the influence of Hegel is more diffuse, as he was attracted more to his 'dissolution of hard-and-fast dividing walls', rather than any particular doctrine, although he was prepared to defend Hegel's criticisms of Kant in his important early essay 'Kant and Philosophic Method' (1884) (see Dewey, J. §1).

Britain. If the pragmatists took Hegel seriously, this was not just because of his impact in America, but also because of the importance of Absolute Idealism in Britain in the 1880s and '90s, which represented the high point of Hegel's influence there.

In Britain, the initial reception of Hegel's work came relatively late. His ideas were given some limited attention in the writings of William Hamilton and James Ferrier, and figured briefly in the historical accounts of German Idealism by J.D. Morrell and G.H. Lewes, while the first translation (of part of Hegel's *Logic*) appeared in 1855. It was not until J.H. Stirling's *The Secret of Hegel* (1865), however, that any substantial sympathetic treatment of Hegel's work became available, and it marks the real beginning of Hegel's influence. While he was aware of the sustained critique of Hegel as a Platonic idealist and essentialist that had gained currency in Germany in the 1840s and '50s, Stirling still adopted this reading, proclaiming that for Hegel 'organic Reason (is) a self-supported, self-maintained, self-moved life, which is the all of things, the ultimate principle, the Absolute' (Stirling (1865) 1898 I: 96).

Stirling's book was followed in 1874 by a translation of Hegel's *Encyclopedia Logic* by William Wallace (1844-97), together with a long introduction entitled *Prolegomena to the Study of Hegel's Philosophy*. Like Stirling, Wallace sought to use Hegel in the critique of positivism and scientific naturalism, and interpreted his

idealism as a kind of thoroughgoing holism, while, like his American contemporaries, he sought to show how Hegel's notion of the dialectic might be used to bring out the rationality of Darwinian evolution. A similar set of concerns is reflected by Edward Caird (1835-1908) in his *Hegel* (1883), for whom 'the task of philosophy is to gain, or rather perhaps to regain, such a view of things as shall reconcile us to the world and to ourselves' (Caird 1892 I: 191). It was this search for unity that Caird found in Hegel's work, particularly in relation to the opposing claims of freedom and necessity, subject and object, God and the universe, and he therefore interpreted Hegel's Absolute as such a reconciling principle.

As well as these published accounts of Hegel's thought, a positive view of Hegel also began to emerge more indirectly, as he was taken up by the important group of idealist thinkers who were becoming increasingly influential at this time. One of the first of these was T.H. Green, who was led to read Hegel by his tutor and later colleague at Balliol, Benjamin Jowett. Green's critique of empiricism had both Kantian and Hegelian elements, while his account of self-consciousness as a single, actively self-distinguishing spiritual principle which expresses itself in temporal human intelligence reflected his understanding of Hegel's conception of *Geist*. None the less, Green declared himself unhappy with Hegel's method for arguing to this conception, stating that 'it must all be done again'. Likewise, while he was clearly helped to his own account of freedom by his reading of Hegel, he remained suspicious of what he took to be the latter's uncritical acceptance of the modern state, in which this freedom was to be realized.

This equivocal attitude is also reflected in the relation of one of the other leading British Idealists to Hegel, F.H. Bradley. Hegel's influence can be traced in Bradley's critique of Kantian ethics in his early *Ethical Studies* (1876); in his hostility to the classical empiricist's view of our experience of reality as divisible into discrete simple elements; in his treatment of judgment, the concrete universal and the problem of relations; and in his conviction that from the perspective of the Absolute, all *aporiai* in our understanding of reality would be overcome. None the less, Bradley remained critical of central aspects of Hegel's thought and method, famously dismissing his *Logic* as an 'unearthly ballet of bloodless categories', and with it the panlogist metaphysics this seemed to represent.

Bradley's contemporary Bernard Bosanquet was less openly critical of Hegel, as he developed Bradley's Hegelian approach to the logical forms of thought (such as judgment and syllogism), in order to show how in these forms, all abstraction from the whole turns out to be incoherent. Bosanquet carried this holism over into what was seen as a Hegelian conception of the individual and society, claiming that for human beings 'their true individuality does not lie in their isolation, but in that distinctive act or service by which they pass into unique contributions to the universal' (Bosanquet [1899] 1923: 170). In his work on aesthetics, Bosanquet focused attention on this aspect of Hegel's system, with his translation of the introduction to Hegel's *Lectures on Aesthetics* (1886), and his account of Hegel in his influential *History of Aesthetics* (1892).

Bosanquet was not alone among the British Idealists in offering interpretative commentaries on Hegel's work, although towards the end of the 1880s, these became increasingly critical and critically informed. A decisive moment came in 1887, with the publication by Andrew Seth (later Andrew Seth Pringle-Pattison) (1856-1931) of *Hegelianism and Personality*, in which he followed Schelling, Trendelenburg and others, and criticized Hegel's apparent panlogicism; following the Left Hegelians, he gave this attack an ethical and political dimension, arguing that by hypostatizing universality, Hegel gives priority to the species over and above the individual, a move which Seth set out to oppose with his own so-called 'Personal Idealism'. For Hegel's followers in Britain Seth therefore represented a parallel to the existentialist critique of his system already developed in Germany, but which had not been properly addressed by the British Idealists before.

In response, interpretations of Hegel emerged which played down his apparent panlogism, and instead began to treat the *Logic* as a kind of category theory. For example, in an influential article on 'Darwin and Hegel' (1890-1), D.G. Ritchie (1853-1903) argued that Hegel does not have to be read as a speculative cosmologist; rather, 'we (will) find that his logic and the whole of his philosophy consist in this perpetual "criticism of categories", i.e. in an analysis of the terms and concepts which ordinary thinking and the various special sciences use as current coin without testing their real value' (Ritchie 1890-1: 61). This approach was most fully developed in the commentaries on Hegel's system by J.M.E. McTaggart. McTaggart argued that the aim of Hegel's dialectic was to show how the categories of ordinary thought provide only partial or imperfect conceptions of the truth, which point towards a highest form of thought - the Absolute Idea - in which these imperfections are finally overcome. Where McTaggart

criticized Hegel was for underestimating the difficulty which we have, as limited intellects, in conceiving of the world in these terms, so that although he accepted the Hegelian claim that a resolution of all *aporiai* must be possible, he questioned whether such a view of reality was achievable by us. This approach to the reading of Hegel led McTaggart to emphasize the many apparent contradictions in how things appear to us (most famously, as events occuring in time), and to claim that therefore these appearances must be unreal, opening the way for him to indulge in extravagant metaphysical theorizing about ultimate reality at odds with our experience of the world.

By the beginning of the First World War, the taste for such theorizing had changed, as the Idealist's claims about the contradictory nature of how things appear to us seemed increasingly spurious, thereby disposing of the need to overcome these contradictions in a view of reality as somehow monistic, atemporal, changeless or immaterial. Anglo-Hegelian idealism was therefore increasingly viewed as irrelevant and poorly grounded by the leaders of the next generation of philosophical thinkers (such as Bertrand Russell and G.E. Moore), while at the same time the 'New Liberals' (such as J.A. Hobson and L.T. Hobhouse) submitted the idealist's theory of the 'organic' state to merciless attack, an attitude which hardened once the war against Germany had begun.

6 Hegelian influence in the twentieth century: Germany, Italy

While towards the end of the nineteenth century it may have appeared that Hegel's philosophy was destined to have only a marginal significance in twentieth-century thought, in fact its impact has been remarkable. This renewed interest in Hegel's position was made possible by a broader understanding of his project, which made many of the standard nineteenth-century criticisms (of panlogism, quietism, anti-individualism and theistic romanticism) appear crude and simplistic, reflecting a misconception of his work.

Germany. During a period in which various forms of positivist naturalism or Neo-Kantian schools dominated the German philosophical scene the Hegelian and idealist legacy generally had found some refuge within the traditional humanistic disciplines which escaped subjection to the methodological canons of the natural sciences. A broadly Hegelian approach thus survived in a largely non-systematic and non-metaphysical hermeneutic form which seemed to offer significant elements at least for the construction of an alternative methodology for the newly developing social and human sciences, the '*Geisteswissenschaften*'. Wilhelm Dilthey was particularly influential in reawakening interest in the world of early German Idealism with his path-breaking study of Hegel's early development, *Die Jugendgeschichte Hegels (The Young Hegel)* (1905). It was also under Dilthey's direct encouragement that his student Hermann Nohl first thoroughly edited and published most of Hegel's surviving early manuscripts of 1790-1800 as *Hegels theologische Jugendschriften* (*Hegel's Early Theological Writings*) in 1907, an event which inaugurated that German resurgence of interest in Hegel's philosophy during the first couple of decades of this century which culminated in the broad movement known as '*Neuhegelianismus*'. Hegel's early writings challenged the image of the systematic rationalist metaphysician of tradition and seemed rather to reveal a thinker passionately concerned with restoring a concrete sense of cultural wholeness and identification with the natural and historical world of lived experience.

The interest in Hegel and the tradition of German Idealism in general did not simply displace the still vigorous forms of Neo-Kantianism but rather entered initially into a complex symbiotic relationship with certain trends within that movement, especially the so-called Southwest School associated principally with Heinrich Rickert (1863-1936) and Wilhelm Windelband (1848-1915) (see Neo-Kantianism §4). Like Dilthey, these thinkers were attempting to develop an appropriate philosophical approach to the entire sphere of cultural and spiritual life as an autonomous domain alongside the sphere of the natural and the exact formal sciences. Many of them believed that the Kantian tradition required significant extension and supplementation to do justice to this dimension of experience and looked to Hegel in particular for intellectual resources adequate to the task. A symptomatic document for the period was Windelband's influential address of 1910, 'Die Erneuerung des Hegelianismus' (The Renewal of Hegelianism). Eventually a fully-fledged neo-Hegelian school began to form as part of a broader cultural project of German intellectual renewal, a process that was actually encouraged rather than weakened by the catastrophic experience of the First World War and the ensuing social and political instability.

An important figure in this development was Georg Lasson (1862-1932) who tirelessly promoted a strongly religious interpretation of Hegel's philosophy as the appropriate antidote to the disintegrative and sceptical tendencies and sense of cultural alienation of the time. The intrinsic philosophical significance of Lasson's work is negligible and often represents little but nationalistic edification, as in *Was heisst Hegelianismus? (What is*

Hegelianism?) (1916), but he performed an extremely important role as an indefatigable editor of Hegel's works.

Other principal figures associated with or broadly sympathetic with *Neuhegelianismus* were Hermann Glockner (1896-1978), also important as an editor of Hegel, with his synoptic monograph *Hegel* (1929-40); Nicolai Hartmann (1882-1950), who drew strongly on Hegel in his own work and provided a classical ontological interpretation of the philosopher in his *Die Philosophie des deutschen Idealismus (The Philosophy of German Idealism)* (1923-9); Richard Kroner (1884-1974) perhaps the purest and most dedicated representative of the movement, who wrote a standard neo-Hegelian history of German Idealism *Von Kant bis Hegel (From Kant to Hegel)* (1921-4) but reverted in his later writings to a more Kantian position influenced by Kierkegaard; Theodor Litt (1880-1962) who remained strongly influenced by Dilthey's philosophy of culture and Heidelberg Neo-Kantianism and later attempted to synthesize contemporary trends in a quasi-Hegelian fashion in *Denken und Sein (Thought and Being)* (1948), *Mensch und Welt (Man and World)* (1948) and *Hegel* (1953). Kroner helped to establish the journal *Logos* which functioned as the organ of the German neo-Hegelians and sympathetic Neo-Kantians during the 1920s. There were also a number of more important and original thinkers on the fringes of the movement who were profoundly influenced by the resurgence of interest in German Idealism and Hegel in particular. These included Georg Simmel, Ernst Cassirer and Franz Rosenzweig, who all engaged with central Hegelian problems in their work and occupied something of an ambiguous and contested space between Kant and Hegel.

By the end of the 1920s, in the context of the German crisis of democracy and the rise of fascism, the vague romantic and undifferentiated aspiration to living 'wholeness' as a supposed alternative to social atomism readily lent itself to ideological mystification and exploitation. Some neo-Hegelians made uncritical appeal to the idea of '*Sittlichkeit*' or concrete ethical life as a model of organic community, but increasingly detached from its original context in Hegel's elaborate conception of the rational modern constitutional state as the climax of the philosophy of history and the evolution of the consciousness of freedom. The Hegelian notion of the '*Volksgeist*' or 'spirit of the people' was also interpreted more in the spirit of Savigny and the 'Historical School' than in that of Montesquieu or even Herder, and the resulting simplification was urged in support of an illiberal communitarian ideology. Certain tendencies in this direction are clearly discernible in the works of Lasson, Glockner and in the monumental study by Theodor Haering (1884-1964) of Hegel's development, *Hegel. Sein Wollen und sein Werk (Hegel: His Project and his Work)* (1929-38). In all these authors romantic over-interpretation and a celebration of the supposedly 'irrational' character of the dialectic almost completely effaces the universalist and rationalist dimension of Hegel's thought and minimizes the significance of his relationship to Kant and eighteenth-century thought. None the less, except for similar interpretations by fascistically inclined legal theorists such as Julius Binder and Karl Larenz the official ideology showed little interest in reclaiming Hegel for the cause of National Socialism.

The significant alternative to the repristination of Hegel under the sign of cultural philosophies of life and value during this entire period was provided by the intellectual renewal of Marxist thought and the emergence of what later became known generically as 'Western Marxism' and 'Critical Theory'. The early work of Karl Korsch (1886-1961), especially his *Marxismus und Philosophie (Marxism and Philosophy)* (1930) and of Georg Lukács, with *Geschichte und Klaßenbewusstsein* (*History and Class Consciousness*) (1923), proved to be the initial stimulus for this development. Both rejected the positivist interpretation of Marxism as a scientific worldview supposedly in secure possession of the 'laws' of social and historical development and regarded the 'dialectics of nature' as a theoretical illusion and a practical irrelevance. Lukács extrapolated from Marx's mature work to his Hegelian origins and outlined a non-deterministic philosophy of praxis and potential self-liberation which owed much to Hegel's *Phenomenology*. Although Lukács later repudiated his earlier work in certain respects as 'idealist revisionism', he continued to emphasize the enduring significance of the Hegelian legacy in Marx against the Marxist-Leninist orthodoxy with *Der junge Hegel* (*The Young Hegel*) (1948), and drew equally heavily on Hegel in his own later works, like the massive study on aesthetics (*Die Eigenart des Asthetischen (The Specificity of the Aesthetic)*) (1963) and the unfinished treatise *Zur Ontologie des gesellschaftlichen Seins* (*The Ontology of Social Being*) (1971-2). The contemporary need to reinvestigate the entire relationship of Hegelian and Marxist thought was also stimulated during this period by the continual publication of previously unknown texts by both Hegel (especially the Jena writings then issued as his *Realphilosophie* in 1933) and Marx (particularly the *Economical and Philosophical Manuscripts* in 1932), writings which did much to confirm the insights of Lukács'

contested interpretation of Marx's debt to Hegel.

Herbert Marcuse was influenced in his early period by Diltheyan philosophy of life, Heidegger's existential phenomenology, and the rediscovery of Hegel's early work. After his study *Hegels Ontologie* (*Hegel's Ontology*) (1932), Marcuse turned explicitly to Marx whose thought he interpreted in a humanist manner in the light of the early Hegelian manuscripts, stressing like Lukács the key concept of alienation and the ineliminable moment of social subjectivity against more standard mechanistic interpretations. Marcuse also defended the Hegelian tradition directly against the charge of totalitarianism and articulated the deep continuity between the thought of Hegel and Marx in *Reason and Revolution* (1941). In his later work Marcuse focused on the question of the aesthetic dimension and its emancipatory potential as a prefiguring of a non-repressive relation to inner and outer nature, attempting to mediate the heritage of classical German philosophy with elements of Freudian thought.

Although Theodor-Wiesengrund Adorno repeatedly made Hegel an object of privileged critique, as in the *Drei studien zu Hegel* (*Three Studies on Hegel*) (1963), he could also be regarded as the most profoundly Hegelian of modern thinkers in terms of the fundamental themes of his philosophy and its elaborate dialectical conceptuality. His major works, *Negative Dialektik* (*Negative Dialectics*) (1966) and *Ästhetische Theorie* (*Aesthetic Theory*) (1970) are a sustained critical engagement in a Marxist spirit with the tradition of Hegel and German Idealist thought and are unintelligible without constant reference to the concepts of totality and dialectic subject-object identity. Adorno sought to reclaim the concept of reconciliation (of social antagonism, spirit and nature, universal and particular) from its apologetic use in speculative philosophy and employ it as a critical measure of existing contradiction and unfreedom. He drew strongly on Hegelian patterns of argument to criticize other thinkers, such as Kierkegaard, Husserl and Heidegger.

Italy. At the turn of the century the tradition of neo-Hegelian thought in Italy was principally represented by Benedetto Croce and Giovanni Gentile. Croce's reception of Hegelian thought was selective and highly reconstructive, in some respects paralleling the initial German renewal of Hegelian studies in Dilthey's wake. Again it was not the metaphysical dimension of Hegel's thought, but rather the doctrine of concrete spiritual agency and its self-objectification in social and cultural life which attracted Croce, as can be seen from his *Ciò che è vivo e ciò che è morto nella filosofia di Hegel* (*What is Living and What is Dead of the Philosophy of Hegel*) (1907). However, Croce's aesthetics owes at least as much to Kant in its emphasis upon the priority of intuition and the total autonomy of the art work; he also rejected the concept of aesthetic genre which was central to Hegel's historical construction of art. He also repudiated the supposed 'death of art' thesis which he influentially took to be implied in Hegel's subordination of art to religion and philosophy. However, Croce entertained no qualms about the apparent supercession of the religious dimension in speculative philosophy and his appropriation of Hegel was thoroughly immanent and humanistic.

If Croce stressed the autonomy of the different domains of spiritual activity, his erstwhile friend and collaborator Gentile followed Hegel more directly in grasping all human activities as interrelated manifestations of creative spirit. Similarly in his philosophy of art Gentile defended a less formalist position than Croce. Gentile's philosophy generally is also marked by a strong voluntarist emphasis and an ardent educational idealism that has affinities with Fichte. Croce finally broke with Gentile when the latter attempted to provide a Hegelian justification of the new, fascist 'Corporate State' as the concrete realization of ethical life. In spite of his political affiliations Gentile's thought continued to exercise a significant influence on Italian thought at both ends of the political spectrum. Hegel's influence was also strongly registered by the Marxist theoretician Antonio Gramsci through the contemporary example of Croce and Gentile. Gramsci developed a philosophy of praxis that closely paralleled Lukács' interpretation of Marx and rejected the quasi-naturalistic conception of dialectical materialism, seeking rather to transcend and preserve the heritage of bourgeois culture and philosophy and endow the Marxist perspective with the potential for cultural hegemony. As with the Western Marxists generally, Gramsci distrusted the mechanical application of any simple basis/superstructure distinction and attempted to grasp the complex mediation between social determinants and the collective self-consciousness of human agents in more dialectical fashion.

7 Hegelian influence in the twentieth century: Britain, America and France

Britain and America. The appropriation of Hegelian idealism by Croce and Gentile influenced R.G. Collingwood, who was one of Hegel's few sympathetic readers in Britain between the wars. Like the Italians, Collingwood

believed that Hegel's Platonism had stopped him properly overcoming the opposition of art and logic, feeling and thought, and in his own method of question and answer he sought to present Hegel's dialectic in less panlogistic, more historicist terms, which did not seek to escape the 'absolute presuppositions' of its time. Moreover, in taking up a Crocean approach to the historical method (summed up in Croce's dictum that 'every true history is contemporary history'), Collingwood drew attention back to Hegel's philosophy of history, from which Croce's was a critical development.

In America in this period, direct interest in Hegel had also waned, although a continuing commitment to the idealist tradition can be found in the work of W.M. Urban (1873-1952) and Brand Blanshard (1892-1966), whose coherence theory of truth refers back to the British school of Absolute Idealism, and thus indirectly to Hegel.

France. As in Germany and Italy, the view of Hegel that emerged in France in the twentieth century no longer set him in opposition to the humanistic, non-metaphysical, anti-essentialist perspective of his critics, but instead treated him as an important precursor and source of this very perspective. Within French thought, the beginnings of this reassessment can be traced back to Jean Wahl's *Le Malheur de la Conscience dans la Philosophie de Hegel (The Unhappy Consciousness in the Philosophy of Hegel)* (1929). In this work, Wahl (1888-1974) attempted to uncover a side to Hegel's thought that was darker, more romantic and less rationalistic than had previously been noticed, and to cast fresh light on the whole direction of his philosophy. He was helped towards this reinterpretation by the publication of Hegel's early writings by Dilthey and Nohl, which revealed to Wahl that Hegel's real preoccupations and concerns were close to those of a Christian existentialist like Kierkegaard, a fact that had been obscured by the speculative approach of the later Encyclopedic system. Wahl was therefore led to look anew at the *Phenomenology of Spirit*, treating it not merely as a prolegomenon to the mature system, but as the highest expression of Hegel's troubled vision; at the centre of his reading of the *Phenomenology*, Wahl placed Hegel's treatment of the Unhappy Consciousness, in which (he argued) the sense of loss is epitomized. Thus, although Wahl himself was not prepared to call Hegel an existentialist, his influential study of the *Phenomenology* showed how existentialist themes could be uncovered in Hegel's thought (see Existentialism).

In the wake of Wahl's study, the Hegel renaissance in France was taken further and given greater impetus by the work of Alexandre Kojève and Jean Hyppolite (1907-68). Kojève gave an important series of seminars on the *Phenomenology* from 1933 to 1939 at the École Pratique des Hautes Études, which was attended by many who were to become leading luminaries of French intellectual life, as well as influential interpreters of Hegel in their own right, including Maurice Merleau-Ponty, Eric Weil (1904-77), Georges Bataille and Jacques Lacan. The text of these seminars was published in 1947, and it remains one of the most challenging readings of Hegel's thought. Equally important were the efforts of Hyppolite, who published the first volume of his magisterial translation of the *Phenomenology* in 1939 and the second in 1941, and in 1946 completed his commentary on the text, entitled *Genèse et structure de la Phénoménologie de l'esprit de Hegel* (*Genesis and Structure of Hegel's Phenomenology of Spirit*).

Kojève made the master-slave dialectic the key to his treatment, into which he wove both Heideggerian and Marxist themes. He cites as an epigraph to his lecture on the *Phenomenology* Marx's comment that 'Hegel… sees *labour* as the *essence* of man, the self-confirming essence, of man' (Marx 1975: 386) and, like Marx, identifies the work of the slave as an essential moment in self-objectification. At the same time, with Heidegger he emphasizes the slave's experience of death, and his recognition of finitude, out of which the slave also feels liberation from the natural world. Kojève therefore interprets Hegel's move to idealism in this light: it is an attempt to show how the human mind can overcome the material world of nature, by creating its own world through the power of speech, language and thought, an ideological realm in which we feel at home and free. This free creativity also has a more tragic aspect, however, as it is limited and defined by an awareness of finitude and death; at the same time the capacity to die represents our liberation from the control of any transcendent creative power, such as God, and is thus the dialectical expression of our highest freedom.

Perhaps Kojève's best-known and most remarkable contribution to the interpretation of Hegel arises directly from the conjunction of Marxist and existentialist aspects in his account: for, drawing on both Heidegger and Marx, Kojève argues that for Hegel history began with the sense of otherness, and can end in the universal satisfaction of the desire for recognition, putting a stop to our urge to negate and overcome all externality. Thus Kojève arrives at a non-metaphysical, secularized conception of Hegel's philosophical history, and reads the end of his system in

anthropological, not theological terms; he therefore takes another step away from the nineteenth-century image of Hegelianism, and offers a new vision of this notoriously problematic aspect of Hegel's work.

For readers of Hegel, however, Kojève's interpretation raises almost as many problems as it solves, and many have felt (with Jean Wahl) that 'it is quite false but very interesting'. Hyppolite's approach is rather more judicious, while he too is influenced in his reading by existentialism and Marx. Like Wahl, he holds that 'unhappy consciousness is the fundamental theme of the *Phenomenology*…. The happy consciousness is either a naïve consciousness which is not yet aware of its misfortune or a consciousness that has overcome its duality and discovered a unity beyond separation. For this reason we find the theme of unhappy consciousness present in various forms throughout the *Phenomenology*' (Hyppolite 1946 I: 184; 1974: 190). Also like Wahl, Hyppolite argues that 'we find (Hegel) in his early works and in the *Phenomenology*, a philosopher much closer to Kierkegaard than might seem credible' (Hyppolite 1971 I: 93): although Hegel admittedly ends in Absolute Knowledge which seems to transcend all diremptions, the journey of consciousness is nevertheless characterized as 'the way of despair'. At the same time, Hyppolite emphasizes Hegel's foreshadowing of Marx's account of alienation, and agrees with Kojève that recognition is capable of overcoming the tension between self and other. None the less, in his later writings on Hegel (such as *Logique et existence (Logic and Existence)* and 'Essai sur la *Logique* de Hegel' ('On the Logic of Hegel')), Hyppolite gave greater weight to the *Logic* than hitherto; for, he argues, the claim to Absolute Knowledge, and the transition to the *Logic* must be made, if 'the phantom of the thing-in-itself' is to be avoided, and with it the sense that we are out of touch with Being. Hyppolite acknowledges, however, that there is a tension between this return to the Logic and metaphysics, and the more humanistic, anthropological method of the *Phenomenology*, a tension which he sees as fundamental to Hegel's thought.

It is partly thanks to this reading of Hegel by Kojève and Hyppolite that Marxism and existentialism became so interlinked in the intellectual life of post-war France; and it is clear that they helped bring about a rapprochement between Hegel and Marx in this period by treating existentialism as a kind of common ground on which Hegelianism and Marxism could be reconciled. While some (such as Althusser) remained hostile to this development, existentialism also served to bring about the same kind of reconciliation outside France, as the themes of alienation, reification and estrangement from nature were discovered in both their works.

8 Contemporary developments

In the last third of the twentieth century, Hegel has continued to have a considerable influence on philosophical thought, both as a major figure within the canon of 'continental' philosophy, and (more recently) within Anglo-American 'analytic' and post-analytic philosophy.

France. Since the end of the 1960s the reception of Hegelian thought in France has been significantly determined by successive waves of intellectual reaction to the previously dominant philosophies. The structuralist movement which partially supplanted the phenomenological and existentialist tradition tended to minimize the Hegelian elements in Marx's thought and emphasize the radical incompatibility of 'idealist' and scientifically 'materialist' approaches to the constitution of social reality. The emergence of a genealogical mode of critical discourse in Michel Foucault, the libidinal materialism of Gilles Deleuze and the postmodern pluralism of Jean Lyotard represented a decided antirealism and anti-foundationalism which questioned the central assumptions of the classical philosophical tradition and its metaphysics of truth. The pervasive influence of Nietzsche and Heidegger, and the perceptible political retreat from a hitherto powerful Marxist tradition, has conspired in the French context to produce something resembling a regnant anti-Hegelianism as a negative mirror image of the era of Kojève, Merleau-Ponty and Sartre. The critique of the metaphysical tradition of ontology and its supposed prioritizing of self-presence has inspired the ethically oriented philosophies of alterity like those of Emanuel Levinas in *Totalité et Infini* (*Totality and Infinity*) (1961) and Jacques Derrida in *Glas* (1974) respectively. For both thinkers, though in subtly distinct ways, Hegel again represents an exemplary case of all-consuming totalizing discourse and consequently a privileged object of critical analysis. What is at stake here is the claim to articulate a logic which can grasp difference positively rather than in terms of opposition, and the rejection of dialectic as an appropriate conceptual resource for this task.

Italy. Since the 1970s, the influence of alternative models of radical philosophy like French structuralism, post-structuralism and deconstruction has partially eclipsed the previous Marxist monopoly on critical social thought in Italy, while as in France, the rejection of 'grand narratives' and supposedly totalizing metaphysical

discourse has led to a developing critique of the idealist tradition as a whole, and Hegelian philosophy in particular. One result has been an increasingly scholarly and interpretative engagement with Hegel and the modern German tradition, but less evidence of any productive appropriation of dialectical thought.

Germany. After the neo-Hegelian movement of the pre-war period in Germany one cannot accurately speak of any Hegelian 'School' of thought. None the less, the significance of Hegel continued to make itself felt indirectly in the hermeneutic version of phenomenology developed by Martin Heidegger. Heidegger's reading of the European metaphysical tradition, supposedly culminating in Hegel, exercised considerable influence upon the interpretation of Hegel's thought. If Heidegger's own attitude to Hegel was problematically ambivalent, his student Hans-Georg Gadamer developed a critical but productive relationship to Hegel mediated by his appropriation of Heidegger's thought as a universal ontological hermeneutics. In *Wahrheit und Methode* (*Truth and Method*) (1960) he drew especially on Hegel's account of experience and endorsed the anti-subjectivist thrust of Hegelian philosophy in his rejection of the psychologistic hermeneutics he associated with Schleiermacher and Dilthey.

It was also in the context of the hermeneutic tradition that a distinct renewal of theological interest in Hegel first arose after the war, a development that was subsequently intensified by the social turn in modern theology with the influence of the Frankfurt School and issues of Marxist-Christian dialogue. Although Karl Barth had always emphasized the autonomy of theological discourse, his evolving thought eventually led him from an initial commitment to a paradoxical dialectic indebted to Kierkegaard towards a position of theological realism, a quasi-Hegelian insistence upon systematic objectivity grounded in the trinitarian nature of God's unreserved self-disclosure. A number of Barth's students and interpreters pursued the critical turn against exclusively existentialist emphasis upon individual subjectivity and an apparent neglect of social reality and historical revelation. Jürgen Moltmann (1926-) responded to impulses from Ernst Bloch and the Frankfurt School and adumbrated a dialectical theology of liberation with his *Theologie der Hoffnung* (*Theology of Hope*) (1965) and *Der gekreuzigte Gott* (*The Crucified God*) (1973). The work of Wolfhart Pannenberg (1928-) reflects the new confidence in systematic theology that draws comprehensively on the classical idealist tradition and Hegel's incarnational metaphysics in particular. Sharing Moltmann's insistence on an open dialectic with a liberatory eschatological dimension, Pannenberg has also attempted a qualified and critical re-appropriation of Hegelian insights. Although profoundly influenced by Heideggerean hermeneutics, Eberhard Jüngel (1934-) also pursues the arguably Hegelian insistence on the radical humanity of God in the later Barth and addresses the question of the 'death of God' through close engagement with German Idealism and the Left Hegelian tradition in his *Gott als Geheimnis der Welt* (*God as the Mystery of the World*) (1977). What unites these theologians, despite significant differences of emphasis, is the attempt to exploit the conceptual resources of the dialectical tradition to overcome the abstract antithesis of atheism and theism and restate the trinitarian character of spirit.

Naturally Hegel also remained a permanent point of reference for the more orthodox strains of Marxism in Germany throughout this period. None the less, the most significant and vital engagement with the Hegelian tradition was still to be found among the heirs of the Frankfurt School and those broadly sympathetic to the aspirations of Critical Theory. Thus Jürgen Habermas has responded intensively to aspects of Hegel's thought in his reformulation of an emancipatory social philosophy. While rejecting Hegel's supposed metaphysical philosophy of identity in favour of a quasi-transcendental account of irreducible constitutive interests, Habermas exploited Hegel's insights into the communicative dimension of social interaction and the centrality of the concept of recognition. In contrast with his predecessors, Habermas has attempted to reorient critical theory through the turn to intersubjectivity as an alternative to traditional philosophy of consciousness. In his later work, *Der philosophische Diskurs der Moderne* (*The Philosophical Discourse of Modernity*) (1985), Habermas recognizes Hegel's contribution to the formulation of the concept of modernity and pursues a post-metaphysical appropriation of aspects of the idealist tradition for a non-foundational universalist ethics.

Karl-Otto Apel (1922-) originally revealed the influences of Heideggerian hermeneutics, philosophical anthropology and Litt's neo-Hegelian idealist philosophy, subsequently modified by an increasing interest in Critical Theory. Like Habermas, Apel's critical relation to the tradition was motivated by the accommodations and failures of pre-war historicist philosophies of culture and ossified Marxism in the face of authoritarianism on right and left. Accepting fundamental features of Hegel's critique of Kant, he has drawn extensively on the American idealist and pragmatic tradition of Royce and Peirce in his *Transformation der Philosophie* (*Towards a Transformation of Philosophy*) (1973) to develop a social conception of the rational community as the ultimate

normative presupposition of enquiry. Apel is committed to the dialectical mediation of abstract alternatives in contemporary philosophy and overcoming the opposition between rational grounding and practical ethical orientation in transcendental self-reflection, employing Kantian and Hegelian elements as mutual correctives of one another. The communicative turn and the question of intersubjectivity is also central for Michael Theunissen (1932-) who has interpreted and productively appropriated aspects of Hegel in sustained interaction with Kierkegaard and constant conjunction with developments in contemporary thought. In his attempt to relate the insights of dialogical and dialectical thought his work reflects all the different currents of existential phenomenology, theology and critical theory in which Hegel has been a latent presence or critical point of reference throughout the century.

Britain and North America. Until recently Hegel was left largely unread by those working within the 'analytic' tradition of Anglo-American philosophy, while the issues raised by Hegelian idealism were only discussed by those at the margins of this movement, such as J.N. Findlay (1903-87) and G.R.G. Mure (1893-1979). Since the 1970s, however, respect for Hegel's thought has grown, partly because of his influence on the communitarian and historicist ideas of Charles Taylor and Alasdair MacIntyre, and partly because Hegel's attack on the Kantian division of 'form' and 'matter' in experience has been echoed by those (such as Donald Davidson, Hilary Putnam and Richard Rorty) who question the scheme/content distinction in epistemology, and who thereby seek to go beyond Kant's 'subjective idealism' in a somewhat Hegelian manner.

Thus, as the 'continental' and 'analytic' traditions have come together over the question of how far metaphysics is possible after the Kantian turn, the current intellectual landscape is characterized by a continuing interest in examining, clarifying and exploiting the conceptual resources of the German Idealist tradition, for which the interpretation, appropriation and contestation of Hegelian philosophy will inevitably represent a permanent point of reference.

See also: Absolute, the; American philosophy in the 18th and 19th centuries §2; Dialectical materialism; Frankfurt school; German idealism; Hegel, G.W.F.; Idealism; Italy, philosophy in; Kyoto School; Marxist philosophy of science; Rosenzweig, F.; Russian Hegelianism; Scandinavia, philosophy in; Western Marxism

ROBERT STERN
NICHOLAS WALKER

List of works

For reasons of space, some of the works cited in the text are not given in the bibliography below, but are cited in the respective biographical entries of the particular author.

1 Primary literature

Stern, R. (ed.) (1993) *G.W.F. Hegel: Critical Assessments*, London: Routledge, 4 vols.(The first two volumes contain a selection of the main writings in Hegel's nineteenth- and twentieth-century reception.)

2 Germany

Adorno, T.W. (1963) *Drei Studien zu Hegel*, Frankfurt: Suhrkamp; trans. S.W. Nicholsen and J.J. Shapiro, *Hegel: Three Studies*, Cambridge, MA, and London: MIT Press, 1993.(Adorno's most explicit critique and appropriation of Hegel.)

Adorno, T.W. (1966) *Negative Dialektik*, Frankfurt: Suhrkamp; trans. E.B. Ashton, *Negative Dialectics*, London: Routledge, 1973.(Adorno's major engagement with Kant, Hegel and Heidegger, informed throughout by his transformation of 'speculative' Hegelian concepts into a critical and 'negative' dialectic.)

Adorno, T.W. (1970) *Ästhetische Theory*, Frankfurt am Main: Suhrkamp; trans. C. Lenhardt, *Aesthetic Theory*, London: Routledge & Kegan Paul, 1984.

Apel, K.-O. (1973) *Transformation der Philosophie*, Frankfurt: Suhrkamp; trans. G. Adey and D. Frisby, *Towards a Transformation of Philosophy*, London: Routledge, 1980.(Extensive essay collection presenting Apel's attempt to mediate the extremes in contemporary philosophy by recourse to the hermeneutic and pragmatic traditions.)

Bauer, B. (1841) *Die Posaune des jüngsten Gerichts über Hegel den Atheisten und Antichristen*, Leipzig: Otto Wiegand; repr. Aalen: Scientia, 1969; trans. L.S. Stepelevich, *The Trumpet of the Last Judgement Against Hegel the Atheist and Antichrist. An Ultimatum*, Lewiston, NY: Mellen, 1989.(Argues for the incompatibility

of Hegelian philosophy with traditional Christian belief.)

Baur, F.C. (1835) *Die christliche Gnosis oder die christliche Religionsphilosophie in ihrer geschichtlichen Entwicklung (Christian Gnosis, or the Christian Philosophy of Religion in its Historical Development)*, Tübingen: Osiander.(Traces the evolution of the Gnostic theological tradition up to and including Hegel and German Idealism.)

Biedermann, A.E. (1849) *Unsere junghegelianishe Weltanschauung oder der sogennante neueste Pantheismus (Our Young Hegelian Worldview, or the so-called latest Pantheism)*, Zurich: F. Schultheiss.(A critical response to a theological polemic directed against the 'Tübingen School' and Hegel's influence.)

Biedermann, A.E. (1868) *Christliche Dogmatik (Christian Dogmatics)*, Berlin: Georg Reimer; 2nd edn, 1884-5. (A late example of a broadly Hegelian and reconstructive exposition of traditional doctrines which questions the 'personality' of God.)

Cieszkowski, A. von (1838) *Prolegomena zur Historiosophie (Prolegomena to The Wisdom of History)*, Berlin: Viet.(Cieszkowski's major study, revising Hegel's philosophy of history.)

Daub, K. (1833) *Die dogmatische Theologie (The Dogmatic Theology of Our Times)*, Heidelberg: J.B. Mohr.(A standard and cautious early Hegelian interpretation of traditional theological topics.)

Dilthey, W. (1905) *Die Jugendgeschichte Hegels (The Young Hegel)*, Abhandlungen der Königlich Preussiche Akademie der Wissenchaften; repr. *Wilhelm Diltheys gesammelte Schriften*, vol. 4, Leipzig and Berlin: D.G. Teubner, 1921.(Led to an important reassessment of Hegel's early writings.)

Engels, F. (1886) *Ludwig Feuerbach und der Ausgang der klassischen deutschen Philosophie*, in *Die neue Zeit*; trans. *Ludwig Feuerbach and the End of Classical German Philosophy*, in K. Marx and F. Engels, *Selected Works in One Volume*, London: Lawrence & Wishart, 1968, 596-632.(Traces the development from Hegel to dialectical materialism.)

Erdmann, J.E. (1834-53) *Versuch einer wissenschaftlichen Darstellung der Geschichte der neueren Philosophie (Attempt at a Scientific Presentation of the History of Modern Philosophy)*, Leipzig: Riga & Dorpat, 6 vols. (One of the most comprehensive and magisterial Hegelian accounts of philosophy from Descartes to Hegel. Very incisively written.)

Erdmann, J.E. (1841) *Grundriß der Logik und Metaphysik (Outline of Logic and Metaphysics)*, Halle: J.H. Lippert; 5th edn, 1875.(Succinct and careful restatement of the essential categories of Hegel's Logic.)

Erdmann, J.E. (1865-7) *Grundriß der Geschichte der Philosophie*, Berlin: Wilhelm Hertz, 2 vols; 4th edn, 1895-6; trans. W.S. Hough, *A History of Philosophy*, London: Swan Sonnenschein, 3 vols, 1890.(Influential classic history of philosophy, thoroughly and carefully written from a Right Hegelian perspective.)

Feuerbach, L. (1839) 'Zur Kritik der Hegelschen Philosophie', in *Sämtliche Werke*, Leipzig: Otto Wigand, 1846-, vol. 2, 185-232; trans. Z. Hanfi, 'Towards a Critique of Hegelian Philosophy', in Z. Hanfi (ed.) *The Fiery Brook: Selected Writings of Ludwig Feuerbach*, New York: Anchor Books, 1972; also in L.S. Stepelevich (ed.) (1983) and R. Stern (ed.) (1993), I: 100-30. (This work established Feuerbach's position as a spokesman for the Hegelian Left.)

Feuerbach, L. (1841) *Das Wesen des Christentums*, in *Sämtliche Werke*, Leipzig: Otto Wigand, 1846-, vol. 7; trans. M.A. Evans (George Eliot), *The Essence of Christianity*, 1854; new edn, New York: Harper & Row, 1957.(Feuerbach's most famous work, in which he puts forward most clearly the Left Hegelian critique of religion and metaphysics.)

Feuerbach, L. (1842) *Vorläufige Thesen zur Reformation der Philosophie (Provisional Theses Towards the Reform of Philosophy)*, in *Sämtliche Werke*, Leipzig: Otto Wigand, 1846-, vol. 2, 244-88.(Develops Feuerbach's humanistic materialism.)

Feuerbach, L. (1843) *Grundsätze der Philosophie der Zukunft*, in *Sämtliche Werke*, Leipzig: Otto Wigand, 1846-, vol. 2, 269-346; trans. M.H. Vogel, *Principles of the Philosophy of the Future*, Indianapolis, IN, and New York: Bobbs-Merrill, 1966.(Develops Feuerbach's materialist and humanistic critique of Hegel.)

Fischer, K. (1852) *System der Logik und Metaphysik oder Wissenschaftslehre (System of Logic and Metaphysics or Doctrine of Science)*, Stuttgart: C.P. Scheitlin; 3rd edn, 1909.(Lucid and brief exposition of Hegel's logical doctrines.)

Fischer, K. (1852-77; 1897-1904) *Geschichte der neueren Philosophie (History of Modern Philosophy)*, 10 vols in 11, Heidelberg: Carl Winters Universitätsbuchhandlung.(A classic, massively detailed history of philosophy from Descartes to Schopenhauer, with the concluding two-volume monograph on Hegel.)

Gabler, G.A. (1827) *Kritik des Bewußtseins (Critique of Consciousness)*, Leiden: A.H. Adriani; 2nd edn, 1901. (One of the very few early examinations of Hegel's *Phenomenology*, written by Hegel's follower and successor in Berlin.)

Gadamer, H.-G. (1960) *Wahrheit und Methode*, Tübingen: J.C.B. Mohr; trans. W. Glen-Doepel, *Truth and*

Method, London: Sheed & Ward, 1979.(Gadamer’s hermeneutic reappropriation of Hegel’s concept of experience from a Heideggerian perspective.)

Gadamer, H.-G. (1971) *Hegels Dialektik: Fünf hermeneutische Studien*, Tübingen: J.C.B. Mohr; trans. P. Christopher Smith, *Hegel’s Dialectic: Five Hermeneutical Studies*, New Haven, CT, and London: Yale University Press, 1976.(Perceptive examination of key themes in Hegel in relation to the classical tradition and Heidegger.)

Gans, E. (1824-35) *Das Erbrecht in weltgeschichtlicher Entwicklung (The Law of Inheritance Considered in its World-Historical Development)*, vols 1 and 2, Berlin: Maurische Buchhandlung; vols 3 and 4, Stuttgart and Tübingen: Cotta.(The major Hegelian contribution to the history of law, written in conscious opposition to the ‘Historical School’.)

Glockner, H. (1929-40) *Hegel*, Stuttgart: Frommann, 2 vols.(Typical expression of vitalist neo-Hegelianism, emphasizing the cultural concreteness and historical richness of Hegelian thought.)

Glockner, H. (1931) ‘Hegelrenaissance und Neuhegelianismus’ (The Hegel Renaissance and Neo-Hegelianism), *Logos* 20; repr. with other contributions in *Hegel-Studien*, Beiheft 2, Bonn: Bouvier.(A historical account of the early twentieth-century Hegel revival in Germany by one of the leading participants.)

Habermas, J. (1968) *Erkenntnis und Interesse*, Frankfurt: Suhrkamp; trans. J.J. Shapiro, *Knowledge and Human Interest*, Oxford: Polity Press, 1987.(Part 1 contrasts Hegel’s critique of knowledge with that of Kant and Marx.)

Habermas, J. (1971) *Theorie und Praxis*, Frankfurt: Suhrkamp, 4th edn; trans. J. Viertel, *Theory and Practice*, Oxford: Polity Press, 1973.(Contains several influential essays on Hegel.)

Habermas, J. (1985) *Der philosophische Diskurs der Moderne*, Frankfurt: Surhkamp; trans. F. Lawrence, *The Philosophical Discourse of Modernity*, Cambridge, MA: MIT Press, 1987.(Traces the critique of modernity as it developed after Hegel.)

Haering, T. (1929-38) *Hegel. Sein Wollen und sein Werk (Hegel: His Life and his Works)*, Leipzig and Berlin, 2 vols; repr. Aalen: Scientia, 1979.(Enormous genetic study of Hegel’s thought from 1790-1807, stressing Hegel’s practical and political concerns. Typical of the organic ‘communitarian’ interpretation of conservative neo-Hegelianism.)

Hartmann, N. (1923, 1929) *Die Philosophie des deutschen Idealismus (The Philosophy of German Idealism)*, Berlin and Leipzig, 2 vols; 3rd edn, repr. Berlin: de Gruyter, 1974.(An important ontological and ‘realist’ reading of Hegel, emphasizing his systematic ambitions and debt to the classical tradition.)

Haym, R. (1857) *Hegel und seine Zeit (Hegel and his Times)*, Berlin: Rudolph Gaertner.(Contains an influential attack on Hegel as an apologist for the Prussian Restoration.)

Henning, L. von (1824) *Prinzipien der Ethik in historischer Entwicklung (Principles of Ethics in Historical Development)*, Berlin: Friedrich August Herbig.(A lucid brief outline of the history of ethics, dedicated to Hegel.)

Hess, M. (1837) *Die heilige Geschichte der Menschheit (The Sacred History of Humanity)*, Stuttgart: Hallberg’sche Verlagshandlung.(Important document of the transformation of Hegel’s philosophy of history as a secularized eschatology.)

Jüngel, E. (1977) *Gott als Geheimnis der Welt*, Tübingen: J.C.B. Mohr, trans. D.L. Guder, *God as the Mystery of the World*, Edinburgh: T. & T. Clark, 1983.(A magisterial contribution to philosophical theology, influenced by Barth and Heidegger, which engages seriously and productively with Hegel and the Left Hegelian tradition.)

Korsch, K. (1930) *Marxismus und Philosophie*, Leipzig: C.L.Hirschfeld, trans. F. Halliday, *Marxism and Philosophy*, London: New Left Books, 1970.(This work reopened the question of Marx’s debt to Hegelian idealism.)

Kroner, R. (1921-4) *Von Kant bis Hegel (From Kant to Hegel)*, Tübingen: Mohr Verlag, 2 vols; 3rd edn, repr. 1977.(The most substantial single product of neo-Hegelianism, interpreting Hegel as the fitting culmination of the entire tradition of classical and idealist thought from a Christian perspective.)

Lasson, G. (1916) *Was heißt Hegelianismus? (What is Hegelianism?)*, Berlin: Reuther & Reichard.(An uncritical and enthusiastic neo-Hegelian manifesto from a rather nationalistic perspective.)

Litt, T. (1948) *Denken und Sein (Thought and Being)*, Zurich: S. Hirzel.(Attempts to apply reformed Hegelian categories to central metaphysical and epistemological questions.)

Litt, T. (1948) *Mensch und Welt (Man and World)*, Heidelberg: Quelle & Meyer, 1961.(A quasi-Hegelian reinterpretation of philosophical anthropology, strongly influenced by Dilthey.)

Litt, T. (1953) *Hegel*, Heidelberg: Quelle and Meyer.(An appreciative assessment and reconstruction of Hegel, but critical of allegedly panlogistic and totalizing elements.)

Lüwith, K. (ed.) (1962) *Die Hegelsche Linke (The Hegelian Left)*, Stuttgart: Frommann-Holzboog.(Selected texts

by members of the Left Hegelian school.)

Lübbe, H. (ed.) (1962) *Die Hegelsche Rechte (The Hegelian Right)*, Stuttgart: Frommann-Holzboog.(Selected texts by members of the Right Hegelian school.)

Lukács, G. (1923) *Geschichte und Klassenbewußtsein*, Berlin: Der Malik-Verlag; repr. in *Werke*, Berlin: Luchterhand, vol. 2, 1968; trans. R. Livingstone, *History and Class Consciousness*, London: Merlin, 1971. (Enormously influential statement of the Hegelian dimension of Marxian thought which broke with the prevailing mechanistic and deterministic approach.)

Lukács, G. (1948) *Der junge Hegel*, Zurich: Europa, 2 vols; trans. R. Livingstone, *The Young Hegel*, London: Merlin, 1975.(Classical Marxist account of Hegel's development, implicitly written to correct vulgar Marxist readings of the idealist background.)

Lukács, G. (1963) *Ästhetik (Aesthetics)*, Berlin: Luchterhand; repr. in *Werke*, Berlin: Luchterhand, 1968, vols 11-12.(The most sustained example of a Marxian appropriation of classical German Idealist aesthetics and the concept of 'mimesis'.)

Lukács, G. (1971-2) *Zur Ontologie des gesellschaftlichen Seins*, Neuwied and Berlin: Luchterhand; trans. D. Fernbach, *The Ontology of Social Being*, London: Merlin, 1978-80, 3 vols.(Part of a projected larger work that was to present a systematic materialist version of the Hegelian legacy.)

Marcuse, H. (1932) *Hegels Ontologie*, Frankfurt: Klostermann; 3rd edn, repr. 1975; trans. S. Benhabib, *Hegel's Ontology*, Cambridge, MA: MIT Press, 1987.(An interesting dynamic and ontological interpretation of Hegel's Logic in the light of the *Phenomenology* and the earlier writings.)

Marcuse, H. (1941) *Reason and Revolution*, London: Oxford University Press.(Emphasizes the fundamental continuity between Hegel and Marx and offers a qualified defence of Hegel's political philosophy against its liberal critics.)

Marheinecke, P.K. (1819) *Die Grundlehren der christlichen Dogmatik als Wissenschaft (The Fundamental Doctrines of Christian Dogmatics as Science)*, Berlin: F. Dummler; 2nd edn, Duncker & Humblot, 1827.(One of the first thoroughgoing applications of Hegel's thought to Christian theology as a whole.)

Marx, K. (1843) 'Zur Kritik der Hegelschen Rechtsphilosophie', *Deutsch-Französische Jahrbücher*, issues 1 and 2; trans. R. Livingstone and G. Benton, *Critique of Hegel's Philosophy of Right*, in L. Colletti (ed.) *Early Writings*, Harmondsworth: Penguin, 1975, 57-198.(Offers a fundamental critique of Hegel from a Left Hegelian perspective.)

Marx, K. (1975) *Early Writings*, trans. R. Livingstone and G. Benton, Harmondsworth: Penguin.(A useful paperback edition, which contains Marx's posthumously published 'Critique of Hegel's Doctrine of the State' and 'Economic and Philosophical Manuscripts' and his (1844), as 'A Contribution to the Critique of Hegel's Philosophy of Right'.)

Michelet, K.L. (1837-8) *Geschichte der letzten Systeme der Philosophie in Deutschland von Kant bis Hegel (History of the Most Recent Systems of Philosophy in Germany from Kant to Hegel)*, Berlin: Dunker and Humblot, 2 vols.(A lucid history of German idealism by a representative of the liberal Hegelian 'centre'.)

Moltmann, J. (1965) *Theologie der Hoffnung*, Munich: Kaiser; trans. J.W. Leitch, *Theology of Hope*, London: SCM Press, 1967.(Indicative of reaction against existentialist dialectics and a renewed theological interest in the social dimension of the German Idealist tradition.)

Moltmann, J. (1972) *Der gekreuzigte Gott*, Munich: Chr Kaiser; trans. R.A. Wilson and J. Bowden, *The Crucified God*, London: SCM Press, 1974.(Approach to a liberation theology of the cross influenced by Marxist-Christian dialogue.)

Pannenberg, W. (1967) *Grundfragen systematischer Theologie (Fundamental Questions of Systematic Theology)*, Göttingen: Vandenhoek & Ruprecht.(Draws on Hegel and idealist thought to articulate a rational philosophical theology and defend the concept of progressive historical revelation.)

Rosenkranz, J.K.F. (1837) *Psychologie, oder die Wissenschaft vom subjektiven Geistes (Psychology, or the Science of Subjective Spirit)*, Königsberg: Bornträger.(A succinct outline of the Hegelian philosophy of mind.)

Rosenkranz, J.K.F. (1840a) *Kritische Erläuterungen des Hegelschen Systems (Critical Exposition of the Hegelian System)*, Königsberg: Bornträger.(Interesting collection of sympathetic essays on various aspects of Hegel, reflecting many of the debates within the school in the 1830-40 period.)

Rosenkranz, J.K.F. (1840b) *Geschichte der Kantischen Philosophie (History of Kantian Philosophy)*, Leipzig: Leopold Voss; repr. in J.K.F. Rosenkranz and F.W. Schubert (eds) *Immanuel Kants sämmtliche Werke*, vol. 12. (A supplement to Rosenkranz's edition of Kant's works indicating the 'completion' of Critical Philosophy in Hegelian idealism.)

Rosenkranz, J.K.F. (1844) *Hegels Leben (Hegel's Life)*, Berlin; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1988.(Rosenkranz's official biography of Hegel.)

Rosenkranz, J.K.F. (1858-9) *Wissenschaft der logischen Idee (Science of the Logical Idea)*, Köningsberg: Borntrager, 2 vols; repr. Osnabruck: Zeller, 1972.(A major critical reworking of Hegel's Logic.)
Rosenkranz, J.K.F. (1870) *Hegel als deutscher Nationalphilosoph (Hegel as German National Philosopher)*, Leipzig: Duncker & Humblot; extracts translated in Stern (ed.) 1993, vol. 1, 256-297.(A contemporary study of Hegel's thought and influence.)
Ruge, A. (1842) 'Hegels Rechtsphilosophie und die Politik unserer Zeit', *Deutscher Jahrbücher* 189, 190; trans. J.A. Massey, 'Hegel's *Philosophy of Right* and the Politics of Our Times', in L.S. Stepelevich (ed.), *The Young Hegelians*, Cambridge: Cambridge University Press, 211-36.(Offers a critique of Hegel from a Young Hegelian standpoint.)
Schelling, F.W.J. (1833-4) *Zur Geschichte der neueren Philosophie*, in *Sämtliche Werke*, ed K.F.A. Schelling, 14 vols, Stuttgart: Cotta, 1856-91, vol. 10: 1-200; trans. A. Bowie, *On the History of Modern Philosophy*, Cambridge: Cambridge University Press, 1994.(Contains the text of Schelling's important critical lectures on Hegel given in Munich and later in Berlin, from the perspective of Schelling's own 'positive philosophy'.)
Stepelevich, L.S. (ed.) (1983) *The Young Hegelians*, Cambridge: Cambridge University Press.(An anthology of writings by Strauss, Cieszkowski, Feuerbach, Bruno Bauer, Ruge, Edgar Bauer, Engels, Marx, Stirner, Hess and Schmidt.)
Stirner, M. (1845) *Der Einzige und sein Eigentum*, Leipzig: Otto Wigand; 2nd edn, 1882; trans. D. Leopold, *The Ego and Its Own*, Cambridge: Cambridge University Press, 1994.(Takes Hegel's concept of the modern subject to an individualist extreme.)
Strauss, D.F. (1835-6) *Das Leben Jesu, kritisch bearbeitet*, Tübingen: Osiander; trans. M.A. Evans (George Eliot), *The Life of Jesus*, London: Chapman Brothers, 1846; repr. Philadelphia: Fortress, 1972.(A seminal work of nineteenth-theology which undermined traditional conceptions of scripture and profoundly influenced theological 'modernism'.)
Strauss, D.F. (1837-8) *Streitschriften zur Verteidigung meiner Schrift (Polemical Writings in Defence of My Work)*, Tübingen: Osiander.(Spirited defence of the *Life of Jesus* which gives a good sense of the theological polemics of the period.)
Theunissen, M. (1970) *Hegels Lehre vom absoluten Geist als theologisch-politischer Traktat (Hegel's Doctrine of Absolute Spirit as Theological-Political Treatise)*, Berlin: de Gruyter.(Major sympathetic exposition of Hegel's religious philosophy which attempts to undercut standard Left and Right interpretations.)
Theunissen, M. (1980) *Sein und Schein (Being and Appearance)*, Frankfurt: Suhrkamp.(Pursues an 'intersubjective' approach to the claims of Hegel's Logic as a potential theory of communicative freedom.)
Trendelenburg, F.A. (1840) *Logische Untersuchungen (Logical Investigations)*, Leipzig: S. Hirzel; 2nd edn, 1862.(An influential critique of Hegel's Logic.)
Trendelenburg, F.A. (1843) *Die logische Frage in Hegels System (The Logical Question in Hegel's System)*, Leipzig: F.A. Brockhaus.(Further develops his critique of Hegel's Logic.)
Vatke, W. (1835) *Die Religion des alten Testaments (The Religion of the Old Testament)*, Berlin: G. Bethge. (Important early treatment of the then-neglected area of Judaism from a Hegelian perspective.)
Vatke, W. (1841) *Die menschliche Freiheit in ihrem Verhältniss zur Sünde und zur göttlichen Gnade (Human Freedom in its Relation to Sin and Divine Forgiveness)*, Berlin: Bethge.(Applies Hegel's philosophy of spirit to elucidate classical theological questions concerning guilt and forgiveness.)
Vischer, F.T. (1837) *Über das Erhabene und Komische (On the Sublime and the Comic)*, ed. W. Oelmüller, Frankfurt: Suhrkamp, 1967.(Reprinted with other essays, this essay gives a thoroughly humanistic Left-Hegelian approach to an area rather neglected by Hegel.)
Vischer, F.T. (1846-54) *Ästhetik oder Wissenschaft des Schönen (Aesthetics or the Science of Beauty)*, Reutlingen & Leipzig: Carl Mäckens, 6 vols.(The largest Hegelian contribution to aesthetics which supplements Hegel by examining the question of natural beauty and the psychology of imagination.)
Windelband, W. (1910) 'Die Erneuerung der Hegelianismus' (The renewal of Hegelianism), *Sitzungsberichte der Heidelberger Akademie der Wissenschaften* 1 (10); repr. in W. Windelband, *Präludien*, Tübingen: J.C.B. Mohr, 6th edn, 1919, 273-89. (An important expression of the German resurgence of interest in Hegel at the beginning of the twentieth century.)
Zeller, E. (1844-52) *Die Philosophie der Griechen (Philosophy of the Greeks)*, 3 vols in 2, Tübingen: Friedrich Fues.(Very influential standard history of ancient philosophy from a Hegelian perspective which was widely studied and translated.)

3 France

Althusser, L. (1970) 'Sur le Rapport de Marx à Hegel', in J. D'Hondt (ed.) *Hegel et la pensée moderne*, Paris:

Presses Universitaires de France; trans. B. Brewster, 'Marx's Relation to Hegel', in *Politics and History*, London and New York: Verso, New Left Books, 1972; repr. in Stern (ed.) (1993) vol. 2, 511-28.(Argues that Marx understood Hegel's dialectic in terms of process.)

Bataille, G. (1955) 'Hegel, le mort et le sacrifice' (Hegel, Death and Sacrifice), *Deucalion* 5: 21-43; trans. J. Strauss, in A. Stoekl (ed.) *On Bataille*, New Haven, CT: Yale University Press, 1990.(A response to Kojève's reading of Hegel.)

Bertholet, R. (1907) 'Thèse: Sur la nécessité, la finalité et la liberté chez Hegel' (Thesis: On Necessity, Finality and Liberty in Hegel), *Bulletin de la Société française de philosophie* 115-140; repr. in R. Bertholet, *Evolutionisme et platonisme*, Paris: Alcan, 1908.(Defends Hegel against the accusation of absolute determinism, integral optimism and panlogicism, and discusses Hegel's influence on Marx.)

Cousin, V. (1828-9) *Cours de l'histoire de la philosophie (Lecture on the History of Philosophy)*, Paris: Pichon & Didier, 3 vols.(The text of Cousin's lectures given in the academic year 1828-9, which drew heavily on Hegelian ideas.)

Derrida, J. (1967) 'De l'économie restreinte à l'économie générale: un hégélianisme sans réserve' (From Restricted to General Economy: A Hegelianism Without Reserve), in *L'Écriture et la différance*, Paris: Éditions du Seuil; trans. A. Bass, *Writing and Difference*, London: Routledge, 1978, 251-77.(Examines Georges Bataille's response to Kojève's reading of Hegel.)

Derrida, J. (1972) 'Le puits et la pyramide: Introduction à la sémiologie de Hegel' (The Pit and the Pyramid: Introduction to Hegel's Semiology), in *Marges de la philosophie*, Paris: Minuit, 79-127; trans. A. Bass, *Margins of Philosophy*, Brighton: Harvester, 1982.(Discusses Hegel's theory of signs.)

Derrida, J. (1974) *Glas*, Paris: Éditions Galilée; trans. J.P. Leavey, Jr and R. Rand, *Glas*, Lincoln, NB: University of Nebraska Press, 1986.(Complex and ambitious reading of various Hegel texts.)

Herr, L. (1894) 'Hegel', *La Grande Encyclopédie* 19: 997-1003; repr. in L. Herr, *Choix d'Écrits*, Paris: Les Éditions Ridier, 1932, vol. 2, 109-40.(Contains a lengthy summary of the Hegelian system, which revived interest in his philosophy in this period.)

Hyppolite, J. (1946) *Genèse et structure de la Phénoménologie de l'esprit de Hegel*, Paris: Aubier, 2 vols; trans. S. Cherniak and J. Heckman, *Genesis and Structure of Hegel's Phenomenology of Spirit*, Evanston, IL: Northwestern University Press, 1974. (Hyppolite's magisterial and influential commentary.)

Hyppolite, J. (1948) *Introduction à la philosophie de l'histoire de Hegel (Introduction to Hegel's Philosophy of History)*, Paris: Rivère; Paris: Editions du Seuil, 1983.(A collection of short studies stressing the relation between the early and later Hegel.)

Hyppolite, J. (1953) *Logique et existence, essai sur la Logique de Hegel (Logic and Existence: Essays on Hegel's Logic)*, Paris: Presses Universitaires de France. (Explores the contested issue of the relation of Hegel's *Logic* to the 'existential' dimension of the *Phenomenology*.)

Hyppolite, J. (1955) *Études sur Marx et Hegel*, Paris: Rivière; trans. J.O. Neill, *Studies on Marx and Hegel*, London: Heinemann, 1969.(Stresses the place of Hegelian themes in Marx's early writings.)

Hyppolite, J. (1971) *Figures de la Pensée Philosophiques*, Paris: Presses Universitaires de France, 2 vols. (Contains all of Hyppolite's articles on Hegel.)

Kojève, A. (1946) 'Hegel, Marx et le christianisme', *Critique* 1: 339-66; trans. H. Gildin, 'Hegel, Marx and Christianity', *Interpretation* (1970) 1: 21-42; repr. in Stern (ed.), vol. 2: 359-82.(Argues for what he describes as an atheistic reading of Hegel.)

Kojève, A. (1947) *Introduction à la lecture de Hegel*, Paris: Gallimard; partially trans. J.H.Nichols in A. Bloom (ed.) *Introduction to the Reading of Hegel*, Ithaca, NY, and London: Cornell University Press, 1969.(Text of Kojève's influential lectures at l'École des Hautes-Études from 1933 to 1939.)

Koyré, A. (1970) *Etudes d'histoire de la pensée philosophique (Historical Studies of Philosophical Thought)*, Paris: Gallimard.(Contains the three most important essays by Koyré on Hegel: 'Note sur la langue et la terminologie hégélienne' ('Note on Hegelian Language and Terminology') (1931) 'Rapport sur l'état des études hégéliennes en France' ('Report on the State of Hegelian Studies in France') (1931) and 'Hegel à Iéna' ('Hegel in Jena') (1934).)

Levinas, E. (1961) *Totalité et Infini*, The Hague: Martinus Nijhoff; trans. A. Lingis, *Totality and Infinity*, The Hague: Martinus Nijhoff, 1969.(An ethical critique of the alleged totalizing ambitions of the metaphysical tradition in general and Hegel in particular.)

Merleau-Ponty, M. (1946) 'L'existentialisme chez Hegel' (Existentialism in Hegel), *Les Temps modernes* 1: 1311-19; repr. in M. Merleau-Ponty, *Sens et non-sens*, Paris: Nagel, 1948; trans. H.L. Dreyfus and P.A. Dreyfus, *Sense and Non-Sense*, Evanston, IL: Northwestern University Press, 1964, 63-70.(Discusses the existentialist themes in Hegel's thought.)

Nöel, G. (1897) *La Logique de Hegel (Hegel's Logic)*, Paris: Alcan; repr. Paris: Vrin, 1933, 1938 and 1967.(A careful and sympathetic analysis of Hegel's *Logic*, which first appeared in instalments in the *Revue de métaphysique et de morale* from 1894 to 1896.)

Renan, E. (1876) *Dialogues et fragments philosophiques (Dialogues and Philosophical Fragments)*, Paris: Calman Lévy.(Written between 1860 and 1871, Renan here attempts to combine Comtean positivism with religion, via an evolutionary reading of Hegel.)

Scherer, E. (1861) 'Hegel et l'hégélianisme' (Hegel and Hegelianism), *Revue des deux mondes* 31: 812-56.(Gives a reasonably sympathetic overview of the attitude to Hegel in this period.)

Sartre, J-P. (1943) *L'Être et le néant*, Paris: Gallimard; trans. H.E. Barnes, *Being and Nothingness*, London: Methuen, 1958.(Follows Kojève's reading of the master-slave dialectic in his account of being-for-others.)

Taine, H.-A. (1857) *Les Philosophes français du XIXe siècle (Nineteenth-Century French Philosophy)*, Paris: Hachette; 3rd edn, 1868, under the title *Les Philosophes classiques du XIXe siècle en France*, Paris: Hachette. (Clearly expresses Taine's admiration for Hegel.)

Vera, A. (1855) *Introduction à la philosophie de Hegel (Introduction to Hegel's Philosophy)*, Paris: A. Franck.(A popular introduction to Hegel's thought.)

Vera, A. (1864) *Essais de philosophie hégélienne (Essays on Hegelian Philosophy)*, Paris: G. Ballère.(The three essays are 'La peine de mort', 'Amour et philosophie', and 'Introduction à la philosophie d'histoire'.)

Wahl, J. (1929) *Le Malheur de la conscience dans la philosophie de Hegel (The Unhappy Consciousness in the Hegel's Philosophy)*, Paris: Rieder, 2nd edn, 1951; repr. New York and London: Garland, 1984; 119-147 trans. R. Northey in Stern (ed.) (1993) vol. 2: 284-310.(An influential study, which treats Hegel's notion of the Unhappy Consciousness as a key to the reading of his work.)

Wahl, J. (1938) *Études kierkegaardiennes (Kierkegaardian Studies)*, Paris: Éditions Montaigne; 2nd edn, Paris: Vrin, 1949, 159-171.(A collection of articles, including 'Hegel et Kierkegaard' and 'La lutte contre l'hégélianisme'.)

Weil, E. (1950a) *Logique de la philosophie (Logic of Philosophy)*. Paris: Vrin.(A major systematic appropriation of Hegel's thought as a comprehensive theory of categories.)

Weil, E. (1950b) *Hegel et l'état (Hegel and the State)* Paris: Vrin.(An important defence of Hegel's political thought against accusations of conservatism.)

4 Northern Europe

Heiberg, J.L. (1824) *Om den menneskelige Frihed (On Human Freedom)*, Kiel: Universitets Boghandlingen.(An early contribution by one of Hegel's students to the theory of the will.)

Kierkegaard, S.A. (1846) *Afsluttende uvidenskabelig Efterskrift*, in A.B. Drachmann, J.L. Heiberg and H.O. Lange (eds) *Samlede Vaerker*, Copenhagen: Gyldendalske Boghandel, 1962-4, vol. 7 of 20; trans. D.F. Swenson and W. Lowrie, *Concluding Unscientific Postscript*, Princeton, NJ: Princeton University Press, 1941. (Contains Kierkegaard's most developed critique of Hegel and Hegelianism.)

Martensen, H.L. (1850) *Den Christlige Dogmatik*, trans. W. Urwick *Christian Dogmatics*, Edinburgh: T. & T. Clark, 1866.(Martensen's major work.)

Nielsen, R. (1864-6) *Grundidéernes Logik (The Logic of Fundamental Ideas)*, Copenhagen.(An application of Hegel's Logic.)

5 Italy

Croce, B. (1907) *Ciò che è vivo e ciò cheè morto nella filosofia di Hegel*, Bari: Laterza; repr. in *Saggio Sullo Hegel*, Bari: Laterza, 1913; trans. D. Ainslie, *What is Living and What is Dead of the Philosophy of Hegel*, London: Macmillan, 1915.(Croce's influential critical study. *Saggio Sullo Hegel* also contains other studies in Hegel and German philosophy.)

Croce, B. (1902) *Estetica*, Bari: Laterza, 4th edn, 1912; Part 1 trans. C. Lyas, *The Aesthetic as the Science of Pure Expression of the Linguistic in General*, Cambridge: Cambridge University Press, 1992.(Draws on the German Idealist tradition, but emphasizes the formal autonomy of the art work.)

Croce, B. (1938) *Storia come pensiero e azione*, Bari: Laterza, 2nd edn; trans. S. Sprigge, *History as the Story of Liberty*, London: Allen & Unwin, 1941.(A liberal version of humanist historicism which de-theologizes the Hegelian idea of history, treating it as the progress of the consciousness of freedom.)

Gentile, G. (1913) *La riforma della dialettica Hegeliana e B. Spaventa*, Messina: Giuseppe Principato; 2nd edn, 1923; Part 1 trans. A.M. Armstrong, 'The Reform of Hegelian Dialectic', *Idealistic Studies* (1981) 11: 187-213. (Gentile's attempt at a critical reconstruction of Hegel's dialectical method as a 'method of immanence'.)

Sanctis, F. de (1870-1) *La storia della letteraria italiana*, Naples: Domenico & Antonio Morano, 2 vols; trans. J. Redfern, *History of Italian Literature*, London: Oxford University Press, 2 vols, 1932.(Important document in the influence of Hegel's ideas in Italy, strongly influenced by Hegel's approach to aesthetics.)

Spaventa, B. (1869) *Studi sull'etica hegeliana (Studies in Hegelian Ethics), Proceedings of the Royal Academy of Moral and Political Sciences of Naples*, vol. 4; repr. in *Opere*, ed. G. Gentile, Florence: Sansoni, 1972, vol. 1, 611-801.(An important and influential early study.)

6 America

Dewey, J. (1884) 'Kant and Philosophic Method', *Journal of Speculative Philosophy* 18: 162-74; repr. in Stern (ed.) (1993) vol. 2, 151-61.(Presents a reading of Kant from a Hegelian perspective.)

Dewey, J. (1930) 'From Absolutism to Experimentalism', in G.P. Adams and W.P. Montague (eds) *Contemporary American Philosophy*, New York, 1930, vol. 2, 13-27; repr. in R.J. Bernstein (ed.) *John Dewey, On Experience, Nature and Freedom*, New York: The Liberal Arts Press, 1960.(An autobiographical essay, in which Dewey expresses his attraction towards Hegelianism when a young man.)

Goetzmann, W. (ed.) (1973) *The American Hegelians: An Intellectual Episode in the History of Western America*, New York: Alfred A. Knopf.(Contains excerpts from the major works of main nineteenth-century American Hegelians, including Brokmeyer, Harris and Stallo.)

Harris, W.T. (1890) *Hegel's Logic, A Book on the Genesis of the Categories of the Mind: A Critical Exposition*, Chicago, IL: S.C. Griggs; New York: Kraus Reprint Co., 1970.(The major work on Hegel by one of the most significant nineteenth-century Hegelians.)

James, W. (1882) 'On Some Hegelisms', *Mind* old series 7: 186-208.(An exuberant attack on Hegel's metaphysics and method, as interpreted by James.)

James, W. (1909) *A Pluralistic Universe*, London: Longman, Green.(Contains James's most sustained critique of Hegel's thought.)

Peirce, C.S. (1868) 'Nominalism versus Realism' and 'What is meant by determined', *Journal of Speculative Philosophy* 2: 57-61, 190-1; repr. in Stern (ed.) (1993) vol. 2: 140-50.(A critical exchange on Hegel with W.T. Harris.)

Peirce, C.S. (1903) *Lectures on Pragmatism*, in C. Hartshorne, P. Weiss and A.W. Burcks (ed.) *Collected Papers of Charles Sanders Peirce*, Cambridge, MA: Belknap Press, vol. 5 of 8, 1-131.(Peirce discusses his relation to Hegel in Lecture 1.)

Royce, J. (1892) *Spirit of Modern Philosophy*, Boston, MA, and New York: Houghton Mifflin. (Contains a vivid and sympathetic account of Hegel's position.)

Royce, J. (1919) *Lectures on Modern Idealism*, ed J. Loewenberg, New Haven, CT: Yale University Press.(Text of Royce's lectures of 1906.)

7 Britain

Bosanquet, B. (1888) *Logic, or the Morphology of Knowledge*, Oxford: Clarendon Press.(Bosanquet's main account of his idealist logic and metaphysics, with clear affinities to Hegel.)

Bosanquet, B. (1899) *The Philosophical Theory of the State*, London and New York: Macmillan; 4th edn, 1923. (Presents Bosanquet's broadly Hegelian account of the State and its relation to the individual.)

Bosanquet, B. (1892) *A History of Aesthetics*, London: Macmillan; 2nd edn, 1904.(Contains chapter on Hegel.)

Bradley, F.H. (1876) *Ethical Studies*, Oxford: Clarendon Press; 2nd edn, 1927.(Shows the impact of Hegel's critique of Kant on Bradley's ethical thinking.)

Bradley, F.H. (1883) *Principles of Logic*, Oxford: Clarendon Press; 2nd edn, 1922.(Reveals Bradley's equivocal relation to Hegel's idealistic logic.)

Bradley, F.H. (1893) *Appearance and Reality*, Oxford: Clarendon Press; 2nd edn, 1897; ninth impression, corrected, 1930.(Bradley's main work, which in some respects has affinities to Hegel's metaphysics.)

Caird, E. (1883) *Hegel*, Edinburgh and London: Blackwood.(A short but well-respected study.)

Caird, E. (1892) *Essays on Literature and Philosophy*, Glasgow: James Maclehose, 2 vols. (Contains several essays with comments on Hegel, informed throughout by Hegelian ideas.)

Collingwood, R.G. (1924) *Speculum Mentis*, Oxford: Clarendon Press.(An attempt to write a phenomenology of forms of experience, influenced by the Italian neo-Hegelians.)

Collingwood, R.G. (1945) *The Idea of Nature*, Oxford: Clarendon Press.(Contains an unusually sympathetic account of Hegel's philosophy of nature.)

Collingwood, R.G. (1946) *The Idea of History*, Oxford: Clarendon Press; revised edn, ed. J. van Dussen, Oxford:

Clarendon Press, 1993.(Contains a sympathetic discussion of Hegel's philosophy of history.)
McTaggart, J.M.E. (1896) *Studies in the Hegelian Dialectic*, Cambridge: Cambridge University Press; 2nd edn, 1922.(A detailed and sympathetic study of Hegel's dialectical method and its place in his metaphysics.)
McTaggart, J.M.E. (1901) *Studies in Hegelian Cosmology*, Cambridge: Cambridge University Press; 2nd edn, 1918.(Contains a discussion of a range of issues, mostly concentrating on Hegel's metaphysics, ethics, political philosophy and philosophy of religion.)
McTaggart, J.M.E. (1910) *A Commentary on Hegel's Logic*, Cambridge: Cambridge University Press.(A detailed commentary.)
Ritchie, D.G. (1890-1) 'Darwin and Hegel', *Proceedings of the Aristotelian Society* 1: 55-74; repr. in Stern (ed.) (1993) vol. 2: 41-59.(Attempts to defend Hegel against Seth's critique.)
Seth, A. (1887) *Hegelianism and Personality*, London: Blackwood.(Contains Seth's most developed and influential critique of Hegel.)
Stirling, J.H. (1865) *The Secret of Hegel*, London: Longman, Roberts & Green, 2 vols; 2nd edn, Edinburgh: Oliver & Boyd, 1898; repr. Bristol: Thoemmes, 1990.(The first major study on Hegel to appear in Britain.)
Wallace, W. (1874) *Prolegomena to the Study of Hegel's Philosophy*, Oxford: Oxford University Press. (Wide-ranging, sympathetic essays, trying to illuminate the background and content of Hegel's Logic.)

Secondary literature

Bellamy, R. (1987) *Modern Italian Social Theory*, Cambridge: Polity Press.(Contains chapters on Labriola, Croce, Gentile and Gramsci, and traces the roots of their thought back to the native idealist tradition.)
Bowie, A. (1993) *Schelling and Modern European Philosophy*, London: Routledge. (Chapter 6 provides a clear account of Schelling's critique of Hegel.)
Bradley, J. (1979) 'Hegel in Britain: A Brief Survey of British Commentary and Attitudes', *Heythrop Journal* 20: 1-24 and 163-82.(A useful survey of Hegel's reception in Britain.)
Brazill, W.J. (1970) *The Young Hegelians*, New Haven, CT: Yale University Press. (Discusses the origins and development of the Young Hegelian school.)
Butler, J. (1987) *Subjects of Desire: Hegelian Reflections in Twentieth Century France*, New York: Columbia University Press.(A specialized treatment of the impact of Hegel on twentieth century French thought.)
Cornehl, P. (1971) *Die Zukunft der Versöhnung (The Future of Reconciliation)*, Göttingen: Vandenhoek & Ruprecht.(An outstanding study of the immediate theological reception of Hegel, focusing on the question of eschatology.)
Derbolav, J. (1969) 'Über die gegenwärtigen Tendenzen der Hegelaneignung in Deutschland' (On the Current Trends in Hegel Reception in Germany), *Hegel-Studien* 5: 267-91.(On the various forms of German Hegel reception since 1945.)
Descombes, V. (1980) *Modern French Philosophy*, Cambridge: Cambridge University Press.(Traces the importance of Kojève's Hegel reception and the eventual reaction to it for an entire generation of contemporary French thinkers.)
Easton, L. (1966) *Hegel's First American Followers: The Ohio Hegelians*, Athens, OH: Ohio University Press.(A useful discussion of Stallo, Kaufman, Conway and Willich, with extracts from their works.)
Fisch, M.H. (1974) 'Hegel and Peirce', in J.J. O'Malley, K.W. Algozin and F.G. Weiss (eds) *Hegel and the History of Philosophy*, The Hague: Martinus Nijhoff, 171-93.(Provides an account of Peirce's view of Hegel.)
Flower, E. and Murphey, M.G. (1977) *A History of Philosophy in America*, New York: Capricorn and Putnam, 2 vols.(Chapter 8 gives a very useful account of the St Louis Hegelians.)
Gasché, R. (1986) *The Tain of the Mirror. Derrida and the Philosophy of Reflection*, Cambridge, MA: Harvard University Press.(A philosophical presentation of Derrida's thought against a Hegelian background.)
Glockner, H. (1965) *Beiträge zum Verständnis und zur Kritik Hegels (Contributions towards the Understanding and Critique of Hegel)*, *Hegel-Studien*, Beiheft 2, Bonn: Bouvier.(Contains 'Hegelrenaissance und Neuhegelianismus' and other relevant essays on neo-Hegelianism.)
Hodgson, P.C. (1966) *The Formation of Historical Theology: A Study of F.C. Baur*, New York: Harper & Row. (On F.C. Baur.)
Jacobitti, E.E. (1981) *Revolutionary Humanism and Historicism in Modern Italy*, New Haven, CT: Yale University Press.(A useful account of Hegel's influence in Italy, including details of his nineteenth-century reception.)
Jarvis, S. (1996) *Adorno*, Cambridge: Polity Press.(A comprehensive reading of Adorno which strongly emphasizes the Hegelian dimension.)
Jay, M. (1984) *Marxism and Totality: The Adventures of a Concept from Lukács to Habermas*, Cambridge: Polity

Press.(Traces the legacy of the Hegelian emphasis upon totality in Western Marxism.)
Kelly, M. (1992) *Hegel in France*, Birmingham: Birmingham Modern Languages Publications.(A thorough survey of Hegel's reception in France from 1800 to the present, with a useful bibliography.)
Kortian, G. (1980) *Metacritique: The Philosophical Argument of Jürgen Habermas*, Cambridge: Polity Press. (Chapter 2 discusses Habermas's reception and engagement with Hegel.)
Löwith, K. (1941) *From Hegel to Nietzsche: The Revolution in Nineteeth-Century Thought*, trans. D.E. Green, London: Constable, 1965.(A classic study of the fate of the Hegelian system and the subsequent polarization of right and left interpretations.)
Mader, J. (1993) *Philosophie in der Revolte: das Ende des Idealismus im 19 Jahrhundert (Philosophy in Revolt: The End of Idealism in the Nineteenth Century)*, Vienna: Universitätsverlag.(A recent account of the break-up of the Hegelian school in the nineteenth century.)
Muirhead, J.H. (1931) *The Platonic Tradition in Anglo-Saxon Philosophy*, London: Allen & Unwin, and New York: Macmillan.(A classic and still useful study of the development of British Idealism.)
Oelmüller, W. (1959) *F.Th.Vischer und das Problem der nachhegelschen Ästhetik (Vischer and the Problem of Posthegelian Aesthetics)*, Stuttgart: Kohlhammer. (Provides a discussion of Vischer and the post-Hegelian legacy in aesthetics.)
Oldrini, G. (1973) *La cultura filosofica napoletana dell'ottocentro (The Neapolitan Culture of the Nineteenth Century)*, Bari: Laterza.(The main Italian study of the Neapolitan Hegelians.)
Riedel, M. (1967) 'Hegel und Gans', in H. Braun and M. Riedel (eds) *Natur und Geschichte. Karl Löwith zum 70. Geburtstag*, Stuttgart: Kohlhammer, 257-73.(On Eduard Gans' relation to Hegel and Hegelianism.)
Roth, M. (1988) *Knowing and History: Appropriations of Hegel in Twentieth Century France*, Ithaca, NY: Cornell University Press.(A useful discussion of twentieth-century French Hegelianism.)
Sass, H.M. and Wartofsky, M.W. (eds) (1978) *Feuerbach, Marx and the Left Hegelians*, *The Philosophical Forum* 8: 1, 2 and 3.(A special issue on the Left Hegelians.)
Schmidt, A. (1962) *Der Begriff der Natur in der Lehre von Marx*, Frankfurt: Europäische Verlagsanstatt; trans. B. Fowkes, *The Concept of Nature in Marx*, London: New Left Books, 1971.(Analyses Marx's complex philosophical relation to Hegel and the Young Hegelians.)
Schmidt, A. (1962) *Geschichte und Struktur*, Munich: Carl Hanser; trans. J. Herf, *History and Structure*, Cambridge, MA: MIT Press, 1981.(Criticizes Althusser and the structuralist tendency to separate an early 'Hegelian' from a mature 'scientific' phase in Marx's thought.)
Stern, R. (1994) 'British Hegelianism: A Non-Metaphysical View?', *European Journal of Philosophy* 2: 293-321. (Examines the impact of Schelling on the readings of Hegel offered by the British Idealists.)
Thulstrup, N. (1967) *Kierkegaard's Relation to Hegel*, trans. G.L. Stengren, Princeton, NJ: Princeton University Press, 1980.(A thorough account of Kierkegaard's acquaintance with Hegel's writings and those of his followers in Denmark.)
Toews, J. (1980) *Hegelianism: The Path Towards Dialectical Humanism, 1805-1841*, Cambridge: Cambridge University Press.(A study of the Hegelian movement before and after Hegel's death.)
Wartofsky, M.W. (1977) *Feuerbach*, Cambridge: Cambridge University Press. (Contains a thorough discussion of Hegel's place in Feuerbach's intellectual development.)

Hegelianism, Russian

In Russian intellectual history the so-called 'remarkable decade' of 1838-48 (P.V. Annenkov's expression) could be characterized as a truly 'philosophical epoch'. Speculative philosophy was seen by then as directly relevant to all important questions of national existence. A similar situation obtained then, in exactly the same years, in the lands of partitioned Poland. In both countries all philosophical discussions revolved around Hegel, whose system was perceived as the culminating point in the development of Western philosophy. In Russia the fascination with Hegelianism was widespread and profound, reaching distant provincial centres and leaving its mark on literature. 'Philosophical notions', wrote Ivan Kireevskii in 1845, 'have become quite commonplace here now. There is scarcely a person who does not use philosophical terminology, nor any young man who is not steeped in reflections on Hegel'. Herzen provides an identical testimony. Hegel's works, he wrote,

> *were discussed incessantly; there was not a paragraph in the three parts of the Logic, in the two of the Aesthetics, the Encyclopaedia and so on, which had not been the subject of desperate disputes for several nights together. People who loved each other avoided each other for weeks at a time because they disagreed about the definition of 'all-embracing spirit', or had taken as a personal insult an opinion on the 'absolute personality and its existence in itself'.*
>
> *(Herzen [1853] 1968: 398)*

This vivid reception of Hegelianism was a socially important phenomenon, meeting several deep-seated psychological demands of the young Russian intelligentsia. First, as in Germany, speculative idealism provided the intelligentsia with a sort of compensation for the paralysis of public life under authoritarian government. Second, Hegelian philosophy was welcomed as an antidote to introspective day-dreaming and attitudes of Romantic revolt; in this context Hegelianism was largely interpreted as a philosophy of 'reconciliation with reality'. Somewhat later this conservative interpretation of Hegelianism was replaced by a Left-Hegelian philosophy of rational and conscious action; at this stage Hegelianism came to be a powerful instrument in the struggle against Slavophile conservative Romanticism. Both as a philosophy of reconciliation and as a philosophy of action Russian Hegelianism was above all a philosophy of reintegration; a philosophy which helped young intellectuals in overcoming their feeling of alienation and in building bridges between their ideals and reality.

1 From reconciliation with reality to philosophy of action

In the 1830s the chief centre of Russian Hegelianism was the Moscow circle of Nikolai Stankevich (1813-40). It counted among its members the literary critic Vissarion Belinskii, the future anarchist Mikhail Bakunin, the liberal historian Timofei Granovskii and even the future Slavophile Konstantin Aksakov; closely associated with the circle was the great Russian writer Ivan Turgenev.

Stankevich's intellectual evolution was a development from Schellingian pantheistic Romanticism (see Schellingianism), combined with Schiller's militant 'subjectivism', to Hegelianism, interpreted initially as a philosophy of reconciliation. In the last year of his life he was greatly impressed by August Cieszkowski's *Die Prolegomena zur Historiosophie (Prolegomena to the Wisdom of History)* (Berlin, 1838) and tried to link Cieszkowski's idea of the 'translation of philosophy into action' (see Cieszkowski, A.) with the Feuerbachian rehabilitation of feelings and senses (see Feuerbach, L.). A similar pattern can be discerned in the intellectual evolution of Bakunin and Belinskii.

After Stankevich left Russia to study philosophy in Berlin (in 1837), the leadership in his circle was taken over by Mikhail Bakunin. His conversion to Hegelianism took place in 1838 and was heavily influenced by the mystical ideas of the German Romantics (see Romanticism, German). In his 'Gimnazicheskie rechi Gegelia. Predislovie perevodchika' (Foreword to Hegel's School Addresses) (1838) Bakunin proclaimed 'the reconciliation with reality in all respects and all spheres of life'. He meant by this a radical overcoming of the abstract intellect (*Verstand* - *rassudok*) whose separation from the historical reason (*Vernunft* - *razum*) had caused the terrible modern disease of subjective revolt, finding expression in Schiller, Fichte and the rebellious poetry of Byron. The most advanced stage of this disease was embodied, in his view, by France - a country mortally sick of a barren revolutionary negation.

In 1840 Bakunin, following Stankevich, went to Berlin - the ‘New Jerusalem’ of the Russian idealists - and very soon became an extreme Left Hegelian, an apostle of a radical, revolutionary ‘negation’ of the ’old world’. In his article in ’Die Reaktion in Deutschland’ (Reaction in Germany) (published under the pseudonym Jules Elysard in *Deutsche Jahrbücher für Wissenschaft und Kunst*) (1842) he elaborated a theoretical foundation for a revolutionary philosophy of action. It was one of the first serious attempts to make a radical ‘Left-wing’ interpretation of the Hegelian dialectic. In contrast to Hegel, Bakunin saw transcending (*Aufhebung*), in other words the final result of dialectic process, as a complete destruction of the past. The essence of contradiction, he argued, is not an equilibrium but ‘the preponderance of the negative’, whose role is decisive. Revolutionary negation includes within itself the totality of the contradiction and so alone has absolute legitimacy. The article ended with the words: ‘The joy of destruction is also a creative joy’. Shortly thereafter Bakunin cut himself off from speculative philosophy and began his ‘search for God’ in revolutionary praxis.

The story of Vissarion Belinskii’s involvement in Hegelianism was especially dramatic (see Belinskii, V.). Before his acquaintance with Hegel he was influenced by the activist idealism of Fichte, interpreted by him as a philosophy of heroic voluntarism. However, towards the end of 1837, he became struck by Hegel’s famous thesis: ‘What is real is rational, what is rational is real’. This led him to proclaim a philosophy of reconciliation, condemning all sorts of voluntaristic ’subjectivism’ in the name of objective and rational laws of history. Unlike Bakunin, Belinskii was far from mystical idealism; hence he interpreted reconciliation in an empirical way, forcing himself to glorify tsarist autocracy and to accept the stifling social realities of Nicolaevan Russia. But this sort of self-indoctrination did not help him to overcome the painful feelings of alienation. In March 1841 he revolted against Hegel’s historiosophical theodicy, vindicating the value of individual freedom and the heritage of the French Enlightenment. Initially it was a purely moral protest but very soon it became philosophically justified by a radical reinterpretation of the notion of ‘rational reality’, stressing the role of ‘dialectical negation’. In this way Belinskii rejected conservative historicism, setting against it the idea of free, autonomous individuality, non-reducible to a mere instrument of ’vast impersonal forces’ (Isaiah Berlin’s expression). At the end of his life he sympathized with the philosophy of Feuerbach, in which he recognized a form of anthropocentrism liberating the individual from the despotic power of ’the absolute spirit’ and rehabilitating the ‘natural side’ of human existence.

Unlike the members of Stankevich’s circle, Herzen was never tempted by the philosophy of reconciliation, seeing it as a form of moral suicide. In 1840 he embarked on a systematic study of Hegelianism with the aim of refuting its right-wing interpretation. He was helped in this endeavour by two Left-Hegelian thinkers: Cieszkowski (whose *Prolegomena* he read in 1838) and Feuerbach.

Herzen’s philosophical essays of the 1840s, especially ‘Buddizm v nauke’ (‘Buddhism in Science’) (1843) and ‘Pis’ma ob izuchenii nauki’ (‘Letters on the Study of Nature’) (1845), develop the themes of the ‘philosophy of action’ and ‘the rehabilitation of the natural’, as two approaches to the problem of personal identity and freedom. In ‘Buddizm v nauke’ Herzen set his philosophy of action against the contemplative ‘panlogism’ of Hegel. He wrote: ‘Action is the personality itself’. This meant that in order to become a personality man must get rid of his ‘natural immediacy’, raise himself to the ‘Buddhist’ realm of impersonal, abstract thought, get lost in the general and then (the third stage) regain his identity and express himself in free, conscious activity, changing the world in the desired direction. In ‘Pis’ma ob izuchenii nauki‘ the problem of personality was shown against the background of the historical development of philosophical attitudes towards Nature. Their leitmotif was the demand for a reconciliation between naturalistic empiricism (identified with materialism) and idealism, with a strong insistence that it was to be realized not only for the sake of science but above all for the benefit of the human personality. Following Feuerbach, Herzen saw idealism as representing ’The General‘, at the expense of ignoring and annihilating the individual. ’Materialism‘, on the other hand, threatened to reduce human beings to their natural, immediate particularism. The future synthesis of both outlooks, Herzen concluded, should vindicate the rights of concrete individuals of flesh and blood but without losing the achievements of idealism, which had raised humans to the level of universalism.

An important theme in Herzen’s critique of Hegel’s philosophy of history was his protest against treating individuals and nations as mere instruments of the Reason of History. Nevertheless, before he went to the West (in 1847) he retained his faith in progress and Hegelian belief in the basic rationality of the historical process as a whole. This belief, however, did not survive the defeat of socialism in the revolutions of 1848. In his book *S togo*

berega (From the Other Shore) (1850) Herzen definitely broke with Hegelianism, asserting that history was but a blind play of chances and refusing to make human sacrifices on the altar of illusionary progress.

2 Belinskii's aesthetics and literary criticism

As a literary critic Belinskii owed a great deal to the German Romantics while supporting the realist tendency in Russian literature. Hegelianism enabled him to overcome Romantic 'subjectivism' and to embrace instead the ideal of 'objective art'. In the period of 'reconciliation' he meant by this that writers should reflect objective reality in all its richness, putting aside their 'subjectivity' and refraining from value-judgments. The perfect embodiments of this ideal were for him the great 'Olympians', Goethe and Pushkin. The rejection of 'reconciliation' entailed a re-examination of this view - a rehabilitation of the artists' right to 'subjectivism'. At this stage, however, 'subjectivism' in art was no longer identified with Romantic whims; Belinskii stressed that legitimate subjectivism does not distort reality but reproduces it faithfully and judges it from the point of view of the forward-moving trend within society. In this way he justified and propagated the ideological commitment of realistic literature.

The influence of Hegel's aesthetics on Belinskii's thought is most visible in Belinskii's article 'The Idea of Art' (1841). It is very characteristic, however, that Belinskii corrected Hegel by placing art, defined as 'the immediate contemplation of truth', above religion, defined as an 'immediate representation of truth'. This correction reflected Belinskii's view that religious mythology was good only for the lowest, infantile stage of the human spirit, and that the development of literature was preparing society for the attainment of maturity in conscious, self-reflecting thought. In this way Belinskii justified his ardent belief in the emancipatory mission of Russian literature.

Belinskii emphasized that artistic generalization should not be confused with a 'logical syllogism' or 'schematic abstraction'. This would be transgressing against the very nature of art, since art was 'thinking in images'. But it was difficult to reconcile the autonomy of art with the conviction that artists, particularly writers, should serve the cause of historical progress. Hence a certain tension in Belinskii's views. In his series of articles on Pushkin (1843-6) he stated that poetry has no purpose beyond itself; in his other articles, however, he used to encourage the progressively minded writers by repeating that there is no such thing as pure art, since every work of art is tendentious. Thus, the legacy of Belinskii's literary criticism was somewhat ambivalent and subject to different interpretations.

3 The Westernizing philosophies of Russian history

Hegel's philosophy of history, as well as his philosophy of law, stressed the necessity and progressive function of the increasing rationalization of social bonds, which was directly relevant for the Russian controversy between Slavophiles and Westernizers. Hegelianism was incompatible with idealization of the Russian past and, for this reason alone, served the cause of Westernism.

The most consistent and prolific Westernizer of the 'remarkable decade' was Belinskii. His contribution consisted in applying to Russia the Hegelian conception of development from 'natural immediacy' to 'rational reality' and the Hegelian distinction between 'nonhistorical peoples' and 'historical nations'.

Pre-Petrine Russia, in Belinskii's view, was a nation in its natural, immediate and patriarchal state; hence it was not a nation in the modern sense of this term (*natsiia*) but only a people (*narod*), and its culture could not have any universal human significance. Only the reforms of Peter the Great raised Russia to the level of nationhood. The price for this was the radical negation of the natural immediacy of ancient Russia and the resulting cleavage between the Westernized 'society' and non-Westernized folk (*narod*). Owing to these reforms the Russian state could become an active subject of universal human history. However, in order to become a fully fledged 'historical nation' the Russians had to develop modern national consciousness and create a national literature of a truly world-historical significance. The decisive step in this direction was made by Pushkin. Thus, the Westernization of Russia laid foundations for its truly national development.

Another Hegelianizing interpretation of Russian history was provided by the young Moscow historian Konstantin Kavelin, a disciple of T.Granovskii. In his study 'Vzgliad na iuridicheskii byt v drevnei Rossii' ('A Brief Survey of Juridical Relations in Ancient Russia') (1847) he developed the argument that the Russian historical process consisted in the gradual dissolution of patriarchal bonds and their replacement by the juridical order of the centralized state, which made room for individual freedom. From this perspective Peter the Great appeared as the

first completely emancipated individual in Russian history. The historical process in Russia led, therefore, to the emancipation of the individual through the rationalization of social relations by means of law. According to Kavelin, the emergence of the centralized Muscovite state played an important part in this process. This view - which sprang from Kavelin’s interpretation of the Muscovite autocracy in terms of the rationalistic Hegelian state - was later to be developed by the ’etatist school’ of Russian historiography. Its representatives, including the Hegelian philosopher Boris Chicherin, argued that in Russia the state had always been the chief organizer of society and the main agent of progress (see Liberalism, Russian §2).

4 Hegelianism and Slavophilism

The leading Slavophile thinkers, Ivan Kireevskii and Aleksei Khomiakov, were definitely opposed to Hegelianism. They saw Hegelian philosophy as the culmination of the anti-ontological tendency in Western rationalism, which transformed living reality into a dialectic of incorporeal notions. They passionately rejected Hegel’s view that the meaning of universal history was the necessary process of the rationalization of society; according to them this was true of the West but did not apply to Russia. On theoretical grounds, against Hegelian dialectical historicism they set a conservative-Romantic historicism, enshrining uninterrupted historical continuity and condemning sudden breaks with the past as a betrayal of ’historicity’.

Nevertheless, two younger Slavophile thinkers - Konstantin Aksakov and Iurii Samarin - tried for some time to combine Slavophilism with Hegelianism, thus giving rise to a curious phenomenon known as ’Orthodox-Christian Hegelianism’ (Herzen’s expression). The main documents of this trend are two books written as master’s dissertations for the University of Moscow: Aksakov’s *Lomonosov in the History of Russian Language and Literature* (published 1846) and Samarin’s *Stepan Iavorskii and Theophan Prokopovich* (defended 1844). The first of them was an ingenious interpretation of the history of Russian language and literature as a process raising the particular to the universal and then overcoming abstract universality. The second developed the view that the lack of rational scholastic theology was in fact the great merit of Orthodox Christianity because, as Hegel had shown, religious truths should not be rationally defined and presented in the form of a quasi-scientific system.

However, Slavophile Hegelianism did not stand the test of time. The two Slavophile Hegelians soon realized that in fact Hegelian philosophy supported the standpoint of the Westernizers and that attempts to make it consonant with Slavophilism were futile and doomed to failure (see Slavophilism).

5 Russian Hegelianism in the second half of the nineteenth century

The ’philosophical epoch’ in Russia ended abruptly in 1848. The revolutionary events of the European Springtime of the Peoples pushed the tsarist government in the direction of the most severe repression of intellectual life. Philosophy, seen as a source of dangerous disease, was banished from universities and the new censorship rules made it impossible even to mention in print the names of Hegel, Feuerbach, or Belinskii. This gloomy period of repression ended only after Russia’s defeat in the Crimean War and the unexpected death of Nicholas I. The first symptom of the forthcoming political ’thaw’ was the public defence of Chernyshevskii’s master’s dissertation ‘Ėsteticheskie otnosheniia iskusstva k deistvitelnosti’ (The Aesthetic Relations Between Art and Reality), in the University of St Petersburg in May 1855 (see Chernyshevskii, N.G.).

The content of Chernyshevskii’s thesis was directly linked with the philosophical discussions of the ’remarkable decade’. Without mentioning the prohibited names, Chernyshevskii discussed in it Hegelian idealism, Feuerbachian anthropological materialism and Belinskii’s views on art: he accused Hegel of establishing ’the tyranny of the universal’, praised Feuerbach for rehabilitating nature and developed Belinskii’s view in a consistently utilitarian direction, stressing especially the didactic function of literature. His utilitarian approach to art, further radicalized by Nikolai Dobroliubov and Dmitrii Pisarev, aroused indignation among the so-called Aesthetic Critics (V. Botkin, P. Annenkov and A. Druzhinin) and caused them to defend Hegelian aesthetics, philosophical idealism in general, and the legacy of the Russian intellectuals of the 1840s. Owing to this Hegelianism came to be associated with the liberal aestheticism, characteristic of the older generation and opposed by the young radicals, who chose to be seen as intransigent enemies of all idealism and aesthetics. This conflict between the two generation of Russian intellectuals is shown in Turgenev’s novel *Fathers and Sons*.

In contrast to the younger radicals, Chernyshevskii himself did not see Hegel’s philosophy as entirely obsolete and

worthless. He made a distinction between Hegel's system and Hegel's dialectical method, rejecting the first but trying to save the second through giving it a naturalistic interpretation. For this reason Soviet philosophers treated him as a precursor of dialectical materialism (see Russian Materialism: 'the 1860s' §3).

Philosophical naturalism and positivistic scientism, which dominated Russian intellectual life after the 1860s, were, of course, deeply inimical to Hegelian idealism. Nevertheless, the Hegelian tradition survived in Russia, although in a marginalized and defensive form. Thus, for instance, Pavel Bakunin, brother of Mikhail Bakunin, elaborated a religious, Romantic interpretation of Hegelianism and presented it in his two books: *Zapozdalyi golos sorokovykh godov (A Belated Voice of the Forties)* (1881) and *Osnovy very i znzniia (Foundations of Faith and Knowledge)* (1886). Another religious reinterpretation of Hegelianism was offered by Nikolai Debolskii, professor at the St Petersburg Theological Academy, whose *Filosofiia fenomenal'nogo formalizma (Philosophy of Phenomenalistic Formalism)* (2 vols, 1892-5) tried to reconcile Hegelian rationalism with traditional Christian theism.

The greatest representative of the Hegelian tradition in the second half of the century was Chicherin, who also tried to reconcile philosophy and religion by identifying universal Reason with a personal God. In his political philosophy he combined Hegelian emphasis on the need of a strong state with conservative liberalism, making use of the Hegelian concept of civil society as a sphere of civil liberties (as distinct from political freedom) and free play of particularistic economic interests. Evgenii Trubetskoi, author of the classical monograph on Vladimir Solov'ëv saw Chicherin as the living bridge between the 'philosophical epoch' of 1838-48 and the Religious-Philosophical Renaissance of the twentieth century.

The legacy of the Russian Hegelians of the 1840s, especially the legacy of Belinskii, was claimed also by a thinker representing the opposite pole of the ideological spectrum: Georgii Plekhanov, 'father of Russian Marxism'. He was greatly impressed by Hegelian dialectical historicism, particularly the concept of rational historical necessity, seeing in it an effective antidote to populist 'subjective sociology' and all sorts of social utopianism. Due to his efforts interest in Hegelian dialectics, in a materialistic interpretation, became a characteristic feature of Russian and Soviet Marxism.

See also: Hegelianism; Il'in, I.A.

ANDRZEJ WALICKI

References and further reading

Bakunin, M.A. (1838) 'Gimnazicheskie rechi Gegelia. Predislovie perevodchika' ('Foreword to Hegel's School Addresses'), in Iu.M. Steklov (ed.), *Sobranie sochinenii i pisem 1828-76*, Moscow: AN SSSR, 1934-5, vol. 2, 166-78.(Referred to in §1.)

Belinskii, V.G. (1841) 'The Idea of Art', in J.M. Edie *et al.* (eds), *Russian Philosophy*, Chicago, IL: Quadrangle Books, 1969, vol. 1, 285-95.(Referred to in §2.)

Berlin, I. (1979) 'A Remarkable Decade', in H. Hardy and A. Kelly (eds), *Russian Thinkers*, New York: Viking Press.(An excellent general account of the 'remarkable decade' in Russian intellectual history.)

Billig, J. (1930) *Der Zusammenbruch des deutschen Idealismus bei den russischen Romantikern. Belinskii, Bakunin (The Collapse of German Idealism among the Russian Romantics. Belinskii, Bakunin)*, Berlin.(A valuable study of Belinskii and young Bakunin, in the context of the crisis of absolute idealism in Germany.)

Chernyshevskii, N.G. (1855) 'Ėsteticheskie otnosheniia iskusstva k deistvitelnosti' ('The Aesthetic Relations Between Art and Reality'), in J.M. Edie *et al*. (eds), *Russian Philosophy*, Chicago, IL: Quadrangle Books, 1969, vol. 2, 16-28. (An abridged translation of Chernyshevskii's master's thesis)

Chizhevskii, D.I. (1939) *Gegel' v Rossii (Hegel in Russia)*, Paris.(The first comprehensive history of Hegelianism in Russia.)

Cieszkowski, A. von (1838) *Die Prolegomena zur Historiosophie (Prolegomena to the Wisdom of History)*, Berlin: Viet.(A dominant influence on Herzen's early Hegelianism.)

Herzen, A.I. (1843) 'Buddizm v nauke'; trans. 'Buddhism in Science', in J.M. Edie *et al.* (eds) *Russian Philosophy*, Chicago, IL: Quadrangle Books, 1969, vol. 2, 16-28.

Herzen, A.I. (1850) *S togo berega*, Paris; trans. M. Budberg, *From the Other Shore*, London: Weidenfeld & Nicolson, 1956.(Herzen's break with the Hegelian belief in the rationality of the historical process.)

Herzen, A.I. (1853) *Byloe i dumy*, London; trans. C. Garnett, *My Past and Thoughts. The Memoirs of Alexander Herzen*, revised by H. Higgens, introduced by I. Berlin, London: Chatto & Windus, 1968, 4 vols.(See ch. 24,

vol. 2 for the reception of Hegel in Russia.)

Jakowenko, B. (1938) *Geschichte des Hegelianismus in Rußland (History of Hegelianism in Russia)*, vol. 1, Prague.(Covers the period 1835-60.)

Koyré, A. (1950) *Études sur l'histoire de la pensée philosophique en Russie (Studies in the History of Philosophical Thought in Russia)*, Paris, 103-70.(A good presentation of Russian Hegelianism of the 'remarkable decade'.)

Planty-Bonjour, G. (1974) *Hegel et la pensée philosophique en Russie 1830-1917. (Hegel and Philosophical Thought in Russia 1820-1917)*, La Haye: M. Nijhoff.(The most recent comprehensive history of the reception of Hegelian ideas in Russia.)

Volodin, A.I. (1973) *Gegel' i russkaia sotsialisticheskaia mysl' xix veka (Hegel and Russian Socialist Thought of the Nineteenth Century)*, Moscow: Mysl'.(Includes bibliography.)

Walicki, A. (1975) *The Slavophile Controversy. History of a Conservative Utopia in Nineteenth-Century Russian Thought*, Oxford: Clarendon Press.(Contains a detailed analysis of the ideas of the 'Slavophile Hegelians' (287-335) and also two chapters dealing with the role of Hegelianism in the controversy between Slavophiles and Westernizers (335-455).)

Walicki, A. (1991) 'Alexander Herzen, August Cieszkowski, and the Philosophy of Action', in *Russia, Poland, and Universal Regeneration. Studies on Russian and Polish Thought of the Romantic Epoch*, Notre Dame, IN: University of Notre Dame Press, 73-106.(An analysis of Cieszkowski's influence on Herzen, as well as a comparative study of both thinkers in a broader context.)

Heidegger, Martin (1889-1976)

Martin Heidegger taught philosophy at Freiburg University (1915-23), Marburg University (1923-8), and again at Freiburg University (1928-45). Early in his career he came under the influence of Edmund Husserl, but he soon broke away to fashion his own philosophy. His most famous work, Sein und Zeit (Being and Time)was published in 1927. Heidegger's energetic support for Hitler in 1933-4 earned him a suspension from teaching from 1945 to 1950. In retirement he published numerous works, including the first volumes of his Collected Edition. His thought has had strong influence on trends in philosophy ranging from existentialism through hermeneutics to deconstruction, as well as on the fields of literary theory and theology.

Heidegger often makes his case in charged and dramatic language that is difficult to convey in summary form. He argues that mortality is our defining moment, that we are thrown into limited worlds of sense shaped by our being-towards-death, and that finite meaning is all the reality we get. He claims that most of us have forgotten the radical finitude of ourselves and the world we live in. The result is the planetary desert called nihilism, with its promise that an ideally omniscient and virtually omnipotent humanity can remake the world in its own image and likeness. None the less, he still holds out the hope of recovering our true human nature, but only at the price of accepting a nothingness darker than the nihilism that now ravishes the globe. To the barely whispered admission, 'I hardly know anymore who and where I am', Heidegger answers: 'None of us knows that, as soon as we stop fooling ourselves' ([1959a] 1966: 62).

Yet he claims to be no pessimist. He merely wants to find out what being as such means, and Being and Time was an attempt at this. He called it a fundamental ontology: a systematic investigation of human being (Dasein) for the purpose of establishing the meaning of being in general. Only half of the book - the part dealing with the finitude and temporality of human being - was published in 1927. Heidegger elaborated the rest of the project in a less systematic form during the decades that followed.

Heidegger distinguishes between an entity (anything that is) and the being of an entity. He calls this distinction the 'ontological difference'. The being of an entity is the meaningful presence of that entity within the range of human experience. Being has to do with the 'is': what an entity is, how it is, and the fact that it is at all. The human entity is distinguished by its awareness of the being of entities, including the being of itself. Heidegger names the human entity 'Dasein' and argues that Dasein's own being is intrinsically temporal, not in the usual chronological sense but in a unique existential sense: Dasein ek-sists (stands-out) towards its future. This ek-sistential temporality refers to the fact that Dasein is always and necessarily becoming itself and ultimately becoming its own death. When used of Dasein, the word 'temporality' indicates not chronological succession but Dasein's finite and mortal becoming.

If Dasein's being is thoroughly temporal, then all of human awareness is conditioned by this temporality, including one's understanding of being. For Dasein, being is always known temporally and indeed is *temporal. The meaning of being is time. The two main theses of Being and Time - that Dasein is temporal and that the meaning of being is time - may be interpreted thus: being is disclosed only finitely within Dasein's radically finite awareness.*

Heidegger arrives at these conclusions through a phenomenological analysis of Dasein as being-in-the-world, that is, as disclosive of being within contexts of significance. He argues that Dasein opens up the arena of significance by anticipating its own death. But this event of disclosure, he says, remains concealed even as it opens the horizon of meaning and lets entities be understood in their being. Disclosure is always finite: we understand entities in their being not fully and immediately but only partially and discursively; we know things not in their eternal essence but only in the meaning they have in a given situation. Finite disclosure - how it comes about, the structure it has, and what it makes possible - is the central topic of Heidegger's thought. 'Time is the meaning of being' was only a provisional way of expressing it.

Dasein tends to overlook the concealed dimension of disclosure and to focus instead on what gets revealed: entities in their being. This overlooking is what Heidegger calls the forgetfulness of the disclosure of being. By that he means the forgetting of the ineluctable hiddenness of the process whereby the being of entities is disclosed. He argues that this forgetfulness characterizes not only everyday 'fallen' human existence but also the entire

history of being, that is, metaphysics from Plato to Nietzsche. He calls for Dasein resolutely to reappropriate its own radical finitude and the finitude of disclosure, and thus to become authentically itself.

1 Life and works

Martin Heidegger was born on 26 September 1889 in Messkirch, Southwest Germany, to Roman Catholic parents of very modest means. From 1899 to 1911 he intended to become a priest, but after two years of theological studies at Freiburg University a recurring heart condition ended those hopes. In 1911 he switched to mathematics and the natural sciences, but finally took his doctorate in philosophy (1913) with a dissertation entitled *Die Lehre vom Urteil im Psychologismus (The Doctrine of Judgment in Psychologism)* (1914). Hoping to get appointed to Freiburg's chair in Catholic philosophy, he wrote a qualifying dissertation in 1915 on a theme in medieval philosophy, *Die Kategorien- und Bedeutungslehre des Duns Scotus (Duns Scotus' Doctrine of Categories and Meaning)* (1916). However, the job went to someone else, and in the autumn of 1915 Heidegger began his teaching career at Freiburg as a lecturer.

At this time Heidegger was known as a Thomist, but his 1915 dissertation was strongly influenced by the founder of phenomenology, Edmund Husserl. When Husserl joined the Freiburg faculty in the spring of 1916, Heidegger came to know him personally, if not well. Their relation would blossom only after the First World War. Heidegger was drafted in 1918 and served as a weatherman on the Ardennes front in the last three months of the war. When he returned to Freiburg his philosophical career took a decisive turn. In a matter of weeks he announced his break with Catholic philosophy (9 January 1919), got himself appointed Husserl's assistant (21 January), and began lecturing on a radical new approach to philosophy (4 February).

Many influences came to bear on Heidegger's early development, including St Paul, Augustine, Meister Eckhart, Kierkegaard, Dilthey and Nietzsche. But the major influences were Husserl and Aristotle. Heidegger was Husserl's protégé in the 1920s, but he never was a faithful disciple. He preferred Husserl's early work, *Logische Untersuchungen (Logical Investigations)* (1900-1), to the exclusion of the master's later developments. Moreover, the things that Heidegger liked about *Logical Investigations* were generally consonant with the traditional scholastic philosophy he had been taught.

First, Husserl's early phenomenology considered the human 'psyche' not as a substantial thing but as an act of revealing (intentionality), one that revealed not only *what* is encountered (the entity) but also the *way in which* it is encountered (the entity's being). Second, the early Husserl held that the central issue of philosophy was not modern subjectivity but rather 'the things themselves', whatever they might happen to be, in their very appearance; and he provided a descriptive method for letting those things show themselves as they are. Third, phenomenology argued that the being of entities is known not by some after-the-fact reflection or transcendental construction but directly and immediately by way of a categorial intuition. In short, for Heidegger, phenomenology was a descriptive method for understanding the being of entities as it is disclosed in intentional acts (see Phenomenological movement).

As Heidegger took it, all this contrasted with Husserl's later commitment to pure consciousness as the presuppositionless 'thing itself' that was to be revealed by various methodological 'reductions'. Heidegger had no use either for the Neo-Kantian turn to transcendental consciousness that found expression in Husserl's *Ideen (Ideas)* (1913) or for his further turn to a form of Cartesianism. Against Husserl's later theory of an unworldly transcendental ego presuppositionlessly conferring meaning on its objects, Heidegger proposed the historical and temporal situatedness of the existential self, 'thrown' into the world, 'fallen' in among entities in their everyday meanings, and 'projecting' ahead towards death.

In the 1920s Heidegger began interpreting the treatises of Aristotle as an implicit phenomenology of everyday life without the obscuring intervention of subjectivity. He took Aristotle's main topic to be 'disclosure' (*alētheia*) on three levels: entities as intrinsically self-disclosive; human *psyche* as co-disclosive of those entities; and especially the human disclosure of entities in discursive, synthetic activity (*logos*), whether that be performed in wordless actions or in articulated sentences. Going beyond Aristotle, Heidegger interpreted this discursive disclosure as grounded in a kind of movement that he named 'temporality', and he argued that this temporality was the very essence of human being.

Using this new understanding of human being, Heidegger reinterpreted how anything at all appears to human

beings. He argued that humans, as intrinsically temporal, have only a temporal understanding of whatever entities they know. But humans understand an entity by knowing it in its being, that is, in terms of how it happens to be present. Therefore, as far as human being goes, all forms of being are known temporally and indeed *are* temporal. The meaning of being is time.

Heidegger developed this thesis gradually, achieving a provisional formulation in *Sein und Zeit (Being and Time)* (1927). In public he dedicated the book to Husserl 'in respect and friendship', but in private he was calling Husserl's philosophy a 'sham' (*Scheinphilosophie*). Meanwhile, in 1923 an unsuspecting Husserl helped Heidegger move from a lecturer's job at Freiburg to a professorship at Marburg University; and when Husserl retired in 1928, he arranged for Heidegger to succeed him in the chair of philosophy at Freiburg. Once Heidegger had settled into the new job, the relationship between mentor and protégé quickly fell apart. If *Being and Time* were not enough, the three works Heidegger published in 1929 - 'Vom Wesen des Grundes' ('On the Essence of Ground'), *Kant und das Problem der Metaphysik (Kant and the Problem of Metaphysics)*, and *Was ist Metaphysik? (What is Metaphysics?)* - confirmed how far apart the two philosophers had grown.

Heidegger's career entered a new phase when the Nazis came to power in Germany. On 30 January 1933 Adolf Hitler was appointed Chancellor, and within a month the German constitution and all-important civil rights were suspended. On 23 March Hitler became dictator of Germany, with absolute power to enact laws, and two weeks later, harsh anti-Semitic measures were promulgated. A conservative nationalist and staunch anti-Communist, Heidegger supported Hitler's policies with great enthusiasm for at least one year, and with quieter conviction for some ten years thereafter. He was elected rector (president) of Freiburg University on 21 April 1933 and joined the Nazi Party on May 1, with the motive, he later claimed, of preventing the politicization of the university. In his inaugural address as rector, *Die Selbstbehauptung der deutschen Universität (The Self-Assertion of the German University)* (27 May 1933), he called for a reorganization of the university along the lines of some aspects of the Nazi revolution. As rector he proved a willing spokesman for, and tool of, Nazi policy both foreign and domestic.

Heidegger resigned the rectorate on 23 April 1934 but continued to support Hitler. His remarks in the classroom indicate that he backed the German war aims, as he knew them, until at least as late as the defeat at Stalingrad in January 1943. The relation, or lack of it, between Heidegger's philosophy and his political sympathies has long been the subject of heated debate.

Heidegger published relatively little during the Nazi period. Instead, he spent those years rethinking his philosophy and setting out the parameters it would have, both in form and focus, for the rest of his life. The revision of his thought is most apparent in three texts he published much later: (1) the working notes from 1936-8 that he gathered into *Beiträge zur Philosophie. Vom Ereignis (Contributions to Philosophy: On Ereignis)*, published posthumously in 1989; (2) the two volumes of his *Nietzsche*, published in 1961, which contains lecture courses and notes dating from 1936 to 1946; and (3) 'Brief über den Humanismus' ('Letter on Humanism'), written in the autumn of 1946 and published in 1947.

After the war Heidegger was suspended from teaching because of his Nazi activities in the 1930s. In 1950, however, he was allowed to resume teaching, and thereafter he occasionally lectured at Freiburg University and elsewhere. Between 1950 and his death he published numerous works, including the first volumes of his massive *Gesamtausgabe (Collected Edition)*. He died at his home in Zähringen, Freiburg, on 26 May 1976 and was buried in his home town of Messkirch. His literary remains are held at the German Literary Archives, at Marbach on the Neckar.

Heidegger, a Catholic, married Elfride Petri (1893-1992), a Lutheran, on 21 March 1917. They had two sons, both of whom served in the *Wehrmacht* during the Second World War and were taken prisoner on the Eastern Front. In February of 1925 Heidegger began a year-long affair with his then student, Hannah Arendt. In February of 1950 they resumed a strong but often stormy friendship that lasted until Arendt's death.

2 Temporality and authenticity

Heidegger was convinced that Western philosophy had misunderstood the nature of being in general and the nature of human being in particular. His life's work was dedicated to getting it right on both scores.

In his view, the two issues are inextricably linked. To be human is to disclose and understand the being of

whatever there is. Correspondingly, the being of an entity is the meaningful presence of that entity within the field of human experience. A proper or improper understanding of human being entails a proper or improper understanding of the being of everything else. In this context 'human being' means what Heidegger designates by his technical term 'Dasein': not consciousness or subjectivity or rationality, but that distinctive kind of entity (which we ourselves always are) whose being consists in disclosing the being both of itself and of other entities. The being of this entity is called 'existence' (see §4).

Heidegger argues that the structure of human being is comprised of three co-equal moments: becoming, alreadiness and presence. (These are usually, and unfortunately, translated as: 'coming towards itself', 'is as having been' and 'making-present'.) As a unity, these three moments constitute the essence of human being, which Heidegger calls 'temporality': opening an arena of meaningful presence by anticipating one's own death. Temporality means being present by becoming what one already is.

Becoming. To be human means that one is not a static entity just 'there' among other things. Rather, being human is always a process of becoming oneself, living into possibilities, into one's future. For Heidegger, such becoming is not optional but necessary. He expresses this claim in various co-equal formulas: (1) The essence of human being is 'existence' understood as 'ek-sistence', an ineluctable 'standing out' into concern about one's own being and into the need to become oneself; (2) the essence of human being is 'factical', always already thrust into concernful openness to itself and thus into the ineluctability of self-becoming; and (3) the essence of being human is 'to be possible' - not just able, but above all needing, to become oneself.

The ultimate possibility into which one lives is the possibility to end all possibilities: one's death. Human beings are essentially finite and necessarily mortal, and so one's becoming is an anticipation of death. Thus, to know oneself as becoming is to know oneself, at least implicitly, as mortal. Heidegger calls this mortal becoming 'being-unto-death'.

Alreadiness. Human being consists in becoming; and this becoming means becoming what one already is. Here the word 'already' means 'essentially', 'necessarily' or 'inevitably'. 'Alreadiness' (*Gewesenheit*) names one's inevitable human essence and specifically one's mortality. In becoming the finitude and mortality that one already is, one gets whatever presence one has.

Presence. Mortal becoming is the way human being (a) is meaningfully present to itself and (b) renders other entities meaningfully present to itself. To put the two together: things are present to human being in so far as human being is present to itself as mortal becoming. In both cases presence is bound up with absence.

How human being is present to itself. Since mortal becoming means becoming one's own death, human being appears as disappearing; it is present to itself as becoming absent. To capture this interplay of presence and absence, we call the essence of human being 'pres-abs-ence', that is, an incomplete presence that shades off into absence. Pres-abs-ence is a name for what classical philosophy called 'movement' in the broad sense: the momentary presence that something has on the basis of its stretch towards the absent.

Pres-abs-ence is an index of finitude. Any entity that appears as disappearing, or that has its current presence by anticipating a future state, has its being not as full self-presence but as finite pres-abs-ence. The movement towards death that defines human being is what Heidegger calls 'temporality'. The quotation marks indicate that 'temporality' does not refer to chronological succession but rather means having one's being as the movement of finite mortal becoming.

How other things are present to human being. Other entities are meaningfully present to human being in so far as human being is temporal, that is, always anticipating its own absence. Hence the meaningful presence of things is also temporal or pres-abs-ent - always partial, incomplete and entailing an absence of its own. Not only is human being temporal but the presence of things to human being is also temporal in its own right.

All of Heidegger's work argues for an intrinsic link between the temporality or pres-abs-ence that defines human being and the temporality or pres-abs-ence that characterizes the meaningful presence of things. But the meaningful presence of things is what Heidegger means by being. Therefore, Heidegger's central thesis is this: as far as human experience goes, all modes of being are temporal. The meaningful presence of things is always imperfect, incomplete, pres-abs-ential. The meaning of being is time.

Heidegger argues that this crucial state of affairs - finite human being as an awareness of the finitude of all modes of being - is overlooked and forgotten both in everyday experience and in philosophy itself. Therefore, his work discusses how one can recover this forgotten state of affairs on both of those levels.

As regards everyday life, Heidegger describes how one might recall this central but forgotten fact and make it one's own again. The act of reappropriating one's own essence - of achieving a personal and concrete grasp of oneself as finite - is called 'resolution' (in other translations, 'resoluteness' or 'resolve'). This personal conversion entails becoming clear about the intrinsic finitude of one's own being, and then choosing to accept and to be that finitude.

Awareness of one's finitude. Human being is always already the process of mortal becoming. However, one is usually so absorbed in the things one encounters ('fallenness') that one forgets the becoming that makes such encounters possible. It takes a peculiar kind of experience, more of a mood than a detached cognition, to wake one up to one's finitude. Heidegger argues that such an awakening comes about in special 'basic moods' (dread, boredom, wonder and so on) in which one experiences not things but that which is not-a-thing or 'no-thing'. Each of these basic moods reveals, in its own particular way, the absential dimension of one's pres-abs-ence.

Heidegger often uses charged metaphors to discuss this experience. For example, he describes dread as a 'call of conscience', where 'conscience' means not a moral faculty but the heretofore dormant, and now awakening, awareness of one's finite nature. What this call of conscience reveals is that one is 'guilty', not of some moral fault but of an ontological defect: the fact of being intrinsically incomplete and on the way to absence. The call of conscience is a call to understand and accept this 'guilt'.

Choosing one's finitude. One may choose either to heed or to ignore this call of conscience. To heed and accept it means to acknowledge oneself as a mortal process of pres-abs-ence and to live accordingly. In that case, one recuperates one's essence and thus attains 'authenticity' by becoming one's proper (or 'authentic') self. To ignore or refuse the call does not mean to cease being finite and mortal but rather to live according to an improper (inauthentic or 'fallen') self-understanding. Only the proper or authentic understanding of oneself as finite admits one to the concrete, experiential understanding that all forms of being, all ways that things can be meaningfully present, are themselves finite.

Summary. The essence of human being is temporality, that is, mortal becoming or pres-abs-ence. To overlook mortal becoming is to live an inauthentic temporality and to be a fallen self. But to acknowledge and choose one's mortal becoming in the act of resolution is to live an authentic temporality and selfhood. It means achieving presence (both the presence of oneself and that of other entities) by truly becoming what one already is. This recuperation of one's own finite being can lead to the understanding that what conditions *all* modes of being is finitude: the very meaning of being is time.

3 Being-in-the-world and hermeneutics

In *Being and Time* Heidegger spells out not only the reasons why, but also the ways in which, things are meaningfully present to human being.

Being-in-the-world. In contrast to theories of human being as a self-contained theoretical ego, Heidegger understands human being as always 'outside' any supposed immanence, absorbed in social intercourse, practical tasks and its own interests. Evidence for this absorption, he argues, is that human being always finds itself caught up in a mood - that is, 'tuned in' to a given set of concerns. The field of such concerns and interests Heidegger calls the 'world'; and the engagement with those needs and purposes and the things that might fulfil them he calls 'being-in-the-world' (or equally 'care').

Heidegger's term 'world' does not mean planet earth, or the vast expanse of space and time, or the sum total of things in existence. Rather, 'world' means a dynamic set of relations, ultimately ordered to human possibilities, which lends meaning or significance to the things that one deals with - as in the phrase 'the world of the artist' or 'the world of the carpenter'. A human being lives in many such worlds, and they often overlap, but what constitutes their essence - what Heidegger calls the worldhood of all such worlds - is the significance that accrues to things by their relatedness to human interests and possibilities. Although being-in and world can be distinguished, they never occur separately. Any set of meaning-giving relations (world) comes about and remains

effective only in so far as human being is engaged with the apposite possibilities (being-in). Being-in holds open and sustains the world.

In *Being and Time* Heidegger studies the world that he considers closest to human beings: the world of everyday activity. The defining moment of such a world is practical purposes ordered to human concerns - for example, the need to build a house for the sake of shelter. A group of things then gets its significance from the direct or indirect relation of those things to that goal. For example, these specific tools get their significance from their usefulness for clearing the ground, those trees get their significance from being suitable for lumber, these plants from their serviceability as thatch. A dynamic set of such relations (such as ‘useful to’, ‘suitable as’, ‘needed for’), all of which refer things to a human task and ultimately to a human possibility, constitutes a ‘world’ and defines the current significance that certain things (for example, tools, trees and reeds) might have.

The significance of things changes according to the interplay of human interests, the relations that they generate, and the availability of material. For example, given the lack of a mallet, the significance of a stone might be its utility for pounding in a tent peg. The stone gets its current significance as a utensil from the world of the camper: the desire for shelter, the need of something to hammer with, and the availability of only a stone. (When the camper finds a mallet, the stone may well lose its former significance.)

Hermeneutical understanding. Heidegger argues that the world of practical experience is the original locus of the understanding of the being of entities. Understanding entails awareness of certain relations: for example, the awareness of this *as* that, or of this *as for* that. The ‘as’ articulates the significance of the thing. In using an implement, one has a practical understanding of the implement’s relation to a task (X as useful for Y). This in turn evidences a practical understanding of the being of the implement: one knows the stone as *being* useful for pounding in a tent peg. In other words, prior to predicative knowledge, which is expressed in sentences of the type ‘S is P’, human beings already have a pre-theoretical or ‘pre-ontological’ understanding of the being of things (this as *being for* that).

Since the ‘as’ articulates how something is understood, and since the Greek verb *hermeneuein* means ‘to make something understandable’, Heidegger calls the ‘as’ that renders things intelligible in practical understanding the ‘hermeneutical as’. This ‘hermeneutical as’ is made possible because human being is a ‘thrown project’, necessarily thrust into possibilities (thrownness) and thereby holding the world open (project).

Hermeneutical understanding - that is, pre-predicatively understanding the ‘hermeneutical as’ by being a thrown project - is the kind of cognition that most befits being-in-the-world. It is the primary way in which humans know the being of things. By contrast, the more detached and objective ‘apophantic’ knowledge that expresses itself in declarative sentences (‘S is P’) is evidence, for Heidegger, of a derivative and flattened-out understanding of being.

Summary. As long as one lives, one is engaged in mortal becoming. This becoming entails having purposes and possibilities. Living into purposes and possibilities is how one has things meaningfully present. The ability to have things meaningfully present by living into possibilities is called being-in-the-world. Being-in-the-world is structured as a thrown project: holding open the possibility of significance (project) by ineluctably living into possibilities (thrownness). This issues in a pre-predicative, hermeneutical understanding of the being of things. Thus mortal becoming *qua* being-in-the-world engenders and sustains all possible significance. In another formulation: temporality determines all the ways that things can have meaningful presence. Time is the meaning of all forms of being.

4 Dasein and disclosure

Heidegger calls human being ‘Dasein’, the entity whose being consists in disclosing and understanding being, whether the being of itself or that of other entities. In so far as Dasein’s being is a disclosure of its *own* being, it is called ‘existence’ or ‘ek-sistence’: self-referential standing-out-unto-itself. Dasein’s very being consists in being related, with understanding and concern, to itself.

But Dasein is not just related to itself. Existence occurs only as being-in-the-world; that is, the openness of human being to itself entails the openness of the world for other entities. One of Heidegger’s neologisms for ‘openness’ is ‘the there’ (*das Da*), which he uses in two interrelated senses. First, human being is its own ‘there’: as a thrown

project, existence sustains its own openness to itself. And second, in so doing, human being also makes possible the world's openness as the 'there' for other entities. Human being's self-disclosure makes possible the disclosure of other entities.

Heidegger calls human being in both these capacities 'being-the-there' - Dasein, or sometimes Da-sein when it refers to the second capacity. In ordinary German *Dasein* means existence in the usual sense: being there in space and time as contrasted with not being at all. However, in Heidegger's usage Dasein means being disclosive of something (whether that be oneself or another entity) in its being. In a word, Dasein is disclosive. And since human being is radically finite, disclosure is radically finite.

The Greek word for disclosure is *alēthēia*, a term composed of the privative prefix *a-* (un- or dis-) and the root *lēthē* (hiddenness or closure). Heidegger finds the finitude of dis-closure inscribed in the word *a-lēthēia*. To disclose something is to momentarily rescue it from (*a-*) some prior unavailability (*lēthē*), and to hold it for a while in presence.

Heidegger discusses three levels of disclosure, ranging from the original to the derivative, each of which involves Dasein: (1) disclosure-as-such, (2) the disclosedness of entities in their being, and (3) disclosure in propositional statements. Heidegger's chief interest is in the first. There, disclosure/*alēthēia* is the original occurrence that issues in meaningful presence (being).

Heidegger argues that levels 1 and 2 are distinct but inseparable and, taken together, make possible level 3. The word 'truth' properly applies only at the third level, where it is a property of statements that correctly represent complex states of affairs. Therefore, to the question 'What is the essence of truth?' - that is, 'What makes the truth of propositions possible at level 3?' - Heidegger answers: Proximally, the disclosure of entities in their being (level 2); and ultimately, disclosure-as-such (level 1). His argument unfolds as follows.

Level 1. Disclosure-as-such is the very opening-up of the field of significance. It is the engendering and sustaining of world on the basis of Dasein's becoming-absent. In so far as it marks the birth of significance and the genesis of being, disclosure-as-such or world-disclosure is the reason why any specific entity can have meaningful presence at all.

There are three corollaries. First, the disclosure of world never happens except in Dasein's being; indeed, without Dasein, there is no openness at all. The engendering and sustaining of the dynamic relations that constitute the very possibility of significance occurs only as long as Dasein exists as mortal becoming. And conversely, wherever there is Dasein, there is world. Second, disclosure-as-such never happens apart from the disclosedness of *entities* as being this or that. In speaking of disclosure 'as such', Heidegger is naming the originating source and general structure of all possible significance that might accrue to any entity at all. The result of disclosure-as-such is the fact that referral-to-mortal-Dasein (that is, significance) is the basic state of whatever entities happen to show up. Third, disclosure-as-such is always prior to and makes possible concrete human action in any specific world. Such concrete actions run the risk of *not* being disclosive (that is, being mistaken about the meaning of something). By contrast, world-disclosure is *always* disclosive in so far as it is the opening-up of the very possibility of significance at all.

Alēthēia/disclosure-as-such - how it comes about, the structure it has, and what it makes possible - is the central topic or 'thing itself' of Heidegger's thought. He sometimes calls it the 'clearing' of being. He also calls it 'being itself' or 'being-as-such' (that is, the very *engendering* of being). Frequently, and inadequately, he calls it the 'truth' of being.

Level 2. What disclosure-as-such makes possible is the pre-predicative availability of entities in their current mode of being. This pre-predicative availability constitutes level 2, the basic, everyday disclosedness of entities as meaningfully present. This disclosedness is always finite, and that entails two things.

First, what disclosure-as-such makes possible is not simply the being of an entity but rather the being of that entity as or as not something: for instance, this stone as not a missile but as a hammer. I know the stone only in terms of one or another of its possibilities: the entity becomes present not fully and immediately but only partially and discursively. Thus the entity's being is always finite, always a matter of synthesis-and-differentiation: being-as-and-as-not. Second, disclosure-as-such lets an entity be present not in its eternal essence but only in its

current meaning in a given situation; moreover, it shows that this specific entity is not the only one that might have this meaning. For example, in the present situation I understand this stone not as a paperweight or a weapon but as a hammer. I also understand it as not the best instrument for the job: a mallet would do better.

Even though it is a matter of synthesis-and-differentiation, this pre-predicative hermeneutical understanding of being requires no thematic articulation, either mental or verbal, and no theoretical knowledge. It usually evidences itself in the mere doing of something. Nevertheless, in a more developed but still pre-predicative moment, such a hermeneutical awareness might evolve into a vague sense of the entity's being-this-or-that ('whatness'), being-in-this-way-or-that ('howness'), and being-available-at-all ('thatness'). Still later, these vague notions might lose the sense of *current* meaningfulness and develop, at level 3, into the explicit metaphysical concepts of the essence, modality and existence of the entity.

The second level of disclosure may be expressed in the following thesis: within any given world, to be an entity is to be always already disclosed as something or other. This corresponds to the traditional doctrine of metaphysics concerning a trans-generic (transcendental) characteristic of anything that is: regardless of its kind or species, every entity is intrinsically disclosed in its being (*omne ens est verum*).

Heidegger argues that while it is based on and is even aware of this second level of disclosure, metaphysics has no explicit understanding of disclosure-as-such or of its source in being-in-the-world. What is more, he claims that the disclosedness of entities-in-their-being (level 2) tends to overlook and obscure the very disclosure-as-such (level 1) that originally makes it possible. He further argues that there is an intrinsic hiddenness about disclosure-as-such, which makes overlooking it virtually inevitable (see §6).

Level 3. Being-in-the-world and the resultant pre-predicative disclosedness of entities as being-thus-and-so make it possible for us to enact the predicative disclosure of entities. At this third level of disclosure we are able to represent correctly to ourselves, in synthetic judgments and declarative sentences, the way things are in the world. A correct synthetic representation of a complex state of affairs (a correct judgment) is 'true', that is, disclosive of things just as they present themselves. Such a predicative, apophantic sentence ('S is P') is able to be true only because world-disclosure has already presented an entity as significant at all and thus allowed it to be taken as thus and so. This already disclosed entity is the binding norm against which the assertion must measure itself.

At level 3, however, it is also possible to *mis*represent things in thought and language, to fail to disclose them just as they present themselves in the world. At level 1 Dasein is always and only disclosive. But with predicative disclosure at level 3 (as analogously with hermeneutical disclosure at level 2) Dasein's representing of matters in propositional statements may be either disclosive or non-disclosive, either true or false.

One of Heidegger's reasons for elaborating the levels of disclosure is to demonstrate that science, metaphysics and reason in general, all of which operate at level 3, are grounded in a more original occurrence of disclosure of which they are structurally unaware. This is what he intends by his claim 'Science does not think'. He does not mean scientists are stupid or their work uninformed, nor is he disparaging reason and its accomplishments. He means that science, by its very nature, is not focused on being-in-the-world, even though being-in-the-world is ultimately responsible for the meaningful presence of the entities against which science measures its propositions.

5 Hiddenness, *Ereignis* and the Turn

Hiddenness. Heidegger claims that disclosure-as-such - the very opening up of significance in Dasein's being - is intrinsically hidden and needs to remain so if entities are to be properly disclosed in their being. This intrinsic concealment of disclosure-as-such is called the 'mystery'. Since Heidegger sometimes calls disclosure-as-such 'being itself', the phrase becomes 'the mystery of being'. The ensuing claim, that the mystery of being conceals itself while revealing entities, has led to much mystification, not least among Heideggerians. Being seems to become a higher but hidden Entity that performs strange acts that only the initiated can comprehend. This misconstrual of Heidegger's intentions is not helpful.

How may we understand the intrinsic concealment of disclosure-as-such? One way is to understand the paradigm of 'movement' that informs Heidegger's discussion of revealing and concealing. Taken in the broad philosophical sense, movement is defined not as mere change of place and the like, but as the very being of entities that are undergoing the process of change. This kind of being consists in anticipating something absent, with the result that

what is absent-but-anticipated determines the entity's present being. Anticipation *is* the being of such entities, and anticipation is determined from the absent-but-anticipated goal. For example, the acorn's being is its becoming an oak tree; and correspondingly the future oak tree, as the goal of the acorn's trajectory, determines the acorn's present being. Likewise, Margaret *is* a graduate student in so far as she is in movement towards her Ph.D. The still-absent degree *qua* anticipated determines her being-a-student.

The absent is, by nature, hidden. But when it is anticipated or intended, the intrinsically hidden, while still remaining absent, becomes quasi-present. It functions as the 'final cause' and *raison d'être* that determines the being of the anticipating entity. That is, even while remaining intrinsically concealed, the absent-as-anticipated 'gives being' (*Es gibt Sein*) to the anticipating entity by disclosing the entity as what it presently is. This pattern of absence-dispensing-presence holds both for the disclosure of Dasein and for the disclosure of the entities Dasein encounters.

It holds pre-eminently for Dasein. Dasein's being is movement, for Dasein exists by anticipating its own absence. Dasein's death remains intrinsically hidden, but when anticipated, the intrinsically hidden becomes quasi-present by determining Dasein's being as mortal becoming. The absent, when anticipated, dispenses Dasein's finite presence.

The same holds for other entities. The anticipated absence determines Dasein's finite being. But Dasein's being is world-disclosive: it holds open the region of meaningful presence in which other entities are disclosed as being-this-or-that. Hence, the intrinsically hidden, when anticipated, determines the presence not only of Dasein but also of the entities Dasein encounters.

Therefore, the very structure of disclosure - that is, the fact that the absent-but-anticipated determines or 'gives' finite presence - entails that its ultimate source remain intrinsically hidden even while disclosing the being of entities. This intrinsic hiddenness at the core of disclosure is what Heidegger calls the 'mystery'. Heidegger argued that the 'mystery' is the ultimate issue in philosophy, and he believed Heraclitus had said as much in his fragment no. 123: 'Disclosure-as-such loves to hide' (Freeman 1971: 33).

Ereignis. The paradigm of movement also explains why Heidegger calls disclosure-as-such '*Ereignis*'. In ordinary German *Ereignis* means 'event', but Heidegger uses it as a word for movement. Playing on the adjective *eigen* ('one's own'), he creates the word *Ereignung*: movement as the process of being drawn into what is one's own. For example, we might imagine that the oak tree as final cause 'pulls' the acorn into what it properly is, by drawing the acorn towards what it is meant to be. This being-pulled is the acorn's movement, its very being. Likewise, Dasein is 'claimed' by death as its final cause and 'pulled forth' by it into mortal becoming. This being-drawn into one's own absence, in such a way that world is engendered and sustained, is what Heidegger calls 'appropriation'. It is what he means by *Ereignis*.

The word '*Ereignis*', along with the image of Dasein being appropriated by the absent, emerges in Heidegger's thought only in the 1930s. However, this later language echoes what Heidegger had earlier called Dasein's thrownness, namely, the fact that Dasein is thrust into possibilities, anticipates its self-absence, and so is 'already' involved in world-disclosure. Both the earlier language of thrown anticipation *of* absence, and the later language of appropriation *by* absence, have the same phenomenon in view: Dasein's alreadiness, its constitutive mortality that makes for world-disclosure.

The paradigm of movement also helps to clarify Heidegger's claim about the concealing-and-revealing, or withdrawing-and-arriving, of being itself (that is, of disclosure-as-such). In a quite typical formulation Heidegger writes: 'Being itself withdraws itself, but as this withdrawal, being is the 'pull' that claims the essence of human being as the place of being's own arrival' (1961: vol. 2, 368). This sentence, which describes the structure of *Ereignis*, may be interpreted as follows:

The 'withdrawal' of disclosure-as-such
(that is, the intrinsic hiddenness of world-disclosive absence)
maintains a relation to Dasein
(which we may call either 'appropriation' or 'thrown anticipation')
that claims Dasein
(by appropriating it into mortal becoming)

so that, *in* Dasein's being,
(in so far as Dasein's being is the openness that is world)
being itself might arrive
(in the form of the relations of significance whereby entities have being-as this-or-that).

The Turn. One can notice a certain shift within Heidegger's work beginning around 1930, both in his style and in the topics he addresses. As regards style, some have claimed that his language becomes more abstruse and poetic, and his thinking less philosophical than mystical. As regards substance, he seems to introduce new topics like 'appropriation' and the 'history of being'.

The problem is to discern whether these and other shifts count as what Heidegger calls the Turn (*die Kehre*). Some argue that beginning in the 1930s Heidegger radically changed his approach and perhaps even his central topic. The early Heidegger, so the argument goes, had understood being itself (that is, disclosure-as-such) from the standpoint of Dasein, whereas the later Heidegger understands Dasein from the standpoint of being itself. But to the contrary it is clear that even the early Heidegger understood Dasein only from the standpoint of being itself.

Heidegger clarifies matters by distinguishing between (1) the Turn and (2) the 'change in thinking' that the Turn demands, both of which are to be kept distinct from (3) the various shifts in form and focus that his philosophy underwent in the 1930s. The point is that, properly speaking, the Turn is not a shift in Heidegger's thinking nor a change in his central topic. The Turn is only a further specification of *Ereignis*. There are three issues here.

First, the 'Turn' is a name for how *Ereignis* operates. *Ereignis* is the appropriation of Dasein for the sake of world-disclosure. For Heidegger, this fact stands over against all theories of the self as an autonomous subject that presuppositionlessly (that is, without a prior world-disclosure) posits its objects in meaning. In opposition to that, *Ereignis* means that Dasein must already be appropriated into world-disclosive absence before anything can be significant at all.

Ereignis also means that Dasein's appropriation by, or thrownness into, world-disclosive absence is the primary and defining moment in Dasein's projection of that disclosure. This reciprocity (*Gegenschwung*) between appropriation/thrownness on the one hand and projection on the other - with the priority going to appropriation/thrownness - constitutes the very structure of *Ereignis* and is what Heidegger calls the Turn. The upshot of this reciprocity is that Dasein must be already pulled into world-disclosive absence (thrown or appropriated into it) if it is to project (that is, hold open) disclosure at all. In a word, the Turn *is Ereignis*.

Second, the 'change in thinking' refers to the personal conversion that the Turn demands. To become aware of the Turn and to accept it as determining one's own being is what Heidegger had earlier called 'resolution' and what he now describes as 'a transformation in human being'. This transformation into an authentic self consists in letting one's own being be defined by the Turn.

Third, the shifts in Heidegger's work in the 1930s - and especially the development and deepening of his insights into thrownness and appropriation - are just that: shifts and developments within a single, continuing project. Important as they are, they are neither the Turn itself nor the change in personal self-understanding that the Turn requires.

6 Forgetfulness, history and metaphysics

Heidegger sees a strong connection between the forgetting of disclosure-as-such, the history of the dispensations of being, and metaphysics.

Forgetting disclosure-as-such. Because disclosure-as-such is intrinsically hidden (this is what is meant by the mystery), it is usually overlooked. When the mystery is overlooked, human being is 'fallen', that is, aware of entities as being-thus-and-so, but oblivious of what it is that 'gives' being to entities. Fallenness is forgetfulness of the mystery. Another term for fallenness is 'errancy', which conveys the image of Dasein 'wandering' among entities-in-their-being without knowing what makes their presence possible. Since disclosure-as-such is sometimes called 'being itself', fallenness is also called 'the forgetfulness of being'.

However, disclosure-as-such need not be forgotten. It is possible, in resolution, to assume one's mortality and become concretely aware of disclosure-as-such in its basic state of hiddenness. Such awareness does not undo the

intrinsic hiddenness of disclosure-as-such or draw it into full presence. Rather, one accepts the concealment of being itself (this is called ‘letting being be’) by resolutely accepting one’s appropriation by absence.

The history of the dispensations of being. Heidegger’s discussions of the ‘history of being’ sometimes verge on the anthropomorphic, and he often uses etymologies that are difficult to carry over into English. Nevertheless, his purpose in all this is clear: to spell out the world-historical dimensions of fallenness.

As we have seen, disclosure-as-such ‘gives’ the being of entities while the ‘giving’ itself remains hidden; and this happens only in so far as Dasein is appropriated by absence. When one forgets the absence that appropriates Dasein, and thus forgets the hidden giving that brings forth the being of entities, fallenness and errancy ensue. Fallen Dasein then focuses on the given (entities-in-their-being) and overlooks the hidden giving (disclosure-as-such). None the less, the hidden giving still goes on giving, but now in a doubly hidden way: it is both intrinsically hidden *and* forgotten. When the hiddenness is forgotten, a disclosure is called a ‘dispensation’ (*Geschick*) of being. The word connotes a portioning-out that holds something back. A certain form of the being of entities is dispensed while the disclosing itself remains both hidden and forgotten.

In German, ‘dispensation’ (*Geschick*) and ‘history’ (*Geschichte*) have their common root in the verb *schicken*, ‘to send’. Playing on those etymologies, Heidegger elaborates a ‘history’ of being, based on the ‘sendings’ or ‘dispensations’ of being. (The usual translations of *Geschick* as ‘fate’ or ‘destiny’ are not helpful here.) In Heidegger’s view each dispensation of being defines a distinct epoch in the history of thought from ancient Greece down to today. He calls the aggregate of such dispensations and epochs the ‘history of being’. Because the whole of these dispensations and epochs is correlative to fallenness, Heidegger seeks to overcome the history of being and return to an awareness of the hidden giving.

Heidegger believes the parameters of each epoch in the history of being can be glimpsed in the name that a major philosopher of the period gave to the being of entities in that age. A non-exhaustive list of such epoch-defining notions of being includes: *idea* in Plato, *energeia* in Aristotle, act in Aquinas, representedness in Descartes, objectivity in Kant, Absolute Spirit in Hegel, and will to power in Nietzsche. What characterizes each such epoch is (1) an understanding of being as some form of the presence of entities and (2) an oblivion of the absence that bestows such presence. None the less, even when forgotten the absence is never abolished, and thus traces of it remain in the various dispensations. Therefore, in studying the texts of classical philosophy Heidegger searches for and retrieves the unexpressed absence (the ‘unsaid’) that hides behind what the text actually expresses (the ‘said’).

Metaphysics. The various ways that presence or being has been dispensed, while absence has been overlooked, are called in their entirety ‘metaphysics’. Heidegger argues that metaphysics as a philosophical position began with Plato and entered its final phase with Nietzsche.

The Greek philosophers who preceded Socrates and Plato were, in Heidegger’s view, *pre*-metaphysical in so far as they had at least a penumbral awareness of disclosure-as-such and at least named it (Heraclitus, for example, called it *logos*, *alētēia*, and *physis*). However, none of these thinkers thematically addressed disclosure-as-such or understood the correlative notions of ek-sistence and Dasein. Heidegger calls the penumbral awareness of disclosure-as-such among archaic Greek thinkers the ‘first beginning’. And he hoped that a ‘new beginning’ would follow the end of metaphysics. If the first beginning was not yet metaphysical, the new beginning will be no longer metaphysical. Heidegger considered his own work a preparation for that new beginning.

But metaphysics persists. The history of the dispensations of being has reached its fullness in the present epoch of technology. As Heidegger uses the word, ‘technology’ refers not to hardware or software or the methods and materials of applied science. Rather, it names a dispensation in the history of metaphysics, in fact the final one. It names the way in which entities-in-their-being are disclosed today.

Heidegger maintains that in the epoch of technology entities are taken as a stockpile of matter that is in principle completely knowable by human reason and wholly available for human use. With this notion metaphysics arrives at its most extreme oblivion of disclosure-as-such. In our time, Heidegger says, the presence of entities has become everything, while the absence that brings about that presence has become nothing. He calls this nil-status of absence ‘nihilism’.

Overcoming metaphysics. None the less, Heidegger sees a glimmer of light in the dark epoch of nihilism. In this

final dispensation of metaphysics, the hidden giving does not cease to function, even when it is completely forgotten. It continues dispensing presence - paradoxically even the nihilistic presence which obscures the absence that gives it. Because the hidden giving goes on giving even when it is forgotten, we can still experience it today (in a mood not unlike dread) and retrieve it. This recovery of world-disclosive absence requires resolution or, as Heidegger now calls it, 'the entrance into *Ereignis*'. To enter *Ereignis* today is to experience a different kind of *nihil* ('nothing') from the one that defines nihilism. The absence that bestows presence is itself a kind of 'nothing' (not-a-thing). This absence is no entity, nor can it be reduced to the being of any specific entity or be present the way an entity is. That is why it is so easily overlooked. Its 'nothingness' is its intrinsic hiddenness.

To enter *Ereignis* is to become aware of and to accept the disclosive *nihil* that rescues one from nihilism. Thereupon, says Heidegger, metaphysics as the history of the dispensations of being ceases and a new beginning takes place - at least for those individuals who achieve authenticity by way of resolution. But metaphysics will continue for those who remain inauthentic, because dispensation is correlative to fallenness.

Summary. The forgetting of disclosure-as-such is metaphysics. Metaphysics knows entities-in-their-being but ignores the very giving of that being. The aggregate of the epochs of metaphysics is the history of the dispensations of being. The history of these dispensations culminates in the epoch of technology and nihilism. But world-disclosive absence can still be retrieved; and when it is retrieved, it ushers in (at least for authentic individuals) a new beginning of ek-sistence and Dasein.

7 The work of art

One of Heidegger's most challenging essays is 'The Origin of the Work of Art', originally drafted in 1935 and published in an expanded version only in 1950. There he distinguishes between the work of art as a specific entity (for example, a poem or a painting) and art itself, the latter being understood not as a collective name for, but rather as the essence and origin of, all works of art. Heidegger asks what art itself is, and he answers that art is a unique kind of disclosure.

Dasein is disclosive of the being of an entity in many ways, some of them ordinary and some of them extraordinary. An outcome common to both kinds of disclosure is that the disclosed entity is seen as what it is: it appears in its form. Examples of ordinary, everyday ways of disclosing the being of entities include showing oneself to be adept at the flute, or moulding clay into a vase, or concluding that the accused is innocent. Each of these ordinary cases of praxis, production and theory does indeed disclose some entity as being this or that, but the focus is on showing what the *entity* is rather than on showing how the entity's *being* is disclosed. On the other hand, extraordinary acts of disclosure bring to attention not only the disclosed entity but above all the event of disclosure of that entity's being. Extraordinary acts of disclosure let us see the very fact that, and the way in which, an entity has become meaningfully present in its being. In these cases not only does an entity appear in its form (as happens in any instance of disclosure) but more importantly the very disclosure of the being of the entity 'is established' (*sich einrichten*) in the entity and is seen there as such.

Heidegger lists five examples of extraordinary disclosure: the constitution of a nation-state; the nearness of god; the giving of one's life for another; the thinker's questioning as revealing that being can be questioned; and the 'installation' (*Sich-ins-Werk-Setzen*) of disclosure in a work of art. Each of these cases discloses, in its own particular way, not just an entity but the very disclosure of that entity's being. Heidegger seeks to understand the particular way in which art itself discloses disclosure by 'installing' disclosure in the work of art.

In his essay Heidegger refers mainly to two works of art: van Gogh's canvas 'Old Shoes', painted in Paris in 1886-7 and now hung in the Stedelijk Museum, Amsterdam; and the 5th century BC Doric Temple of Hera II - the so-called Temple of Poseidon - at Paestum (Lucania), Italy. Let us consider the temple at Paestum as we attempt to answer two questions: what gets disclosed in a work of art and how does it get disclosed?

(1) *What gets disclosed in a work of art?* Heidegger gives three answers. First, a work of art lets us see disclosure in the form of 'world' and 'earth'. A work of art discloses not just an entity or an ensemble of entities but the whole realm of significance whereby an ensemble of entities gets its finite meaning. The temple at Paestum not only houses (and thus discloses) the goddess Hera, but more importantly lets us see the social and historical world - rooted as it was in the natural setting of Lucania - that Hera's presence guaranteed for the Greek colonists. A work of art, Heidegger argues, reveals the very event of disclosure, which event he calls the happening of world

and earth, where ‘earth’ refers not only to nature and natural entities but more broadly to all entities within a specific world.

Second, a work of art lets us see the radical tension that discloses a specific world of significance. Heidegger understands being-in-the-world as a ‘struggle’ (*Streit* or *polemos*) between a given world and its earth, between the self-expanding urge of a set of human possibilities and the rootedness of such possibilities in a specific natural environment. Here, ‘struggle’ is another name for the event of disclosure whereby a particular world is opened up and maintained. What a specific work of art discloses is one particular struggle that discloses one particular world - for instance, the world of the Greek colonists at Paestum.

Third, a work of art shows us disclosure-as-such. The movement of opening up a particular world is only one instance of the general movement of *alēthēia*: the ‘wresting’ of being-at-all from the absolute absence into which Dasein is appropriated. Thus a work of art not only shows us a particular world-disclosive struggle (the way the temple of Hera shows us the earth-world tension at Paestum) but also lets us see the ‘original struggle’ (*Urstreit*) of disclosure-as-such, whereby significance is wrested from the double closure of intrinsic hiddenness and fallenness.

In short, what a work of art reveals is disclosure in three forms: as world and earth; as the struggle that opens up a specific world and lets its entities be meaningful; and as the original struggle that structures all such particular disclosures.

(2) *How does a work of art disclose disclosure?* The specific way that art discloses disclosure is by ‘installing’ it in a given work of art. Here, ‘to install’ means to bring to stability; and ‘to install disclosure’ means to incorporate it into the physical form of a work of art. There are three corollaries:

What the installing is not. Heidegger does not claim that the work of art ‘sets up’ the world and ‘sets forth’ the earth for the first time. That is, installing the disclosure of earth and world in the work of art is not the only or even the first way that earth and world get disclosed. The sanctuary of Hera was not the first to open up the world of Paestum and disclose the fields and flocks for what they are. Tradesmen and farmers had been doing that - that is, the disclosive struggle of world and earth had been bestowing form and meaning - for at least a century before the temple was built.

What the installing is and does. Art discloses, in a new and distinctive way, a disclosure of earth and world that is already operative. Heidegger argues that the temple as disclosive (a) captures and sustains the openness of that world and its rootedness in nature, and (b) shows how, within that world, nature comes forth into the forms of entities while remaining rooted in itself. Heidegger calls these two functions, which happen only in art, the ‘setting up’ of world and the ‘setting forth’ of earth.

The work of art lets us see - directly, experientially and in all its glory - the already operative interplay of human history’s rootedness in nature and nature’s emergence into human history. In Heidegger’s words, art ‘stabilizes’ (*zum Stehen bringen*) the disclosive struggle of world and earth by ‘installing’ it in a particular work of art, such that in and through that medium, disclosure ‘shines forth’ brilliantly in beauty.

The two ways art discloses disclosure, and their unity. Art itself is a specific and distinctive way in which Dasein is disclosive: it discloses disclosure by installing disclosure in the physical form of a work of art. This installation has two moments: the creation and the preservation of the work of art.

Creation is an artist’s Dasein-activity of incorporating disclosure - the world-openness that is already operative - into a material medium (stone, colour, language and so on). This incorporation of disclosure is carried out in such a way that the material medium is not subordinated to anything other than disclosure (for example, it is not subordinated to ‘usefulness’). Rather, the medium becomes, for whoever experiences it, the immediate disclosure of disclosure.

Preservation is the corresponding Dasein-activity of maintaining the power of disclosure in the work of art by resolutely letting disclosure continue to be seen there. Creation and preservation are the two ways that Dasein ‘projects’ (holds open and sustains) the disclosure that is installed in the work of art. The unity of creation and preservation is art itself, which Heidegger calls *Dichtung* - not ‘poetry’ but *poiesis*, the creating-and-preserving installation of disclosure in a disclosive medium.

Disclosure is the central topic of all Heidegger's philosophy, and this fact shines brilliantly through his reflection on the origin of the work of art. Art, both as creation and as preservation, is a specific and distinctive Dasein-activity: the disclosure of disclosure in a medium that is disclosive. In the work of art, as in Heidegger's own work, it's *alēthēia* all the way down.

See also: Hermeneutics §4; Kuki Shūzō; Phenomenological movement §§4-5; Watsuji Tetsurō

THOMAS SHEEHAN

List of works

Heidegger, M. (1975-) *Gesamtausgabe (Collected Edition)*, Frankfurt: Vittorio Klostermann.(The standard edition of Heidegger's works. Over eighty volumes are projected, of which more than forty appeared by 1997. English translations of individual volumes are given at the end of the list of works below.)

Heidegger, M. (1914) *Die Lehre vom Urteil im Psychologismus. Ein kritisch-positiver Beitrag zur Logik (The Doctrine of Judgment in Psychologism: A Critical-positive Contribution to Logic)*, Leipzig: Barth. (Heidegger's doctoral dissertation.)

Heidegger, M. (1916) *Die Kategorien- und Bedeutungslehre des Duns Scotus (Duns Scotus' Doctrine of Categories and Meaning)*, Tübingen: J.C.B. Mohr (Paul Siebeck). (Heidegger's Habilitationsschrift, the qualifying dissertation required for teaching at a university.)

Heidegger, M. (1927) 'Sein und Zeit', *Jahrbuch für Philosophie und phänomenologische Forschung* 8: 1-438; *Sein und Zeit*, Halle an der Salle: Max Niemeyer; trans. J. Macquarrie and E. Robinson, *Being and Time*, New York: Harper & Row, 1962; trans. J. Stambaugh, *Being and Time*, Albany, NY: State University of New York Press, 1996.(Heidegger's most famous work, which treats the structure of Dasein as being-in-the-world and as temporal. The unpublished second half of the work was to have shown that the meaning of being is time.)

Heidegger, M. (1929a) 'Vom Wesen des Grundes', *Jahrbuch für Philosophie und phänomenologische Forschung*, supplement 71-100; trans. T. Malick, *The Essence of Reasons*, Evanston, IL: Northwestern University Press, 1969; trans. W. McNeill, 'On the Essence of Ground', in *Pathmarks*, New York: Cambridge University Press, 1997. (An essay on truth, transcendence and ground, written in honour of Edmund Husserl's 80th birthday, April 1929.)

Heidegger, M. (1929b) *Kant und das Problem der Metaphysik*, Bonn: Friedrich Cohen; trans. R. Taft, *Kant and the Problem of Metaphysics*, Bloomington, IN: Indiana University Press, 1990.(Originally conceived as part of the second half of *Sein und Zeit*, this work argues that the hidden meaning of the transcendental imagination in the 'A' version of Kant's *Critique of Pure Reason* is the temporality of Dasein as presented in *Sein und Zeit*.)

Heidegger, M. (1929c) *Was ist Metaphysik?*, Bonn: Friedrich Cohen; trans. D.F. Krell, 'What is Metaphysics?', in D.F. Krell, *Basic Writings*, revised edn, San Francisco: Harper, 1993; and in *Pathmarks*, New York: Cambridge University Press, 1997.(Heidegger's inaugural address as Husserl's successor at Freiburg University. It discusses boredom as a 'basic mood' and broaches the topic of being as 'the nothing' or 'not-a-thing'.)

Heidegger, M. (1933) *Die Selbstbehauptung der deutschen Universität*, Breslau: Wilhelm Gottlieb Korn; trans. K. Harries, 'The Self-Assertion of the German University', *Review of Metaphysics* (1985) 38: 470-80. (Heidegger's controversial inaugural address as rector of Freiburg University, 27 May 1933.)

Heidegger, M. (1942) 'Platons Lehre von der Wahrheit', *Geistige Überlieferung* 2: 96-124; trans. J. Barlow, 'Plato's Doctrine of Truth', in W. Barrett (ed.) *Philosophy in the Twentieth Century*, New York: Random House, 1962, vol. 2; trans. T. Sheehan, 'Plato's Doctrine of Truth', in *Pathmarks*, New York: Cambridge University Press, 1997.(This close reading of Allegory of the Cave in Plato's *Republic* argues that Plato inaugurates the metaphysical notion of truth as correspondence.)

Heidegger, M. (1943a) *Vom Wesen der Wahrheit*, Frankfurt: Vittorio Klosterman; trans. J. Sallis, 'On the Essence of Truth', in D.F. Krell (ed.) *Basic Writings*, revised edn, San Francisco: Harper, 1993; and 'On the Essence of Truth', in *Pathmarks*, New York: Cambridge University Press, 1997.(Originally drafted in 1930, the essay discusses the levels of disclosure: propositional truth, the manifestness of entities, and disclosure-as-such or *aletheia*.)

Heidegger, M. (1947) 'Brief über den Humanismus', in *Platons Lehre von der Wahrheit. Mit einem Brief über den Humanismus*, Bern: Francke; trans. F.A. Capuzzi, 'Letter on Humanism', in D.F. Krell (ed.) *Basic Writings*, revised edn, San Francisco: Harper, 1993; and in *Pathmarks*, New York: Cambridge University Press, 1997.(Written as an open letter to Jean Beaufret, Heidegger's first publication after the Second World War reveals the shifts his thinking had undergone in the 1930s.)

Heidegger, M. (1950) *Holzwege (Forest Paths)*, Frankfurt: Vittorio Klosterman.(A collection of essays dating from 1936 to 1946.)The following six references are the translations into English of all the essays contained in

Holzwege. Dates within parentheses indicate the original redaction of the German text.

Heidegger, M. (1935) ‘The Origin of the Work of Art’, trans. A. Hofstadter, in D.F. Krell (ed.) *Basic Writings*, revised edn, San Francisco: Harper, 1993.(Originally delivered as a lecture in 1935, the essay argues that art, taken as the essence of any work of art, consists in disclosing disclosure by ‘installing’ it in the physical medium of the work.)

Heidegger, M. (1938) ‘The Age of the World Picture’, in W. Lovitt (trans. and ed.) *The Question Concerning Technology and Other Essays*, New York: Harper & Row, 1977.(Originally delivered as a lecture in 1938, the essay discusses, among other things, the birth of modernity with Descartes’ view of thinking as the representation of entities by the subject as unshakable foundation.)

Heidegger, M. (1942-3) *Hegel’s Concept of Experience*, trans. J.G. Gray, New York: Harper & Row, 1970. (Drawn from seminars Heidegger gave in 1942-3, the essay presents Hegel as the culmination of subject-centred metaphysics.)

Heidegger, M. (1943b) ‘The Word of Nietzsche "God is Dead"’, in W. Lovitt (trans. and ed.) *The Question Concerning Technology and Other Essays*, New York: Harper & Row, 1977.(Originally presented in 1943 (but drawing on lecture courses dating from 1936 to 1940), the essay interprets no. 125 of Nietzsche, *The Gay Science*, and other texts, in order to present Nietzsche’s thought as the culmination of Western metaphysics.)

Heidegger, M. (1946a) ‘What Are Poets For?’, trans. A. Hofstadter, in D.F. Krell (ed.) *Poetry, Language, Thought*, New York: Harper & Row, 1971.(Given as a lecture to commemorate the twentieth anniversary of Rainer Maria Rilke’s death, the essay argues that the poet was both subject to Nietzschean nihilism and attempted to overcome it by recovering the authentic sense of language.)

Heidegger, M. (1946b) ‘The Anaximander Fragment’, trans. D.F. Krell, in D.F. Krell and F.A. Capuzzi (eds) *Early Greek Thinking*, New York: Harper & Row, 1975.(Dating from 1946 but drawing on work done as early as 1932, this essay provides a close reading of Anaximander’s Fragment 1 and offers some remarks on the history of being.)

Heidegger, M. (1951a) *Erläuterungen zu Hölderlins Dichtung (Elucidations of Hölderlin’s Poetry)*, Frankfurt: Vittorio Klostermann; 4th expanded edn, 1971.(Six essays on Hölderlin’s poetry.)Two of these essays on Hölderlin’s poetry have been translated into English and are given below. Dates within parentheses indicate the original redaction of the German text.

Heidegger, M. (1936) ‘Hölderlin and the Essence of Poetry’, in W. Brock (ed.) *Existence and Being*, Chicago: Henry Regnery, 1949.(Delivered as a lecture in Rome in 1936 and first published the following year, the essay presents Hölderlin as the ‘poet of poetry’ and reflects on *poiēsis* as the ‘establishment’ of disclosure.)

Heidegger, M. (1943c) ‘Remembrance of the Poet’, in W. Brock (ed.) *Existence and Being*, Chicago, IL: Henry Regnery, 1949.(Given as a lecture in June 1943, at the centenary celebration of Hölderlin’s death, the essay interprets the poet’s elegy ‘Heimkunft/An den Verwandten’ (‘Homecoming/To the Kinsmen’) and reflects on the poet’s relation to disclosure-as-such, here called ‘the holy’.)

Heidegger, M. (1953a) *Einführung in die Metaphysik*, Tübingen: Max Niemeyer; trans. R. Manheim, *An Introduction to Metaphysics*, New Haven, CT: Yale University Press, 1959.(A lecture course from 1935, treating among other things Heidegger’s interpretation of the meaning of being in Parmenides and Heraclitus.)

Heidegger, M. (1954) *Vorträge und Aufsätze (Lectures and Essays)*, Pfullingen: Günter Neske. (A collection of eleven essays, ranging in date from 1936 to 1954.)Ten of the eleven essays in *Vorträge und Aufsätze* have been translated into English and are given below. Dates within parentheses indicate the original redaction of the German text.

Heidegger, M. (1936-46) ‘Overcoming Metaphysics’, in J. Stambaugh (trans. and ed.) *The End of Philosophy*, New York: Harper & Row, 1973.(Thirty-eight brief notes, dating from 1936 to 1946, on the overcoming of metaphysics.)

Heidegger, M. (1943d) ‘Aletheia’, trans. F.A. Capuzzi, in D.F. Krell and F.A. Capuzzi (eds) *Early Greek Thinking*, New York: Harper & Row, 1975.(Drawing on Heidegger’s 1943 course, ‘The Beginning of Western Thinking (Heraclitus)’, the text, first published in 1954, interprets Heraclitus’ Fragment 16 in the light of what Heidegger calls ‘the clearing’/disclosure-as-such.)

Heidegger, M. (1944) ‘Logos’, trans. D.F. Krell, in D.F. Krell and F.A. Capuzzi (eds) *Early Greek Thinking*, New York: Harper & Row, 1975.(Drawing on Heidegger’s 1944 course ‘Logic (Heraclitus’ Doctrine of *Logos*)’ the text interprets Heraclitus’ Fragment 50 in the light of the relation between disclosure-as-such and Dasein’s correspondence to it.)

Heidegger, M. (1949a) ‘The Thing’, trans. A. Hofstadter, in D.F. Krell (ed.) *Poetry, Language, Thought*, New York: Harper & Row, 1971.(Written in 1949, delivered as a lecture in June 1950, and first published in 1951, the essay offers a phenomenological description of a wine-pitcher as a way of reflecting on the ‘nearness’ of

things.)
Heidegger, M. (1949b) ‘The Question Concerning Technology’, in W. Lovitt (trans. and ed.) *The Question Concerning Technology and Other Essays*, New York: Harper & Row, 1977.(Written and delivered as a lecture in 1949 under the title ‘Das Gestell’, (‘The Enframing’), then delivered under its present title, ‘Die Frage nach der Technik’, in 1953 and published in the following year, the text argues that technology is not primarily something instrumental (a means to an end) but a form of disclosure, and that modern technology, as the demand for complete disclosure, is intrinsically nihilistic.)
Heidegger, M. (1951b) ‘Building Dwelling Thinking’, trans. A. Hofstadter, in D.F. Krell (ed.) *Poetry, Language, Thought*, New York: Harper & Row, 1971.(Delivered as a lecture in 1951 and first published in the following year, the text meditates on the ‘nearness’ of things in terms of the ‘bringing-forth’ (*poiesis*) of things.)
Heidegger, M. (1951c) ‘…Poetically Man Dwells…’, trans. A. Hofstadter, in D.F. Krell (ed.) *Poetry, Language, Thought*, New York: Harper & Row, 1971.(Delivered as a lecture in October 1951 and first published in 1954, the text reflects on Hölderlin and on poetry as a bringing-forth (*poiesis*) of things.
Heidegger, M. (1952) ‘Moira’, trans. F.A. Capuzzi, in D.F. Krell and F.A. Capuzzi (eds) *Early Greek Thinking*, New York: Harper & Row, 1975.(Originally planned as part of Heidegger’s 1951-2 lecture course ‘Was heißt Denken?’ the text interprets Parmenides’ Fragment 8, lines 34-41, and specifically the word ‘*moira*’, as referring to the togetherness of Dasein and disclosure-as-such.)
Heidegger, M. (1953b) ‘Science and Reflection’, in W. Lovitt (trans. and ed.) *The Question Concerning Technology and Other Essays*, New York: Harper & Row, 1977. (Delivered as a lecture in 1953, the text probes the thesis that science is the theory of the real and raises the question of disclosure as the concealed essence of science.)
Heidegger, M. (1953c) ‘Who is Nietzsche’s Zarathustra?’, trans. D.F. Krell, in *Nietzsche*, New York: Harper & Row, 1984, vol. 2.(Delivered as a lecture in May 1953, the essay interprets some major themes in Nietzsche - time and revenge, nihilism, eternal recurrence, will to power - and suggests that Zarathustra represents the togetherness of eternal recurrence and superman.)
Heidegger, M. (1956) *Was heißt Denken?*, Tübingen: Max Niemeyer; trans. F. D. Wieck and J.G. Gray, *What Is Called Thinking?*, New York: Harper & Row, 1968.(The text of Heidegger’s wide-ranging, two-semester lecture course, 1951-2.)
Heidegger, M. (1957a) *Der Satz vom Grund*, Pfullingen: Günther Neske; trans. R. Lilly, *The Principle of Reason*, Bloomington, IN: Indiana University Press, 1991.(The text of Heidegger’s lecture course, 1955-6.)
Heidegger, M. (1957b) *Identität und Differenz*, Pfullingen: Günther Neske; trans. J. Stambaugh, *Identity and Difference*, New York: Harper & Row, 1969.(Two lectures from 1957, one on the principle of identity, the other on the ontotheological structure of metaphysics.)
Heidegger, M. (1958) ‘Vom Wesen und Begriff der Φ[uacgr]σις. Aristoteles Physik B 1’, *Il Pensiero* 3: 131-56; 265-90; trans. T. Sheehan, ‘On the Being and Conception of PHYSIS in Aristotle’s *Physics* B, 1’, *Man and World* (1976) 9: 221-70; and in *Pathmarks*, New York: Cambridge University Press, 1997.(Based on Heidegger’s 1940 seminar, the text examines Aristotle’s understanding of *physis* (nature) and argues that *physis* originally meant disclosure-as-such.)
Heidegger, M. (1959a) *Gelassenheit*, Pfullingen: Günther Neske; trans. J.M. Anderson and E.H. Freund, *Discourse on Thinking*, New York: Harper & Row, 1966.(Two occasional pieces, one an imaginary dialogue dating from 1944, the other a speech commemorating the eighteenth-century composer Conradin Kreutzer, dating from 1955.)
Heidegger, M. (1959b) *Unterwegs zur Sprache*, Pfullingen: Günther Neske; trans. P.D. Hertz and J. Stambaugh, *On the Way to Language*, New York: Harper & Row, 1971.(Four essays and one dialogue, ranging in date from 1950 to 1958, dealing with the question of language.)
Heidegger, M. (1961) *Nietzsche*, Pfullingen: Günther Neske, 2 vols; trans. D.F. Krell and F. Capuzzi, *Nietzsche*, New York: Harper & Row, 1979-87, 4 vols.(Lecture courses and notes on Nietzsche, dating from 1936 to 1946.)
Heidegger, M. (1962) *Die Frage nach dem Ding. Zu Kants Lehre von der transzendentalen Grundsätzen*, Tübingen: Max Niemeyer; trans. W.B. Barton and V. Deutsch, *What Is a Thing?*, Chicago, IL: Henry Regnery, 1967.(The text of Heidegger’s lecture course of 1935-6, which includes a substantial discussion of the ‘Analytic of Principles’ in Kant’s *Critique*.)
Heidegger, M. (1967, 1976) *Wegmarken*, Frankfurt: Vittorio Klostermann; trans. and ed. D.F. Krell, W. McNeill and J. Sallis, *Pathmarks*, New York: Cambridge University Press, 1997.(A collection of fourteen of Heidegger’s most important essays, ranging in date from 1919 to 1961. The essays include Heidegger 1929a, 1929c, 1942, 1943, 1947 and ‘Zur Seinsfrage’ (‘The Question of Being’) 1955.)

Heidegger, M. (1969) *Zur Sache des Denkens*, Tübingen: Max Niemeyer; trans. J. Stambaugh, *On Time and Being*, New York: Harper & Row, 1972.(Four shorter texts dating from 1961 to 1964, including the 1962 lecture 'Zeit und Sein' ('Time and Being').)

Heidegger, M. (1970a) *Heraklit. Seminar Wintersemester 1966/1967*, Frankfurt: Vittorio Klostermann; trans. C.H. Seibert, *Heraclitus Seminar 1966/67*, Tuscaloosa, AL: University of Alabama Press, 1979.(The text of the seminar Heidegger conducted in tandem with Eugen Fink, 1966-7.)

Heidegger, M. (1970b) *Phänomenologie und Theologie*, Frankfurt: Vittorio Klostermann; trans. J.G. Hart and J.C. Maraldo, *The Piety of Thinking*, Bloomington, IN: Indiana University Press, 1994.(Two essays, dated 1927 and 1964 respectively, on the possible relation between theology and Heidegger's thinking.)

Heidegger, M. (1971) *Schellings Handlung Über das Wesen der menschlichen Freiheit (1809)*, Tübingen: Max Niemeyer; trans. J. Stambaugh, *Schelling's Treatise On the Essence of Human Freedom*, Athens, OH: Ohio University Press, 1985.(The text of Heidegger's lecture course of 1936.)

Heidegger, M. (1972) *Frühe Schriften (Early Writings)*, Frankfurt: Vittorio Klostermann.(This volume reprints most notably Heidegger's Ph.D. dissertation of 1914, and his qualifying dissertation of 1916.)

Heidegger, M. (1976) 'Nur noch ein Gott kann uns retten', *Der Spiegel* 23: 193-219; trans. W.J. Richardson, 'Only a God Can Save Us: The *Spiegel* Interview', in T. Sheehan (ed.) *Heidegger, the Man and the Thinker*, New Brunswick, NJ: Rutgers University/Transaction Publishers, 1981.(In this posthumously published interview Heidegger attempts to explain, among other things, his relation to the Nazi regime in 1933-4.)

Heidegger, M. (1977a) *Vier Seminare (Four Seminars)*, Frankfurt: Vittorio Klostermann; 2nd seminar trans. as 'A Heidegger Seminar on Hegel's *Differenzschrift*', *Southwest Journal of Philosophy* (1980) 11.(The records of four informal seminars that Heidegger conducted with friends and colleagues between 1966 and 1973.)

1 English translations of works in the Collected Edition

Heidegger, M. (1975) *Die Grundprobleme der Phänomenologie*, in *Gesamtausgabe*, vol. 24; trans. A. Hofstadter, *The Basic Problems of Phenomenology*, Bloomington, IN: Indiana University Press, 1982.(Lecture course, summer 1927.)

Heidegger, M. (1977b) *Phänomenologische Interpretation von Kants Kritik der reinen Vernunft*, in *Gesamtausgabe*, vol. 25; trans. P. Emad and K. Maly, *Phenomenological Interpretation of Kant's Critique of Pure Reason*, Bloomington, IN: Indiana University Press, 1997.(Lecture course, winter 1927-8.)

Heidegger, M. (1978) *Metaphysische Angangsründe der Logik*, in *Gesamtausgabe*, vol. 26; trans. M. Heim, *The Metaphysical Foundations of Logik*, Bloomington, IN: Indiana University Press, 1984.(Lecture course, summer 1928.)

Heidegger, M. (1979) *Prolegomena zur Geschichte des Zeitbegriffs*, in *Gesamtausgabe*, vol. 20; trans. T. Kisiel, *History of the Concept of Time: Prologomena*, Bloomington, IN: Indiana University Press, 1985.(Lecture course, summer 1925.)

Heidegger, M. (1980) *Hegels Phänomenologie des Geistes*, in *Gesamtausgabe*, vol. 32; trans. P. Emad and K. Maly, *Hegel's Phenomenology of Spirit*, Bloomington, IN: Indiana University Press, 1988.(Lecture course, winter 1930-1.)

Heidegger, M. (1981) *Aristoteles, Metaphysik 1-3*, in *Gesamtausgabe*, vol. 33; trans. W. Brogan, *Aristotle, Metaphysics 1-3: On the Essence and Actuality of Force*, Bloomington, IN: Indiana University Press, 1996. (Lecture course, summer 1931.)

Heidegger, M. (1982) *Parmenides*, in *Gesamtausgabe*, vol. 54; trans. A. Schuwer and R. Rojcewicz, *Parmenides*, Bloomington, IN: Indiana University Press, 1992.(Lecture course, winter 1942-3.)

Heidegger, M. (1983) *Die Grundbegriffe der Metaphysik. Welt - Endlicheit - Einsamkeit*, in *Gesamtausgabe*, vol. 29/30; trans. W. McNeill and N. Walker, *The Fundamental Concepts of Metaphysics: World, Finitude, Solitude*, Bloomington, IN: Indiana University Press, 1995.(Lecture course, winter 1929-30.)

Heidegger, M. (1984a) *Grundfragen der Philosophie. Ausgewählte 'Probleme' der 'Logik'*, in *Gesamtausgabe*, vol. 45; trans. R. Rojcewicz and A. Schuwer, *Basic Questions of Philosophy: Selected 'Problems' of 'Logic'*, Bloomington, IN: Indiana University Press, 1994.(Lecture course, winter 1937-8.)

Heidegger, M. (1984b) *Hölderlins Hymne 'Der Ister'*, in *Gesamtausgabe*, vol. 53; trans. W. McNeill and J. Davis, *Hölderlin's Hymne 'The Ister'*, Bloomington, IN: Indiana University Press, 1996.(Lecture course, summer 1942)

Heidegger, M. (1989) *Beiträge zur Philosophie (Vom Ereignis)*, in *Gesamtausgabe*, vol. 65; trans. P. Emad and K. Maly, *Contributions to Philosophy: On Ereignis*, Bloomington, IN: Indiana University Press, 1989.(Working notes, 1936-8.)

Heidegger, M. (1991) *Grundbegriffe*, in *Gesamtausgabe*, vol. 51; trans. G.A. Aylesworth, *Basic Concepts*,

Bloomington, IN: Indiana University Press, 1993.(Lecture course, summer 1941.)

References and further reading

Freeman, K. (ed.) (1971) *Ancilla to the Pre-Socratic Philosophers*, Cambridge, MA: Harvard University Press. (Heidegger interprets Heraclitus' word *physis* as meaning disclosure-as-such.)

Husserl, E. (1900-1) *Logische Untersuchungen*, Halle an der Salle: Max Niemeyer, 2 vols; trans. J.N. Findlay, *Logical Investigations*, London: Routledge & Kegan Paul, 1970, 2 vols.(Husserl's foundational work in phenomenology.)

Husserl, E. (1913) *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*, vol. 1, *Jahrbuch für Philosophie und phänomenologische Forschung* 1: 1-323; Halle an der Salle: Max Niemeyer; trans. F. Kersten, *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, The Hague: Nijhoff, 1982.(Earliest published evidence of Husserl's turn to transcendental phenomenology and his use of the reductions.)

Kisiel, T. (1993) *The Genesis of Heidegger's Being and Time*, Berkeley, CA: University of California Press. (Exhaustive treatment of Heidegger's development, 1915-26.)

Pöggeler, O. (1987) *Martin Heidegger's Path of Thinking*, Atlantic Highlands, NJ: Humanities Press.(Lucid overview by the leading German commentator.)

Richardson, W.J. (1963) *Heidegger: Through Phenomenology to Thought*, The Hague: Nijhoff.(The classical presentation of the entire *oeuvre* by the pre-eminent Heidegger scholar.)

Sallis, J. (1986) *Delimitations: Phenomenology and the End of Metaphysics*, Bloomington, IN: Indiana University Press; 2nd expanded edn, 1995.(Groundbreaking essays by a major American interpreter.)

Sass, H.-M. (1982) *Martin Heidegger: Bibliography and Glossary*, Bowling Green, OH: Philosophy Documentation Center.(The most comprehensive bibliography in English, but needing to be supplemented by materials found in Sass 1968, 1975.)

Sass, H.-M. (1968) *Heidegger-Bibliographie*, Meisenheim am Glan: Anton Hain.(First comprehensive bibliography of primary and secondary sources up to 1967.)

Sass, H.-M. (1975) *Materialien zur Heidegger-Bibliographie 1917-1972*, Meisenheim am Glan: Anton Hain.(This compliments and revises the information in the previous entry.)

Schürmann, R. (1987) *Heidegger on Being and Acting: From Principles to Anarchy*, Bloomington, IN: Indiana University Press.(Comprehensive interpretation and an argument for postmetaphysical an-archy.)

Taminiaux, J. (1991) *Heidegger and the Project of Fundamental Ontology*, Albany, NY: State University of New York Press.(Essays on Heidegger's early philosophy.)

Van Buren, J. (1994) *The Young Heidegger: Rumor of the Hidden King*, Bloomington, IN: Indiana University Press.(Thorough account of Heidegger's early development.)

Zimmermann, M. (1990) *Heidegger's Confrontation with Modernity: Technology, Politics, Art*, Bloomington, IN: Indiana University Press.(Focuses on the connection between Heidegger's relation to Nazism and his views on technology.)

Heideggerian philosophy of science

Heidegger's importance in the philosophy of science stems less from his scattered remarks about science than from the larger conception of intentionality and ontology that informs them. Heidegger's earliest major work, Being and Time (1927), displayed everyday practical purposive activity as the most fundamental setting for the disclosure of things in the world. Heidegger claimed that the traditional epistemological conception of a subject who represents objects was derivative from and dependent upon such ongoing everyday practical engagement with one's surroundings. Science was then supposed to be the practice that allows things to show themselves shorn of their significance within the 'in-order-to-for-the-sake-of' structure of everyday activity; nevertheless, the sense of scientific claims remained dependent upon the everyday interactions from which they were abstracted.

Shortly after writing Being and Time, Heidegger revised his project in ways that also transformed his account of science. His overall project shifted from describing the transcendental structure of the meaning of being, to interpreting the 'history of being'. Science was reinterpreted as an activity ('research') closely allied with machine technology, and oriented towards more extensive and intensive manipulation and ordering of things. Understood as such, science for Heidegger was an essential manifestation of the modern age. Whereas, earlier, Heidegger thought that science presupposed a philosophical ontology, Heidegger eventually portrayed science and technology as the conclusion of the philosophical tradition. While philosophical metaphysics and epistemology were thus naturalized, Heidegger was concerned with the possibility of a way of thinking outside this convergence of scientific and philosophical metaphysics.

1 Science in Heidegger's early work

Heidegger followed Edmund Husserl's phenomenology in describing the meaning structures through which beings become manifest or *evident* (see Phenomenological movement §5). His inquiry into the 'meaning of being' accepted Husserl's claim that any intentional directedness also incorporates a 'categorial intuition' of being, as an understanding of what could fulfil that intending. For Husserl, however, perception or intuition was the model for intentional fulfilment (see Husserl, E. §§2-5); Heidegger argued instead that using equipment to realize an end was more basic. This shift had several important consequences. Heidegger argued that equipment is encountered holistically: understanding and using one instrument requires understanding others, and what we encounter foremost in successful practice is an interrelated equipmental setting. Equipment use is also not mediated: one encounters the tool itself rather than a representation of it. Finally, using equipment is socially normative. Equipmental complexes are organized around what one does with them and what they are for, and are governed by the enforcement of norms of correct use. Heidegger argued that the subjects of everyday practice are not individuated knowers, but anonymous, undifferentiated doers of what 'one' does.

The most basic form of intentionality is thus not consciousness but agency. We encounter the world as already organized around available *possibilities* for action that make sense of ourselves and our surroundings. We *interpret* the world and ourselves by enacting some of the possibilities disclosed by understanding the world as a setting for action. For Heidegger, one interprets something most appropriately as, for example, a hammer by hammering with it, not by perceiving it or making assertions; one also thereby interprets nails and boards, oneself as a builder, and one's surroundings as a setting in which hammering makes sense. Heidegger thus takes the 'hermeneutical circle' as a general structure of all meaningful activity, and not just of the interpretation of texts. Intentionality is thus fundamentally temporal: present activity is directed towards a future by working out possibilities made available by having been already 'thrown' into a meaningful situation.

Heidegger portrayed science as a distinctive kind of interpretive engagement with the world. Although science also uses equipment for locally intelligible purposes, he ascribed to these local performances a characteristic end, namely the interpretation of things in the world as 'present-at-hand', shorn of their practical significance. For example, Heidegger suggested that 'mass' acquires its sense as a physical property by gradual abstraction from the practical contexts in which things are 'too heavy' or 'too light'. This changeover from a practical to a theoretical understanding can supposedly be initiated whenever equipment is broken, missing or in the way. What can eventually result is an altogether different sense of what it means to be: beings are understood within a mathematized theoretical framework rather than local contexts of practical activity. Heidegger thought that the

intelligibility of this theoretical projection of nature nevertheless presupposed an understanding of the world as practical setting.

This attempt to understand science as decontextualizing things and overlooking their practical significance confronts serious difficulties. First, Heidegger never convincingly described the *genesis* of this ontological shift. The momentary breaking of practical absorption when things malfunction normally results in re-engagement with the equally practical tasks of repair, replacement or removal. Heidegger does not indicate *how* such practical problem-solving yields to decontextualized theorizing. Moreover, Heidegger does not describe how the practical tasks of science (experiment, instrumental manipulation, theoretical problem-solving and calculation) are connected to the disclosure of things as present-at-hand (see Experiment; Scientific method). Philosophers who emphasize these practical dimensions of science may well find a more adequate account of scientific practice in Heidegger's overall account of everyday practical understanding and interpretation, than in his initial understanding of science as abstracting from practical engagement with local equipmental complexes.

2 Science in later Heidegger

Heidegger's philosophical development in the 1930s had striking consequences for his understanding of science. In that later understanding science ceases to be an intelligible possibility whenever one disengages from practical absorption in the world and frames things theoretically. Instead, science makes sense only within the modern age. Science is not opposed to everyday practice, but rather is taken to manifest more clearly what is happening throughout the modern world. This distinctively modern phenomenon of science is still understood as a theoretical (and ultimately mathematical) projection of nature which permits diverse changing phenomena to be understood as instantiations of necessary unchanging laws. But instead of being accomplished all at once by disengagement from practical absorption in the world, this theoretical projection is itself situated within the ongoing activity of scientific research. Research aims to extend the domain within which facts can be exhibited as regulated by law, by working on and with the laws already articulated. This elucidation of scientific laws both extends their scope and justifies earlier laws by displaying their fruitfulness (see Laws, natural §1).

What is distinctive about modern scientific experimentation for Heidegger is its guidance by and conformity to this theoretical projection of nature as lawful:

> experiment begins with the laying down of a law as a basis. To set up an experiment means to represent or conceive the conditions under which a specific series of motions can be made susceptible of being followed in its necessary progression.
>
> (Heidegger [1952] 1977: 121)

Research and experimentation continually and characteristically open up new possibilities for research. The achievements of scientific research are then adapted as resources for further research. Thus for Heidegger science aims at the continual *expansion* of research as ongoing activity; its continuing successful extension to new domains legitimates its prior achievements.

Heidegger connected the expansion of science's capacities for theoretical calculation with the development of machine technologies. Science does not just use machines to produce, measure and interpret phenomena. Heidegger thought that the *theoretical* framework of modern science projected Nature from the start as calculable and manipulable: ' modern science's way of representing pursues and entraps nature as a calculable coherence of forces' (Heidegger [1954] 1977: 21).

Nature is thus manifest neither as Greek *physis*, nor as independent substances, nor as a significant equipmental complex, but instead as a stockpile of interchangeable resources (for example, of energy, chemicals, genes, labour or population).

Heidegger also connected the development of scientific research and machine technology, and the interpretation of Nature as 'stock', to the metaphysical conception of 'world pictures'. The scientific concern to represent the world objectively requires the right *approach*, so it can show itself unchanged by inquiry. Objectivity thus requires the right method, theories or attitudes, in short, the right kind of *subjectivity*. This shift from correct appearance to correct viewing in turn suggests the possibility of alternative views or pictures of the world. What Heidegger found questionable, however, is the underlying metaphysics of the world as depicted by a subject. Heidegger thought that

understanding ourselves as subjects representing and manipulating objects is not itself a picture to accept or reject, or a position to take up, but a disturbing and questionable historical situation we find ourselves in.

Heidegger's questioning of the metaphysics of the world picture was not intended to develop an alternative metaphysics. Heidegger accepted that naturalized metaphysics and epistemology are the appropriate culmination of the philosophical tradition stemming from classical antiquity. He nevertheless asked whether there might still be a way of thinking apart from scientific or philosophical metaphysics. Such thinking would not determine what there is or what we should do, although he thought it might manifest what is at stake in our situation. Heidegger's remarks about thinking are enigmatic. Many commentators emphasize his association of thinking with the poetry of Hölderlin or the fragments of the pre-Socratics. Others argue that his later thinking attempted to purify transcendental inquiry of any metaphysical implications. On either reading, Heidegger associated such thinking with the possibility of salvation from the nihilism he ascribed to the limitless expansion of scientific and technological control. To realize these possibilities willfully, however, would self-defeatingly try to overcome willfulness. Hence the indirection and apparent quietism of Heidegger's later reflections upon science, technology and the modern world.

3 Contemporary issues in Heideggerian philosophy of science

Heidegger's work encourages attention to the local, circumspective understanding of the world manifest in scientific experiment and instrumental work, without reducing that understanding to explicit theoretical representations. Yet Heidegger also emphasizes the theory-ladeness of science. One possible reading of this latter point suggests continuity with early postpositivist emphases upon theory: all understanding and observation are guided by specific theoretical preconceptions (see Observation). Another reading, not incompatible with the first, highlights a specific presumption within scientific practices, that a calculable order in nature is more fully realizable by an interplay between theoretical modelling and the experimental and technological reconstruction of the world.

Heidegger's work challenges both scientific realism and its empiricist and constructivist critics, by questioning the representationalist conception of scientific knowledge that both presuppose (see Scientific realism and antirealism; Constructivism). Perhaps science and technology are better understood as practical engagement with and intervention in the world, rather than systematic representation of its underlying causal structures or their empirical or social manifestations. To some philosophers this Heideggerian perspective suggests a modest and pluralist realism, for which science accurately describes real causal powers in the world without precluding other true descriptions for different purposes. Others take it to undercut the questions that realism and antirealism alike attempt to answer, and shift attention to the cultural and political significance of scientific engagement with the world.

Heidegger also encourages a new perspective on epistemological naturalism (see Naturalized epistemology; Naturalized philosophy of science). Later Heidegger entertains no doubts of science being understood biologically, psychologically or sociologically; he also accepts that such perspectives offer the best answer to traditional epistemological and metaphysical questions about the sciences. He nevertheless wants to formulate another kind of question that naturalized philosophy of science perhaps cannot address, concerning what is at issue and what is at stake in the historical development of scientific practice.

Heidegger's attempt to move beyond naturalized philosophy of science has encouraged some reconsideration of the relation between the natural and the human sciences (see Unity of science). Heidegger's account of understanding and interpretation rejects traditional hermeneuticist defences of the methodological uniqueness of the human sciences, since he takes the hermeneutical circle to characterize all understanding, and not just the interpretation of texts and human action. Some philosophers, however, have used Heidegger's work to distinguish human science from natural science on different grounds. They draw upon Heidegger's account of human existence (*Dasein*) either to argue for methodological constraints upon adequate interpretations of the meaningful world of human interaction, or to argue that a theoretically stable or predictively successful human science is impossible. Others have argued that such adaptations of Heidegger commit a paralogism: they interpret transcendental conditions upon the possibility of interpretation as objective characteristics of interpreters. If this argument is right, Heidegger's hermeneutics imposes no distinctive methodological constraints upon the human sciences. Heideggerian philosophy of science may then suggest, however, that the significance of scientific

interpretation of nature cannot be disentangled from what is at stake (culturally or politically) in specific historical practices.

The transition in Heidegger's thinking about science contributes to the historicizing of the philosophy of science. His concern is not to display the essential structure of science as a form of knowledge or inquiry, but instead to inquire into the historical emergence and future prospects of a particular set of practices whose goals and achievements are at issue in their ongoing development. Heidegger's account of science thereby contributes to the philosophical discourse of modernity, that is, reflection upon how our 'modern' world differs from its predecessors. In Heidegger's own thinking, this interpretation of science sustained a reactionary anti-modernism. Heidegger's work has nevertheless influenced philosophers who take quite different philosophical and political stances toward characteristic features of the modern world.

The connection between Heidegger's reflection upon science and his criticisms of modernity nevertheless raise disturbing questions about the relation between philosophy and political life. The shift in Heidegger's thinking which identified science and technology as essential phenomena of the modern age coincided with his overt identification with National Socialism in Germany in the 1930s. Heidegger's influence upon subsequent philosophy of science can certainly be detached from his personal philosophical allegiance to the Nazis even if his own philosophical project cannot. What cannot be so readily set aside, even today, is the urgency of critical reflection upon the political uses of philosophical work and the dangers as well as hopes attached to politically engaged philosophical reflection upon the modern world.

See also: Heidegger, M.; Pragmatism

JOSEPH ROUSE

References and further reading

Heidegger's works involve extensive technical terminology that present considerable additional difficulties in translation. All other works listed are straightforwardly readable and clearly explicate the technical terms used.

Blattner, W. (1995) 'Decontextualization, Standardization, and Deweyan Science', *Man and World* 28: 321-39. (Assesses different readings of Heidegger's early philosophy of science, and develops parallels to Dewey's pragmatism.)

Caputo, J. (1986) 'Heidegger's Philosophy of Science: The Two Essences of Science', in J. Margolis, M. Krausz and R. Burian (eds) *Rationality, Relativism and the Human Sciences*, Dordrect: Nijhoff, 43-60.(Argues for Heidegger's early philosophy of science as better suited for what is valuable in his later critique of modernity.)

Dreyfus, H. (1986) 'Why Studies of Human Capacities Modeled on Ideal Natural Science Can Never Achieve their Goal', in J. Margolis, M. Krausz and R. Burian (eds) *Rationality, Relativism and the Human Sciences*, Dordrect: Nijhoff, 3-22. (Develops a Heideggerian distinction between natural and human sciences.)

Dreyfus, H. (1991) 'Heidegger's Hermeneutic Realism', in D. Hiley, J. Bohman and R. Shusterman (eds) *The Interpretive Turn*, Ithaca, NY: Cornell University Press. (Heidegger as a realist about science.)

Heidegger, M. (1927) 'Sein und Zeit', *Jahrbuch für Philosophie und phänomenologische Forschung* 8: 1-438; trans. J. Stambaugh, *Being and Time*, Albany, NY: State University of New York Press, 1996.(Heidegger's first and most important work, and principal presentation of his early philosophy of science.)

Heidegger, M. (1952) 'Die Zeit des Weltbildes', in *Holzwege*, Frankfurt: Vittorio Klostermann; trans. W. Lovitt, 'The Age of the World Picture' in *The Question Concerning Technology and Other Essays*, New York: Harper & Row, 1977.(Most detailed discussion of Heidegger's later philosophy of science.)

Heidegger, M. (1954) 'Die Frage nach der Technik', in *Vorträge und Aufsätze*, Pfullingen: Gunther Neske; trans. W. Lovitt, 'The Question Concerning Technology', in *The Question Concerning Technology and Other Essays*, New York: Harper & Row, 1977, 3-35.(Heidegger's most extended later discussion of the relation between science and technology.)

Ihde, D. (1991) *Instrumental Realism*, Bloomington, IN: Indiana University Press. (Develops a Heideggerian interpretation of science emphasizing instrumentation and the relation of science to technology; usefully connects Heidegger to more recent philosophy of science.)

Rouse, J. (1985) 'Heidegger's Later Philosophy of Science', *Southern Journal of Philosophy* 23: 75-92. (Expansion of material of §§1-2 of this entry.)

Rouse, J. (1987) *Knowledge and Power*, Ithaca, NY: Cornell University Press.(Draws extensively upon Heidegger to emphasize the practical and political aspects of science, while opposing both realism and antirealism.)

Zimmerman, M. (1990) *Heidegger's Confrontation with Modernity*, Bloomington, IN: Indiana University Press. (Critically situates Heidegger's discussion of science and technology within his criticism of modernity and his political engagement with National Socialism.)

Heisenberg, Werner (1901-76)

One of the most outstanding of twentieth-century physicists, Werner Heisenberg is famous for the uncertainty, or indeterminacy, principle of quantum mechanics, widely interpreted as implying an irreducibly indeterministic conception of nature. The main proponent of the Copenhagen interpretation after Bohr, Heisenberg conceived of the quantum description as referring not to objective spacetime realities, but merely to the probable outcomes of measurements. Heisenberg's philosophy, containing contradictory positivistic and realistic strands, is best understood in the context of his creative scientific theorizing.

In 1925 Heisenberg developed the foundations for quantum mechanics, proclaiming that a consistent atomic theory should contain only relations between experimentally observable quantities. Heisenberg replaced such 'unobservables' as electrons' positions by an ordered set of data related to the 'directly observable' frequencies and intensities of emitted radiation. An analysis of Heisenberg's work reveals that he used this positivistic principle not as a heuristic guide, but as an a posteriori justification of the successful technical method that *de facto* eliminated electronic positions. Heisenberg's approach, rather than being positivist, is akin to what he defined as 'practical realism'. A 'practical realist' does not resign from the quest for the underlying reality, yet dispenses with those realistic notions which cease to be theoretically fruitful.

In 1927, in an attempt to provide an 'intuitive' interpretation of the abstract quantum formalism, Heisenberg reversed his opinion about the unobservability of electrons' positions. By declaring that the meaning of concepts is identical with the procedure of their measurement, Heisenberg demonstrated agreement between the operational definition of spacetime concepts in thought experiments and the mathematically derived uncertainty formulas. Heisenberg concluded that the precise simultaneous specification of canonically conjugate variables, such as positions and momenta, is impossible in principle. Rather, they are subject to uncertainty relations: $\Delta p \Delta q \sim h$, where Δp, Δq are indeterminacies in the position and momentum measurements, and *h* is Planck's constant. A similar relation, $\Delta E \Delta t \sim h$, holds between energy and time. These formulas imply that in the atomic domain, as opposed to the classical one, the spacetime and energy-momentum descriptions are mutually exclusive, or 'complementary'.

Heisenberg's operational approach in the deduction of the uncertainty principle is often interpreted as a direct continuation of his earlier positivistic stand. Yet we witness here a fruitful epistemological about-face: if in 1925 the directly observable data determined the structure of the theory, in 1927 the theory dictated the possibilities of observation. In 1925 Heisenberg eliminated classical spacetime concepts; in 1927 he resurrected them, restricting their simultaneous applicability. In both cases Heisenberg aimed to secure the acceptance of an unconventional and abstract theory. This strategy of persuasion, which elevated local epistemological moves into overarching principles of knowledge, is a source of confusion in interpretations of Heisenberg's philosophy.

Heisenberg explored the philosophical implications of the uncertainty principle in relation to two central issues: 'indeterminism' and 'objective reality'. Heisenberg declared that the uncertainty principle, necessarily introducing probability into the definition of a system's state, implies a conclusive renunciation of causality. In principle, we cannot know the present with enough precision in order to predict the future with certainty.

Heisenberg's initial explorations of uncertainty contain ambiguities. He did not make clear whether uncertainty signified an epistemological limitation on measurability, or is of an ontological nature - the later Copenhagen stand. Heisenberg did not carefully distinguish between acausality and indeterminism (see Determinism and indeterminism §2), nor did he provide cogent reasons why the uncertainty principle necessarily applies to an individual system, rather than to the statistics of an ensemble of similar systems - a less radical interpretive option.

Another implication of uncertainty concerns the physicist's notion of reality. The probability function does not describe what happens between observations in classical visualizable terms. The physical description, Heisenberg maintained, is no longer about the objective course of nature. Rather than describing 'nature in itself', physicists only specify nature's responses to questions put in experimental set-ups.

Despite positivistic overtones, Heisenberg's position is not identical with Bohr's prohibitions on inquiring about

the reality ‘behind the phenomena’ (see Bohr, N.). Heisenberg describes his position as a neo-Platonic one, which identifies genuine reality with the underlying mathematical forms. Heisenberg’s position that the ‘primal’ reality is mental is revealed by his introduction of a conscious observer into physical description. According to Heisenberg, it is the experiment that ’forces’ an atom to indicate a definite property. While the transition from the ‘possible’ to the ‘actual’ occurs during measurements, the discontinuous change in the system’s probability function takes place only when a conscious observer registers measurement results (see Quantum measurement problem).

Even though Heisenberg and Bohr often presented a united front against the opponents of the Copenhagen interpretation, they differed on many crucial issues. Their disagreement about Bohr’s doctrine of the indispensability of classical concepts - the heart of complementarity philosophy - is a most notable example. Contrary to Bohr, Heisenberg did not believe that a consistent interpretation of the microdomain is possible by using only classical concepts, without recourse to the quantum formalism. Heisenberg’s endorsement of Bohr’s arguments alongside his own dissenting views, resulted in numerous contradictions in Heisenberg’s philosophical writings.

Heisenberg’s philosophy was not formed in an attempt to erect a consistent structure. It was developed in numerous talks to different audiences in changing sociopolitical circumstances. Brilliant and intellectually fearless, Heisenberg was not constrained by logical inconsistencies. Pauli, Heisenberg’s closest collaborator, criticized Heisenberg at the beginning of his scientific career for being ‘very unphilosophical’ (Hermann *et al.* 1979). Many years later, Pauli singled out Heisenberg’s philosophical opportunism as a major source of his astounding scientific creativity.

See also: Logical positivism §4; Operationalism; Quantum mechanics, interpretations of

MARA BELLER

List of works

Heisenberg, W. (1930) *Die Physikalischen Prinzipien der Quantentheorie (Physical Principles of the Quantum Theory)*, Leipzig: Hirzel; trans. C. Eckart and F.C. Hoyt, Chicago, IL: University of Chicago Press, 1930.(A presentation of the foundations of the quantum theory and its interpretation in the ‘Copenhagen Spirit’.)

Heisenberg, W. (1952) *Philosophic Problems of Nuclear Science*, New York: Pantheon Books.(A collection of Heisenberg’s lectures in the 1930s and 1940s, including those to audiences in the occupied territories during the Second World War.)

Heisenberg, W. (1958) *Physics and Philosophy: The Revolution in Modern Science*, New York: Harper & Row. (A collection of philosophical essays which covers Heisenberg’s main interpretive contributions.)

Heisenberg, W. (1969) *Der Teil and das Ganze: Gespräche im Umkreis der Atomphysik (Physics and Beyond: Encounters and Beyond)*, Munich: Piper; trans. A.J. Pomerans, New York: Harper & Row, 1971. (Heisenberg’s memoirs, presented in the form of dialogues, describing the development of atomic physics and its philosophical interpretations.)

Heisenberg, W. (1984-9) *Gesammelte Werke (Collected Works. Series A, B and C)*, ed. W. Blum, H.-P. Dürr and H. Rechenberg, Munich: Piper, and Berlin: Springer.(Contains all Heisenberg’s published works, with series A devoted to original scientific publications, series B containing review articles, lectures and books, and series C containing philosophical and popular writings.)

References and further reading

Beller, M. (1983) ’Matrix Theory Before Schrodinger: Philosophy Problems, Consequences’, *Isis* 74: 469-91.(An analysis of Heisenberg’s original understanding of the new quantum theory, with an emphasis on the elimiantion of spacetime and its justification by the principle of elimination of unobservables.)

Cassidy, D.C. (1992) *Uncertainty. The Life and Work of Werner Heisenberg*, New York: Freeman.(The only comprehensive biography of Heisenberg, this book traces Heisenberg’s intellectual and personal life in the changing sociopolitical context.)

Hermann, A., Meyenn, K.v. and Weisskopf, V.F. (eds) (1979) *Wolfgang Pauli: Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg u.a, Band I: 1919-1929 (Scientific Correspondence with Bohr, Einstein, Heisenberg and others. Volume I: 1919-1929)*, Berlin: Springer.(Contains Heisenberg’s letters to Pauli during crucial phases of the creation and interpretation of quantum mechanics.)

Home, D. and Whitaker, M.A.B. (1992) ’Ensemble Interpretations of Quantum Mechanics. A Modern Perspective’, *Physics Reports* 210: 224-317.(A comprehensive review and re-evaluation of ensemble

interpretations of quantum physics as compared to the orthodox version of Heisenberg and Bohr.)

Jammer, M. (1974) *The Philosophy of Quantum Mechanics. The Interpretation of Quantum Mechanics in Historical Perspective*, New York: Wiley.(Contains an analysis of Heisenberg's uncertainty paper, as well as numerous discussions on the relevance of the uncertainty principle to different interpretations of quantum physics.).

Stapp, H.P. (1972) 'The Copenhagen Interpretation', *American Journal of Physics* 40: 1098-116.(This attempt of a systematic presentation of philisophy of Heisenberg and Bohr is more faithful to Heisenberg's thought than it is to Bohr's.)

Hell

The ancient idea that the dead go to a dark subterranean place gradually evolved into the notion of divinely instituted separate postmortem destinies for the wicked and the righteous. If the former lies behind the Psalms, the latter version appears in apocalyptic works, both canonical and deutero- or non-canonical, and is presupposed by numerous passages in the New Testament. Through the patristic and medieval periods the doctrine gradually achieved ecclesiastical definition, stipulating eternal torment (both physical and spiritual) in a distinctive place for those who die in a state of mortal sin. Most reformers recognized biblical authority for this doctrine. Philosophically, the notion of postmortem survival raises many questions in the philosophy of mind about personal identity. Recent discussion, however, has concentrated on the specialized version of the problem of evil to which the doctrine gives rise.

1 The problem of hell

The doctrine of hell gives rise to a specialized version of the problem of evil because it can be argued that the propositions:

(I) God exists and is essentially omnipotent, omniscient and perfectly good; and
(II) some created persons will be consigned to hell forever,

are logically incompossible, as follows:

(1) if God existed and were omnipotent, he would be able to avoid (II);
(2) if God existed and were omniscient, he would know how to avoid (II);
(3) if God existed and were perfectly good, he would want to avoid (II);
(4) therefore if (I), not (II).

Premise (1) seems obvious because an omnipotent creator could falsify (II) either by altogether refraining from making persons, or by annihilating created persons any time he chose. Further, some traditional theologians (for example, Augustine, Duns Scotus, William of Ockham and Calvin) take divine sovereignty over creation - both nature and soteriology - to imply divine freedom with respect to soteriological policy. For example, God could have legislated annihilation at death for human persons, or temporary purgatory followed by everlasting utopia for all sinners. Premise (2) follows because an omniscient God would know the scope of his powers. Premise (3) rests on the twin intuitions that a perfectly good God would want to be good to any persons he created, whereas hell means decisive ruin for its inhabitants.

The most straightforward solution is to reject either (I) or (II). The problem is that each is deeply entrenched in Christian belief. The Bible represents Christ as teaching that the disobedient and unfaithful will be ‘cast into outer darkness’, where there is ‘weeping and gnashing of teeth’ (Matthew 13: 42, 50; 22: 13), or thrown into the ‘unquenchable fire’ that is ‘prepared for the devil and all his angels’ (Matthew 13: 42, 50; 18: 8-9; 22: 13; see also 3: 10). If faith in Christ implies a belief that the teachings of Christ are true, then such texts seem *prima facie* to commit Christians to (II). Hermeneutical ingenuity will be required to reject (II) without eroding the epistemological foundation for faith in Christ itself. A minority of universalists (who believe that everyone will be saved) and annihilationists (who hold that the wicked will be annihilated at death or after a finite period of postmortem suffering) have shouldered that burden and declared (II) false.

2 Varieties of goodness

The argument in (1)-(4) above will convict Christians of contradiction, however, only if they understand the attribute terms and the doctrine of hell in something like the above-mentioned ways. In fact, Scripture and tradition contain a range of interpretations of divine goodness (see Goodness, perfect §3) and specify a variety of reasons why God might not falsify (II).

Global approaches. Some say that divine goodness finds its expression in the world as a whole, that universal republic for the good of which the wellbeing of individuals may be sacrificed. Thus, some medieval and Reformation theologians affirm that - whether because of the ontological incommensurability of the infinite

relative to the finite, or because God is creator and governor of all else - God has no obligations to creatures and so cannot be deemed to be unjust to them, no matter what he does. Whether in fact or by natural necessity, God orders the world as a whole to advertise his goodness and manifest his glory. Given the Fall, Aquinas explains, God predestines some to show his mercy and damns others to declare his justice. It is certainly epistemically possible that a maximally perfect ordering of the world as a whole should include hell for some created persons.

Metaphysical goodness. Christian Platonist that he is, Augustine assumes that to be is to be good, because it is either to be God (who is goodness itself) or to be somehow God-like. He concludes not only that God is good to any creature simply by conferring the gift of existence, along with other natural endowments, but also (in apparent contradiction of Matthew 26: 24) that the value of such goods to their created possessor trumps any disvalue constituted by deprivations of its wellbeing. Thus, even the damned in hell have reason to give thanks and praise to God!

Conditional goodness to created persons. More intuitive, because more concrete, is the notion that divine goodness to created persons involves God's guaranteeing to each that their life is a great good to them on the whole and in the end. Yet many defenders of the doctrine of hell (Swinburne 1983; Stump 1986; Craig 1989; Walls 1992) insist that taking created freedom seriously forces the qualification 'except through some fault of their own' - which is enough to falsify (3) in the argument in §1.

3 Free-will defences of hell

Since much biblical talk of heaven and hell functions to make ante-mortem created free choices seem momentous, many defend the logical compossibility of (I) and (II) on the basis of the following assumptions:

(A1) created free will is a very great good, whether intrinsically or as a necessary means to God's central purposes in creation;
(A2) God cannot fulfil his purposes for and with free creatures without accepting the possibility that (II) some will qualify for damnation.

(A1) is variously explained in terms of alleged divine desires to confer on created persons the dignity of self-determination with respect to their own (eternal) destinies, or to enter into mutual relations of beatific intimacy with created persons in which their 'yes' to God is their own. (A2) rests on multiple philosophically controversial assumptions. First, it presupposes that created freedom is incompatible with divine determinism; otherwise God could avoid damning creatures simply by causing them to make choices that qualify them for heaven. Second, even granting libertarian or incompatibilist freedom, (A2) holds only if either God lacks middle knowledge (see Providence §3) of what free creatures would do in which circumstances prior in the order of explanation to the divine decision to actualize them in those or other circumstances, or it is logically possible that some creatures are 'transworld damned', in the sense that no matter what circumstances God actualized them in, and no matter what helps of grace compossible with their freedom God furnished, their choices would qualify them for damnation. For if God had such middle knowledge and no creatures suffered from transworld damnation, God could falsify (II) by solving the coordination problem and actualizing created persons only in those circumstances in which none would qualify for hell. Suárez (§1), who is an advocate of divine middle knowledge, affirms the necessity of divine resourcefulness - necessarily, for any possible person and any situation in which it can exist, there are some helps of grace that would (should God supply them) win the creature over without compromising its incompatibilist freedom - and so would pronounce transworld damnation impossible and (A2) false.

Third, (A2) assumes that, given sufficiently sinful exercise of created freedom, God does not have the option of waiving the consequence of damnation. For if (as Duns Scotus and William of Ockham suppose) worthiness of damnation were a statutory policy created by free and contingent divine legislation, it would be within God's power to spare all incompatibilist-free creatures no matter what they did. Even if hell were understood not as external and arbitrarily contrived punishment, but as the intrinsic and natural consequence of ante-mortem sinful choices, divine omnipotence would be as able miraculously to obstruct them as to prevent fire from burning the three young men in the furnace (Daniel 3: 13-20). To defend the impossibility claim in (A2), one would have to maintain that God necessarily owes it to himself to punish such sinners in one of those ways. Otherwise additional reasons will be required to explain God's free and contingent choice.

Contemporary free-will defenders of the doctrine of hell (notably Swinburne, Stump, Craig and Walls) meet this demand by reasserting (A1), the intrinsic value of created freedom and/or the importance of its unfettered exercise to God's purposes. Postmortem miracles transforming the sinner's character to make beatific intimacy with God possible and desirable would rob the creature of the dignity of self-determination with respect to its relationships and destiny. Although the creature would remain metaphysically the same person, the miraculous disruption would produce such discontinuities that it would be difficult for the post-conversion self to identify with the projects and purposes of the ante-mortem character. God pays his respects to created personality precisely by not effecting such miraculous changes, but regretfully giving the creature over to the consequences of its choices and allowing it to remain the person it has become.

By contrast, medieval theologians refused to hold God to a postmortem non-interference policy. They reasoned that if rational creatures were as free to accept or reject God after death as before, the blessed might fall again and the damned might repent. To prevent both, God was said to confirm them in their choices by making it causally impossible for them to deviate from their 'deathbed' orientations. With the elect, the value of divine interference was glossed in terms of a traditional equivocation which saw freedom on the one hand as a self-determining power for opposites, and on the other as an orientation towards intrinsic value for its own sake (see Anselm of Canterbury §6). Ante-mortem careers in which created freedom imitated God with respect to self-determination were brought to rest in a higher freedom that mimicked God's stable orientation to the good. Applying - as they did not - their estimate of the value of libertarian freedom to the fate of the damned, it might seem that if hell is bad enough, the failed effort 'to do it oneself' - like a child's attempt to put a toy aeroplane together - might come to rest in assisted success. Moreover, the transformation could be gradual, carefully engineered by God to approximate the continuity with difference found in dramatic ante-mortem conversions.

Whether divine non-interference constitutes respect for created agency depends on the strength of human agency and the perniciousness of hell. Toning down vivid biblical imagery, some (notably Swinburne and Stump) contend that hell does not rob life of all positive meaning because abandonment to the natural consequences of ante-mortem choices is compatible with many and diverse satisfactions - some perverse (for example, rebellion for its own sake) and all minuscule (for example, debating with Milton's fallen angels the compossibility of divine foreknowledge and created free will) in comparison to the joys of divine company. Others (for example M.M. Adams) reason that since humans were made to find ultimate satisfaction in God, persistent turning away from God, cemented in vice, will eventually unravel the personality into madness, the pain of which is only faintly anticipated by ante-mortem schizophrenia and depression. Could God really respect our dignity by standing idle while we degrade ourselves? Yet if God intervenes to put a floor under our fall, why not to turn us - by a continuity-preserving process - to himself?

A sober look at the reality of human life raises the question whether the vast majority of human beings were ever up to the challenge of deciding their own eternal destinies, of taking it upon themselves to give a final 'yes' or 'no' to God. Humans begin life ignorant, weak and helpless, psychologically so lacking in self-concept as to be incapable of choice. We learn to 'construct' a picture of the world, ourselves and other people only with difficulty over a long period of time and under the extensive influence of non-ideal choosers. We emerge from the long, messy developmental process with entrenched neuroses that colour our cognitive as well as affective 'takes' on the world and thus impair our freedom, while we unconsciously 'act out' doing harm to self and others. Our grasp of hell's grim torments, as of heaven's bliss, is radically inadequate. Even if we could experience either for a short time (for example, burning fire, deep depression, consuming hatred; St Teresa's joyful glimpse of the Godhead), we are unavoidably unable to experience their cumulative effect in advance, and so unable more than superficially to appreciate what is involved in either. The apparent conclusion is that human agents are unavoidably unable to exercise their free choice with fully open eyes, the way free-will defenders sometimes suggest.

The evaluation of ante-mortem human agency is something about which Christians have disagreed for centuries. What may be the majority report tries to reconcile empirical realities with its free-will defence of the doctrine of hell by appealing to the doctrine of Adam's Fall. Adam and Eve, the primordial human couple, had the robust agency required to make the momentous choice. Our damaged agency is an inherited consequence of their sin (see Sin §2). Nevertheless, whether because of a doctrine of collective responsibility (according to which it is just that all members of a family participate in the punitive consequences and/or the guilt of the deeds of its head) or because the damages are alleged not to be disabling, such impairments do not excuse us from the task or the

consequences of making our own momentous decision.

4 Pragmatic arguments

Closely coupled with free-will defences of the doctrine of hell is the pragmatic or moral argument that the threat of hell is a powerful incentive to religious fidelity and moral diligence. Kantians, who may themselves believe in some sort of postmortem retribution, counter that the motive of punishment avoidance undercuts the moral enterprise, which enjoins duty for duty's sake. Moralists of other stripes allow threats a place in the early stages of moral pedagogy or grant moral legitimacy to the self-concerned desire to be rightly related to God, the good, the moral law, and so on. Still others charge that pragmatic arguments cut both ways: the notion that the governor of the universe will consign sinners to the eternal consequences of their choices might just as easily produce rebellion (as, for example, in the case of J.S. Mill), or combine with a pessimistic appraisal of human nature to foster despair!

5 Alternative destinies

While many take biblical apocalyptic language at face value, others see the rhetorical context of exhortation and poetic imagery as opening hermeneutical loopholes that leave the fate of persistent sinners scripturally underdetermined. Positions can be seen to vary along the following five dimensions: the degree of divine initiative involved in setting up and consigning people to hell; whether death is the 'deadline' at which qualifications are assessed, or whether there is a temporary or indefinite period of postmortem opportunity; the duration of hell - whether it begins immediately after death or only after the Last Judgment, and whether it lasts forever or whether souls consigned there eventually wither away; whether the suffering is psychospiritual only or also physical; and whether the fate is irreversible.

See also: Eschatology; Evil, problem of; Heaven; Limbo; Predestination; Purgatory; Reprobation; Resurrection; Soul, nature and immortality of the

MARILYN McCORD ADAMS

References and further reading

Adams, M.M. (1993) 'The Problem of Hell: A Problem of Evil for Christians', in E. Stump (ed.) *Reasoned Faith, A Festschrift for Norman Kretzmann*, Ithaca, NY: Cornell University Press, 301-27.(Defends universalism in preference to doctrines of hell or annihilation.)

Craig, W.L. (1989) 'No Other Name: A Middle Knowledge Perspective on the Exclusivity of Salvation through Christ', *Faith and Philosophy* 6 (2): 172-88.(Uses middle knowledge and the hypothesis of transworld damnation to mount a free-will defence of the doctrine of hell.)

Hick, J. (1966) *Evil and the God of Love*, San Francisco: Harper & Row, 2nd edn, 1978, 87-9, 91-2, 107-14, 177, 341-5.(Argues against traditional doctrines of hell in favour of indefinite postmortem opportunities and eventual universal salvation.)

Stump, E. (1986) 'Dante's Hell, Aquinas' Theory of Morality, and the Love of God', *Canadian Journal of Philosophy* 16: 181-98.(Develops a free-will defence of Dante's hell and defends divine goodness in setting it up.)

Swinburne, R. (1983) 'A Theodicy of Heaven and Hell', in A. Freddoso (ed.) *The Existence and Nature of God*, Notre Dame, IN: University of Notre Dame Press, 37-54.(Offers a free-will defence of annihilation or benevolent alternative environments for those who die with characters unfit for heaven.)

Walls, J. (1992) *Hell: The Logic of Damnation*, Notre Dame, IN: University of Notre Dame Press.(Offers a free-will defence of a 'Methodist' hell modified by the hypothesis that God gives disadvantaged agents a postmortem opportunity to make a decisive choice.)

Hellenistic medical epistemology

During the Hellenistic period (323-31 BC), there arose, largely in Alexandria, a profound debate in medical methodology. The main participants were the Empiricists, committed to an anti-theoretical, practical medicine based on observation and experience and the various Rationalists, such as Herophilus, Erasistratus, and Asclepiades, who held that general theories of physiology and pathology were both attainable and essential to proper medical understanding and practice. Dispute about the nature of scientific inference and the status of causal explanation mirrored and to some extent conditioned the contemporary debate between Stoics and sceptics about epistemology.

1 Rationalists and Empiricists

In the medicine of the post-Classical period, the empiricism already discernible among the Hippocratic doctors (see Hippocratic medicine) gathers strength. Diocles of Carystus (*fl. c.*350 BC) denied that explanation was in all cases possible or even desirable; some things simply have to be accepted as though they were first principles, while experience is the best guide to therapy. The main function of theoretical explanations (in terms of the balance of humours, or whatever) is simply to convince the patient of the doctor's competence.

This empirical approach was taken up by Herophilus (*fl. c.*260 BC), who held that 'the appearances should be stated first even if they are not first' (fr. 50a), and that causes should be accepted 'only hypothetically' (fr. 58). He devised a set of dilemmatic arguments against the existence of causes (fr. 59), for example: causes must be either (a) corporeal of corporeal effects; (b) incorporeal of incorporeal effects; (c) corporeal of incorporeal effects; or (d) incorporeal of corporeal effects; but none of (a)-(d) is possible; hence there are no causes. He performed dissections and vivisections (perhaps even on human subjects: frs 63-74), discovering the distinction between motor and sensory nerves, and described four basic 'powers' responsible for human physiological functioning (frs 131, 184).

His contemporary, Erasistratus, another brilliant anatomist, doubted whether so-called antecedent causes (such as the excessive exposure to sunlight which presages fever) were properly causes at all, since they are not invariably followed by their supposed effects (a view countered by Galen (§3). While paying lip-service to teleology, he developed a severely mechanistic physiology, rejecting specific organic 'faculties' (for example, of attraction for urine by the bladder) to explain particular functions, in favour of a fluid dynamics based on the principle of *horror vacui* (see Strato §3).

Philinus (*fl. c.*250 BC) and Serapion (*fl. c.*225 BC), a pupil of Herophilus, adopted this cautious, empirical attitude and gave it a full-bloodedly sceptical twist, founding the school of medical Empiricism, the fullest account of which is given in Galen's *Outline of Empiricism*. They held that all theoretical knowledge (of the micro-structures of the body, and of causal powers) was unattainable, and in any case unnecessary for successful medical practice, which required only the accumulation of large amounts of empirical data of regular connections between phenomenally observable types of events, in regard both to the prognosis of the course of a disease and its therapy (*Outline of Empiricism* 2). Such evidence was to be gained primarily by personal observation, supplemented by second-hand reports whose reliability was itself subject to empirical testing (*Outline of Empiricism* 3, 8).

Possible remedies might be suggested by inspiration or luck, or the observation of a fortuitously therapeutic set of circumstances (*Outline of Empiricism* 2). Such initially successful therapies must then be subjected to further empirical testing, for 'experience' (a suitably large set of such congruent concatenations) to produce a 'theorem' (an empirically-based therapeutic generalization, for example, 'pomegranates cure diarrhoea'). A third, controversial, feature of their method was 'transition to the similar', a form of analogical reasoning which some of the Empiricists allowed to suggest therapies for hitherto unobserved conditions (*Outline of Empiricism* 4): thus, confronted with an unprecedented ankle-ailment, noting that ankles resemble wrists, one might apply some previously tested wrist-remedy.

The status and applicability of transition was much disputed. It was not emphasized by the earliest Empiricists, and their successors disagreed as to whether it was a proper part of medical practice (hence requiring theoretical justification), or whether doctors simply found themselves using it (*Outline of Empiricism* 4). Some later

Empiricists, notably Cassius 'the Sceptic', refused to allow transition any role at all, and tried to re-establish Empiricism in its original pristine form, unsullied by any such Rationalist accretions.

'Rationalist' here is an umbrella-term for anyone committed to any theoretical physiology and pathology; thus Herophilus, Erasistratus, the Pneumatists (who held that the condition of the *pneuma*, modified inspired air (see Pneuma), was fundamental to health and disease), and the followers of Asclepiades (second century BC), who held that disease was caused 'by the lodgement of theoretical particles in the theoretical pores' (Sextus Empiricus, *Against the Professors* VIII 220) and developed a theory of the internal fluid dynamics of the body on the basis of the principle of 'movement towards the rarefied' (compare Erasistratus' *horror vacui*), are all Rationalists. However disparate their views, they all believed in the possibility of inferring from surface phenomena to the underlying structural reality of things, by a process known as *analogismos*. By contrast, the Empiricists restrict themselves to *epilogismos*, inference from one evident phenomenon to another with which it is regularly associated (*Against the Professors* VIII 7; compare Sceptic and dogmatist on signs: see Pyrrhonism §4).

Galen wrote an early treatise *On Medical Experience* presenting the anti-Empiricist arguments of Asclepiades and the replies of Menodotus (*fl. c.* AD 150), the most sophisticated Empiricist of his day. Asclepiades argues that Empiricists cannot avoid theoretical commitment: their syndromes are supposed to consist of concatenations of phenomena that have been observed to occur similarly in many cases. But what is similarity? Everything both resembles and is distinct from everything else in innumerable ways. What is required is some concept of *relevant* similarity: but that is already a theoretically-laden notion (*On Medical Experience* 4-6). Moreover, how many cases are enough? The Rationalist deploys a Sorites argument (see Vagueness §2) to undermine the coherence of such a concept (*On Medical Experience* 7).

Both charges are answered. First, the Empiricist requires no *theory* of relevant similarity; he is simply guided by the way things appear. He need not say *why* things appear to be alike - they just do. And experience itself shows that this is the way to acquire practical abilities (*On Medical Experience* 9-11). Moreover, although the Rationalists claim to be able to infer from the phenomena to the hidden natures of things, they disagree radically among themselves as to what those natures are - and these disagreements are, by their very nature, undecidable (*On Medical Experience* 12-13). Finally, there is no determinate answer to the question how many cases are enough: that varies from case to case, doctor to doctor. If serious, the Sorites is quite general and will destroy the coherence of innumerable everyday concepts (*On Medical Experience* 16-17, 21). But this is of no concern to the Empiricist, who has no time for such theoretical niceties.

2 Methodists

In the first century AD there arose a third medical sect, that of the Methodists. They held that all disease was reducible to either excessive constriction or excessive dilation of the pores of the body, or to some combination of these conditions. These 'communalities' are directly evident to suitably-trained observation, and require no theoretical inferences. Thus the Methodists agreed with the Empiricists in eschewing theory, but parted company with them over the issue of causes. Rationalists were united by the belief that proper physiology and pathology (and hence proper therapy) demanded a theoretical understanding of the hidden causal mechanisms of nature. Empiricists rejected such speculation about the hidden structure of things as being both inherently ungroundable and therapeutically useless, but were prepared to allow into their syndromes reference to items which the Rationalists at least would describe as antecedent, evident causes. Methodists rejected even appeal to antecedent causes, holding that all that mattered was the current condition of the patient's body, a condition whose aetiology was quite irrelevant.

R.J. HANKINSON

References and further reading

Deichgräber, K. (1965) *Die griechische Empirikerschule: Sammlung der Fragmente und Darstellung der Lehre (The Greek Empiricist School: Collection of Fragments and Teachings of the School)*, 2nd edn, Berlin: Weidmann.(Fragments of the Empiricist school.)

Edelstein, L. (1967) *Ancient Medicine*, Baltimore, MD: Johns Hopkins University Press.(Fine collection of papers, several of which are relevant to the themes of this article.)

Frede, M. (1987) *Essays in Ancient Philosophy*, Oxford: Oxford University Press. (Contains several important

essays on medical epistemology.)
Galen (*c.* AD 150-90) *On the Sects*, *Outline of Empiricism*, and *On Medical Experience*, trans. M. Frede, *Galen: Three Treatises on the Nature of Science*, Indianapolis, IN: Hackett Publishing Company, 1985.(Includes an excellent introduction.)
Hankinson, R.J. (1987) 'Causes and Empiricism', *Phronesis* 32: 329-48.(Analysis of the Empiricist attitude to antecedent causes.)
Herophilus (*fl. c.*260 BC) Fragments, in H. von Staden, *Herophilus: The Art of Medicine in Early Alexandria*, Cambridge: Cambridge University Press, 1989.(Fine edition - from which the fragments of Herophilus are cited - with translation and discussion.)
Vallance, J. (1990) *The Lost Theory of Asclepiades of Bithynia*, Oxford: Oxford University Press.(Good account of Asclepiades' medical system.)

Hellenistic philosophy

The Hellenistic schools dominated the Greco-Roman world from c.*300* BC *to the mid first century* BC*, making it an era of great philosophical brilliance. The principal doctrinal philosophies were Stoicism and Epicureanism, but this was also the age in which scepticism emerged as a philosophical movement. The central issues of debate were the nature and origin of the world, the means to attaining truth and the ethical goal.*

The Hellenistic age is defined negatively, as the interregnum between two empires. It starts with the death of Alexander the Great in 323 BC and the collapse of his empire, and ends with the battle of Actium in 31 BC, the official start of the Roman Empire. Its main positive feature is the ‘Hellenization’ of - that is, the spread of Greek culture to - much of the Mediterranean world, especially the East. A high proportion of the converts to the Hellenistic philosophical schools flocked to Athens from the cities of western Asia, and a sense of excitement about Greek culture was one factor in the emergence of new philosophical schools. A favoured alternative explanation is that the collapse of the old Greek *polis*, or city state, had left people feeling lost and in need of philosophical succour, but it is doubtful whether that hypothesis, even if sustainable, could account for more than a fraction of the new converts.

Although Alexandria eclipsed it as a centre of learning, Athens remained the headquarters of philosophy almost throughout the era. Hence in many ways Hellenistic philosophy would better be dated from 306 BC, when Epicurus founded his school in Athens, to 88 BC, when the Athenian schools were badly weakened in war, never fully to recover their institutional supremacy. The succeeding era was marked by a philosophical diaspora throughout the Roman Empire.

A Hellenistic doctrinal philosophy was a complete system, which had to offer: (1) an understanding of the world’s origins, components and organization, and of our place in it; (2) a methodology of discovery, which included in particular naming one or more ‘criteria of truth’; (3) an account of what the ‘goal’, happiness (see Eudaimonia), consists in. These three areas correspond to what had become the standard tripartition of philosophy into (1) physics, (2) logic and (3) ethics.

The main two doctrinal Hellenistic philosophies can be studied in the entries on Epicureanism and Stoicism. The sceptical philosophers of the New Academy constituted a third main movement, which provided constant critiques of its doctrinal rivals (see Arcesilaus; Carneades). Pyrrhonist Scepticism made a brief appearance at the beginning of the era (see Pyrrho; Timon), but did not become a formal movement until its refoundation by Aenesidemus towards the end of it (see Aenesidemus; Pyrrhonism). Nevertheless, Pyrrhonism is commonly counted as a Hellenistic philosophy.

‘Hellenistic’ is in any case not a purely chronological label. Aristotelianism survived in some form throughout the period (see Theophrastus; Strato; Peripatetics), but did not really become distinctively Hellenistic. The philosophy of the period is sometimes called *post*-Aristotelian. However, although Aristotle did happen to die within a year of Alexander the Great, this label is misleading: it is unclear how far the philosophers of the period recognized Aristotle as a major philosopher to whom they were obliged to respond. Aristotle’s real revival occurred only at the very end of the Hellenistic age.

On the other hand, the main spokesman of Pyrrhonism today is Sextus Empiricus, who is often therefore treated as a Hellenistic thinker, despite his probable date in the second century AD. Likewise, Stoics of the first two centuries AD are commonly thought of as Hellenistic (see Seneca; Musonius Rufus; Epictetus; Marcus Aurelius).

Plato remained a philosophical influence throughout the era, but there was no genuine doctrinal Platonism between *c*.265 and *c*.50 BC. In the early first century BC Antiochus advertised a return to Platonism, but his philosophy is in fact characteristically Hellenistic in its terminology and concepts.

Apart from some short writings of Epicurus, no Hellenistic philosophical treatises have survived intact. It has been the task of modern scholarship to reconstruct the systems and debates from secondary sources. The most important of these are the Latin philosophical dialogues of Cicero, and the third-century AD *Lives of the Philosophers* by Diogenes Laertius.

See also: Academy; Ariston of Chios; Chrysippus; Cleanthes; Hellenistic medical epistemology; Lucretius; Panaetius; Philo of Larissa; Philodemus; Posidonius; Zeno of Citium

DAVID SEDLEY

References and further reading

Annas, J. (1992) *Hellenistic Philosophy of Mind*, Berkeley, CA: University of California Press.(The best study of the topic.)

Annas, J. (1993) *The Morality of Happiness*, Oxford: Oxford University Press.(Major study of issues surrounding eudaimonism in Aristotle and Hellenistic philosophy.)

Barnes, J., Brunschwig, J., Burnyeat M. and Schofield, M. (eds) (1982) *Science and Speculation*, Cambridge: Cambridge University Press.(Collection of articles on Hellenistic philosophy of science.)

Barnes, J. and Mignucci, M. (eds) (1988) *Matter and Metaphysics*, Naples: Bibliopolis.(Collection of articles on Hellenistic metaphysics.)

Brunschwig, J. (1994) *Papers in Hellenistic Philosophy*, Cambridge: Cambridge University Press.(Magisterial anthology of articles.)

Brunschwig, J. and Nussbaum, M.(eds) (1993) *Passions and Perceptions*, Cambridge: Cambridge University Press.(Collection of articles on Hellenistic psychology.)

Cicero (mid 45 BC) *On the Ends of Good and Evil*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1914.(A dialogue in which Hellenistic ethical theories are defended and criticized; includes Latin text.)

Cicero (late 45 BC) *On the Nature of the Gods*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press, and London: Heinemann, 1933. (A dialogue in which Hellenistic theologies are defended and criticized; includes Latin text.)

Cicero (mid 44 BC) *On Fate*, trans. R.W. Sharples, Warminster: Aris & Phillips, 1991. (A classic source on the Hellenistic debate about determinism; includes Latin text.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(Covers the lives and work of most major Hellenistic philosophers.)

Flashar, H. (ed.) (1994) *Die Philosophie der Antike, 4: Die hellenistische Philosophie (Philosophy in Antiquity, 4: Hellenistic Philosophy)* Basle: Schwabe.(Monumental survey, with comprehensive bibliography.)

Frede, M. (1987) *Essays in Ancient Philosophy*, Oxford: Oxford University Press. (Contains several seminal papers on Stoicism and Scepticism.)

Griffin, M. and Barnes, J. (eds) (1989) *Philosophia Togata: Essays on Philosophy and Roman Society*, Oxford: Oxford University Press.(Series of studies on the place of Hellenistic philosophy in Roman life and thought.)

Inwood, B. and Gerson, L.P. (1988) *Hellenistic Philosophy, Introductory Readings*, Indianapolis, IN: Hackett Publishing Company.(Includes a large body of primary texts, in translation.)

Laks, A. and Schofield, M. (eds) (1995) *Justice and Generosity*, Cambridge: Cambridge University Press. (Collection of articles on Hellenistic political philosophy.)

Long, A.A. (1974) *Hellenistic Philosophy*, London: Duckworth.(Superbly accessible and engaging introductory account.)

Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press. (Volume 1 contains sources in translation, with commentary; volume 2 contains the original texts.)

Nussbaum, M. (1994) *The Therapy of Desire*, Princeton, NJ: Princeton University Press.(An eloquent comparison of the different Hellenistic schools' treatments of the emotions.)

Schofield, M., Burnyeat, M. and Barnes, J. (eds) (1980) *Doubt and Dogmatism*, Oxford: Oxford University Press.(Collection of articles on Hellenistic epistemology.)

Schofield, M. and Striker, G. (eds) (1986) *The Norms of Nature*, Cambridge: Cambridge University Press. (Collection of articles on Hellenistic ethics.)

Scott, D. (1995) *Recollection and Experience*, Cambridge: Cambridge University Press.(Contains invaluable analyses of Stoic and Epicurean cognitive psychology.)

Sedley, D. (1977) 'Diodorus Cronus and Hellenistic philosophy', *Proceedings of the Cambridge Philological Society* 203: 74-120.(On the dialectical background to the main Hellenistic philosophies.)

Sextus Empiricus (*c*. AD 200) *Against the Professors*, Books VII-XI trans. R.G. Bury, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1935-6.(Contains extensive comparative reports of Hellenistic philosophical positions and arguments.)

Sharples, R.W. (1996) *Stoics, Epicureans and Sceptics*, London: Routledge.(The best short introduction to Hellenistic philosophy.)

Striker, G. (1996) *Essays on Hellenistic Epistemology and Ethics*, Cambridge: Cambridge University Press.(A set of important studies.)

Helmholtz, Hermann von (1821-94)

In physiology, physics, mathematics, aesthetic theory and epistemology, Helmholtz intervened, and innovated. He contributed to the physiology of perception through work on the central nervous system, followed by work on optics and acoustics. He invented instruments, such as the opthalmoscope and introduced the mathematical principle of the conservation of energy to physics. For geometry, Helmholtz elaborated on the concept of an n*-dimensional manifold. He secured the influence of the 'Berlin physics', introduced Faraday and Maxwell to Germany, refined the theory of electrodynamics and reflected on the role of discrete entities in physics. Having become the most influential representative of German science and its uncontested spokesperson, he repeated the importance of the connection between education and research and the necessity not to separate the natural sciences (*Naturwissenschaften*) from the social sciences (*Geisteswissenschaften*). This monumental body of work is experiencing a revival of interest today as historians of both science and culture consider it in a new light. But the question remains of how to characterize what we might call 'the Helmholtz effect'* in philosophy. *Why was Helmholtz equally influential not only on Cassirer, Husserl, Schlick, Meyerson and Freud, but also on the principal founders of contemporary physics; Einstein, Bohr and Heisenberg? To grasp this, we must understand the constant interaction between science and philosophy which characterized, even permitted, the extraordinary developments in mathematics, physics and physiology in Germany at that time. Here the connection between Helmholtz and Kant is fundamental, since 'the Helmholtz effect' transformed the Kantian heritage. Helmholtz did not write a systematic philosophical work, but in redefining fundamental epistemological concepts and constructing a large part of the conceptual structure in which both philosophy and relativistic and quantum physics developed during the early twentieth century, he modified the very problems of epistemology.*

1 Epistemology after Kant

In its usual nineteenth-century German, philosophical sense, the term epistemology (*Erkenntnistheorie*) designates the theory of knowledge or cognition with a view to the determination of objectifying acts: how can the subject transform *phenomena*, which are given to it, into *objects* of cognition? In this sense epistemology, in the nineteenth century, appeared less as a particular doctrine, as it had been for Immanuel Kant, than as a problem: the problem of 'objectification', which the Kantian critique had recast at the centre of a philosophy founded on the concept of *Vorstellung* (representation). Epistemology was, as Harald Höffding repeated in 1925, the most exact part of philosophy, in contrast with the diverse forms of *Lebensphilosophie* (philosophy of life). But it was, for the followers of Kant, something not only to be repeated (against the idealism of Hegel or Schelling) but also something to be adapted (to the advancement of knowledge) and modified.

Helmholtz clearly indicated that the aim of all his work had been to elaborate a general conception of representation which would permit him to establish 'the sources and the degree of legitimacy of our cognitions', adding that no area of learning could ignore this question with impunity. The *Handbuch der physiologischen Optik (Treatise on Physiological Optics)* (1856, 1860, 1867) already stemmed from his ambition to develop the Kantian transcendental aesthetic in the domain of the natural sciences. The famous lecture of 1878, Die Tatsachen in der Wahrnehmung (Facts in Perception) made clear that the basic problem of the natural sciences is the same as that of epistemology: the problem of knowing in what sense our representations correspond to *das Wirkliche* (the actual real; that is to say, for Kant, that which accords with the material conditions of experience). And in the preface to the 1884 edition of his *Vorträge und Reden (Lectures and Talks)*, Helmholtz wrote: ' I was a committed Kantian at the beginning of my career and I still am'. To understand where our representations come from and how they are organized, Helmholtz wanted to continue Kant's work. But, in so doing, he radically transformed Kantianism.

2 Sensation/intuition, categories, schema/symbol: redefinitions

Perhaps nothing is more difficult to grasp in the history of philosophy than the operation which consists of preserving a problematic while displacing its centre (or its scaffolding), and of changing the concepts of a language without changing its words. It is exactly this operation which defines 'the Helmholtz effect'. To understand this, we will rely upon the internal architecture of Kantian critique by noting the three questions that Helmholtz found in it.

The first question concerns the apprehension of phenomena. How is that which is given in phenomenal

manifestations given to us? Kant had answered this question by distinguishing purely subjective conscious *Empfindung* (perception) from objectively conscious perception or intuition, and by defining pure intuition as the combination of two a priori forms of sensibility (which are thus anterior to all experience): space, the form of external intuitions, and time, the form of internal intuitions. Kant wished, above all, to limit dogmatic metaphysics and to do away with onto-theology; he did not aim to provide the foundation for a psychology and was thus uninterested in analysing sensations. Furthermore, he seemed to assume that the basic constitution of spatiality was correctly described by the axioms of Euclidean geometry; in this sense, whenever a phenomenon is given to us, it is on condition of its localization in ordinary space and time (see Kant, I. §5).

Now, by returning to the questions of sensation and intuition, Helmholtz was first led to recognize the existence of a process of objectification at the level of perception, and then radically to enlarge the concept of intuition. First, with regard to the 'analysis of sensations', until then overshadowed by the transcendental aesthetic, Helmholtz (following his first professor of physiology, J. Müller) showed that sensations are not reflections of real, objective properties of things, but depend essentially on the appropriate sensory receptor, the field of perception, and so on. Extended objects in space only appear to us, in reality, as 'enveloped' in sensations produced by our nervous system. In sensation, there is no relation of resemblance with things, but rather a construction of signs (for example, in depth perception). Correspondence with the external world becomes a matter of *learning* by accumulating 'unconscious inferences'. Second, concerning space and time, Helmholtz showed that what is at issue again are progressively constructed forms of experience. Proposing a new type of empiricism in philosophy of mathematics, he links the structure of spatial intuition to the experience of the free mobility of rigid bodies (this idea of the kinaesthetic constitution of the concept of space was taken up by Edmund Husserl) and he anchors the structure of temporal intuition in that of counting and in the concept of ordinal numbers. From that moment on, the forms of sensibility cease to be a priori, while Euclidean geometry and the axioms of arithmetic lose their transcendental status. Conversely, the concept of intuition must be expanded: for the construction of intuition, it suffices that the impressions of the senses, which may correspond to any mode of observation (including the imaginary), be specified in a determinate and nonambiguous manner. This expanded intuition, which Helmholtz sees as the basis of non-Euclidean geometries, paves the way for Husserlian 'categorial intuition'. In addition, we should note that, for their part, Niels Bohr and Werner Heisenberg adopt the redefinition literally when they come to describe the 'intuitive character' of the new quantum mechanics. In every case, we thus see in Helmholtz a radical displacement of Kantian concepts. For Kant, intuition was an immediate and a priori presentation of phenomena and a specification of the general concept of *Vorstellung* (representation). For Helmholtz, representations become modes of *acquired* presentation, through unconscious inferences (sensations) and systematic work (symbolic languages) at the same time. In addition, intuition, broadly conceived, appears as *interpretation*, which depends simultaneously on the natural environment, the subject's history, the historical condition of languages and forms of thought, and so on.

We can easily see how Helmholtz will transform the second question he takes from Kant: that of the pure concepts of the understanding, the Categories. For Kant, the basic cognitive process is objectification, and objectification takes place by subsuming our intuitions of phenomena under pure (a priori) concepts. The list of these concepts is closed and ahistoric, for the simple fact that it is derived, at least in principle, from general logic. In scientific cognition, the principles of causality (*Ursache und Wirkung*), of the permanence of substance and of reciprocal action are thus the constitutive principles which allow the transformation of phenomena into objects (see Kant, I. §6). Without going into the details, we can say that Helmholtz imposes upon these principles an essential limitation, while interpreting them as simple, *invariant*, methodological rules *and* introducing historicity into them. Helmholtz, in effect, replaces causality with the idea of *Gesetzlichkeit* (lawfulness), that is to say, the requirement that certain relations amongst certain processes have a nomological character (see Laws, natural §2). Lawfulness is, in fact, the only possible 'object' for cognition. In science, we do not know the imminent properties of real objects, but only the nomological regularity of recurring relations. There is no isomorphism between the external world and human cognition; simply put, cognition can only develop by presupposing the existence of a sort of parallelism of lawfulness (Heinrich Hertz develops this idea in his introduction to *Die Prinzipien der Mechanik (The Principles of Mechanics)* (1894), via the concept of a model). On the other hand, in place of the concept of substance, Helmholtz puts that of an invariant element. Most notably, he dissociates substantiality from materiality, which permits him to introduce in 1881 the 'elementary quantum of electricity' (the electron) as a nonmaterial, invariant quantity - a number, not a thing. What we know of the real is what we perceive as invariant,

in a manner which is, every time, historically determined, but without our ever being able to found science on a closed system of pure concepts.

Finally, the third 'Kantian question' which is fundamentally reformulated in Helmholtz concerns the *relation* between intuition and concept. For Kant, this relation was established in two different ways, depending on whether the subject matter was, on the one hand, science, or, on the other hand, art, language, religion or teleology. In scientific cognition (the domain of determinative judgments), what made it possible to link intuitions and concepts was the Schematism, the transcendental imagination's 'hidden work' (see Kant, I. §7). In the altogether different domain of reflective judgments - of taste (where the subject relates only to itself and where there cannot be any direct presentation of concepts in sensible intuition), only what Kant called the Symbolism was operative: an indirect mode of knowledge formation, which proceeds by analogy (see Kant, I. §12). This distinction shaped the entire subsequent history of the relationship between the exact and the human sciences during the nineteenth century in Germany (see Unity of science). It was, however, challenged by the increasingly open suggestion that, in science, as in art or language, there is only one single cognitive process. In this development, which leads from Wilhelm von Humboldt to Schlick, Cassirer and Wittgenstein, Helmholtz plays an essential role because he wants to provide a general theory of signs and symbols - that is, a general theory of cognition in which the Kantian distinction disappears. To be sure, he did no more than propose this: the hints we find in his various articles, and even the lecture of 1878, merely indicate the direction to be followed. But in 1910, Cassirer, in *Substance and Function*, claimed to have found in this theory of signs the idea that, in cognition, humans deal only with 'functional correspondences of one structure to another' - thus making Helmholtz the inspiration for the notion of 'symbolic forms'. And it is difficult not to see the influence of Helmholtz in the generalization of the concept of symbol for twentieth-century philosophers and scholars, from Husserl and Schlick to Bohr and Heisenberg - or even Granger and Goodman.

3 Conclusions

Helmholtz stands at a crossroads of research and ideas of diverse origin. He maintained, in a certain sense, the spirit of the Enlightenment in post-Hegelian Germany. But 'the Helmholtz effect' went beyond this. He touched upon the problems and the language of philosophy in a style both new and enduring, a style which has been clandestinely propagated well into the present. In his effort to reflect on each element of the Kantian account of cognition, Helmholtz accomplished, in effect, an unexpected dissolution of the Kantian critique. He analysed sensations, introduced the idea of unconscious memory and reduced intuition to its elementary processes: this reduction put the *body* in the foreground of the formation of all cognition and by so doing subjected geometry and arithmetic to the human *experience* of movement. Furthermore, Helmholtz expanded the concept of intuition and restricted that of causality by substituting for it the idea of lawfulness: this gives the formation of mathematical or physical concepts complete *freedom* with regard to the constraints of objectification via the tetralogy space-time-causality-continuity (which Bohr later used to define 'classical' physics). In addition, by generalizing the concept of symbol, Helmholtz abolished the fundamental Kantian distinction between science on the one hand and merely analogical knowledge on the other: this opened the way to a global conception of cognition as a construction of systems of signs. For those who read Helmholtz, cognition became a learning process anchored in bodily experience. At the same time, it was freed from the illusion in which concepts reflected stable and objective properties of things, and could thus turn towards the simple search for invariance. Finally, cognition emerged everywhere as a symbolic construction (sensations, language, sciences, art or religions), as a historically determined interpretation of that which is perceived as given for a particular epoch. Helmholtz not only brought the nineteenth century to a close, he set the stage for the twentieth in nearly every discipline.

See also: Geometry, philosophical issues in §4; Logical positivism §1; Neo-Kantianism; Science, 19th century philosophy of; Space; Time

Translated from the French by J. Maskit
CATHERINE CHEVALLEY

List of works

Helmholtz, H. von. (1847) *Über die Erhaltung der Kraft: eine physikalische Abhandlung*, Berlin: G. Reimer; trans. R. Taylor, *H. von Helmholtz, Scientific Memoirs*, London: Tyndall & Francis, 1853.(Introduction of the mathematical principle of conservation of energy.)

Helmholtz, H. von. (1856, 1860, 1867) *Handbuch der physiologischen Optik*, Leipzig: Leopold Voss, 3 vols; 3rd edn, 1909-11; trans. J.P.C. Southhall, *Helmholtz's Treatise on Physiological Optics*, New York: Optical Society of America, 1925; repr. New York: Dover, 1962. (The translation is of the 3rd edn of 1867.)

Helmholtz, H. von. (1863) *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, Brauschweig: Vieweg; trans. A. Ellis, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, London: Longmans, 1875, 1885; repr. New York: Dover, 1954.

Helmholtz, H. von. (1882-95) *Wissenschaftliche Abhandlungen*, Leipzig: J.A. Barth, 3 vols.(With a bibliography.)

Helmholtz, H. von. (1883) *Vorträge und Reden*, Braunschweig: Vieweg, 2 vols; 3rd edn, 1884; 5th edn, 1903. (See especially, on the philosophy of mathematics, 'Über die Tatsachen die der Geometrie zugrunde liegen', 1868, 'Über den Urspring und die Bedeutung der geometrischen Axiome', 1870 and 'Zählen und Messen', 1887; reprinted in M. Schlick and P. Hertz (eds) *Schriften zur Erkenntnistheorie*, Berlin: Springer, 1921; translation by M. Lowe *Epistemological Writings*, Dordrecht: Reidel, 1977. See also, on the general philosophy of knowledge, 'Die Tatsachen in der Wahrnehmung'.)

Kahl, R. (ed.) (1971) *Selected Writings of H. von Helmholtz*, Middletown, CT: Wesleyan University Press.(Recent compilation.)

References and further reading

Cahan, D. (ed.) (1993) *Hermann von Helmholtz and the Foundations of XIX Century Science*, Berkeley, CA: University of California Press.(Good synthesis of Helmholtz. Includes an overall bibliography of secondary literature.)

Koenigsberger, L. (1902-3) *Hermann von Helmholtz*, Braunschweig: F. Vieweg und Sohn, 3 vols; abridged English trans. F.A. Welby, Oxford, 1906; repr. New York, 1965.(The standard biography.)

Turner, S. (1970-80) 'Hermann von Helmholtz', in C. Gillispie (ed.) *Dictionary of Scientific Biography*, New York: Charles Scribner's Sons, 16 vols.(Short treatment of his life and scientific work.)

Helmont, Franciscus Mercurius van (1614-98)

Although he lived in the seventeenth century, van Helmont belongs more to late Renaissance than to modern intellectual culture. He was a larger-than-life figure who, in his prime, had an international reputation as an alchemist and a physician. His metaphysical interests came increasingly to the fore, however, and he became particularly associated with Kabbalistic doctrines. A friend of Locke and Henry More, he was also closely connected with Anne Conway and Leibniz, with whom he shared many intellectual affinities. It is these connections that make his philosophy - in particular, his theodicy and his monadology - of enduring interest.

Franciscus Mercurius van Helmont was born at Vilvoorde, near Brussels. His father, Jean-Baptiste van Helmont, was a famous Paracelsian medic who made a significant contribution to the history of chemistry, especially to the theory of gases. Like his father, Francis became famous as both doctor and alchemist. He too was harassed by the Inquisition, from whom he escaped only by giving up the estate he had inherited. Later, in the freer context of the Protestant Netherlands, he was able to publish many of his father's writings that had been previously suppressed. Indeed he became a publisher of many religiously heterodox works and, with characteristic courage, allowed his name to be associated with some whose actual author was anonymous. For this reason a number of works were mistakenly attributed to him. Even those of which he was the author were often produced by collaboration, or written up by others on the basis of conversations with him.

Van Helmont's first book was concerned with developing the idea of an alphabet of nature, a natural rather than a conventional language originally spoken by Adam. He made practical proposals arising from his theory for teaching those born deaf to speak and to understand speech, which some implemented with success. Over time, however, he became less interested in experiments and more metaphysical than his father had ever been, partly through the influence of Knorr von Rosenruth with whom he collaborated on a German translation of Boethius' *De Consolatione Philosophiae*. He also assisted him in collecting together the mostly Hebrew texts made available for the first time in the Latin *Kabbala denudata* in 1677 and 1684, to which he contributed writings which include his *Cabbalistical Dialogue*, a classic defence of Kabbalistic metaphysics.

Van Helmont was physician to Lady Anne Conway from 1670 until her death in 1679. They came to share not only a commitment to Kabbalistic metaphysics and Biblical interpretation but also an involvement with the Quaker movement. He saw to the publication of her Platonic/Kabbalist *Principles of the Most Ancient and Modern Philosophy*. Although outside the mainstream of Modern philosophy, he became a good friend of Locke whom he met at the Lantern Club in Rotterdam and with whom he spent the winter of 1693-4 during his last visit to England. Like van Helmont, Locke was also a physician, but the bond between them seems to have been a sceptical and liberal attitude concerning speculative matters of religious importance.

Of greater philosophical significance was his friendship with Leibniz, with whom he was associated in the early 1670s and again in the late 1690s. The extent to which they belonged to the same intellectual tradition is shown by their accounts of evil and of matter. For van Helmont and others there is a puzzle as to how God, a wholly perfect being, could have created pain and suffering, supposing these to be evils. There is also a problem of how, if God is pure spirit, he could have produced something as alien to himself as matter. Van Helmont's solution to the first problem lay in a form of optimism in which God is claimed to allow no more suffering than is necessary. As to belief in hell, all punishment is 'medicinal' and no soul is ultimately lost. His opposition to the arbitrary and vindictive god of some of the major Christian denominations is reflected in this view. He also followed the Lurianic Kabbalah in adopting the doctrine of the evolution of human souls through twelve lives, linked to belief in the eventual progress of every soul to eternal happiness. He managed to secure quite a following for this view among the English Quakers, until it was quashed in a revival of orthodoxy in the early 1690s.

Van Helmont's solution to the problem of the existence of matter is analogous to his treatment of the problem of evil, with the counterpart to optimism being a form of idealism, that is an affirmation that only spirit is ultimately real. Matter, being antithetical to God's nature, cannot result from that nature, and so is not so much a reality as a privation. A philosophy, such as Aristotelianism, that affirms the reality of matter must, he thought, deny the existence of a spirit creator. According to van Helmont's Neoplatonic-Kabbalistic theory, the monads emanate as pure spirits from the divinity. But these sparks of divinity eventually become 'dull' and 'sluggish', coalescing to

form what is called matter. Yet the monads, though they are in a degenerate state of ‘privation’, do not entirely lose their individual or spiritual character, which ‘*fundamentally* and *radically*’ they retain. Their material state is only transitory, according to van Helmont, and they will eventually return ‘to a more loose and free state’.

Leibniz’s solutions to these problems, though they have affinities, are more complex and less obviously open to objection (see Leibniz, G.W. §§2-6). Between the early 1670s, when van Helmont impressed Leibniz in long and deep discussions, and the 1690s, when he visited Hanover as an old man, Leibniz himself had changed. He had become too much of a Modern not to be embarrassed by van Helmont. The work published in his friend’s name often lacked rigour, sophistication and paid no attention to recent developments in the sciences. Van Helmont was fluent in Dutch and German but not in Latin or English, and so could only produce works in these languages by recruiting the support of people whose versions of his thought did not, in Leibniz’s view, do him justice. This may be partly why Leibniz himself was willing to help in the production of van Helmont’s last Kabbalistic work on the Book of Genesis, being entirely responsible for the Latin and even being to some extent a co-author. It was shortly after this collaboration that van Helmont returned to the Brabant to settle his affairs. After his death his niece turned to Leibniz for an epitaph, but appears to have ignored his plea to publish some of her uncle’s literary remains.

See also: Kabbalah

STUART BROWN

List of works

Helmont, F.M. van (1667) *Alphabeti vere naturalis Hebraici brevissima Delineatio*, Sultzbach.(Van Helmont’s theory of a natural ‘Adamic’ language.)

Helmont, F.M. van (1682) *A Cabbalistical Dialogue in Answer to the Opinion of a Learned Doctor in Philosophy and Theology that the World was Made of Nothing*, London.(English translation from the Latin edition of Part I of the *Kabbala denudata* which includes *A Rabbinical and Paraphrastical Exposition of Genesis I*. This is the most concise statement of van Helmont’s Kabbalistical theory of the origin of matter.)

Helmont, F.M. van (1684) *Two Hundred Queries Moderately Propounded concerning the Doctrine of the Revolution of Humane Souls and Its Conformity to the Truths of Christianity*, London.(The main account of van Helmont’s theory of transmigration.)

Helmont, F.M. van (1697) *Quaedam praemeditatae et consideratae cogitationes super quatuor priora capita libri primi Moysis*, Amsterdam; trans. as *Some Premeditate and Considerate Thoughts on the first four Chapters of the first book of Moses called Genesis*, London, 1701.(A Kabbalistic interpretation of the Biblical story of the creation.)

References and further reading

Brown, S. (1989) ‘Leibniz and More’s Cabbalistic Circle’, in S. Hutton (ed.) *Of Mysticism and Mechanism: Tercentenary Studies of Henry More (1614 -1687)*, Dordrecht: Nijhoff.(Develops affinities between Leibniz’s monadology and those of van Helmont and others.)

Brown, S. (1997) ‘F.M. van Helmont: his philosophical connections and the reception of his later Cabbalistic philosophy (1677-1699)’, in *Oxford Studies in Seventeenth Century Philosophy*, ed. M.A. Stewart, Oxford: Oxford University Press.(Includes accounts of van Helmont’s relationships with both Locke and Leibniz, as well of his extensive publications in English.)

Coudert, A. (1975) ‘A Cambridge Platonist’s Kabbalist Nightmare’, *Journal of the History of Ideas* 36: 633-52.(A good account of Henry More’s involvement with van Helmont.)

Coudert, A. (1978) ‘Some Theories of a Natural Language from the Renaissance to the Seventeenth Century’, *Studia Leibnitiana Sonderheft* 7: 56-114.(Focuses particularly on the context of van Helmont’s *Alphabet of Nature*.)

Coudert, A. (1995) *Leibniz and the Kabbalah*, Dordrecht: Kluwer.(Argues that van Helmont was an important influence on Leibniz.)

Help and beneficence

Which people are we morally required to help, and to what extent? In a world where the basic needs of many millions remain unmet, this is a philosophical question of great practical urgency. A minimal position is that while it is always praiseworthy to help someone, we are morally required to help only those to whom we stand in some special relation. In addition to the objection that it is too minimal, this view faces difficulties in accounting for emergency cases, in which one could, for example, save a stranger's life at little cost to oneself. More stringent views that place no restrictions on the range of people to be helped do not have these difficulties; they do, however, raise the intractable problem of how much we must sacrifice for the sake of others.

1 Principles of beneficence

A requirement to help or benefit a person to whom one stands in some special relation (such as a family member) is a 'special obligation'. A requirement to benefit people generally is a 'principle of beneficence'. Some principles of beneficence require us to promote other possible aspects of 'the good' in addition to human wellbeing, but our focus is on this central case. Further, as our topic is the general form principles of beneficence might take, we will not consider different possible accounts of wellbeing (see Eudaimonia; Happiness), though of course any actual principle of beneficence requires such an account. Nor will we discuss the problems that arise when a person who could genuinely be helped opposes the intervention (see Autonomy, ethical; Paternalism).

The idea of a general principle of beneficence, as distinct from special obligations, has roots in the West in Judaism and Stoicism, and was developed in Christianity in the form of the virtue of charity (see Charity). We should note here that possession of a virtue involves a certain motivational state, whereas following a duty or obligation need not (see Virtues and vices §§4-5; Duty §3). The scope of this entry, however, does not extend to the question of ethically appropriate motives, nor to the question of the 'natural' motivational sources of beneficent action (see Moral motivation).

Locke understood charity as requiring us to meet the serious needs of any person (1690); this is a more stringent requirement than the duty of beneficence Kant argued for (1797) (see §2). But the modern moral philosophers who take beneficence most seriously are the utilitarians: the whole of morality, on their view, consists in a requirement to promote - indeed to maximize - the good (see Utilitarianism). And the utilitarian tradition has no doubt influenced many non-utilitarians in their attitude to beneficence. It is notable, for example, that W.D. Ross (1930), who argued strongly against the reduction of all of morality to beneficence, accepted, as part of morality, a duty of beneficence of the same stringent maximizing form as the utilitarian principle.

Recognition of a principle of beneficence of some form continues to be common in moral philosophy in the late twentieth century, even among philosophers who reject the primacy of duty and obligation: 'virtue theorists' are likely to discuss a virtue of charity or benevolence (see Virtue ethics). There is, however, an important school of thought, with historical roots in parts of the social contract tradition, that relegates beneficence to the realm of the supererogatory (that is, morally good but not morally required) (see Supererogation §§4-6). This position is at its most plausible in the context of a liberal egalitarian theory of justice, such as that of John Rawls (1971) (see Rawls, J.). The idea is not merely that a just institutional scheme will remove the need for individual acts of beneficence, but that on a proper moral understanding, requirements of justice exhaust the normative domain claimed for requirements of beneficence. An important objection to this view is that it leaves it unclear what is required of a well-off individual in an unjust society; this problem is especially serious in the international sphere, where the prospect for institutions that promote distributive justice is remote. Advocates of the view could reply that unmet human needs do indeed raise an urgent moral issue, but that it is extravagant to respond to this issue by attributing responsibility for the wellbeing of all humanity to each moral agent (see §2). Rather, we should see the responsibility of individuals as mediated: they are required to support reform of the institutional structures that allow the needs to remain unmet.

There is a further difficulty for a moral account in which beneficence is supererogatory. While 'ordinary' moral thought is by no means settled on the issue of beneficence generally, it does seem to be so on the special case of easy assistance in emergency situations. That it would be wrong to walk away from a child drowning in a shallow pond seems so uncontroversial a claim that some philosophers, notably Peter Singer (1972), have seen in such

'rescue' cases firm intuitive foundations for an argument for a general principle of beneficence. This argument could be blocked by offering some principled distinction between rescue and non-rescue cases and by suggesting that the former generate some kind of special obligation. It is by no means clear, however, that a principled distinction is available.

2 The problem of demands

A principle of beneficence makes each of us responsible, in some sense and to some degree, for the wellbeing of everyone else. As we have noted, this can seem to be an extravagant position. The underlying worry here is simply that principles of beneficence require too much from individuals. Applied to utilitarianism, this 'over-demandingness objection' is long-standing, going back at least to Henry Sidgwick. But the issue of over-demandingness has received most attention since the 1970s, stimulated in part by a more general attack on modern moral theories that has centred around the work of Bernard Williams.

It might be thought that the cause of the extreme demands of the utilitarian principle is that it requires us to promote all aspects of the wellbeing of others, and that a more plausible principle of beneficence would require us to meet only the most basic of human needs, such as those for food, shelter, and health care. A principle of beneficence, however, that requires us to do whatever we can to meet even just these needs remains, in the circumstances of the late twentieth century, extremely demanding on individuals. The principle would, of course, be much less demanding if all or most people complied with it, but clearly we cannot assume ideal levels of compliance when assessing actual demands. (That the utilitarian principle is extremely demanding is sometimes denied on the grounds that an individual is rarely in a position to make great sacrifice that will be beneficial on the whole; this claim is typically based on scepticism about, for example, the effectiveness of humanitarian aid programmes.)

The real cause of the extreme demands of the utilitarian principle is its maximizing form: it requires individuals to promote the good up to the point where further sacrifice would burden them as much as it would benefit others. If extreme demands are objectionable, then a possible plausible principle of beneficence would seem to be one that requires sacrifice only up to a certain point. As an example there is the Judaeo-Christo-Islamic idea of giving up a 'tithe' or tenth of what one produces or earns. The problem, however, is to identify the criteria that would justify any particular limit to demands. Indeed it is on reflection difficult to accept that there could be a limit to required sacrifice that remains fixed whatever the changes in circumstances.

Other non-maximizing principles are possible. Thus Kant's duty of beneficence, on some interpretations, requires agents to adopt a 'maxim' of benefiting others, where this is compatible with acting to benefit others only rarely and in minor ways. For someone who takes beneficence seriously, this will be an unacceptably minimal principle. Likewise, the suggestion that no particular degree of beneficence is morally required, that all we need say is 'the more the better', will be seen as too close to the view that beneficence is supererogatory. A third view is motivated by a concern not with the extent of the demands of a maximizing principle, but with the way it imposes demands when not everyone is complying with it. As we have noted, a maximizing principle typically demands more of a complying agent as the number of other people complying decreases; this is to make agents responsible not just for their own 'share' of the demands of beneficence, but also for the shares of everyone else. If this is objectionable we could consider a principle of beneficence that requires agents to sacrifice only as much as it would be optimal for them to sacrifice under full compliance. Among other problems, it is not clear that this suggestion can account for rescue cases.

See also: Consequentialism; Egoism and altruism; Suffering; Vulnerability and finitude; Impartiality

LIAM B. MURPHY

References and further reading

Aiken, W. and La Follette, H. (eds) (1996) *World Hunger and Morality*, Upper Saddle River, NJ: Prentice Hall. (Useful introductory collection on the topic.)

Cohen, J. (1981) 'Who is Starving Whom?', *Theoria* 157: 65-81.(Defends a principle of beneficence requiring people to do only their 'fair share'.)

Kagan, S. (1989) *The Limits of Morality*, Oxford: Clarendon Press.(Defends a maximizing principle of beneficence.)

Kant, I. (1797) *Metaphysische Anfangsgründe der Tugendlehre*, trans. J.W. Ellington, *Metaphysical Principles of Virtue*, Indianapolis, IN: Hackett Publishing Company, 1964.(Advocates less stringent principle of beneficence.)

Locke, J. (1690) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1963. (Suggests we are required to meet the serious needs of any person.)

Murphy, L. (1993) 'The Demands of Beneficence', *Philosophy and Public Affairs* 22 (4): 267-92.(General discussion of the problem of demands and defence of a principle of beneficence requiring people to sacrifice only as much as it would be optimal to sacrifice under full compliance.)

Nagel, T. (1986) *The View From Nowhere*, New York: Oxford University Press, ch. 10.(Influential general discussion of the problem of demands.)

O'Neill, O. (1989) 'The Great Maxims of Justice and Charity', in *Constructions of Reason*, Cambridge: Cambridge University Press.(Defends a broadly Kantian approach, and includes a very helpful discussion of the liberal egalitarian interpretation of benficence.)

Paul, E.F., Miller, F.D. and Paul, J. (eds) (1987) *Beneficence, Philanthropy and the Public Good*, New York: Blackwell.(Collection devoted to beneficence, offering a variety of approaches.)

(eds) (1993) *Altruism*, New York: Cambridge University Press.(Collection featuring different perspectives on altruism.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Influential contemporary theory of justice.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Clarendon Press; repr., Indianapolis, IN: Hackett Publishing Company, 1988.(Advocates the principle of beneficence as one among others.)

Scheffler, S. (1992) *Human Morality*, New York: Oxford University Press.(Defends limits to required sacrifice.)

Singer, P. (1972) 'Famine, Affluence, and Morality', *Philosophy and Public Affairs* 1 (3): 229-43; repr. in W. Aiken and H. La Follette (eds) *World Hunger and Morality*, Upper Saddle River, NJ: Prentice Hall, 1996. (Influential article, claiming that morality requires us to sacrifice our interests for those in the developing world.)

Unger, P. (1996) *Living High and Leting Die*, New York: Oxford University Press. (Extensive discussion of the difficulty of distinguishing between rescue cases and non-rescue cases, and defence of a stringent requirement to help those in need.)

Williams, B. (1981) 'Persons, Character and Morality', in *Moral Luck: Philosophical Papers 1973-80*, Cambridge: Cambridge University Press.(Criticism of the strict impartiality of modern moral theory.)

Helvétius, Claude-Adrien (1715-71)

Helvétius was one of the most noteworthy and notorious figures of the French Enlightenment. In common with his fellow philosophes*, he asserted that all philosophical discussions should be based on the empiricism of Locke's Essay on Human Understanding (1689). But unlike Voltaire, d'Alembert, and the other members of 'the party of humanity', Helvétius took literally the notion that each person is a*tabula rasa *at birth - he boldly argued the case for unabashed environmental determinism. We are what our surroundings have made us, and nothing more.*

Immediately after Helvétius published De l'Esprit in 1758, the Catholic authorities cited his book as definitive proof that the philosophes *were out to destroy religion, throne, family, and all that is sacred. Only the struggle between court and parliament over control of censorship, along with his ties to Madame de Pompadour and the Duc de Choiseul, saved Helvétius. After suffering the indignity of three recantations, he decided upon posthumous publication of his second major work, De l'Homme (1773).*

Not a single philosophe *accepted Helvétius' view that the mind is a completely passive recipient of data received through the senses; nor did any of his comrades second his constantly reiterated claim that all sensibility may be reduced to physical sensations. Some privately expressed their exasperation that Helvétius published so much that seemed to vindicate every charge the Church lodged against them: that they were materialists, advocates of free love, and champions of a scandalous hedonism. Nevertheless at least a few of the* philosophes*, after setting aside the philosophical suppositions of De l'Esprit, came to appreciate that the larger concern of Helvétius was with their own search for the social and political preconditions of an independent intelligentsia, the would-be agents of Enlightenment.*

1 Philosophy

An empiricist, a materialist, and a utilitarian, Helvétius was in every respect a son of the French Enlightenment, yet his fellow *philosophes* remained wary of his writings, which they regarded as a *reductio ad absurdum* of their own.

When Helvétius began his studies with the announcement that all the faculties of the human mind - memory, imagination, judgment, reason - can be reduced to sensation, he undoubtedly believed he was merely repeating what Condillac had set forth in his celebrated *Treatise on Sensations* (1754). Neither Condillac nor any of the *philosophes* had anything but praise for Locke's attack on innate ideas (see Locke, J.). Where Locke went wrong, in Condillac's estimation, was in his retention of the notion of innate mental faculties. Conducting one of the most memorable thought-experiments of the century, Condillac slowly brought a hypothetical statue-man to life, first by endowing it with one sense after another, then by showing how one sense comes to the aid of another, until a being emerges whose mind possesses all the higher mental faculties.

To the *philosophes* the *Treatise on Sensations* was a method of research; it provided the means to disprove all notions of innate ideas, especially as used by the Church to place its views beyond the reach of criticism. Alone among their numbers, and much to their dismay, Helvétius transformed Condillac's method into a system of reductionist philosophy. No longer did the investigation end when painstaking analytical and genetic procedure slowly uncovered the hidden sensual roots of a given faculty of mind. Rather, the books of Helvétius begin with the dogmatic assertion that none of the faculties is anything more than the passive product of sense experience.

No support was forthcoming for the position staked out by Helvétius, not even - as the response of Diderot attests - from other materialists. Overwhelmingly in the eighteenth century, French materialists grounded their position in the newly emerging sciences of life. Sensitive matter, to Diderot, acts on its surroundings no less than the environment shapes matter. Where Helvétius denied the significance of organic constitution, Diderot believed that much of an individual's character is given from the beginning, and will ultimately win out regardless of environmental circumstances. To Helvétius a new environment makes a new person; to Diderot criminals cannot be rehabilitated because it is impossible to override heredity and physical organization.

In ethics Helvétius was a hedonist and a utilitarian, so it is understandable that Bentham regarded *De l'Esprit* and *De l'Homme* as forerunners of his work, all the more so since on more than one occasion Helvétius uttered words virtually identical to Bentham's formula of 'the greatest good of the greatest number'. But had Bentham taken a

closer look, he might well have rejected the notion that Helvétius was his predecessor; for hedonism as understood by Bentham had only the slightest connection with sexual liberation and was in every respect the opposite of a heroic ethic. How shocked, then, Bentham would have been to realize that the primary objective of Helvétius was to recreate the heroic values of antiquity in the modern world, a goal the Frenchman sought to pursue through the lure of sexual rewards.

Surprisingly, it was to ancient Sparta that Helvétius looked for a model of a sexually liberated social existence. Previously Sparta had been regarded as the home of a repressive civic virtue, a city whose citizens were forced to be free, forced to live in accordance with the dictates of their ‘higher selves’. Self-denial and self-overcoming, a constant and painful effort to put the public good above private interest - these themes, long associated with Sparta, were sometimes admired by the *philosophes* but always rejected, because the virtue of Spartans sounded too much like the monkish virtue they despised. Altogether different was the Sparta depicted by Helvétius. Drawing upon Plutarch’s life of Lycurgus, Helvétius conjured up a Sparta in which men eagerly engaged in noble deeds for their country because the greatest citizens were granted the sexual favours of the most beautiful women. ‘Lycurgus made love one of the principal springs of legislation’ (1758, II: 15); he understood that it is great passions that lead to great actions, and was wise enough to stir up grand emotions through the custom of having naked young women dance in front of youthful soldiers, praising the brave men and shaming the cowards.

All the *philosophes* agreed that the reverse side of denying original sin was an affirmation of the joys of sexuality. They also concurred that virtue and self-interest should not be set in opposition to one another; their shared view was that society should be so ordered that everyone has an interest in acting virtuously. However, it was one thing for Diderot to praise the free sexuality of Tahitians in an unpublished essay and quite another for Helvétius to publish and sign his name to an equally audacious proposal, and to imply, unlike Diderot, that his findings were directly applicable to the Europe of his day. The official Diderot, the public spokesman for the cause of enlightenment, was the author of *The Natural Son* (1757) and *The Father of the Family* (17??), two plays that endorsed conventional familial ideals in language that could not have been more exclamatory.

Diderot complained that Helvétius spent so much time attempting to prove that his kennelman, if placed in the proper environment, could have written *De l’Esprit*; he gasped in disbelief when he saw that Helvétius, in reducing everything to physical sensibility, was forever trying to explain the accomplishments of a genius in terms of copulation and defecation. Still, that did not prevent Diderot from placing *De l’Esprit* ‘among the great books of the century’.

What was it that Diderot and some of the other *philosophes*, for all their misgivings, admired in the writings of Helvétius? Above all, they gained a political education from his works; his conviction, that only through better legislation would humans ever have an interest in being virtuous, became theirs as well. ‘Morality is only a frivolous science unless blended with politics and legislation’, wrote Helvétius (1758, II: 15), ‘from which I conclude that, to be useful to the world, philosophers must consider objects from the viewpoint of the legislator’. Originally Diderot and many of the *philosophes* showed very little interest in forms of government; it was from Helvétius, a political thinker from the outset, that they learned how intimately their concerns about literature and the arts were tied to questions about politics.

2 Politics and the Arts

Both *De l’Esprit* and *De l’Homme* are primarily studies of the social situation of the *gens de lettres*, the intellectuals, under different political regimes. The maturity or childishness of the audience, its willingness or refusal to be instructed as well as entertained, the popularity of certain literary genres and the irrelevance of others, the inspiration or desperation of the writer - these matters and more hinge on the type of political regime that rules a country, argued Helvétius.

Helvétius adapted his study of the links between politics and the arts from Montesquieu. But before taking anything from the *Spirit of the Laws* (1748), Helvétius deleted Montesquieu’s chapters on climate. Why rule out in advance the possibility that the peoples of some parts of the world can ever hope to live under better conditions, inquired Helvétius, when political and social explanations suffice to account for their present predicament?

Although he rejected climate as a causal explanation, Helvétius kept intact the entirety of Montesquieu’s typology of political regimes, the division of governments into feudal monarchies, Oriental despotisms, and republics

ancient and modern. Sparta, as we have seen, was the ancient republic most frequently cited by Helvétius; England, called by Montesquieu a ‘republic hiding under the form of a monarchy’, was the country Helvétius constantly alluded to when he wished to draw a contrast between the monarchy France was and the republic he desired it to be.

The constant complaint of Helvétius was that ‘our [French] mores and the form of our government do not permit us to deliver ourselves to strong passions’. Under monarchy, petty intrigue at court to enhance one’s reputation takes the place in politics that under a republic is filled by the ambition of citizens to win fame for doing great deeds for their country. Wherever monarchy is triumphant, there are subjects rather than citizens, and the socially best-placed of these subjects care only for their personal advantage and that of their family name. A good aristocratic father in France will use all his influence to secure a public office for his incompetent son. How different was the world of republican Rome wherein Brutus did not hesitate to sacrifice his sons for the sake of preserving the public good.

From top to bottom of the social scale Helvétius found nothing to admire and much to condemn in the France of his day. So downtrodden were the peasants, so dehumanized were they by the brutality of the nobles, that Helvétius deemed the life of savages preferable to that of the simple folk living in the French countryside. Nor, for that matter, was the existence of Parisians as admirable as foreign visitors were wont to believe. Beneath the glamour and brilliance of operas, dramas, and salons lay a disturbing human reality. To be successful in social life, Helvétius noted, a man must have a pliable character that assumes as many shapes as the number of mansions he visits. Perhaps it is no accident that it was shortly after the publication of *De l’Esprit* that Diderot penned a memorable depiction of a ‘man without character’, who had ‘no greater opposite than himself’, in *Rameau’s Nephew* (1762-74).

The novelty of Helvétius lies in his efforts to apply Montesquieu’s sociopolitical models to the study of literature and the arts. Each type of political regime, the despotic, the monarchical, and the republican, shapes culture in its own image, Helvétius believed. Montesquieu had hinted as much: he had gone so far as to suggest that satirical writings cutting the powerful down to size thrive in England because in that nation society no longer revolves around feudal privilege and legally sanctioned class hierarchy. It was left for Helvétius to convert Montesquieu’s passing suggestions into a systematic treatise on politics and the arts.

About Oriental despotism Helvétius had relatively little to say. He was willing, however, to risk a few statements on the subject of politics and the arts in the non-Western world. It was his contention that Oriental authors, if they ever told the truth, had to present their thoughts in coded form. ‘Under submission to arbitrary power,… it is certain that writers must insensibly contract the habit of thinking allegorically’ (1758, III: 29). Since the historians of despotic countries, unlike the poets, cannot hide behind a veil of allusions and symbols, their account of the past is inevitably a pack of tricks the living play on the dead.

For the most part Helvétius concentrates his energies on drawing a series of sharp contrasts between the vitality of the arts and letters in republican England and the waning of literary glory in monarchical France. Living in a nation that is not politically free, looking up to *grands* who are idle, spoiled, and vain, the French are ‘the most gallant, the most loveable, but the most frivolous people of Europe’ (1758, II: 20). Boileau, repeating Horace, had indicated that the calling of the writer is to instruct as well as to please; Helvétius, however, complained bitterly that the French, ‘by the form of our government, have less need of instruction than of amusement’ (1758, II: 20). Love affairs, flirtations, coquetry, changes of fashion in clothing, and other private matters, none capable of stirring great passions, are the concerns of monarchical subjects. One trivial preoccupation supplants another with remarkable rapidity in France because persons confined to the pettiness of private lives are readily bored. For years the explicit objective of the *philosophes* had been to mould the new phenomenon they referred to as ‘public opinion’; Helvétius’ response was to point out that a public exists only where there is a republic.

‘In London it is a merit to be instructed; in Paris it is ridiculous’ (1758, II: 20). Inevitably, then, the English writer is inspired, the French writer diminished, by the audience.

> In a free state a man conceives the highest thoughts and can express them as vividly as they enter his mind. Such is not the case in monarchical states: in these countries the interest of certain corporations, that of various powerful individuals, and most of all a false and small politics, thwart the *élans* of genius.

(*De l'Esprit*, IV, ch. 4)

Helvétius was typically French in his belief that drama is the highest of the arts, and he was typical of his century in his conviction that the theatre of his day was inferior to that of the *grand siècle*. But he sounded a new note when he offered a political explanation of the decline of French drama. In republican England, he suggested, the grandeur of the tragic genre still holds sway; by contrast, in monarchical France the pettiness of comedy dominates the stage. Tragedies similar to those written by Corneille during a period of sedition and grand passions continue to be well received in England; but in France, beginning with Racine, the corrupted audience has become as indifferent to uplifting public themes as it is eager for love stories.

> I say that in every country where the inhabitants have no part in the management of public affairs, where the words *patrie* and *citoyen* are rarely cited, one does not please the public except in representing on the stage passions agreeable to [private] individuals, such as those of love.

(*De l'Esprit*, II, ch. 19)

Romantic love makes us small; love of country enhances our stature, and there is no better way to promote great civic passion than through offering sexual rewards which satisfy the cravings of physical sensibility. Greek tragedies were as replete with civic lessons as they were devoid of the motif of romantic love. Modern playwrights may yet return to the model set by Sophocles, provided modern legislators precede them in copying the political strategies of Lycurgus. In the posthumous *De l'Homme* Helvétius proposed to convert France into a federation of thirty republics, each animated by civic passion.

3 Conclusion

One of the favourite topics of the *philosophes* was the question of how the intellectuals could become independent and influential voices for enlightenment in a social world based on privilege, wealth and patronage. It was Helvétius who convinced at least a few of the *philosophes*, especially those who remained outside the academies, that only a political solution would suffice. In a civic society writers will be inspired by an enlightened audience and rewarded for their creative efforts. Until the dawn of the new era, the best book will be the one that champions the republican cause.

Evidence of what some *philosophes* eventually borrowed from Helvétius, as well as what they chose to repudiate, may be found in the *Système social* of the Baron d'Holbach, published fifteen years after *De l'Esprit* and at virtually the same time as *De l'Homme*. Almost word for word Holbach repeated the arguments of Helvétius against dramas revolving around the theme of romantic love and in favour of tragedies modelled on those of ancient Greece. Again echoing Helvétius, Holbach complained that the French were a frivolous people, the women especially because they had been miseducated by erotic paintings and literature, in consequence of which the favourite pastime of the *grands* was adultery.

When it came to women Helvétius and Holbach could not have been more similar in their diagnosis, nor more different in their solutions. Holbach advocated that women withdraw from high society to the privacy of their families. Motherhood, fidelity, and a restoration of traditional familial virtues was his message. Only if women accept the sanctity of marriage will men recover from their socially induced corruption. Helvétius, in dramatic contrast, would abolish the remnants of feminine modesty so that 'the favours of women, becoming more common, will appear less precious' (1758, II: 20). Long ago Plato had made a similar suggestion, and Lycurgus - Helvétius believed - had transformed theory into practice.

Helvétius was not a great thinker, but he was surely one of the most daring writers of his age.

See also: D'Alembert, J.; Enlightenment, Continental; Voltaire

MARK HULLIUNG

List of works

Helvétius, C.-A. (1758) *De l'Esprit*, Paris: Arthème Fayard, 1988.
Helvétius, C.-A. (1773) *De l'Homme*, Paris: Arthème Fayard, 1989, 2 vols. (The English translations date from the eighteenth century and are unreliable.)

References and further reading

Andlau, B. (1939) *Helvétius, Seigneur de Voré*, Paris: Fernand Sorlot.(For information about his life and family.)

Smith, D.W. (1965) *Helvétius, a Study in Persecution*, Oxford: Clarendon Press.(For the politics of censorship.)(A good study of the thought of Helvétius has yet to be written.)

Hempel, Carl Gustav (1905-1997)

Hempel's defence of Carnap's and Neurath's physicalism testifies to the presence of certain 'postmodern' themes in logical empiricism (or logical positivism): (1) a textualist turn to sentences from the facts or reality they are said to report; (2) a pragmatic turn from truth to inclusion in the text as the basic scientific concern; and (3) a descriptive turn from logic to empirical sociology of science.

Throughout his philosophical writings Hempel held that the question of what truth-claims mean should be replaced by the question of what criteria we use in deciding whether or not to call sentences true. Granted that 'Whales are mammals' is true if and only if whales are mammals, the question remains of how observation reports bear on whether or not to include a sentence in the text of current knowledge, and how received observation sentences may be dropped from it. For Hempel, then, the problem of analysing the concept of confirmation of sentences by sentences is either the heart of the problem of truth or the successor to that problem. Like Carnap, in 1945 he thought such an analysis necessary to connect the terms 'logical' and 'empiricism'; but four decades later he would conclude that, after all, one must 'leave the decision in matters of confirmation to the intuitive appraisal of the scientist' (1945b: 98).

1 Life

Born in Oranienburg of Prussian stock, Carl Gustav Hempel (called 'Peter' since his school days) studied mathematics, physics and philosophy at Göttingen, Heidelberg, Vienna and Berlin, where he received the Ph.D. in 1934 for work with Hans Reichenbach, lately dismissed from his professorship on account of his Jewish ancestry. In Brussels (1934-7) he was employed by his fellow exile Paul Oppenheim as a philosophical tutor-collaborator. When Oppenheim left Belgium in anticipation of a German invasion, Carnap obtained a research grant for Hempel, taking him to Chicago (1937-8). In time he found teaching posts at City and Queens Colleges (New York, 1939-48) and Yale (1948-55). Much courted, he then taught at Princeton (1955-75) and other universities, notably Pittsburgh (1977-85), and lectured widely (for example, in Beijing in 1982), before finally retiring to Princeton. He has received countless honours, including nine honorary doctorates and is much loved.

2 Confirmation

Hempel's paradox of the ravens has been a persistent challenge for theories of confirmation. Adumbrated (1937: 222) in a critique of Reichenbach's treatment of the probability of laws, it emerged as an illustration of the 'paradox of confirmation' (1945a). The sentences (1) 'All ravens are black' and (2) 'All non-black things are non-ravens' are logically equivalent; each simply denies the existence of non-black ravens. But if reports of black ravens confirm (1), then on the same showing, reports of non-black non-ravens confirm (2). Therefore, since any report confirming a sentence equally confirms any logically equivalent sentence, we have a seeming paradox: reports of non-black non-ravens (for instance, white shoes) confirm the hypothesis that all ravens are black.

Hempel's own account of the conditions under which observation-sentences confirm general sentences (1943, 1945a) can be summarized as follows. Consider an observation report in which certain names appear - perhaps, in the ravens case, just '*a*' and '*b*', where the report is the pair of sentences: '*a* is a black raven', '*b* is a non-black non-raven'. This report is said to *directly* confirm (1) because it logically implies the restriction of (1) to the individuals mentioned in it; and it is said to *indirectly* confirm all other logical consequences of sentences it directly confirms. In the present example the restriction of (1) to '*a*' and '*b*' is the sentence (0) 'If *a* is a raven then *a* is black and if *b* is a raven then *b* is black'; and since the report does logically imply (0), it directly confirms (1), and indirectly confirms (2).

This purely classificatory analysis of confirmation makes no distinction between white shoes and black ravens as sources of confirmation for (1); but perhaps that is because those two sorts of evidence confirm (1) to very different degrees. Hosiasson suggested that while reports of black ravens and white shoes both confirm (1), the latter confirm it to a much lower degree, in the light of our background beliefs about the rarity and homogeneity of ravens (Hosaisson 1940).

Hempel and Oppenheim (1945b) propose a quantitative account of degree of confirmation of hypotheses *H* by evidence-sentences *E*. If the language in which *H* and *E* are stated can classify members of the (finite) population

into non-overlapping, exhaustive categories, then a statistical distribution is an assignment to these categories of definite proportions in the population. Relative to different distributions, a sentence of the language may have different probabilities. Now the degree of confirmation of *H* given *E* is defined as the probability that *H* and *E* are both true divided by the probability that *E* is true - both probabilities being determined by a 'maximum likelihood' distribution, that is, a distribution making *E* as probable as any distribution can (1945b: 112).

In Carnap the role of the maximum likelihood distribution in this definition is played by the average of all possible distributions (Carnap 1950: appendix). Carnap's degrees of confirmation then satisfy all laws of probability, including one violated by the Hempel-Oppenheim definition, namely, the 'multiplication principle', according to which the probability that *A* and *B*, given *E*, is the degree of confirmation of *A*, given *E*, times the degree of confirmation of *B*, given that *E* and *A* are both true. Hempel and Oppenheim preferred their definition to Carnap's on grounds of empiricism; where Carnap assigns equal a priori probabilities to all possible distributions, they choose a distribution by reference to the evidence sentence *E*. They offer their definition as establishing fair betting-odds on *H* given *E*, thus accounting for one aspect of confirmation. Other aspects - number and variety of instances reported in *E* - remain to be accounted for by some other concept, which would determine the amount it would be fair to stake in bets at those odds (1945b: 114). But that view was hard to maintain in the face of Ramsey's and de Finetti's 'Dutch book' arguments for the multiplication principle, showing that 'If anyone's mental condition violated [it], his choice would depend on the precise form in which the options were offered him, which would be absurd' (Ramsey 1931: 182). This would be part of Hempel's later disenchantment with the project of an inductive logic (see Confirmation theory §1; Probability theory and epistemology §1).

3 Law and explanation

From Frege (1879: §12) and Mach (1883: ch. 4, §4.3) through the various phases of logical empiricism, scientific laws were taken to have the same logical form as other factual generalizations; especially, the suggestion was resisted that scientific laws are marked by modal operators, as in 'Necessarily, all Ps are Q' or 'Being P makes things Q'. Where Mach put this as a metaphysical claim - 'In nature there is no cause, no effect; nature is only there once' (1883: 455) - the logical empiricists put it as a decision about the language of science; it is to be extensional. But they saw the function of laws and theories as the systematization of otherwise unwieldy bodies of particular empirical claims, that is, data compression, Mach's 'economy of science'. This would accord well with the puzzling character of fundamental physics that emerged with increasing clarity during the twentieth century, with quantum theory providing predictions of unprecedented accuracy, explicable only as being what the theory provides. Here Hempel's 'covering-law model' of explanation matched the consensus of physicists that accurate systematization is all we can reasonably ask in the way of explanation - 'understanding' being a secondary psychological state that may come with familiarity (see Laws, natural §1; Logical positivism §§2-3).

Hempel identifies laws as those truths of the forms 'Every physical object is so-and-so' and 'Every space-time point is so-and-so' that make no essential reference to particular physical objects or space-time regions (1965: 271). The seemingly limited generalization 'All ravens are black' is really a universal generalization in which 'so-and-so' stands for 'black if a raven' and 'if' is read extensionally, as 'or not'. If that generalization is a law, 'raven' must be definable without essential reference to the planet Earth, as is presumably the case if biology is reducible to physics.

A deductive explanation of an event is a logical deduction of an 'explanandum' sentence, asserting that the event happens, from premises (the 'explanans') which are either general laws or statements of antecedent conditions (1965: 249). Hempel thought that this put explanation on a par with prediction, the difference being attributed to the timing of the explanandum-event, that is to say, after or before the act of prediction or explanation, respectively. In inductive explanations the requirement of logical deductibility is weakened to high logical probability of the explanandum, given the explanans. Since the conditional probability of a statement is highest (that is, 1) given statements that logically imply it, deductive explanations are also inductive explanations - of the strongest sort (see Explanation §§1-2).

4 Theoreticity and cognitive significance

Explanation is systematization. Hempel proposes a measure of the power of a theory *T* to systematize putative data *S* (1965: 287). The idea is that the content of a sentence is determined by the possibilities it rules out, so the

content of a sentence is represented by its denial, and the common content of two sentences by the sentence 'Both are false'. Measuring contents by logical probabilities of denials, Hempel then measures *T*'s power to systematize *S* by the ratio of the measure of their common content to the measure of *S*'s content, that is, by the logical probability of *T*'s falsity, given *S*'s. The idea is applied to 'the theoretician's dilemma': if theoretical terms and laws do their work of deductive systematization in linking observational antecedents to observational consequents, they are otiose, for functionally equivalent rules in a purely observational vocabulary can then establish exactly the same inferential transitions; but this overlooks the role of theoretical terms and principles in *inductive* systematization of observational sentences (1965: 186, 122).

Following Wittgenstein (1922), logical empiricists rejected metaphysics as empty of cognitive content. Early on, this was derived from the so-called verifiability criterion, identifying the content of a sentence or a system with the set of ways in which it might be refuted by experience. Like tautologies, such as 'What will be, will be', metaphysical sentences were thought to be unfalsifiable, and so to be empty of cognitive content. Hempel famously chronicled and sharpened the problems and changes attending attempts at logical definition of this concept, for example, the replacement of the verifiability requirement by confirmability (1965: 101-22). He concludes: 'cognitive significance in a system is a matter of degree: Significant systems range from those whose entire extralogical vocabulary consists of observation terms, through theories whose formulation relies heavily on theoretical constructs, on to systems with hardly any bearing on potential empirical findings' (1965: 117).

5 The career of logical positivism

Logical empiricism was meant to be scientific philosophy, applying the new logic to the structure and methods of all the sciences, as Whitehead and Russell had to mathematics in *Principia Mathematica* (1910-25). The products would be purely logical definitions - of confirmation, explanation, systematic power and so on - but their selection or design would be determined pragmatically, by the needs of science. That was Carnap's view, anyway; and Hempel's, for nearly half a century. But where Carnap's hopes were steady to the end, Hempel's eventually turned from logical analysis to Neurath's old project of *Gelehrtenbehavioristik* (empirical sociology of science) - in loose alliance with Kuhn (1962, 1970), but with fresh concern for the justification of scientific norms.

See also: Meaning and verification §4

R. JEFFREY

List of works

Hempel, C.G. (1934) *Beiträge sur logischen Analyse des Wahrscheinlichkeitsbegriffs (Contributions to the Logical Analysis of the Concept of Probability)*, Jena.(Hempel's Berlin Ph.D. thesis.)

Hempel, C.G. (1935a) 'On the Logical Positivists' Theory of Truth', *Analysis* 2: 49-59. (Exposition and defence of Carnap's and Neurath's physicalism, presented as a quasi-coherence theory of truth.)

Hempel, C.G. (1935b) 'Analyse logique de la psychologie', *Revue de Synthèse* 10: 27-42; trans. 'The Logical Analysis of Psychology', in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Appleton Century; Crofts, 1949, 373-84. ('Psychology is an integral part of physics.')

Hempel, C.G. (1937) 'Le problème de la vérité' (The Problem of Truth), *Theoria* 3: 206-46.(An accessible bridge from the early work on probability and truth to the mature work on confirmation.)

Hempel, C.G. (1943) 'A Purely Syntactical Definition of Confirmation', *Journal of Symbolic Logic* 8: 122-43. (Non-quantitative confirmation as a relation between sentences; a narrowly technical preamble to Hempel 1945a.)

Hempel, C.G. (1945a) 'Studies in the Logic of Confirmation', in *Aspects of Scientific Explanation*, London: Collier-Macmillan, 1965.(Hempel's classical discussion of non-quantitative confirmation and its paradoxes.)

Hempel, C.G. and Oppenheim, P. (1945b) 'A Definition of Degree of Confirmation', *Philosophy of Science* 10: 98-115.(Quantitative confirmation of one sentence by another.)

Hempel, C.G. (1952) *Fundamentals of Concept Formation in Empirical Science*, International Encyclopedia of Unified Science 2, Chicago, IL: University of Chicago Press.(A readable account of the logical empiricist line at the time on definition, reduction and types of measurement.)

Hempel, C.G. (1960) 'Science and Human Values' in R. Spiller (ed.) *Social Control in a Free Society*, Philadelphia, PA: University of Pennsylvania Press, 39-64.(Interactions and parallels between scientific and value judgment.)

Hempel, C.G. (1962) 'Rational action', *Proceedings and Addresses of the American Philosophical Association*

35, Yellow Springs, OH: Antioch, 5-23.(American Philosophical Association presidential address; explanatory uses of the concept of rational action.)

Hempel, C.G. (1965) *Aspects of Scientific Explanation*, London: Collier-Macmillan, and New York: Free Press. (Reproduces some fifteen previously published essays, with commentary, blending and revision. The title essay is original.)

Hempel, C.G. (1966) *Philosophy of Natural Science*, Englewood Cliffs, NJ: Prentice Hall.(A lucid introduction, from a broadly logical empiricist standpoint.)

Hempel, C.G. (1970) 'On the "Standard Conception" of Scientific Theories', in M. Radner and S. Winokur (eds) *Minnesota Studies in the Philosophy of Science* 4, Minneapolis, MN: University of Minnesota Press, 142-263. (How do new theoretical terms get their meanings? Old answers are faulted - as is the question itself.)

Hempel, C.G. (1977) *Aspekte wissenschaftlicher Erklarung*, Berlin: de Gruyter.(Translation and revision of Hempel 1965, with a new section on statistical explanation.)

Hempel, C.G. (1979) 'Scientific Rationality: Analytic vs Pragmatic Perspectives', in T. Geraets (ed.) *Rationality Today/La rationalité aujhourd'hui*, Ottawa, Ont.: University of Ottawa Press, 46-58.(Critique of Kuhn and others, issuing in a view of rational explanation as Janus-headed, at once explanatory and normative.)

Hempel, C.G. (1983) 'Schlick und Neurath: Fundierung *vs*. Koharenz in der wissenschaftlichen Erkenntnis', *Gratzer philosophische Studien*, 16/17: 1-18.(Schlick and Neurath: foundation versus coherence in scientific knowledge. A debate, in Vienna in the 1930s, resonating with Hempel 1979.)

Hempel, C.G. (1988) 'On the Cognitive Status and the Rationale of Scientific Rationality', *Poetics Today* 9: 5-27. (Pragmatic perspectives on induction; methodology as both descriptive and prescriptive.)

Hempel, C.G. (1990) 'The Significance [misprinted 'Signification'] of the Concept of Truth for the Critical Appraisal of Scientific Theories', *Nuova civilita delle macchine* 8: 109-13. (Watered-down title; the manuscript said 'Irrelevance', not 'Signification'.)

Hempel, C.G. (1993) 'Empiricism in the Vienna Circle and in the Berlin Society for Exact Philosophy, Recollections and reflections', in F. Stadtler (ed.) *Scientific Philosophy: Origins and Developments*, Dordrecht: Kluwer.(An eyewitness report.)

References and further reading

Carnap, R. (1928) *Der logische Aufbau der Welt*, Berlin: Felix Meiner; trans. R. George, *The Logical Structure of the World*, Berkeley: University of California Press, 1969. (Phenomenalism formalized; the book that brought Hempel to Vienna.)

Carnap, R. (1934) *Logische Syntax der Sprache*, Vienna; trans. A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul, Trench, Trubner & Co, 1937.(Physicalism formalized.)

Carnap, R. (1950, 1962) *Logical Foundations of Probability*, Chicago, IL: University of Chicago Press. (Probability as a logical magnitude. Crititique of Hempel 1943 and 1945a on pages 468-82.)

Frege, G. (1879) *Begriffsschrift*, Halle: Louis Nebert.(The ultimate source of the logic in logical empiricism.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel*, Cambridge, MA: Harvard University Press.(A valuable source book.)

Horwich, P. (ed.) (1993) *World Changes*, Cambridge, MA: MIT Press.(New light on Kuhn, Hempel and Carnap.)

Hosiasson, J. (1940) 'On confirmation', *Journal of Symbolic Logic* 5: 136-41. (Probabilistic resolution of Hempel's paradoxes of confirmation. Hosaisson was killed by the Gestapo in 1942.)

Kuhn, T. (1962, 1970) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(The book that killed logical empiricism?)

Mach, E. (1883) *Die Mechanik in ihrer Entwickelung*, Leipzig: F.A. Brockhaus; trans. T. McCormack, *The Science of Mechanics*, La Salle, IL: Open Court, 1893.(The economy of science: ch. 4, §4.)

Neurath, O. (1983) *Philosophical Papers 1913-1946*, Dordrecht: Reidel.(Far-left wing of the Vienna Circle.)

Ramsey, F.P. (1931) *The Foundations of Mathematics*, London: Routledge.('Truth and Probability' is one of many gems here. See page 182 for the 'Dutch' book.)

Reichenbach, H. (1935) *Wahrscheinlichkeitslehre*, Leyden: A.W. Sijthoff; trans. *The Theory of Probability*, Berkeley, CA: University of California Press, 1949.(Hempel (1934) was largely concerned with this theory.)

Russell, B.A.W. (1940) *An Inquiry into Meaning and Truth*, London: Allen & Unwin, and New York: Norton, 139-49.(Caricature of Hempel and Neurath on truth.)

Stadtler, F. (ed.) (1993) *Scientific Philosophy: Origins and Developments*, Dordrecht: Kluwer.(Essays relating to the Vienna Circle, with useful references.)

Whitehead, A.N. and Russell, B. (1910-25) *Principia Mathematica*, Cambridge: Cambridge University Press. (The immediate source of the logic in logical empiricism.)

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, London: Routledge.(Mysticism and logic. It electrified the Vienna Circle.)

Henricus Regius (1598-1679)

Henricus Regius supported many of Descartes' doctrines, such as the distinction between mind and body, but disagreed with Descartes' arguments for that distinction, and rejected metaphysics. Their differences led to a celebrated public dispute which lasted until Descartes' death.

Henricus Regius (Hendrik De Roy), one of Descartes' first Dutch disciples, also studied in France and Italy. In 1638 he was appointed professor of Theoretical Medicine and Botany at the University of Utrecht. Through Henricus Reneri he became deeply influenced by Descartes' *Discourse* and *Essays* of 1637, meeting Descartes in 1638 (see Descartes, R. §§1, 8). When Reneri died in 1639, Regius became the primary representative of Cartesianism at Utrecht.

Regius first presented his natural philosophy in a series of disputations in 1641, published as *Physiologia sive cognitio sanitatis*. It incurred the wrath of the Theological Faculty and of the University rector, Gisbertus Voetius. Most controversial were his denial of the scholastic doctrines of matter and form, and his assertion that the human being is an accidental unity. With Descartes' advice, Regius replied to the attacks in 1642. This elicited a formal condemnation of Descartes' philosophy at the University, and Regius was forbidden to teach physics. Descartes responded in the 'Letter to Father Dinet' (1642), which appeared with the Amsterdam edition of his *Meditations* publicly attacking Voetius, who then focused his attention on Descartes.

Throughout this time, Regius was summarizing his own natural philosophy. Descartes disagreed with certain crucial aspects of it and urged him not to publish, but *Fundamenta physices* appeared in spring 1646, followed by *Fundamenta medicinae* in 1647. Descartes consequently denounced Regius in the preface to the new French edition of the *Principles of Philosophy* (1647). Regius responded with *Explicatio mentis humanae* (1647b), emphasizing his points of difference. Descartes retorted with *Notae in programma quoddam* (1648), and finally Regius published *Brevis explicatio mentis humanae sive animae rationalis* (1648).

Bitter as these exchanges were, Descartes is thought to have forgiven Regius before his death in 1650, and Regius continued to be identified as a Cartesian. He remained at Utrecht until his death, a person of some importance in the University. Although he did not write anything new after the 1640s, he published later editions of the *Fundamenta physices* in 1654 and 1661.

Regius rejected scholastic philosophy, embracing a version of the new mechanical philosophy. Like Descartes he distinguished between mind, a thinking substance, and body, an extended substance. The details of his general physics, for example his conception of the laws of nature, are similarly indebted to Descartes, as are his views on human physiology. But there are also notable differences. Descartes' physics was grounded in metaphysics, in God, the human soul and in the validation of reason. Regius rejected metaphysics, beginning his philosophy directly with physics. He also disagreed with the Cartesian arguments for the distinction between mind and body: in the *Fundamenta physices* he argued that the real distinction can be known only by Biblical revelation. In the *Explicatio* (§2), he went farther still: 'so far as the nature of things is concerned… mind could be either a substance or a certain mode of corporeal substance'.

Regius also attacked Descartes' proof of the external world. According to Descartes, God gave us a strong inclination to believe that bodies cause sensations, and no means to correct this strong inclination. Therefore, he would be a deceiver if there were none. Apart from the fact that God might deceive us to punish us, Regius did not think God would be a deceiver if no bodies existed, since we are capable of withholding our assent to their existence, viewing our sensations as mere appearances which may not have bodies as their causes. Again, it is revelation that assures the existence of the external world. Finally, Regius rejected Descartes' innate ideas and all ideas of pure intellection; we have only sensation, imagination and memory - all our ideas come to us from experience. Pierre Gassendi is a possible source of some of Regius' non-Cartesian theses.

See also: Medieval philosophy

DANIEL GARBER

List of works

Regius, H. (1641) *Physiologia sive cognitio sanitatis (Physiology, or, knowledge of health)*, Utrecht.(Collection of disputations which contain the earliest presentation of Regius' natural philosophy.)

Regius, H. (1646) *Fundamenta physices (Foundations of physics)*, Amsterdam.(Regius' main presentation of his natural philosophy.)

Regius, H. (1647a) *Fundamenta medicinae (Foundations of medicine)*, Utrecht.(A further presentation of Regius' natural philosophy.)

Regius, H. (1647b) *Explicatio mentis humanae (Explanation of the human mind)* in *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch, and A. Kenny, Cambridge: Cambridge University Press, vol. 1, 1984-91, 294-6.(Reprinted here with Descartes' response *Notae in programma quoddam*.)

Regius, H. (1648) *Brevis explicatio mentis humanae sive animae rationalis… a Notis Nobil. Cartesii vindicata (Brief explanation of the human mind or rational soul… vindicated from the notes of Descartes)*, Utrecht.(A response to Descartes' *Notae in Programma*.)

Regius, H. (1654) *Philosophia naturalis…(Natural philosophy)*, Amsterdam.(A second edition of the *Fundamenta physices*.)

Regius, H. (1661) *Philosophia naturalis…(Natural philosophy)*, Amsterdam.(A third edition of the *Fundamenta physices*.)

References and further reading

Adam, C. and Tannery, P. (1964-74) 'Descartes et Regius', *Oeuvres de Descartes*, vol. 11, ed. C. Adam and P. Tannery, Paris: CNRS/Vrin, new edn, 11 vols.(A detailed discussion of Descartes' relations with Regius, and Regius' treatment of Descartes and his thought in the *Fundamenta physices*.)

Descartes, R. (1642) 'Letter to Father Dinet', *The Philosophical Writings of Descartes*, vol. 2, (partial trans.), ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91.(First published in the 1642 second edition of the *Meditations*. Although mostly a response to perceived Jesuit attacks against him, it also includes a defence of Regius against Voetius.)

Descartes, R. (1647) 'Author's letter to the translator of the book which may here serve as a preface', *Principles of Philosophy*, Paris; repr. in *The Philosophical Writings of Descartes*, vol. 1, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91.(This preface to the French translation of the *Principles* denounces Regius' *Fundamenta physices*.)

Descartes, R. (1959) *Lettres à Regius et remarques sur l'explication de l'esprit humain*, ed. G. Rodis-Lewis, Paris: Vrin.(Contains the correspondence between Regius and Descartes, both in Latin and in French, the *Notae*, as well as excerpts from others of Regius' writings, an extensive introduction and helpful notes.)

Farina, P. (1977) 'Il corpuscolarismo di Henricus Regius: materialismo e medicina in un Cartesiano olandese del Seicento', in Ugo Baldini *et al*. (eds) *Ricerche sull'Atomismo di Seicento: Atti del Convegno di Studio di Santa Margharita Ligure (14-16 ottobre 1976)*, Florence: La Nuova Italia, 119-78.(A discussion of Regius's physics and medical theories.)

Verbeek, T. (1988) *La querelle d'Utrecht*, Paris: Les impressions nouvelles.(The documents connected with Descartes' troubles in Utrecht, translated into French, with copious commentary.)

Verbeek, T. (1992) *Descartes and the Dutch*, Carbondale, IL: Southern Illinois University Press.(A short but engaging account of Descartes' troubles in Utrecht and Leiden, containing a good account of his relations with Regius.)

Verbeek, T. (1994) 'Regius's Fundamenta Physices', *Journal of the History of Ideas* 55: 533-51.(An account of the history of Regius' main work in natural philosophy.)

Verbeek, T. (ed.) (1993) *Descartes et Regius. Autour de l'explication de l'esprit*, Amsterdam: Rodopi.(A collection of essays concerning Descartes and Regius on the mind, focusing on Regius' *Explicatio* and Descartes' response.)

Vrijer M.J.A. de (1917) *Henricus Regius: een 'Cartesiaansch' hoogleeraar aan de Utrechtsche Hoogeschool*, The Hague: Nijhoff.(Though somewhat dated, still the most complete account of Regius's life and thought currently available.)

Henry of Ghent (early 13th century-1293)

Perhaps the most influential theologian between Thomas Aquinas and Bonaventure in the third quarter of the thirteenth century and John Duns Scotus at the beginning of the fourteenth century, Henry of Ghent stands at a turning point in scholastic philosophy. He was a defender of traditional Neoplatonic positions and has often been seen as the epitome of thirteenth-century Augustinianism. Yet his convoluted metaphysics and a theory of knowledge weaving together Neoplatonic and Aristotelian strands inspired novel philosophical trends in the fourteenth century, particularly among Franciscan thinkers. His work thus constituted the point of departure for scholastic giants like Duns Scotus and William of Ockham, who not only used him as a foil against which to articulate their own system of thought but also absorbed much of his fundamental philosophical outlook and terminology.

Characteristic of Henry's metaphysics was an essentialism so pronounced that critics accused him of positing a realm of essences separate from worldly actuality. In his defence, Henry insisted that essences, though prior to actual existence, were separate only as grounded in the divine exemplars of things, but the Platonism of his approach struck his contemporaries as extraordinary nonetheless. Ironically, Henry's understanding of essence as congruent with intellectual coherence provided an opening for a more logic-based analysis of modality, especially possibility, in succeeding thinkers such as Duns Scotus.

The emphasis on essence re-emerged in Henry's theory of knowledge, and at least in his early writings he offered a vision of knowing truth through divine illumination often taken as paradigmatic of medieval Augustinianism. Even his later attempts to cast epistemology in a more Aristotelian light retained the insistence that true knowledge somehow entails access to the exemplary essences in God's mind. The same essentialism led Henry to formulate what he called an a priori proof for God's existence, best approximation in the thirteenth century to Anselm's ontological argument. Again, however, Henry's Augustinianism provided an unintended springboard for innovation, leading to Duns Scotus' theory of the univocity of being and metaphysical proof of God's existence.

1 Life and writings

Henry was born in Ghent early in the thirteenth century, but we know little about his life before 1275 or 1276, when he assumed a secular chair as master of theology at the University of Paris. By this time he was already working his way up the ladder of ecclesiastical politics in the Low Countries, having been appointed canon at the wealthy episcopal see of Tournai; soon after he was named archdeacon of Bruges, and finally in 1278 or 1279, archdeacon of Tournai. In 1277, Bishop Etienne Tempier of Paris included him in the commission of theologians assigned to examine the orthodoxy of doctrines being taught at the University. Their report led to the list of 219 proscribed propositions in philosophy and theology issued in Tempier's famous condemnation of 1277 (see Averroism; Boethius of Dacia; Siger of Brabant).

By this point, Henry had become aligned with conservative thinkers fearful of the influence of a radical Aristotelianism among Parisian masters of arts and alarmed even by the moderate Aristotelianism of Aquinas and his followers in the school of theology (see Aristotelianism, medieval). Members of this group rallied to the banner of Augustine, and because of the prominence of Franciscans within their ranks they have often been characterized as proponents of a peculiarly Franciscan Augustinianism (see Augustinianism). Ironically, Henry was a bitter opponent of the Franciscans in the political disputes of the late thirteenth century over the conflicting rights and privileges of regular and secular priests.

Henry died in 1293 after an illustrious career at the centre of intellectual debate. His chief works are a series of fifteen *Quodlibeta*, the carefully revised redactions of public disputations which he held nearly annually at the university from 1276 to his withdrawal from academic life the year before his death, and his *Summa* - more properly, *Quaestiones ordinariae (Ordinary Questions)* - a massive compilation of edited classroom debates on questions of philosophy and theology that he left only half finished. Although the latter presents itself as a self-contained and independent work, it was laboriously put together over the same span of time as the *Quodlibeta* and parallels them in chronicling the development of Henry's ideas throughout his career.

2 Metaphysics

Like many of his contemporaries, Henry began with the conviction, promoted by Aquinas, that the most fundamental divide in reality lay between essence and existence. In contrast to Aquinas he interpreted the division so as to give priority to essence. Going back to Avicenna (see Ibn Sina), he claimed that the broadest way of conceiving of an object was as ‘thing’ (*res*), a concept shorn of all consideration of the conditions of being, particular or universal, mental or extramental, actual or hypothetical. Void of ontological significance, ‘thing’ was thus the marker of essentiality alone, and its intelligible content could be represented solely by essential or quidditative descriptions, pointing to the ‘what’ of objects exclusive of the ‘whether’ or ‘how’ of their existence.

Yet despite their formal opposition to being, quidditative descriptions varied according to how dense their ‘thingness’ was: in other words, how suitable they were to represent actual existing objects. The most tenuous - least ‘real’ - kind of thing included all that could be described, even barely imaginable constructs which would not withstand scrutiny for logical coherence. This was the vast domain of ‘thing’, opposed only to that which was absolutely nothing at all. Henry called it *res a reor reris*, by which he meant ‘thing’ insofar as it embraced any content that could be grasped by mind or imagination and which we might translate as ‘imaginable object’. Less thin was ‘thing’ that was not only imaginable but also internally consistent enough to be suitable for becoming an object in the real world. The domain of this kind of ‘thing’ was of course less extensive than *res a reor reris*, since it contained only objects more truly ‘thing-like’ or ‘real’ and thus more worthy of being denominated by the terms ‘essence’, ‘nature’ or ‘quiddity’ in their proper sense. Henry called such objects *res a ratitudine* - that is, ‘things’ as disposed to rational analysis - and among them he located Avicenna’s absolute essence, which he understood to be essence separated from existence but opposed to the less-than-absolute nothingness of incoherent figments of imagination. Finally - and least thin - was the realm of ‘things’ in the world, objects not only inherently suited to existence but in fact existing. This field of things was naturally the narrowest, but it was denser or more ‘real’ than either of the other two. Henry called its constituents actually existing things (*res existentes in actu*) or things of nature (*res naturae*).

Such a hierarchical vision of ‘thing’ was more intricately articulated than any consideration of essence produced before, and in order to accommodate it Henry had to introduce distinctions on the other side of the divide, in the category of ‘being.’ According to him, there were two kinds of being. First was being of essence (*esse essentiae*), a state less than actuality, characterizing objects at the level of *res a ratitudine* and constituting that which separated them from more nothing-like imaginary objects that could not bear logical scrutiny. Second was being of existence (*esse existentiae*), the full being of actuality possessed by all real objects in the temporal present and therefore proper to *res existentes in actu*. From this it followed that there was a two-fold composition of being and essence, the first a combination of absolute essence with the special being distinguishing it from the purely ephemeral whatness of imaginable fictions, the second the joining of such essence with actual existence (see Existence).

The result was that Henry, while maintaining the being-essence distinction, had managed to undermine its radical duality, shifting the analytical balance strongly to the side of essence. By his scheme essences, even when considered ‘absolutely’, already possessed a sort of being that lifted it into the domain of non-fictive things, so that even on a conceptual level essence was never completely denuded of implications of being. Furthermore, essence, taken along with its special sort of being, was clearly prior to existence. One might almost describe it as ‘resident’ in an essential realm ready to receive the actual existence that, in Henry’s words, ‘came to it’ and accounted for its presence in actuality. For Henry, then, being was more like a variable quality of essence in its different states than a fundamental constituent of reality, as it had been for Aquinas. Giles of Rome, Henry’s contemporary at Paris and an extreme defender of the Thomistic real distinction between essence and being, continually attacked Henry on this score. Henry’s response was to invent a distinction intermediate between the purely conceptual and the real - what he called an intentional distinction - and to insist that essence and its two kinds of being were ‘intentionally’ distinct. This tour de force of metaphysical compromise fed into the similarly controversial ‘formal distinction’ of Duns Scotus.

Already strongly Neoplatonic in his essentialism, Henry exaggerated the tilt in his metaphysics by tying his theory of essence to Augustine’s exemplarism. What made an essence an essence, Henry claimed - what explained its being more coherent than an imaginary object and a bearer of being of essence - was the fact that it corresponded to a formal ideal in the divine mind. Essence was essence because it was exemplified, because it was the image of a conceptual exemplar in God. Indeed, the being of essence was reducible to the relation between the exemplified and its exemplar. God’s knowledge made essences what they were, and each was a sort of participation in the

divine being.

Henry took this idea so far as to approach a Platonism astonishing even among thirteenth-century Augustinians (see Platonism, medieval). Contemporaries such as Godfrey of Fontaines accused him of positing a world of essences, eternal and separate from God, since by Henry's terms each essence had its own being of essence different from the existence of a created thing in the actual world. Indeed, it is hard not to read some of his statements about exemplified essences as implying what Godfrey claimed. Yet Henry vigorously rebutted such accusations, insisting that being of essence was not an attribute things could have without also possessing being of existence and thus not a separate state of being. It was rather the metaphysical purchase on actuality that a coherent essential content had as something suitable for existence. In the case of objects not actually existing in the external world, essence was manifested, and possessed its being of essence, just insofar as it was actualized in the intelligible processes of an existing mind. This could be the mind of an intelligent creature, like a human being, considering the essence abstracted from any particular instances external to its thought; it could also be the mind of God. Of course, before creation it was only in God's mind that essence could thus take shape. Eternal essences fell back completely on the ontic grounding of God's eternally active mind and so were identical with the divine ideas, thus substantially identical with God (on this controversial point see Marrone 1985: 122-7).

A system of this sort, isolating essence as an absolute in itself but also grounding it eternally in God, required considerable philosophical elasticity to accept, but it paid handsome metaphysical dividends. For Henry, essences were not purely fictive constructs on the level of *res a reor reris*, but quidditative contents that could be made existent. They constituted, therefore, all simple possibles. Besides providing an exposition of the relation between God and the rest of reality, Henry's theory of essence thus promised to lay bare the foundations of modality as well. In this connection it must be remembered that in Henry's words what made an essence an essence - what constituted it as a possible - was, first, its being represented by an idea in God's mind but, second, its possessing a coherent intelligible content, formally entailed in its corresponding to a divine idea. In the end, Henry was suggesting that what separated possibles on the essential level of *res a ratitudine* from fictive impossibles on the level of *res a reor reris* was internal consistency, with all essences being made up of logically compatible parts. Despite the fact that Duns Scotus criticized Henry's ontology of essence as unworkable and implausible, it provided a major inspiration for his own innovative ideas about modality, which were formally quite similar to those Henry had laid down.

3 Theory of mind and knowledge

Henry's epistemology is often taken as emblematic of thirteenth-century Augustinianism, but it was likewise the area of his thought in which he did most to bring new elements of Aristotelianism into a more traditional Neoplatonic base. This was the side of his philosophy most subject to development throughout his career, where it is especially important to pay attention to the chronology of his ideas.

In the early sections of his *Summa*, Henry laid out a theory of knowledge dependent on a notion of truth as rightness deriving from Anselm and insistent on the inadequacy of the unaided human intellect to achieve fully reliable knowledge without God's help, traditionally described in Neoplatonic terms as an illumination from the divine intelligible light (see Neoplatonism). This first statement of his views about knowledge and truth constitutes the most detailed thirteenth-century exposition of the theory of divine illumination and was taken as the paradigm of illuminationism by later Augustinians such as Duns Scotus, who criticized any such theory as philosophically unsustainable. According to Henry's position, human intellect begins with the data of sensation and works its way by purely natural means to an approximate understanding of an object's essential composition, an understanding neither completely clear nor precise. Henry called this 'knowledge of the true' (*verum*). Only after it had reached this point could the mind call upon its reflexive capacities and God's aid to refine its understanding until it was fully clear and adequate to the theoretical requirements for certain knowledge of the truth.

The first step in the refining process involved taking the cognitive representation of the object achieved in the initial understanding - which Henry identified at that stage of his career as an intelligible species - and comparing it to the external thing again as known through the senses, critically reshaping it into a precise image of the object's essential characteristics. The results of this comparison, Henry said, were preserved in the mind as what he called an expressed concept or 'word' (*verbum*), and it was with this 'word' that mind first reached a grade of cognition worthy of being called knowledge of the truth (*veritas*) and sufficient for the beginnings of scientific

thought. From simple truths gathered in this way, one could construct a system of intellection adequate to the standards Aristotle had established for science in his *Posterior Analytics*.

Yet such truth, and such science, was still imperfect, because the human mind working on its own was subject to error. For absolute certitude, and perfect science, the intellect had to turn to God, calling upon an illumination from above to confirm its insights and correct its mistakes. Although Henry described this second step in various ways, his preference was to see it as a comparison drawn by the mind between the word of truth established at the lower boundaries of science and the eternal exemplar rendered intelligible by God through illumination (see Illumination). The expressed concept, rectified to suit the contours of the divine ideal, was now fashioned into a vehicle for knowledge of the ‘full truth, pure and simple’ (*sincera veritas*), upon which could be constructed an absolutely reliable system of knowledge.

Already evident in this early view was Henry’s intention to make room for Aristotle and the epistemology of the *Analytics* in the rarefied world of illuminationism and exemplarism. Yet in the years 1279-81 he took far greater steps to amplify his commitment to an Aristotelianizing theory of knowledge. In Quodlibets IV and V and in Article 34 of the *Summa*, Henry revisited his description of the process by which intellect moves from apprehension of the true to an understanding of truth worthy of science. He now insisted that the initial apprehension constituted genuine knowledge of the quiddity or essence of an object, although not a precise articulation of all quidditative parts. The way the mind worked its way from this imperfect knowledge to the truth was by applying the Aristotelian procedures of composition and division - what Henry called ‘the art of definition’ (*ars definitiva*) - progressing by means of analysis to a description of essence precise enough to provide the foundation for apodictic science. This precise description, an authentic Aristotelian definition, was manifested in the intellect by what Henry characterized as the refined mental concept or word, which he now conceded to be indicative of the pure truth he had previously reserved for understanding illuminated by God.

At the same time, Henry took parallel steps in his theory of mind. For all his early illuminationism, he had always rejected the notion, advanced by Augustinians like Roger Bacon in the mid-thirteenth century, that God was the agent intellect responsible for human cognition, or the equally Augustinian view that the ideas of things were impressed by God on the mind at its creation. Instead, he held to what he took to be authentic Aristotelianism: the claim that the human intellect possessed its own active power to generate knowledge from sensory data. In these middle years he fleshed out his understanding of this power by combining an Augustinian conviction that the mind was autonomously active and capable of producing knowledge on its own with an understanding of the mechanics of mind drawn from Averroes (see Ibn Rushd). The most striking feature of this synthesis was the notion of a possible intellect rendered active by its initial reception of the concept of the quiddity and working to turn out higher forms of cognition, an idea Henry consciously patterned after Averroes’ ‘speculative intellect’.

Henry carried the emphasis on intellectual activity so far as to reject entirely the notion of intelligible species, which he now saw as tokens of a passivity unworthy of the dynamism of the human soul (Nys 1949). Instead, all knowledge from the very start was contained in his ‘expressed concepts’ or ‘mental words’, products of the immanent agency of intellect as it reached out to grasp the essences of objects, and which could be regarded as those very essences residing in the mind as understood by it. Such a notion of intellectual immediacy was unusual in Henry’s day, but it prepared the ground for the idea of intellectual intuition found later in Duns Scotus as well as the even more radical theories of intellectual agency developed by Ockham and his followers.

Such views enabled Henry, by the end of his career, to find a way of reaffirming the divine presence in human cognition conveyed by his early epistemology of illumination. If mental words were nothing less than the essences of objects brought onto the mind’s cognitive field, they were then the perfect expression of Avicennian absolute essence, shorn of external existence and instantiated solely by the actuality of an intellect’s active consideration. But, of course, absolute essences were eternally present in divine cognition. Here the profound essentialism of Henry’s ontology emerged in full force, leading him to assert that the essence the human mind knew and the essence known by God were the very same on the rarefied level of being of essence. Thus every intelligent agent touched on things at just the point at which they participated in the divine mind, so that knowing the definitive truth in Aristotelian terms was, in a real sense, knowing the object as it bore an ideal relation to God. Henry’s mature understanding of the divine role in human intellection caught Duns Scotus’ attention as much as the literally illuminationist views of Henry’s early years, and he took aim against this as well in his criticism of

traditional Augustinian theories of knowledge.

4 Transcendentals and proof of God’s existence

The essentialism of Henry’s metaphysics and the exemplarism lurking behind even his most Aristotelianizing versions of a theory of knowledge encouraged him to make extraordinary use of the emerging doctrine of transcendentals (see Language, medieval theories of). Although he had insisted that ‘thing’ was the broadest term which could be used to describe an object, Henry agreed with his scholastic predecessors that the first and most general concept available to the human intellect was that of being (*ens* or *esse*). This term, transcendental in the sense that it was logically prior to the generic division of things into the ten categories, brought the mind face to face with the participationist foundations of reality.

Like all thinkers before Duns Scotus, Henry conceded that ‘being’ at its most general was strictly equivocal, disjunctively applicable to the irreducibly distinct realities of God’s being and the being of creatures. Yet although these two ‘beings’ had nothing really in common, they were similar enough - related by means of analogy - to give rise to concepts that were extremely close. Indeed, Henry said that in its initial perception of being, the mind was so deluded by this similarity as to mistake the two concepts referring to the being of God and that of creatures for a single concept.

The mind’s confusion on this point looked to Henry like a most fruitful mistake. After all, his account of the conceptualization of being did not mean simply that the mind took two different concepts or terms to be the same. It actually stipulated that from the start the human intellect was confronted with a dual object: the unlimited being of God and the finite being of creation. From this, Henry concluded that every human intellect had a natural knowledge of God: not a knowledge impressed on the mind at creation, but rather one springing naturally from its very first act of understanding.

Implausible as this position might appear from an Aristotelian point of view, it was fully in line with the cognitive implications of Henry’s own metaphysics, which opened the mind to a world of essence vibrant with the participatory relation between creature and creator. Moreover, it allowed Henry to support what he said was Avicenna’s opinion that the human intellect could naturally construct an a priori proof for the existence of God. Such a proof would be a priori not in the sense that its foundations preceded sensory perception but rather because it arose from a primitive quidditative understanding of the divine being and did not have to be argued a posteriori from the evidence of created existence. If in its first, imprecise efforts to grasp ‘being’ conceptually the mind formed two general concepts so vague and indeterminate as to appear identical, upon reflection it would realize that what it took to be one idea was in fact two. First would be the concept of the ‘negatively indeterminate’ divine being, inherently incompatible with any limitation whatsoever; second, there would be the concept of a ‘privatively indeterminate’ created being, not yet reduced to any genus or species but necessarily determined in that way in every real case outside the mind. For all its vagueness, the first of these two concepts was in fact quidditative and proper to God, adequate to the demands of a priori demonstration.

With this assertion Henry left contemporary theologians behind, most of whom were content with an a posteriori proof for God modelled on the Aristotelianizing inclinations of Aquinas or Averroes. Yet his ideas impressed the following generation. For all his criticism of Henry, Duns Scotus’ radical position on the univocity of the concept of being and the way he worked from it to develop a quidditative argument for God’s existence were inspired by his predecessor’s a priori proof (Marrone 1988: 49-56) (see God, arguments for the existence of).

5 Theory of will

Henry is generally credited with a voluntarist notion of moral action, in contrast to intellectualism of either the robust variety found in Godfrey of Fontaines or the mitigated strain of Aquinas and Giles of Rome. Although in fact no thinker of Henry’s day was voluntarist to the extent of completely severing the will’s choice from the practical intellect’s judgment, it is true that his philosophical position placed the autonomous power of will in sharp relief. For him, will retained free choice even after judgment, its actions entirely uncoerced by what intellect had decided. Although this did not mean that a free will would arbitrarily disregard intellect’s advice, Henry believed it did demand a distinctly non-Aristotelian account of causality in voluntary acts. The will could not act until it was presented with an object by the mind, but it was not the object, the mind’s judgment or any affect

triggered by the intellect that caused the will to act as it did. Instead, the will was entirely self-moved, endowed by God with an immanent power to go from potency to act (see Free will).

Henry went so far as to reject the Thomistic distinction between a primary freedom to act (*libertas exercitii*), independently available to the will, and a secondary freedom to act in a specific way (*libertas specificationis*), constrained by intellect's judgment. Instead, he claimed that the will was always free not just to act or not to act but also to specify the act it would perform. Anything short of this he took to be intolerably deterministic (Macken 1977b: 146-7, 158-9). As with other areas of his philosophy, Henry's vision of the will laid the groundwork for several Augustinian currents in the fourteenth century, feeding into the quite distinct moral theories of Duns Scotus and Ockham.

STEVEN P. MARRONE

List of works

Henry of Ghent (*c*.1275-93) *Opera omnia (Complete Works)*, Louvain: University Press, 1979-.(Published to date are *Quodlibeta* I, II, VI, VII, IX, X, XII, XII q. 31, XIII; *Summa* art. 31-40; *Lectura ordinaria super Sacram Scripturam (Ordinary Lectures on the Bible)*.)

Henry of Ghent (1275-93) *Summa quaestionum ordinariarum (Collected Ordinary Questions)*, Paris: Badius, 1520; St. Bonaventure, NY: Franciscan Institute, 1953, 2 vols.(Intended as an exhaustive exposition of Henry's philosophy and theology; never completed.)

Henry of Ghent (1276-92) *Quodlibeta*, Paris: Badius, 1518; Louvain: Bibliothèque S.J., 1961, 2 vols.(A translation of selected questions appears in R.J. Teske (1993) *Quodlibetal Questions on Free Will*, Milwaukee: Marquette University.)

References and further reading

Bérubé, C. (1983) *De l'homme à Dieu selon Duns Scot, Henri de Gand et Olivi (The Path from Man to God According to Duns Scotus, Henry of Ghent and Olivi)*, Rome: Istituto Storico dei Cappuccini.(Essays on the historical relation between Henry and some of his contemporaries and successors.)

Dumont, S. (1984) 'The quaestio si est and the Metaphysical Proof for the Existence of God according to Henry of Ghent and John Duns Scotus', *Franziskanische Studien* 66: 335-67.(Compares metaphysical foundations of Henry's and Scotus' theory of knowledge with that of Aristotle.)

Gómez Caffarena, J. (1958) *Ser participado y ser subsistente e la metafísica de Enrique de Gante (Participated Being and Subsistent Being in the Metaphysics of Henry of Ghent)*, Rome: Pontificia Università Gregoriana. (Detailed analysis of Henry's ontology.)

Hödl, L. (1964) 'Neue Nachrichten über die Pariser Verurteilungen der thomasischen Formlehre' (New Findings Concerning the Parisian Condemnations of Thomas' Theory of Form), *Scholastik* 39: 178-96.(On Henry's position in early debates over Thomas Aquinas.)

Macken, R. (1972) 'La théorie de l'illumination divine dans la philosophie d'Henri de Gand' (The Theory of Divine Illumination in the Philosophy of Henry of Ghent), *Recherches de Théologie Ancienne et Médiévale* 39: 82-112.(On the development of Henry's theory of knowledge.)

Macken, R. (1973) 'Les corrections d'Henri de Gand à ses Quodlibets' (Henry of Ghent's Corrections to his Quodlibets), *Recherches de Théologie Ancienne et Médiévale* 40: 5-51.(Critical for understanding textual tradition of Henry's works.)

Macken, R. (1977a) 'Les corrections d'Henri de Gand à sa Somme' (Henry of Ghent's Corrections to his Summa), *Recherches de Théologie Ancienne et Médiévale* 44: 55-100.(Critical for understanding textual tradition of Henry's works.)

Macken, R. (1977b) 'Heinrich von Gent im Gespräch mit seinen Zeitgenossen über die menschliche Freiheit' (Henry of Ghent in Dialogue with his Contemporaries over Human Freedom), *Franziskanische Studien* 59: 125-82.(Detailed analysis of Henry on the will, comparing him with contemporaries.)

Marrone, S.P. (1985) *Truth and Scientific Knowledge in the Thought of Henry of Ghent*, Cambridge, MA: Medieval Academy.(On the relation between Henry's changing theory of knowledge and his metaphysics.)

Marrone, S.P. (1988) 'Henry of Ghent and Duns Scotus on the Knowledge of Being', *Speculum* 63: 22-57.(Ties Henry's views on natural knowledge of God to Scotus' position on univocity of being and proof for God's existence.)

Nys, T.V. (1949) *De werking van het menselijk verstand volgens Hendrik van Gent (The Action of the Human Intellect According to Henry of Ghent)*, Louvain: Nauwelaerts. Abbreviated Latin version: *De psychologia*

cognitionis humanae secundum Henricum Gandavensem, Rome: Pontificia Università Gregoriana.(Traces Henry's abandonment of intelligible species.)

Paulus, J. (1935-6) 'Henri de Gand et l'argument ontologique' (Henry of Ghent and the Ontological Argument), *Archives d'Histoire Doctrinale et Littéraire du Moyen Age* 10-11: 265-323.(On metaphysical implications of Henry's proof for existence of God.)

Paulus, J. (1938) *Henri de Gand. Essai sur les tendances de sa métaphysique (Henry of Ghent: Essays on the Contours of his Metaphysics)*, Paris: Vrin.(Most complete study of Henry's metaphysics.)

Porro, P. (1990) *Enrico di Gand. La via delle proposizioni universali (Henry of Ghent: The Proof from Universal Propositions)*, Bari: Levante.(Analysis of ontological and theological significance of Henry's views on natural knowledge of God.)

Henry of Harclay (*c*.1270-1317)

An English philosopher of the early fourteenth century, Harclay moved away from the position of Duns Scotus on the extramental existence of universals and towards the more conceptualist or nominalist stance of William of Ockham. On questions of infinity and continuity, Henry was strongly anti-Aristotelian, holding that there were numbers that were actually infinite and not all equal to each other, and that a continuum was composed of an actual infinity of indivisibles, by which it was properly measured. His position came under powerful mathematical attack.

Henry of Harclay, a member of a noble family from the North of England, received his education in arts at Oxford University and afterwards studied theology at Paris. He was much influenced by Duns Scotus, then teaching at Paris. In 1312, having returned to Oxford, he was elected and confirmed (by the Bishop of Lincoln) as Chancellor of the University. In this capacity in 1317 he journeyed to the papal court at Avignon in pursuance of a university dispute with the Dominican order, where he died.

In his commentary on the *Sentences* of Peter Lombard, representing his lectures at Paris in the course of attaining his theological qualifications, Harclay's teaching was very close to that of Duns Scotus, and he is counted among the first of Scotus' expositors. Later he adopted a more independent position, which has been seen as leading him in the direction of William of Ockham. His later thought is revealed in a number of disputed questions, whose contents have been rather spasmodically preserved in different manuscript codices. In one codex, pride of place is taken by a question as to whether astrologers or other 'calculators' can foretell the time of Christ's Second Coming. In this, Henry may be seen as upholding the claims of rational philosophy against the more mystical and inspirationist views of Arnald of Villanova, although he still allowed that the prophetic abbot Joachim of Fiore had made some true predictions.

On the vexed question of universals, Henry was concerned to diminish their extramental existence (see Universals). He held that each individual thing outside the mind could move the intellect to conceive it either confusedly or distinctly. By the latter conception the mind could distinguish Plato from Socrates, but from the former it received the general notion of man. Harclay's somewhat woolly, if not confused, view was attacked from the realist side by Walter Burley, and from the nominalist camp by William of Ockham, who said that the opinion was false, unintelligible, and neither true nor logical. In the case of relations also, Henry veered towards conceptualism, holding that, if *A* was related to *B*, this did not mean that the relation *inhered* in *A*. This in turn implied that God could be really related to his creatures, since no change was entailed in him. On the question of the univocality of being, Harclay held that there is one kind of being in both God and creatures, and in this he was close to Scotus (although with several differences of nuance) and fervently opposed to Thomas Aquinas, whose positions he often explicitly attacked.

It was in his views on continuity, however, that Henry went most strongly against the mainstream of scholastic philosophy. In these views, Henry built on the opinions of Robert Grosseteste, who had held that a line was properly measured by the number of points that it contained (and it would seem, although some scholars disagree, of which it was composed). To humans such numbers were infinite, but to God they were finite, and they could have between them both rational and irrational ratios. Harclay developed his views in two questions on the eternity of the world, concerning respectively its beginninglessness and its endlessness. Quite probably no schoolman held the heretical position that the world had actually existed from eternity, but there was controversy as to whether the opposite position could be proved by natural reason (see Eternity of the world, medieval views of). Those who held that it was impossible for there to have been an infinite past time were wont to cite paradoxes, such as that, if there had been an infinite past time, then there would have been the same number of past revolutions of the Sun as of the Moon, and in general that a whole would be equal to its part. Harclay's principal strategy in rebutting these claims was to assert that there could be actually infinite numbers, and that one of these could be greater than another. This was contrary to the conventional view that numbers were only potentially infinite, in the sense that there was no limit to how far one could go on counting; the question of comparing different infinite numbers did not then arise.

Harclay extended his arguments into questions concerning the structure of continua, holding that each was

composed of an infinite number of indivisibles, so that a line was made up of an infinite number of points. This again was opposed to the more holistic Aristotelian view that a continuum was infinitely divisible potentially (again in the sense of unlimited acts of division), but always into finite parts. It was certainly not composed of indivisibles, and was only measured by fictive indivisibles, as in treating feet as indivisibles for the measurement of lines. Harclay's position, like that of those who would compose continua out of finite numbers of indivisibles, gave rise to a host of difficulties. Among these was the question of how indivisibles could touch each other, for as Aristotle said, 'Since indivisibles have no parts, they must be in contact with one another as whole with whole' (*Physics* VI.1, 231b3). If that was the case, however, it was hard to see how they could compose something continuous bigger than themselves. Harclay's reply was that they did touch whole to whole, but according to a different position (*secundum distinctum situm*). Not surprisingly, such responses did not convince everyone, and among many attacks on indivisibilist positions, a devastating mathematically-based critique was made by Thomas Bradwardine, who cited Henry by name. However, Bradwardine did admit that he himself could be accused of a *petitio principii* by assuming the truth of the principles of Euclidean geometry.

See also: Cantor's theorem; Duns Scotus, J.

GEORGE MOLLAND

List of works

Henry of Harclay (before 1310) *Commentarius in I et II (?) Sententiarum (Commentary on Books I and [perhaps] II of the Sentences of Peter Lombard).*(A discussion, in Latin, with reference to the manuscript basis, can be found in C. Bali¢, (1959) 'Henricus de Harcley et Ioannes Duns Scotus',*Mélanges offerts à Étienne Gilson*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, and Paris: Vrin, 93-121.)

Henry of Harclay (*c*.1313) *Utrum mundus potuit fuisse ab eterno (Whether the World Could Have Existed From Eternity)*, ed. R.C. Dales, 'Henricus de Harclay: Quaestio "Utrum Mundus Potuit Fuisse ab Eterno"', *Archives d'Histoire Doctrinale et Littéraire du Moyen Age* 50, 1983: 223-55.

Henry of Harclay (*c*.1313) *Utrum universale significet aliquam rem extra animam, aliam a singulari vel supposito (Whether a Universal May Signify Some Thing Outside the Soul Other Than a Singular or Suppositum)*, ed. G. Gál, 'Henricus de Harclay: Quaestio de Significato Conceptus Universalis', *Franciscan Studies* 31, 1971: 178-234.(Includes Latin discussion of the text.)

Henry of Harclay (*c*.1313) *Utrum praedestinatio aeterna qua Deus praedestinavit aliquem ad gloriam et reprobatio aeterna qua Deus aliquem reprobavit ponant in praedestinato vel in reprobato necessitatem aliquam respectu salutis vel repectu damnationis (Whether the Eternal Predestination by which God Predestined One to Glory, and the Eternal Condemnation by which God Condemned Another, imply any Necessity in the Predestined or in the Condemned with Respect to Salvation or Damnation)*, ed. M.G. Henninger, 'Henry of Harclay's Questions on Divine Prescience and Predestination', *Franciscan Studies* 40, 1980: 167-243.(Henninger provides editions of this and the following question.)

Henry of Harclay (*c*.1313) *Utrum in praedestinato sit aliqua causa praevisa a Deo quare iste sit praedestinatus et ille reprobatus (Whether in the Predestined there is any Cause Foreseen by God Whereby One should be Predestined and the other Condemned)*, ed. M.G. Henninger, 'Henry of Harclay's Questions on Divine Prescience and Predestinatio', *Franciscan Studies* 40, 1980: 167-243.(Henninger provides editions of this and the previous question.)

Henry of Harclay (*c*.1313) *Utrum Dei ad Creaturam sit Relatio Realis (Whether there is a Real Relation of God to a Creature)*, ed. M.G. Henninger, 'Henry of Harclay's Question on Relations', *Mediaeval Studies* 49, 1978: 76-123.(Edition of the question.)

Henry of Harclay (*c*.1313) *Utrum Deo et creaturis aliquid sit commune univocum (Whether Anything is Univocally Common to God and Creatures).*(A discussion, without edition, may be found in A. Maurer, 'Henry of Harclay's Question on the Univocity of Being', *Mediaeval Studies* 16, 1954: 1-18.)

Henry of Harclay (*c*.1313) *Utrum aliud a Deo sit simpliciter necesse est (Whether Anything other than God Exists in an Unqualifiedly Necessary Way)*, ed. A. Maurer, 'Henry of Harclay's Questions on Immortality', *Mediaeval Studies* 19, 1957: 79-107.(Edition includes two questions on immortality.)

Henry of Harclay (*c*.1313) *Utrum anima intellectiva sit immortalis (Whether the Intellective Soul is Immortal)*, ed. A. Maurer, 'Henry of Harclay's Questions on Immortality', *Mediaeval Studies* 19, 1957: 79-107.(Edition includes two questions on immortality.)

Henry of Harclay (*c*.1313) *Utrum ad hoc quod Deus cognoscat alia a se oportet ponere in Deo relationes rationis ad absoluta cognita, quae sunt ideae (Whether for God to Know Things other than Himself it is Necessary to Posit in God Relations of Reason to Things Absolutely Known, which are Ideas)*, ed. A. Maurer, 'Henry of

Harclay's Questions on the Divine Ideas', *Mediaeval Studies* 23, 1961: 163-93.(Edition includes two questions on divine ideas.)

Henry of Harclay (*c*.1313) *Utrum ad distinctam cognitionem quam habet Deus ab aeterno de rebus creabilibus requiruntur in eo distinctae rationes cognoscendi, per quas cognoscit ipsas res creabiles (Whether for the Distinct Cognition that God has from Eternity of Creatable Things there are Required in Him Distinct Reasons of Knowing, by which He Knows those Creatable Things)*, ed. A. Maurer, 'Henry of Harclay's Questions on the Divine Ideas', *Mediaeval Studies* 23, 1961: 163-93.(Edition includes two questions on divine ideas.)

Henry of Harclay (*c*.1313) *Utrum in homine sit aliqua forma sustantialis praeter intellectivam (Whether there is any Substantial Form in the Human Being besides the Intellective)*, ed. A.Maurer, 'Henry of Harclay's Disputed Question on the Plurality of Forms', in J.R. O'Donnell (ed.) *Essays in Honour of Anton Charles Pegis*, Toronto,Ont.: Pontifical Institute of Medieval Studies, 1974, 125-159.(Edition of Henry's question on the plurality of forms.)

Henry of Harclay (*c*.1313) *Utrum mundus potuit fuisse ab eterno (Whether the World could have Existed from Eternity)*.(Discussion of the manuscripts and extensive analysis of the doctrines can be found in J.E. Murdoch, 'Henry of Harclay and the Infinite', *Studi sul XIV Secolo in Memoria di Anneliese Maier*, ed. A. Maierù and A. Paravicini Bagliani, Rome: Storia e Letteratura, 1981, 219-261.)

Henry of Harclay (*c*.1313) *Utrum mundum poterit durare in eternum a parte post (Whether the World could Endure to Eternity Hereafter)*.(Discussion of the manuscripts and extensive analysis of the doctrines can be found in J.E. Murdoch, 'Henry of Harclay and the Infinite', *Studi sul XIV Secolo in Memoria di Anneliese Maier*, ed. A. Maierù and A. Paravicini Bagliani, Rome: Storia e Letteratura, 1981, 219-261.)

Henry of Harclay (*c*.1313) *Utrum astrologi vel quicumque calculatores possint probare secundum adventum Christi (Whether Astrologers or any Calculators could Prove the Second Coming of Christ)*, ed. F. Pelster, 'Die Quaestio Heinrichs von Harclay über die Zweite Ankunft Christi und die Erwartung des baldingen Weltendes zu Anfang des XIV Jahrhunderts', *Archivio Italiano per la Storia della Pietà* 1, 1951: 27-82.(Edition of the question.)

References and further reading

Adams, M.M. (1982) 'Universals in the Early Fourteenth Century', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 411-39.(Places Henry's views on universals in the context of late medieval debates.)

Aristotle (*c*.mid 4th century BC) *Physics*, trans. R.P. Hardie and R.K. Gaye in W.D. Ross (ed.) *The Works of Aristotle Translated into English*, Oxford: Clarendon Press, 1930, vol. 2.(Includes Aristotle's writings on indivisibles, discussed by Harclay.)

Dales, R.C. (1984) 'Henry of Harclay on the Infinite', *Journal of the History of Ideas* 45: 295-302.(Discusses the issue with reference to the traditions from which Henry drew.)

Henninger, M.G. (1989) *Relations: Medieval Theories 1250-1325*, Oxford: Clarendon Press.(Devotes a chapter to Henry's views on the subject.)

Maier, A. (1964) *Ausgehendes Mittelalter: Gesammelte Aufsätze zur Geistesgeschichte des 14. Jahrhunderts (The Closing of the Middle Ages: Collected Articles on the Intellectual History of the Fourteenth Century)*, Rome: Storia e Letteratura, vol. I.(Chapter 2 is concerned with early fourteenth-century discussions of the actual infinite, and Chapter 11 contests BaliČcent;'s ascription of a commentary on Book II of the *Sentences* to Henry.)

Murdoch, J.E. (1982) 'Infinity and Continuity', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 564-91.(Valuable for placing Henry's views on the subject in their historical context.)

Pelster, F. (1924) 'Heinrich von Harclay, Kanzler von Oxford, und seine Quaestionen' (Henry of Harclay, Chancellor of Oxford, and His Questions), *Miscellanea Francesco Ehrle*, Rome: Biblioteca Apostolica Vaticana, vol. I, 307-56.(Still the fundamental overall study.)

Thijssen, J.M.N.H. (1990) 'The Response to Thomas Aquinas in the Early Fourteenth Century: Eternity and Infinity in the Works of Henry of Harclay, Thomas of Wilton and William of Alnwick OFM', in J.B.M. Wissink (ed.) *The Eternity of the World in the Thought of Thomas Aquinas and his Contemporaries*, Leiden: Brill, 82-100.(Pays particular attention to Henry's stance towards Aquinas.)

Heraclides of Pontus (4th century BC)

Heraclides, a pupil of Plato, was roughly contemporaneous with Aristotle. Best known in antiquity as a writer of dialogues on moral and religious themes, he also held interesting views about cosmology and the structure of matter. Only fragments of his works survive.

Heraclides was probably born before 385 BC at Heraclea Pontica, on the Black Sea coast of modern Turkey, and lived into the penultimate decade of the fourth century. He received his philosophical training in Plato's Academy, where he may also have been influenced by Speusippus and Aristotle. Heraclides seems to have remained at the Academy under Plato's successor Speusippus, but retired to his home town after being defeated in the contest to succeed him.

Like Aristotle and other pupils of Plato, Heraclides made a transcript of Plato's lecture *On the Good* and wrote independent works on 'dialectic', moral and political philosophy, and religion, as well as literary and historical subjects. Today he is best known for two of his scientific theories. First, he combined a belief in the infinity of the universe with a system of planetary movements partly anticipating that of Tycho Brahe (see Kepler, J.): the earth rotates about its axis from west to east in one synodic day; the sun circles the earth from west to east in a year, presumably in the plane of the ecliptic; and the inner planets, Mercury and Venus, revolve about the sun on epicycles. We are not told his views on the movements of the outer planets.

Second, he believed that matter is made up of two kinds of particle separated by microvoids: 'corpuscles' (*onkoi*) and 'fragments'. Corpuscles themselves are constituted of 'fragments' (*thrausmata*) and may be broken down into fragments; fragments, however, cannot be broken down any further. The corpuscles have at least some of the sensible qualities of macroscopic bodies, but the fragments do not; thus the corpuscles are analogous to the atoms of modern chemistry, the fragments to protons and electrons.

Both theories were inspired by Plato's *Timaeus* (see Plato §16). In this text, the movements of Mercury and Venus were described, but not explained (38d); an axial rotation of the earth was thought by many to be mentioned at 40b-c; and Plato introduced a theory according to which each of the four elements consists of minimum particles shaped like one of the regular solids, and these in turn are constituted of and can be broken down into two kinds of triangle (54d-57c). Heraclides substituted three-dimensional 'fragments' for Plato's two-dimensional triangles; instead of seeing the elementary particles as geometrical constructs, he treated them as aggregates of matter. But his postulating two stages of aggregation distinguished his theory from other contemporary atomic doctrines, and this came to him from Plato.

There was a similar development in his doctrine of the soul. Like Plato, he believed that the soul is immortal and subject to reincarnation, but whereas Plato regarded it as incorporeal, Heraclides taught that it is constituted of 'light', the finest kind of matter; between incarnations, souls not wicked enough to be sent to hell congregate in one part of the heavens, where they form the Milky Way. With this doctrine we leave the field of science and come to that on which Heraclides' popularity rested: ethical and religious dialogues. His outlook was puritanical. He believed that the desire for pleasure corrupts both individuals and societies, with dire results; for the gods observe human behaviour and ensure that crimes are punished. His dialogues were full of such stories, and also revelations, including eschatological ones, miracles and sermons by ancient wise men including Empedocles, Pythagoras and the Scythian *shaman* Abaris. They were heady stuff and found readers until the end of antiquity.

Heraclides left no school. Cicero (§1) admired his style and used his writings as a model for his own dialogues, but the only reader to take his ideas seriously was the physician Asclepiades of Bithynia (see Hellenistic medical epistemology). Asclepiades work was based on Epicurean conceptions, but he adopted Heraclides' corpuscular theory; presumably he found it useful for explaining physiological processes.

H.B. GOTTSCHALK

References and further reading

Gottschalk, H.B. (1980) *Heraclides of Pontus*, Oxford: Clarendon Press.(Includes English translations of most of the important fragments and a full bibliography; in the appendices some new texts are added and others are

shown to be spurious.)
Plato (*c*. 366-60 BC) *Timaeus*, trans. F.M. Cornford, *Plato's Cosmology*, London: Routledge & Kegan Paul, 1937. (Plato's theory of elements is at 53b-6c, his astronomical system at 34a-40d.)
Wehrli, F. (1953) *Die Schule des Aristoteles (The Aristotelian School)*, vol. 7, Basle: Schwabe; repr., 2nd edn, 1969.(Greek text of Heraclides' main fragments, bibliography and German commentary.)

Heraclitus (*c*.540-*c*.480 BC)

No Greek philosopher born before Socrates was more creative and influential than Heraclitus of Ephesus. Around the beginning of the fifth century BC*, in a prose that made him proverbial for obscurity, he criticized conventional opinions about the way things are and attacked the authority of poets and others reputed to be wise. His surviving work consists of more than 100 epigrammatic sentences, complete in themselves and often comparable to the proverbs characteristic of 'wisdom' literature. Notwithstanding their sporadic presentation and transmission, Heraclitus' sentences comprise a philosophy that is clearly focused upon a determinate set of interlocking ideas.*

*As interpreted by the later Greek philosophical tradition, Heraclitus stands primarily for the radical thesis that 'Everything is in flux', like the constant flow of a river. Although it is likely that he took this thesis to be true, universal flux is too simple a phrase to identify his philosophy. His focus shifts continually between two perspectives - the objective and everlasting processes of nature on the one hand and ordinary human beliefs and values on the other. He challenges people to come to terms, theoretically and practically, with the fact that they are living in a world 'that no god or human has made', a world he describes as 'an ever-living fire kindling in measures and going out in measures' (fr. 30). His great truth is that 'All things are one', but this unity, far from excluding difference, opposition and change, actually depends on them, since the universe is in a continuous state of dynamic equilibrium. Day and night, up and down, living and dying, heating and cooling - such pairings of apparent opposites all conform to the everlastingly rational formula (*logos*) that unity consists of opposites; remove day, and night goes too, just as a river will lose its identity if it ceases to flow.*

Heraclitus requires his audience to try to think away their purely personal concerns and view the world from this more detached perspective. By the use of telling examples he highlights the relativity of value judgments. The implication is that unless people reflect on their experience and examine themselves, they are condemned to live a dream-like existence and to remain out of touch with the formula that governs and explains the nature of things. This formula is connected (symbolically and literally) with 'ever-living fire', whose incessant 'transformations' are not only the basic operation of the universe but also essential to the cycle of life and death. Fire constitutes and symbolizes both the processes of nature in general and also the light of intelligence. As the source of life and thought, a 'fiery' soul equips people to look into themselves, to discover the formula of nature and to live accordingly.

The influence of Heraclitus' ideas on other philosophers was extensive. His reputed 'flux' doctrine, as disseminated by his follower Cratylus, helped to shape Plato's cosmology and its changeless metaphysical foundations. The Stoics looked back to Heraclitus as the inspiration for their own conception of divine fire, identifying this with the logos *that he specifies as the world's explanatory principle. Later still, the neo-Pyhrronist Aenesidemus invoked Heraclitus as a partial precursor of scepticism.*

1 Life and work

Heraclitus appears to have spent his life in Ephesus, which had been founded as a Greek colony some 200 years before his birth. According to ancient biography he was an arrogant and surly aristocrat, given to eccentric behaviour, but these anecdotes are largely a fictional construction built out of his own words, in which the tone he adopts in relation to other people is contemptuous. Rather than viewing this as a psychological trait, it is better to treat it as an extreme instance of the way early Greek poets and sages claimed authority for their work. Heraclitus, however, is exceptional in the explicit contempt he expresses for such hallowed authorities as Homer and Hesiod, and also for the contemporary intellectuals Xenophanes, Hecataeus and Pythagoras. He may have been on bad terms with his fellow citizens for political reasons, including perhaps support he received from King Darius of Persia, and it is likely that he was opposed to the democratic constitutions some Greek communities were beginning to adopt.

Although Heraclitus presents himself as uniquely enlightened, he was clearly familiar with the leading thinkers of his time. He draws attention to the relativity of judgments and the difference between humans and animals in ways that recall Xenophanes' critique of religious beliefs (see Xenophanes §3). He almost certainly knew and rejected Pythagoras' doctrine of the transmigration of souls (see Pythagoras §2). His cosmology is both indebted to and a criticism of Milesian science: the criticism appears particularly in his denial of the world's beginning, but his focus

on the law-like processes of nature has clear affinities with Anaximander's celebrated doctrine of cosmic justice (see Anaximander §4).

Heraclitus' work does not survive as a continuous whole. What we have instead is a collection of more than 100 independent sentences, most of which are *ad hoc* citations by authors from the period AD 100-300. Plato and Aristotle rarely cite Heraclitus directly, but their interpretations of him, which are influenced in part by their own preconceptions, shaped the ancient tradition of Heraclitus as exponent of universal flux and of fire as the primary material. Interpretation of Heraclitus is further complicated by the work of his professed follower Cratylus, and still more so by the way Stoics and Pyrrhonists looked back to him as a precursor of their own philosophies (see §6). This afterlife is important as an indication of Heraclitus' complexity and capacity to influence a range of very different thinkers, but modern interpretation of him rightly treats it as secondary to the evidence of his own words.

Heraclitus is credited with writing a book, *On Nature*, and depositing it in a temple of Artemis. Some scholars believe that the book was a later compilation of sayings that he never wrote down, but while many of the fragments are well suited to oral delivery, this scarcely applies to the longest one, with which he is said to have begun his book:

> Of this formula (*logos*) that is so always people are uncomprehending, both before hearing it and once they have heard it. For although all things come to pass in accordance with this formula, people seem lacking in experience even when they experience such words and deeds as I expound, when I distinguish each thing according to its nature and explain how it is. But other people are oblivious of what they do awake just as they forget what they do asleep.
>
> (fr. 1)

2 Methodology

Heraclitus presents himself as the deliverer of a *logos* (see Logos). Any speaker could make a similar claim, but Heraclitus' *logos* is not his personal message or thought. Rather, Heraclitus views himself simply as a conduit for *the logos*, *the* means by which his audience will learn the objective truth about everything. In characterizing the *logos*, Heraclitus claims that 'it is so always'; it is the world's rationale or determining formula, the key to each thing's nature.

The *logos* is accessible to thought and linguistic expression because it is 'common' or 'public'; but 'most people live as though their thinking were private'. Heraclitus uses this contrast, and the contrast between waking and sleeping, to signal the gap between experience people have in common and the erroneous or purely subjective opinions they typically base on that experience. Taken as a set of general truths, the *logos* includes such propositions as the following: 'All things are one', 'All things are an exchange for fire, and fire for all things', 'Nature loves to hide', 'War is father of all and king of all', 'The way up and down is one and the same', 'God is day night, winter summer, war peace, satiety hunger'. Heraclitus does not present arguments for these cryptic statements, nor does he elucidate them. His Delphic style is a deliberate provocation of thought, a way of rousing his audience to solve his riddling account of nature and to discover the *logos* for and in themselves. In order to do so, he suggests, they need to rid themselves of naïve empiricism, to 'expect the unexpected', to connect things that they normally keep separate, to practise self-scrutiny, self-knowledge - in short, they need to discover the 'depth of the soul's own *logos*', which presumably signifies the mind's unlimited capacity to arrive at a complete understanding of the *logos* of nature.

In addition to general statements of the kind illustrated above, Heraclitus uses many other kinds of sentence. These include comments on the lack of wisdom manifested by the supposedly learned as well as by ordinary people, graphic statements concerning the 'unity of opposites' and the relativity of values, instruction concerning the 'measures' of physical change, characterizations of the soul, and numerous injunctions about the wisdom, lawfulness and piety requisite for a good life. One ancient commentator described his book as consisting of three topics: the universe, politics and theology. Another called it 'a guide to conduct', and it was also described as being about 'society, with the treatment of nature being metaphorical'. Such attempts at systematizing Heraclitus are anachronistic, but they have the merit of recognizing his complexity. He was, as he describes himself, an inquirer into the 'nature' of things, but it would be no less correct to call him a moralist and psychologist; what he

derived from his inquiry was an urgent message about the need for human beings to refashion their values and their outlook in accordance with the implications of the *logos*.

3 Unity of opposites and perspectivism

In emphasizing the ‘unity of all things’, Heraclitus was not saying, as his Milesian predecessors had done, that every natural object is the manifestation of a single material nature such as water or air (see Anaximenes). Nor was he denying the obvious phenomenal difference between hot and cold, light and dark, living and dead, up and down. Heraclitus’ world includes the normal range of things that are different from one another. What makes all things one is the fact, as he sees it, that a common formula (*logos*) is at work in everything to which we attribute spatial and/or temporal identity and continuity. Such opposites as those just enumerated are not separate or separable from one another but are co-dependent contributors to the identity of things. In any day (regarded as twenty-four hours), light and dark are combined, as are up and down in any road. Nothing can be alive that will not die, and nothing can be dead that has not lived. If one of these opposites is removed, the other must go too.

Apart from cases where the co-presence or successive presence of opposites is *obviously* essential to things, Heraclitus drew support for the unity of opposites from examples where it could be inferred even though it was less obvious. In the bow and the lyre, he said, there is a ‘back-stretched structure’ (or ‘back-turning structure’ on an alternative reading). The effectiveness of both instruments depends upon tension between a piece of wood and a taut string. These components pull in opposite ways, and thereby they generate the unitary objects, lyre and bow. His famous image, ‘You can’t step twice into the same river’, is probably to be interpreted in a similar way. What gives the river its identity is the continuous flow of successively different water. If ‘All things flow’ is a statement that Heraclitus approved, it is best interpreted as his image for the claim that even apparently stable beings depend for their identity on the interchange or succession of their constituent parts or on conflicting forces. The image need not, and probably should not, be regarded as a claim that nothing has a stable identity, but rather that such stability as things do have is derived from the equipollence of their constitutive opposites.

The unity of opposites formula is also applicable to Heraclitus’ account of the way physical processes occur in the world at large. Changes of temperature and liquidity are constantly taking place, but they do so according to measure and balance. In the cosmic context, ‘All things are one’ obtains because a single thermodynamic principle (as it might be called now) is operative throughout the system: ‘an ever-living fire, kindling in measures and going out in measures’ (fr. 30). Another name for this principle, and another view of the unity of opposites, are evident in the statement that ‘God is day night, winter summer, war peace, satiety hunger. He alters as when [fire] mingled with spices is named according to each person’s pleasure’ (fr. 67). Two implications might be drawn from this passage. One is that, from a divine perspective, the contrary evaluations normally accorded to day *or* night, to war *or* peace, are transcended. The other is that human discriminations between these opposites are purely arbitrary.

Heraclitus certainly endorsed the first implication, but did so for the second only with qualification. The passage does indeed suggest a radical difference between the divine and the human, but Heraclitus’ point is probably that people need to adjust those of their perspectives that are purely personal or subjective and bring them into line with the objective way things are: it will always be the case that sea water is fair *and* foul, ‘fair for fish and foul for human beings’ (fr. 61). It is equally and necessarily the case that non-human animals have different preferences from humans. Yet the unity of opposites is a principle that also requires radical reordering of conventional opinions. Because the world is a dynamic interplay of opposing forces, ‘people must realize that war is common and justice is strife’ (fr. 80). In this difficult saying, Heraclitus is not recommending belligerence and the breakdown of civil society. His point is rather that people must come to terms with the fact that conflicting forces are basic to the way the world is structured - are in fact essential to its balance and order. That is how inanimate physical powers interact. He seems to have inferred, by analogy, that competition, involving winners and losers, is no less essential when accounting for the differences of status within his own day-to-day world. Hence the assertion that: ‘War is father of all and king of all; and some he has shown as gods, others as humans; some he has made slaves and others free’ (fr. 53).

4 Cosmology

Judging from the extant fragments, Heraclitus devoted only a small fraction of his philosophy to cosmology and explanation of physical phenomena. Many of his ancient interpreters, however, starting with Aristotle, interpreted

him as if this were his main purpose. Working from texts in which Heraclitus refers to the transformations of 'ever-living' fire, they concluded that he endorsed the following doctrines: first, that fire is the underlying principle of all things; second, that particular phenomena are composed of and resolved into fire by rarefaction and condensation; and third, that the world as a whole is an everlastingly repeated sequence of temporally limited cycles, each of which emerges out of fire and ends in fire. These 'doctrines', or at least the first and the third of them, are generally treated today as anachronistic attempts to fit Heraclitus' ideas into a pattern more applicable to other thinkers. Although this modern diagnosis is justified in part, ancient philosophers were clearly right in thinking that Heraclitus regarded fire as the world's primary and most explanatory phenomenon. The problem is to understand what he intended by this claim.

Some of his statements suggest that fire is what one should start from and end with in any reflection on natural change. He certainly thought that what would later be called elements or elementary qualities (hot, cold, moist and dry) are interdependent phases of a continuous cycle. Hence his use of statements such as 'The cold gets warm, the warm gets cold, the moist gets dry, the dry gets damp' (fr. 126). This kind of insight is obviously suitable as an illustration of the way 'All things flow' and the unity of opposites, but it does not explain why he said that 'All things are an exchange for fire, and fire for all things, as goods for gold and gold for goods' (fr. 90). What is it that privileges fire in the cycle of change and makes it the one appropriate exchange symbol for everything else?

The answer, authorized by the Aristotelian tradition, is that fire is *the* underlying principle or *the* element. An underlying principle or element should be the unchanging foundation of everything that changes, but Heraclitus explicitly denies that fire is like this. Rather, it is continuously changing, 'kindling in measures and going out in measures'. What is unchanging about Heraclitus' fire is the *way* it changes - according to 'measures'. Heraclitus uses the term 'measures' to indicate cosmic order, balance, proportionality, natural law. He probably inferred by observation that, of all natural phenomena, fire is not only the most dynamic but also the one that is most obviously self-regulating. A fire consumes all the material that is available to it, nothing more and nothing less. What it consumes it also changes. It lives, as it were, by destroying something else; or, viewed from another perspective, it destroys itself by creating something else.

Heraclitus' fire, then, is not merely *one* phase in the cycle of natural changes. As the world's 'currency' or exchange symbol, the sum of fire's activity is completely commensurate with the sum of everything else. Fire's gain is everything else's loss and vice versa. The reciprocity or exchange between fire and the rest of things is Heraclitus' principle of cosmic order. There is a clear similarity here to Anaximander, who invoked 'crime and retribution' as the model to explain how change (hot to cold, wet to dry, and so forth) is balanced and reciprocal. Heraclitus was clearly impressed by his predecessor's bold appeal to 'justice' as the regulating principle of natural processes. But he was entirely original in treating fire as the primary process of nature. Given the conceptual resources available to him, nothing could have served him better as the way to indicate that nature is a dynamic system, that apparently stable things are also processes, that apparent unities are also polarities.

This, then, appears to be what Heraclitus meant by premising his cosmology on 'ever-living fire, kindling in measures and going out in measures'. He probably did not think, as ancient interpreters supposed, that our world is everlastingly recreated and destroyed, but rather, that nothing exists outside the system of fire and its exchanges. The system is self-regulating, and because there is nothing to interrupt its rhythm it persists for ever. Heraclitus is the earliest Greek thinker to postulate an everlasting world, and he is also probably the earliest to apply the term *kosmos* (meaning 'beautiful structure') to the world (see however Pythagoras §2).

Did Heraclitus identify his cosmic fire with the god to whom he sometimes referred? Did he also think of these as the physical correlate or referent of his *logos* formula? Precise answers to these questions are scarcely possible, but his Stoic interpreters were probably not far wrong in answering affirmatively for him. Although he evidently took a huge step in the direction of treating the universe as an autonomic field of counterbalancing forces, he probably also thought of it as a mind-directed system. Some of the fragments seem to allude to a divine plan or purposiveness. Still more suggestive are connections he drew between fire, soul, life and intelligence.

5 Psychology, ethics and religion

No fewer than ten of the fragments are statements about the Psychē. In its primary Greek usage this term signifies a human being's life, but Heraclitus uses the word not only in this way but also to signify mind and intelligence.

He made a major contribution to the idea (immensely important to Greek philosophy from Socrates and Plato onward) that cultivation of the *psychē* is the prerequisite for living well.

‘Eyes and ears’, he said, ‘are poor witnesses to those who have barbarian *psychai*’ (fr. 107). In this dense statement Heraclitus likens the unenlightened majority to ‘barbarians’ - that is, foreigners, who lack the language and culture of genuine Greeks. In order to live authentically, one needs a *psychē* that can interpret empirical evidence correctly - a *psychē* that understands the truths expressed by Heraclitus’ *logos*. The *psychē* as such, he suggests, has unlimited resources, but in order to draw on them it is necessary, as he has done, to ‘inquire into oneself’. Such observations indicate that Heraclitus’ philosophy was simultaneously outward-and inward-looking: his *logos* is both the formula of nature’s processes and the account of a mind or a self thinking and understanding that formula.

Heraclitus forges a link between these two aspects of his *logos* by invoking fire in remarks about the *psychē*. He associates life, intelligence and excellence with fire and dryness; by contrast, it is ‘death for souls to become water’, and a drunken man (that is, someone who has lost his bearings) ‘has a soul that is moist’. We are probably intended to take these statements both literally and symbolically. This complex reading fits Heraclitus’ statements about fire in cosmological contexts, where fire is both a constituent of nature and the currency for understanding all processes. It is also relevant in the case of the *psychē*. Living beings derive their vitality from the ever-living fire, but what a fiery soul contributes to humans is not merely life but the ‘light’ of intelligence. It is entirely in Heraclitus’ manner to exploit the multiple resonance of words, and to demand interpretation that is metaphorical as well as literal.

The intelligence that Heraclitus seeks to kindle should, he suggests, make people rethink their beliefs about life, death, religion and society. He attacks the folly of religious rituals, especially blood sacrifice; he recommends strict respect for civic law and order; he contrasts ‘the best’, who seek immortal fame, with the bovine satisfactions of the majority. He particularly emphasizes people’s inability in general to come to terms with death, suggesting that the way one dies has a bearing both on one’s worth as a person and also on one’s postmortem condition. Does this last point refer to some kind of postmortem existence or simply to a person’s reputation after death? The latter seems the more likely, given his emphasis upon the cyclical phases of natural processes. Nothing seems to have mattered more to Heraclitus than persuading people to regard themselves as integral parts of an ineluctable cosmic order. From this perspective, we should organize our lives in the realization that we are mortal but intelligent phases in the cosmic life of ever-living fire.

6 Influence

Heraclitus’ influence on subsequent philosophers was large and complex. It is probably first evident in Parmenides (§§3-8), whose argument for a wholly changeless and homogeneous reality reads often like a direct echo and refutation of Heraclitus. In response to the Heraclitean philosopher Cratylus, Plato was impressed by the thesis that phenomena are in constant flux, but he rejected this as an account of ‘intelligible’ reality on the grounds that what is in constant flux cannot be known. Aristotle, taking Heraclitus’ combinations of opposites to be a flagrant breach of the principle of non-contradiction, interpreted him as ‘making all statements true’. It seems not to have occurred to either Plato or Aristotle that Heraclitus’ *logos* could be interpreted as the changeless and knowable formula of a world that is incessantly but regularly changing.

Heraclitus’ most penetrating and positive interpreters were the early Stoic philosophers. He inspired basic features of their cosmology, especially their identification of the world’s ‘active principle’ with divine fire, and they also gave their own slant to the Heraclitean *logos*, which they interpreted as the ‘rationality’ embodied in divine fire (see Cleanthes; Stoicism §3). A reading of Cleanthes’ *Hymn to Zeus* is the best way to view Heraclitus’ influence on Stoicism. In that text, Cleanthes reflects Heraclitus not only in his references to fire and *logos* but also in the way he presents the divine principle as reconciling opposites and embodying the laws of nature.

A further aspect of Heraclitus’ influence is to be seen in Aenesidemus, the founder of neo-Pyrrhonist philosophy. Aenesidemus was in the habit of citing Heracliteanism as ‘a route to scepticism’, basing this claim on Heraclitus’ way of predicating opposites of the same subject.

See also: Presocratic philosophy

A.A. LONG

References and further reading

Barnes, J. (1982) *The Presocratic Philosophers*, London: Routledge & Kegan Paul, chaps 4 and 7.(A vigorous treatment, often critical of Heraclitus' logical acumen.)

Cleanthes (*c*.331-232 BC) *Hymn to Zeus*, in A.A. Long and D.N. Sedley, *The Hellenistic Philosophers*, Cambridge: Cambridge University Press, 1987, vol. 1, 326-7, vol. 2: 326-7. (Greek text and translation.)

Guthrie, W.K.C.(1962) *A History of Greek Philosophy*, vol. 1, Cambridge: Cambridge University Press, 6 vols. (Volume 1, chapter 7 gives a detailed and balanced treatment.)

Heraclitus (*c*.540-480 BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 1, 139-82. (The standard collection of the ancient sources; includes Greek texts with translations in German.).

Hussey, E. (1972) *The Presocratics*, London: Duckworth, ch. 3.(Excellent study comparing Heraclitus to Wittgenstein.)

Hussey, E. (1982) 'Epistemology and Meaning in Heraclitus', in M. Schofield and M. Nussbaum (eds) *Language and Logos*, Cambridge: Cambridge University Press, 33-60.(Subtle analysis of Heraclitus' interpretative methodology.)

Kahn, C. (1979) *The Art and Thought of Heraclitus*, Cambridge: Cambridge University Press. (Text, translation and commentary; a fundamental exposition, taking the condition of mortality to be Heraclitus' real subject.)

Kirk, G.S. (1962) *Heraclitus: The Cosmic Fragments*, Cambridge: Cambridge University Press.(Classic study, particularly valuable for assessing the authenticity of key fragments.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(A valuable survey of Presocratic philosophy, including texts and translations; Heraclitus is discussed in chapter 6.)

Long, A.A. (1975-6) 'Heraclitus and Stoicism', *Philosophia* 5-6: 133-56; repr. in A.A. Long, *Stoic Studies*, Cambridge: Cambridge University Press, 1996, 35-57.(Chapter 2 is a study of Heraclitus and Stoicism.)

McKirahan, R.D. (1994) *Philosophy Before Socrates*, Indianapolis, IN: Hackett.(Contains translation and helpful discussion of the fragments in chapter 10.)

Marcovich, M. (1967) *Heraclitus, Greek Text with a Short Commentary*, Merida: Los Andes University Press. (Important scholarly tool.)

Mondolfo, R. and Tarán, L. (1972) *Eraclito: Testimonianze e imitazioni* (Heraclitus: Testimonies and Imitations), Florence: La Nuova Italia.(Contains and comments on Greek texts referring to and imitative of Heraclitus.)

Robinson, T.M. (1987) *Heraclitus, Fragments*, Toronto, Ont.: Toronto University Press. (Includes Greek text of the fragments, with translation and commentary.)

Vlastos, G. (1955) 'On Heraclitus', *American Journal of Philology* 76: 337-68; repr. in *Studies in Greek Philosophy*, Princeton, NJ: Princeton University Press, 1995, vol. 1, 127-50. (Important study of the 'river' fragments.)

Wiggins, D. (1982) 'Heraclitus' Conceptions of Flux, Fire and Material Persistence', in M. Schofield and M. Nussbaum (eds) *Language and Logos*, Cambridge: Cambridge University Press, 1-32.(Sympathetic study by a major modern philosopher.)

Herbart, Johann Friedrich (1776-1841)

From 1798, Herbart developed a 'realistic' alternative to the idealistic philosophy of Kant, Fichte,Schelling andHegel. His theoretical philosophy, which centres around metaphysics and psychology, is sharply critical of the idealistic concept of subjectivity. His practical philosophy rests on ethics and educational theory, each of which presumes the existence of the other.

Herbart laid the foundations for his philosophy by critically examining Fichte's philosophy in the late 1790s. Like Kant, Fichte had distinguished between theoretical and practical philosophy. UnlikeKant, however, he believed he could pinpoint the factor uniting them. Fichte called this uniting factor 'the ego', 'pure, absolute self-consciousness' (see Fichte, J.G. §§3-5).

Herbart responded by stating that it is true that we understand the ego to be identical with self. It is also true that this term suggests itself forcibly to us. However, we cannot conceive of the ego as constituting an identity of this kind, because it is a contradictory concept in itself. If the ego is to be comprehended as an act of 'conceiving of oneself' or as 'conceiving of one's ego', it should be possible for us to know *that* the ego is conceiving something, but not *what* it is conceiving. Fichte's concept of the absolute ego is therefore circular and incomplete.

This leads on to the theoretical part of Herbart's philosophy and into the realms of metaphysics and psychology. For Herbart, metaphysics consists first of a search for and analysis of concepts which are given but not conceivable, because they are contradictory in themselves. Contradictions of this kind present themselves not merely in terms of the concept of 'ego', but also in the concept of an individual thing with many characteristics (the problem of inherence), the concept of change (the problem of causality), and in the concept of matter. In Herbart's view, the second task of metaphysics is to map out a path showing how such contradictions in thinking can be resolved. This task is performed by methodology. In general terms, an attempt is made to demonstrate that the given contradiction in concepts (for example, identity of ego subject and ego object) rests on manifold, composite premises. While these premises are not apparent in the given concepts, the concepts refer to them as a part or outcome. Herbart calls this analysis the 'relations method'. Finally, his view of metaphysics incorporates ontology and 'synechology'. The task of ontology as the science of being is to cast light on the characteristics and structure of being, in other words to determine what is real in the appearances of things. While scepticism may prevail in respect of what we perceive or experience in the world, the fact that we do actually experience something cannot be denied or arbitrarily declared invalid. Experience as a state cannot be negated. For Herbart this implies that being means at any one time an 'absolute position'. Being is positive, *per se* simple, and beyond all relationship. The world we know is based on a multiplicity of simple entities, the 'reals', which are self-sufficient and react to change ('pressure') by seeking self-preservation. Synechology, then, determines the metaphysical manifestations in time and space of the multitude of individual reals which are in themselves simple (line, level, space; movement, speed).

By drawing on the essentials of metaphysics, Herbart was at the same time able to lay the foundations for his psychology. The 'true' basis of the concept of the ego, which is recognized as being contradictory, resides in the fact that Fichte's formula of the conceiving ego being identical with the conceived ego actually refers to a multiplicity of changing states, for which the formulas must be presumed. Herbart's explanation of self-consciousness as the essential task of 'realistic' psychology is as follows. Human consciousness is a mass of different interacting presentations. A study of their interaction gives rise to 'statics and dynamics' of the mind. The conscious separation of the act of conceiving from what is conceived is important for the development of self-consciousness. These two series of 'presentations' are joined by the concept of 'self' which presents itself, above all, in linguistic terms. Just as we say that water carves out its river bed 'for itself', we also project ourselves as the 'subject' of the act of conceiving onto a 'self' in which the conceiving and the conceived ego are deemed to be one and the same thing.

Herbart's practical philosophy also flows from his criticism of ego. Just as there can be no pure ego, there can equally be no pure, transcendentally free will (see Fichte, J.G. §§3-5; Kant, I. §11). A pure will would be a will which does not desire anything. That is a contradiction, however, since we always want something.

Herbart deduces from this that reason does not express itself as will, but rather as an aesthetic judgment. In the sphere of ethics, reason does not pass judgment on natural or artistic beauty, but on the well-balanced nature of human decisions. The fundamental judgment here is that human beings should gain an insight into what is good by dint of their will. This judgment leads to the 'idea of inner freedom' as consisting of a unity between insight and will. Herbart answers the question of which insight is good by presenting four more aesthetic ideas: perfection, benevolence, right and equity. These five ideas are directed at the individual. Similarly, Herbart derives five other ideas for society. In passing its judgment on aesthetic grounds, reason cannot induce morality (as a unity between insight and will). The fact that individuals can follow the judgments of reason is not a matter for moral philosophy or reason itself, but for educational theory. The primary task of educational theory - of 'educative instruction' - is to ensure that the child gains an insight. Its second task is to strengthen the will of the child and lead it to follow its insights - the task of 'discipline'. Ethics is ineffectual without educational theory and is devoid of any moral reality, while educational theory without ethics is devoid of any goal or reason.

See also: Education, history of philosophy of §8; German idealism

ALFRED LANGEWAND

List of works

Herbart, J.F. (1887-) *Sämtliche Werke in chronologischer Reihenfolge (Complete Works)*, ed. K. Kehrbach, Langensalza: Beyer.(The standard complete works.)

Herbart, J.F. (1806a) 'Hauptpunkte der Metaphysik' (Chief Points of Metaphysics), in Herbart, 1887, vol. 2: 175-226.(Herbart's answer to the question put by Kant in his *Critique of Pure Reason*: 'How is cognition possible?'. A very compact survey of Herbart's theoretical philosophy.)

Herbart, J.F. (1806b) 'Allgemeine Pädagogik aus dem Zweck der Erziehung abgeleitet' (General Pedagogy Derived from the Aim of Education), in Herbart, 1887, vol. 2: 1-139.(Herbart's solution of the practical problem of how to make moral ideas real - by education.)

Herbart, J.F. (1808) 'Allgemeine praktische Philosophie' (General Practical Philosophy), in Herbart, 1887, vol. 2: 329-458.(Herbart's proposal concerning the practical question of what, in an ethical sense, is the good.)

Herbart, J.F. (1824, 1825) 'Psychologie als Wissenschaft neu gegründet auf Erfahrung, Metaphysik und Mathematik' (Psychology as a Science Newly Based on Experience, Metaphysics and Mathematics), Part 1 in Herbart, 1887, vol. 5, 177-402; Part 2 in Herbart, 1887, vol. 6, 1-338.(Herbart's fundamental work on the theoretical concept of the self.)

References and further reading

Dunkel, H.D. (1970) *Herbart and Herbartianism: An Educational Ghost Story*, Chicago, IL, and London: University of Chicago.(A systematic introduction to Herbartian pedagogy.)

Frank, M. (1991) 'Fragmente einer Geschichte der Selbstbewußtseinstheorien von Kant bis Sartre' (Historical Fragments of the Theory of the Self from Kant to Sartre), in *Selbstbewußtseinstheorien von Fichte bis Sartre (Theories of the Self from Fichte to Sartre)*, ed. M. Frank, Frankfurt am Main: Suhrkamp, 413-599.(A critical discussion of the concepts of the self since Kant.)

Langewand, A. (1991) *Moralische Verbindlichkeit oder Erziehung. Herbarts frühe Subjektivitätskritik und die Entstehung des ethisch-edukativen Dilemmas (Moral Obligation or Education. Herbart's Early Critique of Subjectivity and the Rise of the Dilemma Between Ethics and Education)*, Munich and Freiburg: Alber.(A study of the genesis of Herbart's theoretical and practical philosophy.)

Herbert, Edward (Baron Herbert of Cherbury) (*c.*1583-1648)

Responding, on the one hand, to religious conflicts over the question of the locus and interpretation of authority for deciding what constitutes authentic belief and, on the other hand, to general philosophical scepticism, Herbert of Cherbury wrote De Veritate (On Truth) in an attempt to determine the character and circumstances of true understanding. In this work, first published in 1624, he sought to enable people to decide for themselves, by the use of their reason, what they ought to hold. According to his thesis the touchstone for such decisions is provided by certain fundamental truths, the 'common notions', which all people recognize to be true once they have become aware of them. In two later works, De Religione Gentilium (On the Religion of the Heathens) (1663) and A Dialogue between a Tutor and his Pupil (1768), both published after his death, Herbert attempted to show that his position is not falsified by the evidence of wide differences among religions. His other writings include an important history of Henry VIII based on research into state papers, an autobiography that tells the story of his life up to 1624, and some poems. While this courtier, adventurer and diplomat was something of a failure as a public figure, and while he is commonly held to have essayed views about innate notions that were to be refuted by Locke's Essay, his writings provide pioneering studies in England in the genres of metaphysics, comparative religion and autobiography. Religiously he has been persistently maligned as 'the father of English deism' although closer consideration suggests that this reputation is not justified. He is, rather, to be considered an independent thinker who wanted to identify a form of religious belief that was rationally warranted and universally perceivable.

1 Life

Edward Herbert, the son of Richard Herbert of Montgomery Castle, was probably born in 1583. His mother was a close friend of John Donne. In 1596 he matriculated at Oxford where he is said to have established a reputation for learning. He attended the court of Queen Elizabeth and was knighted at the accession of James I in 1603. In 1608 he began a series of visits to the continent where, as a courtier and soldier, he observed the consequences of religious disputes. In 1619 he became English ambassador to the French Court. While involved in diplomatic and social duties, he completed his major philosophical treatise, *De Veritate (On Truth)* and, encouraged by Hugo Grotius and Daniel Tilenus, had it printed in Paris in 1624. During that year he was recalled to England and dismissed, probably because his sensible advice to the king was disliked. James I and Charles I failed to reimburse Herbert for the debts that he incurred while ambassador, seeking to satisfy him with titles, first with the Irish peerage of Castle Island in 1624 and then with the barony of Cherbury in 1629. In fruitless attempts to regain royal favour he wrote two historical studies, the first an attempt to vindicate Buckingham's expedition to the Île de Rhé and the other a deservedly respected study of the reign of Henry VIII (both published posthumously, in 1656 and 1649 respectively).

In 1639 Herbert attended the King at Alnwick and in 1640 argued in the Council against a peace treaty with the Scots. Thereafter, however, although his sons served on the royalist side, Herbert's own main aims seem to have been to preserve his property and to continue his studies. In 1644, under threat that his library in London would be seized, he surrendered Montgomery Castle to Parliamentary forces. He moved to London and was given a pension by Parliament. In 1645 he published a revised and enlarged edition of *De Veritate*, copies of which often include his *De Causis Errorum (Concerning the Causes of Errors)* and other pieces, a collection that was also published separately in the same year. In 1647 he visited Pierre Gassendi in Paris. On his deathbed he asked Archbishop Ussher, an old friend, to bring him the sacrament but his request was made in a less than wholly reverential way. Ussher consequently refused and Herbert died in London on 5 August 1648.

2 Metaphysics

Although respected, as well as criticized, by Grotius, Gassendi, Descartes and Locke, Herbert of Cherbury's work has largely been neglected since the beginning of the eighteenth century. This is partly because of the mixture of notions that provide the background to his insights (as well as being widely read in classical works, scholastic literature and works of Renaissance Humanism, he was influenced by Hermetic notions of universal harmony to the extent of suggesting a 'judicial' use of astrology), partly to a lack of order and precision in developing his ideas, and partly to the fact that his work does not allow historians of thought to fit him easily into some school or

pattern of development. If, however, he is to be linked to any contemporary intellectual tradition, it probably should be to that of seventeenth-century Platonism (see Cambridge Platonism; Platonism, Renaissance). Herbert himself, however, was not interested in either supporting or rebutting the views of others. While he culled his ideas from a wide range of sources, he regarded himself as an independent thinker who philosophized 'freely' (*libere*) and who considered 'right reason' (*recta ratio*) as the only authority, although, as Gassendi points out, in some respects his originality may lie more in his terminology than in the content of his thought.

The primary aim of Herbert's philosophical and theological works is well expressed in the full title of his pioneering study *De Veritate, Prout Distinguitur a Revelatione, a Verisimili, a Possibili, et a Falso (On truth in distinction from revelation, probability, possibility, and error)*. His intention in this work is not to develop a new doctrine but to clarify the meaning of 'true' and to determine how any normal person may identify what is true. His method is to assert the canon of 'right reason' in contrast to the appeal to supposed authorities that characterized the religious disputes of his time. Although he sometimes suggests that the assertion of the ultimate status of reason may be justified by reference to God's creative benevolence and overall providence, on the whole he simply presupposes that reason is the indisputable final judge of what is true. He is aware, however, that people differ about what is true. Hence, while he holds that people should trust their own natural instinct (*instinctus naturalis*), he also asserts that their findings should be tested against the standard of universal consent since this is, for him, a sure and available criterion for determining what reason perceives to be true.

Herbert's understanding of how we grasp the truth is based on a doctrine of the faculties. According to his thesis, all people possess a vast number of potential faculties, each of which is 'an internal power of the mind' that corresponds to a differentiable object, whether intellectual or physical, and provides a distinct way of perceiving it. Truth is accordingly defined in relational terms as conformity between objects and their appropriate faculties. Human knowledge of the truth is consequently limited by the range of the available 'faculties' but the existence of a 'faculty' is held to show the existence of a corresponding object. Agreement about what can be known to be true and how it may be so known thus depends upon determining the range, character and proper application of the 'faculties'. This is what Herbert investigates in *De Veritate*, distinguishing between four classes of truth (truth of things as they are in themselves, of the appearance of things, of conceptual distinctions, and of judgments about the correctness of the deliverances of the faculties) and attempting to identify the conditions for the perception of what is true in each class.

Herbert's doctrine of the 'faculties' means that, whereas Locke was to suggest that ideas of objects are basically the product of passive experiences, and whereas Kant's 'critical' revolution was to maintain that objects as known are fashioned by our modes of cognition, Herbert presented a third alternative, namely that there is a fundamental correspondence between what is actually the case and our latent modes of apprehension (see Kant, I. §§4-6). He thus puts forward what may be described as a 'common sense' view of 'truth' according to which the mind has an active role in grasping the reality of what it 'truly' apprehends. In this way, then, Herbert rejects on the one hand the view that the mind is a clean sheet or *tabula rasa* on which objects make their own impressions and which, consequently, receives its ability to deal with objects from objects themselves; on the other hand he accepts the principle that nothing is in the intellect which has not first been in the senses, provided that it is also appreciated that the mind deals with its experiences according to the structures prefigured by the 'faculties'. Contrary to what Locke may suggest, what are innate for Herbert are not propositions but the 'faculties' as pre-established potential modes of thought. The harmonious world system, however, entails that the interaction of the faculties with experiences produces true understanding of reality. Furthermore, Herbert claims that people are so constituted that what he describes as 'the common sense of the inner senses' is only satisfied when the 'faculties' are correctly adjusted to what they are seeking thereby to know. When, therefore, they attempt to apprehend something by what is not the correct corresponding 'faculty', Herbert holds that they experience a particular kind of corrective sensation which tells them that they are falling into error.

As well as holding that a person feels when the correct faculty is not being used to apprehend something, Herbert also maintains that among the intellectual objects represented by the faculties are certain fundamental 'common notions'. These are normative principles implanted by God in every person to ensure that they can distinguish what is true and good from what is false and bad. Accordingly they are to govern understanding and action. Since they are apprehended by natural instincts, they are found in every normal person. The six characteristic qualities of the common notions are *prioritas*, *independentia*, *universalitas*, *certitudo*, *necessitas* and *modus conformationis*

(priority, independence, universality, certainty, necessity - in the sense of being required for human preservation, and immediacy -in that they do not need to be warranted by discursive reasoning). What is *common* about the 'common notions' is that they are latently present in the mind of every person. Individuals only explicitly apprehend them when their minds have been suitably stimulated. Once they have become aware of them, however, normal people everywhere immediately recognize them to be true.

3 Views on religious belief

According to Herbert, certain 'common notions' provide the proper basis for judging what is true in matters of religious belief, including claims about what is held to have been divinely revealed. In *De Veritate* he discusses five 'common notions' that, in his judgment, constitute the foundation of and norm for the true Catholic or universal Church that is the only source of salvation. They are:

1 Some supreme deity exists
2 This supreme deity ought to be worshipped
3 Virtue joined with piety… is and always has been held to be the principal part of religious practice
4 Vices and crimes must be expiated by repentance
5 There is reward or punishment after this life

Herbert does not deny that some truths may have been revealed by God, including ones that may in principle be discovered naturally. He insists, however, that care must be taken to make sure that any alleged revelation is authentic and, furthermore, that no alleged revelation can be regarded as authentic if it contradicts (rather than reasserts or augments) what is laid down in the 'common notions' of religion.

In *De Religione Laici* Herbert develops his understanding of the common notions 'divinely transcribed in the mind itself' to hold that any lay person may, can and should decide between rival systems of religious belief by determining, through the use of reason, which set of doctrines and practices is the closest to those implied by the 'common notions'. Moreover, where beliefs involve claims about what happened in the past, they cannot escape the uncertainty that belongs to all such claims. As for the authority of the Bible, Herbert maintains that each individual has the right to judge the significance of its contents and to accept as 'the very Word of God' needed for salvation only those parts which agree with 'our Catholic Truths, which are like indubitable utterances of God, and matters recorded in the inner court' - by which he means the truths that are contained in the 'common notions' and, as such, that are engraved in the conscience by providence. The products of priestcraft are similarly to be subjected to these criteria. Herbert, however, does not see such an approach to doctrines and practices as hostile to authentic religion, but as upholding it.

While Herbert's views raise questions about unwarranted (and unwarrantable) claims to divine revelation and about the perversion of authentic religion by priestcraft, and while he proposes 'right reason' as the way to determine what is to be believed, it is misleading to describe him (as is widely done) as 'the father of English Deism' (see Deism). Herbert's own remarks and religious practice suggest that he was convinced not only of the actuality of God but also of the gracious activity of God in the lives of individuals, of the possibility of divine revelations, of the efficacy of prayer and of the reality of personal immortality. Moreover, there is little evidence that he was a significant influence on others given the unclear and confusing title of 'deists', despite his treatment as the first English 'deist' in early attacks on 'deism' by Thomas Halyburton (1714), Philip Skelton (1749) and John Leland (1754) in confessedly apologetic works which have provided the basic canon for later studies of this aspect of religious thought in the late seventeenth and eighteenth centuries.

4 Other religions

Two principles guide Herbert's approach to the various religions of the world. The first is his conviction that universal providence is the highest attribute of God. Hence God's benevolence must be regarded as concerned for all people, and salvation as available to all. The second is the need to show empirically that the 'common notions' of religion are indeed universally recognized.

Herbert meets the first principle by holding that atonement for sin is obtained through repentance. This means that neither a believing response to a particular event (such as the death of Jesus as the Christ) nor assent to a particular set of doctrines nor the fulfilment of prescribed rites is necessary for a person to be saved. The knowledge of what

is necessary for salvation is, always has been and always will be available to all people since it is contained within the 'common notions' of religion. Thus God does not unfairly limit saving knowledge to only part of humankind. The providential grace of God extends to all.

The second principle is met by arguing, both in *De Religione Gentilium* and in *A Dialogue between a Tutor and his Pupil*, that, in spite of what the evidence of the doctrines and practices of the world's faiths may seem to indicate, all people do basically recognize the 'common notions' of religion. Herbert deals with apparent evidence to the contrary in three ways. First, he suggests that some reports about the beliefs and practices in other religions have simply been misunderstood, particularly because of ambiguity in the language used. Second, he holds that sometimes symbolic usages have been misinterpreted as literal references. Third, he admits that some doctrines and rites found in other faiths do conflict with the 'common notions' of religion, but only because they have been corrupted by priestcraft. Priests are charged with distorting and perverting religion for their own selfish advantage. Herbert thus maintains that both in principle and practice the religions of the world fundamentally support his understanding of the actuality of the universal providence of God, and that the basic truths of religion are found in the non-Christian religions.

See also: Grace; Hermetism

DAVID A. PAILIN

List of works

Herbert, Edward, Baron of Cherbury (1624) *De Veritate, Prout Distinguitur a Revelatione, a Verisimili, a Possibili, et a Falso (On Truth in Distinction from Revelation, Probability, Possibility, and Error)*, Paris; 3rd edn, London, 1645; partial trans. M.H. Carré, Bristol: University of Bristol, 1937; repr. London and Kinokuniya, Tokyo: Routledge/Thoemmes Press, 1992.(Written in a somewhat unusual Latin and completed while Herbert was ambassador in France, this seeks to identify 'the common nature of the search for truth which exists in every normal human being'. The revised edition of 1645 is best.)

Herbert, Edward, Baron of Cherbury (1645) *De Causis Errorum: Una cum Tractatu de Religione Laici, et Appendice ad Sacerdotes, nec non quibusdam Poematibus (Concerning the Causes of Errors: with a Treatise on the Religion of the Laity, and a Supplement to Priests, also certain Poems)*, London, 1645; ed. and trans. H.R. Hutcheson as *Lord Herbert of Cherbury's De Religione Laici*, New Haven, CT: Yale University Press, 1944.(Copies of the 1645 London edition of *De Veritate* often also contain this work which was also published separately. The Hutcheson translation contains a useful discussion of Herbert's life and thought and a bibliography.)

Herbert, Edward, Baron of Cherbury (1649) *The Life and Raigne of King Henry the Eighth*, London, 1672. (Written between 1632 and 1639, this was long regarded as the standard history of the reign.)

Herbert, Edward, Baron of Cherbury (1656) *Expeditio in Ream Insulam (The Expedition to the Island of Rhé)*, ed. and Latin trans. T. Baldwin, London; original English text ed. Lord Powis, London: Philobiblon Society, 1860.(Completed in 1630, this attempts the difficult task of justifying the conduct of the Duke of Buckingham as leader of an unsuccessful attempt to relieve the Hugenots in La Rochelle in 1627.)

Herbert, Edward, Baron of Cherbury (1663) *De Religione Gentilium (On the Religion of the Heathens)*, ed. I. Vossius, Amsterdam; trans. W. Lewis as *The Antient Religion of the Gentiles*, London, 1705.(Attempts to show that the basic truths of religion are found, albeit sometimes using indirect modes of expression, in non-Christian religions. To date the only English translation is that by Lewis.)

Herbert, Edward, Baron of Cherbury (1764) *The Life of Edward Lord Herbert of Cherbury written by himself*, Strawberry Hill; ed. with notes, appendices and later life, S.L. Lee, London, 1886; ed. with notes, J.M. Shuttleworth, London: Oxford University Press, 1976.(This pioneering work in autobiography (it stops at 1624) is amusing, sometimes frivolously teasing, as well as informative. The edition by Lee includes a useful addition that covers his life after 1624; Shuttleworth takes account of Herbert's corrections and notes significant variant readings.)

Herbert, Edward, Baron of Cherbury (1768) *A Dialogue between a Tutor and his Pupil*, London.(Authorship disputed, but probably at least largely by Herbert. Seeks to show that basic religious truths are recognized by all religions in spite of corruptions due to priestcraft.)

References and further reading

Bedford, R.D. (1979) *The defence of truth: Herbert of Cherbury and the seventeenth century*, Manchester: Manchester University Press.(Readable and comprehensive work that perhaps exaggerates the influence of

hermetic ideas upon Herbert. Select bibliography.)

Halyburton, T. (1714) *Natural religion insufficient; and revealed necessary to man's happiness*, Edinburgh. (Chapter 13 includes an early, brief and hostile 'account' of Herbert as 'the first inventor' of the deist 'religion'.)

Harrison, P. (1990) *'Religion' and the religions in the English enlightenment*, Cambridge: Cambridge University Press, 61-73.(Short but perceptive account, primarily concerned with Herbert's religious ideas and his relationship to 'deism'.)

Leland, J. (1754) *A view of the principal deistical writers*, London: B. Dod, vol. 1, 1-39.(Although hostile to 'deism' and holding Herbert to be 'justly regarded as the most eminent of the deistical writers' (page 4) this is reasonably fair in its critical appraisal of his thought.)

Pailin, D.A. (1983) 'Herbert of Cherbury and the deists', *The Expository Times* 94: 196-200.(Outlines Herbert's religious views and practices to defend him against the charge of being a 'deist'.)

Pailin, D.A. (1988) 'Herbert von Cherbury', in J.-P. Schobinger (ed.) *Grundriss der Geschichte der Philosophie, Band 3: Die Philosophie des 17. Jahrhunderts*, Basel: Schwabe, 224-39, 284-5.(Brief survey in German of Herbert's ideas and of major responses to them.)

Remusat, C.F.M. de (1874) *Lord Herbert de Cherbury. Sa vie et ses oeuvres ou les origines de la philosophie du sens commun et de la théologie naturelle en Angleterre (Lord Herbert of Cherbury. His life and works, or the origins of the philosophy of common sense and of natural theology in England*), Paris.(An old but still useful study in French.)

Rossi, M.M. (1940) 'The nature of truth and Lord Herbert of Cherbury's inquiry', *The Personalist* 21: 243-56, 394-409.(Critical of Herbert's attempt to deal with the question of truth, accusing him of presenting a patchwork of incompatible views.)

Rossi, M.M. (1947) *La vita, le opere, i tempi di Eduardo Herbert di Chirbury (The life, work and times of Edward, Herbert of Cherbury)*, Florence: G.C. Sansoni, 3 vols.(Major, comprehensive study in Italian, if somewhat unsympathetic to its subject.)

Skelton, P. (1749) *Deism revealed or, the attack on christianity candidly reviewed*, London: A. Millar, 2 vols. (Response to what is said to be deism's attack on Christianity that was approved for publication by David Hume and that, in vol. 2, 240-1, describes Herbert as the one whom 'later deists' imitate.)

Sorley, W.R. (1894) 'The philosophy of Herbert of Cherbury', *Mind* n.s. 3: 491-508.(Largely descriptive and balanced account of Herbert's theory of knowledge.)

Webb, C.C.J. (1915) *Studies in the history of natural theology*, Oxford: Clarendon Press, 344-58.(Sound and sympathetic interpretation of Herbert's views.)

Herbrand’s theorem

According to Herbrand’s theorem, each formula F of quantification theory can be associated with a sequence F_1, F_2, F_3,... *of quantifier-free formulas such that F is provable just in case* F_n *is truth-functionally valid for some n. This theorem was the centrepiece of Herbrand’s dissertation, written in 1929 as a contribution to Hilbert’s programme. It provides a finitistically meaningful interpretation of quantification over an infinite domain. Furthermore, it can be applied to yield various consistency and decidability results for formal systems. Herbrand was the first to exploit it in this way, and his work has influenced subsequent research in these areas. While Herbrand’s approach to proof theory has perhaps been overshadowed by the tradition which derives from Gentzen, recent work on automated reasoning continues to draw on his ideas.*

Herbrand’s theorem says that each formula *F* of quantification theory can be associated with a sequence $F_1, F_2, F_3, \ldots$ of quantifier-free formulas such that *F* is provable just in case F_n is truth-functionally valid for some *n*. If *F* is prenex - that is, consists of a quantifier-free matrix *G* preceded by a string of quantifiers - then each F_i will be a disjunction of instances of *G* obtained by substituting terms for free variables. To simplify the exposition we consider only prenex formulas *F*. With each such *F* we can associate an existential formula F^H as follows: (1) substitute new constants for the free variables of *F*; (2) remove each universal quantifier $\forall x$ and replace the resulting free occurrences of *x* by a term $f(y_1, \ldots, y_m)$, where *f* is a new function symbol and $y_1, \ldots, y_m$ are all the variables bound by existential quantifiers to the left of $\forall x$ in *F*. For example, applied to $\forall x \exists y \forall z R(x, y, z)$ (with *R* quantifier-free) this procedure yields $\exists y R(c, y, f(y))$.

Although Herbrand would not have described their relationship in these terms, F^H can be said to express the meaning of *F* in the sense that *F* is true under an interpretation $\mathbb{I}$ just in case F^H is true under any expansion of $\mathbb{I}$ to the new constants and function symbols in F^H. To see this, consider our example. Let α, β range over the domain of $\mathbb{I}$, ϕ over functions on this domain and let ‘ x_α’ indicate that the value α is assigned to the variable *x*. Then the following are equivalent:

$\forall x \exists y \forall z R(x, y, z)$ is true under $\mathbb{I}$.
There is no α such that for all β $\exists z \neg R(x_\alpha, y_\beta, z)$.
There is no α and no function ϕ such that for all β $\neg R(x_\alpha, y_\beta, z_{\phi(\beta)})$.
For every α and every function ϕ there is some β such that $R(x_\alpha, y_\beta, z_{\phi(\beta)})$.
No matter which individual α is chosen to interpret *c* and which function ϕ to interpret *f*, $\exists y R(c, y, f(y))$ is true under $\mathbb{I}$.

This means, in particular, that *F* is true under every interpretation iff F^H is; the latter is called the validity functional form of *F*.

We can use the functional terms appearing in F^H to generate a series of finite domains $\mathbf{D}_i$. These will consist of constants which can themselves be substituted into the quantifier-free matrix G^H of F^H. For convenience, let $\mathbf{D}_1$ consist of a single constant *d*, $\mathbf{D}_2$ of a distinct constant (its ‘denotation’) for each term obtained from one occurring in G^H by substituting *d* for its free variables; in particular, constants occurring in G^H are assigned denotations in $\mathbf{D}_2$. In general, $\mathbf{D}_{n+1}$ consists of distinct denotations for each term obtained from one occurring in G^H by substituting members of $\mathbf{D}_n$ for its free variables. F_n, the Herbrand expansion of *F* of order *n*, is a disjunction whose disjuncts represent all possible ways of substituting members of $\mathbf{D}_n$ for the variables of G^H and replacing functional terms, where possible, by their values in $\mathbf{D}_n$. For example, if we allow each closed term to denote itself, then the first three domains generated by our earlier formula will be $\{d\}$, $\{d, c, f(d)\}$ and $\{d, c, f(d), f(c), f(f(d))\}$. The Herbrand expansions of orders 1, 2 and 3 are:

$$R(c, d, f(d))$$
$$R(c, d, f(d)) \vee R(c, c, f(c)) \vee R(c, f(d), f(f(d)))$$
$$R(c, d, f(d)) \vee R(c, c, f(c)) \vee R(c, f(d), f(f(d))) \vee R(c, f(c), f(f(c))) \vee R(c, f(f(d)), f(f(f(d)))).$$

Notice that F_n is not an interpretation of the quantifiers of F^H in $\mathbf{D}_n$ in the usual sense because some of the

terms occurring in F_n only acquire a denotation in $\mathbf{D}_{n+1}$.

Now, for provable *F*, Herbrand showed how to compute a number *n* such that F_n would be a tautology. He argued by induction on the proof of *F*. The only axioms in his system are the quantifier-free tautologies, so his method is to analyse each rule of inference and show how the order of a tautologous expansion of its conclusion can be calculated from the order(s) of such expansions of its premises. Conversely, he showed how to construct a proof of *F* from any tautologous F_n. Because the only rules required here are forms of generalization and simplification (in particular, *modus ponens* is not used), this yields a normal form for proofs in quantification theory analogous to Gentzen's cut-elimination theorem and lends itself to many of the same applications. It allows a bound to be placed on the complexity of formulas appearing in a normal proof depending on the complexity of its conclusion (and assumptions, if any).

Herbrand's proof contains ideas of great interest but, in contrast to Gentzen's, it is not easily adapted to non-classical systems. Furthermore, it is difficult to follow and turns out to contain a defect. Although this has been repaired by Dreben and his collaborators, the proof has not been reproduced and nowadays the theorem (for prenex formulas only) is often established as a corollary to a refinement of the cut-elimination theorem. (Since a cut-free proof of a prenex formula can be separated into a truth-functional and a quantificational part, by working back from the conclusion we can discover a truth-functionally valid formula from which it follows.)

While he accepted set-theoretic arguments in mathematics, Herbrand was committed to using finitary methods in metamathematical investigations. His work was influenced by Löwenheim (1915), but he rejected the naïve notion of satisfiability in an infinite domain the latter assumed. In fact, Herbrand claimed that his result was just a more rigorous version of Löwenheim's famous theorem to the effect that a sentence is satisfiable if and only if it is satisfiable in a denumerable domain. It would be more accurate, however, to describe it as a finitist variant thereof. To understand why this is so, notice that we can reformulate Herbrand's theorem using a construction dual to the above. We remove existential quantifiers from *F*, replacing the newly freed variables by functional terms to obtain a universal formula (the satisfiability functional form of *F*), which can then be expanded as a series of finite conjunctions F'_n over the $\mathbf{D}_n$. The theorem then states that *F* is irrefutable iff F'_n is truth-functionally satisfiable for every *n*. (This is the original theorem expressed as a criterion for the non-provability of $\neg F$.) Constrained by his finitist scruples, Herbrand explicated validity in terms of provability and, *a fortiori*, satisfiability in terms of irrefutability. He also suggested that *F*'s having truth-functionally satisfiable expansions F'_n for every *n* could serve as a constructive account of what it means for *F* to be satisfiable in a (denumerable) domain. Replacing the informal set-theoretic notions in the statement of Löwenheim's theorem by their finitist analogues transforms it into a statement of Herbrand's theorem - even though the theorems themselves remain distinct.

Herbrand's ideas have yielded numerous metamathematical applications. The normal form for proofs mentioned above supplies a tool for obtaining consistency proofs. Furthermore, the elimination of quantifiers in favour of functional terms suggests how to exhibit the constructive content of the theorems of a formal theory - as in the 'no counterexample' interpretation of arithmetic, for example (Kreisel 1951). Herbrand's theorem also allows us to prove a theory consistent by approximating a model for it, that is, by establishing that, for every (finite) conjunction *F* of the theory's axioms, F'_n is truth-functionally satisfiable for every *n*. Herbrand himself gave a consistency proof for a fragment of arithmetic along these lines, and his methods have been extended to yield proofs for the full system (see Dreben and Denton 1970, Scanlon 1972). These proofs, while no less complex than Gentzen's, yield directly the sort of constructive interpretations alluded to above.

Finally, Herbrand's theorem supplies an approach to the decision problem for classes of quantificational formulas, that is, to discovering an algorithm which, for any formula in the class, computes whether or not it is satisfiable. The idea here is to investigate the relationship between syntactic properties of quantificational formulas and the structure of their expansions, and to discover properties of the latter which ensure the existence of a decision procedure (see Dreben and Goldfarb 1979). His theorem and the methods introduced in its proof are relevant not only to theoretical decidability results, but also to the design of implementable automated reasoning procedures. These employ unification algorithms to improve the efficiency of testing quantifier-free expansions for satisfiability, and Herbrand is credited with being the first to devise such an algorithm (Snyder 1991).

See also: Hilbert's programme and formalism; Logical and Mathematical Terms, Glossary of; Proof theory

A.M. UNGAR

References and further reading

Andrews, P.B. (1986) *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof,* New York: Academic Press.(§35 contains an interesting discussion of Herbrand’s theorem, focusing on its significance for decision problems and automated theorem proving, as well as a proof of one version of it.)

Buss, S.R. (1995) ‘On Herbrand’s Theorem’, in D. Leivant (ed.) *Logic and Computational Complexity*, Berlin: Springer, 195-209.(An up-to-date exposition of the theorem from a proof-theoretic perspective which also sets it in the context of Herbrand’s dissertation.)

Dreben, B. and Denton, J. (1970) ‘Herbrand-Style Consistency Proofs’, in J. Myhill, R.E. Vesley and A. Kino (eds) *Intuitionism and Proof Theory*, Amsterdam: North Holland, 419-33.(Building on ideas to be found in Herbrand, the authors sketch a proof of the consistency of arithmetic and discuss some of its corollaries.)

Dreben, B. and Goldfarb, W.D. (1979) *The Decision Problem: Solvable Classes of Quantificational Formulas*, Reading, MA: Addison-Wesley.(This monograph offers a unified treatment of solvable cases of the decision problem by way of an approach that can be traced back to Herbrand.)

Goldfarb, W.D. (1979) ‘Logic in the Twenties: The Nature of the Quantifier’, *Journal of Symbolic Logic* 44 (3): 351-68.(A very readable discussion of the background to Herbrand’s work.)

Heijenoort, J. van (1985) ‘Jacques Herbrand’s Work in Logic and its Historical Context’, in *Selected Essays*, Naples: Bibliopolis, 99-121.(A readable informal survey of Herbrand’s work.)

Herbrand, J. (1968) *Écrits logiques*, Paris: Presses Universitaires de France; trans. *Logical Writings*, ed. W.D. Goldfarb, Cambridge, MA: Harvard University Press, 1971.(All of Herbrand’s logical writings, with extensive annotations, needed corrections and a helpful introduction by the editor.)

Hilbert, D. and Bernays, P. (1939) *Grundlagen der Mathematik (Foundations of Mathematics)*, vol. 2, Berlin: Springer, 2nd edn, 1970.(A standard reference; Herbrand’s theorem is derived from the first ε-elimination theorem and its metamathematical applications detailed.)

Kreisel, G. (1951, 1952) ‘On the Interpretation of Nonfinitist Proofs’, *Journal of Symbolic Logic* 16 (4): 241-67, 17 (1): 43-58.(The methods of Hilbert and Bernays (1939) are utilized to impart a finitist sense to proofs in arithmetic and a subsystem of analysis.)

Löwenheim, L. (1915) ‘Über Möglichkeiten im Relativkalkül’, *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, ‘On Possibilities in the Calculus of Relatives’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967.(A seminal paper which, because its results are stated and proved in terms of Schröder’s calculus of relatives, is accessible only with difficulty to the modern reader.)

Scanlon, T.M. (1972) ‘The Consistency of Number Theory via Herbrand’s Theorem’, *Journal of Symbolic Logic* 38 (1): 29-58.(This paper presents in some detail an improved and slightly more general version of the proof given by Dreben and Denton (1970).)

Schwichtenberg, H. (1977) ‘Proof Theory: Some Applications of Cut-Elimination’, in J. Barwise (ed.) *Handbook of Mathematical Logic,* Amsterdam: North Holland, 867-95.(Herbrand’s theorem is derived from the cut-elimination theorem and some of its proof theoretic applications are sketched.)

Shoenfield, J. (1967) *Mathematical Logic*, Reading, MA: Addison-Wesley.(Chapter 4 includes an accessible and enlightening proof of Herbrand’s theorem for the predicate calculus with equality.)

Snyder, W. (1991) *A Proof Theory for General Unification*, Boston, MA: Birkhäuser.(A specialized monograph which describes an abstract framework within which to understand different kinds of unification.)

Herder, Johann Gottfried (1744-1803)

Herder was a central figure in the German intellectual renaissance of the late eighteenth century. His achievement spanned virtually every domain of philosophy, and his influence, especially upon Romanticism and German idealism, was immense. In social and political philosophy he played a prominent role in the development of historicism and nationalism. In metaphysics he developed the doctrine of vitalist pantheism, which later became important for Goethe, Schelling and Hegel. In the philosophy of mind he formulated an organic theory of the mind-body relationship, which was crucial for Schelling and Hegel. And in aesthetics he was among the first to defend the value of ethnic poetry and the need for the internal and historical understanding of a text.

Herder's main aim was to extend the powers of naturalistic explanation to the realm of culture, so that characteristic human activities, such as art, religion, law and language, could be included within the scientific worldview. But he also wanted to avoid reductivistic forms of explanation that viewed such activities as nothing more than matter-in-motion or stimulus-response mechanisms. He insisted that explanation in the cultural sphere had to be holistic and internal as well as mechanical and external. An action had to be understood in its historical context and according to the intention of the agent and not simply as another instance of a causal regularity between events. Herder's programme, then, was to develop naturalistic yet non-reductivistic explanations for the realm of culture. He attempted to realize this programme in many spheres, especially language, history, religion and the mind.

1 Intellectual ideals and background

Born into a middle-class family in Mohrungen, East Prussia, Herder spent most of his life as a Lutheran pastor in the provincial cities of Riga, Bueckeburg and Weimar. Herder's intellectual development was dominated by one central fact: during his formative university years in Königsberg from 1762 to 1764, he was the student of both Kant and Hamann. To the same degree, they were a profound influence upon him. But such a rich legacy also proved problematic, for Kant and Hamann represented opposing intellectual movements. While Kant was a champion of the *Aufklärung*, the German Enlightenment, Hamann was a founder of the *Sturm und Drang* (lit.: Storm and Stress), the German literary movement of the 1760s and 1770s that proclaimed personal freedom and glamorized revolt. Although these movements had some common values - individual emancipation, the essential goodness of human nature, contempt for aristocratic privilege and clerical prejudice, hatred of superstition and intolerance - they also had some conflicting ones. The *Aufklärung* stressed the sovereignty of reason, the value of the arts and sciences, and the virtues of cosmopolitanism; the *Sturm und Drang* emphasized the rights of feeling and imagination, the problems of civilization, and the value of local and ethnic identity. Herder struggled to reconcile these opposing ideals. His philosophy could be described as a rich synthesis - or unstable mixture - of these antithetical movements. We might best describe Herder with oxymorons: a turbulent *Aufklärer*, a rational *Stürmer und Dränger*.

What, more specifically, did Herder learn from Kant? All his life Herder adhered to the young Kant's conception of the aim and method of philosophy: its purpose is to develop intellectual autonomy; and its method should be 'analytic', beginning from experience and ascending to general principles. In his later works Herder applied these early Kantian ideals -much to the chagrin of the mature Kant, who regarded them as a relapse into metaphysical dogmatism. Of all Kant's works Herder was most influenced by the *Allgemeine Naturgeschichte und Theorie des Himmels (Universal Natural History and Theory of the Heavens)* (1755). This work was important for Herder for two reasons: (1) its idea of a natural history, that the apparently eternal and static structure of nature is really the product of historical development; (2) its suggestion that human beings are part of nature and also have a natural history. Herder's historicism arose from his application of Kant's idea of natural history to the social, religious and historical spheres.

What, more precisely, did Hamann teach Herder? There were many influential ideas: the value of cultural diversity, the vitality of folk poetry, the power of the passions, the importance of language for reason and of custom for language, and the need to examine the social and historical context of reason. Hamann also impressed upon Herder the limitations of the *Aufklärung*: that religious experience cannot be criticized by reason, that the mind-body relationship is inexplicable by mechanism, and that human beings are indivisible wholes not neatly

analysed into faculties.

The guiding ideal behind Herder’s intellectual development was a synthesis of Kant and Hamann. Throughout his career Herder’s aim was to develop naturalistic and historical explanations for characteristic human activities, such as art, religion, language, law and morality. He wanted to extend the powers of natural explanation, so that all forms of humanity would be brought within the scientific worldview. In this regard he revealed himself as a loyal follower of Kant and the *Allgemeine Naturgeschichte*. Yet there was something unique about Herder’s programme that broke from the heritage of Kant and the *Aufklärung*. Herder insisted that naturalistic explanations do not reduce human activities to matter-in-motion or stimulus-response mechanisms. His critique of mechanism, and his demand for a more holistic and internal form of understanding, betrays the influence of Hamann. Herder’s programme for naturalistic yet non-reductivistic explanation attempted to do justice to the legacy of both Kant and Hamann. We shall now see how Herder attempted to fulfil this programme in virtually every field of philosophy.

2 Theory of language

In 1769 the Berlin Academy of Sciences announced a prize competition for the best essay on the questions ‘If human beings were left with their natural faculties, would they be able to invent language? And, if so, by what means could they invent it?’ Competitors were invited to disagree with a late member of the Academy, J.P. Süssmilch, who argued that a perfect and complete language could have only a divine origin. This quaint question raised an important and broader issue: to what extent is reason, which depends on language, explicable according to natural laws? Should human intelligence be brought within the scientific worldview?

Inspired by the question, Herder wrote a short tract to answer it in December 1770, *Ueber den Ursprung der Sprache (On the Origin of Language)*. This marks the starting point of his philosophical programme, the attempt to provide naturalistic yet non-reductivistic explanations of characteristic human activities. Herder takes a firm stand on behalf of naturalism, dismissing Süssmilch’s theory as so much nonsense. He puts forward two specific theses, each of which is necessary to prove his naturalism. (1) Human reason is sufficient to create language. Since, however, reason itself might have a supernatural or divine origin, he adds a second thesis. 2) The use of reason is natural and necessary because it is the means by which human beings acquire the skills and information necessary for survival. Hence if reason is sufficient to create language, and if its use is natural, then the creation of language will be natural too.

To defend his first thesis, Herder argues that the function of reason is to direct, control and organize our experience. Such a function involves what he calls ‘reflection’ (*Besonnenheit*), the capacity to be selfconscious of our sensations by identifying or remembering them. We can exercise this capacity only by giving signs to specific features of experience; only through the use of signs are we able to identify, remember and control our sensations. In employing signs, though, we have already created language.

To support his second thesis Herder develops a proto-Darwinian account of why reason, and indeed language, is necessary for survival. Since men, unlike animals and insects, do not have instincts to guide them, they must learn how to survive. If each new generation is not to be exposed to the same dangers and perils as past ones, its ancestors must teach them the skills and facts necessary for survival. The most effective means of doing so is, of course, language.

The aim of Herder’s tract is not simply to develop a naturalistic theory of the origin of language. That had been done before by such eminent thinkers as Rousseau and Condillac. Herder’s specific aim is to develop a naturalism that is not reductivist. He rejects the theories of Rousseau and Condillac precisely because they were reductivist, failing to explain the characteristic functions of human reason. In his second *Discours* (1755) Rousseau argued that language originates with the expression of feeling; the first words were cries or ejaculations (see Rousseau, J.-J. §2). In his *Essai sur l’origine des connaissances humaines* (1746) Condillac maintained that language begins with convention, the agreement between speakers about the referents of sounds (see Condillac, É.B. De §2). The purpose of language is to communicate; but people can do this only by agreeing about the meaning of sounds. Herder rejects Rousseau’s theory because it treats human beings as if they were animals, ignoring what is most distinctive about human language, its cognitive dimension. He dismisses Condillac’s theory for its circularity: primitive people cannot make conventions unless they already have a sophisticated concept of language. Both theories are inadequate because they do not consider the characteristic nature of man - reason. Either they reduce

reason to man's animal nature (Rousseau) or they assume it without explaining it (Condillac). Yet it is human rationality, Herder contends, that holds the key to the origin of language. Since men alone possess rationality, and since they also are unique in having language, we must seek its origin in their characteristic nature. That, Herder says, is the guiding assumption behind his whole investigation.

3 Philosophy of mind

In 1774 Herder wrote another treatise for an academic prize, his *Vom Erkennen und Empfinden der menschlichen Seele (On the Knowledge and Sensation of the Human Soul)*, which was eventually published in 1778. He again attempted to answer a question posed by the Academy of Sciences: 'What is the nature of, and relationship between, the two basic faculties of the soul, knowledge and sensation?' This question too raised a wider problem: the relationship between the mind (the source of knowledge) and the body (the source of sensation).

Herder began by criticizing some of the current theories about the mind-body relationship. In the late eighteenth century these usually took two forms: they were dualistic, making the mind into a special substance distinct from the body, or materialistic, explaining the mind as if it were a complicated machine. Herder rejects dualism because it makes the interaction of mind and body mysterious (see Dualism); and he dismisses materialism since it reduces the mind down to a mere stimulus-response mechanism (see Materialism in the philosophy of mind). What is required, he contends, is a theory that avoids both these pitfalls. It should remove the mystery of mind-body interaction while not reducing one to the other. In other words, it must be naturalistic yet non-reductivist. The main task of his tract is to sketch just such a theory.

The central thesis of Herder's organic or vitalist theory is that the mind and body are not heterogeneous substances, nor more or less complicated machines. Rather, they are simply different degrees of organization and development of a single living power. The essence of power consists in self-generating and self-organizing activity, which gradually develops from lower to higher forms of organization and structure. The difference in mind and body is therefore not in kind but only in degree: the mind is organized and developed, the body is amorphous and inchoate, power. Such a theory easily explains interaction since mind and body are both aspects of a single force; but it also does not reduce one to the other because they are different degrees of organization and development.

Herder develops his theory by questioning the fundamental premise behind both dualism and materialism: that matter consists in extension alone. It was this assumption that forced the dualist to separate the mind from the body and that compelled the materialist to reduce it to a machine. If, however, we reject this assumption, then we escape the whole dilemma between materialism and dualism. According to Herder, the latest work in the natural sciences permits us to do just this. The recent discoveries of electricity and magnetism, the phenomenon of irritability investigated by Albrecht von Haller, the concept of *vis viva* developed by Hermann von Boerhaave, and the critique of preformation by Thomas Needham and P.L. Maupertuis, all seem to show that the essence of matter is not extension but force, power or energy. In general, Herder stressed the importance of empirical research for the philosophy of mind. No psychology could succeed, he stressed, unless it followed physiology at every step.

4 Philosophy of history

The most influential aspect of Herder's thought was his philosophy of history, which laid the foundation for much of the historicism of the nineteenth century. He developed his philosophy of history in two main works, *Auch eine Philosophie der Geschichte der Menschheit (Another Philosophy of History of Humanity)* (1774) and *Ideen zur Philosophie der Geschichte der Menschheit (Ideas for a Philosophy of History of Humanity)* (1784-91).

Crucial for the development of Herder's historical thought was his 'genetic method', which he first sketched in his *Fragmente* (1767). This was originally conceived as a method for the proper understanding of a work of literature; but Herder eventually extended it to all human actions and creations. This method consists in two basic principles. (1) To understand an action or creation, it should be seen within its historical context, as the product of a specific time and place. (2) It is necessary to understand the action or creation from within, according to the intention of the agent or creator, and not merely from without, according to its external causes or some purported universal rules. With the first guideline Herder attempted to satisfy the demand for natural explanation; he was applying the general lesson he had learned from Kant's *Allgemeine Naturgeschichte*, that what appears given and eternal is

really the product of history. With the second guideline he stressed the need for non-reductive explanation. Here Herder's immediate target was those literary critics who attempted to evaluate a work according to their own standards without attempting to grasp the purpose of the author.

In his *Auch eine Philosophie der Geschichte* Herder criticized the historiography of the *Aufklärung* and stressed the importance of an internal understanding of the past. The historians of the *Aufklärung*, he claimed, were guilty of ethnocentrism. They judged the past and alien cultures in the light of their own, as if the values of their age are universal and eternal and the very end of history itself. They failed to judge each epoch in its own terms, according to its own values and ideals. If the historian is to comprehend an alien culture or epoch, then he must suspend his preconceptions and feel himself into the standpoints of the agents themselves. In this regard Herder was one of the first to stress the need for a more sympathetic understanding of the Middle Ages, which had been dismissed as an age of darkness, superstition and fanaticism.

Herder's critique of Enlightenment historiography raises the question whether he is admitting, or even advocating, relativism. In several places of his tract he openly states that the values of different cultures are incommensurable, and that each culture is valid in its own terms. Indeed, he argues that the apparently universal and eternal standards of the *Aufklärung* are really only invalid generalizations from their own age.

Nevertheless, despite his flirtations with cultural relativity, Herder was anxious to claim that there are some universal values in history. In an early fragment - *Von der Verschiedenheit des Geschmacks und der Denkart unter den Menschen (On the Difference in Taste and Thinking among Human Beings)* (*c*.1765-8) - he reveals a deep concern to refute sceptical arguments for cultural relativism. In the 1774 tract, he expressly maintains that his aim is to show how, amid all the diversity of cultures, there is still 'progress in a higher sense' and how 'humanity still remains humanity'. Herder's aim, then, was to find some middle path between ethnocentrism and relativism. A closer reading of the 1774 tract indeed shows that Herder stresses the individual worth of each culture not because all values are relative but because each has the value appropriate to humanity *at that stage of its development*. In other words, cultures have *sui generis* values because what is appropriate for a lower stage of development is not so for a higher, just as what is good for a child is not so for an adult. This is not relativism but the very opposite: the belief that all cultures can be graded on a scale of progress and development. There remains in Herder's philosophy of history, then, a profound teleological dimension, an abiding belief in progress and providence. In the 1774 tract, however, Herder is reluctant to identify the goal of history. Such hesitation is understandable, given that any specific definition invites ethnocentrism.

Herder's main work in the philosophy of history is his massive, but still incomplete *Ideen*, which appeared in instalments from 1784 to 1791. It consists in four parts. The first two discuss natural history, human anatomy and anthropology. Only in the third part does Herder begin to treat recorded history, whose course is followed until the Roman Empire. The fourth part continues the history to 1500 AD. The fifth part, which was to conclude with the present age, was never written. Conceived on a grand scale, the *Ideen* attempts to examine all of world history, giving equal weight to the cultures of Egypt, China, India and Greece, and the place of humanity within the cosmos.

The *Ideen* was the crowning work of Herder's philosophical programme. Its aim was to bring human history within the naturalistic worldview. Herder wanted to see history as part of the cosmos as a whole, and to formulate its natural laws. But his naturalism was, again, non-reductivist. He conceived the laws of historical development in organic rather than mechanical terms. For Herder, to explain a culture naturalistically meant to see it as an organism, having a purpose and stages of growth.

The first books of the *Ideen* begin with a grand survey of the whole order of nature. Herder locates the place of the earth in the solar system, then the place of man on earth. Following his dynamic view of matter, he sees all of nature as a hierarchy, as so many stages of organization and development of living force. At the summit is man, the culmination of all the powers of nature. It is a central claim of the *Ideen* that there is no distinction between the realms of culture and nature. A culture is portrayed as a continuous development from nature, a unique adaption to specific natural circumstances, such as terrain and climate. Human rationality is also placed within this naturalistic framework. It is seen as the tool by which man learns the skills and information necessary for survival.

The foundation of Herder's concept of history is his anthropology, which he sketches in the first two parts. Herder

stresses that the characteristic feature of mankind is plasticity, the ability to adapt to the most diverse circumstances. While an insect or animal can live in only a specific climate and terrain, mankind can live almost anywhere upon earth. This plasticity also means that a person is shaped by society. Reason, feeling and volition are determined by education, by assimilating a cultural tradition. With this anthropology, Herder broke with the assumptions of a universal human nature, which were so endemic within the seventeenth and eighteenth centuries.

Although he stresses human plasticity, the formative role of culture and nature in shaping our identity and rationality, Herder again refuses to draw relativistic conclusions. As in the 1774 tract, he maintains that there is a single goal to world history, which each culture attempts to realise in its unique way. Now, however, he is more explicit about the nature of this goal. It is what he calls 'humanity' (*Humanität*). This is Herder's ideal of human perfection, the realization of all characteristic powers, the development of reason, feeling and will into a harmonious whole. Such a vague concept was, however, a mere stopgap against relativism. Although he tries to avoid it, Herder faces the dilemma of relativism or ethnocentrism. If he defines his ideal more specifically, he invites the danger of ethnocentrism; but if he leaves it general and abstract, different cultures can realise it in incompatible ways.

The reception of the *Ideen* was marked by controversy. The first two parts were harshly reviewed by Kant, who saw his erstwhile pupil as a rival. Kant felt that Herder had refused to learn the main lessons of his new critical philosophy. He accused Herder of indulging in metaphysics, for there cannot be any empirical verification for the concept of an organic power. He also took strong exception to Herder's naturalistic conception of rationality, which threatened the noumenal/phenomenal dualism of the critical philosophy. Herder was not without his supporters, however, who wrote counter-reviews to Kant. He was defended by K.L. Reinhold, who later became a prominent spokesman for Kant, and Georg Forster, the German explorer and naturalist. This debate about the limits and validity of an organic concept of nature proved to be an important stimulus for Kant's *Kritik der Urteilskraft (Critique of Judgment)* (1790).

5 Metaphysics

Herder's most explicit statement of his general metaphysical views is his *Gott, Einige Gespräche (God, Some Conversations)*, which appeared in 1787. This work is essentially an attempt to rehabilitate Spinozism by interpreting it in organic or vitalist terms (see Spinoza, B. De). It is Herder's contribution to the famous 'pantheism controversy', which dominated German intellectual life in the late 1780s.

This controversy began in 1786 when Jacobi published his *Briefe über Spinoza (Letters on Spinoza)*. Jacobi argued that Spinozism ends in atheism and fatalism. Since he regarded Spinoza's naturalism as the paradigm of a rational or scientific worldview, he was in effect pointing out the dangerous consequences of reason or science for morality and religion.

Herder's tract was a defence of Spinoza, and indeed scientific naturalism, against Jacobi's criticisms. Herder agreed with Jacobi that Spinoza's philosophy represents the spirit of scientific naturalism; but he denied that it ends of necessity in atheism and fatalism (see Jacobi, F. §§2,4). Such a conclusion follows, he argued, only for Jacobi's very orthodox concepts of God and freedom. If we regard God as personal and supernatural, then Spinozism is indeed atheism, for Spinoza thinks that that God is the whole of nature. And if we assume that freedom involves the power of acting otherwise, then Spinozism is also fatalism, since Spinoza also holds that everything is determined to act of necessity. But, Herder insists, we should reject these concepts of God and freedom in the first place, which are nothing more than relics of mythology. To think of God as transcendent is to make him irrelevant to human concerns; and to ascribe personality to him is to anthropomorphize him, making the infinite merely finite. Like Spinoza, then, we should view God as immanent and impersonal, not as transcendent and personal. Similarly, freedom should be understood not as a mysterious power to act contrary to nature, but as 'acting according to the necessity of our own nature alone', just as Spinoza defines it. Given such an interpretation of the concepts of God and freedom, Spinozism does not undermine but supports morality and religion. Indeed, the great attraction of Spinozism for Herder, and many thinkers of his generation, such as Goethe, Schleiermacher, Hölderlin and Novalis, was that it seemed to reconcile science with religion and morality. Spinozism seemed to be the religion of science, the science of religion.

Herder's account of Spinozism ultimately rested upon his organic or vitalist conception of nature. Since naturalism

for Herder means organicism, he interprets Spinoza's naturalism too in vitalist terms. Accordingly, he sees Spinoza's God not as a static substance but as an active force, the force of forces or primal force. If God is such a living force, then, Herder thinks, we can attribute ends to his activity. This will salvage the concept of providence, removing any further traces of fatalism from Spinoza's system.

Of course, such an interpretation was hardly orthodox Spinozism. Although Spinoza had a dynamic view of substance, he banished teleology and understood explanation in purely mechanical terms. Herder has to admit, therefore, that he is transforming Spinoza more than interpreting him. But he insists that he is only trying to make Spinoza's naturalistic spirit consistent with the latest scientific research. All the results of its investigations - the new chemistry, the concepts of electricty and magnestism, Boerhaave's revival of *vis viva* - show that matter is not extension but force. Spinoza's mechanical conception of nature was simply the legacy of his Cartesianism and the physics of his day. If Spinoza were only alive today, Herder suggests, then he too would have an organic conception of God.

The net result of Herder's *Gott* was his vitalistic pantheism or pantheistic vitalism. Whether this was orthodox Spinozism was merely a side issue. By injecting life into Spinoza's static universe, Herder made Spinozism much more appealing for his generation. The revival of Spinozism in late eighteenth-century Germany is a flowering more of vitalistic pantheism than Spinozism proper. That vital pantheism has much of its roots in Herder's *Gott*.

6 Political thought

Although Herder was not a systematic political thinker, political issues were at the very heart of his concerns. His views are scattered in several writings, especially in various passages of his historical works. The first version of *Briefe, die Humanität betreffend (Letters concerning Humanity)*, written in 1792 but never published, examined the issues raised by the French Revolution.

Herder's most important contribution to modern political thought lay in his nationalism. Although he was not the first spokesman for this doctrine, he did play a prominent part in its formulation and transmission. The essence of Herder's nationalism is his model of government and political association. The foundation of both should be, in Herder's view, not a centralized state but the nation, the culture of a people. The people should be joined together by laws that express their common language, religion, customs and history. Their government should represent and defend, not repress or merely tolerate, their distinctive values and way of life.

Herder's nationalism arose from his profound dissatisfaction with the absolutist states of late eighteenth-century Europe, such as Prussia, Russia and Austria. These states held together people of the most diverse cultures by sheer political force, a centralized bureaucracy run according to the whim of the monarch. The cosmopolitan ideals of these states, Herder firmly believed, were often nothing more than a mask for their oppression. In making the nation rather than the state the basis of political order, Herder's motive was largely populist and democratic. He believed that a government based on the culture of a nation reflected the will of the people.

Although Herder was a pioneer of modern nationalism, his views differ from more recent versions in three fundamental respects. (1) Herder saw the nation as the alternative to, rather than foundation for, the centralized bureaucratic state. (2) He believed in the equal value of all cultures and despised all forms of chauvinism or jingoism. (3) Herder's concept of a nation (*Volk*) is based not on race but culture, whose essential component is language.

In wishing to replace the state with the nation, Herder sometimes leaned toward anarchism. He firmly believed that the people could govern themselves and that the modern state was a persistent threat to that right. The most noble end of the state, he wrote in a censored passage of the *Ideen*, is to become superfluous and disappear. His anarchist sentiments became especially apparent in his polemic against Kant in the *Ideen*. He flatly denies Kant's thesis that the purpose of history is to form a perfect state, insisting instead that its purpose is for the state to disappear. He also takes strong exception to Kant's statement that man is an animal in need of a master; in his view, the very opposite is the case: a man in need of a master is an animal.

Although Herder wanted to abolish the state, he never denied the need for some form of government. All that he opposed is the centralized, bureaucratic modern state. True to the Aristotelian tradition, he insisted that man is a political animal with a natural need for government. That some form of authority is necessary to bridle unruly

passions and to arbitrate disputes he fully recognized. He admitted the need, therefore, for written laws and a representative body to administer civic affairs. Yet the form such authority should take is local self-government, because that alone gives the people maximal opportunity to govern themselves about issues that affect them most. This belief in the value of local self-government appears in Herder's admiration for some of the old German city-states, such as Riga and Hamburg.

In the spectrum of late eighteenth-century German politics, Herder belongs firmly to the left wing. He admired the Revolution in France and embraced its ideals of liberty, equality and fraternity. Like many French radicals, he insisted upon the need for a broad franchise and the responsibility of the government to ensure a minimal standard of welfare and education for the people. All his life he had a contempt for the aristocracy and the royalty of the *ancien régime*. Yet Herder was no radical, still less a revolutionary. He became disillusioned with the Revolution after the execution of Louis XVI. His disillusionment stemmed not from any love for the monarchy, but from his hatred of the Jacobins, who, he rightly feared, wanted to establish another form of centralized bureaucratic state. Like many of his contemporaries, he stressed the need for more education and enlightenment as a precondition of social change.

See also: History, philosophy of; Nation and nationalism; Romanticism, German; Vico, G.

FREDERICK BEISER

List of works

Herder, J.G. (1881-1913) *Sämtliche Werke*, ed. B. Suphan, Berlin: Weidmann.(The standard edition, which contains all works mentioned in the text.)

Herder, J.G. *Briefe, Gesammtausgabe*, ed. W. Dobbek and G. Arnold, Weimar: Bohlausnachfolger, 1979.

References and further reading

Barnard, F.M. (1965) *Herder's Social and Political Thought*, Oxford, Clarendon Press.

Berlin, I. (1976) *Vico and Herder. Two Studies in the History of Ideas*. London, Hogarth Press.

Clark, R.C. (1955) *Herder. His Life and Thought*, Berkeley, CA: University of California Press.(The most comprehensive study in English).

Haym R. (1877-85) *Herder nach seinem Leben und seinen Werken dargestellt*, Berlin: Gaertner.(The most complete and authoritative study).

Norton, R.E. (1991) *Herder's Aesthetics & the Enlightenment*, Ithaca, NY: Cornell University Press.

Hermeneutics

Hermeneutics, the 'art of interpretation', was originally the theory and method of interpreting the Bible and other difficult texts. Wilhelm Dilthey extended it to the interpretation of all human acts and products, including history and the interpretation of a human life. Heidegger, in Being and Time (1927), gave an 'interpretation' of the human being, the being that itself understands and interprets. Under his influence, hermeneutics became a central theme of continental philosophy. Hermeneutics generates several controversies. In interpreting something do we unearth the author's thoughts and intentions, imagining ourselves in his position? Or do we relate it to a wider whole that gives it meaning? The latter view gives rise to the hermeneutic circle: we cannot understand a whole (for example, a text) unless we understand its parts, or the parts unless we understand the whole. Heidegger discovered another circle: as we inevitably bring presuppositions to what we interpret, does this mean that any interpretation is arbitrary, or at least endlessly revisable?

1 The beginnings of hermeneutics

The Greek *hermeneuein* means to express, explain, translate or interpret; *hermeneia* is interpretation and so on, often the interpretation of a sacred message. Plato called poets the *hermenes* - interpreters - of the gods. Philosophers interpreted Homer allegorically. Augustine interpreted the Old Testament as allegory, using Neoplatonic concepts and recording the rise of the soul above the literal and the moral senses of the text to its spiritual sense. Allegorical interpretation remained the norm throughout the Middle Ages. With the Reformation, especially in Germany, *hermeneia* became more explicit and systematic. The word *hermeneutica*, the 'art of interpretation', appeared in the title of J.C. Dannhauer's 1654 work *Hermeneutica sacra sive methodus exponendarum sacrarum litterarum*. Protestants had to interpret the Bible properly; they appealed to it against Roman Catholicism. They rejected allegorical interpretation and insisted on the letter of the text, hoping to retrieve its meaning from distortions introduced by the Church and by scholasticism. Biblical exegesis did not remain isolated from interpretation of other texts. Spinoza, in *Tractatus theologico-politicus* (1670 ch. VII: §94), affirmed that the 'standard of biblical exegesis can only be the light of reason common to all'. For Spinoza, biblical exegesis became biblical criticism, and this involved history. Since reports of miracles fall short of rational standards of belief, we must explain why the authors of the Bible and their contemporaries believed in miracles (see Spinoza, B. de §14).

Johann Ernesti declared in his manual of hermeneutics (1761: 7) that the 'verbal sense of Scripture must be determined in the same way in which we ascertain that of other books'. Other texts in need of interpretation were legal documents and the works of classical antiquity, and these disciplines also contributed to hermeneutics. Significant advances were made by two classicists, Friedrich Ast and Friedrich August Wolf. Ast, in *Grundlinien der Grammatik, Hermeneutik und Kritik (Elements of Grammar, Hermeneutics and Criticism)* (1808), distinguished different levels of understanding a text. The first is 'historical', establishing the authentic text by comparing different manuscripts and deploying knowledge of the history and other writings of the period; to this understanding corresponds the 'hermeneutics of the letter'. The second is grammatical, corresponding to the 'hermeneutics of the sense': we understand the meaning of the words and sentences in the text. The third is spiritual: we ascend from the literal meaning to the spirit (*Geist*) of the author and of their society ('spirit' means 'outlook', 'mentality' or 'worldview'; it need have no theological or psychological connotation). In his lectures on the 'encyclopaedia of classical studies' from 1785 to 1807, Wolf defined hermeneutics as the 'science of the rules by which the meaning of signs is discerned' (1831: 290). Its aim is to 'grasp the written or even merely spoken thoughts of someone else just as he would have them grasped' (1831: 293). This involves not only knowledge of the language of the text, but also historical knowledge, a knowledge of the author's life, of the history and geography of their country. An interpreter should ideally know everything known by the author. Wolf proposed many rules for handling problems of interpretation, but insisted that an interpreter needs a 'lightness of soul' that 'quickly attunes itself to foreign thoughts' (1831: 273). Knowing rules is not enough; we need a skill in applying rules which no rule can guarantee.

2 Schleiermacher

Friedrich Schleiermacher welded these partial theories into a single discipline, embracing the interpretation of all

texts, regardless of genre and doctrines (he interpreted Heraclitus and Plato as well as the Bible.) At each level of interpretation we are involved in a hermeneutical circle. We cannot know the correct reading of a passage in a text unless we know, roughly, the text as a whole; we cannot know the text as a whole unless we know particular passages. We cannot know the meaning of a word unless we know the meanings of surrounding words and of the text as a whole; knowing the meaning of the whole involves knowing the meaning of individual words. We cannot fully understand the text unless we know the author's life and works as a whole, but this requires a knowledge of the texts and other events that constitute his life. We cannot fully understand a text unless we know about the whole culture from which it emerged, but this presupposes a knowledge of the texts and so on that constitute the culture. Not only is there circularity within each level of interpretation, but also between levels. We cannot decide on the correct reading of a particular passage unless we already know something of its meaning and also of the author's life or culture. But how are we to acquire this knowledge, if not from such texts as this?

The hermeneutic circle is less mysterious than often supposed. A text need not be uniformly problematic. A hopelessly corrupt manuscript (or misprinted book) may be undecipherable. But if the manuscripts are reliable over the bulk of the text, the interpreter brings the knowledge of this part to bear on those parts where the manuscripts are corrupt. Not all words and sentences are equally obscure; the relatively transparent supply a clue to the relatively opaque. Again, understanding is a matter of degree. I cannot fully understand a text unless I fully understand each word and sentence, and I cannot fully understand any word or sentence unless I understand the whole. If full understanding and blank incomprehension were the only alternatives, I could not understand a text of any length or complexity. But understanding is not like that: I can roughly understand a text without fully understanding it, and rough understanding enables me to decipher particular parts.

In 1813 Schleiermacher wrote: 'Essentially and inwardly, thought and its expression are completely the same' (1959: 21). This suggests that what we understand is the literal meaning of a text, what the words mean or meant. In 1819 he wrote: 'The art can develop its rules only from a positive formula, and this is: the historical and divinatory, objective and subjective reconstructing of a given utterance' (1959: 87). This suggests that there may be more to an author's thought than the meaning of their words, and that the interpreter has to unearth the thought. The thought may differ from the meaning of the words for several reasons: authors may express themselves badly, a slip of the pen or the tongue, a lack of verbal facility or malapropism. (When someone speaks of 'mitigating against' or writes 'It is worth nothing that…', we assume that they meant 'militating against' and 'It is worth noting that…'.) To understand an utterance fully, we often go beyond the meaning of the words and ask about its author's intentions: Did the author mean it seriously or as a joke? Did they mean the word in this sense or in that? Was the author implicitly criticizing so-and-so? Conversely we may discern more in an author's words than we can plausibly attribute to their conscious thoughts, and invoke their unconscious thoughts or the 'spirit' of the author or his culture. Or we may appeal to the audience. The question 'What does this text mean?' can be expanded in two ways: (1) 'What does/did the author mean by the text?', (2) 'What does/did the text mean to its audience?' These expansions can in turn be interpreted in different ways. What constitutes the author and their meaning? We may restrict this to the author's conscious thoughts and intentions, or we may include their subconscious thoughts and intentions, or even the spirit of their age, implicitly granting this a share in their authorship. Who are the audience? It may be the author's contemporaries, or a later audience such as ourselves. The answers to the two questions are unlikely to coincide, if author and audience belong to different times or cultures. What Shakespeare meant by *Hamlet* is not what *Hamlet* means to a modern audience, unless that audience consists of skilful hermeneuticists. The answers are most likely to coincide if the audience is contemporary with the author; then author and audience share the same 'spirit', even if they do not have the same creative ability.

When Schleiermacher aimed to reconstruct the verbal meaning of a text, in the belief that 'thought and its expression' are identical, he was answering the question 'What did the text mean to its contemporary, cultivated audience?' When he tried to reconstruct the author's thought, in the belief that this need not be the same as its expression, he was answering the question 'What did the author mean by the text?' How can we know what Shakespeare meant (that is, what he had in mind)? Can we know it in much the same way as we know what a contemporary with whom we are conversing has in mind? Our minds are not wholly different from Shakespeare's; there is a 'spiritual affinity' between us. If we acquire sufficient knowledge of his life and works, we can imaginatively step into his shoes, reproducing his thought. It is possible so to step into someone else's shoes;

novelists often do it. It is doubtful whether we need to do this to know what someone is thinking: I may know that a dog wants a bone without imaginatively reproducing this want in myself.

3 Dilthey

Wilhelm Dilthey owed his acquaintance with hermeneutics to his theological training, but he used it to answer the question 'How do the social or human sciences differ from the natural sciences?' While the natural sciences explain (*erklären*), the social sciences understand (*verstehen*). They understand not simply texts and utterances, but any meaningful 'objectification' or 'expression' of human life: gestures, actions, one's own or another's life, paintings, institutions, societies, past events. There are two types of understanding. First, the understanding of simple expressions such as an utterance, an action or a gesture of fear. Here there is no gulf between the expression and the experience expressed: we understand immediately with no inference. Such understanding presupposes a medium 'common to the I and the You', an 'objective spirit' within which expression and understanding occur: a shared language and culture. Secondly, there are 'higher forms of understanding', dealing with complex wholes, such as a life or a work of art. A part has a meaning (*Bedeutung*) grasped by elementary understanding, the whole has a sense (*Sinn*) resulting from the ordered composition of its parts and grasped by higher understanding. Higher understanding is often provoked by a failure of elementary understanding. If I cannot immediately understand a person's action, I explore their culture or their life as a whole. If I cannot understand a sentence, I may interpret the whole book.

Often when elementary understanding fails, this is because a person - the author of a text, a gesture or an action - is unusual, and cannot be understood by the normal canons of objective spirit. To understand what the author says or does, we need to understand them in their individuality. Thus, higher understanding usually involves the understanding of individuals, not simply the general understanding appropriate for daily life. I also understand myself: by elementary understanding I know immediately that I am hungry, jealous and so on without recourse to my expressions. By higher understanding I make sense of myself, of my life as a whole. Higher understanding may again be stimulated by a failure of immediate understanding: how can I be jealous, I wonder, or how can I have done this? In higher understanding of myself, I become aware of my individuality, of what differentiates me from others.

In the 1890s Dilthey regarded psychology as the foundation of the social sciences. Later, hermeneutics displaced psychology. What interests the social sciences is not the 'soul', an individual's psychological processes, but 'spirit', the shared cultural world. The meaning of a play is independent of its author's 'soul'. Even if a work expresses joy or grief, they are states not of the author but of the 'ideal person' in whose mouth he puts his experience. Psychological life, even one's own, is known by the interpretation of its expressions: 'Man knows himself only in history, never by introspection' (1981: 348). Interpretation of history does not capture the essence of humans in a formula. It reveals to us the diverse possibilities of humanity, liberating us from the confines of the present.

4 Heidegger

Historical studies depend, Dilthey held, on our awareness of a human life as a coherent, 'historical' whole, embedded in a historical context. Martin Heidegger also connected questions about the meaning of historical texts with questions about the meaning of life. Texts like St Paul's letters cannot be understood from dictionaries and grammar-books alone; we need to understand the lives and situation of the author and his audience. In the case of any text, but especially those of a difficult philosopher such as Aristotle, we need to explore our 'hermeneutical situation', the situation, shaped by the past, which imposes on us the presuppositions we bring to the understanding of the text. Are the terms in which we interpret Aristotle appropriate? If not, how can we account for this apparent degeneration in our conceptual apparatus? These are questions about the present, and not just about contemporary philosophy, but about contemporary life and our tendency to misconstrue the past. So Heidegger moves on, in his lectures of 1923, to a 'hermeneutic of facticity', an interpretation of the human being ('Dasein') and everyday life.

Heidegger's *Being and Time* (1927) is a 'hermeneutic', in several ways. It explores Dasein's own understanding and interpreting: Dasein understands and interprets, not incidentally and sporadically but essentially and constantly. It understands - it knows its way around in the world as a field for its own activities. It interprets entities within the world - it sees a table as a table, a chair as a chair (see De re/de dicto). Such understanding and

interpretation are prior to the sciences. Before I interpret a document I see it as a document; before I do geology I see rocks as rocks. Interpretation involves presuppositions: to interpret something as a book I must be familiar with a world in which books have a place, a world of rooms, furniture, shelves, readers. Dasein also interprets itself. It regards itself, for instance, as a cobbler or a sailor. It regards its own life in a certain way. Implicitly in everyday life, more explicitly in philosophy, Dasein misinterprets itself - as a rational animal, a thinking substance, or a machine.

Heidegger describes Dasein's essential features, including interpretation and self-interpretation. Since philosophers themselves are Dasein, they display the same tendencies as Dasein in general. Hence, in studying Dasein, they understand and interpret it, continuing, on a higher, conceptual, plane, the self-interpretation that is an inescapable feature of all Dasein. Like all interpretation, Heidegger's involves presuppositions: a preliminary understanding of Dasein such as all Dasein has, a certain way of viewing Dasein (with regard to its 'being', rather than, say, its biological characteristics), and concepts to be applied to it, such as 'existence'.

The interpretation of Dasein and of being in general involves interpretation of texts. Since Dasein misinterprets itself, we must peel away layers of misinterpretation to see it as it is. These misinterpretations occur in their original purity in such philosophers as Kant, Descartes and Aristotle. We study them to see what they got right and where they went astray, to disclose and assess their influence on our hermeneutical situation and, where appropriate, to free ourselves of their influence. Heidegger interpreted such texts mainly in later works, but *Being and Time* foreshadows his procedure. Words do not have fixed, univocal meanings independently of their use and application. Meanings accrue to words from the significant interrelationships that constitute our world. A 'hammer' is not simply an 'implement for banging': the meaning of the word derives from the context of bench, nails, wood, workshop and customers that constitutes the 'world' of the craftsman. What a word means depends on the world of its user: by 'transport', 'freedom' or 'education', Aristotle does not mean the same as we do since he inhabited a different world. To understand a text we need to go beyond dictionaries and grammars to reconstruct the world of its author and the 'possibilities' it offered.

Later, Heidegger avoided the word 'hermeneutics'. But he continued to interpret texts, poetical as well as philosophical, in his quest for the 'meaning of being'. He equivocates on the question whether we can interpret a text definitively. Our interpretations of the past are bound to our hermeneutical situation and open to future revision. *Being and Time* implies that the meaning of an event or of a life, if not of a text, is what it means *for* us/me, depending on the significance we/I confer on it by and in our/my decisions for the future. Conversely, he claims - for all his 'violence' to the literal text - to unearth the meaning of, for example, Aristotle, with scarcely a hint that his own interpretation might later be seen, with equal justification, as yet one more misinterpretation. At all events, the hermeneutical circle now embraces interpreters and their presuppositions, as well as the text, the author and their culture. The prior understanding of the whole, that Schleiermacher and Dilthey saw as required for the interpretation of a part, can come only from the interpreter's own presuppositions. These may however be revised in the course of the interpretation.

5 After Heidegger

Heidegger's hermeneutics has been exploited by Bultmann, Ricoeur and Derrida, but his closest follower is Gadamer. Gadamer also insists that we recapture the context in which an author wrote, taking into account the intended audience and the questions the author was answering. Interpretation presupposes a historically determined 'pre-understanding', a 'horizon'; it involves a 'fusion of horizons', the horizons of the past and of the present. We cannot be sure that our interpretation is correct, or better than previous interpretations. Our interpretation, and our verdict on previous interpretations, is open to future revision. In interpreting a past text we explore our own pre-understanding as much as the text itself.

With Gadamer and others, hermeneutics has returned to its ancient and medieval roots. We no longer seek what the author meant by a text, but what the text means to, or for, us. The medieval justification of this is that God, the ultimate author of the text, can inscribe in it any meaning he chooses, however allegorical or anachronistic. The moderns justify it by appeal to the non-existence, indeterminacy, inaccessibility, or irrelevance of an author's intentions, or to the historically variable presuppositions of interpretation. Dilthey's extension of hermeneutics to lives and to historic events supports this trend. The 'meaning' of the French Revolution cannot be what its author(s) meant by it, or even what it meant to its contemporary audience. It is what it 'means' to successive later

audiences or even what they make of it by their own plans and decisions. Few argue, however, that the interpretation of a text is entirely at the whim of the interpreter. This would subvert the communicability of the hermeneuticist’s scepticism by allowing opponents to interpret its expression as they please.

See also: Chinese Classics; Hermeneutics, biblical; Legal hermeneutics; Midrash; Vedānta

MICHAEL INWOOD

References and further reading

Ast, F. (1808) *Grundlinien der Grammatik, Hermeneutik und Kritik (Elements of grammar, hermeneutics and criticism)*, Landshut: Thomann.(Outlines a universal hermeneutics, based on a single spirit running through human history. It contains the first formulation of the hermeneutic circle.)

Bruns, G.L. (1992) *Hermeneutics Ancient and Modern*, New Haven, CT: Yale University Press.(A good introduction to the subject, covering the whole history of hermeneutics.)

Dannhauer, J.C. (1654) *Hermeneutica sacra sive methodus exponendarum sacrarum litterarum (Sacred hermeneutics: the method of expounding holy scripture)*, Strasbourg: J. Städel.(Apart from its innovative use of the word ‘hermeneutica’ this work proposes rules for the sound interpretation of the Bible.)

Dicenso, J. (1990) *Hermeneutics and the Disclosure of Truth: A Study in the Work of Heidegger, Gadamer and Ricoeur*, Charlottesville, VA: University of Virginia Press. (A sound work on these important figures.)

Dilthey, W. (1981) *Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften (The construction of the historical world in the human sciences)*, ed. M. Reidel, Frankfurt am Main: Suhrkamp.(Dilthey’s main work on the application of hermeneutics to history.)

Ernesti, J.A. (1761) *Institutio interpretis Novi Testamenti*, Leipzig: Weidmann; trans. M. Stuart as *Elements of Interpretation*, Andover: M. Newman, 1827: trans. C.H. Terrot as *Principles of Biblical Interpretation*, Edinburgh: T. Clark, 1832-3.(A work of biblical interpretation which helped to refine grammatical analysis and influenced Schleiermacher’s early lectures on hermeneutics.)

Gadamer, H.-G. (1960) *Wahrheit und Methode*, Tübingen: J.C.B. Mohr; trans. W. Glen-Doepel as *Truth and Method*, London: Sheed & Ward, 1975.(Surveys modern aesthetics and the theory of historical understanding, and presents Gadamer’s own ‘philosophical hermeneutics’.)

Gadamer, H.-G. (1976) *Philosophical Hermeneutics*, trans. D. Linge, Berkeley, CA: University of California Press.(A selection of Gadamer’s essays on hermeneutics, phenomenology and Heidegger.)

Grondin, J. (1994) *Introduction to Philosophical Hermeneutics*, New Haven, CT: Yale University Press.(A wide-ranging historical introduction that considers the major hermeneuticists from antiquity to Habermas.)

Grondin, J. (1995) *Sources of Hermeneutics*, Albany, NY: State University of New York Press.(A difficult but rewarding work on the contributions to hermeneutics of Heidegger and Gadamer.)

Heidegger, M. (1927) *Sein und Zeit*, Halle an der Salle: Max Niemeyer; trans. J. Macquarrie and E. Robinson as *Being and Time*, New York: Harper & Row, 1962; trans. J. Stambaugh, Albany, NY: State University of New York Press, 1996. (Heidegger’s first major work, which initiated the modern interest in philosophical hermeneutics.)

Heidegger, M. (1988) *Ontologie (Hermeneutik der Faktizität) (Ontology (Hermeneutics of facticity))*, ed. K. Bröcker-Oltmans, Frankfurt am Main: Vittorio Klostermann. (Volume 63 of Heidegger’s collected works contains his lectures from the summer of 1923, in which he examines Dasein’s self-interpretation.)

Laks, A. and Neschke, A. (eds) (1990) *La naissance du paradigme hermeneutique: Schleiermacher, Humboldt, Boeckh, Droysen (The birth of the hermeneutical paradigm: Schleiermacher, Humboldt, Boeckh, Droysen)*, Lille: Presses universitaires de Lille.(Collection of essays focusing on the lesser known hermeneuticists, especially the linguist Wilhelm von Humboldt.)

Mueller-Vollmer, K. (ed.) (1986) *The Hermeneutics Reader*, Oxford: Blackwell.(A judicious selection of hermeneutical writings in translation, from Schleiermacher to the present.)

Palmer, R.E. (1969) *Hermeneutics: Interpretation Theory in Schleiermacher, Dilthey, Heidegger and Gadamer*, Evanston, IL: Northwest University Press.(A good historical introduction, which also discusses the philosophical problems in hermeneutics.)

Schleiermacher, F.D. (1959) *Hermeneutik*, ed. H. Kimmerle, Heidelberg: Carl Winter, Universitätsverlag; trans. J. Duke and J. Forstmann as *Hermeneutics: The Handwritten Manuscripts by F.D. Scleiermacher*, Missoula, MT: The Scholars Press, 1977.(The definitive edition of Schleiermacher’s manuscripts on universal hermeneutics.)

Spinoza, B. de (1670) *Tractatus Theologico-politicus (Theological-Political Treatise)*, in *The Chief Works of Benedict de Spinoza*, vol. 1, trans. R.H.M. Elwes, New York: Dover, 1951.(A pioneering work of rational

interpretation and criticism of the Old Testament.)

Tully, J. (ed.) (1988) *Meaning and Context: Quentin Skinner and his Critics*, Cambridge: Polity Press.(A collection of essays by and about Skinner, concerned with interpretation in political philosophy and the history of ideas.)

Wolf, F.A. (1831) *Vorlesung über die Enzyklopädie der Altertumswissenschaft (Lectures on the encyclopedia of classical studies)*, ed. J.D. Gurtler, Leipzig: Lenhold. (A great classical scholar's ideas about the interpretation of ancient texts.)

Hermeneutics, biblical

Hermeneutics has traditionally been defined as the theory of interpretation. Biblical hermeneutics concerns the interpretation of biblical texts. But 'interpretation' tends to reflect the nature of the discipline only from ancient times to about 1960. Increasingly it has come to be seen not as a tool used for difficult or obscure texts, or even for the application of such texts to the present, but as a theory of understanding in the broadest sense. It currently also relates to views of contextual theories of meaning and truth, in contrast to formalist approaches.

From ancient times until about 1800, philosophy played a minimal role in biblical hermeneutics. The subject concerned theology in the context of grammatical, philological, historical and linguistic inquiry. With the work of Schleiermacher in the early years of the nineteenth century, however, hermeneutics entered a new phase. It became a transcendental discipline, seeking to explore the conditions under which the understanding of texts becomes possible at all. In the era following Schleiermacher, theorists drew on the work of Dilthey and Heidegger, among others.

A third phase began with the work of Hans-Georg Gadamer in the early 1960s. Gadamer replaced the Enlightenment preoccupation with 'method' in the context of 'science' and 'reason' with a hermeneutics which took full account of the interpreter's prior situatedness within a given historical tradition. This angle of approach was developed further by Habermas, who noted the role of 'interest' in understanding, and by Ricoeur, who, in dialogue with Freud and others, stressed the role of 'suspicion' as well as 'listening' in hermeneutics. Barthes and Derrida challenged the very notion of a 'given' text, shifting emphasis to construals by society and by readers which reflect motivations not immediately apparent from the supposed messages of texts.

This raises a multitude of fundamental questions for biblical hermeneutics and theology. If texts are no more than shifting constructs, what may still be said about divine grace or revelation? Do sacred texts merely mediate idolatrous constructs? How may hermeneutics serve to unmask interests which interpreters bring to the text and tempt them to use texts manipulatively?

1 The ancient world

Processes of interpretation of earlier biblical texts were already at work in parts of the Hebrew Bible. But hermeneutics entails conscious reflection on what it is to interpret, and in Hebrew tradition this did not explicitly emerge until the era of rabbinic theology, from 70 AD onwards. In the absence of the Temple and its institutions, reflection about the application of ancient biblical texts to the present became a matter of detailed concern. The so-called 'rules of interpretation' of Rabbi Hillel, though, were little more than rule-of-thumb rabbinic procedures (see Hillel ben Samuel of Verona).

Among the ancient Greeks, more conscious hermeneutical reflection arose over the issues of whether the 'classic' text of Homer could be interpreted allegorically, in such a way that stories of the exploits of the gods could be read as information about the world - about the properties of different substances, for example, or about virtues, vices and values. In the *Republic*, Plato expresses reservations about such allegorical 'under-meaning' (*hyponoia*), but many of the Stoics interpreted Homer allegorically and Zeno of Citium read parts of Hesiod in this way. The Jewish writer Philo of Alexandria drew on this tradition to interpret the Old Testament allegorically. For example, God cannot 'plant a garden' (Genesis 2: 8); the verse signifies God's 'planting' of virtue. Philo thus smoothed away both anthropomorphisms and time-bound regulations to open a path to the Hebrew Scriptures for educated Greek and Roman readers.

Whether, or in what way, the New Testament writers interpret the Old Testament allegorically remains a matter of controversy. Certainly they use typology (correspondence of events); it is debatable whether they use allegory (correspondence of ideas). Paul's interpretation of Genesis 16: 15 and 21: 2 in Galatians 4: 24-6 offers one of the most probable examples of allegory, although some insist that it is typology. In the sub-apostolic 1 Clement (*c.* AD 95), the scarlet thread in the house of Rahab (Joshua 2: 18, 21) is seen as an allusion to the blood of the cross. Gnostic writers used allegorical interpretation extensively, not least because it seemed to offer a secret key for initiates. This drove some of the Church Fathers, notably Origen (§1-2), to the first explicitly Christian hermeneutical reflections on method. The difficult question of the status of the Hebrew Old Testament as Christian

scripture was among those considered, not least because Marcion had rejected it in the second century.

2 Medieval, Reformation and post-Reformation issues

Origen argued that although the interpreter began with the literal (that is, semantic, grammatical or lexicographical) meaning, it was legitimate to proceed to a 'spiritual' meaning for pastoral purposes. An ambivalence towards allegorical interpretation emerged in Christian tradition. Origen's reason for being willing to go 'beyond the letter' was different from that of the Gnostics, although an intermediate and ambivalent position was held by Clement of Alexandria. Origen recognized the importance of the public domain rather than 'secret' meanings, and was concerned that 'the rude vessel of words' should be 'brought to life'. Here he anticipated the late-twentieth-century turn from author to reader. In the School of Antioch, however, the fourth-century writer John Chrysostom approached interpretation as a matter of disclosing 'the mind of the writer'. Hence, although he also used allegorical interpretation, he did so mostly in sermons, and to a much lesser extent in didactic expositions. On the other hand, the contrast between Alexandria (Origen and those who shared his method) and Antioch (Chrysostom and others, especially Theodore of Mapsuertia) is often exaggerated in popular works. By the time of Gregory the Great of Rome, in the second half of the sixth century, the notion of 'three senses of Scripture' (literal, moral and spiritual) associated with Origen had become the Gregorian tradition of four senses: the literal (semantic and historical); the allegorical (a theological or spiritual extension); the moral (an ethical application); and the anagogical (that which embraces the future).

The medieval period witnessed the rise of the *Glossa ordinaria* (Standard Gloss), which provided a brief explanatory comment on a given chapter or verse. Among some less learned or more manipulative writers these glosses were sometimes arbitrary or without due historical understanding. But some major thinkers, notably Hugh of St Victor, contributed commentaries (for example, on Lamentations) and biblical homilies (for example, on Ecclesiastes) which placed a renewed emphasis on the historical understanding of the text. Reading based on historical and textual study led on to meditation or spiritual reflection. The ultimate aim, however, was to combine such reading with purity of life, to seek loving contemplation of divine thoughts behind Scripture.

Thomas Aquinas discussed linguistic issues, such as the status of metaphor, raised by the biblical text. He was aware that the Gregorian 'fourfold sense' had the capacity to confuse readers unless its logic was clear. However, since he held that scripture reveals both God's being and God's will and purpose, Aquinas supported the notion of moral, allegorical and anagogical (eschatological) meanings in addition to the semantic or literal meaning, even if this last remained primary. Thus he lent his authority to the prevailing methods of medieval writers.

The Reformers reacted against these 'four senses' with grave suspicion. In his early writings, Luther (§6) followed medieval tradition. But he moved away from it for two main reasons: first, positively, because of his engagement with the linguistic and rational tools of Renaissance learning; and second, negatively, because of his perception that supposed 'spiritual' meanings could be used in manipulatory ways to serve the interests of ecclesiastical authority and to give the Bible 'a wax nose' to be pushed into any shape. In his middle and later years, he regarded all but the 'plain' meaning as 'nothing but rubbish'. For Calvin (§2) even more, the hermeneutical goal is to apprehend 'a true single [*simplex*] sense'.

The post-Reformation era witnessed a steady growth of works on hermeneutics. It seems that the term 'hermeneutics' first explicitly appeared as a title for a work in 1654, with J.C. Dannhauer. Major scholars of the eighteenth century included J.A. Turretinus, J.M. Chladenius, J.A. Ernesti and J.S. Semler.

3 Schleiermacher and Dilthey

Schleiermacher (§9) was nurtured in the tradition of Moravian pietism and became associated with the blossoming Romantic movement. He accepted the transcendental challenge of Kant's three *Critiques*, and perceived clearly the inadequacy and circularity of any system of hermeneutics that served some prior doctrinal understanding of texts. He asked two transcendental questions: how is Christian theology possible? And how do processes of understanding become possible?

Schleiermacher was the first theorist to perceive that the problem of hermeneutics lay in holding together each side of three sets of issues: how understanding in its 'divinatory' or trans-rational dimension interacts with the critical, rational appeal to patterns or regularities of criteria; how understanding the elements or components of texts relates

to a more intuitive anticipation of the text as a whole; and how understanding what is to be understood links or engages with a prior or provisional understanding of that to which the subject matter of the text relates. This last issue became known as the problem of 'pre-understanding', for which the German *Vorverständnis* became a technical term in later hermeneutics. As understanding of the text begins to develop, the revised or fuller understanding feeds back into the 'pre-understanding' to produce more appropriate questions and a better readiness to understand.

Simultaneously, understanding parts of the text depends on a provisional, quasi-intuitive leap into an anticipatory understanding of the whole, but at the same time corrects it, while the reverse process of correction and revision also takes place. Both of these processes constitute different versions of 'the hermeneutical circle'.

In biblical studies this hermeneutical approach had several effects. First, it underlined both the importance and the limits of the rational or 'scientific' aspect of enquiry. Dilthey, writing at the beginning of the twentieth century, took up this issue in his appeal to 'life' rather than to mere 'thought' as the key category in hermeneutical understanding. In his exposition of Schleiermacher, he took up and developed the latter's 'psychological' axis of hermeneutics, namely the need to stand in the shoes of 'the other' in order to understand the other on the basis of shared life and shared humanity. In his criticism of rationalism and empiricism, Dilthey argued that in the veins of the knowing subject (as defined by Locke, Hume and even Kant), no real blood flowed.

Second, the hermeneutical approach led to an emphasis on historical reconstruction in biblical studies, which cohered with the rise of biblical criticism. However, against many popular misunderstandings of Schleiermacher, it must be stressed that this arose not from preoccupation with 'genetic origin', but from the Romantic concern with capturing the living vision of which the textual remainder was a mere deposit. It was for this reason, and not out of mere antiquarian curiosity, that Schleiermacher worked backwards from the text to what lay behind it. It was a side effect in biblical studies that this fostered an approach to texts which used them as mere sources for historical reconstruction. This hermeneutical reductionism and instrumentalism would not have been endorsed by Schleiermacher, who is frequently blamed for it. The level of complexity and sophistication which characterizes his work is vigorously defended in Thiselton (1992).

4 The turn to existentialism

Martin Heidegger (§§2-3) took up and developed further the notions of 'the hermeneutical circle' and a prior 'life relationship' to that which the interpreter seeks to understand. But whereas Dilthey saw 'life' as part of a common humanity, Heidegger stressed the historical finitude and unique 'givenness' or situatedness of the interpreter within pre-given horizons of orientation and practical purpose. In continuity with Kierkegaard and Nietzsche, he called attention to the close connection between finite *Existenz* and the directionality of human will towards its own goals.

In the field of New Testament studies, Rudolf Bultmann (§§2-3) viewed this perspective as fruitful for a hermeneutical engagement with texts. New Testament texts, he argued, are not to inform neutral curiosity, but to challenge the human will. The interpreter may be misled by what appears to speak descriptively of God, Christ or human persons. For this, according to Bultmann, is the language of 'myth', which obscures its genuine purpose. Neither God nor Christ nor persons are to be 'objectified' as being or substance as such. Rather God addresses humanity through the pseudo-descriptive language only to summon humanity from inauthentic existence orientated towards the past and under law and bondage, to an authentic mode of existence made possible through the event of Christ (see Existentialist theology §3).

The relation between operative linguistic function and historical states of affairs for Bultmann remains a complex and controversial issue. 'Right-wing' critics accuse him of too readily collapsing the Bible into Christian existentialism; 'left-wing' critics accuse him of arbitrarily clinging to the factual reality of the cross, or of failing to carry through his 'demythologizing' by exposing 'God' to be no more than a functionally useful linguistic cipher. For this reason, 'the Bultmann School' broke apart in different directions. Ernst Fuchs, for example, provides a bridge between Heidegger and Gadamer, since Fuchs, with Ebeling, identifies the creative eventfulness of language as the key issue in hermeneutics.

5 Issues raised by Gadamer, Habermas, Ricoeur and Derrida

The publication of Hans-Georg Gadamer's *Truth and Method* in 1960 constitutes the most important turning point for the subject since Schleiermacher, who founded the modern discipline (see Gadamer, H.-G.). The term 'method' in the title stands in a negative, or at least ironic, relation to 'truth'. Gadamer attacks as an artificial abstract construct the pure 'reason' and 'method' of Enlightenment thought in Descartes (as opposed to Vico) and in all thinkers who naïvely assume that methods of science can disclose all truth. He agrees with his teacher Heidegger that the individual thinker is shaped by their situatedness within historical tradition. He writes: 'The self-awareness of the individual is only a flickering in the closed circuits of historical life' ([1960] 1989: 276). But this does not exclude the possibility of understanding, for 'pre-judgments' (*Vorurteile*) constitute a 'historical reality' in which the interpreter participates.

Gadamer clarifies the issue by drawing on the analogy of games. The presuppositional stance (that is, the ways in which the rules of the game determine goals and what counts as an appropriate move or strategy) determines 'meaning' and 'reality' for players more than their own conscious reflection. Thus a tradition of 'effective-history' (*Wirkungsgeschichte*) lies beneath the distinctive and indeed unique 'performances' which characterize each game, or each interpretive or communicative event. Gadamer has thus transposed the relation between the whole and the parts which occupied Schleiermacher and Dilthey into a post-Hegelian rather than post-Kantian key. All understanding, he believes, is hermeneutical, not value-neutral. This is what he means by 'the universality of the hermeneutical problem'.

Habermas (§§1-2) was quick to identify a lack of critical social awareness in Gadamer's hermeneutics. Where Gadamer spoke of 'pre-judgment', Habermas spoke of 'interest'. What determines the horizons of understanding, he argues, is social interests (which include power interests) on the part of interpreters and, more especially, their communities of interpretation. But Habermas refuses, as against Richard Rorty, to reduce this to social pragmatism. Through a carefully elaborated theory of communicative action, he denies that communities of interpretation are so radically conditioned by social and historical context that no trans-contextual criteria of meaning and understanding can remain operative. Thus his earlier work, *Knowledge and Human Interests* (1978), leads on to his more complex and valuable work, *The Theory of Communicative Action: The Critique of Functionalist Reason* (1984).

Paul Ricoeur moved from work on human will and fallibility to theories of symbol, and then through hermeneutics to a dialogue between hermeneutics and structuralism, ending with a dialogue between hermeneutics, metaphor and narrative theory. His later work returns to the relation between hermeneutics and personal agency. He is probably best known for his aphorism that 'the symbol gives rise to thought'. His work on hermeneutics led him to explore the phenomena of self-deception and 'double meaning' in Freud's *Interpretation of Dreams*. The element of disguise and deception led to his formulation of 'a hermeneutic of suspicion'. He writes that Freud's work 'compelled me to enlarge my first concept of hermeneutics beyond a mere semantic analysis of double-meaning expressions' ([1975] 1978: 318).

While Ricoeur rejects the 'reductionism' of Freud's theories as an explanation of human life, including religion, he accepts Freud's emphasis on the capacity of the self for self-deception. Hermeneutics, therefore, in Ricoeur's view, performs the double task of 'destroying idols' (that is, unmasking the wishes that condition interpretation and cause projection of one's own desires into texts) and 'listening to symbols'. The first entails a 'vow of rigour', the second a 'vow of obedience'. The symbolic level of meaning appears as metaphor in relation to sentences, and as the projection of narrative-worlds in relation to sentences which combine together to present a structured whole. Some of the more controversial aspects of his hermeneutics concern the 're-figuration' or 'construal' of narrative, and a readiness to speak at times of 'the death of the author' (1976: 29-30).

This attempt to speak of 'the death of the author' for interpreting written texts stems primarily from Roland Barthes. Barthes' semiotics, in turn, paved the way for Derrida. Jacques Derrida stresses that writing (in contrast to oral speech) entails 'the absence of the signatory' and even 'the absence of the referent' ([1967] 1976: 40-1). Further, no text is a 'finished' product or a 'given'. Allegedly on the basis of a theory of semiotics and textuality, he sees textual meaning as endlessly 'deferred'. On the basis of a legacy drawn partly from Freud, Husserl, Nietzsche and Heidegger, he proposes a hermeneutic that has affinities with postmodernism and articulates deconstructionism. In effect, it seems to reduce texts to an ever-shifting flux, which is constantly subject to new forms as new conventions and societal assumptions re-contextualize what on the surface appears to count as

meaning and truth.

6 Post-Gadamerian biblical hermeneutics

Two aspects of Gadamer's work have been of particular importance in the field of biblical studies. First, while he rightly insists that the interpreter must respect the distinctive horizon of the text, Gadamer urges that understanding works towards the goal, at least in principle, of a fusion of the two horizons of meaning and understanding. Second, his work stresses the distinctiveness and uniqueness of each 'performance' as an actualization of interpretation. His emphasis on the 'radical finitude' of each performance must be understood against the background of Hegel and especially Heidegger. Hence, issues of polyvalency of meaning arise. In biblical studies this engages with work on the 'narrative-worlds' of parables and raises questions about a reappraisal of 'tradition' and 'effective-history' as against Enlightenment reason alone.

The unmasking by Habermas of the role of 'interest' in interpretation coheres well with attacks on those interpretations of the Bible that function primarily to serve the interests of oppressive or 'establishment' groups. Thus it adds to the hermeneutical resources of feminist and liberation hermeneutics (see Feminist theology; Liberation theology §3). At the same time, however, the insistence of Habermas that trans-contextual criteria of meaning or truth also have some place firmly separates him from the social pragmatism of Richard Rorty and Stanley Fish. If Rorty's approach is accepted, it invites the oppressed to wield the same kind of instrumental tools as the oppressors, in which case hermeneutics comes to depend on the force of contextual rhetoric, not trans-contextual truth. But if Christian theology is to embody any universal truth-claim, this approach becomes difficult to sustain. Habermas offers a middle ground on which a socially sensitive hermeneutic may operate with intellectual integrity.

Ricoeur has illuminated symbol, metaphor and narrative for biblical hermeneutics. In particular, he constructively stresses both the creative power of 'double-meaning' language in biblical texts and rightly urges suspicion concerning the interpreter's own desire to manipulate them. He has also called attention to the importance of hymnic, narrative and wisdom modes of biblical discourse alongside the prophetic, didactic and directive modes.

It is too soon to chart the ultimate effects of the work of Derrida and deconstructionism on biblical interpretation. Some have attempted to view his approach as an iconoclastic tool which coheres with a 'silent' or 'negative' articulation of divine transcendence. But Derrida's work offers more than a tool. He himself is drawn towards the 'unfinished' vision associated with apocalyptic writings. But interpreters of the Bible have used deconstructive strategies of reading effectively, to date, only on such untypical literature as Ecclesiastes, Job and possibly some parables of reversal which serve only to break the spell of some established view. The best example with regard to the Old Testament is D.J.A. Clines' work on Job: Job 'undermines the… hierarchical oppositions on which it relies' (1990: 65). The most influential examples of work on the New Testament come from J.D. Crossan, who argues that parables subvert the very 'worlds' which they project.

The emphasis on polyvalency in textuality and the role of narrative and of active readers has led to an unprecedented interest in literary theory in biblical hermeneutics. In some quarters, this has led to a partial abandonment of necessary concerns about biblical languages and biblical history. In others, a genuine attempt is made to hold together the best in mainstream methods of enquiry with newer insights which, in some cases, have yet to prove their value fully. At all events, the shift of perspective in biblical interpretation has been considerable over the last twenty years, and biblical hermeneutics has come to occupy a necessary place, after long neglect, in most research universities. Very recently, attention is being given to the ethics of hermeneutics (Patte 1995), the hermeneutics of the self (Ricoeur 1990; Thiselton 1995) and speech-act theory (Wolterstorff 1995), all against the background of the challenges of postmodern thought.

See also: Deconstruction; Hermeneutics; Midrash; Postmodern theology; Postmodernism; Theology, Rabbinic

ANTHONY C. THISELTON

References and further reading

Aquinas, T. (1266-73) *Summa theologiae* Ia, ed. H. McCabe, Oxford: Blackfriars Press, 1964, vol. 3: 47-107. (Q.13, aa.1-12 discuss the language of God.)

Bartsch, H.-W. (1962, 1964) *Kerygma and Myth*, London: Society for Promoting Christian Knowledge.(A sample

of debate about existential interpretation and demythologizing, including the classic essay by Bultmann on this subject.)

Bultmann, R. (1964-5) *Glauben und Verstehen (Faith and Understanding)*, 4 vols, Tübingen: Mohr; vol. 2 trans. as *Essays Philosophical and Theological*, London: SCM Press, 1955; vol. 1 trans. as *Faith and Understanding*, London: SCM Press, 1969.(A collection of classic essays; of particular interest is ‘The Problem of Hermeneutics’ in volume 2.)

Clines, D.J.A. (1990) ‘Deconstructing the Book of Job’, in M. Warner (ed.) *The Bible as Rhetoric*, London: Routledge, 65-80.(A sample of ‘postmodern’ hermeneutics.)

Derrida, J. (1967) *Of Grammatology*, Baltimore, MD: Johns Hopkins University Press, 1976.(A difficult text on hermeneutical theory.)

Dilthey, W. (1976) *Selected Writings*, ed. H.P. Rickman, Cambridge: Cambridge University Press.(This contains translations of parts of volumes 5 and 7 of the standard collected works, in which the ideas described in §3 are developed.)

Fuchs, E. (1964) ‘The New Testament and the Hermeneutical Problem’, in J.M. Robinson and J.B. Cobb (eds) *New Frontiers in Theology*, vol. 2, *The New Hermeneutic*, New York: Harper & Row, 111-45.(The easiest essay by this complex writer; the other essays in the volume offer a useful critique of his work.)

Fuchs, E. (1970) *Hermeneutik*, Tübingen: Mohr, 4th edn.(A classic, although difficult, exposition of the so-called new hermeneutic, which applies perspectives from the late Heidegger and Gadamer to New Testament interpretation.)

Gadamer, H.-G. (1960) *Truth and Method*, London: Sheed & Ward, 2nd English edn, 1989.(A complex but influential work.)

Gadamer, H.-G. (1966) ‘The Universality of the Hermeneutical Problem’, in *Philosophical Hermeneutics*, Berkeley, CA: University of California Press, 1975.(Attacks suppositions about the scope of ‘reason’ or ‘science’.)

Habermas, J. (1978) *Knowledge and Human Interests*, London: Heinemann, 2nd edn.(A classic exposition of his earlier thought prior to more sophisticated work on hermeneutics and language. The appendix to the second edition contains seminal ideas on ‘interest’, later developed more fully in *The Theory of Communicative Action*.)

Habermas, J. (1981) *The Theory of Communicative Action: The Critique of Functionalist Reason*, Cambridge: Polity Press, 1984.(Combines critical hermeneutical views with socio-linguistic theory and social philosophy.)

Jeanrond, W.G. (1991) *Theological Hermeneutics, Development and Significance*, London: Macmillan.(An intermediate-level introduction to the subject.)

Morgan, R. and Barton, J. (1988) *Biblical Interpretation*, Oxford: Oxford University Press.(A basic introduction to some general issues.)

Patte, D. (1995) *Ethics of Biblical Interpretation*, Louisville, KY: Westminster-Knox.(Discusses ethical dilemmas in first-person mode.)

Ricoeur, P. (1975) *The Rule of Metaphor*, London: Routledge & Kegan Paul, 1978. (Reflects the turn to double-meaning in sentences, with useful autobiographical reflectins at the end.)

Ricoeur, P. (1976) *Interpretation-Theory Discourse and the Surplus of Meaning*, Fort Worth, TX: Texas Christian University Press.(A succinct exposition of Ricoeur’s views in mid-development.)

Ricoeur, P. (1983-5) *Time and Narrative*, 3 vols, Chicago, IL: Chicago University Press, 1984-8.(Ricoeur’s *magnum opus* on the hermeneutics of narrative. A work of great stature.)

Ricoeur, P. (1990) *Oneself as Another*, Chicago, IL: Chicago University Press, 1992.(His crowning work on the hermeneutics of the self.)

Schleiermacher, F.D.E. (1977) *Hermeneutics: The Handwritten Manuscripts*, ed. H. Kimmerle, Missoula, MT: Scholars’ Press.(Taken from manuscripts written between about 1805 and 1831, this is the best source for Schleiermacher’s hermeneutical ideas.)

Schneiders, S.M. (1991) *The Revelatory Text: Interpreting the New Testament as Sacred Scripture*, San Francisco, CA: Harper.(Applies hermeneutical theory to some biblical material.)

Tate, W.R. (1991) *Biblical Interpretation: An Integrated Approach*, Peabody, MA: Hendrickson.(Elementary textbook covering only part of the ground.)

Thiselton, A.C. (1980, 1993) *The Two Horizons: New Testament Hermeneutics and Philosophical Description*, Carlisle: Paternoster, and Grand Rapids, MI: Eerdmans. (Detailed exposition and bibliography, with special studies of Heidegger, Bultmann, Gadamer and Wittgenstein.)

Thiselton, A.C. (1992) *New Horizons in Hermeneutics - The Theory and Practice of Transforming Biblical Reading*, London: HarperCollins, and Grand Rapids, MI: Zondervan.(A detailed critical discussion of the

subject, with a comprehensive bibliography; extensive chapters on Schleiermacher, Gadamer, literary theory and specific biblical texts.)

Thiselton, A.C. (1995) *Interpreting God and the Postmodern Self: On Meaning, Manipulation and Promise*, Edinburgh: T. & T. Clark.(This work addresses postmodern tendencies to translate truth-issues into issues of rhetoric and social power. It also includes a hermeneutic of selfhood on pages 47-79. In the final part (121-64) it offers a hermeneutic of biblical and theological approaches to hope and to Trinitarian theology.)

Watson, F. (1994) *Text, Church and World: Biblical Interpretation in Theological Perspective*, Edinburgh: T. & T. Clark.(A useful interactive discussion drawing together biblical studies, theology and postmodernity.)

Wolterstorff, N. (1995) *Divine Discourse*, Cambridge: Cambridge University Press. (A Yale philosopher of religion uses speech-act theory to evaluate the plausibility of the claim that God speaks through the Bible.)

Hermetism

A primarily religious amalgam of Greek philosophy with Egyptian and other Near Eastern elements, Hermetism takes its name from Hermes Trismegistus, 'thrice greatest Hermes', alias the Egyptian god Thoth. Numerous texts on philosophical theology and various occult sciences, ascribed to or associated with this primeval figure, were produced in Greek by Egyptians between roughly AD *100 and 300, and are a major document of late pagan piety. Reintroduced into Western Europe during the Renaissance, they provided considerable inspiration to philosophers, scientists and magicians of the fifteenth and sixteenth centuries.*

Hermetic literature can be divided into *philosophical* treatises, on God, the world and man, and *technical* writings on astrology, alchemy and other branches of occult science. The philosophical Hermetica comprise principally: (1) the *Asclepius* or *Perfect Discourse*, a longish work surviving in a Latin translation; (2) the *Corpus Hermeticum* proper, a Byzantine collection of fourteen treatises, translated into Latin by Marsilio Ficino in 1462-3 and published in 1471 under the title *Pimander* (after *Poemandres*, its first and most important treatise), to which three further pieces were later added; and (3) some twenty-nine extracts in the anthology compiled in the fifth century AD by John Stobaeus. The Stobaeus *Hermetica* vary in length from single sentences (12, 27, 28) to an important extract (23, from the *Korē Kosmou* or *Pupil of the Cosmos*) as long as anything in the *Corpus Hermeticum*.

The philosophical treatises take the form of dialogues, or rather, since disputation and argument are notably absent from them, of expositions, usually although not always by Hermes himself, to one or more trusting disciples. Their scenery and *dramatis personae* - Hermes, his son Tat, Asclepius (alias Imouthes or Imhotep), King Ammon and so forth - are Egyptian and ancient, investing these treatises, like so many other writings of the period, with the authority of primeval revelation. Their philosophy - that is, their cosmology and metaphysics - is a contemporary 'Middle Platonism' (see Platonism, Early and Middle), the only philosophical idiom available in late antiquity to anyone attempting a non-mythological treatment of these subjects. (There are also gnostic and Jewish elements, notably in *Poemandres* and *Corpus Hermeticum* III.) Their purpose, however, is not strictly philosophical. A treatise may start with some standard question of school philosophy - for example, motion (II), death (VIII), or intellection and sensation (IX). But the answer, often garbled, is seldom more than a starting point for meditation and homily. The aim is not to offer some new, coherent and discussible account of God, the world and man so much as to satisfy a religious need, common enough in this period, for a saving 'gnostic' illumination. The purpose of the Hermetist teacher - and the treatises tend to be stylized as lessons in a course of ever more esoteric instruction - is to generate a *gnōsis*, an intuitive knowledge of god and self, vouchsafed to very few, an answer in cosmic terms to the perennial question 'What am I here for? What am I?'. The instruction finds its fulfilment in intellectual illumination, as the pupil becomes aware of being a particle of divine life and light (*Poemandres* 21) and the teacher can say 'You have come to know yourself and our common Father' (*Corpus Hermeticum* XIII 22).

In this context, doctrinal consistency and lucid theory are minor considerations. There are numerous contradictions between the treatises - one text admits as much (*Corpus Hermeticum* XVI 1). Some of these go back to Plato's own works, to the contrast there between the *Timaeus*, with its picture of mankind placed in a good world by a good god - an optimism strongly endorsed by the *Asclepius* and by *Corpus Hermeticum* II, V-VI, VIII-XII, XIV, XVI - and the gloomier account of our human condition in the *Phaedo* and the *Phaedrus*, reflected in the severe pessimism of *Poemandres, Corpus Hermeticum* IV, VII, XIII and *Korē Kosmou*, when they dismiss the material world as a 'totality of evil' (*Corpus Hermeticum* IV 6), into which the soul has fallen as a punishment for original sin (*Korē Kosmou* 24), or in consequence of some primeval blunder (*Poemandres* 14). But the inconsistencies hardly matter. The Hermetic treatises are documents of spirituality, not philosophy. Scholarship has come increasingly to see them as translations, as products of a native Egyptian religious tradition (the very fact that they are attributed to Hermes-Thoth is confirmation of their author's religious loyalties) rewritten in the language of Middle Platonism.

Thoth was, among other things, the god of wisdom, knowledge and science. In Roman Egypt numerous works were ascribed to Hermes Trismegistus on technical subjects such as astrology, alchemy and the hidden properties of plants. His name is constantly invoked in magical papyri. These disciplines all rested on a principle, widely held in late antiquity and briefly sketched in the *Asclepius* (2-7, 19), of cosmic 'sympathy'. Linking things on earth to

each other and to things in heaven is a nexus of largely hidden sympathies and antipathies which can be used to explain, predict and manipulate the course of events. The philosophical Hermetica, where mentioning these occult sciences, give them a high religious colouring. Magic and philosophy alike, says the *Korē Kosmou* (68), nourish the soul. Both are ways to salvation.

Hermes was remembered as a magician, and also as a primeval sage, a younger contemporary of Moses, who foretold the coming of Christianity. During the Middle Ages, numerous works in Arabic and Latin were produced under his name. The arrival of *Corpus Hermeticum* in the West created something of a sensation: Ficino interrupted his life's work on Plato and Plotinus to translate it. A vastly older figure than Plato and a vastly purer exponent of the 'original theology' (*prisca theologia*), Hermes lent authority and respectability to the active interest which Ficino, Pico della Mirandola and others took in magic. The broad Hermetic vision of the world as a network of hidden forces waiting to by discovered and exploited by the magus was to be an inspiration to such luminaries of sixteenth-century science as Paracelsus, whose experiments in alchemy led to the discovery of laudanum, and Giordano Bruno, whose Hermetic interests ended with him burnt at the stake. The antiquity, and hence the authority, of Hermes Trismegistus received a fatal blow in 1614 when Isaac Casaubon demonstrated, on linguistic and other grounds, that the Hermetic writings could only be a late forgery. Hermes still had admirers and readers in the seventeenth century, including the Cambridge Platonists and even Isaac Newton. But Casaubon remained unrefuted; and the Hermetic writings lost their appeal to all save lovers of the occult and, in the twentieth century, historians of religion.

See also: Gnosticism; Mystical philosophy in Islam §2; Renaissance philosophy

JOHN PROCOPÉ

References and further reading

Copenhaver, B.P. (ed.) (1992) *Hermetica: The Greek 'Corpus Hermeticum' and the Latin 'Asclepius'*, Cambridge: Cambridge University Press.(English translation, with notes and a useful introduction.)

Festugière, A.-J. (1944-54) *La révélation d'Hermès Trismégiste*, Paris: Gabalda, 4 vols. (Indispensable in-depth study.)

Fowden, G. (1986) *The Egyptian Hermes*, Cambridge: Cambridge University Press. (Important for the historical and social background of the *Hermetica*.)

Grafton, A. (1983) 'Protestant Versus Prophet: Isaac Casaubon on Hermes Trismegistus', *Journal of the Warburg and Courtauld Institutes* 46: 78-93.(On Casaubon's seminal redating of the *Corpus Hermeticum*.)

Kingsley, P. (1993) 'Poimandres: The Etymology of the Name and the Origins of the Hermetica', *Journal of the Warburg and Courtauld Institutes* 56: 1-24.(Seeks to vindicate as authentically Egyptian the main content of the treatises.)

Nock, A.D. and Festugière, A.-J. (eds) (1945-54) *Hermès Trismégiste*, Budé series, Paris: Belles Lettres, 4 vols. (Text, French translation and notes.)

Yates, F.A. (1964) *Giordano Bruno and the Hermetic Tradition*, London: Routledge & Kegan Paul.(Important study of the Renaissance tradition.)

Herrera, Abraham Cohen de (*c*.1562-*c*.1635)

Herrera was a philosophically oriented Kabbalist who combined Neoplatonism and Kabbalistic knowledge learned from Israel Sarug, a disciple of Isaac Luria. In his Spanish works Puerta del Cielo (The Gate of Heaven), and Casa de la Divinidad (The House of Divinity), he considered Kabbalah as the source of the ancient truth.

Herrera believed that as the intellect is the highest epistemic grade, one must rely on philosophy to reach the divine secrets. He dealt with the concept of Ensoph (his own spelling of Eyn Sof - the Infinite) in terms of its Thomistic and Renaissance definitions - the Infinite that excludes limit. Tzimtzum is the voluntary Contraction of Ensoph, which prepares space for worlds to be created. The space opened up is metaphorically called Adam Kadmon (Primordial Man), the first emanation of which is the Intelligible World and the Tetragrammaton, the principle that shapes the Chain of Being around the Pythagorean number four. Neoplatonically, Adam Kadmon is the Intelligible World. Herrera's account incorporates further complex ideas about First and Second Causes of creation.

In 1699, Herrera drew the attention of Johan Georg Wachter, who, on the basis of Knorr von Rosenroth's Latin translations of Herrera's works in Kabbala Denudata, blamed him for inspiring Spinoza's supposed pantheistic heresy.

1 Life, writings and sources

Born to a *converso* family, probably in Italy, Abraham Cohen (Alonso Nuñez) de Herrera worked in the commercial service of the Sultan al-Mansur of Morocco. In 1596 he was taken prisoner by the Earl of Essex at Cadiz and brought to London, but was finally freed after extended diplomatic negotiations. He met Israel Sarug in Ragusa and delved into Lurianic Kabbalah (see Kabbalah). In Amsterdam he belonged to the synagogue which Spinoza attended in his youth, continued in business, and headed the committee that approved Joseph del Medigo's *Sefer Elim (Book of Deities)* for publication by Menasseh ben Israel. In the latter's own *Conciliador (Conciliator)*, the imprimatur calls him *Philosopho, Theologo y Cabalista* (philosopher, theologist and Kabbalist).

Between 1620 and 1632 Herrera wrote *Casa de la Divinidad* and *Puerta del Cielo*, to introduce other ex-Conversos to Lurianic mysticism. Isaac Aboab da Fonseca translated the works into Hebrew (1655), but lacked his master's philosophical background and failed fully to capture the letter and spirit of the Spanish originals, especially when pagan or Christian sources came into play. To these two works Herrera added two auxiliary booklets, *Epitome de la Logica o Dialectica (Summary of Logic or Dialectics)* and *Libro de Diffiniciones (Book of Definitions)*.

Herrera's library included the chief medieval philosophers, such as Aquinas and other scholastics, but he favoured Marsilio Ficino, Giovanni Pico della Mirandola and their Neoplatonic followers. Besides Maimonides, Albo and Leone Ebreo, he cites Kabbalist sources, including Hayyim Vital's unprinted tractates and sayings heard from Sarug. But he never cites such writings attributed to Sarug as *Limudei Atzilut (Studies of Emanation)* and does not ascribe to him the doctrines of the *malbush* (garment) or the *shi'ashua* (the delighting motion), which are central in his own thought. Indeed, he relies not on Sarug but on Moses Cordovero (1522-70) in dealing philosophically with Kabbalah.

2 Philosophy and Kabbalah

Herrera viewed Kabbalah, stemming from Adam or Abraham, as the fount of *prisca theologia* (ancient theology), standing above philosophy and even theology. He saw no conflict among the three, nor between Lurianic and Cordoveran Kabbalah, but relied on philosophy for aid in deciphering the enigmatic language of the Lurianic Kabbalah, just as Pico deciphered the gnomic symbolism of Plato.

Puerta del Cielo opens with reflections on the Kabbalistic Infinite, here equated with the First Cause of the scholastics. The Ensoph (Infinite) (so transcribed by Herrera himself) was a subject Lurianic Kabbalists avoided, loath to conceptualize God's transcendence. But Herrera gained support from scholastic works like Aquinas' *Summa contra gentiles* and *Quaestiones disputatae* and their Renaissance interpreters. Thus he adopts Aquinas' and Ficino's distinction between the privative infinity of matter, which longs for limit, and the negative infinity of

the First Cause, which excludes limit. Infinity verges towards limit in Tzimtzum, the Contraction of the Ensoph. Adapting Aristotle's principle (*Physics* VII) that every cause is present in its effect, the transcendence of the Ensoph is rendered immanent (see Aristotle). Following Pico's tripartite ontology, Herrera calls the First Cause 'eminent' or 'causal'; the Sephirot (as with Ensoph, his own transliteration of Sefirot - Numerations), or primal hypostases, 'formal' or 'proper'; and the Lower Worlds 'remote' or 'participative'.

Unlike some Kabbalists, Herrera does not imagine Tzimtzum in cathartic, mythical or eschatological terms but conceptualizes it as a voluntary act of Ensoph, preparing space for worlds to be created. What had been a convulsive event within the Divinity, is now a metaphor of the transition to finitude. The space opened up in the Ensoph is metaphorically called Adam Kadmon. This first creation by, from and within Ensoph is the ultimate substrate of potentiality and the first form or actuality.

Herrera treats the doctrine of Adam Kadmon as a central Lurianic theme and labours to find it in pre-Lurianic authorities - the *Sefer Yetzirah (Book of Creation)*, the *Zohar (Book of Splendour)*, Cordovero - arguing that the Keter Eyon (Supernal Crown) is, in fact, Luria's Adam Kadmon. He equates the distinction in *Tikkunei Zohar* between the 'Cause of All Causes' and the 'Cause of Causes' with that between Ensoph and Adam Kadmon. Two generations later this conclusion had dramatic consequences in the theology of the followers of Shabtai Tsvi (Sabbatai Sevi, as the name is sometimes transliterated), which sharply distinguished the First Cause from the Second and equated the latter with the God of Israel.

In a remarkable collation of medieval and renaissance sources, Herrera identifies the Adam Kadmon as a primal hypostasis with the Tetragrammaton, shaping the entire Chain of Being around the Pythagorean number four as the Ineffable name does. Neoplatonically, Adam Kadmon is the Intelligible World and thus the abode of the Sephirot. Responding to the problematics of emanation, Herrera presents the first perfect effect as the first link in an emanative chain, above which no higher perfect effect is possible.

In Cordovero, the Ensoph begins the emanative process by taking pleasure in God, either intellectually or somehow more concretely. But Herrera ascribes the doctrine of *shi'ashua* to Luria and interprets it metaphorically: the delighting motion of sexual congress within the Ensoph leads to a cosmic pregnancy and the birth of the finite from the infinite. Herrera interprets this imagery in terms of an intellectual process within the First Cause that gives definition to its future effects.

Herrera considers the Sephirot, or hypostatic numbers, as the cornerstone of the Kabbalah, which sets numerous hypostases between Ensoph and the Lower Worlds. Relying on the medieval scholastic Cardinal Pierre d'Ailly, he suggests that the First Cause acts the better through Secondary causes. The Sephirot are representations of Ensoph, the world of beauty surrounding the supreme good (see Plato's *Second Letter*) and the formal mode of being that provides our epistemic access to the First Cause. They are thus the *malbush* or garment of Ensoph, but they emanate from Adam Kadmon and lie below it, not at the top of the ontic hierarchy, like the *malbush* in Sarug. As a secondary cause, each Sefira is the supreme perfection of its own genus, heading a certain order of perfections, a supreme *henad*. This Proclean excursion leads Herrera to the audacious description of the Sephirot as secondary deities - a conception that influenced the Sabbatean subordination of the God of Israel.

Transposing the dynamism of Luria's divine lights into the language of ideas, Herrera interprets the Kabbalistic *Mot ha-Melachim* (Death of the Kings) and *Shevirat ha-Kelim* (Breaking of the Vessels) as metaphors for the Platonic fall of the soul into the body. Following Ficino's *Theologia platonica (Platonic Theology)*, he explains that when our soul inclines towards the body it forgets the divine intelligibles. *Tikkun* (restoration) is the return of divine intellects to their former perfection. Intellection, like biological generation, results from the union of the (male) intelligible with (female) intellectual receptivity. The Broken Vessels are restored not by additional light or power but by heightened intellectual receptivity.

In Kabbalistic emanation the soul is represented as a circle, a space produced by the self-contraction of the Infinite. The intellect is a straight line descending through that space. Discussing this idea with Sarug, Herrera had wondered why the circle, a curved line, takes precedence over a straight line. Sarug suggested that Luria's circle was in fact a single point, preceding the line as the principle of life. The line penetrating soul-space descends to produce more and more circles, conjoining soul and intellect, ultimately even in the lowest worlds (see Soul, Nature And immortality of the).

The psychological chapters of *Casa de la Divinidad* elaborate a doctrine of prophecy based on reconciling Lurianic with other Kabbalistic, scholastic and Renaissance ideas. Prophecy, Herrera concludes, is the highest grade of intellection granted by the divine will. Herrera adopts Aquinas' view and rejects that of Maimonides, which finds prophecy only in the most intellectually prepared. Here Herrera's intellectualism bends towards tradition.

3 Legacy

Herrera's conclusions were not wholly accepted, even by his few disciples. By the later seventeenth century his school had become quite unfamiliar with scholastic arguments, and those arguments themselves grew increasingly obsolete. Yet several prominent Kabbalists did adopt the conclusions of Herrera's metaphorical hermeneutic, rejecting in particular the notion that Tzimtzum was a real event within the divinity. Among these Kabbalists were the Sabbatean Miguel Cardoso and such anti-Sabbateans as Joseph Ergas and David Nieto. Herrera's strong opponents on this issue included the Sabbatean Nehemia Hayoun. Most of Herrera's adversaries disliked his reliance on philosophy, and even disciples like Aboab and Zakuto were unfamiliar with Herrera's more technical writings and confined themselves to quoting sayings from his doctrine of the soul. Yet the Hasidic thinker Baruch of Kosov is influenced by Herrera's findings, and the implicitly philosophical discourse of the Kabbalist Moshe Luzatto shows Herrera's distinctive influence.

The influence on Spinoza sometimes ascribed to Herrera is harder to detect. Many modern scholars assume, because of some terminological similarities, that Spinoza knew Herrera's writings, perhaps from his youth. Herrera's Sephirot, for example, each perfect in its own genus, have been linked with Spinoza's theory of God's attributes. Closer analysis shows that this and similar presumed parallels, for example, those regarding creation, reflect a common Judaeo-Spanish background and the influence of the scholastic and Renaissance literature that both men read but each used in his own way.

See also: Ficino, M.; Kabbalah; Platonism, Renaissance

NISSIM YOSHA

List of works

Herrera, A. Cohen de (*c*.1620-32) *Puerta del Cielo (The Gate of Heaven)*, Amsterdam: Jewish-Portuguese Seminary, Ets Haim Library, manuscripts EH 48 A 16 and EH 48 B 19; The Hague: Royal Library, manuscript 131 C 10; New York: Library of Columbia University, manuscript X 86-H 42 Q; Hebrew trans. I.A. da Fonseca, *Sha'ar ha-Shamayim*, Amsterdam, 1655; Latin trans. C. Knorr von Rosenroth, *Kabbala Denudata seu Doctrina Hebraeorum, Apparatus in Librum Sohar, Pars Tertia et Quarta, Quorum prior est Liber Sha'ar ha-Shamayim Seu Porta Coelorum (The Kabbalah Unveiled or the Doctrine of the Hebrews. A Work on the Book of Splendour the third and fourth parts being what was entitled before The Gate of Heaven)*, Sulzbach, 1678; German trans. F. Hausermann with introduction by G. Scholem, *Das Buch Sha'ar ha'Shmayim oder Pforte des Himmels*, Frankfurt am Main: Suhrkamp Verlag, 1974; ed. K. Krabbenhoft, *Puerta del Cielo* (text abridged from manuscript 131 C 10), Madrid: Fundación Universitaria Española, 1987.(Herrera's remarkable effort to translate Lurianic Theosophy - the mystical Kabbalah taught in the sixteenth century by Rabbi Issac ben Shlomo Luria (the ARI) - into philosophical and metaphorical terminology. The work deals mainly with the Divinity, namely the highest ontological hypostases of the Chain of Being and the Intelligible World. These grades can be partially approached with the help of our intellectual efforts.)

Herrera, A. Cohen de (*c*.1620-32) *Casa de la Divinidad (The House of Divinity)*, Amsterdam: Jewish-Portuguese Seminary, Ets Haim Library, manuscript H 48 A 20; Jerusalem: Jewish National and University Library, Varia 106; Hebrew trans. I.A. da Fonseca, *Beit Elohim*, Amsterdam, 1655; Latin trans. C. Knorr von Rosenroth, *Kabbala Denudata Partis Tertiae Tractatus 1 cui nomen Bet Elohim seu Domus Dei (The Kabbalah Unveiled. The name of the first tractate of the third part is The House of Divinity)* Frankfurt, 1684.(The second main work of Herrera deals with other aspects of Lurianic Kabbalah and their philosophical interpretations: angelology and the Upper Worlds under the Divinity; demonology and the Lower Worlds, including our material one; and the epistemic and psychological ways of Kabbalah and philosophy that can help us to reach the divine secrets. His analysis of prophecy favours Aquinas' conclusion that it is a gift of divine grace rather than an inwardly grounded intellectual gift.)

Herrera, A. Cohen de (*c*.1632-5) *Epitome de la Logica o Dialectica (Summary of Logic or Dialectics)*, published originally with *Libro de Diffiniciones* as one booklet, Amsterdam: Rosenthalania (library).(This auxiliary work

was written for those contemporaries of Herrera who had returned to Judaism after leaving the Iberian peninsula. Its aim is that of teaching them the rules of philosophy.)

Herrera, A. Cohen de (*c*.1632-5) *Libro de Diffiniciones (Book of Definitions)*, published originally with *Epitome de la Logica o Dialectica* as one booklet, Amsterdam: Rosenthalania (library). (As with the *Epitome*, this handbook was written following the Aristotelian method of preparing a philosophic *organan*, for those of Herrera's contemporaries who had returned to Judaism.)

References and further reading

Altmann, A. (1982) 'Lurianic Kabbalah in a Platonic Key: Abraham Cohen Herrera's *Puerta del cielo*', *Hebrew Union College Annual* 80: 1-38.(Excellent introduction to Herrera's thought regarding his Platonic and Neoplatonic sources.)

Krabbenhoft, K. (1982) 'The Mystic in Tradition: Abraham Cohen Herrera and Platonic Theology', unpublished Ph.D. dissertation, New York: New York University.(Deals with the Platonic and Neoplatonic influences on Herrera's work.)

Scholem, G. (1978) *Abraham Cohen Herrera ba'al Sha'ar ha-Shamayim - Hayav, Yetsirato ve-hashpa'ato (Abraham Cohen Herrera, Author of the Puerta del Cielo: His Life, Writings, and Influence)*, Jerusalem: Mossad Bialik Publishing House.(Spinoza was accused by eighteenth-centruy philosophers, mainly in Germany, of pantheistic heresy and some of them, Wachter being the first, accused him of following Herrera's Kabbalah into this heresy. Scholem describes brilliantly this polemical debate, which lasted almost a century.)

Saccaro Battisti, G. (1985) 'Herrera and Spinoza on Divine Attributes', *Italia* 4: 21-58.(Finds traces of Herrera's thought in Spinoza's philosophy.)

Saccaro Battisti, G. (1986) 'La cultura filosofica del Rinascimento italiano nella *Puerta del cielo* di Abraham Cohen Herrera', *Italia Judaica* 2: 295-334.(Analyses some of Herrera's Renaissance sources.)

Yosha, N. (1993) 'Abraham Cohen de Herrera: An Outstanding Exponent of *Prisca Theologia* in Early Seventeenth Century Amsterdam', *Dutch Jewish History* 3: 117-26.(Analyses the Renaissance *Prisca Theologia* theory in thought of Herrera, who considered Kabbalah to be the source of the common and eternal truth.)

Yosha, N. (1994) *Mytos u-metaforah - ha-parshanut ha-filisofit shel Abraham Cohen Herrera le-kabbalat ha-Ari (Myth and Metaphor: Abraham Cohen Herrera's Philosophical Interpretation of the Kabbalah of the ARI (R. Isaac ben Solomon Luria))*, Jerusalem: The Magness Press.(In this Hebrew work with an English summary, Yosha addresses Herrera's philosophic interpretation of Lurianic Kabbalah, referring in detail both to Herrera's Kabbalistic and philosophical sources and to its influence on later thinkers.)

Hertz, Heinrich Rudolf (1857-94)

Heinrich Hertz demonstrated the existence of radio waves in research between 1887 and 1888, opening the way for Marconi to develop long-distance radio communication. Hertz's results confirmed Maxwell's electromagnetic theory, and sealed the fate of action-at-a-distance in physics. His theoretical analysis included the famous dictum: 'Maxwell's theory is Maxwell's system of equations'. Hertz also developed a new formulation of Newtonian mechanics using the concepts of mass, length and time, but not force. He presented mechanics as the axiomatic consequence of a single fundamental law: 'every free system persists in its state of rest or of uniform motion in a straightest path'. Hertz's ideas influenced later philosophers of science but were most important as a source for Wittgenstein's Tractatus Logico-Philosophicus, which influenced logical positivism. Hertz's proposal to eliminate the concept of force in physics was an important contribution to the twentieth-century ideal of a philosophical method that does not solve, but rather dissolves, philosophical problems.

Heinrich Hertz began studies under H. von Helmholtz at Berlin in 1878. At that time there was no general agreement on the nature of electricity or its fundamental laws. For Wilhelm Weber electrical currents were counterflows of positive and negative electric particles, with forces between them depending on both their positions and their motions. F.E. Neumann used potential functions and avoided electrical particles. J.C. Maxwell also used potentials, but believed that motions in a universal ether underlay electricity and magnetism (see Electrodynamics). Helmholtz constructed a synthetic theory that led to all three alternatives as special cases. He believed, with Weber and Neumann against Maxwell, that forces acted instantaneously at a distance. He criticized Weber's theory on conceptual grounds - it violated conservation of energy - but failed to produce experimental results that would distinguish between the theories.

In 1879 Hertz showed that the 'inertia' of electricity was either nonexistent or impossibly small, providing the first strong experimental evidence against Weber's theory. A later series of experiments (1887-8) made all theories except Maxwell's untenable. Hertz showed that electrical oscillations in open circuits generated electromagnetic waves in the surrounding space which travelled at the speed of light, and could be refracted, reflected, polarized and diffracted. None of this could be reconciled with forces acting at a distance (Buchwald 1994). Development of Maxwell's theory led directly to Lorentz's classical electron theory (from 1892), which was superseded in turn by quantum mechanics and relativity, in the early twentieth century (Jungnickel and McCormmach 1986).

In the last years of his short life Hertz turned his attention to another theory that existed in several different versions: classical mechanics. As it was impossible to distinguish the versions experimentally, Hertz fell back on philosophical analysis, and it is to this work that he owes his main philosophical influence.

In the introduction to his *Principles of Mechanics* (1894) Hertz compared three 'representations': the traditional version of mechanics presented by Newton and refined by Lagrange; the contemporary Energeticist formulation; and Hertz's own. Each version differed mathematically, and employed different fundamental concepts. Newton and Lagrange employed the concepts of space, time, mass and force. The Energeticists replaced mass and force with the concept of energy. Hertz proposed to eliminate force but to introduce no new concepts. These different representations of mechanics were compared by three standards called by Hertz 'permissibility', 'correctness' and 'appropriateness'. In modern terms these requirements might be expressed as logical adequacy, empirical adequacy and perspicuity. Hertz maintained that the mechanics of Newton and Lagrange had yet to be presented in a form that avoided contradictions, and might therefore be impermissible. Energeticist mechanics was either incorrect or inappropriate. To be appropriate a representation should not exclude anything essential ('distinctness'), and should not include anything inessential ('simplicity').

To maximize 'appropriateness' it was important to eliminate 'empty relations' - concepts with no counterparts in the real world. Examples were the 'idle wheels' introduced by Maxwell in his 1862 theory, to allow cells of ether spinning in opposite directions to rotate freely. Equally 'permissible', 'correct' and 'distinct' representations that did not contain a concept showed it to be superfluous. In Hertz's analysis, 'force' was such a concept. This served Wittgenstein as a model for the elimination of signs from language that fail to mirror the world.

Wittgenstein's picture theory of meaning (see Wittgenstein, L. §3) extends Hertz's account of scientific theories as

representations to language as a whole. The critical method of the *Tractatus Logico-Philosophicus* follows Hertz in attempting to eliminate 'empty relations' and to define its domain of inquiry from within. Wittgenstein's 'objects', and the technique used to place awkward questions outside the domain so they need no answers, are also influenced by Hertz (Barker 1980; Hamilton 1994). The account of science in the *Tractatus* 6.3 covers many issues treated by Hertz in his analysis of mechanics (Barker 1979).

In the *Philosophical Investigations* Wittgenstein continues Hertz's programme of conceptual reform based on scrutiny of our mode of representation, developing methods of dissolving philosophical problems by bringing his audience to see that certain linguistic elements do not deserve their accustomed status. Throughout his mature work Wittgenstein refers to concepts that fail philosophical scrutiny as 'idle wheels'. Thus, Hertz not only provided one of the starting points for philosophy of science, in his account of theory choice, but, through the work of Wittgenstein, Hertz's call to eliminate force from physics contributed to twentieth-century philosophers' attempts to dissolve philosophical problems by eliminating conceptual confusions.

See also: Field theory, classical; Mechanics, classical

PETER BARKER

List of works

Hertz, H.R. (1892) *Untersuchungen über die Ausbreitung der Electrischen Kraft*, Leipzig: Barth; *Gesammelte Werke*, vol. 2, Leipzig: Barth, 1894.(See *Electric Waves*, translated by D.E. Jones, London: Macmillan, 1893. Reprinted in 1962, New York: Dover. The main papers on electromagnetic waves. The introduction is an autobiographical account of their discovery, followed by a searching analysis of Maxwell's theory.)

Hertz, H.R. (1894) *Die Principien der Mechanik*, Leipzig: Barth; *Gesammelte Werke*, vol. 3, Leipzig: Barth, 1894.(See *The Principles of Mechanics*, translated by D.E. Jones and J.T. Walley, London: Macmillan, 1899. Reprinted in 1956, New York: Dover. The introduction is Hertz's main contribution to philosophy.)

Hertz, H.R. (1895) *Schriften vermischten Inhalts*, Leipzig: Barth; *Gesammelte Werke*, vol. 1, Leipzig: Barth, 1896.(See *Miscellaneous Papers*, translated by D.E. Jones and G.A. Schott, London: Macmillan, 1896.)

Hertz, H.R. (1927) *Erinnerungen, briefe, tagbücher*, ed. J. Hertz, Leipzig: Akademische verlagsgesellschaft; repr. M. Hertz and C. Süsskind (eds) *Heinrich Hertz: Memoirs, Letters, Diaries*, San Francisco, CA: San Francisco Press, 1977, 2nd edn, enlarged. (Biographical material, much with scientific content, prepared by one of Hertz's daughters. The 1977 edition contains additional material and has a parallel text translation with German and English on facing pages.)

Hertz, H.G. and Doncel, M. (1995-6) 'Heinrich Hertz's Laboratory Notes of 1887', *Archive for History of Exact Sciences* 45: 149-270.(A facsimile, with transcription and translation, of newly discovered notes from a key period in Hertz's experiments on electromagnetic waves.)

Mulligan, J. (ed.) (1994) *Heinrich Rudolf Hertz (1857-94): A Collection of Articles and Addresses*, New York: Garland.(English versions of most of Hertz's important papers, including the introductions to *Electric Waves* and *Principles of Mechanics*, and remarks on Hertz by Helmholtz, Planck and others.)

References and further reading

With the exception of the first two, which are written for people already familiar with Wittgenstein's *Tractatus*, these items presuppose some knowledge of basic physics, but no special philosophical background.

Barker, P. (1979) 'Untangling the Net Metaphor', *Philosophy Research Archives* 5: 184-99.(Relates Wittgenstein's account of science, *Tractatus* 6.3, to Hertz's discussion of mechanics.)

Barker, P. (1980) 'Hertz and Wittgenstein', *Studies in History and Philosophy of Science* 11: 243-56.(Establishes connections between Hertz's *Principles of Mechanics* and Wittgenstein's *Tractatus*.)

Buchwald, J.Z. (1994) *The Creation of Scientific Effects: Heinrich Hertz and Electric Waves*, Chicago, IL: University of Chicago Press.(An important historical study of Hertz.)

Hamilton, K. (1994) *Hertz and Wittgenstein: The Philosophical Significance of Wittgenstein's Scientific Training*, Ph.D. dissertation, University of Notre Dame, IN. (Presents new historical information about Wittgenstein in a full-length study of his debt to Hertz.)

Jungnickel, C. and McCormmach, R. (1986) *Intellectual Mastery of Nature: Theoretical Physics from Ohm to Einstein*, vol. 1, *The Torch of Mathematics, 1800-1870*, vol. 2, *The Now Mighty Theoretical Physics, 1870-1925*, Chicago, IL: University of Chicago Press.(Presents the theoretical, experimental and institutional development of German physics. Volume 1 ends with Helmholtz's early researches. Volume 2 covers the whole career of Hertz, and the development of physics up to the founding of quantum mechanics and relativity

theory.)

McCormmach, R. (1983) *Night Thoughts of a Classical Physicist*, New York: Avon Books.(A recreation of the world of physics at the time of Hertz, in the form of a short novel.)

Hervaeus Natalis (d. 1323)

The French scholastic philosopher and theologian Hervaeus Natalis was not only one of the most influential early Thomists, but was also an original thinker who made an important contribution to the medieval debate on intentionality. He examined carefully the ontological question of what intentional objects are, and discussed the epistemological problem of how they are generated in a cognitive act. Hervaeus argued that intentional objects are 'third entities' that cannot be reduced to extramental or mental entities, a thesis that sparked controversy in the fourteenth century.

Hervaeus Natalis (Hervé Nédellec, or Nédélec) was born in Brittany in the middle of the thirteenth century and entered the Dominican Order in 1276. He commented on the *Sentences* of Peter Lombard at the University of Paris in 1301-2, obtained the degree of Master in Theology in 1307 and acted as *magister regens* until 1309. From 1309 to 1318 he headed the French province of the Dominican Order, and was elected Master General of the Order in 1318. He died in 1323, shortly before the canonization of Thomas Aquinas.

Hervaeus eagerly defended and developed Aquinas' theology and philosophy. In his *Defensa doctrinae D. Thomae (Defence of Saint Thomas' Teaching)* he expounded Aquinas' theological methodology, and in his *Quodlibeta* he defended the Thomistic epistemology against the objections of Henry of Ghent. Yet he did not slavishly follow Aquinas. In his philosophy of mind, he distinguished himself as an original thinker who exerted a considerable influence on fourteenth-century debates, both on the Continent and in England.

Hervaeus discussed extensively the question of the nature of intentional objects, that is, the objects of directed cognitive acts (see Intentionality). In his *Tractatus de secundis intentionibus (Treatise on Second Intentions)*, the first medieval treatise dealing exclusively with the problem of intentionality, he claimed that these objects are not the extramental things that have real existence, because even if one directs a cognitive act towards a nonexistent object (for example, a chimaera) or towards a not actually existing object (for example, the dead Caesar), the act is directed toward something. Nor are these objects 'intelligible species', that is, representational entities produced by the intellect and existing in the intellect; these species are only the means by which the intellect has access to the intentional objects, but not the objects themselves. Rather, intentional objects are peculiar things with 'objective existence': they have existence only *qua* objects of the intellect and cannot be reduced to the species of things having 'subjective existence', that is, the kind of existence a quality has in a subject. Thus, although several people cognizing a table have their own individual species, which subjectively exist in their individual intellects, all these people are directed toward one and the same intentional table.

Hervaeus emphasized that every intentional object of a cognitive act has objective existence. All that varies is the foundation of such an object. If one is having a cognition of an existing table, the intentional table is immediately founded upon the real table. If, however, one is having a cognition of a chimaera, the intentional chimaera is founded upon the combination of three kinds of real animals. Of course, the cognition of a nonexistent thing is more complex than that of an existing thing because it involves the combination of various real things and the formation of several intelligible species; but the important point is that in cognizing a nonexistent thing, one does have an immediate intentional object.

Hervaeus held that in addition to particular intentional objects (so-called 'first intentions') there are also universal intentional objects. Having cognized a particular thing, such as a human being, one is also able to cognize the universals 'human being' (species) and 'animal' (genus). These universals are abstracted on the basis of particular intentional objects and function as higher-level intentional objects (so-called 'second intentions'). Such an abstraction is possible because universals are instantiated in particular things. This claim makes it clear that Hervaeus' theory of second intentions is closely linked to a realist theory of universals (see Universals).

Hervaeus' introduction of intentional objects with a peculiar ontological status sparked an intense discussion among some of his fourteenth-century contemporaries and successors, including Peter Aureol, William of Ockham, Walter Chatton and Adam Wodeham. Ockham, in his *Ordinatio*, adduced two main arguments against Hervaeus' theory of intentionality. First, the introduction of intentional objects threatens direct realism: if one only has immediate cognition of intentional objects, one never has direct access to extramental reality. Second, if

intentional objects are really distinct from the extramental things, entities will be multiplied without necessity; for each thing with real existence there will be a corresponding thing with objective existence. This doubling of the world is superfluous, Ockham claimed, because the cognitive act is by itself directed toward the extramental thing, without there being any need for an intermediary entity.

Yet, in evaluating Ockham's famous criticism, one should take into account the fact that Hervaeus did not aim to refute direct realism or to posit superfluous entities. Rather, he intended to point out that the thing *qua* cognized thing is not simply the extramental thing, which in itself has no linkage with the intellect, but something founded upon the extramental thing that somehow hooks the intellect up to this thing and determines the cognitive act. Thus, his theory of intentional objects should be understood as an attempt to understand what the content of a mental act is - an attempt that had an impact not only on late medieval discussions, but also on Brentano's philosophy of mind (see Brentano, F.C.).

DOMINIK PERLER

List of works

Hervaeus Natalis (before 1323) *Tractatus de secundis intentionibus*, Venice, 1508.(More or less reliable text; there is no critical edition and no English translation of this treatise.)

Hervaeus Natalis (before 1323) *Defensa doctrinae D. Thomae*, in E. Krebs, *Theologie und Wissenschaft nach der Lehre der Hochscholastik an der Hand der bisher ungedruckten Defensa Doctrinae D. Thomae des Hervaeus Natalis mit Beifügung gedruckter und ungedruckter Paralleltexte*, Beiträge zur Geschichte der Philosophie des Mittelalters 11, Münster: Aschendorff, 1912.(The edition of the *Defensa* is included along with an analysis of Hervaeus' theory of science.)

Hervaeus Natalis (before 1323) *De intellectu et specie (On the Intellect and the Intelligible Species)*, ed. P. Stella, 'La prima critica di Hervaeus Natalis O.P. alla noetica di Enrico di Gand', *Salesianum* 21 (1959): 125-70. (Edition of *De intellectu et specie* with commentary.)

Hervaeus Natalis (before 1323) *Quodlibeta (Quodlibetal Questions)*, Venice, 1513. (No modern editions or translations exist at present.)

References and further reading

Allen, E.B. (1960) 'Hervaeus Natalis. An Early 'Thomist' on the Notion of Being', *Mediaeval Studies* 22: 1-14. (Critical analysis of a key term in Hervaeus' metaphysics.)

Hedwig, K. (1978) 'Der scholastische Kontext des Intentionalen bei Brentano' (The Scholastic Context of the Intentional in Brentano), *Grazer Philosophische Studien* 5: 67-82.(Study dealing with the possible sources of Brentano's thesis that psychic phenomena have an 'intentional inexistence'.)

Käppeli, T. (1970) *Scriptores Ordinis Pradicatorum Medii Aevi (Writers of the Dominican Order in the Middle Ages)*, Rome: S. Sabina, vol. 2, 231-.(A comprehensive list of Hervaeus' works, including all the manuscripts and Renaissance editions.)

Kelley, F.E. (1978) 'Some Observations on the *Fictum* Theory in Ockham and its Relation to Hervaeus Natalis', *Franciscan Studies* 38: 260-82.(Detailed analysis of Hervaeus' objective existence' theory and Ockham's critique.)

Roensch, F.J. (1964) *Early Thomistic School*, Dubuque, IA: The Priory Press.(Short presentation of Hervaeus' life and intellectual activity.)

Schöllgen, W. (1927) *Das Problem der Willensfreiheit bei Heinrich von Gent und Herveus Natalis (The Problem of Freedom of the Will in Henry of Ghent and Hervaeus Natalis)*, Dusseldorf: Pädagogischer Verlag Schwann; repr. 1975.(Introduction to Hervaeus' philosophical psychology and its impact on the problem of freedom.)

Senko, W. (1961) 'Les opinions d'Hervé Nédellec au sujet de l'essence et l'existence' (Hervaeus Natalis' Opinions on Essence and Existence), *Mediaevalia Philosophica Polonorum* 10: 59-74.(Presentation of a fundamental metaphysical issue and comparison with Thomas Aquinas' treatment of this problem.)

William of Ockham (1317-19) *Ordinatio*, Book I, in G. Gál *et al*. (eds) *Opera Theologica*, St Bonaventure, NY: The Franciscan Institute, vol. IV.(Dist. 27 includes a critique of the thesis that intentional objects are a kind of 'third entities'.)

Herzen, Aleksandr Ivanovich (1812-70)

Lauded by Nietzsche as 'a man of every distinctive talent' and admired by Lenin as the founder of the Russian revolutionary movement, Herzen eludes all neat categorizations. As a moral preacher he stands alongside Tolstoi and Dostoevskii (who praised him as a poet). As a philosopher, he was the principal interpreter and popularizer of Hegel's thought in Russia in the first half of the 1840s, while the rebellion against metaphysical systems in his mature work has led him to be seen as a precursor of existentialism. Through the Russian press that he founded in emigration he helped to shape the beginnings of a public opinion in his country and played a major role in debates on Russia's political future on the eve of the emancipation of the serfs, while laying the foundations of the Russian populist movement through his writings on Russian socialism. He is best known in the West for his memoirs, Byloe i dumy (My Past and Thoughts) (1861, 1866), which rank among the great works of Russian literature, and for S togo berega (From the Other Shore) (1850), the most brilliant and original of the works in which he expresses his rejection of all teleological conceptions of history.

1 Life and works

Aleksandr Herzen owed his surname to an affair of the heart between a rich and cultivated Russian landowner, Ivan Iakovlev, and the daughter of a minor German official whom he met when travelling in Europe. Although his parents never married, Aleksandr enjoyed the status of a pampered only son. He received a rounded education in European culture from foreign tutors, and was especially drawn to Rousseau, Voltaire, Schiller and the history of the French Revolution. His illegitimacy undoubtedly contributed to his precocious sense of independence and spirit of protest against the existing order. He identified with the heroes of the Decembrist Revolt of 1825 and (an event celebrated in the history of the Russian revolutionary movement) made a vow with his cousin Nikolai Ogarev at the age of 16 to devote their lives to the struggle against despotism in their country. Like other Russians of his generation, he channelled his frustrated urge for political action into philosophical speculation and was much attracted to the historiosophical schemata of German Idealism. But in an essay written as a student of natural sciences at the University of Moscow he points out that the 'one-sidedness' of pure speculation had been demonstrated by Francis Bacon: only by linking 'the rational method with the empirical' (a combination which Herzen later, in an echo of Feuerbach, calls 'speculative empiricism') would mankind finally achieve an approach to knowledge that would reconcile the claims of the particular and the universal, the individual and the social whole. Such a reconciliation was to be the main concern of Herzen's later thought. It drew him to doctrines of palingenesis and to the humanistic socialism of the Saint-Simonians, whose theories he adopted with Ogarev and their circle in 1833: an event that has been seen as marking the birth of socialism in Russia.

The following year the circle was arrested on suspicion of spreading subversive ideas. Herzen spent nearly a year in prison, followed by five in provincial exile, when (for the last time in his life) he plunged into romantic introspection and religious mysticism. In 1840 he was allowed to return to Moscow, where Hegel's influence was at its peak. He embarked on a systematic study of Hegel's philosophy (during which he was exiled for another year), which resulted in two long cycles of essays: 'Diletantizm v nauke' ('Dilettantism in Science) (1842-3) and 'Pis'ma ob izuchenii prirody' ('Letters on the Study of Nature') (1845). The most important critique of Hegel to appear in Russia in that decade, Herzen's essays were much influenced by the left Hegelian 'philosophy of action' and the anthropocentrism of Feuerbach, which had greatly excited him. In Hegel's vision of dialectical becoming he saw a justification of his own hopes of a socialist transformation. As he put it in the Aesopian language of his essays, in Hegel's philosophy abstract thought had reached its highest achievement by liberating men's minds from the transcendent authorities of the past. The next stage was liberation in life, the move from philosophy to action (see Hegelianism, Russian).

Herzen's vision of the path that action would take was shaped in particular by Proudhon's anarchism, which promised the kind of reconciliation of individual freedom with social harmony that he sought. Unable to discuss political theories in print in Russia, he directed his efforts to fostering the inner moral independence which he saw as the precondition of political liberty, through essays that criticized the human tendency to seek moral direction from external authorities and norms. After 1845 he turned from philosophy to fiction, exploring the problem of moral freedom in short stories and a semi-autobiographical novel *Kto vinovat? (Who is to Blame?)* dealing with a love triangle. In 1847, finding his talents stifled in his country, he emigrated with his family to the West.

His historical optimism was shattered by the European upheavals of 1848. Observation of political events in France from 1847 until 1850 convinced him that parliamentary democracy was a palliative that did not address fundamental issues of social injustice, and he accused both liberals and radical republicans of a worship of political forms combined with indifference to the real needs and aspirations of the masses. In *From the Other Shore*, his philosophical summing-up of the lessons of 1848, he attacked the utopianism of the European left as both empirically unfounded and morally pernicious in subordinating the self-fulfilment of living individuals to the empty abstraction of progress.

With the disappointment of his own utopian hopes of imminent social transformation in the West, Herzen turned his attention to the socialist potential of Russia. The theoretical foundations of Russian populism were laid between 1849 and 1852 in a series of essays (such as 'Ruskii narod i sotsializm' ('The Russian People and Socialism'), addressed to Jules Michelet) in which Herzen sought to acquaint progressive Europeans with the existence of a radical intelligentsia in Russia and the potential of the peasant commune. (His own interest in the commune had been aroused by discussions with the Slavophiles in the early 1840s.) He pointed out that its democratic structure, guaranteeing the right of all its members to the use of a share of its land, contained the seeds of a solution to the problems of pauperization and social atomization currently facing the West.

In 1851-2 Herzen's personal life was shattered by a series of tragedies beginning with the revelation of an affair between his wife and an intimate family friend, the German radical poet Georg Herwegh. After she ended the affair, Herwegh circulated his version of events among the European radical community with the intention of discrediting Herzen, who vainly sought to be publicly vindicated through a revolutionary 'court of honour'. In the midst of these betrayals and humiliations his mother and younger son were drowned at sea. His wife, her health destroyed by the succession of calamities, died a few months later.

The effect of these tragedies was to intensify Herzen's sense of alienation from the West and turn his full attention back to Russia. He began his memoirs, *Byloe i dumy (My Past and Thoughts)*, 'to get into communication with my people', and in 1853 embarked on a new phase of his life with the founding of a Russian Press in London to act as his people's 'uncensored voice' in exerting pressure on the Russian government for reform. In 1856 Ogarev arrived from Russia to join him, and became co-editor of his celebrated periodical *Kolokol (The Bell)*.

Herzen emphasized that his press was not to be the vehicle for a particular political programme or ultimate ideal (his memoirs record that his personal tragedy had destroyed the last remnants of his early utopianism and forced him to accept the dominant role of chance in human affairs). While not renouncing his hopes of a future 'Russian socialism' he saw his political role as mobilizing opinion within Russia to exert pressure for immediate concrete reforms (in particular, emancipation of the serfs, freedom of speech and the end of corporal punishment), and providing a forum for discussion about further steps, in the hope of reaching a consensus on the social form most capable of combining individual liberty with social cohesion. What such a form would be, he stressed, depended on contingent circumstances and could not be predicted in advance.

While *Kolokol* played a major role in shaping public opinion in Russia on the eve of the Great Reforms, after the Emancipation Herzen's open-mindedness on the paths of progress brought him into sharp conflict both with the Russian liberals (who wished Russia to follow a Western path to constitutional democracy), and the radicals who preached an immediate transition to socialism through revolution. Herzen's preference for evolution over revolutionary violence alienated him from the left, while his refusal to exclude revolution as a means of progress was condemned by the liberals.

In a number of essays (notably 'Robert Owen', and those found in 'Kontsy i nachala' ('Ends and Beginnings') and 'K staromu tovarishchu' ('To an Old Comrade')), Herzen developed his view of the open-endedness and unpredictability of the historical process against what he saw as the deterministic utopianism of most Russian progressive thought.

Herzen's complex vision of the historical process met with little sympathy or understanding among the Russian intelligentsia. In 1863 he moved to Geneva in order to have closer contact with the growing population of Russian radical émigrés, but his relations with them soon soured. *Kolokol* ceased publication in 1867 (although it was briefly revived in a French edition), and Herzen's political isolation became complete. His last years were embittered by a stormy relationship with Ogarev's wife Natalie who bore him three children. 'Probably everyone

in Russia will say that Herzen should have died sooner, that he outlived his reputation', Turgenev wrote on his death: but from a historical perspective the isolation of Herzen's last years can be attributed to the fact that his vision was far in advance of his time.

2 Philosophy of history

'Do not look for solutions in this book', Herzen warns the reader of *From the Other Shore*: 'in general, modern man has no solutions…. We do not build, we destroy; we do not proclaim a new revelation, we eliminate the old lie' ([1850] 1956: 3). To a critic who asked him to explain his system, he retorted that his philosophy of history was no system, only 'a scourge to be used against absurd theories…a fermenting agent, no more'. It was, however, systematically consistent in its attack on all systems, dogmas and political eschatologies that claimed to justify or explain the oppression of human beings by reference to idealized abstractions such as freedom, progress, humanity or the common good.

Like most radicals of his time, Herzen was deeply indebted to the anthropocentrism of the Left Hegelians, from whose perspective he approached Hegel's dialectic as 'an algebra of revolution'. But (like Proudhon, who influenced his thought in this respect), he rejected the secular eschatology which substituted Universal Reason for Divine Providence as the guiding force of the historical process, accusing both Hegel and his radical interpreters of evading the logical consequence of his historicism: the denial of all suprahistorical absolutes and universal historical goals.

In his major professions of philosophical faith (*From the Other Shore*, 'Robert Owen', 'Ends and Beginnings'), Herzen defines what Nietzsche would call the 'craving for metaphysical comfort' as the principal obstacle to the fulfilment of human potential in the historical world. He points out that observation and experience do not support the view that history is an ascent towards a final goal, whether pre-programmed by the Creator or inherent in the rational structure of the universe. 'Life has its own embryogenesis which does not coincide with the dialectic of pure reason' ([1850] 1956: 31). It was more like nature than was commonly believed: chance and 'dark forces' constantly triumphed over rational intention, killing off promising beginnings or unpredictably rerouting them into new channels. No ideal could ever be realized in the way in which it had been conceived, no model of human society could be more than a provisional experiment whose life expectancy would depend largely on chance. Herzen attributes the reluctance of mankind to face this truth to the dualism pervading human judgments. Anticipating Nietzsche's critique of otherworldly values that degrade the world of experience, he notes that all the great moral systems have been built on the assumption that process, chance and change are less real than timelessness and stasis, singular events less significant than universal laws and norms. All our images, all our metaphors, are imbued with a dualism that 'sides with one shadow against another, granting spirit the monopoly over matter, species the monopoly over the particular, sacrificing man to the state, the state to humanity' ([1850] 1956: 136). The Enlightenment had replaced Christianity with 'the *Religion des Diesseits*, the religion of science, or universal, hereditary, transcendental Reason'. We seek (against the empirical evidence) to impose a libretto on history; yet if there were such a libretto, we would be mere wheels in a machine. By denying us the consolations of self-transcendence, history maximizes the possibilities of self-creation. Instead of seeking to escape our contingency we should cease to allow preoccupation with an unknown future to distract us from self-fulfilment in the present; but to do so we must revise all our notions about the relation of the individual to himself, society and the state.

Herzen believed that the traditional vocabulary of philosophy was inimical to the perception of reality that was born with Hegel's historicism. He stresses that it is not just a question of revising our categories of ultimate goals and purposes; we must cease to derive the significance of the individual and the contingent from its place in some larger scheme of things. To the '*memento mori*' of metaphysics he opposes the '*memento vivere*' of nature and art. 'Thought must take on flesh': we must counter our tendency to rationalize and universalize by cultivating an aesthetic receptivity to the transient and unrepeatable manifestations of life which will allow us maximally to exploit the multiplicity of possibilities for self-creation.

Herzen's writings on this theme seem to anticipate Nietzsche's notion of *amor fati*: the demand that we face up to the 'beautiful chaos' of contingent existence and joyfully accept the historicity of the human condition with its attendant suffering, finitude and mortality, by shaping random events into the aesthetic unity of a unique self. Herzen's famous claim that 'the truly free man *creates* his own morality' has been likened to Nietzsche's assertion

that 'nothing is true, all is permitted': there is no reality independent of our shifting perspectives on it. But these parallels are misleading. Herzen did not live to engage with Nietzsche's irrationalism, but he unambiguously rejected the pessimism of Nietzsche's mentor, Schopenhauer, arguing that the vision of the world as a chaotic battleground of blind forces was a petulant overreaction to the failure of unjustified idealist expectations, a muddled transitional stage on the way to understanding reality without metaphysical preconceptions. As a corrective both to comforting illusion and exaggerated pessimism Herzen recommends to the student of history a training in the natural sciences that will inculcate 'true nihilism', which he defines as 'science without dogmas…the unconditional submission to experience and the resigned acceptance of all consequences, whatever they may be, if they follow from observation, or are required by reason' ([1861, 1866] 1968: 1764).

Herzen intended his reflections on history to serve as an object-lesson in this approach, spelling out what he regarded as the inescapable, but by no means wholly negative, consequences of the observable parallels between historical development and the chanciness of evolutionary processes. Against the pessimist he argues that experience reveals that chance operates within a framework of laws which allow us to calculate probabilities, to direct and control events within narrow boundaries, and even to progress, although not to utopia. When the anarchist (and former Left Hegelian) Bakunin pronounced the will to destroy as the creative force that would lead mankind to 'absolute liberty', Herzen accused him of a Romantic's disregard for the 'physiology' of historical and social development, the painful slowness of processes of growth in real time (see Bakunin, M. §2).

The shock and disbelief that greeted *On the Origin of Species* (1859), a decade after Herzen published *From the Other Shore*, testifies to the novelty and daring of Herzen's ideas on the dominant role of chance in nature and human life. In the furore over Darwin's opus, Herzen was among the first to grasp clearly the importance of the challenge that it posed to teleological thinking in all branches of intellectual activity, recommending it to Russian readers as an antidote to the extreme conservatism of most contemporary European thought.

Herzen's historical method owed much to the inspiration of an earlier innovator: Francis Bacon. In his philosophical writings of the 1840s he had stressed the significance of Bacon's analysis of those tendencies of the human mind that had to be corrected if knowledge were to be advanced, in particular the 'Idols of the Tribe' - the tendency to impose on external reality an order and a regularity that comes from the human intellect and does not belong to the world itself (see Bacon, F. §5). In his writings on history Herzen frequently cites Bacon's aphorisms on method, his demand that metaphysics be separated from science and his insistence on the barrenness of research into final causes, the '*magnum ignotum*' beyond the realm of knowledge. In 'Sur le libre arbitre' ('On Free Will'), his philosophical *prise de position* (in the form of a letter to his son) on the question of free will, it is Bacon ('an honest thinker') whom he cites in support of his attack on the dogmatically reductionist materialism of mid-nineteenth-century biological scientists. By declaring our sense of moral freedom to be an illusion, he argues, they had exceeded their competence: it was at the very least, an 'anthropological reality', without which we could not function as social beings. The dominance of dualistic systems had ensured that we possessed neither the words nor the categories to discuss the nature of choice in the nexus of contingency and invariant laws. But the purpose could be served by aesthetic metaphor: Herzen expresses the relationship between physical determinism and moral freedom through the distinction between a sound in isolation (a physical phenomenon subject to the laws of acoustics) and a sound in a musical phrase, where it 'acquires for us another value (or existence, if you like)'. This existence does not exempt it from the physical laws to which it is subject - the string may break and the sound will disappear, 'but as long as the string remains unbroken, the sound belongs not just to the realm of vibrations, but also to the realm of harmony where it exists as an aesthetic reality, functioning in a symphony which allows it to resound, dominates and absorbs it, and then abandons it in its wake'.

This unfinished symphony is the historical process, for which Herzen uses a related image in a fine passage in 'Robert Owen':

> having neither programme, set theme nor unavoidable *denouement*, the dishevelled improvisation of history is ready to walk with anyone; anyone can insert into it *his* line of verse and, if it is sonorous, it will remain his line until the poem is torn up, so long as the past ferments in its blood and memory.
>
> ([1861, 1866] 1968: 1245)

This notion of moral freedom bears little resemblance to Nietzsche's concept of the lonely Superman, asserting his will in defiance of history and human societies, which he identifies with the life-denying values of the herd.

Herzen believed that as social animals, human beings could achieve freedom only through a constant adjustment of their demand for self-fulfilment with their need for social solidarity - a model of freedom that has close parallels with Schiller's treatise *On the Aesthetic Education of Man* (1795). Like Schiller, Herzen believed that one should seek to govern one's conduct not by universal rational precepts but by the cultivation of an aesthetic sense of relationships that would allow one to respond to the demands of specific situations without either rigidly adhering to principle or blindly obeying the compulsion of instinct (see Schiller, J.C.F.).

Herzen pronounced Schiller's treatise 'a prophetic work…far in advance of its time'. One of its central themes - the 'fear of freedom' that drives individuals to seek to regulate their conduct by reference to transcendent authorities and universal norms - is the leitmotif of Herzen's writings on the nature of history. It is 'frightening' to abandon religious and secular concepts of predestination, to accept the ominous truth that it is absurd to apply our 'petty household rules', our human demands for intentionality and purpose, to the economy of the universe. The demystification of the world was a deeply painful process: but only those who had passed through that furnace, he maintained, could regard themselves as 'truly modern'.

3 Russian socialism

Herzen's interest in the structure and potential of the Russian peasant commune has been variously, but in general negatively, interpreted by historians. His belief that the commune contained the seeds of an advanced form of socialism has been represented as a messianic faith in flat contradiction with his critique of utopian thought, while his concern in the 1850s and 1860s with such matters as the methods of repartitioning land in the existing village commune has been described as a descent from the sublime to the ridiculous, from the great philosophical issues of *From the Other Shore* to the petty minutiae of peasant life.

Both types of charge were made against Herzen in his lifetime: the first by liberals who interpreted his faith in the commune's potential as a form of Slavophile messianic nationalism, and the second by socialists who demanded of him an inspiring vision of an ideal future. Herzen's response to both groups was to defend his view of the commune as wholly consistent with his view of freedom and the historical process.

Against the liberals he argued that, unlike the Slavophiles, he approached the commune not as a sign of Russia's superior destiny, but as a chance historical survival which might (although only if developed with the aid of Western technology and political theories) offer a solution to the most urgent problems facing both Russia and Europe: how to combine individual liberty with social cohesion, and to achieve economic development without mass pauperisation. While the commune might provide a way for Russia to pass to socialism without the trauma of mass industrialization, there was no guarantee that it would: socialism might be achieved both in Russia and the West by some quite different path, or not at all. He contrasts his 'prosaic' approach, focused on how best to balance individual aspirations and social needs at a given historical juncture, with the doctrinairism of the Russian liberals (who held that Russia's historical destiny was to discard the commune and follow the West along a common path to parliamentary democracy), and of those radicals who lived in expectation of a revolutionary leap to an ideal freedom. In the contempt of many progressive intellectuals for the primitive outlook of the Russian peasant and the imperfections of the existing commune he saw an instance of the teleological thinking that had been the principal target of his philosophical essays. He accused his critics of devaluing all aspects of contemporary experience that did not lead to their chosen goal, and respecting history and human liberty 'only in the future'. In 'K staromu tovarishchu', sometimes described as his political testament, he declares that his aim is to understand the size of the human pace in the present and, by substituting 'tact and inspired improvisation' for doctrinal consistency, 'to keep up with it, not falling behind, but not going so far ahead that it will not follow'.

4 Herzen's legacy

Herzen frequently remarked that his writings consisted of the constant repetition of a few simple truths, but that these truths were so repugnant to his contemporaries that they persisted in misconstruing them.

His message - that there were no ultimate solutions to the fundamental problems of human existence - has proven equally repugnant to Russian intellectuals in this century, leading to interpretations of his work that he would have repudiated. His ideas did not find favour among representatives of the Russian religious and philosophical renaissance of the early 1900s such as S.N. Bulgakov, who argued that his rejection of a transcendent world order

had led him ultimately (and inevitably) to a despairing nihilism. This view is echoed in V.V. Zenkovsky's authoritative history of Russian philosophy, which describes Herzen's thought as a philosophy of 'despair, hopelessness and disbelief'. Exceptions to this consensus were P.B. Struve (see Signposts Movement), who saw in Herzen 'one of our national heroes of the spirit' and the precursor of a new humanism, and Tolstoi, who noted in his diary that Herzen 'awaits his readers in the future. He imparts his thoughts far above the heads of the present crowd to those who will be able to understand him'.

The understanding of Herzen in his own country was not advanced by Lenin's canonization of him as the founder of the Russian revolutionary movement, an interpretation that was faithfully echoed by generations of Soviet historians. In the West his reputation has been slow to grow. As Isaiah Berlin argued in essays that brought Herzen to the attention of a wide English-speaking public, a condescending classification of his views 'as yet another variant of early socialism' left out the attack on utopian systems that was his most arresting contribution to political theory. At the end of the twentieth century we are better placed than Herzen's contemporaries to grasp the significance of that contribution. Herzen was among the first to predict a process with whose consequences we are only beginning to come to terms: the erosion of that faith in a purposeful universe in which all the basic assumptions of the great optimistic systems of his time (including Marxism) were grounded. He was more than just a precursor: his work remains relevant to contemporary debates in philosophy and cultural theory on the effect that the collapse of the 'grand narratives' of progress has had on traditional views of society, morality and the self. Herzen offers a clear alternative to two of the most common philosophical approaches to an inherently contingent world: Schopenhauerian pessimism and Nietzschean aesthetic immoralism. He has been compared to contemporary hermeneutic philosophers who argue that the historicity of the self does not render moral generalizations meaningless, and detect in history the unifying thread of a spiritual quest: the search to articulate our intuitive sense of the good. To those who felt themselves cast adrift in an unprogrammed world, he responded with the invigorating message that not only are freedom and moral responsibility possible in such a world: they are not possible in any other.

See also: Liberalism, Russian; Slavophilism

AILEEN KELLY

List of works

Herzen, A.I. (1954-65) *Sobranie sochinenii (Collected Works)*, Moscow: AN SSSR, 30 vols.(The most complete edition to date, with copious notes and commentaries.)

Herzen, A.I. (1845) *Kto vinovat?*, trans. M.R. Katz, *Who is to Blame?*, Ithaca, NY and London: Cornell University Press, 1984.(Herzen's most important fictional work.)

Herzen, A.I. (1850) *S togo berega*, Paris; trans. M. Budberg, *From the Other Shore*, with an introduction by I. Berlin (ed.), London: Weidenfeld & Nicolson, 1956.

Herzen, A.I. (1852) 'Russkii narod i sotsializm', trans. R. Wollheim, *The Russian People and Socialism*, with an introduction by I. Berlin (ed.), London: Weidenfeld & Nicolson, 1956.(One of Herzen's most important texts.)

Herzen, A.I. (1858) *Letters from France and Italy, 1847-1851*, trans. and ed. J.E. Zimmerman, Pittsburgh, PA: University of Pittsburgh Press, 1995.(Extended reflections on contemporary Europe, a source of much of Herzen's later thought.)

Herzen, A.I. (1861, 1866) *Byloe i dumy*, vols 1 and 2, London; vols 3 and 4 France; first complete publication Berlin: Slovo, 1921, 5 vols; trans. C. Garnett, *My Past and Thoughts. The Memoirs of Alexander Herzen*, revised by H. Higgens, with an introduction by I. Berlin, London: Chatto & Windus, 1968, 4 vols.(Written between 1852 and 1867. Fourth volume includes important essays.)

References and further reading

Acton, E. (1979) *Alexander Herzen and the Role of the Intellectual Revolutionary*, Cambridge: Cambridge University Press.(Focuses on the years 1847-63.)

Berlin, I. (1978) *Russian Thinkers*, London: The Hogarth Press. (Contains two seminal essays on Herzen's view of liberty.)

Carr, E.H. (1968) *The Romantic Exiles. A Nineteenth-Century Portrait Gallery*, Harmondsworth: Penguin.(A study of Herzen, Ogarev and their families in exile.)

Kelly, A. (forthcoming) *Towards Another Shore: Russian Thinkers between Necessity and Chance*, New Haven, CT and London: Yale University Press.(Includes five essays on Herzen as a precursor of developments in

twentieth-century thought.)
Labry, R. (1928a) *Alexandre Ivanovic Herzen, 1812-1870: Essai sur la formation et le développement de ses indées (Aleksandr Ivanovic Herzen, 1812-1870: Essay on the Formation and Development of his Ideas)*, Paris: Bossard.(A good general study.)
Labry, R. (1928b) *Herzen et Proudhon,* Paris: Bossard.(On the most important foreign influence on Herzen's ideas.)
Lampert, E. (1957) *Studies in Rebellion*, London: Routledge.(Includes a chapter on Herzen.)
Malia, M. (1961) *Alexander Herzen and the Birth of Russian Socialism, 1812-1855*, Cambridge, MA: Harvard University Press.(A detailed study of Herzen's formative period.)
Shpet, G. (1921) *Filosofskoe mirovozzrenie Gertsena (Herzen's Philosophical Outlook)*, Petrograd: Kolos.(An exceptionally perceptive study of Herzen's philosophy, which the author describes as 'historical realism'.)

Heschel, Abraham Joshua (1907-72)

Born in Warsaw and educated there and in Berlin, Abraham Joshua Heschel moved to the USA in 1940, where he lived and taught for the rest of his life. His elegantly written books and essays and his striking personality made him a key figure in American Jewish philosophical theology after the Second World War. Written in German, English, Hebrew and Yiddish, his books reflect widely on the Hebrew Bible, Talmud and midrash, and on Jewish mystical writings, continuously engaging with contemporary philosophy and theology.

Born in Warsaw, the scion of a major Hasidic dynasty, Abraham Joshua Heschel was educated thoroughly in Bible, rabbinics, Kabbalah and Hasidism in his youth. After a brief period in Vilnius (Vilna, as it was known in Jewish letters), he studied philosophy at the University of Berlin and taught at the Hochschule für die Wissenschaft des Judentums in Berlin. He moved to the USA in 1940 and first taught at Hebrew Union College (the Reform rabbinical seminary) in Cincinnati and after 1945 at Jewish Theological Seminary (the Conservative rabbinical seminary) in New York for the rest of his life.

In his Berlin doctoral dissertation on prophetic consciousness, *Die Prophetie (Prophecy)* (1937), Heschel creatively used the methods of phenomenology, especially as applied to religious concerns by Max Scheler, to develop the concept of divine pathos that was to be a leitmotif of his subsequent thought. Where many medieval Jewish thinkers had adopted Hellenic ideas of God, most prominently those of Aristotle that assume divine eternity and impassivity, Heschel found the chief emphasis of the Hebrew Bible in God's concern for humanity (see Bible, Hebrew). Timelessness and perfection seemed to preclude any real divine concern with what is essentially temporal and imperfect. But in the God of the Hebrew Scriptures Heschel saw a pathos that was of the essence of prophetic consciousness and its inner sympathy with God's relation to the world.

Heschel's first major English work, *The Sabbath* (1951a), applied his phenomenological method at the core of Jewish piety, pointing out the primacy of time consciousness over space consciousness in Judaism. Heschel showed how the Sabbath addresses sacred time rather than sacred space. Out of step with the more pragmatic and worldly approach of many writers on religion, especially at the time of his writing, Heschel did not seek the value of the Sabbath in its worldly utility. Rather than appeal to secular values by describing the benefits of the Sabbath as a humane institution or a vehicle of socio-cultural continuity, Heschel introduced his readers to the spiritual needs that the Sabbath can address. Pressing the point, he went on to argue that the Sabbath is the only true antidote to the dehumanizing effects of unlimited technology.

In his next English book, *Man Is Not Alone* (1951b), Heschel sought to show the necessity of taking revelation seriously (see Revelation). The opening to this concern was what he called the sense of radical amazement: that is, the human realization that existence itself, our own and that of the world, cannot simply be taken for granted. Only when one sees that the world does not supply the answer to our most basic existential question 'Why is there something and not nothing?' can one be ready for the answer to that question that comes from the Source of the world, who radically transcends its being and all the categories devised to explicate its immanent processes. In thinking along these lines, Heschel shows especially the influence of his long philosophical friendship with Martin Buber.

Heschel's next major book, *God In Search of Man* (1955), contains the most significant statement of his thought and shows Heschel as the theologian he basically was. *Man Is Not Alone* was subtitled A *Philosophy of Revelation*, but the subtitle of *God In Search of Man* is *A Philosophy of Judaism*. He calls his method here 'depth theology', by which he means confrontation of the most basic human questions that religion addresses rather than the presentation of a dogmatic system. Following the fundamental assumption of phenomenology that all thought is about something other than itself, that thought is reflective not self-creative, Heschel presents the key doctrines and norms of classical Judaism as Israel's way of orienting itself towards God. Thus, in discussing the relations between Jewish law (*halakhah*) and reflection (*aggadah*), he argued, too many traditionalists see the law as an end in itself and adopt what Heschel called 'religious behaviourism'. They fail to realize that the religious obligations are part of the human response to the revelation of the transcendent God. Too many liberals, by contrast, see the Law as something arbitrary, to be supplanted in large part by inner religious attitudes. In a striking variation of Kant's famous dictum, Heschel writes: 'Halakha without aggada is dead, aggada without halakha is wild' (see

Halakhah). Throughout his career, Heschel portrayed Judaism as a complex of transcendent intent with immanent content. The categories he developed in *God In Search of Man* have lent themselves to a wide range of applications by later Jewish thinkers.

During his lifetime, Heschel was often described and sometimes dismissed as a mystic. If this means someone who claims to have unusual personal experience of the divine, then calling Heschel a mystic is empty and misleading. But Heschel certainly was influenced by the esoteric Jewish tradition of Kabbalah (see Kabbalah). He incorporated in his mature outlook much of the Kabbalah he learned in his youth, especially the Hasidic versions of it that stressed the human experience of the intimate presence of God, more than the esoteric theosophy of the earlier Kabbalah (see Hasidism). Although in his adult life he was no longer a Hasid in the strict sense of the word, Heschel was always deeply beholden to his spiritual roots. Indeed, his last book was the closest he came to returning to that world. It was a study, in Yiddish, his first spoken and literary language, of the thought of the outspoken nineteenth-century Hasidic master Rabbi Menachem Mendel of Kotsk.

One fundamental Kabbalistic idea that played a key role in Heschel's thought is that of 'divine needs': God not only wills that there be a world, he actually desires it and wants an ongoing relationship with it, as is manifest in the history of the people of Israel. The ultimate source of this divine need is a mystery beyond human ken, but it is the basis of God's reaching out to create. Thus humanity's response to God (through keeping the commandments of the Torah) answers not just to human needs but to those of God. Heschel's recourse to this classic Kabbalistic theme (see Nahmanides) allows him to offer insightful and spiritually moving explanations of many Jewish practices that some liberals found merely bizarre and that many traditionalists simply accept with a dulling kind of positivism.

See also: Bible, Hebrew; Hasidism; Kabbalah

DAVID NOVAK

List of works

Heschel, A.J. (1975) *Between God and Man: An Interpretation of Judaism*, ed. F. Rothschild, New York: Free Press.(An anthology of some of Heschel's more important writings along with a general introduction and shorter introductions to the sections of the anthology.)

Heschel, A.J. (1937) *Die Prophetie (Prophecy)*, Warsaw: Polish Academy of Sciences.(Heschel's doctoral dissertation, a phenomenology of prophetic consciousness.)

Heschel, A.J. (1951a) *The Sabbath*, New York: Abelard-Schuman.(Theological reflections on the spiritual significance of this cornerstone of Jewish religious life.)

Heschel, A.J. (1951b) *Man Is Not Alone*, Philadelphia, PA: Jewish Publication Society.(Lays the theological groundwork for his more philosophical works.)

Heschel, A.J. (1954) *Man's Quest for God*, New York: Charles Scribner's Sons.(Heschel's chief reflection on the Jewish liturgy.)

Heschel, A.J. (1955) *God in Search of Man*, New York: Farrar, Straus & Cudahy.(Heschel's major work in philosophical theology.)

Heschel, A.J. (1962) *The Prophets*, Philadelphia, PA: Jewish Publication Society.(An expanded version of *Die Prophetie* with more extensive illustrations from biblical texts but not developing philosophically the core phenomenological project of the earlier work.)

Heschel, A.J. (1965) *Who Is Man?*, Stanford, CA: Stanford University Press.(Lectures on philosophical anthropology.)

Heschel, A.J. (1966) *The Insecurity of Freedom*, Philadelphia, PA: Jewish Publication Society.(A collection of essays on social, political and religious issues, reflecting his social activism.)

References and further reading

Merkle, J.C. (1985) *The Genesis of Faith: The Depth Theology of Abraham Joshua Heschel*, New York: Macmillan.(The most extensive analysis of Heschel's philosophical theology.)

Merkle, J.C. (ed.) (1986) *Abraham Joshua Heschel: Exploring his Life and Thought*, New York: Macmillan.(A collection of essays by Jewish, Catholic and Protestant scholars.)

Hesiod (*c.*700 BC)

One of the earliest surviving Greek poets, Hesiod was a direct precursor of the first philosophers. He composed one poem on mythic cosmogony and cosmology (Theogony) and another on work and justice in human life (Works and Days).

Little is known of Hesiod’s life, beyond what he tells us: his father emigrated from Asia Minor to Ascra in Boeotia and left him a share of an inheritance over which he quarrelled with his brother Perses, he won first prize in a poetic contest and the Muses themselves initiated him into poetry while he was tending sheep. The *Theogony* and *Works and Days*, which survive intact, are undoubtedly by him. Numerous other works were attributed to him in antiquity.

At least in part because Aristotle systematically distinguishes Hesiod as one of the earlier, mythical, philosophically unserious ’theologians’ from the later, more rational and philosophically interesting thinkers he calls ’physicists,’ most accounts of Greek philosophy tend to separate Hesiod from the Presocratics, assigning Hesiod’s poetry to the primitive phase of mythic thought and locating in Thales the beginning of genuine philosophical reflection in Europe. Yet Hesiod’s questions - What are the origin and structure of things? How can humans achieve success and happiness in their lives? - are the same ones that concern all later philosophers; and his answers, despite their often mythical form, remain of considerable interest.

In the *Theogony*, Hesiod attempts to unify many divergent local versions of Greek mythology by systematizing them all in terms of a coherent account of the origin and legitimacy of the current moral structure of the universe, identified with the rule of Zeus. Conceptual relations like implication, entailment or association are expressed mythologically as procreation or consanguinity, opposition or contradiction as warfare. Although Zeus’ rule was founded upon violence and crime in a succession of divine generations, it is now characterized by justice and order and is a welcome improvement upon earlier conditions. Human beings are a small but not quite negligible part of an extraordinarily complex and by now fairly stable world in which all natural and moral phenomena, including those viewed in a negative light, have a fixed place; cosmic forces of potential disorder remain, but they are at least temporarily under control and the threat of primitive anarchy they represent is remote.

Human existence becomes the central theme in Hesiod’s later *Works and Days*. Once again he uses myths to explain the origins of its permanent features: because of Prometheus labour is necessary; because of Pandora many evils plague human life; the sequence of gold, silver, bronze, heroic, and iron races demonstrates the relation between work, responsibility, and success. Now he adds numerous direct precepts and admonitions concerning social behaviour, religious practice and agriculture, and analyses the concept of *eris* (contention) to show that it can be not only baneful, when it leads to strife and war, but also beneficial, when it provides the impetus towards economic competition. It is Zeus’ rule of justice that distinguishes us from the animals and that ensures that, even though we are obliged to labour incessantly and to worry about an uncertain future, we can console ourselves with the reasonable hope that we not only will survive, but may even flourish, if we choose to face this obligation responsibly rather than try to prosper by immoral means.

See also: Presocratic philosophy

GLENN W. MOST

List of works

Hesiod (*c.*700 BC) *Theogony*, in M.L. West (ed.) *Hesiod. Theogony*, Oxford: Clarendon Press, 1966; H.G. Evelyn-White (ed. and trans.) *Hesiod, The Homeric Hymns and Homerica*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1914, 78-155.(The Evelyn-White edition provides Greek text with an English prose translation.)

Hesiod (*c.*700 BC) *Works and Days*, in M.L. West (ed.) *Hesiod. Theogony*, Oxford: Clarendon Press, 1978; H.G. Evelyn-White (ed. and trans.) *Hesiod, The Homeric Hymns and Homerica*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1914, 2-65.(The Evelyn-White edition provides Greek text with an English prose translation.)

Hesiod (*c.* 700 BC) Fragments, in R. Merkelbach and M.L. West (eds) *Fragmenta Hesiodea*, Oxford: Clarendon

Press, 1967.(A scholarly edition, without translation, of the Greek texts deriving Hesiod’s other works than the *Theogony* and the *Works and Days* and surviving in citations or on papyri.)

References and further reading

Cornford, F.M. (1952) *Principium Sapientiae. The Origins of Greek Philosophical Thought*, ed. W.K.C. Guthrie, Cambridge: Cambridge University Press.(Provides in part 2 pages 159-256 a very stimulating although uneven analysis of the relation between philosophical cosmogony and myth and ritual.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(The discussion of Hesiod at pages 34-46 concentrates on the *Theogony*, with attention to Near Eastern parallels.)

Most, G.W. (1997) ‘Hesiod’s Myth of the Five (or Three or Four Races)’, *Proceedings of the Cambridge Philological Society*.(Interprets the structure of Hesiod’s myth of the races in terms of his view on time, responsibility, and the conditions and chances for human success in the world in which we live.)

Rowe, C.J. (1983) ‘"Archaic" Thought in Hesiod’, *Journal of Hellenic Studies* 103: 124-35. (Argues against the idea of a linear evolution from mythic to philosophical thought.)

Solmsen, F. (1949) *Hesiod and Aeschylus*, Ithaca, NY: Cornell University Press.(Provides a useful general account at pages 3-100, focusing on Hesiod’s relation to his predecessors.)

Stokes, M.C. (1962-3) ’Hesiodic and Milesian Cosmogonies’, *Phronesis* 7: 1-37, and 8: 1-34.(A careful comparison between mythical and philosophical cosmogonies.)

Hess, Moses (1812-75)

Hess was a socialist philosopher, closely connected with the Young Hegelians, who influenced the initial philosophical development of Karl Marx, and later articulated, in the context of a critique of European bourgeois society, one of the first calls for the re-establishment of a Jewish commonwealth in Palestine.

Hess was born to a Jewish family in Bonn, Germany. At an early age he repudiated the religious teachings of his parental home and under the influence of Spinozist, Young Hegelian and Feuerbachian ideas developed a highly original, though philosophically eclectic, socialist philosophy (see Feuerbach, L.A.; Hegelianism §2).

His first book, *Die heilige Geschichte der Menschheit (The Holy History of Mankind)* (1837) proposes a three-stage philosophy of history (God the Father, God the Son and the Holy Spirit) based on the changing relationship between politics and ethics, to culminate in a new, socialist ‘New Jerusalem’. This new order, based on the abolition of private property, will reintroduce into politics those ethical elements banished from human life due to the Christian dichotomy between the material and the spiritual which enabled the emergence of individualist capitalism. In a further publication, *Die Europäische Triarchie (The European Triarchy)* (1841) he sees the cultures of France, Germany and England symbolizing, respectively, radical political, philosophical and industrial innovation, as the foundation of a new progressive alliance against the conservative post-1815 reality in Europe.

Active in the radical intellectual movement which included alsoMarx and Engels,Hess went into exile, to Belgium and France. In numerous essays, such as ’Die Philosophie der Tat’ (’The Philosophy of the Deed’) (1843), *Über das Geldwesen (On Money), Über die sozialistische Bewegung in Deutschland (On the Socialist Movement in Germany)* (1845) and ’Die letzten Philosophen’ (’The Recent Philosophers’) (1845), he continued to develop his brand of ‘ethical’ socialism, slowly differentiating his approach from the ‘materialistic’ approach of Marx and his followers.

After a short return to Germany during the 1848-9 revolution, Hess spent the rest of his life in Paris. There, under the impact of the rise of national movements in Germany and Italy in the wake of the failure of 1848, he also developed his own thoughts on the future of the Jews and Jewish civilization. In his earliest writings he expressed the view that the socialist revolution will also integrate Jewish life into the new revolutionary Europe; but Hess always recognized the salience of national identity in the course of world history. The rise of nationalism in Europe led him to the conclusion that only the renewal of Jewish political life in the ancestral land of Israel could guarantee both an adequate social solution to the Jewish problem and the survival of Jewish culture. Greatly influenced byMazzini’s ideas of a universalist, harmonious nationalism, Hess developed these ideas in his book *Rom und Jerusalem (Rome and Jerusalem)* (1862). The socialist structure of this Jewish commonwealth would be based both on universalist principles and on the strong socially-orientated precepts whichHess, despite his religious agnosticism, found underlying much of the historical Judaic biblical legislation. His book - rather idiosyncratic and hardly noticed at the time of publication - was later to become one of the major intellectual texts of Zionism, especially in its socialist variation.

In his later years Hess tried to develop, on Spinozist lines, a modern philosophy of science with radical, socialist-activist implications; but the work was never finished. He died in Paris on 6 April 1875 and was later to become the only thinker revered by both the socialist and Zionist movements.

Because of his philosophical eclecticism, Hess never produced a systematic body of work comparable to that of other socialist thinkers, such as Saint-Simon or Marx. Yet his interpretation of Spinoza’s monism and pantheism can be seen as the mainstay of both his socialist and proto-Zionist ideas (see Spinoza, B. de). Similarly, because of this lack of a systematic doctrine, Hess was more receptive to a pluralist interpretation of history, which gave a role to issues of national identity and cultural tradition, generally overlooked by more systematic thinkers like Marx.

The unity of matter and spirit which Hess discerns in Spinoza serves him as the foundation of his ethical socialism: economic life cannot be divorced from ethical considerations, nor can ethics be relegated to a separate compartment (‘religion’), purely focused on an other-worldly redemption (whichHess saw as the main characteristic of Christianity). That the historical Jewish legislation institutionalized this unity of politics and

ethics in the concept of a 'holy nation' Hess saw (when shorn of its ethnocentric connotations) as the specifically socialist element embedded in Judaism, in contrast to the dichotomies of Christianity, which have led to competitive individualism, alienation and capitalism.

Many of Hess' ideas on alienation in modern industrial society, on money as 'reified' power dominating the lives of human beings who created it, found their way, via Marx, into the general canon of socialist critique. But because of his eclecticism Hess was free from the reductionist flaws of Marxian thought. Similarly, his recognition of both the autonomy and intrinsic value of culture as embedded in specific national memory and identity saved him from the abstractedness and lack of understanding for national problems which characterized much of socialist thought. Echoes of Hess' thought can be found among the so-called 'Austro-Marxist' thinkers, in the Yugoslav attempt, under Tito, to combine socialism with a multi-ethnic reality, and in socialist Zionism.

See also: Alienation; Socialism; Zionism §1

SHLOMO AVINERI

List of works

Hess, M. (1961) *Philosophische und sozialistische Schriften (Philosophical and Socialist Writings)*, ed. A. Cornu and W. Mönke, Berlin: Akademie Verlag.(Includes 'The Holy History of Mankind', 'The European Triarchy' and all of Hess' socialist writings until 1850.)

(1843) 'Die Philosophie der Tat', trans. 'The Philosophy of the Deed', in A. Fried and R. Saunders (eds) *Socialist Thought*, Garden City, NY: Doubleday, 1964, 249-75.(The major essay in which Hess develops his 'philosophy of praxis' as a critique of classical German Idealist philosophy, which he depicts as quietistic, passive and hence politically reactionary.)

(1845) 'Die letzen Philosophen', trans. 'The Recent Philosophers', in L. Stepelevich (ed.) *The Young Hegelians*, Cambridge: Cambridge University Press, 1982, 355-75.(A critique of idealist philosophy as a secular version of traditional Christian duality between 'spirit' and 'matter' which is, according to Hess, at the root of the development of European capitalist society.)

(1862) *Rom und Jerusalem*, selections trans. as *Rome and Jerusalem*, inA. Hertzberg (ed.) *The Zionist Idea*, New York: Atheneum, 1969, 119-39.(Hess' major proto-Zionist work, in which he criticizes the failure of Jewish emancipation in Europe and calls for the establishment of a Jewish commonwealth in Palestine, to be founded on socialist precepts drawn both from the international socialist movement as well as for the socially oriented content of traditional Judaic law as seen by him.)

(1962) *Ausgewählte Schriften (Selected Writings)*, ed. H. Lademacher, Köln: Melzer Verlag.(Includes a selection of his early writings, and the full text of *Rome and Jerusalem*, with all its notes, appendices, etc.)

References and further reading

Avineri, S. (1985) *Moses Hess: Prophet of Communism and Social Zionism*, New York: New York University Press.(A detailed study of the development ofHess' thought, based on his writings throughout his life. Contains extensive quotations from the writings and a bibliography.)

Berlin, I. (1980) 'The Life and Opinions of Moses Hess', in *Against the Current*, New York: Viking.(The best concise résumé of Hess' work in relation to that of Marx).

McLellan, D. (1969) *The Young Hegelians*, London: Macmillan, 137-59.(PlacesHess in the context of his contemporaries among the Young Hegelians.)

Weiss, J. (1960) *Moses Hess - Utopian Socialist*, Detroit, MI: Wayne State University Press.(A comprehensive study of Hess' critique of capitalist society and his vision of an ethically based socialism.)

Hessen, Sergei Iosifovich (1887-1950)

Sergei Hessen, a disciple of Rickert, has been described as 'the most brilliant and philosophically gifted' representative of Neo-Kantian transcendentalism in Russia on the eve of the Revolution. A co-editor of the St Petersburg-based philosophical journal Logos, he represented a distinctively Westernizing trend, critical in many respects of the openly metaphysical religious idealism of the 'religious-philosophical renaissance' in Russia. However, he readily acknowledged the existence of an ineradicable metaphysical need in human beings and stressed the metaphysical relevance of his philosophy of moral and cultural values. B.V. Iakovenko attributed to him a pronounced 'ontological aspiration' and Hessen's philosophical evolution led him beyond 'pure anthropologism', towards a religiously tinged Platonism.

Like many Russian philosophers, Hessen concentrated on philosophical problems which had direct relevance for practice. His main fields were philosophical pedagogy, which he defined as 'applied philosophy', and political philosophy, with particular emphasis on philosophy of law. Like many other Russian philosophers, he was forced to emigrate, moving from one country to another, which did not help his professional career. Nevertheless, his work Osnovy pedagogiki (Foundations of Pedagogy) (published in Berlin, 1923) was translated into many languages and won him international recognition as the main author of the 'pedagogy of culture'. His original conception of 'rule-of-law socialism' (which deserves to be regarded as the last link in the legal philosophy of Russian liberalism) was much less known, and the manuscripts of his two books on political philosophy perished in the Warsaw uprising of 1944. Still, the UNESCO Committee on the Philosophic Basis of Human Rights invited him to contribute to the preparation of the Universal Declaration of Human Rights of 1948. His article for UNESCO, together with his writings on political philosophy published posthumously in Italy, were a Russian contribution to the philosophy of human rights in modern, liberal democratic states.

1 Philosophy of values

Sergei Iosifovich Hessen was a son of a Russian lawyer who became the editor of the St Petersburg journal *Pravo (Law)* and a prominent member of the Kadet Party. Owing to this Hessen grew up surrounded by the best representatives of Russian liberalism and the legal profession (P. Miliukov, L. Petrażycki and others). He studied philosophy in Germany under the direction of Heinrich Rickert (see Neo-Kantianism §§6-8). His doctoral thesis, *Individuelle Kausalität (Individual Causality)*, was defended *summa cum laude* in 1908 and published in Kantstudien. Having returned to Russia, he edited *Logos* (together with F. Stepun) and lectured at St Petersburg University (see Neo-Kantianism, Russian §6). In 1917 he was a personal secretary of Plekhanov. After the Revolution he served for four years as a professor of philosophy in Tomsk but the increasing ideological pressure made it impossible for him to continue his work in Russia. He emigrated to Berlin and then to Prague, where he lectured in the Russian Pedagogical Institute. In 1935 he moved to Poland, on the invitation of the Free University in Warsaw. During the war he took an active part in the underground university courses in Warsaw. After the war he resumed his professional duties in the newly created University of Łódź.

The intellectual situation of the early twentieth century was defined by Hessen as a double crisis: the crisis of philosophy and the crisis of liberalism, both caused by the undermining of the belief in the objectivity of truth and in the objectivity of values. He wanted to overcome this philosophical and political crisis without returning to naïve scientific and moral dogmatism, and without resuscitating the rationalistic and class-bound illusions of classical liberalism.

In his efforts to do this Hessen made use of a peculiar variant of dialectical method. He attributed its discovery to Plato. In his interpretation dialectic was a means of combating relativism and saving values from destruction: it enabled him to detect continuity and permanence in historical change, to see in the realm of the relative a reflection of the absolute. 'Dialectical comprehension' was to him a combination of the historical approach with phenomenological 'eidetic insight', and the ontological foundation of dialectics was, in his view, the Neoplatonic idea of 'All-Unity' - an idea which, thanks to Solov'ëv, became strongly embedded in Russian religious philosophy.

Under the influence of Roman Jakobson's Linguistic Circle in Prague (see Structuralism in linguistics §2), Hessen turned his attention to structuralism and made 'the principle of structure' an integral part of his dialectical method.

His concept of structure was composed of three elements: (1) the idea of wholeness, levelled against nineteenth-century atomism; (2) the idea of a 'hierarchy of layers', levelled against genetic reductionism; and finally (3) the idea of autonomy, levelled against naturalism and psychologism. All these components were combined in a dialectical relationship which prevented the distorting isolation and absolutization of any, thereby removing the danger of interpreting wholeness as totalism and autonomy as autarchy.

In his *Osnovy pedagogiki (Foundations of Pedagogy)* (1923) Hessen developed a conception of personality as rooted in objective, universal values. Personality, he claimed, is discovered only by participation in suprapersonal tasks. Without having physical existence, values have, nonetheless, 'universal validity'; hence, their mode of being is independent both from the individual psyche and from the supra-individual collective consciousness. Their objectivity and universality consists in their 'inner quality' which can be neither deduced from social conditions nor made dependent on their subjective recognition by a social group. In contrast to the norms of social life, cultural values act not by means of pressure but by means of 'appeal' - an appeal directed to people as free spiritual beings or personalities. The world of values is a hierarchical, multilayered structure; the higher values 'glow through' the lower ones and the lower values are both negated and preserved in the higher. History is a process of realizing values but the 'inner quality' of values is independent of changing historical situations. Philosophy is the self-consciousness of man as a value-realizing or - which means the same - a culture-producing being. The practical application of philosophy is pedagogy, or the science of education.

In later years Hessen tried to transcend this 'purely cultural' standpoint by developing the idea of a supracultural level of human existence. The first outline of this new conception appeared in his book *O sprzecznościach i jedności wychowania (On Contradictions and Unity in Education)* (1939). It divided human existence into four levels: biological, social, cultural (spiritual), and the level of grace. Hessen explained that the highest level was equivalent to the 'Kingdom of God' in the conception of such religious philosophers as Solov'ëv and August Cieszkowski.

The appearance of a religious perspective in Hessen's thought resulted from his deep interest in Russian religious thinkers, above all Solov'ëv and Dostoevskii. From Solov'ëv he took the idea of 'All-Unity', setting it against both atomistic pluralism and ossified monism; his criticism of isolating and absolutizing certain elements of a dialectical whole was directly related to Solov'ëv's criticism of 'abstract principles'. Dostoevskii was for him the greatest Russian thinker, whose works, especially *The Brothers Karamazov*, provided the formulation and intuitive solution of the deepest problems of modern ethics.

Hessen's articles on Dostoevskii and Solov'ëv show both his relative closeness to the broadly conceived 'religious-philosophical renaissance' in Russian thought (see Russian Religious-Philosophical Renaissance) and the peculiarity of his standpoint, clearly distinguishing him from other Russian thinkers inspired by Solov'ëv's philosophy. In his interpretation of Solov'ëv Hessen was a resolute Westernizer, having nothing to say on the 'Russian Idea'. He minimized the importance of Solov'ëv's theocratic utopianism, stressing instead the presence of Kantian motives in his philosophy and crediting him with overcoming utopian tendencies. Similarly, Hessen saw in Dostoevskii not a messianic utopian but rather a profound critic of utopianism, a thinker who powerfully and prophetically foretold the inevitable totalitarian consequences of utopian thinking.

In his essay 'Krushenie utopizma' ('The Breakdown of Utopianism') (1924) Hessen defined utopianism as 'absolutization of the relative': attributing absolute significance to the realization of certain relative values and setting them against the entire historical tradition. This explained, in particular, the nihilistic and purely instrumental attitude towards law, so characteristic of the utopian mentality.

2 Rule-of-law socialism

In Hessen's view, the most important modern utopia was, of course, socialism. He brilliantly analysed the destructive nature of socialist utopianism but stressed, at the same time, that socialism should not be reduced to its utopian features. In the serialized book *Problema pravovogo sotsializma (The Problem of Rule-of-Law Socialism)* (1924-7) he argued for the possibility of overcoming socialist utopianism from within, through combining socialist ideas with the liberal principle of the rule of law.

The first phase of socialism - 'Utopian socialism as the abstract negation of capitalism' - was a reaction against the juridical worldview of the French Revolution. It rejected the rule-of-law state as legalized anarchy, accused the

law of sanctifying social atomization and set against it the principle of the Good. According to Saint-Simon and his disciples law was necessarily subjectivist, pluralist and purely formal, whereas the good was objective, monistic and had a concretely definable content. Socialism, therefore, was conceived not as yet another political system, yet another legal order, but as an organic industrial system in which political power is replaced by the administration of things, and law, as a means of delimiting subjective, egoistic interests, gives way to organic social ties.

The second phase of socialism - classical Marxism - was described as a ‘real’ negation of liberal-capitalist order. It was real, and not merely abstract, because of the scientific character of Marxian analyses of the real contradictions of capitalist development. However, this scientific side of Marxism coexisted with a pronounced hostility towards the rule of law, defined as bourgeois hypocrisy, and with a strong strain of utopianism, culminating in the idea of the ‘withering away of state and law’. Another side of the Marxist utopia was the idea of the disappearance of economy or, rather, its replacement by a purely technical scientific organization of labour.

Russian revolutionary communism represented, in Hessen’s view, the quintessence of Marxist utopianism. Its inner nature was revealed by ‘War Communism’, that is, the subordination of all spheres of life to the militarized control of those in power. The degradation of law in the Soviet state revealed another aspect of the communist utopia: its striving by all possible means to realize an arbitrarily defined and absolutized ideal of positive Good.

The next part of the book is devoted to those currents of socialist thought, both Marxist and non-Marxist, in which the negation of capitalism was not only ‘real’, that is, derived from an analysis of real social contradictions, but also ‘transforming’ and ‘constructive’. Following P.I. Novgorodtsev (see Neo-Kantianism, Russian §4), Hessen discussed in this part both individual thinkers, such as Lassalle, Proudhon and Bernstein, and entire movements, such as French Syndicalism and German social-democratic revisionism. He presented them as different dialectical phases of the process by which the socialist Good was permeated by the principle of law.

In the last part of the book Hessen discussed the views of the British ‘guild socialists’ who defined their socialist ideal as ‘the New Middle Ages’. He welcomed this ideal with important qualifications, stressing that ‘the truth of liberalism’ had to be preserved. Referring to Chicherin (see Liberalism, Russian §2) he pointed out that law should not be confused with morality, as had been the case in the ‘Old Middle Ages’, and assured his readers that in the new medievalism the ‘impenetrability of the person’ and inviolable rights of the individual would be fully respected.

The essence of the new medievalism would consist in the sovereignty of law and plurality of legal orders mutually limiting each other. Law would no longer be identified with the official laws of the state and the state itself would undergo a process of ‘devolution’. Social atomization would be abolished, individuals would participate in the life of different communities without losing their individual autonomy. Property would lose its alienated character and become personalized, conditional and divided. The economy would be ‘bridled’ by law while preserving its autonomy and spontaneity. Human dignity would gain enormously because being dependent on law, rather than on arbitrary political power or blind economic forces, is not a humiliating relationship. In this way socialism would at last fully overcome its utopianism and reveal itself as a new and higher form of the rule of law.

3 Human rights and philosophy of law

In his works written under the Nazi occupation and under the communist rule in Poland (such as his UNESCO article and the booklet *Modern Democracy*) Hessen gave up his idealization of the pluralist legal order of the Middle Ages but further developed his conception of socialism as the last stage in the development of human rights and the rule of law.

In the first stage, represented by the absolute state, human rights took the form of the security of law, which implied the inviolability of the citizen’s person, property and dwelling. In the second stage, that of classical liberalism, security of law was transformed into the liberal principle of limited government, safeguarding the existence of a private sphere in which individuals were free from any interference by the state. The third stage - the modern democratic state - broadened the scope of governmental activity by endowing the state with the task of securing for its citizens certain forms of ‘positive freedom’. This provided justification for welfare policies, universal compulsory education and different kinds of social insurance. Socialism represented the fourth stage in this development. Its task was the further extension of ‘positive freedom’ and democratic rights through endowing all individuals with economic rights and safeguarding their implementation. Its highest principle was to be freedom

from exploitation or the right of everyone to be treated in economic life as a person, not a mere commodity.

The Soviet system, in Hessen's view, was a product of the destructive spirit of communist utopianism. Nevertheless he saw in it an encouraging phenomenon: the gradual overcoming of its initial utopianism, the vindication of some measure of autonomy for the economic sphere, the restoration of law and an increasingly positive attitude towards the historical heritage. He expected therefore that sooner or later the Soviet Union would develop into a democratic and law-respecting socialist state.

Hessen's views on the current state of affairs in the liberal-democratic West stemmed from his conviction that liberalism had entered the phase of its final, insoluble crisis. The Western countries, he argued, had abandoned the Puritan values of the fathers of liberalism; hence the prevalence of consumerist attitudes, a special kind of tolerance, expressing not respect for persons but indifference towards truth, and the peculiar infantilism of mass culture. Hessen saw these phenomena as symptoms of a dangerous, destructive process. It seemed to him that the overcoming of such a crisis could be expected only from an overall regeneration: from religious revival on the one hand and from socialist reconstruction on the other.

Hessen's conception of the changing content of human rights was linked to his general philosophy of law, elaborated in his unpublished book 'On the Essence and Calling of Law'. It placed law between two levels of human existence: the social level and the spiritual/cultural level. Law is below the spiritual level because it lacks *eros* and *caritas*: it replaces moral values by commands, attraction by pressure, autonomy by heteronomy. On the other hand, however, law stands above the social level of existence, since it is a means by which spiritual life influences its social basis. Without law members of a social group cannot become individualized persons. Absolutization of the legalistic outlook leads to atomism but disregard for law in the name of higher moral principles results, as a rule, in descent to the animal level.

In the last two years of his life Hessen was a victim of the enforced Stalinization of Polish intellectual life. He was subject to increasing ideological pressure and forced thereby to withdraw from teaching philosophy and pedagogy, becoming instead a teacher of Russian. After his death several of his works were translated into Italian and influenced Italian educational policies.

ANDRZEJ WALICKI

List of works

S.I. Hessen (1909) *Individuelle Kausalität. Studien zum transzendentalen Empirismus (Individual Causality. Studies in Transcendental Empiricism)*, *Kantstudien*. Ergänzungsheft 15, Berlin.

S.I. Hessen (1910) 'Mistika i metafizika' ('Mysticism and Metaphysics'), *Logos*, 1. St Petersburg.

S.I. Hessen (1912, 1913) 'Filosofiia nakazaniia' ('The Philosophy of Punishment'), *Logos*, 1-2. St Petersburg.

S.I. Hessen (1917) *Politicheskaia svoboda i sotsializm (Political Freedom and Socialism)*, Petrograd.

S.I. Hessen (1923) *Osnovy pedagogiki: Vvedenie v prikladnuiu filosofiiu (Foundations of Pedagogy: Introduction to the Applied Philosophy)*, Berlin; trans. *Fondamenti della Pedagogia come Filosofia Applicata*, introduction by G.L. Radice, Florence, 1936.(Also translated into Latin, Bulgarian, Polish, Serbian and Czech.)

S.I. Hessen (1924) 'Krushenie utopizma' ('The Breakdown of Utopianism'), in *Sovremennye zapiski*, Paris, vol. 19.

S.I. Hessen (1924-7) *Problema pravovogo sotsializma (The Problem of Rule-of-Law Socialism)*, serialized in *Sovremennye zapiski*, Paris, vol. 22 (1924), vol. 23 (1925), vols 27-9 (1926), vol. 30 (1927).

S.I. Hessen (1931) 'Bor'ba utopii i avtonomii dobra v mirovozzrenii F.M. Dostoevskogo i Vl. Solov'ëva' ('The Struggle Between Utopia and the Autonomy of Good in the Worldview of F.M. Dostoevskii and V. Solov'ëv'), in *Sovremennye zapiski*, Paris, vols 45-6.

S.I. Hessen (1938) *Szkola i demokracja na przelomie*, Warsaw.

S.I. Hessen (1939) *O sprzecznościach i jedności wychowania (On Contradictions and Unity in Education)*, Lwów-Warsaw.

S.I. Hessen (1949) 'The Rights of Man in Liberalism, Socialism and Communism', in *Human Rights. Comments and Interpretations. A Symposium*, edited by UNESCO with an introduction by J. Maritain, New York.

S.I. Hessen (1952) *Le virtù platoniche et le virtù evangeliche (Platonic and Evangelical Virtues)*, Rome: Armando Armando.

S.I. Hessen (1957) *Democrazia moderna (Modern Democracy)*, Rome: Armando Armando.

S.I. Hessen (1958) *Diritto e morale (Law and Morality)*, Rome: Armando Armando.

S.I. Hessen (1968) *Studia z filosofii kul'tury (Studies in the Philosophy of Culture)*, ed. A. Walicki, Warsaw: PWN.(Contains a comprehensive bibliography of Hessen's works.)

References and further reading

Broccolini, G. (1966) *Mito e realtà nel pensiero di Sergio Hessen (Myth and Reality in the Thought of Sergius Hessen)*, Rome.(A brief but comprehensive outline of Hessen's philosophy of culture.)

Walicki, A. (1987) 'Sergius Hessen: A Post-Revolutionary Synthesis', in A. Walicki, *Legal Philosophies of Russian Liberalism*, Oxford: Clarendon Press, 404-70.(An outline of Hessen's philosophy and a detailed presentation of his philosophy of law.)

Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, vol. 2; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, vol. 2, 697-702.(The author presents Hessen as the best, most original representative of the Neo-Kantian current in Russian philosophy.)

Heytesbury, William (before 1313-1372/3)

William Heytesbury, an English logician of the mid-fourteenth century, is, with Richard Kilvington, Richard Swineshead, Thomas Bradwardine and John Dumbleton, one of several philosophers known as the Oxford Calculators. In his works, Heytesbury examined mathematical topics related to motion and the continuum as well as paradoxes of self-reference and problems arising from intentional contexts, all within the context of terminist logic, through the resolution of sophismata. He is most noted for developing the mathematics of uniform acceleration, and for his contributions to developing the mathematical treatment of physical qualities such as heat.

William Heytesbury became a fellow of Merton College in 1330. He was a foundation fellow of Queen's College in 1340, but soon returned to Merton, becoming a doctor of theology there in 1348. He served as Chancellor of Oxford University in 1371-2, just before his death. All of his works concern logic and were written in Merton College during his Regency in Arts (1331-9). His writings were particularly popular in fifteenth-century Italy.

Heytesbury's logical investigations centred on sophismata, as his two major works, *Sophismata* and *Regulae solvendi sophismata (Rules for Solving Sophismata)* reveal in their titles. A sophisma is not a sophistical argument, but a statement the truth of which is at issue, given certain assumptions specified within a disputation (see Language, medieval theories of). The respondent must take a position on the statement's truth, and answer his opponent's questions for a certain time without being driven into absurdity or contradiction. Medieval sophismata were intended to be of logical interest, and may pose important philosophical problems.

Regulae solvendi sophismata has six chapters. The first deals with 'insoluble' sentences, self-referential paradoxes such as 'what I am now uttering is false'. The second, 'On knowing and doubting', deals with sentences involving intentional contexts, such as 'You know the king is seated', when the king is seated, and 'you know that sentence A (asserting this) is true, but you do not know what A asserts'. The third chapter deals with sophismata involving relative pronouns, and the fourth with sentences involving the terms 'begins' and 'ceases'. The fifth, 'On maxima and minima', deals with sentences concerning the limits of capacities measured on linear continua, and the sixth with sentences about change and motion.

The first three types of sophismata raise logical questions, but the rest border on mathematics and physical theory in their discussion of continua and motion. Heytesbury first makes a sophisma's logical form clear, using a variety of tools, and then applies the conditions specified to the determination of its truth. As an example, take 'Socrates ceases to know ten propositions', given that he now forgets one of the propositions but continues to know the other nine. The logical form of the sophisma is unclear on the surface, but William 'exposits' it as 'Socrates knew ten propositions immediately before now, but does not know ten propositions now'. Given that exposition, the sophisma follows. The objection that he ceases to know only one proposition, and continues to know the other nine, is handled by pointing out that 'ceases' imposes a compounded, not a divided, sense on 'ten propositions', since it precedes that phrase. The sophisma is not 'ten propositions (each) cease to be known by Socrates', which involves the divided sense and would receive a different exposition.

Again, take a proposition stated in the chapter on relative pronouns, 'Now there begins to be some instant which after the present instant will begin to be'. The exposition yields 'Now there is no instant which after the present instant will begin to be, and immediately after now there will be some instant which after the present instant will begin to be', and so the sophisma is true. The second part of the exposition is taken to mean 'there is no instant after the present instant such that at no instant between it and the present instant there is some instant which after the present instant will begin to be'. In his *De sensu compositio et diviso (Compounded and Divided Senses)*, an important summary of the state of the question on the issues it discusses, and on which medieval logicians afterwards relied, Heytesbury denies that from 'immediately after this there will be some instant', one may conclude that 'there will be some instant immediately after this', relying, as he often does, on word order to make his distinctions. In the first case, 'immediately' comes before 'some instant', and renders its supposition confused, so that what is said is 'there is no instant after the present instant such that at no instant between it and the present instant *there is some instant*'. In the second case 'immediately' comes after 'some instant', so that its supposition is not confused, and the sentence means that, for some specific instant I, 'there is no instant after the present instant such that at no instant between it and the present instant *I is*'. Thus, whereas the first is true, the second is false.

Heytesbury's central concern is often the logic of continual and infinite divisibility, a pursuit we nowadays regard as a part of mathematics. In constructing his sophismata, he makes free use of whatever can happen *secundum imaginationem*, that is, any set of conditions or events involving no formal contradiction, whether or not it is physically or metaphysically possible. (Thus the reductionism characteristic of his nominalism is disabled, and Heytesbury neither attempts to explain how an instant is really identical with things in motion, nor avoids speaking of instants merely because they are not really distinct from such things.) Despite their abstraction, these investigations contributed to the development of the natural sciences. In Aristotelian physics only space, time and motion are considered amenable to mathematical treatment. However, William's sophismata conceive qualities such as heat and whiteness in quantitative terms, while avoiding any metaphysical difficulties in such a conception by proceeding *secundum imaginationem*. As thinkers became used to the notion that anything lying along a continuous range could, in principle, be spoken of quantitatively, the search for a way to measure such things as heat became possible, and so mere logic laid the groundwork for a broadened application of mathematics to the physical world.

In the sixth chapter of his *Regulae*, Heytesbury developed the mathematics of uniform acceleration, proving the mean-speed theorem, in other words that a uniformly accelerated body will, over a given period of time, traverse a distance equal to the distance it would traverse if it were moved continuously in the same period at its mean velocity (one-half the sum of the initial and final velocities) during that period. He derives from this the conclusion that a uniformly accelerated body will, in the second equal time interval, traverse three times the distance it does in the first. Domingo de Soto observed the applicability of the theorem to free fall in 1555. Galileo, although he probably was not acquainted with Heytesbury's work, would have benefitted from the medieval background of which Heytesbury was a part (see Galilei, G.).

See also: Bradwardine, T.; Kilvington, R.; Logic, medieval; Natural philosophy, medieval; Oxford Calculators

JOHN LONGEWAY

List of works

William Heytesbury (1331-9) Works, in *Tractatus de sensu composito et diviso, Regulae solvendi sophismata, et cetera*, Venice: Bonetum Locatellum, 1494.(This includes also *Sophismata* and *De veritate et falsitate propositionis (On True and False Propositions)* and commentaries by Gaetano of Thienne. See Wilson (1960) for additional editions and manuscripts.)

William Heytesbury (1331-9) *Sophismata asina (The Asinine Sophismata)*, ed. F. Pironet, Paris: Vrin, 1944.(An early collection of sophismata each of which proves the proposition, 'You are a donkey' (*Tu es asinas*). Edition, commentary and full manuscripts; the only modern edition of any of Heytesbury's works.)

William Heytesbury (1331-9) *De sensu composito et diviso (The Compounded and Divided Senses)*, trans. in N. Kretzmann and E. Stump, *The Cambridge Translations of Medieval Philosophical Texts*, Vol. 1: *Logic and the Philosophy of Language*, Cambridge: Cambridge University Press, 1988.(An accurate and logically sensitive translation with a brief introduction.)

William Heytesbury (1331-9) *De probationibus conclusionum tractatus regularum solvendi sophismata (On the Proofs of Conclusions from the Treatise of Rules for Resolving Syllogisms)*, Pavia, 1483.(Includes work on the speed of uniformly accelerated bodies.)

William Heytesbury (1335) *Regulae solvendi sophismata (Rules for Solving Sophisms)*: Chapter 1 trans. P.V. Spade, Toronto, Ont.: Pontifical Institute of Medieval Studies, 1979; Chapter 2 trans. 'The Verbs "Know" and "Doubt", in N. Kretzmann and E. Stump, *The Cambridge Translations of Medieval Philosophical Texts*, Vol. 1: *Logic and the Philosophy of Language*, Cambridge: Cambridge University Press, 1988; Chapter 5 trans. J. Longeway in *William Heytesbury: On Maxima and Minima. Chapter 5 of Rules for Solving Sophismata with an anonymous fourteenth-century discussion*, trans. J. Longeway, Dordrecht: Kluwer, 1984.(Chapter 1 deals with self-referential paradoxes; Spade's translation and commentary is intelligent and well-informed. Chapter 2 deals with, among other things, logical problems raised by internal contexts; not much commentary, but the translation is accurate and logically astute. The translation of Chapter 5 contains a study and extensive analytical notes, arguing especially for the mathematical character of the work.)

References and further reading

Sylla, E. (1971) 'Medieval quantifications of qualities: the "Merton school"', *Archive for History of Exact Sciences* 8: 9-39.(An excellent introduction treating Heytesbury within the context of his school.)

Weisheipl, J. (1968) ‘Ockham and some Mertonians’, *Mediaeval Studies* 30: 163-213.(A biographical discussion of the Mertonian School, including Heytesbury.)

Weisheipl, J. (1969) ‘Repertorium Mertonense’, *Mediaeval Studies* 31: 174-224.(A listing of the works of the Mertonian School, including Heytesbury.)

Wilson, C. (1960) *William Heytesbury: Medieval Logic and the Rise of Mathematical Physics*, Madison, WI: University of Wisconsin Press.(An excellent study of the last three chapters of the *Regulae solvendi sophismata* with complete information on Heytesbury.)

Hierocles (2nd century AD)

The Stoic philosopher Hierocles lectured and wrote on ethics. He is important for his defence of the theory of oikeiōsis *(affiliation), a form of self-perception and self-love which becomes the foundation for human commitment to rationality and virtue. Observation of animal and human behaviour, he argues, shows that* oikeiōsis *is innate, rather than learned.*

Little is known of Hierocles' life. He should, however, be distinguished from the later fifth-century Neoplatonist of the same name. When von Arnim demonstrated the common identity of the author of a papyrus text (known now as *Foundations of Ethics*) and the author of ethical excerpts preserved in a late ancient anthology by Stobaeus, it became apparent that Hierocles was active as a philosophical lecturer, in Asia Minor and/or Athens, in the second century AD (a date confirmed by the dating of the papyrus). The excerpts, of which more than a dozen survive, are from discourses on topics in practical ethics (for example, family relations, marriage, civic duty, household management) and on religion, and resemble lectures by other Stoics such as Musonius Rufus and Antipater. *Foundations of Ethics* deals in a more technical manner with the doctrines of self-perception and *oikeiōsis* (affiliation) (see Stoicism §14) and provides philosophical arguments for the view that self-perception is the basis for our affiliation to ourselves.

Foundations preserves evidence about Stoic views on perception and self-awareness, the nature of the soul, and the 'self' as a complex relation of soul to body. It also sheds light on the problem of reconciling the self-directed inclinations of humans (self-love, the natural desire for one's own virtue to thrive) with the social inclinations which are the basis of the other-regarding virtues. Hierocles recognizes three affiliations: to oneself, to other humans, and to appropriate external objects.

One of the discourses (which can be found in von Arnim and Schubart, 1906: 61-2) also sheds light on our relations with others. People should be thought of as being arranged around us (that is, our mind) in a series of concentric circles. In the first and closest circle is our own body and the external things which help to fulfil its needs. The second circle contains our immediate relatives: parents, siblings, wife, children. The next two circles contain other relatives in order of closeness. The circles which follow include the various degrees of civic relationship (deme, tribe, city). Last is the circle of the entire human race. Hierocles proposes that social relations will be improved if we try to treat each group as though it were closer to us than it really is. He also suggests addressing remote relatives by more intimate labels (calling our cousins brothers, for example) in order to demonstrate and to strengthen our feeling for other people. The goal of this exercise is not to treat all human beings as we treat ourselves, but rather to 'reduce, through our own effort, the remoteness of our relationship with each set of people'. Hierocles' theory aims to achieve the closest possible bonds between people, but he falls short of calling for genuine altruism or impartialism.

BRAD INWOOD

List of works

Hierocles (2nd century AD) *Foundations of Ethics*, ed. G. Bastianini and A.A. Long, *Corpus dei papiri filosofici greci e latini*, part 1, *Autori Noti*, Florence: Olschki, 1992, vol. 1, 268-451. (Italian translation and commentary; includes excellent introduction and bibliography.)

Hierocles (2nd century AD) Fragments, ed. H. von Arnim and W. Schubart, *Hierokles: Ethische Elementarlehre*, Berliner Klassikertexte, Heft 4, Berlin: Weidmann, 1906.(Contains the fragments preserved in Stobaeus' anthology; also contains the now obsolete first edition of the papyrus text.)

References and further reading

Inwood, B. (1984) 'Hierocles: Theory and Argument in the Second Century AD', *Oxford Studies in Ancient Philosophy* 2: 151-83.(Discussion of date, identity and nature of argument.)

Isnardi Parente, M. (1989) 'Ierocle Stoico. Oikeiosis e doveri sociali' (Hierocles the Stoic. Oikeiōsis and Social Duties), in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36: 3, 2201-26.(A survey of Hierocles' identity, doctrines and relation to other schools.)

Praechter, K. (1901) *Hierokles der Stoiker*, Leipzig: Dieterich.(Basic work on identity and date.)

Hilbert's programme and formalism

In the first, geometric stage of Hilbert's formalism, his view was that a system of axioms does not express truths particular to a given subject matter but rather expresses a network of logical relations that can (and, ideally, will) be common to other subject matters.

The formalism of Hilbert's arithmetical period extended this view by emptying even the logical terms of contentual meaning. They were treated purely as ideal elements whose purpose was to secure a simple and perspicuous logic for arithmetical reasoning - specifically, a logic preserving the classical patterns of logical inference. Hilbert believed, however, that the use of ideal elements should not lead to inconsistencies. He thus undertook to prove the consistency of ideal arithmetic with its contentual or finitary counterpart and to do so by purely finitary means.

In this, 'Hilbert's programme', Hilbert and his followers were unsuccessful. Work published by Kurt Gödel in 1931 suggested that such failure was perhaps inevitable. In his second incompleteness theorem, Gödel showed that for any consistent formal axiomatic system T strong enough to formalize what was traditionally regarded as finitary reasoning, it is possible to define a sentence that expresses the consistency of T, and is not provable in T. From this it has generally been concluded that the consistency of even the ideal arithmetic of the natural numbers is not finitarily provable and that Hilbert's programme must therefore fail.

Despite problematic elements in this reasoning, post-Gödelian work on Hilbert's programme has generally accepted it and attempted to minimize its effects by proposing various modifications of Hilbert's programme. These have generally taken one of three forms: attempts to extend Hilbert's finitism to stronger constructivist bases capable of proving more than is provable by strictly finitary means; attempts to show that for a significant family of ideal systems there are ways of 'reducing' their consistency problems to those of theories possessing more elementary (if not altogether finitary) justifications; and attempts by the so-called 'reverse mathematics' school to show that the traditionally identified ideal theories do not need to be as strong as they are in order to serve their mathematical purposes. They can therefore be reduced to weaker theories whose consistency problems are more amenable to constructivist (indeed, finitist) treatment.

1 The formalism of the geometric period

Hilbert's conception of mathematics during the period of his work on the foundations of geometry was 'formalist' in the sense that it viewed an axiomatic system as providing not a family of truths specific to a particular subject matter but, rather, a 'theory-form' or hypothetico-deductive framework which potentially and, indeed, preferably served as the form for a variety of different particular sciences.

This doctrine has two distinct moments that it is important to distinguish. One is a 'rigourist' moment which maintains that, for the sake of achieving rigour in mathematical reasoning, axiomatic systems must be formalized to the point of emptying their distinctively mathematical concepts of all intuitional or distinctly mathematical content so that they become mere 'positions' in a network of logical relations. The other moment is a 'schematist' one which says that even for the sake of achieving truth in one's axioms, the non-logical terms in an axiomatic theory do not have to be tied to any particular intuitive content. Conceived in this way, the mathematical terms of an axiomatic theory become 'variables' of a sort (that is, terms that range over a variety of different intuitive contents) rather than 'constants' (that is, terms that signify a unique particular content).

Historically, the rigourist moment had its origins in the seventeenth-century discovery that various proofs in Euclid's *Elements* relied upon assumptions not stated in his axioms, postulates and definitions. In the eighteenth century, Lambert (1766: 162) proposed a remedy for such failures of rigour. He reasoned that since they occur when one makes tacit use of geometric intuition while conducting an inference in a proof, they ought to be eradicable by forcing all inferences to be validated without appeal to geometric intuition. This, he maintained, could be accomplished by requiring the validation of inferences to proceed in abstraction from the intuitive meanings of the mathematical terms involved and to be warranted solely on the basis of their 'symbolic characteristics'. Axioms would thus be treated like 'algebraic equations' in which mathematical terms are manipulated according to the logical positions laid down for them in the axioms and not according to their intuitive meanings.

This, essentially, was the conception of rigour that Pasch (1882) and Hilbert (1899a) put into effect in their axiomatizations of geometry a century later.

Inference, however, is only one part of mathematical practice and rigour, therefore, only one of its goals. Traditionally, it has also been required that the theorems, and, hence, the axioms, be true. To satisfy this goal, however, would seem to require treating the mathematical terms occurring in the axioms from the point of view of their intuitive meanings or contents. Such, at any rate, is the traditional view. The schematist moment of Hilbert's formalism challenges this view by maintaining that, ideally, a system of axioms should have a number of different (important) realizations and that these different realizations should be obtainable from one another by relatively simple schemes of interchange of the meanings of the non-logical terms in the axioms. What is most important, then, is for a system of axioms to have not one particular realization but a variety of different ones.

This interesting view is sometimes distorted and misrepresented by saying that Hilbert defined truth as consistency or that he advocated mere consistency rather than truth (as ordinarily defined) as the goal of mathematical theorizing. Hilbert sometimes said things that encourage such readings (1899b: 39-40, 1900: 448, 1905: 125, 1922: 157). However, more careful consideration of the ideas that influenced his views together with a balanced reading of the texts themselves (1922: 162-5, 1923: 179-80; Hilbert and Bernays 1934: §1) indicates that this was not his position.

This type of formalist view was suggested by a striking phenomenon of nineteenth-century geometry - the so-called 'dualities' that emerged in projective geometry. Generally speaking, a duality is a pair of theorems one of which can be obtained from the other by a simple and uniform scheme of substitution of geometric terms. Among the simpler examples are the following pairs of propositions:

(1) For every two distinct points, there is exactly one line which is incident with both.
(1′) For every two distinct lines, there is exactly one point which is incident with both.
(2) For every three points that are not incident with the same line, there is exactly one plane with which they are all incident.
(2′) For every three planes that are not incident with the same line, there is exactly one point with which they are all incident.

(1′) results from (1) by interchanging 'point' and 'line'. They illustrate a general principle - the principle of duality for the projective plane - that covers many more instances. Similarly, (2) and (2′) illustrate the principle of duality for projective space. (Statements (1′) and (2′) are not true for ordinary or 'metric' geometry, of course, but only for geometries in which the usual elements are augmented by various ideal elements such as points and lines at infinity.)

In addition to these simple and basic dualities, there are many others, some quite striking. They include a duality linking theorems of the Riemannian geometry of the plane to theorems of the Euclidean geometry of the sphere under interchange of the terms 'straight line' and 'great circle', and a duality linking theorems concerning lines in Euclidean three-dimensional geometry to theorems concerning points in Euclidean space of four dimensions under replacement of 'line' with 'point'.

Hilbert, however, was more taken with a duality that had been observed to exist between geometry - specifically, the axioms of linear congruence in Euclidean geometry - and a wholly non-geometric subject - namely, the laws governing proportions of trait-couplings in mutations of certain varieties of fruit flies (Hilbert 1922/23: 84-6). This duality suggested to Hilbert the idea that certain non-logical 'forms' of thought might be so basic or useful to our thinking that they would belong to even its most widely disparate parts (intuitionally speaking). He therefore described it as 'more wonderful' than anything imagined in even the 'boldest fantasy' (Hilbert 1930: 380). Others, too (see Weyl 1927: 26-7, 1944: 635), were moved by dualities between remote subject matters.

The importance of the dualities for understanding Hilbert's formalist view is twofold. First, they suggest that in an axiom system, the mathematical terms do not function as constants but rather as variables of some sort. This in turn suggests that axiom systems are not to be seen as designed to capture a single, favoured interpretation that we think of as 'the truth'. Rather, they are to be seen as structuring a number of different subject matters. They are not designed to serve as theories of a particular subject matter but as 'theory-forms' - 'empty frames' (Pasch 1915:

11), ‘logical moulds’, ‘hypothetico-deductive frameworks’ (Weyl 1927: 25-6) - which provide a logical framework for the description of a variety of different subject matters (Hilbert 1899b: 40-1; see Detlefsen 1993 for more).

2 Axiomatization and the ideal method

The second respect in which the dualities of nineteenth-century projective geometry shed light on Hilbert’s formalism is in their reliance upon the use of so-called ‘ideal’ or ‘imaginary’ elements. These are elements that (need) have no intuitional or perceptual basis and whose sole justification is the simplifying or generalizing effects they have on our thinking about a given subject. Their use is well-illustrated by the duality between (1) and (1′) above. This depends upon the use of ‘points at infinity’ to serve as the points of incidence of pairs of parallel lines. (Without such ideal points, the duality between (1) and (1′) would have to be reduced to one in which ‘exactly’ were replaced by ‘at most’.)

Hilbert embraced the use of ideal elements in axiomatic theorizing. Indeed, he essentially identified the axiomatic method with the method of ideal elements (1926: 383). Specifically, he took the conditions for justified use of the ideal method to be the same as those for justified use of the axiomatic method; namely, that it be consistent with the underlying contentual practice to which it is applied and that it bring simplicity or efficiency to the production of mathematical knowledge (1926: 370, 372-3).

In Hilbert’s view, then, the justification of an axiomatic system depends upon more than its mere consistency. It depends as well upon its promotion of ‘epistemic efficiency’, which is what use of the method of ideal elements has to offer. The geometric dualities illustrate this point well; for every contentual proof, they yield not one but two theorems - one directly provided by the contentual proof, the other by application of the substitution scheme of the duality. One thus, roughly speaking, obtains two theorems for the ‘price’ (that is, the genuine proof) of one. The second, ‘free’ theorem will, of course, imply not only real propositions but also ideal or imaginary propositions (for example, the intersections of parallel lines at infinity). Not everything that it covers thus constitutes a gain in real knowledge. Nevertheless, a great part of what it covers does, and it produces this with superior ease or efficiency.

Hilbert described the axiomatic or ideal method as the expression of an important intellectual freedom - namely, the freedom to create and use imaginary elements. This freedom could be taken as far as one wanted to take it provided that it did not clash with an associated body of contentual practice and that it promoted simplicity or efficiency in the production of real knowledge (1900: 439-40, 1905: 135-6, 1926: 372, 379). Dedekind also spoke of the ‘freedom’ of the axiomatic method. However, he understood it in a sense importantly different from Hilbert’s - namely, as a freedom that contrasts axiomatic thinking not with contentual thinking, but with genetic thinking. Dedekind observed that genetically constructed objects (for example, the finite cardinals constructed as set-theoretic objects) inevitably take on features (for example, the infinity of the individual cardinals thus constructed) that are not relevant to their mathematical functioning. In Dedekind’s view, the axiomatic method provides an alternative by giving us the freedom simply to declare that a set of items has exactly the properties laid down for them in a given set of axioms. Items introduced in this way are ‘free creation(s) of the human intellect’ (1888: §73) which retain only the relevant features of their genetically constructed counterparts.

Axiomatic freedom according to Hilbert, on the other hand, was the freedom to introduce admittedly non-real, imaginary entities into our thinking when doing so increases efficiency (over against purely contentual reasoning) in the production of contentual knowledge (1926: 370-73, 379, 392). He thus assigned ideal elements a role similar to that which Kant assigned the so-called ‘ideas of reason’ in his critical philosophy. Kant saw ideas of reason as non-descriptive, regulative, conservative devices for the efficient development of the judgments of the understanding (*Critique of Pure Reason*, B85-6, 362, 385, 536-7, 825-7). We are free to use them so long as we do not confuse them with objective, descriptive laws applying to an external reality (which confusion leads to inconsistency). Hilbert saw ideal elements as a psychologically efficient (perhaps even indispensable) means of developing our real or contentual knowledge: we are free to use them so long as doing so does not lead to inconsistencies with contentual thinking (1905: 135-6; 1926: 383, 392; 1928: 471).

3 The formalism of the arithmetical period

This basic conception of the ideal or axiomatic method was a key ingredient of Hilbert’s thinking in both the

foundations of geometry and the foundations of arithmetic.

Hilbert's work in the foundations of arithmetic - his so-called 'programme' - was, indeed, nothing other than a defence of a particular application of the ideal method. What distinguished it from his work in the foundations of geometry was that the ideal elements used had a peculiarly 'logical' character, and that the consistency demand for use of these ideal elements could not be satisfied by reducing it to that of some more basic application of the ideal method.

The ideal elements that Hilbert proposed to defend in arithmetic were not imaginary mathematical items (for example, points at infinity, ideal numbers and the like) but imaginary *logical* items - what he referred to as 'ideal propositions'. These are not genuine propositions. They have neither meaning nor truth-value in the ('absolute') sense in which contentual propositions do. They are, rather, imaginary constructions (*Ideenbildungen*) that function only as symbolic or 'algebraic' devices for securing a desirable logic - namely, classical logic - for arithmetical reasoning. (See Hilbert 1922: 164-5; 1923: 179-80; 1926: 379-81; 1928: 467, 469-71; Hilbert and Bernays 1934: §1; see Hallett 1990 for a different interpretation.)

Examples that Hilbert gave of ideal propositions included, on the one hand, denials of true contentual generalizations (for example, '$\neg\forall x(x+1=1+x)$'), and, on the other, unbounded existential generalizations (for example, '$\exists x(\mathrm{Prime}(x)\ \&\ x>p)$', where *p* is some selected natural number) which follow classically from bounded, contentual existential generalizations (for example, '$\exists x(\mathrm{Prime(x)}\ \&\ x>p\ \&\ x<p!)$'). The former are 'ideal' because they do not bound the search for a counter-example to the generalization denied; the latter because they do not bound the search for a witness of the type they assert to exist. However, both help to preserve classical logic as the logic of arithmetical thinking; the first by protecting the law of the excluded middle for true contentual generalizations, the second by protecting the classically valid inference-form from '$\exists x(\phi x\ \&\ \psi x)$' to '$\exists x(\phi x)$', where the former is a contentual truth.

Hilbert's claim is neither that there is no logic of finitary or contentual reasoning nor that, though there is, we cannot know it. Rather, it is that even if such a logic were to be given, we would not want and, perhaps, psychologically speaking, would not be able to use it. For even if given, it would diverge from that logic according to which we most naturally and efficiently reason towards truth in contentual arithmetic (1926: 379-80, 1928: 470-2).

Hilbert thus introduced ideal propositions as purely formal or algebraic instruments for preserving (on a formal level) the desirable patterns of classical logical inference. But this purely formal conception of the ideal propositions induced a similarly formal character for arithmetic as a whole (1926: 376-83, 1928: 467-70; Hilbert and Bernays 1934: §1) - the logical terms as well as the arithmetical terms. The logical terms occur in purely formal devices (namely, ideal propositions) and must therefore themselves be treated from a purely formal point of view. Hilbert's formal conception of ideal propositions thus drove his formalization of arithmetic beyond where his formalization of geometry had gone - namely, to the inclusion of logical as well as non-logical terms (see Detlefsen 1993).

4 The consistency of arithmetic

As we will now see, this radical formalization of ideal arithmetic is what motivated Hilbert's belief in the possibility of a direct or finitary proof of its consistency.

Reasoning from the radical formalization of ideal arithmetic, Hilbert offered a purely syntactic formulation of its consistency problem; namely, to show the nonexistence of a formal proof whose endformula is '$0\neq 0$', or some similarly contentually refutable formula. In like spirit, he believed that there ought to be a purely finitary, contentual resolution of this problem. The problem is to search for a certain type of 'formal object'. The negative resolution of such a problem would, it would seem, lie as much within the province of finitary reasoning as, say, the task of proving the non-existence of two numerals *a* and *b* satisfying the equation '$a^2=2b^2$', a task which is finitarily manageable (1926: 383, 1928: 471).

Hilbert therefore undertook to give such a proof. It would, he believed, fill the final gap in the foundational enterprise by founding elementary ideal arithmetic. It would do so because it would be founded upon the 'absolute' truths of finitary judgment - judgments which neither need nor admit of any more basic justification

(1926: 376-7, 1922: 163; Hilbert and Bernays 1934: 20-1), and because the consistency of the other theories of ideal mathematics (most particularly those in analysis and geometry) can be reduced to that of ideal arithmetic (together with some basic set-theoretic apparatus).

This 'programme' of Hilbert's came under heavy attack in 1930-1 with the discovery of Gödel's incompleteness theorems. The second of these (G2), in particular, was important in this connection (see Gödel's theorems §6; Proof theory §2). It showed, for any formal axiomatic system *T* containing a fragment of arithmetic, how to construct a sentence Con_T of the language of *T* which seemingly 'expresses' the claim that *T* is consistent. It then went on to show that Con_T cannot be proved in *T* so long as *T* is consistent. If, therefore, *T* is rich enough to express all finitary proofs, it would follow that Con_T cannot be proved finitarily. From this it has generally been inferred that a finitary consistency proof is not possible for virtually any interesting portion of classical mathematics. Hilbert's programme therefore fails.

Whether such reasoning is correct depends, of course, upon whether Con_T really does 'express' the consistency of *T* in the appropriate sense. This, therefore, is one issue that must be resolved before the argument against Hilbert's programme from G2 can be conclusively evaluated. However, even if it is granted that Con_T does express the consistency of *T*, a further serious question arises: namely, whether the properties of Con_T that make for its unprovability in *T* (for example, the so-called derivability conditions of Hilbert-Bernays-Löb, or the RE condition of Feferman - see Gödel's theorems §5) are also properties that are indispensable to its ability to 'express' the consistency of *T*. Only if they are can the unprovability in *T* of Con_T be taken as showing the finitary unprovability of *T*'s consistency.

This deep and difficult problem has received little attention in the literature on Gödel's theorems and Hilbert's programme. But there are exceptions - see Detlefsen (1986: ch. 4) for a critical discussion of them. Detlefsen (1986, 1990) discusses what an adequate solution to the problem would look like and establishes the main features of two different types of negative solutions to it. Auerbach (1992) continues this discussion. In the end, the problem leads to fundamental questions concerning the nature of formal systems and to the perhaps even more fundamental question of what should be regarded as the basic unit of epistemic commitment in mathematics (see Detlefsen 1990 for further discussion).

Gödel's first incompleteness theorem (G1) has also been used against Hilbert's programme (see Kreisel 1976; Smorynski 1977, 1985, 1988; Prawitz 1981; Simpson 1988). The usual argument here begins with the basic premise that ideal arithmetic should be required to be a conservative extension of finitary or real arithmetic. G1 is then used to argue that this requirement cannot be met. The reasoning proceeds as follows: G_T (the undecidable sentence for a given ideal system of arithmetic *T*) can be proved by ideal but not by finitary methods; the ideal system in which G_T is provable is therefore not a conservative extension of finitary arithmetic; hence, by the basic premise of the argument, Hilbert's programme cannot be carried out. Detlefsen (1990) critically examines this argument, questioning, in particular, whether Hilbert's programme is rightly seen as requiring that ideal arithmetic be a conservative extension of real arithmetic. See also Webb (1980) for an argument that G1 actually confirms a formalist position like Hilbert's.

The Gödelian challenge to Hilbert's programme has also called forth other types of responses (see Proof theory §§3-4) - responses that concede the legitimacy of the arguments from G1 and G2, but which maintain that Hilbert's programme can none the less survive in a suitably modified form. For the most part, these arguments have taken one of two basic forms. In the first, the key element is weakening the evidential standards for consistency proofs to something less restrictive than Hilbert's original finitary demand. Some in this group (for example, Gentzen, Ackermann and Gödel) have argued that there are types of evidence having the same basic epistemic virtues as finitary evidence that extend well beyond the (traditionally identified) boundaries of finitary reasoning. Others (see Kreisel 1958; Feferman 1988; Sieg 1988) have argued not so much for a reconsideration of what should count as admissible evidence in a consistency proof as for a liberalization of our concept of what a consistency proof should do. They argue that the basic obligation of a consistency proof is to bring about some epistemic gain, and that there are epistemically gainful means of proving the consistency of ideal arithmetic that are not finitary. They therefore urge the replacement of the finitary/non-finitary distinction by a more refined hierarchical scheme that distinguishes grades of constructivity. Doing so, they believe, allows a significant partial realization of Hilbert's programme (see Proof theory §§2-4).

The other main type of response to Gödel's challenge is that represented by the so-called 'reverse mathematics' school of Friedman, Simpson and others. Instead of arguing for an extension of the types of reasoning to be allowed in consistency proofs, it aims at reducing the strength of the systems of ideal reasoning whose consistency needs proving. It begins by identifying more exactly those parts of traditional ideal mathematics that are truly indispensable to the distinctive achievements of classical mathematics. It then seeks axiomatizations that are 'equivalent to' these core parts and that therefore eliminate the extraneous elements of the usual axiomatizations. The idea is that the consistency of these diminished bodies of ideal reasoning might be susceptible to proof of a type to which that of their stronger standard counterparts is not. So far, significant success has been achieved along these lines (see Simpson 1988; Proof theory §4).

See also: Arithmetic, philosophical issues in; Constructivism in mathematics; Geometry, philosophical issues in; Intuitionism; Kronecker, L.; Logical and mathematical terms, glossary of; Mathematics, foundations of

MICHAEL DETLEFSEN

References and further reading

Auerbach, D. (1985) 'Intensionality and the Gödel Theorems', *Philosophical Studies* 48: 337-51. (Discussion of the intensional aspects of G2.)

Auerbach, D. (1992) 'How to Say Things with Formalisms', in M. Detlefsen (ed.) *Proof, Logic and Formalization*, London: Routledge.(Detailed defence of the view that G2 refutes Hilbert's programme.)

Bernays, P. (1930) 'Die Philosophie der Mathematik und die Hilbertsche Beweistheorie' ('The Philosophy of Mathematics and Hilbert's Proof Theory'), *Blätter für Deutsche Philosophie* 3, 4: 326-67; repr. in *Abhandlungen zur Philosophie der Mathematik*, Darmstadt: Wissenschaftliche Buchgesellschaft, 1976. (Philosophical discussion of Hilbert's proof theory.)

Bernays, P. (1935) 'Hilberts Untersuchungen über die Grundlagen der Arithmetik' ('Hilbert's Investigations in the Foundations of Arithmetic'), in D. Hilbert, *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer. (Philosophical and technical discussion of Hilbert's ideas in the foundations of arithmetic by his student and closest collaborator on foundational matters.)

Dedekind, R. (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; trans. W.W. Beman (1901), 'The Nature and Meaning of Numbers', in *Essays on the Theory of Numbers*, New York: Dover, 1963.(§73 includes an interesting statement of Dedekind's philosophical conception of the axiomatic method and its relationship to the older, genetic method.)

Detlefsen, M. (1986) *Hilbert's Program: An Essay on Mathematical Instrumentalism*, Boston, MA, and Dordrecht: Reidel.(Detailed examination of the arguments from Gödel's incompleteness theorems against Hilbert's programme. Focuses primarily on the arguments from G2 and the question of what is required of formulas that are to express consistency.)

Detlefsen, M. (1990) 'On an Alleged Refutation of Hilbert's Program using Gödel's First Incompleteness Theorem', *Journal of Philosophical Logic* 19: 343-77.(Critical appraisal of the arguments against Hilbert's programme from G1.)

Detlefsen, M. (1993) 'Hilbert's Formalism', *Revue internationale de philosophie* 47: 285-304.(Attempt to clarify Hilbert's 'formalist' conception of mathematics and its relation to Kant's views.)

Feferman, S. (1988) 'Hilbert's Program Relativized: Proof-Theoretical and Foundational Reductions', *Journal of Symbolic Logic* 53: 364-84.(Useful introduction to the author's conception of a workable and interesting modification of Hilbert's original programme.)

Hallett, M. (1990) 'Physicalism, Reductionism and Hilbert', in A. Irvine (ed.) *Physicalism in Mathematics*, Dordrecht: Kluwer.(Argues for a non-instrumentalist reading of Hilbert.)

Hallett, M. (1994) 'Hilbert's Axiomatic Method and the Laws of Thought', in A. George (ed.) *Mathematics and Mind*, New York: Oxford University Press.(Discussion of Hilbert's conception of the axiomatic method.)

Hallett, M. (1995) 'Hilbert and Logic', in M. Marion and R. Cohen (eds) *Quebec Studies in the Philosophy of Science*, vol. 1, Dordrecht: Kluwer.(Account of the relationship between Hilbert's conception of logic and his views in the foundations of arithmetic.)

Hilbert, D. (1899a) *Grundlagen der Geometrie*, Leipzig: Teubner, 7th edn, 1930; 2nd edn trans. L. Unger and P. Bernays, *Foundations of Geometry*, La Salle, IL: Open Court, 1971.(Influential treatment of geometry from the widely held formalist perspective of the late nineteenth century.)

Hilbert, D. (1899b) letters to Frege of 12 and 29 December 1899, trans. H. Kaal, in G. Gabriel *et al.* (eds) *Gottlob Frege: Philosophical and Mathematical Correspondence*, Chicago, IL: University of Chicago Press, 1980. (Basic statement of Hilbert's formalist view of theories as logical frameworks of concepts/relations.)

Hilbert, D. (1900) ‘Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongreß zu Paris 1900’, *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse* 253-97; trans. ‘Mathematical Problems’, *Bulletin of the American Mathematical Society* 8: 437-79, 1902.(Includes a number of interesting remarks concerning the nature of mathematical truth and the axiomatic method.)

Hilbert, D. (1905) ‘Über die Grundlagen der Logik und der Arithmetik’, in A. Krazer (ed.) *Verhandlungen des dritten internationalen Mathematiker-Kongreßes in Heidelberg vom 8. bis 13. August 1904*, Leipzig: Teubner, 174-85; trans. ‘On the Foundations of Logic and Arithmetic’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 130-8. (Early statement of Hilbert’s developing views on the foundations of arithmetic. Perhaps most important for its statement of the ‘creative principle’ (135-6), which provides an important insight into his conception of the axiomatic method.)

Hilbert, D. (1918) ‘Axiomatisches Denken’ (Axiomatic Thinking), *Mathematische Annalen* 78: 405-15; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer, 1935.(Transitional essay in which Hilbert attempts to formulate a general conception of the axiomatic method in mathematics and science.)

Hilbert, D. (1922) ‘Neubegründung der Mathematik. Erste Mitteilung’ (’New Foundations for Mathematics: Part 1’), *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 1: 157-77; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer, 1935.(Perhaps the clearest technical statement of Hilbert’s programme for the foundations of arithmetic prior to the statement in Hilbert and Bernays (1934).)

Hilbert, D. (1922/23) ‘Wissen und mathematisches Denken. Vorlesungen 1922/23’ (’Knowledge and Mathematical Thinking’), lecture notes taken by W. Ackermann; repr. Göttingen: Mathematisches Institut, 1988.(Largely philosophical discussion of Hilbert’s foundational ideas.)

Hilbert, D. (1923) ‘Die logischen Grundlagen der Mathematik’ (’The Logical Foundations of Mathematics’), *Mathematische Annalen* 88: 151-65.(Includes Hilbert’s distinction between the kind of truth - ‘absolute truth’ - that characterizes contentual thinking and the ‘truth’ that pertains to axiomatic or ideal thinking.)

Hilbert, D. (1926) ‘Über das Unendliche’, *Mathematische Annalen* 95: 161-90; trans. ‘On the Infinite’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967.(Perhaps the fullest and philosophically most revealing of Hilbert’s writings on the foundations of arithmetic. Points out certain connections between Hilbert’s and Kant’s philosophies of mathematics.)

Hilbert, D. (1928) ‘Die Grundlagen der Mathematik’, *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 6: 65-85; trans. ‘The Foundations of Mathematics’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967.(Mature statement of Hilbert’s views regarding the foundations of arithmetic. Includes an important philosophical characterization of the central philosophical motive of his proof theory (page 475).)

Hilbert, D. (1930) ‘Naturerkennen und Logik’ (’Logic and the Knowledge of Nature’), *Die Naturwissenschaften* 18: 959-63; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer, 1935.(Philosophically revealing statement of Hilbert’s mature thinking on the foundations of arithmetic. Points out respects in which Hilbert’s philosophical thinking represents a modification of Kant’s.)

Hilbert, D. and Bernays, P. (1934) *Grundlagen der Mathematik (Foundations of Mathematics)*, vol. 1, Berlin: Springer, 2nd edn, 1968.(Classic work of mathematical logic; includes valuable statements of Hilbert’s conception of the axiomatic method and finitary reasoning.)

Kreisel, G. (1958) ‘Hilbert’s Programme’, *Dialectica* 12: 346-72; revised version in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (Reflections of a leading contemporary foundational thinker on Hilbert’s programme. Interesting postscript added in 1978.)

Kreisel, G. (1976) ‘What Have we Learnt from Hilbert’s Second Problem?’, in *Proceedings of Symposia in Pure Mathematics*, vol. 28, Providence, RI: American Mathematical Society.(Discussion of what recent developments in mathematical logic (especially proof theory) have to teach us about Hilbert’s programme.)

Lambert, J.H. (1766) ‘Theorie der Parallellinien’ (Theory of Parallel Lines), *Magazin für reine und angewandte Mathematik* 2: 137-64, 3: 325-58, 1786; repr. in P. Stäckel and F. Engel (eds) *Theorie der Parallellinien von Euclid bis auf Gauss*, Leipzig: Teubner, 1895.

Nagel, E. (1939) ‘The Formation of Modern Conceptions of Formal Logic in the Development of Geometry’, *Osiris* 7: 142-224; repr. in *Teleology Revisited and Other Essays in the Philosophy and History of Science*, New York: Columbia University Press, 1982, 195-259.(Partially traces the origins of certain elements of the type of formalist outlook represented in Hilbert’s work in the foundations of geometry.)

Pasch, M. (1882) *Vorlesungen über neuere Geometrie (Lectures on Modern Geometry)*, Leipzig: Teubner; 2nd edn Berlin: Springer, 1926.(Influential modern statement of the view that rigour requires that the inferences in a proof be formalized.)

Pasch, M. (1915) *Mathematik und Logik*, Leipzig: Engelmann, 2nd edn, 1924.(Further statements of Pasch's views regarding the formalization of inference in mathematical proof.)

Peckhaus, V. (1990) *Hilbertprogramm und kritische Philosophie: das Göttinger Modell interdisziplinärer Zusammenheit zwischen Mathematik und Philosophie (Hilbert's Programme and Critical Philosophy: The Göttingen Model of Interdisciplinary Cooperation between Mathematics and Philosophy)*, Göttingen: Vandenhoeck & Ruprecht.(Historical discussion of the influence of the philosophers of the Friesian school at Göttingen on Hilbert's foundational thought.)

Prawitz, D. (1981) 'Philosophical Aspects of Proof Theory', in G. Fløistad (ed.) *Contemporary Philosophy: A New Survey*, vol. 1, The Hague: Martinus Nijhoff.(Useful discussion of the bearing of Gödel's theorems on Hilbert's programme.)

Schlick, M. (1925) *Allgemeine Erkenntnislehre*, Berlin: Springer; trans. A. Blumberg, *General Theory of Knowledge*, Vienna: Springer, 1974.(Chapter 1, §7 includes useful discussion of the nature of axiomatization.)

Sieg, W. (1988) 'Hilbert's Program Sixty Years Later', *Journal of Symbolic Logic* 53: 338-48. (Discussion of the progress and changes in Hilbert's programme since its formulation in the 1920s.)

Sieg, W. (1990) 'Relative Consistency and Accessible Domains', *Synthese* 84: 259-97.(Examination of the connections between Hilbert's foundational ideas and those of Dedekind and others.)

Simpson, S. (1988) 'Partial Realization of Hilbert's Program', *Journal of Symbolic Logic* 53: 349-63. (Survey of the progress made towards realizing the aims of Hilbert's programme via the approach of 'reverse' mathematics.)

Smorynski, C. (1977) 'The Incompleteness Theorems', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland.(Survey of some of the technical material pertaining to Gödel's theorems.)

Smorynski, C. (1985) *Self-Reference and Modal Logic*, New York: Springer.(Includes a clear statement of an argument that G1 destroys Hilbert's programme.)

Smorynski, C. (1988) 'Hilbert's Programme', *Centrum voor Wiskunde en Informatica Quarterly* 1 (4): 3-59. (Fullest presentation of Smorynski's argument that G1 refutes Hilbert's programme.)

Tait, W.W. (1981) 'Finitism', *Journal of Symbolic Logic* 78 (9): 524-46.(Philosophical discussion of the nature of finitary reasoning.)

Toepell, M.-M. (1986) *Über die Entstehung von David Hilberts Grundlagen der Geometrie (On the Development of David Hilbert's Foundations of Geometry)*, Göttingen: Vandenhoeck & Ruprecht.(Historical treatment of various of Hilbert's writings and lectures on the foundations of geometry.)

Torretti, R. (1978) *Philosophy of Geometry from Riemann to Poincaré*, Dordrecht: Reidel.(A survey of philosophical ideas (and their origins) concerning geometry during the period named in the title.)

Webb, J. (1980) *Mechanism, Mentalism, and Metamathematics: An Essay on Finitism*, Boston, MA: Reidel. (Insightful discussion of various of Hilbert's philosophical ideas and the bearing of Gödel's theorems upon them.)

Weyl, H. (1927) 'Philosophie der Mathematik und Naturwissenschaft', in *Handbuch der Philosophie*, Munich: Oldenbourg; revised and expanded trans. *Philosophy of Mathematics and Natural Science*, Princeton, NJ: Princeton University Press, 1949.(Excellent introduction to foundational concerns in mathematics and the natural sciences by one of the most profound thinkers of the twentieth century.)

Weyl, H. (1944) 'David Hilbert and his Mathematical Work', *Bulletin of the American Mathematical Society* 50: 612-54.(Survey of Hilbert's work including comments on his foundational ideas.)

Hildegard of Bingen (1098-1179)

Hildegard of Bingen saw herself as a prophet sent by God to awaken an age in which great troubles were besieging the Church and people no longer understood Scripture. She tried to alleviate the first problem by writing letters to secular and religious leaders and preaching against those she saw as the culprits, and to this end she undertook preaching tours throughout Germany, preaching in cathedrals, monasteries and synods. Her writings, primarily interpretations of her own visions, address the second problem by trying to cast a new light on Christian revelation through illustrating it with original vivid imagery and personifications of abstract concepts. Though her works are not, for the most part, clearly philosophical, Hildegard does show philosophical insight.

Hildegard of Bingen was born to a noble family in Bermersheim in Germany. At the age of eight, she entered a hermitage which soon became a full Benedictine convent (whose abbess she became in 1136). She seems to have had more than the average education given to a noble-born Benedictine nun; she received regular instruction in Latin from the vulgate Bible from the abbess, and was further taught by one of the monks. Nevertheless, Hildegard repeatedly describes herself as ignorant and unschooled. Perhaps she says this to give legitimacy to her own writings: if she was unschooled, then whence could her elaborate texts and interpretations of Scripture come, if not directly from God?

Her claim of divine origin for most of her writings was supported by Bernard of Clairvaux and Pope Eugenius III in 1146-7. Their support launched her public career as a prophetess, and her influence and reputation grew steadily throughout Germany and all of Europe. As a woman receiving visions from God, she advised, admonished and criticized political and religious leaders, and undertook at least three preaching tours through Germany (1160-3). This growing influence enabled her to found her own convent at Mount St Rupert, near Bingen, where she died in 1179.

Hildegard's three major works, *Scivias (Know the Ways)* (1141-51), *Liber vitae meritorum (Book of the Rewards of Life)* (1158-63) and *Liber divinorum operum (Book of Divine Works)* (1163-73) all consist of extremely detailed accounts of her visions followed by allegorical commentaries on them, as well as commentaries on relevant passages of Scripture. *Scivias* and the *Liber divinorum operum* are organized in the same general way: they follow the history of salvation from the creation of the universe and human beings, the fall of the devil, the devil's temptation of Adam and Eve, the fall from Eden, redemption first through the law, the prophets and the synagogue, and then through Christ, the church and the sacraments to the end of times. Although Hildegard organizes her visions around the chronology of the history of salvation, within her interpretation of any single vision her method is to explain every detail of the vision in the order in which it was first described, without regard for thematic organization. The *Liber vitae meritorum* is an examination of thirty-five vices contrasted with their corresponding virtues, stemming from visions of various monsters personifying the different vices. Hildegard also wrote scientific and medical texts, lives of saints, and the music and text for a number of liturgical songs collected under the title *Symphonia*. She believed that all music (vocal and instrumental) brings the human soul back to the original praise Adam raised to God in Eden and which was lost in the Fall.

Hildegard's work is best characterized as theology of an unphilosophical sort. She does not explain Christian doctrine discursively, but rather illustrates it with images and comparisons which she only very rarely explains. For instance, when she discusses virtues and vices she does not explain what they are, nor does she provide the defining characteristics of individual virtues or vices (see Virtues and vices). She also seems oblivious to contradictions; throughout her work, for example, she hesitates between the view that the human body is essentially and incurably evil and the soul necessarily good (forced to sin unwillingly by the body), and the view that the body can be made good (as the tabernacle of the soul) and the soul evil (responsible for sin). Hildegard sometimes does make philosophical distinctions (for example, between different faculties of the human soul: the intellect, the will, reason, passions), and sometimes too she asks distinctly philosophical questions. For instance, in *Scivias* she considers the problem that divine foreknowledge poses for human freedom, but her reply is a heated denunciation of the prideful who suspect God of injustice. However, in her answers to two letters inquiring about fine points of theology, Hildegard gives philosophical characterizations of eternity as timeless existence and of simplicity as indivisibility, and she distinguishes between God's simple eternal nature, and a human being's

understanding of God, which is composite and temporal (see Eternity; Simplicity, divine).

It often seems that Hildegard is fettered by her method of 'vision-commentaries', which leads to difficult and disorganized texts, contradictions and a general lack of clarity in her thought. It was her visions, however, which gave her an authority uncommon to medieval women, and without the visions as legitimizers of her thought, she may not have been able to write as freely and originally as she did. Paradoxically, Hildegard often claimed divine authority for her visions from the fact that she was a 'weak little woman', and that God chooses those who have the appearance of weakness as his prophets, in an age in which those who have the appearance of strength are corrupted. Thus her gender also helped her to establish her claim of divine inspiration.

See also: Mysticism, history of; Natural philosophy, medieval §4; Soul, nature and immortality of

CLAUDIA EISEN MURPHY

List of works

Hildegard of Bingen (1098-1179) *Opera omnia*, ed. J.-P. Migne, *Sanctae Hildegardis Abbatissae Opera Omnia*, Patrologia Cursus Completus Series Latina, vol. 197, Paris, 1855; repr. Turnhout: Brepols, 1976.

Hildegard of Bingen (1141-51) *Scivias (Know the Ways)*, ed. A. Führkötter and A. Carlevaris, *Hildegardis Scivias*, Corpus Christianorum: continuatio mediaevalis 43 and 43a, Turnhout: Brepols, 1978; trans. Mother Columbia Hart and J. Bishop, *Scivias*, New York: Paulist Press, 1990.(Barbara Newman's introduction to this volume is an excellent introduction to important themes in Hildegard's thought, and more especially to the *Scivias*.)

Hildegard of Bingen (1158-63) *Liber vitae meritorum (Book of the Rewards of Life)*, ed. A. Carlevaris, *Liber vitae meritorum*, Corpus Christianorum, Continuatio Medievalis 90, Turnhout: Brepols, 1995; trans. B. Hozeski, *The Book of the Rewards of Life*, New York: Garland, 1994.(Although Hildegard's descriptions of each vice and its corresponding virtue do not include philosophical definitions, they include insightful diagnoses of the psychological conditions that lead to behaviour characteristic of the vices, and of the ways in which someone might justify such behaviour to oneself or to others.)

Hildegard of Bingen (1163-73) *Liber divinorum operum (Book of Divine Works)*, ed A. Derolez and P. Dronke, *Liber divinorum operum*, Corpus Christianorum, Continuatio Mediaevalis 92, Turnhout: Brepols, 1996; trans. (abridged) R. Cunningham, ed. M. Fox, *Book of Divine Works with Letters and Songs*, Santa Fe, NM: Bear & Company, 1987.(Throughout this work, Hildegard hesitates between a dualist view of sin (that is, it is the nature of the soul to be good while it is the nature of the body to be corrupt) and the more orthodox position which she holds, for example in the *Ordo virtutum*, that the soul is responsible both for good and evil. She also hesitates between a fideist view of human reason's ability to understand God, and a more rationalist account according to which God gave human beings reason in part so that they can come to understand him.)

Hildegard of Bingen (1098-1179) Letters, ed. L. Van Acker, *Hildegardis Bingensis Epistolarium*, Pars Prima, Corpus Christianorum, Continuatio Mediaevalis 91, Turnhout: Brepols, 1991; Pars Secunda, Corpus Christianorum, Continuatio Mediaevalis 91A, Turnhout: Brepols, 1993; trans. J. Baird and R. Ehrman, *The Letters of Hildegard of Bingen*, vol. 1, New York: Oxford University Press, 1994.(Some of the letters have clearer philosophical significance than do Hildegard's major works, as she is often asked to answer distinctly philosophical questions as well as pastoral and liturgical ones.)

References and further reading

Dronke, P. (1984) *Women Writers of the Middle Ages*, Cambridge: Cambridge University Press.(A study of Hildegard's understanding of herself in the context of the twelfth century, drawn from a somewhat eclectic selection of materials including biographical fragments, writings in the natural sciences and letters.)

Gössmann, E. (1989) 'Hildegard of Bingen', in M.E. Waithe (ed.) *A History of Women Philosophers*, vol. 2, Dordrecht: Kluwer.(A study of Hildegard as a philosopher and a theologian, with special attention to her views on the place of women.)

Gottfried of St Disibod and Dieter of Echternach (1177-81) *Vita Sanctae Hildegardis*, ed. M. Klaes, Corpus Christianorum, Continuatio Medievalis 126, Turnhout: Brepols, 1993; trans. J. McGrath, ed. M. Palmquist, *The Life of the Holy Hildegard*, Collegeville, MN: Liturgical Press, 1995.(A translation of the earliest life of Hildegard, which contains some important autobiographical fragments. The translation is from A. Führkötter's German, not the Latin text.)

Newman, B. (1987) *Sister of Wisdom, St. Hildegard's Theology of the Feminine*, Berkeley, CA: University of California Press.(A comprehensive study of feminine themes in Hildegard's work, from her scientific view of

women's physiology and sexuality to the identification of some aspects of God as feminine, ranging through traditional themes such as Eve's special role in the Fall of Adam and Mary's role in the redemption of humankind. There are many insightful observations, but the arguments are sometimes unclear, and many of the conclusions are unsupported.)

Hillel ben Samuel of Verona (*c*.1220-95)

Hillel played a crucial role in the response of the philosophers in the Jewish community to the attacks made upon them by their enemies. He stoutly defended Maimonides while at the same time opposing the allegorical interpretation of miracles. Far less radical than Maimonides or Averroes, he tended to follow the approach of Aquinas. He also translated many philosophical texts from Latin into Hebrew. He was influenced by scholastic ideas and especially by the anti-Averroistic controversy. His major work, Tagmule ha-Nefesh (The Rewards of the Soul), completed in Forlì in 1291, deals with the nature of the soul and the intellect and with the spiritual requital of the soul after death. The chief purpose of this work is 'to explain the existence of the soul, its essence and its rational faculty, which continues to exist externally after death'.

1 Life and works

Hillel was a physician, translator, philosopher and Talmudic scholar. He studied medicine at the University of Montpellier and lived for three years (1259-62) in Barcelona. There he would have studied natural sciences. He would have been a pupil of the famous Talmudist Jonah Gerondi, an anti-Maimonidean thinker. Hillel also spent time in Rome, where he became acquainted with Zerahiah ben Isaac Gracian of Barcelona and Isaac ben Mordekhai (Maestro Gaio). He is known to have been in Naples and then in Capua (until after 1270), where he practised medicine and studied philosophy with Abraham Abulafia. After 1287 he retreated to Forlì, occasionally visiting Bologna and Ferrara. Between 1287 and 1289 he defended Maimonides. He sent two letters to that purpose to Maestro Gaio, at that time physician of Pope Nicholas IV. The occasion was the arrival at Ferrara of Solomon ben Abraham of Montpellier (Solomon Petit), the chief instigator of the anti-Maimonidean movement and teacher of Jonah Gerondi. Hillel requested that Maestro Gaio use his influence over the Jewish community of Rome to organize a synod at Alexandria, to which the Jewish communities of Germany and France would be asked to send representatives to discuss whether the reading of Maimonides' works should or should not be permitted. The final decision would be entrusted to a court of Babylonian rabbis.

Between 1289 and 1291 a controversy occurred between Hillel and Zerahiah ben Isaac Gracian over the rationalistic interpretation of the Bible. Hillel supported the historical reality of the supernatural events described in Scripture, such as the encounter between Jacob and the angel and the speech of Balaam's ass, whereas Zerahiah, stressing the twofold sense of the Bible - literal for the masses and allegorical-philosophical for the learned - stated that the miracles had taken place in prophetic visions and hid a philosophical meaning.

2 *Tagmule ha-Nefesh*

Hillel's major work, *Tagmule ha-Nefesh (The Rewards of the Soul)*, is divided into two parts, preceded by an introduction. The first part, divided into seven sections, deals with the nature of the soul and the intellect; the second, divided into three sections, addresses the question of the reward and punishment of the soul after death, addressing and relying for support on quotations from the Bible, the Talmud and midrash. In the introduction, based on the beginning of Aristotle's *De Anima (On the Soul)*, Hillel states that he wishes to write a concise treatise on the soul, because it is 'the most beautiful possession that mankind has' and its happiness constitutes the ultimate purpose of man. By means of a compilation of writings by philosophers, Hillel intends to discuss the existence, nature and recompense of the soul, since it is the soul that defines the human being. He further aims to establish that retribution after death is not corporeal. According to Hillel, a literal interpretation of the Bible and the Talmud can lead to the erroneous conclusion that the retribution of the soul is corporeal, that the soul is a body and that the angels and God, the prime source of emanation, are also bodies.

In the first part (sections I-II) Hillel draws on Avicenna's *al-Nafs (The Soul)* and Dominicus Gundisalvi's *De Anima (On the Soul)*. He then presents *Sheloshah Ma'amarim 'al ha-Sekhel (Three Articles on the Intellect)*, his own Hebrew translation of *Tractatus de Animae Beatitudine (On the Beatitude of the Soul)*, attributed to Averroes. Hillel holds that the soul exists (I) and is not an accident but rather an incorporeal substance. The soul is the substantial form of the human being. Its relation to the body is like that of form to matter. Its 'descent' into union with the body and its 'ascent', in separating from the body are not motions (II). The soul is immutable; furthermore, when united with the body, it is not subject to motion. It is indivisible and is a form between the specific and material forms, united through divine grace to the matter of the human body in order to perfect it (III).

The soul emanates from God through the mediation of the separate Intelligences. It represents the fourth stage in the emanatory process, directly after the 'active intellect'. The latter, one of the ten separate Intelligences, illuminates the rational soul during the cognitive process and is the cause of the passing of the mind from potentiality to actuality (IV). Hillel follows Averroes in asserting that there is only one eternal universal soul, one in substance and number, from which emanate the souls of individuals, like the rays from the sun; the plurality is accidental, a result of the number of subjects that receive it (V) (see Ibn Rushd; Neoplatonism).

The anti-Averroistic controversy, which inflamed the University of Paris in 1270 and 1277, exerted an important influence on *Tagmule ha-Nefesh*, in which special attention is paid to the Averroistic thesis of the oneness of the potential intellect for the whole of mankind (VI) (see Averroism). In order to refute this thesis, the adoption of which would lead to a doctrine of collective immortality, Hillel adapts a denial of faith, he adapts the first chapter of Aquinas' *De Unitate Intellectus contra Averroistas (Article on the Unity of the Intellect against the Averroists)* (VII): the soul is the form of the body, as the intellect is the form of the soul. From this Hillel infers that the rational part of every individual soul persists eternally after it leaves the body. Hillel states that he has not described the soul's faculties, except for the rational one, because 'the treatment of the other, corporeal faculties, such as the nutritive, the augmentative and the generative' is the object of medical science.

In the second part, Hillel states that the soul's reward or punishment after parting from the body is spiritual. The soul derives from a spiritual source to which it returns. Being incorporeal, the soul, cannot be subject to a corporeal reward or punishment (I). God granted humanity the possibility of attaining eternal happiness by means of the three intellects. Hillel describes them in detail, in particular the acquired intellect, which gives the intelligibles to the soul and is intermediary between the potential intellect and the active intellect. The latter watches over the human being in order that he may attain happiness. The quantity of the influence of the active intellect, exerted upon the rational and imaginative powers of the soul, varies with the nature of the individual. In addition to his imaginative and rational faculties, man must also correctly guide his sensitive and appetitive faculties, because on them depend the virtues and vices and the observation and violation of the divine precepts. The individual who has perfected the intellectual and ethical virtues can reach during his life the level of the active intellect and thereby acquire the ability to change the natural course of events by performing miracles. Such a soul, after the separation from the body, reaches a greater perfection than it had before the union. It is able to enjoy the happiness conferred by knowledge. The soul that has sinned, however, descends to a lower level. Through the imaginative power, which, like the intellective power, is immortal, such a soul will then imagine its punishments, and through the intellective power it will rationally know that it will not be able to enjoy spiritual happiness, because of its separation from the divine intellectual light, in accordance with the Avicennian tradition. In closing, Hillel hermeneutically divides the tales of the Rabbis into six classes and states that those passages which refer to corporeal retribution should not be interpreted literally.

See also: Averroism, Jewish; Maimonides, M.; Soul, nature and immortality of the

CATERINA RIGO

List of works

Hillel ben Samuel of Verona (1291) *Tagmule ha-Nefesh (The Rewards of the Soul)*, ed. J. Sermoneta, *Sefer Tagmulei ha-Nefesh le-Hillel ben Shmu'el mi-Verona*, Jerusalem: The Israel Academy of Sciences and Humanities, 1981.(His main philosophical work, which deals with the nature of the soul and the intellect, and with the spiritual requital of the soul after death.)

References and further reading

Davidson, H.A. (1988) 'Averrois Tractatus de Animae Beatitudine', in R. Link-Salinger (ed.) *A Straight Path: Studies on Medieval Philosophy and Culture. Essays in Honor of Arthur Hyman*, Washington, DC: The Catholic University of America Press, 57-73.(A study of the treatise *On the Beatitude of the Soul* attributed to Averroes.)

Elbogen, J. (1935-7) 'Hillel da Verona e la lotta per Mosè Maimonide', *Annuario di studi ebraici* 2: 99-105. (Hillel on the Maimonidean controversy.)

Sermoneta, G. (1962) *Hillel ben Shemuel of Verona and his Philosophical Doctrine*, Ph.D. dissertation, The Hebrew University of Jerusalem.(A monograph on Hillel's philosophy.)

Sermoneta, G. (1974) 'Mappelet ha-Mal'akhim' (The Fall of the Angels), in S. Pines (ed.) *Studies in Memory of*

Jacob Friedman, Jerusalem: The Hebrew University of Jerusalem, 155-203.(A study of Hillel's treatise on the fall of the angels.)

Ibn Rushd (1169-98) *Tractatus de Animae Beatitudine (On the Beatitude of the Soul)*, in *Aristotelis opera… cum Averrois Cordubensis variis in eosdem commentariis (Commentaries of Averroes on Aristotle)*, vol. 9, Venice: Juntas, 1573; repr. Frankfurt: Minerva, 1962.(Commentary on the intellect, attributed to Ibn Rushd, as it appeared in Latin in Christian Europe.)

Hindu philosophy

Hindu philosophy is the longest surviving philosophical tradition in India. We can recognize several historical stages. The earliest, from around 700 BC, was the proto-philosophical period, when karma and liberation theories arose, and the proto-scientific ontological lists in the Upaniṣads were compiled. Next came the classical period, spanning the first millennium AD, in which there was constant philosophical exchange between different Hindu, Buddhist and Jaina schools. During this period, some schools, such as Sāōkhya, Yoga and Vaiśeṣika, fell into oblivion and others, such as Kashmir Saivism, emerged. Finally, after the classical period only two or three schools remained active. The political and economic disturbances caused by repeated Muslim invasions hampered intellectual growth. The schools that survived were the Logic school (Nyāya), especially New Logic (Navya-Nyāya), the grammarians and, above all, the Vedānta schools.

*The central concerns of the Hindu philosophers were metaphysics, epistemological issues, philosophy of language, and moral philosophy. The different schools can be distinguished by their different approaches to reality, but all considered the Vedas (the sacred scriptures) authoritative, and all believed that there is a permanent individual self (*ātman*). They shared with their opponents (Buddhists and Jainas) a belief in the need for liberation. They used similar epistemic tools and methods of argument.*

In contrast to their opponents, who were atheists, Hindu philosophers could be either theists or atheists. Actually we can observe an increased tendency towards theistic ideas near the end of the classical period, with the result that the strictly atheistic teachings, which were more philosophically rigorous and sound, fell into disuse. Hindu metaphysics saw ātman *as part of a larger reality (Brahman).*

Because these views of the world differed, they had to be proved and properly established. Accordingly, logical and epistemological tools were developed and fashioned according to the needs and beliefs of individual philosophers. Most agreed on two or three sources of knowledge: perception and inference, with verbal testimony as a possible third. In this quest for philosophical rigour, there was a need for precision of language, and there were important philosophical developments among the grammarians and the philosophers who explained the Vedas (the Mīmāṃsakas). A culmination of these linguistic efforts can be seen in the philosopher of language Bhartṛhari. One of his greatest accomplishments was the full articulation of the theory that a sentence as a whole is understood in a sudden act of comprehension.

It is customary to name six Hindu schools, of the more than a dozen that existed, thus lumping several into a single school. This is particularly the case with Vedānta. The six are listed in three pairs: Sāōkhya-Yoga; Vedānta-Mīmāṃsā; Nyāya-Vaiśeṣika. This does not take account of the grammarians or Kashmir Saivism.

In their quest for freedom from rebirth, all the Hindu schools operated within the same framework. Their ultimate goal was liberation. How much they were truly engaged in the quest for liberation apart from their philosophical preoccupations is not always clear, yet they never doubted its real possibility.

1 General presuppositions

Hindu schools of philosophy developed in close, lively dialogue with other philosophical trends and schools. As early as 400-300 BC, both Pāṇini and the author of *Manusmṛti* (a third century BC book of laws) identified two major intellectual trends, one involving belief in the sacred texts known as the Vedas, and one involving their rejection. Those who considered the Vedas as their authority later developed into what we know as the Hindu schools.

With a few exceptions, most of the religious and philosophical movements aimed at liberation, complete freedom from life and rebirth. From about the eighth century BC, belief in rebirth was found among most philosophical and religious leaders. At first, the mechanism of rebirth was thought to be prompted by bad actions. It was also believed that by good actions a person became good and by evil actions a person became evil. Since with time this must have come to be perceived as rather simplistic, the idea of rebirth became more complex. A person was reborn just by acts, regardless of whether those acts were good or evil. Liberation from rebirth could be achieved by an absence of desire; desire of any sort, whether a craving for food, say, or for a new thing, entangled a person in the worldly mechanism of repetition (see Karma and rebirth, Indian conceptions of).

The various thinkers and teachers were specifically concerned with effective ways of achieving liberation. This meant establishing the basic presuppositions of the theory, such as what it is that truly exists, how this could be proved, and how liberation was to be viewed, and, moreover, how to promulgate such beliefs. There were constant discussions, an ongoing search for better ways of arguing with opponents. The formal requirements for building an argument were much disputed; each school believed that only its tools for debate were necessary, and that any others were useless.

An axiom held by most followers of the Vedic tradition was that there is a self (*ātman*) which travels from life to life. The 'life' in question need not be human; it can also be that of an animal.

In the early philosophical sources, the Upaniṣads, there is little room for any sort of agency beyond individuals with selves. It is only later that we find the idea of God or gods actively creating the universe and directing individual persons towards liberation or towards realizing some sort of aspiration towards the divine. Sometimes both these aims were combined.

2 Metaphysics

The concept of *ātman* was crucial in many debates, because there were many who either had a different understanding of it or who claimed to need no such concept. Argument helped towards a more precise articulation of the term, although many Hindu thinkers held that knowledge of *ātman* is only a partial understanding of reality; the individual self is only a part of the larger scenario of the universe. The universe was thought to be an all-encompassing spiritual entity, of which *ātman* is a minute fragment. Experiencing this spiritually, through meditative practices, frees a person from the ordinary way of things: such a person is not reborn, and does not repeat the anguish, pain, disease, old age and death of ordinary mortals, but is instead forever free. This can be accomplished through one's own efforts, although often the guidance of a teacher, a guru, is needed. These efforts may need to be extended over several lifetimes in order to work off all the accumulated karmic impressions. Karmic impressions, which may result from physical activities, speech or mental acts, are what actually bind people to the revolving process of rebirth.

Gradually, notions of divine intervention in the process of liberation found their way into numerous teachings. It was a combination of one's own efforts plus divine grace which would grant final deliverance, which was now not only freedom from repeated cycles of lives, but also either an identification with the divine, or companionship with a god as a lover or eternal servant. Some Advaita philosophers postulated a single ultimate principle, whereas others argued for the existence of an ultimate cause of the universe, namely God. Yet the worship of a multitude of gods was still widely practised, as it is to this day (see God, Indian conceptions of).

Udayana (eleventh century), one of the most prominent thinkers of the School of Logic (Nyāya), constructed an elegant set of arguments for the existence of God. Put crudely, his claim is that this multifarious world must obviously be the effect of some cause, and that cause must be nothing other than God. On the other hand, not all thinkers felt a need to trace the world to one primary cause, even though most took it for granted that causal chains are of prime importance in interpreting the world. The nature of the causal relation was much debated. Some claimed that an effect somehow already exists in latent form in its cause, just as yoghurt is potentially already present in milk even before the milk turns sour. In the same way, this whole manifold world somehow pre-existed in an undifferentiated primeval watery mass, into which it will dissolve itself again at the end of its existence. There are repeated existences and dissolutions.

Other philosophers, such as Śaōkara (eighth century), interpreted the relation between cause and effect in a slightly different way. The difference between the two is only apparent, because in reality the universe is only superimposed onto an unchanging, everlasting, universal and undifferentiated principle, the Brahman of the Upaniṣadic thinkers. We superimpose things out of ignorance. A favourite analogy is that of a man walking along the road, half-blinded by the brilliant sunshine at noon. Suddenly he jumps across an elongated shape in the road, out of fear that he may step on a snake. A passer-by laughs and asks, 'Are you afraid of a dirty old piece of rope?' The person who jumped with fear was superimposing a snake onto the old rope. In the same way, we superimpose the whole universe onto Brahman. In reality, there are no causal relations at all. We talk of such concepts to facilitate debate, but they have no place on the ultimate level.

3 Epistemic concerns

Philosophers differed in their views about the number and characteristics of the various means of knowledge (*pramāṇas*). The most widely accepted *pramāṇa* was perception. Inference and verbal testimony (such as an utterance by a competent speaker or a statement from the Vedas) were also considered important. Some schools added analogy and other special kinds of inference (see Knowledge, Indian views of).

Because of their differing ontologies, schools also disagreed about the objects of knowledge. Nyāya held that these were self, body, sense organs, mind, rebirth, pain and freedom (*mokṣa*). Sāōkhya linked the objects of cognition to the sense organs: the eye has colour as its object, the ear has sound, the tongue taste, and so on; inference has as its objects things beyond sensory perception, such as consciousness, the undifferentiated material stuff of the universe, and causal relations.

'Perception' was usually confined to sensory or external perception. Some thinkers also recognized a sort of mental perception for mental states (such as joy or anguish). This was sometimes classed as belonging to a larger category of internal perception which also included yogic perception. Yogic and other types of perception in turn could be classed as 'extraordinary', as in the Nyāya system, especially its later form (Navya-Nyāya, 'New Logic'). Precise definitions of perception varied widely. Some thought it was direct awareness of colour; others argued that it was a cognition arising from the relation of an object with the senses, which is not verbal and not erroneous, but definite. An exchange of ideas arose over whether perception is a direct experience, that is, non-propositional, and whether one can postulate a propositional level of perception. Nyāya and the classical Sāōkhya of Īśvarakṛṣṇa claimed to understand perceptions of two levels: non-propositional (roughly what we call 'sense data') and propositional (naming and attaching concepts to the sense-data) (see Sense perception, Indian views of).

Inference (*anumāna*) was the next most important instrument of knowledge. In the classical period, three kinds of inference were usually enumerated. Their definitions betray a certain confusion between old and new ideas of inference. Inference is used as a source of knowledge in cases where objects cannot be apprehended directly. The basis of inference is the consistent relation between the reason and the thing-to-be-proved (*sādhya*). This requires that we perceive the reason. Here is an example of an argument:

There is fire on the mountain. (Thesis)
Because there is smoke (which I see with my naked eye). (Reason)
Like in the kitchen. (Positive example)
Unlike in the lake. (Negative example)
Therefore there is fire on the mountain. (Conclusion)

The consistent relation is not explicitly stated in the syllogism, although it is quite obvious: where there is smoke, there is fire (see Inference, Indian theories of).

Verbal testimony was thought to be another decisive source of knowledge of things beyond sensory apprehension. A competent testifier is a reliable person who has direct knowledge, wants to communicate it and is also capable of expressing it. It was argued that the revealed sacred literature could be classed as testimony. There was also some discussion about whether to subsume this source of knowledge under inference. Other suggested *pramāṇas* were analogy and presumption. An example of presumption is the following: it is observed that Devadatta is fat; nobody sees him overeating during the day; so (the presumption is that) he must eat all night. Indian philosophers also discussed an argument known as *tarka*, a kind of reasoning that we call *reductio ad absurdum*.

4 Philosophy of language

From the earliest times, there was a preoccupation with language in Indian culture. 'Language' here means Sanskrit. A concern with the power and limitations of language is already clear in the *Ṛg Veda*. Of the six theoretical branches of learning listed in the Vedas, four are closely related to language: grammar, etymology, lexicography and poetry. The level of accomplishment eventually attained in these is demonstrated in Pāṇini's descriptive grammar, the *Aṣṭādhyāyī*. Patañjali's commentary on this, the *Mahābhāṣya* (second century BC) can be thought of as a kind of a bridge between grammatical and philosophical concerns (see Patañjali).

The fifth-century philosopher Bhartṛhari stands out among those who studied language. In his *Vākyapadīya*, he

considers the faculty of speech to be an instinct or intuition. He compares it to animal instinct and does not believe that language is learned. Language, according to Bhartṛhari, accompanies cognition - there is no cognition without language. His understanding of language is rather metaphysical, as he equates language with Brahman.

Bhartṛhari is particularly associated with the theory of *sphoṭa*, although the notion had been formulated some centuries before him. According to this theory, a sentence is an integral unit. Analysing a sentence in terms of phonemes, morphemes or words is useful for learning purposes, but the whole sentence alone is meaningful. When those who know a language hear an utterance in that language, they hear a sentence, not single words or phonemes; only those who do not know the language will hear individual bits of sound. As for the relation between word and meaning, Bhartṛhari holds that it is permanent and natural, not based on convention.

Althought all schools of philosophy were concerned with language, perhaps the earliest was Mīmāṃsā. The followers of this school were concerned with the interpretation of the Vedas, and in particular with the problem of the relation between words and sentences. The Mīmāṃsakas argued that, for a word to be intelligible, each utterance of it has to be identical with an earlier utterance that is now remembered. By extrapolation, words must be eternal. In the same way, the meaning of words is eternal, as is the relation between word and meaning.

Pāṇini had seen the need for a capacity for mutual connection between the meanings of words, and the Mīmāṃsakas similarly developed a set of conditions for meaningful and correct sentences. They named the capacity for mutual connection between the meanings of words ‘mutual expectancy’. For example, ‘he rides an elephant’ fulfils the condition of mutual syntactic expectancy, but a string of words such as ‘elephant, house, riding’ does not. But according to this condition, the sentence ‘he rides a house’ is also a sentence. So another condition, ‘semantic compatibility’, was added. In a sentence like ‘he rides a house’, the semantic compatibility is absent. The Mīmāṃsakas also required that the condition of ‘contiguity’ be fulfilled: words must not be spoken at long intervals or be separated by other words. Another condition was ‘the intention of the speaker’, about which there were varying opinions.

Each of the main branches of the Mīmāṃsā school developed its own theory regarding the semantic relationship between words and sentences. The adherents of Prābhākara believed that the meaning of a sentence arises directly from its collection of words. Conversely, words convey meaning only in the context of a sentence. Each word in a sentence conveys both its isolated meaning and the syntactic meaning. On the other hand, Kumarila Bhāṭṭa and his followers believed that the meaning of a sentence arises indirectly. Each word gives its individual meaning, and this uses up its significative power; therefore the syntactic relation must be obtained by means of a secondary significative power. This view was also shared by the Advaita Vedāntins, who, in order to be able to express truths about the Absolute, could not always use words with their primary meaning, but had instead to use the secondary meaning.

Of course, the Advaitins were not the only ones who distinguished between the primary and secondary meanings of words. This practice was well known among other philosophers, grammarians, and especially literary critics. In his *Mahābhāṣya*, Patañjali distinguished primary and secondary meanings, while Bhartṛhari discussed transfer of meaning (*upacāra*) through such tropes as simile, metonymy, synecdoche and so on. By the ninth century, Ānandavardhana, in his exposition of literary criticism, was discussing the ‘suggestive power’ of words. He observed that a text does not yield its full meaning to every reader, since the ideal reader must be trained in the symbolism and conventions of a text, and familiar with the realities to which the text refers. Such a reader has an intuitive grasp of the text that untrained readers lack.

The school of Nyāya was at first primarily concerned with theories of the relation between a word and its meaning. The Naiyāyikas did not consider this relation to be natural, but saw it as just a matter of convention. This conventional relation is called ‘significative power’ (*śakti*). *Śakti* applies to primary meaning only; although secondary meaning is accepted, it is considered only in terms of its relation to the primary meaning and can apply only to single words, not to whole sentences.

The Naiyāyikas and the grammarians stayed active for many centuries, in the course of which their teachings were transformed. The Naiyāyikas especially developed new terminology and techniques of argument; this change was reflected in their adoption of a modified name, Navya-Nyāya (New Logic) (see Gadādhāra; Gaōgeśa; Language, Indian theories of).

5 Moral issues

Two important principles govern Indian moral philosophy: karma and *dharma*. The theory of karma was articulated early in Upaniṣadic times (which are usually placed from 700 BC onwards, but were possibly earlier). It concerns the causal relation between acts and their results, although neither was always understood in a uniform way. In general, the workings of karma were not interpreted as a fatalistic mechanism. With the exception of a few schools, most Indian thinkers came to conceive of karma in terms of a kind of naturalistic law of causation. The best-known philosopher of the Upaniṣads, Yājñavalkya, was the first one to teach karma, which soon became discernible in almost all intellectual developments, as well as being a governing principle in everyday ethics.

The principle of *dharma* is closely connected with karma. *Dharma* literally means 'to uphold what is correct', what we may call today 'morality'. The precise translation of the term depends on the context. For example, we can translate *dharma* as 'justice' in cases where something that was unlawfully taken away is to be regained. Thus, in the epic *Mahābhārata*, it is justice for the Pāṇḍavas to regain their kingdom, which was illegally taken from them by their cousins, the Kauravas. There is also *dharma* as 'individual duty', according to a person's social and economic status in society. This could be compared to a certain extent with the Kantian idea of duty (duty for duty's sake). Then there is general *dharma* which applies to society as a whole, a guide in moral and social issues.

An important ideal in Hindu moral philosophy, that of the stages of life, is described in the body of literature known as the Dharmaśāstras. This endorses the determination of social status by birth, and prescribes for each individual (at least, each male of the two highest classes, namely priests and royalty) the various stages to progress through in life. The prescribed sequence is as follows: first, the socially responsible person should study and abstain from sexual relations; next, he should marry, bring up offspring and accumulate material possessions; third, he should become a religious seeker, leaving behind the comforts of home, family and riches (although his wife may still provide some familiar comfort); finally, he should leave the companionship of his wife and roam alone as an ascetic until death. Two value systems, one socially engaged, the other with an ascetic tendency, appear to be combined here (see Duty and virtue, Indian conceptions of). Closely related are the four aims of human life (*puruṣārtha*): material wellbeing, pleasure and enjoyment, morality and social responsibility, and, the ultimate goal, liberation from repeated birth. Here, too, two value systems are combined: the first three aims guide the socially engaged, whereas the last is the aim of a person in the final stage of life.

6 Philosophical schools: Vedānta

When speaking of Indian philosophy, it has become a convention to count six schools in the Hindu tradition. This division is rather artificial, especially since it involves putting the schools together in three pairs: Sāōkhya-Yoga; Vedānta-Mīmāṃsā; Nyāya-Vaiśeṣika. In fact, by the fourteenth century we find sixteen philosophical schools discussed in Mādhava's *Sarvadarśanasaṃgraha (Survey of the Major Philosophical Systems)*. Among these are schools opposed to the Hindu ones, such as Buddhism, Jainism and materialism. Mādhava places the materialists at the beginning of his treatise and culminates it with the Vedānta school to which he himself belongs. Including the different schools of Vedānta, we can count about a dozen Hindu schools.

Vedānta still survives and is the most influential school of modern times, having great intellectual and political figures among its adherents. 'Vedānta' describes several schools and numerous thinkers, and means 'the appendage to the Vedas', referring in this way to the body of texts known as the Upaniṣads. The Upaniṣads have been a source of inspiration and dogma since their beginnings around 700 BC. Embedded in them are ideas that came to dominate Indian thought, namely karma, rebirth, and liberation from the ever-revolving cycle of rebirth. The means of liberation is to experience the identity of the individual self (*ātman*) with a larger cosmic entity (Brahman). Individual thinkers each had a different interpretation of these tenets, but all essentially agreed on the means of liberation. Curiously, the development of Vedānta did not take place until more than a millennium after the earlier Upaniṣads. The most influential thinker of whom we know today was Śaōkara.

Śaōkara, like other Vedāntins, built on an earlier tradition. The work to which they all responded was Badarāyaṇa's *Brahmasūtra* (or *Vedāntasūtra*) of around AD 50. It stimulated many interpretive commentaries, which gave occasion for new schools to arise. The most prominent interpretation of the *Brahmasūtra* is known as Advaita Vedānta. It focuses on Brahman, which is understood as identical with *ātman*. Out of ignorance, the material world is superimposed on the ultimately empty Brahman; this superimposition is sometimes described as

an illusory projection (*māyā*). The first prominent name in this tradition is Gauḍapāda, who taught Śaōkara's teacher.

Śaōkara was prolific in his philosophical output. He commented on all the major Upaniṣads and the *Bhagavad Gītā([Upaniṣad, or Secret Teaching] Recited by the Lord Krishna)*, and a number of other works are ascribed to him. His Advaitism can be characterized as a strict nondualism: there is nothing other than Brahman, either real or unreal, and the goal is to know this through a trance-like experience which grants liberation (*mokṣa*) from rebirth. Were it not for this experience of truth, which is the vision of identity between *ātman* and Brahman, we would always superimpose this colourful world on transparent Brahman. This superimposition is an act of mistaking an unreal object for a real one, just as we superimpose silver on a piece of a glittering shell or a snake on a rope (see §2). If we could lift the superimposed object away from the real one, underneath we would find something altogether different.

This doctrine is austere and diminishes the importance of a personal God. It failed to stimulate the imagination of many people, and with time there was a strong reaction to such an abstract portrayal of reality. The form of Vedānta that flourished subsequently tended to have a more theistic cast. The earliest work of theistic Vedānta was Bhāskara's interpretation of the *Brahmasūtra*, whereby the individual self is both different and not different from God (Brahman). This doctrine was called 'the teaching of difference with no difference' (Bhedābhedavāda).

The *Brahmasūtra* was often seen in the light of theology devoted to the god Viṣṇu. The eleventh-century philosopher Rāmānuja, commenting on the *Brahmasūtra* in his *Śrībhāṣya*, claims that everything is Brahman, yet acknowledges the reality of individual selves and the material world. This teaching is called 'qualified monism' (Viśiṣṭādvaitavāda) because Brahman is described as Knowledge and as being merciful, all-powerful and all-pervading. Everything that exists is contained in Brahman, understood as a personal God who should be approached with constant devotion. Other interpreters of the *Brahmasūtra* postulated devotion to God; to many of them, he was some form of Viṣṇu, which indicates that they too had a problem with absolute monism. Therefore they introduced a modified monism: Nimbārka, for instance, combined both dualism and nondualism (Dvaitādvaitavāda).

The extreme position of disavowing monism was taken in the thirteenth century by Madhva (not to be confused with Mādhava), who claimed that there is an absolute difference between Brahman and individual selves (Dvaitavāda). Another extreme position was expressed by Vallabhacārya in his teaching of pure nondualism (Śudhādvaitavāda). Still other thinkers with other interpretations, such as Caitanya (1486-1534) of the Bengal Vaiṣṇavism, did not leave a corpus of literature behind them (see Gauḍīya Vaiṣṇavism; Brahman; Monism, Indian).

7 Other philosophical schools

Sāōkhya and Vedānta are similar with respect to epistemology and some ontological issues. Sāōkhya was an old dualistic school reaching back to the ontologies of Upaniṣadic times. It postulated an irreducible duality of consciousness and material stuff. Originally, the material stuff existed in an undifferentiated form, until it was disturbed by an intangible prodding of consciousness. Once disturbed, it produced twenty-three parts of the universe, with the human individual's parts in preponderance. Altogether, with consciousness and the undifferentiated material stuff, there are twenty-five things that exist. The goal of Sāōkhya was to experience the basic duality in a trance-like state, to discriminate between 'spirit' and 'matter'. Perhaps this dualism reflected vacillation between idealistic/metaphysical tendencies and naturalistic/materialistic tendencies.

Another old school was Vaiśeṣika, which in some respects was close to Sāōkhya. Like Sāōkhya, it strove to list all the things that exist in reality, to name everything there is. Such a proto-scientific enumeration of categories marks the antiquity of these systems. The number of ontological categories according to the classical Vaiśeṣika of Praśastapāda is six. Other philosophers enumerated as many as ten, others only seven. Among these categories, such as substance, quality and activity, we find a category of relation, inherence. Inherence is a relation between things that do not exist in isolation. It holds between qualities and substances, and between particulars and universals; a quality inheres in its substratum, a substance, so that, with a red apple, the red colour is a quality of the substance apple. This red colour cannot exist on its own, but always has to inhere in something, whether an apple or a hibiscus flower. The Vaiśeṣikas are known as the Indian atomists. Motion inheres in the atoms, which,

in their varying compositions as whole objects, are the substratum of motion (see Ontology in Indian philosophy).

The Vaiśeṣika school is frequently lumped together with the Nyāya ('Logic') school. The reason for this might be that later Nyāya philosophers took it upon themselves to comment upon and revise the old atomistic school. The word *nyāya* is often used for a maxim or an example in an argument, which is perhaps why it was adopted for the Logic school; earlier, however, it was used to refer to the system of Mīmāṃsā, yet another school.

The Nyāya school also had a list of basic categories. Their sixteen categories are quite obviously the parts of a rigorous argument: for example, instruments of knowledge, objects of knowledge, doubt, purpose, example, and so on. Thus the list consists of epistemological or proto-epistemological tools. The main preoccupation among the Naiyāyikas was to build proper arguments. They used a five-member syllogism based on the constant relation between logical reason and thing-to-be proved (*sādhya*). This relation came later to be known as concomitance or pervasion (*vyāpti*). The Naiyāyikas also tried to safeguard against possible mishaps in argument by distinguishing three kinds of fallacy in reasoning.

The early Mīmāṃsakas were completely engaged in interpreting the scriptures, which conveyed injunctions for ritual actions such as sacrifices and ceremonies. These actions should be performed because the Vedas say so - the Vedas are authoritative. Other pursuits, such as acquiring knowledge of oneself or engaging in philosophical debates about God, serve no purpose. Furthermore, any philosophical pursuit may give rise to doubts, and the doubts may extend to the authority of the Vedas. It was centuries before Mīmāṃsā was freed from such dogmatism, by the philosophers Kumārila and Prabhākara (both seventh century). They maintained atheistic positions, and made fun of inconsistencies in arguments for the existence of God. Prabhākara put forward a theory of the self-verifying nature of simple (non-propositional) perceptual knowledge which was heavily criticized by philosophers of other schools.

Yoga was a practical discipline of physical postures, breathing techniques and meditation whose origin we are unable to trace to any precise time, place or event. It went hand-in-hand with the ascetic life, no matter what the value system or outlook of its practitioners. Yoga is classically grouped with Sāōkhya. There may be two reasons for this. First, both systems date back to the ancient encounter in Indian history between Aryan nomadic conquerors and a settled agricultural people (which lasted several centuries from 1500 BC onwards). There are speculations that Sāōkhya developed within the newcomers' tradition, whereas Yoga was practiced among the original inhabitants. Perhaps just as the worldviews of these groups grew together, so were the schools assimilated one to the other. The second reason may have been that Yoga needed a theoretical background to become a system. Sāōkhya is the oldest recognizable proto-philosophical bundle of ideas. But what Sāōkhya and Yoga share is not really significant in view of the fact that many other schools adopted Yoga techniques.

According to the *Yogasūtra*, the aim of Yoga is to be able to attain a state of deep concentration (*samādhi*) by stopping the activity of the mind. Yoga differs significantly from Sāōkhya in requiring the grace of God for liberation. Yoga can be seen as a practical discipline capable of being adapted by various theoretical systems, especially those whose adherents are supposed to undergo spiritual experiences to achieve liberation from pain, anguish, longing and rebirth.

See also: Buddhist philosophy, Indian; Jaina philosophy; Mīmāṃsā; Nyāya-Vaiśeṣika; Sāōkhya; Vedānta

EDELTRAUD HARZER CLEAR

References and further reading

Chattopadhyaya, D. (1964) *Indian Philosophy: A Popular Introduction*, Delhi: People's Publishing House; repr. 1972.(Lucid and accessible survey of Indian philosophical schools.)

Coward, H.G. and Kunjunni Raja, K. (eds) (1990) *Encyclopedia of Indian Philosophies*, vol. 5, *The Philosophy of the Grammarians*, Princeton, NJ: Princeton University Press.(A useful handbook. Contains a historical résumé and discussion of the main philosophical concerns. Summaries of works by language philosophers and grammarians comprise most of the volume.)

Larson, G. and Bhattacharya, R.S. (eds) (1987) *Encyclopedia of Indian Philosophies*, vol. 4, *Sāōkhya: A Dualist Tradition in Indian Philosophy*, Princeton, NJ: Princeton University Press.(A handbook containing a historical survey and discussion of Sāōkhya philosophy. Most of the volume consists of summaries of the work of Sāōkhya thinkers.)

Halbfass, W. (1991) *Tradition and Reflection: Explorations in Indian Thought*, New York: State University of New York Press.(Asks new questions and develops novel ideas about important issues. The approach combines history, philology and hermeneutics.)

Halbfass, W. (1992) *On Being and What There is: Classical Vaiśeṣika and the History of Indian Ontology*, New York: State University of New York Press.(A comprehensive study which includes a broad range of texts by the best classical thinkers.)

Mohanty, J. (1992) *Reason and Tradition in Indian Thought: An Essay on the Nature of Indian Philosophical Thinking*, Oxford: Clarendon Press.(A modern philosopher trained in Western phenomenology and analytical philosophy interprets Indian philosophy, with reference to the original sources, for contemporary thinkers.)

Phillips, S.H. (1995) *Classical Indian Metaphysics: Refutations of Realism and the Emergence of 'New Logic'*, La Salle, IL: Open Court.(This book juxtaposes realists and idealists, focusing primarily on the systems of Old and New Logic, and gives responses to the arguments that the idealists directed against the Logicians.)

Potter, K.H. (1963) *Presuppositions of India's Philosophies*, Englewood Cliffs, NJ: Prentice-Hall.(Discusses issues in Indian philosophies, identifying liberation as the ultimate goal for most schools. In a very useful diagram, philosophers are arranged according to their stances on liberation.)

Potter, K.H. (ed.) (1977) *Encyclopedia of Indian Philosophies*, vol. 2, *Indian Metaphysics and Epistemology: The Tradition of Nyāya-Vaiśeṣika up to Gaōgeśa*, Princeton, NJ: Princeton University Press.(An account of the historical background and an excellent discussion of the main issues form the introduction to summaries of works.)

Potter, K.H. (ed.) (1981) *Encyclopedia of Indian Philosophies*, vol. 3, *Advaita Vedānta up to Śaṃkara and his Pupils*, Princeton, NJ: Princeton University Press.(A historical overview is followed by a discussion of the main philosophical issues. Most of the volume consists of summaries of a comprehensive collection of texts.)

Potter, K.H. and Bhattacharyya, S. (eds) (1993) *Encyclopedia of Indian Philosophies*, vol. 6, *Indian Philosophical Analysis: Nyāya-Vaiśeṣika from Gaōgeśa to Raghunātha Śiromaṇi*, Princeton, NJ: Princeton University Press.(A short historical survey is complemented by a table of fifty philosophers and their works. A brief discussion of the philosophical ideas accompanies the summaries of the works.)

Hippias (late 5th century BC)

The Greek Sophist Hippias of Elis is a familiar figure in Plato's dialogues. He served his city as ambassador, and he earned a great deal of money from his lectures. His unusually wide range of expertise included not only rhetoric but also history, literature, mnemonics, mathematics and natural philosophy.

As a contemporary of Socrates, Hippias of Elis appears (as a rather self-important character) in several dialogues of Plato. Because of his learning and eloquence, he served as ambassador for Elis, and he lectured with great success at Athens and elsewhere. His intellectual range was very wide. Plato depicts him as lecturing on Homer (*Hippias Minor* 363a-b) and on astronomy and natural philosophy (*Protagoras* 315c). He also composed poems in several genres, and is credited with an important discovery in geometry, the *quadratrix*, a curve for trisecting angles and squaring the circle. He was one of the earliest specialists in mnemonics, and is reputed to have been able to memorize fifty names at first hearing. His historical research included compiling a list of Olympic victors (an important basis for Greek chronology) and collecting information on genealogies, tribes, famous or legendary persons and the foundation of cities.

Hippias' energy and versatility must have been extraordinary. Plato makes fun of him for appearing once at Olympia dressed in his own handiwork from head to foot, including his home-made signet ring, ornamental belt and sandals. He offered to answer questions on any subject, and he seems to have a prepared speech ready for any occasion (*Protagoras* 347b; *Hippias Minor* 363a-d, 369c).

The most original work of Hippias was a book from which we have the opening sentence:

> Some of these things have perhaps been said by Orpheus, some by Musaeus briefly here and there, some by Hesiod and Homer and other poets, as well as by prose authors, both Greek and barbarian. I will compose a novel and manifold discourse by combining the most important and similar items from all these.
>
> (fr.6)

The book seems to have been an allegorical work in which doctrines from Presocratic philosophy were attributed to the early poets. Hippias is apparently the authority Plato and Aristotle are following when they find the doctrines of Thales and Heraclitus in verses from Homer, Hesiod and Orpheus. It is likely that he was Aristotle's primary source for information on Thales. Hippias thus appears as the first doxographer and hence as the remote precursor for ancient and modern historians of philosophy (see Doxography).

Plato's satirical presentation suggests that he found Hippias' vanity great and his learning superficial. Nevertheless, of all the Sophists Hippias was apparently the most universal in his intellectual activity, a fifth-century precedent for Aristotle.

Hippias is the title of two dialogues in the Platonic corpus, the shorter of which (*Hippias Minor*) is surely by Plato. The authenticity of the longer dialogue (*Hippias Major*) is contested.

See also: Sophists

CHARLES H. KAHN

References and further reading

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, vol. 3, Cambridge: Cambridge University Press.(A full, scholarly account.)

Hippias (late 5th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 7th edn, 1954, vol. 2, 326-34.(The standard collection of the ancient sources; includes Greek texts with translations in German.)

Kerferd, G.B. (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press.(A briefer, more personal interpretation than Guthrie (1969).)

Patzer, A. (1986) *Der Sophist Hippias als Philosophiehistoriker (The Sophist Hippias as Historian of Philosophy)*, Freiberg and Munich: Verlag Karl Alber.(An ambitious, largely convincing reconstruction of Hippias' doxographical work, the *Synagōgē* or 'Gathering-together', as source for many passages in Plato and Aristotle where early poetical texts are cited as anticipations of later philosophical views.)

Plato (*c*.395-387 BC) *Hippias Minor*, trans. H.N. Fowler, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1926.(Short conversation between Socrates and Hippias on Achilles and truthtelling.)

Plato (*c*.386-380 BC) *Protagoras*, trans. C.C.W. Taylor, Oxford: Oxford University Press, 2nd edn, 1991.(Hippias appears in a gathering of famous Sophists and speaks on several subjects.)

Plato (*c*.380-367 BC) *Hippias Major*, trans. P. Woodruff, Oxford: Blackwell, 1982.(Longer conversation between Socrates and Hippias on the nature of the beautiful; Platonic authorship is dubious.)

Snell, B. (1976), 'Die Nachrichten über die Lehren des Thales und die Anfänge der griechischen Philosophie- und Literaturgeschichte' (Reports on the doctrines of Thales and the beginnings of Greek history of philosophy and literature), in C.J. Classen (ed.) *Sophistik*, Darmstadt: Wissenschaftliche Buchgesellschaft, 478-90.(The basic scholarly paper on Hippias as authority for Plato and Aristotle; in German only.)

Sprague, R.K. (ed.) (1972) *The Older Sophists*, Columbia, SC: University of South Carolina Press, 94-105.(Full English translation of the fragments and testimonia from Diels and Kranz (1954).)

Hippocratic medicine

The Hippocratic corpus is a disparate group of texts relating primarily to medical matters composed between c.*450 and* c.*250* BC *and dealing with physiology, therapy, surgery, clinical practice, gynaecology and obstetrics, among other topics. The treatises are (for the most part) notable for their sober naturalism in physiological theory, their rejection of supernatural explanations for disease, and their insistence on the importance of careful observation. Although embodying a variety of different physiological schemes, they are the origin of the enormously influential paradigm of humoral pathology. In antiquity, the authorship of the entire corpus was mistakenly ascribed to the semi-legendary doctor Hippocrates of Cos (*fl. c.*450* BC*).*

1 The Hippocratic corpus

The collection known as the Hippocatic corpus comprises some sixty or more texts, of greatly varying styles, subject matters, lengths and probable origins, written in terse and unattractive Greek. The earliest may date from the mid-fifth century BC; the latest are certainly third century BC. The corpus contains the first treatises in Western rational medicine; and the best of them display the fruits of meditation not only upon practical problems in therapy but also upon fundamental methodological questions in science.

It is unclear how many if any of the Hippocratic texts can be associated with Hippocrates himself. Some of the treatises are clearly working notes of practising doctors (*Epidemics*); others are attempts to supply theoretical frameworks within which such practical medicine may flourish (*Prognosis*). Some treat the nature and aetiology of specific diseases (*On the Sacred Disease*), others the impact of certain environmental features on general epidemiology (*Airs, Waters, Places*); others still are devoted more to practical therapeutics (*Regimen in Acute Diseases*). Some are surgical (*On Joints*), while others (notably the late text *On the Heart*) deal with anatomy; yet others are ethical in nature, notably the famous *Oath* (and *On Decorum*), while some (*On Breaths*, *On the Art*) were written to justify the practice of medicine against its detractors. Also important was the influential anti-theoretical account of medicine in *On Ancient Medicine*, as well as important groups of texts on general pathology, gynaecology and obstetrics.

2 Naturalism

Prominent among the texts is a commitment to sober clinical observation and an attempt to account for diseases on the basis of a naturalistic approach to physiology, epidemiology and therapeutics, eschewing accounts of disease as the results of divine visitation, on the model of Apollo’s plague at the beginning of the *Iliad*.

In the opening chapter of *On the Sacred Disease* (*c.*420 BC), which deals with epilepsy and related seizures, the author affirms roundly that, however terrifying and numinous its appearance, this ailment is no more nor less sacred than any other. He exposes the claims of the ’purifiers’ that it is divinely sent (and hence requires apotropaic ritual treatment) to a cogent and systematic critique. Their ’account’ is simply an attempt to conceal their own ignorance, their ’treatments’ ritual mumbo-jumbo; they take the credit if patients recover, while arming themselves with excuses when they do not. Their taboo prohibition on goat products suggests that people who associate more with goats (such as the Libyans) should be more prone to the disease (which they are not), and that the causes of the disease are implicitly non-supernatural. Moreover, anatomical investigation of afflicted creatures shows them to be suffering from a superfluity of plegm, which suggests that the disease is caused by phlegm (which is cold and moist) blocking crucial pathways in the brain.

This text, which is typical of the corpus, is remarkable for the acuity of its criticism of the opposing model, but equally for the cavalier manner in which its author erects his own preferred physiological model on the flimsiest of evidential bases. Equally optimistic extrapolations from inadequate evidence may be found in *Airs, Waters, Places*, whose author attributes the prevalence of particular types of ailment in certain locales to their aspect, the type of water available, and their prevailing atmosphere. The notion that such environmental influences play some role in epidemic and endemic diseases is no doubt sound and empirically based: the theory that seeks to integrate that evidence is not.

The physiological theories espoused by Hippocratic texts are various, but in general they seek to explain health and disease in terms of the interplay of basic internal elements. Thus *On Regimen* makes fire and water the primary

constituents of the human body, the various ratios between them being responsible for different individuals' differing physiology and psychology, with serious imbalances among them producing disease. By contrast, *On Breaths* makes air fundamental: internal air regulates the body, inspired air nourishes it, noxious air harms it.

On the Nature of Man, rejecting element theory as too remote from the actual facts of human physiology, outlines the theory of the 'four humours' inextricably associated with Hippocrates' name. The four basic physiological constituents are blood, phlegm, yellow bile and black bile, a preponderance of any of which yields one of the four basic character types (sanguine, phlegmatic, bilious, melancholic), and excesses of which cause disease. Cures are thus effected by redressing the balance (almost all of Hippocratic therapeutic theory is allopathic in form, 'opposites cure opposites' being a constantly recurring aphorism). The theory purports to rest upon solid empirical foundations: yet black bile is a theoretical construction, answering (despite its enviable conceptual longevity) to no actual bodily fluid.

In reaction against this type of theorizing, *On Ancient Medicine* takes issue with 'new-fangled hypotheses' as being empirically undeterminable and therapeutically pointless. Medicine consists of seeing what sorts of regimen produce what kinds of illness, and what functions best in their treatment. This text thus anticipates medical Empiricism (see Hellenistic medical epistemology).

Equally, the case histories of the *Epidemics* are apparently untheoretical: they record the onset and progress of diseases, merely noting causal relevant features, such as overwork, fatigue, overheating, over-indulgence in food, drink or sex, as well as location and regimen. But these are precisely the sorts of things that might be expected to disturb a delicate internal humoral balance, not factors to be associated with an infectious model of disease transmission (whom the patient had visited, and so on) which is utterly absent from the Hippocratic corpus (even though Thucydides, for example, was well aware that the great Athenian plague was contagious). The Hippocratics do occasionally point to cuts and wounds as being relevant to the development of certain diseases, although they have little idea how such septic mechanisms might operate. But the notion that pathogens might be directly transmitted from one person to another is implicitly rejected as being incompatible with the dominant model of health and disease as depending upon the state of internal equilibrium or otherwise.

See also: Galen

R.J. HANKINSON

References and further reading

Hippocratic corpus (*c*.450-250 BC) in *Hippocrates*, trans. W.H.S. Jones, E.T. Withington and P. Potter, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 6 vols, 1923, 1931, 1987.(Good selection of the basic Hippocratic texts with parallel Greek/English translation.)

Lloyd, G.E.R. (1979) *Magic, Reason and Experience*, Cambridge: Cambridge University Press.(Magisterial account of the nature of Greek science; chapter 1 is largely devoted to Hippocratic medicine.)

Longrigg, J. (1994) *Greek Rational Medicine*, London: Routledge.(Generally reliable introductory survey.)

Historicism

Historicism, defined as 'the affirmation that life and reality are history alone' by Benedetto Croce (1938: 65), is understood to mean various traditions of historiographical thinking which developed in the nineteenth century, predominantly in Germany. Historicism is an insistence on the historicity of all knowledge and cognition, and on the radical segregation of human from natural history. It is intended as a critique of the normative, allegedly anti-historical, epistemologies of Enlightenment thought, expressly that of Kant. The most significant theorists and historians commonly associated with historicism are Leopold von Ranke, Wilhelm Dilthey, J.G. Droysen, Friedrich Meinecke, Croce and R.G. Collingwood.

The main antecedents for the development of historicism are to be found in two key bodies of work. J.G. Herder's Outlines of a Philosophy of the History of Man (1784) argues against the construction of history as linear progress, stating rather that human history is composed of fundamentally incomparable national cultures or totalities. G.W.H. Hegel's The Philosophy of History (1826) insists on the historical situatedness of each individual consciousness as a particular moment within the total progression of all history towards a final goal. The shifting fusion of these ideas provides the foundation for both the strengths and the problems of historicism. Historicism follows both Herder, in attempting to do justice to objective history in its discontinuity and uniqueness, and Hegel, in attempting to determine general patterns of historical change. Indeed, historicism can perhaps be best termed a Hegelian philosophy of history without an all-encompassing notion of progress.

Rather than constituting a unified intellectual movement, historicism is best known for its elusiveness. Its multifarious quality can be inferred from the variety of critical positions taken up against it. Influential critiques of historicism have been written by Friedrich Nietzsche, Friedrich Rickert, Ernst Troeltsch, Walter Benjamin, Karl Löwith and Karl Popper. Critical engagement with historicism has focused on its alleged relativism, its alleged particularism, its alleged claims to totality, its alleged subjectivism and its alleged objectivism. More positive debates with historicism have significantly influenced the thought of Martin Heidegger, Edmund Husserl and Hans-Georg Gadamer.

1 Historicity and knowledge

For much of the nineteenth century historicism was broadly linked with the great historian, Ranke, whose formulated intention was to describe history 'as it actually happened'. In this, historicism moves close to simple empiricism. However, later readings of historicism have moved substantially away from this conception.

Karl Heussi has argued in one of the most significant critiques of historicism, *Die Krisis des Historismus (The Crisis of Historicism)* (1932), that it possesses four main characteristics: (1) it is historical writing without cognitive analysis; (2) it incorporates a sceptical belief that history cannot lead us to certain judgments; (3) it claims that scientific knowledge is only possible within its own boundaries; and (4) it attempts to reach philosophical and theological truths through an immanent study of history. In contrast, George G. Iggers (1969) has argued that historicism possesses three chief characteristics: (1) the tendency to view the nation-state as an end in itself; (2) a hostility to ethical normativity, that is, the imputation of value to all historical information; and (3) a hostility to conceptuality in historical analysis, that is, an emphasis on intuitive or non-rational means of understanding. Similar points have been made more positively by Friedrich Meinecke in his *Historicism* (1936), the most significant twentieth-century defence of historicism.

These observations, despite clear distinctions, concur in suggesting a definition of historicism as a radical shift away from Enlightenment understandings of history. All of these traits can be viewed as counterpoints to key assumptions of Enlightenment thought: primarily to the identification of personal and national interest with higher morality (natural law) and to the faith in universal norms, both ethical and cognitive (see Enlightenment, Continental).

The vitalist, interested approach which both Heussi and Iggers attribute to historicist thought also derives from a critique of the Enlightenment. It rests on the following assumptions: that humanity has no nature, only history: that is, that experience and cognition are conditioned solely by history; that laws determining human life are not naturally prescribed, but are the products of specific historical contexts; and that the truth-content of cognition is

dependent not on categorial logic, but upon its situatedness in, and constant attentiveness to, history. Implicit in this is the suggestion that the historian and the object of historical study are situated in the same totality of experience and shared history. Historicism thus replaces epistemology and normativity with a notion of a shared, historically shaped, experiential horizon, through which all knowledge must be processed. It has often been claimed, not without justification, that this conception lends itself readily to specifically national historiography, as the experiential totality of history can be easily narrowed to national history (Iggers 1969: 272).

These assumptions have been variously expressed. For example, J.G. Droysen (1858) argues that historian and historical personality both constitute totalities of experience, and that through their interaction the historian may reconstruct the totality of the historical personality through an understanding of surviving expressions and data. In this conception, the norms of historical analysis and value are derived from a process of immanent critique. The immanence of this process is underscored by the assumption of the shared human, expressly not natural, quality of history and historian.

The vitalist or essentialist undercurrent in this historiography is articulated as a hermeneutic procedure by Wilhelm Dilthey (§3) in *Introduction to the Human Sciences* (1883). Here the moment of historical perception is called *Erlebnis* (experience). In this process, the task of the historian is the actualization of the general quality of a particular history through experiential communication with it. For Dilthey, the boundless stream of life is objectified in all historical personalities and artefacts, and can be reactivated by, and for, the present experience and interest of the historian. Historicism thus treats its reflections as context-embedded insights, constantly to be reworked in a cultural synthesis for the historian. The vitalist subjectivism of this construct resurfaces in Georg Simmel's *Problems of the Philosophy of History* (1892) (see Simmel, G.) and, in the context of English thought, in R.G. Collingwood's *The Idea of History* (1946) (see Collingwood, R.G.). For Collingwood, history and narrative literature have near identity: 'Both the novel and the history are self-explanation, self-justifying, the product of an autonomous or self-authorizing activity; and in both cases this activity is the a priori imagination' (Collingwood 1946: 246). Collingwood's 'a priori imagination' or 're-enactment' echoes the historical thought of Croce, in which historiography and aesthetics are also linked: 'Every serious history, and every serious philosophy, ought to be a history and a philosophy "for the occasion", as Goethe said of genuine poetry, though the occasion of poetry is in the passions, that of history in the conduct of life and in morality' (Collingwood 1946: 36).

It is in this radical historicization of knowledge that historicism has achieved its most durable philosophical impact. In the post-historicist thought of Martin Heidegger and Hans-Georg Gadamer (§3), the analytical complexities of historicism are transposed into an ontology of historicity in which the earlier methods of *Verstehen* (understanding) and *Erlebnis* (experience) are assimilated into a universal hermeneutic process, structured around language.

2 Totality and particularity

Particularly in the English-speaking world, historicism is better known for various critiques of its methods than for its own character. However, it is notable that the major critiques of historicism address significantly conflicting ideas of what historicism may mean. Indeed, these critiques are often articulated from positions themselves associated with historicism.

Perhaps the most influential of all critiques of historicism, Nietzsche's *Untimely Meditations* (1876) (see Nietzsche, F.), attacks historicism as an unselective objectivizing reading of the past which is detrimental to the vital impulses of the present. Walter Benjamin, particularly in his 'Theses on the Philosophy of History' in *Illuminations* (1940), rejects historicism as mere reproduction of the oppressive forces of homogenous (capitalist) history. Both these readings clearly see in historicism an early form of historical empiricism. Both, however, launch their attack from within the vitalist tradition of historicist thought.

In *Die Probleme der Geschichtsphilosophie (The Problems of the Philosophy of History)* (1924), the Neo-Kantian Henrich Rickert condemns historicism as 'nihilism' for its supposed lack of ethical values and norms. Karl Löwith, in *Meaning in History* (1949), denounces historicism as a secular product of occidental eschatology. Both these critiques attack radically different notions of historicism. Both however - Rickert in his attempt to ground historical hermeneutics, Löwith in his opposition to redemptive history - are rooted in a territory which historicism, in various guises, has already claimed. Even the theologian Ernst Troeltsch (1922), who proposes a

reading of history underpinned by the supremacy of the evangelical church as a counterpoint to ethical relativism in historicism, remains caught within that which he seeks to undermine.

The most famous critique of historicism in twentieth-century philosophy, Karl Popper's *The Poverty of Historicism* (1957), attacks a construct of historicism overshadowed by the conception that German historical writing, from Hegel through Marx to the 1930s, is sworn to totality (and totalitarianism). Like Löwith, although for wholly different reasons, Popper links historicism with theological, 'utopianist' conceptions of history. In his critique, Popper asserts a theory of historical causality which concentrates on the deduction of specific historical results from specific conditions, and is not reliant on any claim to totalizing perspectives on historical trends or broad-ranging experiential tendencies. Rather, the theory attempts to reintegrate the methodologies of the human and natural sciences. This critique is in fact a reiteration of two different repudiations of historicism, both common in the nineteenth century, on account of its supposed totalizing intention and its supposed historical relativism. Popper's argument culminates in the following: 'I wish to defend the view, so often attacked as old-fashioned by historicists, that history is characterized by its interest in actual, singular, or specific events, rather than in laws or generalizations' (1957: 143).

While Popper's critique certainly proved influential in the English-speaking world, and undoubtedly addressed key general problems of historiography, it is difficult to view this position as a comprehensive investigation into the character of historicism. Popper's analysis has two key deficiencies: (1) it focuses on a tradition of totalizing thought to which much historicism is expressly opposed; (2) it is itself bound into a framework of individualizing empiricism which certain traditions of historicist thought have always sought to make their own. Paradoxically, Popper condemns in historicism precisely what Nietzsche suggests as an alternative to it (see Popper, K. §4).

This diversity both in the defence and critique of historicism can clearly be seen to derive from an aporia implicit in the term itself. The terrain covered by historicism in its protean changes not only evades definition, it encompasses the ground from which antidotes to historicism can be administered. However, it is helpful to understand this aporia as conditioned by the general development of historical thought in the nineteenth century. Indeed, the tensions and contradictions within both pro- and anti-historicist thought can be viewed as energies released by nineteenth-century critique of the Enlightenment. As historicism is articulated as a loose corrective both to the perceived neglect of Enlightenment thought for the non-rational life of the individual subject and to its schematic way of dealing with particular objectivity, it is hardly surprising that historicism should fluctuate between (or seek to reconcile) the extremes of the empiricist particularism and the totalizing subjectivism of early vitalism - both positions seek to guarantee a critical perspective on the Enlightenment. It is within the tensions of critical engagement with the Enlightenment that historicism can be best understood, if not synthetically defined.

See also: History, philosophy of

CHRISTOPHER THORNHILL

References and further reading

Benjamin, W. (1940) 'Theses on the Philosophy of History', in *Illuminations*, London: Cape, 1970.(Recently established as a major work on historical philosophy.)

Collingwood, R.G. (1946) *The Idea of History*, Oxford: Clarendon Press.(Major work in the tradition of English historicism.)

Croce, B. (1938) *History as the Story of Liberty*, London: Allen & Unwin, 1941.(Major work in the tradition of Italian historicism.)

D'Amico, R. (1989) *Historicism and Knowledge*, New York and London: Routledge.(Excellent for its critique of Popper.)

Dilthey, W. (1883) *Introduction to the Human Sciences*, ed. R.A. Makkrees and F. Rodi, Princeton, NJ, and Oxford: Princeton University Press, 1989.(Seminal work in the tradition of philosophical historicism; makes links with hermeneutics.)

Droysen, J.G. (1858) *Historik: Vorlesungen über Enzyklopädie und Methodologie der Geschichte (Historics: Lectures on Encyclopedism and the Methodology of History)*, Munich: Oldenbourg. (Liberal historicism; closely associated with early German parliamentarianism.)

Engel-Jánosi, F. (1944) *The Growth of German Historicism*, Johns Hopkins University studies in historical and political sciences, series 62, no. 2, Baltimore, MD: Johns Hopkins University Press. (A clear overview.)

Gumley, J.E. (1989) *History and Totality: Radical Historicism from Hegel to Foucault*, London and New York:

Routledge.(Interesting scrutiny of emancipatory historicism; imprecise on the definition of historicism.)

Hegel, G.W.H. (1826) *Lectures on the Philosophy of World History*, Cambridge: Cambridge University Press, 1975.(Elaborates history as dialectical teleology; weakest work of a great philosopher.)

Herder, J.G. (1784) *Outlines of a Philosophy of the History of Man*, London: Johnson, 1800. (Seminal work of pre-Hegelian philosophy of history; especially influential in nineteenth-century national theory.)

Heussi, K. (1932) *Die Krisis des Historismus (The Crisis of Historicism)*, Tübingen: Mohr. (Influential critique; not available in translation.)

Iggers, G.G. (1969) *The German Conception of History: The National Tradition of Historical Thought from Herder to the Present*, Middletown, CT: Wesleyan University Press.(Indispensable critique, although rather weak on emancipatory dimensions to historicism.)

Löwith, K. (1949) *Meaning and History: The Theological Implications of the Philosophy of History*, Chicago, IL: Chicago University Press.(Outstanding and original work of critical philosophy.)

Meinecke, F. (1936) *Historicism: The Rise of a New Historical Outlook*, London: Routledge, 1972. (The most extensive defence of historicism ever written.)

Nietzsche, F. (1876) *Untimely Meditations*, Cambridge: Cambridge University Press, 1983.(Most famous of all critiques of historicism; criticizes historicism as the debilitation of the present.)

Popper, K. (1957) *The Poverty of Historicism*, London: Routledge.(Enormously influential, but theoretically misinformed reading of historicism.)

Reill, P.H. (1975) *The German Enlightenment and the Rise of Historicism*, Berkeley and Los Angeles, CA, and London: University of California Press.(Illuminating survey of earliest historicist writings.)

Rickert, H. (1924) *Die Probleme der Geschichtsphilosophie (The Problems of Historical Philosophy)*, Heidelberg: Winter.(Neo-Kantian theory of concept formation, applied to history.)

Schnädelbach, H. (1974) *Geschichtsphilosophie nach Hegel: Die Probleme des Historismus (Historical Philosophy After Hegel: The Problems of Historicism)*, Freiburg and Munich: Karl Alber. (Excellent analysis of the development of nineteenth-century historiography.)

Simmel, G. (1892) *The Problems of the Philosophy of History: An Epistemiological Essay*, New York: Free Press, 1977.(Relatively minor work of a great sociologist.)

Troeltsch, E. (1922) *Der Historismus und seine Probleme (Historicism and its Problems)*, Tübingen: Mohr.(Most painstaking account of historicism. Extremely illuminating exemplification of difficulties experienced by early twentieth-century thinkers in their attempts to overcome historicism; not available in English.)

History, Chinese theories of

The beginnings of Chinese historical writing can be seen in the works of several early thinkers of the sixth through third centuries BC. *The various features which we see in their nascent form in early classical sources were developed, synthesized and found their first mature expression in the composition of what came to be known as the Standard or Official Histories, notably the Spring and Autumn Annals, the Shiji (Records of the Historian) and the Hanshu (History of the Former Han Dynasty).*

Much has been made of the purported 'cyclical' nature of Chinese views of time and the implications this has had on everything from Chinese views of history to the development of science. It is alleged that there was no notion of historical progress among the Chinese, who purportedly held a fatalistic view of infinitely repeating cycles of alternating political order and chaos, beyond human control. However, Chinese thinkers have held a wide range of different views about the pattern and flow of history.

While the amount of historical writing in China is truly staggering, Chinese thinkers have paid relatively little attention to the systematic study of the methodology and nature of history. However, there have been notable exceptions in both critical and speculative history. Critical historians at different times modified and made further suggestions regarding the form and content of the Official Histories. They also criticized the methods by which material for the Official Histories was gathered and controlled, and offered a variety of opinions regarding other forms of historical writing. Chinese philosophers of history developed elegant and sophisticated theories of history which purport to reveal the significance and structure of historical processes.

The nineteenth century marks the beginning of modern Chinese views of history, when Chinese intellectuals were first deeply influenced by Western views on the nature of history and historical method. This was also a time when Chinese society was shaken to its foundations by Western colonialism. Chinese historians of this period responded with impressive syntheses of traditional and western views. Traditional notions - for example the idea that history must fulfil a moral purpose - were recast in terms of the emerging phenomenon of Chinese nationalism. At the same time, new ideas about how objective factors, such as economic forces, shape historical events and the global scale on which these events take shape were incorporated into novel Chinese conceptions of history. This process continued into the early twentieth century when Marxist views became most influential. Early Chinese Marxist theories of history reveal a remarkable mixture of the strange and the familiar. However, Stalinist ideology quickly came to dominate more subtle forms of Marxism. As a result, Chinese Marxist views of history soon became ossified and uninspired. Little new or interesting can be found after the early 1950s.

1 Views of history in the pre-Qin period

The Chinese have shown a remarkable reverence for their past and have expended tremendous effort to record and preserve their history. This intense interest in history may have arisen, at least in part, from the practice of ancestor worship. The early Chinese sought guidance from their ancestors by divining with shell and bone, and engraved their queries and the received responses upon these materials. K.C. Chang (1983: 90) suggests that these inscriptions may have been saved for future reference, thus contributing to the notion that a *record* of past events can serve as a *guide* to proper future action, a central idea in Chinese philosophy of history. E.G. Pulleyblank (Dawson 1964: 144) has noted that ancestor worship required the maintenance of detailed genealogies and that the need to announce important state decisions to royal ancestors may have led to the keeping of state chronicles, one of the earliest forms of history found in China. These chronicles were maintained by court officials called *shi* (scribes), the same character that came to mean 'historian' and 'history'.

These early 'historians' were renowned for unswerving dedication to their calling. Several are said to have died insisting, against the wishes of their lords, that events be fully and accurately set down. These early paragons also purportedly embraced the notion that the historian's ultimate goal is to ensure the just distribution of 'praise' and 'blame'. The deeds of worthy individuals must be preserved, assuring them a kind of immortality and affording uplifting examples to later generations. Despicable acts too must be recorded and their agents held up for perpetual contempt and as a warning to future generations. Chinese history was thus seen as an accurate record of the intentions and actions of human beings, primarily those prominent in the central government, which served as a witness to good and bad deeds and a guide to future generations. These ideas about the proper role of history

explain why biography played such a prominent role in Chinese historical writing, and why these biographies largely concern issues of character.

Confucius believed in a Golden Age brought into being by a series of world-ordering sages, from which subsequent ages had devolved and to which his own age needed to return (see Confucian philosophy, Chinese). He saw his personal mission as preserving and restoring the beliefs, practices and institutions of this past age (see *Analects* 9.5). He himself was merely the messenger, someone who ‘transmits but does not create’ (*Analects* 7.1). In these respects, he can be seen as a figure who epitomizes the ideal of the historian as custodian of a record of ethical guidance, handed down from sagely ancestors.

The Mohists, chief rivals of the Confucians, criticized the traditionalism of the former (see Mohist philosophy). In a chapter entitled ‘Against Confucians’, a Mohist points out that the Confucian teaching of ‘transmitting but not creating’ is self-contradictory, for the divine sages who *created* the culture the Confucians claim to preserve and revere did not themselves practice this ideal. However, the Mohists too appealed to their own version of an idealized past. While they recognized and advocated technological progress, they believed that the proper forms and institutions of society had been discovered by past sages, and these simply needed to be revived.

A third representative of this general belief in a lost Golden Age is the later Daoist thinker Laozi, purported author of the *Daodejing* (see Daodejing). In Chapter 38 of this work, Laozi describes the devolution of the *dao* (way) (see Dao). Human history is the story of the rise of ethical self-consciousness, with a corresponding and proportional loss of innocence and simplicity, resulting in chaos and strife. While the Daoist Golden Age is very different from that of either Confucians or Mohists, the goal - to return to it - is the same. However, unlike the latter, Laozi makes no appeal to written records. In the ideal state, people do not employ writing at all; they keep their accounts by use of ‘knotted cords’. For Laozi, the Way lies in certain pre-rational intuitions that cannot be captured in words. During the time of the Golden Age, people embraced *pu* (unadulterated simplicity), were *wuzhi* (without knowledge) and *wuwei* (without action), living spontaneous lives in accord with natural patterns and tendencies (see Daoist philosophy). The revival of such a state would seem to entail the end of ‘history’.

Another major thinker of this period is the Confucian Mencius, who presents an interesting collection of views regarding history. While he is a strong defender of the Confucian vision, unlike Confucius, he relies more on intuition than tradition and expresses scepticism regarding even revered historical sources: ‘It would be better not to have the *Book of Documents* than to completely trust in it’ (*Mengzi* 7B3). Mencius also seems to have believed in an alternating five-hundred-year cycle of order and chaos (*Mengzi* 2B13). However, it is not clear the degree to which he believed this was an immutable pattern or simply an account of the interval between traditional sage-kings.

A contemporary Daoist thinker, Zhuangzi, recognized that different times and circumstances employ and indeed require different actions and policies. Things ebb and flow in human history but lack any set pattern or structure. Zhuangzi concluded that there was no discernable pattern in history, devolutionary, cyclical or progressive. The one thing one could learn from history is that to insist on patterning oneself on the ways of past sages is a recipe for failure. Thus the records of even the most gifted sages are nothing but their ‘chaff and dregs’.

The next great Confucian of this period, Xunzi, strongly criticized Mencius for believing that human nature contains innate moral tendencies. Xunzi saw this belief as not only wrong but also a threat to the status of the traditional wisdom of the founding Confucian sages. Xunzi regarded the insights of these sages as discoveries that had been achieved through a long and difficult process of experimentation. The founding sages had discovered how to organize society just as potters had discovered how to produce exquisite pottery, through a process of trial and error. Xunzi thus strongly implies an early period in which the *dao* evolved through successive stages until it resulted in the Golden Age whose plan Confucius preserved.

One of Xunzi’s disciples, Han Feizi, ultimately broke with the Confucian tradition, a move which required, and perhaps was in part caused by, his rejection of the Confucian view of history. Han Feizi argued that human history was progressive, moving from more primitive to more advanced stages. He mocked the Confucian call to study the ‘ways of the ancients’ and used historical examples primarily to illustrate the folly of such an attempt (*Han Feizi* 49). Han Feizi did show a strong sympathy with the Daoist belief in a primitive past utopia, but rejected the idea that this state of affairs could be revived. The world had become nastier since then: people were now too numerous

and goods too scarce. All that remained were selfish human desires, which could only be controlled through a system of strict rewards and punishments (see Legalist philosophy, Chinese; Law and ritual in Chinese Philosophy). Hence, while technology and certain social institutions would need continually to evolve, the evolution of human beings had come to a halt after an initial fall from innocence.

2 The Official Histories

The *Spring and Autumn Annals*, along with its various commentaries, served as the earliest models for Chinese historical writing. The text itself is taken to be the court record of Confucius' native state, covering the years 722-481 BC. Confucius purportedly edited it in order to insure that 'praise' and 'blame' were properly assigned. This and another, later classic, the *Shujing (Book of Documents)*, existed in some form during the early period of Chinese philosophy (see §1). Traditional accounts describe the *Shujing* as a 'record of words' - in other words, official proclamations - in contrast to the *Spring and Autumn Annals* which was 'a record of deeds', that is, the actions of officials. The form and theory of the Official Histories was greatly influenced by these two texts and early views regarding them.

A major development with profound and enduring consequences for Chinese views of history and methods of historiography was the writing of the *Shiji (Records of the Historian)*. Its author, Sima Qian (145?-90? BC), completed the work begun by his father, Sima Tan. Both had held the office of *taishi* (grand scribe) but their project was a private, family enterprise. The *Shiji* was a world history, seeking to provide an account of all of Chinese history down to the time of its composition. It presents the past primarily in terms of the lives of important families and individuals, largely though not exclusively those involved directly in the government, relying upon vivid and lively descriptions of their actions and intentions. Its narrative style is designed to allow historical events to speak for themselves; however, the author clearly embraced the 'praise' and 'blame' model of historical writing and believed strongly in the didactic value of history.

The composition of the *Shiji* was followed by another family production when Ban Gu (AD 32-92) completed the *Hanshu (History of the Former Han Dynasty)*. Ban Gu laboured to complete a project begun by his father, and his sister, Ban Zhao, helped to complete the work after his own death. While its structure is largely based upon the *Shiji*, Ban Gu's work differed in being a history of a single dynasty. This affected not only the structure of the work but the view that history should be seen as a unified whole: history now was presented in terms of the rise and fall of a single dynasty. Another difference between this later work and the *Shiji* is that while both began as private family projects, Ban Gu's work was co-opted by the government and received official support and sanction in its later stages. Thus it was the first truly official history, creating a model which began a series of later works that continued until the end of the imperial period in 1911.

While the different official histories vary in detail, they share a remarkable number of general features. As mentioned above, with the exception of the *Shiji*, all concern a single dynastic period; taken together, they present history as a series or, in some conceptions, a repeating cycle of imperial reigns. All rely, almost exclusively, on literary sources to construct their account of the past and employ a cut-and-paste method to arrange these materials into a coherent narrative. Again with the exception of the *Shiji*, all the official histories are retrospective, written by people from succeeding dynasties. By the time of the Tang dynasty (618-906) institutions were in place whereby the reigning dynasty prepared material intended to serve as the primary source for their own future history. Also beginning with the Tang, with the single exception of the *Xin Wudai shi (New History of the Five Dynasties)* by Ouyang Xiu (1007-72), official histories were the product of a team of historians assigned by the government to complete the task.

It was not long before historical writings came to be seen as a distinct and important genre. In the fourfold classification that became standard in the third century, history is ranked immediately after the classics, coming before both philosophy and *belles lettres*. It might seem strange that history was ranked ahead of philosophy, since historical writing owed much of its esteem to its ethical value. However, history, being a record of actual conduct was generally regarded as superior to philosophy, which was by its very nature abstract and speculative. This view was widely held throughout the tradition and precedents for it can be found in texts as early as the the *Zhongyong* (see Zhongyong) and the *Shiji*.

3 Chinese views of historical pattern and flow

We have seen that early Chinese thinkers expressed a wide range of different views concerning the pattern and direction of history (see §1). Some saw history as a repeating cycle of periods of human flourishing and decline. Others believed that history had evolved in earlier periods but that in at least certain important respects - specifically in terms of social institutions and practices - it had reached a stage of completion. Yet other thinkers denied that there was any discernable pattern or rhythm to the flow of history. The most widely held view was that history was a process of devolution from a past Golden Age to present turbid times. The exact nature of the lost Golden Age and the causes of its decline varied widely among different thinkers, but those who advocated such views all advanced an accompanying appeal to revive past institutions and practices in order to return to the Golden Age. This idea was further developed during the succeeding Han dynasty and found clear expression in texts like the *Liji (Book of Rites)*, particularly in the chapter *Liyun (Evolution of the Rites)* (see Chinese Classics), and the *Huainanzi*, particularly Chapter 2 (see Huainanzi). It is important to note that while such views hoped for a revival of the Golden Age, they also clearly entailed the idea that one can fall out of such a state as well as regain it. In other words, these were not millennial movements that saw an end to history in some final, ideal utopia. Some Chinese thinkers did express views like this: for example, the popular Daoist uprisings of the Red Eyebrows in the first century AD and the Yellow Turbans in the second century AD and the Taiping movement of 1851-64 were all founded on such a belief. The modern thinker Kang Youwei (1858-1927) also espoused a similar view (see §5).

Those who believed in a past Golden Age, as well as those who tended to see this age only as a future possibility, often described history in terms of set phases either devolving from or evolving into an age of good order, prosperity and peace. In some cases, such phases were seen as repeating historical cycles. Mencius may have held a view like this and the Han Confucian Dong Zhongshu, possibly developing ideas of an earlier thinker named Zou Yan (*fl.* 4th century BC), described history as a repeating cycle of five elemental phases: wood, fire, earth, metal and water, each of which exerts the dominant influence in their respective age.

Under the influence of Buddhism, later Chinese thinkers developed elaborate cosmic histories (see Buddhist philosophy, Chinese). For example, Shao Yong presented a complex scheme which describes history in terms of the coming into being of the universe, the emergence of living things, the flourishing of human society, the eventual decline of society and human beings, the ultimate end of all life and the extinction of this world system. The process then begins anew. Zhu Xi espoused a similar though much less articulated view, believing that the universe arose out of and would eventually devolve back into a state of undifferentiated *qi*, out of which a new universe would gradually coalesce and take shape (see Qi).

4 Theories of history in traditional China

Liu Zhiji (661-721) composed the first treatise on the writing of history in China, the *Shitong (Survey on History)*. Liu embraced most traditional notions concerning the purpose of historical writing and the idea that history should be viewed in terms of separate dynastic units. However, he criticized the degree to which the bureaucratization of the process of producing the Official Histories had come to influence their accuracy and objectivity. He also offered specific recommendations regarding the form and content of the Official Histories, suggesting that some of the traditional *zhi* (monographs) be dropped and others added. For example, he argued that the monographs on astronomy be eliminated since they concerned a topic that was timeless and hence inappropriate for inclusion in a historical work. He also argued that a new monograph on cities be added, thus expanding the range of material that was to be considered in the official histories. His ideas regarding the incorporation of local material in the writing of the Official Histories and his view that a historian had to possess the specific virtues of insight, literary skill and learning in order to perform his task, influenced Zhang Xuecheng (see below).

While Liu's *Shitong* was a treatise on the theory of historiography, the *Zizhi tongjian (A Comprehensive Mirror for Aid in Government)* of Sima Guang (1019-86) was a world history in chronicle form, covering the period 403 BC-AD 959. In the process of composing this great work, Sima Guang developed and refined his own ideas on the nature and practice of historical writing. As the title of the work suggests, the *Zizhi tongjian* had a very traditional goal: to provide a collection of historical examples of good and bad government that could serve as a resource for contemporary rulers. Sima Guang drew upon a wide variety of different sources for this work, going well beyond those normally used in composing the Official Histories. He began with a vast collection of material organized by period, which he then whittled down in successive drafts. One of his most significant methodological

developments was his effort to verify the authenticity of his sources and the events they recorded. He carefully examined the texts he drew upon and subjected them to strict philological and textual analysis. He would not accept the validity of any purported event if it conflicted with objective facts - such as a firmly established date - or if it failed to cohere with reliable collateral evidence and the dictates of common sense. He developed and maintained the rigorous critical standards of a scientific historian.

The greatest Chinese philosopher of history is undoubtedly Zhang Xuecheng (1738-1801). He made significant contributions to historiography; for example, he wrote and promoted *fangzhi* (local gazetteers) as an important historical genre. However, as he said of himself, his greatest insights concerned not the method of history but its significance. Zhang is renowned for his claim that, 'the six classics are all history'. By this he meant several things, all related to his grand speculative theories concerning history.

Like Hegel, Zhang believed history went through distinct and necessary stages of development. The *dao* (way) came into being as a response to historical necessities. However, unlike Hegel, Zhang believed that the process of historical development - specifically the evolution of social institutions and practices - had reached its highest point in the distant past. The *dao* had attained full form in the Golden Age of the Duke of Zhou. At that point in time, society had become a perfect and harmonious whole in which what we now regard as the separate functions of officials and teachers were united; there were no individual disciplines or schools of learning and all writing was anonymous and *gong* (public), being in the service of some branch of government. Later history has meandered around this lost ideal, with successive ages dominated by an excess of 'philosophy', 'philology' or 'literature'.

Confucius is unique in having been born at the historical moment when the *dao* crested and began to decline. He realized that his historically determined mission was to preserve a record of this ideal society. Later ages wrongly regard the texts he preserved as 'records of the *dao*', but in fact they are simply the records of various official agencies of this ideal time. The *dao* itself cannot be seen or described, only its traces, like the tracks of an invisible passing wheel (see Daoist philosophy §2; Dao). The *dao* takes shape and is revealed in the playing out of historical processes and so the classics, like all texts, are 'history'.

People of later ages also fail to see that their historical situations differ from that of Confucius and mistakenly attempt to emulate him by writing books about the *dao* instead of doing whatever their own historical moment requires. Their true task lies in discerning which of the three errant tendencies is in ascendance in their own age and working to counteract it. In order to succeed at this task, one must be able to read events in the world properly. This requires one to be a certain *kind* of historian. In addition to the three virtues described by Liu Zhiji, one needs to have the *virtue* of a historian: a kind of faculty or sense that allows one to discern how historical events fit into the grand pattern of the *dao*. Zhang's philosophy of history raises history to preeminent status: it is the way - the only true way - to understand the dao.

5 Modern views of history

The story of modern Chinese views of history remains incomplete. However, its opening chapter was written by Kang Youwei (1858-1927). Kang was both a highly respected traditional scholar and one of the first Chinese intellectuals to have a firm grasp and appreciation of Western ideas regarding the nature of history and historical method. This combination led him to propose a revolutionary view of history and in particular, Confucius' role in it. In his *Datongshu (Book of Great Unity)*, Kang rejected the traditional notion of a past Golden Age and the idea that Confucius had merely 'transmitted' its *dao* to later ages. Drawing on and modifying earlier Chinese views of history, such as those found in the *Liyun (Evolution of the Rites)* chapter of the *Liji* (see §2) and the *Gongyang* commentary to the *Spring and Autumn Annals* (see Chinese Classics), and combining these with ideas taken from Buddhism, Christianity and Confucianism, Kang developed a highly original speculative philosophy of history. His view is based upon his theory that the history of the world consists of 'three ages': the past age of Chaos, the present age of Approaching Harmony and future age of Great Harmony. Kang attempted to trace this idea back to early Chinese sources. However, while certain similar ideas can be found in early texts, particularly among the writings of Dong Zhongshu, Kang differed in seeing history as evolutionary and not cyclical. His theory of the three ages described a philosophy of historical progress culminating in a future grand utopia; the Golden Age lay not behind but ahead of us. Moreover, Kang claimed that this was Confucius' real view as well and argued that Confucius, aware of how revolutionary his message was, had cleverly presented his vision of the future as a form

of traditionalism.

Kang’s student Liang Qichao (1873-1929) eventually distanced himself from most of his teacher’s views of history. Liang turned to less speculative projects and wrote the *Zhongguo lishi yanjiu fa (Methodology for the Study of Chinese History)*. This work of practical, critical history attacked traditional Chinese historical method from a western perspective. However, in doing so it relied heavily on earlier Chinese critics of traditional methods, several of which have been discussed above.

Neither Kang nor Liang broke free from the traditional idea that history was to serve a higher moral purpose. Both were trying to relate their philosophies of history to the immediate and pressing concerns of contemporary Chinese nationalism. This led them and their successors to a broader examination of Chinese culture and a much greater appreciation of its variety and richness, particularly its more popular forms, which traditionally had been almost wholly neglected. This development had a profound effect on thinkers like Hu Shi (1891-1962) and Lu Xun (1881-1936) who advocated and contributed to the history and promulgation of popular literature.

The rise and eventual dominance of Chinese Marxism both contributed to and later constricted the development of modern Chinese views of history (see Marxism, Chinese). The Marxist explanation of history as the playing out of economic forces across vast spans of time, helped to introduce new ideas about the role of objective and enduring factors in human history. This broadened the scope and range of Chinese conceptions of history and worked to break down the dynastic view. It also encouraged a more synthetic approach to historical problems. However, cruder forms of Marxist historicism became dominant and the older dynastic view came to be replaced by an equally Procrustean scheme of ‘slave’ and ‘feudal’ periods. Debates then raged as to whether Confucius had lived during a ‘slave’ or ‘feudal’ period. A great deal was seen to turn on this issue: since Confucius defended the ‘feudal system’, if he lived in a slave period he should be viewed as a progressive, someone moving history forward on the ‘inevitable’ course toward its communist future. If however, it turned out that Confucius lived in a feudal period, then he should be viewed as a reactionary, someone holding back the future Golden Age.

In many respects, the Chinese Marxist view of history is both strange and familiar. Its moral temper and utopian nature are familiar features of traditional Chinese conceptions of history; and while its particular content is surely new, its general form and the trajectory it describes can be found in several of the sources and thinkers discussed above. The persistence of these traditional themes suggests that the story of modern Chinese views of history is still being written.

See also: Chinese Classics; Chinese philosophy; Confucian philosophy, Chinese; Daoist philosophy; History, philosophy of; Law and ritual in Chinese philosophy; Legalist philosophy, Chinese; Marxism, Chinese; Mencius; Xunzi; Zhuangzi

PHILIP J. IVANHOE

References and further reading

Beasley, W.G. and Pulleyblank, E.G. (eds) (1962) *Historians of China and Japan*, London: Oxford University Press.(The best anthology available. In particular, see the contributions of van der Loon, Pulleyblank, Demieville and Gray.)

Chang, K.C. (1983) *Art, Myth and Ritual: The Path to Political Authority in Ancient China*, Cambridge, MA: Harvard University Press.(A study of very early Chinese culture from an archaeological perspective.)

Dawson, R. (ed.) (1964) *The Legacy of China*, Oxford: Clarendon Press, 1964. (Solid introduction to Chinese culture.)

Feuerwerker, A. (ed.) (1968) *History in Communist China*, Cambridge, MA: MIT Press.(A collection of articles on various aspects of Marxist historiography.)

Hartwell, R.M. (1971) ‘Historical Analogism, Public Policy and Social Science in Eleventh and Twelfth-Century China’, *American Historical Review* 76 (3): 690-727. (On the role reasoning from historical analogy and precedent played in the formation of governmental policy and how this affected the development of a social scientific approach.)

Grieder, J.B. (1970) *Hu Shih and the Chinese Renaissance*, Cambridge, MA: Harvard University Press. (Intellectual biography of this important modern Chinese philosopher, social activist and intellectual.)

Kang Youwei (1902) *Datong shu (Book of Great Unity)*, trans. L.G. Thompson, *Ta t’ung shu: The One-world Philosophy of K’ang Yu-wei*, London: Allen & Unwin,1958.(The definitive statement of Kang’s utopian

vision.)

Levenson, J.R. (1953) *Liang Ch'i-ch'ao and the Mind of Modern China*. Cambridge, MA: Harvard University Press.(Intellectual biography of Liang Qichao, one of modern China's leading thinkers.)

Liji (Book of Rites) (2nd century BC-2nd century AD), trans. J. Legge, *Li Chi: Book of Rites*, New York: University Books, 1967, 2 vols.(The most important source for traditional ritual practice.)

Needham, J. (1965) *Time and Eastern Man*, Royal Anthropological Institute Occasional Paper 21, Glasgow: The University Press.(Thorough exploration of this topic.)

Nivison, D.S. (1966) *The Life and Thought of Chang Hsüeh-ch'eng (1738-1801)*, Stanford, CA: Stanford University Press.(The definitive work on Zhang Xuecheng and his thought.)

Shujing (Book of Documents) (5th-4th century BC) trans. J. Legge, *The Shoo King or The Book of Historical Documents*, vol. 3 of *The Chinese Classics*, Hong Kong: Hong Kong University Press, 1970.(Translation with facing Chinese text.)

Sima Qian (2nd-1st century BC) *Shiji (Records of the Historian)*, trans. B. Watson, *Records of the Grand Historian of China*, New York: Columbia University Press, 1961, 2 vols.(Selective translation of China's first official history.)

Watson, B. (1963) *Ssu-ma Ch'ien, Grand Historian of China*, New York: Columbia University Press.(A thorough study of the *Shiji* and its author, Sima Qian.)

Zuozhuan (Zuo Annals) (2nd-1st century BC), trans. J. Legge, *The Ch'un Ts'ew with the Tso Chuen*, vol. 5 of *The Chinese Classics*, Hong Kong: Hong Kong University Press, 1970.

History, philosophy of

Philosophy of history is the application of philosophical conceptions and analysis to history in both senses, the study of the past and the past itself. Like most branches of philosophy its intellectual origins are cloudy, but they lie in a refinement of 'sacred' histories, especially those of Judaism and Christianity. The first major philosopher to outline a scheme of world history was Immanuel Kant in The Idea of a Universal History from a Cosmopolitan Point of View (1784), and German Idealism also produced Hegel's Lectures on the Philosophy of World History (1837), a much longer and more ambitious attempt to make philosophical sense of the history of the world as a whole. According to Hegel, history is rational, the working out, in fact, of philosophical understanding itself.

The accelerating success of natural science in the nineteenth century gave rise to a powerful combination of empiricism and logical positivism which produced a philosophical climate highly unfavourable to Hegelian philosophy of history. The belief became widespread among philosophers that Hegel, and Marx after him, had developed a priori theories that ignored historical contingency in favour of historical necessity, and which were empirically unfalsifiable. Karl Popper's philosophy of science was especially influential in converting philosophy of history to a new concern with the methods of historical study rather than with the shape of the past. Two rival conceptions of historical method existed. One tried to model explanation in history on what they took to be the form of explanation in science, and argued for the existence of 'covering laws' by which historians connect the events they seek to explain. The other argued for a distinctive form of explanation in history, whose object was the meaning of human action and whose structure was narrative rather than deductive.

Neither side in this debate was able to claim a convincing victory, with the result that philosophers gradually lost interest in history and began to concern themselves more generally with the nature of human action. This interest, combined with a revival of nineteenth-century German hermeneutics, the study of texts in their social and cultural milieu, in turn revived interest among analytical philosophers in the writings of Hegel and Nietzsche. The impact of continental influences in philosophy, art criticism and social theory was considerable, and reintroduced a historical dimension that had been largely absent from twentieth-century analytical philosophy. In particular, the formation of fundamental philosophical ideas began to be studied as a historical process. The Enlightenment came to be seen as a crucial period in the development of philosophy, and of modernity more generally, and with this understanding came the belief that the contemporary Western world is postmodern. In this way, social theory and the philosophy of culture in fact returned, albeit unawares, to the 'grand narrative' tradition in philosophy of history.

1 Origins

> Is there a logic of history? Is there, beyond all the causal and incalculable elements of the separate events, something that we may call a metaphysical structure of historic humanity, something that is essentially independent of the outward forms - social, spiritual and political - which we see so clearly?… Does world-history present to the seeing eye certain grand traits, again and again, with sufficient constancy to justify certain conclusions?
>
> (Spengler 1918, vol. 1: 3)

So writes Oswald Spengler at the start of his monumental book *The Decline of the West* (1918). Spengler thought that the sort of interest in history which he proposed to take was novel, and to be contrasted with an interest in the 'elements of separate events'. This implies that hitherto (that is before Spengler) all history was concerned with the mere recording of what actually happened, in fact the desire to have a knowledge of the past entirely for its own sake and without practical or theoretical aims, is a relatively sophisticated, and relatively late, intellectual accomplishment. Most historical writing is characterized by a less detached view. Arguably all cultures which lend importance to an awareness of their own past have tried to ascertain within it some meaning, significance or purpose, to cast light upon present events and speculate successfully upon the future. Generally, interest in the past of this practical sort has prevailed over the more academic concern of the study of what has been called 'scientific history'. Scientific history here means the attempt simply to arrive at an accurate account of past events based upon sufficient evidence, without regard to learning lessons or predicting the future course of events. To say that a 'practical' interest in the past has overshadowed a more academic one is not to deny that traces of scientific

history can be found in the ancient world; there is reason to regard Herodotus as one of its earliest and in some ways finest practitioners, and it was this sort of detachment from ultimate meaning that led St Augustine (§15) to regard the 'pagan' history of his own time as seriously inadequate. But more strikingly, and even earlier in Western culture, we find a broader intellectual ambition that can be described as the attempt to look beneath the surface of events and find their inner or ultimate meaning.

The most familiar histories of this kind have been religious. A marked feature of Judaism, for instance, is a belief in the revelatory power of the past to disclose, and to confirm, a relationship between a people and their god. The Jewish religion, as its sacred books reveal, is coloured by the desire to formulate a sacred history. Many of the books of the Old Testament try to move behind such historical events as the Exodus from Egypt or the removal of the Jews to Babylon and detect in them the purposes of God. In Isaiah, for example (which is generally agreed to be the work of at least two authors), the first prophet's principal interest is in explaining the conquests, invasions and natural disasters to which Judea was subjected at one point in its history as the actions of a God highly displeased by the moral and religious waywardness of his chosen people and unmoved by the ritualistic sacrifices they made it their practice to offer (see Hermeneutics, biblical §1).

Jewish theology is sacred history *par excellence*, but in this it is not unique. It has provided a model on which other religions have sought to improve. This is most obviously true of Christianity, which picks out a special set of events from the general flow of the past as the reference points around which a story of divine intentions may be told. This set of events includes many of those from Jewish history, but adds the central event of the Incarnation of God in Jesus, a historical figure of first-century Palestine, through whose life and teaching God can be seen to act directly in a historical time. Among the Church Fathers it is St Augustine who first makes Christian sacred history explicit; indeed we owe the term 'sacred history' to him. But much of the general idea is present in early Christian writings. These make extensive use of historical typology - of patriarch, prophet and messiah, for example.

2 Philosophical history

The idea of a 'sacred history' has its origins in these two religions, and their prominence in Western culture has made its conceptions highly influential in the subsequent development of historical and cultural self-understanding. However, histories of this rather grand sort need not always be concerned with the purposes of God, or even with the invocation of supernatural agencies. Subsequent, secular, thinkers have taken up the same ambition, but transformed it in essentially non-religious ways. The intellectual Enlightenment in the eighteenth century was a considerable stimulus to this. Among the writers who began this process the most notable were J.G. Herder (§4), author of *Outlines of a Philosophy of the History of Man* (1800), and the Marquis de Condorcet, who wrote a *Sketch for an Historical Picture of the Progress of the Human Mind* (1795). Immanuel Kant was the first major philosopher to formulate explicitly *'The Idea of a Universal History from a Cosmopolitan Point of View'* (1784). In this essay Kant maintains that though the empirical study of historical phenomena is essential to understanding the human past, in itself it is insufficient, since the plans of individuals, however rational, cannot on their own explain the emergence of developing forms of thought and civil association. There are, Kant contends, general trends to be observed in the history of humankind. These cannot plausibly be attributed to the purposes of individuals, even in concert, and thus any serious intellectual concern with the past has to supply a different explanation. Their explanation, Kant believed, must impute to the historical process a natural teleology, a purpose or end which nature or providence (he uses both terms) develops through the rational agency of human beings.

Kant's essay is merely an outline of a possible form of inquiry, and in it he looks to the emergence of an equivalent in this sphere of a Kepler or a Newton in physical science, someone intellectually equipped to compose a genuinely universal history which will reveal the underlying rationale of the past in the way that Kepler and Newton uncovered the underlying rationality of the physical universe. Such a figure as Kant anticipates might be thought to have been found in his celebrated successor G.W.F. Hegel (§8), whose *Lectures on the Philosophy of World History* (1837) probably constitute the most sustained effort ever made in this aspect of philosophy. Though Hegel believed himself to be providing a rational exposition of the truths obscurely contained in Christianity, the outcome of his reflections on 'world history' was one in which theological categories were systematically replaced by philosophical conceptions. Thus the work of divine agency is in Hegel converted into the activity of 'reason' itself, and the end of history is not the kingdom of God but the manifestation of absolute mind or spirit.

There remain some mystical elements in Hegel's view, and he himself did not intend that the religious dimension

should be eliminated; it was, rather, to be rationalized by philosophy. But even if Hegel retains some of the essential characteristics of sacred history, during the course of the nineteenth century successor histories of the same 'grand' sort emerged which had no place for religious meaning or the divine. The best known of these is that of Karl Marx (§8). Set out in brief in the *Communist Manifesto* (1848), co-authored with Friedrich Engels, and elaborated at much greater length in the unfinished three-volume work *Capital* (vol. 1, 1867), Marx's theory of history self-consciously abandons religious aspirations as products of false-consciousness and seeks their replacement with purely material (largely economic) alternatives. Marx does not invoke divine intentions but rather historical laws governing economic change - from feudal to capitalist to socialist - and the communistic classless society to which he looks as the future ideal is not a world constituted and governed by universal acknowledgement of divine rule, but one waiting to be created by imaginative human endeavour freed from the constraints of economic necessity.

Marx is not alone with his secular alternative to sacred history. Better known in their own day were those of Auguste Comte and Herbert Spencer. For Comte, regarded by some as the founder of modern sociology, the categories of religious thought which traditional sacred histories employed, represented a primitive 'theological' stage of mind, that the philosophical rationalism of the seventeenth and eighteenth centuries had replaced and hence rendered redundant. This rationalism, which Comte referred to as the 'metaphysical' stage of human history, was in its turn being replaced by the scientific advances of the 'positivist' period in which he was writing. Spencer also saw contemporary science as containing the means of remedying the inadequacies of religious understanding, and set about deploying, as he saw them, the insights of evolutionary theory for the purposes of constructing a new world history.

The accounts of the past elaborated by Comte, or Spencer, have attracted little serious attention in this century, but, no less than Marx's more famous contention that 'all history is the history of class struggle', they exemplify an attempt that has often been made to formulate a grand narrative of human history without the assistance, or restrictions, of revealed religion. These attempts retain a continuity with those that they sought to replace. Between the older sacred histories and their newer secular rivals there are marked differences. Yet there remains in them the same aspiration of looking behind the surface appearances of history, and the same belief that we need not rest content with merely recounting sequences of events in chronological order, but may properly seek to discern a meaningful direction within them.

Alternatives to sacred history, such as those of Hegel, Marx and Comte, have been called 'grand narratives', and in them we find the origins of philosophy of history, to be contrasted with history *simpliciter*. These various and familiar conceptions all illustrate the nature of 'philosophical' as opposed to 'scientific' history. They contend that behind the miasma of historical chance and change, the past has a shape or a direction or a meaning. The task of discerning this shape has attracted various labels. Sacred history is one, grand narrative, world history, philosophical history and universal history are others. But whatever the variety of terms, it is here that philosophy of history proper begins.

3 Analytical philosophy of history

The foregoing is philosophy in rather a grand sense, the sense in which many people who do not know a great deal about philosophy expect it to be conducted. But from an early stage philosophers, especially in the analytical tradition, were sceptical of the ambitions of Hegel and his successors. In fact Hegelian philosophy of history, and with it the whole endeavour to produce a universal history, fell out of favour relatively quickly, and never attracted very widespread support in the English speaking world. There were several reasons for this, one of which was that Hegelian philosophy came to be associated with totalitarianism. A more deep-seated philosophical reason, however, had to do with the success of natural science and the philosophical dominance of the positivistic, empirical conceptions of knowledge to which this success gave rise. The positivist idea that our knowledge can only comprise plain, empirically ascertainable fact is, by its nature, unsympathetic to conceptions of hidden or underlying 'meanings' (see Positivism in the social sciences §1). This is as true of history as of other disciplines.

The most influential elaboration of these doubts for the philosophy of history is to be found in Karl Popper's celebrated book *The Open Society and its Enemies* (1945), in which he develops a line of thought similar to that found in earlier British critics of Hegel, notably L.T. Hobhouse and Bertrand Russell (see Popper, K. §4). In an earlier work Popper had developed a criterion of proper science, by which we should take him to mean real

knowledge. This is the falsifiablity criterion. A theory counts as a genuine contribution to knowledge if it can in principle be falsified by empirical observation, and has withstood attempts to falsify it. According to Popper, Hegel’s philosophy of history, like its less philosophical successor, Marxism, is vitiated by its manipulation of evidence to avoid falsification. Rather than the theory being answerable to the facts scientific history uncovers, the ‘facts’ are determined by the theory. Hegel and Marx do not tell us how the past was, but how, in the light of their theories, it must have been. In this way, though they masquerade as historians, they in fact pass beyond the realms of empirical history, and the claims they make, though they may superficially look like historical theories, are actually unfalsifiable a priori assertions. Thus, when Hegel considers the histories of India or China, which do not fit the general conception with which he is working, they are not taken as refutations, but explained away as aberrations.

In a later book, *The Poverty of Historicism* (1957), Popper applied this criticism to theories of history more broadly, and his attack on what he labelled ‘historicism’ has generally been regarded as both innovative and effective. Marx raises the very same criticism against Hegel, and Hegel himself is not only aware of it, but tries to answer it. Whatever the justice of the criticism, or the cogency of Hegel’s original reply, Popper’s charge stuck for some considerable time, and this caused the philosophy of history to take a different turn.

If the construction of grand narratives is no longer respectable, what interest can philosophy take in history? The answer is ‘methodology’. In this respect philosophy of history in the twentieth century followed the path of philosophy of science a century earlier. With the divorce between natural philosophy and empirical science, philosophers in the nineteenth century ceased to concern themselves directly with explaining the phenomena of the natural world, as philosophers of an earlier period had tried to do, and focused instead upon the methods by which empirical science did this. This is a large part of the purpose of John Stuart Mill’s *System of Logic* (1843), for instance, which has had great influence on philosophers of science, whose aim increasingly became that of making explicit the methods of science (see Mill, J.S.). In a similar fashion, philosophers tried to articulate the methods and presuppositions of empirical history.

Once the subject turned in this direction it was inevitable that comparisons would be drawn between science and history, and hence for a time the predominant ambition was to give history a status equivalent to that of science. Thus, for instance, against a background assumption that natural scientists are value-neutral with respect to their subject matter, it became a question as to whether historians could likewise achieve neutrality with respect to the events they recounted. The attempt to answer this question positively was allied by a movement among historians themselves to abandon ‘Whig’ interpretations of history, according to which the past was conceived as a struggle between the ‘good’ forces of enlightenment and the ‘bad’ forces of reaction.

The high point of philosophy of history modelled on philosophy of science was a famous essay by Carl Hempel, ‘The Function of General Laws in History’ (1942). Having earlier set out a logic of scientific explanation, showing it to require only empirical observation and logical validity, Hempel argued that the same pattern could be found in the investigations of historians, and that the explanation of historical events required general ‘covering laws’ no less than that of physical events. However, Hempel’s application of natural scientific models to history was convincing only so long as it was thought that he had really uncovered the logic of science. In fact, philosophers of science, including Hempel himself, found serious difficulties with the basic model, which, being a purely logical structure, had difficulty in accommodating temporal order and causal dependence. It could, for instance, as easily accommodate the idea that the length of a shadow explains the height of the sun, as that the height of the sun explains the length of a shadow.

4 History and the understanding of action

Despite several attempts like Hempel’s, the difficulty of establishing a convincing parallel between history and natural science raised more serious sceptical doubts: do historians actually explain events? Is an objective knowledge of the past possible at all? These questions drew philosophers of history into yet more general questions of epistemology and metaphysics. What is knowledge? What sort of existence does the past really have? In turn, this prompted an examination of alternative philosophies of knowledge and understanding and, at the hands of philosophers such as Isaiah Berlin, attention began to be paid to German thinkers of the nineteenth century, notably Herder and Dilthey (§3). It also drew attention to the earlier and hitherto neglected author of *The New Science*, the Italian philosopher Giambattista Vico (§6), whose principal intellectual endeavour had been to

capture the distinctiveness of cultural study.

This movement away from the scientific analogy was strengthened by those philosophers of history who had remained convinced that the attempt to model history on science was mistaken. One important thinker in this regard was R.G. Collingwood, an Oxford philosopher, also a distinguished historian in his own right, who had remained in the idealist tradition, largely unaffected by the major currents of the positivistic philosophy of his time. Collingwood's *The Idea of History* (1946) was not published until after his death, but it set out in an influential way both the history of history and a striking alternative account of historical understanding to that of the scientific model.

Collingwood argues, in common with many German philosophers of the nineteenth century, that there is a radical difference between natural processes and historical events. What distinguishes the latter from the former is that they have meaning. That is to say, natural processes, such as volcanoes or chemical reactions, have no significance for the objects or materials which undergo them. This leaves scientists who study them free to identify and order in accordance with their theoretical and explanatory purposes. But the actions of which history is composed are quite different. These have significance and meaning for the agents who performed them; their nature is determined by the cultural context and individual thought that gives rise to them. In the strict sense, natural processes can only be explained; there is nothing about them that requires understanding. By contrast, the events of history have to be understood, and Collingwood marked this difference by saying the historical events have an inner meaning as well as an outer form. The task of the historian, accordingly, is to rethink the thought of the historical agent.

Among philosophers in the analytical tradition, the writer who brought Collingwood's ideas to greatest prominence was W.H. Dray. In his book *Laws and Explanation in History* (1957), Dray directly confronted the Hempel-inspired conception of historical explanation, and argued for a more humanistic conception. In the same spirit, W.H. Walsh argued in favour of 'colligatory' explanations that brought to prominence the idea of historical explanation as narrative. For a time these alternatives fought a vigorous battle in a large number of books and articles. The argument was extended, in fact, into the philosophy of the social sciences, where the gulf between empirical, positivistic theories and hermeneutic, interpretative conceptions of the understanding of human behaviour widened (see Social science, methodology of §4).

However, no clear victory emerged on either side, and the dispute rather suddenly ground to a halt. Interest in Vico, Collingwood and hermeneutics all grew, but explicit interest in the philosophy of history itself declined.

5 The return of 'grand theory'

With the demise of its analytical version, philosophy of history appeared almost to disappear. But this was largely a matter of mere appearance. In other branches of philosophy the introduction of a historical element became increasingly important. In Richard Rorty's *Philosophy and the Mirror of Nature* (1980), for example, the thesis was advanced that the central problems of metaphysics and especially epistemology had to be understood, and resolved, in the context of a historically specific ambition (see Rorty, R.M.). And Alasdair MacIntyre's *After Virtue* (1981) applied the same sort of thinking to the problems of moral philosophy. According to MacIntyre, the philosophical difficulties encountered by attempts to render morality coherent were the outcome of a long historical process of cultural disintegration (see MacIntyre, A.). Charles Taylor, in *Sources of the Self* (1989), was another influential thinker who sought philosophical illumination from the history of ideas (see Taylor, C.).

Rorty, MacIntyre and Taylor, though closely concerned with the analysis and resolution of philosophical problems, all took an approach that set them in the broader context of the general sweep of the past. Their concern was with what had happened to the human mind in the course of developing its ideas. None, with the possible exception of Taylor, were self-consciously followers of Hegel, but the Hegelian character of their approach to philosophy and its past is unmistakable. This powerful movement in analytical philosophy thus found common ground with hitherto alien continental philosophy, which, having always been more historical in its orientation, had gradually focused around the idea that the collapse of Enlightenment rationalism was the key to the demise of the modern mind, or modernity, fashioned by two hundred years of theorizing, and to understanding a postmodern, relativistic world in which values, including even knowledge itself, were in constant flux. Among the most prominent continental theorists are Martin Heidegger, Michel Foucault, Emmanuel Levinas and Jacques Derrida.

These trends in philosophy have been highly influential in legal and social theory, literary criticism and

anthropology. They are not always self-consciously philosophical, however, and still less have they been seen as a revival of speculative philosophy of history. Yet this is what they are - the application of philosophical concepts and analysis to history, and especially cultural history, in the belief that an adequate understanding of contemporary thought requires the mutual illumination of the two. The difference from Hegel's philosophy of world history is in substance rather than style. Whereas Hegel believed that the world of ideas he inhabited was the culmination of rational processes, one in which the human mind had finally arrived at its own intelligibility, postmodernist thinkers, Rorty and MacIntyre no less than continental philosophers such as Foucault and Derrida, generally think of the modern mind as a fragmentary residue of past ideas, and incapable of sustaining the sort of rational intelligibility the Enlightenment aspired to. They are thus 'grand theorists' no less than Herder, Kant or Hegel, though whereas the latter were rationalists and progressivists, the ideas of the former are broadly relativist and pessimistic.

The question therefore arises as to whether these modern grand theories are susceptible once more to the sorts of criticism that turned philosophy in a more methodological direction. And, in a slightly disguised form, this is indeed the direction in which the subject has moved. It is not the illegitimate deployment of the a priori in history that is the chief concern, but whether these new theorists can escape the intellectually crippling implications of radical relativism, and whether indeed they offer a real prospect of genuine knowledge and understanding. If all is in flux, is there any Archimedean point from which this judgment itself is to be made?

In this way philosophy of history, in common with many other branches of philosophy, is best understood as a series of repeated attempts to organize knowledge and insight into a coherent whole, one characterized by a constant dialectic between philosophical criticism and historical understanding.

See also: Hermeneutics; Historicism; History, Chinese theories of; Kyoto School §5

GORDON GRAHAM

References and further reading

Atkinson, R.F. (1978) *Knowledge and Explanation in History*, London: Macmillan. (A good introduction to analytical philosophy of history.)

Augustine (413-26) *The City of God Against the Pagans*, trans. G.E. McCracken, London: Heinemann, 1957.(In this monumental work are the origins of philosophical history.)

Collingwood, R.G. (1946) *The Idea of History*, Oxford: Clarendon Press.(The most influential elaboration of an idealist conception of historical understanding in English-speaking philosophy.)

Comte, A. (1890) *Introduction to Positive Philosophy*, ed. F. Ferré, Indianapolis, IN: Bobbs-Merrill, 1970. (Comte's positivism is not now well regarded.)

Condorcet, A.N. de (1795) *Sketch for an Historical Picture of the Progress of the Human Mind*, trans. J. Barraclough, London: Library of Ideas, 1955.(A good example of the intellectual ambitiousness of the French *philosophe*.)

Croce, B. (1921) *History: Its Theory and Practice*, trans. D. Ainslee, New York: Harcourt Brace.(The influence of this prominent Italian philosopher on philosophy of history is largely through the writings of Collingwood.)

Danto, A.C. (1965) *Analytical Philosophy of History*, Cambridge: Cambridge University Press.(Arguably the high point of analytical philosophy of history.)

Dray, W.H. (1957) *Laws and Explanation in History*, Oxford: Clarendon Press. (The most influential rejection of the 'covering law' theory of historical explanation.)

Gardiner, P. (1952) *The Nature of Historical Explanation*, Oxford: Clarendon Press. (A useful introduction to the central debate among analytical philosophers of history.)

(1959) *Theories of History*, New York: Free Press.(A useful collection of extracts from classic works and essays in analytical philosophy of history.)

Hegel, G.W.F. (1837) *Lectures on the Philosophy of World History*, trans. H.B. Nisbet, Cambridge: Cambridge University Press, 1975.(The best translation of the high point of philosophical history.)

Hempel, C.G. (1942) 'The Function of General Laws in History', *Journal of Philosophy* 38 (1): 35-48.(The definitive statement of the 'covering law' theory of historical explanation.)

Herder, J.G. (1800) *Outlines of a Philosophy of the History of Man*, trans. T.O. Churchill, abridged F.C. Manuel, London: University of Chicago Press, 1968.(This translation originally appeared in 1803.)

Kant, I. (1784) 'The Idea of a Universal History from a Cosmopolitan Point of View', in *Kant on History*, trans. L.W. Beck, R.E. Anchor and E.L. Fackenheim, Indianapolis, IN: Bobbs-Merrill, 1963.(The *locus classicus* of

philosophical history shorn of its theological origins.)
MacIntyre, A.C. (1981) *After Virtue*, London: Duckworth.(One of the most influential books in moral and social philosophy in the twentieth century.)
Marx, K. (1867) *Capital*, trans. B. Fowkes, Harmondsworth: Penguin, 1976.(The outcome of Marx's many years of intellectual labour, only the first volume was published in his lifetime.)
Marx, K. and Engels, F. (1848) The Communist Manifesto, in *Selected Works in Two Volumes*, Moscow: Foreign Languages Publishing House, 1951.(A brilliantly short résumé of the historical materialist theory of history.)
Mill, J.S. (1843) System of Logic: Ratiocinative and Inductive, in *Collected Works of John Stuart Mill*, Toronto, Ont.: University of Toronto Press, vols 7 and 8, 1991. (Mill's account of scientific method held impressive sway for almost a hundred years.)
Popper, K.R. (1945) *The Open Society and its Enemies: Volume 2: The High Tide of Prophecy*, London: Routledge.(A celebrated, though somewhat misguided, attack on Hegel.)
Popper, K.R. (1957) *The Poverty of Historicism*, London: Routledge & Kegan Paul. (Widely regarded as a conclusive refutation of historicist philosophy.)
Rorty, R. (1980) *Philosophy and the Mirror of Nature*, Oxford: Blackwell.(An important modern attempt to dissolve philosophical problems through historical understanding.)
Spengler, O. (1918) *The Decline of the West*, trans. C.F. Atkinson, London: Allen and Unwin, 2 vols, 1926. (Possibly the most ambitious work in philosophical history ever attempted.)
Taylor, C. (1989) *Sources of the Self*, Cambridge: Cambridge University Press. (Taylor may be ranked with Rawls and MacIntyre as among the most influential moral philosophers of the late twentieth century.)
Vico, G. (1744) *The New Science*, trans. T.G. Bergin and M.H. Fisch, Ithaca, NY: Cornell University Press, 1948. (Sir Isaiah Berlin discovered in this hitherto obscure Italian writer the founder of the human sciences.)
Walsh, W.H. (1967) *Introduction to Philosophy of History*, London: Hutchinson, 3rd revised edn.(For many years the standard introduction to the subject.)

Hobbes, Thomas (1588-1679)

Among the figures who were conscious of developing a new science in the seventeenth century, the Englishman Hobbes stands out as an innovator in ethics, politics and psychology. He was active in a number of other fields, notably geometry, ballistics and optics, and seems to have shown considerable acumen as a theorist of light. His contemporaries, especially in Continental Europe, regarded him as a major intellectual figure. Yet he did not earn a living as a scientist or a writer on politics. In 1608 he entered the service of Henry Cavendish, First Earl of Devonshire, and maintained his connections with the family for more than seventy years, working as tutor, translator, travelling companion, business agent and political counsellor. The royalist sympathies of his employers and their circle determined Hobbes' allegiances in the period preceding and during the English Civil War. Hobbes' first political treatise, The Elements of Law (1640), was not intended for publication but was meant as a sort of long briefing paper that royalists in parliament could use to justify actions by the king. Even Leviathan (1651), which is often read as if it is concerned with the perennial questions of political philosophy, betrays its origins in the disputes of the pre-Civil War period in England.

For much of his life the aristocrats who employed Hobbes brought him into contact with the intellectual life of Continental Europe. He found not just the ideas but also the spokesmen congenial. Perhaps as early as 1630 he met Marin Mersenne, then at the centre of a Parisian network of scientists, mathematicians and theologians that included Descartes as a corresponding member. It was to this group that Hobbes attached himself in 1640 when political events in England seemed to him to threaten his safety, causing him to flee to France. He stayed for ten years and succeeded in making a name for himself, particularly as a figure who managed to bring geometrical demonstration into the field of ethics and politics. His De cive, a treatise that has much in common with the Elements of Law, had a very favourable reception in Paris in 1642.

By the time De cive appeared, Hobbes had taught himself enough natural philosophy and mathematics to be taken seriously as a savant in his own right. He had also conceived the plan of producing a large-scale exposition of the 'elements' of philosophy as a whole - from first philosophy, geometry and mechanics through to ethics and politics. De cive would be the third volume of a trilogy entitled The Elements of Philosophy. These books present Hobbes' considered views in metaphysics, physics and psychology against the background of a preferred scheme of science.

Metaphysics, or first philosophy, is primarily a definitional enterprise for Hobbes. It selects the terms whose significations need to be grasped if the principles of the rest of the sciences are to be taught or demonstrated. Foremost among the terms that Hobbes regards as central are 'body' and 'motion'. According to Hobbes, the whole array of natural sciences can be organized according to how each treats of motion. Geometry is the first of these sciences in the 'order of demonstration' - that is, the science whose truths are the most general and on which the truths of all the other natural sciences somehow depend. Mechanics is next in the preferred order of the sciences. It considers 'what effects one body moved worketh upon another'. Physics is the science of sense and the effects of the parts of bodies on sense. Moral philosophy or 'the science of the motions of the mind' comes next, and is informed by physics. It studies such passions as anger, hope and fear, and in doing so informs civil philosophy. Starting from the human emotional make up, civil philosophy works out what agreements between individuals will form commonwealths, and what behaviour is required within commonwealths to make them last.

The behaviour required of the public in order to maintain a commonwealth is absolute submission to a sovereign power. In practice this means abiding by whatever a sovereign declares as law, even if those laws appear to be exacting. Law-abiding behaviour is required so long as, in return, subjects can reasonably expect effective action from the sovereign to secure their safety and wellbeing. With minor variations, this is the theme of all three of Hobbes' political treatises - the Elements of Law, De cive and Leviathan. Government is created through a transfer of right by the many to the one or the few, in whom an unlimited power is vested. The laws of the sovereign power may seem intrusive and restrictive, but what is the alternative to compliance? Hobbes' answer is famous: a life that is solitary, poor, nasty, brutish and short. This conception of life without government is not based on the assumption that human beings are selfish and aggressive but, rather, on the idea that if each is their own judge of what is best, there is no assurance that one's safety and one's possessions will not be at the mercy of other people - a selfish few, a vainglorious minority or even members of a moderate majority who think they have

to take pre-emptive action against a vainglorious or selfish few. It is the general condition of uncertainty, in conditions where people can do anything they like to pursue their wellbeing and secure their safety, that Hobbes calls 'war'.

1 Life

Hobbes was born in Westport, a parish of the town of Malmesbury in Wiltshire, England. His mother came from a yeoman family; his father was a poorly educated vicar who seems to have left his parish in disgrace, deserting his family after having come to blows with another clergyman early in Hobbes' childhood. Hobbes' uncle subsequently supported the family, and it was he who paid for Hobbes' university education. Hobbes was lucky to receive good schooling locally, and he showed an early talent for the classical languages.

In 1602 or 1603, Hobbes began study towards an arts degree at Magdalen Hall, Oxford. From his criticisms of the universities in his published writings, it is sometimes inferred that he disliked his college days, or at least that he disliked the scholasticism of Oxford at that time. (Scholasticism - the fusion of Christian with ancient Greek thought, especially the thought of Aristotle - dominated the curricula of the schools and universities of Europe in the sixteenth and early seventeenth centuries - see Medieval philosophy.) Certainly he disliked the university curriculum in retrospect, as chapter 46 of *Leviathan* (1651) makes clear.

Hobbes completed his degree in 1608, and entered the service of William Cavendish, First Earl of Devonshire, as companion and tutor for his son. Although Hobbes was about the same age as the young Cavendish, he was put in charge of his purse as well as his education. He was the earl's representative at meetings of the Virginia Company, in which the Devonshire family had a considerable financial stake. He also accompanied the earl's son on a grand tour of the Continent in 1610, which allowed Hobbes to improve his command of French and Italian. According to some accounts, he also became acquainted then with criticisms of scholasticism current among Continental intellectuals.

It is unclear how long this grand tour lasted, but Hobbes had returned to England by 1615. At some point during these travels Hobbes seems to have met Fulgenzio Micanzio, the friend and personal assistant of the Venetian writer and politician Paolo Sarpi. He must also have met Marc Antonio de Dominis, who was involved in the translation of Bacon's writings into Italian and who also had connections with Sarpi. Hobbes' own contact with Bacon may have had its stimulus in the requests of the newly befriended Venetians for more details of the Baconian philosophy. The young Cavendish began a correspondence with Micanzio after returning to England in 1615. Hobbes translated this correspondence and through it would have been exposed to Sarpi's theory of the supremacy of temporal rulers rather than spiritual authorities. The theory went against the Papal interdict of 1606, which asserted Rome's right to overrule the decisions of local monarchs and which had encountered much criticism in England. There are apparently strong echoes of this anti-Papal line in Hobbes' own writings.

During his first twenty years of service to the Devonshires, Hobbes seems to have spent his free time immersed in classical poetry and history. His employers had a good library, and Hobbes made use of it. The first fruit of this regrounding in the classics was a translation into English of Thucydides' *History of the Peloponnesian Wars*, published in 1628. Hobbes believed that Thucydides had lessons for those who overvalued democracy and did not see the strengths of monarchy, and it may have been the Petition of Right of 1628 that led to the publication of the translation. The Petition called on Charles I not to levy taxes without the consent of Parliament, not to imprison subjects without due cause, not to billet soldiers in private homes and not to put civilians under martial law.

In 1628 Hobbes chose history as the medium for a political message. Later, in writings like *Leviathan*, he thought science or philosophy was the better vehicle. In the writing of history it is possible for the conventions of the genre to interfere with the communication of wisdom; in the writing of science, he came to believe, the communication of wisdom is assured, if the audience is prepared to pay attention and able to follow a demonstration. He struggled throughout his intellectual life with the problem of combining political rhetoric with political science, and some of his best writings are experimental solutions to this problem. The translation of Thucydides is important as the first of many such experiments.

The year 1628 was a kind of turning-point in Hobbes' career. Apart from publishing the translation of Thucydides, he had to contend with the death of the second earl at the age of 43, resulting in his loss of employment with the Devonshires. Hobbes took up a new post in the house of Sir Gervase Clifton, not far from Hardwick Hall, the

home of the Devonshires. Once again he was engaged as a companion on a grand tour, this one lasting from 1629 to 1631. During this journey Hobbes looked for the first time at Euclid's *Elements*, and fell in love with geometry. There is plenty of evidence in Hobbes' writings that he regarded Euclid's book as one of the supreme examples of a scientific presentation of a subject. Perhaps also during this second journey to the Continent Hobbes was present at a discussion among some well-educated gentlemen about the nature of sense-perception, in which it emerged that none of the participants could say what sense-perception was. Both episodes are significant, because they seem to mark the beginning of Hobbes' transformation from man of letters to man of science.

Perhaps the stimulus for the change was not the second grand tour alone. After his return, Hobbes went back into the service of the Devonshires and became tutor to the young third earl. At about the same time, he came into contact with a branch of his master's family who lived at Welbeck, near Hardwick Hall. The Welbeck Cavendishes were interested in science. The Earl of Newcastle is known to have sent Hobbes on an errand to London to find a book of Galileo's in the early 1630s. The earl's younger brother, Charles, had an even greater interest in science: he acted as something of a patron and distributor of scientific writings. Hobbes was one of those who looked at and gave his opinion of these writings. Charles Cavendish also had contacts among Continental scientists, including Marin Mersenne, a friar in Paris who was at the centre of a circle of scientists and philosophers that included Descartes.

Hobbes' scientific development continued when he embarked with the third earl on yet another grand tour from 1634 to 1636. During this journey he is supposed to have met Galileo in Italy, as well as Mersenne and some members of his circle in Paris in 1636. Hobbes had probably become acquainted with Mersenne five years earlier on the second tour. It is said that on the third grand tour Hobbes was much preoccupied with the nature and effects of motion, and that he started to see for the first time how many natural phenomena depended upon it.

On his return to England, Hobbes kept up with some of the scientific work being produced in Mersenne's circle. Descartes' *Discourse and Essays* were published in 1637. Hobbes was sent a copy and seems to have made a careful study of the first of the *Essays* - on optics - perhaps taking time himself to write something of his own on the same subject. But he was not keeping abreast solely of scientific ideas. Through his association with a circle of clergy, lawyers and aristocrats at Great Tew, near Oxford, he was able to follow the continuing debates surrounding the troubles of Charles I. In 1634 the king started to raise funds for a navy by a ship-money tax levied county by county. This tax-raising met opposition, particularly in non-coastal counties. Besides the ship-money dispute, Charles I had to reckon with the consequences of trying, in 1637, to bring the Scottish Presbyterian prayer book into line with its Anglican counterpart. This provoked a National Covenant in Scotland expressing wholesale opposition to ecclesiastical innovations from England. In 1639 and 1640 the Scots raised armies to back up their opposition, and Charles was forced to recall a parliament he was used to ruling without, and which was extremely hostile to him. When Parliament acted against Stafford, a minister of the King associated with the Earl of Newcastle, Hobbes worked on arguments that could support the royalist position in parliamentary debates. The arguments were produced in a treatise, *The Elements of Law* (1640), not intended for publication but which, in fact, contains much of the doctrine of Hobbes' political philosophy. Fearing that he would be prosecuted for giving the royalists their arguments, he fled to Paris and joined the circle of philosophers and scientists around Mersenne.

Some years after 1640 Hobbes wrote that he had recently conceived a plan for expounding, in three parts, the elements of philosophy or science in general. His exposition would begin with the nature of body and the elements of what we now call physics. It would go on to discuss human nature, in particular perception and motivation, and the third part would be a discussion of moral and civic duty. Perhaps he had already drawn up this plan, and even executed some of it, by the time he reached Paris. What *is* certain, however, is that the first part of the exposition to be published was the last of the three in his outline - the part on morals and politics. Hobbes called this part of his exposition *De cive* and published a very limited edition of it in Latin in 1642.

Hobbes seems to have enjoyed good relations with most of Mersenne's circle. He was at odds with Descartes, however, whose *Meditations* he criticized in a set of 'Objections' - the anonymous 'Third Set' (see Descartes, R. §§4, 6). From 1641 until Mersenne's death in 1648, Hobbes applied himself to the composition of the rest of his three-part exposition of the elements of science. He produced some of the material for the first part of the exposition - on body, the part published in 1655 as *De corpore* - and took up topics that would later occupy the middle part of the exposition. In 1643 he wrote a critical commentary on *De Mundo*, a treatise written by Thomas

White (another Englishman in Paris at that time) which was sympathetic to scholasticism. In 1646 Hobbes composed some arguments about the respects in which freedom is compatible with causal necessitation in nature, arguing once more on this occasion against scholastic positions. He suffered a serious illness in 1647 and almost died. While on his sickbed he rebuffed an attempt by Mersenne to convert him to Roman Catholicism.

In 1648 Mersenne died and the philosophical and scientific activity that had gone on around him ceased to have a focal point. Hobbes now had a place among the royalists in exile, but he was on poor terms with churchmen around England's exiled Charles II in Paris and was receiving his pay rather irregularly. By the autumn of 1649 he seems to have formed the intention of going home.

Leviathan, in which Hobbes attempted to derive from his now well-worked-out political principles the right relation of Church to state, was written at the end of the 1640s, when church government in England began to run on lines of which he approved and at a time when the influence of bishops in the English royal court-in-exile in Paris was, in Hobbes' eyes, too great. In any case, the fact that the theory in the book vindicated Cromwell's policy on church government does not mean that it was a partisan work in favour of Cromwell, calculated to ease Hobbes' return to England. If that had been so, Hobbes would not have made a special presentation copy for the future Charles II. Instead, it seems that the doctrine in *Leviathan* favoured the concentration of all authority in any *de facto* sovereign power, whether republican or royal. To the Paris royalists, mostly strong Anglicans in favour of political powers for bishops, the new book was highly offensive.

By the end of 1651 Hobbes was back in London and all three statements of his political philosophy were available in some form in English. These political works were widely known and read before his exposition of the elements of science was complete. Even though *De cive* had been planned to complete a sequence of three treatises on these elements, it did not depend on the other treatises in order to be understood, and it has always had a readership of its own. When the other two works in the sequence appeared in the 1650s, they did not match *De cive* in quality. The treatise on which Hobbes had been working longest was the opening work of the sequence, *De corpore*, published in 1655. In it Hobbes tried to show how the mature sciences of geometry, mechanics and physics were concerned with the effects of different kinds of motion in matter. Politically motivated critics soon exposed the weaknesses in the mathematical sections of the book, and Hobbes' attempt to vindicate his work involved him in years of fruitless polemic. *De homine*, the second volume of his *Elements* and undoubtedly the least well-integrated of the three, was published in 1658. It was never widely read, and a modern English translation of it has only recently appeared.

In 1660 the monarchy was re-established in England, and on the coat-tails of Charles II there returned to political power many who regarded Hobbes as a traitor to the royalist cause. Charles himself was not hostile, however, and other influential people were also well-disposed towards him. Nevertheless, in 1666 and 1667, Parliament came close to passing a bill outlawing Christian heresy and atheism, and *Leviathan* was specifically investigated as a source of heretical and atheistic views. The danger of imprisonment and exile did not dissipate until the end of the decade. The threats to Hobbes were reflected in additions that he made to an edition of his works published in 1668. He argued that punishment for heresy was illegal under English law and that his materialism was compatible with Christian faith. Two significant works from the 1660s were applications of his political philosophy. There was a history of the English Civil War, *Behemoth* (1668), and the *Dialogue between a Philosopher and a Student of the Common Laws of England*. By the time these works were composed, Hobbes was not permitted to publish, and though he busied himself with some translations of the classics and a few other minor writings of his own in the 1670s, he had come almost to the end of his working life.

In his ninetieth year, Hobbes returned to physics. His *Decameron Physiologicum* (1678) restates some of the methodology and principal results of the physical sections of *De corpore*. For the preceding three years Hobbes had divided his time between the two Devonshire houses of Chatsworth and Hardwick. In December 1679 he died of a urinary complaint. His remains are buried in a small parish church near Hardwick.

2 Science and human improvement

Hobbes' writings are those of an advocate and practitioner of a new science, a system of knowledge of causes that could, as he believed, greatly benefit human life. Yet he formed his ideas during a period when human intellectual powers, including the ability to develop science, were commonly thought to be limited. According to some theories prevalent in the late sixteenth and early seventeenth centuries, the whole human race was involved in a

quite general and unstoppable process of natural decay, so that all the best achievements of human beings belonged to a long-lost golden age. It was a way of understanding things that was consistent with, if not inspired by, the Biblical story of the Fall of Adam and Eve, and the loss of paradise. According to some understandings of that story, Adam's expulsion from paradise cost him not only a life of ease in harmony with God and the rest of nature, but also the gift of a natural insight into the natures of all the things he could name. Regaining the knowledge of those natures might never be possible. In France, in the late sixteenth and early seventeenth centuries, recently popularized Greek sceptical arguments against dogmatism reinforced the view that human intellectual possibilities were limited (see Scepticism, Renaissance §§1, 4). The arguments were directed not only against the traditional learning of the schools but against the idea that any human learning - even untraditional or anti-traditional learning - could amount to a system of genuine knowledge. Hobbes' philosophy of science stands in opposition to much of this gloomy theorizing. It stands in opposition to philosophical scepticism, to the theory of the decay of nature as applied to the human intellect and, to a lesser extent, to pessimistic interpretations of the intellectual costs of the Fall.

Although Hobbes had close friends and intellectual influences among French thinkers who took sceptical arguments seriously, there is little if any solid evidence in his own writing that he studied these arguments closely or took their conclusions to heart. He seems never to have doubted the soundness or scientific status of Euclid's geometry, and he was an early enthusiast for the applied mathematics of Copernicus, Kepler and Galileo. He was also proud of the judgment of some early readers of *De cive* that this book ushered in a demonstrative science of ethics. Hobbes called himself the inventor of civil science and thought that his politics deserved a place alongside Galileo's mechanics. The newly founded natural and moral sciences he regarded not only as great intellectual achievements, but as distinctively modern ones. Contrary to the theory of the decay of nature as applied to the human intellect, natural and civil science had not ceased with the ancients, but had only properly begun with the work of the mathematical astronomers and his own work in *De cive*. However, Hobbes also held that human beings were badly adapted by nature to do science of this kind, and that they had to work very hard to be capable of it. He thus disagrees both with the Cartesian idea that God benignly creates us with the ingredients of science latent in our minds *and* with the Aristotelian idea that knowledge of the natures of things is the unforced and inevitable by-product of repeatedly looking and seeing. To a very significant extent, according to Hobbes, our capacities for natural science are made by us rather than given to us. This position concedes something to both scepticism and the theory that life will always be difficult for Fallen Man. Although sceptics are wrong to claim that we are incapable of science, they would not be wrong to claim that we lack native scientific ability. Neither would they be wrong to deny that human beings are capable of an exalted sort of knowledge, the knowledge of what necessitates effects. For Hobbes, the scientific knowledge of which we are capable rarely rises to the level of knowledge of how effects *must* have been brought about, and it is not taken to extend to all effects. Again, although Hobbes believed that the scientific achievements of his contemporaries and himself were important, and that the then nascent sciences of nature and politics would develop further, he did not, with Descartes, suppose that we might one day complete science. Finally, while he claimed that even in their undeveloped state the sciences had delivered considerable benefits, Hobbes did not expect a more developed science to be the answer to all our problems. Like Bacon he believed that science could not entirely repair the Fall, that it could only act to relieve some of what was bad in human existence. Whatever one did, life would continue to have 'incommodities', but with the development of science life would contain fewer of them.

Science is not, then, for Hobbes, a means of regaining Eden. At best, it is our only way of cutting the costs of losing Eden. In paradise, Adam enjoyed the gift of immortality in conditions of ready abundance. Everything he could properly want was there for the taking. Punished for eating from the tree of knowledge Adam lost his immortality. He lived, as Hobbes puts it in *Leviathan*, under a death sentence ([1651] 1839 III: 438). Adam also lost the abundance of Eden. Banished to a place outside paradise, he had for the first time to work for a living, and to do so in a relatively inhospitable environment. Had they stayed in Eden, Adam and Eve would not have reproduced their kind continually or perhaps at all ([1651] 1839 III: 440). When they left they came under a necessity to multiply that worsened the life of their kind still further. Adam's descendants, the rest of humanity, inherit from him not only their mortality but also life outside Paradise. Thanks to Adam's transgression, human beings in general live in a world that demands ingenuity and hard work for survival. And thanks to human carnality, Adam's descendants have to eke out a living in the company of, and often in competition with, many others of their own kind. These facts of life do not make it easy to do well. In order to flourish in a sometimes

harsh physical environment people have to know which effects they observe are beneficial and which are harmful, and they have to learn to reproduce the beneficial ones and prevent, or at least avoid, the harmful ones. In order to flourish in a heavily populated environment people have to know how to co-operate with one another. Moreover, these problems have to be coped with simultaneously.

As things naturally are, however, the problems are too great for creatures like us. For, being descended from Adam, human beings inherit the cognitive and conative capacities of someone designed to live in paradise, not the harsh world outside. If things had gone according to God's plan, Adam would not have needed to get causal knowledge of nature; he could have satisfied himself with contemplating the diversity and order in nature. Adam would not have had to make nature supply his needs. He would not have had to cope with overpopulation and the demands of co-operation. Made for a life without problems, Adam lacked the means - that is, the science - to solve problems. As Hobbes points out in *Leviathan*, there is no evidence in scripture that Adam had the vocabulary to do science ([1651] 1839 III: 19), and yet without the vocabulary to do science we should be no better off than savages or beasts (1640 pt I, ch. 5: iv; 1658 ch. 10: ii, iii). Either people reconcile themselves to living at the mercy of the elements and of one another - Hobbes thought that this was the course taken by Native Americans - or else they take the long and difficult road to a better life through science. Neither the better life nor the means to it are out of bounds to human beings, but both involve a kind of human reform. To get access to the degrees and varieties of motion that are required to understand nature, human beings need to acquire concepts more general and universal than those for everyday experience, and they need to apply principles involving these concepts in tasks of measurement and manufacture. To solve the problems of peaceful co-operation they need to be able to recognize the consequences of everyone's trying to get what they want. This means more than knowing what moral precepts to follow. It means being able to see what overall good the moral precepts promote - something revealed by Hobbes' civil science - and adjusting one's practical reasoning to the pursuit of that good rather than something nearer and more gratifying.

Although human beings cannot live well without science, science does not come naturally to them. Science depends on the ability to impose names aptly, to join names into propositions, and to join propositions into syllogisms, but not even these prescientific linguistic skills are natural to people. People come into the world being able at most to form sensory representations of things, and to learn from experience. But experience is a far cry from science. In its raw state experience is either a disorganized stream of representations or else a coherent sequence. If it is a coherent sequence, then, according to *Leviathan*, it is 'regulated' by some design or plan, or by curiosity about an observed body's effects ([1651] 1839 III: 13). Regulated in either way, a train of experience is only regulated as past experience allows it to be. Going by its past associations of observed phenomena, the mind will focus on a means to some goal or purpose in hand, or will suggest properties it is accustomed to conjoin with other properties it is now curious about. Once there are words for the things of which the mind has conceptions - words that can be used to signify the elements of experience - the possible ways of juxtaposing the words significantly, of analysing them and drawing consequences, introduce ways of ordering the elements that are not foreshadowed in previous experience.

New ways of regulating thought become possible because, for one thing, it is not necessary for a body spoken about to be present or remembered in order for a train of thought about it to be created. The train of thought can be generated instead by exploiting logical relations or analytic truths to get from one speech or thought to another. Reasoning can thus introduce new possibilities of combining things given separately in experience; it can also introduce ways of taking apart or separating things confounded in experience. Nor are the possibilities confined to the powers of one man's reasoning. Speech enables investigation and reasoning to be carried on co-operatively, and allows one person's explanations to be tested for clarity and coherence by others. The reasonings or explanations of one person can be preserved over time and made the model for those of many other people. A method can even be extracted from the findings of the most successful or conclusive pieces of reasoning, so that conclusive reasonings and explanations - in a word, science - can deepen and spread.

Hobbes believed that science as a whole could be divided into two principal parts, one concerned with natural bodies, the other with bodies politic. Each part of science arrives by reasoning at the causes of the properties of its subject matter; and demonstrates effects, with a view to making the relevant subject matter useful and beneficial to human beings. But the two types of bodies are very different from one another, and pose different scientific problems. Natural bodies are not made by us, and so the causes of their properties have to be worked out by

reasoning from the appearances they present. Since the maker of natural bodies, God, is omnipotent and able to bring about effects in more ways than can be dreamed of in our explanations, only possible causes can be assigned to the appearances they present. Bodies politic *are* made by us - they are human artefacts and so at least in principle we can be certain of the causes of *their* properties. But the philosophical challenge they present is not primarily that of knowing their causes: it is that of knowing how they should be designed so that they last. This means setting out rules by which those involved in the commonwealth - those in government on the one hand and those subject to government on the other - can conduct themselves so that civil peace is ensured. It is doubtful whether the statement of these rules is really a science of a kind of body, as is natural science, and probably the differences between natural and civil science are to be taken more seriously than the supposed analogies. Hobbes claims that civil science is not only more certain, but more widely needed, and more accessible than natural science. On the other hand, natural science is the more fundamental of the two parts of science: its explanatory concepts are more general than those of civil science and are needed if a scientific understanding of the passions and actions of agents in civil life is to be acquired from the ‘first and few’ elements of science as a whole.

3 The elements of philosophy: logic and metaphysics

The relative positions of the two parts of science - natural and civil - are reflected in the organization of Hobbes’ trilogy on the elements of philosophy. The first volume, *De corpore*, expounds first philosophy, geometry, mechanics and physics. It is followed by *De homine*, which is half optics and half psychology; this volume in turn is supposed to prepare the ground for the exposition of the elements of ethics and politics in *De cive*.

The account of the ‘elements’ of science starts in *De corpore* with chapters on the ways in which philosophy depends on names, propositions and methods of reasoning. For Hobbes, logic is nothing more than the right ordering and joining of significant propositions into chains of reasoning. Propositions in turn are no more than coherent concatenations of names with significations of different extents. A name signifies an idea - whatever idea it conveys in the context of a speech to a hearer. But the idea is not what the name refers to or stands for: it refers to or stands for an object. To make a proposition, names have to be put together coherently, and coherent concatenations are concatenations of the same category of name - names of bodies with names of bodies, names of names with names of names. The ‘extent’ of the signification of a name has a bearing on the truth of propositions. The signification of a proper name will extend to an individual, that of a universal name - ‘man’, ‘horse’, ‘tree’ - to each of a plurality of individuals. In the propositions of natural science, names are universal names of bodies. Truth in the propositions of natural sciences is a matter of the inclusion of the extent of a universal subject-term within the extent of a universal predicate-term. Demonstrations are chains of syllogisms, and syllogisms are the stringing together of trios of propositions that share appropriate subjects and predicates. In a sense, then, logic is a technique for working out the consequences of relations between the significations of universal names or their extents. There are also methods belonging to logic for analysing the significations of names, and for arriving at the most general of these significations from a starting point in everyday universal names. Logical analysis of this kind is what is required to locate the terms fundamental to the various branches of science; it also has a role in making scientific questions amenable to resolution. Metaphysics or first philosophy sets out, ideally by means of definitions, the concepts necessary for conducting fruitful enquiries concerning natural bodies and for communicating the results. The relevant concepts include those of body, motion, time, place, cause and effect.

Hobbes composed no full-scale treatise on logic, and no work of his is concerned with metaphysics alone. *De corpore* contains the nearest approximation to a full first philosophy, and even here he is not entirely clear about the borderline between that ‘prior’ science and geometry. There is a chapter on ‘syllogism’, but it is not comprehensive and its relation to traditional syllogistic is never spelled out. As for the chapters on first philosophy, they are more significant for what they deny than for what they affirm. They deny that it makes sense to study ‘being’ in the abstract; they deny that species or genera are things; they deny that ‘substance’ can mean something very different from ‘body’; and they deny that other predicables are more than varieties of sensory appearance caused by bodies. In short, they deny much standard Aristotelian doctrine, including the doctrine defining the subject matter of metaphysics itself - the doctrine that metaphysics studies being *qua* being. As will become clear, the chapters on first philosophy can also be understood to register disagreements between Hobbes and some of the *moderns* -Descartes and Gassendi, among others.

Hobbes’ first philosophy starts with a thought experiment. He imagines that the external world has been

annihilated - all that remains is a single thinker and the traces in memory of the world he previously sensed and perceived. Hobbes claims that the disappearance of the external world would not take away conditions for thought or reasoning, even about the physical world. The annihilation of the world would not even *alter* conditions for such thought and reasoning, since the medium of thought and reasoning is never things themselves but only appearances or phantasms. Hobbes thinks that the annihilation of the external world leaves only the mind and its phantasms in existence. There is no third world of things that exist outside the mind but outside the physical world as well. So Hobbes denies that there exist without the mind abstract natures such as Descartes claims to discover in Meditation V; and, contrary to Gassendi in the *Syntagma*, he does not think that space and time are real independently of the mind. He derives the idea of space from the memory images or phantasms presenting things as if from outside the mind. He derives the idea of time from imagined motion, from succession without existence. He derives the idea of an existing thing from the imagination of an empty space suddenly getting an occupant. Existence is thus restricted to existence in space, which is in turn identified with corporeal existence. These resources allow for only a straitened conception of cause or power, and certainly not for an Aristotelian conception. There are no forms or purposes in nature; but the makings of a conception of efficient cause are available, and that is the only sense of 'cause' recognized by Hobbes' first philosophy.

Most of the concepts that Hobbes thinks are needed for natural science have now been indicated. In Part Two of *De corpore*, after defining 'time' and 'place', he thinks he is in a position to define 'body' and its most general accidents. He then deals with magnitude or real space-occupation, and the spatial relations of continuity and contiguity. Against this background he defines 'motion', and in terms of motion the ideas of length, depth and breadth. After the three spatial dimensions are explicated, he defines quantitative identity and difference for motions and bodies, and then discusses the conditions of qualitative difference between bodies over time. He goes on to consider the causes of qualitative change, concluding with a demonstration of the thesis that all change is motion, that motion is the only cause of motion, and that power (*potentia*) is nothing but motion in so far as it is a cause of motion.

Definitions dominate Hobbes' first philosophy and, officially at least, their purpose is to fix ideas necessary for the business of science proper. Hobbes thought that the mark of a prescientific branch of learning was controversy, and he traced controversy to a failure to define terms and to proceed in orderly fashion from definitions to conclusions. The task of first philosophy is to provide insurance against controversy. It does not do this by coming up with substantive truths that command assent. Instead, Hobbes describes first philosophy as a necessary *preliminary* to the demonstration of substantive truths, where demonstrator and learner are put on one another's wavelength and attach the same significations to their terms, but where their agreement is terminological rather than doctrinal. As Hobbes puts it in the *Six Lessons*, 'he that telleth you in what sense you are to take the appellations of those things which he nameth in his discourse, teacheth you but his language, that afterwards he may teach you his art. But teaching of language is not mathematic, nor logic, nor physic' ([1656b] 1839 VII: 225).

That 'the teaching of language' underdescribes what Hobbes does in practice in stating his first philosophy, and that it suppresses entirely the revisionary character of some of his definitions when compared with Aristotelian ones (and so the controversial nature of the devices that are supposed to pre-empt controversy), should already be clear. But, up to a point, Hobbes' first philosophy is genuinely unassuming. It takes for granted no exotic powers or substances, God included, and it postulates no exotic human capacities for acquiring the concepts that are the key to natural science. The point is not just that Hobbes keeps the relevant concepts to a small number, so that he is economical in the concepts he uses and also in his assumptions about the types of real things there have to be for these concepts to be applicable. Hobbes' first philosophy is also *naturalistic*. Nothing supernatural is assumed to exist in order for natural science to be acquired; indeed, nothing besides matter in motion is postulated. What remains after the annihilation of the world in Hobbes' thought experiment is not the immaterial self of Descartes, but the corporeal body or perhaps the brain, and the motions conserved in its internal parts from past impacts of the external world on the sense-organs.

The denial of immaterialism in Hobbes' first philosophy is anticipated in his Objections to Descartes' *Meditations*. An objection that Hobbes directs at Meditation II sets the tone (see Descartes, R. §5). He accepts that from the fact that I am thinking it follows that I exist, but he wonders whether Descartes can properly conclude, as a corollary, that the I is a mind or an intelligence or a thinking thing. For all the *Cogito* shows, Hobbes says, the I could be corporeal. And not only does the *Cogito* leave open the possibility of the I being corporeal, he goes on, the later

wax argument actually *shows* that the I is corporeal:

> We cannot conceive of jumping without a jumper, or knowing without a knower, or of thinking without a thinker.
>
> It seems to follow from this that a thinking thing is something corporeal. For it seems that the subject of any act can be understood only in terms of something corporeal or in terms of matter, as the author himself shows later [in] his example of the wax: the wax, despite the changes in its colour, hardness, shape and other acts, is still understood to be the same thing, that is, the same matter is the subject of all of these changes.
>
> (Hobbes [1641] 1985 vol. 2: 122)

Descartes' reply concedes that acts need subjects, that it is a *thing* that is hard, changes shape and so on, and also a *thing* that thinks, but he insists that 'thing' in this sense is neutral between the corporeal and the spiritual. He insists, too, that he is non-committal about the nature of the thing that thinks in Meditation II, a claim borne out by his responding agnostically in Meditation II to the question of whether he might be a structure of limbs or a thin vapour. If Hobbes misses that, it may be because he misunderstands the rules of the method of doubt. While implementing the method of doubt, Descartes does assume rather than prove that there is no body for the thinking thing to be or for the thought to inhere in. But this is not a case of begging the question, for the belief in the existence of bodies is reinstated in Meditation VI, and with that the question of whether the subject is essentially immaterial or material.

Hobbes comes at Descartes' immaterialism from another, more revealing direction when he tries to suggest that it is not needed to underpin the distinction between imagination and conception by the mind. In his fourth objection, Hobbes equates imagination in Descartes' sense with having an idea of a thing, and conception in Descartes' sense with reasoning to the conclusion that something exists. Descartes already agrees that imagination is a partly corporeal process resulting from action on the sense organs, but his text suggests that conception by the mind is an altogether different operation. Hobbes puts forward a suggestion that allows the explanation of conception and imagination to be linked, without the postulation of immaterial things. He proposes that reasoning is the process in which labels attached to various things are concatenated into sentences according to conventions agreed by humans.

> Reasoning will depend on names, names will depend on the imagination, and imagination will depend (as I believe it does) merely on the motions of our bodily organs; and so the mind will be nothing more than motion occurring in various parts of an organic body.
>
> (Hobbes [1641] 1985 vol. 2: 126)

The compatibility of this proposal with mechanistic explanation appeals to Hobbes; but Descartes raises some powerful doubts about Hobbes' idea that names alone come into reasoning. Contrary to Hobbes, Descartes takes it that reasoning is a matter of linking together the significations of names, not just the names themselves, and also that the significations of some names cannot be imaged.

As in the case of his objection concerning the subject of the thinking, which seems to overlook the constraints of the method of doubt, Hobbes' objection to Descartes on imagination and conception seems to miss the point. Descartes is not trying to explain the workings of the faculties that result in science, only to find that he has to explain them on immaterialistic principles. He is trying to show that science is possible, that real knowledge of the physical world is possible, because not all of our faculties can coherently be held to be unreliable. Conception by the mind is a case in point. It cannot be held to be unreliable, because it is autonomous and independent of unreliable sense-perception. Hobbes does not see that it is the objectivity of conception rather than the process of conception that Descartes is concerned with. And doubting the objectivity of conception himself, Hobbes does not seek to reconstruct conception as reasoning that might be guaranteed to lead to true conclusions; he wants only to reconstruct it in ways that will not multiply entities beyond those required by mechanistic explanation. The point is that a proof of the objectivity of the conceptions arrived at by science may legitimately be demanded of a metaphysics, and Hobbes' metaphysics does little if anything to meet the demand. The metaphysical economy of materialism will not impress someone who is sceptical of the existence of the external world: the undeniability of the *Cogito* might. One cost of Hobbes' naturalistic approach is that it never attempts the task of legitimizing the scientific enterprise in general, and is probably incapable of doing so.

4 Geometry, optics and physics

First philosophy is a preliminary to natural science proper, and the first of the natural sciences is geometry. Geometry is a natural science for Hobbes, because it studies the effects of moving bodies. It studies the properties of straight lines, for instance, and straight lines are the effects of the motion of a small material thing - a point. He rejects the idea that the geometrical point is an abstraction distinct from some small material mark or other. Being a body, a geometrical point is divisible and not, as Euclid had it, ‘that which hath no part’. A point could no more lack quantity than a line could lack breadth and be constructed by motion. The bodies whose motions geometry studies may not have much quantity, and the quantity may not be relevant to what is being demonstrated of them, but they are bodies all the same. Though geometry is a science of bodies, it is also in a sense an a priori certain science, as physics is not, for the effects are produced by us, and we know in advance by what means they are produced. In this, Hobbes believed, geometry is like politics. In a sense, too, geometry is a very basic science - basic not in the sense that the objects it studies are higher or more real than those in nature, but in the sense that it studies bodies and motion at a very high level of generality, with much that is specific about the bodies left out of account. Hobbes thinks that geometry is also basic in another sense, for its methods of demonstration and analysis are the inspiration for the methods of the other demonstrative sciences.

Hobbes was self-taught in mathematics, and his friend and biographer, John Aubrey, says he did not encounter Euclid until he was middle-aged. Nevertheless, he was taken seriously by very able geometers in Mersenne’s circle, and is even credited with inspiring a proof by Roberval of the equality of arcs of a parabola and an ‘Archimedean’ spiral. He is much better known, however, as a mathematical failure whose attempts at expounding geometry in *De corpore* were ridiculed by the English mathematician John Wallis. Wallis’ attack was motivated by a wish to discredit the anticlerical passages in *Leviathan*, and its attack on the universities. In correspondence with Huygens, Wallis said that Hobbes ‘took his courage’ from mathematics, and so it ‘seemed necessary for some mathematician to show him how little he understands’. Wallis’ attack succeeded - it focused on Hobbes’ doomed enterprise of producing a quadrate of the circle - and was made the more effective by Hobbes’ persistent refusal to concede errors.

Although the geometrical parts of *De corpore* were supposed to present some new results, Hobbes did not claim any great stature for himself in geometry. In optics, on the other hand, he regarded himself as a major figure. It was certainly a subject he turned to early in his transformation from man of letters to man of science. As early as 1636, optical questions were featuring in his correspondence with the Earl of Newcastle, and the treatment of the sensible qualities was mechanical: ‘whereas I use the phrases, the light passes, or the colour passes or diffuseth itselfe, my meaning is that the motion is onely in the medium, and light and colour are but the effects of that motion in the brayne’. Perhaps more accurately, these effects were supposed to be the effects of the motion of the medium transmitted to the animal spirits in the brain.

Exactly how early Hobbes arrived at his conception of the workings of light is not entirely clear. A short treatise dated to 1630 and originally attributed to Hobbes is now thought by some scholars to be the work of someone else in the Cavendish circle. It contains doctrines different from writings that are more certainly ascribed to Hobbes, and which probably belong to the 1640s: the *Tractatus Opticus* I and II. It also contains as ‘principles’ formulations that these later optical writings present as mere hypotheses. Whether the writings of the 1640s represent only a change of mind or whether they are Hobbes’ first extended publications in optics, they show him adopting a mediumistic theory of the propagation of light based on the idea of continuously expanding and contracting light sources. These displace contiguous parts of an ethereal medium of uniform density and set up a chain reaction to the eye. A resistance in the eye caused by a countervailing motion from the brain produces a phantasm of a luminous object - that is, in Hobbes’ terminology, light. Light is propagated instantaneously, as both luminous object and medium expand simultaneously. The account does without the postulation of an emission by luminous objects of species or replicas of themselves which subsequently inform the senses and permit perception. Instead, luminous objects illuminate by radiation: they, so to speak, send out rays or, more precisely, displace the medium along paths called ‘rays’. In their passage from luminous objects to the eye, rays are supposed to describe parallelograms. Hobbes used the properties of other geometrical figures described by rays of light passing through different densities to account for refraction. Colour he regarded as light perturbed by the internal motions of rough or coarse bodies on its way to the eye. The differences between the colours on the spectrum from blue to red he accounts for as the product of refraction plus restraint or reinforcement of the lateral motion of rays of light that

goes with refraction.

At no point in the process that starts with the motion of the luminous objects and ends in the production of the phantasm does Hobbes depart from a mechanical model of the causes of sense-perception. His mature theory of optics is through and through an account in terms of matter and motion. But between the *Tractatus Opticus I* and the *Tractatus Opticus II*, he seems to have revised his ideas about the organs of sensory perception. Phantasms are said to come from the heart, rather than to result from the clash of incoming motions with motions from the brain. By 1646, when Hobbes produced *A Minute or First Draught of the Optiques*, the most polished of his optical treaties, the main lines of his doctrine were settled. In addition to material on light and its propagation, refraction and reflection, there are accounts of various kinds of sensory error.

Physics, understood as the theory of the causes of appearances to sense and of the nature of the objects of sense, is in part an offshoot of optics. It is expounded at the end of *De corpore*. By an 'object' of sense Hobbes means an external body that registers in experience as being the subject of certain qualities, and that sets off the process culminating in an 'act of sense'. The object of sense is not an idea or a sense-datum or a mental image, though such a thing may be the medium in which the object of sense is registered. The greatest of the objects of sense is the world itself, as registered from some point within it. But only a few intelligible questions can be asked about the world, and these cannot be conclusively answered. One can ask whether it is of finite or infinite magnitude, whether it is full or contains empty space, and how long it has lasted. Only the second of these questions is open to a scientific answer, and even then only to a probable conclusion, while the others are for lawfully appointed churchmen to discuss. Hobbes thinks that probably there is no vacuum, that the world is full, but that some of the bodies that make it up are invisible: thus the ether and 'the small atoms disseminated through the whole space between the earth and the stars'. He adopts Copernican and Galilean hypotheses in chapter 26 of *De corpore* to explain the order, motion and relative position of the planets. He also infers explanations of, among other things, the passage of the seasons, the succession of day and night and 'the monthly simple motion of the moon'.

Hobbes goes on to consider the bodies between the earth and the stars. Foremost among these is 'the most fluid ether', which he proposes to regard as if it were first matter. He supposes that its parts only receive motion from bodies that float in them, and impart none of their own. The bodies in the ether are supposed to have some degree of cohesion or hardness and to differ from one another in shape, figure and consistency. Any more specific hypotheses about them Hobbes adopted only to explain particular phenomena. He is, however, willing to venture that many such bodies are 'unspeakably little' or minute, since God's infinite power includes a power infinitely to diminish matter. Assumptions about intersidereal bodies inform his theories of the phantasms appropriate to the different senses, not only light but also heat, sound and odour.

5 Ethics

After *physics*,' Hobbes writes in chapter 6 of *De corpore*, 'we must come to *moral philosophy*; in which we are to consider the motions of the mind, namely, *appetite, aversion, love, benevolence, hope, fear, anger, emulation, envy &c*; what causes they have and of what they be causes' ([1655] 1839 I: 72). The use of the term 'moral philosophy' for the doctrine of the motions of the mind is unfortunate; elsewhere Hobbes says that the precepts of his natural law doctrine add up to a moral philosophy. 'Ethics' is another label he sometimes uses, and it is preferable. The reason ethics comes after physics is that the motions of the mind 'have their causes in sense and imagination, which are the subjects of physical contemplation'. What Hobbes means is that when a body registers in a sensory representation - when, for example, a person sees something - the thing imparts motion to the innermost part of the organ of sight. One effect of the motion is to set up an outward reaction which produces visual experience. But there can be a further after-effect. As Hobbes puts it in chapter 8 of *Elements of Law*, the 'motion and agitation of the brain which we call conception' can be 'continued to the heart, and there be called passion'. The heart governs 'vital motion' in the body, that is the circulation of the blood. In general, when motion derived from an act of sense encourages vital motion, the sentient creature experiences pleasure at the sight, smell or taste of the object and is disposed to move its body so as to prolong or intensify the pleasure. If the object of the pleasure is at a distance, then the creature will typically move towards it. There is a symmetrical account of displeasure. This is an after-effect of the act of sense consisting of a hindrance of vital motion. A creature experiencing a hindrance of vital motion will try to counteract it, typically by retreating from the object of sense. Aversion consists of the small inner movements that initiate the evasive action, just as 'appetite' names the internal

beginnings of approach behaviour.

The pursuit of pleasure and the avoidance of pain are the basic drives recognized by Hobbes' psychology, and they determine the systems of valuation of different individuals. The individual takes as good what it has learned to pursue and regards as bad what it has learned to avoid. In developing a system of valuation, a creature is not discovering an objective distinction in nature between things that are good and things that are bad. Nothing is good or bad independently of its effects on creatures, and the effects may vary from creature to creature. At most, things are good or bad *to* individuals, not good or bad 'simply and absolutely'. In the same vein, Hobbes denies that in the sphere of good and bad things there is one that is the highest and whose attainment constitutes happiness. Instead, there are many different goods for many different individuals. Becoming happy in life is not a matter of being successful in the pursuit of one favoured good, but of being continually successful in the pursuit of many.

Hobbes' account of the constraints on the pursuit of human happiness is the connecting link between his theory of the motions of the mind and his moral and political philosophy proper. To attain happiness, people need to know what goods to pursue and how to pursue them. But in the absence of a science of good and evil, pleasure is their main criterion of the good, displeasure the main criterion of bad. Both pain and pleasure, however, are unreliable guides to the good and bad. A person may find a thing pleasant on one occasion and call it 'good', only to change their mind later. Two people can react differently to the same thing, so that it produces pleasure in one and pain in the other, and is called 'good' and 'not good' simultaneously. Pleasure biases judgment in favour of the nearer and more intense good, even if the cost of pursuing this good is displeasure later, and so on. Part of the correction to these distortions is to judge the good of various things not by how they feel when they are enjoyed or shunned, but by the consequences of enjoying or shunning them. If the costs of the consequences outweigh the present benefits, then a supposed good may be merely an apparent good. Again, if someone detached from the pursuit or avoidance of a thing can judge it good or bad, then it may really be so; while if no-one else can see the attraction or repugnance, it may be illusory. Hobbes thinks only science can supply knowledge of the consequences of actions needed to counterbalance valuations derived from pain and pleasure; and he thinks science does not come naturally to people. Abiding by the value judgments of arbitrators does not come easily either, since people are attached to their valuations and unwilling to lose face by deferring to the judgments of others.

6 Politics: the state of nature

Despite the inconsistency in individual value judgments over time, and between the value judgments of different people at the same time, Hobbes thinks that there are some evils that are so large, and that interfere so markedly with everyone's pursuit of happiness, that practically no-one would knowingly pursue a course of action that resulted in them. War is such an evil, and Hobbes thinks he can show that if everyone makes themselves their own judge of what to pursue in the name of happiness, everyone will be involved in war. His argument to this conclusion is at the same time an argument for people to be guided by a judgment other than their own about what is best for them, the judgment of an existing civil power if they live in an existing commonwealth, the judgment of an as-yet-to-be designated civil power if they live outside any commonwealth.

The argument for the inevitability of war starts with assumptions about what is useful to the achievement of any goal. What is useful, no matter what good is being pursued, no matter whether the good is real or merely apparent, is power - that is, present means to future ends. 'Power' covers the physical capacities of individual agents, and also friends, riches and reputation. Not only is power in any form useful, but there can never be, according to Hobbes, too much power at the disposal of an agent in the nature of things. The reason is not that each agent naturally has an insatiable hunger for power, but that each agent is in competition with other agents for other goods, and any advantage one competitor temporarily has over another can, in principle, be overcome. The naturally strong can be toppled by a number of weak people who join forces; the man who has no enemies can be made into an object of hate with a well-judged campaign of character assassination; the wealthy can be robbed or swindled of their riches, and so on. Not only is it useful to acquire more and more power, but people cannot be blamed for doing so if all that organizes their activity in life is the pursuit of felicity.

Felicity is continual success in one's undertakings, whatever they may be. If what one undertakes is to do down one's competitors, then any means that helps to achieve it will be permissible. Or if, as is more likely, one aims at something else, doing down one's competitors can still often promote one's goal. Even the moderate man who wants only a small share of the good life can have reason to resort to foul means if he thinks he will lose

everything by playing fair with rivals. And he cannot be sure he does not risk losing everything if he plays fair. In general, the goal of felicity requires one to get an advantage and keep it. Disabling others is a means of keeping the advantage; the outright elimination of competitors is even surer. Because these facts can be discovered by everyone, everyone who pursues felicity must bargain for severe insecurity and even worry about survival. Struggling for survival is far removed from felicity, but the pursuit of felicity, no holds barred, can quickly turn into a struggle for survival. Or to put it Hobbes' way, in the state of nature, people who pursue felicity are in a condition of war.

The argument does not depend on the idea that every human being is naturally selfish. It is true that in *De cive* Hobbes paints an unflattering picture of ordinary human behaviour, emphasizing the tendency of people to look out for themselves, to say one thing to other people's faces and another thing behind their backs, the tendency to think very well of their own opinions, but poorly of the views of others, and to fight over trivialities. This is all ordinary human behaviour, but it is not the behaviour of absolutely every human being. That it is so ordinary is enough, in Hobbes' view, to overturn the Aristotelian idea that human beings are by nature fit for society, but he is not claiming that human beings are uniform, or that their behaviour is uniformly antisocial. Hobbes recognizes a variety of temperaments in human beings, and his state of nature encompasses the vainglorious as well as the moderate. The vainglorious will seek to dispossess others because having more than anyone else is an end in itself. Moderates will go on to the attack because they want only a little and fear that the greedy will take even that. Others again will be at odds because they want something that cannot be shared. Whatever the cause, the general effect will be insecurity, and with insecurity goes many unattractive things - not only feelings of fear, but loss of society, loss of production, loss of technology, loss of art, loss of everything that enables human beings to rise above a life of bare subsistence and savagery. Life in the state of war is, in Hobbes' famous phrase, 'solitary, poor, nasty, brutish, and short'.

Is there no such thing as virtue to keep people from pursuing felicity ruthlessly? Hobbes thinks that precepts enjoining the moral virtues - what he calls 'the laws of nature' - are discoverable even in the state of nature, but people are not morally obliged to act on them if they run the risk of dying as a result: the most basic law of nature is to preserve oneself, and there is an inalienable 'right of nature' to be one's own judge of how to secure one's own preservation and wellbeing. This right may be laid down in the interest of self-preservation, but never at the cost of self-preservation. So if one has reason to think that others will take advantage of one's keeping agreements, or of showing gratitude, of not being judge in one's own cause, of being forgiving and so on through the rest of the virtues, one is not obliged to behave in those ways. One is not obliged to act in a way which will advertise one's vulnerability to the unscrupulous. It is enough that one is willing to behave virtuously if it is not too dangerous.

7 Politics: the commonwealth

The answer to the problems of life in the state of nature is an agreement by most in it to delegate their right of nature to a person, or body of persons, empowered to secure the many against physical attack and against the severe deprivations of the state of nature. That person or body of persons is empowered by a collective submission of the wills of the many to the will of the one or few. The many agree to be guided in their behaviour by the laws of a sovereign, on the understanding that this is a more effective way of securing their safety than individual action in the state of nature. The many lend their wills to the sovereign both as potential enforcers of the law against lawbreakers and as an army of defence against foreign invasion. They lend their wills by doing only what is permitted by the sovereign's law and refraining from what the law prohibits. The law in turn expresses the sovereign's judgment regarding who should own what, who should teach what, how trade may be conducted, how wars should be waged, who should be punished and by what method of punishment, and who should be rewarded and the scale of the reward given.

The sovereign's judgment prevails because it, uniquely in the commonwealth, is still allied to a right of nature. Everyone who is subject to the sovereign thereby delegates their right of nature to the sovereign, but not in return for any forfeit or transfer of right by the sovereign himself. It is true that the commonwealth dissolves-that the obligation not to retract the right of nature lapses - if the sovereign is not able to secure the many against life-threatening incursion. But short of a reversion to the state of nature, the state in the person of the sovereign has a claim to expect the compliance of the many. The many owe it to one another to comply, because they agree between themselves to be law-abiding in return for safety if everyone else is law-abiding. They also owe it to the

sovereign, at least for the time that he succeeds in making and keeping the peace, because they voluntarily and publicly submit to the sovereign, signifying to him that they will do what he decrees should be done for their safety.

Hobbes' idea that the sovereign's law can justly reach into every sector of public life had clear application to the questions being debated during the Civil War period in England. Those who complained that it was wrong for Charles I to appropriate property, to billet troops at will, to raise taxes without the consent of parliament, were given a theory that legitimized those actions. According to the theory, the limitations on a king's powers indicate that a state of nature, with its potential for open war, still prevails. Either the powers of government are separated (in which case the contention between, say, king and parliament reproduces the contention between individuals in the pre-political condition), or else the powers of government are not separated, but are limited by the rights of the subjects (in which case the right of nature has not really been transferred, and people are still liable to prefer their own judgment about what is best for them to the judgment that they have agreed to be guided by, with the same potential for slaughter).

Hobbes' theory permits the sovereign to regulate public life very stringently, but his message to sovereigns - there is no doubt that *Leviathan* in particular was intended to be read by heads of state - was not that it was *wise* to regulate public life very stringently. To begin with, there were limits to what laws could do: belief could not be commanded, so a certain tolerance of freedom of thought was inevitable. Again, people could not be expected to risk their lives in order to obey the law, as that would leave them no better off in the state than outside it. So laws that impoverished people to such an extent that they were starving, having to steal in order to live, were ill-conceived. Likewise forced military service, if it were likely to lead to death, might reasonably be seen as unacceptable according to the terms of Hobbes' social contract. Even a regime of law that secured most people from theft and common assault, but that confiscated all income above a measly minimum, could be seen as a failure by the sovereign to come up with what the many bargained for in entering the state. What the many bargain for is safety and, as Hobbes explains in chapter 30 of *Leviathan*, 'safety' signifies more than a 'bare preservation': it means a modicum of wellbeing over and above survival.

The arguments from prudence against overregulation by the sovereign are also arguments against iniquitous practice by the sovereign. Hobbes distinguishes between iniquity and injustice. The sovereign does no injustice to his subjects if he decides to claim as his own all the land in a particular county or all the houses in a village: in creating the sovereign, his subjects give him the power to decree who is the owner of what. For all that, the sovereign may act iniquitously in the sense that he allows his own appetites and interests to count for more than those of anyone else, and so makes himself, for selfish reasons, the owner of more land than anyone else. There is a law of nature against iniquity, and therefore a law that decrees that the sovereign has to *try* to be equitable. But efforts are one thing; actual behaviour is another. The law of nature is not binding on the sovereign's behaviour, since he retains the right of nature and is authoritative about what to do for the best. If, in his opinion, it is for the best to behave iniquitously, then no other free agent, still less one of his subjects, can blame him for behaving accordingly. But the fact that his iniquitous acts are in this sense blameless does not mean that they are wise. If appropriating everyone's land makes people rebellious, albeit unjustly rebellious, then appropriating other people's land may have greater costs than benefits: it is subversive of the sovereign's power, which depends on the willingness of others to obey him.

Regarding the practice of religion, the relationship between church and state is a central preoccupation of *Leviathan*. Hobbes insists there that it is for the sovereign to decide whether people can join together for purposes of worship - that is, whether a given church can lawfully exist in the commonwealth. And he appears not to have been in favour of the establishment of a plurality of churches:

> But seeing a Common-wealth is but one Person, it ought also to exhibite to God but one Worship; which then it doth, when it commandeth it to be exhibited by Private men, Publiquely. And this is Publique Worship; the property whereof, is to be *Uniforme*: For those actions that are done differently, by different men, cannot be said to be a Publique Worship. And therefore, where many sorts of Worship be allowed, proceeding from the different Religions of Private men, it cannot be said there is any Publique Worship, nor that the Commonwealth is of any Religion at all.

(Hobbes [1651] 1839 III: 354)

He goes on to say that 'whereas there be an infinite number of Actions and Gestures, of an indifferent nature; such of them as the Common-wealth shall ordain to be Publiquely and Universally in use, as signes of Honour, and part of Gods Worship, are to be taken and used for such by the Subjects'. It is hard to gather from these passages even tacit approval for a pluralistic form of national religious life. On the contrary, it is strongly implied that unless all members of the commonwealth worship in the same way, it will be doubtful not merely which religion the commonwealth observes but whether it observes any. It is as if Hobbes thinks that in a babble of different religious rites there will be no clear sign of honour from the commonwealth to God. For a clear signal to be sent, the same thing must be transmitted by everyone in the commonwealth. This 'clear-signal' justification for uniformity is not as anti-tolerationist as a justification that holds that all but the appointed religious rites are idolatrous, but it lends support all the same to a highly restrictive form of public religious worship.

As chapter 12 of *Leviathan* shows, Hobbes was aware that people living together but worshipping differently could ridicule or belittle one another's ceremonies and come into conflict. This is another reason for the secular authority to regulate public worship. It is also a reason for worshippers to take religious ceremony out of the public arena altogether, and preserve their differences in private. Hobbes has no quarrel with this sort of privatization of religious practice, so long as it is thoroughgoing: driving it out of the public arena means driving it well out. To obviate regulation, worship must be not only be private (that is, practised openly by a private person) but practised by a private person in secret. As Hobbes says in chapter 31 of *Leviathan*, private worship 'when secret, is Free; but in the sight of the multitude, it is never without some restraint, either from the Laws, or the opinion of men, which is contrary to the nature of Liberty'.

Not only actions required by religious rites can be driven underground if they might disturb the peace; freedom of action in matters indifferent to religion can also be open to regulation, as chapter 31 of *Leviathan* makes clear. Arguably it is indifferent whether prayers are said in Latin or in English; arguably it is indifferent whether services are conducted by married or by celibate men; but notoriously, these are things that people look askance at or insist upon, and about which they can come to blows. For this reason, if for no other, there is a reason for the sovereign to declare what the language shall be, and who shall preside at services.

What is in the sight of the multitude and in the control of the religious is one thing; what is out of sight and uncontrollable is another: 'Internal faith is in its own nature invisible, and consequently exempted from all humane jurisdiction', Hobbes says in chapter 42 of *Leviathan*. Humane jurisdiction is not just secular jurisdiction, but also that of a body charged by a church with the inquisition of believers. Beliefs in general are not subject to the will he says in *De Politico Corpore*, a pirated edition of part two of the *Elements of Law* (1839 IV: 339). And although salvation depends on believing some things and not others, it is hard to be sure which things have to be believed beyond an uncontroversial minimum. For all of these reasons Hobbes is against the policing of religious belief, and against preferment for any one creed. It is in connection with the policing of belief rather than religious practice that his views come close to those of Independents, who in seventeenth-century England favoured a relatively loose, relatively tolerant organization of religious life, in particular a life outside a unitary Church of England. For when he appears in *Leviathan* (chapter 47) to side with the Independents in the Primitive church it is over each person deciding whose preaching to follow, not over many different religions being openly practised ([1651] 1839 III: 695). And in countenancing a variety of religious persuasions Hobbes is not so much showing tolerance as denying the importance to civil order of what goes on below the threshold of visible action.

8 Problems with Hobbes' political theory

In order to legitimize the powers of sovereigns, Hobbes invites his readers to think of sovereigns and states as the creations of free, self-interested people. The condition of subjection to a sovereign, even if it is not entered into by an original contract, can nevertheless be freely endorsed by each subject, since there is a good argument from self-interest for the condition. The argument says that the alternative to subjection is a dangerous chaos, which is infinitely worse than an intrusive but protective civil power. This is the argument directed against people who are already subjects; is the same argument effective when directed at people who do not yet belong to a state, who are in a state of nature? The issue can be sharpened by pointing out that the process of trading the state of nature for the commonwealth involves each person giving something now for the sake of a benefit later. Each person agrees to lay down their right of nature if everyone else will do likewise for the sake of peace. Granting that the condition of peace is better for each than the condition of war, is it not even better for anyone who can get away with it, to

retain their right of nature while others give away theirs? Is it not better to pretend to lay down that right and then to take advantage of those who genuinely do so? If the answer to this question is ‘Yes’, how can the best outcome, from the point of self-interest, be one in which everyone performs and lays down their right of nature?

The question is taken up in a famous passage in chapter 15 of *Leviathan* where Hobbes replies to the fool who pretends that there is no such thing as justice. Commentators have likened it to the question posed by ‘prisoners’ dilemmas’ where, for example, the outcome that would be best for each of two prisoners is for the other to confess and solely take a punishment for a crime, but where it turns out to be rational for each to confess and receive a punishment less severe than the maximum. The question is how this ‘lesser’ outcome can be the better one. In the case of the opportunistic non-performer of covenants discussed by Hobbes, the answer is that there is more security in the performance than in the non-performance of covenants. Someone who takes advantage of another’s laying down the right of nature can only do so once and expect to get away with it. And the temporary advantage they gain may in any case not counterbalance what they will lose by being opposed by all those whose trust is threatened or betrayed.

Another problem with Hobbes’ theory turns on the supposed moral urgency of each person’s laying down the right of nature. Hobbes thinks that the biggest threat to the stability of states is the existence of too much scope for private judgment. The more each person is entitled to think for themselves in matters of wellbeing, the worse it turns out for everyone. This implication is supported in Hobbes’ theory by a supposedly scientific understanding of the diversity of the passions and the way that the passions get the better of judgment in human beings. By delegating their power of judgment to someone who is not affected by the individual passions of the people ruled over, people actually get access to a more effective (because more dispassionate) means of securing themselves than their own individual judgments. But, by the same token, they forgo any intellectual contribution to public life. They do not contribute to the state as citizens but as subjects only: political life for the many consists solely of submission to law. It is by their passivity rather than by the application of their powers of judgment that people promote the public good. This may have seemed persuasive in a time when the Biblical example of Adam and Eve would have been widely understood to illustrate the dangers of private judgment of good and evil, but to contemporary sensibility it verges on the paranoid. In fact, Hobbes’ point is not quite that the judgments of human beings about their wellbeing can never be trusted, but rather that their *prescientific* judgments cannot be trusted. Prescientifically, people are moved by their feelings of pleasure and displeasure to call things ‘good’ and ‘bad’ - few have either the resources or the circumstances to be taught any better. But there is a better conception to be inculcated: Hobbes indicates that it consists of showing how things that are genuinely good, as opposed to pleasant, promote peace or self-preservation, while things that are genuinely bad, as opposed to unpleasant, are conducive to war and self-destruction.

9 The scientific status of Hobbes’ ethics and politics

The main lines of Hobbes’ political philosophy include the idea that the commonwealth is a solution to the ever-present threat of war in the passionate make-up of human beings, and that the commonwealth is made by delegating the right of nature to a sovereign power with unlimited power. This summarizes a theory worked out in very great detail, a theory Hobbes always regarded as ushering in the scientific treatment of morals and politics. What made the theory scientific? A number of answers get support from Hobbes’ writings. The scientific status of politics is sometimes said to be owed to its derivation, in some sense, from Hobbes’ natural science. Again, Hobbes’ use in civil philosophy of a method applicable to natural bodies and bodies politic alike is sometimes thought to be crucial to its scientific status. These answers are consistent with some texts but sit uneasily with others. First, although Hobbes thought that there was a way of approaching the principles of morals and politics from a starting point in the workings of sense and imagination (which were treated of by physics), he consistently denied that civil science *had* to be approached by way of physics. In chapter 6 of *De corpore* he says that people entirely innocent of physics, but who enjoy introspective access to their own passionate states, are able to see in themselves evidence for the truth of the theory of human nature in the civil science. Something similar is said in the Introduction to *Leviathan*. In the same vein there is the explanation of his having been able to publish *De cive*, the third volume in his trilogy, without having first expounded the principles of parts of philosophy that were prior to politics. Hobbes said that this was possible because civil philosophy depended on principles of its own. What ties together all of these remarks is a belief in the autonomy of civil science, a belief that is not seriously called into question by his saying that the two principal parts of civil philosophy were alike in applying a certain sort of

method to the investigation of bodies - bodies politic on the one hand and natural bodies on the other.

When Hobbes says that each part of philosophy deals with bodies, he makes clear that the two kinds of bodies are 'very different' from one another. And there is no evidence that 'body', when applied in the phrase 'body politic', is supposed to mean 'space-occupying thing existing without the mind'. In other words, there is no evidence that bodies politic are bodies in any more than a metaphorical sense. Finally, it is not clear that Hobbes thought that the scientific status of his politics was made more credible by an analogy between bodies politic and natural bodies. It is not as if he thought that natural bodies were well-understood scientifically, and that bodies politic might in principle be as well understood if the methods of physics were applied to them. On the contrary, Hobbes always thought that the properties of human artefacts, such as bodies politic, were much better understood than the properties of natural bodies, which had God's inscrutable will behind them.

It would be a mistake, however, to think that civil science for Hobbes was primarily an exercise in the investigation of the properties of bodies. It was an exercise in putting our judgments about what we ought to do on grounds that were far more solid than pleasure and pain. Good and bad were a matter of what conduced or interfered with self-preservation or peace, not how it felt to do or get this or that. The core of Hobbes' civil science is an attempt to recast the precepts of morality - the laws of nature - as instruments of peace, and to show how the ingredients of war are latent in any project for the pursuit of happiness. The scientific status of the doctrine of the laws of nature - the ground of its claim to be called moral philosophy - was its conforming to the pattern of a deductive system, based on two fundamental laws of nature and the rest derivative. Similarly with the deduction of the rights of sovereigns from the goal of peace. The scientific status of the argument for the inevitability of war consisted in its proceeding from principles about the passions. But these principles were by no means the property of physics or physicists; they were available in each person's introspective self-knowledge.

See also: Human nature; Justice; Public interest

TOM SORELL

List of works

A new edition of Hobbes' works is currently in preparation. Until it is complete, the standard edition will remain that of Sir William Molesworth, which omits many manuscripts and a great deal of Hobbes' correspondence.

Hobbes, T. (1839) *The English Works of Thomas Hobbes*, ed. W. Molesworth, London: John Bohn, 11 vols.(The spelling and punctuation are modernized in this collection.)

Hobbes, T. (1845) *Opera Latina*, London: John Bohn, 5 vols.(Hobbes' works in Latin.)

Hobbes, T. (1628) *Thucydides' History of the Pelopennesian Wars*; repr. ed. R. Schlatter, New Brunswick, NJ: Rutgers University Press, 1975.(The Petition of Right of 1628 may have led to the publication of this translation.)

Hobbes, T. (1630?) *The Short Tract*, in *The Elements of Law Natural and Politic*, ed. F. Tonnies, London: Simpkin & Marshall, 1889.(This work, which appears as Appendix I, is attributed by Tonnies to Hobbes.)

Hobbes, T. (1640) *The Elements of Law Natural and Politic*, ed. F. Tonnies, London: Simpkin & Marshall, 1889. (The correct and intact edition of material also reproduced as *Human Nature* and *De Corpore Politico* in Molesworth, volume 4.)

Hobbes, T. (1641) *The Third Set of Objections to Descartes Meditations*, in *The Philosophical Writings of Descartes*, vol. 2, trans. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1985.(Hobbes wrote these objections as a member of Mersenne's circle in Paris.)

Hobbes, T. (1642) *De cive*, in *The English Works of Thomas Hobbes*, vol. 2, ed. W. Molesworth, London: John Bohn, 1839.(Molesworth follows an English translation, not by Hobbes, of the original Latin.)

Hobbes, T. (1643) *Thomas White's De Mundo Examined*, trans. H Jones, Bradford: Bradford University Press, 1976.(A recently discovered early foray into metaphysics.)

Hobbes, T. (1646) *A Minute or First Draught of the Optiques*.(The most polished of Hobbes' optical treaties)

Hobbes, T. (1651) *Leviathan, or the Matter, Form and Power of Commonwealth, Ecclesiastical and Civil*, in *The English Works of Thomas Hobbes*, vol. 3, ed. W. Molesworth, London: John Bohn, 1839; repr. ed. R. Tuck, Cambridge: Cambridge University Press, 1991; repr. ed. E. Curley, Chicago, IL: Hackett Publishing Company, 1994.(Hobbes' masterpiece.)

Hobbes, T. (1654) *Of Liberty and Necessity*, in *The English Works of Thomas Hobbes*, vol. 4, ed. W. Molesworth, London: John Bohn, 1839.(Classic discussion of freedom and determinism.)

Hobbes, T. (1655) *De corpore*, in *The English Works of Thomas Hobbes*, vol. 1, ed. W. Molesworth, London:

John Bohn, 1839.(Follows an English translation of 1656.)

Hobbes, T. (1656a) *The Questions concerning Liberty, Necessity and Chance, Clearly Stated and Debated between Dr Bramhall, Bishop of Derry, and Thomas Hobbes of Malmesbury*, in *The English Works of Thomas Hobbes,* vol. 4, ed. W. Molesworth, London: John Bohn, 1839.(Bramhall had accused Hobbes of atheism and irreligion.)

Hobbes, T. (1656b) *Six Lessons to the Professors of Mathematics… in the University of Oxford*, in *The English Works of Thomas Hobbes*, vol. 7, ed. W. Molesworth, London: John Bohn, 1839.(Testy reply to some Oxford antagonists.)

Hobbes, T. (1658) *De homine*; trans. and abridged B. Gert, T. Scott-Craig and C.T. Wood as *On Man*, in *Man and Citizen*, New York: Humanities Press, 1972.(A complete translation of *De homine* is in preparation at Cambridge University Press.)

Hobbes, T. (1668) *Behemoth, or the Long Parliament, Dialogue of the Civil Wars of England*, in *The English Works of Thomas Hobbes*, vol. 6, ed. W. Molesworth, London: John Bohn, 1839.(Hobbes' own history of the English Civil War.)

Hobbes, T. (1678) *Decameron Physiologicum*, in *The English Works of Thomas Hobbes*, vol. 7, ed. W. Molesworth, London: John Bohn, 1839.(Restates some of the methodology and principal results of the physical sections of *De corpore*.)

Hobbes, T. (1681) *Dialogue between a Philosopher and a Student of the Common Laws of England*, in *The English Works of Thomas Hobbes*, vol. 6, ed. W. Molesworth, London: John Bohn, 1839.(A discussion of political philosophy.)

References and further reading

Aubrey, J. (1898) *Brief Lives, Chiefly of Contemporaries, Set Down by John Aubrey between the Years 1669 & 1696*, ed. A. Clark, Oxford, 2 vols.(Contains an early biographical sketch of Hobbes.)

Johnston, D. (1986) *The Rhetoric of Leviathan*, Princeton, N.J: Princeton University Press.(A study of the tensions between the scientific pretensions of some of Hobbes' political writings and their persuasive purpose.)

Mintz, S. (1970) *The Hunting of Leviathan*, Cambridge: Cambridge University Press.(Deals with the context and early reception of *Leviathan*.)

Peters, R. (1956) *Hobbes*, Harmondsworth: Penguin.(A clearly written study of all Hobbes' philosophy.)

Rogers, G.A.J. and Ryan, A. (eds) (1988) *Perspectives on Thomas Hobbes*, Oxford: Clarendon Press.(A collection of mostly accessible papers about Hobbes' metaphysics, philosophy of science and politics, published to mark the four-hundredth anniversary of his birth.)

Sommerville, J. (1992) *Thomas Hobbes: Political Ideas in Historical Context*, London: Macmillan.(A very clear and comprehensive account of the historical context of Hobbes' political writings.)

Sorell, T. (1986) *Hobbes*, London: Routledge.(A study of Hobbes' metaphysics and politics against the background of his philosophy of science.)

Sorell, T. (ed.) (1995) *The Cambridge Companion to Hobbes*, Cambridge: Cambridge University Press.(Articles on all aspects of Hobbes' thought, including his mathematics, optics and his theories of law and religion. Contains an authoritative biographical account by Noel Malcolm.)

Tuck, R. (1989) *Hobbes*, Oxford: Oxford University Press.(A short study claiming the importance for Hobbes' thought of philosophical scepticism.)

Zarka, Y. (1987) *La décision métaphysique de Hobbes - Conditions de la politique*, Paris: Vrin.(A study of Hobbes' system, bringing to bear the great sophistication of French scholarship on Hobbes. The author is editor of the Vrin edition of Hobbes.)

Hohfeld, Wesley Newcomb (1879-1918)

W.N. Hohfeld, US law professor and proponent of analytical jurisprudence, was responsible for one of the most influential analyses of the concept of a right in legal and moral philosophy. He offered to resolve all complex legal relations into a few simple and elementary ones, often confusedly referred to as 'rights'.

For Hohfeld, 'right' in its strict sense necessarily correlates with an identical duty owed by a second party. Rights differ from 'privileges' (where a person is not duty-bound), from 'powers' (where a person is legally enabled to bring about change in some legal relationship) and from 'immunities' (where a person is legally protected from another person's purported exercise of legal power designed to effect some change in a given legal relationship).

Hohfeld constructs two sets of interrelated 'correlatives' and 'opposites'. The first set comprises 'right' or 'claim' (sometimes 'claim-right' in the subsequent literature), with its correlative 'duty' and its opposite 'no-right'; 'privilege' (or, preferably according to most commentators, 'liberty') is in turn the opposite of duty and the correlative of no-right. These always concern some action or abstention by a person. One has a duty to act, for example, to make a payment, for another; then, correlatively, that other has a (claim-)right to the payment. The opposite would be if the second party had no right to be paid by the first - and in that case the first would be at liberty to, or would have the privilege to, withhold payment.

The second set concerns power and its correlative 'liability' and opposite 'disability'. Legal power always refers to some legal relationship, for example, a contractual right of A to some payment by B. In this case, the law may enable A in some way to vary the existing legal relationship, for example by voluntarily renouncing right to payment. Then B has as correlative the liability to have the privilege of non-payment conferred on them by A. In the opposite case, we consider the possibility of an absence of power. For example, B normally lacks power to revoke unilaterally A's contractual right. B thus has a 'disability' to bring about this change; A correlatively enjoys an immunity from it.

Notable in this analysis is the atomistic character of each legal relationship; everything is brought down to single relationships of distinct persons in respect of specific actions or abstentions, or particular (purported) power-exercising acts. Hohfeld indeed argues that more complex situations, such as those we encounter in property-ownership or in corporate relations, are resolvable without residue into aggregate sets of simple atomic relationships belonging to his categories; hence his claim that these are the 'fundamental' legal conceptions. This claim has been challenged, and along with it the claim that Hohfeld's analysis merely clarifies or purifies existing legal usages; Hohfeld was more a reformer than an analyst of current legal usage. For those purposes for which it is useful exhaustively to investigate the totality of relations between single individuals for some legal purpose, the analysis provides a valuable tool. But it is less valuable for other purposes, such as elaborating a catalogue of fundamental human or constitutional rights, or constructing a theory of property.

See also: Law, philosophy of; Legal concepts §3; Rights §2

NEIL MacCORMICK

List of works

Hohfeld, W.N. (1919) *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, ed. W.W. Cook, New Haven, CT: Yale University Press, 1966.(Hohfeld's basic statement of position, comprising a series of highly influential and attractively written articles from learned journals.)

References and further reading

Paton, G.W. (1972) *A Text-book of Jurisprudence*, ed. G.W. Paton and D.P. Derham, Oxford, Clarendon Press, 4th edn.(Chapters on rights (18-21) give a good account of Hohfeld in the light of forerunners and successors.)

White, A.R. (1984) *Rights*, Oxford: Clarendon Press.(Contains a root-and-branch critique of Hohfeld and those who have been tempted to follow his doctrines, written in the spirit of linguistic-analytical philosophy.)

Holcot, Robert (*c*.1290-1349)

Earlier scholars labelled Robert Holcot a sceptic, a nominalist and an Ockhamist. In fact, the work of this English Dominican friar is too complex and original for any simple label. What can be said with justice is that Holcot belongs among philosophers who work not at developing their own systematic philosophies but at criticizing the speculative efforts of their predecessors. Holcot questions the scope of Aristotelian logic, the extent of theological knowledge and even the extent to which we can have knowledge of the world around us.

Holcot studied theology at Oxford, delivering his commentary on Peter Lombard's *Sentences* over two academic years there, beginning probably in 1331. This was a fruitful period for philosophical theology at Oxford. Holcot's lectures overlapped with those of William Crathorn, a fellow Dominican but nevertheless a fierce rival, and the influential Franciscan Adam Wodeham. Holcot himself was highly influential, and his biblical commentaries, especially his commentary on the Book of Wisdom, were widely circulated throughout the later Middle Ages and beyond.

Besides the works just mentioned, Holcot also composed a series of quodlibetal questions. The first of these quodlibets is an ideal introduction to Holcot's philosophical views. In it he asks whether theology is a science. His surprising answer is that it is not, or at least not one we can naturally acquire in this life. He means that there are no theological truths we can know naturally with the sort of certainty required for *scientia*; that is, no theological truths are demonstrable (by us, on our own, in this life). First and foremost, we cannot demonstrate that God exists; cosmological arguments rest on contingent facts that are not themselves demonstrable. Moreover, to prove God's existence, one must prove the existence of a thing that satisfies all the articles of the faith: this God must be shown to be infinite, triune, the creator, and so on (see God, arguments for the existence of). Anselm's ontological argument comes in for careful criticism at this point because it takes off from a richer conception of God, one from which many divine attributes might be derived.

In this quodlibet, Holcot notes rather casually that he would extend these claims to all incorporeal things, including the human soul. There is therefore no science of the human soul, and (presumably) we cannot demonstrate even that we have souls. Holcot's discussion of God's causal powers suggests even more strongly that he is a thoroughgoing sceptic. There he acknowledges (as it seems any Christian philosopher must) that God can intervene in the natural order, causing fire to stop giving heat or, as he luridly puts it, causing the entire structure of the world to take on the appearance of a fly. Such possibilities lead Holcot to claim that there is always an element of uncertainty in our empirical beliefs, and in his commentary on the *Sentences* he goes on to conclude that there can be no *scientia* regarding them.

Much has been made of these claims, and if *scientia* is translated as 'knowledge' then Holcot appears to be a quite radical sceptic. However, his position is more subtle than this. The remainder of his argument makes it clear that Holcot does not believe our empirical beliefs lack all warrant: 'It is sufficiently established [*persuasum*], to me, that God has not made such alterations' (*In quatuor libros Sententiarum* IV q.3 ad1); 'that A is the cause of B is held with probability' (*Determinationes* 3). From the fact that such beliefs are not held with complete certainty, we cannot infer that they should not be held at all, and Holcot insists that we have good reasons for believing as we do about the world. He does not go on to ask the questions a modern reader would like to have answered, such as what exactly those good reasons are and whether such beliefs can still count as knowledge. Nevertheless, Holcot is clearly not entirely the thoroughgoing sceptic he might seem to be.

Relatively little of Holcot's work is concerned with establishing the limits of *scientia*. Indeed, he argues in his first quodlibet that theology is still very much worth studying, even if it is not a science, because through it we can acquire a stronger, broader and more explicit faith. True to his words, Holcot involves himself in a wide range of theological and philosophical controversies. One of his more distinctive claims, and one with far-ranging implications, is his rejection of any sort of metaphysical distinction other than a 'real' distinction. This point receives special attention in his discussion of the divine perfections and the Trinity. Following William of Ockham and John Duns Scotus, Holcot rejects out of hand the suggestion of Thomas Aquinas that conceptual distinctions (distinctions 'of reason') account for the differences between, for example, divine wisdom and divine goodness. However, Holcot also rejected the proposals of Scotus and Ockham, each of whom in his own way analysed the

divine essence in terms of ‘formal’ distinctions. ‘All that are formally distinct are distinct things,’ Holcot charges (*Determinationes* 10). By rejecting all non-real distinctions, Holcot places himself in a difficult position as regards the Trinity (see Trinity). Church doctrine was very clear in rejecting any real distinction among the divine Persons, and so Holcot was committed to saying that there are no distinctions at all within the Trinity. If so, however, then the following syllogism should be valid: ‘The Divine Essence is the Father, and the Divine Essence is the Son; therefore, the Son is the Father.’ This syllogism is not valid: the premises are true and the conclusion false. The threat that looms, then, is that Aristotle’s ‘entire syllogistic form would be defective’ (*Determinationes* 10). Holcot’s reply is to claim that Aristotelian logic is not universally applicable: ‘the logic of the faith has to be different from natural logic’ (*In quatuor libros Sententiarum* I q.5, ad1). However, this striking pronouncement, once developed, turns out to be less interesting than it sounds. Holcot does not offer anything like an alternative logic: what he gives us instead are fairly obvious rules for excluding terms that stand for ‘many things and also each one of them’ (*Determinationes* 10), a restriction that rules out terms that stand for God.

See also: Knowledge, concept of; Natural philosophy, medieval

ROBERT PASNAU

List of works

Many of the *Quodlibets* do not have definite numbers assigned to them, as the order of the questions is highly uncertain. Editions of several *Quodlibets* are given here. See also Molteni (1968) and Kennedy (1993) in the list of references and further reading.

Holcot, Robert (1331-3) *In quatuor libros Sententiarum (Commentary on the Four Books of Sentences)*, Lyons, 1518; repr. Frankfurt: Minerva, 1967.(Remains the only printed source for most of Holcot’s philosophical work, including a dozen questions from *Quodlibet* I, printed here as *Determinationes*.)

Holcot, Robert (1332) *Sex articuli (Six Articles)*, ed. F. Hoffmann, *Die ‘Conferentiae’ des Robert Holcot O.P. und die akademischen Auseinandersetzungen an der Universität Oxford 1330-1332*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters NF 5, Münster: Aschendorff, 1993.(Edition of the philosophically rich *Sex articuli*, also known as the *Conferentiae*.)

Holcot, Robert (1333) *Sermon*, ed. J.C. Wey, ‘The *sermo finalis* of Robert Holcot’, *Mediaeval Studies* 11, 1949: 219-23.(A humorous sermon delivered at the end of his Oxford lectures.)

Holcot, Robert (1333) *Quodlibet*, ed. J.T. Muckle, ‘"Utrum Theologia Sit Scientia". A Quodlibet Question of Robert Holcot O.P.’, *Mediaeval Studies* 20, 1958: 127-53. (Edition of *Quodlibet* I q.1.)

Holcot, Robert (1333-4) *Quodlibet*, ed. W. Courtenay, ‘A Revised Text of Robert Holcot’s Quodlibetal Dispute on Whether God is Able to Know More Than He Knows’, *Archiv für Geschichte der Philosophie* 53, 1971: 1-21.(Edits *Quodlibet* I q.6, which contains a detailed discussion of Holcot’s interesting view that the objects of knowledge and belief are not things in the world, but individual tokens of a proposition.)

Holcot, Robert (1333-4) *Quodlibets*, ed. H.G. Gelber, *Exploring the Boundaries of Reason: Three Questions on the Nature of God*, Toronto, Ont.: Pontifical Institute of Medieval Studies, 1983.(Edits *Quodlibeta* I qq.2-4, in which Holcot rejects all non-real distinctions and argues for limitations on the applicability of Aristotelian logic.)

Holcot, Robert (1333-4) *Quodlibets*, ed. P. Streveler *et al*., *Seeing the Future Clearly: Questions on Future Contingents*, Toronto, Ont.: Pontifical Institute of Medieval Studies, 1995. (Edits *Quodlibeta* III qq.1-3, as well as corresponding material from the *Sentences* commentary. K. Tachau’s introduction presents important new historical research.)

References and further reading

Gelber, H.G. (1974) ‘Logic and the Trinity: A Clash of Values in Scholastic Thought, 1300-1335’, Ph.D. thesis, University of Wisconsin.(Never published, but extremely valuable as a survey of the ways in which scholastics from Bonaventure through Holcot attempted to reconcile the Trinity with standard principles of logic.)

Hoffmann, F. (1972) *Die theologische Methode des Oxforder Dominikanerlehrers Robert Holcot (The Theological Method of the Oxford Dominican Scholar Robert Holcot)*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters NF 5, Münster: Aschendorff.(A detailed look at Holcot’s philosophical and theological method, with particular attention to his views on the status of theology as a science.)

Kennedy, L.A. (1993) *The Philosophy of Robert Holcot, Fourteenth-Century Skeptic*, Lewiston, NY: Mellen. (Valuable for the passages that it collects and translates, and for the seven quodlibetal questions it edits.)

Molteni, P. (1968) *Roberto Holcot, O.P. Dottrina della grazia e della giustificazione con due questioni*

quodlibetali inedite, Pinerolo: Editrice Alzani.(Edits *Quodlibeta* I q.9, III q.7.)

Moody, E.A. (1964) ‘A Quodlibetal Question of Robert Holkot, O.P., on the Problem of the Objects of Knowledge and of Belief’, *Speculum* 39: 53-74.(The text has been superseded by Courtenay’s edition, but Moody’s analysis remains well worth reading.)

Oberman, H. (1963) *The Harvest of Medieval Theology: Gabriel Biel and Late Medieval Nominalism*, Cambridge, MA: Harvard University Press. (Perhaps the best starting point for an overview of Holcot’s philosophical tendencies.)

Oberman, H. (1966) *Forerunners of the Reformation: The Shape of Late Medieval Thought*, New York: Holt, Rinehart & Winston.(Translates passages from the Wisdom Commentary concerning grace and justification.)

Schepers, H. (1970-2) ‘Holkot contra dicta Crathorn’ (Holcot Against the Claims of Crathorn), *Philosophisches Jahrbuch* 77: 320-54; 79: 106-136.(Assesses the philosophical dispute between Crathorn and Holcot, with particular reference to their dispute over the objects of knowledge. In German.)

Smalley, B. (1960) *English Friars and Antiquity in the Early Fourteenth Century*, New York: Barnes & Noble. (Offers a colourful account of Holcot’s life and literary methods.)

Hölderlin, Johann Christian Friedrich (1770-1843)

An outstanding German poet, Hölderlin is now widely recognized as one of the most important writers and thinkers of his time. After an initial period of critical neglect and relative public indifference, he eventually came to enjoy a privileged status in German cultural life at the beginning of the twentieth century, and has continued to exercise a profound influence on modern literature and critical thought. Hölderlin's work emerged within the context of Kant's Critical Revolution and developed in constant interaction with the German Idealist speculation this subsequently provoked, against the background of the French Revolution and the fundamental issues raised in its wake for an entire generation. Hölderlin was personally acquainted with many of the leading figures of the period, including Schiller, Goethe, Herder, Novalis, Fichte, Schelling and Hegel. His writings reflect and respond to many of the central philosophical concerns and themes of his time, in a highly original and prescient fashion. However, it is only since Wilhelm Dilthey that the importance of his specific influence on the intellectual development of Schelling and Hegel in particular has been fully recognized, so that he has even been described as the 'Doctor Seraphicus' of German Idealism. His work subsequently provided a frequent point of reference for the Frankfurt School's engagement with the heritage of German Idealism and the problem of the relationship between aesthetics, ethics and politics. More recently, Hölderlin has become a major presence in much contemporary critical and deconstructive theory, largely through the pervasive influence of Martin Heidegger.

1 Life and writings

Hölderlin grew up in a sheltered, intensely Pietist domestic environment. His mother directed his education towards a career in the ministry, a later source of much psychological conflict since he never subsequently gained the financial independence necessary to pursue his chosen literary career. In 1788 Hölderlin entered the Tübingen Stift, ostensibly to study theology, but was soon preoccupied with developments in contemporary literature and philosophy, and the beginnings of the French Revolution. He shed his earlier orthodox religious beliefs under the influence of Kant and the classicizing humanism of Friedrich Schiller, and became friends with his fellow-students Schelling and Hegel. His early poetry celebrated humanistic ideals using a rhyming, hymnic style indebted largely to Schiller. Hölderlin encapsulated his central interests at this time: 'Kant and the Greeks are my only reading' (1946-85 vol. 6: 128). His attempts to mediate his Graecophile enthusiasm with the rigorous republicanism of Fichte, and with the latter's radicalization of Kantian ethics, is documented in the mid 1790s through the various evolving versions of his first major work, the novel *Hyperion* (1798-9).

Almost alone among his writings *Hyperion* enjoyed a certain celebrity in Hölderlin's lifetime, becoming something of a 'sentimental' classic, not entirely unlike Goethe's *Werther*, by which it was probably influenced. Set in the contemporary context of the Greek struggle for national independence, the structure of the first-person epistolary narrative presents the spiritual evolution of its idealistic eponymous protagonist, combining the personal theme of transforming love with broader concerns for the nature of genuine political freedom. The hero moves from an initial pantheistic identification with the totality of life, through painful but educative experiences of failure and disappointment with a recalcitrant human reality, towards a more mediated and differentiated relationship between self and world. *Hyperion* contains numerous allusions to contemporary debates concerning the possibilities of revolutionary social transformation and the appropriate means for its realization. But the narrator also undergoes a process of personal transformation in the act of narration itself, culminating in an emancipatory vision of what it means to become an artist dedicated to the projection of a human future free of despotic coercion. The lyrical, poetic prose celebrates a religio-aesthetic conception of the world as an intrinsically harmonious totality that requires no external theological legitimization, and recognizes the experience of mortality and finitude as a necessary moment of the life of the whole, concluding upon a prospective rather than a resigned note of chastened aspiration.

Hölderlin's next major project, a tragedy on the death of Empedocles, continued and deepened many of these central themes. Hölderlin never brought the play to a satisfactory conclusion, but three fragmentary versions survive, together with an exceedingly compressed and obscure related prose study known as the *Grund zum Empedokles (The Groundplan of Empedocles)*, all composed between 1797 and 1799. Here, too, he concerns himself with the possibility of revolution and the necessity of an accompanying moral-spiritual reformation. But a newer note is also struck through emphasis upon the theme of tragic hubris on the part of the protagonist, and the

poet's increasing concern with the dangers of human identification with the divine - the seductions of pantheism - and the concomitant tendency to exalt human power in a unilaterally dominating relationship with the natural world and other human beings. This arguably represents an incipient critique of the 'Promethean' tendencies of Fichtean idealism and its subordination of nature as object to the subject as free power of total self-determination. None the less, Hölderlin equally emphasizes the importance of 'maturity' and autonomy in the Kantian sense if the potentially authoritarian excesses of revolutionary despotism are to be avoided. If Hölderlin originally conceived his drama as a contribution to contemporary political enlightenment and an encouragement to social renewal, it is unsurprising that he finally abandoned work on it when Napoleon became First Consul in 1799 and the hopes of a peaceful 'German' Revolution receded dramatically.

In the poetry written between 1800 and his apparent mental collapse in 1806, Hölderlin turned increasingly to more explicit engagement with the decisive actualities of European history, particularly the ambivalent relationship between the classical heritage and the claims of the Christian tradition. The implicit and semi-secularized eschatological dimension of Hölderlin's thought now comes insistently to the fore, and he concentrates upon the aporias of the poet's proclamation in a barren age characterized by a thankless and essentially instrumental attitude to the world. The concern with preserving a sense of proper limitation is reflected in the constant reference to 'mortals' and the powerful apocalyptic tone of the final hymns. After his breakdown (arguably exacerbated if not actually precipitated by drastic contemporary 'treatment') Hölderlin was taken under benevolent care until his death. In this last phase he reverted to a more detached and naïve style of composition in which the tensions and problems of his earlier life and work seem fictively resolved and dissolved into an eerie catatonic calm.

2 Fundamental themes

If poetry is 'made of words not ideas', there is an obvious danger in abstracting selected themes from their complex aesthetic context. But Hölderlin 's work articulates rather than illustrates a developing and penetrating body of coherent concerns and a distinctive mode of spiritual perception. In the first instance, it aesthetically prefigures a comprehensive reconciliation of the great polarities and disabling dualisms bequeathed by Kantian thought, principally the bifurcation between subject and object, spirit and nature. Yet the anti-identitarian, literally 'anti-monarchical', moment of Hölderlin's thought also resists the institution of a single controlling principle, that would suppress diversity and differentiation, in favour of an ontology of reciprocal interaction (1946-85 vol. 6: 300-1). The rehabilitation of finitude and fragility, of an acknowledged interdependence of subject and object, art and nature, is typically connected with an ethics of gratitude that is more than an ideology of subservience or deference. Many poems express the thought of an intrinsic dependency of the 'divine' dimension upon the sphere of mortal experience if it is to find fulfilment through an appropriate human 'correspondence' which constitutes neither an identity nor an antithesis. Despite the repudiation of a traditional metaphysical concept of an infinite creator-God independent of the created world, there remains a permanent 'theological' residue, a sublimated Christian element, in Hölderlin's recognition of our inherent 'creatureliness' within the economy of nature, our privileged vocation as the finite site for the self-manifestation and self-enjoyment of life through the indispensable, cultivating addition of human 'art' (1946-85 vol. 6: 326-30). Perhaps Hölderlin's most lasting significance lies in this powerful utopian vision of reconciled dwelling, rooted in a conception of art as liberating praxis that has only become central in our own time. This work is less a unique subjective and personal expression - though it is also that - than a compelling articulation of universal human aspirations for a transformed relationship to inner and outer nature. Aesthetic formalist tendencies can arguably do little justice to work which 'aesthetically' questions the supposed 'autonomy of the aesthetic' as a subjective and sequestered sphere of private experience, and implicitly challenges some of our most cherished modern philosophical assumptions.

See also: German idealism; Romanticism, German

NICHOLAS WALKER

List of works

Hölderlin, J.C.F. (1946-85) *Sämtliche Werke (Große Stuttgarter Ausgabe)*, ed. F. Beissner and A. Beck, Stuttgart: Kohlhammer, 8 vols.(The standard critical edition of Hölderlin's writings with extensive commentary and accompanying documentation - an indispensable reference source.)

Hölderlin, J.C.F. (1975-) *Sämtliche Werke (Frankfurter Ausgabe)*, ed. D.E. Sattler, Frankfurt: Verlag Roter Stern.

(A recent edition of Hölderlin's writings which attempts to present the evolving stages of his poems with less editorial intervention in favour of the supposedly 'final' closed text.)

Hölderlin, J.C.F. (1798) *Hyperion*, trans. W.R. Trask in E.L. Santer (ed.) *Friedrich Hölderlin. Hyperion and Selected Poems*, New York: Continuum, 1994.(English translation of Hölderlin's novel, adapted by D. Schwarz, and bilingual selection of mainly later poems translated by various hands.)

Hölderlin, J.C.F. (1799) *Empedocles*, trans. M. Hamburger in *Hölderlin. Poems and Fragments*, London: Anvil Poets Press, 3rd edn, 1994.(Excellent bilingual edition with representative selection of poems, including the second and third versions of *Empedocles* and the commentated translations of Pindar.)

Hölderlin, J.C.F. (1988) *Friedrich Hölderlin. Essays and Letters on Theory*, ed. and trans. T. Pfau, Albany, NY: State University of New York Press.(Contains the philosophical fragments and essays, the remarks on Greek tragedy and relevant excerpts from the correspondence.)

References and further reading

Hölderlin-Jahrbuch (1944-) Tübingen: Verlag Metzler.(Leading journal which continuously reviews the international secondary literature and publishes articles on all aspects of Hölderlin.)

Adorno, T.W. (1965) 'Parataxis', in *Noten zur Literatur III*, Frankfurt: Suhrkamp, 156-209; trans. S.W. Nicholsen in *Notes to Literature*, New York: Columbia University Press, vol. 2, 109-49, 1992.(Critical of attempts, like those of Heidegger, to minimize Hölderlin's debt to the context of German Idealism.)

Dastur, F. (1992) *Hölderlin, Tragédie et Modernité (Hölderlin: tragedy and modernity)*, Fougères: Encre Marine.(A concentrated and elegant introduction to Hölderlin's thought, focusing on his understanding of the relation between the ancients and moderns.)

Dilthey, W. (1905) 'Hölderlin', in *Das Erlebnis und die Dichtung*, Leipzig: Teubner Verlag, 349-459; trans. J. Ross, in R.A. Makkreel and F. Rodi (eds) *Poetry and Experience*, Princeton, NJ: Princeton University Press, 303-83, 1985.(The first important study of Hölderlin's importance in the evolution of idealist philosophy, with an emphasis upon the earlier works.)

Heidegger, M. (1981) *Erläuterungen zu Hölderlins Dichtung*, vol. 4 of *Gesamtausgabe*, Frankfurt: Klostermann; trans. W. McNeill and J. Davis as *Hölderlin's Hymn to the 'Ister'*, Bloomington and Indianapolis, IN: Indiana University Press, 1996.(Penetrating and influential readings of the poetry in terms of the author's central philosophical preoccupations.)

Henrich, D. (1992) *Der Grund im Bewusstsein, Untersuchungen zu Hölderlins Denken (1794-1795) (Foundations in consciousness. Studies in Hölderlin's thought)*, Stuttgart: Klett Cotta.(Detailed philosophical analysis of Hölderlin's role in developing a new theory of self-consciousness in relation to Fichte and Schelling.)

Henrich, D. (1997) *The Course of Remembrance and other Essays on Hölderlin*, Stanford: Stanford University Press.(Contains a detailed study of one of Hölderlin's late poems, and other essays on Hölderlin in the context of German idealist thought.)

Kurz, G. (1975) *Mittelbarkeit und Vereinigung. Zum Verhältnis von Poesie, Reflexion und Revolution bei Hölderlin (Mediation and unification: the relationship of poetry, reflection and revolution in Hölderlin)*, Stuttgart: Metzler.(General study of Hölderlin in the historical context of early idealism from the perspective of the Frankfurt School of critical theory.)

Ryan, L. (1960) *Hölderlin Lehre vom Wechsel der Töne (Hölderlin's theory concerning the modulation of tones)*, Stuttgart: Metzler.(Detailed study of Hölderlin's aesthetic reflections on poetic genre and register.)

Ryan, L. (1965) *Hölderlins 'Hyperion': Exzentrische Bahn und Dichterberuf (Hölderlin's 'Hyperion': eccentric path and poet's vocation)*, Stuttgart: Metzler.(Important systematic reading of *Hyperion* which explores Hölderlin's response to idealist thought and stresses the philosophical structure of the novel.)

Taminiaux, J. (1977) *La nostalgie de la Grèce à l'aube de l'idéalisme allemand, Kant et les Grecs dans l'itinéraire de Schiller, de Hölderlin et de Hegel (The yearning for Greece and the origin of German idealism: Kant and the Greeks in the intellectual development of Schiller, Hölderlin and Hegel)*, The Hague: Martinus Nijhoff.(Differentiated study of the relationship between early idealist reception of Kant's third Critique and the cult of classical antiquity in Hölderlin, Schiller and Hegel.)

Unger, R. (1975) *Hölderlin's Major Poetry. The Dialectics of Unity*, Bloomington, IN: Indiana University Press.(General study stressing the tension between pantheistic intention and literary realization; good bibliographical references to the entire field.)

Warminski, A. (1987) *Readings in Interpretation: Hölderlin, Hegel, Heidegger*, Minneapolis, MN: University of Minnesota Press.(Comparative analysis of Hölderlin in the light of contemporary deconstructive criticism.)

Holism and individualism in history and social science

Methodological individualists such as Mill, Weber, Schumpeter, Popper, Hayek and Elster argue that all social facts must be explained wholly and exhaustively in terms of the actions, beliefs and desires of individuals. On the other hand, methodological holists, such as Durkheim and Marx, tend in their explanations to bypass individual action. Within this debate, better arguments exist for the view that explanations of social phenomena without the beliefs and desires of agents are deficient. If this is so, individualists appear to have a distinct edge over their adversaries. Indeed, a consensus exists among philosophers and social scientists that holism is implausible or false and individualism, when carefully formulated, is trivially true.

Holists challenge this consensus by first arguing that caricatured formulations of holism that ignore human action must be set aside. They then ask us to re-examine the nature of human action. Action is distinguished from mere behaviour by its intentional character. This much is uncontested between individualists and holists. But against the individualist contention that intentions exist as only psychological states in the heads of individuals, the holist argues that they also lie directly embedded in irreducible social practices, and that the identification of any intention is impossible without examining the social context within which agents think and act. Holists find nothing wrong with the need to unravel the motivations of individuals, but they contend that these motivations cannot be individuated without appeal to the wider beliefs and practices of the community. For instance, the acquiescence of oppressed workers may take the form not of total submission but subtle negotiation that yields them sub-optimal benefits. Insensitivity to social context may blind us to this. Besides, it is not a matter of individual beliefs and preferences that this strategy is adopted. That decisions are taken by subtle strategies of negotiation rather than by explicit bargaining, deployment of force or use of high moral principles is a matter of social practice irreducible to the conscious action of individuals.

Two conclusions follow if the holist claim is true. First, that a reference to a social entity is inescapable even when social facts are explained in terms of individual actions, because of the necessary presence of a social ingredient in all individual intentions and actions. Second, a reference to individual actions is not even necessary when social facts are explained or understood in terms of social practices. Thus, the individualist view that explanation in social science must rely wholly and exhaustively on individual entities is hotly contested and is not as uncontroversial or trivial as it appears.

1 Introduction

Several commentators have noted the extraordinarily muddled character of the debate that ensued in the 1950s between individualists and holists. Why this has been so is one of the enigmas of social science. The main issue was repeatedly derailed, peripheral issues were given singular pre-eminence, different versions of the doctrine remained entangled, and deeper philosophical issues in the philosophy of science were never adequately probed. What is worse, both sides persistently knocked down the weakest, most caricatured formulation of the opponent’s views. Retrospectively, the battle does give the appearance of possessing a large element of sham. No wonder then that when the dust settled, many of the issues that once provoked so much controversy were resolved; motivated prejudice against the other was acknowledged, and soon, more concessive versions, plainly indistinguishable from each other, were formulated. A careful analysis of the debate can identify points of agreement between the two and, more importantly, that something significant and contentious between the two sides persists, therefore the battle is likely to rage in the future. This shall be demonstrated by first unearthing the principal assumptions of individualism and then showing that they can be challenged from a non-individualist viewpoint.

Any further exploration of the subject must begin with a warning: it is unwise to be tempted by the view that a worthy political cause or great moral ideal is at stake in contending methodologies. The passion that Popper (1960) and Hayek brought into the debate immobilized it rather than provided the fresh impetus it badly needed. On this point Schumpeter was right: an advocacy of methodological individualism entails no commitment to liberalism. Likewise, opposing it does not imply abandoning political liberalism or denying the moral value of the individual. So it is not just that methodological individualism should not be confused with political, moral or religious individualism but, more importantly, it must not be seen automatically to enlist itself behind any of them.

It is equally important to remind ourselves that the term ‘methodological’ must be understood here in the broadest

sense possible to include at least three distinct but related ingredients. The first, epistemological, concerns the understanding and explanation of social and individual entities. The second, ontological, turns on whether or not social entities are anything beyond a collection of individual entities. The third, semantics, is about the meaning of words: can a term referring to a social entity be analysed without remainder into terms that refer only to individual entities? The literature on the subject abounds with attempts to turn methodological individualism and holism into purely explanatory doctrines, but disputes over the relevant type of explanatory variable or over correct explanatory form have frequently turned into sharp controversies on ontological and semantic matters. This is hardly surprising, because an explanatory doctrine presupposes or entails ontological and semantic commitments. Therefore, individualism and holism are best viewed as comprehensive strategies with all the three components mentioned above.

2 Ontological reasons

This much clarified, it must be noted that a particular dispute, over one form of semantic individualism, is entirely settled. There are no takers any more for a project inspired by logical constructionism recommending that every single term referring to social entities be analysed as a collection of individual terms. However, for individualists, this failure has no ontological or explanatory ramifications. Second, one ontological dispute must be set aside. Encouraged by careless formulations and deep misunderstanding, it was once thought that holists believed in supra-substances hovering over and above all individuals; that they were carrying forward the Cartesian legacy of substance-dualism by introducing yet another substance called the social. This is unfair to the holist who must be seen instead to be advocating property pluralism and emphasizing the distinctive causal efficacy of social properties.

We may compare this point with an issue in the natural sciences. Non-reductionists in biology often argue that biological organisms display teleological properties that are best understood as irreducible to but supervenient upon physical properties (see Supervenience). The principle of supervenience views the world as stratified, with properties at one level dependent for their existence in a non-causal way on properties lying at the more basic level. However, this relation of ontological dependence is non-reductive because the two sets of properties are never identical. Thus, teleological properties presuppose the existence of physical properties without being reducible to them.

It is in this sense that social properties are said to be supervenient on physical and biological ones. This much ought to be beyond dispute. What is not clear, however, is that this entails a supervenience of the social on the individual. For much depends on what we mean by the individual. If the term ‘individual’ refers to a single identifiable biological organism, then the statement is true. But surely, it is not the pure and abstract biological individual in focus here; rather what concerns us is the individual with distinctive human properties. The key question then is whether these distinctive human properties emerge only when individuals come together or if they already exist, independent of any relations between them. Similarly, it is fairly incontrovertible to claim that any social process is supervenient on the physio-biological movement of individual organisms. Durkheim, as good a non-individualist as any, said so much: ‘It is very true that society comprises no active force other than individuals’ (1982: 251). The holist argument contends only that the specific pattern of this dynamic process is not reducible to the movement of individual biological organisms taken aggregately (see further §5 below).

Another vital issue to register is this: the individualist/holist controversy is not identical to the structure/agency debate in the social sciences, for a number of reasons. First, individualism does not entail a commitment to free human agency. The individualist may believe agency to be restricted too, but locate all constraints on it within the individual psyche. Individualism is compatible with internal constraints and therefore with a full-blooded determinism. Second, it is simply false to assume that anything external to the individual agent is thereby constraining. External structures are not just constraining but also enabling. Indeed a robust view of free and active human agents is compatible with the claim that such capacities depend on the presence of relevant external structures. More importantly, individualists can develop not only a conception of internal but also of external constraints. They too can account for the very constitution of external structures, of rules, norms and conventions constraining human action. Likewise holists can develop a perfectly cogent account of free human action. If this is so, the principal dividing line between the two sides cannot be drawn at the level of free agency and external constraint.

This is not the dominant understanding of those engaged in this debate. Holists accused individualists of ignoring the relations of the individual to the larger social context. Individualists, on their part, blamed holists for treating individuals as mere cogs in the social wheel. One was charged with ultra-voluntarism, the other reproached for supra-determinism. At the end of a long and unsavoury controversy, however, each side realized what should have been obvious from the very start: that neither agency nor the context within which it occurs can be ignored.

Let us identify the points where agreement is undeniable. Holists do not believe in supra-individual, social substances. They emphasize social relations and their undeniable constraining effects. Individualists cannot fault these claims. Individualists, likewise, highlight human agency and show that societies consist in or are a result of people acting. Holists can hardly object to this. Furthermore, both accept that the social does not think; it is individuals who do the thinking. We appear then genuinely to arrive at a consensus on a set of statements so truistic that an explanatory strategy must be deemed mistaken if it fails to incorporate or is incompatible with it.

However, a plausible holist contention is that this agreement is at the price of an even greater intellectual muddle. The seemingly uncontroversial character of these statements obscures the fact that they are open to rival interpretations, one that antecedently shifts the entire balance in favour of the individualist and another which obdurately refuses to do so. The claim that individualism is trivially true may well be a rhetorical device that conceals deeper ontological and methodological issues that divide rival camps. All the dimensions of the debate are not unravelled unless another vital issue, forced upon us less by ontological than by epistemological reasons, is addressed. Sharp differences between holists and individualists are often believed to lie precisely at this point. Close analysis reveals, however, a thicker consensus on this issue. Contentious issues lurk in the background, despite this agreement.

3 Epistemological reasons

The correct move here would first be to identify the relevant explanatory form and then try and show it as ontologically neutral, usable by individualists and non-individualists alike. But contestants in the debate have not approached the issue in this way nor found it easy to identify the correct explanatory form. One reason for this is their insensitivity to another major controversy in social science, namely, the argument between naturalists and anti-naturalists. Each side has often assumed not only the absolute validity of their preferred model of scientific explanation but their monopoly over it. The second reason is the wholly opposite error of identifying the individualist/holist debate with the naturalist/anti-naturalist controversy, as if the settlement of the first dispute rests solely on the successful resolution of the second. By assuming that holism entails scientism and that therefore their victory flows automatically from the defeat of positivist naturalism, individualists such as Hayek (1973) have been prone to this mistake. Clearly, many individualists are no less dogmatic about the claim that human affairs must be studied strictly with the methods of natural science. Conversely, holists can be equally steadfast in their advocacy of a distinctive method for social science. So it is important to see both the relevance of the naturalist/non-naturalist dispute to the individualist/holist issue and not to identify them (see Naturalism in social science).

The other related difficulty is that the debate is conducted without disentangling different explanatory forms. We may begin by distinguishing five versions of individualist and holist explanatory strategies. Three of these deploy the deductive nomological model of explanation (see Explanation §2). The first attempts to explain particular social facts by deducing them from laws of individual action. The second, known as explanatory reduction, explains not particular social facts but social laws by deducing them from laws pertaining to individuals (see Reduction, problems of). Since this reduction is achieved with the help of bridge statements which in turn can take two forms, we obtain two forms of reductionist strategies. The first requiring correlatory laws may be called correlatory reduction; the second, requiring identity statements is called identity-reduction. Identity-reduction is also referred to as micro-reduction because here a macro-entity (the social) is identified with all the micro-entities (the individuals) that compose it.

The difference between each must be maintained because arguments that work against one kind of individualist strategy do not necessarily work against the other. For example, if laws are deemed essential for explanations, and if it is shown that only laws pertaining to individual action can be formulated, then the first non-reductionist deductive-nomological version of individualism is vindicated. Holism fails because no social laws exist to explain particular social facts. Mandelbaum believed precisely this to be the target of individualist attack and therefore

tried to establish the existence of social laws (1973). But an argument establishing social laws does not invalidate individualism because by deducing them from laws pertaining to individuals, one version of individualism is still exonerated. Similarly, showing the impossibility of correlatory laws does not defeat reductionist individualism because the absence of correlatory laws may indicate not the impossibility of reduction but the prior realization of a successful identity-reduction.

A large part of the debate has been traditionally obsessed with these issues. However, this intellectual effort may have been futile if laws pertaining to individuals do not exist. For all these strategies depend not just on the presence of social laws but on general laws concerning individuals. And it is doubtful if such laws of human action exist. Notice that the laws in question must capture the precise features individuating human action. The same entity when subsumed under other descriptions may indeed yield laws but these would be utterly irrelevant. Reductionists are not pursuing laws of physical movement or bodily behaviour. The laws they are after must refer to actions, and therefore, to the beliefs and desires of individuals. However, as Davidson has shown, precisely such strict laws that explain or predict mental phenomena are impossible to obtain (1980). If valid, this is a decisive argument against all deductive-nomological versions of individualism. To revive this aspect of the debate, one has either to defeat this argument or give a conception of laws different from the one normally accepted (see Explanation in history and social science §§1, 3).

Two conclusions may follow from this. First, that explanations in social science are not possible, and second, more importantly, that laws are not necessary for explanation. Therefore we may explore explanatory forms unconnected with the deductive-nomological model. Forms that do not rely on the deductive-nomological model are intentional explanation and intentional understanding. The first, without invoking general laws or theoretical syllogism, explains action by showing that the behaviour of individuals is guided by beliefs and desires (intentions). The second, intentional understanding stresses the pivotal significance of understanding the beliefs and desires of agents. These strategies can be combined to give one overarching model for social science in which the understanding of beliefs and desires go hand in hand with the explanation of action. In what follows, the term ‘intentional explanation’ refers to this model.

4 Individualist intentional explanation

The formal structure of intentional explanation is simple and, up to a point, relatively uncontroversial: *A* desires *X*; *A* believes that doing *Y* will get him *X*; *A* does *Y*. The action, *Y*, is explained by deducing it, via a practical syllogism, from the given beliefs and desires. A crucial feature of beliefs and desires, of all intentional states generally, is that they are directed towards a real or imaginary object and that therefore they have an intentional content. Since these states are complex and in large measure linguistically constituted, we may say that they possess a linguistic content or meaning. Their ascription always involves the use of a ‘that’-clause. For example, *A* does not just believe or desire; *A* believes that the world is menacing and ugly or desires that it be beautiful. Clearly, the individuation of beliefs and desires depends on the meaning of words in these clauses. This presupposes that we capture the oblique occurrence of these expressions, where the oblique occurrence of expression is one in which the substitution of coextensive expression affects the truth-value of the whole containing sentence. For example, if ‘water’ occurs obliquely in ‘*A* believes water to be the most refreshing drink in the world’, then replacing it with ‘H_2O’ will alter the truth-value of the whole sentence.

This feature of obliqueness of expressions is of great significance for intentional explanation. Actions are individuated by their appropriate description, the description under which agents subsume their behaviour. To grasp their perspective or point of view, we must get right the description of their beliefs and desires, and this depends in turn on getting hold of obliquely occurring expressions. Likewise, a situation on which agents hold a view plays no direct role in characterizing their mental states or actions; it contributes to the explanation of actions only as it occurs obliquely, mediated by their points of view. This much is uncontroversial. But contrary to appearance, this by itself does not lead to individualism. The individualist turn comes with the further ontological assumption that these intentional states exist internal to agents, only in their heads. The individualist interpretation appears even more invincible with the second assumption that the meaning of beliefs and desires, their content, is individuated wholly internally, without reference to the external environment of the agent. These Cartesian assumptions turn action into a wholly individualist entity, bolstering the individualist belief that if the correct explanatory form in the social science is intentional explanation, then it just has to imply a victory of the

individualist over the non-individualist.

5 The holist challenge

The widespread acceptance of these assumptions has generated the illusion of consensus. The holist challenge shatters it. It is levelled on two fronts, one at individualist theories of meaning, the other at individualist theories of action. Arguments developed by Putnam and Burge help reinvigorate a non-individualist theory of meaning. Here one begins by the standard disambiguation of meaning into (a) extension and (b) intension (see Sense and reference). Then one shows that traditional theories of meaning are characterized by two assumptions. First, that intension determines extension. This can mean many things but it means at the very least that it helps to fix extension. Also, that sameness of intension always entails sameness of extension. Let us call this the functional character of intension. The second is the ontological assumption, already mentioned, that intension is a psychological entity that exists in the heads of individuals. So, we possess in our heads mental entities or representations that fix extensions of terms.

Putnam's thought experiment about two different substances on earth and twin earth, identical in all respects but their chemical structure, involves a long and complicated argument the conclusion of which, on one possible interpretation, is this: that the functional and the ontological assumptions cannot both be true (1975). Since identical mental states refer to different extensions, we must abandon either the view that meaning is a psychological entity or the claim that intension determines extension. Arguably, Putnam offers sound reasons for dropping the first and retaining the second. In other words, if the conclusion forced by Putnam's thought experiment is true, we must infer that meaning is not a psychological entity in the head of individuals. What then is it? Several answers exist but perhaps the one that challenges the individualist most is the view that it is fixed by an interlocking system of beliefs and actions necessarily involving several individuals. If so, meaning is a social not an individual entity. A related argument developed by Burge shows first that a difference in extension also affects oblique occurrences in 'that'-clauses, and therefore, the contents of mental states, and second, that 'communal practice is a factor in fixing and interpreting the words and attitudes of a person' (1979: 85). In the individuation of beliefs we must refer not just to the internal environment of agents but also to their external context, particularly to relevant social practices. If meanings are inescapably social, then, given that beliefs are individuated by meanings, beliefs too are social, and further, given that beliefs are analytic to action, so are actions. We began by attempting to explain a social entity wholly and exclusively in terms of individual entities, namely actions, beliefs and desires, we now find that all these are inescapably social. A social fact is then explained by another social fact. Individualism is challenged quite fundamentally. Also, holists do not abandon intentional explanation; they give it a non-individualist interpretation. There is no necessary dispute between individualists and non-individualists over which explanatory form is suitable for human affairs.

So, this is the crux of the matter. For individualists, beliefs, desires, intentions and actions are intrinsically individual. This is implausible for the holist because the intention of one person already makes reference to the intention of another; beliefs, desires and actions are already social. But what of the individualist view wherein intentions exist as internal states in the heads of individuals? And what about the ontology of irreducible social practices? To answer this, the holist opens the second front of attack, this time on the traditional theory of action. Put simply, the individualist theory of action is the widely accepted causal view according to which action can be analysed into three components: behaviour, intentional states and the relation of causality that binds them (see Action §1). Action is behaviour caused by intentions. An alternative theory refuses to reduce action to anything more basic. Intentional states separate from behaviour do not exist; rather, agents intend something by their behaviour. On this view, intentions lie directly embedded in behaviour so that we see behaviour as action. Action then is a basic ability of human beings quite independent of the ability to form representations in their heads. The holist accepts this and asks us to go further. If intention can directly inhabit the bodily behaviour of one individual, it can also occupy the bodily behaviours of several individuals.

This ontological issue has consequences for explanatory strategy as well. First, the identification of beliefs and intentions of an individual, crucial for intentional explanation, requires sensitivity to the interlocking intentions of a community of agents. The perspective of the agent, the oblique expressions in which it is embedded, is still crucial but the holist holds no presumption that it is available only or completely to the agent. Second, a grasp of social practice is necessary to situate individual beliefs, desires and actions. This is less a function of mental

operations in individual heads and more of practical know-how. Without practical understanding, as anthropologists testify and as philosophers who have worked on problems of translation and cross-cultural understanding know only too well, intellectual effort, no matter how sincere, is puerile.

See also: Intention; Methodological individualism; Social laws

RAJEEV BHARGAVA

References and further reading

Bhargava, R. (1992) *Individualism in Social Science*, Oxford: Clarendon Press. (Expansion of 2-5 of this entry.)

Burge, T. (1979) 'Individualism and the Mental', in P. French, T. Uhelling and H. Wittestein (eds) *Midwest Studies in Philosophy* 4: 73-121.(Sophisticated defence of holist theory of meaning.)

Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Clarendon Press. (Particularly sophisticated piece against the possibility of psychological laws and on the causal theory of action.)

Durkheim, É. (1982) *The Rules of Sociological Method*, ed. S. Lukes, trans. W.D. Halls, Basingstoke, Macmillan. (Classic sociological defence of holist explanations.)

Elster, J. (1985) *Making Sense of Marx*, Cambridge: Cambridge University Press. (Short statement of methodological individualism known for its defence of an individualist Marxism.)

Garfinkel, A. (1981) *Forms of Explanation*, New Haven, CT: Yale University Press. (Pragmatist defence of holistic explanation.)

Hayek, F.A. (1973) 'From Scientism and the Study of Society', in J. O'Neill (ed.) *Modes of Individualism and Collectivism*, London: Heinemann, 27-67.(Passionate defence of individualism on grounds that holism is committed to scientism.)

James, S. (1984) *The Content of Social Explanation*, Cambridge: Cambridge University Press.(Interprets the individualist/holist controversy as an agency/structure debate.)

Lukes, S. (1973) 'Methodological Individualism Reconsidered', in A. Ryan (ed.) *The Philosophy of Social Explanation*, Oxford: Oxford University Press, 119-29.(One of the best commentaries on the debate of the 1950s-60s.)

Mandelbaum, M. (1973) 'Societal Laws', in J. O'Neill (ed.) *Modes of Individualism and Collectivism*, London: Heinemann, 235-47.(Defends holism on the grounds that irreducible social laws exist.)

Manninen, J. and Tuomela, R. (eds) (1976) *Essays on Explanation and Understanding*, Dordrecht: Reidel.(Fine pieces on the structure of intentional explanation.)

O'Neill, J. (ed.) (1973) *Modes of Individualism and Collectivism*, London: Heinemann.(Best anthology of the individualist/holist controversy in the 1950s-60s.)

Popper, K.R. (1960) *The Poverty of Historicism*, London: Routledge & Kegan Paul. (Passionate defence of methodological individualism. One of the principal originators of the debate.)

Putnam, H. (1975) *Philosophical Papers: Mind, Language and Reality, Vol. 2*, Cambridge: Cambridge University Press.(Particularly original attack on traditional theories of meaning, can be deployed for holist purposes.)

Ruben, D.-H. (1985) *The Metaphysics of the Social World*, London: Routledge & Kegan Paul.(Sophisticated analysis of group/individual relation and supervenience.)

Ryan, A. (1970) *The Philosophy of Social Explanation*, London: Macmillan.(Useful introduction to the main issues in the philosophy of social science, including the individualist/holist controversy.)

Taylor, C. (1985) *Philosophy and the Human Sciences*, Cambridge: Cambridge University Press.(Original and suggestive pieces on holist theories of meaning and action.)

Wright, G.H. von (1971) *Explanation and Understanding*, London: Routledge & Kegan Paul.(Classic study of the difference between nomological explanation and human understanding.)

Holism: mental and semantic

Mental (or semantic) holism is the doctrine that the identity of a belief content (or the meaning of a sentence that expresses it) is determined by its place in the web of beliefs or sentences comprising a whole theory or group of theories. It can be contrasted with two other views: atomism and molecularism. Molecularism characterizes meaning and content in terms of relatively small parts *of the web in a way that allows many different theories to share those parts. For example, the meaning of 'chase' might be said by a molecularist to be 'try to catch'. Atomism characterizes meaning and content in terms of* none *of the web; it says that sentences and beliefs have meaning or content independently of their relations to other sentences or beliefs.*

One major motivation for holism has come from reflections on the natures of confirmation and learning. As Quine observed, claims about the world are confirmed not individually but only in conjunction with theories of which they are a part. And, typically, one cannot come to understand scientific claims without understanding a significant chunk of the theory of which they are a part. For example, in learning the Newtonian concepts of 'force', 'mass', 'kinetic energy' and 'momentum', one does not learn any definitions of these terms in terms that are understood beforehand, for there are no such definitions. Rather, these theoretical terms are all learned together in conjunction with procedures for solving problems.

The major problem with holism is that it threatens to make generalization in psychology virtually impossible. If the content of any state depends on all others, it would be extremely unlikely that any two believers would ever share a state with the same content. Moreover, holism would appear to conflict with our ordinary conception of reasoning. What sentences one accepts influences what one infers. If I accept a sentence and then later reject it, I thereby change the inferential role of that sentence, so the meaning of what I accept would not be the same as the meaning of what I later reject. But then it would be difficult to understand on this view how one could rationally - or even irrationally! - change one's mind. And agreement and translation are also problematic for much the same reason. Holists have responded (1) by proposing that we should think not in terms of 'same/different' meaning but in terms of a gradient of similarity of meaning, (2) by proposing 'two-factor' theories, or (3) by simply accepting the consequence that there is no real difference between changing meanings and changing beliefs.

1 The doctrines

Semantic holism is the view that the meaning of a sentence is determined by its place in the web of sentences comprising a whole theory. Mental holism is the corresponding view for belief content - that the identity of a belief content is determined by its place in the web of beliefs comprising a theory. Sometimes holists advocate a more sweeping view in which the identity of a belief is determined by its relations to a body of theories, or even *the whole of a person's belief system*. In what follows, mental and semantic holism are treated as two aspects of a single view.

Holism can be contrasted with two other views: molecularism and atomism. Molecularism characterizes meaning and content in terms of a relatively small part of the web that many different theories may share. For example, the meaning of 'bachelor' might be said by a molecularist to be 'man who has never married'. And the meaning of 'and' might be given by a molecularist version of inferential role semantics via specifying that the inference from '*p* and *q*' to '*p*', and from '*p*', '*q*' to '*p* and *q*' has a special status (for example, it might be 'primitively compelling', in Peacocke's terms; see Semantics, conceptual role). Atomism characterizes meaning and content in terms of none of the web; it says that sentences and beliefs have meaning or content independently of their relations to any other sentences or beliefs and, therefore, independently of any theories in which they appear.

Note the contrast between the semantic issues that are of concern here and those that concern particular phenomena in particular languages. Semantics in the present sense is concerned with the fundamental nature of meaning and what it is about a person that makes their words mean what they do. We might call the present sense the 'metaphysical' sense. Semantics in the other sense - what we might call the 'linguist's' sense - concerns the issues of how meanings of words fit together to determine the semantic properties and internal structures of sentences. Semantics in the linguist's sense concerns such issues as how many types of pronouns there are and why it is that 'The temperature is rising' and 'The temperature is 60 degrees' does not entail that 60 is rising. There are interactions among the two enterprises, but semantics in the linguist's sense can proceed without taking

much notice of the issue of semantic holism.

2 Motivations for holism

The best-known motivation for semantic/mental holism involves Quine's doctrine of *confirmation* holism, according to which 'Our statements about the external world face the tribunal of sense experience not individually but only as a corporate body' (1953: 41). This view gains its plausibility from the logic of theory revision. An experimental datum confirms (verifies; gives us some reason to believe) a statement only in conjunction with a great number of theoretical ideas, background assumptions about the experiment, and assumptions from logic and mathematics, any one of which could be (and in the history of science often has been) challenged when problems arise (see Confirmation theory). If we combine this confirmation holism with the logical positivist doctrine that the meaning of a sentence is its method of verification or confirmation (see Logical positivism; Meaning and verification), that is, if we combine the doctrine that meaning is confirmation with the claim that confirmation is holistic, we get semantic holism. And this implies that talk of the meaning of a sentence in isolation from other sentences makes no more sense than talk of the meaning of 'of' apart from the contexts in which it occurs.

Positivism and confirmation holism are not the only roads to semantic/mental holism. Another route proceeds from considering how people learn actual scientific theories. For example, one does not learn definitions of 'force', 'mass', 'kinetic energy' or 'momentum' in terms that are understood beforehand, for there are no such definitions. Rather, these terms are learned together (in conjunction with procedures for solving problems). As Quine (1954) and Putnam (1965) argued, local 'definitions' in a scientific theory tend to be mere passing expository devices of no lasting importance for the theory itself. And this is quite ubiquitous in theories, where there is a circle of interdefined theoretical terms none of which is definable in terms outside the theory. This fact motivates Lewis' proposal that scientific terms can be defined functionally in terms of their roles in a whole theory.

3 Must functionalism lead to holism?

Functionalism has become a popular approach in the philosophy of mind generally (see Functionalism). For example, the difference between the belief that one will win the lottery and the desire that one will win the lottery is plausibly a functional difference (a difference in the roles of the states), since one but not the other leads to test-driving a Ferrari. But functionalists go further, claiming that the common *content* of these propositional attitudes can also be functionally defined (in terms of the cognitive roles of states which have these contents in the psychological economy, including links to inputs and outputs). It has often been supposed that the most important feature of the functional role of a belief in determining its content is its role in inference, and for that reason functionalism about content or meaning is sometimes called 'inferential role semantics'. The functional role of a thought includes all sorts of causes and effects that are non-semantic, for example, perhaps depressing thoughts can lower one's immunity, causing one to become ill. Conceptual roles are functional roles minus such non-semantic causes and effects.

A functional theory of the whole mind must make reference to any difference in stimuli or responses that can be mentally significant. The difference between saying 'damn' and 'darn' can be mentally significant (for example, one can have a policy of saying one rather than the other). Your pains lead to 'darn', mine to 'damn', so our pains are functionally different, and likewise our desires to avoid pain, our beliefs that interact with those desires, and so on. So if we functionally define 'pain' in terms of a theory of the whole mind, we are naturally led to the conclusion that two individuals who differ in this way share no mental states. This is why functionalism can lead to holism.

Molecularists object that if you have a fine-grained way of categorizing, you can just coarsen it. But which causes and effects of pain are constitutive and which are not? The form of a solution could be: 'pain = the state constituted by the following causal relations…', where the dots are replaced by a specification of a subset of the mentally significant causal relations into which pain enters. Putnam suggested we look for a normal form for a computational description of pain, and Lycan (1988) and Rey (1997) have suggested that we construct functional theories at different levels, one of which would be suitable to define 'pain' without distinguishing between 'damn' and 'darn'. But after years of discussion, there is no real solution, not even a proposal of something functional common to all and only pains. Lycan and Rey expect the issue to be settled only by an empirical psychology. Moreover, even if one is optimistic about finding a functional definition of pain, one cannot assume that success

will transfer to functionalist accounts of meaning. Success in the case of meaning would seem to require an analytic/synthetic distinction which many have found independently to be problematic.

4 Problems with the analytic/synthetic distinction

Another route to holism arises from considerations involving the analytic/synthetic distinction, that is, the distinction between claims that are true solely in virtue of meaning and claims that depend also on the way the world is. Quineans often hold that the analytic/synthetic distinction is confused. Some philosophers have argued from the idea that there is something wrong with analyticity to holism. We can put the argument in terms of conceptual role semantics. It seems that some inferences (for example, from ‘bachelor’ to ‘unmarried’) are part of meaning-constitutive inferential roles, but others (for example, from ‘bachelor’ to ‘dislikes commitment’) are not. However, if there is no analytic/synthetic distinction, then there is no principled way to draw a line between inferences that constitute meaning and those that do not (see Analyticity). So, the argument concludes, if some inferences are part of meaning-constitutive inferential rules, *all* inferences are part of meaning-constitutive inferential roles, and this is a form of holism (Fodor and LePore 1992; Devitt 1995).

This argument is of course fallacious. A bald man can have some hairs, and there is no principled way of drawing a line between the number or distribution of hairs on a bald man and on a non-bald man. But one would not conclude that everyone is bald. Failure to find a principled way of drawing a line need not require one or the other extreme.

Still, the argument is onto something. How would the molecularist choose those inferences which are meaning-constitutive if the meaning-constitutive must be analytic rather than synthetic but there is no such distinction? In fact, the problem is more general, and far from being an argument *for* holism, it casts doubt on holism too. If meaning-constitutivity entails analyticity, any view - molecularist or holist - that postulates anything meaning-constitutive is in trouble if there is no such thing as analyticity.

One response to this argument has been to ‘challenge’ the principle that a statement or inference that is meaning-constitutive is thereby analytic (Block 1993). There are two very different points of view which see a gap between meaning-constitutivity and analyticity.

One approach to finding a gap between meaning-constitutivity and analyticity derives from the views of Quine and Davidson, on which there is no clear difference between a change of meaning and a change of belief (see Radical translation and radical interpretation). The other appeals to narrow contents (see Content: wide and narrow). Narrow contents are contents that are necessarily shared by ‘twins’, people who are internally as similar as you like, even though their environments differ. Consider the influential example of Putnam’s ‘twin earth’ which is a planet identical to earth in every respect except that wherever earth has H_2O, it has a superficially similar but chemically different substance, XYZ. Arguably, I and my twin on Putnam’s twin earth share a narrow content for ‘water’ despite the different referents of our words. If meaning is narrow in this sense, it is false that meaning-constitutive sentences or inferences are thereby analytic if meaning is narrow. The narrow contents which constitute meaning *themselves* are neither *true* nor *false* and hence cannot be true in virtue of meaning. For example, let us suppose that my twin and I accept the propositions that we express with ‘Water contains hydrogen’. My belief has a true wide content, my twin’s has a false wide content, but the narrow content has to be the same (since we are twins) and is therefore neither true nor false (see Putnam, H. §3). Further, we can even imagine a twin earth in which a putative meaning-constitutive inference is invalid. If there is any inference that is a good candidate for analytically defining ‘water’, it is the inference from ‘water’ to ‘liquid’. But consider a twin earth on which ‘water’ is used as here to refer to H_2O, but where water is very rare, most of the substances referred to as ‘liquids’ being granular solids that look like liquids. So ‘Water is a liquid’ as said by on this twin earth is *false*, even though it is true in our mouths. Perhaps it will be said that what is analytic is not ‘Water is a liquid’ but ‘Water has a liquidish look and feel’. But it is easy to imagine circumstances in which the look and feel of water changes. Perhaps what we should be looking for is not a narrow content that is *true* in virtue of meaning but one that is only *assertible* in virtue of meaning. But it is part of our commitment in the use of natural kind terms that the world plays a part in determining truth-values, so we must regard any appearance of warrant solely in virtue of meaning as superficial.

5 The problem of disagreement and translation

Holism has some weird-sounding consequences. Suppose we say that all of a sentence's inferential links (within a theory or body of theories) are included in its set of meaning-constitutive inferential roles. But what sentences I accept influences what I infer, so how can I reason so as to change my own mind? If I accept a sentence, say, 'Bernini stole the lead from the Pantheon', and then later reject it, I thereby change the inferential role of that sentence, so the meaning of the sentence that I accept is not the same as the one that I later reject. So how can I reason about which of my beliefs should be given up? Along similar lines, one can argue that no two people ever agree or disagree, and that we can never translate anything perfectly from one language to another. The holist owes us a way to reconcile such conclusions with common sense. This section will explore three holistic responses.

Harman (1973) and Block (1986) have argued that we can avoid the problem by replacing the dichotomy between agreement and disagreement with a gradient of similarity of meaning, perhaps multi-dimensional. If I first accept and then reject 'Bernini stole the lead from the Pantheon', it is not as if I have rejected something *utterly unrelated* to what I earlier accepted. This position profits from the analogy with the ordinary dichotomy between believing and disbelieving. Reasoning with this dichotomy can lead to trouble, trouble that is avoided if we substitute a graded notion for the dichotomy. For example, I can have a low degree of belief in a long conjunction even though I have a high degree in each of the conjuncts. But if we put this in terms of the dichotomy between believing and disbelieving, we say that I could believe each conjunct while disbelieving the conjunction, and that is a contradiction. The proposal, then, is that we substitute a graded notion of similarity of meaning for the ordinary notion of same/different meaning. It must be conceded, however, that there are no specific suggestions as to what the dimensions of similarity of meaning are or how they relate to one another.

This approach can be combined with the aforementioned 'two-factor theory', according to which meaning consists of an internal holistic factor and a non-holistic purely referential factor (see Semantics, conceptual role §2). For purposes of translation and communication, the purely referential factor plays the main role in individuating contents. For purposes of psychological explanation, the internal factor plays the main role (see Loar 1987).

There is another (compatible) holistic response to the problem of disagreement which is associated with the views of Quine, Davidson and Putnam, namely that there is something wrong with the terms in which the problem is posed. They explicitly reject the very distinction between disagreeing and changing the subject that is presupposed by the statement of the problem. Putnam (1988) and Stich (1983) have argued, along these lines, that translation is not an objective process; it depends on subjective value-laden decisions as to how to weigh considerations of similarity in reference and social and functional role. It is controversial whether this Quinean response avoids the problem of disagreement only by rendering meaning something unsuitable for science.

Another holistic response is exemplified by Lewis' observation (1995) that there is no need to suppose that a satisfier of a functional description must fit it perfectly - fitting *most* of it is good enough. Lewis proposes that in framing the functional roles, we replace the set of inferences that are the basis for a functionalized account of belief with the disjunction of all the conjunctions of most of them. Thus, if we think there are three inferences, *A*, *B* and *C*, that are closely linked to the meaning of 'if', we might define 'if' as the relation that satisfies either $A \,\&\, B$ or $A \,\&\, C$ or $B \,\&\, C$. (Of course, we thereby increase the danger that more than one relation will satisfy our definition.) Then disagreement will be possible between people who accept most of the inferences that define their subject matters.

I have just been canvassing holistic responses to the problem, but of course atomism and molecularism are also responses. Fodor's version of atomism (1987) construes meanings as purely referential. Fodor goes so far as to insist that there could be punctate minds; minds that have only one belief. This view must, however, find some way of accommodating the insights that motivate holism.

6 Psychological laws

Fodor and LePore (1992) object to holistic accounts of mental content on the ground that they would preclude psychological laws, for example, the belief that one is in immediate danger causes release of adrenaline. According to holism, there is no such thing as 'the' belief that one is in immediate danger because the belief that you designate in this way is not quite the same as the belief that I designate in this way. Beliefs are too fine-grained to be referred to in this way (see Belief; Propositional attitudes). One strategy for dealing with this issue is to observe

that many candidate psychological laws can generalize about contents without actually specifying them. Consider this candidate for a law: for *any* action *a* and any goal *g*, if one wants *g* and also believes that *a* is required for *g*, then one will try to do *a*. This is a universally quantified law (because of the role of 'any'), albeit a trivial one. Universally quantified laws are a good scientific bet, and these can involve holistic content. By quantifying over goals, one can state laws without committing oneself to two agents ever having exactly the same goal. The point just made says that the holist can allow one kind of psychological law (the quantified kind) but not another (the kind that mentions specific contents such as the belief that one is in danger). But the holist may go further, arguing that there is something wrong with the putative laws of specific contents. The point is that 'The belief that one is in immediate danger causes release of adrenaline' stands to psychological law as 'Large slippery rocks on mountain-tops can damage cars on roads below' stands to physical law. Laws should quantify over such specific items, not mention them explicitly (see Laws, natural; Explanation).

However, Fodor and LePore are right that any particular type of holistic state will exist only rarely and transiently. In this respect, holistic mental states are like the states of computers (see Mind, computational theories of). A total computer configuration as specified by the contents of every register in the internal memory and every cell on the hard disk will occur only rarely and transiently. There are deterministic laws of the evolution of total computer states, but they deal with such transient states. So psychological explanation will have to be seen by holists as like explanation of what computers do, in part a matter of fine-grained laws of the evolution of systems, in part coarse-grained accounts of how the systems work that do not have the status of laws.

7 Narrow-content holism

There is a great deal of controversy about whether there is such a thing as narrow content or meaning, but if narrow content exists, there is good reason to think it is holistic. We have already seen one reason having to do with the fact that there is no analytic/synthetic distinction for narrow content. But there is another reason as well that focuses on change of narrow content with learning. Putnam (1983) and Block (1994) give an argument that uses some relatively uncontroversial premises about identity and difference in narrow content at a single time to squeeze out a conclusion to the effect that one's narrow contents can be expected to change whenever one receives substantial new information, however trivial. The argument depends on a variant of the famous 'twin earth' example. Consider twins who grow up in different communities where 'grug' is used to denote different substances, beer in one and whisky in the other, but the difference has not made any difference to the twins. At the age of 10, they are as similar as you like, and so the narrow contents of their 'grugs' are the same. By the age of 12, they know as much about 'grug' as teenagers normally know, including the (different) translations of 'grug' into English. One knows that 'grug' in his language is beer, the other that 'grug' is whisky. The argument motivates the claim that their 'grugs' differ in narrow content at 12 despite being the same at 10, so the information that they have acquired (which is designed to be run-of-the-mill) has changed the narrow contents. (But see Devitt 1995 for a reply.)

Issues about holism continue to be at the heart of debate in philosophy of language and mind. In the mid-1960s, it was widely assumed that to be a holist was to be a sceptic about any science of meaning or content, but thirty years later there has been a spirited debate about whether cognitive science can tolerate it.

NED BLOCK

References and further reading

Block, N. (1986) 'Advertisement for a Semantics for Psychology', in P.A. French *et al.* (eds) *Midwest Studies in Philosophy*, vol. 10, *Studies in the Philosophy of Mind*, Minneapolis, MN: University of Minnesota Press. (Discussion of inferential role semantics.)

Block, N. (1993) 'Holism, Hyper-Analyticity and Hyper-Compositionality', *Mind and Language* 8 (1): 1-27. (Defence of holism from a critique by Fodor and LePore.)

Block, N. (1994) 'An Argument for Holism', *Proceedings of the Aristotelian Society*, new series, 94: 151-69. (Concludes that one's narrow contents can be expected to change whenever one receives substantial new information.)

Devitt, M. (1995) *Coming to Our Senses: A Naturalistic Program for Semantic Localism*, Cambridge: Cambridge University Press.(Spirited defence of molecularism.)

Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT

Press.(Defence of an atomistic view of content and meaning.)
Fodor, J.A. and LePore, E. (1992) *Holism: A Shoppers' Guide*, Oxford: Blackwell.(Critique of arguments for holism.)
Harman, G. (1973) *Thought*, Princeton, NJ: Princeton University Press.(Defence of a language of thought. Uses a functionalist point of view.)
Lewis, D. (1995) 'Lewis, David: Reduction of Mind', in S. Guttenplan (ed.) *A Companion to Philosophy of Mind*, Oxford: Blackwell.(Summary and defence of Lewis' views on mind and meaning.)
Loar, B. (1987) 'Social Content and Psychological Content', in R. Grimm and D. Merrill (eds) *Contents of Thought: Proceedings of the 1985 Oberlin Colloquium in Philosophy*, Tucson, AZ: University of Arizona Press.(Argument that narrow content is the only content.)
Lycan, W. (1988) *Judgment and Justification*, Cambridge: Cambridge University Press.(Includes a defence of empirical functional definitions of psychological terms.)
Putnam, H. (1965) 'The Analytic and the Synthetic', in *Philosophical Papers*, vol. 2, *Mind, Language and Reality*, Cambridge: Cambridge University Press, 33-69.(Discussion of the (in)significance of the analytic/synthetic distinction in science.)
Putnam, H. (1983) *Philosophical Papers*, vol. 3, *Realism and Reason*, Cambridge: Cambridge University Press. (A rich development of a conceptual role semantics.)
Putnam, H. (1988) *Representation and Reality*, Cambridge, MA: MIT Press.(Critique of cognitive science views of mind and meaning.)
Quine, W.V. (1953) 'Two Dogmas of Empiricism', in *From a Logical Point of View*, Cambridge, MA: Harvard University Press.(The classic argument against the analytic/synthetic distinction.)
Quine, W.V. (1954) 'Carnap and Logical Truth', in *The Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1976, 70-132.(Rich examination of the role of definition in logic, mathematics and science.)
Rey, G. (1997) *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Oxford: Blackwell. (Defends the prospect of empirical functional definitions of many mental states.)
Stich, S. (1983) *From Folk Psychology to Cognitive Science*, Cambridge, MA: MIT Press.(Argument against appealing to content in psychological explanation.)

Holmes, Oliver Wendell, Jr (1841-1935)

The most famous judge in the history of the USA, Holmes was also one of the most important US legal theorists. Within US jurisprudence, his work prefigured and stimulated the development of a general distrust of abstractions. Holmes emphasized that, in so far as logical deductions and abstract principles play any role at all in the law, they are side-effects of the ways in which struggles among different interests have been varyingly resolved. For him, the life of the law was not logic but experience; and experience was too diverse and conflict-ridden to be controllable on the basis of any sweeping formulae.

Born in Boston, Massachusetts, Holmes attended Harvard as an undergraduate and (after service in the American Civil War) as a law student. He then practised law for sixteen years, towards the end of which he published his most famous work, *The Common Law* (1881). After a very brief stint as Professor of Law at Harvard, Holmes became a Justice on the Massachusetts Supreme Judicial Court, where he remained for two decades. In 1902 he moved to the United States Supreme Court, where he served as a Justice for three decades.

Having befriended William James, C.S. Peirce and other like-minded men at Harvard, Holmes infused his jurisprudential writings with the spirit of pragmatism and legal positivism. Perhaps the most striking expression of that spirit in Holmes' work was the predictive theory of law which he developed in his famous essay 'The Path of the Law' (1897) and elsewhere. Although Holmes there spoke somewhat strangely of rights and duties as predictions, his analyses make clear that he viewed our *ascriptions* of rights and duties - rather than the rights and duties themselves - as predictive. To ascribe a legal right to someone is to predict that the person will (if necessary) be protected by the coercive power of the state in specified ways. Conversely, to ascribe a legal duty to someone is to predict that the person will (if necessary) be authoritatively punished for performing a forbidden act or for declining to perform a required act. Far from being mysterious entities, rights and duties are positions which people occupy in varying configurations; specifically, rights are positions of protectedness, and duties are positions of threatenedness.

Not only one's attributions of rights and duties, but also any other statements within and about the law, are essentially predictive. Laws and analyses of laws amount to forecasts of the consequences that attend disobedience. Such, at least, is the view which 'The Path of the Law' declares: 'The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by law' (1897: 461). This famous assertion did not bespeak a cynical 'anything goes' attitude towards legal decision-making, but simply flowed from Holmes' legal-positivist belief that official pronouncements and actions - as opposed to extra-legal standards - are the keys to the content and indeed to the very existence of law. Such a positivist stance can perfectly well square with the fact that what judges do in any of the liberal democracies is usually to interpret statutes or other legal materials (including past judicial decisions, of course) in an effort to find and apply relevant laws rather than create them entirely afresh. Though the distinction between finding and creating law is highly problematic, and though the particular means by which any legal system distinguishes finding from creating are decidedly variable, the judicial role in liberal democracies is based on such a distinction. Hence, a reduction of law to 'what courts do in fact' is fully compatible with the thesis that judges apply existent laws scrupulously.

Holmes' affinities with the pragmatists and his emphasis on predictability turn up further in his insistence that the relevant standards of evaluation within the law - standards such as reasonableness - are fundamentally external. Holmes pointed to two kinds of externality. First, the legal standards are brought to bear on people's behaviour rather than on their inner capacities and convictions. The law requires people to tailor their conduct to comply with the norms that regulate their interaction; if some people cannot live up to those norms (because of unintelligence, for example), they will usually be held responsible none the less. Outward deeds, as opposed to inward sentiments, are the overriding concern of the law. A second and partly related way in which the law's standards are external is that they typically flow from the *mores* of the general public rather than from the preferences of a wrongdoer or from the preferences of the decision-maker who passes judgment on the wrongdoer. A judge or a juror should seek to decide cases not on the basis of personal predilections but on the basis of ideals and expectations that are widely shared throughout the community in which the cases arise.

Holmes consistently exhibited a pragmatist distrust of abstractions. With his predictive theory of law, for instance,

he explicated the abstractions of legal parlance by reference to their concrete functioning. There are two basic complaints against a reliance on abstractions in his writings. First, he believed that such a reliance was pointlessly ineffectual. As *The Common Law* proclaims: 'An answer cannot be obtained from any general theory…. The grounds of decision are purely practical, and can never be elicited from grammar or from logic' (1881: 264). Abstract principles in themselves are empty and boundless, and they therefore simply distract us from our genuine problems. Furthermore, even if abstractions could yield determinate guidance, an adherence to them would be undesirable. As Holmes stated: 'To rest upon a formula is a slumber that, prolonged, means death' (1915: 3). Anyone who conjures with abstract formulations is attempting to lock the flux of reality into static moulds. For Holmes, who perceived human existence as a constant and highly salutary struggle among vying interests, the quest for overarching legal/ethical guidelines that could transcend all divisions was a denial of progress and of life itself.

See also: Common law §3; Law, philosophy of; Legal positivism; Legal realism §2; Pragmatism

MATTHEW H. KRAMER

List of works

Holmes, O.W., Jr (1881) *The Common Law*, Boston, MA: Little, Brown.(Holmes' *magnum opus*.)
Holmes, O.W., Jr (1897) 'The Path of the Law', *Harvard Law Review* 10: 457-78.(Predictive theory of law.)
Holmes, O.W., Jr (1915) 'Ideals and Doubts', *Illinois Law Review* 10: 1-4.(Short article on political idealism.)
Holmes, O.W., Jr (1918) 'Natural Law', *Harvard Law Review* 32: 40-44.(Doubts about natural law.)
Holmes, O.W., Jr (1995) *Collected Works*, 3 vols, ed. S. Novick, Chicago, IL: University of Chicago Press. (Thorough collection of Holmes' writings.)

References and further reading

Gordon, R. (ed.) (1992) *The Legacy of Oliver Wendell Holmes, Jr*, Stanford, CA: Stanford University Press. (Collection of essays.)
Grey, T. (1989) 'Holmes and Legal Pragmatism', *Stanford Law Review* 41: 787-870. (Highly readable account.)
Grey, T. (1992) 'Holmes, Pragmatism, and Democracy', *Oregon Law Review* 71: 521- 42.(A lucid analysis highlighting Holmes' political views as well as his pragmatism.)
Rosenblatt, R. (1975) 'Holmes, Peirce, and Legal Pragmatism', *Yale Law Journal* 84: 1123-40.(Introductory-level exposition.)

Holocaust, the

The specific, tragic event of the Holocaust - the mass murder of Jews by the Nazis during the Second World War - raises profound theological and philosophical problems, particularly problems about the existence of God and the meaning of Jewish existence. Among the thinkers who have tried to wrestle with the conceptual challenge posed by the destruction of European Jewry, three who have presented original arguments that can be termed, in a relatively strict sense, philosophical are Richard L. Rubenstein, Emil Fackenheim and Arthur A. Cohen.

Rubenstein has formulated an argument that turns on the theological difficulties raised by the realities of the evil of Auschwitz and Treblinka in a world putatively created and ordered by a benign God. For him, such evil decisively refutes the traditional theological claim that a God possessed of goodness and power exists, and entails the conclusion that 'there is no [traditional] God'. In working out this conceptual position, he uses an unsatisfactory empirical theory of verification concerning religious propositions and too narrow a notion of evidence, both historical and ethical, that ultimately undermines his counterclaims and 'Death of God' affirmations.

Fackenheim seeks not to defend a religious 'explanation' of the Holocaust, but rather to provide a 'response' to it that maintains the reality of God and his continued presence in human, and particularly Jewish, history. To do so, he uses Martin Buber's understanding of dialogical revelation, asserting that revelation is an ever-present possibility, and formulates his own moral-theological demand to the effect that after the Holocaust, Jewish survival is the '614th commandment' (there are 613 commandments in classical Judaism). Fackenheim's defence of this position, however, is philosophically problematic.

Cohen probvides a 'process theology' argument as an explanation of the Holocaust; that is, the Holocaust requires a revision in our understanding of God's nature and action. It forces the theological conclusion that God does not possess the traditional 'omni' predicates; God does not intervene in human affairs in the manner taught by traditional Western theology. However, Cohen's working out of a process theological position in relation to the Holocaust raises as many philosophical problems as it solves.

1 Richard Rubenstein

Rubenstein has been attempting to make sense of the Holocaust since the 1960s, when he published his well-known collection of essays *After Auschwitz* (1966). The theological position he articulated in this first work has changed little over the years: Auschwitz and Treblinka decisively refute any belief in the existence of God or that life is made meaningful by a transcendent source. Had God existed, he would not have permitted the Death Camps, he would have put a stop to the slaughter (see Evil, problem of). Since God does not exist, the universe is a meaningless conglomeration of circumstances without transcendental direction, purpose or warrant. Accordingly, Rubenstein rejects all such traditional Jewish beliefs as the sanctity of the Torah, the normativity of the *mitzvot* (commandments), and divine creation, covenant, revelation and redemption.

Beginning with this series of negations, Rubenstein, like such atheist existentialists as Jean-Paul Sartre, formulates an existential affirmation that creates human meaning without assistance, and so redeems human life from absurdity - to the extent that this is possible. In a world without inherent value, all we have is ourselves and each other; self-confirmation and communal belonging become ultimate values. The living Jewish community, in particular, is not rejected, but it is fundamentally redefined. In place of the divinely chosen, covenanted people of tradition, we find in the Jewish Community a human fellowship that is socially and psychologically supportive despite the denial of its ancient theological message.

Rubenstein gives essentially empirical grounds for his denial of the existence of God: propositions about God are to be straightforwardly confirmed or disconfirmed by appeal to empirical events in the world. On the basis of these events Rubenstein judges that 'God is dead'. The Holocaust becomes the critical empirical test, demonstrating God's nonexistence. In effect Rubenstein argues: there is too much evil in the world, there was too much at Auschwitz, to allow us any longer to affirm that God, as conceived in Judaism, can exist (see Atheism).

Respecting this challenge as an important one, too often lightly dismissed by theologians, and respecting Rubenstein's authentic existential response to an overwhelming reality, one needs to recall that empirical

falsifiability is not generally definitive in theological matters and does not provide the unimpeachable criterion that Rubenstein seeks. History provides evidence both for and against the existence of God, since there is both good and evil in history. The issue may turn on the questionable coherence of verificationism itself. But even one who accepts Rubenstein's assumption that theological judgments can be grounded in historical events, might dispute what counts as crucial empirical or experiential evidence. The State of Israel is also an empirical datum, and some argue that if Auschwitz is evidence against the 'God-Hypothesis', the re-establishment of the State of Israel is evidence in its favour. Neither position is decisively provable but both are equally meaningful.

In recent works like *The Cunning of History* (1975), *The Age of Triage* (1983), and the new edition of *After Auschwitz* (1992), Rubenstein has increasingly turned away from narrowly theological and philosophical concerns to a broader comparative sociological form of inquiry that draws on the thinking of Hegel and Max Weber and emphasizes the important role of the historical development of bureaucracy, technology and capitalism in deciphering not only the Holocaust but modernity at large. Rubenstein now sees the Holocaust as the culmination of a variety of negative processes that have increasingly defined modern European history since the sixteenth century. The Nazi Death Camps, he now contends, were an extension of the logic of New World black slavery, a rational and bureaucratic exploitation of labour taken to the ultimate extreme. The murder of European Jewry was the maximal application of modern notions of surplus population control. These sociological and historical claims have met with considerable critical resistance.

2 Emil Fackenheim

Emil Fackenheim is perhaps the best-known philosophical respondent to the Holocaust. Careful to insist that his work be understood as a 'response' to the Death Camps and not as an explanation of them, he vehemently rejects the notion that suffering is the result of sin (a classical Jewish doctrine sometimes applied to the Holocaust: that is, that it was God's way of punishing Israel for its sins). Fackenheim has sought principally to provide reasons for Jewish survival, to give meaning to ongoing Jewish communal life, after the Holocaust.

He begins with what he acknowledges to be a Buberian model of revelation: revelation is the ever-present possibility of divine-human relationship that speaks anew in every historical and inter-human context. This existential possibility is not subject to objective or public proof, any more than is the existence of God. But, taken on faith, this Buberian dogma, that God cannot be proven but can be met, undergirds Fackenheim's mature reflections on a post-Holocaust Judaism (see Buber, M.).

We witness God's presence in history, Fackenheim argues, in the experience of the people of Israel throughout its existence, and this in two ways: in the 'root experiences' of the people of Israel, and in 'epoch-making events'. The first are creative, extraordinary historical happenings that decisively form the character and continue to influence all future 'presents' of the Jewish people, forever legislating in every future generation of the nation. The Exodus from Egypt, for example, is a historical movement relived every Passover; its power affects each new generation, continually revealing in its yearly re-enactment the saving activity of God. Because past root experiences are lived through as present reality, Jews in every age are 'assured that the past saving God saves still'.

'Epoch-making events' are not formative for Jewish collective consciousness. They are historical challenges to the root experiences, demanding an answer to new and often unprecedented conditions. The occurrences encompassing the destruction of the First and Second Temples were such events. They tested the foundations of Jewish life, specifically the existence of the saving power and commanding providence of God, as previously revealed in the Exodus and at Sinai. But they did not shatter them, as the survival of the Jewish people testifies. The traditional interpretive paradigm has always shown its resiliency, absorbing and surviving the catastrophes that shook its foundations.

What, then, of Auschwitz? Can it be assimilated to Fackenheim's ancient and rather midrashic model? Fackenheim's answer is an unreserved 'yes'. Even Auschwitz does not destroy the 'root experiences' of Israel's faith. God is present even in the Kingdom of Night, commanding Israel still, from within the eye of the storm itself. This extreme reply of unwavering affirmation, in the face of the unprecedented terrors of Auschwitz, is the core of Fackenheim's response and that of others who follow his lead: a Jew cannot, dare not, must not, reject God. Auschwitz itself is revelatory, commanding, and one must learn to sense what God would reveal to Israel

even there.

The commanding word of Auschwitz, which Fackenheim, in a now famous phrase, called the 614th commandment, is this: 'Jews are forbidden to hand Hitler posthumous victories!'. After Auschwitz, Jews are under a sacred obligation to survive; Jewish existence is itself a holy act; Jews have a duty not only to remember the martyrs but to remain open to the possibility of redemption. To succumb to cynicism is to abdicate responsibility for the here and now and to deliver the future over to the forces of evil. Above all, Jews are 'forbidden to despair of the God of Israel, lest Judaism perish'. Hitler's demonic passion was to eradicate Israel from history. For the Jew to despair of the God of Israel as a result of Hitler's monstrous actions would be to do Hitler's work and by a dreadful irony to aid in the accomplishment of Hitler's goal.

The Jewish will for survival is natural enough, but Fackenheim invests it with transcendental significance: precisely because others would eradicate Jews from the earth, Jews are commanded to resist annihilation. Hitler makes Judaism after Auschwitz a necessity. To say 'no' to Hitler is to say 'yes' to the God of Sinai; to say 'no' to the God of Sinai is to say 'yes' to Hitler. Since 1945, every person who has remained a Jew has, from Fackenheim's perspective, responded positively to the commanding voice of God heard at Auschwitz.

What Fackenheim affirms here is only half of the traditional content of Judaism: the God of biblical faith is a commanding as well as a saving God. The crossing of the Red Sea is as central to Jewish history as the revelation at Sinai; both are 'root experiences'. Fackenheim has made much of the commanding presence of Auschwitz, but where is the saving power of the God of the Exodus? Without the crossing of the Red Sea there is no Sinai, as Fackenheim knows. He also knows that to talk of a God of deliverance, no matter how tentatively, after the Holocaust is highly problematic, since God did not work his kindness there and then. The continued existence of the people of Israel, however, and specifically the re-establishment and survival of the State of Israel, allow Fackenheim to risk speaking of hope and the possibility of redemption. The destruction of European Jewry and the birth of the State of Israel are, for him, inseparably linked (see Zionism). What the Holocaust seems to deny, the new state, at least tentatively, affirms, a living testimony to God's continued presence in history that reconfirms the 'root experience' of salvation essential to the survival of Jewish faith.

Despite its interest and wide appeal, Fackenheim's response is not philosophically unproblematic. At least three issues need to be raised: (1) Is Buber's 'I-Thou' dialogical theology an adequate basis for a post-Holocaust religious commitment (see Buber, M. §§3-4)? Can it ground Fackenheim's particular project? (2) What exactly can Fackenheim mean, in light of his Buberian starting point, by referring to a 614th commandment? That is, what are the meaning and force of the notion of a commandment in this context? (3) How can Fackenheim confront the issues of meaning, verification and disconfirmation that are raised but not adequately addressed by his important work?

While raising the level of discussion as regards the philosophical significance of the Holocaust, Fackenheim leaves us with a position that is not entirely conceptually compelling.

3 Arthur A. Cohen

Four highly ramified theses lie at the root of Arthur Cohen's philosophical reflections on the Holocaust, as expressed in his work *The Tremendum* (1981): (1) the Holocaust is unique, and its uniqueness entails particular theological consequences; (2) human thought cannot grasp the enormity of Auschwitz; (3) no 'meaning' is to be found in this genocidal carnage; (4) evil is more real, more consequential, than many, including Cohen himself, had heretofore allowed. These four radical reflections lead Cohen to recognize a need to return to the traditional questions of theology with a new uncertainty, to ask again concerning 'the reality of evil and the existence of God, the extremity of evil and the freedom of man, the presentness of evil and the power of God' (Cohen 1981: 97). They lead him to ask, still more concretely, if,

> like our ancestors we are obliged to decide whether (national) catastrophes are compatible with our traditional notions of a beneficent and providential God. The past generations of Israel decided that they were. The question today is whether the same conclusion may be wrung from the data of the *tremendum*.
>
> (Cohen 1981: 86)

Cohen's reply is that the traditional Jewish 'God concept' cannot survive the Holocaust but requires radical

revision along the lines suggested by Alfred North Whitehead and Charles Hartshorne (see Process theism). God is neither to be blamed for the Holocaust nor censured for not stopping the murder. But the idea of God and of God's relationship to the world needs to be reinterpreted in what Cohen calls a 'di-polar' manner. God, on this redefinition, though loving, is respectful of human freedom and, whether by choice or metaphysical necessity, does not interfere in human history.

Provocative as it is, Cohen's position raises a host of philosophical and theological questions. Is God still God if he is no longer an agent of providence? If he lacks the power to enter history vertically through miracles? Is the dipolar Absolute still the God to whom one prays, the God of salvation? Is the non-interfering God 'whom we no longer fear and from whom we no longer demand' still, as Cohen would have it, worthy of our 'love and honour?' Is this God now reduced to the status of Plato's Demiourgos (see Plato §16), or perhaps closer still to the innocuous and irrelevant God of the Deists (see Deism)? Such a God, it could be argued, makes no difference in how we act and plays no role at all in how history revolves or transpires. If God is indeed so absent from our life and from the historical record, what difference remains for us between this God and no God at all?

Furthermore, the proposed metaphysical reconstruction is founded on no direct phenomenological or experiential procedure, but fashioned, in the present instance, in response to the Holocaust. If so, belief in the dipolar God requires as great a 'leap of faith' as do the theistic affirmations of the ancient Jewish tradition - greater, perhaps, since the new God lacks the support of history and tradition.

It is difficult to see why one would turn towards dipolar theism, given the tragedy of the Holocaust, unless one were already committed to theism to start with. But Cohen has not shown why we should begin with this assumption. The moral dimension of theodicy remains to be dealt with, moreover. For Cohen's dipolar God appears, of necessity, indifferent to human suffering and acts of evil. In short, Cohen's revision of the 'God-concept', and the whole process-theology response to the Holocaust, raises at least as many questions as it answers (see God, concepts of).

4 Conclusion

The Holocaust has shown itself to be an extremely difficult matter to philosophize about. This is because of the enormity and complexity of the issues it raises for all human reflection and for Jewish philosophical deliberation in particular. It has not been shown that the *Shoah* (Holocaust) disconfirms all pre-Holocaust metaphysical and theological systems; but, given the reality that is the *Shoah*, it is correct to require that all philosophical and theological judgments be called upon to measure their conceptual adequacy against the circumstances of Auschwitz and Treblinka.

See also: Anti-Semitism; Evil, problem of; Fackenheim, E.L.; Jewish philosophy, contemporary

STEVEN T. KATZ

References and further reading

Cohen, A.A. (1981) *The Tremendum*, New York: Crossroad Publishing.(An interesting, if ultimately unsatisfactory, attempt to rework post-Holocaust Jewish thought in a Whiteheadian mode.)

Fackenheim, E.L. (1967) *The Religious Dimension in Hegel's Thought*, Bloomington, IN: University of Indiana Press.(A basic, informed study of Hegel's religious thought by one of the most prominent philosophers to focus on the Holocaust. Highly recommended.)

Fackenheim, E.L. (1968) *Quest for Past and Future*, Bloomington, IN: University of Indiana Press.(A collection of Fackenheim's essays.)

Fackenheim, E.L. (1970) *God's Presence in History*, New York: New York University Press.(Argues that we witness God's presence in history through the experiences of the people of Israel.)

Fackenheim, E.L. (1982) *Encounters between Judaism and Modern Philosophy*, New York: Basic Books.(A series of important essays on major modern philosophers including Kant, Hegel and Sartre in relation to Judaism.)

Fackenheim, E.L. (1978) *The Jewish Return into History*, New York: Schocken.(A second valuable collection of essays.)

Fackenheim, E.L. (1994) *To Mend the World*, Bloomington, IN: University of Indiana Press.(Fackenheim's most sustained, but only partially successful, effort to write a post-Holocaust theology.)

Katz, S.T. (1983) *Post-Holocaust Dialogues: Critical Studies in Modern Jewish Thought*, New York: New York

University Press.(Essays on key modern Jewish thinkers.)

Katz, S.T. (1992) *Historicism, The Holocaust, and Zionism: Critical Studies in Modern Jewish Thought and History*, New York: New York University Press.(Essays on the basic issues surrounding the Holocaust and Zionism.)

Katz, S.T. (1994) *The Holocaust in Historical Context*, New York: Oxford University Press, vol. 1; vols 2 and 3 forthcoming.(Discusses the uniqueness of the Holocaust in comparative historical perspective.)

Rubenstein R.L. (1966) *After Auschwitz*, Indianapolis, IN: Bobbs-Merrill; 2nd revised edn, Baltimore, MD: Johns Hopkins University Press, 1992.(Core essays on theology after the Holocaust.)

Rubenstein R.L. (1968) *The Religious Imagination*, Indianapolis, IN: Bobbs-Merrill.(Studies the relationship between psychoanalysis and Jewish theology and examines Jewish theology from a psychoanalytic perspective.)

Rubenstein R.L. (1975) *The Cunning of History*, New York: Harper & Row.(An important theoretical effort to situate the Holocaust within the stream of Western history, using a Weberian model.)

Rubenstein R.L. (1983) *The Age of Triage*, Boston, MA: Beacon Press.(Offers an explanation of the Holocaust within the context of modern culture in general and modern mass murder in particular.)

Schweid, E. (1994) *Wrestling Until Day-Break: Searching for Meaning in the Thinking on the Holocaust*, Lanham, MD: University Press of America.(A valuable critical study of post-Holocaust Jewish theology.)

Tiefel, H.O. (1976) ‘Holocaust Interpretations and Religious Assumptions’, *Judaism* 25: 135-49.(A helpful introductory evaluation of several post-Holocaust thinkers.)

Wyschogrod, M. (1971) ‘Faith and the Holocaust’, *Judaism* 20: 286-94.(Well-known critical appraisal of Emil Fackenheim’s work.)

Home, Henry (Lord Kames) (1696-1782)

Henry Home (better known as Lord Kames, his title as a Scottish judge of the Courts of Session and Justiciary) was an important promoter of and contributor to the Scottish Enlightenment. His philosophy was a search for causes which could serve as the bases of policies to improve most aspects of Scottish life and thought. His writings, like his involvements in clubs, government bodies and improving economic activities, were all means to enlightened and improving ends. Kames was perhaps the most typical thinker of the Scottish Enlightenment, uniting interests in philosophy, science, belles lettres, *history, education and practical improvements of every sort.*

Born in Eccles, Berwickshire, Henry Home, Lord Kames' parentage related him to the Jacobite and Whig gentry but also to the professional classes. Unfortunately, gentility did not confer wealth. Kames' notable interests in improvements were made necessary by his relative poverty which lasted until *circa* 1766 when his wife, Agatha Drummond, inherited the Blair Drummond estate in Stirlingshire. By the end of his career he had been honoured by societies or academies in Manchester, Philadelphia and Berne; he had also been attacked by Voltaire. At home his philosophic views seemed to many overly naturalistic and secular, even atheistic in their tendencies. Such conclusions were relatively easy to draw because Kames tended to deny free will, criticized Newtonian physics (the great prop to design arguments for the existence of God) (see Newton, I.) and extended the scope of sense and reason beyond what the religious were prepared to accept. He never produced a system based on the methods and views of his favourite philosophers - Francis Bacon, Isaac Newton, John Locke, the Third Earl of Shaftesbury and Joseph Butler - but in his works he laid out views which foreshadowed Scottish Common Sense philosophy (see Common Sense School).

Kames, like David Hume, Thomas Reid and other Scots, thought that all knowledge and belief had to be related to human nature, that is the structure and capacities of the mind. His clearest statements of this belief are found in *Essays on the Principles of Morality and Natural Religion* (1751, revised 1758 and 1779), but they are not lacking in other works. Kames found his great guides in sense and feeling. Belief comes from a simple, unanalysable feeling not identical with Hume's 'vivacity of perception' or other modes of awareness. If we have the feeling, we are caused and determined to believe in what we have perceived or remembered or what the critically evaluated testimony of others induces us to believe. Experience shows us a world in which causation is perceived - in ourselves as we will and act, in matter as objects at rest or in motion collide or move in determinable ways apparent to our senses of touch and sight. We do not know why or how objects move, but *motion*, *rest*, *power*, *cause* and *effect* are simple ideas derived from our experiences and requisite to the establishment of both morality and physics.

Kames differed from Newtonians in believing matter to be active and self-moving but bound by God to rules of motion which Newton's physics defined. Surveying nature revealed a lawful and designed world, a beautiful system of related causes and effects. Taken as a whole, it was an effect which led irresistibly to a belief in a creative and benevolent God. Sensations and feelings may mislead us but they do not regularly do so when inductions are carefully made. Moreover, we find in them signs of the substances whose qualities sensations reveal and of which they are correlatives. Through the external senses we perceive external objects; through the external and internal senses we know ourselves and others as enduring and identical persons.

The internal senses are important to Kames' morals and aesthetics. Morality, he thought, rests on instincts, feelings of pleasure and pain and sensations of right or wrong, propriety and impropriety, all of which are produced because we possess moral senses, 'the voice of God within us' (*Essays*, 1751: 63) (see Moral sense theories). This view owes more to Joseph Butler than to the Earl of Shaftesbury, Francis Hutcheson or George Turnbull. Kames may well have developed it independently, and prior to the publication of David Hume's earlier works.

Kames extended these views to law. Criminal laws rest on a sense of impropriety, upon the perception of the improper actor as guilty and one whose actions we resent and try to punish. Sympathy allows us to participate in the feelings of revenge raised in other injured parties. Laws in the various stages of society express all this in ways which depend upon levels of civility. Civil law rests in like manner upon changing notions of property, contract, society and of the wrongness of their violation. The laws of all peoples will, Kames believed, eventually converge, which led him to hope for a more perfect union between England and Scotland.

Kames developed his views of aesthetics in a somewhat similar fashion. A work of art evokes in our internal sense feelings of beauty, sublimity, grandeur, novelty and propriety. Understanding how this comes about and judging the means to their production are the functions of aesthetics and criticism. Kames did this so well that his *Elements of Criticism* (1767) remained a useful textbook in America until the mid-nineteenth century.

Kames' views were unacceptable to many. John Stewart, the pious Edinburgh Newtonian, attacked his science and metaphysics because it allowed for a Godless, self-acting, self-designing world of mere matter. Kames' denial of free will (1751) was condemned by Church courts in 1757 - an event which showed some of the limits to Scottish enlightened thought. His views on law were severely criticized by utilitarians, although the historicism with which they were associated survived (see Historicism). Indeed, his historical views resonated well into the nineteenth century, touching many fields such as sociology and anthropology and contributing to more general views of progress. In philosophy his work set out strategies pursued by the Scottish Common Sense philosophers, who more skilfully defended views he had held. When Lord Kames died many Scots mourned the man who, for contemporaries, was Scotland's most notable philosopher.

See also: Enlightenment, Scottish; Human nature, science of in the 18th century

ROGER L. EMERSON

List of works

Home, H. [Lord Kames] (1751) *Essays on the Principles of Morality and Natural Religion*, Edinburgh: Kincaid & Donaldson; revised London: Hitch & Hawes and Dodsley, Rivington, Fletcher & Richardson, 1758; Edinburgh: Bell & Murray, 1779.

Home, H. [Lord Kames] (1754) 'Essays on the Laws of Motion', *Essays and Observations, Physical and Literary*, Philosophical Society, vol.1, 1-69.(In the same volume see the reply to Kames' views by John Stewart, 'Some Remarks on the Laws of Motion and the Inertia of Matter', 70-140.)

Home, H. [Lord Kames] (1756) *Objections against the Essays on Morality and Natural Religion Examined*, Edinburgh.(This work contains passages which are thought to be by Hugh Blair and/or Robert Wallace, as well as by Home.)

Home, H. [Lord Kames] (1758) *Historical Law Tracts*, Edinburgh: Kincaid & Bell and London: Millar, 2 vols; revised and enlarged, Edinburgh: Bell & Creech and London: Cadell, 1776.

Home, H. [Lord Kames] (1761) *Introduction to the Art of Thinking*, Edinburgh: Kincaid & Bell.

Home, H. [Lord Kames] (1762) *Elements of Criticism*, Edinburgh: Kincaid & Bell and London: Millar; 6th edn with Kames' last additions and corrections, Edinburgh: Bell & Creech and London: Cadell & Robinson, 1785.

Home, H. [Lord Kames] (1774) *Sketches of the History of Man*, Edinburgh: Creech and London: Strahan & Caddell, 2 vols; 3rd edn with Kames' last additions and corrections, Edinburgh: Creech and London: Strahan & Cadell, 1778.(This has sections on the historical evolution of morals and philosophy generally.)

Home, H. [Lord Kames] (1781) *Loose Hints upon Education, Chiefly concerning the Culture of the Heart*, Edinburgh and London: Bell & Murray.

References and further reading

Barfoot, M. (1983) 'James Gregory (1753-1821), and Scottish Scientific Metaphysics, 1750-1800', unpublished Ph.D. dissertation, University of Edinburgh.(Best discussion of Kames' scientific and metaphysical views).

Emerson, R.L. (1991) 'Henry Home, Lord Kames', ed. D.T. Siebert, *Dictionary of Literary Biography*, vol. 104, *British Prose Writers, 1660-1800*, vol. 104, Detroit, MI: Gale Research International.(Most recent short general account with a bibliography and note on manuscript sources.)

McGuinness, A.F. (1970) *Henry Home, Lord Kames*, New York: Twayne Publishers. (General discussion of Kames' aesthetic and critical views.)

Norton, D.F. (1982) *David Hume: Common-Sense Moralist, Sceptical Metaphysician*, Princeton, NJ: Princeton University Press.(This contains an excellent discussion of Kames' views and their background.)

Ross, I.S. (1972) *Lord Kames and the Scotland of His Day*, Oxford: Clarendon Press. (Currently the best biography.)

Stocking, G.W., Jr (1975) "Scotland as the Model of Mankind": Lord Kames's Philosophical View of Civilization', in T.H.H. Thoresen (ed.) *Towards a Science of Man: Essays in the History of Anthropology*, The Hague: Mouton, 65-89.(Interesting discussion of Kames' place in the history of social thought).

Homer (*fl. c.*700 BC)

Author of the Iliad and Odyssey, the Greek poet Homer is probably the earliest surviving, and certainly among the greatest, of European poets. Down to the Renaissance he was considered the source of all scientific and philosophical wisdom; and he still supplies fruitful material for philosophical discussion of moral issues.

For ancient, and many modern, thinkers Homer was most interesting philosophically for the rather primitive cosmological views he no doubt took over from Greek popular belief. 'The Shield of Achilles' (*Iliad* XVIII), with its even-handed depiction of the totality of human life - war and peace, city and country, youth and age, birth and death, dance and trial - all encompassed within nature's temporal cycle of the seasons and the spatial horizon of the heavens and all-surrounding Ocean, sets the bloodily narrow-minded action of the Trojan War in a larger, humanizing context; this image of the largest sources and limits of our human world is filled out by scattered references elsewhere to the brazen heavens, to Ocean, the origin of all things, and to Night's primordial power. Familiarity with an allegorical interpretation of divine myths is implied already in some passages in Homer himself (for example, 'the Battle of the Gods', *Iliad* XX, XXI); this technique went on to become a central way of reinterpreting his authoritative text for later needs: in part fostered by Homer's role as a major author read in schools, this long-lived tradition could claim that numerous philosophical and scientific doctrines were anticipated and legitimized by the poet's words.

In modern times, philosophical attention has tended to focus instead upon Homer's narrative account of human action and responsibility. Homer's poetry, although written, is the culmination of an oral epic tradition, in which many generations of poets gradually refined and improved an inherited body of traditional legendary tales. Homer seems to have inherited stories of divine determination and bloodthirsty war but himself to have preferred to tell them, at least in part, in terms of moral responsibility and humane compassion: the *Iliad*, whose first word is 'wrath', turns out in the end to be no less concerned with pity. The generic tension produced by an oral epic evolving over time through incrementally quite different ethical conceptions ended by posing for Homer the fundamental question to what extent freedom of choice, individual responsibility and feelings of guilt were compatible with the determinative influence of seemingly external factors like gods, oracles, destiny and temporary infatuation. His answer takes the form not of stringent argument but of compelling narrative and, above all, of direct speech in which characters lay bare the wellsprings of their action with breathtaking clarity: external factors turn out always to be involved and to affect the outcome of choice and action, but they never prevent human agents from choosing freely, erring irremediably and being regarded, by themselves and by others, as fully culpable - at least until they choose to accept their human limitations and to acknowledge their culpability in the eyes of their society and, especially, in the eyes of those they have wronged.

GLENN W. MOST

List of works

These cannot be dated securely, but the *Iliad* is generally regarded as earlier than the *Odyssey*.

Homer (*c.*700 BC) *Iliad*, in D.B. Monro and T.W. Allen (eds) *Homeri Opera*, vols 1 and 2, Oxford: Clarendon Press, 3rd edn, 1920; A.T. Murray (ed. and trans.) *Homer, The Iliad*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1924, 2 vols. (Monro and Allen provide the scholarly edition; Murray gives Greek text with an English prose translation.)

Homer (*c.*700 BC) *Odyssey*, in T.W. Allen (ed.) *Homeri Opera*, vols 3 and 4, Oxford: Clarendon Press, 2nd edition, 1917-19; A.T. Murray (ed. and trans.), revised by G.E. Dimock, *Homer, The Odyssey*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1995, 2 vols.(Monro and Allen provide the scholarly edition; Murray gives Greek text with an English prose translation.)

References and further reading

Adkins, A.W.H. (1960) *Merit and Responsibility. A Study in Greek Values*, Oxford: Clarendon Press.(The first four chapters provide one of the most influential and controversial accounts of Homeric ethics.)

Buffière, F. (1956) *Les Mythes d'Homère et la pensée grecque (Homer's Myths and Greek Thought)*, Paris: Les Belles Lettres.(The standard survey of ancient allegorical interpretations of Homer's myths.)

Dodds, E.R. (1951) *The Greeks and the Irrational*, Berkeley and Los Angeles, CA: University of California Press.

(Chapter 1 pages 1-27 is a widely influential study of irrational behaviour in Homer in terms of psychology and anthropology.)

Gaskin, R. (1990) ’Do Homeric Heroes Make Real Decisions?’ *Classical Quarterly* 84, new series, 40: 1-15.(A careful analysis of Homeric decision-making.)

Hernandez-Galiano, M., Heubeck, A. and Russo, J. (1988-92) *A Commentary on Homer’s Odyssey*, Oxford: Clarendon Press, 3 vols.(The commentary is uneven but comprehensive.)

Kirk, G.S. (1985-93) *The Iliad: A Commentary*, Cambridge: Cambridge University Press, 6 vols.(The commentary is uneven but comprehensive.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(Pages 7-17 provide a brief but balanced treatment of Homeric cosmology.)

Lamberton, R. (1986) *Homer the Theologian. Neoplatonist Allegorical Reading and the Growth of the Epic Tradition*, Berkeley and Los Angeles, CA, and London: University of California Press.(A general survey of one important aspect of the tradition of allegorical interpretation of Homer.)

Lamberton, R. and Keaney, J.J. (eds) (1992) *Homer’s Ancient Readers. The Hermeneutics of Greek Epic’s Earliest Exegetes*, Princeton, NJ: Princeton University Press.(A useful collection of essays ranging from Homer to the Renaissance.)

Snell, B. (1953) *The Discovery of Mind*, trans. T.G. Rosenmeyer, Oxford: Oxford University Press.(Chapter 1 is a widely influential presentation of a ’primitivist’ view of Homeric psychology and ethics.)

Williams, B. (1993) *Shame and Necessity*, Berkeley and Los Angeles, CA, and Oxford: University of California Press.(An eloquent defence of the sophistication of Homer’s views on the self and action.)

Honour

Honour consists in living up to the expectations of a group - in particular, in keeping faith, observing promises, and telling truth. This restriction to a particular group is not easy to justify against the background of universalist theories of ethics, and neither does honour accord readily with the central modern concepts duty and utility. Honour requires a social context in which individuals can bind themselves, and has tended to be restricted to free, adult males, who alone have been thought to have the capacity to bind themselves in this way. Older traditions, however, regarded honour as the goal of all virtuous action; and newer thinkers have been rediscovering attractions in this view.

Honour is the disposition to fulfil promises and commitments, whether explicit or implicit, made to other members of a group. Honour has traditionally been important in public affairs, with political and commercial life resting on the notion that commitments must be 'honoured' even without explicit obligation by the person concerned.

Even thieves have honour, however, which seems to show that honour's compatibility with dubious principles makes it a crude guide to moral action. In many respects, indeed, it looks morally atavistic. Hobbes' disquisition about the nature of honour (1651) seems to voice only the crude principles of a society at war (compare Reiner 1972). Today, almost all functional applications of honour have been individualized and legalized. Nobody today fights duels; attempts to vindicate feminine honour would most often, if not always, meet with scorn; politicians no longer tend to resign; conduct in commercial organizations and other traditional concerns of honour are regulated by law. Only at the margins does scope for honour remain, for example *vis-à-vis* gambling debts, which are not recoverable by legal action.

Most modern philosophers would regard honour as superficial to the proper concerns of moral philosophy. Kant's contrast between honour and 'what is in fact useful or dutiful' is probably typical ([1785] 1903: 398). Utilitarianism finds honour inexplicable, for the value of honourable behaviour is not rationally determined by its consequences (see Utilitarianism; Consequentialism). For deontology, the relativity of honour condemns it as cynical (compare the thieves) (see Deontological ethics).

None the less, Aristotle, in *Nicomachean Ethics*, called honour the 'end of virtue' (1115b13); modern virtue ethics more or less explicitly follows this tradition (see, for instance, MacIntyre 1980, Solomon 1993); and even consequentialists seem to be rediscovering its appeal (expressly in Hollis 1992). What has mainstream ethical individualism, with its bare choice between consequentialism and deontology, overlooked?

The structure of honour seems to be as follows. At its heart is a comparatively formal disposition, namely 'fidelity, the observing of promises, and telling truth' (Hume 1741/2: 294). The commitments entered into are not contractual in nature and do not presuppose reciprocal utility. Honour involves a context in which others also assume comparably onerous roles; but despite this background, individuals who assume a role bind in the first instance only themselves.

Only agents considered to have full capacity can bind themselves in this way. Historically, and to a degree even now, children, slaves and women cannot bind themselves, and hence commitments in relation to their roles have to be taken over by proxies (usually males of full age and capacity). Capacity is a condition for the possession of honour.

The content of the commitments assumed by the agent is dependent on social context - which is why thieves can still be honourable. The commitments may coincide with duty (deontologically conceived) or even utility; but they may equally be virtues of a kind not always directly reconstructible in this way, such as continence, bravery, generosity and so forth. They are not generalizable or absolute; the content of honour varies with the honourable society, and is subject to no universal codes.

The fact and degree of honour, by contrast, are amenable to general ascertainment, for the quantity of commitments satisfied is a measure of the agent's capacity. Honour is most evident in those with the means to fulfil conspicuous promises. This is why princes and 'magnificent persons' are favourite illustrations of honour (in Aristotle and Hobbes), and why incapacities such as illness or poverty, even when undeserved, diminish honour.

For Hobbes, honour is by its very nature a display of *power*, and 'to honour' is to acknowledge that power in others.

Honour is the final goal and completion of the virtues, which thus do not have value independently of the moment when they are undertaken in commitment to an honourable life (*Nicomachean Ethics* 1115b13; compare also 1122b7).

There are several obvious difficulties connected with the concept of honour. First, it is sometimes argued that honour is primarily a matter of status or of passively acquired reputation. If, however, honour is the sum of a person's virtues, then personal merit must be involved to a degree. So contrary to Solomon's view (1993), honour does depend on what one has done, or is taken to be prepared to do; simply being 'a member of the group in good standing' is a defective form of honour.

A second problem is the apparent connection between honour and uncritically affirmed collectives, as in 'the honour of the school', and so on. However, although honour requires an interpersonal context, only individuals possess it. Honour is the sum and goal of an individual's virtues; it is not properly attributable to groups (football teams, nations, religions, or even, in the final degeneration of this kind of thinking, races). Commitment to the interests of some specific group, while no doubt essential to honour, does not entail conferring personality on it.

Third, honour reflects the interests and concerns of its social context, and other societies' notions of honour can seem alienating or downright immoral. Nonetheless honour is a formal concept and can in principle be adapted to 'enlightened' concerns. Its central structural motif, that value resides not in individual acts or dispositions, but in their participation in an interpersonally directed whole, is present in a number of modern ethical systems. It is most obvious in Nietzsche, who in his doctrine of the 'will to power' institutes the same equivalence between honour and power as does Hobbes, while at the same time allowing for the virtues to be completed by power in a broader, nonpolitical sense (see Nietzsche, F. §11). It also emerges in G.E. Moore's insistence that virtues have value only when practised for the sake of a higher 'organic unity' (1903: 225, 238; compare *Nicomachean Ethics* 1115b13) (see Moore, G.E.).

It remains an open question, perhaps, whether the nonuniversalism of systems founded on honour is not too restrictive for the scope of modern ethical intuitions, notably in such areas as international human rights (see Universalism in ethics §1). This is a difficulty for all nonuniversalist doctrines such as virtue ethics and communitarianism (see Community and communitarianism).

See also: Bushi philosophy; Desert and merit; Virtues and vices

JULIAN ROBERTS

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985.(Classic ancient discussion of honour and the virtues.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.(Discusses honour in a society at war.)

Hollis, M. (1992) 'Honour Among Thieves', in M. Hollis and W. Vossenkuhl (eds) *Moralische Entscheidung und rationale Wahl*, Oldenbourg: Scientia Nova. (Considers honour in a consequentialist context.)

Hume, D. (1741/2) *Political Essays*, ed. K. Haakonssen, Cambridge: Cambridge University Press, 1994. (Discusses honour and fidelity.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, in *Kants gesammelte Schriften*, ed. Königlichen Preußischen Akademie der Wissenschaften, Berlin: Reimer, vol. 4, 1903; trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(References made to this work in the entry give the page number from the 1903 Berlin Akademie volume; these page numbers are included in the Paton translation. The work provides a classic exposition of a deontological standpoint.)

MacIntyre, A. (1980) *After Virtue*, London: Duckworth.(Influential rejection of the dichotomy beteween deontology and consequentialism.)

Moore, G.E. (1903) *Principia Ethica*, ed. T. Baldwin, Cambridge: Cambridge University Press, 1994.(Discusses value in 'organic unities'.)

Reiner, H. (1972) 'Ehre', in *Historisches Wörterbuch der Philosophie*, vol. 2, Darmstadt: Wissenschaftliche

Buchgesellschaft.(Includes discussion of Hobbes on honour.)

Solomon, R. (1993) *Ethics and Excellence*, Oxford: Oxford University Press, esp. 222.(A recent application of virtue ethics.)

Hooker, Richard (1554-1600)

Hooker's Of the Laws of Ecclesiastical Polity (1593-1662) is the first major work of English prose in the fields of philosophy, theology and political theory. After setting out an entire worldview in terms of the single idea of law, Hooker attempted to justify - and, arguably, to transform - the religious and political institutions of his day. Hooker's work contributed to later, more narrowly political, political thought (Locke cited 'the judicious Hooker' at crucial points in his Second Treatise of Civil Government), but the Laws is chiefly significant for articulating the ideal of a society coherent in and through its religion, a body politic which succeeds in being - not merely having - a church. In Hooker's England this meant that royal authority in religion was extensive, but derived from the community and limited by law. Modern separations of politics from religion and of philosophy from edification have made him difficult to assimilate. More recent critiques of Enlightenment secularism and purely technical philosophy help make him again intelligible.

1 Life

Hooker was born in England, near Exeter. With the help of Bishop John Jewel, the first official apologist for the Elizabethan religious settlement of 1559, Hooker attended Corpus Christi College, Oxford, where he later taught logic and Hebrew. In 1585 his appointment as Master of the Temple made him chief pastor of an important centre of legal studies. His surviving sermons, all from this period or earlier, show notable concern for the crises of faith experienced by reflective Christians during the Reformation. In 1591 he resigned from the Temple to work on *Of the Laws of Ecclesiastical Polity*. The preface and first four books were published in 1593, Book V in 1597. Hooker died in 1600. The politically sensitive last three books were published, in unfinished form, between 1648 and 1662.

2 The kinds of law and other principles

Although formally addressed to Puritan critics of the Elizabethan religious establishment, *Of the Laws of Ecclesiastical Polity* is also directed to the large number of Hooker's compatriots who were reluctant or unwilling to participate in the life of the English church after Pius V's excommunication of the Queen in 1570. Hooker's aim in Book I was thus to lay out normative principles acceptable to the broadest possible range of Christians at a time when violent accentuation of differences was more usual.

The scheme of laws Hooker set forth in Book I comprises (1) a twofold eternal law, in accordance with which (a) God's own voluntary actions are carried out and (b) creatures are directed to their various ends; this second eternal law subdivides, in application to the different kinds of creatures, into (2) a law governing the lives of the angels; (3) laws of purely 'natural' (that is, non-rational) agents; (4) principles of individual and social morality rationally discoverable by human beings; (5) laws made by humans themselves, consisting of (a) 'mixedly' human laws, which set down what we are already obliged to do on rational grounds, and (b) 'merely' human laws, which make obligatory actions which had not previously been obligatory; and (6) divine law, the supernatural law of biblical revelation, by which God directs fallen human nature to eternal blessedness.

It is a mark of Hooker's success in constructing a generally acceptable normative context for his later discussion of disputed issues that Book I can be read as an extraordinarily eloquent but unoriginal distillation of the wisdom of the ages. There is more to it than that. Hooker weaves together some quite diverse past wisdoms, and his elaboration of at least three particular points has substantive significance. First, in defining law as any rule directing to goodness of action, Hooker explicitly refrains from making imposition by a superior essential to law. Further, he argues that human beings are naturally equal with respect to political association. Accordingly, he urges the necessity of consent as a basis for governmental authority and, particularly, for legislative authority (see Authority §3).

The presentation of a lawful cosmos in Book I of the *Laws* has commonly been taken to be Hooker's greatest claim to philosophical merit. Other, more overtly polemical, parts of his work also deserve attention. Before undertaking his account of 'laws and their several kinds in general' in Book I, Hooker had found it necessary to gain breathing space from the Puritan contention that God's law, revealed in the Bible, uncompromisingly commanded major changes in church and society. The preface he composed for this purpose includes an astute analysis of what

Hooker presents as an implicitly revolutionary movement, as well as discussion of the comparative authority of conscience and consensus.

Following his attempt to set out acceptable general norms in Book I, Hooker sought in Books II-IV to refute three principles he saw as underlying the numerous particular Puritan complaints. These principles were (1) that the Bible is the sole rule for everything to be done in this life; (2) that the Bible must of necessity contain full directions for church government; and (3) that the adequacy of a church's reformation could be measured by its distance from the practices of the Roman church. Against (1) Hooker defended (partly on biblical grounds) the competence of human reason. Against (2) he defended the competence of individual national churches to govern their own affairs. Against (3) he developed a historically contextualized account of the scandalous and the edifying in religion.

3 The public duties of religion

The fifth and longest book of the *Laws* is primarily a defence of particular features of the Elizabethan Book of Common Prayer, which was thought by Puritans to be insufficiently Protestant. Hooker begins the book's two main sections, however, with general discussions of the contributions of religion to social well-being. Even in its details, Hooker's championing of public worship as the central setting for personal participation in the life of God is of philosophical interest. It challenges the widespread modern assumption that faith is purely an inward matter or, at its most external, one component of individual ethics - in any case something private rather than public. (Hooker's account of the 'virtue' and 'discipline' of repentance in Book VI also negotiates the border between the subjective and outward sides of religion and does so with a sensitivity unusual in a defender of law and order.)

4 Supreme authority in religion

In the three last books of the *Laws* Hooker turned to the weightiest political problem of his day, the location of supreme human authority in religion. Three positions concerned him: the Presbyterianism advocated by some Puritans as divinely ordained; the papal supremacy advocated as divinely ordained by, among others, Jesuits clandestinely active in England after 1570; and the supremacy of the English crown asserted in parliamentary acts under Henry VIII and Elizabeth I.

Some of Hooker's case against Presbyterianism can be gathered from earlier books, from the preface, and from his defence of episcopal authority in Book VII, but his chief objection both to Presbyterianism and to papal jurisdiction independent of national authority is put most cogently in the first chapter of Book VIII: both positions assume an untenable dualism of 'church' and 'commonwealth' even when the membership of both is identical.

With reliance on Aristotle (§§27-8), Hooker argued that the aims of political association included psychological goods (goods of the soul) and that, since the most important of these was religion, every body politic properly 'cared' for religion. Given the assumption, uncontested in his circumstances, that Christianity is the true religion, Hooker concluded that a commonwealth of Christians is a church, a body politic distinguished from other bodies politic, not by religion but by truth of religion: 'We say that the care of religion being common unto all *Societies* politic, such *Societies* as do embrace the true religion, have the name of the *Church* given unto every of them for distinction from the rest' ([1593-1662] 1989: VIII 1.2).

On this analysis, a royal supremacy of some kind was one way for a (national) community of Christians to care for its Christianity. Hooker's emphasis on the theoretically optional, communally chosen character of the royal supremacy is noteworthy. It may indeed explain why Book VIII was not published during the heyday of divine right monarchy under the early Stuarts. Besides grounding the crown's authority in communal consent, Hooker was also more insistent than any contemporary English author in construing that authority as legally limited. And further, as his recently discovered notes for Book VIII reveal, he intended to conclude the book with a substantial discussion of whether and how a community might correct an errant supreme ruler.

Although Hooker defended a supremacy in religious affairs granted to the civil ruler of the land, the limitations he placed on the royal supremacy, at a time when absolutism was budding in both theory and practice, suggest that Locke's use of Hooker to support the authority of community and law over regal power has more to be said for it than is often supposed. However, while Hooker could still envisage an agreed integration of spiritual and civil life as a feasible social goal, Locke saw the achievement of such coherence as a problem for individuals, who were

likely to solve it in fundamentally different ways.

See also: Bodin, J.; Law, philosophy of; Laws, natural; Locke, J.; Natural law; Political philosophy, history of; Sovereignty

A.S. McGRADE

List of works

Hooker, R. (1582-91) *Tractates and Sermons*, in W.S. Hill (ed.) *The Folger Library Edition of the Works of Richard Hooker*, vol. 5, Cambridge, MA: Harvard University Press, 1990.(Hooker's preaching on issues of faith and conscience that were of central concern during the Reformation.)

Hooker, R. (1593-1662) *Of the Laws of Ecclesiastical Polity*, in W.S. Hill (ed.) *The Folger Library Edition of the Works of Richard Hooker*, vols 1-3, Cambridge, MA: Harvard University Press, 1977-81; vol. 6, Binghamton, NY: Medieval & Renaissance Texts & Studies, 1993.(Volumes 1-3 provide the main text; volume 6 gives discursive introductions and detailed commentary. The Preface and Books I-V of the *Laws* (published 1593-7) criticize the origins and tactics of the Puritans (including their exclusive reliance on biblical authority and unqualified opposition to Roman Catholicism), lay out Hooker's comprehensive normative framework, and defend the forms of public religious life prescribed and embodied in the Elizabethan Book of Common Prayer. Books VI-VIII (published posthumously between 1648-62) discuss the spiritual jurisdiction claimed for lay-elders by those pressing for a Presbyterian church polity, the authority and honour of bishops, and the power of ecclesiastical 'dominion' granted to the crown by English law.)

Hooker, R. (1593-1662) *Of the Laws of Ecclesiastical Polity: Preface, Books I and VIII*, ed. A.S. McGrade, Cambridge Texts in the History of Political Thought, Cambridge: Cambridge University Press, 1989.(Hooker's address to Puritan critics of the Elizabethan religious establishment (with an analysis of their movement), the scheme of norms he proposed as a basis for adjudicating disputed issues, and his reformative defence of the royal supremacy in the English church.)

References and further reading

Hill, W.S. (ed.) (1972) *Studies in Richard Hooker: Essays Preliminary to an Edition of his Works*, Cleveland, OH: Case Western Reserve Press.(Bibliography; includes a perceptive, in the end critical essay by W.D.J. Cargill Thompson on Hooker as a political thinker.)

Kirby, W.J.T. (1990) *Richard Hooker's Doctrine of the Royal Supremacy*, Leiden and New York: Brill.(Argues for Hooker's fidelity to the Magisterial Reformation of Luther and Calvin; see also A.S. McGrade's review in *Journal of Ecclesiastical History* 44 (1993): 315-9.)

Lake, P. (1988) *Anglicans and Puritans? Presbyterian and English Conformist Thought from Whitgift to Hooker*, London: Unwin, Hyman.(Argues for Hooker's originality in relation to previous defenders of the Elizabethan religious settlement.)

McGrade, A.S. (ed.) (1997) *Richard Hooker and the Construction of Christian Community*, Tempe, AZ: Medieval & Renaissance Texts & Studies.(Bibliography; papers on different sides of central issues concerning Hooker's historical and current significance.)

Shuger, D.K. (1990) *Habits of Thought in the English Renaissance: Religion, Politics, and the Dominant Culture*, Berkeley, CA: University of California Press.(Sensitive to tensions within, as well as between, establishment and anti-establishment thought.)

Hope

In Christian theology 'hope' has a central role as one of the three theological virtues. As theology has gradually become separated from moral theory, the inclusion of 'hope' within a theory of ethics has become rare. Hope can be either intentional or dispositional. The former is a specific hope for something, whereas the latter is a state of character. Kant gave a central place to intentional hopes in his moral theory with his doctrine of the postulates. Hope also played an essential role in the moral and political writings of Ernst Bloch and Gabriel Marcel. Bloch regarded hope as concerned with a longing for utopia, whereas Marcel regarded hope as a disposition to rise above situations which tempt one to despair. In each of these writers the Christian connection between hope, on the one hand, and faith and love, on the other, remained, although Kant and Bloch did not oppose these categories to reason, but sought to 'subsume' them under it.

1 Historical outline

Hope first plays a significant role within Augustine's theory of morality where it appears as one of the three theological virtues - faith, hope and charity, or love (see Augustine §8; Theological virtues). According to Augustine, the virtue of hope is subordinate only to faith, and on a level footing with love: 'there is no love without hope, no hope without love, and neither love nor hope without faith' (*Enchiridion*).

As morality became separated from theology during the Enlightenment hope became less important in ethics. There are, however, three notable exceptions to this general trend: Immanuel Kant, Ernst Bloch, and Gabriel Marcel (see §§3-5). In each of these writers the connection between hope, on the one hand, and morality, religion and faith, on the other, is as tight as it was for Augustine. However, both Kant and Bloch interpret religion and faith neither as revealed nor as metaphysically proven, yet as reasoned, while Marcel contrasts both hope and faith to human desiring and calculating, that is, to instrumental rationality.

2 The concept of hope

Hope can be either intentional or dispositional (see Intentionality). Intentional hope is oriented to some desired state of affairs which is believed to be attainable. Hobbes describes such hope as an '*appetite* with an opinion of attaining' (1651: 123; original emphasis). Dispositional hope is a state of being hopeful. This state should be distinguished from optimism and wishful thinking.

A philosophical formulation of optimism is given in Leibniz's work (1710). For Leibniz, the thought that God could have created anything less than the best of all possible worlds is self-contradictory, since it would mean that God would not be the most perfect being, and hence would not be God (see Leibniz, G.W. §§3, 7). Consequently, we can be certain that present suffering is for the greater good and that the future will be as good as it can be (see Evil, problem of §2). This confidence that everything is for the best was famously, and justifiably, ridiculed in Voltaire's *Candide* (1759). But, unlike optimism, being hopeful is not based on the certainty that things will get better. Hope, unlike optimism, presupposes a degree of uncertainty. Descartes noted this when he pointed out that '[w]hen hope is extreme, it changes its nature and is called "confidence" or "assurance"' (1649: 351) (see also Spinoza 1677: 106-7).

Hope should also be distinguished from wishful thinking. It is as impossible to hope legitimately for something which is impossible as it is inappropriate to hope for something which is certain. All wishful thinking is by its nature illusory, whereas only illegitimate hopes are illusory.

3 Hope in Kant's moral philosophy

Like Augustine, Kant held that hope and love presuppose faith. Kant, however, viewed faith as moral faith. Moral faith, unlike what he called ecclesiastical faith, is not based upon belief in metaphysical truth or contingent historical fact, and so available only to those who know of this truth or fact, but is a reasoned corollary of the moral law, and available to all (1793: 93-4). For Kant, the ultimate hope is no longer a hope for union with God, but a hope to attain the highest good (*summum bonum*) which is demanded by the moral law. The three postulates of pure practical reason, as laid out in the *Critik der practischen Vernunft (Critique of Practical Reason)* (1788), namely the hope for the existence of God, for immortality and for the attainment of perfect freedom, are all

subordinate to this superordinate hope, and ultimately to the moral law.

Kant describes the assumption of the validity of the three postulates as ‘a need of reason’ (1788: 149). The use of the term ‘need of reason’ is not meant to suggest that the hope for the existence of God, immortality and freedom are necessary illusions. These postulates are both possible from a theoretical point of view, and objectively justified from a practical point of view by the fact that they are necessary corollaries of the moral law, which itself is based on reason (1788: 127). Since the postulates are the necessary and unavoidable conditions of the possibility of the *summum bonum*, every rational being ought (morally) to hope for the existence of God, immortality and freedom (1788: 149).

This moral argument for the belief (or hope) in God and immortality, has been subjected to a great deal of criticism. Nevertheless it has had an important influence on certain figures in twentieth-century critical theory. Walter Benjamin held that the historical materialist cannot abandon hope for those who have suffered past injustice, and therefore, cannot regard the past as closed. To do so would be implicitly to side with the victors and betray their victims (see §5 for an account of hope as a relation of communion or love).

Peukert also takes up this Kantian/Benjaminian approach to hope in *Science, Action and Fundamental Theology* (1984). Here he argues that utopianism requires hope (see Utopianism). Utopian existence would be destroyed if the individuals who participate in it were burdened with the memory of unredeemed suffering. He asks: ‘How can one hold on to the memory of the conclusive, irretrievable losses of the victims of history - to whom one owes one’s entire happiness - and still be happy?’ (1984: 209). If Peukert’s argument is sound, the very idea of utopia is incoherent without hope for immortality, for the redemption of the victims of history, and the hope that God exists.

4 Hope in Bloch’s thought

Ernst Bloch takes a Hegelian approach to religion. He conceives God not as radically opposed to the human subject, as the absolutely Other, in the way Kierkegaard does (1844: 44-5), but as a human ideal. Bloch interprets faith in God as a hypostatized longing for a utopian form of existence. Once this hypostatization is recognized for what it is, the concept of ‘God becomes the kingdom of God, and the kingdom of God no longer contains a god’ (1954-9: 1196). God is recognized as embodying the hope for an ideal but thoroughly human kingdom.

Bloch’s anthropological interpretation of religion owes a great deal to Feuerbach. He was, however, deeply critical of Feuerbach for eliminating all traces of the transcendent from religion. The human perfection which Feuerbach saw man as projecting onto God is not reappropriated in the recognition of this projection, but is recognized as the promise of reappropriation - that is, as the promise of human perfection. This promise of perfection is the transcendent element which remains in religion, even after its anthropologization (1954-9: 1284-8).

Bloch’s main critics included orthodox Marxists such as Manfred Buhr (1970), who were deeply suspicious of religious utopias. They see religion as tranquillizing, and socially stabilizing, as the ‘opium of the people’. Bloch acknowledged this conservative tendency within religion, but also interpreted the revolutionary tendency of religious hope as the most extreme manifestation of dissatisfaction with, and rebellion against, the present. For Bloch, religious utopias, especially the Christian utopia of the kingdom of God, express revolutionary hope for an utterly different and better form of existence (1954-9: 1193). They are not merely tranquillizing, but in principle transforming.

5 Marcel on hope

For Marcel, hope is a way of overcoming the trials of life. These trials can take the form of, for example, illness, separation, exile or slavery. ‘Hope is situated within the framework of the trial, not only corresponding to it, but constituting our being’s veritable *response*’ (1945: 30; original emphasis). Such trials are a potential cause of despair in which subjects ‘go to pieces’ or lose themselves (1945: 37). So there is no hope without the temptation to despair, and hope is understood by Marcel as ‘the act by which this temptation is actively or victoriously overcome’ (1945: 36).

Marcel is concerned not with intentional hopes, but ultimately with a particular form of dispositional hope which he expresses as ‘I hope in thee for us’ (1945: 60). To place one’s hope in somebody is not the same as to hope for something, that is, for some desired object, although it is usually related to such a hope. I do not want the person in

whom I have placed my hope (although I may hope to gain some desired object by placing my hope in another person).

Hope is here understood as trust without self-interest. It is not like the hope of patients who place their hope in the psychiatrist, which can be expressed as ‘I hope in thee *for me*’. Such hope is still bound up with self-interest. Marcel sees the fundamental form of hope as expressing human dignity as the ability to transcend one’s own desires and to hope for a shared project. Here hope transcends self-interest: my relation to the other person (or persons) in whom I have placed my hope is not instrumental, but a collaborative, ethical relation of communion (1945: 67). Marcel locates such hope against a background of faith in God (1945: 60-1), rather than seeing it as Kant did against the background of the moral law.

Although the ‘thou’ in whom I hope is, for the most part, another person or group, Marcel maintains that we can regard reality itself as a thou. We do this when we cease to regard it as something which can be described exhaustively in terms of a catalogue of facts, and of what can be inferred from these facts, and regard it as an open process, that is, as full of real possibilities which are unconstrained by what is or has been the case. Marcel distinguishes between an objective judgment based on the facts, such as the judgment that there is a chance that I will see my son again, and the simple affirmation, ‘You are coming back’, which is expressed in hope (1945: 66). The former is constrained by the evidence and by what can legitimately be inferred from this, whereas the latter is not so constrained. Such hope is not based upon a calculation, or miscalculation, of probabilities, and hence is not governed by the norms of such calculation. It is rather a refusal to calculate, a refusal to regard the future as wholly delimited by the past and the present. Thus, when hope takes this form, it is beyond objective criticism. It is, however, only beyond objective criticism in so far as it is an expression of love, rather than desire (1945: 66).

See also: Virtues and vices

PHILIP STRATTON-LAKE

References and further reading

Augustine (421) *Enchiridion*, trans. M. Dods, in *The Works of Aurelius Augustinus*, Edinburgh: T. & T. Clark, 1871-6; this translation repr. in V.J. Bourke (ed.) *The Essential Augustine*, New York: New American Library, 1964, 173.(A handbook, or manual, covering the manner in which God is to be worshipped; among the last of Augustine’s works.)

Bloch, E. (1954-9) *Das Prinzip Hoffnung*, Frankfurt: Suhrkamp Verlag, 2 vols; trans. N. Plaice, S. Plaice and P. Knight, *The Principle of Hope*, vol. 3: Oxford: Blackwell, 1986.(Bloch’s major, encyclopedic work expressing his views about hope.)

Buhr, M. (1970) ‘A Critique of Ernst Bloch’s Philosophy of Hope’, *Philosophy Today* 14: 259-71.(Example of the orthodox Marxist response to Bloch’s philosophy.)

Day, J.P. (1969) ‘Hope’, *American Philosophical Quarterly* 6: 89-102.(Clear analysis of goal-directed hope.)

Descartes (1649) *La passions de l’âme*, trans. J. Cottingham, R. Stoothoff and D. Murdoch, *The Passions of the Soul*, in *The Philosophical Writings of Descartes*, Cambridge: Cambridge University Press, 1985.(This work grew out of the correspondence with Princess Elizabeth of Bohemia and addresses the question of how mind and body can interact given that they share nothing in common.)

Godfrey, J.J. (1987) *A Philosophy of Human Hope*, The Hague: Martinus Nijhoff. (An excellent account of different types of hope, and an overview of its role in the writings of Kant, Bloch and Marcel. It is better on Bloch and Marcel than on Kant.)

Godfrey, J.J. (1987) ‘Appraising Marcel on Hope’, *Philosophy Today* 31: 234-40. (Economical introduction to Marcel on hope.)

Hobbes, T. (1651) *Leviathan*, Harmondsworth: Penguin, 1982.(Hobbes’ major work on political philosophy, in which he argues for the absolute power of the sovereign.)

Kant, I. (1788) *Critik der practischen Vernunft*, trans. L.W. Beck, *Critique of Practical Reason*, New York: Macmillan, 1985, book II, ch. 2, sections IV, V and VI.(Kant’s second major work on ethics, containing his doctrine that there are three postulates of practical reason: hope that God exists, hope for immortality and hope for the attainment of perfect freedom.)

Kant, I. (1793) *Die Religion innerhalb der Grenzen der blossen Vernunft*, trans. T.M. Greene and H.H. Hudson, *Religion Within the Limits of Reason Alone*, New York: Harper & Row, 1960.(A late work by Kant which offers a moral interpretation of religion and further develops the role of hope in his philosophy.)

Kierkegaard, S. (1844) *Philosophiske Smuler*, trans. H.V. Hong and E.H. Hong, *Philosophical Fragments*,

Princeton, NJ: Princeton University Press, 1985.(An important work by Kierkegaard in which he argues against any attempt to arrive at faith by means of argument, or objective thinking.)

Leibniz, G.W. (1710) *Essai de Theodicée sur la bonté de Dieu, la liberté de l'homme, et l'origine du mal*, trans. E.M. Huggard, *Theodicy: Essays on the Goodness of God, the Freedom of Man, and the Origin of Evil*, London: Routledge & Kegan Paul, 1952.(An attempt to reconcile God's perfection with the fact that there is evil in the world.)

Marcel, G. (1945) 'Sketch of a Phenomenology and Metaphysic of Hope', in *Homo Viator: prolégomènes à une métaphysique de l'espérance*, Paris: Aubier; trans. E. Crauford, *Homo Viator: Introduction to a Metaphysic of Hope*, New York: Harper & Row, 3rd edn, 1965.(Marcel's most comprehensive statement of his views on hope and its relation to faith and love.)

McCarthy, T. (1986) 'Philosophical Foundations of Political Theology', in L. Rouner (ed.) *Civil Religion and Political Theology*, London: Notre Dame, 23-40.(Makes interesting connections between Kant's defence of the three postulates from Wizenmann's criticism and debates within critical theory.)

Moltmann, J. (1964), *Theologie der Hoffnung*, Munich: Chr. Kaiser Verlag; trans. J.W. Leitch, *Theology of Hope*, London: SCM Press, 1967.(Hard reading for the uninitiated. Good introduction.)

Peukert, H. (1984) *Science, Action and Fundamental Theology: Toward a Theology of Communicative Action*, Cambridge, MA: MIT Press.(Develops a theology based on Jürgen Habermas' critical theory, in which it is argued that that the utopian element in Habermas' thought commits one to a hope for the redemption of the victims of history.)

Philosophy Today 14 (1970).(This volume has a number of clear, short, introductory articles on Bloch's philosophy of Hope.)

Ricoeur, P. (1970) 'Hope and the Structure of Philosophical Systems', *Proceedings of the American Catholic Philosophical Association* 44: 55-69.(Interesting and lucid use of Moltmann's *Theology of Hope* to relate philosophy and theology.)

Spinoza, B. (1677) *Ethica Ordine Geometrico Demonstrata* (Ethics Demonstrated in a Geometrical Manner), trans. E. Curley, *Ethics*, Harmondsworth: Penguin, 1996. (Spinoza's philosophical system presented according to the geometrical method covering God, the mind, the affects, human subjection and human freedom.)

Voltaire, F.M.A. de (1759) *Candide*, trans. D.M. Frame, in *Candide, Zadig and Selected Stories*, New York: New American Library, 1961.(Voltaire's famous satire of Leibniz's optimism.)

Horkheimer, Max (1895-1973)

One of the initiators and founders of the Institute for Social Research in Frankfurt, Germany, Max Horkheimer's philosophical importance derives from his programmatic essays of the 1930s in which he conceptualized the Institute's project of interdisciplinary research, and his later collaboration with Theodor W. Adorno in writing Dialectic of Enlightenment (1947). Horkheimer's vision of a 'critical theory' of society was intended as a reformulation of Marxist theory in which empirical research would be combined with philosophical reflection.

Born near Stuttgart, Max Horkheimer was the only son of a wealthy Jewish family. He studied philosophy at the then new Frankfurt University, writing both his doctoral thesis and *Habilitationsschrift* on Kant's *Critique of Judgment*. He became the director of the newly founded Institute for Social Research in Frankfurt in 1930, and organized the Institute's move first to Geneva and then New York City in 1934. Throughout the 1930s he edited the Institute's innovative journal *Zeitschrift für Sozialforschung (Journal for social research)*, in which appeared important and characteristic essays by Theodor Adorno, Herbert Marcuse and Walter Benjamin. Horkheimer re-established the Institute in Frankfurt after the war, and served as rector of the Frankfurt University from 1951 to 1953. After retiring, he moved to Italy.

Beneath the surface, and in all periods of his work, Horkheimer's philosophy was shaped by the attempt to align Karl Marx's materialist critique of bourgeois society with Arthur Schopenauer's pessimistic vision of human suffering in the midst of a meaningless universe (see Marx, K.; Schopenhauer, A.). Human suffering, the experience of which is the ground of creaturely solidarity, cannot for Horkheimer be transfigured or redeemed; even a future utopian society would carry the curse of the countless victims of history. While compassion for the suffering of other living beings and individuals' longing for happiness form the ethical core of Horkheimer's thought, his most distinctive contribution to philosophy emerges in his attempt to formulate a new Marxist theory of society which would overcome what he saw as the dogmatism and failures of the original.

In his essay 'Traditional and Critical Theory' (1937) Horkheimer attempts to elaborate how the interdisciplinary research programme of the Institute is to be separated from the then standard paradigm of scientific knowledge as a theoretical representation of a wholly independent object domain. What traditional theory overlooks is that both the activity of knowing itself and its (social) objects are historically formed. Because the categories of human existence are historically formed, philosophy needs to be supplemented by social scientific research. Because those same social categories are the bearers of human significance, they require philosophical elucidation. Hence, only a complex interdisciplinary programme integrating philosophy and the social sciences (sociology, history, economics, social psychology and psychoanalysis) can do justice to the complex problems facing contemporary society. While both the actual programme of interdisciplinary research and the hopes for it had faded for Horkheimer by the end of the 1930s, the project was re-launched by Jürgen Habermas, the leading second-generation exponent of critical theory.

For Horkheimer, what makes critical theory 'critical' is its combining of the Marxist 'critique of political economy' with the Kantian 'critique of pure reason'. The importance of the latter for Horkheimer should not be underestimated. In criticizing traditional theory, Horkheimer is assuming the Kantian, 'critical' distinction between the causal understanding of theoretical reason (*Verstand*) and the ends-oriented, holistic thinking of practical reason (*Vernunft*) (see Kant, I. §6). In Horkheimer's lexicon the former appears as 'subjective' or 'instrumental' reason, the latter as 'objective' reason. While in his original programme for critical theory Horkheimer intended the whole complex interdisciplinary practice to be a surrogate for objective reason, with the waning of any hope for radical social change his philosophy became a relentless critique of instrumental reason.

Written with Adorno, *Dialectic of Enlightenment* (1947) was intended to provide a theoretical and historical generalization of the critique of enlightenment found in Hegel's *Phenomenology of Spirit* (see Hegel, G.W.F. §5). The core claim of the *Dialectic* is that myth is already a form of enlightened, instrumental reason - an attempt to control hostile nature - and that the enlightenment critique of myth and superstition reverts to myth. By this latter claim Adorno and Horkheimer mean that the wholly disenchanted modern world is itself under a mythic curse or spell, in which individuals become anonymous ciphers within a wholly functional system. Instrumental rationality pervades the world - as the rationality of capitalist economy right through to the 'culture industry', and as the

reason of a positivistically understood science. Because instrumental rationality is only a part of reason pretending to be the whole of reason, it becomes irrational. It is this thought that lies behind the exorbitant claim that the whole (of society) is untrue (see Adorno, T.W. §3)

In *Eclipse of Reason* (1947), originally presented as lectures at Columbia University in 1944, Horkheimer presented a reprise of the central arguments of *Dialectic of Enlightenment* in a slightly simpler form. Only now is the insistent negativity of his collaborative work softened; throughout history, he claims, myth, religion and philosophy have presented versions of objective reason that have only slowly and recently been 'eclipsed' by subjective reason. While independent thought cannot renew the worn-out ideals of objective reason, it can find equivalents in the affinity of nature and human nature. Alas, these equivalents, given through art, cannot provide the pivot Horkheimer requires for a renewal of reason. At the end Horkheimer was left with only compassion and pity for suffering life, and the longing for sensuous happiness, that had been throughout the mainsprings of his thought.

See also: Critical theory; Frankfurt School

J.M. BERNSTEIN

List of works

Horkheimer, M. (1930-8) *Between Philosophy and Social Science: Selected Early Essays*, trans. G. F. Hunter, M.S. Kramer and J. Torpey, London: MIT Press, 1993. (This selection includes most of Horkheimer's most important essays from the 1930s on moral philosophy, the philosophy of history, and methodology.)

Horkheimer, M. (1932-41) *Critical Theory: Selected Essays*, trans. M.J. O'Connell *et al.*, New York: Seabury Press, 1972.(Some overlap with the above, but importantly includes 'Traditional and Critical Theory' and 'Authority and the Family'.)

Adorno, T.W. and Horkheimer, M. (1947) *Dialectic of Enlightenment*, trans. J. Cumming, London: Allen Lane, New York: Herder & Herder, 1972.(The classic text of first generation Critical Theory, arguing that Enlightenment rationalisation collapses into pure instrumental rationality and a new form of mythology.)

Horkheimer, M. (1947) *Eclipse of Reason*, New York: Continuum, 1992.(Originally given as lectures at Columbia University, a somewhat more accessible version of the previous item, with a more fully articulated conception of objective, subjective and instrumental reason.)

References and further reading

Benhabib, S., Bonß,W. and McCole, J. (eds) (1993) *On Max Horkheimer: New Perspectives*, London: MIT Press.(An excellent collection of fifteen essays covering in detail the most important aspects of Horkheimer's thought. All are worth reading.)

Wiggerhaus, R. (1986) *The Frankfurt School*, trans. M. Robertson, Oxford: Polity Press, Cambridge, MA: MIT Press, 1994.(An authoritative history of the Frankfurt School, which provides a useful account of Horkheimer's activities in relation to the Institute for Social Research.)

Huainanzi (179-122 BC)

Huainanzi is both the honorific name of Liu An, the second king of Huainan and the title of the philosophical work for which he was responsible. The most important surviving text of the academy he established at his court, it consists of twenty-one essays that form a compendium of knowledge the Daoist ruler needs to govern effectively. In this work, the universe is a well-ordered, dynamic and interrelated whole, interfused by the unifying principle of the dao, *that develops according to patterns and processes comprehensible to self-realized human beings. The ruler must cultivate himself fully so that he comprehends these patterns and processes and must establish human society in harmony with them. Embracing the best ideas of earlier philosophers within a Daoist framework, the Huainanzi represents the fullest flowering of the Huang-Lao thought that dominated the early Han dynasty.*

Huainanzi is the honorific name sometimes given to Liu An, a Chinese philosopher-king who was the grandson of Liu Bang, founder of the Han dynasty (206 BC-AD 220). After his father, Liu Chang, was punished for *lèse majesté* and died on the way to exile, Liu An was enfeoffed as the second king of Huainan, a vassal state in what was then considered the southern part of China, today mostly in Anhui province. This area, previously part of the old state of Chu, had long been associated with hermit-mystics and shamans and still retained in the early Han a culture distinct from that of the north, where the new central government resided. By about 150 BC, Liu had established an academy at his court wherein Daoist and Confucian thinkers mingled with scholars of the esoteric arts (*fangshi*), proponents of the 'naturalist' cosmology of *yin* and *yang* and the five phases of *qi*, systematized a century earlier by Zou Yan (see Yin-yang; Qi). Among the many literary products of Liu's academy, the principal survivor is a text called the *Huainanzi*, which Liu presented to his nephew, Emperor Wu, in 139 BC. In 122 BC Liu was accused of sedition and he perished, along with many of his courtiers and his rival cultural centre, at the hands of imperial troops sent to quell his alleged rebellion.

The *Huainanzi* is a work of twenty-one essays on a wide range of topics that was intended to be a compendium of all the knowledge a Daoist sovereign needed to govern effectively. Its topics include cosmology and cosmogony, astronomy and astrology, seasonal ordinances, human nature, psychology and physiology, political theory and the theory of strategic warfare, and mystical self-transformation. Although a few scholars have maintained single authorship by Liu, the predominant opinion is that Liu edited, compiled and arranged the essays and wrote the final essay, which summarizes each of the others and rationalizes their arrangement in the text.

Although initially categorized by bibliographers as an eclectic work because of its selective use of the non-Daoist ideas of such earlier schools as the Confucians and Legalists, the *Huainanzi* is increasingly being recognized as a product of the Daoist syncretism of the early Han that many identify with the Huang-Lao lineage of thought. 'Huang' stands for Huangdi, the mythical Yellow Emperor, and 'Lao' stands for Laozi, the legendary author of the foundational Daoist text, the *Daodejing* (see Daodejing). In light of the mythical connotations of each figure and the ideas associated with this lineage, in the appellation 'Huang-Lao', Huangdi probably symbolizes the coordination of the political with the cosmological and Laozi probably symbolizes the mystical self-realization of the enlightened ruler. Taken together, we find the advocacy of a strong central government whose institutions, daily rituals, and organization of human society parallel greater patterns in the universe of which human beings form an integral part. This government is presided over by a benevolent ruler who is both philosopher and mystic and who is fully self-realized according to Daoist meditation techniques. During the first seven decades of the Han dynasty, the imperial court was dominated by Huang-Lao philosophy, a domination that was coming to an end when Liu An presented his *Huainanzi* to Emperor Wu. As a last-ditch attempt to convince the emperor to abandon the Confucian ideology he eventually chose, the *Huainanzi* represents the final flourishing of the Huang-Lao syncretism that adapted the best ideas of earlier thinkers into a Daoist philosophical framework.

In the cosmology of the *Huainanzi*, the universe is a well-ordered, intricately interrelated and constantly changing dynamic whole, governed by regular patterns and processes (*li*) that can be understood by exceptional human beings (see Li). Precise parallelisms exist between the basic levels of this universe. For example, the roundness of the human head parallels that of heaven, the squareness of the human feet parallels that of earth, heaven's four seasons parallel the four human limbs and so on. All phenomena are constituted of varying densities of *qi*, the vital energy/matter, and each has an inherent nature that determines its physical form and provides the spontaneous

tendencies for growth, development and response that arise during its lifetime (see Qi). Because each thing has a characteristic type of *qi*, one of its 'five phases', it resonates (*ganying*) with other things of the same phase and is thus impelled to develop or act in certain ways when acted upon by other things; when the seasons change, as the planetary bodies revolve and so on. All phenomena, processes and events within the universe on both macrocosmic and microcosmic levels are included in this grand 'correlative' scheme. While all act spontaneously, responses are never random and chaotic; they occur according to the regular and predictable patterns that govern the cosmos. Finally, the entire cosmos is interfused by a singular dynamic force called the *dao* (Way), which is inherent in each phenomenon as its ultimate unifying and guiding principle (see Dao).

Because human beings and their societies are integral parts of this dynamic, interrelated whole, they are subject to its same governing principles. Hence the ruler must know these principles and how the various levels parallel and interact with one another, and must set up governmental and societal institutions to accord with them. Furthermore, after establishing a hierarchical bureaucracy filled by officials whose deeds correspond to their titles, the ruler must cultivate a profound tranquility through meditation that enables him to become unified with the *dao*. Emerging from this mystical unity, the ruler has developed a totally unbiased clarity of mind that enables him to 'take no action (*wuwei*) and leave nothing undone'. With his mental acuity and metaphysical insight into cosmic patterns and inherent natures, the ruler can develop suitable tasks and appropriate customs for his subjects that enable them to realize the fullest potential of their inherent natures, transform his subjects through a mysterious resonance of *qi*, and more generally govern in a benevolent and unbiased manner that brings order and harmony throughout the land. This is how things were in the distant past of the sage-rulers, but the world lost this harmony because people lost the understanding of the value of - and the ability to effect - this profound coordination of the cosmological with the political. The *Huainanzi* was written to provide the ruler with the means to restore this harmony.

See also: Chinese philosophy; Confucian philosophy, Chinese; Daoist philosophy; Guanzi; Legalist philosophy, Chinese; Li; Machiavelli, N.; Political philosophy, History of; Qi

H.D. ROTH

List of works

Liu An [Huainanzi] (179-122 BC) *Huainanzi*, Sibu congkan edition, Shanghai: The Commercial Press, 1920; ed. Haruki Kusuyama, *Enanji*, Shunyaku kambun taikei edition, Tokyo: Meiji Shoin, 1979-88, 3 vols; ed. D.C. Lau and Chen Fong Ching in *A Concordance to the Huainanzi*, The Institute of Chinese Studies Ancient Chinese Texts Concordance Series, Hong Kong: The Commercial Press, 1992.(There are over eighty-five complete and thirty partial editions of the *Huainanzi* in Chinese and Japanese; details are given in Roth (1992). Of the three mentioned here, the Sibu congkan edition is the best of the traditional Chinese editions and is ultimately based on an eleventh-century printed edition of the text. Kusuyama contains an excellent critical Chinese edition together with a modern Japanese translation. Lau and Chen provide a computer-coordinated concordance of all the words in the text as well as one of the best new critical editions. English translations of portions of the Huainanzi are to be found in various works; see Ames 1983, Larre, Robinet and Rochat 1993, LeBlanc 1985, Major 1993, Wallacker 1962.)

References and further reading

Ames, R. (1983) *The Art of Rulership: A Study in Ancient Chinese Political Thought*, Honolulu, HI: University of Hawaii Press.(A careful and readable translation of *Huainanzi* 9, the essay on political theory, that contains a valuable analysis of the philosophical background of its key ideas.)

Graham, A.C. (1986) *Yin-Yang and the Nature of Correlative Thinking*, IEAP Occasional Paper and Monograph Series 6, Singapore: Institute for East Asian Philosophies.(This is a comprehensive and sophisticated analysis of the correlative thinking that produced the cosmology of *yin* and *yang* and the five phases of *qi* that is so important to the *Huainanzi*.)

Larre, C., Robinet, I. and Rochat de la Vallée, E. (1993) *Les grand traités du Huainan zi*, Paris: Les Éditions du Cerf. (Contains translations of the *Huainanzi* chapters 1, 7, 11, 13 and 18.)

LeBlanc, C. (1985) *Huai-Nan Tzu: Philosophical Synthesis in Early Han Thought*, Hong Kong: Hong Kong University Press.(A heavily annotated translation of *Huainanzi* 6, 'Peering Into the Mysterious', that contains an excellent discussion of the concept of resonance).

Major, J.S. (1993) *Heaven and Earth in Early Han Thought*, Albany, NY: State University of New York Press.

(This is an annotated translation of the three most technical essays in the *Huainanzi*, 3, 4 and 5, on the topics of astronomy, geography and seasonal ordinances, accompanied by excellent introductory chapters on the philosophy of the *Huainanzi* in the context of Huang-Lao thought and on early Han cosmology.)

Roth, H.D. (1991) 'Psychology and Self-Cultivation in Early Taoistic Thought', *Harvard Journal of Asiatic Studies* 51 (2): 599-650.(An analysis of the *Huainanzi*'s theories on self-cultivation and their background.)

Roth, H.D. (1992) *The Textual History of the Huai-nan Tzu*, Ann Arbor, MI: Association for Asian Studies Monograph 46.(A comprehensive study of the history and transmission of the text of the *Huainanzi* that includes an introduction to its thought and the circumstances of its creation.)

Wallacker, B. (1962) *Huai-nan Tzu, Book Eleven: Behavior, Culture, and Cosmos*, New Haven, CT: The American Oriental Society.(This is a translation of the *Huainanzi* essay on social customs and human nature with a brief introduction.)

Huet, Pierre-Daniel (1630-1721)

Huet was a French Catholic bishop who wrote important works in theology, philology and literary criticism. In philosophy, he defended an apologetic interpretation of scepticism and opposed Cartesianism, which he thought to be a fanatical form of rationalism. For Huet, faith must guide reason since only the former, received from God, can provide absolute certainty, whereas human knowledge is inevitably fallible. A moderate scepticism is therefore the most appropriate attitude for a philosopher, since it tempers the ambitious claims of reason and extricates the mind from prejudices and false certainties, thus preparing it to receive the divine gift of faith.

1 Life

Pierre-Daniel Huet was considered one of the most learned men of his time. Born in Normandy, he was educated by the Jesuits and acquired a remarkable knowledge of Latin, Greek and mathematics while still very young. In Paris, he studied Hebrew and became deeply interested in the philosophy of Descartes. In 1652, he travelled with his friend Samuel Bochart to Holland and Sweden, where the latter had been invited by Queen Christina, and had the opportunity to meet Henricus Regius, Claude Saumaise and Isaac Vossius. After his return to Paris, he became acquainted with Gassendi, and in 1661 he founded a scientific academy in Caen which is still active. At the age of forty, Huet was elected tutor of the Dauphin Louis, son of Louis XIV, under Jacques Bossuet's direction. He began to edit the famous series of Classics *ad usum Delphini* and became member of the *Académie Française* in 1674. Ordained two years later, he was bishop of Avranches from 1689 when he published the *Censura philosophiae cartesianae*, a meticulous critique of Descartes' philosophy. In 1690, he published the *Alnetanae quaestiones de concordia rationis et fidei*, a theological treatise in which he argues that no truth or ultimate certainty can be reached without the help of faith. This was preparatory to the most famous of Huet's philosophical books, the *Traité philosophique de la faiblesse de l'esprit humain* (1723), which he seems to have considered his best work. Huet died in Paris, and his collection of more than 8,000 volumes, many bearing his annotations, later became part of the Bibliothèque Nationale. His anti-Cartesianism attracted much attention and, soon after his work was published, defences of Cartesianism were written against him by Andreas Petermann, Pierre-Sylvain Régis, Joannes Schotanus, Burchard de Volder and, in 1745, by Lodovico Antonio Muratori.

2 Anti-Cartesianism

Huet treats Cartesianism as merely another philosophical school attempting to restore the old mistakes of dogmatism. The importance Descartes ascribes to the sceptical approach is highly valuable, but his positive principles are all erroneous. First, the firmness of the 'Ego cogito, ergo sum' is deceptive for a number of reasons:

1 It is a mere tautology - either the 'ego' in the expression 'ego cogito' is meaningless, or it refers to something that already exists. But then the whole statement says nothing more than 'I exist and I am something that thinks; whatever exists and is something that thinks, exists; therefore I exist', and this is of course a true, but non-informative tautology.
2 It is not logically prior to, or more simple than, any other form of knowledge - it requires, at least, that one understands the principle 'whatever thinks, exists', and may be capable of using it as the premise of the relevant syllogism.
3 Its acceptance is inconsistent with the strategy of radical and systematic doubt - if Descartes can use the natural light to grasp the truth that whatever thinks exists, then why can the natural light not be employed to affirm other principles equally obvious, as that the whole is bigger than the part?
4 It is not immediate - the *Cogito* is a syllogism, and hence to be formulated requires a process of successive inferences, the correct application of logical principles and a reliable memory.

Second, the criterion of truth is established only via a vicious circle (see Descartes, R. §§5-7). We can trust that our clear and distinct ideas lead from the certainty of a belief to its truth if and only if there is a veridical God, but we can prove as true the existence of a veridical God only by relying on the certainty of our clear and distinct ideas.

Third, it is false that we know the mind before, and better than, the body, since the two are united, and Descartes' intellectualism (we are our minds, not our bodies) is wrong because mind and body are indivisible. The body, and

not the mind, perceives the physical objects, and the mind without the perceptual experience of the body would be empty, since we have no innate ideas, but only particular ideas made more general by abstraction.

Fourth, God's existence cannot be proved by reason. We possess only a finite and imperfect collection of ideas about God's nature, acquired via a negative procedure of amplification of what we know to be positive and exclusion of what we know to be negative in the world. We cannot rise from this inadequate description, whose origin is thoroughly dependent on our minds, to the necessary existence of its reference, with the certainty of a logical proof.

3 Scepticism

The Cartesian edifice therefore collapses and can be replaced by a moderate form of scepticism, to which Huet ascribes an apologetic function. The process of knowing consists of four stages: the Understanding receives perceptual impressions from physical objects, transforms them into image-like ideas, compares ideas and objects, and evaluates their correspondence. We know the truth when we know both the idea and that the idea perfectly corresponds to the object. Thus the level of certainty of our knowledge depends on the level of reliability of each of the previous stages, which Huet claims to be rather low, again for several reasons:

1 Since there is a gap between a physical object, as the source of knowledge, and its idea, as a mental product, knowledge of things is only indirect. Yet (a) everything changes constantly, including the known object, the impressions, and the knowing Understanding; (b) the senses are unreliable; (c) the Understanding is fallible; and (d) the Understanding modifies the impressions coming from the senses (Huet himself anatomized some animals, concluding that the nerve-channels are such that they must modify the impressions when they run through them.) Therefore, it is impossible to verify that internal ideas really correspond to external objects, and any description of the physical world remains at most verisimilar, but is never utterly certain.
2 It is impossible to demonstrate, by mere reflection, that the Understanding is capable of certainties, without falling into a *petitio principii*; but we can analyse the Understanding only by relying on the Understanding itself, thus falling into a vicious circle.
3 Knowledge of something *per genus et differentiam* (for example, to know man as the rational animal) is impossible. We know the genus of *x* if and only if we know both its essence and how it is similar to *y*, where *y* belongs to a different species but the same genus (for example, man and horse are two species belonging to the same genus, animal). Therefore, in order to provide the definition of *x* it is necessary to begin by knowing the essence of both *x* and *y*, but for this purpose it is necessary to know the genus of both, thus falling into a vicious circle. The same applies to the use of the notion of difference: in order to establish that *x* is different from *y*, we must first know both *x* and *y*, but such knowledge is based on the possibility of applying the notion of difference in order to have a definition of both.
4 Nobody can be certain that everybody else has the same impressions, or the same ideas on the same things. Moreover, the very notion of evidence is relative. People in different states (ill, drunk, sleeping, hungry, worried and so forth) perceive the world in different and irreconcilable ways. Perhaps Descartes was right: we do not know whether God - whose omnipotence is interpreted as boundless - has wished to create us so imperfect as to err always. What seems clear is that there is total disagreement about the nature of knowledge and truth even among the Dogmatists.
5 In order to know *x* adequately, we must also know the causes of *x*, but this leads to an infinite regress, for everything depends on something else, and the Understanding is incapable of omniscience.
6 We have no objective and uncontroversial criterion of truth. Huet distinguishes three meanings of criterion: the *criterium duquel* (the knowing subject); the *criterium par lequel* (the senses and the Understanding); and the *criterium selon lequel* (the action of the Understanding, when it applies the *criteria par lequel* in the search of truth.) But we have seen that human nature is unreliable, the senses and the Understanding are not sources of certainty and, with respect to the *criterium selon lequel*, if the faculties are unreliable then so is their action. The problem is that in order to find a proper criterion of truth we must already possess one.
7 Reasonings are always uncertain, and even simple syllogisms beg the question. For example, from the two premises 'all men are rational animals' and 'Peter is a man' we infer that 'Peter is a rational animal', but the first premise depends on inductive reasoning - every man examined so far has been a rational animal. So the inference is hypothetically certain (given the premise the conclusion follows necessarily), but the knowledge so achieved is still questionable, for God could create a man who is not rational. Syllogisms remain the best weapons of the

dogmatists, but they become harmless if one rejects their premises.

From all these arguments, Huet concludes that all forms of *human certainty* - in descending order, the certainty provided by God's words, mathematical certainty, perceptual certainty, and finally moral certainty (the certainty we have about what we have not experienced personally) - are inevitably bound within the limits of our capacities, and relative to our subjective condition. The Understanding can formulate plausible beliefs, but it necessarily remains unknown whether they represent true knowledge. It is possible to achieve total certainty only when there is direct and complete access to the intrinsic nature of things, but this happens only to the blessed souls who, illuminated by God, know with divine certainty and reassurance what, in the best case, we can believe only with limited confidence. While awaiting the acquisition of such a spiritual state, a probabilistic form of scepticism remains the most desirable attitude for a human being. As the origin of all philosophy, scepticism teaches us how to avoid epistemic errors, makes us more critical and independent of intellectual authorities, helps us to accept from every past philosopher whatever looks more acceptable and probable (eclecticism) and, above all, prepares the mind for the divine gift of faith.

Huet is aware of the classic arguments against scepticism. To the philosophers (Plato, Proclus and Descartes) who defend the presence of innate knowledge in the Understanding, he replies that they forget that nothing is in the mind which was not previously in the senses (see Plato §13; Proclus). Universals are acquired from particular ideas: sometimes one particular idea is elected as a general prototype (one particular triangle stands for all triangles), sometimes a general idea is the result of a process of abstraction from a collection of particular ideas (as happens with the idea of movement), but in any case we have no innate knowledge and no ultimate certainties. It is often argued, on the ethical side, that scepticism makes life impossible, is irreligious - because it suggests that God is deceitful, since he makes us incapable of knowledge - and undermines the possibility of developing any sort of belief, including faith in God, thus advancing the moral corruption of man. Huet rejoins that scepticism inclines us to avoid fanaticism and to follow the customs and habits of our time, without fully committing ourselves to anything but a reasonable trust in Christian religion. God has created us limited and weak, but free to give or withhold our assent to opinions. It is only our fault if we give precipitous assent to superficial opinions, and hence err. Scepticism does not lead to atheism. On the contrary, the recognition of the limits of our Understanding helps us to acquire faith. Since knowledge of God is not discursive, acquired by the Understanding with a spontaneous act, but intuitive and given by God to the soul as a free and generous gift, it is not epistemic and hence remains unaffected by sceptical doubts. This explains why, according to Huet, Christian religion appreciates scepticism, although it must be noted that most of the passages he quotes - from Wisdom, Ecclesiastes, Paul, Augustine, Aquinas, and many others - were meant to undermine intellectual pride, not human confidence in ordinary knowledge. Scepticism can be criticized on the epistemological side because it undermines all sciences, the greatest human achievement; because it either begs the question - since it presupposes a criterion to distinguish what is certain from what is only probable, and what is true from what is false, so, after all, the sceptic knows something - or it is self-contradictory, for it denies all knowledge and discredits the validity of all possible reasonings, including its own; and finally, because it does not lead to a systematic philosophy. Huet retorts that faith, not science, is the greatest human achievement, and the sceptic is not to be blamed because he reveals the embarrassing weakness of the Understanding. The dogmatist is hopeless, for he seeks truth through science, whereas the sceptic is happy with his probable opinions, makes no mistakes and is open to the religious message. Indeed, many people have become dogmatists rather than sceptics only because they are lazy, wishing to enjoy their ignorance without searching further. Descartes is one of them. He is ambitious and - although he has taken most of his ideas from ancient philosophers - he has ostracized erudition, cultural tradition and historical knowledge, thus removing the only basis for the collaborative construction of a body of verisimilar opinions. Scepticism is not self-contradictory either, because it eliminates even its own certainties, and follows only the most likely appearances. The sceptic is like Samson, who destroys the temple of dogmatic presuppositions and himself. Finally, scepticism does not give rise to an ordinary sect, but to the 'philosophy of not philosophizing', as Lactantius called it, which can nevertheless become fairly systematic, as Sextus Empiricus showed.

See also: Pascal, B.; Pyrrhonism; Scepticism, Renaissance

LUCIANO FLORIDI

List of works

Huet, P.D. (1661) *De interpretatione libri duo: quorum prior est, de optimo genere interpretandi, alter, de claris interpretibus*, Paris: repr. The Hague, 1683.(Huet defends the compatibility of stylistic ornamentation with the need to communicate, but warning that rhetorical ornaments can be like weeds, which spread and suffocate the good crop, and that conceptual content should have priority over literary form. The second part analyses the works of various translators of the Bible and of classic authors.)

Huet, P.D. (1668) *Origenis in sacram scripturam commentaria*, Rouen.(Edition in 2 volumes of the fragments of Origen's *Commentary on St. Matthew*, discovered by Huet in Stockholm, see J. P. Migne, *Patrologiae cursus completus*, Series Graeca, tom. 17.)

Huet, P.D. (1669) *Traité de l'origine des romans*, Paris; repr. Stuttgart: Metzler, 1966 (first edition, 1670), and Geneve: Slatkine, 1970 (Paris edition, 1799).(One of the first studies of the origin of prose romances. Huet traces the history of romance, providing copious references to earlier romance writers and their works. He defends romance from charges of mere entertainment and immoral temptation. The *Traité* was initially prefixed to the first edition of *Zaïde, histoire espagnole*, of Madame de La Fayette, as a *Lettre de Monsieur Huet a Monsieur de Segrais de l'origins des romans*. The work became a best-seller, was re-edited may times, and translated into several languages, including English, as *A treatise of romances*, London: 1672; *The history of romances*, London: 1715; and *An extract of Monsieur Huet's discours concerning the original of romances*, London: 1729. There have been several modern editions, including *The History of Romances, an annotated text edited, with introduction*, by Robert Clark Malcolm, University of Michigan, Ph.D. Thesis: 1983.)

Huet, P.D. (1679) *Demonstratio evangelica*, Paris; 2nd edn, 1680.(An attempt to prove the principles of Christian religion by means of an axiomatic apparatus. 'Praefatio' section 3, and 'Axiomata' IV sections 2 and 3 contain an interesting, sceptical philosophy of mathematics. The revised 1680 edition is more important.)

Huet, P.D. (1689) *Censura philosophiae cartesianae*, Paris; repr. Hildesheim, New York: G. Olms, 1971.(The sceptical reply to Descartes' *Meditations*. The first edition was printed by Kampen in 1690, but the work was printed and re-edited several times. The fourth revised and augmented edition of 1694 is the most important.)

Huet, P.D. (1691) *Traité de la situation du Paradis terrestre*, Paris.(Paradise is placed at the intersection of the Tigris and the Euphrates. The work was translated in several languages and edited many times. For an English translation by T. Gale see *A treatise of the situation of Paradise*, London: 1694.)

Huet, P.D. (1690) *Alnetanae quaestiones de concordia rationis et fidei*, Paris; repr. Leipzig, 1692.(This work, which takes its name from Alnay, Huet's abbey, is a bridge from Huet's anti-Cartesianism as a form of dogmatic and impious rationalism, to his moderate scepticism, in which the process of systematic doubt becomes preparatory to the acquisition of faith.)

Huet, P.D. (1692) *Nouveaux mémoires, pour servir à l'histoire du cartésianisme*, Paris; important 3rd edn, Paris: R. Mazieres, 1711.(Published under the name of Gilles de l'Aunay, this is a satire in which Huet presents a caricature of Descartes and derides, rather wittily but superficially, his philosophical positions.)

Huet, P.D. (1718) *Commentarius de rebus ad eum pertinentibus*, Amsterdam; trans. as *Memoirs of the life of Peter Daniel Huet, bishop of Avranches, written by himself*, London: Longman, Hurst, Rees, & Orme, and Cadell & Davies, 2 vols, 1810.(Huet's autobiography.)

Huet, P.D. (1723) *Traité philosophique de la faiblesse de l'esprit humain*, Amsterdam; repr. Hildesheim, New York: G. Olms, 1974; trans. as *An essay concerning the weakness of the Human Understanding*, London, 1725; and as *A philosophical treatise concerning human understanding*, London, 1729.(Written in 1690 but published posthumously, this combines ancient, sceptical arguments with the anti-intellectualist perspective of the Pauline tradition and the Neoplatonist distrust of empirical knowledge, to attack Cartesianism as the latest form of dogmatism.)

References and further reading

Alberti, A. (1978) 'Lo Scetticismo Apologetico di Pierre Daniel Huet', *Giornale Critico della Filosofia Italiana* 9: 210-37.(Suggests, but does not develop, the thesis that the *Censura*, the *Alnetanae Questiones* and the *Traité* are three parts of a unique anti-Cartesian and apologetic project. The comparison of Huet's position to Mersenne's gives too little space to the former.)

Aubry, D. (1944) *Daniel Huet*, Paris: La Bonne Presse.(An odd companion, in which Huet's remarks on his life, studies and ideas are selected and reorganized alphabetically and by topic.)

Brown, H. (1934) *Scientific Organization in 17th century France*, Baltimore, MD: Russell & Russell, repr. 1967. (Pages 216-27 discuss Huet's scientific interests.)

Cabeen, D.C. and Brody, J. (eds) (1961) *A Critical Bibliography of French Literature*, vol. 3, *The Seventeenth Century*, ed. N. Edelman, Syracuse, NY: Syracuse University Press.(Some of the most important works published on Huet before 1960 are listed with comments.)

Dini, A. (1987) ‘Anticartesianesimo e apologetica in Pierre-Daniel Huet’, *Rivista di Storia e Letteratura Religiosa* 23: 222-39.(A well-documented analysis of Huet’s interest in Descartes and of his later *Censura*, which subordinates Huet’s scepticism to his anti-Cartesianism.)

Flottes, J.B.M. (1857) *Étude sur Daniel Huet, évêque d’Avranches*, Montpellier: F. Seguin.(Credibly defends Huet from the charge of being a Pyrrhonian. It summarizes Huet’s *Démonstration*, *Censure*, *Questions* and *Traité*, attempting to present him as a systematic writer.)

Guellouz, S. (ed.) (1994) *Pierre-Daniel Huet, 1630-1721: Actes du Colloque de Caen 12-13 novembre 1993*, Biblio 17, *Suppléments aux Papers on French Seventeenth Century Literature*, Paris.(Includes twenty-two contributions, covering many aspects of Huet’s biography, interests, intellectual production and latter fortune, it includes an extended and updated bibliography.)

Henry, C. (1879) *Un érudit homme du monde, homme d’église, homme de cour*, Paris: Librairie Hachette.(A selection of Huet’s correspondence.)

Lux, D. S. (1983) *Royal patronage and seventeenth century science: l’Academie de physique de Caen, 1662-1672*, Ph.D. thesis, University of Michigan.(Presents the institutional history of the Academie de physique de Caen (1662-72) and Pierre-Daniel Huet’s role as the group’s founder and early patron.)

Malbreil, G. (1985) ‘Les droits de la raison et de la foi, la dissociation de la raison, la métamorphose de la foi, selon P.D. Huet’, *Dix-septieme Siècle* 147: 119-33.(Reconstructs various interpretations of Huet’s philosophy and, by focusing on the *Alnetanae Quaestiones*, defends an apologetic rather than a fideistic view of his Christian scepticism.)

Malbreil, G. (1991) ’Descartes censuré par Huet’, *Revue Philosophique de la France et de l’Etranger* 116: 311-28.(Accurate, historical study of Huet’s critique of Descartes, as presented in the *Censura* and the *Nouveaux memoires*, with ample references to the debate between Huet and Regis.)

Massimi, J.-R. (1985) ‘Vérité et histoire chez P.D. Huet’, *Dix-septième Siècle* 147: 167-8.(A review of the *Demonstratio* suggesting that, for Huet, truth is discovered by investigating the sources of our present beliefs and customs in the past.)

McKenna, A. (1985) ‘Huet and Pascal’, *Dix-septième Siècle* 147: 135-42.(A philological and comparative study based on Huet’s remarks on the margins of his copy of Pascal’s *Pensées*. It focuses mainly on the *Demonstratio*.)

Monod, A. (1916) *De Pascal à Chateaubriand: les défenseurs français du christianisme de 1670 à 1802*, Alcan. (Contains an insightful presentation of Huet’s fideism on pp. 83-90.)

Muratori, L. A. (1745) *Delle Forze dell’Intendimento umano, o sia il Pirronismo confutato, trattato [...] opposto al libro del preteso Monsignore Huet intorno alla debolezza dell’umano intendimento*, Venezia: reprinted many times until 1782.(The most radical attempt to refute Huet’s scepticism, written from a Christian perspective. Learning prevails over logical force.)

Neto, J.R.M. and Popkin, R.H. (1995) ‘Bishop Pierre-Daniel Huet’s Remarks on Pascal’, *British Journal for the History of Philosophy* 3: 147-60.(The first half of the paper presents a standard summary of Huet’s scepticism interpreted in terms of religious fideism; the second half contains a more interesting analysis of Huet’s *marginalia*. The authors conclude that Huet’s humanistic and naturalistic Christianity presents a vision of the human situation that is both existentially and religiously much less dramatic than Pascal’ s Augustinianism.)

Petermann, A. (1690) *Philosophiae Cartesianae adversus Censuram Petri Danielis Huetii vindicatio, in qua pleraque intricatiora Cartesii loca clare explanantur*, Lipsiae: Impensis Johannis Caspari Meyeri.(Probably the first attempt to defend Descartes against Huet’s sceptical attack.)

Popkin, R. (1965) ’The High Road to Pyrrhonism’, *American Philosophical Quarterly*, 2: 18-32.(A survey of sceptical thought in the latter half of the 17th century, and an attempt to characterize its major features in terms of epistemological attack against Cartesianism, includes a large section on Huet.)

Regis, P. S. (1691) *Réponse au livre qui a pour titre P. Danielis Huetii censura philosophiae Cartesianae servant d’eclaircissement a toutes les parties de la philosophie, sur tout a la metaphysique*, Paris: Jean Cusson.(On Huet’s reply, unpublished, see Dini (1987).)

Rodis-Lewis, G. (1985) ‘Huet lecteur de Malebranche’, *Dix-septième Siècle* 147: 169-89.(Much information on Huet’s library, especially the philosophical books, together with a philological collection and reconstruction of his remarks on Malebranche.)

Schotanus, J. (1702) *Discussio Censurae Huetianae [...]*, Amsterdam.(Criticism of Huet’s sceptical and anti-Cartesian position.)

Sciacca, G.M. (1968) *Scetticismo Cristiano*, Palermo: Palumbo.(Reconstructs Huet’s cultural milieu, and provides two long paraphrases of his *Traité* and Muratori’s attack.)

Tolmer, L. (1949) ‘Pierre-Daniel Huet (1630-1721) humaniste, physicien’, *Mémoires de l’Académie nationale des*

sciences, arts et belles-lettres de Caen, Nouvelle série, vol. 11.(Probably the most complete intellectual biography written so far, includes a wealth of interesting details on Huet's studies, his cultural milieu and his scientific activities.)

Volder, B. de (1695) *Exercitationes academicae, quibus Ren. Cartesii philosophia defenditur adversus Petri Danielis Huetii […] Censuram philosophiae Cartesianae*. Amsterdam.(A defence of Descartes against Huet's sceptical arguments.)

Hugh of St Victor (d. 1141)

Hugh of St Victor initiated the teaching programme that distinguished the Parisian abbey of St Victor during the twelfth century. His teaching combined an ambitious programme of biblical exegesis with the construction of theological syntheses and the detailed subordination of a comprehensive philosophy to both. Hugh's principal works in theology are biblical commentaries of different kinds and a theological overview based on the notion of sacrament. His philosophical works include Epitome Dindimi in philosophiam (Dindimus' Epitome of Philosophy) and, most importantly, the Didascalicon de studio legendi (Didascalion, or On the Study of Reading). This last book attempts to show how all human knowledge can be used as preparation for the study of the Christian Bible, which in turns leads to the contemplation of God.

Hugh of St Victor initiated the project of teaching and writing that came to distinguish the abbey of St Victor in Paris during the twelfth century. He is, in that sense, the first 'Victorine' theologian. His teaching combined an ambitious program of biblical exegesis with the construction of theological syntheses and the detailed subordination of philosophy to both.

The abbey of St Victor grew out of the community founded by William of Champeaux on his retreat from the Parisian schools of theology in 1108. William brought to the community customs of learning, but no distinctive theological or philosophical project. On his consecration as bishop in 1113, William secured for the small group a significant endowment and a royal charter permitting the free election of an abbot.

When Hugh entered the abbey sometime between 1115 and 1118, it had already begun a thorough reform of monastic life under its first real leader, Gilduin. It was thus ready to receive the kind of intellectual definition Hugh's work and his legacy would bring. Indeed, Hugh was a pioneer in each of the several fields that would become distinctive of the abbey, even as he brought it considerable renown as a school. Hugh himself served as the abbey's prior from 1133, and he died in 1141 within the community.

Hugh's study and teaching ranged widely. Indeed, he could be admired by someone such as Bonaventure just for the breadth of his education. One important group of his writings consists of scriptural interpretations. Hugh wrote three works in which he unfolded the image of Noah's ark into a paradigm of the stages of contemplation. Since the image of the ark is treated in enormous detail as a visual image, Hugh's tropological reading may also have had some influence on the aesthetics of Gothic art. However, Hugh also practised a kind of literal interpretation that put a premium on historical accuracy and made ample use of contemporary Jewish exegesis. One instance of his literal interpretation is a *Chronica (Chronicle)* in which he outlines the rational order of Hebrew history. This style of reading would be carried forward by Hugh's student, Andrew of St Victor. Hugh further undertook the construction of a unified or 'systematic' theology in his *De sacramentis christianae fidei (On the Sacraments of the Christian Faith)*. This impulse to theological construction was continued by another of Hugh's followers, Richard of St Victor, who knew of Hugh only through his writings. Hugh also produced a number of spiritual writings in more traditional genres, as well as sermons and letters.

Hugh produced several works that can be counted philosophical. There are, for example, introductory treatises on the liberal arts. One of these, *Epitome Dindimi in philosophiam (Dindimus' Epitome of Philosophy)*, is a short dialogue that lays out definitions and divisions of philosophy for the beginner. It can serve to show how broadly Hugh understands philosophy. He divides philosophy into four parts: logic, ethics, theory and mechanics. Logic comprises grammar and the rule of inference (*ratio disserendi*), which is in turn divided into probable, necessary and sophistical reasoning. Ethics comprises the solitary or individual, the domestic or economic and the public or political. Theory covers mathematics, physics and theology, which here means not Christian doctrine but the study of immaterial beings. In last place, Hugh incorporates what he calls 'mechanics'. Here he means not only technologies such as weaving or techniques such as navigation, but medicine and even theatre. Hugh seems the first Latin author to make 'mechanics' a part of philosophy (see Natural philosophy, medieval).

The disposition of the parts of philosophy and their subordination to Christian ends is accomplished more fully in Hugh's *Didascalicon de studio legendi (Didascalion, or On the Study of Reading)*, which must be regarded as his masterpiece in philosophy. In accord with current taste, Hugh borrows for his title the Greek term *didascalicon*,

which he probably takes to mean the art of teaching. His *Didascalicon* is, indeed, a work that teaches what, when, and how to read for the sake of a well-formed wisdom. The *Didascalicon* is divided into two parts, each of which has three Books. The first part concerns philosophy and all its parts, which are distinguished much as they were in the *Epitome*. The second part, which finds no correspondence in the *Epitome*, concerns the reading of 'holy writings', which are principally but not exclusively found in the Christian Bible.

The treatment of philosophy in the *Didascalicon* is considerably more detailed than in the *Epitome*. Here Hugh provides not just descriptions of each part of philosophy, but remarks on their founders or principal authors and on the sequence of study. Throughout, he follows traditional sources. He seems mostly uninfluenced by the pieces of Arabic science that were beginning to find their way into Latin learning (see Translators). More importantly, Hugh here reiterates an Augustinian account of how the various moments in the study of philosophy adumbrate a unified wisdom in which 'the form of the complete good' is grasped (see Augustinianism). This wisdom is, of course, Christian. The study of 'holy writings' both completes and corrects the study of philosophy. Hugh introduces the second part of the *Didascalicon* by comparing the writings of philosophers to a whitewashed wall that covers over the clay of error with seeming truth. The Christian Bible, by contrast, is like a honeycomb: dry on the outside, filled with sweetness within.

The second part of the *Didascalicon* is in this way something more than an introduction to biblical study. It does contain a number of chapters on the levels of scriptural meaning, as on the canon and the rules for resolving ambiguous interpretations. However, it directs these materials to a practice of meditation in which scriptural reading is enacted as communion with God. This communion is the Christian's philosophy. Indeed, the person whom Hugh calls the 'Christian philosopher' turns out to be a monk reading Scripture prayerfully. 'The monk's simplicity is his philosophy', and the monk's 'philosophizing' is not academic disputation, but delight in learning from Scripture the life of virtue.

See also: Richard of St Victor

MARK D. JORDAN

List of works

Hugh of St Victor (1118-25?) *Practica geometriae (Practice of Geometry)*, in R. Baron (ed.) *Hugonis de Sancto Victore Opera propaedeutica*, Notre Dame, IN: University of Notre Dame Press, 1966; trans. F.A. Homann, *Practical Geometry*, Milwaukee, WI: Marquette University Press, 1991.(An introduction to applied plane geometry.)

Hugh of St Victor (1118-25?) *De grammatica (On Grammar)*, in R. Baron (ed.) *Hugonis de Sancto Victore Opera propaedeutica*, Notre Dame, IN: University of Notre Dame Press, 1966.(An introduction to the elements of Latin grammar.)

Hugh of St Victor (1118-25?) *Epitome Dindimi in philosophiam (Dindimus' Epitome of Philosophy)*, in R. Baron (ed.) *Hugonis de Sancto Victore Opera propaedeutica*, Notre Dame, IN: University of Notre Dame Press, 1966. (An introduction to the scope and rudimentary principles of philosophy most broadly conceived.)

Hugh of St Victor (1120-5) *Didascalicon de studio legendi (Didascalicon, or On the Study of Reading)*, ed. C.H. Buttimer, Washington, DC: Catholic University of America Press, 1939; trans. J. Taylor, *Didascalicon: A Medieval Guide to the Arts*, New York: Columbia University Press, 1961.(Taylor's translation contains an ample introduction and very rich notes.)

Hugh of St Victor (1130-3) *De sacramentis christianae fidei (On the Sacraments of the Christian Faith)*, Patrologia Latina 176: 173-618; trans. R.J. Deferrari, *On the Sacraments of the Christian Faith*, Cambridge, MA: Mediaeval Academy of America, 1951.(Hugh's magisterial survey of the foundations of Christian faith and life.)

Hugh of St Victor (before 1141) *Chronica (Chronicle)*, partial edn in L.B. Mortensen, 'Hugh of St Victor on Secular History: A Preliminary Edition of Chapters from His *Chronica*', *Cahiers de l'Institut du Moyen Âge Grec et Latin* 62, 1992: 3-30.(An edition of a portion of Chapter 3 and Chapter 9 from a single manuscript.)

References and further reading

Baron, R. (1957) *Science et Sagesse chez Hugues de Saint-Victor (Knowledge and Wisdom in Hugh of St Victor)*, Paris: Lethielleux. (A general introduction to Hugh's writings.)

Châtillon, J. (1992) *Le mouvement canonial au Moyen Age: Réforme de l'Église, spiritualité et culture (The Movement of Canons in the Middle Ages: Reform of the Church, Spirituality and Culture)*, ed. P. Sicard, Paris

and Turnhout: Brepols.(An anthology of earlier essays, several of them devoted to Hugh.)
Ehlers, J. (1973) *Hugo von St Victor: Studien zum Geschichtsdenken und der Geschichtsschreibung des 12. Jahrhunderts (Hugh of St Victor: Studies on the Concepts of History and Historiography in the Twelfth Century),* Wiesbaden: Steiner.(Studies of Hugh's concept of history and practice of historiography.)
Girolimon, M.T. (1994) 'Hugh of St Victor's *De sacramentis Christianae fidei*', *Journal of Religious History* 18 (2): 127-38.(A useful introduction to the intention of Hugh's work on the sacraments.)
Goy, R. (1976) *Die Überlieferung der Werke Hugos von St Viktor (The Transmission of the Works of Hugh of St Victor),* Stuttgart: Hiersemann, 1976.(A very detailed scholarly study of the authenticity of the corpus attributed to Hugh.)
Illich, I. (1993) *In the Vineyard of the Text: A Commentary to Hugh's Didascalicon*, Chicago, IL: University of Chicago Press.(A free meditation on Hugh's *Didascalicon*; not to be used as an accurate paraphrase of the text.)
Van den Eynde, D. (1960) *Essai sur la succession et la date des écrits de Hugues de St-Victor (Essay on the Sequence and Date of Writings of Hugh of St Victor)*, Rome: Pontificium Athenaeum Antonianum.(A carefully reasoned hypothesis about the chronology of Hugh's writings; not universally accepted.)

Human nature

Every political philosophy takes for granted a view of human nature, and every view of human nature is controversial. Political philosophers have responded to this conundrum in a variety of ways. Some have defended particular views of human nature, while others have sought to develop political philosophies that are compatible with many different views of human nature, or, alternatively, which rest on as few controversial assumptions about human nature as possible. Some political philosophers have taken the view that human nature is an immutable given, others that it is shaped (in varying degrees) by culture and circumstance. Differences about the basic attitudes of human beings toward one another - whether selfish, altruistic or some combination - have also exercised political philosophers. Although none of these questions has been settled definitively, various advances have been made in thinking systematically about them.

Four prominent debates concern: (1) the differences between perfectionist *views, in which human nature is seen as malleable, and* constraining *views, in which it is not; (2) the* nature/nurture *controversy, which revolves around the degree to which human nature is a consequence of biology as opposed to social influence, and the implications of this question for political philosophy; (3) the opposition between* self-referential *and* other-referential *conceptions of human nature and motivation - whether we are more affected by our own condition considered in itself, or by comparisons between our own condition and that of others; and (4) attempts to detach philosophical thought about political association from all controversial assumptions about human nature.*

1 Perfectionist versus constraining views of human nature

The most celebrated proponent of perfectionist views of human nature is Aristotle, and his most elaborate discussion of the question appears in *The Nicomachean Ethics* (*c*.330 bc). Aristotle's view developed there is sometimes described as teleological because it was defined by reference to a fundamental contrast between untutored or brute human nature on the one hand, and human nature as it could be if we realized our essential nature or *telos* on the other. Ethics, for Aristotle, is the science that instructs us how to get from the one to the other. When correctly employed, ethical precepts thus have the potential to transform human beings from their rude and untutored states into the best kinds of beings they can become.

Although there is both room and need for human nature to develop and change on Aristotle's view, his is a *naturalist* one in that both the accounts of brute human nature and perfected human nature are rooted in a philosophical psychology, a theory of natural human needs and potentials. Naturalist views are generally distinguished from *anti-naturalist* accounts, in which the sources of meaning and value for human beings are externally given, whether in Platonic forms, transcendental arguments about the nature of knowledge and obligation, or some other extrinsic origin. On anti-naturalist views what is good or right is good or right regardless of actual human needs and desires, whereas for naturalists the gulf between 'is' and 'ought' is bridged via an account of human needs or psychology (see Naturalism in ethics; Needs and interests). Naturalist perfectionism can be secular or theologically-based; indeed, much medieval Christian Aristotelianism rested on the adaptation of the story of the Garden of Eden, the Fall from Grace, and the possibility of redemption to Aristotle's ethical categories. Marxism, on the other hand, is a secular version of naturalist perfectionism. For Marx, human beings as we find them are alienated from their 'true' selves, their potential stunted by a variety of malevolent forces. But these forces are in some sense artificial impediments to human flourishing that can be understood and ultimately vanquished, opening the way for authentic human development (Marx 1875) (see Alienation §6).

Although perfectionist views can be naturalist or anti-naturalist, constraining views of human nature are invariably naturalist in character. Proponents of constraining views regard human nature as an immutable given, holding that social and political institutions must take account of it but that have no effect on it. Perhaps the most extreme proponent of this view in the history of Western philosophy was Jeremy Bentham (§3), who made it an axiom of his utilitarian system that nature: 'has placed mankind under the governance of two sovereign masters, *pain* and *pleasure*' (1948: 125). For Bentham, pleasure seeking and pain avoidance 'govern us in all we do, in all we say, in all we think: every effort we can make to throw off our subjection, will serve but to demonstrate and confirm it' (1948: 125). Bentham believed that the utilitarian system he advocated was uniquely scientific in that it recognized our inescapable subjection to the pleasure/pain calculus.

Although Bentham offers one of the most forthright and clear accounts of a constraining view of human nature, his is by no means the first or the only such view in the history of Western political theory. In their modern form such views are traceable at least to Hobbes' insistence, *contra* Aristotle, that human nature is rigidly fixed. For Hobbes human beings are inescapably driven by a primordial fear of death, which produces in them a 'relentless desire for power after power'. Basic human impulses can be channelled in more and less productive ways, and to some extent people can be educated to see that certain political allegiances are better than others given their basic impulses, but the impulses themselves cannot be altered (Hobbes 1651) (see Hobbes, T. § 6).

Utilitarians since Bentham have generally regarded human nature as given and external to the operations of political and cultural institutions. Even if utilitarians have differed greatly from one another concerning the content of human nature and even how much we can reasonably aspire to know about its content, they have generally accepted Hume's dictum that 'reason is the slave of the passions'. Whereas Hume, like Bentham, thought we are all driven by the same passions in the same ways, neoclassical utilitarians of the late nineteenth and early twentieth centuries (such as Pareto 1970) questioned this assumption, taking a view of human nature that was perhaps most decisively stated by the emotivist philosopher Charles Stevenson in his critique of Hume: that even if we are convinced that all human beings are driven by needs for gratification of some sort, there is no decisive reason to think we all have the same such needs, or even that we can intelligently compare different desires experienced by different individuals with one another (Stevenson 1944). This is the standard view in modern economics and rational choice theories of politics: Preferences are thought of as exogenously determined, but the possibility of meaningful interpersonal comparisons of preferences across individuals is denied (see Utilitarianism; Economics and ethics).

2 The nature/nurture controversy

Since time immemorial scientists and philosophers have debated the question whether human nature is biologically given or rather the result of contingencies of cultural and historical circumstance, and these debates have been thought to be pregnant with political significance. Rousseau (§2), for example, was the first in a long line of political theorists who have argued that Hobbes' political system was based on a confusion of parochial traits that were the product of seventeenth-century English life with enduring features of the human condition. In Rousseau's view, Hobbes wrongly thought the impulse toward self-preservation was incompatible with a desire to preserve others: 'because of [his] having improperly included in the savage man's care of self-preservation the need to satisfy a multitude of passions which are the product of society' (1964: 129). Rousseau's own account of the state of nature has been criticized on analogous grounds however. Indeed it has become a standard move in modern political argument to question every assertion about natural human traits and propensities, however qualified or historicized, by arguing that the traits in question are socially produced or 'socially constructed'. Such debates can become exceedingly heated when they become embroiled in discussions of the determinants of intelligence and such things as ethnic, racial and gender differences (see Sociobiology).

The nature/nurture controversy is often confused with the issue of whether or not human nature is alterable by human design. It is often said, for instance, that much or all of human nature is socially constructed rather than naturally given, and that for this reason it should not be regarded as immutable. In fact, whether a given human characteristic is a product of nature or culture may have little to do with the degree to which it is alterable by conscious human design. On the one hand, there are many features of the natural world that human beings have effectively altered and will alter more in the future; the science of genetic engineering, for example, is presumably in its infancy. On the other hand, there are many pieces of human reality that, while indisputably the product of human action, we often seem powerless to influence. Ethnic hatred may be learned, for example, yet it may be impossible to get people to unlearn it or even to stop them passing it on to the next generation.

It is sometimes thought that cultural creations can be more easily understood than natural creations because they are the products of the human will. This view was common in the seventeenth century, for example, where philosophers like Hobbes and Locke, under the strong influence of the Cartesian idea that the reflective individual has privileged access to the contents of their own mind, held that only this individual can really know 'the ghost in the machine' in Gilbert Ryle's memorable phrase. Knowing one's own motivations and understanding one's behaviour guarantees nothing, however, about one's understanding of the complexities of human interaction in the social and political arena. In any case, in a post-Freudian age the Cartesian view must be regarded as questionable.

We might be confused, or wrong, or we might deceive ourselves about our motivations and purposes. For all we know Bentham might have been right when he insisted that, 'it is with the anatomy of the human mind as with the anatomy and physiology of the human body: the rare case is, not that of a man's being unconversant, but of his being conversant with it' (1954, vol. III: 425). In short, the human capacity to shape human nature is a product of how well we can understand and control the causal mechanisms that lead human traits to be what they are. This is contingently related, if it is related at all, to whether a particular trait is a product of biology, culture, some combination of the two, or some third thing.

For all the attention it has received from political philosophers over the centuries, the nature/nurture controversy is a red herring in a second respect as far as politics is concerned. Since human beings are naturally conventional creatures who often achieve biological adaptation through social learning, the distinction between what is natural and what is cultural in human behaviour is not merely difficult to pin down in practice, but impossible to get at in principle. Perhaps partly for this reason, in recent years much of the discussion in Anglo-American political and moral philosophy has turned away from the nature/nurture debate and towards requiring principled justifications for any differences among persons that have political consequences, regardless of their origin. John Rawls is perhaps more responsible for this than any other single figure. In *A Theory of Justice* (1971) he makes the case that differences in talents and abilities among persons, whether natural or cultural in origin, are 'morally arbitrary'; that is, they are products of luck either in the genetic lottery or in the milieu into which one happens to be born. In either case, the differences are not chosen or produced by actions for which the relevant agents can reasonably be held responsible. By the same token, it seems unfair that benefits should accrue to persons differentially in virtue of differences in their talents and abilities, whether such differences are rooted in nature, culture or both.

Making the Rawlsian move has the effect of socializing human capacities regardless of their origins (at least as a normative matter), and although it is exceedingly difficult to find a principled basis on which to resist his reasoning, it generates problems that are in many ways as difficult as those that it resolves. A person might find it both rationally undeniable and psychologically impossible to accept that nothing they do or achieve is the autonomous result of their own efforts. Analogously, it might be argued that for both individual and species some fictions about individual responsibility for different outcomes may be required for human reproduction and wellbeing, even if we know them to be fictions. Such beliefs may be indispensable to the basic integrity of the human psyche and necessary for generating and sustaining the incentive to work on which human beings are, after all, critically reliant. As a result, although facts about moral luck and socially produced productive capacities conspire - when confronted - to enfeeble the idea of individual responsibility for outcomes, people may none the less be powerless to abandon it (see Responsibility). One interesting strand of modern scholarship by Ronald Dworkin, Amartya Sen, Richard Arneson and others attempts to grapple with the issues raised by Rawls' discussion of moral arbitrariness, but it has not converged on any single or compelling conclusion (see Dworkin, R.).

3 Self-referential versus other-referential conceptions of human nature

A third dimension along which conceptions of human nature vary concerns the degree to which people's interests are linked to the fortunes and perceptions of others. Contrast Ronald Reagan's 1984 election slogan: 'Are you better off now than you were four years ago?' with an employee who says to his employer 'I don't care what I get paid, so long as its more than Jones down the corridor'. The Reagan slogan is based on a *self-referential* conception of human nature; it assumes that people are interested in improving their own fortunes without any necessary reference to anyone else. The latter example is *other-referential*: a person's perception of their interests is intrinsically reliant on the welfare of others. We cannot make a judgment about their welfare without reference to the welfare of others. The distinction between self- and other-referential views operates independently of whether one is a subjectivist and of the basic metric of value along which welfare is judged. For instance, Marx's measure of exploitation is radically other-referential in that it is critically reliant on the proportion of the social surplus that the workers receive as compared with employers. Thus his technical measure of the rate of exploitation may increase even if real wages are increasing and the worker's subjective sense is that his welfare is improving. Rawls, on the other hand, works with a hybrid of self- and other-referential views of human nature in that he argues that rational people will always want to improve the lot of the worst-off individual in society, relative to what it was before, even if this means that who is the worst off changes.

Whether one thinks human nature is basically self- or other-referential has substantial implications for politics and distributive justice. Utilitarians, both classical and neoclassical, tend to think of people as self-referential maximizers who want to get on as high an indifference curve as possible without reference to anyone else, although there is no necessary reason that this be so. Aristotelians like Alasdair MacIntyre (1981), on the other hand, are necessarily committed to other-referential views. For them what people value above all is to be valued by others whom they value (see MacIntyre, A.; Recognition). 'She's a cellist's cellist' is an appellation we intuitively grasp, and to the extent that it captures something fundamental about what motivates people, models of individual maximization that ignore it will be descriptively misleading and morally unsatisfying. In this connection it is worth noting that in Hegel's (1807) *Phenomenology* the argument is made that slavery is an unstable set of social arrangements not merely because the slave can eventually be expected to resist oppression, but also because the master will find it unsatisfying; he needs recognition from someone whom he values (see Hegel, G.W.F.).

Whether and to what extent human nature is substantially self-referential is not to be confused with whether and to what degree people are selfish. Although self-referentially motivated people may typically be indifferent - at best - to the fortunes of others, other-referential motivations may be more or less selfish. The demand to be paid 'more than Jones', 'the same as Jones' or 'less than Jones' are all other-referential in character, although they vary considerably on the dimension between selfish and altruistic motivation. Sadists and rapists exhibit what economists refer to as 'interdependent negative utilities' in that their wellbeing is intrinsically linked to the suffering of others; so too does the divorcing spouse whose utility increases if a dollar is taken out of the pocket of their spouse and burned. These are all other-referential types of motivation, no more or less than are the motivations of those who seek and experience increased happiness at another's happiness or success. In short, the self-referential/other-referential dimension refers to the structure of human psychology rather than its content. This is not to say that the structure has no implications for content. In particular, to the extent that people are other-referentially motivated, their behaviour and aspirations will tend not to be captured by rational agent models adapted from microeconomics, which rest exclusively on self-referential assumptions about human psychology. People who find arguments based on rational agent models unsatisfying descriptively or morally tend to be sceptical of the extent to which self-referential assumptions capture important dimensions of the human condition. Whether or not such scepticism is warranted is in the end an empirical question; it will not be settled by armchair reflection (see Rational choice theory).

4 'Political' versus 'metaphysical' conceptions of human nature

Given the enduring - perhaps endemic - disagreements about human nature it is perhaps unsurprising that some political philosophers have sought to develop political theories assuming as little as possible about it. One strategy that has been extensively explored since the 1960s is to try to seek political principles that are neutral between different views of human nature and conceptions of the human good. Some of Rawls' earlier formulations of his project were understood thus, and Bruce Ackerman, Charles Larmore and others subsequently attempted to formulate different neutrality arguments (see Neutrality, political §§1, 4). By the 1990s a fairly broad consensus had begun to emerge that neutrality arguments fail, that there is no set of political principles that can be genuinely neutral among different conceptions of human nature and the human good, and that a little digging into allegedly neutral arguments will inevitably bring their assumptions to the surface.

A subtler response to the problem is to abandon the search for neutral political standards, but retain the ambition to develop political principles that may be compatible with as many different views of the human condition as possible. Here the key idea is to try to limit what one might reasonably expect of justification in political theory, and in particular to abandon all aspirations to get at the true or right theory of human nature. This is the strategy adopted by the later Rawls (1992), who asks: which political principles does it make most sense to agree on, given that we know we will never all agree on basic metaphysical questions about the human condition? The answer to which he is drawn is the one that tolerates as many conceptions of the human good as possible, consistent with a 'like liberty for all'. This approach explicitly abandons the agenda of coming up with right answers to questions about human nature, seeking only to tolerate as many answers as possible. It courts the possibility that it may actually lead us not to tolerate the 'true' view of human nature if, for instance, this is embedded only in a fundamentalist religion that also requires that no other views be tolerated. On a 'political not metaphysical' approach, we are not expected - as political theorists - to take any view on the question whether the fundamentalist view is correct; for all we know it might be. However, since we can never know that it is the correct view and we

do know that there will always be competing views, it is never rational to accept it as governing our politics. Various difficulties confront this and other attempts to move in the ‘political not metaphysical’ direction, and some will find the philosophical abdication implied by this strategy impossible to live with. But in a world in which assumptions about human nature are both indispensable to politics and endemically controversial, it is difficult to see how the impulse to turn to institutional, rather than philosophical, solutions to the problems it creates can reasonably be resisted.

See also: Xing; Political philosophy, nature of

IAN SHAPIRO

References and further reading

Ackerman, B. (1980) *Social Justice and the Liberal State*, New Haven, CT: Yale University Press.(Defence of a theory of justice that is intended to be neutral among different conceptions of human nature and the human good. Written in dialogue form, generally accessible to the nonspecialist.)

Aristotle (perhaps *c.*330BC) *The Nicomachean Ethics*, trans. J.A.K. Thompson, Harmondsworth: Penguin, 1977. (The most mature and elaborate statement of Aristotle’s ethical naturalism. A difficult work.)

Arneson, R. (1989) ‘Equality and Equal Opportunity for Welfare’, *Philosophical Studies* 56 (1): 77-93.(A utilitarian response to the moral arbitrariness inherent in the distribution of natural capacities. Moderately difficult.)

Ashcraft, R. (1987) *Locke’s Two Treatises of Government*, London: Allen & Unwin.(An accessible introduction to Locke’s views on human nature and natural law.)

Bentham, J. (1948) *A Fragment on Government and An Introduction to the Principles of Morals and Legislation*, Oxford: Blackwell.(The *locus classicus* of Bentham’s utilitarianism. Contains a programmatic statement of his assumptions about human nature, as well as applications in the fields of politics, law and economics. Easy to read despite arcane English.)

(1954) *Jeremy Bentham’s Economic Writings*, ed. W. Stark, London: Allen & Unwin, 3 vols.(The complete collection of all Bentham’s writings on economics and economic psychology. Accessible.)

Dworkin, R. (1981) ‘What is Equality? Part I: Equality of Elfare’, *Philosophy and Public Affairs* 10 (3): 185-246; and ‘What is Equality? Part II: Equality of Resources’, *Philosophy and Public Affairs* 10 (4): 283-345.(A discussion of the moral arbitrariness of the distribution of natural capacities, and a proposal to deal with this in thinking about distributive justice. Moderately difficult.)

Hegel G.W.F. (1807) *Phenomenology of Spirit*, trans. A.V. Miller, Oxford: Oxford University Press.(Chapter 4 contains the famous discussion of ‘Lordship and Bondage’ commented on in §3. A difficult work.)

Hobbes, T. (1651) *Leviathan*, London: Pelican, 1968.(Part one contains the fullest statement of Hobbes’ views about human nature. Easy to read despite arcane English.)

Larmore, C. (1987) *Patterns of Moral Complexity*, Cambridge: Cambridge University Press.(Defence of a theory of justice that is intended to be neutral among different conceptions of human nature and the human good. A difficult book written for the specialist in political philosophy.)

Locke, J. (1664) *Essays on the Law of Nature*, Oxford: Clarendon Press, 1958.(The most elaborate statement of Locke’s views on natural law. The essays deal with his views of the immutability of natural law, the nature of its force as law and its accessibility to human beings. A difficult work.)

(1690) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1960.(The most elaborate statement of Locke’s views on human nature and politics. Reasonably accessible.)

MacIntyre, A. (1981) *After Virtue: An Essay in Moral Theory*, London: Duckworth.(A neo-Aristotelian theory of politics and human nature. Reasonably accessible.)

Marx, K. (1867) *Capital*, vol. 1, Harmondsworth: Penguin, 1976.(Chapter 9 contains the discussion of exploitation mentioned in this entry. A difficult work.)

(1875) *Critique of the Gotha Program*, Beijing: Foreign Language Publishers, 1972.(Contains Marx’s most mature and elaborate attempt to distinguish human needs from human wants, and to develop the implications of this distinction for distributive justice under capitalism, socialism and communism. An accessible work.)

Pareto, V. (1970) *Manual of Political Economy*, New York: Augustus Kelly.(The most elaborate statement ever developed of the logical bases of neoclassical utilitarianism. An exceedingly difficult work.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(The most elaborate statement of Rawls’ theory. The pertinent discussion of assumptions about human nature appears in chapter 7. A difficult work.)

Rawls, J. (1992) *Political Liberalism*, New York: Columbia University Press.(Contains Rawls’ attempt to detach

his theory more completely than in his earlier work from controversial assumptions about human nature. A difficult work.)

Riley, P. (1982) *Will and Political Legitimacy*, Cambridge, MA: Harvard University Press.(A useful and accessible survey of different views of the human will, and their implications for politics, in the Western tradition since the seventeenth century.)

Rousseau, J.-J. (1964) *The First and Second Discourses*, ed. R. Masters, New York: St Martin's Press.(Contains Rousseau's celebrated critique of Hobbes' view of human nature as fixed, and defence of a developmental view in its stead.)

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.(A seminal critique of the Cartesian view of the mind as a 'ghost in the machine'. A difficult work.)

Sen, A. (1985) 'Well-being, Agency and Freedom', The Dewey Lectures 1984, *The Journal of Philosophy* LXXXII (4).(An attempt to replace the concept of human nature with that of wellbeing as a basis for thinking about social justice. Moderately difficult.)

Shapiro, I. (1986) *The Evolution of Rights in Liberal Theory*, Cambridge: Cambridge University Press.(A history of changing conceptions of human nature in the Anglo-American tradition since the seventeenth century, and their implications for liberal theories of individual rights. Moderately accessible.)

(1991) 'Resources, Capacities and Ownership: The Workmanship Ideal and Distributive Justice', *Political Theory* 19 (1): 47-72.(A discussion of the origins and implications of the 'socialization of human capacities' discussed in this entry. Moderately accessible.)

Stevenson, C.L. (1944) *Ethics and Language*, New Haven, CT: Yale University Press.(A critique of Hume's assumption that human nature is invariant across individuals, developed in the course elaborating the controversial view that ethical propositions are no more than expressions of emotion. A moderately difficult work.)

Human nature, science of, in the 18th century

Eighteenth-century speculation on human nature is distinguishable by its approach and underlying assumptions. Taking their cue from Francis Bacon and Isaac Newton, many philosophers of the Enlightenment endeavoured to extend the methods of natural science to the moral sciences. Perhaps the most explicit of such endeavours was David Hume's ambition for a 'science of man', but he was not alone. There was a general convergence on the idea that human nature is constant and uniform in its operating principles - that is, its determining motives (passions), its source of knowledge (sense experience) and its mode of operation (association of ideas). By virtue of this constancy human nature was predictable, so that once it was scientifically understood, then social institutions could be designed to effect desired outcomes.

David Hume regarded his proposed 'science of man' as fundamental because upon it depended not only logic, morals, criticism and politics but also even mathematics, natural philosophy and natural religion. Hence he believed that if that science is conquered then a relatively easy victory over the others can be expected. We know from Hume's own *Abstract* (1740) of the *Treatise of Human Nature* (1739-40) that he himself thought the most original element was the use made of the principle of the association of ideas. Hume identified three principles of association - resemblance, contiguity and causation (see Hume, D. §2). Starting from the Lockean premise that any inquiry into human nature must rely entirely upon experience, these principles operate uniformly for humans to bind the universe together. Just as there is, accordingly, a general course of nature with respect to the sun and climate, so there is in human actions; and just as there has been great success in the understanding of natural phenomena, so Hume aimed to emulate that success in the moral sciences through adopting its 'experimental' procedures. While this was Baconian in provenance (see Bacon, F.), it was Newton's achievements that were more generally paradigmatic (see Newton, I.).

Newton himself had indicated the appropriateness of the endeavour when in Question 31 (1953: 179) of the *Opticks* (1730) he remarked that if the analytical method of natural philosophy is perfected then 'the bounds of moral philosophy will also be enlarged'. Since for Newton the key to natural philosophy lay in mathematics, one characteristic of the eighteenth-century attempt to establish a science of human nature was to subject the phenomena of human nature to quantification. Francis Hutcheson, whom Hume identified as one of his predecessors in the scientific quest, flirted with this when he developed a calculus derived from certain simple axioms whereby $M=B \times A$ or the moment of good equals benevolence times ability (see Hutcheson 1725).

Although Hutcheson excised this calculus from later editions of the *Inquiry*, this quantitative and calculative concern persisted. Jeremy Bentham stated that we are governed by the two sovereign principles of pleasure and pain, and that it is the principle of utility that recognizes this subjection and employs it as the foundation of a system. The object of this system is to promote felicity, and to achieve that end with precision it is necessary to calculate which policies will produce a net balance of pleasure over pain. Armed with this scientific knowledge, the legislator will be able to pass laws that can be guaranteed to further the interests of the community. The view that this science should be useful is typical. Cesare Beccaria, in his argument for a reformed penal system (see Beccaria 1764), held that punishment should be determined with geometric precision, and that the probability of guilt amounts to moral certainty - something humans necessarily presuppose in daily life. The computation of such probability was undertaken by the Marquis de Condorcet, by far the most sophisticated eighteenth-century social scientist.

Condorcet stands at the end of a significant French tradition, which, although there is a Cartesian ingredient, has a similar genealogy to that of the British. This is evident in Etienne Bonnot de Condillac, the most systematic French writer. He too believed that moral reasoning was susceptible to the same exactitude as geometry, and also identified the association of ideas as the key to the study of human nature. Condillac sought to trace all 'ideas' to sensation thus rendering Locke's 'ideas of reflection' derivative. In his *Traité des sensations (Treatise of Sensations)* of 1754 he traced the development of a statue, such that by the addition of each of the senses in turn it came to possess the full range of human mental ability.

Any account of human nature seemingly has to distinguish the nature of humans from that of animals. Georges, Comte de Buffon, who wrote the fullest scientific natural history (1749-89) in the eighteenth century, thought

humans resembled animals materially but differed by virtue of possessing a soul, reason and language. It was a common task of the science of human nature to account for language 'naturally' by tracing its evolution from onomatopoeic interjections to grammatical abstractions, without recourse to divine interposition or rational direction. While Buffon retained a traditional account of human differentia, Julien Offroy de La Mettrie argued that if animals were machines then, contrary to Descartes, there was no need to withhold that designation from humans. More explicitly Newtonian, David Hartley attempted to provide a physiological foundation for the association of ideas.

Leaving aside Hartley's own particular, ultimately religious, agenda, there is no reason to think Hume would find any incompatibility between that attempt and his own ambitions for a science of human nature; and it is those same ambitions that lie at the roots of contemporary social science.

See also: Enlightenment, Continental; Enlightenment, Scottish; Human nature

CHRISTOPHER J. BERRY

References and further reading

Aarsleff, H. (1974) 'The Tradition of Condillac', in H. Hymes (ed.) *Studies in the History of Linguistics*, Bloomington, IN: Indiana University Press, 93-156.(An excellent survey of eighteenth-century speculation on language.)

Baker, K.M. (1975) *Condorcet: From Natural Philosophy to Social Mathematics*, Chicago, IL: University of Chicago Press.(Not only the best book on Condorcet but one of the richest on Enlightenment thought in general and social science in particular.)

Beccaria, C. (1764) *On Crimes and Punishment*, trans. H. Paolucci, Indianapolis, IN: Library of Liberal Arts, 1963.(Most influential work on penology in the eighteenth century, especially for its attack on the death penalty and torture.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. W. Harrison, Oxford: Blackwell, 1948.(His first attempt at systematizing and applying the 'principle of utility' to the end of social reform.)

Berry, C.J. (1982) *Hume, Hegel and Human Nature*, The Hague: Nijhoff.(Chapter 1 & Part II discuss respectively the account of human nature propounded by the Enlightenment and by Hume.)

Buffon, G.L. de (1749-89) *Natural History: General and Particular*, trans. W. Smellie, London, 1812, 20 vols.(A compendium of data with the emphasis on description which is critical of Linnaean classification.)

Condillac, E.B. de (1740) *Essai sur l'origine des connoissances humaines (Essay on the Origin of Human Knowledge)*, in G. Le Roy (ed.) *Oeuvres Philosophiques de Condillac*, Paris: Presses Universitaires de France, 1947.(The first systematic French development of Locke's empiricism.)

Condillac, E.B. de (1754) *Traité des Sensations (Treatise on Sensations)*, in G. Le Roy (ed.) *Oeuvres philosophiques*, Paris: Presses Universitaires de France, 1947. (A more rigorous advance on the argument of the Essay and expounds a thoroughgoing sensationalist psychology.)

Gay, P. (1967-70) *The Enlightenment*, London: Weidenfeld & Nicolson, 2 vols.(The most generally informative and readable treatment of the Enlightenment, the second volume deals with issues closest to this article; both volumes contain lengthy bibliographic essays.)

Hartley, D. (1749) *Observations on Man, His Frame, His Duty and His Expectations*, London: Leake & Frederick, 2 vols.(Expounds a physiological theory of the mind with changes in the brain in form of vibrations corresponding to changes in ideas.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised P.H. Nidditch, Oxford: Clarendon Press, 1978.(One of philosophy's great books that Hume himself later dismissed in preference for his subsequent writing.)

Hume, D. (1740) *An Abstract of a book lately published entitled A Treatise of Human Nature*, repr. in *An Inquiry concerning Human Understanding*, 1748, ed. C. Hendel, Indianapolis, IL: Library of Liberal Arts, 1955. (Hume's own anonymous review of the *Treatise*.)

Hutcheson, F. (1725) *An Inquiry into the Original of our Ideas of Beauty and Virtue: in Two Treatises*, London and Dublin.(Contains his first formulation of the philosophy of moral sense.)

Knight, I.F. (1968) *The Geometric Spirit: The Abbé Condillac and the French Enlightenment*, New Haven, CN: Yale University Press.(A good exposition of Condillac's thought and the fullest available in English.)

La Mettrie, J.O. de (1758) *Man and Machine*, trans. and ed. G. Bussey, La Salle, IL: Open Court Press, 1912. (An uncompromising enunciation of materialism.)

Newton, I. (1730) *Opticks*, 4th edn, in H. Thayer (ed.) *Newton's Philosophy of Nature*, New York: Hafner Library of Classics, 1953.(Contains selections of the *Opticks*, which along with his *Principia Mathematica* (1687) established Newton's reputation, especially once experiments supported his calculations rather than Descartes.)

Humanism

The philosophical term 'humanism' refers to a series of interrelated concepts about the nature, defining characteristics, powers, education and values of human persons. In one sense humanism is a coherent and recognizable philosophical system that advances substantive ontological, epistemological, anthropological, educational, aesthetic, ethical and political claims. In another sense humanism is understood more as a method and a series of loosely connected questions about the nature and character of human persons.

From the fourteenth century to the end of the nineteenth century, humanism minimally meant: (1) an educational programme founded on the classical authors and concentrating on the study of grammar, rhetoric, history, poetry and moral philosophy; (2) a commitment to the perspective, interests and centrality of human persons; (3) a belief in reason and autonomy as foundational aspects of human existence; (4) a belief that reason, scepticism and the scientific method are the only appropriate instruments for discovering truth and structuring the human community; (5) a belief that the foundations for ethics and society are to be found in autonomy and moral equality. From the end of the nineteenth century, humanism has been defined, in addition to the above, by the way in which particular aspects of core humanist belief such as human uniqueness, scientific method, reason and autonomy have been utilized in such philosophical systems as existentialism, Marxism and pragmatism.

1 The problem of humanism

Like other terms that have had parallel careers in philosophy, intellectual history and popular usage (for example, liberalism or idealism), the concept of humanism suffers from an overabundance of definitions such that when considering them together it is not uncommon to ask whether there is anyone who is not a humanist. One can, for instance, define humanism by historical period: classical humanism, Renaissance humanism, Enlightenment humanism. Or one can adopt a definition that links humanism with a particular philosopher or philosophical movement, such as Marxist humanism, Heideggerian humanism or existentialist humanism. Or again one might, like the 1928 *Oxford English Dictionary*, approach the task of defining humanism by its constituent themes or claims: humanism centred on humane virtues like benevolence, humanity - independent persons as opposed to persons conceived within a transcendent divine order - and humane learning as structured within the classical *studia humanitatis*.

In each of these definitional approaches there is a risk of producing an account either so general as to be unhelpful or so particular as to contradict other legitimate usages. Thus, because of the ambiguity and controversy surrounding such terms as 'Renaissance' and 'Enlightenment', it is doubtful whether it helps in defining humanism to speak of Renaissance or Enlightenment humanism. Similarly, to speak of humanism in Pope's sense as 'the proper study of mankind is man' is to bring everything - scientific, non-scientific, religious, anti-religious, for instance - within its orbit. Nor is it apparent that we make any significant gain by talking about Marxist or Christian humanism as if they shared any substantial resemblance. In short, to be helpful, a definition of humanism must be as much alive to what it excludes as to what it includes.

Such an overabundance of definitional signposts, many of them pointing in different directions, suggests that humanism is as much a series of interrelated definitional and methodological questions as of substantive questions about the plausibility of a particular set of philosophical doctrines. This means that perhaps the most coherent understanding of humanism can be found by thinking about it in terms of a series of concentric circles in which the innermost and smallest circle contains the most limited and unproblematic sense of the term, the next and larger circle contains a more comprehensive and contested sense, and the outer and largest circle contains the most recent and controversial usages of the concept (see Figure 1).

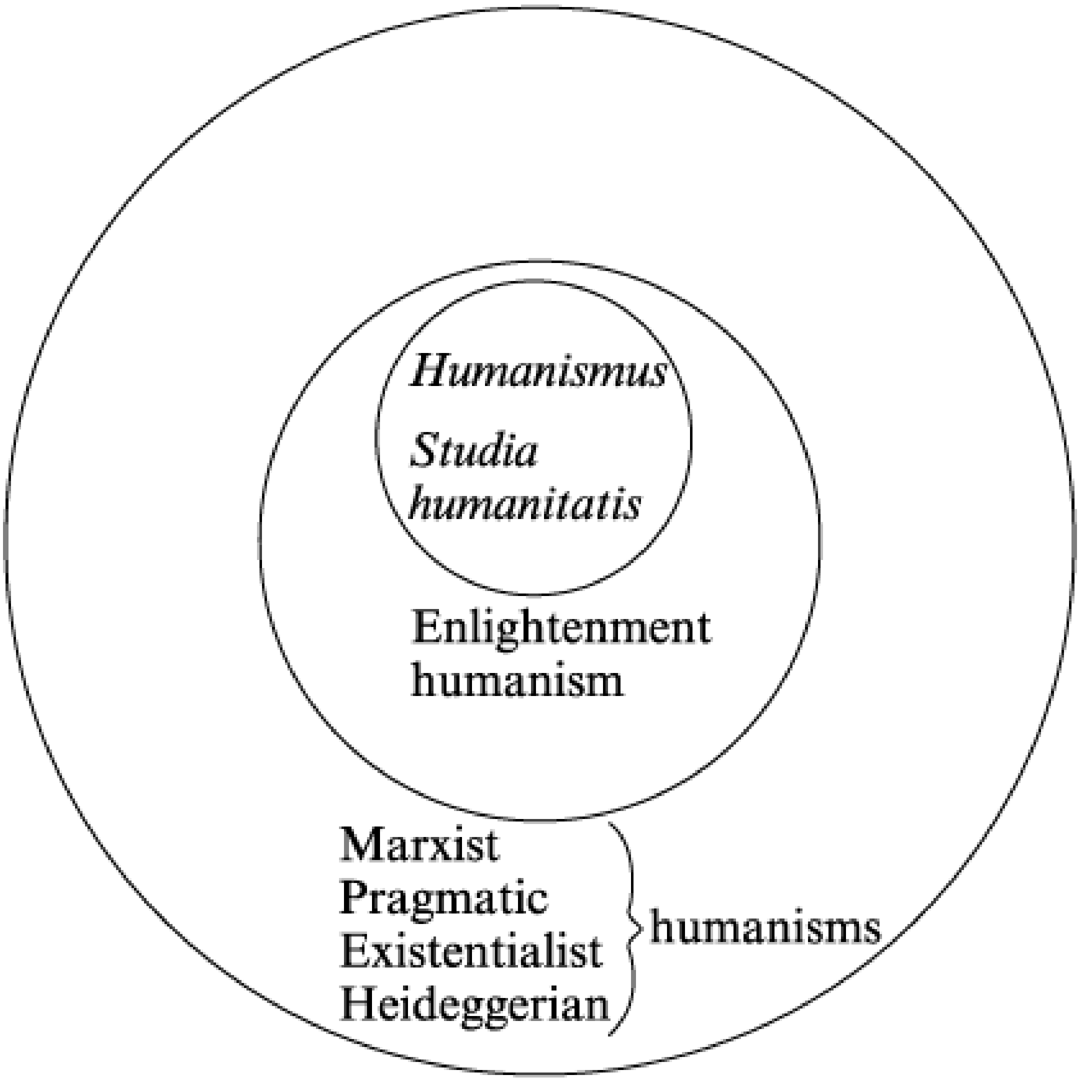

Figure 1

2 *Studia humanitatis*, *Humanista* and *Humanismus*

At the centre of the humanist circle are two definitions, one modern, the other originally ancient and subsequently revived in the Italian Renaissance, both of which define humanism in terms of an educational programme, *studia humanitatis*.

The word ‘Humanismus’ was created to differentiate the classical curriculum of secondary education with its emphasis on Greek and Latin from the modern obsession with scientific, mathematical and practical education. The word was then borrowed by nineteenth century historians (see George Voigt’s 1859 *The Revival of Classical Antiquity* or *The First Century of Humanism*) to describe the educational and cultural programme of a group of Renaissance writers, historians and philosophers, such as Petrarch, Valla, Ficino, and Pomponazzi.

Humanismus is thus derived from *humanista* or humanist, which along with its equivalents in English, French and Italian referred to teachers in schools or universities and students of the humanities. Paul Kristeller (1961: 9) suggested that ‘to judge from its earliest appearance known so far, [the word] seems to have originated in the student slang of the Italian universities, where the professor of the humanities came to be called *umanista*, after the analogy of his colleagues in the older disciplines’. The Renaissance term *Humanista* is itself derived from a more ancient usage *studia humanitatis*, the educational programme of the Greeks and Romans that centred on grammar, rhetoric, history, poetry and moral philosophy. It is important to note that the *studia humanitatis* did not include logic, natural philosophy, metaphysics or mathematics. *Humanismus*, *humanista* and *studia humanitatis* refer not

to some philosophical doctrine but to a broad cultural and educational programme that focused, through close study of the ancients, on the development of elegant writing and speaking.

3 Enlightenment humanism

In the second and wider circle one finds Enlightenment humanism which is closest to the sense that commonly attaches today to the terms ‘secular’ or ‘scientific’ humanism. By Enlightenment humanism one means the men of letters, philosophy and science (for instance, Voltaire, Rousseau, Montesquieu, Diderot, D’Alembert, Locke, Hume, Condorcet and, more problematically, Kant) who lived in the seventeenth and eighteenth centuries and believed that the central concern of human existence was not the discovery of God’s will, but the shaping of human life and society according to reason. For these thinkers, human dignity was not a function of man’s allegedly divine origin, but of the ordering and rational possibilities of earthly existence. The end of persons is neither the immediate adoration of God nor the heavenly city of the blessed, but rather the realization of those projects appropriate to this world, suggested by both reason and imagination.

What drives the project of Enlightenment humanism is the attempt to replace the traditional belief that the concept of a human person makes sense only within the context of talk about created souls, a transcendent, divine order and faith, with the conviction that the idea of persons is correctly to be understood within the context of reason, freedom for self-creation and fundamental scepticism. Enlightenment humanism is at foundation the re-defining of the nature of persons within a conceptual framework of their own making rather than one imposed from the outside. In this sense to speak of Christian humanism seems to be a fundamental contradiction inasmuch as it assumes a prior ontology of divinity within and against which the essence of human persons must be situated. For the Enlightenment humanist there is no such framework, no such transcendent God, no such pre-ordered account of the good life of individual persons or human societies.

To this general focus on the world of self-creating persons must be added a second feature of Enlightenment humanism, namely its conception of reason. Though immensely diverse in both interests and intellectual programme, in substance and nuance, what unites the Enlightenment humanists in a way quite different from either Renaissance humanists or subsequent modern humanism is their enormous confidence in the power of human reason to describe and order the world, and in their ready acceptance of responsibility for the consequences of reason’s ordering.

This confidence and its consequences are frequently caricatured. Enlightenment reason and Enlightenment humanism are often portrayed as cold, passionless, abstract and inappropriately systematic. But this is to mistake the mask for the face. At its core what drives reason for the Enlightenment humanist is passion, a passion for exposing error, for facts based on sound observation and for the coherent ordering of those facts in a fashion that would produce not so much a system as genuine insight into the character of existence even if the product of such insight were not so much illumination as perplexity. As d’Alembert put it in his summation of the eighteenth century ‘ everything has been discussed and analysed, or at least mentioned. The fruit or sequel of this general effervescence of minds has been to cast new light on some matters and new shadows on others’ (1759 (4): 5-6).

Reason, for the Enlightenment humanist, is driven by criticism and scepticism, and if we are to find the most consistent feature of Enlightenment reason it would be its insistence that all claims are falsifiable, and that no one has a privileged position in determining truth by virtue of their authority. For the humanist, sceptical reason thus ensures that all knowledge claims are publicly tested and that no person or institution can claim infallibility. For the Enlightenment humanist the importance of science lies primarily not in its advances in knowledge but in the energy, force and universality of its method.

Reason is central for the Enlightenment humanist not merely because of its critical and sceptical force but equally because of its objective character. Unlike revelation and dogma, which links truth to a particular authoritative person, reason for the humanist acquires its power precisely because the persons advancing knowledge claims are in a crucial sense interchangeable. Reason has the capacity, the humanist argues, to produce a truth that is quite literally subject-free: its claims hold only if they are accessible and make sense to all. Reason in this sense is genuinely emancipatory in at least three ways. First, it liberates individuals from what Kant termed the tutelage of others, the imposed dictates of external authority. Second, it opens up the process of truth to everyone, for truth is not a function of position or perspective. Third, it has the capacity to liberate persons from their own subjectively

imposed prejudices by forcing them to abandon the inadequacy of their merely private viewpoints for the widest, most objective perspective. Reason is thus an instrument of emancipation - it produces not just truth but also autonomy and equality. And it is this connection with autonomy and equality that provides the foundation for the moral and political content of Enlightenment humanism.

As with the Enlightenment humanists' account of reason, so with their conceptions of ethics and politics there is a strong temptation to reduce an enormous diversity and complexity to the simplicity of caricature. Thus some accounts of Enlightenment humanism find within its central tenets sweeping accounts of historical progress, grand schemes for educational reform and radical changes for political institutions. While there are elements of such beliefs and programmes in certain humanists of this period, the general content of humanist ethical and political thought is more restrained.

What emerges most clearly from the humanist emphasis on autonomy and equality is the open, tolerant and diverse character of humanist society. If reason is genuinely the gift of all and if criticism and scepticism and the public search for truth are the task of all, then the number of absolute truths that will structure the social life of society is very small and the circle of those competent to participate actively in the life of running society is very large - humanist reason serves to decrease the amount of societal certainty while increasing the number of qualified members of a political society. Questions of the good personal and collective life are by definition questions that people will solve, as autonomous beings, in different and (in some instances) competing ways. Enlightenment humanism comes closest to agreement in its ethical and political maxims in its commitment to the discipline of reason in which truth must emerge, both publicly and privately, not through dogma but through argument and counter-argument, and in its commitment to the virtues of beneficence and non-malfeasance. In this sense it is not incorrect to claim that much of what today is described as a liberal society - basic liberties, free institutions, diversity and tolerance, the sense, to use Bertrand Russell's phrase, of 'order without authority' - can plausibly be traced to the Enlightenment humanist's conception of reason.

Enlightenment humanism can be approached from any number of directions to come to terms with its intellectual and emotional core: through its compassion for humanity and social reform, through its interest in material and intellectual progress, through its championing of literary freedom, its interest in the classical past, or its belief in the centrality of science and scientific method. At the end, however, its defining characteristic is its confidence in and celebration of human autonomy, an autonomy secure through the critical and creative powers of reason.

4 Modern humanism

There are at least four modern versions of humanism, three of which claim some affinity with the themes of Enlightenment humanism, and one which explicitly disavows any connection with it. These four, Marxist humanism, pragmatic humanism, existentialist humanism and Heideggerian humanism, do not exhaust the humanisms of the modern age, for in popular usage one will frequently find reference to scientific humanism, Christian humanism and secular humanism.

Of the four, pragmatic humanism is closest to the humanism of the Enlightenment (see Pragmatism). Rooted in the American philosophy of pragmatism as developed by Charles Peirce, William James, John Dewey and F.C.S. Schiller, pragmatic humanism shares the Enlightenment humanist concerns for human primacy, the validity of the scientific method and the incompatibility of human dignity with religious belief. While championing the humanist themes of dubitability and scepticism, pragmatic humanism is at odds with Enlightenment humanism on the belief that reason is an unchanging and defining essence of human nature and on the nature of human freedom. Most pragmatic humanists would extend their scepticism to the claim that there is a human essence, and certainly to the belief that reason occupies a privileged place in defining 'humanness'. These humanists would prefer to speak of persons being what they do, with the anthropological emphasis placed squarely on the changing ensemble of environmental conditions that create persons. Similarly, such pragmatists are uncomfortable with the confident Enlightenment talk of human freedom and autonomy, though they do speak of freedom in terms of variation, innovation and foresight.

Closely aligned with pragmatic humanism in its denial of a universal human essence is existentialist humanism, particularly as formulated by J.-P. Sartre in his lecture '*Existentialism is a Humanism*' (1946) (see Existentialism). Inasmuch as 'existence precedes essence', according to Sartre, there is no objective and permanent thing called

human nature. It is nothing else than what persons do, how they choose to act. There is no paradigm of human against which action can be measured. But while concurring with the pragmatic humanists with respect to the essence of human persons, and registering, unlike Enlightenment humanists, a fair degree of scepticism about the capabilities of reason itself, existentialist humanism shares unreservedly in the Enlightenment humanist's celebration of human autonomy. While existentialists like Sartre would allow, at one level, that natural explanations of human actions are appropriate, at another level they would insist that human actions, individual and collective, must be subject to another type of explanation - the explanation of freedom or self-causation. Of the first two types of modern humanism, each has accepted but part of the Enlightenment humanist programme. Whereas pragmatic humanism concurs with the Enlightenment's celebration of human reason but rejects its emphasis on human freedom, existentialist humanism makes human freedom central to its affirmation of humanness but doubts the Enlightenment's emphasis on the primacy of reason.

Any discussion of Marxist humanism labours under the considerable disadvantage of having to come to terms with a complicated variety of Marxes and Marxism (see Marx, K.). Despite this, the general outline of Marxist humanism seems relatively clear with respect to the paradigm of Enlightenment humanism. Although Marx often used the word humanism during 1843-4, it occurs very rarely in his post-1845 works. The reason for this is that Marx came to three insights that subsequently led to his rejection of much of Enlightenment humanism: first, that the development of history is really the constant creation of human nature; second, that if we can speak of a real essence of persons it can be nothing more than the complete ensemble of an individual's social relations; and third, that humanism in the traditional sense is simply another ideology that masks the true nature of social relations by conferring a spurious legitimacy on the status quo. Thus, while accepting Enlightenment humanists' rejection of theism, and their instrumental rationality, Marxist humanism is not prepared to accept their essentialist reading of human nature.

Equally problematic for the Marxist humanist is the Enlightenment's emphasis on individual autonomy, for there is a strong deterministic, or at least inevitabilist element within much Marxist thought. Perhaps the most significant contribution of Marxist humanism, however, is its attempt to give concrete definition to the social ideals that are often only given the vaguest mention in Enlightenment humanism. Apart from its claims about the ideological character of humanism, the unique character of Marxist humanism lies in its practical attention to the Enlightenment's ideas of autonomy and equality.

The other variety of modern humanism is that of Martin Heidegger in his famous *Letter on Humanism* (1947). The letter is a wide-ranging analysis of the humanisms of the past, particularly those of Marx and Sartre. Heidegger's thesis is that traditional humanism (including Enlightenment humanism) in what he calls metaphysical humanism, mistakenly locates human essence in the rational animality of persons, and thus fails to recognize that the genuine source of human essence is in existence, in the human's primordial relationship with Being. Heidegger's rejection of traditional humanism is not based simply on the fact that such rooting of human essence is wrong but that it necessarily corrupts humanism by investing it with a metaphysical character. The goal is thus to 'think the truth of Being at the same time… to think the humanity of *homo humanus*. What counts is *humanitas* in the service of the truth of Being, built without humanism in the metaphysical sense' ([1947] 1977: 21). Heideggerian humanism thus explicitly rejects metaphysical humanism, yet retains both reason and thinking as central to its account of persons and Being. At the same time, however, as this rejection of metaphysical humanism occurs, there is a residual concern for 'meditating and caring, that man be human and not inhumane' ([1947] 1977: 200), in effect an ethical concern, that is surprisingly consistent not only with Enlightenment and every other type of metaphysical humanism, but with every other modern humanism as well.

See also: Humanism, Renaissance; Enlightenment, Continental

JOHN C. LUIK

References and further reading

While many of these works require both some knowledge of the history of philosophy and the development of intellectual history and a facility for often difficult argument, they do not require any technical knowledge.

Cassirer, E. (1951) *The Philosophy of the Enlightenment*, Boston, MA: Beacon Press.(A philosophically sophisticated and at times difficult work which centres on the coherence of Enlightenment humanism.)

D'Alembert, J. (1759) 'Élements de Philosophie', in *Mélanges de littérature, d'historie et de philosophie*, novelle

édition, Amsterdam, 6 vols.(One of the best introductions to what might be termed the ‘mind of the mid-eighteenth century’. Particularly useful in that it provides an analysis of the development of Enlightenment thought in the context of the three hundred years preceding the eighteenth century and the critical perspective of the author’s own philosophy.)

Gay, P. (1959) *The Party of Humanity*, New York: Norton.(An elegant discussion of the Enlightenment concept of humanism described in §3 above.)

Gay, P. (1966) *The Enlightenment: An Interpretation*, New York: Vintage Books, 1969.(One of the best accounts of the development, scope and logic of Enlightenment humanism in all of its dimensions.)

Goicoechea, D., Luik, J. and Madigan T. (eds) (1991) *The Question of Humanism: Challenges and Possibilities*, Buffalo, NY: Prometheus Books.(A comprehensive analysis, by a collection of authors, of the career of humanism from the ancient world to the twentieth century.)

Heidegger, M. (1947) *Letter on Humanism*, trans. J.G. Grey and F.A. Capuzzi, New York: Harper & Row, 1977.(Like much of Heidegger, a very dense work that demands some understanding both of the development of humanism and of Heidegger’s general philosophy.)

Kant, I. (1784) *What is Enlightenment?*, trans. L. White Beck, New York: Bobbs-Merrill, 1963.(A classic account of the tenets of an Enlightenment humanist.)

Kristeller, P.O. (1961) *Renaissance Thought: The Classic, Scholastic and Humanist Strains*, New York: Harper & Row.(Examines the humanism discussed in §2 above.)

Kristeller, P.O. (1965) *Renaissance Thought II: Papers on Humanism and the Arts*, New York: Harper & Row.(As with the previous work, examines the humanism discussed above in §2.)

Manuel, F. and Manuel, F. (1979) *Utopian Thought in the Western World*, Boston, MA: Harvard University Press.(Humanism within the context of the larger issues of utopianism and human perfectability.)

Marx, K. (1932) *Economic and Philosophical Manuscripts*, trans. R. Livingstone and G. Benton, Harmondsworth: Penguin, 1975.(A difficult discussion of the basis of Marx’s humanism.)

Passmore, J. (1970) *The Perfectibility of Man*, New York: Scribner.(Humanism within the context of the larger issues of utopianism and human perfectibility.)

Petrosyan, M. (1972) *Humanism: Its Philosophical, Ethical and Sociological Aspects*, Moscow: Progress.(A contemporary explication and defence of Marxist humanism which presupposes a fair understanding of Marxist texts and interpretation.)

Sartre, J.-P. (1946) *L’Existentialisme est un humanisme*, Paris: Nagel; trans. B. Frechtmen as *Existentialism*, New York: Philosophical Library, 1947.(A key text of existentialist humanism but one which requires some persistence.)

Voigt, G. (1859) *The Revival of Classical Antiquity* or *The First Century of Humanism*; repr. in D. Goicoechea, J. Luik and T. Madigan (eds) *The Question of Humanism: Challenges and Possibilities*, Buffalo, NY: Prometheus, 1991.(Mentioned above in §2.)

Humanism, Renaissance

The early nineteenth-century German educator, F.J. Niethammer, coined the word 'humanism', meaning an education based on the Greek and Latin classics. The Renaissance (for our purposes, Europe from about 1350 to about 1650) knew no such term. The Renaissance had, instead, the Latin phrase studia humanitatis *(literally 'the studies of humanity'), best translated 'the humanities'. The Renaissance borrowed the phrase from classical antiquity. Cicero used it a few times, but it was the later grammarian Aulus Gellius who clearly equated the Latin word* humanitas *with Greek* paideia, *that is, with the classical Greek education of liberal learning, especially literature and rhetoric, which was believed to develop the intellectual, moral and aesthetic capacities of a child (*pais *in Greek; hence* paideia*).*

Renaissance humanists understood by studia humanitatis *a cycle of five subjects: grammar, rhetoric, poetry, history and moral philosophy, all based on the Greek and Latin classics. A humanist was an expert in the* studia humanitatis*. The dominant discipline was rhetoric. Eloquence was the highest professional accomplishment of the Renaissance humanists, and rhetorical interests coloured humanists' approach to the other parts of the* studia humanitatis*. The Renaissance humanists were the successors of the medieval rhetorical tradition and the resuscitators of the classical rhetorical tradition. Renaissance humanism was, in the words of P.O. Kristeller, 'a characteristic phase in what may be called the rhetorical tradition in Western culture' ([1955] 1961: 11).*

Renaissance humanism was neither a philosophy nor an ideology. It reflected no fixed position towards religion, the state, or society. Rather it was a cultural movement centred on rhetoric, literature and history. Its leading protagonists held jobs primarily as teachers of grammar and literature. Outside academia, they served as secretaries, ambassadors and bureaucrats. Some were jurists. The Renaissance humanists reasserted the importance of the humanities against the overwhelming dominance of philosophy and science in medieval higher education. As humanism penetrated the wider culture, it was combined with other disciplinary interests and professions so that one found humanist philosophers, physicians, theologians, lawyers, mathematicians and so forth.

Ideologically humanists were a varied lot. Some were pious, some were not. Some were interested in philosophy, most were not. Some became Protestants, others remained Catholic. Some scorned the vernacular while others made important contributions to it. Humanism influenced virtually every aspect of high culture in the West during the Renaissance. Depending on the humanist under discussion, one can legitimately speak of Christian humanism, lay humanism, civic humanism, Aristotelian humanism and other combinations.

Humanism had a profound effect on philosophy. Writing outside the philosophical establishment, humanists sought to make philosophy more literary in presentation and more amenable to rhetorical concerns. No less importantly, they recovered and translated into Latin a large reservoir of Greek classical texts unknown or ignored in the Middle Ages. Platonism, Stoicism, Epicureanism and scepticism all experienced revivals. The humanists challenged medieval Aristotelianism by offering new Latin translations of Aristotle that in some respects amounted to fresh interpretations. They also significantly enriched the Aristotelian corpus by translating the Poetics and the late ancient Greek commentators on Aristotle.

Renaissance humanism arose out of the peculiar social and cultural circumstances of thirteenth-century Italy. It came to maturity in Italy in the fifteenth century and spread to the rest of Europe in the sixteenth. It gradually lost its vitality in the seventeenth and eighteenth centuries as its focus on Latin eloquence became out of date in a world increasingly won over to the vernacular literatures and new science. In the nineteenth century, it did not so much die as become metamorphosed. Renaissance humanism sloughed off its rhetorical impulse and became modern scholarly classicism. Today the word humanism has taken on new connotations, but the heritage of Renaissance humanism runs deep in our culture. As long as we continue to value literature and history, and the functional skills and cultural perspective attached to these disciplines, every educated person by training will be a humanist in the Renaissance sense.

1 Origins

Education reflects the culture that it serves. In antiquity, it was not philosophers and scientists but rhetoricians who

provided the most common form of higher education. Literary-rhetorical training sufficed to qualify one as an educated person. In the cultural disasters that marked the passing of antiquity, monasteries became for a while the main depositories of learning in the West. But in the great medieval resurgence of the eleventh and twelfth centuries, cathedral schools in northern and central France emerged as centres for the education of the clergy. They were, in modern terminology, seminaries. These schools taught the seven liberal arts, a curriculum in which the core subject was grammar and the main reading the classical Latin authors. Thus, at one point in the Middle Ages literary studies constituted, as they had once before in antiquity, the basis of professional education.

Towards the end of the twelfth century, however, the medievals began to find logic and the newly translated Aristotelian corpus, especially the *Physics*, far more valuable (see Aristotelianism, medieval). By the early thirteenth century, a new institution had emerged in northern Europe, the university, where the fundamental faculty was that of arts, and where arts meant the study of Aristotelian logic, philosophy and science. Arts graduates, having received this scientifico-philosophic education, could then proceed to one of the three professional faculties: law, medicine or theology. At no time in Western history have science and philosophy dominated educated culture to the degree they did in the later Middle Ages.

Italy differed from the rest of Europe. North of Rome it was a land not of agrarian feudal principalities, but of city states. Urban interests and urban culture dominated, and the universities differed accordingly from their counterparts elsewhere. Italy’s first university, Bologna, began at the end of the twelfth century as a law school; an arts faculty dominated by the medical faculty quickly followed. The Bolognese two-faculty university became the standard form of Italian universities: in medieval Italian universities the two secular professions of law, with its own faculty, and medicine, within the arts faculty, achieved hegemonic status.

One subject closely connected with law but found in the arts faculty was the *ars dictaminis*, the art of letter writing. This *ars* had emerged in the eleventh century to answer a functional need of the new burgeoning urban society of Italy. The manuals provided instruction on how to write different kinds of letters to different classes of people, along with a generous dose of sample letters, salutations and formulas. The practitioners and teachers of the *ars dictaminis* were the *dictatores*. Their business was eloquence and persuasion. They were medieval rhetoricians, hired by chanceries all over Italy. By the early thirteenth century, *dictatores* were teaching in the arts faculty of the University of Bologna.

Medieval Italy also created another practical form of secular rhetoric, the *ars arengandi*, the art of haranguing. Public speeches played a role not only in politics, but also in a variety of social functions, such as weddings and the awarding of a university degree. In the thirteenth century, Bolognese *dictatores* began to produce model speeches for public occasions; the first extant *artes arengandi* date from the end of the century. Medieval Italy produced a rich literature of *arenge* (secular speeches in Latin) and *dicerie* (secular speeches in Italian). The *ars dictaminis* spread to the rest of Europe, but the *ars arengandi* remained unique to Italy, and in combination with the *ars dictaminis* set medieval Italy apart as the home of a lively culture of secular oratory and rhetoric.

During the thirteenth century, *dictamen* entered into the training of notaries by being linked to the *ars notaria*. The *ars notaria* focused on document writing. Even more than lawyers, notaries gained fame, especially through their skill in composition. Notaries filled the Italian chanceries and the most skilled of them rose to be chancellors of Italian governments and secretaries of popes.

Thus, by the thirteenth century, Italy had a large class of lay professionals associated with law who took a special interest in eloquence and rhetoric. Not by chance, notaries, lawyers and bureaucrats loom inordinately large in the development of medieval Italian literature. The first school of Italian poetry arose in the chancery of Emperor Frederick II (1215-50), one of whose leaders was Giacomo ‘the Notary’. Two of the most eminent Italian poets before Dante were Guittone of Arezzo, chancellor of his home town, and the Bolognese judge Guido Giunicelli. The earliest extant examples of Italian literary prose are the model speeches found in a work of the late thirteenth-century *dictator* Guido Faba. Of the three great fourteenth-century Florentine authors, Dante was a nobleman enrolled in the apothecary guild (see Alighieri, Dante), but Petrarch (see Petrarca, F.) was the son of a notary and was himself trained in law at Bologna, while Boccaccio actually took a law degree.

In the thirteenth century, the Italian rhetorical tradition began to absorb the classicism of the dying cathedral school culture of France. The wellsprings of Latin eloquence, it was believed, lay in the Latin classics.

Commentaries on classical rhetoric that were no longer being copied in France, began to be copied in Italy in the thirteenth century. By the fourteenth century, Italians were writing their own commentaries. *Dictatores*, such Giovanni del Virgilio at Bologna, a correspondent of Dante, began to lecture on classical authors at the universities. Late in the century, the Paduan judge Lovato de'Lovati studied the metre of Seneca and gathered about himself a circle of lawyers interested in classical studies. A younger contemporary, the notary Albertino Mussato became chancellor of Padua and gained eminence because of his Latin histories and epic poem *Ecernis*. In 1315 the Paduans crowned him poet laureate.

This classicization of the medieval Italian rhetorical tradition produced Renaissance humanism. The person who made the movement self-conscious was the Florentine expatriate Petrarch (1304-74).

2 Francis Petrarch

Read today by the general public almost exclusively for his Italian poetry, Petrarch was the greatest classical scholar and Latinist of his time. He was also a cultural icon. Popes at Avignon tried to make him a papal secretary. The republic of Venice furnished him with a free house in the vain hope of acquiring his library. The despots of Parma, Padua and Milan gave him hospitality, and for these princely patrons he several times indited letters and wrote orations in the tradition of the medieval *dictatores*. He had himself crowned poet laureate at Rome in 1337, but not before having the King of Naples 'test' him and give him a diploma (probably written by Petrarch himself) qualifying him as a poet and a historian.

Petrarch promoted the ideal of classical eloquence. The embodiment of that eloquence, he believed, was Cicero. In 1333 Petrarch inaugurated the Renaissance's reappropriation of classical oratory with the discovery at Liège of Cicero's oration *Pro Archia*. In 1345 at Verona he discovered Cicero's *Letters to Atticus*. Petrarch's interests extended to Greek literature. He acquired a Greek codex of Plato's dialogues and another of Homer, though he never learned to read Greek.

Imitating his classical models, Petrarch composed moral treatises and biographies of classical heroes, wrote an introspective dialogue (entitled *The Secret*), produced several collections of his Latin letters, drafted and delivered orations, gained fame as a Latin poet, exposed medieval forgeries of classical documents, and worked on correcting the corrupted texts of classical authors. Petrarch shared the medieval admiration for St Augustine, but reflecting his new cultural perspective, the Augustine he admired was not the theologian but the psychologist and doctor of souls he found in the *Confessiones (Confessions)*.

The contrast between Petrarch's humanist perspective and the philosophico-scientific culture regnant in the universities emerges with special clarity in his invective *De Sui Ipsius et Multorum Ignorantia (On His Own Ignorance and the Ignorance of the Many)* (completed 1371). About 1366, already a cultural icon, Petrarch met four young Venetians who decided, it later transpired, that Petrarch was a good man, but ignorant. If one grants their premise, that learning is defined by the scientific and philosophical interests of the medieval university, then, despite Petrarch's classical learning and literary greatness, the young men were right. In their naïvety, they had accurately drawn the line between humanism and the medieval scientific culture. Petrarch's defence was pretty weak. Apart from using Socratic ignorance as a rhetorical stratagem ('at least I know that I am ignorant'), what he had to say was more a programme than an argument. He denigrated the value of scholastic learning, noted Aristotle's less than dominant position in antiquity, expressed his preference for Plato as being more in harmony with Christianity, and portrayed himself as a true Christian in opposition to the contemporary Aristotelians who worshipped the pagan Aristotle and his Arab commentator Averroes. Petrarch did not even have a name for the outlook he espoused. His successors would remedy this failing and make good his programme (see Petrarca, F.).

3 A name and an image

In the first half of the fifteenth century the new movement had established itself in all the major cultural centres of Italy. Its image-conscious exponents came up with an attractive name for it: the *studia humanitatis* ('the humanities'), and gave its products a new appearance: roman and italic script.

Coluccio Salutati, an Italian notary and a great admirer of Petrarch, was using the phrase *studia humanitatis* as early as the 1360s. Salutati is important because he eventually became the chancellor of Florence and gathered about himself a brilliant circle of younger humanists. The most illustrious were the notaries Leonardo Bruni and

Poggio Bracciolini, who became in turn not only famous humanists, but also chancellors of Florence. Salutati did not identify a fixed set of subjects as constituting ‘the studies of humanity’, but by the 1440s Tommaso Parentucelli (the future Pope Nicholas V) could draw up for Cosimo de’Medici, the political leader of Florence, a plan for an ideal library collection in which he defined the *studia humanitatis* as works of grammar, rhetoric, history, poetry and moral philosophy. These five subjects were in fact those which the humanists taught and on which they wrote with professional competence.

The word *humanista* originated as student slang at the Italian universities in the late fifteenth century as the name of teachers who taught the *studia humanitatis*. No component of the new humanistic programme was in itself new. Ethics was the one part of philosophy which had special relevance to students of literature and rhetoric (see Aristotelianism, Renaissance §7). In the case of history the humanists were innovative in that they were the first to make it an academic subject, not just a literary genre. (Significantly, Petrarch wished to have his diploma qualify him as a historian, not just a poet.) Furthermore, they gave grammar, rhetoric and poetry the quite specific meaning of classical grammar, rhetoric and poetry. The humanists despised medieval Latin. They strove to recover classical Latin and they studied classical Greek. They cultivated neoclassical Latin in their own writings and assiduously imitated classical literary forms. Ironically, a purist humanist would not use the word *humanista* because it was an unclassical neologism.

Along with the name, humanism also acquired a script. Poggio Bracciolini created the new hand in the early fifteenth century under the misconception that in rejecting the angular, densely abbreviated gothic script of the later Middle Ages, he was restoring antique script. He was in fact imitating not Roman manuscripts, but Carolingian manuscripts of the early Middle Ages. Nonetheless, Bracciolini called his creation *littera antiqua*, and we have gone along with his misnomer by calling his new hand ‘roman’. Italic seems to have been the brainchild of Bracciolini’s associate in the Salutati circle Niccolò Niccoli, who sought a cursive version of Bracciolini’s new humanist script. When at the end of the fifteenth century the internationally famous press of Aldo Manuzio began using this humanist cursive in its books, the new script gained the name by which it is known in many languages: ‘italic’. Humanism thus acquired a distinctive look to go along with its propaganda-laden name.

4 Language

Language was at the heart of the humanist movement (see Language, Renaissance philosophy of). Indeed, the distinguishing mark of the humanists was not any philosophy or ideology but their desire for eloquence based on a classicizing Latin. A humanist incapable of writing neoclassical Latin was a contradiction in terms. Humanists enthusiastically revived the classical prose genre of the oration and the letter. They imitated as best they could classical poetic forms such as the epic and the epigram. They composed neoclassical tragedies and comedies. And they wrote manuals of poetics, oratory, history and epistolography with classical models constantly in mind.

Petrarch complained about medieval Latin, but by the early fifteenth century even his Latin no longer met humanist standards. Leonardo Bruni became the new model of contemporary Latin scholarship and others would follow, such as the great philologist Angelo Poliziano (d. 1494) at the end of the century and the Christian humanist Desiderius Erasmus (d. 1536) in the first half of the sixteenth century. Since the humanists took over the teaching of Latin, they achieved near total victory over medieval Latin. Early in the sixteenth century, medieval Latin grammars ceased to be printed. Eventually the only Latin one could learn in the schools was the neoclassical Latin of the humanists. Medieval Latin survives today mainly in specialized vocabulary encountered outside the Latin class, for example, the phrase *a priori* in philosophy, which should be written *a priore* according to classical grammar.

It would be a mistake to think that the humanists collectively were hostile to the vernacular. Some were, but many (starting with Petrarch) wrote in the vernacular as well as in Latin. In the case of Thomas More, his Latin *Utopia* was a far more original work than any of his English writings (see Utopianism), whereas some humanists, such as Pietro Bembo (1470-1547), were actually instrumental in formulating the standards of vernacular literature. Michel de Montaigne, the greatest French author of the second half of the sixteenth century, even had Latin as his first language, because his father would only let him hear Latin until he was old enough to go to school. In general, however, the replacement of Latin by the vernacular as the standard language of literature and learning continued apace in the Renaissance. But even into the eighteenth century, Latin remained an important medium for literary and scientific expression.

Humanist literary values transferred not merely to the vernacular, but also to non-humanistic disciplines. In the Renaissance science and philosophy became more literary in presentation and increasingly vernacular in language. By the mid-seventeenth century, it was hardly scandalous that Galileo published his most substantial scientific works in the form of vernacular dialogues and that Descartes would describe his method of philosophy in a consciously limpid style of French. The *philosophes* of the Enlightenment are in this sense the heirs of the Renaissance humanists, and every academic discipline today presumes at least some minimum literary standard in expository writing.

Petrarch had intuited the importance of Greek for Latin literature. His successors put that intuition into effect. The first decisive moment came in 1397 when the city of Florence hired the Byzantine teacher and sometimes diplomat Manuel Chrysoloras to satisfy the desire of Coluccio Salutati and his circle to learn Greek. Some humanists, such as the famous educator Guarino Veronese, travelled in the other direction to learn Greek by living in Constantinople. Very rapidly in the fifteenth century it was established that excellence as a humanist demanded a thorough knowledge of Greek. Only a minority of humanists ever achieved this ideal, though all praised it. Oddly enough, when other Byzantine scholars followed Chrysoloras to Italy, such as George of Trebizond of Crete about 1416 and Theodore Gaza of Constantinople about 1439, they initially had little impact on Greek learning in Italy. Some of the scholars that arrived after the fall of Constantinople in 1453, especially Johannes Argyropoulos at Florence, did become important cultural forces, but these later Greeks did not shape what had already become a mature humanist movement. They simply served its interests.

5 Learning: the classical heritage

The impulse that drove the humanists to recover the classical languages also drove them to discover lost or unknown classical texts. The Middle Ages had experienced a transformatory infusion of classical texts, but nothing on the scale of the Renaissance revival. Petrarch's discoveries have already been mentioned, but the most spectacularly successful discoverer was Poggio Bracciolini, who in series of forays into German and French libraries during the Council of Constance (1414-7) unearthed a breathtaking number of classical Latin texts. The process of discovery was ongoing from the fourteenth century into the seventeenth century. By the mid-seventeenth century, the Renaissance had recovered the overwhelmingly larger part of the classical literary heritage available now, and in the process had transformed Western learning and literature.

In the case of Latin texts, the humanists essentially discovered manuscripts copied in the early Middle Ages and ignored thereafter. Among their more notable finds were the historians Tacitus and Ammianus Marcellinus as well as new decades of Livy, the poets Catullus and Lucretius, new comedies of Plautus, the architectural authority Vitruvius, the orations, letters and some of the major rhetorical works of Cicero, the complete text of Quintilian, the cookbook of Apicius, the *Astronomicon* of Manilius, the medical author Celsus Cornelius and ten books of the letters of Pliny the Younger.

As varied as these Latin finds were, they pale in comparison to the great mass and riches of the Greek classical heritage exposed by the humanists. Greek texts involved the further difficulty that they had be translated in addition to being found. By the mid-seventeenth century the humanists had put into Latin virtually every Greek text known to them. Some of these translations have remained the only one in any language up to today. Leonardo Bruni, the star pupil of Chrysoloras in Florence, led the way in the first decades of the fifteenth century with his very popular translations of Plato, Aristotle, Demosthenes, Plutarch and St Basil the Great (see Aristotelianism, Renaissance §7; Platonism, Renaissance §§1, 4). His work proved to be only the beginning of a huge expansion of translation from the Greek during the Renaissance.

Translating became, in fact, a major enterprise of the humanists. As classroom teachers they focused on the humanities, but as translators and classical scholars they took in the whole of the Christian as well as pagan classical heritage. They recovered and translated for the first time not only works by the Greek orators, dramatists, historians, poets, rhetoricians, literary critics (including the *Poetics* of Aristotle) and essayists, but also the Greek mathematicians, astronomers, geographers, philosophical writers (especially Plato, Plotinus, Sextus Empiricus and the commentators on Aristotle), medical authors and military tacticians. They retranslated virtually every Greek scientific and philosophical text known in the Middle Ages, including Aristotle. They also set about finding and translating the vast number of writings left by the Greek church fathers, retranslating the Bible itself and

publishing it in its original languages. The humanists created a patristic and biblical renaissance in addition to a pagan literary renaissance.

Before the West could surpass antiquity, it had to know and make available what the ancients knew and did. Bringing that about in literature, science, philosophy and theology was one of the great achievements of the Renaissance humanists (see Aristotelianism, Renaissance; Platonism, Renaissance).

6 Learning: philology

The humanists did much more than introduce classical texts: they came to understand them better. They could do so, first, because of their superior knowledge of classical history and languages. The first landmark work in this regard was Lorenzo Valla's *Elegantiae Linguae Latinae (The Fine Points of the Latin Language)* completed in the 1440s, where Valla taught the West to understand many of the words, idioms, phrases and grammatical constructions of classical Latin correctly(see Valla, Lorenzo §1). Second, the humanists began to become professionally concerned with the emendation of texts. The key event here was the printing press.

Before printing, the reproduction of a text meant copying. Any substantial copying invariably corrupted the text and, were copying to be repeated often enough or carelessly enough, such corruption could become serious indeed. The early humanists were acutely aware of the need to emend the classical texts they cherished. Petrarch and Valla, for instance, attempted to emend the text of Livy. In the 1440s at Naples, Valla even became involved in a quarrel with other humanists concerning the emendation of Livy.

The introduction of printing into Italy in the late 1460s transformed this situation. Whereas every manuscript copy of a text was different from every other, every copy of a printed edition was at least theoretically identical with every other. Humanists quickly became editors and correctors of classical texts for the new presses. They also began to argue furiously about the principles and practice of textual emendation. One humanist, Niccolò Perotti, at Rome in 1470, even urged the pope to censure the printing of books not for the purposes of religion but for the purpose of maintaining quality control in the reproduction of classical texts.

Humanists had found a new professional outlet; they had become philologists, the editors and correctors of texts. The most sophisticated of them early on was Angelo Poliziano (1454-94), whose *Miscellanea* summed up many of the principles developed by previous humanists (for example, going back to the Greek models of Latin texts, understanding the linguistic habits of an author, and seeking out the oldest manuscripts) and added some new ones (including the need to understand the genealogy of a manuscript and historical sources, and the need to cite sources precisely).

Lorenzo Valla provided a prime example of how this emerging philological sophistication could be put to use for more than purely grammatical purposes. In his clamorous *Oratio de falso credita et ementita Constantini Donatione (Oration on the Falsely Believed and Fabricated Donation of Constantine)* of 1440, he investigated the authenticity of a document that had been cited for centuries to prove that the Emperor Constantine gave the whole western half of the empire to the papacy. Using linguistic, historical and numismatic evidence Valla proved that it was in fact a medieval forgery. This sort of historical and philological knowledge had a wide application. In law it meant a reinterpretation of Justinian's *Corpus Iuris Civilis*, which had served as the basic text of secular law in the universities. The medieval commentators had frequently made the *Corpus* applicable to the new society of medieval Europe by interpreting it anachronistically. The new historical methods of the humanists totally undermined this medieval approach. Ironically, the humanist historical understanding of the *Corpus* failed to take hold in Italy, where the traditionalists, called the Bartolists (after the famous medieval jurist Bartolo of Sassoferrato), were able to retain control of the law faculties. Instead the Italian historico-philological approach to law received the name *mos Gallicus* ('the French custom') because it was the sixteenth-century French jurists who appropriated it and took the lead in applying it.

The new humanist philology had an even greater impact on religion. The Italian humanists translated many Greek church fathers, especially under the influence of Pope Nicholas V (1447-55), formerly Tommaso Parentucelli (see §3). The Florentine businessman turned humanist Gianozzo Manetti (d. 1459) made a fresh translation of the New Testament from the Greek and the Psalms from the Hebrew. In the 1440s, Lorenzo Valla prepared a comparison of the Greek text of the New Testament with the current Latin translation, called the Vulgate, and found the latter faulty in hundreds of places. Manetti's and Valla's work had little resonance in Italy because as a group Italian

humanists took no professional interest in theology. In the sixteenth century, however, the discovery of Valla's *Annotationes in Novum Testamentum (Annotations on the New Testament)* inspired the humanist religious reformer Desiderius Erasmus (§6) of Rotterdam to become a scripture scholar, and in 1516 he published the first printed Greek edition of the New Testament. He also made a new Latin translation and wrote a commentary where he drew theological conclusions from his philological and historical understanding of the Greek text. Humanist philology when applied to religion had resulted in scriptural theology. Humanists now clearly had a professional role to play in theology alongside the traditional doctrinal theologists.

Erasmus' timing could not have been better. The year after his edition, the appearance of Luther's Ninety-Five Theses ushered in the Reformation. Protestants made Scripture the sole source of religious authority. Martin Luther was in many ways a very traditional dogmatic theologian, but he learned Greek and translated the New Testament into German from Erasmus' edition. Henceforth humanist philological knowledge was necessary in any serious treatment of the text of Scripture on the Catholic as well as the Protestant side of the struggle. Moreover, since both sides claimed continuity with the early church and wished to harness to their positions the authority of the church fathers, the historical and linguistic expertise of the humanists became all the more crucial to official religion, as reflected in not only biblical scholarship, but also the myriad editions of the church fathers and the monumental cooperative historical projects, such as *The Magdeburg Centuries* of the Protestants and the *Acta Sanctorum* of Jean Bolland and his fellow Jesuits.

Philosophy also began to blend with humanism, most notably with regard to Platonism. As we have seen, Petrarch expressed preference for Plato. Leonardo Bruni considered himself an Aristotelian, but his translation of some of Plato's dialogues inaugurated the Renaissance reception of Plato. The process culminated in the great Renaissance Platonist Marsilio Ficino who combined competence in Greek and humanist literary interests with the profession of a philosopher. Only such a combination allowed Ficino to make Platonism a powerful element in Renaissance philosophy. Similarly, the humanistically trained Gianfrancesco Pico della Mirandola was able to introduce classical scepticism into Renaissance philosophical discourse because his anti-Aristotelian *Examen vanitatis doctrinae gentium (An Examination of the Emptiness of Pagan Learning)* of 1520 incorporated large tracts of the then still untranslated works of Sextus Empiricus. The humanists Cosma Raimondi and Lorenzo Valla became the chief proponents of Epicureanism in the fifteenth century.

Humanism combined less significantly with Aristotelianism (see Aristotelianism, Renaissance; Logic, Renaissance §3). Early on, several important humanists, such as Leonardo Bruni, George of Trebizond, and Theodore Gaza, were avowed Aristotelians. Bruni and Gaza consciously strove to replace in their translations the traditional terminology of medieval Aristotelianism. By the early sixteenth century one of the leading traditional Aristotelians in Italy, Agostino Nifo, thought it necessary to learn Greek. By the later sixteenth century humanistic Aristotelianism had become a powerful force, as documented by the popularity of the bilingual Aristotelian editions of Giulio Pace (1550-1635) and most of all by the bilingual editions and commentaries of the Jesuits of the Portuguese University of Coimbra (see Collegium Conimbricense).

7 Evolution of humanism

Different humanists had different interests and served different masters. Petrarch at times did the bidding of tyrants and sang the praises of the solitary life. In fifteenth-century Italy, the tyrants, popes and Aragonese kings of southern Italy had large numbers of humanists serving them. So did the republics of Florence and Venice. Leonardo Bruni, even before he became chancellor of Florence, glorified participation in the republic and has been hailed in more recent times as the founder of civic humanism (a brand of humanism that stressed republican values). Some humanists, such as Giannozzo Manetti, Bartolomeo Fazio and Erasmus emphasized the moral freedom of humans while others, such as Coluccio Salutati, Lorenzo Valla and all Lutheran humanists denied or severely limited the range of human free will. The Aristotelian George of Trebizond launched a crusade against Platonism at Rome in the 1450s and 1460s on the eve of the emergence of Florentine Platonism under Marsilio Ficino. Petrarch and Valla attacked medieval Aristotelianism in order, they said, to defend Christianity. Most Italian humanists, on the other hand, did not involve themselves with theology. Starting in the early sixteenth century, under the leadership of Erasmus, Christian humanism became a potent force for religious reform. In the religious crisis of the sixteenth century, humanists could be found on all sides of the theological arguments. Eventually Catholics as well as the major Protestant churches made humanism the basis of their respective

educational curricula. Humanists vented their fury on matters of Latin style. Some, such as George of Trebizond, Paolo Cortese, Pietro Bembo and the Jesuit schoolmasters insisted on the imitation of Cicero. Others, such as Angelo Poliziano, Erasmus and Petrus Ramus (§2) bitterly opposed the Ciceronians and called for an eclectic approach to imitation. Other models were also proposed: Valla favoured Quintilian, Cortese towards the end of his life converted to the imitation of Apuleius, and late in the sixteenth century Justus Lipsius popularized the imitation of Seneca. In short, apart from their commitment to Latin eloquence and classical literature, Renaissance humanists could differ considerably on philosophical, religious and cultural issues.

The people involved with humanism also changed over time. As the fifteenth century advanced, the link between notaries and humanists in Italy faded with the humanists becoming a distinct group of educated professionals, mainly occupied with teaching. The prototypical fifteenth-century Italian Renaissance educators had been Vittorino da Feltre and Guarino Veronese who ran humanist schools for the princely courts of Mantua and Ferrara respectively. But humanists also made careers in secular governments and the church. A few were businessmen, such as Niccolò Niccoli (d. 1437), a great collector of books. After Niccoli's death, Cosimo de'Medici took control of his library (in return for paying his debts) and made it the nucleus of the first public library since antiquity, the Biblioteca Laurenziana in Florence. Some humanists were even businessmen, such as Aldo Manuzio, whose printing press at Venice from the 1490s onwards became famous for its editions of the Greek and Latin classics. The Aldine Press remains today the best-known publishing house of the Renaissance.

Almost all humanists were men, but some women were also involved. The first female humanist of importance was Isotta Nogarola (1418-66), a noblewoman of Verona, whose Latin writings fill two volumes in their modern edition. At the other end of the chronological and geographical spectrum the pious Dutchwoman Anna Maria van Schurman (1607-78) was significant; she was best known for her treatise defending the education of women and extraordinary for her command of languages (Latin, Greek, Hebrew, Aramaic, Syrian, Arabic, Turkish, Ethiopian, Dutch, Spanish and English). The problem for women in the Renaissance was that they were denied access to the universities and learned professions.

Towards the end of the fifteenth century humanism began to penetrate northern Europe, which differed from Italy in some important respects. The north lacked Italy's tradition of secular oratory, and its universities were more involved in preparing students for a clerical career than those of Italy. Consequently, though northern Europe had its share of secular-minded humanists, it is no accident that Christian humanism became a movement first in the north and that the first humanist manuals for sacred oratory emanated from northern humanists. As in Italy, the vast majority of northern humanists were by profession teachers, along with a significant number of lawyers (including Thomas More), diplomats, bureaucrats and clergymen.

In the course of the sixteenth century a revolution took place in education (see Melanchthon, P.). The Jesuit colleges and the Protestant academies established a system of secondary schools, unknown to the Middle Ages. The basis of these new schools was the humanist curriculum. They were popular because they worked. Their students were better prepared for the university than those students who did not attend them. By the mid-seventeenth century, the victory of humanism was virtually complete. To be educated meant that one had received along the way a humanistic education.

That victory cost Renaissance humanism dearly. The humanists had captured the grammar schools and the new secondary schools and lost the universities. By the mid-seventeenth century, humanists had come to be mainly primary and secondary school teachers. In Catholic countries, for lay humanists even this career path was limited by the Jesuit colleges. Furthermore, in a world where the language of learning and letters had become the vernacular and where science and philosophy had left behind the classical systems, Renaissance humanist education appeared increasingly irrelevant and provoked calls for educational reform. Paradoxically, this turn away from Renaissance humanism proved liberating. In the nineteenth century, humanism was able to slough off its Renaissance rhetorical mission and become instead a training in culture and taste. The educated public rediscovered antiquity as a source of aesthetic and cultural values. No less importantly, the new humanist became a scientific discoverer, that is, the university professor opening up new frontiers of knowledge. The new humanism emanating from Germany had its own professional education, scientific methods and prestigious *cursus honoris*. Humanism once again took the lead in innovative learning. It engendered the research seminar and made research the standard for promotion within higher education. Indeed, as humanism became scientific classicism, it showed

the way for the vernacular literatures and the arts to become in their own right academic subjects at all levels of education. The Renaissance *studia humanitatis* had evolved into the modern humanities.

See also: Aristotelianism, Renaissance; Language, Renaissance philosophy of; Messer Leon; Petrarca, F.; Platonism, Renaissance; Renaissance philosophy; Scepticism, Renaissance

JOHN MONFASANI

References and further reading

Baron, H. (1966) *The Crisis of the Early Italian Renaissance: Civic Humanism and Republican Liberty in an Age of Classicism and Tyranny*, Princeton, NJ: Princeton University Press.(Revised edition of the 1955 publication; responsible for popularizing the notion of civic humanism.)

Chomarat, J. (1981) *Grammaire et rhétorique chez Erasmus (Erasmus' Grammar and Rhetoric)*, Paris: Les Belles Lettres, 2 vols.(The fundamental work on the humanism of the most important northern humanist of the Renaissance.)

Ferguson, W.K. (1948) *The Renaissance in Historical Thought*, Cambridge, MA: Harvard University Press.(The best study on the idea of the Renaissance with many references to humanism.)

Hankins, J. (1990) *Plato in the Renaissance*, Leiden: Brill, 2 vols.(A wide-ranging study of an important cultural current.)

Ijsewijn, J. (1990) *Companion to Neo-Latin Studies*, Louvain: Peeters, 2nd edn.(The best survey of Neo-Latin authors and literature.)

Kraye, J. (ed.) (1996) *The Cambridge Companion to Renaissance Humanism*, Cambridge: Cambridge University Press.(A collection of essays covering scholarship, teaching, the Bible, political thought, dialectic and science. Good bibliography, organized by topic.)

Kristeller, P.O. (1956-85) *Studies in Renaissance Thought and Letters*, Rome: Storia e Letteratura, 2 vols. (Contains many of Kristeller's important articles.)

Kristeller, P.O. (1955) 'The Humanist Movement', in *Renaissance Thought: The Classic, Scholastic, and Humanistic Strains*, New York: Harper, 1961, 3-23.(This was the first of Kristeller's celebrated Martin Classical Lectures of 1954 (published in 1955). The lecture also appears in *Renaissance Thought and Its Sources* on pages 21-32.)

Kristeller, P.O. (1979) *Renaissance Thought and Its Sources*, New York: Columbia University Press.(The handiest collection of Kristeller's 1945 article 'Humanism and Scholasticism in the Italian Renaissance' and his 1954 Martin Classical Lectures, which are the classic expositions of his interpretation of Renaissance humanism, presented in this entry. Also contains his magisterial overview of rhetoric and philosophy from antiquity to the Renaissance.)

Monfasani, J. (1976) *George of Trebizond: A Biography and a Study of His Rhetoric and Logic*, Leiden: Brill. (George of Trebizond is important for humanist rhetoric, logic, philosophy, translations, literature and the social setting.)

Rabil, A., Jr (ed.) (1988) *Renaissance Humanism: Foundations, Forms, and Legacy*, Philadelphia, PA: University of Pennsylvania Press, 3 vols.(A rich collection of articles by many specialists.)

Rose, P.L. (1975) *The Italian Renaissance of Mathematics*, Geneva: Droz.(A study of the importance of humanism to a non-humanistic discipline in the Renaissance.)

Schmitt, C.B. (1983) *Aristotle and the Renaissance*, Cambridge, MA: Harvard University Press.(Brings out the significance of humanistic Aristotelianism.)

Seigel, J.E. (1968) *Rhetoric and Philosophy in Renaissance Humanism: The Union of Eloquence and Wisdom, Petrarch to Valla*, Princeton, NJ: Princeton University Press. (A polemical assertion of the Kristeller thesis and an attack on the Baron thesis of civic humanism.)

Trinkaus, C. (1970) *In Our Image and Likeness*, Chicago, IL, and London: University of Chicago Press, 2 vols. (Extensive discussion of major humanist thinkers.)

Voigt, G. (1893) *Die Wiederbelebung des classischen Alterthums (The Revival of Classical Antiquity)*, ed. M. Lehnerdt, Berlin: Reimer, 3rd edn, 2 vols; Italian trans. of 2nd edn, G. Zippel (ed.) and D. Valbusa (trans.) *Il risorgimento dell'antichità classica*, Florence: Sansoni, 1888-97, 3 vols.(The first great study of Renaissance humanism; it has not yet lost its usefulness.)

Weinberg, B. (1961) *A History of Literary Criticism in the Italian Renaissance*, Chicago, IL: University of Chicago Press, 2 vols.(A fundamental study of a major interest of the humanists.)

Weiss, R. (1969) *The Renaissance Discovery of Classical Antiquity*, Oxford: Blackwell.(A guide to humanist antiquarian researches.)

Humboldt, Wilhelm von (1767-1835)

Along with Schiller and Goethe, Humboldt was one of the chief representatives of Weimar classicism, a movement that aspired to revive German culture along the lines of ancient Greece. Humboldt's philosophical significance resides mainly in two areas: political theory and the philosophy of language. In political theory he was one of the founders of modern liberalism; and in the philosophy of language, he was among the first to stress the importance of language for thought, and of culture for language.

Born into an aristocratic family in Prussia, Humboldt received a private education by some luminaries of the Berlin Enlightenment (see Enlightenment, Continental). As a young man he frequented the literary salons of Rahel Levin and Henriette Herz, where he cavorted with some of the leaders of the Romantic movement. Humboldt spent much of his career in the service of the Prussian state. From 1802 to 1810 he was the Minister of Education in the Prussian Reform Administration of Baron von Stein; and from 1813 to 1815 he acted as Prussian representative at the Congresses of Prague and Vienna. His most important achievements were as Minister of Education. In this role he founded the University of Berlin, and created the first unified school system in Germany, whose broad outlines still exist.

Like Kant,F.H. Jacobi and Georg Forster, Humboldt was one of the founders of German liberalism. His *Ideen zu einem Versuch, die Gränzen der Wirksamkeit des Staats zu bestimmen (The Limits of State Action)*, which was written in 1792 but not published until 1851, later became a classic of the liberal cause. Strictly speaking, it is anachronistic to regard Humboldt as a spokesman for liberalism, given that liberalism was not a self-conscious and organized movement in Germany until the 1840s. Nevertheless, his ideas proved influential for later liberals. The *Ideen* inspired much of J.S. Mill's *On Liberty* and Edouard Laboulaye's *L'État et ses limites (The State and Its Limits)*.

The distinguishing feature of Humboldt's liberalism is its humanism, its ethic of perfection, the development of individuality (*Eigentümlichkeit*) and wholeness. Humboldt contended that the main aim of the state should not be to promote happiness, but excellence or individuality. The central argument of the *Ideen* is that the best state to achieve this end is the minimal one that permits everyone to do whatever they want, as long as they do not interfere with a similar liberty for others. Since the chief condition of individuality and excellence is freedom of choice, the best state is the smallest.

It is important not to confuse Humboldt's view with other doctrines later associated with liberalism. Unlike most German liberals of the 1840s, Humboldt had little sympathy for democracy, and held that the best government for the protection of liberty is a constitutional monarchy. Although Humboldt approved of a market economy, his liberalism was also not a defence of *laissez-faire*. He deplored the doctrine that the main aim of the state should be to create wealth or prosperity, for he feared that such materialism undermined the pursuit of excellence. Finally, unlike later liberalism,Humboldt's individualism was not at the expense of communal values, because he insisted that people develop their powers only through cooperation with others.

Humboldt's role in the development of the modern philosophy of language has been the subject of dispute. It is too much to claim that he is 'the father of modern linguistics' (Ernst Cassirer), since this ignores the many earlier contributions to this field (F. Bopp, J. Degerando, E. Condillac, F. Schlegel andJ. Herder). Humboldt's philosophy synthesized, however, many of the most advanced ideas on the nature of language in the early nineteenth century. At the very least,Humboldt was one of the leading figures behind the development of comparative linguistics, the detailed study of the different forms of language.

The characteristic conception behind Humboldt's theory is his view of language as an organism. Like many early nineteenth-century thinkers, Humboldt believed that human activities are best explained in organic rather than mechanical terms. He held that language, like any organism, is an indivisible whole. It is a whole in two respects: the main unit of meaning is not the word but the sentence; and each language is unique, having a distinctive meaning that cannot be completely translated into another language. Humboldt also maintained that, like any living thing, language is dynamic. We cannot understand it simply as a static collection of rules and words, because it is an activity, the attempt to express thoughts in symbols. Because it is an activity, language is

constantly evolving and changing.

Rejecting the older rationalist view that words are only arbitrary symbols to designate already fixed concepts, Humboldt stressed the unity of thought and language. He insisted that we discover, define and develop our concepts only through words. Thinking consists in a dynamic interchange between words and concepts. After expressing an idea in a sign, that sign takes on an existence of its own that also partially determines the shape of our ideas. Both concept and sign shape each other until there is at least an approximate correspondence between them.

Humboldt also emphasized the cultural dimension of language. Each language is a unique form of thought, expressing the whole way of life of a people. This view led him to a doctrine of linguistic relativity, which he never fully explored. According to this doctrine, there are as many ways of thinking about the world as there are cultures.

The most controversial aspect of Humboldt's philosophy of language is his typology, which evaluates languages according to the degree that their matter is subordinate to form. He imagined a hierarchy beginning with the purely agglutinative languages, where matter dominates form, and ending with the inflected languages, where form subjugates matter. Such a typology, however, presupposes that there is some ideal or universal language, which is incompatible with the cultural dimension of language. It also clashes with two facts: that some peoples who have an agglutinative language have a highly developed culture (the Chinese); and that some inflected languages, such as English, tend to become formless by losing their endings. Humboldt attempted to meet these difficulties by revising and refining his typology. In the end, however, he never perfected his theory, which remains scattered in many unfinished writings spanning several decades.

See also: Liberalism

FREDERICK BEISER

List of works

Humboldt, W. (1903-36) *Gesammelte Schriften (Complete Works)*, ed. Prussian Academy of Sciences, Berlin: Behrs Verlag.(The most complete and authoritative edition of Humboldt's works.)

Humboldt, W. (1981) *Werke in Fünf Bänden (Complete Works, in Five Volumes)*, ed. A. Flitner and K. Giel, Darmstadt: Wissenschaftliche Buchgesellschaft.(The most accessible and commonly used edition.)

Humboldt, W. (1792) *Ideen zu einem Versuch, die Gränzen der Wirksamkeit des Staats zu bestimmen* (Ideas for an Essay to Determine the Limits of State Activity), in *Gesammelte Werke (Complete Works)*, ed. C. Brandes, Berlin: Reimer, 1851, (7), 1-197; trans. J.W. Burrow, *The Limits of State Action*, Cambridge: Cambridge University Press, 1969.(Humboldt's central work of political philosophy, in which he emphasizes the importance of the minimal state in the promotion of individuality and excellence.)

Humboldt, W. (1963) *Humanist without Portfolio: An Anthology of the Writings of Wilhelm von Humboldt*, trans. M. Cowan, Detroit, MI: Wayne State University Press.(Selected translations of Humboldt's writings.)

Humboldt, W. (1973) *Schriften zur Sprache (Writings on Language)*, ed. M. Böhler, Stuttgart: Reclam.(A useful anthology of all the essential writings.)

Humboldt, W. (1988) *On Language: The Diversity of Human Language-Structure and its Influence on the Mental Development of Mankind*, trans. P. Heath and with intro. by H. Aarsleff, Cambridge: Cambridge University Press.(A good anthology and translation of Humboldt's writings on the philosophy of language, with a useful introduction.)

References and further reading

Beiser, F. (1992) *Enlightenment, Revolution and Romanticism*, Cambridge, MA: Harvard University Press, 111-37.(A detailed summary of Humboldt's early political theory.)

Brown, R. (1967) *Wilhelm von Humboldt's Conception of Linguistic Relativity*, Hague: Mouton, Janua linguarum, series minor 65.(A clear explanation of Humboldt's philosophy of language.)

Haym, R. (1856) *Wilhelm von Humboldt: Lebensbild und Charakteristik (Humboldt: His Life and Character)*, Berlin: Gaertner.(The classic biography of Humboldt.)

Kaehler, S.A. (1927) *Wilhelm von Humboldt und der der Staat (Humboldt and the State)*, Munich: Oldenbourg.(A detailed account of Humboldt's politics.)

Lammers, W. (1936) *Wilhelm von Humboldt's Weg zur Sprachforschung 1785-1801 (Humboldt's Path towards*

the Investigation of Language 1785-1801), Berlin: Junker & Dünnhaupt.(A detailed account of the development of Humboldt's philosophy of language.)

Leroux, R. (1932) *Guillaume de Humboldt. La formation de sa pensée jusqu'en 1794 (Humboldt: The Development of His Thought up to 1794)*, Paris: Vrin.(The clearest and most detailed account of his early intellectual development.)

Hume, David (1711-76)

Hume's philosophy has often been treated as the culmination of the empiricist tradition of Locke and Berkeley, but it can also be seen to continue the sceptical tradition, and, even more strikingly, the naturalist tradition of Epicurus, Lucretius, Hobbes and Spinoza. Hume challenges orthodox religious conceptions of human nature. He presents us as part of a larger nature, sharing our basic cognitive and affective capacities with the higher animals. Our 'reason' is not some God-given privileged access to truth, but simply our language-affected variant of 'reason in animals'. The negative, anti-rationalist arguments of his first work, A Treatise of Human Nature, where he attacks the views of theistic rationalists, are more muted in his later writings, but the anti-religious arguments become ever more explicit. The importance of his philosophy lies in the thoroughness of his naturalistic project. He tries to show that neither knowledge nor ethics nor the political order needs any sort of religious foundations, and also to explain why so many thinkers had mistakenly held that they did.

To the end of his life Hume called himself a sceptic, but his scepticism was in the service of his secular reform of culture, and always 'mitigated' by his recognition that a 'true sceptic' would be as diffident of his doubts as of his convictions. Sceptical arguments are found useful, however, to cut down the pretensions of dogmatic religious and rationalist claims. His essay 'The Sceptic', although purporting only to portray one sort of philosopher, is often read as a self-portrait.

His first and now most famous work, his Treatise, was 'of human nature'*, which he takes to include our understanding, our passions, and what drives our moral and political life. Much of Book I of the Treatise, 'Of the Understanding', is devoted to showing how many of our beliefs are owing to our 'imagination', rather than to our 'reason'. Book II analyses our passions, their foundation in pleasure and pain, their various idea 'causes' and 'objects', and their communication by sympathy. One of its most famous claims is that passions are needed to motivate action. Hume, whose project is similar to that of Hobbes in Human Nature or the Fundamental Elements of Policy, follows him in taking will to be simply the transition from belief-informed passions to action, and takes voluntary human action to be in principle predictable, like everything else. Book III examines our moral evaluations of actions and passions.*

In the Abstract of the Treatise Hume underscores the importance of psychological association in our passions, our belief formation, and, more generally, in the workings of our imagination. Where John Locke and Francis Hutcheson had found associative thinking a disease of the mind, Hume makes it the norm. Association of ideas by 'resemblance', by temporal and spatial 'contiguity', and by 'causation' gently guides our minds in our spontaneous thinking and fantasizing, and association by causation guides us less gently in our inferences. Like Locke, Hume takes it for granted that 'anatomists' will have some explanation, say in terms of our nervous system and the physical proximity of memory traces in the brain, to account for the psychological phenomena.

Hume emphasizes the influence of experienced repetition on our beliefs and our passions. His famous account of causal inference, which Immanuel Kant claimed woke him from his 'dogmatic slumber', makes it the more or less instinctive extrapolation into the future of regularities and frequencies that have been experienced in the past, along with a tendency to project the felt 'determination of the mind', in its causal inferences, onto the subject matter of those inferences, giving us the idea of causal necessity. Even when made explicit in language, our causal inferences cannot be made into 'demonstrations', in which the conclusion follows with logical necessity from premises known to be true. Hume is credited with discovering 'the problem of induction', although Pascal may have helped in that discovery. Problematic or not, induction is relied on in Hume's own account of our nature. Indeed he believed that, unless we did rely on it in ordinary life, we would 'perish and go to ruin'.

Hume's general theory of human nature, and of the basic capacities and passions which we share with other mammals, is put to work in his account of what is distinctive to us, namely morality, religion, art, politics, and criticism. Like Shaftesbury and Hutcheson before him, he takes morality to involve a special pleasure taken in some human character traits. As moral judges, we find various character traits pleasing and displeasing, and we evaluate forms of social, cultural, and political organization which express or encourage such traits. Our moral sense is a reflexive form of our more basic capacity to take pleasure and displeasure in a range of different things. Morality, as Hume analyses it, is in no way dependent on that other distinctively human phenomenon, religion.

In his writings on religion he gave great offence to believers of his day, both by his diagnosis of the causes of religious fervour, and by his claims about its 'pernicious' effects. Like Pascal, he saw the basic cause of religion to lie in our anxiety about our own fate, but unlike him, did not endorse the religious response to this anxiety. Our ability to think about the future, combined with the limits of our success in finding natural causes for events that affect our happiness, and the intensity of our concern for that, lead us to postulate gods, invisible intelligent causal forces, super-persons, whom we try to please and placate by our prayers and devotions. Different religions and different forms of monotheism naturally develop, and religious persecution and religious wars are the result.

This entirely naturalistic account of reason, morality, politics, culture, and religion made Hume a hero for later Darwinians, such as Huxley, and the toast of the free-thinking Paris salons of his day. But in Britain, despite the fact that his political views are often taken to be conservative, his views on religion were seen by his contemporaries as dangerously radical, as an attack on the very foundations of his own culture.

1 Life and reputation

Hume was born in Edinburgh on 26 April 1711. He left us his own brief account of his life, where he highlights the 'good family' he came from (lawyers and landowners on both sides of the family), his 'slender' patrimony from his father, Joseph Home of Ninewells, who died in his infancy, the devotion of his mother, Katherine Falconer, to her three children, his own disinclination to study law, as was thought suitable for him as younger son, and his life-long 'passion for literature'.

After study at Edinburgh University, and beginning the study of law, he began working on what became the *Treatise of Human Nature* while living at Ninewells (Chirnside, near Berwick, Scotland) with his mother, brother and sister. His mother's only recorded comment about her famous son, 'Our Davie is a fine gude-natured crater [creature], but uncommon wake-minded [weak-minded, or weak of will]', may have its explanation in his failure to stick to his law studies. He left home in 1734 and completed the *Treatise* in La Flèche, 'a country retreat' in France, where he had the use of the library of the Jesuit college (where Descartes had been educated), and could try out his arguments about miracles on the priests. The book was published in London in 1739-40, and Hume was bitterly disappointed by its reception. He published essays during the next few years, and his *Essays Concerning Human Understanding*, (later the *Enquiry Concerning Human Understanding*), which restates some themes of the *Treatise*, in 1748. He had written to his friend Francis Hutcheson that the *Treatise* would offend the religious, but he still seems to have been dismayed when the offence it gave ruled him out as an applicant for the chair of Moral Philosophy at Edinburgh University in 1745. His radical views also cost him the Chair of Logic at Glasgow; the only teaching position he ever held was as tutor to a mad English nobleman, for one year, 1745-6. Later he served as a military and diplomatic secretary, as librarian to the Advocates' Library in Edinburgh (while he was working on his *History of England*), and as an Undersecretary of State in London, 1767-9. As Secretary to the British Ambassador in Paris, 1763-6, he met many French intellectuals and earned the title 'le bon David'. On his return journey he escorted Rousseau to England to escape persecution in France and Switzerland. This brief acquaintance ended in a famous falling out.

During the 1750s he had published the *Enquiry Concerning the Principles of Morals*, which Hume believed was his best work, the *Political Discourses*, and the *Four Dissertations* (containing his controversial *Natural History of Religion*), as well as volumes of his *History of England*. It was chiefly as a historian and essayist that he achieved fame in his lifetime. In 1769 he retired to Edinburgh, built a house in the new St Andrews Square, and could live 'opulently' on his earnings. He never married, but had several close friendships with women, and took a great interest in the education of the children of his older brother and of some of his friends. He and his sister entertained a large circle of Scottish friends (including Adam Smith, the architects Robert and James Adam, William Robertson, Adam Ferguson, Hugh Blair, and many ministers of the Church of Scotland), and occasional guests from overseas, such as Benjamin Franklin, who stayed with Hume for a month. Until his death in 1776 Hume continued adding corrections to new editions of his published writings and polishing his *Dialogues Concerning Natural Religion*, published posthumously in 1779. As he was dying he was visited by James Boswell, curious to find out if, when death was close at hand, the 'great infidel' could preserve his calm Epicurean acceptance of the view that death was annihilation. (His publisher, William Strahan, also expressed the same curiosity in a letter.) Hume joked with Boswell about death and the possibility of an afterlife, and then changed the subject to recommend that Boswell read Smith's *Wealth of Nations*. Boswell left 'with impressions which disturbed me for

some time'.

Hume continued to disturb the religious from the grave. Leslie Stephen, writing in 1876, remarked that Hume's writings were so subversive to religion that the general reluctance to face up to their arguments had led to a neglect of his philosophy. However Thomas Reid and Immanuel Kant had not neglected it, and nor had Jeremy Bentham and John Stuart Mill. T.H. Huxley saw Hume as an ally in the cause of a naturalistic science of human nature, and even T.H. Green, who found in his philosophy little except an object lesson in the errors of subjectivism and empiricism, did him the valuable service of re-editing his writings.

To his first biographer, T.E. Richie, Hume was a great historian who had dabbled in philosophy in his youth. For much of the twentieth century he was seen as a great sceptical philosopher who, having shown how little philosophy could hope to do, turned away from philosophy and passed his time writing a sort of '1066 and all that'. Increasingly he is coming to be seen as a scientist of human nature whose 'science' included studying the origins and development of his own culture. His *History* is found to be interestingly philosophical, his philosophy from the start to have been intent on 'a cautious observation of human life …men's behaviour in company, in affairs, and in their pleasures' (*Treatise*: xix). The sceptical strain in his thought, especially when combined with his ambitions as a scientist of human nature, continues to fascinate commentators. Increasing attention is being paid to Hume's essays and later work, and to their relationship to his earlier more 'abstruse' writings.

Strahan, in a letter to Hume's nephew who was arranging the posthumous publication of the *Dialogues Concerning Natural Religion*, wrote that this long-delayed publication 'will probably make some noise in the world; and its tendency be considered in different lights by different men'. Not only the *Dialogues*, but the entire corpus of Hume's writings continues to make some noise in the world, and the tendency of his thought continues to be considered in different lights by different interpreters of it.

2 Hume's account of knowledge

Knowledge and probability. In Hume's early works, the term 'knowledge' is restricted to what is certain, and the term 'probability' (in a wide sense) is used in the *Treatise* for all factual beliefs which might get revised. (Later Hume is more willing to talk with the vulgar, and use 'knowledge' less strictly, 'probability' more narrowly.) Knowledge in the strictest sense is confined to current sense impressions, along with 'intuitions' about the relations between currently sense-perceived qualities, and certain 'relations of ideas', namely those that are 'demonstrably certain'. 'Demonstration' shows that some claims must be affirmed if we are to avoid self-contradiction. Factual claims are never of this 'demonstrated' sort. They are either, if restricted to current sense certainties, known immediately, or, if they make a claim about what goes beyond the evidence of current sensation, then they will be less than certain. Some past tense factual claims will be based on memory (and we can of course misremember), but most factual claims depend upon some inference which takes us from the evidence of current sense impressions and memories to a belief about what we have not, or not yet, ourselves observed. Hume's main concern in Book I of the *Treatise* and in the *Enquiry Concerning Human Understanding* is to examine these latter inferences, and distinguish the relatively well-supported from the less well-supported of them. Of particular interest to him are those that involve reliance on other people's testimony, and his famous discussion of the acceptance of reports of miracles (*Enquiry Concerning Human Understanding*: X) puts all the main principles of his epistemology to work. (This discussion was originally intended for inclusion in the *Treatise*.)

Impressions and ideas. Hume's most general epistemological principle is that all 'ideas', the contents of our thought, derive from more lively 'impressions', the contents of our sense experience and emotional experience. This is put forward at first as an empirical thesis, but it is also used as a normative principle. Impressions are either 'of sense' or, like passions, 'of reflection'. Ideas for which impression-credentials cannot be produced are dubbed 'false', 'imagined', or 'pretended' ideas. Hume takes this line with the idea of a vacuum, and of duration without change. We may *think* we have such ideas, he argues, but we cannot really have them. It is as if ideas, simply by being ideas, make some implicit claim to having impressions as their source, and, if such claims cannot be made good, they are dismissed as false pretenders. Ideas which go well beyond their impression data are termed 'fictions'. These include the idea of the continuous existence through time of physical things we have only intermittently observed, and of the identity of our own minds, with their everchanging perceptions. In accepting such ideas we 'disguise the variation' and 'remove the interruption' (*Treatise*: 254) in the data on which such ideas are based. This constructive contribution of the human mind, in its interpretation of the data of its experience, was

highlighted also by Immanuel Kant, who promoted Hume's 'fictions' into a priori 'categories'.

Abstract ideas. Hume follows Berkeley in taking every idea to be fully particular or 'individual', and to refer to something of which we have had or might have an impression. Abstractness and generality get into ideas by the use to which they are frequently put in our thinking. A particular idea becomes general, on a given occasion for a given thinker, if, for the purposes in hand, any other idea from a certain set could be substituted for it - if it is serving as a representative of that group of resembling ideas. The words of the thinker's language label these sets (are 'annexed' to them), and fix the sort of 'resemblance' determining their membership. So if our purpose is to consider a claim about all triangles, we will begin with an 'image' before our minds of, say, an equilateral triangle, but we may and should substitute other ideas of other types of triangles to test the claim. As Hume puts it, it is as if the thinker had 'a whole intellectual world of ideas' available to them, all of them ideas of particulars, but 'collected together' in groups and 'plac'd under a general term with a view to that resemblance which they bear to each other'. We show a capacity to 'pick out such as were most proper for our purpose' (*Treatise*: 24). Hume terms this instinct for relevance 'a kind of magical faculty in the soul'. At times he speaks as if the crucial resemblances are recognized before the resembling ideas are 'plac'd under a general term'; at other times it seems as if the prior availability of such terms is crucial to the operation of abstraction, as he analyses it. To treat language as a great aid to abstraction and generalization, rather than an indispensable precondition, would sit better with Hume's views about the role of generalization in causal inference, since he attributes causal inference to animals who lack language. He says that 'men…surpass animals in reasoning', and that 'books and conversations enlarge the sphere of one man's experience and thought' (*Enquiries*: 107).

Causal inference to new beliefs. Hume defines belief as 'a strong and lively idea deriv'd from a present impression related to it' (*Treatise*: 105). The derivation is by causal inference; the relation is causation. The inference is a move of the mind from some impression of sense (or memory) to a conclusion about what is not (or not yet) observed. The conclusion inherits liveliness, 'vivacity', or 'the belief feeling' from its impression-premise. An inference such as 'this is bread, so it will nourish' is never a 'demonstration', since there would be no logical self-contradiction in affirming the premise but denying the conclusion. Memory may assure us that in the past bread has nourished us, and current sense impressions may assure us that this seems to be bread, but what assures us that this bread will act on us as past bread has acted? We can try to convert the causal inference 'this is bread; past bread has nourished us; so this will nourish us' into a 'demonstration' by supplying an extra premise claiming that 'instances of which we have no experience must resemble those of which we have had experience' (*Treatise*: 89); but to claim that we *know* this, that it is certain, would be false pretence. It itself cannot be 'demonstrated'. Moreover, if we try to make it a matter of 'probability', itself a conclusion of an inference whose premise is that *in the past* instances which we had not yet experienced turned out to resemble those we had already experienced, we would have an inference of the same form as the original one, from evidence about the observed to a conclusion about the unobserved. To try to support the 'presumption' that 'the course of nature always continues uniformly the same' by pointing out that it *has* so continued is to give an answer that 'gives still occasion to a question of the same kind, even *in infinitum*' (*Treatise*: 91). We cannot turn our original inference into a sound demonstration. Either we add a premise about what nature 'always' does, which is merely dogmatically asserted, and so get a valid demonstration one of whose premises is a mere 'presumption', or we add less general premises for which we can provide empirical verification, but which leave the inference non-demonstrative in form.

Hume in *Enquiry Concerning Human Understanding* presents this argument as if it provides 'sceptical doubts concerning the operations of the understanding'. In the *Treatise* it is simply a phase of a longer positive argument to the conclusion that ''tis only so far as it [causation] is a natural relation, and produces an union among our ideas, that we are able to reason upon it, or draw any inference from it' (*Treatise*: 94). This is the substance of what the *Enquiry Concerning Human Understanding* calls the 'sceptical solution' to the 'sceptical doubts'. For Hume the scientist of human nature it is simply a fact about us, and a fortunate one, that like other animals we do make causal inferences; unlike them, however, we are able to realize exactly what we are doing, and to formulate rules to help us do it better, so that our predictions will be less often falsified.

The *Treatise* account of our belief formation through causal inference is not only explanatory but also normative, since it ends in rules for improving our inferences. Hume shows how we can see causal inferences as cases of mental association, and explains just what is special about association by 'causation'. Originally in the *Treatise* (Bk I, Pt I, Sect. IV), he had simply listed causation as one of the three 'natural relations' or associative principles.

But its effects are found to be special. Only causation, among the forms of association, can form a new belief - not merely, like the other principles of association, revive an old one, or make it more vivid. Hume undertakes to analyse the causal relation as we perceive it, so as to show how it can guide inference to new factual beliefs. He takes it to require contiguity, rather than distance, in space (one of the other principles of association), contiguity in time in the form of immediate priority in time, and some other more mysterious element that we perceive as 'necessary connexion'.

It is not until near the end of his long account that Hume can demystify that important ingredient in our idea of cause, although 'hints' had been given eight sections earlier, when Hume had 'discover'd' a 'new relation betwixt cause and effect, …*Constant Conjunction*' (*Treatise*: 87; original emphasis). Although 'from the mere repetition of any past impression [or successions of impressions?], even to infinity, there never will arise any new original idea, such as that of necessary connexion', he had written, 'it wou'd be folly to despair too soon', and in the end it turns out, on his story, that the experience of constancy of temporal conjunction does produce something from which the idea of necessity can be derived, namely inference, or constrained belief, what Hume calls 'the determination of the mind'. His prediction that 'Perhaps t'will appear in the end, that the necessary connexion depends upon the inference, instead of the inference's depending on the necessary connexion' (*Treatise*: 88) turns out to be correct. Necessity 'lies only in the act of the understanding' (*Treatise*: 166). Causal necessity has been reduced to the inferability of one member of a constantly conjoined pair of events from the other. So Hume then offers us two alternative definitions of causation, one as constant conjunction, the other as inferability.

In Section V of the *Enquiry Concerning Human Understanding* Hume says that the 'custom' that determines our causal inferences has 'equal weight and authority' with the 'reason' which is at work in demonstrative inferences. Yet he calls his naturalistic solution 'sceptical', and in Section XII of the *Enquiry Concerning Human Understanding* refers to the sceptic's 'triumph' in such an analysis of causal inference. In the *Treatise* the 'rules' offered for improving our causal inferences are characterized as having the status only of 'unphilosophical probability', or prejudice, albeit a reflexive case of prejudice. Sceptics may be pleased, Hume writes, when they observe 'a new and signal contradiction in our reason…the following of general rules is a very unphilosophical species of probability; and yet 'tis only by following them that we can correct this, and all other unphilosophical probabilities' (*Treatise*: 150). Scholars continue to disagree about the sense of 'sceptic' in which Hume sees himself as one of these sceptics, and about what most pleased him in his account of causal reasoning.

Testimony, credulity, and miracle reports. Hume's descriptive account covers bad as well as good causal inferences and probability estimations, and it is as part of his attempt to separate the better from the worse inferences that he looks at our reliance on human testimony, the inference from the fact that someone says they witnessed something to the conclusion that they did witness it, or, more generally from the fact that they say that they believe something, and have good evidence to do so, to the conclusion that they do believe it, and do have the evidence they claim for it. A large part of what we call 'education', Hume says, is simply believing what we are told, especially if it is repeated over and over again. (This for him confirms how sensitive we are to repetition.) Such uncritical credulity is out of conformity to Hume's epistemic rules. 'The wise man…proportions his belief to the evidence' (*Enquiries*: 110). The evidence on human lying, and on human error, should give us pause before believing all we hear. When what we hear goes counter to the most widely established 'constant conjunctions', when it is something that it suits someone's purpose that we believe, when it caters to our own suspect 'inclination to the marvellous', and especially when all these factors are present, the wise person is on their guard. Miracles, claimed violations of laws of nature on which religions are founded, do combine all these factors. 'What greater temptation than to appear a missionary, a prophet, an ambassador from heaven?' (*Enquiries*: 125). Self-delusion and 'pious frauds' are to be expected in such cases, from both reporter and hearer of the report. To anyone who looks calmly at the human record of miracle reports and at the rival religions founded on such reports, and who then weighs the probability that a reporter of such a miracle is lying or deluded against the 'proof' (constancy of conjunction), which is the support for the 'law of nature' that supposedly has been violated, the conclusion will be clear. 'We may conclude that the Christian Religion not only was at first attended by miracles, but even at this day cannot be believed by any reasonable person without one.' Hume here does his typical reflexive turn - as he applies causal reasoning to itself, so he here applies his account of belief in miracles to that very belief itself. (This idea is celebrated in the title of a modern challenge to theism, John Mackie's *The Miracle of Theism*.)

Bodies and minds. Book I, Part IV of the *Treatise* explores 'sceptical and other systems of philosophy'. The influence of Berkeley is apparent in Hume's discussion of our belief in lasting physical things, but Hume, unlike Berkeley, cannot appeal to God's notions of physical things to explain their individuation and identity. Belief in lasting bodies pervades our interpretation of current sensations, and so infects the terminology in which most of our causal inferences are formulated. So if these beliefs are, as Hume claims, mere 'fictions', based on 'false suppositions' (*Treatise*: 217), then our causal inferences will also be contaminated. Belief in lasting material objects ('bodies'), Hume claims, helps us to shore up our belief in constant conjunctions, since we just assume that although we may not have observed some particular cause or effect, it nevertheless happened. Lasting things serve as homes for unperceived events, and we need to believe in them to sustain our faith in 'nature's' regular course of action. But some uses of causal analysis, in particular those used to examine the causal relation between physical stimuli and our sense organs, and the supposed relation of 'secondary' to 'primary' properties, can seem to destroy our belief that in sense perception we get reliable information about lasting bodies. There is an unstable relation between our trust in causal inference and our presumption that there are lasting physical things that are involved in causal relations with one another, and with our minds. The first part of the conclusion to Book I of the *Treatise* explores this instability.

Hume's account in the *Treatise* (Bk I, Pt IV, Sect. VI) of what we believe about the identity of our own minds over time, an account supplemented in Book II of the *Treatise*, when he turns to personal identity 'as it regards our passions' (*Treatise*: 253), is less sceptical than the account of belief in lasting mind-independent bodies. However, in the 'Appendix' he reports that his hopes that a contradiction-free account could be given of our self-conceptions and the sort of identity we have reason to ascribe to ourselves (*Treatise*: 633) had been dashed when he realized that, in the section on personal identity, he had involved himself in a 'labyrinth'. Commentators do not agree on just what the perceived difficulty was. He had rejected the idea of a simple self, possessor of its perceptions, since there is no impression from which such an idea could be derived. The 'true idea' of any human mind is of a 'system of different perceptions or different existences, which are link'd together by the relation of cause and effect, and mutually produce, destroy, influence, and modify each other' (*Treatise*: 261). The identity of a mind is 'only a fictitious one, and of a like kind to that which we ascribe to vegetables and animal bodies' (*Treatise*: 259). The most proper analogy, however, is not to bodies but to a 'republic or commonwealth… as the same individual republic may not only change its members, but also its laws and constitution; in like manner the same person may vary his character and disposition, as well as his impressions and ideas, without losing his identity' (*Treatise*: 261). This account of the 'fiction' we tell ourselves about our (and our nation's) identity, and of the factual supports which give it believability, had seemed a satisfactory substitute for the rejected rationalist view that we know ourselves to be simple thinking substances, with an unchanging essence, persisting uninterruptedly through time. But the 'Appendix' finds the account given of 'the principle of connexion' that binds the 'different existences' together to be 'very defective'. Hume's 'scepticism' (*Treatise*: 633) about his original account of the achievable coherence and empirical well-groundedness of our ideas of ourselves continues to spawn a wealth of interpretative literature.

On Hume's own account, if it takes very 'refin'd and elaborate' or 'metaphysical' reasonings (*Treatise*: 268) to see the difficulties the philosophical sceptic finds, then even the mind that has seen them usually 'quickly forgets' them. His own purported attitude to our questionable beliefs about the world and ourselves is, by the end of the conclusion of Book I of the *Treatise*, that of the 'true sceptic', who, while not forgetting his philosophical doubts, is nevertheless as diffident about them as about his earlier convictions. In both cases he defers to his readers to point out to him any errors they find in his reasoning. Plenty of readers have taken up Hume's invitation, especially when it comes to his *Treatise* writings on personal identity. In the *Enquiry Concerning Human Understanding* he dropped this 'difficult' topic, and very much curtailed his treatment of belief in lasting bodies and a 'material world' (Sect. XII, Pt I). 'A blind and very powerful instinct of nature' is said to make us have and preserve a belief in lasting external objects, and trust our own senses to give us reliable information about such objects. 'The profounder and more philosophical sceptics' will be able to shake our blind faith, producing 'momentary amazement and confusion', but if we ask the sceptic 'what he proposes by all these curious researches?' 'if any durable good or benefit to society could ever be expected to result' (*Enquiries*: 159) from dwelling on his arguments, then, Hume says, the sceptic is the one who will be embarrassed.

Social epistemology. Hume's discussions of human knowledge in his essays and in the *History of England* keep

the question ‘Do you expect durable good and benefit for society to come from your epistemological researches, or are you merely amusing yourself?’ very much in the forefront of their author’s mind. His question becomes ‘When and how do various branches of knowledge grow?’ not ‘Is knowledge possible?’ Where his earlier epistemology was ‘abstruse’, analytical, and general, his later epistemology is much less general, and not at all abstruse. His interest shifts to social epistemology. Do the sciences thrive better in a republic than in a monarchy? What sort of representation does a guinea make, and would it make a different representation if it were made of copper, not gold? What relation is there between commercial trade between nations and exchange of ideas? What has the invention of the printing press done to the life of the mind, and what other inventions make or might make comparable changes? Epistemology merges into social philosophy and ethics.

3 Human passions and human actions

Passions and sentiments. Hume takes us to share our basic passion-repertoire, as well as our ability to learn from experience, with the higher animals. They may lack moral, aesthetic, religious, and philosophical ‘sentiments’, but they do, if Hume is right, love, hate, feel pride and shame, as well as desire, enjoy, suffer, hope, and fear. Their passions have the same sorts of objects that ours do, and we can sympathize with some of their passions, as some of them can with some of ours. ‘Passions’, with the exception of a few instinctive appetities, are ‘impressions of reflection’, reactions of pleasure or displeasure to some perception of our situation.

All passions have ‘objects’, what Brentano called ‘intentional directedness’ and others have termed ‘aboutness’ (see Intentionality). ‘Direct’ passions, such as desire, joy, sorrow, hope and fear, are caused by their ‘objects’. What Hume calls the ‘indirect’ passions, such as pride, involve both the thought of something that pleases, such as a fine cloak, and also the recognition of that good thing as belonging to a particular person, bringing a consequent pleasure in that person. Should the particular person be oneself, the pleasure one feels will be pride. Should it be another person, the pleasure will be affection or esteem for that person. So the basic ‘causes’ of all passions are ‘agreeable’ pleasure and ‘uneasy’ pain or distress, and thoughts about their causes or occasions. Since they are ‘impressions *of reflection*’, they are equally founded on our thoughts about what has given or would give us such pleasure and pain. Throughout Parts I and II of Book II of the *Treatise*, Hume traces ‘the double relations of impressions and ideas’ which are responsible for the varieties of pride and humility, love and hate, and various mixtures of them. He invokes association by resemblance to explain the transitions in our emotional life from one pleasure to another (from being pleased by some witticism to being pleased with the friend who uttered it), and he invokes association of more varied sorts to explain the thought transitions involved in such sequences of passions.

When Hume rewrote his account of human passions, as given in Book II of the *Treatise*, in one of his *Four Dissertations*, ‘Of Passions’, he blunted the distinction between direct passions such as joy and indirect passions such as pride, but he is still fairly boastful about the explanatory power of his principles of association. The *Dissertation* reverses the order of treatment in the *Treatise*, beginning with hope and fear and moving on to pride, humility, love, and hate, which he terms not ‘indirect’, but simply ‘passions of a more complicated nature’, ones which ‘imply more than one view or consideration’. The substance of the account is largely unaltered, and the ‘mechanics’ of association are still very prominent, but the order of treatment and the terminology have changed.

In both *Treatise* and *Dissertation* a distinction is drawn between ‘calm’ and ‘violent’ passions. This cuts across the distinction between relatively simple (or ‘direct’) passions and relatively complicated ones. Any passion can become violent. A passion, say a desire, is said to be violent when it produces ‘sensible emotion’ (*Treatise*: 417) or ‘sensible agitation’. Passions tend to become violent when one ‘swallows up’ another, thereby getting ‘new force and violence’, when a passion encounters opposition or delay in gratification, when uncertainty produces ‘agitation’ in the mind, including the case where the object of a passion is temporarily absent, or partially concealed or veiled. Although violence is not the same as strength, still the ‘force’ it involves usually strengthens a passion, so that ‘when we wou’d govern a man, and push him to any action, t’will commonly be better policy to work upon the violent than the calm passions’ (*Treatise*: 419). Custom, by giving us ‘facility’, can strengthen a passion despite the fact that it removes some causes of violence, namely agitating novelty, challenge, and uncertainty. Hume ends the *Dissertation* with the expressed hope that he has shown that ‘in the production and conduct of the passions, there is a certain regular mechanism, which is as susceptible of as accurate a disquisition, as the laws of motion, optics, hydrostatics, or any part of natural philosophy’. (Here he seems to echo Spinoza in his preface to *Ethics* III.) The ‘mechanics’ of violence in passions, however, seems a little less than fully

systematised by Hume. Since, in the *Treatise*, he wants to explain the errors of rationalists in ethics in part by the hypothesis that they confuse calm passions with reason, any inadequacies in his account of calm and violence has repercussions for his case against the rationalists.

Sympathy. All passions can be and often are communicated from one animal or person to another, through their understood bodily expression and our response to that in what Hume calls ‘sympathy’. Sympathy enlivens a mere idea of another’s passion into an impression. ‘The howlings and lamentations of a dog produce a sensible concern in his fellows’ (*Treatise*: 398). It is not only distress which is thus communicated, but any expressed passion, or even opinion. ‘This is not only conspicuous in children, who implicitly embrace every opinion propos’d to them; but also in men of the greatest judgment and understanding, who find it very difficult to follow their own reason or inclination, in opposition to that of their friends and daily companions’ (*Treatise*: 316). We spontaneously imitate the expressed state of mind of those around us, Hume believes, and this communication of feeling, along with its extension by our ability to imagine what others would feel, in various circumstances, is essential for the possibility of ‘the moral sentiment’. But there is also a ‘principle’, or basic tendency, which interferes with the workings of sympathy. This is ‘comparison’, which leads us to ask if we are doing better or worse than others. We can welcome another’s misfortune, rather than feel compassion for them, if that misfortune points up our own better fortune. ‘Comparison’ is invoked by Hume to explain malice and envy. He treats remorse as a case of malice against oneself, ‘an irregular appetite for evil’. He finds envy to be typically felt for those close in position to ourselves, and to be felt even when we are ‘superior’, should our inferiors be perceived to be advancing.

Sympathy can be blocked by the operation of ‘comparison’, but is facilitated by the perception of any sort of ‘natural relation’ between ourselves and others. ‘Similarity in our manners, or character, or country, or language…facilitates the sympathy’ (*Treatise*: 318). (Blood ties, and spatial contiguity also facilitate it.) Sympathy is sharing the feelings of others perceived to be like ourselves, or related to us, and so is felt differentially. Hume takes the moral sentiment to correct for such natural partialities in our sympathy.

The will, passion, and action. Hume includes an account of the will, and its freedom or lack of it, in Book II of the *Treatise*. Although he takes will to be the transition from thought and passion to action, it is discussed in his book on the passions, since ‘the full understanding of its nature and properties is necessary to the understanding of them [the passions]’ (*Treatise*: 399). The ‘mechanical’ nature of human passions is found to be corroborated by the ‘uniformity and regularity’ of human conduct. Our decisions and behaviour are just as predictable (in principle) as what happens in inanimate matter. We have at best ‘the liberty of spontaneity’, on those occasions when our behaviour has determining causes internal to us as persons, rather than ‘violent’ external causes, overriding our own conscious wishes. In the *Enquiry Concerning Human Understanding* Hume defines liberty as ‘the power of acting or not acting according to the determinations of the will’, (*Enquiries*: 94) and the will’s determinations are taken themselves to have their own psychological determining causes. This makes liberty a property not of the will, but of the agent (see Will, the). In this, as in his determinism, Hume agrees with Hobbes.

His account of our motivation, then, is that the complicated play of our passions, closely affected as they are by our memories, beliefs, and imaginings, lead to our intentions, ‘the determinations of our will’. These, as long as we are at liberty and do not forget or change our minds, get realized in our actions. Because passions are ‘impressions’, they are active psychological causes, sufficient to cause action. Because they are ‘impressions of reflection’, they incorporate the information about our situation that our memory, our beliefs, our reason and our reflection have given us. They need not be blind, however partial their ‘views’ often are. Hume’s account of the passions which motivate us makes them intrinsically reflective and thought-informed.

Because Hume takes our passions to determine our actions, the sort of conversation and company we provide, and the sorts of institutions, books, and artworks we are likely to produce, the evaluation of our passions becomes very important. His essays ‘Of Tragedy’, and ‘Of the Standard of Taste’ explore the passions literature arouses, our pleasure in it, and evaluation of it. Moral evaluation, he holds, is always the evaluation of ‘character’, of ‘principles in the mind and temper’ (*Treatise*: 477). His careful exploration of different sorts of sensitivity or ‘delicacy’, in the essay ‘Of the Delicacy of Taste and Passion’, like his investigations into relative ‘violence’ in passions, and into the role played in our passions by sympathy and by ‘the principle of comparison’, as well as his discussion of liberty and necessity, all feed into his account of moral evaluation.

Hume’s ‘moral psychology’ bears a striking similarity to Spinoza’s. Whether he had read Spinoza’s *Ethics* or just

Bayle's version of it, he certainly shows agreement with Spinoza's deterministic version of our psychology, and with particular points of detail, such as the effects of vacillation. His account of the effect of opposition on the violence of a passion (so that we naturally 'desire what is forbid') and his negative treatment of remorse, also repeat theses that Spinoza had advanced (see Spinoza, B. de; Bayle, P.).

4 Ethics

The virtues. Hume takes it to be agreed that moral judgment is primarily a judgment about human character traits, a recognition of 'virtues' and 'vices' (see Virtues and vices). He has some controversial views both about what enlightened moral judges will and will not include in their lists of virtues, and about how they do their judging.

In his 'catalogue of virtues' he distinguishes the 'natural virtues' from what in the *Treatise* he calls 'artificial', or convention-dependent, virtues. The latter consist in conformity to some beneficial social scheme of cooperation, such as a particular form of government. The artificial virtues, for him, include 'justice' (taken to include respect for traditional property rights, and 'fidelity to promises'), 'allegiance' (to magistrates), female modesty and chastity (preparation for and conformity to the role given to women in the artifice of marriage), and the duties of sovereign states to keep treaties, to respect each others' territorial boundaries, to give protection to ambassadors, and otherwise conform to 'the law of nations'. All the artificial virtues will be expected to take different specific forms in different societies and historical conditions.

In contrast to the artificial virtues, the 'natural' virtues are expected to be fairly invariant across cultures. Hume includes among the natural virtues compassion, generosity, gratitude, friendship, fidelity, charity, beneficence, clemency, equity, prudence, temperance, frugality, industry, courage, ambition, due pride (duly concealed so as not to offend others), due modesty (awareness of one's weaknesses), due self-assertiveness, good sense, wit and humour, perseverance, patience, courage, parental devotion, good nature, cleanliness, articulateness, responsiveness to poetry, decorum, and 'a certain *je-ne-sais-quoi* of agreeable and handsome', which 'render a person lovely or valuable' (*Treatise*: 611-12). The controversial features of his catalogue lie in part in the substitution of pride for humility, in part in the inclusion of qualities which do not depend on their possessor's will, and in part in the particular glosses he gives to traditional virtues such as courage and charity.

Courage is to be exercised, if possible, in non-military contexts, so that it need not involve killing or 'the sack of cities'. About charity Hume says:

> Giving alms to common beggars is naturally praised; because it seems to carry relief to the distressed and indigent: but when we observe the encouragement thence arising to idleness and debauchery, we regard that species of charity rather as a weakness than a virtue.
>
> (*Enquiries*: 180)

From the truly beneficent person, the hungry may indeed 'receive food, the naked clothing', but mainly because from such a person they will also receive 'skill and industry' (*Enquiries*: 178). Hume's later essays on economics develop this theme.

Hume's list of virtues is a self-conscious rejection of a puritan morality, and to some extent of Christian morality. Not only, as he wrote to Francis Hutcheson, does he not take his 'catalogue of virtues' from *The Whole Duty of Man* (a Protestant religious tract he had been encouraged to read as a child), but he finds some of the virtues listed there to be vices, from the point of view of a morality which has freed itself from 'the delusive glosses of superstition and false religion'. Hume's ethics are avowedly hedonist (see Hedonism). All virtues, he generalizes, are qualities that please from a moral point of view either because they prove 'agreeable' (tend to bring pleasure) to their possessors or to others, or else because they are seen to have 'utility' for their possessors or for others. 'Celibacy, fasting, penance, mortification, self denial, humility, silence, solitude, and the whole train of monkish virtues' are, he finds, neither agreeable nor useful; they 'stupify the understanding and harden the heart, obscure the fancy and sour the temper' (*Enquiries*: 270). *The Whole Duty of Man* had included, in its list of breaches of duty, various breaches of the duty of humility, and such other sins as 'eating too much', 'heightening of lust by pampering the body', 'not labouring to subdue it by Fasting, and other Severities', and 'not assigning any Set or Solemn time for Humiliation and Confession, or too seldom'. It was not merely the 'monks' whom Hume was opposing when listing the virtues, but the Calvinists who had preached to him, and the other Protestant divines

whose tracts he had been given as a child to help him learn to recognize vice.

But there are very many of the duties listed in *The Whole Duty of Man* that Hume includes in his catalogue, and that repeat those listed in Cicero's *Offices*, which he told Hutcheson was his preferred source book on morals (see Cicero). 'Not loving peace', and 'going to law on slight occasion', as well as theft, ingratitude, lying, malice, and oppression are all condemned in *The Whole Duty of Man*, and Hume would have little quarrel with these disapprobations. He told Boswell that, as a child, he skipped such things as theft and murder, in the list of sins provided by the *Whole Duty*, 'having no inclination to perform them'. As far as murder goes, not only did he omit to think about it, when examining his conscience as a child, but he omitted to consider it carefully enough in his moral philosophy. After citing parricide as a 'horrid crime' which rationalist moralists may have difficulty explaining, he then proceeds to say all but nothing in his own voice about what is wrong with homicide (as distinct from cruelty). He defends suicide (in the posthumously published 'Of Suicide'), and discusses varying tolerance of it and of tyrannicide in 'A Dialogue', (published with *Enquiries Concerning the Principles of Morals*) but gives us no extended discussion either of any artifice-grounded right to life, or of any natural virtue consisting of respect for human life.

Which forms of killing are vicious, and why? The rationalist answer may well have weaknesses, or even incoherencies, but Hume gives us no explicit answer. At most he gives us hints, such as 'While we are ignorant whether a man were an aggressor or not, how can we determine whether the man who killed him be criminal or innocent?' (*Enquiries*: 290). There are also many discussions of particular killings in the *History of England*, where Hume charts the shifting penalties for murder, their variation depending on the rank and sex of the victim, and the various degrees of cruelty in methods used by official or semi-official killers. (See his accounts of the killing of the exiled Queen Elgiva, of King Edward II, of Joan of Arc, as well as of the followers of Wycliff and other martyrs at the hands of various church and state authorities.)

Hume calls it a verbal matter that he includes among the virtues qualities of mind, such as wit and good sense, and 'involuntary' as well as 'voluntary' abilities. But it is no *merely* verbal matter that he refuses to call the 'involuntary' excellences 'talents', and to separate them from moral virtues. A sharp distinction between the voluntary and the involuntary is needed, he wrote in the *Treatise*, only by those who are handing out rewards and punishments, hoping thereby to regulate voluntary actions (*Treatise*: 609). But expressed moral approbation and disapprobation, as he understands them, are not deliberate reward and punishment, nor need they be preludes to reward and punishment. They are the more or less spontaneous expression of moral sentiments, themselves no more wholly voluntary than their objects. 'Philosophers, or rather divines under that disguise, treating all morals on a like footing with civil laws, were necessarily led to render this circumstance, *voluntary* or *involuntary*, the foundation of their whole theory' (*Enquiries*: 322). Hume does not treat morals this way, and so he is just as 'necessarily led' to resist the restriction of moral judgment to the voluntary.

Artificial virtues: justice. The part of morals that Hume does treat as being on a like footing with civil laws, and so as sanction-backed, is the area covered by 'the artificial virtues'. These consist in conformity to some generally beneficial convention, where the benefit accrues not act by act, but from the 'whole scheme of actions'. Since general conformity is needed for the benefit to be obtained, pressure is deliberately brought to bear, on adults as on children, to get them to show honesty in matters of property, to keep their promises, to show allegiance to lawful magistrates, and to show chastity and marital fidelity, at least if they are women.

'Justice', as rendering each their due, is taken by Hume to be primarily respect for property rights. His brief discussion of what we would call retributive justice is to be found in his discussion of liberty of the will (and throughout the *History of England*), not in his account of what he calls 'justice'. Although he frequently refers to equity, it is not clear what relation he took it to have to justice. (He includes it in a list of natural virtues.) What he argues about 'justice', in his sense, is that we need to have a convention establishing property, and establishing particular property rights, before anything could count as respect for such rights. Similarly, he will argue, we must have the social institution of promise (or contract, taken as mutual promise) before anything could count as keeping or breaking promises, and must have the institution of government before anyone can count as a superior, to whom allegiance is owed. To get the needed distinctions between possession and property, or between telling someone what one will do and promising to do it, or doing what one is told and showing allegiance to one's rightful superior, we must be able to turn to some social 'convention' which promotes some cases of possession

into ownership, some transfers into barter, gifts, loans or sales, some statements of intention into promises, some issuing of sentences in the imperative mood into lawful commands.

In the *History of England*, when reporting John Ball's 'seditious preaching' in the late fourteenth century, Hume describes the popularity of Ball's doctrine of 'the first origin of mankind in one common stock, their equal right to liberty and to all the goods of nature, the tyranny of artificial distinctions'. He says that such doctrines are bound to be popular since they are 'so conformable to ideas of primitive equality, which are engraven in the hearts of all men' (Ch. XVII). His own story, which he owes to Lucretius, about the origins of justice is not so different from Ball's, except that he emphasizes the inevitability of 'artifice', and does not regard it as necessarily bringing tyranny. The 'primitive equality' he recognizes lies in the need to get the agreement of each person to limiting the equal access of all to 'all the goods of nature'. Until there is a general agreement ending the primitive scramble to grab what one can from the goods of nature, then nothing will count as theft, as unjust taking, as wrongful possession. On Hume's story, only when every person sees, or seems to see, that they will be better off if present possession is frozen into rightful possession, into property, and only when there is an agreement to take only with owner's consent, can it cease to be all right for people to take what they please and can. Hume's story, like Ball's and like Hobbes' before him, and like that of Rousseau after him, presupposes the postulation of a primitive equality and liberty. The 'convention' needed to create private property rights, Hume writes, depends upon a 'general sense of common interest', on the perception that although any single property-respecting act may not bring advantages, indeed may involve loss, still 'the whole system of actions, concurr'd in by the whole society, is infinitely advantageous to the whole and to every part' (*Treatise*: 498). Unlike the Lollards and Levellers, however, Hume thought that rational and long-sighted convenors would see that equality of property, however theoretically desirable, is impracticable: 'Render possession ever so equal, men's different degrees of art, care, and industry will immediately break that equality. Or, if you check these virtues, you reduce society to the most extreme indigence' (*Enquiries*: 194). 'The most rigorous inquisition too is requisite to watch every inequality on its first appearance, to punish and redress it…so much authority must soon degenerate into tyranny' (*Enquiries*: 194). Where Ball saw inequality to depend upon tyranny, Hume sees the threat of tyranny to accompany the attempt to prevent inequality. But his sympathy with the egalitarians' ideals is as clear as his disagreement with them concerning the costs of equality.

Artificial virtues: fidelity to promises. Hume's analysis of promises and promissory rights follows the same lines as the analysis of property and property rights. Until a convention fixes which assurances are binding assurances, there will be no promises or contracts. The fixing, as in the case of property, will contain some elements of arbitrariness. Societies vary in such matters as the need for witnesses, for the contract to be written and signed, for there to be some 'consideration'. Hume takes the function of promise to be to extend security in transfers of goods and services to include future delivery and future services. He takes there to be a special form of words, 'I promise…' used when such binding assurances are given, and he takes their force to include the promisor's acceptance of the appropriateness of penalty in the form of withdrawal of trust (and thus disablement as a party to binding agreements) if the promise is broken. This informal enforcement of the sacredness of promissory obligations gives promises their verbal magic. Hume likens them to the priest's words in the mass, to the magic of transubstantiation, with the significant difference that 'the obligation [binding force] of promises is an invention for the interest of society', while 'priestly inventions…have no public interest in view' (*Treatise*: 524).

Hume denies that promises need to be given when any convention is accepted. Because he regards the obligation to keep promises and contracts as itself 'artificial', it would be absurd, he finds, to rest all artifice on promise. 'We are not surely bound to keep our word because we have given our word to keep it' (*Enquiries*: 306). Hume likens the conventions on which property and promissory rights rest to the tacit agreement we give to the conventions of our native language, or to the use of the monetary currency we find in use around us as 'measures of exchange', and to the agreement of rowers who both want to get in one boat across a river, and so coordinate their strokes. Recent contractarians have claimed Hume as one of them, despite his explicitly dismissing as absurd the suggestion that the agreement on which the obligations of justice rest could possibly itself be a mutual promise, and his rejection, both in the *Treatise* and in his essay 'Of the Original Contract', of the view that the authority of magistrates must rest on some promise or contract tying governor to governed. He agrees with the contractarians that the origins of justice lay in some sort of agreement that it was in everyone's interest to make with everyone else, but he would probably reject the label 'contractarian', and he certainly would not agree with those who try to

make all of morality, not only justice, rest on a hypothetical self-interested agreement (see Contractarianism).

Artificial virtues: chastity and modesty. Hume includes the virtues of female modesty and chastity among the artificial virtues, since he sees them to depend upon the existence of the institution of marriage, and in particular on the 'agreement' that husbands should not be expected to contribute to the upbringing of children they did not father. The only way men can have any confidence that the children born to their wives are 'really their own' (*Treatise*: 571) is by taking as wives only those women trained from childhood to an unnatural 'modesty', 'some preceding backwardness or dread' of sexual activity, and then by imposing the 'punishment of bad fame or reputation' on any wives suspected of infidelity. Hume stresses the 'unnaturalness', not just the social contrivance and usefulness, of these demands made on wives and would-be wives. It is in the interests of children that they receive care from a male as well as a female parent, and in general men can only be 'induc'd' to contribute to the care of children whom they take to be 'their own', so it becomes 'reasonable, and even necessary, to give them some security in this particular'. All men, Hume notes, including 'batchelors, however debauch'd', are shocked at 'lewdness or impudence in women', or at least in marriage-bent women. Is there irony in his account of the 'reasonableness' of a double standard? His early essays 'Of Love and Marriage', 'Of Polygamy and Divorces', and 'Of Moral Prejudices', continue his examination of the forms of marriage with which he was familiar, but readers do not agree on what these discussions reveal of Hume's own views.

How artificial virtue is inculcated. Hume takes it that we have a self-interested motive to adopt the policy of respecting property and keeping our promises and marriage vows since these 'inventions' are so designed that general conformity to their rules does bring advantages to 'the whole and every part' of society. And we have good moral reason to approve of such conformity. But, especially with property rights as they develop over a long period of time, and are affected by the transfers that have occurred through contract and through the invention of money, it is easy for a person to lose sight of the 'remote' and long term interest they may have in respect for established property rights. 'The artifices of politicians' may be needed to bolster their motivation. But Hume believes that we will usually see well enough what is wrong with disrespect for property rights when we are the victims of it. We see such disrespect as a 'vice', and feel disapprobation. We feel sympathy with other victims of theft, and our 'sympathy with public interest' leads us to condemn all breaches of property rights, however 'obscure' it may have become just what personal advantage we each get from a policy of honesty in preference to occasional judicious dishonesty. Hume grants that there may be no convincing answer to the 'sensible knave', who successfully conceals his dishonesty, and believes he does better than the scrupulously honest person. If educators, moralists and politicians have failed to give him 'an antipathy to treachery and roguery', if all he cares about is 'profit or pecuniary advantage', then no argument the moralist can provide is likely to change his mind. The moralist, on behalf of 'the party of humankind' (some of whom the clever knave is cheating and defrauding), may sincerely believe that the knaves are 'the greatest dupes', since they 'have sacrificed the invaluable enjoyment of a character, with themselves at least, for the acquisition of worthless toys and geegaws' (*Enquiries*: 283), but the knave will have his own sincere views about what is and is not worthless, and who are the dupes.

The moral sentiment, moral points of view. Hume takes virtues to be recognized as such by the moral sentiment, a special pleasure taken in agreeable and useful character traits, when these are surveyed from a 'general and steady' point of view. One of his most famous theses is that moral distinctions are not made by 'reason alone', but by the special sort of pleasure and displeasure we take in character traits and in the 'manners', or ways of behaving, that express them. Our approbation and disapprobation are what makes the approved traits virtues, the disapproved ones vices. 'We do not infer a character to be virtuous, because it pleases: But in feeling that it pleases after such a particular manner, we in effect feel that it is virtuous' (*Treatise*: 471). The 'particular manner' in which character traits must please to be virtues stems from the special 'point of view' which must be taken, and which we take in order to 'converse together on any reasonable terms' (*Treatise*: 581) about human merit. It is a 'steady and general' point of view which we have reason to expect others to be able to take, and from which agreement is in theory possible. In order to 'overlook our own interest', and get such a 'general' view, we must be capable both of sympathizing with other people's private viewpoints, and of correcting for natural bias in our sympathy. We consider how a given trait, say our own ambition, affects others, and try to see it as they see it. We do not ignore our own interest, but we look beyond that to the interests and concerns of everyone affected. When what we are judging is some military leader's courage, we will sympathetically consider both its effects on the armies he led, and his own nation, and also its effects on those against whom he led his forces, 'the subversion of

empires, the devastation of provinces, the sack of cities' (*Treatise*: 601). Moral approbation is a pleasure taken in a character trait, all things considered, and much preparation, including fact-finding, is needed to get to a point of view which really can claim to be general, which can expect to be 'steady', and expect to be shared.

Moral disagreement. Our expectation that our moral judgment will be shared with other moral judges is usually tempered by experience of moral controversy. Can Hume account for apparent moral disagreement? Even when we try to speak for 'the party of humankind', rather than just for ourselves, our class, or our nation, we often find other would-be representatives for humankind contradicting us.

Hume discusses apparent moral disagreement in 'A Dialogue'. 'Of the Standard of Taste' addresses what he sees as the closely similar topic of apparent disagreement about literary merit. In both these discussions Hume explains away the appearance of disagreement as due to some confusion, or lack of proper preparation or competence in some of the disagreeing judges. 'In moral decisions, all the circumstances and relations must be previously known; and the mind, from the contemplation of the whole, feels some new impression of affection or disgust, esteem or contempt, approbation or blame' (*Enquiries*: 290). Since very many judgments which purport to be moral are based on imperfect knowledge of circumstances, on narrow sympathies, on a contemplation of much less than 'the whole', it is not surprising that there is less than unanimity in such judgments. Whereas Hume takes all of us to be capable of becoming competent moral judges (if we are able and willing to rid our minds and hearts of 'the illusions of religious superstition and philosophical enthusiasm'), he took literary criticism to be a more elite skill, requiring both a 'delicacy of taste' and an extensive reading which not all could be expected to have. One interesting feature of his specification of the competent literary critic is that they show due allowance for difference of religion from writer to writer, so not condemn a book simply because of the false religion it presents, but show no such toleration on moral matters. The critic will be justly 'jealous' of their moral standards.

The arguments against the rationalists. Hume, in the *Treatise*, tries to refute the claims of rationalists such as Samuel Clarke that moral distinctions can be discerned by pure reason. His argument has two main parts. The first is that the conclusions of reason are 'ideas', but only 'impressions' can motivate, and moral distinctions are supposed to motivate. This argument relies on the sharp contrasts, drawn earlier in the *Treatise*, between impressions and less vivacious ideas, and between sense impressions and the passions they would need to 'concur' with, in order to motivate. Hume suggests that the rationalist may be mistaking calm passions, which do not agitate us, for passionless reason. The rationalist who disagrees with Hume's distinctions will be unmoved by his first argument.

The second part of the argument is that, in any case, there do not seem to be any rationally discernible facts or relations which would establish the sort of conclusion which the rationalist moralist expects to be able to draw, for example that certain sorts of behaviour (incest, killing a parent) are wrong if done by human beings, but not if done by animals or plants. This argument attempts to show that the wrongness of murder or incest is not a matter of rationally discernible relations. Hume cites the supposedly exhaustive list of 'philosophical' (cognitively discernible) relations of ideas given in Book I of the *Treatise*, and the claim of the *Treatise* (Bk I, Pt III, Sect. I) that only four of these, resemblance, contrariety, degrees in quality, and proportion in quantity or number, can be traced by 'demonstration'. Hume challenges any rationalist who disputes the completeness of his list of 'demonstrable relations' to point out the extra morally relevant relation. He requires that such a relation relate inner actions or states of mind to mind-external 'objects' (since he takes it that we all agree that moral judgment is restricted to expressed character traits), and that such a relation or relations determine what is 'forcible and obligatory' for all those capable of discerning them, that is for all rational beings. He takes it that, in order to found obligations, the relations would have to be shown to have a necessary effect on the will of all those who are obligated, and he seems to think that no rationalist could meet this challenge.

He adds, to his main anti-rationalist arguments, a famous 'observation' which he thought would 'subvert all the vulgar systems of morality' (*Treatise*: 470). This observation is that those who want to move from some factual claim to some conclusion about how we ought to behave owe us an explanation of how to derive the 'ought' of the conclusion from the 'is' of the premise (see Fact/value distinction). This is a challenge similar to that issued earlier to the rationalist who wants to found obligations on rationally discernible relations - to show a connexion to the will of the obligated, to show why that relation, or that fact, should fix for them what to do. He has not shown, nor does he claim to have shown, that there is no adequate answer to his challenges. 'Vulgar' rationalist systems may

be subverted by his observations, along with other vulgar systems, but more refined versions of such systems can be depended upon to find a way around the 'difficulties' which he challenges them to face.

5 Political and social philosophy

Nature and point of government. In the *Treatise* governmental authority is a social 'artifice', added to those of property, promise, and marriage. For it to come to exist, there must be a felt need that it satisfies, some 'convention' creating the right to govern, and bestowing it on some person or some group of persons. The need for government arises, Hume believed in the *Treatise*, only when other artifices, such as promise and contract, have enabled some people to accumulate so much property that they cannot adequately guard it against marauders, and to acquire it in ways suspected of fraudulence, so that others resent their possession of it. 'Acts of injustice' become frequent, and those who continue to respect property and promissory rights fear becoming 'cullies of their integrity'. What becomes needed is some security for the 'just' against exploitation by the 'unjust'. By creating a special job, that of 'magistrates', who will be responsible for 'the execution and decision of justice', such security is provided. The security is not only for the just against the unjust, but for the just impulses in any person against any temptation to injustice and inequity. Hume follows Hobbes in finding it a universal human weakness to see the attractions of the good that is close more easily than the attractions of 'remoter' goods, even when they are real and often greater. Without enforcement of property rules and contracts, 'the consequences of every breach of equity seem to lie very remote, and are not able to counterballance any immediate advantage, that may be reap'd from it' (*Treatise*: 535). The remedy, Hume writes, is to create rule-enforcers, so that there will come to be an immediate interest (avoiding punishment) as well as a remote interest (good character and reputation for it, social order and the benefits one gets from it) in avoiding acts of injustice. 'We are, therefore, to look upon all the vast apparatus of our government as having ultimately no other object or purpose but the distribution of justice, or, in other words, the support of twelve judges' (*Essays*: 27). These are Hume's words in 'Of the Origin of Government', a relatively late essay. In the *Treatise* he gave magistrates an additional role, namely coordinating new beneficial cooperative schemes which private enterprise might have neither motive to embark on nor the needed supervisory ability to bring to completion. Hume became more of a prophet for limited government as he grew older, but even in his late works he saw it as the governors' task not merely to declare and protect rights which pre-existed their own right to govern, but 'to point out the decrees of equity, to punish transgressors, to correct fraud and violence, and to oblige men, however reluctant, to consult their own real and permanent interests' (*Essays*: 38).

Who should govern. Hume is, of course, acutely aware of the difficulty of finding competent and reliable 'magistrates' or governors. In the *History of England* he praises the 'mixed' character of the English government, even in its earliest days providing some protection against tyranny. To avoid tyranny, powers must be balanced against each other, and the main tasks of politics as a 'science' are to find the best such balance, to design institutions that work well even when bad people are office-holders, and to provide for orderly transitions of power. The party system, the different roles of electorate, of the legislature, of monarch or other executive, of the judiciary, of the press, all should contribute to the goal of providing an ongoing government that really does serve the real interests of the governed.

Hume emphazises that there is some inevitable arbitrariness in the selection of governors. 'The magistrate may often be negligent, or partial, or unjust in his administration', and yet be a lawful magistrate, who is owed allegiance. The right to govern, like a property right, can be based on any of several different grounds. The hypothetical first magistrate might well have derived his authority from the consent of the governed, who might well have promised him allegiance:

> But when government has been establish'd on this footing for some considerable time, and the separate interest, which we have in submission, has produc'd a separate sentiment of morality, the case is entirely alter'd.
>
> (*Treatise*: 554)

> The same interest, therefore, which causes us to submit to magistracy, makes us renounce itself in the choice of our magistrates, and binds us down to a certain form of government, and to particular persons, without allowing us to aspire to the utmost perfection in either. The case is here the same as in that law of nature concerning the stability of possession.
>
> (*Treatise*: 555)

As the initial property rights were, Hume supposes, determined by ‘present possession’, to which in time get added ‘prescription’ or ‘long possession’, ‘accession’, ‘succession’, and ‘transfer by consent’ from former owners, so with governmental authority - a parallel plurality of grounds can be taken to establish it. What our ‘interest’ in government dictates is that the transitions from one governor to the next be smooth and nonviolent, and this paramount interest means that laws dictating how the next governor is to be selected are usually in place. ‘Positive law’, as well as ‘original contract’, ‘long possession’, ‘present possession’, and ‘succession’ can select the magistrates.

In his essay ‘Of the Original Contract’, in other essays, and throughout the *History of England*, Hume continues this *Treatise* discussion of the dangerousness and impracticality of being determined to ‘aspire to the utmost perfection’ in government, especially given widely different opinions of what method of selecting a magistrate is the more perfect. Consent or ‘original contract’ can get government started, but once any of the original contractors (subjects or sovereign), dies, or new subjects are born, it is wildly impractical and dangerous to demand renewed contracts. ‘Did one generation of men go off the stage at once, and another succeed, as is the case with silkworms and butterflies, the new race…might voluntarily, by general consent, establish their own form of civil polity, without any regard to the laws and precedents, which prevailed among their ancestors’ (*Essays*: 476). Hume has no quarrel with those who believe that ‘the consent of the people’ is ‘the best and most sacred’ of any claim to authority (*Essays*: 474), just as he has no quarrel with those who believe that equality would be the ideal distribution of property. In both cases, it is the means of trying to sustain the ideal state that pose the problems and the dangers. Even when nations do not aim so impossibly high as to found their governments on the consent of each and every subject, but appeal, as England did, to royal ‘succession’, disputes easily arise, and such disputes provide Hume with much of the storyline of the *History of England*. Hume’s political philosophy is realistic and pragmatic in character. Sympathetic though he is with the ideals of equality and government by consent, he sees the dangers of violence, disorder, and tyranny very vividly, and sees them to be the usual costs of aspiring to perfection in government. He did however write one very interesting essay on ‘The Idea of a Perfect Commonwealth’ in which his own republican ideals are developed, and he supported the cause of independence for the American colonies.

Hume’s later writings do not significantly alter his *Treatise* account of the function and authority of governments. They add mainly reflections on the relative wisdom of different governors and rebels against them, rather than any new points about the basis of governmental authority. The *History of England* tries to rebut the Whig claim that the English constitution, protecting their ‘ancient liberties’, dated from the Magna Carta. Hume describes a slow evolution of the ‘plan of liberty’ which eighteenth-century Britons enjoyed, with many periods of great but tolerated tyranny (under Edward III, under Elizabeth), some periods of lesser tyranny which nevertheless led to civil war and the king’s execution (the reign of Charles I), with rebellions occurring for all sorts of reasons and on all sorts of pretext, sometimes leading to improvements, sometimes to worse tyranny, sometimes to unintended advances in the direction of ‘a regular and equitable plan of liberty’. The study of English history, he wrote at the end of his discussion of the reign of Richard III, shows ‘the great mixture of accident which commonly concurs with a small ingredient of wisdom and foresight, in erecting the complicated fabric of the most perfect government’.

Social philosophy. There is virtually nothing that Hume wrote, after Book I of the*Treatise* that could not properly be classed as ‘social philosophy’. Even the *Enquiry Concerning Human Understanding*, mainly concerned with fairly ‘abstruse’ epistemological topics, keeps its eye on the social role of the epistemologist, and the social use made of, say, miracle reports. ‘Man is a sociable, no less than a reasonable being’ (*Essays*: 8), and Hume expresses the hope that even relatively abstruse philosophy may ‘diffuse itself throughout the whole society’, so that, among other effects, ‘the politician may acquire greater foresight and subtility, in the subdividing and balancing of power’ (*Essays*: 10). His essays on economics combine fairly detailed discussion of particular aspects of the then current economic policy in Britain with discussions of the highest generality about the role of money, the determinants of its value at a given time, the unimportance of what metal or non-metal is used as currency, the importance of a national debt. The *History of England* is crammed with interesting observations on such matters as the changing attitude to usury, what was being imported and exported, what taxes were being levied and with what purported authority, what role the church was playing, what the relationship was between canon law and civil law.

6 Philosophy of religion

Causes and effects of religion. Hume believes that religions, especially primitive religions, spring from the ‘trembling curiosity’ with which men, ‘agitated by hopes and fears’, scrutinize the ‘course of future causes’ affecting their life. Prominent among these fears is the fear of death and what may lie beyond it, and so it is part of Hume’s proposed cure for what he sees as the evils of ‘religion as it has been commonly found in the world’ that this fear be conquered, that he persuade us both that death is ‘annihilation’, which ‘entirely destroys this self’, and that the thought of such future annihilation need be no more repugnant than the thought that, before conception and birth, we were nothing. Hume here repeats Epicurus.

The ‘trembling curiosity’ into future causes affecting our life gets its anxious quality in part from the intensity of the hopes and fears which are involved, in part from uncertainty, in part from the sort of causes which are postulated. Religion involves ‘belief of invisible intelligent power’. ‘Invisible’ signalizes our lack of real understanding of this power; ‘intelligent’ signalizes our determination to make it comprehensible, to treat it as if it were a person, one to whom we can make petitions, who might be placated by our praises and sacrifices. It is our ‘curiosity about causes’, and its imperfect satisfaction by our empirical inquiries, which is the cognitive component of the religious impulse. It is our concern for our own welfare which provides the affective component. Religion evidences both our will to know the causes affecting our welfare, and our only partial success in discovering the determinants of our happiness and misery by natural means, let alone in controlling them.

As scientific understanding of these causal factors increases, religion might be expected to have a less vital role to play, but then, if Hume is right, religion’s role is never purely cognitive. It is our anxiety that needs a palliative, and uncertainty is only part of the cause of our anxiety. There is also our concern for our own future, and science may fail to cater to that. Hume sees the natural development of religion, once natural knowledge grows, to be a shift from polytheism to more ‘rational’ monotheism. Monotheism fits better with a more or less unified account of an orderly universe, whose regularities can be regarded as divine laws. But when one all powerful God is substituted for the many gods of more primitive religions, the reasons for fear of this God are increased, not diminished. Belief in one all-powerful God ‘is apt, when joined to superstitious terrors, to sink the human mind into the lowest submission and abasement’ (*The Natural History of Religion*: 52). What is more, different religious traditions develop slightly different versions of this one God, who becomes a jealous God, requiring his worshippers not merely to abase themselves, but to go to war against the infidel, and to burn the heretic. Hume sees monotheistic religions to be intellectually less ‘ridiculous’ than the polytheism from which they develop, but to be, morally speaking, much worse. The roots of religion are our ignorance of causes affecting our happiness, our tendency to anthropomorphism, and our desperate fears for our own future. To the extent that ignorance of causes is reduced, without any abatement of our fear of death or our infantile wish that some super-person be in charge of our fate, religion will simply adapt itself to scientific knowledge, not be banished by it. Hume’s thesis in the *Treatise*, that reason serves the passions, gets special application in the *Natural History*, where he tries to diagnose the particular passions served by theology and by religious beliefs, and finds that ‘the love of truth’ is ‘too refined’ a motive to explain such beliefs. They are not caused nor can they be cured by rational argument, however refined. If the roots of religion lie as much in our anxious passions and in our will to anthropomorphism as in our interest in causal forces at work in the world, then, to counter religion, more than science and argument will always be needed. The clever arguments put into the sceptic Philo’s mouth in the *Dialogues on Natural Religion*, and not answered by stronger counterarguments, seem in the end not to convince even Philo himself.

Argument from design. ‘The argument from design’ is put forward in the *Dialogues* by Cleanthes, a very calm and enthusiasm-free believer, who is of the opinion that ‘religion, however corrupted, is better than no religion at all’, (*Dialogues Concerning Natural Religion*: 219) and who defends the hypothesis that the universe, with the order it displays to our eyes and minds, is the work of an intelligent creator. Philo argues, against Cleanthes, that the hypothesis of an intelligent cause of the universe is only one of several that intelligent human thinkers have come up with, and has no better empirical support than the others. No hypothesis about a supposedly unique cause of a unique effect can have support of the usual inductive sort. Since the universe was not formed under our eye, all we can do is speculate and suggest analogies, taken from the ‘corner’ of the universe we have observed. We know by experience that an orderly effect (that is, one that strikes us as well-ordered) can come about in several ways. It may, like a building, have an intelligent builder as its cause. It may, like a spider’s web, be spun instinctively from the belly of an insect. It may, like a well-formed calf, have come to be by animal ‘generation’. It may, like a well

shaped turnip, have grown from seeds, by 'vegetation'. It may, like patterns in the sand on the beach, have come about by sheer chance. All these local causes of 'order' could be used as analogies to explain the order we perceive in the whole universe (or as much of it as we have any knowledge of). None of these known causes of local order are known as ultimate causes. In particular human thought is not. The builder's plans may have come into his head by observing and imitating natural order, or by copying earlier builders' work, and by cooperation with his workmates. Orderly thought itself calls for a causal explanation. Thought is often disorderly, and human designs are often botched. So our thought provides a very dubious analogy on which to model a worship-worthy universe-cause. 'What peculiar privilege has this little agitation in the brain which we call thought, that we must thus make it the model of the whole universe?' (*Dialogues Concerning Natural Religion*: 148).

Philo eventually answers his own question. Thought has privilege because the question of the cause of the universe arises only for thinkers. So 'a purpose, an intention, a design strikes everywhere the most careless, the most stupid thinker' (*Dialogues on Natural Religion*: 214). Our incurable anthropomorphism, what Hume in the *Treatise* called our mind's propensity to spread itself on external objects, ensures the privilege of thought among the many equally well- or ill-founded hypotheses about the cause of the universe. Philo's conclusion about what human reason can establish is 'one simple, though somewhat ambiguous, or at least undefined proposition, *that the cause or causes of order in the universe probably bear some remote analogy to human intelligence*' (*Dialogues on Natural Religion*: 227; original emphasis). The analogy is not said to be less remote than the others which Philo had proposed. 'Human intelligence' is undefined. It may still be being taken as a matter of agitation in the brain. And 'cause or causes' is carefully noncommital - reason cannot even establish monotheism over polytheism, let alone establish one intelligent *and just* cause of order in the universe. The remote analogy is to human intelligence, not to human benevolence or justice.

After this estimate of how little human thought at its least careless and stupid can conclude about the cause or causes of an orderly universe, and after a vivid catalogue of the 'pernicious consequences' of 'religion as it has commonly been found in the world', Philo purports to fly for alleviation of his ignorance to the revealed truths of Christianity (which, in the *Natural History*, Hume had argued was effectively polytheistic). Shortly before his dramatic flight, Philo had asked, 'When we have to do with a man, who makes a great profession of religion and devotion; has this any effect on those who pass for prudent, than to put them on their guard, lest they be cheated and deceived by him?' (*Dialogues Concerning Natural Religion*: 221). We have been put on our guard, but, as readers of Hume's *Dialogues*, we are still left in reasonable doubt about how to interpret the intentions of his characters, let alone of their author. Hume left his readers with a real enigma in the character of Philo, and, more generally, in these *Dialogues* and in the *Natural History*. As he wrote to Adam Smith when trying, shortly before his death, to arrange for the publication of the *Dialogues*, 'nothing can be more cautiously and more artfully written', and as he wrote at the end of the *Natural History*, 'the whole is a riddle, an aenigma, an inextricable mystery'.

Our prospects. In the *Natural History*, Hume's optimistic hypothesis was that 'the first religious principles must be secondary', or derivative. In some or even most conditions, human nature leads to religion, to varying religions, and so to religious zeal, religious persecution, religious wars. But religious belief and devotion are not 'so universal as to admit of no exceptions', and so do not spring from 'an original instinct', as Hume believed that, say, 'love of progeny' does. His aim was to diagnose the 'first principles' which in so many conditions had religion as their 'secondary' manifestation. Part of what we can see him to be doing in his *History of England* is studying the actual variations in religious (or irreligious) sentiment in one country's history, as a supplement to the less historical (more 'natural-historical') treatment of the varieties of religion that he gives us in the *Natural History*. His work as a scientist of human nature combined, in his lifelong attention to human religions, with his advocacy of the cause of 'the party of humankind', whom he saw to be so threatened by 'sacred zeal and rancor'. He spoke occasionally of the 'true religion'. Analogously to 'true scepticism', which turns scepticism on scepticism, this can be taken to refer to whatever benign reflection-improved form can be taken by the 'primary principles' that get their pathological secondary expression in superstition and in religious enthusiasm.

Hume is no optimist about the human condition. He is well aware of our bloodthirsty history, our record of inhumanity, and in his *History* he contributes to its recording. Where the Christians see religious devotion as the proper response to acknowledgment of the radical evil in our natures, Hume sees religion itself to be a near-inevitable evil propensity in our natures, one that increases our capacity for cruelty and inhumanity, and calls

for some secular salvation. This turning of the tables yields a perhaps over-simplified diagnosis of the main causes of human misery, but its originality, its daring, and its considerable explanatory power can scarcely be gainsaid.

See also: Artistic taste; Causation; Common Sense School; Empiricism; Enlightenment, Scottish; Induction, epistemic issues in; Miracles; Moral sense theories; Natural theology; Naturalism in ethics; Rationalism; Scepticism; Testimony

ANNETTE BAIER

List of works

Hume, D. (1875) *The Philosophical Works of David Hume*, ed. T.H. Green and T.H. Grose, London: Longman, Green, 4 vols.(At present the most complete edition of Hume's philosophical writings. It contains 'A Dialogue', and all the following works except the *History of England*, the 'Abstract' of the *Treatise*, and *The Letter from a Gentleman*. A new edition of Hume's writings, to be published by Clarendon Press, is in progress, edited by T.L. Beauchamp, D.F. Norton and M.A. Stewart.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press, 1975.(Includes the *Abstract*.)

Hume, D. (1740) *Abstract of A Treatise of Human Nature*, ed. J.M. Keynes and P. Sraffa, Cambridge: Cambridge University Press, 1938.

Hume, D. (1741-77) *Essays, Moral, Political and Literary*, ed. E.F. Miller, Indianapolis, IN: Liberty Classics, 1985.

Hume, D. (1745) *A Letter from a Gentleman to His Friend in Edinburgh: Containing Some Observations on A Specimen of the Principles Concerning Religion and Morality, said to be maintain'd in a Book lately publish'd, intituled, A Treatise of Human Nature*, ed. E.C. Mossner and J.V. Price, Edinburgh: Edinburgh University Press, 1967.

Hume, D. (1748, 1751) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press, 1978.(Does not contain 'A Dialogue'. Page references in the entry given as *Enquiries* refer to the 1978 joint edition.)

Hume, D. (1754-62) *The History of England, from the Invasion of Julius Caesar to the Revolution in 1688*, Edinburgh and London, 6 vols.(Most easily obtainable now from Indianapolis, IN: Liberty Classics, 1983.)

Hume, D. (1757) *The Natural History of Religion*, ed. H.E. Root, Stanford, CA: Stanford University Press, 1957.

Hume, D. (1779) *My Own Life*, in T.H. Green and T.H. Grose (eds) *Works*, vol. 3, London: Longman, Green. (Also titled *The Life of David Hume, Written by Himself*.)

Hume, D. (1779) *Dialogues Concerning Natural Religion*, ed. N.K. Smith, London: Collier Macmillan Publishers; New York: Macmillan Publishing Company, 1947.

Hume, D. (1932) *The Letters of David Hume*, ed. J.Y.T. Greig, Oxford: Clarendon Press, 2 vols.

Hume, D. (1954) *New Letters of David Hume*, ed. R. Klibansky and E.C. Mossner, Oxford: Clarendon Press.

References and further reading

(Anthologies are not included.)

Árdal, P.S. (1966) *Passion and Value in Hume's 'Treatise'*, Edinburgh: Edinburgh University Press.(An important study of Book II of the *Treatise* in relation to Book III.)

Baier, A.C. (1991) *A Progress of Sentiments: Reflections on Hume's 'Treatise'*, Cambridge, MA: Harvard University Press.(Emphasizes the dialectical unity of the *Treatise*.)

Beauchamp, T.L. and Rosenberg, A. (1981) *Hume and the Problem of Causation*, New York: Oxford University Press.(The standard work on Hume's views about the causal relation.)

Box, M.A. (1990) *The Suasive Art of David Hume*, Princeton, NJ: Princeton University Press.(A study of Hume's changes of style after what he saw as the failure of the *Treatise* to convey his message.)

Bricke, J. (1996) *Hume's Moral Psychology*, New York: Oxford University Press.(Emphasizes Hume's account of motivation, and relates his views to those of Donald Davidson.)

Capaldi, N. (1989) *Hume's Place in Moral Philosophy*, New York: Peter Lang.(Takes Hume to have effected a 'Copernican revolution' in ethics by turning away from the 'I think' perspective of his predecessors to a more social and pragmatic 'we do' perspective, from which cultural practice commands attention.)

Danforth, J.W. (1990) *Hume and the Problem of Reason: Recovering the Human Sciences*, New Haven, CT: Yale University Press.(A study of Hume's philosophy of culture.)

Deleuze, G. (1953) *Empiricism and Subjectivity*, trans. C.V. Boudas. London; New York, 1980.(Emphasizes the importance of Hume's account of the imagination, and the influence of general rules. Not for beginners.)

Flew, A. (1961) *Hume's Philosophy of Belief*, New York: Humanities Press.(A close examination of the *Enquiry Concerning Human Understanding*)

Fogelin, R.J. (1985) *Hume's Skepticism in the 'Treatise of Human Nature'*, London, and Boston, MA: Routledge & Kegan Paul.(A vivid presentation of Hume's *Treatise* sceptical arguments, taken to include his account of moral judgment.)

Forbes, D. (1975) *Hume's Philosophical Politics*, Cambridge: Cambridge University Press.(A mine of information about the complex background of Hume's political philosophy.)

Gaskin, J.C.A. (1978) *Hume's Philosophy of Religion*, New York: Macmillan.(The standard work on Hume's views on religion.)

Hall, R. (1978) *Fifty Years of Hume Scholarship: A Bibliographical Guide*, Edinburgh: Edinburgh University Press.(Supplementary updates are published in *Hume Studies*, the journal of The Hume Society.)

Hendel, C. (1925) *Studies in the Philosophy of David Hume*, Princeton, NJ: Princeton University Press. (Sympathetic and scholarly reading of some themes in Hume's writings, a good counterweight to T.H. Green's treatment, in his and Grose's edition of Hume's works.)

Huxley, T.H. (1894) *Hume with Helps to the Study of Berkeley*, London: Macmillan.(Of some historical importance. Hume's views on liberty and necessity are highlighted, as are his views about religion. Hume is praised for the 'sagacity' of treatment of human capacities as continuous with those of the higher animals. A brief but beguiling biography is included.)

Ikeda, S. (1986, 1988) *David Hume and Eighteenth Century British Thought*, Tokyo: Chuo University Library, 2 vols.(A valuable bibliography.)

Jones, P. (1982) *Hume's Sentiments: Their Ciceronian and French Context*, Edinburgh: Edinburgh University Press.(The book lives up to its title.)

Kemp Smith, Norman (1941) *The Philosophy of David Hume*, London: Macmillan.(Emphasizes the influence of Hutcheson, the priority of Hume's moral over his epistemological interests, and of his naturalism over his scepticism. A landmark work.)

Laird, J. (1932) *Hume's Philosophy of Human Nature*, London: Methuen.(A good critical survey of Hume's philosophy. Broad in coverage of topics, fair in criticism.)

Lecaldano, E. (1991) *Hume e la nascita dell'etica contemporanea (Hume and the Origin of Modern Ethics)*, Laterza.(Relates Hume's ethics and social philosophy to the contemporary contractarian tradition.)

Livingston, D. (1984) *Hume's Philosophy of Common Life*, Chicago, IL: University of Chicago Press.(Emphasizes the importance of Hume's treatment of time, and the 'narrative structure' of thought.)

Mackie, J. (1980) *Hume's Moral Theory*, London: Routledge & Kegan Paul.(An acute analysis and sympathetic criticism of Hume's moral theory.)

Malherbe, M. (1976) *La Philosophie Empiriste de David Hume (The Empiricist Philosophy of David Hume)*, Paris.(A general introduction.)

Miller, D. (1981) *Philosophy and Ideology in Hume's Political Thought*, Oxford: Clarendon Press.(A helpful introduction not only to Hume's political theory, but to his philosophy in general.)

Mossner, E.C. (1954) *The Life of David Hume*, Oxford: Clarendon Press.(Goes well beyond the earlier biographies by T.E. Richie (1907), John Hill Burton (1846), and J.Y.T. Greig (1931) in the wealth of fascinating information offered.)

Norton, D.F. (1982) *David Hume: Common-Sense Moralist, Sceptical Metaphysician*, Princeton, NJ: Princeton University Press.(Discusses Hume's relation to Hutcheson, and the restrictions of his scepticism.)

Norton, D.F. (ed.) (1993) *The Cambridge Companion to Hume*, New York: Cambridge University Press.(Essays on all aspects of Hume's thought, with extensive bibliography.)

Noxon, J. (1973) *Hume's Philosophical Development: A Study of His Methods*, Oxford: Clarendon Press.(About the only study of Hume's philosophical development.)

Passmore, J.A. (1952) *Hume's Intentions*, Cambridge: Cambridge University Press.(A clear and scholarly book which stresses the many different strains in Hume's thought, including associationism, positivism and scepticism.)

Pears, D. (1990) *Hume's System: An Examination of the First Book of His 'Treatise'*, Oxford: Oxford University Press.(Particularly helpful on Hume's *Treatise* account of causal inference.)

Penelhum, T.H. (1975) *Hume*, London: Macmillan.(A short clear book, especially interesting and influential on personal identity.)

Penelhum, T.H. (1992) *David Hume: An Introduction to His Philosophical System*, West Lafayette, IN: Purdue University Press.(A very helpful commentary on selected texts from Hume's *Enquiries*. Contains a fine bibliographical guide.)

Phillipson, N. (1989) *Hume*, London: Weidenfeld & Nicolson.(A study of Hume as historian, emphasizing his criticisms of religion as a social force.)

Price, H.H. (1940) *Hume's Theory of the External World*, Oxford: Clarendon Press.(A close examination of Book I, Part IV of the *Treatise*. Ingenious, influential, idiosyncratic.)

Snare, F. (1991) *Morals, Motivation and Convention: Hume's Influential Doctrines*, Cambridge: Cambridge University Press.(A close critical scrutiny of Hume's *Treatise* views in social philosophy and their relevance for contemporary discussion.)

Stewart, J.B. (1992) *Opinion and Reform in Hume's Political Philosophy*, Princeton, NJ: Princeton University Press.(Challenges the view that Hume is a 'conservative' in his politics.)

Strawson, G. (1989) *The Secret Connexion: Causation, Realism, and David Hume*, Oxford: Clarendon Press. (Challenges the usual reading of Hume's views about causes in nature.)

Streminger, G. (1994) *David Hume: Sein Leben und Sein Werk (David Hume: His Life and Work)*, Paderborn-München-Wien-Zürich: Ferdinand Schöningh.(A lavishly illustrated comprehensive account of Hume's life and writings, particularly good on the cultural background.)

Stroud, B. (1977) *Hume*, London: Routledge & Kegan Paul.(A deeply philosophical exploration of the interplay of scepticism and naturalism in Hume's philosophy.)

Whelan, F.G. (1985) *Order and Artifice in Hume's Political Philosophy*, Princeton, NJ: Princeton University Press.(Has become the standard work on Hume's political philosophy in the *Treatise*.)

Wright, J. (1983) *The Sceptical Realism of David Hume*, Manchester: Manchester University Press.(Helpful on Hume's debt, in his epistemology and metaphysics, to Descartes and Malebranche.)

Yandell, K.E. (1990) *Hume's 'Inexplicable Mystery': His Views on Religion*, Philadelphia, PA: Temple University Press.(Relates Hume's views on religious belief to his *Treatise* account of belief in general.)

Humour

What is meant by saying that something is humorous or funny? It is clear that humorousness must be elucidated in terms of the characteristic response to humour, namely humorous amusement, or mirth. It is plausible to define humour in this way: for something to be humorous is for it to be disposed to elicit mirth in appropriate people through their awareness or cognition of it, and not for ulterior reasons. But this invites the question, 'What is mirth?' The three leading ideas in philosophical theories of humour are those of incongruity, superiority and relief or release. Although the perception of incongruity is often involved in finding something funny, and the resolution of a perceived incongruity plays an important role in good humour, none of these, in themselves or combined with others, is capable of capturing the concept of mirth. Mirth is not identical with the pleasurable perception of an incongruity, pleasure in feeling superior, the relief of tension or release of accumulated mental energy, or any combination of these elements. A better account of mirth is that it is a certain kind of pleasurable reaction which tends to issue in laughter if the reaction is sufficiently intense. So something is funny if it in itself pleases appropriate people through being grasped, where the pleasure is of the sort that leads, though not inevitably, to laughter.

1 The main question

What is humour? Or alternatively, what makes something funny? This basic query can be interpreted as a call for either conceptual elucidation or causal explanation. Taken the first way, the question asks, 'What does it mean to say, or what are the truth conditions of saying, that a given item is humorous or funny?' In other words, how is 'humorous' or 'funny' to be defined in a philosophically rigorous fashion? Taken the second way, the question becomes, 'What is it about a humorous or funny item that is responsible for its being humorous or funny?' On the first construal, we seek to understand exactly when something counts as humorous; on the second, we seek to ascertain in virtue of what a given item succeeds in being humorous. These questions have not been sharply separated in traditional theories, which sometimes seem concerned to address the one and sometimes the other. Whether in the last analysis the questions *can* be sharply separated, though, is not entirely clear. Ultimately it comes down to asking whether the identifying response to humorousness can itself be identified without specifying what occasions it, either externally, in terms of features of the occasioning item, or internally, in terms of features of the subject's representation of the item. This identifying response is usually labelled 'amusement', but 'amusement' is here to be understood in its specifically humour-related sense, and not in the sense of general entertainment or diversion.

Although the basic query has here been formulated in the objective mode - that is, it asks what it is for something to *be* humorous - there is, perhaps, the prior question, in the subjective mode, of what it is for someone to *find* something humorous. These, however, can be plausibly related as follows: something is humorous if and only if it is found humorous by appropriate (or intended) audiences under favourable conditions, including cognitive, attitudinal and emotional ones. The context should make clear under which of these modes the basic question is being pursued at a given point. Some objectivity about humour - some degree of true-or-false-ness in the attributions of humour - is presupposed in considering the question in the objective mode, but this appears to be justified: items are regularly and sustainedly classified as humorous, and not just as humorous to a particular person on a particular occasion. Humour, though patently a response-dependent phenomenon, seems to have at least as much objectivity as, for example, the properties of beauty or virtue.

2 Traditional theories of humour

There are three main philosophical traditions of accounting for humour: the incongruity tradition, the superiority tradition and the relief or release tradition.

The hallmark of incongruity theory is that it locates the humorous in some incongruity presented by or perceived in some item. The humorous item may be itself incongruous, relative to some assumed other object, or it may involve or contain incongruity. There have been various interpretations of the incongruity of items or elements, ranging from logical impossibility or paradoxicality, to absurdity and irrelevance, to unexpectedness and general inappropriateness.

Incongruity theorists include Schopenhauer, Hazlitt, Kierkegaard, A. Koestler, D.H. Monro and perhaps Kant. Kant held the humorous to consist in 'the sudden transformation of a strained expectation into nothing' (1790). The incongruity here, if any, is between the expectation, or what it points towards, and its deflation, or what it issues in. Schopenhauer gave incongruity theory a clearer, perhaps canonical, formulation: the essence of the ludicrous, he claimed, lies in the incongruity between concepts, the vehicles of abstract thought, and concrete objects, apprehended in perception, when the incongruity is grasped of a sudden. The mismatch of thought and perception can appear from either of two directions: a single concept can be applied to two very different objects, which only awkwardly encompasses them both ('wit'), or two objects originally ranged under a given concept can be subsequently realized to be fundamentally disparate ('folly'). Koestler's version of incongruity theory, a descendant and elaboration ofSchopenhauer's, holds that humour arises from the bisociation (double association) of an item in respect of two different and incompatible reference frames or interpretive matrices at once.

Recent incongruity theorists have generally held the perception of incongruity to be the core of the response to something as humorous, but not the whole of it. Perceived incongruity is taken as necessary, but not sufficient, for the occurrence of humorous amusement, and the reasons for this are manifest. Incongruity is on the face of it an undesirable property and its confrontation usually fails to elicit pleasure straightforwardly; anxiety or bewilderment, or at best curiosity, are more likely to result. In addition, not all pleasure taken in incongruity appears to constitute amusement, as opposed to aesthetic or other forms of satisfaction.

What more, then, is required? Some theorists hold that the perceived incongruity must be enjoyed for its own sake, some that it must be enjoyed as such but not aesthetically, some that it must not give rise to negative emotions (such as fear or disgust), some that it not engage practical concerns (as for knowledge or safety), and some that it must have a tendency to issue in laughter. Some stress the temporal structure needed for perceived incongruity to be found humorous, while others insist it is not perceived incongruity itself that is the source of amusement, but only the consequent resolution of such incongruity. Still others underline the fulfilment of background conditions, such as being in fun, or the absence of egoistic or sympathetic concern for the object of humour.

Nevertheless, we may be justifiably sceptical of the claim that all intuitively accountable instances of humour, for example, mimicry, satire, sarcasm, slapstick or sexual ribaldry, turn on the perception of incongruity. It has not been demonstrated that properties other than incongruity cannot sensibly figure as what one is explicitly amused by. Thus, doubt remains over whether the incongruity theory, however qualified or supplemented, can be correct as a conceptual elucidation of humour.

Superiority theorists, who include Aristotle, Thomas Hobbes, Alexander Bain and Henri Bergson (§6), construe humour as rooted in the subject's awareness of superiority, in some respect, to the humorous object. Hobbes famously declared humorous laughter to be the result of a 'sudden glory' in one's eminence or fortune, in contrast with another or one's former self. Bergson (1899) theorized the comic as essentially 'the encrustation of the mechanical on the living' - a falling-off from the human ideal of flexibility, suppleness and accommodation. The observer of this, for Bergson, accordingly feels superior, takes pleasure in so feeling, and manifests their pleasure naturally in laughter at the imperfectly human. Comedy and emotion are held to be incompatible, since comic engagement by its very nature short-circuits emotional involvement. In addition,Bergson maintains that comic laughter is a social corrective that chastises and hopefully reforms the socially undesirable rigid behaviour at which it is directed.

Spencer and Freud, the most well-known relief theorists, locate the essence of the humorous in the relief from psychic constraint or the release of accumulated mental energy that it affords (see Freud §4). Spencer (1911) felt it important to investigate not only the features of humour, but why it is specifically laughter that humour induces, thus necessitating a physiological explanation. The explanation he offers emphasizes nervous tension and its bodily manifestation when suddenly excessive or redundant.Freud's striking account of the pleasure taken in jokes, influenced bySpencer's, is well worked out, as is his extensive typology of jokes in terms of their structures and techniques. Freud (1905) viewed the enjoyment of jokes as rooted in an economy of psychic energy, namely, that of inhibition or repression. With innocent jokes, the inhibition is against nonsense and pure play, while with tendentious jokes, the inhibition is against a display of aggression or sexuality, but in both cases the energy of inhibition thus freed up manifests itself as pleasure.

Whatever truth they contain, superiority and release theories lack the generality of incongruity theory. In addition,

they seem more concerned with concomitants or mechanisms of the humorous reaction than with its conceptual core. Thus these competitors of incongruity theory are currently seen as even less able to provide an adequate answer to the basic question.

The above classification of theorists involves a good deal of oversimplification, for strands of each of the three guiding traditions of reflection on humour can be uncovered in almost every major theorist. Thus, Kant might just as easily be called a relief theorist as an incongruity theorist, in virtue of the stress he placed on the animation of the body through the quick release of tension built up in expectation of what does not arrive. Bergson could justifiably be classified as an incongruity theorist rather than a superiority theorist. The incongruity that defines the comic in his analysis is posited between a human or human-like being and various quintessentially anti-human automatisms and rigidities with which they are afflicted. Schopenhauer's account (1844), though obviously bringing incongruity to the fore, also includes a strain of superiority theory; for Schopenhauer, the phenomenon of humour exemplifies an important truth that we have independent reasons for acknowledging, namely, the superiority of perceptual to conceptual modes of knowing the world. Part of our pleasure in humour is in direct consequence of its affirmation of this truth. Spencer's account, which conceives the humorous reaction as a sudden release of nervous energy, also posits that this results from a 'descending incongruity', in a manner reminiscent of Kant. Koestler's account (1964), borrowing from Spencer, combines the postulation of a collision of incompatible frames or matrices - a form of the incongruity idea - with the notion of an emotional mass and its explosive diversion as laughter when deserted by thought, in the tradition of release theories.

3 The analysis of humour

A number of considerations must be borne in mind when formulating an adequate analysis of humour, that is, an answer to the question 'What is humour?' construed conceptually. Most of these concern the proper relationship of humour to other phenomena, such as laughter, emotion, pleasure and aesthetic appreciation. First, humour and laughter are not coextensive, that is, not all laughter, by any means, is occasioned by humour. Laughter can result from, among other things, tickling, nitrous oxide, organic disorder, joy, embarrassment or vengeful exultation. Second, not all humour is productive of laughter, even in appropriate subjects; humour may engender amusement without any behavioural manifestation, or with only the lesser one of smiling. Third, humour does not always produce amusement, its characteristic pleasure, even in appropriate subjects; certain background conditions of mood or psychic preparedness also need to be met. Fourth, humour seems to have both a cognitive and an affective component, which are bound up together in the response.

It might be thought that incongruity theory is clearly aimed at answering the conceptual query, while release theory is clearly aimed at the causal one: noticing incongruity at least appears explicative of what finding something funny consists in, whereas the release of tension seems to concern the mechanism whereby finding something funny generates pleasure or some other effect. Perceived incongruity is a plausible intentional object of amusement - what it is directed upon - while release of tension is not.

But what of superiority theory? Though one is surely not amused at a quick release of nervous energy, it seems not impossible that one might be amused at one's evident superiority to some less fortunate person, when it is suddenly noted, in addition to or as opposed to whatever incongruity such misfortune may present. It is not clear that the pleasure I take in someone's accidentally slipping on a banana skin without serious harm cannot be accounted part of humorous enjoyment as such, but only as something distinct. It is not clear that an item's reinforcement of one's good fortune, or its deflation of expectations, or its presentation of ambiguity, or its surprisingness, or its strangeness, or some other more specific property, cannot itself be what is relished, and even the whole of what is relished, in certain cases of amusement. If so, then it may be a mistake to regard perceived incongruity as conceptually requisite to humorousness.

Hence there seem to be two choices for proceeding with the analysis. On one hand, if all cases of humour can be demonstrated to involve perceived incongruity, and to do so non-accidentally, then perceived incongruity should figure in an elucidation of the notion of humour, with apparent cases of humour without incongruity being shown to be cases either of non-humour or of humour in which non-humorous pleasure, derived from other sources, overshadows what proper humour-pleasure is present. On the other hand, if apparent cases of humour without incongruity are to be accepted as genuine instances of humour, then an analysis is needed which elucidates humour without reference to perceived incongruity. Perhaps the only plausible way of doing so would be in terms of a

distinctive and recognizable effect on perceivers and one arising through cognition of the item in question. On such a perspective, apparent incongruity would be the most common, but not the necessary, focus of humorousness; other properties might so figure, on other occasions.

Suppose, believing that even incongruity theory unjustifiably limits the possible objects of amusement, we begin with this general idea: the humorous is that which makes one laugh by thinking or perceiving it. Then a natural refinement would be as follows: an item x is humorous (or funny) if and only if x has the disposition to elicit, through mere cognition of it, and not for ulterior reasons, a certain kind of pleasurable reaction in appropriate subjects generally (that is, informationally, attitudinally and emotionally prepared subjects), where this pleasurable reaction (amusement or mirth) is identified by its own disposition to induce, at moderate or higher degrees, a further phenomenon, namely, laughter. Thus described, the humorous cannot be detached from all felt inclination, however faint, toward the convulsive bodily expression we call laughing. The propensity of the state of amusement to issue in laughter is arguably what is essential to its identity, and underpins the widespread intuition that humour and laughter, though not coextensive, are nevertheless intimately related. The connection between amusement and laughter, then, would be this: the mental state of amusement is partly identified by its disposition (universal in humans, if ultimately contingent) to issue in laughter if sufficiently intense.

What of the idea that amusement is not amusement unless it both arises in a certain way and has a certain intentionality? The present analysis acknowledges this in its own fashion, for it entails that a reaction to x is not amusement unless, in addition to being pleasurable and characteristically leading to laughter, it comes about in virtue of cognition of x, and is also directed on x. However, *pace* certain theorists, such asClark (1970), amusement may not have a formal object - a description under which an object must be seen if it is to amuse - beyond the minimal ‘that which is amusing’.

‘What makes x funny?’, when taken as a question about why speakers count x as funny, is directly answered by such an analysis. Roughly, x is funny in that, or because, the cognition of x amuses people with an appropriate mental set, that is, pleases them in such a way that they have some inclination to laugh at x. But ‘What makes x funny?’, taken as a question regarding what it is about x that underwrites or contributes to its being funny or eliciting amusement, can be answered only by empirical investigation or reflective survey. There may be many factors that enter into the explanation of an item’s possessing the power of humorousness, though we can be confident that presenting appearances of incongruity will figure largely among them.

It may then be useful to recognize two analyses of humorousness: one ‘thin’, one ‘thick’. The ‘thinner’ analysis, which we have just been sketching, denies there is any necessary focus or intentional object of humour, and holds an item’s humorousness to be merely its power to raise via cognition a certain pleasurable effect, identified through its connection with laughter, in appropriate subjects. The ‘thicker’ analysis assumes that all cases of humour can be shown conceptually to involve the perception of incongruity, and so adds that to the specification of the thought through which a humorous item must cause a pleasurable effect or produce enjoyment. But as we have seen, adopting an analysis of that sort runs the danger of prematurely foreclosing on the possible objects of amusement, a reaction that may well be characterizable without recourse to a focus on incongruity.

4 Incongruity reconsidered

Even if incongruity is not a necessary condition or component of humorousness, no account of humour can fail to accord it a special status. Beyond being the most common focus of humour, its special status may consist in the following. First, there is reason to think that superior forms of humour - those which are most satisfying, intellectually and emotionally - all rely on incongruity in one way or another. Second, there may be categories of humour (jokes, for instance) that are inconceivable in the absence of incongruity, even if there are categories (farce, perhaps) that have another basis. Third, the quality of incongruity-based humour may be tied to a further feature, one obviously presupposing such incongruity, namely, the nature and extent of the resolution of the incongruity that the humorous item embodies or presents.

The pleasure afforded by incongruity-based humour characteristically seems to require that the apparent incongruity be in some sense *resolved* by the subject. Such resolution can be an object of conscious awareness on the subject’s part, and can take various forms, including justification, rationalization, unification, or dissolution, but is perhaps best understood as the *grasping* of the rationale of the incongruity the humorous item presents. The

appreciation of incongruity-based humour can be likened to the solution of a puzzle, though a puzzle where insight is attained in a relatively immediate and effortless way. The resolution of a joke's incongruity is easily related to or identified with the experience known as 'getting' the joke. In good incongruity humour - a clever pun, for example - one is made to see the 'why' of the incongruity in addition to the 'what'.

The idea of grasping the incongruity in an instance of humour might be taken further: to resolve the incongruity in an item of humour and thus to be in a position to appreciate its humorousness is to grasp the basis of the incongruity involved, and, at the same time, an aspect of congruity as well often residing in the humorous vehicle itself. Without such a double grasp, of both the 'fit' and the 'non-fit' involved in a piece of humour, amusement of a high order is unlikely. Good incongruity-humour offers 'the pleasure of finding connections where none were thought to exist' (Monro 1951).

We may thus propose that model instances of incongruity-humour involve an underlying unification of their disparate contents, a tying together in the humorous vehicle of incongruous elements, rather than just their brute juxtaposition. In other words, the best such humour always has a pivot on which the humour turns, which rationalizes the apparently incongruous elements that have been brought together.

5 Other issues

Other issues of philosophic interest involving humour include: the status of amusement *vis-à-vis* emotion; the relation of humour and aesthetic experience; the relation between producing and consuming humour; the role or use of humour in human life; the nature of the sense of humour; the distinctive features of jokes; humour and society; humour and seriousness; humour and creativity; the ethics of humour.

See also: Comedy; Emotion in response to art

JERROLD LEVINSON

References and further reading

Bergson, H.-L. (1899) *Le rire: essai sur la signification du comique*, Paris; trans. 'Laughter', in W. Sypher (ed.) *Comedy*, Garden City, NY: Doubleday, 1956.(The most fecund and imaginative of traditional theories of humour, combining elements of the superiority and incongruity traditions, and paying special attention to the social dimensions and purposes of humour.)

Carroll, N. (1991) 'On Jokes', *Midwest Studies in Philosophy* 16: 280-301.(Criticizes Freud's account of jokes and proposes an alternative account in terms of puzzle-solving.)

Clark, M. (1970) 'Humour and Incongruity', *Philosophy* 45: 20-32.(A seminal essay that analyses humour as involving the enjoyment of perceived incongruity for its own sake.)

Cohen, T. (1983) 'Jokes', inE. Schaper (ed.) *Pleasure, Preference, and Value*, Cambridge: Cambridge University Press: 120-36.(Explores some of the presuppositions and implications, including ethical ones, of making and enjoying jokes.)

Dauer, F.W. (1988) 'The Picture as the Medium of Humorous Incongruity', *American Philosophical Quarterly* 25: 241-51.(Argues that pictures, not propositions, are the fundamental vehicle of humour.)

Freud, S. (1905) *Der Witz und seine Beziehung zum Unbewussten*, Leipzig/Vienna: Dueticke; trans. J. Strachey, *Jokes and their Relation to the Unconscious*, Harmondsworth: Penguin, 1956.(Presents an elaborate categorization of jokes and a unified account of their functioning, with the pleasure of all jokes based on an economy of psychic energy.)

Hartz, G. (1991) 'Humor: The Beauty and the Beast', *American Philosophical Quarterly* 28: 299-309.(Argues that humour is a special sort of emotion, akin to aesthetic emotions.)

Kant, I. (1790) *Kritik der Urteilskraft*, Berlin: Lagarde; trans. J.H. Bernard, *Critique of Judgment*, New York: Hafner, 1951.(Kant offers a compressed, though suggestive, proposal along incongruity lines, seeing the essence of humour in a sudden deflation of expectation.)

Kierkegaard, S.A. (1846) *Concluding Unscientific Postscript*, trans. D.F. Swenson, Princeton, NJ: Princeton University Press, 1941.(A particular version of incongruity theory, relating issues about humour to the demarcation of the aesthetic, the ethical and the religious.)

Koestler, A. (1964) *The Act of Creation*, New York: Macmillan.(Perhaps the most sophisticated recent form of incongruity theory, coupled with elements of relief theory.)

Kulka, T. (1990) 'The Incongruity of Incongruity Theories of Humour', *IYYUN, The Jerusalem Philosophical Quarterly* 39: 223-35.(Argues that humour derives from resolution of incongruity, as opposed to incongruity

itself.)
LaFollette, H. and Shanks, N. (1993) 'Belief and the Basis of Humor', *American Philosophical Quarterly*, 30: 329-39.(Relates the problem of humour to others more central in current philosophy.)
McGhee, P. (1979) *Humor: Its Origin and Development*, San Francisco, CA: W.H. Freeman. (A comprehensive survey of its questions by a leading psychologist of humour.)
Martin, M.W. (1983) 'Humour and the Aesthetic Enjoyment of Incongruities', *British Journal of Aesthetics* 23: 74-84.(Argues that enjoyment of perceived incongruity for its own sake is insufficient for humorous amusement.)
Monro, D.H. (1951) *The Argument of Laughter*, Melbourne: Melbourne University Press. (Comprehensive classification and survey of theories of humour.)
Morreall, J. (1983) *Taking Laughter Seriously*, Albany, NY: State University of New York Press.(Monograph by the most prominent contemporary philosophical writer on humour, which explains laughter as the result of a pleasant psychological shift.)
(1983) 'Humor and Emotion', *American Philosophical Quarterly*, 20: 297-304.(Argues that amusement is not usefully categorized as an emotion, but should be contrasted with it.)
(1989) 'Enjoying Incongruity', *Humor*, 2: 1-18.(Explores evolutionary reasons for the development of the sense of humour in humans.)
Mulkay, M. (1988) *On Humour*, Oxford: Blackwell.(General survey of the topic from a sociological point of view.)
Raskin, V. (1985) *Semantic Mechanisms of Humor*, Dordrecht: Reidel.(An impressive treatise bringing to bear insights from linguistics and communication theory.)
Santayana, G. (1896) *The Sense of Beauty*, New York: Charles Scribner's Sons.(Offers criticisms both of incongruity and superiority theories.)
Schopenhauer, A. (1819, 1844) *Die Welt als Wille und Vorstellung*, vols 1 and 2, Leipzig: Brockhaus; trans. E.F.J. Payne, *The World as Will and Representation*, vols 1 and 2, New York: Dover, 1966.(An early and influential formulation of incongruity theory, which sees humorous incongruity as invariably rooted in the discrepancy between the concreteness of percepts and the abstractness of concepts.)
Scruton, R. (1982) 'Laughter', *Proceedings of the Aristotelian Society*, supplementary vol. 56: 197-212. (Questions whether humour necessarily focuses on incongruity, stressing instead its devaluational or demolitional aspect.)
Sousa, R. de (1987) *The Rationality of Emotion*, Cambridge, MA: MIT Press.(§11 explores the ethics of 'dark' humour and laughter.)
Spencer, H. (1911) 'The Physiology of Laughter', in *Essays on Education, Etc.*, London: Dent.(Formulation of the relief theory, in terms of 'mental hydraulics'.)
Swabey, M. (1961) *Comic Laughter*, New Haven, CT: Yale University Press.(Defends an incongruity account of humour of broad scope.)

Hungary, philosophy in

The situation of Hungarian philosophy can be best illustrated by two sayings: 'there are Hungarian philosophers, but there is no Hungarian philosophy', and 'a certain period of Hungarian philosophy stretches from Descartes to Kant'. The two ideas are closely connected. Thus on the one hand, there is such a thing as Hungarian philosophy: there are scientific-educational institutions in philosophical life and there are philosophers working in these institutions. On the other hand, there is no such thing as Hungarian philosophy: it is a history of adoption, largely consisting of attempts to introduce and embrace the great trends of Western thought.

After some preliminaries in the medieval and early-modern periods, Hungarian philosophy started to develop at the beginning of the nineteenth century. As a result of the reception of German idealism - the so-called Kant debate and Hegel debate - the problems of philosophy were formulated as independent problems for the first time, and a philosophical language began to evolve. After an attempt to create a 'national philosophy' - and after some outstanding individual achievements - the institutionalization of Hungarian philosophy accelerated at the end of the century. The early years of the twentieth century brought the first heyday of philosophy to Hungary, with the rapid reception of new idealist trends and notable original contributions. In the period between the two wars the development stopped: many philosophers were forced to emigrate, and Geistesgeschichte *(the history of thought) became prevalent in philosophical life. Following the communist take-over, the institutions of 'bourgeois' philosophy were eliminated, and Marxism-Leninism, which legitimated political power, took a monopolistic position. During this period, the only significant works created were in the tradition of critical Marxism and philosophical opposition. The changes in 1989 regenerated the institutional system, and the articulation of international contemporary trends - analytic philosophy, hermeneutic tradition and postmodernism - came to the fore.*

Besides some works by thinkers in exile, Hungarian philosophy has produced only one achievement which can be considered significant at an international level: the oeuvre *of György (Georg) Lukács.*

1 The beginnings

The origins of Hungarian philosophical thinking can be traced back to the Middle Ages. The products of the Latin-speaking culture of the age were defined by scholasticism. Following a number of anonymous or insignificant authors, the only important and influential thinker appeared on the scene in the fifteenth century: Pelbárt Temesvári, Franciscan theologian, was a philosopher of European standard, representative of the medieval spirit. His collection of sermons, containing drafts of speeches, based on the strict logical order of the scholastic division-system and illustrated by literary stories, became highly popular both within and outside the country.

In the Hungarian history of ideas, the influence of scholasticism continues well beyond the Middle Ages: it was the predominant philosophy until the nineteenth century, taught in Catholic high-schools. The effects of the different trends of Renaissance, Enlightenment and German Idealism found their place within this tradition (see Renaissance philosophy; Enlightenment, Continental; German idealism).

The author of the first Hungarian philosophical work, János Apáczai Csere, was the most prominent among the Protestant students who had been educated at Western universities. His main work, the *Magyar encyclopaedia (Hungarian Encyclopedia)* (1655) printed in Utrecht, summarizes the scholarship of the age for the school curriculum. Its significance for cultural history comes from the fact that this was the first work to expound the ideas of the time (through Cartesian epistemology) in Hungarian. Apáczai's attempt, however, did not have followers and Hungarian philosophy showed no notable development for another hundred years.

2 Debates and achievements

After these sporadic and isolated preliminaries, real development was initiated at the beginning of nineteenth century. The starting point was the reception of German Idealism, with the two debates spanning through the first half of the century: the Kant and the Hegel debates (see Kant, I.; Hegelianism).

The two important features of the Kant debate were the lack of an autonomous philosophical milieu on the one hand, and the reductive character of the Kant interpretation on the other. The debate of Protestant ministers - who

had visited German universities and become acquainted with Kantianism - was a philosophical debate at a time when no philosophical life was to be found. The reception of philosophical thought, therefore, lacked an independent sphere, and its treatment was defined by different - mainly moral and political - dimensions. Kantians saw Kant as a supporter of religion, morals and political status quo; anti-Kantians viewed his work as an attack on these. Accordingly, the discussion was seriously deformed: Kant's *oeuvre* was reduced to ethical-theological aspects. The fact that the epistemological and logical aspects of the critical philosophy took precedence over the ethical-ideological aspects both genetically and structurally was hardly acknowledged.

The Hegel debate was also mainly conducted by Protestant ministers, but was effected less by denominational characteristics. Its driving force was the awakening of national spirit, typical of the early romantic age, whose main purpose was to promote scholarship - including philosophizing - in the national language. The Hegelian participants of the debate attempted to introduce the most influential contemporary philosophy, namely Hegelianism, to Hungary; the anti-Hegelians, however, judged the abstract Hegelian system of categories to be alien to the national spirit. The discussion had a dual result. On the one hand, this was the first time that a philosophical debate gained broader publicity: following articles in literary journals, Hegelianism became a widespread topic of discussion. On the other hand, the discussion initiated a considerable development in Hungarian philosophical language as the translation of Hegelian terms contributed to the establishment of Hungarian philosophical terminology.

The Hegel debate forms the background to the evolution of the so-called 'reconciliation' school of philosophy, which had a dominant role from the late 1830s until the mid-1850s. The school set itself the task of creating a characteristically Hungarian 'national philosophy'. Although its representatives took sides against German Idealism, their philosophy was in fact nothing but the adoption of ideas of some German thinkers - mainly those of Krug and Fries, who interpreted Kant from an empirical-psychological point of view. The significance of the eclectic, rather low-level reconciliation school is to be found not so much in its intellectual contribution as in its culture-founding role. Following the example of the conventional school, the next century brought several more attempts to create a 'national philosophy' as an answer to the challenge which arose from the consciousness of both national and philosophical underdevelopment.

Around the middle of the century, two thinkers did significant work. János Erdélyi, aesthete and littérateur, represented Hegelianism on the *niveau* of the age. His polemical treatise, *A hazai bölcsészet jelene (The Present State of Domestic Philosophy)* (1857) criticizes the national philosophy of the reconciliation school by appealing to the universal character of philosophy. His monograph about the early history of Hungarian philosophy has been the most useful summary on the topic ever since. Writer and politician József Eötvös's reflections on liberal ideas yielded quite modern insights in contemporary terms. His monograph, *Der Einfluß der herrschenden Ideen des 19. Jahrhunderts auf den Staat (The Influence of the Predominant Ideas of the Nineteenth Century on the State)* (1851-4) (published originally in German), analysed the concepts of liberty, equality and nationality in sociological terms, and thereby became one of the first manifestations of a conservative reinterpretation of liberalism (see Equality; Freedom and liberty; Liberalism; Nations and nationalism).

3 The establishment and functioning of the institutional system

The full development of the institutional system of Hungarian philosophy started in the last third of the nineteenth century. The infrastructure of philosophizing became consolidated, and philosophy secured an autonomous position within intellectual life.

Bernát Alexander, a prominent figure of the turn of the century, played a major role in this process. The author of some eclectic works in the history of philosophy, psychology, aesthetics and literary studies, using mainly elements of positivism and Neo-Kantianism (see Neo-Kantianism), his significance is due to his activity as organizer of philosophical life. He had a considerable influence as a teacher, an editor and publicist, and he translated many works into Hungarian. He was the editor of the series called 'Collection of Philosophical Writers', published in twenty-nine volumes between 1881 and 1919 by the Hungarian Academy of Science. This series played a fundamental role in spreading philosophical scholarship in Hungary: for the first time, many classics of the history of philosophy - including the works of Plato and Aristotle, Bruno, Descartes, Pascal, Spinoza, Hume, Diderot and Kant - were published in Hungarian, and some of these translations are still in use today.

The first Hungarian philosophical periodical was the *Magyar Philosophiai Szemle (Hungarian Philosophical Review)*, founded in 1882; from 1892, it was replaced by *Athenaeum*. The Hungarian Philosophical Society was founded in 1901, first as an opposition to official philosophical scholarship, then becoming a part of the status quo.

Supported by the institutional bases, a significant flowering occurred in philosophy at the beginning of the twentieth century, and especially after 1910: this was the first heyday of Hungarian philosophy. The somewhat late positivist orientation was followed by the reception of various directions of new idealism. Axiological Neo-Kantianism, Diltheyian philosophy of life, Bolzano's logic, Husserlian phenomenology and Meinongian theory of objects all soon found supporters (see Dilthey, W.; Bolzano, B.; Husserl, E.; Meinong, A.). Two groups played a determining role in introducing the new idealist paradigm: the Böhm-students around the turn of century, and the Lukács-circle in the 1910s. Károly Böhm first tried to find a compromise between positivism and Neo-Kantianism, then worked on an axiology-based philosophy, akin to the efforts of the Baden school of Neo-Kantianism. His students at Kolozsvár - members of the 'Transylvanian School' - exemplified a specifically Hungarian branch of Neo-Kantianism. The young Lukács, with his short-lived journal *A Szellem (The Spirit)* and his intellectual group, the Sunday Circle, represented an ethical idealism embedded in a partly Kantian, partly philosophy-of-life-based conceptual framework (see Lukács, G.). Some members of the Circle subsequently emigrated from Hungary and made a name for themselves, such as art historian Frigyes Antal, art sociologist Arnold Hauser, and the founder of sociology of knowledge, Károly (Karl) Mannheim. The intellectual prosperity of the beginning of the century formed the background of the early intellectual development of both the philosopher of economics Károly (Karl) Polányi and the philosopher of science Mihály (Michael) Polányi (see Polányi, M.).

After the First World War and the revolution, development stopped. Many figures who had been important at the beginning of the century were forced to emigrate. The most significant character between the two wars was Ákos Pauler. Beginning as a positivist, then embracing the different tendencies of new idealism, Pauler built a system that reflected the Platonism of pure logic. His works - the two most important being the *Bevezetés a filozífiába (Introduction to Philosophy)* (1920) and *Logika (Logic)* (1925) - still breathe an air of cold perfection, but they were considered somewhat anachronistic even in their time.

The dominant trend of the period was the *Geistesgeschichte* approach which transformed new idealism into philosophy of history and culture. Its most prominent representative, Lajos Prohászka, combined this direction with neo-Hegelianism. His work on national characterology *A vándor és a bujdosó (The Wanderer and the Exile)* (1934), born under German influences, became the major philosophical bestseller of the age. Beside the broadly interpreted *Geistesgeschichte* tendency, all other schools lacked influence: this was the case with Neo-Kantianism, which had gradually lost ground, as well as with German existentialism, which occasionally appeared in Hungary.

4 Decades of communism

Following the few years of democratic digression immediately after the end of the Second World War, Hungary was ruled by communist regimes for about four decades. This essentially determined the institutional system, the dominant direction and the achievements of Hungarian philosophy.

The institutional system underwent a radical transformation. The traditional 'bourgeois' institutions of philosophy were closed. The Hungarian Philosophical Society fell apart, and the publication of *Athenaeum* stopped. The institutional field of philosophical activity was provided by the system of Marxism-Leninism departments (a compulsory subject at all universities), and by the Philosophical Institute founded within the Hungarian Academy of Sciences (see Marx, K.; Dialectical materialism; Marxism, Western). The only periodical which was devoted entirely to philosophical scholarship was the *Magyar Filozófiai Szemle (Hungarian Philosophical Review)* (started in 1957). The philosophical life which developed within the institutional system was usually at a very low level, was imbued with ideology, and yet functioned with shared subjects, a set terminology and a unified audience.

The dominant philosophical trend was Soviet-Marxism, which commanded a monopolistic position at the beginning of the period but lost force after 1956 in the post-Stalinist period (see Marxist philosophy, Russian and Soviet). As for its function, this was the ideology that served as legitimization of political power. Its tenets were to refer to the works of Marx and Engels, and were called 'scientific'; any criticism of the official worldview was

condemned as ‘revisionist’ or ‘anti-scientific’. It consisted of two parts: dialectical materialism - the ontology providing cosmology and natural philosophy; and historical materialism, the history of philosophy yielding a teleological explanation of history.

In this period, there were considerable philosophical accomplishments, but mainly on the margins and outside official philosophy. The former case is exemplified by the activity of the Budapest School, which formed around the later György Lukács. The critical Marxist group’s watchword was ‘the renaissance of Marxism’, announced by Lukács in the 1960s (see Lukács, G.). They produced notable works which, however, eventually resulted not so much in the renewal, but rather in the deconstruction of Marxism. Of the circle, Ágnes Heller published ethical works; Ferenc Fehér wrote aesthetic studies; György Márkus attempted to introduce analytic philosophy and expounded the notion of human essence in the works of the young Marx; and Mihály Vajda worked on the Marxist reception of phenomenology and on an analysis of fascist movements. The book *Hogyan lehetséges kritikai gazdaságtan? (How Can Critical Economics be Possible?)* (1970-72) - written by Márkus with his students, György Bence and János Kis and not published at the time - criticizes the foundations of Marxian socialism. Another book, *Dictatorship over Needs* (1981), co-authored by Márkus, Heller and Fehér after their emigration, provides a critical Marxist explanatory theory of Soviet socialism. The philosopher of science Imre Lakatos who became famous after his emigration, was also a member of the circle for some time. Two works outside official philosophy deserve mention. Political thinker István Bibó attempted to reconcile liberalism and socialism (see Socialism). He was generally condemned to silence, but at the time of the postcommunist transition his writings served as the most influential ideological reference. The philosopher Béla Hamvas, who started with a philosophy of crisis and ended up with a new sacral metaphysics, also relied on oriental traditions. His enigmatic essays, published years after they had been written, were a major inspiration to literary philosophical essay-writing.

5 After 1989

As far as philosophy is concerned, the changes of 1989 were preceded by earlier developments. In the last decade, the dominance of Marxist-Leninist philosophy became merely apparent. There was no opportunity for open criticism of official ideology, but members of the philosophical elite all joined contemporary Western trends. The significance of the changes in 1989 is due to the fact that these endeavours could become public within the new institutional system.

During the 1980s, ideological research was no longer compulsory in the Institute of Philosophy, and the Institute finally became a place for ideology-free philosophizing in official terms. In 1987, the social organization for philosophy, the Hungarian Philosophical Society, was reorganized. Since 1989, Marxism-Leninism has not been a compulsory subject at the universities; the departments that had earlier taught dialectic and historical materialism here started to give courses on the history of philosophy, introductions to philosophy and applied philosophy adjusted to the profile of their university. The publication of philosophical periodicals also was revitalized. About a dozen new philosophy-oriented periodicals were founded, the most important of which being the *Athenaeum*, which contains collections of translations on given topics and *Gond*, which publishes philosophical essays. To compensate for the loss of several years a number of new editions of older translations have been published, together with new translations of classic and contemporary works and the writings of Hungarian authors blacklisted for ideological and political reasons. Following the renewal in the institutional system, there has been a renaissance in philosophy: philosophy seems to have become a significant area of contemporary Hungarian intellectual life.

The three most important schools in the country today appear to be analytic philosophy, hermeneutics and the postmodern approach (see Analytic philosophy; Hermeneutics; Postmodernism). Scholars of analytic philosophy are interested in the philosophy of language, the philosophy of science and the history of philosophy. They are primarily influenced by Wittgenstein and Kuhn. Those who work in the hermeneutic tradition are mainly attached to Heidegger, and to Gadamer in aesthetic research. The most prominent representatives of postmodernism are members of the former Budapest School who returned from external or internal emigration, and (following some US and French examples) renewed the tradition of philosophical essay-writing. The rediscovery of a Hungarian philosophical tradition is an important element of this philosophical renaissance.

LÁSZLÓ PERECZ

References and further reading

Alexander, B. (1896) ‘Magyar filozófia története’ (History of Philosophy in Hungary), in *Pallas Nagy Lexikona (Pallas Great Encylopedia)*, Budapest: Pallas, vol. XII, 16-22.(Relevant to §§1-3. A concise and useful summary about the development to the beginning of the twentieth century.)

Apáczai, Cs.J. (1655) *Magar encyclopaedia (Hungarian Encylopedia)*, Utrecht: Ex officina Joannis Waesberg; Bucharest: Kriterion Könyvkiadó, 1977.(The first philosophical work in Hungarian.)

Bence, Gy., Kis, J. and Márkus, Gy. (1970-2) *Hogyan lehetséges kritikai gazdaságtan? (How Can Critical Economics be Possible?)*, Budapest: T-Twins-Lukás Archívum, 1992.(Critical analysis of the foundations of Marxian socialism.)

Eötvös, J. (1851-4) *Der Einfluss der herrschenden Ideen des 19. Jahrhunderts auf den Staat (The Influence of the Predominant Ideas of the Nineteenth Century on the State)*, Vienna: Jasper Hügel u Manz, 1851; Leipzig: Brockhaus, 1854; Budapest: Magyar Helikon, 1981. (One of the first manifestations of a conservative reinterpretation of liberalism in contemporary terms.)

Erdélyi, J. (1857) *A hazai bölcsészet jelene (The Present State of Domestic Philosophy)*, Sárospatak: Főiskolai Nyomda; reprinted in *Filozófiai és esztétikai írások (Philosophical and Aesthetical Writings)*, Budapest: Akadémiai Kiadó, 1981, 25-102.(A polemical treatise representing Hegelianism on the niveau of the age.)

Erdélyi, J. (1981) *A hazai bölcsészet múltja (The Past of Domestic Philosophy)* in *Filozófiai és esztétikai írások (Philosophical and aesthetical writings)*, Budapest: Akadémiai Kiadó, 142-299.(Relevant to §1. Written in the nineteenth century, but still the most thorough account of the early history of philosophy in Hungary today.)

Fehér, F., Márkus, G. and Heller, Á. (1981) *Dictatorship over Needs: An Anaylsis of Soviet Societies*, Oxford: Blackwell, 1983.(A critical Marxist explanatory theory of Soviet socialism.)

Hanák, T. (1976) *Die marxistische Philosophie und Soziologie in Ungarn (Marxist Philosophy and Sociology in Hungary)*, Stuttgart: Ferdinand Enke Verlag.(Relevant to §4, a characterization of the institutional system and achievements of official Marxism.)

Hanák, T. (1982) *Az elfelejtett reneszánsz. A magyar filozófiai gondolkodás századunk elsö felében (The Forgotten Renaissance. Hungarian Philosophical Thinking in the First Half of This Century)*, Bern: EPMSZ. (Relevant to §3, a monograph on the non-Marxist philosophy.)

Hanák, T. (1990) *Geschichte der Philosophie in Ungarn. Ein Grundriß (History of Philosophy in Hungary: An Outline)*, Munich: Dr Rudolf Trofenik Verlag.(The most up-to-date monograph on the subject.)

Hanák, T. and Lendvai, L.F. (1996) ‘Ungarn’ (Hungary), in H. Dahm and A. Ignatow (eds) *Geschichte der philosophischen Traditionen Osteuropas (The History of Eastern European Traditional Philosophy)*, Darmstadt: Wissenschaftliche Buchgesellschaft, 451-513.(An overall view of non-Marxist philosophical development.)

Hermann, I. (ed.) (1982) *A magyar filozófiai gondolkodás a két világháború között (Hungarian Philosophical Thinking Between the Two World Wars)*, Budapest: Kossuth Könyvkiadó.(A collection of essays on topics related to §3.)

Kiss, E. and Nyíri, K. (eds) (1977) *A magyar filozófiai gondolkodás a századelõn (Hungarian Philosophical Thinking at the Beginning of the Century)*, Budapest: Kossuth Könyvkiadó.(A collection of essays on topics related to §3.)

Kornis, Gy. (1907) ‘A magyar bölcseleti mũnyelv fejlõdése’ (‘The development of Hungarian philosophical terminology’), *Magyar Nyelv (Hungarian language)* III: 97-104, 145-53, 193-201, 241-50, 301-7, 348-56. (Relevant to §§1-2. The history of the formation of Hungarian philosophical terminology.)

Kovesi, J. (1967) ‘Hungarian Philosophy’, in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York and London: Macmillan, vol. 4, 93-5.(A short summary in English.)

Nyíri, J.C. (1986) ‘Österreich und Ungarn: Eine philosophisch-soziologische Skizze’ (Austria and Hungary: a Socio-philosophical Account), in *Gefühl und Gefüge. Studien zum Entstehen der Philosophie Wittgensteins (Feeling and Construct: Studies in the Origination of Wittgenstein’s Philosophy)*, Amsterdam: Editions Rodopi, 11-30.(Relevant to §3. An interpretation of developments at the turn of the century from the point of view of the sociology of knowledge.)

Nyíri, J.C. (1994) ‘Tradition and Bureaucratic Lore: Lessons from Hungary’, in B. Smith (ed.) *Philosophy and Political Change in Eastern Europe*, La Salle, IL: The Monist Library of Philosophy.(Relevant to §5; an apt characterization of post-Marxist tendencies.)

Pauler, Á. (1920) *Bevezetés a filozófiába (Introduction to Philosophy)*, Budapest: Pantheon. (The main work of the most significant thinker between the two world wars.)

Pauler, Á. (1925) *Logika (Logic)*, Budapest: Eggenberger.(A Platonistic work of pure logic.)

Perecz, L. (1994) ‘Két kísérlet. Az egyezményesek és Alexander Bernát a nemzeti filozófiáról’ (‘Two attempts:

the reconciliation school and Bernát Alexander on the national philosophy'), *Gond (Care)* 5-6: 104-28. (Relevant to §2. A sketch of the history of the 'national philosophy' idea.)

Perecz, L. (1995) 'Az elsõ másfél évtized. A Magyar Filozófiai Társaság indulása' ('The first one and a half decades: the beginning of the Hungarian Philosophical Society'), *Magyar Tudomány (Hungarian science)* CII: 735-42.(Relevant to §3. A description of the foundation of the Hungarian Philosophical Society as a part of the institutionalization of Hungarian Philosophy.)

Prohászka, L. (1936) *A vándor és a bujdosó (The Wanderer and the Exile)*, Budapest: Dunántúli Pécsi Egyetemi Könyvkiadó éd Nyomda; reprinted Szeged: Universum Kiadó, 1990.(A national characterology, conceived under German influences.)

Rácz, L. (1928) 'Die Ungarische Philosophie' ('Hungarian Philosophy') in K. Oesterreich (ed.) *Friedrich Ueberwegs Grundriß der Geschichte der Philosophie, Fünfter Teil, Die Philosophie des Auslandes*, Berlin: Mittler & Sohn, 348-57.(Relevant to §§1-3, a summary with much information.)

Sándor P. (1973) *A magyar filozófia története, 1900-1945 (The History of Philosophy in Hungary)*, vols I-II, Budapest: Magvetõ Könyvkiadó.(Relevant to §3. Describes the philosophy of the first half of the twentieth century, with much information but with an ideologically biased view.)

Steindler, L. (1988) *Ungarische Philosophie im Spiegel ihrer Geschichtsschreibung (Hungarian Philosophy Through its Historical Writings)*, Freiburg and Munich: Verlag Karl Alber.(An up-to-date monograph on research of the history of philosophy in Hungary.)

Tamás, G.M. (1994) 'A filozófia és az új Magyarország' ('The philosophy of the new Hungary'), in *Másvilág (Other-world)*, Budapest: Új Mandátum Kiadó, 211-18.(Relevant to §§4-5, a witty account of Marxist and post-Marxist philosophy.)

Tankó, B. (1934-5) 'Hungarian Philosophy', *Acta literarum ac scientiarium Regiae Universitatis Hungariae Francisco Josephinae, Sectio Philosophica*, V: 119-36.(Relevant to §§1-3, an account in English from a *Geistesgeschichte* approach.)

Veres I. and Mezei B. (eds) (1994) *Gondolatok gondolatokról. Elöadások a magyarorsz gi filozófia történetérõl (Ideas on Ideas: Lectures on the History of Philosophy in Hungary)*, Miskolc: Felsõmagyarország Kiadó.(A collection of papers with the most recent results of research into the history of Hungarian philosophy.)

Hus, Jan (*c*.1369-1415)

From his appointment as rector of the Bethlehem chapel in Prague in 1402 until his execution at the Council of Constance in 1415, Jan Hus advanced the goals of an ecclesiastical reform movement with Czech national overtones. Hus' ministerial and academic posts provided a broad platform for his leadership. He preached tenaciously against clerical abuses. At the University of Prague he taught philosophical and ecclesiological doctrines which, his opponents charged, were taken from the radical Oxford reformer, John Wyclif. Whereas Wyclif's philosophical realism (for example, the indestructibility of 'being', led him to adopt several positions, condemned as heretical, Hus' polemic, in which he castigated the fiscalization and bureaucratization of the papacy, sprang more from his ideals of evangelical minority and apostolic poverty.

Hus came from Husinec in southern Bohemia. He was trained at the University of Prague, earning his BA in 1393 and his MA in 1396. Divisions in the university faculty ran along national origins, German and Czech, and were reflected in schools of philosophy. Hus sided with his compatriots who adhered to the realism of Wyclif as opposed to the nominalism embraced by the German counterparts. Along with his teaching duties, Hus preached in the large Bethlehem chapel. He followed in the footsteps of John Milic of Kromeríz and Matthew of Janov, forerunners of the Czech reform effort, committed to the reformation of the church according to the ideal of the early apostolic community.

The conflict between the faculties escalated into denunciations of heresy. Accused of Wycliffite errors, Hus and his fellow Czech academics responded by overturning the university voting procedure which had favoured the German students. After King Wenceslaus IV of Bohemia confirmed this reversal by mandate in 1409, the Germans left *en masse* for other German universities, where they continued to challenge the 'Wycliffites' by seeking their condemnation at the imperial court and the Roman Curia. Moreover, Hus encountered stiff resistance from the indigenous clerical hierarchy, stung by his strident accusations of careerism and moral laxity. Ultimately, he broke with the archbishop of Prague, Zbynek Zajíc, over the papal election at the Council of Pisa in 1409. The archbishop, although at first sympathetic to reform, maintained obedience to Pope Gregory XII of Rome, while Hus supported the newly elected Alexander Zbynek; in return the Pisan pope decreed that preaching cease in all Prague chapels. Hus ignored the pope's prohibition; even more, he publicly defended condemned articles of Wyclif. Outraged at Hus' defiance, Zbynek declared him excommunicate.

Hus alienated King Wenceslaus when Alexander's successor, John XXIII, called for a crusade in 1411 against the supporters of Gregory XII. The campaign was to be financed by the sale of indulgences, the profits of which Wenceslaus shared. Hus decried the crusade as fratricide; the King withdrew his protection for Hus. Meanwhile, Hus' antagonists at the Pisan curia secured papal excommunication against him, extended by an interdict on Prague.

In deference to his sovereign, Hus left the city in 1412 and found refuge among nobles in southern Bohemia. From there he defended his position in Latin and Czech treatises, the most important titled *Tractatus de ecclesia (On the Church)*. Like Wyclif before him, Hus defines the universal Church as the 'totality of the predestined'. The Church on earth consists of the elect and those foreknown by God to be damned. This militant Church includes the Roman church with the pope and the cardinals who form 'the chief part of its dignity', provided they adhere to Christ's example. Hus rejects papal claims to headship; Christ alone remains head of the Church. Anyone not prepared to follow Christ in word and deed is a 'disciple of the Antichrist'.

When summoned to appear before the general council at Constance (1414-18), Hus received from the emperor Sigismund a guarantee of safe conduct to and from Constance. Instead of being allowed to present his case, Hus was imprisoned. Confronted with a series of loose quotes, deemed Wycliffite, that were gleaned from his works, Hus responded that he would abandon any position if it could be proven heretical on the basis of the Scriptures. Herein lay the main cause for his condemnation: Hus repudiated all pronouncements, either papalist or conciliarist, formulated without scriptural foundation. Refusing to recant, Hus was burned as a heretic on 6 July 1415. His death rallied his Czech countrymen and ignited a broad-based reform movement. Its conservative wing wrested major concessions from the Council of Basel in 1431-49, which were acknowledged in the Compacts of Prague (1436).

See also: Luther, M.; Wyclif, J.

CURTIS V. BOSTICK

List of works

Jan Hus (*c*. 1393-1415) *Magistri Iohannis Hus, Opera Omnia*, ed. V. Flajshans, Prague: J. Bursík and J.R. Vilímek, 1903-7, 8 vols.(A collection of Hus' academic works including his commentary on Lombard's *Sentences*.)

Jan Hus (*c*.1393-1415) *Documenta Magistri Johannis Hus*, ed. F. Palacky, Prague: F. Tempsky, 1869.(Collection of correspondence and documents related to Hus' career.)

Jan Hus (1413) *Tractatus de ecclesia (On the Church)*, ed. S.H. Thomson, *Magistri Johannis Hus, Tractatus de Ecclesia*, Boulder, CA: University of Colorado Press, 1956.(The critical edition of Hus' most significant and controversial work.)

References and further reading

De Vooght, P. (1960) *L'hérésie de Jean Huss (The Heresy of Jan Hus)*, Louvain: Publications Universitaires de Louvain.(For a counter-balancing perspective.)

Kaminsky, H. (1967) *A History of the Hussite Revolution*, Berkeley and Los Angeles, CA: University of California Press.(For further study of the Czech revolution after Hus' execution.)

Novotný, V., and Kybal, V. (1919-31) *M. Jan Hus, Zivot a Ucení (The Life and Teachings of Master Jan Hus)*, Prague: J. Laichter, 5 vols.(The foundational biography for any in-depth study.)

Spinka, M. (1968) *John Hus, a Biography*, Princeton, NJ: Princeton University Press. (Well-written but partisan.)

Husserl, Edmund (1859-1938)

Through his creation of phenomenology, Edmund Husserl was one of the most influential philosophers of our century. He was decisive for most of contemporary continental philosophy, and he anticipated many issues and views in the recent philosophy of mind and cognitive science. However, his works were not reader-friendly, and he is more talked about than read.

Husserl was born in Moravia, received a Ph.D. in mathematics while working with Weierstraß, and then turned to philosophy under the influence of Franz Brentano. He was particularly engaged by Brentano's view on intentionality and developed it further into what was to become phenomenology. His first phenomenological work was Logische Untersuchungen (Logical Investigations) (1900-1). It was followed by Ideen (Ideas) (1913), which is the first work to give a full and systematic presentation of phenomenology. Husserl's later works, notably Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins (On the Phenomenology of the Consciousness of Internal Time) (1928), Formale und transzendentale Logik (Formal and Transcendental Logic) (1929), Kartesianische Meditationen (Cartesian Meditations) (1931) and Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie (Crisis of the European Sciences) (partly published in 1936), remain largely within the framework of the Ideas. They take up topics that Husserl only dealt with briefly or were not even mentioned in the Ideas, such as the status of the subject, intersubjectivity, time and the lifeworld.

Brentano had characterized intentionality as a special kind of directedness upon an object. This leads to difficulties in cases of hallucination and serious misperception, where there is no object. Also, it leaves open the question of what the directedness of consciousness consists in. Husserl therefore endeavours to give a detailed analysis of those features of consciousness that make it as if of an object. The collection of all these features Husserl calls the act's 'noema'. The noema unifies the consciousness we have at a certain time into an act that is seemingly directed towards an object. The noema is hence not the object that the act is directed towards, but is the structure that makes our consciousness be as if of such an object.

The noemata are akin to Frege's 'third world' objects, that is, the meanings of linguistic expressions. According to Husserl, 'the noema is nothing but a generalization of the notion of meaning [Bedeutung] to the field of all acts' ([1913] 1950: 3, 89). Just as distinguishing between an expression's meaning and its reference enables one to account for the meaningful use of expressions that fail to refer, so, according to Husserl, can the distinction between an act's noema and its object help us overcome Brentano's problem of acts without an object.

In an act of perception the noema we can have is restricted by what goes on at our sensory surfaces, but this constraint does not narrow our possibilities down to just one. Thus in a given situation I may perceive a man, but later come to see that it was a mannequin, with a corresponding shift of noema. Such a shift of noema is always possible, corresponding to the fact that perception is always fallible. These boundary conditions, which constrain the noemata we can have, Husserl calls 'hyle'. The hyle are not objects experienced by us, but are experiences of a kind which we typically have when our sense organs are affected, but also can have in other cases, for example under the influence of fever or drugs.

In our natural attitude we are absorbed in physical objects and events and in their general features, such as their colour and shape. These general features, which can be shared by several objects, Husserl calls essences, or 'eidos' (Wesen). Essences are studied in the eidetic sciences, of which mathematics is the most highly developed. We get to them by turning our attention away from the concrete individuals and focusing on what they have in common. This change of attention Husserl calls 'the eidetic reduction', since it leads us to the eidos. However, we may also more radically leave the natural attitude altogether, put the objects we were concerned with there in brackets and instead reflect on our own consciousness and its structures. This reflection Husserl calls 'the transcendental reduction', or 'epoché'. Husserl uses the label 'the phenomenological reduction' for a combination of the eidetic and the transcendental reduction. This leads us to the phenomena studied in phenomenology, that is, primarily, the noemata.

The noemata are rich objects, with an inexhaustible pattern of components. The noema of an act contains constituents corresponding to all the features, perceived and unperceived, that we attribute to the object, and moreover constituents corresponding to features that we rarely think about and are normally not aware of,

features that are often due to our culture. All these latter features Husserl calls the 'horizon' of the act. The noema is influenced by our living together with other subjects where we mutually adapt to one another and come to conceive the world as a common world in which we all live, but experience from different perspectives. This adaptation, through empathy (Einfühlung), was extensively studied by Husserl.

Husserl emphasizes that our perspectives and anticipations are not predominantly factual: 'this world is there for me not only as a world of mere things, but also with the same immediacy as a world of values, a world of goods, a practical world' ([1913] 1950: 3, 1, 58). Further, the anticipations are not merely beliefs - about factual properties, value properties and functional features - but they also involve our bodily habits and skills.

The world in which we find ourselves living, with its open horizon of objects, values, and other features, Husserl calls the 'lifeworld'. It was the main theme of his last major work, The Crisis of the European Sciences, of which a part was published in 1936. The lifeworld plays an important role in his view on justification, which anticipates ideas of Goodman and Rawls.

1 Life

Edmund Gustav Albrecht Husserl was born of Jewish parents in Prossnitz (now Prostejov in the Czech Republic) in Moravia, in what was then Austria-Hungary on 8 April 1859. He was thus of the same age as Dewey and Bergson.

Husserl's early interests lay in the direction of mathematics and science. In 1876 he began studying mathematics and astronomy at the University of Leipzig. After three semesters he transferred to the University of Berlin in order to study with Weierstraß, Kronecker and Kummer, a trio that made Berlin a centre in the mathematical world during that period. After three years in Berlin he left for Vienna, where he received his doctorate in January 1883. He then returned to Berlin in order to become an assistant for Weierstraß. However, Weierstraß became ill, and after just one semester in Berlin Husserl entered military service for a year, spending most of it in Vienna. A growing interest in religious questions made him decide in 1884 to study philosophy with Franz Brentano in Vienna, who inspired him to go into philosophy full-time and exerted a decisive influence on his later phenomenology.

Husserl studied with Brentano until 1886, when Brentano advised him to go to Halle, where one of Brentano's earlier students, Carl Stumpf, was teaching philosophy and psychology. Husserl habilitated in Halle in 1887 and remained there as a *Privatdozent* until 1901, when he became Associate Professor (*außerordentlicher Professor*) in Göttingen, and in 1906 Full Professor. In 1916 he went to Freiburg, where he taught until he retired in 1928. He died in Freiburg on 27 April 1938.

Husserl's first philosophical work was his Habilitation dissertation, *On the Concept of Number*, which was printed, but not published, in 1887. This was incorporated into the first three chapters of his *Philosophy of Arithmetic*, whose first volume was published in 1891. A second volume was announced, but never came. Instead, Husserl underwent a radical philosophical reorientation. He gave up his main project in *Philosophy of Arithmetic*, which had been to base mathematics on psychology. Instead, he developed his lasting philosophical achievement, phenomenology, which was first presented in *Logische Untersuchungen (Logical Investigations)*, arriving in two volumes in 1900 and 1901. In 1905-7 he introduced the idea of a transcendental reduction and gave phenomenology a turn towards transcendental idealism. This new version of phenomenology was expounded in *Ideen (Ideas)* (1913), and is the most systematic presentation of phenomenology.

Husserl's notable later works were *Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins (On the Phenomenology of the Consciousness of Internal Time)* (1928), *Formale und transzendentale Logik (Formal and Transcendental Logic)* (1929), which Husserl characterized as his most mature work, and *Kartesianische Meditationen (Cartesian Meditations)* (1931). The first part of his *Krisis der europäischen Wissenschaften (Crisis of the European Sciences)* was published in 1936, but the main part of this work and about 40,000 pages of manuscripts were left after his death. These manuscripts, together with Husserl's family and his library, were rescued from Germany by the Belgian Franciscan Van Breda, who established the Husserl Archive in Louvain, where the material is now accessible to researchers. Copies of the manuscripts are kept in other Husserl archives in various parts of the world. Gradually, the most important parts of Husserl's papers and scholarly editions of his published works are being published in the series *Husserliana*. In addition, *Erfahrung und Urteil (Experience and*

Judgment) was prepared by Husserl's assistant Ludwig Landgrebe in consultation with Husserl, and appeared shortly after Husserl's death in 1938. Husserl's main works are available in good English translations.

2 Intentionality

The central theme of phenomenology is intentionality. All of phenomenology can be regarded as an unfolding of the idea of intentionality (see Intentionality). Husserl's interest in intentionality was inspired by his teacher, Franz Brentano. However, there are many differences between Husserl's treatment of this notion and that of Brentano. This section deals first with these differences, then goes on to further features of Husserl's notion of intentionality reaching beyond the issues considered by Brentano.

Husserl retains the following basic idea of Brentano's: 'We understand by intentionality the peculiarity of experiences to be "consciousness of something"' ([1913] 1950: 3, 1, 188; Husserl's emphasis). Husserl's formulation comes close to Brentano's oft-quoted passage from *Psychology from an Empirical Point of View*:

> Every mental phenomenon is characterized by what the scholastics in the Middle Ages called the intentional (and also mental) inexistence of an object, and what we could also call, although in not entirely unambiguous terms, the reference to a content, a direction upon an object.
>
> (1874: 1, 2, 85)

However, there is already an important difference between Brentano and Husserl at this starting-point. While Brentano says straightforwardly that for every act there is an object towards which it is directed, Husserl focuses on the 'of'-ness of the act. There are two reasons for this difference: First, Husserl wants to get around the difficulties connected with acts that lack an object. Second, he aims to throwing light on what it means for an act to be 'of' or 'about' something. Let us begin by discussing these two differences.

Acts that lack an object. Brentano's thesis may seem unproblematic in the examples Brentano considered: just as when we love there is somebody or something that we love, so there is something that we sense when we sense, something we think of when we think, and so on. However, what is the object of our consciousness when we hallucinate, or when we think of a centaur? Brentano insisted that even in such cases our mental activity, our sensing or thinking, is directed towards some object. The directedness has nothing to do with the reality of the object, he held. The object is contained in our mental activity, 'intentionally' contained in it. And Brentano defined mental phenomena as 'phenomena which contain an object intentionally'.

Not all of Brentano's students found this lucid or satisfactory, and the problem continued to disturb both them and Brentano. Brentano struggled with it for the rest of his life, and suggested, among other things, a translation theory, giving Leibniz credit for the idea: when we describe an act of hallucination, or of thinking of a centaur, we are only apparently referring to an object. The apparent reference to an object can be translated away in such a way that in the full, unabbreviated description of the act there is no reference to any problematic object. There are two weaknesses of Brentano's proposal. First, unlike Russell later, Brentano does not specify in detail how the translation is to be carried out (see Russell, B. §9). Second, if such a translation can be carried out in the case of hallucinations and so on, then why not carry it out everywhere, even in cases of normal perception? What then happens to the doctrine of intentionality as directedness upon an object?

One of Brentano's students, Alexius Meinong (§§2-4), suggested a simple way out. In his *Gegenstandstheorie* Meinong maintained that there are two kinds of objects, those that exist and those that do not exist. Hallucinations, like normal perception, are directed towards objects, but these objects do not exist. Brentano was not happy with this proposal. He objected that, like Kant, he could not make sense of existence as a property that some objects have and others lack.

Husserl's solution was, as noted, to emphasize the 'of'. Consciousness is always consciousness *of* something. Or better, consciousness is always *as if of* an object. What matters is not whether or not there is an object, but what the features are of consciousness that makes it always be as if of an object. These three words, 'as if of' are the key to Husserl's notion of intentionality. To account for the directedness of consciousness by saying only that it is directed towards an object leaves us in the dark with regard to what that directedness is. This leads us to the second reason for why Husserl diverged from Brentano. Husserl wanted to throw light on just this issue: what does the directedness of consciousness consists in? He made it a theme for a new discipline: the discipline of

phenomenology.

What is directedness? To get a grip on what the directedness of consciousness consists in - to understand better the word ‘of’, which Husserl emphasized in his definition of intentionality quoted at the beginning of §2 above - let us note that for Husserl intentionality does not simply consist in consciousness directing itself towards objects that are already there. Intentionality for Husserl means that consciousness in a certain way ‘brings it about’ that there are objects. Consciousness ‘constitutes’ objects, Husserl said, borrowing a word from the German Idealists, but using it in a different sense. Above, the phrase ‘bringing about’ was put in quotation marks to indicate that Husserl does not mean that we create or cause the world and its objects. ‘Intentionality’ means merely that the various components of our consciousness are interconnected in such a way that we have an experience as of one object. To quote Husserl:

> an object ‘constitutes’ itself - ‘whether or not it is actual’ - in certain concatenations of consciousness which in themselves bear a discernible unity in so far as they, by virtue of their essence, carry with themselves the consciousness of an identical X.
>
> ([1913] 1950: 3, 1, 313; translation emended)

Husserl’s use, here and in many other places, of the reflexive form ‘an object constitutes itself’, reflects his view that he did not regard the object as being produced by consciousness. Husserl considered phenomenology as the first strictly scientific version of transcendental idealism, but he also held that phenomenology transcends the traditional distinction between idealism and realism, and in 1934 he wrote in a letter to Abbé Baudin: ‘No ordinary "realist" has ever been as realistic and concrete as I, the phenomenological "idealist" (a word which by the way I no longer use)’ (Kern 1964: 276). In the preface to the first English edition of the *Ideas* (1931), Husserl stated:

> Phenomenological idealism does not deny the factual [*wirklich*] existence of the real [*real*] world (and in the first instance nature) as if it deemed it an illusion …. Its only task and accomplishment is to clarify the sense [*Sinn*] of this world, just that sense in which we all regard it as really existing and as really valid. That the world exists…is quite indubitable. Another matter is to understand this indubitability which is the basis for life and science and clarify the basis for its claim.
>
> (1950: 5, 152-3)

To see more clearly what Husserl is after, consider Jastrow and Wittgenstein’s duck/rabbit picture. In order to come closer to Husserl we should modify the example and consider not a picture, but a silhouette of the real animal against the sky. When we see such a silhouette against the sky, we may see a duck or a rabbit. What reaches our eyes is the same in both cases, so the difference must be something coming from us. We structure what we see, and we can do so in different ways. The impulses that reach us from the outside are insufficient to determine uniquely which object we experience; something more gets added.

3 Noema

The structure that makes up the directedness of consciousness, Husserl called the ‘noema’. More accurately, the noema has two main components. First, the ‘object meaning’ that integrates the various constituents of our experience into experiences of the various features of *one* object, and second, the ‘thetic’ component that differentiates acts of different kinds, for example, the act of perceiving an object from the act of remembering it or thinking about it. The thetic component is thereby crucial for the reality-character which we ascribe to the object.

Our consciousness structures what we experience (see Kant, I.). How it structures it depends on our previous experiences, the whole setting of our present experience and a number of other factors. If we had grown up surrounded by ducks, but had never heard of rabbits, we would have been more likely to see a duck when confronted with the duck/rabbit silhouette; the idea of a rabbit would not have occurred to us.

The structuring always takes place in such a way that the many different features of the object are experienced as connected with one another, as features of one and the same object. When, for example, we see a rabbit, we do not merely see a collection of coloured patches, various shades of brown spread out over our field of vision (incidentally, even seeing coloured patches involves intentionality, since a patch is also a kind of object, but a different kind of object from a rabbit). We see a rabbit, with a determinate shape and a determinate colour, with the

ability to eat, jump and so on. It has a side that is turned towards us and one that is turned away from us. We do not see the other side from where we are, but we see something which has another side.

That seeing is intentional, or object-directed, means just this, that it is as if of an object: the near side of the object we have in front of us is regarded as a side of a thing, and the thing we see has other sides and features that are co-intended, in the sense that the thing is regarded as more than just this one side. The object meaning of the noema is the comprehensive system of determinations that gives unity to this manifold of features and makes them aspects of one and the same object.

It is important at this point to note that the various sides, appearances or perspectives of the object are constituted together with the object. There are no sides and perspectives floating around before we start perceiving, which are then synthesized into objects when intentionality sets in. There are no objects of any kind, whether they be physical objects, sides of objects, appearances of objects or perspectives of objects without intentionality. And intentionality does not work in steps. We do not start by constituting six sides and then synthesize these into a die; we constitute the die and the six sides of it in one step.

We should also note that when we experience a person, we do not experience a physical object, a body, and then infer that a person is there. We experience a fully fledged person, we are encountering somebody who structures the world, experiences it from their own perspective. Our noema is a noema of a person; no inference is involved. Seeing persons is no more mysterious than seeing physical objects, and no inference is involved in either case. When we see a physical object we do not see sense-data or the like and then infer that there is a physical object there, but our noema is the noema of a physical object. Similarly, when we see an action, what we see is a fully fledged action, not a bodily movement from which we infer that there is an action.

The word ‘object’ must hence be taken in a very broad sense. It comprises not only physical things, but also, as we have seen, animals, and likewise persons, events, actions and processes, and sides, aspects and appearances of such entities.

Essences. Husserl distinguishes between physical objects and processes, which are temporal and normally also spatial, and essences (*Wesen*) or eidos, which are features that the object can share with other objects, such as the triangularity of a triangle or the greenness of a tree. For Husserl, an object’s essence is therefore not something unique to that object, as it is for many other philosophers. Mathematics is the most highly developed study of essences.

Noema and meaning. The features of the noema that we have mentioned, in particular the role it plays in the analysis of acts without objects and the way it accounts for the object-directedness of acts, make it natural to compare the noema to the meaning of linguistic expressions. This comparison and the ensuing way of reading Husserl has been contested. However, it is well supported by textual and systematic considerations, and it is now often regarded as the standard way of interpreting Husserl. One factor contributing to this has been Husserl’s own statement, in a manuscript, that ‘the noema is nothing but a generalization of the notion of meaning (*Bedeutung*) to the field of all acts’ ([1913] 1950: 5, 89).

Noesis. The noema is an abstract structure that can in principle be the same from act to act, in the unlikely case that at two different occasions we should have the same kind of experience of the same object from the same point of view, with exactly the same anticipations, and so on. An act has a noema in virtue of comprising a kind of experience that Husserl calls a ‘noesis’. The noema is the meaning given in an act, Husserl says, while the noesis is the meaning-giving aspect of the act. There is hence a close parallelism between noema and noesis. The relation between noema and noesis bears some similarity to the type/token relation in Peirce (see Type/token distinction). The noesis is a temporal process, in which the noema ‘dwells’.

4 Hyle; filling; evidence

In acts of perception, the noema that we can have is restricted by what goes on at our sensory surfaces, but the restriction does not narrow our possibilities down to just one. Thus in a given situation I may perceive a man, but later come to see that the man was a mannequin, with a corresponding shift of noema. Such a shift of noema is always possible, corresponding to the fact that perception is always fallible. These boundary conditions, which constrain the noemata we can have, Husserl calls ‘hyle’. The hyle are not objects experienced by us, but are

experiences of a kind which we typically have when our sense organs are affected, but also can have in other cases, for example, under the influence of fever or drugs.

In the case of an act of perception, its noema can also be characterized as a very complex set of expectations or anticipations concerning what kind of experiences we will have when we move around the object and perceive it, using our various senses. We anticipate different further experiences when we see a duck and when we see a rabbit. In the first case we anticipate, for example, that we will feel feathers when we touch the object, while in the latter case we expect to find fur. When we get the experiences we anticipate, the corresponding component of the noema is said to be ‘filled’. In all perception there will be some filling: the components of the noema that correspond to what presently ‘meets the eye’ are filled, and similarly for the other senses.

Such anticipation and filling is what distinguishes perception from other modes of consciousness, such as imagination or remembering. If we merely imagine things, our noema can be of anything whatsoever. In perception, however, our sensory experiences are involved; the noema has to fit in with our sensory experiences. This eliminates a number of noemata which I could have had if I were just imagining. In your present situation you can probably not have a noema corresponding to the perception of an elephant. This does not reduce the number of perceptual noemata you can have just now to one, for example, of having a book in front of you.

It is a central point in Husserl’s phenomenology that I can have a variety of different perceptual noemata that are compatible with the present impingements upon my sensory surfaces. In the duck/rabbit case this was obvious, for we could go back and forth at will between having the noema of a duck and having the noema of a rabbit. In most cases, however, we are not aware of this possibility. Only when something untoward happens, when I encounter a ‘recalcitrant’ experience that does not fit in with the anticipations in my noema, do I start seeing a different object from the one I thought I saw earlier. My noema ‘explodes’, to use Husserl’s phrase, and I come to have a noema quite different from the previous one, with new anticipations. This is always possible, he says. Perception always involves anticipations that go beyond what presently ‘meets the eye’, and there is always a risk that we may go wrong, regardless of how confident and certain we might feel.

When some components of the noema are filled, we have ‘evidence’. Evidence comes in degrees, depending on how much of the noema is filled. Husserl discusses two kinds of perfect evidence: ‘adequate’ evidence, where every component in the noema is filled, with no unfilled anticipations, and ‘apodictic’ evidence, where the negation of what seems to be the case is self-contradictory. After some vacillation Husserl ended up holding that we can never attain any of these kinds of perfect evidence - we are always fallible.

5 Intuition

Husserl uses the term ‘intuition’ (*Anschauung*) for any act where an object is experienced as ‘given’, that is, as really there. Earlier philosophers have used the word ‘intuition’ in a variety of ways, mostly about some sort of direct, non-inferential insight. Perception has usually been classified as a kind of intuition. A key issue in medieval philosophy as well as in rationalism and empiricism was whether there are other sorts of such insight. Kant defined ‘intuition’ as a representation which ‘relates immediately to its object and is singular’ (*Critique of Pure Reason* 1781/87: A320; B376-7). Bernard Bolzano developed this idea with great precision. For Husserl, an intuition is an act where we are constrained in how we constitute its objects, such as we typically are in perception, which is one of his two varieties of intuition. He calls the other variety ‘essential insight’ (*Wesensschau*). The object is here a general feature, an essence. For Husserl, as for Kant, intuition is a key kind of evidence in mathematics. This, then, is what Husserl means by the mysterious-sounding term ‘*Wesensschau*’. One might still claim that there is no such thing, but it is difficult to reject the notion once one agrees that the object of an act is underdetermined by what reaches our senses, and one accepts the correlated idea of intentionality.

6 The reductions; phenomenology

Husserl distinguishes between several so-called ‘reductions’. First, there is the ‘eidetic’ reduction, which we perform each time we pass from focusing on an individual physical object to focusing on one of its essences (eidos). This kind of reduction has been carried out in mathematics since its beginning, and Husserl conceived of other eidetic sciences in addition to mathematics. Second, a reduction that is distinctive for phenomenology is a special kind of reflection. Instead of focusing on the normal objects of our acts, be they physical objects, actions,

persons or general features that many objects can have in common, we reflect on the structures of our own consciousness and study the noemata, the noeses or the hyle. The noemata, the noeses and the hyle have two important features: we are normally not aware of them, and they are a *sine qua non* for the appearance of a world. Entities with these two features are called ‘transcendental’. The reduction that leads to them, where the ordinary objects are bracketed, is therefore called the ‘transcendental’ reduction. Husserl also calls it the ‘epoché’, using a word that the ancient sceptics used for refraining from taking a stand. We study the features of the act that make it seem to have an object and do not ask whether or not it actually has one. Husserl got the idea of the transcendental reduction in 1905. It marks the transition from the early phenomenology of the *Logical Investigations* to the ‘idealist’ phenomenology of the *Ideas* and later works.

The ‘phenomenological’ reduction, finally, is the combination of the eidetic reduction and the transcendental reduction. That is, it is a reduction that leads us from acts directed towards physical objects via acts directed towards essences to acts directed towards the noema, noesis and hyle of acts directed towards essences. Husserl sometimes takes the two steps in the inverse order, starting with a transcendental reduction and then focusing on the essential traits of the noema, noesis and hyle. The end product is not quite the same, but the phenomenological reduction can presumably be either.

Phenomenology is the study of the transcendental elements in our experience that are uncovered through the phenomenological reduction: the noema, the noesis and the hyle. In phenomenology, all these three elements are studied, with emphasis on the noematic/noetic structures. Husserl carried out detailed analyses of temporal structures and how they are constituted, in *On the Phenomenology of the Consciousness of Internal Time*, on the structures that are basic to logic and mathematics, in *Formal and Transcendental Logic* and *Experience and Judgment*, and on intersubjectivity and the processes whereby we come to constitute a common world, in *Cartesian Meditations* and in thousands of pages of manuscripts, the most important of which have been collected by Iso Kern in *Husserliana*, vols 13-15.

For Husserl, phenomenology is a study of the subjective perspective. In science one aims for objectivity and endeavours to arrange observations and experiments in such a way as to minimize differences between different observers. Phenomenology focuses on the subjective, on the manner in which each subject structures or ‘constitutes’ the world differently, on the basis of different experiences and cultural background, but also on the basis of adaptation to other subjects through interaction and communication.

7 The past

We constitute not only the different properties of things, but also the relation of the thing to other objects. If, for example, I see a tree, the tree is conceived of as something which is in front of me, as perhaps situated among other trees, as seen by other people than myself, and so on. It is also conceived of as something which has a history: it was there before I saw it, it will remain after I have left, or perhaps it will eventually be cut down and transported to some other place. However, like all material things, it does not simply disappear from the world.

My consciousness of the tree is in this way also a consciousness of the world in space and time in which the tree is located. My consciousness constitutes the tree, but at the same time it constitutes the world in which the tree and I are living. If my further experience makes me give up the belief that I have a tree ahead of me because, for example, I do not find a tree-like far side or because some of my other expectations prove false, this affects not only my conception of what there is, but also my conception of what has been and what will be. Thus in this case, not just the present, but also the past and the future are reconstituted by me. To illustrate how changes in my present perception lead me to reconstitute not just the present, but also the past, Husserl uses an example of a ball which I initially take to be red all over and spherical. As it turns, I discover that it is green on the other side and has a dent:

> the sense of the perception is not only changed in the momentary new stretch of perception; the noematic modification streams back in the form of a retroactive cancellation in the retentional sphere and modifies the production of sense stemming from earlier phases of the perception. The earlier apperception, which was attuned to the harmonious development of the ‘red and uniformly round’, is implicitly ‘reinterpreted’ to ‘green on one side and dented’.
>
> (1938: 96)

Husserl held that time and space are constituted. In *On the Phenomenology of the Consciousness of Internal Time* and various manuscripts that have been published in Volume 10 of *Husserliana* he gives a highly interesting analysis of the way objective time is constituted (Miller 1984).

8 Values; practical function

So far we have focused on the factual properties of things. However, things also have *value* properties, and these properties are constituted in a corresponding manner. The world within which we live is experienced as a world in which certain things and actions have a positive value, others a negative. Our norms and values, like our beliefs, are subject to change. Changes in our views on matters of fact are often accompanied by changes in our evaluations.

Husserl emphasizes that our perspectives and anticipations are not predominantly factual. We are not living a purely theoretical life. According to Husserl, we encounter the world around us primarily ‘in the attitude of the natural pursuit of life’, as ‘living functioning subjects involved in the circle of other functioning subjects’ (1950: 4, 375). Husserl says this in a manuscript from 1917, but he has similar ideas about the practical both earlier and later. Thus in the *Ideas* he says: ‘ this world is there for me not only as a world of mere things, but also with the same immediacy as a world of values, a world of goods, a practical world’ ([1913] 1950: 3, 1, 58).

In later manuscripts, particularly from 1917 onwards, Husserl focused more and more on the role of the practical and the body in our constitution of the world. Just as he never held that we first perceive sense-data, or perspectives or appearances, which are then synthesized into physical objects, or that we first perceive bodies and bodily movements and then infer that there are persons and actions, so it would be a grave misunderstanding of Husserl to attribute to him the view that we first perceive objects that have merely physical properties and then assign a value or a practical function to them. Things are directly experienced by us as having the features - functional and evaluational as well as factual - that are of concern for us in our natural pursuit of life.

In our discussion of the hyle we characterized the noema of an act of perception as a very complex set of expectations or anticipations concerning what kind of experiences we will have when we move around the object and perceive it. We should note that these experiences depend not only on our sensory organs, but also on the movements of our body, on our bodily skills and our familiarity with various kinds of practical activities. In numerous passages Husserl talks about practical anticipations and the role of kinesthesis in perception and bodily activity (Føllesdal 1979).

9 Horizon

When we are experiencing an object, our consciousness is focused on this object, and the rest of the world and its various objects are there in the background as something we ‘believe in’ but are not presently paying attention to. The same holds for most of the inexhaustibly many features of the object itself. All these further features of the object, together with the world in which it is set, make up what Husserl calls the ‘horizon’ of that experience. The various features of the object, which are co-intended, or also-meant, but not at the focus of our attention, Husserl calls the ‘inner horizon’, while the realm of other objects and the world to which they all belong, he calls the ‘outer horizon’.

The horizon is of crucial importance for Husserl’s concept of justification, which we shall discuss later. What is particularly significant is the hidden nature of the horizon. As we noted, the horizon is that which is not attended to. Take as an example our ‘expectation’ that we will find a floor when we enter a room. Usually, we have not even thought about there being a floor. Typically, we cannot even recall when we first acquired the corresponding ‘belief’ or ‘anticipation’. According to Husserl, there may never have been any occasion when we actually judged there to be a floor in some particular room. Still we have come to ‘anticipate’ a floor, not in the sense of consciously expecting one, but in the sense that if we entered the room and there were none, we would be astonished. In this example we would easily be able to tell what was missing, in other cases our ‘anticipations’ are so imperceptible that we just may feel that something has gone awry, but not be able to tell what it is.

Words like ‘belief’ and ‘anticipate’ are clearly not the proper ones here, since they have overtones of something being conscious and thought about. Both English and German seem to lack words for what we want to get at here: Husserl uses the words ‘*antizipieren*’, ‘*hinausmeinen*’ and ‘*vorzeichnen*’.

10 Intersubjectivity

Throughout his life, Husserl emphasized that the world we intend and thereby constitute is not our own private world, but an intersubjective world, common to and accessible to all of us. Thus in the *Ideas* he writes:

> I continually find at hand as something confronting me a spatiotemporal reality [*Wirklichkeit*] to which I belong like all other human beings who are to be found in it and who are related to it as I am.
>
> ([1913] 1950: 3, 1, 61)

Husserl's studies of intersubjectivity focus in particular on the processes by which we experience others as experiencing subjects, like ourselves, and adapt our anticipations to those that we take them to have. Thanks to this, our way of constituting the world is not solipsistic, but we constitute the world as a shared world, which we each experience from our different perspective. A notion of objectivity arises, we may come to regard ourselves as deviant, for example, as colour-blind or as cognitively biased, and we also experience ourselves as confronted with a reality to which our beliefs and anticipations have to adapt. In works that remain largely unpublished, Husserl started to develop an ethics based in part on a study of the objectifying processes whereby objective ethical principles and norms arise from our subjective likes and dislikes.

Husserl stresses the shared, intersubjective nature of the world, particularly in §29 of the *Ideas*, which he entitles 'The "Other" Ego-subjects and the Intersubjective Natural Surrounding World'. There he says:

> I take their surrounding world and mine Objectively as one and the same world of which we are conscious, only in different ways [*Weise*].… For all that, we come to an understanding with our fellow human beings and together with them posit an Objective spatiotemporal reality.
>
> ([1913] 1950: 3, 1, 60)

In the later works one finds similar ideas, particularly in the many texts that have been collected by Iso Kern in the three volumes of the *Husserliana* devoted to intersubjectivity, but also in many other works, for example in the *Crisis*:

> Thus in general the world exists not only for isolated men but for the community of men; and this is due to the fact that even what is straightforwardly perceptual is communal.
>
> (1936, 1954: 6, 166)

Husserl discusses in great detail empathy and the many other varieties of intersubjective adaptation that enable us to intend a common, intersubjective world. (See the three volumes on intersubjectivity referred to above.)

11 Existence

The passages quoted in §10 above express a further feature of Husserl's notion of intentionality which is rarely discussed, in spite of its importance: intentionality does not just involve directedness upon an object, but also a 'positing' of the object, corresponding to the two components of the noema discussed in §3 above. The object is experienced as real and present, as remembered, or as merely imagined, and so on. In the passages just quoted, Husserl said, 'I continually find at hand as something confronting me a spatiotemporal reality', and 'we come to an understanding with our fellow human beings and together with them posit an Objective spatiotemporal reality'. The same point is stressed also when he discusses the lifeworld in the Crisis:

> the lifeworld, for us who wakingly live in it, is always there, existing in advance for us, the 'ground' of all praxis, whether theoretical or extratheoretical. The world is pregiven to us, the waking, always somehow practically interested subjects, not occasionally but always and necessarily as the universal field of all actual and possible praxis, as horizon. To live is always to live-in-certainty-of-the-world.
>
> (1936, 1954: 6, 145)

Husserl discusses this *thetic* character of intentionality, and, correspondingly, of the noema, in many of his books and manuscripts. He was particularly concerned with what gives reality-character to the world. Like William James, whom he had read already when he made the transition to phenomenology in the mid-1890s, he stressed the importance of the body, and the inflictions upon our body, for our sense of reality. As James put it: 'Sensible

vividness or pungency is then the vital factor in reality' (1890: 2, 301). Husserl could also have subscribed to James's observation that 'the *fons et origo* of all reality, whether from the absolute or the practical point of view, is thus subjective, is ourselves' (1890: 2, 296-7).

This latter passage from James gets a double meaning in Husserl which expresses the core of his view of the reality of the world: the subjective (ourselves) is the *fons et origo* of all reality in two senses, a transcendental and an empirical: we constitute the world as real through our intentionality, and the reality-character we give it is derived from our being not merely transcendental subjects, but empirical subjects with a body immersed in a physical world.

12 The lifeworld

The idea of Husserl's that has become most widely known is that of the lifeworld. In particular, the word 'lifeworld' (*Lebenswelt*) itself has gained wide currency. It was used by Simmel and others before Husserl. After the Second World War it became a favourite word of many social scientists, who used it in many different senses. Several of them refer to Husserl without seeming to have studied his philosophy and therefore without knowing the many important features that the lifeworld has in his thought.

The first place Husserl uses the word 'lifeworld' in print is in his latest work, the *Crisis*, of which the first two parts were published in 1936. The rest of this unfinished work, containing the important third part, with the main discussion of the lifeworld, was not published until 1954, but it was known to some of Husserl's students and followers, including Maurice Merleau-Ponty, who came to the Husserl Archives in Louvain to study this part in April 1939.

Interpreters of Husserl differ widely in their views on the lifeworld. It is often thought that it constitutes a major break in Husserl's development, from the 'early' Husserl of the *Ideas* to the 'late' Husserl of the *Crisis*. Is it such a break? And second, what exactly is the lifeworld and what role does it play in phenomenology? On the former question the answer is a definite 'No'. The lifeworld is fully compatible with Husserl's earlier philosophy, and there is even a definite place for it in his phenomenology from its beginning. Husserl touches upon the lifeworld repeatedly in his earlier work and he gradually deepens and modifies his views on it, as he did with everything else in his phenomenology. Instead of regarding the lifeworld as a break with Husserl's earlier philosophy, we should view it as intimately connected with the other main themes in phenomenology. Properly to understand the lifeworld with all its nuances it is important to appreciate fully the connection between it and the rest of Husserl's philosophy.

The lifeworld arises from the distinction between the natural attitude and the transcendental or phenomenological attitude, which Husserl introduced in 1905. The first appearance of the notion for which he later introduced the term 'lifeworld' occurs shortly thereafter, in his lectures 'Fundamental Problems in Phenomenology' in 1910-11, that is, already before the *Ideas*. Husserl begins these lectures with an extended discussion of 'the natural attitude and the "natural world concept"'. Here he says:

> It could also be shown that philosophical interests of the highest dignity require a complete and comprehensive description of the so-called *natural world concept*, that of the natural attitude, on the other hand also that an accurate and profound description of this kind is not easily carried out, but on the contrary would require exceptionally difficult reflections.
>
> (1950: 13, 124-5)

Husserl here borrows the phrase 'natural world concept', which he emphasizes, from Richard Avenarius, whom he discusses later in the lecture. In a manuscript from 1915, Husserl describes this world in the following way (following Avenarius):

> All opinions, justified or unjustified, popular, superstitious, scientific, all relate to the already *pregiven world*.… All theory relates to this immediate givenness and can have a legitimate sense only when it forms thoughts which do not offend against the general sense of the immediately given. No theorizing may offend against this sense.
>
> (1950: 13, 196; emphasis added)

In the following years, Husserl repeatedly returns to this and related themes, using various labels that sometimes allude to other philosophers who had propounded similar ideas, such as Nietzsche. Quite often he uses Avenarius' phrase 'natural world'. In a manuscript from 1917, which appears to be the first place where he uses the word 'lifeworld', he introduces this new word as equivalent to the former: 'The lifeworld is the natural world - in the attitude of the natural pursuit of life are we living functioning subjects involved in the circle of other functioning subjects' (1950: 4, 375; the manuscript dates from 1917, but was copied during the first half of the 1920s, and it is possible that the word 'lifeworld' appeared then).

Gradually during the 1920s and especially in the 1930s the lifeworld becomes a central theme in Husserl's writings, until his discussion culminates in the *Crisis* in 1936. One aim of this work was to provide a new and better access to phenomenology, through the notion of the lifeworld. The lifeworld is for Husserl our natural world, the world we live in and are absorbed by in our everyday activities. A main aim of phenomenology is to make us reflect on this world and make us see how it is constituted by us. Through the phenomenological reduction phenomenology will take us out of our natural attitude where we are absorbed by the world around us, into the phenomenological, transcendental attitude, where we focus on the noemata of our acts - on our structuring of reality.

Pregivenness. In the passage just quoted from Husserl's 1915 manuscript, Husserl says that the world is pregiven (*vorgegeben*). This point is also discussed in the *Ideas*, where Husserl notes that

> In my waking consciousness I find myself in this manner at all times, and without ever being able to alter the fact, in relation to the world which remains one and the same, though changing with respect to the composition of its contents. It is continually 'on hand' for me and I myself am a member of it.
>
> ([1913] 1950: 3, 1, 58)

and a few pages later the passage that was quoted earlier, in the section on intersubjectivity:

> I continually find at hand as something confronting me a spatiotemporal reality [*Wirklichkeit*] to which I belong like all other human beings who are to be found in it and who are related to it as I am.
>
> ([1913] 1950: 3, 1, 61)

Also the passage from §37 of the *Crisis* that was quoted in the section on existence above expresses this same idea:

> The lifeworld…is always there, existing in advance for us, the 'ground' of all praxis, whether theoretical or extratheoretical. The world is pregiven to us…
>
> (1936, 1954: 6, 145)

Science and the lifeworld. A contested point in Husserl scholarship is the relation between the lifeworld and the sciences. Many interpreters of Husserl like to find an opposition to the sciences in the lifeworld. However, such a disdain for the sciences is out of character with Husserl's background in and continued interest in mathematics and science. It also accords poorly with the texts, which give us a different and more intriguing picture. According to Husserl, the lifeworld and the sciences are intimately connected, in three different ways:

(1) The sciences are *part* of the lifeworld. This comes out most explicitly and clearly in *Experience and Judgment*, where Husserl says:

> everything which contemporary natural science has furnished as determinations of what exists also belong to us, to the world, as this world is pregiven to the adults of our time. And even if we are not personally interested in natural science, and even if we know nothing of its results, still, what exists is pregiven to us in advance as determined in such a way that we at least grasp it as being in principle scientifically determinable.
>
> (1938: 39)

Similar statements are also found elsewhere in Husserl's work, for example in the *Crisis*: 'Now the scientific world - [the subject matter of] systematic theory - …like all the worlds of ends "belongs" to the lifeworld' (1936, 1954: 6, 460).

(2) Scientific statements get their *meaning* by being embedded in the lifeworld. This was stressed by Husserl already in the manuscript from 1915, quoted in §12 above:

> All opinions, justified or unjustified, popular, superstitious, scientific, all relate to the already *pregiven world*.… All theory relates to this immediate givenness and can have a legitimate *sense* only when it forms thoughts which do not offend against the general sense of the immediately given. No theorizing may offend against this sense.
>
> (1950: 13, 196; emphasis added)

(3) The sciences are *justified* through the lifeworld. There is an interplay between this point and point (1) above; the sciences are justified because they belong to the lifeworld, and at the same time they belong to the lifeworld because they are conceived of as describing the world, as claiming to be true:

> Though the peculiar accomplishment of our modern objective science may still not be understood, nothing changes the fact that it is a validity for the lifeworld, arising out of particular activities, and that it belongs itself to the concreteness of the lifeworld.
>
> (1936, 1954: 6, 136)

And similarly:

> all these theoretical results have the character of validities for the lifeworld, adding themselves as such to its own composition and belonging to it even before that as a horizon of possible accomplishments for developing science. The concrete lifeworld, then, is the grounding soil [*der gründende Boden*] of the 'scientifically true' world and at the same time encompasses it in its own universal concreteness.
>
> (1936, 1954: 6, 134)

13 Ultimate justification

This brings us to the final theme of this presentation of Husserl's phenomenology: the role of the lifeworld in justification (see Justification). The traditional interpretation of Husserl attributes to him a 'foundationalist' position: he is alleged to hold that we can reach absolute certainty with regard to a number of matters, particularly in philosophy. However, there is considerable evidence that Husserl had a view on justification similar to that of Goodman and Rawls (Føllesdal 1988). An opinion is justified by being brought into 'reflective equilibrium' with the *doxa* of our lifeworld. This holds even for mathematics: 'mathematical evidence has its source of meaning and of legitimacy in the evidence of the lifeworld' (1936, 1954: 6, 143).

A major puzzle that many see in this idea of justification is, 'How can appeal to the subjective-relative doxa provide any kind of justification for anything? It may help to resolve disagreements, but how can it serve as justification?' Husserl answers by pointing out that there is no other way of justifying anything, and that his way is satisfactory:

> What is actually first is the 'merely subjective-relative' intuition of prescientific world-life. For us, to be sure, this 'merely' has, as an old inheritance, the disdainful colouring of the *doxa*. In prescientific life itself, of course, it has nothing of this; there it is a realm of good verification and, based upon this, of well-verified predicative cognitions and of truths which are just as secure as is necessary for the practical projects of life that determine their sense. The disdain with which everything 'merely subjective and relative' is treated by those scientists who pursue the modern ideal of objectivity changes nothing of its own manner of being, just as it does not change the fact that the scientist himself must be satisfied with this realm whenever he has recourse, as he unavoidably must have recourse, to it.
>
> (1936, 1954: 6, 127-8)

So far, this is a mere claim. However, Husserl elaborates his view in other parts of his work. His key observation, which is an intriguing contribution to our contemporary discussion of ultimate justification, is that the 'beliefs', 'expectations' or 'acceptances' on which we ultimately fall back are unconsidered, and in most cases have never been considered. Every claim to validity and truth rests upon this 'iceberg' of unconsidered prejudgmental acceptances discussed earlier. One would think that this would make things even worse. Not only do we fall back on something that is uncertain, but on something that we have not even thought about, and have therefore never subjected to conscious testing. Husserl argues, however, that it is just the unconsidered nature of the lifeworld that makes it the ultimate ground of justification. 'Acceptance' and 'belief' are not attitudes that we decide to have

through any act of judicative decision. What we accept, and the phenomenon of acceptance itself, are integral to our lifeworld, and there is no way of starting from scratch, or 'to evade the issue here through a preoccupation with aporia and argumentation nourished by Kant or Hegel, Aristotle or Thomas' (1936, 1954: 6, 134). Only the lifeworld can be an ultimate court of appeal: 'Thus alone can that ultimate understanding of the world be attained, behind which, since it is ultimate, there is nothing more that can be sensefully inquired for, nothing more to understand' ([1929] 1974: 17, 249) (see Phenomenology, epistemic issues in).

14 Influence

Husserl's phenomenology has been a major influence on philosophy in our century, primarily on the continent, but since the 1970s also in the United States, Britain and several other countries. Husserl's immediate successor in Freiburg, Martin Heidegger, conceived of *Being and Time* (1927) as a phenomenological study and dedicated it to Husserl. Also Jean-Paul Sartre received strong impulses from Husserl, particularly from Husserl's idea that our material surroundings do not uniquely determine our noema. Sartre developed this idea into a philosophy of freedom, notably in *Being and Nothingness* (1943), which has the subtitle 'A Phenomenological Essay on Ontology'. Also Emmanuel Levinas, Paul Ricoeur and several other French philosophers were heavily influenced by Husserl. A new generation of young French and German philosophers is now combining Husserl scholarship with work on systematic issues in epistemology, philosophy of language and philosophy of mind.

Husserl's conception of the lifeworld become important for the so-called 'new hermeneutics' (Heidegger and Gadamer; see Hermeneutics) and for the methodology of the humanities and the social sciences (Schütz, Luckmann), largely because it provides a framework for discussing the subjective perspective and the many features of our way of structuring the world of which we are unaware and that often reflect the culture in which we have grown up. The issues connected with intersubjectivity and Husserl's exploration of the various ways in which we adapt to one another and come to conceive the world as a common world were pursued by several of his students, notably Edith Stein, in her dissertation *On the Problem of Empathy* (1917). His ideas about the role of the body, of kinesthesis and of practical activity recur in different versions in Heidegger's existentialism and in Merleau-Ponty's phenomenology. Merleau-Ponty in particular is generous in the credit he gives Husserl.

Husserl's many students and followers explored a number of other themes in Husserl and applied his ideas in a variety of fields. Thus Roman Ingarden used them in aesthetics, Aron Gurwitsch and several others in the study of perception. Husserl's views have led to new developments in psychology and psychotherapy. They have influenced philosophers of mathematics, including Gödel (see Føllesdal 1995), and they are beginning to have an impact on the philosophy of mind and on cognitive science.

See also: Phenomenological movement

DAGFINN FØLLESDAL

List of works

Husserl, E. (1950-) *Husserliana*, ed. H.L. Van Breda, The Hague: Nijhoff; since 1989, Dordrecht: Kluwer.(The standard critical edition of Husserl's work. References to Husserl's works are by volume, page number and line number in this edition.)

Husserl, E. (1891) *Philosophie der Arithmetik. Logische und psychologische Untersuchungen*, vol. 1, in *Husserliana*, vol. 12, 1970.(Husserl's first book, where he attempted to found arithmetic on psychology. He gave up this project, and the planned second volume of the work never appeared.)

Husserl, E. (1900-1) *Logische Untersuchungen*, in *Husserliana*, vols 18-19, 1950; trans. J.N. Findlay, *Logical Investigations*, London: Routledge & Kegan Paul, 2 vols, 1970.(Husserl's first phenomenological work. Volume 1 is a criticism of psychologism, volume 2 consists of six studies of key themes in phenomenology.)

Husserl, E. (1913) *Ideen*, 3 vols, in *Husserliana*, vols 3-5, 1950; revised edn in two parts, 1976; *Ideas*, vol. 1, trans. F. Kersten, The Hague: Nijhoff, 1982; vol. 2, trans. R. Rojcewicz and A. Schuwer, Dordrecht: Kluwer, 1989; vol. 3, trans. T.E. Klein and W.E. Pohl, Dordrecht: Kluwer, 1980.(Husserl's main systematic presentation of phenomenology. Difficult, but *the* book to read if one should read only one book by Husserl.)

Husserl, E. (1928) *Vorlesungen zur Phänomenologie des inneren Zeitbewußtseins*, in *Husserliana*, vol. 10, 1966; trans. J.B. Brough, *On the Phenomenology of the Consciousness of Internal Time (1893-1917)*, The Hague: Nijhoff, 1991.(Husserl's main texts on time and time-consciousness.)

Husserl, E. (1929) *Formale und transzendentale Logik*, in *Husserliana* vol. 17, 1974; trans. D. Cairns, *Formal*

and Transcendental Logic, The Hague: Nijhoff, 1969. (This book, which Husserl characterized as one of his most mature works, aims at providing a phenomenological foundation for formal logic.)

Husserl, E. (1931) *Kartesianische Meditationen*, in *Husserliana* vol. 1, 1950; 2nd edn, 1973; improved edn, ed. E. Ströker, *Philosophische Bibliothek*, vol. 291, Hamburg: Felix Meiner, 1977; trans. D. Cairns, *Cartesian Meditations*, The Hague: Nijhoff, 1960.(A 'Cartesian' approach to phenomenology. Focuses especially on the ego and on intersubjectivity. Easiest to read among Husserl's later works.)

Husserl, E. (1936, 1954) *Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*, in *Husserliana* vol. 6, 1954; trans. D. Carr, *The Crisis of European Sciences and Transcendental Phenomenology*, Evanston, IL: Northwestern University Press, 1970.(Presents Husserl's theory of the lifeworld.)

Husserl, E. (1938) *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, ed. L. Landgrebe, Prague: Akademia Verlagsbuchhandlung; trans. J.S. Churchill and K. Amerik, *Experience and Judgment*, Evanston, IL: Northwestern University Press, 1973.(Compiled by Husserl's assistant Ludwig Landgrebe, this work throws light on the lifeworld and several of the themes dealt with in *Formal and Transcendental Logic*.)

Schuhmann, K. and Schuhmann, E. (eds) (1994) *Edmund Husserl Briefwechsel*, The Hague: Nijhoff, 10 vols.(A critical edition of Husserl's correspondence.)

References and further reading

Avenarius, R. (1891) *Der menschliche Weltbegriff*, Leipzig: Reisland.(One source of Husserl's concept of the lifeworld.)

Bell, D. (1990) *Husserl, The Arguments of the Philosophers*, London: Routledge. (Good on the early Husserl.)

Bernet, R., Kern, I. and Marbach, E. (1993) *An Introduction to Husserlian Phenomenology, Studies in Phenomenology and Existential Philosophy*, Evanston, IL: Northwestern University Press.(A good introduction to phenomenology by three eminent Husserl scholars.)

Boehm, R. (1968, 1981) *Vom Gesichtspunkt der Phänomenologie*, in *Phenomenologica* 26 and 83, The Hague: Nijhoff, 2 vols.(Two collections of essays on the epoché and other central topics in Husserl by a most knowledgeable Husserl editor and scholar.)

Brentano, F. (1874) *Psychologie vom empirischen Standpunkt (Psychology from an Empirical Point of View)*, Leipzig: Duncker & Humblot.(This work by Husserl's main teacher had a strong influence on Husserl and many of his contemporaries.)

Dreyfus, H. (ed.) (1982) *Husserl, Intentionality and Cognitive Science*, Cambridge, MA: MIT Press.(A selection of articles on Husserl and cognitive science.)

Føllesdal, D. (1969) 'Husserl's Notion of Noema', *Journal of Philosophy* 66: 680-7. (This and the following two articles present evidence for the interpretation of Husserl on which this entry is based.)

Føllesdal, D. (1979) 'Husserl and Heidegger on the Role of Actions in the Constitution of the World', in E. Saarinen, R. Hilpinen, I. Niiniluoto and M. Provence Hintikka (eds) *Essays in Honour of Jaakko Hintikka*, Dordrecht: Reidel, 365-78.

Føllesdal, D. (1988) 'Husserl on Evidence and Justification', in R. Sokolowski (ed.) *Edmund Husserl and the Phenomenological Tradition: Essays in Phenomenology* (*Studies in Philosophy and the History of Philosophy*, vol. 18), Washington, DC: Catholic University of America Press, 107-29.

Føllesdal, D. (1995) 'Gödel and Husserl', in J. Hintikka (ed.) *From Dedekind to Gödel: Essays on the Development of the Foundations of Mathematics*, Dordrecht: Kluwer, 427-46.

Gurwitsch, A. (1966) *Studies in Phenomenology and Psychology*, Evanston, IL: Northwestern University Press. (Collection of significant articles, particularly on consciousness and on perception.)

Haaparanta, L. (ed.) (1994) *Mind, Meaning and Mathematics. Essays on the Philosophical Views of Husserl and Frege*, Dordrecht: Kluwer.(Brings together several contributions on the interesting relationship between Frege and Husserl.)

Heidegger, M. (1927) *Being and Time*, trans. J. Macquarrie and E. Robinson, *Being and Time*, New York: Harper & Row, 1962.(Heidegger's main work, a phenomenological study dedicated to Husserl.)

Ingarden, R.W. (1931) *The Literary Work of Art*, Evanston, IL: Northwestern University Press, 1973.(The first of Ingarden's many influential studies of the philosophy of art.)

James, W. (1890) *The Principles of Psychology*, New York: Holt.(Husserl studied this work in the 1890s and felt that James had said what he himself wanted to say.)

Kern, I. (1964) *Husserl und Kant*, in *Phenomenologica* 16, The Hague: Nijhoff.(*The* book on Husserl's study of Kant and on the similarities and differences between Husserl's philosophy and that of Kant.)

Künne, W. (1986) 'Edmund Husserl: Intentionalität', in J. Speck (ed.) *Grundprobleme der großen Philosophen, Philosophie der Neuzeit IV*, Göttingen: Vandenhoeck & Ruprecht, 165-215.(A lucid presentation of Husserl's

thought by a philosopher who combines systematic insight with scholarship.)
Lapointe, F.H. (1980) *Edmund Husserl and His Critics. An International Bibliography (1884-1979)*, Bowling Green, OH: Philosophy Documentation Center.(The most comprehensive bibliography on Husserl. However, literature on Husserl is growing rapidly, and bibliographies quickly become outdated.)
Marbach, E. (1993) *Mental Representation and Consciousness. Towards a Phenomenological Theory of Representation and Reference*, Dordrecht: Kluwer.(A discussion of Husserlian notions that have a bearing on cognitive science.)
Meinong, A. (1904) 'Über Gegenstandstheorie', in A. Meinong (ed.) *Untersuchungen zur Gegenstandstheorie und Psychologie*, Leipzig: Barth, 1-50; trans. and ed. R.M. Chisholm, 'The Theory of Objects', in *Realism and the Background of Phenomenology*, Atascadero, CA: Ridgeview, 1982, 76-117.(Presents Meinong's theory of objects. Meinong's ideas also became influential in the English-speaking world through Russell's generous review articles in *Mind* (1894) of Meinong's *On Assumptions* (1902).)
Miller, I. (1984) *Husserl, Perception, and Temporal Awareness*, Cambridge, MA: MIT Press.(A clear and pedagogical discussion of Husserl's view on perception and its relation to time.)
Mohanty, J. (1989) *Transcendental Phenomenology: An Analytic Account*, Oxford: Blackwell.(One of several books by a prominent Husserl scholar who is also familiar with other traditions in philosophy.)
Sartre, J.-P. (1943) *Being and Nothingness: An Essay of Phenomenological Ontology*, New York: Philosophical Library, 1956.(Sartre's first major philosophical work, with the subtitle *An Essay on Phenomenological Ontology*, reflects his study of phenomenology, notably in Germany in 1933-4, encouraged by Raymond Aron.)
Schuhmann, K. (1973) *Die Dialektik der Phänomenologie*, in *Phenomenologica* 56/7, The Hague: Nijhoff, 2 vols. (Two collections of essays by the editor of *Husserl-Chronik* and Husserl's *Briefwechsel*, as well as of the new, improved edition of the *Ideen*.)
Schuhmann, K. (1977) *Husserl-Chronik. Denk- und Levensweg Edmund Husserls*, The Hague: Nijhoff.(A detailed chronology of Husserl's life and activities.)
Smith, D. and McIntyre, R. (1982) *Husserl and Intentionality: A Study of Mind, Meaning and Language*, Dordrecht: Reidel.(Relates Husserl's work to other recent approaches to intentionality. It was the most cited Husserl book during the first years after it was published.)
Smith, B. and Woodruff Smith, D. (eds) (1995) *The Cambridge Companion to Husserl*, Cambridge: Cambridge University Press.(Essays by several prominent Husserl scholars on various aspects of Husserl's thought, followed by useful bibliographies.)
Sokolowski, R. (1970) *The Formation of Husserl's Concept of Constitution*, in *Phenomenologica* 18, The Hague: Nijhoff.(A classic study in English of one of the key notions in phenomenology.)
Sokolowski, R. (ed.) (1988) *Edmund Husserl and the Phenomenological Tradition: Essays in Phenomenology (Studies in Philosophy and the History of Philosophy, vol. 18)*, Washington DC: Catholic University of America Press.(The essays stem from a lecture series at the Catholic University of America in the Fall of 1985.)
Spiegelberg, H. (1960) *The Phenomenological Movement: A Historical Introduction*, in *Phenomenologica* 5/6, The Hague: Martinus Nijhoff, 2 vols; 3rd edn, 1982.(The standard historical survey of the development of the phenomenological tradition from Brentano and Stumpf through Husserl to its mid-century representatives.)
Stein, E. (1917) *Zum Problem der Einfühlung*, Halle; trans. W. Stein, *On the Problem of Empathy*, The Hague: Nijhoff, 1970.(The dissertation of Edith Stein, Husserl's assitant, who later became a Carmelite nun and died in Auschwitz in 1942.)
Ströker, E. (1993) *Husserl's Transcendental Phenomenology*, Stanford, CA: Stanford University Press.(A presentation of the development of Husserl's thought by a leading Husserl scholar.)

Hutcheson, Francis (1694-1746)

Francis Hutcheson is best known for his contributions to moral theory, but he also contributed to the development of aesthetics. Although his philosophy owes much to John Locke's empiricist approach to ideas and knowledge, Hutcheson was sharply critical of Locke's account of two important normative ideas, those of beauty and virtue. He rejected Locke's claim that these ideas are mere constructs of the mind that neither copy nor make reference to anything objective. He also complained that Locke's account of human pleasure and pain was too narrowly focused. There are pleasures and pains other than those that arise in conjunction with ordinary sensations; there are, in fact, more than five senses. Two additional senses, the sense of beauty and the moral sense, give rise to distinctive pleasures and pains that enable us to make aesthetic and moral distinctions and evaluations.

Hutcheson's theory of the moral sense emphasizes two fundamental features of human nature. First, in contrast to Thomas Hobbes and other egoists, Hutcheson argues that human nature includes a disposition to benevolence. This characteristic enables us to be, sometimes, genuinely virtuous. It enables us to act from benevolent motives, whereas Hutcheson identifies virtue with just such motivations. Second, we are said to have a perceptual faculty, a moral sense, that enables us to perceive moral differences. When confronted with cases of benevolently motivated behaviour (virtue), we naturally respond with a feeling of approbation, a special kind of pleasure. Confronted with maliciously motivated behaviour (vice), we naturally respond with a feeling of disapprobation, a special kind of pain. In short, certain distinctive feelings of normal observers serve to distinguish between virtue and vice. Hutcheson was careful, however, not to identify virtue and vice with these feelings. The feelings are perceptions (elements in the mind of observers) that function as signs of virtue and vice (qualities of agents). Virtue is benevolence, and vice malice (or, sometimes, indifference); our moral feelings serve as signs of these characteristics.

Hutcheson's rationalist critics charged him with making morality relative to the features human nature happens at present to have. Suppose, they said, that our nature were different. Suppose we felt approbation where we now feel disapprobation. In that event, what we now call 'vice' would be called 'virtue', and what we call 'virtue' would be called 'vice'. The moral sense theory must be wrong because virtue and vice are immutable. In response, Hutcheson insisted that, as our Creator is unchanging and intrinsically good, the dispositions and faculties we have can be taken to be permanent and even necessary. Consequently, although it in one sense depends upon human nature, morality is immutable because it is permanently determined by the nature of the Deity.

Hutcheson's views were widely discussed throughout the middle decades of the eighteenth century. He knew and advised David Hume, and, while Professor of Moral Philosophy at Glasgow, taught Adam Smith. Immanuel Kant and Jeremy Bentham, among other philosophers, also responded to his work, while in colonial America his political theory was widely seen as providing grounds for rebellion against Britain.

1 Life and works

Francis Hutcheson was born on 8 August 1694 near Saintfield, County Down, Ireland. Although often taken to be the founder of the Scottish Enlightenment, he always considered himself an Irishman. Hutcheson studied first at a classical school in Saintfield, then at an academy in Killyleagh, and finally, for two years, at Glasgow College. Ordained as a minister in the Presbyterian Church of Ireland, Hutcheson followed instead an academic career. In the early 1720s he established a dissenting academy in Dublin, where he remained until called to Glasgow as Professor of Moral Philosophy in 1729. This position he held until his death, having in 1745 declined an offer of a similar position at Edinburgh.

In Dublin, Hutcheson came under the influence of Robert Molesworth, himself a philosophical disciple of the Third Earl of Shaftesbury. In the mid 1720s Hutcheson published papers outlining some of his own views, and others criticizing Thomas Hobbes and Bernard Mandeville, as well as his first book, *An Inquiry into the Original of our Ideas of Beauty and Virtue* (1725a). This work he initially described as a defence of Shaftesbury against an attack by Mandeville. His *Essay on the Nature and Conduct of the Passions and Affections: with Illustrations on the Moral Sense* appeared in 1728. His next work was probably *A System of Moral Philosophy*, written by 1738 but published only posthumously in 1755. His last major work was his *Philosophiae moralis institutio compendiaria* (1742a), a translation (Hutcheson himself was probably the translator) of which, *A Short*

Introduction to Moral Philosophy, appeared in 1747, the year of his death.

Hutcheson corresponded with, and probably met, David Hume, and gave Hume advice, some of which he took, regarding the third volume of his *Treatise of Human Nature*. Notwithstanding these connections, Hutcheson apparently opposed Hume's efforts to be appointed (in 1745) to the chair of moral philosophy in Edinburgh. Hutcheson's students at Glasgow included Adam Smith, author of *The Theory of Moral Sentiments* (1759) and *The Wealth of Nations* (1776).

2 The foundations of morality and the moral sense

Much of Hutcheson's early work may be seen as a contribution to a long-standing debate about the foundations of morality. For over a century before Hutcheson joined the debate, moral theorists had offered fundamentally incompatible accounts of the origin and nature of morality. Every participant in this debate accepted the fact that there are moral phenomena to explain. No participant denied, for example, that there is a set of moral terms (such terms as, in English, 'good', 'evil', 'virtue', 'vice', 'right', 'wrong', 'just', 'unjust') that are competently used by ordinary speakers. Even those philosophers who were said to have denied that morality has a foundation assumed that it is to rational beings (principally humans) and their actions that this set of terms applies, and supposed that ordinary humans do so apply the terms, however much they may disagree about which term to use in any given situation. The controversy raged, however, over the proper characterization of such moral phenomena. For many writers, it was not merely a matter of providing a causal explanation of these phenomena. Even cynics and sceptics could do that. Rather, these writers, who tended to think of themselves as moral realists, demanded that a proper understanding of morality be a part of this explanation. Having concluded that moral differences are both real and unique, they insisted that one could be said to have given an account of the foundations of morality only if one could trace these real and unique moral differences to some set of objective and unique natural or transcendental features adequate to ground such differences in a non-reductive way.

Hutcheson's work illustrates this latter demand. In a preview of his influential *Inquiry*, he says that his new work will include an essay on the foundations of morality, a needed antidote to the socially poisonous views of those (most notably Hobbes, Samuel Pufendorf and John Locke, as we later learn) who suppose that the 'foundation of virtue' is nothing more than fear of punishment. In the *Inquiry* itself Hutcheson develops his criticism of these 'selfish moralists' (egoists, as we would say), and also makes explicit his deeply felt objections to Mandeville's claim that what is called virtue is simply 'the Political Offspring which Flattery begot upon Pride' (1725a).

Although his philosophy owes much to Locke's empiricist approach to ideas and knowledge, Hutcheson was far from satisfied with Locke's account of our moral ideas and our moral psychology. According to Locke, our normative ideas - of beauty and virtue, for example - are complex ideas of mixed modes, and, although formed out of the materials of experience, have no objective reference. These ideas, Locke says, are constructed by our minds, and are neither copies of anything real, nor even made according to the pattern of any real existence. Hutcheson found this anti-realist account of the origin of our moral ideas seriously flawed. Moreover, he also complained that Locke's account of human pleasure and pain was too narrowly focused. Locke had failed to note that there are pleasures and pains other than those that arise in conjunction with ordinary sensations. Indeed, Locke had failed to note that there are more than five senses, and that our additional senses - the sense of beauty and the moral sense - give rise to distinctive pleasures and pains (to approbations and disapprobations) that enable us to make moral distinctions and moral evaluations. Human nature is considerably more complex than Locke had supposed.

As to Hutcheson's disagreement with Hobbes and Mandeville, Hutcheson can be seen to have rejected their pessimistic, cynical view of human nature - in effect, that humans are inherently corrupt - and to have adopted in its place the more optimistic view that human nature incorporates a substantial element of goodness. More particularly, while Hobbes and Mandeville argue that all human acts are motivated by self-interest, Hutcheson argues that humans have, and actually do act from, other-regarding motives, and that the 'selfish theory' - the view that all motivations are self-interested - cannot account for many features of our moral experience. Hutcheson sees the selfish theorists as maintaining that we act only from a regard for our own pleasure, and hence that those things that we call 'good' are simply those things that give us pleasure, and that the actions we call 'morally good' or 'just' are simply those actions that correspond to whatever laws politicians happen to have promulgated. As a matter of pure speculation these claims may be comprehensible, but a more careful survey of our moral evaluations shows them to be false.

Hutcheson notes in the *Inquiry*, for example (1) that although both a generous action and a productive field give us pleasure, there is a significant difference in the pleasures derived from these two things. Our approbation of them differs in such a way that we would think it quite odd or senseless to say that the field is virtuous, but entirely sensible to say that the action is virtuous (1725a: 114-5, 119, 140-41); (2) that while reading history we learn of temporally distant individuals who cannot contribute to our interests or pleasure, and that, as the actions of these individuals vary, so do our responses: we feel approbation toward some, and disapprobation or indifference toward others; (3) that while we benefit from the actions of an individual who treasonously betrays his country to ours, we none the less morally disapprove of that individual and his actions; (4) that while we ourselves may be bribed to perform an action which we think to be morally wrong, we cannot be bribed to *feel* that this same action is right or that we are right to undertake it; (5) that we sometimes feel a moral indignation toward a person who has caused us no injury at all, while on other occasions we find that, although someone has acted in such a way as to injure us, we feel no moral indignation.

Facts of this sort, Hutcheson concludes, establish beyond doubt that the selfish theory is mistaken. They show that there are natural (immediate and unlearned) differences in our responses to actions or events, differences that cannot be accounted for by this theory. Our moral approvals and disapprovals are more subtle than these philosophers have thought. We find that we naturally and routinely make reliable distinctions (1) between *natural good* and *moral good*; (2) between *moral good* and *moral evil*; and (3) between things that are *immediately good* and those only 'advantageous' or *instrumentally good*. Given these important facts, the question becomes, as Hutcheson typically frames it: 'What feature of human nature is presupposed by the fact that we can and do make these distinctions?' His answer: 'The moral sense'. Had we, he writes in the *Inquiry*, no moral sense enabling us to perceive the moral qualities of agents and actions, we could not respond to them as we do. Without a moral sense we might have developed an abstract idea of virtue, but we would not, as we do, actively approve and esteem those who reveal themselves to be virtuous. Had we 'no Sense of moral Good in Humanity…Self-Love, and our Sense of natural Good' would cause us, contrary to fact, always to approve, for example, of the traitor who benefits our cause, and always to disapprove the courageous patriot who harms our cause. But that is not the case; the facts do not fit the egoists' theory. Rather, such facts reveal that 'there is in human Nature a disinterested ultimate Desire of the Happiness of others', an inherent benevolent, and hence moral, concern.

Humans have also an inherent cognitive power that enables them to respond differently to benevolence and self-interest. The human mind is formed in such a way that it can approve or condemn actions or agents without concern for its own pleasure or interest. Just as the Creator has 'determin'd us to receive, by our external Senses, pleasant or disagreeable Ideas of Objects, according as they are useful or hurtful to our Bodys …[so] he has given us a moral Sense, to direct our Actions, and to give us still nobler Pleasures'. Thus if two individuals contribute in similar ways to our wellbeing, but the one acts 'from an ultimate Desire of our Happiness, or Good-will toward us; and the other from Views of Self-Interest, or by Constraint', we have, we find, significantly different responses to these two individuals and their acts. In response to the one we feel gratitude and approbation; to the other we are indifferent. Or, if we know that an individual has benevolent dispositions but has been prevented from exercising them, we count that individual as morally good even though they have not been able to act - even though they have done nothing to benefit us. The nature and complexity of these responses show that we must have a perceptual power, a moral sense, for without such a sense we would assess fields and agents or patriots and traitors only with regard to our own interests and wellbeing.

In the third edition (1729) of the *Inquiry* Hutcheson explicitly denied that he meant to identify virtue and vice with feelings or sentiment. The moral sense relies upon feelings to distinguish virtue or vice, but moral qualities are themselves independent of the observer who feels approbation or disapprobation of them. The 'admired Quality', he says, is a quality of the agent judged, and entirely distinct from the approbation or pleasure of either the approving observer or the agent, and the moral perceptions (the idea or concept) involved 'plainly represents something quite distinct from this Pleasure'. Feelings play a cognitive and a motivating role, but virtue is constituted by the benevolent disposition that gives rise to approbation, and vice by the malevolent or sometimes indifferent dispositions that give rise to disapprobation.

Gilbert Burnet and John Balguy, two early rationalist critics, pronounced themselves satisfied with Hutcheson's good intentions and even with his fundamental conclusion, namely that virtue and vice are, contrary to Hobbes, Locke, and Mandeville, fundamentally real. Burnet, in his Letter to the *London Journal* (1969-71, vol. 7: 6) could

not agree, however, that Hutcheson had found 'the true and solid foundation' of morality, while Balguy (1728-9), all praise for Hutcheson's good nature and good sense, regrets that he makes serious mistakes - mistakes that 'lie at the Foundations of Morality, and like Failures in Ground-work, affect the whole Building'.

Burnet's fundamental complaint is that Hutcheson's account of moral good and evil explains these notions only relatively - only as good or evil things relate to us or affect us - and gives us no guarantee that good and evil have an immutable foundation in the nature of things. Balguy objected that Hutcheson had rested virtue or morality on two features of human nature, a natural affection (a concern for others) and an instinct (the moral sense). These are, he granted, features of humankind, and it is right of Hutcheson to try to represent virtue as something that flows 'unalterably from the Nature of Men and Things'. But, he went on, Hutcheson portrays morality as something arbitrary, dependent upon human features 'that might originally have been otherwise, or even contrary to what they now are; and [that] may at any time be alter'd, or inverted, if the Creator pleases' (1728: 292). Moreover, in Hutcheson's hands morality is resolved into 'mere *Instinct*'. If we are motivated to what we call virtuous acts by an instinct, of what moral merit are the resulting actions? These actions seem to be necessitated, while virtuous acts are always free acts. If Hutcheson should reply that these instincts do not '*force* the Mind, but only *incline* it', then, says Balguy, it will be reason, and not the moral sense, that decides our actions and thus serves as the foundation of morality.

Hutcheson undertook to meet these objections in a series of letters to the *London Journal* (1725b) and in the second part of his *Essay* of 1728. As an ordained clergyman of the Westminster Confession, Hutcheson was unwilling to deny the creative freedom of the Deity. Consequently, he could not in one sense deny that his theory of the moral sense made morality dependent upon the free choice of the Deity. But he could and did argue that, were the Deity's basic nature not in its own way similar to our kindest affections and best moral nature, then he would not have been motivated to create us in the particular manner in which he did actually create us. On the other hand, if we suppose that the Deity does have an analogously kind and moral nature, then we see at once that this would have motivated him to create us as he did. As long as we are satisfied that the Deity is unchanging, and that his is the best nature possible, we can see that our own natures are nothing like the merely arbitrary result of some divine whim. Our natures are, in any relevant sense, necessary and necessarily fixed or unchanging.

To the charge that he had made morality dependent on an instinct that effectively eliminates the free choice required of moral behaviour, Hutcheson replied by distinguishing between instincts of body and certain mental powers or 'affections'. The latter, he insists, are no more destructive of morality than is the '*Determination to pursue Fitness*' that, according to his opponents, characterizes the divine will. Virtue, Hutcheson argues, can be real and meritorious even though perceived by a sense, and chosen because of an affection or instinct. It is not necessary that reason play these roles. Hutcheson's critics were not satisfied by these replies (Richard Price (1758) was to pose similar objections some thirty years later; see Price, R.), but Hutcheson's responses show how far he was from being the moral noncognitivist some commentators have supposed. He undertook to determine the origin of our moral ideas in order 'to prove what we call the *Reality of Virtue*' (1725a: xi). That his inquiry led him to argue that our moral concerns and evaluations depend on certain fundamental dispositions and feelings - that morality has its origin in certain senses and affections found in human nature - should not prevent us from recognizing this realist intent.

3 Practical ethics and influence

Although Hutcheson's later works, his *System* and his *Short Introduction*, reveal his continuing commitment to the moral sense theory and all that presupposes, these works more noticeably reflect his need, after 1729, to offer lectures on a broad range of moral issues - issues of practice as well as principle - to largely adolescent audiences. Making extensive use of writers in the modern natural law tradition (Grotius, Pufendorf, and Richard Cumberland, for example), as well as of Cicero, Hutcheson in these works focuses principally on such practical issues as our duties to God, to humanity, and to ourselves; the law of nature and the rights of individuals; property and contracts; marriage and parenting; and civil government.

Hutcheson's contemporaries and students remembered him as earnestly concerned with civil and religious liberty. He was, for example, an outspoken critic of the Aristotelian or classical theory of slavery, and also of the justification of slavery by conquest. Arguing that there are no natural slaves nor ought there to be slaves by conquest, Hutcheson insists that the '*natural equality of men*' consists chiefly in the fact that 'natural rights belong

equally to all', and that the laws of God and nature prohibit even the most powerful from depriving the least powerful of their rights, or from inflicting any harm on them. The least talented humans have the use of reason and thus have vastly greater capacity for happiness or misery than do animals. All humans 'have strong desires of liberty and property, have notions of right, and strong natural impulses to marriage, families, and offspring, and earnest desires of their safety…. We must therefore conclude, that no endowments, natural or acquired, can give a perfect right to assume powers over others, without their consent'. This, he adds, 'is intended against the doctrine of Aristotle, and some others of the ancients, "that some men are naturally slaves"' (1755, I: 299-301). Equally abhorrent is the view that those taken prisoner in a just war may be justly enslaved as punishment or security against further offence. No 'damage done or crime committed can change a rational creature into a piece of goods void of all right, and incapable of acquiring any [rights]' (1755, II: 202-3). Those who claim that Africans were better off as slaves than they would have been if left in Africa have let custom and the prospect of profit stupefy their consciences until they have lost all sense of natural justice (1755, II: 84-5).

At the heart of Hutcheson's political theory is his endorsement of the principle that the safety of the people is the supreme law. He insists that there are specifiable limits to the powers of the state, and that citizens retain the right to resist the excesses of any form of government and even the right to overthrow and replace a government. Consequently, any government that fails to function for the 'safety and happiness of the whole body' can be legitimately abolished (1742b: 303). Hutcheson explicitly applied these principles to colonies. Colonial subjects also have a right to legitimate - that is, beneficial - government. If they fail to receive such government, and are oppressed, they may justly overthrow their oppressor: 'the people's right of resistance is unquestionable' (1742b: 292).

Hutcheson's writings had a substantial influence in the eighteenth century. His claim that ill-governed colonies have the right to rebel was widely and effectively repeated in colonial north America, and may rightly be taken as having provided many patriots with a philosophical rationale for rebellion. His published views on natural equality and natural rights were reprinted in colonial Philadelphia, where they added philosophical weight to the anti-slavery movement. His moral sense ethics, although much criticized by rationalists, gained partial allegiance from David Hume. His best-known student, Adam Smith, learned from him both moral and economic theory, and credits Hutcheson as 'being the first who distinguished with any degree of precision in what respect all moral distinctions … are founded upon immediate sense and feeling' (Smith 1759: 321). Even Immanuel Kant, a moralist so apparently different, was initially attracted by, and always respectful of, Hutcheson's contributions to moral theory. And if Jeremy Bentham thought the theory of the moral sense was simply totally unconvincing, he none the less adopted as fundamental to his utilitarian theory a principle that Hutcheson had enunciated, namely 'that Action is best, which procures the greatest Happiness for the greatest Numbers' (1725a: 181).

See also: Beauty; Egoism and altruism; Enlightenment, Scottish; Hume, D.; Moral sense theories; Shaftesbury; Slavery; Virtues and vices

DAVID FATE NORTON

List of works

Hutcheson, F. (1969-71) *Collected Works*, ed. B. Fabian, Hildesheim: Olms, 7 vols. (A facsimile edition of the works listed below, with the exception of the items found in (1724).)

Hutcheson, F. (1724) 'Reflections on the Common Systems of Morality', *The London Journal*; repr. in *Francis Hutcheson: Two Texts on Human Nature*, ed. T. Mautner, Cambridge: Cambridge University Press, 1993.(The two texts are Hutcheson's brief account of the systems of morality (1724), and a translation of his inaugural lecture on the social nature of man (1730).)

Hutcheson, F. (1725a) *An Inquiry into the Original of our Ideas of Beauty and Virtue; In Two Treatises*, London and Dublin; revised 1726, 1729, 1738.(The two treatises are: I, *An Inquiry concerning Beauty, Order &c*. and II, *An Inquiry concerning Moral Good and Evil*. These constitute Hutcheson's initial attempt to show that we are equipped with aesthetic and moral senses that enable us to recognize beauty and virtue.)

Hutcheson, F. (1725b) Letters to *The London Journal*.(Hutcheson's letters of June and October were in response to critical letters by Gilbert Burnet the Younger, who objected to Hutcheson's attempt to found morality in human nature. The complete exchange is reprinted in *Collected Works*, 1969-71, vol. 7.)

Hutcheson, F. (1725c) 'Reflections upon Laughter', *The Dublin Weekly Journal*. repr. in *Collected Works*, 1969-71, vol. 7.(Three letters criticizing Hobbes' account of laughter and of human nature.)

Hutcheson, F. (1725d) ‘Remarks upon the Fable of the Bees’, *The Dublin Weekly Journal*; repr. in *Collected Works*, 1969-71, vol. 7.(Three letters criticizing Mandeville’s egoistic account of morality.)

Hutcheson, F. (1728) *An Essay on the Nature and Conduct of the Passions and Affections: With Illustrations on the Moral Sense*, London and Dublin, revised, 1728, 1730, 1742.(Hutcheson’s account of the passions and a further explication and defence of the moral sense theory.)

Hutcheson, F. (1730) *De naturali hominum Socialitate Oratio Inauguralis (Inaugural Lecture on the Social Nature of Man)*, Glasgow; repr. in *Francis Hutcheson: Two Texts on Human Nature*, ed. T. Mautner, Cambridge: Cambridge University Press, 1993.(1993 is a recent translation, with helpful notes.)

Hutcheson, F. (1742a) *Metaphysicae synopsis: ontologiam et pneumatologiam complectens (Synopsis of Metaphysics: Comprising ontology and pneumatology)*, Glasgow; repr. in *Collected Works*, vol. 7, 1969-71. (Apparently lecture notes from Hutcheson course on metaphysics, published without his consent; revised by Hutcheson 1744-5.)

Hutcheson, F. (1742b) *Philosophiae moralis institutio compendiaria*, Glasgow; trans. *A Short Introduction to Moral Philosophy*, Glasgow, 1747. (A comprehensive survey, based on Hutcheson’s lectures, of the elements of ethics and the law of nature. The translation is thought most likely to be by Hutcheson or his colleague and friend James Moor.)

Hutcheson, F. (1755) *A System of Moral Philosophy*, Glasgow.(Written by 1738, but published posthumously by Hutcheson’s son, Francis the Younger. A comprehensive account of morality in three books. Includes discussions of human nature, the supreme good, and of our natural and civil duties.)

References and further reading

Balguy, J. (1728-9) *The Foundation of Moral Goodness: or a Further Inquiry into the Original of our Idea of Virtue*, London, 1728; Part II 1729; facsimile repr. Part I and Part II, New York: Garland Press, 1976.(Includes Balguy’s objections to Hutcheson’s moral theory as well as his alternative rationalist account of the foundations of morality.)

Frankena, W. (1955) ‘Hutcheson’s Moral Sense Theory’, *Journal of the History of Ideas* 16: 356-75.(Argues that Hutcheson is an emotivist.)

Haakonssen, K. (1990) ‘Natural Law and Moral Realism: The Scottish Synthesis’, in *Oxford Studies in the History of Philosophy I* (*Studies in the Philosophy of the Scottish Enlightenment*), 61-85.(Important effort to assess the character of Hutcheson’s moral realism.)

Henrick, D. (1957) ‘Hutcheson und Kant’, *Kant-Studien* 49: 49-69.(An account of Kant’s interest in, and disagreements with, Hutcheson.)

Hope, V. (1989) *Virtue by Consensus: The Moral Philosophy of Hutcheson, Hume, and Smith*, Oxford: Clarendon Press.(An essentially comparative study of Hutcheson, Hume and Adam Smith.)

Hume, D. (1739-40) *A Treatise of Human Nature*, London.(Book III, a classic work of moral theory, was revised in response to Hutcheson’s criticisms.)

Jensen, H. (1971) *Motivation and the Moral Sense in Francis Hutcheson’s Moral Theory*, The Hague: Martinus Nijhoff.(Aims to show the problems arising from Hutcheson’s accounts of motivation and the moral sense.)

Kivy, P. (1976) *The Seventh Sense: A Study of Francis Hutcheson’s Aesthetics and its Influence in 18th-century Britain*, New York: Franklin.(A comprehensive study of Hutcheson’s aesthetic theory.)

Leidhold, W. (1985) *Ethik und Politik bei Francis Hutcheson (Ethics and Politics in Frances Hutcheson)*, Munich: Alber.(A comprehensive study of Hutcheson’s moral philosophy.)

Mandeville, B. (1714-23) *The Fable of the Bees: or, Private Vices Publick Benefits*, London. (Includes several works supportive of Mandeville’s cynical, egoistic account of the origins of morality.)

Moore, J. (1990) ‘The Two Systems of Francis Hutcheson: On the Origins of the Scottish Enlightenment’, in *Oxford Studies in the History of Philosophy I* (*Studies in the Philosophy of the Scottish Enlightenment*), 37-59. (Argues for a distinct change of view between Hutcheson’s earlier and later writings.)

Moore, J. (1994) ‘Hume and Hutcheson’, in M.A. Stewart and J.W. Wright (eds) *Hume and Hume’s Connexions*, Edinburgh: Edinburgh University Press.(A somewhat speculative account of the relationship between Hutcheson and Hume.)

Norton, D.F. (1974) ‘Hutcheson’s Moral Sense Theory Reconsidered’, *Dialogue* 12: 3-23.(A rebuttal of Frankena’s emotivist interpretation of Hutcheson.)

Norton, D.F. (1976) ‘Francis Hutcheson in America’, *Studies in Voltaire and the Eighteenth Century* 151 (5): 1547-68.(A discussion of Hutcheson’s influence in colonial north America.)

Norton, D.F. (1982) *David Hume: Common-Sense Moralist, Sceptical Metaphysician*, Princeton, NJ: Princeton University Press, 1982.(Chapter 2 discusses Hutcheson; Chapter 3 includes a comparison of important aspects

of the moral theories of Hutcheson and Hume.)
Norton, D.F. (1985) ‘Hutcheson’s Moral Realism’, *Journal of the History of Philosophy* 23: 397-418.(Argues against Winkler (1984) that Hutcheson is plausibly interpreted as one kind of moral realist.)
Price, R. (1758) *A Review of the Principal Questions in Morals*, London.(Includes criticisms of Hutcheson’s account of the foundations of morality.)
Raphael, D.D. (1947) *The Moral Sense*, Oxford: Oxford University Press.(Includes a chapter on Hutcheson’s moral theory.)
Robbins, C. (1954) ‘When it is that Colonies May Turn Independent: An Analysis of the Environment and Politics of Francis Hutcheson’, *William and Mary Quarterly*, 3rd series, 11: 214-51.(On Hutcheson’s political philosophy and its influence in colonial America.)
Scott, F.R. (1900) *Francis Hutcheson: His Life, Teaching and Position in the History of Philosophy*, Cambridge: Cambridge University Press.(Outdated, but the only substantial biography of Hutcheson available.)
Smith, A. (1759) *The Theory of Moral Sentiments*, London.(A much-modified account of the moral sense, by Hutcheson’s most famous student.)
Smyth, D. (ed.) (1992) *Francis Hutcheson*, *Fortnight* 308, supplement, Belfast.(Includes brief articles by D.D. Raphael, J. Moore, D.F. Norton, and seven others, on Hutcheson’s life, philosophy and influence.)
Winkler, K. (1984) ‘Hutcheson’s Alleged Realism’, *Journal of the History of Philosophy* 23: 179-94.(Criticism of Norton’s realist interpretation of Hutcheson.)

Huxley, Thomas Henry (1825-95)

Huxley, an English zoologist with strong philosophical interests, originally influenced by K.E. von Baer's embryological typology, became an authority first in invertebrate zoology and then in vertebrate palaeontology. After the publication of Darwin's On the Origin of Species, he proclaimed his acceptance of the theory of evolution, but disagreed on important points and applied common descent - but not natural selection - in his scientific works only after reading Ernst Haeckel's Generelle Morphologie (1866). He published extensively in anthropology, ethnology, philosophy, religion, politics and ethics, and was a great popularizer of science.

Thomas Henry Huxley, a leading exponent of nineteenth-century evolutionism, was born in Ealing, near London. After studying medicine he sailed, in 1846, as assistant surgeon on board HMS *Rattlesnake* in the southern hemisphere. His brilliant voyage observations would further his scientific career, but he secured a stable job only in 1854 at the School of Mines in London. His fame and authority grew with time and by 1883 he was President of the Royal Society, but had to resign in 1886 because of a breakdown.

The traditional view of Huxley as a staunch Darwinist was challenged by scholars in the 1970s, and now he is generally seen as an evolutionist but not of the Darwinian kind. An admirer of German culture and conversant with the language, he interpreted morphology in the light of Baer's embryological typology - animal forms classified according to four separate structural plans determined through embryological development.

Desirous to combat both the Platonic interpretation of animal type and the influence in scientific circles of Richard Owen, the leading British morphologist of the time, Huxley and he were often at odds. Before the publication of Darwin's *On the Origin of Species* (1859) Huxley opposed evolutionary views, both Lamarck's and those of Robert Chambers, the 'anonymous' author of the popular *Vestiges of the Natural History of Creation*.

When the *Origin* appeared, Huxley wrote two predominantly positive reviews of it, after which he often assumed in public the role of official spokesman ('bulldog', he called it) of Darwin's group. He saw two reasons for supporting Darwin's views: (a) it identified a vera causa of the production of species; (b) it was a valid hypothesis of the order of nature based on natural processes. But it was a hypothesis and not yet the theory of species, since there was no decisive experimental proof that natural selection had produced, from one species, another infertile with it. Huxley conceived the value of a theory as 'absolute'; a theory can be isolated and individually assessed in a context of truth and falsity provided by experimentally ascertained facts. For Darwin, on the other hand, if the consequences of a hypothesis are in accordance with a number of facts to be explained, the hypothesis gains explanatory power which justifies its retention. Huxley constantly played down natural selection, and never accepted Darwin's gradualism, but supported a saltatory view of evolution, probably a remnant of his originally discontinuous view of type organization. After his famous Oxford clash with Bishop Wilberforce on human descent in 1860, Huxley became popularly known as a champion of Darwinism, but his scientific production does not show any deployment of evolutionary reasoning until 1868. Even his 1863 'Man's Place in Nature', though showing anthropomorphic apes morphologically closer to humans than to lower primates, and thus providing evidence for Darwin's views, made little reference to evolution, being mainly an attack on Owen's systematics. Only after reading Haeckel's *Generelle Morphologie* (1866), which presented a reinterpretation of embryological typology in evolutionary terms and emphasized descent rather than natural selection, did Huxley introduce evolutionism into his science (see Haeckel, E.H.). Thus Huxley related evolution to the scientific tradition to which he belonged.

In ethnology, Huxley denied a role for philology in human taxonomy, and fought the polygenic view of human origins. Huxley popularized science and devoted much time to teaching. He followed with interest philosophical debates, and wrote on Descartes with appreciation for his view of knowledge and of animals as automata. He respected Mill, but loathed the positivist movement, both French and British. Possibly it was through his acquaintance with Darwin's views that he approached Hume's philosophy, on which he wrote (1878). He applauded Hume's view of the connection between common sense and scientific discourse, and of the untenability of the evidence for miracles (see Hume, D. §2), a subject on which Huxley argued with Gladstone. The term 'agnostic' was coined by Huxley to denote the scientific refusal to make statements in fields which, like theology, were outside the scope of his research. His latter publications concerned the debate on evolution and ethics, in

which he opposed Spencer's naturalistic ethics (see Spencer, H.), arguing for a complete separation of knowledge from values, opposing*laissez-faire* policy in favour of state intervention, but critical at the same time of socialism. He was concerned at the extreme individualism of his time and believed that only through science was the progress of humankind possible; that human suffering could not be explained or justified scientifically. Huxley thought that in the course of evolution, higher degrees of consciousness had arisen with humankind, and with them, a higher awareness of suffering.

See also: Darwin, C.R.; Evolution and ethics; Evolution, theory of; Philosophy of science in the 19th century

MARIO A. DI GREGORIO

List of works

Huxley, T.H. (1893-4) *Collected Essays*, London: Macmillan, 9 vols.(Huxley edited these volumes that contain most of his works.)

Huxley, T.H. (1898-1903) *The Scientific Memoirs of Thomas Henry Huxley*, eds G. Foster and E.R. Lankester, London, 4 vols and suppplement.(Contains original writings; hereafter referred to as *Scientific Memoirs*.)

Huxley, T.H. (1849) 'On the Anatomy and Affinities of the Medusae', *Philosophical Transactions of the Royal Society of London* 2: 413; repr. in *Scientific Memoirs*, 1: 9-32.(Huxley determined the homology of jellyfish and established the ground for the germ-layer theory later developed by Alexander Kowalewsky, Ernst Haeckel and E.R. Lankester).

Huxley, T.H. (1853) 'On the Morphology of the Cephalous Mollusca, as illustrated by the Anatomy of Certain Heteropoda and Pteropoda collected during the Voyage of HMS *Rattlesnake* in 1846-50', *Philosophical Transactions of the Royal Society of London* 1: 43, Part 1, 29-66; *Scientific Memoirs* 1: 152-93.(Huxley endeavoured to interpret the type, then the basic concept in morphology, in an empirical manner, as denoting a form which summarises and encapsulates the characteristics of a zoological group, thus distancing himself from Platonic metaphysics.)

Huxley, T.H. (1857-9) 'On the Theory of the Vertebrate Skull', *Proceedings of the Royal Society of London* 9: 381-57; repr. in *Scientific Memoirs*, 1: 538-606.(Huxley attacked Owen's view of the vertebral theory of the skull. With the support of comparative embryology he endeavoured to show the untenability of Owen's anatomical interpretation of the skull as formed by expanded vertebrae.)

Huxley, T.H. (1859) 'The Darwinian Hypothesis', *The Times* 26 December; repr. in *Collected Essays* 2: 1-21.(A review of Darwin's *Origin*.)

Huxley, T.H. (1860) 'The Origin of Species', *Westminster Review* 17: 541-70; repr. *Collected Essays* 2: 22-79. (Another review of Darwin's *Origin*.)

Huxley, T.H. (1863) *Evidence as to Man's Place in Nature*, London: Williams & Norgate; repr. in *Collected Essays* 7: 1-208.(This book, in fact a collection of papers, sprang from Huxley's controversy with Owen on the presence of the hippocampus minor in the cerebellum of both man and ape, which Owen denied and Huxley maintained. Owen intended to separate man and apes into different subclasses, while Huxley defended Linnaeus' traditional classification, which was in no way evolutionary.)

Huxley, T.H. (1863) 'On Our Knowledge of the Causes of the Phenomena of Organic Nature'; repr. in *Collected Essays* 2: 303-475.(A series of popular lectures on various aspects of natural science including Darwin's theory).

Huxley, T.H. (1868) 'On the Animals which are Most Nearly Intermediate between Birds and Reptiles', *Geological Magazine* 5: 357-65; repr. in *Scientific Memoirs* 3: 303-13.(Dinosaurs as link between birds and reptiles).

Huxley, T.H. (1869) 'The Scientific Aspects of Positivism', *Fortnightly Review* 5: 653-70.(Huxley called Comte's philosophy 'Catholicism minus Christianity'.)

Huxley, T.H. (1874) 'On the Hypothesis that Animals are Automata, and its History', *Nature* 10: 362-6; repr in *Collected Essays* 1: 199-250.(Huxley discusses Descartes' view that animals are automata.)

Huxley, T.H. (1874) 'Universities, Actual and Ideal', *Contemporary Review* 23: 657-79; repr. in *Collected Essays*, 3: 189-234.(On teaching and research in the reorganization of British universities.)

Huxley, T.H. (1878) *Hume*, London: English Men of Letters Series; repr. in *Collected Essays* 6: 3-240.(A thorough analysis of Hume's philosophy.)

Huxley, T.H. (1889) 'Agnosticism', *Nineteenth Century* 25: 169-94; repr. in *Collected Essays* 5: 209-62. (Huxley's refusal of both Atheism and traditional Christian theology.)

Huxley, T.H. (1893) 'Evolution and Ethics' *Collected Essays* 9: 46-116.(Huxley's criticism of naturalistic ethics.)

Huxley, T.H. (1894) ‘Prolegomena to "Evolution and Ethics"’, *Collected Essays* 9: 1-45.(Huxley elaborates on the themes discussed in *Evolution and Ethics*.)

References and further reading

Bartholomew, M. (1975) ‘Huxley’s Defence of Darwin’, *Annals of Science* 32: 525-35.(Huxley did not apply Darwin’s theory to his scientific works.)
Chambers, R. (1844) *Vestiges of the Natural History of Creation*, London: Churchill.(Published anonymously, a popular book which propounded a progressive view of evolution.)
Collie, M. (1991) *Huxley at Work*, London: Macmillan.(Huxley’s methods and results.)
Darwin, C. (1859) *On the Origin of Species*, London: Murray.(Darwin’s celebrated book on the origin of species by natural selection.)
Desmond, A. (1994, 1996) *Huxley, The Devil’s Disciple*, London: Michael Joseph, 2 vols.(A biography of Huxley in his social and political context.)
Di Gregorio, M.A. (1984) *T.H. Huxley’s Place in Natural Science*, New Haven, CT, and London: Yale University Press.(The contents and significance of Huxley’s science.)
Haeckel, E. (1866) *Generelle Morphologie der Organismen (The General Morphology of Organisms)*, Berlin: Reimer, 2 vols.(On evolutionary morphology.)
Huxley, L. (1900) *The Life and Letters of Thomas Henry Huxley*, London and New York: Macmillan, 2 vols. (Huxley’s biography based on his correspondence arranged by his son.)
Jensen, J.V. (1991) *Thomas Henry Huxley: Communications for Science*, Newark, NJ: University of Delaware Press.(The cultural aspects of Huxley’s work.)
Paradis, J.G. (1978) *T.H. Huxley: ‘Man’s Place in Nature’*, Lincoln, NE, and London: University of Nebraska Press.(Huxley in his Victorian milieu).

Hypatia (*c.* AD 370-415)

The Greek philosopher Hypatia was a Neoplatonist. She was famous for her public talks on philosophy and astronomy, and her forthright attitude to sex. Alhough concerned with higher knowledge, she was also a political animal and had a keen sense of practical virtue. She was killed by a Christian mob, and has remained since a martyr to the cause of philosophy.

Hypatia was born in Alexandria around AD 370. However, according to one source she was 'aged' (*palaia*) in AD 415, which has led some to propose she was born around AD 355 (Dzielska 1995: 68). She was educated by her father Theon, who wrote important treatises on mathematical astronomy. But in what field did she flourish? This presents the first of several fascinating problems of interpretation. The works ascribed to her (none survives) are exclusively mathematical commentaries: *On Diophantus* (a third-century Alexandrian mathematician and the first to use algebra), *On the Conic Sections of Apollonius* and *On the Astronomical Canon* (of Ptolemy). Furthermore, when Damascius (a sixth-century Alexandrian Neoplatonist, who became head of the Athenian school) compared her to his teacher, he deemed her to be 'much different from him, not only as woman is to man, but also as a geometer is to a real philosopher' (Damascius, *Life of Isidore*, fr. 164.1-2). On the other hand, 'she was more courageous than her father and was not content with her mathematical education but boldly reached up to the other philosophy' (*Suda* U166.13-5). (For this passage from astronomy to knowledge of the Forms, see Plato, *Republic* 529a9.) 'She outdid the philosophers of her time…. Those interested in philosophy run to her from everywhere' (Socrates, *Ecclesiastical History* 7.15), and for Hypatia's eminent student, Synesius (first pagan, then Christian bishop of Cyrene and statesman), she surpassed the Athenian 'sophists' (Synesius, *Letters* 136.17-21). So, her intellectual home may have been the mathematical quadrivium - traditionally part of Platonism - but she dared to do exceptionally well in metaphysics.

What was her philosophy? According to one statement, 'she succeeded to the Platonic school that descended from Plotinus' (*Socrates' Ecclesiastical History*, 7.15.5). According to another, 'she explained Aristotle or Plato or some other philosopher' (*Suda* U166.17-9). Furthermore, 'the woman wore a rough mantle (*tribōn*) and made public appearances at the centre of the city' (U166.16). However, the last two tell nothing special. The rough mantle, although originally of the Cynics, was often the habit of the plain-living philosopher (for example, the Stoics). Interpretation of Aristotle, Plato or others was carried out by most philosophers, including the Neoplatonists and Ammonius Saccas, the Alexandrian teacher of Plotinus. So, Hypatia must have followed Neoplatonism, probably as a result of Porphyry's dissemination (see Neoplatonism). Her students included both pagans and Christians.

What was her stand on gender issues? Men extolled her 'exceptional' beauty, her flair for public speaking and her capacity to do things unexpected of a woman in that society. Male followers were enamoured of her intellectually and emotionally. How did she cope? On one celebrated occasion, to temper a love-sick student she removed her bloody sanitary towel, threw it at him and told him 'this is what you really want, young man, and it is no good' (Damascius, *Life of Isidore* 102). This agrees with other descriptions of her as 'solemn' and 'virgin/unmarried to the end'. But, these were the very attitudes cultivated equally by select male philosophers (see, for example, Proclus §1). Hypatia upheld the Platonic, or generally philosophical, freedom from bodily concerns, to rise above sex altogether.

These elements, and others, are brought into sharp focus when the possible reasons for her gruesome murder are considered. Some jealousy was to blame, fuelled by her political prominence when secular civic authority was being challenged by the ambitious bishop Cyril. Was the acrimony purely local? Was it intellectual or religious (shared belief in astrology and paganism)? Was it personal (the result of Hypatia's friendship with the city's Governor)? Whatever the complex truth, Hypatia soon came to be seen as martyr in the hands of the Christian mob. In modern times, criticism of the Victorian novel *Hypatia* by C. Kingsley (London, 1853) has brought to the fore the gender aspects, and, more recently, 'Hypatia' has become the name of two philosophy journals on feminism.

LUCAS SIORVANES

References and further reading

Damascius (*c.* AD 462-540) *Life of Isidorus*, ed. C. Zintzen, *Damascii vitae Isidori reliquiae,* Hildesheim: Olms, 1976.(Greek text; a source for many late Neoplatonists.)
Dzielska, M. (1995) *Hypatia the Alexandrian*, Cambridge, MA: Harvard University Press. (Extensive account of Hypatia's life, historical and ideological context.)
Kingsley, C. (1853) *Hypatia: or: New Foes with an Old Face*, London: Parker.(A morality novel by the chaplain to Queen Victoria.)
Socrates (5th century AD) *Socrates' Ecclesiastical History*, ed. W. Bright, Oxford: Clarendon Press, 2nd edn, 1893.(The church historian Socrates is the main source for Hypatia's life.)
Suda (10th century AD) *Suidae Lexicon*, ed. E. Adler, Leipzig: Teubner, 1931; repr. Stuttgart, 1967.(Greek text of the dictionary; often the only place where ancient fragments survive; for Hypatia.)
Synesius (4th century AD) *Letters of Synesius of Cyrene*, trans. A. Fitzgerald, Oxford: Clarendon Press, 1926.(For Hypatia, see letter 54.)
Wider, K. (1986) 'Women Philosophers in the Ancient Greek World: Donning the Mantle', *Hypatia* 1: 21-62. (Women philosophers and gender bias in the reports about them.)

Iamblichus (*c.* AD 242-327)

The late ancient philosopher Iamblichus was, alongside Plotinus and Porphyry, a founder of Neoplatonism. He established a new curriculum for the teaching of philosophy and formulated many distinctions that pervaded later Neoplatonic metaphysics. He began to mathematize all fields of philosophical concern. Most of all, he asserted that acts of transcendence, not contemplation, secure union with the divine, because it can only be reached by an equally divine faculty, which is present in every individual. Matter, soul and mind contain images of the divine and so are genuine participants in salvation.

Details of Iamblichus' life are scant. He was born at Chalcis, a descendant from an ancient line of priest-kings of Syria. In the AD 270s he studied at Caesarea (where many Christian theologians flourished). Perhaps he was still there when he heard of Porphyry, and may have visited him in Italy. They split in the AD 290s over theurgy, and Iamblichus' chief surviving work, *On the Mysteries*, is a record of their disagreement. He went to teach philosophy at Apamea (Syria), where the Neo-Pythagorean Numenius, and more recently Plotinus' chief follower Amelius, had flourished. The school he set up became very popular. From Pergamum his envoys converted the emperor Julian from Christianity, influencing the Athenian school of philosophy, and through it the Alexandrian school.

Iamblichus took further the programme of 'harmonization' of Greek philosophies that had been started by Porphyry: Aristotle agreed with Plato even on the Theory of Forms and on learning as recollection. He designed a curriculum for the teaching of Plato which had Aristotle as a prerequisite (see Neoplatonism §2). He also wrote grand commentaries on the two philosophers. However, he considered Pythagoras the supreme authority. Iamblichus saw number as the essence of Forms and things, and promised to make philosophy and theurgy scientific. His many treatises on Pythagoreanism mark a revival of the movement. Eight or so titles of his large number of works survive, along with reports in Proclus, Damascius and Simplicius. Later Neoplatonists were wary of Iamblichus' claims about Aristotle and Plato, but maintained his curriculum, his metaphysical distinctions and his emphasis on mathematics.

Iamblichus introduced radical revisions to Plotinus' philosophy, multiplying the ranks of being (see Plotinus §§3-4). Plotinus saw the metaphysical levels as areas of the soul's activity, with flexible and sometimes conflicting properties. Iamblichus separated psychology from ontology, and looked for distinct properties and substances in the anatomy of things. He also distinguished the 'imparticipable' from the 'participated'. Every class of things has a first principle, which retains its pure definition by being exempt from its members.

The Neoplatonic One was inferred by stripping away all attributions, but was also the first cause of being. How could the same thing transcend its products, and also be in a causal relation to them? Iamblichus promptly introduced two Ones (later Neoplatonists reverted to a single One). At the level of intellect (see Nous), he distinguished the acts of thought - the 'intellective' - from their objects - the 'intelligible'. At the level of soul (*psychē*), he separated the individual soul from the higher: in contrast to Plotinus, for Iamblichus the individual soul descends entire into the body (see Psychē). Similarly Iamblichus distinguished not only between time and timelessness, but also between a pure, eternal time and ordinary time, which is flowing.

To describe the essential nature of things, and the 'flow' or 'procession' from unity to diversity, Iamblichus turned to Neo-Pythagorean metaphysics of number. He believed that things are organized by number and relate to each other in mathematical proportion. The 'divine numbers' distribute the unity of the One to things and render them coherent. The 'intelligible numbers' stand as paradigms of the many species of reality. Intellect is organized by 'intellective numbers'. Soul exemplifies the 'mathematical numbers'. Last are the 'physical numbers', the forms immanent in matter.

Iamblichus' conclusion that the soul descends entire into matter had profound consequences. Individual souls are dominated by physical necessities (which fits with Aristotle's definition of soul as the form of biological body). But soul is still essentially divine and rational. So the embodied soul presents an uncomfortable mystery. It contains an immortal, intelligent, divine nature, but is genuinely part of a mortal, concrete, imperfect domain. The personal soul has lost touch with its deeper nature and has become self-alienated.

Iamblichus' analysis was that the transcendent cannot be grasped with mental contemplation, because the

transcendent is supra-rational. Theurgy, literally 'divine-working', is a series of rituals and operations aimed at recovering the transcendent essence by retracing the divine 'signatures' through the layers of being. Education is important for comprehending the scheme of things as presented by Aristotle, Plato and Pythagoras but also by the Chaldaean Oracles (see Chaldaean Oracles). The theurgist works 'like with like': at the material level, with physical symbols and magic; at the higher level, with mental and purely spiritual practices. Starting with correspondences of the divine in matter, the theurgist eventually reaches the level where the soul's inner divinity unites with God.

See also: Neo-Pythagoreanism

LUCAS SIORVANES

List of works

Iamblichus (late 3rd century AD) *On the Mysteries*, ed. E. des Places, *Jamblique. Les mystères d' Égypte*, Paris: Les Belles Lettres, 1989; trans. T. Taylor, *On the Mysteries*, London, 1821; repr. Hastings: Chthonios Books, 1989.(The former is parallel Greek text and French translation; the latter is English translation of old edition.)

Iamblichus (late 3rd century AD) *On the Pythagorean Life*, trans. E.G. Clark, Liverpool: Liverpool University Press, 1989.(English translation and annotation.)

Iamblichus (early 4th century AD) *Commentaries on Plato*, ed. J.M. Dillon, *Iamblichi Chalcidensis in Platonis dialogos commentariorum fragmenta*, Leiden: Brill, 1973.(Greek text with English translation of the fragments and extensive notes.)

Iamblichus (early 4th century AD) *Exhortation: Protrepticus*, ed. L. Pistelli, Leipzig: Teubner, 1888; repr. Stuttgart, 1975; trans. and ed. E. des Places, *Jamblique: Protreptique*, Paris: Les Belles Lettres, 1989; trans. T. Johnson, 1907; repr. in S. Neuville (ed.), *Iamblichus. Exhortation to Philosophy*, Grand Rapids, MI: Phanes Press, 1988.(The former is parallel Greek text and translation; Neuville is an English translation of Johnson.)

Iamblichus (early 4th century AD) *Common Mathematical Theory*, ed. N. Festa, *Iamblichi de communi mathematica scientia*, Leipzig: Teubner, 1891; repr. Stuttgart, 1975.(Greek text.)

References and further reading

Blumenthal, H.J. and Clark, E.G. (eds) (1993) *The Divine Iamblichus*, London: Bristol Classical Press. (Collection of articles, including some on soul, political theory and theurgy.)

Finamore, J.F. (1985) *Iamblichus and the Theory of the Vehicle of the Soul*, Chico, CA: Scholars Press.(Covers aspects of Iamblichus' psychology, metaphysics and mysticism.)

Gersh, S. (1978) *From Iamblichus to Eriugena*, Leiden: Brill.(Detailed account of later Neoplatonic metaphysics and the transition to medieval theology, including pseudo-Dionysius; major themes are illustrated with diagrams.)

Larsen, B.D. (1972) *Jamblique de Chalcis*, Aarchus: Aarchus University.(Comprehensive study of Iamblichus' thought, and treatment of Plato and Aristotle, including the scale of virtues.)

O'Meara, D.J. (1982) *Pythagoras Revived: Mathematics and Philosophy in Late Antiquity*, Oxford: Clarendon Press.(Important study on the programme of mathematization and relation to metaphysics and epistemology.)

Plass, P.C. (1977) 'Timeless Time in Neoplatonism', *Modern Schoolman* 55: 1-19.(On Iamblichus' conception of time.)

Saffrey, H.D. (1990) *Recherches sur le néoplatonisme après Plotin (Studies of Neoplatonism After Plotinus)*, Paris: Vrin.(Collected articles about Porphyry, Iamblichus and Proclus on theological themes, by a leading scholar in the field.)

Shaw, G. (1995) *Theurgy and the Soul*, University Park, PA: Pennsylvania State University Press.(A new account of the religious value of theurgy, and the relation of soul to matter.)

Steel, C. (1978) *The Changing Self: A Study on the Soul in Later Neoplatonism: Iamblichus, Damascius, Simplicius*. Brussels: Royal Academy.(Seminal study of Iamblichean psychology.)

Ibn ‘Adi, Yahya (893-974)

Following in the footsteps of the Greek philosophers, Ibn ‘Adi concerned himself with the ultimate human end, happiness, which he found in knowledge. However, he was primarily occupied with defending the compatibility between the concept of God’s unity and that of the trinity. He reasoned that a thing can be one in one respect and many in another. Therefore, there is no inconsistency in holding that God is both one and three. Ibn ‘Adi can best be described as the Christian philosopher of unity, as he devoted most of his career and used all his logical skills to defend the concept of God’s unity and its consistency with the concept of trinity.

Ash-Shaykh Abu Zakariyya’ Yahya ibn ‘Adi was a Jacobite Christian who lived in Iraq. Born in Takrit, he moved as a youth to Baghdad, one of the most important centres of learning in the tenth century. Of Syriac origin, he was Arabized like many other Syriacs at that time. He learned logic and philosophy with the well-known logicians, Abu Bishr Matta ibn Yunis and al-Farabi, and after their deaths he became the leading logician of his time. He translated Greek philosophical works from Syriac into Arabic, wrote a number of logical, philosophical and theological treatises - the most important of which are *Tahdhib al-akhlaq (Refinement of Character)* and *Maqala fi at-tawhid (Essay on Unity)* - and established the Aristotelian school at Baghdad. His students, a mixture of Muslims and Christians, included Ibn Miskawayh, Ibn al-Khammar, who wrote a treatise on the harmony between philosophy and dogma (which may have influenced Ibn Rushd in his treatment of the same subject), and Ibn Zur‘a.

In *Tahdhib al-akhlaq*, Ibn ‘Adi sets out his ethical philosophy along Greek lines (see Ethics in Islamic philosophy). It is based on his view of the human soul, which is divided into three parts or souls: the appetitive, the courageous and the rational. The first is the lowest and is shared by humans with other animals. The last is the noblest and the distinguishing mark of being human. To follow the first is to fall into ignorance and evil; to follow the last is to adhere to goodness and happiness. While all human beings have the natural capacity for reasoning, some have the skill to reason and some do not. Those who do not may acquire it by learning the rational sciences.

Ibn ‘Adi’s concern with God’s unity, however, was his main preoccupation. He identifies the meaning of the word ‘one’ as it applies to God, and investigates whether God is one in all respects or one and many. Asserting that the one is that in which there is no otherness inasmuch as it is one, he classifies six things as one: the genus, the species, the relation, the continuous, the indivisible and the definition. He then offers arguments to show that God cannot be one except in the sense of definition (definition being a descriptive statement of the essence of a thing, in that it gives the essence as it is). Since every definition mirrors an essence, God must also be one in essence.

Most Greek and medieval philosophers considered God indefinable. A definition, they say, includes a genus and a difference, such as ‘animality’ and ‘rationality’ for ‘humanity’. God, being simple, cannot be included in any genus because there is nothing above him. But if he has no genus, he cannot have any difference, for the difference is what differentiates a genus. Without a genus and a difference, God cannot be defined. While agreeing that a definition requires a genus and at least one difference, Ibn ‘Adi believes that God is definable because he falls under the genus ‘substance’, for like every substance he does not reside in anything. This is interesting because it places something, the genus ‘substance’, above God and makes him the cause of the substance out of which he is made. This follows from Ibn ‘Adi’s view that the genus is caused by its members and that God is the cause of everything. The difficulty then arises that there was a time when God had no substance.

Ibn ‘Adi asserts that God is one in one respect and multiple in another. As stated earlier, God is one in definition and in essence. However, a definition is a statement, and every statement has its parts. Such parts signify separate meanings. As such, these parts are separate definitions. Thus there is no contradiction between saying that God is one in definition and multiple in definition. The multiple definitions of God signify the three attributes, goodness, power and wisdom, referred to in Ibn ‘Adi’s later works as intellect, act of intellection and object of intellection, respectively. Goodness signifies the Father; power signifies the Holy Spirit; and wisdom signifies Christ. The same thing may be reflected in two mirrors differently, in one as a cause and in the other as an effect. Similarly, God and the Trinity are one in essence but different as individuals. Observation of God’s creatures demonstrates these attributes.

The third century AH (ninth century AD) Muslim philosopher al-Kindi wrote a treatise attempting to refute the

Christian concept of the Trinity purely on the basis of logical reasoning. This treatise is, as far as we know, preserved only in the writings of Ibn ‘Adi. Al-Kindi asserts that, according to Christianity, the three figures of the Trinity share substance or essence, but each is an individual by virtue of a specific property. The substance plus this property makes every one of them a composite. However, every composite is caused, and whatever is caused cannot be eternal. Ibn ‘Adi responds that it is true that the three figures are composite individuals by virtue of their substance and properties, but their being composite does not make them caused, because it is possible for a thing to be composite eternally if the parts were not separate before the composition.

See also: Ethics in Islamic philosophy; God, concepts of; Logic in Islamic philosophy; Trinity

SHAMS C. INATI

List of works

Ibn ‘Adi (893-974) *Tahdhib al-akhlaq (Refinement of Character)*, ed. N. at-Takriti, Beirut: ‘Uweidat, 1978.(An Arabic edition of *Tahdhib al-akhlaq* and a critical study of this treatise, showing the significant impact it had on Christian and Islamic circles. Unfortunately, the English sections suffer from a number of typographical errors and inaccuracies, and must therefore be read with caution.)

Ibn ‘Adi (893-974) *Maqala fi at-tawhid (Essay on Unity)*, ed. S. Khalil, Juni: Al-Maktaba al-Bulisiyya; Rome: al-Ma‘had al-Babawi ash-Sharqi, 1980.(The best and most comprehensive treatise of Ibn ‘Adi on the concept of God’s unity.)

Ibn ‘Adi (893-974) *Maqalat Yahya Ibn ‘Adi al-falsafiyya (Yahya ibn ‘Adi’s Philosophical Essays)*, ed. S. Khulayfat, ‘Amman: Manshurat al-Jami‘a al-Urduniyya, 1988. (Includes twenty-four of Ibn ‘Adi’s essays carefully edited and annotated. The introduction examines Ibn ‘Adi’s general thought and its influence on various students and later philosophers. This work is a fairly reliable source for Ibn ‘Adi’s logical, philosophical and theological thought.)

References and further reading

Endress, G. (1977) *The Works of Yahya Ibn ‘Adi: An Analytical Inventory*, Wiesbaden: Reichert.(The best available classification of Ibn ‘Adi’s works.)

Platti, E. (1983) *Yahya Ibn ‘Adi, théologien chrétien et philosophe arabe: sa théolgie de l’Incarnation (Yahya Ibn ‘Adi, Christian Theologian and Arab Philosopher: His Theology of the Incarnation)*, Orientalia Lovanensia analecta vol. 14, Leuven: Departement Orientalistiek.(Deals with a major theme in Ibn ‘Adi’s thought.)

Ibn al-‘Arabi, Muhyi al-Din (1164-1240)

Ibn al-‘Arabi was a mystic who drew on the writings of Sufis, Islamic theologians and philosophers in order to elaborate a complex theosophical system akin to that of Plotinus. He was born in Murcia (in southeast Spain) in AH 560/AD 1164, and died in Damascus in AH 638/AD 1240. Of several hundred works attributed to him the most famous are al-Futuhat al-makkiyya (The Meccan Illuminations) and Fusus al-hikam (The Bezels of Wisdom). The Futuhat is an encyclopedic discussion of Islamic lore viewed from the perspective of the stages of the mystic path. It exists in two editions, both completed in Damascus - one in AH 629/AD 1231 and the other in AH 636/AD 1238 - but the work was conceived in Mecca many years earlier, in the course of a vision which Ibn al-‘Arabi experienced near the Kaaba, the cube-shaped House of God which Muslims visit on pilgrimage. Because of its length, this work has been relatively neglected. The Fusus, which is much shorter, comprises twenty-seven chapters named after prophets who epitomize different spiritual types. Ibn al-‘Arabi claimed that he received it directly from Muhammad, who appeared to him in Damascus in AH 627/AD 1229. It has been the subject of over forty commentaries.

*Although Ibn al-‘Arabi was primarily a mystic who believed that he possessed superior divinely-bestowed knowledge, his work is of interest to the philosopher because of the way in which he used philosophical terminology in an attempt to explain his inner experience. He held that whereas the divine Essence is absolutely unknowable, the cosmos as a whole is the locus of manifestation of all God’s attributes. Moreover, since these attributes require the creation for their expression, the One is continually driven to transform itself into Many. The goal of spiritual realization is therefore to penetrate beyond the exterior multiplicity of phenomena to a consciousness of what subsequent writers have termed the ‘unity of existence’. This entails the abolition of the ego or ‘passing away from self’ (*fana’*) in which one becomes aware of absolute unity, followed by ‘perpetuation’ (*baqa’*) in which one sees the world as at once One and Many, and one is able to see God in the creature and the creature in God.*

1 Epistemology

While still an unbearded youth, Ibn al-‘Arabi was introduced by his father to the celebrated philosopher Ibn Rushd, who eagerly questioned him about his spiritual experiences. Ibn al-‘Arabi describes the interview as follows:

> He said, ‘How did you find the situation in unveiling and divine effusion? Is it what rational consideration gives us?’ I replied, ‘Yes and no. Between the yes and the no spirits fly from their matter and heads from their bodies.’
>
> (*al-Futuhat al-makkiya* I.154, in Chittick 1989: xiii)

This cryptic answer, which reputedly made the philosopher turn pale and tremble, implies the existence of divinely-bestowed knowledge which is superior to knowledge gained by ‘rational consideration’ (*nazar*). But what precisely is the relationship between them? Elsewhere, Ibn al-‘Arabi speaks not of two levels of knowledge, but of three (for example, in *al-Futuhat al-makkiya* I.31). First, there is ‘knowledge based on reason’ (*‘ilm al-‘aql*), that is, knowledge which can be acquired by rational consideration. Here he probably has in mind the principal tenets of Muslim theology rather than the a priori self-evident propositions of logic and mathematics. Second, there is knowledge based on states (*‘ilm al-ahwal*), which is what we would call empirical knowledge. He gives as examples the sweetness of honey, the bitterness of aloes and the pleasure of sexual intercourse, none of which can be known without ‘tasting’ them. Third, there is ‘knowledge of mysteries’ (*‘ulum al-asrar*) - sometimes called ‘gnosis’ (*ma‘rifa*) - which is specific to prophets and saints and is akin to Spinoza’s *scientia intuitiva* (see Spinoza, B. de) It is futile to strive for this third type of knowledge, for it lies concealed in every man but is only unveiled when the divine light is effused into the hearts of those who are predisposed to receive it. ‘Knowledge of mysteries’ includes knowledge of the first type, except that it is acquired without reflection, and knowledge of the second type but pertaining to higher states not experienced by lesser mortals; potentially it embraces everything except the unknowable Essence. It is nothing short of divine knowledge for, in the words of the celebrated *hadith qudsi* (extra-Qur’anic revelation), when God loves his servant he becomes ‘the hearing with which he hears, the sight with which he sees’.

The distinctly subordinate role given to reason in Ibn al-‘Arabi’s epistemology appears at first to be out of step with the Qur’an, which repeatedly urges man to engage in ‘rational consideration’ and ‘reflection’ (see Epistemology in Islamic philosophy). In his view however, there is no real tension because the main purpose of considering and reflecting is to lead man to the realization that he cannot reach knowledge of God through his unaided reason. This is illustrated in Chapter 167 of *al-Futuhat al-makkiyya (The Meccan Illuminations)*, in which the progressive journey of the gnostic and the philosopher towards the truth is depicted in terms of a heavenly ascent akin to that experienced by Muhammad (the *mi‘raj*). As they pass through each of the celestial spheres the gnostic is addressed by their spirits - the prophets who inhabit each sphere - and perceives their inner reality. The philosopher, on the other hand, learns only the phenomenal or apparent and becomes increasingly perplexed and sceptical until he finally becomes a Muslim and follows the path of the gnostic (see Gnosticism).

2 Theology

Although Ibn al-‘Arabi has often been labelled a pantheist, he was far too subtle a thinker to have subscribed to the doctrine that God is everything and everything is God (see Pantheism). He believed that God *per se* - whom he called ‘the Real’ (*al-Haqq*) or ‘the Essence’ (*al-dhat*) - is absolutely unknowable because he transcends all humanly conceivable qualifications. God’s ‘names’ (*asma’*) or ‘attributes’ (*sifat*), on the other hand, are the relationships which can be discerned between the Essence and the cosmos. They are known to God because he knows every object of knowledge, but they are not existent entities or ontological qualities, for this would imply plurality in the godhead. It may help to understand the status of the ‘names’ (which of course must not be confused with ‘the names of the names’ known to us from the Qur’an and Islamic tradition) if we draw an analogy with the complex web of interpersonal human relationships. One and the same individual may be known to others variously as teacher, pupil, friend, enemy, father, son, brother, husband, lover and so forth. A man who knows another as his friend genuinely knows him, but does not necessarily know him as his teacher or his father and cannot know him as he is in himself without regard to others. Similarly, to know God as the All-Merciful, for instance, does not entail knowing him as the Vengeful or the Abaser, nor does it mean knowing his Essence even though each of the names denotes the Essence. The analogy must not be pressed, however, because unlike human beings, whose relationships are temporal, God has possessed the divine names from all eternity.

Thus far Ibn al-‘Arabi’s theology remains within the confines of Islamic *kalam*, although he seems in some ways to have more affinity with the Mu‘tazila than with the Ash‘ariyya (see Ash‘ariyya and Mu‘tazila; Islamic theology). However, he differs markedly from both in holding that, not withstanding the fact that in his Essence, God is independent of the cosmos, his names none the less seek the creation, for without it they would remain virtualities. The cosmos as a whole is the locus of their manifestation and it is only through it that their properties can be seen and understood. Moreover, since all the creatures in the cosmos and everything which they make or do are God’s ‘acts’ (*af‘al*), God is present everywhere and the all-inclusive name Allah may be used to denote the sum total of all things: divine Essence, divine names and divine acts. Nevertheless, since creatures have only relative existence (see §3), it also has to be said that he is nowhere to be found (see God, concepts of).

3 Ontology

In agreement with the Islamic philosophers, Ibn al-‘Arabi distinguishes between the essence or ‘quiddity’ (*mahiyya*, literally the ‘what is it?’) of a thing and its ‘existence’ (*wujud*, literally its ‘being found’), holding that the former is mentally separable from the latter. That is, one can define the nature of things (unicorns, pink elephants and so forth) regardless of whether or not they are actually found as phenomena. He also accepts the distinction between necessary being (*wajib al-wujud*), impossible things and possible things. The necessary being is God, the one reality who cannot not exist because his quiddity is being (see Being). Impossible things are things which cannot exist as phenomena although they may subsist in the imagination (see §5). Possible things are things which become ‘existent entities’ (*a‘yan mawjuda*) when God chooses to give them existence; their existence or non-existence at any given time depends on his will. They have, however, been known to him eternally as ‘immutable entities’ (*a‘yan thabita*). This latter term is rendered by Affifi (1938: 47) as ‘fixed prototypes’ and Izutsu (1983: 159) as ‘permanent archetypes’, expressions which suggest that like the Platonic Ideas they are the original models of which objects in the phenomenal world are multiple copies (see Universals). Chittick (1989: 84) objects to this on the grounds that although the ‘immutable entities’ actually become the ‘existent entities’, they are the things themselves prior to their being given existence in the world. Although I have adopted his translation

as closer to the meaning of the Arabic, I have reservations about his interpretation, for there are passages in Ibn al-‘Arabi’s writings which seem to imply that some of the immutable entities are universals. Nevertheless it may well be the case that Ibn al-‘Arabi’s views on this subject were not entirely consistent.

Another problem area is Ibn al-‘Arabi’s lack of precision in using the word *wujud* to mean both ‘existence’ and ‘Being’. It was stated above that possible things become existent entities when God gives them existence, but it must be stressed that the existence which he gives them is relative existence. They become the ‘loci of manifestation’ (*mazahir*) of the names of God, who, as Essence, is alone Being in the strict sense. Another way of putting it is to say that each entity becomes a receptacle for Being, but that since entities differ from one another, they differ also in their capacity to function as vehicles of his self-manifestation.

4 The ‘perfect man’ and the Muhammadan reality

The first chapter of the *Fusus al-hikam (The Bezels of Wisdom)* is entitled ‘The Wisdom of Divinity in the Word of Adam’. It begins with the assertion that the Real created the cosmos as an all-inclusive object in which he could contemplate the entities of his names, but that until he created Adam and breathed his spirit into him, the cosmos remained like an unpolished mirror. Here Ibn al-‘Arabi’s idea seems to be that the cosmos as a whole - the totality of existent entities - manifests all the divine names but does so in a diffuse way, whereas man, as a microcosm endowed with consciousness, brings them into sharp focus as a unity. Potentially every man is a microcosm, but in practice men differ in their polishing of the cosmic mirror, with only a select few realizing their primordial nature. These are the prophets and saints, all of whom belong to the category of ‘the perfect man’ (*al-insan al-kamil*). They alone assume the character traits of God, which are latent in all human beings, and manifest them in perfect equilibrium.

Muhammad is the ‘perfect man’ *par excellence*. Basing his argument on the *hadith* (sayings of the Prophet), ‘I was a prophet when Adam was between water and clay’, Ibn al-‘Arabi propounds the view that as ‘the Muhammadan reality’ (*al-haqiqa al-Muhammadiyya*), Muhammad is identical with ‘the first intellect’ (*al-‘aql al-awwal*), the eternal principle unifying the immutable entities. All the other prophets, beginning with Adam, only became prophets during their historical mission; each was the bearer of a fragment of this Muhammadan reality in a particular place and time, a bezel in which a jewel of the divine wisdom was displayed. None the less, after their mission the prophets continued to exert an influence through the saints who were their spiritual heirs.

5 Imagination and mysticism

The *al-Futuhat* contains a good deal of autobiographical material in anecdotal form, some of which strains credulity. For instance Ibn al-‘Arabi, who was in no doubt that he himself was one of the most important saints in the history of Islam, tells us that he met and conversed with the prophets of old including Moses, Jesus and Muhammad. On one moonlit night, when on board a ship in the port of Tunis, he allegedly encountered Moses’ spiritual guide al-Khidr, who came to him walking on the water without getting his feet wet, before going off to a lighthouse over two miles away, which he reached in two or three steps (*al-Futuhat al-makkiyya* I: 186). It is tempting to dismiss these visions as hallucinations induced by extreme ascetic practices or illness - on the occasion when he saw al-Khidr he had gone to the side of the ship because of a stomach pain which prevented him from sleeping - but Ibn al-‘Arabi offers a different explanation based on his perception of the nature of the cosmos.

In his view, the cosmos comprises a hierarchy of three distinct worlds or levels: the ‘world of spirits’, ‘the world of images’ and ‘the world of bodies’. The second of these - ‘the world of images’ (*‘alam al-amthal*), also called ‘the world of imagination’ (*‘alam al-khayal*) - plays a key role because of its intermediate position. It is the isthmus (*barzakh*) between the world of spirits and the world of bodies, the realm in which spirits are corporealized and bodies are spiritualized. The world of images is a really existent world, but in the waking state we are generally unaware of it; in our dreams, when our souls are no longer distracted by sensory input from the world of bodies, we function at this level, conversing with the departed and with those normally separated from us by geographical distances. What ordinary human beings experience only in their dreams, the mystic may experience at other times. Thus for example, when al-Khidr appeared to Ibn al-‘Arabi, this took place in the world of images, al-Khidr - who belongs to the world of spirits - being corporealized for the occasion.

The supposition of this intermediate world of images also furnishes the key to understanding both the miracles

performed by prophets and saints and some of the more bizarre descriptions of the hereafter in the *hadith*. As regards the miracles, Ibn al-‘Arabi’s starting point is the observation that we all can create things in our imagination or imagine things happening as we would like them to happen. The ‘perfect man’ is in addition endowed with extraordinary spiritual energy or *himma*, which enables him to bring the creatures of imagination out of the world of images into the world of bodies thus giving them existence. However, far from acting like a superman, he exercises restraint only employing his miraculous powers when commanded by God to do so. As regards the traditional descriptions of the hereafter, Ibn al-‘Arabi maintains that they should be understood as comparable to dream imagery. In a celebrated dream, the Prophet was given a cup of milk to drink, which in the waking state he subsequently interpreted as knowledge. What is impossible in the world of bodies - the corporealization of milk as knowledge - is perfectly possible in the world of images. Similarly in the hereafter, our works will be weighed in the scales and death will be brought in the form of a salt-coloured ram. We who are resurrected will really see these things, but we will see them in the world of images.

6 Assessment

In view of the volume of Ibn al-‘Arabi’s work, much of which is still unpublished, any assessment of his philosophy must remain highly tentative. Although he was influenced by earlier Sufis and was conversant with the works of the *falasifa* and the disputes between the Mu‘tazila and Ash‘ariyya, the dominant influence on his thought seems to have been the Neoplatonism of Plotinus as mediated by the *Epistles* of the Brethren of Purity (Affifi 1938: 174-94) (see Ikhwan al-Safa’; Neoplatonism in Islamic philosophy). Nevertheless he differs from them in at least two important respects. First, despite his use of emanationist language, it is clear that for him ‘emanation’ (*fayd*) is a figure of speech for what is more accurately described as self-revelation. Second, he does not simply take from the Qur’an and hadith convenient pegs on which to hang his doctrine, but rather offers what amounts to a profound esoteric commentary on both.

See also: Islamic theology; Mystical philosophy in Islam

NEAL ROBINSON

List of works

Ibn al-‘Arabi (after 1229) *Fusus al-hikam (The Bezels of Wisdom)*, ed. A. Affifi, Cairo, 1946; trans. R.W.J. Austin, *The Bezels of Wisdom*, New York: Paulist Press, 1980.(A late work which contains the quintessence of Ibn al-‘Arabi’s spiritual doctrine in the form of twenty-seven brief chapters named after prophets who epitomize different spiritual types.)

Ibn al-‘Arabi (*c.*1231-8) *al-Futuhat al-makkiyya (The Meccan Illuminations)*, Cairo, 1911; partial trans. M. Chodkiewicz *et al.*, *Les Illuminations de la Mecque: The Meccan Illuminations*, Textes choisis/Selected Texts, Paris: Sindbad, 1988.(The definitive synthesis of Ibn al-‘Arabi’s teaching, comprising 560 chapters which deal with every aspect of mystical knowledge.)

References and further reading

Addas, C. (1989) *Ibn ‘Arabi ou La quête du Soufre Rouge (Ibn al-‘Arabi and The Quest for Red Sulphur)*, Paris: Gallimard; trans. P. Kingsley, *Quest for Red Sulphur: The Life of Ibn ‘Arabi*, Cambridge: Islamic Texts Society, 1993.(Critical biography.)

Affifi, A.E. (1938) *The Mystical Philosophy of Muhyid Din-Ibnul Arabi*, Cambridge: Cambridge University Press. (Pioneering work, still useful although polemical and somewhat dated.)

Austin, R.W. J. (1971) *Sufis of Andalusia*, London, George Allen & Unwin. (Biographical essay and partial translations of *Ruh al-quds* and *al-Durrat al-fakhirah*, which give valuable insight into the spiritual milieu of Ibn al-‘Arabi’s early years.)

Chittick, W.C. (1989) *The Sufi Path of Knowledge: Ibn al-‘Arabi’s Metaphysics of Imagination*, Albany, NY: State University of New York Press.(Exposition of Ibn al- ‘Arabi’s thought, based primarily on *The Meccan Illuminations* with extensive excerpts translated by the author.)

Chittick, W.C. (1996a) ‘Ibn ‘Arabi’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 30, 497-509.(Clear and perceptive discussion of Ibn al-‘Arabi.)

Chittick, W.C. (1996b) ‘The School of Ibn ‘Arabi’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 31, 510-23.(Discussion of Ibn al-‘Arabi’s school.)

Chodkiewicz, M. (1992) *Un océan sans rivage: Ibn al-‘Arabi le livre et la loi (An Ocean Without Shore: Ibn*

al-‘Arabi, the Book and the Law), Paris, Seuil; trans. D. Streight, *Ocean Without Shore*, Albany, NY: State University of New York Press, 1993.(A study of the hermeneutical principles which govern Ibn al-‘Arabi’s approach to the Qur’an.)

Izutsu, T. (1983) *Sufism and Taoism*, Los Angeles, CA: University of California Press.(Pages 1-283 contain the classic account of Ibn al-‘Arabi’s ontology, based primarily on *The Bezels of Wisdom* as interpreted by early Muslim commentators. The book is a revised version of *A Comparative Study of the Key Philosophical Concepts in Sufism and Taoism*, 1966-7.)

Ibn ar-Rawandi (*c*.910?)

A highly enigmatic and controversial figure in the history of Islamic thought, Ibn ar-Rawandi wavered between a number of Islamic sects and then abandoned all of them in favour of atheism. As an atheist, he used reason to destroy religious beliefs, especially those of Islam. He compared prophets to unnecessary magicians, God to a human being in terms of knowledge and emotion, and the Qur'an to an ordinary book. Contrary to Islamic belief, he advocated that the world is without a beginning and that heaven is nothing special.

Medieval biographical dictionaries agree that Ibn al-Husain Ahmad ibn Yahya ibn Ishaq ar-Rawandi lived in Baghdad, but differ as to the form of his name and the date of his death, and indicate that he was intellectually unstable and that very little was known about his real thought. While he is best known as ar-Rawandi, he is also referred to as ar-Rindi, ar-Rawindi and ar-Riwindi. Also he is said to have died at a number of different dates, ranging from AH 243 to AH 301. The most accepted view is that he died about AH 245/AD 910 at the age of forty.

Islamic sources mention that Ibn ar-Rawandi was first a Mu'tazilite, then a Shi'ite and later an atheist. The same sources indicate that the Jews had warned Muslims that, as his father converted to Islam from Judaism after trying to refute the Torah, Ibn ar-Rawandi would attempt to refute the Qur'an after abandoning Islam. These sources agree that this happened, and that he wrote books for the Jews, Christians and even dualists (idolators) in which he attacked Islamic beliefs. He was said to have told the Jews to inform Muslims that Moses had said that there will not be any prophet after him. With the exception of Ibn al-Murtada (d. AH 436/AD 1044) and Ibn Khallikan (d. AH 680/AD 1283), Muslim authors distanced themselves from him and many called him 'the cursed', including Abu al-Hayyan al-Tawhidi and Abu al-'Ala' al-Ma'arri (d. AH 449/AD 1057), who themselves were considered as atheists in Islamic religious circles.

The reasons for his abandoning Mu'tazilism and later Islam entirely were a matter of controversy. Some believed that poverty pushed him to earn some money by writing books for the opponents of Mu'tazilism and Islam in general. The *Ma'ahid at-tansis (Known Citations)* of al-'Abbasi (d. AH 960/AD 1556) mentions that for four hundred dirhams Ibn ar-Rawandi wrote a book for the Jews, criticizing Islam. After he received the money he wished to refute it, but agreed not to do so after receiving one hundred dirhams more. Others were of the opinion that he abandoned Mu'tazilism because he did not reach the high positions in Mu'tazilite circles to which he aspired. Still others contended that the sense of rejection and loneliness he felt after having been isolated by the Mu'tazilites forced him to seek refuge in their opponents' circles. He himself claimed that he affiliated himself with different schools of thought, including atheism, in order to familiarize himself with their doctrines and learn from them.

Ibn ar-Rawandi's real thought remained somewhat unknown primarily because in the Middle Ages the authorities discouraged the reading of his books and banned their circulation. Most of the one hundred and fourteen books he wrote have been lost. Only parts of three of his works are extant. *Fadihat al-mu'tazila (The Scandal of the Mu'tazilites)* was preserved almost in its entirety and responded to by al-Khayyat (d. AH 300/AD 912) in *Kitab al-intisar (The Book of Victory)*. *Fadihat al-mu'tazila* is a response to *Fadilat al-mu'tazila (The Virtue of the Mu'tazilites)*, a work by al-Jahiz (d. AH 254/AD 868), in which the latter pointed out the vices of their opponents in addition to the virtues of the Mu'tazilites themselves. Following the heyday of the Mu'tazilite movement during the early Abbasid rule of al-Ma'mun, al-Mu'tasim and al-Wathiq (see Ash'ariyya and Mu'tazila), the movement felt the need to defend itself against attacks by various opponents; al-Jahiz was one of its defenders. In *Fadihat al-mu'tazila*, Ibn ar-Rawandi presents the views of all the major Mu'tazilite thinkers and tries to show that they suffered from inconsistencies. Many fragments of *ad-Damigh (A Refutation)*, another work of Ibn ar-Rawandi, are extant in Ibn al-Jawzi's *al-Muntazam fi at-tarikh (Organization in History)*. In *ad-Damigh*, Ibn ar-Rawandi attacks the Qur'an. Finally, parts of *az-Zumurrud (Diamond)* are also extant in the *Majalis (Councils)* of al-Mu'ayyad fi al-Din (d. AH 369/AD 979). In *az-Zumurrud*, Ibn ar-Rawandi focuses on proving the falsehood of prophets and prophecy, which he rejects in Islam and in general.

Ibn ar-Rawandi's tremendous courage in pursuing a rational path in religious debates forced him to reach conclusions not accepted by mainstream Islam. Thus he was attacked severely by the major Muslim thinkers as early as the fourth century AH (tenth century AD), including al-Kindi, al-Khayyat, Abu al-Hasan al-Ash'ari, Abu

‘Ali al-Jubba’i and al-Farabi.

In most of his later works, Ibn ar-Rawandi advocated rejection of religious doctrines, which he considered unacceptable to reason. Thus, he attacked the prophets and certain traditional interpretations and concepts of the Qur’an. Among his teachings were the ideas that prophets make the same kind of claims that magicians make, and that the world is eternal and its events do not prove that they have a first cause. The Qur’an, in his view, is not the eternal word of God, nor is its language miraculous; some human beings, such as al-Aktham ibn Saifi, made better statements than some of those found in the Qur’an. God was without knowledge until he created his knowledge: God is like an angry enemy who can remedy things only by imposing punishment on others and who is capable of wrongdoing. Since he can do these things directly, he needs no holy book and no prophet. However, a God who treats his creatures in this way is not wise. His lack of wisdom is also revealed in his requiring his creatures to obey him when he knows that they will not do so, and in placing them in hell for eternity if they disobey him. Heaven, as described in the Qur’an, has nothing desirable.

Ibn ar-Rawandi had a gloomy outlook on life. This is best expressed in some of his verses where he says: ‘The calamities of life are numerous and continuous. Its joy, on the other hand, comes to you as do holidays.’

See also: Ash‘ariyya and Mu‘tazila

SHAMS C. INATI

List of works

Ibn ar-Rawandi (before 910?) *Fadihat al-mu‘tazila (The Scandal of the Mu‘tazilites)*. (What remains of this work can be found in al-Khayyat, *Kitab al-Intisar war-radd ‘ala Ibn ar-Rawandi al-Mulhid*, ed. H.S. Nyberg, Cairo: Matba‘at al-Kutub al-Masriyya, 1925; trans. A.N. Nadir, *Le livre du triomphe et de la réfutation d’Ibn al Rawandi l’hérétique*, Beirut: Catholic Press, 1957. This also contains al-Khayyat’s response to the work. This is the best primary source for understanding the debates between Ibn ar-Rawandi and the Mu‘tazilites.)

Ibn ar-Rawandi (before 910?) *ad-Damigh (A Refutation)*.(Portions of this work can be found in Ibn al-Jawzi, *al-Muntazam fi tarikh al-muluk wal-umam*, ed. M.A.Q. ‘Ata and M.A.Q. ‘Ata, Beirut: Dar al-Kutub al-‘Ilmiyya, 1992.)

Ibn ar-Rawandi (before 910?) *az-Zumurrud (Diamond)*.(Parts of this work can be found in al-Mu‘ayyad fi al-Din, *al-Majalis al-mu’ayyadiyya lil-Mu’ayyad fi ad-Din ash-Shirazi*, ed. M.A.Q. ‘Abd al-Nasir, Cairo: Dar ath-Thaqafa, 1975.)

References and further reading

al-‘Abbasi (before 1556) *Ma‘ahid at-tansis (Known Citations)*, ed. M. ‘Abd al-Hamid, Cairo: Matba‘at as-Sa‘ada, 1947.(Includes a brief account of Ibn ar-Rawandi’s life, works and poetry.)

al-A‘sam, A. (1975) *History of Ibn Ar-Riwandi the Heretic*, Beirut: Dar al-Afaq al-Jadida.(A collection of the most important medieval Islamic sources, including those of Ibn al-Jawzi and al-Mu’ayyad fi ad-Din, that mention Ibn ar-Rawandi. It shows that the majority of Muslim thinkers in the Middle Ages rejected his views and tried to distance themselves from him.)

al-A‘sam, A. (1978-9) *Ibn ar-Riwandi fi al-maraji‘ al-‘arabiyya al-haditha (Ibn ar-Rawandi in Modern Arabic Sources)*, Beirut: Dar al-Afaq al-Jadida.(Another collection of Islamic sources, dating from the fourteenth century to the present. It reveals that Muslim thinkers continued to consider Ibn ar-Rawandi a heretic. Like the above work of al-A‘sam, it is helpful in presenting the most important views about Ibn ar-Rawandi, as they circulated in chronological order.)

al-Mu’ayyad fi al-Din (before 979) *al-Majalis al-mu’ayyadiyya lil-Mu’ayyad fi ad-Din ash-Shirazi (The Mu’ayyadiyya Councils by al-Mu’ayyad fi ad-Din of Shiraz)*, ed. M.A.Q. ‘Abd al-Nasir, Cairo: Dar ath-Thaqafa, 1975.(Includes parts of the *az-Zumurrud* of Ibn ar-Rawandi.)

Ibn al-Jawzi (before 1201) *al-Muntazam fi tarikh al-muluk wal-umam (Organization in the History of Kings and Nations)*, ed. M.A.Q. ‘Ata and M.A.Q. ‘Ata, Beirut: Dar al-Kutub al-‘Ilmiyya, 1992.(Includes portions of the *ad-Damigh* of Ibn ar-Rawandi.)

Stroumsa, S. (1985) ‘The Barahima in Early Kalam’, *Jerusalem Studies in Arabic and Islam* 6: 229-41.(Offers a non-traditional interpretation of Ibn ar-Rawandi’s thought, which helps in drawing attention to his early views that are usually neglected by the majority of Islamic sources.)

Ibn Bajja, Abu Bakr Muhammad ibn Yahya ibn as-Say'igh (d. 1138)

Ibn Bajja's philosophy may be summed up in two words; al-ittisal *(conjunction) and* al-tawahhud *(solitude). Conjunction is union with the divine realm, a union that reveals the eternal and innermost aspects of the universe. Through this union or knowledge, one is completed as a human being, and in this completion the ultimate human end, happiness, is achieved. Solitude, on the other hand, is separation from a society that is lacking in knowledge. Once united with the eternal aspects of the universe, one must isolate oneself from those who are not in the same state, who may therefore distract one from the supernatural realm through their ignorance and corruption.*

1 Life and works

Abu Bakr Muhammad ibn Yahya ibn as-Say'igh, known as Ibn Bajja (or Avempace in the West), was born in Saragossa, Spain, at an unknown date and died in Fez in North Africa in AH 537/AD 1138. In *Akhbar al-hukama' (Information About Wise People)*, al-Qifti mentions that Ibn Bajja died from being poisoned by rivals in the field of medicine. He was the teacher of Ibn al-Imam and Ibn Rushd (Averroes). His prominence was the result of his being the first in the West to show deep understanding of the views of some of his predecessors, such as Plato, Aristotle, al-Farabi, Ibn Sina (though Ibn Bajja never directly mentions him) and al-Ghazali. Thus he served as a link between the East and the West.

Ibn al-Imam edited his teacher's works in AH 534/AD 1135. They include medical works, commentaries on Aristotle and al-Farabi and original philosophical treatises. The most important of these treatises are *Tadbir al-mutawahhid (Management of the Solitary)*, *Risalat al-wada' (Essay on Bidding Farewell)* and *Risalat al-ittisal al-'aql al fa''al bil-insan (Essay on the Conjunction of the Intellect with Human Beings)*.

In spite of the criticism directed against Ibn Bajja by some of his contemporaries, such as Ibn Tufayl, he was highly regarded even by those critics themselves. Ibn Tufayl described him as possessing 'the sharpest mind', 'the soundest reasoning' and 'the most valid opinion' of those following the first generation of thinkers. The reference here is to the Arab/Spanish thinkers who lived from the earlier part of the second half of the tenth century to the end of the eleventh century, which marks the emergence of the second generation of Arab/Spanish thinkers. The first generation is characterized by interest in learning about ancient logic and philosophy; the second, beginning with Ibn Bajja, is characterized by originality of philosophical writings. His student Ibn al-Imam describes him as the marvel of his time in depth of philosophical knowledge.

2 The human soul

The two most essential pillars on which Ibn Bajja's philosophy rests are *al-ittisal* (conjunction) and *al-tawahhud* (solitude or union). *Al-ittisal* is that of the philosopher with the agent intellect (the lowest celestial intelligence and home to the universals), and *al-tawahhud*, when used in the sense of solitude, is that of the philosopher in society. *Al-tawahhud* is also used in the sense of *al-ittisal* (union). Like his Eastern predecessors, al-Farabi and Ibn Sina, Ibn Bajja was most concerned with the ultimate human objective, the intellectual or philosophical ideal, which in turn is in conjunction with the agent intellect through grasping the universals (see Universals). This conjunction results in self-completion, which is the same as happiness. To understand the ultimate human objective and the instruments through which it is attained, Ibn Bajja first traces the development of the human soul, the only means to conjunction.

The human soul, he believes, develops from the plant to the animal and finally to the rational life. The plant life is the embryonic life, which provides one with nourishment and growth. With the progress from the plant to the animal life, which is the sensitive life, one moves from mere vegetation to sensation, movement and desire. Sensation is acquired either by the five external senses or by the internal senses, the common sense, the imagination and memory. By acquiring thought, or the highest human state, one rises to the level of rational speculation. While the human soul incorporates these three states, human nature or essence as such is described as *'aql* (reason or intellect). In the tradition of al-Farabi and Ibn Sina, Ibn Bajja teaches that the intellect is either potential or actual. When it is potential, it has the capacity for acquiring its proper object, the intelligible form (*as-sura al-'aqliyya*) or, as Ibn Bajja is fond of calling it, the spiritual form (*as-sura ar-ruhaniyya*), the form that belongs to the soul. When it is actual, it is identified with its object (see Soul in Islamic philosophy).

3 Intelligible forms

Four types of intelligible forms in themselves are distinguished: those of the bodies that have an eternal circular motion; those of the agent and acquired intellects (the acquired intellect being the highest level of the human intellect, which results from conjunction with the agent intellect); those of the material world, which are stripped by the external senses from external particular things; and those in the internal senses. The first are in all respects immaterial, that is, lacking any necessary relation to matter. The second are in themselves immaterial; they have only an inessential relation to matter, the agent intellect by virtue of causing the material forms and the acquired intellect by virtue of completing them. The third are essentially linked to matter; they exist in matter and are made intelligible only through the mediation of the external senses. The fourth lie between the second and the third and are therefore in part material and in part not. Since, to Ibn Bajja, immateriality necessitates universality and materiality necessitates particularity, the following conclusion is drawn: the first, second and, in part, fourth types of intelligible forms are universal, while the third and, in part, fourth are particular.

Along Greek and Islamic lines, Ibn Bajja insists that the completion of every nature is the best for that nature and its highest objective. However, the completion of a nature requires that nature to acquire its proper objects. Since the nature of being human is reason, and since the proper objects of reason are the separate forms or universals which reside in the agent intellect, to acquire these objects is to complete human nature. A human being grasps the purely material forms through the external senses and benefits from them, and grasps the forms that are somewhat purified from matter through the internal senses and also benefits from them. Grasping the separate forms through reason is, however, the most befitting to human nature, and hence is best for it. On the basis of the kind of power one uses to grasp the intelligibles, Ibn Bajja divides people into three groups: the multitudes, the theorists and the philosophers. The multitudes grasp the intelligibles with the external senses, the theorists grasp them with the internal senses and the philosophers grasp them with reason. Only the philosophers can be classified as happy, for they acquire the universals in themselves, the objects of reason.

In *al-Ittisal*, Ibn Bajja states that it is as if the multitudes grasp the sun as reflected in the mirror after its having been reflected in water. The theorists grasp it as reflected in water; the philosophers grasp it in itself. He compares the multitudes to people in a cave in which the sun never shines. If they are in the very inside of the cave, objects appear to them in a state of darkness; if instead they are at the entrance of the cave, objects appear to them in the shade. As the people of the cave, regardless of their place, have no idea of what it means to see the sun, so also the multitudes have no idea of what it means to grasp the intelligibles rationally. He compares the theorists, on the other hand, to those who have gone out of the cave, where they can see the sun shedding its light on things and making the colours of things visible in themselves. He does not believe, however, that the happy ones can be compared to any beings with physical vision 'since they and the thing they grasp become one'. Being in the happy state is like having vision itself transformed into light. In other words, the multitudes grasp the reflection of the reflection of a thing, the theorists grasp the reflection of a thing and the philosophers grasp the thing itself.

4 The state of conjunction

To be a philosopher, or to have conjunction with the universals in the agent intellect, is to have ultimate human happiness and to experience 'witnessing' of the truth. The happy ones are incorruptible, eternal and 'numerically one with no difference among them in themselves whatsoever'. Their instruments, or their bodies, are the only things that differentiate them from each other. They are incorruptible and eternal because the intelligibles with which they are identified are so, and are numerically one because they are all identified with the same intelligibles.

Ibn Bajja rejects the Sufi concept that the ultimate human end is the pleasure (*al-ladhdha*) which results from witnessing (*mushahadat*) the divine world internally, in a higher sensible form as presented by the common sense, imagination and memory. According to him, this amounts to saying that 'having pleasure internally is the ultimate objective of knowing the Truth through the internal senses'. However, this is not the case since this pleasure is not sought for its own sake. In support of his view, Ibn Bajja mentions among other things that if pleasure of the internal senses were the ultimate human end, then reason (which is a higher power than the internal senses) as well as its knowledge would be superfluous and futile.

Because knowledge of the internal senses is higher than that of the external senses, the objects of the former being

more enduring than those of the latter, the pleasure of the internal senses is higher than that of the external senses. The assertion that the former objects are more enduring than the latter is demonstrated by the fact that one can imagine the existence of something that has ceased to exist externally. However, even knowledge of the internal senses still falls short of reaching the sublimity of the knowledge of reason since the objects of the former do not endure as much as those of the latter. Only the objects of the latter endure permanently, unaffected by the forgetfulness or even the removal of their subject. Knowledge of these permanent objects gives the knower a permanent status since the knower and the known in this case are one. It also gives the highest and most permanent pleasure. The state of happiness is one which cannot be described in language, owing to its nobility, pleasure, beauty and goodness. When human beings reach this ultimate end, they become simple intellects of which it is true to say that they are nothing but divine.

5 Solitude

The knower, or happy person, may exist in society in either a virtuous or a nonvirtuous city. A virtuous city is one whose members are all complete in knowledge, while in a nonvirtuous city the contrary is the case.

If perfected people exist in a nonvirtuous city, they must live in isolation from the rest of society, for their complete knowledge makes them 'strangers' or 'weeds', that is, those whose true opinions are contrary to the opinions of society. While isolation from society is not natural or essential for a human being in the natural or virtuous city, it is accidental to one's nature and must be practised in order to preserve oneself from the corruption of the nonvirtuous cities.

See also: Epistemology in Islamic philosophy; Ethics in Islamic philosophy; Soul in Islamic philosophy

SHAMS C. INATI

List of works

Ibn Bajja (before 1138) *Rasa'il ibn Bajja al-ilahiyya (Ibn Bajja's Metaphysical Essays)*, ed. M. Fakhry, Beirut: Dar al-Jil, 1992.(The best available collection of Ibn Bajja's sociopolitical and metaphysical treatises. It includes the most important of such treatises, *Tadbir al-mutawahhid (Management of the Solitary)*, *al-Ittisal al-'aql al-fa''al bil-insan (Essay on the Conjunction of the Intellect with Human Beings)* and *Risalat al-wada' (Essay on Bidding Farewell)*.)

Ibn Bajja (before 1138) Commentary on Aristotle's *Physics*, ed. P. Lettinck, *Aristotle's Physics and Its Perception in the Arabic World with an Edition of the Unpublished Parts of Ibn Bajja's Commentary on the Physics*, Leiden: Brill, 1994.(Includes a survey of Ibn Bajja's commentary on Aristotle's *Physics*, his longest commentary, and its influence on Ibn Rushd.)

References and further reading

Al-'Alawi, J.D. (1983) *Mu'allafat Ibn Bajja (Ibn Bajja's Works)*, Beirut: Dar ath-Thaqafa.(The best available classification of Ibn Bajja's works, including a discussion of the sources and influences of their ideas.)

Goodman, L. (1996) 'Ibn Bajjah', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 21, 294-312.(Discussion of the thinker and his period, describing in detail the cultural context within which he worked.)

Ibn Rushd (1126-98) *Talkhis kitab an-nafs (Summary of the Book on the Soul)*, ed. A.F. al-Ahwani, Cairo: Maktabat an-Nahda, 1950.(Includes Ibn Bajja's *Risalat al-ittisal*. Useful, not only in offering Ibn Bajja's view on the soul, but also in comparing it with those of Ishaq ibn Hunayn, al-Kindi and Ibn Rushd.)

al-Qifti, A. (*c*.1172) *Akhbar al-hukama' (Information About Wise People)*, ed. J. Lippert, Leipzig: Maktabat al-Mutanabbi, 1903.(Includes a section on Ibn Bajja's life and works.)

Zainaty, G. (1979) *La morale d'Avempace (The Ethics of Avempace)*, Paris: Vrin. (Useful overview of Ibn Bajja's metaphysics, psychology and ethics.)

Ibn Daud, Abraham (*c*.1110-*c*.1180)

Ibn Daud was born in Cordoba and died in Toledo. In Jewish texts he is known as Rabad, an acronym of his Hebrew name, Rabbi Abraham ben David. He was known to medieval Christian philosophers by a variety of names, including Avendauth and possibly John of Spain as well. His Sefer ha-Kabbalah (The Book of Tradition), regarded by some scholars as the first comprehensive study of Jewish history, is an extended argument for the authority of rabbinic Judaism on the grounds that it is an unbroken tradition of authentic sources, from the Mosaic origins through the first two Jewish commonwealths, the exile, and down to the author's time. His major work in philosophy is Al-'Aqida al-Rafi'a (The Exalted Faith), composed in 1160 in Judaeo-Arabic, the form of Arabic written in Hebrew characters that was commonly used by Jewish scholars and thinkers in the Muslim milieu. It survives only in two late fourteenth-century Hebrew translations, one by Samuel Motot, entitled Ha-Emunah ha-Nissa'ah, and the other, better known, by Solomon ibn Labi, entitled Ha-Emunah ha-Ramah.

1 Aristotelianism

Abraham ibn Daud was the first philosopher to apply the diverse elements of Aristotelian philosophy to forge a philosophy of Judaism. He initiated a new direction in Jewish philosophy that continued in the writings of Maimonides (1135-1204) and culminated in the philosophy of Gersonides (1288-1344). The relationship of these Jewish Aristotelians might be compared to that of Locke, Berkeley and Hume in British philosophy. If Ibn Daud is not always as rigorous as Maimonides or Gersonides, it is because they build on the foundations he laid down. But no Jewish Aristotelian's work is as comprehensive as Ibn Daud's, and on many topics his treatment is more thorough or philosophically more sophisticated than that of Maimonides. For example, the concept of the soul and its faculties is of central importance to Maimonides in the *Dalalat al-Ha'irin (Moreh Nevukhim, Guide to the Perplexed)*. Yet he presents no clear theory of the soul in the *Dalalat al-Ha'irin (Moreh Nevukhim, Guide to the Perplexed)*; and, where he does deal with psychology at length, as in *Shemonah Perakim (Eight Chapters)*, many of his assertions become clear only with the aid of Ibn Daud's discussions in *Ha-Emunah ha-Ramah*. For instance, Maimonides states that while souls have multiple powers there are not multiple souls in each individual; and a particular faculty in one species is not the same as the corresponding faculty in another, even though the faculties bear the same name. Thus, both animals and humans exercise nutritive functions and have nutritive 'souls', but the term names a faculty, not a separate being; and the nutritive function in animals is not the same as that of humans. Ibn Daud develops and defends these views at length; Maimonides merely asserts them.

2 *Ha-Emunah ha-Ramah*

Ha-Emunah ha-Ramah comprises three books. The first explains the presuppositions of Aristotelianism. The intended audience consists of those cultured Jews who know of, but little about, the new Aristotelian science. In the Western lands of Muslim civilization that science was now displacing the older atomism of the *kalam* (see Islamic theology). In the second book Ibn Daud lists six basic principles of Judaism and explains them in the light of the new science. In the third book he applies to ethics the presuppositions and principles he has laid out.

The status of the third book, entitled *Barefuah ha-Nafshiyah (The Healing of the Soul)*, within the whole is problematic. Ibn Daud tells us that his goal in composing the entire work was to solve the problem of human choice and determinism. That question was dealt with directly in the final chapter of the second book. Given his stated intention, this is where *Ha-Emunah ha-Ramah* should end. Everything that goes before can be seen as material arrayed in support of this final discussion; but the internal structure of all the material presented in every known manuscript of the third book is incoherent. For example, it is supposed to comprise two chapters, but all the existing manuscripts contain only a first chapter that deals with a potpourri of issues in ethics. At best Book Three is an addendum, and one that has an unfinished look compared to the organized state of what went before. For these reasons the following summary is limited to the first two books.

Book One contains eight chapters. The first three define key technical terms in the new Aristotelian science: substance, accident, and the ten categories (ch. 1), form and matter (ch. 2), and motion (ch. 3). There follow two chapters on physics. Here Ibn Daud spells out the theses that material bodies possess neither actual nor potential infinity (ch. 4) and that all motion comes from a mover and requires a first mover (ch. 5). In two chapters on rational psychology he describes the nature and powers of the soul (ch. 6) and defends the claim that the rational

power is nonphysical (ch. 7). The last chapter deals with astronomy. Here Ibn Daud argues for the claim, critical to his solution of the problem of determinism, that the heavens are rational, living organisms, their motion intentional.

Book Two uses the ideas of the first book to explain what Ibn Daud deems to be the basic principles of the Jewish faith and religious law (see Halakhah). The first four principles address the existence and nature of God. The fifth defends the authority of the rabbinic tradition. The sixth turns to the target Ibn Daud specified in his introductory abstract, the problem of human choice.

The first two principles are that God is a necessary being and that he is one. Only a necessary being can be truly one (ch. 1). God's unity is essential, not accidental and admits no plurality of any kind (ch. 2). Ibn Daud's third principle is that all affirmative characterizations of God, even the ascription to him of unity and necessity, are equivocal, bearing different senses from those of ordinary usage. As applied to God, they express either a negation or a relation. Like Maimonides, then, Ibn Daud espouses a negative theology (see Negative theology). But his acceptance of relational statements about God saves him from the kind of attack that Maimonides' theology later received from such Aristotelians as Thomas Aquinas and Gersonides.

Ibn Daud's fourth principle explains divine actions in terms of the mediation of the beings that Scripture calls angels. Ibn Daud identifies them with the separate intellects of Aristotelian cosmology. He devotes two chapters to proofs of their existence. The first argues from epistemology and the powers of the soul. The second is based on physics and astronomy. A third chapter surveys the entities in the universe and the kinds of angels/separate intellects who govern them on behalf of God.

The fifth principle is that the Hebrew Scriptures as interpreted by the rabbis (not the Karaites, Muslims or Christians) are an authoritative source of truth. An introductory abstract defends the general thesis that authentic traditions can make veridical claims. The next two chapters argue that rabbinic Judaism is such a tradition. The first of these deals with the nature of the prophecy recorded in Scripture and categorizes its degrees, laying a foundation for the argument that the prophecy of Moses, as recorded in the Torah, is an unimpeachable witness to the word of God. The second argues that the transmission of Moses' report through the early and later rabbinic sages was faithful to his original testimony. Thus statements in the rabbinic tradition preserve the same epistemic status as direct reports of sense experience or reliable traditions of such reports. This chapter recaps the more extended argument of Ibn Daud's earlier work on the continuity of Jewish ideas, *Sefer ha-Kabbalah (The Book of Tradition)*.

The final principle is that human choice is not obviated by divine or natural necessity. The problem as Ibn Daud states it: if God rules over all things, no human being would have any choices. But that cannot be. For God issues commands and punishes disobedience, and no one can be justly punished for acts that were not free. Nor can one be commanded to do where one has no choice. So God's rule does not exclude human choice. But how can there be anything over which God does not have dominion? Scripture only exacerbates the dilemma. For some texts seem to say that God determines everything, while others imply that we do have choices.

Ibn Daud devotes two chapters to the problem. The first grounds his solution in his earlier discussion of God's attributes. Since all terms predicated of God are equivocal, no statements about divine power and rule can be understood literally. The second chapter presents Ibn Daud's answer. Everything is determined by God through his ordering of the universe, but this ordering allows us to make choices. So people are morally responsible and are subject to divine providence, punishment and reward.

Ibn Daud does not list creation as a fundamental principle of Judaism, although his predecessor, Saadiah Gaon al-Fayyumi (882-942), made it the cornerstone of Jewish belief. Perhaps Ibn Daud omitted creation because it was the one central belief in Judaism that could not be explained or defended from the standpoint of the new Aristotelianism. The conflict between Aristotelian eternalism and biblical creationism was to become a central theme in the philosophies of Maimonides and Gersonides.

3 Influence

Beyond their intrinsic worth and interest as contributions to philosophy, Ibn Daud's ideas had a powerful impact on later Jewish thought. While everything he said was rooted in rabbinic tradition, his formulations, applications

and syntheses were highly original and presaged the issues that were to dominate the history of Jewish Aristotelianism. A key example is his understanding of the relationship between religion and science. For Ibn Daud and his Jewish Aristotelian successors, there is no conflict between religious and scientific truth. Truth is one. Religion no less than science makes truth claims; and these claims, if true, must cohere with those of science. Furthermore, since Jewish faith cannot be confined to one part of one's life, no understanding of Judaism that excludes the insights of science can be 'Torah true'. This challenging model of the relationship between science and religion allowed Ibn Daud to incorporate the new Aristotelian science into the outlook of rabbinic Judaism and set the agenda for later Jewish philosophers from Maimonides through Gersonides. The synthesis he inaugurated survived until its eventual overthrow by Hasdai Crescas (1340-1412) and Benedict de Spinoza (1634-77), when Aristotelianism was itself surpassed.

See also: Aristotelianism, medieval; Halakhah; Maimonides, M.

NORBERT M. SAMUELSON

List of works

Ibn Daud, A. (1160) *Sefer ha-Kabbalah*, trans. G.D. Cohen, *The Book of Tradition*, Philadelphia, PA: Jewish Publication Society, 1967.(An attack on the contentions of Karaism and a justification of rabbinical Judaism by the establishment of a chain of traditions from Moses to the author's own time.)

Ibn Daud, A. (1160) *Ha-Emunah ha-Ramah*, trans. N.M. Samuelson, ed. N.M. Samuelson and G. Weiss, *The Exalted Faith*, Rutherford, NJ: Fairleigh Dickinson University Press, 1986.(Written originally in Judaeo-Arabic as *Al-'Aqida al-Rafi'a*. The work is now known by the title cited here, which is that of the Hebrew translation made in the late fourteenth-century by Solomon ibn Labi. The translated Hebrew title of the third part of the work is *Barefuah ha-Nafshiyah (The Healing of the Soul)*. The work introduced the phase of Jewish philosophy that led to Maimonides. This phase differs from earlier ones in its more thorough systematic from derived in large measure from Aristotle.)

References and further reading

Arfa, M. (1954) 'Abraham ibn Daud and the Beginnings of Medieval Jewish Aristotelianism', unpublished Ph.D dissertation, New York: Columbia University, 1954.(An interesting read.)

Elbogen, I. (1915) 'Abraham ibn Daud als geschichtsschreiber', in *Festschrift zum Siebzigsten Geburtsage Jabok Guttmanns*, Leipzig: G. Fock.(An examination of the importance of Ibn Daud as a historical writer.)

Maimonides, Moses (1168) *Shemonah Perakim (Eight Chapters)*, trans. with notes by J. Gorfinkle, *The Eight Chapters of Maimonides on Ethics*, New York, 1912; repr. New York: AMS, 1966.(Part of Maimonides' commentary on the Mishnah, *Kitab al-Siraj* (translated into Hebrew as *Sefer ha-Maor, hu Perush ha-Mishnah*). The work forms a thematic introduction to the mishnaic collection of ethical aphorisms known as *Pirkei Avot (The Sayings of the Fathers)*.)

Maimonides, Moses (*c*.1190) *Dalalat al-Ha'irin (Moreh Nevukhim, Guide to the Perplexed)*, ed. S. Munk, *Le Guide des Égarés*, Arabic text, critically edited, with annotated French translation, Paris, 1856-66, 3 vols; repr. Osnabrück: Zeller, 1964; trans. S. Pines, with an introductory essay by L. Strauss, *The Guide of the Perplexed*, Chicago, IL: Chicago University Press, 1963, 1969, 2 vols.(Maimonides' best-known work, which takes the form of a letter to a disciple who is confused by the seeming disparity between biblical and scientific/philosophical ideas.)

Samuelson, N.M. (1977) 'Ibn Daud's Conception of Prophecy', *Journal of the American Academy of Religion* 45: 883-900.

Samuelson, N.M. (1979) 'Causation and Choice in the Philosophy of ibn Daud', in *The Solomon Goldman lectures*, vol. 2, Chicago, IL: Spertus College.(A study of Ibn Daud's account of free will and determinism.)

Samuelson, N.M. (1982) 'Ibn Daud and Franz Rosenzweig on Other Religions: A Contrast Between Medieval and Modern Jewish Philosophy', in *Proceedings of the Eighth World Congress of Jewish Studies, Division C: Talmud and Midrash, Philosophy and Mysticism, Hebrew and Yiddish Literature*, Jerusalem: World Congress of Jewish Studies, 75-80.(A comparative study of what these two major Jewish philosophers have to say as Jewish thinkers about other religions.)

Ibn Ezra, Abraham (1089-1164)

The philosophy of Ibn Ezra attained broad influence in Jewish literature through his Bible commentaries, included to this day in rabbinic Bibles. Born in Tudela, Spain, he was forced in later life (1140 until his death) to wander widely, at length settling in Rome and Lucca, where he composed some of his greatest works. A friend and, by some traditions, son-in-law of the poet-philosopher Judah Halevi, whom he mentions occasionally, he was himself a poet and wrote prolifically on grammar, exegesis, philosophy, medicine, astronomy and astrology. The many editions and manuscripts of his works attest their popularity, and some, especially on astronomy and astrology, were translated into Latin and then into French, Spanish, English and German.

Neoplatonic in orientation, with a Neo-Pythagorean fascination with numerology, Abraham ben Meir ibn Ezra's philosophy is often difficult to follow, since most of his works are unsystematic in exposition and elliptical in style. Frequently he engages in technical issues of grammar or literary analysis. His commentaries sometimes note 'this is a mystery (*sod*)' or 'the intelligent (*maskil*) should keep silent'. Such comments can signal the profundity of an issue, but at times Ibn Ezra seems deliberately to mute his radical conclusions.

Reason (*sekhel*) for Ibn Ezra is at the core of revelation, which is not just a historical event but a rational process. A proper understanding of Scripture thus demands a rational exegetical method grounded in Hebrew grammar and seeking the plain sense (*peshat*) of the text, not subordinate to homiletics or allegories. Yet revelation cannot contradict reason. So when a literal interpretation contradicts our rational or empirical knowledge, we must read nonliterally. Ibn Ezra generally upholds rabbinic tradition, but he is critical of rabbinic homiletical liberties (for example, in reconciling the variant wordings of the Decalogue in Exodus 20 and Deuteronomy 5) and of Karaite literalism (regarding the *lex talionis* in Exodus 21, for instance). Of both he says, 'reason cannot tolerate (*ein ha-da'at sovelet*)' such views. For 'reason is implanted in the heart by divine wisdom…. The judgement of reason is the foundation, and the Torah was not given to those who lack reason. The angel between man and his God is his reason'.

Spinoza, who founded modern critical readings of Scripture, saw Ibn Ezra as his forerunner in the belief that 'it was not Moses who wrote the Pentateuch, but someone who lived long after him'. 'Aben Ezra,' he writes (1670), '…a man of enlightened intelligence, and no small learning, who was the first, so far as I know, to treat of this opinion, dared not express his meaning openly, but confined himself to dark hints' (1670: ch. 8).

Ibn Ezra does question such passages as Genesis 12: 6, 22: 14, 36: 31-6, Exodus 6: 28, Numbers 21: 1-3, Deuteronomy 1: 2, and 34: 1-12, but his conclusion is not Spinoza's. Faithful to rabbinic tradition, he affirms as revealed the authority of the later prophets, the 'Oral Torah' of the rabbis and even the textual work of the medieval Masoretes (see Halakhah). Occasional interpolations do not undermine the authority of revelation but prove it, since revelation must make sense, and it would be absurd to ascribe, say to Moses, statements he could not meaningfully have made.

Ibn Ezra's interest in astrology was shared by some Jewish philosophers like Abraham Bar Hayya and Levi Ben Gershom (see Gersonides) but vigorously opposed by others such as Maimonides. For Ibn Ezra astrology can explain terrestrial phenomena and, so, some biblical phenomena scientifically (see Commentaries on Genesis 31: 19, 1 Samuel 19: 13, and Numbers 21: 8). Thus, unlike Maimonides, Ibn Ezra draws no sharp line between scientific astronomy and judicial astrology.

His Neoplatonic assumption that the terrestial realm is subject to influences from the incorruptible celestial spheres, which in turn are governed from the supernal realm of the rational soul (*neshamah*) and the angels, all under the sovereignty of God, the 'absolute One' (see Commentary on Exodus 3: 15, and 6: 3), prompts Ibn Ezra to seek astral explanations of terrestial, especially human, affairs. But he stoutly rejects star worship. The celestial bodies are 'servants' (*mershartim*), without independent will or conscious purpose; their activity is mechanical (see Commentary on Exodus 33: 21, citing the Epistles of the Sincere Brethren of Basra) (see Ikhwan al-Safa').

Ibn Ezra mitigates the determinism of astrology by arguing that astral influence is not merely a function of the celestial configuration (*ma'arekhet*) but also of the physical consititution (*toledet*) of the recipient (*meqabbel*) of influence (Introduction to Ecclesiastes). This explains why one astral agent can produce diverse terrestrial effects.

It also leaves room for free will. Indeed, it is the predictability of astral effects that enables one to take measures to avoid their impact. The event will occur, but those who understand can escape its effect. Reason, after all, transcends the intermediate realm of the stars.

The Torah, deriving from the supernal rational realm, affords a systematic avenue of escape from astral determinations to those Jews who observe it. Building on the Talmudic dictum ‘Israel has no constellar sign (*mazal*)’, Ibn Ezra reasons: ‘it is well established that every nation has a known star or constellation. But God gave Israel a great superiority, since He rather than a star is their guide - for "Israel is God’s portion"’ (Commentary on Deuteronomy 4: 19). The Torah imparts knowledge to its adherents, enabling them to escape astral influences. They do so naturalistically, not theurgically.

For Ibn Ezra revelation and reason are ultimately perfectly congruent. His critical reading of the biblical text and his astrological interpretations of some biblical passages arise from his consistent application of a naturalistic and rationalistic exegetical method and express his commitment to the view that rationality is inherent in revelation itself.

See also: Cosmology; Halakhah; Neoplatonism in Islamic philosophy

RAPHAEL JOSPE

List of works

Ibn Ezra, A. (1970) *Kitvei R. Avraham ibn Ezra*, Jerusalem: Makor, 4 vols.(Collected free-standing works edited by several hands in the nineteenth century and photomechanically reprinted here.)

Ibn Ezra, A. (1975) *The Religious Poems of Abraham Ibn Ezra*, Jerusalem: Israel Academy of Sciences and Humanities, 2 vols.(A critical, annotated edition.)

Ibn Ezra, A. (1985) *Abraham Ibn Ezra Reader*, ed. L. Israel, New York and Tel Aviv: Israel Matz.(In the original Hebrew, this collection comprises valuable annotated texts of selected works, including poetry, Bible commentaries, philosophy and theology.)

Ibn Ezra, A. (1986) *Ibn Ezra’s Commentaries on the Pentateuch*, ed. A. Weiser, originally published as a separate volume; repr. in *Humash Torat Hayyim*, Jerusalem: Mosad ha-Rav Kook, 1986.(In Hebrew. Although not without errors, this edition is far more reliable than the standard texts printed in traditional rabbinic Bibles.)

Ibn Ezra, A. (before 1164) *Perush Rabbenu Avraham ibn Ezra al Yishayahu (Commentary of Ibn Ezra on Isaiah)*, ed. M. Friedländer, *The Commentary of Ibn Ezra on Isaiah*, London: Society of Hebrew Literature, Truebner and Co., 1873.(Classic presentation of an important text, with English translation and commentary.)

Ibn Ezra, A. (before 1164) *Reshit Hokhmah (The Beginning of Wisdom)*, ed. and trans. R. Levy and F. Cantera, *The Beginning of Wisdom: An Astrological Treatise by Abraham ibn Ezra*, Baltimore, MD: Johns Hopkins University Press, 1939.(An important study of Ibn Ezra’s astrological theories.)

Ibn Ezra, A. (before 1164) *Sefer ha-Yesodot (Book of Foundations)*, French trans. J. Halbronn, *Le Livre des Fondements Astrologiques*, Paris: Retz, 1977.(An important study of Ibn Ezra’s astrological theories.)

References and further reading

Assaf, S. *et al*. (1969) ‘Ibn Ezra, Abraham’, in *Encyclopedia Hebraica*, Jerusalem: Encyclopedia Publishers, vol. 1, 210-15.(A study of Ibn Ezra’s life, thought and contributions to biblical exegesis, Hebrew poetry and grammar.)

Assaf, S. *et al*. (1972) ‘Ibn Ezra, Abraham’, in *Encyclopedia Judaica*, Jerusalem: Keter, vol. 8, 1163-70.(Account of Ibn Ezra’s life, beliefs and contributions to Hebrew poetry, grammar and biblical exegesis.)

Cohen, J. (1983) Mishnato ha-pilosofit Datit Shel Rabbi Avraham ibn Ezra (The Philosophical-Religious Teaching of Rabbi Abraham ibn Ezra), unpublished Ph.D. dissertation, Israel: Bar Ilan University.(The most comprehensive recent overview of Ibn Ezra’s philosophy; with many valuable references.)

Bacher, W. (1904-12) ‘Ibn Ezra, Abraham’, in *Jewish Encyclopedia*, New York: Funk and Wagnells, vol. 6, 520-4.(Still a useful survey of Ibn Ezra’s life and works.)

Esteban, F.D. (ed.) (1990) *Abraham ibn Ezra and His Age: Proceedings of the International Symposium*, Madrid: Association Espanola de Orientalistas.(An important collection of recent international scholarship on diverse aspects of Ibn Ezra’s work and thought.)

Friedländer, M. (1877) *Essays in the Writings of Abraham ibn Ezra*.(Despite a conservative approach to Ibn Ezra’s philosophy and to his critical approach towards biblical exegesis, this remains the most complete treatment of his life and work.)

Goldstein, D. (ed.) (1971) *The Jewish Poets of Spain: 900-1250*, Harmondsworth: Penguin, 153-62.(Selected poems by Ibn Ezra in English translation.)

Guttman, J. (1964) *Philosophies of Judaism*, New York: Holt Rinehart, Winston, 118-20.(Places Ibn Ezra's work in its Jewish philosophical context.)

Halbronn, J. (1985) *Le Monde Juif et l'Astrologie*, Milan: Arche Milano.(Includes examination of Ibn Ezra's astrological ideas.)

Jospe, R. (1994) 'The Torah and Astrology According to Abraham ibn Ezra', *Proceedings of the Eleventh World Congress of Jewish Studies* (Division C) 2: 17-24; a much fuller version, in Hebrew, appears in *Da'at* 32-3: 31-52.(A study of the place of the Torah in Ibn Ezra's astrological theory.)

Jospe, R. (1996) 'Biblical Exegesis as a Philosophical Literary Genre: Abraham ibn Ezra and Moses Mendelssohn', in E. Fackenheim and R. Jospe (eds) *Jewish Philosophy and the Academy*, Rutherford, NJ: Fairleigh Dickinson University Press.(Discusses philosophic issues in Ibn Ezra's exegesis.)

Kiener, R. (1987) 'The Status of Astrology in the Early Kabbalah: From the Sefer Yesirah to the Zohar', in *Jerusalem Studies in Jewish Thought* 6: 1-42.(Places Ibn Ezra's astrology in the context of the medieval Jewish astrological tradition.)

Simon, U. (1985) 'Ibn Ezra Between Medievalism and Modernism: The Case of Isaiah XL-LXVI', in *Supplements to Vetus Testamentum* 36: 257-71.(Important treatment of Ibn Ezra's exegetical methods.)

Simon, U. (1982) 'Bible: Exegesis, Abraham Ibn Ezra', in *Encyclopedia Biblica*, Jerusalem: Bialik Institute, vol. 8, 671-80.(A study of Ibn Ezra's methodology as a Bible exegete.)

Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Cambridge: Cambridge University Press, 93-112.(Valuable especially because it makes use of manuscript materials; includes discussion of Ibn Ezra's astrological interests.)

Spinoza, B. de (1670) Tractatus Theologico-politicus (Theological-Political Treatise), in *The Chief Works of Benedict de Spinoza*, vol. 1, trans. S. Shirley, Leiden: Brill, 1991.(Spinoza regarded Ibn Ezra highly and credited him with being the first critic of the traditional religious view of Moses as the author of the entire Pentateuch.)

Twersky, I. and Harris, J. (eds) (1993) *Rabbi Abraham ibn Ezra: Studies in the Writings of a Twelfth Century Jewish Polymath*, Cambridge, MA: Harvard University Press.(An up-to-date collection of papers by leading scholars on the thought and work of Ibn Ezra.)

Wolfson, E. (1990) 'God, the Demiurge and the Intellect: On the Usage of the Word *Kol* in Abraham ibn Ezra', in *Revue des Études Juives* 149 (1-3): 77-111.(Philosophical cosmology in the exegetical writings.)

Ibn Ezra, Moses ben Jacob (*fl.* 1055-1135)

Ibn Ezra was an exegete, Jewish scholar and one of the foremost Hebrew poets of medieval Spain. Although none of his systematic biblical commentaries have been preserved, two important works survive in Judaeo-Arabic prose, both dealing with biblical literary theory, rhetoric and philosophy. The literary dimension of his work makes Ibn Ezra a forerunner of modern biblical criticism. His speculative system, deeply influenced by Neplatonism, was to have a profound impact on the early Spanish Kabbalists.

Born in Granada into a family of government officials, Ibn Ezra was a disciple of the famous Talmudist Isaac ibn Ghiyath (1038-88) at Lucena, where he received a thorough Hebrew and Arabic education. In his youth, Ibn Ezra befriended the foremost literati of his time, especially the poet and philosopher Judah Halevi, whose creative talent he encouraged. During the Almoravid invasion of Granada in 1090 he fled to Christain Spain, where he led a life of exile, constantly pining for his native Andalusia, which he never regained. Ibn Ezra is rightly considered one of the greatest Hebrew poets of the Andalusian Golden Age. His Arabic prose works, containing his views on literary history, theory and criticism, rhetorics and philosophy are not as well known. Yet his *Kitab al-muhadara wal-mudakara (Book of Conversation and Discussion)* and his *Maqalat al-hadiqa fî ma'na l-magaz wal-haqiqa (The Treatise of Garden on Metaphorical and Literal Meaning)* rank amongst the classics of Judaeo-Arabic literature. The former is divided into eight chapters, containing a discussion of the nature of poetry and a treatise on Jewish literary history. The last and longest chapter examines the figures of thought and expression used to embellish Hebrew poetry. In addition to its didactic - and even polemical - content as a kind of Hebrew *ars poetica*, the book has recently been described as an attempt to provide biblical legitimation to the 'profane' Hebrew poetry of the Andalusian school. Its application of rhetoric to Scripture is an early step towards the modern literary criticism of the Hebrew Bible.

The *Maqalat al-hadiqa fî ma'na l-magaz wal-haqiqa (The Treatise of Garden on Metaphorical and Literal Meaning)*, although not a systematic philosophical treatise in the strict sense, contains much speculative material. Inititially conceived of as a disquisition on the biblical use of metaphorical language, the work is divided into two parts, the first of which proposes a variety of philosophical themes for the consideration of the reader. Suffused with anecdotes, poetic quotations, wise sayings and scientific observations, the work preserves important extracts of works now lost, such as the Arabic original of Ibn Gabirol's *Fons Vitae*. It adopts a genuinely humanistic perspective, expressed in part in its anthropocentrism. The book as a whole reflects both the influence of the Mu'tazilite school and the strong Neoplatonic current that entered Spain with the works of the Brethren of Purity (see Ash'ariyya and Mu'tazila; Ikhwan al-Safa'). The eleven chapters of the first part discuss the fundamental themes of Mu'tazilite theodicy, coloured with Neoplatonic insights: divine unity and unknowability, negative and apophatic theology, creation, revelation, nature, intellect and the human soul. Drawing on the Neoplatonic systems of Pseudo-Empedocles, Isaac Israeli, Dunash ben Tamim, Ibn Gabirol and, above all, the Brethren of Purity, Ibn Ezra adopts the doctrine of emanation and makes the 'active intellect' the first divine creation, an entity that by his account proceded from the divine Will. This intellect is a pure and simple substance containing the forms of all existents (see Neoplatonism). The author also adopts Ibn Gabirol's hylomorphism. On the Greco-Arabic side, besides the ancient authors, such as pseudo-Aristotle, Plato and Hippocrates, the most often quoted are the Brethren of Purity and al-Farabi, mentioned here for the first time in Jewish philosophy.

The second part takes the form of a glossary of terms designating humanity, viewed from physical and spiritual perspectives, and examines their literal and figurative use. Ibn Ezra's exegesis belongs to the Andalusian rationalistic tradition. The Hebrew version of the philosophical section of the book, translated in the Middle Ages by the celebrated poet and translator, Judah al-Harizi, was to have an abiding impact even among Kabbalists, who used it in the development of their speculative systems. It was an important link in the transmission of Neoplatonic elements to the early Spanish Kabbalah (see Kabbalah). However, the work is important more as a transmittor than as a fount of original thinking.

See also: Ikhwan al-Safa'; Neoplatonism

PAUL B. FENTON

List of works

Ibn Ezra, M. (*c*.1055-1135) *Kitab al-muhadara wal-mudakara (Book of Conversation and Discussion)*, ed. A.S. Halkin, Jerusalem: Mekize Nirdamim, 1975. (Edition of original Arabic text with a modern Hebrew translation of this pioneering medieval work on Biblical poetics.)

Ibn Ezra, M. (*c*.1055-1135) *Maqalat al-hadiqa fî ma'na l-magaz wal-haqiqa (The Treatise of Garden on Metaphorical and Literal Meaning)*, ed. P. Fenton, Jerusalem: Mekize Nirdamim, 1989.(Edition contains the original Arabic plus a modern Hebrew translation. Long extracts in French can also be found in Fenton (1997a).)

References and further reading

Dana, J. (1982) *Poetics of Mediaeval Hebrew Literature acording to Moshe Ibn Ezra*, Jerusalem-Tel Aviv: Dana, Dvir.(Analysis, in Hebrew, of the *Muhadara* in the light of Arabic rhetoric.)

Diez Macho, A. (1953) 'Mose Ibn Ezra como poeta y preceptista' (Moses ibn Ezra as a Poet), Madrid-Barcelona. (Introduction to the poetry of Ibn Ezra with a biographical essay.)

Fenton, P. (1976) 'Gleanings from Moseh Ibn 'Ezra's *Maqalat al-hadiqa*', *Sefarad* 36: 285-98.(An account of an important philosophical text.)

Fenton, P. (1997a) *Philosophie et exégèse dans le Jardin de la Métaphore de Moïse Ibn 'Ezra (Philosophy and Exegesis in Moses ibn Ezra's Garden of Metaphor)*, Leiden: Brill.(Analysis of the philosophical and exegetical material in the *Maqalat al-hadiqa*, including extensive passages in French translation, and an edition of Arabic extracts from the *Fons Vitae*.)

Fenton, P. (1997b) 'Traces of Moses Ibn 'Ezra's *'Arugat ha-bosem* in the Writings of the Early Spanish Qabbalists', in I. Twersky (ed.) *Studies in Medieval Jewish History and Literature III*, Cambridge, MA: Harvard University, Center for Jewish Studies.(His links with the Kabbalists.)

Pinès, S. (1957) 'Fragments of the Arabic Original of *Fons Vitae* in Moses Ibn 'Ezra's *'Arugat ha-bosem*', *Tarbiz* 27: 218-33.(Study, in Hebrew, of the quotations in the *Maqalat* from the lost Arabic original of Ibn Gabirol's *Fons Vitae*.)

Ibn Falaquera, Shem Tov (1223/8-after 1290)

A prolific author with a clear and precise Hebrew style, Ibn Falaquera wrote both original works and Hebrew translations of Arabic works of philosophy and science. His writings include encyclopedias, Bible commentaries, the first commentary on Maimonides' *Guide to the Perplexed and, by his own account, some twenty thousand verses of poetry. Unlike Maimonides, who wrote for the intelligentsia, Ibn Falaquera wrote most of his works with the stated aim of raising the cultural level of the Jewish people. Most of his prose works survive, many in multiple editions or manuscripts and several in European translations, a testimony to their popularity. A consistent theme in his works is the harmony of faith and reason.*

Shem Tov ben Joseph ibn Falaquera was born in Spain between 1223 and 1228; his last known work refers to events in 1290. Various etymologies have been suggested for the name Falaquera, for which we have diverse spellings in Hebrew and European languages, and which was the name of a prominent Jewish family in Tudela. Most of his prose works survive, and many can be found in multiple editions or manuscripts. More than half of his youthful poetry was lost, he tells us, and in later life he abandoned his poetic career but continued to intersperse poetry, some of it humorous, with his prose works.

We know of eighteen works by Ibn Falaquera. In probable chronological order they are:

(1) *Batei Hanhagat Guf ha-Bari, Batei Hanhagat ha-Nefesh (Verses on the Regimen of the Healthy Body and Soul)*;
(2) *Iggeret ha-Musar (Epistle on Ethics)*, a *maqama* (prose narrative interspersed with verse) replete with with Jewish and Arabic ethical maxims, recounting the adventures of a youth in search of wisdom;
(3) *Zeri ha-Yagon (Balm for Sorrow)*, another *maqama*, containing rabbinic and philosophic consolations;
(4) *Megillat ha-Zikkaron (The Scroll of Remembrance)*, a lost work, presumably chronicling Jewish sufferings;
(5) *Iggeret ha-Vikuah (Epistle of the Debate)*, a debate between a pious traditionalist Jew and a philosopher, demonstrating that philosophy is the twin sister of Torah and fully harmonious with Judaism;
(6) *Reshit Hokhmah (The Beginning of Wisdom)*, an introduction to the sciences, including much material based on the works of al-Farabi;
(7) *Sefer ha-Ma'alot (Book of Degrees)*, an ethical work on the corporeal, spiritual and divine degrees of human perfection (*ma'alot* (degrees) is the Hebrew term for the virtues);
(8) *Sefer ha-Mevaqqesh (Book of the Seeker)*, composed in 1263, when Ibn Falaquera says he was past 35 and approaching 40, a *maqama* expanding on the theme of the youthful seeker after wisdom, surveying the arts, professions and sciences, culminating in philosophy;
(9) *De'ot ha-Philosofim (The Opinions of the Philosophers)*, a voluminous encyclopedia of the sciences intended to propagate philosophy among Jews and quoting extensively from the Arabic sources;
(10) *Sefer ha-Nefesh (Book of the Soul)*, the first systematic Hebrew work of psychology;
(11) *Shelemut ha-Ma'asim (The Perfection of Actions)*, based on the *Nicomachean Ethics* and Arabic ethical literature;
(12) *Iggeret ha-Halom (Treatise of the Dream)*, on physical and spiritual well-being;
(13) *Sefer ha-Derash (The Book of Interpretation)*, no longer extant, probably a commentary on *aggadic* passages in the Talmud and midrash;
(14) *Perush (Bible Commentary)*; only fragments (often philosophical in orientation) survive, in citations by later authors;
(15) *Moreh ha-Moreh (Guide to the Guide)*, completed in 1280, the first commentary on Maimonides' *Guide to the Perplexed*, including original Hebrew translations of the passages discussed, and comparing Maimonides (sometimes critically) to numerous classical, Arabic and Jewish philosophers;
(16) *Liqqutim mi-Sefer Meqor Hayyim*, selections from the *Fons Vitae* in Hebrew translation, eliminating the dialogue format of Ibn Gabirol's original (see Ibn Gabirol);
(17) *Liqqutim mi-Sefer ha-'Azamim ha-Hamishah (Selections from the Book of the Five Substances)*, Hebrew translation of passages from a Pseudo-Empedoclean work;
(18) *Mikhtav 'al Devar ha-Moreh (Letter on the Matter of the Guide)*, Ibn Falaquera's last known work, a defence of Maimonides' *Guide* against anti-philosophical critics.

A consistent theme in Ibn Falaquera’s works is the harmony of faith and reason. Philosophy and Torah, when both are properly understood, are ‘sisters’ and ‘twins’. The rabbinic saying, ‘Rabbi Meir found a pomegranate; he ate what was within and discarded the peel’ (Hagigah 15b), means that one must accept what is true and concordant with Torah in philosophy. Reason can verify religious truth, and faith perfects reason.

To reject philosophy because some philosophers have erred is like denying water to a person dying of thirst, just because some people have drowned. A Jew must learn the truth from any source, as one takes honey from a bee. For ‘all nations share in the sciences; they are not peculiar to one people’ (*Sefer ha-Ma‘alot*). Thus: ‘Accept the truth from whoever utters it; look at the content, not the speaker’ (*Sefer ha-Ma‘alot*). Since true human perfection is intellectual, dissemination of philosophy in Hebrew and rebuttal of its detractors serve a religious as well as a cultural need.

Ibn Falaquera’s rationalism is manifest throughout his works, including his Biblical exegesis and his specific theses. He equates his Platonizing doctrine of creation with the account of Genesis, and he reads his intellectualism into Biblical ethics to derive an extreme asceticism which does not stop short of misogyny: ethical and social commitments are ‘impediments’ to true, intellectual fulfillment. Ibn Falaquera combines his deep concern for the education of his people with the belief that genuine felicity is attained by the individual who is ‘solitary’ (*mitboded*), isolated not physically but spiritually from the external distractions of society and the internal promptings of the appetites. Knowledge of God begins with self-knowledge, knowledge of one’s own soul: ‘Know your soul, O man, and you will know your Creator’. Psychology is thus prior to all other sciences: ‘Knowledge of the soul is prior to the knowledge of God, and… is the most excellent form of knowledge after the knowledge of God’ (*Sefer ha-Nefesh*; *De‘ot ha-Philosofim* VI:A:1).

Ibn Falaquera was not an original thinker of the first order, but the breadth and depth of his knowledge of Judaism, philosophy and science make him an important figure in the history of Jewish philosophy. The pioneering philosophical efforts of earlier luminaries attained an enduring impact through their consolidation and popularization by philosophers like Ibn Falaquera, whose contribution is no less important for the fact that their light was often a reflected one.

See also: Maimonides, M.

RAPHAEL JOSPE

List of works

Most of Falaquera’s works exist in nineteenth or twentieth century Hebrew editions; some, accompanied by introductions or translations into European languages. See Jospe 1988: 485-97 for a more detailed bibliography.

Ibn Falaquera, Shem Tov (1223/8-after 1290) *Sefer ha-Ma‘alot (Book of Degrees)*, ed. L. Venetianer, Berlin, 1894; photo repr. in *Kitvei Rabbi Shem Tov Falaquera*, Jerusalem: Maqor, 1970.(An ethical work on the degrees of human perfection.)

Ibn Falaquera, Shem Tov (1223/8-after 1290) *Sefer ha-Nefesh (Book of the Soul)*, Warsaw, 1924; photo repr. in *Kitvei Rabbi Shem Tov Falaquera*, Jerusalem: Maqor, 1970; ed. and trans. R. Jospe in *Torah and Sophia: The Life and Thought of Shem Tov ibn Falaquera*, Cincinnati, OH: Hebrew Union College Press.(The first systematic Hebrew work on psychology.)

Ibn Falaquera, Shem Tov (1223/8-after 1290) *Shelemut ha-Ma‘asim (The Perfection of Actions)*, ed. R. Jospe in *Torah and Sophia: The Life and Thought of Shem Tov ibn Falaquera*, Cincinnati, OH: Hebrew Union College Press.(Ethical work based on the *Nicomachean Ethics* and Arabic ethical literature.)

Ibn Falaquera, Shem Tov (1223/8-after 1290) *Zeri ha-Yagon (Balm for Sorrow)*, trans. R.K. Barkan, ‘Shem Tob ben Joseph ibn Falaquera’s *Sori Yagon* or "Balm for Assuaging Grief", Its Literary Sources and Traditions’, Ph.D. thesis, Columbia University, 1971; repr. Ann Arbor, MI: University Microfilms, 1981.(A valuable study of the sources in classical and Arabic consolation literature.)

Ibn Falaquera, Shem Tov (1223/8-after 1290) *Iggeret ha-Vikuah (Epistle of the Debate)*, ed. A. Jellinek, Vienna, 1875; photo repr. in *Kitvei Rabbi Shem Tov Falaquera*, Jerusalem: Maqor, 1970; ed. G. Dahan, ‘Epistola Dialogi: Une Traduction Latine de L’*Igeret Ha-Vikuah* de Shemtov ibn Falaquera. Étude et Edition’, *Sefarad* 39 (1979): 1-112; trans. S. Harvey, *Falaquera’s Epistle of the Debate: An Introduction to Jewish Philosophy*, Cambridge, MA: Harvard University Press, 1987.(Dahan is a study of the Latin version of this popular work. Harvey is a clear annotated English translation with valuable appendices.)

Ibn Falaquera, Shem Tov (1223/8-after 1290) *Reshit Hokhmah (The Beginning of Wisdom)*, ed. M. David,

Berlin, 1902; photo repr. in *Kitvei Rabbi Shem Tov Falaquera*, Jerusalem: Maqor, 1970; trans. I. Efros in ‘Palquera’s *Reshit Hokmah* and Alfarabi’s *Ihsa al‘Ulum*’, *Jewish Quarterly Review*, new series 25, 1934-5: 227-35.(An important annotated translation of this major work, citing Arabic sources.)

Ibn Falaquera, Shem Tov (1223/8-after 1290) *Perush (Bible Commentary)*, fragments trans. R. Jospe and D. Schwartz, ‘Shem Tov ibn Falaquera’s *Lost Bible Commentary*’, *Hebrew Union College Annual* 64, 1993: 167-200.(New manuscript material with English translation and notes.)

Ibn Falaquera, Shem Tov (1223/8-after 1290) *Iggeret ha-Halom (Treatise of the Dream)*, ed. H. Malter, ‘Shem Tob ben Joseph Palquera: His Treatise of the Dream’, *Jewish Quarterly Review*, new series 1, 1910-11: 451-501; photo repr. in *Kitvei Rabbi Shem Tov Falaquera*, Jerusalem: Maqor, 1970.(A useful edition of this work.)

Ibn Falaquera, Shem Tov (1223/8-after 1290) *De‘ot ha-Philosofim (The Opinions of the Philosophers)*, introduction ed. M. Zonta, ‘Un Dizionario Filosofico Ebraico del XIII Secolo: L’introduzione al ‘Sefer De‘ot ha-Filosofim’ di Shem Tob ibn Falaquera’, Turin: Silvio Zamorani Editore, 1992.(A carefully annotated presentation of the Introduction to Falaquera’s unpublished philosophic encyclopedia.)

Ibn Falaquera, Shem Tov (1263) *Sefer ha-Mevaqqesh (Book of the Seeker)*, ed. M. Tamah, The Hague, 1778; photo repr. in *Kitvei Rabbi Shem Tov Falaquera*, Jerusalem: Maqor, 1970; trans. M.H. Levine, *Falaquera’s Book of the Seeker*, New York: Yeshiva University Press, 1976.(English translation, revised from the author’s 1954 Columbia University Ph.D. thesis, which is still worth consulting in its own right (published Ann Arbor, MI: University Microfilms, 1971).)

Ibn Falaquera, Shem Tov (1280) *Moreh ha-Moreh (Guide to the Guide)*, ed. Mordecai Leib Bisliches, Pressburg, 1837; photo repr. in *Sheloshah Qadmonei Mefarshei ha-Moreh*, Jerusalem, 1961; ed. Y. Shiffman, ‘Rabbi Shem Tov ben Joseph Falaquera’s *More ha-More*: A Philosophical and Philological Analysis’, unpublished Ph.D. thesis, Hebrew University of Jerusalem, 1990.(Shiffman is the definitive, annotated critical edition of Ibn Falaquera’s Commentary on Maimonides’ *Guide to the Perplexed*, focusing on his Arabic sources.)

References and further reading

Harvey, S. (1992) ‘Falaquera’s Epistle of the Debate and the Maimonidean Controversy of the 1230s’, in R. Link-Salinger (ed.) *Torah and Wisdom: Studies in Jewish Philosophy, Kabbalah, and Halacha: Essays in Honor of Arthur Hyman*, New York: Shengold, 75-86.(A helpful study of Ibn Falaquera in the context of the controversy over philosophy in the century after Maimonides.)

Jospe, R. (1986) ‘Rejecting Moral Virtue as the Ultimate Human End’, in W.M. Brinner and S.D. Ricks (eds) *Studies in Islamic and Judaic Traditions*, Atlanta, GA: Scholars Press, 185-204.(A study of Ibn Falaquera’s ethical theory, including his tendency towards asecticism, in light of tensions within Maimonides’ thought.)

Jospe, R. (1988) *Torah and Sophia: The Life and Thought of Shem Tov ibn Falaquera*, Cincinnati, OH: Hebrew Union College Press.(The most complete study to date of Ibn Falaquera’s life and philosophy, including a critical edition and annotated translation of *Sefer ha-Nefesh* and a critical edition of *Shelemut ha-Ma‘asim*.)

Malter, H. (1910-11) ‘Shem Tob ben Joseph Palquera: A Thinker of the Thirteenth Century’, *Jewish Quarterly Review*, new series 1: 151-81.(A still useful early study of Ibn Falaquera.)

Melamed, A. (forthcoming) *The Philosopher King in Medieval and Renaissance Jewish Political Thought*, Atlanta, GA: Scholars Press.(Includes an important study of Ibn Falaquera in the context of classical, Arabic and Western political philosophy.)

Munk, S. (1857) *Mélanges de Philosophie Juive et Arabe*, Paris; repr. Paris: Vrin, 1927.(A pioneering study of Ibn Falaquera on the basis of which Munk identified Solomon ibn Gabirol as the Avicebrol of the Latin *Fons Vitae*, until then known only as a translation from the Arabic.)

Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Paris: Éditions de la Maison des Sciences de l’Homme; Cambridge: Cambridge University Press, 234-8. (Includes much previously unpublished manuscript material and recent scholarship.)

Sitskin, L. (1979) *Eight Jewish Philosophers in the Tradition of Personalism*, New York: Feldheim, 134-40. (Treats Ibn Falaquera as a significant figure in the history of Jewish philosophy.)

Zonta, M. (1995) *Un Interprete Ebreo della Filosofia di Galeno: Gli Scritti Filosofici di Galeno Nell’opera di Shem Tob ibn Falaquera (A Hebrew Interpretation of Galen’s Philosophy According to the Works of Shem Tov ibn Falaquera)*, Turin: Silvio Zamorani.(A careful study of Ibn Falaquera in the medieval tradition of Galen’s philosophy.)

Ibn Gabirol, Solomon (1021/2-57/8)

Ibn Gabirol was an outstanding exemplar of the Judaeo-Arabic symbiosis of medieval Muslim Spain, a poet as well as the author of prose works in both Hebrew and Arabic. His philosophical masterwork, the Mekor Hayyim (Fountain of Life), was well known to the Latin scholastics in its twelfth century Latin translation, the Fons Vitae. The work presents a Neoplatonic conception of reality, with a creator God at the apex. The universal hylomorphism that pervades the created order, both spiritual and corporeal, has divine will as the intermediary between God and creation, allowing Ibn Gabirol to avoid the rigidly determinist emanationism of his Greek predecessors. The Fons Vitae challenged such philosophers as Thomas Aquinas and Duns Scotus to critical reflections regarding individuation and personal immortality.

1 Life and writings

Solomon ben Judah Ibn Gabirol was born in Malaga, Spain, and reared in Saragossa. He spent his brief life under the enlightened reign of the Umayyad caliphate in Saragossa, Granada and finally Valencia, where he died between 1054 and 1070, most likely in 1057/8. Within a century of his death less tolerant regimes would hold sway, but for the moment Jews took an active part in the heady intellectual and literary life of Andalusia. Ibn Gabirol's poetry addresses both religious and secular themes. His *Keter Malkhut (Kingly Crown)* is a paean to God and creation. Its vivid expression of the poet's philosophical cosmology continues in use in some traditions of Jewish liturgy. Two philosophical prose works survive, both written originally in Arabic: the *Mekor Hayyim (Fountain of Life)*, translated into Latin as *Fons Vitae*, and the *Tikkun Middot ha-Nefesh (On the Improvement of Moral Qualities)*. The latter, written in 1045, survives in a fourteenth century Judaeo-Arabic manuscript; the former is represented in its original form only fragmentarily, but the twelfth-century Latin version is extant and was influential in Christian circles.

2 Philosophical thought

Ibn Gabirol's Neoplatonism is influenced by the *Liber de causis*, a ninth-century Arabic composition based on Proclus' *Elements of Theology* (see Liber de causis). More proximately, his work is influenced by the writings of Isaac Israeli and the pseudo-Empedoclean *Book of the Five Substances*. His philosophical poetry is expressive of Neoplatonic themes such as the soul/body dualism that undergirds what might be called Ibn Gabirol's rational mysticism: corporeal existence is devalued, in so far as it is deemed to block the path to knowledge and true felicity.

The *Tikkun Middot ha-Nefesh* comprises five parts, each of four chapters. It traces the development of some twenty moral traits to roots in specific physical senses. Pride and humility are assigned to sight; love and hate, to hearing; joy and sorrow, to taste; anger and jealousy, to smell; liberality and niggardliness, to touch. The connections are hardly perspicuous, and Ibn Gabirol does not explain them thoroughly except by introducing biblical passages suggestive of his theme. Yet it is clear that for him the body as well as the soul plays a role in the formation of virtue and thus in the attainment of happiness, which depends on achieving the proper balance among the bodily humours.

Ibn Gabirol's interest in the duality of body and soul is equally marked in the *Mekor Hayyim*, with its central polarity of matter and form. Written as a dialogue between master and pupil, the five books or treatises of the work propose a Neoplatonic cosmology and ontology along with a complementary epistemology that makes knowledge of the external world (and thus natural science) not merely dependent on self-knowledge but in fact a recapitulation of our self-awareness. As in Plotinus (*Enneads* 1.6: 8-9), the macrocosm is contained within the microcosm.

Ibn Gabirol's ontology is triadic: God, the divine will and the hylomorphic substances are its prime components. All created substances, whether spiritual or corporeal, are conjuncts of matter and form. Spiritual substances are individuated by intelligible matter; and corporeal substances, by virtue of their place in the hierarchy of being, are composed of a plurality of forms. Both doctrines, that of intelligible matter and that of the plurality of forms, occasioned controversy among the Christian scholastics. Intelligible matter, upheld by some to account for the individuation of incorporeal entities and to differentiate such entities from God, seemed to others an incoherent

notion that collapsed the very duality upon which Neoplatonic metaphysics is based. The plurality of forms, required by the emanationist programme and intended to account for the essence of corporeal entities, seemed to endanger the *unity* of the individual. A question arose whether living beings are composed of one soul or many; and if the latter, as Ibn Gabirol held, it was asked, how can such a composite being can be a unitary entity?

Just as the Greek Neoplatonists link the One to nature via Mind (see Neoplatonism), Ibn Gabirol links God to the world through a series of emanations, via the divine will. The relation of this will to God, the so-called first essence, remains somewhat problematic. For this will is the creative power of God, and as such is *in* God and inseparable from his being; yet it is also the first emanation from God and so would seem to be not an attribute but a separate hypostasis. In either case, however, the divine will is the cause of all created substances, both spiritual and corporeal.

Ibn Gabirol's reliance on the creative role of the divine will and thus on the voluntariness of creation contrasts dramatically with the tactics of his Greek philosophical predecessors, whose emanationism seemed excessively deterministic to philosophers of a monotheistic stamp. Ibn Gabirol and many of his Jewish successors exalt the voluntariness of the creative act, divine or human. They see the world not as a sequence of predetermined steps or a 'timeless flowing forth of necessity', as a recent commentator has put it (Goodman 1992), but as the free act of God. Ibn Gabirol's postulation of a will that created the world is a Jewish Neoplatonist's way of safeguarding the divine from the necessitarianism of Neoplatonic emanationism. One corollary of this rejection of reductionism is the corresponding safeguarding of human freedom. If creation itself is an act of divine free will, human freedom need not be reduced to the mere product of a rigidly deterministic causal sequence. The freedom of the microcosm recapitulates that of the macrocosm.

Within the created order, matter and form are irreducible to one another (see Matter). They emanate ultimately from different sources and are present at all levels of creation. Indeed, Ibn Gabirol (like Proclus) gives a striking and un-Aristotelian priority to matter, whose ultimate source is God, the first essence. Form derives from the divine will; its source is thus one rung removed from the deity. There is a fine irony here, in giving matter, the almost canonical source of evil, an immediate divine origin. However, the priority and divine origin of matter are driven by the assumption that matter is a genus made determinate by the imposition of successive differentiae.

3 Influence

Ibn Gabirol's philosophical work was destined to have comparatively little direct influence on later Jewish thinkers. Within a century of his death, Aristotelian trends interpenetrated with, altered and displaced Neoplatonic ones. Ibn Gabirol's philosophical influence waned, although it is strikingly visible among some later Jewish Neoplatonists and in Kabbalistic circles. His influence on Christian thinkers, however, was great. Translated into Latin by John of Spain (who may be the Jewish writer Ibn Daud) in collaboration with Dominicus Gundissalinus, the *Fons Vitae* was read and criticized by the Dominican Aristotelians Albertus Magnus (see Albert the Great) and Thomas Aquinas, who rejected outright its doctrine of intelligible matter on the grounds noted above. Franciscans such as Bonaventure and Duns Scotus were more sympathetic, and Ibn Gabirol's hylomorphism seems to have fed the classic scholastic debates on identity and individuation, specifically over whether personal immortality presupposes intelligible matter, or whether form by itself can individuate particulars.

Ibn Gabirol's influence, however, was all but anonymous. His text bore no biblical or rabbinic citations as a sign of its Jewish authorship, and it came down under the (Arabicized) name of Avicebrol (or Avicebron or Avencebrol), a corruption of Ibn Gabirol. The identity of its author with the well-known Hebrew poet was rediscovered only in the nineteenth century by the French scholar Salomon Munk, who recognized fragments quoted from the work in the writings of Shem Tov Ibn Falaquera. Perhaps by historical accident, Ibn Gabirol's philosophy becomes a paradigm of philosophical ecumenism.

See also: Neoplatonism; Platonism, Medieval

DANIEL H. FRANK

List of works

Ibn Gabirol, S. (1045) *Tikkun Middot ha-Nefesh (On the Improvement of Moral Qualities)*, ed. and trans. S. Wise, *The Improvement of the Moral Qualities*, New York: AMS Press, 1966.(Hylomorphism in the context of moral

philosophy.)
Ibn Gabirol, S. (before 1057/8?) *Mekor Hayyim (Fountain of Life)*, Latin trans. *Fons Vitae*, ed. C. Baeumker, *Beiträge zur Geschichte der Philosophie des Mittelalters* 1, parts 2-4, Münster: Aschendorff, 1892-5.(The standard edition of the surviving Latin translation of Ibn Gabirol's philosophical masterwork.)
Ibn Gabirol, S. (before 1057/8?) *Keter Malkhut (The Kingly Crown)*, trans. I. Zangwill in I. Davidson (ed.) *Selected Religious Poems of Solomon ibn Gabirol*, New York: Arno Press, 1973; trans. R. Loewe in R. Loewe, *Ibn Gabirol*, New York: Grove Weidenfeld, 1990.(A prayer to God, creator of the universe, and a celebration of divine creation.)

References and further reading

Brunner, F. (1965) *Platonisme et Aristotelisme: La Critique d'Ibn Gabirol par St. Thomas d'Aquin (Platonism and Aristotelianism: the Critique of Ibn Gabirol by St Thomas Aquinas)*, Louvain: Publications universitaires de Louvain.(Details Aquinas' virulent critique of Ibn Gabirol's Neoplatonism.)
Goodman, L.E. (ed.) (1992) *Neoplatonism and Jewish Thought*, Albany, NY: State University of New York Press.(See especially Goodman's introduction and the papers by J. Dillon, B. McGinn and A. Hyman.)
Guttmann, J. (1933) *Philosophies of Judaism*, New York: Schocken Books, 2nd edn, 1973, 101-17.(A good brief account of the *Fons Vitae* in a standard history of Jewish philosophy.)
Husik, I. (1916) *A History of Medieval Jewish Philosophy*, New York: Athenaeum, 2nd edn, 1976, 59-79.(A good brief account of both the *Fons Vitae* and the *Tikkun Middot ha-Nefesh*.)
Rudavsky, T. (1978) 'Conflicting Motifs: Ibn Gabirol on Matter and Evil', *New Scholasticism* 52: 54-71.(An important recent article detailing Ibn Gabirol's disparate, perhaps inconsistent views on matter and its relation to evil.)
Schlanger, J. (1968) *La Philosophie de Salomon ibn Gabirol*, Leiden: Brill.(The most important recent and comprehensive study.)
Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Cambridge: Cambridge University Press, 68-81.(A short account that focuses on the metaphysics of both the *Fons Vitae* and the philosophical poem *Keter Malkhut*.)

Ibn Hazm, Abu Muhammad ‘Ali (994-1063)

Ibn Hazm was the originator of a school of interpretation which based its understanding of religious texts on the apparent meaning of scriptural concepts as opposed to their hidden meaning. He argued that there is a place for reason in the understanding of scripture, but that it has to be used within the context of revelation and is severely limited in terms of what it can demonstrate. His approach is based on the idea that the language and context of religious texts are sufficient for their readers to understand them, and that there is no need to use concepts such as analogy.

Abu Muhammad ‘Ali ibn Hazm was born into an important Andalusian family and went on to have a rather tumultuous political career, being imprisoned three times and banished from Cordoba on several occasions. He is best known for his writings on jurisprudence, and also for his charming *Tawq al-hamama (The Dove’s Neck Ring)*, which deals with the concept of love. In it he analyses the concept and differentiates between divine love, which is placed at the highest level, and affection, which is the lowest. Clearly influenced by Plato’s *Phaedrus* and *Symposium*, he regards love as the coming together of otherwise incomplete beings (see Plato). Genuine love occurs when the lover sees beneath the surface of the appearance something which presents an idea of his own nature, and thus becomes strongly attracted by it. Weaker forms of affection result when the individual is limited to the form of the appearance, but perceives nothing deeper beyond it (see Love).

Ibn Hazm is the leading exponent of the Zahirite school of jurisprudence. This school adheres to the exoteric or apparent (*zahir*) meaning of the religious text, in accordance with the principles of grammar, the *hadith* (traditions) of the Prophet and the consensus (*ijma‘*) of the community. The main opponents of this view are those Muslims who appeal to the esoteric meaning - such as the Isma‘ilis, for example - because they think that one needs to look beyond the surface of the text to discover what it really means, and also the philosophers, who insist that reason is a vital means of gaining access to the meaning of scripture. This is more than a dispute about jurisprudence; it affects the understanding of the way in which texts are to be interpreted. Ibn Hazm attacks the notion that one can understand the meaning of a text by using principles such as analogy, as the Mu‘tazilites do, to acquire some grasp of the nature of God. The latter argue that we can understand the *sifat*, the names or qualities of God, by analogy from our understanding of our own characteristics; so, for example, we can grasp what it means for God to be just if we understand what human justice is (see Ash‘ariyya and Mu‘tazila). According to Ibn Hazm, however, God is a unique being whose qualities cannot be grasped rationally but have to be accepted through faith. There is no objective standard of justice with which God has to concur. He could have obliged us to act in impossible ways, and set out to punish angels and reward the evil. To deny this is to anthropomorphize the concept of God, which is going beyond the nature of the language in Islam about God. The Zahirite uses reason to get an accurate view of the language of the relevant texts and the supplementary hermeneutical material, and stops there (see God, concepts of; Islamic theology).

What role does reason play, then, in Ibn Hazm’s understanding of the meaning of important concepts? Reason is important, and essential in any understanding of the facts. We have to use reason to work out what the facts of a particular situation are, but we cannot use it to identify its ethical or religious character. Within the context of revelation, reason has a role to play, but it is an essentially subsidiary role. In comprehending religious language we have to use reason to interpret the text, but we must be aware of the dangers of overelaborating and departing from the apparent meaning. Sometimes people are impressed with the ability of reason to delve into the mysteries of reality, yet all that emerges are disputes about what texts mean and what the nature of the law is. We can avoid this, according to Ibn Hazm, if we stick to the apparent meaning of the text and maintain the autonomy of God. God can do anything at all, he is absolutely free, and we are very limited in our ability to use reason to encompass him. We can use the laws of logic, Arabic grammar and the evidence of our senses, but that is all.

By the time of his death, Ibn Hazm had succeeded in establishing the Zahiri school of interpretation, which followed his particular approach to hermeneutics and which was solidly within the Asha‘rite and Sunni tradition. The most distinguished follower of this form of thought was al-Ghazali, who was clearly heavily influenced by Ibn Hazm.

See also: Ash‘ariyya and Mu‘tazila; God, concepts of; Islamic theology; Law, Islamic philosophy of

OLIVER LEAMAN
SALMAN ALBDOUR

List of works

Ibn Hazm (994-1063) *Tawq al-hamama (The Dove’s Neck Ring)*, trans. A. Arberry, *The Ring of the Dove*, London: Luzac, 1953.(A systematic treatment of love and affection, combining metaphysics, social commentary and psychology.)

Ibn Hazm (994-1063) *al-Fisal fi’l milal wa’l ahwa’ wa’n nihal (Treatise on Religions and Schools of Thought)*, Cairo: Maktabat al-Khanji, no date; trans. M. Asín Palacios, *Abenházam de Córdoba y su historia crítica de las ideas religiosas*, Madrid: Real Academica de la Historia, 1927-32.(A treatise on the religions and beliefs which presents the first comparative work of its kind in Arabic.)

Ibn Hazm (994-1063) *Mudawat an-nufus (Character and Behaviour)*, ed. and trans. N. Tomiche, Beirut, 1961; trans. M. Asín Palacios, *Los caracteres y la conducta, tratado de moral práctica por Abenházam de Córdoba*, Madrid: Junta para ampliación de estudios y investigaciones científicas, 1916. (Ibn Hazm’s philosophical psychology, and an account of ethical development, with many interesting comments on contemporary Andalusian society.)

References and further reading

Arnaldez, R. (1984) *Grammaire et théologie chez Ibn Hazm de Cordoue: essai sur la structure et les conditions de la pensée musulmane (The Grammar and Theology of Ibn Hazm of Cordoba: Essay on the Nature and Structure of Muslim Thought)*, Paris: Vrin.(A brilliant analysis of Ibn Hazm’s philosophy of language as the key to his thought.)

Chejne, A. (1982) *Ibn Hazm*, Chicago: Kazi Publications.(A general account of his life and thought.)

Goldziher, I. (1884) *Die Zahiriten, ihr Lehrsystem und ihre Geschichte: Beitrag zur Geschichte der muhammedanischen Theologie (The Zahirites, their School and its History: Contribution to the History of Islamic Theology)*, Leipzig: O. Schulze.(Still a very useful source of information about the school of theology which Ibn Hazm founded.)

Hourani, G. (1985) ‘Reason and Revelation in Ibn Hazm’s Ethical Thought’, in G. Hourani (ed.) *Reason and Tradition in Islamic Ethics*, Cambridge: Cambridge University Press, 167-89.(Detailed exposition of the philosophical basis of Ibn Hazm’s ethics.)

Pavlin, J. (1996) ‘Sunni *Kalam* and Theological Controversies’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy,* London: Routledge, ch. 7, 105-18.(Account of some of the most important Sunni theologians, including Ibn Hazm.)

Ibn Kammuna (d. 1284)

Physician and man of letters, Ibn Kammuna left a number of writings on philosophy and religion. His treatise comparing Judaism, Christianity and Islam caused major rioting in Baghdad, forcing him to flee that city in secret. His commentary on al-Suhrawardi's Talwihat, the major text of Islamic Illuminationist philosophy remains one of the clearest and most thorough expositions of that branch of thought.

1 Comparative religion

Of the major writings of 'Izz al-Dawla Sa'd bin Mansur ibn Kammuna, only the two that compare the views of religious communities have been published thus far. The longer one, the *Tanqih al-abhath fi akhbar al-milal al-thalath (An Overview of Investigations into the Views of the Three Faiths)* is *sui generis* in medieval literature. It begins with a extended investigation of prophecy, aiming to establish in a manner acceptable to adherents of all faiths (not just the prophetic ones) that revelation does occur. Here as elsewhere, Ibn Kammuna combines ideas culled from highly diverse sources including Moses Maimonides, Judah Halevi, al-Ghazali and Fakhr al-Din al-Razi. The next next three sections examine the most important prophetic traditions: Judaism, Christianity and Islam. Each section opens with an exposition of the basic tenets of the faith concerning its revealed Scripture, followed by a series of queries or objections, and the answers that a defender of that faith may be expected to give.

The presentation is dispassionate and eschews any polemical tone. Ibn Kammuna perhaps hoped that philosophical commitments shared by readers from the different faiths, and the unique political situation - the rulers were pagan Mongols bearing allegiance to none of the three faiths - would allow a calm examination of the scriptures sacred to each. The rioting against his book was perhaps incited by religious leaders, but it is not hard to understand why both Christians and Muslims were moved to write refutations of the book. It gives considerably more space to criticism of the sacred scriptures of the two 'daughter religions' than to the Torah, and it dwells on sectarian and other internal differences in Christianity and Islam but not in Judaism - although elsewhere Ibn Kammuna wrote an entire treatise on the differences between Rabbanites and Karaites.

It is not clear whether Ibn Kammuna attempted to formulate positions reflecting an actual consensus within each community (which would lend even greater historical interest to his text), or whether he forged his own synthesis from a melange of doctrines taken from existing literature. His fusion of the Halevi's chronicle of divine revelation from Adam onwards with Maimonides' rationalistic explanation of the ancient Israelite temple cult is not attested from other sources. It suggests that the book displays an original concatenation rather than an empirical study of views then normative in the three communities.

The treatise on the two major trends within Judaism in his day is considerably shorter. It devotes about twice as much space to Rabbanite views as to the Karaites, leading Leon Nemoy to conclude that, 'dispassionate critic though he was, his sympathies nevertheless remained with the mother synagogue' (Nemoy 1968: 109).

2 The Ishraqi tradition

Ibn Kammuna's commentary on Ibn Sina's *al-Isharat wa 'l-tanbihat* is essentially a paraphrase, reflecting in a few places a somewhat different structure than the published text of Ibn Sina. Among its distinctive features are its division of the sciences, particularly the characterization of mathematical sciences as 'proto-physical' (*ma qabla al-tabiy'a*), balancing the accepted term for metaphysics, as 'what comes after the physics'. Ibn Kammuna offers his own interpretation of the story of Ibn Sina's Salman and Absal (the former symbolizes the rational soul, the latter the speculative intellect) and offers other unusual insights into the gnosis which Ibn Sina sketches in the third part of the book.

Ibn Kammuna's commentary on al-Suhrawardi's *Talwihat* is his longest work and, to judge from the number of surviving manuscripts, his most widely read. It seems to have played no small role in the development the 'illuminationist' (*ishraqi*) school of philosophy. Corbin noted the value of its exposition as well as its critical importance for establishing the text of the *Talwihat*. Nonetheless, about all that is available in print of this work are some brief quotations in Corbin's study and translations of selected passages by Shlomo Pines in footnotes to his studies of Abu 'l Barakat al-Baghdadi. The commentary is encyclopaedic, its discussions much fuller than those in the commentary on Avicenna. It promises to be a rich source for the exposition of Ishraqi philosophy and

for a broad range of topics debated by thinkers of the period. The legitimacy of the so-called fourth figure of the syllogism, the possibility of alchemical transmutation of metals, and the nature of time are among the topics on which Ibn Kammuna offers lengthy disquisitions.

3 Impact

We know from manuscript catalogues that Ibn Kammuna wrote a number of other treatises in addition to those discussed above, but the only one published thus far is a short work on the immortality of the soul, a typically thorough survey. His impact on Jewish thought was minimal at best. Only his treatises on comparative religion and al-Suhrawardi were transcribed into Hebrew characters, none was widely diffused, and no citations of his writings have been found in the work of any other Jewish thinker. His works seem to have addressed the general, rather than the Jewish, public; the most influential of them, as noted, being his commentary on al-Suhrawardi.

See also: Illuminationist philosophy; al-Suhrawardi

Y. TZVI LANGERMANN

List of works

Ibn Kammuna (before 1284) *The Arabic Treatise on the Immortality of the Soul by Sa'd ibn Mansur ibn Kammuna, facsimile reproduction of the only known manuscript… with a bibliographical note by Leon Nemoy*, ed. L. Nemoy, New Haven, CT: Yale University Library, 1944; trans. L. Nemoy, 'Ibn Kammuna's Treatise on the Immortality of the Soul', in S. Löwinger, A. Scheiber and J. Somogyi (eds) *Ignace Goldziher Memorial Volume*, Jerusalem: Rubin Mass, 1958, part II, 83-99.(The original Arabic work had no title: 'Treatise on the Immortality of the Soul' is the title given to the work by Nimoy.)

Ibn Kammuna (before 1284) *Tanqih al-abhath fi akhbar al-milal al-thalath (An Overview of Investigations into the Views of the Three Faiths)*, ed. M. Perlman, *Ibn Kammuna's Examination of the Inquiries into the Three Faiths*, Berkeley, CA: University of California, 1967; trans. M. Perlman, *Ibn Kammuna's Examination of the Three Faiths*, Berkeley, CA: University of California, 1971.(The 1971 translation contains Arabic text and English translation, both with a short introduction, and brief notes.)

Ibn Kammuna (before 1284) Commentary on Ibn Sina's *al-Isharat wa 'l-tanbihat*.(No studies of this work have been undertaken and there are at present no editions; it exists as a manuscript in the British Library, I.O. Loth 484.)

Ibn Kammuna (before 1284) 'Ibn Kammunah's Treatise on the Differences between the Rabbanites and the Karaites', ed. L. Nemoy, *Proceedings of the American Academy for Jewish Research* 36, 1968: 107-65.(Text, transcribed into Arabic characters, translation, and introduction. The original work had no title; 'Treatise on Differences' is the title given to the work by Nemoy.)

References and further reading

Corbin, H. (1945) *Şihabaddin Yahya as-Suhrawardi Opera Metaphysica et Mystica (Metaphysical and Mystical Works of al-Suhrawardi)*, Istanbul: Maarif Matbaasi, vol. 1, pp. lxii-lxxiii.(Extensive discussion of the commentary on *al-Talwihat*, especially with regard to the establishment of the text.)

Baneth, D.H. (1925) 'Ibn Kammuna', *Monatsschrift für Geschichte und Wissenschaft des Judenthums* 69: 295-311.(Thorough refutation of Steinschneider's contention that Ibn Kammuna converted to Islam; contains many additional insights.)

Nemoy, L. (1968) 'New Data for the Biography of Sa'd ibn Kammunah', *Revue des études juives* 123: 507-10. (Some biographical facts, with additional confimation that Ibn Kammuna never abandoned Judaism.)

Steinschneider, M. (1902) *Die arabische Literatur der Juden (The Arabic Literature of the Jews)*, Frankfurt: J. Kaufman, 239-40.(List of works and manuscripts.)

Ibn Khaldun, ‘Abd al-Rahman (1332-1406)

Ibn Khaldun’s work on the philosophy of history is a landmark of social thought. Many historians - Greek, Roman, Muslim and other - had written valuable historiography, but here we have brilliant reflections on the meaning, pattern and laws of history and society, as well as profound insights into the nature of social processes and the interconnections between phenomena in such diverse fields as politics, economics, sociology and education. By any reckoning, Ibn Khaldun was the outstanding figure in the social sciences between Aristotle and Machiavelli, and one of the greatest philosophers of history of all time.

His most important philosophical work is the Muqaddima, the introduction to a much longer history of the Arabs and Berbers. In this work, Ibn Khaldun clearly defines a science of culture and expounds on the nature of human society and on political and social cycles. Different social groups, nomads, townspeople and traders, interact with and affect one another in a continuous pattern. Religion played an important part in Ibn Khaldun’s conception of the state, and he followed al-Ghazali rather than Ibn Rushd as a surer guide to the truth.

1 Life and cultural context

Abu Zayd ‘Abd al-Rahman ibn Khaldun al-Hadrami was born in Tunis in AH 732/AD 1332. He was deeply rooted in his Islamic background, occupying high government posts in Granada, Morocco, Algeria, Tunisia and Egypt. He spent four years among the Bedouins, and negotiated with both Pedro the Cruel of Spain in Seville and with Timur Lenk (Tamurlane) on the outskirts of Damascus. He was deeply versed in Arabic literature, theology, historiography, jurisprudence and philosophy, and was particularly influenced in the latter by Averroes or Ibn Rushd. He died in Cairo in AH 808/AD 1406.

Ibn Khaldun lived at a time when it was possible to reflect upon a long and profound period of Islamic thought, and he seems to have felt that part of his function as a writer was to sum up this period, with the further aim of pointing towards the future of Islamic intellectual enquiry. As one would expect, he used the terms and concepts of his time, and some have argued that he was a culturally-specific phenomenon (al-Azmeh 1981), so that any attempt at interpreting his thought in Western terms must distort it fatally. This is an error. Like all great thinkers, Ibn Khaldun’s thought contains both specific and universal elements, and the latter can readily be conveyed to modern readers with no more than the usual difficulties of translation from one cultural and historical period to another.

2 Philosophy of history

The work on which Ibn Khaldun’s reputation chiefly rests is the *Muqaddima*, the introduction to his great history of the Arabs and Berbers, the *Kitab al-‘ibar*, which is divided into a further six books. In the first book he presents a general account of sociology, in the second and third a sociology of politics, in the fourth a sociology of urban life, in the fifth a sociology of economics and in the sixth a sociology of knowledge. The whole work is studded with brilliant observations. Thus in the field of economics, Ibn Khaldun understands very clearly the supply and demand factors which affect price, the interdependence of prices and the ripple effects on successive stages of production of a fall in prices, and the nature and function of money and its tendency to circulate from country to country according to demand and the level of activity. In his writings on public finance, he shows why at the beginning of a dynasty taxation yields a large revenue from low rates of assessment, but at the end a small revenue from high rates of assessment. Elsewhere his observations on the evolution of the Arabic language and script are masterly examples of sociological analysis, and his remarks on the difference between acquiring a skill in a language and learning its grammar, and on the use of intuition as opposed to logic in solving difficult problems, can still be read with profit.

However, it is Ibn Khaldun’s views on the nature of the state and society which reveal most clearly both his profundity and the originality that marks him off so sharply from his Muslim predecessors and successors. Ibn Khaldun fully realised that he had created a new discipline, *‘ilm al-‘umran*, the science of culture, and regarded it as surprising that no one had done so before and demarcated it from other disciplines. This science can be of great help to the historian by creating a standard by which to judge accounts of past events. Through the study of human society, one can distinguish between the possible and the impossible, and so distinguish between those of its

phenomena which are essential and those which are merely accidental, and also those which cannot occur at all. He analysed in detail the sources of error in historical writings, in particular partisanship, overconfidence in sources, failure to understand what is intended, a mistaken belief in the truth, the inability to place an event in its real context, the desire to gain the favour of those in high rank, exaggeration, and what he regarded as the most important of all, ignorance of the laws governing the transformation of human society. Ibn Khaldun’s attitude to the study of social phenomena is suffused with a spirit which has caused several commentators to call him the founder of sociology. His attempt at creating a theoretical structure for the analysis of history is a very impressive contribution to the philosophy of history (see History, philosophy of; Society, concept of).

For Ibn Khaldun, human society is necessary since the individual acting alone could acquire neither the necessary food nor security. Only the division of labour, in and through society, makes this possible. The state arises through the need of a restraining force to curb the natural aggression of humanity. A state is inconceivable without a society, while a society is well-nigh impossible without a state (see Political philosophy in classical Islam). Social phenomena seem to obey laws which, while not as absolute as those governing natural phenomena, are sufficiently constant to cause social events to follow regular and well-defined patterns and sequences. Hence a grasp of these laws enables the sociologist to understand the trend of events. These laws operate on masses and cannot be significantly influenced by isolated individuals. There is very little talk of ‘great men’ in Ibn Khaldun’s books; while individuals do affect the course of events, their influence is very limited.

The overwhelming impression given by Ibn Khaldun’s writings is that society is an organism that obeys its own inner laws. These laws can be discovered by applying human reason to data either culled from historical records or obtained by direct observation. These data are fitted into an implicit framework derived from his views on human and social nature, his religious beliefs and the legal precepts and philosophical principles to which he adheres. He argues that more or less the same set of laws operates across societies with the same kind of structure, so that his remarks about nomads apply equally well to Arab Bedouins, both contemporary and pre-Islamic, and to Berbers, Turkomen and Kurds. These laws are explicable sociologically, and are not a mere reflection of biological impulses or physical factors. To be sure, facts such as climate and food are important, but he attributes greater influence to such purely social factors as cohesion, occupation and wealth. This comes out very clearly in his discussion of national characters, for example of Arabs, Persians and Jews, where he is careful to point out that what are regarded as characteristic features can be explained by sociological factors such as nomadism, urbanization and oppression. Similarly, different social groups, such as townspeople, nomads and traders, have their own characteristics derived from their occupations.

Ibn Khaldun sees the historical process as one of constant cyclical change, due mainly to the interaction of two groups, nomads and townspeople. These form the two poles of his mental map; peasants are in between, supplying the towns with food and tax revenue and taking handicrafts in return. Nomads are rough, savage and uncultured, and their presence is always inimical to civilization; however, they are hardy, frugal, uncorrupt in morals, freedom-loving and self-reliant, and so make excellent fighters. In addition, they have a strong sense of *‘asabiya*, which can be translated as ‘group cohesion’ or ‘social solidarity’. This greatly enhances their military potential. Towns, by contrast, are the seats of the crafts, the sciences, the arts and culture. Yet luxury corrupts them, and as a result they become a liability to the state, like women and children who need to be protected. Solidarity is completely relaxed and the arts of defending oneself and of attacking the enemy are forgotten, so they are no match for conquering nomads.

Ibn Khaldun then traces very clearly the political and social cycle. Nomads conquer territories and their leaders establish a new dynasty. At first the new rulers retain their tribal virtues and solidarity, but soon they seek to concentrate all authority in their own hands. Increasingly they rule through a bureaucracy of clients - often foreigners. As their former supporters lose their military virtues there is an increasing use of mercenaries, and soldiers come to be more important than civilians. Luxury corrupts ethical life, and the population decreases. Rising expenditure demands higher taxes, which discourage production and eventually result in lower revenues. The ruler and his clients become isolated from the groups that originally brought them to power. Such a process of decline is taken to last three generations, or about one hundred and twenty years. Religion can influence the nature of such a model; when *‘asabiya* is reinforced by religion its strength is multiplied, and great empires can be founded. Religion can also reinforce the cohesion of an established state. Yet the endless cycle of flowering and decay shows no evolution or progress except for that from the primitive to civilized society.

Ibn Khaldun does occasionally refer to the existence of turning points in history, and thought that he was himself witnessing one of them. The main cause for this great change was the Black Death, which had a profound effect upon Muslim society, together with the Mongol invasions; and he may also have been impressed by the development of Europe, whose merchants and ships thronged the seaports of North Africa and whose soldiers served as mercenaries in the Muslim armies. He suggests that a general change in conditions can produce an entirely new social and political scene, rather as if a new world had been created.

3 Critique of Islamic philosophy

Ibn Khaldun wrote on other topics apart from history, although in his autobiography he is rather coy about admitting it. In his *Shifa' al-sa'il (The Healing of the Seeker)*, he responds to the question as to whether it is possible to attain mystical knowledge without the help of a Sufi master leading the novice along the path. Ibn Khaldun tends to follow al-Ghazali (§3) in reconciling mysticism with theology, but he goes further than the latter in bringing mysticism completely within the purview of the jurisprudent (*faqih*) and in developing a model of the Sufi *shaykh*, or master, as rather similar to the theologian. The fourteenth century, in which Ibn Khaldun was working, was very strongly influenced by what Fakhry (1970) calls 'neo-Hanbalism', which brought with it a strong suspicion of the claims of both mysticism and philosophy. Philosophy was regarded as going beyond its appropriate level of discourse, in that 'the intellect should not be used to weigh such matters as the oneness of God, the other world, the truth of prophecy, the real character of the divine attributes, or anything else that lies beyond the level of the intellect' (*Muqaddima* 3, 38). He refers to the intellect as like a balance which is meant for gold, but which is sometimes inappropriately used for weighing mountains. Logic cannot be applied to this area of enquiry, and must be restricted to non-theological topics (see Logic in Islamic philosophy).

Ibn Khaldun is also critical of Neoplatonic philosophy (see Neoplatonism in Islamic philosophy). The main object of his criticism is the notion of a hierarchy of being, according to which human thought can be progressively purified until it encompasses the First Intellect which is identified with the necessary being, that is, God. He argued that this process is inconceivable without the participation of revelation, so that it is impossible for human beings to achieve the highest level of understanding and happiness through the use of reason alone. Interestingly, the basis of his argument here rests on the irreducibility of the empirical nature of our knowledge of facts, which cannot then be converted into abstract and pure concepts at a higher level of human consciousness.

Ibn Khaldun also had little respect for the political theories of thinkers like al-Farabi (§4), with their notions of rational government being based upon an ideal prophetic law. He saw little point in using theories which dealt with ideals that have nothing to do with the practicalities of contemporary political life. Although Ibn Khaldun rarely agrees with Ibn Rushd, there is no doubt that his thought is strongly marked by the controversy between him and al-Ghazali, the latter being acknowledged as the surer guide to the truth. The basis of Ibn Khaldun's critique of philosophy is his adherence to the notion of the state. Religion has a vital role in society, and any argument that it can be identified with either reason or contact with God is to threaten that function. This is doubtless the basis of his attack on Islamic philosophy and on mysticism.

Although Ibn Khaldun is hostile to a version of Islamic philosophy, his discussion of society is full of observations and ideas which clearly have as their source philosophical distinctions. For example, his account of the three stages in the development of the state, from the nomadic to the militant and finally to the luxurious and decadent is modelled on the three types of soul in Greek thought (see Soul in Islamic philosophy §2), as is his notion of *'asabiya*, of the spirit of cohesion, as a point of equilibrium between different aspects of the soul. One of the features of Ibn Khaldun's work which makes it so thought-provoking is the tension, which he never finally resolved, between a concern to acknowledge the facts of historical change while at the same time bringing those facts under very general theoretical principles. His contribution to the philosophy of history is outstanding.

See also: Culture; History, philosophy of; Political philosophy, history of; Political philosophy in classical Islam

CHARLES ISSAWI
OLIVER LEAMAN

List of works

Ibn Khaldun (1332-1406) *Shifa' al-sa'il li-tahdib al-masa'il (The Healing of the Seeker)*, ed. M. al-Tanji, 1957; trans. R. Pérez, *La voie et la loi ou le maître et le juriste*, Paris: Sindbad, 1991.(Ibn Khaldun's work on

Sufism.)

Ibn Khaldun (1332-1406) *Muqaddima*, ed. and trans. F. Rosenthal, *The Muqaddimah*, Princeton, NJ: Princeton University Press, 1967.(Outstanding translation of this key introduction to Ibn Khaldun’s history of the Arabs and Berbers, the *Kitab al-‘ibar*.)

References and further reading

al-Azmeh, A. (1981) *Ibn Khaldun in Modern Scholarship: A Study in Orientalism*, London: Third World Centre. (A lively defence of the view that most previous commentators have misunderstood Ibn Khaldun.)

al-Azmeh, A. (1990) *Ibn Khaldun: An Essay in Reinterpretation*, London: Routledge.(A guide to the general thought of Ibn Khaldun.)

Fakhry, M. (1970) *A History of Islamic Philosophy*, New York: Columbia University Press; repr. 1985.(A useful account of Ibn Khaldun’s position in Islamic philosophy as a whole.)

Issawi, C. (1986) *An Arab Philosophy of History*, Princeton, NJ: Princeton University Press.(Translation and commentary on some important texts.)

Issawi, C. (1994) *Ibn Khaldun on Roman History: A Study in Sources*, Princeton, NJ: Princeton Papers in Near Eastern Studies.(Ibn Khaldun’s treatment of Roman history.)

Lakhsassi, A. (1996) ‘Ibn Khaldun’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 25, 350-64.(Comprehensive guide to Ibn Khaldun’s thought, emphasizing the links between his social theories and philosophical ideas.)

Mahdi, M. (1957) *Ibn Khaldun’s Philosophy of History*, London: Allen & Unwin.(A study of the philosophical foundations of his theory of culture.)

Rosenthal, E. (1956) ‘The Theory of the Power-State: Ibn Khaldun’s Study of Civilization’, in E. Rosenthal, *Political Thought in Medieval Islam*, Cambridge: Cambridge University Press.(A masterly summary of Ibn Khaldun’s political philosophy.)

Ibn Masarra, Muhammad ibn ‘Abd Allah (883-931)

Muhammad ibn Masarra is said to be responsible for the first structuring of Andalusian Spanish Muslim philosophy. The thrust of his philosophy was to show the agreement between reason and revelation. The two paths taken by honest philosophers and prophets lead to the same goal of reaching the knowledge of the oneness of God. We can only know that God exists but not what His nature is. Ibn Masarra held that the divine attributes of knowledge, will and power are a distinct aspect of the simple and ineffable essence of God, and the Neoplatonic theory that all beings have emanated from him through the First Intellect and are either invisible or apparent. There are two sciences, one of the invisible, transcendental world, the other of the apparent and sensible world. The inner meanings in the sciences can be learned through the science of letters. By studying the enigmatic letters at the beginning of the Qur'anic surahs, *one can decipher the secret knowledge of the truth symbolized by them.*

1 Life and times

Muhammad ibn ‘Abd Allah ibn Masarra was born in Cordoba, Spain, in AH 269/AD 883 and died in AH 319/AD 931. In a hermitage he had founded for his friends and disciples in the Sierra of Cordoba, Ibn Masarra undertook to instruct them in his doctrines, to initiate them into the use of esoteric knowledge and to practice *zuhd* (asceticism) through acts of penance and devotion. His success came from a Socratic style of pedagogy as well as a charismatic personality and skill in communication. After his death the jurists carried out a veritable persecution of his disciples; who had formed themselves into an ascetic order, the Masarriya, in Cordoba and later in Almeria.

Two of Ibn Masarra’s four works, *Kitab al-i‘tibar (On Reflection)* and *Kitab khawass al-huruf (Characteristics of Letters)*, were published in 1982. Both are short tracts which have provided a better understanding of his thought, but because of their conciseness they raise new questions. It is still not possible to reconstruct his philosophical system until the remaining works are found, especially his *Tawhid al-muqinin (The Certain Profession of the Oneness of God)*, where he discussed God’s attributes.

2 Doctrines

M. Asín Palacios, the Spanish scholar who first reconstructed an integral account of Ibn Masarra’s life and thought, concluded that he was the first Andalusian to structure Spanish Islamic philosophy (*hikma*) and that he conveyed his doctrines in a series of *batini* (inward) esoteric images and symbols (Asín Palacios 1972). The centrepiece of Asín’s thesis, however, was the elaboration of a whole theory of Ibn Masarra’s inspiration from a pseudo-Empedocles, who had developed a peculiar form of Plotinian ideas on the One and the five eternal substances of Primal Matter, Intellect, Soul, Nature and Secondary Matter. According to Asín, Ibn Masarra was the founder of a philosophical-mystical school which influenced Jewish, Christian and Muslim medieval philosophers. Andalusian Sufism from Isma‘il al-Ru‘ayni (d. AH 555/AD 1268) to Ibn al-‘Arabi by way of Ibn al-‘Arif (d. AH 536/AD 1141) sprang from the Masarri school.

The thrust of Ibn Masarra’s philosophy is to demonstrate the agreement of reason and revelation. Each takes a different path leading to the same goal, *al-tawhid*, the knowledge of the oneness of God. By using *‘aql*, the intellect with which God endowed human beings, they reflect on God’s signs and rise step by step to the knowledge of the Truth. Those who ascend by way of reason proceed from the bottom up and discover the same truth the Prophets have brought down from on high. In fact, the Qur’an invites us to reflect on the signs of his creation. Reflection (*i‘tibar*) only confirms prophecy; what is learned by authority (*sama‘*) is confirmed by investigation. Ibn Masarra admits, however, that the philosophers and the ancients had attained the knowledge of the true One well before the age of prophecy and without its mediation, a position not acceptable to the religious scholars.

Ibn Masarra conceives of two sciences both created by God. One, the science of the invisible and intelligible reality (*‘ilm al-ghayb*), which cannot be grasped by the senses, is created whole, entire and at once. The other is the science of the apparent and sensible reality (*‘ilm al-shahada*) (Surah 6: 73). The Qur’an, the speech of God, is one whole in its divine essence, but diversified (*mufassal*) with respect to creation. It displays three aspects, each the subject of a different science: the science of divinity (*rububiya*), its signs, evidences and certainty; the science of prophecy with its demonstrative arguments, signs and necessity; and the science of tribulation (*mihna*) with its

laws, promises and threats.

3 God’s attributes

God transcends all human thought and all we can know about his nature is that he exists. His attributes are distinct from him, that is, from his essence (*dhat*). They are, however, related to each other. Strangely enough, Ibn Masarra concludes from that relationship the finitude or creation of the attributes. Like many Mu‘tazilite theologians of his time in the East, he distinguished between the attributes of the essence, which are eternal, and the attributes of action, which are created. This was a way to which the Mu‘tazilites resorted in order to assert the oneness and ineffability of God while maintaining human free will and responsibility. God’s knowledge is only of universals; were he to know particulars, his oneness would be jeopardized and our moral responsibility denied.

In making the distinction between God’s essence and his action, Ibn Masarra established three hierarchical attributes, the highest of which is connected to God’s essence and the other two to his actions. These are divinity (*aluhiya*), royalty (*mulk*) and grace (*ni‘ma*) or creation (*khalq*), through which God the Artificer (*al-sani‘*) is manifested. This hierarchy is reflected in the way human society is organized.

4 God’s creation

All beings are divided into four categories, some nobler than others in accordance with the following scheme. First, there is the Being, or essence of God (*dhat*), separate, unique, ineffable, infinite and motionless; it is the ultimate, the visible and the invisible. The remaining beings are the signs that point to Him. Second is the Universal Intellect (*al-‘aql al-kulli*), which is the conception or idea of things. It is spiritual by nature and permanent. It is the Mother of the Book (*umm al-kitab*) (Surah 3: 6), and the Preserved Tablet (*al-lawh al-mahfuz*) (Surah 85: 21) on which all things are inscribed. The totality of what is in the Book is the idea (*mithal*, Eidos) of the universe, whatever was, is or shall be. It is also the Throne (Surah 10: 2) which incites motion in response to God’s volition and will. The relationship of the Intellect to God is similar to the relation of the sun’s light to the sun. Third is the Great Soul (*al-nafs al-kubra*) that carries the body of the universe. The relation of the Soul to the Intellect is like the light of the moon to that of the sun. Through this Soul, immersed in materiality, Royalty (*mulk*) is constituted and the celestial spheres are held. To Royalty are predicated government and politics. Finally, lower than the Great Soul is the Physical Soul (*al-nafs al-tabi’iya*), which is completely immersed in corporeality and is the efficient cause of corporeal beings. The Throne encloses the invisible world (*‘alam al-ghayb*) and the Great Soul encloses the visible (*‘alam al-shahada*).

The origination of the cosmos has been achieved in time by the command ‘Be’ (*kun*), expressing the volition and will according to knowledge. When the One wants to do something, he causes it to appear in the Preserved Tablet. This in itself is the command (*amr*) to set the idea into action by his willing. God, according to him, is concealed from creation by two veils from the perspective of his creation, inasmuch as nothing can conceal him from the perspective of his essence. Motion is then set by the Throne, since no action *ad extra* can be attributed to the One.

Unlike the pseudo-Empedocles, who conceived of love and discord as the driving force of creation, Ibn Masarra talked about capacity and power designating them as truth (*haqq*) (Surah 2: 72). In the final analysis God is the Aristotelian unmoved mover, but in Neoplatonic style all creation emanates from him. Unlike Plotinus, Ibn Massara finds that the processes of emanation and creation are the results of God’s will (*irada*) and deliberate action. Both he and Plotinus agree on the intermediary roles played by the Intellect and the Soul in the creation of the material world, but whereas Plotinus believed in involuntary emanation, Ibn Massara retained the Islamic view of voluntary creation (see Neoplatonism; Neoplatonism in Islamic philosophy).

The principles from which all creatures have come are fourteen in number, ten of which are in the sublunar world: chaos (*al-‘ama’*), primordial dust (*al-haba’*), which is considered by some as the *materia prima*, air, wind, atmosphere, water, fire, light, darkness and clay. The remaining four, the Pen (*qalam*), the Tablet, the Command and the Spiritual locus (*makan*) exist in the world above. From the fourteen are made the Throne, paradise, hell, the seven heavens, the earth, the angels, the jinn, human beings, animals and vegetation.

5 Esoterism and mysticism

In his work *Kitab khawass al-huruf (Book of the Characteristics of Letters)*, Ibn Masarra appears as an esoteric

(*batini*) philosopher investigating the esoteric meanings of the *nuraniya*, the fourteen separate letters which introduce certain surahs of the Qur’an, basically following the tradition of Islamic gnosis. The mysterious letters, according to the Batini school, represented the universe so that its entirety is a book whose letters are God’s words. The ‘science of letters’ followed by Ibn Masarra had nothing to do with divination or magic; it is merely a path to the discovery of the truths hidden behind the symbols. In this he was inspired by the work of Sahl al-Tustari (d. AH 283/AD 896), the author of a similar work on the science of letters.

Reflection (*i‘tibar*) allows us to decipher the principles of all beings. The basic idea is to show that the different degrees that constitute beings in general correspond to the surah’s *fawatih* (opening letters) as well as to the order of being. The letters are twenty-eight in number, equal to the length of the lunar phases. Fourteen are exoteric and the remaining fourteen are esoteric. These are used by God to manifest his knowledge: their secret meanings have been bestowed upon the Prophet Muhammad as expressed in the Qur’an, and consequently the Qur’an is the source of all knowledge, old and new. The steps leading to paradise and salvation are equal in number to the Qur’anic verses and to the number of God’s beautiful names, excepting the great name of Allah.

The first letter, *alif*, is the alpha and omega of all letters inasmuch as it represents the principle of all things. It is the first manifestation of God and his will; it is a metaphor for the production of things (*takwin*), the emergence of justice and the permanent and unchanging primordial decree (*al-qada’ al-awwal*). It never rests and continuously causes generation and corruption. This decree has two aspects. A prior aspect (*sabiq*) is connected to the Preserved Tablet, the tablet of the Universal Intellect where all things are inscribed. This is a decree that does not respond to invocation. The second aspect, the diversifier (*mufassil*), particularizes all things that are not permanent. Like the other attributes, the two decrees manifesting God’s knowledge and power are other than God, although they are not created in time. The concept of *huduth*, or coming to be, is realized only in time; but God’s knowledge, according to Ibn Masarra, whether it is knowledge of the universals or particulars, is not in time. Coming to be in time is the particularization of beings found in a locus performed by that decree that particularizes things and responds to invocation.

Human salvation can be achieved through either the *via reflectiva* or the *via prophetica*, an idea considered heretical by most Muslim theologians (see Islamic theology). In both cases, individuals have to follow certain rules in order to free their souls from the bondage of materiality. Ibn Masarra distinguishes clearly between the soul (*ruh*) and spirit (*nafs*), with the latter being the prototype (*mithal*) of the first.

There is a tradition in Andalusian literature to the effect that Ibn Masarra enjoyed great respect and veneration in spite of the fact that his teachings were criticized and refuted. On the other hand, his disciples were persecuted. Transformed into an ascetic society, his disciples first in Cordova and later in Almeria put into practice his Sufi and esoteric teachings. He is certainly one of the first mystical-philosophical Andalusians. His Sufi teachings as well as his works continued to circulate and to be studied for centuries. His influence on Ibn al-‘Arabi is attested by the many references to him in the latter’s works and by similarity in a number of ideas, especially in the continuous use of similes of light and illumination to describe the essence of God.

See also: Ibn al-‘Arabi; Mystical philosophy in Islam; Neoplatonism in Islamic philosophy

GEORGE N. ATIYEH

List of works

Ibn Masarra (883-931) *Kitab al-i‘tibar (On Reflection)*, ed. M.K. Ja‘far, *Min al-turath al-falsafi li-ibn Masarrah: 1. Risalat al-i‘tibar, 2. Khawass al-huruf*, Cairo, 1982.(One of Ibn Massara’s two surviving works.)

Ibn Masarra (883-931) *Kitab khawass al-huruf (Characteristics of Letters)*, ed. M.K. Ja‘far, *Min al-turath al-falsafi li-ibn Masarrah: 1. Risalat al-i‘tibar, 2. Khawass al-huruf*, Cairo, 1982.(One of Ibn Massara’s two surviving works.)

References and further reading

Addas, C. (1992) ‘Andalusi Mysticism and the Rise of Ibn Arabi’, in S.K. Jayyusi (ed.) *The Legacy of Muslim Spain*, Leiden: Brill.(A well-written study on Sufism in Spain including a sizeable section on Ibn Masarrah’s mystical teachings.)

Asín Palacios, M. (1972) *The Mystical Philosophy of Ibn Masarra and His Followers*, trans. E.H. Douglas and H.W. Yoder, Leiden: Brill.(Still the major work on Ibn Masarra, although some of the conclusions have been

challenged.)

Arnaldez, R. (1960) ‘Ibn Masarra’, in *The Encyclopedia of Islam*, 2nd edn, vol. 3: 868-72.(Concise account of Ibn Masarra’s life and thought.)

Cruz Hernández, M. (1981) ‘La Persecución anti-Masarri durante el reinado de ‘Abd al-Rahman al-Nasir li-Din Allah según Ibn Hayyan’ (The Anti-Masarri Persecution During the Reign of ‘Abd al-Rahman According to Ibn Hayyan), *al-Qantara* II (182): 51-67.(An analysis of the account of Ibn Hayyan’s *al-Muqtabas* concerning the persecution of the disciples of Ibn Masarrah, concluding that it does not differ substantially from that of Asín.)

Goodman, L. (1996) ‘Ibn Masarra’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 20, 277-93.(The role of Ibn Masarra in creating a distinctive philosophy and form of mysticism in al-Andalus.)

Stern, S.M. (1983) ‘Ibn Masarra, Follower of Pseudo-Empedocles - An Illusion’, in F. Zimmerman (ed.) *Medieval Arabic and Hebrew Thought*, London: Variorum.(A criticism of Asín’s theory.)

Ternero, E. (1993) ‘Noticia sobre la publicación de obras inéditas de ibn Masarra’ (Review of the Publication of Ibn Masarra’s Unedited Works), *al-Qantara* XIV: 47-64.(A summary of the two works of Ibn Masarra, published for the first time in 1982.)

Ibn Miskawayh, Ahmad ibn Muhammad (*c.*940-1030)

Like so many of his contemporaries in the fourth and fifth centuries AH *(tenth and eleventh centuries* AD*) Ibn Miskawayh was eclectic in philosophy, basing his approach upon the rich variety of Greek philosophy that had been translated into Arabic. Although he applied that philosophy to specifically Islamic problems, he rarely used religion to modify philosophy, and so came to be known as very much an Islamic humanist. He represents the tendency in Islamic philosophy to fit Islam into a wider system of rational practices common to all humanity.*

Ibn Miskawayh's Neoplatonism has both a practical and a theoretical side. He provides rules for the preservation of moral health based on a view of the cultivation of character. These describe the ways in which the various parts of the soul can be brought together into harmony, so achieving happiness. It is the role of the moral philosopher to prescribe rules for moral health, just as the doctor prescribes rules for physical health. Moral health is based upon a combination of intellectual development and practical action.

1 Metaphysics

Like so many of his philosophical contemporaries, Ahmad ibn Muhammad ibn Miskawayh, born in Rayy in Persia *c.*AH 320/AD 940, combined an active political career with an important philosophical role. A historian as well as a philosopher, he served as a Buwayhid official at Baghdad, Isfahan and Rayy. He was a member of the distinguished group of intellectuals including al-Tawhidi and al-Sijistani. He died in AH 421/AD 1030. Although not an important figure on the creative side of Islamic philosophy, he is a very interesting adaptor of existing ideas, especially those arising out of the Neoplatonic tradition in the Islamic world (see Neoplatonism in Islamic philosophy).

Ibn Miskawayh wrote on a wide variety of topics, ranging from history to psychology and chemistry, but in philosophy his metaphysics seems to have been generally informed by a version of Neoplatonism. He avoids the problem of reconciling religion with philosophy by claiming that the Greek philosophers were in no doubt concerning the unity and existence of God. He goes so far as to suggest that Aristotle's identification of the creator with an unmoved mover is a powerful argument in favour of a creator acceptable to Islam, since the very distinct nature of such a being prevents our normal categories of description from making sense. Such a creator can only be described in terms of negative concepts, an interesting prefigurement of the tradition of the *via negativa* in philosophy. He has an unusual account of emanation, wherein the deity produces the active intellect, the soul and the heavens without intermediaries, making one suspect that he did not have a firm grasp of the distinction between emanation and creation. The normal Neoplatonic account of emanation then current in Islamic philosophy used the notion of a scale of being that separates these different divine products far more radically. It is difficult to see how Ibn Miskawayh really reconciles metaphysical difficulties at this point.

2 Ethical writings

Ibn Miskawayh's work on ethics, however, is of a much higher order, and does show evidence of considerable conceptual complexity. In his *Taharat al-a'raq (Purity of Dispositions)*, better known as *Tahdhib al-akhlaq (Cultivation of Morals)* - which is not to be confused by the work of the same name by Yahya Ibn 'Adi - he sets out to show how we might acquire the right dispositions to perform morally correct actions in an organized and systematic manner.

The basis of his argument is his account, adopted from Plato, of the nature of the soul, which he sees as a self-subsisting entity or substance, in marked contrast to the Aristotelian notion (see Soul, nature and immortality of the). The soul distinguishes us from animals, from other human beings and from things, and it uses the body and the parts of the body to attempt to come into contact with more spiritual realms of being. The soul cannot be an accident (or property of the body) because it has the power to distinguish between accidents and essential concepts and is not limited to awareness of accidental things by the senses. Rather, it can apprehend a great variety of immaterial and abstract entities. If the soul were only an accident it could do none of these things, but could only perform in the limited way of the physical parts of the body. The soul is not an accident, and when we want to concentrate upon abstract issues the body is actually an obstruction that we must avoid if we are to make contact with intelligible reality. The soul, then, is an immortal and independent substance that controls the body. It has an

essence opposite to that of the body, and so cannot die; it is involved in an eternal and circular motion, replicated by the organization of the heavens. This motion takes two directions, either upwards towards reason and the active intellect or downwards towards matter. Our happiness arises through upwards movement, our misfortunes through movement in the opposite direction.

Ibn Miskawayh's discussion of virtue combines Aristotelian with Platonic ideas (see Virtue ethics). Virtue is the perfection of the aspect of the soul (that is, human reason) that represents the essence of humanity and distinguishes it from lower forms of existence. Our virtue increases in so far as we develop and improve our ability to deliberate and apply reason to our lives. We should do this in accordance with the mean, the point most distant from two extremes, and justice results when we manage to achieve this. Ibn Miskawayh combines the Platonic division of virtues with an Aristotelian understanding of what virtue actually is, and adds to this the idea that the more these virtues can be treated as a unity, the better. This is because, he argues, that unity is equivalent to perfection, while multiplicity is equivalent to a meaningless plurality of physical objects. This idea is not just based upon a Pythagorean aesthetic (see Pythagoreanism). Ibn Miskawayh argues further that the notion of justice when it deals with eternal and immaterial principles is a simple idea, while human justice by contrast is variable and depends upon the changing nature of particular states and communities. The law of the state is based upon the contingent features of the time, while the divine law specifies what is to be done everywhere and at every time.

Ibn Miskawayh uses the notion of friendship to distinguish between those relationships that are essentially transitory and variable (in particular those based upon pleasure) and those based upon the intellect, which are also pleasurable but not in a physical way. Our souls can recognize similarly perfected souls, and as a result enjoy intense intellectual delight. This is very different from the normal kind of friendship, in which people form relationships with each other because they want to get something out of it. Still, even those capable of the most perfect form of relationship have to involve themselves in the less perfect levels of friendship, since they must live in society if they are to achieve perfection, and so must satisfy at least some of the expectations of society (see Friendship). The highest form of happiness exists when we can abandon the requirements of this world and are able to receive the emanations flowing from above that will perfect our intellects and enable us to be illuminated by divine light. The eventual aim seems to be the throwing off of the trappings of our physical existence and following entirely spiritual aims in mystical contemplation of the deity.

3 Practical ethics and humanism

This mystical level of happiness seems to rank higher than mere intellectual perfection, yet Ibn Miskawayh is particularly interesting in the practical advice he gives on how to develop our ordinary capacity for virtue. He regards the cultivation of our moral health in a very Aristotelian way as akin to the cultivation of physical health, requiring measures to preserve our moral equilibrium (see Aristotle §§23-6). We ought to keep our emotions under control and carry out practices that help both to restrain us on particular occasions and also to develop personality traits that will maintain that level of restraint throughout our lives. To eradicate faults, we must investigate their ultimate causes and seek to replace these with more helpful alternatives. Take the fear of death, for example: this is a baseless fear, since the soul is immortal and cannot die. Our bodies will perish, but they must do so since we are contingent; to acknowledge that contingency and also to wish that we were not thus contingent is some sort of contradiction. If we are worried by the pain involved in dying, then it is the pain we fear, not death itself. Ibn Miskawayh argues, along with al-Kindi (§3) and the Cynics and Stoics who no doubt influenced him on this issue, that to reconcile ourselves to reality we have to understand the real nature of our feelings (see Cynics; Stoicism). We have to use reason to work out what we should do and feel, since otherwise we are at the mercy of our feelings and the varying influences that come to us from outside ourselves (see Death).

This emphasis upon the capacity of the human mind to use reason to help us determine what we should do and who we are has led the most distinguished commentator on Ibn Miskawayh, Mohammed Arkoun (1970), to call him a humanist and part of the general humanist movement of his time. It is certainly true that religion plays a small part in Ibn Miskawayh's writings, and when he does consider Islam he often gives its religious practices a rather instrumental rationale. Al-Ghazali (§2) was infuriated by Ibn Miskawayh's suggestion that the point of communal prayer is to base religion upon the natural gregariousness of human beings in society. This seemed to al-Ghazali to disparage the religious enterprise, since he argued that the significance of religious rituals is that they are specified by the religion, and there can be no other reason. Their rationale is that they are unreasonable. God indicates the

huge gap that exists between him and us by setting us unpleasant and difficult tasks. For Ibn Miskawayh, the reason for the ritual is that it has a part to play in helping us adapt to religious life, using the dispositions that are natural to us, so that the rules and customs of religion are essentially reasonable. A whole range of authorities may be consulted to help us understand our religious duties concerning how we are to live and what we are to believe; some of these are Islamic, while others are not. Ibn Miskawayh seems on the whole to accord greater respect to Greek rather than specifically Islamic authorities.

Ibn Miskawayh's thought proved to be influential. His style, combining abstract thought with practical observations, is attractive and remained popular long after his death. Sometimes he merely presents aspects of 'wisdom' literature from previous cultures; sometimes he provides practical comments upon moral problems that are entirely unanalytical. At its best, however, his philosophy is highly analytical and maintains a high degree of coherence and consistency. The fact that he mixes together aspects of Plato, Aristotle, Pythagoras, Galen and other thinkers influenced by Greek philosophy is not an indication of cultural looting but rather a creative attempt at using these different approaches to cast light upon important issues.

Ibn Miskawayh shows how possible it is to combine a Platonic conception of the soul with an Aristotelian account of moral development. The idea of a still higher realm of being at which the soul comes into contact with divine reality is a perfectly feasible addition to the account he gives of social and intellectual life. He never imports the notion of revelation to resolve theoretical difficulties, and we have seen how his approach both annoyed and stimulated al-Ghazali. It is perhaps the combination in Ibn Miskawayh of elegance of style, practical relevance and philosophical rigour that prolonged his influence in the Islamic world.

See also: Ethics in Islamic philosophy; Neoplatonism in Islamic philosophy; Soul in Islamic philosophy

OLIVER LEAMAN

List of works

Ibn Miskawayh (before 1030) *Tahdhib al-akhlaq (Cultivation of Morals)*, ed. C. Zurayk, Beirut: American University of Beirut Centennial Publications, 1966; trans. C. Zurayk, *The Refinement of Character*, Beirut: American University of Beirut, 1968.(A summary of Ibn Miskawayh's ethical system. This work is also known as *Taharat al-a'raq (Purity of Dispositions)*.)

References and further reading

Arkoun, M. (1961-2) 'Deux épîtres de Miskawayh' (Two Treatises of Miskawayh), *Bulletin d'Études Orientales* (Institut Français de Damas) 17: 7-74.(A clear account of Ibn Miskawayh's general metaphysics as well as his ethics.)

Arkoun, M. (1970) *Contribution à l'Étude de l'humanisme arabe au IVe/Xe siècle: Miskawayh, philosophe et historien (320/325-421) = (932/936-1030) (Contribution to the Study of Arab Humanism in the 4th/10th Century: Miskawayh, Philosopher and Historian)*, Paris: Vrin; revised 2nd edn, 1982. (The standard exegesis of Ibn Miskawayh's contribution to philosophy and history.)

Fakhry, M. (1975) 'The Platonism of Miskawayh and its Implications for his Ethics', *Studia Islamica* 43: 39-57. (A careful account of the Platonic and Neoplatonic influences on Ibn Miskawayh.)

Goodman, L. (1996) 'Friendship in Aristotle, Miskawayh and al-Ghazali', in O. Leaman (ed.) *Friendship East and West: Philosophical Perspectives*, Richmond: Curzon, 164-91.(A range of views on friendship, and their philosophical significance explained.)

Kraemer, J. (1984) 'Humanism in the Renaissance of Islam: a Preliminary Study', *Journal of the American Oriental Society* 104 (1): 135-64.(An account of Ibn Miskawayh's place in the culture of Islamic humanism.)

Leaman, O. (1996a) 'Ibn Miskawayh', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 252-7.(An account of the context within which Ibn Miskawayh worked and the influence of his views.)

Leaman, O. (1996b) 'Islamic Humanism in the Fourth/Tenth Century', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 155-61.(Survey of a group of thinkers including Ibn Miskawayh, al-Tawhidi and al-Sijistani.)

Leaman, O. (1996c) 'Secular Friendship and Religious Devotion', in O. Leaman (ed.) *Friendship East and West: Philosophical Perspectives*, Richmond: Curzon.(Account of Ibn Miskawayh's notion of friendship and comparison with contrary views.)

Ibn Paquda, Bahya (*fl.* early 12th century)

Bahya ibn Paquda, the chief exponent of Jewish pietism, gave that ecumenical strand of thought and practice a markedly philosophical cast, preferring the intellectual to the fideistic side of pietist tradition, and embracing rationalism as the ally of faith rather than rejecting it as an enemy. Drawing selectively from Muslim as well as Jewish sources, Bahya's spiritual vademecum, *al-Hidaya ila fara'id al-qulub (The Book of Guidance to the Duties of the Heart), was widely studied ever since its composition, especially in its medieval Hebrew translation, and parts of it are even included in the liturgical meditations of the Ten Days of Penitence. In it, Bahya thematizes his materials carefully, using his own sense of the reasonable to structure pietism as a philosophical system, controlling the monistic penchant of mysticism and disciplining the ascetic tendencies of the devotional cast of mind. Maimonides found Bahya's asceticism excessive and rejected Bahya's related leanings towards predestinarianism and resignation; but he quietly adopted Bahya's moral and intellectual interpretation of the mystic quest for unity with God, fell into step with his predilection for spiritual immortality as distinguished from bodily resurrection, echoed his affirmation of God's absolute unity and simplicity, and concurred in his admiration for negative theology, the theology of divine transcendence.*

1 General character of Bahya's thought

Born in the later eleventh century, Bahya ibn Paquda flourished in Saragossa and served as a rabbinic judge. He is the probable author of some twenty Hebrew *piyyutim*, or hymns, extant under the name Bahya, but his best known work is the fountainhead of Jewish philosophical pietism, *al-Hidaya ila fara'id al-qulub (The Book of Guidance to the Duties of the Heart)*, widely circulated since medieval times in the Hebrew translation (1161) of Judah ibn Tibbon and thus familiar to traditional Jews as *Hovot ha-Levavot*. This version, although its struggle for literalism sacrifices the fluidity of the original, was widely copied and abstracted in manuscript and first printed at Naples in 1489, again at Venice in 1548, and more critically at Mantua in 1559. It attracted Hebrew commentaries by various hands including those published in 1691, 1774, 1790, 1803 and 1836, and was translated into Portuguese, Italian, German, Spanish, Ladino, Yiddish, Arabic(!), French and English.

Steeped in Biblical and Rabbinic learning and the philosophy of Saadiah Gaon, Bahya also knows the Greek philosophers and Galen in their Arabic texts. He draws upon *kalam* and Sufi writers and the Shi‘ite group known as the Sincere Brethren of Basra (see Ikhwan al-Safa'), whose cosmopolitan and humanistic pietism is akin to his own. Bahya does not seem to know the work of Avicenna; he shows numerous affinities to the writings of al-Ghazali, but exhibits a milder spirit and far greater confidence in the philosophical tradition. It is no longer believed that al-Ghazali was a source for him; rather, the two seem to share common sources in Sufi literature.

2 The duties of the heart

Complaining of the externalism prominent in Jewish religiosity, attention to the 'duties of the limbs' rather to the detriment of the inward or spiritual 'duties of the heart', Bahya announces at the outset of his magnum opus that he intends to redress the balance. The heart here is our spiritual identity, celebrated by Pascal but named as early as the charge in Deuteronomy that we must love God with all our hearts. As Bahya's Muslim contemporary al-Ghazali carefully notes, the heart in this instance is not the physical organ. Bahya does not contrast reason with the promptings of the heart. On the contrary, the heart for him includes the mind. He argues dialectically that external obligations depend for their motivation on the commitment of the heart and that intention, which is rooted in heart, is of essence in determining the status of our acts, even at so behavioural a level as in the determination of the difference between murder and manslaughter. But the roots of sound intention must lie in understanding, and those who would attain the ideal of obedience to God's commandments must therefore pursue wisdom.

Wisdom comprises physics, mathematics and theology, the three canonical branches of philosophical exploration. It is theology that is crucial in religion and that makes philosophy a religious obligation. The weak-minded rightly rely on tradition and authority to frame their outlook, but those of more penetrating mind have an obligation to explore the foundations of faith and action, lest their actions be at odds with their intentions, through the incoherency or inadequacy of their understanding. In matters of law, it was long ago established that the majority decision of the learned is determinative of practice; here, convention, grounded in scriptural revelation, will suffice. But in theology we must probe the tradition, seeking not merely uniformity but the comprehension that

will inform our lives. The argument is a pietist's counterpart to the Socratic thesis that the unexamined life is not worth living (see Socrates §4).

Bahya organizes his treatise around ten pietist virtues, a thematic easily lost sight of if we translate the title of the first of his ten chapters as 'On the Unity of God'. The title, *Tawhid*, in fact means monotheism, the affirmation of God's unity, and it is here in the first instance that Bahya seeks to bring philosophy to the aid of piety by clarifying the content of monotheistic belief. Such belief is to be affirmed by the heart, not merely uttered by the tongue. Its affirmation is our primal religious obligation, because from it flow all the rest. But if commitment to God is to be a ground of our intentions, that commitment must be conceptually lucid and firmly understood. Ordinary Israelites may content themselves with the apologetics of *kalam* (dialectical theology), but those who are capable of the understanding ordained by Deuteronomy (4: 39, 'lay it to heart') are obliged to delve beneath rote faith. They must seek deeper understanding not only for the sake of countrymen who need philosophical aid, but also for the nations of the world. In this way they contribute to the fulfilment of the prophecy (Isaiah 2: 3) that God's word will issue forth from Zion.

Such inquirers will find grounds for their belief in the beauty and order of the cosmos and in the recognition that nothing can create itself. Rather, the whole ontic hierarchy, the architecture of heaven and earth rises to the utter simplicity of God and the primal source of differentiation, in his will. To Moses, God revealed himself conceptually as 'I am that I am', but to lesser minds, he is revealed as a figure of tradition, 'God of Abraham, God of Isaac, and God of Jacob'.

The second virtue is reflectiveness (*'i'tibar*), leading to fruitful meditations on creation and God's grace toward it. The study of nature is not merely a way of recognizing God's creative powers but also a way of exercising our own God-given potential, raising ourselves above the animal level. When Isaiah (40: 26) exhorts us 'Lift up your eyes on high and behold who created these things', we can learn from rabbinic teaching (Shabbat 75a) that one who is able to study astronomy but does not is among those who 'regard not the work of the Lord' (Isaiah 5: 12).

The boundedness of nature, the microcosm of human anatomy, the utility of the human faculties, the usefulness of all creatures in sustaining the world at large and ourselves in particular, the system of the arts and sciences, and the beauties of God's law are all objects of intellectual wonder, instruction and delight that draw the mind upward toward God. Man is the chief but not the sole object of creation, and it is through self-examination, as practised in philosophy that we gain our clearest apprehension of God's work and grace in nature.

The third virtue is obedience to God, a reasonable response to the grace we have received. Even if our performance is imperfect, we should try to make our intentions pure, although God alone acts entirely without self-interest. Human nature rests on a deep duality that generates inevitable tensions. We have a desire and appetite to populate the earth. However, we reach understanding only through self-denial, that is, denial of the lower self, the ego, which pietists sharply distinguish from the seat of pure intentions, the heart. The Law provides the discipline we require in mediating the demands of our earthly natures and our heaven-bound souls, so ill at ease in this world. However, study of the Law also fortifies the mind and is itself a spiritual ladder.

We may manifest our obedience to God through acts of submission expressive of fear and dread or through words of praise and glorification. External acts are significant enough to be subject to reward and punishment, but only inner devotion is truly free and sincere. The finite obligations of the Law are thus only the gateway to the open-ended obligations of the heart, the preparatory discipline for a rise typified by the infinite scope that opens in the quest for understanding, which grows 'day unto day' (Psalms 19: 3).

External duties are precisely specified; they discipline the appetites of both the many and the wise, but to follow the Law solely for the sake of a reward is tantamount to polytheism. The wisest follow God's commands for his sake, not their own; that is, they entrust their all to God, not for any gain but simply from the love of God.

The mind's persuasion is God's inspiration, Bahya argues. He illustrates this thesis by inserting a dialogue into his text in which the mind teaches the soul the nature of the obedience she owes God and shows her that philosophy can strengthen her resolve in ways that tradition cannot. Continuing their dialogue, the mind shows the soul that the conflict between free will and predestination is insoluble by reason or scriptural authority. The pious need not resolve this antinomy. They will, as their character calls upon them to do, accept full responsibility with regard to their own moral and spiritual obligations and treat all worldly outcomes as decrees of divine wisdom. Our

‘inclination’ urges just the opposite, pressing us to assume control in worldly matters (to take credit when they succeed, to feel anxiety when they hang in the balance, and to suffer regret when they fail) but to abdicate our moral and spiritual obligations by pleading the excuse of destiny, fate or nature. Pursuing a pietist stance, Bahya has here accurately pinioned much that passes for metaphysical discussion of the problem of free will (see Free will; Pietism).

The fourth virtue is reliance upon God alone. Bahya interprets this important pietist virtue as an appeal against worldliness and in behalf of peace of mind. He contrasts the pious man with the alchemist, whose Faustian desire for control drives him to hermetic studies, unending travels, and the constant worldly anxieties of *homo faber*. Mitigating the anti-causal doctrine of Islamic occasionalists, Bahya argues that natural causes are too weak to produce their effects but must be deemed to act at God’s behest (see Occasionalism). God works through nature, not despite it. So it is permissible for us to exert ourselves in our own behalf, paradigmatically, by the use of medicine. But we must constantly remember that it is God on whom we are reliant. We must keep our eye on the Source, not the instrumentality through which grace is manifested.

The fifth virtue is *ikhlas*, purity of intention. This sincerity is a material rather than a mere formal virtue; that is, sincerity here means focus on the highest ideal, direction of all our actions towards God. It is not a matter of mere faithfulness to our own intentions. The tranquillity that comes from trust, the clarity that comes to philosophical monotheism, and the perfect absence of hypocrisy that was Bahya’s original goal are all products of the virtue of *ikhlas*. It is the essence of self-rule, and thus of freedom.

The sixth virtue is humility, necessary to counter the pride that might beset even our piety. Humility manifests itself bodily in softspokenness, clemency and self-restraint. It is attained through reflection on our fragility and abjectness before God and manifested in patience, benevolence and forbearance, modesty in the face of well-founded praise, and firm demurral in the face of unfounded praise. Mastery belongs only to God, and one cannot attain any part of piety without first abandoning all claims to mastery.

The seventh virtue is penitence, a special gift from God in recognition of the inadequacy of human deeds. The key to penitence is self-knowledge; thus the eighth virtue is self-scrutiny. This, like any intellectual virtue, imposes obligations that vary in accordance with our intellectual capacities. At its lowest it means simply reflection on the wretchedness of the human body and its vile origins. At a higher level, it involves reflections on God’s grace; but ultimately it means constant, probing examination of our own intentions and the adequacy of our pursuit of the highest of them, allying ourselves firmly with others who can aid us in that pursuit.

The ninth virtue is abstemiousness (*zuhd*), a Sufi virtue and part of the legacy that all three monotheistic religions inherit from Hellenistic asceticism. All living beings, Bahya argues, need asceticism as a check upon their appetites and impulses, an inner counterpart to the economic and social discipline of the polity. Every nation has its own specialization in one or another art or science, through which it contributes to the world. Asceticism, which fosters the mind’s dominance over the entire soul, is Israel’s forte, allowing its people to overcome their powerful appetites and passions and become the spiritual physicians for the world, as it is said, ‘The fruit of the righteous is a tree of life’ (Proverbs 11: 30). Restoring the ideal of renunciation to its Platonic roots in the conception of sophrosyne, Bahya follows up on Sufi and other pietist elaborations of the ascetic ideal to find Jewish application for the notion of abstemiousness in the social sphere, through a cheerful and gentle comportment towards others; in the bodily sphere, through the discipline of the Torah’s ritual laws, with their complex system of permitted and forbidden pleasures, where we are urged to progress to the point where we regard permitted pleasures as on a par with forbidden ones (but not to the point that we harm ourselves) and in the inward or spiritual sphere, through the conquest of covetousness and worldly ambition.

The final virtue is the love of God, the summation of all that has gone before, a sincere desire to cleave to God, in gratitude, penitence and awe, pursuing a mystic intimacy attained not by ecstatic exercises but by meditations on one’s own history, on Scripture and on nature, by practice of the divine commandments, acceptance of God’s sovereignty, and living the life of the pious and abstemious. Those who truly love God will give up property, body and soul for him, as Abraham did, first by placing his worldly goods at the service of his guests, so as to draw them to God; second by undergoing circumcision; and third by his willingness to sacrifice Isaac in obedience to God’s command. The last is a superhuman piety, accessible to us only through God’s aid. Fasts and vigils are part of intimacy with God; but so are justice, neighbourly reproof and the public institution of justice, as we read in

Avot 5.18: ‘Moses was meritorious and made the many meritorious, and the merit of the many came to them through him, as it is written, ‘he worked the justice of the Lord and His righteousness with Israel’ (Deuteronomy 33: 21). That is, Moses shared in God’s attribute of justice by becoming its instrumentality in the world.

See also: Pietism; Virtues and vices

L.E. GOODMAN

List of works

Ibn Pakuda, Bahya (before 1161) *al-Hidaya ila fara’id al-qulub (The Book of Guidance to the Duties of the Heart)*, ed. A.S. Yahuda, Leiden: Brill, 1912; trans. M. Mansoor, *The Book of Direction to the Duties of the Heart*, London: Routledge, 1973. (Judah ibn Tibbon’s medieval translation is edited with English translation by M. Hyamson, New York, 1925-47, 5 vols; repr. Jerusalem: Boys Town Publishers, 1965.)

References and further reading

Eisenberg, Y. (1981) ‘Reason and Emotion in *Duties of the Heart*’, *Daat* 7: 5-35. (Examines the interplay of rationality with the heart in Bahya’s pietist psychology.)

Goodman, L.E. (1983) ‘Bahya on the Antinomy of Free Will and Predestination’, *Journal of the History of Ideas* 44: 115-30.(Explores the pietist approach to the philosophical problem of free will and determinism and parallels between Bahya’s approach and that of Kant.)

Vajda, G. (1937) ‘Le Dialogue de l’âme et de la raison dans les *Devoirs des Coeurs* de Bahya Ibn Paquda’ (The Dialogue of the Soul and Reason in Bahya ibn Paquda’s *Duties of the Heart*), *Revue des Études Juives* 102: 93-104.(Deals with Bahya’s literary device of setting up an internal device between reason and the soul.)

Vajda, G. (1947) *La Théologie Ascetique de Bahya ibn Paquda (The Ascetic Theology of Bahya ibn Paquda)*, Paris: Cahiers de la Société Asiatique, vol. 7.(Surveys Bahya’s efforts to create a balanced asceticism on the foundations of his pietist spirituality.)

Ibn Rushd, Abu’l Walid Muhammad (1126-98)

Ibn Rushd (Averroes) is regarded by many as the most important of the Islamic philosophers. A product of twelfth-century Islamic Spain, he set out to integrate Aristotelian philosophy with Islamic thought. A common theme throughout his writings is that there is no incompatibility between religion and philosophy when both are properly understood. His contributions to philosophy took many forms, ranging from his detailed commentaries on Aristotle, his defence of philosophy against the attacks of those who condemned it as contrary to Islam and his construction of a form of Aristotelianism which cleansed it, as far as was possible at the time, of Neoplatonic influences.

His thought is genuinely creative and highly controversial, producing powerful arguments that were to puzzle his philosophical successors in the Jewish and Christian worlds. He seems to argue that there are two forms of truth, a religious form and a philosophical form, and that it does not matter if they point in different directions. He also appears to be doubtful about the possibility of personal immortality or of God’s being able to know that particular events have taken place. There is much in his work also which suggests that religion is inferior to philosophy as a means of attaining knowledge, and that the understanding of religion which ordinary believers can have is very different and impoverished when compared with that available to the philosopher.

When discussing political philosophy he advocates a leading role in the state for philosophers, and is generally disparaging of the qualities of theologians as political figures. Ibn Rushd’s philosophy is seen to be based upon a complex and original philosophy of languages which expresses his critique of the accepted methods of argument in Islamic philosophy up to his time.

1 Commentaries

Abu’l Walid Muhammad ibn Ahmad ibn Muhammad ibn Rushd, often known as Averroes (the Latinized version of his name), was born in AH 520/AD 1126 in Cordoba. He came from a distinguished line of jurists and theologians, who like him served as public officials. As a result of royal patronage he became both royal physician and *qadi* (judge) of Cordoba in succession to his father. Due to the political turmoil in Andalus (Islamic Spain) at the time, he was not always in favour, and was banished to North Africa when he was seventy during a period of persecution of philosophy. He died in AH 595/AD 1198 after having been rehabilitated, but his religious orthodoxy still seems to have been suspected by the public.

There is a famous story that when Ibn Rushd was about forty-two there was a meeting between the caliph and Ibn Rushd, at which the latter was asked to summarize the works of Aristotle in order that the ideas of that thinker might be better understood by the caliph himself, and no doubt also by the intellectual community. Ibn Rushd’s reported nervousness at accepting this commission was well-founded, since changing political circumstances had in the past - and would in the future - put Aristotle and those influenced by him under a theological cloud; the interest of a ruler in philosophy could quite easily turn into hostility. Over the next twenty-six years, however, Ibn Rushd wrote commentaries on most of Aristotle’s works. These commentaries took a variety of forms. Often he would write a summary, medium commentary and long commentary of the same text, thus presenting the ideas of Aristotle to a variety of audiences; those who were seeking a detailed discussion of the whole text would look to the long commentary, while those who wanted just to get a flavour of the original could be satisfied with the paraphrase. As Aristotle’s *Politics* was not available to him, he used Plato’s *Republic* instead for his commentary on a political text.

The remarkable feature of these commentaries is the way in which Ibn Rushd tried to get back to the original arguments of Aristotle, cleansed of the Neoplatonic accretions which had developed. This was very difficult to do, since a long and well-developed tradition of Neoplatonic commentary had very much set the agenda over the previous centuries in the Islamic world (see Neoplatonism in Islamic philosophy). However, Ibn Rushd was often able to distinguish between the points which Aristotle was trying to make and those which had been imposed upon him by the commentators. He certainly respected some of the classical commentators such as Alexander of Aphrodisias and Themistius, as well as some of the Islamic philosophers and especially his own countryman Ibn Bajja, but the style of his commentaries is to try to understand the text anew and to reconstruct the Aristotelian argument in a way which represents Aristotle’s original view. Sometimes he is more successful than at others, and

he was not averse to adding his own comments on the text when he felt this would be useful. The paraphrases are certainly a very loose summary of the originals, and often give Ibn Rushd the opportunity to express his own views on an Aristotelian theme. However, the long commentaries are very impressive analyses of the text, especially given the nature of the translations with which Ibn Rushd was working, and they came to wield great influence in the Christian and Jewish worlds (see Averroism; Averroism, Jewish.)

2 God and the world

Although Ibn Rushd did discuss theological topics in his commentaries on occasion, he usually reserved them for his more polemical works, where he has a more contemporary philosopher in mind. His *Tahafut al-tahafut (Incoherence of the Incoherence)* is a response to an earlier attack upon philosophy, the *Tahafut al-falasifa (Incoherence of the Philosophers)* written by al-Ghazali, who had argued in this work that there are two major problems with Islamic philosophy. The first problem is that it misapplies the very philosophical techniques which it advocates; that is, its arguments fall foul of the criteria for validity which philosophy itself advocates. The other problem is that the conclusions of philosophy go against the principles of Islam, which the philosophers pretend they are supporting. Al-Ghazali produced accurate descriptions of philosophical arguments and then set about demolishing them, using the same philosophical principles which his opponents try to employ. He argued that although the philosophers purport to prove that philosophy is merely a more sophisticated analysis of the nature of reality than that available to ordinary Muslims, the philosophers are in fact involved in dismantling the religious notion of God, the afterlife and creation in the guise of merely analysing these ideas. Although the object of his attack is primarily the work of al-Farabi and Ibn Sina, Ibn Rushd perceived that the whole peripatetic approach to philosophy was being challenged by al-Ghazali, and he rushed to its defence.

The nub of al-Ghazali's attack on philosophy is what he regards as its misguided interpretation of the relationship between God and the world. The Qur'an is full of references to the creation of the world and to its eventual destruction should the deity feel it appropriate, yet Islamic philosophy tends to argue that the world is eternal. If God really is an agent, al-Ghazali asks, why cannot he just create the world *ex nihilo* and then later destroy it? Ibn Rushd replies that temporal and eternal agents act very differently. We can decide to do something, we can wait for a certain time before acting, we can wonder about our future actions; but such possibilities cannot arise for God. In his case there is no gap between desire and action, nothing stands in the way of his activity; and yet we are told by al-Ghazali that God suddenly created the world. What differentiates one time from another for God? What could motivate him to create the world at one particular time as opposed to another? For us, different times are different because they have different qualitative aspects, yet before the creation of the world, when there was nothing around to characterize one time as distinct from another, there is nothing to characterize one time over another as *the* time for creation to take place (see Eternity §4).

Al-Ghazali argues that such a response is evidence of mental laziness. Even we can choose between two alternatives which appear to be identical in every respect except position. He gives the example of a hungry man being confronted by only two dates, where he is able to take just one. Since they are to all intents and purposes identical, it would seem to follow that if the philosophers are right he must just stand there and starve since there is no difference between them. Ibn Rushd criticizes this analogy, since it is not really about a choice as to which date to eat but about a choice between eating and not eating. What al-Ghazali is trying to do is establish some scope for divine action and decision-making which represents God as a real agent and not just as a cipher for natural events which would take place anyway. Ibn Rushd comments that the difficulty lies in distinguishing between the divine will and knowledge. Since an omniscient God knows exactly how the universe should be organized to produce the optimal arrangement, Ibn Rushd insists that there is no point in thinking of a gap in time existing between that conception and its instantiation. An omnipotent God does not need to wait for the appropriate moment to create the universe since nothing exists which could oblige him to wait, and he does not require time to bring about the creation. Ibn Rushd argues that given God's nature, we cannot think of his acting in any different way from that represented by the organization of the world. This does not imply a lack of freedom or ability to choose, but is merely a reflection of God's perfect nature (see Omnipotence).

Al-Ghazali followed the Ash'arites in being so concerned to emphasize the power and ubiquity of God that he refused to accept that the ordinary world really consists of stable material objects between which there are relationships of natural necessity (see Ash'ariyya and Mu'tazila §5; Causality and necessity in Islamic thought §2).

Ibn Rushd claims that this theory leads to a denial of the possibility of knowledge of the world. Since al-Ghazali accepts causality as a practical guide to our everyday lives, one might wonder what point Ibn Rushd is trying to make here. The point is that the nexus between a term and its causal properties is not merely contingent, but is really one of meaning. Al-Ghazali gives the example of a decapitated person acting just like an ordinary human being, except for the absence of a head. No such event has ever taken place, but if God wills it it could happen, because God is omnipotent and we can imagine such a possibility. If God wants to activate a headless person, he could do so. This shows that the connection between having a head and being an active human being is merely contingent, without necessity. Ibn Rushd wonders whether this change to our conceptual scheme is really possible. There are some properties which are significant aspects of the meaning of the thing of which they are the properties, and there is a necessary relationship between what a thing is and what it does. The advantage of Ibn Rushd's response is that it provides an account of how naming is possible. We can set about naming things because we can identify relatively stable entities with lawlike patterns of behaviour with other things. We may often go awry in our naming, but if we could not be sure that on the whole our names correspond with stable and fixed essences, naming itself would be an empty procedure.

Along with his insistence that the deity is a real agent, al-Ghazali was concerned to provide God with real knowledge of the everyday events of the world he created. Ibn Sina argued that God is limited to knowing only very general and abstract features of the world, since any other sort of knowledge would diminish him as an eternal and immaterial being. Al-Ghazali objects that any God which is acceptable to Islam must know the everyday events of our world. Ibn Rushd suggests that on the contrary, this would make God into someone very like his creatures and would provide him with knowledge that is beneath his dignity. God's knowledge is superior and unique because he is not limited to receiving information from the world, as is the case with finite creatures like human beings. He is the creator of the objects in the world, and he knows them in a more perfect and complete way than we can hope to attain. This suggests that God cannot know individuals as such. The best knowledge is abstract and universal, and this is the sort of knowledge which God can be thought to enjoy.

One might expect Ibn Rushd to share Ibn Sina's view that God's knowledge is limited to universal judgments, but he does not adopt this line, arguing rather that God's knowledge is neither universal nor individual, although it is more like the latter than the former. Our knowledge is the result of what God has brought about, whereas God's knowledge is produced by that which he himself has brought about, a reality which he has constructed. The organization of the universe is a reflection of God's thought, and through thinking about his own being he is at the same time thinking about the organization of the world which mirrors that essence. He cannot really be identical with contingent and accidental phenomena, yet his essence is not totally unconnected with such phenomena. They represent contingent aspects of the necessary and essential relationships which he has established. To take an example, God knows which physical laws govern the universe, but he does not need to observe any moving objects to understand the principles of movement. Such observations are only appropriate objects of knowledge of sentient creatures with sensory apparatus and are far beneath the dignity of the creator. Ibn Rushd argues that this is not to diminish God's knowledge, but rather emphasizes the distinctness of the deity from his creatures and their ways of finding things out (see God, concepts of).

3 The soul

Another charge which al-Ghazali brought against philosophy was that it fails to allow the physical resurrection of human beings and the provision of physical rewards and punishments appropriate to their behaviour during their lives. He has in mind here the Aristotelian notion of the soul, which makes the idea of an afterlife difficult to grasp. This is because the soul is the form of the living being, an aspect of the being itself, and there is no point in talking about the matter existing without the form when we are considering living creatures. Persons are combinations of soul and body, and in the absence of the latter there are no persons left (see Soul in Islamic philosophy).

Ibn Rushd appears to argue that as we become more involved with immortal and eternal knowledge, and with universal and abstract principles, our mind becomes identical to a degree with those objects of knowledge. So, once we have perfected ourselves intellectually and know everything that there is to know about the formal structure of reality, there is no longer really any 'us' around to do the knowing. Ibn Rushd regards our progress in knowledge as equivalent to a lessening of our ties with our material and individual human characteristics, with the

radical result that if anything survives death, it must be the species and not the individual. Temporal and finite creatures are destructible, but as members of a species we are permanent, although only the species itself is entirely free from destruction.

This seems even more incompatible with the traditional religious view of the afterlife than the position which al-Ghazali attacks. Ibn Rushd follows this with a political account of the function of the religious language, describing the afterlife as providing ordinary believers with a motive for virtuous action and dissuading them from immorality. He does not entirely rule out the possibility of the sort of physical afterlife on which al-Ghazali insists, but it is clear from his work that he regards such a possibility as wildly unlikely. The only meaning which can be given to such a notion is political, and there is nothing irreligious about such an interpretation, according to Ibn Rushd. It is difficult for unsophisticated believers to understand that it is worthwhile to act well and avoid evil, or that their actions have a wider reference than the immediate community of acquaintances, so any religion which is able to motivate them must address them in ways that they comprehend and in a language which strikes an emotional chord. Richly descriptive accounts of the afterlife, of God seeing everything which happens and of his creation of the world out of nothing, help adherence by the majority to the principles of religion and are the only sort of language which most members of the community can understand. The arguments which Ibn Rushd presents for hedging in the notion of the immortality of the individual soul would not mean much to the unsophisticated believer, while the more intellectually alert are expected by Ibn Rushd to understand how that notion fits in with the basic principles of Islam.

4 Moral and political philosophy

Ibn Rushd presents a firm critique of the Ash'arite theory of moral language, which interprets rightness and wrongness entirely in conformity with the commands of God. The purpose of that theory is to emphasize the power and authority of the deity over everything, even over the meaning of ethical terms (see Ethics in Islamic philosophy). What we ought to do then is simply equivalent to God's commands, and we ought to do it because God has commanded it, so that everything we need to know about moral behaviour is encapsulated in Islam. Ibn Rushd argued that on the contrary, a distinction should be drawn between moral notions and divine commands. Here he follows an Aristotelian approach. Since everything has a nature, and this nature defines its end, we as things also have natures and ends at which our behaviour is directed. The purpose of a plant is to grow and the aim of a saw is to cut, but what is the purpose of a human being? One of our ultimate aims is to be happy and to avoid actions which lead to unhappiness. It is not difficult here to align Islamic and Aristotelian principles: moral virtue leads to happiness since, if we do what we should in accordance with our nature, we will be able to achieve happiness. This happiness may be interpreted in a number of ways, either as a mixture of social and religious activities or as an entirely intellectual ideal. However, the latter is possible only for a very few, and neither religion nor philosophy would approve of it as the ultimate aim for the majority of the community. There is an essential social dimension to human happiness which makes the identification of happiness with correct moral and religious behaviour much easier to establish. It is conceivable that someone would try to live completely apart from the community to concentrate upon entirely intellectual pursuits, but this way of living is inferior to a life in which there is a concentration upon intellectual thought combined with integration within the practices of a particular society.

One might expect that a thinker such as Ibn Rushd, who was working within an Islamic context, would identify happiness and misery with some aspect of the afterlife, but as we have seen, he was unable to accept the traditional view of the afterlife as containing surviving individuals like ourselves. What the notion of the afterlife is supposed to achieve is an understanding that the scope of personal action is wider than might immediately appear to be the case. Without religious language and imagery, ordinary believers may find it difficult to grasp that our moral actions affect not only ourselves but the happiness of the whole community, not just at a particular time or in a particular place but as a species. When we behave badly we damage our own chances of human flourishing, and this affects our personal opportunities for achieving happiness and growing as people. It also affects our relationships with other people, resulting in a weakening of society. While it is possibly true that the misery consequent upon evil-doing may not follow us personally after our death, it may well follow the community. The importance of the notion of an afterlife is that it points to the wider terms of reference in which moral action has life.

In his commentary on Plato's *Republic*, Ibn Rushd modifies Plato in terms of his own Aristotelian views and applies the text to the contemporary state. He uses Plato's idea of the transformation and deterioration of the ideal state into four imperfect states to illustrate aspects of past and contemporary political organization in the Islamic world (see Political philosophy in classical Islam). He takes mischievous pleasure in comparing the theologians of his own time, the *mutakallimun*, to Plato's sophists (see Islamic theology §2). He describes the theologians as a genuine danger to the state and to the purity of Islam, and suggests to the ruler that a ban on the publicizing of their activities is advisable. In this and many of his other works, Ibn Rushd stresses the importance of a careful understanding of the relationship between religion and philosophy in the state. Revelation is superior to philosophy in that it makes its message more widely available than is possible for philosophy. The prophet can do things which the philosopher cannot, such as teaching the masses, understanding the future, establishing religious laws and contributing to the happiness of the whole of humanity. Through divine revelation or inspiration, the prophet establishes laws which make it possible for people to attain an understanding of how they should behave. The credentials of the prophet are to be established by political skill. Miracles are irrelevant here; only legislative abilities count.

The philosopher has all the theoretical knowledge which the prophet has, but only the latter can embody this knowledge in a law and persuade the general public that this is a law which must be obeyed (see Law, Islamic philosophy of). What the prophet has is practical knowledge as well as the theoretical knowledge which he shares with the philosopher, and so the content of the prophetic law (*shari'a*) is no different from the content of the philosophical law (*namus*). The prophet is much better at putting this content across to the community, and can transform abstract ideas about human happiness into political ideas and social norms which then are capable of regulating the life of the community. However, it is worth emphasizing that the only advantage which religion has over reason is that the former involves a practical form of knowledge which is not necessarily possessed by the latter. The issue of the relationship between philosophy and religion fascinated the Islamic philosophers (see Epistemology in Islamic philosophy §5), and Ibn Rushd was no exception in this respect. He tried to refine this issue time and time again throughout his works.

5 The role of philosophy

The role of the philosopher in the state was a topic of continual interest for Ibn Rushd. He noticed that Aristotle (§26) seemed to hesitate between the view that the prime constituent of the good life is intellectual thought and the alternative, based upon a broader collection of virtues. These two alternatives have very different implications, especially within the context of a religious philosophy. The identification of a more social notion of happiness as living in accordance with a general mixture of virtues would make happiness more generally available to the public, since it would mean that the unsophisticated but dutiful believer could achieve a high level of perfection in their life. The idea that intellectual excellence is the highest form of human wellbeing or happiness implies that the great majority of the community, unable or disinclined to concentrate completely on intellectual issues, is thereby deprived of the very best form of life. No religion such as Islam with its claims to universality could tolerate such a confining restriction on human happiness. Ibn Rushd thinks he can avoid this dilemma. The basis to his solution is the argument that religion and philosophy are not incompatible. Islam is a rational system of beliefs and it requires its adherents to attend to rational arguments concerning how they are to behave and think. The rational arguments are there in the Qur'an and other places for those who can follow them, and for those who cannot there are other forms of presentation of the truth which are easier to understand.

This might seem a patronizing way to describe the faith of the ordinary believer, but Ibn Rushd suggests that if we look at examples from law and medicine we shall see how acceptable it is. Lawyers may study in detail the principles behind legislation, yet most of the community just follows the law without thinking deeply about its rationale. Those who work in medical fields have a good understanding of how the body works and how different forms of treatment affect the health of the individual. The ordinary person does not understand much of this, and just goes along with what they are told by the medical experts. There is nothing wrong with this; there is no necessity for everyone in the community to be either a lawyer or a doctor. Different people have different attitudes to both the law and medicine, some based upon real understanding and some based upon casual acquaintance, and these differences do not interfere with the ability of everyone in the community to live in an organized and healthy society.

Any religion with claims to general acceptability must present its message in a suitable form for the particular audience it is addressing. Ibn Rushd argues that Islam is an especially excellent religion because it has the ability to present the important issues to the greatest variety of people. Some people will be attracted to Islam and strengthened in their faith if the philosophical arguments for being a Muslim are pursued and developed. Others, perhaps the majority, cannot really understand such arguments but can understand simpler arguments and parables which describe in simple terms what is wrong with other religions and why Islam is superior to them. Still others will not even be able to grasp such simple arguments and so must be persuaded by rhetorical devices, which include a grain of logical force but mainly consist of persuasive imagery and exhortation. The way in which Ibn Rushd makes this distinction has led some commentators to think that his real view is that philosophy alone reveals the truth, and religion is only suitable for the intellectually weak who have to be satisfied with stories and doctrines which are, strictly speaking, false. Such a disingenuous interpretation is not required, however. Ibn Rushd is trying to highlight the fact that there are a variety of ways of coming to know something, some of which are surer than others, but all of which are acceptable. Once the object of knowledge is acquired then it is known, however that knowledge has been achieved. We know religious truths in different ways, but we really do know exactly the same thing.

One of the excellences of Islam, according to Ibn Rushd, is its accessibility to a wide range of adherents. In many of his works, and especially in his *Fasl al-maqal (Decisive Treatise)*, he argues that the highest form of demonstrative reasoning cannot clash with the principles of religion. He claims here that philosophers are best able to understand properly the allegorical passages in the Qur'an on the basis of their logical training, and that there is no religious stipulation that all such passages have to be interpreted literally. Where demonstrative reasoning appears to conflict with the sense of Scripture, then those capable of demonstration (the philosophers) know that the passages must be interpreted allegorically so as to cohere with the demonstrative truths. Philosophers should be careful when they do this not to offend the religious sensibilities of the less sophisticated, in sharp contrast with the practice of the theologians. The latter frequently interpret such passages so crudely that they either throw doubt on religion itself, or threaten the pursuit of philosophy by raising doubts in people's minds concerning the orthodoxy of the conclusions reached by the philosophers. Language should be seen as a sophisticated vehicle for communicating information to different categories of audience. Religion is a means for the easy comprehension of the majority of the people, and where a hidden meaning exists it is up to the philosophers to discover it and keep it to themselves, while the rest of the community must accept the literalness of Scripture.

6 Philosophy of language

Ibn Rushd is in a difficult position when trying to respond to al-Ghazali's attacks upon philosophy, since the former tried at the same time to distance himself from the sort of Neoplatonic approach to theoretical issues which Ibn Sina advocated, and it was Ibn Sina who was the direct object of al-Ghazali's critique. One of the most significant methodological disputes between Ibn Rushd and Ibn Sina lies in their differing analyses of the relationship between essence and existence, and this has an important influence upon Ibn Rushd's approach to meaning. Ibn Sina held that a state of affairs is possible if and only if something else acts to bring it into existence, with the sole exception of the deity. Ibn Rushd characterizes this view, quite correctly, as one in which possible states of affairs are nonexistent in themselves, until their existence is brought about by some cause. The possible is that whose essence does not include its existence and so must depend upon a cause which makes its actuality necessary, but only necessary relative to that cause. In this modal system there are really only two kinds of being, that necessary through another and that necessary in itself (that is, God), so that the realm of the possible becomes identical with both the actual and the necessary.

Both Ibn Rushd and Ibn Sina maintain that there is a logical distinction between essence and existence, but the former accuses the latter of conflating the order of thought with the order of things, the logical order with the ontological order. Ibn Sina does indeed start with the logical distinction between essence and existence and then proceeds via his theory of emanation to show how existence comes to essence from the necessarily acting Necessary Being (see Neoplatonism in Islamic philosophy §3). The occasionalism of al-Ghazali is like the theory of emanation of Ibn Sina, in that both doctrines interpret the contingent world as radically dependent upon something else. The account of essence and existence provided by Ibn Sina is perfectly acceptable to al-Ghazali, with the proviso that direct divine intervention is required to bring existence to the essences. Ibn Sina divides up the world into existing things and essences, into what we can think about and what really exists, and into things

which are necessary through another and are possible in themselves. These distinctions throw doubt on the sort of realism and emphasis upon substance that is so important for Ibn Rushd and his form of Aristotelianism. This latter is based upon a model of the world as one entity, as a single order of nature with no impenetrable barriers to human understanding and investigation. This leads Ibn Rushd to argue that although a logical distinction can be drawn between the existence and essence of a thing, there is nonetheless a necessary relationship between existence and essence. Without such a relationship, one could conceive of all sorts of things happening to essences without regard to how they are actually instantiated - the sorts of thought-experiments which al-Ghazali advocates - which Ibn Rushd argues seriously misrepresents the nature of philosophy. The meaning of the name of a thing is intimately connected with the way in which it is instantiated, and it is a radical error in the philosophy of language to separate essence and existence (see Existence; Meaning in Islamic philosophy §3).

To understand Ibn Rushd's account of a variety of paths to the truth, we have to grasp his theory of meaning. He emphasizes the importance of notions such as equivocation and ambiguity in language because he thinks it is important to be able to explain how names can be used in similar ways in different contexts. Ibn Rushd agrees with Aristotle that there can be no priority or posteriority within the same genus, and so he develops an account of meaning which is based upon the *pros hen* rather than the genus-species relation. If the latter were used, meaning would come out as univocal and al-Ghazali would be entirely justified in expecting the philosophers to account for God and his activity in the same sort of language as we use to describe ourselves. If meaning is expressed in terms of *pros hen* equivocals (*bi nisba ila shay' wahid*), then we can look for some similarity in the objects which form the basis to the sharing of the name, but we do not have to insist that exactly the same name be used in its different contexts with precisely the same meaning. We can also insist that the different contexts in which a name is used have to be taken into account when we come to ask for the meaning of the name. For al-Ghazali, abstract terms have a meaning which is independent of their reference in the external world. The meaning of such terms is equivalent to the series of pictures or images in which the events they describe are characterized in particular ways. All that we have to do to conceive of God miraculously creating something out of nothing is to imagine it happening, and so it is possible. Ibn Rushd argues that, on the contrary, it is not enough to have a series of images in one's mind to establish the meaningfulness of that combination of images. A meaningful use of language is possible only through the connection of linguistic terms and ideas with a framework in which they make sense, and such a framework is connected to the varying uses of the terms and to the way in which the world is.

The concept which Ibn Rushd wants his account of language to characterize is that of a point of view. In Ibn Rushd's thought there is a continual contrast between different points of view, not just a distinction between God's point of view and the human point of view, but also a differentiation of the standpoints of the whole of humanity based upon their forms of reasoning. For example, in the *Fasl al-maqal* he talks about demonstrative, dialectical, rhetorical and sophistical people, all of whom are using similar language to discuss what is important to them, namely their faith, morality, the next life and so on. This language is not identical regardless of the way in which it is used, nor is it completely equivocal. There are connections between different applications of the same name, and these connections are strong enough for it to make sense to say that these uses are of the same name; so we can talk about there being a variety of routes to the same destination, a variety of views based upon the same ideas and beliefs, and a variety of ways of living which together add up to a morally and religiously desirable form of life.

Ibn Rushd extends the use of the notion of *ijma'* (consensus) from its theological role of establishing what is acceptable within Islam to an even more important role, that of establishing what words mean. If there is agreement in the community that particular scriptural passages are clear, then they are clear and that is the end of the matter. If it is felt by some that there is ambiguity in some passages, then there is ambiguity which has to be resolved in some way if practice is not to suffer. Those who feel that there is ambiguity have to try to resolve that ambiguity in a way which enables them to follow the route to salvation. They must do this without challenging the views of the rest of the community, since to do so would threaten the ordinary meanings of the terms which are being used. Ibn Rushd suggests that if the theologians publicize their confused thoughts about the meaning of the Qur'an, ordinary believers would doubt that they understood the texts they originally thought they knew. If doctors were to do this sort of thing, then their patients would come to think that there is no such thing as health and illness.

Ibn Rushd argues that we know from our everyday experience that there exists health and illness, and that religious texts contain important information as to how we should behave. We also have to pay attention to the different

ways in which different people relate to these facts. There exists a whole variety of different views on a particular issue, and this variety of views is represented by the variety of language which is available to describe this continuum of views, ranging from the entirely demonstrative to the purely poetic and expressive. Equivocation in language is not something to be challenged; rather it is to be accepted, since it represents a feature of our lives as different people living in a community with a whole range of ends and purposes. We should respect the different uses of the same word because they represent different points of view, different perspectives on the same thing. When Ibn Rushd tries to reconcile apparently contradictory views his approach is to argue that all these views are acceptable as different aspects of one thing. Throughout his philosophy he tries to show how it is possible for one thing to be described in a variety of ways.

See also: Aristotelianism in Islamic philosophy; Averroism; Averroism, Jewish; al-Ghazali; Ibn Sina; Islamic philosophy: transmission into Western Europe; Meaning in Islamic philosophy; Neoplatonism in Islamic philosophy

OLIVER LEAMAN

List of works

Some of Ibn Rushd's works are now only extant in Hebrew or Latin, and some not at all. The most useful bibliography is Rosemann, P. (1988) 'Ibn Rushd: A Catalogue of Editions and Scholarly Writings from 1821 onwards', *Bulletin de philosophie médiévale* 30: 153-215.

Ibn Rushd (1169-98) Commentaries on Aristotle, *Aristotelis opera… cum Averrois Cordubensis variis in eosdem commentariis*, Venice: Juntas, 1562-74; repr. Frankfurt: Minerva, 1962.(Ibn Rushd's commentaries as they appeared in Latin and formed part of the approach to Aristotle in Christian Europe.)

Ibn Rushd (*c*.1174) Middle Commentaries on Aristotle, ed. C. Butterworth, *Averroes' Middle Commentaries on Aristotle's Categories and De Interpretatione*, Princeton, NJ: Princeton University Press, 1983.(Translation and commentary on two of Ibn Rushd's major works on philosophical logic and language.)

Ibn Rushd (before 1175) Short Commentaries on Aristotle, ed. C. Butterworth, *Averroes' Three Short Commentaries on Aristotle's 'Topics', 'Rhetoric' and 'Poetics'*, Albany, NY: State University of New York Press, 1977.(A translation and commentary on three of Ibn Rushd's main discussions of different forms of language.)

Ibn Rushd (1179-80) *Fasl al-maqal (Decisive Treatise)*, ed. G. Hourani, *Averroes on the Harmony of Religion and Philosophy*, London: Luzac, 1961; repr. 1976.(Translation and discussion of the *Fasl al-maqal* and two other short pieces on the same topic.)

Ibn Rushd (1180) *Tahafut al-tahafut (The Incoherence of the Incoherence)*, ed. S. Van den Bergh, *Averroes' Tahafut al-Tahafut (The Incoherence of the Incoherence)*, London: Luzac, 1954; repr. 1978.(The standard translation of Ibn Rushd's response to al-Ghazali, incorporating the latter's text.)

Ibn Rushd (*c*.1190) Long Commentary on Aristotle's *Metaphysics*, ed. C. Genequand, *Ibn Rushd's Metaphysics*, Leiden: Brill, 1984.(A translation and commentary of Ibn Rushd's commentary on Aristotle's *Metaphysics*, Book Lambda.)

Ibn Rushd (1194) Middle Commentary on Plato's *Republic*, ed. R. Lerner, *Averroes on Plato's 'Republic'*, Ithaca, NY: Cornell University Press, 1974.(The most modern translation with extensive commentary of Ibn Rushd's commentary on Plato's *Republic*.)

References and further reading

Allard, M. (1952-4) 'Le Rationalisme d'Averroès d'après une Étude sur la création' (Averroes' Rationalism in his Study on the Creation), *Bulletin d'Études Orientales* 14: 7-59.(An account of some of the stresses between the philosophical and theological approaches to creation.)

Fakhry, M. (1958) *Islamic Occasionalism and Its Critique by Averroes and Aquinas*, London: Allen & Unwin.(The way in which Ibn Rushd's approach to the topic of causality became part of wider philosophical thought in Christian Europe.)

Hayoun, M.-R. and Libera, A. de (1991) *Averroès et l'averroïsme*, Paris: Presses Universitaires de France.(A concise but comprehensive description of Ibn Rushd's thought and its philosophical impact through the Averroist movement.)

Kogan, B. (1985) *Averroes and the Metaphysics of Creation*, Albany, NY: State University of New York Press.(Comprehensive treatment of Ibn Rushd on causal necessity, miracles, God's knowledge and emanation.)

Leaman, O. (1988) *Averroes and His Philosophy*, Oxford: Clarendon Press; 2nd edn, Richmond: Curzon, 1997.(A

general account of his philosophy.)
Leaman, O. (1994) 'Was Averroes an Averroist?', in F. Niewöhner and L. Sturlese (eds) *Averroismus im Mittelalter und in der Renaissance*, Zurich: Spur Verlag, 9-22.(A discussion of the links between the thought of Averroes and the Averroist movement.)
Leaman, O. (1995) 'Averroes', in F. Niewöhner (ed.) *Klassiker der Religionsphilosophie*, Munich: Beck, 142-62. (Concise account of the contribution of Averroes to philosophy.)
Leaman, O. (1996) 'Averroes and the West', in M. Wahba and M. Abousenna (eds) *Averroes and the Enlightenment*, New York: Prometheus, 53-67.(The links between Ibn Rushd and Averroism should be acknowledged as close, as should the role of Ibn Rushd in the growth of modernity in the West.)
Urvoy, D. (1991) *Ibn Rushd (Averroes)*, London: Routledge.(An account of his thought which lays particular emphasis upon contemporary events in Andalus.)
Urvoy, D. (1996) 'Ibn Rushd' in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 330-45.(General account of his thought, with particular attention to the context within which he was writing.)
Wolfson, H. (1961) 'The Twice-Revealed Averroes', *Speculum* 36: 373-92.(A highly influential and suggestive summary of Ibn Rushd's standing in the history of philosophy.)

Ibn Sab'in, Muhammad ibn 'Abd al-Haqq (1217-68)

Ibn Sab'in is well-known in Islamic philosophy for presenting perhaps the most radical form of Sufism. He argued that everything is really just one thing, part of the deity, and that breaking up reality into different units is to deny the nature of creation. He was hostile as a result to the attempts of the philosophers who were inspired by Aristotle to develop logic as a means to understand reality. The best way to attain the truth is the mystical path, and this is achieved by appreciating the unity of everything, not by analysing reality into separable concepts.

Ibn Sab'in came from Murcia (southeast Spain) and embraced the same type of philosophical thinking and writing as the main mystic of the twelfth century, Ibn al-'Arabi. Ibn Sab'in moved to Mecca after the year AH 642/AD 1245, where he remained until he put an end to his life. His pre-eminence in Sufi circles won him the title 'Qutb ad-Din' (the Pole of Religion). He belongs to a school of Sufism which views existence, both in its divine and worldly forms, as one indivisible unit. An important work is *Asrar al-hikma al-mashriqiyya (The Secrets of Illuminationist Philosophy)*. The most important source of Ibn Sab'in's philosophy, however, is his book *Budd al-'arif (Escape of the Gnostic)* which deals with the path to knowledge, which he wrote in AH 643/AD 1245 after moving from Spain to Morocco. In this book, Ibn Sab'in poses the fundamental question of how a Sufi can reach truth and prepare for the reception and comprehension of divine perfection. In answering these questions, he discusses the opinions of the philosophical schools that preceded him, proving the inability of those schools to reach the truth.

In his writings, Ibn Sab'in attempted to enter the realm of Sufism by way of philosophy, explaining in his presentation of the history of Islamic thought that such disciplines as philology, scholastic theology (*kalam*) and philosophy are but milestones along the Sufi's path to perfection. Thus the discipline of the Sufi who has achieved perfection is the essence of all the other disciplines. In his attempt to determine the manner in which a Sufi attains unity with God, Ibn Sab'in examines the manner in which a person achieves knowledge. By knowledge, he means the discipline of logic and the ability of such a discipline to help one to achieve divine knowledge (see Logic in Islamic philosophy). He then deals with a host of concepts with a special focus on the 'intellect' and the 'self' as being the tools for achieving knowledge and as having a primary role in bringing the Sufi closer to God. Ibn Sab'in is close to most Muslim Sufi thinkers in their initial emphasis on other disciplines that a Sufi should master, particularly logic. He differs from them, however, in his conclusions about the role of logic. Whereas most philosophers, particularly the Aristotelians, view logic as a tool which helps us to know the world and which founds the theory of knowledge (see Aristotelianism in Islamic philosophy), Ibn Sab'in views the role of logic as being random and worthless, since for him knowledge is only knowledge of God. Such knowledge is subject in its meaning to one consideration, a consideration of the internal experience based on our stimulation and sensations, sensations with standards and bases superior to the standards and methods of logic.

Since according to Ibn Sab'in the unity of existence is the criterion for understanding existence, examination of the various phenomena in existence would be incompatible with such unity and hence would mean the postulation of the idea of a God superior to and separate from his creations. In Ibn Sab'in's concept of absolute unity of existence, following that of Ibn al-'Arabi, the separation of God from his creation is inadmissible. He presents the 'intellect' and the 'self' as the means which lead us to the divine Being. In this theory, Ibn Sab'in is critical of his predecessors who viewed the mind or intellect as no more than a means for the acquisition of knowledge; he views the intellect as being of divine origin. His defence of this theory is based on the Prophetic *hadith* (tradition) that 'the first thing God created was the intellect; God then told it to come forward which it did, and then told it to go away which it also did.' Ibn Sab'in's attempt to demonstrate the possibility of the intellect contacting the divine is based on this relationship, namely that the intellect is a divine creation and as such the mind can directly and without any mediation communicate with its origin. The philosophers' attempt to reach a higher level of knowledge through making contact with the active intellect is a less ambitious attempt than that of the Sufis, who make unity with God their chief concern.

Like the intellect, the self is not merely a means by which perception takes place but should rather be perceived as a goal in itself, a goal which constitutes the knowledge that the Sufi seeks. In Ibn Sab'in's theory the self, like the intellect, is a divine creation. The path that leads to unification with God, then, is based on the discovery of the self or, more correctly, discovery of the secret which God has entrusted to us. Unification with God is thus an internal

experience which does not follow the philosophers’ accounts. For whereas the philosophers saw the path to God as a matter of proof, Ibn Sab‘in saw such a path as a matter of experience based on a particular discovery through which the truth becomes evident, so that the Sufi feels that God is closer to him than his own jugular vein.

See also: Mystical philosophy in Islam

ELSAYED M. H. OMRAN

List of works

Ibn Sab‘in (1217-68) *Budd al-‘arif (Escape of the Gnostic)*, ed. G. Kattura, Beirut: Dar al-Andalus, 1978.(Most complete summary of Ibn Sab‘in’s critique of philosophy and defence of Sufism.)

Ibn Sab‘in (1217-68) *Asrar al-hikma al-mashriqiyya (The Secrets of Illuminationist Philosophy)*.(There is no published edition of this work.)

Ibn Sab‘in (1217-68) *al-Kalam ‘ala’l masa’il al-siqliyyah (Philosophical Correspondence with the Emperor Frederick II Hohenstaufen)*, ed. S. Taltkaya, *Correspondance philosophique avec l’Empereur Frédéric II de Hohenstaufen*, Paris, 1943.(His discussions with the king of Sicily, the Emperor Frederick II, about the main principles of Aristotelianism.)

References and further reading

Taftazani, A. and Leaman, O. (1996) ‘Ibn Sab‘in’, in S.H. Nasr and O. Leaman (eds) *The History of Islamic Philosophy*, London: Routledge, 346-9.(General account of Ibn Sab‘in’s thought and times.)

Ibn Sina, Abu ‘Ali al-Husayn (980-1037)

Ibn Sina (Avicenna) is one of the foremost philosophers in the Medieval Hellenistic Islamic tradition that also includes al-Farabi and Ibn Rushd. His philosophical theory is a comprehensive, detailed and rationalistic account of the nature of God and Being, in which he finds a systematic place for the corporeal world, spirit, insight, and the varieties of logical thought including dialectic, rhetoric and poetry.

Central to Ibn Sina’s philosophy is his concept of reality and reasoning. Reason, in his scheme, can allow progress through various levels of understanding and can finally lead to God, the ultimate truth. He stresses the importance of gaining knowledge, and develops a theory of knowledge based on four faculties: sense perception, retention, imagination and estimation. Imagination has the principal role in intellection, as it can compare and construct images which give it access to universals. Again the ultimate object of knowledge is God, the pure intellect.

In metaphysics, Ibn Sina makes a distinction between essence and existence; essence considers only the nature of things, and should be considered apart from their mental and physical realization. This distinction applies to all things except God, whom Ibn Sina identifies as the first cause and therefore both essence and existence. He also argued that the soul is incorporeal and cannot be destroyed. The soul, in his view, is an agent with choice in this world between good and evil, which in turn leads to reward or punishment.

Reference has sometimes been made to Ibn Sina’s supposed mysticism, but this would appear to be based on a misreading by Western philosophers of parts of his work. As one of the most important practitioners of philosophy, Ibn Sina exercised a strong influence over both other Islamic philosophers and medieval Europe. His work was one of the main targets of al-Ghazali’s attack on Hellenistic influences in Islam. In Latin translations, his works influenced many Christian philosophers, most notably Thomas Aquinas.

1 Biography

Ibn Sina was born in AH 370/AD 980 near Bukhara in Central Asia, where his father governed a village in one of the royal estates. At thirteen, Ibn Sina began a study of medicine that resulted in ‘distinguished physicians… reading the science of medicine under [him]’ (Sirat al-shaykh al-ra’is (The Life of Ibn Sina): 27). His medical expertise brought him to the attention of the Sultan of Bukhara, Nuh ibn Mansur, whom he treated successfully; as a result he was given permission to use the sultan’s library and its rare manuscripts, allowing him to continue his research into modes of knowledge.

When the sultan died, the heir to the throne, ‘Ali ibn Shams al-Dawla, asked Ibn Sina to continue as vizier, but the philosopher was negotiating to join the forces of another son of the late king, Ala al-Dawla, and so went into hiding. During this time he composed his major philosophical treatise, *Kitab al-shifa’ (Book of Healing)*, a comprehensive account of learning that ranges from logic and mathematics to metaphysics and the afterlife. While he was writing the section on logic Ibn Sina was arrested and imprisoned, but he escaped to Isfahan, disguised as a Sufi, and joined Ala al-Dawla. While in the service of the latter he completed *al-Shifa’* and produced the *Kitab al-najat (Book of Salvation)*, an abridgment of *al-Shifa’*. He also produced at least two major works on logic: one, *al-Mantiq*, translated as *The Propositional Logic of Ibn Sina*, was a commentary on Aristotle’s *Prior Analytics* and forms part of *al-Shifa’*; the other, *al-Isharat wa-’l-tanbihat (Remarks and Admonitions)*, seems to be written in the ‘indicative mode’, where the reader must participate by working out the steps leading from the stated premises to proposed conclusions. He also produced a treatise on definitions and a summary of the theoretical sciences, together with a number of psychological, religious and other works; the latter include works on astronomy, medicine, philology and zoology, as well as poems and an allegorical work, *Hayy ibn Yaqzan (The Living Son of the Vigilant)*. His biographer also mentions numerous short works on logic and metaphysics, and a book on ‘Fair Judgment’ that was lost when his prince’s fortunes suffered a turn. Ibn Sina’s philosophical and medical work and his political involvement continued until his death.

2 Reason and reality

Ibn Sina’s autobiography parallels his allegorical work, *Hayy ibn Yaqzan*. Both clarify how it is possible for individuals by themselves to arrive at the ultimate truths about reality, being and God. The autobiography shows

how Ibn Sina more or less taught himself, although with particular kinds of help at significant moments, and proceeded through various levels of sophistication until he arrived at ultimate truths.

Such progress was possible because of Ibn Sina’s conception of reality and reasoning. He maintains that God, the principle of all existence, is pure intellect, from whom other existing things such as minds, bodies and other objects all emanate, and therefore to whom they are all necessarily related. That necessity, once it is fully understood, is rational and allows existents to be inferred from each other and, ultimately, from God. In effect, the totality of intelligibles is structured syllogistically and human knowledge consists of the mind’s reception and grasp of intelligible being. Since knowledge consists of grasping syllogistically structured intelligibles, it requires the use of reasoning to follow the relations between intelligibles. Among these intelligibles are first principles that include both concepts such as ‘the existent’, ‘the thing’ and ‘the necessary’, that make up the categories, and the truths of logic, including the first-figure syllogistics, all of which are basic, primitive and obvious. They cannot be explained further since all explanation and thought proceeds only on their basis. The rules of logic are also crucial to human development.

Ibn Sina’s stand on the fundamental nature of categorical concepts and logical forms follows central features of Aristotle’s thought in the *Prior Analytics* (see Aristotle §§4-7). Borrowing from Aristotle, he also singles out a capacity for a mental act in which the knower spontaneously hits upon the middle term of a syllogism. Since rational arguments proceed syllogistically, the ability to hit upon the middle term is the ability to move an argument forward by seeing how given premises yield appropriate conclusions. It allows the person possessing this ability to develop arguments, to recognize the inferential relations between syllogisms. Moreover, since reality is structured syllogistically, the ability to hit upon the middle term and to develop arguments is crucial to moving knowledge of reality forward.

Ibn Sina holds that it is important to gain knowledge. Grasp of the intelligibles determines the fate of the rational soul in the hereafter, and therefore is crucial to human activity. When the human intellect grasps these intelligibles it comes into contact with the Active Intellect, a level of being that emanates ultimately from God, and receives a ‘divine effluence’. People may be ordered according to their capacity for gaining knowledge, and thus by their possession and development of the capacity for hitting on the middle term. At the highest point is the prophet, who knows the intelligibles all at once, or nearly so. He has a pure rational soul and can know the intelligibles in their proper syllogistic order, including their middle terms. At the other end lies the impure person lacking in the capacity for developing arguments. Most people are in between these extremes, but they may improve their capacity for grasping the middle term by developing a balanced temperament and purity of soul (see Logic in Islamic philosophy §1).

In relation to the older debate about the respective scopes of grammar and logic, Ibn Sina argues that since logic deals with concepts that can be abstracted from sensible material, it also escapes the contingencies of the latter. Language and grammar govern sensible material and therefore have a different domain; indeed, languages are various and their rules of operation, their grasp of sensible material, are likewise articulated variously (see Language, philosophy of). Nevertheless, languages make available the abstracted concepts whose operation is governed by logic; yet if language deals with contingencies, it is not clear how it can grasp or make available the objects of logic. At times, as for example in *al-Isharat*, Ibn Sina suggests that languages generally share a structure.

3 Theory of knowledge

In his theory of knowledge, Ibn Sina identifies the mental faculties of the soul in terms of their epistemological function. As the discussion of logic in §2 has already suggested, knowledge begins with abstraction. Sense perception, being already mental, is the form of the object perceived (see Sense and reference §1). Sense perception responds to the particular with its given form and material accidents. As a mental event, being a perception of an object rather than the object itself, perception occurs in the particular. To analyse this response, classifying its formal features in abstraction from material accidents, we must both retain the images given by sensation and also manipulate them by disconnecting parts and aligning them according to their formal and other properties. However, retention and manipulation are distinct epistemological functions, and cannot depend on the same psychological faculty; therefore Ibn Sina distinguishes faculties of relation and manipulation as appropriate to those diverse epistemological functions (see Epistemology in Islamic philosophy §4).

Ibn Sina identifies the retentive faculty as 'representation' and charges the imagination with the task of reproducing and manipulating images. To conceptualize our experience and to order it according to its qualities, we must have and be able to reinvoke images of what we experienced but is now absent. For this we need sensation and representation at least; in addition, to order and classify the content of representation, we must be able to discriminate, separate out and recombine parts of images, and therefore must possess imagination and reason. To think about a black flag we must be able to analyse its colour, separating this quality from others, or its part in the image from other images, and classify it with other black things, thereby showing that the concept of black applies to all such objects and their images. Imagination carries out this manipulation, allowing us to produce images of objects we have not seen in fact out of the images of things we have experienced, and thereby also generating images for intelligibles and prophecies.

Beyond sense perception, retention and imagination, Ibn Sina locates estimation (*wahm*). This is a faculty for perceiving non-sensible 'intentions that exist in the individual sensible objects'. A sheep flees a wolf because it estimates that the animal may do it harm; this estimation is more than representation and imagination, since it includes an intention that is additional to the perceived and abstracted form and concept of the animal. Finally, there may be a faculty that retains the content of *wahm*, the meanings of images. Ibn Sina also relies on a faculty of common sense, involving awareness of the work and products of all the other faculties, which interrelates these features.

Of these faculties, imagination has a principal role in intellection. Its comparison and construction of images with given meanings gives it access to universals in that it is able to think of the universal by manipulating images (see Universals). However, Ibn Sina explains this process of grasping the universal, this emergence of the universal in the human mind, as the result of an action on the mind by the Active Intellect. This intellect is the last of ten cosmic intellects that stand below God. In other words, the manipulation of images does not by itself procure a grasp of universals so much as train the mind to think the universals when they are given to the mind by the Active Intellect. Once achieved, the processes undergone in training inform the mind so that the latter can attend directly to the Active Intellect when required. Such direct access is crucial since the soul lacks any faculty for retaining universals and therefore repeatedly needs fresh access to the Active Intellect.

As the highest point above the Active Intellect, God, the pure intellect, is also the highest object of human knowledge. All sense experience, logic and the faculties of the human soul are therefore directed at grasping the fundamental structure of reality as it emanates from that source and, through various levels of being down to the Active Intellect, becomes available to human thought through reason or, in the case of prophets, intuition. By this conception, then, there is a close relation between logic, thought, experience, the grasp of the ultimate structure of reality and an understanding of God. As the highest and purest intellect, God is the source of all the existent things in the world. The latter emanate from that pure high intellect, and they are ordered according to a necessity that we can grasp by the use of rational conceptual thought (see Neoplatonism in Islamic philosophy). These interconnections become clearer in Ibn Sina's metaphysics.

4 Metaphysics

Metaphysics examines existence as such, 'absolute existence' (*al-wujud al-mutlaq*) or existence so far as it exists. Ibn Sina relies on the one hand on the distinction in Aristotle's *Prior Analytics* between the principles basic to a scientific or mathematical grasp of the world, including the four causes, and on the other hand the subject of metaphysics, the prime or ultimate cause of all things - God. In relation to the first issue, Ibn Sina recognizes that observation of regularities in nature fails to establish their necessity. At best it evinces the existence of a relation of concomitance between events. To establish the necessity implicated in causality, we must recognize that merely accidental regularities would be unlikely to occur always, or even at all, and certainly not with the regularity that events can exhibit (see Causality and necessity in Islamic thought). Thus, we may expect that such regularities must be the necessary result of the essential properties of the objects in question.

In developing this distinction between the principles and subject of metaphysics, Ibn Sina makes another distinction between essence and existence, one that applies to everything except God. Essence and existence are distinct in that we cannot infer from the essence of something that it must exist (see Existence). Essence considers only the nature of things, and while this may be realized in particular real circumstances or as an item in the mind

with its attendant conditions, nevertheless essence can be considered for itself apart from that mental and physical realization. Essences exist in supra-human intelligences and also in the human mind. Further, if essence is distinct from existence in the way Ibn Sina is proposing, then both the existence and the non-existence of the essence may occur, and each may call for explanation.

5 The existence of God

The above distinctions enter into the central subject matter of metaphysics, that is, God and the proof of his existence. Scholars propose that the most detailed and comprehensive of Ibn Sina’s arguments for God’s existence occurs in the ‘Metaphysics’ section of *al-Shifa’* (Gutas 1988; Mamura 1962; Morewedge 1972). We know from the *Categories* of Aristotle that existence is either necessary or possible. If an existence were only possible, then we could argue that it would presuppose a necessary existence, for as a merely possible existence, it need not have existed and would need some additional factor to bring about its existence rather than its non-existence. That is, the possible existence, in order to be existent, must have been necessitated by something else. Yet that something else cannot be another merely possible existence since the latter would itself stand in need of some other necessitation in order to bring it about, or would lead to an infinite regress without explaining why the merely possible existence does exist. From this point, Ibn Sina proposes that an essential cause and its effect will coexist and cannot be part of an infinite chain; the nexus of causes and effects must have a first cause, which exists necessarily for itself: God (see God, arguments for the existence of §1).

From his proof of God’s existence, Ibn Sina goes on to explain how the world and its order emanates from God. Whereas Aristotle (§16) himself did not relate the Active Intellect that may be implied in *On the Soul* III with the first, ever-thinking cause of the universal found in Book XII of his Metaphysics, later commentators on his work (for example, Alexander of Aphrodisias) identified the two, making the Active Intellect, the principle that brings about the passage of the human intellect from possibility to actuality, into the first cause of the universe. Together with this is the proof of God’s existence that sees him not only as the prime mover but also as the first existent. God’s self-knowledge consist in an eternal act that results in or brings about a first intelligence or awareness. This first intelligence conceives or cognizes the necessity of God’s existence, the necessity of its own existence, and its own existence as possible. From these acts of conception, other existents arise: another intelligence, a celestial soul and a celestial body, respectively. The last constitutes the first sphere of the universe, and when the second intelligence engages in its own cognitive act, it constitutes the level of fixed stars as well as another level of intelligence that, in turn, produces another intelligence and another level of body. The last such intelligence that emanates from the successive acts of knowing is the Active Intellect, that produces our world. Such emanation cannot continue indefinitely; although being may proceed from intelligence, not every intelligence containing the same aspects will produce the same effects. Successive intelligences have diminished power, and the active intellect, standing tenth in the hierarchy, no longer possesses the power to emanate eternal beings.

None of these proposals by Ibn Sina give grounds for supposing that he was committed to mysticism (for an opposing view, see Mystical philosophy in Islam §1). His so called ‘Eastern philosophy’, usually understood to contain his mystical doctrines, seems to be an entirely Western invention that over the last two hundred years has been read into Ibn Sina’s work (see Gutas 1988). Nevertheless, Ibn Sina combines his Aristotelianism with a religious interest, seeking to explain prophecy as having its basis in a direct openness of the prophet’s mind to the Active Intellect, through which the middle terms of syllogisms, the syllogisms themselves and their conclusions become available without the procedure of working out proofs. Sometimes the prophet gains insight through imagination, and expresses his insight in figurative terms. It is also possible for the imagination to gain contact with the souls of the higher spheres, allowing the prophet to envisage the future in some figurative form. There may also be other varieties of prophecy.

6 The soul

In all these dealings with prophecy, knowledge and metaphysics, Ibn Sina takes it that the entity involved is the human soul. In *al-Shifa’*, he proposes that the soul must be an incorporeal substance because intellectual thoughts themselves are indivisible. Presumably he means that a coherent thought, involving concepts in some determinate order, cannot be had in parts by different intellects and still remain a single coherent thought. In order to be a coherent single unity, a coherent thought must be had by a single, unified intellect rather than, for example, one intellect having one part of the thought, another soul a separate part of the thought and yet a third intellect having a

third distinct part of the same thought. In other words, a coherent thought is indivisible and can be present as such only to an intellect that is similarly unified or indivisible. However, corporeal matter is divisible; therefore the indivisible intellect that is necessary for coherent thought cannot be corporeal. It must therefore be incorporeal, since those are the only two available possibilities.

For Ibn Sina, that the soul is incorporeal implies also that it must be immortal: the decay and destruction of the body does not affect the soul. There are basically three relations to the corporeal body that might also threaten the soul but, Ibn Sina proposes, none of these relations holds true of the incorporeal soul, which therefore must be immortal. If the body were a cause of the soul's existence, or if body and soul depended on each other necessarily for their existence, or if the soul logically depended on the body, then the destruction or decay of the body would determine the existence of the soul. However, the body is not a cause of the soul in any of the four senses of cause; both are substances, corporeal and incorporeal, and therefore as substances they must be independent of each other; and the body changes and decays as a result of its independent causes and substances, not because of changes in the soul, and therefore it does not follow that any change in the body, including death, must determine the existence of the soul. Even if the emergence of the human soul implies a role for the body, the role of this corporeal matter is only accidental.

To this explanation that the destruction of the body does not entail or cause the destruction of the soul, Ibn Sina adds an argument that the destruction of the soul cannot be caused by anything. Composite existing objects are subject to destruction; by contrast, the soul as a simple incorporeal being is not subject to destruction. Moreover, since the soul is not a compound of matter and form, it may be generated but it does not suffer the destruction that afflicts all generated things that are composed of form and matter. Similarly, even if we could identify the soul as a compound, for it to have unity that compound must itself be integrated as a unity, and the principle of this unity of the soul must be simple; and, so far as the principle involves an ontological commitment to existence, being simple and incorporeal it must therefore be indestructible (see Soul in Islamic philosophy).

7 Reward and punishment

From the indestructibility of the soul arise questions about the character of the soul, what the soul may expect in a world emanating from God, and what its position will be in the cosmic system. Since Ibn Sina maintains that souls retain their identity into immortality, we may also ask about their destiny and how this is determined. Finally, since Ibn Sina also wants to ascribe punishment and reward to such souls, he needs to explain how there may be both destiny and punishment.

The need for punishment depends on the possibility of evil, and Ibn Sina's examination maintains that moral and other evils afflict individuals rather than species. Evils are usually an accidental result of things that otherwise produce good. God produces more good than evil when he produces this sublunary world, and abandoning an overwhelmingly good practice because of a 'rare evil' would be a privation of good. For example, fire is useful and therefore good, even if it harms people on occasion (see Evil, problem of). God might have created a world in another existence that was entirely free of the evil present in this one, but that would preclude all the greater goods available in this world, despite the rare evil it also contains. Thus, God generates a world that contains good and evil and the agent, the soul, acts in this world; the rewards and punishments it gains in its existence beyond this world are the result of its choices in this world, and there can be both destiny and punishment because the world and its order are precisely what give souls a choice between good and evil.

8 Poetry, character and society

Identifying poetic language as imaginative, Ibn Sina relies on the ability of the faculty of imagination to construct images to argue that poetic language can bear a distinction between premises, argument and conclusion, and allows for a conception of poetic syllogism. Aristotle's definition of a syllogism was that if certain statements are accepted, then certain other statements must also necessarily be accepted (see Aristotle §5). To explain this syllogistic structure of poetic language, Ibn Sina first identifies poetic premises as resemblances formed by poets that produce 'an astonishing effect of distress or pleasure' (see Poetry).

The resemblances essayed by poets and the comparisons they put forward in poems, when these are striking, original and so on, produce an 'astonishing effect' or 'feeling of wonder' in the listener or reader. 'The evening of

life’ compares the spans of a day and a life, bringing the connotations of the day to explain some characteristics of a lifespan. To find this use of poetic language meaningful, the suggestion is that we need to see the comparison as the conclusion of a syllogism. A premise of this syllogism would be that days have a span that resembles or is comparable to the progression of a life. This resemblance is striking, novel and insightful, and understanding its juxtaposition of days and lives leads subjects to feel wonder or astonishment. Next, pleasure occurs in this consideration of the poetic syllogism as the basis of our imaginative assent, paralleling assent in, for example, the demonstrative syllogism: once we have accepted the premise, we are led to accept the associations and imaginative constructions that result; once we accept the comparison between days and lives, we can understand and appreciate the comparison between old age and evening. Ibn Sina also finds other parallels between poetic language and meaningful arguments, showing that pleasure in imaginative assent can be expected of other subjects; assent is therefore more than an expression of personal preferences. This validity of poetic language makes it possible for Ibn Sina to argue that beauty in poetic language has a moral value that sustains and depends on relations of justice between autonomous members of a community. In his commentary on Aristotle’s *Poetics*, however, he combines this with a claim that different kinds of poetic language will suit different kinds of characters. Comedy suits people who are base and uncouth, while tragedy attracts an audience of noble characters (see Aesthetics in Islamic philosophy).

9 Links to the West

Latin versions of some of Ibn Sina’s works began to appear in the early thirteenth century. The best known philosophical work to be translated was his *Kitab al-shifa’*, although the translation did not include the sections on mathematics or large sections of the logic. Translations made at Toledo include the *Kitab al-najat* and the *Kitab al-ilahiyat (Metaphysics)* in its entirety. Other sections on natural science were translated at Burgos and for the King of Sicily. Gerard of Cremona translated Ibn Sina’s *al-Qanun fi’l-tibb (Canon on Medicine)*. At Barcelona, another philosophical work, part of the *Kitab al-nafs (Book of the Soul)*, was translated early in the fourteenth century. His late work on logic, *al-Isharat wa-’l-tanbihat*, seems to have been translated in part and is cited in other works. His commentaries on *On the Soul* were known to Thomas Aquinas and Albert the Great, who cite them extensively in their own discussions.

These and other translations of Ibn Sina’s works made up the core of a body of literature that was available for study. By the early thirteenth century, his works were studied not only in relation to Neoplatonists such as Augustine and Duns Scotus, but were used also in study of Aristotle. Consequently, they were banned in 1210 when the synod at Paris prohibited the reading of Aristotle and of ‘summae’ and ‘commenta’ of his work. The force of the ban was local and only covered the teaching of this subject: the texts were read and taught at Toulouse in 1229. As late as the sixteenth century there were other translations of short works by Ibn Sina into Latin, for example by Andrea Alpago of Belluno (see Aristotelianism, medieval §3; Islamic philosophy: transmission into Western Europe; Translators).

See also: Aesthetics in Islamic philosophy; Aristotelianism in Islamic philosophy; Epistemology in Islamic philosophy; Logic in Islamic philosophy; Soul in Islamic philosophy; Islamic philosophy: transmission into Western Europe

SALIM KEMAL

List of works

Ibn Sina (980-1037) *Sirat al-shaykh al-ra’is (The Life of Ibn Sina)*, ed. and trans. W.E. Gohlman, Albany, NY: State University of New York Press, 1974.(The only critical edition of Ibn Sina’s autobiography, supplemented with material from a biography by his student Abu ‘Ubayd al-Juzjani. A more recent translation of the Autobiography appears in D. Gutas, *Avicenna and the Aristotelian Tradition: Introduction to Reading Avicenna’s Philosophical Works*, Leiden: Brill, 1988.)

Ibn Sina (980-1037) *al-Isharat wa-’l-tanbihat (Remarks and Admonitions)*, ed. S. Dunya, Cairo, 1960; parts translated by S.C. Inati, *Remarks and Admonitions, Part One: Logic*, Toronto, Ont.: Pontifical Institute for Mediaeval Studies, 1984, and *Ibn Sina and Mysticism, Remarks and Admonitions: Part 4*, London: Kegan Paul International, 1996.(The English translation is very useful for what it shows of the philosopher’s conception of logic, the varieties of syllogism, premises and so on.)

Ibn Sina (980-1037) *al-Qanun fi’l-tibb (Canon on Medicine)*, ed. I. a-Qashsh, Cairo, 1987.(Ibn Sina’s work on

medicine.)

Ibn Sina (980-1037) *Risalah fi sirr al-qadar (Essay on the Secret of Destiny)*, trans. G. Hourani in *Reason and Tradition in Islamic Ethics*, Cambridge: Cambridge University Press, 1985.(Provides insights into a neglected area of Ibn Sina’s thought.)

Ibn Sina (980-1037) *Danishnama-i ‘ala’i (The Book of Scientific Knowledge)*, ed. and trans. P. Morewedge, *The Metaphysics of Avicenna*, London: Routledge and Kegan Paul, 1973.(This is a translation of a metaphysical work in Persian.)

Ibn Sina (*c.*1014-20) *al-Shifa’ (Healing)*.(Ibn Sina’s major work on philosophy. He probably began to compose *al-Shifa’* in 1014, and completed it in 1020. Critical editions of the Arabic text have been published in Cairo, 1952-83, originally under the supervision of I. Madkour; some of these editions are given below.)

Ibn Sina (*c.*1014-20) *al-Mantiq (Logic)*, Part 1, *al-Madkhal (Isagōgē)*, ed. G. Anawati, M. El-Khodeiri and F. al-Ahwani, Cairo: al-Matba‘ah al-Amiriyah, 1952; trans. N. Shehaby, *The Propositional Logic of Ibn Sina*, Dordrecht: Reidel, 1973.(Volume I, Part 1 of *al-Shifa’*.)

Ibn Sina (*c.*1014-20) *al-‘Ibarah (Interpretation)*, ed. M. El-Khodeiri, Cairo: Dar al-Katib al-‘Arabi, 1970.(Volume I, Part 3 of *al-Shifa’*.)

Ibn Sina (*c.*1014-20) *al-Qiyas (Syllogism)*, ed. S. Zayed and I. Madkour, Cairo: Organisme Général des Imprimeries Gouvernementales, 1964. (Volume I, Part 4 of *al-Shifa’*.)

Ibn Sina (*c.*1014-20) *al-Burhan (Demonstration)*, ed. A.E. Affifi, Cairo: Organisme Général des Imprimeries Gouvernementales, 1956.(Volume I, Part 5 of *al-Shifa’*.)

Ibn Sina (*c.*1014-20) *al-Jadal (Dialectic)*, ed. A.F. Al-Ehwany, Cairo: Organisme Général des Imprimeries Gouvernementales, 1965.(Volume I, Part 7 of *al-Shifa’*.)

Ibn Sina (*c.*1014-20) *al-Khatabah (Rhetoric)*, ed. S. Salim, Cairo: Imprimerie Nationale, 1954.(Volume I, Part 8 of *al-Shifa’*.)

Ibn Sina (*c.*1014-20) *al-Ilahiyat (Theology)*, ed. M.Y. Moussa, S. Dunya and S. Zayed, Cairo: Organisme Général des Imprimeries Gouvernementales, 1960; ed. and trans. R.M. Savory and D.A. Agius, ‘Ibn Sina on Primary Concepts in the *Metaphysics* of *al-Shifa’*’, in *Logikos Islamikos*, Toronto, Ont.: Pontifical Institute for Mediaeval Studies, 1984; trans. G.C. Anawati, *La métaphysique du Shifa’*, Études Musulmanes 21, 27, Paris: Vrin, 1978, 1985.(This is the metaphysics of *al-Shifa’*, Volume I, Book 5.)

Ibn Sina (*c.*1014-20) *al-Nafs (The Soul)*, ed. G.C. Anawati and S. Zayed, Cairo: Organisme Général des Imprimeries Gouvernementales, 1975; ed. F. Rahman, *Avicenna’s De Anima, Being the Psychological Part of Kitab al-Shifa’*, London: Oxford University Press, 1959.(Volume I, part 6 of *al-Shifa’*.)

Ibn Sina (*c.*1014-20) *Kitab al-najat (The Book of Salvation)*, trans. F. Rahman, *Avicenna’s Psychology: An English Translation of Kitab al-Najat, Book II, Chapter VI with Historical-philosophical Notes and Textual Improvements on the Cairo Edition*, Oxford: Oxford University Press, 1952.(The pyschology of *al-Shifa’*.)

References and further reading

Alexander of Aphrodisias (*c.*200) *De anima (On the Soul)*, in *Scripta minora* 2.1, ed. I. Bruns, Berlin, 1887; ed. A.P. Fontinis, *The De Anima of Alexander of Aphrodisias*, Washington, DC: University Press of America, 1979.(Important later commentary on Aristotle.)

Davidson, H.A. (1992) *Alfarabi, Avicenna and Averroes on Intellect: Their Cosmologies, Theories of the Active Intellect, and Theories of the Human Intellect*, New York: Oxford University Press.(A thorough consideration of Ibn Sina’s theory of the intellects in relation to Hellenistic and Arabic philosophers.)

Fakhry, M. (1993) *Ethical Theories in Islam*, 2nd edn, Leiden: Brill. (Contains material on Ibn Sina’s ethical thought.)

Goodman, L. (1992) *Avicenna*, London: Routledge.(A useful introduction to central features of Ibn Sina’s philosophical theories.)

Gutas, D. (1988) *Avicenna and the Aristotelian Tradition, Introduction to Reading Avicenna’s Philosophical Works*, Leiden: Brill.(An excellent account of the considerations that entered into the construction of Ibn Sina’s *corpus*, the book contains translations of a number of smaller texts, a careful consideration of method and sharp criticisms of, among other things, ascriptions of mysticism to Ibn Sina. This is probably the most useful guide to an engagement with the philosopher’s work currently available in English.)

Inati, S. (1996) ‘Ibn Sina’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 16, 231-46.(Comprehensive guide to his analytical thought.)

Janssens, J.L. (1991) *An Annotated Bibliography on Ibn Sina (1970-1989), Including Arabic and Persian Publications and Turkish and Russian references*, Leuven: University of Leuven Press.(An indispensible tool for study of Ibn Sina and recent work on the philosopher, though it will soon need to be updated.)

Kemal, S. (1991) *The Poetics of Alfarabi and Avicenna*, Leiden: Brill.(A philosophical study of Ibn Sina’s philosophical poetics and its relation to epistemology and morality.)

Mamura, M.E. (1962) ‘Some Aspects of Avicenna’s Theory of God’s Knowledge of Particulars’, *Journal of the American Oriental Society* 82: 299-312.(This paper, along with those of Morewedge (1972) and Rahman (1958), are seminal to contemporary understanding of Ibn Sina’s thought.)

Mamura, M.E. (1980) ‘Avicenna’s Proof from Contingency for God’s Existence in the *Metaphysics* of *al Shifa’*, *Medieval Studies* 42: 337-52.(A clear exposition of the proof.)

Morewedge, P. (1972) ‘Philosophical Analysis and Ibn Sina’s "Essence-Existence" distinction’, *Journal of the American Oriental Society* 92: 425-35.(A welcome explanation of the implications of a distinction central to Ibn Sina’s proof of God’s existence.)

Nasr, S.H. (1996) ‘Ibn Sina’s Oriental Philosophy’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 17, 247-51.(Concise and interesting defence of the idea that Ibn Sina really did have distinctive system of mystical philosophy.)

Rahman, F. (1958) ‘Essence and Existence in Avicenna’, *Medieval and Renaissance Studies* 4: 1-16.(A version also appears in *Hamdard Islamicus* 4 (1): 3-14. The paper considers the philosophical usefulness of the distinction of essence from existence.)

Ibn Taymiyya, Taqi al-Din (1263-1328)

Ibn Taymiyya was a staunch defender of Sunni Islam based on strict adherence to the Qur'an and authentic sunna *(practices) of the Prophet Muhammad. He believed that these two sources contain all the religious and spiritual guidance necessary for our salvation in the hereafter. Thus he rejected the arguments and ideas of both philosophers and Sufis regarding religious knowledge, spiritual experiences and ritual practices. He believed that logic is not a reliable means of attaining religious truth and that the intellect must be subservient to revealed truth. He also came into conflict with many of his fellow Sunni scholars because of his rejection of the rigidity of the schools of jurisprudence in Islam. He believed that the four accepted schools of jurisprudence had become stagnant and sectarian, and also that they were being improperly influenced by aspects of Greek logic and thought as well as Sufi mysticism. His challenge to the leading scholars of the day was to return to an understanding of Islam in practice and in faith, based solely on the Qur'an and* sunna*.*

Ibn Taymiyya was born in Harran, Syria, and died in Damascus in AH 728/AD 1328. He lived in a time when the Islamic world was suffering from external aggression and internal strife. The crusaders had not been fully expelled from the Holy Land, and the Mongols had all but destroyed the eastern Islamic empire when they captured Baghdad in AH 656/AD 1258. In Egypt, the Mamluks had just come to power and were consolidating their hold over Syria. Within Muslim society, Sufi orders were spreading beliefs and practices not condoned by orthodox Islam, while the orthodox schools of jurisprudence were stagnant in religious thought and practice. It was in this setting of turmoil and conflict that Ibn Taymiyya formulated his views on the causes of the weakness of the Muslim nations and on the need to return to the Qur'an and *sunna* (practices) as the only means for revival.

Although Ibn Taymiyya was educated in the Hanbali school of thought, he soon reached a level of scholarship beyond the confines of that school. He was fully versed in the opinions of the four schools, which helped lead him to the conclusion that blind adherence to one school would bring a Muslim into conflict with the letter and spirit of Islamic law based on the Qur'an and *sunna*. Similarly, he had acquired a deep understanding of philosophical and mystical texts. In particular, he focused on the works of Ibn Sina and Ibn al-'Arabi as examples of philosophical and mystical deviation in Islam, respectively. Both of these trends had come to exert strong influence on Muslim scholars and lay people alike.

Ibn Taymiyya placed primary importance on revelation as the only reliable source of knowledge about God and about a person's religious duties towards him. The human intellect (*'aql*) and its powers of reason must be subservient to revelation. According to Ibn Taymiyya, the only proper use of *'aql* was to understand Islam in the way the Prophet and his companions did, and then to defend it against deviant sects. When discussing the nature of God, he argued, one must accept the descriptions found in the Qur'an and *sunna* and apply the orthodox view of not asking how (*bi-la kayf*) particular attributes exist in God. This means that one believes in all of the attributes of God mentioned in the Qur'an and *sunna* without investigating the nature of these, because the human mind is incapable of understanding the eternal God. For example, one accepts that God is mounted upon a throne above the heavens without questioning how this is possible. This same attitude is held for all of God's attributes such as his sight, his hearing or his hand.

This view is very much opposed to the philosophical view of God as First Cause and as being devoid of attributes. Thus the philosophical argument that the oneness of God precludes a multiplicity of attributes was not acceptable to Ibn Taymiyya, because God says that he is one and that he has various attributes. This denial of the attributes of God based on rationalism was adopted by the Mu'tazila (see Ash'ariyya and Mu'tazila), of whom Ibn Taymiyya was especially critical. Even the more orthodox views of the Ash'aris, who accepted seven attributes basic to God, were criticized by Ibn Taymiyya. However, he did not go so far as to declare these two groups heretical, for they deviated only in their interpretation of God's nature. But he did not spare the label of apostate for those philosophers such as al-Farabi and Ibn Sina who, in addition to the denial of God's attributes, also denied the createdness of the world and believed in the emanation of the universe from God.

Ibn Taymiyya attacked the idea of emanation not only in its philosophical but also in its mystical context, as adopted by the Sufis (see Mystical philosophy in Islam). He felt that the beliefs and practices of the Sufis were far more dangerous than were the ideas of the philosophers. The latter were a small elite group that had little direct

effect on the masses. The Sufis, however, were widespread and had a large popular following. However, Ibn Taymiyya saw a link between the ideas of the philosophers and those of the Sufis, even though apparently they had little in common.

The main tenet of Sufi thought as propounded by Ibn al-‘Arabi is the concept of the oneness of existence (*wahdat al-wujud*). Through this belief, Sufis think they are able to effect a merging of their souls with God’s essence. That is, when God reveals his truth to an individual, that person realizes that there is no difference between God and the self. Ibn Taymiyya saw a link between the Sufi belief of *wahdat al-wujud* and the philosophical concept of emanation. Although the philosopher would deny that a human soul could flow into, and thus be, the First Cause, the mystical experience of the Sufis took them beyond the realm of intellectual discourse. According to the mystic, a merging occurred but could not be expressed in rational terms. For Ibn Taymiyya, both the philosopher and the mystic were deluded, the former by reliance on a limited human intellect and the latter by excessive emotions.

Ibn Taymiyya’s argument against the Sufis is on two levels. First, there is the theological position that God has attributes and that one of these attributes is God as creator. Ibn Taymiyya believed that the Qur’an firmly establishes that God is the one who created, originated and gave form to the universe. Thus there exists a distinction between God the creator and the created beings. This is an absolute distinction with no possibility of merging. He then went on to say that those who strip God of his attributes and deny that he is the creator are just one step away from falling into the belief of *wahdat al-wujud*. This is the basis for the second part of his argument. Ibn Taymiyya believed that a Sufi is simply someone who is overcome by an outburst of emotion. For example, someone may deny God’s attributes but could then be overwhelmed by a feeling of love for God. However, the basis of that person’s knowledge is not the authentic information from the Qur’an, and so their weak intellectual foundation collapses with the onslaught of emotion. For according to Ibn Taymiyya, sense perception and emotions cannot be trusted, and the likelihood of being led astray by them is compounded when one has a basis of knowledge which is itself errant and deviant. One holds a proper belief in God and maintains a proper relationship with him, Ibn Taymiyya argued, by establishing a foundation of knowledge based on the Qur’an and authentic *sunna*.

See also: Ibn al-‘Arabi; Islamic theology; Law, Islamic philosophy of; Mystical philosophy in Islam; Neoplatonism in Islamic philosophy

JAMES PAVLIN

List of works

Ibn Taymiyya (1263-1328) *Mas’ala fi al-‘aql wa al-nafs (Concerning the Matter of the Intellect and the Soul)*, in A.A.M. Qasim and M.A.A. Qasim (eds) *Majmu‘ fatawa Shaykh al-Islam Ibn Taymiyya*, vol. 9, Riyad: Matba‘ah al-Hukama, 1996.(This is a short essay in which Ibn Taymiyya summarizes his views on the relationship between the intellect and the soul.)

Ibn Taymiyya (1263-1328) *al-‘Ubudiyya fi al-Islam (The Concept of Worship in Islam)*, Cairo: al-Matba‘ah al-Salafiyya.(This is one of Ibn Taymiyya’s most important statements concerning issues of faith and belief in Islam. He speaks extensively on matters of predestination, love for God and Sufi concepts of the annihilation of the soul.)

(1263-1328) *al-Jawab al-sahih li-man baddala din al-masah (The Correct Answer to the One Who Changed the Religion of the Messiah)*, trans. T.F. Michel, *A Muslim Theologian’s Response to Christianity*, Delmar, NY: Caravan Books, 1984.(This is an abridged translation, with an excellent introduction to Ibn Taymiyya’s polemics against various groups and an extensive bibliography.)

References and further reading

Bell, J.N. (1979) *Love Theory in Later Hanbalite Islam*, Albany, NY: State University of New York Press.(This work investigates the role of love in the thinking of Hanbali scholars and shows how they defined it in opposition to philosophers and mystics.)

Hallaq, W.B. (1993) *Ibn Taymiyya against the Greek Logicians*, Oxford: Clarendon Press.(An excellent translation of Ibn Taymiyya’s most important arguments against Greek logic. The introduction and notes give depth and perspective to this very difficult topic. It also contains an extensive bibliography.)

Izutsu Toshihiko (1965) *The Concept of Belief in Islamic Theology: A Semantic Analysis of Iman and Islam*, Yokohama: Yurindo Publishing Company.(Although this work focuses on the concept of belief in early Islam,

the author makes extensive use of Ibn Taymiyya's theories to explain how orthodox scholars came to understand this term.)

Pavlin, J. (1996) 'Sunni Kalam and Theological Controversies', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 7, 105-18.(Includes a discussion of Ibn Taymiyya's view.)

Ibn Tufayl, Abu Bakr Muhammad (before 1110-85)

Ibn Tufayl's thought can be captured in his only extant work, Hayy Ibn Yaqzan (The Living Son of the Vigilant), a philosophical treatise in a charming literary form. It relates the story of human knowledge, as it rises from a blank slate to a mystical or direct experience of God after passing through the necessary natural experiences. The focal point of the story is that human reason, unaided by society and its conventions or by religion, can achieve scientific knowledge, preparing the way to the mystical or highest form of human knowledge. The story also seeks to show that, while religious truth is the same as that of philosophy, the former is conveyed through symbols, which are suitable for the understanding of the multitude, and the latter is conveyed in its inner meanings apart from any symbolism. Since people have different capacities of understanding that require the use of different instruments, there is no point in trying to convey the truth to people except through means suitable for their understanding.

1 Life and works

Abu Bakr Muhammad Ibn 'Abd al-Malik Ibn Muhammad Ibn Muhammad Ibn Tufayl al-Qaysi is known to the West as Abubacer. It can be estimated that he was born in the first decade of the sixth century AH (twelfth century AD), based on the fact that he was in his sixties when he met Ibn Rushd in AH 564/AD 1169. Born in Wadi Ash (Guadix), a small town in Spain about sixty kilometres northeast of Granada, he died in Morocco in AH 581/AD 1185. Ibn Tufayl was the second most important Muslim philosopher in the West, the first being Ibn Bajja.

With the exception of some fragments of poetry, his only extant work is *Hayy Ibn Yaqzan (The Living Son of the Vigilant)*. The title and names of characters of this work are borrowed from two of Ibn Sina's philosophical treatises, *Hayy Ibn Yaqzan* and *Salaman and Absal*, and its framework is borrowed from an ancient eastern tale, *The Story of the Idol and of the King and His Daughter*. The title is taken from the name of the main character, Hayy Ibn Yaqzan. In the introduction and conclusion, the author addresses the reader directly; in other parts of the work, he uses a 'thin veil', a symbolic form, a story to express his philosophical views.

2 Ibn Tufayl's introduction to *Hayy Ibn Yaqzan*

In the introduction the author presents some of the views of his predecessors, al-Farabi, Ibn Sina, al-Ghazali and Ibn Bajja. Al-Farabi is strongly criticized for what is said to be his inconsistent view concerning the afterlife. No criticism of Ibn Sina is given; on the contrary, it is said that Ibn Sina's oriental wisdom will be expounded in the rest of the work. Ibn Bajja's views are said to be incomplete, mentioning the highest speculative state but not the state above it, that of 'witnessing' or mystical experience (see Ibn Bajja). While al-Ghazali's mystical experience is not in doubt (see al-Ghazali), none of his works on mystical knowledge are said to have reached the author. The introduction is intended to announce the author's intention, namely the elaboration of Ibn Sina's oriental wisdom and to show how the work differs from those of his predecessors.

3 Hayy's birth and rational progress

The story of Hayy Ibn Yaqzan takes place on an equatorial island uninhabited by human beings. There Hayy is found alone as an infant. Philosophers were of the opinion that he was born spontaneously when the mixture of elements reached an equilibrium state, making it possible for this mixture to receive a human soul from the divine world. Traditionalists believed that he was the son of a woman who chose to keep her marriage to her relative, Yaqzan, secret from her brother who ruled a neighbouring island and did not find any man qualified to marry his sister. After breastfeeding Hayy well, she put him in a box and threw it into the waters, which took him to the uninhabited island.

A deer who had just lost her son and was still experiencing the feelings of motherhood heard Hayy's cries. She suckled him, protected him from harmful things and took care of him until she died when he was seven years of age. By then he had learned to imitate other animals in speech, and he covered parts of his body with leaves after noticing that those animal parts are covered with hair or feathers. The deer's death transformed Hayy's life from one of dependency to one of exploration and discovery.

In an effort to find out the reason for the deer's death, a reason which he could not locate by observing her

appearance, he dissected her with sharp stones and dry reeds. Noticing that every bodily organ has a proper function and that the left cavity of her heart was empty, he concluded that the source of life must have been in this cavity, and must have abandoned it. He reflected on the nature of this vital thing, its link to the body, its source, the place to which it has departed, the manner of its departure and so on. He realized that it was not the body but this vital entity that was the deer and the source of its actions. With this realization he lost interest in the deer's body, which he then viewed as a mere instrument. While he could not decipher the nature of this vital thing, he observed that the shape of all deer was similar to that of his mother. From this he concluded that all deer were managed by something similar to the vital thing that managed his mother's life.

After his discovery of life, he came across a fire. He noticed that, contrary to other natural objects, which move downward, fire moves upward. This indicated to him that the essence of fire is other than that of natural things. He continued to investigate other parts of nature: animal organs, their arrangement, number, size and position, as well as the qualities that animals, plants and inanimate things have in common and those that are proper to each of them. Through continued reasoning he grasped the concepts of matter and form, cause and effect, unity and multiplicity, as well as other general concepts concerning the earth and the heavens. Concluding that the universe is one in spite of its multiple objects, he moved on to consider whether it is created or eternal. Through highly sophisticated reasoning, he found that neither the idea of creation nor that of eternity is immune to objection. Though he could not rationally decide whether the universe is created or eternal, he concluded that it must have a cause on which it remains dependent and that this cause or necessary being is non-physical and above it in essence, even if not in time.

He also concluded that the thing in him which knew this cause must also be non-physical. The more detached this non-physical thing in him was from sensory perceptions, the clearer was its vision of this cause, a vision that gave the highest joy. Even though sensations obstructed this vision, he felt obliged to imitate animals by experiencing sensations to preserve his animal soul, which would enable him to imitate the heavenly bodies. Imitating the heavenly bodies by doing things like circular movement provided him with continuous but impure vision, for attention in this type of imitation is still paid to the self.

By knowledge of the necessary being, Hayy sought to imitate this being's positive attributes; by an attempt to transcend the physical world, he sought to imitate the negative ones. Imitation of the necessary being for the sake of this being involved no attention to the self and hence provided him with pure vision. Not only was Hayy's self or essence obliterated in this state, but so also was everything other than the necessary being. No human sight, hearing or speech could grasp this state, as it lies beyond the world of nature and sense experience. Therefore no explanation of necessary being can be given, only mere signs, as Ibn Sina contends in *al-Isharat wa-'l-tanbihat (Remarks and Admonitions)*. One who seeks an explanation of this state is like one who seeks 'the taste of colours inasmuch as they are colours'. Verification requires direct experience. Using human language, which is described as an inadequate instrument, to hint at the truth Hayy is said to have witnessed in this state, the necessary being is said to pervade the universe as sunlight pervades the physical world. Trying to express the inexpressible, the author says that Hayy realized in this state that the whole is one, even though unity and multiplicity, like other contraries, exist only for sense perception. The Neoplatonic pantheistic tendency is here obvious (see Necessary being; Neoplatonism).

4 Harmony of Hayy's philosophy with revealed religion

On a neighbouring island a group of people, including the king, Salaman, practised a religion which was sound yet provided the masses with symbols, not direct truths. Absal, a friend of Salaman, observed the rituals of this religion but, contrary to others who adhered to its literal meaning, he delved into its inner truths. Being naturally inclined to solitude, which was in agreement with certain passages of the Scripture, Absal moved to the island on which Hayy lived. When he encountered Hayy he was frightened, until Hayy made it clear that he intended no harm. Absal then taught Hayy human language by pointing to objects while uttering the corresponding words.

With the acquisition of language, Hayy was able to explain to Absal his development in knowledge. At hearing this, Absal realised that what Hayy had witnessed were the realities described in his own religion: God, the angels, the holy books, prophets, afterlife and so on. When Absal discussed the truths as detailed in his religion, Hayy too found these truths in agreement with what he had come to know. However, Hayy could not understand why Absal's religion resorted to symbols and permitted indulgence in material things.

5 Suitability of religion in its outward aspect to the majority of people

Hayy expressed interest in visiting the neighbouring island to explain to its people the pure truth. Absal, who knew their nature, reluctantly accompanied him. Addressing the most intelligent group on this island, Hayy was shown respect until he tried to go beyond the literal meaning of their Scripture. The people then shunned him, distracting themselves from the truth by commercial activity. Hayy understood then that such people are incapable of grasping the direct truth and that religion is necessary for their social stability and protection. Social stability and protection, however, in no way secure happiness in the afterlife. Only preoccupation with the divine, which is rare among people of this kind, can provide such security. In contrast, the preoccupation with this world in which the majority of people indulge results in darkness or hell. While the truths of reason and revelation are the same, the majority of those adhering to the latter do so for worldly success and hence achieve eternal misery. Realizing that an attempt to enlighten those incapable of vision will only destabilize them without preparing them for happiness, Hayy asked people to continue practising their religion, warning them only against indulgence in worldly matters. Hayy and Absal returned then to the deserted island to practise their mysticism in isolation.

Ibn Tufayl ends the work by describing it as ‘containing a piece of discourse not found in a book nor heard in ordinary speech’. How is this to be understood when he had already told us in the introduction that the work is an elaboration of Ibn Sina’s oriental wisdom? Perhaps the answer can be found in Ibn Tufayl’s emphasis on the novelty of a certain ‘discourse’ or ‘speech’, not on the novelty of its content. If so, the originality of the work would seem to lie only in its form.

See also: Ibn Sina, abu ‘Ali al-Husayn; Mysticism in Islamic philosophy; Necessary being; Neoplatonism in Islamic philosophy

SHAMS C. INATI

List of works

Ibn Tufayl (before 1185) *Hayy Ibn Yaqzan (The Living Son of the Vigilant)*, ed. L. Gauthier, Beirut: Catholic Press, 1936; trans. L. Goodman, *Ibn Tufayl’s Hayy Ibn Yaqzan, a Philosophical Tale*, New York: Twain Publishers, 1972.(Ibn Tufayl’s only extant work, this book captures his main philosophical thought.)

References and further reading

Conradi, L.I. (ed.) (1996) *The World of Ibn Tufayl: Interdisciplinary Perspectives on Hayy Ibn Yaczan*, Islamic Philosophy, Theology and Sciences Series, vol. 24, Leiden: Brill.(Contains a large bibliography of works on Ibn Tufayl.)

Goodman, L. (1996) ‘Ibn Tufayl’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 22, 313-29.(Good examination of Ibn Tufayl’s life and thought.)

Hawi, S. (1973) ‘Ibn Tufayl’s Hayy Ibn Yaqzan, Its Structure, Literary Aspects and Methods’, *Islamic Culture* 47: 191-211.(Focuses on the most essential elements of the work, insisting that it is not a ‘symbolic expression’ but a ‘philosophical discourse’.)

Hawi, S. (1974a) *Islamic Naturalism and Mysticism: A Philosophical Study of Ibn Tufayl’s Hayy Yaqzan*, Leiden: Brill.(Study of *Hayy Ibn Yaqzan*.)

Hawi, S. (1974b) ‘Beyond Naturalism: A Brief Study of Ibn Tufayl’s Hayy Ibn Yaqzan’, *Journal of the Pakistan Historical Society* 22: 249-67.(Lucid and shows a good grasp of Ibn Tufayl’s view of mystical experience.)

Hawi, S. (1976) ‘Ibn Tufayl’s Appraisal of His Predecessors and Their Influence on His Thought’, *International Journal of Middle East Studies* 7: 89-121.(An attempt to show Ibn Tufayl’s originality and the influence on him of al-Farabi and al-Ghazali, rather than Ibn Sina.)

Hourani, G. (1956) ‘The Principal Subject of Ibn Tufayl’s Hayy Ibn Yaqzan’, *Journal of Near Eastern Studies* 15 (1): 40-46.(An excellent article, which reconsiders the principal subject of Ibn Tufayl’s *Hayy Ibn Yaqzan*. It rejects Gauthier’s thesis that the essential subject of the work is the harmony of religion and philosophy, arguing instead that it is the ascent of unaided human reason from elementary to mystical knowledge.)

Ibn Tzaddik, Joseph ben Jacob (d. 1149)

Joseph ibn Tzaddik was a thinker firmly within the Neoplatonic tradition of Jewish philosophy. He argued that through knowledge of our own body we understand the natural world, and through knowledge of our soul the spiritual world. He identified prophecy with philosophy and suggested that we need to employ both philosophy and the religious commandments in order to worship God. Not everyone can understand philosophy, but everyone can follow the commandments and thus approach God. Human beings will receive their deserts in the next world. The pure soul will rise to the realm of spirituality, while the evil soul will be heavy and sink into matter, never achieving repose but continually caught up in the movement of the spheres.

1 Life and works

Of Ibn Tzaddik's life we know very little. He seems to have been a well-known poet. From 1138 he was a rabbinical judge at Cordoba. He died in 1149. He was clearly influenced by Isaac Israeli and Solomon Ibn Gabirol. His major work, *al-'Alam al-Saghir*, has survived only in an anonymous Hebrew translation entitled *Ha-Olam ha-Katan (The Microcosm)*, a title indicative of the theme that unites the work. For Ibn Tzaddik follows Israeli's conception of philosophy as human self-knowledge. Man is a miniature of the world; and the world, in turn, a human being writ large. Humans are a microcosm because they have in themselves all elements of the universe.

The text is divided into four parts or discourses. The first deals with physics, the principles and constitution of the corporeal world. The second discourse, developing the idea that man is a microcosm of the larger world, deals with anthropology and psychology. Discourses three and four reflect the familiar concerns of *kalam* (dialectical theology), namely the existence, unity and attributes of God; theodicy; freedom of the will; and reward and punishment. Influenced by Saadiah Gaon, Bahya Ibn Paquda, Ibn Gabirol and Islamic *kalam* theologians, Ibn Tzaddik combines familiar Jewish Neoplatonic themes along with Aristotelian and *kalam* influences. Two requisites are necessary for philosophic understanding: knowledge of God and performance of his will. But in order to acquire this level of philosophy, one must have a knowledge of everything else as well. And self-knowledge is the key to the requisite knowledge of all things.

2 Why human beings study themselves

Part One of the text starts with knowledge of the physical world, which ultimately reflects back upon the individual knower. Humans perceive things in two ways, through sense (*ha-regesh*) and through intellect (*ha-sekhel*). The five senses yield knowledge of accidental qualities, or 'shells'; reason penetrates to the true nature or essence of a thing. Knowledge can be classified as either self-evident (immediate) or demonstrated (mediate). Correspondingly, there are four kinds of objects of knowledge: percepts of senses (*murgashot*); self-evident truths (*mefursamot*); traditions (*mekubalot*); and axioms (*muskalot*). According to Ibn Tzaddik, each of these can be traced back to either rational or sensory knowledge. The former is superior, for it distinguishes humans from animals.

The second discourse treats of ontology, and it displays the influence of both Israeli (§2) and Ibn Gabirol (§2). Matter is the foundation (*ha-yesod*) and principle of all things (see Matter). The common matter of the four elements is prime matter, which is endowed with the form of corporeality (*etzem gishmi*). Matter and form are relative to each other. Spiritual things are also composed of matter and form. In spiritual things, we may compare genus to matter, species to form, specific difference to efficient cause and individual to final cause. It is here that Ibn Gabirol's influence is most evident. Everything exists either in itself (*omed be'atzmo*) or in something else (*omed be-zulatto*). Thus matter exists in itself, whereas form exists in matter. After matter assumes a form, matter becomes an actual substance. However, matter and form can be separated only in thought, not in reality. Substance is is capable of bearing opposite and changing qualities. A substance can be the opposite of another only through its accidents. Absolute substance (*ha-etzem hamuhlat*) is pure and spiritual. It is what remains of a corporeal substance when we take everything away, and is similar to Ibn Gabirol's substance which supports the categories (see Substance). Ibn Tzaddik then turns to the corporeal world (*ha-'olam hagishmi*), namely the spheres, the four elements and the three natures. The sphere (*galgal*) differs from other bodies in matter, form and qualities. Moving in a circle, the sphere has the most perfect of motions. For spherical motion has no beginning or end. The sphere

also has knowledge of God. The four elements - fire, air, water, earth - are simple bodies and have no qualities. They can change into each other. The basis of the elements is a substance filling place as a result of its assuming the form of corporeality. The three natures - plant, animal and mineral - are composed of the four elements. The general process of the sublunar world is genesis and dissolution. Thus, the world is not permanent, for the basis of its processes is change. The human body corresponds to the corporeal world in that it too is subject to genesis and decay. It is composed of elements and has powers of growth and sustenance like plants. In true Neoplatonic fashion, Ibn Tzaddik claims that humans are superior to all other beings in that they comprise all of them. In so far as the human is a 'celestial plant', its head is directed heavenward (see Natural philosophy, medieval).

3 The activity of the soul

Part Two of *Ha-Olam ha-Katan* deals with the different types of souls. Clearly combining Platonic and Aristotelian themes in an uncritical fashion, Ibn Tzaddik distinguishes three types of soul: the vegetative, the animal, and the rational (see Plato §14). The faculty of the plant soul is appetition, and the seat of the faculty is the liver. All of its powers derive from universal powers in the upper world. The animal soul is seated in the heart; it is borne in the blood. Its functions are motion and sensation. Life is the effect of the animal soul. Death is the separation of rational soul (*ha-nefesh ha-chokhma*) from the body. Death results from an imbalance of the four humours in the heart, or by disease or injury to the brain. The rational soul is incorporeal; thus it is not located in the body. Like Aristotle (§17), Ibn Tzaddik holds that the soul is a substance, not an accident, for it is permanent - reason is essential to humans. Moreover soul is superior to body, so it must be a spiritual substance, since the higher entity, it is reasoned, could not be a mere accident while the lower is a substance. All three souls are spiritual powers. The rational soul and the intellect have a common matter. When the soul is perfected it becomes intellect; the only difference between the two is one of degree and excellence, inasmuch as the intellect comes straight from God without intermediary. The activity of the rational soul is knowing, exploring the unknown. The rational soul of one who studies is destined for the spiritual world. In order to study, a person's animal impulses must be deadened. The person then comes to know first the corporeal world, then the spiritual world, and finally the Creator. Knowledge of God is the highest kind of knowledge and the basis of human perfection. Ibn Tzaddik states that those who have no such knowledge are doomed to error. The existence of many individual souls is evidence that there must be a universal or world soul. The universal soul is received into all living bodies; just as objects receive the sun's light, so too splitting of the world soul into many souls is due to the plurality of bodies which absorb it.

4 The unity of God

How do we achieve knowledge of God? Following Bahya Ibn Paquda (§2), and contrary to the practitioners of *kalam*, Ibn Tzaddik argues that God's essence is unknowable. To know a thing, we must investigate its four causes, but with God we can only know whether he *is*, not what he is. Ibn Tzaddik offers a rudimentary proof for the existence of God as follows: if substance and accident are not eternal, something must have brought them into being. This something is God. Furthermore, since the cause of the many must be one, God is one. Ibn Tzaddik goes on to show, by means of *kalam* arguments, that there cannot be two eternal beings.

The troubling question is why God created the world at all. Clearly this is not because God lacked anything. For God is complete and needs nothing. Creation is an act of will, but Ibn Tzaddik regards God's will as eternal, not created. What then is the relation of God's will to God? As Guttman points out (1966: 131), Ibn Tzaddik's answer is ambiguous. Divine will is identical with divine essence, yet Ibn Tzaddik does not explain how this will and essence are compatible with God's immutability. According to Ibn Tzaddik, God is beyond space and time. So when we say that God's will created the world, we cannot mean either that creation is taking place now or that it took place at any definite time. Ibn Tzaddik simply declares that God created the world *ex nihilo* and that it is perfect. When he speaks of creation as a mystery, he suggests that not all men can understand the secrets of philosophy. Only a hint of these matters should be given, and intelligent individuals will comprehend of their own accord.

With respect to God's attributes, Ibn Tzaddik claims that they are different from all other attributes, in that they are all identical with his essence. The very notion of divine attributes can be applied only figuratively. The predicates assigned to God express either action or essence. But, in the end, both sorts are aspects of the same attribute. Divine attributes become models of moral action. We derive our knowledge of God from his effects, but ultimately

we can apply them only negatively.

5 Theodicy

Ibn Tzaddik's distinction between rational and traditional commandments is similar to those of Ibn Paquda (§2) and Saadiah Gaon (§2). These commandments are for our own good, so that we may be happy in the next life. Ibn Tzaddik's discussion of the four virtues (wisdom, courage, temperance and justice) is clearly influenced by Plato's four virtues. For Ibn Tzaddik, as for many of his contemporaries, knowledge of a suprasensual world is a prerequisite for eternal happiness. As recent scholars have pointed out, the identification of prophecy and philosophy causes problems because philosophy includes a number of sciences hard to envisage as having been revealed on Mount Sinai. Ibn Tzaddik attempts to resolve this problem by affirming that at the time of the Torah, God bestowed prophecy on the whole people, for such was his will; but since at the present time no one can attain philosophy, that is, prophecy, except through the mediation of science, all must successively acquire the various degrees of science. Science and the desire urging man towards God are common to all, but the aptitude for science depends essentially on climatic conditions. The good soul is knowledgeable and survives in the upper world. The bad soul loses its spirituality and revolves forever with the spheres in the world of fire. When the messiah comes, the saints will be brought back to life and never die again; the wicked souls will be rejoined to their bodies and burnt.

See also: Ibn Gabirol; Ibn Paquda; Saadiah Gaon; Soul, nature and immortality of the

TAMAR RUDAVSKY

List of works

Ibn Tzaddik, Joseph ben Jacob (before 1149) *al-'Alam al-Saghir (Ha-Olam ha-Katan, The Microcosm)*, ed. A. Jellinck, *Ha-Olam ha-Katan*, Leipzig, 1854; ed. S. Horovitz, *Sefer ha-Olam ha-Qatan*, Breslau: Druck von Th. Shatzky, 1903; trans. with notes and introduction by J. Haberman, *The Microcosm*, Doctoral Dissertation Series, no. 8676, Ann Arbor, MI: University Microfilms, 1954.(Written in Arabic, this work has survived only in an anonymous Hebrew translation. It is an eclectic Neoplatonic work with Aristotelian and *kalam* influences, written as a handbook for beginners.)

References and further reading

Guttman, J. (1966) *Philosophies of Judaism*, New York: Anchor Books.(A history of Jewish philosophy.)

Vajda, G. (1949) 'La philosophie et la théologie de Joseph ibn Zaddiq', *Archives d'histoire doctrinale et littéraire du moyen âge* 24: 93-181.(An extensive, detailed examination of Ibn Tzaddik's cosmology.)

Vajda, G. (1957) *L'Amour de Dieu dans la théologie juive du moyen âge (The Love of God in Jewish Theology in the Middle Ages)*, Paris: Librairie Philosophique, J. Vrin.(A study of how medieval Jewish philosophers and theologians, including Ibn Tzaddik, viewed the love of God.)

Weinsberg, L. (1888) 'Der Mikrokosmos: Ein angeblich im 12, Jahrhundert von dem Cordubenser Josef ibn Tzaddik verfaßtes philosophisches System', repr. in S.T. Katz (ed.) *Jewish Neoplatonism*, New York: Arno Press, 1980.(A philosophical study of *Ha-Olam ha-Katan*.)

Wolfson, H.A. (1965) 'Joseph Ibn Saddik on Divine Attributes', *Jewish Quarterly Review* 55: 277-98. (Examination of the issue of divine predication.)

Idealism

Idealism is now usually understood in philosophy as the view that mind is the most basic reality and that the physical world exists only as an appearance to or expression of mind, or as somehow mental in its inner essence. However, a philosophy which makes the physical world dependent upon mind is usually also called idealist even if it postulates some further hidden, more basic reality behind the mental and physical scenes (for example, Kant's things-in-themselves). There is also a certain tendency to restrict the term 'idealism' to systems for which what is basic is mind of a somewhat lofty nature, so that 'spiritual values' are the ultimate shapers of reality. (An older and broader use counts as idealist any view for which the physical world is somehow unreal compared with some more ultimate, not necessarily mental, reality conceived as the source of value, for example Platonic forms.)

The founding fathers of idealism in Western thought are Berkeley (theistic idealism), Kant (transcendental idealism) and Hegel (absolute idealism). Although the precise sense in which Hegel was an idealist is problematic, his influence on subsequent absolute or monistic idealism was enormous. In the US and the UK idealism, especially of the absolute kind, was the dominating philosophy of the late nineteenth and early twentieth century, receiving its most forceful expression with F.H. Bradley. It declined, without dying, under the influence of G.E. Moore and Bertrand Russell, and later of the logical positivists. Not a few philosophers believe, however, that it has a future.

1 The general case for idealism

As the term will be used here, a philosopher is an idealist if and only if they believe that the physical world exists *either* (1) only as an object for mind, *or* (2) only as a content of mind, *or* (3) only as something itself somehow mental in its true character, a disjunction we shall sum up as the thesis that the physical is derivative from mind. Particular idealists may go further and say that everything whatever is derivative from mind except mind itself, but this would not be affirmed by, for example, Kant, who believed in things-in-themselves which may be neither mental nor mind-derivative; neither, perhaps, would it be accepted by Schopenhauer, for whom Kant's things-in-themselves become an unconscious cosmic Will. Moreover, there is no one view of the status of so-called abstract objects or universals which seems required of an idealist (see Abstract objects).

The mind-dependence of the physical has been argued for and developed in widely varying ways. For example, the idealist may be a monist or a pluralist about the mind(s) from which the physical is derivative. Very significant too is the contrast between idealisms which are more ontological and those which are more epistemological in their approach. The two great exemplars of each are George Berkeley and Immanuel Kant, the founding fathers of Western idealism and sources of most subsequent arguments in its favour.

Ontological idealism affirms that a certain view of reality, in which the physical is mind-dependent, is absolutely true, and regards such elements of common sense or science as seem to conflict with this either as wrong, or as only seemingly incompatible. Epistemological idealism is concerned, rather, to show that the most acceptable views of the physical world, which doubtless include the claim that it is not mind-dependent, are, indeed, only true-for-us, but that truth-for-us is the only kind of truth it makes sense to seek. (A more qualified epistemological idealism may allow that chinks of a more absolute truth may suggest themselves and be important, but hardly belong to the main body of what we should call knowledge.) Thus, for idealism of the second kind, the mind-dependence of the physical is not so much a claim as to what is true about it, as about the sort of truth which truth about it is.

2 Berkeleian ontological idealism

According to Berkeley, there are only two types of existent - spirits (or minds) and ideas. Physical objects, as we ordinarily conceive them, are collections of sensory ideas (sense impressions). Thus an apple is simply a collection of such sensory appearances as we are immediately aware of when we say that we are perceiving it (including the sensation of eating it). As for things, or those aspects of things, which are not perceived by any finite mind, they are there *either*: in the secondary sense that they would come into our minds if we took appropriate steps (gave ourselves appropriate impressions of moving in certain ways) to have a look at them, a sniff of them or whatever; *or* they are being perceived by an infinite mind. The second alternative brings in God immediately, the first is only

explicable by saying that they are ideas which God would produce in us as a result of our taking those steps. Either way, the idealist truth that physical objects are collections of ideas, taken together with the obvious fact that everything is as though they continued their existence when unobserved by finite minds, appears to Berkeley an incontestable proof of God’s existence.

Two of the main reasons why Berkeley thought that the physical world must consist of ideas were:

1 It is only if physical objects are conceived as collections of ideas which hang together in experience that we have any empirical evidence for their existence.
2 It is generally admitted that the so-called secondary qualities of physical things only exist as ideas in our minds (see Primary-secondary distinction). Moreover, it could be proved by the way in which secondary qualities vary with the state of the observer, and the way in which they are inseparable from sensations of pleasure and pain. But the considerations which show that secondary qualities are mind-dependent show equally that the primary ones are too. (Presented shape varies with conditions of observation as much as colour.) Moreover, no one can conceive of primary qualities existing in the absence of secondary qualities, so that they can only exist tied up with the admittedly mind-dependent.

It is usual to say that Berkeley’s line of thought works only if one already accepts doctrines which he adopted uncritically from Locke (as he understood him), namely that all we ever perceive are ideas, and that secondary qualities are mind-dependent. It is, therefore, worth emphasizing that arguments of an essentially Berkeleian sort can be presented, and have been influential, which do not depend upon this Lockean inheritance.

The core of these arguments will be: physical objects, as they present themselves to our senses, do so with qualities which we cannot suppose to exist except for a perceiving mind. Indeed, we cannot even conceive them lacking all such qualities. These qualities, with which things present themselves to the senses, include what we may call all their perspectival qualities (the thing is given with features which reflect the position from which it is seen or the way in which it is felt and so on), also hedonic and aesthetic qualities, and finally an organization of the perceptual field into foreground and background, and into certain *Gestalten*. However much you try to imagine a thing as it is in itself, apart from any observer, you will find yourself imagining it as having features which represent the rough position of an observer of it, how they feel about it, and how they organize their perceptual field. In short you can only imagine it with features which it could only have as a presence to some observer. Such reasoning continues to persuade those of a Berkeleian cast of mind that one cannot form any genuine conception of a physical world existing except as an object for an observer.

All this is likely to invite two objections. First, it may be said that you should distinguish between the representation (such as an image in your mind) and what that representation represents for you. Only certain features of the image serve a representative function. Now the fact that the image may have some of the features which an actual sense impression of it would have only if the thing were perceived in a certain way, does not mean that these features must be regarded as belonging to what is represented. To this it may be replied by the idealist that they do not deny that, by ignoring certain features of the image, you can regard only the others as playing a role in picturing the object; and that these need not include those which obviously imply presence to a subject. What they deny, in contrast, is that one can form any sort of representation which will, so-to-speak, positively depict the thing as existing without subject-implying features. And unless one can do this one has no real sense of what an unperceived thing could be like.

The second objection is that one can conceive what one cannot imagine. Surely you can *conceive* a physical thing without these subject-implying features even if you cannot *imagine* it. To this the idealist may reply that you do not really understand what you are thinking if you only think about it in words (and doubtless this is what the objector means by conceiving it). Really to bring before your mind the character of the situation you believe in requires that, using the expression broadly, you must *imagine* it, and this you cannot do except by imagining it as it would present itself to a certain observer.

Such a line of thought, though not precisely Berkeley’s in detail, is Berkeleian in spirit and inspiration and it is likely to be a main plank of an ontological idealism which claims that unperceived physical reality is an impossibility. What positive view of the world can be based upon such reflections? For Berkeley it showed that there must be a God who is responsible for those ideas which (after acting in a certain way) we have no choice but

to experience and who keeps the whole system of ideas available to each individual spirit in conformity with a universal system of laws determining the appearances available to each.

However, there have been philosophers who put forward a phenomenalism supposed not to imply the existence of God. According to them, one can speak meaningfully of physical things as existing unperceived. However, these only exist in a secondary sense as compared with those which are actually perceived, and their existence in this secondary sense is only the fact that they are available for perception. That is, there are definite facts for each of us (according to what we would ordinarily call our position in space) which determine what perceptions are available or compulsory for us in response to what we do or suffer (what sensations of movement we give ourselves or are given). There is no need to suppose that there is some explanation for this; it must just be accepted as a brute fact.

This phenomenalism is often not classed as idealist because its reductive account of the physical is divorced not only from theism, but from any other conception of the world as shaped by Reason or other higher forms of Mind. It is a puzzle of intellectual history that some of those most influenced by Berkeley's views of physical reality have been among the most atheistic and, in the popular sense, most 'materialist' of thinkers (for example, T.H. Huxley, A.J. Ayer and, with qualifications, J.S. Mill).

3 Kantian transcendental idealism

A simple version of Kant might present him as a phenomenalist who supplemented his phenomenalism with the admission that there must be some explanation of why the sense experiences available to us are what they are, yet who regarded this explanation as unavailable to us except as the thesis that they result from unconscious operations which we (as we really are rather than as we appear to ourselves) conduct upon things-in-themselves of whose real character we can know nothing (except that it cannot be that of anything properly called physical). However, Kant's reasoning for his transcendental idealism is largely different from those deriving from, or inspired by, Berkeley (see Kant, I. §5).

For Kant there are two striking facts about our knowledge of the world which only his transcendental idealism can explain ('transcendental' means 'having to do with our cognitive powers'). First, we have a great deal of 'synthetic a priori' knowledge about it (see A priori; Analyticity). Thus we know that arithmetic and the axioms of Euclid apply to the physical world as a whole, that every physical or mental process occurs in conformity with universal causal laws, and that change requires a permanent substratum of matter which remains quantitatively the same. Second, neither a priori nor empirical knowledge can answer the great questions of human destiny, such as whether God exists and whether we are immortal. The only possible explanation of our synthetic a priori knowledge about the physical and, indeed, mental worlds is that it is really our knowledge of our own cognitive nature. Space and time are the forms of our perceptual intuition and the categories of causation, substance and accident and so on are the categories by which we construct the unitary world of our actual and possible experience out of unconscious stimuli which reach the hidden self from things-in-themselves (or 'noumena'), of whose character we must remain ignorant. And it is because we are ignorant of things-in-themselves that we cannot *know* the answers to the questions about God and immortality, for these concern absolute truth rather than that truth-for-us which is all that is available for knowledge. On the other hand, just because we cannot *know* the answers to these questions, we may have *faith* that they would suit our moral natures and show that, in spite of the causal determinism holding in the phenomenal world, we are responsible at some 'noumenal' level for our own adherence or otherwise to the categorical imperatives of morality.

Some of the details of Kant's theory are outmoded by the fact that science seems no longer committed to some of his supposedly synthetic a priori truths such as the axioms of Euclid and the universality of causation. However, the idea that the world as we know it owes, to an incalculable extent, its general character to our particular modes of perception and thought still has great force. In Berkeley there was no suggestion that what we know is created by our knowledge of it. The ideas which constitute the physical world are simply the ones which God has chosen to give himself and us and to organize in a certain way. Our knowledge of those we perceive is a fully accurate knowledge by direct acquaintance and the existence and character of others, as actualities or possibilities, is known by induction. In Kant, knowledge itself to a great extent creates its objects by unconscious operations upon unconscious stimuli reaching us from things-in-themselves whose real nature it leaves in darkness.

The distinction is somewhat subtle, since both the Berkeleian and Kantian, in effect, regard facts about the physical world as facts about the perceptions we may obtain through sensations of movement in certain directions. However, the Berkeleian inheritance has mainly been to insist on the way in which the physical world cannot be conceived without sensory qualities which can only occur as contents of experience, while the Kantian inheritance has mainly been to insist on the way in which our cognition of the physical world interprets it by concepts which it *brings to* experience rather than *abstracts from* it.

In fact, Kant's position is nearer to Berkeley's than he himself allowed. According to Kant his idealism is transcendental, whereas Berkeley's is empirical. What this comes to is that Berkeley's idealism professes to give the absolute truth about the physical world, as a corrective to a realism which regarded it as existing independently of mind, while Kant accepted such realism, but claimed that it was only true for us, and that, as for the absolute truth about things which underlie it, we know nothing beyond the mere fact that there must be such an absolute truth (in the moral and theistic significance of which we may have faith).

4 German absolute idealism

The great figures in German absolute idealism were J.G. Fichte, G.W.F. Hegel and F.W.J. von Schelling (see Absolute, the). (The character of their considerable political influence cannot be considered here.) In effect, each agreed with Kant that ordinary common sense and 'scientific' (in our usual sense, not theirs, in which it referred to their own philosophic conclusions) truth about the physical world is only truth for us. But they went beyond Kant in holding that philosophy can put this in the context of an absolute and rationally demonstrated truth about an essentially spiritual world. In fact, Kant's attempt to close the door on attempts to know the ultimate truth of things opened it to some of the most robust claims ever made to have probed the mysteries of the universe.

For Kant the physical world only exists for us, and our knowledge of it is only truth for us. However, we can recognize that there must be two hidden determinants of it, modes of cognition which take place in our own hidden depths, and the unconscious non-physical stimuli from mysterious things-in-themselves out of which they make the familiar physical world. Fichte thought the postulation of such things-in-themselves quite unnecessary. If the knowable physical world is something whose form we construct unconsciously why should not the matter be something we determine unconsciously too? Thereby we avoid the nebulous hypothesis of things-in-themselves, and are left simply with our own indubitable existence and hidden depths thereof, of which we dimly sense the presence. Of course there is an external world or non-ego, but it exists only as something which the ego posits and does so for reasons the general character of which can be deciphered. For the ego wants to live a life of moral worth and this it can only do if it has obstacles to overcome; thus the external world it posits consists precisely in those obstacles whose over-coming is most morally valuable at its current stage of development.

But how is it that each ego shares a non-ego, as it evidently does, with other egos? Fichte has two related answers. One is that moral development is something which can only occur in a community, so that the different egos need to posit a shared non-ego giving them a common environment in which to work out their moral destiny. Second, as his thought developed, Fichte became clearer that the ego which is working out its moral destiny in each of us is really a single world-spirit living out an apparent multiplicity of lives. Fichte developed this account by way of a dialectical method which became the hallmark of German idealism (inspired by Kant for whom, however, it was rather a source of illusion than a means to truth) in which apparently opposed truths are successively reconciled in higher syntheses until absolute truth is reached. Thus was born absolute idealism, in which the reality behind both nature and finite mind is a single absolute mind or self in process of self-discovery or development. However, Fichte's brand of absolute idealism is sometimes also called 'subjective idealism', because it regards the natural world as existing only for the subjective experience of finite individuals, expressions of a single world self though they may be.

Schelling was originally a follower of Fichte, but his continually shifting versions of idealism tended to become more 'objective' or at least more positively concerned with nature for its own sake. The Absolute or universal self does not simply dream the physical world as the scene of moral endeavour but rather expresses itself in a parallel dialectic, both 'really' in the nature from which the mind arises and 'ideally' in the mind for which nature exists. The two come together eventually in philosophical understanding and, more concretely, in art.

Absolute idealism, and the dialectical method and ontology, reached its historically most important form in the

philosophy of Hegel. For Hegel, the world consists in a series of terms each surpassing its (only sometimes temporal) predecessors by incorporating what was satisfactory in them, in a manner which reconciles in a higher synthesis that in which they contradicted each other. The series begins with pure concepts, leading on to actual natural and then humanly historical processes and terminates in a community in the free service of which each individual can find themselves fulfilled and in the consciousness, in the minds of philosophers, of its total nature. Thus, everything exists as path to, and as fodder for, a rich communal spiritual life, but how far this means that nothing really exists except as a component within or object for consciousness or spirit, is controversial. Therefore, it is unclear how far Hegel was an idealist in our sense (as opposed to the broader sense mentioned parenthetically above).

Hegel and Schelling had originally seen themselves as partners in developing a new philosophy, but Hegel soon surpassed his at first better known associate in fame and influence. However, Schelling had his turn again on Hegel's death, developing a new so-called positive philosophy in which he rejected the high a priori road to the nature of existence which both thinkers had taken previously. Absolute idealism must appeal partly to empirical features of the world rather than merely cite them sometimes as illustrations of what reason can independently prove must be so. This more traditionally Christian philosophy sought to give God and man a freedom effectively denied them by Hegelianism. Its somewhat bizarre ontology also seemed to many to show German idealism in its death throes.

In different ways Fichte, Schelling and Hegel each held that the world could only be understood through realizing that it is the concrete actualization of concepts whose proper home is in the mind. This binds them to Kant, but they sought to go beyond him in explaining why the relevant categories are just as they are and why there is a real unity of experience common to apparently different minds: namely that, in the end, the world is the construction of one universal Mind or Reason. Each saw himself as drawing on Spinoza as well as Kant, but as substituting an ultimate self or subject for Spinoza's substance.

Standing quite apart from these absolute idealists is the lonely but immensely influential figure of Arthur Schopenhauer. Arguably the closest metaphysically, if not in mood, to Kant, and accepting in the main his transcendental idealism, he claimed to have discovered the true nature of the realm of things-in-themselves, regarding them as aspects of a single universal Will, manifesting itself as object for a subject (which was its own self fallen into a state of wretched self-assertion), from which it can escape only by a culmination of that denial of the will to live, characteristic, as he saw it, of sainthood.

5 Anglo-American absolute idealism

As German philosophers moved away from idealism in the later part of the nineteenth century, idealism of an essentially absolute kind became the dominant mode of philosophy in the UK and the USA (where, however, there were more serious rivals to it). This was motivated partly by the search for a form of religious belief which would be less vulnerable to Lyell and Darwin than traditional Christianity had been, and by an ethical viewpoint which would be rather nobler in its conception of the possibilities of human life than Benthamite utilitarianism. Some of these philosophers (for example John and Edward Caird, and William Wallace), were doctrinal Hegelians, utilizing Hegelianism to save Christianity.

More importantly original philosophers of an idealist persuasion during this period were T.H. Green and F.H. Bradley in the UK (also the very like-minded, though more Hegelian, Bernard Bosanquet), and Josiah Royce in the USA. We can only mention in passing the very distinctive idealism already advanced by J.F. Ferrier in Scotland, which draws both on Berkeley and on German idealism. These thinkers were to various extents influenced by Kant and Hegel and the other German idealists, but in the case of Bradley, at least, something of the Berkeleian tradition is, perhaps unconsciously, present.

Green was anxious above all to show that the development of human life from animal origins could not be explained purely by way of natural selection, or indeed in any naturalistic way. Rather, must it be recognized as the gradual unfolding of the life of a universal spirit aspiring to fulfilment in an eventually virtuous form of human life. For empiricism and naturalism cannot explain the connectedness of the world, and the ability of the human mind to synthesize events of different times into a unitary history. This is only possible if the world is the expression of a single universal spirit of which each of us is an actualization in which it becomes aware of itself.

The general upshot is quite Hegelian, but there is little use of Hegelian dialectic.

Bradley's metaphysics derives from two main reflections: first, that nothing is genuinely conceivable except experience with its various modes and contents; second, that what we describe as distinct things in relation to each other can only be adequately conceived as abstractions from a higher unity. In the end all things must, therefore, be abstractions from one single Cosmic Experience. With his denial of time's reality and his claim that Reality is really a single cosmic *Nunc Stans* whose ingredients only seem to be passing away in time, Bradley strikes a note which is perhaps more Platonic than Hegelian. Royce's absolute idealism has a good deal in common with Bradley, but whereas for Bradley God was only a rather superior 'appearance' along with the ordinary things of daily life, for Royce the Absolute was God, being personal in a way that Bradley's Absolute was not.

6 Panpsychism

One of the main charges against idealism is that of 'cosmic impiety' (Santayana). Its tendency is to make the vast realm of nature simply a representation in a mind observing or thinking of it. This can hardly do justice either to its obstinacy (surely not primarily of our own making, whatever Fichte may have thought) or to its wonderfulness. Such reflections have led some of those who are persuaded of the basic idealist claim that unexperienced reality is impossible, to hold the panpsychist position that nature is composed of units which feel their own existence and relation to other things, just as truly, if less articulately, as we do (see Panpsychism). This was the view of Royce, and Bradley thought it might be true. It was a main plank, somewhat eccentrically developed, of the German idealist Gustav Fechner (and is perhaps adumbrated in Schelling); also of Leibniz, who in this respect can be called an idealist.

Panpsychism of this sort has been most fully developed in recent times in the work of A.N. Whitehead and of Charles Hartshorne. It is sometimes regarded as a synthesis of realism and idealism; realist because it gives the ultimate units of nature (whatever they are) a reality in themselves (as what they are for themselves); idealist because it denies unexperienced reality. When the inner sentient life of (the rest of) nature is thought of as unified with the subjective life of humans and animals (as it must be for a Bradley or a Royce) in one absolute consciousness, we have a form of absolute or objective idealism which quite avoids the anthropocentric character it had in the work of thinkers such as Fichte.

7 Personal idealism

Many thinkers of an idealist persuasion in the English-speaking world bridled somewhat at the downplaying of individual persons by absolute idealism, especially Bradley and to a lesser extent Royce. This led to the development, as the nineteenth century closed, of some forms of personal idealism for which reality is a community of independently real spirits (with or without a God as a *primus inter pares*) and the physical world their common object or construction. There is no great figure here, with the possible exception of J.M.E. McTaggart who espoused a highly individual form of pluralistic idealism. Otherwise the main proponent of personal idealism was the US philosopher, G.H. Howison, although eight Oxford philosophers published a manifesto under this label in 1902 (see Personalism).

Anglo-American idealism was, for a time, widely thought to have been refuted by the work of G.E. Moore and Bertrand Russell in the UK and such pragmatists as James and Dewey in the USA (though there were certainly idealist features to the thought of these two Americans), but a contrary judgement is now not uncommon. Edmund Husserl's phenomenology remains influential in some quarters, and some agree with his eventual view that it implies a form of transcendental idealism (see Husserl, E.). Some regard the antirealism associated with Michael Dummett as idealist in spirit (see Realism and antirealism), while some of the continuing school of Wittgensteinians regard the thought of Wittgenstein as a form of social idealism. Much closer to traditional idealism, however, is the conceptual idealism of the important US philosopher Nicholas Rescher (which synthesizes idealism and pragmatism) and idealist positions (not, it must be admitted, so far very influential) advocated in the UK by John Foster and, if he may say so, by the author of this entry.

T.L.S. SPRIGGE

References and further reading

Berkeley, G. (1710) *A Treatise Concerning the Principles of Human Knowledge*, in *The Works of George*

Berkeley, vol. 2, ed. T.E. Jessop, London: Thomas Nelson & Sons Ltd, 1949.(This and the following work give the complete statement of Berkeley’s theistic idealism. Berkeley’s introduction is essential reading.)

Berkeley, G. (1713) *Three Dialogues between Hylas and Philonous*, in *The Works of George Berkeley*, vol. 2, ed. T.E. Jessop, London: Thomas Nelson & Sons Ltd, 1949.(The second of Berkeley’s two most important works. Many other satisfactory editions of these works exist.)

Bradley, F.H. (1897) *Appearance and Reality*, Oxford: Oxford University Press, 1930.(This work represents the high point of British idealism in the nineteenth century.)

Ferrier, J.F. (1854) *The Institutes of Metaphysics*, Edinburgh.(A fine, much-neglected work, presenting an idealism that lies somewhere between that of Berkeley and the German Idealists.)

Fichte, J.G. (1794, 1797) *The Science of Knowledge*, trans. P. Heath and J. Lachs, Cambridge: Cambridge University Press, 1982.(Develops Fichte’s conception of the world as consisting of self and not-self, the latter a posit of the former as the scene for its ethical self-development.)

Findlay, J.N. (1970) *Ascent to the Absolute*, London: George Allen & Unwin.(Wide-ranging essays by an idealistically inclined philosopher.)

Foster, J. (1982) *The Case for Idealism*, London: Routledge & Kegan Paul.(A difficult but important argument for an idealist conception of the physical world.)

Green, T.H. (1883) *Prolegomena to Ethics*, Oxford: Oxford University Press.(Argues that a naturalistic, in particular Darwinian, account of the origin of mind in the universe can never explain how it can have anything beyond the momentary as its object. This ability shows that it is a stage in the self-realization of an infinite mind.)

Hartshorne, C. (1962) *The Logic of Perfection*, Peru, IL: Open Court.(A set of essays sufficient for a general grasp of Hartshorne’s process philosophy which he regards as a form of idealism.)

Hegel, G.W.F. (1807) *Phenomenology of Spirit*, trans. A.V. Miller, Oxford: Oxford University Press, 1977. (Difficult like all Hegel’s works, this charts the stages by which spirit or consciousness comes to consciousness of itself as the ultimate reality.)

Howison, G.H. (1901) *The Limits of Evolution, and Other Essays Illustrating the Metaphysical Theory of Personal Idealism*, New York: Macmillan.(Howison’s personal idealism is presented as religiously more satisfactory than the then-still-dominant philosophy of absolute idealism.)

Husserl, E. (1913) *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, trans. F. Kersten, Dordrecht: Kluwer, 1982.(The first book is the most relevant. Shows that phenomenology is a route to transcendental idealism.)

Kant, I. (1781) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1933. (The classic statement of Kant’s transcendental idealism, one of the greatest and most influential works in the history of philosophy.)

McTaggart, J.M.E. (1921, 1927) *The Nature of Existence*, Cambridge: Cambridge University Press, 2 vols.(The closely argued main statement of his doctrine of the Absolute as a society of spirits linked to each other by love.)

Schopenhauer, A. (1818) *The World as Will and Representation*, trans. E.F.J. Payne, New York: Dover Publications, 1969, 2 vols.(The main work in which Schopenhauer develops his view that the natural world is the way in which the single World Will appears to itself at the level at which it actualizes itself as human consciousness.)

Royce, J. (1885) *The Religious Aspect of Philosophy*, Gloucester, MA: Peter Smith, 1965. (Royce’s first work, perhaps the best statement of his form of absolute idealism.)

Royce, J. (1919) *Lectures on Modern Idealism*, New Haven, CT and London: Yale University Press, 1964. (Lectures delivered 1906: a perfect introduction to German Idealism.)

Schelling, F.W.J. (1797) *System of Transcendental Idealism*, trans. P. Heath, Charlottesville, VA: University Press of Virginia, 1978.(The classic statement of Schelling’s idealism in one of its earlier phases.)

Sprigge, T.L.S. (1983) *The Vindication of Absolute Idealism*, Edinburgh: Edinburgh University Press.(Argues that the inner being of nature is experiential and that all experience is united in an absolute world consciousness.)

Sturt, H. (ed.) (1902) *Personal Idealism: Philosophical Essays by Eight Members of the University of Oxford*, London: Macmillan.(Personal idealist manifesto against the submerging of the individual by absolute idealism. See especially the contributions of F.C.S. Schiller and Hastings Rashdall.)

Whitehead, A.N. (1929) *Process and Reality*, New York: The Free Press, 1978.(Classic statement of Whitehead’s process philosophy which remains one of the great alternative conceptions of how things really are and purportedly combines what is true in realism and idealism.)

Idealizations

Scientific analyses of particular phenomena are invariably simplified or idealized. The universe does not contain only two bodies as assumed in Newton's derivation of Kepler's laws, or only one body as assumed in Schwarzschild's relativistic update; real economic agents do not act exclusively to maximize expected utilities, the surfaces of ordinary plate condensers are not infinitely extended planes, and the sine of an angle is not equal in measure to the angle itself. There are many reasons for the use of such misdescriptions. First and foremost is the need to achieve mathematical tractability. Science gets nowhere unless numbers, or numerical constraints, are produced that can form the basis of predictions and explanations. Idealizations may also be required because of the unavailability of certain data or because of the absence of necessary auxiliary theories.

The philosophical problem is to make normative sense of this common but complex scientific practice. For example, how can theories be tested given that they connect to the world only through the intermediary of idealized descriptions? In what sense can there be scientific explanations if what is to be explained must be misdescribed before theory can be brought to bear? The fact that idealizations can often be improved, with corresponding salutary effect on the accuracy of prediction or usefulness of explanation, suggests that idealizations should be understood as part of some sort of convergent process.

1 The falsity of idealizations

Idealizations are false. This much is clear. And not only false, but such that their use, by and large, must lead to incorrect observational consequences. In order to generate a numerical prediction, Michelson and Morley described their interferometer as being perfectly rigid and the optical media as being stable, even though they knew this not to be true. The example is not an isolated exception. This causes serious problems for conventional hypothetico-deductive and Bayesian views of theory confirmation (see Confirmation theory). Let *t* represent some underlying or fundamental theory, *i* the idealizing assumptions made, and *p* some actually derivable prediction. Philosophical and scientific common sense has it that if *p* is found to be true there is confirmation, or at least the satisfaction of a necessary condition for confirmation, and if *p* is found to be false, disconfirmation. But the falsity of *i* blocks such inferences. Consider first disconfirmation. From the premises $t\&i \Rightarrow p$ and $\sim p$ all that follows is that $\sim (t\&i)$, which is equivalent to $\sim t \vee \sim i$ (that is, not *t* and/or not *i*). But this conclusion follows directly from the falsity of *i*. Therefore, there is no need to engage in the expense of experimentation if the conclusion sought is simply that either theory or idealizations are false. A similar problem holds for confirmation.

Bayesian accounts fare no better. Distinguishing, as above, between the underlying theory and the idealizations needed for an actual calculation, Bayes' theorem takes the form,

$$\mathrm{P}(t \;\&\; i \mid e) = \frac{\mathrm{P}(e|t \;\&\; i)\mathrm{P}(t \;\&\; i)}{\mathrm{P}(e)}$$

where *e* represents the empirical evidence. Because the idealizations are false, $\mathrm{P}(t \;\&\; i) = 0$. Therefore, the evidence cannot affect the probability of $(t \;\&\; i)$. This is just the Bayesian analogue of the problem caused by the use of idealizations for hypothetico-deductivism. Consider also the typical case in science where the theory-produced prediction is false because of the distortion introduced by the idealizations. In such a case $\mathrm{P}(e \mid t\&i)$ will be zero. So once again, there will be no change in the probability of $(t \;\&\; i)$. Trying to avoid such difficulties by separating *i* from *t* and conjoining it with *e* leads to similar disappointments for Bayesians.

2 Idealizations as approximations

One popular approach to idealizations is to view their use as a form of approximation. The basic idea is to think of individual idealizations as elements of some larger structure that is convergent on the truth. This way of thinking is naturally inspired by the existence in science of sequences, such as the following, of increasingly less idealized descriptive equations.

$$to\mathrm{ml}^2 \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = -mgl\theta(1) ml^2 \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = -mgl\sin\theta(2) ml^2 \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = -cl\frac{\mathrm{d}\theta}{\mathrm{d}t} - mgl\sin\theta(3)$$

These equations apply to simple pendulums where *m* is the mass of the pendulum bob, *l* the length of the suspension cord, *g* the gravitational strength, *c* the coefficient of resistance, and the angular displacement. A

natural ordering exists insofar as (3) approaches (2) as the coefficient of resistance is made to approach zero, and (2) reduces to (1) if $\sin\theta$ is replaced by . Focusing first on the reduction of (2) to (1), the difference between $\sin\theta$ and will be ‘small’ if is itself ‘small’. For example, if is 0.0873 radian, then $\sin\theta$ is 0.0872. And in general, as approaches zero the difference between $\sin\theta$ and also approaches zero. The hope therefore is that if is kept small, the difference in the solution to (1) will differ from that of (2) by a correspondingly small amount. And because this hope is satisfied, it can be said that approximates $\sin\theta$. The relationship between (2) and (3) is similar: if the coefficient of resistance *c* is small, then so too, one hopes, will be the difference between calculated periods. What is illustrated in these equations is a transformational sense of approximation: the substitution of some term or expression X' for *X* in a mathematical function or equation, where the effect of the transformation will be correspondingly small if the difference between *X* and X' is small. And if the transformations are made in the appropriate order, a sequence such as that illustrated is formed of increasingly less idealized and more realistic descriptive equations.

The above approach distinguishes between the process of idealization, the misdescription of things, and the hope that this misdescription will yield a successful transformational approximation. In the case of equation (2), the natural inclination is to say that a causal factor has been ignored or misdescribed, namely, the resistance. Examples such as this suggest that idealization involves the deliberate ignoring or misdescription of causally relevant factors. Whether or not such misdescriptions can be interpreted as yielding transformational approximations becomes a separate mathematical question. But as the approximation of $\sin\theta$ by suggests, there may be cases where the process of idealization seems motivated by exclusively mathematical considerations. It must also be kept in mind that the closeness of an original quantity and its idealized surrogate is by itself no guarantee that functional outputs or solutions will be correspondingly close. The specifics of the mathematics involved must be examined in order to determine whether closeness is preserved.

The construction of sequences that are increasingly less idealized is a major challenge for science. The pendulum example also illustrates an important experimental challenge. And that is the construction and development of experiments that more closely approach existing idealized analyses. So, for example, assuming the correctness of (3), eliminating the resistance and keeping the oscillations small is a way of making an actual pendulum approach its counterfactual cousins. There are two responses therefore that typically are made to the use of idealizations. First, one can try to bring the analysis to the phenomenon by making the analysis less idealized. Second, one can try to bring the phenomenon to the description by experimental refinement.

3 Monotonically convergent sequences and targets of approximation

In the case of the pendulum the identification of with $\sin\theta$ and the substitution of zero for resistance could be readily interpreted as transformational approximations because equation (3) served as a reference or target. But where do such targets come from? How completely need they be specified in order to determine that some idealization is a transformational approximation? Is the apparent convergence of a sequence of idealized analyses a reliable indication that a target exists? In order to appreciate the difficulties such questions pose, imagine that *t* is the true fundamental theory applicable in some domain of interest, and that there exists a sequence of input conditions and their idealizations $\langle i_1, \ldots, i_i, \ldots\rangle$ such that $t(\langle i_1, \ldots, i_i, \ldots\rangle)$ is convergent on the actual effect. (Henceforth, the sequence will be denoted $\langle i_i\rangle$, and functional application as $t\langle i_i\rangle$.) So, for example, imagine that there are one hundred forces, f_1 through f_{100}, all aligned along a common axis, operating on some particle. And let the first ninety-nine be of equal magnitude but alternating direction, and the remaining force be twice the magnitude of the others. Now imagine that scientist *S* arrives on the scene, gradually learns of the existence of the forces in the sequence $\langle f_1, f_2, \ldots, f_1, f_{i+1}\rangle$, and applies standard Newtonian theory, that the acceleration of the particle will be equal to the product of its mass and the vectorial sum of the forces. (The complications that would result if *S* were to use idealized descriptions of the forces will be ignored.) The resultant summations can naturally be described as becoming increasingly more realistic and less idealized. *S*’s ongoing efforts, however, will be to little avail because improving things this way does not result in the gradual convergence of prediction to actual acceleration. Convergence occurs all at once when the enumeration is complete. The target becomes clear only once one has arrived and it is known that one has arrived. It must be known that the enumeration is complete. That $t\langle i_i\rangle$ is convergent is by itself not of much value because all that is assured is that if *S* works long enough, experimental fit will get better and remain so. But how long is long enough is not specified; only that there exists at least one such point. Consider the consequences of this observation for the testing of scientific theories. Say, for

example, that *S* wishes to test whether the underlying Newtonian theory is true. Failure to achieve improvement of experimental fit by the use of increasingly more realistic idealizations would not be evidence of its falsity. By contrast, assume the existence of a world where the input can be ordered such that $t\langle i_i \rangle$ is strictly monotonic (that is, predictive output is always getting better). In such a world, the use of *t* would be rewarded by constantly improving experimental fit as increasingly more realistic idealizations were employed - assuming, of course, that the appropriate input sequence is employed. If such continual improvement did not occur, then it would be known that the theory was not correct. So in a situation where strict monotonicity holds, disconfirmation is possible and the use of a correct theory will be rewarded in the sense that experimental fit will continually improve as increasingly more realistic idealizations are used. But only on the assumption that one has access to the strictly monotonic sequence of input conditions. How, though, is this to be assured? Or that such a sequence even exists? It is hard to see how such determinations could be made without assuming the truth of the very theory to be tested. Say, for example, that it is discovered that *t* and some set of idealizations, i_1, yield more accurate predictions than the combination of *t* and some other set i_2 of idealizations. Presumably, this lends some support for i_1 being placed later in the sequence than i_2. Such an appraisal, however, relies on the assumption that *t* is true. If so, then it is the idealizations that are being tested by means of experiment and not the theory itself! So it looks as though theories can be tested if it is assumed that idealizations are monotonically ordered, and idealizations can be ordered if the corresponding theory is assumed. But there seems no way to test experimentally both theory and idealizations.

One proposal for avoiding this result is to employ some sort of bootstrapping procedure. Briefly, the idea is to begin the confirmational process by using the theory along with experimentation to tentatively order the idealizations used in some particular domain of application. Attention is then shifted to a different domain or experimental situation, but it must be one that is not so different that the previously obtained measures of relative realism will not be expected to hold. If the new theoretical predictions do not retain the ordering of the original domain (assuming appropriate compensation can somehow be made for relevant differences), then suspicion is cast on the theory. So, for example, predictions of pendulum period are more accurate when hydrostatic effects are taken into account. This superiority is expected to continue in the case of compound pendulums. Sometimes enough is known about the interfering causes and the mathematical properties of possible idealizations that it can be formally shown that monotonically convergent sequences cannot exist for the theory in question. For example, with respect to the kinetic theory of gases, it was shown that the equipartition theorem entailed that making molecule descriptions more realistic (by adding more degrees of freedom) would lead to predicted values for specific heats that would be progressively worse than those predicted on the basis of simple models.

4 Developments and consequences

Systematic development of the themes presented above has proceeded along a broad front. Much effort has been expended to develop abstract but mathematically precise ways of characterizing scientific theories that will take into account idealizations and approximations. One of the anticipated benefits of such characterizations is being able to analyse theories themselves as idealizations or approximations of other theories. There has also been considerable formal development of appropriate measures for nearness or closeness of fit. The choice of such measures becomes problematic when more is involved than simply determining how far one number is from another and there are multiple criteria for successful fit. Another line of investigation involves understanding the scientific use of idealizations and approximations in terms of more general notions of computation. It has long been believed that there is a fruitful analogy to be developed between the development of science and finding equational solutions by means of successive approximations. Other subjects of active investigation include the differences between the uses of idealizations in the pure and applied sciences, the relevance of the use of idealizations for the debate among realists and instrumentalists (see Scientific realism and antirealism), and the importance of non-linear dynamics (see Chaos theory) for understanding the successful use of idealizations and approximations.

See also: Experiment; Mechanics, classical; Models; Theories, scientific

RONALD LAYMON

References and further reading

Balzer, W., Moulines, C.U. and Sneed, J.D. (1987) *An Architectonic for Science: The Structuralist Program*,

Dordrecht: Reidel.(Develops a formal taxonomy for scientific theories. See chapter VII for connections with approximation.)

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Clarendon Press. (Discusses many interesting scientific cases and presents a very original view of the relevance of the use of idealizations for scientific realism and instrumentalism.)

Cartwright, N. (1989) *Nature's Capacities and their Measurement*, Oxford: Clarendon Press. (Further developments of positions presented in her earlier book.)

Laymon, R. (1985) 'Idealizations and the Testing of Theories by Experimentation', in P. Achinstein and O. Hannaway (eds) *Experiment and Observation in Modern Science*, Boston, MA: MIT Press-Bradford Books, 147-73.(A general informal introduction to the subject. Contains many historical examples.)

Laymon, R. (1990) 'Computer Simulations, Idealization and Approximation', in A. Fine, M. Forbes and L. Wessels (eds) *PSA 1990*, East Lansing, MI: Philosophy of Science Association, vol. 2, 519-34.(Develops connections with formal theories of computation and computer semantics. Discusses some of the implications of chaos theory.)

Niiniluoto, I. (1984) *Is Science Progressive?*, Dordrecht: Reidel.(Covers most of the items discussed in §4. Especially valuable is the work on measures of approximation. Also contains extensive literature reviews and bibliographies.)

Ideals

Ideals are models of excellence. They can be moral or nonmoral, and either 'substantive' or 'deliberative'. Substantive ideals present models of excellence against which things in a relevant class can be assessed, such as models of the just society or the good person. Deliberative ideals present models of excellent deliberation, leading to correct or warranted ethical conclusions. Ideals figure in ethics in two opposed ways. Most centrally, ideals serve to justify ethical judgments and to guide people in how to live. Sometimes, however, ideals may conflict with moral demands, thereby testing the limits of morality.

Reliance upon ideals in the development of ethical theories seems unavoidable but raises difficult questions. How can the choice of a particular ideal be justified? How might conflicts between ideals and other values, especially moral demands, be resolved?

1 Substantive and deliberative ideals

Substantive ideals delineate the features that something or someone must possess in order to be excellent in a specific regard. Such ideals can provide a highly detailed model or a more general framework. Whether specific or general, however, substantive ideals logically imply certain evaluative judgments in conjunction with the facts about whether and to what degree something or someone possesses the relevant features. A person can hold universal ideals of human excellence or human social life, as well as more personal ideals regarding who to be and how to live. Common universal ideals include ideals of the just or virtuous person, ideals of human rationality, the liberal ideal of a free society, and hierarchical ideals of societies structured to realize human excellence. Prevalent personal ideals include those of being an exceptional parent, a connoisseur of high culture, a brave soldier or a devout follower of God.

Substantive ideals provide standards that guide the development of character and commitments, aims and attitudes, social relationships and institutions. When persons hold an ideal of human excellence, for instance, they regard it as central to their identity. They are normally motivated to cultivate the emotions, qualities and pursuits their ideal upholds, and they feel shame, guilt or self-contempt when they fail to live up to their ideal. To hold an ideal thus differs from merely wanting or preferring something. Because ideals provide a basis for assessment, holding an ideal is more akin to having a second-order desire, a desire to have or not to have certain other desires.

Each moral system or theory has its corresponding ideals, but whether each actually derives from an ideal is a matter of some controversy. Nevertheless, substantive ideals have historically been central to many different moral conceptions. In his *Republic*, Plato famously constructs a blueprint of the just state (see Plato §14). Aristotle's account of human happiness in *Nicomachean Ethics* rests upon a substantive ideal of excellent human functioning (see Aristotle §§22-7). G.E. Moore, in *Principia Ethica* (1903), presents an account of 'the Absolute Good or Ideal', which is made up of the intrinsic goods of personal affections and aesthetic enjoyments (see Moore, G.E. §1). Other important substantive ideals include Kant's 'kingdom of ends' (1785), an ideal union between self-legislating rational beings (see Kant, I. §§9-10; Kantian ethics); Nietzsche's ideal of the 'overman' (*Übermensch*) (1883-5), the creative, self-determining man who lives without fear of himself or others (see Nietzsche, F. §10); and John Rawls' Kantian ideal of the person (1980) (see Rawls, J.).

Substantive ideals continue to be important to ethical theorizing. But philosophers in the twentieth century have largely avoided explicit appeal to substantive ideals, perhaps in recognition of the apparent plurality of defensible values (see Moral pluralism; Values). Instead, they often rely on deliberative ideals, seeking to develop a critical perspective on our ordinary desires and ethical assessments that is sensitive to value pluralism.

Deliberative ideals specify optimal conditions for reflection on ethical questions. They form the basis of broadly counterfactual accounts of moral and nonmoral value, most notably, in 'ideal observer' theories and contractarian theories. The deliberative ideals developed in contractarian theories figure critically in attempted derivations of principles of justice or morality. But as employed in ideal observer theories, deliberative ideals do not logically support evaluative conclusions: evaluative conclusions are determined by an observer's reactions - usually some sort of 'pro' or 'con' attitude - under ideal conditions. In either case, though, deliberative ideals function to justify ethical conclusions by incorporating conditions designed to ensure both that a deliberator's conclusions or

reactions have been subjected to rational, critical scrutiny and that they address the concerns of actual deliberators. Conditions are thereby also designed to preserve the recommending force of ethical judgments and the connection that normally holds between ethical judgment and motivation (see Moral motivation §1). Deliberative ideals typically enhance the rationality of deliberators and either markedly increase or relevantly constrain their knowledge. They may also improve a deliberator's motivations, for instance, by eliminating emotional distortions (see Morality and emotions §1).

Deliberative ideals are sometimes offered as analyses or reforming definitions of ethical terms. The paradigm ideal observer theory, developed by Roderick Firth (1952), analyses the meanings of ethical words in terms of the 'ethically significant reactions' of an ideal observer, one who is disinterested, dispassionate, omniscient, omnipercipient, consistent and 'otherwise normal'. Certain ethical naturalists, notably Richard Brandt (1979) and Peter Railton (1986), have offered structurally similar 'reforming definitions' of ethical terms. And R.M. Hare's account of 'critical thinking', based on the 'logic' of the moral concepts, parallels Firth's view in many respects (see Hare, R.M. §3).

John Rawls' account of the 'original position', in *A Theory of Justice* (1971), is perhaps the most famous contractarian deliberative ideal. According to Rawls, fair choice of principles of justice occurs when rational, self-interested, mutually disinterested persons select principles under conditions of limited shared information from behind a 'veil of ignorance'. Deliberative ideals related to Rawls' can be found in the contractarian views of Thomas Scanlon (1982) and David Gauthier (1986), for example, as well as in Jürgen Habermas' 'discourse ethics' (1990) (see Rawls, J.; Contractarianism §9; Habermas, J. §3).

Deliberative ideals are often thought to avoid substantive evaluative commitments, at least controversial ones. Certain deliberative ideals, however, may reflect substantive ideals, as the deliberative ideal embodied in Rawls' original position reflects his Kantian ideal of the person. Thus, deliberative ideals may not be fully distinct from substantive ideals.

2 Selecting ideals

Given the important justificatory role of ideals in ethical theories, how might choice of a particular substantive or deliberative ideal be justified? Various answers have been given, each of which encounters difficulties.

Plato held that knowledge of the Forms of Goodness and Justice is innate and recovered through a process of recollection. But such claims regarding innate knowledge seem contradicted by ethical disagreement. We might, following Aristotle, look to human nature to support choice of ideals (see Human nature §1). But we lack a convincing account of human nature, and the evident diversity of human talents and reasonable aims suggests that human nature will not support any single ideal. Moore determines the content of 'the Ideal' by considering items in complete isolation in order to ascertain their intrinsic value. This 'method of isolation', however, assumes implausibly that we can directly intuit an item's value. Firth claims to select the characteristics of an ideal observer 'by examining the procedures which we actually regard, implicitly or explicitly, as the rational ones for *deciding* ethical questions' (1952: 332; original emphasis). But our ordinary view of which procedures are rational may be faulty and less determinate than Firth supposes. Brandt's and Railton's reforming definitions incorporate purely epistemic ideal conditions, thereby apparently remaining neutral as between substantive ideals and desires. Yet conditions that are only epistemically ideal may lack normative force. For the reactions individuals would have under such conditions will depend, in part, on their own motivational systems, which may themselves warrant normative criticism. Finally, Rawls argues that his original position incorporates only weak and widely shared assumptions about what makes for a fair choice. Critics of Rawls have argued, however, that his assumptions are controversial.

Choice between ideals used to support competing conceptions of moral and nonmoral value can be settled to some degree by considering such things as the relative coherence and consistency of the ideals, as well as their compatibility with empirical information. But as Samuel Scheffler regretfully remarks, there may be no 'ideal-independent, trans-theoretical moral perspective which condemns the ideal we abhor and favours the ideal we prize' (1979: 300) (see Moral scepticism).

3 Ideals and morality

In some cases, conflicts may arise between a person's ideals (moral or nonmoral) and other values, in particular, alleged moral obligations (see Duty). These conflicts reveal important questions about the divide between morality and nonmoral value and about what limits may exist to the demands of morality.

Bernard Williams has considered such conflicts (1973), arguing that utilitarianism in particular and impartial morality more generally may be destructive of individual integrity, since they demand that persons set aside their deepest commitments in favour of what morality may require (see Impartiality; Williams, B.A.O. §2). Williams does not use the language of ideals, but the commitments which give individuals a sense of what their lives are about will surely include commitments to particular ideals. The conflicts examined by Williams expose difficult questions about morality. Is it purely impartial or does it include 'agent-centred prerogatives' (Scheffler 1982)? If morality is purely impartial, are its demands always overriding? In answering these questions, we see how ideals may function not only to ground morality, but to fix its contours and limit its demands.

See also: Axiology; Perfectionism; Solidarity; Virtue ethics

CONNIE S. ROSATI

References and further reading

Anderson, E. (1993) *Value in Ethics and Economics*, Cambridge, MA: Harvard University Press.(Develops a clear and accessible ideal-based theory of value.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. W.D. Ross, Oxford: Oxford University Press, 1975. (Aristotle's account of happiness and excellent human functioning.)

Brandt, R. (1979) *A Theory of the Good and the Right*, Oxford: Clarendon Press. (Presents reforming definitions of ethical terms, such as 'good' and 'morally wrong', that employ deliberative ideals.)

Firth, R. (1952) 'Ethical Absolutism and the Ideal Observer', *Philosophy and Phenomenological Research* 12: 317-45.(Classic example of a deliberative ideal offered as an analysis of ethical terms and judgments.)

Gauthier, D. (1986) *Morals by Agreement*, Oxford: Clarendon Press.(Advanced. Argues that morality is a set of rational principles for choice that rational persons would accept in a fully voluntary *ex ante* agreement.)

Habermas, J. (1990) 'Discourse Ethics: Notes on a Philosophical Program of Justification', in *Moral Consciousness and Communicative Action*, Cambridge, MA: MIT Press.(Explication and programmatic justification of discourse ethics, according to which the practical discourse that leads to valid norms involves communication that adequately approximates ideal conditions.)

Hare, R.M. (1963) *Freedom and Reason*, Oxford: Oxford University Press.(Contains discussion of conflicts between ideals and interests and how this may limit the scope of moral argument as Hare depicts it.)

Hare, R.M. (1981) *Moral Thinking: Its Method, Level, and Point*, Oxford: Oxford University Press.(Develops an ideal of moral reasoning based on Hare's account of the logic of moral concepts.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964, ch. 3.(Develops Kant's ideal of a kingdom of ends.)

Moore, G. (1903) *Principia Ethica*, ed. T. Baldwin, Cambridge: Cambridge University Press, 1993, ch. 6. (Complex discussion of intrinsic goods and evils.)

Nietzsche, F. (1883-5) *Also sprach Zarathustra*, trans. and ed. W. Kauffman, *Thus Spoke Zarathustra*, in *The Portable Nietzsche*, New York: Viking Penguin, 1982.(Presents Nietzsche's conception of the 'overman'.)

Plato (*c*.380-367 BC) *Republic*, trans. P. Shorey, Cambridge, MA: Loeb Classical Library, Harvard University, 1930; repr. in *The Collected Dialogues including Letters*, ed. E. Hamilton and H. Cairns, Oxford: Oxford University Press, 1973. (Develops Plato's substantive ideals of justice in the state and in the individual.)

Railton, P. (1986) 'Facts and Values', *Philosophical Topics* 14: 5-31.(Develops a deliberative ideal as part of a reforming account of an individual's nonmoral good.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Main statement of Rawls' deliberative ideal and theory of justice.)

Rawls, J. (1980) 'Kantian Constructivism in Moral Theory' (The Dewey Lectures), *Journal of Philosophy* 77: 515-72.(Advanced discussion of Rawls' Kantian ideal of the person.)

Scanlon, T.M. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press.(Presents a contractualist account of the nature of morality that incorporates a deliberative ideal.)

Scheffler, S. (1979) 'Moral Scepticism and Ideals of the Person', *Monist* 62: 288-363.(Discussion of the connection between moral scepticism and ideals of the person.)

Scheffler, S. (1982) *The Rejection of Consequentialism*, Oxford: Clarendon Press. (Addresses the challenge raised by Williams, defending the existence of 'agent-centred prerogatives'.)

Scheffler, S. (1982) 'Ethics, Personal Identity, and Ideals of the Person', *Canadian Journal of Philosophy* 12 (2): 229-62.(Suggests that moral theories rest on commitment to an ideal of the person, and offers this as a basis for replying to the sceptic.)

Williams, B. (1973) 'A Critique of Utilitarianism', in J.J.C. Smart and B. Williams, *Utilitarianism For and Against*, Cambridge: Cambridge University Press.(Important critique of utilitarianism.)

Identity

Anything whatsoever has the relation of identity to itself, and to nothing else. Things are identical if they are one thing, not two. We can refute the claim that they are identical if we can find a property of one that is not simultaneously a property of the other. The concept of identity is fundamental to logic. Without it, counting would be impossible, for we could not distinguish in principle between counting one thing twice and counting two different things. When we have acquired the concept, it can still be difficult to make this distinction in practice. Misjudgments of identity are possible because one thing can be presented in many guises.

Identity judgments often involve assumptions about the nature of things. The identity of the present mature tree with the past sapling implies persistence through change. The non-identity of the actual child of one couple with the hypothetical child of a different couple is implied by the claim that ancestry is an essential property. Knowledge of what directions are involves knowledge that parallel lines have identical directions. Many controversies over identity concern the nature of the things in question. Others concern challenges to the orthodox conception just sketched of identity itself.

1 Exposition of a popular view

Identity is the relation that, necessarily, each thing has to itself and to nothing else. Thus Constantinople has the identity relation to Istanbul because Constantinople *is* Istanbul, the very same city. This relation is often called 'numerical' identity, to distinguish it from 'qualitative' identity, the relation of exact similarity. Although 'identical' twins might be qualitatively identical, they are not numerically identical, for there are two of them, not one. The formula '$x = y$' says that *x* and *y* are (numerically) identical.

Identity is governed by two basic logical principles, reflexivity and Leibniz's Law (see Leibniz, G.W. §11). To say that the identity relation is reflexive is to say that for each thing *x*, $x = x$. Leibniz's Law says that if $x = y$ then whatever is true of *x* is also true of *y*; it is used in arguments such as 'Jack the Ripper was in Whitechapel last night, the Prince of Wales was not in Whitechapel last night; therefore Jack the Ripper is not the Prince of Wales'. Leibniz (1704) was not the first to formulate this law, which was known in antiquity. It is sometimes called the 'indiscernibility of identicals', not to be confused with Leibniz's principle of the 'identity of indiscernibles', a kind of converse, which says (implausibly) that qualitative identity implies numerical identity (see Identity of indiscernibles). Reflexivity and Leibniz's Law characterize identity uniquely: it is provably impossible for two non-equivalent relations to satisfy both principles. The two principles also entail that identity is symmetric (if $x = y$ then $y = x$) and transitive (if $x = y$ and $y = z$ then $x = z$).

If $a = b$, Leibniz's Law says that whatever is true of *a* is true of *b*. However, this permits the replacement of '*a*' by '*b*' only in contexts in which the expressions merely specify which thing is being talked about. For example, it is invalid to argue from 'Jocasta = the mother of Oedipus' and 'Oedipus knows that he married Jocasta' to 'Oedipus knows that he married the mother of Oedipus', for here 'the mother of Oedipus' does not merely specify a person; it specifies the description under which Oedipus is said to know that he married her (see Propositional attitudes). Looking at photographs of a mature tree and a sapling, one cannot use Leibniz's Law to refute the hypothesis 'The mature tree = the sapling' on the grounds that 'The mature tree is tall' is true and 'The sapling is tall' is false. The noun phrases 'the mature tree' and 'the sapling' do not merely specify trees; they indicate the times with respect to which tallness is predicated. A correct understanding of Leibniz's Law is needed if identity through change is not to seem contradictory.

A genuine consequence of Leibniz's Law is the necessity of identity: things that are in fact identical could not have been distinct (see Modal logic, philosophical issues in §4). For suppose that $x = y$. Since *x* could not have been distinct from itself, *x* could not have been distinct from *y*. Thus *x* and *y* cannot be contingently identical: they must be identical in all circumstances. A more complex argument concludes that they cannot be contingently distinct. They are either necessarily identical or necessarily distinct. Consequently, they are either always identical or always distinct. For example, if my headache is identical with an event in my brain, then that headache has to be that event; neither could exist without the other. Of course, a sentence such as 'Rome = the capital of Italy' is contingently true, but that is because the description 'the capital of Italy' need not have specified the city it actually specifies. The example is consistent with the necessity of identity, however, for it does not imply that one

city could have been each of two.

An argument like that for the necessity of identity can be used to refute the idea that questions of identity need have no right answer. For example, it is sometimes held to be indeterminate whether a given mass of rock and ice is Everest (see Vagueness §1). However, it is determinate whether Everest is Everest. If Leibniz's Law and ordinary logic apply in this context, it follows that Everest is distinct from that mass, and the identity question has a right answer - a negative one. Thus the hypothesis of indeterminacy contradicts itself.

Statements of identity in natural languages often include a noun answering the question 'same what?', for example, 'Istanbul is the same city as Constantinople'. Since identity is uniquely characterized by its logic, the role of the noun is not to disambiguate 'same'; '*a* is the same *F* as *b*' is equivalent to '*a* is an *F* and $a = b$' (from which '*b* is an *F*' follows by Leibniz's Law). Identity is not defined kind by kind. Rather, we use the pregiven notion of identity in defining kinds. For example, we explain the difference between rivers and collections of water molecules by saying that the same river contains different collections of water molecules at different times.

A 'criterion of identity' for a kind is a necessary and sufficient condition for members of the kind to be identical. Frege's criterion of identity for directions is that the directions of lines are identical if and only if the lines are parallel (1884). The criterion of identity for numbers is that the number of *F* things is the number of *G* things if and only if there is a one-to-one correlation between the *F* things and the *G* things. In such cases, members of the kind can be presented in various guises: directions as the directions of various lines; numbers as the numbers of various pluralities. The criterion of identity states the condition for two guises to be guises of one member of the kind. Without a grasp of these conditions, one would not know what directions or numbers were.

A criterion of identity for a spatiotemporal kind is expected to give the condition for a member of the kind at one place and time to be identical with a member of the kind at another place and time: for example, it may have to follow a continuous trajectory between these space-time points. An account may also be required of an object's identity across hypothetical circumstances. For example, one may hold that a member of the kind could not have originated at different places and times in different possible worlds. Spatiotemporal objects can be exactly similar without being identical, for they can originate at different times or places. In contrast, purely abstract objects cannot be exactly similar without being identical.

2 Alternatives

Every aspect of the preceding view has been questioned. Although it is popular, the philosophers who accept it all may be in a minority.

That identity is a relation has been denied, both on metaphysical grounds (it cannot relate two things) and on grammatical grounds (some assimilate 'is' in 'Constantinople is Istanbul' to the 'is' of predication as in 'Constantinople is crowded'). However, the logic of 'Čequals;' does single out a unique class of ordered pairs of objects to which it applies, which suffices to make identity a relation in some minimal sense.

The contrast between identity and indiscernibility has been challenged, on the grounds that indiscernibles satisfy the same descriptions and since *a* satisfies the description 'identical with *a*', so does anything indiscernible from *a*. The obvious reply is that indiscernibles merely satisfy the same *intrinsic* descriptions, but it is hard to explain what 'intrinsicness' is.

Many applications of Leibniz's Law are problematic. Some deny its applicability to contexts that treat non-actual possibilities or even non-present times, thus excluding the derivation of the necessity or permanence of identity. The intention is, for example, to permit a pot to be contingently identical with the clay of which it is made, or the clay to be temporarily identical with the pot. One question is whether the envisaged restrictions on Leibniz's Law are *ad hoc*. They would not be if so-called identity between objects in different possible worlds or at different times could not be taken at face value but was somehow reducible to relations of qualitative similarity among counterpart objects each of which was confined to a single possible world or time. Such views have been supported by appeal to the difficulty of specifying what is essential to an object, for example, to what extent a particular ship could originally have been made from different planks of wood, or how much it could change without ceasing to exist. However, the proposed reductions are both complex and hard to reconcile with ordinary assumptions about the nature of everyday objects.

Identity has also been regarded as sometimes indeterminate. Although it is usually conceded that if things are identical then it is determinate whether they are identical, in some nonstandard logics it does not follow that if it is indeterminate whether they are identical then they are not identical. Such logics postulate an intermediate status that propositions can have between truth and falsity. One challenge to this view is to explain what it means for a proposition to be not true without being false.

Yet another nonstandard view is the doctrine that identity is always relative to an answer to the question ‘same what?’. For example, if *a* and *b* are copies of *Middlemarch*, this view disputes the inference from ‘*a* is the same edition as *b*’ and ‘*a* is a copy’ to ‘*a* is the same copy as *b*’. The relation of being the same edition will not satisfy Leibniz’s Law unrestrictedly, but the view needs to show that no other relation (for example, being the same copy) could satisfy Leibniz’s Law unrestrictedly, for such a relation would be a case of non-relative identity.

Many disputes about criteria of identity for particular kinds of entity (see Personal identity) concern the nature of those entities, not of identity itself. But even the concept of a criterion of identity is itself problematic. For example, can one give an adequate criterion of identity for events by saying that events are identical if and only if they have the same causes and effects? The problem is that the causes and effects of events include other events. The criterion is in some sense circular, but it is hard to state the requirement of non-circularity clearly. Even if that could be done, it is unclear why every kind of object should have a non-circular criterion of identity. There may be nothing more basic than identity to which identity could be reduced. The standard logic of identity demands no such reduction.

Perhaps identity will come to be regarded as a logical constant, no more problematic than, say, conjunction. If so, many of the issues mentioned above will remain difficult, but their difficulty will concern the nature of various kinds of object, not the relation of identity.

See also: Continuants §1; Logical and mathematical terms, glossary of

TIMOTHY WILLIAMSON

References and further reading

Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl*, trans. J.L. Austin, *Foundations of Arithmetic*, Oxford: Blackwell, 2nd edn, 1980, §§64-73.(On criteria of identity.)

Kripke, S.A. (1980) *Naming and Necessity*, Oxford: Blackwell.(A basic work on the connection between identity and necessity.)

Leibniz, G.W. (1704) *New Essays on Human Understanding*, trans. and ed. P. Remnant and J. Bennett, Cambridge: Cambridge University Press, 1981, bk 2, ch. 27.(Gives Leibniz’s view of identity and contrasts it with that of John Locke.)

Noonan, H.W. (ed.) (1993) *Identity*, Aldershot: Dartmouth.(Collects many relevant articles.)

Wiggins, D.R.P. (1980) *Sameness and Substance*, Oxford: Blackwell.(On Leibniz’s Law and the ‘same what?’ question.)

Williamson, T. (1990) *Identity and Discrimination*, Oxford: Blackwell.(On supposed cases of indeterminate identity.)

Identity of indiscernibles

The principle of the identity of indiscernibles states that objects which are alike in all respects are identical. It is sometimes called Leibniz's Law. This name is also frequently used for the converse principle, the indiscernibility of identicals, that objects which are identical are alike in all respects. Both principles together are sometimes taken to define the concept of identity. Unlike the indiscernibility of identicals, which is widely accepted as a logical truth, the identity of indiscernibles principle has frequently been doubted and rejected. The principle is susceptible of more precise formulation in a number of ways, some more dubitable than others.

1 Leibniz's principle

In his metaphysics, Leibniz frequently stated a principle that no two separate individual things could differ only numerically, that is, resemble one another in all their properties or 'intrinsic denominations' and yet not be one and the same thing. According to various scholastics, two things may differ merely numerically, meaning they could have the same nature, but be distinguished by their matter. Immaterial individuals such as angels must each have a distinct nature, and Leibniz, rejecting matter, extended this view to all individuals. Sometimes he justified it theologically: God never does anything without a reason, and if two things were exactly alike, he would have no reason to create one rather than the other. Sometimes Leibniz referred to empirical evidence, recounting the story of the nobleman he set vainly to searching a castle garden to find two leaves exactly alike, though of course the mere failure to find a counter-instance on a restricted search is no serious test. Leibniz used his principle to argue against empty space and time.

In his logical writings Leibniz defined a relation of sameness between terms as follows: 'Terms which can be substituted for one another whenever we please without changing the truth of any statement (*salva veritate*), are the same (*eadem*) or coincident (*coincidentia*)'. When terms are coincident they denote the same thing or things. Leibniz's definition covers general as well as singular terms, as his triangle/trilateral example makes clear. It would perhaps be better to say that the things are the same when their names are coincident, but Leibniz's meaning is clear.

Leibniz's definition of identity was adopted by Frege, who restricted it to singular terms. In second-order predicate logic, singular identity is often defined as follows:

$$a = b \equiv_{\mathrm{Df.}} \forall \mathrm{F}(\mathrm{F}(\mathrm{a}) \equiv \mathrm{F}(\mathrm{b})).$$

The left-to-right implication is uncontroversial, but needs careful formulation to exclude non-extensional contexts. For example, in 'John believes that *x* defeated Mark Antony', substituting the names 'Octavian' and 'Augustus' for *x* may yield different truth-values if John does not realize that Augustus and Octavian are the same person, yet his ignorance does not impugn their identity (see Propositional attitudes). *F* is taken to range over properties or predicate extensions determined by extensional sentential contexts only. It is thus important to distinguish the substitution principle, which is about expressions, and requires such restrictions, from the indiscernibility principle, which is about the properties of individuals (see Identity §1).

It is the right-to-left implication that is the modern symbolic form of Leibniz's identity of indiscernibles, and this is the half of the definition that has attracted most criticism. There is a version of the principle which is trivially true. Assume $\forall F(F(a) \equiv F(b))$. Substituting '$a = x$' for '$F(x)$' we have $(a = a) \equiv (a = b)$, and since it is logically true that $a = a$, it follows that $a = b$.

Leibniz hardly intended his principle to be understood thus: had it been a logical triviality, he would not have laboured it; indeed he did not consider it a logical truth. This version of the principle is not in doubt. To rescue it from triviality and make it fit to define identity, we must screen off identity and all properties involving it from consideration. Call properties that pass the supposed screening-off test 'material' properties. The most obvious material properties are qualities, so a more substantive version says that no two distinct individuals have all their qualities in common. This is closer to Leibniz's intentions, but readily refuted. For very simple and numerous individuals such as fundamental particles the qualities attributable to them are few, and it is unlikely that two such particles are never exactly alike with respect to the few qualities they do have.

Even allowing that a particle has a history and may change its qualities, we cannot exclude the possibility of two particles having exactly similar qualitative histories. Even if two things never in fact resemble one another exactly, it is not logically impossible that two distinct things be exactly alike. So the qualitative version of the principle seems clearly false.

2 Temporal and modal versions

Evidently there are several indiscernibility principles, differing in plausibility. The question becomes: which one is neither so weak that it is trivial nor so strong that it is false, but just strong enough to characterize identity?

Two objects may be 'temporarily' indiscernible and yet be distinct because they are discernible at some other time. A golden cup and the gold out of which it is made occupy the same place, have the same weight, thermal conductivity, stand in the same spatial and causal relations to other contemporaneous things, and so on. But the gold existed before the cup was fashioned out of it.

Suppose two objects came into being simultaneously and have been always indiscernible to date. Are they identical? Not if their histories diverge later. If a cat *C* and the mereological sum of its body and tail $B+T$ have been indiscernible until now, but later the cat loses its tail, then the histories of *C* and $B+T$ diverge, for *C* loses *T* as a part but $B+T$ does not (it either becomes scattered or ceases to exist), so *C* was not identical with $B+T$: they were merely indiscernible prior to the loss of the tail. The property by which they differed was future-related, just as the history of the gold before it came to constitute the cup is past-related.

Neither temporary indiscernibility nor indiscernibility to date suffice for identity, but if individuals are 'permanently' indiscernible, that is, exist and have the same properties at the same times, are they identical? It seems not. For suppose *C* never loses its tail. *C* remains coincident and indiscernible from $B+T$ at all times it exists. But *C*, unlike $B+T$, could have lost *T* and carried on existing: cats, unlike body-tail sums, can exist without tails. So it seems that *C* is 'modally' discernible from $B+T$. (See Essentialism §4.)

Does identity require indiscernibility in all modal properties? This seems too strong. For example, it is necessary to the first Roman Emperor that he have subjects, but not necessary to Octavian that he have subjects, although Octavian was the first Roman Emperor. This example is inconclusive, since 'the first Roman Emperor' is a definite description and the necessity is *de dicto*: it is not necessary *de re* to the Emperor, that man, that he have subjects (see De re/de dicto §1; Descriptions §2). This may seem to rebound on the cat example: '$B+T$' is also a definite description, 'the sum of *B* and *T*', so the fact that *C* might not have coincided with $B+T$ does not entail that they are distinct, since by the theory of descriptions '$C = \text{the } A$' just means

$$\forall x(x \text{ is an } A \equiv (x = C)).$$

But the cases differ, for while it is true of Octavian that he was an emperor, it is not obviously true of the cat that it is a sum of body and tail. The difference is again modal: no emperor is essentially an emperor, but sums are essentially sums, and *C* is not essentially a sum (see Essentialism §4).

This gives us a more precise modal version: if *a* and *b* are indiscernible with respect to all their non-modal properties and all their 'essential' properties then *a* = *b*. What then is an essential property? One which *a* must have if it exists. But being identical with *a* is such a property, so triviality is again a problem. Furthermore, deciding whether a property is essential or not may turn on deciding identity questions. While this indicates that matters of essence are closely connected with identity, we have to restrict attention to material properties again, and modality muddies considerations.

3 Relations and thought experiments

Once we admit relational properties, that is, consider how individuals stand in relation to others, we radically enlarge the range we quantify over and enhance what can be discerned. Though two electrons might be qualitatively indiscernible, if one is 1 metre away from a third particle while the other is not, they are distinguished by this relational property. The Thomistic view that matter individuates bodies which are alike in nature is also relational: $a \neq b$ because *m* is *a*'s matter and not *b*'s.

This suggests a more promising non-modal version: individuals which are 'materially' indiscernible, that is,

indiscernible with respect to all material properties (monadic and relational), are identical. Can there be two distinct materially indiscernible individuals? Take the two electrons mentioned above. If the third particle is a proton and the electron that is not 1 m from it is 1 m from no other proton, then the electrons differ in their material relations. If both were 1 m from qualitatively indiscernible protons the electrons would not thereby differ materially, for though this electron is 1 m from this proton and that electron 1 m from that proton, we can no longer use this information to discern them, since it turns on this proton's being distinct from that one.

To find numerically distinct but materially indiscernible individuals, philosophers have concocted thought experiments with symmetrical universes. Max Black (1952) envisaged a universe consisting solely of two qualitatively indistinguishable spheres with indiscernible histories. Each sphere and each part of a sphere is materially indiscernible from a numerically distinct sphere or part of a sphere. Except for symmetrical individuals, every part of the universe has a numerically distinct *doppelgänger*. Black's universe appears possible, so either the spheres would have to be merely formally distinct, violating the principle, or there would have to be some haecceitas or individualizer to distinguish the spheres, each being uniquely related to its own haecceitas. But to posit such unobservable things seems unacceptably metaphysical.

Those who uphold the principle of material indiscernibility argue by *modus tollens* that there can be no symmetrical universe, for the spheres would be one sphere, and since the sphere is spherically symmetrical it would 'collapse' to a line, whose points, being identical, would collapse to a point: the only symmetrical universe would be monistic. Since our universe is apparently not symmetrical, it is hard to know how to resolve this dispute.

4 Empirical counterexamples

Harder and more disturbing evidence against Leibniz's principle in almost all versions comes from physics. Two electrons in an atomic shell are discernible from one another while both are in it, having opposite spins, but if an electron enters the shell and an electron later leaves it, there is no way to answer the question whether the same electron left as previously arrived. It is not merely that we cannot observe the proceedings closely enough: there is seemingly no fact of the matter. But electrons are fermions and so must always be discernible at a point in time by their quantum states. The Leibnizian may therefore retreat to holding that identity for electrons is only synchronic, and that they are only momentary, though lone electrons seem to persist happily, only gregarious ones 'losing their identity'. The case is worse for photons, which are bosons and so can be temporarily indiscernible in all their material properties, as well as their position and relations. They can become superposed 'soup': a laser is just very homogeneous photon soup. Yet the intensity of a laser beam depends on how many photons are in it: coincident photons do not collapse into one. A Leibnizian can rescue indiscernibility here only by denying that photons are individuals: talk of photons must then be a physicist's dispensable *façon de parler*.

So although there is dispute about what the indiscernibility of identicals amounts to, and whether in this or that version it is true, we appear comfortable conceiving of things as individuals only to the extent that they are in some sense discernible.

See also: Logical and mathematical terms, glossary of

PETER SIMONS

References and further reading

Black, M. (1952) 'The Identity of Indiscernibles', *Mind* 61: 153-64. (A famous dialogue on the spheres example.)

Brody, B.A. (1980) *Identity and Essence*, Princeton, NJ: Princeton University Press. (Tenaciously defends and exploits the trivial version of Leibniz's principle.)

Fraassen, B.C. van (1991) *Quantum Mechanics: An Empiricist View*, Oxford: Clarendon Press. (See chapters 11, 12 for a logically astute philosopher's discussion of the 'problem of identical particles' in physics.)

Leibniz, G.W. (1956) *Philosophical Papers and Letters*, ed. L.E. Loemker, Dordrecht: Reidel, 1975. (Follow the index entries under 'Identity'.)

Mates, B. (1986) *The Philosophy of Leibniz: Metaphysics and Language*, New York: Oxford University Press. (Chapter 7, §3 discusses the identity of indiscernibles principle in Leibniz.)

Wiggins, D. (1980) *Sameness and Substance*, Oxford: Blackwell. (Chapter 2, §2 discusses (and rejects) the identity of indiscernibles principle amid a general theory of identity.)

Ideology

An ideology is a set of ideas, beliefs and attitudes, consciously or unconsciously held, which reflects or shapes understandings or misconceptions of the social and political world. It serves to recommend, justify or endorse collective action aimed at preserving or changing political practices and institutions. The concept of ideology is split almost irreconcilably between two major senses. The first is pejorative, denoting particular, historically distorted (political) thought which reinforces certain relationships of domination and in respect of which ideology functions as a critical unmasking concept. The second is a non-pejorative assertion about the different families of cultural symbols and ideas human beings employ in perceiving, comprehending and evaluating social and political realities in general, often within a systemic framework. Those families perform significant mapping and integrating functions.

A major division exists within this latter category. Some analysts claim that the study of ideology can be non-evaluative in establishing scientific facts about the way political beliefs reflect the social world and propel people to specific action within it. Others hold that ideology injects specific politically value-laden meanings into conceptualizations of the social world which are inevitably indeterminate, and is consequently a means of constructing rather than reflecting that world. This also applies to interpretations undertaken by the analysts of ideology themselves.

1 Marxist approaches

The study of ideologies initially arose from the attempt of the French writer Antoine Destutt de Tracy to create a systematic science of ideas, distinct from prejudices. That meaning of ‘ideology’ is no longer of importance in contemporary debate. Of far greater consequence was the Marxist conception of ideology, which has had a significant impact on the inquiry into, and understanding of, ideology, as well as directing its analysis into an intellectual cul-de-sac. In the nineteenth century Karl Marx (§§6, 7), assisted by Friedrich Engels, converted ’ideology’ into a critical concept (Marx and Engels 1846). The great value of Marx’s contribution lay in relating ideology decisively to the socio-economic practices from which it sprang - specifically to capitalist production - and in pointing out how effective its thought edifices could be in (mis)directing human energies and shaping human institutions. Ontologically, ideology could be contained in Marx’s foundational materialist axiom that being conditions consciousness. Epistemologically, it emerged as a special case when human consciousness reflected the alienated, dehumanized and partial existence of human beings, through a distorted representation of that existence (see Alienation §§3, 5). Hence not all thought was ideological. Ideology arose historically as an inverted reflection of the material contradictions of the capitalist mode of production and denoted only one type of socially produced thought, albeit a very pervasive one. The misrepresentation of the real character of social relations, ’the distorted language of the actual world’, was thus both a historical necessity - because of the conditioned nature of human thought - and a function of ideology (now interpreted as concealment). In identifying this latter aspect, Marx emphasized ideology as a reflection, and in itself a form, of the power and exploitation embedded in the social contradictions to which material conditions give rise. The dominant class - the bourgeoisie - was the beneficiary of ideology, precisely because ideology served its interest - namely, the maintenance of its domination - through inverting the real facts about capitalism. Originally Marx employed the analogy of the mirror image of the *camera obscura* to illustrate that distortion and to distinguish ideology from error or illusion. However, his analysis was more sophisticated in that he did not see ideology as a direct reflection of a distorted reality but as an inversion of that distortion, so that the repressive elements of capitalism, such as entrapping people in commodity exchange relationships, could be presented as free trade.

One consequence of Marx’s analysis of ideology lies in its opposition to true consciousness. The function of ideology is concealment, but its critical identification as such is a crucial step towards its overcoming. Ideology is hence an ephemeral phenomenon, and once the practices which give rise to it are negated, it too will disappear. A second consequence is its association with power and domination; indeed, for many analysts the ‘dominant ideology’ thesis locates its central feature in the superimposition of a class or particular world view on a society. Thus for the Italian Marxist Antonio Gramsci (§3), ideologies were superstructures, consolidated by intellectuals (Gramsci 1971). This replaced the need for a forced class rule through providing an integrating culture in the broadest sense. Such ideologies included art, religion, literature and law, avoiding reduction to economic analysis.

They ensured the hegemony of a dominant class over the masses by means of a consensual historical bloc. The important notions of practice and action - the organizational attributes of ideology - are incorporated in this view which, at the same time, loses its ephemeral link to specific historical circumstances.

Louis Althusser further encouraged Marxists to analyse ideology on its own merits (Althusser 1984). He retained its dissimulative function but emphasized its integrative one, 'interpellating' individuals and bestowing on them a subjective (ostensibly free), as well as subjected, identity within their society. Thus recognition and misrecognition operate side by side to sustain a given society. All classes produce ideology, which is now seen as a permanent material phenomenon in societies, existing objectively in the form and substance of social practices and their rituals. In that sense, ideology is neither a distortion nor an illusion. It is an 'imaginary' representation of the real, yet also a 'lived' relation between individuals and their conditions of existence, a 'new reality'. Elaborating on earlier Marxist analyses, the state in particular was seen to employ 'ideological state apparatuses' in order to ensure the dominance of the established order and its relations of production.

2 Generalizing ideology and the problem of knowledge

The limitations of the Marxist family of theories in respect of ideology lay in restricting it to a highly specific mode of thinking about politics and society. The concept of ideology thus became attached to a number of perspectives: to a reflection in thought of historically situated classes; to an inverted or misperceived form of consciousness, detached from social and material reality; and to the pejorative converse of social truth, which would either replace ideology or existed, as scientific knowledge, in parallel to it. Although none of these perspectives was unequivocally held by any one thinker in the Marxist tradition, all were used to some effect within more 'vulgar' *marxisant* discourses about ideology. Outside the Marxist tradition, although in partial reaction to it, these themes were picked up or discarded in different measures. Althusser importantly recognized the relative detachment of ideology from a material base, a move which has both been welcomed and deplored. His views thus coalesced with those who were analysing ideology as a phenomenon worthy of study in its own rights rather than merely as an epiphenomenon.

Karl Mannheim had already been taking a similar direction, although for him it grew out of dissatisfaction with some of the problems incumbent in Marxism. In his seminal *Ideology and Utopia* (1929) he retained the Marxist insight into the social and historical origins of thought, linking the meaning of an idea not to the laws of logic but to its genesis. However, rather than subscribing to the view that ideology was a reflection of a particular historical distortion, Mannheim suggested that it was a pluralistic product of diverse social groups undergoing common experiences. Knowledge was a cooperative process, but political discussion was further characterized by an unmasking of rationalized situational motivations which were attributes of a collective unconscious. These could adopt two forms: ideology and utopia. Ideology referred to the interest-bound thought of ruling groups and had conservative, stabilizing consequences. It was divided into two conceptions: a particular conception, primarily on the psychological level, in which ideology is consciously recognized as a lie or an error; and a total conception, on a socio-historical level, in which the entire *Weltanschauung* of a group is involved. Utopia referred to an emphasis on the transformation of a society and, unwittingly, on misdiagnosing it by identifying only its negative features (see Utopianism §1). Mannheim's distinction is itself debatable, as both ideology and utopia could be complementary forms of an unconscious, action-oriented interpretation of a historical reality.

Mannheim raised the problem of analysing ideology as a fundamental methodological issue of the social sciences. Having abandoned the Marxist ontology of a true consciousness which would emerge once social contradictions were negated, and having unmasked Marxism itself as ideological in the total sense, Mannheim's general theory of ideology paved the way for a new epistemology expressing the notion that 'all historical knowledge is relational knowledge, and can only be formulated with reference to the position of the observer' (1929: 71). The problem became how to overcome the relativism implicit in the recognition that all social points of view were ideological, while eschewing any adherence to ultimate values (see Relativism). That was made especially difficult because an older epistemology was still attached to relativism, an epistemology which assessed each assertion from the intrinsic perspective of the logical and universal truths it contained. By replacing relativism with relationism - an appreciation, influenced by the holism of psychological Gestalt theory, of the systemically reciprocal interconnections among all historically and spatially located thinking - an evaluative procedure could emerge which surmounted the limitations of ideology. Intellectuals, those who were able to cut loose from their historical

and social situatedness, could incorporate conflicting viewpoints in a flexible and dynamic relationism, assessing their scope and validity. An initially non-evaluative relativism, incapable of discriminating among various static and eternal views, would make way for an inescapably evaluative sociology of knowledge. The latter would embrace a more modest conception of truth, critically held and approximate, based on the multiple perspectives in a given society and seeking comprehensiveness and fruitfulness of understanding (see Sociology of knowledge).

3 Ideology, political science and culture

Ultimately, Mannheim too sought to transcend ideology, accepting its inevitability but arguing, somewhat unsatisfactorily, for a new objectivity alongside it. Mid-twentieth-century political scientists, while heavily influenced by Mannheim’s concentration on ideology as an ubiquitous social and political given, had no such epistemological and critical aspirations, and approached their subject-matter both as a conscious, empirically ascertainable product of politico-social groups, and as capable of rational assessment by external analysts. Furthermore, the location of ideology as the product of groups in specific historical situations was played down through employing a neo-positivist, individualist methodology. In particular, US political science adopted a functional approach, concentrating on the concrete phenomenon of ideologies rather than the category of ideology, and regarding them as necessary, often socially beneficial, and distinguishable from other types of thinking about society or in society (see Functionalism in social science §2). Political scientists, however, exhibited two different interpretations of ideologies. One, as exemplified in the work of Daniel Bell (1960) and Giovanni Sartori (1969), restricted them to doctrinaire, highly consistent and closed systems of ideas cum ’social levers’, usually infused with extremist passion. These rationalist, deductive constructs, frequently having rhetorical purposes, were contrasted with pragmatic and empirical political belief systems. Given that distinction, it was possible to foresee the ’end of ideology’ as passion and imperviousness to temporal and spatial influences apparently gave way, at least in the West, to pluralism and an exhaustion with the great ’isms’. This approach underplayed both the ideological components of so-called ’pragmatism’ and the emotive elements endemic to political argument.

The second and more prevalent interpretation conceived of ideologies as a general and omnipresent category encompassing all relatively coherent sets of cultural symbols - ideas, beliefs and attitudes - that are action-oriented, and whose function it is to interpret the political system and to direct and justify public policy aimed towards preserving or changing political institutions and processes. That approach emphasizes ideologies as consciously-held views of aggregates of individuals or of ruling groups who impose them on such aggregates. Individual conduct is thus organized and integrated by cognitive rational constructs - regarded as simplified selection rather than distortion - abetted by evaluation of the options for action in the social world. This return to the scientific study of ideology as a value-free activity differs from the Marxist juxtaposition of ideology and science, because the truth-distortion dimension is removed from the consideration of ideology, and science is employed to comprehend, not to replace, ideology. Such a concrete and empirical view of ideologies involves typologies frequently based on a left-right continuum, ranging from communism through socialism, social democracy, liberalism, conservatism and fascism. Newer belief systems such as feminism, nationalism or ecologism fit uneasily into that rubric, nor can they always claim the comprehensiveness expected of systemic political ideologies. Others, such as anarchism or libertarianism, cut across conventional boundaries. Psychological offshoots of this approach relate to the formation of attitudes and their classification.

The structural anthropology of Claude Lévi-Strauss (§6) and cultural anthropology of Clifford Geertz (1964) demonstrate affinity, although not identity, with such systemic, self-substantiating conceptions of ideology, as does Althusser’s analysis. For Lévi-Strauss myths, and ideology as modern myth, manifest an internal, self-contained logic and are endowed with unconsciously held meaning. Unlike Althusser, he regarded ideology as a ’thought-of’ order external to objective reality, but like Althusser, he argued that they made sense only in their relationships with ’lived-in’ orders. For Geertz, an ideology is an ordered system of cultural symbols, both cognitive and expressive, that is interpenetrated by social and psychological processes and in turn organizes them into patterns of meaning. Its systemic features are ’maps of problematic social reality’ which enable purposive action and perform an integrative function, irrespective of whether those maps are accurate or not. The relationship of an ideology to the real world is no longer as centrally challenging, methodologically speaking, as it is for Marxist theories. It is primarily assessed as a cultural whole performing necessary functions.

4 Words, concepts and interpretation

In the USA political science developed its conceptions of ideology in almost complete isolation, not only from a consideration of the unconscious elements of ideology which separate political actors and their thoughts from the perception of the analyst, but also from hermeneutic arguments which were beginning to affect the concept of ideology through a denial of the very epistemological validity of such separation. Hermeneutics transformed the view of ideology by refusing to ask questions concerning how to gain access to reality, or how to avoid the illusions of human perception (see Hermeneutics). Instead, social and historical actions and utterances are seen as bereft of objective meaning, and are to be experienced through the interpretive medium of the observer. The polysemy of words necessitates an appreciation of the dual contexts of both their authors and their readers or consumers in attaching a determinate meaning to texts. In allowing for a socio-cultural decontextualization of texts 'as wordless and authorless' objects, and for the 'surplus of meaning' that words carry, theorists such as Paul Ricoeur recognize the function of ideology as effecting an inevitable closure of interpretation without which consciousness, self-representation and meaning, as well as social integration, are unattainable. Although it is possible to achieve critical distance from ideological closure, it will always - echoing Mannheim - be incomplete. The treatment of ideology is now virtually detached from the circumstances of its production, but is enriched by a sophisticated appreciation of the coexistence of intentional and unconscious import. Here, ideological unconsciousness is not necessarily distorted consciousness, the permanence of ideologies represents the infinite variety of the human imagination, yet the tentativeness of understanding is underlined by a consciousness of its own value-laden temporality.

The relationship of hermeneutics to the examination of language is crucial. The analysis of ideology was advanced by a harnessing of the insights of Ludwig Wittgenstein (§11) and Ferdinand de Saussure. Wittgenstein's notion of language games assisted in crystallizing the view that words could shape, rather than reflect, reality, albeit within a network ineluctably reflecting grammatical rules. Saussure's semiology - the study of signs - likewise suggested that the relationship between signs and meaning was arbitrary, but that meaning could be imposed on words by organizing them in specific patterns. In this systemic perspective, the mutual proximity of words established their meanings. The critical function of ideology was diminished by regarding it as a socially produced text in which meaning was related to usage and to form, impacting on, not merely representing, practices in the 'real' social and historical world. However, ideology was itself re-established as a symbolic human practice of direct interest to scholarship. Discourse analysis concentrates on these findings, but also links the study of language as a social phenomenon with the Marxist theme of ideology as domination. Successful closures of meaning in language reflect the power relationships of its users, and language is hence a medium through which political legitimacy may be accorded to groups, overtly or frequently through unconscious dissimulation. This method can be applied not only to the broad ideological families that attempt to control public policy making but to a wide range of discourse situations in which meanings are exchanged and displaced in specific, everyday, conversational contexts (see Discourse semantics).

From the perspective of political theory it is possible to combine the focus of political scientists on the overt and systemic contents of ideological debate, with an interest in language and meaning and in the decoding of unconscious ideological messages. W.B. Gallie's notion of the essentially contested nature of political concepts presented them as indeterminate because of the impossibility of agreeing on their normative components (Gallie 1955-6). However, they are also indeterminate due to their inability to contain all their logically entailed meanings simultaneously. Hence, ideologies may be seen as clustered patterns of a wide spectrum of political concepts, such as liberty, justice or power, each of which is necessarily decontested, reflecting a specific set of historical and spatially located social meanings without which political decision making cannot be effected. A potentially unrestricted universe of meanings is also limited by logical and morphological constraints operating in all ideological systems, as political concepts are decontested by their users through deliberately or unconsciously arranging their proximity to other political concepts. Different conceptual patterns may be produced by any social group, although these still may be usefully classed according to loose family resemblances that approximate major ideological traditions such as liberalism, conservatism or socialism. Decontesting also expresses the inevitability of political power which, importantly, need not be exploitative. Ideologies hence complete over their claims to confer legitimate meanings on political concepts and words, and map and recommend thought practices which in turn affect political action. Ideology nevertheless is not coterminous with political thought, but is one of its analytical dimensions: a particular interpretive handling of political concepts which are suspended between logical indeterminacy and cultural determinacy.

MICHAEL FREEDEN

References and further reading

Althusser, L. (1984) *Essays on Ideology*, London: Verso.(Referred to in §1. A significant reappraisal of ideology within the Marxist tradition.)

Bell, D. (1960) *The End of Ideology: On the Exhaustion of Political Ideas in the Fifties*, Glencoe, IL: Free Press. (The most prominent statement on the subject.)

Freeden, M. (1996) *Ideologies and Political Theory: A Conceptual Approach*, Oxford: Clarendon Press.(An elaboration of the argument in §4.)

Gallie, W.B. (1955-6) ’Essentially Contested Concepts’, *Proceedings of the Aristotelian Society* 56: 167-98.(A seminal analysis of political concepts referred to in §4.)

Geertz, C. (1964) ’Ideology as a Cultural System’, in D.E. Apter (ed.) *Ideology and Discontent*, New York: Free Press, 47-76.(A groundbreaking article, referred to in §3.)

Gramsci, A. (1971) *Selections from Prison Notebooks*, ed. Q. Hoare and G. Newell-Smith, London: Lawrence & Wishart.(The influential text on ideology as hegemony.)

Larrain, J. (1979) *The Concept of Ideology*, London: Hutchinson.(A sophisticated overview, supportive of the critical function of ideology.)

Mannheim, K. (1929) *Ideology and Utopia*, London: Routledge, 1936.(A modern classic on the subject, despite a confusing usage of terminology.)

Marx, K. and Engels, F. (1846) *The German Ideology*, ed. C.J. Arthur, London: Lawrence & Wishart, 1974.(The foundational Marxist analysis of ideology.)

Mullins, W.A. (1972) ’On the Concept of Ideology in Political Science’, *American Political Science Review* 66: 498-510.(A good illustration of the approach of contemporary political science.)

Sartori, G. (1969) ’Politics, Ideology, and Belief Systems’, *American Political Science Review* 63: 398-411.(An example of confining ideology to closed, non-empirical systems.)

Thompson, J.B. (1984) *Studies in the Theory of Ideology*, Oxford: Polity Press.(A perceptive assessment of a range of current views on ideology.)

Ikhwan al-Safa’

The philosophy of the group of Arab philosophers of the fourth or fifth century AH *(tenth or eleventh century* AD*) known as the Ikhwan al-Safa’ (Brethren of Purity) is a curious but fascinating mixture of the* Qur’anic*, the Aristotelian and the Neoplatonic. The group wrote fifty-two epistles, which are encyclopedic in range, covering matters as diverse as arithmetic, theology, magic and embryology. Their numerology owes a debt to Pythagoras, their metaphysics are Aristotelian and Neoplatonic and they incorporate also a few Platonic notions into their philosophy. The latter, however, is more than a mere synthesis of elements from Greek philosophy, for it is underpinned by a considerable* Qur’anic *substratum. There are profound links between the epistemology and the soteriology (doctrine of salvation) of the Ikhwan, and it would not be an exaggeration to say that the former feeds the latter. In the history of Islamic philosophy the Ikhwan illustrate a group where the Aristotelian and the Neoplatonic clash head-on and where no attempt is made to reconcile competing and contradictory notions of God, whom the Epistles treat in both* Qur’anic *and Neoplatonic fashion. The final goal of the Ikhwan is salvation; their Brotherhood is the ship of that salvation, and they foster a spirit of asceticism and good living accompanied by ‘actual knowledge’ as aids to that longed-for salvation.*

1 Life and works

The Arabic name Ikhwan al-Safa’ has been translated as both ‘Brethren of Purity’ and ‘Brethren of Sincerity’. Both are possible, though the former is probably to be preferred because of the emphasis throughout the group’s writings on the concept of purity achieved via a life of asceticism and virtuous living.

Little firm information is available about their exact identities, their lives and the precise time during which they flourished. Most scholars agree, however, that they lived in Basra in the fourth or fifth century AH (tenth or eleventh century AD); beyond that there has been much diverse speculation. Their own thought and philosophy is enshrined in fifty-two epistles (*rasa’il*) of varying lengths which are encyclopedic in their scope and cover a vast number of topics. Formally, these epistles divide into four major sections: the first fourteen deal with the mathematical sciences, the next seventeen are on the natural sciences, a further ten deal with the psychological and rational sciences and the final eleven come under the heading of theological sciences. It should be noted that the Ikhwan’s usage of these divisions is much broader in range than might be expected at first sight. For example, the last of the epistles grouped under the heading ‘theological sciences’ deals with magic and related subjects. What may broadly be said to link all the epistles, however, is a mixed Aristotelian and Neoplatonic substratum, though it must be stressed here that the epistles of the Ikhwan al-Safa’ are more than just a synthesis of Aristotelian and dominant Neoplatonic themes. The incorporation by the Ikhwan of syncretic philosophical and theological themes, motifs, elements and doctrines in their writings was done with a particular soteriological purpose (see §4). Their eclectic borrowing was done with a view to bolstering the doctrine of purity which their name so neatly reflects.

2 Metaphysics

The metaphysics of the Ikhwan al-Safa’ are built upon those of Aristotle and Plotinus, though it must be emphasized that it is a Middle Eastern version of Aristotelianism and Neoplatonism which we encounter when we read the *rasa’il* of the Ikhwan. In the first place, their terminology is infused with such terms as matter and form, substance and accidents, the four causes and potentiality and actuality. Their usage of such terms, however, does not always adhere to the classical Aristotelian paradigm or usage. The development of terminology is often in a Neoplatonic direction. For example, the Ikhwan held in one place that substance was something which was self-existent and capable of receiving attributes. We recognize here a description akin to Aristotle’s usage of the word ‘substance’ in the *Metaphysics*. But elsewhere, confusingly, form is divided into two kinds, constituting and completing; constituting forms are called substances and completing forms are called accidents. Similarly, the Ikhwan adopted a fourfold terminology of causes - material, formal, efficient and final - but the shades of astrology and Neoplatonism hang heavily over at least two of the examples of these four causes which they provide. They say that the material cause of plants is the four elements of fire, air, water and earth, and that their final cause is to provide food for animals: both of these ideas are recognizably Aristotelian in their orientation, but the Ikhwan then go on to suggest that the efficient cause of plants is the power of the Universal Soul and that their formal cause has complicated astral elements!

It is, however, the Neoplatonic elements which dominate the articulation of all thought in the writings of the Ikhwan al-Safa’ and their metaphysics are no exception. The latter are imbued in particular with the Neoplatonic concepts of emanation and hierarchy. By contrast with the simple triad of Plotinus, which comprised the three hypostases of The One or The Good, Intellect and Soul, with the lower eternally emanating from the higher entity, the Ikhwan elaborated this into an emanationist hierarchy of nine ‘members’, hypostases or levels of being, as follows: the Creator, the Intellect, the Soul, Prime Matter, Nature, the Absolute Body, the Sphere, the Four Elements and the Beings of this world in the three divisions of mineral, plant and animal. In such a hierarchical profusion we can perhaps see the ghosts of Iamblichus and Proclus, who also multiplied the hypostases about which they wrote. It is noteworthy that for the Ikhwan, and in contrast to the view of Plotinus, matter becomes a full part of the emanationist hierarchy and is regarded in a positive light. Furthermore, and this time in a very Neoplatonic way, God in the Ikhwan’s scheme entrusts the movement of the world and the spheres to the Universal Soul, and it is the latter which channels God’s gifts finally into Matter itself (see Neoplatonism).

The Neoplatonic dimensions of the thought of the Ikhwan have profound implications for their view of God. The picture which they present of the deity in their epistles is a confused and ultimately contradictory one. No attempt is made to reconcile what is in fact irreconcilable. On the one hand, the Ikhwan present a God at the top of a complex emanationist hierarchy who is unknowable in the classic Neoplatonic sense. On the other hand, the Ikhwan present a Qur’anic God who is a guide and a help, and who is invoked at the end of many of the epistles as one who will grant success in correct action and show his people the path of righteousness. The majority of epistles also invoke God with the traditional Islamic *basmala*, ‘In the name of God, the Merciful, the Compassionate’. However, God’s power, as noted above, seems to be ‘shared’ in some way when it is exercised via the Universal Soul. To what extent, one may reasonably ask, does that compromise the traditional Islamic view of God? Furthermore, to what extent do the recognizably Islamic features in the Ikhwan’s portrait of God prevent that deity being considered as a total mirror of Plotinus’ One?

The metaphysics of the Ikhwan al-Safa’ must therefore be regarded as *sui generis*. Their mixing of Aristotelian and Neoplatonic elements had profound implications both for their theology and the coherence of their philosophy. Contradictions abound; if reasons be sought for this, it is worth remembering one theory, promulgated by A.L. Tibawi (1955), that the epistles are akin to the minutes taken during the deliberations of a learned society, meeting on many occasions over a period of years. This would account for both contradiction and repetition. We know from the epistles themselves that the authors urged their brothers to meet specially at set times, in closed sessions.

3 Epistemology

Thus far in this article, nothing has been said about the impact of Platonic thought on the epistles of the Ikhwan. This is because the Brethren revere the Platonic *hero* rather more than they revere purely Platonic *philosophy*. Socrates is admired as a great and wise philosopher who knew how to meet death bravely. However, some Platonic imagery does permeate the epistles (see Platonism in Islamic philosophy). The most notable image is that of the body constituting a prison for the soul. The Ikhwan indeed compare the soul in the body to the state of a man imprisoned inside a lavatory: the body’s blemishes and sins are like the filth in the lavatory. It is clear that the Ikhwan were familiar with Plato’s doctrine of Forms or *ideai*, since they quote a speaker saying that the different types of animal in the world simply mirror those in the world of the spheres and the heavens (see Plato). However, this is not a doctrine for which the Ikhwan seem to have had much use, for they neither discuss nor elaborate upon it.

It is, therefore, unsurprising that the Ikhwan’s epistemology differs quite radically from that of Plato. The latter looked forward to a state of real knowledge achieved when the soul was separated from the body; but in the soteriology and epistemology of the Ikhwan, one could gain some knowledge of the divine in this world to help one reach Paradise. Indeed, they present their epistles to the world as a body of just such knowledge. For them, learning was much more than mere recollection or reminiscence. They held that the soul was ‘potentially knowledgeable’ and, with instruction, could become ‘actually knowledgeable’. That instruction should be via the senses, the intellect and logical deduction, and they stressed that we could know nothing without the senses. This is indeed a far cry from Plato’s well-known suspicion of evidence or knowledge gleaned via the senses, and his overwhelming exaltation of the intellect.

4 Soteriology

The mass of information - philosophical, theological and other - adumbrated in such an encyclopedic manner in the epistles of the Ikhwan al Safa’ is probably incomprehensible as a totality unless one bears in mind the driving force which lies at the heart of the epistles themselves. The Ikhwan did not compile the epistles from a pure love of knowledge and for no other reason. The magpie eclecticism with which they surveyed and utilized elements from the philosophies of Pythagoras, Plato, Aristotle and Plotinus, and religions such as Christianity, Judaism and Hinduism, was not an early attempt at ecumenism or interfaith dialogue. Their accumulation of knowledge was ordered towards the sublime goal of salvation. To use their own image, they perceived their Brotherhood, to which they invited others, as a ‘Ship of Salvation’ that would float free from the sea of matter; the Ikhwan, with their doctrines of mutual cooperation, asceticism and righteous living, would reach the gates of Paradise in its care.

What, then, did it mean for the Ikhwan al-Safa’ to ‘do philosophy’? It did not mean to throw off the religious constraints of the Qur’an and to become pure rationalists. Though they often use the Qur’an as a cloak to disguise their Neoplatonism, one cannot ignore the massive Qur’anic substratum elsewhere in their writings, which has no such intent. ‘Doing philosophy’ did not mean either the uncritical acceptance of the data from a variety of sources such as Pythagoras, Plato and Aristotle, not to mention Plotinus, even though they were profoundly influenced by at least three of these four ancient masters and it is no misnomer to describe the Ikhwan as ‘Muslim Neoplatonists’. Philosophy, for the Ikhwan, was still the handmaiden of a precise theological goal: salvation for the soul. Their eclecticism and tolerance provided them with a unique methodology for the achievement of that goal. Thus they searched out the texts of other creeds and the philosophies of non-Muslim sages in search of materials which might bolster their own ethics of purity and asceticism. Their intellectual heroes were Socrates and Jesus as well as Muhammad. Above all, knowledge and philosophy were always soteriological tools and never ends in themselves.

See also: Islamic theology; Mystical philosophy in Islam; Neoplatonism in Islamic philosophy

IAN RICHARD NETTON

References and further reading

Diwald, S. (1975) *Arabische Philosophie und Wissenschaft in der Enzyklopädie Kitab Ihwan as-Safa’ III: Die Lehre von Seele und Intellekt (Arab Philosophy and Science in the Book of the Brethren of Purity III: Teachings on the Soul and the Intellect)*, Wiesbaden: Harrassowitz.(A German translation of the third section of the *Rasa’il* with extensive notes.)

Goodman, L.E. (1978) *The Case of the Animals versus Man Before the King of the Jinn*, Boston, MA: Twayne Publishers.(An English translation of a major portion of the second section of the *Rasa’il* with extensive introduction and notes.)

Ikhwan al-Safa’ (1957) *Rasa’il Ikhwan al-Safa’ (Epistles of the Brethren of Purity)*, Beirut: Dar Sadir, 4 vols. (The complete text of the fifty-two epistles in the original Arabic.)

Marquet, Y. (1975) *La philosophie des Ihwan al-Safa’ (The Philosophy of the Brethren of Purity)*, Algiers: Société Nationale d’Édition et de Diffusion.(A major study by France’s leading expert in the field.)

Nasr, S.H. (1978) *An Introduction to Islamic Cosmological Doctrines*, revised edn, London: Thames & Hudson. (Contains a major section on the cosmology of the Ikhwan.)

Netton, I.R. (1982) *Muslim Neoplatonists: An Introduction to the Thought of the Brethren of Purity (Ikhwan al-Safa’)*, London: Allen & Unwin; paperback edn, Edinburgh: Edinburgh University Press, 1991.(A major introduction in English to the thought of the Ikhwan.)

Netton, I.R. (1996) ‘The Brethren of Purity’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 15, 222-30.(Concise and clear outline of their views.)

Tibawi, A.L. (1955) ‘Ikhwan as-Safa’ and Their Rasa’il: A Critical Review of a Century and a Half of Research’, *Islamic Quarterly* 2 (1): 28-46.(A very useful and neat survey of Ikhwan scholarship up to 1955.)

Il’enkov, Eval’d Vasil’evich (1924-79)

Eval’d Il’enkov advanced a distinctive brand of Hegelian Marxism that was influential in the rejuvenation of Soviet philosophy after Stalin. Il’enkov draws on Hegel and Marx to argue that non-material phenomena can exist as genuine features of objective reality independent of the consciousness and will of individuals. Il’enkov argues that the existence of such phenomena, conceived as objectifications of human social activity, is central to the explanation of the nature and possibility of the human mind. The world becomes a possible object of thought through its ’idealization’ by activity, and children attain mental capacities in the full sense only through the appropriation of the ideal as it exists objectified in ’humanity’s spiritual culture’.

Il’enkov’s defence of the reality of culture represents a critique of positivism and scientism, a critique he pursued in many other writings. A tireless opponent of reductionist theories of mind and ’biological determinist’ theories of human development, he advanced a view of persons as socially constituted beings and stressed socialism’s obligation to create the circumstances in which human beings may develop their almost limitless potential. Like many in the post-Stalin era, his criticism of the Soviet philosophical establishment takes the form of a call for a genuinely orthodox form of Marxism, faithful to the spirit of Marx’s thought.

1 Life

Perhaps the most inventive of Russian Marxist philosophers, Eval’d Il’enkov made a significant contribution to the renewal of Soviet philosophy after the Stalin period. He advanced a form of Hegelian Marxism that stressed the active powers of human beings to transform themselves and their world through the objectification of culture. He was a relentless critic of reductionism and scientism, which he believed fed the dehumanizing elements of Soviet ideology, and his works upheld the importance of critical philosophical reflection in opposition to dogma.

Il’enkov was born in Smolensk, though he lived most of his life in Moscow. He was educated at the Moscow Institute of Philosophical and Literary Studies (IFLI) and Moscow State University (MGU), where he defended his ’candidate’s dissertation’ in 1954. During the ’thaw’ in the early 1960s, he published the works which established him as a leading voice in the Soviet philosophical community. They include a treatise on Marx’s method, *Dialektika abstraktnogo i konkretnogo v ‘Kapitale’ Marksa (The Dialectics of the Abstract and the Concrete in Marx’s ’Capital’)* (1960), and an influential essay, ‘Ideal’noe’ (’The Ideal’) (1962), in which Il’enkov argues that non-material phenomena are genuine constituents of objective reality and advances a distinctive account of the relation of mind and world. Il’enkov continued to explore these themes in many subsequent writings spanning a variety of fields, including history of philosophy, education and aesthetics. Like many of his contemporaries, Il’enkov experienced political difficulties. Accused of the heresy of ’epistemologism’, he lost his position at MGU in 1955. The rest of his career was spent at Moscow’s Institute of Philosophy, though he continued to have an uneasy relation with the Soviet philosophical establishment, which was suspicious of the critical and humanistic elements of his work. He died in 1979, by his own hand.

2 Dialectics

Il’enkov’s early work on dialectics develops the so-called ’method of ascent from the abstract to the concrete’ sketched by Marx in his ‘Introduction to a Critique of Political Economy’. Il’enkov argues that in both natural- and social-scientific inquiry, the objects of cognition are ’organic’ wholes, consisting of parts standing in relations of mutual determination. These wholes develop through the tensions, or dialectical contradictions, between their parts. Il’enkov argues that cognition proceeds by isolating the ’principle of development’ of the whole. This phenomenon - the ’cell’ or ’unit’ - represents the part, the evolution of which necessitates the development and mutual determination of the other parts. The ’unit’ is thus a ’concrete universal’ that explains the whole’s development through the necessary interrelations of its developing parts. Cognition therefore represents a movement from the *abstract* (from the abstraction of the unit) to the *concrete* (to a conception of the whole as a ’unity in diversity’ of essentially related components). In this process, historical and logical analyses coincide, for to represent the history of the object is to chart the necessary logic of its evolution. Il’enkov argues that Marx’s analysis of capitalism in *Capital* illustrates this method. By isolating the concept of commodity and tracing its development, Marx reconstructs the logic of capitalism, the principles of its existence and necessary transformation.

Such esoteric epistemology may seem an unlikely focus for scholars seeking to reanimate Marxism in Eastern Europe. But for Il’enkov (and others such as A.A. Zinov’ev and the Czech philosopher Jindřich Zelený), the analysis of ’Marx’s logic’ was an important catalyst because it demanded that Marx be treated, not as an authoritative source of state ideology, but as a philosopher whose ideas could be understood only through a critical engagement with the history of Western traditions in philosophy. Approaching Marx this way showed the virtues of scholarship, rather than partisanship. In addition, Il’enkov’s work on dialectics had a political subtext. Il’enkov contrasts Marx’s approach with empiricism. The latter typically construes the knowing subject as arriving at a conception of the world by a movement from concrete to abstract: the subject makes sense of perceptual data through the formation and application of abstract concepts and general laws that facilitate the organization and prediction of experience. Il’enkov was convinced that empiricist conceptions, often ill-conceived and poorly articulated, were prevalent in Soviet philosophy and contributed to positivistic infatuations with science as an all-powerful force. Il’enkov’s work on method thus represents a critique of empiricism in Soviet thought and scientism in Soviet culture.

3 The problem of the ideal

Il’enkov’s most original contribution to philosophy is his treatment of ‘the problem of the ideal’. What place should a materialist find for meaning, value and universals in the material world? Many of Il’enkov’s Soviet contemporaries portrayed such phenomena as mental in kind and argued that mental states were ultimately reducible to physical states of the brain: everything seemingly non-material is in the head. In contrast, Il’enkov maintains that ideal phenomena can have a supra-individual existence as aspects of the world as it is independent of individual consciousness. Drawing on Hegel and Marx, Il’enkov insists that there is nothing mystical in the objective existence of the ideal, for ideal phenomena are objectifications of human activity. In the course of the transformation of nature by human action, meaning and value are written into nature:

> ‘Ideality’ is like a peculiar stamp impressed on the substance of nature by social human life activity; it is the form of the functioning of a physical thing in the process of social human life activity. All things incorporated into the social process acquire a new ’form of existence’, an ideal form, quite distinct from their physical nature.
>
> (1979: 148)

Il’enkov argues that our natural environment is organized or ‘humanized’ by activity: our world is full of physical entities - artefacts, signs, symbols, models - made meaningful through action. The edifice of objectified ideality represents (in Marx’s words) humanity’s ’inorganic body’ or ’spiritual culture’, embodied in our environment as ’thought in its otherness’:

> Outside the individual and independently of his consciousness and will exists not only *nature*, but also the socio-historical environment, the world of things, created by human labour, and the system of human relations, formed in the process of labour. In other words, outside the individual lies not only nature as such (‘in itself’), but also *humanized* nature, nature re-made by human labour. From the point of view of the individual, ’nature’ and ’humanized nature’ merge together into the ’surrounding world’.
>
> (1964: 41-2)

This view is part of a dynamic vision of human beings creating themselves through the creation of culture. Transformed by activity, our environment confronts us as rich with significance; in virtue of this significance, we act further to transform our world, endowing it with new meaning and precipitating further transformations. Human beings must constantly adapt to their changing environment, acquiring skills necessary to orientate themselves within it. And in this dialectic, human beings gradually understand and harness the forces which define them and their world. Il’enkov sees communism as the project of attaining control over these powers.

Il’enkov’s account of ideality is intended to solve fundamental issues of the relation of mind and world. He argues that ‘nature "as such" is given to the individual only insofar as it is transformed into an object [*predmet*], into the material or means of production of material life’ (1964: 42). There is thus a sense in which all of nature is ’idealized’ by human activity: no object speaks to us unless it is brought within the compass of humanity’s spiritual culture. The primary object of thought is thus nature humanized or idealized.

In addition, Il’enkov’s position yields a distinctive view of thought itself, which is conceived as the capacity to inhabit an idealized environment: a thinking body is one which shapes its activity to the norms that constitute our humanized environment. Thought is thus construed as a species of movement, the ability to conform to and manipulate meanings as they are formed in the flux of social being. Il’enkov significantly does not represent this ability as innate in the individual: it is as socially constituted as the environment the thinking thing inhabits. Human children become thinking things in so far as they acquire the ability to orientate themselves in their environment by internalizing the forms of activity of their community (note here the continuity with Vygotskii’s psychology - see Vygotskii, L.S. §2). Thus, in contrast to the individualism of traditional epistemology, Il’enkov does not believe that each individual mind finds the world anew for itself. We enter a world which history has made cognizable and learn to find our way through the agency of others who help us appropriate our spiritual culture.

Il’enkov’s solution to the problem of the ideal may thus be seen to follow from the application of his dialectical method to the question of the relation of thinking and being. For Il’enkov, human activity plays the role of the ’unit’, the exploration of which reveals the nature and possibility of both halves of the cognitive relation - object (nature humanized by activity) and subject (the ’thinking body’ that inhabits the humanized environment). Each is born of activity and their unity is sustained in and through activity.

4 Popular and polemical writings

Il’enkov believed that philosophers should be moralists, and he produced a significant number of popular and polemical writings. Many address questions of education. Developing his view that human individuals are socially constituted beings, he argued that a socialist government has an overriding responsibility to create a culture which nurtures the abilities of all individuals. Il’enkov was critical of the trend towards specialization in the Soviet education system, arguing that it should be guided by Marx’s vision of the ’all-round individual’ whose capacities are universal in nature. Some of Il’enkov’s most striking writings on education were inspired by Aleksandr Meshcheriakov’s work on the education of blind-deaf children, which Il’enkov took to illuminate the social construction of cognitive processes, as well as the duties of a radical education system.

As fascination with the ’scientific-technological revolution’ grew in the USSR through the 1960s and 1970s, Il’enkov’s hostility to positivism became yet more pronounced. He was scathingly critical of thinkers who portray our mental lives as nothing more than a series of events in our brains, and he attacked the idea that the talents and aptitudes of human beings are genetically determined. The thirst to represent human beings as sophisticated machines, he suggested, was a symptom of the increasing technologization of social life. The latter was supposedly to be the theme of his final book, *Leninskaia dialektika i metafizika positivizma (Leninist Dialectics and the Metaphysics of Positivism)* (1980), which was to argue that although Soviet philosophers constantly reiterated Lenin’s critique of the Russian empiriocriticist, Aleksandr Bogdanov, the technocratic vision of social organization advanced by Bogdanov in fact pervaded Soviet culture. Unfortunately, this message was obscured when the book was cut for posthumous publication.

Il’enkov’s popular and polemical writings have aged badly, and though they have historical interest, they are not philosophically deep. Moreover, not all of these texts can be portrayed as progressive in orientation. His Hegelian attacks on formal logic, for example, provoked censure from thinkers anxious to animate Soviet philosophy by forging links with the analytic tradition. However, despite this, and the strident tone of much of his writing, with its call for a return to true orthodoxy after the perversions of Stalinism, there is no doubting the depth of Il’enkov’s philosophical vision, particularly in his early work, and his importance in preserving a cultural tradition of philosophical criticism in Soviet Russia.

See also: Marxist philosophy, Russian and Soviet

DAVID BAKHURST

List of works

Il’enkov, E.V. (1960) *Dialektika abstraktnogo i konkretnogo v ’Kapitale’ Marksa*, Moscow: Akademiia nauk; trans. S. Syrovatkin, *The Dialectics of the Abstract and the Concrete in Marx’s ’Capital’*, Moscow: Progress, 1982.(Il’enkov’s influential work on Marx’s method; said to have been heavily censored before publication.)

Il’enkov, E.V. (1962) ’Ideal’noe’ (The Ideal), in *Filosofskaya ėntsiklopediia (Philosophical Encyclopedia)*,

Moscow: Sov’etskaya ėntsiklopediia, vol. 2, 219-27; amended repr. in Il’enkov (1974), 183-210; and (1991), 212-28.(The first presentation of Il’enkov’s philosophical anthropology in which the concept of activity is invoked to explain the very possibilty of the relation between subject and object.)

Il’enkov, E.V. (1964) ’Vopros o tozhdestve myshleniia i bytiia v domarksistskoi filosofii’ (’The Question of the Relation of Thinking and Being in Pre-Marxist Philosophy’), in *Dialektika - Teoriia poznaniia (Dialectics - The Theory of Knowledge)*, Moscow: Nauka, 21-54. (A history of conceptions of the relation of mind and world from Descartes and Spinoza through the German idealists to Marx.)

Il’enkov, E.V. (1968) *Ob idolakh i idealakh (Of Idols and Ideals)*, Moscow: Politizdat.(A collection of previously published writings, popular in tone. After attacking scientism in the opening chapters, Il’enkov develops a humanistic philosophy of education.)

Il’enkov, E.V. (1974) *Dialekticheskaia logika*, Moscow: Politizdat; 2nd edn, revised and enlarged, 1984; 1st edn trans. H. Campbell-Creighton, *Dialectical Logic*, Moscow: Progress, 1977.(A history of philosophy from Descartes to Marx; akin to the materialist phenomenology of spirit.)

Il’enkov, E.V. (1979) ‘Problema ideal’nogo’ (’The Problem of the Ideal’), *Voprosy filosofii (Questions of Philosophy)*, 6: 145-58; 7: 128-40; amended repr. as ’Dialektika ideal’nogo’ (’The Dialectic of the Ideal’) in Il’enkov (1984), 8-77; and (1991), 229-70; abridged trans. R. Daglish, ’The Concept of the Ideal’ in *Philosophy in the USSR: Problems of Dialectical Materialism*, Moscow: Progress, 1977: 71-99.(Il’enkov’s fullest statement of his conception of activity and its role in the constitution of the non-material.)

Il’enkov, E.V. (1980) *Leninskaia dialektika i metafizika positivizma*, Moscow: Politizdat; trans. B. Pearce, *Leninist Dialectics and the Metaphysics of Positivism*, London: New Park, 1982.(An account of Lenin’s controversy with Bogdanov, from an unfinished manuscript. Intended as a critique of scientism and technologism in Soviet culture, though heavy censorship obscures this theme.)

Il’enkov, E.V. (1984) *Iskusstvo i kommunisticheskii ideal (Art and the Communist Ideal)*, Moscow: Iskusstvo.(A collection of writings on art, ideality and ideals; includes Il’enkov’s notes on Wagner, his favourite composer.)

Il’enkov, E.V. (1991) *Filosofiia i kul’tura (Philosophy and Culture)*, Moscow: Politizdat.(A comprehensive anthology of Il’enkov’s writings; annotated.)

References and further reading

Bakhurst, D.J. (1991) *Consciousness and Revolution in Soviet Philosophy. From the Bolsheviks to Evald Ilyenkov*, Cambridge: Cambridge University Press, 1991.(An account of Il’enkov’s thought, set in the context of the history of the Soviet philosophical tradition.)

Korovikov, V.I. (1990) ’Nachalo i pervyi pogrom’ (’The Beginning and the First Pogrom’), in *Voprosy filosofii* 2: 65-8.(A description of the climate in which Il’enkov’s career began at MGU.)

Mikhailov, F.T. (1990) ’Slovo ob Il’enkove’ (’A Word about Il’enkov’), in *Voprosy filosofii* 2: 56-64.(A discussion of the significance of Il’enkov’s contribution by a friend and colleague.)

Novokhat’ko, A.G. (1991) ’Fenomen Il’enkova’ (’The Phenomenon of Il’enkov’), in Il’enkov (1991) 5-16.(A detailed account of Il’enkov’s life and work.)

Il’in, Ivan Aleksandrovich (1883-1954)

Educated in law and philosophy in the first years of the twentieth century at Moscow University and several Western European universities, Il’in produced an important two-volume commentary on Hegel’s philosophy (1918), and a number of substantial works in political and legal theory, ethics and religious thought, aesthetics and literary criticism in later years. As a resolute foe of the Bolsheviks before and after the Revolution of 1917, he was exiled by them in 1922, living in Berlin until 1938, and subsequently in Switzerland until his death. Throughout his exile he remained deeply devoted to his Russian homeland, circulating extensive proposals for the eventual reconstruction of the Russian state, Church and society following the collapse of the Soviet regime (see Nashi zadachi (Our Tasks) (1956a)). He developed his own distinctive theory of monarchy as an ideal political form, grounded in a doctrine of natural right, and advocated it as the most appropriate choice for Russia in the best case, though he withheld judgment as to whether it would prove historically possible to implement it in the conditions prevailing after the demise of the Soviet system. His writing also focused to a significant extent upon moral and spiritual discipline and renewal, for societies as well as individuals, as necessary conditions of the well-ordered state and the well-lived life.

1 Life

Il’in was born in Moscow; his father was a native Muscovite, and his mother was of German descent. He completed gymnasium in the spring of 1901, and entered the Faculty of Law of Moscow University in autumn of the same year, receiving his diploma in 1906. In the Faculty of Law he encountered two unusually inspiring philosopher-jurists, Pavel I. Novgorodtsev (from the beginning) and Prince Evgenii N. Trubetskoi (from the end of 1905), among his professors. They provided a solid grounding in the history of philosophy in addition to legal studies. Novgorodtsev conveyed a lively interest in natural right and the rule of law to his students, a lesson which Il’in absorbed thoroughly but interpreted in a distinctive way. Upon completion of his diploma Il’in was retained at the university (on the recommendation of both Novgorodtsev and Trubetskoi) to undertake the programme of studies leading to the master’s degree, was appointed *privat-dozent* in the Faculty of Law (1909) and eventually (1918) received both the master’s and the doctoral degrees at once for the two volumes of his dissertation on Hegel.

Il’in undertook an intense study of Hegel beginning in 1908 and continuing until 1916, by which time his dissertation was essentially completed. During this period he spent two years (1911-12) engaged in study and research in Germany, Italy and France, including periods of work at Heidelberg (with Jellinek), Freiburg (with Rickert), Göttingen (with Husserl and Nelson) and Berlin (with Simmel).

With the outbreak of the First World War, Il’in began to take an active role in advocating the defence of tsarist Russia from enemies within and without. Following the events of February and October 1917 he became a particularly vigorous and public opponent of the Bolsheviks who arrested him six times between 1918 and 1922, finally exiling him in September 1922, under threat of execution, along with many other scholars, philosophers, theologians and writers irreconcilably opposed to the new regime.

Settling in Berlin, he became a professor at the newly opened Russian Academic Institute and taught there until the Nazis removed him in 1934 for refusal to cooperate with their propaganda aims. Forbidden all employment in Nazi Germany, deprived of the right to speak publicly, and watched by the Gestapo, he finally succeeded in leaving Germany for Switzerland in July 1938, where he remained until his death. During the difficult years of exile he produced a substantial number of highly original works on ethics, political theory, aesthetics, religion, and moral and spiritual discipline (see Poltoratzky 1989; Lisitsa 1993).

2 Commentary on Hegel

Il’in’s commentary, *Filosofiia Gegelia kak uchenie o konkretnosti Boga i cheloveka (The Philosophy of Hegel as a Doctrine of the Concreteness of God and Man)* (1918) was published in two volumes and its structure reflects its intended purpose as a dissertation to be submitted to the Faculty of Law of Moscow University. Volume 2, *Uchenie o cheloveke (The Doctrine of Man)* was, broadly speaking, a study of Hegel’s theory of the state, while volume one, *Uchenie o Boge (The Doctrine of God)*, provided an overview of Hegel’s philosophical position.Il’in

declined to engage in the most usual form of Hegel commentary, the *explication de texte* retracing the author's exposition section by section through a single work. His commentary is rather a *reconstruction*, theme by theme, of the major distinctive elements of Hegel's position, often drawing simultaneously on several of Hegel's texts, while duplicating the rhetorical structure of none of them.

Volume 1 of *Logic*. It is of special interest in that Il'in explicated Hegel's doctrine of *speculative concreteness* with a degree of clarity and detail unmatched by any commentator before him. This volume of the commentary is itself structured in terms of the distinctive doctrine of speculative concreteness, and demonstrates that this doctrine, and the criterion of reality formulated in terms of it, also structure the central claims of Hegel's *Logic* as a whole.

The second volume of Il'in's commentary offers an extended reflection upon the central themes involved in Hegel's *Philosophy of Right*; thus 'freedom', 'will', 'right', 'morality' and 'the ethical' are all treated in separate chapters there, as well as 'humanity' and 'the state'. In this study, Il'in focused significant attention on Hegel's concept of God, most especially arguing that in Hegel's conception, the divine nature as a whole is subject to Becoming, thus capable of undergoing the tragic experience of suffering and death. Il'in himself held to the traditional doctrine of the divine nature as impassible, and so he personally rejected the view he ascribed to Hegel. He argued that in working out the details of his ethical and political doctrine Hegel was forced to 'compromise' the original intent of his philosophy which, according to Il'in, was above all to produce a theodicy (see Grier 1997).

Both volumes taken together constitute a profound, subtle and sustained reflection on the theological implications of Hegel's philosophy, and the influence of Il'in's commentary has been greatest in this area. Both Hans Küng's *Menschwerdung Gottes* (1970) (translated as *The Incarnation of God* (1987)) and Cyril O'Regan 's *The Heterodox Hegel* (1994) were significantly influenced by Il'in's commentary.

It should be pointed out, however, that it is only the drastically shortened German translation, *Die Philosophie Hegels als kontemplative Gotteslehre* (1946) produced by Il'in himself, that has been known to non Russian-speaking scholars. Owing to Il'in's failing health and his extreme reluctance to take time away from other unfinished projects, eight of the ten chapters of the original Russian volume 2 were left out of this edition.

3 Political theory

In formulating his own theory of right, authority, law and the state Il'in demonstrated the originality and independence of his thought to a striking extent. He departed from the settled paths and terms of modern European political theory to such a degree that all of his works on the subject must be read attentively in their entirety to avoid misunderstanding. He set aside the wisdom of the post-Lockean liberal democratic synthesis (which he labels generically 'the republican principle') as a conception aspiring merely to avoid the worst excesses of governmental tyranny, always threatened by degeneration into an unmanageable anarchy of extreme individualism (1979, chaps 6, 7). In its place he offered a theory of maximum aspiration, a conception in which the ultimate justification of state authority would be the development in the citizenry of a moral, legal and spiritual culture in which the requirements of natural right would be so widely exemplified in human conduct as to make genuine self-government a reality. He conceives of 'monarchy' (in his own distinctive meaning of the term) as the most appropriate form for the realization of such a state (1979, chaps 1-3).

'*Pravosoznanie*', from *pravo* (right or law) and *soznanie* (consciousness), is Il'in's distinctive term for an awareness (and acceptance) by the individual of the system of obligations and rights specified in natural (and positive) law. Conduct guided by *pravosoznanie* is necessary for any human community that provides the conditions for the emergence of spiritual life, the highest good for the individual, and *pravosoznanie* is itself one of the constituent elements of that highest good. A rational system of positive law would reflect the structure of natural law (1956b, chaps 5, 6).

The development of *pravosoznanie* in the population remains a fundamental obligation and goal of state authority. The obligations laid upon the monarchical state itself by *pravosoznanie* are conceived as fundamental, and where the monarch fails to meet these obligations Il'in sees not merely a right, but an *obligation* of disobedience on the part of the population (1979: 222-3). The authority of the monarch is thus strictly limited by natural law, but within those bounds is depicted as properly autocratic, patriarchal, grounded in spiritual authority, and giving rise to gradations of rank among the population. Paradoxically, the ultimate justification of monarchical authority lies in the fulfilment of its obligation to render itself superfluous by the development of *pravosoznanie* to the point of

genuine self-governance by the population. He supposed not that the ideal of such a state would be appropriate for every nation, but that it articulated something essential to the original Russian idea, a critically important ideal, he believed, if that nation was ever to be rescued from Bolshevism and reconstructed.

4 Ethics, aesthetics and religious thought

In 1925 Il’in published the very controversial *O soprotivlenii zlu siloiu (On Opposing Evil with Force)*, carefully reviewing the fundamental tenets of Christian moral theory and defending the claim that under certain circumstances one’s Christian duty might include using the sword to oppose evil. Having satisfied himself that the argument for this conclusion was sound, he forcefully and scornfully attacked the Tolstoian doctrine of non-resistance to evil so popular among many turn-of-the-century Russian intellectuals, characterizing it as a product of intellectual inadequacy and moral weakness of will and arguing that it was responsible in significant measure for the failure of the Russian state to defend itself against Bolshevism. The publication of this work produced a furore in the Russian émigré community, causing a division of opinion for and against Il’in that endured for many years (Il’in 1925, Poltoratzky 1975).

His principles of aesthetic criticism were first presented in *Osnovy khudozhestva: O sovershennom v iskusstve (Fundamentals of Art: On Perfection in the Arts)*(1937), and applied to some of the major Russian émigré writers in *O t’me i prosvetlenii (On Darkness and Illumination)*(1959). He regarded art as a kind of spiritual service and source of joy for the artist as well as the audience, a mode of religious experience for those capable of truly participating in it. A collection of his occasional essays and lectures on Russian writers and other aesthetic topics has also been published (Il’in, 1973).

As a deeply religious thinker and individual, Il’in tended to see a connection between most aspects of successful human endeavour and an actively lived faith. Philosophy, art, morality, law, politics and government, even science - each of these dimensions of human activity needed to be understood in connection with the divine in order to reach its perfected form. In this sense nearly all of Il’in’s major writings touched on religion in some respect. Among those devoted more exclusively to the topic of religion various emphases can be seen, such as (1) numerous occasional writings on the fate of the Russian Orthodox Church; (2) instruction and inspirational writing on the life of spirit, such as *Put’ dukhovnogo obnovleniia (The Path of Spiritual Renewal)* (1935), *Poiushchee serdtse (The Singing Heart)* (1958) and *Put’ k ochevidnosti (The Path to Manifest Truth)* (1957); and (3) his monumental two-volume work on the fundamental presuppositions of religious experience *Aksiomy religioznogo opyta (The Axioms of Religious Experience)* (1953). In this last work Il’in sought to specify the objective requirements of the fundamental ’religious act’ inherent in the Eastern Orthodox faith, and the specific ’structure of the soul’ presupposed by it and developed in it. He felt that these elementary ’presuppositions’ of faith had been left largely unexamined by theologians and were subject to degeneration and distorted comprehension in the contemporary world. Il’in’s overriding concern in all of these works was to trace the spiritual roots of the general crisis affecting Western and Russian civilization in the twentieth century and point out paths to the restoration of spiritual health for individuals and for nations.

PHILIP T. GRIER

List of works

Il’in, I.A. (1993) *Sobranie sochinenii v desiati tomakh (Collected Works in Ten Volumes)*, ed. Iu.T. Lisitsa, Moscow: Russkaia kniga. (Vol. 1 contains a bibliography listing about 350 publications by Il’in, including books, brochures, journal articles and newspaper articles, in three languages. A still more detailed and corrected bibliography can be found in vol. 2 of *Sochineniia v dvukh tomakh (Works in Two Volumes)* compiled by Iu.T. Lisitsa in the series *Iz istorii otechestvennoi filosofskoi mysli* published by Voprosy filosofii, Moscow: Medium, 1993, 1994.)

Il’in, I.A. (1918) *Filosofiia Gegelia kak uchenie o konkretnosti Boga i cheloveka (The Philosophy of Hegel as a Doctrine of the Concreteness of God and Man)*, vol. 1, *Uchenie o Boge (The Doctrine of God)*, vol. 2, *Uchenie o cheloveke (The Doctrine of Man)*, Moscow: Izd. G.A. Lemana i S.I. Sakharova.(The author himself translated all of vol. 1 and the last two chapters of vol. 2 of the original Russian edn into German and published it as a single volume: *Die Philosophie Hegels als kontemplative Gotteslehre (The Philosophy of Hegel as a Contemplative Doctrine of God)*, Berne: A. Francke, 1946. The original 1918 Russian edition has recently been republished in its entirety as a single volume (St Petersburg: Nauka, 1994).)

Il’in, I.A. (1925) *O soprotivlenii zlu siloiu (On Opposing Evil With Force)*, Berlin: privately published; 2nd edn London, Canada: Zaria, 1975.(Il’in’s argument that under certain circumstances Christian moral doctrine does justify the use of the sword to oppose evil, and his attack on Tolstoi’s doctrine of non-resistance to evil, and its consequences.)

Il’in, I.A. (1935) *Put’ dukhovnogo obnovleniia (The Path of Spiritual Renewal)*, Belgrade: Russkaia biblioteka, Kniga 43; 2nd edn Munich: privately published by N.N. Il’ina, 1962. (The original edition did not contain the last three chapters included in the second publication.)

Il’in, I.A. (1937) *Osnovy khudozhestva: O sovershennom v iskusstve (Fundamentals of Art: On Perfection in the Arts)*, Riga: Russkoe akademicheskoe izd.(The primary statement of Il’in’s aesthetic doctrine, in which he holds that great works of art and literature contain a religious/spiritual dimension of which the artist may not have been fully conscious.)

Il’in, I.A. (1953) *Aksiomy religioznogo opyta (Axioms of Religious Experience)*, Paris: privately published, 2 vols. (Il’in’s most important religious work; an exploration of the fundamental ‘religious act’ inherent in the Eastern Orthodox faith, and the specific ’structure of the soul’ required by it.)

Il’in, I.A. (1956a) *Nashi zadachi: Stat’i 1948-1954 (Our Tasks: Articles 1948-1954)*, Paris: Izd. Russkogo Obshche-Voinskogo Soiuza, 2 vols.(Il’in’s proposals for the reconstruction of the Russian state, Church and society following the collapse of Soviet regime, originally circulated as a series of 215 anonymous ‘bulletins’ by a Russian émigré organization called the Russian Joint Military Union, which collected all of the bulletins and published them together after Il’in’s death.)

Il’in, I.A. (1956b) *O sushchnosti pravosoznaniia (On the Essence of Consciousness of Right)*, Munich: privately published by N.N. Il’ina.(The initial version (1919) was revised throughout Il’in’s life. This is a work of political and legal philosophy in which Il’in explained the doctrine of ’consciousness of right/law’ and its connection with natural law as an essential element of the political/legal arrangements he thought appropriate as an ideal for a future Russian state.)

Il’in, I.A. (1957) *Put’ k ochevidnosti (The Path to Manifest Truth)*, Munich: privately published by N.N. Il’ina. (An inquiry and guide into the problem of self-evidence, or what is required for penetrating to the objective essence of a problem - of knowledge, of ethics, of spiritual life - and avoiding the blindness induced by surface appearances.)

Il’in, I.A. (1958) *Poiushchee serdtse. Kniga tikhikh sozertsanii (The Singing Heart. A Book of Quiet Contemplations)*, Munich: privately published by N.N. Il’ina, 1943.(Il’in described this as ’devoted not to theology, but to a quiet, philosophical praising of God’, to the development of a new philosophy, Christian in spirit and style.)

Il’in, I.A. (1959) *O t’me i prosvetlenii: Kniga khudozhestvennoi kritiki. Bunin-Remizov-Shmelev (On Darkness and Illumination: A Book of Aesthetic Criticism. Bunin-Remizov-Shmelev)*, Munich: privately published by N.N. Il’ina.(Il’in’s commentary on these three important writers of the emigration, illustrating his own principle of aesthetic criticism. Written 1939.)

Il’in, I.A. (1973) *Russkie pisateli, literatura i khudozhestvo: Sbornik statei, rechei i lektsii (Russian Writers, Literature and Art: An Anthology of Articles, Speeches and Lectures)*, Redaktsiia, predislovie i primechaniia N.P. Poltoratskogo, Washington: Izd. Russkogo knizhnogo dela v SShA, Kamkin.(A collection of Il’in’s brochures, lectures and speeches devoted to the literary genius of Pushkin, the spiritual art of Shmelev, literary criticism of Merezhkovskii, the Russian folktale, Russian poetry and the idea of Russia contained in it, and questions of artistic creativity.)

Il’in, I.A. (1979) *O monarkhii i respublike (On Monarchy and the Republic)*, Redaktsiia, predislovie i prilozhenie N.P. Poltoratskogo, New York: Sodruzhestvo.(This is Il’in’s primary work on forms of the political state. Not completed in his lifetime, it was assembled and edited by Poltoratzky.)

References and further reading

Evlampiev, I.I. (1992) ‘‘‘Drama mirotvoriashchego bozhestva’’: Bog i chelovek v filosofii Ivana Il’ina’ (‘‘‘The Drama of the Worldcreating Divinity'': God and Man in the Philosophy of Ivan Il’in’), *Stupeni*, 3: 66-89.(An attempt to identify distinct ‘periods’ in the development of Il’in’s outlook.)

Grier, P.T. (1994) ‘The Complex Legacy of Ivan Il’in’, in *Russian Thought After Communism: The Recovery of a Philosophical Heritage*, ed. J.P. Scanlan, Armonk, NY: M.E. Sharpe, 165-86.(A discussion of the contemporary revival and reception of Il’in’s writings in Russia.)

Grier, P.T. (1997) ‘The Speculative Concrete: I.A. Il’in’s Interpretation of Hegel’, in *Hegel, History and Interpretation*, ed. S. Gallagher, Albany, NY: State University of New York Press, 169-93.(An examination of distinguishing features of Il’in’s interpretation of Hegel. Expansion of §2.)

Lisitsa, Iu.T. (1991) ‘I.A. Il’in kak pravoved i gosudarstvoved’ (’I.A. Il’in as a Theorist of Law and the State’), *Voprosy filosofii* 5: 146-58. (A careful, reliable discussion of aspects of Il’in’s political theory.)

Lisitsa, Iu.T. (1993) ‘Ivan Aleksandrovich Il’in. Istoriko-biograficheskii ocherk’ (Ivan Aleksandrovich Il’in. A Biographical and Historical Essay), in *I.A. Il’in. Sobranie sochinenii v desiati tomakh*, Moscow: Russkaia kniga, vol. 1, 5-36.(The most recent, authoritative biographical essay on Il’in. A source for §1.)

Offermanns, W. (1979) *Mensch, werde wesentlich! Das Lebenswerk des russischen religiösen Denkers Ivan Iljin für die Erneuerung der geistigen Grundlagen der Menschheit (Man, Become Your Essence! The Lifework of the Russian Religious Thinker Ivan Il’in for the Renewal of the Spiritual Foundation of Humanity)*, Band 11 von OIKONOMIA, Quellen und Studien zur orthodoxen Theologie unter Mitarbeit von Erich Bryner und Karl Christian Felmy, Erlangen: Hrg. von Fairy v. Lilienfeld.(The most important monograph-length study of Il’in’s religious thought to date.)

Poltoratzky, N. (1975) *I.A. Il’in i polemika vokrug ego idei o soprotivlenii zlu siloiu (I.A. Il’in and the Controversy Over his Idea of Opposing Evil with Force)*, London, Canada: Zaria; also published as an appendix to the second edition of *O soprotivlenii zly siloiu*, London and Canada: Zaria, 1975.(Poltoratzky’s compilation and review of the critical reaction to Il’in’s publication.)

Poltoratzky, N. (1979) *Monarkhiia i respublika v vospriiatii I.A. Il’ina (Monarchy and the Republic in the Perception of I.A. Il’in),* New York: Sodruzhestvo.(An authoritative overview of Il’in’s political thought. A source for §3. Also published as an appendix to Il’in’s *O monarkhii i respublike*.)

Poltoratzky, N. (1989) *Ivan Aleksandrovich Il’in. Zhizn’, trudy, mirovozzrenie. Sbornik statei* (Ivan Aleksandrovich Il’in. Life, Works, Worldview. A Collection of Articles), Tenafly, NJ: Ermitazh. (Nikolai Petrovich Poltoratzky (1921-90) was the literary executor for Il’in, the compiler of the Il’in papers at Michigan State University, and the leading authority on Il’in. An important biographical work; a source for §1.)

Illuminati

Begun in 1776 in Bavaria, the Illuminati were an overtly political as well as morally orientated secret organization that imitated the forms of freemasonry. While masonic lodges forbade discussion of politics and religion at their meetings, the Illuminati did the reverse. They were openly yet paradoxically secret about their irreligion and their devotion to the radical French Enlightenment; and they wanted reform in the absolutist states of Central Europe. The authorities arrested and persecuted them, but their activism foreshadows the French Revolution and a desire for more representative systems of government in Continental Europe. Despite the notoriety of the Illuminati, no more than about 600 members have been identified.

Freemasonry always promised more than it delivered. Its rules and the constitutions of Grand Lodges throughout Europe spoke of all men meeting upon the level, and of merit as the only requirement for advancement. Yet in lodge after lodge, hierarchy and social position reasserted themselves, and the Grand Lodges were, without exception, led by aristocrats. In addition the masonic lodges, particularly in the German-speaking lands, had been used by enlightened absolutists as extensions of their social influence. Frederick II in Berlin, and later Joseph II in Vienna, openly used membership as a symbol of loyalty to themselves, and of a man's willingness to work for the interests of the state before those of the clergy or the local power of old elites. It is little wonder that someone, somewhere would adopt the form and ideology of freemasonry and radicalize both. That is precisely what the Illuminati did.

Founded, not accidentally, in 1776 - the year of revolutionary upheaval in the new world - by Weishaupt (1748-1830), a professor of law at the Bavarian university at Ingolstadt, the Illuminati were overtly and dangerously political. Unlike the lodges which forbade open discussion of politics - and then subtly assumed political personae - the lodges of the Illuminati openly called for political and philosophical discussions, less hierarchy, and attention to issues of reform. Weishaupt adopted the form of the masonic lodge, issued 'statutes' as would a lodge, embraced secrecy, and called for moral reform: *'Mocht sich die alten und neuern System der Moral, Philosophie als stoischen, epikureischen…'* ('you should embrace the old and new System of Morality, Philosophy such as the stoic and the epicurean…'). Thus Enlightenment paganism was harnassed to political reform. Lessing's masonic dialogue, *Ernst und Falk* (1778), may have been partly inspired by Illuminati ideals; certainly it argued for democracy and for a set of reforms that would make societies and states less hierarchical, less sectarian, and less bellicose (see Lessing, G.E.). Lessing's argument in the dialogue centred on the notion that all he was doing was bringing freemasonry to its logical conclusion. The masonic goal of virtue and individual moral improvement was turned outward, and the mirror reflected poorly upon the absolutist principalities or the clergy favoured in them.

Again, it was not accidental that Weishaupt had been educated by the Jesuits. There had long been tension between the order and the masonic lodges. In the 1740s the leading Amsterdam freemason, Jean Rousset de Missy said that masters of a lodge would admit any man - except a Jesuit. The Illuminati were quickly locked in combat with the Jesuits who spied in them another example of masonic evil. Weishaupt argued that his brothers should do as had the Jesuits, make themselves indispensable as servants of the state. The point of this infiltration, however, would be to drastically alter the state; not to support it as had the Jesuits. Other German intellectuals were attracted to his arguments, notably the freemason Adolf Freiherrn von Knigge, who became a theorist of illuminism. The debate begun by the Illuminati and the Jesuits can be found echoed or discussed by Fichte, Schiller, and Herder.

The League of the Illuminati were strongest in Munich where von Knigge sought to use it as a vehicle to reform freemasonry and to extend its influence throughout Germany. At the height of its fame the League had no more than 600 members of which most were court and administrative officials, clergymen and military officers. They swore an oath that said 'the sole aim of the League is education, not by declamatory means but by favouring and rewarding virtue. The order of the day is to put an end to the machinations of the purveyors of injustice'.

The goals of the League provoked fear on the part of the authorities who sought to infiltrate their ranks, and eventually they arrested and imprisoned their leaders. It was said that Jesuits were also attempting to spy on the Illuminati, and what began as a small effort by a young professor in Bavaria suddenly captured the attention of the European press. The Illuminati became symbol and reality of the crisis of absolutism seen in Berlin, Paris and

Vienna during the 1780s. At its heart lay the contradictions within enlightened absolutism: How to reform the state without unleashing uncontrollable forces or allowing potential revolutionaries to gain the upper hand? The Illuminati became the Anabaptists of their era, indeed Catholic propaganda openly compared them to the radicals of Münster. For their part the German freemasons moved to distance themselves from this new hybrid. The Bavarian response foreshadows by a decade the emergence of conspiracy theory during the French Revolution. As early as the autumn of 1789 the freemasons were being blamed for events in Paris, but the background lay in the furore caused by the Illuminati and the efforts made by the German states to exterminate them.

Something discernibly modern emerged in Bavaria in 1776, only to be suppressed by 1786: groups of young men intent upon planning for radical political reform that could include the dismantling of established political authority. Equally modern was the response: conspiracy theorists surfaced who defined the world as a place threatened by dark forces possessed of extraordinary power solely by virtue of their secrecy and silence. In the process the state had come to be seen as a human institution, made by the people and not by divine will. As early as 1780 Illuminati showed how fragile the authorities of the absolutist state perceived their power and control to be; the decade of the 1790s would prove their fears to have been justified.

See also: Enlightenment, Continental

MARGARET C. JACOB

List of works

Von Knigge, A.F. (1799) *Practical Philosophy of Social Life*, London.
Von Knigge, A.F. (1788) *Sämtliche Werke (Complete Works)*, Leichtenstein: KTO Press, 1978.
Weishaupt, A. (1788) *Über die Gründe der menschlichen Erkenntniss (On the Foundations of Human Knowledge)* Berlin; Brussels, 1969.

References and further reading

Abbott, S. (1991) *Fictions of Freemasonry and the German Novel*, Detroit, MI: Wayne State University Press.(A useful collection of primary source material.)
Jacob M.C., (1991) *Living the Enlightenment. Freemasonry and Politics in Eighteenth Century Europe*, New York: Oxford University Press.(A survey of European freemasonry from its foundation in England in 1717 to the French revolution.)
Neugebauer-Wölk, M. (1995) *Esoterische Bünde und Bürgerliche Gesellschaft, Entwicklungslinien zur modernen Welt im Geheimbundwesen des 18. Jahrhunderts (Esoteric Societies and Citizens' Groups, Lines of Development to the Modern World from the Secret Societies of the Eighteenth Century)*, Wolfenbüttel: Lessing Academie.
Van Dülmen, R. (1992) *The Society of the Enlightenment, The Rise of the Middle Class and Enlightenment Culture in Germany*, trans. Anthony Williams, Cambridge: Polity Press.(A solid historical work on the emergence of civil society in the eighteenth century.)

Illumination

The most influential theories of illumination explain certain features of our knowledge by developing an analogy with ordinary sensory vision and the role played in it by light. According to theories of this sort, our knowledge of necessary and immutable objects and truths requires the activity of a kind of intelligible light illumining objects that are purely intelligible, thereby making them 'visible' to our mind. Plato held that this light comes from the Form of the Good, Augustine that it comes from God, and others that it is intrinsic to reason itself.

The peculiar nature and behaviour of light has provided a model not just for theories of knowledge but also for philosophical accounts of fundamental features of reality. Neoplatonist cosmologies, for example, liken the genesis of the universe to the emanation of rays of light from a light source. Theories of illumination therefore can be metaphysical as well as epistemological. Both kinds of theory have their historical roots in Platonism broadly construed.

1 Epistemological illuminationism

Platonist epistemologies characteristically focus on the fact that we understand certain immutable, eternal natures (for example, the objects of geometry, such as the circle and the triangle) and know certain necessary truths (for example, that the interior angles of a triangle are equal to two right angles). They argue that we cannot have acquired knowledge and understanding of this sort through sense perception since no objects of sense perception are of the right sort to account for the necessity and immutability distinctive of this sort of knowledge. They conclude that if such knowledge is to be possible, there must be necessary and immutable objects epistemically accessible to us by means other than sense perception. Epistemologies of this sort hold that these objects are purely intelligible and that we make epistemic contact with them through reason, intellect or mind. Our understanding of the nature of a triangle (for example) and our certain knowledge that its interior angles are equal to two right angles are to be explained by our being (or having been) directly acquainted with the relevant purely intelligible object(s).

The model for this epistemically fundamental notion of direct intellective acquaintance is sensory vision. Our access to purely intelligible objects by means of reason or mind is analogous to our access to ordinary visible objects through sight. According to this analogy, there are three essential components in both the sensory and intellective cognitive processes. First, we must possess the relevant cognitive power (sight in the case of sensory vision; intellect in the case of intellective cognition). Second, there must be an appropriate object for that power (a coloured material body; a purely intelligible object). And third, there must be an agent whose activity enables the cognitive power and its object to make contact. In sensory vision, this third component - the enabling agent of cognition - is light: corporeal light (from the sun, for example) illumines material objects, making them not merely potentially but actually visible to us. In intellective cognition, there must be an analogous intellectual light whose activity makes intelligible objects actually and not merely potentially intelligible to us. Our direct acquaintance with the nature of a triangle (for example) depends essentially, therefore, not only on our intellective powers and the existence of the relevant intelligible object(s) but also on the activity of an intellectual light.

The analogy between intellectual understanding and sensory vision is superficially compelling (our ordinary ways of talking about knowledge and understanding rely heavily on it), but the metaphor of intellectual light gives rise to difficult philosophical questions. What exactly is the nature of this intellectual light? Is it something ontologically independent of us or is it identical with or merely a function of our intellectual powers and capacities? Plato's famous analogy of the sun suggests that the intellectual light has as its source the Form of the Good:

> Say, then, that it is the sun which I call the offspring of the Good, which the Good begot as analogous to itself. What the Good itself is in the world of thought in relation to the intelligence and things known, the sun is in the visible world in relation to sight and things seen…. [W]henever one's eyes are turned upon objects brightened by sunshine, they see clearly…. So too understand the eye of the soul: whenever it is fixed upon that upon which truth and reality shine, it understands and knows, and seems to have intelligence…. Say that what gives truth to the objects of knowledge, and to the knowing mind the power to know, is the Form of the Good.
>
> (*Republic* VI, 508b-e)

Later Platonists develop and extend what we might call the supernaturalist elements of Plato’s analogy. Augustine (§4), for example, typically describes the source of intellectual illumination not as the Form of the Good but as Truth itself, which he identifies with God. He holds, then, that our knowledge depends essentially on God’s activity in our souls. For this reason, his view is typically characterized as a doctrine of divine illumination.

By developing Plato’s analogy in an explicitly theological direction, Augustine intended to deepen it and weave its epistemological elements into his broad, theistic account of reality. But medieval philosophers attracted by Augustine’s Christian Platonism saw two general difficulties in his doctrine of divine illumination. First, Augustine had suggested that in so far as any particular truth is necessary, immutable and eternal, it must be, or in some way be a part of, God, since God alone is truly necessary, immutable and eternal. But in that case it seems that in grasping a truth of this sort, ordinary human knowers are in direct epistemic contact with the divine nature, a state that Christian doctrine takes to be virtually unattainable by human beings in this life. Second, in so far as Augustine’s account identified illumination with God’s activity within our minds, it seemed to make genuine knowledge a product of supernatural intervention in our cognitive lives, and so not genuinely human at all.

Medieval philosophers and theologians typically try to steer a middle course between an interpretation of Augustine’s doctrine of divine illumination that is unacceptably supernaturalist and an interpretation that leaves no room at all for divine activity in human knowing. Many preserve the doctrine essentially intact, introducing minor distinctions designed to blunt the most serious objections. Bonaventure (§7), for example, argues that our grasp of the eternal natures that are in God can be a matter of degree:

> Since, then, certain knowledge belongs to the rational spirit, insofar as it is the image of God, it follows that in this knowledge the spirit attains to the eternal reasons [that are in God]. But since, as long as it is in the wayfaring state, [the rational spirit] is not fully deiform, it does not attain to them clearly and fully and distinctly…. It is to be granted, then,… that in all certain knowledge those principles of knowledge are attained by the knower. They are reached in one way, however, by the wayfarer, and in another way by him who enjoys the vision of God [in the next life].
>
> (*Disputed Questions Concerning Christ’s Knowledge*, q.4)

But some important late medieval thinkers give the doctrine of divine illumination an essentially naturalistic interpretation; this is especially the case with those philosophers - such as Bonaventure’s contemporary, Thomas Aquinas (§11) - who are committed to an Aristotelian analysis of mind and cognition. Aquinas argues that the intellectual light that makes immutable natures actually intelligible to us must be identified with a power *in us*, otherwise the resultant knowing would not be ours. That power, Aquinas claims, is the Aristotelian agent intellect, the power of the human soul that abstracts universal natures from the material and individuating conditions that characterize the presentations of sense perception. He adapts the terminology of the doctrine of illumination to fit the structure of his Aristotelian account, calling the agent intellect a ‘light’ and thinking of its activity of abstracting universal forms as a kind of ‘illumination’. In doing so, Aquinas preserves the letter of Augustine’s position if not its entire spirit.

This naturalistic rendering of the doctrine of illumination survives into early modern philosophy in the work of philosophers such as René Descartes (§7). The foundational role he assigns to propositions that are clearly and distinctly perceived to be true (or that are shown to be true by the natural light of reason) marks Descartes’ theory of knowledge as part of the legacy of Platonist epistemological illuminationism.

2 Metaphysical illuminationism

Light’s seeming propagation of itself in all directions from its source provides the model for emanationist cosmologies and related metaphysical accounts. Plotinus (§§3, 5), for example, described the emanation of all reality from the single, unified source of being (the One) as like the emanation of rays of light from a light source. Medieval Christian philosophers, who rejected emanationism as an account of the genesis of the universe, nevertheless incorporated the notion of emanation into their explanations of various aspects of the relation between creatures and their creator. According to Albert the Great (§4), for example, metaphysical universals are like rays radiating from God, the primary intelligence and creator of all forms. These universal forms, which are in themselves simple, fall on matter (producing particular individuals) and on souls (for which they are the principles of cognition). Henry of Ghent (§§2-3) argued that a creature’s essence is related to its existence in the way a ray of

sunlight is related to the light it possesses: the ray receives its light from the sun (and so is in that respect distinct from and dependent on it), but it receives it in such a way that the ray's very nature and essence is light (and so is in this respect not distinct from it). Accounts of this sort can be thought of as developing the metaphysical side of the metaphor of illumination.

See also: Chinul; Epistemology, history of §§1-4; Grosseteste, R. §3; Illuminationist philosophy; Medieval philosophy

SCOTT MacDONALD

References and further reading

Albert the Great (1240s) 'On Universals' (*Liber de praedicabilibus*, tr. II, ch. 6), trans. S. MacDonald, *Cambridge Translations of Medieval Philosophical Texts*, vol. 2, *Metaphysics*, Cambridge: Cambridge University Press, forthcoming.(Albert shapes a Neoplatonist metaphysics of emanation to fit his theistic conception of God.)

Aquinas, T. (1266-73) *Summa theologica* Ia, trans. Fathers of the English Dominican Province, Westminster, MD: Christian Classics, 1981.(Questions 79 and 84 show Aquinas developing his own Aristotelian account of intellectual cognition with a clear awareness of the Augustinian Platonist tradition of illumination.)

Augustine (388, 391-5) *On Free Choice of the Will*, trans. T. Williams, Indianapolis, IN: Hackett Publishing Company, 1993.(Book II contains one of Augustine's more extended developments of his doctrine of divine illumination.)

Bonaventure (1256) *Disputed Questions Concerning Christ's Knowledge* (*Quaestiones disputatae de scientia Christi*), q.4, trans. E.R. Fairweather, *A Scholastic Miscellany: Anselm to Ockham*, New York: Macmillan, 1970. (Bonaventure's Augustinian sympathies lead him to try to resolve difficulties inherent in Augustine's doctrine of illumination.)

Descartes, R. (1641) *Meditations on First Philosophy*, trans. G.E.M. Anscombe and P.T. Geach, *Philosophical Writings*, Indianapolis, IN: Bobbs-Merrill, 1971. (Descartes' epistemology extends the Augustinian heritage into modern philosophy.)

Henry of Ghent (1276) 'Is a Creature's Essence its Being?' (*Quodlibet* I, q.9), trans. S. MacDonald, *Cambridge Translations of Medieval Philosophical Texts*, vol. 2, *Metaphysics*, Cambridge: Cambridge University Press, forthcoming. (Henry finds in what he takes to be peculiar features of the rays of light emanating from a light source a useful analogy for the relation between essence and existence.)

Plato (*c*.390 BC) *Republic*, trans. G.M.A. Grube, Indianapolis, IN: Hackett Publishing Company, 1974.(Plato's parable of the sun, in book VI, is an enormously influential source of epistemological illuminationism.)

Plotinus (*c*.260) *Enneads*, trans. A.H. Armstrong, Cambridge, MA: Harvard University Press, 1966-88, 7 vols. (Plotinus develops Platonist themes into full-blown cosmological emanationism.)

Illuminationist philosophy

Illuminationist philosophy started in twelfth-century Persia, and has been an important force in Islamic, especially Persian, philosophy right up to the present day. It presents a critique of some of the leading ideas of Aristotelianism, as represented by the philosophy of Ibn Sina (Avicenna), and argues that many of the distinctions which are crucial to the character of that form of philosophy are misguided. Illuminationists develop a view of reality in accordance with which essence is more important than existence, and intuitive knowledge is more significant than scientific knowledge. They use the notion of light, as the name suggests, as a way of exploring the links between God, the Light of Lights, and his creation. The result is a view of the whole of reality as a continuum, with the physical world being an aspect of the divine. This sort of language proved to be very suggestive for mystical philosophers, and Illuminationism quickly became identified with Islamic mysticism.

1 Origins

Illuminationist philosophy stems from the Arabic term *ishraq*, meaning ‘rising’, in particular the ‘rising of the sun’. The term is also linked to the Arabic for ‘East’, and has come to represent a specifically Eastern form of philosophical thought. It is used, especially within the context of Persian poetic literature, to represent a form of thought which contrasts with cognitive reason (*‘aql*); that is, it is taken to be intuitive, immediate and atemporal knowledge. The source of this form of thought is often identified with Ibn Sina’s ‘Eastern Philosophy’ (*al-hikma al-mashriqiyya*), a text about which there is a great deal of dispute and discussion, and which may never have actually existed. It is supposed to represent Ibn Sina’s departure from Peripateticism and his attempt to construct a new and deeper philosophical system (see Islam, concept of philosophy in §2).

The real originator of illuminationist philosophy is al-Suhrawardi, a Persian philosopher of the twelfth century AD, who composed over fifty works but who is chiefly remembered for his brief *Hikmat al-ishraq (The Philosophy of Illumination)*. In this book, al-Suhrawardi (in Persian, Sohravardi) adopts some of the main principles of Peripateticism (*al-falsafa al-mashsha’iyya*), but also sets out to challenge others. He criticizes Peripatetic approaches to a wide variety of topics, in particular quantification, the confusion between ‘term’ and ‘utterance’, the notion of amphiboly, petitio principii and many other issues (see Aristotelianism in Islamic philosophy). There is a marked similarity between his critique and that of William of Ockham, who also identifies in his *Summa logicae* what he regards as ten fallacies in Aristotelian logic. Both al-Suhrawardi and Ockham rearrange the parts of the Organon and omit from their discussions some of the books.

2 Logic and semantics

Illuminationist philosophy challenges the Peripatetic position of the absolute, unchanging and universal validity of the truths discoverable by Aristotelian methodology (see Peripatetics). It sets out instead to construct a system applicable to the whole continuum of being, including what is called ‘immediate knowledge’. Al-Suhrawardi rejects Aristotle’s theory of definition, arguing that there is no criterion for the parts of a definition (the genus and the differentia), and so the species is defined in terms of something less known than itself. He goes on to claim that some of the Aristotelian categories are superfluous, since action and passion are modes of motion, and possession and posture are kinds of relation. Thus we need only five categories instead of ten, leaving substance, quantity, quality, relation and motion.

The basis of al-Suhrawardi’s approach is really Stoic and Megaric (see Stoicism; Megarian school). According to this approach, the denotation (the external object) should be compared with the thing, the sign should be compared with the utterance and what is signified should be compared with the meaning. These semantic notions are used to define the relation between the first atemporal act of thought and the second temporally-extended grasp of the thing known, its essence (Ziai 1990: 42-). This involves the development of a theory of types of signification, relation of class names to constituents of the class, types of inclusion of members in classes, and a well-defined theory of supposition.

In the illuminationist view of logic, a conclusion reached by using a formally established syllogism has no epistemological value as a starting point in philosophical construction. For a universal affirmative proposition to have philosophical value as a foundation of scientific knowledge, it must be ‘necessary and always true’. Yet if we

introduce the mode 'possibility' and give it an extension in time as in 'future possibility', the universal affirmative proposition cannot be 'necessarily true always'. This is because of the impossibility of knowing or deducing all possible future instances. The epistemological implication of this logical position is that formal validity ranks lower than the certitude obtained by the self-conscious subject who, when alerted to a future possible event through 'knowledge by presence', will simply 'know' it. The future event cannot be deduced at the present time *and* given universal validity (see Logic in Islamic philosophy).

3 Epistemology and ontology

The crucial notion for Illuminationist epistemology is knowledge-by-presence (*al-'ilm al-huduri*). This identifies an epistemological position prior to acquired or representational knowledge (*al-'ilm al-husuli*). This has often been related to intuitive knowledge, and results in attempts to unravel the mysteries of nature not through the principles of physics but through the metaphysical world and the realm of myths, dreams, fantasy and truths known through inspiration. The distinction between scientific knowledge and knowledge-by-presence is crucial for al-Suhrawardi, who claims that the essence of human beings lies in their self-awareness, through the luminosity of their own inner existence (see Epistemology in Islamic philosophy; Science in Islamic philosophy).

This approach also has implications for ontology. Illuminationism defends the 'primacy of quiddity'; it sets itself up against both Aristotle and Ibn Sina in upholding the priority of essence over existence. Some philosophers uphold the primacy of being or existence, and consider essence to be a derived mental concept, while those who adhere to the primacy of quiddity consider existence to be a derived mental concept. If existence has a reality outside the mind, then the real must consist of the principle of the reality of existence and the being of existence, which requires a referent outside the mind. Its referent outside the mind must also consist of two things, which can in turn be subdivided, and so on ad infinitum. That is, if 'existence' denotes an existent, then there must be another 'existence' connected to it which makes it real, and if so then this would also apply to the second 'existence', which leads to a vicious regress. To avoid this absurdity, we must regard existence as an abstract and derived mental concept; existence cannot signify an actual entity. If there were a distinction between a substance external to the mind and its existence, it would exist by accident, since two external substances must have different essences and cannot be distinguished by being existents. In that case, existence is nothing but a mental idea and cannot be defined. Since existence is attributable to many things, it must be mental (see Existence).

4 Light

Illuminationism is distinguishable from Peripateticism through its semantics, logic, epistemology, ontology, the priority of the intuitive over the purely noetic, and also its use of a language of light entities to describe the whole continuum of reality. The latter consists of four things: intellect, soul, matter and a fourth realm named the *'alam al-khayal*, which is similar to Platonic Forms except that entities in it are continuous with the whole of reality. This fourth realm (translated by Corbin as the '*mundus imaginalis*' (Corbin 1971)) is describable as that of 'things as ideas' prior to taking on shape, that is, before they receive 'luminosity' from the One Source, the Light of Lights. The light received is essentially the same, and the luminous thing differs from other light entities only in respect of degrees of intensity. Luminosity flows eternally, and gives shape to the forms, thus making the entity 'visible' and known. The difference between things, then, lies not in their essences but in terms of the degrees of intensity of the shared essences of the things. All luminous things constitute an aggregate whole and are coeternal with the Light of Lights. The Light of Lights is one, but is neither beyond being nor nonbeing, nor does it have a will. Everything in the continuum is generated from the Light of Lights and shares a degree of light similarity. The Light of Lights is one with respect to all possible modes, known or discovered subsequently.

These highly suggestive references to light were taken up by a large number of later thinkers who developed it in different but connected ways. According to Nasafi, the existence of God is an infinite light, the existence of which is equivalent to its essence, and everything which exists is a face or expression of this light. God is the ultimate reality of everything which exists in the universe. Baba Afdal Kashani argues that the notion of being is more general than the notion of existence, since we can wonder whether a thing actually exists; being is then prior to existence, and experience of the light of God's creation is comprehended solely through an internal illumination of the soul. This results not in knowledge of a fact or thing, but rather in a way of relating to God, a way which maintains one's status as part of the deity. One of the key aspects of Illuminationism is its disinclination to make a sharp distinction between God and that which God has produced. This is what has made Illuminationist philosophy

seem so close to mysticism at times, and it leads to a sharp differentiation from aspects of Ash‘arite thought, such as adherence to the atom as the basic constituent of physical reality.

5 The character of Illuminationist philosophy

The influence of Illuminationist philosophy on the Islamic world persists to this day. A wide range of important thinkers including Nasir al-Din al-Tusi, Shams al-Din Shahrazuri, Sa‘d Ibn Kammūna, Qutb al-Din al-Shirazi, Jalal al-Din al-Dawani, the School of Isfahan and, right up to our own day, Ha’iri Yazdi, are clearly within this tradition of philosophy. When it comes to issues of interpretation, there is a controversy as to how close this form of philosophy really is to mysticism. Some writers such as Izutsu (1971), Rahman (1975) and Ziai (1990), stress its links with analytical thought and deny that there is anything particularly mystical about it. It is certainly true that the greatest interest has been focused on a relatively small number of al-Suhrawardi’s works which have more of an esoteric nature, while his more technical and strictly logical works have tended to be ignored.

The work of Henry Corbin (1971) and Seyyed Hossein Nasr (1964), on the other hand, emphasizes the mystical contribution which Illuminationism makes, and they see the esoteric aspects of this form of thought as being of leading significance. There is no doubt that the general use to which Illuminationist philosophy has been put often involves mysticism, and there is little difficulty in combining it with the thought of Ibn al-‘Arabi, for example, which later philosophers were to do. It is certainly true that some of the leading texts by al-Suhrawardi are entirely technical and deal with issues of philosophy which have no mystical dimensions, but it must be admitted that when one examines his general approach to metaphysics, it clearly fits in with many of the ideas which mystics like to use. The terminology of light points to a view of the nature of reality which is far removed from that presented by the Peripatetics, or even from Ibn Sina in his more suggestive and mystical moods. Illuminationism is not just a critique of Aristotle and Ibn Sina, but it is also the development of an original metaphysical model which has subsequently proved very fruitful within the Islamic world.

See also: Aristotelianism in Islamic philosophy; Epistemology in Islamic philosophy; Illumination; Islamic philosophy, modern; Mystical philosophy in Islam; al-Suhrawardi

HOSSEIN ZIAI
OLIVER LEAMAN

References and further reading

Corbin, H. (1971) *En Islam iranien (Islam in Iran)*, Paris: Gallimard.(The main interpreter of illuminationism in the West and the esoteric approach to it.)

Ha’iri Yazdi, M. (1992) *The Principles of Epistemology in Islamic Philosophy: Knowledge by Presence*, Albany, NY: State University of New York Press.(Masterly analysis of this key notion in illuminationist philosophy, treated from the perspective of analytical philosophy.)

Izutsu Toshihiko (1971) *The Concept and Reality of Existence*, Tokyo: Keio Institute. (Detailed discussion of the notion of existence in illuminationism.)

Nasr, S.H. (1964) *Three Muslim Sages*, Cambridge, MA.: Harvard University Press. (Exposition of the Corbin form of interpretation with respect to illuminationism.)

Netton, I. (1989) *Allah Transcendent: Studies in the Structure and Semiotics of Islamic Philosophy, Theology and Cosmology*, London: Routledge.(Very clear account of the metaphysics of illuminationism.)

Rahman, F. (1975) *The Philosophy of Mulla Sadra*, Albany, NY: State University of New York Press.(An account of how later thinkers took up and developed illuminationism.)

al-Suhrawardi (1183-91) *Oeuvres philosophiques et mystiques*, vols I and II, ed. H. Corbin, Tehran and Paris: Adrien-Maisonneuve, 1976; vol. III, ed. S.H. Nasr, Tehran and Paris: Adrien-Maisonneuve, 1977.(A vitally important collection of the basic principles of illuminationism. An English translation of some of the works in Volume 3 can be found in *The Mystical and Visionary Treatises of Shihabuddin Yahya Suhrawardi*, trans. W. Thackston, London: Octagon Press, 1982.)

Walbridge, J. (1992) *The Science of Mystic Lights: Qutb al-Din Shirazi and the Illuminationist Tradition in Islamic Philosophy*, Cambridge, MA: Harvard University Press.(A commentaary on al-Suhrawardi and Ibn Sina, with an excellent discussion of the leading principles of illuminationism.)

Ziai, H. (1990) *Knowledge and Illumination: A Study of Suhrawardi’s Hikmat al-Ishraq*, Atlanta, GA: Scholars Press.(Study of the more analytical parts of al-Suhrawardi’s philosophy.)

Ziai, H. (1992) ‘Source and Nature of Authority: A Study of Suhrawardi’s Illuminationist Political Doctrine’, in

C. Butterworth (ed.) *The Political Aspects of Islamic Philosophy*, Cambridge, MA: Harvard University Press, 304-44.(Discussion of the political implications of lluminationism.)

Ziai, H. (1996) ‘The Illuminationist Tradition’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 29, 465-96.(A clear description of the topic, with the emphasis on the analytic aspect of illuminationism.)

Imagery

Most philosophers prior to the twentieth century thought of mental images as inner pictures, along lines suggested by introspection. But there are obvious differences between mental images and pictures. The former have no objective size or shape, for example. These differences have led some philosophers to argue that mental images are more like linguistic descriptions. The descriptional view of images is also taken by some cognitive psychologists. Other psychologists maintain that the pictorial conception of images provides the best explanation for the results of a number of intriguing experiments on imagery.

Are the front legs of a kangaroo shorter than its back legs? Questions like these typically prompt people to form mental images, which they then inspect before answering. But just what are mental images?

Introspection suggests that images are inner pictures. And that is how most philosophers prior to the twentieth century thought of them. According to this view, held by Aristotle, Descartes and the British Empiricists, visual images are significantly like pictures in the way that they represent things in the world (see Empiricism). Notwithstanding its widespread acceptance, the picture theory of mental images was left largely unexplained in the traditional philosophical literature. Images were taken somehow to resemble the things that they represent, but little more was said.

In the twentieth century the pictorial view of mental images has been the target of numerous philosophical objections. It has been noted that mental images are not seen with real eyes; they have no objective weight or colour. What, then, can it mean to say that images are pictorial? Moreover, mental images seem indeterminate in ways that pictures are not: for example, a mental image of a striped tiger need not represent any definite number of stripes. But an ordinary picture of a striped tiger must be determinate with respect to the number of stripes.

Finally, there have been attacks on the evidential underpinnings of the pictorial theory. The introspections on which belief in images is often based are taken to reveal much less about the mind than was traditionally supposed. In particular, it has been argued that what introspection really shows about visual images is not that they are pictorial, but only that what goes on in imagery is experientially much like what goes on in seeing (see Block 1983; Introspection, psychology of).

Perhaps the most influential alternative to the pictorial view in philosophy has been an approach that is now standardly known as 'descriptionalism'. The basic thesis of descriptionalism is that mental images represent in the manner of linguistic descriptions. They are complex representations, structurally like descriptions, within some neural code - representations that are the objects of computational procedures. So, descriptionalism about images is often taken to go hand in hand with the classical computational model of the mind (see Language of thought; Mind, computational theories of).

Descriptionalism remains popular in philosophy today (see, for example, Dennett 1981), and it also has significant support in contemporary psychology, where it has received its clearest articulation (see, for example, Pylyshyn 1981). Nevertheless in recent years the pictorial view has been making a comeback. In response to a large corpus of experimental data on imagery, psychologists such as Stephen Kosslyn (1980, 1994) have developed an empirical version of the pictorial view that seems more promising than any of its philosophical predecessors.

Kosslyn hypothesizes that the imagery medium, which he calls 'the visual buffer', is shared with visual perception. Each cell in the medium represents a tiny, just noticeable patch of object surface. He begins by conceiving mental images on the model of displays on a cathode ray tube screen attached to a computer. Such displays are generated on the screen by the computer from information that is stored in the computer's memory. This display is spatial and it is made up of a large number of basic units or cells, some of which are illuminated to form a picture. Analogously, according to Kosslyn, mental images may be regarded as functional pictures in a functional spatial medium made up of a large number of basic units or cells, that is, they function with respect to the representation of relative distance relations *as if* they were screen displays, even though, in reality, they need not be laid out in the brain with just the same spatial characteristics as real displays.

This last point is significant, if Kosslyn's view is to be properly understood. Even though the initial model for

mental images is provided by screen displays, there is an important respect in which images are *not* like such displays. The displays make no difference to the functioning of the machine. If the screen is shattered, for example, the machine will go on computing just as it did before. The displays are there for *us*, not for the machine. But that is not the case for mental images. They are critical to many cognitive tasks. On Kosslyn's view, mental images are really more like the internal arrays or matrix-like representational structures that computers sometimes manipulate, rather than the external screen displays. So, according to Kosslyn, there is nothing in the computational model of mind that commands adherence to descriptionalism. To be sure, computers manipulate internal descriptions, but they also manipulate internal, picture-like representations too.

This view is motivated by many intriguing experiments, of which I can mention only two. Kosslyn found that when people are told to form mental images of a map, which they have studied carefully, and to focus on one part of the image, the time it takes for them to shift their focus to another part of the image increases linearly with the distance apart of the objects on the map corresponding to the image parts. Kosslyn claims that this experiment shows that mental images can be scanned at fixed speed. This, in turn, is taken to support the pictorialist hypothesis (see Tye 1991).

In another experiment, Shepard and Metzler (1971) discovered that when people were shown drawings of pairs of block figures at different orientations, and were asked whether the members of each pair were congruent, the time it took the subjects to respond increased linearly with the angular separation of the figures. This experiment strongly suggests that mental images can be rotated (again at fixed speeds). This also fits in nicely with the view of images as having a picture-like (or array-like) character.

Not all cognitive scientists agree with pictorialism, however. According to Pylyshyn, Kosslyn's experiments on imagery can be explained by reference to the task demands placed on subjects by the experimenter's instructions together with facts the subjects already tacitly know, for example, the fact that it takes longer to scan greater distances with the eyes. No inner picture-like entity is needed.

It is far from clear that the appeal to tacit knowledge can explain all the imagery data, contrary to Pylyshyn's claim (see Kosslyn 1981). There are, however, other descriptional theories of imagery that have no need of the doctrine of tacit knowledge (for example, Hinton 1979). These versions also do well in explaining the results of Shepard and Cooper's rotation experiment. So, descriptionalism cannot be refuted simply by attacking the doctrine of tacit knowledge.

See also: Vision

MICHAEL TYE

References and further reading

Block, N. (ed.) (1981) *Imagery*, Cambridge, MA: MIT Press.(Useful anthology of central papers on both the philosophy and psychology of images, including many of the papers referred to here.)

Block, N. (1983) 'Mental Pictures and Cognitive Science', *Philosophical Review* 92: 499-542.(Philosophically sophisticated discussion of the current imagery debate.)

Dennett, D. (1981) 'The Nature of Images and the Introspective Trap', in Block (1981), 51-61.(Imaginative effort to undermine the introspective appeal of positing inner pictures.)

Hinton, G. (1979) 'Some Demonstrations of the Effects of Structural Descriptions in Mental Imagery', *Behavioral and Brain Sciences* 3: 231-50.(Defence of descriptionalism without an appeal to tacit knowledge of facts about visual processing.)

Kosslyn, S. (1980) *Image and Mind*, Cambridge, MA: Harvard University Press.(Early discussion by the major proponent in psychology of pictorial imagistic representations.)

Kosslyn, S. (1981) 'The Medium and the Message in Mental Imagery', in Block (1981), 207-44. (Pictorialist's reply to descriptionalism.)

Kosslyn, S. (1994) *Image and Brain: The Resolution of the Imagery Debate*, Cambridge, MA: MIT Press.(Recent defence of imagery in psychology.)

Pylyshyn, Z. (1981) 'The Imagery Debate: Analog Media Versus Tacit Knowledge', *Psychological Review* 88: 16-45; repr. in Block (1981), 151-206.(Psychologist's defence of descriptionalism.)

Shepard, R. and Metzler, N. (1971) 'Mental Rotation of Three-Dimensional Objects', *Science* 171: 701-3.(One of the earliest of the experiments adducing non-introspective evidence of pictorial representations.)

Tye, M. (1991) *The Imagery Debate*, Cambridge, MA: MIT Press.(Summary of the debate between pictorialism and descriptionalism and elaboration of an intermediate position.)

Imagination

'Imagination' and 'imagine' enjoy a family of meanings, only some of which imply the use of mental imagery. If I ask you to imagine a red flower, I will likely be inviting you to form an image. But if, for example, I say that I imagine that I will go to the party after taking a nap, I am not obviously giving voice to mental imagery. A variety of questions has arisen concerning imagination in its various forms, of which the following four are central. How do internal acts of imagining come to be about particular external objects and states of affairs, actual and non-actual? How are perceptual acts similar to and different from the central cases of imagining? To what extent does routine perception and cognition use similar cognitive resources to creative imagination? Are there any cognitive pursuits in which imagination can play a justificatory role?

1 The intentionality of imagination

The noun 'imagination' and the verb 'imagine' have a family of meanings rather than a single, sharp, meaning. We may use 'imagination' to indicate an ability to form images of things. Alternatively, 'imagination' can take on a more normative use, to indicate a measure of inventiveness, as when we say that something was done with imagination. Turning to 'imagine', we sometimes use that verb to describe image formation, sometimes as an indication of speculation ('I'd imagine that Jack the Ripper was responsible'), sometimes as an expression of surprise ('Imagine!'), sometimes to attribute error ('He imagines that Mars is the planet closest to the sun'), sometimes interchangeably with 'conceive'. Clearly, not all of these uses directly connote the use of mental imagery.

Exercises of the imagination have intentionality - that is, they are about objects, events and situations, actual and non-actual. We can imagine our best friend, a golden mountain, falling off a ladder, or that Martians visit Earth in the year 3000 (see Intentionality). Imaginings have not taken centre stage in twentieth-century philosophy of mind, which, perhaps regrettably, has tended increasingly to focus on belief and desire in its discussions of intentionality. Yet philosophers directly concerned with imagination have typically felt the need to provide an account of its intentionality. Taking those exercises of imagination that involve mental imagery as central, some have tried to explain the intentionality of imaginings in terms of the intrinsic properties of mental images, usually their pictorial properties. This approach is widely recognized as problematic. Even those who are willing to reify mental images recognize, first, that one cannot explain their intentionality by appealing to pictorial properties - pictorial properties cannot explain my imagining *my* cottage rather than another one that looks exactly like it - and, second, that there are plenty of mental acts that we call imaginings where mental images do not seem to be at play at all. As a result, accounts of the intentionality of imaginings have, increasingly, appealed to features other then the intrinsic properties of images - some to intentions that lie behind images, some to the mental activity of entertaining propositions, some to similarities between episodes of imagining and episodes of seeing, so that one imagines x just in case one is in a situation that is relevantly similar to seeing x. Of course, since 'imagine' has a variety of meanings, we need not require or expect that a single account will cover all exercises of imagination.

Many imaginative actions also have intentionality. For example, we can, by acting a certain way, make believe that a broom is a horse. It is natural to think that private mental episodes of imagining are conceptually primary, and that we can understand the intentionality of such actions in terms of the private imaginative episodes that cause them. This natural view has not gone unchallenged, however. Those of a more behaviouristic bent tend to recoil at the idea of private mental imaginings that are relatively inaccessible from a third person, and which could, in principle, be divorced from any link whatever to behaviour (see Behaviourism, Analytic; Privacy). An important source of broadly behaviourist ideas, Gilbert Ryle's *Concept of Mind* (1949), offers a package that eschews the existence of mental images before the mind's eye and instead takes *activities* of make-believe and pretence as the starting point for an account of imagination.

2 Imagining and perceiving

Paradigmatic cases of imagining seem to be akin to the act of seeing something. Aristotle (*On the Soul; On Dreams*: *c*.360s bc-320s bc) was sufficiently struck by the similarities between imagining and perceiving to claim that both were movements within the single faculty of sense perception.

One similarity that has drawn the attention of philosophers is that of perspective. Paradigmatically, we adopt a particular perspective when we imagine things. Just as it is difficult to see the front and back of a house at the same time, it is difficult to imagine the front and back of a house at the same time. This has been an important datum in discussions of imagination. It has been used in favour of the reification of mental images with pictorial properties, whereby the need for perspective in imagination can be likened to the need for perspective in a painting. And it has been used against the view that paradigmatic cases of imagining can be thought of as the mere entertaining of a set of propositions, for the reason that the entertaining of a set of propositions seems no more restricted to a single perspective than a set of architectural descriptions in a book.

While obviously similar, there are, equally obviously, differences between paradigmatic cases of imagining something and perception. The most obvious difference is external to the agent: when one perceives a thing, that thing will tend to figure in the explanation of the perceiving, but when one imagines a thing, it will typically be inappropriate to invoke that thing in an explanation of the imagining. But there are also differences of a phenomenological kind between imaginings and perceivings, differences that can be sensed from the first person point of view. One difference that has often been highlighted - by Aristotle (*On The Soul; On Dreams*), Sartre (1948) and Wittgenstein (1953), for example - is this: imaginings seem subject to the will, perceivings not. This point can be pushed too far. For there do seem to be cases where one's imagination runs riot (to the extent that one's imaginings may be involuntary), and yet there is something quite right about the idea that I can ask you to imagine something in a way that I cannot ask you to perceive or believe something. Other differences between imaginings and perception have been explored. A common idea is that imaginings involve the entertaining of thoughts that are neither affirmed or denied, whereas both perceiving and believing involve taking something as true. Sartre offers a rather different account. When one perceives x, he explains, the presence of x is posited by the perception. When one imagines x, not only is the presence not posited, but its *absence* is posited. For Sartre this contrast is an important clue to the nature of consciousness: the presence posited by perception is used to account for our sense of 'being-in-the-world', the posited absence by imagination being the fundamental exercise of freedom, whereby one withdraws from the world.

3 The imaginative and the routine

When a child comes to recognize dogs as dogs, does it use similar cognitive resources to those used by an artist who comes to see a cloud as a camel? An important theme among some philosophers of the imagination has been the idea that the cognitive faculties, skills or phenomena at work in our adventures of the imagination are also very much involved in the most routine episodes of cognition. In Hume, for example, ordinary categorization of things into kinds and the belief in independent continuous objects are attributed to the very faculty - imagination - that is responsible for storytelling. This faculty, we are told, makes faint copies of perceptions, recombining them in accordance with certain associative habits that Hume sets out to specify.

In Kant, the current theme is even more at work: perceptual experience cannot properly be thought of as a passive registering of information but, rather, requires that one actively apply concepts to sensation so as to see the world as this way or that, an activity for which imagination is responsible (though no implication of illusion is intended). Taking one's sensations as perceptions those of a dog is apparently treated by Kant as importantly analogous to the child's playfully seeing a broom as a horse. Kant's imagination has two species: reproductive, whereby one sees something as being of a kind with something one has previously seen; and productive, whereby a priori concepts are brought to bear upon experience.

In finding continuities between the creative and the routine, one should not obliterate that distinction. One particular kind of exercise of the imagination that calls out for special treatment is the aesthetic variety. While giving imagination an extremely broad role to play, Kant was well aware that there is something distinctive about the aesthetic exercise of the imagination and sought to explain what that was. His central idea was that aesthetic imagination, while thought-provoking, takes us beyond the cognitive domain provided by our concepts. This idea proved crucially important to those Romantic thinkers after Kant who sought to use the aesthetic imagination as the grounding for fundamental metaphysics (see Romanticism, German §3).

Both Kant and Hume use 'imagination' in an extended and thus unusual way in order to emphasis the putative continuities between the creative and the routine. A number of later writers such as Peter Strawson (1970) and Mary Warnock (1976) have also, in various ways, explored the continuity of imagination with humdrum

perception and cognition, though there has been a move away from the idea of explaining the continuity in terms of a single faculty - an organ of the mind - that does various sorts of creative and humdrum work.

4 The justificatory role of imagination

It is important to cognitive excellence that we cultivate imagination in all areas of inquiry. Even in science, it is clear that fruitful inquiry requires an ability to imagine what an explanation for some given phenomenon might look like, imaginatively to see familiar things as somehow representative of theoretical phenomena, and so on.

Granting that imagination is a source of cognitive inspiration, a further question concerns the extent to which an exercise of the imagination can *justify* various kinds of beliefs.

One interesting test case here is provided by geometry, where some thinkers, including Kant and J.S. Mill (1843), have accorded spatial imagination an important justificatory role, while others have insisted that it is merely a useful aid to the geometer.

Meanwhile, in discussions of possibility and necessity, considerable attention has been given to Hume's idea that whatever can be imagined is possible. Hume himself wielded this idea against any putatively necessary link between cause and effect, arguing that one can always at least imagine the cause without the effect. The crucial issue here is whether there is a sense of 'imagine' that can hold the various strands of the Humean view together. If 'imagine' means 'form a mental picture', as in a story, we can well doubt whether an ability to imagine something can justify our believing it possible, because the standards of propriety that we apply when telling stories are quite different to those we apply when the question of genuine possibility is at stake. We may also note in passing that it seems clear that there are possibilities we cannot picture. Meanwhile, there is a use of 'imagine' that is conceptually linked to possibility, but in that sense it is far from clear that I can imagine, say, eating every living creature on this planet.

Finally, in ethics, some (for example Johnson 1985) have suggested that an exercise in imagination is important to the rational evaluation of competing ethical claims. The idea is that the weighting and choice of values, as well as, perhaps, the concrete deployment of abstract or metaphorical moral principles, requires some sort of imaginative engagement with relevant possible situations.

See also: Imagery

J. O'LEARY-HAWTHORNE

References and further reading

None of these works is very technical. Russow and Strawson are recommended as introductory pieces.

Aristotle (*c*.360s BC-320s BC) *On the Soul* and *On Dreams*, in (ed.) R.McKeon *The Basic Works of Aristotle*, New York: Random House, 1941, 535-603 and 618- 25.(Mentioned in §3.)

Brann, E. (1991) *The World of Imagination*, Lanham, MA: Rowman & Littlefield.(An encyclopedic and cross-disciplinary study of our image-making ability.)

Casey, E. (1976) *Imagining: A Phenomenological Study*, Bloomington IN: University of Indiana Press.(A prominent philosopher of imagination in the phenomenological tradition.)

Hume, D. (1739-40) *A Treatise on Human Nature*, ed. L.A. Selby-Bigge. Oxford: Clarendon Press, 1978. (Discussed in §4 and 5. See especially book I, part I and book I, part IV, §2.)

Johnson, M. (1985) 'Imagination in Moral Judgment', *Philosophy and Phenomenological Research* 46: 265-80. (Explores a topic mentioned in §5.)

Kant, I. (1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1929.(Discussed in §4, mentioned in §5. See especially 'Transcendental Deduction of the Pure Concepts of Understanding'.)

Kant, I. (1790) *Critique of Judgment*, trans. J. C. Meredith, Oxford: Clarendon Press, 1952.(Discussed in 4. Contains discussion of aesthetic deployment of imagination.)

Mill, J.S. (1843) A System of Logic in *The Collected Works of J.S. Mill*, vol. 7, book II, ch.5.(Mentioned in §5.)

Russow, L. (1978) 'Some Recent Work on Imagination', *American Philosophical Quarterly* 15: 57-66.(Useful survey and bibliography.)

Ryle, G. (1949) *The Concept of Mind*, New York: Barnes and Noble, ch. 8. (Discussed in §2.)

Sartre, J.-P. (1948) *The Psychology of Imagination*, New York: Philosophical Library.(Discussed in §3.)

Strawson, P.F. (1970) 'Imagination and Perception' in L. Foster and J.W. Swanson (eds) *Experience and Theory*,

31-54.(Contains excellent discussion of Hume and Kant.)
Warnock, M. (1976) *Imagination*, London: Faber & Faber.(Useful history of the concept of imagination from Hume to the twentieth century, aimed at bringing out the importance for education of cultivating the imagination.)
Wittgenstein. L. (1953) *Philosophical Investigations*, trans. E. Anscombe, Oxford: Blackwell.(Mentioned in §3. A lot of suggestive and much-discussed material on imagination.)

Immanuel ben Solomon of Rome (*c*.1261-before 1336)

Immanuel of Rome wrote commentaries on the Bible and poems of a religious, philosophical or jocular nature, concerning the most varied themes. His main topics are conjunction with the 'active intellect' and the superiority of theoretical knowledge. He participated in many of the leading controversies of his time and is interesting for the ways in which he managed to express theoretical ideas poetically.

Immanuel ben Solomon ben Moses ben Jekhutiel was born in Rome and probably studied at the school (*bet midrash*) of Zerahiah ben Isaac Gracian of Barcelona. In 1290-1, during a controversy between Zerahiah Gracian and Hillel ben Samuel of Verona on the origin of human language and the rationalistic interpretation of biblical miracles, he sided with the former, writing a polemical letter to Hillel. Following the loss of his property, he left Rome and lived in various towns including Ancona, Perugia and Camerino. At Fermo he found shelter at the house of a patron, who requested that he assemble his scattered verses in a book. After a short time, however, he left for reasons unclear. Immanuel's father-in-law, Rabbi Samuel, was a well-known representative of the Jewish community of Rome. He was probably part of the mission sent by the community to obtain from the Pope, based at that time in Avignon, revocation of the decree of expulsion of the Jews from the Church's territory. After the murder of Rabbi Samuel in 1321, Immanuel sojourned in the towns of Perugia, Orvieto and Ancona and also visited Gubbio and the court of Can Grande della Scala in Verona. Following the revocation of the papal decree, he passed several years in Rome and then returned in 1328 to the house of his patron in Fermo.

In his vast literary-exegetic production Immanuel displays an encyclopedic knowledge, from the biblical and rabbinic literature and the oldest Hebrew mystical works, to such disciplines as poetry, grammar, rhetoric, music, astronomy, mathematics, medicine and philosophy. Belonging to the Aristotelian school, although introducing certain ideas drawn from the Neoplatonists, he deals with the entire gamut of philosophical topics in his exegetic commentaries. If from a literary point of view he is influenced by his immediate environment, in particular the Italian school of *dolce stil nuovo*, be it in the ideas or in style, introducing the fourteen-line Petrarchian sonnet to Hebrew poetry (while applying Arabic metre to the Italian verse), from the philosophic point of view, as is seen in his vast exegetic production, he is part of the Maimonidean-Tibbonian movement.

In his exegetical works, he continues in the tradition of the Tibbonids, adopting some of the formal components of the European university teaching method of the twelfth and thirteenth centuries: a great part of each commentary is actually devoted to the work of previous scholars and only a small part to original exegesis of the text, in accordance with the fourfold division of the role of the writer defined by the Latin scholastics (*scriptor*, *compilator*, *commentator*, *auctor*). In his *Perush ha-Torah (Commentary on the Pentateuch)*, Immanuel, under the influence of Abraham ibn Ezra's hermeneutical method, gives a fivefold interpretation of the verses: the Aramaic translation, a grammatical analysis, a literal interpretation, a philosophical interpretation, and a mystical or esoteric explanation. In his other biblical commentaries as well, he begins with a grammatical analysis of the verse, followed by an exposition of the literal and then the philosophical-allegorical meaning, frequently offering several interpretations for the same verse. In the philosophical section of *Perush le-be-Reshit (Commentary on Genesis)*, Immanuel interprets the creation of the sublunar world naturalistically. He also addresses topics such as time, motion, the nature of the spheres, the nature of light as identified with the celestial matter (Aristotle's fifth element), the nature of the prime matter, the concept of the creation existing for the sake of man, the origin of human language, the superiority of the Hebrew language, the doctrine of prophecy, and the role of the imagination in the prophetic revelations.

There are links between his poetical and exegetical-philosophical works. In his best-known literary work, *Machberot (Compositions)*, he often expresses himself in allegorical language, characteristic of the Maimonidean-Tibbonian school, in order to convey the same philosophical themes that occupy much of his exegetical commentaries, such as the influence of the heavenly bodies, the cognitive process of the human rational soul and the immortality of the human soul. The *Machberot* are organized as a dialogue between the author and a prince. Scholars are inclined to see in the prince who converses with Immanuel a symbol signifying the 'active intellect', with which the potential intellect of the poet strives to unite. These topics return frequently in the exegetic writings, in which Immanuel repeatedly emphasizes the superiority of intelligible knowledge, of which

the highest degree of perfection is represented by the conjunction of the potential intellect with the last of the separate intelligences, namely, the active intellect. It is precisely this conjunction of the potential intellect with the active intellect, allegorically symbolized by the union between the lover and the beloved, that constitutes the main topic of *Perush le-Shir ha-Shirim*, his philosophical commentary on the Song of Songs.

CATERINA RIGO

List of works

Immanuel ben Solomon of Rome (before 1328) *Perush le-be-Reshit (Commentary on Genesis)*, trans. and ed. F. Michelini Tocci, *Il commento di Emanuele Romano al Capitolo I della Genesi*, Rome: Centro di studi semitici, Istituto di studi del vicino Oriente, 1963.(Immanuel's commentary on the first chapter of Genesis. This Italian edition is introduced by the editor.)

Immanuel ben Solomon of Rome (before 1328) *Perush le-Megillat Shir ha-Shirim-ha-Cheleq ha-Filosofi (Commentary to the Song of Songs - the Philosophical Part)*, ed. with notes by Y. Ravitzky, Jerusalem: Hebrew University, 1970.(Written as an MA thesis, this edition comprises the Hebrew text of the philosophical section of Immanuel's commentary on the Song of Songs and critical study of it.)

Immanuel ben Solomon of Rome (before 1328) *Sefer Mishle ʿim Perush Immanuel ha-Romi (The Book of Proverbs with the Commentary of Immanuel of Rome)*, ed. D. Goldstein, Jerusalem: The Jewish National and University Library Press and The Magnes Press, 1981.(This edition, which is accompanied by an introduction by the editor, is a facsimile of the Naples edition, *circa* 1487. It is the only completely published edition of a commentary by Immanuel on a biblical text.)

Immanuel ben Solomon of Rome (1328) *Machberot Immanuel ha-Romi (The Compositions of Immanuel of Rome)*, ed. D. Jarden, Jerusalem: Mossad Bialik, 1957, 2 vols.(Critical edition of Immanuel's major literary work. Jarden provides an introduction to and commentary on the work, and a comprehensive bibliography.)

References and further reading

Cassuto, U. (1971) 'Immanuel (ben Solomon) of Rome', *Encyclopedia Judaica*, Jerusalem: Keter, vol. 8, 1295-8. (Account of his life and works.)

Goldstein, D. (1971) 'Longevity, the Rainbow and Immanuel of Rome', *Hebrew Union College Annual* 42: 243-50.(Analysis of two passages of Immanuel's commentary on Genesis.)

Ravitzky, A. (1981) 'Immanuel of Rome's Commentary on the Book of Proverbs - its Sources', *Kiryat Sefer* 56: 726-39.(Analysis of the sources of Immanuel's commentary on Proverbs.)

Schechterman, D. (1984) 'The Philosophy of Immanuel of Rome in Light of his Commentary on the Book of Genesis', unpublished Ph.D. dissertation, Jerusalem: Hebrew University.(A study of the philosophical section of Immanuel's commentary on Genesis.)

Sermoneta, G. (1965) 'La dottrina dell'intelletto e la "fede filosofica" di Jehudàh e Immanuel Romano', *Studi Medievali* series 3, 6 (2): 3-78.(Compares Immanuel's doctrine of conjunction with the active intellect with that of Judah ben Moses.)

Sermoneta, G. (1976) 'Yehudah and Immanuel ha-Romi, "Rationalism Culminating in Mystical Faith"', in M. Hallamish and M. Schwarz (eds) *Revelation, Faith, Reason*, Ramat Gan: Bar Ilan University, 54-70.(Examines Immanuel's doctrine of conjunction with the active intellect.)

Shiloah, A. (1993) 'A Passage by Immanuel ha-Romi on the Science of Music', *Italia* 10: 9-18.(An analysis of a passage of Immanuel's on the science of music included in his commentary on Genesis.)

Immutability

The doctrine of divine immutability consists in the assertion that God cannot undergo real change. Plato and Boethius infer divine immutability from God's perfection, Aristotle from God's being the first cause of change, Augustine from God's having created time. Aquinas derives divine immutability from God's simplicity, his having no parts or attributes which are distinct from himself. All of these arguments finally appeal to aspects of God's perfection; thus, the doctrine of divine immutability grew from a convergence of intuitions about perfection. These intuitions dominated Western thought about God well into the nineteenth century.

The doctrine's foes argue that God's power, providence and knowledge require its rejection. Their arguments contend that since the world does in fact change through time, this must entail change in God. If God responds to changing historical circumstances and to prayers, that would seem to require some sort of change in him (from not responding to responding). And if he does not intervene to prevent a war, for example, then after the war, he will have lost the power to prevent it (assuming, as many do, that God cannot alter the past), so again there is a change of state. Finally, it is argued that God's knowledge of tensed truths (for example, 'it is now noon') must change as what time 'now' is changes. Some responses to these arguments appeal to the claim that God is in some sense outside time.

1 Introduction

From Philo of Alexandria to the nineteenth century, 'classical theism' ruled Western thought about God, at times without challenge. Prominent in classical theism was the doctrine of divine immutability, that God cannot change in nature, character, intentions, knowledge or any other respect. In such writers as Augustine and Aquinas, God is eternal because he is immutable, and eternality is God's distinctive mode of being. So the doctrine of divine immutability is at the roots of such writers' understandings of God's nature.

The doctrine consists in the assertion that God cannot undergo real change, such as the change involved in learning, growing or reddening. Divine immutability does not entail that God cannot begin or cease to exist, if these are not real changes in the thing which begins or ceases. Nor does it rule out purely extrinsic changes in God, changes such as becoming admired. A thing changes purely extrinsically when it gains a property due entirely to real change in some other thing (see Change §1). The doctrine allows that God becomes creator of a new creature when Moses is born. But it insists that this change in God is purely extrinsic. God always willed that Moses be born at *t*. When *t* arrived, his will took effect in real changes in Moses, his parents, and so on. But the only change in God was that he came to deserve a new title, Creator of Moses.

The doctrine of divine immutability is distinct from the doctrine of divine impassibility. Nothing external can affect an impassible entity. Something could be impassible but mutable if it could change itself, but nothing else could change or affect it. Again, God could be immutable but not impassible. An unchangeable God could feel such responsive emotions as love, pity and compassion. But he would feel them without change and so always feel them. If temporal, such a God would pity our pain before, while and after we suffer. If timeless, such a God would pity our pain timelessly - but responsively, because of our pain. Thus, the doctrine of divine immutability need not 'depersonalize' God.

Western Scriptures insist that God is unchanging in some respects: if he makes a promise, he never alters his intent to keep it, and his basic character (benevolence, holiness, great power) never alters. But Western Scriptures seem to conflict with full divine immutability. Some scriptural texts depict human sin as saddening God and so changing his feelings (for example, Genesis 6: 6), then bringing God to new decisions - for example, to flood the world. According to John, 'the Word became flesh' (1: 14), that is, God took on a human nature he did not always have. The doctrine of divine immutability appears in theologies which take Western Scriptures as authorities. It thus is worth asking why Western theism has so strongly endorsed the doctrine, given that its scriptural roots seem to deny it.

2 The case for immutability

The answer lies in the philosophical sources and methods of Western religious thought. Western thinkers have largely followed the method of perfect-being theology, filling out the concept of God by ascribing to him the

properties he must have to count as absolutely perfect (see God, concepts of §§2-6). God's perfection seems to rule out many sorts of change. If perfect, God is all-knowing. If God learns something new, then before that he was not all-knowing. If God changes his plans, either he did not anticipate what makes him change them and so was not all-knowing, or he was not perfectly wise in making his plans, or he is not perfectly wise in changing them now. So if God is *necessarily* all-knowing and perfectly wise, God *cannot* change in these ways. More general arguments convince adherents of perfect-being theology that God cannot change in any way.

Plato's Theory of Forms inclined many to believe in divine immutability (see Plato §§10-14). In the *Phaedo* and elsewhere, Plato takes the Forms to be perfect cases of themselves - for example, the Good is a perfectly good thing - and associates the Forms' perfection with their immutability. Theists' liking for Plato's views disposed them to see perfection Plato's way. Plato argues specifically that God cannot improve or deteriorate: if God is already perfect, God cannot change for the better, and being perfect includes being immune to change for the worse (*Republic* II 381b-c). Plato ignores value-neutral changes. But there may be some; is it good or bad to become a nanosecond older? Still, God would undergo value-neutral change either voluntarily or involuntarily. If involuntarily, then either God wishes not to change but cannot avoid it, or God does not care whether he changes, but is subject to forces or conditions he does not control. Either way, God's power seems less than perfect. If voluntarily, God is less than perfectly rational, since he does or undergoes something with literally no value, and hence for which there is no good reason. (If God creates time and thereafter grows older, the good for which he creates time is reason to let himself grow older, and so growing older is not value-neutral.)

Many medieval theists accepted Aristotle's 'cosmological' argument for God's existence (see Aristotle §16; God, arguments for the existence of §1). Aristotle reasoned that if change occurs, change has a final source, an eternally unchanged changer (*Physics*). He also argued that something is eternally unchanged only if unchangeable (*De caelo*). Later theists thought the role of first cause of change too lofty not to be God's. Writers who took Aristotle's argument or its variants to prove God's existence found themselves committed to divine immutability.

Augustine (§§7-8) developed Plato's thoughts into an ontology of degrees of existence. As he thought that to exist is to be present, his ontology was also of degrees of presence. Augustine thought immutable things most present. For a thing which changes realizes its nature gradually and never at any point has all the attributes it ever has; a human gains adult stature only by losing the distinctive charms of childhood. A thing which *can* change never actually has all the attributes it possibly has. If a thing is immutable, its entire character and nature are present at any moment: it always is all it can be. Thus, it is more fully present than changing things. So if existence is presentness, immutable things have a higher degree of existence than mutable things. As only highest existence is appropriate to a perfect being, perfect-being theology thus led Augustine to divine immutability.

Augustine also argued specifically (in *The Literal Meaning of Genesis*) that by creating temporal things, God created time itself. As free creator of time, God is intrinsically outside time (as Philo of Alexandria (§5) too saw, in *On the Unchangeableness of God*). But only things which exist in time can change. Hence, if God is intrinsically beyond time, God is intrinsically changeless. Augustine's understanding of God's role as Creator, then, led him to divine timelessness and so to divine immutability. One might move to divine immutability via timelessness today from the widely accepted premises that God is not located in space and (on a standard reading of Special Relativity) things are in time only if in space (see Time §3).

Temporal beings no longer live the past parts of their lives, losing them forever, and still lack the future parts of their lives. A.M.S. Boethius (§5) saw these traits as defects. So he inferred via perfect-being theology that God has no past or future, living his life in a pure present (see Eternity). What has no past or future does not change. For what changes goes from what it was to what it then was going to be, and so has a past and a future. Hence, for Boethius, perfection required changelessness. If it does, necessary perfection - which is better than contingent perfection, and so by perfect-being theology is God's - requires being immutable.

Aquinas (like Augustine) derived divine immutability from the deeper classical theist doctrine of divine simplicity (*Summa theologiae* Ia 9.1). If God is simple, God has no parts of any sort, including distinct attributes. Something which changes becomes partly different (else there was no change) and stays partly the same (else there was not change in one selfsame surviving thing; instead, one thing disappeared and another replaced it). So only things with parts can change. If so, a simple God cannot change. Divine immutability's connection with divine simplicity and the classical theist theory of God's perfection which centres on divine simplicity is one of the deepest reasons

for its broad historical appeal; hence one cannot fully explain what moved thinkers to accept it without also treating the motivation for the doctrine of divine simplicity (see Simplicity, divine).

The doctrine of divine immutability was accepted, then, as intuitions about God's perfection and creatorhood came together with the seeming availability of arguments proving an unchangeable First Cause's existence. This convergence of intuitions and arguments from varied sources was the equivalent in philosophy of confirmation from independent tests in science: it strongly warranted belief in divine immutability.

3 The case against immutability

Direct challenges to the doctrine of divine immutability have rested on Scripture and on God's providence, power and knowledge. Scripture depicts God as changing, and so some who read it literally reject divine immutability. Facing such texts, friends of the doctrine defuse the appearance of divine change by appeal to doctrines less speculative and theoretical than divine immutability. Thus, Philo argues from God's foreknowledge of the future and constancy of character that God cannot repent or feel regret, as the Flood story suggests (*On the Unchangeableness of God*). Divine immutability's friends also stress the strengths of the overall theory (classical theism) that includes the doctrine; theories' overall virtues can be good reason to interpret data in a way one might not otherwise consider.

Providence. Believers in divine providence think that God answers prayers and acts in history. Parting the Red Sea (for instance) was a new act in response to historical developments. God first was not doing it, then did it. Thus, God's providence seems to involve him in change, partly because he responds to changes (see Providence §1).

Divine immutability's partisans again respond to this by appealing to less theoretical theistic doctrines. They argue that before making the world, God knew what would or might occur and what would or might be prayed if he made it. Thus, God never has new decisions to make. As all-wise, God had full plans to meet events even before making the world. So in creating the world, God built into it 'trigger' mechanisms, which would bring about the responses God willed given the right events. Thus, God made his providential contribution all at once, without changing himself. God willed changing providential responses without having to change his will. God was spontaneous and creative, but expressed his creativity all at once rather than sequentially. Does this mean (oddly) that God answered our prayers before we prayed them? Not if God is also timeless; a timeless God's actions are not *earlier than* anything.

The Incarnation was an especially knotty problem for divine immutability's Christian friends. In general, these argued that all the change it involved occurred in the human nature which God the Son assumed rather than in God; God was eternally ready to be incarnate, and eternally had those experiences of the earthly Christ which the Incarnation makes part of God's life. Through changes in Mary and the infant she bore, what was eternally in God eventually took place on earth (see Incarnation and Christology §§1-2).

Power. Before the Ice Age, God could ensure that there would never be an Ice Age. God has this power now only if he can alter the past. Few think he can (see Omnipotence §§1, 3). So events seem to change God's power. (If they do, God also is passible; other things affect him. Further, if God once had and no longer has a power, God has a past, and so is in time.) Defenders of divine immutability reply that any change here is purely extrinsic. God has the intrinsic power he always has. He has lost a chance to use it, and so we no longer want to *call* his power a power to prevent an Ice Age. But God is intrinsically as able as ever to do so.

Omniscience. Tensed facts and truths are those we state in tensed sentences such as 'today is Christmas' or 'it is now noon'. An argument that divine immutability is incompatible with God's being omniscient and there being tensed facts and truths can be traced back at least to Islamic Mu'tazilite theologians and Ibn Sina. The argument is this: if God is omniscient, then he knows what time it is now. Now, at t, what God knows is that it is now t. Only at t can God know that it is now t, since only at t is this true. Later, at $t+1$, what God knows is that it is now $t+1$. God can know that it is now $t+1$ only at $t+1$. So the content of God's knowledge changes. So God changes (see Omniscience §2). In reply, some argue that:

(a) No truths are *irreducibly* tensed. Rather, 'it is now t' really reports (say) that 'at t, it is t'. The latter truth is tenseless. It is true at all times. So one can know it at all times, without change. If God knows tensed truth by knowing tenseless truths, he knows at each time what time it is, without changing.

(b) God can know in other ways the facts we express in irreducibly tensed truths. This is enough to qualify him as omniscient.
(c) What the tensed facts are depends on one's temporal standpoint. If so, then if God is not just immutable but timeless, his omniscience need encompass only tenseless facts plus whatever tensed truths there are in eternity. The latter do not change.

One may also wonder whether the nature of God's mind undercuts the argument. On some externalist theories of mental content, my knowledge that *P* is a complex consisting of my inner mental state plus certain items in the world. Those items are where and when they are, yet the knowledge is 'in my mind', and so (one presumes) where and when *I* am. If God is timeless and his mind's contents are as externalism suggests, his state of knowledge might be a complex consisting of his mental state plus temporal items; this knowledge might be 'where and when' God is - that is, timeless - and, as timeless, never change.

See also: Ash'ariyya and Mu'tazila §4; Natural theology; Necessary being; Negative theology; Process theism

BRIAN LEFTOW

References and further reading

Aquinas, Thomas (1266-73) *Summa theologiae*, New York: Benziger Brothers, 1948. (*Locus classicus* of classical theism.)
Aristotle (mid 4th century BC) *De caelo*, trans. R. Hardie and R. Gaye, in R. McKeon (ed.) *The Basic Works of Aristotle*, New York: Random House, 1941.(A theory of physical change; argues that only what cannot change can be eternally changeless.)
Aristotle (mid 4th century BC) *Physics*, trans. J. Stocks, in R. McKeon (ed.) *The Basic Works of Aristotle*, New York: Random House, 1941.(Treats the nature of change, and argues for God's existence.)
Augustine (401-15) *De genesi ad litteram (The Literal Meaning of Genesis)*, trans. J.H. Taylor, New York: Newman, 1982, 2 vols.(Clear, interesting source for many aspects of his theology and metaphysics; asserts that God creates time (V.5.12).)
Boethius, A.M.S. (*c*.522-4) *The Consolation of Philosophy*, trans. H. Stewart, in H. Stewart and E.K. Rand (trans.) *Boethius: The Theological Tractates*, Cambridge, MA: Loeb Classical Library, Harvard University Press, 1936.(Includes his most extensive treatment of time, change and eternity.)
Gale, R. (1986) 'Omniscience-Immutability Arguments', *American Philosophical Quarterly* 23 (4): 319-35.(Very full statement of the argument, with intricate exploration of possible replies. Rigorous but not technical.)
Hallman, J. (1981) 'The Mutability of God: Tertullian to Lactantius', *Theological Studies* 42 (3): 373-93. (Historical survey, mostly on divine emotion; blurs the distinction between immutability and impassibility.)
Hartshorne, C. (1948) *The Divine Relativity*, New Haven, CT: Yale University Press.(Freewheeling, many-faceted critique of the doctrine of divine immutability, and an argument for divine passibility.)
Helm, P. (1988) *Eternal God*, New York: Oxford University Press.(Includes a nice treatment of the omniscience-immutability issue.)
Kretzmann, N. (1966) 'Omniscience and Immutability', *Journal of Philosophy* 63 (14): 409-21.(Independent rediscovery of the medieval omniscience-immutability argument.)
Leftow, B. (1991) *Time and Eternity*, Ithaca, NY: Cornell University Press.(Treats divine immutability and related issues in defending divine timelessness. Rigorous but not technical.)
Mann, W. (1987) 'Immutability and Predication', *International Journal for Philosophy of Religion* 22 (1): 21-39. (Elegant defence; traces the doctrine of divine immutability to divine simplicity.)
Philo of Alexandria (first half of 1st century AD) *On the Unchangeableness of God*, trans.F. Colson and G. Whitaker, in F. Colson and G. Whitaker (trans.) *Philo*, vol. 3, Cambridge, MA: Loeb Classical Library, Harvard University Press, 1960.(Pioneering defence of divine immutability in the face of recalcitrant scriptural texts.)
Plato (*c*.386-380 BC) *Phaedo*, trans. G. Grube, Indianapolis, IN: Hackett, 1977. (Develops his theory of Forms.)
(*c*.380-367 BC) *Republic*, trans. G. Grube and C. Reeve, Indianapolis, IN: Hackett, 1992.(Treats Forms and divine immutability (II 381b-c) during a broad discussion of human nature and political theory.)
Sorabji, R. (1983) *Time, Creation and the Continuum*, Ithaca, NY: Cornell University Press.(Valuable, clear treatment of many issues related to divine immutability.)

Impartiality

On the one hand, most of us feel that we are permitted, even required, to give special consideration to the interests of ourselves and our loved ones; on the other hand, we also recognize the appeal of a more detached perspective which demands equal consideration for the interests of all. Among writers in the utilitarian tradition, some insist that the strictly impartial perspective is the only one that is ethically tenable, while others argue that a measure of institutionalized partiality can be justified as a means to maximizing welfare. An alternative tradition, stemming from Kant, sees the demand for impartiality as deriving from the importance of fairness and equal respect for persons, but tends to leave open the degree of partiality permitted. Finally, the Aristotelian conception of ethics offers a justification of partiality based on the structure of those virtuous dispositions of character (such as those involved in friendship and self-esteem) which are required for developing our distinctively human potentialities.

1 The conflict between impartiality and partiality

Impartiality commands widespread approval. When an action is exposed as an instance of bias we disapprove precisely because the agent has allowed special ties to particular individuals or groups to affect what ought to have been a fair and impartial decision. Some philosophers, indeed, see impartiality as part of the essence of the moral outlook, believing that to adopt that outlook is, in Adam Smith's famous phrase (1759), to try to put oneself in the position of an 'impartial spectator' of mankind. Pulling in the opposite direction, however, are the special commitments that most human beings have towards their close friends and loved ones (see Family, ethics and the; Friendship; Love §2). For a parent to treat a child with no more than the impartial consideration owed to any human being seems cold and unloving. If morality is concerned with the good for humankind, then it must surely find a legitimate place for the development of those special, inherently partialistic relationships which are virtually inseparable from our human nature. A major task for moral philosophy is to examine, perhaps to try to resolve, the apparent conflict between those of our ethical intuitions which seem to demand an impartial stance, and those which apparently permit, or even require, a large measure of partiality.

The importance of impartiality in moral thinking is given force by the thought that although my own life may appear 'special' from my personal perspective, that specialness seems to evaporate from the wider perspective of humanity as a whole. In assessing the rightness of an action, by reference to its consequences for general welfare, Jeremy Bentham, as quoted by J.S. Mill (1861), insisted on the maxim 'everybody to count for one, nobody for more than one', which seems to rule out skewing the calculation in favour of particular individuals or groups. Hence if benefiting X produces a greater overall good than benefiting Y, it is right to benefit X; the fact that I have some special tie to Y (who could be, for example, my brother, or a member of my club, clan or race) ought presumably to make no difference. William Godwin (1793) devised a famous case to illustrate this: if two people (a philanthropic archbishop and a chambermaid) are trapped in a burning building, and I can rescue only one, then I should rescue the one who can do most good for mankind as a whole. Given that this is the archbishop, then it is he who should be rescued; I should resolutely set aside the fact that the chambermaid happens to be my mother, for 'what magic is in the pronoun "my" that should justify us in overturning the decisions of impartial truth?' ([1793] 1985: 170). Such resolute impartialism offers a potentially salutary challenge to many of our unreflective patterns of conduct, but, for its critics, it risks setting an unrealistic standard for action which few if any are capable of incorporating systematically into their daily lives.

2 Utilitarian approaches

Utilitarians have sometimes seemed quite prepared to embrace the austere consequences of the impartialist perspective (see Utilitarianism). Thus Peter Singer (1979) suggests that if I am dishing out food to hungry children, then I should make my own children wait in line; though favouring them might be psychologically understandable, from the ethical point of view it is suspect. There are two serious objections to any approach which yields this result. The first has the form of a 'slippery slope' argument. If all personal favouritism is ethically dubious, then we are apparently required to forego any good whatsoever for ourselves and our loved ones when an assignment of the relevant resources to others would produce greater global utility. Yet this implies not just (as most of us might on reflection accept) that I should give up luxuries when the money might be spent on famine relief, but that even the decisions to spend time and money on providing a home for my family, or on working at my own chosen

career, are ethically tainted. In short, the insistence on impartialism seems to entail the unbearably demanding result of uncompromising ethical ‘globalism’: as long as one surplus penny remains which could benefit those in greater need, nothing short of a life of near total self-sacrifice will stand up to ethical scrutiny (see Help and beneficence). Defenders of strict impartialism, such as Shelly Kagan (1989), would reply that the fact we feel uncomfortable with this result is no reason to reject the underlying theory, for who ever imagined that to act morally was supposed to be easy?

Here, however, a second objection comes to the fore. There seems a risk that what the impartialist apparently requires us to give up will include those very ties and commitments which make human life worth living, the personal projects (raising a family, developing long-term friendships, organizing a career) which are the very key to realizing our human potentialities. The point goes beyond the empirical observation that all these projects in fact require an enormous expenditure of time and energy on localized and individual (as opposed to general and global) welfare. For it appears, as Bernard Williams (1973) has argued persuasively, that our very individuality, our integrity as persons, depends largely on a self-oriented conception of ourselves (and by extension our loved ones) as special. Without that sense of specialness, and the partiality towards ourselves and those near to us which it implies, our very humanity is threatened: a being who lacked this orientation would be a mere worker bee in the utility hive, not a recognizable human being (see Morality and identity §4).

These observations pave the way for a response structured by an indirect (institutional or ‘rule’) version of utilitarianism. If human happiness indeed requires us to develop self-oriented projects and ties, then the way to maximize global utility will turn out to be not to require austerely impartial behaviour from all, but rather to allow humans to develop those institutions, practices and relationships which permit them to fulfil their human nature (see Utilitarianism §§3, 5). The argument will then run in a way analogous to that which led Adam Smith to suggest that the pursuit of individual self-interest is in fact the best way to secure general happiness. Given this framework, the utilitarian may well recommend that each of us devote time and attention to self-oriented projects, such as raising families and forming close personal relationships; but at a more general and abstract level, the justification for such social structures will still be a firmly consequentialist one, namely that the fostering of relevant institutions is the best way for global happiness to be secured. Such an approach has the advantage of providing a critical framework for trying to distinguish those types of partiality which are permissible from those which are not. Thus, public institutions which foster cronyism, nepotism and patterns of preference based on gender or race may be condemned as diminishing the prospects for the long-term maximization of global happiness, while family loyalties and ties of friendship may be seen as benign. A further advantage is to allow room for preserving intuitions about the appeal of impartiality as an ethical ideal; for the ‘benign’ nature of those institutions which are justified by this type of argument will hinge precisely on the fact that they are the kinds of institution which would be approved by an impartial and detached spectator of the global human predicament.

3 Impartiality, fairness and respect

A problem with trying to link the value of impartiality to the demands of welfare maximization is that it seems possible to imagine that global utility might be maximized by partialistic institutions of an (intuitively) objectionable kind. A master race might secure enormous benefits by preserving entrenched institutions according systematic preferential treatment to their own kind; and the resulting highly partialistic system might be one which would have to commend itself to an impartial spectator interested in global utility. This line of thought suggests the need for a conception of impartiality that goes beyond purely aggregative and utilitarian concerns. On one such conception, a truly impartial evaluator would require not just the best overall result, but one that would be recognized by all parties as acceptable, irrespective of their position in the social order. This is the thought behind John Rawls’ device of the ‘veil of ignorance’ (1971), designed to secure fairness by ensuring that social institutions are ones we could in principle approve of even if we had to change places with the least advantaged of our fellow citizens (see Rawls, J. §1). It is not, however, clear that the application of these conditions yields one uniquely ‘rational’ choice. Just as two equally impartial judges might favour, respectively, stern retribution and maximum leniency, so one impartial deliberator might prescribe highly competitive social institutions which allowed the weak to go to the wall, while another might favour heavy taxation to fund lavish welfare benefits. Pure impartiality, in short, appears to be merely a ‘meta-requirement’, demanding consistency in the application of rules to all concerned, but not providing clear decisions about the ethical credentials of various possible social institutions.

The approach of Rawls owes much to the tradition in ethics started by Kant (1785), which, with its stress on respect for persons, would prohibit any action or institution which treats individuals merely as a means to an end (see Kantian ethics; Respect for persons). Although this ‘categorical imperative’ is silent on the ethical status of personal ties and commitments, it suggests room for a compromise between partialistic ethics and impersonal duty: love and other partialistic emotions and patterns of conduct are permitted, so long as they do not violate (impartially defined) Kantian rules of justice and respect for others. But while such a compromise might operate satisfactorily in an economy of abundance, the preferential assignments demanded by the partialist will, in an economy of scarcity, inevitably raise questions about our ethical obligations to relieve suffering elsewhere. In devoting resources to myself and my loved ones, perhaps I may not formally be violating Kantian rules against treating others without respect, but I am certainly reserving for myself goods which might be used to enable others to reach the level of dignity and self-respect which I take for granted in pursuing my own partialistic concerns and projects. The conflict between the autocentric perspective of the partialist and the more universal stance of both deontological and consequentialist ethics seems inescapable (see Deontological ethics).

4 The perspective of virtue ethics

The tradition of so-called ‘virtue ethics’ approaches these questions from a rather different standpoint (see Virtue ethics). For Aristotle, the dispositions of character to be fostered are those which develop our potentiality for human fulfilment, and though impartial virtues such as justice figure importantly in the resulting blueprint for the good life, many of the other virtues, most notably those connected with love and friendship, seem inherently partialistic, involving special commitments to those to whom I am bound by close personal ties (*Nicomachean Ethics*) (see Aristotle §§25-6; Virtues and vices). Aristotle draws several interesting comparisons between my attitude to my friends and my attitude to myself, and it is striking that he allows a perfectly legitimate ethical space for self-love, or self-esteem. Aristotelian virtue thus allows me, in the assignment of goods, to give special preference to a particular individual, myself. Several twentieth-century philosophers have laid great stress on universalizability as a test for morality, holding that there is no place in ethics for principles which contain an ineliminable reference to particular individuals (see Universalism in ethics). It seems, though, that there is nothing to stop the advocate of Aristotelian-style partialism allowing that every human being is entitled to accord special preference to themselves and their loved ones, so that there is a sense in which the ethics of partiality do, after all, pass the test of universalizability. The difficulty remains, however, that the assignment of time and resources necessary to develop a rich and fulfilling life for myself and my close friends will inevitably mean that less remains for those outside the charmed partialistic circle; Aristotle himself was largely unbothered by slavery. That said, it remains an attractive feature of the Aristotelian approach that it allows due ethical value to those deep personal concerns and emotional commitments that appear to be crucial ingredients in any plausible recipe for a good human life. Ethical excellence, is, Aristotle insisted, concerned with feelings as well as actions; a significant part of our fulfilment lies in the cultivation of harmonious and enriching habits of feeling (see Morality and emotions §3). Since so much that is of value in our emotional life seems to have an ineradicable dimension of particularity and partiality, any system of morality which fails to allow scope for that dimension risks being seriously defective.

See also: Equality; Feminist ethics; Justice; Morality and identity §2

JOHN COTTINGHAM

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985, books II, VIII, IX.(Book II sets out the framework for virtue ethics, while books VIII and IX are particularly concerned with its application to friendship and self-love.)

Becker, L. ***et al***. (1991) ‘Symposium on Impartiality and Ethical Theory’, *Ethics* 101 (4).(A collection of papers discussing various aspects of the debate between partialists and impartialists. Contributors include Lawrence Blum, Barbara Herman, Marcia Baron and John Cottingham.)

Godwin, W. (1793) *An Enquiry Concerning Political Justice*, London: G.G. and J. Robinson, 3rd edn, 1798; 3rd edn, ed. I. Kramnick, Harmondsworth: Penguin, 1985, book II, ch. 2.(Introduces the dilemma of whether to rescue a philanthropic archbishop or a chambermaid, who happens to be my mother, from a burning building.)

Hare, R.M. (1981) *Moral Thinking*, Oxford: Oxford University Press, ch. 8. (Discusses the various levels at which moral thinking occurs, contrasting a ground floor level at which partialistic commitments might operate with a

more detached perspective where impartiality is paramount.)
Kagan, S. (1989) *The Limits of Morality*, Oxford: Clarendon Press. (Uncompromising impartialism defended from a consequentialist perspective.)
Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964, ch. 2. (Presents the highly influential theory of the 'categorical imperative' and the principle of respect for persons.)
Mill, J.S. (1861) *Utilitarianism*, ed. M. Warnock, London: Fontana, 1962, 319.(Mill here cites Bentham's dictum 'everybody to count for one, nobody for more than one'.)
Nagel, T. (1991) *Equality and Partiality*, New York: Oxford University Press.(A compelling account of the conflict between the impersonal and the personal perspectives in ethics.)
Rawls, J. (1971) *A Theory of Justice*, Cambridge MA: Harvard University Press, part I.(Outlines the 'veil of ignorance' as a device for ensuring ethical principles are decided from an impartial perspective.)
Singer, P. (1979) *Practical Ethics*, Cambridge: Cambridge University Press; 2nd edn, 1993, chaps 1, 8.(Strict impartiality defended as a fundamental feature of the ethical point of view. For the 'bowl of rice' example, which suggests that if a person is dishing out food to hungry children, then they should make their own children wait in line like all the others, see the 2nd edn, 233.)
Smith, A. (1759) *The Theory of Moral Sentiments*, ed. D.D. Raphael and A.L. Macfie, Oxford: Clarendon Press, 1976, part I, section I, ch. 5, part II, section I, ch. 2. (Identifies the moral perspective with that of the 'impartial spectator'.)
Williams, B. and Smart, J.J.C. (1973) *Utilitarianism, For and Against*, Cambridge: Cambridge University Press, part II.(Debate between a supporter and a critic of Utilitarianism. For Williams' objection that our integrity as persons depends largely on a self-oriented conception of ourselves, see section 5, 108-18.)

Imperative logic

Imperatives lie at the heart of both practical and moral reasoning, yet they have been overshadowed by propositions and relegated by many philosophers to the status of exclamations. One reason for this is that a sentence's having literal meaning seems to require its having truth-conditions and 'Keep your promises!' appears to lack such conditions, just as 'Ouch!' does. One reductionist attempt to develop a logic of imperatives translates them into declaratives and construes inferential relations among the former in terms of inferential relations among the latter. Since no such reduction seems fully to capture the meaning of imperatives, others have expanded our notion of inference to include not just truth - but also satisfaction - preservation, according to which an imperative is satisfied just in case what it enjoins is brought about.

A logic capturing what is distinctive about imperatives may shed light on the question whether an 'ought' is derivable from an 'is'; and may elucidate the claim that morality is, or comprises, a system of hypothetical imperatives. Furthermore, instructions, which are often formulated as imperatives ('Take two tablets on an empty stomach!'), are crucial to the construction of plans of action. A proper understanding of imperatives and their inferential properties may thus also illuminate practical reasoning.

1 Jörgensen's dilemma and Ross's 'paradox'

Any attempt to codify the logical properties of imperatives must grapple with the following dilemma, first formulated by Jörgensen (1938). On the one hand there appear to be cogent inferences involving imperatives, as in:

Keep your promises!
This is a promise of yours.
Keep this promise!

On the other hand, the only widely agreed-upon conception of good inference is in terms of truth-preservation: a set of sentences Γ implies a sentence ϕ just in case there is no possible situation in which all members of Γ are true and ϕ is false. Yet evidently this conception of cogency in terms of truth-preservation cannot be used to assess the validity of arguments utilizing imperatives, for imperatives are not, *prima facie*, either true or false. Hence either we deny, implausibly, that there are good inferences involving imperatives; or we grant it and then face the question, to what conception of inference are we appealing?

This challenge is to be distinguished from another, less urgent concern first formulated by Ross (1944) and echoed by Williams (1963). According to the classical logic of truth functions, any proposition *P* implies $P \vee Q$, for any *Q*. Similarly, it would seem that even the most conservative imperative logic would countenance the inference

Post this letter!
Post this letter or burn it!

Demanding that the auditor either post the letter or burn it may be seen as granting permission to do either. But then if anyone committed to the premise is thereby committed to the conclusion above, it may seem that demanding that the auditor post the letter is implicitly granting permission to burn it. This reasoning vacillates on the concept of commitment. We may distinguish between commitment to the content of a sentence and commitment to performing the speech act of uttering that sentence with its characteristic illocutionary force. One who utters the premise of the above inference is not thereby committed to performing the speech act of demanding that the auditor either post the letter or burn it, just as one who asserts *P* is not thereby committed to asserting $P \vee Q$. (Asserting *P* merely commits one to the truth of $P \vee Q$.) Yet it is only the performance of the speech act of uttering an imperative that grants permission to the auditor to satisfy that sentence by satisfying any disjunct that it contains: commitment merely to the content of that sentence does not suffice to grant permission of any kind.

To others a logic of imperatives may seem trivially simple. It is common to distinguish between two elements in an imperative sentence, one of which is shared with indicative sentences and the other not: from both 'Shut the door!' and 'The door is shut' we can abstract a reference to the state of affairs of the door's being shut. The former

sentence enjoins that state of affairs while the latter asserts that it obtains. Following Stenius (1967), let us say that what is common to the two sentences is the ‘sentence-radical’, while what distinguishes them is their ‘modal element’. One might urge that the validity of arguments involving imperatives should be understood just in terms of sentence-radicals, since these require no modification in our ‘truth-preservation’ conception of validity. But this leaves us with no account of when and where modal elements can occur in valid arguments. It is unclear, for instance, how without *ad hoc* postulation we are to avoid countenancing:

Shut the door!
The door is shut.

Approaches to the logic of imperatives may be seen as grappling with one or other horn of Jörgensen’s dilemma.

2 Reductionist approaches

Reductionist approaches attempt to show that there is no good inference that crucially involves imperatives; at best, any inference that seems to involve imperatives can be replaced with an equivalent one comprising only indicatives, to which the truth-preservation conception of validity applies. To the objection that only propositions appear to be truth-bearers, reductionists might reply that the difference in grammatical mood between imperatives and declaratives is an idiosyncrasy of idiom; a mere reflection of the fact that these two types of sentence are typically used to perform different kinds of speech acts. The difference of mood need be of no more interest to the logician than is the difference between the active and passive voice in declarative propositions.

So with what propositions are imperatives to be identified? On one proposal, ‘Shut the door!’ is equivalent to ‘You will shut the door’. This equation is perhaps suggested by the fact that the latter type of sentence is often used to convey a command; if correct then imperative inference may be understood in terms of inferential relations among statements in the future tense (see Tense and temporal logic). Yet this identification falls foul of our ability to make predictions without formulating any kind of directive: a parole officer might say to a repeat offender, ‘You will steal again’, without enjoining him to crime. Others have contended that ‘Shut the door!’ is equivalent to ‘I want you to shut the door’; here imperative inference is a direct consequence of the logical properties of attitude locutions (see Propositional attitude statements §§1-2). But one can, and may have to, command something one does not want. Furthermore, recipes and instructions are often formulated with imperatives but appear not to be reports of anybody’s desires.

It has also been proposed that ‘Shut the door!’ is equivalent to ‘Either you are going to shut the door, or *X* will happen’, where *X* is something bad for the person being addressed. Yet a speaker might reasonably utter an imperative such as ‘Keep your promises!’ knowing full well that no harm will come the auditor if they fail to obey. Second, the theory implies that for any action *Y*, either ‘Do *Y*!’ or ‘Don’t do *Y*!’ is true. A distinct, ‘performative’ analysis of the imperative has it that ‘Shut the door!’ is equivalent to ‘I order you to shut the door’ (see Performatives). This proposal requires an imperative to be as multiply ambiguous as there are distinct illocutionary acts that can be performed with that sentence; on a given occasion of utterance the same sentence might mean, ‘I request that you shut the door’. This theory is thus committed to a multiplication of senses for what seem to be univocal sentences.

The most popular reduction may be traced to Kant (1785) and equates ‘Shut the door!’ with the ‘deontic’ proposition ‘You should (or ought to) shut the door’. On this approach imperative inference is a matter of inferential relations among deontic locutions (see Deontic logic). We may distinguish between two species of imperative - fiats and directives - only the latter of which are directed to a particular person or group of persons. Clearly, directives are the much more common form and for these it would seem that being directed to a particular individual is part of their content. This is perhaps why imperatives are so naturally issued with vocatives, as in ‘Mary, make sure John does his homework’. In contrast to an imperative, the truth of a deontic proposition does not, as such, lay obligations upon any one person rather than another. ‘It is obligatory that *A*’ does not discriminate among the various people who might be obliged to make it the case that *A*. This even holds for ‘John should see to it that *A*’: a speaker might want to lay responsibility on Mary for making John see to it that *A*, but will not be able to do so with just deontic propositions. A deontic reduction lacks the resources to capture the directedness of imperatives.

3 Non-reductionist approaches

Grappling with the other horn of Jörgensen's dilemma involves formulating a notion of inference to include things other than truth-bearers. Since validity as here characterized is a semantic notion, an account of valid reasoning with imperatives would require assigning semantic values to imperative sentences. One proposal stems from Lemmon (1965), according to whom the basic semantic value of an imperative is to be that of its being 'in force' (or not). Thus when a commanding officer says, 'Clean the latrine!' to a soldier, the command is typically in force, whereas if the soldier were to say the same to the commanding officer this would typically not be the case. This suggests the following criterion of validity:

An argument is valid if and only if, given that all the directive premises are in force and all the declarative premises are true, then the conclusion is in force if it is a directive and true if it is a declarative.

This definition of validity sheds no light on the question whether, for instance, the being in force of 'Take all the boxes!' implies the being in force of 'Take this thing!' when the speaker is unaware that this thing is one of the boxes. Lemmon's proposal is thus a mere schema for a definition. A more explicit proposal flows from the work of Hofstadter and McKinsey (1939), Rescher (1966) and Sosa (1967). Say that an imperative is 'satisfied' just in case its sentence-radical (see above) is true. One might suggest that an argument is valid if and only if, whenever its imperative premises are satisfied and its indicative premises are true, then its conclusion is satisfied if imperative and true if indicative. This definition would explain the apparent validity of the inference with which §1 of this entry began. On the other hand, it countenances the inference from 'Shut the door!' to 'The door is shut'.

Sosa attempts a definition in terms of satisfaction-preservation for arguments consisting only of imperatives, some of which may be hypothetical in form. Let us say that imperatives such as 'If it rains, close the window!' are 'satisfied' if it rains and the window is closed; 'violated' if it rains and the window is not closed; and 'neutral' if it does not rain. (All imperatives can be construed as hypothetical if we imagine $P \vee \sim P$ as antecedent when none is otherwise specified.) We now say:

A directive argument is valid just in case (1) if its premises are satisfied then the conclusion is satisfied and (2) if the conclusion is violated then at least one of its premises is violated.

This definition can account for the validity of

If you read this book, come to see me!
Read this book!
Come to see me!

On the other hand, the definition renders invalid the apparently correct

If it rains, close the window!
If it rains and thunders, close the window!

A cognate definition given by Rescher founders on its inability to treat the relation 'is validly inferable from' as being transitive.

There is today no accepted account of imperative inference. Instead research focuses upon the semantics of imperatives while stopping short of formulating a calculus. Yet if these semantic questions can be answered it may be hoped that new proposals for understanding imperative inference may grow in the light the answers shed. Until that time reductionist approaches will continue to be attractive.

See also: Emotive meaning; Logical and mathematical terms, glossary of; Pragmatics; Questions; Speech acts

MITCHELL GREEN

References and further reading

Austin, J.L. (1962) *How to Do Things with Words*, Oxford: Clarendon Press.(A classic work in the theory of speech acts. Defends a version of the 'I order you to…' species of reduction.)

Beardsley, E. (1944) 'Imperative Sentences in Relation to Indicatives', *Philosophical Review* 53: 175-85.(Early proponent of non-reductionism.)

Belnap, N. (1989) 'Declaratives Are Not Enough', *Philosophical Studies* 59: 1-30. (Attacks the sentence-radical/modal element distinction as failing to capture the distinctive content of imperatives.)

Bohnert, H. (1945) 'On the Semiotic Status of Commands', *Philosophy of Science* 12: 302-15.(Argues for the reduction of imperatives to sentences of the form 'Either you conform or something bad will happen'.)

Castañeda, H.-N. (1975) *Thinking and Doing: The Philosophical Foundations of Institutions*, Dordrecht: Reidel. (Wide-ranging study of imperatives, particularly their semantics and their role in practical reasoning.)

Chellas, B. (1969) 'Imperatives', *Theoria* 37: 114-29.(Formulates a logic for imperatives, treating them as reducible to deontic propositions.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 2nd edn, 1960.(Historically important formulation of a distinction between illocutionary force and semantic content.)

Hamblin, C.L. (1987) *Imperatives*, New York: Blackwell.(Best full-length treatment to date of the grammar, semantics and pragmatics of imperatives. Includes an extensive bibliography.)

Hare, R.M. (1952) *The Language of Morals*, Oxford: Clarendon Press.(Propounds the centrality of imperatives in ethical discourse. Distinguishes between phrastic and neustic.)

Hofstadter, A. and McKinsey, J.C.C. (1939) 'On the Logic of Imperatives', *Philosophy of Science* 6: 446-57. (Mentioned in §3. Early formulation of the notion of satisfaction as a distinctive semantic value for imperatives.)

Huntley, M. (1984) 'The Semantics of English Imperatives', *Linguistics and Philosophy* 7: 103-33.(Sophisticated defence of a semantic treatment of imperatives as differing from propositions in lacking an indexical element.)

Jörgensen, J. (1938) 'Imperatives and Logic', *Erkenntnis* 7: 288-98.(Mentioned in §1. Explains the central challenge facing attempts to formulate an imperative logic.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. H.J. Paton, *Groundwork of the Metaphysics of Morals*, London: Harper Perennial, 1989. (Mentioned in §2. Defends a reduction of imperatives to deontic propositions.)

Lemmon, E.J. (1965) 'Deontic Logic and the Logic of Imperatives', *Logique et Analyse* 8: 39-71.(Mentioned in §3. Sketches a conception of imperative inference in terms of the concept of an imperative's being in force.)

Lewis, D.K. (1970) 'General Semantics', repr. in *Philosophical Papers*, vol. 1, Oxford: Oxford University Press, 1983.(A classic paper in the philosophy of language. Defends a version of the 'I order you to…' species of reduction.)

Rescher, N. (1966) *The Logic of Commands*, London: Routledge & Kegan Paul. (Mentioned in §3. Highly readable general treatment of imperative logic as well as an excellent introduction to the topic. Defends a notion of imperative inference in terms of satisfaction-preservation.)

Ross, A. (1944) 'Imperatives and Logic', *Philosophy of Science* 11: 30-46. (Mentioned in §1. Argues for the impossibility of a non-trivial imperative logic.)

Searle, J.R. and Vanderveken, D. (1985) *Foundations of Illocutionary Logic*, Cambridge: Cambridge University Press.(Develops some logical features of devices used to indicate illocutionary force and inferential relations among various types of speech act.)

Sosa, E. (1967) 'The Semantics of Imperatives', *American Philosophical Quarterly* 4: 57-64.(Mentioned in §3. Defends a notion of imperative inference in terms of satisfaction-preservation.)

Stenius, E. (1967) 'Mood and Language-Game', *Synthese* 17: 254-74.(Mentioned in §1. Defends a distinction between the sentence-radical and modal element.)

Williams, B.A.O. (1963) 'Imperative Inference', *Analysis,* supplement 23: 30-6. (Mentioned in §1. Following Ross, argues for the impossibility of an imperative logic.)

Implicature

A term used in philosophy, logic and linguistics (especially pragmatics) to denote the act of meaning or implying something by saying something else. A girl who says 'I have to study' in response to 'Can you go to the movies?' has implicated *(the technical verb for making an implicature) that she cannot go. Implicatures may depend on the conversational context, as in this example, or on conventions, as when a speaker says 'He was clever but poor', thereby implying - thanks to the conventional usage of the word 'but' - that poverty is unexpected given intelligence. Implicature gained importance through the work of H.P. Grice. Grice proposed that conversational implicatures depend on a general principle of rational cooperation stating that people normally try to further the accepted purpose of the conversation by conveying what is true, informative, relevant and perspicuous. The extent and nature of the dependence, and the precise maxims involved, are matters of controversy. Other issues include whether certain implications are implicatures rather than presuppositions or parts of the senses (literal meanings) of the words used.*

1 Speaker meaning and implication

H.P. Grice (1989: chaps 5, 6 and 14) drew an important distinction between what a speaker means or implies, and what a sentence or other expression means or implies. The sentence 'The aeroplane is a mile long' means that the aeroplane is 5,280 feet long, and implies that it is over 3,279 feet long. Nevertheless, it would be most unlikely for a speaker uttering the sentence to mean or imply such things unless he were joking. More likely, the speaker would be exaggerating, meaning simply that the plane is enormous compared with typical planes. A speaker using the sentence in a coded message might mean something completely unrelated to its English meaning.

As a first approximation, word meaning may be characterized as *conventional* speaker meaning. Whereas we could conceivably use 'bachelor' to mean just about anything, we conventionally mean 'unmarried male', which is what the word means. Speaker meaning is determined by the intentions of the individual speaker (see Communication and intention). While the exact intentions required are a matter of controversy, it is plausible that a speaker *S* means that the plane is enormous only if *S* used the sentence to express the belief that the plane is enormous. Typically, *S* would intend to produce that belief in the audience. Speaker implication is *indirect* speaker meaning: meaning one thing (for example, that one cannot go out) by meaning another (that one has to study).

2 Speaker implicature

Grice was the first to systematically study cases in which speaker meaning differs from sentence meaning. He introduced the verb 'implicate' and the cognate noun 'implicature' as technical terms denoting 'the act of meaning or implying something by saying something else' (Grice 1989: chaps 2 and 3). Consider the following dialogue:

(1) Ann: Where can I get petrol?
Bob: There's a petrol station around the corner.

We may suppose Bob to have implied that Ann can get petrol at the petrol station. Nevertheless, Bob did not actually say that Ann can get petrol there. So Bob has 'implicated' it. What Bob said, and therefore did not implicate, is just that there is a petrol station around the corner. By 'saying', Grice meant not mere uttering, but saying *that* something is the case. (As Grice realized, 'say' can be used more or less narrowly. The less speakers are counted as saying, the more they are counted as implicating.)

Example (1) shows that saying is more closely related to conventional meaning than to speaker meaning. The implication that Ann can get petrol at the petrol station is no part of the meaning of 'There's a petrol station around the corner'. If '*e*' means that *p* on a given occasion, then the speaker says that *p*. The converse fails for several reasons. First, a speaker who utters 'Jack and Jill went up the hill' has said that Jack went up the hill; but that is only part of what the sentence means. Second, a speaker who utters 'He is liberal' might have said that President Clinton is liberal; but (being indexical) 'He is liberal' does not mean 'President Clinton is liberal', even on that occasion (see Demonstratives and indexicals). In general, though, an implicature may be described as something the speaker means or implies that is not part of what the sentence literally means.

If '*e*' means that *p*, or if *S* says that *p* by uttering '*e*', then '*p*' is a truth-condition of '*e*', either in general, or as

used on that occasion. Truth-conditions are logical or a priori implications. Assuming particular interpretations, ‘*p*’ is a truth-condition of sentence ‘*e*’ if and only if it is absolutely impossible for ‘*e*’ to be true without ‘*p*’ being true. ‘There is a petrol station’ is a truth-condition of ‘There’s a petrol station around the corner’. For the latter cannot be true unless the former is true. But ‘Ann can get petrol at the petrol station’ is not a truth-condition of ‘There’s a petrol station around the corner’. The petrol station might be out of petrol or closed. It follows that an implicature need not be a truth-condition of the sentence uttered. The two notions are not mutually exclusive, however, as the following dialogue between an auditor and tax payer illustrates.

(2) Alan: Is it true that you or your spouse is 65 or older or blind?
Bill: I am 67.

By saying ‘I am 67’, Bill implicated (he implied but did not say) that he or his spouse is 65 or older or blind. Furthermore, he could not be 67 unless he or his spouse were 65 or older or blind. So *S*’s implicature is also a truth-condition of the sentence *S* uttered on that occasion.

Empirically necessary conditions, let us say, are ‘natural’ implications. ‘The power is on’ is a natural implication of ‘The lights are on’. If ‘*e*’ naturally implies ‘*p*’, then ‘*e*’ means that *p* in Grice’s natural sense. Whether natural implications are implicatures depends on the speaker’s intentions. In dialogue (3), we would expect Beth to have implicated that the power is on.

(3) Alice: Is the power on?
Beth: The lights are on.

But if Beth does not know what power is, and gives that answer out of desperation, then she would not have implicated that the power is on.

Speakers who know that an implicature is false have misled their audience. But they have not lied unless the implicature happens to be a truth-condition. If Bob knows there is no petrol at the petrol station, then he has misled Ann in dialogue (1). But Bob has not lied, since he did not actually say that there was petrol there. However, if Bill knows that what he has implicated is false in dialogue (2), then he has to be lying.

3 Utterance and sentence implicature

An utterance implicates whatever the utterer implicates. Thus in (1), Bob’s utterance implicated that Ann can get petrol. A sentence implicates, roughly, what speakers following linguistic conventions would normally use it to implicate. Despite what Bob implicated in (1), the sentence Bob used does not itself implicate ‘Ann can get petrol at the petrol station’. For in most contexts, there would be no impropriety in using ‘There’s a petrol station around the corner’ without that implicature. In contrast, (4a) itself implicates (4b):

(4a) Bill is sick, so he should rest.
(4b) Bill’s being sick implies that he should rest.

For a speaker could not properly use (4a) without implicating (4b). And (5a) implicates (5b):

(5a) Some died.
(5b) Not all died.

For speakers would normally use (5a) with that implicature.

4 Conventional versus conversational implicature

Implicatures generated by the conventional meaning of the words uttered, as in (4), are classified as ‘conventional’ (see, for example, Grice 1989: 25-6). Non-conventional implicatures, as in (5), are termed ‘conversational’, since they depend on the conversational context. Normally, if one uses a sentence aware of the falsity of some of its conventional implicatures or logical implications, one uses the sentence improperly. If one uses a sentence aware of the falsity of a conversational implicature, one’s utterance is at most misleading. Conversational implicatures differ further in being cancellable. The implicature of (5a) can be explicitly cancelled by adding ‘indeed, all did’. ‘some died, indeed all did’ is perfectly consistent, and does not imply ‘Not all died’ in

any way. In contrast, since ‘All died’ logically implies ‘some died’, ‘All died, indeed none did’ is self-contradictory. ‘Bill is sick, so he should rest, but Bill’s being sick in no way implies that he should rest’ is incoherent in a different way. Conventional implicatures can also be implicitly cancelled by the context. Thus (5a) would not have its customary implication if uttered in full view of a garden in which every single plant was obviously dead. Conversely, conversational implicatures are reinforceable. Conjoining a sentence with one of its conversational implicatures is not redundant, and logically implies what the original sentence implicates; witness ‘some died, but not all did’; in contrast, ‘All died, and some did’ is redundant.

5 Conversational maxims and the cooperative principle

Besides identifying the phenomenon of implicature, Grice formulated a theory in terms of which he classified different sorts of conversational implicature, and tried to explain how they arise and are understood. It is common knowledge, he asserted, that people generally follow certain rules for efficient communication. Grice’s rules (1989: 26-9) included one general ‘principle of cooperation’, and four ‘maxims’ specifying how to be cooperative:

The cooperative principle. Contribute what is required by the accepted purpose of the conversation.
The maxim of quality. Make your contribution true, so do not convey what you believe false or unjustified.
The maxim of quantity. Be as informative as required, neither more nor less so.
The maxim of relevance. Be relevant.
The maxim of manner. Be perspicuous, so avoid obscurity and ambiguity, and be brief and orderly.

Generalizations of these rules govern rational, cooperative behaviour in general. If I am helping a man change the oil, I will hand him a can of oil rather than a jack (relevance), a barrel of oil (quantity), or fake oil (quality), and I will not take all day doing it (manner). The maxims do not, therefore, have the characteristic arbitrariness of linguistic conventions. Much of the literature is devoted to clarifying and strengthening the maxims, relating them, and extending the list. But Grice’s formulations are still dominant.

To sketch how the principles are thought to explain the way implicatures arise and are understood, consider (1). Bob would have infringed the maxim of relevance unless he believed Ann could get petrol at the petrol station. Since Bob was trying to cooperate, his utterance indicated that belief. Thus Bob could imply that Ann can get petrol there and expect Ann to recognize that implication. Speakers using (5a) would violate the maxim of quantity if they knew that all died. So by not making the stronger statement, they can imply that not all died, and expect hearers to recognize the implication. Unlike conventional implicatures, conversational implicatures are cancelled when the maxims are in abeyance. Under cross-examination of an uncooperative witness, (5a) would not implicate (5b).

Some implicatures depend on flouting the maxims, Grice proposed (1989: 30-1). This occurs when what a cooperative speaker says so obviously fails to obey the maxims that the hearer must assume the speaker to be meaning something different. Irony and metaphor appear to depend on flouting the maxim of quality. Hearing ‘Fine day, isn’t it?’ in the middle of a blizzard, we would recognize that the speaker cannot really believe it to be a fine day, and means the opposite. Hearing ‘The music danced lightly on his ears’, we would recognize that the speaker believes no such thing, and is using metaphor. The speaker who says ‘The singer produced a series of sounds closely corresponding to the score of *Oklahoma*’, thereby meaning that the singing was terrible, is flouting the maxim of manner.

Many of Grice’s theoretical claims about the exact role played by the maxims in implicature are implausible. For example, Grice suggested that *S* conversationally implicates ‘*p*’ only if: (i) *S* is presumed to be observing the cooperative principle (‘cooperative presumption’); (ii) the supposition that *S* believes ‘*p*’ is *required* to make *S*’s utterance consistent with the cooperative principle (‘determinacy’); and (iii) *S* believes or knows, and expects the hearer *H* to believe, that *S* believes *H* is able to determine that (ii) is true (‘mutual knowledge’). The cooperative presumption would fail if, as is possible, Beth implied what she did in case (3) in defiance of Alice’s mistrust, or if Beth implicated that she cannot answer the question by saying ‘I’m late for an appointment’.

Determinacy fails outright in case (1), as presumably does mutual knowledge: the assumption that Bob believed Ann could at least learn where to get petrol would also reconcile his utterance with the maxims, as would the more fantastic assumption that Bob believed the petrol station was giving away new diesel Volvos. Indeterminacy is of

legendary proportions in figures of speech. Mutual knowledge may even fail independently. In (3), for example, Beth may have contemptuously implied that the power is on, believing that Alice would not be smart enough to figure it out. Some authors have nevertheless strengthened the mutual knowledge condition to say that *S* knows that *H* knows that *S* knows that *H* knows *ad infinitum*, a requirement both psychologically unrealistic and theoretically unnecessary.

While the maxims undeniably provide insight into conversational implicature, their predictive and explanatory power is extremely limited. Consider the account of irony. Grice suggests that *S* implicates 'It is a lousy day' when saying 'Fine day, isn't it?' because otherwise *S* would be violating the maxim of quality and thereby the cooperative principle. But *S* is still violating that principle, by violating the maxim of manner: figures of speech are seldom the most perspicuous ways of communicating information. Consider quantity implicatures. If (5a) implicates (5b) because of the maxim of quantity, then why does (5a) not implicate 'It is not the case that two per cent died'? Why does 'The repairs will take some time' not implicate 'They will not take a long time'? Why does 'Ann said Bill was responsible' not implicate 'Ann did not insist (or predict, or argue, or shout, or…) that Bill was responsible'? Consider finally the sequential interpretation of conjunctions. Grice explained why a conjunction like (6a) implies (6b) in terms of the submaxim 'Be orderly'.

(6a) Bill fell ill and saw the doctor.
(6b) Bill fell ill and therefore saw the doctor.
(6c) Bill saw the doctor and therefore fell ill.
(6d) It is more likely that Bill fell ill than that he saw the doctor.
(6e) Bill did not see the doctor because he fell ill.

But reverse temporal or causal order is just as orderly as forward order (witness 'Bill died; he had cancer'). So why should (6a) not imply (6c)? It would also be quite orderly to express the conjuncts from most to least probable. So why is (6d) not implied? And since '*B* because *A*' is stronger than '*A* and *B*', why is (6e) not implied in virtue of the quantity maxim? (For a comprehensive critique of Gricean theory, see Davis 1998.)

6 Theoretical importance

Difficulties in Gricean theory do not cast doubt on the existence or theoretical importance of implicature. Speakers cannot be fully understood without knowing what they implicate. It does not suffice to know the truth-conditions (or in some cases even the meaning) of all the sentences uttered, or what is said. Implicature must be considered even in truth-conditional semantics. Here are two examples.

Ambiguity. Readily available evidence suggests that the word 'and' is ambiguous in English, having a strong sense (literal meaning) connoting temporal or causal order in addition to the weak sense connoting joint truth alone. But before this can be accepted, the competing hypothesis that the sequential connotation is merely an implicature must be rejected (see Ambiguity §3). Many facts support implicature over ambiguity. For example, 'Bill fell ill and saw a doctor, but not necessarily in that order' does not have a contradictory reading. Furthermore, 'It is not the case that Bill fell ill and saw a doctor' would be interpreted as false when 'Bill fell ill' and 'Bill saw a doctor' are both true, no matter what the sequence of events.

Grice (1989: 47-50) proposed a version of 'Ockham's razor' according to which 'senses are not to be multiplied beyond necessity'. But the same can be said for implicatures. It is often claimed that since implicatures can be accounted for in terms of general psycho-social principles of conversation, the postulation of senses rather than implicatures results in a more complex overall theory. But this overstates the explanatory powers of any known principles of conversation.

Presupposition. '*e*' is said to (semantically) 'presuppose' '*p*' provided the truth of '*p*' is necessary for '*e*' to be either true or false (see Presupposition). Following P.F. Strawson, and opposing Bertrand Russell, many have argued that a sentence like (7a) presupposes (7b):

(7a) Your crimes are inexcusable.
(7b) You have committed crimes.
(7c) Your crimes are not inexcusable.

The presupposition hypothesis is plausible since (7c) seems to imply (7b) just as strongly as (7a) does. Accepting presuppositions seriously complicates logical theory, however. For example, if some declarative sentences are neither true nor false, then standard formulations of the 'law of excluded middle' must be modified. One strategy for avoiding these complications hypothesizes that the implication of the negation is merely a conventional implicature (see, for example, Karttunen and Peters 1979). On this view, (7c) conventionally implicates (7b) without logically implying it. The fact that a sentence may be true when an implicature is false allows the Russellian to account for the felt implications, while insisting that (7c) is true and (7a) false when (7b) is false. Supporting Strawson, though, is the strong intuition that 'Are your crimes excusable or not?' is a loaded question, which cannot be answered if you are innocent. An outstanding problem for either approach is to describe how implications of complex propositions (whether presuppositions or implicatures) are related to those of their components.

Implicature has also been invoked in accounts of lexical gaps, language change, indirect speech acts, textual coherence, discourse analysis and even syntax.

See also: Meaning and communication; Pragmatics; Semantics; Speech acts

WAYNE A. DAVIS

References and further reading

Blakemore, D. (1992) *Understanding Utterances*, Oxford: Blackwell.(A readable introduction to the theory of implicature developed in Sperber and Wilson (1986).)

Davis, W.A. (1998) *Implicature: The Failure of Gricean Theory*, Cambridge: Cambridge University Press.(A comprehensive critique of Grice's theory of implicature, and of similar theories. Argues that conversational implicatures arise not from conversational principles, but from intentions and conventions.)

Gazdar, G. (1979) *Pragmatics: Implicature, Presupposition, and Logical Form*, New York: Academic Press. (Attempts a partial formalization of conversational implicatures, focusing on quantity implicatures, and uses it to provide a pragmatic theory of presupposition.)

Grice, H.P. (1989) *Studies in the Ways of Words*, Cambridge, MA: Harvard University Press.(Contains all of Grice's work on meaning and implicature, plus an introduction and retrospective epilogue.)

Harnish, R.M. (1976) 'Logical Form and Implicature', repr. in S. Davis (ed.) *Pragmatics: A Reader*, Oxford: Oxford University Press, 1991, 316-64.(Presents many problems for Gricean theory, suggesting solutions and extensions.)

Horn, L.R. (1989) *A Natural History of Negation*, Chicago, IL: University of Chicago Press.(A remarkably comprehensive and detailed account of negation, focusing *inter alia* on its effect on implicature, and explanation of a wide variety of linguistic facts about negation in terms of implicature.)

Karttunen, L. and Peters, S. (1979) 'Conversational Implicature', in C.-K. Oh and D.A. Dinneen, *Syntax and Semantics*, vol. 11, *Presupposition*, New York: Academic Press, 1979, 1-56.(A detailed, compelling argument that different sorts of presupposition can be accounted for as different types of implicature, plus a representation of conventional implicature using model-theoretic semantics, specifically, Montague grammar.)

Leech, G. (1983) *Principles of Pragmatics*, London: Longman.(A leading introduction to pragmatics stressing implicature. Develops Grice's suggestion that another maxim is 'Be polite', showing that it motivates many implicatures.)

Levinson, S.C. (1983) *Pragmatics*, Cambridge: Cambridge University Press.(A leading introduction to pragmatics, with an extensive discussion of conversational implicature and presupposition from a Gricean point of view.)

Martinich, A.P. (1984) 'A Theory for Metaphor', *Journal of Literary Semantics* 13: 35-56; repr. in S. Davis (ed.) *Pragmatics: A Reader*, Oxford: Oxford University Press, 507-18.(An extended Gricean treatment of metaphor.)

Sadock, J.M. (1978) 'On Testing for Conversational Implicature', in S. Davis (ed.) *Pragmatics: A Reader*, Oxford: Oxford University Press.(Points out problems with all of Grice's tests for conversational implicature.)

Sperber, D. and Wilson, D. (1986) *Relevance: Communication and Cognition*, Oxford: Blackwell.(Develops a theory of implicature in terms of a single 'principle of relevance' quite distinct from Grice's 'maxim of relevance'.)

Incarnation and Christology

It is a central and essential dogma of Christianity that Jesus of Nazareth, who was crucified in Judea during the procuratorship (AD 26-36) of Pontius Pilate, and God, the eternal and omnipresent creator of the universe, were in some very strong sense 'one'. The department of Christian theology that is devoted to the study of the nature and implications of this 'oneness' is called Christology. Orthodox Christology (unlike certain heretical Christologies) sees this oneness as a oneness of person, as consisting in the co-presence of two natures, the divine and the human, in one person, Jesus Christ. To speak plainly, orthodox Christology holds that there is someone, Jesus Christ, who is both divine and human. Because God pre-existed and is superior to every human being, orthodox theologians have found it natural to speak of the union of the divine and human natures in the person of Jesus Christ as something that happened to *the pre-existent divine nature: at a certain point in time, at the moment of the conception of Jesus, it 'took on flesh' or 'became incarnate'; in the words of the Athanasian Creed, the union of the two natures was accomplished 'not by conversion of the Godhead [*divinitas*] into flesh, but by taking of the manhood into God'. This event, and the continuing union it established, are called 'the Incarnation'. The Incarnation was not, according to Christian teaching, undone by Christ's death (his corpse - a* human *corpse - continued to be united with the divine nature by the same bond by which the living man had been united) or by his 'Ascension' (his 'withdrawal' from the everyday world of space and time forty days after the Resurrection), and it will never be undone: the Incarnation is eternal.*

The primary statements of the dogma of the Incarnation are the Definition issued by the Council of Chalcedon (AD 451) and the Athanasian Creed (fifth century; its origins are obscure). The creed issued by the Council of Nicaea (AD 325) and the longer, revised version of this creed that is today used liturgically (and commonly called 'the Nicene Creed') contain nothing of substance that is not found in the two later statements.

1 The doctrine of the Incarnation

Books present portraits, generally inadvertent, of their authors and their intended audiences. The New Testament presents a portrait of Christians who were not quite sure what to say about the relation between Jesus Christ and God. On the one hand, the earliest Christians could hardly deny that Christ was a man (had he not frequently referred to himself as 'the Son of Man'?), a human being who had been born and had died (albeit he was not dead for very long), who ate and drank and spoke and slept and left prints in the dust of Palestine. On the other hand, they could hardly speak of Christ without mentioning God in the same breath, they called him 'the Son of God', and they were unreflectively willing to ascribe to him honours held traditionally in their cultures to be due to God alone. Explicit statements about the relation between Christ and God are rare in the New Testament. There are, however, a few passages in which this relation is described, and all of these imply, or come very close to implying, a 'high' Christology - a Christology that in some sense identifies Christ with God. Thus (the translations are those of the New American Standard Bible):

> In the beginning was the Word [*logos*], and the Word was with God, and the Word was God. He was in the beginning with God. All things came into being by him, and apart from him nothing came into being that has come into being…. And the Word became flesh and dwelt among us, and we beheld his glory…
>
> (John 1: 1-3, 14)

> He [Christ] is the image of the invisible God, the first-born of all creation [that is, has an authority over the created world comparable to that which, under Jewish law, a first-born son had over his living father's estate]. For by him all things were created, both in the heavens and on earth, visible and invisible,…- all things have been created by him and for him. And he is before all things and in him all things hold together [*sunistemi*]…. For in him all the fullness of Deity dwells in bodily form.
>
> (Colossians 1: 15-17; 2: 9)

> In these last days, [God] has spoken to us in his Son, whom he appointed heir of all things, through whom also he made the world [literally: 'made the ages']. And he is the radiance of his [God's] glory, and the exact representation of his nature, and upholds all things by the word of his power.
>
> (Hebrews 1: 2, 3)

It should also be noted that in John 8: 58, and possibly in John 10: 22-38, the author represents Jesus as affirming his own deity, and that in John 20: 28, the apostle Thomas addresses the risen Jesus as ‘My Lord and my God’.

The New Testament, although it is rich in Christological suggestion, contains no systematic Christology. The development of a systematic Christology was the work of the first five Christian centuries. The relevant biblical passages - those quoted above, and a few others, such as Philippians 2: 5-11 - have been treated by most theologians not as explicit statements of doctrine but as data to which explicit statements of doctrine must be responsible. The ‘developed’ doctrine, the doctrine of the fifth century, is contained in the following two passages.

> …we all unanimously teach that we should confess that our Lord Jesus Christ is one and the same Son, the same perfect in Godhead and the same perfect in manhood, truly God and truly man, the same of a rational soul and body, consubstantial with the Father in Godhead, and the same consubstantial with us in manhood, like us in all things except sin; begotten from the Father before the ages as regards his Godhead, and in the last days, the same, because of us and because of our salvation begotten from the Virgin Mary, the *Theotokos* [God-bearer], as regards his manhood; one and the same Christ, Son, Lord, only-begotten, made known in two natures without confusion, without change, without division, without separation, the difference of the natures being by no means removed because of the union, but the property of each nature being preserved and coalescing in one *prosopon* [person] and one *hypostasis* [subsistence], not parted or divided into two *prosopa*, but one and the same Son, only-begotten, divine Word, the Lord Jesus Christ, as the prophets of old and Jesus Christ himself have taught us about him and the creed of our fathers has handed down.
>
> (Definition of Chalcedon; see Kelly 1960)

> …[N]ow the right faith is that we should believe and confess that our Lord Jesus Christ, the Son of God, is equally both God and man.
>
> He is God from the Father’s substance, begotten before time; and he is man from his mother’s substance, born in time. Perfect God, perfect man composed of a rational soul and human flesh, equal to the Father in respect of his divinity, less than the Father in respect of his humanity.
>
> Who, although he is God and man, is nevertheless not two but one Christ. He is one, however, not by the transformation of his divinity into flesh, but by the taking up of his humanity into God; one certainly not by confusion of substance but by oneness of person. For just as rational soul and flesh are a single man, so God and man are a single Christ.
>
> (Athanasian Creed; see Kelly 1964)

These passages seem to say, and rather insistently, that there is a single *prosopon* or *hypostasis* or *persona*, Jesus Christ, who both has a beginning in time and has no beginning in time. More generally, he has all of the properties - attributes, features or characteristics - appropriate to a (sinless) human being, and at the same time has all of the properties that the Christian faith ascribes to God.

In order to grasp the orthodox doctrine expressed in these statements, it is advisable to have some conception of unorthodoxy, of the great Christological heresies of the first five Christian centuries, for these heresies were present in the minds of the framers of the statements and are to a large extent responsible for the details of the wording. The most important heresies for this purpose are the following three.

Nestorianism denies that the two natures belong to the same person: according to Nestorianism - if not to the eponymous Nestorius (d. *circa* 451), who may not actually have held this - the divine Christ and the human Christ are numerically distinct persons. That is, however intimate the union - or ‘conjunction’ (*sunápheia*), the word preferred by the Nestorians - of the human and the divine Christs, there is no one who can say truly both ‘I am God’ and ‘I am a human being’. According to Nestorians, the ‘conjunction’ of the divine and the human natures was effected between God and a pre-existent human being. Since God and Jesus each existed before the conjunction, and each was then who he was, there could not have been only one person ‘present’ after the conjunction unless the human person had somehow ceased to be - which is obviously unacceptable. (Orthodox Christology holds that the human being Jesus of Nazareth did not exist before the Incarnation, and that, in modern terminology, there is no possible world in which he exists, even for the briefest instant, without God’s being incarnate in him.)

Monophysitism ('one-naturism') holds that there is only one nature, the divine nature, in the incarnate Christ. The human attributes of Jesus of Nazareth are somehow taken up into or made to be contained within the divine nature; at any rate, they do not constitute a distinct, subsistent human nature - that is, a human being. (It should be noted that early Christological writings do not always make it clear whether 'a nature' (*physis*, *natura*) is a 'predicable', like the *attributes* divinity and humanity, or a 'first substance', an impredicable subject of predication, like God and Jesus of Nazareth. Some authors write as if the distinction did not exist or was of no importance; others seem perversely to take their opponents to mean one of these things when they should pretty clearly be taken to mean the other.)

Apollinarianism (after Apollinarius (*c*.310-*c*.390)) holds that Christ did not have a human mind or spirit or rational soul - that he lacked something that is essential to human nature - and that God or some 'aspect' of God (such as the divine *logos*) was united to the human body of Jesus of Nazareth in such a way as to 'be a substitute for' or perform the function of the human mind or soul or spirit. This is perhaps the most important of the heresies for the task of understanding orthodoxy: it is certainly very frequently suggested by the language of popular Christianity. Orthodoxy insists, however, that whatever is present in our common human nature (other than sin) is present in Christ. The reason for this, briefly, is that the saving work of Christ is to heal our ruined human nature, and (in the words of Gregory of Nazianzus), 'what [Christ] has not assumed he has not healed' (Letter 101, *Patrologia Graeca*, vol. 37). The conviction that 'what he has not assumed he has not healed' is one of the main pressures that guided the development of the orthodox 'Chalcedonian' Christology. It was this conviction that made 'fully human' to be for the 'Chalcedonian' party what in the 1960s was called a 'non-negotiable demand'.

At the same time, other pressures made 'fully divine' non-negotiable for most of the parties, including the Chalcedonian party, to the great Christological controversies. One might cite (1) the conviction that the situation of fallen humanity was so desperate that any effective saviour of humanity must be divine, (2) the fact that from the earliest days of Christianity, Christians had offered honours to Christ that it would be blasphemous to offer to a creature, and (3) the realization that the biblical texts quoted above - which at the very least seem to represent Christ as 'very like' God - could not be interpreted as describing Christ as a being who was close to being God but was not quite God, or even as a being who was ontologically intermediate between God and man: no such 'not quite God', no such intermediate, is possible, because any being who is not God is finite, and any finite being must be infinitely closer ontologically to any other finite being than it is to God.

2 Logical problems

The main philosophical problems facing the doctrine of the Incarnation are logical: the doctrine implies, or seems to imply, that there is an object that has various pairs of incompatible properties - or worse (if worse is possible), that there is an object and a property such that that object both has and does not have that property. (God is eternal; Jesus is not eternal; God is identical with Jesus; hence there is something that both is and is not eternal.) Some theologians have held, apparently, that the doctrine does have these features and is therefore internally inconsistent, but is nevertheless to be believed. Having said this, they proceed to deprecate 'merely human logic'. Their point is not that the doctrine *seems* to be inconsistent owing to the deficiencies of merely human logic; it is rather that it is only because of the deficiencies of merely human logic that inconsistency (at least in theology) seems objectionable (see Morris 1986: 24-5 for a nice selection of quotations on this subject). This position has (to be gentle) little to recommend it. It perhaps rests on a failure to see clearly that a truth cannot be inconsistent with a truth.

It would, of course, be possible to maintain that although the orthodox doctrine of the Incarnation seems to be inconsistent, it is in fact consistent, the explanation of the fact that it is, despite appearances, consistent being beyond human understanding. This position will not be discussed in the present entry, for the simple reason that there is little if anything that can be said about it. Rather, various attempts to solve the problem that faces orthodox Christology owing to the *prima facie* inconsistency of the properties it ascribes to the incarnate Christ will be discussed. Philosophically promising solutions will be examined, but with no attempt to trace their historical roots. These solutions are (as perhaps any solution must be) of one or the other of two types:

(1) The orthodox doctrine does not imply all the statements it appears to imply. One or the other of each of the mutually contradictory pairs of statements the doctrine appears to imply it does not in fact imply. (For example, the doctrine appears to imply that Christ began to exist when Herod was king in Judea, and also appears to imply

that Christ's existence had no beginning. But it in fact implies only the latter. Those who have concluded that the doctrine implies the former statement have reached this conclusion on the basis of a superficial understanding of what is contained in the concept of humanity.)
(2) The doctrine does imply all the statements it appears to imply, but these statements are not, as they appear to be, inconsistent. (For example, the doctrine implies that Christ began to exist when Herod was king in Judea, and also implies that Christ's existence had no beginning. But these two statements are consistent. Those who have concluded that they are inconsistent have reached this conclusion on the basis of a superficial understanding of their logical form.)

We shall call solutions of the former type 'attributive' and solutions of the latter type 'predicative'. These words are no more than convenient labels. Neither corresponds to any school or division in the history of theology.

3 Attributive solutions

Is it not possible that we sometimes read more into certain of the attributes that orthodoxy ascribes to Christ than is really there? We tend to assume that a human being must have a beginning in time, and that this is part of the concept of a human being. But what justifies this assumption? We tend to assume that whatever is a human being is essentially a human being. But, again, what justifies this assumption? Even if Solomon and Catherine the Great and all other 'ordinary' human beings have a temporal origin - even if they all have this property essentially - and even if they are all essentially human, does it follow that there could not be an eternal being who acquired the attribute of humanity at a certain point in time? Is there something about the concept of humanity that makes the idea of such a being a conceptual impossibility? Or is there something about the attribute of humanity that makes the idea of such a being a metaphysical impossibility? We may think so - but how do we know? Are we prepared to assert confidently that an alleged divine revelation that implied the existence of such a being would have to be judged a fraud or fantasy of merely human origin? (No doubt we ought to be prepared to assert this in respect of an alleged divine revelation that implied that the walls of the New Jerusalem would be both square and circular.) Or consider omnipresence. A human being must be locally present somewhere, but does it follow that a human being cannot be omnipresent? We must remember that local presence and omnipresence are two different modes of presence. It is indeed impossible for one and the same being to be both locally present everywhere (like the luminiferous ether) and locally present only at a certain spot on the shores of the Sea of Galilee - but omnipresence is not local presence everywhere (see Omnipresence §4). Again: is it really impossible for a human being to be omnipotent? If an omnipotent being took on a set of properties such that whatever had those properties would be human, it would thereby acquire a certain set of powers or abilities, a set that might be the whole set of the powers had by an 'ordinary' human being; but might this set not be simply a (rather small) subset of the set of all its powers - might that being not continue to be omnipotent? Might it not continue to be able to move mountains? (Only, of course, in the way it had been able to move mountains before it took on its 'new' set of properties and powers: it would be unable to move mountains by using its limbs to exert physical pressure on them, even as Solomon and Catherine were unable to do this.) Similar questions can be asked in respect of omniscience.

If a human being can indeed be omnipresent, omnipotent and omniscient, then one might address the problem of the apparent logical inconsistency of the doctrine of the Incarnation simply by denying it reality and asserting, without qualification, that Jesus of Nazareth was omnipresent, omnipotent and omniscient. Whether this solution is philosophically coherent could be debated interminably. But there is also the question whether it is in fact usable by Christians (who are, of course, the only people who have a use for it), for it is certainly arguable that it is inconsistent with the data of the New Testament, in which Jesus is sometimes represented as unable to do certain things and as learning things of which he had hitherto been ignorant and, in one case, as simply not knowing the exact day and hour of the end of the age (Matthew 24: 36; Mark 13: 32). It would no doubt be possible to insist that the passages in the New Testament that represent Jesus as subject to many of the limitations of ordinary human existence can be reconciled with the thesis that he is omnipotent, omniscient and omnipresent - rather as many have found it possible to insist that the doctrine of a timeless, impassible God can be reconciled with the narratives of the Pentateuch. But it would seem that a solution to the problem would be more attractive to most Christians if it allowed Jesus to share our human limitations.

There is an attributive solution that claims to have this feature. It is called *kenoticism*, from the Greek *kenosis* ('emptying'), an account of the nature of the Incarnation that is based on the statement (Philippians 2: 7) that, in

becoming incarnate, Christ 'emptied' himself. Whatever the correct interpretation of the difficult passage in which this statement occurs may be, kenoticism holds that, in becoming incarnate, Christ relinquished omnipotence, omniscience, omnipresence and various other of what are commonly called the divine attributes (although it continued to be true of him that he was morally perfect, had no beginning in time and was not a created being). Kenoticists do not, however, hold that when Christ had become incarnate *no one* was omnipotent and omnipresent. This entry has not so far touched on the relation of the doctrine of the Incarnation to the doctrine of the Trinity, but the two doctrines intersect on the following point: it was the second person of the Trinity alone, God the Son, who became incarnate. According to the kenotic theory, God the Son 'emptied himself' of omnipotence and omnipresence, but God the Father and God the Holy Spirit continued to possess these attributes.

Kenoticism, it will be noted, requires a 'rethinking' of divinity and various of the divine attributes as well as of humanity. We are inclined to think that a being that had attributes such as omniscience and omnipotence would have to have them essentially - and hence that that being could not relinquish them. But how, the kenoticist asks, do we know this? We are also inclined to think that a being who was not omniscient or omnipotent at *t* would not be at *t* a divine being. But how do we know this? How do we know that, for example, omnipotence is really a divine attribute, really entailed by divinity? Might it not be that it is not strictly omnipotence that is entailed by divinity, but rather some 'weaker' attribute - perhaps 'omnipotence unless omnipotence is voluntarily relinquished'?

It is very doubtful whether kenoticism can be reconciled with orthodoxy. The following two difficulties are apparent. First, the theory does not mesh well with orthodox Trinitarian theology, for it seems to imply that the persons of the Trinity are distinct substances, distinct *beings*, and hence to imply tritheism (see Trinity §§1-2). Second, the theory allows the incarnate Christ to be divine only by considerably 'weakening' the concept of divinity. The incarnate Christ can indeed say truly 'I am a divine being', but only because, for the kenoticist, it is possible to be a divine being at a certain moment without then being omnipotent, omniscient or omnipresent. It is doubtful whether kenoticism can be reconciled with the following two requirements: 'the property [that is, defining properties or essential features] of each nature being preserved' (Definition of Chalcedon); 'not by the transformation of his divinity into flesh' (Athanasian Creed). Even if kenoticism can, by artful interpretations of their language, be reconciled with the letter of the two documents, it certainly cannot be reconciled with their spirit. (Kenoticism is a nineteenth-century invention and did not, therefore, influence the wording of the Definition or the Creed. There can be no real doubt that they would have been so worded as to exclude it if it had existed in the fifth century.)

Thomas Morris (1986) has defended an attributive solution that both takes into account the biblical data concerning the human limitations of Christ and retains a robust concept of divinity. Morris accounts for the biblical data by suggesting that Christ, in becoming incarnate, acquired a human mind without thereby relinquishing his divine mind. (The divine mind that Christ retains is not, according to Morris, 'divine' only in some etiolated, kenotic sense: it remains, for example, omniscient.) Between the two minds there exists an 'asymmetrical accessing relation': only a minuscule segment of what is present in the divine mind is accessible to the human mind. Christ's human limitations are to be traced to the limitations of his human mind, and it is this mind whose thoughts and sufferings are recorded in the Gospels.

Morris' solution is certainly to be preferred both to the 'unqualified' solution and to kenoticism. One might wonder, however, whether it is not a form of monophysitism. (This point could also have been brought against the 'unqualified' solution.) Morris' solution represents Christ's human mind as a 'subsystem' of his divine mind - which is, after all, really just his mind, his mind *simpliciter*. And although Christ has (on the physical side) weight and shape and local presence - all the physical properties that someone needs to be fully human - these are simply properties that God has acquired, 'additions' to the properties that he had before the Incarnation. It would seem, therefore, that Morris' solution inherits the following feature of the 'unqualified' solution: it represents the incarnate Christ as a single substance, a divine substance that, by becoming incarnate, acquired certain properties that would otherwise belong only to created beings. Morris would certainly vigorously affirm the presence of two natures in the incarnate Christ, but his account of what this means seems to be very like the account given by those monophysites who were willing to accept the 'two natures' terminology: 'It is the claim of orthodoxy that Jesus had all of the [essential] properties of humanity, and all the [essential] properties of divinity, and thus existed (and continues still to exist) in two natures' (Morris 1986: 40). The definitional statements that an advocate of Morris'

solution would have to pay special attention to are these: 'made known in two natures without confusion [without running together]' (Definition of Chalcedon); 'not by confusion of substance' (Athanasian Creed).

4 Predicative solutions

Predicative solutions concede that pairs of predicates like 'is eternal' and 'was born of a human mother in 6 BC' are inconsistent. But advocates of these solutions maintain that the real logical form of some or all sentences of the superficial form 'Christ is *F*' is not what that superficial form suggests. There are two, or perhaps three, ways a sentence of this form can be construed: 'Christ is *F qua* God (as regards his divinity)', 'Christ is *F qua* man', and, possibly, 'Christ is *F* without qualification (*simpliciter*)'. Predicative solutions contend, moreover, that pairs of sentences like 'Christ is eternal *qua* God' and 'Christ was born of a human mother in 6 BC *qua* man' are consistent. (It would seem natural to suppose that proponents of predicative solutions - since they concede that 'is eternal' and 'was born…' are inconsistent - must regard 'Christ is eternal *simpliciter*' and 'Christ was born of a human mother in 6 BC *simpliciter*' as inconsistent. Indeed, it would seem natural to suppose that they would have to say that it was false that Christ had either his 'exclusively divine' or his 'exclusively human' attributes *simpliciter*, and would allow only that he had each of them *qua* God or *qua* man, as appropriate - although they might hold that Christ had certain attributes, such as moral perfection, *simpliciter*. But some 'predicativists' may refuse to recognize the '*simpliciter*' mode of predication. And at least one 'predicativist' author explicitly treats such pairs of sentences as consistent, for reasons that will be given below.)

'Predicativism', so understood, is unquestionably orthodox. But it is doubtful whether it constitutes a solution to the problem of the apparent inconsistency of orthodox two-natures Christology. A satisfactory predicative solution must supplement the abstract theses of the preceding paragraph with some sort of reply to the following challenge:

Where *F* and *G* are incompatible properties, and K_1 and K_2 are 'kinds', what does it mean to say of something that it is *F qua* K_1 but *G qua* K_2? - or that it is *F qua* K_1 but is not *F qua* K_2? And can any more or less uncontroversial examples of such pairs of statements be found?

R.T. Herbert (1979) has offered an example of such a pair of statements. Consider ambiguous figures, like the familiar 'duck-rabbit'. One could point at such a figure and say truly both, 'That has ears *qua* rabbit' and 'That has no ears *qua* duck'. But this example immediately suggests a question: what does the word 'that' refer to? Certainly not to the ambiguous figure. The *figure* belongs neither to the kind 'rabbit' nor to the kind 'duck', and therefore, presumably, has no properties *qua* rabbit or *qua* duck. (And if 'that' does not refer to the figure, what does it refer to?) It may be true that the figure has the property 'representing something with ears' *qua* rabbit-representation and lacks this same property *qua* duck-representation. If so, this would be an example of a pair of '*qua*' statements of the forms the above challenge has demanded. But it does not seem that this example will be of use in Christology, for the property the figure has '*qua* something' and lacks '*qua* something else' is a representational property, and Christ, not being a drawing or a statue, has no representational properties. (It is not accidental to the example that the property it involves is representational, and the example provides the inquirer with no clue as to how to construct an example involving a property that is not representational.) It is true that one biblical quotation describes Christ as 'the exact representation [*charaktér*] of [God's] nature' (Hebrews 1: 3). But this means that Christ is a 'perfect copy' of God, and does not imply that any of Christ's properties is a representational property. (A daughter who is the 'spit and image' of her mother is not a representation of her mother in the way a portrait of her mother is.)

Peter van Inwagen (1994) has presented a very comprehensive and general predicative solution to the problem of the apparent inconsistency of the doctrine of the Incarnation. It has been observed by several authors that Peter Geach's thesis of the 'relativity of identity' can be employed to solve the 'Leibniz's Law' problems faced by the doctrine of the Trinity (see Trinity §3). Van Inwagen has shown that the techniques employed in this solution can be extended to solve the similar problems faced by the doctrine of the Incarnation. The following is a simplified version of van Inwagen's solution.

Suppose that, although God is not (of course) the same substance or being as the human being Jesus of Nazareth, he is nevertheless the same *person* - the same 'I' or 'thou' or 'he'. (This assumption has two closely related presuppositions: that it is possible for *x* to be the same person as *y* but not the same substance; and that if *x* is the same person as *y*, and *x* has the property *F*, it does not follow that *y* has *F*.) Let the adjective 'Nazarene' represent

some conjunction of 'human' properties that uniquely identify the human being Jesus of Nazareth. We may now offer the following three definitions:

Jesus Christ is *F simpliciter*:

Something *x* is such that: something divine is the same person as *x*; and something Nazarene is the same person as *x*; and *x* is *F*.

Jesus Christ is *F qua* God:

Something *x* is such that: something divine is the same person as *x*; and something Nazarene is the same person as *x*; and *x* is divine; and *x* is *F*.

Jesus Christ is *F qua* man:

Something *x* is such that: something divine is the same person as *x*; and something Nazarene is the same person as *x*; and *x* is human; and *x* is *F*.

The two latter definitions have the expected and desired consequences. The following two consequences of the first, however, although not unorthodox, are certainly somewhat counterintuitive: Jesus Christ is eternal *simpliciter* (since something divine and something Nazarene are the same person, and the former is eternal), and Jesus Christ had a beginning in time *simpliciter* (since something divine and something Nazarene are the same person, and the latter had a beginning in time). Using constructions like those illustrated in the three definitions, van Inwagen has shown how to correlate each statement endorsed by the orthodox doctrine of the Incarnation with a statement in a formal language comprising only the two 'relative identity' predicates 'is the same being as' and 'is the same person as' and a small stock of additional predicates. (The formal language contains no names or descriptions; it contains no terms but variables.) He has shown that the set of statements in the formal language that are correlated with the set of statements that orthodoxy endorses are formally consistent (given a certain explicitly stated set of rules that defines valid inference in the formal language). This 'solution' to the problem of the apparent logical inconsistency of the doctrine of the Incarnation, unlike the other solutions we have examined, has no ontological content. Van Inwagen makes no attempt at a metaphysical analysis of divinity or humanity; his project is simply to set out a formal representation of the orthodox doctrine that is provably formally consistent.

His solution raises at least two serious questions (if, indeed, metaphysically empty as it is, it is proper to call it a solution). First, do the two 'closely related presuppositions' mentioned above (that it is possible for *x* to be the same person as *y* but not the same substance; and that if *x* is the same person as *y*, and *x* has the property *F*, it does not follow that *y* has *F*) really make any logical sense? Second, van Inwagen's constructions represent the statements of traditional Incarnational theology as having very different logical forms from those they appear to have. (For example, given the appropriate theological assumptions, the sentence 'Jesus Christ is eternal *qua* God' is true on his analysis - but not because 'Jesus Christ' is a singular referring expression that denotes an object that has, on those assumptions, the property expressed by the predicate 'is eternal *qua* God'.) In view of this fact, can these constructions plausibly be held to represent the intended content of the traditional statements?

See also: Atonement; Identity; Immutability §3; Simplicity, divine

PETER VAN INWAGEN

References and further reading

Bettenson, H. (ed.) (1963) *Documents of the Christian Church*, Oxford and London: Oxford University Press, 2nd edn.(A useful collection. Contains materials relating to Apollinarianism, Nestorianism and - under the heading 'Eutychianism'- monophysitism.)

Geach, P.T. (1977) *Providence and Evil*, Cambridge: Cambridge University Press. (An interesting discussion of the logic of 'predication *qua*', with applications to Christology; see especially pages 24-8.)

Gregory of Nazianzus (mid-to-late 4th century) 'Letter 101', in J.P. Migne (ed.) *Patrologia Graeca*, vol. 37, Paris, 1857-66, 162 vols; also in A.E. McGrath (ed.) *The Christian Theology Reader*, Oxford: Blackwell, 1995, 141-2.(The source for the quotation discussed in §1, a famous line standardly quoted in discussions of Apollinarianism. McGrath's book contains an English translation of relevant parts of Gregory's letter.)

Herbert, R.T. (1979) *Paradox and Identity in Theology*, Ithaca, NY: Cornell University Press.(An excellent

discussion of the logical problems of the Incarnation, which contains Herbert's 'ambiguous figure' model; see especially chapter 4, 'The Absolute Paradox: The God-Man'.)

Kelly, J.N.D. (1960) *Early Christian Doctrines*, London: A. & C. Black, 2nd edn. (The source of the translation of the Definition of Chalcedon quoted in the text; see especially pages 339-41.)

Kelly, J.N.D. (1964) *The Athanasian Creed*, London: A. & C. Black.(The source of the translation quoted in the text; see especially pages 19-20.)

McGrath, A.E. (1994) *Christian Theology: An Introduction*, Oxford: Blackwell. (Recommended for readers with no background in theology or Church history. Clear and reliable. See especially chapter 9, 'The Doctrine of the Person of Christ'.)

Morris, T.V. (1986) *The Logic of God Incarnate*, Ithaca, NY: Cornell University Press.(The most important recent work on the Incarnation by a philosopher at time of writing. Contains discussions of most of the issues addressed in this article, as well as a presentation of Morris' 'two minds' Christology. May be consulted for references to a wide range of recent philosophical work on the Incarnation.)

Van Inwagen, P. (1994) 'Not by Confusion of Substance, but by Unity of Person', in A.G. Padgett (ed.) *Reason and the Christian Religion: Essays in Honour of Richard Swinburne*, Oxford: Clarendon Press.(An application of the 'logic of relative identity' to the problems of Christology. Contains what could be described as an analysis of '*qua* God' and '*qua* man' in terms of relative identities.)

Incommensurability

When one scientific theory or tradition is replaced by another in a scientific revolution, the concepts involved often change in fundamental ways. For example, among other differences, in Newtonian mechanics an object's mass is independent of its velocity, while in relativity mechanics, mass increases as the velocity approaches that of light. Earlier philosophers of science maintained that Einsteinian mechanics reduces *to Newtonian mechanics in the limit of high velocities. However, Thomas Kuhn (1962) and Paul Feyerabend (1962, 1965) introduced a rival view. Kuhn argued that different scientific traditions are defined by their adherence to different* paradigms, *fundamental perspectives which shape or determine not only substantive beliefs about the world, but also methods, problems, standards of solution or explanation, and even what counts as an observation or fact.* Scientific revolutions *(changes of paradigm) alter all these profoundly, leading to perspectives so different that the meanings of words looking and sounding the same become utterly distinct in the pre- and post-revolutionary traditions. Thus, according to both Kuhn and Feyerabend, the concepts of mass employed in the Newtonian and Einsteinian traditions are incommensurable with one another, too radically different to be compared at all. The thesis that terms in different scientific traditions and communities are radically distinct, and the modifications that have stemmed from that thesis, became known as the thesis of incommensurability.*

1 Early phase of the debate

The thesis of incommensurability stemmed from new historical research indicating to some that ' once current views of nature were, as a whole, neither less scientific nor more the product of human idiosyncrasy than those current today' (Kuhn 1962: 2). Transitions from one theory to another seemed no longer understandable as resulting from new evidence forcing the rejection of one theory and the acceptance of a new one. The positivistic idea of evidence common to the rejected view and its replacement, of observations which remained invariant under theoretical change, had become suspect. More radical sorts of change seemed involved, implicating even scientific standards, methods and goals.

In the introduction to *The Structure of Scientific Revolutions* (1962) Kuhn framed the incommensurability claim in terms of ' paradigms', bodies of 'universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners'. However, the paradigm itself is more 'global' than formulations in terms of particular theories or achievements. During its period of dominance, the paradigm governs research, shaping the community's ' methods, problem-field, and standards of solution' (1962: 102). Different paradigms shape different conceptions of all these, and also of observations or facts, methods, goals of science, and therefore the definition of the corresponding science. There remain no extra-paradigmatic criteria in terms of which two different paradigms can be compared or judged. Consequently, despite the retention in successive paradigms of many of the same terms (for example, 'mass' in Newtonian and Einsteinian physics), the concepts (presumably, the meanings of the terms) are so different as to be incomparable. Different paradigms are ' not only incompatible but often actually incommensurable with that which has gone before' (1962: 102). Communication between adherents of different paradigms is impossible (see Kuhn, T.S.).

Feyerabend's (1962, 1965) formulation of the thesis is similar to Kuhn's in many respects. He speaks of 'high-level background theories', rather than more 'global' paradigms, which shape the way science is done, and is more explicit than Kuhn in asserting that high-level background theories lie behind languages. Nevertheless, he agrees with Kuhn that at least some fundamental theories are incommensurable. Whereas the early Kuhn was at best ambivalent about the extent of incommensurability between paradigms (see §3), Feyerabend throughout his career asserted that incommensurability is rare, occurring only under special circumstances, the nature of which he never clarified. (Kuhn later came to agree that incommensurability is only 'local'.) The major difference between the two is that while Kuhn believed a scientific community or tradition to be governed by a common paradigm, Feyerabend argued that scientists should try to develop alternative hypotheses, the more radically different from the accepted one the better (see Feyerabend, P.K.).

2 Criticisms of early forms of the thesis

These early forms of the incommensurability thesis were criticized on four general grounds.

First, the thesis as presented is vague and ambiguous. There are problems with the claim itself and what it has to do with. As to the claim itself, it relied on the vague terms ‘paradigm’ and ‘high-level background theory’; but what is a paradigm exactly, especially if it is more ‘global’ than any attempted formulation of it? (Masterman (see Lakatos and Musgrave 1970) pinpoints twenty-one different uses of the word ‘paradigm’ in Kuhn’s 1962 edition of *Scientific Revolutions*.) How, precisely, does it shape or determine the ways members of a tradition or community think - the standards, methods, and so on? Are the ways of thinking deduced from the paradigm, or does the paradigm exercise merely a psychological (rather than logical) control? These and other vagaries infected the incommensurability thesis, and were only aggravated by Kuhn’s frequent but never-explained qualifications of his more radical contentions. Thus he declared, without further comment, that it is only ‘often’ that paradigm changes result in incommensurability. Elsewhere he maintained that ‘ proponents of competing paradigms practise their trades in different worlds … [and] see different things when they look from the same point in the same direction’ (1962: 150). Yet he also alleged that they ‘see different things’ in ‘different areas’, presumably of the same world.

The thesis is also unclear about what it is about different paradigms that is incommensurable. Kuhn himself spoke of three types of incommensurability: conceptual, observational and methodological. Much discussion, both on Kuhn’s part and that of his critics, has centred on the first of these: how is the claim of conceptual incommensurability to be understood? Much of Kuhn’s talk suggests that it has to do with terms like ‘mass’ and their incomparability in different paradigm traditions. But what is it about such terms that is incommensurable? Is it their *meanings* that are so radically different, or their *referents*? Focusing on their meanings (for reference, see §3), what is supposed to be part of the incomparable meaning of a term? There are obviously similarities between the Newtonian and Einsteinian meanings of ‘mass’; if the meanings themselves are held incomparable, the similarities are presumably relegated to the non-meaning portions of their use. Kuhn does not discuss this distinction. Further, are all the terms of a paradigm tradition incommensurable with all those of another, or does incommensurability arise only for some of the terms? Certainly Feyerabend believed that ‘ the meaning of every term we use depends upon the theoretical context in which it occurs’ (1965: 180). Kuhn, too, sometimes wrote as if proponents of different paradigms are ‘always at cross-purposes’, that communication across paradigms is ‘inevitable’, though sometimes he qualified this claim by saying that they are only ‘partially’ so, ‘in some areas’ (1962: 149). But which is it, partial in all areas, complete in some areas, or both? Also, if the paradigm logically *entails* all the ways of thinking within a tradition, everything, presumably including all meanings, will be affected by the profound differences between paradigms; and even if determination by paradigm is only psychological rather than logical, the limits, if any, of such determination are left unclear. We are left wondering what it is that is incomparable, and whether there is anything that is not, and why.

Second, the thesis is incoherent, even self-contradictory. Feyerabend defines ‘incommensurability’ as: ‘ Two theories will be called incommensurable when the meanings of their main descriptive terms depend on mutually inconsistent principles’ (1965: 227). This definition is puzzling, since for two propositions to be inconsistent with one another is for one to assert A and the other not A; but this is to share A and be comparable. In general there is ambiguity as to whether two fundamental theories *compete with*, *contradict* one another, or are strictly incomparable. This ambiguity also occurs in Kuhn when he claims that different paradigms explain in different ways; for this claim seems to require meta-paradigmatic criteria for identifying what counts as an explanation in each of the two paradigms - for saying that they differ in how they do something that can be compared, namely explain.

Third, the thesis leads to relativism - and merely through its vagueness and ambiguity, not through the revelations of its case studies. If a paradigm determines, in some sense, the standards, methods, goals, meanings of terms, and so on, used in a tradition or community, and if two paradigms are incommensurable with regard to these, then what criteria are left for assessing the relative merits of two different paradigms? *Reasons* become paradigm-relative, and there remain no paradigm-transcending reasons to accept one paradigm over another. Furthermore, these relativist consequences follow merely from the fact that paradigm scope and control are left so ambiguous. Much of the text suggests that paradigms control in the sense of determining, either logically or psychologically, all that goes on within a paradigm tradition. If, as Kuhn claimed in later years, this was never his intent, what he did want to say on these points was far from clear (see Epistemic relativism).

Fourth, the thesis is falsified by actual science and its history. A host of historical case studies have purported to

refute (or confirm) the incommensurability thesis. Many such studies, however, uncritically take it for granted that the thesis is clear and unambiguous, and that it need only be confronted with solid historical evidence to be refuted or confirmed.

3 Later phase of the debate

To the 1970 edition of *The Structure of Scientific Revolutions*, Kuhn appended a 'Postscript' replacing the term 'paradigm' with 'disciplinary matrix' and distinguishing four components of this matrix: 'symbolic generalizations', 'metaphysical paradigms', 'values' and 'exemplars' (1970: 182-6). These are transparadigmatic and can be used in comparison and evaluation of paradigms. The 'values' include ' accuracy of prediction…; the balance between esoteric and everyday matter; and the number of different problems solved' (1970: 205-6). However, ' such reasons function as values and…they can thus be differently applied…by men who concur in honoring them' (1970: 199-200). Thus they appeared to be agreed-on standards of comparison in name only, doing little to clarify and defend the thesis of incommensurability.

Kuhn's later writings (for example, 1983) speak decreasingly of paradigms and increasingly of theories and languages. Otherwise, these later papers alter his previous ideas in at least three major respects. First, he advocates a theory of *local incommensurability*, according to which, rather than all meanings changing in revolutions, ' Only for a small subgroup of (usually interdefined) terms and for sentences containing them do problems of translatability arise' (1983: 670-1). Second, he distinguishes a 'narrow sense' of translation, piecemeal if not word-for-word, from 'interpretation', in which piecemeal translation is impossible. Third, combining these first two points, he argues that the interdefined concepts cause translation problems because they divide the world up in distinctive ways, different from the structurings of other theories or languages. This structuring can only be grasped holistically, by becoming familiar with the whole network of interrelations between the concepts and with the assumptions about the world which are made in that interlocked network. This constitutes *interpretation*, through which it is possible to gain an understanding of radically different theories even though *translation in the narrow sense* is impossible. Some critics maintain that the allegedly indissoluble connections of concepts can be disentangled in one way or another, for example, by attending to similarities and differences in the uses of *tokens* of terms (Kitcher) or by introducing a notion of *partial denotation* (Field). Kuhn's response to such arguments is that recognizing similarities and differences of denotation is only a first step in gaining adequate understanding of the alien theory or language. In any case, either on Kuhn's view or the views of such critics, this claim that two different theories can be mutually intelligible removes most of the sting that originally accompanied the charge of incommensurability. Understanding and communication are possible, even perhaps to a degree approximating that attainable through translation.

Other critics remain convinced that incommensurability in some radical sense holds for the *content of belief* of different traditions or fundamental theories, and have therefore turned to other aspects of the total scientific picture for ways of comparing and judging theories. Thus Laudan (1977: 143-5) has emphasized relative problem-solving ability, rather than factual disagreement, as the basis of theory comparison and evaluation. However, Laudan's view requires some unprovided principle of individuation for counting problems, and for giving them different weights on the basis of relative importance, ease of solution, and so on. Other writers have had recourse to formal and evaluative standards or internal consistency.

In recent years, attempts to resolve problems of incommensurability by utilizing the causal theory of reference has been the most influential approach to the incommensurability claim. The causal theory of reference holds that the reference of a term is established not by a set of descriptive conditions having to do with the meaning of the term, but by ostension (roughly, pointing) in an original 'baptismal' ceremony, which reference is passed on to succeeding generations (see Reference §4). Such constant reference can then serve as the basis for saying that very different traditions are continuous and comparable with one another, 'talking about the same thing', despite radical differences that may have come about between their descriptive concepts. There are many variations on this theme. Some writers have suggested that many successive 'baptismal events' can take place for a particular term, adding to or otherwise changing the original reference and extension; again, it has been objected that the references of theoretical terms in science cannot usually be established by pointing, and that some important types of reference must appeal to descriptive criteria. Sankey (1994) incorporates these and other modifications in a version of the causal theory that is consistent with incommensurability but allows comparability, thus harmonizing with Kuhn's

later views. Again, however, this is an extremely weak version of incommensurability. In any case, the near-mythical nature of the baptismal event, the possibility of a succession of these, and the difficulties of satisfactorily interpreting theoretical terms along the lines of the causal theory of reference, all make it unlikely that appeal to that theory can help resolve disputes about the continuity or discontinuity of science.

A further possible solution of worries about incommensurability (Shapere 1989) consists in an analysis of the *reasons* advanced for altering a reference-fixing condition (for example, the particular shape of a track in a cloud chamber under particular conditions) or for making changes in attributions of descriptive properties to theoretical entities (for example, neutrinos). On this view, changes come about in understandable ways, and can be in either reference or meaning. This approach places the burden of analysis on the concept of a reason, rather than on meaning or reference, but it is perhaps closer to the ways scientists actually proceed.

DUDLEY SHAPERE

References and further readings

Feyerabend, P.K. (1962) 'Explanation, Reduction and Empiricism', in H. Feigl and G. Maxwell (eds) *Minnesota Studies in the Philosophy of Science*, Minneapolis, MI: University of Minnesota Press, vol. III, 28-97.(One of two early papers in which Feyerabend presented his views of incommensurability.)

Feyerabend, P.K. (1965) 'Problems of Empiricism', in R.G. Colodny (ed.) *Beyond the Edge of Certainty*, Englewood Cliffs, NJ: Prentice Hall, 145-260.(The second of two early papers in which Feyerabend presented his views of incommensurability.)

Field, H. (1973) 'Theory Change and the Indeterminacy of Reference', *Journal of Philosophy* 70: 462-81.(Argues that there may be 'denotational refinement' of terms in transitions from one theoretical context to a later one.)

Gutting, G. (ed.) (1980) *Paradigms and Revolutions: Applications and Appraisals of Thomas Kuhn's Philosophy of Science*, Notre Dame, IN: Notre Dame University Press.(Articles on Kuhn, including both critical reviews and case studies.)

Hacking, I. (ed.) (1981) *Scientific Revolutions*, London: Oxford University Press. (Useful collection of articles.)

Hoyningen-Huene, P. (1993) *Reconstructing Scientific Revolutions: Thomas S. Kuhn's Philosophy of Science*, Chicago, IL: University of Chicago Press.(Careful study of Kuhn's views, usually interpreting him in a favourable light. Comprehensive bibliography.)

Kitcher, P. (1978) 'Theories, Theorists and Theoretical Change', *The Philosophical Review* 87: 519-47. (Approaches incommensurability problem via a distinction between types and tokens of a term.)

Kuhn, T.S. (1962, 1970) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press. (Second edition contains the important 'Postscript' in which Kuhn clarifies - or alters - his concept of a paradigm.)

Kuhn, T.S. (1983) 'Commensurability, Comparability, Communicability', in P.D. Asquith and T. Nickles (eds) *PSA 1982*, East Lansing, MI: Philosophy of Science Association, vol. 2, 669-88.(Contains comments by P. Kitcher and M. Hesse with responses by Kuhn. Symposium discussion revealing evolution of Kuhn's thought.)

Lakatos, I. and Musgrave, A. (eds) (1970) *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press.(Post-*Structure* Kuhn and his critics, with an initial and a final article by Kuhn from the same period as the 'Postscript' to *Structure*. See especially the article by Masterman mentioned in §2.)

Laudan, L. (1977) *Progress and its Problems*, Berkeley, CA: University of California Press.(See §3).

Sankey, H. (1994) *The Incommensurability Thesis*, Sydney: Averbury Press. (Critical review of attempts to deal with the problem, with the author's own views. Comprehensive bibliography.)

Shapere, D. (1989) 'Evolution and Continuity in Scientific Change', *Philosophy of Science*, 419-37.(Scientific change and continuity seen in terms of reasons rather than in terms of linguistic categories of meaning and reference.)

Shapere, D. (1971) 'The Paradigm Concept', *Science* 172: 706-9.(Review of the second edition of *Structure*, focusing on the 'Postscript' and on Kuhn's papers in Lakatos and Musgrave 1970, and supplementing this author's review reprinted in Gutting 1980.)

Indian and Tibetan philosophy

The people of South Asia have been grappling with philosophical issues, and writing down their thoughts, for at least as long as the Europeans and the Chinese. When Hellenistic philosophers accompanied Alexander the Great on his military campaigns into the Indus valley, on the western edge of what is now the Republic of India, they expressed delight and amazement upon encountering Indians who thought as they thought and lived the sort of reflective life that they recommended living.

Nearly all philosophical contributions in India were made by people writing (or speaking) commentaries on already existing texts; to be a philosopher was to interpret a text and to be part of a more or less well-defined textual tradition. It is common, therefore, when speaking of Indian philosophers, to identify them as belonging to one school or another. To belong to a school of philosophy was a matter of having an interpretation of the principal texts that defined that school. At the broadest level of generalization, Indians of the classical period were either Hindus, Buddhists or Jainas (see Buddhist philosophy, Indian; Hindu philosophy; Jaina philosophy). In addition to these three schools, all of which were in some sense religious, there was a more secular school in the classical period, whose tenets were materialistic and hedonistic (see Materialism, Indian school of). The end of the classical period in Indian philosophy is customarily marked by the arrival of Muslims from Turkey and Persia at the close of the first millennium. The contributions of Indian Muslims added to the richness of Indian philosophy during the medieval period (see Islamic philosophy).

Writing was introduced into Tibet not long after the arrival of Buddhism from India in the seventh century. The earliest literature of Tibet was made up mostly of Buddhist texts, translated from Indian languages and from Chinese. Eventually, ideas associated with Bon, the indigenous religion of Tibet, were also written down. Tibetan philosophers followed the habit of Indians in that they made their principal contributions by writing commentaries on earlier texts (see Tibetan philosophy). Key Buddhist philosophers from Tibet are Sa skya Paṇḍita (1182-1251), Tsong kha pa Blo bzang grags pa (1357-1419), rGyal tshab dar ma rin chen (1364-1432), mKhas grub dge legs dpal bzang po (1385-1438) and Mi bskyod rdo rje (1507-54).

1 Hindu philosophy

The philosophical schools associated with what we now call Hinduism all had in common respect for the authority of the Veda (‘Knowledge’), scriptures accepted as a revealed body of wisdom, cosmological information and codes of societal obligations. The textual schools that systematized disciplines derived from the Veda were the Mīmāṃsā, the Nyāya, the Vaiśeṣika, the Sāōkhya and the various Vedānta schools (see Mīmāṃsā; Nyāya-Vaiśeṣika; Sāōkhya; Vedānta). Concerned as all these schools were with correct interpretation of the Veda, it is natural that questions of language were of paramount importance in Indian philosophy (see Language, Indian theories of; Meaning, Indian theories of). These involved detailed investigation into how subjects are to be defined and how texts are to be interpreted (see Definition, Indian concepts of; Interpretation, Indian theories of).

Closely related to questions of language were questions of knowledge in general and its sources (see Epistemology, Indian schools of; Knowledge, Indian views of). The two most important sources of knowledge that Indian philosophers discussed were sensation and inference, the theory of inference being important to the development of logic in India (see Sense perception, Indian views of; Inference, Indian theories of). Another topic about which Indian thinkers had much to say was the problem of how absences are known (see Negative facts in classical Indian philosophy). Because of the importance of scriptures and religious teachers, epistemologists in India discussed the issue of the authority of texts and the question of the reliability of information conveyed through human language (see Testimony in Indian philosophy). The questions associated with epistemology are in Indian philosophy often closely connected with questions of human psychology (see Awareness in Indian thought; Error and illusion, Indian conceptions of).

Most schools of Indian philosophy offered not only an epistemology but also an ontology (see Ontology in Indian philosophy). Many posited a personal creator god or an impersonal godhead (see God, Indian conceptions of; Brahman; Monism, Indian). Just how particular things come into being through creative agency or through impersonal natural laws was a matter of considerable debate (see Causation, Indian theories of; Cosmology and cosmogony, Indian theories of). Indian thinkers also debated the precise nature of matter, the ontological status of

universals, and how potentials become actualities (see Matter, Indian conceptions of; Universals, Indian theories of; Potentiality, Indian theories of).

In addition to epistemology and metaphysics, a third area that Indian systematic philosophers nearly always commented upon were issues concerning the nature of the human being (see Self, Indian theories of; Mind, Indian philosophy of). This included thoughts on a variety of ethical questions and the rewards for living an ethical life (see Duty and virtue, Indian conceptions of; Karma and rebirth, Indian conceptions of; Fatalism, Indian; Heaven, Indian conceptions of). While most thinkers dealt with individual ethics, some also gave attention to the question of collective behaviour and policy (see Political philosophy, Indian).

The Hindu tradition produced a number of important individual philosophers. Among the earliest extant philosophers from India are the political theorist Kauṭilya (fourth century BC) and the grammarian and philosopher of language Patañjali (second century BC). The legendary founder of the Nyāya school, Akṣapāda Gautama, is traditionally regarded as the author of a set of aphorisms that modern scholars believe were composed in the second or third century. These aphorisms present the basic ontological categories and epistemological principles that were followed not only by the Nyāya school but by many others as well. The philosopher of language Bhartṛhari (fifth century) developed the intriguing idea that the basic stuff of which all the universe is made is an intelligence in the form of a readiness to use language. Vātsyāyana (fifth century) and Uddyotakara (sixth century) were both commentators on Gautama. The Vedānta systematist Śaōkara (eighth century) wrote that realizing the underlying unity of all things in the form of Brahman could set one free. The aesthetician Abhinavagupta (tenth-eleventh century) made the education of the emotions through the cultivation of aesthetic sensitivity the basis of liberation from the turmoil of life. Udayana (eleventh century) of the Nyāya school developed important arguments for the existence of God. Rāmānuja (eleventh-twelfth century) and Madhva (thirteenth century), both Vedāntins, offered systems that became serious rivals to Śaōkara's monism. The work of the logician Gaōgeśa (fourteenth century), who revised the classical system of logic and epistemology, became the foundation for an important new school of thought, Navya-Nyāya ('New Nyāya'). Mādhava (fourteenth century) and Vallabhācārya (fifteenth-sixteenth century) made important contributions to Vedāntin philosophy. Gadādhara (seventeenth century) continued making advances in logical theory by building on the work of Gaōgeśa. Also important in the sixteenth century were several thinkers who commented upon the religious thinker Caitanya (see GauḌĪya VaiṢṆavism). Finally, there were several thinkers and movements in the nineteenth and twentieth centuries, a period during which Indian intellectuals struggled to reconcile traditional Indian ways of thinking with European and especially British influences (see Aurobindo Ghose; Gandhi, M.K.; Radhakrishnan, S.; Tagore, R.; Arya Samaj; Brahmo Samaj; Ramakrishna Movement).

2 Buddhist and Jaina philosophy

As was the case for Hindu philosophy, Buddhist and Jaina Philosophy in India tended to proceed through commentaries on already existing texts. Jainism was founded by Mahāvīra and is best known for its method of seeing every issue from every possible point of view (see Manifoldness, Jaina theory of). The principal Buddhist traditions that incorporated significant philosophical discussions were those that tried to systematize doctrines contained in various corpora of texts believed to be the words of the Buddha (see Buddhism, Ābhidharmika schools of; Buddhism, Mādhyamika: India and Tibet; Buddhism, Yogācāra school of). An important issue for Buddhist thinkers, as for most Indian philosophers, was analysing the causes of discontent and suggesting a method for eliminating unhappiness, the cessation of suffering being a condition known as *nirvāṇa* (see Suffering, Buddhist views of origination of; Nirvāṇa). A doctrine of special interest to the Mādhyamika school was that everything is conditioned and therefore lacking independence (see Buddhist concept of emptiness). Some Buddhists developed the view that the conditioned world is so transitory that it disappears and is recreated in every moment (see Momentariness, Buddhist doctrine of). In the area of epistemology and philosophy of language, some Buddhists repudiated the Hindu confidence in the authority of the Veda (see Nominalism, Buddhist doctrine of).

The Buddhist tradition gave India a number of important philosophers, beginning with the founder of the religion, the Buddha (fifth century BC). The first important Buddhist philosopher to write in Sanskrit and the man traditionally regarded as the founder of the Mādhyamika school was Nāgārjuna (second century). A key commentator in both the Ābhidharmika schools and in the Yogācāra school was Vasubandhu (fifth century). Two key Buddhist epistemologists and logicians were Dignāga (fifth century) and Dharmakīrti (seventh century).

Buddhism disappeared from northern India in the twelfth century and from southern India a few centuries later. In the twentieth century, there has been an effort to revive it, especially among the community formerly known as ‘untouchables’. A remarkable leader of this community was Bhimrao Ramji Ambedkar.

3 Pronunciation of Sanskrit words

Sanskrit is an Indo-European language, closely related to Greek and Latin. In India, it is written in a variety of phonetic scripts, and in the West it is customary to write it in roman script. Many letters used to write Sanskrit are pronounced almost as they are in English; k, g, j, t, d, n, p, b, m, y, r, l, s and h can be pronounced as in English without too much distortion. The sound of the first consonant in the English word ‘church’ is represented by a simple ‘c’ in Sanskrit. In addition to these consonants there is a class of retroflex consonants, so called because the tongue is bent back so that the bottom side of the tongue touches the roof of the mouth. These sounds are represented by letters with dots under them: ṭ, ḍ, ṇ and ṣ. As in English, some consonants are heavily aspirated, so that they are pronounced with a slight puff of air. These consonants are represented by single letters in Indian scripts but by two-letter combinations in roman script; thus ‘kh’ is pronounced as the ‘k’ in English ‘kill’, ‘th’ as ‘t’ in ‘tame’ (never as ‘th’ in ‘thin’ or ‘there’), ‘dh’ as in ‘mudhouse’, and ‘ph’ as ‘p’ in ‘pat’ (never as ‘ph’ in ‘philosophy’). The letter ‘ś’ is approximately like ‘sh’ in ‘shingle’. The letter ‘ō’ is like ‘ng’ in ‘finger’ or ‘nk’ in ‘sink’, while ‘ñ’ is approximately like ‘ny’ in ‘canyon’.

Vowels are pronounced approximately as in Spanish or Italian. Vowels with a macron over them (ā, ī and ū) are pronounced for twice as much time as their unmarked equivalents. The vowel ‘ṛ’ is pronounced with the tip of the tongue elevated towards the roof of the mouth, very much like the ‘er’ in the American pronunciation of ‘carter’. The diphthongs ‘ai’ and ‘au’ are pronounced as ‘i’ in ‘kite’ and ‘ou’ in ‘scout’ (or almost as ‘ei’ and ‘au’ are pronounced in German) respectively. Accent tends to be on the third syllable from the end; thus the name ‘Ś āōkara’ sounds like ‘SHANG-ka-ra’, not ‘shang-KA-ra’. If the second syllable from the end is long, then it is accented; ‘Dignāga’ is pronounced ‘dig-NAA-ga’.

4 Pronunciation of Tibetan words

Tibetan is a language of the Sino-Tibetan family, which includes various languages spoken in China as well as Burmese and Thai. It is written in a phonetic alphabet derived from the Brahmi script of India, from which most modern Indian scripts, as well as the alphabets used to write Sinhalese, Thai and Mon, are also derived. There are many different systems commonly used by Europeans to transliterate the spelling of Tibetan words. In this encyclopedia, the system designed by T. Wylie is used for transliteration, and a system used at the University of Virginia is used to indicate approximate pronunciation of names.

The spelling of Tibetan words was fixed over a millennium ago and has not changed since. Pronunciation, however, has shifted. Unfortunately, it has not shifted in exactly the same way in every region of Tibet, with the result that the same written word may be pronounced quite differently in the east of Tibet from the way it is pronounced in the west and the central region. The University of Virginia system of indicating pronunciation captures the dialects of central Tibet, which have shifted the greatest distance from the pronunciations of a millennium ago. Consequently, many combinations of letters are not pronounced at all as they once were, and numerous letters have become silent in modern central Tibetan dialects. Given all these changes, the pronunciation of some Tibetan words can be surprisingly different from what one might expect from their spelling.

Many single letters and combinations are pronounced about as in Sanskrit, as described above; so k, kh, g, ō, c, j, ñ, t, th, d, n, p, ph, b, m, y, r, l, s and h can be pronounced as described there. The pairs of letters ‘ts’ and ‘dz’ represent single letters in the Tibetan alphabet and are pronounced as they would be in English ‘cats’ and ‘adze’ respectively. The letters ‘tsh’ represent a single Tibetan letter that is pronounced like an aspirated version of ‘ts’. The combination ‘sh’ is used to represent a Tibetan letter that is pronounced about like ‘sh’ in ‘show’. Some combinations of Tibetan letters are no longer pronounced as they were when spelling was fixed. Examples of this are ‘kr’, ‘tr’ and ‘pr’, all of which are now pronounced the same way, approximately as the ‘tr’ in ‘trick’. Similarly, ‘gr’, ‘dr’ and ‘br’ are all pronounced about like ‘dr’ in ‘drink’. When the letters ‘g’, ‘b’, ‘m’, ‘r’, ‘l’ and ‘s’ occur at the beginning of a syllable and are followed immediately by any consonant other than ‘r’ or ‘l’, they are usually silent. The letter ‘s’ at the end of a syllable is usually silent. Thus ‘bsdigs’ is pronounced somewhere between English ‘dig’ and ‘dick’.

The Tibetan script does not have upper-case and lower-case letters, so there is no custom of writing proper names any differently from ordinary words. In roman transliteration, however, it is customary to capitalize the first pronounced letter of a name. In the name 'rGyal tshab', for example, the silent 'r' is not capitalized. Similarly, in 'mKhas grub' the silent 'm' is not upper case.

Tibetan consonants are pronounced with no trace of aspiration or with heavy aspiration. To the English ear attuned to hearing aspiration about midway between that used in Tibetan consonants, Tibetan 't' can sound like English 'd' and vice versa. Similarly, Tibetan 'k' and 'p' can sound like English 'g' and 'b' respectively. At the end of words, Tibetan 'g' and 'b' may sound like English 'k' and 'p' respectively. It is for this reason that the Virginia phonetic system renders 'Tsong kha pa' as 'Dzong-ka-ba' and 'rGyal tshab' as 'Gyel-tsap'. In the name 'mKhas grub rje' we can see many of the principles discussed above represented in its Virginia rendering as 'kay-drup-jay'.

See also: Buddhist philosophy, Chinese; Iqbal, M.; Mysticism, history of; Reincarnation; Salvation; Shah Wali Allah (Qutb al-Din Ahmad al-Rahim)

RICHARD P. HAYES

Further reading

Mohanty, J.N. (1992) *Reason and Tradition in Indian Thought. An Essay on the Nature of Indian Philosophical Thinking*, Oxford: Clarendon Press.(A thoughtful exploration of the principal issues of Indian philosophy.)

Powers, J. (1995) *Introduction to Tibetan Buddhism*, Ithaca, NY: Snow Lion Publications.(A useful survey of the main schools of Tibetan Buddhism, as well as of the Bon tradition.)

Raju, P.T. (1985) *Structural Depths of Indian Thought*, Albany, NY: State University of New York Press.(A good survey of the different schools of Indian philosophy from ancient times to the present.)

Indicative conditionals

Examples of indicative conditionals are 'If it rained, then the match was cancelled' and 'If Alex plays, Carlton will win'. The contrast is with subjunctive or counterfactual conditionals, such as 'If it had rained, then the match would have been cancelled', and categoricals, such as 'It will rain'.

Despite the ease with which we use and understand indicative conditionals, the correct account of them has proved to be very difficult. Some say that 'If it rained, the match was cancelled' is equivalent to 'Either it did not rain, or the match was cancelled'. Some say that the sentence asserts that the result of 'adding' the supposition that it rained to the actual situation is to give a situation in which the match was cancelled. Some say that to assert that if it rained then the match was cancelled is to make a commitment to inferring that the match was cancelled should one learn that it rained. This last view is often combined with the view that indicative conditionals are not, strictly speaking, true or false; rather, they are more or less assertible or acceptable.

1 Indicative and material conditionals

In general an 'indicative conditional' has the form 'If *A* then *C*', where *A* is called the antecedent and *C* the consequent. A central issue is the relationship between the truth-value of a conditional and the truth-values of its antecedent and consequent. This much is immediately plausible: if *A* is true and *C* is false, then the conditional is false. If I say 'If it rained, the match was cancelled', and what happened was that it rained but the match went ahead, then what I say is clearly false. There are three other possibilities: *A* and *C* are both true, *A* is false and *C* is true, and *A* and *C* are both false. There are a number of arguments designed to show that in each of these cases the conditional is true. Here is one. 'If *A* then *C*' is intuitively equivalent to the disjunction 'Either not-*A*, or *A* and *C*'. (Instead of saying that if it rained the match was cancelled, I could have said 'Either it did not rain, or it did and the match was cancelled'.) But the latter is true in all three cases: in each, either the first disjunct ('not-*A*') or the second disjunct ('*A* and *C*') is true. A conditional that is false when its antecedent is true and its consequent false, and true in all other cases, is called a 'material conditional', and is symbolized '$A \supset C$' (read '*A* hook *C*'). '$A \supset C$' is definitionally equivalent to 'Not-*A* or *C*'. Hence the argument's conclusion is that indicative conditionals are equivalent to material conditionals. This conclusion has the virtue of validating the two most obviously valid inferences governing conditionals - *modus ponens*: '*A*, if *A* then *C*, therefore, *C*', and *modus tollens*: 'Not-*C*, if *A* then *C*, therefore, not-*A*'.

Despite the appeal of the argument, there are serious problems for the equivalence thesis that indicative conditionals are material conditionals. It entails that any conditional with a false antecedent is true regardless of its consequent, that is, the account validates 'Not-*A*, therefore, if *A* then *C*'. This is implausible. 'If I live in London then I live in Scotland' strikes us as false (it is rather 'If I live in London then I do not live in Scotland' which is true), but because I do not live in London, it is true on the equivalence thesis. Also, it entails that any conditional with a true consequent is true: the account validates '*C*, therefore, if *A* then *C*'. But 'If I live in London, then I live in Australia' strikes us as false even after we learn that I do in fact live in Australia. These two results are known as the paradoxes of material implication ('material implication' being the name of the relation between *A* and *C* when '$A \supset C$' is true).

2 Possible worlds treatments

An obvious response to the paradoxes is to insist that more than the truth of '$A \supset C$' is required in order for 'If *A* then *C*' to be true; that the truth of '$A \supset C$' is a necessary but insufficient condition. One way to do this is to require a connection between antecedent and consequent; to hold that part of what makes 'If it rains, the grass will grow' true is the connection between rain and grass growing. This blocks the paradoxes. However, we sometimes use conditionals precisely to deny that there is a connection between antecedent and consequent. A doctor who says that if you go to bed and take an aspirin you will get better in a week, whereas if you go to work it will take seven days, is saying that there is no connection between going to bed and getting better.

A more promising approach is to require that the material conditional not only be true the way things actually are, but be true in the closest possible worlds where the antecedent is true (see Semantics, possible worlds). Roughly, the account is: 'If *A* then *C*' is true iff '$A \supset C$' is true, that is, true the ways things actually are *and also* true in

the closest possible worlds where *A* is true. Another way of saying the same thing is: 'If *A* then *C*' is true iff the closest worlds in which *A* is true are worlds in which *C* is also true. (This is because any world where *A* and '$A \supset C$' are true is a world where *C* is true.) The appeal of this approach derives from the appeal of the idea that when we evaluate a conditional we look at the way things actually are and 'in imagination' add the antecedent and then see whether, with only the changes forced by adding the antecedent, the consequent comes out true.

This approach is very attractive for subjunctive or counterfactual conditionals (see Counterfactual conditionals) but faces a problem when applied to indicative conditionals. (Interestingly, one architect of the possible worlds approach to conditionals, David Lewis (1973), only ever intended it to apply to subjunctive conditionals. Robert Stalnaker, the other architect (1968), intended it to apply to both.) It construes indicative conditionals as being in part about the way things might have been but are not in fact. That seems exactly right for subjunctives. When I say that had I invested in Western Mining ten years ago, I would now be rich, I am saying how things are in some possible but regrettably non-actual world. Indeed, I might express this by saying that had I invested in Western Mining ten years ago, things would now be different - and better - than they actually are. But I cannot say this in the indicative mood: 'If I invested in Western Mining ten years ago, things are different from the way they actually are' is nonsense. Or consider the difference between 'If Oswald had not shot Kennedy, Kennedy would have won a second term' and 'If Oswald did not shoot Kennedy, Kennedy won a second term'. The reason we reject the second, despite accepting the first, is that we know that the way things actually are Kennedy did not live to win a second term. Subjunctive conditionals typically concern non-actual worlds, whereas indicative conditionals concern the actual world under various hypotheses about what it is like. Thus we say that if Oswald did not shoot Kennedy, someone else did, because that is the only thing to think about the actual world (in which we know that Kennedy was shot) given that Oswald was not responsible.

3 Adams and Lewis

What about the acceptance or justified assertion conditions, as opposed to truth-conditions, for indicative conditionals? Many, including particularly Ernest Adams (1975), have urged that it is justified to assert 'If *A* then *C*' to the extent that *C* is probable given *A*. Because the probability of *C* given *A* - in symbols, $Pr(C|A)$ - is (roughly) the probability we ascribe to *C* on learning *A* (see Probability, interpretations of §5), this suggestion fits nicely with the plausible idea that one is prepared to assert 'If *A* then *C*' to the extent that one is prepared to infer *C* on learning *A*. It also explains our reluctance to assert together 'If *A* then *C*' and 'If *A* then not-*C*' when *A* is consistent, for when *A* is consistent, $Pr(C|A) = 1 - Pr(\text{not-}C|A)$, so they cannot be high together.

Can we explain this assertibility condition in terms of truth-conditions? David Lewis (1976) showed that it cannot be the case that the truth-conditions of 'If *A* then *C*' are such that *Pr*(If *A* then *C*) = $Pr(C|A)$. For then this equality would hold for all probability functions *Pr*, and that leads to trouble as follows:

$$Pr(\text{If } A \text{ then } C) = Pr(\text{If } A \text{ then } C|C).Pr(C) + Pr(\text{If } A \text{ then } C|\text{not-}C).Pr(\text{not-}C),$$

by expansion by cases. But if *Pr*(If *A* then *C*) = $Pr(C|A)$ holds for all *Pr*, then it holds for $Pr(-|C)$ and $Pr(-|\text{not-}C)$, as the class of probability functions is closed under conditionalization. This means that

$$Pr(\text{If } A \text{ then } C|C) = Pr(C|A.C)$$
$$Pr(\text{If } A \text{ then } C|\text{not-}C) = Pr(C|A.\text{not-}C)$$

(where $Pr(C|A.C)$ is the probability of *C* given the conjunction of *A* and *C*). Hence we have

$$\begin{aligned} Pr(\text{If } A \text{ then } C) &= Pr(C|A.C).Pr(C) + Pr(C|A.\text{not-}C).Pr(\text{not-}C) \\ &= 1.Pr(C) + 0.Pr(\text{not-}C) \\ &= Pr(C) \end{aligned}$$

But then, by the claim under discussion, $Pr(C) = Pr(C|A)$. This is a 'reductio', for in general the probability of *C* is not independent of that of *A*. (This is the essence of the simplest of the proofs Lewis offered. For the substantial developments see Lewis (1986).)

Many respond to this proof by holding that indicative conditionals do not have truth-conditions at all;they only have assertion or acceptance conditions. This theory undoubtedly has its attractions but faces two problems. What

should be said about the powerful intuition that a conditional with a true antecedent and a false consequent is false? And how should the relevant notion of assertibility be elucidated? Not in terms of likelihood of truth, obviously.

An alternative strategy, called the supplemented equivalence theory, is to return to the equivalence thesis and argue that there is a convention governing the assertion of ‘If *A* then *C*’ to the effect that it should only be asserted when it would be right to infer *C* on learning *A*. This convention is like that governing the use of ‘but’. ‘*A* but *C*’ has the same truth-conditions as ‘*A* and *C*’, but the use of the former conventionally implicates, in H.P. Grice’s terminology, a contrast (see Grice 1975; Implicature). Likewise, runs the suggestion, ‘If *A* then *C*’ has the same truth-conditions as ‘ $A \supset C$’ but its use carries the implicature that my reasons for ‘ $A \supset C$’ are such that it would be right, on learning *A*, to infer *C* (that is, to use *modus ponens*). Now that will be the case just if the probability of ‘ $A \supset C$’ would not be unduly diminished on learning that *A* is true - otherwise it would not then be available as a probably true premise to combine with *A* on the way to inferring *C*. It follows that it will be right to assert ‘If *A* then *C*’ to the extent that (1) ‘ $A \supset C$’ is probable, and (2) ‘ $A \supset C$’ is probable given *A*. It is an elementary exercise in probability theory to show that this two-fold condition is satisfied to the extent that $Pr(C|A)$ is high. The supplemented equivalence theory therefore explains the assertibility condition noted by Adams.

What do supplemented equivalence theorists say about our earlier example ‘If I live in London, then I live in Scotland’, which comes out true on the equivalence theory and so on the supplemented equivalence theory? They say that it seems false not because it is false but because it has very low assertibility - *Pr*(I live in Scotland/I live in London) = 0 - and in general that our intuitions about truth and validity of inferences involving indicative conditionals are governed by responses to assertibility rather than truth.

Warning: almost everything about indicative conditionals is controversial, including whether they are well labelled by the term ‘indicative’, and some even deny the validity of *modus ponens*!

See also: Logical and mathematical terms, glossary of; Relevance logic and entailment

FRANK JACKSON

References and further reading

Adams, E. (1975) *The Logic of Conditionals*, Dordrecht: Reidel.(A detailed defence of an approach to conditionals that denies them truth-conditions and shows how to develop a logic of conditionals based on his views about their assertibility.)

Appiah, A. (1985) *Assertion and Conditionals*, Cambridge: Cambridge University Press.(A detailed exposition of an account of indicative conditionals in terms of assertibility, set within the philosophy of mind.)

Edgington, D. (1995) ‘On Conditionals’, *Mind* 104: 235-329.(Excellent survey article that contains interesting material in its own right; includes a comprehensive and up-to-date reading list.)

Grice, H.P. (1975) ‘Logic and Conversation’, in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991.(Explains implicature and the distinction between conversational and conventional implicature.)

Jackson, F. (1987) *Conditionals*, Oxford: Blackwell.(A detailed defence of the supplemented equivalence theory with criticisms of various alternative accounts of indicative conditionals.)

Lewis, D.K. (1973) *Counterfactuals*, Oxford: Blackwell.(The book which, along with Stalnaker’s paper, launched the possible worlds approach to conditionals.)

Lewis, D.K. (1976) ‘Probabilities of Conditionals and Conditional Probabilities’, in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991.(Gives the full version of the proof sketched in this entry.)

Lewis, D.K. (1986) ‘Probabilities of Conditionals and Conditional Probabilities II’, in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991.(More on the proof sketched in this entry.)

Stalnaker, R. (1968) ‘A Theory of Conditionals’, in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991.(The paper which, along with Lewis’s book, launched the possible worlds approach to conditionals.)

Indirect discourse

Indirect discourse is a mode of speech-reporting whereby a speaker conveys the content of someone's utterance without quoting the actual words. Thus, if Pierre says, 'Paris est belle', an English speaker might truly say,

(1) Pierre said that Paris is beautiful.

In English, sentences of indirect discourse often have the form 'A said that s', where 'A' refers to a person and 's' is often called the 'content sentence' of the report.

Sentences of indirect discourse have been classed with attributions of belief (and other psychological states) in view of an apparent conflict with the 'principle of the intersubstitutability of coreferring terms', which states that the truth-value of a sentence does not alter if one term in a sentence is replaced with another referring to the same thing. If (1) is true and 'Paris' and 'the City of Light' refer to the same thing, (2) may still be false:

(2) Pierre said that the City of Light is beautiful.

1 Frege's theory

In 'indirect discourse' a speaker reports someone's utterance without quoting the actual words used (often using the form '*A* said that *s*'). Sentences of indirect discourse give rise to an apparent conflict with the 'principle of the intersubstitutability of coreferring terms', which states that the truth-value of a sentence does not alter if one term is replaced with another which refers to the same thing. Gottlob Frege (1892) put forward an important theory to deal with this problem. He proposed that 'said' alters what the words in the 'content sentence' ('*s*' above) refer to. For example, consider the following statements:

(1) Pierre said that Paris is beautiful.
(2) Pierre said that the City of Light is beautiful.

'Paris' and 'the City of Light' both usually refer to the actual city, Paris. The phrase 'is beautiful' usually refers to the set of beautiful things. But when words appear after the phrase 'said that', they refer to the meanings they normally express - which Frege called their 'senses' - which are not the same as their referents (see Frege, G. §3).

The sense of an expression is a 'mode of presentation' of its referent. The sense associated with 'the City of Light' presents Paris in a particular way: as the City of Light. To understand an expression is to grasp its sense (see Sense and reference). Frege called the sense of a sentence a 'thought'. To have a propositional attitude is to stand in a certain relation to a thought. For example, to believe that Paris is beautiful is to stand in the believing relation to the thought that Paris is beautiful. Frege claimed that senses and thoughts are not psychological, mind-dependent things, but rather are Platonic, abstract objects. If two people believe that Paris is beautiful, then it is the very same thought that they both have. And this thought would exist even if there had been no minds (see Frege, G. §4).

The Fregean theory explains away the apparent failures of intersubstitution: the 'senses' of 'Paris' and 'the City of Light' are different. Hence the words refer to different things in (1) and (2) and so no violation of the substitution principle occurs.

2 Davidson's theory

Donald Davidson (1968/69) has provided a very different account of indirect discourse. Davidson proposed that, at the level of logical form (see Logical form), the analysis of speech reports should be composed of two separate sentences, the first ending in the demonstrative pronoun 'that'. The analysis of (1) above would then be:

(3) Pierre said that. Paris is beautiful.

On Davidson's theory the two sentences are logically distinct (not connected by any logical particle, such as 'and'). However, on specific occasions of use the demonstrative 'that' ending the first sentence refers to the reporter's utterance of the second.

According to this theory, the direct object of the verb 'said' is the reporter's own token utterance. So, suppose that

someone, *A*, utters (1). This speech act is composed of two token utterances: one of ‘Pierre said that’ and one of ‘Paris is beautiful’. Call this latter utterance ‘*U*’. Then, what Pierre would be reported as saying is *U*.

This feature is surprising, since it does not seem that Pierre could have said *U* (the reporter’s utterance) itself. But Davidson provides an account of what ‘said’ means which addresses the difficulty. He introduces the notion of ‘samesaying’: roughly, two individuals are samesayers if and only if they produce utterances which have relevantly similar semantic contents. Davidson has provided a general theory of meaning and interpretation which underwrites his appeal to semantic contents in this context (1984). We are then to understand (1) along the following lines:

(4) There was some utterance, *u*, of Pierre’s, and *u* and my next utterance make us samesayers. Paris is beautiful.

The idea is that the relation expressed by ‘said’ in *A*’s utterance of (1) can be understood in terms of the more complex relation spelled out in (4). But (4) is not intended to be any part of a technical semantic theory. It is just an informal, extra-theoretic guide to readers to help them understand the theory. In a semantic theory in Davidson’s style (see Davidson, D. §4), the axiom for ‘said’ might be:

(5) (*x*)(*y*)(‘said’ is true of $\langle x, y\rangle$ iff *x* said *y*).

The point is that since ‘said’ is supposed to relate reportees to utterances of reporters, some informal explanation of this saying relation is required. (4) provides this explanation.

Davidson’s theory explains the apparent failures of substitution such as that illustrated by (1) and (2). In an utterance of (1) the main verb is ‘said’ and its syntactic object is the demonstrative ‘that’. The ensuing utterance of ‘Paris is beautiful’ only serves as a referent for the demonstrative. It is not in any other way a semantically relevant component of the report. An utterance of (1) and an utterance of (2) resemble each other in that each has the form ‘Pierre said that’. But the referents of the two instances of ‘that’ are different utterances. So (1) does not entail (2). No violation of the intersubstitutability of coreferring terms occurs. The latter principle still applies to the content sentences of (1) and (2):

(6) Paris is beautiful.
(7) The City of Light is beautiful.

(6) does entail (7), given the coreference of ‘Paris’ and ‘the City of Light’. But this has no direct bearing on the logical relations between (1) and (2).

Davidson’s theory differs importantly from Frege’s. The former does not entail any shifts of reference for the words in the content sentence, which has exactly the same semantic features as it would have if uttered in isolation. This feature of the theory implements a natural and elegant idea of how speech reporting works. When *A* reports on the speech of *B*, *A* produces a new sentence, *S*, as content sentence of the report. *S* in *A* ’s mouth expresses the content of *B*’s speech act. The utterance of *S* then serves as a model of *B*’s original speech act. And this exhausts its role.

Various objections have been made to Davidson’s proposal. Only two are discussed here. The first objection is due to Tyler Burge (1986). Burge points out that on Davidson’s theory, an utterance of (8) below would be true only if there were some utterance of the reporter’s and some utterance of Pierre’s that made them samesayers:

(8) Pierre said something.

This result follows from the functioning of the word ‘said’ on the theory. But the result is wrong: I might utter (8) truly without myself ever making any utterance that would make Pierre and me samesayers. The objection is perhaps not conclusive. But Burge points out that a solution to the difficulty would be to allow the objects of speech reports to be abstract objects, rather than token utterances, and offers several further arguments that this solution is correct. It seems natural to take the relevant abstract objects to be sentence-types. If this suggestion is adopted, then Burge’s objection forces no major revisions of Davidson’s theory. The original account of ‘said’ in terms of samesaying no longer applies. But an analogous explanation of the functioning of ‘said’ remains available. Instead of understanding (1) along the lines of (4), we would understand it along the lines of:

(9) There was some sentence, *s*, which Pierre uttered, and *s* meant in Pierre's mouth then what the following means now in mine: Paris is beautiful.

The second objection is due to James Higginbotham (1986). On Davidson's account, (10) and (11) ought to have identical semantic features:

(10) Every girl said that her mother was kind.
(11) Every girl said that. Her mother was kind.

This is because (11) merely makes explicit the analysis that Davidson would attribute to (10). However, (10) and (11) are semantically distinguishable. (10) is ambiguous: on the first reading, (10) would be true if and only if each girl said of some specific contextually indicated person, *x*, that *x*'s mother was kind; on the second reading, it would be true if and only if each girl said that her own mother was kind. But (11) is only subject to the first reading. This objection shows that Davidson's theory is not correct as it stands, and will have to modified in very serious ways if it is to accommodate interactions between words inside and outside content sentences.

See also: Logical and mathematical terms, glossary of; Propositional attitude statements

GABRIEL SEGAL

References and further reading

Burge, T. (1986) 'On Davidson's "On Saying That"', in E. Lepore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, New York: Blackwell, 191-208.(A critical article on Davidson's theory of indirect discourse.)

Davidson, D. (1968/69) 'On Saying That', *Synthese* 19: 130-46.(The classic work of Davidson's theory of indirect discourse, including criticisms of Frege's theory.)

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.(A collection of Davidson's articles in the philosophy of language, including some seminal works on the theory of meaning and interpretation.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 1962, 56-78; 3rd edn, 1980.(The first clear statement of the problems of apparent failures of intersubstitutability from the standpoint of technical semantic theory. One of the most influential articles in philosophy of language.)

Geach, P.T. (1957) *Mental Acts*, London: Routledge & Kegan Paul.(A general work in the philosophies of mind and language, including extended discussion of the relationship between saying and propositional attitudes such as believing or judging.)

Higginbotham, J. (1986) 'Linguistic Theory and Davidson's Program in Semantics', in E. Lepore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, New York: Blackwell, 29-48.(Argues for the incorporation of Davidson's main technical innovations in philosophical semantic theory into empirical linguistics and includes a sophisticated theory of indirect speech that retains some core aspects of Davidson's proposal but avoids the main objections to it.)

Scheffler, I. (1954) 'An Inscriptional Approach to Indirect Quotation', *Analysis* 10: 83-90. (An influential nominalist account of indirect discourse.)

Segal, G. and Speas, M. (1986) 'On Saying ðČbepsi;t', *Mind and Language* 1 (2): 124-32. (An article criticizing Davidson's theory of indirect discourse on empirical grounds.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press, 1964.(A general work in systematic philosophy including arguments critical of semantic theories, such as Frege's, that use a mentalistic or abstract notion of meaning; and discussion of various approaches to indirect discourse and propositional attitude statements.)

Induction, epistemic issues in

Consider the following:

(1) Emeralds have been regularly dug up and observed for centuries; while there are still emeralds yet to be observed, every one observed so far has been green.

It is easy to see why we regard (1) as evidence, if true, that:

(2) Every emerald observed up until 100 years ago was green.

(1) logically implies (2): there is no way (1) could be true without (2) being true as well. It is less easy to see why we should think that (1), if true, is any evidence at all that:

(3) All hitherto unobserved emeralds are green as well.

(1) does not logically imply (3): it is consistent with (1) that (3) be false - that the string of exclusively green emeralds is about to come to an end. None the less, we do *regard (1) as evidence, if true, that (3). What, if anything, justifies our doing so?*

To answer this question would be to take a first step towards solving what is known as the 'problem of induction'. But only a first step. There is, at least on the surface, a wide variety of arguments that share the salient features of the argument from (1) to (3): their premises do not logically imply their conclusions, yet we think that their premises, if true, constitute at least some evidence that their conclusions are true. A fully fledged solution to the problem of induction would have to tell us, for each of these arguments, what justifies our regarding its premises as evidence that its conclusion is true. Still, the question as to how this step might be taken has been the focus of intense philosophical scrutiny, and the approaches outlined in this entry have been among the most important.

1 The presupposition approach

Considering the above, one proposal that leaps immediately to mind is that what justifies our regarding (1) as evidence that (3) is that we presuppose that the world is so constructed that observed samples are representative of the populations from which they are drawn - in Hume's words, that 'the future is conformable to the past' or (as others have put it) that 'nature is uniform' (see Hume, D. §2). We can express the presupposition thus (where *F* and *G* are any predicates):

(4) Whenever at least some *F*s are observed and *n* per cent of all the *F*s observed are *G*, *n* per cent of the remaining *F*s are *G*.

Given this presupposition, it is no more difficult to see why we should regard (1) as evidence that (3) than it is to see why we should regard (1) as evidence that (2): the conjunction of (1) and (4) logically implies (3). But there are three problems.

(a) *(1) and (4) imply too much*. Let us say that an object is *grue* just if *either* it has been observed by now and is green *or* it has yet to be observed and is blue. By this definition, (1) logically implies that:

(5) Emeralds have been regularly dug up and observed for centuries; while there are still emeralds yet to be observed, every one observed so far has been grue.

And, just as (1) and (4) logically imply (3), so (5) and (4) logically imply:

(6) All the hitherto unobserved emeralds are grue as well.

Thus (1) and (4) logically imply *both* (3) and (6). But, given that there are emeralds yet to be observed, (6) is incompatible with (3): unobserved emeralds that are grue are, by definition, blue. Now, the predicate 'grue' seems so obviously jerry-rigged for the occasion that it looks easy to circumvent the difficulty just described. Simply distinguish between jerry-rigged predicates and the rest - between *unprojectible* predicates and *projectible* ones - and say that what we presuppose is that observed samples are representative, *with respect to projectible predicates*,

of the populations from which they are drawn. That is:

(4′) Where *F* and *G* are projectible, whenever at least some *F*s are observed and *n* per cent of all the *F*s observed are *G*, *n* per cent of the remaining *F*s are *G*.

But only in concert with a premise asserting that ‘emerald’ and ‘green’ are projectible predicates and ‘grue’ is not, will this modification work. On pain of being entirely *ad hoc* this premise would need the support of some principled way of distinguishing between projectible and unprojectible predicates - something that has proved extremely elusive. This much, however, is worth noting. The problem just outlined (known as the ‘new riddle of induction’ - see Goodman, N. §3) is not peculiar to the presupposition approach: it is a problem that any non-sceptical approach to the problem of induction will also have to solve.

(b) *Observed samples are not, as a matter of fact, always representative of the populations from which they are drawn*. This is particularly common with small samples, but can and does occur even with large ones; nature sometimes throws us a curve. Most of this difficulty can be circumvented by moving to a yet more modest version of (4). For example, we can say that what we presuppose is that the world is so constructed that observed samples are *usually* representative of the populations from which they are drawn, that:

(4″) Where *F* and *G* are projectible, it is typically the case that when at least some *F*s are observed and n per cent of all the *F*s observed are *G*, *n* per cent of the remaining *F*s are *G*.

Of course, (1) and (4″) do not logically imply (3). But it is open to us to claim that what justifies our regarding (1) as evidence that (3) is that, in so far as we think that (4″) is true, we think that the argument from (1) to (3) is an instance of an argument form whose instances typically have a true conclusion when the premise is true. And, while further modification would be required to allay fully the worry about small samples - (4″) does not seem very credible in a case in which only one *F* has been observed as an instance - this does not seem in principle impossible.

(c) *It is hard to see how, by the lights of the presupposition approach itself, we could be justified in believing (4″) (or (4′) suitably modified)*. (4″) makes a very strong claim about the world. (4″) says, not only that samples so far observed have typically been representative, with respect to projectible predicates, of the populations from which they are drawn, but also that samples not yet observed are typically representative in this way as well. This latter claim is just of the sort we presumably are not justified in believing except on the basis of evidence - in this case, evidence that samples so far observed typically have, indeed, been representative. But, given the justificatory scheme advanced by the presupposition approach, we cannot regard the claim that observed samples have typically been representative as evidence that unobserved samples are typically representative unless we already believe (4″). Thus it would seem that very claim to which the presupposition approach would have us appeal to justify our regarding (1) as evidence that (3), is one that both requires justification and is (given this approach) incapable of receiving any.

2 The pragmatic approach

But is it so important that we believe (4″)? Clearly, if (4″) is true then the following argument form will be reliable:

(7) At least some *F*s have been observed and *n* per cent of them have been *G* (where *F* and *G* are projectible). Therefore, *n* per cent of the remaining *F*s are also *G*.

That is, an instance of (7) will typically have a true conclusion when it has a true premise. And just as clearly, if (4″) is false, (7) will be unreliable. But, some have (in effect) suggested, if (4″) is false, then *no* competing projective argument form (no other argument form that predicts what percentage of the remaining *F*s are *G* from the percentage in the observed sample) will be reliable. More precisely, they have claimed that (for any projectible *F* and *G* where the number of *F*s is infinite) if there is a limit to the relative frequency with which observed *F*s are *G* (roughly, if there is a point in an infinite sequence of observations beyond which the percentage of observed *F*s that are *G* remains fairly stable) then the repeated use of (7) will eventually predict that relative frequency: that is, they have claimed that, in such a case, (7) will eventually find the true relative frequency if any projective

argument form will. Thus the justification for our regarding (1), if true, as evidence that (3): the argument from (1) to (3) is an instance of (7), an argument form which is (in this sense) reliable if any projective argument form is (see Reichenbach, H. §3). There are, however, the following problems:

(a) *The pragmatic approach justifies too much*. Consider the infinite number of variants of the following argument form:

(7′) At least some *F*s have been observed and *n* per cent of them have been *G* (where *F* and *G* are projectible). Therefore, $(n + f(m))$ per cent of the remaining *F*s are also *G*,

where *m* is the number of *F*s that have so far been observed and where the values of the function *f*(*m*), while non-zero up to rather large finite values of *m* (in particular, for values of *m* large enough to represent accurately the number of emeralds that have in fact thus far been observed), converge towards zero as *m* approaches infinity. No less than (7), all these argument forms will find the true relative frequency if any projective form will. Yet, among them are argument forms that have, as an instance, the inference from (1) to:

(8) Fifty per cent of the hitherto unobserved emeralds are green as well.

(Suppose that *n* is the number of emeralds that have in fact been observed and $f(m) = -\frac{1}{2}$, where $m \leq n$ and *f*(*m*) converges to zero thereafter.) We do not (indeed, it would seem positively counterinductive to) regard (1) as evidence that (8). Yet the pragmatist approach provides no justification for regarding (1) as evidence that (3) that is not readily available to justify our regarding (1) as evidence that (8).

(b) *Even were (7) the only argument form that will eventually find the true relative frequency if any projective argument form will, this could not justify our regarding (1) as evidence that (3)*. For there to be a limit to the relative frequency with which observed emeralds are green, there needs to be an infinite number of emeralds to observe. But the quantity of emeralds is, presumably, finite. Once we know there is no limit of the relative frequency, we know that (7)’s boast, that it will find the limit if there is one, is a boast which can be equally made by any argument form whatsoever. Under such circumstances, it is hard to see why we would think that the fact that inference from (1) to (3) is an instance of (7) justifies our regarding (1) as evidence that (3).

3 The inductive approach

Some have suggested another way of circumventing the third difficulty with the presupposition approach. It is not really necessary, in order for us to be justified in regarding (1) as evidence that (3), that we believe (4″). It will suffice that we believe that:

(9) (4″) is true when *F* and *G* predicate properties of physical objects.

This, of course, is not a modest claim; if (3) requires justification, then certainly (9) does as well. But (it has been argued) that justification is readily available from the fact that instances of (7) employing the sort of predicates described in (9) (call these *type 1 instances of (7)*) have been observed to be reliable in the past, and from our belief that:

(10) (4″) is true when *F* and *G* predicate properties of type 1 instances of (7).

And, since (9) and (10) are logically independent, no begging of the question is involved when we appeal to our belief in (10) to justify our belief in (9).

Of course, (10) itself stands as much in need of justification as (9) did. But this, too, is available in so far as we have observed that instances of (7) employing predicable expressions of the sort described in (10) (call these *type 2 instances of (7)*) have been reliable in the past and in so far as we believe that:

(11) (4″) is true when *F* and *G* predicate properties of type 2 instances of (7).

This justification, again, begs no question. And, by appeal to the track record of instances of the sort described in (11) and to the truth of (4″) when *F* and *G* predicate properties of such instances, (11) can in turn be justified. The sequence of ever higher-level justification can be extended indefinitely. But then there are the following

difficulties:

(a) *The proposed justification can never be produced, since it involves an infinite regress of justifying beliefs*. To that extent, the inductive approach would seem to exchange the defect of circularity for another defect just as serious.
(b) *We can just as easily justify regarding (1) as evidence that not-(3)*. Modify (4″) so that it reads, '*It is not the case that n* per cent of the remaining *F*s are *G*'; modify (7) so that its conclusion reads, '*It is not the case that n* per cent of the remaining *F*s are also *G*'. Call the results (4″)* and (7)* respectively. Substitute (4″)* for (4″) and (7)* for (7) in the foregoing, and what you get is every bit as good a justification for regarding (1) as evidence that not-(3). But (1) presumably can not be at the same time evidence that (3) *and* evidence that not-(3).

4 The probabilist approach

This approach differs from those so far canvassed in that, while it also seeks to show how we can be justified in regarding one hypothesis as evidence for the truth of another when the first does not logically imply the second, it does not rely on any appeal to (4″) or the reliability of (7). It appeals instead to the propriety of probabilism.

Probabilism is the view that (a) confidence comes in degrees, and (b) where *Z* is any set of hypotheses closed under truth-functional operations, your degree of confidence assignment to the hypotheses in *Z* (a measure of how confident you are in the truth of each of the hypotheses in *Z*) is open to rational criticism if it fails to satisfy the following axioms of probability: for any propositions *P* and *Q* in *Z*, (i) $\mathrm{prob}(P) \geq 0$; (ii) if *P* is a tautology, $\mathrm{prob}(P) = 1$; (iii) if *P* and *Q* are logically incompatible,

$$\mathrm{prob}(P \vee Q) = \mathrm{prob}(P) + \mathrm{prob}(Q)$$

Suppose you have a degree of confidence assignment to the hypotheses in some such *Z*, where that assignment satisfies the foregoing axioms. For any hypotheses *P* and *Q* in *Z*, let us define your personal odds for *P* as against *Q* as just the ratio your degree of confidence in *P* bears to your degree of confidence in *Q*. Since your degree of confidence that *P* and your degree of confidence that not-*P* must sum to 1 (if they are to satisfy the Kolmogorov axioms), the greater your personal odds are for *P* as against not-*P*, the more confident you are that *P*.

Suppose, now, that for some *P* and *Q* in *Z* of whose truth values you are uncertain, your personal odds for $P \,\&\, Q$ as against $\text{not-}P \,\&\, Q$ are greater than your personal odds for *P* as against not-*P* - that is, your personal odds for *P* as against not-*P* are greater given *Q* than they are otherwise. Then, probabilists hold, you must surely be counted as regarding *Q* as a positive indication that *P* - as a hypothesis that (if true) is evidence that *P*. But if so, it is possible for us to say how we can be justified in regarding (1) as evidence that the conjunction of (1) and (3) is true - even though (1) does not logically imply that conjunction. It suffices that the conjunction logically implies (1). That is because, if *P* logically implies *Q*, then, so long as your degree of confidence assignment satisfies the foregoing axioms and you are certain neither of the truth-value of *P* nor of the truth-value of *Q*, your personal odds for $P \,\&\, Q$ as against $\text{not-}P \,\&\, Q$ must be greater than your personal odds for *P* as against not-*P* - you must regard *Q* as evidence that *P* (see Probability theory and epistemology). But, again, there are difficulties.

(a) *Probabilism, and its doctrine about the relation between personal odds and the assessment of evidential relevance, are controversial*. Critics have argued that confidence does not (nor is there epistemic reason why it should) always come in degrees; that there are difficulties both with the doctrine that degrees of confidence are open to criticism if they do not satisfy the foregoing axioms and with the arguments probabilists have offered in support of that doctrine; that the probabilist's account of the relation between personal odds and assessments of evidential relevance has counterintuitive consequences. Probabilists have powerful responses to all these criticisms (and, indeed, the proliferation of more modest variants of probabilism that do not presuppose that we have, or ought to have, precise degrees of confidence has rendered the first criticism moot), but not everyone is convinced.
(b) *Since (1) is logically implied by the conjunction of (5) and (6), probabilism requires us to regard (1) as evidence that this conjunction - which is equivalent to the claim that all emeralds are grue - is true*. Not that this is obviously a mistake. After all, given (1), both the conjunction of (1) and (3) and the conjunction of (5) and (6) are guaranteed to be right about the emeralds already observed; otherwise they are not. But it does indicate that only because the truth of (1) forecloses the possibility that these conjunctions are wrong about the emeralds that have already been observed are we bound to regard (1) as evidence in favour of the conjunctions. What colour the conjunctions attribute to the emeralds that are still unobserved has nothing to do with it.

(c) *Probabilism does not justify our regarding (1) as evidence that (3)*. (3) does not logically imply (1). So, even granting the propriety of what probabilists say about evidence (that is, even granting that probabilists have succeeded in showing how we can in some cases be justified in regarding one hypothesis as evidence for another it does not logically imply - no mean feat), probabilism provides no justification for regarding (1) as evidence that (3).

5 The reflective-equilibrium approach

The reflective-equilibrium approach, like the probabilist one, also renounces any appeal to (4″). Advocates of this approach argue that the only way to go about justifying our regarding (1) as evidence that (3), is to appeal to some argument form such as (7). But, convinced that there is no cogent non-question-begging way to argue for the reliability of such an argument form, the advocates of the reflective-equilibrium approach hold that the best we can do to justify our appeal to the argument form is to show that each of its instances is acceptable to us, that each of its instances is such that we are prepared on reflection to regard its premise as evidence that its conclusion is true. That is, the best we can do to justify our regarding (1) as evidence that (3) is to show that the inference from (1) to (3) is an instance of an argument form with respect to which we are in *reflective equilibrium*. But there are the following worries:

(a) *There is nothing in the procedure for achieving reflective equilibrium with respect to an argument form that offers even the slightest guarantee that the argument form is not (unknown to us) entirely fallacious*. Consider a person who regards this coin's landing tails on the first ten tosses as evidence that it will land heads on the next. This person is, of course, quite mistaken. If the coin is fair, its performance on its first ten tosses is no evidence at all of how it will perform on the eleventh: the outcomes of the tosses are independent. But the fact that this person is mistaken is entirely compatible with their being able to achieve reflective equilibrium with respect to the following (fallacious) argument form:

(12) *C* is a fair coin that has landed heads on its first ten tosses.
Therefore, *C* will land tails on its eleventh toss.

The procedure for achieving reflective equilibrium with respect to an argument does not guarantee that the form is not fallacious - only that those who have achieved this reflective equilibrium do not *regard* the form as fallacious. It would thus be disappointing if the best we can do to justify our regarding (1) as evidence that (3) is to engage in such a procedure. One would have hoped we could do (and one might even think a true justification would require us to do) something better.

(b) *Achieving reflective equilibrium requires too much*. The task of furnishing the sort of justification called for by the reflective-equilibrium approach is an extremely ambitious one, involving the crafting of an argument form (of which the argument from (1) to (3) is an instance) that can survive an extensive test of its adequacy to our intuitions about its instances. A fully adequate argument form has yet to be articulated. But if this task of justification is thus not one we have yet actually accomplished, then whatever it is that justifies our regarding (1) as evidence that (3), it cannot be that we have placed our so regarding (1) in reflective equilibrium with an argument form and its instances. We have done no such thing.

6 The naturalist approach

All the approaches canvassed so far assume that what justifies our regarding (1) as evidence that (3) is that there is available to us some sort of argument in favour of the claim that (1) is evidence that (3). But to some philosophers, this idea seems misguided. Humans constitute just one species of creatures whose survival depends upon their following inductive practices. The rest of these species are quite incapable of understanding *any* argument in support of such practices; no arguments are available to *them*. Yet (these philosophers maintain) it still seems perfectly appropriate to say of a dog that it is justified in regarding its master's picking up its leash as evidence that it is about to go for a walk. Thus (they conclude) the availability of argument is not necessary for justification: to the dog, there is no argument available.

Nor (they continue) does the mere availability of argument seem sufficient for justification. Consider a person who sets no store by our canons of good argument and only regards (1) as evidence that (3) because they like the sound

of the utterance, '(1) is evidence that (3)'. It is compatible with this that there is just as good an argument available to them as there is to you or me for regarding (1) as evidence that (3). But surely (these philosophers maintain), given that they set no store by that argument, we can hardly want to call their regarding (1) as evidence that (3) *justified*.

What qualifies the dog's state of opinion as justified, but not the person's, has nothing to do with the availability of argument; it has to do, instead, with what caused their respective states of opinion. Whether a state of opinion is justified is determined entirely by features of the process that is causally responsible for being in that state - for example, the features (such as reliability in producing true beliefs) that make these processes conducive to our survival. What the holder of the state of opinion is in a position to *say* about the features of the process that causes their belief is, in general, beside the point (see Reliabilism; Internalism and externalism in epistemology). But, again, there are worries.

(a) *The claim that being able to offer a justification for one's opinion is neither necessary nor sufficient to render the opinion justified is difficult to square with our practice*. Were the claim true, it is hard to see why we would care in the least whether we could offer even the most meagre justification for any of our beliefs.
(b) *If it is the aetiology of our opinions that determines whether they are justified, then we are ill-placed to tell, with respect to most of them, why they are worth holding*. We are not usually in any position to say by exactly what processes our opinions have been caused - consider, for example, beliefs in theories, beliefs held since childhood - let alone whether the processes that have caused them possess whatever features are necessary and sufficient for them to confer justification on the beliefs they produce. But, if so, by the lights of the naturalist approach it would seem that we are usually in no position to say that our opinions are justified - that they are any more worth holding than those we would acquire by reading tea leaves.

7 The meaning approach

All the foregoing approaches presuppose that our regarding (1) as evidence that (3) *requires* some substantive justification. But not all philosophers think that this is so. Some have argued that no justification is called for: it is simply part of what we *mean* by 'evidence' that (1) constitutes evidence that (3). This approach has some difficulties:

(a) *To use a meaning claim in this way seems dogmatic*. In the wake of work (by Quine and others) to the effect that there is no bona fide distinction to be drawn between true claims that are true by virtue of meaning and those that are not (see Analyticity), this attempt to settle an apparently open question by appeal to meaning claims looks to be just a form of dogmatism.
(b) *The meaning claim, even if granted, is ineffective*. It may well be that, for at least some people, it is part of what they mean by 'evidence' that the failure of a fair coin to land heads on one toss is evidence that it will land heads on the next one. But that can hardly show that they are justified in regarding the coin's having landed tails the first time as evidence that it will land heads the next. And, if so, the fact (if it is a fact) that it is part of what *we* mean by 'evidence' that (1) constitutes evidence that (3) cannot show that *we* are justified in regarding (1) as evidence that (3).

8 The ordinary-language approach

But, to some, there is still something perverse in the claim that our regarding (1) as evidence that (3) requires justification. To say that an opinion is one that requires justification is to say that it is an opinion that a person ought not to hold unless they can produce justification. But we cannot seriously mean that until they can successfully produce the requisite justification for so regarding (1), a person should not regard (1) as evidence that (3). This is a condition that no ordinary person - and possibly no philosopher - can satisfy. Anyone who insisted upon its being met - anyone who habitually demanded of any person who asserts that (1) is evidence that (3) that they either justify that claim or withdraw it - would rightly be counted as mad. But there is the following problem:

The ordinary-language approach appears to mistake the context in which we mean to be saying that our so regarding (1) requires justification. Granted, in an ordinary context, it would indeed be perverse to say of a person who regarding (1) as evidence that (3), that their so regarding (1) requires justification. But it does not follow that it is a perverse thing to say when our purposes are philosophical - that is, when we are in the context of

philosophical inquiry. Yet, it is precisely for such purposes, and in such a context, that we mean to be asking what, if anything, justifies our regarding (1) as evidence that (3). It is precisely for such purposes, and in such a context, that we mean to be saying that our so regarding (1) requires justification.

9 The Wittgensteinian approach

Of course, this reply to the advocates of the ordinary-language approach presupposes that language can be freely wrested from the ordinary context in which it is used and thrust into a philosophical context without cost. But some (and Wittgenstein in particular) have argued that it cannot.

To appreciate how their reasoning might apply to the case at hand, suppose we say, with the critics of the ordinary-language approach, that, while our regarding (1) as evidence that (3) requires no justification for ordinary purposes in an ordinary context, it does require justification for philosophical purposes in a philosophical context. How are we to tell whether an adequate justification has been produced? For example, are justifications that contain circular arguments or infinite regresses admissible? Were we free to consult the standards of justification we employ in ordinary circumstances, we could easily see that the answer to this last question is ‘no’. But we are not. The whole point of our reply to the advocates of the ordinary-language approach is to say that those standards of justification - and their verdict that our regarding (1) as evidence that (3) does not require justification - are to be set aside in a philosophical context. But what other standards do we have? None.

We are thus faced with a dilemma. The standards of justification we employ in ordinary contexts forbid us to say that our regarding (1) as evidence that (3) requires justification. But if we turn our backs on the standards of justification we use in ordinary contexts, as it is imagined we do in the philosophical context, we deprive ourselves of any source of insight into what would *count* as justifying our regarding (1) as evidence that (3). What sustains philosophical attempts to say what justifies our regarding (1) as evidence that (3) thus looks like nothing more than a failure to face up to this dilemma. There is one major difficulty:

Wittgensteinian doctrines are very controversial. This is all the more true for the fact that the morals drawn from them so often place apparently respectable philosophical enterprises in a poor light. In this case, it may be difficult to see how a sophisticated, orderly philosophical inquiry of more than two centuries’ duration could possibly be the intellectual sham the advocates of the Wittgensteinian approach make it appear. Of course, it is part and parcel of Wittgensteinian doctrine that one cannot expect those who have toiled in the philosophical enterprise criticized to find it easy to see, even once it is pointed out to them, the way in which that enterprise is fundamentally ill-conceived: language bewitches us. But this consideration has done little to convince the many philosophers who (rightly or wrongly) see Wittgensteinian approaches to philosophical problems (such as this one) as fundamentally anti-philosophical, as attempts to hold philosophy to such standards as would make philosophy impossible.

See also: Confirmation theory; Inductive inference; Meaning, Indian theories of

MARK KAPLAN

References and further reading

Black, M. (1954) *Problems of Analysis*, Ithaca, NY: Cornell University Press.(Prominent defence of the inductive approach. Chapter 11 is especially relevant.)

Edwards, P. (1949) ‘Russell’s Doubts about Induction’, *Mind* 68: 141-63.(Prominent example of the ordinary-language approach.)

Goodman, N. (1955) *Fact, Fiction and Forecast*, Indianapolis, IN: Hackett Publishing Company, 3rd edn, 1979. (The classic presentation of both the reflective-equilibrium approach and the new riddle of induction.)

Hume, D. (1739) *A Treatise of Human Nature*, 2nd edn, ed. L.A. Selby-Bigge, revised P.H. Nidditch, Oxford: Clarendon Press, 1978.(The problem of induction is first introduced, and the presupposition approach to its solution is first put forward and criticized.)

Kaplan, M. (1996) *Decision Theory as Philosophy*, New York: Cambridge University Press.(Chapters 1 and 2 provide an introduction to - and defence of a modest variant on - probabilism and a critical discussion of the probabilist approach.)

Kitcher, P. (1992) ‘The Naturalists Return’, *The Philosophical Review* 101: 53-114.(Survey and defence of the naturalist approach.)

Skyrms, B. (1975) *Choice and Chance*, Encino, CA: Wadsworth, 2nd edn, 1985. (Chapter 2 provides an excellent

brief survey of the major approaches to the problem of induction.)
Salmon, W. (1963) 'Inductive Inference', in B. Baumrin (ed.) *Philosophy of Science: The Delaware Seminar*, New York: Interscience Publishers, 353-70.(Defence of the pragmatic approach.)
Strawson, P.F. (1952) *Introduction to Logical Theory*, New York: John Wiley & Sons.(A defence of the meaning approach is provided in chapter 9.)
Swinburne, R. (ed.) (1974) *The Justification of Induction*, Oxford: Oxford University Press.(Collection of important papers on the topic, particularly on the pragmatic, inductive and ordinary-language approaches.)
Travis, C. (1991) 'Annals of Analysis', *Mind* 100: 237-64.(Section 1 contains a defence of Wittgensteinian misgivings about our capacity to evaluate the truth and falsity of what is asserted outside our ordinary contexts.)
Wittgenstein, L. (1953) *Philosophical Investigations*, New York: Macmillan, 3rd edn, 1958.(The source of the line of thought behind the Wittgensteinian approach.)

Inductive definitions and proofs

An inductive definition *of a predicate R characterizes the Rs as the smallest class which satisfies a* basis clause *of the form* $(\beta(x) \rightarrow Rx)$*, telling us that certain things satisfy* R, *together with one or more* closure clauses *of the form* $(\Phi(x, R) \rightarrow Rx)$*, which tell us that, if certain other things satisfy R, x satisfies R as well. 'Smallest' here means that the class of* Rs *is included in every other class which satisfies the basis and closure clauses.*

Inductive definitions are useful because of inductive proofs. To show that every R has property P, show that the class of Rs that have P satisfies the basis and closure clauses.

The closure clauses tell us that if certain things satisfy R, x satisfies R as well. Thus satisfaction of the condition $\Phi(x, R)$ *should be ensured by positive information to the effect that certain things satisfy R, and not also require negative informative that certain things fail to satisfy R. In other words, the condition* $\Phi(x, R)$ *should be* monotone*, so that, if* $R \subseteq S$ *and* $\Phi(x, R)$*, then* $\Phi(x, S)$*; otherwise, we would have no assurance of the existence of a smallest class satisfying the basis and closure conditions.*

While inductive definitions can take many forms, they have been studied most usefully in the special case in which the basis and closure clauses are formulated within the predicate calculus. Initiated by Yiannis Moschovakis, the study of such definitions has yielded an especially rich and elegant theory.

1 Inductive definitions in general

Inductive definitions occur throughout mathematics. For example, one aspect of the reduction of classical mathematics to set theory has been to treat natural numbers as sets, identifying a natural number with the set of its predecessors, so that they comprise the smallest class such that:

$\emptyset$ is a natural number
If $x = y \cup \{y\}$, for some natural number *y*, *x* is a natural number.

Having this inductive definition, we can give proofs by *mathematical induction*. To show that every number has property *P*, show that $\emptyset$ has the property and that, if *y* is a natural number with property *P*, so is $y \cup \{y\}$.

One inductive definition begets others. We define 'is a sum of' as the smallest ternary relation such that:

If *x* is a natural number, *x* is a sum of *x* and $\emptyset$.
If *u* is a sum of *y* and *v*, $x = u \cup \{u\}$ and $z = v \cup \{v\}$, then *x* is a sum of *y* and *z*.

An inductive proof shows that this stipulation uniquely determines a binary operation on the natural numbers.

We can extend arithmetical methods beyond the finite, identifying a limit ordinal with the set of its predecessors, so that the ordinals are the smallest class such that:

$\emptyset$ is an ordinal.
If $x = y \cup \{y\}$, for some ordinal *y*, then *x* is an ordinal.
If the set *x* is a union of ordinals, *x* is an ordinal.

Again one inductive definition leads to another. Given a function *G*, to define a function *F* on the ordinals such that:

$F(\alpha) = G(\alpha, F \mid \alpha)$

(where $F \mid \alpha = \{< \beta, F(\beta) >: \beta < \alpha\}$), take the smallest class of ordered pairs such that:

$< 0, G(0, \emptyset) >$ is in *F*.
If *h* is a function defined on the ordinals $< \alpha$ such that $< \beta, h \mid \beta >$ is in *F*, for each $\beta < \alpha$, then $G(\alpha, h)$ is in *F*.

Speaking abstractly, an inductive definition describes the smallest relation such that:

If $\beta(x_1, \ldots, x_n)$, then $Rx_1 \ldots, x_n$.
If $\Phi_1(x_1, \ldots, x_n, R)$, then $Rx_1 \ldots x_n$.
. . . .
If $\Phi_m(x_1, \ldots, x_n, R)$, then $Rx_1 \ldots x_n$.

Here β is an *n*-ary relation, and the conditions $\Phi_1, \ldots, \Phi_m$ are monotone relations between *n*-tuples and *n*-ary relations.

Since the disjunction of two monotone conditions is a monotone condition, we can consolidate the basis and inductive clauses thus:

If $\beta(x_1, \ldots, x_n)$ or $\Phi_1(x_1, \ldots, x_n, R)$ or ... or $\Phi_m(x_1, \ldots, x_n, R)$, then $Rx_1, \ldots, x_n$.

Consequently, there will be no loss of generality if we only consider inductive definitions that describe the smallest class such that:

If $\Phi(x_1, \ldots, x_n, R)$, then $Rx_1, \ldots, x_n$,

where Φ is monotone.

Setting the example of the ordinals aside, let us restrict our attention to situations in which there is some set *A* such that, for each $x_1, \ldots, x_n$ and *S*, if $\Phi_1(x_1, \ldots, x_n, S)$, then each $x_i \in A$. Then, we shall not have to be concerned about the existence conditions for proper classes.

It is only because of the requirement that Φ is monotone that we can be confident that there is a smallest class which satisfies the closure clause, as we can see if we try to find the smallest class such that:

If either $x = 0$ and $\neg R1$ or $x = 1$ and $\neg R0$, then Rx.

Because of the monotony requirement, however, the intersection of all the sets which satisfy the closure clause satisfies the closure clause. It also satisfies the converse of the closure clause, so that (setting $n = 1$):

$(\forall x)(Rx \leftrightarrow \Phi(x, R))$.

We have inductive proofs: if we know $(\forall x)((Rx \wedge \Phi(x, P)) \rightarrow Px)$, we can conclude $(\forall x)(Rx \rightarrow Px)$.

Inductively defined classes are naturally arrayed in a well-ordered sequence of layers, for we can define the level of a member of *R* to be the number of applications of the closure clause that are needed to capture it. More precisely, we define, for each ordinal α, R_α to be $\{x{:}\, \Phi(x, \cup_{\beta<\alpha} R_\beta)\}$. Then the R_αs form a nonstrictly increasing sequence (that is, if $\beta < \alpha$, then $R_\beta \subseteq R_\alpha$); their union is *R*. Because we have stipulated that the R_αs are all contained within a set *A*, the sequence cannot be strictly increasing, since, if each R_α introduced new members of *R*, there would be as many members of *R* as there are ordinals. Consequently, there must exist an ordinal α with cardinality at most the cardinality of *A* such that $R_\alpha = R_{\alpha+1} = R$. The least such α is the *height* of the inductive definition.

2 First-order positive inductive definitions

An occurrence of a predicate *R* in a formula of the predicate calculus without '$\rightarrow$' or '$\leftrightarrow$' is *positive* if it occurs within the scope of an even number of negation signs. A first-order formula $\Phi(x_1, \ldots, x_n, R)$ in which *R* occurs only positively describes a monotone condition. *First-order positive inductive definitions*, in which the antecedent of the closure clause is a first-order formula in which the defined term occurs only positively, have been extensively and fruitfully studied.

Let A be a model for a language $\mathcal{L}$ for the first-order predicate calculus, and let $\mathcal{L}_\mathrm{A}$ be the language derived from $\mathcal{L}$ by adding a new constant $\underline{a}$ for each element *a* of the universe *A* of A. Apart from the predicate being defined, our first-order positive inductive definitions will consist of symbols of $\mathcal{L}_\mathrm{A}$.

We can explicitly describe how an inductively defined set is built. Suppose *C* is inductively defined as the smallest set such that:

If $\Phi(x, R)$, then Rx.

Then *a* is in *C* if and only if $R\underline{a}$ is derivable from the axiom:

$(\forall x)(\Phi(x, R) \rightarrow Rx)$

in A-logic, that is to say, $R\underline{a}$ is derivable from the axiom together with atomic and negated atomic sentences true in A by first-order logic supplemented by the A-rule: from $\psi(\underline{a})$, for each *a* in *A*, to infer $(\forall x)\psi(x)$.

We shall restrict our attention to cases in which A is *acceptable*, that is, A contains an isomorphic copy of the natural number system and A can describe the formation of finite sequences. Acceptability of A will make it possible for $\mathcal{L}_{\mathrm{A}}$ to describe its own syntax. We also suppose that $\mathcal{L}$ is built up from a finite vocabulary. Similar but more complicated analogues to the results described here hold without these restrictions.

It turns out that, to get a satisfactory theory, we need to look, not at the separate behaviour of individual inductive definitions but at their combinations. A relation *C* is *inductively definable* (or simply *inductive*) over A if *C* can be defined in a language derived from $\mathcal{L}_{\mathrm{A}}$ by adding one or more inductively defined predicates; in the formula defining *C* the new predicates are allowed to occur only positively. More concisely, *C* is inductively definable over A if and only if there exist an $n + m$-ary relation *D* and members $b_1, \ldots, b_m$ of the universe *A* of A such that *D* is defined by a first-order positive inductive definition and *C* is equal to $< a_1, \ldots, a_n >: < a_1, \ldots, a_n, b_1 \ldots, b_m > \in D$. Showing this equivalence is a matter of seeing how several inductive definitions can be consolidated.

An inductively defined set whose complement is also inductively defined is *hyperelementary*.

A set *C* is inductive if and only if there is a finitely axiomatized theory Δ and formula $\psi(x)$ in a language extending $\mathcal{L}_{\mathrm{A}}$ such that, for any *a*, *a* is an element of *C* if and only if $\psi(\underline{a})$ is derivable from Δ in A-logic. Just in case *C* is hyperelementary, we can arrange things so that, if *a* is in *C* then $\psi(\underline{a})$ is derivable from Δ, whereas if *a* is not in *C*, $\neg\psi(\underline{a})$ is derivable.

The results cited in the last paragraph would also obtain if the word 'inductive' were replaced by 'recursively enumerable', 'hyperelementary' by 'recursive' and 'derivable in A-logic' by 'derivable in first-order logic'. This observation illustrates a far-reaching analogy between structural properties of inductively definable and recursively enumerable sets.

Another example is the following: inductive sets, like recursively enumerable sets, enjoy the *reduction property*, that is, for any inductive sets *B* and *C*, there exist nonoverlapping inductive sets $B' \subseteq B$ and $C' \subseteq C$ with $B' \cup C' = B \cup C$. We get them by comparing the stages in the relevant inductive definitions at which the members of *B* and *C* are introduced, putting a member of $B \cap C$ into B' just in case it is put into *B* before being placed into *C*.

Again, there are inductively defined sets *C* which are *complete*, meaning that, for each inductive set *D*, there is a hyperelementary function *F* such that, for any *a*, $a \in D$ if and only if $F(a) \in C$. The best-known complete set, at least among philosophers, is the minimal fixed point of Saul Kripke's theory of truth. The theory of truth is robust enough to encode every inductive definition.

There is a version of the recursion theorem for inductive sets. Also, one can do priority arguments (see Sacks 1990).

The height of the monotone operator which generates an inductively defined set is a measure of the complexity of the set. Indeed, the principal ideas of the theory of inductive sets originate in S.C. Kleene's investigations of how to extend the arithmetical hierarchy into the transfinite. The *order* of A, $o(\mathrm{A})$, is the supremum of the heights of all first-order positive inductive definitions over A. This supremum is reached by the complete inductive sets, among others. A subset of *A* is hyperelementary if and only if it is defined by a first-order positive inductive definition whose height is less than $o(\mathrm{A})$.

3 $\prod_1^1$ and inductively defined sets

An element *a* of *A* will be in the smallest class which satisfies the closure clause $(\Phi(x, R) \rightarrow Rx)$ just in case the sentence $(\forall R)((\forall x)(\Phi(x, R \rightarrow Rx) \rightarrow R\underline{a})$ is true. It follows that every inductive set is $\prod_1^1$,that is, it is

definable by a formula consisting of initial second-order universal quantifiers followed by a first-order formula.

If *A* is countable, the converse holds. $(\forall R)\psi(R,\underline{a})$ is true if and only if the sentence $\psi(R,\underline{a})$ is a member of the smallest class of sentences which:

contains the atomic and negated atomic sentences true in A;
contains the theorems of first-order logic;
contains θ whenever it contains $(\chi \rightarrow \theta)$ and χ; and
contains $(\forall v)\chi(x)$ whenever it contains each $\chi(\underline{a})$.

Because A is acceptable, so that it can represent its own syntax, this can be written as a first-order positive inductive definition.

One proves this by proving a completeness theorem, showing that, for *A* countable, $(\forall R)\theta(R)$ is true if and only if $\theta(R)$ is derivable in A-logic.

4 Inductive and admissible sets

To get a picture of the place of inductively defined sets within the universe of set theory, we look at sets that can be built up by taking the elements of *A* as basic building blocks or *urelements*:

$V_0 = A$.
$V_{\alpha+1}$ = the set of subsets of V_α.
$V_\lambda = \cup_{\alpha<\lambda} V_\alpha$, for λ a limit.

The *constructible sets* form a subclass of the V_αs:

$L_0 = A$.
$L_{\alpha+1}$ = the set of subsets of L_α which are definable in the language derived from $\mathcal{L}_\mathrm{A}$ by adding a predicate '*U*', whose extension is *A*, a predicate '$\in$', whose extension is the elementhood relation, restricted to L_α, and names for all the members of L_α.
$L_\lambda = \cup_{\alpha<\lambda} L_\alpha$, for λ a limit.

We are interested in the initial stages of the constructible hierarchy. First, let us introduce a little more notation. The *bounded formulas* are those in which every universal quantifier takes the form $(\forall x)(x \in t \rightarrow \ldots)$, and every existential quantifier takes the form $(\exists x)(x \in t \wedge \ldots)$. The Σ_1 *formulas* are those obtained by prefixing one or more existential quantifiers to a bounded formula.

One can show that $L_{0(\mathrm{A})}$ constitutes the smallest set which:

(1) contains *A* and the sets definable in A;
(2) contains x, y and the union of *x* whenever it contains *x* and *y*;
(3) satisfies the universal closure of the separation axiom schema, $(\exists y)(\forall z)(z \in y \leftrightarrow (z \in x \wedge \phi(z)))$, restricted to bounded formulas ϕ; and
(4) satisfies the universal closure of $(\forall x)(x \in t \rightarrow (\exists z)\phi(x,z)) \rightarrow (\exists y)(\forall x)(x \in t \rightarrow (\exists z)(z \in y \wedge (x,z))$, for ϕ bounded.

Subsets of the class of $V_\alpha s$ which satisfy these four conditions are said to be *admissible* above A. The first-order axioms which describe, in a natural way, a well-founded set which satisfies conditions (2) through (4) are called 'KPU' (for 'Kripke-Platek set theory with urelements', after the thinkers who first studied admissible sets).

The Barwise-Moschovakis-Gandy theorem states that a subset of *A* is inductive if and only if it is definable by a Σ_1 formula on $L_{o(\mathrm{A})}$. While the proof is complicated, it is possible to give a summary. To show that every inductively definable set is Σ_1 definable on $L_{o(\mathrm{A})}$, show that, if there is a proof in A-logic of a given sentence, then such a proof exists in $L_{o(\mathrm{A})}$. For the converse, show that, if *S* is Σ_1 definable, there is a formula σ such that $a \in S$ if and only if $\sigma(\underline{a})$ is derivable from KPU in A-logic.

It follows that a set is hyperelementary over A if and only if it is an element of $L_{o(\mathrm{A})}$.

The Barwise-Moschovakis-Gandy theorem explains the parallels between recursively enumerable and inductive sets. Recursively enumerable = Σ_1 definable in the smallest set admissible above $\emptyset$. Inductive over $\mathrm{A} = \Sigma_1$ definable in the smallest set admissible above A.

Admissible sets have found widespread applications to the study of infinitary logic, but here it is possible to do no more than refer to Barwise (1975).

The results cited here are standard, and proofs can be found in Moschovakis (1974) and Barwise (1975).

See also: Constructible universe; Definition; Infinitary logics; Logical and mathematical terms, glossary of; Turing reducibility and Turing degrees

VANN McGEE

References and further reading

Barwise, J. (1975) *Admissible Sets and Structures: An Approach to Definability Theory*, Berlin and New York: Springer.(A clear and readable development of the theory of admissible sets, with particular attention to infinitary logic.)

Moschovakis, Y.N. (1974) *Elementary Induction on Abstract Structures*, Amsterdam: North Holland.(This is the fundamental work on inductively defined sets. Lucid and readable.)

Sacks, G.E. (1990) *Higher Recursion Theory*, New York: Springer-Verlag.(Uses the methods of recursion theory, particularly priority arguments, to describe the structure of inductively defined sets.)

Inductive inference

According to a long tradition, an inductive inference is an inference from a premise of the form 'all observed A *are* B*' to a conclusion of the form 'all* A *are* B*'. Such inferences are not deductively valid, that is, even if the premise is true it is possible that the conclusion is false, since unobserved* A*s may differ from observed ones. Nevertheless, it has been held that the premise can make it reasonable to believe the conclusion, even though it does not guarantee that the conclusion is true.*

It is now generally allowed that there are many other patterns of inference that can also provide reasonable grounds for believing their conclusions, even though their premises do not guarantee the truth of their conclusions. In current usage, it is common to call all such inferences inductive. It has been widely thought that all knowledge of matters of fact that we have not observed must be based on inductive inferences from what we have observed. In particular, all knowledge of the future is, on this view, based on induction.

1 Paradigms of induction

The inference from 'all observed *A* are *B*' to 'all *A* are *B*', which was once taken to be the pattern for all inductive inference, is called (universal) *inductive generalization* or *enumerative induction*. A standard example is the inference from all observed ravens being black to all ravens being black. The fallibility of such inference is illustrated by the fact that, although all the swans seen by Europeans of the eighteenth century were white, black swans existed in Australia.

Some writers, such as J.S. Mill, argued that inductive generalization is the only legitimate kind of induction. However, as Mill was well aware, others have thought that there are other ways of making inferences from the observed to the unobserved. One method that has played an important role in philosophy of science is the method of hypothesis. In this method, the premises are that (i) hypothesis *H* implies a proposition *E* describing observable phenomena, and (ii) *E* is observed to be true; the conclusion is that *H* is true. In modern discussions this method is often called the hypothetico-deductive method, because *E* is supposed to be deduced from the hypothesis *H*.

The method of hypothesis is not deductively valid, since false hypotheses can have true consequences. Its defenders have argued that its premises can nevertheless make it reasonable to believe its conclusion. Descartes offered an example to support this position: if we find that a message in code makes sense when for each letter we substitute the following letter (B for A, C for B, and so on), we will be practically certain that this gives the true meaning of the message, especially if the message contains many words.

The method of hypothesis licenses conclusions that can never be reached by inductive generalization. For example, from the fact that observable phenomena are as they would be if matter consists of atoms, one may infer by the method of hypothesis that matter does consist of atoms. This conclusion could be reached by inductive generalization only if one had observed instances of matter composed of atoms, something that had not been done at the time that most scientists accepted the atomic theory of matter.

Other commonly recognized types of inductive inference are: (1) *statistical inductive generalization*, in which the premise is that *x* per cent of observed *A*s have been *B* and the conclusion is that about *x* per cent of all *A*s are *B*; (2) *predictive inference*, in which the premises are that *x* per cent of observed *A*s have been *B* and that *a* is *A*, the conclusion being that *a* is *B*; (3) *direct inference*, in which the premises are that *x* per cent of all *A*s are *B* and that *a* is *A*, the conclusion being that *a* is *B*; and (4) *inference by analogy*, in which the premises are that certain individuals have properties $F_1, \ldots, F_n$ and *a* also has $F_1, \ldots F_{n-1}$, the conclusion being that *a* also has F_n.

2 Induction in practice

It is generally allowed that the cogency of an inductive inference is greater the more observations have been made and the more varied these observations have been. Mill also stressed that the reliability of an inductive generalization is greatly increased by finding that other similar generalizations hold up; conversely, he maintained that our knowledge of the variability of bird-colouring undermined the inductive generalization to the conclusion that all ravens are black. But the relevance of such considerations means that inductive inferences are not adequately represented by the paradigm forms listed above, since relevant information ought to be included in the

premises.

Even if we added to the paradigm forms the sort of information just mentioned, they would still not adequately represent the inductive inferences that are made in everyday life and in science. For example, consider the inductive generalization to the conclusion that all ravens are black. The premise was that all observed ravens have been black. But observed by whom? If we say observed by anyone, then we cannot be sure that the premise is correct, and this ought to affect our confidence in the conclusion, though the argument does not provide for that. If we say observed by me, then the premise leaves out the very relevant information of what I know from the testimony of others. Further, even when the premise is limited to my own observation, I cannot be sure it is true; I could very well have observed a white raven and mistakenly inferred from the fact that it was white that it was not a raven. Thus my confidence in the conclusion ought to depend on how closely I have examined both what I take to be ravens and what I take to be non-black things, though this is not reflected in the premise.

Another example: suppose we have weighed an object three times and obtained measurements of 4.9, 5.0 and 5.1 grams. From these results we might infer that the true weight of the object is between 4.7 and 5.3 grams. If this is to fit one of the standard forms, it will have to be the method of hypothesis. But the hypothesis that the true weight is between 4.7 and 5.3 grams does not entail that the weighings will give the results they did. In modern presentations of the hypothetico-deductive method it is commonly said that the premises include not only the hypothesis itself but also initial conditions, such as that the object was weighed three times, and auxiliary hypotheses, such as that the balance is accurate; but even with these additional premises the evidence does not follow deductively.

Since tests of hypotheses in science typically involve measurement, the preceding example is sufficient to show that science does not have much use for a method that is truly ‘hypothetico-deductive’. But if the evidence only follows from the hypothesis with some probability, then it makes a difference what that probability is, though the method of hypothesis as usually presented does not include this information in the premises. As we will see below, it also makes a difference how probable the conclusion is on competing hypotheses, another relevant consideration that is not incorporated in the premises of the method of hypothesis as standardly conceived. Finally, our background information almost always gives us at least some relevant information; for example, we usually have some approximate idea what an object should weigh, and measurements that diverge too far from this will be taken to show that the balance is faulty. If the object that was weighed at 4.9, 5.0 and 5.1 grams was a loaded truck, we will conclude that there is something seriously wrong with the measurement process. The method of hypothesis does not provide a way to incorporate such relevant information.

3 Cogency

The premises of an inductive argument may provide more or less support for the conclusion. We would like to be able to identify the factors that determine the strength of this support. Some platitudes to this effect were mentioned above, but they have several defects. In particular, despite their vagueness, they are not true in general. For example, increasing the number of *A*s that have been observed to be *B* does not always make it more probable that all *A* are *B*; thus a person who died at age 149 is a person who died before age 150, but evidence that there is such a person would *reduce* the probability that all people die before age 150. In any case, those platitudes are limited to the paradigm forms of inductive inference, which we have seen inadequate to represent actual inductive inferences. For these reasons, many contemporary theorists have looked for a more rigorous yet flexible framework for discussing the degree to which inductive premises support their conclusions. Probability theory has often been seen as providing such a framework.

Let *H* be a hypothesis and *E* some evidence. Let $\mathrm{P}(H)$ be the probability that *H* is true given the other information that we have besides *E*. Then the probability that we should give to *H* after learning *E* is the probability of *H* conditional on *E*, which we write as $\mathrm{P}(H \mid E)$. A theorem of probability, called Bayes’ theorem (see Probability, interpretations of §5), says the following (where $\bar{H}$ means *H* is false):

$$\mathrm{P}(H \mid E) = \frac{\mathrm{P}(E|H)}{\mathrm{P}(E|H)\mathrm{P}(H)+\mathrm{P}(E|\bar{H})\mathrm{P}(\bar{H})}\mathrm{P}(H)$$

We say that *E confirms H* if $\mathrm{P}(H \mid E) > \mathrm{P}(H)$. Assuming $\mathrm{P}(H) > 0$, *E* confirms *H* just in case the fraction on the right hand side of the above equation is greater than 1. Assuming $\mathrm{P}(H) < 1$, this condition will hold just in

case $\mathrm{P}(E \mid H) > \mathrm{P}(E \mid \bar{H})$. Thus for evidence to confirm a hypothesis, it is not necessary that the evidence be entailed by the hypothesis, or even that the evidence be very probable given the hypothesis. Evidence that is quite improbable supposing the hypothesis to be true will confirm that hypothesis if the evidence is even less likely on the supposition that the hypothesis is false.

Another implication of the above equation is that even if evidence *E* does strongly confirm hypothesis *H*, it does not follow that the hypothesis has a high probability of being true, for its probability given other information may be quite low; that is, $\mathrm{P}(H \mid E)/\mathrm{P}(H)$ may be large and yet $\mathrm{P}(H \mid E)$ small, because $\mathrm{P}(H)$ is small. For example, when *H* is the hypothesis that a loaded truck weighs about 5 grams, $\mathrm{P}(H)$ is infinitesimal.

Philosophers often refer to the method of hypothesis as ‘inference to the best explanation’. This terminology has the merit of recording the fact that the inference depends not just on the relation between hypothesis and evidence but also on how other hypotheses relate to the evidence. However, it is wrong to suggest that *explanation* plays a fundamental role here; what is important in the connection between evidence and hypothesis is the probability of the evidence given the hypothesis, and this probability is not a measure of the degree to which the hypothesis explains the evidence (see Explanation; Inference to the best explanation). Furthermore, the terminology misleadingly suggests that the cogency of an inference is determined by the relation between the various hypotheses and the evidence, when in fact it depends also on the prior probability of the hypotheses.

4 Inference

Inductive inference has traditionally been understood as inference in the usual sense: on the basis of premises that are categorically accepted one comes to categorically accept another statement, the conclusion. For writers such as Bacon, Whewell and Mill, this conception was unproblematic. They thought that inductive inference could provide practical certainty on substantive matters of science and everyday life, and that where such certainty was lacking we should withhold assent. However, various factors, in particular the failure of such well-supported scientific theories as Newtonian mechanics, have convinced contemporary writers that substantive inductive conclusions always have some uncertainty, the best-supported scientific theories not excepted. This presents a dilemma: either we accept only those conclusions that are certain, in which case it seems we can accept very little and there will be almost no inductive inference, or else we accept conclusions that might be wrong.

Some writers on probability and induction, notably Carnap and Jeffrey, have embraced the first horn of this dilemma. They hold that induction should be conceived, not as a process by which we pass from some accepted statements to others, but rather as a process by which we assign probabilities to various hypotheses in the light of our evidence. On this view, we ought virtually never to make inductive inferences, as these have been traditionally conceived.

This rejection of inductive inference presupposes that acceptance of a hypothesis involves treating it as if it were certainly true. Some writers have argued that we can preserve the legitimacy of inductive inference by abandoning the assumption that someone who accepts a hypothesis must treat it as certainly true. A popular suggestion is that accepting a hypothesis is merely a matter of being highly confident of it, for example, of giving it a probability greater than some threshold. One objection to this view is that the set of propositions that have high probability is not consistent. For example, in a large unbiased lottery, the probability that any particular ticket will not win is high, though we also know that some ticket will win. Thus if we accepted every proposition with high probability we would accept an inconsistent set of propositions. This is called the ‘lottery paradox’ (see Paradoxes, epistemic). Another objection to equating acceptance with high probability is that informativeness is a reason for accepting a hypothesis but is not a reason for giving the hypothesis high probability.

Several writers have suggested that we ought to think of acceptance as a risky decision and use decision theory to evaluate when acceptance is rational (see Decision and game theory §2). The thought is that acceptance has a certain cognitive utility, which is greater if the proposition accepted is true than if it is false; acceptance of a hypothesis is rational, at least so far as cognitive goals are concerned, if it maximizes expected cognitive utility. The cognitive utility of accepting a true hypothesis is higher the more informative the hypothesis is, and thus this account provides a place for both probability and informativeness in determining the rationality of inductive inference. Some versions of this approach take acceptance to imply certainty while others do not.

See also: Chinese philosophy §§1-2; Confirmation theory; Confucian philosophy, Chinese §1; Induction,

epistemic issues in; Learning; Rationality of belief; Reichenbach, H. §3; Science, 19th-century philosophy of; Scientific method; Statistics

PATRICK MAHER

References and further reading

Bacon, F. (1620) *Novum Organum*; repr. ed. T. Fowler as *Bacon's Novum Organum*, Oxford: Clarendon Press, 1888; trans. P. Urbach and J. Gibson, La Salle, IL: Open Court, 1994(Referred to in §4. Book I argues that a new inductive methodology is needed and the incomplete Book II attempts to provide it.)

Carnap, R. (1950) *Logical Foundations of Probability*, Chicago, IL: University of Chicago Press; 2nd edn, 1962. (Carnap's conception of inductive inference, referred to in §4 above, is presented in §44B.)

Descartes, R. (1644) *Principles of Philosophy*; trans. V.R. Miller and R.P. Miller, Dordrecht: Reidel, 1983. (Descartes' defence of the method of hypothesis, mentioned in §1 above, is in §205.)

Fraassen, B.C. van (1989) *Laws and Symmetry*, Oxford: Clarendon Press.(Chapters 6 and 7 discuss inference to the best explanation.)

Jeffrey, R.C. (1956) 'The valuation and acceptance of scientific hypotheses', *Philosophy of Science* 23: 237-46. (Jeffrey's argument against acceptance of uncertain conclusions, mentioned in §4.)

Levi, I. (1991) *The Fixation of Belief and its Undoing*, Cambridge: Cambridge University Press.(A version of the decision-theoretic account of rational acceptance referred to in §4, with acceptance taken to imply certainty.)

Maher, P. (1993) *Betting on Theories*, Cambridge: Cambridge University Press.(A version of the decision-theoretic account of rational acceptance referred to in §4, with acceptance taken to be compatible with uncertainty.)

Mill, J.S. (1843) *A System of Logic*, London: Longman; 8th edn, 1872.(Referred to in §§1, 2 and 4. Mill's theory of induction can be studied by reading the introduction, Chapter I of Book II, and Book III.)

Milton, J.R. (1987) 'Induction before Hume', *British Journal for the Philosophy of Science* 38: 49-74.(Second section surveys thought on induction from Aristotle to Locke.)

Skyrms, B. (1986) *Choice and Chance*, Belmont, CA: Wadsworth, 3rd edn.(An elementary contemporary introduction to induction, probability and decision theory.)

Whewell, W. (1840) *Philosophy of the Inductive Sciences*, London: Parker; 2nd edn, 1847.(Referred to in §4. Advocates the method of hypothesis; presents a conception of induction deeply influenced by Kant.)

Inference to the best explanation

Inference to the best explanation is the procedure of choosing the hypothesis or theory that best explains the available data. The factors that make one explanation better than another may include depth, comprehensiveness, simplicity and unifying power. According to Harman (1965), explanatory inference plays a central role in both everyday and scientific thinking. In ordinary life, a person might make the inference that a fuse has blown to explain why several kitchen appliances stopped working all at once. Scientists also seem to engage in inference to the best explanation; for example, astronomers concluded that another planet must exist in order to account for aberrations in the orbit of Uranus. However, despite the suggestiveness of cases like these, the extent to which we do and should rely on inference to the best explanation is highly controversial.

1 The legitimacy of inference to the best explanation

If inference to the best explanation is legitimate, then a person is entitled to accept a hypothesis provided it meets certain minimal standards in accounting for the relevant data, and explains the data better than any other hypothesis available. But, presumably, when we make an inference from a body of evidence, our goal is to arrive at a further or wider grasp of the truth. Inference to the best explanation will advance this goal only if the satisfaction of explanatory desiderata makes a hypothesis likelier to be true. The trouble is that explanatory virtues (like breadth or simplicity) and truth appear to be unrelated (see Theoretical (epistemic) virtues). Under these circumstances, believing a hypothesis because of its explanatory value would not be much better than believing it because someone thought it up on your birthday. It seems that such a procedure will, if anything, hinder the search for truth, not promote it.

Proponents of inference to the best explanation may respond to this line of criticism in a number of ways:

(1) They could deny that truth is the goal of inquiry (see Scientific realism and antirealism §§1-2). This reply may become more difficult to sustain when inference to the best explanation is used in everyday life. For example, it would be uncomfortable to think that a jury which rejects a contention on explanatory grounds (say, for having too many loose ends) is not seeking the truth.
(2) Another response would be to deny that the goal of inquiry is *solely* to amass a body of truths. Rather, we seek (true) explanations of phenomena. Explanatory goodness thus stands alongside likelihood of truth as a distinct but legitimate basis for evaluating hypotheses. This suggestion, however, does not seem to dispel the concern that there may be a *conflict* between pursuit of truth and pursuit of explanatory goodness.
(3) Let us say that a hypothesis is *tested* in so far as it, or its consequences, is compared with observed data, and let us grant that the successful testing of a hypothesis by observations (*confirmation*) increases the likelihood that the hypothesis is true. One might then try to argue that better explanations are more testable by a body of data than inferior ones. In that event, the superior explanation will be better confirmed by the data, and favouring the superior explanation would at least indirectly further the pursuit of truth.

Consider, for illustrative purposes, the following all-too-crude way of linking simplicity and testability. Suppose hypothesis *H* explains a given phenomenon by positing one mechanism *A*, while its competitor *H** accounts for the same phenomenon by positing the joint action of two different mechanisms *B* and *C*. The simpler hypothesis *H* will be more readily tested because we need evidence to support only one independent claim (‘*A* is at work’), instead of two (‘*B* is at work’ and ‘*C* is at work’). A difficulty here is that one might just as well say that *H* is *less* testable than *H**, since *H** can be *disconfirmed* by finding evidence against either of its components. Friedman (1983) and others argue with much more sophistication and plausibility that: (a) explanatory hypotheses are ones that unify the data; and (b) to the extent that hypotheses participate in multifarious, unifying relations to the data, they are more thoroughly exposed to testing by that data. Hence, superior explanations are more strongly confirmed by observations, and have better claims to truth.

2 Inference to the best explanation and enumerative induction

Inference to the best explanation is *ampliative*. The conclusion one reaches is not a mere summary of the data on hand - one comes to believe something further which explains the data. We also recognize a pattern of ampliative inference called ‘enumerative induction’. This is the extrapolation of observed regularities to universal

generalizations or to conclusions about particular unobserved cases. Thus, given that every calico cat you have observed has been female, you may infer that the next calico cat you see will be female. This conclusion certainly goes beyond the data you possess, so enumerative induction is ampliative.

With these points in hand, we can consider a further response to the challenge raised in the previous section: (4) There is indeed no logical connection between the satisfaction of explanatory criteria and truth, but this gap does not impair the legitimacy of inference to the best explanation. For, arguably, no pattern of ampliative inference can be shown antecedently to lead to true conclusions (see Induction, epistemic issues in). If the lack of such a guarantee were to make a pattern of ampliative inference illegitimate, it would be improper for us to rely on enumerative induction. Since this result is absurd, there can be no force to the demand that explanatory goodness be demonstrably associated with truth.

However, not all patterns of ampliative inference are reasonable or acceptable. Hence, (4) as such does not establish that inference to the best explanation, in particular, has anything to recommend it. The outcome would be more telling if inference to the best explanation could be more closely linked with enumerative induction. Proceeding along these lines, Gilbert Harman has argued forcefully that all enumerative induction really *is* inference to the best explanation (at least implicitly). If so, the legitimacy we accord to induction cannot be wholly denied to inference to the best explanation.

A difficulty for this view is that we seem to make inductive inferences that in no way serve to explain what has been observed. For example, you might notice that there have been fewer people in the supermarket on Tuesdays than on other days. You conclude by induction that Tuesday is in general the least crowded day at the market. Yet, you may have no explanation at all for what you have observed, no idea why the store has had fewer customers on Tuesdays. It is therefore difficult to see how inference to the best explanation enters into your inductive reasoning. This objection follows Ennis (1968).

Still, there is something suggestive about the notion that inference to the best explanation is of a piece with enumerative induction. Suppose that inference to the best explanation favours simple hypotheses. The inductive extrapolation of observed regularities to other times and places also displays a methodological preference for certain kinds of uniformity or simplicity. This point is highlighted by various cases which can equally well be treated as instances of explanatory inference or of enumerative induction. Suppose you are trying to determine the functional dependence of two quantities, X and Y, and you have collected data points like (1,1), (2.5, 2.5), (4,4), and (5,5). It would be natural and reasonable to suppose that the function is $X = Y$, representable as a straight line passing through the data points. Reaching the conclusion $X = Y$ can be thought of as a simple extrapolation from the data given, that is, as an example of enumerative induction. There are, however, indefinitely many wavier curves that pass through the given data points - for example, the graph of the function $Y = \sin(2\pi X) + X$. The choice of a straight line to connect the points is sometimes cited as a paradigm of inference to the best explanation; one is opting for the simplest hypothesis (in this case, the simplest curve or function) that accounts for the data (see Simplicity (in scientific theories)). If curve-fitting cases provide reason to think that enumerative induction and inference to the best explanation really merge or overlap, a critic will be unable to accept one wholeheartedly while rejecting the other across the board.

3 The scope and success of inference to the best explanation

The discussion to this point has treated the legitimacy of inference to the best explanation as an all-or-nothing affair. However, explanatory inference seems to raise the greatest doubts when it is used to support hypotheses about *theoretical* entities - that is, about entities that are unobservable in principle. The basic concern seems to be that theory is underdetermined by evidence: if you can formulate one theory about unobservables to account for a body of evidence, you can formulate indefinitely many others. Inference to the best explanation will select one theory from among these as the correct theory if it best meets our criteria of explanatory adequacy. How, though, are we ever to verify that this choice is in fact the correct one? The candidate theories agree in the observations they predict, so no evidence we can collect will settle the matter.

The reservation here would appear to be that the use of inference to the best explanation in theoretical contexts exacerbates the possibility that explanatory value and truth may diverge. For, if inference to the best explanation does lead us astray in some theoretical context, we will be unable to discover the error by observation, and we will

have gone *hopelessly* astray. One response to this heightened concern would be retrenchment. For example, Reichenbach (1976) held that it is indeed proper to choose the simplest curve that fits one's data, as above - but only when that choice can be tested against additional data-points provided by observation. So, more generally, one might restrict the use of inference to the best explanation to non-theoretical contexts.

Another response by proponents of inference to the best explanation is to become more ambitious rather than less. Boyd (1991) and others maintain that science has historically demonstrated an impressive amount of *convergence*. That is, successor theories have tended to retain some important part of their predecessors' content. The best explanation of this fact is supposedly that our theories have become progressively more accurate. Now, if inference to the best explanation has had a large role to play in selecting scientific theories, the phenomenon of convergence provides evidence that inference to the best explanation leads to the truth, even in theoretical contexts.

Opinion about this ambitious line of thought has been sharply divided. One prominent objection is that, in so far as the argument purports to legitimate inference to the best explanation, it is unacceptably circular, as the thesis that inference to the best explanation reliably leads to the truth is itself established by inference to the best explanation (see Scientific realism and antirealism). Proponents of the argument deny that this circularity is objectionable. Also, altogether different defences of inference to the best explanation may sanction its use in theoretical settings.

4 Further issues

Bas van Fraassen (1989) and others have written critically about inference to the best explanation from a Bayesian or probabilist perspective. They hold that necessary and sufficient conditions for justified belief are provided by the probability calculus and the requirement that a subject's beliefs should be updated by the process of conditionalization (see Probability theory and epistemology §2). To the extent that inference to the best explanation imposes some further qualitative condition on the way one updates one's beliefs, such inference would lead one to accept, irrationally, bets that are guaranteed to lose. One way to try to meet this challenge would be to show how explanatory inference could be incorporated within a probabilist setting. Another response would be to argue that these critics have somehow misconstrued the way inference to the best explanation works.

There are various important aspects of inference to the best explanation that have not even been touched upon here. These include its implications for 'explanatory coherence' theories, the relation between inference to the best explanation and causal inference, and the utility of inference to the best explanation in responding to sceptical arguments (see the further reading suggestions offered below for references on these topics).

See also: Rational beliefs; Scepticism

JONATHAN VOGEL

References and further reading

Boyd, R. (1991) 'Observations, explanatory power, and simplicity: toward a non-Humean account', in R. Boyd, P. Gasper and J.D. Trout (eds) *The Philosophy of Science*, Cambridge, MA: MIT Press.(Rich treatment of many of the issues discussed in this entry.)

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Oxford University Press.(See especially chapters 4 and 5; against inference to the best explanation, with an interesting comparison to causal inference.)

Ennis, R.H. (1968) 'Enumerative induction and best explanation', *Journal of Philosophy* 65 (18): 523-9. (Criticizes Harman's thesis that enumerative induction is implicitly inference to the best explanation.)

Friedman, M. (1983) *Foundations of Space-Time Theories*, Princeton, NJ: Princeton University Press.(Advanced work, of which chapters 6 and 7 are most relevant; makes points that are both supportive and critical of inference to the best explanation.)

Harman, G.H. (1965) 'The inference to the best explanation', *Philosophical Review* 74 (1): 88-95.(A seminal article.)

Harman, G.H. (1968) 'Enumerative induction as inference to the best explanation', *Journal of Philosophy* 65 (18): 529-33.(Harman's response to Ennis.)

Lipton, P. (1991) *Inference to the Best Explanation*, London: Routledge.(A good and approachable monograph.)

Peirce, C.S. (1960) *Pragmatism and Pragmaticism*, in *Collected Papers of Charles Sanders Peirce*, vol. 5, ed. C. Hartshorne and P. Weiss, Cambridge, MA: Harvard University Press.(Important source for Peirce's views

about inference to the best explanation, which he calls ‘abduction’ or ‘hypothesis’.)
Railton, P. (1989) ‘Explanation and metaphysical controversy’, P. Kitcher and W.C. Salmon (eds) *Scientific Explanation*, Minnesota Studies in the Philosophy of Science, vol. 13, Minneapolis, MN: University of Minnesota Press.(Masterly discussion of some issues raised in §3 above.)
Reichenbach, H. (1976) *Experience and Prediction*, Chicago, IL: University of Chicago Press.(Classic treatment of induction, simplicity and the curve-fitting problem.)
Sellars, W.S. (1979) ‘More on givenness and explanatory coherence’, in G.S. Pappas (ed.) *Justification and Knowledge*, Dordrecht: Reidel.(Difficult but essential discussion of explanatory coherence.)
van Fraassen, B.C. (1989) *Laws and Symmetry*, Oxford: Oxford University Press. (Advanced work that raises important objections from a probabilist perspective.)
Vogel, J.M. (1990) ‘Cartesian skepticism and inference to the best explanation’, *Journal of Philosophy* 87 (11): 658-66.(Discusses the use of inference to the best explanation to solve the problem of scepticism about the external world.)

Inference, Indian theories of

The use of argument in rational inquiry in India reaches almost as far back in time as its oldest extant literature. Even in very early texts, one finds the deliberate use of modus tollens*, for example, to refute positions thought to be false. In light of such practice, it is not surprising to discover that Indian thinkers came to identify certain forms of reasoning and to study them systematically.*

The study of inference in India is, as Karl Potter (1977) has emphasized, not the study of valid reasoning as reflected in linguistic or paralinguistic forms, but the study of the circumstances in which knowledge of some facts permits knowledge of another fact, and of when acceptance by one person of some state of affairs as a fact requires that that person accept another as a fact. Still, the form of inference which came to be systematically investigated in India can be given schematically (see below).

*At the core of the study of inference in India is the use of a naïve realist ontology. The world consists of individual substances or things (*dravya*), universals (*sāmānya*), and relations between them. The fundamental relation is the one of occurrence (*vṛtti*). The relata of this relation are known as substratum (*dharmin*) and superstratum (*dharma*) respectively. The relation has two forms: contact (*saṃyoga*) and inherence (*samavāya*). So, for example, one individual substance, say a pot, may occur on another, say the ground, by the relation of contact. In this case, the pot is the superstratum and the ground is the substratum. Or a universal, say brownness, may occur in an individual substance, say a pot, by the relation of inherence. Here, brownness, the superstratum, inheres in the pot, the substratum. The converse of the relation of occurrence is the relation of possession.*

*Another important relation is the relation that one superstratum bears to another. This relation, known as pervasion (*vyāpti*), can be defined in terms of the occurrence relation. One superstratum pervades another just in case wherever the second occurs the first occurs. The converse of the pervasion relation is the concomitance relation. As a result of these relations, the world embodies a structure: if one superstratum* H *is concomitant with another superstratum* S*, and if a particular substratum* p *possesses the former superstratum, then it possesses the second. This structure is captured in this inferential schema:*

Pakṣa *(thesis):* p *has* S.
Hetu *(ground):* p *has* H.
Vyāpti *(pervasion): Whatever has* H *has* S.

Here are two paradigmatic cases of such an inference:

Pakṣa *(thesis):* p *has fire.*
Hetu *(ground):* p *has smoke.*
Vyāpti *(pervasion): Whatever has smoke has fire.*

Pakṣa *(thesis):* p *is a tree (that is, has tree-ness).*
Hetu *(ground):* p *is an oak (that is, has oak-ness).*
Vyāpti *(pervasion): Whatever is an oak (that is, has oak-ness) is a tree (that is, has tree-ness).*

1 Inference and rational inquiry

In tracing out the origin and development of the science of reasoning in general, and of inference (*anumāna*) in particular, one must be careful to distinguish the exercise and implicit recognition of forms of reasoning in thought and inquiry from their explicit articulation and systematic study. This distinction is well exemplified by classical Greek mathematicians: the forms of arguments used in their proofs far exceeded Aristotle's syllogistic and Stoic propositional logic combined. In a similar way, classical Indian thinkers did not articulate many forms of argument which they instinctively recognized as valid and routinely deployed in rational inquiry.

While such principles of logic as the law of non-contradiction, the law of double negation, the law of excluded middle, De Morgan's laws and the equivalence of a proposition and its contrapositive were widely and routinely used by Indian thinkers, their formulation was not taken up for systematic study. Hence, the use of these principles of logic is not discussed in the present entry. Instead, attention is confined to how Indian thinkers came to identify

inference as captured in the inferential scheme and how they made such inferences the object of systematic study.

The history of thought in India can be divided into three periods: the pre-classical period (up to the fourth century BC), the classical period (up to the tenth century AD) and the scholastic period (from the eleventh century on). During the pre-classical period the practice of rational inquiry and debate developed, furnishing the raw material for the formulation of works on the science of reason - *hetu-śāstra* or *nyāya*, as the Indian thinkers called it - which appeared at the beginning of the classical period and played an important role in every major school of philosophical and empirical thought.

Rational inquiry comprises the search for reasons for publicly accepted facts, subject to public and rational scrutiny. It is an activity which involves people both severally and collectively. It involves people severally in so far as individual people are the locus of the exercise of reason. It involves people collectively in so far as the exercise of reason is sharpened by the scrutiny of others. Inference is central to both its individual and collective exercise.

Undoubtedly, the Indian science of reason has its origins in the emergence of rational inquiry and debate. However, the origins of rational inquiry and debate are not clear, the relevant literature being scant and the references therein scanty. None the less, a sketch of the earliest period can be constructed.

2 Seeds of rational inquiry

The seeds of rational inquiry in India are found in appendices to its earliest extant body of literature, the Vedas (Wisdom, Knowledge), composed between the thirteenth and eleventh centuries BC, and comprising four collections (*samhitās*): the *Ṛg Veda (Wisdom of the Verses)*, hymns dedicated to the gods worshipped by the Indo-Aryans; the *Yajur Veda (Wisdom of the Sacrificial Formulas)*, sacred formulas recited at rituals; the *Sāma Veda (Wisdom of the Chants)*, an anthology of chants used by priests in their rituals; and the *Atharva Veda (Wisdom of the Atharva Priests)*, a compendium of hymns, incantations and magic charms used by priests.

Their appendices include the Brāhmaṇas (Books Pertaining to Prayer, dating from between the eleventh and eighth centuries BC), and the Āraṇyakas (Books Studied in the Forest, from between the ninth and seventh centuries BC). In them one finds speculation about the causal relations involved in sacrificial rites. Thus, one encounters the concept of *ṛta*, the natural order of things, which underwrites the efficacy of sacrifice, that is, the capacity of the performance of sacrificial rites to bring about prescribed effects. One also encounters the concept of karma, whereby deeds performed at one point in one's life have consequences which manifest themselves in terms of pain and pleasure experienced later in one's life or in one's later life (see Karma and rebirth, Indian conceptions of). Moreover, speculation about the origins of the universe is also found in the Aitareya Āraṇyaka.

The kind of speculation just described appears to have become the object of public contests in which contestants showed off their knowledge of ritual and their proto-cosmological speculations as to its basis. Inevitably, the views of these contestants differed, and indeed conflicted. Such differences undoubtedly gave rise to questions of how to establish the correctness of one view and how to adjudicate between conflicting views. In this way, according to some scholars (for example, Witzel 1987), arose the practice of public debates, so common in the classical and scholastic periods.

In any event, over the course of time the range of objects of speculation grew to include not only questions of cosmology but also questions of metaphysics, especially the question of what there is, and the naturally ensuing question of how one knows or can justify one's claims about what there is. Thus, one finds views purporting to assist humankind in the alleviation of the unhappiness it suffers, views about what comprises happiness, what constitutes matter, whether or not there is an eternal soul, whether or not there is free will, and whether or not knowledge is possible. The diversity of views includes not only views defended in the Brahmanical literature, in particular the Upaniṣads (dating from the eighth to sixth centuries BC) and, in the Buddhist canon, the *Tipiṭika (Three Baskets*, dating from after the sixth century BC), but also dissenting views reported in the same literature, such as those of the materialist and atomist Ajita Kesakambalin, of the determinist Makkhali Gosāla, and of the sceptic Sañjaya Belatthiputta.

Central to the pursuit of these questions is the concept of causation, which itself fell within the purview of rational inquiry. Thus, for example, in the Śvetaśvatara Upaniṣad (1.2), a number of views pertaining to causation are

reported: determinism (*niyativāda*), indeterminism (*yadṛcchāvāda*) and latency (*svabhāva*).

While the range of rational inquiry increased, it also deepened, for rational inquiry (*anvīkṣikī*, *parīkṣā*) itself became the object of rational inquiry, that is, questions were raised about how facts are known. Thus, one finds as early as the Taittirīya Āraṇyaka enumerations of the sources of knowledge (*pramāṇas*), namely perception (*pratyakṣa*), inference (*anumāna*), religious authority (*smṛti*) and tradition (*aitihya*).

In short, during the pre-classical period, three practices important to the development of the study of inference make their appearance. Rational inquiry into the causes of things, rational inquiry into how things are known, and public contests which perhaps anticipate public debates.

3 Growth of rational inquiry

By the fifth century BC, great social change was taking place in India and a period of intense intellectual activity began. Rational inquiry had taken hold and its topics were not confined to cosmology and soteriology. True, apart from Pāṇini's *Aṣṭādhyāyī*, the world's earliest extant grammar, no works devoted to such topics as agriculture, architecture, astronomy, grammar, law, mathematics, medicine, phonology and statecraft date from this period. None the less, scholars agree that this period contains the roots of many works not redacted until around the turn of the millennium, including *Kṛṣiśāstra (Treatise on Agriculture)*, *Śilpaśāstra (Treatise on Architecture)*, *Jyotiṣaśāstra (Treatise on Astronomy)*, *Dharmaśāstra (Treatise on Law)*, *Carakasamhitā(Caraka's Collection*, a medical treatise) and *Arthaśāstra (Treatise on Wealth*, a political treatise).

Allusions to questions about the number of sources of knowledge, their nature and their objects are found here and there in the Upaniṣads (for example, Maitrī Upaniṣad). Similar issues find elaboration in the Buddhist literature dating from the same period.

Many of the dialogues of the Buddha, Siddhārtha Gautama, show that some version of the law of non-contradiction is to be respected, that is, they show adherence to the view that consistency is a necessary condition for truth. Moreover, many dialogues suggest that truth consists in correspondence to facts, independent of observers. Indeed, a persistent message of the Buddha is that knowledge is grounded in personal observation and inference, where observation is taken to include paranormal observation (though observation none the less). In addition, Buddhists of this period recognize that inference itself must be grounded in causal explanation.

Indeed, much energy is devoted to the development of causal explanations, not only with respect to grounding soteriological claims, for example, as reflected in the doctrine of dependent origination (*pratītya-samutpada*), but even more broadly. Thus, for example, the *Paṭṭhāna* (Causes), one of the seven works of the *Abhidhamma piṭaka* (Basket of Further Doctrine), itself one of the three 'baskets' making up the Buddhist canon, classifies intentional states in terms of twenty-four kinds of causal relations. Causal relations are invoked to explain perceptual states as well. Indeed causal relations are routinely presupposed in discussions of physical events and processes, such as how seeds give rise to shoots (see Causation, Indian theories of).

Moreover, these thinkers not only used causal relations to construct explanations, they were clearly sensitive to the problem of how to establish the existence of causal relations. Passages in the Pāli Canon indicate awareness of the problem of drawing general conclusions from limited observations. Indeed, this period seems to have witnessed the formulation of Mill's method of agreement and difference, known to Indian thinkers as the method of *anvaya-vyatireka*. Such a method, as G. Cardona (1967) has shown, was clearly known to Kātyāyana (*fl.* third century BC), the author of the earliest extant commentary on the *Aṣṭādhyāyī*, and seems to have been known to the Lokāyatikas, who also date from this early period (see Materialism, Indian school of).

Publicly organized debates (*pariṣad*) were commonplace, alluded to in various Upaniṣads (Chāndogya Upaniṣad 5.3.1; Bṛhadāraṇyaka Upaniṣad 6.2.1) and exemplified in over half of the thirty-four *suttas* of the Buddhist *Dīgha Nikāya* (Collection of Long Discourses), which are a part of the *Sutta Piṭaka (Basket of Discourses)*, one of the three 'baskets' of the Buddhist canon. A better-known, but much later, example of such engagements is the Buddhist work *Milindapañho (Questions of King Milinda)*.

The earliest example of protracted systematic rational inquiry into metaphysical issues is found in Moggalīputta Tissa's *Kathāvatthu (Points of Controversy)*, a Buddhist work of the third century BC. It sets out the refutation of some two hundred propositions over which the Theravādins, one of the Buddhist schools, disagreed with other

Buddhist schools. The treatment of each point comprises a debate between a proponent and an opponent. The refutations turn on demonstrating the inconsistency of a set of propositions.

Not only is the form of the debate canonical but the form of the reasoning is as well. Thus, for example, one finds the following pattern throughout Book 1, Chapter 1:

Proponent: Is *A B*?
Opponent: Yes.
Proponent: Is *C D*?
Opponent: No.
Proponent: Acknowledge defeat, since if *A* is *B*, then *C* is *D*.

It is clear that refutation hinges on inconsistency as exhibited by the following propositional schema: p, $\neg q$, $p \rightarrow q$. Though, as I. Bochenski (1956) has remarked, no propositional forms are identified as such, it is none the less clear that inconsistency of a certain form was considered a decisive refutation.

4 Appearance of the canonical inference

In the classical period, as stated above, begins the redaction of works codifying the results of systematic, rational inquiry begun earlier, not only into such fields as grammar, architecture, mathematics, law and politics, but also into the enduring philosophical questions, including the nature of rational inquiry itself. This is made clear in a brief passage in the *Arthaśāstra* (1.2), a work on statecraft attributed to Kauṭilya (*fl.* fourth century BC).

The texts surviving from this period have diverse origins. There are two important passages in the *Carakasamhitā*, a medical text which was redacted in its current form in the first century AD. There are also texts from the philosophical literature, both Brahmanical and non-Brahmanical. The best-known texts pertaining to inference are in a work attributed to Akṣapāda Gautama, the *Nyāyasūtra*, a Brahmanical treatise on rational inquiry, whose actual redaction is thought to date to the third century AD (see Gautama, Akṣapāda). Another Brahmanical work touching upon inference is the *Vaiśeṣikasūtra*, a work on speculative ontology attributed to Kaṇāda; this was redacted in its current form perhaps at the beginning of the second century AD, though perhaps some of its passages date back to the second century BC. Still another Brahmanical work, which survives only in fragments (which have been collected by E. Frauwallner), is a Sāōkhya one entitled *Ṣaṣṭitantra* (Sixty Doctrines), attributed by some to Pañcaśikha (*fl.* second century BC) and by others to Varṣaṇya (*fl.* after second century AD).

The remaining texts are found in the Buddhist philosophical literature. To begin with, there are four passages in four works by the Buddhist idealist Asaōga (*fl.* fourth century AD): one in his *Abhidharmasangītiśāstra (Treatise on the Proclamation of the Abhidharma)*, one in his *Abhidharmasamuccaya (Compendium of the Abhidharma)*, one at the end of a chapter of his *Yogācārabhūmiśāstra (Treatise on the Stages of Yogic Practice)* and one in a work which survives only in Chinese translation, *Shun Zheng Lun (Treatise on According with What is Correct)*. In addition, Vasubandhu (*fl.* fifth century AD), another Buddhist idealist, thought to be the younger brother of Asaōga, wrote at least three works on debate: *Ru Shi Lun (Treatise on Resembling the Facts*, reconstructed in Sanskrit as *Tarkaśāstra*, or Treatise on Reasoning), which survives only in Chinese translation, the *Vādavidhi (Rules of Debate)*, whose Sanskrit fragments have been collected by E. Frauwallner, and the *Vādavidhāna (Precepts of Debate)*. Finally, there are two other works of unknown author and unknown date which survive only in Chinese translation: *Fang Bian Xin Lun* (reconstructed in Sanskrit by G. Tucci as *Upāyahṛdaya*, or Essential Methods), and *Xian Zhang Sheng Jiao Lun* (reconstructed in Sanskrit by G. Tucci as *Prakaraṇāryavācāśāstra*, or Treatise on the Noble Words of Explanation).

With the notable exception of the *Vaiśeṣikasūtra*, which treats inference only as an epistemic process, the greater part of each text is devoted to inference as argument in debate. These texts typically enumerate, define or classify public discussions, propositions as they are used in public discussions, parts of arguments, qualities which either enhance or detract from a discussant's performance, and statements or actions by a discussant which warrant his being considered defeated, including the uttering of various fallacies.

These texts identify the form of argument used in debate. Thus, for example, the *Nyāyasūtra* (1.1.32), like the *Carakasamhitā* (2.8.31), defines an argument as having five parts: the thesis (pakṣa), or proposition (*pratijñā*); the ground, or reason (*hetu*); the corroboration (*dṛṣṭānta*); the application (*upanaya*); and the conclusion (*nigamana*).

The *Carakasamhitā* furnishes the following example:

Thesis: The soul is non-eternal.
Ground: Because it is detectable by the senses.
Corroboration: It is like a pot.
Application: As a pot is detectable by the senses, and is non-eternal, so is the soul detectable by the senses.
______________________________________ Conclusion: Therefore, the soul is non-eternal.

This form of argument clearly reflects the debate situation. First, one propounds a proposition, that is, one sets forth a proposition to be proved. One then states the ground or reason for the proposition. Next, one corroborates with an example the implicit connection between the property mentioned in the proposition and the property adduced as its ground. The immediately ensuing step, the application, spells out the analogy between the example and the subject of the proposition. Notice that this part of the argument retains vestiges of the analogical reasoning that is no doubt its predecessor. Finally, one asserts the proposition.

As was obvious to these thinkers, not all arguments of this form are good arguments. Not surprisingly, the texts catalogue bad arguments. Grounds adduced in arguments catalogued as bad are referred to as non-grounds (*ahetu*) or pseudo-grounds (*hetu-ābhāsa*). It is difficult to be sure what the basis of the classification was. In the case of the *Nyāyasūtra*, the author gives neither a definition nor an example. Even in cases where definitions and examples are given, as in the *Carakasamhitā*, the modern reader is not always sure what is intended. In all likelihood, included here are cases where the premises of the argument can be true but the conclusion false - formal fallacies - and cases where the argument, though formally valid, is none the less unpersuasive, for example, because its ground (*hetu*) is as controversial as its conclusion.

These very same texts, as well as the *Vaiśeṣikasūtra*, touch on inference as an epistemic process. While the examples of inference furnished all comprise expressions corresponding with the thesis (*pakṣa*) and the ground (*hetu*) of the inferential schema, not all the texts are equally explicit in identifying the form of inference. In particular, both the *Carakasamhitā* (1.11.21-2) and the *Nyāyasūtra* (1.1.5) define inference as knowledge of one fact on the basis of knowledge of another, leaving unmentioned any knowledge of a relation linking the two. Moreover, these texts classify inferences on the basis of characteristics completely extrinsic to the logical features of the inferences adduced. Inferences appear to be classified according to the temporal order of the occurrences of the properties corresponding with those in the thesis (*pakṣa*) and the ground (*hetu*) of the inferential schema.

5 Formalization of the canonical inference

Improved definitions, which mention not only what corresponds with the thesis (*pakṣa*) and the ground (*hetu*) of the inferential schema, but also what corresponds with its pervasion (*vyāpti*), are found in the *Ṣaṣṭitantra*, the *Vaiśeṣikasūtra* and the *Abhidharmasamuccaya* (Further Doctrine Collection), where knowledge of a relation is explicitly included in their definitions of inference. However, what corresponds with pervasion (*vyāpti*) in the inferential schema is not a formal relation, but one of a miscellany of material relations instead. The *Ṣaṣṭitantra* enumerates seven such relations, while the *Vaiśeṣikasūtra* (9.20) enumerates five: the relation of cause to effect, of effect to cause, of contact, of exclusion, and of inherence. In each of these texts, the miscellany of material relations serves to classify inferences. Thus, although in these three works the parts of an inference are made explicit, the formal connection between them remains implicit.

As Katsura (1986) has observed, the earliest extant texts to grapple explicitly with making the formal connection between the parts of an inference apparent are those of Vasubandhu. To begin with, in his *Vādavidhi*, Vasubandhu makes it clear that the relation, knowledge of which is necessary for inference, is not just any in a miscellany of material relations, but a formal relation, which he designates in some places as *avinābhāva* (‘being a *sine qua non*’) and in others as *nāntarīyakatva* (‘being unmediated’).

At the same time, Vasubandhu exploits the idea, ascribed by Asaōga in his *Shun Zheng Lun* to an unknown school (thought by at least one scholar to be the Sāōkhya school), that a ground (*hetu*) in an inference is a proper one if and only if it satisfies three conditions - the so-called *trirūpahetu*, or the grounding property (*hetu*) in its three forms. The first form is that the grounding property should occur in the subject of an inference (*pakṣa*). The second is that the grounding property should occur in those substrata similar to the subject in so far as they have the

property to be established (*sādhya*). And third, the grounding property should not occur in any of those substrata dissimilar from the subject in so far as they lack the property to be established. The first form corresponds to the inferential schema's ground (*hetu*), while the second and the third correspond to its pervasion (*vyāpti*). With the formulation of the *trirūpahetu*, the formal nature of the classical Indian inference is at last fully circumscribed. This form is succinctly displayed by the following schema:

Pakṣa (thesis): *p* has *S*.
Hetu (ground): *p* has *H*.
Vyāpti (pervasion): Whatever has *H* has *S*.

An insightful distinction found in Vasubandhu's work is the one between inference for oneself and inference for others, which makes explicit what had previously been only implicit, namely that inference, the cognitive process whereby one increases one's knowledge, and argument, the device of persuasion, are but two sides of a single coin.

Dignāga (*fl.* beginning of the sixth century), another Buddhist philosopher, consolidates and systematizes the insights into the formal basis of inference found in Vasubandhu's works. First, he makes the distinction between inference for oneself and inference for others the organizational cornerstone of his treatment of inference. Second, he undertakes to make the three forms of the grounding property, the so-called *trirūpahetu*, more precise, pressing into service the Sanskrit particle *eva* (only). And third, and perhaps most strikingly, he devises the *hetucakra* ('wheel of reasons'), a three-by-three matrix set up to classify pseudo-grounds in the light of the last two forms of the *trirūpahetu*. On the one hand, there are the three cases of the grounding property occurring in some, none or all of the substrata where the property to be established occurs. On the other hand, there are the three cases of the grounding property occurring in some, none or all of the substrata where the property to be established does not occur.

6 Developments after formalization

Dignāga's works set the framework within which subsequent Buddhist thinkers address philosophical issues pertaining to inference and debate. Thus, Śankarasvāmin (*fl.* sixth century) writes a brief manual of inference for Buddhists, called the *Nyāyapraveśa (Beginning Logic)*, based directly on Dignāga's work. Not long thereafter, Dharmakīrti (*c.*600-60), the great Buddhist metaphysician, also elaborates his views on inference and debate within the framework found in Dignāga.

Dharmakīrti makes at least two contributions to the treatment of inference. Recall that one of the developments found in Vasubandhu's work was the identification of the formal contribution of what corresponds with the inferential schema's pervasion (*vyāpti*), making it explicit that the corresponding relation is a formal one. One of Vasubandhu's terms for it, namely *avinābhāva*, or 'being a *sine qua non*', made it clear that inference involves some form of necessity. The question raised and systematically investigated by Dharmakīrti is: in what does this necessity consist? His answer is that it consists in two things: the causation relation (*tadutpatti*) and the identity relation (*tādātmya*). The second contribution by Dharmakīrti is his attempt to bring knowledge of absences, or roughly negative facts, within the purview of inference (see Negative facts in classical Indian philosophy §§1-2).

Another important Buddhist thinker who treated inference is Dharmottara (*fl.* eighth century). He wrote a useful commentary on Dharmakīrti's widely read *Nyāyabindu (Drop of Logic)*.

Dignāga not only had a profound influence on his Buddhist followers, he also influenced his non-Buddhist contemporaries and their followers. It would be wrong, however, to conclude that every adoption of ideas similar to those used by Dignāga in his works should be attributed to him. After all, we cannot be certain that his contemporaries did not arrive at similar ideas independently or get their ideas from sources common to them and Dignāga. Thus, for example, in his *Nyāyāvatāra* (Descent into Logic), the Jaina thinker Siddhasena Divākara (*c.* fifth century), a contemporary of Vasubandhu, also speaks of the formal relation *antarvyāpti*, or inner pervasion, which corresponds to *vyāpti*, or pervasion, in the inferential schema. In addition, Praśastapāda (*fl.* sixth century), an adherent of the Vaiśeṣika school and near contemporary of Dignāga, also defines inference in a way which not only makes clear its formal nature but also uses the quantificational adjective *sarva* (all) to make the formal connection precise.

At the same time, some authors of this period seem to have retained a view of inference akin to the one found in the *Ṣaṣṭitantra* and the *Vaiśeṣikasūtra*, in which the formal role of what corresponds with the inferential schema's pervasion (*vyāpti*) has yet to have been identified. This is true both of Vātsyāyana (*fl.* fifth century), the author of the earliest extant commentary on the *Nyāyasūtra*, and of Śabara (*fl.* sixth century), the author of the earliest extant commentary on Jaimini's *Mīmāṃsāsūtra*.

However, it was not long before the advocates of both Nyāya and Mīmāṃsāadapted to the formal view of inference. On the one hand, one finds that the Mīmāṃsāthinker Kumārila Bhaṭṭa (*fl.* early seventh century) adopts, without special comment, the formal perspective. On the other hand, though the Nyāya thinker Uddyotakara (*fl.* late sixth century) argues vigorously against many of Dignāga's views, he none the less advocates a view which presupposes the formalization found in them. For example, Uddyotakara rejects the suitability of the quantifying particle *eva* (only) in the formulation of the three forms of the grounding property (*trirūpahetu*), yet he himself introduces a classification of inference which presupposes Dignāga's formalization. Thus, Uddyotakara classifies grounds (*hetu*) as: concomitant (*anvaya*), where nothing distinct from a particular substratum *p* (in the inferential schema) fails to have the property *S*; exclusive (*vyatireka*), where nothing distinct from *p* has *S*; and both concomitant and exclusive, where some things distinct from *p* have *S* and some fail to have *S*. This classification becomes the standard classification for the adherents of Nyāya during the scholastic period.

The developments of the scholastic period are confined almost entirely to the so-called Navya-Nyāya school. Its contributors to the treatment of inference include Gaōgeśa (*fl.* fourteenth century), perhaps the most outstanding of all the thinkers of this school, Raghunātha (*fl.* early sixteenth century), Jagadiśa (*fl.* later seventeenth century) and Gadādhāra (1604-1709).

We conclude with two warnings. First, the treatment of Indian inference here has centred exclusively on tracing out the unfolding of the inferential schema in the history of Indian thought. It has only hinted at the rich philosophical debates that arose among those Indian thinkers eager to harmonize their insights into the formal nature of inference with their views pertaining to metaphysical and linguistic questions. Second, the treatment has not prepared the reader for the intricacies and pitfalls of the terminology used by Indian thinkers in their discussions of inference. The terminology is both ambiguous and inconstant, not just as one goes from the work of one thinker to that of another, but even within the work of a single thinker. Annoying as the contemporary reader may find such ambiguity, it is necessary to remember that this is an inevitable by-product of intellectual fermentation.

See also: Definition, Indian concepts of; Knowledge, Indian views of; Nyāya-Vaiśeṣika

BRENDAN S. GILLON

References and further reading

Annambhaṭṭa (*c.*1600) *Tarkasamgraha*, trans. M.R. Bodas, Poona: Bhandarkar Oriental Research Institute, 1918. (A very clear manual of logic for the Nyāya school.)

Bochenski, I.M. (1956) *Formale Logik*, trans. I. Thomas, *A History of Formal Logic*, Nôtre Dame, IN: University of Nôtre Dame Press, 1961.(Part 6 is a forty-page history of Indian inference.)

Cardona, G. (1967) 'Anvaya and Vyatireka in Indian Grammar', *Adyar Library Bulletin* 5 (31-2): 313-52.(The author describes the origins of Mill's method of agreement and difference in early Indian thought.)

Dharmakīrti (*c.*650) *Nyāyabindu (Drop of Logic)*, trans. M. Gangopadhyaya, Calcutta: Indian Studies Past and Present, 1971.(A good introduction to Dharmakīrti's metaphysics and logic.)

Dignāga (*c.*525) *Nyāyamukha (Introduction to Systematic Reasoning)*, trans. G. Tucci, *The Nyāyamukha of Dignāga*, Heidelberg: Materialien zur Kunde des Buddhismus, 1930.(This very brief text summarizes Dignāga's views on inference and debate.)

Gillon, B.S. (1986) 'Dharmakīrti and his Theory of Inference', in B.K. Matilal and D. Evans (eds) *Buddhist Logic and Epistemology*, Dordrecht: Reidel, 77-87.(A self-contained explanation of how Dharmakīrti treats knowledge of absences as a form of inference.)

Gillon, B.S. and Hayes, R. (1982) 'The Role of the Particle *Eva* in (Logical) Quantification in Sanskrit', *Wiener Zeitschrift für die Kunde Südasiens* 26: 195-203. (The authors show how the Sanskrit particle *eva* ('only') is used to express logical quantification by Indian logicians.)

Gillon, B.S. and Love, M.L. (1980) 'Indian Logic Revisited: *Nyāyapraveṣa* Reviewed', *Journal of Indian Philosophy* 8: 349-84.(The authors set out the logical and grammatical details of Indian inference as found in

the *Nyāyapraveśa*.)

Hayes, R. (1980) 'Diōnāga's Views on Reasoning (*Svārthānumāna*)', *Journal of Indian Philosophy* 8: 219-77.(A thorough and scholarly treatment of Dignāga's theory of inference.)

Jayatillike, K.N. (1963) *Early Buddhist Theory of Knowledge*, London: Allen & Unwin.(Thorough survey of the intellectual climate at the time of the Buddha.)

Katsura Shoryu (1986) 'On the Origin and Development of the Concept of *Vyāpti* in Indian Logic', *Hiroshima Tetsugakkai* 38: 1-16.(The author traces the origin of the Indian form of inference.)

Kauṭilya (4th century BC) *Arthaśāta*, ed. and trans. R.P. Kangle, Delhi: Motilal Banarsidass, 1988, 3 vols.(A work on statecraft.)

Matilal, B.K. (1985) *Logic, Language, and Reality: An Introduction to Indian Philosophical Studies*, Delhi: Motilal Banarsidass.(The first two chapters address many of the outstanding philosophical and philological problems pertaining to the study of inference in Indian philosophy. This is a difficult but important work by one of the twentieth century's major contributors to Indian philosophy.)

Matilal, B.K. and Evans, D. (eds) (1986) *Buddhist Logic and Epistemology*, Dordrecht: Reidel.(A collection of articles by experts on Buddhist logic and epistemology.)

Milindapañho (The Questions of King Milinda) (*c*.100 AD), The Sacred Books of the East 35-6, trans. T.W. Rhys-Davids, Oxford: Clarendon Press, 1890-4, 2 vols; repr. New York: Dover, 1963.(A work on Buddhist metaphysics set out as a dialogue between the senior Buddhist monk, Nāgasena, and the Indo-Greek king Menander.)

Moggalīputta Tissa (*c*.255 BC) *Kathāvatthu (Points of Controversy)*, trans. S.Z. Aung and C.A.F. Rhys Davids, London: The Pali Text Society, 1915.(The author sets out the refutation of some 200 propositions over which the Sthavīravādins, one of the Buddhist schools, disagreed with the Sarvāstivādins, another Buddhist school. The text illustrates both the canonical forms of debate and reasoning.)

Mokṣākaragupta (*c*.1000) *Tarkabhāṣā*, trans. G. Jhā, Poona: Oriental Book Agency, 1967.(A very clear manual of logic for Buddhists.)

Nyāyasūtra (*c*.3rd century AD), trans. G. Jhā, Delhi: Motilal Banarsidass, 1984.(A fundamental work in the history of Indian inference.)

Potter, K.H. (1977) 'Introduction to the Philosophy of Nyāya-Vaiśeṣika', in *Encyclopedia of Indian Philosophies*, vol. 2, New Delhi: Motilal Banarsidass. (Chapter 9 of Part 1 gives a clear thirty-page survey of the Nyāya theory of inference.)

Śankarasvāmin (*c*.525) *Nyāyapraveśa*, trans. M. Tachikawa, *Journal of Indian Philosophy* 1: 111-45.(A clear and concise manual of logic for Buddhists.)

Staal, F. (1973) 'The Concept of *Pakṣa* in Indian Logic', *Journal of Indian Philosophy* 2: 156-66.(Useful and clear discussion of the difficulties of Indian inferential terminology.)

Stcherbatsky, T. (1931) *Buddhist Logic*, New York: Dover, 2 vols.(The first volume is a comprehensive introduction to Dharmakīrti's theory of knowledge and the second volume is a translation of Dharmakīrti's *Nyāyabindu*, together with the commentary by Dharmottara. The work is a remarkable achievement by one of the truly great scholars of the twentieth century. However, it is very difficult and somewhat dated.)

Uddyotakara (*c*.600) *Nyāyavārttika*, trans. G. Jhā, Delhi: Motilal Banarsidass, 1984. (Sets down one of the most important commentaries on the *Nyāyasūtra*.)

Upaniṣads (8th-6th century BC), *The Upaniṣads*, The Sacred Books of the East 1, 15, trans. F. Max Müller, Oxford: Clarendon Press, 1879-84; repr. New York: Dover, 1962.(These texts record views pertaining to a variety of matters, including cosmological, soteriological and epistemological matters, and are ascribed to thinkers thought to have lived between 1000 and 500 BC.)

Vātsyāyana (*c*.450) *Nyāyabhāṣya*, trans. G. Jhā, Delhi: Motilal Banarsidass, 1984. (The author sets down the earliest extant commentary on the *Nyāyasūtra*.)

Vaiśeṣikasūtra (1st century BC), ed. and trans. A.E. Gough, *The Vaiśeshika Aphorisms of Kaṇāda*, Varanasi: E.J. Lazarus, 1873; repr. Delhi: Munshiram Manoharlal, 1975.(The earliest extant work on speculative ontology within the Brahmanical tradition.)

Vidyabhusana, S.C. (1921) *A History of Indian Logic*, Delhi: Motilal Banarsidass, 1971, 2 vols.(This is the standard, though dated, reference on the history of Indian logic.)

Viśvanātha Pañcanana (*c*.1650) *Bhāṣāpariccheda*, trans. Mādhavānanda, Almora: Advaita Ashrama, 1940.(A widely used manual of logic for the Nyāya school.)

Witzel, M. (1987) 'The Case of the Shattered Head', *Studien zur Indologie and Iranistik* 13/14: 363-415.(This article describes public contests which are thought to be the precursors of public debate.)

Yaśovijaya (*c*.1650) *Jaina Tarkabhāṣā*, trans. D. Bhargaya, Delhi: Motilal Banarsidass, 1973.(A manual of logic

for Jainas.)

Infinitary logics

An infinitary logic arises from ordinary first-order logic when one or more of its finitary properties is allowed to become infinite, for example, by admitting infinitely long formulas or infinitely long or infinitely branched proof figures. The need to extend first-order logic became pressing in the late 1950s when it was realized that many of the fundamental notions of mathematics cannot be expressed in first-order logic in a way that would allow for their logical analysis. Because infinitary logics often do not suffer the same limitation, they have become an essential tool in mathematical logic.

Traditionally, logic - considered as the analysis of valid human reasoning, or something intimately connected with it - was seen as having to be finitary in its expressions, inferences and proofs, owing to the finitude of the human mind. This situation began to change in the second half of the nineteenth century, for two principal reasons. First, in philosophy, the Neo-Kantian conception of ‘*Geltung*’ (roughly, ‘validity’) gave rise both to a logical reconstructionist view in the philosophy of science - represented most prominently by Carnap - and to an anti-psychologistic strain of logical theory - represented most prominently by Frege. Second, in mathematics, although increased attention to foundations led to the elimination from analysis of a direct appeal to the infinite, it also led to increased emphasis on the infinite through the development of set theory.

In Löwenheim’s proof of his famous theorem (1915), we find perhaps the first deliberate use of infinitely long formulas and strings of quantifiers (see Löwenheim-Skolem theorems and nonstandard models). But development was interrupted for some thirty years by the influence of Hilbert, who stressed the finitary standpoint and focused on (the proof theory of) first-order logic and elementary recursion theory (see Hilbert’s programme and formalism).

In the early 1950s older modes of research (for example, investigation by logical reconstruction relying on infinitary means whenever necessary) were knit together again and model theory became fruitful. However, with the success of first-order model theory the deficiencies of this language for the expression of mathematical concepts (for example, compactness arguments) also became apparent. Thus, by the end of the decade, an extended model theory which searched for stronger logics began to emerge. Infinitary logics (ILs), in addition to fragments of *n*th-order logic, new quantifiers, and so on, proved to be a useful tool in increasing the expressive power of logical language.

Karp, Henkin, Tarski and others pioneered the development of ILs. At roughly the same time, ILs were introduced into proof theory as a means of streamlining and extending known techniques for consistency proofs. Since the Lindström theorems (see Lindström 1969) showed that interesting properties of first-order logic (for example, compactness; see Hanf 1964) must be lost if one tries to extend it, a new phase of research began when it was discovered that restriction to ‘admissible’ parts of a language not only preserves the usefulness of ILs but also yields suitable variants of lost properties (for example, Barwise compactness; see Barwise 1975).

Let $\mathcal{L}_{\kappa\lambda}$, with κ, λ (regular) infinite cardinals, denote an otherwise first-order language that allows for disjunctions (conjunctions) with fewer than κ disjuncts (conjuncts) and fewer than λ free variables (and hence fewer than λ quantifiers). (If κ or λ is arbitrary, we write ∞ instead.) Recall that ω is the cardinality of the natural numbers and ω_1 the least cardinal bigger than ω, that is, the cardinality of the reals in case the continuum hypothesis is true. In this notation ordinary first-order language $\mathcal{L}^1$, allowing for only finitely many disjuncts and free variables, becomes the finitary $\mathcal{L}_{\omega\omega}$.

Suppose we supplement a first-order language $\mathcal{L}_{\omega\omega}$ with an infinitary disjunctor, $\bigvee$, and add the formation rule: if Φ is a countable set of first-order formulas $\{\phi_n\colon n \in \omega\}$, then $\bigvee \Phi$ is a formula of $\mathcal{L}_{\omega_1\omega}$, being the countable disjunction of all $\phi \in \Phi$. Similarly for the infinitary conjunctor $\bigwedge \Phi$, though it might also be defined by De Morgan’s law as $\neg \bigvee \{\neg\phi_n\colon n \in \omega\}$.

To obtain a semantics for $\mathcal{L}_{\omega_1\omega}$, add to the standard semantics the clause

$$\mathrm{F} \models \bigvee \Phi \text{ iff there is some } \phi \in \Phi \text{ such that } \mathrm{F} \models \phi$$

and the corresponding clause for $\bigwedge \Phi$. Complementing the deduction rules of $\mathcal{L}_{\omega\omega}$ with the obvious variants for

$\bigvee$ (respectively, $\bigwedge$) results in a sound and complete calculus for $\mathcal{L}_{\omega_1\omega}$. Thus, for example, one might add the following rules:

(1) $\vdash \phi \rightarrow \bigvee \Phi$, for $\phi \in \Phi$
(2) If $\vdash \phi \rightarrow \psi$, for all $\phi \in \Phi$, then $\vdash \bigvee \Phi \rightarrow \psi$.

Two cautionary notes are in order here. First, completeness holds only with respect to the valid sentences and does not generally hold with regard to consequence (specifically, if the set of premises is not countable; see Scott 1965). Second, (2) shows that infinitely long derivations may occur, transforming the formal system into a semi-formal one.

In $\mathcal{L}_{\omega_1\omega}$ mathematical concepts that are not axiomatizable in ordinary $\mathcal{L}_{\omega\omega}$ become so. Simple examples include: the class of torsion groups, with the addition to the group axioms of the axiom

$$\forall x \bigvee \{\underbrace{x \circ \ldots \circ x}_{n \text{ times}} = e\colon\ 1 \leq n \in \omega\}$$

(where *e* denotes the neutral element of a group); the class of fields with characteristic unequal zero, with the addition to the field axioms of

$$\forall x \bigvee \{\underbrace{1 + \ldots + 1}_{n \text{ times}} = 0\colon\ 2 \leq n \in \omega\};$$

and the class of structures isomorphic to the natural numbers, with the addition to the first two Peano axioms of

$$\forall x \bigvee \{\underbrace{\mathrm{S} \ldots \mathrm{S}}_{n \text{ times}}0 = x\colon\ 0 \leq n \in \omega\}$$

(where S denotes the successor function). Among the logical metatheorems, downward Löwenheim-Skolem holds. Thus, if a sentence $\phi \in \mathcal{L}_{\omega_1\omega}$ is satisfiable, then it is countably satisfiable. The upward theorem is more difficult, requiring the use of the Hanf number $\mathbf{H}_\alpha$. If ϕ is satisfiable in all cardinalities $\mathbf{H}_\alpha$, with $\alpha < \omega_1$ and $\mathbf{H}$ a kind of superexponentiation of the alephs, then ϕ is satisfiable in all infinite cardinalities. Among other metatheorems that hold, Lopez-Escobar (1965), using a complete Gentzen system, was able to establish Craig's interpolation theorem and Beth's definability theorem (see also Karp 1964, Keisler 1971; Beth's theorem and Craig's theorem).

For languages between $\mathcal{L}_{\omega_1\omega}$ and $\mathcal{L}_{\infty\infty}$, as expected, the expressive power increases as the language becomes more infinitary. Thus, for example, each well-ordering is characterizable up to isomorphism by a sentence ϕ of $\mathcal{L}_{\infty\omega}$ and the notion of well-ordering itself becomes definable in $\mathcal{L}_{\omega_1\omega_1}$. Large languages thus admit more applications to mathematics (especially in algebraic subjects), even though many fundamental second-order concepts (for example, the class of topological spaces) remain inexpressible in such a language as $\mathcal{L}_{\infty\infty}$. In addition, as the language becomes larger, the rather coherent framework of metatheoretic results that still hold for $\mathcal{L}_{\omega_1\omega}$ dissolves (see Dickmann 1975, 1985).

In proof theory, the Gentzen-type consistency proofs of the early 1950s (see Proof theory) using ω-logic gained considerably in clarity by being reformulated in $\mathcal{L}_{\omega_1\omega}$ (see Tait 1968). Another kind of IL, ordinary $\mathcal{L}^1$ with infinitely long terms, was used to define functionals of finite type and was first employed by Tait (1965) to recast Gödel's *Dialectica* interpretation (producing a second link between infinitary logic and recursion theory in addition to that forged by the introduction of admissible sets; see Intuitionism §6).

See also: Logical and mathematical terms, glossary of

BERND BULDT

References and further reading

Barwise, J. (1975) *Admissible Sets and Structures: An Approach to Definability Theory*, New York: Springer. (Reference work concerning infinitary logic and admissible set theory; accessible.)

Barwise, J. (1980) 'Infinitary Logics', in E. Agazzi (ed.) *Modern Logic: A Survey*, Dordrecht: Reidel, 93- 112.(A mostly historical survey, which complements Moore (1990); very readable.)

Barwise, J. (1981) 'The Role of the Omitting Types Theorem in Infinitary Logic', *Archiv für mathematische Logik und Grundlagenforschung* 21: 55-68.(Surveys a metatheorem not treated here; useful.)

Barwise, J. and Feferman, S. (eds) (1985) *Model-Theoretic Logics*, New York: Springer, part A, chaps 2 and 3, and part C.(These first-rate surveys give ample information on infinitary logics in model theory, including a bibliography; advanced.)

Dickmann, M.A. (1975) *Large Infinitary Languages: Model Theory*, Amsterdam: North Holland. (Textbook; advanced.)

Dickmann, M.A. (1985) 'Larger Infinitary Languages', in J. Barwise and S. Feferman (eds) *Model-Theoretic Logics*, New York: Springer, 1985, 317-63.(Valuable survey; advanced.)

Ebbinghaus, H.-D. and Flum, J. (1995) *Finite Model Theory*, Berlin: Springer.(Textbook which also covers the use of infinitary logics in the growing field of finite model theory; very readable.)

Ebbinghaus, H.-D., Flum, J. and Thomas, W. (1984) *Mathematische Logik*, Heidelberg, Berlin and Oxford: Springer, 4th revised edn, 1996; trans. *Mathematical Logic*, Berlin: Springer, 2nd edn, 1994.(Introductory textbook including some infinitary logic and proofs of the Lindström theorems; very readable.)

Feferman, S. (1977) 'Theories of Finite Type Related to Mathematical Practice', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, ch. D4.(§8.3 provides some information on infinite terms; advanced.)

Hanf, W. (1964) 'Incompactness in Languages With Infinitely Long Expressions', *Fundamenta Mathematicae* 53: 309-24.(A milestone proving that ordinary compactness is lost in infinitary logics; advanced.)

Henkin, L.A. (1961) 'Some Remarks on Infinitely Long Formulas', in *Infinitistic Methods: Proceedings of the Symposium on Foundations of Mathematics*, New York: Pergamon Press, 167-83. (One of the earliest papers on the subject; readable.)

Hugly, P. and Sayward, C. (1983) 'Can a Language Have Indenumerably Many Expressions?', *History and Philosophy of Logic* 4: 73-82.(The authors claim not.)

Karp, C. (1964) *Languages With Expressions of Infinite Length*, Amsterdam: North Holland.(The first textbook on the subject; proceeds algebraically. Austerely written.)

Keisler, H.J. (1971) *Model Theory for Infinitary Logic: Logic With Countable Conjunctions and Finite Quantifiers*, Amsterdam: North Holland.(Proceeds along familiar model theoretic lines; very readable.)

Lindström, P. (1969) 'On Extensions of Elementary Logic', *Theoria* 35: 1-11.(The paper in which the two famous Lindström theorems are proved; advanced.)

Lopez-Escobar, E.G.K. (1965) 'An Interpolation Theorem for Denumerably Long Formulas', *Fundamenta Mathematicae* 57: 253-72.(Accessible.)

Moore, G.H. (1990) 'Proof and the Infinite', *Interchange* 21 (2): 46-60.(Survey of the history of infinitary logic from Aristotle to the early 1960s, which complements Barwise (1980); recommended.)

Pohlers, W. (1989) *Proof Theory: An Introduction*, Berlin: Springer.(Textbook using infinitary languages to obtain Gentzen-style consistency proofs; readable.)

Scott, D. (1965) 'Logic With Denumerably Long Formulas and Finite Strings of Quantifiers', in J.W. Addison *et al*. (eds) *The Theory of Models*, Amsterdam: North Holland, 329-41.(Influential paper arguing convincingly to the effect that $\mathcal{L}_{\omega_1 \omega}$ is the 'truest' first-order extension, whereas $\mathcal{L}_{\omega_1 \omega_1}$ is already second-order logic in disguise.)

Scott, D. and Tarski, A. (1958) 'The Sentential Calculus With Infinitely Long Expressions', *Colloquium Mathematicum* 6: 164-70.(One of the earliest papers on the subject; readable.)

Tait, W.W. (1965) 'Infinitely Long Terms of Transfinite Type', in J.N. Crossley and M.A.E. Dummett (eds) *Formal Systems and Recursive Functions*, Amsterdam: North Holland, 176-85.(Advanced.)

Tait, W.W. (1968) 'Normal Derivability in Classical Logic', in J. Barwise (ed.) *The Syntax and Semantics of Infinitary Languages*, New York: Springer, 204-36.(Milestone introducing $\mathcal{L}_{\omega_1 \omega}$ to proof theory; advanced.)

Tarski, A. (1958) 'Remarks on Predicate Logic With Infinitely Long Expressions', *Colloquium Mathematicum* 6: 171-6.(One of the earliest papers on the subject; readable. Note 5 gives some historical remarks.)

Infinity

The infinite is standardly conceived as that which is endless, unlimited, immeasurable. It also has theological connotations of absoluteness and perfection. From the dawn of civilization, it has held a special fascination: people have been captivated by the boundlessness of space and time, by the mystery of numbers going on forever, by the paradoxes of endless divisibility and by the riddles of divine perfection.

The infinite is of profound importance to mathematics. Nevertheless, the relationship between the two has been a curiously ambivalent one. It is clear that mathematics in some sense presupposes the infinite, for instance in the fact that there is no largest integer. But the idea that the infinite should itself be an object of mathematical study has time and again been subjected to ridicule. In the nineteenth century this orthodoxy was challenged, with the advent of 'transfinite arithmetic'. Many, however, have remained sceptical, believing that the infinite is inherently beyond our grasp.

Perhaps their scepticism should be trained on the infinite itself: perhaps the concept is ultimately incoherent. It is certainly riddled with paradoxes. Yet we cannot simply jettison it. This is why the paradoxes are so acute. The roots of these paradoxes lie in our own finitude: it is self-conscious awareness of that finitude which gives us our initial sense of a contrasting infinite, and, at the same time, makes us despair of knowing anything about it, or having any kind of grasp of it. This creates a tension. We feel pressure to acknowledge the infinite, and we feel pressure not to. In trying to come to terms with the infinite, we are trying to come to terms with a basic conflict in ourselves.

1 Early Greek thought

The Greek word *peras* is usually translated as 'limit' or 'bound'. *To apeiron* denotes that which has no *peras*, the unlimited or unbounded: the infinite. *To apeiron* made its first significant appearance in early Greek thought with Anaximander of Miletus in the sixth century BC (see Anaximander §2). He thought of it as the boundless, imperishable, ultimate source of everything that is. He also thought of it as that to which all things must eventually return in order to atone for the injustices and disharmony which result from their transitory existence.

Anaximander was something of an exception, however. On the whole, the Greeks abhorred the infinite (as the old adage has it). More typical of that era were the Pythagoreans, a religious society founded by Pythagoras (see Pythagoreanism §2). They believed in two basic cosmological principles, *Peras* and *Apeiron*, the former subsuming all that was good, the latter all that was bad. They held further that the whole of creation was to be understood in terms of, and indeed was ultimately constituted by, the positive integers $1, 2, 3, \ldots$; and that this was made possible by the fact that *Peras* was continuously subjugating *Apeiron* (the integers themselves, of course, are each finite). The Pythagoreans were followed to some extent in these beliefs by Plato, who also held that it was the imposition of limits on the unlimited that accounted for all the numerically definable phenomena that surround us.

However, the Pythagoreans soon learned to their dismay that they could not simply relegate the infinite to the role of cosmic villain. This was because of Pythagoras' own discovery that the square on the hypotenuse of a right-angled triangle is equal to the sum of the squares on the other two sides. Given this theorem, the ratio of a square's diagonal to each side is $\sqrt{2} : 1$. There are some good approximations to this ratio: it is a little more than $7 : 5$, for example, and a little less than $17 : 12$. Indeed there are approximations of any desired degree of accuracy. Nevertheless, given the basic tenets of Pythagoreanism, it ought to be *exactly* $p : q$, for some pair of positive integers *p* and *q*. The problem was that they discovered a proof that it is not, which they regarded as nothing short of catastrophic. According to legend, one of them was shipwrecked at sea for revealing the discovery to their enemies. The Pythagoreans had stumbled across the 'irrational' within mathematics. They had seen the limitations of the positive integers, and had thereby been forced to acknowledge the infinite in their very midst.

At around the same time, Zeno of Elea was formulating various celebrated paradoxes connected with the infinite (see Zeno of Elea §6). Best known of these is the paradox of Achilles and the tortoise: Achilles, who runs much faster than the tortoise, cannot overtake it in a race if he lets it start a certain distance ahead of him. For in order to do so he must first reach the point from which the tortoise started, by which time the tortoise will have advanced a fraction of the distance initially separating them; he must then make up this new distance, by which time the

tortoise will have advanced again; and so on *ad infinitum*. Such paradoxes, as well as having a profound impact on the history of thought about infinity, did much to reinforce early Greek hostility to the concept.

2 Aristotle

Aristotle's understanding of the infinite was an essentially modern one in so far as he defined it as the untraversable or never-ending. But he perceived a basic dilemma. On the one hand Zeno's paradoxes, along with a host of other considerations, show that the concept of the infinite really does resist a certain kind of application to reality. On the other hand there seems to be no prospect of doing without the concept, as the Pythagoreans had effectively realized. As well as $\sqrt{2}$, time seems to be infinite, numbers seem to go on *ad infinitum*, and space, time and matter all seem to be infinitely divisible.

Aristotle's solution to this dilemma was masterly. It has dominated all subsequent thought on the infinite, and until very recently was adopted by almost everyone who considered the topic. Aristotle distinguished between the 'actual infinite' and the 'potential infinite'. The actual infinite is that whose infinitude exists, or is given, at some point in time. The potential infinite is that whose infinitude exists, or is given, *over* time. All objections to the infinite, Aristotle insisted, are objections to the actual infinite. The potential infinite is a fundamental feature of reality. It is there to be acknowledged in any process which can never end: in the process of counting, for example, in various processes of division, or in the passage of time itself. The reason why paradoxes such as Zeno's arise is that we pay insufficient heed to this distinction. Having seen, for example, that there can be no end to the process of dividing a given racecourse, we somehow imagine that all those possible future divisions are already in effect there. We come to view the racecourse as already divided into infinitely many parts, and it is easy then for the paradoxes to take hold.

Even those later thinkers who did not share Aristotle's animosity towards the actual infinite tended to recognize the importance of his distinction. Often, though, Aristotle's reference to time was taken as a metaphor for something deeper and more abstract. This in turn usually proved to be something grammatical. Thus certain medieval thinkers distinguished between categorematic and syncategorematic uses of the word 'infinite'. Putting it very roughly, to use the word categorematically is to say that there is something with a property that surpasses any finite measure; to use the word syncategorematically is to say that, given any finite measure, there is something with a property that surpasses it. In the former case the infinite has to be instantiated 'all at once'. In the latter case it does not.

The categorematic-syncategorematic distinction heralds another distinction, whose importance to the infinite is hard to exaggerate. This is the distinction between saying that there is something of kind *X* to which each thing of kind *Y* stands in relation *R*, and saying that each thing of kind *Y* stands in relation *R* to something of kind *X* (not necessarily the same thing each time). This is referred to below as the 'Scope Distinction' (see Scope).

But Aristotle himself was not thinking in these very abstract terms. He took the references to time in his own account of the actual-potential distinction quite literally, and this gave rise to his most serious difficulty. He held that time (unlike space) is infinite. He also held that time involves constant activity, as exemplified in the revolution of the heavens. When our attention is focused on the future, there is no obvious problem with this. Past revolutions, however, because they are past, seem to have an infinitude which is by now completely given to us, and hence which is actual. This difficulty, in various different guises, has been a continual aggravation for philosophers who have wanted to see the infinite in broadly Aristotelian terms.

3 The rationalists and the empiricists

For over two thousand years Aristotle's conception of the infinite was regarded as orthodoxy. Often this conception was motivated by a kind of empiricism: the actual infinite was spurned on the grounds that we can never encounter it in experience. But does the potential infinite fare any better in this respect? Is experience of an infinitude that is given over time any less problematic than experience of an infinitude that is given all at once? The more extreme of the British empiricists were hostile to the infinite in all its guises. Where Aristotle had felt able to accept that space and time were infinitely divisible, Berkeley and Hume denied even that. They thought that the concept of the infinite was one that we could, and should, do without (see Empiricism).

This was partly a backlash against their rationalist predecessors. The rationalists had argued that we could form an

idea of the infinite, even though we could neither experience it nor imagine it. They thought that this idea was an innate one, and that it constituted, or helped to constitute, a vital insight into reality. They did not see any difficulty in this view. As Descartes put it, the fact that we cannot grasp the infinite does not preclude our touching it with our thoughts, any more than the fact that we cannot grasp a mountain precludes our touching it (see Rationalism §2).

Descartes believed that our idea of the infinite had been implanted in our minds by God (see Descartes, R. §6). Indeed this was the basis of one of his proofs of God's existence. Only a truly infinite being, Descartes argued, *could* have implanted such an idea in our minds. Note here the assimilation of the infinite to the divine: this was a legacy of medieval thought which is nowadays quite commonplace. But when the assimilation was first made, at the end of antiquity - most famously, by the Neo-Platonist Plotinus-it marked something of a turning point in the history of thought about the infinite (see Plotinus §§3, 4). Until then there had been a tendency to hear 'infinite' as a derogatory term. Henceforth, it was quite the opposite.

The empiricists, meanwhile, needed to defend their rejection of the infinite against the charge that it invalidated contemporary mathematics. They had more or less sophisticated ways of doing this, though in the case of geometry, where the problem was at its most acute, Hume took the rather cavalier step of simply denying certain crucial principles which mathematicians took for granted. (Berkeley's chief concern was with the use of infinitesimals in the recently invented calculus. In fact, his reservations were perfectly justified: it was a century before they were properly addressed.)

4 Kant

Kant played his characteristic role of conciliator in the debate on the infinite (see Kant §§2, 5, 8). He had an empiricist scepticism about the infinite, based on the fact that we cannot directly experience it. Nevertheless, he sided with the rationalists by insisting that there are certain formal or structural features of what we experience, which are accessible a priori and which *do* involve the infinite. Thus he thought that space and time were infinite (both in the sense of being infinitely extended and in the sense of being infinitely divisible): it is written into the form of whatever we experience that there can also be experience of how things are further out, further in, earlier or later. These, on Kant's view, were mathematical truths, a priori and unassailable.

But there is a question about how the topology of space and time *can* be a priori. Kant's celebrated reply was that space and time are not features of 'things in themselves'; they are part of an a priori framework which we contribute to our experience of things. What then of the *contents* of space and time, the physical universe as a whole? This was different. Kant did not think that what was physical was constructed a priori. Nor, on the other hand, did he think that it was ultimately real, that is to say real in a way that transcends any possible access we have to it. It had no features, on Kant's view, that exceed what we are capable of grasping through experience. So here the concept of the infinite did resist application. It still had what Kant regarded as a legitimate *regulative* use. That is, we could proceed *as if* the physical universe as a whole were infinite, thereby encouraging ourselves never to give up in our explorations. But we ultimately had no way of making sense of such infinitude. Kant was forced to take an extreme empiricist line by denying that the physical universe as a whole is infinitely big, that it has infinitely many parts and, going this time beyond Aristotle (thus bypassing the difficulty that had beset Aristotle himself), that it is infinitely old.

However, there was a dilemma. Kant was also forced to deny that the physical universe is *finite* in each of these three respects. Apart from anything else, to postulate infinite, empty space or time beyond the confines of the physical universe is itself to postulate that which exceeds what we are capable of grasping through experience.

This dilemma looks acute. Kant himself presented it in the form of a pair of 'antinomies'. These antinomies consisted of the principal arguments against the physical universe's being infinite in each of the specified respects, and the principal arguments against its being finite. But he believed that the dilemma contained the seeds of its own solution. If what is physical is not ultimately real - if there is no more to it than what we are capable of experiencing - then we are at liberty to deny that there is any such thing as the physical universe as a whole. There are only the finite physical things that are accessible to us through experience. The physical universe as a whole is neither infinite nor finite. It does not exist.

Kant's solution involved him in a direct application of the Scope Distinction (see §2 above). On the one hand he

affirmed that any finite physical thing is contained within something physical. On the other hand he denied that there is something physical within which any finite physical thing is contained. Both of these, the affirmation and the denial, were grounded in the fact that there is nothing we can identify in space and time such that we cannot identify more. This is fundamentally a fact about us: the fact that we are finite. Our identifications are always incomplete. What Kant added, in an idealist vein, was that what we cannot identify does not exist. Here, as in so many other places, we see how deeply involved with human finitude Kant's philosophy was, and how seriously he took it.

5 Post-Kantian metaphysics of the infinite

Metaphysical thought about the infinite since Kant has continued to be just as deeply involved with human finitude. Existentialists in particular have been greatly exercised by it, especially in its guise of mortality. But they have also for the most part recognized an element of the infinite within us. This too is Kantian. Kant believed that we are free rational agents, and that when our agency is properly exercised, it has an unconditioned autonomy that bears all the hallmarks of the truly infinite. For Kant, this was something which exalts us. But for many of the existentialists, still preoccupied with the fundamental fact of human finitude, it is something which is responsible for the deepest tensions within us, and thus for the absurdity of human existence (see Existentialism §2).

Hegel agreed with Kant that the truly infinite is to be found in the free exercise of reason (see Hegel, G.W.F. §8). But he took this further than Kant. He argued that reason is the infinite ground of everything. Everything that happens, on Hegel's view, can be understood as the activity of a kind of world-spirit, and this spirit *is* reason.

This led Hegel to a very non-Aristotelian conception of the infinite. For Hegel, the infinite was the complete, the whole, the unified. Aristotle's conception of the infinite as the never-ending was in Hegel's view quite wrong. He explained this conception as arising from our finite attempts to assimilate the truly infinite. And he described Aristotelian infinity as a 'spurious', or 'bad', infinity - a mere succession of finite elements, each bounded by the next, but never complete and never properly held together in unity. Such 'infinitude' seemed to Hegel at turns nightmarish, then bizarre, then simply tedious, but always a pale, inadequate reflection of the truly infinite.

6 Modern mathematics of the infinite

Despite Kant's influence on Hegel, and despite his own commitment to infinite reason (as well as to infinite space), Kant certainly helped to propagate the Aristotelian tradition of treating the actual infinite with hostility and suspicion. As this tradition prevailed, the actual infinite came increasingly to be understood in the more general, non-temporal sense indicated in §2 above. Eventually, exception was being taken to any categorematic use of the word 'infinite'. The most serious challenge to this tradition, at least in a mathematical context, was not mounted until the nineteenth century, by Cantor, whose mathematical contribution to this topic is unsurpassed (see Cantor, G.).

Objections to the actual infinite had tended to be of two kinds. The first kind we have already seen: objections based on the fact that we can never encounter the actual infinite in experience. Objections of the second kind were based on the paradoxes to which the actual infinite gives rise. These paradoxes fall into two groups. The first group consists of Zeno's paradoxes and their variants. By the time Cantor was writing, however, the calculus (which had then reached full maturity) had done a great deal to mitigate these. Of more concern by then were the paradoxes in the second group, which had been known since medieval times. These were paradoxes of equinumerosity. They derive from the following principle: if (and only if) it is possible to pair off all the members of one set with all those of another, then the two sets must have just as many members as each other. For example, in a non-polygamous society, there must be just as many husbands as wives. This principle looks incontestable. However, if it is applied to infinite sets, it seems to flout Euclid's notion that the whole is greater than the part. For instance, it is possible to pair off all the positive integers with those that are even: 1 with 2, 2 with 4, 3 with 6 and so on.

Cantor accepted this principle. And, consistently with that, he accepted that there are just as many even positive integers as there are positive integers altogether. Far from being worried by this, he defined precisely what is going on in such cases, and then incorporated his definitions into a coherent, systematic and rigorous theory of the actual infinite, ready to be laid before any sceptical gaze.

It might be expected that, on this understanding, all infinite sets are the same size. (If they are, that is not unduly paradoxical.) But much of the revolutionary impact of Cantor's work came in his demonstration that they are not. There are different infinite sizes. This is a consequence of what is known as Cantor's theorem: no set, and in particular no infinite set, has as many members as it has subsets. In other words, no set is as big as the set of its subsets. If it were, then it would be possible to pair off all its members with all its subsets. But this is not possible. Suppose there *were* such a pairing and consider the set of members paired off with subsets not containing them. Whichever member was paired off with *this* subset would belong to it if and only if it did not belong to it (see Cantor's theorem).

In the course of developing these ideas, Cantor laid down some of the basic principles of the set theory which underlay them; he devised precise methods for measuring how big infinite sets are; and he formulated ways of calculating with these measures. In short, he established transfinite arithmetic.

Even so, there are many who remain suspicious of his work and who continue to think of the infinite in broadly Aristotelian terms. Cantor himself was forced to admit that there are some collections, including the collection of all things, which are *so* big that they cannot be assigned any determinate magnitude (their members cannot be given 'all at once'). Concerning such collections, he even sometimes said that they are 'truly' infinite. There is in fact a real irony here: Cantor's work can in many ways be regarded as corroborating Aristotelian orthodoxy.

Brouwer believed that Cantor had gone wrong in not showing sufficient respect for the first kind of objection to the actual infinite: that we cannot encounter it in experience. All Cantor had done, in Brouwer's view, was to demonstrate certain tricks that can be played with (finite) symbols, without addressing the question of how these tricks answer to experience. The relevant experience here - the experience to which any meaningful mathematical statement must answer, according to Brouwer and other members of his intuitionistic school - is our experience of time. It is by recognizing the possibility of separating time into parts, and then indefinitely repeating that operation over time, that we arrive at our idea of the infinite. And such infinitude is potential, not actual - in the most literal sense (see Intuitionism §1).

There was a very different critique of Cantor's ideas in the work of Wittgenstein, though it led to similar results (see Wittgenstein §14). Wittgenstein believed that insufficient attention had been paid (at least by those interpreting Cantor's work, if not by Cantor himself) to what he called the 'grammar' of the infinite, that is to certain fundamental constraints on what could count as a meaningful use of the vocabulary associated with infinity. In effect, Wittgenstein argued that the word 'infinite' could not be used categorematically.

7 Human finitude

Problems about the infinite, we have seen, are grounded in our own finitude. On the one hand our finitude prevents us from being able to think of anything, including the whole of reality, as truly infinite. On the other hand it also prevents us from being able to think of anything finite - anything to that extent within our grasp - as the whole of reality. One way to reconcile these would be to deny that there is any such thing as the whole of reality and to argue that there are only bits of reality, each a part of some other. Here once again we see application of the Scope Distinction: every bit of reality is a part of something, but there is nothing of which every bit of reality is a part. Aristotle, Kant and even to an extent Cantor played out variations on this theme.

But one of the most pressing questions of philosophy still remains: in what exactly does our finitude consist? Some of the most striking features of that finitude are conditioned by our temporality. In particular, of course, there is the fact of our death. How are we to view death? Among the many subsidiary questions that this raises, there are two in particular which are superficially equivalent but between which it is important to distinguish. Putting them in the crudest possible terms (their refinement would be a large part of addressing them): (1) Is death a 'bad thing'? (2) Would immortality be preferable to mortality?

It can easily look as if these questions must receive the same answer. True, no sooner does one begin refining them than one sees all sorts of ways in which a full, qualified response to one can differ from a full, qualified response to the other. But it is in any case important to see how, even at this crude level, there is scope for answering 'yes' and 'no' respectively. Putting it very roughly, death is a bad thing because it closes off possibilities, but immortality would not be preferable to mortality because mortality is what gives life its most basic structure and, therewith, the possibility of meaning.

To answer ‘yes’ to (1) and ‘no’ to (2) in this way is once again to invoke the Scope Distinction. It is to affirm that at each time there is reason to carry on living for longer, while denying that there is reason to carry on living forever. Meaning, for self-conscious beings such as us, can extend further than any given limits. But it cannot extend further than them all.

If it is true that, in some sense, at some level and with all the myriad qualifications that are called for, the answer to (1) is ‘yes’ and the answer to (2) is ‘no’, then, coherent though that is, it points to a basic conflict in us: while it would not be good never to die, it is nevertheless never good to die. That conflict is one of the tragedies of human existence. It is also a version of the original conflict which underlies all our attempts to come to terms with the infinite. In thinking about the infinite, we are thinking, at a very deep level, about ourselves.

See also: Continuum hypothesis, the §1; Death

A.W. MOORE

References and further reading

My thanks are due to Dartmouth publishers for permission to re-use material from the introduction to my book *Infinity*.

Aristotle (*c.* mid 4th century BC) *Physics*, Books III and IV, trans. E. Hussey, Oxford: Oxford University Press, 1983.(Book III, sects 4-8, presents the main elements of Aristotle’s account of the infinite.)

Benardete, J.A. (1964) *Infinity: An Essay in Metaphysics*, Oxford: Oxford University Press.(Entertaining and wide-ranging discussion of the infinite, with particular emphasis on its paradoxes.)

Bennett, J. (1971) ‘The Age and Size of the World’, in A.W. Moore (ed.) *Infinity*, Aldershot: Dartmouth, 1993. (Excellent discussion of Kant’s antinomies.)

Bolzano, B. (1851) *Paradoxes of the Infinite*, trans. D.A. Steele, ed. F. Prihonsky, London: Routledge & Kegan Paul, 1950.(Anticipates some of Cantor’s ideas, though with much less rigour. Historically significant.)

Brouwer, L.E.J. (1913) ‘Intuitionism and Formalism’, trans. A. Dresden, in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 1983.(Classic statement of some of the fundamental tenets of intuitionism.)

Cantor, G. (1895-7) *Contributions to the Founding of the Theory of Transfinite Numbers*, trans. P.E.B. Jourdain, New York: Dover, 1955.(Cantor’s second major publication, in which he establishes transfinite arithmetic. Of ground-breaking importance, but very technical.)

Hegel, G.W.F. (1812-16) *Science of Logic*, trans. A.V. Miller, London: Allen & Unwin, 1969.(Pages 116-57 and 225-38 provide the main elements of Hegel’s views.)

Heidegger, M. (1927) *Being and Time*, trans. J. Macquarrie and E. Robinson, Oxford: Blackwell, 1978.(*Locus classicus* of existential thought on human finitude, but exceedingly difficult.)

Hilbert, D. (1925) ‘On the Infinite’, trans. S. Bauer-Mengelberg, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967.(A modern classic which defends ‘finitism’, a position according to which all references to the infinite are strictly meaningless though they can serve a useful function.)

Kant, I. (1781) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1933.(The section entitled ‘Transcendental Dialectic’ up to Book II, ch. 2 and omitting Book II, ch. 1 (A293/B349-A340/B398 and A405/B432-A567/B595) presents the antinomies and their solution.)

Lear, J. (1979-80) ‘Aristotelian Infinity’, in A.W. Moore (ed.) *Infinity*, Aldershot: Dartmouth, 1993.(Helpful and instructive account of Aristotle’s views, including a superb discussion of the problem of infinite past time.)

Moore, A.W. (1990) *The Infinite*, London: Routledge.(Introductory and partly historical study of all aspects of the infinite.)

Moore, A.W. (ed.) (1993) *Infinity*, Aldershot: Dartmouth.(Collection of the most important and influential articles on infinity published since 1950, with an extensive annotated bibliography and an introduction which expands on the material in this entry.)

Owen, G.E.L. (1957-8) ‘Zeno and the Mathematicians’, in A.W. Moore (ed.) *Infinity*, Aldershot: Dartmouth, 1993.(Thorough and scholarly discussion of Zeno’s paradoxes.)

Rucker, R. (1982) *Infinity and the Mind: The Science and Philosophy of the Infinite*, Sussex: Harvester Wheatsheaf.(Lively and fascinating account of the more mathematical aspects of the infinite. Defends a kind of mysticism.)

Russell, B. (1926) *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*, London: Allen & Unwin.(Chapters 5-7 provide a vigorous defence of Cantor.)

Salmon, W.C. (ed.) (1970) *Zeno's Paradoxes*, Indianapolis, IN: Bobbs-Merrill. (Collection of some of the best known articles on Zeno's paradoxes.)

Williams, B. (1973) 'The Makropulos Case: Reflections on the Tedium of Immortality', in A.W. Moore (ed.) *Infinity*, Aldershot: Dartmouth, 1993.(Superb defence of the ambivalent attitude to death advocated in §7 above.)

Wittgenstein, L. (1975) *Philosophical Remarks*, trans. R. Hargreaves and R. White, ed. R. Rhees, Oxford: Blackwell.(Section XII and pages 304-14 provide a good representative sample of Wittgenstein's views.)

Information technology and ethics

Information technology ethics is the study of the ethical issues arising out of the use and development of electronic technologies. Its goal is to identify and formulate answers to questions about the moral basis of individual responsibilities and actions, as well as the moral underpinnings of public policy.

Information technology ethics raises new and unique moral problems because information technology itself has brought about dramatic social, political, and conceptual change. Because information technology affects not only how we do things but how we think about them, it challenges some of the basic organizing concepts of moral and political philosophy such as property, privacy, the distribution of power, basic liberties and moral responsibility.

Specific questions include the following. What are the moral responsibilities of computer professionals? Who is to blame when computer software failure causes harm? Is computer hacking immoral? Is it immoral to make unauthorized copies of software? Questions related to public policy include: what constitutes just policy with respect to freedom of speech, association, and the exercise of other civil liberties over computer networks? What determines the extent and limits of property rights over computer software and electronic information? What policies adequately protect a right to privacy?

The list of questions shifts in response to developments in information technology. One noteworthy example is the rise in prominence of questions about communication and information in response to the explosive growth of high-speed digital networks. This shift has subsumed the field commonly called 'computer ethics' under the broader rubric of 'information technology ethics'.

1 The philosophical study of information technology ethics

Information technology gives rise to distinctive ethical questions. These questions challenge traditional moral concepts such as property, privacy, and responsibility, as described below by John Perry Barlow:

> Imagine a place where trespassers leave no footprints, where goods can be stolen an infinite number of times and yet remain in the possession of their original owners, where businesses you never heard of can own the history of your personal affairs, where only children feel completely at home, where the physics is that of thought rather than things, and where everyone is as virtual as the shadows in Plato's cave.
>
> (1991: 23)

In an influential essay, 'What is Computer Ethics?' (1985), James Moor characterized the challenge as a 'conceptual vacuum'. The vacuum is generated when ethical problems raised by information technology are inaccessible to concepts and principles of traditional theory (see Morality and ethics). For example, Barlow's goods that 'can be stolen an infinite number of times' challenge a traditional understanding of property (see §2). In order to resolve problems in information technology ethics, according to Moor, one must not only deliberate about values, but also fill the conceptual vacuum; that is, one must develop ethical concepts and principles so that they cover novel situations. Philosophical work in the area of information technology ethics has been drawn particularly to these problems.

Below, three examples of problems calling for conceptual re-examination are discussed: ownership of computer software and digital information; privacy in the face of information technology; and risks and responsibilities in an age of computerization. Each has attracted considerable attention from philosophers and others writing about the social, ethical, and political impacts of information technology.

2 Property and information technology

Ownership of computer software and of digitalized information is a controversial matter. On the question of ownership over computer software some, like the computer scientist Richard Stallman, have argued that private ownership is morally indefensible. They contend that social welfare and progress in the field of computer science are both better served by allowing the free flow of ideas than by protecting the special interests of corporate and commercial developers with strong property rights (see Property). Like laws of physics and mathematical theorems, computer programs ought to remain in the public domain. A system of rewards and incentives could be

devised, even in the absence of property rights, to motivate programmers to continue producing good work.

Most oppose Stallman's position in favour of at least some form and some degree of private ownership. They diverge, however, over the nature and extent of these ownership rights, puzzling over whether the existing legal framework for protecting intellectual ownership is appropriate to the unique metaphysical character of software. Some contend that one or another of the traditional categories of copyright and patent can and ought to be extended to cover software; others contend that a new *sui generis* form of ownership ought to be devised especially for it.

The growth of global, high-speed networks raises yet a different set of ownership issues, not only about computer programs but about all information in digital form, including computerized databases, creative works of writing and music, visual images, even computerized mailing lists and records of computerized transactions. Many have held that the electronic medium poses both a practical challenge to the exercise of traditional property rights, as well as a fundamental challenge to the very idea of ownership. These sceptics say that the digital electronic medium is importantly different because information can be reached via networks virtually instantaneously by vast numbers of people; it can be transmitted and copied with no deterioration in quality; and can be used and enjoyed by others without reducing access by owners. Some see this as a practical problem only. In time, the technical means will be devised to enable owners of digitally stored work to control access to it and exact payment. Others believe that applying old ideas of ownership to this new form of property is not only practically difficult, but morally indefensible. They suggest that society will be able to reap the full benefits of digital technology only by discarding the constraints of traditional intellectual ownership and encouraging the free flow of ideas and information. Strong proprietary holds may give a stranglehold to a few wealthy and powerful individuals over a valuable social resource and may even obstruct the right to free exchange of ideas by limiting quotation of copyrighted sources.

3 Privacy and information technology

Information technology enables the storage, retrieval, manipulation and communication of vast amounts of information. When the information is about persons, this raises serious and perplexing questions about personal privacy (see Privacy).

A great deal has been written about the effects of information technology on privacy. These works, which consider practical questions of immediate policy as well as theoretical questions about the moral foundations of privacy, have concentrated on two broadly defined areas. One deals with databases and 'information privacy', the other with electronic networks and 'communications privacy'.

Most discussions about information privacy concede that if individuals choose to live social lives in a civil society information about them is needed both for the effective running of government and for normal personal dealings. The type and extent of information required remains controversial. Information technology has made the question ever more urgent because of its capacities for storing and using information. Enhanced organization, communication, and retrieval mean that more information can be obtained and more sources tapped. No longer can we count on the *de facto* protections of privacy offered not only by the limitations of human memory but by inefficiencies in record-keeping and information-gathering practices in times before computers were commonly used for this.

In the 1960s, the greatest concern was with the role of governments in creating and using databases about citizens. Although record-keeping was not new to government, computerized records vastly enhanced the degree to which individuals could be tracked. Databases replaced existing paper records, multiplying in scope and magnitude as government agencies eagerly conceived new uses for them. The benefits of computerized databases were clear: greater efficiency, fairer allocation of benefits, and more effective control of crime, among others. Critics, by contrast, brandishing sinister images of 'Big Brother' from George Orwell's novel, *Nineteen Eighty-Four*, urged strict limits. They warned that unchecked collection of data threatened the privacy of individuals, as well as their autonomy and liberty (see Freedom and liberty). Through computerized databases, a government could acquire inordinate power over its citizens.

One way of placing limits is through legislation. In the United States, for example, the Privacy Protection Act of 1974 prohibited a centralized, comprehensive database, secret databases, and databases whose function is not

explicitly stated and justified. Other countries have gone further, creating agencies with powers to regulate and enforce privacy protection in both government and the private sector. Nevertheless, many critics maintain that legislation has not gone far enough in protecting privacy.

As a consequence of technical improvements, greater ease of use, and a drop in prices, the private sector has played a growing role in creating personal databases. Information about persons has become a commodity of great value. Databases include comprehensive records of an individual's lifetime activities held by companies such as credit reporting agencies, medical insurance and mail order companies, as well as specialized databases containing one or two records. Because, in the private sector, political autonomy is not a driving force, the strength of the moral claims of privacy are less clear. One formulation of the dispute pits the rights of database subjects whose data is held frequently without their knowledge and against their will, against the rights claimed by owners of databases to create these databases and profit from them (see Rights).

Philosophical discussion of privacy, suffering from Moor's 'conceptual vacuum', offers only partial insight. The connection some philosophers have demonstrated between personality development, autonomy, intimacy and liberty, on the one hand, and privacy, on the other, has helped justify special protection for confidential information and against invasive means of acquiring information. Philosophical images of privacy, however, remain insufficiently richly textured to draw fine distinctions between a general conception of morally allowable versus morally reprehensible intrusions on privacy.

On the separate question of communications privacy, discussion has focused on public policy. Some argue that it is reasonable to expect the very highest degrees of privacy, guaranteed by unlimited access to encryption technology. Others place interests, such as law enforcement and accountability for communications, above privacy. Progress in this area is likely to involve careful balancing of these competing interests and claims.

4 Risk and responsibility

As with other technologies, information technology introduces a risk of harm as a result of design, system failure or abuse. Because information technology is used so widely, such harm may be extensive and range from mere inconvenience to loss of wealth or even life, from loss of property to loss of reputation and peace of mind. Specialists in the field of computer system reliability have argued that the unique character of computer technology makes it dangerous in unique ways. 'Small' errors which are hard to detect can lead to disastrous outcomes; the enormous complexity of computer systems makes them opaque even to those who have created their components. Where security is not strong, information systems are attractive targets of abuse. The usual philosophical and moral problems of risk apply to information technology as well (see Risk).

Closely related to the problem of risk is that of responsibility (see Responsibility). One set of questions about responsibility looks at the professional responsibility of the producers of information technology, including computer scientists, programmers, and engineers (see Engineering and ethics; Professional ethics). Themes raised in connection with computing overlap substantially with those prevalent in the context of other professions, such as the special responsibilities and privileges of computer professionals, the efficacy of codes of ethics as a means both to encourage and to enforce ethical behaviour, the problem of conflicting loyalties to clients, employers and society at large, and the duty of professional honesty. One consideration not typically an issue with traditional professions such as law and medicine is whether computer professionals are professionals in the full-blown sense at all.

Another set of questions about responsibility is concerned with moral blameworthiness for harms (see Praise and blame). This is not an area unique to information technology ethics, but the technology makes an already thorny issue more complex. A tendency to accept the presence of errors (known as 'bugs') as a normal part of software because they are ubiquitous means they may not be considered as worthy of blame. Furthermore, because of the mystique of computers as intelligent machines, which carry out tasks performed previously by humans in positions of responsibility, there is a tendency to absolve humans from blame in computer-related incidents. Finally, because computer systems and the institutions responsible for producing them are large and complex, blame is easily obscured.

The advent of computer networks raises another set of issues mainly having to do with harms caused by certain uses and abuses of information. Examples include defamatory, offensive and abusive speech and unauthorized

access to proprietary information and software. In response, philosophers, legal theorists, and social commentators have discussed morally and legally defensible lines of responsibility for these actions.

5 Other issues

Balancing the value of answerability for harms are other, sometimes conflicting, moral and political values, including civil liberties such as the freedoms of speech and association, and the right to anonymity (see Freedom of speech). Allowing users of computer networks to post anonymous notices on electronic bulletin boards, for example, creates the possibility of harmful speech for which there is no direct recourse. In another case, free association on-line enables the gathering of groups with immoral intent.

Also of considerable interest to computer ethics and computer law are activities known colloquially as ‘hacking’ and ‘cracking’. These activities range from harmless pranks to unauthorized entry into computer systems to serious destruction of systems and data. Responses to hacking are varied. Some judge it immoral and say it should be outlawed, others not only tolerate forms of hacking that do not result in egregious harm but consider it worthy of both technical and ideological respect. They sympathize with the position frequently expressed by hackers that computing ought to be free and accessible and not hoarded by a privileged few.

The discussion of hacking ties into broader questions of distributive justice (see Justice). Many commentators argue that just societies, in recognition of information technology as a substantial good, ought to guarantee to their citizens reasonable and equitable access to its benefits.

Another important approach to information technology ethics, distinct from the philosophical approach described in this entry, carves the area into quite different categories. It is organized around contexts rather than moral issues. Researchers taking this approach have, for example, examined the effects of computing on the workplace, on education, on social power and democracy, on communication, on information science, on cultures and on communities. Their work is descriptive as well as normative.

See also: Applied ethics; Technology and ethics

HELEN NISSENBAUM

References and further reading

Barlow, J.P. (1991) ‘Coming Into the Country’, *Communications of the ACM* 34 (3): 19-21.(Suggestive short article on the unique nature of computing and information technology.)

Gavison, R. (1980) ‘Privacy and the Limits of Law’, *Yale Law Journal* 89: 421-71. (An article providing a definition of privacy useful for understanding information technology.)

Johnson, D.G. (1994) *Computer Ethics*, Englewood Cliffs, NJ: Prentice Hall.(Clear introductory text written from a philosophical perspective.)

Johnson, D.G. and Nissenbaum, H. (eds) (1995) *Computers, Ethics, and Social Responsibility*, Englewood Cliffs, NJ: Prentice Hall.(Comprehensive set of readings from various disciplines.)

Mitcham, C. (1995) ‘Computers, Information and Ethics: A Review of Issues and Literature’, *Science and Engineering Ethics* 1 (2): 113-26(Good historical review with wide-ranging bibliography.)

Moor, J. (1985) ‘What is Computer Ethics?’, *Metaphilosophy* 16 (4): 266-75.(An essay defining the scope of computer ethics with special attention to its unique elements.)

Regan, P. (1995) *Legislating Privacy: Technology, Social Values and Public Policy*, Chapel Hill, NC: University of North Carolina Press.(A thorough discussion of contemporary privacy issues in the context of the United States.)

Westin, A.F. (1967) *Privacy and Freedom*, New York: Atheneum.(Classic work on the nature and value of privacy in a technological age.)

Weizenbaum, J. (1976) *Computer Power and Human Reason*, New York: W.H. Freeman and Company.(An early classic by a computer scientist which points to the dangers of over-dependence on computers.)

Information theory

Information theory was established in 1948 by Claude Shannon as a statistical analysis of factors pertaining to the transmission of messages through communication channels. Among basic concepts defined within the theory are information (the amount of uncertainty removed by the occurrence of an event), entropy (the average amount of information represented by events at the source of a channel), and equivocation (the 'noise' that impedes faithful transmission of a message through a channel). Information theory has proved essential to the development of space probes, high-speed computing machinery and modern communication systems.

The information studied by Shannon is sharply distinct from information in the sense of knowledge or of propositional content. It is also distinct from most uses of the term in the popular press ('information retrieval', 'information processing', 'information highway', and so on). While Shannon's work has strongly influenced academic psychology and philosophy, its reception in these disciplines has been largely impressionistic. A major problem for contemporary philosophy is to relate the statistical conceptions of information theory to information in the semantic sense of knowledge and content.

1 The technical concept of information

Information theory (less commonly called 'statistical communication theory') was first formulated systematically by Claude Shannon in 1948, in his paper 'A Mathematical Theory of Communication' (see Sloane and Wyner 1993). Its conceptual basis, however, was provided by previous engineering studies of efficiency in the transmission of messages over electrical channels. In 1924 Harry Nyquist proposed a logarithmic measure of speed in the 'transmission of intelligence', and had shown that this variable depends both upon the speed of the signal and upon the number of different signal elements employed. R.V.L. Hartley used the same mathematical construct in 1928 as a quantitative measure of what he called 'information', and showed how this measure relates also to the frequency range of the transmitting medium. Shannon extended the theory to include other factors, such as the effects of noise, the statistical structure of the message and the reliability of the channel through which it is transmitted. His treatment was sufficiently rigorous to permit the derivation of several basic theorems that have proved essential to the design of present-day communication and computational systems.

The commonplace intuition upon which Hartley's measure was founded is that information is conveyed by relatively improbable events, and that the amount of information conveyed by an event (*E*) is inversely related to its probability. In order to make the information of associated events additive, the measure chosen was $\log 1/\mathrm{P}(E)$ - that is, the logarithm of the reciprocal of the initial probability of the event in question. This measure admits further specification, inasmuch as logarithms may be based on any positive number. Three different bases have been used in the literature. When the base chosen is ten, the quantity $\log_{10} 1/\mathrm{P}(E)$ is the unit of information, named a 'Hartley' in honour of the originator of the measure. Natural logarithms also have been used, with a unit called 'nat' for natural unit. Most commonly, however, logarithms to the base two are employed, with a unit called a 'bit' (for binary unit) following the usage of Shannon. One advantage of the latter is the intuitive understanding it provides of the mathematical measure $\log 1/\mathrm{P}(E)$: the number of bits of information represented in the occurrence of a given *E* is equal to the number of times its initial probability must be doubled to reach unity. Thus the occurrence of heads in the flip of an unbiased coin (50 per cent probable) would yield one bit of information, the occurrence of an event initially 25 per cent probable would yield two bits, that of an event initially 60 per cent probable would yield 0.74 bits, and so on. Another advantage of a measure based on logarithms to the base two is its convenient application to factors involved in the design of digital computers. Logarithms to the base two will be assumed in the remainder of this entry.

Information, in the sense (henceforth called 'info(t)') germane to technical information theory, is the quantity measured by $\log 1/\mathrm{P}(E)$. What is measured thereby is the difference between *E*'s initial probability and the probability of 100 per cent that marks its actual occurrence. The difference between a probability before occurrence of 25 per cent and a probability upon occurrence of 100 per cent, as noted above, is a quantity characterized by two bits of information. The information thus characterized is just the difference in probability involved. Insofar as events can be characterized as statistically more or less uncertain prior to occurrence, info(t) is sometimes defined also as the amount of uncertainty removed by the occurrence of *E*. This conception of info(t) as

the removal of statistical uncertainty is useful for interpreting the theory, but must be carefully guarded against psychological or 'subjective' connotations. Information in the relevant sense is entirely quantitative, and is defined with respect to a mathematical measure. Pending further discussion below, this in itself is enough to show that info(t) cannot be equated with either knowledge or semantic content.

2 Information channels

In its bare essentials, an information channel consists of two ensembles of statistically interrelated events. One is the *input ensemble*, consisting of events presented at the input of the channel. The other is the *output ensemble*, consisting of events at the terminus that are to some extent indicative of occurrences at the input. These two ensembles may be represented respectively as *A*, consisting of *n* symbols $a_1, a_2, \ldots, a_n$, and *B*, comprising *m* symbols $b_1, b_2, \ldots, b_m$. Both *A* and *B* are assumed to consist of mutually exclusive events, with combined probability of occurrence equal to unity. The relationship between input and output may then be described by a set of conditional probabilities $\mathrm{P}(b_j/a_i)$, specifying for each output event b_j the probability of its occurrence in association with each input event a_i. This set of conditional probabilities characterizes the *channel* itself, in distinction from its input and output ensembles.

A typical realization of an information channel is a telephone circuit connecting two modems, with events at the sender and the receiver constituting *A* and *B* respectively, and their association through a physical medium represented by an appropriate set of conditional probabilities. Further characterization of the circuit might mention such features as the medium of transmission (satellite, fibre optics, and so on), the intervening switching mechanisms, and the temporal lag between events at input and output. Such features, however, are not essential to an information channel, and may be replaced by others in an equivalent physical system. It is even conceivable that there might be information channels without physically connecting media, such as in communication by 'mental telepathy'. The only general requirement for an information channel is two ensembles related by conditional probabilities.

The quantity of info(t) represented by the occurrence of an event *a* with probability $\mathrm{P}(a)$ is $\log 1/\mathrm{P}(a)$. Unless all input events are equiprobable, this quantity will vary for different members of the input ensemble. The average amount of info(t) represented by occurrences within the input ensemble is the sum of the quantities $\log 1/\mathrm{P}(a)$ for individual events in *A* each multiplied by the probability of its occurrence. This average (which varies with different input distributions $\mathrm{P}(a)$) is called the *entropy* of *A*, and is commonly designated ' $\mathrm{H}(A)$'. If the members of *A* are statistically independent (a 'zero-memory' input), its entropy is given by the formula:

$$\mathrm{H}(A) = \textstyle\sum_A \mathrm{P}(a) \log 1/\mathrm{P}(a)$$

A corresponding average $\mathrm{H}(B)$ can be calculated for the ensemble of output events. In general, the closer the individual events within an ensemble approach equiprobability, the closer the entropy of that ensemble approaches maximum value.

The value of $\mathrm{H}(A)$ does not reflect the statistical interaction between *A* and *B* in an information channel. Because of this interaction, the probability of a given event's occurring in *A* may be affected by the occurrence of an event in *B*, and vice versa. $\mathrm{H}(A)$ thus is sometimes called the a priori entropy of *A*, in contrast to the a posteriori entropy of *A* conditional upon events in *B*. The a posteriori entropy of *A* given a particular b_j is the average info(t) represented by an event in *A* when b_j has occurred in *B*:

$$\mathrm{H}(A/b_j) = \textstyle\sum_A \mathrm{P}(a/b_j) \log 1/\mathrm{P}(a/b_j)$$

Summing $\mathrm{H}(A/b_j)$ for all b_j, we then have the average a posteriori entropy of *A*:

$$\mathrm{H}(A/B) = \textstyle\sum_B \mathrm{P}(b)\mathrm{H}(A/b)$$

$$= \textstyle\sum_B \mathrm{P}(b) \sum_A \mathrm{P}(a/b) \log 1/\mathrm{P}(a/b)$$

$$= \textstyle\sum_{A,B} (a,b) \log 1/\mathrm{P}(a/b)$$

$\mathrm{H}(A/B)$ is called the *equivocation* of *A* with respect to *B*. As this label suggests, the amount of uncertainty left in *A* given the occurrence of an event in *B* is a measure of the 'ambiguity' of *A* with respect to that event, where 'ambiguity' varies indirectly with the reliability of that event as an indication of events in *A*. The above equalities

show how the equivocation of *A* with respect to *B* may be conceived as the average 'ambiguity' of *A* with respect to events in *B*. If we think of *B* as capable of yielding 'information about' *A* in proportion to its general reliability, then that capacity is inversely proportionate to $\mathrm{H}(A/B)$.

There is an intuitive sense in which an information channel with low equivocation is characterized by a representation at the output of approximately the same occurrences as those presented at the input. Conversely, the higher a channel's equivocation, the less reliable are events at its output as representations of input occurrences. The capacity of a channel to convey info(t) thus varies inversely with the equivocation of input to output. This overall capacity also varies directly with the amount of info(t) that can be entered into the channel - that is, with the entropy of the input ensemble. The difference between the entropy of its input and the equivocation of its input with respect to the output thus measures the capacity of the channel as a reliable conveyor of info(t) given a particular input distribution $\mathrm{P}(a)$. This quantity

$$\mathrm{I}(A;B) = \mathrm{H}(A) - \mathrm{H}(A/B)$$

is called the *mutual information* of the information channel. *Channel capacity* in general is defined as the maximum of $\mathrm{I}(A;B)$ over all possible input distributions.

Barring limiting cases, an information channel is characterized by equivocation in either direction, which is to say that both $\mathrm{H}(A/B)$ and $\mathrm{H}(B/A)$ possess positive values. One limiting case is the *noiseless channel*, in which each event at the output indicates a single input event with perfect reliability. A noiseless channel is one with $\mathrm{H}(A/B) = 0$. A *deterministic channel*, on the other hand, is one in which each input event is associated invariably with a single event at the output, and in which accordingly $\mathrm{H}(B/A) = 0$. A channel that is noiseless from input to output thus is deterministic in the opposite direction, and of course vice versa.

3 Two basic theorems of information theory

Practical use of an information channel typically involves both a sender (a source) and a receiver, and communication systems are usually designed with the needs of these users in view. Among factors influencing the design of such systems are considerations of efficiency in the coding of the sender's message. If the anticipated use of a digital communication system is to transmit alphabetically expressed messages, for example, then a system must be devised for coding the letters of the (Latin) alphabet into binary digital form. One such system would employ progressive series of 1s in combination to mark the ordinal place of each letter in the alphabet ('1' for a, '11' for b, '111' for c, and so on), and with '0' to mark the end of a symbol. Simple calculations show that the code symbols assigned to letters in this system would average over thirteen digits in length. A much more efficient system would encode letters in terms of a progressive ordering of 1s and 0s in a five digit sequence ('00000' for a, '00001' for b, '00010' for c, …, '10000' for f, '10001' for g, and so on), with unambiguous decodability assured (independently of synchronization between input and output) by the fact that no complete symbol is a prefix of another symbol. The average symbol length in this system of course is five. One of Shannon's major results is the so-called 'noiseless coding theorem', which establishes limits in the efficiency of coding systems. According to this theorem, the average symbol length of a uniquely decodable coding system can be no less than the entropy of the message source (measured similarly to the input entropy). A special problem arises in the not uncommon case of a source that issues messages in a continuous (for example, not discrete) form, such as music and pictures involving variable pitch and colour. In order to code these messages in discrete binary symbols, their original form must be quantized into small discontinuous segments (as music is recorded on a compact disc). Shannon's noiseless coding theorem also quantifies the cost of fidelity in such transformations in terms of entropy characteristics of the quantized source.

Noise is constituted by any factor that contributes to equivocation $\mathrm{H}(A/B)$, and thus prevents a channel from being noiseless from input to output. Communication engineers have various devices available for reducing noise, including coding systems incorporating redundancy and other 'error correction' features. A simple example of the latter is sheer repetition, as when a specific code element (0 or 1) is transmitted several times in sequence. An unavoidable by-product of such an expedient, however, is an increase in the time required to transmit a given message. Another basic theorem (Shannon's second, also called the 'noisy channel coding theorem') relates requirements of error correction to channel capacity and rate of message transmission through the channel. In very general terms, what this theorem states is that a (sufficiently long) sequence of messages can be transmitted with

an error rate arbitrarily close to zero, as long as the rate of transmission is less than overall channel capacity.

4 The entropy of information theory related to that of thermodynamics

The entropy of an ensemble *A* can be expressed equivalently in the form:

$$\mathrm{H}(A) = -\sum_A \mathrm{P}(a) \log \mathrm{P}(a)$$

which bears an oft-noted similarity to Boltzmann's formulation of thermodynamic entropy (*S*):

$$S = k \log_n \mathrm{P}$$

Both H and *S* vary with the function $\log \mathrm{P}$, where *P* measures the probability of a given system state. Their difference in sign, however, indicates that they are inversely related. An intuitive grasp of this relationship may be provided in terms of the conception of thermodynamic entropy as departure from order. According to the second law of thermodynamics, a closed system tends through time to increase in entropy, which is to lose energy capable of doing work or (equivalently) to lose structure in its overall configuration (see Thermodynamics). Since order and structure are improbable conditions, this means that the detectable configurations (macrostates) in which the system actually exists from moment to moment become increasingly predictable, and that the initially improbable configurations (which provide structure and energy) become even less likely to occur. If we view the total set of possible macrostates (probable and improbable) as an ensemble with the properties of an information source, this means in turn that the possible configurations depart increasingly from equiprobability. Since an information source has maximum H when its states are equally probable, the result is that H decreases with an increase in *S*. This inverse relation between H and *S* has led some information theorists (for example, Brillouin 1962) to refer to info(t), along with energy and structure, as forms of negentropy.

Detailed technical discussions of the relation between H and *S* can be found in Brillouin and in Tribus (1961). An intriguing claim of the latter is that thermodynamics can be developed deductively from quantum physics by application of the methods of information theory.

5 Ramifications of philosophic interest

The term 'information' has various senses in ordinary discourse, including knowledge ('*N*'s information about birds comes from actual experience') and propositional content ('*N*'s report contained information that a golden eagle had been spotted'). Information in these senses has semantic features (call it 'info(s)'), such as reference and truth-value. In contrast with info(t), as already noted, info(s) is not quantitative and cannot be characterized mathematically. Another difference is that communication of info(s) is a three-term transaction, involving sender, receiver and intended referent, whereas the communication of info(t) in relevant respects requires only sender and receiver. Another way of putting this point of difference is that intention is not a factor in the description of channels for communicating info(t).

Despite the insistence of information theorists (including Shannon) from the start that info(t) has nothing to do with semantics, there have been persistent attempts by philosophers and cognitive psychologists to found theories of semantic content on an information-theoretic basis. Shortly following Shannon's original paper, the physicist D.M. MacKay initiated a series of publications purporting to show that information theory deals with representation (a semantic relation; see the bibliography in *Information, Mechanism and Meaning*, 1969). MacKay's purported results stand behind subsequent influential work in information-processing psychology (for example, that of D.E. Broadbent and W.R. Garner), and may be in part responsible for the assumption found in recent cognitive psychology that information theory itself provides a theory of info(s).

An independent approach to semantic issues through information theory among professional philosophers began with D.M. Armstrong (1968) and D.C. Dennett (1969). This essentially materialist approach yielded, in F.I. Dretske (1981), a clearly articulated account of how info(s) might be produced by certain causal (hence, non-information-theoretic) properties of info(t)-processing systems. Apart from their other merits, however, none of these studies maintains fidelity to a precise understanding of information theory as developed by Shannon. An attempt to found an account of info(s) on a precise understanding of Shannon's theory is pursued in K.M. Sayre (1976).

See also: Computability and information; Information theory and epistemology

KENNETH M. SAYRE

References and further reading

All except Brillouin (1962), Tribus (1961) and the Shannon collection (Sloane and Wyner 1993) can be read profitably without specialized training.

Abramson, N. (1963) *Information Theory and Coding*, New York: McGraw-Hill.(A standard introduction accessible to non-specialists.)

Armstrong, D.M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.(An analysis of mind-making extensive use of the concept of information.)

Brillouin, L. (1962) *Science and Information Theory*, New York: Academic Press. (Integrates thermodynamics and information theory. Written for readers with broad interests.)

Dennett, D.C. (1969) *Content and Consciousness*, New York: Humanities Press. (An early attempt to bring the concept of information to bear in philosophic analysis.).

Dretske, F.I. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(A carefully argued and easily readable account of knowledge and related concepts, based on an interpretation of information theory deviating substantially from Shannon's.)

Hamming, R.W. (1980) *Coding and Information Theory*, Englewood Cliffs, NJ: Prentice Hall.(A more recent introduction of comparable difficulty to Abramson (1963). Hamming is a major contributor to information theory.)

MacKay, D.M. (1969) *Information, Mechanism and Meaning*, Cambridge, MA: MIT Press.(A collection of essays documenting the author's influential efforts to inject 'meaning' into technical information theory.)

Sayre, K.M. (1976) *Cybernetics and the Philosophy of Mind*, London: Routledge & Kegan Paul.(An attempt to establish a naturalistic view of the world on the basis of information theory.)

Sayre, K.M. (1986) 'Intentionality and Information Processing: An Alternative Model for Cognitive Science', *Behavioral and Brain Sciences* 9: 121-65.(Tries to show how technical information theory can be employed as the basis of a theory of intentionality.)

Sloane, N.J.A. and Wyner, A.D. (eds) (1993) *Claude Elwood Shannon: Collected Papers*, New York: IEEE Press.(A complete collection of Shannon's technical papers, including two delightful biographical essays.)

Tribus, M. (1961) *Thermostatics and Thermodynamics*, Princeton, NJ: Van Nostrand.(A carefully executed attempt to relate quantum theory and thermodynamics by the statistical methods of information theory.)

Information theory and epistemology

The mathematical theory of information (also called communication theory) defines a quantity called mutual information *that exists between a source, s, and receiver, r. Mutual information is a statistical construct, a quantity defined in terms of conditional probabilities between the events occurring at r and s. If what happens at r depends on what happens at s to some degree, then there is a communication 'channel' between r and s, and mutual information at r about s. If, on the other hand, the events at two points are statistically independent, there is zero mutual information.*

Philosophers and psychologists are attracted to information theory because of its potential as a useful tool in describing an organism's cognitive relations to the world. The attractions are especially great for those who seek a naturalistic account of knowledge, an account that avoids normative - and, therefore, scientifically unusable - ideas such as rational warrant, sufficient reason and adequate justification. According to this approach, philosophically problematic notions like evidence, knowledge, recognition and perception - perhaps even meaning - can be understood in communication terms. Perceptual knowledge, for instance, might best be rendered in terms of a brain (r) receiving mutual information about a worldly source (s) via sensory channels. When incoming signals carry appropriate information, suitably equipped brains 'decode' these signals, extract information and thereby come to know what is happening in the outside world. Perception becomes information-produced belief.

1 Objective information

The mathematical theory of information has its primary application to telecommunications systems. It deals exclusively with *amounts* of information, either the amount generated at a source - the more unlikely an event is the more information its occurrence generates - or the amount of mutual information existing between source and receiver. Mutual information between two points is just a useful measure of the statistical dependence of events occurring at these two points: the greater the independence between events occurring at *r* and *s*, the less mutual information there is between them (see Information theory).

If signals are understood to be information-carrying events (sounds, gestures, electric currents), then the mathematical theory is concerned with the statistical properties of these events, not with the information they happen to carry in any ordinary sense of the word 'information' (news or message). It is concerned with how probable or improbable these events are, not with what, if anything, they say or mean. For this reason it has sometimes been said that communication theory is not really a theory about information - *semantic* information - at all. There is some truth in this charge; mutual information is a statistical quantity and it must be carefully distinguished from the semantic content - the news or message - that events carry.

Nonetheless, though the theory is primarily concerned with the statistical properties of signals, not with the semantic information (if any) they carry, the statistical properties are relevant to a signal's capacity to carry semantic information. Unless there is, in a communication theory sense, a statistically reliable channel of communication between *s* and *r*, the signals reaching *r* from *s* cannot inform, tell or teach one what is happening at *s*. A broken instrument, one whose pointer (*r*) does not bear the right statistical relations to the quantity (*s*) it is supposed to measure, cannot convey semantic information about the value of that quantity. A broken clock cannot tell you the time of day even if it happens (coincidentally) to register the correct time. For the same reason, one receives little or no information from unpredictable liars even when they happen to be telling the truth. Hence, information theory, though not itself directly addressing semantic topics, does, it seems, formulate a possible objective condition on the flow of semantic information. No signal can carry semantic information unless it satisfies the appropriate statistical constraints of communication theory.

It is this fact about communication theory - that it can be interpreted as relating something important about the conditions that are needed for the transmission of information as ordinarily understood - that has tempted philosophers to use it in developing accounts of empirical knowledge. If coming to know *p* is equivalent to being 'informed' that *p*, and being informed that *p* is understood as a way of being made to believe that *p* by the information that *p*, then empirical knowledge is information-produced belief. If, in addition, information is taken to be constrained by, if not equivalent to, mutual information between world and believer, then one is close to having an objective theory of knowledge: knowing that *p* is being caused to believe that *p* by information-bearing signals

arriving over a channel of communication.

Information-based accounts of knowledge, perception, memory and so on are variants of reliability theories of knowledge (see Reliabilism). The basic idea behind such accounts of knowledge is that one knows if and only if one's belief - or (on some accounts) the evidence on which one's belief is based - exhibits a reliable connection with the conditions one believes to exist. Information theory is merely a convenient way of expressing these facts about the required reliability of the belief-formation process.

Information theory would not be useful in epistemology if it embodied - implicitly - the very same ideas that philosophers use it to analyse. If it turned out that the probability relations used by this theory to define information were themselves ideas that presupposed - either in definition or application - knowledge and meaning, then it would be circular to use mutual information as a tool to analyse knowledge and meaning. For this reason, it is important to understand that the 'mutual information' that epistemology borrows from communication theory is supposed to be an objective commodity, one whose existence is independent of our knowledge or understanding. The probabilities defining mutual information cannot, therefore, be what are called *subjective* probabilities, probabilities that depend on what people know or believe about the likelihood of events at *s* given events at *r* (see Probability theory and epistemology§1). If mutual information is to serve as a useful tool in epistemological and semantic studies, then it must be conceived of as an objective measure of correlation or lawful dependency between events. No one need ever know, understand or believe that such correlations exist for them - and, hence, for *mutual* information - to exist.

2 Pure and impure theories

If epistemological concepts like knowledge, justification and evidence are interpreted in informational terms, one can distinguish pure and impure versions of such interpretations. An impure theory of knowledge, for example, would contain a mixture of conditions on knowledge, some involving information (to know that *p* one must stand in the right informational relation to the fact that *p*), others, perhaps, involving subjective conditions (to know that *p*, one must also be subjectively justified in thinking that the appropriate informational relations exist). A pure theory, on the other hand, forgoes all subjective requirements (except belief) and identifies knowledge of the facts with beliefs that stand in the appropriate informational relations to the facts. A pure theory, for example, identifies perceptual knowledge with beliefs aroused by information-carrying sensations. Subjects need not know or be justified in believing that their sensations do convey the requisite information. All that a pure theory requires is that sensations, in fact, carry the right information. In order to know, by perceptual means, that something is moving, one need not know, or be justified in believing, that one is receiving information. All that is required is that one *be* receiving it.

A pure theory of knowledge, in this sense, is most plausible in the case of direct observational knowledge. It enjoys greatest plausibility here because one seldom troubles to determine *how* reliable sensations are - and, thus, how much mutual information they carry - before one trusts them to 'inform' (give one knowledge about) the world one lives in. We see, and we promptly believe; there is precious little (if any) conscious thought, reasoning, justification or inference that comes between the seeing and the believing. We classify the resulting beliefs as knowledge - in both ourselves and animals - as long as we are convinced that the sensory mechanisms are operating reliably, as long as we believe they are channels of mutual information. According to its advocates, then, this is a strong endorsement of informationally pure theories of knowledge. If we do not, in actual everyday practice, require more than information-caused belief, then ordinary everyday knowledge requires nothing more, it is nothing more, than information-caused belief.

Advocates of a pure theory of knowledge also argue that it represents the best answer - indeed, according to some, the *only* answer - to philosophical scepticism. If knowledge of the world required not only information about the world but also that one had to somehow justify the claim that one's senses were providing genuine information, then the philosophical sceptic would be right: knowledge of the world would be impossible. For the only way we have of getting the information that we are, in fact, getting genuine information through the senses is through the very senses whose reliability is in question. This is a circle from which there is no escape. You cannot expect instruments to provide information about their own reliability (see Scepticism §1).

Pure theories are at their weakest when applied to less direct forms of knowledge. When one comes to know that *p*

by reasoning and inference, by using informants, instruments, tests and procedures, it seems implausible to suppose that one needs no information about the channels through which information arrives. Knowledge that *p* may not require knowledge that one is getting genuine information about *p*, but an impure theorist will insist that it requires some evidence, reason or justification for thinking one's instruments, methods and informants are reliable (see Internalism and externalism in epistemology).

3 How much information is enough?

Since mutual information is a statistical quantity that comes in varying amounts, questions arise about how much of it is enough. Even cheap or badly calibrated instruments carry some mutual information about the quantities they are used to measure. Is it enough to know? Another way of putting this question is to ask how much mutual information a signal must carry about a condition at a source in order to carry the semantic information that that condition exists at the source.

Some philosophers have argued that perfect reliability is needed. If, given the sort of signals reaching *r* from source *s*, there is a chance, however small, that condition *q* rather than *p* exists at *s*, then the signals do not carry the information that *p* exists at *s* (though they might carry the information that *p* or *q* exists at *s*). The signals are equivocal - not reliable enough to enable someone at *r* to know that *p* exists at *s*. Others, assuming that such perfect reliability is seldom, if ever, achieved, have argued that to avoid scepticism, information and knowledge must be understood as requiring something less than a probability of 1. Otherwise, for any condition *p*, we would never get the information that *p* exists and, therefore (if knowledge requires information) never know that *p*.

Despite the apparent sceptical consequences of requiring perfect reliability for knowledge and information, there are theoretical reasons favouring this option. For if one could know that *p* when the informational sources are less than completely reliable, then knowledge would cease to 'add up' - one might know that *p* and know that *q* but fail to know that *p* and *q*. For suppose that a probability of only .98 was required to know. Then one might know that *p* (the probability of *p* being .98), know that *q* (probability also .98), but fail to know that *p* and *q* since the probability of *both* *p* and *q* is the product (.96) of their individual probabilities. This, however, seems like an unacceptable result. Knowledge (it is said) adds up: anyone who knows that *p* and knows that *q* *thereby* knows that *p* and *q*. Since the 'unacceptable' result occurs *whenever* the probability required for knowledge is less than 1, it seems that to avoid it, the probability required for knowledge must be set equal to 1.

For these and other reasons, some philosophers despair of finding general answers to questions about reliability conditions on knowledge. They argue that knowledge and semantic information are relative notions. What is *enough* depends on a context of inquiry. When a drug, if known to be safe, will be given to human infants, it takes more (mutual) information to know it is safe than it does when it will be administered to rats in a laboratory test. For the same reason, whether a newspaper contains semantic information about a death, for example, depends on whose death it is and what are the consequences of readers believing it.

See also: Knowledge, causal theory of; Knowledge, concept of; Naturalized Epistemology; Perception, epistemic issues in

FRED DRETSKE

References and further reading

Cherry, C. (1957) *On Human Communication*, Cambridge, MA: MIT Press.(Excellent history and exploration of the attempts to apply information theory in a variety of areas - including psychology.)

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(Develops an information-based theory of knowledge, perception and meaning, and attempts to bridge the gap between mutual information (of communication theory) and semantic information. Contains an elementary exposition of the mathematical theory.)

Dretske, F. (1983) 'Multiple book review of *Knowledge and the Flow of Information*', *Behavioral and Brain Sciences* 6 (1): 55-89.(Especially useful for criticisms and commentary by twenty philosophers and psychologists.)

Sayre, K. (1965) *Recognition: A Study in the Philosophy of Artificial Intelligence*, South Bend, IN: University of Notre Dame Press.(Early, interesting attempt by a philosopher to apply information-theoretical ideas to epistemology.)

Shannon, C. (1948) ‘The mathematical theory of communication’, *Bell Systems Technical Journal* 27: 379-423, 623-56; repr. with intro. by W. Weaver, Urbana, IL: The University of Illinois Press, 1949.(Mathematically demanding book, but the classic source for the engineer’s concept of information.)

Ingarden, Roman Witold (1893-1970)

Ingarden was a leading exponent of phenomenology and one of the most outstanding Polish philosophers. Representing an objectivist approach within phenomenology he stressed that phenomenology employs a variety of methods, according to the variety of objects, and aspires to achieve an original cognitive apprehension of these objects. Its aim is to reach the essence of an object by analysing the contents of appropriate ideas and to convey the results of this analysis in clear language. Ingarden applied his methods in many areas of philosophy. He developed a pluralist theory of being and an epistemology which makes it possible to practise this discipline in an undogmatic manner and to defend the value of human knowledge. In the theory of values he developed an inspiring approach to the analysis of traditionally problematic areas. He was best known for his work in aesthetics, in which he analysed the structure of various kinds of works of art, the nature of aesthetic experience, the cognition of works of art and the objective character of aesthetic values. In general, he gave phenomenology a lucid and precise shape.

In the interwar period Ingarden was the main opponent in Poland of the dominant Lwów-Warsaw School (Polish Analytic School), which had a minimalistic orientation. The main lines of his own investigations emerged largely as a result of his regular debates with Husserl, in particular those concerning Husserl's transcendental idealism. Ingarden's best-known work, Das literarische Kunstwerk (The Literary Work of Art) (1931a) has its origins in this debate.

1 Life

Roman Ingarden was born in 1893 in Cracow and died there in 1970. He studied in Lwów and Göttingen (1912-14), taking philosophy under Edmund Husserl, mathematics under David Hilbert and psychology under G.E. Müller. In 1916 he followed Husserl to Freiburg, where in 1918 he took a doctor's degree, then taught for some time at high schools in Lublin, Toruńand Lwów. He received his habilitation at Lwów University in June 1924, was appointed *dozent* there in 1925, and after Twardowski's retirement obtained the chair of philosophy in 1933. He stayed in Lwów until 1941; during the Soviet occupation he did not teach philosophy but German studies. He was also active in the Polish underground University. From 1945 to 1950, and from 1957 until his retirement in 1963, he was professor at the Jagiellonian University in Cracow.

Most of Ingarden's students from his Lwów period were either killed during the Second World War or became Marxists. After the war another group of his followers emerged in Cracow, though after Ingarden's death many of them questioned elements of his thought, some developing it, some reshaping it and others rejecting it.

2 Ontology

Ingarden held that philosophy divides into ontology and metaphysics. Ontology is an autonomous discipline in which we discover and establish the necessary connections between pure ideal qualities by intuitive analysis of the contents of ideas. This is an indispensable preparation for metaphysics, which aims to elucidate the necessary truths of factual existence. Each section of philosophy - theory of knowledge, philosophy of man, philosophy of nature and so on - has ontological and metaphysical aspects.

Ingarden argues that every being is a triple unity of matter (contents), form (of the matter) and existence (in a certain mode). Accordingly, ontology as a whole is divided into material, formal and existential ontology. Existence is neither a property nor one of the material or formal moments of an object; it is always the existence *of something* and what exists determines by its essence a mode of being which belongs to it. Modes of being are constituted from existential 'moments', of which Ingarden distinguishes the following opposite pairs: originality-derivativity, autonomy-heteronomy, distinctiveness-connectiveness and independence-dependence. Taking into account the modes of being thus constituted, there are four basic spheres of being: absolute (supratemporal), ideal (timeless), real (temporal - it has the most numerous forms) and purely intentional (atemporal, sometimes seemingly in time). Ingarden also draws a distinction between three domains: pure ideal qualities, ideas and individual objects. Each individual object is formally a subject of properties whose identity is determined by its constitutive nature. Individual objects of higher order, such as organisms, may be superstructured on autonomous individual objects. Ideas and purely intentional beings have a two-sided formal

constitution - besides their own structure they also have contents (in the case of ideas it is constituted by constants and variables, and in the case of purely intentional beings by places of indeterminateness).

Analyses of being in time, of the stream of consciousness and of the world show that their existence is derivative and depends on their relation to original (absolute) being. The foundation of being is placed either in its essence (and ultimately in the content of some idea) or is purely factual in its character. In his analysis of the controversy over the existence of the world, Ingarden first formulates Husserl's transcendental starting point, and then demonstrates and states precisely its assumptions concerning the two elements of initial relation: the real world and the stream of consciousness, together with a subject which belongs to it (pure ego). These considerations lead Ingarden to reject both Husserl's solution and his way of setting the question.

What is real appears in three temporal phases: the future, the present and the past. Objects determined in time include objects enduring in time, processes and events. A human being is an object enduring in time and constituted by a soul, which comprises an ego together with a stream of consciousness, and a body (with a subsystem constituting 'the gate of consciousness'). Living on the border of two spheres, the real (nature, animality), and the ideal (values), human beings create a third sphere of culture. Thus their need to transcend this fragility by a process of self-formation that is subordinated to values makes them prone to tragedy.

3 Aesthetics

Ingarden gave special attention to questions concerning aesthetic and moral values. Art plays a significant role in human life, and Ingarden develops a complex theory of works of art: a work of art exists in a purely intentional manner, though on the basis of some physical foundation of its identity; it is formally two-sided (it has some contents, for example, a represented world in certain types of arts); it can be analysed in the respect of its phases and its 'layers'; it is given (in its contents) in quasi-perception or in imagining ('imagining-thinking'); if it is of linguistic character (such as a literary work), sentences in it form quasi-propositions, and not propositions in the sense of logic; because of its schematic character it allows for various interpretations, but its main aim is to make aesthetic interpretation (the construction of an aesthetic object) possible. Ingarden draws a distinction between artistic values, which belong to the work of art, and aesthetic values, which belong to the aesthetic object. A succesful work of art may realize not only an aesthetic value but, in connection with this value, it may also 'flare-up' with a 'metaphysical' quality (that is, an existentially significant quality) and make visible an essential connection between some concrete situation in life (embodied in the represented world) and an existential quality or qualities. Aesthetic experience has many phases and it consists of various factors - perceptual, intellectual, imaginative, emotional and volitional, for instance. In a developed aesthetic experience there are four phases: the preliminary emotion, the adoption of an aesthetic attitude, aesthetic interpretation (the construction of an aesthetic object), and the approval or disapproval of the aesthetic value apparent in the aesthetic object. Ingarden maintains that there are many different aesthetic values (see Art, value of §1).

4 Conclusion

Ingarden contributed to many areas of philosophy, but his most significant achievements include the following: the formulation of conditions for practising an autonomous theory of knowledge, emphasizing the role in it of the immediate 'awareness of an experience'; the distinction between a 'mode of being' and an 'existential moment'; a new theory of ideas as two-sided structures (that is, containing 'constants' and 'variables') which allows him to refute arguments against 'general objects' as self-contradictory; an extensive analysis of the controversy over the existence of the external world as well as an analysis of Husserl's arguments on this subject; an account of the foundations of aesthetics by detailed studies of the structure and modes of existence of the main types of works of art and of the ways in which we perceive them (the theory of aesthetic experience); and the formulation of the concept of a relatively isolated object and its application to the theory of the physical world, the theory of organisms and the problem of the freedom of human action. Ingarden's understanding of the philosophy of language also led him to make a significant attack on the Vienna Circle in Prague in 1934. It was in aesthetics, however, that his influence proved to be deepest and most durable.

See also: Phenomenological movement §2

Translated from the Polish by Piotr Gutowski

ANTONI B. STĘPIEŃ

List of works

Ingarden, R.W. (1957-95) *Dzieła filozoficne (Philosophical Works)*, Warsaw: PWN, 15 vols.(This edition contains all of Ingarden's important works.)

Ingarden, R.W. (1921) 'Über die Gefahr einer Petitio Principii in der Erkenntnistheorie' (On the Danger of a Petitio Principii in the Theory of Knowledge), *Jahrbuch für Philosophie und Phänomenologische Forschung* 4: 545-68.(An attempt to avoid the 'petitio principii' mistake by an appeal to the immediate awareness of experience.)

Ingarden, R.W. (1922) 'Intuition und Intellekt bei Henri Bergson. Darstellung und Versuch einer Kritik' (Intuition and Intellect by Henri Bergson), *Jahrbuch für Philosophie und Phänomenologische Forschung* 5: 286-461.(Contains critical analysis of Bergson's views and considerations on conditions for building a theory of knowledge.)

Ingarden, R.W. (1925) 'Essentiale Fragen. Ein Beitrag zum Problem des Wesens' (Essential Questions: A Contribution to the Problem of Essence), *Jahrbuch für Philosophie und Phänomenologische Forschung* 7: 125-304.(A new theory of ideas with reference to Husserl and Hering.)

Ingarden, R.W. (1929) 'Bemerkungen zum Problem "Idealismus-Realismus"' (Remarks on the Problem of 'Idealism-Realism'), *Jahrbuch für Philosophie und Phänomenologische Forschung*, supplement 10: 159-90. (Analysis of various aspects of the idealism-realism problem.)

Ingarden, R.W. (1931a) *Das literarische Kunstwerk*, Halle an der Saale: Niemeyer, 1960, 1965; trans. G.G. Grabowicz, *The Literary Work of Art*, Evanston, IL: Northwestern University Press, 1973.(Ingarden's classic work which began a revolution in aesthetics. It contains, among other things, analyses of four layers of literary work.)

Ingarden, R.W. (1931b) 'Niektóre założenia idealizmu Berkeleya' (Some Assumptions of Berkeley's Idealism), in *Księga Pamiątkowa Polskiego Towarzystwa Filozoficznego*, Lwów, 215-58.(A continuation of Ingarden 1929, and further preparation to Ingarden 1947-8, 'Controversy Over the Existence of the World'.)

Ingarden, R.W. (1935) 'L'essai logistique d'une refonte de la philosophie', *Revue Philosophique* 60 (120): 137-59.(French version of the debate with the Vienna Circle.)

Ingarden, R.W. (1937) *O poznawaniu dzieła literackiego*, Lwów; extended edn, Warsaw, 1976; German translation, *Vom Erkennen das literarischen Kunstwerk*, Tübingen, 1968; English trans., *The Cognition of the Literary Work of Art*, Evanston, IL: Northwestern University Press, 1973.(Supplements Ingarden 1931a, from the point of view of descriptive psychology.)

Ingarden, R.W. (1947-8) *Spór o istnienie świata (The Controversy Over the Existence of the World)*, vols 1 and 2, Cracow, vol. 3, Warsaw, 1981; vol. 1 partially trans. H.R. Michejda, *Time and Modes of Being*, Springfield, IL, 1964; German trans. *Der Streit um die Existenz der Welt*, Tübingen, 1964, 1965, 1974.(Ingarden's fundamental work. After an outline of his conception of philosophy he begins systematic debate with Husserl's idealism using sophisticated analyses within existential and formal ontology. He gives theory of being, theory of world, theory of pure conciousness and its object.)

Ingarden, R.W. (1958) *Studia z estetyki (Studies in Aesthetics)*, vols 1 and 2, Warsaw; vol. 3, Warsaw: PWN, 1970.(A series of analyses of various works of art (mainly literature, painting and music), aesthetic experience and the problem of values.)

Ingarden, R.W. (1957-8) 'The Hypothetical Proposition', *Philosophy and Phenomenological Research* 18 (4): 435-50.(Excellent analyses of hypothetical propositions.)

Ingarden, R.W. (1960-1) 'Aesthetic Experience and Aesthetic Object', *Philosophy and Phenomenological Research* 21 (3): 289-313.(Analysis of aesthetic experience and aesthetic object as opposed to the work of art.)

Ingarden, R.W. (1962) *Untersuchungen zur Ontologie der Kunst, Musikwerk, Bild, Architektur, Film (Investigations into the Ontology of Art, Musical Works, Pictures, Architecture and Film)*, Tübingen: Niemeyer.(It undertakes the problem of the structure of various kinds of works of art.)

Ingarden, R.W. (1964) 'Artistic and Aesthetic Values', trans. H. Osborne, *British Journal of Aesthetics* 4 (3): 198-213.(It develops the distinction between the values of the work of art and the values of the aesthetic object.)

Ingarden, R.W. (1970) *Über die Verantwortung. Ihre ontischen Fundamente (On Responsibility: Its Ontical Foundation)*, Stuttgart: Philip Reclam Jun.(An analysis of responsibility and its ontic fundaments. An outline of the theory of man.)

Ingarden, R.W. (1971) *U podstaw teorii poznania (At the Foundation of the Theory of Knowledge)*, Cracow: Polish Academy of Sciences.(Critical analysis of psychophysiological theory of knowledge. It states the condition of possibility for a theory of knowledge.)

Ingarden, R.W. (1974) *Wstęp do fenomenologii Husserla (Introduction to the Phenomenology of Husserl)*, Warsaw: PWN.(Short presentation of Husserl's views.)

Ingarden, R.W. (1989) *Wykłady z etyki (Lectures in Ethics)*, Warsaw: PWN.(Lectures concerning the foundations of ethics.)

Ingarden, R.W. (1994) *Frühe Schriften zur Erkenntnistheorie*, ed. W. Galewicz (vol. 6, *Ingarden gesammelte Werke*, ed. R. Fieguth and G. Küng), Tübingen: Niemeyer.(Several articles concernign the conception and condition of possibility of a theory of knowledge.)

References and further reading

Dufrenne, M. (1953) *Phénomenologie de l'expérience esthétique*, Paris, vol. 1, 266-73.(Critical analysis of Ingarden's theory of aesthetic experience and aesthetic object.)

Galarowicz, J. (1983) 'U podstaw sporu Ingardena z Husserlem' (Foundations of the Controversy between Ingarden and Husserl), *Studia Filozoficzne* 3-4: 79-86; 5-6: 83-95.(An analysis of the controversy about the interpretation of Husserl's idealism.)

Gierulanka, D. (1989) 'Kształt metody fenomenologicznej w filozofii Romana Ingardena' (The Shape of Phenomenological Method in Ingarden's Philosophy), in J. Perzanowski (ed.) *Jak filozofować? Studia z metodologii filozofii*, Warsaw, 96-110. (An original analysis of Ingarden's method.)

Graff, P. and Krzemień-Ojak, S. (ed.) (1975) *Roman Ingarden and Contemporary Polish Aesthetics*, Warsaw.

Ingarden, R. (1964) *Szkice filozoficzne Romanowi Ingardenowi w darze*, various eds, Warsaw/Cracow.(Contains essays dedicated to Ingarden as well as a presentation of his philosophy and a bibliography of his works.)

Ingarden, R. (1990) 'Roman Ingarden - metafizyk wolności', *Studia Filozoficzne*, various eds, 2-3: 85-97.(An interesting interpretation of perspectives on metaphysics in Ingarden's late work.)

Kajta, S. (1984) 'Teoria organizmu w ujęciu Romana Ingardena' (Ingarden's Theory of Organism'), *Z zagadnień filozofii przyrodoznawstwa i filozofii przyrody*, vol. VI, Warsaw, 163-267.

Kiereś, H. (1986) 'Recepcja estetyki Romana Ingardena w filozofii analitycznej' (Reception of Ingarden's Aesthetics in Analytic Philosophy), *Zeszyty Naukowe Katolickiego Uniwersytetu Lubelskiego*, 29/2 (1986): 13-17.

Kowalski, J. (1966) 'Romana Ingardena koncepcja materii i formy. Próba analizy krytycznej' (Ingarden's Conception of Matter and Form. A Critical Analysis), *Roczniki Filozoficzne* 14 (3): 99-123.(Detailed analysis of Ingarden's views in comparison with those of Aristotle and Aquinas.)

Paśniczek, J., 'Struktura ontologiczna przedmiotów nie istniejących: Meinong a Ingarden' (Ontological Structure of Non-Existing Objects: Meinong and Ingarden), *Studia Filozoficzne* 4 (1984): 27-41.

Półtawski, A. (1975) 'Ingarden's Way to Realism and His Idea of Man', *Dialectics and Humanism*, 2 (3).(An attempt to find a way to a realistic solution based on Ingarden's later remarks.)

Półtawski, A., 'Roman Ingarden - metafizyk wolności', *Studia Filozoficzne* 2-3 (1990): 85-97.(An interesting interpretation of perspectives on metaphysics in the late Ingarden)

Rudnick, H.H. (ed.) (1990) *Ingardeniana II*, Dordrecht; *Analecta Husserliana* vol. 30. (Contains an international primary and secondary bibliography of Ingarden's works by H.H. Rudnick, J.W. Wawrzycka, 225-96.)

Sosnowski, L. (ed.) (1993) *Estetyka Romana Ingardena. Problemy i perspektywy* (Ingarden's Aesthetics: Problems and Perspectives), Cracow.(An outline of Ingarden's aesthetic views.)

Steinbach, H., 'Ist Ontologie als Phänomenologie möglisch? Kritische Betrachtungen zu Ingardens Existenzialontologie', *Zeitschrift für Philosophische Forschung*, 22/1 (1968): 78-100.

Stępień, A., 'Ingardenowska koncepcja teorii poznania' ('Ingarden's Conception of Theory of Knowledge'), *Roczniki Filozoficzne* 12/1 (1964): 77-92.

Stępień, A., 'Fenomenologia w Polsce' ('Phenomenology in Poland'), *Studia Philosophiae Christianae* 2/1(1966): 29-47.

(1976) 'Bemerkungen zu Ingardens Konzeption der Ontologie', in *S. Thomas Aquinas 700th Anniversary of His Death. Modern Interpretation of His Philosophy*, Lublin, 127-39.(Points out difficulties within Ingarden's ontology and problems connected with relation of ontology to metaphysics.)

Stróżewski, W., 'Gli Studi di estetica di Roman Ingarden', *Rivista di Estetica*, 8/1 (1963): 131-42.

Studia Filozoficzne (1972) 'Fenomenologia Romana Ingardena', special issue, Warsaw.(Contains a bibliography of Ingarden's works from 1915 to 1971, prepared by A. Półtawski, 19-54.)

Szczepańska, A. (1989) *Estetyka Romana Ingardena (Roman Ingarden's Aesthetics)*, Warsaw.(An account of Ingarden's aesthetics.)

Tymieniecka, A.T. (1957) *Essence et existence. Étude à propos de la philosophie de Nicolai Hartman et Roman Ingarden*, Paris.(Comparative analysis of the two authors.)

Tymieniecka, A.T. (ed.) (1976) *Ingardeniana*, Dordrecht; *Analecta Husserliana* vol. 4.(Volume devoted to various aspects of Ingarden's thought.)

Tymieniecka, A.T. (ed.) (1990) *Ingardeniana III*, Dordrecht; *Analecta Husserliana* vol. 33. (Volume devoted to various aspects of Ingarden's thought.)

Tymieniecka, A.T.*et al.* (1959) *For Roman Ingarden: Nine Essays in Phenomenology*, S-Gravenhage.(Articles concerning various aspects of phenomenological thought.)

Wojtysiak, J. (1993) 'R. Ingardena koncepcja ontologii a koncepcja metafizyki tomizmu egzystencjalnego', *Studia metafilozoficzne*, Lublin, vol. 1, 167-211.(This article is a detailed examination of Ingarden's theory of being in comparison to ideas of Polish and American existential Thomists.)

Żegleń, U., 'Próba analizy formalnej fragmentu ontologii R. Ingardena' ('Formal analysis of a Fragment of Ingardens Ontology'), *Roczniki Filozoficzne* 29/1 (1981): 51-63.

Żegleń, U., 'Analiza semantyczna relacji występujBących w ontologii Romana Ingardena' ('Semantic Analysis of Relations in Ingarden's Ontology'), *Studia Semiotyczne* 13 (1983): 35-47.

Inge, William Ralph (1860-1954)

Inge, a philosopher and theologian, was a Christian Platonist. Platonic philosophy emphasized knowledge of necessary truths that it held to be grounded in abstract objects; its deity was a designer limited by the properties of a matter it did not create. Christian theology emphasizes a God who, both Creator and Providence, became incarnate in Jesus Christ. For Platonism, nothing historical could have ultimate significance or importance; for Christianity, the birth, death, and resurrection of Christ have ultimate significance. As his thought developed, Inge increasingly was able to retain much of Platonism while slowly coming to accept the consequences of the Christian emphasis on history.

William Ralph Inge was born in 1860 in Crayke, Yorkshire, and educated at Eton and Cambridge. He was a Fellow of Hertford College, Oxford, from 1889 to 1905, vicar of All Saints' Church, Knightsbridge, from 1905 to 1907, and became a professor of divinity at Cambridge in 1907. In 1911, he was appointed Dean of St Paul's Cathedral, from which he retired in 1934. The remaining twenty years of his life were spent in writing and reflection. He was influenced by the British neo-Hegelians (T.H. Green, Edward Caird, John Caird, F.H. Bradley and Bernard Bosanquet) and, especially, by Baron Friedrich von Hügel, whose emphasis on divine transcendence Inge entirely embraced. A bombastic speaker and writer, always a source of 'good copy', he defined pragmatism as an act of violence exercised by the will upon the intellect to make it accept what we find it helpful to believe, and replied by postcard to an invitation to an elaborate ecclesiastical ceremony planned in his honour, 'What do you think I retired for?'

Inge firmly identified himself as a Christian Platonist, belonging thereby to a tradition that extends from the first Christian century through Augustine to our own times. He held that any adequate view of the world must take modern science into account, and he spoke and wrote on social and moral issues. He consciously endeavoured to combine Platonic philosophy with Christian theology. Plato argued for the existence of Forms - abstract objects that are the truth conditions of necessary truths and the referent and ground of correct standards for moral and aesthetic values. There is, for Plato, also a deity who designs the world from matter whose recalcitrance limits his ability to form it exactly after the pattern provided by the Forms.

Inge held that religion involves belief in an eternal, spiritual, ideal world of which the natural, temporal world is an imperfect copy. Agreeing with Plato that there are eternal, unchangeable standards for truth, goodness and beauty, but holding to Christian monotheism, he held that Truth, Goodness and Beauty are not Forms, but attributes of a personal God who in no way depends for existence or perfection on the existence of a temporal world. The attributes themselves provide the only ontological basis for absolute values, and our awareness of these attributes provides us with the only knowledge of absolute values, which are discovered not made, ends and not means. Our wellbeing resides in awareness of, and conformity to, these ideals. Inge contrasted, then, the temporal world, in which evil is an inseparable condition of good, with the eternal world, which contains no evil. The hope that religion provides envisages our leaving the temporal world for the eternal.

Inge rejected natural theology, holding that there can be no successful inference from things that do exist but might not have existed to a being that exists with necessary independence. He also held that the doctrines of the virgin birth and bodily resurrection of Jesus Christ relate to scientific matters that are not of religious concern. The former entails that Mary bore a child without ceasing to be a virgin, and the latter entails that the dead body of Jesus was placed in a tomb and later restored to life, albeit as a body with powers not possessed by human bodies generally. Inge's Platonist, nonhistorical perspective led him to think of such things as biological and historical, not religious or theological. If Inge sometimes reshaped Platonism to fit his Christianity, here he arguably trimmed his Christianity to fit his Platonism.

In the *Outspoken Essays* (1922), Inge incautiously remarks that should the necessity arise, he could do without the Old Testament prophets and without Christ, and make do with just Plotinus. But in his Gifford Lectures, *The Philosophy of Plotinus*, he writes that:

> So far as I can see, nothing but a personal Incarnation, and the self-sacrifice of the Incarnate, could either adequately reveal the love of God for man, or call forth the love of man for God…. Vicarious suffering - the suffering of the sinless for the sinful - remained a stumbling block for the non-Christian world; and it is only in

this doctrine that the sting of the world's sorrow and injustice is really drawn.

(1918, vol. 2: 227)

The contrasting emphases of these remarks perhaps illustrate the conflicting perspective of a philosophy which can find no ultimate significance in anything not eternally true and a religion whose central doctrines make ineliminable reference to particular events and a particular historical person. At any rate, they illustrate the difficulty one who wishes to combine the two can have in doing so consistently.

KEITH E. YANDELL

List of works

Inge, W.R. (1899) *Christian Mysticism*, London: Methuen.(Contains Inge's Bampton Lectures; emphasizes the importance of religious experience as contrasted to texts and institutions.)

Inge, W.R. (1918) *The Philosophy of Plotinus*, London: Longmans, Green & Company.(Two series of Gifford Lectures, given in 1917 and 1918, by a sympathetic interpreter of Plotinus.)

Inge, W.R. (1919, 1922) *Outspoken Essays*, London: Longmans, Green & Company, two series.(Vintage, iconoclastic Inge.)

Inge, W.R. (1926) *The Platonic Tradition in English Religious Thought*, London: Longmans, Green & Company. (Inge's discussion of the tradition with which he was most in sympathy.)

Inge, W.R. (1930) *Christian Ethics and Moral Problems*, London: Hodder & Stoughton. (Gives Inge's distinctive perspective on issues in ethics.)

Inge, W.R. (1933) *God and the Astronomers*, London: Longmans, Green & Company. (Contains Inge's Wharburton Lectures concerning the problem of time, God in history, God and the world of nature, and the eternal world.)

References and further reading

Elliot-Binns, J.E. (1956) *English Thought 1860-1900: The Theological Aspect*, London: Longmans, Green & Company.(Discusses the historical, philosophical, theological and social setting of English theology in the indicated period.)

Horton, W.M. (date unknown) *Contemporary English Theology*, London: SCM Press.(Treats pre- and post-First-World-War English theology from a perceptive American perspective.)

Langford, T. (1969) *In Search of Foundations: English Theology 1900-1920*, Nashville, TN: Abington Press. (Presentation of English theological thought in its cultural and philosophical context; good bibliography.)

Mozley, J.K. (1951) *Some Tendencies in British Theology*, London: SPCK. (Discussion of British theology from *Lux Mundi* (1889) to the thought of John Oman.)

Innate knowledge

If innate knowledge exists, there must be innate beliefs and those beliefs must count as knowledge. In consequence, the problem of clarifying the concept of innate knowledge divides in two: to explain what it is for a belief to be innate and then to connect that account with a characterization of what knowledge is. Modern biology requires changes in traditional philosophical conceptions of innateness; and two quite different theories of knowledge entail that innate beliefs will often fail to count as knowledge.

1 Innateness

Within some species of bird, individuals produce their characteristic song even if they are reared in silence. Among others, a bird sings its song only if it first hears that song performed. And among still others, a bird produces its characteristic song only if it hears some sort of song or other (Gould and Marler 1991). We are inclined to apply the concept of innateness to the first type of bird and to withhold it from the second. But what of the third? Should we say that these birds learn their songs by hearing any of a set of quite dissimilar songs? Or should we say that the song is innate - that it is ‘in’ the bird, only awaiting an environmental trigger for its release?

Both options seem unsatisfactory. How can a given song be learned from hearing a quite different song, if the target song would have emerged regardless of which song the bird had heard? The option of judging the song innate also has its pitfalls. What does it mean to say that the song is ‘in’ the bird from the start? This spatial metaphor sounds suspiciously like the preformationist doctrines of eighteenth century embryology, according to which foetal development simply involves an increase in size of the fully formed individuals found in newly fertilized eggs. Arguably, the song is ‘in’ the bird from birth no more than adolescent pubic hair is ‘in’ a new-born baby.

It is important not to confuse the claim that the song is ‘in’ the bird from birth with the much less controversial claim that the new-born bird is disposed to acquire the song once the bird passes various developmental landmarks and experiences various environmental cues. The latter claim does not entail the former. Indeed, the latter formulation offers a completely de-natured reading of the idea of innateness; it becomes trivially true that all beliefs are innate. Surely this formulation short-circuits the controversy rather than solving it.

(In *Notes on a Certain Program*, Descartes suggests that ideas are innate in precisely the same sense that diseases are sometimes innate; this does not mean that foetuses suffer from diseases *in utero*, but that individuals are born with a certain ‘disposition’ to contract them. The challenge to innatism, laid down by Locke in the *Essay Concerning Human Understanding*, is to explain what ‘disposition’ means so that some traits, but not others, turn out to be innate. Leibniz takes up this challenge in the *New Essays*.)

Contemporary biology does not exploit the philosopher’s traditional distinction between innate and learned. First of all, the biological opposite of innate is not ‘learned’, but ‘acquired’. Learning is one way to acquire a characteristic, but there are others; sunburns are acquired, but they are not learned. It is worth asking what distinguishes learning (and *mis*learning) from other processes of belief acquisition; if beliefs could be acquired by swallowing pills, arguably that would not count as learning. In any event, the present point is that biologists do not think of ‘innate’ and ‘learned’ as exhaustive alternatives.

Perhaps the most important contrast between the biological and the traditional philosophical concept is that biology does not treat the difference between innate and acquired as a dichotomy; rather, it is a matter of degree. Traits differ with respect to how dependent their emergence is on environmental details. The three groups of birds described before can be arrayed on a continuum; there is no need to draw a line that separates two of them from the third.

It is also important to distinguish the initial appearance of a trait from its subsequent maintenance. A trait may emerge across a wide range of environments and yet be modifiable once in place. For example, an Egyptian vulture, when first confronted with an ostrich egg and a stone, will break the egg with the stone. However, if the vulture repeatedly discovers that the eggs it breaks open are empty, eventually it will stop breaking eggs (Pulliam and Dunford 1980). In this case, the behavioural disposition might be termed innate. All the same, ‘innate’ does

not mean ‘unmodifiable’.

Yet another gap between the traditional philosophical concept and modern biology concerns the issue of universality. Philosophers often expect innate beliefs to be universal in our species; they are part of ‘human nature’ and so should be present in all individuals who have enjoyed a normal development. (Thus we find Locke, in the *Essay* (bk I, ch. III, §§8-18), arguing that the idea of God is not innate because it is not found in children and ‘savages’.) A related expectation is that if two people have a trait, then the trait must be innate for both if it is innate for either.

Neither of these assumptions is unproblematic. The biologist’s concept of a ‘norm of reaction’ describes how the phenotype which a genotype will develop depends on the environment it experiences. For example, consider a particular fruitfly genotype; let us ask how that genotype’s attainment of a particular phenotype (for example, bristle number) depends on some environmental variable (for example, nutrition). Different genotypes often exhibit different degrees of environmental contingency with respect to the same phenotype. One fruitfly genotype may produce the same bristle number regardless of variation in nutrition, whereas another may produce different numbers of bristles in response to different amounts of food. Because of this, it is quite possible for two flies to have the same number of bristles, but to have achieved that result by quite different developmental pathways. Bristle number may be ‘innate’ for one and ‘acquired’ for the other.

In view of all this, it is a reasonable hypothesis that the most that can be salvaged from the ancient concept of innateness is this: *a phenotypic trait is innate for a given genotype if and only if that phenotype will emerge in all of a range of developmental environments*. What counts as the appropriate range of environments is left open in this proposal. Perhaps there is a uniquely correct answer to this question; then again, maybe the range is determined pragmatically. It is difficult to see how the latter conclusion can be evaded.

2 Innateness and the a priori

The above definition opens a large gap between the concepts of innateness and a prioricity. A proposition is a priori if it can be known or justified independently of experience (see A priori). A prioricity does not concern what causes an organism to formulate a belief; rather, it has to do with the status the belief enjoys once it is formulated. Arguably, the belief that all bachelors are unmarried is not innate; it emerges in some social and linguistic environments, but not in others. Still, once an individual has this belief, it may be true that the belief can be justified without appeal to experience.

Conversely, it is also possible for a belief to be both a posteriori and innate (see A posteriori). Perhaps human cognitive development inevitably leads people to believe that a physical world exists external to their own minds. This is as natural and invariant a developmental outcome as the acquisition of pubic hair during adolescence. Yet, it remains for the philosopher to judge whether solipsism is a priori false. Perhaps our natural anti-solipsism is either unjustifiable or is justifiable as an inference to the best explanation; maybe the postulate of an external world makes sense because it explains regularities that obtain in the flow of experience. If so, we would have here the case of an innate belief that is a posteriori (see Inference to the best explanation).

3 Knowledge

If we have beliefs whose emergence is relatively invariant across a range of developmental backgrounds, will those beliefs count as knowledge? To begin with, if *S* knows that *p*, then *p* must be true. Beyond that, what does knowledge require? Let us explore the suggestion that *S*’s belief that *p* counts as knowledge only if *S* is justified in believing *p*.

The justification that *S* has for believing *p* presumably must come from other beliefs (or thoughts or sensations) that *S* has. In particular, it is not enough that those other beliefs (or other items) bear some abstract relation of justification to the proposition *p*; after all, it might be true that *S* believes *q* and that *q* is evidence for *p* even though *S* believes *p* for no reason at all or for bad reasons. What is required is that *S*’s belief in *p* be related ‘in the right way’ to *S*’s justifying belief in *q*.

What might this right way be? One possibility is that *S*’s belief in *q* causes *S* to start believing *p*. Another is that *S* continues to believe in *p* because *S* believes *q*. A third is that *S*’s belief in *q* makes it harder for *S* to stop believing

p. The list could continue (see Justification, epistemic; Knowledge, concept of §§1-6).

Regardless of how this condition is interpreted, it seems clear that innate true beliefs can fail to count as knowledge. Consider a belief that people inevitably have and cannot abandon; perhaps our spontaneous anti-solipsism is an example. Such propositions may admit of inventive justifications, but this does not mean that philosophical pronouncements can influence whether ordinary folks believe in an external world.

Let us turn, instead, to a concept of knowledge that abandons the requirement of justification. Reliability theories of knowledge demand that the act of believing should be related to the world in the right way. If *S* knows that *p*, then it must be true that, in the circumstances that *S* occupied, *S* would not have believed *p* unless *p* had been true. If *S* knows that *p*, then the requirement is that the probability that *p* is true given that *S* believes that *p*, is equal to 1.0. *S* must be a perfectly reliable detector of the truth of *p*, if *S*'s belief that *p* is to count as knowledge (Dretske 1981) (see Reliabilism).

It is easy to see why an innate belief can evolve so as to violate the requirement that the probability that *p* is true given that *S* believes that *p*, is equal to 1.0. For example, suppose that snakes are usually, although not always, dangerous. An organism that confronts a snake thus faces a problem of decision-making under uncertainty. It can believe without further ado that the snake is dangerous; or, it can try to find some further observational evidence on which to base its decision. In the former case, the organism might have the innate belief that snakes are dangerous. In the latter, its opinion will depend on the details of present and past experience.

Imagine that the organism in question is not very good at determining from observable cues whether a snake is dangerous. Imagine further that the organism does better by erring on the side of caution (Stich 1990). If the snake is dangerous, it is very important to believe that this is so; but if the snake is not dangerous, it is of much smaller moment whether one realizes this fact. These assumptions all favour the evolution of the innate belief that snakes are dangerous. Innateness will be a better strategy than learning in this instance (Sober 1994).

Notice that the factors that predict the evolution of innateness provide no guarantee of the reliability of the resulting beliefs. The organism believes of each snake encountered that it is dangerous. The organism therefore has true beliefs only as often as snakes in fact are dangerous. This need not be a terribly high frequency; it clearly need not equal 100 per cent. For this reason, it is easy to see how innate beliefs can fail to count as knowledge from the point of view of the reliability theory of knowledge. Indeed, there is no guarantee that the belief must be true more often than not; pervasive error may be a consequence of the evolution of innate beliefs.

See also: Cognitive development; Language, innateness of; Nativism; Rationalism §2

ELLIOTT SOBER

References and further reading

Descartes, R. (1648) *Notes on a Certain Programme*, in *The Philosophical Works of Descartes*, trans. E.S. Haldane and G.R.T. Ross, Cambridge, Cambridge University Press, 1931, 431-50.(Descartes responds to various misunderstandings of his views, including his position on innate ideas.)

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(Elaboration of the reliability theory of knowledge.)

Fodor, J. (1981) 'The Current Status of the Innateness Controversy', in *Representations*, Cambridge, MA: MIT Press, 257-316.(Argues that the dispute between empiricists and rationalists should be understood in terms of whether various simple beliefs are learned or triggered.)

Gould, J. and Marler, P. (1991) 'Learning by Instinct', in D. Mock (ed.) *Behavior and Evolution of Birds*, San Francisco, CA: Freeman, 4-19.(Describes a variety of learning strategies that birds use to acquire their songs.)

Leibniz, G.W. (1703-4) *New Essays Concerning Human Understanding*, trans. with notes by A.G. Langley, La Salle, IL: Open Court, 3rd edn.(Point-by-point commentary on Locke's essay by the greatest exponent of continental rationalism.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press; New York: Oxford University Press.(The founding document of British empiricism, accessible to the general reader.)

Pulliam, H. and Dunford, C. (1980) *Programmed to Learn*, New York: Columbia University Press.(Argues that different learning strategies will evolve depending on how the environment is structured.)

Sober, E. (1993) *Philosophy of Biology*, Boulder, CO: Westview Press.(Clarifies concepts such as norm of reaction and heritability.)

Sober, E. (1994) 'The Adaptive Advantage of Learning and *A Priori* Prejudice', in *From a Biological Point of View: Essays in Evolutionary Philosophy*, Cambridge: Cambridge University Press.(Presents an evolutionary model for predicting when a proposition will be an innate belief and when it will depend on experience for its emergence.)

Stich, S. (1975) *Innate Ideas*, Berkeley, CA: University of California Press.(Anthology of historical and contemporary essays, some of which address issues concerning Chomskian linguistics. Of special interest are the essays by A. Goldman relating the causal theory of knowledge to the question of whether innate beliefs can count as knowledge; by W. Hart, arguing that innateness is irrelevant to the issue of a priori knowledge; and by S. Stich, which proposes a clarification of what innateness means.)

Stich, S. (1990) *The Fragmentation of Reason*, Cambridge, MA: MIT Press.(Argues that evolutionary considerations make it plausible to suspect large amounts of irrationality and unreliability in belief formation.)

Innocence

In its most general sense, innocence refers to the state of being without sin. A more restricted meaning is attributed to the word in the legal sphere, where people who are found not to be guilty of a particular crime are described as innocent. In official teaching of the Roman Catholic church, all direct killing of the innocent is forbidden. Here the word refers to people who are not harmful. Nowadays, this can be taken to mean 'non-aggressors'. According to Rousseau, humans possess original goodness. Corrupting influences come from outside them.

Innocence, in perhaps the oldest sense of the term, is the state of being sinless or unacquainted with evil (see Evil; Sin). Within various branches of Christianity such is believed to have been the original state of the first humans. However, in view of the fact that nobody on this planet appears to be in such an immaculate condition now, the term 'innocence' usually has a more restricted meaning. It can be used thus in the sphere of ethics, although it would seem to be more commonly employed in legal circles, where people who are found not to be guilty of a particular crime are described as innocent. The fact that they may have committed innumerable other misdemeanours which do not fall within the ambit of the law is of no consequence in this regard. It is often said in such cases that the people concerned are not guilty of wrongdoing because they either did not perform a certain act, or, having indeed performed it, were not blameworthy in so doing. It is worth noting, however, that, in courts of law, the label 'innocent' can be, and often is, applied to people who have acted in accordance with the requirements of a particular law, even in cases in which, from an ethical point of view, breaking the law would have been the right thing to do. Nevertheless, there can also be innocence in the moral sense in such cases if the people concerned sincerely believe that they have acted rightly in obeying the law, unless, of course, their belief is based on culpable ignorance (see Responsibility §2).

Having said that, it is worth noting that the term 'innocence' is, in fact, sometimes used to describe the state of people who are ignorant of certain evils. The adjective 'innocent' occasionally replaces 'naïve' in such cases. In spite of what has already been said about culpable ignorance, such innocence can be an endearing, although also dangerous, quality, as is often the case with small children who are open and trusting, never having been exposed to more than a few minor forms of malevolence. Here we can say that the state of ignorance or innocence results from a lack of experience of the harsh realities of life.

In some spheres of ethics, most notably in Roman Catholic moral theology, the term 'innocence' and its adjective have been attributed a good deal of importance in debate about the rightness or wrongness of killing people. According to official Roman Catholic teaching, no direct killing of the innocent can ever be condoned. Here again a restricted meaning is adopted. In current Catholic literature the word 'innocent' is used to refer to non-aggressors, an obvious example being noncombatants in time of war. One may kill an aggressor in a just war, or, even outside the war situation, in an act of self-defence or when acting in defence of a third party, provided that there is a proportionate reason for doing so. Non-aggressors (innocent people), however, may be killed only indirectly, in accordance with the requirements of the principle of double effect (see Double effect, principle of). This might happen during a just war, as a result of the bombardment of an enemy installation, near or in which are some noncombatants. Their deaths could be accepted as a regrettable but unavoidable side effect of the bombardment, provided that killing them was not the aim of those effecting the bombardment, that the good intended was not brought about by means of those deaths, and there was a proportionate reason for effecting the bombardment in the first place (see War and peace, philosophy of §§2-3).

The same kind of thinking is applied in the sphere of medical ethics (see Medical ethics §§2, 4). Official Catholic teaching prohibits all direct abortion, but permits such therapeutic procedures as hysterectomies performed on pregnant women suffering from cancer of the uterus. The foetuses - described as innocent persons - are aborted indirectly. Something similar can be said about official Catholic teaching regarding euthanasia. Even a person who asks to be killed remains innocent in the sense under discussion, and may not, therefore, be killed directly. Analgesics, however, may be administered to such a person to relieve intolerable pain, even though it is foreseen that they will have the side effect of shortening their life. Again, in this last case, an innocent person is killed, but indirectly.

A lack of innocence in this sense is also attributed to at least some people found guilty of murder. Official Roman Catholic teaching permits capital punishment in such cases, but only if less violent ways of protecting human life and public order from such aggressors are unavailable.

The unrestricted meaning of innocence would seem to be what is referred to in the writings of Jean-Jacques Rousseau. In his opinion (1762), the fact that we see that people are wicked in no way conflicts with the fact that humans are naturally good, possessing original goodness. In their original state, there is not a single vice in their hearts, no perversity at all. Evil must have an external source. In bringing up our children, therefore, we need, according to Rousseau's way of thinking, to keep away from them all corrupting external influences in order that nature may be allowed to take its course. Children must have a natural education, that is, an education that interferes as little as possible with their free, natural development.

BERNARD HOOSE

References and further reading

McCormick, R.A. (1989) *The Critical Calling: Reflections on Moral Dilemmas Since Vatican II*, Washington, DC: Georgetown University Press, ch. 12, esp. 220-32.(A detailed examination of the Roman Catholic teaching prohibiting direct killing of the innocent.)

Roche, K.F. (1974) *Rousseau: Stoic and Romantic*, London: Methuen, 39-51. (Explanation of Rousseau's thought, as presented in his *Émile*.)

Rousseau, J-J. (1762) *Émile: ou, de l'éducation*, trans. A. Bloom, *Emile: or, On Education*, Harmondsworth: Penguin, 1991.(Rousseau claims that every child is born innocent.)

Seago, P. (1994) *Criminal Law*, London: Sweet & Maxwell, esp. 28-101. (A consideration of the general principles of criminal liability.)

Institutionalism in law

'Institutionalism' is the name for an approach to the theory of law worked out in the late nineteenth and early twentieth centuries by a number of scholars from continental Europe, working mainly in independence from each other. Their common characteristics can be stated only in rather generic and negative terms. They are all critical of statalism, that is, too readily identifying law and state; of voluntarism, that is, treating will as an essential element of law; and of normativism, defining law as a body of norms. In positive terms, they have in common a generic emphasis on the social character of law, and a sense of the need to take a view of legal experience broader than that defined by its traditional boundaries, and to extend the 'official' catalogue of the sources of law.

Nowadays, there is new talk of institutionalism in the philosophy of law, but in a different sense, with reference to the idea of law as made up of institutional facts, that is, facts whose meaning depends on norms. This neo-institutionalism is normativist, analytical and hermeneutic in approach and has only the most tenuous links with classical institutionalism.

1 1 Institutions and institutionalism in law

Legal institutionalism conceives of law as a kind of 'institution'. No less evasive a definition could be given without opening the question of the rival merits of the various institutionalist theories and of the notion of 'institution' that each adopts. 'Institution' is an intrinsically ambiguous term, used in a notably wide variety of senses, to the point of appearing useless to many thinkers. As well as being a term in general use, it is also a technical term in social science, particularly in sociology, where it is often simply a synonym for social organization or for some element of social structure; elsewhere, it has come to be used, in an equally general way, to denote any sort of conduct or belief sanctioned and practised in a stable way by a community. In a more restricted sense again, 'institution' is used to signify the normative aspect of social organizations; in this sense, for example, 'marriage' would be the institution corresponding to the family as a social order.

This variety of meanings is reflected, obviously enough, in law. In legal culture, however, yet other senses of the term 'institution' are to be found. This expression is customarily used to signify appointment to an office, or constitution of an heir, or, in its substantive use, an entity or organization personified by the law. In the plural, one speaks of institutions to signify the fundamental elements of a discipline (here remembering the *Institutiones* of the Roman jurist Gaius). Jurists in some linguistic traditions are accustomed to make a (somewhat variable) distinction between 'institution' and 'institute', often using the former to refer to underlying social regularities (for example, capitalist ownership) while the latter denotes a standard set of legal relations conceived in abstract terms (for example, property). This distinction highlights the normativism and formalism implicit in common juristic thought. It is no accident that it is a distinction given prominence by the neo-institutionalism discussed in §5 below.

'Classical' legal institutionalism is however more indebted to the sociological than to the juridical usage of the term 'institution'. Historically, institutionalism was a family of jurisprudential approaches that arose in the late nineteenth and early twentieth century in France and Italy. The historical phenomenon that most influenced its theoretical development was the swift transformation of the nineteenth-century liberal state through the rise of intermediate groups between individuals and the state, such as industrial corporations and trade unions. The concept 'institution', because of its very vagueness and ambiguity, has been found serviceable, in law and elsewhere, as the key to analysis and even justification of the changes in social relationships and power-relations that have taken place (see Social theory and law §2).

Thus it is easy enough to determine the extension of the term 'legal institutionalism' by pointing out the approaches that have been called 'institutionalist'. What is not so easy is to determine its intension, by identifying the characteristics common to the various legal institutionalisms; some even doubt the possibility of doing so. A common characteristic does however exist, namely, a generically realist tendency, a propensity to regard the law as a social phenomenon (see Legal realism §2). The use by institutionalist approaches of the terminological apparatus thrown up by the nascent social sciences should not however deceive us into supposing that the theories in question sought to establish an empiricist view of law; on the contrary, legal institutionalism arose by and large from ethico-political impulses (this is particularly true of Renard and Gurvitch; see §3), and, even when driven by epistemological tenets, as in the case of Santi Romano (see §4), it was inspired more by a formalistic legal-science

model than by sociological empiricism. Beyond the rather weak unifying element of a generic realism, one can go no further. For the institutionalist authors have neither a shared cultural formation nor shared philosophical reference points in the way of common ontological or epistemological ideas; nor have they even shared ethico-political attitudes.

2 2 Maurice Hauriou

Maurice Hauriou, French theorist of public law, has come to be regarded as the founding legal institutionalist. He set himself against two targets: extreme subjectivism fixated on the notions of will and of contract (here the critique is of Rousseau and certain nineteenth-century German theories of public law), and extreme objectivism premised on the notion of the rule of law (here criticized are Duguit and Durkheim). (see Durkheim, É.; Rousseau, J.-J.). For Hauriou, institutions are both social facts and legal configurations; they both pre-exist and produce law. They can be either institution-persons (corporate bodies such as the state, associations, companies) or institution-things (such as the standing rules of the law). His earlier work stresses power as the unifying factor of institutions, and he defines them in terms of power-relations (Hauriou 1910); later, he accords greater importance to what he calls the '*idée directrice*' ('directing idea') or '*idée d'oeuvre*' ('operative idea'). 'Among those interested in the idea there are produced manifestations of communion directed by its organs of authority and regulated by a definite procedure' (Hauriou 1925). The elements of the institution are thus the '*idée directrice*', the power organized through realization of the idea, and the communal activities which are produced in the social group as reflections of the idea and its realization. The profoundly mysterious notion of the 'directing idea' belongs to an organicist social theory, permeated with vitalism (the subtitle of Hauriou's best-known work is much to the point: 'an essay in social vitalism'); a theory in which the idea of institution remains nebulous and with the vaguest of boundaries. Hauriou's philosophical ideas are, to be sure, eclectic and unsystematic (Bergson's influence can be seen), and in any event they have the sound of *ex post* justifications for a theory that remains at base an essentially juristic one.

3 3 The development of French institutionalism: Renard and Gurvitch

Hauriou's work has aroused few echoes in common law countries, or in the German-speaking countries; yet it has been highly influential in the Hispanic and Italian worlds, and of course in France. In France, Georges Renard and Georges Gurvitch were both pupils of Hauriou's who reacted against voluntaristic dogmas and juridical individualism and statalism, and shared an ambition to broaden the horizons of legal experience.

A critic of voluntarism and of individualism, Renard (1876-1943) shared Hauriou's opposition to identification of law with enacted law. The essence of institutions in his view lies in the way in which activities cohere in the pursuit of a common aim, with a preordained internal structure; such a structure is always based on authority-relationships (Renard 1930). He tries to subsume this institutionalist conception of his, standing as it does midway between sociology and jurisprudence, under a Thomistic metaphysics. This leads Renard on the one hand to interpret the social philosophy of Aquinas as an institutionalist theory before its time, while on the other hand treating institutions as all essentially directed towards realization of the common good (see Thomism; Natural law §§3-4).

Gurvitch (1894-1965), a Russian lawyer who became a naturalized French citizen, was a sociologist of law with social-democratic convictions. At the centre of his thought lay his idea of 'normative fact', which he treated as equivalent to Hauriou's 'institution'. Law was not founded on will or on some norm or norms, but on the very fact of union or of associational community, that is, on a 'normative fact'. Gurvitch (1932) also brings to bear on this pluralistic and anti-formalist conception of law a thoroughly value-laden philosophical approach, which he calls 'transpersonalism'. This is based on an exaltation of 'social law' and 'social rights', whose essential characteristic is found in integrative relationships, and which stand in opposition to individual rights, with their contrasting character of being coordinative and subordinative. Social rights, in Gurvitch's definition, are autonomous communitarian rights, which provide an objective integration of every really active collectivity, whether or not organized.

4 4 Italian legal institutionalism: Santi Romano

Santi Romano (1875-1947) was an Italian public lawyer and author of an institutionalist theory which enjoyed

great currency in his own country, and was applied in several branches of law, including international law, ecclesiastical law and labour law. His thought was little if at all influenced by Hauriou's work, of which he had some knowledge and appreciation, but which he thought rather limited in scope and insufficiently jurisprudential. He was a pure jurist and the creator of a general theory that was rigorously anti-philosophical and pure of any sociological influence. Moreover, as against the French institutionalists, he was a critic of natural law and a defender of the thesis of the separation between law and morals; in this sense he can be considered a legal positivist (see Legal positivism §1).

Romano's polemical writings have as their targets normativism and the common definitions of law that are based on examples drawn from private law and hence inadequate for the explanation of the other branches of law (Romano 1917-18). This gives rise to his definition of law as a juridical order, a notion which in turn leads back to that of institution. Romano takes 'institution' to signify any social entity or body. This entails an extreme form of legal pluralism, and leads Romano to the point of treating as legal even organizations that are prohibited by state-law, and indeed to considering revolutionary organizations as juridical orders. Romano, for whom every social organization, however ephemeral and transitory, counted as a juridical order, tended to represent all social relations in a legal guise. Such an approach in the final analysis has the effect of blanking out social conflicts, since these come to be depicted as instances of relationships between orders, and the orders in their turn are seen as having a spontaneous tendency toward orderliness.

Romano's institutionalism won the approbation of the German legal thinker Carl Schmitt, theorist of National Socialism, who saw in it an exemplification of his own conception of law as a concrete order. This conception, along with normativism and decisionism, in Schmitt's view constituted the three basic types of pure legal thought (Schmitt 1934). Extremist in thought as he was, Schmitt saw with perfect clarity that institutionalism, if carried to its logical conclusions, could not but end up in a radical realism in which the reality of the law becomes reduced to mere relations of power and of decision-making.

5 5 Contemporary legal institutionalism: MacCormick and Weinberger

The classical institutionalist theories were all ethico-political and were characterized in general by a holistic and organicist view of society and an aversion from conflict-models of society and law. Each emphasized the values of order, of neutralization of conflicts, of harmony and social integration.

Taken as legal theories in the strict sense, the varieties of institutionalism have the considerable merit of having neutralized the extremer versions of opposed opinions, and in particular normativism, though they did not refute it outright. In truth, the anti-normativist polemics convince only in relation to a narrow Austinian notion of the legal norm, understood as the imperatival expression of an act of human will (see Austin, J.). This notion is now completely discredited. Institutionalism in its moderate, non-Schmittian version cannot coherently be completely anti-normativist. For the very ideas of social organization on which it is founded presuppose norms that regulate and empower individuals (see Norms, legal). The anti-normativist polemics of the institutionalists have without doubt contributed to bringing legal theory up to date. In particular, the emphasis on the notion of organization has brought about general acceptance of a conception of the law as a normative order, and has led to the elaboration of a typology of norms better suited to representing law in terms of structured ordering, giving due recognition to organizational norms, or secondary norms.

Today, classical institutionalism has lost its innovative drive but jurists do still use many ideas drawn from institutionalist theories to give theoretical respectability to their own doctrinal writings. Suffice it here to mention the case of the *lex mercatoria*, or 'New Law Merchant', the so called 'common law of commerce', which has come to be construed as a form of legal order coexisting with the laws of the states and with international law (Osman 1992).

But at present there is also new talk of institutionalism on the part of two contemporary philosophers of law, Neil MacCormick in Scotland and Ota Weinberger the Czech scholar (see Weinberger, O.). MacCormick and Weinberger's neo-institutionalism however has declaredly little in common with classical institutionalism, on several counts. In truth theirs is a conception of law worked out not from the standpoint of the pure jurist nor from that of the sociologist of law, but from that of legal philosophy, indeed that of philosophers with an analytico-linguistic sensitivity and training. Its basis is a confessedly empiricist metaphysics. It is neither

anti-statalist nor anti-formalist in its treatment of the sources of law; moreover, far from criticizing normativism, it treats norms as being at the core of the concept of an institution; and finally, it is a positivistic theory of law, critical of natural law theory (in this having affinity with the ideas only of Romano among the classical school).

The institutionalism of MacCormick and Weinberger starts out from the notion of 'institutional fact' and the related opposition between 'brute facts' and 'institutional facts' (Anscombe 1958; Searle 1969). The law is composed of institutions (contract, wills, statutes and so on) created through the implementation of legal rules. Such institutions are nothing other than abstract concepts through which it is possible to ascribe meaning to, and to interpret, congeries of human actions. Institutions in this sense belong, however, to a philosophical and semiotic notion, and have little in common with the various sociological notions of institution. The ideas of MacCormick and Weinberger (1986) do however disclose an institutional aspect in the classical sense to the extent that they give emphasis to the relationship that always exists between institutions in the philosophical sense and institutions as sociological realities. According to these authors the law is in fact always at the same time (1) an ensemble of abstract meaning-concepts; (2) an ensemble of institutional acts and facts understandable only in the light of the network of rules constitutive of the concepts; and (3) an ensemble of organizations in the sociological sense. However, to grasp the reality of the legal it is indispensable to adopt a hermeneutic outlook whereby it is possible to coordinate the contributions of lawyers, sociologists and philosophers of law (see Legal hermeneutics).

Weinberger's work in particular brings together the notions of institutions and institutional facts and squares them with a more all-embracing philosophico-jurisprudential view, comprehending anthropology, ontology and a formal theory of action.

See also: Hart, H.L.A.; Law, philosophy of; Legal concepts; Social theory and law §§1-3

Translated by D.N. MacCormick
ANNA PINTORE

References and further reading

Anscombe, G.E.M. (1958) 'On Brute Facts', *Analysis* 18: 69-72.(A brilliant and concise pioneering treatment of the distinction between brute facts and institutional facts.)

Biscaretti di Ruffia, P. (ed.) (1977) *Le dottrine giuridiche oggi e l'insegnamento di Santi Romano*, Milan: Giuffré.(Particularly useful for anyone interested in the 'fate' of Romano's ideas in Italy; a collection of papers by philosophers and legal theorists, constitutional lawyers, administrative lawyers, international lawyers and ecclesiastical lawyers.)

Broderick, A. and Welling, M. (eds) (1970) *The French Institutionalists: Maurice Hauriou, Georges Renard, T. Delos*, Cambridge, MA: Harvard University Press. (Scholarly studies of French institutionalists from a standpoint broadly sympathetic to theirs; a helpfully constructive critique.)

Gurvitch, G. (1932) *L'Idée du droit social. Notion et système du droit social. Histoire doctrinale depuis le XVII^e^ siècle jusqu'à la fin du XIX^e^ siècle*, Paris: Sirey. (Contains a notable historical section reconstructing the idea of a social right from the time of Grotius onwards.)

Hauriou, M. (1910) *Principes de droit public (Principles of Public Law)*, Paris: Sirey, 2nd edn, 1916.(This little book on principles of public law is a student's introduction to the subject, in which Hauriou's theoretical approach is also introduced.)

Hauriou, M. (1925) 'La Théorie de l'institution et de la fondation. Essai de vitalisme social' (On the Theory of the Institution and the Foundation), *Cahiers de la nouvelle journée* 4.(This essay is Hauriou's best-known theoretical statement; it is aimed more widely than at a purely juristic audience.)

Jennings, W.I. (1933) 'The Institutional Theory', in W.I. Jennings (ed.), *Modern Theories of Law*, London: Oxford University Press, ch. 5: 68-85.(An account of Hauriou and his school by the leading contemporary English public lawyer, considerably, and perhaps excessively, stressing the Catholic element in institutionalist thought.)

MacCormick, N. and Weinberger, O. (1986) *An Institutional Theory of Law: New Approaches to Legal Positivism*, Dordrecht: Reidel.(This is the 'manifesto' of neo-institutionalism, comprising a collection of essays independently written by the two authors, together with a jointly-written introduction.)

Osman, F. (1992) *Les Principes généraux de la lex mercatoria: contribution a l'étude d'un ordre juridique anational*, Paris: Librairie Générale de Droit et de Jurisprudence.(An application of Santi Romano's pluralistic theory of legal orders to the concept of the *lex mercatoria* or 'Law Merchant' in its modern form.)

Renard, G. (1930) *La Théorie de l'institution. Essai d'ontologie juridique*, Paris: Sirey.(A prolix and sprawling

piece of work, not an easy read.)

Romano, S. (1917-18) *L'ordinamento giuridico (Juridical Order)*, Florence: Sansoni, 2nd edn, 1946; 3rd edn, 1967.(This is Romano's chief work on 'juridical order', a concise and beautifully written account of his theory of the institution and its place in our understanding of legal order. The second edition is identical with the first, save for an up-dating note; the third is simply a reprint of the second.)

Schmitt, C. (1934) *Über drei Arten des rechtswissenschaftlichen Denkens*, Hamburg. (A lucid and straightforward exposition of the fundamentals of Schmitt's decisionism, through an account of what he took to be the three basic types of purely legal thinking. Indispensable reading.)

Searle, J.R. (1969) *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.(This is the 'classical' account of institutional facts and constitutive rules.)

Intensional entities

Intensional entities are such things as concepts, propositions and properties. What makes them 'intensional' is that they violate the principle of extensionality; the principle that equivalence implies identity. For example, the concept of being a (well-formed) creature with a kidney and the concept of being a (well-formed) creature with a heart are equivalent in so far as they apply to the same things, but they are different concepts. Likewise, although the proposition that creatures with kidneys have kidneys and the proposition that creatures with hearts have kidneys are equivalent (both are true), they are not identical. Intensional entities are contrasted with extensional entities such as sets, which do satisfy the principle of extensionality. For example, the set of creatures with kidneys and the set of creatures with hearts are equivalent in so far as they have the same members and, accordingly, are identical. By this standard criterion, each of the following philosophically important types of entity is intensional: qualities, attributes, properties, relations, conditions, states, concepts, ideas, notions, propositions and thoughts.

All (or most) of these intensional entities have been classified at one time or another as kinds of universals. Accordingly, standard traditional views about the ontological status of universals carry over to intensional entities. Nominalists hold that they do not really exist. Conceptualists accept their existence but deem it to be mind-dependent. Realists hold that they are mind-independent. Ante rem *realists hold that they exist independently of being true of anything;* in re *realists require that they be true of something.*

1 History

Contemporary use of the term 'intension' derives from the traditional logical doctrine that an idea has both an extension and an intension (or comprehension). This doctrine is explicit in most modern logicians of the nineteenth century, implicit in many medieval logicians and, arguably, present in Porphyry and Aristotle. Although there is divergence in formulation, most of these thinkers accept that the extension of an idea (or concept) consists of the subjects to which the idea applies, and the intension consists of the attributes implied by the idea. The Port Royalists (see Arnauld 1662) tell us, 'the comprehension of the idea of triangle includes extension, figure, three lines, three angles, and the equality of these three angles to two right angles, and so forth'. This yields the 'law of inverse ratio': the larger the intension, the smaller the extension. (Note that these traditional doctrines were expressed with the use of plurals and without explicit commitment to classes or sets.) If the extension consists of the individuals (as opposed to species) to which an idea applies, it is evident that two ideas could have the same extension but different intensions. This is the tie to Russell's subsequent notion of an intensional propositional function (1910-13) and leads to our contemporary usage of 'intensional entity' and 'extensional entity'.

In contemporary philosophy, it is linguistic expressions, rather than concepts, that are said to have intensions and extensions. The intension is the concept expressed by the expression, and the extension is the set of items to which the expression applies. On the standard picture, that set is the same as the set of items to which the concept applies, hence the standard view that intension determines extension. This usage resembles Mill's use of 'denotation' and 'connotation' and Frege's use of '*Bedeutung*' and '*Sinn*' (see Mill, J.S. §§2-3; Frege, G. §3).

The systematic study of intensional entities has been pursued largely in the context of intensional logic; that part of logic in which the principle of substitutivity of equivalent expressions fails. For example, 'Necessarily, creatures with kidneys have kidneys' is true whereas 'Necessarily, creatures with hearts have kidneys' is false. Following Frege (1892), Whitehead and Russell (1910-13), Carnap (1947) and Church (1951), the now standard explanation of this failure of substitutivity is this: the truth-value of these sentences is determined, not (just) by the extension of 'creature with kidney' and 'creature with heart' (that is, the set of creatures with kidneys) but (also) by the intensions of these expressions, that is, by the intensional entities which they express (the concept of being a creature with a kidney and the concept of being a creature with a heart). Because these intensions are different, the sentences can have different truth-values.

This account yields a method for determining precise intensional identity conditions: intensional entities should be as finely discriminated as is necessary for explaining associated substitutivity failures in intensional logic. For example, this method has led most philosophers to hold that logically equivalent concepts can be distinct, and even that in a correct definition the definiens and the definiendum can express different concepts. This is not to say that the properties corresponding to those concepts would be distinct; on the contrary, the definition would be correct

only if the properties were identical. Such considerations lead to the view that there are also more coarsely grained intensional entities (properties, relations, conditions in the world). Independent arguments in metaphysics, epistemology, philosophy of science and aesthetics seem to support the same conclusion (see, for example, Armstrong 1978, Bealer 1982, Lewis 1986). On this view, properties and relations (as opposed to concepts) play a primary role in the non-arbitrary categorization and identification of objects, in the constitution of experience, in description and explanation of change, in the theory of inductive inference, in the analysis of objective similarities and in the statement of supervenience principles. In the resulting picture there are both fine-grained intensions, which play the role of cognitive and linguistic contents, and coarse-grained intensions, which play a constitutive role in the structure of the world.

Systematic theories of intensional entities have often incorporated some form of extensional reductionism. The leading reductions (possible worlds, propositional complexes, propositional functions) are critically reviewed below; non-reductionist approaches to intensional entities are then discussed. At issue is the question of what intensional entities *are*. Are they identical to extensional functions, ordered sets, sequences and so on? Or are they *sui generis* entities, belonging to an altogether new category?

2 Extensional reductions

On the possible worlds reduction (see Lewis 1986; Stalnaker 1984), a proposition is either a set of possible worlds or a function from possible worlds to truth-values, and properties are functions from possible worlds to sets of possible (usually non-actual) objects (see Semantics, possible worlds §9). Many people find the possible worlds theory intuitively implausible. Are familiar sensible properties (for example, colours, shapes, aromas) really functions? When I am aware that I am in pain, is a set (function) of possible worlds really the object of my awareness? Besides this intuitive objection, there are familiar epistemological and metaphysical objections to a theory that is truly committed to the existence of things that do not exist actually. Finally, the possible worlds reduction implies that necessarily equivalent intensions are identical, but as seen above this is implausible in the case of concepts and propositions. Certain possible worlds theorists (for example, Cresswell 1985) have responded to this problem by holding that concepts and propositions are ordered sets (sequences, abstract trees) whose elements are possible worlds constructs. This revisionary view is a variation of the propositional-complex reduction.

According to this reduction (see Perry and Barwise 1983), concepts and propositions are nothing but ordered sets (sequences, abstract trees, partial functions) whose ultimate elements are properties, relations and perhaps individuals. For example, the proposition that you are running is the ordered set $\langle \mathrm{running,\ you} \rangle$; the proposition that you are running and I am walking is

$\langle \mathrm{conjunction,\ \langle running,\ you\rangle, \langle walking,\ me\rangle} \rangle;$

and so forth. As with the possible worlds theory, this theory clashes with intuition. When I am aware that I am in pain, is an ordered set the object of my awareness? When I see that you are running, do I see an ordered set? Furthermore, there is in principle no way to determine which ordered set I allegedly see. Is it $\langle \mathrm{running,\ you} \rangle$? Or is it $\langle \mathrm{you,\ running} \rangle$? The choice seems utterly arbitrary, but if the reduction were correct, there would have to be a fact of the matter.

A rather different kind of difficulty arises in connection with ‘transmodal quantification’. Consider the following intuitively true sentence:

(1) Every *x* is such that, necessarily, for every *y*, it is either possible or impossible that $x = y$.

In symbols,

(2) $(\forall x)\Box(\forall y)(\mathrm{Possible}\ [x = y] \vee \mathrm{Impossible}\ [x = y]).$

By the propositional-complex theory, this is equivalent to

(3) $(\forall x)\Box(\forall y)(\mathrm{Possible}\ \langle x, \mathrm{identity}, y\rangle \vee \mathrm{Impossible}\ \langle x, \mathrm{identity}, y\rangle).$

The singular term ‘ $\langle x, \mathrm{identity}, y\rangle$’ may be thought of as a definite description, ‘the ordered set whose elements

are *x*, identity and *y*'. Accordingly, there are two ways to understand (3). On the first construal (corresponding to the narrow scope reading of the definite description), (3) implies

$$(\forall x)\Box(\forall y)(\exists v)v = \langle x, \text{identity}, y\rangle.$$

However, by the principle that, necessarily, a set exists only if its elements exist, this implies

$$(\forall x)\Box(\exists v)v = x.$$

That is, everything necessarily exists. A manifest falsehood. On the alternative construal of (3), corresponding to the wide scope reading of the definite description, 'the ordered set $\langle x, \text{identity}, y\rangle$', (3) implies that every *x* is such that, necessarily, for all *y*, there exists an actual set $\langle x, \text{identity}, y\rangle$. That is

$$(\forall x)\Box(\forall y)(\exists_{\text{actual}}\, v)v = \langle x, \text{identity}, y\rangle.$$

But, by the principle that, necessarily, a set is actual only if its elements are actual, this implies

$\Box(\forall y)y$ is actual.

That is, necessarily, everything (including everything that could exist) is already actual. Another manifest falsehood. So either way, (3) implies something false. But (3) is the propositional-complex theorists' way of representing the true sentence (2). So the propositional-complex theory appears unable to handle intuitively true sentences such as (2). (This style of argument can be exploited in a defence of *ante rem* realism.)

According to the propositional-function theory (see Whitehead and Russell 1910-13), a property is a function from objects to propositions, where propositions are taken to be primitive entities. For example, the property 'being red' is the function (λx)(the proposition that *x* is red). For any given object *x*, the proposition that *x* is red is the result of applying this function to the argument *x*. That is, the proposition that *x* is red is (λx)(the proposition that *x* is red)(*x*). But are the familiar sensible properties really functions? Once again, this is highly implausible and should be rejected on this ground alone if an acceptable alternative exists.

Besides this sort of intuitive difficulty, there are several technical difficulties. Consider an illustration involving properties of integers. Being even = being an *x* such that *x* is divisible by two; and being self-divisible = being an *x* such that *x* is divisible by *x*. Given the propositional-function theory, the following identities hold:

$$\begin{aligned}\text{that two is even} &= (\lambda x)(\text{that } x \text{ is even})(\text{two})\\ &= (\lambda x)(\text{that } x \text{ is divisible by two})(\text{two})\\ &= \text{that two is divisible by two}\\ &= (\lambda x)(\text{that } x \text{ is divisible by } x)(\text{two})\\ &= (\lambda x)(\text{that } x \text{ is self-divisible})(\text{two})\\ &= \text{that two is self-divisible.}\end{aligned}$$

However, that two is even and that two is self-divisible are plainly different propositions: certainly someone could be consciously and explicitly thinking the former while not consciously and explicitly thinking the latter. Indeed, someone who is thinking that two is even might never have considered the concept of self-divisibility. (Analogous difficulties arise for Church's somewhat similar propositional-function theory of concepts (1951). Incidentally, since properties are not propositional functions, using function abstracts to denote them invites confusion. A growing practice is to use $\ulcorner[v\colon\ A]\urcorner$ to denote the property of being a *v* such that *A*, just as in set theory $\ulcorner\{v\colon\ A\}\urcorner$ denotes the set of things *v* such that *A*.)

3 Non-reductionist approaches

Each of the preceding problems arises from the attempt to reduce intensional entities of one kind or another to extensional entities; either extensional functions or sets. These difficulties have led some theorists to adopt non-reductionist approaches (see Bealer 1979, 1982). Consider the following truisms. The proposition that $A\&B$ is the conjunction of the proposition that *A* and the proposition that *B*. The proposition that not *A* is the negation of the proposition that *A*. The proposition that *Fx* is the predication of the property *F* of *x*. The proposition that there

exists an *F* is the existential generalization of the property *F*, and so on. These truisms tell us what these propositions essentially are: they are by nature conjunctions, negations, singular predications, existential generalizations and so on. These are rudimentary facts which require no further explanation and for which no further explanation is possible. Until the advent of extensionalism, this was the standard view.

By adapting techniques developed in the algebraic tradition in extensional logic, one is able to develop this non-reductionistic approach. Examples such as those just given isolate fundamental logical operations: conjunction, negation, singular predication, existential generalization and so on. Intensional entities are then taken as *sui generis* entities; the aim is to analyse their behaviour with respect to the fundamental logical operations.

An intensional algebra is a structure consisting of a domain *D*, a set of logical operations and a set of possible extensionalization functions. The domain divides into subdomains: particulars, propositions, properties, binary relations, ternary relations and so on. The set of logical operations includes those listed above plus certain auxiliary operations. The possible extensionalization functions assign a possible extension to relevant items in the domain: each proposition is assigned a truth-value; each property is assigned a set of items in *D*; each binary relation is assigned a set of ordered pairs of items in *D*, and so on. One extensionalization function is singled out as the actual extensionalization function: the propositions which are true relative to it are the propositions which are actually true, and so on. (More formally, an intensional algebra is a structure $\langle D, \tau, K, G \rangle$. *D* divides into subdomains: $D_{-1}, D_0, D_1, D_2, \ldots$. D_{-1} consists of particulars; D_0, propositions; D_1, properties; D_2, binary relations, and so on. τ is a set of logical operations on *D*. *K* is a set of extensionalization functions. *G* is a distinguished function in *K* which is the actual extensionalization function.)

To illustrate how this approach works, consider the operation of conjunction, conj. Let *H* be an extensionalization function. Then, conj satisfies the following: for all propositions *p* and *q* in *D*, *H*(conj(*p*, *q*)) = true iff *H*(*p*) = true and *H*(*q*) = true. Similarly, if neg is the operation of negation, then for all propositions *p* in *D*, *H*(neg(*p*)) = true iff *H*(*p*) = false. Likewise, for singular predication preds, which takes properties *F* in *D* and items *y* in *D* to propositions in *D*: *H*(pred_s(*F*, *y*)) = true iff *y* is in the extension *H*(*F*).

This non-reductionistic approach can be extended to more complex settings in which, for example, both fine-grained intensional entities (concepts, propositions) and coarse-grained intensional entities (properties, relations, conditions) are treated concurrently and in which a relation of correspondence between the two types of intensional entities can be characterized in terms of the fundamental logical operations. The non-reductionist approach can thus accommodate the fine-grained intensional entities which serve as cognitive and linguistic contents and also the more coarsely grained intensional entities which play a constitutive role in the structure of the world.

See also: Nominalism; Intensional logics; Logical and mathematical terms, glossary of; Modal logic, philosophical issues in; Second-order logic, philosophical issues in; Property theory

GEORGE BEALER

References and further reading

Aczel, P. (1980) ‘Frege Structures and the Notions of Propositions, Truth and Set’, in J. Keisler, J. Barwise and K. Kunen (eds) *The Kleene Symposium*, Amsterdam: North Holland, 31-59.(Development of the propositional-function view.)

Aczel, P. (1989) ‘Algebraic Semantics for Intensional Logics, I’, in G. Chierchia, B. Partee and R. Turner (eds) *Property Theories, Type Theories, and Semantics*, Dordrecht: Kluwer, 17-46.(Further development of the propositional-function view.)

Armstrong, D.M. (1978) *A Theory of Universals*, Cambridge: Cambridge University Press.(Study of coarse-grained intensional entities, discussed at close of §1.)

Arnauld, A. (1662) *Logic, or the Art of Thinking: The Port-Royal Logic*, trans. T.S. Baynes, Edinburgh: Sutherland & Knox, 1850.(Presentation of the traditional doctrine of intension and extension, discussed in §1.)

Bealer, G. (1979) ‘Theories of Properties, Relations, and Propositions’, *Journal of Philosophy* 76: 643-8. (Development of the non-reductionist view discussed in §3.)

Bealer, G. (1982) *Quality and Concept*, Oxford: Clarendon Press.(Expansion of the material in §1 and development of the non-reductionist view discussed in §3.)

Bealer, G. (1993a) ‘Universals’, *Journal of Philosophy* 90: 5-32.(Defence of the *ante rem* theory of intensional

entities.)
Bealer, G. (1993b) ‘A Solution to Frege’s Puzzle’, in J. Tomberlin (ed.) *Philosophical Perspectives*, Atascadero, CA: Ridgeview Press, vol. 7, 17-61.(Application of the non-reductionist view to some outstanding problems in the logic for intensional entities.)
Bealer, G. and Mönnich, U. (1989) ‘Property Theories’, in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 4, 133-257.
Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press.(Early presentation of the contemporary use of ‘intension’ and ‘extension’ and precursor to the possible worlds theory.)
Church, A. (1951) ‘A Formulation of the Logic of Sense and Denotation’, in P. Henle, H.H. Kallen, S.K. Langer (eds) *Structure, Method, and Meaning: Essays in Honor of Henry M. Scheffer*, New York: Liberal Arts Press, 3-24.(Development of a Fregean version of the propositional-function view.)
Cresswell, M.J. (1985) *Structured Meanings*, Cambridge, MA: MIT Press.(A combination of the possible worlds theory and the propositional-complex theory.)
Frege, G. (1892) ‘Über Sinn und Bedeutung’, *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, ‘On Sense and Meaning’, in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980.(Defence of the theory that meaningful expressions typically have both a sense and a reference.)
Frisch, J.C. (1969) *Extension and Comprehension in Logic*, New York: Philosophical Library.
Lewis, D.K. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.(Defence of coarse-grained intensions and the possible worlds theory.)
Lewis, D.K. (1983) ‘New Work for a Theory of Universals’, *Australasian Journal of Philosophy* 61: 343-77.
Menzel, C. (1986) ‘A Complete Type-Free "Second-Order" Logic and its Philosophical Foundations’, in *Report No. CSLI-86-40*, Stanford, CA: Center for the Study of Language and Information, Stanford University. (Development of the non-reductionist view discussed in §3.)
Mönnich, U. (1983) ‘Toward a Calculus of Concepts as a Semantical Metalanguage’, in R. Bäuerle, C. Schwarze and A. von Stechow (eds) *Meaning, Use, and Interpretation of Language*, Berlin: de Gruyter, 342-60.
Parsons, T. (1980) *Nonexistent Objects*, New Haven, CT: Yale University Press.(Development of the non-reductionist view in combination with a Meinongian theory of nonexistent objects.)
Perry, J. and Barwise, J. (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press.(Development of the propositional-complex view.)
Richard, M. (1990) *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*, New York: Cambridge University Press.
Salmon, N. (1986) *Frege’s Puzzle*, Cambridge, MA: MIT Press.(Development of the propositional-complex view.)
Soames, S. (1987) ‘Direct Reference, Propositional Attitudes, and Semantic Content’, *Philosophical Topics* 15: 47-87.
Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press.(Defence of the possible worlds view.)
Turner, R. (1987) ‘A Theory of Properties’, *Journal of Symbolic Logic* 52: 455-72.
Turner, R. (1991) *Truth and Modality for Knowledge Representation*, Cambridge, MA: MIT Press.(Development of the propositional-function view.)
Turner, R. and Chierchia, G. (1988) ‘Semantics and Property Theory’, *Linguistics and Philosophy* 11: 261-302.
Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(A classic of mathematical logic, which includes a development of intensional logic based on the propositional-function theory.)
Zalta, E. (1983) *Abstract Objects: An Introduction to Axiomatic Metaphysics*, Dordrecht: Reidel.
Zalta, E. (1988) *Intensional Logic and the Metaphysics of Intentionality*, Cambridge, MA: MIT Press. (Development of the non-reductionist view in combination with a Meinongian theory of nonexistent objects.)

Intensional logics

Intensional logics are systems that distinguish an expression's intension (roughly, its sense or meaning) from its extension (reference, denotation). The purpose of bringing intensions into logic is to explain the logical behaviour of so-called intensional expressions. Intensional expressions create contexts which violate a cluster of standard principles of logic, the most notable of which is the law of substitution of identities - the law that from a $=$ *b and P(a) it follows that P(b). For example, 'obviously' is intensional because the following instance of the law of substitution is invalid (at least on one reading): Scott* $=$ *the author of* Waverley*; obviously Scott* $=$ *Scott; so, obviously Scott* $=$ *the author of* Waverley*. By providing an analysis of meaning, intensional logics attempt to explain the logical behaviour of expressions such as 'obviously'. On the assumption that it is intensions and not extensions which matter in intensional contexts, the failure of substitution and related anomalies can be understood.*

Alonzo Church pioneered intensional logic, basing it on his theory of types. However, the widespread application of intensional logic to linguistics and philosophy began with the work of Richard Montague, who crafted a number of systems designed to capture the expressive power of natural languages. One important feature of Montague's work was the application of possible worlds semantics to the analysis of intensional logic. The most difficult problems concerning intensional logic concern the treatment of propositional attitude verbs, such as 'believes', 'desires' and 'knows'. Such expressions pose difficulties for the possible worlds treatment, and have thus spawned alternative approaches.

1 Intension and extension

Gottlob Frege believed that expressions of language have both an extension (reference, denotation) and an intension (sense, meaning; see Sense and reference). He distinguished intensions from extensions to explain a failure of the law of substitution of identity. From 'The author of *Waverley* $=$ Scott' and 'John admires the author of *Waverley*' it does not follow that 'John admires Scott', for John may think that Scott wrote *Ulysses*, which he hates. (Frege illustrated the point with 'the morning star' and 'the evening star'.) One way out of the puzzle is to assume that the second sentence reports a relation (of admiration) between John and the meaning (or intension) of '*Waverley*', rather than a relation between John and the extension of 'the author of *Waverley*', that is, Scott. The fact that the meanings of 'Scott' and 'the author of *Waverley*' differ can help explain why substitution fails.

The failure of substitution in the preceding example is caused by the word 'admires'. Such intensional expressions violate several other standard principles of logic. A notable case is existential generalization: $P(a)$; therefore there exists an *x* such that $P(x)$. Existential generalization fails because, for example, 'John admires Superman' does not entail that Superman actually exists. The stock of intensional expressions of English is very large. It includes phrases such as 'it is necessary that', 'it used to be the case that', 'believes that', 'it ought to be the case that', and many more.

Under a broad construal, intensional logic includes any logic that accounts for the deductive behaviour of one or more of the intensional expressions. So modal logic, tense logic and logics of belief and obligation all count as intensional logics from this point of view. However, the term 'intensional logic' more commonly refers to those systems in which some explicit symbolic device is introduced for distinguishing intensions from extensions. These are the systems described here.

2 Possible worlds semantics

Despite its failings, possible worlds semantics has come to play a central role in intensional logic (see Possible worlds). The basic idea goes back to Rudolf Carnap. Why do the meanings (or intensions) of 'Scott' and 'the author of *Waverley*' differ? Because there are possible worlds where the author of *Waverley* was not Scott. We can imagine a world, for example, where it was James Joyce or Charles Dickens who wrote *Waverley*. If we consider the extensions that 'the author of *Waverley*' takes across different possible worlds, we will see variations from one world to the next. It makes sense, then, to identify the *intension* of the expression with the pattern of its *extensions* across different possible worlds. Seen in this way, its intension is a function that takes each possible world into a corresponding extension. The value of the function is Scott in the real world, Joyce in another world, and Dickens

in another. The intension for ‘Scott’, on the other hand, is the constant function that has the value Scott in each possible world.

Intensions for sentences, predicates and expressions in other grammatical categories can be treated in a similar way. The extension of a sentence is its truth-value. The intension is identified with the pattern of truth-values the sentence takes across the different possible worlds. It is therefore a function that assigns to each possible world a corresponding truth-value. The same tactic can be used for predicates. The extension of the predicate ‘is blue’ is traditionally taken to be the set of objects that are blue. Its intension is a function that takes each of the possible worlds into a corresponding set of objects (the set of blue objects in that world).

3 Montague’s intensional logic

Alonzo Church (1951) was the pioneer in intensional logic, basing it on his theory of types, but Richard Montague’s work has been crucial to its widespread application in philosophy and linguistics (see Montague 1974). The system IL described here is a variant of Montague’s approach, due to Gallin (1975). Intensions and extensions for sentences, names, predicates, adjectives, prepositions and so on are different sorts of things. Therefore we need to distinguish a rich variety of ‘types’ (grammatical categories or parts of speech). Most intensional logics are constructed from some form of type theory, a higher-order logic that allows a completely general formation of types (see Theory of types; Second- and higher-order logics).

In IL there are two basic types: *t* for truth-values and *e* for entities (or objects). Extensions of sentences belong to type *t* and extensions of proper names belong to type *e*. An unlimited variety of complex types can be defined from the two basic types as follows. If α and β are types then (α, β) is another type, and if α is a type then so is (s, α), where the letter ‘*s*’ is used to indicate the set of possible worlds (or situations). The type (α, β) is for the set of all possible functions which assign to each thing with type α a value with type β. For example, if α is the type for the extension of some expression, then (s, α) is the type for its intension.

Extensions and intensions for the different grammatical categories of English can be assigned one of these types in the following way. Extensions of sentences and proper names are of types *t* and *e*, respectively, and the types for their intensions are (s, t) and (s, e), respectively. The type for predicate extensions is (e, t): the type for functions that take entities into truth-values. The idea is that the extension for ‘is blue’ is a function that assigns the truth-value T (for true) to those entities that are blue and F (for false) to those that are not. This has the same effect as treating the extension of ‘is blue’ as a set of entities, but it simplifies the definition of the complex types. Extensions for transitive verbs should take a pair of entities (one for the subject and one for the object of the verb) into a truth-value, so transitive verb extensions are assigned to type $(e, (e, t))$. The types for intensions of intransitive and transitive verbs would be $(s, (e, t))$ and $(s, (e, (e, t)))$, respectively.

IL has both constants and variables in all the possible types. It also contains the logical symbols ‘ $\equiv$’, ‘ λ’, ‘ $^\wedge$ ’, ‘ $^\vee$ ’, ‘[’ and ‘]’. These symbols allow the construction of complex terms (or expressions). The symbols ‘ $^\wedge$ ’ and ‘ $^\vee$ ’ (called ‘cap’ and ‘cup’) indicate the intension and extension of a term, respectively. So when *A* has type α, $^\wedge A$ is the intension of *A* and must have type (s, α). When *A* has type (s, α), then $^\vee A$ is *A*’s extension and belongs to type α. The symbol ‘ $\equiv$’ indicates identity. When *A* and *B* are two terms of the same type, the expression ‘ $A \equiv B$’ belongs to the *t* (truth-value) type, with the understanding that $A \equiv B$ is true if and only if *A* and *B* name the same thing. $[A \;\; B]$ indicates the result of applying the function that *A* names to whatever *B* names. So when *A* belongs to type (α, β) and *B* to type α, $[A \;\; B]$ belongs to type β. For example, if *A* is ‘runs’ with type (e, t) and *B* is ‘John’ with type *e*, then $[\mathrm{runs} \;\; \mathrm{John}]$ has type *t*, and symbolizes the extension of the English sentence ‘John runs’. When *x* is a variable with type α, and *A* is an expression with type β, then $\lambda x A$ is an expression with type (α, β), indicating functional abstraction. For example, when $A =$ ‘Mary loves *x*’, $\lambda x A$ names the function that assigns to each entity *o* either T or F depending on whether Mary loves *o* or not. Hence, in this example, $\lambda x A$ corresponds to the predicate ‘is loved by Mary’.

IL has strong expressive powers. All of the symbols of predicate logic can be defined in it - including propositional connectives (negation, conjunction and so on) and quantifiers in all the types. Furthermore, the presence of ‘ $^\wedge$ ’ ensures that an S5 strength modal operator ‘ $\Box$’ (read ‘it is necessary that’) can be defined (for a description of S5, see Modal logic). IL is on a firm technical footing. Gallin has devised rules and axioms for IL and has shown how to provide a (generalized) semantics for this system for which IL is both consistent and complete.

Intensional logics such as IL have been applied to tenses, necessity, obligation and many other intensional constructions. The same ideas can be used to analyse indexicals such as ‘I’, ‘it’ and ‘that’, with the understanding that possible worlds include contexts of utterance that help fix the extensions of these terms. A hint at how IL is applied to language can be given by showing how to explain the failure of substitution in the case of John’s admiring the author of *Waverley*. Assume that each of the words of English is assigned an appropriate type of IL. The verb ‘admires’ is intensional, meaning that the object of the verb is an *intension* of type (s, e). So the extension of ‘admires’ falls in type $(e, (\underline{(s, e)}, t))$, not $(e, (e, t))$. It follows that ‘John admires Scott’ is represented in IL by $[[\mathrm{admires} \quad \mathrm{John}] \quad {}^\wedge S]$, where *S* abbreviates ‘Scott’. The cap ‘ $^\wedge$ ’ makes it clear that it is the intension of ‘Scott’ that is at issue. Similarly, $[[\mathrm{admires} \quad \mathrm{John}] \quad {}^\wedge W]$ represents ‘John admires the author of *Waverley*’. The claim that Scott is the author of *Waverley* is written ‘ $S \equiv W$’. Since $S \equiv W$ only says that *S* and *W* have the same *extension*, the inference from $S \equiv W$ and $[[\mathrm{admires} \quad \mathrm{John}] \quad {}^\wedge W]$ to $[[\mathrm{admires} \quad \mathrm{John}] \quad {}^\wedge S]$ is invalid. An identity involving *S* and *W* does not warrant substitution of ${}^\wedge S$ for ${}^\wedge W$.

4 The problem of propositional attitudes

While intensional logics in the style of IL appear to work nicely in the case of tense, necessity, obligation, indexicals and the like, serious problems arise in certain cases when IL is applied to so-called propositional attitude verbs, such as ‘believes’, ‘wants’, ‘knows’ and ‘admires’ (see Propositional attitude statements). The fundamental difficulty is that systems based on IL validate substitution of *necessary* identities: if expressions *A* and *B* have the same extensions in every possible world, then an occurrence of *A* in one expression may be replaced with an occurrence of *B* in another. But such substitution is not valid for propositional attitude verbs. For example, John can know that 17×3 is composite (not prime) but fail to know that 51 is composite, even though in every possible world $17 \times 3 = 51$.

The difficulty is more than a quirk about mathematical expressions. It is widely held that proper names and natural kind terms are rigid designators, that is, that they have the same extension (or none) in every possible world. So substitution of pairs of names (‘Cicero’, ‘Tully’) or natural kind terms (‘water’, ‘ $\mathrm{H_2O}$’) that have the same extensions should be correct in all contexts. However, John can admire Cicero without admiring Tully (even though Cicero *was* Tully), and he can know that water is safe to drink, without knowing that $\mathrm{H_2O}$ is safe to drink (even though water *is* $\mathrm{H_2O}$). The distinctions between intensions defined by possible worlds semantics appear to be too crude to explain the failure of substitution in propositional attitude contexts.

5 Alternative approaches

Alternatives to the possible worlds approach in intensional logic have grown out of the desire to resolve problems including that of propositional attitudes. One natural reaction is to dispense with talk of intensions altogether, and to explain John’s different reactions to the ideas that 17×3 and 51 are composite as due to his different relationships with the symbols ‘ 17×3’ and ‘51’. However, ‘John knows 17×3 is composite’ cannot report a relation between John and the symbol ‘ 17×3’, because this sentence can be true even when John has never seen the symbols ‘ 17×3’.

Max Cresswell’s solution to the problem, following Carnap, is to bring in what he calls structured meanings (1985). The meaning of atomic expressions can be given by a possible worlds account. For example, the meaning of ‘51’ can be the function that picks out 51 in every possible world. For complex expressions, such as ‘ 17×3’, the meaning is a structure, $\langle \mathrm{times}, 17, 3 \rangle$, composed of the meanings of the sub-expressions ‘times’, ‘17’ and ‘3’. Note that $\langle \mathrm{times}, 17, 3 \rangle$ is not itself a symbol, but does preserve the structure of ‘ 17×3’, thereby allowing the meanings of ‘ 17×3’ and ‘51’ to differ.

Since neither ‘Cicero’ nor ‘Tully’ is complex, Cresswell must use different tactics to manage this case. One strategy he discusses is to analyse ‘John admires Cicero’ as ‘John admires the person he calls Cicero’. (However, since Cresswell believes propositional attitude statements are essentially ambiguous, this is not his final word on the topic.)

Other approaches abandon entirely the possible worlds account of intensions. Anderson, for example, presents an intensional logic whose semantics is based on the notion of synonymy (1984). The idea is that the rules of a language allow us to define the class of synonyms for an atomic expression (that is, the set of expressions with the

same intension), called its ‘concept’. Concepts for complex expressions can be built up from the concepts for simple expressions. Expressions of the language are indexed so that A_1 indicates the concept for expression *A*. The problem of the failure of substitution of necessary identities in propositional attitude contexts is solved as follows. We assume that the verb ‘admires’ is intensional, so it must take expressions for concepts, not expressions for extensions. So ‘Cicero is Tully’ is represented by $c = t$, while ‘John admires Cicero’ is represented by Ajc_1, where the index 1 indicates that it is the *concept* of Cicero (c_1) that is at issue. But now Ajt_1 (for ‘John admires Tully’) does not follow, since $c = t$ does not warrant substitution of t_1 for c_1. On the plausible assumption that ‘Cicero’ and ‘Tully’ are not synonymous (that is, do not have the same intension), the substitution can be blocked as desired. Note, however, that ordinary intuitions rule that ‘water’ is synonymous with ‘ H_2O’, and that ‘ 17×3’ is synonymous with ‘51’. Hence the ordinary notion of synonymy cannot explain the failure of substitution in these cases. For a general solution to the problem, Anderson must employ a notion of synonymy that is at odds with ordinary intuitions.

Zalta (1988) lays synonymy aside, and introduces abstract objects to serve as the meanings of intensional expressions. Abstract objects may be thought of as visual images or other mental encodings used to represent features of the world. An extensional term *c* and its corresponding intensional term c_p belong to the same type, and the subscript *p* indicates which person does the encoding. Using this machinery, ‘John admires Cicero’ is represented by $A(j, c_j)$, where c_j indicates John’s encoding of Cicero. Note that the encodings named by c_j (‘Cicero’) and t_j (‘Tully’) may be distinct. Since the notation for ‘Cicero $=$ Tully’ ($c = t$) lacks the index *j*, this claim does not warrant substitution of t_j for c_j in $A(j, c_j)$.

See also: Intensional entities; Logical and mathematical terms, glossary of

JAMES W. GARSON

References and further reading

Anderson, C.A. (1984) ‘General Intensional Logic’, in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 2, 355-85.(A clear and useful introduction to intensional logics, including an account of IL. See §§3, 5.)

Barwise, J. and Perry, J. (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press.(An enjoyable introduction to situation semantics, an influential outgrowth of possible worlds semantics for intensional logic.)

Church, A. (1951) ‘A Formulation of the Logic of Sense and Denotation’, in *Structure, Method, and Meaning*, New York: Liberal Arts Press.(The seminal work in intensional logic, mentioned in §3.)

Cresswell, M.J. (1985) *Structured Meanings*, Cambridge, MA: MIT Press.(Referred to in §5. An accessible account of an intensional logic based on structured meanings.)

Gallin, D. (1975) *Intensional and Higher-Order Modal Logic*, Amsterdam: North Holland.(The classic formulation of Montague’s approach to intensional logic. Somewhat technical.)

Montague, R. (1974) *Formal Philosophy: Selected Papers of Richard Montague*, ed. R.H. Thomason, New Haven, CT: Yale University Press.(A collection of Richard Montague’s papers prefaced by a very helpful introduction describing the application of intensional logic to natural language.)

Zalta, E.N. (1988) *Intensional Logic and the Metaphysics of Intensionality*, Cambridge, MA: MIT Press.(An accessible account of Zalta’s approach to intensional logic described in §5.)

Intensionality

The truth or falsity of many sentences depends only on which things are being talked about. Within intensional contexts, however, truth values also depend on how *those things are talked about, not just on which things they are. Philosophers and logicians have offered different analyses of intensional contexts and the behaviour of terms occurring within them.*

The extension of a term is the thing or things it picks out: for instance, the extension of 'the Big Dipper' is the stellar constellation itself. The intension of a term can be thought of as the way in which it picks out its extension. 'The Big Dipper' and 'the Plough' have the same extension - a particular constellation - but pick it out in different ways. The two terms have different intensions. There is much debate about how we should actually understand the notion of an intension (see Intensional entities).

For many classes of sentence, the substitution of one term for another with the same extension leaves the truth value unchanged, irrespective of whether the two terms also have the same intension. This is called intersubstitution *salva veritate* (preserving truth). For instance, if it is true that the Big Dipper consists of seven stars, then it is also true that the Plough consists of seven stars. In a number of important cases, however, terms with the same extension but different intensions are not intersubstituable *salva veritate*. Consider, for instance:

1 Tom believes that the Big Dipper consists of seven stars.
2 Tom believes that the Plough consists of seven stars.

It is quite possible that Tom mistakenly believes that 'the Plough' refers to a totally different constellation consisting of only six stars, in which case (1) could be true and (2) false. Co-extensive terms cannot be intersubstituted *salva veritate* within the scope of the verb 'believes'. Such contexts are called intensional.

Other verbs which, like 'believe', refer to propositional attitudes - verbs like 'hope', 'desire', 'fear' - also create intensional contexts (see Propositional attitude statements). Propositional attitudes are *intentional* states (note the 't') which has led some to consider the possible connections between inten*S*ionality and inten*T*ionality (see Intentionality §2). The two must be distinguished carefully, however, not least because intensional contexts are also created by non-psychological terms, most importantly modal terms like 'necessarily' and 'possibly'. The intensionality of modal contexts leads to particular difficulties when they also involve quantification (see Modal logic, philosophical issues in §3); these problems have led to controversies over the interpretation not only of modal terms but also of the quantifiers (see Modal operators; Quantifiers, substitutional and objectual §1).

The failure of intersubstitutability *salva veritate* raises questions about the behaviour of terms in intensional contexts. Frege (1892) argues that, in intensional contexts, the term 'the Big Dipper' in (1) refers not to the Big Dipper as usual but to its own intension (or, in Frege's terminology, its sense; see Indirect discourse §1; Proper names §5; Sense and reference §5). Quine (1956, 1961) describes intensional contexts as *referentially opaque*, arguing that terms occurring within them do not refer at all (see Propositional attitude statements §2; Modal operators §1). Davidson's (1969) *paratactic* analysis of propositional attitude statements and of indirect discourse, another intensional context, tries to preserve our intuition that terms in intensional contexts work in the same way as in any other contexts (see Indirect discourse §2; Propositional attitude statements §3).

Many philosophers think that logic should have nothing to do with intensions. But the intensionality of many natural language contexts is hard to deny, and the development of intensional logics suggests that a more tolerant attitude might well pay off (see Intensional logics).

See also: Concepts; Reference; Semantics, possible worlds

SIMON CHRISTMAS

References and further reading

Davidson, D. (1969) 'On Saying That', *Synthèse* 19: 130-46; repr. in D. Davidson and J. Hintikka (eds) *Words and Objections: Essays on the Work of W.V. Quine*, Dordrecht: Reidel, 158-74.(Davidson's paratactic analysis of indirect discourse.)

Frege, G. (1892) ‘Über Sinn und Bedeutung’, *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black as ‘On Sense and Reference’, *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 1962, 56-78.(Frege’s seminal discussion of failures of intersubstitution and distinction between sense and reference.)

Quine, W.V. (1956) ‘Quantifiers and the Propositional Attitudes’, *The Ways of Paradox*, Cambridge MA: Harvard University Press, 1976.(Quine on referential opacity.)

Quine, W.V. (1961) ‘Reference and Modality’, *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 2nd edn, 1980; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971.(Quine on the referential opacity of modal contexts. The Linsky collection contains some fairly difficult discussions.)

Searle, J.R. (1983) *Intentionality*, Cambridge: Cambridge University Press.(Chapter 1 contains a useful discussion of the relationship between inten*S*ionality and inten*T*ionality.)

Intention

Suppose that Kevin intends to brush up on his predicate logic, and acts on this intention, because he wants to conduct a good tutorial and he believes that some preparatory revision will help him to do so. In an example like this, we explain why Kevin intends to revise his logic, and why he (intentionally) does revise it, by appealing to the belief and desire which provide his reasons both for his intention and his corresponding intentional action. But how does Kevin's intention to act, coming between his reasons and his action, help to explain what he does? Central questions in the theory of intention include the following: Are intentions distinct mental attitudes or are they analysable in terms of other mental attitudes - such as beliefs and desires? How is intending to do something related to judging that it is best to do it? What distinctive roles, if any, do intentions play in getting us to act? Are foreseen but undesired consequences of an intentional action intended?

1 Intention and belief

Is intending to act a belief that one (probably) will? The identification of intending to act in a certain way with a belief that one (probably) will so act is misguided since even the weaker thesis that intending to act in a certain way implies the belief that one (probably) will is too strong. Thus Kevin may intend to go to the gym (after revising his logic) while harbouring serious doubts about whether he will, which leave him agnostic about the future - for he may know only too well his tendency to balk at doing painful things when the time for action comes, or even to forget his good intentions in such cases. A principal motive for holding the thesis that intending to act implies believing that one (probably) will has been the concern to distinguish intentions from mere hopes or wishes, but such counterexamples show that the distinction will have to be drawn in some other way. Some philosophers hold that there is a minimal positive belief constraint on intention: intention implies the belief that one has some chance of achieving what one intends (see Davidson 1985). Others think that at most one can insist on a negative constraint: intention implies that one does not believe that one (probably) will not do as one intends. Still other philosophers prefer to interpret such a negative belief constraint as a rationality condition on intention. It is not possible to adjudicate these issues here.

Even where an intention to act is matched by a belief that one (probably) will so act, the two are clearly not identical. Suppose an agent who intends to *A* actually does believe that they (probably) will *A*. The agent's belief (insofar as it is rational) will depend upon their estimate of their chances of *A*-ing given that they intend to *A*. So, in such examples, the agent's intention, as a cause of the belief about its presence, will belong, as a distinct item, in the causal background of the (tascit or explicit) reasoning which issues in their belief about what they (probably) will do. Some philosophers have thought that when an intention to act in a certain way corresponds to a belief about what one will do, the belief is not properly construed as based on evidence. Indeed, some even hold the belief in such cases to be noninferential. If the above account is correct, such a belief has perfectly familiar sorts of justifying grounds, namely evidential considerations that bear on one's chances of acting in a certain way, including (saliently) one's belief that one intends to act in that way.

2 Intention and desire

Is intending to do something the same as desiring most to do it? Every intention is or incorporates a desire or want in the broadest sense of these terms, since every intention is or incorporates a motivational state. This concession, however, raises a critical issue in the theory of intention. For some philosophers contend that intentions motivate actions because they are or essentially include predominant desires. This cannot be right since there can be intention without predominant desire, and the converse. Plausible examples here typically exploit a functional difference between intention and predominant desire: the former, unlike the latter, characteristically *settles* the question of what to do for an agent. Thus Kevin's intention may track what he thinks it is overall best to do despite his present strongest desire - as when he forms an intention to go to the gym, which he thinks it best to do, despite wanting most to go to the pub. And conversely, Kevin may continue to treat a question of what to do as open, and so not yet have formed an intention about what to do, despite his present strongest desire - as when he continues to wonder about whether to go to the gym, his predominant desire to go to the pub notwithstanding.

Once the conceptual wedge between intention and predominant desire is granted, however, there remains the issue of how to characterize the role of intentions as motivators of intentional actions, given that that role is not

reducible to desire strength. A natural move, at this point, is to develop the notion that intentions settle for agents, or commit them on, questions of what to do. Such accounts suggest an intuitive contrast between intentions as executive desires and other mere desires. Intentions involve first-order dispositions to act (intentionally), whereas mere desires involve second-order dispositions to act (intentionally). That is, mere desires are or include dispositions to acquire such first-order dispositions to act as are involved in intentions.

Theories of intention which emphasize the notion that intentions settle for agents, or commit them on, questions of what to do have a further virtue. They are especially well-suited to do justice to the several important functional roles which mark out intentions as a group. These functions are that intentions initiate, sustain and guide intentional actions, are elements in coordinative plans, and prompt and terminate practical reasoning.

3 Intention and evaluation

Is intending to act an evaluative attitude? Donald Davidson (1978) construes desires and intentions as evaluative attitudes. Desires are conditional evaluations, the contents of which are expressible in judgments of the form: ‘It is desirable for me to *A* insofar as my *A*-ing is *F*.’ Thus the content of Kevin’s desire to go to the gym would be expressible as (say): ‘It is desirable for me to go to the gym insofar as my going to the gym will help me to keep fit.’ Intentions are unconditional or all-out evaluations, the contents of which are expressible in judgments of the form: ‘It is desirable (intention-worthy) for me to *A*.’ Thus the content of Kevin’s intention to go to the gym would be expressible as: ‘It is desirable (intention-worthy) for me to go to the gym.’ One of Davidson’s chief concerns is how to set out schematically the form of practical reasoning, which contains a mix of cognitive and conative propositional attitudes. By distinguishing among the propositional expressions of attitudes, Davidson hopes to mark differences among the attitudes. His economical proposal is to let ordinary indicative sentences express the contents of ’factual’ beliefs and to let indicatives with explicitly evaluative words express the contents of conative attitudes like desires and intentions. Intentions are then distinguished from mere desires by the all-out or unconditional form of the evaluative sentences which express them. On Davidson’s account, the inferential step wherein an agent comes to have an intention (forms an all-out evaluation by detachment from a corresponding conditional one) is treated as analogous to the inferential step wherein a believer comes to have an inductive belief that a proposition is simply true or false (by detachment from a corresponding subjective probability assessment).

Davidson’s evaluative theory of intention has proved contentious. One issue is whether he is right about the logical form of evaluative sentences. For example, Davidson construes the conditional judgment that it is desirable for me to *A* insofar as my *A*-ing is *F* as having the logical form: ‘It is prima facie desirable for me to *A* insofar as my *A*-ing is *F*’, which he symbolizes as: pf (It is desirable for me to *A*, my *A*-ing is *F*) to bring out the point that these judgments are thought of as the result of applying a ‘prima facie’ operator to a pair of sentences related as evaluation and ground. Such judgments supposedly parallel probabilistic ones of the form: ‘it is probable that it will rain insofar as rain is forecast.’

A second issue is whether, even if he is correct about logical form, Davidson is right to assimilate detached, unconditional evaluations to intentions to act. There are, in fact, good general reasons to resist the assimilation of conative attitudes to evaluative ones. Such reasons have to do with the explanatory value in philosophical psychology of being able to admit fractures between evaluation and corresponding desire or intention (see Akrasia). Yet there is a need for caution here with respect to one of Davidson’s chief preoccupations - namely, the issue of how to represent the form of practical reasoning. For, even if desires and intentions are different from evaluative attitudes, it remains a viable option that sentences expressing the latter are appropriate for marking the former in schematic representations of practical reasoning. The symmetrical line of thought here would be that just as sentences about what is true (belief-worthy) are appropriate for marking beliefs in theoretical cases, sentences about what is desirable in some way or intention-worthy are appropriate for marking conative attitudes in practical cases. Of course, a difference would also have to be conceded, once the difference between conative and evaluative attitudes was conceded. Sentences about what is true (or their equivalents) would count as expressing the contents of beliefs; whereas sentences about what is desirable or intention-worthy would count as expressing the contents of justifying evaluative judgments which corresponded to (without being identical with) conative attitudes.

4 Intentions as self-referential

Are intentions self-referential? When an agent intends to act in a certain way, does the content of the intention make essential reference to the intention itself?

On some proposals the self-referentiality thesis can be characterized (roughly) as follows: an intention to *A* is the intention that, because of this very intention, one will *A*. So, if Kevin intends to brush up on his predicate logic, then (more fully) he intends that, as a result of this very intention, he will so act. Such proposals have an obvious curious feature: they tend to conflate standard cases of intending with special cases of self-control where an agent forms an intention with the higher-order intention of thereby bringing it about that they will act in a certain way.

One standard defence of the self-referentiality thesis is that it explains why some intended behaviour falls short of intentional action (see the discussion of deviant causal chains in Action; Reasons And Causes §3). Consider just one kind of example. Suppose that Kevin intends to shoot an intruder. He aims his gun but, at the crucial moment, is startled by a moving shadow, which causes him to contract his finger, so that he shoots the intruder - but not intentionally. Advocates of self-referentiality have made claims like the following: Kevin does not shoot the intruder intentionally in such a case because he does not do what he intends - since his intention to shoot the intruder is the intention that that very intention will lead him to do the shooting at the crucial moment, which does not happen. Such a defence of the self-referentiality thesis is unconvincing since there is a clear, alternative reason why Kevin does not count as an intentional agent in the example, which has nothing to do with the supposed self-referentiality of intentions. Quite simply, the relevant bodily movement - Kevin's contracting his trigger finger (or perhaps better, the contracting of Kevin's trigger finger) - is not even an action; rather, it is an involuntary movement caused by Kevin's being startled. Or, if Kevin does contract his finger as an action, the action is not even caused by his intention to shoot - he contracts his finger as a result of being startled.

John Searle's version (Searle 1983) of the thesis that intentions are self-referential deserves special comment. Searle distinguishes between prior intentions (intentions that precede actions) and intentions in action (intentions that are contemporary with and part constituents of actions). On Searle's view, both prior intentions and intentions in action are self-referential. A prior intention (say) to raise one's arm has the content: that I perform the action of raising my arm by way of carrying out this intention. An intention in action, when one intentionally raises one's arm, has the content: that my arm go up as a result of this intention in action.

Searle's view that intentions are causally self-referential relies on the idea that the specification of the content of an intention should coincide with the specification of its conditions of satisfaction. But why should this be conceded? Searle seems merely to trade on the apparent coincidence of conditions of satisfaction and content in the case of certain other propositional attitudes (like beliefs and desires) and some speech acts (like statements).

Some philosophers have argued that, in the case of intentions, conditions of satisfaction go beyond content in that there is a distinct requirement on specifications of the former that they include a reference to intentions as causes. Alfred Mele (1992) advances a plausible alternative to the self-referentiality theory of content: the content of an agent's intention to *A* is identical with the intention's plan, and the plan simply consists in the agent's representation of their prospective *A*-ing and of the route, if any, that they intend to take. So the content (plan component) of Kevin's intention to revise his predicate logic simply is that he revise his predicate logic and that he do so by (say) working his way through the relevant chapters of his logic textbook.

5 Intention and intentional action

Is it the case that an agent intentionally *A*s only if the agent intends to *A*? Michael Bratman (1987) has dubbed the thesis that all intentional actions are intended 'the Simple View'. The thesis enjoys initial plausibility, though it does not survive scrutiny.

Sudden or impulsive intentional actions are held to provide counterexamples to the Simple View. The problem in such cases, supposedly, is that there is action without enough time for forethought. This difficulty, however, does not seem very serious. For just how long need it take to form (come to have) an intention? More serious is the problem raised by subsidiary actions, where routine action parts of encompassing intentional and intended actions - like the various steps one takes when walking to the theatre - are apparently intentional without being intended. Such examples do seem to defeat the Simple View.

Another objection concerns examples where rational agents apparently perform intentional actions which it would

be irrational for them to intend to do straightout, given what they believe about the circumstances in which they act. I refer the reader to the debate between Bratman and McCann (1991). It is noteworthy that the examples in question raise several vexed issues. One of these issues is whether it can be rational to try to *A* without its being rational to intend to *A*. Another issue is whether one can intend to try to *A* (where the intention represents *A*-ing as a desired goal, aim or purpose) without intending to *A*.

Consider, finally, the objection to the Simple View that there are intentional actions which are the foreseen, *non-intended* behavioural side effects of other intentional actions. Two concerns come into focus in evaluating this range of counterexamples. The first - and more basic - concern is: when, if ever, do expected behavioural side effects of intentional actions properly count as non-intended? For holists about intention urge that all foreseen side effects of an intentional action - or all seriously considered ones - properly count as intended. The second concern, which accepts that there are expected, *non-intended* behavioural side effects of intentional actions, is the following: when, if ever, are such side effects themselves intentional actions? A common view holds that when expected, non-intended side effects enter - or ought to enter - into one's practical reasoning as (significant) reasons against the action one settles on, such side effects count as intentional actions. On this view, for example, someone who foresees that leaving the party early will insult their host and reluctantly goes ahead - in order to get home to a sick friend by an agreed time - counts as intentionally insulting the host, even though they do not in the least intend to do this. If the common view in question is acceptable, the Simple View is false.

6 Intentions and reasons

When philosophers discuss reasons for intending to act, they normally concentrate on object-related considerations, such as ones which support the claim that the intention is an intention to do something which is of value in some way. Such reasons, which coincide with reasons for acting, do not exhaust the kinds of reasons there are for intending to act, however. For there are also attitude-related reasons, such as considerations that support the claim that the intention is one which it is in one's interest to have. This distinction between object- and attitude-related reasons applies across the range of propositional attitudes - as when, say, evidential reasons for believing are contrasted with prudential or hedonic ones. The distinction bears on rationality, since it suggests that, corresponding to the two broad kinds of reasons, the rationality of propositional attitudes like intentions can be assessed along two different dimensions.

See also: Action; Belief; Communication and Intention; Desire; Propositional Attitudes; Rationality, practical

ROBERT DUNN

References and further reading

Audi, R. (1986) 'Intending, Intentional Action, and Desire', in J. Marks (ed.) *The Ways of Desire*, Chicago, IL: Precedent.(A good representative of Audi's reductive theory of intention, which analyses intentions in terms of beliefs about what one (probably) will do and predominant desire. Relevant to §1-2.)

Bratman, M. (1985) 'Davidson's Theory of Intention' (with appendix), in E. LePore and B. P. McLaughlin (eds) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell, 14-28. (Contains Bratman's criticisms of Davidson's evaluative theory of intention, together with his response to Davidson's rejoinder. Relevant to §3.)

Bratman, M. (1987) *Intention, Plans and Practical Reason*, Cambridge, MA: Harvard University Press.(For discussion relevant to: §1, see chapter 3; §2, see chapters 2 and 7; §5, see chapters 8 and 10.)

Davidson, D. (1978) 'Intending', in *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, 83-102. (Davidson's classic article on intentions as all-out evaluative judgments.)

Davidson, D. (1985) 'Replies to Essays I-IX', in B. Vermazen and M. Hintikka (eds) *Essays on Davidson: Actions and Events*, Oxford: Clarendon Press.(For discussion relevant to: §1, see 'Reply to David Pears'; §3, see 'Reply to Michael Bratman', 'Reply to Paul Grice and Judith Baker' and 'Reply to Christopher Peacocke'.)

Dunn, R. (1992) 'Akratic Attitudes and Rationality', *Australasian Journal of Philosophy* 70: 24-39.(Applies the distinction between object- and attitude-related reasons to an account of the rationality, and then of *akrasia*, of propositional attitudes like intentions. Relevant to §6.)

Hampshire, S. (1975) *Freedom of the Individual*, Princeton, NJ: Princeton University Press.(Argues for a contrast between non-inductive and inductive knowledge of the future - the former being had in virtue of one's having formed intentions to act. Relevant to §1.)

Harman, G. (1976) 'Practical Reasoning', *Review of Metaphysics* 79: 431-63. (Argues that intentions are

self-referential. Relevant to §4.)

Harman, G. (1983) 'Rational Action and the Extent of Intentions', *Social Theory and Practice* 9: 123-41. (Criticizes the version of holism about intention which holds that all foreseen side effects of an intentional action properly count as intended. See §5.)

Harman, G. (1986) *Change in View*, Cambridge, MA: MIT Press.(Revised version of Harman (1983) appears as chapter 9. For discussion relevant to various sections in this entry, see also chapter 8.)

McCann, H. (1991) 'Settled Objectives and Rational Constraints', *American Philosophical Quarterly* 28: 25-36. (A good representative of McCann's defence of the Simple View. Referred to in §5.)

Mele, A. R. (1992) *Springs of Action*, New York: Oxford University Press. (Canvasses most central issues in the causal theory of intention and includes a comprehensive bibliography. For discussion relevant to: §1, see chapter 8; §2, see chapters 3, 8, 9 and 10; §4, see chapter 11; §5, see chapters 8 and 10.)

Mele, A. R. (1992a) 'Recent Work on Intentional Action', *American Philosophical Quarterly* 29: 199-217. (Excellent survery of important literature (1980-92) relating to most sections in this entry.)

Peacocke, C. (1985) 'Intention and *Akrasia*', in B. Vermazen and M. Hintikka (eds) *Essays on Davidson: Actions and Events*, Oxford: Clarendon Press, 51-73. (Includes Peacocke's criticisms of Davidson's 'pf' judgments. Relevant to §3.)

Peacocke, C. (1986) 'Intention and Practical Reasoning: A Reply to Donald Davidson', *Analysis* 46: 45-9. (Contains Peacocke's response to Davidson's rejoinder to Peacocke (1985). Relevant to §3.)

Searle, J. (1983) *Intentionality*, Cambridge: Cambridge University Press. (Presents the theory that intentions are self-referential in chapter 3. See §4.)

Intentionality

Intentionality is the mind's capacity to direct itself on things. Mental states like thoughts, beliefs, desires, hopes (and others) exhibit intentionality in the sense that they are always directed on, or at, something: if you hope, believe or desire, you must hope, believe or desire something. Hope, belief, desire and any other mental state which is directed at something, are known as intentional states. Intentionality in this sense has only a peripheral connection to the ordinary ideas of intention and intending. An intention to do something is an intentional state, since one cannot intend without intending something; but intentions are only one of many kinds of intentional mental states.

The terminology of intentionality derives from the scholastic philosophy of the Middle Ages, and was revived by Brentano in 1874. Brentano characterized intentionality in terms of the mind's direction upon an object, and emphasized that the object need not exist. He also claimed that it is the intentionality of mental phenomena that distinguishes them from physical phenomena. These ideas of Brentano's provide the background to twentieth-century discussions of intentionality, in both the phenomenological and analytic traditions. Among these discussions, we can distinguish two general projects. The first is to characterize the essential features of intentionality. For example, is intentionality a relation? If it is, what does it relate, if the object of an intentional state need not exist in order to be thought about? The second is to explain how intentionality can occur in the natural world. How can merely biological creatures exhibit intentionality? The aim of this second project is to explain intentionality in non-intentional terms.

1 The history of the concept of intentionality

The term 'intentionality' derives from the medieval Latin *intentio*. Literally, this means a tension or stretching, but it is used by scholastic philosophers of the thirteenth and fourteenth centuries as a technical term for a concept. This technical term was a translation of two Arabic terms: *ma' qul*, Al-Farabi's translation of the Greek *noema*; and *ma' na*, Avicenna's term for what is before the mind in thought (see al-Farabi §3; Ibn Sina §3). In this context, the terms *noema*, *ma' qul*,*ma' na* and *intentio* can be considered broadly synonymous: they are all intended as terms for concepts, notions or whatever it is which is before the mind in thought (see Knudsen 1982). Scholars translate *intentio* into English as 'intention' - but it should be borne in mind throughout that this is not meant to have the connotations of the everyday notion of intention.

Medieval logicians followed al-Farabi in distinguishing between first and second intentions. First intentions are concepts which concern things outside the mind, ordinary objects and features of objects. Second intentions are concepts which concern other intentions. So, for example, the concept *horse* is a first intention since it is concerned with horses, but the concept *species* is a second intention, since it is concerned with first intentions like the intention *horse* (because of the nominalism prevalent at the time, the distinction between the concept/intention *horse* and the property of being a horse is not always clearly made). Many of the medieval philosophers, including Roger Bacon, Thomas Aquinas and John Duns Scotus, followed Avicenna in holding that second intentions were the subject matter of logic (see Logic, medieval §4).

Some of these philosophers developed detailed theories about how intentions were connected to the things they concerned - what we would now call theories of intentionality. One of the most influential theories was that of Aquinas, whose starting point was Aristotle's theory of thought and perception. According to Aristotle, in thought and perception the mind takes on the form of the thing perceived, without receiving its matter. When I think about or perceive a horse, my mind receives the form of horse (see Sorabji 1991; see Aristotle §18). Aquinas developed Aristotle's view. When I think about a horse, the form of horse exists in my mind. But the form has a different kind of existence in my mind than it does in a real horse. In a real horse, the form of horse has *esse naturale* or existence in nature; but in my thought of a horse, the form of horse has *esse intentionale* or intentional existence (see Anscombe and Geach 1961; Kenny 1984). The heart of Aquinas' view is that what makes my thought of an X a thought of an X is the very same thing which makes an X an X: the occurrence of the form of X. The difference is the way in which the form occurs (see Aquinas §11).

These scholastic terms largely disappeared from use during the Renaissance and the modern period. Empiricist and rationalist philosophers were of course concerned with the nature of thought and how it relates to its objects, but

their discussions were not cast in the terminology of intentionality. The terminology was revived in 1874 by Franz Brentano, in his *Psychology from an Empirical Standpoint*. In a well-known passage, Brentano claimed that:

> Every mental phenomenon is characterized by what the scholastics of the Middle Ages referred to as the intentional (and also mental) inexistence of the object, and what we, although with not quite unambiguous expressions, would call relation to a content, direction upon an object (which is not here to be understood as a reality) or immanent objectivity.
>
> (Brentano [1874] 1973: 88)

A few clarifications of this passage are needed. First, Brentano is not particularly concerned to distinguish between a mental state's (or as he called it, a mental act's) relation to a content and its relation to an object - although as we shall see in §2, later writers find a related distinction useful. And second, intentional inexistence does not itself mean that the objects of thought need not exist - although as we shall see, this is a relatively uncontroversial feature of intentionality. What inexistence means is rather that one thing - the object of thought - exists in another, as the object of the mental state itself (see Bell 1990, ch. 1).

Brentano's account of intentionality was developed by his student Edmund Husserl, who reintroduced the Greek term *noema* (plural: *noemata*) for that which accounts for the directedness of mental states. *Noemata* are neither part of the thinking subject's mind nor the objects thought about, but abstract structures that facilitate the intentional relation between subject and object. So *noemata* are not the objects on which intentional states are directed, but it is in virtue of being related to a *noema* that any intentional state is directed on an object at all. In this respect the concept of a *noema* resembles Frege's concept of sense: senses are not what our words are about, but it is in virtue of expressing a sense that words are about things at all (see Frege, G. §3). In other respects, however, senses and *noemata* differ - for instance, *noemata*, unlike Frege's senses, can be individuated in terms of perceptual experiences (see Dreyfus 1984). The point of Husserl's phenomenological reduction was to provide an account of the structure of *noemata* (see Phenomenology, epistemic issues in).

A striking claim of Brentano's is that intentionality is what distinguishes mental from physical phenomena:

> This intentional inexistence is exclusively characteristic of mental phenomena. No physical phenomenon manifests anything similar. Consequently, we can define mental phenomena by saying that they are such phenomena as include an object intentionally within themselves.
>
> (Brentano [1874] 1973: 88)

However, it is important to stress that by 'physical phenomena', Brentano does not mean physical objects. Phenomena are what are given to the mind, and Brentano does not believe that physical objects are given to the mind (see Brentano [1874] 1973: 77-78). The distinction he is making is among the data of consciousness, not among entities in the world: among these data, mental phenomena are those which exhibit intentionality, and physical phenomena are those which do not.

However, in analytic philosophy in the second half of the twentieth century, Brentano's distinction came to be interpreted as a distinction between entities in the world. This was chiefly because of this period's prevailing realism. An important figure in this revival of interest in Brentano's notion of intentionality was R.M. Chisholm. In chapter 11 of *Perceiving* (1957), Chisholm argued against the behaviourism that was popular at the time by showing that it is not possible to give a behaviouristic account of, for example, belief, since in order to say how belief leads to behaviour one has to mention other intentional states (such as desires) whose connections with behaviour must themselves be specified in terms of belief and other intentional states (see Behaviourism, analytic). This suggests that we should postulate an irreducible category of intentional mental entities: reductive physicalism must be false. However, the argument can be taken in another way, as W.V. Quine argued: if we assume reductive physicalism, we can take the irreducibility of intentionality to demonstrate the 'baselessness of intentional idioms and the emptiness of a science of intention' (Quine 1960: 221). Work on intentionality in the analytic tradition in the 1980s and 1990s has attempted to resolve this dilemma. For example, Fodor (1987), Dretske (1980) and others have attempted to reconcile physicalism with the existence of intentionality by explaining it in non-intentional terms.

2 The nature of intentionality

Despite the interest in intentionality in twentieth-century philosophy, there is still controversy about how to characterize it. All writers agree that intentionality is the directedness of the mind upon something, or the aboutness of mental states, but the disagreements start when we try to explain these ideas in more detail.

To begin with, calling intentionality ’directedness’ makes it look as if it is a relation between the mind and the thing on which the mind is directed. After all, if A is directed on B, then A and B are related - if an arrow is directed on a target, the arrow and the target are related. But if the arrow is genuinely related to the target, then the arrow and the target must exist. And similarly with other relations: if Antony kisses Cleopatra, Antony and Cleopatra must exist. But this is not so with intentionality, as Brentano observed (Brentano [1874] 1973: appendix). I can desire to possess a phoenix without there being any such thing. So what am I related to when I am in an intentional state?

One reaction is to postulate that intentional relations are relations to intentional objects. A phoenix is not the material object of my desire, but it is the desire’s intentional object. However, it is not obvious that this really solves any problems. For what are intentional objects? Are they real objects? Brentano and Husserl both thought so: intentional objects are just ordinary objects. But if this is so, then what does it mean to say that intentional objects need not exist? Alexius Meinong, on the other hand, thought that intentional objects have a different kind of existence from real material objects. But this seems to misrepresent intentionality: if I want a phoenix, I want a real phoenix, with wings and feathers - not something with a different kind of existence. In any case, the idea that there are different kinds of existence is of dubious coherence. So whatever we say about intentional objects, they do not offer a satisfactory explanation of intentionality (see Scruton |1970-1).

A second important difference between intentionality and other relations is that with other relations, the way you describe the relata does not affect whether the relation holds. But with intentionality this is not so: you can believe that George Orwell wrote *Animal Farm* without believing that Eric Blair wrote *Animal Farm*, simply because you do not know that Orwell is Blair. But since Orwell is Blair, then your belief surely relates you to the same thing - so how can the obtaining of the relation (belief) depend on how the thing is described?

For these reasons, it seems impossible to regard intentionality as a relation at all. One way of avoiding these difficulties is to distinguish, as Brentano did not, between the intentional object of a state and its intentional content. Intentional content (like Husserl’s *noema* or Frege’s *thought*) is what makes it possible for a mental state to be directed on an object. Thus understood, intentional contents are not representations. Rather, they are what constitute something’s being a representation: it is in virtue of the fact that a mental state has an intentional content that it represents what it does. It is in virtue of the fact that my belief that pigs fly involves some relation between me and an intentional content - the proposition or Fregean thought that pigs fly - that it represents what it does. This is what is meant by saying that intentional states, or propositional attitudes, are relations to propositions or contents (see Propositional attitudes).

Although beliefs, desires and other intentional states are sometimes described in this way - as relations to propositions or contents - this idea should be sharply distinguished from the idea, just discussed, that intentional directedness is a relation. The intentional content expressed by the sentence ‘Pigs fly’ is not what my belief that pigs fly is directed on: the belief is directed on pigs and flying. Some have thought that all intentional directedness can ultimately be reformulated in terms of relations to intentional contents: an intentional state is directed on an object X in virtue of the fact that it is a relation to an intentional content concerning X. However, this thesis has difficulty dealing with certain intentional phenomena, most notably the phenomenon of loving: no one has given a satisfactory reformulation of the notion ’X loves Y’ in terms of X’s relations to intentional contents.

There is much controversy about exactly what intentional contents are and how we should individuate them (see Salmon and Soames 1988). Some philosophers attempt to clarify (or even sidestep) these ontological and epistemological difficulties by adopting what Quine calls ’semantic ascent’: they examine sentences which report intentionality rather than intentionality itself. A distinctive feature of many sentences reporting intentional states is that their constituent words do not play their normal referential role. Part of what this means is that the apparently uncontroversial logical principles of existential generalization (from Fa infer $(\exists x)Fx$) and Leibniz’s Law (from Fa and $a=b$ infer Fb) fail to apply to all sentences reporting intentionality. For example, from ‘I want a phoenix’ we cannot infer that there exists a phoenix that I want; and from ‘Vladimir believes that Orwell wrote *Animal Farm*’ and ‘Orwell is Blair’ we cannot infer that ‘Vladimir believes that Blair wrote *Animal Farm*’ (see Propositional

attitude statements).

Contexts where these two principles fail to hold are known as ’non-extensional’ contexts - their semantic properties depend on more than just the extensions of the words they contain. They are also called ’intensional’ contexts, or contexts which exhibit intensionality (see Intensionality). The connection between intensionality and intentionality is not merely typographical: the failure of existential generalization in intensional contexts is the logical or linguistic analogue of the fact that intentional states can be about things which do not exist. And the failure of Leibniz’s Law is the logical or linguistic analogue of the fact that the obtaining of an intentional relation depends on the way the relata are characterized.

However, the notion of intensionality must be distinguished from the notion of intentionality, not least because there are intensional contexts which are nothing to do with the direction of the mind on an object. Prominent among these are modal contexts: for example, from ‘Necessarily, Orwell is Orwell’ and ‘Orwell is the author of *Animal Farm*’ we cannot infer ‘Necessarily, Orwell is the author of *Animal Farm*’. Other concepts which can create intensional contexts are the concepts of probability, explanation and dispositionality. But it is very controversial to hold that these concepts have anything to do with intentionality.

Another (more controversial) reason for distinguishing between intensionality and intentionality is that intentionality can be reported in sentences which are extensional. Some philosophers have argued that the context ’x sees y’ is like this. Seeing seems to be a paradigm case of the direction of the mind on an object. But if Vladimir sees Orwell, then there is someone whom he sees; moreover, if Vladimir sees Orwell, then surely he also sees Blair, and he also sees the author of *Animal Farm*, and so on. So although seeing is intentional, ’x sees y’ seems to be an extensional context.

3 Intentionality as the mark of the mental

Thus, the notion of intensionality cannot provide a purely logical or semantic criterion of intentionality. We should be content with the psychological criterion: intentionality is the directedness of the mind upon something. As I remarked earlier, Brentano thought that intentionality was the mark of the mental: all and only mental phenomena exhibit intentionality. In discussing this claim - often called Brentano’s thesis - I shall follow analytic philosophers in ignoring Brentano’s own quasi-idealistic use of the term ’phenomenon’. Brentano’s thesis shall be taken as a thesis about the distinction between mental and physical entities in the world.

Is Brentano’s thesis true? We can divide this question into two sub-questions: (1) Do all mental states exhibit intentionality? (2) Do only mental states exhibit intentionality?

(1) It is natural at first sight to think that there are many kinds of mental state which do not have any intentionality. For instance, there are states like undirected anxiety, depression and elation (see Searle 1983: 2). On what are these states directed? Well, I can be anxious without being anxious about anything in particular - but this anxiety is at least directed at myself. Other popular examples of supposedly non-intentional mental states are sensations like pain. But while it may be true that pains are not propositional attitudes - if propositional attitudes are states reportable by sentences of the form ’X φs that p’, where φis a psychological verb - this does not mean that pains are not directed on anything. I could have two pains, one in each hand, which felt exactly the same, except that one felt to be in my right hand, and the other felt to be in my left hand. This is a difference in intentionality - in what the mental state is directed on - so it is not true that pains exhibit no intentionality (see Bodily sensations §2).

However, there are properties of pains which do seem to be wholly non-intentional, such as the naggingness of a toothache (see Qualia). And these properties seem to be essential to pains. This shows that the distinction we need is between those mental states whose whole nature is exhausted by their intentionality, and those whose whole nature is not. Pains are in the latter category, since they seem to have essential non-intentional properties: there are elements of pains which are not exhausted by whatever intentionality those pains may have.

(2) So much, then, for the idea that all mental states exhibit intentionality. But is intentionality only exhibited by mental states? That is: is it true that if something exhibits intentionality, then that thing is a mind? Are minds the only things in the world that have intentionality?

To hold that minds are not the only things that have intentionality, we need an example of something that has intentionality but does not have a mind. This may seem easy. Take books: books contains sentences which have

meaning and are therefore directed at things other than themselves. But books do not have minds.

The natural reply to this is to say that the book's sentences do not have intentionality in themselves - they do not have what some call 'original' intentionality - but only because they are interpreted by the readers and writer of the book. The intentionality of the book's sentences is derived from the original intentionality of the states of mind of the author and reader who interpret those sentences (for this distinction, see Haugeland 1990).

So we can reframe our question as follows: can anything other than minds exhibit original intentionality? One problem with this question is that if we encountered something that exhibited original intentionality, it is hard to see how it could be a further question whether that thing had a mind. The notion of intentionality is so closely bound up with mentality that it is hard to conceive of a genuine case of original intentionality that is not also a case of mentality. If, for example, we could establish that computers were capable of original intentionality, it would be natural to describe this as a case where a computer has a mind.

However, there is an interesting way in which original intentionality and mentality could come apart. Some philosophers want to locate the basis of intentionality among certain non-mental causal patterns in nature. So on this view, there would be a sense in which original intentionality is manifested by things other than minds. This is the hope of those philosophers who attempt to reduce the intentional to the non-intentional: the hope summed up by Jerry Fodor's quip that 'if aboutness is real, it must really be something else' (Fodor 1987: 97).

These philosophers are in effect trying to steer a course between the two horns of the dilemma presented by the passage from Quine's *Word and Object* quoted in §1: you can respond to the Chisholm-Brentano thesis of the irreducibility of intentionality either by accepting an autonomous theory of intentionality and rejecting physicalism, or by denying the reality of intentionality. There are those who are eliminative materialists and who deny the reality of intentionality (see Eliminativism), and there are those who are prepared to accept intentionality as an unanalysed, primitive phenomenon. But the orthodox line among late twentieth-century analytic philosophers is to reconcile the existence of intentionality with a physicalist (or naturalist) world view. This reconciliation normally takes the form of a theory of content: a specification in non-intentional terms of the conditions under which an intentional state has the intentional content it does, or concerns the object(s) it does. A common style of theory of content spells out these conditions in terms of hypothesized law-like causal relations between intentional states and their objects. The model here is the simple kind of representation or meaning found in nature: the sense in which clouds mean rain, and smoke means fire (see Dretske 1980). Causal theories of content hope to explain how the intentionality of mental states is underpinned by simple regularities like these. These theories have had great difficulty accounting for misrepresentation and the normative elements of mental states, and it is this problem that has received most attention in contemporary discussions of intentionality (see Semantics, informational; Semantics, teleological).

See also: Belief; Desire; Intention; Perception; Imagination; Emotions, nature of

TIM CRANE

References and further reading

Anscombe, G.E.M. and Geach, P.T. (1961) *Three Philosophers*, Oxford: Blackwell.(The chapter on Aquinas gives a very clear and nontechnical account of his views on intentionality.)

Bell, D. (1990) *Husserl*, London: Routledge(Clear and critical account of Husserl's philosophy. Chapter 1 is about Brentano.)

Brentano, F. (1874) *Psychology from an Empirical Standpoint*, trans. A.C. Rancurello, D.B. Terrell and L.L. McAlister, London: Routledge & Kegan Paul, 1973.(The classic work that revived the concept of intentionality when it was first published in 1874. Difficult reading.)

Chisholm, R.M. (1957) *Perceiving: A Philosophical Study*, Ithaca, NY: Cornell University Press.(Chapter 11 contains the argument for the irreducibility of intentionality described in §1.)

Dretske, F.I. (1980) 'The Intentionality of Cognitive States', in P.A. French *et al*. (eds) *Midwest Studies in Philosophy* V, Minneapolis, MN: University of Minnesota Press.(Useful and accessible summary of Dretske's reductive theory of content.)

Dreyfus, H.L, with Hall, H. (eds) (1984) *Husserl, Intentionality and Cognitive Science*, Cambridge, MA: MIT Press.(Wide-ranging collection of essays on Husserl's ideas on intentionality and their relation to cognitive science.)

Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.(Influential account of intentionality from a physicalist perspective. Chapter 4 contains Fodor's theory. Original but not excessively technical.)

Haugeland, J. (1990) 'The Intentionality All-Stars', in J. Tomberlin (ed.) *Philosophical Perspectives 4: Action Theory and the Philosophy of Mind*, Atascadero, CA: Ridgeview.(Clear and readable survey of research in intentionality in the analytic tradition.)

Kenny, A. (1984) 'Aquinas: Intentionality', in T.Honderich (ed.) *Philosophy Through Its Past*, Harmondsworth: Penguin.(A straightforward discussion of Aquinas' theory of intentionality and an attempt to relate it to Wittgenstein's views.)

Knudsen, C. (1982) 'Intentions and impositions', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press. (Survey of the idea of an *intentio* as employed by the leading medieval philosophers.)

Quine, W.V.O. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Chapter 6 contains Quine's influential discussion of intentionality and intensionality discussed in §2. A classic, but a difficult work for the beginner.)

Salmon, N. and Soames, S. (eds) (1988) *Propositions and Attitudes*, Oxford: Oxford University Press.(Useful collection of essays, many of them classics, on the notions of proposition and intentional content. Some technical logical material.)

Scruton, Roger (1970-1) 'Intensional and Intentional Objects', *Proceedings of the Aristotelian Society* 92, 187-207.(Clear and thorough discussion of the notion of an intentional object.)

Searle, J.R. (1983) *Intentionality*, Cambridge: Cambridge University Press.(A complete theory of intentionality, with material relating to the issues discussed in §§2-3. Chapter 1 is a good introduction.)

Sorabji, R. (1991) 'From Aristotle to Brentano: The Development of the Concept of Intentionality', in H. Blumenthal and H. Robinson (eds) *Oxford Studies in Ancient Philosophy*, supplementary vol. Oxford: Oxford University Press.(Informative discussion of how the concept of intentionality developed, particularly in later Greek philosophy, also with reference to Islamic and scholastic writers.)

Spiegelberg, Herbert (1976) '"Intention" and "Intentionality" in the Scholastics, Brentano and Husserl', trans. L. McAlister, in L. McAlister (ed.) *The Philosophy of Brentano* London: Routledge & Kegan Paul, 1976, ch. 9.(A standard source for the origins of the different uses of these terms.)

Internalism and externalism in epistemology

The internalism-externalism distinction is usually applied to the epistemic justification of belief. The most common form of internalism (accessibility internalism) holds that only what the subject can easily become aware of (by reflection, for example) can have a bearing on justification. We may think of externalism as simply the denial of this constraint.

The strong intuitive appeal of internalism is due to the sense that we should be able to determine whether we are justified in believing something just by carefully considering the question, without the need for any further investigation. Then there is the idea that we can successfully reply to sceptical doubts about the possibility of knowledge or justified beliefs only if we can determine the epistemic status of our beliefs without presupposing anything about which sceptical doubts could be raised - the external world for example.

The main objections to internalism are: (1) It assumes an unrealistic confidence in the efficacy of armchair reflection, which is often not up to surveying our entire repertoire of beliefs and other possible grounds of belief and determining the extent to which they support a given belief. (2) If we confine ourselves to what we can ascertain on reflection, there is no guarantee that the beliefs that are thus approved as justified are likely to be true. And the truth-promoting character of justification is the main source of its value.

Externalism lifts this accessibility constraint, but in its most general sense it embodies no particular positive view. The most common way of further specifying externalism is reliabilism, the view that a belief is justified if and only if it was produced and/or sustained by a reliable process, one that would produce mostly true beliefs in the long run. This is a form of externalism because whether a particular belief-forming process is reliable is not something we can ascertain just on reflection. The main objections to externalism draw on internalist intuitions: (1) If the world were governed by an evil demon who sees to it that our beliefs are generally false, even though we have the kind of bases for them we do in fact have, then our beliefs would still be justified, even though formed unreliably. (2) If a reliable clairvoyant (ones who 'sees' things at a great distance) forms beliefs on this basis without having any reason for thinking that they are reliably formed, those beliefs would not be justified, even though they pass the reliability test.

1 1 Forms of the distinction

Although this entry is restricted to epistemology, the terms 'internalism' and 'externalism' are also used to mark distinctions in ethics and philosophy of mind (see Content: wide and narrow; Moral motivation §1). Within epistemology itself the terms are used variously. The most basic distinction is between views concerning the epistemic justification of belief and views concerning knowledge. As applied to either, *internalism* is construed variously - for example, as the irreducible normativeness of justification (knowledge), and as the view that justification or knowledge always depends on the subject's belief system. The most common understanding, however, is Access Internalism - the view that only what is cognitively accessible to the subject in some strong fashion can have any bearing on justification. We may think of *externalism* as simply the denial of this restriction. Strong cognitive access to a fact is variously conceived; the most common version is that the fact be ascertainable by the subject just by reflecting on the matter. If my justified belief that Susie is in Panama is to justify my belief that she will not be at the meeting tonight, then it must be that I can ascertain by armchair reflection that I justifiably believe that she is in Panama (Ginet 1975: 34; Pollock 1986: Ch. 5, §4.1 offers a quite different way of thinking of the internalist requirement.) Note that this constraint does not require that individuals actually have the knowledge that the alleged justifier obtains, but only that they be capable of acquiring it on reflection.

A crucially important, but usually ignored, distinction is between access to the justifier and access to the epistemic efficacy of the justifier. Consider the perceptual belief that there is snow in front of my house. The justifier, let us say, is my visual presentation as I look out the front window. That would seem unproblematically to satisfy the internalist constraint. But in order for me to know that the belief is justified, it is not enough that I know that my current visual presentation is of such-and-such a character. I must also know that this presentation suffices to justify the belief (that it possesses that 'justificatory efficacy'). And it is far from unproblematic that this is accessible just on reflection. Indeed, many philosophers have struggled in vain to show that our ordinary perceptual beliefs about the immediate environment are justified by the sensory experiences on which they are

typically based. As we shall see, the most common arguments for accessibility internalism give equal support to both kinds of internalist constraint. If one advocates only one kind, we have a mixed view of the sort to be described below.

2 Justification internalism - pro and con

The most explicit arguments for accessibility internalism have proceeded from a deontological conception of justification, according to which a belief is justified only if having the belief does not violate any intellectual obligation or requirement, only if one is permitted to have the belief. In Ginet (1975) the argument goes as follows:

(1) One should believe that *p* only if one is (epistemically) permitted to do so.
(2) One is permitted to believe that *p* only if one has a justification for the belief.
(3) But a condition cannot be required for permission unless one is able to determine whether it obtains.
(4) Hence one can always determine whether one has a justification for a belief.

Note that this argument is designed to support both kinds of accessibility requirement. For to determine 'whether one has a justification for a belief' one must be able to spot both the allegedly justifying condition and also determine whether it suffices to do the job.

One trouble with this argument is that, even if cogent, it does not show that justification must be ascertainable simply on reflection. Its conclusion is only that the subject must be able to know of it somehow or other. In addition, there are serious questions about the deontological conception of justification on which the argument is based: it would seem to assume an unwarranted degree of voluntary control of beliefs (Alston 1989: 115-52).

To be sure, accessibility internalism has a strong intuitive appeal. It has seemed obvious to many philosophers through the centuries that we should be able to determine the epistemic status of our beliefs just by armchair reflection, and that has been the standard method, at least since Descartes shut himself up in a small room to determine whether there was anything he knew with certainty. But intuitive plausibility, even when combined with hallowed tradition, is not enough to bear the weight of so strong a restriction. Moreover, this plausibility stems partly from confusions, particularly that between the activity of justifying a belief and the state of being justified in believing that *p*. The former does seem to presuppose reflective access. At least I cannot adduce a support in justifying (arguing for) my belief that *p* unless I can become aware of that support. But we cannot infer from this that there is the same requirement for being justified in believing that *p*. Obviously, I am justified in many beliefs that I have never engaged in justifying. Otherwise I would have precious few justified beliefs, since I spend little time justifying my beliefs.

Another common motive for embracing internalism is the supposition that it removes an obstacle to giving an effective response to the sceptical denial or doubt that we have any knowledge or justified beliefs. If I have to rely on what I think I have learned from perception, induction, explanation or scientific theorizing, in order to show that certain of my beliefs are justified, then my attempt is open to sceptical challenges to those sources. If I only need to rely on reflection, it is supposed, I do not have this worry. But this line of thought deals only with certain kinds of scepticism. If my only appeal is to reflection, I am immune to challenge from scepticism about perception and induction. But that leaves scepticism about reflection. Though historically most sceptics have concentrated on perception, induction and high level reasoning, some of the classical sceptical arguments would apply to reflection as much as to anything else. Hence internalism provides only a limited guarantee against sceptical doubts (see Scepticism).

The very features of internalism that make it attractive to many also give rise to some serious liabilities. For one thing the position would seem to be much too sanguine concerning the cognitive powers of mere reflection. Here the above distinction between access to the justifier and access to its justificatory efficacy is relevant. Claims to reflective access to the latter are very dubious. Some epistemologists (Chisholm [1966] 1989: 7, 76) exhibit confidence that, just by thinking about the matter, one can tell what justifies what, but I find this very questionable. It partly depends on how we think of justification. The deontological conception may seem more friendly to internalism here. It is easy to convince oneself that one can tell what is permitted to one or required of one just by carefully considering the matter. But if being justified in believing that *p* involves believing in such a way as to be in a strong position to get the truth, then we must deny, must we not, that we are generally able to determine *that*

just by raising the question? Will that tell us, for example, how much and what kind of evidence will put us in a position to get the truth in our scientific theorizing or our religious beliefs?

This issue will recur in the second objection. For now we should note also that not everything is clear sailing *vis-à-vis* access to the justifier. For example, a common type of alleged justifier consists in other justified beliefs of the subject (my belief that my wife is not at home is justified because it is based on my justified belief that her car is not in the garage). But then to know that I have a justifier of this sort I would not only have to know that I have this belief (which may well be accessible on reflection) but also that it is justified. And that brings us back to justificatory efficacy again, namely the issue of what it takes to justify the belief that my wife's car is not in the garage. Again, suppose that a certain belief is justified only if my total evidence (the set of all my justified beliefs) renders that belief probable. Then we also run into the difficulty of determining just by reflection the entire range of my beliefs, and it is doubtful that I can accurately survey *all* my beliefs just by thinking about the matter.

The second objection involves the point just noted that it seems clearly wrong to suppose that we can determine by armchair reflection whether certain (allegedly justifying) conditions render a belief likely to be true. If a belief's being justified implies the probability of its truth, then it follows that anything we can ascertain by reflection will not amount to justification in this sense. And surely this implication of the probability of truth is an essential part of what makes it desirable and important that our beliefs be justified. If this implication is lacking why should we care whether our beliefs are justified? After all, the basic aim of cognition is to believe what is true and to avoid believing what is false. Thus internalism would seem disconnected from any understanding of justification that renders it a major epistemic desideratum.

In the face of these difficulties we should note that there is a weaker brand of internalism on the market, according to which what bears on the justification of belief is restricted to the evidence, reasons or experiences of the subject, in contrast to what is external to the subject's cognitive states (Feldman and Conee 1985). This position is not subject to the objections we have been surveying. But the opposite side of that coin is that it cannot draw on the intuitive support we have seen to accrue to accessibility internalism.

3 Justification externalism - pro and con

In the most general sense, externalism is simply a rejection of the internalist constraint on access. (Externalism can allow reflectively accessible factors, but it denies that what affects justificatory status is *restricted* to them.) Externalism in this generic sense does not include any particular positive view of justification, but it is able to consider various possibilities that internalism blocks. In fact, externalism is always associated with 'strong position' conceptions of justification (believing in such a way as to be in a strong position to thereby get the truth). Since it can recognize conditions of justification that are not reflectively accessible, it is able uninhibitedly to carry out the idea that justification entails truth-conducivity.

The most prominent form of externalism is reliabilism, and in its most common form it holds that a belief is justified if and only if it was produced and/or sustained by a reliable belief-forming process (Goldman 1986) (see Reliabilism §2). A particular perceptual belief, for example, is justified provided it was formed from sensory experience in such a way that when a belief with a content like that is formed from an experience related in that way to the belief content, it will usually be true. A belief formed by inferring a fact from observable indications (believing that a party is going on because one notes a lot of lights on and hears a lot of noise) is justified provided beliefs with that sort of content formed by indications like that will generally be correct. Since the truth-making tendency of a certain process is not something we can ascertain just on reflection, access internalists are not in a position to embrace reliabilism, but externalists are free to do so.

There are many problems one encounters in working out a reliabilist view in detail. For example, how does one assign a particular belief-forming process to a class or type of processes? (note the use of 'like that' in the above examples) (Feldman 1985). And is the reliability a matter of actual track record or a matter of what *would* happen in situations of certain kinds? I will just note here that some views of this kind are formulated in terms of probability rather than in terms of processes. Thus the second example could be put in terms of the probability of a party occurring given that there is a lot of noise, and so on (Swain 1981; Dretske 1981) (see Information theory and epistemology; Probability theory and epistemology §1).

Since externalism amounts to a denial of internalism, it is to be expected that its main supports come from the

difficulties in internalism. The main attraction of externalism lies in the facts that: (1) it enables us to retain the truth-conducivity implication of justification, as internalism does not; and (2) it does not require us to make exaggerated claims for the powers of rational reflection.

The chief criticisms of externalism are based on internalist intuitions and are typically directed against reliabilism. First, there is the consideration of 'demon worlds', worlds controlled by an evil omnipotent demon who sees to it that our beliefs are generally false, even when they seem most obviously true. In such a world one's perceptual, inductive, and mathematical beliefs are false even though one has the same bases for them that we have in the actual world. The denizens of this world believe that there is a tree in front of them when they seem to be seeing a tree in front of them. They believe that 3 + 2 = 5 because it seems self-evident to them, and so on. There is a considerable tendency to judge that these unfortunates would be justified in their beliefs, just as much as we are in ours, even though they are formed in a very unreliable manner. Since they have the same grounds as we do, how can our beliefs be justified and theirs not? But then, reliability of formation is not necessary for knowledge (Foley 1987: Ch. 3). Although this argument is couched in terms of a fantastic scenario, the underlying idea is simple and sober. Since it is conceivable that we have what are recognized as very strong grounds for a belief even though it is not formed reliably, reliability cannot be *necessary* for justification.

Second, consider someone who forms beliefs in a reliable way but who has no, or insufficient, reason for supposing it to be reliable. Again, the argument is frequently presented in terms of *outré* examples like clairvoyance (BonJour: Ch. 3), but more standard cases will do as well. Suppose Jim infers emotional states from outward demeanour and behaviour in the standard way, but has no reason for thinking this to be a reliable way of making such judgments. There is a strong tendency to think that he has no justification for the beliefs so formed, given that this mode of belief formation could be completely unreliable so far as he can tell. But then the mere fact that the way of forming beliefs is reliable cannot be *sufficient* for justification.

The best counter-move for the reliabilist is to deny the 'intuitions' adduced by the critic. No doubt the demon world denizen has something going for them, epistemically, but, says the reliabilist, it is not *justification*, so long as there is nothing about the belief and the way it is formed or held that makes it likely to be true. And since the reliability condition is satisfied for Jim *vis-à-vis* beliefs formed in the usual way about the emotional states of others, that will suffice to render the beliefs justified, despite the lack of higher-level knowledge (justified belief) concerning how reliable that way is. This is to take truth-conducivity as both necessary and sufficient for justification and, as we have seen, this is the heart of epistemic justification from an externalist perspective.

What if Jim not only is not justified in supposing this way of forming beliefs to be reliable, but is justified in supposing it to be unreliable (even though it is reliable)? In that case the reliabilist could recognize Jim's beliefs about the emotional states of others to be *unjustified* - by taking reliability to be sufficient only for *prima facie* justified belief, belief that is justified provided the justification is not overridden in one way or another. In this case, Jim's justification for supposing that way of forming beliefs to be unreliable would be a sufficient overrider. To modify reliabilism in this way would not involve abandoning the basic thrust of the position. Reliability of belief formation would still be at the centre of the picture. It is just that other considerations are given a secondary role.

Various blends of internalism and externalism have been suggested. The distinction between 'justifier' and 'justificatory efficacy thereof' provides an obvious basis for such a combination. Thus Alston (1989: Ch. 9) develops a view ('Internalist Externalism'), according to which a condition can be a justifier only if it is the sort of thing that is typically available to reflection (this is the internalist part), but there is no such reflective access required for the justificatory efficacy of the condition. That efficacy depends on whether the alleged justifier is in fact a reliable indication of the truth of the belief in question (this is the externalist part).

4 The distinction applied to knowledge

As for knowledge, internalism would seem to be less plausible than in application to justification, while externalism is more plausible. This is because of doubts that justification of belief is a necessary condition of knowledge. The dominant view of knowledge in the twentieth century has been that it consists of justified true belief. Since Gettier posed his famous difficulty for this view, those who take true justified belief to be at least sufficient for knowledge have tried to evade these difficulties by adding a fourth condition, but with indifferent

success so far (see Gettier problems §§1-2). If justification were necessary for knowledge, then any constraint on the former would equally be a constraint on the latter, and so accessibility internalism could not be less plausible for knowledge than for justification. But if justification of belief is not necessary for knowledge, we have a different story. More than one prominent externalist has developed a conception of knowledge in which justification is not mentioned at all (Dretske 1981; Nozick 1981). Something like reliability, or probability of the belief on its grounds, is taken as sufficient to turn true belief into knowledge. For example, on Nozick's view the following four conditions are individually necessary and jointly necessary for *S* knowing, via method *M*, that *p*.

(1) *p* is true.
(2) *S* believes, via *M*, that *p*.
(3) If *p* were not true and *S* were to use *M* to arrive at a belief whether (or not) *p*, then *S* would not believe, via *M*, that *p*.
(4) If *P* were true and *S* were to use *M* to arrive at a belief whether (or not) *p*, then *S* would believe, via *M*, that *p*.

Nozick puts this by saying that a true belief counts as knowledge when it 'tracks' the truth, that is, when its being held or not varies with the truth value of the proposition believed. This form of externalism is cut loose from dependence on the right way of thinking of justification, though, of course, its success does depend on the right way of thinking of knowledge. (For some criticisms of Nozick's proposals see Luper-Foy 1987.)

See also: Justification, epistemic; Knowledge, concept of; Rational beliefs; Reasons for belief

WILLIAM P. ALSTON

References and further reading

Alston, W.P. (1989) *Epistemic Justification: Essays in the Theory of Knowledge*, Ithaca, NY: Cornell University Press.(Advanced essays on some of the main topics of current interest.)

BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press. (Accessible defence of coherence theory and excellent criticism of alternatives.)

Chisholm, R.M. (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall; 3rd edn, 1989.(Classic presentation of foundationalism.)

Dretske, F.I. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(Major statement of an externalist theory of knowledge.)

Feldman, R. (1985) 'Reliability and Justification', *The Monist* 68 (2): 159-74. (Classic presentation of the difficulty in determining the type to which to assign a particular belief-forming process.)

Feldman, R. and Conee, E. (1985) 'Evidentialism', *Philosophical Studies* 48 (1): 15-34.(A moderate form of internalism.)

Foley, R. (1987) *The Theory of Epistemic Rationality*, Cambridge, MA: Harvard University Press.(A relatively subjectivist form of internalism.)

Gettier, E. (1963) 'Is Justified True Belief Knowledge?', *Analysis* 23 (6): 121-3. (The classic argument for a negative answer.)

Ginet, C. (1975) *Knowledge, Perception, and Memory*, Dordrecht: Reidel. (Another outstanding version of internalism.)

Goldman, A.I. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(Definitive statement of reliabilism with connections to cognitive science.)

Luper-Foy, S. (1987) *The Possibility of Knowledge: Nozick and His Critics*, Totowa, NJ: Rowman & Littlefield. (Collection of essays critical of Nozick's theory of knowledge.)

Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press.(Influential version of externalism.)

Pollock, J. (1986) *Contemporary Theories of Knowledge*, Totowa, NJ: Rowman & Littlefield.(Advanced textbook and presentation of an unusual form of internalism.)

Swain, M. (1981) *Reasons and Knowledge*, Ithaca, NY: Cornell University Press. (Version of reliabilism in terms of probability.)

International relations, philosophy of

The philosophy of international relations - or more precisely its political *philosophy - embraces problems about morality in diplomacy and war, the justice of international practices and institutions bearing on economic welfare and the global environment, human rights, and the relationship between sectional loyalties such as patriotism and global moral commitments.*

Not everyone believes that such a subject can exist, or rather, that it can have significant ethical content. According to political realism - a widely-held view among Anglo-American students of international relations - moral considerations have no place in decisions about foreign affairs and international behaviour. The most extreme varieties of realism deny that moral judgment can have meaning or force in international affairs; more moderate versions acknowledge the meaningfulness of such judgments but hold either that leaders have no responsibility to attend to the morality of their actions in foreign affairs (because their overriding responsibility is to advance the interests of their constituents), or that the direct pursuit of moral goals in international relations is likely to be self-defeating.

Leaving aside the more sceptical kinds of political realism, the most influential orientations to substantive international morality can be arrayed on a continuum. Distinctions are made on the basis of the degree of privilege, if any, extended to the citizens of a state to act on their own behalf at the potential expense of the liberty and wellbeing of persons elsewhere. 'The morality of states', at one extreme, holds that states have rights of autonomy analogous to those of individuals within domestic society, which secure them against external interference in their internal affairs and guarantee their ownership and control of the natural and human resources within their borders. At the other end of the continuum, one finds cosmopolitan views which deny that states enjoy any special privilege; these views hold that individuals rather than states are the ultimate subjects of morality, and that value judgments concerning international conduct should take equally seriously the wellbeing of each person potentially affected by a decision, whether compatriot or foreigner. Cosmopolitan views may acknowledge that states (and similar entities) have morally significant features, but analysis of the significance of these features must connect them with considerations of individual wellbeing. Intermediate views are possible; for example, a conception of the privileged character of the state can be combined with a conception of the international realm as weakly normative (that is, governed by principles which demand that states adhere to minimum conditions of peaceful coexistence).

The theoretical difference between the morality of states and a fully cosmopolitan morality is reflected in practical differences about the justifiability of intervention in the internal affairs of other states, the basis and content of human rights, and the extent, if any, of our obligations as individuals and as citizens of states to help redress the welfare effects of international inequalities.

1 The scope of the subject

The philosophy of international relations is the branch of political philosophy devoted to the examination of principles of conduct for the international realm. Historically, such philosophical attention as has been paid to international relations has taken place mainly in three genres: (1) treatises on natural law, 'the law of nations' and international law; (2) the development of a doctrine of 'just' war (often in the context of the law of nations); and (3) the articulation of peace plans - that is, institutional schemes aimed at ending war. There are significant (albeit brief) passages concerning international relations in works devoted primarily to the political theory of the state by most of the important modern writers, including Machiavelli, Hobbes, Locke, Hume, Kant, Hegel, Bentham, Mill and Sidgwick. Among the few works of great stature devoted mainly to international relations, Kant's *Perpetual Peace* (1795) is arguably the most important, but its brevity and epigrammatic style render its normative content elusive.

The paucity of philosophical thought about international relations is not easy to explain, although the burden must fall on the central distinction of the international realm *vis-à-vis* the domestic - namely, the absence of a state-like structure, and consequently of an institutional focus for philosophical speculation. In recent decades there has been an increase in interest in the subject, doubtless reflecting the fundamental changes in the character of international politics and society that have occurred in the mid- and late twentieth century. An intensification of international

and transnational activity, particularly in finance and trade, has given rise to a vast increase in the number of non-state participants in international affairs - including both international organizations (that is, organizations of states) and non-governmental organizations. ‘Total war’ (characterized by the mobilization of entire societies) and the invention of chemical and nuclear weapons have rendered it virtually impossible to contain the harms of war to those who participate in combat, so that the traditional distinction between soldier and civilian has broken down (see War and peace, philosophy of §6). The exploration of space, the exploitation of ocean resources and growing political concern about the environment have spawned policy issues which are essentially trans- (rather than inter-) national. For all of these reasons, the occasion for international philosophy is greater now than at any earlier point in the modern age.

Regarded as a normative enterprise, the philosophy of international relations might be seen as an outgrowth of speculation about the basis and content of the ‘law of nations.’ But this would convey an excessively narrow conception of the subject matter of international philosophy. The law of nations was a body of norms applicable to what was conceived as a society of states whose principal forms of interaction were diplomacy and war. International relations consists of more than the diplomatic and military engagements of states, however. The actors include international and transnational organizations (the latter composed of members which are not states), business firms, and sometimes individuals, and their forms of interaction extend well beyond diplomacy and war - they include commerce, financial transactions and the exchange of information and culture. Such variety gives rise to problems of choice of great complexity and potential moral significance. The consequences of transnational interactions for the global distribution of income and wealth, the movement of human capital in the form of immigration flows, and the quality of the global environment are of particular importance for human welfare. A normative philosophy of international relations adequate to its empirical referent would embrace all of these subject matters.

2 Political realism and international scepticism

Political realism is the name of a collection of views which are often traced to Thucydides, Machiavelli and Hobbes, but whose prominence in modern international thought is due largely to a small number of recent and contemporary students of international relations such as Hans Morgenthau (1946) and George Kennan (1951). The common element is a denial that moral considerations should carry weight in decisions about foreign policy.

The most impressive statement of the realist position was given by Hobbes (1651), who compared the international realm with the state of nature in respect of the absence of a central authority, and concluded that states in international relations have the same freedom from the obligation to follow moral principles (‘laws of nature’) as do individuals in the state of nature. On the Hobbesian view, states are not obliged to comply with moral principles because they have no reason to do so in the absence of a common enforcement agency or effective conventions of reciprocal compliance. At the base of this view is a conception of morality as a mutual benefit scheme, in which individuals comply with principles restricting the direct pursuit of self-interest when and because general compliance with these principles is mutually beneficial in comparison to mutual noncompliance. In this extreme form, political realism is no more than a special case of a familiar type of moral scepticism, which denies that moral judgments can be either reasonable or motivationally efficacious if they are not backed up by considerations of advantage (see Moral scepticism; Moral motivation). If one were to reject this type of moral scepticism, then the form of political realism associated with it would fall by the wayside.

Another form of realism, found more often in the literature of policy than philosophy, holds that leaders of states are not obligated to follow higher-level moral principles because the controlling obligation of their office is to advance the state’s interest. Views of this kind are likely to seem particularly plausible in democratic states, where leaders are elected by the people and their responsibilities are understood as the advancement of the interests of those for whom they act as agents. Unlike the form of political realism considered earlier, this version is not in any obvious sense a type of moral scepticism. Indeed, it acknowledges that leaders of states, and by inference, their constituents can have moral obligations. The challenge faced by anyone who wishes to defend this view is to explain why it should be thought that leaders of states may justifiably do for their constituents what their constituents, as a group, may not justifiably do for themselves.

Another non-sceptical kind of realism, and perhaps the most persuasive of all versions of the view, is best described as heuristic. It does not deny that outcomes in foreign affairs can be appraised in terms of the good and

evil, or that actions and policies can be evaluated in terms of their rightness or wrongness. It holds, however, that outcomes which are desirable from a moral point of view are more likely to occur when foreign policy decisions are screened from the direct influence of moral considerations. The argument for this position is essentially historical: typically, that moral considerations have tended to distort practical reasoning in foreign affairs in a more or less systematic way, and that reasoning which is blocked from the direct influence of such considerations, and whose scope is limited to considerations of interest, has more often led to morally desirable results. In his Lindley Lectures (1951) - a defining document of political realism - George Kennan made an argument of this kind with respect to US diplomacy between the wars. The historical thesis plainly cannot be examined here, but it is worth observing that, even if the thesis turns out to be true, there would be no theoretical challenge to the possibility of international philosophy.

3 The morality of states and the domestic analogy

Assuming that the sceptical challenge can be overcome, the most important problem in international philosophy is the moral status of the state and the character of its relationship to other agents in the world. This problem is implicated in most of the leading normative issues in international relations - for example, the meaning of sovereignty, the permissibility of intervention, the basis and content of human rights, the extent of acceptable restrictions on the movement of persons across borders, entitlements to natural resources, and the obligations of states and their people to provide material assistance to others.

Writing about immigration policy, Sidgwick described a ‘general conflict between the cosmopolitan and the national ideals of political organization’. According to the national ideal, foreign policy should ‘promote the interests of a determinate group of human beings, bound together by the tie of a common nationality’, according to the cosmopolitan ideal, it should strive impartially to promote the interests of everyone, regardless of location or citizenship (1919: 309). The distinction between ‘national’ and ‘cosmopolitan’ ideals is fundamental in contemporary discourse about international ethics.

The dominant view in modern political theory, enduring well into the twentieth century, has been one or another variation of Sidgwick’s ‘national ideal’, or what we might more accurately call ‘the morality of states’. This conception has four distinguishing features. First, the international realm is a ‘society of states’ in which states, rather than individuals, are the principal actors and therefore the subjects of the major rights and duties. Second, states have a moral status analogous to that of individuals in domestic society. In particular, states have rights of territorial integrity and political sovereignty, analogous to the right of individual liberty, which secure their governments against external interference in the exercise of political authority over their territories and populations. Third, states are not generally responsible for the circumstances of outsiders or of other states, and within limits set by the principle of sovereignty, any individual state is entitled to assign priority to advancing the wellbeing of its inhabitants in preference to the equal wellbeing of others. Fourth, there is a high level of tolerance for diversity within the international order: states are not held to a single standard of political legitimacy, and, except perhaps in extreme cases, neither individual states nor the international community are authorized to intervene to protect a state’s people against their own government or to bring about domestic political reform.

The morality of states is a genuine morality: like individual liberty in the domestic case, sovereignty is hardly the same as licence (see Sovereignty). It is, however, peculiar in presuming that a discontinuity exists between the moral order among the individuals who cohabit a single domestic society and the more comprehensive moral order among all the individuals in the world. What can be said for this presumption?

We might begin with the picture of the international realm as a ‘society’ in which states play the roles played by persons in domestic society. This ‘domestic analogy’ is the most influential form in which the morality of states has been articulated, due mainly to the writings of eighteenth-century international jurists like Vattel (1758) and Wolff (1749) (see Wolff §7). Many find the analogy persuasive even today; for example, it is the organizing principle of Michael Walzer’s formulation of the doctrine of the just war (Walzer 1977).

Although influential, the domestic analogy faces notorious difficulties, among which the most serious is that individual persons and the collectivities to which they are compared are such fundamentally different kinds of entities. States, for example, lack the unity of consciousness and the capacity for moral personality which we presume individual moral persons to possess, and so cannot be said to be self-determining in the same sense as

individuals. Indeed, as a general matter, states may not even be identified with the societies they presume to organize and govern; leaving aside the special (and historically unusual) cases of the representative democracies, an indigenous government may stand in as arbitrary a relationship to its own people as would a foreign conqueror. Because states and persons have such different properties, it would be a surprise if it turned out that the reasons for valuing *individual* autonomy also apply to the *state*; and if they do not, the domestic analogy will not do much philosophical work in justifying the morality of states (see State, the).

Most contemporary writers who rely on the domestic analogy to describe an international order in which states have a privileged ethical status appeal to considerations other than the analogy itself to defend the view. Two contrasting approaches deserve particular attention: first, the liberal idea that states' rights of political sovereignty and territorial integrity derive from the underlying individual rights and liberties of their members; second, the communitarian argument that respect for sovereignty enables political communities to preserve their distinctive histories and cultures, and thus to serve the human interests of their inhabitants.

For the first of these arguments, it is necessary to describe the extent and character of the individual rights and liberties which are supposed to serve as the basis of the derivative rights of states. Such a view emerges most naturally from voluntaristic conceptions of the social contract, in which the state is pictured as a free association based on either a historical contract or an ongoing implicit agreement (see Contractarianism). The state's authority to interfere in individual lives would, on such a view, be bounded by the terms of the underlying agreement; and the collective right against interference by outside forces in the internal life of the state would simply be the outward expression of the rights of its members not to be interfered with in their individual lives without their own consent. But the machinery of social contract theory is not really essential. Taking their lead from J.S. Mill (in 'A Few Words on Non-intervention', 1859), those who believe there is an important connection between individual liberty and the sovereignty of states need hold no more than that sovereignty makes individual liberty possible - if only because states serve as bulwarks against conquest by other states and protect a zone where free societies can flourish.

Liberal notions of the state's special status face two kinds of difficulties, which on examination reduce to one. Voluntaristic conceptions are open to the familiar criticism that very few actually existing states can realistically be described as voluntary associations; either the original contract lies too far in the past to exert any normative force in the present, or it is not possible to demonstrate that the prevailing terms of association are the object of a tacit or implicit agreement. Non-voluntaristic conceptions risk overgeneralizing by supposing that individual liberties are better protected through a system of more or less unconditional respect for the sovereignty of states than in a system in which the right of sovereignty is restricted.

The communitarian alternative links the moral status of the state with that of the underlying society. The view develops in two stages. First, it is argued that individuals are unlikely to flourish (or to realize their goods) except as members of communities characterized by at least a minimum level of mutual identification and fellow-feeling. This is only in part because individual human good may involve intrinsically social elements; it is also because the instrumental values of society, such as stability of expectations, can only be realized when there is an unforced disposition to cooperate, and this in turn can only be sustained when individuals identify themselves and each other as members of a common enterprise. Second, it is claimed that individuals are not likely to develop or sustain the requisite level of mutual identification if they are not governed through institutions which they can regard as the political expression of their own community. Again, this is only in part because the motivational force of communal loyalty may arise from a conception of the autonomous community as having intrinsic value; it is also because a community's capacity to defend itself from others and to elicit the necessary degree of contribution from its members will certainly depend on the people's acceptance of their governing institutions (see Community and communitarianism; Nation and nationalism).

The communitarian view assumes that there is some meaningful sense in which states can be seen as expressions of the values of their underlying societies, but this is controversial. Particularly in societies without representative institutions, there may simply be no way to determine whether the existing institutions are expressions of widely shared political values or, instead, legacies of force and fraud. There is a further problem: the communitarian argument trades on an image of domestic societies as reflecting a moral consensus in their communities, a condition that is surely, today, more the exception than the rule. As David Luban has argued (1980), both the

liberal and the communitarian accounts of the state's special status often seem to reflect a romantic idealization that obscures the diversity of state forms in the modern world and the great variation in their degrees of respect for individual rights and liberties.

A doctrine of international morality in which states occupy a privileged position might be defended without relying on any view about the morally special character of the state itself. In fact, some of the most influential views about international morality have this structure. They argue that a world order of states is more likely, on balance, to achieve values of general importance to human beings - a relatively high level of stability and conditions in which groups and societies will face the most favourable circumstances for their own prosperity. Such a generalization would not necessarily be undermined by examples of states whose governments are repressively authoritarian or which enforce exploitative domestic economic arrangements, as long as these examples could be shown to be exceptional. Moreover, views as highly consequentialist as this would have no difficulty, in principle, accommodating principles establishing minimum global standards of political legitimacy such as those found in the international doctrine of human rights.

4 Cosmopolitan morality

Cosmopolitanism stands in contrast to the notion that the boundaries between states, nations or societies have deep moral significance. It holds that each person is equally a subject of moral concern (or, that in the justification of choices of action or policy, the interests of each person affected should be taken equally into account), and that spatial proximity or shared membership are not in themselves sources of moral privilege. On a cosmopolitan view, there is no fundamental moral discontinuity between domestic and international society because states *qua* states have no special standing; if individuals have more extensive responsibilities to their own compatriots than to foreigners, or if states are entitled to be treated with some special respect by others in the international arena, this should be explained in a way that is consistent with the basic conception of a single moral realm in which each individual is equally worthy of concern and respect (see Impartiality).

There is a strong *prima facie* case for cosmopolitanism. For one thing, ordinary morality appears to presuppose its essential premise. Cosmopolitanism applies to the world the maxim that what we should do, or what institutions we should establish, should be decided on the basis of an impartial consideration of the claims of each person who would be affected by our choices. In one or another form, this maxim, although not beyond philosophical criticism, is a mainstay of the political moralities found in virtually all contemporary democratic cultures. Moreover, as the discussion in §3 suggests, the morality of states - cosmopolitanism's major rival - is difficult to maintain as a distinct view once the special status assigned to the state is brought into question; indeed, the most plausible explanation of the right of sovereignty may be an application of cosmopolitan principle rather than an alternative to it.

Against these considerations, there are at least three sources of doubt about cosmopolitan moralities. First, such views can seem either pallid or otherworldly in their apparent failure to connect with recognizable sources of motivation (for example, local affiliations or patriotic loyalties). Second, and relatedly, cosmopolitanism can appear to be too demanding, imposing requirements on individual conduct which may properly be described as heroic. Finally, these views can appear politically innocent, inviting an effort to build global political structures (a 'world government') that would do more harm than good.

The most common reply to these doubts is to argue they rest on an excessively simple interpretation of cosmopolitanism. In response to the first point, for example, it might be noted that cosmopolitan views are no more remote than most other moral doctrines from recognizable sources of motivation; although there is certainly a question about the capacity of cosmopolitan reasons for action to influence the will, this question is not different or more complex than the same question raised about the more familiar versions of Kantianism and utilitarianism. It is not possible to reply to the second objection without considering the normative consequences of cosmopolitanism; but it is worth noting that the view need not be interpreted as a doctrine of individual conduct, and indeed, that its most natural interpretation is instead as a view (or family of views) about institutional policy, in which case the objection is out of place. Finally, cosmopolitanism need not make any assumptions at all about the best political structure for international affairs; whether there should be an overarching, global political organization, and if so, how authority should be divided between the global organization and its subordinate political elements, is properly understood as a problem for normative political science rather than for political

philosophy itself. Indeed, cosmopolitanism is consistent with a conception of the world in which states constitute the principal forms of human social and political organization; the central question is whether this conception or some feasible alternative would be preferred when the matter is regarded from a perspective in which the interests of all are equally represented.

Although formally consistent with a state-based conception of world order, any plausible version of cosmopolitanism is likely to differ substantively from familiar versions of the morality of states. Consider, for example, the question of intervention. Because it assigns a high value to state sovereignty, the morality of states takes it as a settled matter that states, coalitions of states, and international agencies should refrain from interfering in the internal affairs of individual states, except in certain well-defined, exceptional cases, chief among which is intervention in self-defence. In the aftermath of Nuremberg, intervention to end shocking and egregious practices such as genocide has come to be accepted as well (although such an exception, because it presupposes an attenuated principle of sovereignty, fits uncomfortably with the morality of states). Cosmopolitanism, on the other hand, is likely to be less suspicious of intervention in principle, and to admit the possibility that intervention for protective and remedial purposes might be justified; on a cosmopolitan view, the most important questions about intervention are pragmatic, having to do with the chances that intervention would accomplish its legitimate purposes without doing unacceptable collateral harm.

Or consider the question of international distributive justice (see Justice, international). In its traditional form, the morality of states must regard a state’s involvement in the relief of material suffering elsewhere as a matter of charity or mutual aid. This is because the principle of sovereignty functions not only to protect a society against interference in its own political and social affairs, but also as a kind of collective property right: it secures a state against non-consensual deprivations of the resources and wealth to be found on (or under) its territory. Understood more abstractly, it conveys a conception of domestic societies as bearing the primary responsibility for their own development, and of social wealth as deriving more significantly from the collective efforts of previous generations in a single society than from that society’s good fortune in the distribution of natural and genetic resources and in the favourable prior course of political and social history. A cosmopolitan view, again in contrast, would begin with a conception of the world as a single human community. It would be unlikely to regard states as having privileged claims to the accumulation of social wealth unless institutions embodying such a recognition were the best feasible means to realize the requirements of whatever principles of distributive justice, themselves global in their scope, would be preferred from a point of view in which everyone’s interests were represented.

Between the morality of states and moral cosmopolitanism, a variety of intermediate positions might be identified. The most interesting of these are two-tiered views which combine acceptance of a minimum global standard of conduct with tolerance of substantial diversity above the minimum. Such views vary according to the content of the global minimum - whether, for example, it includes significant standards of domestic legitimacy (are states required to respect the human rights of their people?) or significant expectations of participation in schemes involving the international transfer of wealth (are states required to contribute to famine relief or international development efforts?) A good example of such a view is the revisionist doctrine of a ‘law of peoples’ recently set forth by John Rawls (1993).

5 Human rights

As an element of the discourse of international affairs, the doctrine of universal human rights is a legacy of the settlement of the Second World War. As a philosophical matter, however, the notion of a human right is ancient, its universalistic strand running to the Stoics and its deontological structure dating at least to the early modern notion of natural right.

The underlying idea is that all human beings, simply because they are human beings, are entitled to be treated according to certain minimum standards. Very extensive statements of the content of this doctrine can be found in the Universal Declaration of Human Rights, adopted by the United Nations in 1948, and in several human rights covenants adopted by the United Nations and various regional organizations. There is little dispute in contemporary public discourse about either the grounding or the content of human rights doctrine.

There is, however, very little agreement among philosophers about any of the major theoretical questions concerning human rights (see Rights §5). Philosophers disagree, for example, about the grounding of human

rights, particularly the sense in which these rights can be said to pertain to human beings 'as such'; their extent, particularly whether human rights include, in addition to the standard individual liberties, civil and political rights and rights to a minimum standard of living; about the priorities among human rights, a question which is especially vexed if their extent is taken to be relatively broad; and about the relationship between human rights and other moral values. The doctrine of human rights did not receive sustained theoretical attention until the late 1970s and 1980s; the most notable contributions are due to Henry Shue (1980) and R.J. Vincent (1986).

The doctrine of human rights is chiefly important in international philosophy as an embodiment of the notion that there are universal minimum standards of political legitimacy - that is, standards to which all societies are responsible. To put it slightly differently, the doctrine of human rights sets a limit to the extent of acceptable diversity among the political constitutions of states. As such, human rights doctrine might be seen from one perspective as the entering wedge of cosmopolitan thought, or, from another, as the price that statist conceptions must pay for their plausibility. Either way, human rights doctrine is the most familiar form in which two-tiered theories of international morality present themselves, and perhaps the most likely avenue for the future development of international philosophy.

See also: State, the

CHARLES R. BEITZ

References and further reading

Beitz, C.R. (1979) *Political Theory and International Relations*, Princeton, NJ: Princeton University Press.(A critical analysis of political realism and the morality of states, and an effort to set forth a cosmopolitan theory of international justice; extensive bibliography.)

Brierly, J.L. (1963) *The Law of Nations: An Introduction to the International Law of Peace*, ed. Sir H. Waldock, Oxford: Oxford University Press, 6th edn.(Historical introduction to international law.)

Brownlie, I. (ed.) (1993) *Basic Documents on Human Rights*, 3rd ed. Oxford: Oxford University Press. (Comprehensive compilation of authoritative international documents on human rights.)

Hobbes, T. (1651) *Leviathan. The English Works of Thomas Hobbes*, vol. 3, ed. Sir W. Molesworth, London: John Bohn, 1839.(Chapters 21 and 30 referred to in §2.)

Kant, I. (1795) *Perpetual Peace*, trans. H. Nisbet in *Political Writings*, 2nd edn ed. H. Reiss, Cambridge: Cambridge University Press, 1991.(Referred to in §1.)

Kennan, G. (1951) *American Diplomacy 1900-1950*, Chicago, IL: University of Chicago Press. (The Lindley Lectures referred to in §2 above; a defining document of American political realism.)

Luban, D. (1980) 'Just War and Human Rights', *Philosophy and Public Affairs* 9 (2): 160-81, and 'The Romance of the Nation-State', *Philosophy and Public Affairs* 9 (4): 392-97.(Cosmopolitan critique of the morality of states.)

Mill, J.S. (1859) 'A Few Words on Non-intervention', *Dissertations and Discussions: Political, Philosophical, and Historical*, London: Longmans, Green, Reader and Dyer, 1867, vol. 3, 153-78. (Historically important statement of the liberal theory of special status of the state; referred to in §3.)

Morgenthau, H. (1946) *Scientific Man versus Power Politics*, Chicago, IL: University of Chicago Press. (Influential defence of political realism.)

Nardin, T. (1983) *Law, Morality, and the Relations of States*, Princeton, NJ: Princeton University Press.(Argues that the aim of international law and morality is to regulate a pluralistic international society consisting of states with differing internal moralities; comprehensive bibliography.)

Nickel, J.W. (1987) *Making Sense of Human Rights*, Berkeley and Los Angeles, CA, and London: University of California Press.(Documentary appendix.)

Rawls, J. (1993) 'The Law of Peoples', in S. Shute and S. Hurley (eds) *On Human Rights*, New York: Basic Books, 41-82.(Example of a 'two-tiered' view; referred to in §3.)

Shue, H. (1980) *Basic Rights: Subsistence, Affluence, and U.S. Foreign Policy*, Princeton, NJ: Princeton University Press.(An argument for an expansive doctrine of universal human rights.)

Sidgwick, H. (1919) *The Elements of Politics*, London: Macmillan, 4th edn.('National' and 'cosmopolitan' ideals contrasted in connection with immigration policy; referred to in §3.)

Vattel, E. de (1758) *The Law of Nations or the Principles of Natural Law*, trans C.G. Fenwick, The Classics of International Law no. 4, vol. 3, Washington, DC: The Carnegie Institution, 1916. (Important source of domestic analogy; referred to in §3.)

Vincent, R.J. (1986) *Human Rights and International Relations*, Cambridge: Cambridge University Press, in

association with the Royal Institute of International Affairs.(A survey of theoretical and political issues in human rights doctrine; extensive references.)

Walzer, M. (1977) *Just and Unjust Wars*, New York: Basic Books.(Most influential contemporary work on morality in war, with extensive references; referred to in §3.)

Wolff, C. (1749) *The Law of Nations Treated According to the Scientific Method*, trans. J.H. Drake, The Classics of International Law no. 13, vol. 2, Washington, DC: The Carnegie Institution, 1934. (Important source of domestic analogy referred to in §3.)

Interpretation, Indian theories of

Need for interpretation of texts was felt already during the ancient period of Vedic texts in India. Vedic texts were orally transmitted for over a thousand years. During this period, the change in locations of people reciting the texts and the mother tongues of the reciters led to a widening gap between the preserved sacred texts and their interpreters. Additionally, there was a notion that the sacred language was a mystery which was only partially understood by the common people. This led to the early development of exegetical tools to assist the interpretation of the sacred literature.

Later, grammarians and etymologists developed sophisticated exegetical tools and theories of interpretation. These generally led to a deeper understanding of the structure of language. The priestly tradition developed its own canons of interpretation, which are manifest in the system of Mīmāṃsā. Here we have the first fully developed theory of discourse and context.

The categories developed by Mīmāṃsāwere used by other schools, especially by the school of Dharmaśāstra, or Hindu religious law. Both Mīmāṃsāand Dharmaśāstra created sets of hierarchical principles for authoritative guidance in interpretation. Other philosophical and religious traditions developed categories of their own to deal with problems of interpretation. A major problem was created when the literature accepted as authoritative by a tradition contained apparently contradictory passages. The traditions had to deal with this problem and find ways of explaining away those passages which did not quite fit with their own view of truth. For this purpose, a whole set of categories were employed. At a later period, several ingenious principles of interpretation were used for texts in general. Here, significant contributions were made by the traditions of Sanskrit grammar and poetics.

1 Text and interpreter

Concern for the interpretation of the scriptural language was already being voiced in India from the earliest period of the Vedic scriptures (*c.*1500 BC). The primary reason for such concern lies in the real or perceived distance between the language of a sacred text and the contemporary language of the person trying to comprehend and use that text. Real distance between the language of a sacred text and the language of its interpreter or user can be caused by various factors, such as temporal and geographical distance, linguistic change in the sacred language, shifts in the mother tongues of the text's preservers, and corruption and loss in the text in the course of oral or written transmission. The middle Vedic texts already express a concern that the users of language are unable to pronounce proper accents. The loss of this ability contributed to improper use of sacred formulas, as well as to misinterpretation of their meaning. We are told a story about a demon who seeks a son who will kill the god Indra. He should ask for a son who will be *indraśatrú*, 'killer of Indra'. However, the ignorant demon uses a wrong accent and asks for a son who will be *índraśatru*, 'he whose killer is Indra'. This story is later repeated by Sanskrit grammarians and etymologists to illustrate the importance of a proper understanding of accents. Similarly, in the oral tradition, long stretches were memorized and recited in an unbroken fashion. Besides actual cases of mutilation and alteration of the text, such recitation resulted in multiple possibilities for dividing up the text.

Ideas regarding the nature of the sacred language contributed an additional feeling of distance between text and interpreter. The *Ṛg Veda* tells us that language, in reality, is measured in four quarters, all of which are known to the wise, insightful priests. However, three quarters of speech are hidden away from common people, who have access only to a single quarter. The goddess of speech is said to reveal her true form only to a person of her own choice. Another important notion is that the gods do not like the direct mode of expression (*pratyakṣa*). They would rather have sacred expressions in an indirect mode (*parokṣa*). Middle Vedic prose texts often explain the so-called indirect expressions by paraphrasing them with expressions which have structures more transparent to their audience.

2 Early tools to assist interpretation

In the middle (*c.*1000 BC) and late (*c.*700-500 BC) Vedic periods, we find a growing concern for proper understanding of older texts and the beginning of concerted efforts to develop technical tools towards that goal. For the Brāhmaṇa texts, the highest perfection (*rūpasamṛddhi*) of ritual results when the recited Vedic text matches the ritual action. This places a high premium on comprehending the meaning of the recited texts before

one decides how and when they are to be used in ritual. The texts repeatedly state that the man who performs a ritual ‘knowingly’ (*ya evaṃ vidvān*) will be rewarded.

There was also a serious effort by priest-scholars such as Māṇḍūkeya and Śākalya to put together the Padapāṭhas, ‘word-by-word’ versions of Vedic texts. The late Vedic text *Aitareya-Āraṇyaka* speaks of *nirbhuja* (‘unsegmented’) versions versus *pratṛṇṇa* (‘segmented’, ‘broken-down’) versions. Later on, the segmented versions were used as the basis for reproducing continuous versions (*saṃhitā*) which were deemed to match the originals. These activities provided an enormous opportunity to build formal tools to assist interpretation, a task taken up later by the grammarians, etymologists and phoneticians of ancient India.

3 Etymologists and grammarians

The goal of the etymologists was to provide insight into the meaning of an expression (*nirvacana*) by providing an analytical breakdown of it. Yāska, the author of *Nirukta (Analytical Exegesis*, *c*.500 BC), the only surviving ancient text on etymology, says that comprehension of meaning is the highest goal of his science. Without comprehension of meaning, one cannot break down a continuous Vedic text into its constituent words, nor can one know the proper ritual application for such a text. He criticizes the reciters of Vedic texts who do not care for meaning. He refers to Kautsa, who argued that the Vedas are devoid of meaning. Yāska refutes Kautsa’s arguments and proposes a number of methods for etymology:

> Now we shall deal with etymology. In this context, words whose accent and grammatical form are regular and have a transparent modification of the formative elements should be explained in the regular manner. However, [for other words,] where the meaning is unclear and the modifications of the formative elements are not transparent, one should always analyse them by focusing on their meaning, by the analogy of some shared action [with another word]. If no such analogy is found, one should explain [words] even by the commonality of a single syllable or sound; but one should never give up an attempt at etymology.
>
> (*Nirukta* 2.1)

While the etymologists were busy offering a breakdown of difficult Vedic expressions, they were clearly aware of a different tradition, namely the tradition of grammar (*vyākaraṇa*). In Yāska’s terminology, this reverse process, called *saṃskāra*, consists of building up words by fusing their formative elements and putting them through a series of modifications (*vikāra*). This is the pattern followed in the famous ancient grammar of Pāṇini. It extensively uses the term *artha*, ‘meaning’, and yet its goal is not interpretation *per se*. It is an encoding grammar rather than a decoding grammar. It begins with the meaning to be communicated and with morphological primes such as roots, stems and affixes. Through a series of steps involving affixation, substitution, modification, and so on, it builds up a surface expression in Sanskrit that corresponds to the intended meaning.

However, Pāṇini’s successors, Kātyāyana and Patañjali, clearly see the value of his grammar as a tool for the preservation and comprehension of the Vedic scriptures. They argue that the usage of Sanskrit backed by an understanding of grammar leads to success in life here and hereafter. The grammarians are dealing with a stage of Sanskrit when it is no longer a first language, and needs to be acquired by a deliberate study of grammar. Among the purposes of the study of Sanskrit grammar, Patañjali includes several which relate to one’s ability to interpret the scriptural texts properly. One of these purposes is *asandeha*, ‘removal of doubts’. Besides mentioning the old *indraśatrú* example, Patañjali cites one more case, *sthūlapṛṣatī*. This word describes a cow that is supposed to be sacrificed. If it is accented as *sthūlapṛṣatī́*, then it can be interpreted as a Karmadhāraya compound to mean ‘a cow that is fat and spotted’. However, if it is accented as *sthūlápṛṣatī*, then it becomes a Bahuvrīhi compound and means ‘a cow that has fat spots’. A person who does not know the accents or what the difference in accent leads to will not be able to interpret the ritual prescription properly. For Patañjali, studying the Vedas without comprehending their meaning is as futile as throwing firewood away from fire. How can it possibly catch fire?

4 Mīmāṃsāprinciples of interpretation

The tradition of Mīmāṃsāsystematizes the rules of interpretation as they apply to the ritual use of Vedic texts, as well as to the performance of ritual as understood from those texts. Here the term ‘Veda’ applies to both the mantras and the prose Brāhmaṇa commentaries. For Mīmāṃsā, the Vedas are entirely uncreated, not authored either by humans or by God, yet are fully meaningful and authoritative (see MīmāṃsāČsect;3). Given this eternal

innate authority of the Vedic texts, Mīmāṃsāprovides a set of six hierarchical principles of interpretation, the earlier ones overriding the later ones: (1) direct statement (*śruti*); (2) word-meaning (*liōga*); (3) syntactical connection (*vākya*); (4) context (*prakaraṇa*); (5) position in the text (*sthāna*); (6) name (*samākhyā*).

Let us consider the conflict between the first two guiding principles. Verse 8.51.7 of the *Ṛg Veda* describes the divinity Indra. On the basis of the word-meaning (*liōga*), one would think that this verse should be used while making an offering to Indra. However, a direct statement (*śruti*) in the *Taittirīya-Saṃhitā* (1.5.8), another Vedic text, says that one makes an offering to Gārhapatya fire by using a verse addressed to Indra. This direct statement overrides what one would otherwise decide on the basis of the meaning of the particular verse.

Besides these six principles, Mīmāṃsāalso deals with numerous important issues relating to Vedic interpretation, such as cases where one finds seemingly contradictory statements in the Vedic corpus. How, for example, does one deal with a statement authorizing a Niṣāda chief to perform a certain sacrifice? The tribal people referred to by the term ‘Niṣāda’ are generally beyond the pale of those eligible to perform sacrifices to Vedic gods. If the word ‘Niṣāda-chief’ is taken to mean ‘one who is a Niṣāda and a chief’, this authorization conflicts with the general prohibition of low-caste people taking part in Vedic sacrifices. However, if it is taken to mean ‘a chief of the Niṣāda’, it could possibly apply to a person who is not a Niṣāda, yet is their chief. The Mīmāṃsakas try to resolve such issues. Where there is ultimately no way to choose between contradictory Vedic injunctions, Mīmāṃsāadvises an optional choice, for example, between ‘One uses the Ṣoḍaśīcup in the Atirātra sacrifice’ and ‘One does not use the Ṣoḍaśīcup in the Atirātra sacrifice’. Considering the entire Vedic corpus to be fully authoritative leads to rejection of its historical and geographical diversity. This rejection gives rise to many more instances of apparently contradictory statements.

5 Other philosophical traditions

Acceptance of the entire Vedic corpus as authoritative requires the reconciliation of a large number of apparently contradictory statements. However, the problem is not unique to Mīmāṃsā. Bādarāyaṇa, in his *Brahmasūtra*, uses the term *samanvaya* (‘coordination’, ‘reconciliation’) for the technique of reconciliation. While Mīmāṃsādeals largely with ritual injunctions, Bādarāyaṇa is concerned with the apparently divergent views regarding the ultimate reality of Brahman in the Upaniṣads. The commentators on the *Brahmasūtra* are faced with the same dilemmas, which get more complicated with every generation. Beginning with Śaōkara, the tradition of Vedānta adopts the view that it is founded on three authoritative sources (*prasthānatrayī*), namely the Upaniṣads, the *Brahmasūtra* and the *Bhagavad Gītā*. The principle of *samanvaya* needs to be expanded to all these texts. With Rāmānuja and Madhva, the Purāṇas and the texts of the Pāñcarātra tradition are also added to this pool of authoritative texts. Commentators often create a deliberate hierarchy among different texts, or among different passages from the same text. Then they conveniently argue that some of these texts are to be taken literally (*mukhyārtha*), while others are to be interpreted metaphorically (*lakṣaṇā*). Another dichotomy among texts is stipulated by suggesting that different texts or passages are meant for different audiences, and that these audiences, having different intellectual and spiritual abilities, need suitably different teachings. Some teachings are meant for those who are truly capable, while others are meant to be provisional teachings for those who have not yet arrived at the same high level. Such distinctions are found also in the Jaina and Buddhist traditions (for example, the *upāya*/*upeya*, *vyāvahārika*/*pāramārthika* and *neyārtha*/*nītārtha* distinctions, all of which correspond respectively to ‘provisional’ and ‘true’). With such tools, one can extract a uniformity of teaching from texts which apparently do not have it. The tradition of Dharmaśāstra, religious law, attempted to eliminate similar apparent conflicts between various texts by saying that some were restricted to certain periods of time, regions or social classes.

6 Principles of interpretation in Hindu law

Dharmaśāstra provides an important set of hierarchical guidelines for interpration. Authorities include the Vedas, the Smṛtis (law books), the behaviour of the elites in the community, and, finally, one’s own judgment. According to this tradition, and most of the wider Indian tradition, the authority of the Vedic texts overrides that of the Smṛtis. While the Vedic texts are either uncreated or created by God, the Smṛtis are said to be authored by human scholars and are believed to be based on the Vedic texts. The behaviour of the elite leaders of the community is presumed to be based on the Vedas and the Smṛtis. When there is no access to any other authority, one has recourse to one’s own conscience (see Duty and virtue, Indian conceptions of §2).

In fact, some early Dharmaśāstra texts say that there were no prescriptions in the Vedas (*śrutyabhāvāt*) regarding the laws that govern different regional, caste and family traditions, and, for that reason, the ancient lawgiver Manu explained these topics. However, along with the tradition of Mīmāṃsā, Dharmaśāstra argues that this simply means that the Vedic texts dealing with such topics are lost, and not that they did not exist. The existence of lost Vedic texts (*anumitaśruti*) is inferred on the basis of statements in the Smṛtis. In general, while the Dharmaśāstra theory lays great emphasis on the Vedas as the ultimate basis for authority, in reality the interpretation of the Vedas gets pulled in the direction of contemporary beliefs and practices.

7 Generalized principles of textual interpretation

Coming back to the wider context of textual interpretation, we find a number of important initiatives. In the tradition of Sanskrit grammar and in the traditions relating to Brahmanical ritual and law, there is a large literature called Paribhāṣā, 'Maxims of Interpretation' (see Abhyankar 1968). In the context of rules of Sanskrit grammar, these maxims help account for the derivation of certain forms, by effectively extending or narrowing the scope of a rule through interpretation. Generally there is great reluctance to alter the wording of a rule or offer a new rule. The tradition would rather get a new ruling by reinterpreting an old one. In doing this, the commentators have recourse to a great many interpretive techniques (see Kielhorn 1887). Related to the Paribhāṣāliterature is the device called *nyāya*. In this context, the term refers to a proverbial statement of colloquial (*laukika*) or technical (*śāstrīya*) wisdom. For example, when two events have no genuine relationship, but are merely coincidental, one cites the *kākatālīyanyāya*, which refers to a story of coincidences, such as the following: a person is resting under a palm tree; a crow flying over the tree comes and sits on a palm fruit; the weight of the crow causes the fruit to fall and crack the person's head. There are a large number of these traditional *nyāyas* (see Kane 1977).

The fifth-century grammarian Bhartṛhari, in his *Vākyapadīya (On Sentences and Words)*, provides a catalogue of many guiding principles of interpretation:

> Meanings of words are differentiated on the basis of the sentence [in which they occur], the context, the meaning [of other words in the context], propriety, time and place [of utterance], and not merely on the basis of the form of a word. Connection, separation, association, opposition, meaning, context, indication, presence of another word, suitability, propriety, place, time, gender, and accent, etc. are the factors which help to determine a specific meaning, in the absence of [natural] clarity of meaning.
>
> (*Vākyapadīya* 2.314-16)

Many of these ideas are carried over into the tradition of Sanskrit poetics, which aims at explaining how one understands the aesthetically pleasing significance of poetry. The notion of interpretation is crucial in all discussions in this tradition. While the lexical meaning of words may be taken for granted, there is a wide gap between lexical meaning and intended meaning. In discussing the phrase 'the cowherd colony is on the river Ganges', we are told that we move away from the literal meaning because of the difficulty of construing the literal meaning (*anvayānupapatti*) and the difficulty of justifying the literal interpretation in view of the intention of the speaker (*tātparyānupapatti*). Nudged by these two factors, we move to the meaning 'the cowherd colony is on the bank of the river Ganges'. Beyond this comes the level of implied or suggested meaning, which may or may not match the meaning intended by the speaker. A listener often understands far more than what the speaker intends. Such levels of suggested and implied meanings are further accounted for through the suggestive function of language and/or inference.

See also: Language, Indian theories of; Meaning, Indian theories of

MADHAV M. DESHPANDE

References and further reading

Abhyankar, K.V. (ed.) (1968) *Paribhāṣāsaṃgraha: a Collection of Paribhāṣā (Texts from Various Systems of Sanskrit Grammar*, Poona: Bhandarkar Oriental Research Institute.(A comprehensive collection of interpretive rules in different schools of Sanskrit grammar.)

Bhartṛhari (5th century) *Vākyapadīya (On Sentences and Words)*, ed. W. Rau, *Abhandlungen für die Kunde des Morgenlands* 42 (4), 1977.(A critical edition of the verse portion of the *Vākyapadīya*; there is no English translation of this, but see Iyer 1969.)

Deshpande, M.M. (1992) *The Meaning of Nouns: Semantic Theory in Classical and Medieval India*, Dordrecht:

Kluwer Academic Publishers.(The bulk of the book is an annotated translation of a seventeenth-century Sanskrit text on the meaning of nouns. The introduction covers the history of a number of semantic theories in Sanskrit grammar and philosophy.)

Gächter, O. (1983) *Hermeneutics and Language in Pūrvamīmāṃsā*, Delhi: Motilal Banarsidass.(An important modern analysis of Mīmāṃsā (theories of language and interpretation.)

Iyer, K.A.S. (1969) *Bhartṛhari: A Study of the Vākyapadīya in the Light of the Ancient Commentaries*, Poona: Deccan College.(A very comprehensive study, though somewhat dated.)

Kane, P.V. (1962) 'On Mīmāṃsā (Principles and Rules of Interpretation in Relation to Dharmaśāstra', in *History of Dharmaśāstra*, Poona: Bhandarkar Oriental Research Institute, 2nd edn, 1977, vol. 5, part 2: 1283-1351.(A seminal study relating the interpretive rules of Mīmāṃsā (to Hindu law.)

Kielhorn, F. (1887) 'Notes on the Mahābhāṣya, No. 7, Some Devices of Indian Grammarians', *The Indian Antiquary* 16: 244-52.(A famous study of the principles of interpretation as used by Sanskrit grammarians.)

Raja, K. (1963) *Indian Theories of Meaning*, Adyar Library Series 91, Madras: Adyar Library and Research Centre.(A somewhat dated, but still very useful, account.)

Yāska (*c.*500 BC) *Nirukta (Analytical Exegesis)*, ed. and trans. L. Sarup, *The Nighaṇṭu and the Nirukta: the Oldest Indian Treatise on Etymology, Philology and Semantics*, London: Oxford University Press, 1920.(Critical edition and translation of an important work attributed to Yāska.)

Introspection, epistemology of

If we wish to know what is going on in someone else's mind, we must observe their behaviour; on the basis of what we observe, we may sometimes reasonably draw a conclusion about the person's mental state. Thus, for example, on seeing someone smile, we infer that they are happy; on seeing someone scowl, we infer that they are upset. But this is not, at least typically, the way in which we come to know our own mental states. We do not need to examine our own behaviour in order to know how we feel, what we believe, what we want and so on. Our understanding of these things is more direct than our understanding of the mental states of others, it seems. The term used to describe this special mode of access which we seem to have to our own mental states is 'introspection'.

A view which takes its inspiration from Descartes holds that introspection provides us with infallible and complete access to our states of mind. On this view, introspection provides us with a foundation for our knowledge of the physical world. On this view we come to know the physical world by first coming to recognize certain features of our mind, namely, the sensations which physical objects excite in us, and then drawing conclusions about the likely source of these mental states. Our knowledge of the physical world is thus indirect; it is grounded in the direct knowledge we have of our own minds. The view that introspection provides an infallible and complete picture of the mind, however, is no longer widely accepted.

Introspection has also been called upon to support various metaphysical conclusions. Descartes argued for dualism on the basis of introspective evidence, and certain contemporary philosophers have argued in much the same spirit. Hume noted that introspection does not reveal the presence of an enduring self, but only a series of fleeting perceptions; some have concluded, therefore, that there is no enduring self.

Philosophers concerned with self-improvement, whether epistemological or moral, have frequently called upon introspection. Introspection has been thought to aid in forming beliefs on the basis of adequate evidence, and it has been used as a tool of self-scrutiny by those concerned to understand and refine their motivations and characters.

1 The complete transparency view

Precisely what Descartes believed about introspection is a subject of some controversy (see Descartes, R.). Nevertheless, we may describe a view for which there is substantial textual support and which has historically played an important role. The powers of introspection, on this view, are considerable. First, introspection is held to be infallible. Thus, for example, if I believe that I am experiencing a mild headache, then I am experiencing a mild headache. Second, this infallibility is not merely a matter of psychological law, something which could have been otherwise; instead, it is a matter of necessity. The very idea of introspective error is incoherent. Finally, introspection provides one with a complete picture of the contents of one's mind. There can be nothing in the mind which may escape its notice. The picture one gets of one's mind when one introspects is thus both complete and completely accurate.

In all of these respects, introspection contrasts quite sharply with our ability to know the world outside our minds. We can and, of course, do make errors in our judgments about the world around us. We may sometimes think that a certain book is on the shelf, but on trying to take it down discover that what seemed to be a book is merely an empty dust-jacket. When it comes to physical objects, appearances sometimes deceive. Our mental states, however, cannot fail to be precisely as they appear to us. By the same token, our picture of the world around us is fundamentally incomplete; we simply do not know everything there is to know about it. Our mental life, on the other hand, is wholly transparent to introspection. It would be incoherent, on Descartes' view, to suppose that one might be in a certain mental state without its being accessible to introspective attention.

Introspection is not merely superior to perception in Descartes' view, however; instead, introspection is a prerequisite for perceptual knowledge. We perceive the external world only indirectly. By way of introspection, we gain direct access to our mental states. Having made note of our mental states, we may then attempt to infer what external objects, if any, might have made such impressions upon us. The self-knowledge achieved through introspection thus forms an infallible foundation upon which we may attempt to build our understanding of the external world. The evidence we have for the existence and character of the external world is fully provided for us

by way of introspection. Introspective knowledge is, on this account, epistemically prior to all knowledge of the external world.

This view of introspection gives rise to a number of epistemological problems which have been pursued with great vigour since the time of Descartes. First, and most obviously, there is the problem of explaining how it is that the knowledge we have of our own mental states can, by itself, provide sufficient evidence for beliefs about an external world of physical objects (see Scepticism). Second, there is the problem of other minds, that is, the problem of explaining how the knowledge we have of the behaviour of others can provide sufficient evidence for beliefs about their mental states (see Other minds). Third, there is a problem about our knowledge of the past and the future. Since introspection provides us with knowledge of our current mental states alone, we must somehow find sufficient evidence in our knowledge of the present instant to ground claims about other times. Although Descartes raised and addressed only the first of these three problems, they are really variations on a common theme. In each case, there is thought to be only indirect access to certain potential objects of knowledge: in the first, physical objects; in the second, mental states of other people; in the third, the past and the future. One needs to develop hypotheses about these things, but the truth or falsity of one's hypotheses cannot be directly checked against the facts, since one has no direct access to them (see Inference to the best explanation). Some have argued that these problems are resolvable, that indirect evidence of the relevant facts is evidence enough. Others have argued that the problems are insurmountable, and that a very broad scepticism is thus forced upon us; we are left with knowledge of nothing but our current mental states. Still others have suggested that because the Cartesian approach inevitably leads to such radical scepticism, we should reject the premises which raised these difficulties in the first place.

2 Transparency and the contemporary scientific view of the mind

There are, however, problems for the transparency view in any case. It is quite clear that introspection does not give us total access to our current mental states. Moreover, once we recognize that the picture introspection presents is inevitably incomplete, it is but a small step to recognizing the fallibility of introspection as well.

The strengths and limitations of Descartes' view may best be understood by first examining a view which attributes even greater powers to introspection than did Descartes. Let us thus consider the view that mental phenomena are not merely accessible to introspection, but instead that they must actually be present to consciousness if they are to exist. On this very strong view, the idea that one might have memories (in the mind) which are not current objects of introspective awareness is a mistake. But how can one then account for the ability to recall past events? The simple phenomenon of memory retrieval becomes mysterious if one cannot allow for mental occurrences which are not conscious.

Far more plausible is Descartes' suggestion that mental phenomena are essentially introspectable, that is, that they are potential objects of introspection, rather than that they must actually be introspected in order to exist. On this view one may countenance the existence of memories and other mental states generally, which are not current objects of introspective awareness, so long as they may be brought to consciousness. It is clear enough that some memories may only be brought to consciousness with great difficulty. The mere inability to introspect some mental item on a particular occasion, even after great effort, is thus insufficient to show that it does not exist. Mental phenomena might easily escape one's introspective notice on such a view; they need only be accessible to introspection in principle.

Even this view, however, is almost certainly false, although the reasons for thinking so were not available to Descartes. Work in experimental psychology has for some time now required the postulation of mental states which are not accessible to introspection. Two brief examples should suffice to make the point. First, in order to explain our responsiveness to features of the environment of which we are not consciously aware, it is necessary to postulate mental mechanisms which register information about the environment and integrate this information with other facts to which we are sensitive. No amount of introspective attention can acquaint us with the operation of these mental processes or the contents of the mental states which register the information on which they draw. Second, as Noam Chomsky's work has shown, the process of language acquisition involves the generation and testing of detailed hypotheses about the structure of the language spoken around one; language use requires that the rules of one's language be mentally represented (see Chomsky, N). These hypotheses and rules, however, while explicitly represented in the mind, are not accessible to introspection. No amount of introspective scrutiny

can acquaint us with any of these features of the mind.

The picture of the mind which is thus emerging from work in experimental psychology and psycholinguistics is quite different from that portrayed by Descartes. On Descartes' view, theorizing about the mind is unnecessary because the data provided by introspection already encompass all that there is to the mind. In this respect, psychology contrasts with the other sciences where theories need to be constructed to explain the data collected, and entities and processes need to be postulated to account for the regularities in observed phenomena. Current work in psychology, however, presents the mind as an object of study as fit for theorizing as those in any other science and, indeed, as requiring theoretical investigation if it is to be properly understood. Indeed, once we recognize the existence of memories which are not current objects of awareness, we have taken the first step in theorizing about the mind. On current accounts, introspection gives us access to only a very small corner of the mental world - comparable to the access unaided vision provides us to chemical phenomena. By far the greatest part of our mental life is simply inaccessible to introspection.

One might think that although current work in psychology undermines Descartes' view that introspection gives us access to a complete account of the mind, it leaves his claim that introspection provides an infallible guide to the mind, at least as far is it goes, untouched. This is not, however, the case. Once it is acknowledged that the mind is a fit object of scientific theorizing, it becomes impossible to insulate the judgments of introspection from the encroachment of theory. Introspection provides people with a certain perspective on their mental states. It is but one perspective, however. Our best available theories of the mind draw on other sources of information as well, and may in principle provide us with reasons for thinking that the perspective provided by introspection is not always an accurate one, just as our best available theories of vision may allow us to recognize situations in which visual appearances are likely to mislead. In the case of introspection, as with vision, one need not appeal to subtle work in experimental psychology to see its fallibility at work. The fallibility of introspection is not merely a theoretical possibility, but a fact of everyday life.

Works of literature have long portrayed characters whose introspective understanding of their own mental states fall far short of perfection, not merely by being incomplete, but by being inaccurate as well. Failure to recognize accurately one's own emotions, desires, beliefs, motivations and traits of character are all familiar. Indeed, it is simply a part of the common-sense conception of the mental that introspection is subject to error in making these kinds of judgments. The current experimental literature in psychology both supports and deepens this common-sense view. Introspective judgments of current mental states simply are not infallible. This is not to say, of course, that introspection cannot provide us with knowledge of our own mental states, for, at least on most views, knowledge does not require infallibility.

The limitations of introspection were recognized by Kant in the *Critique of Pure Reason* (1781/1787): introspection, according to Kant, 'represents to consciousness even our own selves only as we appear to ourselves, not as we are in ourselves' (B153).

3 Cartesian projects in contemporary perspective

Even if we acknowledge that introspection provides neither a complete nor a completely accurate account of the mind, it is not obvious what implications this has for the Cartesian doctrine of the epistemic priority of self-knowledge - that is, the view that knowledge of one's own mental states is in an important sense prior to knowledge of the physical world. Many have suggested that this last Cartesian doctrine survives the downfall of the others, and that the projects to which it gives rise must still be carried out, although perhaps with some minor modifications. Others have argued that the doctrine of the epistemic priority of self-knowledge must fall with the other Cartesian doctrines, and that the projects to which it gives rise are thereby completely undermined.

One project which is clearly undermined is the constructive project of radical foundationalism. The radical foundationalists, following Descartes, sought to provide empirical knowledge with an infallible foundation. On this view, a belief is shown to be justified either by showing it to be infallible or by appropriately basing it on a foundation of infallible beliefs. Because the foundation consists of beliefs about which mistakes are impossible, it can need neither revision nor correction in light of new information. Introspection was thought to provide such a foundation of judgments about the current contents of one's mind. The fallibility of introspection, however, clearly vitiates this project by robbing it of the required foundation. This need not, of course, be taken to show that radical

foundationalism provides a mistaken account of the requirements for justified belief. Rather, it shows that the requirements which radical foundationalists set out cannot be met. Thus, either radical foundationalism does not correctly lay out the requirements for justified belief, or scepticism about the external world is inevitable.

The constructive projects associated with weaker versions of foundationalism, however, are not so obviously undermined. Introspection might be thought to provide one with a fallible foundation for one's beliefs about the physical world. Such a foundation would be subject to revision, but one would still need to show, as with radical foundationalism, how to ground judgments about the physical world in the deliverances of introspection. Many of the justificatory projects faced by radical foundationalism will have analogues within this more modest foundationalist enterprise; the problems of the external world, of other minds, and of the past and the future, for example, will present the modest foundationalist with justificatory projects (see Foundationalism).

Coherence theorists - who hold, roughly, that a belief is justified if it fits in with one's other beliefs - may also accept the doctrine of the priority of self-knowledge. Indeed, on one natural way of presenting a coherence theory, agents who wish to justify their beliefs must begin by introspecting in order to discover what beliefs they have; only then can the project of justifying one's beliefs begin. In holding such an account of justification, one need not hold that introspection gives one complete or infallible access to one's beliefs. Introspection is used here, as in modest foundationalism, simply to provide a starting point for the task of justification. Having provided such a starting point, the deliverances of introspection may subsequently be revised in light of considerations of overall coherence (see Knowledge and justification, coherence theory of).

These versions of foundationalism and the coherence theory share a commitment to the doctrine of the epistemic priority of self-knowledge. On these versions of both views, knowledge of the physical world is possible only when appropriately founded in knowledge of (certain of) one's own mental states. Some have argued, however, that both foundationalism and the coherence theory thus understood are misguided for precisely this reason, for it may be thought that the doctrine of the epistemic priority of self-knowledge is also undermined once introspection is seen as a source of only an incomplete and partially accurate view of the mind. On such a view, the special epistemic status of self-knowledge depends upon the infallibility of introspection. If introspection is a source of knowledge which is both incomplete and fallible, then it fails to distinguish itself from perception of the external world as a privileged source of belief. Introspection may be, on the whole, a source of accurate information about the mind, but on this view it can occupy no special place in a theory of justification once the claim of infallibility is rejected. There is no more reason to begin the project of justifying one's beliefs with the deliverances of introspection than with the deliverances of sense perception. The reflective believer, on this view, does not begin reflection by focusing exclusively on the contents of their own mind, and then determining what can be justified in light of such contents; instead, perceptual judgments are taken at face value from the beginning, just as introspective judgments are, and the project of justification proceeds against the background of one's total body of belief. Such a view makes a complete break with Cartesian doctrine by assigning no special status whatsoever to introspection.

4 Introspection and the metaphysics of mind

Thus far, we have examined the significance of introspection for epistemology, but introspection has also been pressed into service to support important metaphysical doctrines. Descartes argues in the Sixth Meditation that we recognize our existence as thinking beings by way of introspection, and this provides the key premise in an argument for dualism (see Dualism). Once again, there is substantial scholarly debate on the precise nature of the argument, but this much is clear: Descartes points out that he can conceive that the world could be such that he should have no body and yet, at the same time, his introspective understanding should guarantee that he exists as a thinking thing. Thinking alone, he argues, is essential to him. From these premises, he concludes that the mind may exist without the body; it is an immaterial substance.

This argument was attacked from its first presentation. Descartes' critics contend that the mere fact that one can conceive of the mind existing without the body does not show that, in fact, it can exist without the body. I can imagine, for example, that my computer should continue to work even if its plug should be removed from the outlet, but this of course does not show that it is a genuine possibility that it should continue to work in such circumstances. My ability to conceive of my mind existing without my body may similarly reflect no genuine possibility. Although the argument is clearly defective when presented in just this way, it continues to remain a

source of interest in discussion of the mind and its nature, and many still argue that some version of this argument, or considerations in some way similar to those Descartes brought forth, may be used to defend dualism. Thus, a number of authors have recently argued that a complete physical description of the universe inevitably leaves something out which introspection reveals, namely the qualitative nature of mental states. There is something it is like to be a bat, for example, but any physical description of the bat's brain, central nervous system, sensory equipment and so on, inevitably leaves this out. This shows, it is argued, that a physical account of the bat is thus incomplete. Some have taken this to show that a proper account of the mind must inevitably appeal to something nonphysical. Once again, introspective evidence is called upon to support dualism. More than this, the argument assumes, with Descartes, that introspection provides us with some deep insight into the essential properties of our mental states, properties which are fundamentally incompatible with those of the physical realm.

5 Introspection and self-improvement

Just as introspection has long been thought to provide a special route and a special insight into the self, it has also been thought to provide an indispensable tool in projects of self-improvement, both epistemological and moral. On the epistemological side, the special connection thought by Cartesians to hold between introspective access to one's mental states and the justification of one's beliefs dictates that those who wish to improve their epistemic situation must allow introspection to play a central role in that project. By the same token, many have thought that the first step on the road to moral self-improvement involves introspective self-scrutiny, examining one's motivations and character traits. Indeed, it has frequently been held that introspective self-monitoring is a crucial constituent of a moral life. One might reasonably wonder, however, given the possibility of self-deception, whether introspection is really the best tool for carrying out such a project of moral self-evaluation. It has even been argued, on empirical grounds, that introspection is not well suited to the task of epistemic self-evaluation because introspection may be especially insensitive in locating and identifying beliefs held on inadequate grounds. Even the most radical anti-Cartesians, however, grant that introspection must play some role in these projects. While it can no longer be claimed that introspection presents the final word on the character of one's mind, there can be no doubt that it plays an important role in the project of self-understanding.

See also: Consciousness; Introspection, psychology of; Naturalized epistemology; Unconscious mental states

HILARY KORNBLITH

References and further reading

Alston, W.P. (1971) 'Varieties of privileged access', *American Philosophical Quarterly* 8 (3): 223-41.(Important discussion of the wide spectrum of views available on the kind of special access which introspection might provide.)

Armstrong, D.M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.(Defends a perceptual model of introspection, arguing that introspective judgment is fallible.)

Chomsky, N. (1980) *Rules and Representations*, New York: Columbia University Press.(Gives an account of the reasons stemming from linguistic theory, discussed in §2 above, for thinking that a large portion of the mind is inaccessible to introspection.)

Churchland, P.M. (1985) 'Reduction, qualia and the direct introspection of brain states', *Journal of Philosophy* 82 (1): 8-28.(Responds to the kind of argument for dualism, discussed above in §4, which is based on introspection.)

Descartes, R. (1641) *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 3 vols, 1985.(The tranparency view may be found in Meditation Two, the argument for dualism is in Meditation Six.)

Gopnik, A. (1993) 'How we know our minds: The illusion of first-person knowledge of intentionality', *Behavioral and Brain Sciences* 16 (1): 1-14.(A vigorous defence of a radically anti-Cartesian view of introspection. Contains an excellent bibliography of the experimental literature and is followed by comments from a large number of philosophers and psychologists of varying perspectives.)

Hume, D. (1739) *A Treatise of Human Nature*, Oxford: Clarendon Press, 1978. (Hume's discussion of personal identity can be found at bk I, pt 4, §6.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N.K. Smith, New York: St Martin's Press, 1965.(Kant's discussion of introspection appears in B131-6; B152-8; A341-66; B406-32.)

Kornblith, H. (1989) 'Introspection and misdirection', *Australasian Journal of Philosophy* 67 (4): 410-22.(Argues on empirical grounds that introspection is either useless or counterproductive in providing a check on the

manner in which we arrive at our beliefs.)
Lehrer, K. (1991) *Metamind*, Oxford: Oxford University Press.(Presents a Cartesian perspective on the importance of introspective understanding for a wide range of rational activities.)
Lyons, W. (1986) *The Disappearance of Introspection*, Cambridge, MA: MIT Press.(Critically examines accounts of introspection in psychology and philosophy of mind.)
Nagel, T. (1974) 'What is it like to be a bat?', *Philosophical Review* 83 (4): 435-50. (Presents the classic statement of the difficulties, discussed in §4 above, which introspective understanding presents for a physical account of the world. Nagel stops short of endorsing dualism.)
Nisbett, R. and Wilson, T. (1977) 'Telling more than we can know: Verbal reports on mental processes', *Psychological Review* 84 (3): 231-59.(Important contribution to the experimental literature on the kinds of errors to which introspection is susceptible.)
Ryle, G. (1949) *The Concept of Mind*, New York: Barnes & Noble.(Offers a behavioural account of the mind and mental concepts generally, including introspection.)
Shoemaker, S. (1988) 'On knowing one's own mind', *Philosophical Perspectives* 2: 183-209.(Defends the Cartesian view that the essence of certain mental states lies in our special introspective access to them.)

Introspection, psychology of

Introspection is the process of directly examining one's own conscious mental states and processes. Since the seventeenth century, there has been considerable disagreement on the scope, nature and epistemic status of introspection. Descartes held that all our mental states are subject to introspection; that it is sufficient to have a mental state to be aware of it; and that when we introspect, we cannot be mistaken about what we 'see'. Each of these views has been disputed. Nineteenth-century psychology relied heavily on introspection, but with the exception of important work in psychophysics and perception, it is now primarily of historical interest. Recently, Nisbett and Wilson have argued that when people attempt to report on the processes mediating the effects of a stimulus on a response, they do not do so on the basis of introspection but, rather, on the basis of an implicit common-sense 'theory'. Ericsson and Simon have developed a model of the mechanisms by which 'introspective' reports are generated and have used that model to identify the conditions under which such reports are reliable.

1 The classical conception

The term 'introspection' comes from the Latin 'spicere' (to look) and 'intra' (within). In ordinary English, it means 'the observation of one's own mental states or processes'. As we shall see in §3, whether introspection truly constitutes a form of observation is a matter of dispute. However, it is generally agreed that introspection involves awareness of certain aspects of one's mind, specifically, an awareness that involves a 'higher-order' belief or representation that one is in some 'lower-order' mental state. (A thought about cats is a 'first-order' mental state, whereas a belief *that one has a thought about cats* is 'second-order'.)

The first important treatment of introspection is that of Descartes (1641), who held that there is nothing to the mind but what is in consciousness and available to introspection (see Consciousness §7; Introspection, epistemology of). Introspective awareness of our mental states and processes might be thought to be the result of a separate higher-order process trained on lower-order states and processes. However, for Descartes, the connection is much closer than that: mental states and processes are, on his view, 'self-intimating'. It is sufficient simply for a mental state or process to be 'within us' for us to be aware of it. Indeed, Descartes held that the beliefs that result from introspection are infallible. As long as we consider our mental states only as they are 'in themselves', they cannot be false.

2 The scope of introspection

Does the set of states and processes that can be introspected exhaust the mental? Philosophical responses to this question fall into two traditions. Until well into the nineteenth century, most Anglo-American philosophers, including, in particular, the British Empiricists (Locke, J.,Hume, D., Mill, J.) and the American, William James, followed Descartes in answering in the affirmative (see Empiricism; Sense-data). In contrast, many German philosophers, beginning with Leibniz and including Kant and Schopenhauer, have argued that there are aspects of mind that are hidden from introspective awareness.

The current view is largely anti-Cartesian. Both Freudian psychoanalytic theories and contemporary cognitive science posit the existence of unconscious mental states and processes, that is, mental states and processes to which introspection has no access (see Unconscious mental states). A popular view among psychologists, in fact, is that introspective accessibility is not simply a black and white affair but, rather, that there are multiple levels of consciousness, each with its own conditions of accessibility.

3 The nature of introspection

There is substantial disagreement about the role of introspection and the kind of process introspection might be. Many disagree, for example, about whether one can be in a conscious mental state without being introspectively aware that one is in that state. The received view (see, for example, Armstrong 1968; Rosenthal 1986; Dretske 1993) is that one can; Shoemaker (1994) has argued to the contrary.

Second, there is disagreement over what 'order' of belief or representation is required for introspective awareness. Most philosophers who adopt the 'higher-order thought' view opt for a second-order thought as sufficient.

Rosenthal, however, suggests that second-order thoughts are simply what makes a state conscious; introspective awareness requires that we be conscious of those second-order thoughts and, hence, that there be a *third*-order thought that one is having the second-order ones.

Finally, there is considerable discussion over whether introspection is correctly considered to be a form of perception or observation, a view associated with Locke (1689). For example, in perception, there is both object-awareness and fact-awareness. I perceive the tree and perceive that there is a tree. Are there also two forms of awareness in introspection? Do I both perceive my pain and perceive that I am in pain? Shoemaker and Dretske argue that introspection does not involve awareness of any (mental) object such as a ‘pain’. Furthermore, in perception, the relation between the objects perceived and the state of perceiving is causal. Is the same thing true for introspection, or is the relation also conceptual, as Shoemaker has argued? In other words, are there mental states whose introspectibility is essential to their very nature? This would be the case if one adopts Descartes’ view that mental states are self-intimating or Shoemaker’s view that mental states are defined by their functional role, and this role includes producing introspective awareness of the states in question (see Consciousness §10).

4 Epistemic status and scientific role

Many philosophers have taken issue with the Cartesian view that introspection is infallible. For example, James held that because introspection involves conceptualization, it suffers from fallibility just as any mode of observation does. Armstrong agrees that introspective awareness is fallible, but for a different reason, namely, that the state being introspected and the introspecting state are ‘distinct existences’, and, hence, connected only contingently.

Psychologists have focused not so much on the epistemic status of introspective awareness *per se* but on whether introspection can constitute a reliable method for the scientific study of the human mind. There were considerable efforts in the nineteenth century to found psychology on careful introspective reports, and the work of Weber, Fechner and Helmholtz gave rise to important work on the ‘psychophysics’ of sensation (or the relation of introspected to physical magnitudes of, for example, light and sound) that continues to the present day (see Stevens 1951; Sekuler and Blake 1994). In contrast, the ‘structuralist’ programme of Wundt and Titchener, which sought to discover the basic elements of thought and the laws by which they combine into more complex mental experiences, was far less successful.

In the early nineteenth century, Comte (1830) issued two challenges to a general introspective psychology. First, he asserted that introspection of intellectual (as opposed to emotive) activity is impossible. The conscious self is, in some sense, essentially unified. Hence, it cannot ‘split’ itself into two, one portion engaged in first-order intellectual activity with a second portion looking on. Second, even supposing that introspection is possible, it is worthless, presumably, because it is unreliable (see Consciousness §6).

Psychologists and philosophers interested in scientific psychology have responded to these challenges in a variety of ways. For example, James (1890) agreed with Comte that introspective awareness of ‘on-line’ intellectual activity is impossible. However, since we clearly are often aware of what passes in our minds, we must ask how this can be. His view, following Mill (1865), was that introspection is actually retrospection. We are not aware of our occurrent states as we have them but, rather, of our *immediate memories* of those states. Since these memories are not part of the stream of consciousness, the impossible split in the conscious self is avoided (see Consciousness §10).

Comte’s second challenge is, perhaps, the more serious of the two. For it might well be the case that although introspection is possible, the process of introspecting distorts the first-order mental activity being introspected. The threat of such distortion was a serious problem for the introspectionist psychologists of the late nineteenth and early twentieth century. In fact, some have argued that Titchener’s failure to take the problem of unreliability sufficiently to heart contributed not only to the downfall of introspectionist psychology in particular, but also to the radical behaviourist ‘revolution’ and the eclipse of any mentalistic psychology for much of the twentieth century (see Behaviourism, methodological and scientific).

Behaviourism has itself now been overthrown in favour of an approach that, once again, takes the mental seriously. As a consequence, psychologists are struggling with the question of whether and to what extent introspective awareness can give us information about our mental states and processes. A contemporary challenge

to the reliability of introspection comes from the work of Nisbett and Wilson (1977) who considered the psychological causes of our judgments and behaviour. They argued that people have little or no privileged introspective access to such causal information. Instead, their reports are based on implicit, cultural 'theories' that they apply to themselves as they do to others.

In one classic study cited by Nisbett and Wilson, subjects were asked to tie together the ends of two cords hung from the ceiling that were placed far enough apart so that one cord could not be reached while holding on to the other. To solve the problem, subjects were permitted to use any of the various objects strewn about the room. There were several solutions, one of which involved attaching a weight to the end of one cord, and then swinging it so that it could be caught while holding on to the second cord. Most subjects could not discover this solution on their own. However, when cued by the experimenter casually putting one of the cords in motion in the periphery of the subjects' vision, they typically solved the problem within 45 seconds. Subsequently, though, when subjects were asked how they arrived at this solution, the experimenter's hint was mentioned by less than a third of the subjects.

The most powerful defence of introspection in recent years comes from two cognitive scientists, Ericsson and Simon (1980). They argued that verbal reports of mental activities constitute important data, but that we can only assess the epistemic value of this data if we understand the mechanisms by which the reports are generated. They then proposed a model according to which cognitive information can be stored in various memories, including short-term memory (STM), which has a limited capacity and a short duration, and long-term memory (LTM), which has a very large capacity and is relatively permanent. Information that is attended to is kept in STM and is directly accessible for verbal report. Information stored in LTM must be retrieved before it can be reported (see Memory).

There are many different kinds of verbal report. First, we can distinguish reports that are concurrent with the execution of the primary task (such as solving a problem in mathematics) from reports that are retrospective. Second, some verbal reports ('level I') require no additional processing. This is so when information used in the primary task is already in verbal form (in STM) and simply requires articulation. In contrast, there are verbal reports that do require additional processing, either recoding, say, from visual to verbal form ('level II'), or filtering and providing information (such as the reasons for certain behaviour or choices) to which attention would not ordinarily be paid ('level III').

The Ericsson and Simon model predicts what a host of experimental studies have borne out; that each of these different types of verbal report will differ in accuracy. In particular, concurrent level I reports neither change nor slow the course or structure of the primary-task cognitive process. Concurrent level II reports may slow the primary-task cognitive processes, but also do not change them. However, for level III reports, there may well be substantial effects on task performance. Finally, the accuracy of retrospective reports depends not only on the level involved but also on the duration of the task and the lag between completion and the report. As both the duration and lag time increases, accuracy will decrease.

Ericsson and Simon claim that their model is quite consistent with Nisbett and Wilson's findings. In particular, in the studies reviewed by Nisbett and Wilson, the retrospective procedures used to elicit the actors' reports were not those the model identifies as eliciting valid reports. For example, in the cord-tying task, subjects do not report the causal role of the hint because by the time they are queried, this information is no longer in short-term memory.

See also: Consciousness; Introspection, epistemology of; Qualia; Unconscious mental states

BARBARA VON ECKHARDT

References and further reading

Armstrong, D.M. (1968) *A Materialist Theory of the Mind*, London: Routledge.(Discussion of the nature and epistemic status of introspection in the context of a defence of central state materialism. Presents arguments mentioned in §§3, 4.)

Churchland, P.M. (1985) *Matter and Consciousness*, Cambridge, MA: MIT Press.(Chapter 4 includes several arguments against the infallibility of introspective awareness.)

Comte, A. (1830) *Introduction to Positive Philosophy*, trans. and ed. F. Ferre, Indianapolis, IN: Hackett Publishing Company, 1988.(The classic statement of objections to the possibility and reliability of

introspection.)
Danziger, K. (1980) ‘The History of Introspection Reconsidered’, *Journal of the History of the Behavioral Sciences* 16: 241-62.(A useful discussion of the history of introspective psychology with emphasis on the period from 1880 to 1914.)
Dennett, D.C. (1978) ‘Toward a Cognitive Theory of Consciousness’, in *Brainstorms*, Cambridge, MA: MIT Press.(Sketch of an information-processing model of consciousness, including an account of introspection.)
Descartes, R. (1641) ‘Meditationes de prima philosophia’, in *Oeuvres de Descartes*, vol. 7, ed. C. Adam and P. Tannery, Paris: CNRS/Vrin, 1964; trans. ‘Meditations on First Philosophy, including Objections and Replies’, in *The Philosophical Writings of Descartes*, vol. 2, ed. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1984.(The classical philosophical treatment of consciousness and introspection.)
Dretske, F. (1993) ‘Conscious Experience’, *Mind* 102: 263-83.(Presents argument mentioned in §3.)
Ericsson, K.A. and Simon, H.A. (1980) *Protocol Analysis: Verbal Reports as Data*, Cambridge, MA: MIT Press; revised edn, 1993.(A classical cognitive science treatment of the reliability of introspective verbal reports.)
Farthing, G.W. (1992) *The Psychology of Consciousness*, Englewood Cliffs, NJ: Prentice Hall.(A useful psychology text on consciousness, with two chapters on introspection.)
Hill, C. (1991) *Sensations: A Defense of Type Materialism*, Cambridge: Cambridge University Press.(An interesting treatment of introspection which emphasizes its active nature.)
Howe, R.B.K. (1991) ‘Introspection: A Reassessment’, *New Ideas in Psychology* 9: 25-44.(A defence of introspection as a method in psychology.)
James, W. (1890) *The Principles of Psychology*, New York: Holt.(Chapter 7, volume 2 includes James’ response to Comte, discussed in §4.)
Locke, J. (1689) *An Essay concerning Human Understanding*, with notes by A.C. Fraser, New York: Dover, 1959. (The classical statement of the view that introspection is a form of perception.)
Lycan, W. (1995) ‘Consciousness as Internal Monitoring’, in *Consciousness and Experience*, Cambridge, MA: MIT Press, ch. 2.(Defends the Lockean view that introspection is inner sense, mentioned in §3, against numerous objections.)
Lyons, W. (1986) *The Disappearance of Introspection*, Cambridge, MA: MIT Press.(A useful overview of the various historical and contemporary views on introspection.)
Mill, J.S. (1865) *Auguste Comte and Positivism*, Ann Arbor, MI: University of Michigan Press, 1968.(Includes Mill’s response to Comte’s arguments mentioned in §4.)
Nisbett, R. and Wilson, T. (1977) ‘Telling More Than We Can Know: Verbal Reports on Mental Processes’, *Psychological Review* 84 (3): 231-59.(An extensive review of the psychological literature that argues that people are often unaware of what causes their actions.)
Rosenthal, D.M. (1986) ‘Two Concepts of Consciousness’, *Philosophical Studies* 49: 329-59.(Contrasts the Cartesian and a non-Cartesian conception of consciousness.)
Rosenthal, D.M.(1993) ‘Thinking that One Thinks’, in M. Davies (ed.) *Consciousness: Psychological and Philosophical Essays*, Oxford: Blackwell.(Develops an argument for the higher-order thought view of consciousness.)
Sekuler, R. and Blake, R. (1994) *Perception*, New York: McGraw-Hill.(A good recent textbook on psychophysics.)
Shoemaker, S. (1994) ‘Self-Knowledge and "Inner Sense"’, *Philosophy and Phenomenological Research* 54 (2): 249-315.(An extended argument against the view that introspection is a form of perception. Presents arguments mentioned in §3.)
Stevens, S.S. (1951) *A Handbook of Experimental Psychology*, New York: Wiley & Sons.(A classic contemporary anthology including works in psychophysics.)
White, P.A. (1988) ‘Knowing More About What We Can Tell: "Introspective Access" and Causal Report Accuracy 10 Years Later’, *British Journal of Psychology* 79: 13-45.(A review of the literature published in response to Nisbett and Wilson (1977) that calls some of their conclusions into question.)

Intuitionism

Ultimately, mathematical intuitionism gets its name and its epistemological parentage from a conviction of Kant: that intuition reveals basic mathematical principles as true a priori. Intuitionism's mathematical lineage is that of radical constructivism: constructive in requiring proofs of existential claims to yield provable instances of those claims; radical in seeking a wholesale reconstruction of mathematics. Although partly inspired by Kronecker and Poincaré, twentieth-century intuitionism is dominated by the 'neo-intuitionism' of the Dutch mathematician L.E.J. Brouwer. Brouwer's reworking of analysis, paradigmatic for intuitionism, broke the bounds on traditional constructivism by embracing real numbers given by free choice sequences. Brouwer's theorem - that every real-valued function on a closed, bounded interval is uniformly continuous - brings intuitionism into seeming conflict with results of conventional mathematics.

Despite Brouwer's distaste for logic, formal systems for intuitionism were devised and developments in intuitionistic mathematics began to parallel those in metamathematics. A. Heyting was the first to formalize both intuitionistic logic and arithmetic and to interpret the logic over types of abstract proofs. Tarski, Beth and Kripke each constructed a distinctive class of models for intuitionistic logic. Gödel, in his Dialectica *interpretation, showed how to view formal intuitionistic arithmetic as a calculus of higher-order functions. S.C. Kleene gave a 'realizability' interpretation to the same theory using codes of recursive functions. In the last decades of the twentieth century, applications of intuitionistic higher-order logic and type theory to category theory and computer science have made these systems objects of intense study. At the same time, philosophers and logicians, under the influence of M. Dummett, have sought to enlist intuitionism under the banner of general antirealist semantics for natural languages.*

1 Kantian epistemology, constructivism and 'semi-intuitionism'

The thesis from which mathematical intuitionism draws its name is Kant's: that we exercise a pure intuition to apprehend the correctness of basic mathematical claims directly and a priori. Kant defended the thesis - for arithmetic and geometry - in the 'Transcendental Aesthetic' of his first *Critique*. Twentieth-century intuitionists, under the leadership of L.E.J. Brouwer, founder of Dutch or neo-intuitionism, adopted a modification of Kant's view. Intuition in elementary arithmetic and analysis was to be identified with a pure intuition of the subdivision of time intervals. Kant's account of geometry was refused and only elementary arithmetic retained intuitive primacy (see Mathematics, foundations of). By Brouwer's lights, the role of intuition cannot be understudied by empirical inquiry, linguistic understanding or logical reasoning. Brouwer's intuitionism, then, opposes both empiricism and logicism. Moreover, since intuitionists think to know intuitively principles which are infinitary and contentful, intuitionism is also at odds with finitism and formalism.

Deprecation of logic's influence on mathematics has been characteristic of intuitionism and is reminiscent of the attitudes Schopenhauer expressed in *Die Welt als Wille und Vorstellung (The World as Will and Representation)*. According to Schopenhauer, mathematics is not best developed *more geometrico*, since deductions from axioms obscure the true grounds for mathematical results. He argued that each mathematical truth *S* has its own intuitable ground of being *G* and that the premier means to intuit the truth of *S* would be to uncover *G* and construct it in time without the aid of explicit deduction. Schopenhauer recommended that mathematics be reformulated accordingly.

Since before the time of Brouwer, intuitionistic mathematics has been constructive. In general, the demand for constructivism is the demand that E be respected:

(E) The correctness of existential claim $\exists x A(x)$ is to be guaranteed by warrants from which both an object *a* and a further warrant for $A(a)$ are constructible.

Constructivism is commonly deemed revisionary of conventional mathematics, requiring alteration in standard notions of mathematical objects, canons of adequate proof and fundamental principles of reasoning (see Constructivism in mathematics). Intuitionists trace their constructive lineage at least as far back as Leopold Kronecker (1823-91), who initiated a programme for arithmetizing higher algebra; in this, he demanded for arithmetic a primacy irreducible to natural science or logic and refused to countenance non-constructive existence proofs.

Writings of the so-called ‘semi-intuitionists’, particularly Poincaré and Borel, exerted a strong influence on Brouwer and his followers. Poincaré, considered the leading European mathematician of his day, also rejected the idea that geometric truths express synthetic judgments known a priori but embraced the intuitive aprioricity of elementary arithmetic, with mathematical induction holding pride of place (see Poincaré, J.H. §§3, 5). In common with the Dutch intuitionists, Poincaré construed infinite totalities as potential, rather than completed. Accordingly, he rejected, along with Russell’s type theory, Zermelo’s axioms for set theory as guarantors of transfinite arithmetic. Borel was a prominent member of the ‘Paris school’ of mathematicians, which included Baire and Lebesgue. He advocated constructivism to the extent that existence was to be tied to explicit definability: specific functions and real numbers were only to be accepted once they were fully defined. Borel distinguished, as did the young Brouwer, a ‘geometric’ from a ‘practical’ continuum. The former was deemed the object of a general intuition, not analysable into individual reals. The latter was the totality of definable real numbers. It was probably in writings by Borel (or an earlier book by Du Bois Reymond) that Brouwer came upon the idea that real numbers be given by choice sequences of rationals generated randomly, as if by successive throws of a die.

2 Brouwer and Dutch intuitionism

The intuitionism of the twentieth century is that of Brouwer: a radical extension of constructive mathematics incorporating anti-classical principles of logic and analysis. Brouwer held the chair in set theory and axiomatics at the University of Amsterdam from 1912 until 1955. He first achieved fame for discoveries in algebraic topology, particularly proofs of the invariance of dimension and of the fixed-point theorem that bears his name. In articles published between 1907 and 1930, Brouwer combined a battery of influences with an austere personal philosophy of life to create a programme for intuitionistic mathematics. The programme called for trenchant criticism of presuppositions of conventional mathematics as well as thorough reconstruction of arithmetic, analysis and set theory so as to free them of those presuppositions.

In accordance with his subjectivistic, even solipsistic, personal philosophy, Brouwer insisted that intuitionistic mathematics be a private study of mental mathematical constructions. Language and interpersonal communication, indeed all intellectual activity, played diminished roles in Brouwer’s scheme of things. He thought language an instrument so defective as to make mathematicians incapable of conveying the precise sense of their constructions to others. Hence, if mathematics is to retain certainty and exactness, it was not to consist in linguistic representations but in languageless construction. Brouwerian mathematics would not be able to derive its authority from the application to mathematics of a universal logic, valid for all domains of thought.

After Brouwer, the principal exponent of Dutch intuitionism was his student and colleague, Heyting. In addition to major contributions to the formalization of intuitionistic logic, arithmetic and set theory and to the interpretation of that formalism, Heyting pursued studies in intuitionistic algebra, geometry and the theory of Hilbert space. Among early intuitionists were M.J. Belinfante, G.F.C. Griss, H. Freudenthal, B. van Rootselaar and, for a time, H. Weyl. The most distinguished proponents of Dutch intuitionism during the last quarter of the twentieth century are students of Heyting: A.S. Troelstra and D. van Dalen. Notable as well has been a group of intuitionists from Nijmegen including H. de Swart and W. Veldman.

3 Elements of Brouwerian intuitionism

Already in his 1907 dissertation, ‘On the Foundations of Mathematics’, Brouwer was advancing the intuitionistic idea that mathematics is a ‘free creation’ of the mind unfettered by empirical data or linguistic constraints. Beginning with ‘The Unreliability of the Logical Principles’ (1908 [1975]), Brouwer took critical aim at the ground rules of conventional mathematics, challenging the validity of the law of the excluded middle (LEM). (The 1908 article may also reflect views held by Brouwer’s teacher and colleague G. Mannoury.)

The validity of LEM does conflict with a natural view of constructivity. At present, it is unknown whether there exist an infinite number of twin primes, that is, pairs of prime numbers that differ by 2. Let the predicate $P(x)$ hold of *p* just in case $\langle p, p+2\rangle$ is the greatest twin prime pair, if such a pair exists, or of 1, if there are an infinite number of twin primes. If LEM is valid, there exists a determinate *m* such that $P(m)$. However, short of solving the problem of twin primes, no one can provide, in accordance with E, a recipe for computing *m* in decimal notation. It is noteworthy, first, that Brouwer did not think that this style of reasoning proved the negation of LEM. Rather, he maintained that it affords LEM a ‘weak counterexample’, a proof that the intuitionist will, most likely,

never be able to endorse that law. (Quantified versions of LEM are refuted by intuitionistic continuity principles, described below.) Second, a direct mathematical connection is here forged between the truth of disjunctions and the availability of recipes for deciding an issue.

Brouwer also constructed weak counterexamples to such conventional theorems as that every real is either rational or irrational and that the continuum is totally ordered. In the case of the former, generate in parallel the decimal expansions of π and of ρ, where ρ's expansion begins $0.33333\ldots$ but is to be halted when a string of seven 7s appears in the expansion of π. Clearly, ρ cannot be irrational, for then the desired string neither appears nor fails to appear. But, to assert that ρ is rational, E requires that one locate integers *m* and *n* such that ρ is their quotient. This is not possible, for we would then know where a string of seven 7s appears in the expansion of π, something now unknown.

Prior to 1920, Brouwer adopted the notion of 'choice sequences', also called 'infinitely proceeding' sequences. These need be neither law-like - governed by computable recipes for generating terms - nor even fully determinate in advance. Brouwer permitted restrictions imposed by a spread law (see below) but, beyond that, nothing about the future course of the sequence may be known, other than the fact that its terms are freely and independently chosen.

Brouwer employed two intuitionistic analogues for the conventional notion of a class of numbers. These are the concepts now called 'species' and 'spread'. Species are properties determined by extensional mathematical predicates and are themselves individuated extensionally: pairs of species are identical when they apply to the same items. Spreads are either species of choice sequences constrained to obey a spread law or species of sequences obtained from these via continuous mappings. A 'spread law' is a decidable property of finite sequences which holds of the empty sequence and culls out from the tree of all sequences a subtree containing branches unbounded in length. A choice sequence is an 'element' of a spread *S* when it determines a branch in *S*'s tree. A spread with a tree which always branches finitely often was called by Brouwer a 'fan'.

These notions in hand, Brouwer proved the most celebrated result of intuitionistic mathematics, now known as 'Brouwer's theorem': every real-valued function of a real variable defined on a closed, bounded interval is uniformly continuous. This theorem is known to follow from two distinctively intuitionistic principles. The first, originally isolated by Heyting, is weak continuity (WC-N). Let 'α' and 'β' range over choice sequences and '*n*' over natural numbers. Neighbourhoods of α are determined by its initial segments *s*: β lies in the *s*-neighbourhood of α ($\beta \in s$) whenever it shares *s* with α.

(WC-N) If $\forall\alpha\exists nA(\alpha,n)$, then, $\forall\alpha\exists n\exists$ neighbourhood *s* of α $\forall\beta\in sA(\beta,n)$.

E and the indeterminacy of choice sequences are thought to secure plausibility for WC-N. When we have $\forall\alpha\exists nA(\alpha,n)$, there is to hand a uniform method *M* for constructing, from each choice sequence α, an *n* such that $A(\alpha,n)$. Since *M* is uniform, it can take cognizance, in calculating a specific *n* from α, of no more than an initial segment of α. After all, we may know nothing more of the course of α than the terms already calculated. Hence, *M* must yield a proof, for that *n*, of $\forall\beta\in sA(\beta,n)$. WC-N is alone sufficient to yield strong counterexamples to quantified forms of LEM. Contemporary proofs of Brouwer's theorem also rely on the 'fan theorem', an intuitionistic correlate of König's lemma, which asserts that, with respect to the neighbourhoods just described, every fan is constructively a compact topological space.

In such articles as 'Über Definitionsbereiche von Funktionen' (On Domains of Definitions for Functions, 1927 [1975]), Brouwer attempted a detailed proof of Brouwer's theorem and explained how to deduce the fan theorem from another assertion about spreads, the 'bar theorem'. The latter is, in effect, a principle of 'reverse induction' up from the bottom nodes in a countable, well-founded tree. Let *S* be a spread. A set of finite sequences *B* in *S* is a 'bar' when all α in *S* have initial segments in *B*. A set of finite sequences *A* in *S* is 'hereditary' if *A* holds of sequence *s* in *S* whenever it holds of all *s*'s descendants in *S*. The bar theorem asserts that, whenever *A* is hereditary and contains a decidable bar *B*, the empty sequence belongs to *A*. Unfortunately, Brouwer's attempt to prove it suffered from a non sequitur and is widely considered inconclusive.

Most controversial of Brouwer's innovations, even for fellow intuitionists, was his theory of the 'creative subject'. In articles published after 1947, Brouwer put to work, in defining choice sequences, a representation in temporal

series of the discoveries of an idealized mathematician. From this representation, Brouwer looked to derive strong counterexamples - direct refutations - for findings of conventional analysis, among them that every linear equation with non-zero coefficients has a real solution. Some see the creative subject as a child of Brouwer's subjectivistic philosophy because of Brouwer's assumption that any claim which is permanently unknown to the ideal mathematician is false. Due to difficulties attending its formal analysis, the creative subject remains a subject of debate.

4 Heyting's logic and the proof interpretation

To win a prize offered by the Dutch Mathematical Association, Heyting made the first efforts at a comprehensive formalization for Brouwerian logic, arithmetic and set theory. His system, published in 1930 and 1931, was anticipated in part by the incomplete systems of A.N. Kolmogorov and V.I. Glivenko. Notational details aside, Heyting's predicate logic, now treated as definitive of intuitionistic logic, is equivalent to Gentzen's later natural deduction system NJ. Standard formalizations of intuitionistic predicate logic and elementary arithmetic are still called 'Heyting predicate logic' - or HPL, for short - and 'Heyting arithmetic' - HA - in honour of his achievement.

In a monograph of 1934, Heyting advanced an interpretation of the logic which, he believed, expressed its Brouwerian meaning. He gave, by recursion on claims *A*, conditions on abstract intuitionistic proofs of *A*. The 'proof interpretations' for ' $\vee$', ' $\rightarrow$' and ' $\forall$' statements are, roughly, as follows.

($\vee$) A proof of $A \vee B$ consists in both a number and a further proof such that, if the number has value 0, the further proof is one of *A* and, if the number has value 1, the proof is of *B*.
($\rightarrow$) A proof of $A \rightarrow B$ affords a construction converting any proof of *A* into one of *B*.
($\forall$) A proof of $\forall n A(n)$ affords a construction *f* such that, for any number *m*, $f(m)$ proves $A(m)$.

If constructions are computable operations, logics sound for this interpretation may not validate LEM. Let $A(n)$ define an undecidable set of natural numbers. Then, $\forall n(A(n) \vee \neg A(n))$ plausibly lacks proof. Were there such a proof - one satisfying ($\rightarrow$) and ($\forall$) - it would yield, computably in *n*, an $f(n)$ correctly marking the truth or falsity of $A(n)$, thus contradicting $A(n)$'s undecidability.

The idea of testing logical principles with abstract constructions and proofs is visible in Brouwer's writings. In 1932, Kolmogorov had offered, independently of Heyting, a structurally similar interpretation of intuitionistic reasoning as a 'calculus of problems'. Consequently, the proof interpretation is also known as the 'BHK interpretation', after Brouwer, Heyting and Kolmogorov. The adequacy of certain aspects of the BHK has been a matter of academic dispute. For example, if condition ($\rightarrow$) defines the concept 'abstract proof of $A \rightarrow B$', then the definition is impredicative: it contains a quantification - 'any proof of A' - over all proofs (see Constructivism in mathematics).

5 Metamathematics: intuitionistic systems and translation theorems

In the first decades after 1945, S.C. Kleene and G. Kreisel, together with their students and colleagues, led a host of metamathematical explorers into traditional intuitionistic territory. Kleene's system for Brouwer's analysis, 'FIM' after Kleene's book with R.E. Vesley, *The Foundations of Intuitionistic Mathematics* (1965), brought to it standard formalization. Among Kreisel's innumerable contributions are formal systems for choice sequences, higher-order arithmetic and the theory of constructions. Of special interest to post-war proof theorists was the system T invented by Gödel to express, for his *Dialectica* interpretation, the behaviours of primitive recursive functionals of finite type (see §6 below).

The 1970s brought out a bumper crop of intuitionistic formal theories for sets, types and higher-order functions. Troelstra (1973a) subjected HAS, a second-order version of Heyting arithmetic containing variables for species and an impredicative comprehension axiom, to logical examination. For computer scientists (see §7), Martin-Löf's unconventional ML type theories acquired a certain celebrity. The ML systems were not invented, as was FIM, to formalize Brouwer's analysis but to represent a mixture of traditional ideas with Martin-Löf's own approach to constructivism. Among the ideas ML assimilates are the BHK interpretation and Frege's distinction between thought and assertion as well as such proof-theoretic mainstays as normalization and W. Howard's 'formulas-as-types'.

As for set theories, IZF - 'intuitionistic Zermelo-Fraenkel' - was a formalism of choice for studying universes of Heyting-valued sets (see §6). H. Friedman's system B was intended to formalize E. Bishop's constructive analysis but is no stronger, proof-theoretically, than HA. P. Aczel's 'constructive Zermelo-Fraenkel' CZF is translatable into an ML system with a type for well-founded trees. CZF is close kin to J. Myhill's constructive set theory CST.

In 1933, Gödel - and, independently, Gentzen and Bernays - showed how to embed the theorems of conventional Peano arithmetic (PA) faithfully into the negative ($\vee, \exists$-free) fragment of HA. So, PA is seen to be conservative over HA for negative formulas and, as far as consistency goes, HA is commensurate with its classical cousin. Unknown to Gödel and Gentzen, Kolmogorov had already proposed, in 1925, a similar negative translation for a fragment of HPL. More recently, negative translations have been discovered for higher-order arithmetic, type theory and set theory. Gödel also constructed (in 1933) a translation from Heyting's propositional logic into the modal calculus S4. H. Rasiowa and R. Sikorski later applied Gödel's idea to quantified systems. During the 1980s, modal interpretations for intuitionistic arithmetic motivated N. Goodman and S. Shapiro to recommend systems for 'intensional mathematics'.

6 Metamathematics: algebraic, functional and realizability interpretations

Just as conventional logical signs represent Boolean operations, the logical signs in HPL can also be read algebraically. The relevant algebraic operations act on the open sets of a topological space or, more generally, on elements of Heyting algebras - kinds of distributive lattices with greatest element T. So construed, every formula of HPL gets assigned a topological or lattice-theoretic 'truth-value' such that HPL-theorems always evaluate to T. In 1936, S. Jaśkowski seems to have been the first to interpret Heyting's propositional logic over finite distributive lattices. In the next two years, M. Stone and A. Tarski independently interpreted the logic over topological spaces. Tarski's article (1938) includes a proof, in conventional metalogic, that every propositional formula which is not an intuitionistic theorem has a countermodel over the open subsets of $\mathbb{R}$. A. Mostowski gave topological semantics to HPL. D.S. Scott (1968) showed how to extend topological interpretations to intuitionistic analysis, thereby constructing models for Brouwer's theorem. In the 1970s, Scott was among the prime movers in the study of Heyting-valued algebraic semantics for intuitionistic higher-order logic and set theory.

Close relatives of topological models (and of Cohen forcing) are Beth and Kripke models. In 1947, E.W. Beth introduced the idea of interpreting HPL over domains attached to nodes in finitely branching trees. He later argued that Heyting's logic is sound and complete with respect to his notion of model and that this fact is provable intuitionistically. (In the latter, Beth was shown to have erred.) S. Kripke (1965) proposed, in effect, that formulas be assigned upward closed subsets of partially ordered sets as 'truth-values'. He gave conventional proofs that HPL is sound and complete with respect to these assignments and that validity in Kripke structures and in Beth structures are equivalent. Both types of models have been extended to other intuitionistic systems, including set theory. Later researchers saw that sets of Kripke or of Beth models can be 'glued' together to prove that HA, not to mention other systems, possesses the disjunction property (DP) and explicit definability for numbers (EDN). A system *S* has DP just in case, whenever *A* and *B* are sentences and *S* derives $A \vee B$, then *S* derives *A* or *S* derives *B*. *S* has EDN if, whenever $\exists n A(n)$ is a sentence that *S* derives, there is a number term *t* such that *S* derives $A(t)$.

The premier functional interpretation for HA was Gödel's 1959 *Dialectica*, named for the journal in which it first appeared. The point of the interpretation, conceived by Gödel as early as 1941, was to obtain a consistency proof for arithmetic via an extension of Hilbertian finitistic reasoning to higher types. Here, each arithmetic formula is replaced by a quantifier-free translation in the formal system T. The logic of HA is thereby subsumed within combinatorics for higher-type objects. Gödel maintained that the correctness of T is sufficiently evident that the interpretation gives a convincing proof for HA's consistency. Extensions and applications of Gödel's ideas were soon forthcoming: in the late 1950s and early 1960s, Kreisel and C. Spector constructed *Dialectica*-style interpretations for analysis; J.-Y. Girard found functional interpretations for HAS in 1971.

Generally, a 'realizability interpretation' is an association between formulas *A* and items *i* such that *i* serves, intuitively, to encode information available from a constructive proof of *A*. When *i* is so associated with *A*, it is said to 'realize' *A* or to be a realizability witness for *A*. On Kleene's original realizability interpretation from 1945, witnesses are natural numbers that encode the computational information given by *A*. Witnesses of conditionals, for example, are Turing machine indices such that, if $[[A]]$ and $[[B]]$ are the sets of witnesses for *A* and *B*,

respectively, then $[[A \rightarrow B]]$ is the set of all indices of machines which, given inputs from $[[A]]$, output in $[[B]]$. Kleene proved each HA-theorem to have a non-empty set of witnesses. Realizability also shows HA to be consistent with Church's thesis, the assertion that every arithmetical function is Turing computable. Kleene later devised a realizability interpretation for FIM in which witnesses are functions rather than numbers. He also enlarged the idea of realizability to incorporate his *q*-realizability and 'slash' interpretations; these permit straightforward proofs of DP and EDN for HA. Many forms of realizability have been invented, extending the concept to most intuitionistic systems.

7 Category theory and computer science

In the 1970s and 1980s, applications of formal systems to category theory and to computer science stimulated new interest in intuitionism. During 1969, F.W. Lawvere isolated axioms for categories which are elementary topoi - generalizations of Grothendieck topoi. Together with M. Tierney, he verified that topoi represent categorial versions of set-theoretic universes which are intuitionistic, rather than conventional, in their logic. The set of truth-values within a topos is always a Heyting, rather than a Boolean, algebra. Research has shown that the theory of interpretations over categories encompasses not only algebraic and topological models but also Kripke models, Beth models and realizability. (See Category theory, applications to the foundations of mathematics.)

Computer scientists have put formal constructive proofs to work extracting programs. As E would have it, a proof *p* of *A* is constructive to the extent that there are methods *M* for extracting from *p* computational information relevant to *A*. For instance, if *p* is a constructive proof of $\forall n \exists m A(n, m)$, *M* should extract from *p* a recipe guaranteed to compute a function *f* such that, for all *n*, $A(n, f(n))$. The power of high-level programming languages makes it possible to ask for software to play the role of methods *M*: to extract automatically, from formal proofs *p*, programs guaranteed to compute functions with the behaviours *p* specifies. Such software would convert a formal constructive proof that a specification is satisfiable into a program certain to satisfy it. During the 1980s, a number of researchers looked to implement constructive program extraction, among them R.L. Constable and his associates at Cornell (1986), creators of the PRL ('proof refinement logic') systems inspired by Martin-Löf's ML.

8 Intuitionism and antirealism

Since the 1970s, Michael Dummett has been principal sponsor for the view that foundational issues for intuitionism and its logic are properly addressed in philosophical discussions over programmatic 'antirealist' semantics for natural language. Dummett proposed that debates between proponents of conventional and of intuitionistic logics gain a solid footing only when viewed as one skirmish within a wider battle between proponents of realist and of antirealist semantics. Dummett's semantic realist insists that the meaning of statements be explicated by theories associating each statement with bivalent truth-conditions which obtain or fail independently of human knowledge. Dummett's conjectured antirealist would have it that the meanings of statements can only be given in terms of operations people have been taught to employ in verifying or falsifying claims, and not in terms of truth-conditions which may prove ineffable. Intuitionism, then, is supposed to represent an expression of antirealism for which the relevant semantic operations would be those entering into the fashioning of intuitionistic proofs. (See Antirealism in the philosophy of mathematics §2.)

Obstacles to the antirealist programme include the difficulties inherent in designing, even in outline, a coherent antirealist semantics that covers all the territory Dummett has suggested it might cover, including antirealism about psychology and the past. Also, it is in no way clear that features of intuitionistic logic and mathematics can be well understood from the vantage of a semantic theory. It remains to be shown that intuitionistic mathematics, rather than finitism or Markovian constructivism, is the true ally of antirealist semantics. Lastly, it is a concern that Dummett's ideas flout one of the main planks of Brouwer's intuitionism: that mathematics must be independent of metaphysics. For traditional intuitionists, there can be no final explanation for intuitionistic mathematics other than the mathematics itself.

See also Logical and mathematical terms, glossary of

DAVID CHARLES McCARTY

References and further reading

Beeson, M.J. (1985) *Foundations of Constructive Mathematics*, Berlin: Springer.(An overview of the metamathematics of constructive formal systems; requires considerable formal experience.)

Brouwer, L.E.J. (1975) *Collected Works*, vol. 1, ed. A. Heyting, Amsterdam: North Holland.(An invaluable compendium of Brouwer's writings, some in English and others in German; the articles 'The Unreliability of the Logical Principles' and 'Über Definitionsbereiche von Funktionen' (On Domains of Definitions for Functions) and sections of his dissertation appear. The level of technical difficulty is high.)

Brouwer, L.E.J. (1981) *Brouwer's Cambridge Lectures on Intuitionism*, ed. D. van Dalen, Cambridge: Cambridge University Press.(A compact sample of Brouwer's thinking; well edited but quite technical.)

Constable, R.L. *et al*. (1986) *Implementing Mathematics With the NuPrl Proof Development System*, Englewood Cliffs, NJ: Prentice-Hall.(An admirable introduction to the computer system but no substitute for 'hands-on' experience.)

Dummett, M.A.E. (1977) *Elements of Intuitionism*, Oxford: Clarendon Press, 1990.(An influential text in which great emphasis is placed upon the metalogic of intuitionism.)

Dummett, M.A.E. (1978) *Truth and Other Enigmas*, London: Duckworth, and Cambridge, MA: Harvard University Press.(Essential reading on intuitionism and antirealism; the preface is especially recommended.)

Fourman, M.P. and Scott, D.S. (1979) 'Sheaves and Logic', in M.P. Fourman *et al*. (eds) *Applications of Sheaves*, Springer Lecture Notes in Mathematics, vol. 753, Berlin: Springer, 302-401.(A mathematically sophisticated but lucid introduction to Heyting-valued interpretations of intuitionistic systems.)

Gödel, K. (1986-95) *Collected Works*, ed. S. Feferman *et al.*, New York and Oxford: Oxford University Press, 3 vols.(Authoritative compendium of Gödel's published works and very fine source books. The clear and painstaking introductions are especially recommended.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879- 1931*, Cambridge, MA: Harvard University Press.(An invaluable collection of significant articles by Gödel, Hilbert, Brouwer, Skolem and others; technical knowledge often required.)

Heyting, A. (1956) *Intuitionism: An Introduction*, Amsterdam: North Holland; 3rd revised edn, 1971. (The first - and still the most charming - popular technical account of Brouwer's intuitionism.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, New York: St Martin's Press, 1965.(The 'Transcendental Aesthetic' section includes Kant's highly influential views on a priori intuition.)

Kleene, S.C. (1945) 'On the Interpretation of Intuitionistic Number Theory', *Journal of Symbolic Logic* 10: 109-24.(Technical but wonderfully clear and amazingly prescient.)

Kleene, S.C. and Vesley, R.E. (1965) *The Foundations of Intuitionistic Mathematics*, Amsterdam: North Holland. (Mathematically exacting, historically important but dauntingly formal.)

Kreisel, G. (1965) 'Mathematical Logic', in T.L. Saaty (ed.) *Lectures in Modern Mathematics*, vol. 3, New York: Wiley, 95-195.(A challenging sampler of Kreisel's - and others' - thoughts and theorems on the foundations of mathematics.)

Kripke, S.A. (1965) 'Semantical Analyses of Intuitionistic Logic', in J. Crossley and M.A.E. Dummett (eds) *Formal Systems and Recursive Functions*, Amsterdam: North Holland, 92-130.(Includes formal results on the completeness of intuitionistic logic but also nontechnical explications of the model theory.)

Lambek, J. and Scott, P. (1986) *Introduction to Higher-Order Categorical Logic*, Cambridge: Cambridge University Press.(A fine introduction to a highly technical subject.)

Martin-Löf, P. (1984) *Intuitionistic Type Theory*, Naples: Bibliopolis.(Based on lectures delivered by the author. Offers a helpful introduction to a detailed subject.)

Rasiowa, H. and Sikorski, R. (1963) *The Mathematics of Metamathematics*, Warsaw: PWN.(A standard reference on algebraic semantics; requires technical proficiency.)

Schopenhauer, A. (1818, 1844) *Die Welt als Wille und Vorstellung*, trans. E.F.J. Payne, *The World as Will and Representation*, New York: Dover, 1969, vol. 1, §15.(Schopenhauer's reflections on mathematics.)

Scott, D.S. (1968) 'Extending the Topological Interpretation to Intuitionistic Analysis I', *Compositio Mathematica* 20: 194-210.(A topological treatment of Brouwerian analysis that calls for metamathematical experience.)

Smorynski, C. (1973) 'Applications of Kripke Models', in A.S. Troelstra (ed.) *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, Lecture Notes in Mathematics, vol. 344, Berlin: Springer, 324-91.(A lucid technical introductory overview to the model theory of intuitionistic arithmetic.)

Tarski, A. (1938) 'Der Aussagenkalkül und die Topologie' (The Propositional Calculus and Topology), *Fundamenta Mathematica* 31: 103-34; repr. and trans. in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 1983, 421-54.(Characteristically careful and clear exposition of early results on topological interpretations.)

Troelstra, A.S. (1973a) 'Notes on Intuitionistic Second-Order Arithmetic', in A. Mathias and H. Rogers (eds)

Cambridge Summer School in Mathematical Logic, Berlin: Springer, 171-205.(Remains a fine example of the metamathematics of higher-order intuitionistic systems. A companion piece to Troelstra (1973b).)

Troelstra, A.S. (ed.) (1973b) *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, Lecture Notes in Mathematics, vol. 344, Berlin: Springer.(Terse and highly technical. For metamathematicians of the 1970s and 1980s, the bible of intuitionistic systems.)

Troelstra, A.S. and Dalen, D. van (1988) *Constructivism in Mathematics: An Introduction*, Amsterdam: North Holland, 2 vols.(A comprehensive technical overview of the current state of constructive mathematics, with emphasis on intuitionism and mathematical logic.)

Intuitionism in ethics

To intuit something is to apprehend it directly, without recourse to reasoning processes such as deduction or induction. Intuitionism in ethics proposes that we have a capacity for intuition and that some of the facts or properties that we intuit are irreducibly ethical. Traditionally, intuitionism also advances the important thesis that beliefs arising from intuition have direct justification, and therefore do not need to be justified by appeal to other beliefs or facts. So, while intuitionism in ethics is about the apprehension of ethical facts or properties, traditional intuitionism is principally a view about how beliefs, including ethical beliefs are justified. Varieties of intuitionism differ over what is intuited (for example, rightness or goodness?); whether what is intuited is general and abstract or concrete and particular; the degree of justification offered by intuition; and the nature of the intuitive capacity. The rejection of intuitionism is usually a result of rejecting one of the views that lie behind it.

Note that 'intuition' can refer to the thing intuited as well as the process of intuiting. Also, somewhat confusingly, intuitionism is sometimes identified with pluralism, the view that there is a plurality of fundamental ethical properties or principles. This identification probably occurs because pluralists often accept the epistemological version of intuitionism.

1 The motivation for intuitionism

Traditional intuitionism in ethics is not so much an attempt to establish that we have justification for ethical beliefs as an attempt to explain how ethical beliefs get their justification. That some ethical beliefs have justification is taken as a datum to be explained.

The form this explanation takes is determined by an explicit or implicit commitment to three views: moral realism, the autonomy of ethics and foundationalism in the theory of knowledge. Moral realism is the view that there are ethical facts or properties that are in some way objective or mind-independent, although the exact nature of the objectivity or mind-independence is controversial (see Moral realism). That ethics is autonomous means that ethical facts cannot be 'reduced' to, or fully explained in terms of, nonethical ones (see Autonomy, ethical §5). For example, this would preclude something's goodness being analysable in terms of someone's desire for it. Foundationalism is a general theory about the justification of beliefs, asserting that all justification has as a foundation beliefs that have direct justification (see Foundationalism). Nonfoundational beliefs must rely for their ultimate justification on beliefs that have direct justification.

If it is accepted that we sometimes have justification for ethical beliefs and the explanation of their justification has to be consistent with these three views, then traditional intuitionism is almost inevitably the result. It should be noted, however, that some modern philosophers, such as Jonathan Dancy (1983), hold that some ethical facts or properties must be apprehended (intuited) directly, but reject foundationalism and the view that beliefs arising from intuition have direct justification. In what follows I shall focus on traditional intuitionism.

Moral realism, the autonomy of ethics, and foundationalism lead to traditional intuitionism as follows. The combination of moral realism and the autonomy of ethics ensures that ethical beliefs are about irreducibly ethical facts or properties and, thus, that ethical beliefs are irreducibly ethical. According to foundationalism, justification of beliefs about these irreducibly ethical facts or properties has as a foundation beliefs that have direct justification (that is, they are self-evident or need no evidence). Irreducibly ethical beliefs cannot have nonethical beliefs as their foundation. For if they could, the contents of the ethical beliefs would have to be so intimately related to the contents of the foundational nonethical beliefs as to be inconsistent with the autonomy of ethics. For example, if beliefs about desires provide the foundation for beliefs about goodness, then desiring and goodness would have to be intimately related. So, for any ethical beliefs to have justification, there must be some ethical beliefs that have direct justification and arise from direct apprehension (intuition). (If the foundational beliefs could arise from indirect apprehension, they could also be justified indirectly.)

Adding the 'datum' that we do sometimes have justification for ethical beliefs, we can conclude that some of our ethical beliefs arise from intuition and have direct justification. This is traditional intuitionism in ethics.

2 Varieties of intuitionism

Although elements of intuitionism have always had a place in ethical debate, it was only in the seventeenth century that it became a distinctive view (see Cambridge Platonism). More accurately, it became the distinctive core of a group of views, which has been developed in different ways.

For a start, in developing this core, we can ask what intuitions are about. For example, are they about rightness and wrongness, as according to H.A. Prichard (1949) and Joseph Butler (§3), goodness and badness, as held by G.E. Moore (1903), or defeasible moral reasons for performing and not performing actions, as W.D. Ross argued (1930; 1939) (see Prichard, H.A. §2; Moore, G.E.; Ross, W.D.)? Is what we intuit general and abstract, as Richard Price (§2) and Henry Sidgwick (§2) maintained, or concrete and particular (Prichard and Butler)? For example, if intuitions are about the rightness of actions, would we intuit that it is morally right to perform any action of a certain general type or would we intuit that a particular act we are contemplating is morally right?

Since justification comes in degrees, there can be divergent views about the degree of justification conferred on beliefs arising from intuition (see Moral knowledge §3). Although intuition has usually been taken to confer justification at least sufficient for knowledge, it could be held to confer any degree of justification from certainty (which guarantees truth) to a mere presumption in favour of the fact intuited. If complete justification is not conferred on a belief by intuition, then, of course, it may acquire additional justification on other grounds, such as the degree to which it coheres with other beliefs. Also, intuitionists can take different sides on the externalism/internalism question about justification (see Internalism and externalism in epistemology). Typically, internalists say that there has to be an awareness of the features that give a belief justification, while externalists deny this. Intuitionists who are internalists might say, for example, that a belief arising from intuition has justification only if the believer holds it to be indubitable (impossible not to believe) or, alternatively, merely difficult not to believe. On the other hand, externalists deny that such psychological features are required for the justification of beliefs arising from intuition.

There are also differing views about the nature of the intuitive capacity. It can be taken to be an intellectual capacity, as it was by Price and Moore, or a capacity for something more like perception, as it was by Prichard and Thomas Reid. Those who take the intuitive capacity to be intellectual usually hold that the contents of the intuitions are general or abstract principles. Those who take it to be more like perception usually believe that intuitions have particular, concrete contents.

It could be held that we have the capacity for both kinds of intuition, intellectual and perceptual, and each kind gives us foundational beliefs. It is more usual, however, to hold that we have the capacity for only one kind of intuition, that its contents (general or particular) are foundational and beliefs of the other kind are indirectly justified, if at all.

If we have the capacity only for intellectual intuition and we intuit general principles, then particular ethical beliefs might be justified by recognizing that some situation involves an instance of the general principle. For instance, if we intuit a general principle like ‘There is some ethical reason not to do any act that results in harm to anyone’, we might recognize that a particular act falls under the principle and conclude that we have an ethical reason not to do it. Alternatively, if the capacity is taken to be perceptual and the intuitions to be about particular situations, then the general principles may be justified by some kind of induction on the particular cases. For example, we might intuit something like ‘The harm this action would cause gives me an ethical reason not to do it’ and use some form of induction to arrive at the general principle.

In any case, it is important to distinguish between the order of justification and the order of discovery. It could be that general principles are discovered through consideration of particular cases, but are justified directly, while the particular beliefs are justified only indirectly, as instances of the general principle or content. For example, we may have to experience a number of instances of lying before we can intuit the general principle that lying is, *ceteris paribus*, wrong, while we are justified in believing that a particular instance of lying is, *ceteris paribus*, wrong, only by applying the general principle.

Finally, the capacity for ethical intuition needs grounding in some mechanism. Those who believe that the intuitive capacity is intellectual and the contents of intuition general usually hold that these contents are necessary, synthetic truths, that there are other such truths, such as those of mathematics, and that the same capacity and mechanism is used in apprehending all such truths (Price and Ross). Some who hold that that intuition is particular

and similar to perception postulate a 'moral sense' (Reid and Prichard) (see Moral sense theories). In neither case is the mechanism well understood.

3 Why not intuitionism?

The rejection of intuitionism is usually a consequence of rejecting one or more of the views that motivate it: moral realism, the autonomy of ethics and, in the case of traditional intuitionism, foundationalism. Although each of these views is controversial, the combination of moral realism and the autonomy of ethics is the source of objections to any form of intuitionism in ethics.

Intuitionism, and its commitment to irreducibly ethical facts or properties that we can apprehend directly, is incompatible with a broadly scientific, or naturalistic, picture of the world and how we learn about it. According to this picture, irreducibly ethical facts or properties would have to be occult. They would not be the sort of things that we can have scientific theories about. Matters are only made worse by claiming that we have some faculty that allows us to apprehend directly these irreducibly ethical properties or facts, to say nothing of the claim that beliefs arising from this direct apprehension are justified directly. Since the properties or facts are occult, so too must be the faculty that allows our awareness of them. Since persons would have an occult faculty, we could not have a complete scientific theory about persons. Indeed, the fact that moral realism and the autonomy of ethics lead to the claim that we have a capacity for intuition does not provide support for intuitionism; rather, it gives us ample reason to believe that at least one of the views leading to it must be wrong. Or so goes the objection.

Those who reject intuitionism by rejecting the autonomy of ethics usually argue that ethical properties or facts are actually nonethical ones, whose apprehension is compatible with the scientific picture of the world (see Naturalism in ethics). For example, it might be argued that to say that something is morally required is to say that it brings about more pleasure than its alternatives. Note that such views involve moral realism and may involve foundationalism.

Rejecting intuitionism by rejecting moral realism is consistent with accepting the autonomy of ethics. For example, it could be held that what seem to be ethical judgments are actually irreducibly ethical prescriptions, the expression of irreducibly ethical attitudes or the projection of such attitudes (see Analytic ethics; Emotivism; Projectivism).

Another possibility is to accept that our moral practices and beliefs are committed to intuitionism and that intuitionism's incompatibility with the scientific picture of the world shows that our ethical practices and beliefs involve fundamental error (see Moral scepticism §3; Moral realism §7).

There are several possible responses to the claim that intuitionism is incompatible with a scientific picture of the world. The compatibility of these outlooks rests, it can be argued, on a proper understanding of the scientific picture of the world. In the same vein, this compatibility can be said to depend on a proper understanding of intuition. A further possible response is to say 'so much for the scientific picture'.

See also: Common sense ethics; Epistemology and ethics; Moral judgment

ROBERT L. FRAZIER

References and further reading

Audi, R. (1996) 'Intuitionism, pluralism, and the foundations of ethics', in W. Sinnott-Armstrong and M. Timmons (eds) *Moral Knowledge?: New Readings in Moral Epistemology*, Oxford: Oxford University Press, 101-36.(A modern defence of a view which is quasi-intuitionistic.)

Dancy, J. (1983) 'Ethical particularism and morally relevant properties', *Mind* 92: 530-47.(An example of nontraditional intuitionism, arguing that what we intuit is particular and concrete.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press, esp. chaps 5, 6.(Holds that intuitions are general, that the truth of utilitarianism is self-evident and that we can intuit the goodness of some properties.)

Prichard, H.A. (1949) *Moral Obligation*, Oxford: Oxford University Press.(Argues that we intuit the rightness of particular actions.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Clarendon Press, esp. ch. 2.(Classic exposition of pluralist intuitionism.)

Ross, W.D. (1939) *The Foundations of Ethics*, Oxford: Clarendon Press, esp. 79-86 and ch. 8.(Holds that we intuit

general principles about the moral relevance of types of actions.)
Schneewind, J.B. (1990) *Moral Philosophy from Montaigne to Kant*, Cambridge: Cambridge University Press, 2 vols.(Contains descriptions of, and extracts from, many of the early intuitionists, including Richard Cumberland, Samuel Clarke, Shaftesbury, Francis Hutcheson, Joseph Butler, Richard Price and Thomas Reid.)

Intuitionistic logic and antirealism

The law of excluded middle (LEM) says that every sentence of the form $A \vee \neg A$ ('A or not A') is logically true. This law is accepted in classical logic, but not in intuitionistic logic. The reason for this difference over logical validity is a deeper difference about truth and meaning. In classical logic, the meanings of the logical connectives are explained by means of the truth tables, and these explanations justify LEM. However, the truth table explanations involve acceptance of the principle of bivalence, that is, the principle that every sentence is either true or false. The intuitionist does not accept bivalence, at least not in mathematics. The reason is the view that mathematical sentences are made true and false by proofs which mathematicians construct. On this view, bivalence can be assumed only if we have a guarantee that for each mathematical sentence, either there is a proof of the truth of the sentence, or a proof of its falsity. But we have no such guarantee. Therefore bivalence is not intuitionistically acceptable, and then neither is LEM.

A realist about mathematics thinks that if a mathematical sentence is true, then it is rendered true by the obtaining of some particular state of affairs, whether or not we can know about it, and if that state of affairs does not obtain, then the sentence is false. The realist further thinks that mathematical reality is fully determinate, in that every mathematical state of affairs determinately either obtains or does not obtain. As a result, the principle of bivalence is taken to hold for mathematical sentences. The intuitionist is usually an antirealist *about mathematics, rejecting the idea of a fully determinate, mind-independent mathematical reality.*

The intuitionist's view about the truth-conditions of mathematical sentences is not obviously incompatible with realism about mathematical states of affairs. According to Michael Dummett, however, the view about truth-conditions implies antirealism. In Dummett's view, a conflict over realism is fundamentally a conflict about what makes sentences true, and therefore about semantics, for there is no further question about, for example, the existence of a mathematical reality than as a truth ground for mathematical sentences. In this vein Dummett has proposed to take acceptance of bivalence as actually defining a realist position.

If this is right, then both the choice between classical and intuitionistic logic and questions of realism are fundamentally questions of semantics, for whether or not bivalence holds depends on the proper semantics. The question of the proper semantics, in turn, belongs to the theory of meaning. Within the theory of meaning Dummett has laid down general principles, from which he argues that meaning cannot in general consist in bivalent truth-conditions. The principles concern the need for, and the possibility of, manifesting one's knowledge of meaning to other speakers, and the nature of such manifestations. If Dummett's argument is sound, then bivalence cannot be justified directly from semantics, and may not be justifiable at all.

1 Standard intuitionistic logic

The term 'intuitionistic logic' is used in two related senses. In the first sense, intuitionistic logic is the set of logical principles which are valid according to general intuitionism, as originally proposed by the Dutch mathematician L.E.J. Brouwer (1908). In the second, more standard, sense, intuitionistic logic is the logic determined by any one of several equivalent sets of logical rules and axioms. Such an axiom system, motivated by Brouwer's ideas, was first proposed by the Dutch logician Arend Heyting (1930) (see Intuitionism).

Usually, the phrase 'intuitionistic logic' is taken to mean the logic you get by accepting classical logic and withdrawing your acceptance of the law of excluded middle (LEM). LEM is the law according to which every sentence of the form $A \vee \neg A$ is logically true. Intuitionistically, this law is rejected. This means that the intuitionist is not allowed to assert a sentence $A \vee \neg A$ purely on logical grounds. It does not mean that the negation of that sentence may be asserted. Intuitionistically, as much as classically, the sentence $\neg(A \vee \neg A)$ is a contradiction.

Because of the inconsistency of $\neg(A \vee \neg A)$ we are allowed to infer its negation, $\neg\neg(A \vee \neg A)$. Classically, we could now infer $A \vee \neg A$, because the rule of double negation elimination (DNE), that is, the rule that from $\neg\neg A$ you may infer *A*, is valid in classical logic. Therefore, DNE is also not intuitionistically valid. By equally simple, intuitionistically acceptable reasoning, you can infer DNE from LEM. In fact, in natural deduction, systems for intuitionistic logic differ from their classical counterparts precisely by not containing DNE.

Classically, LEM is valid in virtue of the truth table explanations of the propositional constants. By the explanation of ‘ $\neg$’, either *A* is true or $\neg A$ is true, and by the explanation of ‘ $\vee$’, nothing more is needed for the truth of $A \vee \neg A$. So those who reject LEM must understand these logical constants in a different way.

2 Logic from the intuitionistic viewpoint

L.E.J. Brouwer is commonly regarded as the father of intuitionism, but he himself traces the intuitionist tradition back at least to Kant, who thought of intuitions of space and time as basic to mathematics (see Kant, I. §5). In Heyting (1934) the following two theses characterize intuitionism:

(1) Mathematics has not only formal significance, but also content.
(2) Mathematical objects are immediately grasped by the thinking subject; therefore mathematical knowledge is independent of experience.

Heyting then subdivides intuitionism into ‘half-intuitionism’ and intuitionism proper. The half-intuitionist holds that mathematical objects exist independently of our thinking, but can be known by means of our mathematical constructions. The full-blown intuitionist denies this, or at least rejects the idea that the thesis of mind-independent existence of mathematical objects can be used as a means of mathematical proof.

How can the realistic view of mathematical objects be used as a means of proof? Precisely because that view licenses the use of the law of excluded middle. If mathematical objects exist independently, and have all their properties independently, then, for instance, either, as a matter of fact, the sequence of digits 0123456789 does occur somewhere in the decimal expansion of π, or, as a matter of fact, it does not. For each mathematical statement, either mathematical reality renders that statement true, or else it does not, in which case the statement is false. In this way mathematical realism justifies LEM. LEM, in turn, is used as a means of proof.

Brouwer rejected mathematical realism. He thought that mathematics is an activity where propositions are proved by means of mental constructions, and above all that such mental constructions are precisely what make mathematical propositions true. The statement that 0123456789 does occur somewhere in the decimal expansion of π is rendered true by a construction, for example, by the construction of the decimal expansion up to the point where the sequence occurs, and is rendered false by a construction which proves that it cannot ever occur. And (roughly) a disjunction is rendered true by having one of its disjuncts rendered true. On this view, we are not in advance justified in asserting that the sequence either does or does not occur, for we have no guarantee, before a proof is actually found, that either we can carry out the verifying construction or else we can carry out the falsifying construction.

On these grounds Brouwer rejected LEM. He went further, in that, in sharpest contrast to the formalist school in the philosophy of mathematics (see Hilbert’s programme and formalism), he refused to accept logic at all as a separate source of knowledge in mathematics. In particular, that a set of mathematical axioms is consistent is not in itself a good reason, according to Brouwer, for accepting them. The relation between logic and mathematics is the opposite. Logical validity depends on insight in the nature of mathematical truth. And this leads to an understanding of the logical operators themselves.

3 Bivalence and the meaning of the logical constants

The truth table explanations give the meaning of the classical propositional constants. For instance, the meaning of ‘ $\rightarrow$’ is given by the fact, or stipulation, that $A \rightarrow B$ is false if *A* is true and *B* false, and true in the other three cases. This gives the classical meaning of ‘ $\rightarrow$’ for the reason that these four truth-value combinations exhaust all possibilities, which, in turn, involves the principle of *bivalence*, the principle that each statement is either true or false.

Under default assumptions, bivalence holds if and only if LEM is valid. The default assumptions are two: first, that $(A \vee B)$ is true if *A* is true or *B* is true. Second, that $\neg A$ is true if *A* is false.

The intuitionist does not accept the principle of bivalence. Again, this means that the intuitionist is not allowed, without further justification, to assert of a sentence *A* that it is either true or false. It does not mean that the intuitionist may assert that there are sentences which are neither true nor false. In fact, the question whether there are such sentences is a quite separate issue (see Vagueness §3; Presupposition §2; Many-valued logics,

philosophical issues in). However, the intuitionist as well as the classical logician may adopt the view that some sentences are neither true nor false, for example, because of non-referring singular terms. The modified classical logician will then claim that each sentence is either true, or false, or neither true nor false, and the modified intuitionist will not accept this principle.

The intuitionist also accepts the truth tables, but for the intuitionist they do not explain the meaning of the constants. Since the principle of bivalence is not accepted, it cannot be assumed that the entries in the truth tables exhaust all possibilities. The intuitionist may be allowed to assert $A \rightarrow B$, on the basis of a deduction of *B* from *A*, without being allowed to assert that either *A* is false or *B* is true.

On the intuitionistic conception, a mathematical proposition is true only if there is a proof of it. Moreover, the *meaning* of a mathematical sentence is closely tied to what counts as a proof of it. This is the leading idea in the Brouwer-Heyting or Brouwer-Heyting-Kolmogorov interpretation of the logical constants. Heyting once suggested that a mathematical sentence expresses an expectation of finding a proof of it. The meaning of '$\vee$' is then recursively explained by saying that a sentence $A \vee B$ expresses an expectation which is fulfilled just in case at least one of the expectations expressed by *A* and by *B* is fulfilled.

The modern version of this kind of explanation is due primarily to Michael Dummett (1973). Dummett suggests that we can correlate forms of sentence to forms of proof. There is no one form that is common to all proofs of a sentence $A \vee B$, but we can make a distinction between 'canonical' and 'non-canonical' proofs, proposed by both Dummett and the Swedish logician Dag Prawitz (1974), such that a canonical proof of $A \vee B$ is a proof consisting of either a proof of *A* or a proof of *B*. Non-canonical proofs are constructions which give us *effective* methods of finding canonical proofs. An effective method is a method which is guaranteed to yield the result after a finite number of elementary steps.

Then a proof of $A \vee B$ either consists of a proof of *A* or a proof of *B*, or of an effective method of finding the one or the other. A canonical proof of $A \,\&\, B$ is a pair of proofs, of *A* and of *B* respectively. A canonical proof of $A \rightarrow B$ is an effective method of transforming any proof of *A* into a proof of *B*. A canonical proof of $\neg A$ is an effective method of transforming any proof of *A* into a proof of '$\bot$' (absurdity - a sentence constant that is false by definition), and therefore a proof of the impossibility of a proof of *A*. A canonical proof of $\forall x A$ is an effective method for finding, for any term *t*, a proof of *At*. A canonical proof of $\exists x A$ is a proof of *At*, for some term *t*.

These are the intended intuitionistic interpretations of the logical constants. It follows from these meaning explanations that a non-canonical proof of $A \vee \neg A$ is an effective method for finding either a proof of *A* or a proof of $\neg A$. If there is such a non-canonical proof of $A \vee \neg A$, then *A* is called 'effectively decidable' (in Dummett's terminology, which we follow in §6; more technically, the term applies to predicates, or infinite classes of sentences). However, the notion of an absolutely undecidable sentence, that is, of a sentence *A* such that neither *A* nor $\neg A$ can be proved at all, does not make sense intuitionistically. *A* cannot be proved only if it can be proved that *A* cannot be proved (since truth amounts to provability), but such a proof constitutes a proof of $\neg A$, contradicting the second assumption that $\neg A$ cannot be proved.

A semantic idea that is closely related to the Brouwer-Heyting-Kolmogorov interpretation is the idea that the meaning of the logical constants is given by the rules for them in natural deduction, as first devised by Gerhard Gentzen (1934) and later developed by Prawitz (1965) (see Logical constants §4). In particular there is a close analogy between the introduction rules in natural deduction and the forms of canonical proofs. For instance, the introduction rule for 'Čamp;' is that $A \,\&\, B$ may be inferred from the joint premises of *A* and *B*. From this alternative perspective, introduction rules are valid by definition, and all other rules are to be justified with respect to these.

From both these kinds of meaning-explanation, standard intuitionistic logic is justifiable and LEM is not. But the question is whether further rules than the standard ones are valid. This could be answered in the negative by a proof of semantic completeness, that is, of the statement that all intuitionistically valid principles are derivable from standard intuitionistic logic. Such a proof would have to invoke the intended interpretations rather than some other semantics, and it would have to employ only intuitionistically valid modes of reasoning. At the time of writing, it is not clear that such a proof can be found.

4 Dummett on antirealism and semantics

Brouwer rejected mathematical realism, rejected the law of excluded middle and adopted the idea that mathematical statements are true in virtue of mental constructions. These views naturally go together, but their interrelations are non-trivial.

Antirealism, that is, rejection of realism, comes in many forms. The most direct form is a denial of the existence of entities, for example, the existence of universals. Brouwer denied the existence of extra-mental mathematical objects. If this is denied, then there is a question of evaluating mathematical sentences. Since we want to preserve standard arithmetical judgments, we must understand a sentence such as ‘ $2 + 2 = 4$’ so that neither its meaningfulness, nor its truth, depends on the existence of entities of the rejected kind. One possibility is to interpret the sentence in terms of mental constructions, as Brouwer did, but it is not the only one. You can also reinterpret arithmetical sentences, for example, so as to make them true or false by virtue of facts about numerals and their operators. You can even do it in such a way that, if you are a realist about numerals, the principle of bivalence remains valid. So not every form of antirealism will necessitate a change of logic. It will always depend on the way the sentences are reinterpreted.

The form of view suggested, combining a denial of the existence of entities (numbers) with an affirmation of bivalence for (arithmetical) sentences, has been categorized by Dummett as a form of sophisticated realism, rather than a form of antirealism. On Dummett’s view, belief in the determinacy of facts is more central to realism than belief in the existence of entities. A mathematical antirealist can accept the existence of numbers and reject the statement that every mathematical state of affairs either obtains or does not obtain. On this view we just cannot take for granted that every mathematical question has an answer in mathematical reality.

These reflections suggest a central place for the principle of bivalence in questions of realism, and that is just what Dummett has proposed. However, there are two more specific features of Dummett’s views on bivalence and realism. First, Dummett places bivalence, as a criterion of realism, in an epistemic perspective. In this perspective, realism is in the first place opposed to the knowability principle, that is, the view that if a sentence is true, then it can be known to be true (see Meaning and verification §5). If you accept bivalence, then, from the epistemic perspective, what is important is that you hold every sentence to be either true or false whether or not it is possible to find out which. And if you reject bivalence, the default reason, from the epistemic perspective, is that on the one hand you accept the knowability principle and, on the other, you do not think there is any guarantee that each sentence can be either known to be true or known to be false.

Second, Dummett regards acceptance of bivalence not just as characteristic of the realist, but as actually defining what it is to be a realist. Similarly, rejection of bivalence by definition makes you an antirealist. Although Dummett recognizes further aspects of realism and antirealism, bivalence is taken as the main dividing line. In this way realism becomes a semantic issue (hence the terms ‘semantic realism’ and ‘semantic antirealism’ in discussion of Dummett’s views).

There are three interconnected reasons for this change of perspective. First, Dummett holds that traditional metaphysical disputes are framed in pictorial language, with pictures like that of an ethereal realm of abstract entities. These pictures should be replaced by more precisely stated claims. Secondly, the conflict about, for example, the reality of the past, is a conflict about what makes our sentences about the past true, and therefore the question of the reality of the past will involve the question of the proper semantics for sentences about the past. The realist thinks that a sentence can be made true by past facts even if it is impossible for us to know what they are, and the antirealist rejects this as incoherent. Similarly for other disputes over realism. Third, if the conflict is taken as a conflict about semantics, then it can actually be resolved. It will be resolved once it has been shown what a correct meaning theory for the language in question must look like.

5 The relation between realism and bivalence

However, it may seem that Dummett has overstated the role of bivalence in metaphysics. Semantics is what connects bivalence, for some range of sentences, and the determinacy of facts in the corresponding area. What connection there is depends on the kind of semantics, that is, on how the sentences are interpreted. But it also depends on how well the meanings of the sentences are determined.

Suppose that you interpret sentences about the past according to a correspondence model, so that they are made true or false by past facts or events. The sentence ‘George Washington crossed the Potomac’ is then true, if it is true, in virtue of the event of Washington’s crossing the Potomac. Suppose further that all sentences concerned have a well-determined meaning. If a sentence has a well-determined meaning, then it depends only on reality whether or not the sentence is true.

Under these assumptions realism and bivalence are equivalent. If past reality is fully determinate, so that all past states of affairs either obtained or did not obtain, and the meanings of the sentences are such that this and nothing else is needed for making them true or false, then each sentence is true or false. And similarly, if bivalence holds, so does realism.

You may doubt that reality is fully determinate, suspecting that there might be gaps, as it were, and take this as a reason for accepting the knowability principle, thinking that our only guarantee that some state of affairs does or does not obtain is that we can get to know either that it obtains or that it does not. In that case, under the same initial assumptions, bivalence will fail. Conversely, if you accept the knowability principle for another reason, then you have a reason to reject bivalence, since there is no guarantee that each sentence can be known either to be true or to be false. Again under the same assumptions, you must reject realism.

However, if one of these two assumptions is dropped, then the equivalence is lost. Suppose that you give up the second assumption, about the determinacy of meaning. You may think, for example, that the application of a predicate such as ‘…crossed…’ to a pair of objects, such as Washington and the Potomac, is determinately either correct or incorrect only if it is within our present capacity to decide whether it is the one or the other. If we do not have that capacity, then, on this view, the meaning of the predicate is underspecified for these arguments. This view leads to the rejection of bivalence. We cannot assert that each sentence is either true or false, since if we cannot either verify or falsify a particular sentence, then, on this view, its truth-value is left indeterminate. In this case, however, there are no consequences for reality. Reality may be fully determinate. The failure of bivalence depends on meaning only, not on the world.

We get the same result if the first assumption is dropped. Suppose that we adopt a kind of verificationism for sentences about the past. We interpret them so that they are made true or false not by past facts or events, but by present evidence, that is, by such present facts as we take as grounds for asserting those sentences, such as facts about written documents or memories. Then again we have a reason to reject bivalence, for there is no guarantee that we will, for each sentence about the past, be in possession either of truth-making evidence or of falsity-making evidence. And again nothing seems to follow about realism. If sentences about the past are not made true or false by past facts, then failure of bivalence implies nothing about the determinacy of past reality.

The natural reply to the objection is that we have no conception whatsoever about the reality of the past save as that which is represented by our sentences about the past. If these sentences cannot reasonably be given a semantics according to which they are true or false in virtue of past facts, then the notion of an independent past reality must be rejected, for it is only as the truth-maker of our sentences that we have a reason to be realists about it. Moreover, we cannot then use the sentence ‘Past facts are fully determinate’ for stating an independent metaphysical truth, for this sentence must itself be evaluated according to the proposed alternative semantics.

If this reply is correct, then questions of metaphysics are *eo ipso* questions of semantics. And then, as Dummett thinks, metaphysical conflicts are to be settled within the theory of meaning. Indeed, according to Dummett (1991), expressions of metaphysical pictures of reality have no content at all in addition to the semantic views that go with them.

6 Dummett’s antirealist argument

We have seen that the principle of bivalence is of central importance both for realism and for classical logic. Bivalence itself, for any kind of sentence, or for some particular kind, is to be justified by the proper semantics for the kind of sentence in question. And what the proper semantics is, is a question in the general theory of meaning.

Within the theory of meaning Dummett has, notably in ‘What is a Theory of Meaning? (II)’ (1976), proposed general principles, from which an argument is drawn to the conclusion that a semantics of bivalent truth-conditions cannot be correct.

The first step in Dummett's reasoning is the observation that an account of linguistic meaning must be an account of linguistic understanding, that is, of knowledge of meaning. The thesis that meaning consists of truth-conditions is acceptable if and only if it is acceptable that understanding a sentence consists of knowing its truth-conditions, and similarly for other accounts of meaning.

Second, understanding must be publicly manifestable. That is, if I understand a particular sentence, then I must be able to reveal to my interlocutors that I do. This is because language is a means of communication. If there were ingredients in meaning such that knowledge of them could not be made manifest to others, then speakers could not know whether they attached the same meanings to the same expressions, and then communication, as far as those ingredients in meaning are concerned, would be impossible (see Private states and language).

Thus, if meaning consists of truth-conditions, knowledge of truth-conditions must be manifestable. But how do you manifest your knowledge of the truth-conditions of a sentence? Sometimes you can manifest your understanding by means of a verbal explanation of the meaning of the sentence, but in this way the problem is transferred to the question of your knowledge of the meaning of the sentences you use in your explanation. When you cannot, without circularity, explain further, then you must manifest your understanding in a more primitive way. Dummett says that such a manifestation is a manifestation of a particular capacity, namely, the capacity to get to know that the sentence is true, if it is true, and to get to know that it is false, if it is false. According to Dummett, this is what counts as manifesting knowledge of truth-conditions.

The next step concerns the possibility of such manifestations. If the sentence in question is effectively decidable (see §3 above), then manifesting knowledge of its truth-conditions is possible in principle, but perhaps not in practice. If a sentence is practically decidable, that is, deciding it is within the scope of ordinary human abilities, then I can manifest my knowledge of its truth-conditions by applying a decision method, like making an observation or a small number of calculations, and arrive at a knowledge of its truth-value. If the sentence is effectively decidable, but not in practice, for example, because the number of calculations needed is too large, then I can at least manifest my knowledge of a decision method for it.

The problem appears, according to Dummett, when we come to sentences which are not effectively decidable. A sentence such as 'Two million days before the birth of Benjamin Franklin, an even number of birds landed on Mount St Helen' is not effectively decidable, for we have no method today which is guaranteed to give the right answer. In this case I do not have a capacity to get to know the truth-value, and therefore I cannot manifest my knowledge of its truth-conditions.

The sentence is clearly meaningful, and I clearly understand it. Since knowledge of meaning must be manifestable, and I cannot manifest my knowledge of bivalent truth-conditions in this case, knowledge of its meaning cannot be knowledge of bivalent truth-conditions. Hence, Dummett concludes, meaning does not in general consist of bivalent truth-conditions.

If this argument is sound, then the principle of bivalence cannot be justified in the most natural way, that is, directly from the proper semantics of the sentences it concerns, for that is not a semantics of bivalent truth-conditions. As Dummett points out, this does not in itself invalidate the bivalence principle, since the argument allows that bivalence may be justified in some other way. It seems plausible, however, that bivalence cannot be justified in any way allowed for by Dummett's meaning-theoretical principles.

If bivalence cannot be justified for some particular class of sentences, then neither can semantic realism for that class of sentences. And, since logical principles are perfectly general and topic neutral, if there is some class for which bivalence fails, then, under the default assumptions about the relation between bivalence and LEM (§3), LEM is not a logical law.

This is the destructive aspect of Dummett's enterprise. It has been criticized by a number of philosophers. One objection, due to McDowell (1987), is that the requirement of (full) manifestability is too strong, and another is that Dummett has a rigid and unjustified view of how knowledge of truth-conditions may be manifested.

The constructive aspect is a project of outlining a kind of semantics that satisfies Dummett's own requirements. The model for such a semantics, according to Dummett, is precisely the intuitionistic interpretation of the logical constants. The question is whether this idea can be generalized, in particular from logic and mathematics to

empirical sentences. Such a generalization would amount to some form of verificationism, that is, a view according to which the meaning of a sentence is somehow given by the method, or methods, of verifying the sentence. There are, however, many ways of verifying a verifiable sentence, and only some are directly connected with its meaning. In accordance with the intuitionistic interpretation of the logical constants we should therefore distinguish between canonical and non-canonical, or direct and indirect, verifications of a sentence. Then we can say that the meaning of a declarative sentence is given by what counts as a direct verification of it. On this model, understanding of a sentence would be manifested by exercising an ability to decide, for every proposed piece of evidence, whether or not it constitutes a verification of the sentence. The point is that we can be credited with such an ability even in cases where we in fact do not know how to find a verification. It is enough that we can recognize it if we find it. Therefore the requirement of manifestability is taken to be fulfilled.

However, it has been claimed by Wright (1987) that if full manifestability is required, then problems begin already with sentences not decidable in practice, for example, sentences whose shortest proofs are too long to be surveyable by humans. But, as Wright points out, a finite upper bound on the length of proofs cannot be acknowledged within intuitionism.

Dummett rejects such finite upper bounds, but in fact leaves it open whether intuitionistic logic meets all the meaning-theoretic requirements. He does not, however, see any *principled* reason for doubting it, as he does in the case of classical logic, with its accompanying realism.

See also: Antirealism in the philosophy of mathematics; Logical and mathematical terms, glossary of; Meaning and truth; Realism and antirealism; Realism in the philosophy of mathematics

PETER PAGIN

References and further reading

Brouwer, L.E.J. (1908) 'De onbetrouwbaarheid der logische principes', *Tijdschrift voor wijsbegeerte* 2: 152-8; trans. A. Heyting, 'The Unreliability of the Logical Principles', in *Collected Works*, vol. 1, *Philosophy and Foundations of Mathematics*, ed. A. Heyting, Amsterdam and Oxford: North Holland, 1975. (Brouwer's original proposal that intuitionistic logic is the set of logical principles which are valid according to general intuitionism.)

Brouwer, L.E.J. (1913) 'Intuitionisme en formalisme', *Wiskundig Tijdschrift* 9: 180-211; trans. A. Dresden, 'Intuitionism and Formalism', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge, MA: Cambridge University Press, 1964. (One of several papers on the nature of intuitionism and logic, originally read in 1912.)

Dummett, M. (1973) 'The Philosophical Basis of Intuitionistic Logic', in *Truth and Other Enigmas*, London: Duckworth, 1978. (This paper includes the canonical proof kind of interpretation of the logical constants and presents the antirealist argument.)

Dummett, M. (1976) 'What is a Theory of Meaning? (II)', in G. Evans and J. McDowell (eds) *Truth and Meaning: Essays in Semantics*, Oxford: Clarendon Press. (Dummett's classic paper on the theory of meaning, including a presentation of the antirealist argument.)

Dummett, M. (1977) *Elements of Intuitionism*, Oxford: Clarendon Press. (A thorough introduction to intuitionistic mathematics and logic. Densely written.)

Dummett, M. (1991) *The Logical Basis of Metaphysics*, London: Duckworth. (A comprehensive treatment of logic, meaning and metaphysics; partly technical, partly difficult.)

Gentzen, G. (1934) 'Untersuchungen über das logische Schliessen', *Mathematische Annalen* 39: 176-210; trans. 'Investigations into Logical Deduction', in *Collected Papers*, ed. M.E. Szabó, Amsterdam: North Holland, 1969. (Gentzen's presentation of natural deduction.)

Heyting, A. (1930) 'Die formalen Regeln der intuitionistischen Logik' ('The Formal Rules of Intuitionistic Logic'), in *Sitzungsberichte der Preußische Akademie der Wissenschaften*, Berlin. (Heyting's formalization of intuitionistic propositional logic.)

Heyting, A. (1934) *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie (Foundations of Mathematics. Intuitionism. Proof Theory)*, Berlin: Springer; repr. 1974. (Includes an excellent 25-page introduction to intuitionism, as known at the time.)

McDowell, J. (1987) 'In Defence of Modesty', in B. Taylor (ed.) *Michael Dummett: Contributions to Philosophy*, Dordrecht: Nijhoff. (A criticism of Dummett's requirements on meaning theories, including the requirements on manifestability. The book contains other good papers on Dummett, with extensive replies.)

McGuinness, B. and Oliveri, G. (eds) (1994) *The Philosophy of Michael Dummett*, Dordrecht: Kluwer. (A good collection of critical comments on Dummett, with extensive replies.)

Prawitz, D. (1965) *Natural Deduction: A Proof Theoretic Study*, Stockholm: Almqvist & Wiksell. (A classic study of properties of natural deduction systems for both classical and intuitionistic logic.)

Prawitz, D. (1974) 'On the Idea of a General Proof Theory', *Synthese* 27: 63-77. (Introduces the notion of a canonical argument in the context of defining the concept of a valid argument.)

Tennant, N. (1987) *Antirealism and Logic: Truth as Eternal*, Oxford: Clarendon Press. (A well-written and useful introduction to, and overview of, the issues, both philosophical and formal, and a defence of antirealism.)

Wright, C. (1987) 'Strict Finitism', in *Realism, Meaning and Truth*, Oxford: Blackwell. (This paper concerns among other things whether the strict finitist has not got equally good arguments against the intuitionist as the intuitionist has against classical logic. The book as a whole is to a great extent concerned with Dummett's projects, both the negative and the positive.)

Iqbal, Muhammad (1877-1938)

Muhammad Iqbal was an outstanding poet-philosopher, perhaps the most influential Muslim thinker of the twentieth century. His philosophy, though eclectic and showing the influence of Muslims thinkers such as al-Ghazali and Rumi as well as Western thinkers such as Nietzsche and Bergson, was rooted fundamentally in the Qur'an, which Iqbal read with the sensitivity of a poet and the insight of a mystic. Iqbal's philosophy is known as the philosophy of khudi *or Selfhood. Rejecting the idea of a 'Fall' from Eden or original sin, Iqbal regards the advent of human beings on earth as a glorious event, since Adam was designated by God to be God's viceregent on earth. Human beings are not mere accidents in the process of evolution. The cosmos exists in order to make possible the emergence and perfection of the Self. The purpose of life is the development of the Self, which occurs as human beings gain greater knowledge of what lies within them as well as of the external world. Iqbal's philosophy is essentially a philosophy of action, and it is concerned primarily with motivating human beings to strive to actualize their God-given potential to the fullest degree.*

1 Life

Muhammad Iqbal was born at Sialkot in India in 1877. His ancestors were Kashmiri Brahmans; his forefathers had a predilection for mysticism, and both his father, Nur Muhammad, and his mother, Imam Bibi, had a reputation for piety. An outstanding student, Iqbal won many distinctions throughout his academic career. He passed the intermediate examination from the Scotch Mission School in Sialkot in 1893 and then moved to the Government College in Lahore, where he graduated in 1897. In 1899 he obtained a master's degree in Arabic. Having been deeply influenced by Sir Thomas Arnold, the well-known scholar of Islam, while he was at the Government College, Iqbal followed his advice and proceeded to Trinity College, Cambridge in October 1905, graduating in 1907 having studied philosophy under J.M.E. McTaggart. His doctorate was taken at Munich University, with a thesis entitled *The Development of Metaphysics in Persia*.

On his return from Europe in July 1908, Iqbal took up the post of Professor of Philosophy and English Literature at the Government College in Lahore, and also began his law practice. In 1911 he gave up his teaching career because he felt that he had a message to deliver and could do it better if he adopted an independent profession such as law. However, he always remained interested in education and was associated with the Oriental College, the Government College and the Islamia College in Lahore, and with the Jami'a Millia in Delhi. During the sessions of the Round Table Conferences in London he worked on the various committees connected with educational reforms. In 1933 Iqbal, along with some others, was invited by the Afghan government to visit the country and advise the government and Kabul University on educational matters.

Iqbal also took an interest in the workings of the Muslim League, but did not participate actively in politics. During the period 1910-23 he tried instead to create political consciousness and bring about an awakening of Indian Muslims. In 1924 Iqbal became a member of the National Liberal League of Lahore; in 1926 he was elected a member of the Punjab Legislative Assembly, and in 1930 was elected president of the All-India Muslim League, where he delivered a historic address. He took part in the Second and Third Round Table Conferences held in London, and was most disappointed with the outcome. Iqbal was knighted on 1 January 1923. His last years were clouded with ill health. After his death in 1938 he was buried near the gate of the Badshahi Mosque in Lahore, with many attending and millions in mourning.

2 Works

A precocious youth, Iqbal began to write poetry at a very early age, and soon after he came to Lahore he became known through his participation in poetic symposia. As a young poet, he came under the influence of Mirza Dagh Dehalvi, one of the renowned exponents of Urdu poetry. An organization to which Iqbal was devoted all his life was the Anjuman-e-Himayat-e-Islam (Society for the Support of Islam). The annual sessions of the Anjuman fulfilled an acute emotional need of Indian Muslims and became national festivals. Iqbal read his poems regularly at these sessions, and in fact his poems were the main attraction for the thousands who flocked to Lahore, almost on an errand of pilgrimage, to see and hear him. It was at an Anjuman meeting in April 1911 that Iqbal read his famous 'Shikwa' (The Complaint), a poem which commands such a unique place in Urdu literature that Iqbal's fame could rest securely on it alone.

The publication of *Asrar-e-khudi (The Secrets of the Self)* in 1915 was a significant event. It was followed by *Rumuz-e-bekhudi (The Mysteries of Selflessness)*, which dealt with the development of the communal ego, in 1918; *Payam-e-mashriq (The Message of the East)*, Iqbal's answer to Goethe's *West-Östlicher Divan*, in 1923 (see Goethe, J.W. von); *Zabur-e-'ajam (The Persian Psalms)* in 1927; *Javid nama (The Pilgrimage of Eternity)*, Iqbal's magnum opus modelled on Dante's *Divine Comedy*, in 1932; *Musafir (The Traveller)* in 1934; *Bal-e-Jibril (Gabriel's Wing)* in 1935; *Zarb-e-Kalim (The Stroke of Moses)* and *Pas che bayad kard ai aqwam-i-sharq? (So What Should be Done, O Nations of the East?)* in 1936. *Armaghan-e-hijaz (The Gift of Hijaz)*, containing both Persian and Urdu verse, appeared posthumously in 1938. As well as his poetical works, Iqbal wrote three works in prose. *'Ilm-ul-iqtisad (The Study of Economics)*, which was the first book on political economy to be published in Urdu, appeared in 1903; Iqbal's doctoral thesis, *The Development of Metaphysics in Persia* was published in 1908; and his lectures, *The Reconstruction of Religious Thought in Islam*, were first published in 1930. Iqbal also wrote numerous articles in Urdu and English in various journals and newspapers.

3 Philosophy of the self

Iqbal's philosophy is often described as the philosophy of *khudi*, or the Self. For him, the fundamental fact of human life is the absolute and irrefutable consciousness of one's own being. For Iqbal, the advent of humanity on earth is a great and glorious event, not an event signifying human sinfulness and degradation. He points out that according to the Qur'an, the earth is humanity's 'dwelling-place' and 'a source of profit' to it. Iqbal does not think that having been created by God, human beings were placed in a supersensual paradise from which they were expelled on account of an act of disobedience to God. Pointing out that the term 'Adam' functions as the symbol of self-conscious humanity rather than as the name of an individual in the Qur'an, Iqbal describes the 'Fall' as a transition from 'a primitive state of instinctive appetite to the conscious possession of a free self, capable of doubt and disobedience' (Iqbal 1930: 85). For Iqbal, Adam's story is not the story of the 'First Man' but the ethical experience, in symbolic form, of every human being. Following the Qur'anic teaching that though human beings come from the earth, God's spirit has been breathed into them, Iqbal holds on the one hand that human beings are divinely created, and on the other hand that they have evolved from matter. Unlike dualists, Iqbal sees no impassable gulf between matter and spirit, nor does he see human beings as a mere episode or accident in the huge evolutionary process. On the contrary, the whole cosmos is there to serve as the basis and ground for the emergence and perfection of the Ego. Humanity's evolution has not come to an end, for the destiny of human beings lies 'beyond the stars'.

The purpose of life is the development of the Self. In order that they may achieve the fullest possible development, it is essential for human beings to possess knowledge. Following the Qur'an, Iqbal maintains that there are two sources of knowledge: the inner consciousness of human beings and the outer world of nature. Starting with the intuition of the Self, human beings become aware of the Not-Self, the confronting 'other' which provides a constant challenge for them. Nature, however, does not confront God in the same way as it confronts humanity, since it is a phase of God's consciousness. God is immanent since God comprehends the whole universe, but also transcendent since God is not identical with the created world. All life is individual. There is a gradually rising scale of selfness running from the almost inert to God who is the Ultimate Ego. God is not immobile nor is the universe a fixed product; God is constantly creative and dynamic and the process of Creation still goes on. The Qur'anic saying, 'Toward God is your limit' (Surah 53: 42), gives Iqbal an infinite worldview, and he applies it to every aspect of the life of humanity and the universe.

Iqbal distinguishes between two aspects of the Self, the efficient and the appreciative. The efficient self is that which is concerned with, and is itself partially formed by, the physical world. It apprehends the succession of impressions and discloses itself as a series of specific, and consequently numerable, states. The appreciative self is the deeper self, of which one becomes aware only in moments of profound meditation when the efficient self is in abeyance. The unity of the appreciative self is that in it, each experience permeates the whole. The multiplicity of its elements is unlike that of the efficient self. There is change and movement, but this change and movement are indivisible; their elements interpenetrate and are wholly non-serial in character.

Corresponding to the two aspects of the Self are the two levels of time, serial time and pure duration. Serial time is spatialized or clock time, whereas pure duration is a ceaseless continuous flow in which all things live and move and have their being. As human beings perfect their egohood, they cast off the girdle of serial time and gain a

measure of eternity.

Iqbal believes ardently that human beings are the makers of their own destiny and that the key to destiny lies in one's character. He constantly refers to the Qur'anic verse, 'Verily God will not change the condition of a people till they change what is in themselves' (Surah 13: 12). Humanity's mission on earth is not only to win greater freedom but also to gain immortality, which according to Iqbal 'is not ours by right; it is to be achieved by personal effort. Man is only a candidate for it' (Iqbal 1930: 119).

Though humanity is the pivot around which Iqbal's philosophy revolves, yet as pointed out by Schimmel, Iqbal's 'revaluation of Man is not that of Man qua Man, but of Man in relation to God' (Schimmel 1963: 382). Iqbal's Ideal Person is the Servant of God. The relation between humanity and God is a personal one; hence the great importance of prayer in the thought of Iqbal. The belief in the one living God gives humanity freedom from all false deities and fortifies it against forces of disintegration. Iqbal sees his concept of the Ideal Person realized in the Prophet of Islam, whose life exemplifies all the principles dearest to Iqbal's heart. In his view art, religion and ethics must be judged from the standpoint of the Self. That which strengthens the Self is good and that which weakens it is bad. Iqbal does not admit the absolute existence of evil but regards it as being necessary for the actualization of moral purpose as vital activity in the world. His Iblis (or *al-Shaytan*, Satan) is the counterpart to his Ideal Person (see Free will; Self, Indian theories of).

4 Epistemology

Iqbal cannot be easily or exclusively classified as an empiricist, rationalist or intuitionist since he combines sense-perception, reason and intuition in his theory of knowledge (see Epistemology in Islamic philosophy). He defines knowledge as 'sense-perception elaborated by understanding' ('understanding' here does not stand exclusively for 'reason' but for all non-perceptual modes of knowledge). There are two ways of establishing connections with the Reality that confronts us. The direct way is by means of observation and sense-perception; the other way is through direct association with that Reality as it reveals itself within.

Iqbal compares the classical spirit with its contempt for sense-perception with the empirical attitude of the Qur'an 'which sees in the humble bee a recipient of Divine inspiration and constantly calls upon the reader to observe the perpetual change of the winds, the alterations of day and night, the clouds, the starry heavens and the planets swimming through infinite space' (Iqbal 1930: 91) The cultures of the ancient world failed, says Iqbal, because their approach to reality was entirely introspective as they moved from within outwards. This gave them theory without power, and no durable civilization can be based on mere theory.

Iqbal distinguishes between 'logical understanding', which has a sectional nature and 'the deeper movement of thought', which is identical with intuition. He frequently points out (particularly in his poetry) the limitations of the former, but this does not mean that he was an anti-rationalist or anti-intellectual. Iqbal cites enthusiastically the Qur'anic verses (2: 28, 31) which state that Adam's superiority over angels lay in his power to 'name' things, that is, to form concepts. Concepts are not abstract logical entities: they are based on, and indissolubly linked with, facts of sensation. It is the knowledge of things and their inherent nature that exalted Adam over celestial creatures, and it is only through an unceasing struggle to attain the knowledge of things that humanity can maintain its superiority with justice in the world. Without discursive 'intellect' science would be impossible, and without science very little progress would be made in the material sphere. Iqbal believed strongly in the power and utility of science, but he did not regard science as the measure of all things. Science seeks to establish uniformities of experience, that is, the laws of mechanistic repetition, but does not take account of feelings, purposes and values. In Iqbal's opinion, the predicament of present-day humanity is that its life is wholly overshadowed by the results of its intellectual activity and it has ceased to live soulfully, or from within, having been cut off from the springs of life.

Like the existentialists, Iqbal sounds a warning that an idolatrous attitude towards reason and science leads in the direction of dehumanization. Since he puts great emphasis on intuition as a mode of knowledge, there has been much debate on the apparent conflict between reason and intuition in his works. However, Iqbal regarded reason and intuition as organically related and considered both to be necessary for the fulfilment of human destiny. Equating scientific knowledge with *'aql* (reason) and mystic knowledge with *'ishq* (love), Iqbal struggles constantly against separating the former from the latter since he believes that without love, reason becomes

demonic.

5 Political philosophy

Of all the parts of his thought, Iqbal's political philosophy is perhaps the most commonly misunderstood. This misunderstanding is largely the result of dividing his political philosophy into phases, such as the nationalistic phase, the pan-Islamic phase and the last phase in which he pioneered the Muslim independence movement. By regarding each phase as being quite different from and independent of the other phases, one almost always reaches the conclusion that either Iqbal's political views changed with astonishing rapidity or that he could not make up his mind and was inconsistent. One can indeed see Iqbal first as a young poet with rather narrow parochial sympathies which gradually widened into love of homeland, and then gave way to love of Islam which later became transformed into love of humanity. However, he can also be seen as a visionary, whose ideal from first to last was the realization of God's Kingdom on Earth, who believed in the interrelatedness, equality and freedom of human beings, and who strove at all times to achieve these goals; and by viewing Iqbal in this one light, one attains a much better understanding of his political philosophy.

Iqbal's interest in politics was secondary not primary. In his historic address at Lahore, in 1932, he made this clear:

> Politics have their roots in the spiritual life of Man. It is my belief that Islam is not a matter of private opinion. It is a society, or if you like, a civic Church. It is because present-day political ideals, as they appear to be shaping themselves in India, may affect its original structure and character that I find myself interested in politics.
>
> (Iqbal 1964: 288)

Iqbal's impact on the political situation of the Muslims in India was so great that he is hailed as the 'spiritual' founder of Pakistan. Undoubtedly, there was much focus on the Islamic community in his major works written between 1908 and 1938. Nevertheless, when accused by Lowes Dickinson of being exclusive in his thinking, Iqbal denied the allegation and said:

> The humanitarian ideal is always universal in poetry and philosophy, but if you make it an effective ideal…you must start, not with poets and philosophers, but with a society exclusive in the sense of having a creed and well-defined outline…. Such a society according to my belief is Islam. This society has so far proved itself a more successful opponent of the race-idea which is probably the hardest barrier in the way of the humanitarian ideal…. All men and not Muslims alone are meant for the Kingdom of God on earth, provided they say goodbye to their idols of race and nationality and treat one another as personalities. The object of my Persian poems is not to make out a case for Islam: my aim is simply to discover a universal social reconstruction, and in this endeavour, I find it philosophically impossible to ignore a social system which exists with the express object of doing away with all the distinctions of caste, rank and race.
>
> (Iqbal 1964: 98-9)

6 Critical evaluation

Iqbal undertook the task of uniting faith and knowledge, love and reason, heart and mind. In the case of a writer at once so prolific and so provocative, there is bound to be considerable controversy in the evaluation of the measure of his success. Some regard Iqbal's thought as mainly eclectic, while others regard it as exciting and original. But surely it is true that if any thinker has succeeded - to whatever degree - in the task of building a bridge between East and West, it is Iqbal.

One of the most important questions to be asked regarding Iqbal's work as a philosopher is, from what point of view is it to be judged? Inevitably the work of every philosopher must be subjected to the test of coherence and consistency. On the whole, Iqbal's philosophy sustains this test. Like any other philosopher he has first principles which seem to him self-evident, and which he therefore does not seek to defend. Like most other philosophers, there are times and places where he is not very clear or is evasive and unwilling to commit himself. Many writers have also seen a number of contradictions (which appear in some specific part of his thought, usually in his socio-political philosophy) disappear when viewed in the larger context of his total philosophy.

See also: Islamic philosophy, modern; Soul in Islamic philosophy

RIFFAT HASSAN

List of works

Iqbal, M. (1903) *'Ilm-ul-iqtisad (The Science of Economics)*, Lahore: Iqbal Academy, 1961.(Iqbal's first published book in Urdu was also the first book on economics to be published in Urdu, and contains a number of important socioeconomic ideas which were to become important elements of his philosophy of the individual and communal ego.)

Iqbal, M. (1908) *The Development of Metaphysics in Persia*, Cambridge: Cambridge University Press; Lahore: Bazm-Iqbal, 1964.(Iqbal's doctoral dissertation in which he traces the logical continuity of Persian thought.)

Iqbal, M. (1915) *Asrar-e-khudi (The Secrets of the Self)*, Lahore: Shaikh Ghulam Ali and Sons; trans. R.A. Nicholson, Lahore: Shaikh Muhammad Ashraf, 1950.(Iqbal's first volume of poetry in Farsi, in which he laid the foundations of his philosophy of life based on the mysticism of the struggle, the continuing endeavour of the self to fully actualize its potential.)

Iqbal, M. (1918) *Rumuz-e-bekhudi (The Mysteries of Selflessness)*, Lahore: Shaikh Ghulam Ali and Sons; trans. A. Arberry, London: John Murray, 1953.(This volume of Farsi poetry is concerned with the role of the individual in a community and forms the basis of Iqbal's social and political philosophy.)

Iqbal, M. (1923) *Payam-e-mashriq (The Message of the East)*, Lahore: Shaikh Muhammad Ashraf.(This volume of Farsi poetry was inspired by Goethe's *West-Östlicher Divan*, and highlights those social and religious ideas which have a bearing on the spiritual development of individuals and communities.)

Iqbal, M. (1924) *Bang-e-dara (The Sound of the Caravan Bell)*, Lahore: Shaikh Ghulam Ali and Sons.(Iqbal's first collection of poetry in Urdu, which contains some of his best known and most inspiring poems such as 'Shikwa' (The Complaint) and 'Jawab-e-shikwa' (The Response to the Complaint).)

Iqbal, M. (1927) *Zabur-e-'ajam (The Persian Psalms)*, Lahore: Shaikh Ghulam Ali and Sons.(In this volume of Farsi poetry, Iqbal's philosophy of love finds exquisite expression.)

Iqbal, M. (1930) *The Reconstruction of Religious Thought in Islam*, Lahore: Shaikh Muhammd Ashraf. (This is one of the most important works of philosophy in modern Islam, in which Iqbal formulates a new Muslim metaphysics in the light of the philosophical tradition of Islam and recent developments in various domains of human knowledge.)

Iqbal, M. (1932) *Javid nama (The Pilgrimage of Eternity)*, Lahore: Shaikh Ghulam Ali and Sons.(This volume of Farsi poetry is Iqbal's magnum opus, modelled on the Prophet's *me'raj* or ascension to the presence of God, which is a symbol of self-realization.)

Iqbal, M. (1933) *Musafir (The Traveller)*, Lahore: Shaikh Mubarak Ali.(This volume of Farsi poetry records Iqbal's visit to Afghanistan at the invitation of King Nadir Shah.)

Iqbal, M. (1935) *Bal-e-Jibril (Gabriel's Wing)*, Lahore: Shaikh Mubarak Ali.(Iqbal's second volume of poetry in Urdu is an acknowledged masterpiece in terms of its literary quality and its prophetic call to action.)

Iqbal, M. (1936) *Zarb-e-alim (The Stroke of Moses)*, Lahore: Shaikh Ghulam Ali and Sons.(This volume of Urdu poetry contains Iqbal's critique of the modern age in the light of the symbol of Moses, a seeker of God who insisted on seeing God.)

Iqbal, M. (1938) *Armaghan-i-hijaz (The Gift of Hijaz)*, Lahore: Shaikh Ghulam Ali and Sons.(This volume, containing both Farsi and Urdu poetry, was published posthumously and sums up what Iqbal believed to be the teaching of the Prophet from the Hijaz, the symbol of the desert which represents a life of struggle and austerity.)

Iqbal, M. (1948) *Speeches and Statements*, ed. Shamloo, Lahore: Al-Manar Academy. (Collection of shorter works on a variety of subjects.)

Iqbal, M. (1961) *Stray Reflections*, ed. J. Iqbal, Lahore: Shaikh Ghulam Ali and Sons.(Collection of Iqbal's reflections on a number of topics, ranging from the philosophical to the practical.)

Iqbal, M. (1964) *Thoughts and Reflections of Iqbal*, ed. S.A. Vahid, Lahore: Shaikh Muhammad Ashraf. (Contains Iqbal's historic addresses and significant writings on a number of subjects.)

References and further reading

Arnold, T.W. (1928) *The Islamic Faith*, London: Benn.(An important work on Islam by Iqbal's teacher, mentor and friend.)

Hassan, R. (ed.) (1977) *The Sword and the Sceptre*, Lahore: Iqbal Academy.(A collection of historic writings on the life and work of Iqbal.)

Hassan, R. (1979) *An Iqbal Primer*, Lahore: Aziz.(An introduction to Iqbal's philosophy, analysing the salient ideas in each of his works.)

Schimmel, A.M. (1963) *Gabriel's Wing*, Leiden: Brill.(By far the best treatment of Iqbal's poetic metaphysics.)

Sinha, S. (1947) *Iqbal, the Poet and his Message*, Allahabad: R.N. Lal.(Interesting discussion of Iqbal's poetic style and its philosophical relevance.)

Vahid, S.A. (1959) *Iqbal: His Art and Thought*, London: John Murray.(A comprehensive overview of Iqbal's literary and philosophical ideas by a noted Iqbal scholar.)

Irigaray, Luce (1930-)

Luce Irigaray holds doctorates in both linguistics and philosophy, and has practised as a psychoanalyst for many years. Author of over twenty books, she has established a reputation as a pre-eminent theorist of sexual difference - a term she would prefer to 'feminist'. The latter carries with it the history of feminism as a struggle for equality, whereas Irigaray sees herself more as a feminist of difference, emphasizing the need to differentiate women from men over and above the need to establish parity between the sexes.

Speculum de l'autre femme (1974) (Speculum of the Other Woman) (1985), the book that earned her international recognition, fuses philosophy with psychoanalysis, and employs a lyrical 'mimesis', or mimicry, that parodies and undercuts philosophical pretensions to universality. While adopting the standpoint of universality, objectivity and uniformity, the philosophical tradition in fact reflects a partial view of the world, one which is informed by those largely responsible for writing it: men. Without the material, maternal and nurturing succour provided by women as mothers and homemakers, men would not have had the freedom to reflect, the peace to think, or the time to write the philosophy that has shaped our culture. As such, women are suppressed and unacknowledged; femininity is the unthought ground of philosophy - philosophy's other.

1 Life and works

Born in Belgium, Luce Irigaray first studied literature, focusing on Paul Valéry, and her early training was at Louvain. She completed her Doctorate of Letters at the University of Paris VIII, and has lived in France since that time. Her thesis was later published in 1974 as *Speculum de l'autre femme* (1974) (*Speculum of the Other Woman*, 1985). In 1977 a collection of articles and interviews, *Ce sexe qui n'est pas un* (1977) (*This Sex Which is Not One*, 1985), appeared. Translations of both books were published in 1985, and these are the two early works for which she is still best known in the English-speaking world. She is better known in Italy, where she has established a large following, has contributed frequently to a Communist Party paper, and has engaged with public dialogues with Renzo Imbeni, an Italian politician to whom her book *J'aime à toi* (1992) is dedicated.

While at times she insists on the importance of her philosophical contribution, her work is also heavily indebted to Jacques Lacan, whose psychoanalytic theory remains an important influence in her more recent work, such as *Sexes et parentés* (1987) (*Sexes and Genealogies*, 1987) - despite the break with Lacan after the appearance of *Speculum* in 1974. Luce Irigaray is the Director of the Centre National de la Recherche Scientifique, where she established an international team to work on linguistic patterns in relation to gender, the focus of her most recent work.

In a 1994 interview (Hirsh and Olson 1995) Luce Irigaray divides her work into three periods. The first stage includes *Speculum of the Other Woman* and *This Sex Which is Not One*, and is characterized as a critique of the way in which 'a single subject, traditionally the masculine subject, had constructed the world and interpreted the world according to a single perspective' (see Subject, postmodern critique of the§2). The second stage, in which one might include *Sexes and Genealogies*, was one in which she defined 'a second subject', that of a 'feminine subjectivity'. The third phase, represented by*Je, tu, nous* (1990) (*Je, Tu, Nous: Toward a Culture of Difference*, 1993), is one in which she constructs 'an intersubjectivity respecting sexual difference'. Here I shall focus on the first phase of Irigaray's tripartite division, as the most overtly philosophical period of her work, noting only that the third phase contains some more accessible works such as *Je, tu, nous*. It is also the period during which Irigaray began to receive a more sympathetic hearing from critics, who in the 1980s tended to dismiss her, with Julia Kristeva, as an essentialist.

2 'Luce' and 'Irigaray'

Luce Irigaray prefers to be referred to by her given name, Luce, as well as her paternal name, Irigaray. This preference reflects several fundamental issues with which her work is centrally concerned, the most obvious of which is a rejection of the primacy of the name of the father, and an insistence upon the distinction between femininity and masculinity. By insisting on the difference between her first name and her surname, Luce Irigaray draws attention to the act of naming and its presuppositions, and calls for a thoughtful relation to the language and traditions that we often unthinkingly inhabit. At the same time, she reclaims authority from the paternal law that

governs language, refusing to capitulate to the dictates of a system of representation that has repressed and marginalized women. In doing so, she recreates and transforms the significance that we usually assume in the process of naming, in a gesture that is both negative and positive. On the one hand Luce Irigaray challenges the traditional place of women as defined by mainstream Western society, and on the other she offers a new interpretation of women's roles. This double gesture opens the way for a third contribution, namely the re-enactment of sexual difference as a relationship between men and women that is not governed by a privileging of masculinity.

3 *Speculum of the Other Woman*

In *Speculum*, Luce Irigaray provides close textual readings of philosophical figures ranging from Plato to Freud, and including Descartes, Hegel and Plotinus. In her own words, *Speculum* treats 'the problem of the Other as woman in Western culture'. In 1985, along with *This Sex Which is Not One*, the English translation of *Speculum* appeared, marking the beginning of what has proved to be a turbulent relationship between Luce Irigaray and her critics. One reason for the controversial reception her work has elicited from English-speaking readers is the difficulties her writing - which is full of word play - poses for translators. By the word 'speculum' Luce Irigaray intended an allusion not only to 'the mirror in a simple sense', in which 'one sees oneself', but also to works that speak of '"*speculum mundi*" - that is, the "mirror of the world"'. Thus, says Luce Irigaray, at issue is how 'it's possible to give an account of the world within a discourse'.

The problems of translation that Luce Irigaray's work precipitates are not limited to her sometimes elliptical allusions. The translation of *Speculum: de l'autre femme* as *Speculum of the Other Woman* led to the assumption, says Luce Irigaray, that she was concerned with 'the image of the other woman', or with 'an empirical relation between two women, for example. This is absolutely not the project of *Speculum*.… it should have been put, *Speculum* on *the Other Woman* or *On the Other: Woman*. That would have been best' (Hirsh and Olson 1995).

By documenting the ways in which diverse texts of the Western philosophical tradition adhered to a systematic representation of women as other, Luce Irigaray demonstrated the sense in which the tradition achieved a certain unity and coherence on the basis of that which it excluded: women. Disparate in other respects, the philosophers she explores in *Speculum* share in common their dismissal, neglect and denigration of women.

Deploying a strategy of mimicry, imitating the methods of the very texts she examines, Luce Irigaray exhumes what remains of women, once the tradition has had its say, putting together the disparate parts, the shattered remnants, in order to create for the first time an image of women that goes beyond the mirroring function they have served in order to 'constitute a masculine subject' and knowledge.

Thus, for example, she retrieves the tragic and mythical figure of Antigone from the hands of Hegel ([1807] 1952) not by repudiating from the start the Hegelian analysis of this Sophoclean heroine, but precisely by taking his analysis to its limits, pushing the boundaries of his own presuppositions until they pass into their opposite, and yielding conclusions that contradict Hegel's own premises. Thus Hegel, the master of dialectic, is dialectically forced to say what he both says and cannot say: that women are 'the eternal irony of the community', and that their disruption of the ethical system of the state is to be celebrated and desired, rather than contained and repressed. Luce Irigaray thus sublates the conclusions of his absolute knowledge. She perfects and masters Hegel's own methods, even as she undermines and erases their results.

Antigone represents for Hegel the purest tragic figure, but she is not conscious of the meaning of her own deed. Hegel will grant Antigone only an immediate and intuitive understanding of ethical action, denying her the ability to reflect rationally on the significance of her action in the state as a whole. He thereby privileges the community over the individual, rationality over intuition, and abstract political duty over the exigencies of body, blood, and kinship. He upholds Creon's male claim to represent the *polis*, over Antigone's female adherence to the only tasks that ancient Thebes allows her as a woman - the care of the body and allegiance to family. Insisting on performing the rites of burial for her traitorous brother, Antigone will care for her family members in death, as in life. But whereas Hegel, along with Creon, upbraids her for this, Luce Irigaray rejoices in her refusal to obey an edict that forbids her to do the only task that is truly hers as a woman - banned from public life as her sex is.

4 Lips that speak together

Luce Irigaray is best known for the image that she presents in her essay 'When Our Lips Speak Together' (in *Ce sexe qui n'en est pas un*, 1977). Read as an ironic counterpart to Lacan's 'phallus', Luce Irigaray's view of the subject is what one critic calls 'lipcentric', but in a way that seeks to uncover multiple 'phalluses' rather than make any foundational or univocal claims for its own privilege. Luce Irigaray's work creates a 'feminine imaginary' to counteract that of the dominant culture, a masculine imaginary. A feminine or female imaginary can be understood both as the shards or fragments that remain of woman in the Western philosophical unconscious, and the creative space that these open for future re-mapping of women's identification. Luce Irigaray's texts flout academic conventions, insisting on a plurality of meanings, and blurring the boundaries separating commentary from the original texts she interprets, in order to make the point that the authority of the canon often derives from a source of unacknowledged sustenance, namely the work of women.

TINA CHANTER

List of works

Irigaray, L. (1974) *Speculum de l'autre femme*, Paris: Éditions de Minuit; trans. G.C. Gill, *Speculum of the Other Woman*, Ithaca, NY: Cornell University Press, 1985. (Exposes the phallogocentric mythology pervading philosophical texts of the Western tradition from Plato to Freud.)

Irigaray, L. (1977) *Ce sexe qui n'en est pas un*, Paris: Éditions de Minuit; trans. C. Porter with C. Burke, *This Sex Which is Not One*, Ithaca, NY: Cornell University Press, 1985.(A collection of essays and interviews expounding the ways in which women have functioned as objects of exchange among a male economy, as the unacknowledged resource that supports a phallic symbolic, and suggesting an alternative view of interaction both among women and between women and men.)

Irigaray, L. (1980) *Amante marine. De Friedrich Nietzsche*, Paris: Éditions de Minuit; trans. G. Gill, *Marine Lover of Friedrich Nietzsche*, New York: Columbia University Press, 1991.(A study of the ways in which Nietzsche's work operates by excluding one of the four elements, water.)

Irigaray, L. (1982) *Passions élémentaires*, Paris: Éditions de Minuit; trans. J. Collie and J. Still, *Elemental Passions*, New York: Routledge, 1992.(A poetic meditation on the role of earth, air, fire and water, emphasizing the fluidity of these categories.)

Irigaray, L. (1983) *L'Oubli de l'air: chez Martin Heidegger*, Paris: Éditions de Minuit.(A study of Heidegger's work in the light of his repression of air as an element, which is sacrificed in favour of the other elements.)

Irigaray, L. (1984) *Ethique de la différence sexuelle*, Paris: Éditions de Minuit; trans. C. Burke and G. Gill, *An Ethics of Sexual Difference*, Ithaca, NY: Cornell University Press, 1993.(Irigaray's most philosophical work since Speculum, containing essays on Plato, Aristotle, Merleau-Ponty and Levinas, among others.)

Irigaray, L. (1985) *Parler n'est jamais neutre*, Paris: Éditions de Minuit.(Argues against the neutrality of language in relation to gender, and includes discussion of Derrida among others.)

Irigaray, L. (1987) *Sexes et parentés*, Paris: Éditions de Minuit; trans. G. Gill, *Sexes and Genealogies*, New York: Columbia University Press, 1993.(Psychoanalytic studies of feminine sexuality in relation to myth and religion.)

Irigaray, L. (1989) *Le Temps de la différence*, Paris: Éditions Générale Française; trans. A. Martin, *Thinking the Difference*, New York: Routledge, forthcoming.(Accessible essay on cultural and political measures that need to be taken in order to ensure the recognition of sexual difference.)

Irigaray, L. (1990) *Je, tu, nous*, Paris: Grasset & Fasquelle; trans. A. Martin, *Je, Tu, Nous: Toward a Culture of Difference*, London: Routledge, 1993.(Comprised of essays first presented in oral form, this brief text provides a useful introductory text to Irigaray's recent thought.)

Irigaray, L. (1991) *The Irigaray Reader*, ed. M. Whitford, Oxford: Blackwell.(Offers a collection of representative essays and abridged versions of texts from Irigaray's early and middle periods.)

Irigaray, L. (1992) *J'aime à toi: esquisse d'une félicité dans l'histoire*, Paris: Bernard Grasset; trans. A. Martin, *I Love To You*, London: Routledge, 1996.(Explores positive ways in which men and women can establish relationships based upon respect for each other in their difference.)

References and further reading

Berg, M. (1991) 'Luce Irigaray's "Contradictions": Poststructuralism and Feminism', *Signs* 17 (1).(One of the best discussions of Luce Irigaray, particularly good on the relationship between the phallus and the figure of the lips.)

Chanter, T. (1995) *Ethics of Eros: Irigaray's Rewriting of the Philosophers*, London: Routledge.(Concentrates on Luce Irigaray's relation to de Beauvoir, Derrida, Hegel, Heidegger and Levinas.)

Hegel, G.W.F. ([1807]1952) *Phänomenologie des Geistes*, ed. J. Hoffmeister, Hamburg: Felix Meiner; trans. A.V. Miller, *Phenomenology of Spirit*, Oxford: Clarendon Press, 1979.(Classic text in the history of philosophy which includes a discussion of Sophocles' *Antigone*, the text that forms the basis of Irigaray's essay on women as, in Hegel's phrase, 'the eternal irony of the community' in *Speculum*.)

Hirsh, E. and Olson, G.A. (1995) '"Je - Luce Irigaray": A Meeting with Luce Irigaray', *Feminist Ethics and Social Policy*, Part II, special issue of *Hypatia* 10 (2).(A useful interview covering the tripartite division of Luce Irigaray's work and the difficulty of translation.)

Whitford, M. (1991) *Luce Irigaray: Philosophy in the Feminine*, London and New York: Routledge.(The first monograph on Luce Irigaray's work to appear in English; a comprehensive and scholarly discussion of her work, focusing on its psychoanalytic aspects. Very good bibliography.)

Isaac of Stella (d. *c*.1177)

Like other twelfth-century Cistercians, Isaac of Stella was well versed in secular learning. Centrally engaged with the contemplative life, he expresses his spiritual insights in terms of the science of his day, and combines a spiritual psychology derived from Johannes Scottus Eriugena and Hugh and Richard of St Victor with an anthropology grounded in Stoic physics, Greek and Arab medicine, and a cosmic model derived from Plato's Timaeus. A unifying theme of his writings is the relation between the physical and spiritual dimensions of human experience.

Born in England in the early twelfth century, Isaac studied in France. Though he speaks of the boldness and brilliance of his teachers, he mentions no names, but his writings reveal the influence of Thierry of Chartres, Gilbert of Poitiers and William of Conches, and indicate a solid grounding in the liberal arts and the writings of secular authors. He was also throughly versed in the Augustinian tradition represented by Hugh of St Victor (see Augustinianism), and in Greek patristic thought as transmitted by Eriugena. He had presumably been a monk for some time when, in 1147, he became abbot of the monastery of l'Étoile (Latin *Stella*), near Poitiers, which had recently been joined to the Cistercian Order. After twenty years during which he helped to found further Cistercian houses, and apparently aided Thomas Becket in his exile from England in the mid-1160s, Isaac removed, probably in 1167, to the island of Ré near La Rochelle. Here he assisted in founding a new monastery, Notre-Dame des Chateliers, where he evidently remained until his death around 1177.

Isaac's secular training appears plainly in his sermons, which are in effect brief theological essays dealing, frequently in technical language, with human nature, the relation of God to the universe, the Trinity and the stages of the soul's ascent to God. The distinctive features of his thought are synthesized in his *Epistola de anima (Letter on the Soul)*, a short treatise which considers the soul as an intermediary between the material world and the divine. Like Plato in the *Timaeus*, Isaac develops the idea that human nature in its complexity mirrors the universe at large, but a human being is more than a 'lesser universe', for the human soul has the power to comprehend its cosmic environment and consider it in relation to the divine Wisdom it manifests. The soul is the eye of the cosmic body, and at the same time bears a unique relation to God. Its three fundamental faculties, the rational, concupiscible and irascible (*rationabilitas*, *concupiscibilitas*, *irascibilitas*), correspond to the Persons of the Trinity conceived as Wisdom, Love and Power.

Set between God and the material world, the soul, sustained and inspired by divine grace, has the capacity to resolve the opposition between body and spirit. *Rationabilitas*, in which the soul resembles the divine Wisdom, and *conscupiscibilitas*, which emulates the all-embracing love of the Holy Spirit, reach out to the created world by means of the senses, whose perceptions are adapted to spiritual understanding by reason and imagination. When intimately engaged with the material world, the soul's imagining power (*phantasticum animae*) conceives images that are 'almost bodily', while sensory imagination for its part can form images of an 'almost spiritual' purity. At this level, Isaac declares, drawing on the Stoic concept of a 'fiery' power which informs physical and mental activity at all levels (see Stoicism), bodily and spiritual perception can collaborate, 'joined in a personal union without confusion of nature'. This linkage, through which material reality becomes accessible to the divine faculties of the soul, is analogous to the relationship whereby the soul itself bears the likeness of God and has the capacity to be 'taken up even into personal oneness' with him. From this analogy, Isaac proceeds to consider other ways in which the 'likeness' between physical and spiritual allows perception to transcend the limits of bodily existence by bringing all levels of our nature into collaboration with 'the gift of creating grace'.

Though the circulation of Isaac's writings seems to have been limited, the *Letter on the Soul* had a considerable influence through its contribution to the widely circulated treatise *On the Spirit and the Soul*. This compilation, wholly unoriginal and occasionally confused, nonetheless exercised a wide influence, partly because of its attribution to Augustine, and brought elements of Isaac's spiritual psychology to bear on the *summae* of the thirteenth-century pioneers of systematic theology, as well as the contemplative writers of the later Middle Ages.

See also: Chartres, School of; Soul, nature and immortality of

WINTHROP WETHERBEE

List of works

Isaac of Stella (before 1177) *Opera (Works)*, ed. B. Tissier, Bibliotheca Patrum Cisterciensum, Bonnefontaine, 1664; repr. in J.-P. Migne (ed.) Patrologia Latina 194, Paris, cols. 1689-1786.(Edition of collected works.)

Isaac of Stella (before 1177) *Epistola de anima (Letter on the Soul)*; trans. B. McGinn, *Three Treatises on Man: A Cistercian Anthropology*, Cistercian Fathers Series 24, Kalamazoo, MI: Cistercian Publications, 1977, 155-77. (This collection also contains the anonymous treatise *On the Spirit and the Soul*, 181-288.)

Isaac of Stella (before 1177) *Sermones (Sermons)*, ed. A. Hoste, G. Raciti and G. Salet, Sources Chrétiennes nos. 130, 207, 339, Paris: Éditions du Cerf, 1967-87, 3 vols.(With French translation. English versions of Sermons 1-26 appear in *Isaac of Stella: Sermons for the Christian Year*, vol. I, trans. H. McCaffrey, introduction by B. McGinn, Cistercian Fathers Series 11, Kalamazoo, MI: Cistercian Publications, 1979.)

References and further reading

McGinn, B. (1969) *The Golden Chain. A Study in the Theological Anthropology of Isaac of Stella*, Cistercian Studies Series 15, Washington, DC: Cistercian Publications, Consortium Press. (Places Isaac's writings in the context of twelfth-century thought and learning.)

Raciti, G. (1971) 'Isaac de l'Étoile', in *Dictionnaire de Spiritualité*, Paris: Beauchesne, vol. 7, cols. 2011-38.(A careful review of the evidence for Isaac's life and work, and a lucid summary of his theology and anthropology.)

Islam, concept of philosophy in

There is no generally accepted definition of what Islamic philosophy is, and the term will be used here to mean the sort of philosophy which arose within the culture of Islam. There are several main strands to Islamic philosophy. Peripatetic philosophy follows broadly the Greek tradition, while Sufism uses the principle of mystical knowledge as its leading idea. Some would argue that Islamic philosophy has never lost its concentration on the Qur'an and other significant Muslim texts, and that throughout its history it has sought to understand the essence of the realities both of the Sacred Book and of the created world. The decline of Peripatetic philosophy in the Islamic world did not mean the decline of philosophy as such, which continued to flourish and develop in other forms. Although it is sometimes argued that philosophy is not a proper activity for Muslims, since they already have a perfect guide to action and knowledge in the Qur'an, there are good reasons for thinking that Islamic philosophy is not intrinsically objectionable on religious grounds.

1 Nature and origins of Islamic philosophy

One of the interesting features of Islamic philosophy is that there is controversy as to what it actually is. Is it primarily the sort of philosophy produced by Muslims? This is unsatisfactory, since many Muslims who work as philosophers do not deal with Islamic issues in their philosophical work. Also, there are plenty of philosophers who are not Muslims and yet whose work is clearly in the area of Islamic philosophy. Could we call 'Islamic philosophy' philosophy which is written in Arabic? Certainly not, since a great deal of Islamic philosophy, perhaps the majority of it, is written in other languages, in particular Persian. Is Islamic philosophy then philosophy which examines the conceptual features of specifically Islamic issues? Not necessarily, since there are many thinkers whose work on logic and grammar, for example, is part of Islamic philosophy, even though there is no direct religious relevance in their work. Some commentators have tried to develop a central agenda which everyone who can be called an Islamic philosopher must share; they then have the difficulty of fitting everything in Islamic philosophy into that framework, a task which ultimately tends to fail (Leaman 1980). Perhaps the best way of specifying the nature of Islamic philosophy is to say that it is the tradition of philosophy which arose out of Islamic culture, with the latter term understood in its widest sense.

When did Islamic philosophy start? This is also a difficult question to answer, since from the early years of Islam a whole variety of legal and theological problems arose which are clearly philosophical, or at least use philosophical arguments in their elucidation. For example, there were heated debates about the acceptability of anthropomorphic language to describe the deity, and about the roles of free will and determination in the lives of human beings. Philosophy in its fullest sense began in the third century of the *hijra*. (The *hijra* was in 622 AD, when the Prophet Muhammad moved to Medina and set up a political community there; it is the first year according to the Muslim calendar, represented as AH 1.) The supremacy of the Abbasids over the Umayyads had led to an eastward movement of the Islamic empire, with the capital moving from Damascus to Baghdad. By this time also, Islam dominated such areas as Egypt, Syria and Persia, all places which were thoroughly immersed in Greek culture. The new rulers sought to apply the learning which existed in the empire to their own purposes. Much of this knowledge was very practical, being based on medicine, astrology, astronomy, mathematics and engineering. The caliph al-Ma'mun founded in Baghdad the *bayt al-hikma*, the House of Wisdom, in AH 217/AD 832, which served as an observatory and, more importantly, as a library and centre for the translation of Greek texts into Arabic. Many of the translators were Christians, who translated texts first from Greek into Syriac and then into Arabic (see Greek philosophy: impact on Islamic philosophy). In addition to the influence of the many translations of Greek texts, there was also an important transmission of Indian and Persian literature into Arabic, which undoubtedly had an influence on the development of Islamic philosophy.

It should not be thought that these translations were uncontroversial. Many Muslims questioned the necessity for Muslims to study philosophy at all. After all, Islam presents a complete practical and theoretical model of the nature of reality, and the 'first sciences' of the Greeks often seemed unnecessary and even opposed to Islam. Muslims had not only the Qur'an to help them regulate their lives and theoretical queries, they had also the *hadith*, the traditional sayings of the Prophet and the righteous caliphs (his immediate successors and companions) and the *sunna*, the practices of the community. There was further the system of *fiqh*, Islamic law, which discussed particular problems concerning how Muslims ought to behave, and the science of grammar, which explained how

the Arabic language ought to be understood. There was also by this time a well-developed system of *kalam*, theology, which dealt with the less obvious passages of the Qur'an, and which sought conceptual unity in apparent difficulties arising from the combinations of different canonical texts (see Islamic theology). What need was there then for the sort of philosophy which existed in Greek, which originated with non-Muslims and was initially transmitted into Arabic by non-Muslims?

This would not have been such a heated issue had philosophy not seemed to be so antagonistic to Islam on so many points. The philosophy that was transmitted into Arabic at this time was profoundly Neoplatonic (see Neoplatonism in Islamic philosophy); it tended to agree with Aristotle (§16) that the world is eternal, that there is a hierarchy of being with the intellect at the summit and the world of generation and corruption at the bottom, and recommended a rather ascetic system of ethics (see Neoplatonism §3). Even more crucial was the criterion of validity which the philosophers used. This was based on reason, as opposed to revelation, and naturally brought into question the significance of religious revelation. Thus philosophy came to be seen not so much as an alternative formulation of religious truths but as a rival and competing system of thought, one which required opposition by Islam. Those Muslims who worked as philosophers had to justify themselves, and they did so in a number of ways.

The first philosopher of the Arabs, al-Kindi, tended to argue that there is no basic inconsistency between Islam and philosophy, just as there is no basic inconsistency between Plato and Aristotle. Philosophy helps the Muslim to understand the truth using different techniques from those directly provided through Islam. Once philosophy became better established, however, it managed to sever the link with religion altogether, as we can see from al-Farabi onwards. Religion is then taken to represent the route to truth available to the unsophisticated and simple believer; when compared to philosophy it is seen as a version of the truth, albeit perhaps of poorer conceptual quality. The most determined defender of this view is undoubtedly Ibn Rushd (Averroes), with whom this form of philosophy largely came to an end in the sixth century AH (twelfth century AD).

2 *Falsafa* and *hikma*

Peripatetic philosophy in the Islamic world came to have considerable importance for a fairly limited period, from the third to sixth centuries AH (ninth to twelfth centuries AD). Sometimes the distinctness of this form of reasoning from traditional Islamic methodologies was emphasised by the use of the term *falsafa*, an Arabic neologism designed to represent the Greek *philosophia*. Often, however, the familiar Arabic term *hikma* was used. *Hikma* means 'wisdom', and has a much wider meaning than *falsafa*. A good deal of *kalam* (theology) would be classed as *hikma*, as would mysticism or Sufism (see Mystical philosophy in Islam §1). Whereas much *falsafa* is defined as the knowledge of existents, wider conceptions of the discipline tend to use the term *hikma*. Al-Suhrawardi, the creator of illuminationist philosophy, called it *hikmat al-ishraq*, a title which was taken up later by Mulla Sadra, and which is often translated in English as theosophy (see Illuminationist philosophy; Theosophy). This sort of philosophy involves study of reality which transforms the soul and is never really separated from spiritual purity and religious sanctity.

Philosophy as *hikma* has the advantage of referring to a wide range of conceptual issues within Islam. Philosophy can then deal both with the exoteric aspects of the Qur'anic revelation and the esoteric dimensions which lie at the heart of religion. Both the Qur'an and the universe are often viewed as aspects of divine revelation which require interpretation, and philosophy in its widest sense has a vital role here. Western commentators have tended to overemphasize the Greek background of Islamic philosophy, yet most of the major Islamic philosophers wrote extensively on the Qur'an and saw the role of philosophy as lying chiefly in the hermeneutic investigation of holy texts. This is particularly the case with the philosophers in Persia and India, who continued the philosophical tradition after it largely came to an end in its Peripatetic form. Islamic philosophy is then essentially 'prophetic philosophy', since it is based on the interpretation of a sacred text which is the result of revelation. It deals with human beings and their entelechy, with the One or Pure Being, and the grades of the universal hierarchy, with the universe and the final return of all things to God. An important aspect of this view is that it sees Islamic philosophy not as a transitory phenomenon but as a continuing tradition in the Islamic world, not as something largely imported from an alien culture but as an essential aspect of Islamic civilization.

A good example of this wider notion of philosophy lies in the controversy over the 'oriental philosophy' (*al-hikmat al-mashriqiyya*) of Ibn Sina (Avicenna). Ibn Sina is well known as a creator of a Peripatetic

philosophical system, one which came to have considerable significance within both Islamic and Western philosophy. His book *Mantiq al-mashriqiyyin (Logic of the Orientals)* deals largely with logical differences between him and Aristotle, but also includes a reference to other of his own works in which he claims to have gone in an entirely different direction from that of the Peripatetic (*mashsha'i*) thinkers. This book is not extant; perhaps the *Mantiq* is the first part of it. From what we find in his surviving works, a picture of the 'oriental philosophy' can be constructed. The Aristotelian universe becomes transformed, reason is linked to the intellect, the external universe becomes interiorized, facts become symbols and philosophy itself becomes a type of gnosis or *sophia*. The aim of philosophy is not only the theoretical knowledge of the substances and accidents of the universe, but also the experience of their presence and instantiation in such a way as to enable the soul to free itself from the confines of the universe. The universe is experienced not as something external to be understood but rather as a succession of stages along a path on which one is travelling. The notion of this 'oriental philosophy' has played an important part in the development of future illuminationist and Sufi forms of philosophy which not only seek to understand the universe rationally but also analyse the wonder we feel when we contemplate the divine mystery of that universe.

An advantage of seeing Islamic philosophy as broadly *hikma* rather than as the more narrow *falsafa* is that it avoids the danger of regarding it as predominantly an unoriginal and transmitted form of thought. This has often been the form of interpretation favoured by Western commentators, who are interested in seeing how originally Greek (and sometimes Indian and Persian) ideas reach the Islamic world and then form part of alternative systems of philosophy. There is no doubt that an important part of Islamic philosophy does follow this path, and the study of it is perhaps more appropriately a part of the history of ideas than of philosophy. Yet it should not be forgotten that by far the larger part of Islamic philosophy does not deal with the concerns of Peripatetic philosophy as such, but is firmly directed to the issues which arise within the context of an Islamic perspective on the nature of reality. Peripatetic philosophy, *falsafa*, may well enter this process, but it is far from the uncritical application of Greek ideas to Islamic issues. Although the central principles of *falsafa* have their origin in Greek philosophy, they were so radically transformed and developed within Islamic philosophy that there is no justification in thinking that the latter is merely a result of the transmission of ideas from outside Islam.

3 Heresy and the decline of Peripatetic philosophy

A highly influential attack on the role of philosophy as part of Islam was carried out by al-Ghazali in his *Tahafut al-falasifa (The Incoherence of the Philosophers)*. According to al-Ghazali, the Peripatetic philosophers (he was thinking in particular ofIbn Sina) present as truths theses which are often either heretical (*kufr*) or innovatory (*bid'a*). One might have expected him to go on to argue that these philosophical theses are therefore unacceptable on those grounds alone, but he does not do so. Instead, al-Ghazali criticizes these theses because, he argues, they do not follow from the arguments which the philosophers themselves give. These arguments are philosophically weak, and so need not be accepted. It is a happy consequence of the failure of these arguments that the principles of Islam are seen to rest on solid rational principles, at least in the sense that their contraries are untrue. Although al-Ghazali is often regarded as the archenemy of philosophy, it is evident on closer inspection of many of his texts that he himself seems to adhere to many of the leading principles of Ibn Sina's thought. Also, in common with many other opponents of philosophy, he had a high regard for logic (which was regarded as a tool of philosophy rather than a part of it) and insisted on the application of logic to organized thought about religion. Some opponents of philosophy such as Ibn Taymiyya even went so far as to criticize logic itself, but on the whole as the Peripatetic tradition of philosophy declined in the Sunni Islamic world it nonetheless entered other areas of Islamic life, such as theology and jurisprudence, and continued to have an influence until it was revived within the last century as part of the Islamic renaissance (*nahda*) (see Islamic philosophy, modern §1).

Philosophy persisted in the Shi'i world far more easily, and there has been a continuing tradition of respect for philosophy in Persia and other Shi'i communities up to today. Sunni Muslims tend to accept that the door to *ijtihad* (independent judgment) is now closed, and we must seek resolution of any theoretical and practical difficulties by referring to a series of canonical texts and to the consensus of the community. Shi'i Muslims appeal also to the authority of the imams, and especially in the case of some Shi'is to the 'hidden' or twelfth imam, as being in the continuing line of religious descent from the Prophet and from his son-in-law Imam 'Ali. Since the bases of religious authority are more fluid for the *shi'a*, it tends to be more receptive to philosophy than is the case with the Sunnis. The legal definitions of what constitutes heresy and unbelief are sometimes much looser (as with the

Isma'ilis, for example), and the openness to a diversity of ideas and approaches has marked many Shi'i communities and countries. While Peripatetic philosophy went into a sharp decline in the Sunni world after the sixth century AH/twelfth century AD, it continued as part of a variety of philosophical approaches in the Shi'i world, either on its own or combined with elements of illuminationist (*ishraqi*) philosophy, and developed into more and more complex theoretical systems. Of course, philosophy continued to flourish in both the Sunni and Shi'i worlds in the sense of mystical philosophy or Sufism, which has been a persistent aspect of Islamic philosophy throughout its life.

4 Reason and revelation

Many Western commentators on Islamic philosophy take the conflict between reason and revelation as its central issue. This is often symbolized as the struggle between Athens and Jerusalem, or between philosophy and religion. While this is far too crude to be an accurate description, it does raise an important issue which has been discussed ever since Islamic philosophy began and which is still a live issue today in the Islamic world. If revelation tells believers everything they need to know, why bother to explore the same topics with reason? There are a number of answers to this question. First of all, the Qur'an itself speaks not only to Muslims, but to everyone who is able to read and understand it. It constantly urges the reader to consider rationally the evidences for Islam, and so places a high value on reason (Leaman 1985). This is not to suggest that there is no role for faith, nor that faith will not be necessary at some stage in order to approach God, but the Qur'an does offer rational indications of the truth of what it is advocating in terms of signs and proofs. This is certainly not an argument for free enquiry in the modern sense of the term, but it is an approach which places high value on the notion of independent reason, which might be seen as sympathetic also to the practice of philosophy itself.

According to Islam, the Prophet Muhammad is the last prophet. This implies that from that time on, no messenger can claim divine authority. We are reliant upon the correct interpretation of the *ayat* (signs) in both the Qur'an and in the universe. The ending of prophecy means that God expects human beings to use their reason to seek to understand the nature of reality, albeit reason which is guided by the principles of Islam. As the Qu'ran has it, 'We will show them our signs in all the parts of the earth and in their own souls, until they clearly see what is true' (Surah 41: 53). It is not as though there is competition between prophecy and philosophy, since the latter should be seen as supplementing and explaining the former. There are good grounds, then, for thinking that there is no basic incompatibility between the pursuit of reason and the pursuit of religion, at least not in Islam.

See also: Aesthetics in Islamic philosophy; Aristotelianism in Islamic philosophy; Causality and necessity in Islamic thought; Epistemology in Islamic philosophy; Ethics in Islamic philosophy; Faith; Hellenistic philosophy; Illuminationist philosophy; Islamic fundamentalism; Islamic theology; Logic in Islamic philosophy; Meaning in Islamic philosophy; Islamic philosophy, modern; Mystical philosophy in Islam; Neoplatonism; Law, Islamic philosophy of; Science in Islamic philosophy; Platonism in Islamic philosophy; Political philosophy in classical Islam; Revelation; Greek philosophy: impact on Islamic philosophy

OLIVER LEAMAN

References and further reading

Akhtar, S. (1995) 'The Possibility of a Philosophy of Islam', in S.H. Nasr and O. Leaman (eds) *The History of Islamic Philosophy*, London: Routledge, 1162-9. (Defence of the view that there is scope for philosophy within an Islamic context.)

Corbin, H. (1993) *History of Islamic Philosophy*, trans. L. Sherrard, London: Kegan Paul International.(Very important defence of the significance of mysticism and Persian thought as key aspects of Islamic philosophy.)

al-Ghazali (1095) *Tahafut al-falasifa (Incoherence of the Philosophers)*, in S. Van Den Bergh (trans.) *Averroes' Tahafut al-tahafut*, London: Luzac, 1978.(The classic attack on Peripatetic philosophy in Islam.)

Leaman, O. (1980) 'Does the Interpretation of Islamic Philosophy rest on a Mistake?', *International Journal of MidEastern Studies* 12: 525-38.(Critique of the view that all Islamic philosophy is about the reason versus revelation issue.)

Leaman, O. (1985) 'Introduction', in O. Leaman (ed.) *An Introduction to Medieval Islamic Philosophy*, Cambridge: Cambridge University Press, 1-21.(Account of the relationship between Islam and rationality.)

Leaman, O. (1992) 'Philosophy vs. Mysticism: An Islamic Controversy', in M. McGhee (ed.) *Philosophy, Religion and the Spiritual Life*, Cambridge: Cambridge University Press, 177-88.(A discussion of two different ways of interpreting Islamic philosophy.)

Leaman, O. (1996a) ‘Introduction’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 1-10.(Elucidation of different definitions of Islamic philosophy.)

Leaman, O. (1996b) ‘Orientalism and Islamic philosophy’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 1143-8.(Critique of Western approaches to Islamic philosophy.)

Nasr, S.H. (1996c) ‘Introduction’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 11-18.(Nasr has also produced a number of important chapters in this work, including ‘The Meaning and Concept of Philosophy in Islam’, 21-6; ‘The Qur’an and *Hadith* as Source and Inspiration of Islamic Philosophy’, 27-39; and ‘Ibn Sina’s "Oriental Philosophy"’, 247-52.)

Islamic fundamentalism

The philosophical roots of Islamic fundamentalism are largely the result of a conscious attempt to revive and restate the theoretical relevance of Islam in the modern world. The writings of three twentieth-century Muslim thinkers and activists - Sayyid Qutb, Ayatollah Ruhollah al-Khumayni and Abu al-'Ala al-Mawdudi - provide authoritative guidelines delineating the philosophical discourse of Islamic fundamentalism. However, whereas al-Khumayni and al-Mawdudi made original contributions towards formulating a new Islamic political theory, it was Qutb who offered a coherent exposition of Islam as a philosophical system.

Qutb's philosophical system postulated a qualitative contradiction between Western culture and the religion of Islam. Its emphasis on Islam as a sui generis *and transcendental set of beliefs excluded the validity of all other values and concepts. It also marked the differences between the doctrinal foundations of Islam and modern philosophical currents. Consequently Islamic fundamentalism is opposed to the Enlightenment, secularism, democracy, nationalism, Marxism and relativism. Its most original contribution resides in the formulation of the concept of God's sovereignty or lordship. This concept is the keystone of its philosophical structure.*

The premises of Islamic fundamentalism are rooted in an essentialist world view whereby innate qualities and attributes apply to individuals and human societies, irrespective of time, historical change or political circumstances. Hence, an immutable substance governs human existence and determines its outward movement.

1 Essentialism and dualism

Paganism (*jahiliyya*) is the generic designation given to all systems of thought other than Islam, both ancient and modern. According to Islamic fundamentalism, since the dawn of history human society has been a battleground between belief and unbelief, right and wrong, religious faith and idolatry. Individuals and their beliefs may carry different names in different ages, but this duality remains essentially the same.

The definition of paganism is thus stretched to encompass Greek philosophy in the ancient world as well as utilitarianism and existentialism in the modern age. To Sayyid Qutb, for example, paganism is deemed to be present wherever 'people's hearts are devoid of a divine doctrine that governs their thought and concomitant legal rules to regulate their lives' (Qutb 1982: 510). Moreover, although outward manifestations may differ from age to age, the nature and attributes of paganism remain permanent. On the other hand, religion operates throughout the ages within constant perimeters, rotating around a fixed axis. Furthermore, religion and the cosmic order reflect God's will in its harmonious design (see Cosmology).

In this scheme of things, human nature and the cosmos are substances which retain their identities while undergoing change. A substance generates properties and assigns them a function peculiar to their qualities. Properties inhere in substances and are dependent for their existence and persistence on them. Such properties are not incidental, but form an identifiable structure quite distinct from other structures. These properties are therefore not transferable, in that once transferred they lose their function or significance (see Substance).

According to Islamic fundamentalism, the essential nature of human beings is religious and atheism is an aberration. Throughout human history there have been only two methods of organizing human life: one that declares God to be the sole sovereign and source of legislation, and another that rejects God, either as a force in the universe or as the lord and administrator of society. These two methods are irreconcilable: the first denotes Islam, the second paganism. Once human beings accept legislation to be dependent on the will of an individual, a minority or a majority, and not as the prerogative of God alone, they lapse into a type of paganism, be it a dictatorship, capitalism, theocracy or communism.

However, human history is an emanation of a doctrinal concept that is implanted by God in human beings in their capacity as his designated lieutenants on this earth. The lieutenancy (*khilafa*) of a human being is to carry out the commands of God. According to this line of reasoning, most human societies in the twentieth century resemble in their way of life the state of affairs that existed before the rise of Islam. In order to re-establish Islam as a system of government, it is thus of primary importance to discover anew the fundamental constituents of its doctrine. Such an honourable task falls to a well-disciplined group of believers. These pioneers, dubbed 'the vanguard' by Qutb, 'the Revolutionary Party' by Abu al-'Ala al-Mawdudi and 'the holy warriors' by Ayatollah Ruhollah

al-Khumayni, are called upon to undertake the reinstatement of Islam in both its doctrinal and political dimensions. The method of struggle is often referred to as *jihad*, or holy war launched in the path of God.

2 The fundamental principles of Islam

The fundamental principles of Islam and the injunctions of its laws are one seamless garment woven by God for his creatures. Whereas Greek thought, particularly Aristotelian thought, asserted that we are political animals by nature, Islamic fundamentalism contends that the basic instinct of human beings is intrinsically religious. Religion is understood in this context to be Islam itself (see Islamic theology).

Islam has its own constant, immutable and clearly defined nature. Its underlying aim is to change the process of history and create a new human being, unfettered by subservience to other human beings or institutions. To be a Muslim is to believe in the fundamental principles of Islam in their entirety. Moreover, the doctrinal principles of Islam are not to be studied theologically, metaphysically or philosophically. Their study is primarily a practical endeavour aimed at discovering the base on which an all-embracing system is to be erected for the benefit of humanity. Theory and praxis go hand in hand; knowledge is simply a prelude to social action and political engagement.

For Qutb, al-Mawdudi and al-Khumayni, the doctrine of Islam forms an organic unity. A description of its constituent parts is therefore a mere analytical device, which should under all circumstances indicate the interdependence and complementarity of these parts. Once a part is detached and treated on its own it loses its significance, depriving the harmonious totality of its beauty and truth. The true nature of divinity, for example, cannot be understood apart from its direct efficacy in regulating the movement of the universe and in all its physical and spiritual connotations. Thus God’s divinity ensures the harmonious essence of cosmic laws: God sustains, guards and regulates the universe according to fixed laws. Nevertheless, his absolute will fashions every movement or event without being bound by them. These laws are not self-regulatory in that they persist as a result of the immediate act of God, and are thus created anew every moment. The world was created in time, a fact denoting a temporal beginning rather than an eternal existence.

In classical Islam, God’s attributes were enumerated and discussed by a number of theologians and philosophers, but his essence was deemed to lie beyond human knowledge. Islamic fundamentalism, as represented by Qutb’s system, shifted the debate to Islam’s essence and attributes. Hence the fundamental principles of Islam were considered by Qutb to consist in their delineation of God’s divinity as well as human servitude in carrying out the tenets of the message as handed down to the seal of the prophets, Muhammad. These fundamentals spell out God’s divinity (*uluhiyya*) and the servitude of animate and inanimate objects to God (*‘ubudiyya*), in addition to the true essence of the cosmos, life and humanity (see God, concepts of). Moreover, the visible and invisible worlds are both an integral part of this doctrine and should be present in treating the vicissitudes of human existence.

These fundamentals are not the result of an exertion by the human mind. Rather, the human mind receives them in their entirety once it is freed of its *a priori* conceptions. One does so by adhering to the sound linguistic or conventional meaning of the text in which such principles are propounded. The human mind has no function other than to understand the exact meaning of the text, irrespective of its conformity to the axioms of reason. Hence, one must accept the existence of angels, *jinns*, resurrection, hell and paradise without equivocation.

According to Qutb, the principle of divinity is the primary and most efficacious essence in the formation of the Islamic doctrine. The existence of such an essence, being absolute and eternal, does not stand in need of external evidence. The innate nature of human beings recognizes this existence, unless it is encumbered by corrupt beliefs that render it incapable of receiving this single fact. Furthermore, the methodology of the Qur’an itself is not concerned with affirming the existence of divine power. Rather, it concentrates on describing its true quality in order to rectify the distorted views of other creeds. This rectification is not confined to distortions which prevailed before the rise of Islam; its scope covers all deviant beliefs down to the present. It thus shows the aberrance or the negativity of Aristotle’s God (see Aristotle §16), that of Plato’s Forms and their adoption in Schopenhauer’s unconscious will (see Plato; Schopenhauer, A.), and the series of emanations elaborated by Plotinus and taken over by so-called Islamic philosophy (see Neoplatonism in Islamic philosophy). It also rectifies the dualism of Descartes as well as Bergson’s vital power (see Bergson, H.-L. §7), in addition to the materialism of Parmenides in the ancient world and that of Karl Marx in our modern period (see Materialism). The Qur’anic methodology is first

and foremost concerned with the question of monotheism rather than existence (see Monotheism). Its main aim is to show the simple, indivisible and unique essence of God; it also asserts the attributes of God in their utter uniqueness and splendour.

3 Islam's attributes

Whereas orthodox Islamic philosophy and theology (*'ilm al-kalam*) were largely concerned with defining and elaborating God's attributes, Islamic fundamentalism shifted its focus to the attributes of Islam itself. In other words, Islam became a substantive quality with certain characteristics which could rival in their structures and functions other modern ideologies, such as fascism and Marxism. This is not to say that the divine attributes were ignored, but their significance was made a function of the predicative characteristics of a new Islamic theory.

It is well known that in mainstream Islamic theology, as propounded in the tenth century AD by Abu al-Hasan al-Ash'ari, only those attributes denoting God's acts are considered to lie within human knowledge (see Islamic theology §§2-3; Ash'ariyya and Mu'tazila §1). These acts were held by al-Ash'ari to be seven in number. It is in this context that Qutb's doctrinal work, *Khasa'is al-tasawwur al-islami (The Characteristics of the Islamic Conception)*, gains in significance. Although Qutb contended that Islamic theology and philosophy were outdated modes of knowledge, tainted by their reliance on categories derived from classical Greek thought, he aspired to inaugurate a new Islamic vision using an amalgamation of ancient and modern ideas. He claimed, for example, that his new interpretation consisted of a direct act of understanding the Qur'an. This receptivity is said to be unmediated and based on an immediate grasp of Qur'anic verses.

However, Qutb's binary division of Islam into 'characteristics' and 'fundamentals' is reminiscent of orthodox debates on the essence and attributes of God. It is also worth mentioning that in enumerating the characteristics of Islam, Qutb devised a new list which nevertheless, in a manner reminiscent of al-Ash'ari, included seven attributes: lordship, constancy, comprehensiveness, balance, positivity, realism and absolute unity or monotheism. These attributes of Islam emanate from God's will and specify certain rules and modes of behaviour incumbent on all believers.

4 Knowledge, causation and faith

In Islamic fundamentalism, the affinity between philosophy and natural science, an axiom of classical and medieval thought (see Natural philosophy, medieval), is ruptured and deemed to be unwarranted. Scientific knowledge is then confined to technical details and superficial alterations, a fact that renders its concepts temporary, relative and liable to change. Science is linked with experimental knowledge rather than the discovery of underlying principles.

Islam continues to be credited with stimulating the promotion of experimental methodologies that were appropriated by European scholars after the Renaissance. Nevertheless, Islamic fundamentalism, while placing the Qur'an outside the scope of modern science and philosophical debates, persists in alluding to the shortcomings of Western theories and trends of thought. Qutb, for example, highlights the fact that life itself is not inherent in the nature of matter or the universe (see Life, origin of); rather, it was infused by God into dead substances. This statement allows him to refute Darwin's theory of evolution in so far as it leaves aside supernatural factors in explaining the emergence of living beings (See Darwin, C.R.). He also calls Karl Marx's interpretation of social progress by means of purely economic laws an arbitrary idea; so also is Bergson's concept of life as a willed or vital creation (see Marx, K. §8; Bergson, H-L. §7).

While Islamic fundamentalism rejects the atomist theory of orthodox Muslim theologians, it retains the idea of God as the real cause of events (see Islamic theology). Thus the connection between a cause and its effect is assumed to be the result of God's action. The metaphor used by al-Ghazali to show that combustibility, in the case of a flame coming into contact with a piece of cotton, has no other cause but God, is reiterated by Qutb. A piece of cotton is not set alight because of an act performed by a flame, but as a result of God's will to render the piece of cotton combustible. Moreover, God may decide to suspend the common course of nature, and miracles occur as an indication of the divine interruption of fixed laws. Such a miracle, Qutb points out, is mentioned in the Qur'an in relation to Abraham when a burning flame failed to set him alight. It is for this reason that the use of empirical evidence in order to demonstrate causality becomes an arbitrary human construct.

See also: Ash‘ariyya and Mu‘tazila; Causality and necessity in Islamic thought; al-Ghazali; Islamic philosophy, modern; Islamic theology; Monotheism

YOUSSEF CHOUEIRI

References and further reading

Choueiri, Y.M. (1990) *Islamic Fundamentalism*, London: Pinter Publishers; 2nd revised edn, London and Washington, DC: Cassell-Pinter, 1997.(An essential guide to the philosophical and political discourse of al-Mawdudi, al-Khumayni and Qutb.)

al-Khumayni, Imam R. (1981) *Islam and Revolution*, trans. and annotated H. Algar, Berkeley, CA: Mizan Press. (An essential collection of articles, essays and speeches outlining the theoretical and political background to Khumayni’s contribution.)

al-Mawdudi, S. (1932) *Towards Understanding Islam*, Beirut: The Holy Koran Publishing House, 1980.(A succinct exposition of the theoretical foundations of Islamic fundamentalism. The work was originally published in Urdu under the title *Risala-e-Diniyat*.)

Qutb, S. (1978) *Ma‘alim fi al-tariq (Milestones)*, Beirut: The Holy Koran Publishing House.(One of the most powerful accounts of the doctrine of Islamic fundamentalism.)

em2;__ (1980) *Khasa’is al-tasawwur al-islami (The Characteristics of the Islamic Conception)*, Beirut: Dar al-Shuruq.(One of the most comprehensive accounts of the theoretical assumptions of Islamic fundamentalism.)

em2;__ (1982) *Fi zilal al-Qur’an (Under the Aegis of the Qur’an)*, Beirut: Dar al-Shuruq, vol. 1: 510.(This work forms part of a multi-volume exegesis that is considered the essential theoretical work on modern Islamic fundamentalism.)

em2;__ (1986) *Muqawimat al-tasawwur al-islami (Foundations of the Islamic Conception)*, Beirut: Dar al-Shuruq. (An indispensable work on the philosophy of Islamic fundamentalism.)

Islamic philosophy

Islamic philosophy may be defined in a number of different ways, but the perspective taken here is that it represents the style of philosophy produced within the framework of Islamic culture. This description does not suggest that it is necessarily concerned with religious issues, nor even that it is exclusively produced by Muslims (see Islam, concept of philosophy in).

1 The early years of Islamic philosophy

Islamic philosophy is intimately connected with Greek philosophy, although this is a relationship which can be exaggerated. Theoretical questions were raised right from the beginning of Islam, questions which could to a certain extent be answered by reference to Islamic texts such as the Qur'an, the practices of the community and the traditional sayings of the Prophet and his Companions. On this initial basis a whole range of what came to be known as the Islamic sciences came to be produced, and these consisted largely of religious law, the Arabic language and forms of theology which represented differing understandings of Islam.

The early conquests of the Muslims brought them into close contact with centres of civilization heavily influenced by Christianity and Judaism, and also by Greek culture. Many rulers wished to understand and use the Greek forms of knowledge, some practical and some theoretical, and a large translation project started which saw official support for the assimilation of Greek culture (see Greek philosophy: impact on Islamic philosophy). This had a powerful impact upon all areas of Islamic philosophy. Neoplatonism definitely became the prevalent school of thought (see Neoplatonism in Islamic philosophy), following closely the curriculum of Greek (Peripatetic) philosophy which was initially transmitted to the Islamic world. This stressed agreement between Plato and Aristotle on a range of issues, and incorporated the work of some Neoplatonic authors. A leading group of Neoplatonic thinkers were the Ikhwan al-Safa' (Brethren of Purity), who presented an eclectic philosophy designed to facilitate spiritual liberation through philosophical perfection (see Ikhwan al-Safa'). However, there was also a development of Aristotelianism in Islamic philosophy, especially by those thinkers who were impressed by the logical and metaphysical thought of Aristotle, and Platonism was inspired by the personality of Socrates and the apparently more spiritual nature of Plato as compared with Aristotle (see Aristotelianism in Islamic philosophy; Platonism in Islamic philosophy). There were even thinkers who seem to have been influenced by Greek scepticism, which they turned largely against religion, and Ibn ar-Rawandi and Muhammad ibn Zakariyya' al-Razi presented a thoroughgoing critique of many of the leading supernatural ideas of Islam.

Al-Kindi is often called the first philosopher of the Arabs, and he followed a broadly Neoplatonic approach. One of the earliest of the philosophers in Baghdad was in fact a Christian, Yahya Ibn 'Adi, and his pupil al-Farabi created much of the agenda for the next four centuries of work. Al-Farabi argued that the works of Aristotle raise important issues for the understanding of the nature of the universe, in particular its origination. Aristotle suggested that the world is eternal, which seems to be in contradiction with the implication in the Qur'an that God created the world out of nothing. Al-Farabi used as his principle of creation the process of emanation, the idea that reality continually flows out of the source of perfection, so that the world was not created at a particular time. He also did an enormous amount of work on Greek logic, arguing that behind natural language lies logic, so that an understanding of the latter is a deeper and more significant achievement than a grasp of the former. This also seemed to threaten the significance of language, in particular the language - Arabic - in which God transmitted the Qur'an to the Prophet Muhammad. A large school of thinkers was strongly influenced by al-Farabi, including al-'Amiri, al-Sijistani and al-Tawhidi, and this surely played an important part in making his ideas and methodology so crucial for the following centuries of Islamic philosophy.

Ibn Sina went on to develop this form of thought in a much more creative way, and he presented a view of the universe as consisting of entirely necessitated events, with the exception of God (see Causality and necessity in Islamic thought). This led to a powerful reaction from al-Ghazali, who in his critique of Peripatetic philosophy argued that it was both incompatible with religion, and also invalid on its own principles. He managed to point to some of the major difficulties with the developments of Neoplatonism which had taken place in Islamic philosophy, and he argued that while philosophy should be rejected, logic as a conceptual tool should be retained. This view became very influential in much of the Islamic world, and philosophy came under a cloud until the

nineteenth century.

2 Philosophy in Spain and North Africa

A particularly rich blend of philosophy flourished in al-Andalus (the Islamic part of the Iberian penninsula), and in North Africa. Ibn Masarra defended a form of mysticism, and this type of thinking was important for both Ibn Tufayl and Ibn Bajja, for whom the contrast between the individual in society and the individual who primarily relates to God became very much of a theme. The argument was often that a higher level of understanding of reality can be attained by those prepared to develop their religious consciousness outside of the framework of traditional religion, a view which was supported and became part of a highly sophisticated account of the links between religion and reason as created by Ibn Rushd. He set out to defend philosophy strenuously from the attacks of al-Ghazali, and also to present a more Aristotelian account than had been managed by Ibn Sina. He argued that there are a variety of routes to God, all equally valid, and that the route which the philosopher can take is one based on the independent use of reason, while the ordinary member of society has to be satisfied with the sayings and obligations of religion. Ibn Sab‘in, by contrast, argued that Aristotelian philosophy and logic were useless in trying to understand reality since those ideas fail to mirror the basic unity which is implicit in reality, a unity which stems from the unity of God, and so we require an entirely new form of thinking which is adequate to the task of representing the oneness of the world. A thinker better known perhaps for his work on history and sociology than in philosophy is Ibn Khaldun, who was nonetheless a significant philosophical writer; he presents an excellent summary of preceding philosophical movements within the Islamic world, albeit from a conservative (Ash‘arite) point of view.

3 Mystical philosophy

Mystical philosophy in Islam represents a persistent tradition of working philosophically within the Islamic world (see Mystical philosophy in Islam). Some philosophers managed to combine mysticism with Peripatetic thought, while others saw mysticism as in opposition to Peripateticism. Al-Ghazali had great influence in making mysticism in its Sufi form respectable, but it is really other thinkers such as al-Suhrawardi and Ibn al-‘Arabi who produced actual systematic mystical thought. They created, albeit in different ways, accounts of how to do philosophy which accord with mystical approaches to reality, and which self-consciously go in opposite directions to Peripateticism. Ibn al-‘Arabi concentrated on analysing the different levels of reality and the links which exist between them, while al-Suhrawardi is the main progenitor of Illuminationist philosophy (see Illuminationist philosophy). This tries to replace Aristotelian logic and metaphysics with an alternative based on the relationship between light as the main principle of creation and knowledge, and that which is lit up - the rest of reality. This tradition has had many followers, including al-Tusi, Mulla Sadra, Mir Damad and al-Sabzawari, and has been popular in the Persian world right up to today. Shah Wali Allah extended this school of thought to the Indian subcontinent.

4 Islamic philosophy and the Islamic sciences

Islamic philosophy has always had a rather difficult relationship with the Islamic sciences, those techniques for answering theoretical questions which are closely linked with the religion of Islam, comprising law, theology, language and the study of the religious texts themselves. Many theologians such as Ibn Hazm, al-Juwayni and Fakhr al-Din al-Razi presented accounts of Islamic theology which argued for a particular theory of how to interpret religious texts (see Islamic theology). They tended to advocate a restricted approach to interpretation, rejecting the use of analogy and also the idea that philosophy is an objective system of enquiry which can be applied to anything at all. Most theologians were Ash‘arites (see Ash‘ariyya and Mu‘tazila), which meant that they were opposed to the idea that ethical and religious ideas could be objectively true. What makes such ideas true, the Ash‘arites argued, is that God says that they are true, and there are no other grounds for accepting them than this. This had a particularly strong influence on ethics (see Ethics in Islamic philosophy), where there was much debate between objectivists and subjectivists, with the latter arguing that an action is just if and only if God says that it is just. Many thinkers wrote about how to reconcile the social virtues, which involve being part of a community and following the rules of religion, with the intellectual virtues, which tend to involve a more solitary lifestyle. Ibn Miskawayh and Al-Tusi developed complex accounts of the apparent conflict between these different sets of virtues.

Political philosophy in Islam looked to Greek thinkers for ways of understanding the nature of the state, yet also

generally linked Platonic ideas of the state to Qur'anic notions, which is not difficult given the basically hierarchical nature of both types of account (see Political philosophy in classical Islam). Even thinkers attracted to Illuminationist philosophy such as al-Dawani wrote on political philosophy, arguing that the structure of the state should represent the material and spiritual aspects of the citizens. Through a strict differentiation of role in the state, and through leadership by those skilled in religious and philosophical knowledge, everyone would find an acceptable place in society and scope for spiritual perfection to an appropriate degree.

Particular problems arose in the discussions concerning the nature of the soul (see Soul in Islamic philosophy). According to the version of Aristotle which was generally used by the Islamic philosophers, the soul is an integral part of the person as its form, and once the individual dies the soul disappears also. This appears to contravene the notion of an afterlife which is so important a part of Islam. Even Platonic views of the soul seem to insist on its spirituality, as compared with the very physical accounts of the Islamic afterlife. Many of the philosophers tried to get around this by arguing that the religious language discussing the soul is only allegorical, and is intended to impress upon the community at large that there is a wider context within which their lives take place, which extends further than those lives themselves. They could argue in this way because of theories which presented a sophisticated view of different types of meaning that a statement may have in order to appeal to different audiences and carry out a number of different functions (see Meaning in Islamic philosophy). Only the philosopher really has the ability to understand this range of meanings, and those who work in the Islamic sciences do not know how to deal with these issues which come outside of their area of expertise. While those skilled in dealing with the law will know how to adjudicate between different legal judgements, we need an understanding of the philosophy of law in Islam if we are to have access to what might be called the deep structure of law itself (see Law, Islamic philosophy of). Similarly, although the Qur'an encourages its followers to discover facts about the world, it is through the philosophy of science that we can understand the theoretical principles which lie behind that physical reality (see Science in Islamic philosophy).

Many of the problems of religion versus philosophy arose in the area of aesthetics (see Aesthetics in Islamic philosophy). The rules of poetry which traditionally existed in the Arabic tradition came up against the application of Aristotle's *Poetics* to that poetry. One of the interesting aspects of Islamic aesthetics is that it treated poetry as a logical form, albeit of a very low demonstrative value, along the continuum of logical forms which lie behind all our language and practices. This is explained in studies of both epistemology and logic (see Epistemology in Islamic philosophy; Logic in Islamic philosophy). Logic came to play an enormous role in Islamic philosophy, and the idea that logic represents a basic set of techniques which lies behind what we think and what we do was felt to be very exciting and provocative. Many theologians who attacked philosophy were staunch defenders of logic as a tool for disputation, and Ibn Taymiyya is unusual in the strong critique which he provided of Aristotelian logic. He argued that the logic entails Aristotelian metaphysics, and so should be abandoned by anyone who wishes to avoid philosophical infection.

However, the general respect for logic provides the framework for the notion that there is a range of logical approaches which are available to different people, each of which is appropriate to different levels of society. For the theologian and the lawyer, for instance, dialectic is appropriate, since this works logically from generally accepted propositions to conclusions which are established as valid, but only within the limits set by those premises. This means that within the context of theology, for example, if we accept the truth of the Qur'an, then certain conclusions follow if we use the principles of theology; but if we do not accept the truth of the Qur'an, then the acceptability of those conclusions is dubious. Philosophers are distinguished from everyone else in that they are the only people who use entirely certain and universal premises, and so their conclusions have total universality as well as validity. When it comes to knowledge we find a similar contrast. Ordinary people can know something of what is around them and also of the spiritual nature of reality, but they are limited to the images and allegories of religion and the scope of their senses. Philosophers, by contrast, can attain much higher levels of knowledge through their application of logic and through their ability to perfect their understanding and establish contact with the principles which underlie the whole of reality.

5 Islamic philosophy in the modern world

After the death of Ibn Rushd, Islamic philosophy in the Peripatetic style went out of fashion in the Arab world, although the transmission of Islamic philosophy into Western Europe started at this time and had an important

influence upon the direction which medieval and Renaissance Europe was to take (see Averroism; Averroism, Jewish; Translators; Islamic philosophy: transmission into Western Europe). In the Persian-speaking world, Islamic philosophy has continued to follow a largely Illuminationist curriculum right up to today; but in the Arab world it fell into something of a decline, at least in its Peripatetic form, until the nineteenth century. Mystical philosophy, by contrast, continued to flourish, although no thinkers matched the creativity of Ibn al-'Arabi or Ibn Sab'in. Al-Afghani and Muhammad 'Abduh sought to find rational principles which would establish a form of thought which is both distinctively Islamic and also appropriate for life in modern scientific societies, a debate which is continuing within Islamic philosophy today (see Islamic philosophy, modern). Iqbal provided a rather eclectic mixture of Islamic and European philosophy, and some thinkers reacted to the phenomenon of modernity by developing Islamic fundamentalism (see Islamic fundamentalism). This resuscitated the earlier antagonism to philosophy by arguing for a return to the original principles of Islam and rejected modernity as a Western imperialist instrusion. The impact of Western scholarship on Islamic philosophy has not always been helpful, and Orientalism has sometimes led to an overemphasis of the dependence of Islamic philosophy on Greek thought, and to a refusal to regard Islamic philosophy as real philosophy (see Orientalism and Islamic philosophy). That is, in much of the exegetical literature there has been too much concern dealing with the historical conditions under which the philosophy was produced as compared with the status of the ideas themselves. While there are still many disputes concerning the ways in which Islamic philosophy should be pursued, as is the case with all kinds of philosophy, there can be little doubt about its major achievements and continuing significance.

See also: Ancient philosophy; Jewish philosophy; Medieval philosophy; Renaissance philosophy

OLIVER LEAMAN

References and further reading

Corbin, H. (1993) *History of Islamic Philosophy*, trans. L. Sherrard, London: Kegan Paul International.(An authoritative account of most Islamic philosophy, stressing in particular the illuminationist and mystical trends.)

Gibb, H. *et al.* (eds) (1960-) *Encyclopaedia of Islam: New Edition*, Leiden: Brill.(A continuing project which provides exhaustive accounts of all the main philosophical figures and concepts.)

Nasr, S. and Leaman, O. (eds) (1996) *History of Islamic Philosophy*, London: Routledge.(A comprehensive account of the different schools of thought, thinkers and concepts.)

Islamic philosophy: transmission into Western Europe

The Arabs took on the mantle of late antique philosophy and passed it on to both Latin scholars and Jewish scholars in Western Europe in the Middle Ages. The debates among Islamic scholars between rationalism and fideism also provided texts and models for Christian and Jewish debates. In this assimilation of Islamic thought, several stages can be observed. First, there was an interest in Neoplatonic cosmology and psychology in the latter half of the twelfth century, which fostered the translation of texts by al-Kindi, al-Farabi, the Ikhwan al-Safa' and, especially, Avicenna (Ibn Sina). Second, the desire to understand Aristotle's philosophy resulted in the translation of the commentaries and epitomes of Averroes (Ibn Rushd) in the second quarter of the thirteenth century. Jewish scholars participated in both these movements, and from the second quarter of the thirteenth century they took the initiative in translating and commenting upon Arabic texts. Thus when, in the late fifteenth century, a renewed interest in the ancient texts led scholars to search out the most accurate interpretations of these texts, it was to Jewish scholars that they turned for new translations or retranslations of Avicenna and, in particular, Averroes. From the early sixteenth century, Arabic philosophical texts were again translated directly into Latin, Arabic speakers began to collaborate with Christian scholars and the foundations for the teaching of Arabic were being laid. With the establishment of Arabic chairs in European universities, the rich variety of Islamic thought began to be revealed. This process has lasted until the present day.

1 Early translations: twelfth century

Some seventy works were translated from Arabic by Gerard of Cremona, nicknamed 'the Master' (*dictus magister*), at the cathedral of Toledo. These included Aristotle's *Posterior Analytics*, *Physics*, *On Generation and Corruption* and *Meteorology* I-III (see Aristotle), as well as four short tracts on natural science by Alexander of Aphrodisias. To these originally Greek works, Gerard added four philosophical letters of al-Kindi, a letter on proof by the Ikhwan al-Safa' and al-Farabi's *Kitab ihsa' al-'ulum (On the Classification of the Sciences)* (see al-Farabi).

Also at this time, the importance of the *al-Shifa' (Healing)* of Avicenna (Ibn Sina) was brought to the notice of Archbishop John of Toledo by a Jewish scholar called 'Avendauth' (perhaps the same as Abraham Ibn Daud, the author of *al-'Aqida al-rafi'a (The Exalted Faith)*). Portions of the text, including those on universals, physics (in part), the soul and metaphysics, were translated by Avendauth, Dominicus Gundissalinus (an archdeacon in the cathedral, *fl.* 1161-81) and a certain '*magister* John of Spain'. Members of this team also translated Algazel's *Maqasid al-falasifa (The Aims of the Philosophers)* (the first part of a two-part work, the second part being *Tahafut al-falasifa (Incoherence of the Philosophers)*) (see al-Ghazali); the *Mekor Hayyim (Fountain of Life)* of the Jewish philosopher Ibn Gabirol, translated into Latin as *Fons vitae*; and the *Liber de causis*, a cento of propositions from the *Elements of Theology* of Proclus, assembled in Arabic (see Liber de causis). Unattributed are translations of *On the Rise of the Sciences* (said in the Latin version to be by al-Farabi), and a compendium of sixteen questions on Aristotle's *On the Heavens* made by Hunayn ibn Ishaq and added to the selections from the *al-Shifa'* in the Latin tradition (*Collectiones expositionum ab antiquis grecis in libro Aristotelis De mundo qui dicitur liber caeli et mundi*).

Many of these works have a distinctly Neoplatonic tone, which is reflected in the original works of Gundissalinus and the anonymous author of the *Liber de causis primis et secundis et… qui consequitur eas (Book of the First and Second Causes and… Which Follows Them)*, who joins Avicenna to the radically Neoplatonic John Scottus Eriugena. Some of these works were being read in the late twelfth century at Oxford University, amongst whose scholars was, in all probability, the translator Alfred of Shareshill (responsible for the translation from Arabic of Nicholas of Damascus' *De plantis* and the sections from *al-Shifa'* on mineralogy, on inundations and, perhaps, on botany) whose commentaries influenced the style of philosophical teaching in the embryonic university (see Translators §3).

2 Thirteenth-century translations

The translation of *al-Shifa'* continued in the thirteenth century: Aristotle's *Zoology* was translated by Michael Scot, and the *Physics* (starting from the point where the earlier translation left off), *On the Heavens*, *On Generation and Corruption*, *On Actions and Passions*, *Meteorology* and perhaps the *Botany* were translated by Juan Gonsalvez de Burgos and a Jew called Salomon for Gonsalvez García de Gudiel, Bishop of Burgos

(1275-80), apparently from a single manuscript deposited in Toledo cathedral.

However, with the exception of the *Zoology*, these translations were hardly read. Rather, a new climate in which Aristotle's texts were on the one hand being accepted as the foundation for the arts curricula in the newly-founded universities, but on the other hand appeared to present some views dangerous to Christianity, made the interpretative works of Averroes (see Ibn Rushd) particularly relevant (see Aristotelianism, medieval). Many of the latter's commentaries on Aristotle's texts were translated. Michael Scot (d. before 1236), who moved from Toledo to the court of Frederick II in Sicily, translated Averroes' 'large commentaries' on the *On the Heavens* and probably those on the *On the Soul*, *Physics* and *Metaphysics* and others as well. Theodore of Antioch, the court philosopher, translated Averroes' *Proemium to the Physics*, and William of Luna, perhaps another member of Frederick II's circle, translated the 'middle commentaries' on the *Isagōgē* of Porphyry, and on Aristotle's *Categories* and *De interpretatione*.

Hermann the German (see Translators §3) translated the 'middle commentaries' on the *Rhetoric* (partial, together with excerpts on rhetoric from *al-Shifa'* and by al-Farabi), *Poetics* (in 1256) and *Nicomachean Ethics* (in 1240) in Toledo, but may also have been patronized by Frederick II's son Manfred, if he is the same as the 'translator Manfredi' mentioned by Roger Bacon (*Opus tertium*, cap. 25). Hermann's translation of Averroes' 'middle commentary' on the *Poetics* was the only form in which Aristotle's *Poetics* was known to Latin readers in the Middle Ages: William of Moerbeke made a Greek-Latin translation in 1278, but this was lost until the 1930s. The letter of Manfred to the philosophers of Paris, accompanying translations of 'Aristotle and other philosophers from Greek and Arabic', indicates one route by which these texts may have been transmitted, although the 'large commentaries' on *On the Soul*, *Physics* and *Metaphysics* were already known in Paris and Oxford in the late 1220s.

The fact that only authors whose writings have been translated from Arabic are included amongst the philosophers whose errors are listed by Giles of Rome - who mentions Aristotle, Averroes, Avicenna, Algazel, al-Kindi and Maimonides - gives some idea of the extent of the penetration of these texts. The margins of manuscripts of Aristotle's works on natural science from Oxford University in the same period, which are crammed with citations from Averroes' commentaries, also testify to this penetration.

The Arabic contribution to the Latin rationalist-fideist controversy began with the translation of Algazel's introduction to his double volume *Maqasid al-falasifa* and *Tahafut al-falasifa* by Gundissalinus' circle (see al-Ghazali). This introduction only survives in one manuscript, but was known to Roger Bacon. Maimonides' *Dalalat al-ha'irin (Guide to the Perplexed)*, originally written in Arabic (see Maimonides), which was read by and was most likely translated in the circle of Frederick II, provided material for the debate; both Algazel's work and Averroes' *Tahafut al-tahafut (Incoherence of the Incoherence)*, along with other texts, were used in the Arabic original by the Franciscan Ramón Llull, and in the polemical work *Pugio fidei (The Dagger of Faith)*, written in 1278 by the Dominican Ramón Martí. The doctrines of the Mu'tazila and Ash'ariyya, and especially the dialectical theologians (*mutakallimun*) were reported in the *Dalahat al-ha'irin* and the *Tahafut al-tahafut*, and Averroes' 'large commentary' on the *Physics* included the ideas of Ibn Bajja as well as of the Greek philosophers Themistius, Alexander of Aphrodisias and John Philoponus.

If the edict promoted by Llull at the Church Council of Vienne in 1312, calling for chairs in Arabic to be set up in four universities and the Papal curia, had been put into effect, Arabic studies in Europe might have had a different history. As it is, the teaching of Arabic in the Dominican mission schools and the use of newly-translated Arabic philosophical works by Dominican preachers against the Cathar heresy may have contributed to the interest in Arabic authorities shown by the Dominican masters, Albert Magnus (see Albert the Great) and Thomas Aquinas.

3 The contribution of Jewish scholars

Jewish scholars had often acted as interpreters of the Arabic texts for Christian translators or patrons. This practice continued in the fourteenth century when Calonymus ben Calonymus ben Meir translated Averroes' *Tahafut al-tahafut* from Arabic for Robert of Anjou, King of Naples. However, from the late thirteenth century onwards, most of the Latin translations of Arabic authors were made from Hebrew versions. When Jewish scholars in the West started to use Hebrew rather than Arabic as the language of learning, they began to translate a large number of Arabic philosophical texts into Hebrew. At least thirty-eight of Averroes' commentaries were translated into

Hebrew; the earliest of these is the 'middle commentary' on *De interpretatione*, translated at the court of Frederick II by Jacob Anatoli at the same time as William of Luna translated it into Latin. Thereafter in the thirteenth and fourteenth centuries several scholars in the south of France, Catalonia and Italy, including Moses ibn Tibbon, Shem Tov Ibn Falaquera, Levi ben Gerson (see Gersonides) and Moses of Narbonne translated Averroes' commentaries and other works, and wrote 'super-commentaries' on them. It was to these Hebrew versions that humanist scholars turned in the late fifteenth century, just at the time when the works of Aristotle, Avicenna and Averroes were being set in print (see Humanism, renaissance).

The new translations, patronized especially by Pico della Mirandola, Cardinal Domenico Grimani and Pope Leo X, were made by Elias del Medigo ('middle commentary' on *Meteorology* I-II, *Quaestiones* on *Prior Analytics*, the preface to Book Lambda of the *Metaphysics*, the 'middle commentary' on *Metaphysics* I-VII and the 'middle commentary' on Plato's *Republic*) (see Delmedigo, E.); by Paulus Israelita ('middle commentary' on *On the Heavens*); Abram de Balmes (epitomes of the Organon, 'middle commentaries' on *Topics*, *Sophistical Refutations*, *Rhetoric* and *Poetics*, the 'large commentary' on the *Posterior Analytics* and the *Epistola expeditionis* of Avempace (see Ibn Bajja)); Johannes Burana (epitome and 'middle commentary' on *Prior Analytics*, 'middle' and 'large commentary' on *Posterior Analytics*); Vitalis Nissus (epitome on *On Generation and Corruption*) and above all, Jacob Mantinus. This hard-working Jewish doctor from Tortosa had apparently been charged by Girolamo Bagolino, Romolo Fabi and Marco degli Oddi, the ambitious editors of Aristotle's complete works with Averroes' commentaries (finally published by Tommaso Giunta in Venice in 1550-2), with the complete overhaul of the translations of Averroes' commentaries. However, having made ten new translations of Averroes' texts (including another translation of the 'middle commentary' on Plato's *Republic*) and of Levi ben Gerson's 'super-commentary' on the Organon, he died in 1549 on a journey to Damascus. From the late 1530s onwards, the old translations were replaced or at least printed side by side with the new, as can be seen especially in the Giunta printing of 1550-2 and the subsequent Aristotle - Averroes editions of 1562 and 1573-6.

4 The beginnings of Arabic scholarship

The early sixteenth century also saw the beginning of a revival of the study of Arabic texts directly and, of providing the means to do this. A grammar and dictionary of Granadan Arabic by Pedro Alcalà was published in 1505. The kidnapped Arab scholar who took the name Leo Africanus, from Pope Leo X, provided biographies of Arabic authors after 1518. An Arabic manuscript of the so-called *Theology of Aristotle* (in reality parts of Plotinus' *Enneads* 4-6) (see Plotinus) was found in Damascus and, in the translation by Moses Arovas and Pier Nicolas Castellani, was published in Rome in 1519. Also in Damascus, Andrew Alpago revised the medieval translation of Avicenna's medical *Canon*, introduced two new works on the soul by Avicenna and described Sufi rituals from first-hand experience. In 1584, an Arabic press was set up in Rome by Giovan Battista Raimondi under the aegis of the Medici. Arabic studies became established through the efforts of Guillaume Postel, who took the first European chair in Arabic in Paris in 1535, and wrote a grammar, Franciscus Raphelengius, the author of the first dictionary of classical Arabic, Thomas Erpenius, the first to hold a chair in Arabic in Leiden in 1613 and the author of what became the standard grammar of Arabic, Jacob Golius and Edward Pococke. Pococke's *Specimen historiae Arabum* (including sketches of the ideas of a wide range of Islamic philosophers) and translation of Ibn Tufayl's philosophical novel, *Hayy ibn Yaqzan* (to which his son, of the same name, added an edition of the Arabic text), are indicative of the new, academic study of Arabic philosophy which was to flourish in European universities and which has culminated in the multi-volume publications of the Egyptian Academy's edition of the Arabic Avicenna, Simone Van Riet's *Avicenna Latinus* and the *Corpus Averroicum*. But at the same time, these works mark the end of a period in which Arabic authors were regarded as important for assimilating the knowledge which made humanity wise.

See also: Aristotelianism, medieval; Aristotelianism, renaissance; Averroism; Averroism, Jewish; al-Farabi; Gerard of Cremona; al-Ghazali; Greek philosophy: impact on Islamic philosophy; Ibn Rushd; Ibn Sina; Liber de causis; Translators

CHARLES BURNETT

References and further reading

Bertola, E. (1962) 'Le traduzioni delle opere filosofiche arabo-giudache nei secoli XII e XIII' (The Translations of Arabic-Jewish Philosophical Works in the Twelfth and Thirteenth Centuries), in *Studi di filosofia e di storia*

della filosofia in onore di Francesco Olgiati, Milan: Società editrice vita e pensiero.(A useful list of works translated and their manuscripts or printed editions.)

Butterworth, C.E. and Kessel, B.A. (1994) *The Introduction of Arabic Philosophy into Europe*, Leiden: Brill. (Articles deal respectively with Spain (before 1200), Paris in the thirteenth century, British schools in the Middle Ages, Oxford in the seventeenth century, Louvain from the sixteenth century to the present day, Budapest, Cracow and Kiev, and with the reception of Sufism in the West.)

Callus, D.A. (1943) 'Introduction of Aristotelian Learning to Oxford', *Proceedings of the British Academy* 29: 229-81.(Shows the influence of Avicenna and Averroes on the doctrine and format of Latin philosophical texts.)

Cranz, F.E. (1976) 'Editions of the Latin Aristotle Accompanied by the Commentaries of Averroes', in E.P. Mahoney (ed.) *Philosophy and Humanism: Renaissance Essays in Honor of Paul Oskar Kristeller*, Leiden: Brill, 116-28.(Explores the motives for making new translations of Averroes in the fifteenth and sixteenth centuries and the publication of the Aristotle - Averroes editions of 1550-2, 1562 and 1573-6.)

D'Alverny, M-T. (1993) *Avicenne en Occident (Avicenna in the West)*, Paris: Vrin. (Articles concerning both the medieval and the Renaissance translations of Avicenna.)

Halkin, A. (1971) 'Translation and Translators (Medieval)', *Encyclopedia Judaica* 15: 1318-29.(Summarizes in English the monumental research of M. Steinschneider.)

Hamesse, J. and Fattori, M. (eds) (1990) *Rencontres de cultures dans la philosophie médiévale: traductions et traducteurs de l'antiquité tardive au XIV^e siècle (Meetings of Cultures in Medieval Philosophy: Translations and Translators from Late Antiquity to the Fourteenth Century)*, Louvain: Brepols.(See especially the articles by A.L. Ivry on 'Philosophical Translations from the Arabic in Hebrew during the Middle Ages', and H. Daiber, 'Lateinische Übersetzungen arabischer Texte zur Philosophie und ihre Bedeutung für die Scholastik des Mittelalters', the latter of which includes a comprehensive bibliography.)

Jolivet, J. (1988) 'The Arabic Inheritance', in P. Dronke (ed.) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press, 113-47. (Expertly covers the period discussed in §1 above.)

L'Averroismo in Italia (Averroism in Italy) (1979), Atti dei convegni Lincei 40, Rome: Accademia nazionale dei Lincei.(Includes discussions on the court of Frederick II by R. Manselli, the introduction of Averroes into the West by F. Van Steenberghen and Averroism in the Renaissance by C.B. Schmitt.)

Leaman, O. (1996) 'Jewish Averroism', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 46, 769-80.(An account of the influence of Ibn Rushd on Jewish philosophy in the Middle Ages.)

Marenbon, J. (1996) 'Medieval Christian and Jewish Europe', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 58, 1001-12.(Short survey of the transmission and interpretation of Islamic philosophy in medieval Europe.)

Russell, G.A. (1994) *The 'Arabick' Interest of the Natural Philosophers in Seventeenth-Century England*, Leiden: Brill.(On the beginnings of academic research into Arabic learning.)

Stroumsa, G.G. (1992) 'Anti-Cathar Polemics and the *Liber de duobus principiis*', in B. Lewis and F. Niewöhner (eds) *Religionsgespräche im Mittelalter*, Wolfenbütteler Mittelalter-Studien 4, Wiesbaden: Harrassowitz, 169-83.(Discusses the use of new translations from Arabic and Hebrew in combatting the Cathar heresy.)

Wilson, C. (1996) 'Modern Western Philosophy', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 59, 1013-29.(Intriguing account of the ways in which modern Western philosophy came into contact with Islamic philosophy, and the impact which this contact may have had.)

Wolfson, H.A. (1961) 'The Twice-Revealed Averroes', *Speculum* 36: 373-92.(A readable introduction to the thirteenth century Arabic-Latin and fifteenth/sixteenth-century Hebrew-Latin translations of the commentaries of Averroes, and the main doctrinal debates that they fuelled.)

Zonta, M. (1996) *La filosofica antica nel medioevo ebraico (Ancient Philosophy in the Hebrew Middle Ages)*, Brescia: Paideia.(An authoritative survey of all the forms in which ancient philosophy was transmitted in the Hebrew language.)

Islamic philosophy, modern

There are a number of major trends in modern Islamic philosophy. First, there is the challenge of the West to traditional Islamic philosophical and cultural principles and the desire to establish a form of thought which is distinctive. From the mid-nineteenth century onwards, Islamic philosophers have attempted to redefine Islamic philosophy; some, such as Hasan Hanafi and Ali Mazrui, have sought to give modern Islamic philosophy a global significance and provide an agenda for world unity.

Second, there is a continuing tradition of interest in illuminationist and mystical thought, especially in Iran where the influence of Mulla Sadra and al-Suhrawardi has remained strong. The influence of the latter can be seen in the works of Henry Corbin and Seyyed Hossein Nasr; Mulla Sadra has exercised an influence over figures such as Mahdi Ha'iri Yazdi and the members of Qom School, notably Ayatollah Ruhollah Khomeini. The philosopher Abdul Soroush has introduced a number of concepts from Western philosophy into Iran.

Finally, there have been many thinkers who have adapted and employed philosophical ideas which are originally non-Islamic as part of the normal philosophical process of seeking to understand conceptual problems. This is a particularly active area, with a number of philosophers from many parts of the Islamic world investigating the relevance to Islam of concepts such as Hegelianism and existentialism. At the same time, mystical philosophy continues to exercise an important influence. Modern Islamic philosophy is thus quite diverse, employing a wide variety of techniques and approaches to its subject.

1 Reactions to the West

There has been a tendency in the Islamic world since the late nineteenth century to explore the issue of the relative decline or decadence of Arabic intellectual thought and science as compared with its Western equivalent. During the Christian medieval period the Islamic world was in its cultural and political ascendancy, and was at the centre of theoretical work in both science and philosophy. However, by the nineteenth century an enormous gap had opened between the Islamic world and the West. A wide variety of explanations for this decline have since been sought.

The realization that this gap existed led to the *Nahda* (rebirth or renaissance) movement between 1850 and 1914. Beginning in Syria but developed largely in Egypt, the movement sought to incorporate the main achievements of modern European civilization while at the same time reviving classical Islamic culture which predated imperialism and the centuries of decadence.

The main problem facing the *Nahda* thinkers had was how to interpret the Islamic cultural tradition, including philosophy, in an environment dominated by the West. Jamal al-Din al-Afghani and Muhammad 'Abduh both argued that Islam is inherently rational and need not be abandoned in the face of the encroachment of Western forms of scientific and cultural thought. The Egyptian philosopher Mustafa 'Abd al-Raziq also argued that it is possible to demonstrate the authenticity of traditional Islamic philosophical work and its modern relevance within Islamic society. He posits an inseparable link connecting rationalism and revelation in Islam, and he defends the traditional Islamic sciences as compatible with science and rationality. In this he constitutes what might be thought of as a more conservative position than his predecessor 'Abduh, who was more dubious about the values of some of the Islamic schools of thought, in particular of Sufism (see Mystical philosophy in Islam).

Muhammad 'Abid al-Jabiri suggests that a viable Arab future can only come about through a deconstruction and critique of the reasons for the decline of the Arab world. He criticizes the dichotomy between the Islamicists, who hark back to a Golden Age in the past, and the liberal Westernizers, who praise the principles of the European Renaissance from which colonialism originated. The solution he offers is the freeing of modern Arabic thought from both the language and the theological limitations of the past. The Arab mind has become very much part and parcel of traditional ways of exploring the world, and is restricted in its potential if it remains too closely wedded to its Islamic heritage.

Fu'ad Zakariyya' argued that the Arab world declined due to its inability to historicize the past and its dependence on tradition, while Zaki Najib Mahmud brought out the importance of philosophy in taking us from the known to the unknown, and was critical of the ability of religion to interfere with this movement in thought. Hasan Hanafi

presents a form of phenomenology which argues that a new concept of *tawhid* (divine unity) should be developed which will involve a principle of unity and equality for all people. Hanafi also throws the charge of decadence back at the West, suggesting that the West is now entering a period of decadence and will require an infusion of ideas and energy from the East. He uses the language of liberation theology, which holds that revelation is adaptable to the language of each age (see Liberation theology). The original revelation was suited to the time and place of the Prophet and not necessarily of the current world. Modern Muslims should reinterpret revelation in modern language and in accordance with present demands; fossilized conservatism is a misinterpretation of the true dynamic and dialectical spirit of Islam.

Fazlur Rahman also contends that Islamic conservatism contradicts the essence of Islam. Islam's aims are economic reform and the establishment of a just social order (see Islamic theology §6). According to the Qur'an, he argues, moral and economic decline are related events. Therefore Islamic societies should turn away from petrified conservatism and educate their children in the new technologies. Islam should not be limited to communities of the faithful, but should seek a prominent place in the new ethical and social world order.

Another movement in Islamic political philosophy depicts Muslims not as the antagonists of Western culture, but rather as being in the vanguard of the globalization of peace and social justice. The most popular thinker of this school in the USA is Malcolm X, who began his career as an isolationist minister for the Nation of Islam movement. At first he used Islam to separate African-Americans from white people, but later he preached an internationalized Islam that reaches beyond racial and national differences.

An important African thinker in this tradition is Ali Mazrui, who tries to harmonize several interdependent factors in Islamic theology with current global realities. Mazrui proposes a marriage between the Islamic monotheistic *jihad* (universal struggle), Islam's anti-racist and humanist agenda, and the need for global economic cooperation; he employs culture as a vehicle for social change through his integration of multiculturalism, the politics of pan-Islamicism and the need for globalism. He takes Islam to be the first Protestant revolution in Christianity. Moreover, he suggests that Islam's economic message turns monotheism from isolated spirituality to communitarian humanism - in the form of a Muslim world order among a community of faithful (*umma*) - through global economic cooperation, social justice and the brotherhood of all. The essence of a multicultural perspective implies the acknowledgement that cultures project their own biases onto their perceptions of other societies. In a world which demands global economic policy-making and increasing interdependency, Mazrui believes that Muslims should see their religion of 'all is Godism' as a type of globalism. His innovation (*ijtihad*) interprets the Islamic *jihad* as an agenda of global peace and justice, thereby transforming what is taken to be a negative image of Islam into a signal for economic unity and world peace.

2 The Persian approach to philosophy

The area of the Islamic world which continued most forcefully the Islamic tradition in philosophy after the decline of Peripateticism is undoubtedly Persia (see Islam, concept of philosophy in §§3-4). Interestingly, one of the most staunch advocates of the form of thought which might be called neo-Illuminationism, and which stems from the *ishraqi* principles of al-Suhrawardi, is Henry Corbin (1903-78), a French philosopher who worked in Iran. Corbin was active in translating and interpreting post-Avicennan Islamic philosophy with an emphasis on shi'ism, *ishraqi* thought and the mysticism of Ibn al-'Arabi. He posited the existence of a perennial school of philosophical wisdom, which can be detected through the recurrence of archetypal symbols such as the icon of light. Such icons exist in the works of Shihab al-Din al-Suhrawardi in the early twelfth century AD, and have their source in Eastern (*ishraqi*) traditions such as Zoroastrianism, Hermeticism and Manicheism. (The term *ishraq*, which signifies 'light', also means 'East' or 'Orient'.)

For Corbin, '*ishraq*' designates not only a static spatial direction but a prescriptive invitation for a hermeneutic reorientation, whereby persons scrutinize their spiritual needs and points of return to archetypal origins. Corbin also discusses the role of the imagination, a faculty which exists between the senses and the intellect. While the senses perceive discrete data and the intellect categorizes, imagination is concerned with the world of archetypes (*'alam at-mithal*). For example, the notion of the perfect person (*al-insan al-kamil*) is an icon for the psychic centre. This centre signifies peace and the perfection of the self-realization process. Corbin asserts that by means of a series of epistemic states - which include revelation (*kashf*) and recollection (or archetypal memory) (*dhikr*) - one may return to the eternal origin. This process describes a cycle, thereby reasserting the Islamic theme of the

unity of being (*al-wahdat al-wujud*).

Corbin's followers, such as Hermann Landolt, William Chittick and Seyyed Hossein Nasr, have developed his ideas in a variety of different ways. The latter is the best known contemporary Islamic philosopher. According to him, people share a spiritual component that cannot be actualized by either descriptive or pragmatic accounts of nature. Nasr's world perspective includes a normative element which integrates people in the same way as earlier religions and cosmologies (Nasr 1993). In the past, everyone considered their religion to be the true religion; today, however, we are confronted with a plurality of religions. How can a Muslim attain a workable relation with the sacred in such an environment? Nasr employs Sufism to refer to the archetypal dimensions common to all religions; it is through the realm of mysticism that different forms of spirituality meet. The contemporary world creates the need for followers of different creeds and cultures to communicate.

Islam must coexist with the Western world, but this does not imply an Islamic surrender to all the practices of Western society. Nasr's views on Western scientific progress show his dissatisfaction with many Western perspectives. Citing the ecological disasters of overpopulation and pollution, Nasr criticizes the value of Western technological advances. According to him, the fault lies in the mistaken metascientific presupposition that an innate nature exists which is disconnected from humanity and can be investigated separately and controlled. Moreover, the increasingly pervasive quantitative perspective supplied by units of measurement - like that by which the size of a building might be described - is an incomplete outlook because it does not articulate the qualitative effects of what it describes on the surrounding environment. By contrast, Nasr holds that Islamic and Eastern perspectives on science and technology are integrative and harmonious. They stress unity in their studies of nature, thereby acknowledging the long-term ecological significance of development. Unless religious and spiritual values are embedded in a technological agenda, ecological disasters as well as a general lack of a sense of meaning in life are inevitable. Western science and its technological consequences are of ecological import to modern civilization. Consequently, neither science nor technology can consider itself irrelevant to environmental ethics (see Environmental ethics). Philosophy along Neoplatonic lines should be pursued, since only this form of analysis does justice to the spiritual wholeness of humanity.

The main emphasis in recent Persian philosophy has been on the thought of Mulla Sadra and al-Suhrawardi. Islamic philosophy has moved from the *madrasa* (traditional school) system and became an important part of the university curriculum. One of the most interesting thinkers is Mahdi Ha'iri Yazdi, whose work on knowledge by presence (*'ilm al-huduri*) provides an example of the fruitful combination of ideas from Western analytical philosophy and the *ishraqi* tradition in order to elucidate metaphysical and epistemological problems (Ha'iri Yazdi 1992). Recent Shi'ite theologians, as students of the work of Mulla Sadra, were versed in the dialectics of time and change. 'Ali Shari'ati, another student of Corbin, is an important social thinker whose work advocates a social process of Islamization. He rejects both the Peripatetic philosophers and the mystical thinkers, claiming that the existential being of each person contains a determination formed through mutual trust and compassion between them and God as their essence. This presumption is the ground for each person's being and the very core of each subject's potential for therapeutic unity (*tawhid*); its purpose is justice in both the providential and the social contexts. Islamization is achieved through an existential empathy and a phenomenological assimilation of exemplary people - such as the Imam Hossein (the Prophet's grandson) or Fatima (his daughter and the wife of Imam 'Ali) - into archetypal memory. The martyrdom of 'Ali or Hossein is a paradigmatic message, not for sorrow but for the assimilation of their characters into the self. Further, Shari'ati depicts history as a dialectical process which does not exclude economic and material realities, Islam as a practical religion or people as potential agents of justice. He replaces the Platonic theory of epistemic recollection with a theory of normative archetypal recollection. One may gain normative knowledge through the archetypal recollection of a religion's most exemplary mythical figures. Religion provides social ideals, and yet it demands not a withdrawal to a secret realm but a social revolution in the everyday world.

A creative commentary on Mulla Sadra was produced by Ruhollah Khomeini, who argues that people are primarily social as well as private citizens. Thus religious teachings relate not only to the personal morality of individuals but also to their social responsibilities and political actions (see Social sciences, philosophy of). In practice, these ideas imply a theocracy that does not distinguish between politics and religion. Bringing such a dominion into existence, he claims, requires an internal revolution from the masses directed against the existing ruling class, but this revolution must be guided by the directives of the religious authorities. He modifies Islamic

theology with the notion of the religious jurist-ruler’s guardianship (*velayat-e-faqih*), whose role is to guide the community of faithful in their universal struggle (*jihad*). This *jihad* is not essentially military, but is largely educational and seeks the expansion of monotheistic (that is, Islamic) ethics (see Ethics in Islamic philosophy).

Khomeini was a member of the School of Qom, based on the college in that city, which also produced Muhammad Hossein Tabataba’i, Murtaza Mutahheri and Muhammad Taqi Misbah Yazdi, all of whom have directed their influential thought at confronting the challenge to Islamic philosophy coming from the West. It should not be thought that this is an essentially reactionary strategy, however; Misbah has encouraged many of his students to study in the West and to take seriously scientific and logical thought as practised in the West. Also, although much of Misbah’s work has been on Mulla Sadra, he has been far from uncritical of the latter. In particular, he criticizes the notion of prime matter, which Mulla Sadra (§§1-2) identifies as the pure potentiality for existents. He questions the principle that a potentiality for existents exists prior to existents themselves; after all, there is nothing but existents. Misbah argues further that many relations are not truly essences. For example, in the mind-dependent realm, we may ascribe ‘below’ as a relation between a table and book, but this subject-directed ascription does not imply that below is an essence in the actual world.

An interesting and quite recent controversy in Persian philosophy has been that between Abdul Soroush on the one hand, and the philosophers of the school of Qom, as well as those influenced by the Corbin school, on the other. Soroush introduced a number of concepts from Western philosophy into Iran, in particular the leading ideas of Popper, Moore, Berlin and Wittgenstein. This led him to suggest that we should use a notion like that of collective reason to understand and interpret religious ideas. Collective reason is the best way of dealing with theoretical and practical problems, and is preferable to relying solely on solutions attainable through the efforts of the jurisprudents and religious authorities. Not surprisingly, this aroused the ire of the school of Qom philosophers, and their representative Sadiq Larijani engaged Soroush in a debate which largely dealt with the correct interpretation of thinkers such as Popper, Watkins and Stalnaker, and in particular Hempel’s paradoxes of confirmation (see Hempel, C.G. §2). Soroush was also attacked by the Corbin circle, whose basic philosophical approach relies very much on Heidegger along with traditional Islamic philosophy, and who were quite out of sympathy with the analytical nature of Soroush’s books. This controversy is interesting in that it brings out the fact that philosophers in Iran are generally familiar now not only with traditional forms of Islamic philosophy but also with the current philosophical ideas of the West. Modern philosophers do not entirely reject Western views, but neither are they completely taken over by the West; they are prepared to examine Western views with a critical sympathy.

3 Modern trends

A very vibrant area in Islamic philosophy is the history of philosophy, in particular the Greek tradition in Islamic philosophy. There exists both in the West and in the Islamic world a large number of scholars who have developed accounts of this close relationship and who continue to edit, translate and work on important texts in order to get some idea of the nature of the philosophical material which was produced in the early centuries of Islam. In addition, many philosophers in the Islamic world have adapted Western philosophy so as to make sense of the philosophical problems in which they are interested. C.A. Qadir in Pakistan developed an account of Islamic philosophy which he thought was in line with logical positivism, while ‘Abd al-Rahman Badawi applied existentialism to Arab society. Zaki Najib Mahmud followed William James in presenting a pragmatic account of philosophy. Some thinkers applied particular techniques in the Islamic tradition to philosophy, so that ‘Ali Sami al-Nashshar for example based his work on Ash‘arite theology (see Ash‘ariyya and Mu‘tazila), while Muhammad ‘Aziz Lahbabi (1954) used Hegelianism to develop a theory of being which is quite unusual within the context of Islamic ontology. Hichem Djait (1986) combines Hegelianism with existentialism. He argues that only dialectical epistemology can be used to understand the modern situation of the Arab world, and that the apparent opposites of decadence/renaissance, Arab/non-Arab, orthodox/heterodox, tradition/modernity need to be transcended if we are to understand the present nature of Islamic culture. Abdallah Laroui (1976) and Muhammad Arkoun (1985) both stress the contrast between Islam and modernity, and the former advocates the adoption of Westernization as the appropriate strategy for the Islamic world. In his approach to the Qur’an, Arkoun uses the semiotic ideas current in modern French literature to argue that Islam has always been changing and developing, so that there is no point in referring to a particular constant orthodoxy.

While many of these thinkers are hostile to mysticism and its Islamic form, Sufism, there can be little doubt that the latter represents a very potent framework for a good deal of present Islamic philosophy. The tradition of Sufism presents both a way of life which avoids many of the rigidities of traditional Islam and also a complex conceptual system which enables the philosopher to develop ideas and arguments which are intellectually satisfying. Modern Islamic philosophy employs a wide variety of different techniques and approaches to the subject.

See also: ‘Abduh, M.; Al-Afghani; Illuminationist philosophy; Islam, concept of philosophy in; Islamic fundamentalism; Islamic theology; Mystical philosophy in Islam

PARVIZ MOREWEDGE
OLIVER LEAMAN

References and further reading

Arkoun, M. (1985) *La pensée arabe (Arab Thought)*, Paris: Presses Universitaires de France.(Account of how theoretical concepts in the Arab world have changed in response to influences from the West.)

Brown, S., Collinson, D. and Wilkinson, R. (eds) (1995) *Biographical Dictionary of Twentieth-Century Philosophers*, London: Routledge.(Contains information on a number of modern Islamic philosophers. Relevant entries include ‘Arkoun, Mohammed’ (30-1), ‘Corbin, Henry’ (159-60), ‘Hanafi, Hasan’ (305-6), ‘Lahbabi, Muhammad Aziz’ (431), ‘Nagib Mahmud, Zaki’ (562), ‘Nasr, Seyyed Hossein’ (563-4), ‘Qadir, C.A.’ (641), ‘Rahman, Fazlur’ (645-6), and ‘Yazdi, Mehdi Hairi’ (859). These and many other thinkers are also discussed in Nasr and Leaman (1996).)

Corbin, H. (1993) *History of Islamic Philosophy*, with the collaboration of S.H. Nasr and O. Yahya, trans. P. Sherrard, London: Kegan Paul International.(Discussion of the links between Islamic philosophy and contemporary Persian thought.)

Clarke, J.H. (1993) *Malcolm X: The Man and his Times*, Trenton, NJ: Africa World Press.(Study of the Nation of Islam leader and his thought.)

Djait, H. (1986) *Europe and Islam: Cultures and Modernity*, Berkeley, CA: University of California Press. (Discussion of how the Islamic renaissance came about through contact with the West, and how it has revived Arab culture.)

Ha’iri Yazdi, M. (1992) *The Principles of Epistemology in Islamic Philosophy: Knowledge by Presence*, Albany, NY: State University of New York Press.(A combination of modern Western epistemology with illuminationist philosophy into a creative and interesting synthesis, representing the openness of modern Islamic philosophy to Western thought.)

Khomeini, R. (1981) *Islam and Revolution: Writings and Declarations*, trans. H. Algar, Berkeley, CA: Mizan Press.(An account of the theological and philosophical bases of the Shi‘i notion of the Islamic state.)

Lahbabi, M. (1954) *Le personalisme musulman (Muslim Personalism)*, Paris: Presses Universitaires de France. (The application of Hegelianism and existentialism to Islamic thought, together with the argument that the latter has to develop in accordance with changing cultural and material trends.)

Laroui, A. (1976) *The Crisis of the Arab Intelligentsia: Traditionalism or Historicism?*, Berkeley, CA: University of California Press.(Argument for the replacing of traditional Islamic issues with Western ones, since the Islamic world needs to transcend its past to come into real contact with modernity.)

Morewedge, P. (1990) ‘The Onyx Crescent: Ali A. Mazrui on the Islamic-Africa Axis’, in O.H. Kokole (ed.) *The Global African: A Portrait of Ali A. Mazrui*, Trenton, NJ: Africa World Press, 217-65.(On Mazrui’s African-Islamic philosophy.)

Morewedge, P. (1995a) *Essays in Islamic Philosophy, Theology and Mysticism*, Oneonta, NY: Oneonta Philosophy Series.(Interesting account of the basic ideas of Islamic philosophy, theology and mysticism.)

Morewedge, P. (1995b) ‘Theology’, in J.L. Esposito (ed.) *The Oxford Encyclopedia of the Modern Islamic World*, Oxford: Oxford University Press, vol. 4, 214-24.(Description of modern theological issues, with their underlying assumptions from Islamic philosophy.)

Nasr, S.H. (1993) *The Need for a Sacred Science*, Albany, NY: State University of New York Press.(Defence of the notion of a firm of knowledge which is thoroughly based on contact with the sacred.)

Nasr, S.H. (1996) ‘Islamic Philosophical Activity in Contemporary Persia: A Survey of Activity in the 50s and 60s’, in M.A. Razavi (ed.) *The Islamic Intellectual Tradition in Persia*, Richmond, VA: Curzon.(Useful account of the forms of philosophical thought in Iran during this period.)

Nasr, S.H. and Leaman, O. (eds) (1996) *History of Islamic Philosophy*, London: Routledge.(See the section ‘The Modern Islamic World’, 1037-1169, in particular M. Aminrazavi, ‘Persia’, 1037-50; M. Suheyl Umar,

‘Pakistan’, 1076-80; I. Abu-Rabi‘, ‘The Arab World’, 1082-1114; M. Campanini, ‘Egypt’, 1115-28; Z. Moris, ‘South-East Asia’, 1134-40; P. Lory, ‘Henry Corbin’, 1149-55; S. Akhtar, ‘The Possibility of a Philosophy of Islam’, 1162-69. This section contains discussions of all the thinkers mentioned in this entry.)

Rahman, F. (1982) *Islam and Modernity: Transformation of an Intellectual Tradition*, Chicago, IL: University of Chicago Press.(A defence of modernism and the importance of independent judgment, and an emphasis on ethics as opposed to metaphysics in philosophy.)

Shari‘ati, A. (1980) *Marxism and Other Western Fallacies: An Islamic Critique*, trans. H. Algar, Berkeley, CA: Mizan Press.(An advocate of the significance of the people of the Third World using their culture to overthrow imperialism, arguing for an Islamic state on different and more liberal principles as compared to Khomeini.)

Islamic theology

'Ilm al-kalam *(literally 'the science of debate') denotes a discipline of Islamic thought generally referred to as 'theology' or (even less accurately) as 'scholastic theology'. The discipline, which evolved from the political and religious controversies that engulfed the Muslim community in its formative years, deals with interpretations of religious doctrine and the defence of these interpretations by means of discursive arguments.*

The rise of kalam *came to be closely associated with the Mu'tazila, a rationalist school that emerged at the beginning of the second century AH (seventh century AD) and rose to prominence in the following century. The failure of the Mu'tazila to follow up their initial intellectual and political ascendancy by imposing their views as official state doctrine seriously discredited rationalism, leading to a resurgence of traditionalism and later to the emergence of the Ash'ariyya school, which attempted to present itself as a compromise between the two opposing extremes. The Ash'arite school gained acceptability within mainstream (Sunni) Islam. However,* kalam *continued to be condemned, even in this 'orthodox' garb, by the dominant traditionally-inclined schools.*

In its later stages, kalam *attempted to assimilate philosophical themes and questions, but the subtle shift in this direction was not completely successful. The decline of* kalam *appeared to be irreversible, shunned as it was by traditionalists and rationalists alike. Although* kalam *texts continued to be discussed and even taught in some form,* kalam *ceased to be a living science as early as the ninth century AH (fifteenth century AD). Attempts by reformers to revive it, beginning in the nineteenth century, have yet to bear fruit.*

1 The pre-Mu'tazilite groups

The term *kalam* has usually been translated as 'word' or 'speech', but a more appropriate rendering in this context would be 'discussion', 'argument' or 'debate'. Those who engaged in these discussions or debates were referred to as *mutakallimun* (those who practise *kalam* or debate). The term has special significance in that traditionalists disapproved of these discussions, arguing that the early Muslims were not known to have indulged in them. Those who dabbled in such debates were said to have 'spoken about' or 'discussed' (*takallma fi*) 'forbidden' topics. The proponents of *kalam* also liked to refer to it as *'ilm al-usul* (the science of basic principles) or *'ilm al-tawhid* (the science of [affirming God's] unity), and it is under this latter name that some of its topics continue to be taught and discussed in Muslim educational institutions today.

The rise of *'ilm al-kalam* was a result of the many controversies that had divided the Muslim community in its early years. Although the emergence of Islam was characterized by polemics with polytheists and followers of earlier revelations, controversies over fundamental religious questions were deemed irreverent by early Muslims, especially during the lifetime of the Prophet. However, disputes (mainly political) broke out immediately following the death of the Prophet, and again following the tragic events that led to the murder of the third Caliph Othman in AH 35/AD 656, this time heralding the breakdown of the political system established after the Prophet's death.

In a community that defined itself in terms of its religious identity, political disputes inevitably turned into theological ones. The political struggles over who should lead the Muslim community gave rise to three major competing groups: the Khawarij, who opposed the fourth Caliph 'Ali and rejected the compromises he made with his opponents; the Shi'a, who supported 'Ali; and the Murjiya, who tried to remain neutral. These groups attempted to influence a wider Muslim community dominated by a loose grouping of mainstream schools, mainly conservative or traditionalist, known collectively as *ahl al-sunna wa'l-jama'a* (the proponents of the [Prophet's] traditions and consensus).

The term *khawarij* (literally 'rebels') first referred to a group of dissidents who rebelled against the leadership of 'Ali following the inconclusive battle of Siffin (AH 37/AD 658) between 'Ali and his challenger, Mu'awiya, and later evolved into a distinct antiestablishment tendency. The Khawarij had neither a unified leadership nor a settled doctrine, and was primarily a militant political tendency with an uncompromising attitude. The core of their views revolved around the nature of legitimate leadership and the conditions for salvation. Although the Khawarij's uncompromising views condemned them to a marginal existence, their impact on the general body of the Muslim community was significant. Most of the major schools of thought that emerged did so in response to one or other

of their assertions, especially on the issues of leadership and the 'status of sinners'.

At the opposite pole stood the Shi'a (party) of 'Ali. Unlike the Khawarij, who defied all authority, the Shi'a believed in the undisputed authority of the divinely ordained *imam* (leader). The position of 'Ali as *imam* and successor to the Prophet was vouchsafed by revelation and was not a matter of opinion. Each *imam* would then designate his successor by virtue of the divine authority vested in him. In theory, Shi'ism should not have encouraged much theological speculation, since it sought to perpetuate and reproduce the authority of the Prophet and vest it in the person of the living *imam*, who had direct access to the divine truth. In practice, however, Shi'ism did indulge in theological speculation, especially with the emergence of the doctrine of the Absent *Imam*, which referred the burden of seeking the truth back to the community.

In between these two extremes, a large number of intermediary positions were espoused, notably that of the Murjiya. This group refused to condemn the perpetrators of grave sins (a euphemism for usurpers of power) as unbelievers, but neither did it want to absolve them, arguing that the matter should be left to God to judge in the hereafter. Murjiism was also associated with political neutrality, and an implied tacit support for the *status quo*.

While the above three groups were political in origin, adopting theological arguments to support their politics, there were also groups of which the primary focus was on theology. The earliest of these was the Qadariyya (the name, meaning proponents of *qadar*, or predestination, was a misnomer for this school which supported freedom of the will). This school argued for the absolute freedom of the will. God, its members said, would not put us human beings under obligation to act righteously if we did not possess the power to choose our course of action.

Diametrically opposed to this school were the Jabriyya (determinists). Their most prominent spokesman was Jahm ibn Safwan (d. AH 128/AD 746), who taught that no attributes could be predicated of God except for creation, power and action, since any attribute that could be predicated of creatures was not fit to be predicated of the Creator. As God is the sole Creator and actor, our actions are also authored by him alone; therefore, we as persons have no control over our actions and no free will. Jahm also said that since God could not be described as a speaker, the Qur'an could not be said to be his word, except in the sense of having been created by him.

2 Mu'tazila and rise of *kalam*

These earlier schools were amorphous groupings, very fluid both in membership and doctrine. With the exception of the Shi'a, who later developed into a number of coherent sects, these tendencies either faded away or merged into other tendencies. The rise of a systematic theological discourse had to await the emergence of the Mu'tazila. The association of *kalam* with the Mu'tazila, who were characterized by their elitism and their militant rationalism, determined its course and its eventual fate. The Mu'tazila attempted to systematize religious doctrine into a rational schema centred on the affirmation of God's absolute unity and absolute justice (see Ash'ariyya and Mu'tazila).

However, the Mu'tazila's elitism and their irreverent quest for 'a reason for everything,' to paraphrase al-Shahrastani, alienated the more conservative mainstream tendencies. The latter questioned the very possibility of a theological discourse of the type advocated by the Mu'tazila, regardless of content, viewing such discourse as at best superfluous and at worst a heretical deviation. This attitude was expressed succinctly by Malik ibn Anas (d. AH 179/AD 795), the leading jurist of Medina, when asked to explain how God could be said to have 'established himself on the Throne' as mentioned in the Qur'an: 'The establishment is known, the modality is unknown, the belief in it is obligatory and asking questions about it is an unwarranted innovation.' On the question of divine justice, the traditionalists rejected the Mu'tazila's attempts to impose human and rational concepts of justice on God. It would be meaningless to speak of justice in this context, since God was the absolute sovereign and absolute master of all His creation, which meant that anything which He did was by definition just (see Omnipotence §5).

The struggle between the two trends came to a head in the 'creation of the Qur'an' controversy which erupted in the first half of the third century AH (ninth century AD). The 'Inquisition' which the Mu'tazila instigated with the help of the ruler of the day, the Caliph al-Ma'mun (AH 198-218/AD 813-33), to enforce this and related doctrines proved disastrous, not only for the Mu'tazila but also for the discipline of *kalam* itself. In spite of being reclaimed for orthodoxy by Abu'l-Hassan al-Ash'ari and others, the science and art of arguing matters of faith with appeal to unaided human reason fell into disrepute and went into a decline from the start of the sixth century.

3 Main themes

Classical definitions tended to emphasize the apologetic function of *kalam*, probably in order to appease traditionalist critics. Al-Iji speaks of 'a science which makes it possible to prove the truth of religious doctrines by marshalling arguments and repelling doubts' (*al-Mawaqif*: 7). *Kalam*, however, has also been the arena on which battles over what constituted true religious doctrine were fought between rival schools.

The subject of *kalam*, to quote al-Iji again, was 'knowledge on which the proofs of religious doctrines depend, directly or indirectly' (*al-Mawaqif*: 7). It was also said to deal with *'usul* (basics), as opposed to *furu'* (subsidiary issues). These included the fundamentals of religious belief, such as God, his attributes and acts, the proofs of religious doctrines, the nature of the universe and our place in it.

The first issue that divided Muslims into opposing schools was the question of political authority and its legitimacy. Most traditionalist and mainstream schools accepted the actual procedures adopted to elect the first four caliphs as normative, thus affirming that a ruler gains legitimacy by being freely elected by the influential members of the community. The Khawarij accepted the procedures up to the election of the third caliph, but then added that even an elected caliph should be removed if he deviated from his mandate. The Khawarij also held that any qualified individual was fit to be caliph, provided the community at large approved of him. The traditionalists narrowed the field of selection to the Prophet's tribe, Quraysh, while the Shi'a narrowed it still further to the Prophet's family, in particular his son-in-law 'Ali and the latter's descendants. Shi'ism argued that political leadership, being the most important religious institution, could not be left for human reason to determine.

The second major issue to be discussed within *kalam* was the status of the grave sinner. The Khawarij started this debate by arguing, contrary to mainstream opinion, that any person who committed a grave sin automatically became a non-believer, thus forfeiting all rights and protections afforded by Islamic law. The Murjiya argued for the withholding of judgment while tending to widen the interpretation of who could qualify as a believer; the Mu'tazila held that such a person was in an intermediate position, being neither a Muslim nor an unbeliever.

The third major issue discussed in *kalam* was freedom of the will. The Mu'tazila and Qadariyya both came out unequivocally in support of freedom of the will. They held that we are the creators of our own acts, for otherwise God would be committing a grave injustice if he were to punish those who had no choice in what they did (see Evil, problem of §3; Free will). At the other extreme, the Jabriyya held that man could not have any control over his actions, since God was the sole creator and actor. Most other groups tried to strike a balance between these two poles. The Shi'a tended to affirm the freedom of the will and some of them, such as the Zaydiyya, agreed completely with the Mu'tazila on this. Some Shi'a factions, however, qualified their stance by affirming that we are in part compelled because of the chain of causation that triggered our acts. The Khawarij accepted the idea of predestination, holding that God was the Creator of the acts of people, and that nothing occurs which he did not will.

This was also the view of mainstream orthodox and traditionalist groups, who affirmed that the will of God was supreme and that he was the creator of all human acts, whether evil or good; nothing could happen on earth that contradicted his will. This position was later given some nuances by al-Ash'ari, who argued that God created human acts, but we acquired (*kasaba*) these acts by willing them prior to their creation.

The fourth major issue discussed in *kalam* was the question of divine attributes. The Jabriyya used the affirmation of the uniqueness of God's attributes to deny the existence of free will. The Mu'tazila developed the idea further, arguing that God could not have attributes in addition to his essence, for this would mean a multiplicity of eternal entities. Later Mu'tazilites, such as Abu'l-Hudhayl al-'Allaf (d. AH 227/AD 842), added that the divine attributes are identical with the divine essence. God's knowledge is not an attribute added to his essence, but is identical with that essence.

Early Shi'ite theologians opposed the Mu'tazila, affirming God's immanence in space and denying his immutability and transcendence of time and space. They held that God's will was also mutable, and ascribed motion to him. God could also be the locus of accidents (*hawadith*) and was corporeal in some sense. God's knowledge and will could not be eternal, for this would negate human freedom and make accountability redundant. It could also imply the eternal existence of things. Later Shi'ite theologians, however, especially the Zaydiyya,

repudiated most of the anthropomorphisms of their predecessors and veered towards Mu‘tazilite positions.

Traditionalists (who include the Ash‘ariyya) affirmed the reality of God’s eternal attributes, which they said were neither identical with his Essence nor distinct from it. They also affirmed the literal sense of apparently anthropomorphic Qur’anic references, such as those to God’s ‘face’, ‘hands’ and ‘eyes’, adding that the exact nature of these limbs could not be known.

Related to the issue of divine attributes was the issue of the Qur’an’s creation. The Mu‘tazila denied that God’s words were eternal and affirmed that the Qur’an had to be created; this idea was accepted also by the Khawarij. However, the bulk of the traditionalists (and Ash‘ariyya) rejected this view, arguing that one could not describe God’s speech as created because this would mean that God was subject to changing states. Speech (*kalam*) was one of God’s eternal attributes, and the Qur’an, being God’s word, could not be said to be created or uncreated. Some early Shi‘i theologians, in particular Hisham ibn al-Hakam (d. *c.* AH 200/AD 816), developed a more sophisticated version of the latter argument, saying that the Qur’an (or God’s word) could not be described as creator, created or uncreated, because an attribute, being an adjective, could not have another adjective predicated of it. Similarly, one could not say about God’s attributes that they were eternal or contingent.

Besides these main themes, *kalam* touched on related issues such as whether God could be seen in the hereafter (with the Mu‘tazila rejecting this, while their opponents affirmed beatific vision), the nature and limits of faith, whether hellfire and paradise were everlasting, and the nature and limits of God’s knowledge, will and power. Starting with ‘Allaf, some philosophical themes were introduced into *kalam*, in particular the discussion of such questions as the nature and classification of knowledge and the nature of movement, bodies and things. It even went on to discuss questions belonging to other sciences, such as biology, psychology and chemistry, as well as various logical investigations. However, this expansion of the scope of *kalam* coincided with its decline and did not lead to significant advances in any of these areas.

4 Methodological tendencies

Kalam generally dealt either with attempting to justify religious beliefs to reason, or with employing reason to draw new conclusions and consequences from these beliefs. Its doctrines comprise three major components: the articulation of what a school regarded as fundamental beliefs; the construction of the speculative framework within which these beliefs must be understood; and the attempt to give coherence to these views within the accepted speculative framework.

The various schools of *kalam* agreed with the traditionalists in accepting the authority of texts as the basis of the first component. They disagreed, however, about the extent to which these texts should be subjected to ‘rational’ analysis. Traditionalists had always suspected that the ‘reason’ being referred to was in fact the suspect intellect of infidel heretics; why else would a believer want to drag the articles of faith in front of the court of human reason, fallible and limited as it was? The traditionalist suspicion of non-Islamic influences behind every early *kalam*-ist ‘heresy’ has been reproduced by modern researchers, who seek an alien origin for every idea expressed in *kalam* (see Orientialism and Islamic philosophy). However, the impact of non-Islamic influences on the evolution of the schools of *kalam*, though undeniable, could easily be exaggerated. Many of *kalam*’s early themes, such as the status of the sinner or the question of political legitimacy, appear to have arisen within a purely Islamic context.

Regarding the second component, the speculative framework, the early groups did not erect elaborate systems. It is with the Mu‘tazila that we find the first attempt to construct such a system, based on their five principles (divine unity, divine justice, divine warnings, the intermediary status and the enjoining of virtue and discouragement of vice). The Mu‘tazila also brought with them an attitude of absolute confidence in human reason and a consequent lack of reverence for the authority of texts, which they regularly challenged.

The third component, the cohesion of views within the speculative framework, also came into prominence with the Mu‘tazila, who tried to systematize the body of religious beliefs and harmonize its components, provoking intense controversy as they attempted to reinterpret key elements of orthodoxy in order to achieve this. The attempts at systematization inevitably led to the raising of philosophical questions. Later Mu‘tazilite thinkers, such as al-‘Allaf and Ibrahim al-Nazzam (d. AH 231/AD 846), reflected in their theses the influence of translated Greek philosophical texts and propagated a worldview influenced by Hellenistic speculation (see Greek philosophy:

impact on Islamic philosophy). The Ash‘arite school, especially al-Juwayni and al-Ghazali, formally introduced the tools of Aristotelian logic into the methodology of *kalam* (see Logic in Islamic philosophy).

This introduction of philosophical themes and methods and the employment of formal logic in the Aristotelian tradition represented a significant development in *kalam*. Prior to that, *kalam* arguments had used textual and linguistic analysis as their central tools. However, in spite of these forays into philosophical speculation and the employment of Aristotelian logic, *kalam* remained firmly anchored in a specifically Islamic framework. Authoritative texts were routinely cited to clinch an argument, while an accusation of heresy was thought to be a conclusive refutation of any argument.

Even without the help of philosophy, however, Ash‘arism brought to *kalam* a trenchant scepticism that had a healthy impact on the field of rational argument. This scepticism was carried to great lengths by al-Ghazali, who used it to demolish the confused Neoplatonism of the Hellenizing philosophers. This approach had the potential to contribute much more to the advancement of knowledge than the dogmatic reiteration of philosophical theses, but that potential was not to be realized because the *kalam* practitioners were more interested in demolishing their opponents' arguments than in constructing viable alternatives.

5 Later evolution and decline

The decline of *kalam* proceeded apace from the fifth century AH (eleventh century AD), settling by the ninth century AH (fifteenth century AD) into ossified dogmatic texts that, to paraphrase al-Ghazali, taught the dogma as well as its formal 'proof', which was not the same thing as proving it to be true. This decline of *kalam* became too apparent to ignore even by its practitioners. Al-Iji comments that the aversion to the discipline in his time meant that engaging in it had become 'among the majority a reprehensible thing' (*al-Mawaqif*: 4). Ibn Khaldun, another Ash‘arite writing during the same period (*c.* AH 779/AD 1377), deplored the fact that *kalam* had deteriorated and become confused with philosophy, on top of being redundant because the heresies it was meant to combat had become extinct.

However, *kalam*'s problem was not so much its fusion with philosophy as its failure to evolve into a fully-fledged philosophical system with its own complete frame of reference. The possible evolution in this direction had been interrupted by a number of factors. First, there was the rift that developed between *kalam* as a discipline and philosophy proper; this was caused in part by the decline of the Mu‘tazila, the natural allies of philosophy. In addition, the failure of the Mu‘tazila to develop a common language with their opponents, thus turning *kalam* into a kind of sectarian pursuit rather than a discipline, was duplicated by the philosophers. The quasi-religious reverence shown by early Muslim philosophers to Greek texts put them at odds with mainstream thought, causing them to behave like just another sect. This limited the interaction between *kalam* and philosophy, as each treated its basic principles and texts as 'sacred' rather than as theses which could themselves be questioned. The rise of philosophy thus came both at the expense of *kalam* and in opposition to it, and this antagonism damaged both.

Kalam was also undermined by the rise of pro-traditionalist tendencies within the discipline itself. It was difficult to reconcile vigorous rationalist discourse with the traditionalist position, which discouraged questioning in many key areas and even counselled the acquiescence in apparent contradictions. At another level, the resurgence of traditionalism under Ahmad ibn Hanbal and subsequent revivals under Ibn Taymiyya and his disciples was anti-*kalam*, rejecting not only its theses but its methods as anathema. Rearguard actions fought by Ash‘arite and Maturidi scholars of the fifth to eighth centuries AH (eleventh to fourteenth centuries AD), failed to stem this tide and revive *kalam*. Finally, complementing the effect of traditionalism was the rise and popularity of Sufi mysticism (see Mystical philosophy in Islam). Although opposed by traditionalism, Sufism was also anti-rationalist and had also grown at the expense of *kalam* and philosophy.

With all these powerful forces deployed against it, the decline of *kalam* was inevitable. The early schools of *kalam* all became extinct, but traces of their teachings remain embedded within the doctrines of the six main schools of Islamic jurisprudence. The two main Shi‘a schools (the Ithna ‘Ashriyya and Zaydiyya) have inherited some aspects of Mu‘tazilite rationalism and doctrines. Shi‘ism has also been more successful in assimilating Sufi tendencies and more reconciled to philosophical discourse. The Hannafiyya became closely associated with the Maturidi school of *kalam*. The Shafi‘iyya espoused Ash‘arism as a general rule, as did the Malikiyya, although with less enthusiasm. The Hanbalites favoured an anti-rationalist and anthropomorphic position, distrusting *kalam*

altogether.

6 Conclusion

The manifest failure by the various schools of *'ilm al-kalam* either to create for itself a secure niche among the religious sciences, or to attain the status of a philosophical system independent of religious dogma, was not merely the result of the arrogant elitism of the Mu'tazila and their political opportunism. A deeper malaise afflicted the rationalist schools, reflected in their methodological confusion and, simultaneously, their militant dogmatism. Ironically, it was left to the traditionalist theologians, notably al-Ash'ari, al-Ghazali and Ibn Taymiyya, to introduce some healthy scepticism into the discourse by revealing some of the more glaring self-contradictions of the rationalist dogmas. However, the traditionalists not only inherited some of the confusion of their opponents, they also added some of their own.

An interesting example of this confusion was the uncritical acceptance by all schools of *kalam* of the Neoplatonic premise that the perfection of God as an eternal being meant that he could not be the locus of accidents (*hawadith*), while rejecting its logical consequence: God's remoteness from his creation and the impossibility of his day-to-day involvement with it. The confusion which this self-contradiction generated was then cited by many as a proof of how inadequate reason was in dealing with matters of faith. The choice offered the community was thus between rationalists who discredited themselves by their manifest errors, and traditionalists who exploited these errors and confusion to discredit rational thought as such.

The attacks of self-doubt brought by turmoil of modern times have created an atmosphere for a revival of Islamic theology and philosophy (see Islamic philosophy, modern). The pioneers of the modern Islamic revival, such as Jamal al-Din al-Afghani and Muhammad 'Abduh, tried to revive Islamic philosophy and *kalam*; al-Afghani indeed insisted that the revival of philosophy was an indispensable precondition for any Islamic revival. A century later, the tides of revival have drowned all attempts at philosophizing. On the face of it, the vibrancy and capacity for self-regeneration of the Islamic faith seem to be proportionately resistant to the emergence of systematic theologies and philosophies.

However, in spite of the self-satisfaction on the part of orthodoxy, on the grounds that history has condemned the systems rejected by Islam as fatally flawed and confused, there can be no substitute for setting up a viable worldview and a defensible theology, which would remain fallible and incomplete but still an essential guide for life. It would seem that if Islam is to continue as a living system, *'ilm al-kalam* (or something like it) may need to be revived, so that progress towards Muslim self-understanding, interrupted some six centuries or so ago, can be resumed.

See also: Ash'ariyya and Mu'tazila; Greek philosophy: impact on Islamic philosophy; Islamic fundamentalism; Mystical philosophy in Islam; Political philosophy in classical Islam; Religion, history of philosophy of; Religion, philosophy of; Revelation; Soul in Islamic philosophy

ABDELWAHAB EL-AFFENDI

References and further readings

Abdel Haleem, M. (1996) 'Early *Kalam*', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 5, 71-88.(Description of some of the variety among early theologians in Islam.)

'Abduh, M. (1954) *Risalat al-tawhid (Treatise on Divine Unity)*, Cairo: Dar al-Manar; French trans. by B. Michel and M. Abdel Razik, Paris: Librairie Orientaliste Paul Gauthner, 1925.(An early modern attempt at reformulating *kalam* textbooks.)

Anawati, G. and Gardet, L. (1950) *Introduction à la théologie musulmane (Introduction to Muslim Theology)*, Paris: Vrin.(One of the most thorough discussions of *kalam* in a Western language.)

Corbin, H. (1993) *History of Islamic Philosophy*, trans. L. Sherrard, London: Kegan Paul International.(While most introductions tend to neglect Shi'i contributions to *kalam*, this one redresses the imbalance, with an extensive bibliography.)

al-Farabi (*c*.870-950) *Ihsa' al-'ulum (Enumeration of Sciences)*, ed. A. González Palencia, with Spanish translation (*Catálogo de las ciencias*), Madrid: Maestre, 1932. (A survey of the state of learning in the fourth century AH (tenth century AD), offering an assessment of *kalam* in its heyday from a philosophical perspective.)

Farrukh, O. (1979) *Tarikh al-fikr al-'arabi ila ayyam Ibn Khaldun (The History of Arab Thought up to the Time*

of Ibn Khaldun), Beirut: Dar al-‘Ilm li’l-Malayin.(A general introduction to *kalam* and Islamic philosophy with a comprehensive bibliography of Arabic sources.)

al-Ghazali (*c.*1107) *al-Munqidh min al-dalal (The Deliverer from Error)*, ed. and trans. F. Jabre, *Erreur et déliverance*, Beirut: al-Lajnah al-Lubnaniyyah li-Tarjamat al-Rawai‘, 1965; trans. W. Montgomery Watt, *The Faith and Practice of al-Ghazali*, London: George Allen & Unwin, 1953.(A celebrated intellectual autobiography by the most famous Ash‘arite author, telling of how he had sought salvation in *kalam*, in vain.)

al-Ghazali (*c.*1111) *Iljam al-‘awam ‘an ‘ilm al-kalam (Restraining Commoners from Kalam)*, ed. M. al-Baghdadi, Beirut: Dar al-Kitab al-‘Arabi, 1985.(Al-Ghazali’s final attacks on *kalam*, arguing that it would be harmful for most people to indulge in it.)

Goldziher, I. (1910) *Vorlesungen über den Islam (Introduction to Islam)*, Heidelberg; trans. A. Hamori and R. Hamori, *Introduction to Islamic Theology and Law*, Princeton, NJ: Princeton University Press, 1981.(An early introduction that still stands out in spite of minor flaws that are a feature of its times.)

Hallaq, W. (1993) *Ibn Taymiyya Against the Logicians*, Oxford: Clarendon Press.(A translation of a summary of Ibn Taymiyya’s *al-Radd ‘ala al-mantiqiyyin* with a good introduction to the latter’s crusade against logic and philosophy.)

al-Iji (before 1355) *Al-mawaqif fi ‘ilm al-kalam(Book of Stations on Kalam)*, Cairo: Dar al-’Ulum.(Textbook which reflects the culmination of the rapprochement between philosophy and *kalam* in later Maturidi thought.)

Macdonald, D. (1903) *Muslim Theology, Jurisprudence and Constitutional Theory*, New York: Scribner’s.(A scholarly introduction that deserves its status as a classic in the field.)

Madelung, W. (1985) *Religious Schools and Sects in Medieval Islam*, London: Variorum Reprints.(A reprint of essays published by the author over many years covering various aspects of *kalam* and the evolution of religious thought in Islam.)

Montgomery Watt, W. (1948) *Free Will and Predestination in Early Islam*, Edinburgh: Edinburgh University Press.(A discussion of the beginning of theological thinking in Islam.)

Montgomery Watt, W. (1962) *Islamic Philosophy and Theology*, Edinburgh: Edinburgh University Press.(A broader introduction to *kalam* and Islamic philosophy.)

Morewedge, P. (ed.) (1979) *Islamic Philosophical Theology*, Albany, NY: State University of New York Press.(A collection of essays by leading writers in the field.)

Muhajarani, A. (1996) ‘Twelve-Imam Shi‘ite Theological and Philosophical Thought’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 8, 119-43.(Discussion of the main principles of Shi‘ite theologians.)

al-Nashshar, A. (1984) *Manahij al-bahth ‘ind mufakkiri al-Islam (The Research Methodologies of Muslim Thinkers)*, Beirut: Dar al-Nahda al-‘Arabiyyah.(An attempt to shed light on the evolution of methodological approaches in philosophy, theology and science in medieval Islam.)

Pavlin, J. (1996) ‘Sunni *Kalam* and Theological Controversies’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 7, 105-18.(Account of some of the most important Sunni theologians.)

al-Tusi (*c.*1270) *Talkhis al-muhassal (Summary of [al-Razi’s] Muhassal)*, ed. T.A. Sa’d, Beirut: Dar al-Kitab al-‘Arabi.(A summary and commentary on al-Razi by a leading Shi‘ite scholar who favoured Neoplatonist philosophy more than did al-Razi and other practitioners of *kalam*.)

Wensinck, A. (1932) *The Muslim Creed*, Cambridge: Cambridge University Press. (An early introduction which influenced a large number of writers on the subject.)

Israeli, Isaac ben Solomon (*c*.855-955)

A pioneering Jewish philosopher and a physician, Isaac Israeli was among the very first medieval Jewish writers to formulate a philosophy employing Greek sources. He based his metaphysics chiefly on the Neoplatonic theory of emanation. However, like many later Jewish philosophers, Israeli understood God as a voluntary agent who acted through power and will. He combined a commitment to the traditional doctrine of creation ex nihilo *with the idea of emanation: the first level of the ontic hierarchy was formed by divine creation but lower levels emerged by emanation. This original synthesis was Israeli's most prominent philosophical innovation.*

Israeli's anthropology was at base largely Neoplatonic, teaching that the soul can ascend the emanatory ladder back to God. The first Jew to give a psychological account of prophecy, Israeli, unlike later Jewish thinkers, did not discuss at length the relation between his philosophical ideas and traditional sources. His pioneering work did not have a decisive impact on later generations of Jewish thinkers; he was known in Latin Europe, but mainly for his medical writings.

1 Life and career

Little is known about the life and background of Isaac ben Solomon Israeli, a Jewish philosopher and physician who lived in North Africa. He was one of the earliest medieval Jewish philosophers and a distinguished physician in the court of the Fatimid caliph in Kairouan (modern Tunisia). His disciple, Dunash ibn Tamim, mentions a philosophical correspondence with Saadiah, later known as Saadiah Gaon - who, like Israeli, was of Egyptian origin - in which Saadiah presented questions to his older contemporary.

In spite of his long life, Israeli never married. He is said to have remarked that his books would be a greater guarantor of immortality than children. His numerous writings include such philosophical works as the *Kitab al-hudud wa 'l-rusum (Book of Definitions)*, *Kitab al-ustuqussat (Book on the Elements)*, *Kitab al-jawahir (Book of Substances)*, *The Book on Spirit and Soul* and the *Chapter on the Elements*, and such medical works as *Kitab al-Hummayat (Book of Fevers)*, *Kitab al-adwiya al-mufrada wal-aghdiya (Book of Foodstuffs and Drugs)*, and *Kitab al-baul (Book of Urine)*. Although written in Arabic, these books are known to us mainly in Hebrew and Latin translations.

Israeli was not on the whole an original philosopher; he tended to the eclectic, but he was one of the first Jewish thinkers to recognize the value of the Greek philosophical tradition, known to him through Arabic translations. Adopting ideas from both Aristotle and Plato, he assumed agreement between the two, as did his Muslim contemporary al-Farabi. Israeli was most directly influenced by al-Kindi, the first major Muslim philosopher, and by an unknown Neoplatonic work, traces of which are found in later Jewish philosophical writings. He made concerted efforts to integrate ideas from his sources into a unified, rational outlook, but did not attempt explicitly and systematically to reconcile philosophical methods with Jewish tradition, as Saadiah later sought to do.

Despite his varied sources and his quest to bridge opposing viewpoints, Israeli is clearly an early Jewish Neoplatonist, as can be seen in his emanationism, his ideas of illumination and unification with the divine light, and the general Neoplatonic cast of his imagery (see Neoplatonism). When Jewish thought turned toward Aristotelianism, Israeli's views were largely forgotten. Maimonides dismissed him as 'only a physician', and called his books 'futile and vain'.

2 Metaphysics

The central questions of Neoplatonism, as laid out by Plotinus, were first, how and why plurality entered the world from the One, that is, God; and second, what was the source of matter if the One is absolutely incorporeal (for if it were corporeal, it would not be absolutely One). The Neoplatonists found their answer in the idea of emanation: the One timelessly produces a series of hypostases. Each flows from that which precedes it in a procession that transcends the limits of any merely natural process, and seemingly any volitional process as well. The further a hypostasis is from the Source, the more plurality enters into it until, ultimately, the level of sheer matter is reached. Plotinus follows emanation from the One, to the divine and universal Mind, to the world Soul, and then to Nature. However, the Neoplatonic tradition that reached Israeli proposed a more complex scheme, partly as a result of efforts to introduce a dimension of voluntarism into the process of emanation (see Voluntarism, Jewish).

Israeli envisioned the following series:

The Creator (with will and power);
First form (wisdom) and first matter - intellect;
The rational soul;
The animal soul;
The vegetative soul;
Nature (= sphere = heavenly bodies);
The four elements.

God acts as a voluntary creator. Power and will are aspects of God, not independent hypostases (as contrasted with the view of the later Jewish Neoplatonist Solomon ibn Gabirol). In traditional Neoplatonism, all entities emanate from the One (see Neoplatonism), but Israeli taught that the first form and the first matter were created *ex nihilo*. The remaining spiritual substances arise by emanation from these. Hence there is a double creation: a voluntary act of God at the start of creation, and emanation as its continuation. This unconventional and original theory clearly seeks to to unite Neoplatonic premises with Jewish tradition.

Israeli deviated from other Plotinian ideas in describing the remaining hypostases too, although these deviations were generally not original to Israeli but were derived from the Neoplatonic tradition. For example, Israeli located matter in the beginning of creation. He also identified the first form with the divine wisdom which together with the first matter, forms the intellect. He divided the hypostatic soul, in accordance with the Aristotelian scheme, into rational, animal and vegetative souls (see Aristotle), and replaced Plotinian nature, with the Sphere, that is, the heavenly bodies from which the four elements proceed.

3 Anthropology

Israeli adopted the Neoplatonic doctrine of the return of the soul through purification and illumination. The goal of the ascent is unification with the supernal light, not directly with God. Some human beings, whose evil nature and impurity prevent them from ascending to the spiritual substances, are left below the sphere, tormented by heat and fire. Thus Israeli combines rabbinic teachings concerning heaven and hell with Neoplatonic doctrines about the destiny of noble and base souls.

Analysing the nature of prophecy, without direct reference to the Bible or the rabbinic sources, Israeli used Neoplatonic accounts of the intellect and the soul to forge the first Jewish psychological account of prophecy (see Prophecy). A prophet, who is also a philosopher, seeks to direct people toward righteousness. Prophetic inspiration is manifested mainly in dreams, when the powers of the soul apprehend spiritual objects. Prophets transmit what they receive to the people, sometimes clearly, sometimes through obscure expressions which need to be explained. Although he did not explicitly discuss such distinctively Jewish topics as the commandments or the revelation at Mount Sinai, Israeli does obliquely indicate his sensitivity to these theological themes.

The Jewish Neoplatonists who followed Israeli were to a certain extent influenced by him, as were the Kabbalists of the school of Gerona (see Kabbalah). Yet it is difficult to conclude that Israeli had a decisive influence on later Jewish thought. His work did not address the core issues of theology that occupied so many other Jewish thinkers, and his Neoplatonism came to seem irrelevant or primitive. He did, however, win the immortality he sought through his books, which are still studied a thousand years after his death.

See also: Neoplatonism; Saadiah Gaon; Voluntarism, Jewish

DANIEL J. LASKER

List of works

Israeli, Isaac ben Solomon (before 955) Collected Works, ed. A. Altmann and S.M. Stern, *Isaac Israeli. A Neoplatonic Philosopher of the Early Tenth Century: His Works Translated with Comments and an Outline of his Philosophy*, London: Oxford University Press, 1958.(English translations of Israeli's major philosophical works, and a thorough discussion of his philosophy, with extensive bibliography.)

Israeli, Isaac ben Solomon (before 955) *Kitab al-hudud wa 'l-rusum (Book of Definitions)*, German and Hebrew edition in *Festschrift zum 80. Geburtstag M. Steinschneiders*, Leipzig: Hartwig Hirschfeld, 1896.(An early, not

entirely accurate edition.)
Israeli, Isaac ben Solomon (before 955) *The Book on Spirit and Soul*, ed. M. Steinschneider in *Ha-Karmel*, 1871, 400-5.(An early edition.)
Israeli, Isaac ben Solomon (before 955) *Chapter on the Elements*, ed. A. Altmann, ‘Isaac Israeli’s "Chapter on the Elements" (Ms. Mantua)’, *Journal of Jewish Studies* 7, 1956: 31-57.(Relatively recent edition of this short work.)
Israeli, Isaac ben Solomon (before 955) *Kitab al-jawahir (Book of Substances)*, ed. S.M. Stern, ‘The Fragments of Isaac Israeli’s "Book of Substances"’, *Journal of Jewish Studies* 7, 1956: 13-29.(Only fragments of this work have survived.)
Israeli, Isaac ben Solomon (before 955) (The Book on the Elements), ed. S. Fried, *The Book on the Elements (Sefer Ha-Yesodot)*, Drohobycz, 1900.(A Hebrew version of Israeli’s longest philosophical work.)

References and further reading

Altmann, A. (1979) ‘Creation and Emanation in Isaac Israeli: A Reappraisal’, in I. Twersky (ed.) *Studies in Medieval Jewish History and Literature*, Cambridge, MA: Harvard University Press, 1-15.(Rebuts an attack on Altmann’s understanding of Israeli’s doctrine of creation.)

Italy, philosophy in

Since the Renaissance, Italian philosophy has been rooted in the humanist and historical tradition, stemming from the rediscovery of Greek philosophy in the fifteenth century and the resumption of classical studies at that time. However, the momentous cultural and religious transformations of sixteenth-century Europe caused a reaction which greatly restrained the innovative impact of Renaissance thought. Living through these two periods, Bruno, Campanella and Galileo experienced all the conflicts of the Counter-Reformation, when philosophy especially felt the weight of the Roman Catholic Church which exercised a tight control on the culture of the time. The best traditions of the Renaissance were inherited and maintained solely by Vico, the most important figure of this latter period.

As the spirit of the Counter-Reformation declined, Italian philosophy acquired a new vigour in the eighteenth and nineteenth centuries. This occurred especially in the north of the country where contacts with modern European culture influenced a political situation dominated by the question of national unity (a problem debated since the time of Machiavelli). Philosophical thought during this period tended to be strongly political (as in the philosophy of Cattaneo), although it also assumed a deeply religious tone (for example, Mazzini and Gioberti). These themes were united in Hegelian Idealism, promoted by Spaventa among others. The neo-Idealists of the early twentieth century, such as Croce and Gentile, distinguished themselves by assuming a political role and by opposing empiricism and positivism, philosophical heirs of the Enlightenment which were traditionally considered alien although popular at that time.

The aftermath of the Second World War brought to gradual dissolution both philosophical and political neo-Idealism, a movement which was rooted especially at the universities of Rome and Naples; it also saw the rise of Marxist thought (Gramsci) and Existentialism, especially at the universities of Bologna, Milan and Turin. Among the most important figures in contemporary Italian philosophy we find Bobbio (political philosophy), Pareyson, Vattimo and Eco. While both French and German philosophy still play a major role in the philosophical debate, in recent years increasing attention has been paid to Anglo-Saxon philosophy.

1 From Counter-Reformation to Enlightenment

The great philosophical blossoming of the fifteenth and sixteenth centuries, which in Italy had started with the revival of the *humanae litterae*, came to a decisive and lasting halt with the Counter-Reformation movement which began with the Council of Trent (1545-63). The new cultural climate - inspired by an idea of humanity and human destiny which bore little resemblance to the spirit of liberty, creativity and loyalty typical of philosophers and artists such as Pico della Mirandola, Marsilio Ficino, Leonardo da Vinci and Michelangelo - prevailed in the second half of the sixteenth century and especially in the seventeenth century. Indeed, the general lack of cultural confidence expressed itself as censorship, especially of the three main philosophers of the sixteenth and seventeenth centuries: Bruno and Campanella (both heirs to Telesio's naturalistic tradition, combined with Neo-Platonist, magical and cabalistic ideas), and Galileo.

Having fled Italy to avoid being persecuted by the Church for his philosophical ideas, Giordano Bruno lived in England between 1583 and 1585, where he wrote and published works which were later gathered under the title of *Dialoghi Italiani (Italian Dialogues)*. When he returned to Italy, he was condemned by the Church authorities to die at the stake in 1600. Central to his philosophy is the idea of the unity and infinity of nature which the Copernican system, through Neo-Platonist metaphysics, attempts to demonstrate. The universe is one, open and infinite, and may contain many worlds. The unity of nature means that it contains within itself the principle of its own movement. Bruno thus adapts, and uses against Aristotle, the hylozoistic idea of the Presocratics to which he often refers. Corresponding to this naturalistic concept is an ethical vision that unites contemplation and activity through a radical criticism of medieval virtues and the defence of worldly virtues. The *eroico furore* (heroic fury) is the synthesis of theoretical knowledge and practical activity which allows the individual to grasp the dynamical unity of nature, a composition of opposing tensions (*coincidentia oppositorum*) that Cusano had previously attributed solely to God.

Tommaso Campanella, a contemporary of Bruno, further undermined the influence of Aristotelian philosophy (still prominent at the University of Padua). His epistemology is a combination of Augustinism and empiricism:

knowledge is sensation, but its justification is grounded on the internal awareness of self (*sensus inditus*) By referring to Augustine (§§4, 5), Campanella anticipates Descartes. His naturalism is full of references to magic (*De sensu rerum et magia*, 1620), on which he builds his religious, ethical and political vision - the ultimate goal of magic is ‘to give laws to men’. In his *Città del sole (City of the Sun)* (1623) Campanella describes the organization of an ideal city (like Thomas More’s *Utopia* or Plato’s *Republic*). The three principal officers (Captain, Wisdom and Love) and supreme head (Metaphysic or Sun) are personifications of metaphysical and religious principles derived from the synthesis of Christianity with a natural religion of a strongly naturalistic and magical character, which in its essence tries to be ‘rational’. The extent of Campanella’s awareness of the developments in scientific thought is evident from his *Apologia pro Galilaeo* (1622).

Galileo is credited with having been the first to formulate what has gone down in history as the ‘Galilean scientific method’ (see Galilei, Galileo). However, unlike Bacon, he never actually explicitly formulated this method - it was derived from his scientific treatises, including the *Saggiatore (The Verifier)* (1623), the *Dialogo sui massimi sistemi (Dialogue on the Chief World Systems)* (1632), and *Discorsi e dimostrazioni matematiche intorno a due nuove scienze (Dialogues and Mathematical Demonstrations about Two New Sciences)* (1638). As a result of this method, using some fundamental concepts (such as causality) and ideas from the traditions of Ockham and of Naturalism (which had undermined medieval ways of thought), Galileo made scientific discoveries of enormous importance in both astronomy and in physics. Unlike Bruno and Campanella, he based his study of nature on mathematics and excluded anything metaphysical or magical. Only what is measurable (what Locke was to call ‘primary qualities’) can be the subject of investigation: this establishes a distinction between religion/philosophy and science, and guarantees the autonomy of the latter from the former.

While Galileo was not a philosopher in the strict sense of the word, the cultural importance of his personal vicissitudes, and his trial and conviction by the Inquisition, is especially significant. Bruno and Campanella remained isolated in the tragedy of their personal circumstances because the philosophy of each of them was closely tied to his personality. In the case of Galileo, the condemnation did not manage to stem the spread of his ideas, because they were based on a new methodology which was destined to dominate on its own merits. The immediate influence of Galileo was therefore mainly confined to the scientific sphere, where his pupils initiated an important series of studies, especially in the field of mathematics. Bonaventura Cavalieri, for example, sought to replace the method of indivisibility with the method of exhaustibility in the calculation of volumes and areas: Evangelista Torricelli, possibly Galileo’s greatest pupil, built on these discoveries. Vincenzo Viviani is known for his work on geometry, and Michelangelo Ricci was the author of a book on the problem of tangents and of maxima and minima.

The Counter-Reformation movement sought to reunite philosophy, science and religion, whose compartmentalization had given rise to the new science based on experimentation (Galileo himself tried throughout his life to show that the three disciplines were compatible with each other). Campanella more than anybody had tried to achieve this goal: the *Città del Sole* is in fact the idealization of a world in which religion, science, philosophy and politics find a harmonious synthesis. Less idealistic, on the other hand, is the political thought of the Jesuit Giovanni Botero, author of *Della ragion di stato (About the Reason of State)* (1589). In this work he opposes Machiavelli by attempting to bring ethics back into politics, believing justice to be the prince’s supreme virtue and claiming that politics is ultimately subordinate to the counsels of religion. At this time Alberico Gentili, professor at Oxford and one of the first theoreticians of natural law, maintained in his *De jure belli* the existence of rights that apply even in times of war (respect for prisoners, women and children and so on), holding that the only just war is carried on in self-defence.

Continuing Counter-Reformation preoccupations meant that Italian philosophy of the seventeenth and eighteenth centuries was limited to repeating scholastic arguments. The thought of Descartes, Hobbes (who had visited Galileo during his exile at Arcetri), as well as that of Leibniz and Berkeley (both of whom had also visited Italy), had little impact (but see Fardella, M.). The first figure of note in this period took an open stance *against* European thought, continuing in that tradition of historical Humanism typical of Italian thought (especially of the Neapolitan school down to the present day). Gianbattista Vico was a seventeenth-century humanist who, like all other humanists, regarded the Middle Ages a time of barbarism and considered the main path of knowledge to lie in the synthesis of philology and philosophy. In contrast to Descartes, Vico denies that evidence is the criterion of truth or that the *Cogito* could convince sceptics. In fact, one can only know what one has done oneself. Just as God

knows the universe because he created it, humans can know their own works, but they cannot know nature or know themselves, only their history.

The identity between explanatory knowledge of *x* and direct experience of the causes that have brought about *x* (the *Verum et factum convertuntur* hypothesis) is the great principle of the *nuova scienza* (new science) which justifies including Vico in the history of the contemporary hermeneutic movement. In his main works *De antiquissima Italorum sapientia* (1710), and in *Scienza nova* (which went to three editions), Vico outlines the foundations of historical science. Against Descartes' followers, he extols the superiority of both historical-juridical studies and of the rhetoric over the exact sciences. History he divides into three eras: the age of gods (when imagination plays an important part in the origin of language), the age of heroes, and the age of humans. These three phases are cyclical, so human history is a repetition of historical phenomena, guided by a providence that makes it 'a reasoned civil theology of the divine providence'.

Mario Pagano, a follower of Vico, was representative of the Neapolitan Enlightenment that opened up to European culture from the middle of the eighteenth century. Traces of Rousseau are evident in his *Saggi politici (Political Essays)*, just as echoes of Montesquieu can be found in Gaetano Filangieri's *Scienza della legislazione (Science of Legislation)*. The high opinion in which these men are held by J. Schumpeter (he gives a place of honour in the field of pre-Smith systemizers to Italian economists of the eighteenth century), is due to their analytical capacity and to their interest in a civil economy which reconciles 'public good' and 'utilitarian happiness' in the concept of 'public happiness'. Many of these economists were in public employment as ambassadors and administrators, for example L.A. Muratori, author of *Della pubblica felicità (Concerning Public Happiness)*; Antonio Genovesi, who in his *Meditazioni filosofiche sulla religione e sulla morale (Philosophical Meditations on Religion and Morals)* expounded an ethical concept inspired by Condillac and Helvetius; and Ferdinando Galiani, author of the treatise *Della moneta (On Money)*.

The influence of the French Enlightenment was strong in the Neapolitan School (see Enlightenment, continental). The Italian translation of the Encylopédie was published in Lucca and Livorno between 1758 and 1779. Lombardy was another centre for the diffusion of Enlightenment thought. Its geographical position and economic and political situation (it was part of the Austro-Hungarian Empire, whose rulers Maria Theresa and Joseph II were among the more enlightened sovereigns), placed it in a strategic position for contact with the rest of Europe. Gathered around the 'Società dei Pugni' (Society of Fists) founded in 1761, whose mouthpiece was 'Il Caffé' (a newspaper inspired by England's 'Spectator'), the exponents of the Enlightenment in Lombardy were, like those in Naples, mainly interested in economic and political studies. Pietro Verri's *Discorso sulla felicità (Discourse on Happiness)* (1763) proposed an ethical theory which merged sensationalistic and utilitarian themes; Cesare Beccaria, author of the well-known *Dei delitti e delle pene (On crimes and punishments)* (1764) which was condemned and banned by the Catholic Church in 1766, questioned the justice and use of the death penalty and of torture. For him punishment must defend public welfare - punishment is just only when it guarantees the inviolability and freedom of the people.

2 The nineteenth century

During the nineteenth century, philosophy in Italy was closely connected to political events in which most philosophers played an active part. The influence of Vico was especially strong and, due to Vincenzo Cuoco, permeated even the Milanese school which traditionally tended more towards the Enlightenment. The attention to historical detail and to Italian tradition which are evident in Cuoco's work *Platone in Italia (Plato in Italy)* (1805) was also characteristic of the work of Carlo Cattaneo, a proponent of Federalism along with Giuseppe Ferrari. But Cattaneo's work is more philosophically aware of wider issues in European culture. He was the author of historical and economic studies (in 1844 he predicted the famine that was to hit Ireland during 1845-8) and the driving force behind the group based around the journal *Il Politecnico*, which he also managed on many occasions. His original theory of knowledge, outlined in the *Psicologia delle menti associate (Psychology of associated minds)* (1859-66), justifies Kant's theory of phenomena in terms of historical development and the efficacy of science. Works closer to Enlightenment thought from this time include those of Francesco Soave, critic of Kantianism; Melchiorre Gioia who, in *Filosofia della statistica (Philosophy of Statistics)* (1826), worked out an economic theory influenced by Bentham; and Giandomenco Romagnosi who put forward a naturalistic conception of society.

The major European philosophies of the time - Kantianism and Idealism - were combined in Italy with elements

derived from the long-standing humanist and religious traditions. Pasquale Galluppi contributed to the spread of the philosophy of Kant, especially in Naples (which later became an important centre for Hegelian philosophy). In his *Saggio filosofico sulla critica della conoscenza (Philosophical essay on the critique of knowledge)* (1819), Galluppi maintained the necessity of going beyond sensationalism (such as that of Condillac) and Kantian subjectivism, but his critics charged him with falling into an idealism more inconsistent than that of Kant.

Strongly spiritualistic elements are found in the romantic idealism of authors such as Giuseppe Mazzini, Antonio Rosmini and Vincenzo Gioberti. Mazzini advocated ideas that inspired the struggle for national unity (*Risorgimento*), and the slogan '*Dio e popolo*' (God and the people) summarizes his thought: supreme authority is of the people, but the aims they pursue are inspired by God. The Italian people, he believed, had been called to begin a new era founded not on rights, like the French Revolution, but on duties, on the awareness that each person belongs to Humanity. Mazzini, an opponent of Socialist International, proposed a lay spiritualist philosophy and attempted to give it European significance by creating an association called '*Giovane Europa*' (Young Europe), following the domestic '*Giovane Italia*' movement. The philosophy of Antonio Rosmini is even more strongly spiritualist (see Rosmini-Serbati, A.). A priest, he elaborated a philosophy (suspected of being heterodox) which was meant to supersede that of Kant. He developed an ontology founded on the primacy of the idea of being as prior to any other knowledge or category. The sole category is possibility and knowledge is rooted in consciousness of self (*Nuovo saggio sull'origine delle idee (New essay on the origin of ideas)* (1830); *Logica (Logic)* (1854)).

If, in spite of his polemics, Rosmini is considered the Italian Kant, Vincenzo Gioberti is the Italian Hegel. A critic of Rosmini, Gioberti summarized his own philosophy with the formula '*L'ente crea l'esistente, l'esistente ritorna all'ente*' (being creates existence, existence returns to being). This describes a process of imitation and participation clearly inspired by Neoplatonism, which brought accusations of immanentism. Also a priest, Gioberti like Rosmini was motivated by an apologetical intention, but his Catholic philosophy does have some original points. Apart from strictly philosophical works such as *Degli errori filosofici di A. Rosmini (On the philosophical errors of A. Rosmini)* (1841) and *Protologia (Protology)* (1857), Gioberti is also remembered for his civil commitment and political theories. In *Del primate morale e civil degli italiani (On the moral and civic superiority of the Italians)* (1842-3) he supports a neo-Guelphian idea which sees the Roman Catholic Church as the guide of the Italian people, helping them to regain their historical pre-eminence. Rosmini's and Gioberti's Catholic spirituality met with hostility from the Church which, in the 1879 encyclical *Aeterni Patris* of Leo XIII, declared the philosophy of Thomas Aquinas to be the orthodox doctrine of the Roman Catholic Church. This date marks the beginning of neo-scholastic philosophy which became widespread in the twentieth century, especially at the Pontifical University of Rome and Catholic University of Milan.

The legacy of the Enlightenment continued, particularly in the positivist school in the North which comprised more scientists than philosophers, and which was interested in the 'sciences of life', by which was meant anthropology, psychology, sociology and so on. The major figure of this movement was Roberto Ardigò, an ex-priest and academic at Padua (a well-known centre for heterodoxy since the Renaissance). He wrote an essay on Pomponazzi and maintained a naturalistic positivism which was a cross between Spencer's Evolutionism (Spencer was the best-known positivist philosopher in Italy) and Bruno's Naturalism. For Ardigò, the fact is divine, the absolute to which every theory must be referred. Although not without a certain dogmatism, this theory is more radical than Spencer's because it refuses to recognize the unknowable. Ardigò's work on ethics, *La morale dei positivisti (The morality of positivists)* (1879), met with great success, especially in the field of criminal law, while his *Psicologia come scienza positiva (Psychology as a positive science)* (1870) is important in the development of psychology as an autonomous science.

Naples was the centre of Hegelianism in Italy, and two of the most influential Italian philosophers of the beginning of the twentieth century, Croce and Gentile, were educated there. Hegelianism in Italy is not regarded as an imported school of thought - rather, the roots of European philosophy are sought in the Italian and especially the Neapolitan tradition. Bertrando Spaventa's theory of the circulation of ideas, outlined in his *La filosofia italiana nelle sue relazioni con la filosofia europea (Italian philosophy and its relation to European philosophy)* (1862), maintains that the origin of modern philosophy is to be found in the Italian Renaissance: Spinoza, Kant, Fichte and Hegel he regards as disciples of Bruno, Campanella and Vico. This interpretation of history, in spite of its one-sidedness (it ignores Galileo and the Enlightenment), stimulated important historical studies, especially on the

Renaissance and on German philosophy which was seen as the link with the Renaissance. Such one-sidedness is typical of Italian philosophy of the nineteenth century, where Positivists and Idealists alike viewed theoretical progress in science with indifference, leaving thinkers of European stature practically isolated: these included Giuseppe Peano, for instance, who completed the work of Weierstrass and whose discoveries in mathematics and geometry place him among the greatest scientists of the century; and Federigo Enriques, the greatest representative of the Italian school of algebraic geometry and supporter of an epistemological theory similar to that of Lakatos (see Lakatos, I.).

3 The twentieth century

In the early twentieth century, Italian philosophy was dominated by a lively anti-positivist controversy, led by Idealist, Marxist and by Pragmatic thinkers. The major figures of neo-Idealism were Benedetto Croce and Giovanni Gentile, both committed to reforming Hegelianism. Initially close to Marxism, Croce later embraced a philosophy of Spirit clearly of Hegelian stamp. For the triad of the dialectic movement (thesis, antithesis and synthesis), Croce substituted a dialectic of opposites articulated in four distinct categories: aesthetics, theoretics, economics and ethics. The first two make up theoretical activity, the others the practical *(Logica come scienza del concetto puro (Logic as the science of pure concept)* (1905); *Filosofia della pratica, economia e etica (Philosophy of practice, economics and ethics)* (1909)). The distinct categories are mutually irreducible, though they are not independent. Croce is especially famous for his aesthetic theory, outlined in *Estetica come scienza dell'espressione e Linguistica generale (Aesthetics as the science of expression and general linguistics)* (1902), which was founded on the identification of both intuition and expression. He also provoked philosophical debate through the journal *La Critica*, a forum for cultural, civil and social discussion which was open to European influences.

If the Idealism of Croce is characterized by a devaluation of scientific thought (whose concepts he regarded as 'pseudo-concepts'), the Idealism of Giovanni Gentile classified science in the dogmatic manner of Fichte. The reform of Hegelianism proposed by Gentile in his *Teoria generale dello spirito come atto puro (General theory of the spirit as pure act)* (1916) and *Sistema di logica (System of logic)* (1917-22) eliminates any presupposition (both logic and philosophy of nature) of the life of the spirit. The life of the spirit is consequently reduced to pure act - beyond the actuality of thought there is nothing. As in Fichte, the ego is self-creation. Whereas Croce viewed philosophy as history in a wide sense (economic, political and so on), for Gentile philosophy is merely the history of thought, it is the history of philosophy and the history of the spirit is its own education. As Minister for Education in the Fascist Government Gentile had enormous influence on Italian culture. He was instrumental in the education reforms of 1924 and directed the *Enciclopedia Italiana*, one of the greatest productions of Italian culture from that time. Gentile's involvement with Fascism, however, caused a bitter rift with Croce, who had distanced himself from Mussolini. Similarly, Piero Martinetti was strongly anti-Fascist. He reinterpreted Kant's philosophy in Platonic and Spiritualistic terms and, arguing against positivism, he maintained the superiority of a philosophy whose result is faith, which he regarded as the most radical fruit of reason. Giovanni Vailati also argued against positivism: his pragmatism is independent of the corresponding American movement and is original in its attention to language through the study of semantic and syntactic structure, and the demand for logical rigour as a tool for conceptual clarification.

The prevailing climate of Idealism at the beginning of the twentieth century meant that Marxism had little support in Italy. It is thanks to Antonio Labriola that the philosophies of Marx and Engels became known, thereby opening a debate with Positivism and Idealism. A pupil of Spaventa, Labriola developed a philosophy whose central element is a social psychology which considers the ego not in an abstract sense but in the context of its historical condition. Only after the Second World War did Marxism permeate cultural life more deeply due to the publication of the works of Antonio Gramsci. A vigorous opponent of Fascism and a founder of the Italian Communist Party, Gramsci was imprisoned and wrote the majority of his works while thus incarcerated. These were published posthumously under the title *Quaderni dal carcere (Notebooks from Prison)* (1948-51). His philosophy is a philosophy of practice - in opposition to the speculative historicism of Croce, it eliminates all transcendency and thus is absolute Historicism and Humanism. The intellectual is not separated from the rest of society (a renaissance ideal still valid for Croce and Gentile), but rooted in it, an 'organic intellectual' with a fundamental role to play in the revolution. Gramsci reversed the Marxist relation between structure and superstructure. He regarded it as the duty of intellectuals to guide the masses to the revolution, to permeate civilian

society, and to undertake a role of cultural hegemony in the awareness that the Party is the ‘modern Prince’. Unlike Machiavelli, for Gramsci this prince is merely a body expressing a collective will. The influence of Lenin is evident in the demand for a synthesis between theory and practice, the duty to inaugurate a new society falling on intellectuals and on the Party (see Lenin, V.I.).

After the Second World War, Italian philosophy became more open to contemporary European culture, especially to German phenomenology and hermeneutic existentialism. This phase was dominated by the Turin school, where some of the major present-day Italian philosophers taught or were educated. Nicola Abbagnano was among the first Italians to develop existentialist themes, but was also known for his historical studies on ancient scepticism and Hume. His ‘positive’ existentialism, which is strongly influenced by Dewey’s Pragmatism, regards existence as a possibility or project within a finite horizon. The existentialism of Luigi Pareyson has personal and hermeneutical elements as outlined in his *Verità e Interpretazione (Truth and interpretation)* (1970): the human person is an organ of truth, which is interpretation, although not exhausting itself in interpretation. These themes were developed towards a theory of ‘formativeness’ in Pareyson’s aesthetics, whereby artistic activity becomes object and manner of its own action (*Estetica, Teoria della formatività (Aesthetics: A Theory of Formativeness)* (1954)). The final phase of his philosophy is strongly religious, offering a meditation on mythical and revelatory language.

Norberto Bobbio, a political philosopher initially closer to phenomenology, has moved towards Anglo-American trends, from the analytical philosophy expressed in his *Giusnaturalismo e positivismo giuridico (Natural law doctrine and legal positivism)* (1964) to the formalism of Kelsen which he revised in a functionalist manner more appropriate to the description of democratic societies in *Dalla struttura alla funzione (From structure to function)* (1977). Well-known for his anti-Fascist views and for his civic commitment (he was made a senator for life), Bobbio represents a new spirit of enlightenment. He has contributed to widening the horizons of political studies in Italy, freeing them from Idealism, and has contributed greatly to the theory of law with works such as *Teoria della norma giuridica (Legal Theory)* (1968), *La teoria delle forme di governo (Theory of the forms of government)* (1976), and *Destra e sinistra (Right and left)* (1994).

The founder of the phenomenological school in Italy is Enzo Paci whose thought centres mainly on Husserl’s ideas, trying to bridge the gap between the categorial and pre-categorial world. Umberto Eco, a pupil of Pareyson and professor of semiotics at Bologna, is well-known not only as a philosopher but also as a writer (*Il nome della rosa (The Name of the Rose)* is his most famous novel). Applying Pareyson’s aesthetic theories to modern art, Eco has developed a semiological theory which sees meaning as indefinite and open (*Opera aperta (Open Work)* (1962)), which does not mean absent, as he particularly points out in his recent works against American deconstructionism, including *I limiti della interpretazione (The Limits of Interpretation)* (1990).

Another pupil of Pareyson is Gianni Vattimo, a professor in Turin and Italy’s most important representative of hermeneutic philosophy based on Heidegger and Gadamer. In works such as *La fine della modernità (The end of modernity)* (1985), Vattimo regards hermeneutics as the philosophy consequent upon the end of metaphysics and modernity. The acknowledgement of the end of foundational systems is seen to bring a liberating and emancipating effect (*La società trasparente (Transparent society)* (1989)). His hermeneutic way of thinking follows Nietzsche and Heidegger: *Pensiero debole (Weak Thought)* (1983) posits the end of the stable structure of being, noting that any relation to being is within linguistic boundaries which are historically determined.

The philosophies of Emanuele Severino and Massimo Cacciari (both professors at Venice) are developed on original lines. Severino criticizes the nihilism of Western philosophy which, by reducing being to nothing (going from being to becoming), creates a philosophical justification for its project of domination, whose final outcome is technology (*Essenza del nichilismo (The Essence of Nihilism)* (1972)). Against this tendency Severino proposes a return to Parmenides, to an ontology of identity that unmasks the illusion of becoming and makes it possible to act free from domination. Through an original interpretation of Nietzsche, Heidegger and Wittgenstein, Cacciari sees in ‘negative thought’ (technology and conventionalism) the origin of the modern crisis of foundations, whose outcome is disenchantment and whose theoretical premises are found in the religious tradition of the West. His works include *Krisis* (1976), *L’angelo necessario (The necessary angel)* (1986), and *Dell’inizio (On origins)* (1990).

Italian philosophy has always considered philosophical movements and fields such as empiricism, analytical

philosophy, logic and philosophy of science to be of minor importance, its strong humanist and historical tradition aligning more with French and German thought. This has become less true in recent times. The example of Ludovico Geymonat, who held the first chair of Philosophy of Science in Milan, is indicative. Having studied with Reichenbach, Schlick and Carnap, Geymonat developed an epistemological theory sensitive to the historical evolution of science, attempting to bridge the traditional gap between humanist and scientific culture. His studies of Galileo and of philosophical and scientific thought are also directed to this end.

See also: Humanism, Renaissance; Renaissance philosophy

Translated by Anna Gannon
GAETANO CHIURAZZI

References and further reading

Ardigò, R. (1870) *Psicologia come scienza positiva (Psychology as a positivist science)*, Mantova: Guastalla. (Outlines Ardigò's idea of psychology as a science. Includes a chapter about the relationship between positivist psychology and philosophy.)

Ardigò, R. (1879) *La morale dei positivisti (The morality of positivists)*, ed. G. Giannini, Milan: Marzorati, 1973. (The ethical doctrine of Ardigò, in open debate with idealism, with an introduction and bibliography by the editor.)

Bausola, A. ***et al.*** (1985) *La filosofia italiana dal dopoguerra a oggi (Italian philosophy from the end of the Second World War to today)*, Rome and Bari: Laterza. (Essays by Garin, Dal Pra, Pera, Bedeschi, Bausola and Verra on the most significant movements in contemporary Italian philosophy.)

Beccaria, C. (1764) *Dei delitti e delle pene (On crime and punishment)*, ed. F. Venturi, Turin: Einaudi, 1992. (Introduction and bibliographical notes by the editor. Includes an appendix of letters and documents relating to the composition of the work and to the arguments it gave rise to in eighteenth-century Europe.)

Bobbio, N. (1964) *Giusnaturalismo e positivismo giuridico (Natural law and legal positivism)*, Milan: Comunità, 1977.(Collection of essays from 1958 to 1964 on the difference between natural rights and positivist rights.)

Bobbio, N. (1968) *Teoria della norma giuridica (Theory of judicial standards)*, Turin: Giappichelli.(University lectures on the theory of rights from the normative viewpoint.)

Bobbio, N. (1976) *La teoria delle forme di governo (Theory of forms of government)*, Turin: Giappichelli. (University of Turin lectures, 1975-6, on political theories from Plato to Marx.)

Bobbio, N. (1994) *Destra e sinistra (Right and left)*, Rome: Donzelli.(Brief essay on the causes and significance of the political distinction between right and left, with a bibliography by P. Polito following the debate about the left in Italy.)

Bruno, G. (1584-5) *Dialoghi italiani (Italian dialogues)*, ed. G. Aquilecchia, Florence: Sansoni, 1958, repr. 1985. (Critical edition of Bruno's metaphysical and moral dialogues, with a note on the translations into foreign languages in the preface by G. Gentile.)

Botero, G. (1589) *Della ragion di stato (The reason of state)*, ed. L. Lirpo, Turin: U.T.E.T., 1948.(Critical edition with an introduction, biographical note and list of all Botero's works with an annotated list of essential reading.)

Cacciari, M. (1976) *Krisis*, Milan: Feltrinelli.(An examination of negative and critical-radical thought in contemporary culture, with particular reference to the German-speaking world, especially Nietzsche and Wittgenstein.)

Cacciari, M. (1986) *L'angelo necessario (The necessary angel)*, Milan: Adelphi.(The angel in Western culture regarded as a hermeneutic bridging the visible and the invisible.)

Cacciari, M. (1990) *Dell'inizio (On origins)*, Milan: Adelphi.(Cacciari's most important and difficult theoretical work which re-examines the philosophical problem of origins through a combination of philosophy and theology.)

Campanella, T. (1620) *Del senso delle cose e della magia (On material things and magic)*, ed. A. Bruers, Bari: Laterza, 1925.(Unedited Italian text with a critical introduction.)

Campanella, T. (1622) *Apologia di Galileo (Apology for Galileo)*, ed. L. Firpo, Turin: U.T.E.T., 1968.(Useful for understanding the relation between Campanella and Galileo. Text, with a copy of the 1632 Latin edition and a brief historical-bibliographical note.)

Campanella, T. (1623) La *Città del sole (The city of the sun)*, ed. N. Bobbio, Turin: Einaudi, 1941.(Latin text with historical-critical note on the different editions of this work in which Campanella sets out his political doctrine. Extensive introduction and essential reading list.)

Cattaneo, C. (1859-66) *Psicologia delle menti associate (Psychology of associated minds)*, in *Scritti filosofici*

(Philosophical writings), vol. 1, 407-479, ed. N. Bobbio, Florence: Le Monnier, 1960.(Introduction and critical notes by the editor. There is a psycho-sociological interpretation of Kantianism which re-evaluates the latter's idealistic-transcendental aspect.)

Croce, B. (1902) *Estetica come scienza dell'espressione e Linguistica generale (Aesthetics as the science of expression and general linguistics)*, ed. G. Galasso, Milan: Adelphi, 1990.(Systematic exposition of aesthetics, divided into theoretical and historical parts. Extensive bibliography and critical note by the editor.)

Croce, B. (1905) *Logica come scienza del concetto puro (Logic as the science of pure concept)*, Bari: Laterza, 1971.(Fundamental text on historical idealism in which Croce sets out his idea of philosophy as Logic and its relationships with experience and history. Not for beginners.)

Croce, B. (1909) *Filosofia della practica. Economia e etica (Philosophy of practice, economics and ethics)*, Bari: Laterza, 1945.(Sets out Croce's systematic theory of the forms of spiritual, economic and ethical activity.)

Cuoco, V. (1805) *Platone in Italia (Plato in Italy)*, ed. F. Nicolini, Bari: Laterza, 1969.(Translated from the Greek, this describes a hypothetical journey by Plato to Italy.)

Eco, U. (1962) *Opera aperta (Open work)*, Milan: Bompiani.(Essays, 1958-60, examining form and indeterminism in contemporary poetics.)

Eco, U. (1990) *I limiti dell'interpretazione (The limits of interpretation)*, Milan: Bompiani.(Collection of essays about his own hermeneutic anarchism, and in particular of American deconstructionism. The interpretation is free, but follows a principle of falsification which is its limit.)

Filangieri, G. (1784) *Scienza della legislazione (The science of legislation)*, in *Scritti (Writings)*, ed. F. Venturi, Turin: Einaudi, 1976, 3-109.(Sets out Filangieri's political and economic theory. Introduction and good bibliography.)

Galiani, F. (1780) *Della moneta (On money)*, ed. F. Nicolini, Bari: Laterza, 1915.(Work of economic theory on the value of money.)

Galilei, G. (1964) *Opere (Works)*, ed. F. Brunetti, Turin: U.T.E.T., 2 vols.(Critical edition of Galileo's scientific works and scientific-philosophical polemics, including *Saggiatore* (vol. 1), *Dialogo sui massimi sistemi* and *Discorsi e dimostrazioni matematiche intorno a due nuove scienze* (both vol. 2). Each work has introductory notes and biographical and bibliographical references. Useful on Galileo's scientific discoveries but rather specialized.)

Galluppi, P. (1819) *Saggio filosofico sulla critica della conoscenza (Philosophical essay on the critique of knowledge)*, ed. N. Abbagnano, Turin: Paravia, 1943. (Attempts to reduce the sceptical aspects of previous theories, from Berkeley and Hume to Condillac and Kant. Introduction and critical notes by the editor.)

Garin, E. (1970) *Dal Rinascimento all'Illuminismo. Studi e Ricerche. (From the Renaissance to the Enlightenment. Studies and research)*, Pisa: Nistri-Lischi.(Essays giving an overview of Italian philosophy in the 1500s to 1700s. Attention is drawn particularly to the essays on the period from Campanello to Vico and to that on the reception of Hobbes in the 1700s.)

Genovesi, A. (1977) *Scritti (Writings)*, ed. F. Venturini, Turin: Einaudi.(Economic writings and letters on various subjects, with an introduction and good bibliography.)

Gentile (1916) *Teoria generale dello spirito come atto puro (General theory of the spirit as pure act)*, in *Scritti*, vol. 3, Florence: Sansoni, 1959.(Fundamental text of Gentile's actualism in which the influence of Berkeley and Fichte is evident.)

Gentile (1917-22) *Sistema di logica (System of logic)*, in Scritti, vol. 5, Florence: Sansoni, 1964.(Systematic exposition of Gentile's logic as the logic of the Identical; difficult text.)

Gentili, A. (1877) *Del diritto di guerra (On the rules of war)*, ed. A. Fiorini, Livorno: Vigo.(Italian translation of *De jure belli*, holding that the only just war is carried out in self-defence.)

Geymonat, L. (1970-2) *Storia della pensiero filosofico e scientifico (History of philosophical and scientific thought)*, Milan: Garzanti, 6 vols.(University-level manual with particular reference to the history of scientific thought, structured by topic, with specific chapters on Italian philosophy and a bibliography up to the 1970s.)

Gioberti, V. (1841) *Degli errori filosofici di A. Rosmini (On Rosmini's philosophical errors)*, in *Opere (Works)*, vols 9 and 10, ed. U. Redano, Milan: Bocca, 1939.(Documents the dispute between Rosmini and Gioberti; with introduction and bibliographical notes.)

Gioberti, V. (1842-3) *Del primato morale e civile degli italiani (On the moral and civil superiority of the Italians)*, in *Opere (Works)*, vols 1, 2 and 3, ed. E. Castelli, Milan: Bocca, 1938.(Gioberti sets out his political ideas recalling the Italians to their historic mission; preface and essential reading list.)

Gioberti, V. (1857) *Protologia (Protology)*, 4 vols, ed. G. Bonafede, Padua: Cedam, 1983-4.(Essential for understanding Gioberti's idealism. Extensive introduction by the editor.)

Gioia, M. (1826) *Filosofia della statistica (Philosophy of statistics)*, Turin: Tip. Economica, 1852.(The first

Piedmontese edition of Gioia’s work. Contains brief historical notes on his life and works.)

Gramsci, A. (1948-51) *Quaderni dal carcere (Notebooks from prison)*, ed. V. Gerratana, Turin: Einaudi, 1975, 4 vols.(Critical edition from the Gramsci Institute with introduction and biography by the editor. The fourth volume provides an index.)

Muratori, L.A. (1749) *Della pubblica felicità (On public happiness)*, ed. B. Brunelli, Bologna: Zanichelli, 1941. (Muratori’s ethical-political theory, with an introduction and brief bibliographical note by the editor.)

Pagano, F.M. (1800) *Saggi politici (Political essays)*, ed. F. Collotti, Bologna: Cappelli, 1936.(Political essays on the origins and differences between different societies with an introduction and brief bibliographical note.)

Pareyson, L. (1954) *Estetica. Teoria della formatività (Aesthetics: a theory of formativeness)*, Turin: Edizioni di Filosofia.(Sets out Pareyson’s aesthetic theory in which art is not form but formativity, a doing which gives form. The discussion develops in the context of different modern theories of aesthetics, in particular from Kant onwards.)

Pareyson, L. (1970) *Verità e interpretazione (Truth and interpretation)*, Milan: Mursia, 1971.(Collection of essays from 1964-70 on the hermeneutic concept of truth and the relationship between religion, science and philosophy.)

Rosmini, A. (1830) *Nuovo saggio sull’origine delle idee (New essay on the origin of ideas)*, ed. F. Orestano, Rome: An. Romana editor, 3 vols, 1934.(Systematic account of Rosmini’s epistemology and ontology, with a preliminary critique of previous philosophers from Plato to Kant.)

Rosmini, A. (1854) *Logica (Logic)*, ed. V. Sala, Centro Internazionale di Studi Rosmini, Strese: Città Nuova Editrice, 1984.(Critical edition of Rosmini’s work centred on the relative and innate nature of the idea of being, and on the objectivity of knowledge as opposed to Kantian and idealistic subjectivity.)

Severino, E. (1972) *Essenza del nichilismo (The essence of nihilism)*, Brescia: Paideia.(The idea of being as becoming. Severino proposes a ‘return to Parmenides’, rethinking his fundamental idea. Not an easy text.)

Spaventa, B. (1862) *La filosofia italiana nelle sue relazioni con la filosofia europea (Italian philosophy and its relation to European philosophy)*, ed. G. Gentile, Florence: Sansoni, 1943.(Account of Italian philosophy in an idealistic key from Bruno to Gioberti, tracing its roots in European philosophy up to Hegel.)

Vattimo, G. (1985) *La fine della modernità (The end of modernity)*, Milan: Garzanti.(Essays from 1981-4 on nihilism and hermeneutics in postmodern culture, with particular reference to Nietzche and Heidegger.)

Vattimo, G. (1989) *La società trasparente (Transparent society)*, Milan: Garzanti.(Analysis of contemporary postmodern society as the society of communication in which the mass-media, with its pervasive, complex and chaotic qualities, provides a possibility of emancipation.)

Vattimo, G. and Rovatti, P.A. (eds) (1983) *Il pensiero debole (Weak thought)*, Milan: Feltrinelli.(Collection of essays by various authors including Vattimo, Rovatti, Eco, Marconi and Carchia, on the ‘crisis of reason’ in contemporary philosophy which sees in the end of ‘strong reason’ the emergence of a more flexible and tolerant truth.)

Verri, P. (1763) *Discorsi… sull’indole del piacere e del dolore, sulla felicità… (Discourses… on the balance of pleasure and pain, on happiness…)*, ed. A. Plebe, Milan: Marzorati, 1972.(Brief text illustrating Verri’s concept of ethics. Introduction, critical interpretation and bibliographical note by the editor.)

Viano, C.A. and Rossi, P. (1991) *Filosofia italiana e filosofie straniere nel dopoguerra (Italian philosophy and foreign philosophies in the period following the Second World War)*, Bologna: il Mulino.(Collection of essays by various authors, on topics such as phenomenology, existentialism, hermeneutics, pragmatism, positivism, Marxism and structuralism. Provides a clear and coherent view of contemporary Italian philosophy.)

Vico, G. (1710) *De antiquissima Italorum sapientia (On the ancient knowledge of the Italians)*, in *Opere filosofiche (Philosophical works)*, ed. P. Cristofolini, Florence: Sansoni, 1971, 55-131.(Advances Vico’s principal doctrine in which truth and fact are mutually interchangeable, thus founding the new historical science.)

Vico, G. (1744) *Scienza nova (New science)*, in *Opere filosofiche (Philosophical works)*, ed. P. Cristofolini, Florence: Sansoni, 1971, 377-702.(Advances Vico’s philosophy of history.)

Itō Jinsai (1627-1705)

Itō Jinsai, along with his contemporary Yamaga Sokō, pioneered the kogaku, *or 'Ancient Learning', philosophical movement of Tokugawa Japan.* Kogaku *reacted against the allegedly stifling and excessively metaphysical ideas of Zhu Xi's neo-Confucianism. In making his call for a return to the ancient Confucian teachings, Jinsai produced one of the first and most systematic visions of Confucian philosophy.*

A lifelong resident of Kyoto, the imperial capital, Jinsai made his call for a return to the ancient Confucian teachings of the *Analects* and the *Mengzi* (see Confucius; Mencius) in isolation from the *samurai* regime based in Edo (now Tokyo). He published little in his day, being content with a life of dignified poverty teaching numerous students, including townspeople and court nobles, at his *Kogidō*, or 'School of Ancient Semantics'. Nevertheless, Jinsai's ideas circulated nationally through pirated editions of his writings. One such unauthorized publication sparked the interest of Ogyū Sorai, then living in Edo, in *kogaku*. Ogyū Sorai wrote to Jinsai, praising his ideas and requesting a teacher-disciple relationship, but Jinsai, then within a year of his own death, never responded. About a decade later, Ogyū Sorai formulated his own *kogaku* system, which systematically attacked Jinsai's ideas even as it advanced upon the trail Jinsai had blazed. Despite Ogyū Sorai's critiques, Jinsai's ideas, as popularized and published by his son and successor Itō Tōgai, prevailed in the eighteenth century as the most humane variety of classical Confucianism.

Jinsai's philological methodology derived largely from the *Xingli ziyi (The Meanings of Neo-Confucian Terms)* of Chen Beixi (1159-1223), a philosophical lexicon systematically defining some twenty-five key concepts as understood by Zhu Xi, the master synthesizer of neo-Confucian philosophy in East Asia. The *Xingli ziyi*, first promoted in Japan by Hayashi Razan (1583-1657), appeared in several seventeenth-century editions as Tokugawa scholars sought to fathom the new metaphysical philosophy of neo-Confucianism. No sooner had Chen Beixi's lexicon conveyed the essence of Zhu Xi's ideas, than its philologico-lexicographical methodology gave rise to systematic critiques of Zhu Xi's philosophical semantics by Sokō, Jinsai and then Sorai.

Jinsai's *Gomōjigi (The Meanings of Terms in the Analects and the Mencius)* was patterned after the *Xingli ziyi*. The word *jigi* in Jinsai's title is the Japanese for *ziyi* (the meanings of terms), part of Chen Beixi's title. However, the *Gomōjigi* repeatedly criticizes Chen Beixi, taking issue with him even as it appropriates his methods. Themes developed in the *Gomōjigi* owe much to the ideas of Luo Qinshun (1465-1547), a Ming dynasty scholar who espoused loyalty to Zhu Xi's neo-Confucianism even while criticizing it. The critiques of Luo Qinshun and Jinsai in turn reflected Zhu Xi's open-minded call for critical scrutiny in philosophy. The *Jinsilu (Reflections on Things at Hand)*, edited by Zhu Xi, quips, 'By doubting the indubitable, one greatly advances'.

Jinsai's ideas ultimately transcended their Chinese sources, producing one of the first and most systematic, philologically-based visions of Confucian philosophy ever articulated in East Asian intellectual history. Japanese scholars often note that Jinsai's ideas predated, by nearly a century, those of Dai Zhen, the famed Qing dynasty scholar of 'Evidential Learning' whose critiques of Zhu Xi took much the same form as those of Jinsai.

Breaking with the neo-Confucian dualism of principle and material force, Jinsai proclaimed that all things consist of a unitary, primal material force, and that alone. His metaphysics assumed vivacious activity, unlike Zhu Xi who claimed that the original nature of humanity could only be glimpsed in quietistic meditation (see Xing). Jinsai saw the human mind as an energetic, active endowment, not as a quiescent, motionless entity (see Mind, philosophy of; Xin). He valued human feelings, contending that they are the innate roots of moral behaviour. He praised *jin*, or empathetic moral concern, along with other ancient Confucian virtues as integral elements of *michi*, or the true Way of human conduct. Perfection of the latter made one a gentleman, the ideal proffered by Jinsai in place of Zhu Xi's notion of sagehood. Religiously, Jinsai did affirm the neo-Confucian line in explaining ghosts and spirits as the activities of the psycho-material forces of *yin* and *yang* (see Yin-yang). Though generally a tolerant and broadminded thinker, Jinsai argued in a famous essay that the *Daxue (Great Learning)* was not an authentic Confucian text because it devalued the emotions in ways that Confucius never had (see Daxue).

See also: Confucian philosophy, Japanese; Dai Zhen; Japanese philosophy; Neo-Confucian philosophy; Ogyū Sorai; Zhu Xi

JOHN ALLEN TUCKER

List of works

Itō Jinsai (1683-1705) *Gomōjigi (The Meanings of Terms in the Analects and Mencius)*, in Yoshikawa Kōjirō and Shimizu Shigeru (eds) *Itō Jinsai/Itō Tōgai*, Nihon shisō taikei vol. 33, Tokyo: Iwanami shoten, 1971.(This source includes annotated Japanese texts of not only the *Gomōjigi* but also of Jinsai's other major work, *Kogaku sensei bunshū (Collected Prose Works of the Master of Ancient Learning)* along with explanatory essays. Itō Tōgai's *Kokin gakuhen (The Shift in Learning from Ancient to Modern Times)* is also included.)

References and further reading

de Bary, W.T. and Bloom, I. (eds) (1979) *Principle and Practicality: Essays in Neo-Confucianism and Practical Learning*, New York: Columbia University Press. (Contains several well-researched essays pertaining to Jinsai.)

Ishida Ichirō (1960) *Itō Jinsai*, Tokyo: Yoshikawa Kōbunkan.(A somewhat dated but useful biography of Jinsai in Japanese.)

Katō Nihei (1940) *Itō Jinsai no gakumon to kyōiku (The Scholarship and Education of Itō Jinsai)*, Tokyo: Meguro shoten.(The only book-length study of Jinsai's pedagogical practices.)

Kimura Eiichi (ed.) (1970) *Itō Jinsai shū (The Works of Itō Jinsai)*, Nihon no shisō vol. 11, Tokyo: Chikuma shoten.(Includes major texts by Jinsai and an analytic essay by Kimura.)

Koyasu Nobukuni (1982) *Itō Jinsai-jinrinteki sekai no shisō*, Tokyo: Tokyo daigaku shuppankai.(A biography in the style of intellectual history.)

Koyasu Nobukuni (1986) *Itō Jinsai kenkyū (A Study of Itō Jinsai)*, in *Osaka daigaku daigaku kiyō*, vol. 26.(A concept-by-concept study of Jinsai's thought.)

Miyake Masahiko (1987) *Kyōto chōshū Itō Jinsai no shisō keisei*, Tokyo: Shibunkaku shuppan.(Explores Jinsai's relations as a thinker to Kyoto and its populace. Also pioneers textual studies distinguishing Jinsai's thought from that of Tōgai.)

Sakai Naoki (1992) *Voices of the Past: The Status of Language in Eighteenth-Century Japanese Discourse*, Ithaca, NY: Cornell University Press.(Argues that from the late-seventeenth century, Japanese discourse began conceptualizing the world in diverse ways, all of which had political consequences.)

Spae, J.J. (1948) *Itō Jinsai*, New York: Paragon; repr. 1967.(The only biography of Jinsai in English.)

Yamashita, S. (1983) 'The Early Life and Thought of Itō Jinsai', *Harvard Journal of Asiatic Studies* 43: 453-80. (Surveys Jinsai's transition from neo-Confucianism to Ancient Learning philosophy.)

Yoshikawa Kōjirō (1983) *Jinsai, Sorai, Norinaga: Three Classical Philologists of Mid-Tokugawa Japan*, Tokyo: Tōhō Gakkai.(Presents an English translation of Yoshikawa's seminal 'Jinsai gakuan'.)

Jacobi, Friedrich Heinrich (1743-1819)

Polemicist and literary figure, Jacobi was an outspoken and effective defender of individualism. He accused philosophers of conceptualizing existence according to the requirements of explanation, thus allowing no room for individual freedom or for a personal God. In a series of polemics that influenced the reception of Kant, Jacobi applied his formula, 'Consistent philosophy is Spinozist, hence pantheist, fatalist and atheist', first to Enlightenment philosophy and then to idealism. Jacobi was not however opposed to reason; in 'faith' and 'feeling' he sought to recover the intuitive power of reason philosophers ignored.

Jacobi also criticized the literary movement spearheaded by the young Goethe, because of its latent fatalism. He dramatized in two novels the problem of reconciling individualism with social obligations. An exponent of British economic and political liberalism, Jacobi was an early critic of the French revolution which he considered the product of rationalism.

1 Life and work

Born in Düsseldorf, of well-to-do merchant family, Jacobi studied in Geneva where he became acquainted with Rousseau and Bonnet. Destined to a business career, after marrying Elisabeth (Betty) von Clermont he entrusted his affairs to his brother-in-law and dedicated himself to literary and social activities. Together with Betty he made his villa at Pempelfort a meeting place for the personalities of the day and a centre of liberal ideas. He was instrumental to the founding of the journal *der Teutsche Merkur*. In the 1770s he held political positions in the duchies of Julich-Berg and, briefly in 1779, in Bavaria. In 1805 he moved to Munich to take up the post of president of the Academy. He spent his last years there, supervising the edition of his works.

Widely regarded as an anti-Enlightenment figure in his own lifetime and afterwards, Jacobi in fact sought to defend individualistic values which were just as much part of the Enlightenment as the rationalism he fought all his life. He claimed that philosophy artificially abstracts from the individuality of existence yet takes its empty conceptual artifacts to define reality itself. For the sake of explanation philosophy subordinates conditions of existence to conditions of thought, thereby subverting the possibility of choice and action. Hence Jacobi accused philosophy of 'nihilism' (a term he popularized).

2 Critique of philosophy

In a masterstroke of philosophical propaganda, Jacobi propelled his position (and himself) to the centre of attention in 1785, by publishing with commentary letters he had recently exchanged with Moses Mendelssohn (*Concerning the Doctrine of Spinoza in Letters to Herr Moses Mendelssohn*; second, much enlarged edition, 1789) (see Mendelssohn, M.). The purpose of the correspondence had been to clarify Lessing's declaration, allegedly made to Jacobi in a conversation shortly before Lessing's death, that he was a Spinozist (see Lessing, G.E.; Spinoza, B. de). In that conversation Jacobi had claimed that philosophy inexorably leads to Spinoza's 'substance', and this abstract concept, when set up as first principle, undermines individual distinctions, most of all the distinction between God and creatures. Lessing had responded by declaring his sympathy for Spinoza, and Jacobi had thereupon urged him to perform a *salto mortale* - a spiritual somersault in virtue of which, through an act of faith, Lessing would simply declare himself *for* a personal God, and *for* freedom, and thereby rejoin common sense.

The book caused an uproar. Jacobi had succeeded in dramatizing his cause against philosophy by identifying the latter with Spinozism (widely regarded at the time as synonymous with atheism), and, by associating Spinoza with Lessing, he had cast doubt on Lessing's typical Enlightenment belief that reason can save the essential truths of religion and morality. Jacobi had at the same time challenged the authority of Mendelssohn, universally acknowledged as Lessing's intimate friend and a witness to reason's ability to establish truths transcending ethnic and religious differences. The equation Jacobi then drew, 'Consistent philosophyČequals;SpinozismČequals;pantheismČequals;fatalismČequals;atheism', is one he defended to the end, although he soon had to adapt it to the idealism of Kant and his followers.

3 Critique of idealism

Jacobi's objections to Kant's critique of reason became commonplace in the sceptical reaction to Kant (see the

Appendix to *David Hume on Faith, or Idealism and Realism: A Dialogue*, 1787). Jacobi argued that it is inconsistent to accept the 'thing in itself' yet disclaim knowledge of it; that in a priori 'space-time' the individuality of determination is a mere illusion; that the categories are empty forms artificially imposed upon experience, and the transcendental 'I' a counterfeit subject. His later negative reactions to Fichte and Schelling were even stronger. Jacobi accused Fichte of inverted Spinozism - of parading as 'subject' and 'subjectivity' an indeterminate source of activity (the self-positing 'I') in fact just as anonymous and impervious to individuation as Spinoza's substance. He made this accusation in *Jacobi to Fichte* (1799), at a time when Fichte was under suspicion of atheism. The attack on Schelling - fellow member of the Munich Academy whose views on art and nature Jacobi feared would corrupt the young - came in 1811 (*Of Divine Things and Their Revelation*). The ensuing bitter public debate is known as the 'pantheism dispute'.

4 Positive doctrine

Despite his polemics, Jacobi insisted that he was not against reason but that, on the contrary, his aim was to unmask the irrationalism latent in rationalistic philosophy. In *David Hume* - philosophically his most interesting work - he claimed that he used 'faith' in the sense of Hume's 'belief' (both *Glaube* in German). He however rejected Hume's phenomenalism and in the same work defended a realism of the senses not unlike that of Thomas Reid (by whom he was influenced, as also by Adam Ferguson). He concluded with the suggestion that reason is a higher and more reflective form of sensibility - a position which Jacobi associated with Leibniz and which had obvious vitalistic implications. Yet Jacobi later rejected the vitalism of Herder because of its pantheism (Appendices IV-V to *Doctrine of Spinoza*, 1789).

Jacobi's ambiguity towards philosophy - to which he was attracted but which he feared because of its possible nihilistic effects - was reflected in his attitude to the philosophers he criticized. He admired Spinoza for his consistency and his recognition that truth is selfrevelatory (hence transcends ratiocination). He felt kinship with Kant, with whom he shared belief in the priority of existence over thought and in the finitude of reason. Retrospectively he used Kant's distinction between reason and understanding to define what he had always meant by 'faith' or the 'feeling for truth'. He identified the 'reason' he had criticized as Kant's 'understanding'- a faculty by nature bent on cataloguing and exploiting finite, sensible things. But above understanding there is true reason with its intuitive power to apprehend higher moral truths. By 'faith', Jacobi claimed, he had always meant this true reason. Kant had rightly distinguished it from understanding but had failed to recognize its intuitive power, thereby subordinating it to understanding ('On the Undertaking of Critique to Reduce Reason to the Understanding', 1802). Jacobi also hinted that in his *David Hume* he had offered an alternative to Kant's deduction of the categories, and that Kant had based his refutation of idealism on his own early claim that no 'I' is possible without a 'Thou'. Despite ambiguities and fluctuations, Jacobi always defended his two fundamental theses: (1) Existence is radically individualized and precedes thought possibilities; (2) Truth is selfrevelatory and is best intuited in the relations between individuals and between individuals and their God.

5 Literary work

Inspired by Goethe, with whom he had a long but uneven relationship, Jacobi wrote two novels which he published, beginning in 1775, under various forms: *Allwill*, (final form 1792) and *Woldemar* (final form 1796). Goethe had also figured in Jacobi's report of his controversial conversation with Lessing and, implicitly, was just as much a target of attack in the ensuing controversy as Mendelssohn. Jacobi feared Goethe's enthusiasm for Spinoza, and the literary glorification of nature he promoted. In *Allwill* he portrayed the seductive and ultimately destructive effects of a lively poetic imagination undisciplined by social and moral constrains. In *Woldemar* he explored the problem of balancing nature with art and virtue, and of establishing social bonds between individuals richly endowed with subjectivity. The man-woman relationship theme is central to both novels.

6 Political ideas

In early essays Jacobi defended free trade and attacked the theorists of state absolute power. Passionately interested in anything political, he criticized the French revolution which he regarded as an unrealistic attempt to reshape human institutions to suit abstract reason, without regard for history or nature.

7 Influence

Jacobi's personalism and his religiosity (which drew inspiration from many sources, notably Hamann) had wide influence. In his early *Kantian Letters* Reinhold cast Kant in the role of mediator between Mendelssohn's reason and Jacobi's faith, with far-reaching consequences for the reception of Kant. Jacobi's influence also spread through personal contacts and a voluminous correspondence. Although not himself a product of pietism, late in his life Jacobi inspired pietist educated writers such as Schleiermacher and Fries, and through their writings his influence passed on to the nineteenth century (see Pietism).

See also: Enlightenment, Continental

GEORGE DI GIOVANNI

List of works

Jacobi, F.H. (1785) *Über die Lehre von Spinoza an den Herrn Moses Mendelssohn (Concerning the Doctrine of Spinoza in Letters to Herr Moses Mendelssohn)*, Breslau: Löwe; 2nd edn, 1789; repr. in H. Scholz (ed.) *Die Hauptschriften zum Pantheismusstreit zwischen Jacobi und Mendelssohn*, Berlin: Reuther & Reichard, 1916. (1789 edition is much enlarged. 1916 edition also contains other writings relating to the Jacobi-Mendelssohn controversy.)

Jacobi, F.H. (1787) *David Hume über den Glauben, oder Idealismus und Realismus. Ein Gespräch (David Hume on Faith, or Idealism and Realism: A Dialogue)*, Breslau: Löwe; 2nd edn, 1815; 1st edn repr. New York and London: Garland, 1983.(1815 edition was part of the *Werke* and was significantly edited, with a new lengthy introduction. 1983 edition includes this 1815 new Preface, with an English introduction by Hamilton Beck.)

Jacobi, F.H. (1792) *Eduard Allwills Briefsammlung, herausgegeben von Friedrich Heinrich Jacobi, mit einer Zugabe von eigenen Briefen (Edward Allwill's Collection of Letters, edited by Friedrich Heinrich Jacobi, with an Addition from Letters of His Own)*, Königsberg: Nicolovius; repr. J.U. Terpstra (ed.), Groningen: Djakarta, 1957.(This is the final text of a novel, parts of which had previously been published starting from 1775. 1957 edition is a critical edition.)

Jacobi, F.H. (1796) *Woldemar*, Königsberg: Nicolovius.(This is a somewhat edited version of the 1794 final text of a novel, parts of which had previously been published starting from 1779.)

Jacobi, F.H. (1799) *Jacobi an Fichte (Jacobi to Fichte)*, Hamburg: Perthes; repr. in R. Lauth and H. Gliwitzky (eds) (1972) *Fichte-Gesamtausgabe*, Stuttgart and Bad Cannstat: Frommann, series 3, vol. 3.(An open letter to Fichte publicly rejecting his idealism. 1972 edition is a critical edition.)

Jacobi, F.H. (1802) 'Über das Unternehmen des Kriticismus, die Vernunft zu Verstande zu bringen, und der Philosophie überhaupt eine neue Absicht zu geben' (On the Undertaking of Critique to Reduce Reason to the Understanding, and to Give a New Purpose to Philosophy in General), in *Beyträge zur leichtern übersicht des Zustandes der Philosophie beym Anfange des 19. Jahrhunderts* 3: 1-110.(The Beyträge were a series of volumes edited by K.L. Reinhold.)

Jacobi, F.H. (1811) *Von den Göttlichen Dingen und ihrer Offenbarung (Of Divine Things and Their Revelation)*, Leipzig: Fleischer; repr. in W. Weischedel (ed.) (1967) *Streit um die göttlichen Dinge. Die Auseinandersetzung zwischen Jacobi und Schelling*, Darmstadt: Wissenschaftliche Buchgesellschaft.(Jacobi's attack on Schelling. 1967 edition includes other writings relating to the Jacobi-Schelling controversy.)

Jacobi, F.H. (1812-25) *Friedrich Heinrich Jacobi's Werke*, eds J.F. Köppen and C.J.F. Roth, Leipzig: Fleischer, 6 vols; easily available in photo-mechanical reproduction, Darmstadt: Wissenschaftliche Buchgesellschaft, 1968. (1812-25 edition was supervised by Jacobi before his death. It is not reliable as a document of Jacobi's earlier production.)

Brüggen, M. and Sudhof, S. (eds) (1981-) *Briefwechsel*, Stuttgart and Bad Constatt: Frommann-Holzboog.(The first volumes of a projected critical edition so far covering the early correspondence. Published to date: part 1, vols 1-3, part 2, vol. 1.)

di Giovanni, G. (ed. and trans.) (1994) *The Main Philosophical Writings and the Novel 'Allwill'*, Kingston and Montreal: McGill-Queen.(It includes complete translations of *Concerning the Doctrine of Spinoza* (1785), *David Hume* (1787), *Allwill, Jacobi to Fichte*, the introduction to the *David Hume* of 1815, and excerpts from other works. It also includes a long study on Jacobi and his age, and a detailed bibliography.)

References and further readings

Baum, G. (1969) *Vernunft und Erkenntnis. Die Philosophie F.H. Jacobis*, Bonn: Bouvier.(A study of Jacobi's epistemology.)

Beiser, F.C. (1987) *The Fate of Reason*, Cambridge, MA: Harvard University Press.(Includes a very readable interpretation of Jacobi as proto-existentialist.)

di Giovanni, G. (1989) ‘From Jacobi’s Philosophical Novel to Fichte’s Idealism: Some Comments on the 1798-99 "Atheism Dispute"’, *Journal of the History of Philosophy* 27 (1): 75-100.(Readable and informative.)

Ford, L.S. (1965) ‘The Controversy Between Schelling and Jacobi’, *Journal of the History of Philosophy* 3 (1): 75-89.(Readable and informative.)

Hammacher, K. (ed.) (1971) *Friedrich Heinrich Jacobi. Philosoph und Literat der Goethezeit*, Frankfurt: Klostermann.(A collection of important but highly specialized studies.)

Hammacher, K. (1969) *Kritik und Leben II. Die Philosophie Friedrich Heinrich Jacobis*, Munich: Fink.(A scholarly treatment of all aspects of Jacobi’s philosophy.)

Hammacher, K. and Hirsch, H. (1993) *Die Wirtschaftspolitik des Philosophen Friedrich Heinrich Jacobi*, Amsterdam and Atlanta, GA: Editions Rodopi.(A study of Jacobi’s economic policy.)

Lévy-Bruhl, L. (1894) *La philosophie de Jacobi*, Paris: Alcan.(A dated but still very readable study.)

Nicolai, H. (1965) *Goethe und Jacobi, Studien zur Geschichte ihrer Freundschaft*, Stuttgart: Metzler.(A fascinating study of Jacobi’s relationship to Goethe; it throws light on the whole period.)

Verra, V. (1963) *F.H. Jacobi. Dall’Illuminismo all’Idealismo*, Turin: Edizioni di ‘Filosofia’.(Still the best introduction to Jacobi and his age.)

Jaina philosophy

The issues in Jaina philosophy developed concurrently with those that emerged in Buddhist and Hindu philosophy. The period from the second century BC *to about the tenth century* AD *evinces a tremendous interaction between the schools of thought and even an exchange of ideas, borne out especially in the rich commentary literature on the basic philosophical works of the respective systems. Jaina philosophy shares with Buddhism and Hinduism the aim of striving, within its own metaphysical presuppositions, for absolute liberation (*mokṣa *or* nirvāṇa*) from the factors which bind human existence. For the philosophical systems of Indian thought, ignorance (of one's own nature, of the nature of the world and of one's role in the world) is one of the chief such factors, and Jainism offers its own insights into what constitutes the knowledge that has the soteriological function of overcoming ignorance. Jainism is not exempt from the problem of distinguishing the religious and/or mystical from the 'philosophical'; the Indian tradition has no exact equivalents for these categories as they are usually employed in Western thought.*

The significance ascribed to knowledge is reflected in the attention given to epistemology and logic by Jaina philosophers. The first systematic account was given by the fourth- or fifth-century philosopher Umāsvāti, who distinguished two types of knowledge: partial knowledge, which is obtained from particular standpoints, and comprehensive knowledge, which is of five kinds - sensory knowledge, scriptural knowledge, clairvoyance, telepathy and omniscience. Of these, the first two are held to be indirect (consisting in, or analogous to, inference) and the remainder are direct; Jainism is unique among Indian philosophies in characterizing sensory knowledge as indirect. The aim of the treatises on knowledge is to present what the Jainas believe would be known in the state of omniscience, as taught by Mahāvīra. Omniscience is an intrinsic condition of all souls; however, due to the influence of karma since beginningless time this essential quality of the soul is inhibited.

The Jaina interest in logic arose, as with the other schools, through a consideration of inference as a mode of knowledge. The methods and terminology of the Nyāya school were heavily drawn upon; this is evident in Siddhasena Divākara's Nyāyāvatāra (The Descent of Logic) (c. *fifth century), one of the first detailed presentations of Jaina logic. The Jainas used logic to criticize other schools and defend their own. The acquaintance with other traditions that this implies is a notable aspect of classical Jainism; their interest in other schools, coupled with their belief in collecting and preserving manuscripts, makes the Jaina corpus very important for the study of classical Indian thought.*

According to Jaina ontology, reality is divided into the two basic principles of sentience and non-sentience, neither of which is reducible to the other. The former is manifested in souls, of which there are an infinite number, and the latter in the five basic substances, which are matter, dharma *and* adharma *(factors posited to explain movement and rest), space and time. Matter consists of atoms; as it becomes associated with the soul, it gets attached to it, becomes transformed into karma and thereby restricts the functions of consciousness. This pernicious process can only be reversed through ascetic practices, which ultimately lead to liberation.*

Ascetic practices constitute the basis of Jaina ethics, the framework of which are the 'five great vows', according to which the ascetics vow to live. These are: nonviolence towards all forms of life, abstinence from lying, not taking what is not given, celibacy and renunciation of property. Nonviolence is strongly emphasized, since violence produces the greatest amount of karma. Hence great care has to be exercised at all times, especially because injury of life forms should be avoided also in plants, water, fire, etc. The minimization of physical activity to avoid injury is therefore an important ideal of Jaina asceticism. Inspiration for constant ethical behaviour is provided by a contemplation of the lives of the twenty-four Jinas, of whom Mahāvīra was the last. Though human, these 'conquerors of the passions' are worshipped as divine beings because of their conduct in the world and knowledge of the nature of ultimate reality.

1 Ontology and metaphysics

Jaina thought has retained several elements of ancient natural philosophy, such as the theory of atoms, explaining the nature of the universe without recourse to a creator god, and the interpretation of karma as particles of subtle matter (see Karma and rebirth, Indian conceptions of §6). The Jainas hold an ancient theory in which ultimate reality is divided into two basic categories, one which represents the principle of sentience or consciousness as

such, usually translated as 'soul' (*jīva*), and one which represents the non-sentient principle (*ajīva*). In their intrinsic natures they are exclusive categories in the sense that, despite their coexistence (as in human beings), the one cannot take on the nature of the other. Both are ontological substances (*dravya*) in that reality or existence as such can ultimately be reduced to these two eternal, uncreated and indestructible categories (see Ontology in Indian philosophy §1). The soul is best described as a monad, of which there are a countless number, each independent but with the same principle of sentience as its chief characteristic, and each able to expand or contract according to the size of the body which it inhabits. Souls intrinsically possess unlimited qualities of bliss and energy, and are omniscient.

The non-sentient principle is a generic term for the following five substances. First, matter (*pudgala*) represents the basic stuff or raw material out of which the universe is composed. Matter is made up of atoms, each of which occupies one point of space, and which can be with or without form and can be inhabited by souls. Objects come into being through a combination or aggregation of atoms and change is explained not through destruction or creation of atoms but as a rearrangement of the basic stuff. Together with the view that atoms can have the qualities of colour, taste, touch and smell, Jaina philosophy evolved the theory of substance, quality and mode (see §4) to explain the world in terms of identity and difference, depending on the standpoint and purpose. Matter is crucial for the philosophical anthropology of Jainism because it is matter that directly affects the nature of the soul in the sense that, by becoming converted into karma, it restricts the intrinsic functions of the soul, inhibiting right faith, right knowledge and right conduct. Precisely how this operates is explained later in this section with reference to the so-called 'seven basic truths' which form the basis of Jaina metaphysics.

The second and third substances, called *dharma* and *adharma*, are posited for technical reasons. They stand respectively for the factors that assist movement and rest, without themselves being set into motion or coming to rest. They are formless, coextensive with worldly space and without qualities. The function of these two substances is usually explained on the analogy of water and fish. Just as water is not responsible for the fish's movement or its being stationary but makes these possible, so too *dharma* and *adharma* offer the condition for the possibility of movement or rest to souls and matter.

The fourth substance, ether or space (*ākāśa*), does not itself occupy space but is space itself, giving the other substances the place they require in which to exist. According to Jaina cosmology, space is the only substance that encompasses both the world (*loka*) and the nonworld (*aloka*). The world is divided into the underworld or hells, in which souls reap the fruits of their unmeritorious deeds, the middle world, which corresponds to our world and in which only human beings can perform the necessary austerities for absolute freedom, the world of different kinds of heaven, in which souls reap the fruits of their meritorious acts and, finally, the world above the heavens, to which a liberated soul (one that has completely rid itself of the burden of all its positive and negative karma) rises, to remain there in its pristine nature. The nonworld serves no purpose at all and is described as being empty. However, it demarcates the area along which liberated souls eternally exist (see Cosmology and cosmogony, Indian theories of §3).

Time (*kāla*) is the fifth non-sentient substance; it has no spatial points, its smallest unit being called a moment (*samaya*), out of which the other units of time - minutes, hours, and so on - are then derived. Time enables change to take place without itself changing. On the absolute level, time exists as an undifferentiated continuum for the liberated souls. On the cosmological level, it has two broad phases or eras, and is described on the analogy of a wheel with twelve spokes. The world goes through six phases of gradual descent, representing the decline of values, then through six phases of gradual ascent, representing a steady improvement of values. Both phases involve fantastic spans of time. The Śvetāmbara tradition hesitates to accord time the status of an independent substance, though it uses it as a category necessary to explain the course of the world. The Digambaras and the Śvetāmbaras, the two main groups of Jainism, differ in several philosophically minor issues.

What the above ontological structure of reality does not do is explain the original cause of the association of the souls with the other substances, and especially how karma originally attaches itself to the soul to its detriment. This problem applies *mutatis mutandis* to all schools of Indian thought which speak of the possibility of liberation from the bondage of karma, where the basic issue is simply presupposed and is either left unanswered or euphemistically avoided by saying, as the Jainas also do, that the association between the souls and the world has been so since beginningless time (*anādi*). The karma theory was accepted a priori without the need to explain a

first cause and the issue does not seem to be regarded as a problem. Moreover, Jainism shares with Buddhism and Hinduism the basic assumption that the human situation is characterized by suffering and pain; more important than seeking the first cause is what can be done to free oneself from this predicament. Past deeds cannot be undone, but once the basic presuppositions are accepted - especially that the process of karma, which is responsible for the cycles of existence, can be interfered with - the individual is free to opt for the path that promises the goal of absolute freedom. The Jainas do not have recourse to a creator god who sets the universe and life in the world in motion. Everything operates through its own inner dynamics and the role of the individual is to meditate on the basic truths of reality.

Jaina metaphysics is based on seven truths or fundamental principles (*tattva*) which are an attempt to explain the human predicament. The first two are the two ontological categories of the soul and the non-soul discussed above, namely the truth that they exist. For the remaining five truths the factor of karma has to be brought into play because all are connected to it. The third truth is that through the interaction, technically called *yoga*, between the two substances, soul and non-soul, matter flows into (*āsrava*) the soul, clings to it, becomes converted into karma and - the fourth truth - acts as a factor of bondage (*bandha*), restricting the manifestation of the consciousness intrinsic to it. The fifth truth states that a stoppage (*saṃvara*) of new karma is possible through asceticism. An intensification of this burns up the existing karma - this sixth truth is expressed by the word *nirjarā*. The final truth is that when the soul is freed from the influence of karma, it reaches the goal of Jaina teaching, which is liberation (*mokṣa*).

It must be noted that the later tradition included two more categories after bondage (*bandha*), namely *puṇya* and *pāpa*, which may be translated as ‘good’ and ’bad’. They refer to two kinds of bondage caused by one’s deeds. The ethical implications of these terms underscore the significance of Jaina ethics for Jaina soteriology.

2 Epistemology and logic

Epistemology and logic are closely linked in the Indian tradition because logical issues are largely associated with problems related to inference which, in turn, is regarded as a means or instrument of knowledge. The occupation with abstract, logical ideas related to inference was the favourite topic of the Nyāya school, particularly in its Navya-Nyāya form, which flourished between about the thirteenth and the seventeenth century. Jaina logic adopts much of the Nyāya language and method (see Nyāya-Vaiśeṣika §6).

The history of Indian logic can be traced back at least to pre-Christian works on Indian medicine in which physicians are advised to know about a long list of categories, epistemological and logical, when participating in public debates. In the commentary literature to the basic philosophical treatises, especially after the third century AD, epistemology takes a prominent position, each school presenting its own theory of knowledge after criticizing other views. Jainism made its own unique contribution to this mainstream development by also occupying itself with the basic epistemological issues, namely, with those concerning the nature of knowledge, how knowledge is derived, and in what way knowledge can be said to be reliable. These issues feature in the Indian context in the form of the consistent attention paid to what exactly constitutes cognition, the means of cognition, and the validity of cognition. The problems dealt with in epistemology broadly served two functions. They represented an attempt to provide the basis for an intelligible discourse on matters of common, everyday experience and, albeit indirectly, to distinguish this area of discourse from what constitutes the knowledge of ultimate reality. Thus Indian epistemology entails an implicit metaphysical concern when it is employed in the context not only of the knowledge of one’s intrinsic nature but also of reality as such.

Knowledge for the Jainas takes place in the soul, which, without the limiting factor of karma, is omniscient. Human beings have partial knowledge - the object of knowledge is known partially and the means of knowledge do not operate to their full capacity. The Jainas have an intricate theory of knowledge concerning the fundamental principles (*tattva*), and the first systematic presentation of it was given by Umāsvāti (see §4) in the first chapter of his *Tattvārthasūtra* (fourth or fifth century). The following is a summary of his theory; the bracketed numbers are the relevant aphorisms. The knowledge of the basic Jaina truths is said to be obtained through means or instruments of knowledge (*pramāṇa*) which can yield a comprehensive knowledge of an object, and through particular standpoints which yield a partial knowledge (*naya*; 6) (see Manifoldness, Jaina theory of). The first type of knowledge is of five kinds: sensory knowledge, scriptural knowledge, clairvoyance, telepathy and omniscience (9-10), among which again the first two are described as being indirect means of knowledge (*parokṣa*), with the

others furnishing direct knowledge (*pratyakṣa*; 11-12), by which it is meant that the object is known directly by the soul (not, for example, as with inferring fire through the cognition of smoke). Jainism is unique in the Indian tradition in regarding sensory perception as an indirect means of knowledge, because it does not take place directly through the soul, although some notable Jaina thinkers (such as Akalaōka and Māṇikyanandin) are exceptions.

Synonyms for sensory knowledge are remembrance, recognition, 'induction' and 'deduction', all caused by the senses and the mind (13-14). Scriptural knowledge, which is based on sensory knowledge, is of two, several or twelve kinds (20), depending on which scriptural works are taken as authoritative. The range of sensory and scriptural knowledge extends to all the six substances (soul and the five non-sentient substances), but not in all their modes (26). Clairvoyance is possessed by divine and infernal beings, and can be obtained by human beings and animals through a destruction and/or subsidence of the karma which hinders it (21-2). Its scope is all entities that have form (27). Telepathy is of two kinds, which are distinguished on the basis of purity and infallibility (23-4). Its scope is infinitely greater than that of clairvoyance (28). The scope of omniscience, on the other hand, extends to all substances in all their modes simultaneously (29). Sensory knowledge, scriptural knowledge and clairvoyance are explicitly referred to as means of knowledge that may also be erroneous (31).

The basic division of knowledge into the direct and indirect types is generally retained by all Jaina thinkers, though with differences in the way in which they are further classified. For example, Akalaōka (*c.* eighth century) regarded sensory knowledge as a direct means of knowledge, and Māṇikyanandin (*c.* ninth century) evolved a much more complex system with several more subvarieties. When a description of inference as a means of knowledge is given, the occasion is used to discuss logical issues in detail. Jaina logic applies the Nyāya language within its own metaphysical presuppositions. Some of the basic features of Indian logic which the Jainas also adopt involve: (1) the use of inference in the accepted standard form, to convince others of the validity of an argument, clearly distinguishing the three basic terms of the major, the middle and the minor; (2) an example or instance, which can also be in a negative form, based on the invariable concomitance of or inseparable connection between the middle term of an argument and all cases in which the major term obtains; and (3) a concern with types of fallacies based not only on pervasion but also on the relation of the different terms with each other and the locus in which they apply (see Inference, Indian theories of).

With his *Nyāyāvatāra (The Descent of Logic)*, Siddhasena Divākara (*c.* fifth century) is perhaps the first Jaina thinker who deals with these and other issues in Jaina logic. A few selected examples will show how Jainism imbibed the logic of the Nyāya school, the structure of which it applied to prove its own arguments and to show errors in the arguments of others. The classical example for an understanding of the abstract ideas expressed here is that of smoke and fire, the basic elements of which, as they feature in a five-step inferential argument, are: (1) the theory - the hill (minor term) possesses fire (major term); (2) the cause or reason - because of smoke (middle term); (3) the example based on invariable concomitance - wherever there is smoke there is fire (as in a kitchen); (4) the application of the theory - similarly the mountain has smoke; and (5) the conclusion - therefore it is the same, that is, has fire.

> A statement expressive of the reason (*i.e.*, the mark or the middle term, called *hetu*) which is inseparably connected with that which is to be proved (*i.e.*, the major term, called *sādhya*) having been composed of the minor term (called *pakṣa*, signifying a side or place), etc., is called an inference for the sake of others (*parārthānumāna*).
>
> (*Nyāyāvatāra*, stanza 13; trans. S.C. Vidyabhusan)

> Where the inseparable connection of the major term (*sādhya*) and the middle term (*sādhana* or *hetu*) is shown by homogeneousness (*sādharmya*), the example is called a homogeneous one, on account of the connection (between those terms) being recollected.
>
> (stanza 18)

> The reason (*i.e.*, the middle term called *hetu*) has been defined as that which cannot exist except in connection with the major term (*sādhya*); the fallacy of the reason (*hetvābhāsa*) arises from nonconception, doubt or misconception about it (the middle term).
>
> (stanza 22)

The Indians did not develop a system for presenting their abstract ideas symbolically; natural language

(pre-eminently Sanskrit), used in such a technical way that only the specialist can decipher the arguments, served the function. However, there have been several successful contemporary attempts to translate Indian logic into the terminology of modern Western logic (Ingalls 1951; Matilal 1968, 1971; Staal 1960).

3 Ethics

Jaina ethics evolved out of the rules for the ascetic, which served as the model, with necessary changes, for the laity as well. The basic ascetic rules are encapsulated in the so-called five great vows (*mahāvratas*) ascribed to Mahāvīra, which seem to be a summary of Indian asceticism as a whole from ancient times. The first and foremost of these is nonviolence, which entails total abstinence in thought, word and deed from injury to all life forms. The principle of life is the souls which inhabit atoms, so Jainism emphasizes extreme care with reference not only to plant and animal life forms but also to those in earth, water, fire and air. The vow of nonviolence is extended to include not making another perform acts of violence and not approving them in any way. Further, ascetics, who are usually wandering mendicants, are required to stay in one area during the monsoon to avoid unintentionally disturbing and harming life forms in and as a result of the abundance of water. To avoid inadvertently injuring insects, certain groups of ascetics even cover their mouths and noses, and carry a whisk to keep insects away. The vow of not eating after sunset for the same reason is an ascetic rule and one which is considered to be ethically meritorious when practised by householders.

The other vows are: to abstain from lying, and to take care not to use violent or harmful speech; not to take what is not given; to lead a life of celibacy; and to renounce attachment to the objects of the world, that is, to renounce possession of property.

Jainism's extreme emphasis on nonviolence is grounded in its metaphysics. Violence is responsible for the maximum amount of karma that can be accumulated by the soul, and since liberation is possible only when karma is completely destroyed, the task is reduced through an avoidance of violent deeds. Physical activity *per se* is responsible for the accumulation of karma; abstinence from it is symbolized by the famous iconic representations of ascetics standing upright over such long periods that vines grow up their legs - physical control represents the mental control that is also necessary to avoid subtle, inner movement. With the axiom 'nonviolence is the highest religion', Jainism summarizes the basis of its ethics and religious life.

The ascetic vows are mirrored in the rules for what the laity should do to exemplify an ethical life. The vow of chastity is relaxed for householders, with sexual contact restricted to the married partner; bearing in mind the ascetic's great vow of chastity, however, restraint is enjoined as a virtue. Other vows which are included in the religious life of devotees include: nonattachment to property and possessions, and the aim of leading a simple life; religious giving or donation; eschewing excessive and unnecessary travel; fasting on auspicious days of the Jaina calendar.

The mutual reliance of ascetics and laity is evident throughout the history of Jainism. The ascetics do not cook, and rely on the devotees for their daily subsistence, and the laity require the ascetics for their religious teaching and advice. Confession of ethical transgressions belongs to the religious life of both ascetics and laity, who atone for them by penances of religious purification, the aim being to reduce or even completely annihilate the karmic effect of the infringement.

Jainism, together with Buddhism, shows how a religious and virtuous life is possible without the idea of a creator god to whom one can turn, one who is ultimately made responsible for the human condition. Models for ethical life in Jainism are provided by the biographies of the twenty-four Jinas, the conquerors of the passions, of whom Mahāvīra was the last. Indeed, they are worshipped as divine beings, even though the tradition represents them as human beings who through their extreme asceticism gained an insight into the nature of reality, on account of which they are regarded as omniscient. Their lives serve as a guiding principle and, according to the tradition, an emulation of their virtues can lead one to the same goal of liberation that they achieved.

Under the rubric of ethics the issue of voluntary death may be mentioned. Inscriptional evidence records its occurrence throughout the history of Jainism and even in contemporary times, though it has been rare since about the twelfth century. This kind of death, open to both ascetics and laity, is a death that literally makes 'the physical body and the internal passions emaciated' (*sallekhanā*; Pūjyapāda on *Tattvārthasūtra* V, 22); often it serves to accelerate the death process already in progress. Jainism contrasts this with death that occurs through suicide,

which the Jainas eschew because passions such as 'attachment, aversion or infatuation' are involved in suicidal death. By virtue of its excellence, the passionless death (which is performed under strict conditions), is regarded as the most effective ascetic practice to rid the soul of binding passions and to terminate an ethical life.

4 Kundakunda, Umāsvāti and Siddhasena Divākara

These three thinkers are the pioneers of Jaina philosophy whose basic ideas set the trend for most later thinkers. Biographical details of all of them are mixed with legend and there are differences of opinion as to whether they really wrote all the works ascribed to them; there is also a lack of consensus about their dates.

Kundakunda. If recent research is correct in considering him to have belonged to the second or third century AD, then this would make him the first significant and independent thinker of the post-canonical period whose views are accepted as representing the essence of Jaina thought. Although he was a pioneering Digambara thinker, probably from South India, appreciation for his views also comes from the Śvetāmbara sect of Jainism. He was also known as Padmanandi. The name Gṛdhrapiccha, erroneously used for him since about the fourteenth century, has led to confusion because it is also an alias for Umāsvāti.

A total of eighty-four works on various themes are ascribed to Kundakunda, of which fifteen are extant and three may be said to be philosophical masterpieces, all written in the Prakrit language. These are the *Pañcāstikāyasāra (Essence of the Five Existents)*, the *Pravacanasāra (Essence of the Scripture)* and the *Samayasāra (Essence of the Doctrine)*. The *Pañcāstikāyasāra* is an elementary work dealing with the Jaina substances (excluding time because it does not occupy any spatial points) and the fundamental truths, to which two additional categories are added, namely the meritorious and demeritorious acts related to karma (*puṇya* and *pāpa*). The *Samayasāra* emphasizes, among other things, two standpoints mentioned in the canonical literature which seem to have no relation to the standard sevenfold standpoint (see Manifoldness, Jaina Theory of). These are the 'definitive' standpoint (*niścayanaya*), used synonymously with the 'pure' or 'transcendental' (*śuddha* or *paramārtha*) standpoint, and the 'mundane' standpoint (*vyavahārikanaya*). It is an illuminating work dealing with the nature of the soul and its contamination by matter, and whether the soul's intrinsic nature is in any way affected or changed through karma bondage in so far as it is the doer and enjoyer of activities. An attempt is made to reconcile these problems, solutions to which depend on the standpoint from which one approaches the issues. The *Pravacanasāra* is an insightful work whose three sections clearly delineate its scope: knowledge, the objects of knowledge, and conduct. The problem of substance, quality and mode, is one of the pivotal issues in Jaina philosophy and a few points are outlined below in order to show how Kundakunda deals with it. It forms the subject matter of the second section of the *Pravacanasāra*, which the tenth-century commentator Amṛtacandra says Kundakunda 'properly discusses'.

The problem is basically that of how change in the world may be explained given the permanent, eternal nature of the two basic substances of ultimate reality; this has obvious implications for the essential nature of the soul. Kundakunda begins the section with the statement: 'The object of knowledge is made up of substances, which are said to be characterized by qualities, and with which, moreover, are (associated) the modifications' (*Pravacanasāra* II, 1; trans. A.N. Upadhye). The basic problem is then evident when he says: 'There can be no origination without destruction, nor is there destruction without origination; origination and destruction are not possible in the absence of permanent substantiality' (II, 8). How 'origination' and 'destruction', which in fact refer to change, are to be understood is expressed by Kundakunda in typical Jaina language in II, 19: 'The substance forever retains its position, its own nature, as endowed with positive and negative conditions according as it is looked at from the substantial and the modificational viewpoints.' This is further elaborated:

> All substances are nondifferent from the substantial viewpoint, but again they are different from the modificational view-point, because of the individual modification pervading it for the time being. According to some modification or the other it is stated that a substance exists, does not exist, is indescribable, is both or otherwise.
>
> (II, 22-3)

What Kundakunda means by origination and destruction is distinctively Jainist and is clarified later in the same section: 'In this world, in which modifications originate and pass away at every moment, nothing is absolutely produced or destroyed; what is the production of one modification is the destruction of another; and thus

origination and destruction are different' (II, 27). The change that occurs in matter is understandable on the analogy of objects and colour. Just as gold (regarded here as a substance), for example, can have not only different shades of colour (with colour being its basic quality) but also different forms (with the object made out of gold being its modification), so too all substances retain their substantiality despite the apparent destruction of their qualities and modes. The situation is more complicated with the soul substance. The problem is technical, and relates to two 'operations' (*upayogas*) ascribed to the soul, namely 'indeterminate intuition' and 'determinate knowledge'; these operations are described as two qualities (*guṇas*) of the soul. The concern is with the unity or identity of the soul and involves the question of whether the two *upayogas* operate in the soul simultaneously or in succession, and if in succession, which is first, and whether they maintain their distinctness in the state of omniscience. Kundakunda maintains that they operate successively at the mundane level and simultaneously at the transcendental level of omniscience. His view, which is also held by Umāsvāti, represents the attitude of the Digambara sect and is opposed, for example, by Siddhasena Divākara, who, in regarding quality and mode as synonyms, says that they are not separate operations in the state of omniscience. His view represents the general Śvetāmbara standpoint, based on the fact that the canonical literature distinguishes only substance and quality, without mentioning the standpoint connected with the modifications of a substance.

Umāsvāti. He is famous for the first Jaina work written in Sanskrit, called the *Tattvārthasūtra* or *Tattvārthādhigamasūtra (Mnemonics on the Meaning of the Fundamental Principles)*. Again, biographical details are scanty and both the sects of Jainism claim him as one of their own (with the Digambaras also calling him Umāsvāmī) and regard his work, in traditional Indian manner, as authoritive for Jaina thought. His dates vary from the second to the fifth centuries AD, with recent preference for the fourth or fifth centuries. Of the five works ascribed to him, the *Praśamaratiprakaraṇa (Treatise on the Love for Tranquility)* - a popular work dealing with ethical issues and addressed to ascetics and householders - and the *Tattvārthasūtra* are philosophically important. There is an ongoing debate about whether, as the Śvetāmbaras believe, a commentary on the *Tattvārthasūtra* was written by Umāsvāti himself, or whether, as the Digambaras believe, the first commentary on it is Pūjyapūda's fifth-century *Sarvārthasiddhi (Attainment of the Meaning of Everything)*. There are two versions of the work by the two sects, with hardly any philosophically significant differences.

The *Tattvārthasūtra* contains a series of aphorisms, divided into ten chapters, which are understandable only with a commentary. The value of the work is evident from the fact that throughout the history of Jaina philosophy, every major thinker has written a commentary on it. Until around the tenth century, the Digambara thinkers, such as Akalaōka (*c.* eighth century) and Vidyānandin (*c.* ninth century), took centre stage. They wrote in a difficult style and hardly any research related exclusively to their writings has been done.

Commentators often took the opportunity to criticize other views and defend the Jaina standpoint. For example, in his commentary on Umāsvāti's aphorism on the means of knowledge and the standpoints (*Tattvārthasūtra* I, 6), Vidyānandin enters into an interestingly detailed debate with the Buddhists (even quoting from Dharmakīrti's seventh-century work on the means of knowledge, *Pramāṇavārttika*) regarding the knowledge of an object, namely, whether an object is cognized as a whole or in parts. The issue is raised in the context of the Buddhist view that an object as a whole does not exist, against which it is asserted that a part of an object, which can be understood to be a whole in itself, cannot then be cognized, putting the Buddhist standpoint in jeopardy. Vidyānandin, applying the Jaina theory of standpoints, says that both a part of an object and the object as a whole are cognized from different standpoints.

Umāsvāti's contribution lies in presenting the basic issues in Jaina philosophy in a systematic form, so much so that his work usually forms the standard for Jaina thought as a whole. The ontology, metaphysics and epistemology summarized in §§1-2 are based on his *Tattvārthasūtra*.

Siddhasena Divākara. It seems certain that Siddhasena lived in the fifth century AD and that he wrote in Sanskrit and Prakrit. He is generally considered to have belonged to the Śvetāmbara sect, although he is claimed by both the sects of Jainism as one of their own, serving as the first Jaina logician for both. Apart from the canonical literature, there seems to be little influence from other sources in his writings, despite similarities of ideas between him and, for example, Kundakunda. Reference to Siddhasena's work on logic, the *Nyāyāvatāra*, has been made in §2. His second most important work on philosophy, written in Prakrit, is the *Sammaisutta (Mnemonics on Proper Understanding*; Sanskrit, *Sanmatisūtra*), dealing with the seven Jaina standpoints, knowledge and the objects of

knowledge. It is in the last section that he clearly discusses the issue of the standpoints, taking into consideration the quality of a substance, which according to him is not sanctioned by Jaina scripture as a separate category.

Siddhasena is also credited with having written twenty-one short compositions, each consisting of thirty-two verses (and simply called the ‘thirty-twos’), on a variety of themes, including eulogies to Mahāvīra, critiques of Buddhist and Hindu schools, and an exposition of Jaina concepts. The work on logic is also written in this form.

5 Other notable philosophers

A host of philosophers, some already referred to, have followed Kundakunda, Umāsvati and Siddhasena. Suffice it to make a few remarks on the contribution of five further thinkers, to cover the period until the seventeenth century. Haribhadra, perhaps of the eighth century, belongs to the Śvetāmbara sect and is praised as a great teacher of Jainism. Apart from works on classical yoga as a system of meditational practice, which he applies in the Jaina context, he is famous for his compendium of six systems, the *Ṣaḍdarśanasamuccaya*, in which he includes Jaina philosophy together with Buddhism and four Hindu systems. He evinces a trait common among Jaina thinkers and evident already in Akalaōka, namely an excellent acquaintance with other systems of Indian thought so as to enter into authentic debate with them. Some scholars (notably Halbfass 1981) have seen a connection between this and the emergence of the theory of manifoldness which became the hallmark of Jainism.

There is a relative consensus about the dates of the polyhistor Hemacandra (1089-1172), whose contribution lies especially in systematizing and upholding the Śvetāmbara tradition of Jainism. Although his influence is stronger in the field of Jaina literature, insightful attacks on other philosophical views are also evident, for example, in his *Anyayogavyavacchedikā(Critique of Other Schools)*. His treatise on yoga (*Yogaśāstra*), dealing with Jaina ethics, is still an exemplary work.

Abhayadeva, active in the eleventh century, needs to be mentioned for his commentaries on the Jaina canonical works. The value of the commentary literature in India lies not only in the explication and clarification of basic texts, but also in the fact that commentators often present views which they might not otherwise have found occasion to express. In his commentary on the *Sthānāōgasūtra*, for example, a canonical text which is a compendium of Jaina doctrine, ethics and cosmology, Abhayadeva takes the opportunity to discuss epistemological issues.

Vādidevasūri (twelfth century) is not only famous as a Śvetāmbara thinker who defeated in a public debate the Digambara thinker Kumudacandra, but for his excellent *Syādvādaratnākara (The Ocean of Manifoldness)*. This is a commentary on his own work on knowledge and standpoints, inspired by the work of the ninth-century Digambara thinker Māṇikyanandin.

Yaśovijaya (1624-88) is perhaps the last intellectual giant of Jainism, praised not only for his acumen as a logician, but also for his vast knowledge of Jainism and other traditions. This erudite Śvetāmbara scholar is credited with up to a hundred works, including an attack on the great logician of Navya-Nyāya, Raghunātha Śiromaṇi (early sixteenth century), and a commentary on a work by Vidyānandin.

Jaina philosophy never really took the centre of the philosophical stage in India, but that does not reduce its significance. Royal patronage from Mahāvīra’s time till about the sixteenth century, especially in Gujarat and Karnataka, and the architectural masterpieces of its temples bear witness to the social impact it has had throughout its history. Jaina monasteries were and are centres of learning, and research undertaken since the early 1980s is gradually reaping the benefit of the Jaina virtue of collecting and copying manuscripts of all disciplines for the sake of acquiring religious merit. It is thanks to this that a Jaina scholar, Sukhlalji Sanghvi, was able to publish in 1940 an original eighth-century work on Indian materialism by one Jayarāśi, based on the only extant manuscript. Further, the European Indologist Erich Frauwallner, in a posthumously published work (Frauwallner 1984), convincingly proves the original atheistic beginnings of the Vaiśeṣika system on the basis of the information about it from Haribhadra. The Jaina libraries contain a wealth of information relevant not only to philosophy but also to various aspects of Indian studies which have yet to be researched. Unfortunately, studies in Jainism suffer from a lack of published texts, though an earnest attempt has been made in the West and by the various Jaina institutions in India to make available critical editions, translations and studies. Clearly this has important implications for the study of the history of Indian philosophy.

See also: Mahāvīra

JAYANDRA SONI

References and further reading

Bhatt, B. (1974) 'Vyavahāra-naya and Niścaya-naya in Kundakunda's Works', *Zeitschrift der Deutschen Morgenlandischen Gesellschaft*, supplementary vol. 2: 279-91.(Deals with the two standpoints mentioned in §4.)

Folkert, K.W. (1933) *Scripture and Community: Collected Essays on the Jains*, ed. J.E. Cort, Atlanta, GA: Scholars Press.(Deals not only with aspects of Jaina philosophy, but also Jaina scripture, monastic praxis, lay-monastic relations, rituals and community.)

Frauwallner, E. (1984) *Nachgelassene Werke I. Aufsätze, Beiträge, Skizzen (Posthumous Works I. Essays, Contributions and Outlines)*, ed. E. Steinkellner, Vienna: Verlag der österreichischen Akademie der Wissenschaften; trans. J. Soni, *Erich Frauwallner's Posthumous Essays*, New Delhi: Aditya Prakasan, 1994. (Frauwallner refers to the Jaina polyhistor Haribhadra to establish the original atheism of Vaiśeṣika philosophy.)

Halbfass, W. (1981) *Indien und Europa. Perspektiven ihrer geistigen Begegnung*, Basle and Stuttgart: Schwabe & Co.; *India and Europe: An Essay in Philosophical Understanding*, Delhi: Motilal Banarsidass, 1990.(See the chapter on doxography for his view described in §5; the 1990 edition contains additional material.)

Haribhadra (*c.* 8th century) *Ṣaḍdarśanasamuccaya (Compendium of Six Systems)*, ed. M.K. Jain, Delhi: Bharatiya Jnanpith Publication, 1970, 3rd edn 1989; trans. K.S. Murphy, *A Compendium of Six Philosophies*, Tenali: Tagore Publishing House, 1957. (The Jain edition includes an explanatory translation of the text, and the commentaries of Guṇaratna and Somatilaka.)

Hemacandra (1089-1172) *Anyayogavyavacchedikā(Critique of Other Schools)*, Sanskrit with Hindi trans. by P. Sāhityācārya, Agas: Rājacandra Jaina Śāstramālā, 1970; trans. F.W. Thomas, in *The Flower-Spray of the Quodammodo Doctrine*, Delhi: Motilal Banarsidass, 1968.(This work was made famous in the thirteenth century through Malliṣeṇa's commentary on it, the *Syādvādamañjarī*, and is usually published with it.)

Hemacandra (1089-1172) *Yogaśāstra*, trans. A.S. Gopani, *The Yoga Shastra of Hemchandracharya (A 12th Century Guide to Jain Yoga)*, Jaipur: Prakrit Bharti Academy, 1989.(Gopani's edition also includes the Sanskrit text.)

Ingalls, D.H. (1951) *Materials for the Study of Navya-Nyāya Logic*, Harvard Oriental Series 40, Cambridge: Harvard University Press; repr. Delhi: Motilal Banarsidass, 1987.(One of the earliest serious attempts to present Indian logic in terms of Western logic.)

Jaini, P.S. (1979) *The Jaina Path of Purification*, Delhi: Motilal Banarsidass.(An authoritative work dealing with the nature of reality, the mechanism of bondage, the mendicant path and the attainment of the goal, among other matters.)

Jayarāśi (*c.* 9th century) *Tattvopaplavasiṃha (The Lion that Destroys Philosophical Categories)*, ed. S. Sanghavi and R. Parikh, Gaekward Oriental Series 87, Baroda: University of Baroda, 1940; trans. E. Franco, *Perception, Knowledge and Disbelief: A Study of Jayarāśi's Scepticism*, Wiesbaden: Franz Steiner Verlag, 1987; 2nd edn, Delhi: Motilal Banarsidass, 1994.(The only extant text of the Indian materialist school called Cārvāka, known for its radical scepticism.)

Kundakunda (2nd or 3rd century) *Pravacanasāra (Essence of the Scripture)*, ed. and trans. A.N. Upadhye, Gujarat: Shrimad Rajachandra Ashrama, 1984.(The Prakrit text, critically edited with Sanskrit and Hindi commentaries, together with an English translation and a scholarly introduction on the linguistic and philosophical aspects of this and other works by Kundakunda.)

Kundakunda (2nd or 3rd century) *Pañcāstikāyasāra (Essence of the Five Existents)*, ed. A.N. Upadhye, *Pañcāstikāyasāra: The Building of the Cosmos*, Delhi: Bharatiya Jnanpith Publication, 1975.(Contains the Prakrit text, a Sanskrit translation by the tenth-century commentator Amṛtacandra, an English translation by A. Chakravartinayanar, and an introduction to Jaina philosophy.)

Kundakunda (2nd or 3rd century) *Samayasāra (Essence of the Doctrine)*, ed. Pannalāla Sāhityācārya, Varanasi: ŚrīGaṇeśaprasāda VarṇīGranthamālā, 1969.(Prakrit text with Hindi translation and commentary, with an introduction including Kundakunda's biography.)

Māṇikyanandin (9th century) *Parīkṣāmukham (Gateway of Investigation)*, trans. and ed. S.C. Ghosal, *Parīkṣāmukham by Māṇikyanandī*, Lucknow: Ajitasram, 1940.(The text of this influential ninth-century Jaina thinker is translated into English, with comments; the introduction contains an excellent survey of Jaina epistemology.)

Matilal, B.K. (1968) *The Navya-Nyāya Doctrine of Negation: The Semantics and Ontology of Negative*

Statements in Navya-Nyāya Philosophy, Harvard Oriental Series 46, Cambridge, MA: Harvard University Press.(Presents Indian logic in terms of Western logic.)

Matilal, B.K. (1971) *Epistemology, Logic and Grammar in Indian Philosophical Analysis*, The Hague: Mouton. (Presents Indian logic in Western terms, but not restricted to one school as in Matilal 1968.)

Ohira, S. (1982) *A Study of the Tattvārthasūtra with Bhaṣya. With Special Reference to Authorship and Date*, Ahmedabad: L.D. Institute of Indology. (Deals with the issues of date and authorship of Umāsvāti's pioneering work on Jaina philosophy, including the debate over the authorship of an auto-commentary.)

Pūjyapāda (*c.* 5th century) *Sarvārthasiddhi (Attainment of the Meaning of Everything)*, trans. S.A. Jain, *Reality*, Madras: Jwalamalini Trust, 1960, repr. 1992. (This is the first commentary, according to the Digambara sect, on Umāsvāti's *Tattvārthasūtra*, the standard work on Jaina philosophy.)

Settar, S. (1986) *Inviting Death: Historical Experiments on Sepulchral Hill*, Dharwad: Karnatak University.(Deals with the 'passionless death' mentioned in §3.)

Settar, S. (1990) *Pursuing Death: Philosophy and Practice of Voluntary Termination of Life*, Dharwad: Karnatak University.(Deals with the 'passionless death' mentioned in §3.)

Siddhasena (5th century) *Nyāyāvatāra (The Descent of Logic)*, trans. S.C. Vidyabhusan, in A.N. Upadhye (ed.) *Siddhasena's Nyāyāvatāra and Other Works*, Bombay: Jaina Sāhitya Vikāsa Maṇḍala.(A collection of works containing the text of the *Nyāyāvatāra*, with English translation and notes, the texts of two other works by Siddhasena, *Twenty-one Dvātriṃśikās* and the *Sammaisuttam (Mnemonics on Proper Understanding)*, an informative introduction and a bibliographic survey.)

Sikdar, J.C. (1991) *Theory of Reality in Jaina Philosophy*, Varanasi: P.V. Research Institute. (Deals with Jaina ontology on the basis of original texts, comparing the Jaina view with other schools of Indian thought.)

Singh, R.J. (1974) *The Jaina Concept of Omniscience*, Ahmedabad: L.D. Institute of Indology.(Examines the traditional arguments for and against omniscience.)

Soni, J. (1991) '*Dravya*, *Guṇa* and *Paryāya* in Jaina Thought', *Journal of Indian Philosophy* 19: 75-81.(Places in philosophical context the issue of substance, quality and mode in Jainism; a more detailed discussion of the question of change and permanence according to Kundakunda described in §4.)

Soni, J. (1996a) *Aspects of Jaina Philosophy*, Madras: Research Foundation for Jainology, for University of Madras Department of Jainism.(Deals with karma theory and Jaina ethics, the Jaina theory of manifoldness and Vidyānandin's commentary on *Tattvarthasūtra* I, 6, on the means of knowledge.)

Soni, J. (1996b) *The Notion of Āpta in Jaina Philosophy*, Toronto, Ont.: University of Toronto, Centre for South Asia Studies.(Discusses the Jaina view that a trustworthy person (*āpta*) is a source of knowledge and compares the Jaina view of reality with that of the Śaiva Siddhānta tradition.)

Soni, J. (forthcoming) 'Aspects of Jaina Epistemology with Special Reference to Vidyānandin', paper presented at conference 'Approaches to Jain Studies', Toronto March 31-April 2 1995, to be published in conference volume.(Deals especially with Vidyānandin's debate with the Buddhists - in which he even quotes Dharmakīrti - over the issue of parts and wholes, or universals and particulars, as mentioned in §3.)

Staal, J.F. (1960) 'Correlations Between Language and Logic in Indian Thought', *Bulletin of the School of African and Oriental Studies* 23: 109-22.(Presents Indian logic in terms of Western logic, as do some of Staal's subsequent articles.)

Suklalji, P. (1974) *Tattvārthasūtra of Vācaka Umāsvāti*, trans. K.K. Dixit, Ahmedabad: L.D. Institute of Indology. (A modern commentary by a renowned authority on Jaina philosophy, with an introduction about the original author and commentaries on the work.)

Tatia, N. (1951) *Studies in Jaina Philosophy*, Varanasi: P.V. Research Institute.(A systematic interpretation of Jaina philosophy based on a critique of other schools.)

Umāsvāti (*c.* 2nd century) *Tattvārthādhigamasūtra (Mnemonics on the Meaning of the Fundamental Principles)*, trans. and ed. J.L. Jaini, Sacred Books of the Jainas 2, Arrah: The Central Jaina Publishing House, 1920; repr. New Delhi: Today and Tomorrow's Printers and Publishers, 1990; trans. N. Tatia, *Tattvārtha Sūtra: That Which Is*, London: HarperCollins, 1994.(The first work in Sanskrit which summarizes the whole of Jaina philosophy in traditional, classical style. Jaini's edition supplies the differences between the Digambara and Śvetāmbara versions of the text. Tatia's translation takes into consideration the commentaries of Umāsvāti, Pūjyapāda and Sīddhasenagaṇi.)

Umāsvāti (*c.* 2nd century) *Praśamaratiprakaraṇa (Treatise on the Love for Tranquility)*, trans. and ed. Y.S. Shastri, Ahmedabad: L.D. Institute of Indology, 1989.(A popular and simple work presenting the basics of Jaina thought.)

Vādidevasūri (12th century) *Syādvādaratnākara (The Ocean of Manifoldness)*, Delhi: Bhāratīya Book Corporation, 1988.(Sanskrit text. An influential Śvetāmbara thinker deals with various aspects of Jaina

philosophy.)

Yaśovijaya (17th century) *Jaina Tarka Bhāṣā(The Language of Jaina Logic)*, trans. and ed. D. Bhargava, *Mahopādhyāya Yaśovijaya's Jaina Tarka Bhāṣā*, Delhi: Motilal Banarsidass, 1973.(A critical edition of a popular treatise on Jaina epistemology serving also as a manual of Jaina logic.)

Zydenbos, R.J. (1983) *Mokṣa in Jainism According to Umāsvāti*, Wiesbaden: Franz Steiner Verlag.(Deals with liberation, the subject of chapter 10 of Umāsvāti's *Tattvārthasūtra*; includes text and translation, a commentary attributed to Umāsvāti and an introduction to the problem.)

James of Viterbo (*c.*1255-1308)

James of Viterbo's writings reveal a loyalty to Augustine combined with an interest in Neoplatonic sources such as Proclus, Pseudo-Dionysius and Boethius. He also reveals a strong interest in the Greek commentators on Aristotle, in Simplicius as regards predication and language and in Themistius and John Philoponus as regards the nature of the intellect and intellectual cognition. Frequently, after presenting various positions on a topic and noting how they differ, he proposes to state what seems probable to him. He thus proposes, for example, a theory of innatism. In contrast, he shows no little certitude in his De regimine christiano (On Christian Rule), where he presents a markedly papalist political theory.

James entered the Order of the Hermits of Saint Augustine around 1272, and studied at Paris from 1275 to 1282 and again after 1285. He later succeeded to the chair of his teacher, Giles of Rome, which he held until 1300. He then taught at the Augustinian *studium generale* in Naples until 1302. He ended his career involved in active ministry as the Archbishop of Naples.

The major source for our knowledge of James' philosophical views is his *Disputationes de quolibet (Quodlibetal Disputations)*, written some time between 1293 and 1297 at Paris. James accepts Augustine's concept of seminal reasons present in matter. He takes them to be the beginning (*inchoatio*) of or aptitude (*aptitudo*) for the form that will actually exist when something is generated. Nonetheless, James emphasizes that the human soul, unlike other forms, comes to exist through its direct creation by God. He does not consider Aristotle's position on the origin of soul, that is, whether or not it comes from outside the fetus, to be evident (see Soul, Nature and Immortality of the).

When James analyzes the nature of the agent intellect, he adopts a theory of innate knowledge that resembles his theory of seminal reasons. He sets forth the positions presented by Philoponus and Themistius, as well as views taken from Augustine, Boethius and his own contemporaries. Although James accuses ancient and recent interpreters of Aristotle of attempting to shape his words to suit their own positions, he then does the same. He proposes that when God creates the soul he endows the human intellect, which is itself a general aptitude, with special aptitudes or fitnesses (*idoneitates*) that are only incomplete actualities. These are the potential contents of the cognitive acts of our intellect. Sense experience simply excites the human intellect, which then moves itself to actual knowledge. James sets forth a similar theory regarding special aptitudes in the human will, which is itself a general aptitude, in order to explain individual acts of volition and to maintain that the will, too, is fundamentally active in nature and not passive to anything external.

Some of James' metaphysical views merit attention. In his early *Quaestiones de divinis praedicamentis (Questions on Divine Categories)*, he discusses problems concerning language and predication regarding the Trinity, referring constantly to Simplicius' commentary on Aristotle's *Categories* (see Simplicius). James opts for his own version of analogical language. In his *Disputationes de quolibet*, after analyzing the differing views of Giles of Rome, Henry of Ghent and Godfrey of Fontaines on essence and existence, he proposes to put forth what appears probable to him. Basically, essence and existence are related as the abstract to the concrete. Although they are identical as regards the principal thing that they signify, namely the essence, they differ by a real composition, since existence signifies secondarily and concretely all the things that are joined to the essence and that essence needs in order to exist (see Existence). In like fashion, James pursues a question which also much interested Henry and Godfrey as well as John Duns Scotus, namely whether all beings comprise an essential order which must come to a stop (*status*). James appears to favor the view that the series stops with a highest creature.

James occasionally alludes to a widely adopted scheme of metaphysical hierarchy. Many of his contemporaries held that both God and matter (or non-being) serve as measures of all the grades within the order of things according as things approach to or recede from them. James appears to concentrate almost exclusively on God as the measure. None the less, the concept of metaphysical hierarchy enables him to account for the destiny of human beings. In order to communicate and represent divine goodness, God established different grades among created things. The human being is the one creature that is composed of both the spiritual and the bodily and that thereby serves, as it were, as the medium and bond (*medium et vinculum*) of both. The human soul perfects the body rather than being perfected by the body. Indeed, the end of the soul is not the perfection of the body but union with

something better than either the body or the soul, namely God.

See also: Aristotelianism, medieval; Augustinianism; Henry of Ghent; Language, medieval theories of; Neoplatonism

EDWARD P. MAHONEY

List of works

James of Viterbo (*c*.1255-1308) *Disputationes de quolibet (Quodlibetal Disputations)*, ed. E. Ypma, Cassiciacum, Supplementband I-IV, Würzburg: Augustinus-Verlag, 1968-75, 4 vols.(The major source for James' views.)

James of Viterbo (*c*.1255-1308) *Prima quaestio disputata de verbo (First Disputed Question on the Word)*, in C. Scanzillo, 'La prima quaestio disputata "De verbo" del Codice A. 971 della Biblioteca dell'Archiginnasio di Bologna', *Asprenas* new series 19, 1972, 25-61.(Disputed question on language.)

James of Viterbo (*c*.1255-1308) *Quaestiones de divinis praedicamentis (Questions on Divine Categories)*, q. I-X and XI-XVII ed. E. Ypma, *Corpus Scriptorum Augustinianorum*, Vol. V, Parts 1-2, Rome: Augustinianum, 1983, 1986.(Ypma has continued to publish the remaining questions in *Augustiniana*: q. XVIII, *Augustiniana* 38, 1988, 67-98; q. XIX, *Augustiniana* 39, 1989, 154-85; q. XX, *Augustiniana* 42, 1992, 351-78; q. XXI, *Augustiniana* 44, 1994, 177-208; q. XXII, *Augustiniana* 45, 1995, 299-318; q. XXIII, *Augustiniana* 46, 1996, 147-76; q. XXIV, *Augustiniana* 46, 1996, 339-69. There are thirty-two questions in the work.)

James of Viterbo (*c*.1255-1308) *De regimine christiano (On Christian Rule)* in H.-X. Arquillière, *Le plus ancien traité de l'Église: Jacques de Viterbe 'De regimine christiano' (1301-1302). Étude des sources et édition critique*, Paris: Gabriel Beauchesne, 1926; ed. and trans. R.W. Dyson, *On Christian Government: De Regimine Christiano*, Woodbridge: The Boydell Press, 1995.(James' work on political philosophy.)

References and further reading

Casado, F. (1951-3) 'El pensiamento filosófico del Beato Santiago de Viterbo' (The Philosophy of the Blessed St James of Viterbo), *La Ciudad de Dios* 163, 437-54; 164, 301-31; 165, 103-44, 283-302, 489-500. (Informative general overview.)

Mahoney, E.P. (1973) 'Themistius and the the Agent Intellect in James of Viterbo and Other Thirteenth Century Philosophers' (Saint Thomas, Siger of Brabant and Henry Bate), *Augustiniana* XXIII: 422-67.(Discusses James's use of Greek commentators on Aristotle, Themistius, Simplicius and Philoponus, and also his innatism.)

Mahoney, E.P. (1980) 'Metaphysical Foundations of the Hierarchy of Being according to Some Late Medieval and Renaissance Philosophers', in P. Morewedge (ed.) *Philosophies of Existence: Ancient and Medieval*, New York: Fordham University Press, 165-257. (Conceptual scheme for the Great Chain of Being.)

Mahoney, E.P. (1995) 'Duns Scotus and Medieval Discussions of Metaphysical Hierarchy: The Background of Scotus's "Essential Order" in Henry of Ghent, Godfrey of Fontaines and James of Viterbo', *Via Scoti: Methodologica ad mentem Joannis Duns Scoti, Atti Congresso Scotistico Internazionale Roma 9-11 marzo 1993*, vol. 1, ed. L. Sileo, Rome: PAA-Edizioni Antonianum, 359-74.(Evidence that others besides Scotus worried about the notion of an "essential order" and whether it must come to a stop.)

Phelps, M. (1980) 'The Theory of Seminal Reasons in James of Viterbo', *Augustiniana* 30: 271-83.(Clear statement of his theory together with medieval criticisms.)

Ruello, F. (1974-5) 'Les fondements de la liberté humaine selon Jacques de Viterbe O.E.S.A. Disputatio 1ª de Quolibet, q. VII (1293)' (The Foundations of Human Freedom According to James of Viterbo), *Augustiniana* 24, 283-347; 25, 114-42.(Close detailed analysis.)

Wippel, J.F. (1982) 'The Relationship between Essence and Existence in Late Thirteenth-Century Thought: Giles of Rome, Henry of Ghent, Godfrey of Fontaines, and James of Viterbo', in P. Morewedge (ed.) *Philosophies of Existence: Ancient and Medieval*, New York: Fordham University Press, 131-64.(Shows importance of James's theory in context of contemporary discussions.)

Ypma, E. (1974) 'Recherches sur la carrière scolaire et la bibliothèque de Jacques de Viterbe †1308' (The Scholarly Career and Library of James of Viterbo), *Augustiniana* 30: 247-82.(Fundamental examination of problem of dating and sources.)

Ypma, E. (1975) 'Recherches sur la productivité littéraire de Jacques de Viterbe jusqu'à 1300' (The Literary Productivity of James of Viterbo Before 1300), *Augustiniana* 25: 223-82.(Further on dating.)

James, William (1842-1910)

The American William James was motivated to philosophize by a desire to provide a philosophical ground for moral action. Moral effort presupposes that one has free will, that the world is not already the best of all possible worlds, and, for maximum effort, according to James, the belief that there is a God who is also on the side of good.

In his famous, often misunderstood paper 'The Will to Believe', *James defended one's right to believe in advance of the evidence when one's belief has momentous consequences for one's conduct and success, and a decision cannot be postponed. One such belief is the belief in objective values. Generally, a belief is objective if it meets a standard independent of the believer's own thought. In morals, objective values emerge from each person's subjective valuings, whatever their psychological source, when these valuings become the values of a community of persons who care for one another. Still, even in such a community there will be conflicting claims, and the obligations generated by these claims will need to be ranked and conflicts resolved. James' solution is to say that the more inclusive claim - the claim that can be satisfied with the lesser cost of unsatisfied claims - is to be ranked higher. This is not to be mistaken for utilitarianism: James is not a hedonist, and it is not clear what he means by the most inclusive claim.*

A concern for others makes sense only if there are others who inhabit with us a common world. Pragmatism, which he co-founded with C.S. Peirce, and radical empiricism provide James' answer to those who would be sceptics concerning the existence of the common-sense world. Pragmatism is both a theory of meaning and a theory of truth. As a theory of meaning it aims at clarity; our thoughts of an object are clear when we know what effects it will have and what reactions we are to prepare. As a theory of truth, pragmatism makes clear what is meant by 'agreement' in the common formula that a belief is true if it agrees with reality. Only in the simplest cases can we verify a belief directly - for example, we can verify that the soup is too salty by tasting it - and a belief is indirectly verified if one acts on it and that action does not lead to unanticipated consequences. Contrary to a widespread misunderstanding, this does not mean that James defines truth as that which is useful; rather, he points out that it is, in fact, useful to believe what is true.

James rejects the dualism of common sense and of many philosophers, but he is neither a materialist nor an idealist, rather what he calls a 'pure experience' (for example, your seeing this page) can be taken as an event in your (mental) history or as an event in the page's (physical) history. But there is no 'substance' called 'pure experience': there are only many different pure experiences. You and I can experience the same page, because an event in your mental history and an event in mine can be taken to be events in the same physical history of the page; James may even have been tempted to say that a pure experience can be taken to belong to more than one mental history.

According to James, pragmatism mediates the so-called conflict between science and religion. James took religious experiences very seriously both from a psychologist's perspective and as evidence for the reality of the divine.

1 Life

William James, brother of the novelist Henry James, Jr, was the oldest child of Henry James, Sr, a man of independent means, who in early manhood experienced a major psychological crisis. Under the influence of the writings of the eighteenth-century Swedish mystic Emanuel Swedenborg, Henry Sr came to see redemption in self-surrender, allowing an in-flowing of divine love that led to a concern with social justice, an interest in Fourierist socialism, and a large number of writings and public lectures. William experienced a similar crisis in 1870, but for him the return to normality was stimulated by reading the French philosopher Charles Renouvier; it consisted not in self-surrender but in affirming his belief in free will and in the resolve to acquire intellectual habits that would lead to daring acts of thought. However, James shared his father's concern for social justice. The effects of this experience are to be seen in many of James' writings, from the long discussion of the will and the emphasis on habit in *The Principles of Psychology* (1890), and the often misunderstood essay 'The Will to Believe', to his widely read *The Varieties of Religious Experience* (1902) and *A Pluralistic Universe* (1909).

After an unorthodox but rich education at home and in Europe, William found it difficult to decide on a career. Eventually he earned an MD, but never practised. In 1873 he taught a course in physiology at Harvard, where he was to continue to teach until his retirement in 1907. In 1875, James set up the first psychological laboratory in America and taught his first seminar in psychology; in 1878 he married and, in the same year, published his first articles in philosophy. James taught both psychology and philosophy, advancing to a full professorship, and in 1890 published the monumental, justly renowned two-volume *Principles of Psychology*; in abridged form, it became a widely used text.

From 1879 on James presented single lectures before a variety of audiences, some of which were published in 1897 in *The Will to Believe, and Other Essays in Popular Philosophy*. In its preface, James announces that he regards all claims concerning matters of fact as hypotheses subject to revision in the light of subsequent experience, and extends this empirical attitude to metaphysical hypotheses. *Talks to Teachers on Psychology and to Students on Some of Life's Ideals* appeared two years later. He concludes the preface to that volume, after condemning American imperialism, 'Religiously and philosophically, our ancient national doctrine of live and let live may prove to have a far deeper meaning than our people now seem to imagine it to possess'.

James retired in January 1907 having just given the Lowell lectures. These were published under the title *Pragmatism: A New Name for Some Old Ways of Thinking* and brought together virtually all of James' philosophical concerns. While the term 'pragmatism' was first introduced by James' friend C.S. Peirce in 1878 and reintroduced by James in a lecture given at Berkeley in 1898, after which it became widely used, it is fair to say that in *Pragmatism* this philosophy receives its full elaboration, at any rate as it is understood and embraced by James (see §5 below). The central idea is James' account of truth; its acceptance, James thought, would bring about a revolution in epistemology. Instead it met with misunderstanding and opposition to which James responded in 1909 in *The Meaning of Truth: A Sequel to 'Pragmatism'*, a collection of articles, the first of which was written as early as 1884 but several of which were composed specifically for that volume.

William James died on the 26 August 1910. There were several posthumous publications, in particular *Essays in Radical Empiricism* (1912) which contains James' most technical philosophy (see §7). While James claims in the preface to *Pragmatism* that the doctrine of radical empiricism is logically independent of that of pragmatism, John J. McDermott is surely correct when he claims that one will misunderstand James' views in the works published in his lifetime unless one takes his radical empiricism into account (introduction to *Essays in Radical Empiricism*).

2 Psychology

Although he taught both subjects, James is one of those to be credited with the separation of psychology from philosophy. In the preface to *The Principles of Psychology* (1890), James attempts to distinguish sharply between metaphysics and psychology. 'I have kept to the point of view of natural science throughout the book. Every natural science accepts certain data uncritically…. Psychology, the science of finite individual minds, assumes as its data (1) *thoughts and feelings*, and (2) *a physical world* in time and space with which they coexist and which (3) *they know*. Of course, these data are themselves discussible; but the discussion of them (as of other elements) is called metaphysics and falls outside the province of this book' (1890: v-vi). Yet James cannot avoid metaphysics entirely and admits as much. Psychology, he holds, accepts the mind-body dualism of common sense but rejects the metaphysical dogma that there can be no mind-body interaction. For if mind and body are radically different, the biological evolution of mind can be explained only on the hypothesis that there have been 'mental atoms' as well as physical ones at the origin of the universe, and that the former must be able to form complex aggregates just as physical atoms do. But in *The Principles of Psychology*, James denies precisely this: 'We cannot mix feelings as such, though we may mix the objects we feel, and from *their* mixture get new feelings' (1890: 157). Moreover, feelings, sensations and thoughts cannot combine themselves; they form new wholes only from the perspective of an outside observer. Confusing one's own standpoint with that of the mental fact observed is, for James, the 'Psychologist's Fallacy' (1890: 196). Of course, James does not deny the existence of 'higher' states of mind, say, understanding a five-word sentence, but what he asserts is that understanding the sentence is a mental fact other than, and in addition to, the five facts of understanding each word separately. But on this James was to change his mind. In his presidential address to the American Psychological Association in December 1894, 'The Knowing of Things Together', he concludes that mental contents can be complex, that we can speak of the parts of a field of consciousness, though such parts are not separable. In the same lecture he also rejects the dualism that

had been a working assumption of *The Principles of Psychology* (see §6). Always, however, the mind is not a mere knower: we know in order to act.

Two chapters of *The Principles of Psychology* are of particular philosophical interest, namely, ‘The Stream of Thought’ (later called ‘stream of consciousness’) and ‘The Consciousness of Self’. In the former, James offers five facts. (1) Every thought (taking thought to include feelings and sensations) appears to belong to some self. (2) No state of mind can recur in exactly the same way. Although we may experience the same object again, the sensation will be different the second time, for the condition of the brain will have been modified by the earlier sensation. (3) Introspection shows that consciousness does not jump from one discrete content to the next, but changes in a continuous manner. Even after interruptions, such as in sleep, we are aware of being the same person, and no sudden change in sensation comes as a total break; it is only a more rapid change. Another element in the continuity of consciousness is what James called ‘fringe’. The fringe is what leads one’s mind from thought to thought, rejecting this one and accepting another until one reaches a conclusion. Later James identified the ‘fringe’ with the subconscious, and that was to play an important role in his explanation of religious experience (see §4). (4) We take it that our thoughts refer to independently existing objects because many thoughts (both of one person and of different persons) appear to have the same object. (5) We pay selective attention to some parts of the field of consciousness, and over the millennia we and our ancestors have in this manner ‘extracted’ the everyday world out of ‘the primordial chaos of sensations’ 1890: 277) (see §5).

While we agree to a remarkable degree in what we find interesting, our own selves are for each of us of unique importance. James distinguishes several senses of self. The material self consists of one’s body and at least some of one’s material possessions. The social self, or rather the social selves, are the various personalities one presents to various others; these, to a large extent, dictate one’s behaviour. One’s spiritual self is not the whole of one’s inner life but that portion of which we say that it *has* the thoughts, feelings, etc. that make up our stream of consciousness. Although James calls this self ‘spiritual’, he does not identify it with a continuous soul substance. Introspection reveals only a stream of thoughts, of which the present one is the judgment that I am the same self I was yesterday, a judgment based on resemblance of present bodily feelings to past ones and on the continuity of this thought with its predecessors. The present thought appropriates (selectively!) past thoughts of the stream; it is the ‘thinker’. It is pointless to appeal to a Kantian pure ego or a Cartesian soul, for no one has explained how such an entity would hold the stream of thought together (see Kant, I. §6; Descartes, R.).

3 Moral philosophy

James’ writings in moral philosophy, though sparse, provide the best entrance into his thinking. These writings fall into two categories: those that address directly some relatively concrete problem and those of a more systematic nature. The former range from letters to newspapers protesting against America’s policy in the Philippines, to essays like ‘The Moral Equivalent of War’, suggesting a struggle against nature as a substitute for the senseless struggle of nations against nations, and ‘The Social Value of the College-Bred’, seeking to allay the ennui of young women who saw no use for the education they had acquired. ‘What Makes a Life Significant’ and ‘On a Certain Blindness of Human Beings’ are essays that speak to a precondition of the moral life - the ability to appreciate what another’s life is like, what makes it worth living or intolerable - and form a bridge to the ethical theory found in the essay ‘The Moral Philosopher and the Moral Life’ (1891) and in various passages in *The Principles of Psychology*.

Leading a moral life requires, according to James, a purpose that gives the life significance, will, courage and determination, a belief in objective values and a belief in free will; here I shall concentrate on the last two items. In ‘The Moral Philosopher and the Moral Life’, James distinguished three questions: (1) where do our moral ideas come from? (2) To what do they refer? (3) How do we weigh various goods and evils, various obligations against one another?

Our valuings have multiple psychological sources: some can be traced back to bodily pains and pleasures, others are the result of habituation or indoctrination, while still others are due to an inborn higher moral sensibility that varies somewhat from person to person. Examples of the last are the demand that like cases be treated alike and the abhorrence with which we respond to the thought of sacrificing an innocent child for the happiness of many. These subjective valuings become objective, that is, coercive, whenever there are creatures who care about one another or acknowledge the claims they make upon each other. Our sympathetic instincts are not only the source of our

higher moral sensibilities, but the basis of objectivity in ethics. Conflicts of claim are, however, unavoidable; how are they to be resolved?

Moral agents encounter conflicts of serious claims in moments of crisis. In *The Principles of Psychology*, James emphasizes the loneliness of such moments and also their momentous nature. What one chooses in such a moment is not so much what one is to do as what sort of person one is to become, and what guides one here is the inner voice of one's higher moral sensibilities. That voice, according to 'The Moral Philosopher and the Moral Life', will tell us to 'seek incessantly, with fear and trembling, so to vote and to act as to bring about the very largest total universe of good which we can see' (1897: 209). Analogously, moral philosophers are to fashion out of the values they find in the world the most inclusive coherent system. And to this task they are to bring no value of their own (this is how the philosopher differs from the ordinary moral agent) other than these 'logical' values of coherence and inclusiveness. But this, surely, is an impossible task. Since there is no common unit to which all goods can be reduced and yet all goods cannot be jointly realized, moral philosophers must appeal to their own moral sensibilities in weighing one ideal against another. The impossible task seemingly given to moral philosophers is perhaps an ironic description of what other philosophers have attempted, rather than a task for us, or James, to accomplish. Ethics, like physics, is for James an empirical science. It rests on claims actually made, and what the most inclusive realizable good will be depends both on that and on the nature of the universe; finally, since we are fallible, we must be prepared at all times to revise our judgments in response to complaints from those whose demands have gone unmet. When James demands that we pursue ideals that can be realized at least cost to other ideals, he does not simply say that in our own lives we must be prepared to sacrifice lesser goods for greater ones, nor that a minimal sympathy and a minimal sense of justice suggest that we are at times to sacrifice our own lesser goods to the greater goods of a larger community. He articulates here his commitment to democracy and liberalism, to tolerance of anything other than intolerance, and his struggle against that blindness in human beings mentioned above.

The moral life consists not only of moments of crisis; for the most part, guided by our aspirations that give it significance, it runs along in quite habitual ways. James provides a series of maxims on how to acquire and keep good habits, for good habits are what make up a moral character. There is nothing so repulsive, he thought, as a sentimental character who never acts forthrightly.

All the above assumes that we can and do choose among alternatives that make a real difference to the world. As a psychologist, James simply records that deliberation usually comes to an end when one has only one idea of one action before one's mind; action follows. But when one's moral impulses conflict with more habitual or instinctive ones, countervailing reasons and motives tend to be still before the mind; one has a sense of effort and we speak of moral victories. What has been said so far is compatible both with the hypothesis of free will and its denial. Indeed, neither logic nor empirical evidence can decide the question, and since for James it seem to have momentous consequences, he opts for free will, for a universe of alternative possible futures. James rejects both a facile optimism that believes nothing in the world needs fixing and a pessimistic resignation that believes nothing can be fixed. Indeed only in an open world are there objectively good and bad actions, and only someone believing in such a world will be able to lead the strenuous moral life. Additional moral energy, James holds, flows from the belief that the help of a God makes the victory of good over evil possible.

4 Religion

The question of the nature of God and of His existence occupied James' thought throughout his life. Although he was not a conventional Christian and quite anticlerical, he was a deeply religious person. In *The Varieties of Religious Experience*, he wrote with regret regarding mystical states, 'my own constitution shuts me out from their enjoyment almost entirely' (1902: 299). Yet twice in his life he suffered the sense of utter worthlessness and despair, indeed panic-fear, that often precedes a religious experience.

James took seriously the question whether, or to what extent, religious experiences provide evidence for the existence of the divine. In reply to positivistic scientists who claim a priori that religious experience is non-veridical and impugn it by pointing to its organic bases, he notes that all experience has an organic basis, and maintains that a religious experience's 'spiritual value' depends on its 'immediate luminousness', on how it fits into the rest of our beliefs and on its 'moral helpfulness' (1902: 33). All our conceptual systems are fashioned by us to deal with experience from the perspective of some interest or sets of interests; we have no reason to think that

only one of them agrees with reality. In particular, both science and religion (better: particular sciences as well as particular religious beliefs and practices) are ‘genuine keys to unlocking the treasure-house of the universe to him who can use either of them practically’, and the use of one does not exclude that of the other (1902: 110).

After a rich and detailed survey of religious personalities and religious experiences, James then turns to the questions of spiritual value and practical moral consequences.

> What I then propose to do is, briefly stated, to test saintliness by common sense, to use human standards to help us decide how far the religious life commends itself as an ideal kind of human activity. If it commends itself, then any theological beliefs that may inspire it, in so far forth will stand accredited. If not, then they will be discredited, and all without reference to anything but human working principles.
>
> (1902: 264)

James’ conclusion is well modulated. In a world in which not everyone is a saint, saintly turning of the other cheek will lead to the victory of evil. Yet both the faith of saints in a kingdom of heaven already here and the utopias of socialists and anarchists ‘help to break the edge of the general reign of hardness and are slow leavens of a better order’ (1902: 285).

There remains the question whether the religious person’s sense of a divine presence is objectively true. Because mystical experiences are intensely private and vary from person to person, and because the arguments of philosophers and theologians are unconvincing, James concludes that any attempt to demonstrate the existence of a deity is hopeless. Philosophy can distinguish, however, between a common religious core and the various ‘over-beliefs’ found both in institutionalized creeds and in the faith of individuals. The core is objectively true and consists in an awareness that one’s conscious self is part of a wider self. That wider self is understood as the source of one’s higher moral ideals and of one’s religious experiences, but may be merely one’s subconscious. However, James’ own over-beliefs include a belief in a God who is not only the source of our highest ideals but can produce real changes in us. He makes the victory of good over evil possible but (here James differs from most religious persons) not certain. This over-belief is a hypothesis and entails predictions - in a world with such a God, miracles will happen, and a different conduct is required than would be in a godless world; for, James speculates, it may well be that our efforts aid God in His tasks as He aids us in ours. Finally, James notes, religious experience requires neither that the deity be infinite nor that there be only one.

5 Pragmatism

James’ major contributions to philosophy are pragmatism and radical empiricism. Pragmatism for James consists in both a theory of meaning, which he ascribes rightly to C.S. Peirce, and a theory of truth, which he ascribes too generously to John Dewey and F.S.C. Schiller (see Truth, pragmatic theories of §2).

James gives in one paragraph two formulations of the Pragmatic Maxim: ‘… to develop a thought’s meaning, we need only determine what conduct it is fitted to produce: that conduct is for us its sole significance’. And, virtually quoting Peirce,

> To attain perfect clearness in our thoughts of an object, then, we need only consider what conceivable effects of a practical kind the object may involve - what sensations we are to expect from it, and what reactions we must prepare. Our conception of these effects, whether immediate or remote, is then for us the whole of our conception of the object, so far as that conception has positive significance at all.
>
> (1907: 26)

This maxim must not be confused with the verifiability criterion of meaning of the logical positivists (see Meaning and verification). Neither Peirce nor James used the maxim to reject metaphysics, although, as noted in §2, James used it to point out that the metaphysical hypothesis of a substantial soul merely reasserts, but fails to explain, one’s sense of being a continuous self. Again, unlike the logical positivists, James includes among sensations all kinds of feelings and emotional reactions; ‘experience’ would have been a less misleading choice of word. Thus, as mentioned in §4, religious hypotheses have meaning for James: believing them makes a difference to one’s moral energy, and some religious experiences are, for those who have them, overwhelmingly evidential. Finally, unlike the logical positivists, James, as discussed in §3, takes it that value judgments, in particular moral judgments, are true or false, not mere expressions of the speaker’s state of mind, nor mere attempts to sway an

audience (see Emotivism; Prescriptivism).

As a form of empiricism, pragmatism emphasizes the fallibility of all knowledge claims, but unlike other empiricisms it does not conceive of the mind as a passive receptacle. Rather, guided by our interests and the concepts we already have, we pay selective attention to our sensory inputs. Since human beings have, through evolution and through upbringing, to a very large extent identical interests, we have come to share a conceptual framework, the framework of common sense. Our common conceptual scheme makes our common lives possible, though we tend to notice our disagreements rather than the larger-scale agreements in which they are embedded. But we could not even disagree, nor would our disagreements matter, if we did not live in a common world, if any one person's beliefs and actions did not have practical consequences for others.

We move, with greater or lesser ease, between conceptual systems: in addition to those of common sense and of religion, there are multiple scientific frameworks. These ways of ordering experiences were developed in response to pressing interests and are modified or rejected if they fail to be useful. Seen in this light, scientific theories are not transcriptions of reality but rather 'instruments' that summarize old facts and lead us to new ones. They are true in so far as the new facts we are led to expect do in fact materialize, or, as James would prefer to say, they *become* true in so far as they become verified. Does he really mean that there are no unverified truths? According to James' student and biographer, Ralph Barton Perry, while he often defends the existence of an unexperienced reality, he fails to explain what this existence consists in ((Perry 1935: 591-2).

James points out that all our knowledge of the world is a product of the world and the human mind. Most of us take the basic concepts of common sense to correspond to reality, while physicists and some philosophers say this rather of the concepts of relativistic quantum physics. But these and any other conceptual frameworks are all equally human contributions to our knowledge. There is no reason to think that one of these will emerge as ultimately triumphant, nor that all of them will prevail as our knowledge grows. Some philosophers find this idea, the idea that the trail of the human serpent is over every knowledge claim, distressing, as if we 'made up' the world. But we only make up the concepts, and whether our beliefs using these concepts are true depends on the world.

Pragmatists accept the traditional view that truth is an agreement between our beliefs or propositions and reality but try to clarify the notion of agreement by asking how agreement is verified. I can see (verify directly) that there is a clock on the wall. But my belief that it tells the time correctly is taken 'on credit'; it will be verified if I act on it and that action does not lead to surprises, for example, a missed appointment. The vast majority of our beliefs are verified even more indirectly. Yet truth is tremendously important. Our existence in the world is precarious; true beliefs lead us to avoid dangers and take advantage of opportunities, while false beliefs may prove fatal. Thus true beliefs are useful and good to believe, but that does not mean that any belief that proves temporarily useful or makes us feel good is true. Not only must any one of our beliefs harmonize with the whole body of beliefs already held - past experience and logic between them leave very little play to the mind - but our new beliefs must prove themselves not merely at the present moment but for the indefinite future. Both common sense and scientific beliefs may have to be modified radically in response to future experiences.

Although James held that the truth is useful to believe, critics who accuse him of identifying truth with usefulness are careless readers. The misreading is, perhaps, encouraged by a hasty reading of his essay 'The Will to Believe' as well as by its unfortunate title. There are certain times when one has to decide to believe 'ahead of the evidence'. We must decide to trust a new acquaintance before we have any basis for trust, else the opportunity of forming a friendship will be lost for ever. What characterizes such situations is that the alternatives before us are equally 'live' for us (we can see ourselves making either choice), that the choice is 'momentous' (much hangs on it; we would not care to toss a coin), and that there is no time to wait for further evidence (waiting is making the choice; in the example, it is to lose the opportunity for friendship). In such cases, and only in such cases, James insists that we have the right to believe. He exercised that right when he chose to believe in free will; as do scientists who commit themselves to investigate as yet untested hypotheses and political reformers who stake their lives on realizing their programmes. It is perhaps worth mentioning that beliefs thus chosen are nevertheless subject to the same verification/falsification processes as are any other beliefs.

6 Radical empiricism

Radical empiricism, found primarily in the posthumously published *Essays in Radical Empiricism* (1912), is James' ontology, his theory of perception and his theory of intentionality. Although James claimed, rightly, that one could be a pragmatist without being a radical empiricist, pragmatism must be able to explain how it is that we live in a world about which we can communicate, which we can change by joint actions, and which any one person's action can change for everyone. The question is urgent because philosophy since Descartes tends to lead to 'a congeries of solipsisms, out of which in strict logic only a God could compose a universe even of discourse' (1912: 37-8). Beliefs in the existence of other minds and of bodies are instinctive, noninferential beliefs, part of the common-sense framework within which we infer what another thinks or what kind of bodies there are. These common sense beliefs do not, however, survive philosophical criticism unaltered. James holds that there is 'one primal stuff or material in the world, a stuff of which everything is composed' but 'there is no *general* stuff of which experience at large is made' 1912: 4, 14). This is not simply a rejection of mind-matter dualism, even talking of a 'stuff' out of which everything is made is misleading; pure experience - the 'primal stuff' - has no essential properties. While the word 'experience' tempts one to understand radical empiricism as a version of Berkeleyan idealism, the view resembles rather a neutral monism, but that it is a label James would have abhorred (see Neutral Monism). In all his writings James emphasized pluralism: a plurality of possible futures, a plurality of irreducible values, a plurality of useful conceptual schemes, a plurality of pure experiences.

Radical empiricism attempts to answer two related questions: how is it that what I take to be my percept of this page of this book is indeed *of* this page (and not of any one of the thousands of exactly similar pages of this book elsewhere), and how is it that my percept and your percept are of the same page when, as we normally say, we are looking at the same page? Analogous questions arise with respect to objects merely thought about. When I see a pen, the experience (call it E) taken in itself, apart from all conceptualization - to the extent that that is possible - is what James calls a 'pure' experience, and the crucial claim made by James is that E is as much an event in the pen's history as it is an event in mine. If we take E with what James calls its 'energetic associates' - with how the pen came to be here and what happens to it hereafter - we have the pen, and if we take E with my prior and subsequent thoughts, feelings, etc. we have my stream of consciousness. This is, first of all, a direct realism. I do not have a sense datum (idea, impression, representation) from which I infer (or 'construct') the pen: I see the pen. And I see its surroundings, I handle it, etc. That is how the mystery of how I can 'mean' this pen rather than any one of the thousands of virtually identical other pens is answered. But who does the 'taking'? It can only be a later item in the stream of consciousness, but just what that item would be is unclear.

To say, as James does, that the context defines the referent when you and I are talking about some absent person presupposes that we live in a common world - that we experience, for example, the same pen as you hand it to me. But how is this possible? No experience of mine is an experience of yours. My-seeing-the-pen is an experience of mine, hence it is not an experience of yours. How then can the-pen-being-seen-by-me be the-pen-being-seen-by-you? James responds that thinking of the pen as my percept (or yours) is a later experience; it is not the 'pure experience' that James calls 'the pen'. *That* pure experience may be appropriated by you as well as by me. As if he realized that this answer might not convince his readers, James adds that 'the decisive reason in favour of our minds meeting in *some* common objects at least is that, unless I knew that supposition, I have no motive for assuming that your mind exists at all…and for me to speak of you is folly' (1912: 38). And, he continues, these objects would be there even if one or more of our minds were destroyed. Indeed, James agrees with common sense that the earth existed before there were any sentient beings and that there will be stars after the last human being has vanished. Once again we are driven to wonder what James can mean by an unexperienced reality. We can easily grasp the idea of a pure experience that does not 'belong' to a physical object (for instance, a hallucination), but what can we make of the idea of a pure experience that is not experienced at all? James himself, perhaps due to this difficulty, toyed with the idea of panpsychism, but this appears to be an unnecessary concession to idealism.

See also: Empiricism; Idealism; Panpsychism; Pragmatism; Pragmatism in ethics; Psychology, theories of §§1-2; Religious experience; Religion and morality

RUTH ANNA PUTNAM

List of works

James, W. (1975-88) *The Works of William James*, Cambridge, MA, and London: Harvard University Press, 17 vols.(This edition includes all previously published as well as many previously unpublished writings by James.

Several of his writings are, however, readily available in paperback. References in the text are to the editions listed below.)

James, W. (1878) 'Remarks on Spencer's Definition of Mind as Correspondence', in *Essays in Philosophy*, Cambridge, MA and London: Harvard University Press, 1978.(The first statement of James' view that the mind is not a passive recipient but actively contributes to what is known.)

James, W. (1890) *The Principles of Psychology*, New York: Dover, 1950, 2 vols.(A monumental work; when it was published and for many years thereafter, it was widely used as a text in psychology courses. Today it is of interest for its many philosophical reflections.)

James, W. (1891, 1897) 'The Moral Philosopher and the Moral Life' and 'The Will to Believe' in *The Will to Believe and Other Essays in Popular Philosophy*, New York: Dover, 1956.(A collection of essays almost all of which were originally delivered as lectures to student audiences. This is one of James' most accessible books, and the second essay mentioned is one of his best known.)

James, W. (1895) 'The Knowing of Things Together', in *Psychological Review* 2; repr. in *Essays in Philosophy*, Cambridge, MA, and London: Harvard University Press, 1987.(This volume contains 21 essays from 1876-1910 covering a wide range of topics.)

James, W. (1898) 'Philosophical Conceptions and Practical Results', Appendix to *Pragmatism*, Cambridge, MA and London: Harvard University Press, 1975, 255.(In this lecture to the Philosophical Union at the University of California, Berkeley, James reintroduced Pragmatism twenty years after Peirce's original essays on the subject.)

James, W. (1899) 'What Makes a Life Significant' and 'On a Certain Blindness in Human Beings', in *Talks to Teachers on Psychology and to Students on Some of Life's Ideals*, Cambridge, MA, and London: Harvard University Press, 1983.(The talks to teachers cover topics discussed in *The Principles of Psychology* that are of particular interest to school teachers, but the two essays mentioned are of enduring human interest and do not presuppose any knowledge of either philosophy or psychology.)

James, W. (1902) *The Varieties of Religious Experience*, New York: Collier Books; repr. London: Collier Macmillan Publishers, 1961.(James' most popular book is a careful account and analysis of the various types of religious experience and religious personality, mostly Christian.)

James, W. (1907) *Pragmatism: A New Name for Some Old Ways of Thinking*, Indianapolis, IN, and Cambridge: Hackett Publishing Company, 1907.(A series of eight lectures for educated laypersons; an exuberant presentation of the essentials of pragmatism as James understood it.)

James, W. (1908) 'The Social Value of the College-Bred', in *Essays, Comments and Reviews*, Cambridge, MA, and London: Harvard University Press, 1987.(An address to a meeting of the Association of American Alumnae at Radcliffe College, the essay was published in McClure's Magazine in 1908. The volume is of interest primarily because it contains all of James' writings on matters of political concern.)

James, W. (1909) *The Meaning of Truth*, Cambridge, MA, and London: Harvard University Press, 1975.(This sequel to *Pragmatism* consists of replies to critics and attempts at clarification. *Pragmatism* and *The Meaning of Truth* were published bound together and with an introduction by A.J. Ayer by Harvard University Press in 1978; this is available in paperback.)

James, W. (1910) 'The Moral Equivalent of War', in *Essays in Religion and Morality*, Cambridge, MA, and London: Harvard University Press.(The essay was published in various journals and magazines in 1910 and reprinted in *Memories and Studies* (ed. Henry James, Jr) in 1912. The volume contains twelve essays on various topics.)

James, W. (1912) *Essays in Radical Empiricism*, Cambridge, MA, and London: Harvard University Press, 1976. (This volume, first published posthumously, contains James' most sustained attempt to deal with epistemology, ontology and the theory of intentionality. It is difficult reading, but intellectually highly rewarding.)

References and further reading

Barzun, J. (1983) *A Stroll with William James*, New York: Harper & Row.(An easy-to-read, loving introduction to James, the man and the thinker.)

Bird, G. (1987) *William James*, London and New York: Routledge & Kegan Paul.(A volume in the *Arguments of the Philosophers* series; an excellent and very readable introduction to James.)

Edie, J.M. (1987) *William James and Phenomenology*, Bloomington, IN, and Indianapolis, IN: Indiana University Press.(An approach to James' philosophy from a phenomenological (Husserlian) perspective.)

James, H. (ed.) (1920) *The Letters of William James*, Boston, MA: Atlantic Monthly Press, 2 vols.(These letters, edited by James' son Henry, provide an opportunity to become acquainted with William James, the man, as well as offering some insights into his philosophizing.)

Levinson, H.S. (1981) *The Religious Investigations of William James*, Chapel Hill, NC: University of North Carolina Press.(A thorough but readable examination of James' inquiries into religious matters.)

Lewis, R.W.B. (1991) *The Jameses: A Family Narrative*, New York: Doubleday.(A thorough and thoroughly enjoyable biography of the family, from William James' grandfather to his children.)

Myers, G.E. (1986) *William James: His Life and Thought*, New Haven, CT, and London: Yale University Press. (A monumental scholarly study of all of James' writings. A worthy successor to Perry (1935).)

Perry, R.B. (1935) *The Thought and Character of William James*, Boston, MA: Little, Brown, 2 vols.(Perry uses hundreds of unpublished letters by James and scores addressed to him as well as many manuscripts (then unpublished) in this masterly account of James' life and thought.)

Perry, R.B. (1938) *In the Spirit of William James*, New Haven, CT: Yale University Press. (A brief, thoughtful, appreciative yet critical exposition and discussion of James' main ideas.)

Putnam, R.A. (ed.) (1997) *The Companion to William James*, Cambridge: Cambridge University Press.(Eighteen informative and provocative articles on various aspects of James' thought provide a thorough but accessible introduction to his thought. An extensive bibliography is provided.)

Skrupskelis, I.K. and Berkeley, E.M. (eds) (1992-4) *The Correspondence of William James, vols 1-3: William and Henry*, Charlottesville, VA, and London: University Press of Virginia.(All the extant letters exchanged by these famous brothers provide a wonderful insight into their personalities and relationship.)

Sprigge, T.L.S. (1993) *James and Bradley: American Truth and British Reality*, Chicago and La Salle, IL: Open Court.(An excellent, quite technical, sympathetic but critical examination of the views of these important philosophers.)

Wilshire, B. (1968) *William James and Phenomenology: A Study of the Principles of Psychology*, Bloomington, IN: Indiana University Press.(A thorough study of James' *The Principles of Psychology* for advanced readers.)

Japanese philosophy

The most distinctive characteristic of Japanese philosophy is how it has assimilated and adapted foreign philosophies to its native worldview. As an isolated island nation, Japan successfully resisted foreign invasion until 1945 and, although it borrowed ideas freely throughout its history, was able to do so without the imposition of a foreign military or colonial presence. Japanese philosophy thus bears the imprint of a variety of foreign traditions, but there is always a distinctively Japanese cultural context. In order to understand the dynamics of Japanese thought, therefore, it is necessary to examine both the influence of various foreign philosophies through Japanese history and the underlying or continuing cultural orientation that set the stage for which ideas would be assimilated and in what way.

The major philosophical traditions to influence Japan from abroad have been Confucianism, Buddhism, neo-Confucianism and Western philosophy. Daoism also had an impact, but more in the areas of alchemy, prognostication and folk medicine than in philosophy. Although these traditions often overlapped, each also had distinctive influences.

In its literary forms, Japanese philosophy began about fourteen centuries ago. Confucian thought entered Japan around the fifth century AD*. Through the centuries the imprint of Confucianism has been most noticeable in the areas of social structure, government organization and ethics. Philosophically speaking, the social self in Japan has its roots mainly in Confucian ideals, blended since the sixteenth century with certain indigenous ideas of loyalty and honour developed within the Japanese* samurai *or warrior class.*

The philosophical impact of Buddhism, introduced around the same time as Confucianism, has been primarily in three areas: psychology, metaphysics and aesthetics. With its emphasis on disciplined contemplation and introspective analysis, Buddhism has helped define the various Japanese senses of the inner, rather than social, self. In metaphysics, Buddhist esotericism has been most dominant; through esoteric Buddhist philosophy, the Japanese gave a rational structure to their indigenous beliefs that spirituality is immanent rather than transcendent, that mind and body (like humanity and nature) are continuous rather than separate, and that expressive power is shared by things as well as human thought or speech. This metaphysical principle of expression has combined with the introspective psychology and emphasis on discipline to form the foundation of the various aesthetic theories that have been so well developed in Japanese history.

Neo-Confucianism became most prominent in Japan in the sixteenth century. Like classical Confucianism, it contributed much to the Japanese understanding of virtue and the nature of the social self. Unlike classical Confucianism in Japan, however, neo-Confucianism also had a metaphysical and epistemological influence. Its emphasis on investigating the principle or configuration of things stimulated the Japanese study of the natural world. This reinforced a tendency initiated with the very limited introduction of Western practical sciences and medicine in the sixteenth century.

Western philosophy, along with Western science and technology, has had its major impact in Japan only since the middle of the nineteenth century. The process of modernization forced Japanese philosophers to reconsider fundamental issues in epistemology, social philosophy and philosophical anthropology. As it has assimilated Asian traditions of thought in the past - absorbing, modifying and incorporating aspects into its culture - so Japan has been consciously assimilating Western thought since the early twentieth century. The process continues today.

What in all this is distinctively Japanese? On the superficial level, it might seem that Japan has drawn eclectically from a variety of traditions without any inherent sense of intellectual direction. A more careful analysis, however, shows that Japanese thinkers have seldom adopted any foreign philosophy without simultaneously adapting it. For example, the Japanese philosophical tradition never fully accepted the emphasis on propriety or the mandate of heaven so characteristic of Chinese Confucianism. It rejected the Buddhist idea that impermanence is a reality to which one must be resigned, and instead made the appreciation of impermanence into an aesthetic. It criticized the neo-Confucian and Western philosophical tendencies toward rationalism and positivism, even while accepting many ideas from those traditions. In short, there has always been a complex selection process at work beneath the apparent absorption of foreign ideas.

Both historically and in the present, some Japanese philosophers and cultural critics have tried to identify this selection process with Shintō, but Shintō itself has also been profoundly shaped by foreign influences. The selection process has shaped Shintō as much as Shintō has shaped it. In any case, we can isolate a few axiological orientations that have seemed to persist or recur throughout the history of Japanese thought. First, there has been a tendency to emphasize immanence over transcendence in defining spirituality. Second, contextual pragmatism has generally won out over attempts to establish universal principles that apply to all situations. Third, reason has often been combined with affect as the basis of knowledge or insight. Fourth, theory is seldom formulated in isolation from a praxis used to learn the theory. Fifth, although textual authority has often been important, it has not been as singular in its focus as in many other cultures. Thus, the Japanese have not typically identified a single text such as the Bible, the Analects, the Qur'an or the Bhagavad Gītā as foundational to their culture. Although there have been exceptions to these general orientations, they do nonetheless help define the broader cultural backdrop against which the drama of Japanese philosophy has been played out through history.

1 Archaic spirituality

The earliest accounts of Japan by Chinese visitors, archaeological remains of the prehistoric culture and the earliest recorded prayers and songs all suggest that Japan was originally an animistic culture with shamanistic qualities. The world was understood to be full of *kami*, sacred presences in the form of awe-inspiring natural objects, personal deities, ghosts and clannish guardian spirits. The ancient rituals were apparently designed for appeasing the *kami* so that humans might live in harmony with them and benefit from their powers. The early poems, recorded in such court-sponsored compilations as the *Man'yōshū*, indicate an internal relationship between humanity and nature. That is, the ancients understood humanity and nature to be parts of each other, not independently existing entities related as subject and object. The ancient myths describe the creation of Japan through the fortuitous actions of the deities. For this reason, the world is infused with *kami* or sacred presence.

According to the myths, natural objects such as rocks and streams were originally endowed with speech, a power taken from them because of their noisy bickering and querulous nature. Although natural things lost their voice, they did not necessarily lose their expressiveness. Human beings, if properly attuned to the natural world, could voice that expressiveness in thoughts, words and artefact. In ancient Japanese, the term for this expressive possibility was *kotodama*, the 'spirit' (*dama*) of 'word' (*koto*) and/or 'thing' (also *koto*).

In short, the ancient Japanese worldview understood the gods, the natural world and humanity to be an ontological continuum. It is not precise to say that rocks and trees were the dwelling place of spirits, because that would establish a bifurcation between the spiritual and material instead of a continuity. The term *kami*, therefore, applied to any object where a sacred presence was particularly manifest or concentrated: the Sun Goddess Amaterasu, Mount Fuji, a special tree or waterfall, the emperor or the vengeful spirit of a fallen warrior. Even the sword of that warrior might be treated as *kami*. This represents one enduring idea in Japanese culture, the emphasis on spiritual immanence instead of transcendence: the sacred permeates the everyday world.

2 The importation of Confucianism and Buddhism

The Chinese writing system was introduced into Japan at about the beginning of the fifth century AD, but it was not until the eighth century that a viable adaptation was devised for rendering Japanese in written form. Therefore early Japanese thought was expressed in Chinese, and in fact many philosophical intellectuals continued to write in Chinese (or a Japanized version of Chinese) as late as the nineteenth century.

When the Chinese writing system was first introduced, the various clans had begun to form a central government under the leadership of what would become the imperial family. The government coalesced in Yamato, a large plain adjoining what is today Kyoto, Nara and Osaka. With the introduction of Chinese literacy, the Japanese elite gained access to more than a millennium of Confucian and Buddhist philosophy. These ideas were immediately put to use in organizing the state.

Confucianism gave Japan a hierarchical model for social and political order. It focused on personal interaction, explaining the responsibilities and duties relevant to the five basic dyadic relations: master-servant, parent-child, husband-wife, elder sibling-younger sibling and friend-friend. When the dyadic relationships are hierarchical, the person in the superior position is to care for the person in the lower and the person in the lower position is to be

loyal to the superior. The imperial family used this system to institute a vertical bureaucracy. Although Confucianism supplied a social structure to the state, the ancient Japanese showed little interest in developing Confucian philosophy *per se* (see Confucian philosophy, Japanese).

Buddhism, on the other hand, was initially most attractive to the Japanese for its aesthetic and thaumaturgic qualities. Buddhist artisans, often immigrants from Korea, brought new techniques of grand architecture, painting, sculpture and music. Using these elegant accoutrements in its rituals for healing, prosperity and protection of the state, Buddhism sometimes competed with the indigenous religious practices addressed to the *kami*.

From the philosophical perspective, however, the most important impact of Buddhism was its psychology. Through its meditative techniques and advanced analyses of the human predicament, Buddhism heightened the Japanese awareness of the workings of heart and mind. Buddhism teaches that egoism is the primary cause of human anguish and dissatisfaction. The ego seeks permanence and control in a world of continuing flux. By controlling the desires and eliminating egoism, one can achieve peace and inner harmony. These Buddhist teachings brought a dimension of inner awareness and psychological analysis to a culture that had formerly operated only within the concepts of taboo, purification and animistic appeasement (see Buddhist philosophy, Japanese).

At the same time, the indigenous religion began to define itself in relation to its rival, Buddhism. It took the name *kami no michi or shintō*, the 'way of the *kami*'. The state helped to systematize a series of myths in the eighth century, explaining the relationships among the tutelary *kami* of the various powerful clans and, presumably, the political relationships among the clans themselves. Most critically, through its familial relation with the chief *kami*, the Sun Goddess Amaterasu, the throne established itself as the blood tie between the celestial *kami* and the Japanese people (see Shintō).

A major goal of philosophy in the seventh and eighth centuries, therefore, was to integrate the available ideas, both foreign and native, into a systematic worldview in the service of political stability. In this light the Seventeen Article Constitution of AD 604 is one of the first philosophical documents of ancient Japan (see Shōtoku Constitution). Attributed to Prince Shōtoku (574-622), the Constitution is really more a set of guidelines for bureaucrats than a set of laws defining political structure. In it, however, we find the early impact of Chinese thinking and its adaptation to the Japanese context of the time.

The Constitution's first article opened with a quotation from Confucius about the importance of maintaining 'harmony'. As noted already, in traditional Confucianism one achieves harmony primarily through performing actions appropriate to one's relationships in the society. Rather than discussing such Confucian principles, however, a large portion of the Constitution discussed human frailty and the need to develop a sympathetic attitude. The Constitution admonished against hypocrisy, preferential treatment, envy and egocentric motives. On the positive side, it advocated consensus and open-mindedness. In short, while the document aimed for a harmonious Confucian social order, it also drew on Buddhist psychology to explain the obstacles to harmony and to suggest an introspective understanding of personal motivations. Although the Constitution itself lacked any detailed philosophical argument, it marks an early attempt to draw on multiple philosophical traditions in a coherent manner. In effect, it advocated a Confucian social and government order (see Confucian philosophy, Japanese) supported by Buddhist practice and the insights of Buddhist psychology. This philosophy of government remained dominant in Japan for at least a millennium.

In the Nara period (710-94), Japanese scholar-monks secured court support to accumulate and study more texts in Buddhist philosophy. They organized themselves loosely around major traditions from the mainland, and became the Six Nara Schools of Buddhism: Ritsu, Kusha, Jōjitsu, Hossō, Sanron and Kegon (see Buddhist philosophy, Japanese). The first primarily concentrated on the study of Vinaya, the precepts and regulations for ordering monastic life. The Kusha and Jōjitsu were schools of Abhidharma Buddhism emphasizing the detailed analysis of dharmas, the basic constituents of reality or consciousness (see Buddhism, Abhidharmika schools of). The Hossō was primarily based in the Indian Yogācāra tradition and Sanron in the Mādhyamika (see Buddhism, Yogācāra school of; Buddhism, Madhyāmika: India and Tibet). Kegon represented the tradition known in China as Huayan. Although the Six Nara Schools played an important role in both education and court politics, there is little evidence that they were philosophically creative centres. Their historical role was mainly to introduce Buddhist analysis and doctrine into Japanese culture. They provided the intellectual raw material for the later philosophical

developments of the Heian period.

3 Metaphysical vision of ancient Japanese esoteric Buddhism

Although already a significant presence in the Nara period, only in the Heian period (794-1185) did Buddhism undergo a profound process of philosophical development and Japanization. Two Buddhist thinkers were particularly influential: Kūkai (774-835; posthumous title, Kōbō Daishi) and Saichō (767-822; posthumous title, Dengyō Daishi). Of the two, Kūkai's philosophical contribution was the more comprehensive. He went to China in 804 to study esoteric Buddhism and, upon his return two years later, founded Japanese Shingon Buddhism. The analytic and systematic character of Kūkai's writings may well qualify him as the first true philosopher in Japanese history.

For Kūkai, reality is fundamentally a person. The entire cosmos is no more than the thoughts, words and deeds of the Buddha called Dainichi (literally the 'Great Sun'). Dainichi is not the creator of the universe; Dainichi is the universe. In a perpetual state of enlightened meditation, Dainichi performs the three great practices of esoteric Buddhism: the chanting of sacred syllables (*mantras*), the visualization of geometrical arrays of symbols (*mandalas*) and the performance of sacred postures or hand gestures (*mudras*). These three activities define the nature of the universe. The *mantras* are microcosmic resonances or vibrating states of matter-energy that constitute the basic elements. The *mandalas* define the essential structure, and the *mudras* constitute the patterns of change. By performing the rituals of *mantra*, *mandala* and *mudra*, a person achieves immediate insight into the nature of the cosmos. By introspection on the nature of their own thoughts, words and deeds, the Shingon Buddhist is said to achieve insight into the thoughts, words and deeds of the Buddha, Dainichi. By understanding one's own person, one understands the person that is all of reality.

Within the framework of this metaphysical system, Kūkai developed a comprehensive philosophy addressing several major philosophical issues. For example, he criticized the idea that insight or enlightenment could be purely mental or intellectual. Because the universe itself consists of thought, word and deed (or structure, resonance and patterned change), it can only be grasped by unified praxis of the whole person: mind, speech and body. To know reality is to participate in it fully, in all three of its dimensions.

Another philosophical issue of interest to Kūkai was the nature of expression. Because the cosmos is a person, reality is the expressive style of that person. Every entity is, therefore, a symbol or imprint of Dainichi's mental, somatic and verbal activity. Yet, because we contain the element of volitional consciousness, we humans can also ignore the true source of all activity, construing ourselves as independent entities. We can interpret things through the superimposition of our own imprints, covering over their (and our) more fundamental nature. This delusion is the source of human anguish and ignorance. The escape from delusion lies in recognizing and participating in the self-expressive nature of reality.

Kūkai's contemporary and major competitor, Saichō, founded Japanese Tendai Buddhism. Although Saichō's primary goal in visiting China was to bring back to Japan the teaching of the Tiantai Buddhist tradition (see Buddhist philosophy, Chinese §7), he also had a chance encounter with a teacher of esotericism. When he returned to Japan, therefore, his Tendai school incorporated esoteric Buddhist elements as well as exoteric Tiantai teachings. Through the exchange of Shingon and Tendai disciples in Japan, some of Kūkai's esoteric teachings also found their way into Tendai. The result was that by the end of the ninth century, the two dominant forms of Japanese Buddhism were both at least partially esoteric in nature.

We can explain the influence of the esoteric Buddhist theory of reality from both the standpoint of cultural history and the history of philosophy. Culturally, it is important that esoteric Buddhism did not either doctrinally or politically oppose the presence of the indigenous religion, Shintō. Esoteric Buddhism reinforced rather than challenged many aspects of the indigenous religious worldview: the ubiquity of the sacred, the expressiveness of nature and the nonduality of matter and spirit, for example. Furthermore, both Kūkai and Saichō correlated the various buddhas of esotericism with the various *kami* of the traditional Japanese religion. By this process, Japanese archaism was, in effect, defended by a philosophically sophisticated system of Buddhist thought imported from the mainland and adapted to the Japanese context. Buddhism and Shintō could therefore be practised alongside each other without contradiction.

This development set the metaphysical backdrop against which later Japanese thought would develop. One might

say that esoteric Buddhism did for Japanese philosophy what Plato and Aristotle did for Western philosophy. It laid out a set of assumptions and a *Problematik* that had a profound influence on the thought to follow. Two assumptions were particularly influential.

First, esotericism has a distinctive view of the relation between part and whole. The whole is recursively manifest or reflected in the part. It is not that the parts constitute the whole nor that the whole is more than the sum of its parts; rather, since the part is what it is by virtue of the whole, if we truly understand the part, we find the whole imprinted in it. In Shingon's case, for example, since any individual thing is an expression of the cosmos as Dainichi, when we truly understand the part (the individual thing), we encounter the whole (Dainichi) as well.

With this orientation as a cultural presupposition, later Japanese philosophers would seldom endorse either atomistic analysis or individualism (see Atomism, ancient; Methodological individualism). Atomistic analysis had been introduced into Japan via the Nara Schools of Jōjitsu and Kusha. Kūkai explicitly ranked them as philosophical 'mindsets' far below Kegon and Tendai, in part because only the latter endorsed this theory of whole-as-part. Individualism, with attendant theories of social contract, entered Japan via the West only in the late nineteenth century. Since it viewed the social whole as constituted by the parts, it ran counter to this esoteric assumption. Not surprisingly, individualism has never taken hold in Japan as a basis for social, ethical or political theory.

Second, esotericism's metaphysics argued that reality is self-expressive. Human beings are, of course, part of reality. When humans speak authentically or truly, therefore, they do not refer to reality, but rather are part of its self-expression. This position undermines any philosophical tendencies toward idealism (reality as a production of mind), realism (reality as pre-existing our expressions and truth as matching our expressions with that reality) or radical nominalism (expressions refer primarily to other expressions without necessary connection to non-linguistic reality) (see Idealism; Realism and antirealism; Nominalism). In the Nara Schools there were variants of idealism (Hossō), realism (Kusha and Jōjitsu) and nominalism (Sanron). Again Kūkai ranked them all below Kegon and Tendai, and certainly below what he believed to be the only comprehensive mind-set, esotericism.

This presupposition about the self-expressive nature of reality underlies most Japanese aesthetic theories as well. The Heian period was the first to develop a detailed set of aesthetic terms, for example, *miyabi* (cultured, refined elegance) and *mono no aware* (the 'ah-ness of things' or the aesthetic tinge of sadness arising from an appreciation of the evanescent). In later periods, other aesthetic terms such as *yūgen* (hidden sublimity or depth), *sabi* (the 'loneliness' of an elegance allowed to age) and *wabi* (rusticity) came to the fore. These terms have rich connotations not readily translated into English. In general, though, they signify equally a quality of the object and the response of the aesthete or artist. The aesthetic is understood to be a self-expressive resonance between the object and artist. The world expresses itself through the artist as the work of art (see Aesthetics, Japanese).

4 Medieval philosophical anthropology: Pure Land Buddhism

During the Heian period the highly literate and elegant culture of the Kyoto court was at its peak. The aesthetically pleasing rituals of esoteric Buddhism found a receptive audience among the aristocrats and clergy. They alone enjoyed the leisure time, education and resources to devote their lives to its study and practice. The general populace were left to their folk religions, an amalgamation of practices with roots in Buddhism, Shintō and Daoist alchemy.

By the early twelfth century, the court had become so politically effete that the provincial aristocracy began to vie for power and the newly risen *samurai* fought for control of territory. Plagues, famines and earthquakes were also unusually devastating. In short, to any sensitive observer of the times, it was easy to see the decadence of the social order, the harshness of nature and the corruption of the human spirit. There was little time for, or consolation in, metaphysical speculations. The philosophers turned their analyses to this world and their imaginations to wondering what failing in humanity had caused such suffering. By 1185 the Minamoto clan was victorious and in 1192 Yoritomo became *shōgun*, establishing his centre of government in Kamakura. Thus began the Kamakura period (1185-1333) and a new set of philosophical and religious orientations.

The Kamakura period philosophers such as Hōnen (1133-1212), Shinran (1173-1262), Yōsai or Eisai (1141-1215), Dōgen (1200-53) and Nichiren (1222-82) responded to the decay and suffering of their times. Each developed his

own interpretation of the human predicament with an accompanying solution. All had originally trained as Tendai monks, but these reformers eventually left the establishment and founded new Buddhist sects that served the masses as well as the elite echelons of Japanese society. Of the Kamakura schools of Buddhism, Pure Land and Zen have been the most influential in their theories of human being. Shinran was the founder of Shin Buddhism or the 'True Pure Land School; Dōgen was the founder of Sōtō Zen. Each developed most fully the philosophical foundations for his own tradition (see Buddhist philosophy, Japanese).

These two schools differ radically in their philosophical anthropologies. As Buddhists, both Shinran and Dōgen accepted the general Buddhist analysis about the source of ignorance: egoism. Egoism defines the self as an independent agent that initiates actions and has experiences. The Buddhist view, on the other hand, maintains that 'I' is no more than a name for related actions and events, not something that lies behind them. This implies that the boundaries of the self are fuzzy rather than sharply delineated. For example, from a distance we can readily identify the general course of a river, but if we move up close enough, we cannot specify exactly where the river ends and the river bank begins. If the river were self-conscious and tried to specify for itself 'my' boundaries as opposed to 'its' boundaries, the river would lose sight of the very processes that bring it into being and help define it. Analogously, Buddhists generally maintain that egoism, by attempting to define, delineate and protect the self, ignores the self's broader context. It overlooks the self's dependence on what egoism considers to be outside and separate from it.

As a disciple of Hōnen, the founder of the Japanese Pure Land School, Shinran also accepted the specific ideas of that tradition, including the theory that humanity had entered a degenerate period in history. Left to our own devices, we are presumed to be doomed to live and relive an existence of anguish, dissatisfaction and despair. Pure Land Buddhists believe, however, that a celestial buddha called Amida has seen our situation and taken pity on us. He has vowed that if we entrust ourselves totally to his compassion and call on his name, we can be assured rebirth in his Pure Land where the conditions are conducive to Buddhist practice. In his Pure Land we can perform the necessary spiritual disciplines and then return to this world to attain enlightenment. In so doing, we can be a spiritual aid to others in this world. Such was the basic tradition of Pure Land Buddhism that Shinran accepted. Within that traditional framework, however, he developed his own distinctive philosophical analysis of human being.

Shinran believed he had found a contradiction in how most Buddhists understood practice. Specifically, they self-consciously undertook various disciplines (meditation, reading texts, chanting mantras) as a means of eliminating egoism and attaining insight into reality. The implication is that one can overcome egoism by one's own power, by 'self-power.' People believed they were 'earning merit' by their religious practices. Shinran argued that if people practice as a means of achieving merit, then the actions only feed, rather than eradicate egoism: discipline is understood as a means by which *I* can improve *myself*. Shinran considered this emphasis on self-power to be the psychological foundation of the degenerate age in which he believed he lived. Because people misunderstood the relation between self and Buddhist practice, the teachings of Buddhism had indeed become unintelligible and the prospect for insight had disappeared.

Although Shinran did not explicitly deny the ontological reality of the Pure Land or the account of Amida's vow as historical, his philosophy focused more on the psychological and logical implications of the Pure Land position. For him the fundamental point was that people must surrender their egoistic senses of self by adopting an 'entrusting heart-and-mind' (*shinjin*). By completely abandoning the ego, people can have faith in the power of Amida's vow to help. They turn from 'self-power' to 'other-power.' Assured birth in the Pure Land, the realm devoid of egoism, the sense of a discrete or independent self will disappear. Yet, Shinran reasoned, if there is no self at that point, logically there can be no discrete 'other' either. There is just 'naturalness' and Amida in effect disappears as well. *Shinjin* continues as the entrusting that opens itself to this naturalness. In that egoless state, one can spiritually help other people. It can be said that one returns from the Pure Land to be reborn in this world. Yet, by assisting others in this world, the ego may once again be constituted as the 'I' who self-consciously helps 'others.' As soon as one begins to think that good deeds are done by virtue of one's own power, the whole process must be renewed. One must again see the delusion of the ego, again turn oneself over to Amida's power, again be reborn in the Pure Land, again allow self and Amida to disappear, and again return to this world from the Pure Land.

In short, Shinran agreed that to eliminate ignorance one must eliminate egoism. This can be accomplished, he believed, only by the complete renunciation of the notion that one can help oneself. Only by despairing of the efficacy of self-power and by entrusting oneself to Amida's power can one become naturally what one truly is (see Self-realization). In that egoless state one can understand reality for what it is and act freely in the world as a compassionate being. In addressing the same issue, Zen Buddhism took an almost diametrically opposed approach, however.

5 Medieval philosophical anthropology: Zen Buddhism

Like the Pure Land Buddhists, Zen Master Dōgen believed there is a fundamental flaw in the usual interpretation of Buddhist practice as the means to enlightenment. Rather than arguing that practice must be abandoned, however, he instead maintained that practice or self-discipline is an end in itself. He rejected the popular theory that his was a degenerate age in which enlightenment was no longer possible. Instead, Dōgen maintained everyone is already enlightened, but that enlightenment was not being manifested or expressed in their actions. The goal, therefore, is to authenticate what we already are.

Trained in the Zen (Chan) Buddhist tradition in China, Dōgen was sensitive to the limitations of language and mistrustful of certain types of thinking. Like other Buddhists, he understood the problem to be egoism. By hypostatizing the ego, one falls into a desire for reality to be a specific way. One seeks permanence in both self and in one's own worldview. Therefore, it is easy to project interpretations on experience, interpretations that shape the experience to meet our presuppositions, expectations and desires.

Dōgen believed that experiential immediacy is possible. In Zen meditation, one quiets the mind and merely lets phenomena appear. Dōgen called this a state of 'without-thinking' as opposed to either 'thinking' or 'not-thinking'. Thinking, for Dōgen, included any form of sustained conceptualization whether fantasy, cogitation, believing, denying, wishing, desiring or whatever. Not-thinking is the effort to blank the mind and empty it of all awareness. In without-thinking, however, there is the awareness of brute phenomena but no sustained act of bestowing meaning. There is no consciousness of a self having an experience. Furthermore, since no meaning at all is projected on the event, it is free of the distortions found in ordinary, ego-driven forms of experience. Dōgen simply called this 'the presencing of things as they are'.

Dōgen claimed this form of meditation was not a means to enlightenment. Instead, precisely because it is egoless, it is enlightenment itself. Yet, this meditation-enlightenment event is always accessible. It is, as it were, at the root of all experience, even thinking and not-thinking. In this respect we are all already enlightened but we have not authenticated that fact. Of course, to authenticate it, we must return to a state of without-thinking.

Without-thinking is, therefore, a not-yet-conceptualized immediacy. Since it is without concepts, however, it is intrinsically meaningless; but it is impossible for humans to live a life without meanings. Enlightenment must not only be authenticated in meditation. It must also be expressed in everyday life. How can this be possible?

Dōgen claimed that meaning is always contextual. He noted that the ocean has a different meaning to a fish swimming in it, a person in a boat out at sea and a deity looking down at it from the heavens. To the fish, the ocean is a translucent palace; to the person it is a great circle extending to the horizon in all directions; to the deity it is a string of jewel-like lights glittering in the sunshine. If the deity were to interpret the ocean as a circle or the person in the boat were to interpret it as a palace, however, they would not be expressing what is actually in front of them. Their interpretation would be false. Therefore, the key to truth in meaning is the appropriateness of the context.

According to Dōgen, meditative without-thinking is the 'touchstone' for determining whether the context is appropriate. Context is continually shifting and giving rise to new meanings as we live out our lives. What is 'tree branch' at one point may be 'firewood' at another and 'weapon' in a third. Dōgen referred to these points as 'occasions' of 'being-time'. For Dōgen, the problem with egoism is its resistance to accepting flux. Egoism tries to make a set of previous meanings into a fixed worldview, into *my* reality. Therefore, one projects contexts that are not actually present in the current phenomena. Through meditation, however, one can break the closed cycle of self-verifying projections. One can return to the presence of things as they are before meaning, before they are embedded in any particular context. Then, as one returns to the expressive world of everyday action, it is easier to verify the appropriateness of the contextualizing process that generates meaning (see Meaning and verification).

The philosophical anthropologies of Pure Land and Zen address the common problem of egoism, but their solutions to the problem are fundamentally different. In Pure Land Buddhism, recognizing the inefficacy of egoism leads to a psychodynamic of despair, entrusting and naturalness. In the Pure Land philosophy of human being, the ego is rejected in favour of a model of dependence or interdependence. It is a process of self-effacement and surrender. By contrast, Zen overcomes the negative effects of egoism not by self-effacement, but by self-analysis. One studies the dynamics of consciousness and grounds oneself in pure, but meaningless, presence. Pure Land and Zen agree that self-discipline is not a means to enlightenment. Yet, for Pure Land this entails the rejection of self-discipline; in Zen, on the other hand, it entails the acceptance of self-discipline as an end in itself.

These two philosophical anthropologies exemplify how Kamakura philosophers generally focused on the nature of praxis as part of an analysis of human existence. In the Heian period, the pressing philosophical question seems to have been the nature of the cosmos. In the Kamakura period it shifted to the nature of the self. Shinran and Dōgen produced particularly impressive analyses of human motivation and the structure of consciousness, and their models of the self remain influential in Japanese culture up to the present. It is significant, however, that their focus was primarily on the inner self instead of the social self. The social dimension was to become a major concern in the next major period in the development of Japanese philosophy.

6 Neo-Confucianism, the *samurai* code and Tokugawa society

Following the Kamakura period there were more periods of intermittent warfare and internal strife. A long lasting, nationwide peace arrived only with the establishment of the Tokugawa family's regime as *shogun*s. For nearly all of the Tokugawa period (1600-1868) Japan closed itself off from most interaction with outside world. For example, Christianity, which had been introduced by missionaries in the sixteenth century, was formally proscribed. The Tokugawa *shogun*s established a highly bureaucratic government, giving them unprecedented control over Japanese society from its system of education to its business practices and religious institutions. In this context, much of philosophy turned to the interests of the state and the definition of social responsibility.

During the fifteenth and sixteenth centuries, Japan had again received a strong infusion of foreign thought. In particular, Zen Buddhist monks who had visited the mainland brought back to Japan texts of the neo-Confucian traditions established in China by Zhu Xi and Wang Yangming. Since Japanese Buddhist philosophy taught little about social responsibility in secular contexts, the ethical dimension of these texts attracted increasing attention. Neo-Confucianism went further than traditional Confucianism by adding a metaphysical level to explain the natural world and how it could be known (see Neo-Confucian philosophy).

From ancient times Confucianism had played a major role in the social, bureaucratic and ethical structures of Japanese culture. In trying to organize and stabilize the government after centuries of warfare, the Tokugawa *shogun*s were naturally intrigued by this new and more comprehensive form of the social philosophy that had already served Japan in the past. Furthermore, wary of Buddhism's popularity, they probably welcomed neo-Confucianism's challenge to the near hegemony that Buddhism had established in Japanese philosophy. In any case, from early in the seventeenth century, the Tokugawa *shogun*s gave special status and support to neo-Confucianism, especially to the school called Shushigaku, the Japanese school of Zhu Xi.

With the increased peace and prosperity of the Tokugawa period, there was a new market for philosophical education, especially in the great urban centres of Kyoto, Osaka and Edo (now Tokyo). The rising merchant class wanted the social polish of an upper class education. Furthermore, because it was peacetime, many unemployed *samurai* wanted a classical education to qualify for positions in the government bureaucracy. The result was an increase in independent schools and a proliferation of teachers with different philosophical approaches.

Generally we find major philosophical development during the Tokugawa period in two areas: naturalistic metaphysics and social philosophy or ethics. The Shushigaku school introduced a theory of reality or metaphysics foreign to the Buddhist theories so entrenched in Japanese thought. In particular, it analysed reality in terms of the dynamic between 'configuration' or 'principle' (*ri*; in Chinese, *li*) and 'material energy' or 'vital force' (*ki*; in Chinese, *qi*) (see Li; Qi). According to Shushigaku, *ri* gives the universe its structure and, since *ri* is also in the mind, it is the foundation of knowledge. By 'investigating the nature of things' we come to know *ri*, both in ourselves and in the things we study. *Ki*, on the other hand, was considered the basic stuff that is ordered by *ri*.

Although the notion of *ri* was known to the Japanese through Tendai and Kegon Buddhism, the neo-Confucians

gave the term a distinctive emphasis. They embedded it into the broader enterprise of understanding of the natural world. During the Tokugawa period there was a practical interest in better understanding nature; in the sixteenth-century traders, and missionaries from Europe had introduced some Western science. With the closure of Japan this contact was severely limited, although the occasional Dutch treatise on practical science or medicine did find its way into Japan.

For the most part, Japanese philosophers found the Shushigaku emphasis on *ri* to be overly abstract. To many, it seemed that *ri* was an unnecessary transcendent realm behind physical reality that could be known only through some mysterious half-contemplative, half-empirical study. In response, many Japanese thinkers took a more phenomenalistic approach. Kaibara Ekken (1630-1714), for example, argued for the primacy of *ki*. To him, *ki* was the basic constituent of reality and should be studied directly; *ri* was no more than the name for the patterns one could abstract from the behaviour of *ki*. Certainly from the perspectives of both medicine and the martial arts, *ki* became the more important category in Japan.

Other naturalistic philosophers such as Miura Baien (1723-89) developed intricate systems of their own for categorizing natural phenomena. Such indigenous concerns for observation and classification of the natural world may not have developed into a full-blown science in the Western sense, but the orientation did show an increasing Japanese concern for observing and understanding the natural world. This phenomenalist tendency would serve Japan well in the nineteenth century when Western science and technology were reintroduced.

In the field of social or moral philosophy, an important development was the emergence of the school of ‘Ancient Learning’. Led by Yamaga Sōkō (1622-85), Itō Jinsai (1627-1705) and Ogyū Sorai (1666-1728), these philosophers rejected the metaphysical speculations of the neo-Confucianists and tried to return to the early classics of the Confucian tradition, especially the *Analects* (see Chinese philosophy; Confucian philosophy, Chinese; Confucius). They developed sophisticated philological and exegetical skills as tools for attempting to discover the original meanings of those texts. Their goal was to clarify traditional Confucian social philosophy so that it could become the basis for Japanese society. In this regard, the school of Ancient Learning put its emphasis on the nature of virtue and the development of character. Ogyū Sorai had an especially broad impact on society for his theories about education and moral training.

Some philosophers, like Yamaga Sōkō of the Ancient Learning School, mixed Confucian values with warrior values about loyalty and honour. Yamaga tried to develop a warrior mentality for service to the state that would be appropriate to peace time. Furthermore, in their unemployment, many *samurai* entered the various Buddhist orders, especially Zen, where they found a familiar emphasis on discipline and regimentation. The combined result was an idealized code of the warrior (*bushidō*) as a way of life, even for non-*samurai* and even in times of peace (see Bushi philosophy).

7 Native studies: religio-aesthetic foundation of the Shintō state

The school of Ancient Learning’s return to the original classics of Chinese Confucianism was mirrored in a movement to return to the early texts of Japan, the school of ‘Native Studies’ or ‘National Studies’ (*kokugaku*). Originally a literary and philological group, Native Studies scholars like Keichū (1640-1701), Kada no Azumamaro (1669-1736) and Kamo no Mabuchi (1697-1769) analysed the language and worldview of Nara and Heian poetry and prose. The school expanded beyond these literary goals, however, as it turned to questions of religion and national identity. In this development the philosophy of Motoori Norinaga (1730-1801) played a pivotal role.

The major shift in Native Studies began with Motoori’s decision as a philologist to decode the antiquated writing system of the *Kojiki (The Record of Ancient Matters)*. An eighth-century text, the *Kojiki* was supposedly the written version of what had formerly been an oral tradition. Mixing myth and history, it discussed the origins of the world, the formation of Japan and the succession of Japanese emperors from the beginning of time up to its present. There was a twin work in Chinese written at the same time, *Nihon shoki (The Chronicle of Japan)*. Because the orthography of the *Kojiki* died out as a writing system for Japanese shortly after the Nara period, the *Nihon shoki* became the more used text, supposedly containing the same information. The *Kojiki* had in fact become virtually unintelligible even to the educated Japanese of Motoori’s time. By decoding the text, therefore, Motoori hoped to bring to light the original Japanese worldview.

A devout follower of Shintō, Motoori's task assumed a profoundly religious dimension as well. He believed the *Kojiki* was a written account of what had been orally transmitted word for word from the time of creation. The *Kojiki* contained the very words of the deities who had created the world. Furthermore, since the text was written in an orthography that had soon fallen into disuse, Motoori believed the written text was uncorrupted by later interpreters, making it superior to the adulterated cosmogonies of other cultures. This firm belief sustained Motoori's lifelong devotion to the enormously complex task of decoding the text.

Based on his readings of Japanese classics, Motoori also developed a philosophy of poetic or religious expression: his theory of 'heart' (*kokoro*). As a technical term, *kokoro* designates the seat of thinking and feeling; it is the basis of sensitivity (see Kokoro). Heart is not, however, limited to people; things and words also have heart. If poets have 'genuine heart,' they will be touched by the heart of things and the heart of words. The poetically or religiously expressive act, therefore, is an act of the heart, something shared by the things, the person and the words. Since the genuine heart is also a goal of Shintō purification, Motoori saw in this theory the basis of religious language as well. This in turn influenced his understanding of the significance of the *Kojiki*.

If the *Kojiki* represents the original words of the deities at the time of creation, to read or study the text is virtually a ritualistic re-enactment of creation. The implication for Motoori was that the ancient Japanese language of the text is not only the language of the deities, but also the most pure intimation of the heart of things. By this line of reasoning, Motoori made the *Kojiki* into the sacred scripture of Shintō. Based on his reading of the text, Motoori founded a philosophy of Shintō supposedly free of Buddhist influence. For virtually the first time, Shintō could develop a formal doctrinal system of its own (see Shintō).

This line of thought readily supported a nationalist ideology. If the ancient Japanese language was the protolanguage of all languages, if that language were most purely resonant with the heart of things and if the Japanese emperor was the special link between the deities and humanity, obviously Japan would have a special place in the world. This sense of national superiority became especially strong in the next generation of Native Studies scholars such as Hirata Atsutane (1776-1843) and helped contribute to the movement to overthrow the Tokugawa shogunate and restore the emperor as the true leader of Japan. In this effort, it found an intellectual ally in the Mito School, a Shintō-Confucian synthesis, that argued for the centrality of the emperor as the 'body of the state'. The restoration of the emperor was completed in 1868, the beginning of Japan's modern era.

In summary, the introduction of neo-Confucianism from China and the establishment of a stable state under Tokugawa rule created a new fertile environment for Japanese philosophy. Ultimately, neo-Confucianism itself did not become a dominant philosophy in Japan, but its presence challenged the dominant Buddhist philosophies. In this new context Japanese philosophy grew more concerned with social ethics, the study of the natural world and cultural identity. In this period developed the idea of the warrior-turned-bureaucrat, as fiercely loyal to the organization as formerly to the lord. As there was more interest in studying and classifying the physical processes of nature, there was also a newly defined aesthetic of sensitivity and poetic expression (see Aesthetics, Japanese). For the first time, Shintō became a major intellectual force in Japanese thought and there was an attendant sense of the uniqueness and superior quality of Japanese culture. All these factors became the intellectual background for the emergence of the modern era after 1868, and its related philosophies.

8 Modern Japanese philosophy and its critique of Western philosophy

The Tokugawa policy of seclusion ended with the appearance of US gunboats in 1854 and their demand that Japan open itself to international trade. To protect its sovereignty from infringement by the Western powers, Japan believed it had to become a modern industrial and military power in its own right. The government sent its brightest young intellectuals to Europe and the United States to study what was needed for modernization, such as medicine, engineering, agriculture, postal systems and education. This effort included the study of Western thought as a means to understanding Western society and the ideas behind its science and technology. Although there was some sustained interest in American pragmatism (see Pragmatism), most Japanese philosophers turned toward Germany for their inspiration (see German idealism).

Throughout the nineteenth century, most Japanese leaders hoped Japan could superimpose Western science and technology on a society that remained true to Asian cultural values. By the first decade of the twentieth century, Japan had successfully developed the technology and military might needed to defeat both Russia and China in

wars. Many Japanese intellectuals feared, however, that this was at the expense of traditional values. The ideal of detached objectivity in Western science threatened the tradition of apprenticed learning through imitation of the master. The Buddhist and Confucian theories of reality were in jeopardy of being overwhelmed by Western scientism. The new egalitarian ideals of education so helpful in developing a technological society were also part of a democratic worldview that emphasized the individual as basic unit of society and threatened the traditional Confucian virtues and social hierarchy. How to negotiate the differences between traditional Asian and modern Western values became a major concern among Japanese philosophers in the first half of the twentieth century.

The most influential development in modern Japanese philosophy was the emergence of the Kyoto School of thought (see Kyoto School). By the early twentieth century, philosophy had become an academic study in Japanese universities. An influential circle of philosophers clustered around Nishida Kitarō (1870-1945), a professor at Kyoto University. This group tended to address problems about the meaning of self, the nature of knowledge, the role of spirituality and the place of both ethical and aesthetic value.

Nishida was the single most influential philosopher in the prewar period. His philosophical goal was to locate empiricism and scientific thinking within a larger system that would also give value judgements a non-subordinate place. *Zen no kenkyū (An Inquiry into the Good)*, his first major work, developed the notion of ‘pure experience’, an idea adapted from William James and perhaps developed in light of Nishida’s own Zen Buddhist practice. The book’s theme is that there is a thrust toward unity in all experience. Thought arises out of the disruption of the unity of immediacy and serves as a means to establishing a more comprehensive unity. In Nishida’s phrase, pure experience is the ‘alpha and omega of thought’.

Nishida himself subsequently decided this early effort was too ‘psychologistic’ and ‘mystical’, and developed a different philosophical system in the 1920s and 1930s that emphasized the ‘logic of place (or *topos*)’. According to Nishida, every judgement is restricted by the logic of its context, which in turn derives from a broader experiential domain that it cannot explain in its own terms (see Logic in Japan). An empirical judgement, for example, excludes the subject of the experience (see Empiricism). Its internal logic precludes the consideration of the self. Yet, of course, there can be no empirical data without an experiencing subject; so, the logical place within which empirical judgements are made is within a broader experiential context that assumes the function of the self. If that broader context is then made the logical domain for judgements, we have idealism. In turn, according to Nishida, the experiential locus that makes idealist judgements possible cannot be spoken of logically within the domain of idealism. Nishida calls this experiential locus ‘place of absolute nothingness’, the ground of ‘acting-intuiting’. This region cannot be expressed in any logical form, but is the basis of all logical expression. It is also the ground of value: spiritual, ethical and aesthetic.

In this way, Nishida argued that the realm of empirical judgement is necessarily grounded experientially in a realm of value that it cannot analyse from its own standpoint. Nishida’s system attempted to grant Western science its logical place while showing that its experiential ground was what traditional Asian values had affirmed all along. Religion, at least in its Asian forms, was not antagonistic to science, nor was it endangered by science. On the contrary, Nishida argued that spiritual experience is what makes science logically possible.

Nishida argued for the synthesis of Eastern values and Western values by analysing the logic of epistemology. He was joined in this logical or epistemological approach by other philosophers connected in some way with the Kyoto School such as Tanabe Hajime (1885-1962) and Nishitani Keiji (1900-90). Other Kyoto School philosophers addressed the issue from the other direction, by analysing values. Hisamatsu Shin’ichi (1889-1980), for example, analysed the religio-aesthetic worldview of Zen Buddhism, advocating it as the basis for a style of life. Miki Kiyoshi (1897-1945), on the other hand, developed a ‘logic of creativity’ inspired by both Buddhist and Marxist theories of praxis.

Among the modern Japanese ethicists, the most influential was Watsuji Tetsurō (1889-1960), a professor at Tokyo University and not technically a member of the Kyoto School. Watsuji explained that Western ethics takes the individual for its prime locus. Western ethics is constituted *vis-à-vis* individual needs and the focus of morality is the individual agent. In contrast, Watsuji said, Confucian ethics takes the social as its prime locus, being constituted out of the primary social relations. Watsuji maintained both traditions are faulty in seeing only one dimension of the whole. As an alternative, he developed a philosophical anthropology emphasising ‘betweenness’, a dialectical tension between the individual and the collective. The collective establishes norms within which one

can act in a given society, whereas the individual serves as the locus of freedom. If unqualified by the opposing pole, the collective suppresses freedom and the individual rejects the objective validity of norms.

Watsuji concluded, therefore, that true ethical behaviour is possible only as a ‘double negation’ that rejects both poles without settling in either. In this ‘betweenness’ we find the dialectical tension between the social and the individual, the fundamental definition of our human being. So, the nature of ethics follows from the definition of human being, and ethics is the fundamental way of realizing our humanity. Watsuji developed these ideas first in his *Ningen no gaku to shite no rinrigaku (Ethics as the Study of Human Being)* (1934) and then more fully in his *magnum opus*, a three-volume work called simply *Rinrigaku (Ethics)* (1937-49).

9 Postwar developments

The defeat in the Second World War caused many philosophers to rethink their positions. Tanabe and Watsuji, for example, explicitly repented some of the nationalistic implications of their earlier writings. In his *Zangedō to shite no tetsugaku (Philosophy as Metanoia)* in 1946, Tanabe developed an intricate, self-critical dialectical method to check the emergence of philosophical ideologies. In conceiving this method he drew inspiration from the philosophical analysis of Shinran’s thought done by one of his students, Takeuchi Yoshinori. The connection with religious philosophy has become a recurrent theme in the further development of the Kyoto school as exemplified in the works of Nishitani, Takeuchi and Ueda Shizuteru. In a spirit reminiscent of the Kamakura Buddhist thinkers, many of the postwar philosophers have turned inward to re-examine the nature of human existence, now able to be formulated in relation to the problematics of existentialism as well as Buddhism.

At the same time some Japanese philosophers have continued to specialize in the scholarly study of Western philosophy. In many Japanese universities there are departments of philosophy where much of the work is indistinguishable from what might find in a philosophy department in Europe or North America. In general, the Continental traditions of philosophy, rather than the British analytic traditions, continue to dominate.

Lastly, and especially since the 1960s, there are individuals and groups of philosophers who have explored new provocative directions, drawing their ideas from a wide variety of sources including Western science, psychoanalysis and phenomenology as well as traditional Asian thought and medicine. This phenomenon is another example of a recurrent pattern in the history of Japanese philosophy: the assimilation and adaptation of foreign ideas against the background of an ongoing tradition.

See also: Aesthetics, Japanese; Buddhist philosophy, Chinese; Buddhist philosophy, Japanese; Bushi philosophy; Confucian philosophy, Chinese; Confucian philosophy, Japanese; Literature, philosophy in modern Japanese; Kokoro; Kyoto School; Logic in Japan; Shintō

THOMAS P. KASULIS

References and further reading

Maruyama Masao (1952) *Nihon seiji shisō shi kenkyū*, trans. Mikiso Hane, *Studies in the Intellectual History of Tokugawa Japan*, Tokyo: University of Tokyo Press, 1974. (English translation of Maruyama’s classic study of the relation between Tokugawa thought and politics, including good summaries of the basic philosophical positions.)

Matsunaga Daigan and Matsunaga, A. (1974-6) *Foundation of Japanese Buddhism*, Los Angeles, CA: Buddhist Books International, 2 vols.(A technical, detailed exposition of the Buddhist philosophical systems of Nara, Heian and Kamakura Japan.)

Moore, C.A. (ed.) (1967) *The Japanese Mind: Essentials of Japanese Philosophy and Culture*, Honolulu, HI: University of Hawaii Press.(A provocative, but somewhat uneven, collection of fifteen essays by Japanese philosophers who participated in East-West Philosophers’ Conferences during 1939-64.)

Nakamura Hajime (1969) *A History of the Development of Japanese Thought*, Tokyo: Kokusai Bunka Shinkokai (Japan Cultural Society), 2 vols, 2nd edn.(Not an integrated book but a set of seven excellent essays on different periods of Japanese philosophy. Out-of-print, but still one of the most insightful discussions available in English.)

Nishida Kitarō (1911) *Zen no kenkyū*, trans. Masao Abe and C. Ives, *An Inquiry into the Good*, New Haven, CN: Yale University Press, 1990.(Translation of Nishida’s first major work; probably the most famous philosophical work of modern Japan.)

Piovesana, G.K. (1969) *Contemporary Japanese Philosophical Thought*, New York: St John’s University Press.

(Good survey of major Japanese philosophers 1862-1962. Includes a good, but dated, bibliography.)

Tanabe Hajime (1946) *Zangedō to shite no tetsugaku*, trans. Takeuchi Yoshinori *et al*., *Philosophy as Metanoetics*, Berkeley, CA: University of California Press, 1987. (Translation of Tanabe's dense but profound reflection on the nature of philosophy.)

Tsunoda, R., de Bary, W.T. and Keene, D. (eds) (1964) *Sources of Japanese Traditions*, New York: Columbia University Press, 2 vols.(Extensive collection of short excepts from philosophical texts in translation, including brief but useful introductions to each writer.)

Watsuji Tetsurō (1937-49) *Rinrigaku (Ethics)*, Tokyo: Iwanami shoten, 3 vols. (Watsuji's fully developed ethical system with critiques of other systems, East and West. The best available edition is in vols 10-11 of *Watsuji Tetsurō zenshū (Watsuji's Collected Works)*, Tokyo: Iwanami shoten, 1961-3, 20 vols.)

Watsuji Tetsurō (1934) *Ningen no gaku toshite no rinrigaku (Ethics as the Study of Human Being)*, Tokyo: Iwanami shoten.(The best source is vol. 9 of *Watsuji Tetsurō zenshū (Watsuji's Collected Works)*, Tokyo: Iwanami shoten, 1961-3, 20 vols. About half of this work has been translated by S. Yamamoto and R.E. Carter, *Watsuji Tetsuro's Rinrigaku: Ethics in Japan*, Albany, NY: State University of New York Press, 1996.)

Jaspers, Karl (1883-1969)

Karl Jaspers is generally known as an existentialist, but he also developed interesting conceptions in other fields of philosophy: in philosophy of religion, the concepts of Transcendence, cipher and philosophical faith; in philosophy of history, the thesis of an Axial Period in history; in political philosophy, the idea of a new, reasonable politics. His existentialism deals mainly with personal moral attitudes and private aspects of individual self-realization in boundary situations and intimate interpersonal communication. His political philosophy concentrates on controversial political affairs and some of the urgent problems of his age (for example, the possibility of extinguishing all life on earth by the atom bomb, or of establishing a world-wide totalitarian regime).

1 Life

Karl Jaspers studied medicine and became a psychiatrist and psychologist before turning to philosophy at the University of Heidelberg in 1920. His first major book, *Allgemeine Psychopathologie (General Psychopathology)* (1913), is a study in the methodology of psychiatry focusing on the relevance of phenomenological and hermeneutical approaches for the methods of psychopathology, while his second major book, *Psychologie der Weltanschauungen* (1919), sets forth a psychology of worldviews. This book shows clearly how much he was influenced by Immanuel Kant, Søren Kierkegaard, Friedrich Nietzsche and Max Weber. Other formative influences on his philosophy stem from Plato, Plotinus, Friedrich Hegel, Friedrich Schelling, Wilhelm Dilthey and Edmund Husserl. Jaspers' existentialism is developed extensively in the three-volume *Philosophie* published in 1932. During the Nazi regime in Germany he was denied the right to teach and forbidden to publish. In 1948 he moved from Heidelberg to Switzerland where he earned a professorship at the University of Basel. The experience of Nazi terrorism led to a break in the continuity of his philosophizing: he no longer understood his philosophy as an existential philosophy, but as one of reason, reflecting on actual political concerns. This change in Jaspers' philosophical perspective becomes evident in his major book in political philosophy, *Die Atombombe und die Zukunft des Menschen (The Atom Bomb and the Future of Man)* (1958). During this period, Jaspers also published his main works in philosophy of religion, *Der philosophische Glaube (The Perennial Scope of Philosophy)* (1948) and *Der philosophische Glaube angesichts der Offenbarung (Philosophical Faith and Revelation)* (1962) and began developing a new and original conception of world-philosophy. When he died, only a single volume of a history of world-philosophy had been published.

2 Existentialism

Jaspers' existentialism is grounded on a conception of the human being as a dual project of self-realization. Combining some ideas of Kant and Kierkegaard, he sees a human being as an empirical phenomenon and as a phenomenon with a non-empirical dimension. In the empirical dimension he distinguishes three modes of human being or self-realization: (1) vital existence or naïve vitality (*bloßes Dasein*); (2) consciousness-as-such (*Bewußtsein überhaupt*), that is, the dimension of logical thinking and rationality; (3) spirit (*Geist*), that is, the dimension of ideas which becomes manifest in personal ideals, creative conceptions of the arts, and so on. These three modes of self-realization are the empirical media for another mode which is non-empirical and non-objective, which Jaspers calls '*Existenz*'. *Existenz* is the authentic ground of human being, the intimate dimension of personal autonomy, existential freedom and true selfhood. No empirical studies or doctrines of ontology, philosophical anthropology, or ethics can give an adequate understanding of this dimension of subjectivity and humanity. Such an understanding is possible only by realizing this dimension in one's own life or by elucidating it through philosophizing.

But how can one realize one's personal autonomy and *Existenz*? Jaspers' answer is that *Existenz* cannot be planned or managed at all because it is a gift from a non-objectifiable being which he calls 'Transcendence'. Furthermore, realization of *Existenz* in specific moments of life is possible only under two conditions: first, the experience of boundary situations and, second, the experience of mutual existential communication with a partner. Boundary situations like death, suffering, struggle, and guilt cannot be overcome by using the objective and rational knowledge that helps us solve everyday problems. They require a radical change in our attitudes and common ways of thinking. In *Psychologie der Weltanschauungen* and volume 2 of *Philosophie* Jaspers presents a remarkable range of deep psychological insights into boundary situations drawn from his experience as a

psychiatrist and psychologist. He shows that many of the usual reactions to such situations prevent us from realizing our *Existenz*. The boundary situation of death, for example, can be the source of fear, anxiety and nihilistic despair, but it can also evoke the urgency of living authentically without delay and self-deception.

Jaspers distinguishes four types of communication operative in the four modes of being by which we exist. In the dimension of naïve vitality a human being lives in primitive communities with other human beings using other people only to reach vital ends. The second dimension, consciousness-as-such, is linked to a type of intellectual communication based on logical and rational categories. In the third dimension, spirit, a human being experiences communication in the close community of an 'idea of a whole', a religion, family, university, party, society, state and so on. The highest and most valuable form of communication, where the human being realizes *Existenz*, is existential communication. This is a non-objective, very intimate personal relationship between two persons, friends, lovers, a married couple, parent and child, or teacher and student. This type of communication cannot be adequately described and verbally communicated in an objectifying language. In Jaspers' subtle analyses of different reactions to boundary situations, as well as in his reflections on existential communication, a number of moral attitudes are mentioned which are highly significant for his concept of *Existenz*: courage without self-deception, composure, self-possession, dignity, fidelity, existential solidarity, responsibility, autonomy. These values constitute a specific ethos of humanity in Jaspers' existentialism (see Existentialism).

3 Metaphysics and philosophy of religion

Self-realization as *Existenz* is experienced as a gift from a transcendent source. This experience directs the subject to a dimension of being which Jaspers calls 'Transcendence', absolute being, being-as-such, God and the Encompassing. His metaphysics rejects every type of metaphysical system or speculation that aims to bridge the gap between immanence and Transcendence. The task of metaphysics is instead to appeal to every human being to 'read' the cipher-script or the 'language' of Transcendence for its own sake. A cipher is a subjective and intuitive symbol that always remains open in its meaning and functions as a signpost to Transcendence. Everything in the world is able to become a cipher. Jaspers explicitly mentions: nature, creations of the arts, metaphysical systems, myths and religions. He emphasizes the experience of foundering (*Scheitern*) as a special cipher and signpost to Transcendence.

Jaspers' philosophy of religion does not accept an objective historical revelation of God. God is hidden forever and cannot speak to us in an objectifying language through mediators. We can only be aware of Transcendence as *Existenz* in a non-objective act of cipher-'reading'. But no cipher can give any information about Transcendence beyond the horizon of historicity. It is a necessary consequence of this position that Jaspers understands Jesus as one of the 'examplary men' in human history alongside Confucius, Buddha and Socrates. Because of the paradigmatic way that Jesus realized existential possibilities, such as the ability to love and capacity for suffering, he is relevant to human beings at all times. In place of any religious faith grounded on a revelation of God, Jaspers holds out the conception of philosophical faith. This sort of faith has no objectively guaranteed proof of the existence of Transcendence and is not bound to rituals, churches, priests and theologians, who claim to be the interpreters of God's revelation. Philosophical faith is an optimistic credo involving confidence in the possibility of freedom and humanity and in the existence of a meta-empirical dimension of being (Transcendence, God, the Encompassing). It claims neither absolute certainty with respect to the possession of moral truth, nor an absolutely true knowledge of this dimension of being (see Existentialist theology).

4 Philosophy of reason and political thinking

The concept of reason became increasingly dominant in Jaspers' later philosophy. His comprehensive book, *Von der Wahrheit* (1947), is an extensive study of the dimensions of reason explicated in the context of the doctrine of the Encompassing. Reason is the bond of all modes of the Encompassing, a dynamic principle that expresses an anti-dogmatic and pluralistic intention, on the one hand, and a universalizing and communicative intention, on the other. These basic intentions of reason influenced Jaspers' political philosophy after the Second World War. During this period, he was concerned with the nature of German guilt for the rise of the Nazi movement in *Die Schuldfrage (The Question of German Guilt)* (1946), and with the problem of German reunification, as well as with global problems such as an emerging world order, the dangers of totalitarianism, the chance for worldwide communication among different nations in the face of the atom bomb and the cold war, the importance and deficiencies of the United Nations Organisation, and the necessary conditions of world peace. Reason works in this

context as a dynamic principle to prevent the dogmatization of political worldviews. It is also the task of reason to give normative impulses to traditional politics, that is, 'Realpolitik', as a means for gaining power and domination over others. Reason transfers the ethos of humanity and moral ideals characteristic of Jaspers' concept of *Existenz* from the private sphere of life to the public sphere. In this way a new reasonable politics based on moral attitudes (for example, intellectual integrity, open-mindedness, tolerance, respect for other persons, groups and people in their cultural and ethnic diversity) may help overcome the two universal political-boundary situations of the modern age: the possibility of annihilating all life on Earth with the atom bomb and the possibility of establishing a worldwide totalitarian regime.

5 Philosophy of history and history of philosophy

Jaspers' philosophy of history, which he develops in *Vom Ursprung und Ziel der Geschichte (The Origin and Goal of History)* (1949), argues for the existence of an Axial Period in history (800-200 BC). During the Axial age the basic categories of human thought, culture and religion were developed in China, India and the West. Those categories have enabled the self-understanding of mankind to continue up to the modern period. The historical consciousness of the Axial Period reminds us not only of certain cultural highlights of mankind (in China, the thoughts of Confucius and Laozi; in India, the Upanishads and Buddha; in Palestine, the prophets; in Greece, the great philosophers and tragedians), but also of the threefold origin of the culture of mankind and of the importance of pluralism in universal history. This historical consciousness provides impulses for a worldwide communication beyond all differences of creed and political opinion in our time. Jaspers' philosophy of history is nondeterministic: history is not a teleological process whereby individuals are determined by historical laws, inevitable fate, or socioeconomic mechanisms immune from human will and intention. Rather, human beings help direct the historical process through their freedom without ever being able completely to predict future tendencies by means of total planning.

For Jaspers, studying the history of philosophy entails an existential act of encounter with great human beings in the history of mankind. Great philosophers can influence life orientations and modes of thinking in all times because they make us aware of basic existential possibilities and fundamental attitudes of thinking as such. Jaspers intended to write a world history of philosophy based on five distinct approaches to the history of philosophy (historical, thematic, genetic, practical, dynamic). This project remained unfinished. During Jaspers' lifetime, only the book *Die Großen Philosophen (The Great Philosophers)* (1957) appeared; but two volumes with fragments of this immense project were published in 1981, long after his death.

KURT SALAMUN

List of works

Jaspers, K. (1913) *Allgemeine Psychopathologie*, Berlin: Springer; 7th edn, trans. J. Hoenig and M.W. Hamilton, *General Psychopathology*, Chicago, IL: University of Chicago Press, 1963.(A methodological work which focuses on the hermeneutic approach in psychiatry.)

Jaspers, K (1919) *Psychologie der Weltanschauungen (Psychology of Worldviews)*, Berlin: Springer; 6th edn, Munich, 1985.(A psychological analysis of worldviews which anticipates basic ideas of Jaspers' later existentialism.)

Jaspers, K (1932) *Philosophie*, Berlin: Springer, 3 vols; trans. E.B. Ashton, *Philosophy*, Chicago, IL: University of Chicago Press, 3 vols, 1969-71.(Jaspers' main work in existentialism, comprising three volumes: World-Orientation, Existential Elucidation and Metaphysics.)

Jaspers, K (1935) *Vernunft und Existenz*, Groningen: J.W. Wolters; trans. W. Earle, *Reason and Existenz*, London: Routledge & Kegan Paul, 1956.(A collection of five lectures concerning the relation between Existenz and reason.)

Jaspers, K (1946) *Die Schuldfrage*, Heidelberg/Zurich; trans. E.B. Ashton, *The Question of German Guilt*, New York: Dial Press, 1947.(This book treats the question of German guilt in relation to the rise of the Nazi regime.)

Jaspers, K (1947) *Von der Wahrheit (On Truth)*, Munich: R. Piper.(In this book Jaspers' conception of the Encompassing and his philosophy of reason are developed.)

Jaspers, K (1948) *Der philosophische Glaube*, Zurich; trans. R. Manheim, *The Perennial Scope of Philosophy*, London: Routledge & Kegan Paul, 1950.(The first explication of Jaspers' conception of philosophical faith.)

Jaspers, K (1949) *Vom Ursprung und Ziel der Geschichte*, Zurich: Artemis; trans. M. Bullock, *The Origin and*

Goal of History, London: Routledge & Kegan Paul, 1953.(A representation of Jaspers' philosophy of history in which he sets forth his idea of an Axial Age in world history.)

Jaspers, K. and Bultmann, R. (1954) *Die Frage der Entmythologisierung*, Munich; trans. N. Guterman, *Myth and Christianity*, New York: Noonday Press, 1958.(The question of the inevitable function of myth in Christianity is discussed in opposition to Bultmann's call for demythologization.)

Jaspers, K. (1957) *Die Großen Philosophen Erster Band*, Munich; ed. H. Arendt, trans. R. Manheim, *The Great Philosophers*, New York: Harcourt Brace, 2 vols, 1962, 1966.(The first volume of Jaspers' enormous project of a world history of philosophy.)

Jaspers, K. (1958) *Die Atombombe und die Zukunft des Menschen*, Munich; trans. E.B. Ashton, *The Future of Mankind* (also as *The Atom Bomb and the Future of Man*), Chicago, IL: University of Chicago Press, 1961. (Jaspers' basic work in political philosophy which gives a thorough analysis of the implications of human existence after the advent of atomic weapons.)

Jaspers, K. (1962) *Der philosophische Glaube angesichts der Offenbarung*, Munich: R. Piper; trans. E.B. Ashton, *Philosophical Faith and Revelation*, Chicago, IL: University of Chicago Press, 1967.(A critical engagement of Jaspers' conception of philosophical faith with religious faith based on revelation.)

Jaspers, K. (1981) *Die Großen Philosophen Nachlaß 1 und 2*, Munich; ed. M. Ermarth and L.H. Ehrlich, trans. E. Ehrlich and L.H. Ehrlich, *The Great Philosophers Vol. III*, New York: Harcourt Brace, 1993.(Fragments of Jaspers' project of a world history of philosophy.)

Jaspers, K. (1986) *Karl Jaspers: Basic Philosophical Writings; Selections*, ed., trans. and with intro. by L.H. Ehrlich, E. Ehrlich and G.B. Pepper, Athens, Ohio and London: Ohio University Press; 2nd edn, Atlantic Highlands, NJ: Humanities Press, 1994.(An excellent selection of key texts with competent introductions and translations.)

References and further reading

Ehrlich, L.H. (1975) *Philosophy as Faith*, Amherst, MA: University of Massachusetts Press.(Focusing on Jaspers' philosophy of religion.)

Ehrlich, L.H. and Wisser, R. (eds) (1988) *Karl Jaspers Today: Philosophy at the Threshold of the Future*, Lanham, MD: University Press of America.(A collection of papers initially presented on the occasion of the centenary year of Jaspers' birth at the 17th World Congress of Philosophy.)

Hersch, J. (1986) *Karl Jaspers: Eine Einführung in sein Werk*, Munich: R. Piper.(An introduction to the main ideas of Jaspers' philosophy.)

Köhler, L. and Saner, H. (eds) (1985) *Hannah Arendt Karl Jaspers Briefwechsel 1926-1969*, Munich; trans. R. Kimber and R. Kimber, *Hannah Arendt Karl Jaspers Correspondence 1926-1969*, New York: Harcourt, 1992. (Gives fruitful insights into the close relationship and friendship of both philosophers.)

Olson, A.M. (1979) *Transcendence and Hermeneutics: An Interpretation of Karl Jaspers*, The Hague: Nijhoff. (Focuses on hermeneutical aspects of Jaspers' philosophy that are relevant to the philosophy of religion.)

Olson, A.M. (ed.) (1994) *Heidegger and Jaspers*, Philadelphia, PA: Temple University Press.(A collection of essays that incorporates the results of investigations into the Jaspers-Heidegger relationship presented at conferences of the Jaspers Society of North America.)

Penzo, G. (1985) *Il comprendere in Karl Jaspers e il problema dell' ermeneutica*, Rome: Armando editore.(Points out the hermeneutical dimension of Jaspers' philosophizing.)

Salamun, K. (1985) *Karl Jaspers*, Munich: C.H. Beck.(A general introduction with critical arguments concerning methodological positions in Jaspers' existentialism.)

Salamun, K. (ed.) (1991) *Karl Jaspers - Zur Aktualität seines Denkens*, Munich: R. Piper.(A collection of articles demonstrating the continuing importance of Jaspers' philosophy for the present age.)

Salamun-Hybasek, E. and Salamun, K. (eds) (1988-95) *Jahrbuch der Österreichischen Karl-Jaspers-Gesellschaft (Yearbook of the Austrian Karl-Jaspers-Society)* 1-8, Innsbruck, and Vienna: Österr. Studienverlag.(Periodical concerning Jaspers published by the Austrian Karl-Jaspers-Society.)

Samay, S. (1971) *Reason Revisited: The Philosophy of Karl Jaspers*, Notre Dame, IN: University of Notre Dame Press. (Treats the relevance of Jaspers' concept of reason for his conceptions of philosophy, science and metaphysics.)

Saner, H. (1970) *Karl Jaspers in Selbstzeugnissen und Bilddokumenten*, Reinbeck: Rowohlt.(Chronological introduction stressing also biographical facts.)

Saner, H. (ed.) (1970) *Karl Jaspers in der Diskussion*, Munich: R. Piper.(A collection of articles and reviews dealing with several fields of Jaspers' philosophical thinking from a critical point of view.)

Schilpp, P.A. (ed.) (1981) *The Philosophy of Karl Jaspers*, La Salle, IL: Open Court.(The essays in this collection

address a wide range of basic ideas in Jaspers' philosophizing. Includes a 'Reply' by Jaspers to his critics.)

Schrag, O.O. (1971) *Existence, Existenz, and Transcendence*, Pittsburgh, PA: Duquesne University Press. (Discusses Jaspers' position in philosophy of religion, especially his concept of cipher.)

Tilliette, X. (1959) *Karl Jaspers: Theorié de la vérité, Metaphysique des chiffres, Foi philosophique*, Paris: Aubier.(Provides fruitful insights into the metaphysical components of Jaspers' philosophy.)

Walraff, C.F. (1970) *Karl Jaspers: An Introduction to His Philosophy*, Princeton, NJ: Princeton University Press. (A short and clear introduction to basic themes of Jaspers' existential philosophy, his conception of philosophy, and his idea of the Encompassing.)

Walters, G.J. (1988) *Karl Jaspers and the Role of 'Conversion' in the Nuclear Age*, Lanham, MD: University Press of America.(An interesting application of Jaspers' existentialist and political idea of 'conversion' to the problems of nuclear war and disarmament.)

Walters, G.J. (ed.) (1995) *The Tasks of Truth: Essays on Karl Jaspers' Idea of the University*, Frankfurt and New York: Peter Lang.(A collection of articles about Jaspers' philosophy of education.)

Wisser, R. and Ehrlich, L.H. (eds) (1993) *Karl Jaspers: Philosopher among Philosophers*, Würzburg: Königshausen & Neumann.(A collection with contributions concerning the relation of Jaspers to other philosophers.)

Young-Bruehl, E. (1981) *Freedom and Karl Jaspers' Philosophy*, New Haven, CT: Yale University Press.(An analysis of the various aspects of Jaspers' concept of freedom.)

Jefferson, Thomas (1743-1826)

Thomas Jefferson came from a privileged background, began his public career as a lawyer, and rose to hold the governorship of his home state of Virginia and to serve his country successively as minister to France, secretary of state, vice-president and president. His proudest declared achievements were to have been author of the Declaration of Independence, and of the Statute of Virginia for religious freedom, and founder of the University of Virginia. An architect, inventor, scientist, educator and writer, he was one of the most versatile and brilliant men of his generation. In philosophy his main contribution was to political theory, where he supported a social contract theory and a doctrine of natural rights.

1 Life

Thomas Jefferson was born on 13 April 1743 at Shadwell in Virginia, heir of a wealthy family. Educated (1760-62) at William and Mary College, he became a lawyer, served (from 1769) in the Virginia House of Burgesses, and in 1775 and 1776 in the Continental Congress. In June 1776 he was appointed head of a committee of five to prepare the Declaration of Independence, of which he was the principal author. From 1779 onwards he held a number of major political posts, including that of US minister to France (where he succeeded Benjamin Franklin in 1785), the vice-presidency of the USA from 1796 to 1800, and the presidency from 1800 to 1808.

2 Philosophical outlook

Jefferson's intellectual debts were particularly to John Locke, Isaac Newton, Dugald Stewart, Destutt de Tracy, Pierre-Jean Cabanis, and J.P. Flourens. He was a materialist whose model of reality derived from Newton, but also a deist who considered Jesus a great moral teacher but not divine. From Stewart and the Scots he absorbed the moral sense theory, but his realism owed more to the French Ideologues than the Scots (see Enlightenment, Continental).

Jefferson's philosophical contribution was in political philosophy. He was a firm believer in the social contract theory and in natural rights, which he believed were based in the biological constitution of the individual and God's benevolence in making possible the satisfaction of those needs with which individuals were created. Jefferson's view of the political order was normative rather than descriptive, but it was empirical in that it was founded on the moral sense - violations of natural rights are immediately known to be wrong. Majority rule, while essential to democracy, cannot therefore involve infringement of the natural rights of the minority. Believing that sovereignty lay undivided in the people, he thought the people could employ multiple political agents to perform their will. This led to a hierarchy of political units - individual, local, state, and Federal - on the principle that each should be responsible for those functions which it can perform alone, delegating to the larger unit those which it cannot.

Although concerned with the machinery of government and the methods of balancing the units to prevent concentration of power, Jefferson believed that structure was not enough; the success of democracy lay in the values and attributes of the people. But these in turn, he thought, were influenced by society and education. For Jefferson, a free society had to be one in which people were independent, not only politically but also economically, through the ownership of property. In the society of yeomen farmers he saw the nearest realization of his ideal of a nation of free, responsible, and moral individuals. Free elections, he believed, would result in leadership being furnished by the 'natural aristocracy' of talent and virtue, for the cultivation of which he proposed a system of public education. But Jefferson's democracy did not include blacks; although he condemned slavery, he held slaves and regarded blacks as inferior.

See also: American Philosophy in the 18th and 19th centuries; Enlightenment, Scottish; Franklin, B.

MURRAY G. MURPHEY

List of works

Jefferson, T. (1892-9) *The Writings of Thomas Jefferson*, ed. P. Leicester Ford, New York: Putnam, 10 vols.(An older edition, but covering the span of Jefferson's life.)

Jefferson, T. (1950-92) *The Papers of Thomas Jefferson*, vols. 1-19 ed. Julian Boyd et al, vols. 20-23, ed. C.

Cullen, vols. 24-6, ed. J. Catanzariti, Princeton, NJ: Princeton University Press.(The best modern scholarly edition, done in chronological order, but still in progress.)

Jefferson, T. (1984) 'The Declaration of Independence' in *Thomas Jefferson: Writings*, ed. M. D. Peterson, New York, Library of Classics of the United States Inc., 19-24.(This text gives the original version of the Declaration as Jefferson wrote it, together with the changes made before its adoption and promulgation on July 4, 1776. The Declaration of Independence is, together with the Constitution, the most important state paper in American history.)

References and further reading

Boorstin, D. (1948) *The Lost World of Thomas Jefferson*, New York: Holt.(Excellent study of the Jeffersonian intellectual world.)

Flower, E. and Murphey, M.G. (1977) *A History of Philosophy in America*, New York: Putnam, vol. 1, chap. 5. (A brief study of Jefferson's philosophy.)

Koch, A. (1943) *The Philosophy of Thomas Jefferson*, New York: Columbia University Press.(Dated but useful study of Jefferson's thought.)

Sheldon, G.W. (1991) *The Political Philosophy of Thomas Jefferson*, Baltimore: John Hopkins Press.(Relying on the recent work of J.G.A. Pocock, Joyce Appleby and others, Sheldon attempts to fit Jefferson's thought into the current debates over Republicanism, Liberalism and Classicism.)

Wills, G. (1978) *Inventing America: Jefferson's Declaration of Independence*, Garden City, NY: Doubleday. (Particularly good on the Scottish influence on Jefferson.)

Jewish philosophy

Jewish philosophy is philosophical inquiry informed by the texts, traditions and experiences of the Jewish people. Its concerns range from the farthest reaches of cosmological speculation to the most intimate theatres of ethical choice and the most exigent fora of political debate. What distinguishes it as Jewish is the confidence of its practitioners that the literary catena of Jewish tradition contains insights and articulates values of lasting philosophical import. One mark of the enduring import of these ideas and values is their articulation in a variety of idioms, from the mythic and archetypal discourse of the Book of Genesis to the ethical and legislative prescriptions of the Pentateuch at large, to the admonitions of the Prophets, the juridical and allegorical midrash and dialectics of the Rabbis, and to the systematic demonstrations, flights of imagination, existential declarations and apercus of philosophers in the modern or the medieval mode.

1 The nature of Jewish philosophy

Students of Jewish philosophy, especially those who aspire to contribute a window or a wing to the edifice, must learn many languages, to read and listen to voices very different from their own. Just as the writers of the Genesis narratives or of the Pentateuch had to recast and reinvent the ancient creation myths and the ancient Babylonian laws to express the distinctively universal ethical demands and aesthetic standards of their God, and just as the Deuteronomist had to rediscover the ethical core in the original Mosaic legislation, hearing God’s commands now as urgent reminders through the very human voice of Moses, so in every generation new interpreters are needed, to rediscover what is essential and living in the tradition. Such interpreters have always needed to negotiate the rapids of historical change - not just with regard to idiom but also with regard to content, refocusing and restructuring the living tradition, sculpting it philosophically with their own moieties of reason. Such thinkers have worked always with a view to the continuity of the tradition; that is, to the faithfulness of its future to its past, but also to the vitality and vivacity of what they found timeless in the tradition and therefore capable of acquiring new meanings and new spheres of application in the present.

The confidence of the practitioners of Jewish philosophy in the conceptual vitality and continually renewed moral and spiritual relevance of the tradition is typically the reflex of an existential commitment to that tradition and to the people who are its bearers. That confidence, and its repeated vindication by the richness of the tradition itself, is also a wellspring of renewal and encouragement for the commitment that energizes it - even, and especially, in times of historical crisis and external pressures, which have rarely confined themselves to sheerly intellectual challenges. Symptomatic of that commitment is the prominence and recurrence of the philosophy of Judaism among the concerns of Jewish philosophy. However, the two should not be confused. The philosophy of Judaism is inquiry into the nature and meaning of Jewish existence. Its questions address the sense to be given to the idea of a covenant between the universal God and the people of Israel, the meaning of that people’s mission, their chosenness, their distinctive laws, customs and rituals and the relation of those norms to the more widely recognized norms of humanity, of which the Prophets of Israel were early and insistent messengers.

The philosophy of Judaism wants to understand Zionism, the Holocaust, the Jewish Diaspora and the historical vicissitudes that gave shape to Jewish experience over the millennia, from the age of the biblical patriarchs to the destruction of the first and second temples in Jerusalem, to the exile of the Jewish people and the return of many, after a hundred generations, to the land they had been promised and in which they had prospered, a land which some had never left but which most, for centuries, had pictured only through the sublimating lenses of sacred history, apocalypse and philosophy. The philosophy of Judaism wants to understand the ancient Jewish liturgy, the exegetical practices and hermeneutical standards of the Jewish exegetes. Like Freud, it wants to understand Jewish humour. Like Pico della Mirandola, it wants to understand Kabbalah, Jewish mysticism, and like Buber, it wants to understand Hasidism. The concerns of the philosophy of Judaism touch every aspect of Jewish experience, just as the concerns of philosophy at large touch every aspect of experience in general. But the concerns of Jewish philosophy, like those of general philosophy, do not confine themselves to Jewish experience. They are, in fact, the same concerns as those of general philosophy, rendered distinctively Jewish by their steady recourse to the resources of the tradition, and sustained as philosophical by an insistence on critical receptivity, responsible but creative appropriation of ideas and values that withstand the scrutiny of reason and indeed grow and give fruit in its light.

2 Strengths and weaknesses

There are two weaknesses in Jewish philosophy as practised today. One is a tendency to historicism, that is, the equivocal equation of norm with facticity and facticity with norm that leads to an abdication of philosophical engagement for a detached clinical posture or an equally unwholesome surrender of judgment to the flow of events. Historicism is a natural by-product of respect for tradition, or of expectation of progress. It becomes particularly debilitating under the pressure of positivism, whether of the logical empiricist sort that dominated philosophy for much of the early twentieth century, or of the more endemic sort that thrives on the sheer givenness of any system of law and ritual or that allows itself to be overwhelmed by the press of history itself. It is not unusual, even today, when logical positivism is widely thought to be long dead, to find scholars of Jewish thought who substitute historical descriptions for philosophical investigations, often in the process begging or slighting the key philosophical questions. Nor is it unusual among those of more traditional stamp for scholars to be found who imagine that a faithful description of the contents of authentic Jewish documents constitutes doing Jewish philosophy - as though faithfulness to the tradition were somehow a substitute for critical grappling with the issues and problems, and as though the question as to what constitutes faithfulness to the tradition, conceptually, historically, morally and spiritually, were not itself among the most crucial of those issues and problems.

The second weakness is a narrowing of the gaze, a tendency to substitute philosophy of Judaism for the wider discourse of Jewish philosophy, as though the resources of the tradition had nothing (or nothing more) to contribute to ethics, or natural theology, or metaphysics and logic, for that matter. The work of the great practitioners of Jewish philosophy has repeatedly given the lie to such narrow expectations. In every epoch of its existence, Jewish philosophy has played an active role in the philosophical conversation of humankind - which is a universal conversation precisely because and to the extent that those who take part speak every language and bring to the conversation experiences that are universal as well as those that are unique.

But if two weaknesses are to be mentioned here, at least one strength should be cited as well: Jewish philosophy, although intimately engaged throughout its history with the philosophical traditions of the West, has also been a tradition apart. The open access of most of its practitioners to the Hebrew (and Aramaic) Jewish sources has afforded a perspective that is distinctive and that can be corrective of biases found in other branchings of the tree of philosophical learning. The early access of medieval Jewish philosophers to Arabic philosophical and scientific writings, and to the Greek works preserved in Arabic, enriched and broadened their philosophical repertoire. The scholastic learning of later medieval Jewish philosophers and their collaboration with scholastic thinkers made them at once participants and observers of in the lively philosophical debates of their day. The immersion and active participation of Renaissance and Enlightenment Jewish philosophers in the movements that spawned modernity gave them a similar philosophical vantage point. All philosophers must be, to some degree, alien to their society - Socrates and Nietzsche, and for that matter even Plato, Aristotle and Descartes were, to some degree, intellectual outsiders in their own times - not so alien as to have no word or thought in common with their contemporaries, but not so well integrated as to become mere apologists, or complacent and unquestioning acquiescors in the given. Jewish philosophy has long made and continues to make a distinctive, if today underutilized, contribution to cosmopolitan philosophical discourse in this regard. It shares the problematic of Western philosophy but typically offers a distinctive slant or perspective that calls into question accepted verities and thus enhances the critical edge of philosophical work for those who study it.

3 Movements and important figures

Jewish philosophy has over the course of its history been the source of a number of different types of study based on the philosophically relevant ideas of the Hebrew Bible, Rabbinic Law (Halakhah), Rabbinic theology and Rabbinic homiletics, exegesis and hermeneutics (midrash) (see Bible, Hebrew; Halakhah; Theology, Rabbinic; Midrash). The anti-Rabbinical, biblicist movement known as Karaism and the mystical tradition of the Kabbalah are examples of differing types of movements which have emerged (see Karaism; Kabbalah), while Jewish voluntarism and Jewish Averroism were fields for the rivalry between intellectualist and less deterministic, more empiricist views of theology as it was played out among Jewish thinkers (see Voluntarism, Jewish; Averroism, Jewish). More modern movements include the Jewish pietist movement founded by Israel Baal Shem Tov and known as Hasidism, the Jewish Enlightenment movement known as the Haskalah, and Zionism, the movement that led to the establishment of the modern State of Israel (see Hasidism; Enlightenment, Jewish; Zionism).

The first exponent of Jewish philosophy was Philo of Alexandria, a major contributor to the synthesis of Stoicism, Middle Platonism and monotheistic ideas that helped forge the tradition of scriptural philosophy in the West. Other early figures include Daud al-Muqammas and Isaac Israeli, two of the first figures of medieval Jewish philosophical theology. Al-Fayyumi Saadiah Gaon (882-942), the first systematic Jewish philosopher, was also a major biblical translator and exegete, a grammarian, lexicographer and authority on Jewish religious law and ritual. The rationalism, pluralism and intellectual honesty evident in his work made it a model of Jewish philosophy for all who came after him. Solomon ibn Gabirol (*c.*1020-*c.*1057), long known as a Hebrew poet, was discovered in the nineteenth century to have been the author as well of the famous Neoplatonic philosophical work, preserved in Latin as the *Fons Vitae*. Moses ibn Ezra (*c.*1055-after 1135) is notable for his poetic and philosophic conributions. Abraham ibn Ezra (*c.*1089-1164) is likewise noted for his hermeneutical ideas and methods; his forthright approach to the Hebrew Bible was a critical influence on the thinking of Jewish philosophers from the Middle Ages to Spinoza and beyond. A less familiar figure is Abu 'l-Barakat al-Baghdadi (*fl. c.*1200-50), a brilliant Jewish thinker who converted to Islam late in life. He developed highly independent views about the nature of time, human consciousness, space, matter and motion. His work undercuts the notion that the medieval period was simply an age of faith and static commitment to a faith community.

A polymath of rather different spirit was Abraham bar Hayya in the eleventh century, who wrote on astronomy, mathematics, geography, optics and music as well as philosophy and who collaborated on scientific translations with the Christian scholar Plato of Tivoli, the transmitter of the Ptolemaic system to the Latin world. Bar Hayya's *Meditation of the Sad Soul* expresses the forlornness of human life in exile from the world of the divine, a forlornness tinged with the hope of future glory. Joseph ibn Tzaddik (d. 1149) similarly developed Neoplatonic ideas around the theme of the human being as a microcosm.

Bahya ibn Pakuda (early twelfth century) wrote as a pietist philosopher. He placed philosophical understanding and critical thinking at the core of the spiritual devotion called for by the sincerest form of piety. Judah Halevi (before 1075-1141), probably the greatest Hebrew poet after the Psalms, wrote a cogently argued philosophical dialogue best known as the *Kuzari*, but more formally titled, *A Defence and an Argument in behalf of the Abased Religion*. Set in the Khazar kingdom, whose king, historically, had converted to Judaism, the work mounts a trenchant critique of the intellectualism of the prevalent philosophical school and the spiritualizing and universalizing ascetic pietism that was its counterpart. Calling for a robust recovery of Jewish life and peoplehood in the Land of Israel, the work is not only a striking anticipation of Zionist ideas but a remarkable expression of the need to reintegrate the spiritual, intellectual, moral and physical dimensions of Jewish life.

Abraham ibn Daud (*c.*1110-80), a historian as well as a philosopher, used his historiography to argue for the providential continuity of the Jewish intellectual and religious tradition. His philosophical work laid the technical foundations that made possible the philosophical achievement of Moses Maimonides (1138-1204), the greatest of the philosophers committed to the Jewish tradition. Besides his medical writings and his extensive juridical corpus, which includes the authoritative fourteen-volume code of Jewish law, the *Mishneh Torah*, Maimonides was the author of the famous *Guide to the Perplexed*. Written in Arabic and intended for an inquirer puzzled by the apparent discrepancies between traditional Judaism and Aristotelian-Neoplatonic philosophy, the *Guide* is a paradigm in the theology of transcendence, addressing questions ranging from the overt anthropomorphism of the scriptural text to the purposes of the Mosaic legislation, to the controversy over the creation or eternity of the world, the problem of evil, and the sense that can be made of the ideas of revelation, providence, divine knowledge and human perfectibility. Like Halevi's *Kuzari* and Bahya's *Duties of the Heart*, the *Guide to the Perplexed* continues to be studied to this day by Jews and non-Jews for its philosophical insights.

Abraham ben Moses Maimonides (1186-1237), the son of the great philosopher and jurist, began his scholarly life as a defender of his father's work against the many critics who feared Maimonidean rationalism. In his mature work he became the exponent of a mystical, pietist and ascetic movement, largely influenced by Sufism. Moses Nahmanides (1194-1270), exegete, theologian and a founding figure of the Kabbalistic theosophy, championed Judaism in the infamous Barcelona Disputation of 1263 and played a leading role in the Maimonidean controversy. He struggled to harmonize his conservative and reactive tendencies with his respect for reason and the unvarnished sense of the biblical text.

Ibn Kammuna (d. 1284) was a pioneer in other areas. Besides his work in the Ishraqi or Illuminationist tradition of

theosophy, laid out in commentary on the Muslim philosopher Ibn Sina (Avicenna), he wrote a distinctively dispassionate study of comparative religions, favouring Judaism but fairly and unpolemically presenting the Christian and Muslim alternatives.

Shem Tov ibn Falaquera (*c*.1225-*c*.1295) was a warm exponent of Maimonidean rationalism and an ardent believer in the interdependence of faith and reason. His selections in Hebrew from the lost Arabic original of Ibn Gabirol's magnum opus allowed modern scholars to identify Ibn Gabirol as the Avicebrol of the surviving Latin text, the *Fons Vitae*.

Hillel ben Samuel of Verona (*c*.1220-95), physician, translator, Talmudist and philosopher was a Maimonist who introduced numerous scholastic ideas into Hebrew philosophical discourse. Immanuel of Rome (*c*.1261-before 1336) was a prolific author of philosophical poetry and exegesis, often praising reason and intellectual love. Judah ben Moses of Rome (*c*.1292-after 1330), known as Judah Romano, was an active bridge person between the Judaeo-Arabic and the scholastic tradition of philosophical theology.

Levi ben Gershom, known as Gersonides (1288-1344), was an important astronomer and mathematician as well as a biblical exegete and philosopher. His *Wars of the Lord* grappled with the problems of creation, providence, divine knowledge, human freedom and immortality. Aiming to defend his ancestral faith, Gersonides followed courageously where the argument led, often into radical and creative departures from traditional views.

Hasdai Crescas (1340-1410), an ardent defender of Judaism against Christian conversionary pressures, was among the most creative figures of Jewish philosophy, challenging many of the givens of Aristotelianism, including the idea that the cosmos must be finite in extent. Crescas' student Joseph Albo (*c*.1360-1444) sought to organize Jewish theology into an axiomatic system, in part to render Jewish thought defensible against hostile critics.

Profiat Duran (d. *c.* 1414), also known as Efodi, used his extensive understanding of Christian culture to criticize Christianity from a Jewish perspective. Deeply influenced by Moses Maimonides and Abraham ibn Ezra and by Neoplatonic and astrological ideas, he sought to balance the practical with the intellectual aspects of the Torah. Simeon ben Tzemach Duran (1361-1444) contributed an original approach to the project of Jewish dogmatics and an implicit critical examination of that project.

The Shem Tov family included four thinkers active in fifteenth century Spain (see Shem Tov family). Their works follow the persecution of 1391 and the ensuing mass apostasy of Spanish Jews and seek to rethink the relations of philosophy to Judaism. Shem Tov, the *paterfamilias*, criticized Maimonides and endorsed Kabbalah, but his sons Joseph, a court physician and auditor of royal accounts at Castile, and Isaac, a popular teacher of Aristotelian philosophy, and Joseph's son, again named Shem Tov, wrote numerous Peripatetic commentaries. These offspring charted a more moderate course that enabled Jewish intellectuals to cultivate philosophy and the kindred arts and sciences while asserting the ultimate primacy of their revealed faith.

Isaac ben Moses Arama (*c*.1420-94), like Nahmanides, was critical of Maimonidean and Aristotelian rationalism but did not discard reason, seeing in it a crucial exegetical tool and an avenue toward understanding miracles and providence. Isaac Abravanel (1437-1508), leader of the Jews whom Ferdinand and Isabella exiled from Spain in 1492, like Arama criticised Maimonidean rationalism in the interest of traditional Judaism as he saw it, but at the same time put forward a theistic vision of history and strikingly modern views about politics and the state. His son, Judah ben Isaac Abravanel, also known as Leone Ebreo (*c*.1460-*c*.1521), wrote the *Dialoghi d'amore*. Couched in the language of courtly love, the work explores the idea that love is the animating force of the cosmos. The work stands out as a brilliant dialectical exploration of the differences and complementarities of the Platonic and Aristotelian approaches to philosophy.

Judah Messer Leon (*c*.1425-*c*.1495) was a philosopher, physician, jurist, communal leader, poet and orator. Awarded a doctorate in medicine and philosophy by the Emperor Frederick III, he could confer doctoral degrees in those subjects on the students in his *yeshivah*. He saw logic as the key to harmonizing religion and philosophy and favored scholastic logic over the Arabic logical works. His encyclopedia became a popular textbook, and his systematic elicitation of Hebrew rhetoric from the biblical text, in *The Book of the Honeycomb's Flow*, one of the first Hebrew books to be printed, was a masterpiece of cross-cultural humanistic scholarship. But Messer Leon failed to curb the spread of Kabbalah, whose underlying Platonic metaphysics he abhorred and whose appropriation by Christian Platonists he held in deep suspicion. Indeed, his own son turned toward the Kabbalah

and sought to combine its teachings with the Aristotelianism favored by his father.

Yohanan ben Isaac Alemanno (1433/4-after 1503/4) brought together in his thinking Averroist, Kabbalistic, Neoplatonic and Renaissance humanist themes. He instructed Pico della Mirandola in Hebrew and in Kabbalah, bringing to birth what became a Christian, syncretic Kabbalism. Elijah Delmedigo (*c*.1460-93) was an Aristotelian and Averroist. He translated works into Latin for Pico della Mirandola and developed a subtle critique of the kabbalistic ideas that in his time were rivaling and often displacing what he saw as more disciplined philosophical thinking. Abraham Cohen de Herrera (*c*.1562-*c*.1635) was a philosophically oriented kabbalist of Spanish origin. His Spanish writings, in Latin translation, were blamed for inspiring Spinoza's views.

4 Movements and important figures (cont.)

Moses Mendelssohn (1729-86), a leading figure of the European Enlightenment, spread Enlightenment ideas to Hebrew literature, fought for Jewish civil rights and did pioneering conceptual work on political theory, especially with regard to religious liberty in his Jerusalem. Solomon Maimon (1753/4-1800) took his name in honour of Moses Maimonides. Trained as a rabbi, he pursued secular and scientific learning and became an important and original critic of the philosophy of Kant. Nachman Krochmal (1785-1840), a leader of the Jewish Enlightenment in Galicia, found anticipations of Kant, Hegel and Schelling in the ancient Jewish writings. His work shows how a thinker whose underlying assumptions differ from those of the idealist philosophers could take their views in quite a different direction from the one they chose.

Hermann Cohen (1842-1918), a major Kantian philosopher and one of the first non-baptized Jews to hold an important academic post in Germany, applied his own distinctive version of critical idealism to the understanding of Judaism as a spiritual and ethical system. Franz Rosenzweig (1886-1929), an important Hegelian thinker, went on to formulate a Jewish existential philosophy that deeply influenced many of the most prominent Jewish thinkers of the twentieth century. Martin Buber (1878-1965), Zionist advocate of accommodation with the Palestinian Arabs and an admiring student of Hasidic traditions, added his own stamp to the continental tradition of Jewish philosophy by developing a widely influential dialogical philosophy that privileged relationships experientially and celebrated the I-thou, a mode of relation that allows for authentic encounter.

A number of twentieth century philosophers of Judaism have grasped at diverse threads of the Jewish experience, illustrating both the attractions of the tradition and the fragmentation produced by centuries of persecution that would culminate in the Holocaust, only to be accentuated by the centrifugal tendencies of Jewish life in post-Holocaust liberal societies. Ahad Ha'Am, the pen name of Asher Ginzberg (1856-1927), was an essayist who argued that the creation of a 'spiritual centre' of Jewish culture in Palestine would provide the sustenance needed to preserve the diaspora Jewry from the threat of assimilation. No state was needed. David Baumgardt (1890-1963) was a philosopher who sought to reconcile ethical naturalism with the ideals he found in the Jewish sources, but, unlike Hermann Cohen, Baumgardt did not explore those sources in close detail. Mordecai Kaplan (1881-1981) sought to devise a social mission and communal identity for Jews without reliance on many of the core beliefs and practices that had shaped that identity in the past. Abraham Joshua Heschel (1907-72) sought to salvage the spiritual dimensions of Jewish experience, which found expression both in ritual and in ethical and social action. Joseph Soloveitchik (1903-93) gave canonical expression to Orthodox ideals by focusing on the intellectual and ritual rigours of his archetypal figures, Halakhic man and the Lonely Man of Faith. Yeshayahu Leibowitz (1903-94), an influential Israeli thinker, struggled for the disengagement of authentic and committed religious observance from the toils of governmental officialdom. Jews are mandated, he argued, to observance, as a community. That imperative is not to be put aside. Neither can the observant pretend to ignore the State of Israel. But the State can give no mandate to religious observance, and religious faithfulness can impart none of its aura to the State. For it is essential not to place God in the service of politics. Emil Fackenheim (1916-) seeks an authentic response to the Holocaust, which he formulates in an intentionally inclusionary way, as a '614th' commandment, not to hand Hitler a posthumous victory but to find some way, that might vary from individual to individual, of keeping alive Jewish ideas, practices and commitments.

See also: Anti-Semitism; Bioethics, Jewish; Enlightenment, Jewish; Holocaust, the; Islamic philosophy; Jewish philosophy in the early 19th century; Jewish philosophy, contemporary; Medieval philosophy; Renaissance philosophy; Religion, philosophy of; Zionism

L.E. GOODMAN

Jewish philosophy in the early 19th century

Although Jewish philosophy flourished in the Middle Ages, it underwent a serious decline in 1492, when the Jews were expelled from Spain. The period following Kant and Mendelssohn witnessed an attempt to reintegrate Jewish philosophy into the mainstream of Western culture. The strategy of reintegration consisted of two elements: (1) showing that there is more to Judaism than the study of Scripture and (2) arguing that some of the ideas that won favour in the Enlightenment were anticipated by Jews centuries earlier.

The most central idea that Jewish thinkers claimed as their own is a shift in focus from theoretical issues to practical ones. As important features of the medieval worldview fell into disrepute, many philosophers began to ask whether there are limits to what human reason can know. Can it really prove that God exists or that the soul is immortal? One response was to argue that even if no proof can be found, there are still grounds for believing, particularly if our understanding of ourselves as moral agents makes no sense without them.

But it did not take long for people to argue that moral agency requires not just a God but a free and transcendent God capable of issuing commands to free agents created in the divine image. Here too, Jewish thinkers claimed that the idea of a God who is not limited by nature and insists on mercy and justice for all people was an integral part of monotheism as understood by the Hebrew prophets.

1 The legacy of Mendelssohn and Kant

It is generally agreed that Jewish Emancipation and modern Jewish philosophy began with Moses Mendelssohn. A product of, and participant in, the Enlightenment, Mendelssohn saw no incompatibility between his commitment to Judaism and his integration into modern European society (see Enlightenment, Continental). According to Mendelssohn (1783), the truths on which religion is based - the existence of God, divine providence, and the immortality of the soul - can be grasped by all people through the aid of reason. Since these truths are required for the pursuit of happiness, revelation is not only unnecessary but unjustifiable. Why would God reveal the secrets of happiness to one nation and leave the rest of humanity in the dark? Thus Mendelssohn recognizes no eternal truths other than those comprehensible to human reason and capable of being verified by it. With respect to Judaism, Mendelssohn points out that the Torah contains no commandment that says ‘Thou shalt believe…’. Rather than articles of faith, Judaism is based on revealed legislation. Although this legislation presupposes truths accessible to all humanity, the only thing incumbent on Jews alone is a system of behaviour.

Mendelssohn’s stress on the practical nature of Jewish spirituality bears a strong resemblance to the critical philosophy of Immanuel Kant. While Kant is often credited with the destruction of metaphysics, it would be more accurate to say that he attempts to move metaphysical ideas from a theoretical context to a practical one. Though reason cannot prove the existence of God, the immortality of the soul, or personal freedom, we must presuppose these foundational ideas if we are to make sense of our duties under the moral law. The result is that our idea of God is not a ‘demonstrated dogma’ but a necessary postulate. As Kant says in the introduction to the *Kritik der Reinen Vernunft (Critique of Pure Reason)*, his intention is to limit knowledge in order to make room for faith (*Glaube*).

Not surprisingly, Kant concludes that the only content we can ascribe to our idea of God is moral. Even if we could prove that a necessary Being exists, we could not get a satisfactory idea of divinity because the crux of that idea is the notion of a perfect or holy will. It could be said, therefore, that Kant’s view of God is much closer to the prophetic model of a judge and law-giver than it is to the idea of a first cause. This is important because scientific speculation is not indigenous to Judaism. The prophets sought God not by reasoning from effect to cause but by trying to awaken people’s conviction that the world could be better and that it is our obligation to make it so. In short, they sought God in ideas like commandment and repentance, both of which play a central role in Kant’s understanding of morality.

Kant argues for an idealized form of Christianity, which is to say a religion which recognizes all moral duties as divine commands. Looking beyond sectarian quarrels, he stresses the importance of trying to overcome the propensity of the human heart to do evil. Religion becomes spurious when people believe they can please God in ways that do nothing to improve behaviour, that simply wearing the right clothes or performing the right ritual is

more important than purifying one's motives. Kant's picture of a religion that rejects intolerance and dogmatic pronouncements could not help but appeal to Jews, even though Kant himself did not think much of Judaism as a religion. Yet Kant's attempt to formulate an idealized form of Christianity raised the question of whether the same could be done for Judaism.

2 S.D. Luzzatto (1800-65)

Luzzatto agrees with Mendelssohn that Judaism is not based on articles of faith but carries this conclusion much further. Sharply critical of the medieval philosophers who tried to justify the principles of Judaism with rational arguments, Luzzatto argues that metaphysics is impossible and trust in human reason, unfounded. In fact, the whole attempt to view Judaism as a source of truth about God is characteristic of what Luzzatto terms 'Atticism'. In his view, Greek ethics was an attempt to achieve happiness through wisdom. But experience shows that the Greek view of human behaviour is fundamentally flawed: intelligence alone will not yield lasting wellbeing or cure human ills. Luzzatto therefore proposes that we understand wisdom not as contemplation of abstract truths but as a condition in which people act with care and thoughtfulness, trusting in the saving grace of God (see Grace).

The positive side of Luzzatto's thought is his emphasis on compassion. All that is abiding in human life is the joy of performing acts of kindness and love, putting aside self-interest to help others. In Luzzatto's view (1880), these emotions, ignored by rationalist philosophers, permeate every aspect of the legislation of the Torah. In the days of Abraham, Isaac and Jacob, Luzzatto claims, religion was not enforced by a system of rewards and punishments. People understood the difference between right and wrong and invariably chose the former. Without proofs or arguments, they believed in God and worshipped in a natural, self-initiated fashion. It was only when the Israelites became a nation and began to mix with other nations that they needed the laws, rituals and educational imperatives revealed to Moses. Thus Judaism is autonomous before and heteronomous after Sinai.

Luzzatto's scepticism about human reason goes well beyond anything found in Mendelssohn and Kant. In fact, his emphasis on moral emotions and natural religion is suggestive of Rousseau. Although his influence was not nearly as great as Mendelssohn's, Luzzatto was one of the first Jewish thinkers to break with medieval Jewish rationalism.

3 Moritz Lazarus (1824-1903)

Closer in spirit to Mendelssohn and Kant is the thought of Moritz Lazarus. In a two volume work entitled *Die Ethik des Judentums (The Ethics of Judaism)* (1901), he argues that the Bible does not expound dogmatic teachings about the nature of God. Obedience to God is not based on omnipotence or omnipresence but on God's moral perfection. God commands an act because it is right; it is not right because God commands it. It is not God's command but its own moral necessity that makes a law binding. 'Law' and 'morality' here mean Kant's categorical imperative. In Lazarus' view, this principle underlies all the commandments, sayings and moral insights scattered through the Torah and Talmud.

Lazarus argues that Judaism is committed to the autonomy of the moral subject. When Kant says that morally we are subject only to a law of which we can regard ourselves as author, the point is not that we have created the law ourselves but that our moral self recognizes and accepts the validity of the law. God may be the one who gives the law, but it is the recognition of its validity that determines its moral character. Sanctification, as Lazarus says several times, is moralization. Not surprisingly, Lazarus goes to great length to stress the universal features of Jewish law. All people are created by God and formed in the image of God. Thus the Bible commands the Israelite to 'love thy neighbour as thyself' (Leviticus 19: 18) and to ensure that there is 'one law for you and for the stranger in your midst' (Leviticus 24: 22; Numbers 15: 15). The election of Israel has nothing to do with blood or moral superiority but with the fact that Israel was commanded to follow a universal moral law.

Book Two of *Die Ethik des Judentums* is mainly a study of the psychological foundations of Jewish ethics. The purpose of ethics is not to negate human nature and promote asceticism but to assign boundaries to the various instincts clamouring for satisfaction and to harmonize the opposing claims that arise in social intercourse. Lazarus maintains that Jewish teachings foster sympathy, charity and love of one's fellow human beings. His focus on the psychological foundation of Jewish ethics is both the strong point and the weak point of his system. The strong point, because he sheds considerable light on Jewish ritual: each practice either curbs an instinct or fosters a virtue

(see Ritual). The weak point, because Kantian ethics is supposed to be based on the objective necessity of the moral law, not on the feelings it engenders or the social harmony it promotes.

4 Solomon Ludwig Steinheim (1789-1866)

A very different reaction to Kantian philosophy can be found in the thought of Solomon Ludwig Steinheim, whose major work, *Die Offenbarung nach dem Lehrbegriffe der Synagoge (Revelation according to the Doctrine of the Synagogue)* (1835), is a critique of all attempts to justify religion with arguments derived from philosophy. The crux of Steinheim's argument is that it would be silly and superfluous for God to reveal a doctrine that we are capable of discovering on our own. Steinheim therefore rejects scholastic attempts to argue that reason and revelation are alternative ways of arriving at the same truth. This does not mean that the contents of revelation are secret, ineffable, or unfathomable. On the contrary, Steinheim insists that revelation offers us a doctrine that is both communicable and comprehensible. His point is that it must give us something new, something beyond what our reason, conscience, or feelings already provide.

By 'revelation', Steinheim means a historical event, not an evolutionary process or rational discovery (see Revelation). Although our understanding of revelation may change and grow over time, the contents do not. Revelation is needed because reason cannot free itself from antinomies regarding freedom and necessity. Reason conceives of the world according to necessary connections. Whether God is understood as the first member of a causal chain or as a being whose essence implies existence, it is still true that the God of reason must submit to the rule of necessity. According to Steinheim, this means that the God of reason is not a free and spontaneous agent who creates the world but a mechanical force that resides within it. The problem is that reason has no choice but to uphold the principle *ex nihilo nihil fit*, from nothing comes nothing. If this principle is true, every event must be preceded by a cause. It follows that reason cannot conceive of anything that is spontaneous, least of all a creation that is genuinely *ex nihilo*. The only creation it can subsume under causal laws is the ordering of an eternal, formless matter. But, in Steinheim's opinion, this sort of creation is no creation at all. So reason cannot account for the existence of the world in time.

What is true of creation is also true of human action. If every event must be preceded by a cause, individual moral responsibility is impossible. Like creation, the initiation of action requires spontaneity, a coming into being from nothing. Freedom is an undeniable fact, the basis of all ethical judgments. But without the teaching of revelation, this fact also would be simply an event without a cause.

In Steinheim's view, we have no choice but to admit the superiority of revelation: that is, a creating God, human agents made in the image of God, and the immortality of the soul. Although Steinheim is sometimes read in light of Kierkegaard, the comparison is wide of the mark. We do not give up reason in order to embrace absurdity but go beyond reason in order to avoid determinism. Note, however, that Steinheim's view of the contents of revelation corresponds to Kant's idea of the postulates of pure practical reason. So while Steinheim may be right in saying that reason, understood as a faculty that moves from cause to effect, cannot account for spontaneity, it does not follow that we have no choice but to affirm revelation as a historical fact. The Kantian retort is that reason also has a practical employment compatible with freedom.

5 Solomon Formstecher (1808-89)

If the thought of Lazarus and Steinheim is a reflection on Kant, that of Formstecher is a reflection on Hegel and Schelling. In *Die Religion des Geistes (The Religion of Spirit)* (1841), Formstecher emphasizes that Judaism is based not on the apprehension of timeless truths given to Moses but on an evolving conception of spirit that achieves greater clarity over time.

Formstecher's philosophy begins with the idea of a world soul (*pneumatikon*) that manifests itself either as spirit or as nature. The world soul is a unity that underlies the world and exists independently in the sense that the world soul could exist without the world but not vice versa. Formstecher insists, however, that the world is not created by a necessary causal process but by free choice. Put otherwise, God, the bearer of the world soul, is a free, self-conscious, self-determining entity, separate from the world but responsible for its existence. Formstecher thus rejects any form of pantheism. Of the two spheres of reality, only spirit is conscious both of itself and of nature. Thus the world soul can be conceived only under the guise of spirit.

The duality of nature and spirit gives rise to two forms of religion: one based on necessity, the other on freedom. Formstecher identifies the first type with paganism; the second, with Judaism. Paganism deifies nature by worshipping objects within it or identifying God with the forces that control it. The outgrowth of paganism is detached contemplation of the beauty of nature and the attempt to decipher its secrets. But a religion of spirit stresses freedom over necessity and goodness over beauty. It is here that knowledge becomes more than awareness of an external object and attains self-consciousness in the true sense. The essence of Judaism is ethical monotheism: God stands apart from nature, and our purpose in life is not to achieve unity with God but to emulate God's ways by striving for the moral ideal.

Formstecher treats revelation not as a single event but as the process by which spirit becomes aware of itself. A liberal, he looks forward to the day when Judaism will de-emphasize ritual and stake its claim on greater understanding of the process of spiritual development. In Formstecher's view, Christianity is an attempt to bring the religion of spirit to the nations of the world. But where Hegel saw the doctrine of incarnation as an attempt to bridge the gap between an infinite creator and finite creation, Formstecher sees it as a concession to paganism.

6 Samuel Hirsch (1815-89)

Like Formstecher, Hirsch interpreted the history of religion in terms of the historical development of spirit. His major work, *Die Religionsphilosophie der Juden (The Religious Philosophy of the Jews)* (1842), argues that while European society has begun to tolerate Jews, it has failed to recognize the unique contribution of Judaism to world culture. Thus Kant did not regard Judaism as a religion at all, and Hegel thought of it as an inferior religion superseded by Christianity.

At the root of Hirsch's philosophy is the act of self-consciousness or recognition of the self as an 'I'. It is in this act that we separate ourselves from nature and come to experience freedom. But, being abstract, this freedom is empty. Religious life is the process by which true freedom is realized. For freedom to become concrete, it must acquire content by facing the choice between real freedom, which implies virtue, and illusory freedom, which implies sin (see Freedom and liberty).

According to Hirsch, freedom is experienced as a gift, so there must be a divine dispenser of freedom who is absolutely free. Humans are the image of God, and history is the human quest for perfect freedom. Hirsch took strong exception to the Pauline doctrine of original sin, on the grounds that it compromises the idea of freedom. I am not guilty of sin until I exercise my own free choice (see Sin). Like Formstecher, Hirsch sees Christianity as an attempt to bring the truth about God to the nations of the world. But, unlike Formstecher, he does not argue for the inherent superiority of Judaism. In his view, both religions are evolving and will reach their perfect or absolute state in the days of the Messiah. With respect to Judaism, Hirsch played a decisive role in the founding of the Reform movement.

7 Conclusion

Despite their differences, all the philosophers discussed in this article tried to integrate Judaism into Western culture or what Formstecher called 'the universal religion of civilized people'. All stress the importance of human freedom, the need to undertake moral action, and the existence of a transcendent God. As the nineteenth century drew to a close, the influence of Hegel began to wane and rationalist thinkers returned to Kant. After the First World War, the primary debate in Jewish philosophy became that between Kantians like Hermann Cohen and existentialists like Martin Buber and Franz Rosenzweig.

See also: Jewish philosophy, contemporary; Kant, I.; Mendelssohn, M.

KENNETH SEESKIN

References and further reading

Cohen, H. (1919) *Die Religion der Vernunft aus den Quellen des Judentums*, Leipzig: Fock; trans. S. Kaplan, *Religion of Reason out of the Sources of Judaism*, New York: Frederick Ungar, 1972.(Cohen was the greatest of the Jewish Neo-Kantians. His work can be viewed as the culmination of nineteenth-century Jewish philosophy or the beginning of the twentieth.)

Formstecher, S. (1841) *Die Religion des Geistes (The Religion of Spirit)*, Frankfurt: Hermann'sche Buchhandlung.(Emphasizes the idea of spirit and historical evolution. Although clearly in the Hegelian

tradition, Formstecher still believed in a God who is separate from the world but responsible for its existence.)
Guttmann, J. (1933) *Die Philosophie des Judentums*; trans. D.W. Silverman, *Philosophies of Judaism*, New York: Schocken, 1973.(The best one-volume history of Jewish philosophy available. Keep in mind, however, that Guttman was sympathetic to Kant and Cohen.)
Hirsch, S. (1842) *Die Religionsphilosophie der Juden (The Religious Philosophy of the Jews)*, Leipzig: Hunger. (Like Formstecher, Hirsch was influenced by Hegel. Also like Formstecher, he emphasized the idea of freedom and believed in a God who is separate from the world.)
Lazarus, M. (1901) *Die Ethik des Judentums*, Frankfurt: J. Kauffman; trans. H. Szold, *The Ethics of Judaism*, Philadelphia, PA: Jewish Publication Society of America, 1900.(Lazarus was one of the first of the Jewish Neo-Kantians, but his work was overshadowed by that of Cohen.)
Luzzatto, S.D. (1880) *Yesodei ha-Torah*; trans. N.H. Rosenbloom, ‘The Foundation of the Torah’, in N.H. Rosenbloom (ed.), *Luzzatto’s Ethico-Psychological Interpretation of Judaism*, New York: Yeshivah University, 1965.(Luzzatto was one of the first people to break with scholasticism, particularly the thought of Maimonides.)
Mendelssohn, M. (1783) *Jerusalem, oder über religiöse Macht und Judentum (Jerusalem, or on Religious Power and Judaism)*, trans. with notes by A. Arkush, *Jerusalem*, Hanover, NH: University Press of New England, 1983.(A religious and rhetorical masterpiece. This is the book that launched modern Jewish philosophy and made pioneering contributions to the theory of the separation of Church and state.)
Rotenstreich, N. (1968) *Jewish Philosophy in Modern Times*, New York: Holt, Rinehart & Winston.(A critical study of Jewish philosophy from Mendelssohn to Rosenzweig.)
Schwarzschild, S. (1990) *The Pursuit of the Ideal*, ed. M. Kellner, Albany, NY: State University of New York Press.(A collection of distinguished essays written by a modern philosopher in the Kantian tradition.)
Steinheim, S.L. (1835) *Die Offenbarung nach dem Lehrbegriffe der Synagoge (Revelation according to the Doctrine of the Synagogue)*; trans. J.O. Haberman, ‘The Revelation according to the Doctrine of Judaism: A Criterion’, in J.O. Haberman (ed.) *Philosopher of Revelation: The Life and Thought of S.L. Steinheim*, Philadelphia, PA: Jewish Publication Society of America, 1990.(An excellent translation and commentary on Steinheim’s work. Steinheim emphasized the superiority of revelation to reason in order to preserve the idea of human freedom.)

Jewish philosophy, contemporary

Jewish philosophy is pursued by committed Jews seeking to understand Judaism and the world in one another's light. In this broad sense, contemporary Jewish philosophy maintains the central focus of classical, medieval and Enlightenment Jewish philosophy. But a certain kind of traditionalism distinguishes many contemporary Jewish philosophers from their predecessors: an effort to show how Judaism maintains continuity and coherence despite historical change. Jewish thinkers who are traditionalists in this sense are no longer preoccupied with showing non-Jewish philosophers how Judaism fares when evaluated by universal reason, as their classical, medieval and modern predecessors were. Nor is their chief concern with exhibiting the good reasons for remaining Jewish and not converting to Christianity or Islam, as was that of many earlier Jewish thinkers. One work that sets an agenda for many of these traditionalists is Franz Rosenzweig's Der Stern der Erlösung (The Star of Redemption) (1921). Like Rosenzweig, (1) they often reject the Enlightenment demand for a transcendental propaedeutic as a prelude to asking substantive questions. Instead, they address Jewish thought, ethics and experience head on. (2) None among this group thinks of his work as beholden primarily and inevitably to standards of thought articulated first and foremost outside distinctively Jewish experience. (3) The six points of Der Stern der Erlösung (The Star of Redemption) - Creation, Revelation, Redemption, God, Israel, and the World - mark the large themes they aim to define or the categories through which they propose, explore and defend their claims. Besides traditionalism thus understood, contemporary Jewish philosophy, particularly among philosophers with analytic training, is marked by efforts philosophically to reanimate the classic texts of medieval Jewish philosophy, especially the work of Moses Maimonides.

1 Classical and medieval background

What counts as contemporary Jewish philosophy? Posing the question historically is virtually inevitable, because the discipline recurrently contends with its own past. Locating Jewish philosophy in the context of classical and medieval thought seems relatively easy, especially with the aid of hindsight: We might describe such philosophy as a reasoned and disciplined investigation by committed Jews of humankind-in-God's-world, or as their search 'for the deepest truth the world has to offer' (Novak 1995: 6). Typically, the philosophers engage extensively with traditional Jewish texts, especially the Hebrew Scriptures and Talmud. Philo of Alexandria, Al-Fayyumi Saadiah Gaon, Judah Halevi and Moses Maimonides are salient examples.

For most classical and medieval Jewish philosophers, the confluence of Jewish with non-Jewish sources - for example, Plato and Aristotle - is decisive. Thus Jewish philosophy is syncretistic all the way down. With the destruction of an independent Jewish state and the emergence of the cosmopolitan imperial cultures in which the key figures of classical and medieval Jewish philosophy typically wrote, a central question motivating many Jewish philosophers well versed in the philosophical canon of the cosmopolitan culture was this: how can learned and reasoning people embrace the claims that Jewish texts seem to make? A tension was all but inevitable between views rooted in Jewish history and tradition and what non-Jewish philosophers typically saw and presented as universal truths. Are Jewish claims about humankind-in-the-world compatible with, even deducible from, the principles and claims of universal reason, as articulated, say, in the Graeco-Roman philosophical tradition?

2 Modern background

Classical and medieval Jewish philosophers, living in cosmopolitan Alexandria, in Muslim-governed countries, or the ghettoes of Catholic Europe, saw themselves as free to pursue their basic questions in ways that allowed them to live, act and think openly as Jews. Their concerns over possible tensions between particularity and universality persist into the modern period. But the writing of many modern thinkers takes a new turn, under the threat of a Protestant theocracy, as in the case of Spinoza, or under what might be called the tyranny of liberal Protestant civility, as with Mendelssohn, Cohen, Rosenzweig, Buber and others.

Responding to his intellectual and spiritual environment, Spinoza treated both Judaism and Christianity as significant, but human, cultural constructions with no privileged authority *vis-à-vis* the universe. He could thus lay conceptual foundations for a liberal republic welcoming to a plurality of reasonable religions that would benefit the public welfare by inculcating moral constraints and, as Jefferson would later put it, 'neither pick [our] pocket[s] nor break [our] leg[s]'.

Much influenced by Spinoza, Mendelssohn (1783) shared a liberal, republican agenda with many of the Protestant thinkers of his day. But he could not share a vision of grace and law that is so heavily laden as theirs often is with ancient anti-Hebraic grudges and disparaging misreadings of Judaism. He had to rebuff his Protestant admirers when they urged him, in the name of his own philosophy, to abandon Judaism for Christianity. He demanded separation of church and state and dismissed claims to political control over the human mind and spirit. Jewish life and thought, he argued, are compatible with universal reason and at least as conducive to liberal republican polity and culture as Christianity. But they depend necessarily on freedom of conscience.

Cohen went a step further. In *Die Religion der Vernunft aus den Quellen des Judentums (Religion of Reason out of the Sources of Judaism)* (1919) he argues that Judaism, more than other religious options, offers the morality, piety and spirituality that a thinking, liberal person requires. Buber's *Ich und Du (I and Thou)* (1923) similarly celebrates the Jewish covenant as central to the universal values of human life at its best. He actually dismisses any vestige of Jewish law, so as to render Judaism a paradigm of the 'grace-ful' existential attitude that is conducive to a liberal democratic community, open to divinity - the ultimate Thou - and capable of thriving among the nations (see Buber, M.).

Franz Rosenzweig was at first attracted to liberal Protestantism but ultimately rejected its blandishments. He turned the tables on Buber's Jewish existentialism, arguing for a religious philosophy that renders any and every humanism, including every kind of liberal or socialist doctrine, dependent on a prior theology. The core of this theology is the conception of God's word, embodied eternally, outside the historical life of the nations, by the Jewish people, even as it is embodied historically in the Christian Church(es) and their militant quest for peace and understanding among the nations.

Rosenzweig presents himself as a *ba'al teshuvah*, a penitent returning to Judaism from the theologically liberal tradition that runs from Spinoza through Mendelssohn to his own teacher, Hermann Cohen. Rosenzweig's stance as a returning Jew is crucial in understanding contemporary Jewish philosophy. At least for many paradigm cases, a similar attitude of return sets the tone and orients much of the programme - certainly for Fackenheim (1973, 1982), Levinas (1969, 1990), Novak (1995) and Wyschogrod (1983). How is this so?

3 Characteristics of contemporary Jewish philosophy

First, philosophers as diverse as Fackenheim, Hartman, Leibowitz, Levinas, Soloveitchik and Wyschogrod have learned the lesson Rosenzweig taught: that it is possible to drop the Kantian (or Lockean) transcendentalist demand for a philosophical propaedeutic before first-order questions can be asked. None of them takes the transcendental turn before turning to traditional religious issues - as Rosenzweig himself still did in Part One of *Der Stern der Erlösung* (1921), seeking to refute, not simply reject, transcendentalism. None of these figures asks such Lockean or Kantian questions as 'What are the conditions for the possibility of thought, language, knowledge, science, ethics, or experience' in general, or of religious thought, language, knowledge, ethics, or experience in particular (see Locke, J.; Kant, I.)? Rather, each investigates Jewish thought, ethics and experience head on, asking in particular the following questions. How good are these things? How can they help us shape our lives? What resources do they give us to conceptualize problems of meaning that threaten our human capacity to endure or triumph, physically, cognitively, morally, in the face and wake of suffering, absurdity and evil? How do they fare when compared with challengers or competitors, opposers, or nay-sayers? Perhaps most often addressed is this question: how did Enlightenment philosophy thin out the rich tradition of Judaism that once gave Jews conceptual and practical tools for understanding and acquitting themselves with grace, mercy and justice in God's vexed and vexing world? Even more significantly, how can Jewish thought, ethics, and experience be restored to their cultural, social and personal thickness, their traditional complexity, profundity and open-ended creative dynamic?

These preoccupations mark many contemporary Jewish philosophers as traditionalists in the sense that they are seeking, as William James might have put it, 'a minimum of jolt and a maximum of continuity' in Jewish ways of thinking and living. As David Novak puts it in his own case (1995), they embrace a traditionalism that tends to avoid both historicism and fundamentalism, two orientations that had already rejected the Enlightenment project. Historicism, as they see it, undercuts the coherence of Jewish tradition by rendering it into discrete, ultimately incommensurable epochal, or time-sliced, cultures. Fundamentalism, by contrast, assumes too much invariance in

the tradition. It transforms Jewish continuity and coherence into a monolith and blunts its vital and vibrant diversity - just as what we might call Enlightenment fundamentalism assumed too much invariance in rationality.

Historicists are left with many Judaisms but no tradition. Fundamentalists are left with a purportedly essential Jewish tradition that lacks historical dynamic or nuance. Traditionalists, seeking a way between this Scylla and Charybdis, contend that the task for a living Jewish philosophy is to discover how Judaism maintains continuity and coherence through all the vicissitudes it has suffered and all the changes it has undergone - not merely despite them. This approach, they claim, reflects the open-textured character of Judaism; it also permits a full-fledged philosophy of Judaism and clarifies Judaism's contributions to philosophy at large.

If this characterization of such thinkers as Fackenheim, Hartman, Leibowitz, Novak, Soloveitchik and Wyschogrod makes sense, a second Rosenzweigian characteristic of their work readily comes into view: None among these contemporary traditionalists thinks of his work as beholden, primarily or inevitably, to standards of thought and action articulated first and foremost outside distinctively Jewish experience. To be sure, like Rosenzweig, each learns from mentors and gleans arguments where or when he finds them; and, like him, some, at least, will be much concerned with relations between Judaism and Christianity - although none will adopt Rosenzweig's duplex theology, dividing the work of true religion between Jewish spirituality and Christian worldliness. There is no sense among them that Jewish ideas, as such, must be checked for their validity, as a matter of sound principle, against any purportedly universal philosophical standard, whether Greek, Cartesian, Kantian, Hegelian, or existentialist. They abandon any variety of Enlightenment fundamentalism. This marks a departure from the practice of Jewish philosophers from Philo to Buber. Indeed, much of the Jewish philosophical writing of these contemporary traditionalists reflects on and continues the work of Jewish sources, often using commentary with a marked family resemblance to Jewish midrashic and *aggadic* discourse.

Third, as we have suggested, for these traditionalists, Rosenzweig's *Der Stern der Erlösung* outlines the basic points of concern. The six points of this work - (1) Creation, (2) a Revealed Covenant for an elect People, (3) the promise of Redemption, (4) God as the Creator and Revealer of Covenant, (5) Israel as God's Elect and (6) the World as the beneficiary of God's Promise of Redemption - mark out the ground they aim to defend and provide the categories through which they explore, discover and urge their claims.

Fourth, and linked with this agenda, if this group shares *bêtes noires*, they are thinkers who despoil or reject any of the above six points. Thus, perhaps no *bête* is more *noire* than Spinoza (see Fackenheim 1982; Hartman 1985; Levinas 1969; Novak 1995). Spinoza here is the great and baleful alternative to Rosenzweig's return. He is the classic renegade who abandons Judaism and rejects its deepest truths. And he is the father of academic anti-Hebraism, the watershed figure who teaches the likes of Kant and Hegel to portray Judaism as spiritually crippled, fanatically nationalistic, and superstitiously, slavishly loyal to 'the Law'.

For the contemporary traditionalists, Spinoza was the first and perhaps most challenging philosopher to attempt a complete dismantling of transcendence; he directly attacks all six points of *Der Stern der Erlösung*. Where Judaism contends that we live in a created world, Spinoza argued that the idea of creation explains nothing. Where Judaism contends that the world is governed by a God who especially cares for humankind and reveals that care in the covenant by which he elects the people Israel, Spinoza maintained that the world is explainable through natural principles that relate only randomly to human wish and will (see Fackenheim 1982; Hartman 1985). For Spinoza, although everything is grounded in divine power, God has no particular role in establishing the Hebrew covenant. The Hebrews legislated that covenant for their own local purposes, in essence electing God (see Novak 1995). Where Judaism finds in God's world profound reasons for hope or the expectation of redemption, since God, in cooperation with Israel and all humankind, works to realize, in due time, the good life and the establishment of his kingdom, Spinoza seems to picture humankind as an incident within and among the many incidents of a self-perpetuating and unending process, a motiveless but all-encompassing system of physical or natural cause and effect (see Fackenheim 1982; Hartman 1985; Levinas 1969; Soloveitchik 1944).

Finally, like Rosenzweig, the modern traditionalists turn the tables on the mainly Protestant liberal philosophical theologies that emerged in the intellectual and cultural wakes made by Kant and Hegel: Paradigmatically, they reject the anthropologizing of religion they find apparent in Kant's *Die Religion innerhalb der Grenzen der blossen Vernunft (Religion within the Limits of Reason Alone)* (see Kant, I. §14) (see Fackenheim 1973; Hartman 1985; Leibowitz 1992; Novak 1985; Soloveitchik 1944; Wyschogrod 1983). Parallel with neo-orthodox Protestant

thinkers like Karl Barth, they are confident of reaching perfectly reasonable religious decisions and conclusions out of unabashedly covenantal sources and theocentric lines of thinking (for parallels with Barth, see Novak 1985 and Wyschogrod 1983). As part of this aspect of their 'return', they reject Kant's insistence on the self-legislation - and in that sense, autonomy - of ethics, attempting to restore what might be described as Micah's view, that justice and mercy towards the neighbour and stranger are grounded in the humble love of God and are virtually inseparable from it (see Fackenheim 1973, 1982; Levinas 1969).

Naturally, they find in Kant's academic anti-Hebraism an unappreciative and distorted reading of Jewish sources that reduces the mutuality of the covenant to tyrannical legalism and that caricatures the *mitzvot* as onerous and alienating obedience (see Fackenheim 1973; Soloveitchik 1944). They reject the Protestant/Hegelian triumphalist history-of-religions narratives at work in liberal theology, the Hegelian presumption picked up by such writers as Otto, Wach, and Eliade, that irreversible, unidirectional progress from religious particularism to religious universalism (of a decisively Christian incarnational kind) is either somehow historically inevitable or normatively optimal for genuinely religious people or peoples (again, see Fackenheim 1973, but also Soloveitchik 1944).

The contemporary traditionalists barely bother to contend any longer with Enlightenment attacks on transcendence or demands for epistemological or metaphysical propaedeutics. The theological liberalisms that emerged out of the German Enlightenment, which reduced full-bodied - and fully bodied - Jewish tradition to some thin, vague, relatively nondescript and typically secularized confession are hardly engaged, even as straw men. The single issue emerging from the German Enlightenment that still divides traditionalists is that of religious pluralism (see Religious pluralism). Some, like Fackenheim, Hartman and Novak, construe an appreciation for other religious traditions as a kind of moral, social, or intellectual requirement. Others, like Soloveitchik, find such appreciations tangential to *halakhic* life. Still others, like Leibowitz, find such pluralism actually beside the point, covenantally speaking.

Most exponents of contemporary traditionalism find four episodes in twentieth-century Jewish history decisive in the emergence of their outlook: the Nazi destruction of European Jewry, the wartime abandonment of the Jews by the North Atlantic democracies, the re-establishment of the State of Israel in the wake of the Second World War, and the disaffection of increasing numbers of American Jews from traditional Jewish life, in response to the attractions of secular affluence and assimilation. These events locate Jewish philosophy socially and culturally in a context vastly different from that of its classical, medieval, or Enlightenment counterparts. They seem to render almost irrelevant the particularist/universalist, faith/reason problematics of the medievals and the more distinctively modern challenges of liberal Christian civility. It is widely supposed that if the First World War did not render the German Enlightenment tradition quaint, then surely the Third Reich's Final Solution and the moral abdication of the North Atlantic republics exposed in its implementation a depth of hypocrisy that soured the entire enterprise (see Holocaust, the).

Liberal, secular and assimilationist paths seem to many clearly to spell the extinction of the ancient tradition. And many, especially in the United States, suppose that cultural Zionism, or a diaspora sensibility centred on Israel can alone render appealing the commitments of Jewish traditional life (see Zionism). But these assumptions are highly debatable, if rarely argued pro or con with much rigour. In fact, Jewish philosophers, like virtually all other philosophers as the century ends, have barely begun to come to grips with the profound events affecting Jewish life throughout the twentieth century.

4 Contemporary problems and prospects

Many characteristics of contemporary traditionalism reflect the course philosophy itself has taken, especially in western Europe, notably, with the emergence of existential, phenomenological, hermeneutic, Marxist, feminist and other cultural critiques of the Enlightenment. The works of Fackenheim, Levinas, Soloveitchik, Wyschogrod and others must be read in this context. So too must Israelis like Halbertal and Margalit 1992), who address traditional Jewish concepts like idolatry and worship in analytically and phenomenologically rigorous ways. Here, too, we have philosophers abandoning the Enlightenment thinning of tradition and reduction of religion to, say, the presumptive categories of science (error and accuracy), ethics (duties and inclinations), and aesthetics (appropriateness and inappropriateness). For many of the new traditionalists, authenticity means giving a full-blooded religious culture an honest and appreciative hearing in something like its own terms.

Paralleling the Continental developments, English-speaking philosophers, especially Wittgensteinians, and analytic pragmatists like Quine, Sellars, Davidson, Putnam and Rorty, have subjected basic Enlightenment assumptions (such as that about 'the given' or 'the very idea of a conceptual scheme') and distinctions (such as those between analytic and synthetic judgments, or between things and phenomena) to severe criticism. Many of the English-speaking philosophers have themselves essayed what might be construed as forms of traditionalism. But no contemporary Jewish philosopher has attempted to construct a Wittgensteinian philosophy of Judaism, and virtually none has come to grips with the pragmatists - despite Harry Austryn Wolfson's pronouncement (1912) that the tangles between the Halevian and Maimonidean approaches in philosophy prefigure what Wolfson considered the major philosophical debate of his time (which surely overlaps with ours): the debate between pragmatism and eternalism - or what we might now call realism.

Among English-speaking religious philosophers, the vacuum left by the collapse of modern Enlightenment projects has been filled largely by analytically trained thinkers, many of whom study classical, medieval, or scholastic texts and issues. Here an asymmetry must be noted between the works of Christian and those of Jewish philosophers. Christian philosophers like Alvin Plantinga, William Alston, or James Ross engage in historical studies of Christian scholasticism (broadly construed) but go on to converse about various issues in the wider Christian intellectual community, and finally to contribute variations on scholastic Christian arguments to current epistemological, semantic and metaphysical debate. Analogous Jewish philosophers, by and large, have not followed suit. Unlike their traditionalist Jewish peers, most have focused on reorienting and reanimating the history of medieval Jewish philosophy itself, especially as represented in the work of Moses Maimonides, treating his thought with the philosophical sophistication it deserves and opening up avenues for renewed philosophical engagement with it. Such engagement may both reflect and stimulate a constructive turn in Jewish philosophy that is more appreciative of the role of (a restrained, chastened, or more epistemically catholic) reason in religious thought and more sanguine about the light it can shed on Judaism and on the world - witness the recent ambitious work by Lenn Goodman (1991, 1996). Still it remains true that few have yet sought to help revitalize or heighten the tenor of intellectual concern and philosophical conversation in the general Jewish community; and few have yet contributed arguments to current epistemological, semantic, or metaphysical debate out of distinctively Jewish sources or commitments. How this asymmetry between Jewish and Christian philosophy is to be explained - by the negligence or diffidence of the potential practitioners or their marginalization by the intellectual community - and whether it will disappear in years to come are questions that remain to be answered.

See also: Enlightenment, continental; Enlightenment, Jewish; Jewish philosophy in the early 19th century; Maimonides, M.

HENRY S. LEVINSON
JONATHAN W. MALINO

References and further reading

Buber, M. (1923) *Ich und Du*, Leipzig: Insel Verlag; trans. R.G. Smith, *I and Thou*, New York: Charles Scribner's Sons, 2nd edn, 1958; trans. W. Kaufman, *I and Thou*, New York: Charles Scribner's Sons, 1970. (The classic exposition of Buber's dialogical philosophy.)

Cohen, H. (1919) *Die Religion der Vernunft aus den Quellen des Judentums*, Leipzig: Fock; trans. S. Kaplan with an introduction by L. Strauss, *Religion of Reason out of the Sources of Judaism*, New York: Frederick Ungar, 1972.(A systematic exposition of Judaism by the founder of German Neo-Kantianism.)

Cohen, A. and Mendes-Flohr, P. (eds) (1987) *Contemporary Jewish Religious Thought*, New York: Charles Scribner's Sons.(Original essays on critical concepts, movements and beliefs of contemporary Jewish thinkers.)

Fackenheim, E. (1973) *Encounters between Judaism and Modern Philosophy*, New York: Basic Books.(Critical consideration of modern philosphers from Hume through Heidgger in terms of Fackenheim's own construal of Judaism.)

Fackenheim, E. (1982) *To Mend the World*, New York: Schocken.(A neo-Hegelian philosophy of Judaism which emerges out of criticism of Spinoza, modern empiricism, Kant, Hegel, Sartre and Heidegger.)

Goodman, L.E. (1991) *On Justice: An Essay in Jewish Philosophy*, New Haven, CT: Yale University Press.(An ontological theory of justice based on the biblical idea of deserts.)

Goodman, L.E. (1996) *God of Abraham*, New York: Oxford University Press.(Reflections on the nexus between natural theology and the values represented in the Jewish tradition.)

Halbertal, M. and Margalit, A. (1992) *Idolatry*, Cambridge, MA: Harvard University Press.(Philosophically

sophisticated conceptual and historical account of Jewish understandings of idolatry.)
Hartman, D. (1985) *A Living Covenant*, New York: Free Press.(A neo-Maimonidean defence of Jewish covenantalism in dialogue with Soloveitchik and Leibowitz.)
Leibowitz, Y. (1992) *Judaism, Human Values, and the Jewish State*, Cambridge, MA: Harvard University Press. (Collection of provocative occasional pieces on Jewish religion and politics.)
Levinas, E. (1969) *Totalité et infini*, The Hague: Martinus Nijhoff; trans. A. Lingis, *Totality and Infinity*, Pittsburgh, PA: Duquesne University Press, 1969.(The classic expression Levinas' views of ethics, rather than of metaphysics, as first philosophy.)
Levinas, E. (1990) *Nine Talmudic Readings*, trans. A. Aronowicz, Bloomington, IN: Indiana University Press. (Seminal works by the leading French Jewish phenomenologist, concerned to replace ontology with ethics.)
Mendelssohn, M. (1783) *Jerusalem, oder über religiöse Macht und Judentum (Jerusalem, or on Religious Power and Judaism)*, trans. with notes by A. Arkush, *Jerusalem*, Hanover, NH: University Press of New England, 1983.(Inaugural work of German Enlightenment Jewish philosophy.)
Novak, D. (1995) *The Election of Israel*, Cambridge: Cambridge University Press.(Systematic effort to retrieve philosophically the biblical notion of election.)
Pines, S. and Yovel, Y. (eds) (1986) *Maimonides and Philosophy*, Dordrecht: Martinus Nijhoff.(Collection of philosophical essays on Moses Maimonides.)
Rosenzweig, F. (1921) *Der Stern der Erlösung*, Frankfurt: Kauffmann; trans. W.W. Hallo, *The Star of Redemption*, Boston, MA: Beacon Press, 1970.(The highly influential twentieth-century philosophy of Judaism opposing all Enlightenment Jewish philosophy.)
Seeskin, K. (1991) 'Jewish Philosophy in the 1980's', *Modern Judaism* 11: 157-72.(A helpful summary and enumeration of recent work on Jewish Philosophy.)
Soloveitchik, J. (1944) *Ish ha-Halakhah*; trans. L. Kaplan, *Halakhic Man*, Philadelphia, PA: Jewish Publications Society, 1983.(An existentialist presentation of *halakhic* Judaism.)
Spinoza, B. (1985) *Collected Works*; trans. E. Curley, Princeton, NJ: Princeton University Press, 2 vols.(Includes Spinoza's celebrated naturalistic account of salvation as the intellectual love of God, the inaugural work in modern biblical criticism, and an early defence of political and religious liberalism.)
Wolfson, H.A. (1912) 'Maimonides and Halevi', in I. Twersky and G.H. Williams (eds) *Studies in the History of Philosophy and Religion*, vol. 2, Cambridge, MA: Harvard University Press, 1977.(An exposition of Maimonidean and Halevian attitudes towards Greek philosophy with allusions to contemporary debates in philosophy.)
Wyschogrod, M. (1983) *The Body of Faith*, New York: Harper & Row.(Exceptionally well-written account of the relationship between philosophy and Judaism; includes a spirited defence of the election of Israel and an attack on liberal tendencies to spiritualize Judaism.)

Jhering, Rudolf von (1818-92)

Jhering saw law as a mechanism for achieving current purposes, supplying the compulsion needed where other levers were insufficient to secure the conditions of social life. In this, he took issue with those other nineteenth-century German jurists who regarded law as a settled hierarchy of rules and concepts. He had considerable influence on the development of jurisprudential thought both in Germany and in the USA.

Born in Aurich, a lawyer by education, Jhering taught at various universities, including Basle and Vienna, spending sixteen years at Giessen and twenty at Göttingen, where he died in 1892. He was a gifted lecturer, a prolific and often entertaining writer, a robust debater, an enthusiastic supporter of Bismarck, a once-failed parliamentary candidate and a twice-married family man who enjoyed social life.

Jhering's first research was in Roman law, which remained a lifelong interest. Soon, however, he was concentrating on its evolutionary and comparative aspects, publishing, from 1852-8, several volumes on *Geist des römischen Rechts (The Spirit of Roman Law)*. The work was never completed, as he turned to the development of his own philosophical approach.

This was largely a reaction to the views of two allied and influential schools of German thought. One saw law as something that grew unconsciously, reflecting the spirit of a people (see Jurisprudence, historical; Savigny, F.K. von). The other treated it as a settled hierarchy of rules and concepts, the terms used being sufficient to resolve any problem that might arise. Jhering also took issue with English utilitarianism. Though espousing the notion that law has a purpose and certain of their ideas about human motivation (Jhering expresses these in ways also reminiscent of Adam Smith), he rejected Bentham and Mill's individualism.

Jhering's own views are found in *The Struggle for Law* (1872), arguing that law develops as the result of a 'conscious struggle for claims', and in *Law as a Means to an End* (1877-83). Also unfinished, this presented law as directed to the fulfilling of human interests or purposes, which involve securing 'not merely the conditions of physical existence, but all those goods and pleasures which in the judgment of the subject give life its true value' (1877-83: 331). This requires cooperation, individuals connecting their own interests with those of others and 'the converging of all interests at the same point'. Certain social levers are needed to bring this about. Where duty, love or economic reward prove insufficient, the law steps in to provide the compulsion required.

The connections Jhering made between law, social utility and interests were influential for, though not entirely endorsed in, US thought (see Pound, R.) and for the development of *Interessenjurisprudenz* (the jurisprudence of interests) in Germany, while his analysis of legal coercion seems to foreshadow the ideas of Hans Kelsen.

Some sociologists, however, maintain that his statist and coercive conception of law inadequately reflects social facts. He is also criticized for endorsing such a conception, for taking relativistic views of social utility and justice, and for subordinating individual to social interests. On another understanding, he was actually arguing that it is only possible to secure individual interests in a 'realized partnership of the individual and society'.

See also: Law, philosophy of; Legal idealism §3; Roman law/civilian tradition §3

ELSPETH ATTWOOLL

List of works

Jhering, R. von (1852-8) *Geist des römischen Rechts (The Spirit of Roman Law)*, Leipzig: Breitkopf und Härtel; repr. Aalen: Scientia Verlag, 1968.(Usually referred to as 'The Spirit of Roman Law', there is no English version but it has been translated into various other languages.)

Jhering, R. von (1869) *Über den Grund des Besitzesschutzes (On the Reason for the Protection of Possession)*, Jena: Mauke; repr. Aalen: Scientia Verlag, 1968.(Analysing Roman law on possession.)

Jhering, R. von (1870) *Die Jurisprudenz des tägliches Lebens*, Jena: H. Dufft; trans. H. Goudy, *Law in Daily Life*, Oxford: Clarendon Press, 1904; repr. Littleton, CO: F.B. Rothman, 1985.(A collection of questions and cases for discussion.)

Jhering, R. von (1872) *Der Kampf um's Recht*, Vienna: G.J. Manz; trans. from 5th edn by J.J. Lalor, *The Struggle for Law*, 2nd edn with intro. by A. Koucourek, Chicago, IL: Callaghan and Co., 1915.(Regarded as one of his

two main jurisprudential works.)

Jhering, R. von (1877-83) *Der Zweck im Recht*, Leipzig: Breitkopf und Härtel; trans. I. Husik, *Law as a Means to an End*, New York: Macmillan, 1924.(Regarded as the second of his two major jurisprudential works.)

Jhering, R. von (1879) *Vermischte Schriften juristischen Inhalts (Miscellaneous Writings with a Legal Content)*, Leipzig: Breitkopf und Härtel; repr. Aalen: Scientia Verlag, 1968.(Includes his 1842 doctoral dissertation in Latin on the possession of an inheritance.)

Jhering, R. von (1882) *Das Trinkgeld (On Gratuities)*, Braunschweig: G. Westermann.(In which he argues that gratuities are a form of organized begging.)

Jhering, R. von (1885) *Scherz und Ernst in der Jurisprudenz (Jest and Earnest in Jurisprudence)*, Leipzig: Breitkopf und Härtel, repr. Darmstadt: Wissenschaftliche Buchges, 1992. (Collection of essays and articles of which the most famous is 'Im juristischen Begriffshimmel: ein Phantasiebild', for which see trans. C.L. Levy, 'In the Heaven for Legal Concepts: a Fantasy', in *Temple Law Quarterly* 58 (1985): 799.)

Jhering, R. von (1889) *Der Besitzwille (On Intention to Possess)*, Jena: G. Fischer; repr. Aalen: Scientia Verlag, 1968.(With a critique of prevailing juristic method.)

Jhering, R. von (1894) *Vorgeschichte der Indo-Europaër und die Entwickelungsgeschichte des römischen Rechts (A Prehistory of the Indo-Europeans and an Account of the Historical Development of Roman Law)*, Leipzig: Breitkopf und Härtel, trans. A. Drucker, *The Evolution of the Aryan*, London: S. Sonnenschein & Co., 1897. (Posthumously published.)

Jhering, R. von (1913) *Rudolph von Jhering in Briefen an seine Freunde (Jhering in Letters to his Friends)*, Leipzig: Breitkopf und Härtel; repr. Aalen: Scientia Verlag, 1971.(With an introduction by Helene Ehrenberg, née Jhering.)

Jhering, R. von (1988) *Briefe an Windscheid (Letters to Windscheid)*, Göttingen: Vandehoeck & Ruprecht Verlag.(Of interest to theorists; in German.)

References and further reading

Dreier, R. (1993) 'Jherings Rechtstheorie - eine Theorie evolutionärer Rechtsvernunft' and the response by O. Behrends, 'War Jhering ein Rechtspositivist?', in *Privatrecht heute und Jherings evolutionäres Rechtsdenken (Private Law Today and Jhering's Evolutionary Legal Thought)*, Cologne: Dr Otto Schmidt.(A collection of essays by contemporary scholars of Jhering; in German.)

Jenkins, I. (1960) 'Rudolph von Jhering', *Vanderbilt Law Review* 14: 169-90. (Suggesting an interpretation of Jhering that places him closer to idealism than understood by his followers.)

Losano, M.G. (1984) *Studien zu Jhering und Gerber (Studies on Jhering and Gerber)* and *Der Briefwechsel zwischen Jhering und Gerber (The Correspondence between Jhering and Gerber)*, Ebelsbach: Verlag Rolf Gremer.(Both volumes reviewed by S.L. Paulson *American Journal of Comparative Law* 37 (1989): 171-84.)

Macdonell, Sir J. and Manson E. (eds) (1914) 'Rudolph von Jhering', in *Great Jurists of the World*, Continental Legal History Series, vol. 2, Boston, MA: Little, Brown and Co.; repr. New York: Augustus M. Kelley, 1968: 590.(A brief biography.)

Smith, M. (1895) 'Four German Jurists', *Political Science Quarterly* 10: 664.(On Bruns, Windscheid, Jhering, Gneist.)

Stone, J. (1965) *Human Law and Human Justice*, London: Stevens & Co., esp. 147- 59.(An account and assessment of Jhering, with many useful references.)

Wieacker, F. (1942) *Rudolph von Jhering*, Leipzig: Kohler & Ameling; repr. Stuttgart: Kohler, 1968.(A remembrance on the 50th anniversary of Jhering's death, in German.)

Wieacker, F. and Wollschlägen, C. (eds) (1969) *Jherings Erbe: Göttinger Symposium zur 150 Wiederkehr des Geburtstags von Rudolph von Jhering*, Göttingen: Vandenhoeck & Ruprecht Verlag.(The proceedings of a symposium celebrating the 150th anniversary in 1968 of Jhering's birth, to which the contribution of H.L.A. Hart on 'Jhering's Heaven of Legal Concepts and Modern Analytical Jurisprudence' is reprinted in his *Essays in Jurisprudence and Legal Philosophy*, Oxford: Clarendon Press, (1983), and that of K. Zweigert in an extended English version by Zweigert and Siehr in *American Journal of Comparative Law* 19 (1971): 215.)

Jia Yi (201-169 BC)

Jia Yi, in forging a brilliant synthesis of classical Legalist, Daoist and Confucian doctrines, fashioned a coherent political and educational philosophy in support of strong central government. Jia's impact on Chinese history can hardly be overestimated; his philosophy significantly shaped the imperial institutions of the Han dynasty (202 BC-AD 220), which became the models for successive dynasties of imperial China.

Jia Yi's writings show him to have been a forceful, often iconoclastic thinker. At 20, in recognition of his remarkable erudition, Jia was appointed academician (*boshi*), the highest scholastic honour usually reserved for senior teachers, at the court of Emperor Wen of Han (179-157 BC). By that time, Han rule had been established for some twenty years. Fearing that the court had been lulled into a false sense of security by self-serving advisors, Jia insisted that the empire was neither at peace nor well-ruled. According to Jia's most famous essay, 'Faulting the Qin', an empire may be won by arms but it can only be preserved by 'benevolent government' for: 'The common people in every single case are the basis for the state. … Calamity and good fortune [for the ruler] does not depend upon Heaven, but upon the [will] of the people.' To win their support, good government should (1) establish more lenient penal, tax and *corvée* codes; (2) encourage the 'basic' occupations of farming and weaving, while prohibiting the wasteful production of luxury items; and (3) adopt a new calendar and new ritual standards to further distinguish Han rule from that of the Qin, the previous dynasty, which was widely regarded as oppressive. Once the ruler has won the approval of the masses, Jia argued, he can then inculcate in them the social values that make for stable community. In the process, the ruler maximizes the number of trustworthy subjects from whom he may gather necessary information and advice.

To reduce the likelihood of civil and foreign wars, which inevitably impose the greatest hardship upon the common people, Jia proposed a series of daring initiatives. In effect, the semi-feudal institutions adopted by the Han founder would be dismantled and replaced with centralized institutions. At the same time, Jia devised an insidious plan by which to sap the strength of China's nomadic enemies: Chinese luxury goods, rather than crack troops, would be sent deep into Xiongnu territory. Though senior advisors at court, stung by Jia's criticisms and jealous of his undeniable charisma, persuaded Emperor Wen to send him into virtual exile, most of Jia's policy proposals were implemented soon after his untimely death.

Jia's masterwork, the *Xinshu (New Book)* contains a number of chapters that are less overtly political. For example, the essay 'Arts of the *Dao*' builds upon early Mohist attempts to define systematically fifty-six ethical terms and their antonyms. Three essays identify and consider six patterns inherent in human nature that will lead to virtue if fostered by a classical education: the Way (*dao*), grace, disposition, divine charisma, brilliance and destiny. In addition, several chapters devoted to education attribute the development of the moral sense less to human nature than to intelligent social engineering. Therefore, the skilled use of ritual precepts (especially sumptuary regulations) designed to counter the profit motive can promote a society in which 'no man who is treacherous will benefit, and so all will be good'. Convinced that the single most effective educator in the state is the ruler, whose suasive example in turn insures the survival of the dynasty, Jia Yi also details the proper training of the heir apparent in the ritual and legal precedents that make for strong, yet benevolent government.

Jia's work shows a philosophical coherence far beyond mere rhetorical skill, despite a fundamental tension between his desire to consolidate imperial power and his drive to inculcate ethical values in order to maintain a workable balance of power within the state. With Jia's sophisticated analysis tracing the success of public policy to the quality of the contemporary ritual and educational systems, his writings in these areas are still of considerable interest today.

See also: Legalist philosophy, Chinese; Confucian philosophy, Chinese; Daoist philosophy; Logic in China; Mohist philosophy

MICHAEL NYLAN

List of works

Jia Yi (207-169 BC) *Xinshu (New Book)*, ed. and trans. Tsuruda Kyūsaku, *Kokuyaku kambun taisei*, Tokyo: Kokumin bunko, 1924, vol. 18.(A collection of essays attributed to Jia. A review of the grammar, vocabulary

and content of the received text supports the traditional attribution, despite objections regarding its content and style posed by certain sceptics.)

References and further reading

Emmerich, R. (1991) Untersuchungen zu Jia Yi (200-168 v.Chr.), PhD. thesis, University of Hamburg. (General introduction, with an emphasis on Jia Yi’s political thought.)

Hsiao Kung-ch’üan (1979) *A History of Chinese Political Thought*, trans. F.W. Mote, Princeton, NJ: Princeton University Press, vol. I, 469-83.(Still the standard overview on Jia Yi’s political programmes.)

Qi Yurang (1974) *Jiazi Xinshu jiaoshi (Collated Explanations for Jia Yi’s Xinshu)*, Taibei: Dongya.(Provides background information on the life and works of Jia Yi.)

SatōAkira (1981), ‘Shinsho yōkei hen ni tsuite’ (On the ‘Rongjing’ Chapter in the Xinshu), *Chūgoku kankei ronsetsu shiryō* 23: 309-16.(Argues the authenticity of the ‘Rongjing’ (Classic on Demeanour) chapter in the *Xinshu*).

Svarverud, R. (1992) *The Art of Defining Ethical Terms in Ancient China*, Oslo: Serica Osloensia, no. 5.(A preliminary study of the ‘Daoshu’ (Arts of the Dao) chapter in the *Xinshu*.)

Wang Zhouming (1982) ‘Xinshu fei weishu kao’ (An Investigation into the Authenticity of the *Xinshu*), *Wenxue yichan* 2: 17-28.(Defends the authenticity of the *Xinshu*).

Joachim of Fiore (*c.*1135-1202)

Joachim was a charismatic monastic reformer and inventive scriptural exegete whose study of the Bible led him to propound complex theories of history. Especially interested in the Apocalypse as a guide to history, he believed that the advent of the Antichrist and a violent end of the age were imminent. Contemporaries considered him a prophet, and this reputation - furthered by spurious works attributed to him - endured for four centuries. His theology, however, was widely criticized by such authoritative thinkers as Aquinas, Bonaventure and Bradwardine.

Joachim was born near Cosenza in Calabria, Italy. Educated as a notary, he began his career at the court of the Norman Kingdom of Sicily. An interior conversion around 1167 prompted him to devote himself to religious life. After false starts as a hermit, itinerant preacher and ordained priest, in 1171 Joachim became a monk (and later abbot) at the Benedictine monastery of Corazzo, where he began his intensive study of the Bible and the Church Fathers. From 1184, he became increasingly preoccupied with his writing, and with the pursuit of a more rigorous monasticism, eventually founding a new monastic order and monastery, San Giovanni in Fiore, around 1190.

Joachim's four major works - *Liber de concordia novi ac veteris testamenti (The Book on the Concord of the New and Old Testaments)*, *Expositio in apocalypsim (Commentary on the Apocalpyse)*, *Psalterium decem chordarum (The Ten-Stringed Psaltery)* and *Tractatus super quatuor evangelia (Treatise on the Four Gospels)* - all begun in or after 1183, interweave novel principles of exegetical theory with the exegesis itself. Except for the *Psalterium*, a study of the Trinity in the form of a commentary on the Psalms, each of these works shows his concern with the patterns of history. The *Liber de concordia* expounds the correspondence of persons and events in the Old Testament to those in the New; the *Expositio* charts the history of the church in the events of the Apocalypse; and the *Tractatus* shows the correspondence of events in the Gospels with events in the shared history of the Jews and the Greek and Latin churches. The premise of all three works is that an understanding of relations between past and present, as read through Scripture, reveals something of future history.

Joachim simultaneously adhered to several theories of history, including the Augustinian seven-age scheme (see Augustinianism) and the division of all history into two periods, one before Christ's incarnation and one after. His own theory, informed by Trinitarian theology, divides all history into three overlapping ages (*status*). Each age, in order, is identified with one person of the Trinity - Father, Son and Holy Spirit - and the relations between the ages reflect relations within the Trinity (see Trinity). Thus, the second age arises out of the first, and the third has its sources in both the first and the second. In the fullness of the third age, the Holy Spirit is to bring peace and religious perfection to the world through a new type of monks, a renewed church and a perfect 'spiritual understanding' of the Old and New Testaments. Joachim believed that tribulations traditionally associated with the end of human history and Christ's return would instead mark the transition to the third age, which he expected by the mid-thirteenth century.

Though he was an idiosyncratic exegete, Joachim's doctrinal polemic against Peter Lombard shows that he was not entirely isolated from the scholastic currents of his day. Joachim's *De unitate et essentia trinitatis* is no longer extant, but we know he accused Lombard of transforming the Trinity into a 'quaternity' by making the divine essence into something really distinct from the three persons. According to Lateran Council IV (1215), which vindicated Lombard and condemned *De unitate*, Joachim erroneously explained the divine essence as a composition of the three persons, thereby compromising divine simplicity (see Simplicity, divine).

Joachim's two other important theological treatises, *De prescientia Dei et predestinatione electorum (On the Foreknowledge of God and the Predestination of the Elect)* and *De articulis fidei (On the Articles of Faith)*, contrast strikingly with his four major works. Setting aside his theories of history, he concentrates on explaining Christian doctrine. *De prescientia* attempts to show that human freedom is compatible with divine grace and foreknowledge. Salvation is impossible without grace, but grace is available to all. It is a matter of free choice whether a person accepts and remains in grace, or rejects it. He emphasizes humility as the virtue by which one can best see and accept God's offer of grace (see Grace; Salvation). Joachim carefully distinguishes between divine foreknowledge and predestination, arguing that God foreknows, but does not predestine, all things. Foreknowledge, he argues, does not compromise human freedom, because God's knowledge is not the antecedent

cause of a future human action, but is itself dependent on that future action's being chosen (see Omniscience).

See also: History, philosophy of; Mysticism, history of

SEAN EISEN MURPHY

List of works

Joachim of Fiore (1183-?) *Liber de concordia novi ac veteris testamenti (The Book on the Concord of the New and Old Testaments)*, ed. E.R. Daniel, *Transactions of the American Philosophical Society* 73 (8), 1983.(Books I-IV only. For Book V, see the Venice edition, 1519, reprinted Frankfurt: Minerva, 1964. An English translation of Book II, Part 1, Chapters 2-12 can be found in B. McGinn, *Apocalyptic Spirituality*, New York: Paulist Press, 1979. The *Liber de concordia* includes a systematic development of Joachim's two chief theories of history.)

Joachim of Fiore (1183-?) *Expositio in apocalypsim (Commentary on the Apocalypse)*, Venice, 1527; repr. Frankfurt: Minerva, 1964.(The most important source for his ideas about the imminent end of the age.)

Joachim of Fiore (1183-?) *Psalterium decem chordarum (The Ten-Stringed Psaltery)*, Venice, 1527; repr. Frankfurt: Minerva, 1965.(Essential for understanding Joachim's trinitarian thought.)

Joachim of Fiore (1183-?) *Tractatus super quatuor evangelia (Treatise on the Four Gospels)*, ed. E. Buonaiuti, *Fonti per la Storia d'Italia* 67, Rome: Tipografia del Senato, 1930.(The last of his major exegetical works, left unfinished at his death.)

Joachim of Fiore (1183-?) *De articulis fidei (On the Articles of the Faith)*, ed. E. Buonaiuti, *Scritti minori di Gioacchino da Fiore, Fonti per la Storia d'Italia* 78, Rome: Tipografia del Senato, 1936.(Briefly analyses and explains central Christian doctrines, including the Trinity, the Incarnation and the sacraments; also shows his practical interest in the pursuit of religious perfections.)

Joachim of Fiore (1183-?) *De prescientia Dei et predestinatione electorum dialogi (Dialogues on the Foreknowledge of God and the Predestination of the Elect)*, ed. P. De Leo, *Gioacchino da Fiore: Aspetti inediti della Vita e delle Opere*, Soveria Manelli: Rubbettino Editore, 1988.(The most philosophical of Joachim's works; makes original contributions to longstanding debates over divine foreknowledge and predestination.)

References and further reading

Baraut, C. (1974) 'Joachim de Flore', *Dictionnaire de Spiritualité*, Tome VIII, Paris: Beauchesne, 1179-1201. (The best short introduction to Joachim's thought.)

McGinn, B. (1985) *The Calabrian Abbot: Joachim of Fiore in the History of Western Thought*, New York: Macmillan.(Useful chapters on Joachim's exegesis, Trinitarian theology and treatment by Aquinas and Bonaventure.)

Reeves, M. (1969) *The Influence of Prophecy in the Later Middle Ages: A Study in Joachimism*, Oxford: Clarendon.(Indispensable as a study of Joachim's influence, though very brief in its exposition of his thought and works. Reprinted with an updated bibliography, Notre Dame, IN: University of Notre Dame Press, 1993.)

West, D.C. and Zimdars-Swartz, S. (1983) *Joachim of Fiore: A Study in Spiritual Perception and History*, Bloomington, IN: Indiana University Press.(A good introduction to Joachim because it assumes no prior knowledge and follows the texts very closely.)

John of Damascus (*c*.675-*c*.750)

John of Damascus, who lived in the seventh and eighth centuries, is known for his Fount of Knowledge, which became the standard textbook of theology in the Eastern Orthodox tradition, and his opposition to the Iconoclasts, who opposed the use of images in Christian worship. The Fount of Knowledge, which drew heavily on patristic sources, was translated in the West by Robert Grossteste, where it had an influence on writers such as Peter Lombard.

John Mansur, also known as Chrysorrhoas (Golden-tongued), was born of a well-to-do Christian family in Damascus about 675 AD. He was educated by Cosmas, a ransomed Sicilian prisoner. The city was under Arab rule, but John rose in the Muslim administration until the anti-Christian Caliph Malik came to power in 685, when John retired with Cosmas to become a monk at Mar Saba near Jerusalem. He died about 750.

John composed a summa of theology, the *Fount of Knowledge* (after 743), divided into three parts. The first part, translated in the West by Robert Grossteste as the *Dialectica*, is a philosophical dictionary intended as an aid to the study of theology. It defines key terms in logic and metaphysics, many with theological implications, and generally follows the Eastern patristic tradition and responds to its concerns. The logical definitions draw heavily on Porphyry's *Isagōgē*, Ammonius Hermeae's commentary on the same and Aristotle's *Categories* (see Ammonius; Aristotle; Porphyry). The second part, also translated by Grosseteste, is a history of heresies, following verbatim the *Panarion* of Epiphanius of Constantia in most of it, and some other work for more recent heresies. In the last three chapters, which are of John's own composition, he discusses the Muslims, the Iconoclasts and a sect, otherwise unknown, that rejected the sacraments and the priesthood. The third part, the *Exposition of the Orthodox Faith*, was highly regarded in the East, where it has the same status as Aquinas' *Summa theologiae* in the West. Four translations were made of the work in the medieval West, where it was known as *De fide orthodoxa (On the Orthodox Faith)*, the most popular being those of Burgundio (1153-4), and Grosseteste's revision of Burgundio's translation (1235-40). PeterLombard used Burgundio's translation in his *Sentences*. John also wrote polemical tracts against the Muslims, Manichees and various Christian heretics, and *Sacred Parallels*, a collection of biblical and patristic texts chiefly on moral questions, arranged in alphabetical order by topic. Of his devotional and ascetic works, the three sermons on the Assumption of Mary are most famous.

John's writing is a patchwork of selections, generally uncredited, from earlier authors, though his own opinion usually emerges in his selection, omissions and commentary. The lost *Apology* of Aristides was discovered as a speech in his hagiographical romance, *Barlaam and Josaphat*. His philosophical and theological convictions are essentially those of the Cappadocian Fathers (see Patristic philosophy) and Pseudo-Dionysius. Thus he emphasizes the incomprehensibility of God's nature, particularly in connection with the Trinity, where the distinction of Persons is only thought, but nonetheless objective. He shows no knowledge of Augustine, or any other Western writer except Leo the Great. In the West, Walter of St Victor identified John as an opponent of the Augustinian *filioque* doctrine, which holds that the Spirit proceeds from both the Father and the Son, and John did hold, with Gregory of Nyssa, that the Son and the Spirit 'co-originate' from the Father, and though the Son is first-born, the nature of this priority cannot be grasped intellectually. In particular, though the Spirit proceeds from ([eacgr]k) the Father ('from' indicating a causal relation), it arises and is manifested through (δι[aacgr]) the Son (see Trinity).

John is most noted for his role in the Iconoclastic controversy, supporting the veneration of images in three *Discourses Against the Iconcolasts* (726-30) that formed the basis for the orthodox view in the Eastern Church. He could oppose the Iconoclastic Emperor Leo III in part because he lived outside the Byzantine Empire. Though condemned in the Iconoclastic Synod of 754, John was vindicated by the second Nicaean Council of 787, and has been considered a saint since the ninth century. He argued that God is beyond both being and knowledge, and so before the Incarnation it was suitable that he not be represented by any earthly image. However, he is now just as suitably worshipped in his manifestation as Jesus Christ, for just as Christ is truly God and truly man, and even the fleshly man Jesus is worthy of worship, so an image of this man is truly an image of God. It represents the now visible God, even when it represents suffering and ignominy. Therefore, whoever confesses the genuineness of the Incarnation ought to accept the veneration of images.

See also: Aristotle Commentators; Byzantine Philosophy; God, concepts of; Patristic philosophy

JOHN LONGEWAY

List of works

John of Damascus (*c.*675-*c.*750) Collected Works, ed. P.B. Kotter, *Die Schriften des Johannes von Damaskos*, Patristische Texte und Studien, Bd. 7, 12, 17, 22, 29, Berlin: de Gruyter, 1969-87, 5 vols.(This is the standard edition in Greek.)

John of Damascus (726-30) *Apologetic Discourses Against Those who Slander the Holy Images*, ed. P.B. Kotter in *Die Schriften des Johannes von Damaskos*, Berlin: de Gruyter, vol. 1; trans. D. Anderson, *On the Divine Images*, Crestwood, NY: St. Vladimir's Seminary Press, 1980; trans. M.H. Allies, *St. John Damascene on Holy Images: Followed by Three Sermons on the Assumption*, London: T. Baker, 1898.(John's most notable polemic work.)

John of Damascus (after 743) *Fount of Knowledge: The Philosophical Chapters*, ed. P.B. Kotter in *Die Schriften des Johannes von Damaskos*, Berlin: de Gruyter, vol. 1; medieval Latin translation ed. O.A. Colligan, *Dialectica, Version of Robert Grosseteste*, St Bonaventure, NY: The Franciscan Institute, 1955; trans. F.H. Chase in *John of Damascus: Writings*, Fathers of the Church 37, New York: Fathers of the Church, 1958; trans. G. Richter, *Philosophische Kapitel*, Stuttgart: Anton Hiersemann, 1982.(Richter contains a German translation embedded in extensive scholarly notes on John's sources.)

John of Damascus (after 743) *Fount of Knowledge: On Heresies*, ed. P.B. Kotter in *Die Schriften des Johannes von Damaskos*, Berlin: de Gruyter, vol. 4; trans. F.H. Chase in *John of Damascus: Writings*, Fathers of the Church 37, New York: Fathers of the Church, 1958.(This is largely taken verbatim from earlier writers and provides no independent information about heretical movements, but John is also interested in attacking contemporary hereies, and Islam, and the last chapters of the work are interesting and original.)

(after 743) *Fount of Knowledge: An Exact Exposition of the Orthodox Faith*, ed. P.B. Kotter in *Die Schriften des Johannes von Damaskos*, Berlin: de Gruyter, vol. 2; medieval Latin translation ed. E.M. Buytaert, *De Fide Orthodoxa, Versions of Burgundio and Cerbanus*, St Bonaventure, NY: The Franciscan Institute, 1955; trans. F.H. Chase in *John of Damascus: Writings*, Fathers of the Church 37, New York: Fathers of the Church, 1958. (An extremely influential work in the Eastern Church.)

John of Damascus (*c.*675-*c.*750) *Dialogue Between a Saracen and a Christian*, trans. D.J. Sahas, Leiden: Brill, 1972.(An interesting attack on Islam. Sahas also includes a translation of Chapter 100 of *Fount of Knowledge: On Heresies*.)

John of Damascus (*c.*675-*c.*750) *Encomium on the Assumption of Holy Mary, Mother of God*, ed. P.B. Kotter in *Die Schriften des Johannes von Damaskos*, Berlin: de Gruyter, vol. 5; trans. M.H. Allies, *St. John Damascene on Holy Images: Followed by Three Sermons on the Assumption*, London: T. Baker, 1898.(Perhaps John's most important contribution to more popular religious thought.)

References and further reading

Florovsky, G. (1987) *The Byzantine Fathers of the Sixth to Eighth Century*, trans. R. Miller and A.M. Dollinger-Labriolle, Collected Works of Georges Florovsky, Vaduz: Bücher vertriebs anstalt, vol. 9, 254-92. (An excellent discussion of John's theology.)

John of Jandun (*c.*1280/9-1328)

John of Jandun was the most important medieval philosopher in the Latin West to consider Averroes the true interpreter of the thought of Aristotle. He considered Aristotle to be 'the prince of philosophers', and Averroes to be the best philosopher after him. Jandun's defense of Averroes' attributing to Aristotle the doctrine that the intellect is one for all humans, and his own interpretation of various doctrines of Averroes, were much debated and criticized in Italy during the late fifteenth and the sixteenth centuries.

John of Jandun was a native of the small town of Jandun, near Rheims. Since he was a master of arts at Paris by the year 1310, he must have been born sometime before 1290, probably between 1280 and 1289. His earliest work was a question on the agent sense written in reply to a sophisma on the subject by Bartholomew of Bruges. He taught at the new College of Navarre from 1315 onwards. Jandun was closely identified with Marsilius of Padua, and fled Paris with him in 1326 when Marsilius' authorship of the *Defensor pacis* became known. Both went to the court of Emperor Ludwig of Bavaria. Excommunicated in 1327, Jandun died the following year.

In his questions on Aristotle's *Physics*, which were finished in 1315, Jandun insists that philosophical reasoning must be based on self-evident principles that agree with what is experienced by the senses. The first part of natural philosophy, namely physics, has as its subject matter mobile being or mobile body. It treats of things in motion, motion itself and the principles of motion. This science is in no way subordinated to metaphysics. Indeed, it is necessary for human happiness, since without it humans cannot know God perfectly in this life. It belongs to the natural philosopher and not to the metaphysician to demonstrate the existence of God and separated substances (the Intelligences), and to prove that they and God are immaterial (see Natural philosophy, medieval). The metaphysician's proper task is to determine the quiddity of God and the separated substances. Jandun denies that philosophers such as Aristotle and Averroes (see Ibn Rushd) could have known creation from nothing, since their knowledge starts with what can be known through the senses. That all things were created by God, including the first man, must be held both according to Christian faith and according to truth. In like fashion, while the resurrection of numerically the same individual human being is impossible according to the philosophers, it should be asserted on the basis of faith.

Jandun's psychological views are to be found especially in his questions on Aristotle's *De anima*. (There are two versions of Books I and II but only one of Book III.) However, he also discusses psychological topics in other question-commentaries on Aristotle and in independent questions. Although he does take Aristotle and Averroes to be the best and the next best philosopher and spends much effort on explicating their common position on the soul and intellect, he makes clear that he does not think that they knew the truth. Against their views, Jandun proclaims that he accepts as true that the human soul is created from nothing by God, is the form of the human body that gives existence to the body, and has immaterial powers of an agent intellect and a potential possible intellect. However, he insists that none of these things was held by Aristotle, nor can any of them be demonstrated by natural reason. On the contrary, since the miraculous is involved, these things are held only on faith.

Against Albert the Great and Thomas Aquinas, he maintains the plurality of forms in the human being. Jandun is insistent that if the intellective soul were the act of the body, giving it existence (*esse*) and united to it in existence - as Aquinas held - it would be corruptible. There are in fact two forms of the human body. The one is the sensitive or cogitative soul which gives existence to the body; the other is the intellective soul that is a single separate substance and one for all humans. It is a form only in an extended meaning of the term. Borrowing from the *de anima intellectiva* of Siger of Brabant, whom he mentions by name, Jandun explains that the intellective soul is an active cause that operates within the body (*operans intrinsecum*). It is united to the body in its operation as a sailor is united with his ship. Jandun attempts to answer two objections, namely that his account of Aristotle and Averroes would mean that the individual human being (a) does not really think and (b) lacks unity. The latter objection he admits to be the greatest difficulty and the strongest point of those who attack Averroes (see Soul, nature and immortality of the).

The intellective soul is itself composed of two different essential parts, an agent intellect and a potential (possible) intellect. Jandun denies that the agent intellect is God for Averroes, and against Thomas Wilton's proposal that the potential intellect is active, Jandun insists that it is not. He takes the phantasm to be the immediate active principle

of the intelligible species produced in the potential intellect, while the agent intellect is simply the immediate active principle of the act of thinking. The agent intellect's 'abstracting' is not a process of universalizing the content of the phantasm, but rather simply causes the act of thinking. Jandun's attributing the notion of intelligible species to Averroes was later attacked by Renaissance Aristotelians such as Agostino Nifo.

Jandun maintains that after the human being has developed its speculative intellect and has been united to all the intelligible species in the possible intellect, it knows the separate agent intellect intuitively and then ascends through the different separate substances until it enjoys an intuitive knowledge of God.

In the realm of metaphysics, Jandun shows his allegiance to Averroes in his denial that there is any distinction of essence and existence. Unlike Siger of Brabant, he does not appear to accept the doctrine of participation as central to his metaphysics. In various writings, Jandun accuses Aquinas of attempting to contradict Averroes but considers his labours to have been in vain (*Physics* VII, q.2). Jandun's different treatises on the agent sense would mark him as one of the major sources of the long discussion of this topic that lasted until the end of the sixteenth century.

See also: Aristotelianism, medieval; Aristotelianism, Renaissance; Averroism; Ibn Rushd; Marsilius of Padua; Siger of Brabant; Soul, nature and immortality of

EDWARD P. MAHONEY

List of works

John of Jandun (before 1328) *In libros Aristotelis De coelo et mundo quae estant quaestiones subtilissimae (Commentary on Aristotle's On the Heavens and Earth)*, Venice: Iuntas, 1552.(This edition also contains a work on the *De substantia orbis* that is not by Jandun.)

John of Jandun (before 1328) *Quaestio de habitu intellectus (Question on the Disposition of the Intellect)*, in Z. Kuksewicz, 'Les trois "Quaestiones de habitu" dans le ms. Vat. Ottob. 318. Editions des textes de Jean de Jandun, Guillaume Alnwick et Anselme de Côme (?)', *Mediaevalia philosophica Polonorum* 9, 1961: 3-30. (Edition of one of Jandun's Questions.)

John of Jandun (before 1328) *Quaestio de infinitate vigoris Dei (Question on God's Power)*, in Z. Kuksewicz, '*De infinitate vigoris Dei* des Pariser Averroisten Johannes de Janduno', *Manuscripta* 9, 1965: 167-70.(Edition of one of Jandun's Questions.)

John of Jandun (before 1328) *Quaestio de notioritate universalium (Questions on Universals as Better Known [than Particulars])*, in Z. Kuksewicz, 'La *quaestio de notioritate universalium* de Jean de Jandun', *Mediaevalia philosophica Polonorum* 14, 1970: 87-97.(Edition of one of Jandun's Questions.)

John of Jandun (before 1328) *Quaestio de principio individuationis (Question on the Principle of Individuation)*, in Z. Kuksewicz, '*De principio individuationis* de Jean de Jandun', *Mediaevalia philosophica Polonorum* 11, 1963: 93-106.(Edition of one of Jandun's Questions.)

John of Jandun (before 1328) *Quaestio utrum aeternis repugnet habere causam efficientem (Question: Is Having an Efficient Cause Incompatible with Eternal Things)*, in A. Maurer, 'John of Jandun and the Divine Causality', *Mediaeval Studies* 17, 1980: 198-207; repr. in A. Maurer, *Being and Knowing: Studies in Thomas Aquinas and Later Medieval Philosophers*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1980, 275-308.(Edition of one of Jandun's Questions.)

John of Jandun (before 1328) *Quaestio utrum forma substantalis perficiens materiam sit corruptibilis (Question: Is the Substantial Form that Perfects Matter Corruptible?)*, in Z. Kuksewicz (ed.) *Averroïsme bolonais au XIVe siècle: Edition des textes*, Ossolineum: Editions de l'Academie Polonaise des Sciences, 1965.(Edition of one of Jandun's Questions.)

John of Jandun (before 1328) *Quaestiones in duodecim libros Metaphysicae iuxta Aristotelis et magni Commentatoris intentionem disputatae (Disputed Questions on Aristotle's Metaphysics and the Great Commentator's Intention)*, Venice: Hieronymus Scotus, 1553; repr. Frankfurt: Minerva, 1966.(Edition of the disputed questions.)

John of Jandun (before 1328) *Quaestiones super metaphysicam (Questions on the Metaphysics)*, Book IV, q. 6, Book VI, q. 8 and Book VII, q. 6, are reprinted in R. Lambertini and A. Tabarroni, 'Le *Quaestiones super Metaphysicam* attribuite a Giovanni di Jandun. Osservazioni e problemi', *Medioevo* 10, 1984: 41-104. (Questions on Aristotle.)

John of Jandun (before 1328) *Sophisma de sensu agente (Sophisma on the Agent Sense)*, in A. Pattin (ed.) *Pour l'histoire du sens agent: La controverse entre Barthélemy de Bruges et Jean de Jandun, ses antécédents et son évolution*, Louvain: Presses Universitaires de Louvain, 1988, 118-65.(Jandun on the intellect.)

John of Jandun (before 1328) *Super libros Aristotelis De anima subtilissimae quaestiones (Questions on Aristotle's On the Soul)*, Venice: Heirs of Hieronymus Scotus, 1587; repr. Frankfurt: Minerva, 1966. (Questions on Aristotle, Jandun's major psychological work.)

John of Jandun (before 1328) *Super octo libros Aristotelis de physico auditu subtilissimae quaestiones (Questions on Aristotle's Physics)*, Venice: Iuntas, 1587; repr. Frankfurt: Minerva, 1969.(Questions on Aristotle.)

John of Jandun (before 1328) *Super parvis naturalibus Aristotelis quaestiones perutiles ac eleganter discussae (Questions on Aristotle's Parva naturalia)*, ed. M. Zimara, Venice: Heirs of Octavianus Scotus, 1505. (Questions on Aristotle.)

John of Jandun (before 1328) *Tractatus de laudibus Parisius (A Treatise on the Praises of Paris)*, in *Paris et ses historiens aux XIV^e et XV^e siècles*, ed. Le Roux de Lincy and L.M. Tisserand, Paris: Imprimerie Impériale, 1867, 32-79.(Jandun on the city of Paris.)

John of Jandun (before 1328) *Tractatus de sensu agente (A Treatise on the Agent Sense)*, in A. Pattin (ed.) (1988) *Pour l'histoire du sens agent: La controverse entre Barthélemy de Bruges et Jean de Jandun, ses antécédents et son évolution*, Louvain: Presses Universitaires de Louvain, 166-222.(Jandun on the intellect.)

References and further reading

Duhem, P. (1954) *Le système du monde*, VI, Paris: Librairie Scientifique Hermann, 534-75.(A systematic summary of Jandun's thought.)

Ermatinger, C.J. (1969) 'John of Jandun in His Relations with Arts Masters and Theologians', in *Arts libéraux et philosophie au moyen âge*, Montreal, Que.: Institut d'études médiévales, and Paris: Vrin, 1173-84.(Indicates topics that Jandun debated with such contemporaries at Paris as Bartholomew of Bruges, Thomas Wilton and William Alnwick.)

Grignaschi, M. (1958) 'Il pensiero politico e religioso di Giovanni di Jandun' (The Political and Religious Thought of John of Jandun), *Bullettino del'Istituto Storico Italiano per il Medio evo e Archivio Muratoriano* 70: 425-96.(Bases himself on the questions on the *Metaphysics* and *Rhetoric*.)

Jung-Palczewska, E. (1987) 'Jean de Jandun, a-t-il affirmé la nature active de l'intellect possible? Problème d'un contradiction dans la question 6 du livre III "De anima" de Jean de Jandun' (Did John of Jandun Posit an Active Nature for the Possible Intellect? The Problem of Contradiction in Qu. 6 of Book III of John of Jandun's Questions on Aristotle's *On the Soul*), *Mediaevalia philosophica polonorum* 27: 15-20.(Solves seeming contradiction in Jandun's remarks about whether the possible intellect is active.)

Kuksewicz, Z. (1968) *De Siger de Brabant à Jacques de Plaisance: La théorie de l'intellect chez les averroïstes latins des XIII^e et XIV^e siècles (From Siger of Brabant to Jacques de Plaisance: The Theory of the Intellect among the Latin Averroists of the Thirteenth and Fourteenth Centuries)*, Wroclaw: Editions de l'Academie Polonaise des Sciences, 202-43, 292-310.(A detailed overview of Jandun's psychology.)

MacClintock, S. (1956) *Perversity and Error: Studies on the 'Averroist' John of Jandun*, Indiana University Publications, Humanities Series 37, Bloomington: Indiana University Press.(Study of Jandun's position on the soul, the intellect and the agent sense. Many references to manuscripts.)

Mahoney, E.P. (1986) 'John of Jandun and Agostino Nifo on Human Felicity (*status*)', in C. Wenin (ed.) *L'homme et son univers au moyen âge*, Philosophes médiévaux 26, Louvain-la-Neuve: Editions de l'Institut Superieur de philosophie, vol. I, 465-77. (Compares Jandun's and Nifo's views on union of the human being to the separate intellect.)

Mahoney, E.P. (1987) 'Themes and Problems in the Psychology of John of Jandun', in J.F. Wippel (ed.) *Studies in Medieval Philosophy*, Studies in Philosophy and the History of Philosophy 17, Washington DC: The Catholic University of America, 273-88.(A synthesis of Jandun's remarks in different works regarding the nature of the soul, the intellect and cognition.)

Maurer, A. (1955) 'John of Jandun and the Divine Causality', *Mediaeval Studies* 17: 185-207, repr. in Maurer (1980) *Being and Knowing: Studies in Thomas Aquinas and Later Medieval Philosophers*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, pp. 275-308.(On efficient causality, creation and divine causality.)

Pacchi, A. (1958-60) 'Note sul commento al "De anima" di Giovanni di Jandun' (Notes on John of Jandun's Commentary on *On the Soul*), *Rivista critica di storia della filosofia* 13: 372-83, 14: 437-57, 15: 354-75.(A comprehensive summary of Jandun's psychological views including the agent sense.)

Schmugge, L. (1966) *Johannes von Jandun (1285/89-1328): Untersuchungen zur Biographie und Sozialtheorie eines lateinischen Averroisten (John of Jandun: Studies in the Biography and Social Theory of a Latin Averroist)*, Pariser Historische Studien 5, Stuttgart: Anton Hiersemann.(Close study of Jandun's life and the chronology of his writings, his relation to Marsilius of Padua, and his views on virtue, the social community

and political structures.)

Vitali, M.C. and Kuksewicz, Z. (1984) ‘Notes sur les deux rédactions des "Quaestiones de anima" de Jean de Jandun’ (Notes on the Two Versions of John of Jandun’s Questions on Aristotle’s *On the Soul*), *Mediaevalia Philosophica Polonorum* 27: 3-24.(Careful and detailed examination of this work that includes a dating of various of Jandun’s writings.)

John of La Rochelle (d.1245)

John of La Rochelle was one of the first generation of Franciscan theologians at the University of Paris. What little is known of his life places him as a close partner to the greatest of the early Franciscan teachers, Alexander of Hales. John collaborated with Alexander not only in steering the Franciscan order towards some sort of institutional equilibrium, but also in the elaboration of the first synthesis of Franciscan theology, the Summa Fratris Alexandri (Summa of Brother Alexander). Whole sections of the Summa were written by John, most notably the treatment of moral law. John wrote in his own voice a number of theological works and two books on the soul, the Tractatus de divisione multiplici potentiarum animae (Treatise on the Multiple Division of the Soul's Power) and the Summa de anima (Summa on the Soul).

John seems to have entered the Franciscan order around 1230. He then studied theology at Paris, perhaps under Alexander of Hales, the most important Franciscan theologian. John became a Master of Theology in 1236, but he remained associated with Alexander in his work and in the public imagination. Indeed, he joined with Alexander in helping to depose Brother Elias as minister general of the Franciscans. He then collaborated with Alexander and two other Franciscans to write an important commentary on the Franciscan rule in 1240-1. John was also, and more influentially, a member of the team of editors that composed a massive summa of theology, which was attributed as a whole to Alexander, from affection or filial piety, under the title *Summa Fratris Alexandri (Summa of Brother Alexander)*. Significant sections of the *Summa* were written by John, most notably the treatment of moral law. He also wrote in his own name in the genres typically used by Parisian masters of theology, producing sermons, a treatise on preaching, disputed questions and various short treatises (on grace, the articles of faith, the sacraments and topics in moral theology). John was praised by his contemporaries for his gift of clarifying inherited material by organizing it around central theological themes.

John's interest in philosophical topics seem to have been focused on the human soul, its nature and its moral dispositions. In this interest, John shows himself the heir of several twelfth-century genres, most notably treatises on natural philosophy, summaries of virtues or vices and handbooks of spiritual psychology. John also inherits Aristotle's *On the Soul* and the annexed biological books, as well as the ample erudition of newly translated Arabic medicine. He is, like his contemporary Vincent of Beauvais, within reach of an ever-expanding philosophical library. The temptation merely to accumulate citations was real, but not fatal. Indeed, John's two works on the soul move from encyclopedism to a more disciplined consideration of the soul. The first work, the *Tractatus de divisione multiplici potentiarum animae (Treatise on the Many Divisions of the Soul's Powers)*, is encyclopedic in its selection and handling of authorities. The second work, the *Summa de anima (Summa on the Soul)*, excludes a number of topics and precedent texts in order to present a tighter arrangement of its teaching. Unfortunately, only the *Tractatus* is readily available in a modern edition. Fortunately, it shows more clearly than the *Summa* how John appropriates and begins to transform his philosophical sources.

According to John's own explanation, the *Tractatus* is divided into three parts that correspond to three ways of considering the soul: according to definition, according to division and according to completion or perfection. The parts are not equal in length nor, one presumes, in importance. The division of the soul's powers is by far the longest; the consideration according to definition by far the shortest. There is also marked variety in the complexity of internal organization. The first part proposes and explains in linear sequence eleven different definitions of the soul. The second part can be broken into five sections: philosophical divisions of the soul's powers, with special attention to Avicenna (see Ibn Sina); medical divisions; a division by John Damascene (see John of Damascus); a division by Augustine; and a review of some individual powers. These powers include what John distinguishes as reason, intellect, spirit (in both intellectual and physiological senses), will, intelligence and imagination. His third part can be understood as treating three large topics: grace and its divisions; the virtues, both philosophical and theological; and beatitude.

Even this brief summary will suggest that John has luxuriant vocabularies for discriminating among internal 'faculties'. His problem is rather to simplify than to complicate the resulting descriptions. John's typical procedure in each of the three parts of the *Tractatus* is to assemble authoritative texts on the point at hand and then to analyze their competing vocabularies. These authorities cover an enormous range, and are by no means exclusively

philosophical. Definitions of the soul are taken from Plato, Aristotle, Seneca, the Book of Genesis, Nemesius of Emesa (cited as ‘Remigius’), John Damascene and the anonymous Cistercian work *On Spirit and Soul*. Among the five different divisions of the soul considered, there are medical schemata taken from Johannitius (Hunayn ibn Ishaq) and Avicenna’s *Canon of Medicine*. Even when John moves on from these schemata to consider some of the soul’s powers individually, he seems principally concerned to catalogue various descriptions of each. But John is no mere copyist, even in the *Tractatus*. He often uses one or another text to provoke an extended discussion in the style of a disputed question. For example, a definition taken from Boethius occasions a string of questions on what kind of existence ultimate happiness can have in the soul. John replies with careful distinctions among virtue, passion, habit and power, as well as with technical efforts to situate the gift of blessedness with respect to each power’s sphere of action. In this way, John’s *Tractatus* shows the labour of conceptual clarification required during the first half of the thirteenth century as masters of theology read their ways into the flood of newly translated philosophy.

See also: Alexander of Hales; Aristotelianism, medieval

MARK D. JORDAN

List of works

John of La Rochelle (1233-9) *Tractatus de divisione multiplici potentiarum animae (Treatise on the Many Divisions of the Soul’s Powers)*, ed. P. Michaud-Quantin, Paris: Vrin, 1964.(A critical edition of John’s *Treatise*.)

John of La Rochelle (1240?) *Summa de anima (Summa on the Soul)*, ed. T. Domenichelli, *La Summa de anima di Frate Giovanni della Rochelle dell’Ordine de’minori*, Prato: Giachetti, 1882.(An uncritical edition of John’s *Summa*.)

John of La Rochelle (before 1245) *Quaestiones de gratia (Questions on Grace)*, ed. L. Hödl, *Die neuen Quästionen der Gnadentheologie des Johannes von Rupella OM (†1245)*, Munich: M. Hüber, 1964.(Contains on pages 51-81 a critical edition of John’s disputed questions on grace.)

References and further reading

Bougerol, J.-G. (1994) ‘Jean de la Rochelle - les oeuvres et les manuscrits’ (John of La Rochelle - Works and Manuscripts), *Archivum Franciscanum Historicum* 87 (3-4): 205-15.(A survey of some aspects of the transmission of John’s works.)

Glorieux, P. (1943) *Répertoire des Maîtres en Théologie de Paris au XIIIe siècle*, Paris: Vrin, no. 302, vol. 2: 27. (A list of John’s works, both philosophical and theological.)

Lottin, O. (1942-60) *Psychologie et morale aux XIIe et XIIIe siècles*, Louvain: Abbey of Mont César, vol. 6: 181-223.(On John’s definitions and divisions of the soul.)

Michaud-Quantin, P. (1949) ‘Les puissances de l’âme chez Jean de la Rochelle’, *Antonianum* 24: 489-505.(A summary of John’s several discussions of the soul’s powers.)

Salman, H.D. (1947-8) ‘Jean de la Rochelle et l’averroïsme latin’, *Archives d’histoire doctrinale et littéraire et du Moyen Age* 16: 133-42.(Shows John’s dependence on intermediate sources for his knowledge of Aristotle’s natural works.)

Smalley, B. (1985) *The Gospels in the Schools, c.1100-c.1280*, London: Hambledon Press, 171-190.(On John’s Gospel commentaries, with particular attention to their Franciscan elements.)

Smalley, B. (1974) ‘William of Auvergne, John of La Rochelle, and St. Thomas Aquinas on the Old Law’, in *St. Thomas Aquinas, 1274-1974: Commemorative Studies*, Toronto, Ont.: Pontifical Institute of Medieval Studies, vol. 2: 11-71.(On John’s role in the constitution of the debate about the rationality of Old Testament Law.)

John of Mirecourt (*fl. c.*1345)

The traditional view that John of Mirecourt was condemned because he was a radical sceptic has been brought into question by more extensive research on his writings. For example, it appears that he did not doubt the existence of God as he was accused of doing. John was, however, greatly interested in describing as accurately as possible the kinds of evidence that lead to knowledge and the means by which they are produced.

John of Mirecourt, a Cistercian monk, lectured on the *Sentences* of Peter Lombard at the University of Paris during 1344-5. Propositions drawn from his *Sentences* commentary were condemned by the chancellor of the university in 1347. Against an initial list of condemned propositions, Mirecourt issued an apology in 63 articles, and then later, in reply to the official list of condemned propositions, he issued a further apology in 41 articles. Study of Mirecourt's thought has been hampered by the lack of a critical edition of the full *Sentences* commentary. The edition of the two apologies includes brief excerpts from the commentary, and questions I 2-6 (on knowledge) and I 13-16 (on enjoyment of God) have been edited. In addition, a few scholars have based their discussions of specific issues on examinations of unedited questions. However, as the table of contents printed in Tessier (1974) indicates, much more remains to be done before Mirecourt's thought may be properly assessed.

As is true of several of his contemporaries at Paris, Mirecourt's areas of special interest and the methods of inquiry he uses to pursue them show the influence of debates at Oxford earlier in the century. For example, Mirecourt displays interest in specifying and carefully describing the kinds and sources of human knowledge. He distinguishes suspicion or opinion (based on 'inevident assent' given with some concern about actual or potential error) from knowledge proper, relying on 'evident assent' given without fear of actual or potential error. Two sorts of evidence lead to evident assent: 'special' and 'natural'. Only the principle of non-contradiction, conclusions based upon it and propositions reducible to it are specially evident. Propositions derived from experience are nearly always proven only by natural evidence; the only exceptions (an example Mirecourt derives from Augustine) are those referring to the speaker's own existence, since one who doubts such a proposition would contradict one's intuitive awareness of one's own existence.

Mirecourt's careful investigation of the cognitive processes by which we acquire natural evidence shows not scepticism about the reliability of natural compared with special evidence, but deep interest in the debates at Oxford about the mechanism and respective priority of intuitive and abstractive cognition. His familiarity with these debates is revealed as a result of his style of argument: he tends to include a wide range of possible positions and to give detailed arguments for and against each one. While it is not always easy to determine Mirecourt's own view - he does not always indicate which side (if either) of an argument he prefers - on the whole it appears that Mirecourt tightens up the conditions for the acts of cognition that precede assent to natural evidence so as to reduce or eliminate intermediaries between the object of cognition and its subject. Like Adam Wodeham, Mirecourt rejects any extension of intuitive cognition to cover cases other than simple apprehension in the presence of its object: he ascribes Scotus' and Ockham's imperfect intuitive cognition (the cognition of an existent object not present at the time) to memory, and appears also to reject any intuitive cognition of non-existents (see Duns Scotus, J.; William of Ockham). While Mirecourt agrees with the usual view that abstractive cognition does not depend on the existence and presence of its object, he advances a number of arguments that propose (as Ockham did) that it does not rely on intermediary *species*, but instead results more directly from a prior intuitive cognition or from a *habitus* (that is, a disposition produced by one or several cognitions in the past).

Besides the compendious tendency of his questions, another characteristic of fourteenth-century English philosophy found in Mirecourt is his interest in using new logical and scientific techniques in order to discuss theology. For example, he uses proportional analysis to explain the conventional distinction between *usus* and *fruitio*, the 'use' and 'enjoyment' of an achieved end, and discusses questions of human will and merit using theories of intension and remission and of extrinsic limits (see Natural philosophy, medieval; Oxford Calculators). It may have been the style and methodology of Mirecourt's commentary more than the views it expressed that got him into trouble with the authorities at Paris: if, indeed, the condemnation was motivated by anything other than the simple malice to which Ceffons and d'Ailly attribute it (for the suggestion that the real controversy producing the Paris condemnations was over methodology, and references to the remarks of Peter Ceffons and Pierre d'Ailly,

see Tachau (1988): 379-80, and Courtenay (1973): 173).

See also: Oxford Calculators

FIONA SOMERSET

List of works

John of Mirecourt (1344-6) *In I librum Sententiarum (Commentary on Book I of the Sentences)*.(Portions appear in A. Franzinelli (1958) 'Questioni inedite di Giovanni di Mirecourt sulla conoscenza (Sent I q. 2-6)' (Unedited Questions by John of Mirecourt on Cognition), *Rivista Critica di Storia della Filosofia* 13: 319-40, 415-49; and M. Parodi (1978) 'Questioni inedite tratte dal I libro del Commento alle Sentenze di Giovanni di Mirecourt (q. 13-16)' (Unedited Questions Drawn from Book I of the Commentary on the Sentences by John of Mirecourt), *Medioevo* 3: 237-84.)

John of Mirecourt (1347) *Apologiae*, ed. F. Stegmüller, 'Die zwei Apologien des Jean de Mirecourt' (The Two Apologies of John of Mirecourt), *Recherches de théologie ancienne et médiévale* 5, 1933, 40-78, 192-204. (Includes both *apologiae*, the *Iohannes de Mercuria Apologia Prima* and *Iohannes de Mercuria Apologia Altera*. Excerpts from the relevant passages of the *Sentences* commentary are included in the notes.)

References and further reading

Copleston, F.C. (1963) *A History of Philosophy*, Garden City, NY: Image Books, vol. 3, 127-34.(An introductory account largely based on the two apologies.)

Courtenay, W. (1972-3) 'John of Mirecourt and Gregory of Rimini on Whether God can Undo the Past', *Recherches de théologie ancienne et médiévale* 39: 224-56; 40: 147-74a.(Discusses I 39, suggesting Mirecourt was not as radical as he has often been considered.)

Murdoch, J.E. (1978) '*Subtilitates Anglicanae* in Fourteenth Century Paris: John of Mirecourt and Peter Ceffons', in M.P. Cosman and B. Chandler (eds) *Machaut's World: Science and Art in the Fourteenth Century*, New York: New York Academy of Science.(Examines the use of scientific concepts derived from Oxford in III 4, 6-11 and I 10.)

Parodi, M. (1984) 'Il linguaggio delle proportiones nella distinctio prima di Giovanni di Mirecourt (The Language of Proportions in the first distinction of John of Mirecourt)', *Rivista Critica di Storia della Filosofia* 39 (4): 657-86.(Shows how I 13-15 use mathematical ratios.)

Tachau, K. (1988) *Vision and Certitude in the Age of Ockham*, Leiden: Brill, 365, 371-3 and *passim*.(The best guide to the intellectual background to Mirecourt's theory of cognition in late thirteenth- and early fourteenth-century Oxford. Examines the situation in Paris only briefly and in conclusion. Shows the influence of Wodeham on I 3-5.)

Tessier, G. (1974) 'Jean de Mirecourt, Philosophe et Théologien' (John of Mirecourt, Philosopher and Theologian), *Histoire Litteraire de la France* 40: 1-52.(Discusses I 21 of the *Sentences* commentary, and prints a table of contents for the whole work.)

Van Neste, R. (1976) 'The Epistemology of John of Mirecourt: A Reinterpretation', *Citeaux* 27: 5-28.(The only commentary on I 3-5 available. To be used with caution, preferably in conjunction with Tachau (1988).)

John of Paris (*c*.1260-1306)

John of Paris was a prominent Dominican theologian at Paris at the end of thirteenth century. He began his career with polemical works in defense of Thomist positions. In them, he asserts the distinction between essence and existence, the unity of substantial form and the function of matter as principle of bodily individuation. John later took part in wider controversies, including those between the French crown and the papacy. His best known work is a treatise on the mutual independence of secular and spiritual authority.

John of Paris (or John Quidort) spent much of his life as a controversialist. After the deaths of Thomas Aquinas in 1274 and Albert the Great in 1280, the Dominicans were pressed to explain their appropriation of Aristotelian doctrines in theology (see Aristotelianism, medieval). However, the strife between Thomists and anti-Thomists, or ‘Aristotelians’ and ‘Augustinians’, was symptomatic of deeper tensions in the theological use of philosophy. The controversies produced by these tensions dominated John’s career as a teacher and writer.

The year of John’s birth is uncertain. Probably he was studying as a Dominican in Paris by the end of the 1270s. What is certain is that John’s first large-scale work, written in the early 1280s, was a work by a Dominican on behalf of Dominicans. It had its origin in a controversy that threatened the standing of the order generally and of its Parisian house in particular: the controversy over Thomas Aquinas’ adaptations of Aristotle.

After March 1277 and before August 1279, the Franciscan William de La Mare wrote his *Correctorium fratris Thomae (Correctory of Brother Thomas)*, a collection of about 118 passages from Aquinas’s works, mostly from the first part of the *Summa theologiae*. William described, criticized and refuted each passage by authoritative texts drawn from the Bible and other authorities. In 1282, the Franciscan order decreed that Aquinas’ *Summa* was not to be read in Franciscan houses except when it was accompanied by William’s ‘declarations’. However, the Dominicans, who dubbed William’s work the ‘Corruptory’, had already begun to produce rebuttals. John of Paris undertook to write one of these, which is known by its opening word as *Sciendum*. His eclectic defence of Thomas is marked by an emphasis on the logic of terms and demonstrations, and is concerned primarily with showing flaws in the complaints of his opponents.

In 1292, John began his commentary on the *Sentences* of Peter Lombard, in which he takes up and augments Thomist positions. It is perhaps not surprising, then, that sixteen theses in it were singled out as heterodox. John’s point-by-point replies to the charge seem scattered and pusillanimous. Still, John also defended himself in a public disputation on the distinction between essence and existence. His defence contains several striking passages of metaphysical analysis.

John’s best known and most influential work is the short treatise *De potestate regia et papali (On Royal and Papal Power)*, his contribution to the quarrel between Philip IV of France and Pope Boniface VIII. The quarrel began in 1296 over the crown’s right to tax clerical property without seeking permission from the papacy. The resolution of the immediate dispute did not settle its principles, the most pertinent of which was the king’s claim that in the temporal regimen of his kingdom he was not subject to the rule of the pope. Boniface asserted the papacy’s right to judge the actions of all, kings included. The French party replied with a call for Boniface’s trial before a general council of the Church on charges of heresy and simony. In this atmosphere, but before agents of the French king finally arrested the pope, John wrote his measured defence of the separation of royal and papal power.

The treatise traces something like the dialectic of a medieval disputed question. Its first ten chapters describe general principles of secular and ecclesiastical power, including the principle that bishops have as such no civil jurisdiction. In the eleventh chapter, John rehearses forty-two arguments for the opposing, papal side. He then turns in Chapters 12 and 13 to deciding the question, to weigh the merits of the two sides. He argues that God gave to the Church a spiritual power comprising the capacity to perform its sacraments, the authority to teach the faithful and to coerce those who despise the sacraments, and a proper provision both for the differentiation and support of its ministers. God did not give the Church as such any direct temporal power or jurisdiction. Having set forth this determination, John proceeds in Chapters 14-20 to reply to each of the forty-two arguments advanced from the other side. He then adds five chapters in something like an appendix, addressing the so-called ‘donation of Constantine’, the claim that the papacy was immune from criticism and the issue of papal abdication. John’s

treatise proved quite influential, especially in the Gallican movement of the seventeenth century.

Determinatio de modo existendi corporis Christi in sacramento altaris (Determination on the Manner of Existing of Christ's Body in the Sacrament of the Altar), John's last and in some ways most controversial work, was an attempt to offer an alternative to Aquinas' understanding of the transubstantiation of bread and wine in the celebration of the Eucharist, namely that Christ assumes the bread and wine as he assumed a human body. This view was immediately contested. In 1305, John was censured by an episcopal commission for his alternative teaching. He died at Bordeaux in September 1306 while awaiting a ruling on his appeal to the papacy.

See also: Aquinas, T.; Aristotelianism, medieval; Political philosophy, history of

MARK D. JORDAN

List of works

John of Paris (1283-4) *Correctorium corruptorii (Correctory of the Corruptory)*, ed. J.-P. Muller, *Le Correctorium corruptorii 'Circa' de Jean Quidort de Paris*, Rome: Herder, Studia Anselmianum, 1941.(This is John's detailed rebuttal of charges against the teaching of Thomas Aquinas.)

John of Paris (1292-6) *Super Sententias (Commentary on the Sentences)*, ed. J.-P. Muller, *Commentaire sur les Sentences*, Rome: Herder, Studia Anselmianam, 1961-4.(Books 1-2 only.)

John of Paris (1296-9) *Apologia*, ed. J.-P. Muller, 'À propos du Mémoire justificatif de Jean Quidort', *Recherches de théologie ancienne et médiévale* 19, 1952: 344-51.(On John's reply to the charges against his commentary on the *Sentences*.)

John of Paris (1296-9) *Quaestio de unitate esse et essentiae (Question on the Unity of Being and Essence)*, ed. P. Glorieux, 'Jean Quidort et la distinction réelle de l'essence et de l'existence', *Recherches de théologie ancienne et médiévale* 18, 1951: 156-7.(The 'real distinction' was a characteristic position for Thomists.)

John of Paris (1302-3) *De potestate regia et papali (On Royal and Papal Power)*, ed. and trans. F. Bleienstein, *Johannes Quidort von Paris Über königliche und päapstliche Gewalt*, Stuttgart: Klett, 1969; trans. J.A. Watt, *On Royal and Papal Power*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1971.(Watt's translation contains a full and helpful introduction.)

John of Paris (1304-5) *Determinatio de modo existendi corporis Christi in sacramento altaris (Determination on the Manner of Existing of Christ's Body in the Sacrament of the Altar)*, ed. and trans. J.H. Martin, 'The Eucharistic Treatise of John Quidort of Paris', *Viator* 6, 1975: 214-40.(This had been one of the most contested topics in the dispute between Thomists and anti-Thomists.)

References and further reading

Coleman, J. (1983) 'Medieval Discussions of Property: *Ratio* and *Dominium* according to John of Paris and Marsilius of Padua', *History of Political Thought* 4: 209-28.(Connects John's discussions with the more notorious and influential views of Marsilius of Padua.)

Kaeppeli, T. (1975) 'Ioannes Parisiensis Quidort (Monoculus)', in *Scriptores ordinis praedicatorum medii aevi* 2: 517-24.(The most comprehensive bio-bibliography.)

La Mare, William de (1277-9) *Correctorium fratris Thomae (Correctory of Brother Thomas)*, in P. Glorieux (ed.) *Les premières poléemiques thomistes 1: Le Correctorium corruptorii 'Quare'*, Bibliothèque Thomiste 9, Kain, 1927.(William's text is found *passim*, interspersed among the replies to his various articles.)

Leclercq, J. (1942) *Jean de Paris et l'écclésiologie du XIIIe siècle (John of Paris and the Ecclesiology of the Thirteenth Century)*, Paris: Vrin.(Still the standard introduction to John's teaching on the relations of ecclesiastical and secular authority.)

Roensch, F.J. (1964) *Early Thomistic School*, Dubuque: Priory Press, 98-104.(The most synoptic treatment in English.)

John of Salisbury (1115/20-1180)

John of Salisbury is one of the most learned and penetrating of twelfth-century Latin writers on moral and political matters. In his style as in his teaching, John represents a style of medieval philosophy heavily indebted to Roman models of rhetorical education. His interests in grammar, dialectic, politics and ethics are subordinated to an over-arching concern for moral formation. Three of John's works stand out. The Entheticus de dogmate philosophorum (Entheticus of the Teaching of the Philosophers) is a satire on the pretensions and immoralities of those who divorce eloquence from philosophy in order to pursue power. The Metalogicon defends the traditional arts of the trivium *and asserts the unity of eloquence and the other verbal arts with philosophy. By far the most important is the* Policraticus*, a sustained argument for philosophic wisdom against the vanities of worldly success, especially in politics.*

1 Life

John of Salisbury is one of the most learned, acerbic, reactionary and penetrating of twelfth-century Latin writers of moral and political philosophy. His writing and teaching represent the culmination of a style of medieval philosophy heavily indebted to Roman models of rhetorical formation, especially to Cicero and Seneca. His discussions of grammar, dialectic, politics and ethics are subordinated to an over-arching concern for moral formation. Indeed, John may well be considered one of the pre-eminent figures for the history of moral philosophy in the Latin West before the re-introduction of Aristotle's *Nicomachean Ethics*.

Little is known of John's early life. He traveled in 1136 to Paris to study at Mont-Saint-Geneviève, then the site of numerous schools of the liberal arts and theology. From the *Metalogicon* we learn that John studied with Peter Abelard, William of Conches, Thierry of Chartres, Adam of Petit-Pont and Gilbert of Poitiers. In 1147, John joined the household of Theobald, Archbishop of Canterbury, traveling as his envoy to the Continent and managing his complex domestic affairs. Among Theobald's other aides, and so among John's friends, the most notable figure was Thomas Becket. In 1156, Henry II banished John from the English court for two years for being over-loud in defence of Church rights against the crown. He regained favour, only to lose it again by supporting Becket's cause after 1162. He spent much of the 1160s in exile on the Continent, lobbying for the rights of the English church and trying to find some middle ground between the arrogance of the crown and the recalcitrance of his friend, the archbishop. After Becket's murder, John returned to England and to new ecclesiastical offices. In 1176, he was called back to the Continent as bishop of Chartres, where he died in 1180.

John complained of the pressures of his ministerial occupations, but found time for a brilliant literary career. His writings fall into two classes. Before the struggles of the 1260s, John wrote a number of works in the liberal arts and philosophy. Three of these are notable for their philosophic teaching: *Entheticus de dogmate philosophorum (Entheticus of the Teaching of the Philosophers)*, completed probably in 1156, the *Metalogicon* in 1159, and the *Policraticus*, also in 1159. Later, John turned to ecclesiastical history, writing his memoirs of the papal court, the *Pontifical History*, and very polished lives of Anselm of Canterbury and Becket.

2 Works

The *Entheticus* is the earliest of John's major works, and it sketches out in poetic form some of the concerns that would preoccupy him in all of them. There is evidence that the *Entheticus* was never finished and, hence, never intended for circulation outside a small circle of friends. It is hard to imagine a satire of such specific ferocity ever being offered to the public. The work begins by defending the alliance of reason and eloquence against the decadence of John's contemporaries. It next extols philosophy, bringing forward the principal teachers of antiquity. The conclusion of the hortatory argument is the superiority of Christian wisdom and the folly of an exclusive infatuation with logic. The poem then breaks off to describe, in harsh terms, the little book's imagined voyage to the royal court. It ends with a brief epilogue on grace. Readers have found it hard to discern any single plan in the *Entheticus*; John took his patterns of composition from the dialectical, digressive masters of Roman oratory. He is less concerned with demonstration or analysis than with producing rhetorical effects by which the reader may be persuaded to a virtuous life.

The *Metalogicon* is often pillaged for what it records of the Parisian schools in the twelfth century. It deserves

better. With the *Metalogicon*, John first achieves the kind of scope needed for his pedagogical project. The work presents itself as a reply to unnamed opponents who have, once again, arrogantly denied the importance of eloquence to philosophy, thereby upending the whole order of liberal arts and so wrecking the foundation of philosophy. John undertakes to show what unites the verbal arts with each other and why they are all necessary for philosophy. Much of what follows in the *Metalogicon* is an erudite rehearsal of lessons about the interrelations of grammar and logic, including all of the pieces of Aristotle's *Organon*. Shorter, more elliptical lessons are recalled about philology and eloquence. However, John interrupts or enriches his rehearsal with explanations of the sources of cognition, the nature of moral virtue and the limitations of philosophy in regard to faith. The *Metalogicon* is not so much a handbook of the verbal arts, then, as an extended argument for the instrumental use of a unified *trivium* in the service of moral philosophy.

The *Policraticus*, John's last and most famous philosophical book, is typically studied as a work of political philosophy, best known for some remarks on the justice of killing tyrants, but the work is considerably more. John intends it as an extended invitation to take up philosophic life, by which he means the pursuit of the moral good. The basic contrast of the *Policraticus* is between the 'frivolity of courtiers' and the 'traces of the philosophers', between the evident emptiness of lives squandered on worldly ambition and the alluring reminders of lives devoted to the pursuit of truth. For example, Book 1 shows how the talented are corrupted by cheap gifts into the pursuit of any number of unsuitable ends: hunting, gambling, theatre and especially the shams of magic. Book 2 then distinguishes useless superstition from the pious interpretation of natural phenomena as signs of divine will. This dialectical progression, of invective and instruction, of misuse and right use, runs throughout the *Policraticus*.

For that reason, the work cannot be mapped in a small space. Indeed, John takes different maps for different parts of his text. *Policraticus* 4, for example, is loosely ordered as a commentary on Deuteronomy 17: 14-20, a passage introduced to support the claim that princes ought to be subservient to the divine law. John not only dilates the commentary over eight chapters, he omits, perhaps deliberately, certain parts of the biblical text, such as its prescriptions for the selection of rulers. Book 5 develops an analogy between the human body and human government: the prince is the head, the senate the heart, and so on. Present throughout the work are bits of moral invective, poetic quotations and historical examples, both pagan and biblical. The examples are probably the most important element in John's notion of moral teaching. He seems to find in these carefully chosen tales not just colour or incident, but the deeper logic of moral consequences.

If John's larger structures are loose, they are not negligible. In one crucial case, the structure of his argument settles a vexed problem of interpretation: whether John does or does not advocate the killing of tyrants. The claim that he does usually relies on an extended discussion of tyranny in *Policraticus* 8, though the text seems very unsatisfying as a guide to assassination. It is a description of tyranny as a moral anti-type, as a negation of what should be sought in a philosophically ordered life. Tyrants live deformed lives and so meet with hideous deaths, which they deserve because their lives deform what is fundamental in human morality. The discussion of tyranny is a striking and even terrifying depiction of the anti-type of morality. The reader can learn this from Book 8's rhetorical structure. It begins with a critique of the Epicureans, who, John says, taught self-love and self-indulgence, the fundamental vices from which tyranny springs and also, not coincidentally, the typical vices of courtiers. Hence the critique of Epicureanism, the antithesis of philosophy, is properly followed by the depiction of the tyrant and then by a moral peroration in which the reader is asked for a last time to reject the failed life of Epicureanism, the vanity of courtiers, in favour of the morally good life taught by philosophy (see Epicureanism).

What remains truly puzzling in the *Policraticus* is John's integral appropriation of pagan philosophy even as he professes the Christian faith. He is fond of juxtaposing the two: he compares Christ with Odysseus, and supplements Socrates with the Bible. He does sometimes stop to justify at least a piece of his practice, as when he argues that he can use pagan examples because Paul used them. However, John seems most often to proceed as if Christianity could complete pagan philosophy without conflict and without remainder. If he censures pagan superstition, he immediately praises pagan wisdom, and he frequently quotes the Roman poets as moral authorities. In these details as much as in his professed teaching, John shows a characteristic confidence in the unity of human learning. He also shows his antipathy to the separation of philosophy either from the liberal arts that lead to it or from the Christian theology that completes it.

See also: Chartres, School of; Political philosophy, history of

MARK D. JORDAN

List of works

John of Salisbury (1155-9) *Entheticus de dogmate philosophophorum (Entheticus major)*, ed. and trans. J. van Laarhoven, Leiden: Brill, 1987.(A critical edition with a well-annotated and well introduced translation.)

John of Salisbury (1159) *Metalogicon*, ed. J.B. Hall and K.S.B. Keats-Rohan, Turnhout: Brepols, 1991; ed. and trans. K.S.B. Keats-Rohan, Oxford: Clarendon, 1990.(Discusses the *trivium* and asserts the unity of the verbal arts with philosophy.)

John of Salisbury (1159) *Policraticus*, ed. C.C.J. Webb, Oxford: Clarendon Press, 1909; *Policraticus* I-IV, ed. K.S.B. Keats-Rohan, with the *Entheticus minor*, Turnhout: Brepols, 1993.(Salisbury's major work of political philosophy. A translation of the whole of the *Policraticus* can be pieced together from J. Dickinson, *Policraticus: The Statesman's Book*, New York: Knopf, 1927, and J.B. Pike, *The Frivolities of Courtiers and the Footprints of the Philosophers*, Minneapolis: University of Minnesota Press, 1938. Selections have been translated by C.J. Nederman, Cambridge: Cambridge University Press, 1990.)

References and further reading

Liebeschütz, H. (1968) *Mediaeval Humanism in the Life and Writings of John of Salisbury*, revised edn, Nendeln: Kraus.(A standard intellectual biography, now somewhat dated.)

Nederman, C.J. (1988) 'Nature, Sin, and the Origins of Society: The Ciceronian Tradition in Medieval Political Thought', *Journal of the History of Ideas* 49: 3-26.(The impact of Cicero on medieval philosophy.)

Nederman, C.J. (1991) 'Aristotelianism and the Origins of 'Political Science' in the Twelfth Century', *Journal of the History of Ideas* 52: 179-94.(The development of political philosophy in the Middle Ages.)

Wilks, M. (ed.) (1984) *The World of John of Salisbury*, Oxford: Blackwell.(An anthology of twenty-four papers, with a bibliography of works on John for 1953-82 on pages 445-57.)

John of St Thomas (1589-1644)

The seventeenth-century Portuguese Dominican, John of St Thomas or John Poinsot, was a major figure in late scholastic philosophy and theology. Educated at Coimbra and Louvain, he taught both disciplines in Spain: at Madrid, Plasencia and Alcalá. Aspiring to be a faithful disciple of Thomas Aquinas, he published a three-volume Cursus philosophicus thomisticus (Thomistic Philosophical Course) and before he died began the publication of a Cursus theologicus (Theological Course). His philosophical writing was explicitly on logic and natural philosophy. However, in both his philosophical and theological works, he treated many metaphysical, epistemological and ethical issues. His logic is divided into two parts, formal and material. Of particular interest is his semiotic doctrine which appears in the second part. In natural philosophy, he explained Aristotle with a Thomistic slant. While following Aquinas in theology, John at times developed his master's doctrine along new lines. Both in his own time and after he has had considerable authority within scholasticism, especially for Thomists. Among those whom he has influenced in twentieth-century Thomism are Joseph Gredt, Reginald Garrigou-Lagrange, Santiago Ramirez, Jacques Maritain and Yves Simon.

1 Life and works

John of St Thomas was born John Poinsot in Lisbon. He studied humanities and philosophy at the University of Coimbra, where in 1605 he received a bachelor of arts degree. While studying philosophy, he probably attended lectures given by the Jesuits who directed the College of Arts at Coimbra and whose course in philosophy was published in five volumes which appeared successively between 1562 and 1606 (see Collegium Conimbricense). In October 1605, John began the study of theology at Coimbra and the following year he continued this at Louvain where in 1608 he gained his bachelor's degree in biblical studies. On 17 July 1609, he entered the Order of Preachers (the Dominicans) at the Convent of Our Lady of Atocha in Madrid, where he spent the next year. On 17 July 1610, he made his formal profession as a Dominican and at this time chose the name by which he would be known in religion: John of St Thomas. From 1610 to 1615 he studied philosophy and theology at the Atocha convent where from 1615 to 1620 he taught arts. In 1620, he went as lecturer in theology to the Dominican priory at Plasencia. Between 1625 and 1630 he was a lecturer in arts and theology at the Dominican College of St Thomas in Alcalá. In 1626, he became regent of studies there and in 1627 was appointed a consultant for the Supreme Tribunal of the Spanish Inquisition. From 1630 to 1640, he held the *cátedra de vísperas* (chair of vespers) in theology at the University of Alcalá. During this period he received his doctorate in theology in 1633 and he published, at Alcalá and Madrid, his logic and natural philosophy, which, printed together at Rome in 1637-8 and at Cologne in 1638, became his *Cursus philosophicus thomisticus (Thomistic Philosophical Course)*. At Alcalá in 1637, he began to publish his *Cursus theologicus (Theological Course)*, which was completed posthumously from its author's manuscripts edited by Diego Ramirez, Jacques Quetif and Francis Combefis. In 1641 he attained at Alcalá the theological *cátedra de prima* (chair of prime), which he held until 1643. In that year, he was appointed confessor to King Philip IV. Accompanying the king on an expedition to Catalonia, John was stricken by fever at the siege of Lerida and died at Fraga in Aragon.

Most of all, John of St Thomas aimed at being a true disciple of Thomas Aquinas. As he tells us in his *Cursus theologicus* (I d.2 a.5), this entailed two things: (1) to follow Aquinas' doctrine as true and Catholic, and (2) to develop it as much as possible. Since beyond Aquinas in philosophy there was Aristotle, John's own philosophical work was also very much in the vein of Aristotelian commentary (see Aristotelianism, Renaissance). Explicitly, as published, that work was in logic and natural philosophy. However, in both his philosophical and theological writings, he covered numerous issues in metaphysics, epistemology and ethics. On such issues he almost invariably sought to defend Thomistic doctrine against its adversaries. Chief among these, as John saw them, were the Jesuits, Gabriel Vásquez (1549-1604) and Francisco Suárez.

2 Logic

John of St Thomas' logic is divided into two parts. In the first part, his concern is with the formal theory of correct thinking. This part, which includes such medieval items as supposition, exponibles and consequences, corresponds especially to the *Summulae Logicales* of Peter of Spain and the *Prior Analytics* of Aristotle. In the larger second part, which corresponds to the *Isagōgē* of Porphyry, plus the *Categories* and the *Posterior Analytics* of Aristotle,

he deals with material logic, which is a general theory of scientific demonstrations and the necessary connections they involve depending upon their content. Organized according to the three operations of the intellect (simple apprehension, judgment and reasoning) the first part treats of terms, propositions, consequences and syllogisms, along the way attending to definitions, divisions and their various facets. In the second part, John reflects upon the nature of logic itself, which he thinks is at once an art and a science. As science, he says, 'logic is essentially and absolutely in virtue of its own principles speculative, but it takes on the manner of a practical science in so far as it offers rules and direction for speculation itself' (*Ars Logica*, II q.1 a.4). The object of logic is beings of reason, such as species, genus, subject, predicate, antecedent, consequent and so on, formed by the mind's reflections upon its own operations. Such beings of reason, which have some foundation in reality outside the mind, fall under a wider notion that includes beings of reason like chimeras. These last lack such a foundation because they are impossible (that is, self-contradictory) objects. Contrasted with both sorts of beings of reason are real beings which are divided into the various Aristotelian categories (see Logic, medieval §4).

The categories are the central concern of the second part of John's logic. He rejects Duns Scotus' teaching that *ens* (being) is said univocally of the categories (see Duns Scotus, J. §5). Rather, as Cajetan (§2) thought, 'being' as said of the categories is 'formally analogous with an analogy of proper proportionality' (*Ars Logica*, II q.14 a.3). Passing through each category in succession he concurs with Cajetan's interpretation of Aquinas (*Summa theologiae* Ia.28.1) to the effect that relation is unique because it can be found in the order either of real being or being of reason. Only relation, since it is not just 'in' something but also 'towards' something, can transcend categorial status and be conceived apart from real existence either in itself (that is, the mode of existence proper to a substance) or in something else (the mode proper to an accident). This thought was to be central to John's doctrine of signs.

The doctrine of signs itself, which in part at least reflects the influence of John's Jesuit teachers at Coimbra, forms a treatise which runs over questions 21-3 in the second part of his logic. Essentially relational, signs are divided first into formal and instrumental, and second into natural, conventional and customary. All signs make something else known: formal signs (for example, a concept or an impressed species) do so without themselves first being known; instrumental signs (such as smoke or a spoken word) are themselves first known and then lead to the knowledge of something else. Natural signs (smoke) differ from conventional signs (words) inasmuch as the former simply arise from causal connections in the natural order while the latter result from human choice. Customary signs also result from choice but can be natural when a custom leads us naturally to its cause. Signs involve two non-reciprocal relations: between the sign and the significate (as in the relation of measured to measure) and between the sign and the cognitive power of a knower (as in the relation of measure to measured). Considered just as such, these relations are not in the line of physical causality but rather in that of intentionality in which the sign substitutes for the significate. Within this line of intentionality the ability of relation to transcend real being and being of reason becomes the basis of a unified semiotic doctrine respecting all signs of whatever sort they are (see Collegium Conimbricense; Language, Renaissance philosophy of §1).

3 Natural philosophy

John originally planned his natural philosophy in parts corresponding to four subject areas: mobile being in general, incorruptible mobile being, corruptible mobile being and animated mobile being. The second part, a tract on astronomy, may have existed in manuscript but was never published. The first part contains twenty-six questions principally expounding the matter of Aristotle's *Physics*. Thus after a question about the science of natural philosophy, its formal object (mobile being), its unitary character and way of proceeding, he proceeds to explain the principles of natural things (matter, form and privation). Going beyond Aristotle, he here *Naturalis philosophiae*, I q.7) discusses subsistence and existence, and he defends the Thomistic real (*a parte rei*) distinction of essence and existence in creatures. Continuing with natural philosophy, he follows Aristotle to explain nature versus both art and what is forced or unnatural, causality in general, the four basic causes (material, formal, efficient and final), motion, the infinite, place, the void and time. Near the end of this part (*Naturalis philosophiae*, I q.24 a.3) he treats Aristotle's Prime Mover. This he identifies with God whose attributes of infinity, immateriality, intelligence, ubiquity, eternity and unity he deduces from the fact that God is pure act. The third part corresponds to Aristotle's *On Generation and Corruption* and *Meteorology*. Through twelve questions on the *On Generation*, it deals with such topics as substantial generation, corruption, alteration, intension, remission, action and reaction, mixing, condensation and rarefaction, nutrition and growth. This last provokes the question:

‘How does a living thing remain the same during the whole time of its nutrition and growth?’ (*Naturalis philosophiae*, III q.9 a.6). To this John replies with Aquinas that living flesh remains the same not according to that within it which is material but according to that which is formal, that is, according to a disposition continuing over time to retain a certain figure, temperament, and order in respect of other parts and the form of the whole. The fourth part contains John’s philosophical psychology. After a brief summary of Book 1 of Aristotle’s *On the Soul*, it discusses issues raised in Books 2 and 3 about the soul in general, the vegetative soul, the sensitive soul with its external and internal sense powers, and finally the rational soul. In a more Thomistic than Aristotelian way, he defends (*Naturalis philosophiae*, IV q.9 a.1) the immortality of this last on the basis of Scripture and with two philosophical arguments. These are, first, that the operational independence of the soul from the body indicates its ability to survive the death of the body, and second, that a need for ethical sanctions here and now suggests the soul’s survival to receive reward or punishment in an afterlife (see Natural philosophy, medieval §§8-9).

4 Metaphysics

In his *Cursus theologicus* John addressed a variety of metaphysical questions about the existence, nature and operations of God. Like Aquinas (§13), he rejects the assertion that God’s existence is self-evident. Instead, a demonstration must be mounted from creatures as effects to God as their cause. In this, he retraces and explains the famous ‘five ways’ of Aquinas. He also follows Aquinas in his general deduction of the attributes of God. Distinctive, however, is his interpretation of Aquinas to the effect that it is the very act of intellection itself which is ‘formally constitutive of the divine nature’ (*Cursus theologicus*, I d.16 a.2). Also distinctive, but perhaps more a matter of updating, is John’s elaborate confrontation, here (d.18 a.4) as well as earlier in his logic (*Ars Logica*, II q.2 a.4), of the late scholastic question of whether God knows impossible beings of reason. In contrast with Vásquez, he maintains that while God may not form such beings of reason, he can know them as formed by us. Both the question itself and his discussion of it have an important bearing on scholastic intentionality teaching. Finally in his moral doctrine, yet another difference with Aquinas which John most likely saw as an updating was in his embrace of ‘probabilism’, the theory that one may resolve a moral doubt by accepting any opinion proposed by a prudent moral authority (*Cursus theologicus*, IV d.12 a.3) (see God, concepts of).

See also: Aquinas, T.; Aristotelianism in the 17th century; Aristotelianism, Renaissance; Collegium Conimbricense; Language, Renaissance philosophy of; Logic, Renaissance

JOHN P. DOYLE

List of works

John of St Thomas (1631-2) *Cursus philosophicus: Ars Logica (Philosophical Course: Logical Art)*, ed. B. Reiser, vol. 1, Turin: Marietti, 1930; trans. F.C. Wade, *Outlines of Formal Logic*, Milwaukee, WI: Marquette University Press, 1955; trans. Y.R. Simon, J.J. Granville and G.D. Hollenhorst, *The Material Logic of John of St Thomas*, Chicago, IL: University of Chicago Press, 1955.(Contains the two parts of John’s formal and material logic. The 1955 editions are partial translations.)

John of St Thomas (1632) *Tractatus de Signis: The Semiotic of John Poinsot*, trans. and ed. J.N. Deely with R. Austin Powell, Berkeley, CA: University of California Press, 1985.(A translation and interpretative arrangement of passages from the *Ars Logica*, including Part II, Questions 21-3.)

John of St Thomas (1633-5) *Cursus philosophicus: Naturalis philosophiae partes I, III, et IV (Philosophical Course: Natural Philosophy)*, ed. B. Reiser, Turin: Marietti, vol. 2, 1933; vol. 3, 1937.(Contains John’s natural philosophy, including his philosophical psychology in volume three.)

John of St Thomas (1637-45) *Cursus theologicus (Theological Course)*, ed. Solemnes, vols 1-4, Paris: Desclée, 1931-46; vol. 5, Matiscone: Protat Frères, 1964.(Contains John’s commentary questions on Aquinas’ *Summa theologiae*.)

John of St Thomas (1649-67) *Cursus theologicus (Theological Course)*, vols 6-8, ed. D. Ramirez and F. Combefis, Paris: Louis Vivès, 1885.(Completed posthumously from John’s manuscripts, this edition continues his commentary on Aquinas’ theology.)

References and further reading

Ashworth, E.J. (1988) ‘The Historical Origins of John Poinsot’s *Treatise on Signs*’, *Semiotica* 69: 129-47.(A critical review of Deely’s translation and interpretation of the *Tractatus de Signis*.)

Deely, J. (1983) ‘Neglected Figures in the History of Semiotic Inquiry: John Poinsot’, in A. Eschbach and J.

Trabant (eds) *History of Semiotics*, Amsterdam: John Benjamins Publishing Company, 115-26.(Treatment of John's doctrine of signs.)

Deely, J. (1994) *New Beginnings. Early Modern Philosophy and Postmodern Thought*, Toronto, Ont.: University of Toronto Press.(John's semiotic doctrine as central to a new approach to the history of philosophy.)

Deely, J. (ed.) (1994) 'John Poinsot', special issue of *American Catholic Philosophical Quarterly* 67 (3).(Contains articles on various aspects of John's work.)

Lohr, C.H. (1987) *Latin Aristotle Commentaries. II Renaissance Authors*, Florence: Olschki, 204-5.(Contains biographical and bibliographical information on John.)

Maritain, J. (1938) 'Signe et symbole', *Revue Thomiste* 44: 299-330; trans. H. Binsse, 'Sign and Symbol', in *Redeeming the Time*, London: Geoffrey Bles, 1943, 191-224, 268-78.(Presents John's doctrine of signs.)

Ramirez, J.M. (1924) 'Jean de St Thomas', *Dictionnaire de théologie catholique*, Paris: Letouzey & Ane, vol. 8, cols 803-8.(Contains information on John's life, writings and theological doctrine.)

Johnson, Alexander Bryan (1786-1867)

Johnson was a self-taught philosophic genius who, completely alone, set out in the 1820s to analyse the nature and limits of language. He was the first thinker who consciously and systematically based his whole approach to philosophical problems on a critique of language. According to Johnson, a knowledge of the nature of language is important because it 'bears the same relation to all speculative learning as a knowledge of the qualities of drugs bears to the practice of medicine or as a knowledge of perspective and colours bears to painting' (1854).

Johnson tried to show that most of the propositions and questions of speculative philosophers are neither true nor false, but 'insignificant'. They arise from our 'misunderstanding of the nature of language'. He found the source of the misunderstanding in our tendency to interpret nature by language. 'My lectures', Johnson wrote, 'will endeavour to subordinate language to nature - to make nature the expositor of words, instead of making words the expositor of nature. If I succeed, the success will ultimately accomplish a great revolution in every branch of learning' (Rynin 1947: 300). Johnson's work did not attract any attention until the 'linguistic turn' in philosophy in the twentieth century made his rediscovery possible.

Born in 1786 in Gosport, England, of Dutch-Jewish ancestry, Johnson emigrated to the USA in 1801 and settled in Utica, New York, where he embarked on a long and successful career in banking. He published books on banking, on politics, and on religion and morality. However, what engrossed him most was language and philosophy. He published three books on language: *The Philosophy of Human Knowledge* (1828), *A Treatise on Language* (1836) and *The Meaning of Words* (1854). He also published two books on the theory of knowledge: *The Physiology of the Senses* (1856) and *Deep Sea Soundings and Explorations of the Bottom* (1861). In his unpublished autobiography Johnson briefly summarized what he thought his writings had accomplished:

> The verbal systems of speculative philosophy are as interminable as the different tunes that can be formed out of the notes of a piano, and realizing how barren such philosophies have ever been for the settlement of the questions for which such philosophies are employed, I have essayed the new system as an ultimate and fixed limit of all speculative knowledge; and which will in time…cause an abandonment of the old and endless speculations of the verbal philosopher.
>
> (Johnson's unpublished autobiography, date of writing unknown)

Johnson conceived of his philosophical task as that of setting limits to our knowledge: in *The Physiology of the Senses*, the limits of what we can perceive; in his books on language, the limits of what we can say; and in *Deep Sea Soundings*, the limits of what we can think. These limits are determined by the nature of reality and the nature of human beings. Human beings, according to Johnson, are physical, emotional and intellectual (he called it 'Man's Triplicity') and reality as it appears to us is divisible into three irreducible classes: the physical (sights, sounds, tastes, feels and smells), the emotional (the emotions of joy, pain, fear, awe, and so on) and the intellectual (concepts such as cause, identity, infinity, and so forth, which Johnson called 'intellections'). Johnson considered the failure to discriminate clearly these three irreducible classes the main cause of philosophical confusions.

> Properly discriminated into its organic classes, all our knowledge is free from mystery, or equally mysterious if we prefer to esteem it mysterious; but we add an unnecessary mystery at our inability to discover sensibly what is not sensible. Human knowledge can be analyzed no farther than into the organisms to which it pertains; and all attempts to delve below or beyond this boundary are founded in ignorance of the inconvertibility into each other, of the information yielded by our several organisms.
>
> (Johnson 1861: 10)

Johnson's conception of the nature and tasks of philosophy is amazingly similar to that of Wittgenstein. According to Johnson, speculative philosophers are 'misled by language', or 'deceived by the forms of language'; they 'play bo-peep with words', and their 'misuse of language' causes intellectual 'perplexity', 'amazement' and 'mystery'. Thus, Johnson's verdict on previous philosophy is: 'much of what is esteemed as profound philosophy is nothing but a disputatious criticism on the meaning of words' (1836: 282) and 'Nearly all the perplexing questions of speculation are produced by verbal equivokes… Such equivokes pass usually for profound mysteries; but all

mystery vanishes' when we understand the nature of language correctly (1854: 231-2). Like Wittgenstein, Johnson's goal in philosophy is to make philosophical problems 'vanish' or to 'terminate the mystification'.

To make all philosophical perplexities 'vanish', Johnson concentrated much of his efforts on the study of language, especially its inherent defects. On the title page of *The Meaning of Words* we find the statement: 'Four ineradicable fallacies are concealed in the structure of language: it identifies what unverbally are diverse, assimilates what unverbally are heterogeneous, makes a unit of what unverbally are multifarious, and transmutes into each other what unverbally are untransmutable'. This structural defect is one of the main causes of philosophical problems. Language consists of a limited number of words, but they are used to talk about an infinite number of things. And because we use one word to refer to many things, we assume there must be something in common among these things. As Johnson puts it, 'In all our speculation, we estimate created existences by the oneness of their name' (1836: 47). For example: 'I am speaking, I am standing, several persons are present. Each of these assertions is a truth. But if we seek among these truths for truth itself, believing it to be a unit, we are seeking a nature for what is merely a contrivance of language' (1836: 73).

Johnson used his insights on language to throw light on a whole host of traditional philosophical puzzles and in the process anticipated many ideas which became fashionable one hundred years later. For example: 'Words have no inherent signification; but as many meanings as they possess applications to different phenomena' (1836: 73); and 'If we wish to know the meaning of the word truth in any given case, we must examine the circumstances to which the word is applied' (1836: 73); another example: 'After hearing all that he can adduce in proof of the earth's sphericity, consider the prepositions significant of these proofs. If you deem it significant beyond them, you are deceived by the forms of language' (1836: 129). Similarly, a theory 'is significant of nothing but the data which are adduced in proof of the theory' (1836: 236).

K.T. FANN

List of works

Johnson, A.B. (1828) *The Philosophy of Human Knowledge, or a Treatise on Language*, New York: G. & C. Carvill.(A version of this treatise, collated with *A Treatise on Language*, was reprinted in Rynin 1947.)

Johnson, A.B. (1836) *A Treatise on Language, or The Relation which Words Bear to Things*, New York: Harper & Bros.(Reprinted in Rynin 1947, in a version collated with *Philosophy of Human Knowledge*.)

Johnson, A.B. (1854) *The Meaning of Words*, New York: Appleton & Co.

Johnson, A.B. (1856) *The Physiology of the Senses*, New York: Derby & Jackson.

Johnson, A.B. (1861) *Deep Sea Soundings and Explorations of the Bottom: or The Ultimate Analysis of Human Knowledge*, Boston, MA.(This work was published privately by the author.)

References and further reading

Rynin, D. (ed.) (1947) *Alexander Bryan Johnson: A Treatise on Language*, ed. D. Rynin, Berkeley, CA: University of California Press, 1947.(A collated version of Johnson's *Treatises* of 1828 and 1836, with a critical essay on his philosophy of language.)

Todd, C.L. and Blackwood, R.T. (eds) (1969) *Language and Value: Proceedings of the Centennial Conference on the Life and Works of Alexander Bryan Johnson*, New York: Greenwood Publishing Co.

Johnson, Samuel (1696-1772)

Johnson was the first important philosopher in colonial America and author of the first philosophy textbook published there. He derived his views largely from others, combining in one system elements from diverse sources. He followed the empiricists in holding that knowledge begins with sensations but held the Augustinian view that knowledge of necessary truths comes only from the mind's illumination by divine light. With Berkeley, he denied matter's existence, viewing bodies as collections of ideas. He held that these ideas are 'faint copies' of God's archetypal ideas, which he thought of in much the same way as had Malebranche and John Norris. His ethical views, influenced by William Wollaston, take happiness to be the supreme good, stressing that human beings should seek a happiness consonant with their nature as rational, immortal and social beings.

1 Life

Born at Guilford, Connecticut in Colonial America, Johnson was brought up as a Calvinist and educated at Yale in the Ramist philosophy favoured by the Puritans. At eighteen, he wrote 'An Encyclopedia of Philosophy', a compendium of learning set forth in 1,271 propositions. At this stage of his intellectual development he considered the masters of modern philosophy to be Ramus and his followers Richardson and Ames. While still a student at Yale, Johnson's outlook underwent a twofold revolution: after reading works of Bacon, Locke, Newton and John Norris, he became a keen defender of the 'new philosophers'; and after reading works of Anglican apologetics, he converted to Anglicanism and became, in 1722, a priest. During the next four decades, he produced successive versions of his compendium of philosophical learning (not all appeared in print), first revising his 'Encyclopedia' in light of his new views (1716), then composing a 'Logic' (1720), several versions of *An Introduction to the Study of Philosophy* (1731) and finally *Elementa philosophica* (1752). The *Elementa*, his most important work, consists of two parts, 'Noetica: Or the First Principles of Human Knowledge' and 'Ethica: Or the First Principles of Moral Philosophy' (the latter a revision of his 1746 *A New System of Morality*). Locke and Newton were important influences on Johnson once he was 'wholly changed to the New Learning'. But from 1730 onwards, his dominant influence was Berkeley, whom he visited during the latter's residence in Rhode Island, with whom he carried on a notable correspondence, and to whom he dedicated *Elementa philosophica*. Johnson was a founder of King's College (today Columbia University) and its first president.

2 Epistemology and metaphysics

With Locke, Johnson held that knowledge begins with 'ideas' received from the senses, ideas that are either simple (for example, red, sweet, hard) or complex (such as our idea of an apple). But he followed Berkeley in insisting that these ideas do not *represent* bodies to the mind; rather, a complex idea of sensation - for example, a certain combination of colour, shape, odour and so on - *is* a body; again like Berkeley, he held that, since bodies are combinations of ideas and ideas cannot exist without a mind that perceives them, bodies exist only if perceived.

Having got ideas from sensation, the mind, reflecting on its own operations and states, gets 'simple notions' of perceiving, remembering, willing and the like, and the complex notion of itself as a 'spirit' (mind). Self-reflection further discloses to the mind that, although it passively receives ideas from the senses, it is essentially active and capable of self-determination - a view that led Johnson to attack belief in predestination (see Predestination). From sensation and reflection we thus get notions of two kinds of being: bodies (passive beings that exist only because they are perceived) and spirits (active, perceiving beings). Only an active being can be a causal agent, so spirits are the sole true causes. We are immediately aware only of our own minds and of the ideas we perceive, but we infer, from their bodily behaviour, the existence of other finite spirits. Further, we have abundant evidence of the existence of God, the infinite spirit. (Johnson gave several arguments for God's existence, one inspired by Locke's argument in *Essay concerning Human Understanding* IV.10; one taken from Berkeley's *Principles of Human Knowledge* §§146-7; and one a version of the Augustinian argument that infers, from the existence of 'eternal and necessary truths', the existence of an eternal mind that knows them - see Augustine).

Johnson acknowledged Berkeley as his chief inspiration, but he also expressed high regard for Locke, Wollaston, Norris, Malebranche and Fénelon. By the last three he was persuaded of the Augustinian doctrine that we can know 'eternal truths' of logic, mathematics, metaphysics and morality only because God, by 'divine light', directly illuminates our 'pure intellect', disclosing to it an 'intelligible realm'. To Johnson, the doctrine that God

immediately causes our knowledge of the intelligible order seemed the complement of Berkeley's doctrine that God immediately causes our perception of the sensible order. And he found a use for Malebranche's and Norris' doctrine that there is in the divine mind an 'ideal world' - the intelligible, non-sensuous model of which the natural or sensible world is a copy. Berkeley had held that there are, in God's mind, archetypal ideas that are somehow correlated with our sensible ideas, but he never gave a very full account of his theory of archetypes, though Johnson several times, in his letters, asked him to do so. With Johnson, the doctrine of archetypes became part of a theory of representative perception: our impermanent, private sensible ideas are 'faint copies or images' that represent to us a stable, ideal world that exists, independently of our minds, in the divine mind.

3 Ethics

Happiness, for Johnson, is the 'natural good', the end all conscious beings strive for; but the happiness we seek should be consonant with our nature. Hence, ethics must begin with the study of human nature. We have senses like the animals, argued Johnson, but we also have reason; our bodies will perish after a time, but our souls are immortal; we have private interests but are also by nature social and mutually dependent. We must seek, therefore, a happiness consonant with our rational, immortal, social nature. Hedonists fall into error not in taking happiness to be our natural end but in failing to discover our true nature. These considerations led Johnson to a threefold account of duty: we have duties to ourselves, to God and to our fellow creatures. To ourselves, our duty is to pursue a kind of happiness befitting our nature; we fail in this duty if we gratify our senses at the expense of reason, temporal interests at the expense of eternal ones, private interests at the expense of public ones. Our duty to God is to love and worship him, both because we owe that to the being from whom our existence derives and because it is in his power to make us happy. To others, our duty is to increase their happiness and relieve their suffering when we can, both because reason shows us that happiness is also their natural end, and because our own happiness is bound up with theirs. One who fulfils each of these duties has the *virtues*, respectively, of temperance, piety and benevolence; one who fulfils them all has the virtue of justice, rendering to all persons what is their due.

Elementa philosophica was the philosophy textbook used at the colonial colleges that became the University of Pennsylvania and Columbia University. With his pupil Jonathan Edwards, Johnson was one of the earliest American defenders of idealism, a species of philosophy that, in one form or another, has been a recurring motif in American thought.

See also: Idealism

CHARLES J. McCRACKEN

List of works

Johnson, S. (1929) *Samuel Johnson: President of King's College, His Career and Writings*, ed. H.W. Schneider and C. Schneider, New York: Columbia University Press, 4 vols.(The standard edition of Johnson's works; volume 2 includes his previously unpublished 'Encyclopedia of Philosophy' and 'Logic', and his philosophical correspondence with Berkeley and others.)

Johnson, S. (1731) *An Introduction to the Study of Philosophy*, London; 2nd edn, London, 1744.(A short work first published in the periodical *Republic of Letters*, May 1731; the second edition is an enlarged version published as a small volume for use of students at Yale. Excerpts are included in *Samuel Johnson* (1929), volume 2.)

Johnson, S. (1746) *A New System of Morality*, Boston.(Also bears the title *Ethices elementa. Or, the First Principles of Moral Philosophy*; a revised version of this work became Part II of *Elementa philosophica*.)

Johnson, S. (1752) *Elementa philosophica*, Philadelphia.(Johnson's chief work, setting out his metaphysical, epistemological, and ethical views, and incorporating material from the two preceding works. It is included in its entirety in *Samuel Johnson* (1929), volume 2.)

References and further reading

Berkeley, G. (1949) *The Works of George Berkeley, Bishop of Cloyne*, vol. 2, ed. A.A. Luce and T.E. Jessop, Edinburgh: Thomas Nelson.(This volume contains the two philosophical works that most deeply influenced Johnson, *A Treatise concerning the Principles of Human Knowledge* and *Three Dialogues between Hylas and Philonous*; it also contains the philosophical correspondence between Berkeley and Samuel Johnson (1729-30),

in which Berkeley replies to a number of penetrating questions asked by Johnson about those works.)
Carroll, P.N. (1978) *The Other Samuel Johnson: A Psychohistory of Early New England*, Rutherford, NJ: Fairleigh Dickinson University Press.(A psychoanalytically-oriented biography in the manner of Erik Erikson.)
Ellis, J. (1973) *The New England Mind in Transition: Samuel Johnson of Connecticut*, New Haven, CT: Yale University Press.(A good general account of Johnson's life and thought.)
Flower, E. and Murphey, M.G. (1977) *A History of Philosophy in America*, vol. 1, New York: G.P. Putnam's Sons, 81-99.(A brief, accurate account of Johnson's philosophy.)
Locke, J. (1689) *An Essay concerning Human Understanding*, ed P.H. Nidditch, Oxford: Clarendon Press, 1975. (Johnson's early 'Logic' closely followed the Essay's account in Book 2 of the origin and kinds of ideas, and in Book 4 of the degrees of knowledge and of probable assent. The Essay's influence is still evident much later in *Elementa philosophica*.)
Schneider, H.W. (1958) *The Puritan Mind*, Ann Arbor, MI: University of Michigan Press.(Chapter 5 contains a good account of Johnson's development by a leading historian of American philosophy.)

Johnson, Samuel (1709-84)

Famous as a man of letters and lexicographer, Johnson was no formal academic philosopher - indeed he was suspicious of abstractions. His works perfectly embody the darker side of the eighteenth-century mind, with its distrust of theoretical reason and system-mongering, and a profound sensitivity to the imperfections of a human existence in which there was more to be endured than to be enjoyed.

Born on 18 September 1709 in Lichfield, Staffordshire, the child of ill-matched parents, Samuel Johnson was a sickly child. At the age of three he was taken by his mother to Queen Anne in London to be cured of a scrofula through the ‘royal touch’.

He matriculated at Pembroke College, Oxford, but poverty forced him to leave after only two years. For the next two decades his life was a struggle with penury and with psychological traits (awkwardness, anxiety, anger) that left him prey to melancholy, and which manifested themselves in a profusion of personal eccentricities. His melancholy coloured his views of human nature and destiny. A pious, even superstitious Anglican, he found little solace in a faith that left him with a profound sense of the misery of human existence in general, and of his personal shortcomings and culpability.

Having moved to London in 1737, he obtained employment on the *Gentleman’s Magazine* as a parliamentary reporter and hack writer. His poem *London*, a satire on the great city, appeared anonymously in 1738. In 1747 he was commissioned by London’s main booksellers to prepare a *Dictionary of the English Language*, in two volumes. It was eventually published in 1755. Johnson had no belief in any such thing as an ideal language; nevertheless, in his view it was one of the lexicographer’s duties to regulate and stabilize language.

Meanwhile in January 1749 he published *The Vanity of Human Wishes*, an imitation of the tenth satire of Juvenal and an expression of his own vision of human folly. Around the same time, his tragedy, *Irene*, was finally staged by his erstwhile pupil, David Garrick.

About a year later, he began to publish *The Rambler*, a magazine in the manner of the *Spectator* containing short essays on morals, manners and literature. This was followed in 1758 by *The Idler*. Johnson’s essays were sober, moralistic and instructive, commenting insightfully on human failings. In the meantime, he published a review of Soame Jenyns’ *A Free Enquiry into the Nature and Origin of Evil* (1757). Jenyns had argued, in a rather Leibnizian way, in favour of the doctrines of cosmic optimism and plenitude: all evils (for example, illness and disability) were only apparent, part of a larger scheme of benevolence created by a superfecund Deity. Johnson countered that Jenyns’ views were specious rationalizations that trivialized pain and suffering and discredited philosophy.

Meditations on malevolence and the vanity of human wishes formed the core of Johnson’s only extended work of fiction, *Rasselas*, published in 1759. Its protagonists, Rasselas and Imlac, Nekayah and Pekuah, leave their ‘happy valley’ in Abyssinia to see the world, but find only suffering, pride, fatuity and self-deception in a series of exploits that parallel *Candide*, the contemporary work by Voltaire. Johnson’s work ends in a ‘Conclusion, in which nothing is Concluded’. For rather traditional Christian reasons, Johnson could not easily subscribe to Enlightenment notions of human goodness and progress.

From the 1760s, Johnson was befriended by the Scot James Boswell. Thanks to Boswell, Johnson is remembered as a conversationalist, famous for his trenchant, peremptory, but often deeply humane, sentiments.

In the 1770s Johnson wrote as a political polemicist, first in defence of the British title to the Falkland Islands, and then on the side of the Crown in the dispute with the American colonies. *Taxation no Tyranny* shows Johnson as an authoritarian Tory, though in his off-the-cuff remarks he often supported the underdog against authority. In 1765 his edition of Shakespeare had appeared; the first four volumes of the *Lives of the Most Eminent English Poets* were published in 1779, the remaining six in 1781.

ROY PORTER

List of works

Johnson, S. (1750-52) *The Rambler*, London: Payne & Bouquet; ed. W.J. Bate and A.B. Strauss, New Haven, CT: Yale University Press, 1969, 3 vols.(Magazine containing short essays on morals, manners and literature.)
Johnson, S. (1755) *A Dictionary of the English Language*, London: Strahan, 2 vols.(Not, as sometimes imagined, the first major English dictionary, but the first with copious instances from literature of the proper use of the words in question.)
(1757) 'A Review of Soame Jenyns' *A Free Enquiry into the Nature and Origin of Evil*', in B. Bronson (ed.), *Samuel Johnson, Rasselas, Poems and Selected Prose*, San Francisco, CA: Rinehart Press, 1971, 219-28. (Johnson's attack on Jenyns was perhaps the most biting and devastating attack on the vacant moral optimism in fashion in the mid eighteenth century.)
(1758-60) *The Idler*, London: Newbery; ed. W.J. Bate, J.M. Bullitt and L.F. Powell, New Haven, CT: Yale University Press, 1963.(Together Johnson's *The Rambler* and *The Idler* essays take the form of the brief essay earlier pioneered by Addison and Steele and give the genre an altogether more serious moral twist.)
(1759) *The Prince of Abissinia: A Tale*, London: Dodsley; repr., ed. J.P. Hardy, Oxford: Oxford University Press, 1988.(Popularly known today as *Rasselas*. Rightly regarded as the English equivalent to Voltaire's exactly-contemporaneous *Candide*, Johnson's moral tale poses the question of the choice of life before ending with a conclusion in which nothing is concluded.).
(1779-81) *The Lives of the Most Eminent English Poets*, London: Bathurst; repr. ed. A. Waugh, London: Oxford University Press, 1952.(Johnson's book helped to establish the 'canon' of great English poetry while offering sometimes provocative judgments, as with his rather negative verdicts on Milton.)
(1968) *Selected Writings*, ed. P. Cruttwell, Harmondsworth: Penguin.(Includes Johnson's 1738 poem *London*, published anonymously and his *Vanity of Human Wishes* of 1749.)

References and further reading

Boswell, J. (1791) *The Life of Samuel Johnson*, ed. G.B. Hill, 6 vols, Oxford: Clarendon Press, 1934-50.(The classic biography.)
Chapman, R.W. (ed.) (1984) *The Letters of Samuel Johnson*, 3 vols, Oxford: Clarendon Press.(Though no copious correspondent, Johnson's letters convey the personal sources of his wider moral philosophy.)
Gross, G.S. (1992) *This Invisible Riot of the Mind: Samuel Johnson's Psychological Theory,* Baltimore, MD: University of Pennsylvania Press.(Explores Johnson's view of human nature as well as his own psychological weaknesses.)
Pierce, C. (1983) *The Religious Life of Samuel Johnson*, London: Athlone.(Emphasizes Johnson's strong religious guilt.)
Reddick, A. (1990) *The Making of Johnson's Dictionary 1746-73,* Cambridge: Cambridge University Press. (Contains useful material on the philosophical writings rifled by Johnson in compiling his *Dictionary*.)
Schwarz, R.B. (1975) *Samuel Johnson and the Problem of Evil*, Madison, WN: University of Wisconsin Press. (Contextualizes Johnson's attack on Jenyns.)
Wain, J. (1980) *Samuel Johnson*, London: Macmillan.(A lively and readable modern biography.)

Journalism, ethics of

It is sometimes suggested that ethical principles, even fundamental ones like nonmaleficence and beneficence, are totally out of place in journalism, and that it should be shaped solely by market forces. This suggestion should be resisted. One reason why journalism should be ethical is that in a democracy it is expected to serve the public interest, which means that it should accept the responsibility to circulate the information and opinion without which a democracy could not operate, and to enable it to do this the freedom of the press is acknowledged.

If journalism is to serve the public interest, then a commitment to truth-telling is fundamental. Journalists should also be fair and accurate in reporting news, should publish corrections, should offer a right of reply. They should avoid discrimination, deception, harassment, betraying confidences and invasions of privacy. But ethical journalism is more than lists of requirements and prohibitions. In investigative journalism, for example, some deception or intrusion into privacy could be justified in order to uncover corruption. Ethical journalism is therefore reflective understanding of the underlying principles of harm and benefit and the public interest, and an ability to apply them in particular cases.

1 The very idea of ethics in journalism

There are good reasons why journalism should be regarded as a practice governed by ethics, but also good reasons why this is problematic. It is expected of any group of practitioners who aspire to professional status that they base their practice in ethics, often by adhering to an explicit code of practice that specifies the expected behaviour in accordance with an acknowledged set of ethical principles ruled by the twin principles of nonmaleficence (do no harm) and beneficence (do good) (see Professional ethics). Those in breach of the professional code may be penalized in some way, with expulsion from the profession as the ultimate sanction. Clear examples are regulated professions like medicine, nursing, social work, law and teaching, in which the practitioner provides care or service for the client (patient, student) with individual and identifiable needs (see Medical ethics; Nursing ethics).

But can journalism be fitted into this model? Whereas it is easy to say that journalism too should be governed by nonmaleficence and beneficence - because these principles are all-encompassing - it is not so easy to see further analogies between journalism and the traditional professions, for two related reasons.

First, journalism is not a regulated profession, and perhaps not a profession at all: anyone who can get a relevant job in the media industry is a journalist. Second, even if it is a profession, journalism is certainly not a ‘caring’ profession with individual ‘clients’, but an occupation that takes place within a very competitive commercial framework. Journalists increasingly work for organizations which seek power and profits, and which are concerned with quantitative measures like audience figures and advertising revenue rather than with qualitative issues of ethics.

As a result the professional advancement of journalists depends (with rare exceptions) not on their adherence to ethical principles but on their contribution to the commercial success of their organization. Indeed, the ‘scoop’ is all, so success is more likely to follow from ignoring ethical considerations like confidentiality or privacy. This leads to the sceptical claim that there simply is no place for ethics in journalism. Even the principles of nonmaleficence and beneficence would be ludicrously out of place, because on this view journalism is motivated and shaped solely by market forces (see Market, ethics of the).

2 Journalism and democracy

But before the irrelevance of ethics to journalism can be confirmed, we need to look at the traditional political justification of ethical journalism, as this is ignored rather than refuted by the sceptics.

Liberal theory places a high value on freedom of thought and freedom of speech (see Freedom of speech). One argument for this is in terms of *rights*: rational individuals have a right to think as they find and to speak as they think. Another argument, due especially to Mill, is in terms of *consequences*: truth is beneficial to individuals and society, and it is most likely to be arrived at through free expression and open discussion, through the unfettered critical clash and contention of ideas (see Mill, J.S. §12).

Democratic theory adds to these goals and arguments (see Democracy). An informed citizenry is a necessary

condition of democracy, and this gives the media their special role with both rights and responsibilities. The rights are the privileges of press freedom, and the responsibilities are the duties to provide the accurate information and balanced comment without which people cannot fulfil their roles as rational citizens.

This link between free speech, freedom of the press and liberal democracy famously appears in the First Amendment to the Constitution of the United States of America: ‘Congress shall make no law…abridging the freedom of speech, or of the press’. It has since become enshrined in international charters such as the Universal Declaration of Human Rights: ‘Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media regardless of frontiers’ (Article 19).

To the extent that there is anything in this democratic justification of journalism, the abandonment of journalism to market forces must be rejected, and journalism emerges as an ethical practice. This is because it can serve the required democratic role only if it constitutes itself on the basis of ethical concepts like truth, objectivity and fairness. Journalism which is ethical in this sense rejects a purely commercial ethos and recognizes a commitment to serving the public interest.

The more specific principles of journalistic ethics, concerning objectivity, confidentiality and the like, follow from this overall commitment that journalism should do no harm to but should further the interests of a democratic polity and its individual citizens. Although there is scope for a legal framework on such matters as defamation, obscenity and privacy, legislation has a merely restrictive effect on freedom of expression and does not actively promote journalism of high quality. It is ethical engagement by journalists, and not legally backed censorship, which is a necessary condition of journalistic quality.

3 The ethics of journalism

Although ethical issues in journalism can be classified in various ways, the most straightforward is to distinguish between what is ethically required and what is forbidden. (This is not, however, an absolute distinction, and some injunctions, depending on how they are worded, can be in either category.)

On the side of what is required, it follows from the democratic justification that truth is fundamental, and this leads to the often-made claim that truth-telling is constitutive of journalism (see Truthfulness). Therefore, honesty and accuracy in presenting factual information is required, and, since no one is free from error, a commitment to publish corrections and to offer a right of reply to those who are criticized. In practice it is not as simple as this, as there is often no agreement about what is a factual matter or about what the facts in a particular case are, and a right of reply could easily be abused by anyone with a trivial complaint against the press.

There is also the problem that although journalists might have a commitment to the truth and nothing but the truth, it is impossible to publish the whole truth. Selection of material is always necessary, and so further principles like objectivity and impartiality, fairness and balance, are called up. It follows that there is never an algorithmic solution to ethical problems in journalism. The broad principles of honesty, objectivity, fairness and the like require both further specification and detailed consideration in the context of particular cases, and journalists must exercise ethically-informed judgments in the light of a commitment to the public interest.

The difficulties are compounded if we turn from factual journalism to comment and opinion, and especially to political journalism. Whereas public service broadcasting is expected to conform to the principles of impartiality and balance, the commercial press normally has a political commitment or ‘bias’. There is nothing wrong with individual journalists being partisan so long as they are honestly and openly partisan, but when does partisanship become manipulation or propaganda? And to what extent do editors have a responsibility to try to ensure that all varieties of opinion are represented in their paper? To what extent does the press as a whole have such a responsibility? It is in this area that the ethical ideals that lie behind the democratic justification of press freedom are diminished by commercial realities.

What is ethically forbidden in journalism includes sexism, racism and other forms of discrimination, deception, harassment, invasions of privacy, exploiting children who are in the news, and buying the stories of criminals (see Discrimination). Again, in practice these prohibitions are problematic, and a good deal of reflection on fundamental principles of good, harm and the public interest is required.

This is so especially in the area of the relations between journalists and their sources or potential sources. The confidentiality of willing sources, including whistle-blowers, should be respected, but investigative journalism presents special difficulties. Journalists often want to find out things which those who are involved prefer to keep hidden. Investigative journalism is part of the truth-telling function of journalism, and journalists have a responsibility to uncover scandal and corruption, especially where harm is being done to innocent victims. So investigative journalism can serve the public interest, and in doing so there may be justification for some deception or intrusion into privacy. But where in such cases to strike a balance is the most difficult issue in the ethics of journalism.

Privacy is especially problematic. Whereas it is easy to agree that journalists should not invade privacy, it is not at all easy to draw a satisfactory boundary between private and public (see Privacy). The truly private citizen who lives a life of public insignificance is in little danger from intrusive journalists, but sometimes such citizens are forced by good fortune (winning a lottery) or bad (being a disaster victim) into the public eye. In such cases requests for privacy should be respected and harassment frowned on. There is a difference between the public interest and what the public is interested in.

But should public figures in politics, business, entertainment, sport, be entitled to privacy? There is more than one issue involved here. Many such figures choose to live by publicity and are not in a morally strong position to object if the publicity is not exactly what they want. But are politicians, for example, entitled to a private life? The usual answer is to say that those areas of their lives that do not affect the performance of public duties are properly private, and that therefore financial scandals can legitimately be exposed in the media, but not sexual scandals. But it is doubtful whether this distinction, which involves dividing one person into two separate parts, the private and the public, is satisfactory on psychological or ethical grounds. And there is a further problem, in that appeals to ‘privacy’ can be question-begging. Investigative journalism often shows that what the corrupt politician claims is private is not private at all but is legitimately a matter of public concern.

What this brief survey of ethics in journalism shows is that there are no easy answers. Although it is possible to talk of what is required and what is forbidden in ethical journalism, more than a list of positive and negative injunctions is necessary. This is one reason why various formulations of codes of practice for journalists have not proved satisfactory. What is needed is a rational understanding of the underlying principles of harm and benefit, of individual rights and the democratic public interest. What is needed is the incorporation of reflective ethical practice into journalism, so that, as Klaidman and Beauchamp so well express it (1987), every journalist becomes ‘the virtuous journalist’.

See also: Applied ethics; Responsibilities of scientists and intellectuals

ANDREW BELSEY

References and further reading

Belsey, A. and Chadwick, R. (eds) (1992) *Ethical Issues in Journalism and the Media*, London: Routledge.(A collection of introductory essays on various aspects of media ethics.)

Belsey, A. and Chadwick, R. (1995) ‘Ethics as a vehicle for media quality’, *European Journal of Communication* 10: 461-73.(Part of a special issue on media ethics.)

Klaidman, S. and Beauchamp, T.L. (1987) *The Virtuous Journalist*, New York: Oxford University Press. (Discusses ethical journalism and its democratic justification in the context of the American Constitution’s guarantee of press freedom. The best introduction to media ethics.)

Lee, S. (1990) *The Cost of Free Speech*, London: Faber & Faber.(A close critical examination of the usual justifications of free speech.)

Stephenson, H. (1994) *Media Freedom and Media Regulation: An Alternative White Paper*, Birmingham: Association of British Editors, Guild of Editors and International Press Institute.(Argues against further legal restrictions on the press in Britain, especially a privacy law.)

Judah ben Moses of Rome (*c*.1292-after 1330)

Judah Romano, translator and Maimonidean philosopher, participated in the intellectual climate of Latin scholasticism, introducing a large number of the ideas of the Christian philosophers into Hebrew. He gained his knowledge in the various branches of philosophy through meticulous study of the writings of Aristotle and the scholastic commentaries to his works (particularly those of Albert the Great, Aquinas and Giles of Rome), of the Maimonidean-Tibbonian school, and of some Arabic philosophical treatises in Latin translation. Influenced by Albert the Great, Aquinas and Maimonides, he discussed all the major philosophical topics, and especially the idea of creation through intermediaries, the origins of the human soul, the cognitive process, which represents the real purpose of man, the nature of prophecy and prayer.

1 Creation

Judah may well have held an official position at Naples, at the court of Robert II of Anjou. Since Robert was then Senator of Urbe, one cannot exclude the possibility that he actually worked in Rome. The terminology of his exegetical writings almost entirely Aristotelian, but Judah fuses Aristotle's doctrines with others of Neoplatonic origin. His *Be'ur Ma'aseh Be-re'sit (Commentary on Genesis)* interprets philosophically the verses which open the Bible and argues that the creation cannot be proven apodeictically. Judah adopts the doctrine of creation through intermediaries: creation is seen as a process of emanation from the First Cause. The first emanation is light, identified with existence, which is the first perfection of a being (see Neoplatonism).

Judah goes on to develop a fully-fledged metaphysics of light and to discuss a number of astronomical questions regarding such as matters as the stellar motions, the positions and dimensions of the planets, and solar and lunar eclipses. The First Cause is called the subsisting Being. It is the simplest being, differing from created beings, which are compounded of essence and existence (in the case of the separate Intelligences) or of matter and form (in the case of the creatures of the sublunar world).

Judah also expounds on the creation of the vegetative, sensitive and rational souls, following the approach of Albert the Great: the rational soul represents the ultimate and most elevated degree of a dynamic process which starts from the active element, present in the potentiality of the prime matter. The creation of the rational soul, unlike that of the vegetative and sensitive souls, the causes of which are organic and natural factors joined to celestial virtue, takes place as a result of the direct intervention of the First Cause, through the influence of the 'active intellect'. It is because of the intellect, which is the substantial form of man, that man belongs to the superior world of the Intelligences. It is for this reason that man is said to be created 'in the image of God'. In agreement with Moses Maimonides, and contrary to the teachings of the Latin scholastics, Judah does not hold the active intellect to be the active element intrinsic in the the human rational soul, but rather the tenth separate Intelligence.

Adam, once created and endowed with an intellect capable of receiving the *intelligibilia*, was put under the dominion of the active intellect. Knowledge attained with the aid of the active intellect represents the ultimate purpose of man. Even after the expulsion from Eden, the purpose of man is to devote oneself to the *intelligibilia*, in order to be able to conjoin oneself with the active intellect in the highest degree of the cognitive process.

2 Prophecy

In *Sa'arim (Gates to the Books of Prophecy)*, Judah fuses Maimonides' and Aquinas' doctrines of prophecy, claiming for prophecy the character of a science on a par with other philosophical disciplines (see Aquinas, T.). He analyses its properties and assigns to it the peculiarities attributed by the Latin scholastics to theology, at the service of which the secular sciences are placed. The science of prophecy, which emanates from God's knowledge, differs from the natural sciences by its supernatural source of inspiration. It is the most elevated among the intellectual and ethical sciences with regard to veracity. The content of some prophetic revelations is superior to that of metaphysics, the highest science, although its final purpose, the ultimate realization of the human being, is not superior to that of ethics. Prophetic revelations were necessary in order to reveal supernatural truths to humanity. as well as certain natural truths, which would otherwise have remained concealed from the majority of mankind.

Judah analyses the content of the prophetic revelations and the purpose and nature of prophecy. He reaches a conclusion at a middle position between that of Maimonides and that of Aquinas: the prophet is a philosopher, prepared in all the branches of knowledge, whose rational and imaginative faculties are perfect. The preparation of the prophet is a prerequisite for prophecy, but prophecy is not a purely natural phenomenon. Some prophecies are natural and others miraculous: in the natural revelations the prophet perceives the content of the revelation, which is similar to that of the various branches of science, by means of a cognitive intellectual-imaginative process. The content of the miraculous revelations, however, surpasses human capacities, or the prophetical cognitive process takes place without the mediation of the imagination.

Prophecy consists of: (1) intellection, (2) transmission of the prophetic message to others (3) the performing of miracles. Only the first element, however, is necessary; by means of the other two elements, the prophet functions as a guide. For Judah, influenced by Maimonides, prophecy is an inspiration from the First Cause, through the mediation of the active intellect, to the intellect and then to the imaginative faculty of the prophet (see Prophecy).

3 Prayer and the universe

In *Be'ur ha-Qaddis we-ha-Qedusah (Commentary to the Liturgy),* Judah interprets philosophically the key-words of prayer, using the *Liber de Causis (Book of Causes)* as his guide. The First Cause is the universal, infinite, most simple and transcendent cause, beyond all definition; all beings, according to their nature, are influenced and 'preserved' by it. The universe is organized hierarchically, and the world of the Intelligences exerts a real causal impact on the sublunar world. The First Cause is the cause of eternity; the Intelligences are eternal - that is, timeless and motionless - and the human soul, identified by Judah as the 'noble soul', belongs by nature to the eternal intellectual world, although operating within time.

Through a rationalistic analysis in which he is influenced by Aquinas, Judah discerns two components of prayer: (1) the uplifting of the human intellect to God by means of the cognitive process and (2) the 'request of suitable things': the content of the prayer expresses the agreement of the will of the human being with his intellectual representation, whereby the individual renders his words ('the external language') suitable and asks God for suitable things. Through prayer supplicants reach the point where their intellect conforms to the divine will and, thanks to the perfection of their intellection, are the cause of the acceptance of their own prayer. Concerning the distinction between individual prayer and public prayer, Judah asserts that the former is not necessarily expressed in words: for one who has reached a high level of perfection, prayer is an intellectual act, superior to any form of prayer expressed by means of the voice. Individual prayer, in its most elevated expression, is then a cognitive and contemplative act. Public prayer, however, is necessarily 'vocal', since the officiant in the synagogue prays for the whole community. Such prayer has the 'political' function of strengthening the faith of the masses.

Following Albert the Great, Judah stresses repeatedly that the individual must advance gradually in the acquisition of the *intelligibilia*, in order to achieve the highest level of intellection. The most elevated intelligible concepts are, in fact, God and the separate Intelligences. The real purpose of man is to 'purify' his intellection. This purpose can be partially realized starting in this life, while the soul is united to the body. It is completely attainable, however, only in the hereafter. Then the human rational soul, free from corporeal bonds, will unceasingly enjoy the *intelligibilia* gleaned during this life. This will constitute its eternal happiness. For its condition, at that point, will be like that of the separate Intelligences.

See also: Albert the Great; Aquinas, T.; Maimonides, M.; Neoplatonism; Prayer; Prophecy

CATERINA RIGO

List of works

Judah Ben Moses of Rome (*c*.1320-50) 'Sancti Thomae de Aquino *Opusculum de ente et essentia* a Rabbì Jehudàh ben Mosèh ben Dani'èl Romano primum hebraice redditum (saec. XIV incipiente)', ed. G. Sermoneta, in A.Z. Bar-On (ed.) *From Parmenides to Contemporary Thinkers: Readings in Ontology*, Jerusalem, I: The Magnes Press, 1978, vol. 1: 184-214.(Judah's version of an Aquinas text.)

Judah Ben Moses of Rome (*c*.1320-50) 'Il *De substantia orbis* di Averroè: edizione della versione latino-ebraica con commento di Yehudah b. Mosheh Romano', ed. with introduction and commentary by C. Rigo, Ph.D. dissertation, Turin: Università degli Studi di Torino, 2 vols, 1989-92.(Critical edition of Judah's translation of Averroes' *De substantia orbis*.)

References and further reading

Rigo, C. (1993a) ‘Un antologia filosofica di Yehuda b. Mosheh Romano’, *Italia* 10: 73-104.(A philosophical anthology of Judah ben Moses.)

Rigo, C. (1993b) ‘Yehudah b. Mosheh Romano traduttore di Alberto Magno (commento al *De Anima* III, II, 16)’, *Henoch* 15: 65-91.(A critical edition and analysis of two different translations of a passage from Albert the Great’s commentary on Aristotle’s *De Anima*.)

Rigo, C. (1993c) ‘Le traduzioni dei commenti scolastici al *De Anima* eseguite da Yehudah b. Mosheh nella tradizione filosofica ebraico-italiana dei secoli XIII-XIV’, in F. Vattioni (ed.) *Atti della VII settimana (Roma, 25-30 novembre 1991): Sangue e antropologia nel Medioevo*, Rome: Pia Umione Preziosissimo Sangue, 1073-95. (Account of his translations of scholastic commentaries to Aristotle’s *De Anima*.)

Rigo, C. (1994) ‘Egliolio Romano nella cultura edraica: le versioni di Yehudah b. Mosheh Romano’, *Documenti e studi sulla tradizione filosofica medievale* 5: 397-437.(Examines Judah’s role as a translator of Giles of Rome.)

Rigo, C. (1995) ‘Yehudah b. Mosheh Romano traduttore degli Scolastici latini’, *Henoch* 17: 141-70.(Focuses on Judah as a translator of Latin scholastics.)

Rigo, C. (1996) ‘The Be’urim on the Bible of R. Yehudah Romano: the Philosophical Method which comes out of them, their sources in the Jewish Philosophy and in the Christian Scholasticism’, Ph.D. dissertation, Jerusalem: The Hebrew University of Jerusalem, 2 vols.(A study of Judah’s philosophical interpretations of biblical verses.)

Sermoneta, G. (1965) ‘La dottrina dell’intelletto e la "fede filosofica" di Jehudàh e Immanuel Romano’, *Studi Medievali*, series 3, 6 (2): 3-78.(Comparison of his doctrine of conjunction with the active intellect with that of Immanuel of Rome.)

Sermoneta, G. (1969) ‘The Commentary to the First Weekly Reading in *Genesis* by Judah Romano’, *Fourth World Congress of Jewish Studies* 2: 341-3.(An analysis of Judah’s commentary to the first weekly portion in Genesis.)

Sermoneta, G. (1976) ‘Yehudah and Immanuel ha-Romi, "Rationalism Culminating in Mystical Faith"’, in M. Hallamish and M. Schwarz (eds) *Revelation, Faith, Reason*, Ramat Gan: Bar Ilan University, 54-70.(Analyses Judah’s doctrine of conjunction with the active intellect.)

Sermoneta, G. (1980) ‘Jehudah ben Moše ben Danìel Romano, traducteur de Saint Thomas’, in G. Nahon and Ch. Touati (eds) *Hommage à Georges Vajda: Études d’histoire et de pensée juive*, Louvain: Éditions Peeters, 235-62.(Looks at Judah’s role as a translator of Aquinas.)

Sermoneta, G. (1984) ‘Prophecy in the Writings of R. Yehuda Romano’, in I. Twersky (ed.) *Studies in Medieval Jewish History and Literature*, Cambridge, MA: Harvard University Press, vol. 2, 337-74.(A study of Judah’s doctrine of prophecy.)

Sermoneta, G. (1990) ‘"Thine Ointments Have a Goodly Fragrance": Rabbi Judah Romano and the Open Text Method’, in ‘Shlomo Pines Jubilee Volume on the Occasion of his Eightieth Birthday’, *Jerusalem Studies in Jewish Thought* special issue 9 (2): 77-113.(Discussion of his hermeneutical approach.)

Sermoneta, G. (1994) ‘The Light - its Definition and Function in the Commentary to the First Reading in Genesis by R. Yehudah ben Mosheh ben Daniel Romano’, M. Oron and A. Goldreich (eds) *Massu’ot: Studies in Kabbalistic Literature and Jewish Philosophy in memory of Prof. Ephraim Gottleib*, Jerusalem: The Bialik Institute, 343-60.(Examines Judah’s metaphysics of light.)

Steinschneider, M. (1870) ‘Guida Romano’, *Il Buonarroti*, series 2, 5: 3-12.(Account of his works.)

Jung, Carl Gustav (1875-1961)

Jung was among the leaders in the development of depth psychology at the beginning of the twentieth century. An early follower of Sigmund Freud, he broke with the founder of psychoanalysis in 1913 and established his own school of analytical psychology.

Jung's theoretical development originated in his work on the word association test and the theory of feeling toned complexes. As he continued to explore the workings of the unconscious, he postulated the existence of instinctual patterns of cognition and behaviour which he termed 'archetypes'. Archetypal patterns are, according to Jung, common throughout the human species and constitute an inherited 'collective unconscious'.

Jung's approach to psychology was eclectic. He accepted the psychological importance of any phenomenon, even if it conflicted with current thinking in other fields. This attitude led to a deep investigation of the psychological significance of occult phenomena and alchemy, which Jung viewed as expressions of the unconscious that anticipated modern psychology. Later in life, Jung turned increasingly to considerations of the contemporary cultural expressions of psychological forces, writing extensively on what he viewed to be a deepening spiritual crisis in Western civilization.

1 Early work and relationship with Freud

Jung's father was a Protestant minister and Hebraist, while his paternal grandfather had been a distinguished physician and university reformer. On his mother's side, Protestant theologians predominated. Given this milieu, Jung was exposed, from an early age, to the main currents of nineteenth-century theology and philosophy, most notably the works of Kant, Schopenhauer and Nietzsche. Jung attended gymnasium in Basel and went on to medical school at the University of Basel. Following graduation, he undertook postgraduate training in psychiatry at the Burghölzli Hospital in Zurich, under the direction of Eugen Bleuler. At about the same time, Jung married Emma Rauschenbach, the daughter of a wealthy German industrialist. They had five children together. From 1906 to about 1913, Jung was deeply involved with Sigmund Freud in the development of psychoanalysis. Following their break, Jung undertook to develop his own theories about the workings of the psyche. The theory of archetypes and the collective unconscious, for which he is best known, prompted Jung to travel widely, after the First World War, in search of evidence to support his understanding of the commonality of psychological functioning throughout the species. In the 1930s, Jung was briefly in contact with figures connected to the National Socialist takeover of German medicine. He soon broke contact, however, and deeply regretted having had any relationship with them.

Jung was surrounded, throughout his adult life, by a large circle of students and admirers. Eventually, this circle developed into the C.G. Jung Institute at Kusnacht, outside Zurich, where Jung had his home.

Although Jung is best known for his theories of the archetypes of the collective unconscious, he first attracted international attention for his work on the word association test. The test was conducted by reading off a list of stimulus words and noting the characteristics of the responses on the part of the subject. Jung significantly refined the measurement of the anomalies in association, such as delayed response, cardiopulmonary function, and galvanic skin response. These refinements allowed for more precise interpretation of test responses. Out of this experimental work, but also under the influence of Pierre Janet in France and Théodore Flournoy in Switzerland, Jung developed the theory of 'feeling toned complexes', or affect-laden automatisms within the larger psychic structure of the individual. This theory formed the foundation for all subsequent developments in Jung's system of psychology.

The refined word association test provided the first experimental evidence supporting Sigmund Freud's theory of repression. In 1906, Jung wrote to Freud, sending copies of his most recent work on the test, only to learn that Freud already had the papers. Thus began one of the most famous and ill-fated collaborations in modern science.

Although a host of factors contributed to the eventual break between Jung and Freud, several points of theory distinguished the two from the beginning. Freud based his work on a definition of libido that was exclusively sexual in nature, and the conviction that psychopathology derived from infantile sexual conflicts. The task of psychoanalysis was to uncover the sexual etiology of the disorder and thereby dissipate its effect. Jung's work

with the association test, however, had convinced him that libido was not exclusively sexual and that not all complexes could be reduced to a single etiologic source.

Other points of theory further separated the two. Based on work done in his doctoral dissertation on a case of somnambulism or mediumship, Jung argued that at least some of the workings of the unconscious were motivated by the anticipated development of the individual. This teleological view of the unconscious prompted Jung to question the role of *repression*, which for Freud was the fundamental mechanism of psychodynamics, in the psychic economy. For Jung, *projection* of psychic contents from the unconscious, without prior repression, took on increasing significance. Finally, based on his extensive experience with dementia praecox (schizophrenia) at the Burghölzli hospital, Jung concluded that there were what he termed ‘psychological dominants’ at work in all individuals, and that these dominants were evolutionary in origin, and relatively consistent throughout the species. Although Freud believed that the Oedipus complex was common throughout humanity, he resisted the notion that there could be other complexes with equal status.

2 The theory of archetypes and personality theory

Following the break with Freud in 1913, Jung went into a period of self-analysis and theory building. At this time, Jung developed his method of active imagination which used the images presented in dreams and waking reveries to explore the individual psyche. Out of this work with the imagination he developed his characteristic vocabulary for the structure of the psyche, and his more refined understanding of the psychological dominants, now referred to as ‘archetypes’. The consciously presented personality was referred to as the ‘persona’, from the mask worn by actors in the ancient Greek theatre. Those parts of the personality that were not allowed into conscious presentation formed the individual’s ‘shadow’.

In addition to persona and shadow, Jung posited the existence of images of the counter-sexual in both men and women. These he referred to as ‘anima’ and ‘animus’, respectively. The images of the counter-sexual formed a bridge between the individual unconscious and the ‘collective unconscious’ which contained the archetypes. The term collective unconscious refers to the inherited instinctual substrate of human psychological functioning.

The theory of archetypes has been the subject of various interpretations, and Jung was not always precise in his application of the term to a psychological event. One common misconception, however, is that Jung thought of the archetypes in purely Platonic terms, as existing in some transcendent realm of ideas. Much of the confusion arises from a failure to recognize Jung’s distinction between *archetypes* and *archetypal images*. Although there are places where Jung appears to discuss the archetypes in transcendental terms, perhaps informed by his extensive reading of Kant, his most consistent position is that the archetypes are biologically inherited, instinctual patterns of cognition and behaviour. It is this level of shared cognitive and behavioural patterns that constitutes the *collective unconscious*.

‘Archetypal images’, on the other hand, are more or less conscious representations of the archetypal or instinctual structures. Thus, instinctual seeking of the breast in the new-born infant demonstrates an archetypal form of behaviour associated with the mother. The instinctual relationship to the mother, however, also drives the formation of individual fantasies and images representative of the infant’s experience of the mother, and the culture at large contributes to the further elaboration of these fantasies by means of myths, rituals, fairy tales and other forms of collective representation associated with the mother.

Jung was also at work on a theory of personality or ‘psychological types’. Jung posited two essential orientations toward the world, introversion and extroversion, to which were added the functions of thinking, feeling, intuiting and sensing. In essence, the extrovert focuses on the outside world for information and inspiration, while the introvert focuses on interior states. By the same token, the functions are paired in such a way that one operates consciously while the other works unconsciously. Thus a thinking type will have an unconscious feeling function that attaches itself to contents of the unconscious.

3 Occultism, alchemy and projection

Jung had a lifelong interest in occultism and esoteric traditions. He was by no means alone in this interest, as research on spiritualism was widespread at the time. Jung and Freud both met with William James during a trip to the USA in 1909, and James came away deeply impressed by Jung, in part, no doubt because of their shared

interest in spiritualism. For his part, Jung considered James to be a profound influence on his development.

Interest in how causally unrelated events could be meaningfully connected, as when a series of numbers seem to repeat themselves several times during a single day, led Jung to develop his theory of synchronicity. As with many of Jung's theories, it is difficult to distinguish empirical research from the formulation of a theory of psychological functioning. Thus, on its surface, the theory of synchronicity attempts to provide statistical measures of some paranormal phenomena. By and large, however, these measures are not significant. On the other hand, the role of projection in the formation of meaningful associations is clearly demonstrated. This understanding of the primacy of projection in psychic functioning laid the foundation for the development of such standard projective tests as the Rorschach and the Thematic Apperception Test (TAT).

Jung's long-standing fascination with esoteric traditions, particularly Gnosticism (see Gnosticism), led him into an extensive study of alchemy. Alchemy was not, on Jung's reading, a precursor to modern natural science. Rather, it was essentially a Western form of psychological and spiritual discipline resembling the great meditative traditions of India and East Asia. The operations of the alchemists in their work and writings were manifestations of psychic projection, and the search for a means to transform base metals into gold was a spiritual quest for self-perfection (see Alchemy).

The investigation of projection eventually led Jung to consider the epistemological implications of his system of psychology. To the extent that projection from the unconscious informed his interpretation of cognitive phenomena, it became increasingly difficult to maintain a purely empirical view of those phenomena. In the end, Jung spoke of a 'psychoid world' where the boundaries between psychic events and objective reality break down.

4 The critique of culture

While implicit in some aspects of his early work, Jung's late work (1939-61) focused increasingly on what he saw as a crisis in Western civilization. In many respects, Jung's critique resembles those of Husserl and Heidegger, although it focuses more on issues of spiritual collapse and renewal. Additionally, Jung thought that it was possible to assume a therapeutic stance in relation to the cultural crisis.

Jung focused his critique of culture, to a substantial degree, on the historical role of Christianity. He approached this problem from within the frame of his psychological theories, however, rather than from either a historical or theological standpoint. This meant that the body of historical and textual materials that make up the Christian tradition were treated, by Jung, much as he would treat the dreams and fantasies of a patient. Jung viewed the tradition as increasingly maladaptive, and therefore in need of therapy. At the same time, he viewed such doctrinal moves as the proclamation of the bodily assumption of the Virgin as attempts to overcome maladaptive aspects of dogmatic Christianity by incorporating previously denigrated aspects of human experience such as the female body.

Jung exercised an influence on the practice of psychotherapy and on fields ranging from literature to industrial psychology that far exceeds what is usually recognized. The circumstances surrounding his break with Freud, as well as his brief association with the Nazis, did considerable damage to his reputation. An idiosyncratic vocabulary added to his image as an undisciplined thinker. Nevertheless, his system of psychology, beginning as it does with the word association test, and emphasizing the biological foundations of cognition and behaviour, thereby anticipating more recent developments in cognitive theory and evolutionary psychology, rests on far stronger empirical foundations than most other depth psychologies.

See also: Indian and Tibetan philosophy; Psychoanalysis, methodological issues in; Psychoanalysis, post-Freudian; Psychology, theories of; Reductionism in the philosophy of mind

GEORGE B. HOGENSON

List of works

Jung, C.G. (1953-91) *The Collected Works of C.G. Jung*, eds G. Adler, M. Fordham, and H. Read, executive ed. W. McGuire, trans. R. Hull, Princeton, NJ: Princeton University Press, and London: Routledge.(The standard collection of Jung's works.)

Jung, C.G. (1973) *Memories, Dreams, Reflections*, recorded and edited by A. Jaffé, trans. R. and C. Winston, New York: Pantheon Books.(Jung's autobiography, important not only for the light it sheds on his life, but also

as a study in autobiographical technique.)
Jung, C.G. (1974) *The Freud/Jung Letters: The Correspondence Between Sigmund Freud and C.G. Jung*, ed. W. McGuire, trans. R. Manheim and R. Hull, Princeton, NJ: Princeton University Press.(The indispensable source on the relationship between Jung and Freud.)
Jung, C.G. (1988) *Nietzsche's Zarathustra: Notes on the Seminar Given in 1934-1939 by C.G. Jung in Two Volumes*, ed. J. Jarrett, Princeton, NJ: Princeton University Press. (A transcription of Jung's Nietzsche seminar. Of interest mainly as an example of Jung at work on a text and in dialogue with his students.)

References and further reading

Brome, V. (1978) *Jung*, New York: Atheneum.(The most reliable survey of Jung's life.)
Brooke, R. (1991) *Jung and Phenomenology*, London: Routledge.(A clear and compelling view of Jung in relation to the phenomenological tradition.)
Franz, M.-L., von (1975) *C.G. Jung: His Myth in Our Time*, trans. W. Kennedy, New York: G.P. Putnam's Sons. (An intellectual biography of Jung by one of his closest followers. While often hagiographic, it nevertheless captures the more spiritual aspects of Jung's work.)
Hogenson, G. (1994) *Jung's Struggle with Freud*, Wilmette: Chiron Publications, revised edn.(A view of the philosophical implications of the conflict between Jung and Freud with particular emphasis on the problem of authority.)
Homans, P. (1979) *Jung in Context: Modernity and the Making of a Psychology*, Chicago, IL: University of Chicago Press.(A view of Jung's development in light of modern psychoanalytic theory, particularly Kohut's self-psychology.)
Kerr, J. (1993) *A Most Dangerous Method: The Story of Jung, Freud, and Sabina Spielrein*, New York: Alfred A. Knopf.(A historical overview of the conflict between Jung and Freud which also provides a discussion of the important role played by Sabina Spielrein in the early development of psychoanalysis. Contains an excellent annotated bibliography.)
Maidenbaum, A. and Martin, S. (eds) (1991) *Lingering Shadows: Jungians, Freudians and Anti-Semitism*, Boston, MA: Shambhala.(A collection of essays concerned with Jung's alleged anti-Semitism and his relation to the Nazis.)
Samuels, A. (1985) *Jung and the Post-Jungians*, London: Routledge & Kegan Paul. (A survey of Jung's theories and of subsequent developments in Jungian psychology. Contains a comprehensive bibliography.)
Stein, M. (1985) *Jung's Treatment of Christianity: The Psychotherapy of a Religious Tradition*, Wilmette: Chiron Publications.(An analysis of Jung's hermeneutics and his critique of culture.)
Young-Eisendrath, P. and Dawson, S. (eds) (1997) *The Cambridge Companion to Jung*, Cambridge: Cambridge University Press.(A comprehensive set of essays on Jung and developments in Jungian theory.)

Jungius, Joachim (1587-1657)

Joachim Jungius was one of the most important seventeenth-century reformers of Aristotelian logic. Through critical assessment of Suárez and by recourse to Ramus, Zabarella and Melanchthon, he tried to replace Aristotelian syllogistics with a logic based on empirical judgment. This differed sharply from other contemporary attempts to reform logic in Protestant Europe. Jungius was a pioneer in the development of the modern concept of 'element' in chemistry, and made an important contribution to the classification of plants.

1 Life and works

As a schoolboy in the German city of Lübeck, Jungius showed his exceptional talent by teaching his classmates the dialectic of Petrus Ramus. Thus in 1606, when he began to study logic and metaphysics in Rostock, his stance towards the neoscholastic system of Suárez and the logic of Coimbra was already influenced by his thorough knowledge of Ramus (see Collegium Conimbricense). Soon he turned to mathematics and continued his studies at the recently-founded University of Giessen, receiving a Master's degree in 1608. Having taught there as a professor of mathematics, he moved back to Rostock in 1616, began to study medicine, and - after a scholarly tour through Europe - qualified as *Doctor medicinae* in Padua. There he worked mainly on botany and zoology before returning to Rostock in 1619. He joined Johann Valentin Andreae's circle and helped to found the *societas ereunetica/zetetica* (*c*.1623). However, this society lasted only until 1625 because of the Thirty Years' War. Its aims were on the one hand the rejection of Jesuit metaphysics and the critique of sophistics, on the other the promotion of mathematical studies and a natural science regulated by observation and experience. The relation between *ratio* and *experientia* should be determined in such a way that the sciences could be based on demonstrable certainty. After professorships in mathematics in Rostock (1624) and medicine in Helmstedt (1625), work as a doctor in Braunschweig and return to his professorship in Rostock (1626), Jungius took up simultaneously the rectorships at the Johanneum and the academic *Gymnasium* in Hamburg. Disputes with the politically influential Lutheran orthodoxy about the style of New Testament Greek forced him to retire as rector of the Johanneum in 1640.

Jungius did not publish much during his lifetime, but his ideas were promoted to a certain extent by Comenius and by the British mathematician Charles Cavendysshe. His immediate influence was in part based on the disputations and dissertations he supervised when rector at the academic *Gymnasium* (for example on stoicheiometry, doctrine of physical principles, theory of motion and other subjects). In 1691 all but about a quarter of Jungius' originally voluminous posthumous papers were destroyed by fire (Meinel 1984b).

2 *Logica Hamburgensis*

Written in the early 1630s as a textbook for the academic *Gymnasium*, the *Logica Hamburgensis* is based on the idea of a continuous improvement of the sciences. In this respect it directly follows Francis Bacon's concept of science and the pedagogical-utopian ideas of Johann Valentin Andreae (see Bacon, F.). According to Jungius, one can discern a progress in the sciences despite the Deluge, the migration of nations and the Thirty Years' War, and this progress is due to the improvement of logic. As the editor of the second edition explained, the aim of the *Logica Hamburgensis* was to complete and amend Aristotelian logic - a very accurate description of the syncretistic character of Jungius' logic, which links up conceptually with Zabarella's definition of logic as 'habitus mentis organicus (instrumentalis)' (see Zabarella, J. §2) and with Ramus' concept of logic as the art of proper reasoning. Accordingly, the structure of the *Logica Hamburgensis* follows the Ramistic distinction between *inventio* and *iudicium* which Jungius relates to the three activities of the human mind, namely (1) *notio* or *conceptus*, (2) *enuntiatio* and (3) *discursus* or *ratiocinatio*. Following this basic structure, Jungius deals in the first three books (*Logica generalis*) with the doctrine of the predicables (*De notionibus*), the doctrine of the predicables' modes of expression (*De enuntiatione*), and finally their application to syllogistic figures (*De ratiocinatione*). Books 3 to 6 (*Logica specialis*) deal with the range of our a posteriori cognition within the limits of apodictic certainty (*Logica apodictica*), with dialectics, and sophistics (particularly the theory of fallacies). Given this heterogeneous internal structure, the merit of the *Logica Hamburgensis* lies certainly in its strongly epistemological concept of logic and the value it ascribes to empirical concepts. Although Jungius' logic was sometimes harshly criticized for its attempt at a moderate reform of Aristotelian syllogistics, it remained the basis

for the teaching of logic at the academic *Gymnasium* in Hamburg for many decades and even influenced Leibniz's studies of logic.

3 Writings on natural history

Jungius' writings on natural history, based on lectures and published posthumously, are of an even more didactic nature than his logic textbook. Unlike Descartes, who had tried to found philosophy anew seemingly without any presuppositions, Jungius apparently thought that progress in philosophy through the study of nature could be achieved only by examination of traditional doctrines. By critically assessing the physics of Zabarella, Jean Fernel, Daniel Sennert and others, Jungius defined natural science in the *Doxoscopiae physicae minores* (1662) as a mode of cognition which does not adjust the phenomena to preconceived opinions but the hypotheses to the phenomena. In the Aristotelian-Theophrastic-Galenic tradition, on which he explicitly based his methods, he replaced - in contrast to the metaphysics of Coimbra - the question of the essence of things with that of their function and mode of action. According to Jungius, the natural scientist does not define 'change' or 'motion', but rather what it means 'to be changed' or 'to be moved'. Therefore he focused on the concept of a *materia substantialis* which, together with a critical assessment of the ancient doctrine of principles, Anaxagoras' homoeomeries, Democritus' theory of atoms, and other theories, led him to the discovery of material 'elements'. This theory of elements was not only different from the ancient doctrine of the four elements and the hermetic doctrine of the three principles, but also from Robert Boyle's corpuscular theory, and was a major step towards the modern notion of element.

Among his contributions to almost every aspect of natural history and its applications, from mineralogy to entomology, chemistry, mining, mechanics and so on, Jungius established a classification of plants into classes, genera and species according to a uniform principle (*ratio constans*). He made a distinction between specific differences (for example, leaves) and accidental differences which have no influence on the definition of the species (such as spines, smell, medical use, place, time of germination, number of petals and so on), thereby anticipating Linnaeus' system of classification.

See also: Aristotelianism in the 17th century

Translated from the German by Peter Schnyder
RALPH HÄFNER

List of works

Jungius, J. (1613) *Kurtzer Bericht von der Didactica, oder Lehrkunst Wolfgangi Ratichii (Short treatise concerning the Didactica, or Art of Learning, according to Wolfgang Ratichius)*, Frankfurt.(Written with Christoph Helvich, this is a positive assessment of Wolfgang Ratichius' inductive method of teaching; at least five editions were published up until 1621.)

Jungius, J. (1627) *Geometria empirica (Empirical geometry)*, Rostock.(Early work on geometry.)

Jungius, J. (1638) *Logica Hamburgensis, hoc est, institutiones logicae in usum scholae Hamburgensis conscriiptae (Logica Hamburgensis, that is institutions of logic, written for the use of the Academic Gymnasium in Hamburg)*, Hamburg: Barthold Offermans; 2nd edn, ed. J. Vagetius, Hamburg, 1681; German trans. R.W. Meyer, Hamburg: Augustin, 1957.(Books 1-3 had already been published in 1635.)

Jungius, J. (1662) *Doxoscopiae physicae minores, sive isagoge physica doxoscopica. In qua praecipue opiniones in physica passim receptae breviter quidem, sed accuratissime examinantur (The Pupils of physical doctrine, or, Introduction to the physical doctrines, in which the traditional opinions in physics are shortly, but accurately examined)*, ed. M. Fogel, Hamburg: Johannes Nauman.(Reprinted Hamburg, 1679.)

Jungius, J. (1678a) *Harmonica (Harmony)*.(Investigates the doctrines of harmony and introduces the theory of music.)

Jungius, J. (1678b) *Isagoge phytoscopica (Introduction to the doctrines and the theory of plants)*, Hamburg. (Investigates the received doctrines of plants on the basis of empirical observations.)

Jungius, J. (1691) *Historia vermium (Natural history of worms and insects)*, ed. J. Vagetius, Hamburg.(Contains mainly contributions to entomology.)

Jungius, J. (1747) *Opuscula botanico-physica (Short treatises on botany and natural history)*, ed. M. Fogel and J. Vagetius, Coburg: Georg Otto.(Collection of Jungius' botanical studies, based on his private lectures. Reproduces some pieces already published in his *Doxoscopiae physicae minores*.)

Jungius, J. (1863) *Briefwechsel mit seinen Schülern und Freunden. Ein Beitrag zur Kenntniß des großen Jungius und der wissenschaftlichen wie sozialen Zustände zur Zeit des dreißigjährigen Krieges, aus den Manuscripten

der Hamburger Stadtbibliothek zusammengestellt (Correspondence with his pupils and friends. Information on Jungius and the scientific and social circumstances at the time of the Thiry Years' War, compiled from manuscripts in the Hamburg City Library), ed. R.C.B. Avé-Lallemant, Lübeck: Friedrich Asschenfeldt. (Contains an essential part of Jungius' correspondence.)

Jungius, J. (1977) *Logicae Hamburgensis additamenta*, ed. W. Risse, Veröffentlichungen der Joachim-Jungius-Gesellschaft der Wissenschaften, BD 25, Göttingen:Vandenhoeck & Ruprecht.(Unpublished additions to the *Logicae Hamburgensis*.)

Jungius, J. (1982) *Praelectiones physicae*, ed. C. Meinel, Veröffentlichungen der Joachim-Jungius-Gesellschaft der Wissenschaften, BD 45, Göttingen:Vandenhoeck & Ruprecht.(Critical edition of Jungius' lectures on physics.)

Jungius, J. (1988) *Disputationes Hamburgenses*, ed. C. Müller-Glauser, Veröffentlichungen der Joachim-Jungius-Gesellschaft der Wissenschaften, BD 59, Göttingen:Vandenhoeck & Ruprecht.(Critical edition of Jungius' disputations and dissertations.)

References and further reading

Avé-Lallemant, R.C.B. (1882) *Das Leben des Dr. med. Joachim Jungius aus Lübeck (The life of Doctor of medicine Joachim Jungius from Lübeck)*, Breslau: Hirt. (Interesting work on Jungius' Life.)

Fogel, M. (1657) *Memoria Joachimi Jungii (Memorial to the late Joachim Jungius)*, Hamburg.(First biographical survey, published at the time of Jungius' death.)

Guhrauer, G.E. (1850) *Joachim Jungius und dein Zeitalter. Nebst Goethe's Fragmenten über Jungius (Joachim Jungius and his time; including Goethe's fragments on Jungius)*, Stuttgart and Tübingen: J.G. Cotta.(Thorough study of Jungius' life, work and its reception.)

Kangro, H. (1968) *Joachim Jungius' Experimente und Gedanken zur Begründung der Chemie als Wissenschaft. Ein Beitrag zur Geistesgeschichte des 17. Jahrhunderts (Joachim Jungius' experiments and thoughts to found chemistry on scientific grounds. A contribution to the intellectual history of the seventeenth century)*, Boethius, Bd. VII, Wiesbaden: F. Steiner.(Important work with a very helpful bibliography.)

Kangro, H. (1969) 'Joachim Jungius und Gottfried Wilhelm Leibniz. Ein Beitrag zum geistigen Verhältnis beider Gelehrten'(Joachim Jungius and Gottfried Wilhelm Leibniz. A contribution to the intellectual relationship between them), *Studia Leibnitiana* 1: 175-207.(Study of the influence of Jungius on Leibniz.)

Meinel, C. (1982) 'Der Begriff des chemischen Elementes bei Joachim Jungius' (The concept of the chemical element in Joachim Jungius' thought), *Sudhoffs Archiv* 66: 313-38.(Important contribution to the notion of a chemical element.)

Meinel, C. (1984a) *In physicis futurum saeculum respicio. Joachim Jungius und die Naturwissenschaftliche Revolution des 17. Jahrhunderts (In physics I see the coming age. Joachim Jungius and the scientific revolution in the seventeenth century)*, Veröffentlichungen der Joachim Jungius-Gesellschaft der Wissenschaften, Bd. 52, Göttingen: Vandenhoeck & Ruprecht.(Investigates Jungius' contribution to the history of science.)

Meinel, C. (1984b) *Der handschriftliche Nachlaß von Joachim Jungius in der Staats- und Universitätsbibliothek Hamburg (The manuscript heritage of Joachim Jungius in the State and University Library of Hamburg)*, Katalog der Handschriften der Staats- und Universitätsbibliothek Hamburg, Bd.IX, Stuttgart: Hauswedell.(Catalogue of Jungius' remaining manuscripts.)

Meinel, C. (1992) *Die Bibliothek des Joachim Jungius. Ein Beitrag zur Historia litteraria der frühen Neuzeit (The library of Joachim Jungius. A contribution to the historia litteraria in the early modern period)*, Veröffentlichungen der Joachim Jungius-Gesellschaft der Wissenschaften, Bd.67, Göttingen: Vandenhoeck & Ruprecht. (Study of Jungius' library.)

Meyer, R.W. (1957) 'Joachim Jungius und die Philosophie seiner Zeit' (Joachim Jungius and the philosophy of his age), in *Die Entfaltung der Wissenschaft. Zum Gedenken an Joachim Jungius*, Hamburg: Augustin. (Examines the position of Jungius within the seventeenth-century philosophical tradition.)

Trevisani, F. (1978) 'Geometria e logica nel metodo di Joachim Jungius' (Geometry and logic as contained in the method of Joachim Jungius), *Rivista critica di storia della filosofia* 33: 171-208.(Interesting work on the mathematical basis of Jungius' scientific method.)

Jurisprudence, historical

Historical jurisprudence is the title usually given to a group of theories, which flourished mainly in the nineteenth century, that explain law as the product of predetermined patterns of change based on social and economic change. It is thus opposed both to theories that see law as essentially an expression of the will of those holding political power (positivist theories) and to those that see it as an expression of principles that are part of man's nature and so applicable in any kind of society (natural law theories). The writers of the Scottish Enlightenment first connected the historical development of law with economic changes. In the nineteenth century, Savigny and Maine postulated grand evolutionary schemes, which purported to be applicable universally. They were, however, based on the development of ancient Roman law and could only with difficulty be applied to other systems. These schemes are now discredited, but in the twentieth century more modest studies have successfully related particular kinds of law to particular sets of social circumstances.

1 1 The historical school of law

Montesquieu in his influential *The Spirit of the Laws* (1748) first showed the connection between law and the circumstances of society. He accepted the natural lawyers' view that law must reflect the nature of things but insisted that the nature of things is not the same in every society. Rather law reflects the spirit of each society, which is an amalgam of various components, some physical, such as its climate and the qualities of its land, and some moral, such as religion and social customs. Montesquieu recognized that some societies were more developed than others and that therefore their laws had to be more sophisticated, but he imposed no scheme of historical progression on his material. Adam Smith in his *Lectures on Jurisprudence* (1762-3) argued that a society's law changes according to the progress of society through four stages based on the prevailing mode of subsistence, namely, hunting, pastoral, agricultural and commercial. The first three were reflected in developments in the law of property and the last in the law of contract. The writers of the Scottish Enlightenment accepted as axiomatic that as a society's mode of subsistence varies, so its laws must be different, and thus established the connection between legal and economic development.

In the nineteenth century these ideas formed the basis for theories of legal evolution, which held that law develops on predetermined lines from its own internal forces. The implication was that changes should be left to these natural social mechanisms and therefore that reform legislation and codification were contrary to the nature of law. Its theorists generalized from the history of ancient Roman law, which had a continuous development of over a thousand years, and appeared to be relatively unaffected by legislation. From its infancy as the law of a tribal society in the sixth century BC, it grew into the law of a world empire, reached its technical peak in the second and early third centuries AD and thereafter declined until its codification by the Emperor Justinian in the sixth century AD.

The main German protagonist of historical jurisprudence was F.K. von Savigny, whose view was published in (1814 as a polemic against the proposal to codify German law on the model of the Code Napoléon. Eighteenth-century proponents of codes had combined the positivist and the naturalist traditions by arguing that what would make the codes law was legislation but their substance would be dictated by what was congruous with man's nature. Savigny insisted that a people's law must fit its particular circumstances. He took over the idea that the history of Roman law was divided into three periods, dressed it in the metaphor of organic growth, which Edmund Burke had used to describe the unwritten constitution of Britain, and declared that law among the 'nobler' nations (by which Savigny meant essentially the Roman and the German) goes through three periods: growth, maturity and decline. In the first period it is manifested directly in the practices of the people. As society grows, some parts of law become more technical and are adapted by expert jurists, acting as delegates of the people; but whether it is based on custom or on juristic discussion, it is developed not by legislation but through silent forces within itself. In its infancy it is not technical enough for legislation and in its decline it is not worthy to be codified. The only suitable period for codification would be the period of maturity, but then, with the maximum participation of the people and juristic science at its zenith, legislation would be an unnecessary intrusion.

Savigny's followers in the German Historical School accepted the organic and inevitable connection between the

special character of a people and its language and its law. For them the spirit of a particular people (*Volksgeist*) was something more than Montesquieu's spirit. The latter comprised a number of quantifiable components. The *Volksgeist* had a mystical dimension, much in tune with the romanticism of the first half of the nineteenth century. The enthusiasm generated by the historical school produced a series of distinguished studies of legal history, both in the Roman law and in the Germanic law area.

2 2 Maine's ancient law

In France Charles Comte 1827) produced *Traité de legislation ou exposition des lois générales suivant lesquelles les peuples prospèrent, dépérissent ou restent stationnaires* (Treatise on Legislation or Exposition of the General Laws According to which Peoples Prosper, Decline or Remain Stationary). He put special emphasis on geographical and racial factors as accounting for the progress of some societies and lack of progress of others and their effect on such societies' law. Relying on the lectures of the English physiologist William Lawrence, he asserted the superiority of people of Caucasian race.

Comte's distinction between progressive and stationary societies was combined with Savigny's scheme of legal evolution by the principal English exponent of historical jurisprudence, Sir Henry Maine, in *Ancient Law* (1861). Maine was concerned to provide a theory which would explain the law of 'progressive' societies (principally ancient Rome and England) and act as a counterweight to the positivism of Bentham and Austin, which held that all true law was the product of legislation by a political sovereign. He was equally concerned to refute those who based law on a priori abstractions such as the law of nature, and sought to offer an explanation of law which was both empirical and universally true in the sense that natural science was true.

> If it be truth at all, it must be scientific truth. There can be no essential difference between the truths of the Astronomer, of the Physiologist and of the Historian. The great principle that underlies all our knowledge of the physical world, that Nature is ever consistent with herself, must also be true of human nature… if indeed history be true, it must teach that which every other science teaches - continuous sequence, inflexible order, and eternal law.

(Maine (1876: 265-6)

Maine's starting point was sound. By viewing all law as legislation, Bentham and his followers obscured the nature of legal change. As Maine insisted, the farther back in time we go, the farther we are removed from the Benthamite notion that law is made expressly by legislation. Rather it evolves imperceptibly. The science which Maine used as his model was geology: the rudimentary ideas of law found in the literature of early societies 'are to the jurist what the primary crusts of the earth are to the geologist' Maine 1861: 3).

Ancient Law, however, although purporting to be concerned with 'the early history of society', was not concerned with primitive societies, such as Native Americans. It was essentially limited to Indo-European societies and was structured round the development of the institutions of Roman law. Thus the account of the law of the earliest period of society starts with divinely inspired kings handing down isolated judgments, or 'themistes'. Later they lose their sacred power and are replaced by a small group of aristocrats who have a monopoly of knowledge of the traditional customs. They abuse their power and popular agitation demands the recording of the customs in 'Ancient Codes'. This is a generalization of what happened in Rome when the monarchy gave way to the Republic and the plebeians demanded the enactment of the Twelve Tables. The scheme is not readily discernible elsewhere and in particular it has little application to England.

On the other hand Maine introduced certain ideas which have entered the common currency of jurisprudential discourse. Examples are the recognition of fictions and equity as early 'agencies of legal change', which precede legislation; the predominance of procedural rules over substantive rules in early law; and especially his famous generalization: 'the movement of progressive societies has hitherto been a movement from Status to Contract' (Maine 1861: 170).

Maine's work fitted the mood of the age of biological evolution, produced by Darwin's *Origin of Species* (1859), but its ideas were challenged both by anthropologists, who showed that they did not apply to early non-Indo-European societies, and by legal historians, who objected to the inaccuracies of his generalizations. Politically Maine's ideas led to a legal conservatism that saw the movement from status to contract as leading to

ever greater self-determination for the individual, so that the imposition of any form of strict liability on someone without fault was condemned as contrary to the principles of legal evolution.

Loss of faith in the applicability of the methods of the natural sciences to social phenomena and in the idea of the progress of society reduced interest in Maine's theories. However, the notion that a society's law must be viewed as closely related to its social and economic circumstances and should be studied not in isolation but in comparison with the laws of other societies with similar or different circumstances became permanently rooted in legal thought and is now commonplace.

3 3 Maine's heirs

Two followers of Maine, James Bryce and Sir Paul Vinogradoff, attempted modifications of his ideas. Bryce (1901) developed the comparative method. For him the law of any society was always a compromise between tradition and convenience. The study of similarities and differences in different legal systems can lead to the identification of frequently recurring phenomena, which can be treated as universals and thus in turn elucidate the nature of law in general. Vinogradoff (1920-22) more ambitiously extrapolated from the history of European societies six stages of legal progress, from totemistic law through medieval law, marked by the tension between feudal and canon law, to individualistic jurisprudence and finally socialist jurisprudence.

The immediate heirs of historical jurisprudence can be found in various disciplines. In the area of social theory, its influence can be seen in the work of two significant writers. Ferdinand Tönnies (1887), starting from the movement from status to contract, developed the idea of two contrasting forms of social grouping, the *Gemeinschaft*, or community of those sharing to some degree a common way of life, and the *Gesellschaft*, a more limited relationship, based on the parties' will to achieve a particular purpose, usually by contract. In the same tradition Max Weber (1922) introduced a typology of legal authority, contrasting rational authority with traditional and charismatic types of authority. He popularized the notion of 'ideal' model types, in the sense of social constructs, not actually found in reality, but useful as hypotheses to enable the scholar to highlight similarities and differences in actual systems. These types provide a way of coping with the theorists' perennial dilemma of achieving their need to generalize while at the same time maintaining fidelity to accuracy (see Social theory and law §§1-3).

4 4 Historical jurisprudence today

Although much of what would once have been called historical jurisprudence is now carried on under such labels as legal anthropology or comparative legal history, there are signs of a revival of the traditional search for general patterns of legal change, but with more modest objectives than those of the nineteenth-century theorists. Progress is no longer seen in terms of continuous improvement, but recurring patterns of legal change, corresponding to social and economic changes, can be identified. By shifting the notion of causal connection from one of necessity to one of probability, one can postulate likely connections between forms of society and the kinds of law they have and this process can result in some, albeit provisional, general propositions. Hitherto theories were concerned almost exclusively with the institutions of private law, but it is possible fruitfully to use model types to relate the growth of a distinct public law to a society with particular characteristics (such as France) and so criticize its introduction in a society (such as England) which lacks those features (Allison (1996).

Roberto Unger 1976) has emphasized the need to identify the different conceptions of the law and of what may be expected from it that characterize different types of society. This has enabled him to distinguish various types of society on the basis of such interpretive explanations. Thus in 'savage' societies customary patterns of interaction among individuals and groups are based on a recognition that such patterns produce reciprocal expectations of conduct. In 'bureaucratic' societies, the state has become the controlling institution, in which the dominant hierarchy begins to take a critical attitude to social customs and seeks to control them through the publication of express rules, seen as emanations of the human will. The idea that such rules should apply equally to all members of the society only emerges, however, in societies of the 'legal order', as exemplified by the liberal societies of post-feudal Europe. Such societies are composed of different groups with different interests, no one of which is dominant, so that the law is seen as a cohesive force, binding all groups equally.

Another manifestation of what may be called historical jurisprudence concentrates on the identification of

recurring patterns in the technical development of law, irrespective of their social context. Such studies emphasize similarities between the legal techniques of the laws of different societies, in such matters as the relations between procedure and substantive law, for example, the effect of use of laymen, such as civil juries, on the nature of the law, or the relations between written law, in which the text is fixed but subject to recurring techniques of interpretation, and unwritten law, in which the rules are not finally formulated but elucidated on a case by case basis by authoritative judges or jurists.

See also: Evolution and ethics; Evolutionary theory and social science; Law, philosophy of

PETER STEIN

References and further reading

Allison, J.W.F. (1996) *A Continental Distinction in the Common Law: A Historical and Comparative Perspective on English Public Law*, Oxford: Oxford University Press.

Bryce, J., Viscount (1901) *Studies in History and Jurisprudence*, 2 vols, Oxford: Clarendon Press.

Comte, C. (1827) *Traité de legislation ou exposition des lois générales suivant lesquelles les peuples prospèrent, dépérissent ou restent stationnaires (Treatise on Legislation or Exposition of the General Laws According to which Peoples Prosper, Decline or Remain Stationary)*, 4 vols, Paris.

Diamond, A. (ed.) (1991) *The Victorian Achievement of Sir Henry Maine: a Centennial Re-appraisal*, Cambridge: Cambridge University Press.(Various studies of Maine's influence, of which those of J.W. Burrow, E. Shils and C. Woodard are specially relevant.)

MacCormack, G. (1985) 'Historical Jurisprudence', *Legal Studies* 5: 251-60.(Current state of the movement.)

Maine, H.S. (1861) *Ancient Law*, London: Murray.

(1876) *Village Communities in the East and West*, 3rd edn, London: Murray.

Montesquieu, C. Baron de (1748) *De l'esprit des lois* (*The Spirit of the Laws*), trans. T. Nugent, 2 vols, London, 1750.

Savigny, F.K. von (1814) *Zum Beruf unsrer Zeit für Gesetzgebung und Rechtswissenschaft*, trans. A. Hayward, *On the Vocation of Our Age for Legislation and Jurisprudence*, London: Littlewood, 1831.

Smith, A. (1762-3) *Lectures on Jurisprudence*, R.L. Meek, D.D. Raphael and P.G. Stein (eds), Oxford: Oxford University Press, 1978.

Stein, P. (1980) *Legal Evolution: The Story of an Idea*, Cambridge: Cambridge University Press.(General survey of theories of evolution of law through successive stages.)

(1986) 'The Tasks of Historical Jurisprudence', in N. MacCormick and P. Birks (eds), *The Legal Mind: Essays for Tony Honore*, Oxford: Clarendon Press.(A possible future for historical jurisprudence.)

Tönnies, F. (1887) *Gemeinschaft und Gesellschaft*, trans. C.P. Loomis, *Community and Society*, London: Routledge, 1955.

Unger, R. (1976) *Law in Modern Society*, New York: Free Press, 1976.

Vinogradoff, Sir P. (1920-22) *Outlines of Historical Jurisprudence*, 2 vols, Oxford: Oxford University Press.

Weber, M. (1922) *Wirtschaft und Gesellschaft* (*Max Weber on Law and Society*), M. Rheinstein (ed.), Cambridge, MA: Harvard University Press, 1954.

Justice

*The idea of justice lies at the heart of moral and political philosophy. It is a necessary virtue of individuals in their interactions with others, and the principal virtue of social institutions, although not the only one. Just as an individual can display qualities such as integrity, charity and loyalty, so a society can also be more or less economically prosperous, artistically cultivated, and so on. Traditionally defined by the Latin tag '*suum cuique tribuere*' - to allocate to each his own - justice has always been closely connected to the ideas of desert and equality. Rewards and punishments are justly distributed if they go to those who deserve them. But in the absence of different desert claims, justice demands equal treatment.*

A common division of the topic distinguishes between corrective and distributive justice. Corrective justice covers that which is due to a person as punishment, distributive that which is due by way of benefits and burdens other than punishments. Within the sphere of corrective justice there is disagreement about the justification of punishment itself. But there has been - and is - widespread agreement on the criteria for just punishment: just punishments must be properly imposed and the quantum of punishment must reflect the seriousness of the offence.

There has been no such agreement about the content of just principles for the distribution of benefits and (non-punitive) burdens. Conventionalists claim that what is due to each person is given by the laws, customs and shared understandings of the community of which the person is a member. Teleologists believe that an account can be given of the good for human beings and that justice is the ordering principle through which a society (or humanity) pursues that good. Justice as mutual advantage proposes that the rules of justice can be derived from the rational agreement of each agent to cooperate with others to further their own self-interest. Theorists of what may be called justice as fairness believe that justice is a thin concept which provides a fair framework within which each person is enabled to pursue their own good.

1 Corrective justice

There is general agreement that a just punishment should meet the following criteria. First, it should be imposed only on a properly convicted wrongdoer. Second, the quantum of suffering should satisfy the principle of ordinal proportionality. This means, as Hirsch (1990) has put it, that 'persons convicted of crimes of comparable gravity should receive punishments of comparable severity' except where mitigating or aggravating circumstances alter the culpability of the offender. Third, the quantum of suffering should satisfy the principle of cardinal proportionality: there should be a vertical ranking of crimes and penalties by seriousness.

There is, however, disagreement over the justification for punishment, and this makes it controversial how 'seriousness', 'severity' and 'culpability' are assessed and how the scale of penalties should be fixed. Those who appeal to deterrence may regard a widespread and socially disruptive crime as serious (whatever its degree of moral wrongness), and favour a scale of penalties designed to deter criminality. Those who favour retribution, however, have traditionally regarded seriousness as a factor of moral culpability and the scale of penalties as being derived from some notion of desert (see Crime and punishment).

2 Conventionalism

Turning to social, or distributive, justice, the attraction of some form of conventionalist approach is clear. Since there are institutions, conventions and systems of law that determine what is due to whom, resolving issues of justice may be thought merely to require reading off the correct answer from such sources. The earliest extant statement of a conventionalist view of justice is offered by Socrates' interlocutors, Cephalus and Polemarchus, in Book I of Plato's *Republic*. Polemarchus states that justice is giving a man his due, or what is appropriate to him, and it is clear that for Polemarchus what is appropriate to each person is dictated by the conventions prevalent in contemporary Athenian society.

A more complicated statement of conventionalism has been offered by Michael Walzer (1983). He argues that every social good (for example, health care, wealth, income and political rights) has an appropriate criterion of distribution which is internally related to how that good is understood by society. For example, in the UK (as elsewhere), health care is understood essentially to concern itself with illness and the restoration of health. This shared understanding of health care seems to entail a distributive criterion: medical need. Anyone, therefore, who

claims that health care in the UK (and many other societies) ought to be distributed in accordance with, say, status has either failed to grasp the nature of the good of health care or falls outside the community which is united and defined by its shared understandings. The only universal principle of distributive justice is the demand that respect be given to different shared understandings: no community ought to impose its own understanding of a given good, and its criterion for the distribution of that good, on any other community with different views.

The dependence of this theory on shared understandings has led to its being criticized on both empirical and theoretical grounds. Empirically, it is doubtful that any society is so homogenous as to boast a single, coherent and uncontested understanding of each of its social goods. (Is the freezing of embryos so that women might have children later in life, after they have established themselves in a career, meeting a medical need?) Theoretically, the account is flawed because the proposed universal principle of justice is not found in any society. A belief may be ‘ours’ in that it defines justice *according* to us; but it does not follow that we believe it is ‘ours’ in the sense of *applying* only to us. If justice is internally related to social understandings, there can be no perspective from which anyone might, from outside a given set of understandings, condemn the understandings or practices of a society as unjust. Yet there clearly is such a perspective. For example, in a hierarchical society it may be that the local conception of justice would require gross inequalities based on ascription at birth; this could be regarded as unjust.

The *prima facie* attraction of conventionalism is immediately undone when one asks why any given set of conventions, laws, or shared understandings should determine the distribution of benefits and burdens. Once this question is asked there is no reason why the answer has to flow from within the narrow resources of the community. It is at this point, when one asks why things are as they are, that the philosophical problem of justice really begins.

3 Teleology

In the history of thought about justice, the most common justification of any given set of laws, conventions or practices has been to appeal to an external, usually divine, authority. In the natural law tradition, to make an act just it is not enough that it complies with the society’s positive law. The positive law must itself be in accordance with a natural law which is knowable through the faculty of human reason (see Natural law). This tradition, owing its origins to the Greek Stoics, found its most lucid interpreter in Cicero (§§2, 4) and was given its definitive Christian form by Aquinas (§13). ‘True law’, writes Cicero, ‘is right reason in agreement with nature; it is of universal application, unchanging and everlasting’(*c*.54-51 bc: III, XXII, 211). The link to human nature via human reason is important, for it then follows that human beings reach their true end, or realize their true natures, only by living in accordance with natural law. What justice is and why it is a good are, thus, answered at the same time.

A major problem for this account is its reliance on an external source. Cicero is typical in claiming in the same passage that it is God who ‘is the author of this law, its promulgator, and its enforcing judge’. Natural law theory faces the difficulty of having to give an account of the existence and verifiability of the ‘true and unchanging’ moral order. Commonly, the natural law was conveniently said to underwrite the existing positive law: this may reflect the role of the powerful both in formulating positive law and defining natural law.

Suppose we are attracted to the notion that human institutions are to be justified by their contribution to human good, but do not believe that human reason is capable of discerning a divine plan. Then we may naturally arrive at the secular alternative embodied in utilitarianism: the idea that the ground of justification is human wellbeing, happiness or ‘utility’. When the utility of different people conflicts, the criterion for bringing their interests into relationship with one another is that aggregate utility is to be maximized (see Utilitarianism).

Utilitarianism can be characterized as a theory of justice if we simply define justice as the principal virtue of institutions. Utilitarians claim that, in the last analysis, there is only one virtue that matters: maximizing utility. However, utilitarianism is liable to violate elementary demands of justice as they are commonly understood: two people with identical deserts will be treated very differently if the utilitarian calculus happens to produce the finding that utility would be maximized by so doing. The common belief is that what is due to a person ought to be related to something about that person, not derived from a calculation about what would most effectively further some overall desirable state of affairs.

An important line of thought within utilitarianism has attempted to head off such criticisms by arguing that in practice the dictates of utilitarianism would underwrite the practices and institutions that are usually thought of as being required by justice. Thus, David Hume (§4), in *A Treatise of Human Nature* (1739-40), described justice as an 'artificial virtue' in that individual acts of justice contributed to utility not directly (as an act of benevolence would) but indirectly *qua* adherence to an institution that was on the whole beneficial. Hume's examples were respect for property, chastity (in women), allegiance to the government and promise-keeping. For Hume, then, justice was a convention - but it made sense to ask what good was served by following it. On somewhat parallel lines, J.S. Mill (§11) argued in 'Utilitarianism' (1861) that 'justice' is the name we give to those precepts whose strict observance is important for the furtherance of the utilitarian end. Thus, utilitarians argue that the arbitrary departures from social rules summoned up by anti-utilitarians as an implication of the doctrine are, in the long term, not really for the general good. Opponents of utilitarianism however have claimed that situations might still arise in which injustice (as normally understood) would be for the general good. Further, they complain that the foundations of the theory remain unsatisfactory because it does not (in John Rawls' phrase) 'take seriously the separateness of persons'.

4 Justice as mutual advantage

Utilitarianism is based on the assumption that the good of different individuals can be in some sense lumped together, and the pursuit of aggregate utility proposed as the objective of everyone. But, in the absence of an external lawgiver, how can an (not naturally benevolent) individual be encouraged to adopt the maximization of total utility as a binding demand? If we doubt that any satisfactory answer can be given, we may be tempted by a theory of justice that takes as its starting point the assumption that each person has a conception of their own good, and that justice must be shown to contribute to the attainment of that good. Justice is thus the terms of a *modus vivendi*: it gives everyone the best chance of achieving their good that they can reasonably expect, given that others are simultaneously trying to achieve their (different) good. Versions of 'justice as mutual advantage' can be found in Thrasymachus' 'might is right' argument in Book I of Plato's *Republic* and in the fraudulent social contract identified by Rousseau (1755: Part II) as having been perpetrated by the rich on the poor. But the *locus classicus* of this theory is undoubtedly *Leviathan* by Thomas Hobbes (§7).

If the terms of agreement are to be to the advantage of each (compared with unrestrained conflict), they must reflect the relative bargaining strengths of the cooperators. The strong and talented have little to gain (or to fear) from the weak or infirm, and the latter may even 'fall beyond the pale' of morality if the strong have no reason for taking their interests into account (Gauthier 1986: 268). Intuitively, it may seem perverse to call this a theory of justice. It is true that justice as mutual advantage has much of the structure of a theory of justice in that it results in rules that constrain the pursuit of self-interest. But the content of those rules will correspond to those of ordinary ideas of justice only if a rough equality of power holds between all the parties.

Even if this objection is not regarded as decisive, the theory suffers from internal problems. These concern the determinacy of the rules and their stability. The determinacy problem arises because of the necessity for all to start from a common view of the relative bargaining strengths of the participants. This is an immensely demanding condition, given the information about resources required and the different predictions that are liable to be made about the outcome of conflict. Even after rules have been agreed, some parties will have reason to press for changes if their bargaining power increases. Justice as mutual advantage results in rules which are no more than truces and, like truces, they are unlikely to be stable if there are changes in the balance of power between the sides (Barry 1995: 41).

Stability will also be challenged by the problem of non-compliance. Justice as mutual advantage appeals to the self-interest of each and does so by establishing rules that, if generally complied with, will further the interests of all individuals. However, this gives the agent no reason to comply with a given rule when there is greater advantage to be had by breaking it. This applies especially when the agent can free ride on the compliance of others. All that can be attempted is to increase the costs of non-compliance by increasing the sanctions if the agent is detected. To run a society using only self-interest and sanctions, however, would mean using a degree of coercion and of 'policing' hitherto unthought of in even the most totalitarian society.

5 Justice as fairness

Theories of justice as fairness start, like those of justice as mutual advantage, from the premise that the role of justice is to provide a framework within which people with competing ideas of the good can live together without conflict. However, justice as fairness seeks a framework that can be said to be fair to all the parties. It aspires to allocate to each person a fair set of opportunities to pursue their idea of the good life. The problem is that if what is due to each person is not determined by convention, by some overarching theory of the good, or by mutual advantage, how is it to be determined? On this question there is widespread disagreement. What is agreed is that there is an initial equal claim to consideration (see Kymlicka 1990). The root of disagreement lies in determining the scope, grounds and nature of this equality.

The range of accounts of the content of justice which is compatible with justice as fairness is very great: Robert Nozick's entitlement theory and John Rawls' theory of justice, for example, come to very different conclusions about what justice demands, even though both start with the basic idea that justice is to regulate the interactions of free and equal persons. Nozick believes that 'individuals have rights, and there are things no person or group may do to them (without violating their rights)' (1974: ix). From this starting point he generates what he calls an entitlement theory of justice (see Nozick, R. §2). The just pattern of distribution is that which would result from voluntary transfers, given that holdings were justly acquired in the first place (by just transfer or by an appropriation that makes no one else worse off). Nozick, then, regards the claim to an equal set of absolute rights as defining the limits of justice: any actions which interfere with those rights (such as redistribution) are unjust no matter what the pattern or outcome of the entitlement theory. Nozick does not, however, offer any account of the existence of such robust rights, and his arguments from intuition to show that any interference with individual rights is unjust are unconvincing. It is plausible that injustice may result from a large number of individual transactions each of which taken separately seems just.

John Rawls (1971) has argued that justice requires the provision of equal basic liberties and fair opportunities for all, and that social and economic inequality can only be justified where it is to the benefit of the least advantaged. These principles are derived by arguing that they would be chosen by free persons in an 'original position', the specifications of which prevent people from making unfair use of their natural and social advantages (see Contractarianism §7; Rawls, J. §1). Rawls then, in contrast to Nozick, believes that justice requires us to do much more for each agent than merely provide them with absolute property rights. Rather, the pattern of distribution is set at that which will maximally benefit the worst off. The question is one of determining the content of justice: when is it appropriate to move away from the concept of equality? 'when have we done enough' for a person to be able to say that variations in outcome are 'deserved'? (Scanlon 1988: 187) (see Equality; Desert and merit)

Recent work in Anglo-American political theory has been dominated by the development of theories between these positions. Ronald Dworkin (1981), for example, has argued that Rawls goes too far in treating all characteristics that make people more or less productive as 'morally arbitrary', and thus fails to allow for the justice of rewarding enterprise and ambition. Dworkin, in his own work, seeks to accommodate differences in ambition while retaining the feature of endowment insensitivity (see Dworkin, R.).

A different line of argument is that Rawls' original position is not a bargaining environment because the veil of ignorance entails that the participants are identical. Moreover, Rawls secures the two principles only by building in a number of ad hoc requirements (for example, that the participants are risk averse). The alternative proposed is to posit participants who are aware of their identities and *morally* motivated to seek agreement. Such a position has been proposed by the philosopher Thomas Scanlon (1982) and developed into a theory of justice by Brian Barry (1995) (see Contractarianism §9).

All theories of justice as fairness face the problem of grounding the commitment to the fundamental equality of persons and giving an account of each agent's motivation to behave justly. Rawls offers two justifications for his principles. The first is that the principles match our 'considered moral judgements' about justice: we come to a 'reflective equilibrium' in which the principles reflect and organize our moral intuitions. The second justification is offered as a Kantian interpretation. On this account the original position provides a 'procedural interpretation' of Kant's realm of the 'kingdom of ends' (see Kant, I. §9). The original position and the choice of the principles are viewed as an attempt to replicate Kant's reduction of morality to autonomy and autonomy to rationality. Thus, by living in accordance with justice we realize our true natures as autonomous beings. This provides the motivation required.

In later papers and in his second book, *Political Liberalism* (1993), Rawls has moved away from the Kantian interpretation on the grounds that it requires a controversial metaphysics and that it commits him to a particular view of the good life as autonomy. Instead Rawls now emphasizes that his theory relies on nothing more than ideas 'latent in the public political culture' of modern Western democratic states. This move aligns him with the conventionalist position discussed in §2, and is open to the criticisms raised there. Without the Kantian interpretation it is also difficult for Rawls to give an account of moral motivation. What is left is a theory that claims to give a content to the idea of justice but which cannot, when the demands of justice conflict with the self-interest of the agent, provide the agent with a reason to be moral (Nagel 1991).

Scanlon's and Barry's accounts assume morally motivated individuals who are concerned to find terms of agreement which reflect the equality of persons. Barry has argued that the demands of fairness can be generated from the lack of an authoritative account of the good; but this argument still relies on a prior commitment to freedom and equality. There is, however, no reason why such theories should not claim that such a commitment has universalist implications: an appeal to what we believe does not mean that what we believe applies only to us. Similarly, theorists of justice as fairness need not make the illegitimate claim (much more often attributed to them than made by them) that justice as fairness is neutral between all ways of life (no matter how illiberal) (see Neutrality, political).

Justice as fairness can claim to give a content to the idea of justice by telling us what is justly due to whom, and can do so in a way that matches our fundamental intuitions about the nature not only of justice but of morality generally - namely, that each agent is a locus of equal value. That they cannot provide the agent with a decisive reason to behave morally when to do so conflicts with the agent's self-interest or conception of the good may not reflect a defect of justice as fairness. Rather, it may be an inevitable aspect of moral theorizing without recourse to a divine order or a single, comprehensive, idea of the good (see Moral motivation).

6 Critics of justice

The discussion so far has proceeded on the assumption that justice is the principal virtue of institutions. The theories of justice examined here would explain this primacy in different ways; by appealing to the most important shared understandings, the most stringent demands of Nature or God, the conduciveness of justice to utility or civil peace, or the role of justice in providing a fair framework for the pursuit of different conceptions of the good. But all agree that, where justice conflicts with other values, those other values must give way.

This consensus has been challenged on the grounds that under ideal conditions justice would be unnecessary, and appeals to it would actually destroy valuable social relationships. Thus, a marriage in which the spouses were constantly arguing in terms of rights and duties would be less good than one in which mutual love created spontaneous harmony. By an extension of this sentimental line of thought, an ideal community would be one in which justice had been transcended by a spirit of what used to be called (until feminist scholars objected) fraternity. This is one strand in the thought of Marx, and it recurs in the work of some contemporary feminist and communitarian writers (see Community and communitarianism; Feminist political philosophy §4).

The theorists of justice discussed above would not necessarily dispute such claims. Both Hume and Rawls argued that there are 'circumstances of justice' that make justice necessary. These are precisely the conditions - conflicting demands for material goods and unreconcilable aspirations - that the critics of justice believe would be transcended by a sufficiently strong community spirit. The disagreement is not analytical but turns on the view taken of the possibility and the desirability of creating a community in which justice ceased to be the first virtue. The partisans of justice can point out that the theoretical assault on 'bourgeois morality' has provided the supposed justification for the most appalling violations of rights (for example, in China, Cambodia and the former USSR), and ask if there is any reason to suppose that other social experiments driven by the same animus would be any more benign.

BRIAN BARRY
MATT MATRAVERS

References and further reading

Aristotle (*c*.mid 4th century BC) *Nichomachean Ethics*, trans. D. Ross, revised J.L. Ackrill and J.O. Urmson, Oxford: Oxford University Press, 1980.(Famous discussion of ethics. Book V contains an extended analysis of

justice including its division into distributive and rectificatory.)
Barry, B. (1995) *Justice as Impartiality*, vol. 2, *A Treatise on Social Justice*, Oxford: Clarendon Press.(Author's work. Develops the approach introduced by Scanlon and also contains useful discussions of justice as mutual advantage and of Rawls' theory. Referred to in § §4 and 5.)
Cicero, M.T. (*c*.54-51 BC) *De Re Publica*, trans. C.W. Keyes, Cambridge, MA: Harvard University Press and London: Heinemann, 1928.(Parallel Latin text and English translation. Important statement of neo-Stoic and natural law theory. Referred to in §3.)
D'Entrèves, A.P. (ed.) (1948) *Aquinas: Selected Political Writings*, Oxford: Blackwell.(Standard English edition of Aquinas' political theory, relevant to the natural law position discussed in §3.)
Dworkin, R. (1981) 'What is Equality? Part 1: Equality of Welfare; Part 2: Equality of Resources', *Philosophy and Public Affairs* 10 (3 and 4): 185-246, 283-345.(Quite difficult discussion of whether equal treatment requires equality of welfare or of resources. Also contains Dworkin's theory of justice referred to in §5.)
Gauthier, D. (1986) *Morals By Agreement*, Oxford: Clarendon Press.(Modern neo-Hobbesian account of justice as mutual advantage. Referred to in §4.)
Hirsch, A. von (1990) 'Proportionality in the Philosophy of Punishment: From "Why Punish?" to "How Much?"', *Criminal Law Forum* 1 (Winter): 259-90.(Referred to in §1.)
Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.(Referred to in §4.)
Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1978.(Referred to in §3. See especially Book 3 Part 2 for a discussion of justice.)
Hume, D. (1748 and 1751) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1975.(Published originally as two works, hence the dates.)
Kymlicka, W. (1990) *Contemporary Political Philosophy: An Introduction*, Oxford: Clarendon Press.(Very useful introductory work covering many of the positions dealt with here. Referred to in §5.)
Marx, K. and Engels, F. (1968) *Marx/Engels: Selected Works in One Volume*, London: Lawrence & Wishart. (Contains extracts of many of Marx's most famous tracts. For a denunciation of the whole idea of justice see the 'Critique of the Gotha Programme', 311-31.)
Mill, J.S. (1861) 'Utilitarianism', *Collected Works of John Stuart Mill*, vol. X, *Essays on Ethics, Religion and Society*, ed. J.M. Robson, Toronto: University of Toronto Press, 1969.(Discussion of utilitarianism. Chapter V contains the argument for a utilitarian theory of justice referred to in §3.)
Nagel, T. (1991) *Equality and Partiality*, New York: Oxford University Press.(Referred to in §5.)
Noddings, N. (1984) *Caring: A Feminine Approach to Ethics and Moral Education*, Berkeley, CA: University of California Press.(Leading feminist theorist who argues for the replacing of the ethic of justice with an ethic of care.)
Nozick, R. (1974) *Anarchy, State, And Utopia*, New York: Basic Books.(Discussed in §5.)
Plato (*c*.380-367 BC) *Republic*, trans. R. Waterfield, Oxford: Oxford University Press, 1993.(Classic discussion of justice. Book 1 is referred to in §2.)
Rawls, J.B. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(Discussed in §5.)
Rawls, J.B. (1993) *Political Liberalism*, New York: Columbia University Press.(Rawls' revised theory, referred to in §5.)
Rousseau, J.-J. (1755) *A Discourse on Inequality*, trans. M. Cranston, Harmondsworth: Penguin, 1984.(Referred to in §4.)
Sandel, M.J. (1982) *Liberalism and the Limits of Justice*, Cambridge: Cambridge University Press.(A communitarian critique of Rawls and of the centrality of justice.)
Scanlon, T.M. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press, 103-28.(Discussed in §5.)
Scanlon, T.M. (1988) 'The Significance of Choice', in S.M. McMurrin (ed.) *The Tanner Lectures on Human Values*, vol. 8, Salt Lake City, UT: University of Utah Press, 151-216.(Referred to in §5.)
Waldron, J. (ed.) (1987) *Nonsense on Stilts*, London: Methuen.(Reprints some attacks on rights and contains a valuable discussion of these critiques by Waldron.)
Walzer, M. (1983) *Spheres of Justice: A Defence of Pluralism and Equality*, New York: Basic Books; London and Oxford: Blackwell.(Discussed in §2.)

Justice, equity and law

Laws are intended to achieve justice, but the application of an otherwise just law may yield an injustice in the circumstances of a particular case. This is because laws are framed in terms of general rules which cannot adequately provide in advance for all possible variations in relevant circumstances. Equity modifies the rigid application of the law in such cases in order to secure justice in the light of all the relevant circumstances. So, in Aquinas' example, the law may justly require the closure of the city gates after a certain hour, but officials may equitably decree the opening of the gates during the legal hours of closure in order to save fighters defending the city who are being pursued by an enemy. In this sense, the equitable decision is not distinct from justice, but rather secures justice in the particular case by remedying the deficiencies of positive law retrospectively at the point of application. The question of the proper relationship between justice, equity and law has been explored both by a rich philosophical tradition that finds its classic statement in the writings of Aristotle, and by the world's major legal traditions.

At least two major problems arise for this complex of ideas. First, that of the 'decadence' of equity. As equity is incorporated into formal processes of legal adjudication, for instance in the form of 'maxims' of equity or equitable 'doctrines', it comes to acquire the generality of positive law. This creates the problem that the so-called equitable maxims or doctrines may themselves then be applied in a strict way that leads to injustice in the particular case, which is precisely the problem equity is meant to remedy. Second, in securing justice by the deployment of discretion in the particular case equity also threatens an injustice. For it seems to involve a departure from the principle of legality, that is, the duty to apply pre-existing laws declared beforehand to those subject to them. In being adversely affected by the retrospectively operative discretion of the adjudicator, the party who would have benefited from the strict application of the law may regard the resort to equity as itself unjust.

1 Positive law and justice

A fundamental objective of positive law is widely thought to be that of achieving justice. Even contemporary legal positivists hold that it is an essential feature of law that it claims to conform with precepts of justice (see Legal positivism §4). In its broadest sense, justice refers to those moral requirements that apply in any interpersonal context. It is that part of morality that governs the relations and interactions between morally significant beings. It is in this vein that John Rawls refers to an abstract *concept* of justice - understood as a 'set of principles for assigning rights and duties and for determining… a proper distribution of the benefits and burdens of social cooperation' (Rawls (1971: 5) - that is shared by most people who engage in discourse about justice, but regarding which there may be vastly divergent and competing specific interpretations or *conceptions*.

However justice is understood, it is a further widely held belief that positive laws may fail to achieve justice. Those who deny the moral fallibility of law are the proponents of a rarely held version of normative ethical relativism, according to which the laws of one's community are always just or according to which justice is to be identified with the standards enshrined in those laws. But this sort of relativism conflicts with the pervasive fact that people do submit laws to critical scrutiny in terms of independent standards of justice and often intelligibly purport to find those laws to be unjust for breaching those standards.

Given that positive laws are morally fallible, injustice can arise in various ways (Finnis 1980: 352-4). The lawmaker may have been motivated by improper considerations; the law may have been made outside the scope of the lawmaker's authority; the law may have been enacted in contravention of the formal requirements of the rule of law, for example, publicity, generality, prospectivity, and so on; and, finally, the law may be substantively unjust, because it posits a morally defective assignment of rights and duties or benefits and burdens.

2 Equity between law and justice

The study of the relationship between justice, equity and law belongs in a normative theory of legal adjudication (see Legal reasoning and interpretation). This is concerned with articulating the nature of ethically sound legal adjudication. One of its central aims is to determine whether and how a judge or other legal decision-maker may rely on the deliverances of their sense of justice in order to adjust, or depart from, the strict application of a law to avoid doing injustice in the circumstances of a particular case. It is in responding to this question that the nexus

between justice, equity and law has been invoked, both by an illustrious philosophical tradition and by the world's major legal systems.

Equity is a form of judgment that remedies the injustice of positive law at the point of application by attending to the particular features of the persons and circumstances involved. Thus, equity is not concerned with defects in intention, author and form or with the substantive injustice of the general rule itself. Even if a positive law avoids these kinds of injustice it may still yield a result that is substantively unjust when it is applied to a particular set of facts. The source of the injustice is the inherent generality of legal rules. We can trace the philosophical elaboration of the idea of equity back to ancient Greece. The canonical text is Aristotle's *Nicomachean Ethics*, especially Book V, as supplemented by his discussion in the *The Art of Rhetoric*, although Aristotle's account is prefigured in earlier writings both literary and philosophical, for example by Plato (see Aristotle; Plato).

For Aristotle justice, both as an individual virtue and as an ordering of social life, is only possible within the context of the institutional arrangements, especially the laws, of a political community. He attributes three general characteristics to laws (see Yack 1993: 180-4). First, laws are expressed in the form of general rules intended to apply to an indefinite multiplicity of future situations. This is what distinguishes them from decrees (*psephismata*) about specific individuals and circumstances. Second, they apply to all members of the political community, including its leaders. Third, laws originate in the practical wisdom (*phronēsis*) of individuals and communities. The latter feature does not mean that laws are inherently good or just, but rather that they represent a fallible attempt on the part of humans' practical reason to cope with the problems of establishing a morally sound ordering of social life.

Aristotle introduces his analysis of equity (*epieikeia*) by means of an apparent paradox: equity seems neither the same as justice nor different from it in kind. On the one hand, we regard equity as a praiseworthy quality in both individuals and judgments. But it seems odd to value the equitable if it does not coincide with the just. Hence the dilemma: either we must cease regarding the just or the equitable as good; or, if they are both good, we must regard them as identical. Aristotle's solution is to define equity as a kind of justice, but a kind that is superior to and often opposed to another kind, strict legal justice. So, equity is not generically different from justice. But it is ideal justice - taking into account all the relevant circumstances of a given case - as opposed to justice strictly according to law.

The need for equity arises because the law is framed in general rules that apply prospectively to a vast number of future situations picked out by their general features. Only the general features thus identified (for example, regarding the type of person to which the rule applies, the type of activity it regulates, and so on) are relevant to the application of the rule. These comprise the only material facts to be taken into account in applying the rule, with all other facts being ignored. But, when applied to a particular situation in accordance with its terms, such a general rule may work an injustice. This will be either because the general rule is silent with respect to a particular situation, so that a just resolution is not secured, or because its application leads to an unjust result. This deficiency is not the fault of the lawmaker, but resides in the inherent variability of the circumstances of human life, which are such that the demands of practical wisdom (*phronēsis*) cannot be adequately captured in any set of general rules capable of a mechanical and unerring application.

In such cases equity justifies a departure from the strict letter of the legal rule, and a resort to its animating purpose or intention, in order to secure justice. Equity thus mediates between positive law and justice by importing an element of flexibility and context-sensitivity into legal adjudication. Hence the essential nature of equity as 'a rectification of law in so far as law is defective on account of its generality' (*Nicomachean Ethics* 1137b26-7; or 'right going beyond the written law', *The Art of Rhetoric* 1374a27-8).

The benefits of equity go beyond the attainment of justice in the particular case. On the one hand, equitable discretion admits into the adjudicative process ameliorative values such as mercy that mitigate the harshness of legal rules (see Forgiveness and mercy §3). The emergence of equity is therefore historically bound up with a heightened sensitivity to the plight of individuals, as opposed to an unbending concern with the maintenance of an impersonal moral order (Nussbaum 1993: 87-92). On the other hand, where the unforeseen situations to which equity responds reflect important changes in society generally, equity can be seen as a mechanism for the progressive and piecemeal adaptation of law to evolution in social conditions and moral belief. This reforming process is augmented by the tendency of the grounds on which equitable judgments are made to be

'systematized' into rules of positive law (see Maine 1930: 15).

It is clear that Aristotle regards equity as constituting a general virtue of character that also operates outside the context of formal legal adjudication. The equitable person 'is not unduly insistent upon his rights, but accepts less than his share, although he has the law on his side' (*Nicomachean Ethics* 1137b34-1138a3; *The Art of Rhetoric* 128). He also differentiates crimes from errors and misfortunes, since the latter are not a product of human wickedness; forgives human failings; judges actions in terms of the purposes of their agents and the entire context in which they occur; and recollects the good rather than the bad treatment they have received and endures being wronged. But the connection between equity and mercy is not yet fully established in Aristotle's analysis, for the latter can involve a leniency in assigning and exacting punishment that tempers the demands of particularized justice itself. Aristotle's concern is that such leniency may express an unacceptable self-depreciation in the face of injustices committed by others. A fuller elaboration of the connection between equity and mercy is, instead, given by the Roman Stoics (see Stoicism), especially Seneca, and the medieval canonists.

3 Contemporary influence

Although Aristotle is the key figure in the equity tradition in philosophy, this general complex of ideas has a line of inheritance that includes the Greek and Roman Stoics, the Roman orators, St Thomas Aquinas and the medieval canonists and casuists. In contemporary jurisprudence, however, this tradition has been somewhat sidelined by a preoccupation with the logically prior question of whether in fact judges ever do, or should, possess discretion. There are, however, three contemporary movements which embody essential themes of the equity tradition. Despite their many differences, they are united by a rejection of the merely mechanical application of general legal rules as an adequate normative account of legal adjudication and by an emphasis on the need for sound legal reasoning to be attentive to the normative significance of the constituent features of the particular case.

First, there is the neo-Aristotelian approach to legal reasoning (for example, Kronman 1993: Part 1). This portrays legal reasoning as a matter of Aristotelian practical wisdom, a mode of judgment that implicates the character of the legal reasoner in response to the salient particularities of the situation they are adjudicating upon, and not as the expert manipulation of a body of legal norms in a merely 'technical' or 'scientific' fashion. More generally, the emergence of neo-Aristotelian theories of practical reasoning illustrates the same line of thought. These urge the inadequacy of a purely rule-bound conception of practical reasoning - commonly associated with the Kantian and utilitarian traditions (see Kantian ethics; Utilitarianism) - and defend instead a model of practical reasoning as involving a quasi-perceptual appreciation of the normative demands generated by the particular context of agency and deliberation. Proponents of neo-Aristotelian theories of practical reasoning include John McDowell 1979), Martha Nussbaum, David Wiggins (1975-6) and Alasdair MacIntyre. The latter explicitly contends that Aristotle's account of the need for equity in purely legal contexts holds of practical life and reasoning in general (MacIntyre 1988: 120).

Second, the equity tradition resonates with key themes in feminist legal theory. For much feminist theory posits a stark opposition between a 'masculine' approach to legal reasoning, one that prizes the rigorous application of general rules that inevitably abstract from the rich complexity of concrete situations, and a 'feminine' approach that is highly particularistic, arriving at a judgment of substantive justice that reflects the totality of the circumstances (see Gilligan 1982). Feminist theory tends to employ this opposition in an antinomian way, privileging the substantive justice attainable through the exercise of unconstrained discretion over the idea of rule-following. Here it departs from the equity tradition, which views equitable discretion as a supplement to rule-governed reasoning in law (see Feminist jurisprudence).

Third, the 'law and literature' movement advances the reading and appreciation of literary texts as a model for legal adjudication. In the hands of one of its major exponents - Martha Nussbaum - the connection with the equity tradition is made central. Nussbaum stresses the importance Aristotle accords to *sungnome*, or 'judging with', in his analysis of equity. This is the sympathetic understanding of 'human things' which is tied up with the disposition to appraise actions in relation to the motives and intentions that animate them as well as the broader personal and social context in which they occur. Nussbaum associates this account of equitable judgment with Aristotle's theory of tragedy, which involves the spectator attending to the particular features of a tragic hero's narrative and, by doing so, sympathetically identifying with their plight. In Nussbaum's own analysis, however, the paradigmatic model of equitable judgment is not that of the spectatorship of tragic plays, but rather the

sensitive reading of a novel, since the latter genre exceeds tragic drama in its formal commitment to 'following complex life histories, looking at the minute details of motive and intention and their social formation' (Nussbaum 1993: 105).

4 The 'systematization' and 'decadence' of equity

The equity tradition in philosophy has had an impact on the two major Western legal traditions: Roman law and English common law. Equity is recognized in Roman law in the distinction between *ius* (law) and *aequitas* (equity), in the Roman orators' maxim *summum ius summa iniuria* (extreme right is extreme injury), and, most importantly, in the role of the *praetor*, an official empowered to grant a remedy where no legal remedy previously existed by exercising an equitable discretion. But perhaps the most conspicuous attempt formally to incorporate the notion of the equitable mediation between legal rules and justice is provided by the English common law.

From the fifteenth century until the late nineteenth century the English legal system was unique in maintaining an institutional distinction between courts that administered the common law, and those that administered equity (since then the bifurcated system has been abolished, and all courts are empowered to apply both law and equity). The court dispensing equity - the Court of Chancery - dealt with petitions from those claiming that the common law system was not capable of achieving a just result in their particular case. Equity was viewed as supplementing the common law and remedying some of its imperfections by developing discretionary remedies to deal with situations where the rigorous application of common law rules would lead to an injustice. By the sixteenth century, the dominant justification for this sort of modification of the application of general legal rules in specific cases was sought in Aristotle's writings on equity.

As in the Roman case, a chief concern of English equity was to give effect to the substance of transactions, despite failures of legal form. This coheres with the Aristotelian notion that equity looks to the substance of the law and to the intentions of the parties. So, for example, the Court of Chancery could relieve a party from meeting a legal requirement that a contract be put in writing if it would be against conscience for the other party to rely on the lack of writing, for example, when the first party has already carried out his part of the bargain. Equity also recognizes that the common law remedy for breach of contract - monetary damages - does not yield a just result in certain cases, and allows for an order that the contract be performed specifically. Similarly, equity intervened in other situations where strict reliance on legal rules was productive of injustice: for example, where there is fraud or mistake in formal transactions, to protect people incapable of looking after their own interests, or to protect relationships of trust and confidence. In addition to these doctrines, a number of principles were formulated - 'maxims of equity' - to underwrite and justify the use of equitable discretion, for example, 'he who comes to equity must have clean hands', 'equity regards the substance and not the form', 'equity is equality'.

With the systematization of equity into doctrines - such as that of specific performance, trusts and so on - and maxims that came to have the force of general rules of law, the Court of Chancery in turn came to be almost as rule-bound in its activities as the common law courts. Equity thereby ceased to be - as on the Aristotelian model - a discretionary modification of the application of legal rules in response to the particular circumstances of the case, and became instead a separate body of law itself expressible in general rules.

But with this systematization there arises the danger of what Roscoe Pound called 'the decadence of equity' (Pound 1905: 25). As equity becomes a system of legal rules, those rules are susceptible to being applied in a strict way that leads to injustice in the particular case, which is precisely the problem equity was meant to remedy. Pound's prescription was not to resist systematization, but rather to adopt an attitude of vigilance (see Pound, R.).

5 The conflict with legality

The idea of equity as a discretionary rectification of positive law at the point of application in the interests of substantive justice has, however, been subjected to a number of criticisms (see Lucas 1966: 214-22). Most of these come down to the objection that equity conflicts with the principle of legality. This is a normative principle requiring adjudicators to resolve disputes by applying legal rules that have been declared beforehand, and not to alter the legal situation retrospectively by discretionary departures from established law. The principle of legality, therefore, provides a basis for claiming that the deployment of equitable discretion, although intended to secure justice in the particular case, in fact produces injustice to the party adversely affected by it. But whence derives the

normative force of the principle of legality itself?

Most importantly, it derives from the fact that it expresses a requirement of the rule of law as it bears on adjudicators (see Rule of law (rechtsstaat)). The rule of law requires that people - including the legal officials of a state - be ruled by the law and obey it. But this whole ideal will be undercut if adjudicative bodies fail to apply pre-existing law in accordance with its clear and settled meaning. The sort of retrospective modification of pre-existing legal rules that is authorized by equity threatens this principle. The rule of law itself is upheld because it secures a number of central values. First, it achieves certainty and predictability in legal administration, so that people will know to what laws (and, in particular, what legal penalties) they will be subject. This enables them to plan their affairs, which is itself a key manifestation of a form of 'negative' liberty. Second, it protects human dignity. General legal rules create rational expectations among those subject to them. For an adjudicator to frustrate those expectations retrospectively seems akin - as Bentham argued - to the arbitrary way in which a man makes law for his dog: not declaring beforehand what conduct is required but rather waiting until the animal has done something he thinks wrong, and beating it. That equity is in tension with the rule of law, that its essentially discretionary and retrospective character imports an unacceptable element of uncertainty and arbitrariness into the legal system, is a common complaint, one expressed famously in the English legal context by the jibe that equity varied with the length of the Lord Chancellor's foot.

It is the attraction of the principle of legality which explains in large part the tendency to crystallize the grounds of equitable intervention into general legal rules, which in turn leads to the problem of the 'decadence' of equity mentioned above. More generally, it seems that we are confronted here with a conflict between 'formal' legal justice (which may be expressed in the injunction to 'treat like cases alike', where the criterion of likeness is determined by features picked out in advance by general legal rules) and 'substantive' justice in the particular case, which is the concern of equity. The sense of a clash between basic values is heightened by the realization that the 'generality' of legal rules - which is the structural source of the injustice that equity is meant to correct - is itself a requirement of the rule of law as it bears on the activities of legislatures (disallowing Acts of Attainder, for example).

6 Responses to the conflict

The classical equity tradition does not provide us with a ready response to such criticisms, mainly because the rule of law (in its connection with certainty and autonomy, at least) is a distinctively modern value that did not have a prominent role in the classical tradition represented by Aristotle and Aquinas. However, a response to such criticism might be to argue that the notion of equity can be understood as grounded in existing law, and therefore not discretionary in any way that seriously impugns the rule of law. There are at least two versions of this reply.

First, it might be claimed that in dispensing equity, the adjudicator is not departing from existing law nor making new law, but rather complying with the intention (as opposed to the letter) of existing law, which is its true meaning. This in fact is Aristotle's construal: the equitable adjudicator decides the case as the legislator would have decided it had he been aware of those particular circumstances. This response suffers from two main defects. On the one hand, it may be unworkable, given that it may not be possible to find any, or the right kind of, underlying legislative intention in a given case. In particular, all the serious difficulties involved in constructing a counterfactual intention arise here. On the other hand, to seek to avoid this problem through a presumption that legislators always intend to secure justice is to fall into vacuity, for what will then be at work will be a discretionary judgment dressed up as the realization of the legislator's intention.

The second version holds that the equity is grounded in ethical principles which are part of the law. One might here deploy Dworkin's theory of adjudication, according to which judges must strive to identify the set of ethical principles that provide the best interpretation of the back letter legal rules (Dworkin 1977). These principles are for Dworkin part of the law, and yield a single right answer in each case, thus eliminating the need for extra-legal discretion (see Dworkin, R.). But again there are at least two problems. First, if they are part of the law, it is not at all clear that these principles really perform the classic function ascribed to equity. Indeed, the principles to which Dworkin refers are not usually thought of as equitable but rather as expressing the rationale behind a series of more particular rules, and hence legitimately controlling the application of those rules in particular cases. Second, in so far as principles are legal standards with a determinate content, they fall victim to the possibility of the 'decadence' of equity.

A more promising line of response is to reject the demand to make an all-or-nothing choice between legality and equity (formal justice and substantive justice, rule and discretion) as posing a false dilemma. What is needed in any ethically sound legal system is an appropriate mix of both of these values, not a decision single-mindedly to pursue one at the complete expense of the other. Indeed, the equity tradition conceives of equity as a crucial supplement and corrective to legality in certain circumstances, not something that licenses the substitution of an unfettered and arbitrary discretion for the application of legal rules. Further, combining both legality and equity, rule and discretion, in a legal system seems practically unavoidable. This is because the idea of pure legality, of a legal code so determinate in its meaning that it calls for no discretionary judgment on the part of the adjudicator, seems illusory. The leeway for discretion will arise from at least two sources, in addition to the source (generality) already mentioned as giving rise to equity. These are: (1) indeterminacy in the interpretation of rules; and (2) indeterminacy in the proper characterization of the fact situation to which it will apply (see Lucas 1966: 24-5).

What is more difficult to ascertain is how the institutional balance between legality and equity is to be rightly struck. There is no reason to believe that there is a single best solution to this problem. But the equity tradition as discussed here does offer guidance on the reasonable solution in concrete circumstances (see Tasioulas 1996: 462-6); the rule of law tradition and the equity tradition need to be read together and held in balance.

See also: Justice; Law, philosophy of

JOHN TASIOULAS

References and further reading

Aquinas, T. (1266-73) Law and Political Theory, in *Summa theologiae*, vol. 28, IaIIae 90-97, London: Eyre and Spottiswoode, 1966.

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. J.A.K. Thomson and revised with notes and appendices H. Tredennick, Harmondsworth: Penguin, 1976.(Canonical text, especially Book V, on the idea of equity.)

Aristotle (*c*. mid 4th century BC) *The Art of Rhetoric*, trans. with intro. and notes H. Lawson-Tancred, Harmondsworth: Penguin, 1991. (Supplement to Aristotle's discussion of equity in *Nicomachean Ethics*.)

Dworkin, R.M. (1977) *Taking Rights Seriously*, London: Duckworth.(Chapter 2 contains a classic exposition of the role of principles in legal reasoning.)

Finnis, J. (1980) *Natural Law and Natural Rights*, Oxford: Clarendon Press.(A major contemporary restatement of natural law theory.)

Gilligan, C. (1982) *In a Different Voice: Psychological Theory and Women's Development*, Cambridge, MA: Harvard University Press.(Extremely influential psychological analysis of gender differences in ethical reasoning.)

Kronman, A.T. (1993) *The Lost Lawyer: Failing Ideals of the Legal Profession*, Cambridge, MA: Harvard University Press.(Presents a neo-Aristotelian approach to legal reasoning.)

Lucas, J. (1966) *The Principles of Politics*, Oxford: Oxford University Press.(A very effective discussion of the inter relations of justice, equity and law.)

McDowell, J. (1979) 'Virtue and Reason', *The Monist* 62: 331-50.(Defends an Aristotelian approach to practical reasoning.)

MacIntyre, A. (1988) *Whose Justice? Which Rationality?*, Notre Dame, IN: University of Notre Dame Press. (Defends a broadly Aristotelian-Thomistic approach to practical reason and justice.)

Maine, H.S. (1930) *Ancient Law: Its Connection with the Early History of Society and its Relation to Modern Ideas*, ed. F. Pollock, London: John Murray.(Chapters 2 and 3 contain a classic account of the history and social function of equity.)

Newman, R.A. (1973) *Equity in the World's Legal Systems: A Comparative Study*, ed. R.A. Newman, Brussels: Établissements Émile Bruylant.(Contains an exhaustive historical and comparative survey of the role of equity in the world's major legal systems.)

Nussbaum, M. (1990) 'The Discernment of Perception: An Aristotelian Conception of Private and Public Rationality', in *Love's Knowledge: Essays on Philosophy and Literature*, Oxford: Oxford University Press. (Defends an Aristotelian conception of practical rationality as involving 'perception' of what is rationally required in particular situations.)

Nussbaum, M. (1993) 'Equity and Mercy', *Philosophy and Public Affairs* 22: 83-125.(A discussion of the classical roots of the equity tradition and of its contemporary relevance, from the point of view of both

neo-Aristotelianism and the law and literature movement.)
Pound, R. (1905) 'The Decadence of Equity', *Columbia Law Review* 5: 20-35. (Relates to the discussion of 'decadence' in §4.)
Pound, R. (1954) *An Introduction to the Philosophy of Law*, New Haven, CT: Yale University Press.(Chapter 3 is particularly effective in stressing how the appropriateness of equitable judgment varies according to the area of law in question.)
Rawls, J. (1971) *A Theory of Justice*, Oxford: Oxford University Press.(The classic contemporary discussion of justice.)
Stein, P. and Shand, J. (1974) *Legal Values in Western Society*, Edinburgh: Edinburgh University Press. (Chapters 3 and 4 present a general account of justice, equity and law with illuminating legal illustrations.)
Tasioulas, J. (1996) 'The Paradox of Equity', *Cambridge Law Journal* 55: 456-69.(See discussion §6.)
Wiggins, D. (1975-6) 'Deliberation and Practical Reason', *Proceedings of the Aristotelian Society* 76: 29-51. (Defends an Aristotelian account of practical reason.)
Yack, B. (1993) *The Problems of a Political Animal: Community, Justice, and Conflict in Aristotelian Political Thought*, Berkeley, CA: University of California Press.(Chapters 5-6 are a vigorous presentation of the Aristotelian understanding of justice, equity and law that distances it from the concerns of classical natural law theory.)

Justice, international

Although it has been denied (by, for example, F.A. Hayek 1976) that the concept of distributive justice has application within states, it is not controversial that there can be unjust laws and unjust behaviour by individuals and organizations. It has, however, been argued that it makes no sense to speak of justice and injustice beyond the boundaries of states, either because the lack of an international sovereign entails that the conditions for justice do not exist, or because the state constitutes the maximal moral community. Both arguments are flawed. Without them, we are naturally led to ask what are the implications of the widely-held idea of fundamental human equality, the belief that in some sense human beings are of equal value. This cannot be coherently deployed in a way that restricts its application to within-state relations. In either a utilitarian or Kantian form it generates extensive international obligations. An objection that is often made to this conclusion is that the obligations derived are so stringent that compliance cannot reasonably be asked under current political conditions. But this shows (if true) that current political conditions are incompatible with international justice.

1 The scope of international justice

In its most literal sense, ‘international justice’ refers to justice between nations - ‘nations’ in this context being equivalent to ‘states’, as in ‘international law’, ‘national sovereignty’, and so on. Following common usage in political philosophy, it will here be understood in a more extended sense. In this sense, ‘international justice’ refers to justice in any context other than that of the purely internal affairs of a state.

It is possible that an inquiry into international justice will yield the conclusion that justice can be predicated properly only of relations between states. But formulating the scope of the inquiry in this way leaves open a variety of other possibilities. Thus, for example, it may be concluded that states have obligations of justice to individuals outside their borders (for example, to give asylum), that individuals have obligations of justice to states other than their own (for example, not to engage in terrorism), and that international organizations, transnational corporations and collective actors other than states have obligations of justice that should constrain their behaviour. The conclusion may also be that a world organized into states is inherently incapable of satisfying the demands of justice, and that only a single worldwide state holds out any possibility of realizing global justice.

2 Scepticism about international justice

International justice has in common with some other topics (for example, theology) the peculiarity that there is some controversy about the existence of its object. One reason for scepticism is that international agreements, including those setting up international organizations, are (in Hobbes’ words) ‘covenants without the sword’. However, granting the premise that the international arena is a Hobbesian ‘state of nature’ (see Hobbes, T.§6), it still does not follow that considerations of justice are irrelevant to it. We should bear in mind the conclusion that Hobbes himself drew from the insecurity of a ‘state of nature’: that a sovereign was needed to supply the security that would render it safe to keep covenants and thus ensure their validity. This, as Hobbes recognized, leaves the sovereigns in a ‘state of nature’ with one another. But a more thoroughgoing Hobbesian might plausibly derive the conclusion that the lack of an international sovereign threatens precisely the calamities that states were supposed to protect against: violent death in war and lack of the conditions of ‘commodious living’. Even if the threat of war is receding, military expenditure diverts resources away from life-saving alternative uses. Moreover, the faltering international response to such issues as global warming and biodiversity may suggest that only international-level coercion is capable of maintaining the long-term conditions of life on the planet.

It is also worth recalling that, according to Hobbes, there can be obligations of justice even in a ‘state of nature’, since one of the ‘laws of nature’ is ‘that men perform their covenants made’ and justice consists in keeping covenants. Admittedly, the obligation is said to hold only when the other party has performed its part and there is no ‘new cause of fear’ arising subsequent to the making of the covenant. But this is enough to establish that, in general, states behave unjustly in reneging on an international agreement that the other signatories have adhered to.

A second line of argument derives from the romantic nationalist movement of the nineteenth century. This assumes that each state is the home of a community whose members share a certain view of the world which gives meaning to their lives (see Nation and nationalism). In its strongest (and most logically coherent) form, this doctrine

concludes that, since there is no moral community beyond the state, there is no basis for saying that anything a state does can be unjust. A weaker version (especially associated with Michael Walzer (1977)) maintains that the notion of distributive justice makes sense only within each community, since this consists in the distribution of each kind of good in accordance with the meaning that the community attributes to that good (see Justice §2). Clearly, there is no global community with shared understandings of the meanings of goods, so it follows that the notion of distributive justice has no application outside a state.

An obvious objection to both versions is that scarcely any states are moral communities of the postulated sort. Moreover, many people do in fact have views (albeit often somewhat inchoate) about what international justice requires. It is not necessary to have very sophisticated ideas about the shared meanings of goods to arrive at the view that there is something fundamentally wrong with a world in which millions of people are starving to death and dying of easily preventable diseases while others are reduced to buying jewelled necklaces for their pets to find a way of disposing of their incomes.

3 Cosmopolitan justice

The American Declaration of Independence, the French Declaration of the Rights of Man and the Citizen and the Universal Declaration of Human Rights all proclaim the equal value of all human beings. The two main strands of modern moral philosophy, Kantianism and utilitarianism, are likewise based on an axiom of equal value: for Kant everyone is a participant in the 'kingdom of ends' and entitled to equal respect, while for Bentham 'everybody counts for one' in the sense that all units of utility are given the same weight regardless of the identity of their owners.

In practice, however, many philosophers operate as if the axiom's validity extended only to fellow citizens, and much popular discourse follows the same lines. Yet a universal principle cannot simply have its application arbitrarily truncated at a state boundary. In revulsion against this, it is tempting to adopt an equally absolutist cosmopolitan line, according to which each of us should regard ourselves as citizens of the world, with no special obligations to the fellow citizens of our own state. But if (see §2) justice entails keeping covenants, may we not be said to be bound to our fellow citizens by innumerable implicit covenants? If you have contributed throughout your adult life to the national system of social insurance, for example, you have a claim on your fellow citizens for support in old age or in the event of becoming disabled and incapable of working. Your co-nationals owe this to you in virtue of a formal system of mutual aid: they do not have the same obligation to somebody (perhaps equally deserving or needy) who has not participated in the scheme.

There is nothing in this which is inconsistent with cosmopolitanism, understood as the extension to all human beings of the axiom of fundamental equality. Utilitarians normally recognize that the universal good is likely to be most effectively pursued indirectly. Similarly, a Kantian can agree that an element of particularism may consistently be willed as a universal law. The limits of this particularism are, however, quite tightly drawn. The value of a family to its members depends on a certain material base: formal autonomy without adequate resources is a cruel cheat. In a similar way, the picture of a world in which people have special obligations to compatriots loses its moral attractiveness unless each country has enough material resources to provide at least the basic essentials in the way of food, shelter, sanitation and medical care to its inhabitants (see International relations, philosophy of §§3, 4).

A utilitarian will have little difficulty in recognizing that, however inefficient the transfer process might be, a situation in which the richest fifth of the world's population is sixty times better off than the poorest fifth must be one in which utility would be increased by a shift in resources. The Kantian legacy leaves more room for interpretation. All that can be said here is that contemporary neo-contractarianism seems naturally to lead to an equally pressing obligation to shift resources from the richest to the poorest countries. Thus, if we imagine people from everywhere in the world, ignorant of their personal identities, meeting in a Rawlsian 'original position' to choose principles of global justice, we must regard it as inconceivable that they would not seek to guard against suffering the fate of the majority of inhabitants of the poorest countries (see Rawls, J. §1). And if we follow Thomas Scanlon's (1982) proposal that we ask what principles could not reasonably be rejected by people seeking agreement with others under conditions that ruled out the exercise of bargaining power, we would have to say that those from the poorest countries could reasonably reject principles giving rise to a world order that left so many living in degrading deprivation (see Contractarianism; Justice).

4 Problems with cosmopolitan justice

The absorptive capacities of poor countries set severe limits on the amount of global redistribution that would be possible. Nevertheless, let us suppose that cosmopolitan principles would require rich countries to transfer ten per cent of their incomes to poor ones. This would imply a tenfold increase in aid from the Scandinavian countries, and an increase of fifty times and more from the current laggards such as the UK and USA. It is widely assumed not only that this is 'politically infeasible' but that this infeasibility somehow casts doubt on either the premises or the reasoning process leading to the conclusion about what justice demands. Yet many countries must on the same criteria be judged grossly unjust domestically, and this is not normally felt to cast doubt on the conclusion. Is there any rationale for the difference in attitude?

Thomas Nagel, in *Equality and Partiality*, has argued that lying behind the 'political infeasibility' is a moral consideration: that people in rich countries 'have sufficient reason to resist if they can' a 'radical drop in the standard of living of [themselves] and [their] famil[ies]' (1991: 174). Nagel denies that this reason amounts only to 'pure selfishness', but all he claims is that 'personal interests and commitments' are relevant (1991: 172). It is, however, hard to see how any morally compelling commitments could not be met with a modestly reduced standard of living, so 'personal interests' (that is, the unwillingness to do what is just) seems to be all that is left. Even if some commitments such as private schooling or a car as a graduation present had to be abandoned, it seems grotesque to mention this in the same breath as the plight of parents in poor countries who have to watch helplessly while their children die of diseases caused by malnutrition and lack of sanitation.

For Nagel, international redistribution does not create problems that are distinctive in principle: 'some poor countries such as India and Mexico have wealthy minorities' who can, apparently, equally legitimately resist redistribution (1991: 170). Other people, however, seem to regard the problem of international justice as different. A possible rationale might run as follows. States have a coercive apparatus at their command which *could* be used to bring about justice even if it is not so used and (given the constellation of political forces) is in many cases not likely to be so used. There is no international agency with the coercive power to create international justice, even under ideal political conditions. This means that moves towards international justice will require voluntary action by governments. We cannot reasonably expect the electorates of rich countries to vote for politicians who promise if elected to take steps to reduce their standard of living significantly. But if we cannot reasonably expect somebody to do something, we can scarcely mount a serious criticism of that person for not doing it. This must suggest that failing to do it is not appropriately described as unjust.

This argument is manifestly a close cousin of the Hobbesian one dissected in §2 in the link it makes between motivation, institutions and justice. As before, we can respond that the correct conclusion to draw is that justice requires appropriate coercive institutions. If domestic taxation were purely voluntary, it would no doubt be regarded as unreasonable to expect people to pay a large proportion of their incomes to the government, yet nobody thinks much of governments taking in over forty per cent of the national income for the provision of public services and cash benefits. If rich countries could be coerced by an international organization, there would, similarly, be no reason for a tax rate of, say, ten per cent to be regarded as extraordinary.

The absence of such a coercive institution has meant that the populations of rich countries have had little self-interested motivation to correct the situation. However, as Annette Baier has argued, 'people grow more reckless of the lives of others as their own lives become more wretched, insecure, and intolerable' (1991: 34). And while it would be better in every respect if rich countries acted under their own momentum to transfer resources to poor countries, the development and proliferation of devastating nuclear and (relatively portable) chemical and biological weapons means that not to do so 'may be not just inhumanity [but] *folly* if the excluded do feel resentment and have power to make it felt' (1991: 54; original emphasis).

BRIAN BARRY
MATT MATRAVERS

References and further reading

Baier, A. (1991) 'Violent Demonstrations', in R.G. Frey and C.W. Morris *Violence, Terrorism, and Justice*, Cambridge: Cambridge University Press, 33-58.(One of a several essays discussing the problem of terrorism; referred to in §4.)

Barry, B. (1989) *Theories of Justice*, vol. 1, *A Treatise on Social Justice*. Hemel Hempstead: Harvester Wheatsheaf, Berkeley and Los Angeles, CA: University of California Press, 183-9.(Discussion of Rawls, with special reference to international justice.)

Barry, B. (1995a) *Justice as Impartiality*, vol 2, *A Treatise on Social Justice*, Oxford: Clarendon Press.(Develops the approach introduced by Scanlon; see 67-72 for initial exposition.)

Barry, B. (1995b) 'Spherical Justice and Global Injustice', in D. Miller and M. Walzer (eds) *Pluralism, Justice, and Equality*, Oxford: Oxford University Press, 67-80.(Discussion of Walzer, including implications for international justice.)

Beitz, C. (1979) *Political Theory and International Relations*, Princeton, NJ: Princeton University Press, part III, 125-76.(Contains an argument for the global application of Rawls' 'difference principle', which prescribes making the worst-off group as well off as possible.)

Brown, C. (1992) *International Relations Theory: New Normative Approaches*, Hemel Hempstead: Harvester Wheatsheaf.(A lucid explanation of the 'liberal versus communitarian' theme in relation to international normative issues; good discussion of Kant and Hegel.)

Brown, C. (ed.) (1994) *Political Restructuring in Europe: Ethical Perspectives*, London: Routledge.(Part II, pp. 69-184, contains a useful and accessible cosmopolitan/anti-cosmopolitan debate, with contributions by O. O'Neill, T.W. Pogge, C.R. Beitz, D. Miller and C. Brown.)

Hayek, F.A. (1976) *The Mirage of Social Justice*, vol. II, *Law, Legislation and Liberty*, Chicago, IL: University of Chicago Press.(Referred to at the beginning of the entry.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991, ch. 13.(Justice in the 'state of nature', referred to in §2.)

Honderich, T. (1994) *Hierarchic Democracy and the Necessity of Civil Disobedience*, London: South Place Ethical Society.(The 69th Conway Memorial Lecture, this argues that the scale of inequality within and between countries demands civil disobedience and would justify violence if there were any prospect of its succeeding.)

Nagel, T. (1991) *Equality and Partiality*, New York: Oxford University Press, ch. 15, 169-79.(Quoted in §4.)

O'Neill, O. (1986) *Faces of Hunger: An Essay on Poverty, Development and Justice*. London: George Allen & Unwin.(A Kantian argument for the redistribution of material resources.)

Pogge, T.W. (1989) *Realizing Rawls*, Ithaca, NY: Cornell University Press, part III, 211-80.(Discussion of possible international application of Rawls' principles.)

Rawls, J. B. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, ch. 3, 118-92.(Sets out the idea of an 'original position' mentioned in §3.)

Rogers, P. (1995) 'Security Means More Now Than Defence'. Parliamentary Brief 3, no. 5: 30-1.(Makes the point about 'suitcase bombs' and their implications for the security of rich countries in line with the end of §4.)

Scanlon, T.M. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds) *Utilitarianism and Beyond*. Cambridge: Cambridge University Press, 103-28.(Discussed in §3.)

Shue, H. (1995) 'Avoidable Necessity: Global Warming, International Fairness and Alternative Energy', in I. Shapiro and J.W. DeCew *Theory and Practice*, NOMOS XXXVII. New York: New York University Press, 239-64.(To avoid potentially catastrophic global warming, sharp reductions in the emission of 'greenhouse gases' are required. What would be a fair distribution of the burden internationally, and how does 'political feasibility' enter in to it?)

Singer, P. (1972) 'Famine, Affluence and Morality', *Philosophy and Public Affairs* 1 (3): 229-43.(Famous article making a strong case on utilitarian premises for large-scale global redistribution.)

UNICEF (1992) *The State of the World's Children*, Oxford: Oxford University Press.(New editions each year; this one includes the estimate of global military expenditures in relation to the national incomes of poor countries given in §2.)

Walzer, M. (1977) *Spheres of Justice: A Defense of Pluralism and Equality*. New York: Basic Books; Oxford: Blackwell.(This contains the theory of distributive justice associated with Walzer and discussed in §2.)

Justification, epistemic

The term 'justification' belongs to a cluster of normative terms that also includes 'rational', 'reasonable' and 'warranted'. All these are commonly used in epistemology, but there is no generally agreed way of understanding them, nor is there even agreement as to whether they are synonymous. Some epistemologists employ them interchangeably; others distinguish among them. It is generally assumed, however, that belief is the target psychological state of these terms; epistemologists are concerned with what it takes for a belief to be justified, rational, reasonable or warranted. Propositions, statements, claims, hypotheses and theories are also said to be justified, but these uses are best understood as derivative; to say, for example, that a theory is justified for an individual is to say that were that individual to believe the theory (perhaps for the right reasons), the belief would be justified.

Historically, the two most important accounts of epistemic justification are foundationalism and coherentism. Foundationalists say that justification has a tiered structure; some beliefs are self-justifying, and other beliefs are justified in so far as they are supported by these basic beliefs. Coherentists deny that any beliefs are self-justifying and propose instead that beliefs are justified in so far as they belong to a system of beliefs that are mutually supportive. Most foundationalists and coherentists are internalists; they claim that the conditions that determine whether or not a belief is justified are primarily internal psychological conditions (for example, what beliefs and experiences one has). In the last quarter of the twentieth century, externalism emerged as an important alternative to internalism. Externalists argue that one cannot determine whether a belief is justified without looking at the believer's external environment. The most influential form of externalism is reliabilism.

Another challenge to traditional foundationalism and coherentism comes from probabilists, who argue that belief should not be treated as an all-or-nothing phenomenon: belief comes in degrees. Moreover, one's degrees of beliefs, construed as subjective probabilities, are justified only if they do not violate any of the axioms of the probability calculus. Another approach is proposed by those who advocate a naturalization of epistemology. They fault foundationalists, coherentists and probabilists for an overemphasis on a priori theorizing and a corresponding lack of concern with the practices and findings of science. The most radical naturalized epistemologists recommend that the traditional questions of epistemology be recast into forms that can be answered by science.

An important question to ask with respect to any approach to epistemology is, 'what implications does it have for scepticism?' Some accounts of epistemic justification preclude, while others do not preclude, one's beliefs being justified but mostly false. Another issue is the degree to which the beliefs of other people affect what an individual is justified in believing. All theories of epistemic justification must find a way of acknowledging that much of what each of us knows derives from what others have told us. However, some epistemologists insist that the bulk of the history of epistemology is overly individualistic and that social conditions enter into questions of justification in a more fundamental way than standard accounts acknowledge.

1 Epistemic and non-epistemic justification

Decisions, actions, policies, procedures, punishments, laws, rules and host of other things can be justified or unjustified. They are justified only if there are adequate reasons for them, and unjustified if there are not. Epistemic justification is concerned with the justification of beliefs, and hence with reasons for believing. However, being epistemically justified is not simply a matter of having adequate reasons for believing, since there can be reasons for believing a claim that are not epistemic. If I offer you a million dollars to believe that the earth is flat, I have given you a reason to have this belief, but it is not an epistemic reason. The most controversial use of this point was made by Pascal and James, each of whom argued that we have such reasons to believe in God (see James, W. §4; Pascal, B. §6).

A straightforward way to understand the distinction between epistemic and non-epistemic reasons for belief is in terms of the end, or goal, that the reason aims at promoting. If *X* is a valuable end and if bringing about *Y* increases the likelihood of *X* occurring then, all else being equal, one has a reason to bring about *Y*. Different kinds of reasons are distinguished in terms of different kinds of ends. If *X* is an economic end, one has an economic reason to engage in *Y*. If *X* has broader pragmatic value - it improves your health, enhances your friendships and so on -

one has a more broadly pragmatic reason to engage in *Y*. Analogously, if *X* is epistemically valuable, one has an epistemic reason for *Y* (see Reasons for belief).

The question of what makes a reason an epistemic reason thus reduces to the question of what constitutes an epistemic end. A commonly accepted answer is that an end is epistemic in so far as it is concerned with believing what is true and not believing what is false. Once epistemic reasons are distinguished from non-epistemic ones, a question arises as to whether the latter can ever override former in cases where the two conflict. W.K. Clifford (1879), arguing against William James, said that they can not, claiming that we have an absolute obligation to believe only that for which we have adequate evidence. This is a difficult position to defend, however. Suppose a catastrophe could be avoided by believing a claim for which there is little or no evidence - for example, a terrorist threatens to detonate a bomb that will kill thousands of people unless you somehow get yourself to believe that the earth is flat. Clifford is committed, implausibly, to saying that under such conditions you have an obligation not to believe that the earth is flat.

Although it may be possible for non-epistemic reasons for belief to override epistemic ones, it is also rare. One explanation of its rarity is that even if we become convinced that we have non-epistemic reasons to believe a claim, this is ordinarily not enough to prompt belief. By contrast, if we become convinced that there is good evidence (that is, good epistemic reasons) for a claim, we usually automatically come to believe it. This suggests that if we are to acquire beliefs for which we realize we lack adequate evidence, we may have to go to extraordinary lengths. For example, one might have to resort to post-hypnotic suggestion or some equally extreme measure. However, such measures will typically have unfortunate consequences, especially since beliefs cannot be altered in a piecemeal fashion. Believing the earth is flat, for instance, will involve dismissing as misleading or fabricated a huge amount of information that on the face of it indicates that the earth is spherical. Dismissing such information in turn will adversely affect the overall accuracy of one’s belief system and with it the overall effectiveness of one’s decision making, with the likely result that one will be less able to achieve one’s goals in a wide variety of contexts. The lesson, then, is that except in the most unusual kinds of cases - the case of the terrorist, for example - the costs associated with an effort to get oneself to believe that for which one lacks adequate evidence will not be worth the effort. Thus, what we have reasons to believe all things considered - that is, when all our ends are taken into account - is usually identical with what we have epistemic reasons to believe.

2 Foundationalism and coherentism

According to foundationalists, epistemic justification has a hierarchical structure. Some beliefs are self-justifying and as such constitute one’s evidence base. Others are justified only if they are appropriately supported by these basic beliefs. Foundationalists differ among themselves as to what conditions have to be met in order for a belief to be basic and what conditions have to be met in order for other beliefs to be appropriately supported by basic beliefs. The strictest versions of foundationalism require that a belief be infallible, indubitable or incorrigible if it is to be self-justifying. A belief is infallible if it is impossible to have the belief and for it to be false; a belief is indubitable if it is impossible to doubt whether it is true; and a belief is incorrigible if it is impossible to have good reasons for thinking it is false. The strictest versions of foundationalism also impose stringent constraints on the support relation, restricting it to logical implication and enumerative induction; a non-basic belief is justified only if it is implied or inductively supported by one’s basic beliefs. Many of the most influential figures in the history of epistemology are strict foundationalists (for example, Descartes, Hume, Berkeley and Locke) as are many of the most important epistemologists in the twentieth century (for example, Russell, Ayer and Carnap).

Modest versions of foundationalism relax the requirements of justification. For a belief to be basic, it does not have to be infallible, indubitable or incorrigible, but it does have to be intrinsically probable; the fact that one has the belief must itself make it likely that the belief is true. Similarly, the support relation is broadened beyond logical implication and enumerative induction to include inferences to the best explanation and other support relations (see Inference to the best explanation; Theoretical (epistemic) virtues).

Another distinction is between those foundationalists who insist on objective conditions of justification and those who permit subjective conditions. For example, Foley (1993) argues that one’s basic beliefs are those that are noncontroversial for one to use as evidence, given one’s other beliefs and one’s own deepest epistemic standards. He adds that one’s other beliefs are justified in so far as one would agree on ideal reflection that one’s basic beliefs make them probable. This is a version of foundationalism, but it is a subjective version.

For the foundationalist, the reigning metaphor for epistemology is that of a building whose foundation of basic beliefs supports additional stories of non-basic beliefs. For the coherentist, the reigning metaphor is that of a web made up of, and deriving its strength from, mutually supportive beliefs.

According to coherentists, the primary objects of justification are not individual beliefs but, rather, belief systems. A belief system is justified if its component beliefs cohere in an appropriate way. Individual beliefs are justified by virtue of belonging to such a set of beliefs. Thus, for the coherentist, epistemic justification is a holistic notion rather than a hierarchical one. The picture is not of basic beliefs being intrinsically justified and then passing on their justification to other beliefs; it is, rather, of justification emerging when one's belief system hangs together, or coheres.

Nineteenth-century philosophers such as Hegel and Bradley were coherentists, and in the twentieth century it was critiques of foundationalism that led to a resurgence of coherentism. Especially influential was the work of Wilfrid Sellars, who argued that no belief is self-justifying; beliefs can be justified only by appeal to other beliefs. Sellars was particularly forceful in his criticism of what he called the 'myth of the given', which is the notion that beliefs about our own sense experiences cannot be mistaken, since these experiences are directly presented, or given, to us.

It was then a small step from the claim that no belief is self-justifying to a coherence theory of epistemic justification. The step was made with the help of the observation that at least some beliefs are justified by reference to other beliefs that are themselves justified. This observation invites the question 'what justifies these latter beliefs?', and if the answer is 'still other justified beliefs', we can ask of the latter, 'what justifies them?' If pursued, this line of questioning suggests there are only three different ways of thinking about the structure of epistemic justification. The first is that there is an infinite regress of justifiers. Belief *A* is justified by belief *B*, *B* is justified by *C* and so on *ad infinitum*. However, most philosophers have thought that if there is no way to stop this regress, none of the early members of the series is justified and scepticism is the result; nothing is epistemically justified (see Scepticism). A second option is the one endorsed by foundationalists; belief *A* is justified by belief *B*, *B* by *C* and so on until a self-justifying belief ends the regress. The critiques of Sellars and others caused many to reject this second option, and thus they seemed to be left with no option but to endorse circular justification of some sort. Belief *A* is justified by belief *B*, *B* by *C* and so on until we reach a belief that is justified by *A* and other beliefs early in the series. In other words, the set of beliefs is mutually supportive, with belief *A* being justified by beliefs that *A* itself helps to justify.

Coherence among beliefs is at least a matter of consistency. If a set of beliefs is inconsistent, it is impossible for all the beliefs in the set to be true, and hence they are not mutually supportive. However, consistency is not enough for coherence; beliefs that are altogether unrelated to one another are consistent, but they are not mutually supportive. Some coherentists, F.H. Bradley for example, suggest that mutual entailment is required for coherence; every member of a coherent set should be deducible from other members of the set. However, most coherentists propose far less stringent conditions. BonJour (1985), for example, thinks of coherence as more than mere consistency but less than mutual entailment; it comes in degrees, with the degree increasing with the number of inferential connections among the component beliefs of the set and decreasing with the number of unexplained anomalies.

An objection to coherentism is that even the minimal condition of coherence - namely, consistency - is not an absolutely necessary condition of justification. It is ordinarily epistemically desirable to avoid inconsistency, but sometimes the costs are too high. The case of the lottery can be used to illustrate this.

Imagine a lottery of a million tickets, and suppose you are justified in believing that the lottery is fair and as such has only one winning ticket. Suppose also you have no reason to distinguish among the tickets concerning their chances of winning. So, the probability of ticket number 1 losing is .999999, the probability of ticket number 2 losing is .999999, and similarly for each of the other tickets. Thus, you have extraordinarily strong reasons to believe of ticket number 1 that it will not win, and accordingly it seems that you are justified in believing this. If we deny that you are justified, we will be hard-pressed to claim that very many of your other beliefs are justified, since the chances of their being in error are at least as great. However, your position with respect to ticket number 2, ticket number 3 and right through to ticket number 1,000,000 is the same as your position with respect to ticket

number 1. So, if you are justified in believing that number 1 will lose, it seems that you are justified in believing this of each of the other individual tickets as well. But then, if you also justifiably believe that one ticket (you know not which) will win, you have justified but inconsistent beliefs. Hence, consistency is not an absolutely necessary condition of epistemic justification, contrary to what coherentists claim (see Paradoxes, epistemic §1).

3 Internalism and externalism

Most foundationalists and coherentists share a commitment to internalism. In its strongest form, internalism is the view that one can always determine by careful introspection whether one's beliefs are justified. A motivation for strong internalism - found in epistemologists as otherwise diverse as Descartes, Locke, Chisholm and BonJour - is that we have a *prima facie* duty to have epistemically justified beliefs. If so, we are *prima facie* blameworthy if we have unjustified beliefs. But then, internalists argue, we must have access to the conditions that determine whether our beliefs are justified. Otherwise, we could do our epistemic best and hence be blameless but still have unjustified beliefs (see Epistemology and ethics).

Modest internalism is any view that has epistemic justification supervening upon internal psychological conditions, whether or not these conditions are introspectively accessible to the believer. To say that epistemic justification supervenes on psychological conditions is to say that whenever a belief *p* is epistemically justified, there is a set of psychological properties β such that necessarily if one has β and believes *p*, then one's belief is justified (see Supervenience).

A number of epistemologists (for example, BonJour 1985; Chisholm [1966] 1989; Foley 1993 and Lehrer 1974) propose conditions of epistemic justification that are internal psychological conditions but that are also complex enough for it to be implausible to think that we can always determine by introspection whether or not they obtain. Such epistemologists are best regarded as modest internalists. By contrast, externalists insist that some of the key prerequisites of epistemic justification obtain outside the skin of the believer, for example, those concerning the believer's environment, history or social context.

Externalist accounts of epistemic justification are encouraged by the presupposition that when epistemic justification is added to true belief, knowledge is the result if there are no Gettier problems present (see Gettier problems). Alvin Goldman (1986) is an especially instructive figure in this regard. He was one of the earliest champions of a reliabilist account of knowledge. According to reliabilists, one knows that *p*, at least in the simplest cases, if one has a true belief *p* that is the product of a generally reliable cognitive process (see Reliabilism). With this account of knowledge in hand, Goldman argued that because epistemic justification is by definition that which has to be added to true belief to get knowledge, epistemic justification must essentially be a matter of one's beliefs having been produced by highly reliable cognitive processes. Since the reliability of a cognitive process is in part dependent on the external environment in which one is operating, this is an externalist account of epistemic justification.

Reliabilism is the most influential form of externalism, and the most influential reliabilists include Goldman, Dretske (1981), Armstrong (1973) and Sosa (1991). Despite their differences, these epistemologists share a commitment to the view that justification is basically a matter of there being a correct fit between one's environment on the one hand and one's cognitive processes and intellectual practices on the other, where a fit is correct if it has a tendency to produce true beliefs and not produce false beliefs. An offshoot of reliabilism is a view, championed by Alvin Plantinga (1992), that makes epistemic justification a matter of having beliefs that are the products of a properly functioning cognitive system, where a system is properly functioning if it is functioning in the way it was designed to function (either by natural selection or God) in the environment that it was designed for.

The disputes between internalists and externalists illustrate that the notion of epistemic justification is used by epistemologists in different senses. One way to determine how the notion is being used is to inquire whether the epistemologist is assuming that justification is by definition what has to be added to true belief in order to get a serious candidate for knowledge. Many externalists have been explicit about defining justification in this way. For others it is an implicit working assumption; if a proposed account of justification is such that justified true beliefs are not always good candidates for knowledge, the proposed account is for that reason inadequate. By contrast, many internalists understand justification in some independent way, leaving it an open question whether

justification when added to true belief generates a serious candidate for knowledge.

Confusion can occur when epistemologists use the notion in more than one way. For example, they may want to think of epistemic justification in terms of what a responsible believer would believe and also in terms of what has to be added to true belief to get knowledge. Confusions of this sort are not entirely unexpected, given that some of the most influential figures in epistemology have thought that one and the same notion could capture both ideas. Descartes and Locke, for example, were concerned with what is involved in being a responsible believer, and each thought that this was in large part a matter of believing only that for which we have adequate (internal) evidence. However, they also thought that by being responsible believers, we can be assured of acquiring knowledge.

Few epistemologists are so sanguine any more, but the lesson is not that the one or the other aspect of Cartesian and Lockean projects have to be abandoned. The lesson, rather, is that there are different projects for epistemologists to pursue. One project, roughly put, is to explore what is required for us to put our own intellectual house in order. Another is to explore what is required for us to stand in a relation of knowledge to our environment. It is not unnatural to report the results of both kinds of exploration using the language of justification and rationality, given the history of these terms. But this means, in turn, that the terms have distinct senses, one tending to be externalist, objective and closely connected with knowledge, the other tending to be internalist, more subjective and closely connected with responsible believing.

4 Probabilism

Whereas reliabilism is an externalist alternative to traditional foundationalism and coherentism, probabilism is an internalist alternative. Probabilists note that one's doxastic options with respect to a proposition are not limited to believing it, disbelieving it or withholding judgment on it. Rather, one can have varying degrees of belief, or confidence, in the truth of a proposition. Probabilists often add that these degrees of belief can be represented as subjective probabilities, with a subjective probability of 1 representing maximum confidence in the truth of a proposition, a subjective probability of 0 representing no confidence in its truth, and the numbers in the interval (0,1) representing degrees of confidence between the two extremes. Probabilists then argue that one's degrees of belief must be coherent if they are to be justified, and that they are coherent if they obey the probability calculus.

Consider some simple examples. According to the probability calculus, a proposition can be no more probable than a proposition it implies. Thus, if *P* implies *Q* and if one believes *P* with more confidence than *Q*, one's degrees of belief are incoherent and hence unjustified. Similarly, if *P* and *Q* are mutually exclusive propositions - that is, if it is not possible for both to be true - the probability of (*P* or *Q*) is equal to the probability of the first disjunct plus the probability of the second disjunct. Suppose, then, that one believes *P* with degree of confidence *x*, *Q* with degree of confidence *y*, and the disjunction (*P* or *Q*) with degree of confidence *z*. If $z \neq x + y$, one's degrees of belief are incoherent and as a result unjustified.

'Dutch book arguments' are sometimes used by probabilists to illustrate the undesirability of having incoherent degrees of belief. These arguments establish that if you are willing to post betting odds on the truth of propositions in accordance with your degrees of belief - for example, if you are twice as confident of the truth of *P* as you are of its negation, you are willing post 2:1 odds on the truth of *P*, giving an opponent the option of betting on *P* or against it - you will be vulnerable to having someone make a dutch book against you if your degrees of belief are incoherent. A dutch book is made against you when your opponent makes a series of bets such that you will suffer a net loss no matter how the events you are betting on turn out.

The most influential form of probabilism is Bayesianism, which proposes both an account of synchronic rationality (what is required for one's degrees of belief to be justified at any given moment) and also an account of diachronic rationality (what is required for one's degree of belief to change rationally over time). With respect to former, Bayesians insist on coherence. With respect to the latter, they insist that one should modify one's beliefs over time, in response to new information, in accordance with Bayes' Theorem (see Probability theory and epistemology).

One advantage of probabilism is that it sidesteps the lottery problem. Probabilists simply observe that it is rational to have a high degree of belief in the individual propositions in the lottery - that ticket number 1 will lose, that ticket number 2 will lose and so on - but an extremely low degree of confidence in the conjunction of these propositions, a conjunction that is equivalent to the proposition that no ticket will win. They are then free to leave

the matter at that, without having to decide whether it is rational to believe *simpliciter* the individual propositions.

Nevertheless, the most persistent objection against probabilism is analogous to the objection that the lottery problem raises for traditional coherentism. Most generally expressed, the objection against traditional coherentism is that sometimes we can be justified in having beliefs that we know are less than ideally accurate. In particular, we can be justified in having beliefs that we know to be inconsistent and hence that we know cannot all be true. Similarly, we sometimes can be justified in having degrees of belief that we know are incoherent and hence cannot possibly reflect the objective probabilities. The clearest examples are non-contingent propositions, whose objective probability we know is either 1 or 0. For example, Goldbach's conjecture, which asserts that every even number can be expressed as the sum of two primes, is either necessarily true or necessarily false. If the former, it has an objective probability of 1, if the latter, it has an objective probability of 0. But since the conjecture has never been proven or disproven, we can be justified in having a degree of belief in the conjecture that is neither 1 nor 0.

5 Naturalized epistemology

Naturalized epistemology is the name given to a cluster of views that insist upon a close relationship between epistemology and the practices and findings of science. Naturalized epistemologists criticize foundationalists, coherentists and probabilists for their overemphasis on a priori theorizing. If we want to know how cognition actually works, why it is usually reliable, and when it has a tendency to go wrong, they say we need to turn to science. Psychology and neurobiology can tell us how cognition works; biology and especially evolutionary theory can provide us with an explanation as to why cognition is generally reliable; and cognitive psychology can give us information about the kinds of intellectual mistakes we have a tendency to make, so that we can guard against them.

Quine (1969) coined the phrase 'naturalized epistemology' and is one of its most influential proponents. In his most trenchant moments, he recommends that epistemology should become a chapter of psychology. This is to be accomplished by recasting the traditional questions of epistemology into forms that can be answered by psychology. For example, the question 'Does sensory experience give us good reasons to have beliefs about the external world?' is replaced with 'How do sensory experiences in fact cause us to believe what we do about the external world?'. The idea is to make epistemology a part of natural science and hence, in Quine's view, respectable.

In this radical form, naturalized epistemology threatens to become non-normative. It describes how a relatively restricted sensory input produces a rich array of beliefs, but it does not assess which procedures, methods and practices we are justified in using. Naturalized epistemologists, including Quine himself, have been eager to respond that they are not abandoning the normative element within epistemology. One response, emphasized by Alvin Goldman (1986), is that a naturalized approach to epistemology can coexist comfortably with a reliabilist account of epistemic justification. Epistemology, on this conception, is normative, since it tells us we are justified in employing an intellectual procedure, method or practice in so far as the practice is reliable; but it is also naturalized, since science that tells us which procedures, methods and practices are reliable.

6 Scepticism

One way of marking differences among various epistemological theories is to ask what, if anything, does the theory imply about radical sceptical hypotheses, which imply that most of one's beliefs are mistaken, not just in detail but seriously mistaken. For example, the evil demon hypothesis, which Descartes tried to refute, is that unknown to me I am under the control of a demon who is manipulating me and my environment in such a way that almost all my beliefs about the external world are false. Such hypotheses seem far-fetched, but they do illustrate in a dramatic way what seems to be a legitimate question, namely, whether or not our beliefs might be largely in error.

In so far as naturalized epistemologists feel free to bring the findings of science to their epistemologies, they are presupposing the falsity of such sceptical worries. By contrast, reliabilists can be neutral about sceptical hypotheses. They assert that our beliefs must be the products of reliable cognitive processes in order to be justified, but they need not take a stand on whether the cognitive processes we actually employ are reliable. Strong foundationalism purports to provide a refutation of sceptical hypotheses. Descartes, for example, thought that his

method of doubt could establish with certainty a set of truths, including the existence of a benevolent and omnipotent God, which could then be used to disprove the evil demon hypothesis and other such sceptical hypotheses (see Descartes, R. §4). Modest versions of foundationalism are usually regarded as more plausible than Descartes' version, but they also provide correspondingly weaker responses to the sceptic. The same is true of coherentism; it is difficult to argue that it is altogether impossible for coherent beliefs to be drastically mistaken.

One stance to take towards sceptical worries, available to coherentists, moderate foundationalists and reliabilists alike, is to concede that it is part of our human intellectual predicament that such worries cannot be definitively banished; we have no choice but to live with them and the uncertainty they entail. It has proved difficult for epistemologists to be content with this uncertainty, and this in turn has motivated attempts to show that sceptical worries are self-referentially incoherent. The argument is that, in raising radical worries, would-be sceptics inevitably make use of the very intellectual faculties and methods about which they are raising doubts. In doing so, they are presupposing the general reliability of these faculties and methods. Hence, it is incoherent for them to entertain the idea that they might be unreliable. An objection against this anti-sceptical argument is that it fails to appreciate that the argumentative strategy of sceptics can be entirely negative. Their aim need not be to establish any positive thesis, not even the thesis that our faculties and methods are untrustworthy. Rather, their strategy can be to assume for the sake of argument that our favourite faculties, procedures and methods are reliable, and then to illustrate that these faculties, procedures and methods, if applied rigorously, will generate evidence that undermines their pretence of reliability.

Another anti-sceptical position is that sceptical hypotheses are metaphysically impossible. For example, Davidson (1986) argues that the nature of belief is such as to rule out the possibility that our beliefs are largely in error since, at least in the most simple cases, the objects of our beliefs are the causes of them. Hence, our beliefs must be largely true (see Davidson, D.). In a similar spirit, Putnam (1987) argues that, in thinking about the world, it is impossible to separate our conceptual contributions from what is 'objectively there', and that plausible theories of reference and truth must acknowledge this impossibility. But in turn, Putnam says, this implies that the world cannot be entirely different from what our beliefs represent it to be (see Putnam, H.).

Such theories of truth, belief and reference are interesting and debatable as metaphysical positions, but for the purpose of laying sceptical worries to rest they are of limited value, especially since the intricate arguments used to defend these theories are themselves subject to sceptical doubts. Regardless of how we marshal our intellectual resources, there can be no non-question-begging assurances that the resulting inquiry is reliable, and this applies to metaphysical inquiries into the nature of truth, belief and reference as much it does to any other kind of inquiry.

7 Testimony and social epistemology

Another issue that marks differences among various approaches to epistemology is the degree to which the beliefs of other people enter into questions of what an individual is justified in believing. It is not seriously debatable that much of what each of us believes derives from others. So, all theories of epistemic justification must find a way to acknowledge the social interdependence of knowledge and justification. What separates epistemologists is how to do justice to this interdependence.

One way is to focus on the importance of testimony in our intellectual lives. A historically important approach to testimony, associated with Hume, claims that we all have good inductive reasons for thinking that others are reliable. Hence, we are generally justified in relying on what others tell us. There are at least two problems with this approach. First, it does not seem adequate to the breadth and depth of our dependence on others. As children, we acquire a huge set of beliefs from others, mostly without thought and long before we have the capacity to judge whether we have inductive reasons to trust them. Second, it is notoriously difficult to reconstruct the good inductive reasons we are supposed to have for trusting others. We rely on the testimony of strangers whose track record for truthfulness is unknown to us. Similarly, we rely of the opinions of experts in fields of inquiry about which we know so little that we are unable to verify reliability of the experts.

Another approach is to claim that we are always *prima facie* justified in accepting the testimony of others, even when we do not have good inductive reasons for trusting them. This approach is associated with Thomas Reid, who said that we all have a sense of credulity which disposes us to believe what others tell us. Reid regarded credulity as a cognitive faculty, analogous to vision, hearing, touch and memory. We are born with this faculty

and, like other native cognitive faculties, we are justified in relying on it unless problems arise, that is, unless we have concrete reasons to distrust it (see Reid, T.).

A third approach is that, in so far as I am entitled to trust in general my own cognitive functioning, I am also entitled to trust in general the cognitive functioning of others, since their cognitive faculties are broadly similar to mine and likewise their environment is similar to mine. This is not to deny there are important differences among people and their environments, but the differences arise against an enormous backdrop of similarity. Thus, it would be incoherent for me to trust my cognitive functioning unless I am also willing to trust that of others; consistency demands this.

Some epistemologists insist that all the above approaches are overly individualistic, since they try to account for the social interdependence of justification by focusing on how individuals should treat the testimony of others. Social conditions enter into questions of justification in a more fundamental way than this, these critics say. For example, a position hinted at by both Wittgenstein and Rorty is that one is justified in believing *p* only if the manner in which one comes to believe *p* is intellectually acceptable in one's community (see Wittgenstein, L. §12; Rorty, R.M.).

A claim that has even more radical implications for epistemology is that the target states of epistemic justification are not the beliefs of individual people but rather the opinions of social groups, where these opinions cannot be reduced to the opinions of the individuals comprising the group. The beliefs of individuals are justified only in a derivative sense. Either they belong to the set of opinions justifiably held by one's community, or they are appropriately related to this set of opinions.

See also: Epistemology, history of; Induction, epistemic issues in; Internalism and externalism in epistemology; Knowledge and justification, coherence theory of; Knowledge, concept of; Naturalized epistemology; Rational beliefs; Reasons for belief; Scepticism; Social epistemology; Testimony

RICHARD FOLEY

References and further reading

Alston, W. (1989) *Epistemic Justification*, Ithaca, NY: Cornell University Press.(Collection of important essays on foundationalism, externalism and internalism, reliabilism and other epistemological issues.)

Armstrong, D.M. (1973) *Belief, Truth, and Knowledge*, Cambridge: Cambridge University Press.(Early defence of a reliabilist theory of knowledge.)

BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.(Clear, concise criticism of foundationalism and a defence of coherentism.)

Chisholm, R. (1966) *The Theory of Knowledge*, Englewood Cliffs, NJ: Prentice-Hall, 3rd edn, 1989.(Influential and detailed defence of foundationalism and internalism.)

Clifford, W.K. (1879) 'The Ethics of Belief', in *Lectures and Essays*, vol. 2, London: Macmillan.(Defence of the view that it is always wrong to believe something for which there is insufficient evidence.)

Davidson, D. (1986) 'A Coherence Theory of Truth and Knowledge', in E. LePore (ed.) *The Philosophy of Donald Davidson*, London: Blackwell.(The most accessible of several articles in which Davidson argues that the nature of belief and reference makes it impossible for our beliefs to be radically incorrect.)

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(Information-theoretic defence of reliabilism.)

Foley, R. (1993) *Working Without A Net*, New York: Oxford University Press. (Discussions of subjective foundationalism, epistemic and non-epistemic reasons for belief, coherentism, probabilism and scepticism.)

Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(Influential and detailed defence of reliabilism.)

Kornblith, H. (ed.) (1994) *Naturalizing Epistemology*, Cambridge, MA: MIT Press. (Collection of influential articles by proponents of naturalized epistemology plus a helpful introduction by Kornblith.)

Lehrer, K. (1974) *Knowledge*, Oxford: Oxford University Press.(Important defence of coherentism and attack on foundationalism.)

Plantinga, A. (1992) *Warrant: The Current Debate*, New York: Oxford University Press.(Discussions of foundationalism, coherentism, probabilism and reliabilism, and a defence of Plantinga's own proper functioning account.)

Putnam, H. (1987) *The Many Faces of Realism*, LaSalle, IL: Open Court.(Defence of Putnam's internal realism

and its anti-sceptical implications.)
Quine, W.V. (1969) 'Epistemology Naturalized', in *Ontological Relativity and Other Essays*, New York: Columbia University Press.(The article most responsible for beginning the naturalized epistemology movement.)
Quine, W.V. (1990) *The Pursuit of Truth*, Cambridge, MA: Harvard University Press.(The first chapter contains a succinct and accessible account of Quine's epistemology.)
Schmitt, F. (ed.) (1994) *Socializing Epistemology*, London: Rowan & Littlefield.(Collection of articles on testimony and the social epistemology.)
Sellars, W. (1963) *Science, Perception, and Reality*, London: Routledge & Kegan Paul.(Difficult but influential statement of Sellars's epistemological and metaphysical positions.)
Sosa, E. (1991) *Knowledge in Perspective*, Cambridge: Cambridge University Press. (Collection of essays on justification, knowledge, reliabilism, foundationalism and coherentism.)

Justification, religious

Justification is about the restoration of human beings after Adam's Fall, by the life, death and resurrection of Jesus, and the beginning of a new life that anticipates the glory of heaven. According to the Roman Catholic Church, justification has two aspects: forgiveness of sin and the infusion of grace that makes Christians just (innocent). It is the beginning of a new life of grace, in which the gifts of faith, hope and charity enable one to perform meritorious works. However, the restoration is never complete in this life and concupiscence remains; a fall from the state of grace is thus possible, but this is reversible through penance.

A central feature of the Protestant Reformation was a dispute with the Roman Catholic Church over how justification should be understood. According to Luther, one does not become renewed (innocent) in justification. Rather, one is forgiven because the righteousness or justice of Christ is imputed to those who have faith in God's promise of redemption; however, one remains a sinner. More recent thought, however, has pointed to the fact that in both Lutheran and Catholic conceptions of justification, there is a sense of incompleteness, that it is just part of the process of redemption. There has also been interest in the idea of justification involving an indwelling of God rather than a gift from God to the individual; this has interesting affinities with Eastern Orthodox beliefs.

1 The Bible

In the Old Testament the word 'justification' has a juridical connotation. It refers to the obtaining of justice for one who is unjustly accused, whereby they are vindicated or shown to be not guilty. Quite often the context is one in which the innocence of a person before God is made plain by God through a public tribunal.

One exceptional but important use is in Isaiah: 'The righteous one, my servant, shall make many righteous, and he shall bear their iniquities' (53: 11). A servant by his suffering death obtains pardon and remission of sin for others. His suffering justifies, makes the guilty righteous. If this passage is interpreted in terms of the juridical understanding of justification, it must mean that the guilty are actually made innocent. In the New Testament, Jesus is identified as the suffering servant, whose life, death and resurrection justify the sinner.

Jesus used the term 'just' in a very wide sense to mean a life of innocence or holiness before God. For him, the righteousness of the kingdom of God was very different from that of the Pharisees. In one parable, a Pharisee thanks God for his freedom from sin, while a tax collector prays for God to have mercy for his sinfulness; 'I tell you, this man went home justified rather than the other; for all who exalt themselves will be humbled, but all who humble themselves will be exalted' (Luke 18: 9-14). In the parable of the prodigal son (Luke 15: 11-32), Jesus made it clear that God seeks a deep bond of love, not just external conformity, as the basis of his relationship with both the prodigal and the conventional son. Covenant, a sacred relation established by God, precedes law, and is the foundation of life with God, rather than conformity to law apart from the covenantal relation.

According to Paul, who explicated Jesus' teaching on justification more systematically than any other biblical writer, the insistence that Gentiles (non-Jewish converts) should observe the Jewish law threatened to destroy the new covenant established by Jesus. The controversy forced Paul to explore more deeply the significance of Jesus' death and resurrection: 'We ourselves are Jews by birth and not Gentile sinners; yet we know that a person is justified not by the works of the law but through faith in Jesus Christ' (Galatians 2: 15). Life in God is the acceptance of the gratuitous gift of justification through faith in Christ. The justified believer lives a new life because they are united with the risen Lord.

2 Augustine

Augustine's teaching on justification had far-reaching influence in the Latin (Western) Church (see Augustine §6, 13). It was occasioned by the teachings of Pelagius, who denied that one needed divine grace to do good and right (see Pelagianism). This denial of original sin, and hence of the need of redemption through Christ, aroused Augustine to reaffirm that all are born in sin and cannot become just except through the grace of Christ. God saves those who would justly be condemned. All the initiative in justification comes from God, so that free cooperation with God in justification (manifest in faith, hope and charity) is possible only by God's grace. Later, Augustine rejected semi-Pelagianism, in which it was taught that God's justification awaits a person's good action. Augustine stressed that a person's very assent to the message of salvation is God's gift. Grace is necessary to restore free will

(see Grace).

Augustine set the terms for the discussion of justification in the Western Church. The way the medieval Church developed Augustine's views on the remission of sin and good works led Luther to reject its teaching on justification. Augustine's stress on the total gratuity of justification led to controversy among Calvinists and to a deep division between Calvinists and Methodists over the role of free will; it also gave rise to the Jansenist controversy within Roman Catholicism in the seventeenth century.

3 Medieval theology

Medieval theologians agreed with Augustine that there are two aspects of justification: forgiveness of sin and an infusion of grace that overcomes one of the effects of the Fall, so that one becomes in fact just or innocent. The gift of integrity, whereby appetites are fully controlled by reason, is not given. Concupiscence therefore remains, but, as Augustine taught, it is not itself sin and there is room for growth in the gifts of faith, hope and charity; one can move towards integrity. Justification can be lost by mortal sin, but can be regained through the sacrament of penance.

Divisions occurred over the relation between the two aspects of justification, forgiveness of sin and the infusion of grace. They were prompted, in part, by the scholastic desire for precision, and this discussion of justification has the flavour of academic speculation, rather than of practical and pastoral concerns as in Augustine. The application of Aristotelian philosophy, particularly the notion of change in terms of formal and material causality, to theological issues shaped the way the question of the two aspects of justification was treated.

There were two primary schools of thought regarding the changeover from the state of sin to the state of grace. According to Aquinas, there is a necessary connection between the forgiveness of sin and the infusion of grace, or between forgiveness and restoration to a condition of justice or innocence. This stems from Aristotle's notion of one form replacing another form: the change involves the form of grace expelling the form of sin. Both aspects of justification occur together because they are necessarily connected to the reception of the form of grace.

The other school of thought rejected this Thomist intellectualism in favour of the primacy of divine will. Duns Scotus said that forgiveness of sin and infusion of grace are connected in fact because of God's will, but not by their nature. William of Ockham went even further by saying that not only is God's decree or will the only reason for the actual connection of forgiveness of sin and the condition of grace, but that there would be no logical contradiction if they were not connected, so that God could decree one without the other (see William of Ockham §10; Voluntarism).

It is generally held by Roman Catholic theologians that Duns Scotus' voluntarism and Ockham's nominalism, by rejecting any intrinsic connection between remission of sin and grace, enabled Martin Luther to claim that in justification people remain sinners. That is, when forgiven by God, people remain sinners, but God declares that they are just. However much Luther may have been influenced in his theology by voluntarism and nominalism, he appealed to Scripture for his views on justification. This is also true of later reformers, such as John Calvin (§4).

The sixteenth-century Council of Trent by and large endorsed the Thomistic view of justification. Justification implies an actual removal of sin, not a pardon of sin only (a mere forensic declaration by God that one is forgiven, but with sin still remaining). There is an interior renewal by the infusion of grace, whether or not it is experienced psychologically.

4 Luther

Luther's view that in justification one is *simul iustus et peccator* (both just and sinner) does indeed mean that for him justification is forensic (not an actual regeneration). Because of Christ, God does not impute one's sins to one and so remits punishment. One is not internally renewed or made pure, but is saved by faith in God's promise in Christ to bring his followers into his eternal kingdom. Since one is saved unconditionally, there is no role for good works in the attainment of salvation. Good works flow from one's faith.

Luther's views on justification can easily be misconstrued by both Roman Catholics and Protestants if his approach is not kept firmly in mind. Luther believed that he was following Paul in stressing that one is not justified *according* to the law, but *apart* from the law. According to the law, all are sinners. But by imputing Christ's

righteousness to sinners, God forgives them and declares them to be just. In relation to God, but not according to the law, they are just. *Simul iustus et peccator* is to be understood in the context of a justification *apart* from the law. If one retains the framework of justification according to the law, then one can understand the reasoning that insists that one must have undergone a total inner transformation, for only so can one actually be just according to the law. But for Luther this means retaining the Old Testament principle noted in §1 above, namely that in justification the person who is justified is found to be innocent of sin. For Luther, to claim that in justification one becomes inherently innocent of sin by the infusion of grace is to remain within the context of justification according to the law. None the less, he insists that the law is not abolished and that one is not to become lawless; rather, one is justified apart from the law.

Luther's approach does not settle the question of whether the two aspects of justification, forgiveness of sin and total renewal, can be separated, as he and Ockham (for different reasons) claimed. But to understand Luther's approach enables one to see more accurately the differences between the Roman Catholic Church and Luther. According to the Roman Catholic Church, one is made pure (innocent) by infused grace, and one is just or justified because one is innocent. According to Luther, one is saved or justified by faith in God's forgiveness in Christ, not because one is innocent. He claimed that the need to insist that one is actually innocent is because justification is conceived within the framework of the law, even though the alleged innocence is achieved by infused grace. For both Luther and the Roman Catholic Church, good works are only possible because of divine grace, but for the Church, good works are needed because one must actually be good to deserve justification. For Luther, good works flow from our justification.

Protestants have frequently failed to grasp properly Luther's approach. The stress that justification is by faith alone, a completely gratuitous gift, led Protestants to wrestle in various ways with the question raised by Paul in Romans 6: 1. After Paul has explained that one is justified apart from the law, he asks 'Should we continue in sin in order that grace may abound?' If God has done everything for one's salvation, what is there for one to do? Why should one obey the law? Protestants have oscillated between laxity, on the ground that justification is apart from good deeds, and rigorism, on the ground that since good works flow from one's faith, unless one performs them, one cannot have been saved.

The tendency to laxity stems from a failure to see that, for Luther, justification is the *beginning* of renewal or inner transformation by the Holy Spirit. (Seeberg (1961) and Forde (1984) insist that, for Luther, justification is the beginning of holiness.) To infer from justification apart from the law that one is to be lawless is to repudiate justification itself, since justification is the beginning of the reception of God's promise in Christ eventually to establish his kingdom or rule in his followers. Rigorism, in its reaction to lawlessness, lapses into the framework of justification according to the law.

5 Possible reconciliation

Recent post-Tridentine and ecumenical thought has softened the opposition between Roman Catholicism and the Protestants who, in various denominations, follow Luther's teaching on justification. In the Roman Catholic view, justification does not deliver one wholly from the effects of the Fall because concupiscence, which inclines one to sin, remains. It is also clear that Luther's claim that in justification one is *simul iustus et peccator* includes the notion that justification is the beginning of one's renewal (holiness). Since in both views of justification there is newness and incompleteness, the difference is not absolute.

The Roman Catholic Church is now raising into prominence the teaching of the indwelling or the presence of God in one's salvation. The earlier stress on infused grace tended to overlook the fact that such grace is relative to or linked to the divine presence (uncreated grace). The stress on infused grace enabled one to focus on one's condition in isolation, so that even though one's state of grace and good works are possible only because of infused grace, one's state and deeds are one's own and are meritorious. The indwelling of God mitigates such an overly external relation between God and those whom God sanctifies. One's state and one's good works are the result of the continuing presence of God, so that one can never attend to oneself apart from God.

Among the Fathers of the Eastern Orthodox Church, for whom the indwelling of God and one's elevation into the life of God are central, there are no controversies over justification. Some Roman Catholic theologians (such as Garrigou-Lagrange 1957 and Congar 1962) are aware that their recent emphasis on the central concerns of the

Orthodox Fathers has important ecumenical promise, not only for their relation to Orthodoxy, but also for their relation to those Protestants who are increasingly aware that Luther's view of justification may be harmonious with that of the Fathers. In time, there may be a growing convergence of opinion, not only on justification, but also on the nature of the Christian life.

See also: Atonement; Faith; Forgiveness and mercy; Salvation; Sanctification

DIOGENES ALLEN

References and further reading

Aquinas, T. (1266-73) *Summa theologiae*, New York: McGraw-Hill, 1964. (Aquinas' views on the necessary connection of forgiveness and regeneration; section 4 of book III is particularly useful.)

Augustine (412-15; 428-9) *The Nicene and Post-Nicene Fathers*, vol. 5, *Saint Augustine: Anti-Pelagian Writings*, ed. P. Schaff, Grand Rapids, MI: Eerdmans, 1956. (A collection of Augustine's influential views on justification; the earlier writings are against Pelagianism, the later ones against semi-Pelagianism.)

Cassirer, H. (1988) *Grace and Law*, Grand Rapids, MI: Eerdmans.(A study of Paul, Kant and the Hebrew prophets by a Kantian scholar.)

Congar, Y.M.J. (1962) *The Mystery of the Temple*, London: Burns & Oates.(Treats of divine indwelling.)

Forde, G.O. (1984) 'The Christian Life', in C.E. Braaten and R.W. Jensen (eds) *Christian Dogmatics*, vol. 2, Philadelphia, PA: Fortress.(A thorough discussion of Luther and Lutheranism on justification and sanctification.)

Froude, J.A. (1896) *Lectures on the Council of Trent*, New York: Scribners.(A classic, sympathetic Protestant account.)

Garrigou-Lagrange, R. (1957) *Grace*, St Louis, MO: B. Herder Book Co.(Treats of how we are partakers of the divine nature.)

Küng, H. (1964) *Justification*, New York: Thomas Nelson & Sons.(A careful comparison of Karl Barth's work and Roman Catholic teaching on justification; includes an ample bibliography.)

Luther, M. (1515-16) *Commentary on the Epistle to the Romans*, ed. J. Theodore Mueller, Grand Rapids, MI: Zondervan, 1954.(A major account of Luther's development of what became the Protestant view of justification.)

McGrath, A. (1986) *Iustitia Dei*, New York: Cambridge University Press, 2 vols.(A history of work on the doctrine of justification and sanctification.)

Seeberg, R. (1961) *Textbook of the History of Doctrines*, trans. C.E. Hay, Grand Rapids, MI: Baker Book House. (A brief, valuable account of the Augsburg Confession of 1530, the main Lutheran confession.)

Stump, E. (1989) 'Justification and Atonement', in R. Feenstra and C. Plantinga (eds) *Trinity, Incarnation, and Atonement*, Notre Dame, IN: University of Notre Dame Press.(An examination of the connection between justification by faith and atonement.)

Justinian (AD 482-565)

There was a late Roman renaissance during Justinian's reign in the sixth century AD. Its high point was the compilation by his minister Tribonian of a huge restatement of Roman law in four works, the Institutes, Digest, Code *and* Novels*, preserving a selection of its achievements. Called by medieval lawyers the* Corpus iuris civilis*, it is the basic material for studying Roman law and the source of much of Europe's legal thinking.*

Born Petrus Sabbatius in Macedonia, Justinian (Iustinianius Flavius) became Emperor at Constantinople in 527. Eager for fame, he wanted to restore and renew the Roman world. His generals reconquered the western empire while his architects adorned the capital with buildings such as St Sophia. Roman law, too, represented the glorious past: his minister Tribonian was put in charge of an elaborate programme for its restatement.

This began in 529 with the first *Code*, a collection of earlier imperial enactments, superseded in 534 by the second and only existing version. Subsequent legislation by Justinian and others is preserved in a sixth-century compilation called the *Novels*. The *Digest*, begun in 530 and completed, according to Bluhme's theory (1820), by three committees working furiously in parallel in only three years, was Justinian's largest work, about one and a half times the size of the Bible. An anthology of excerpts from the great jurists of the classical age (*c*. 50 BC to AD 250) such as Papinian, Paul, Ulpian and Gaius, it contains virtually all that has survived of them (see Gaius). However, its compilers could and did change the original texts: scholars have been pointing out these 'interpolations' for centuries.

Though providing a ready-made law library in one book, the *Digest* was a vast and shapeless collection, in which the detailed rules on any topic were hard to find and fit into an overall pattern. Justinian's *Institutes*, a short introductory textbook published in 533, was an attempt to solve this problem for first-year law students at Constantinople. It was largely a second edition of the classical *Institutes* of Gaius, from which it took its so-called institutional order and much of its contents, with later reforms tacked on to each section.

Justinian intended his restatement to be final, with further commentary banned. But medieval and later scholars in Europe called it the *Corpus iuris civilis*, the basic source-material for an enormous body of secondary literature on Roman law (see Legal hermeneutics §1). These studies were the starting point for much of Europe's legal thinking. Through them most European countries experienced a 'reception' of Roman law, becoming what are called civil law systems. Typically the *Digest*, *Code* and *Novels* supplied them with many specific rules and institutions, while the *Institutes* provided a clear scheme for systematizing and later codifying their law. Whatever its imperfections, Justinian's compilation gave the world Roman law (see Roman law).

See also: Law, philosophy of

GRANT McLEOD

List of works

Justinian (533) *The Digest of Justinian*, T. Mommsen, P. Krueger and A. Watson (eds), Philadelphia, PA: University of Pennsylvania Press, 1985.(The Latin text by Mommsen and Krueger with a readable translation by various writers edited by A. Watson.)

Justinian (533) *Justinian's Institutes*, P. Birks and G. McLeod, London: Duckworth, 1987.(The Latin text by Krueger with a good modern translation and useful introduction.)

Justinian (534) *Codex Iustinianus*, in *Corpus iuris civilis*, vol. 2, ed. P. Krueger, 13th stereotype edn, Berlin, 1954.(The Latin and Greek text of Justinian's *Code*; there is no English translation.)

Justinian (after 578) *Iustiniani Novellae*, in *Corpus Iuris Civilis*, vol. 3, eds R. Schoell and G. Kroll, 6th edn, Berlin, 1954.(The predominantly Greek text of the *Novels* of Justinian and his successor; there is no English translation.)

References and further reading

Honoré, T. (1978) *Tribonian*, London: Duckworth.(A brilliant though sometimes technical account of the whole of Justinian's work.)

Stein, P.G. and Lewis, A.D.E. (eds) (1983) *Studies in Justinian's Institutes in Memory of J.A.C. Thomas*, London:

Sweet and Maxwell.(A collection of detailed studies on aspects of the *Institutes*.)

Kabbalah

Kabbalah is the body of Jewish mystical writings which became important at the end of the twelfth century in Provence and has been taken up with varying degrees of enthusiasm in an attempt to explore the esoteric side of Judaism. There are two main forms of Kabbalah: one which concentrates on gaining knowledge of God through study of his name, and a theosophical tradition that approaches God through his impact on creation. On both accounts God is linked to the world through ten Sefirot, *hypostatic numbers which mediate between the Infinite and this world and thus (among other functions) help to explain how a being who is entirely ineffable can produce so much variety as is observed in nature. God's willingness to relate to the world gives his creatures the possibility of personal knowledge of him, although this can be acquired only through difficult and strenuous spiritual exercises. The variety of works which the Kabbalists produced are a blend of philosophical and mystical ideas which attempt to explore the inner meaning of faith and represent a creative and influential stream that both draws upon and contributes to Jewish philosophy.*

1 The nature of Jewish mysticism

There are broadly two types of mysticism within the Jewish tradition. One identifies mysticism with esoteric knowledge and links mysticism with philosophy. This provides the adept with access to a level of understanding unavailable to ordinary believers. The mystical knowledge is passed on from teacher to student, generally orally. In some cases, however, there is a gnosis that is achieved through some variety of divine revelation. Reliance on esoteric teachings and privileged gnosis is not taken to show that the traditional Jewish texts such as the Torah are not themselves sources of deep insight into the nature of reality. But the Kabbalistic tendency is to assume that such texts need to be interpreted in the right way if their inner essence is to be extracted. For this the reader needs a key, which Kabbalah (literally, tradition) can provide, through the appropriate training, to those capable of grasping it.

The other type of mysticism places the emphasis on experience, in particular ecstatic experience. The route to this experience is frequently allied to magic. It may involve the recitation of the divine name, specific rituals or particular types of meditation.

Study of the Kabbalah has often been controversial. Some who saw the Torah as consisting of secrets felt that once they understood these secrets they need no longer obey the laws. And there were a variety of heresies that challenged orthodoxy by the interpretations that they put on mystical experience.Yet within certain Jewish communities of great piety the Kabbalah became widely popular, especially because of its claim to elucidate (and facilitate) the linkage between this world and the divine. The ideas of the Kabbalah were also taken up by a variety of Christian thinkers who were interested in mysticism. These included Ramon Lull, Pico della Mirandola, Agrippa Von Nettesheim and Henry More.

2 Merkavah and Hekhalot mysticism

Merkavah (chariot) and Hekhalot (palace) mysticism have a common form. They deal with accounts of the ascent of the individual through the celestial realms, and end with the attainment of an ecstatic vision of the throne in the seventh palace of the seventh heaven. The reference to the chariot comes originally from the Book of Ezekiel, where the prophet describes his vision of the glory of the enthroned Godhead. Descriptions of this experience frequently refer to the experiencing (or perhaps better, the experiencing again) of God at Sinai, one of the crucial revelatory episodes in Judaism. There are two linked features of this form of mysticism. There is the ascent which results in a mystical experience, and there is also the acquisition of the angelic powers, perhaps through the adept's clothing himself with the divine name.

The notion of experiencing God is problematic in a religion like Judaism, with its frequent emphasis on God's transcendence. Thus there arises an interesting tension in Jewish mysticism, with its aim of seeing a God who at the same time cannot be seen. One way of addressing this tension is through poetic discourse that characterizes the mystical experience in erotic terms. Thus, the throne is often represented as female and the glory on it as male, so that the successful mystic makes contact with the holy union that lies at the core of reality.

What is the process by which mystics attain their experience? They progress through the seven heavens of Jewish

cosmology and the six palaces of the last and seventh heaven, and then enter the seventh palace, where the last throne exists. They stand before the enthroned glory and repeat the appropriate words with the angels. This may give them access either to the throne of glory itself or to a neighbouring throne. There they have a vision of the shining form of God. Although they never become the same as God, they do acquire some of the divine characteristics - spiritual or angelic properties. Angels participate in both divinity and humanity and thus permit the mystic to approach God without becoming God. Once mystics acquire angelic properties (through purification and asceticism), they can understand the hidden nature of reality, which was not available to them in their mundane existence.

3 God as a physical being

One of the aspects of God which the mystic is, in principle, capable of grasping is his physicality. Usually this is not expressed in terms of the asciption to God himself of physical proportions. But such proportions are ascribed to a secondary power, sometimes called the Metatron. This device enabled mystics to provide a visionary interpretation of the Talmudic notion, for example, that God wears phylacteries (Berachot 6a). It also allowed them to weave a visual symbolism on verses such as ‘Let us make man in our image’ (Genesis 1: 26). Such verses could not apply unproblematically and literally to God himself, since God transcends physicality. But they could be applied to a lower, angelic being who shares some physical characteristics with creatures. There were many different calculations of the size of God and the Metatron, based on the numerical values of Hebrew letters in crucial biblical passages (see God, concepts of).

4 Esoteric cosmology

The outstanding Kabbalistic work that deals with cosmology is the *Sefer Yetzirah (The Book of Creation)*. The work has obviously experienced changes and accretions over the centuries, but it seems to have been constructed between the third and ninth centuries. Its main assertion is that the scope for divine creativity involves the thirty-two paths of wisdom which make up ten basic forces (Sefirot) and the twenty-two letters of the Hebrew language. The Sefirot are to be seen as the attributes of God. They are sometimes identified with the heavenly creatures who support the throne. The notion of Sefirot helps the mystic understand the concept of God as enthroned, surrounded by his power. The Sefirot have a neo-Pythagorean origin, so it is natural for them to be identified with numbers, but they are also the basic constituents of physical matter; and they can be equated with other elemental notions, such as east, west, south, north, height, depth, good, evil, the end and the beginning. They are thus the underlying principles of the temporal, spatial and moral order of the world. Accordingly, they are sometimes seen as the four directions, spirit, water, fire, height, depth and the divine spirit. A common representation of the Sefirot takes the form shown in the figure.

Keter (crown)
Binah (understanding) Chochmah (wisdom)
Din (judgment) Chesed (mercy)
Tiferet (beauty)
Hod (splendour) Netzach (victory)
Yesod (foundation)
Malchut (kingdom)

The first triad of the Sefirot represent thought, understanding and wisdom, the three intellectual aspects of the first emanation. The first nine are regarded as masculine, while the last, kingdom, is identified with the *shekhinah*, the divine presence, immanence, or indwelling, and is feminine. The divine powers present in the Sefirot are part of the continuous process of emanation, by which the world is created.

The Sefirot represent the expressivity of divine creation. Tzimtzum (contraction) is the divine self-contraction - for the differentiation brought about by God also implies not only emanation but also limitation. The Sefirot represent the capacity to create. But Tzimtzum is the precondition of that creation. For without self-limitation, God would not be able to enter into relationship with us. Sometimes this idea of divine self-limitation is expressed in terms of the divine light’s having to go into vessels in order to be weak enough to make contact with mortal creatures. Unless God were prepared to limit the extent of the light, it would not be suitable for the finitude of creation but would overwhelm all that was to be with the divine effulgence. Creation and revelation are parallel here. So the

specificity of the divine Law is another outcome made possible only by divine self-limitation. Furthermore, since God is infinite, when God creates he does not create something outside of himself. Rather, he makes room within his own self for a greater diversity. Isaac Luria (1534-72) argued that the vessels in which the light is restricted represent the commandments (*mitzvot*) of the Torah. God's restricting himself to the commandments does not bring him into the power of human beings, nor does it confine him under the categories which apply to our notion of how things are. His self-limitation is an internal action, and the appropriate response of human beings is to carry out his laws in order that we may become colleagues with God in an inner sphere of activity. As an ineffable and infinite being, God transcends multiplicity while at the same time linking himself to everything through the Sefirot. He freely chooses, through self-contraction to allow multiplicity to emerge. This gave the Kabbalists an interesting route to reconciling freedom with divine omniscience. God could know what was going to happen in the future, but he wills not to, in order to allow human beings to make their own decisions freely (see Omniscience).

One intriguing aspect of the *Sefer Yetzirah* is its discussion of letter symbolism as both the form of divine creativity and the content of reality. A thing is brought to life through the appropriate combination of letters, since the letters both represent the potentiality of things and the materiality of existence. Each letter has an effect on three levels: space, time and the microcosm. The human body, like everything else, is composed basically of letters.

5 German pietism

Although Jewish mystical thought has a long history, it received a fresh impulse from the thinkers in the Rhineland in the twelfth and thirteenth centuries. Judah ben Samuel ben Kalonymus of Regensburg (d.1217) and his pupil Eleazar ben Judah ben Kalonymus of Worms (d.1240) combined the Kabbalah with the contributions of philosophers such as Al-Fayyumi Saadiah Gaon and Abraham Ibn Ezra to concentrate on the nature of divine glory. This concept was felt to be problematic, in that prayer implies that God has characteristics which Judaism in other places denies, and yet for prayer to be possible (it often seemed) there should be some scope for visualization of the object of prayer. Rhineland mystics distinguished between a lower and a higher glory, the latter being the *shekhinah* (presence or indwelling) while the former is to be identified with the prophetic or mystical imagination. The lower glory varies in its images, depending on the will of God and the capacities of the individual worshipper. There is certainly likely to be a heavy use of anthropomorphic imagery, in particular the notion of enthronement, with the proviso that the one who sits on the throne should not be regarded as a body. This point is emphasized by the tendency to apply the measurements which previously had been applied to God to the form which is constituted within the imagination, the lesser glory. How is the mystical experience to be achieved? Of great importance here is contemplation of the divine name, which progresses via a series of rituals of purification, and the recitation of the divine and angelic names and combinations of letters in order to bring about a state of religious ecstasy similar to prophecy.

6 Kabbalah in Spain and Provence

Kabbalistic thought developed at much the same time in Provence and Spain, although the lines of transmission are not clear. The main text here is the *Sefer ha-Bahir (Book of Bahir)*, which interprets biblical and rabbinic language into images of divine activity. The Sefirot in this work are seen as either like a tree or a human figure. We again need to return to Genesis 1: 26 and the idea of humanity being created in God's image, and this produces the notion of the divine body as paralleling the human frame - often characterized as Adam Kadmon, the primordial man. Through obeying the religious law, the individual strengthens the divine limb, while disobeying the law weakens it. Some Kabbalists see the Torah itself as the shape of God, so that each of the laws corresponds to, or is derived from, a particular limb in the perfect form of man, the form of Adam. On the other hand, they frequently warned against taking this imagery too literally, insisting that language about divine limbs was really about the divine powers, the Sefirot. Such language is entirely symbolic, and we cannot know the nature of divine reality. On the other hand, it must be admitted that much Kabbalistic writing is intensely anthropomorphic, and some Kabbalists were not frightened by the prospect of the interpretation of their language as more than exercises of imagination.

7 Light and darkness

The Sefirot are frequently identified with light, and are seen as the ten spiritual lights which emanate from the Eyn Sof, the Infinite. Although the Sefirot are many, they all originate in the One and indeed constitute a unity, expressed in an image popular in a variety of mystical philosophies. Light comes from one thing, and enables many things to be noticed, yet it is essentially just one thing. Neoplatonic language dominates here, and the emanation of the Sefirot from the Infinite One resembles the rays of light from the sun (see Neoplatonism). Like such rays, their production does not in any way diminish the power or the unity of the source. This sort of language led to an interesting debate between Kabbalists as to whether the Sefirot represent the essence of the deity or rather are the instruments into which God pours his ineffable light.

Questions about the nature of the opposite of light, namely, darkness, naturally arose within the Kabbalah. The *Zohar: Mishnat ha-Zohar (The Book of Splendour)* is the work that has the most detailed account of the struggle between the holy and the unholy. Sometimes this is represented as a battle between the forces of light and the forces of darkness, and sometimes both good and evil are seen as existing along a continuum of reality. On the latter view it is probably a mistake to see evil as a force which is in principle eradicable from the world, since it is a formal feature of reality and the best that can be achieved is to contain it.

The notion of light served as more than just a metaphysical notion; it is also a guide to the view of the Kabbalists as to how to achieve enlightenment. The divine form can be visualized through study of the Torah, and the esoteric approach to the latter as containing secrets means that these need to be elucidated, lit up as it were, so as to reveal what they hide. It was not regarded as coincidental that the Hebrew for ‘secret’ (*raz*) and ‘light’ (*or*) have the same numerical value. The letters of the Torah are literally luminous to those who know how to read them. The techniques of the Kabbalists often involved imagining colours while praying or meditating upon particular passages of the Bible, and the role of light is clearly of the first importance here.

8 Abraham Abulafia and prophetic Kabbalah

Abraham Abulafia (*c*.1240-*c*.1291) produced a different account of Kabbalah, which sees as the purpose of mysticism union with the deity. Abulafia criticizes the Kabbalists who see the Sefirot as divine powers, since such a view seemed to limit our access to the divine. If the Sefirot are seen as Neoplatonic separate intellects then we can, through perfecting ourselves, make contact with them, and thus become one with the emanation of reality and truth that flows from God. The route to this end is, among other activities, contemplation of the divine name. This sort of approach is called prophetic because it uses the language of prophecy made popular by the writings of Moses Maimonides and heavily relied upon in Islamic philosophy. For the authors who relied on these Neoplatonic themes, the overflow from the ‘active intellect’ leads to prophecy, and the prophet as a result comes into contact with the active intellect, the lowest of the ten separate intellects below God. It is when the prophet transcends his materiality and fulfills his spirituality that he may be able to rise to the level of the active intellect. Unlike the Islamic philosophers and Maimonides, who set forth an intellectual path for those who would attain prophecy, Abulafia recommended concentration on the permutations of the names of God and the letters constituting those names. The prophet has succeeded when he is able to see both the letters of the divine name and the being of God simultaneously.

Prophetic Kabbalah exercised much influence from the end of the thirteenth century in Palestine. Kabbalists in both Safed and Jerusalem derive much of their doctrine from Abulafia. They range from Isaac of Acre to Solomon Alkabetz, Moses Cordovero, Elijah de Vidas and Chayyim Vital, and even extend up to the eighteenth century and the works of the Hasidim (see Hasidism).

9 Jewish mysticism and Sufism

There can be little doubt that Sufism had a powerful impact on Jewish mysticism and Kabbalah (see Mystical philosophy in Islam). In the Iberian peninsula there were often close intercultural exchanges between Jews and Muslims, and both Ibn Masarra (886-931) and his disciple Sahl al-Tustari placed great importance on the mystical significance of the Arabic alphabet. As we have seen, there was a long-standing interest among Jewish thinkers in the arithmetical properties of Hebrew letters, and while this certainly predated Islam itself in the Talmud, it could well have been encouraged when Jewish philosophers became involved with Islamic Neoplatonism (see Neoplatonism in Islamic philosophy). Bahya ibn Paqudah’s *Fara’id al-qulub (Duties of the Hearts)* contains an individualistic description of a spiritual journey, leading the soul through contemplation and love to union with the

divine light, very much following the progressive spiritual stages of the path as described in Sufi pietistic books (see Ibn Paquda, Bahya). Bahya replaces Qur'anic motifs with Jewish ones and does not advocate extreme asceticism, as do some of the Sufis, yet the Sufi influences on his work are quite clear.

Sufism became important as a part of Islamic religious life in Egypt, which country was a haven for many Jews fleeing the Almohad persecution in the West and the Crusaders in the East. There are many Sufi works in the Cairo Genizah, the room in an old synagogue which contained thousands of writings going back to the Middle Ages. They were not destroyed, since it is forbidden to destroy anything on which the divine name may be written. There are many manuscripts of Muslim Sufi writings either in Arabic characters or in Hebrew (obviously translated for Jewish readers), or Sufi-inspired works written by Jews. The latter often include works on traditional Jewish themes, describing biblical figures as masters of the Sufi way. The most famous Jewish mystic in Egypt was undoubtedly the son of Moses Maimonides, Abraham (1186-1237) who is quite explicit in his use of Sufi material and authors in his *Kifayat al-'Abidin (Complete Guide for Devotees)* (see Maimonides, A.). Abraham stressed the spiritual significance of the Torah and the mysteries it contains, claiming that the Sufis had managed to hold on to these mysteries which then needed to be rediscovered by the Jews, who had lost them during the tragedy of the exile. Abraham seems to have initiated ceremonial practices in synagogues which were closely linked with Muslim practices in the mosque. He also advocated solitary meditation in a dark and isolated place, and reliance on the guidance of a spiritual master. Abraham's son, Obadyah, even went so far as to write rather disparagingly of marriage and family responsibilities, which is to go very far from the mainstream Jewish view of social responsibilities and the importance of family life.

The Kabbalist communities in Palestine were also close in their practices to Sufism, which was no doubt present in the wider community of the area. They advocated meditation and breath control in quiet dark places in order to attain illumination of the soul. The system of techniques used during these retreats included ritual purity, silence, fasting, restrictions of sleep and food, deep trust in God and the constant repetition of the divine name as aspects of the path to ecstasy and personal fulfilment. Sufism may even have influenced the types of mysticism followed in the Hasidism of the eighteenth century, which, it must be remembered, first developed in what had been a Turkish province of Poland. Shabbetai Zvi (d.1675) came into close contact with the Sufis even before his conversion to Islam, and his followers in the Ottoman empire continued to follow a form of worship which incorporated many Sufi practices. The close contact between Islamic and Jewish culture, which existed over a long and especially rich period, undoubtedly played its part in defining the specific character of Jewish mysticism.

See also: Ficino, M.; Hasidism; Hermetism; Mystical philosophy in Islam; Mysticism, history of; Platonism, Renaissance

OLIVER LEAMAN

References and further reading

Dan, J. (1986a) *The Early Kabbalah*, New York: Paulist Press.(Standard introduction to the early Kabbalah.)

Dan, J. (1986b) *Jewish Mysticism and Jewish Ethics*, Seattle, WA: University of Washington Press.(An account of the link between mysticism and ethics.)

Fenton, P. (1995) 'Judaism and Sufism', in S. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 755-68.(Excellent account of the many links between Jewish and Islamic mysticism in the Middle Ages.)

Gikatilla, J. (*c*.1300) *Sha'arey Orah*, Mantua, 1561; trans. with introduction by A. Weinstein, *Gates of Light*, London: HarperCollins, 1994.(An important Kabbalistic work by a thirteenth-century thinker who tried to make Kabbalah more accessible).

Idel, M. (1988a) *Kabbalah: New Perspectives*, New Haven, CT: Yale University Press.(The best account of the Kabbalah as a whole.)

Idel, M. (1988b) *Studies in Ecstatic Kabbalah*, Albany, NY: State University of New York Press.(Standard text on the ecstatic school of Kabbalah.)

Idel, M. (1989) *Language, Torah, and Hermeneutics in Abraham Abulafia*, Albany, NY: State University of New York Press.(Excellent account of the significance of Abulafia as a Kabbalistic thinker.)

Schafer, P. (1992) *The Hidden and Manifest God: Some Major Themes in Early Jewish Mysticism*, Albany, NY: State University of New York Press.(An acute discussion of the controversies about the precise nature of Kabbalah.)

Scholem, G. (1961) *Major Trends in Jewish Mysticism*, New York: Schocken.(Very influential text in the modern history of Kabbalah scholarship.)
Scholem, G. (1969) *On the Kabbalah and its Symbolism*, trans. R. Manheim, New York: Schocken.(A definitive guide to the language of the Kabbalah.)
Scholem, G. (1973) *Sabbatai Sevi: The Mystical Messiah*, trans. R. Werblowski, Princeton, NJ: Princeton University Press.(A magisterial account of his links with Kabbalah and general influence on Jewish thought.)
Scholem, G. (1991) *On the Mystical Shape of the Godhead: Basic Concepts in the Kabbalah*, New York: Schocken.(Summary of the views of this century's most famous Kabbalah scholar.)
Sefer Yetzirah (3rd-6th century AD) trans. I. Friedman, *The Book of Creation*, New York: Schocken, 1977.(The outstanding Kabbalistic work that deals with cosmology. Its main assertion is that the scope for divine creativity involves the thirty-two paths of wisdom which make up ten basic forces (Sefirot) and the twenty-two letters of the Hebrew language.)
Wolfson, E. (1994) *Through a Speculum that Shines: Vision and Imagination in Medieval Jewish Mysticism*, Princeton, NJ: Princeton University Press.(Exploration of some of the main concepts used in medieval Kabbalah.)
Wolfson, E. (1996) 'Jewish Mysticism: a Philosophical Overview', in D. Frank and O. Leaman (eds) *History of Jewish Philosophy*, London: Routledge, 450-98.(Clear survey of the area.)
Zohar: Mishnat ha-Zohar (*c*.13th century) ed. I. Tishby, Jerusalem: Mosad Bialik, 1957; trans. and ed. G. Scholem, *Zohar, Book of Splendour: Basic Readings from the Kabbalah*, New York: Schocken.(One of the main Kabbalistic texts.)

Kaibara Ekken (1630-1714)

*Kaibara Ekken was a leading Japanese scholar in the school of neo-Confucianism established by the renowned twelfth century Chinese synthesizer, Zhu Xi. As a thinker and a scholar Ekken, embraced a wide variety of topics from highly specialized neo-Confucian philosophy to the need to popularize Confucian ethics and to assist the society through practical learning (*jitsugaku*).*

Born in Fukuoka in southern Japan, Ekken studied for seven years in Kyoto with the principal scholars of the period. He spent much of his life as an advisor and teacher to the Kuroda *daimyō* family in Fukuoka, but traveled frequently to Kyoto and the capital, Edo (Tokyo). Although Ekken did not establish his own school as did other neo-Confucian scholars, he became through his writings one of the most influential figures in early modern Japan. He lived during a period when Japan isolated itself from the rest of the world while developing a rich complex of Confucian, neo-Confucian, Shintōand Buddhist schools of thought. Neo-Confucianism in particular made a significant contribution to education during this period with the spread of schools and a curriculum largely based on Confucian and neo-Confucian texts and commentaries (see Confucian philosophy, Japanese; Neo-Confucian philosophy). It was in this context that Ekken's ideas had a particular impact in both their breadth and depth of concerns.

Ekken was especially concerned to spread neo-Confucian moral teachings and self-cultivation to society at large, and he encouraged the education of all classes. He felt that neo-Confucian ideas could contribute to harmony of both self and society. To accomplish this, he wrote numerous popular treatises on learning for women, for children and for the family. Many of these treatises were reprinted frequently throughout the Tokugawa period (1603-1868). He also wrote on such practical subjects as healthcare, botany, agriculture, topography and history. His work on healthcare is still consulted, and his botanical studies and classification earned him the title of 'the Aristotle of Japan'.

These broad interests in both education and practical learning were based on a carefully developed philosophical position. By the age of thirty he had rejected Buddhism as tending toward quietism, and Wang Yangming's neo-Confucianism as potentially antinomian (see Wang Yangming). He argued for the importance of Zhu Xi's neo-Confucianism (see Zhu Xi) but with some significant adaptations. Philosophically, Ekken embraced the unity of principle (*li*) and material force (*qi*) (see Li; Qi). His position, which was indebted to the Chinese neo-Confucian Luo Qinshun, is described as a monism of *qi*. In a significant treatise, the *Taigiroku (Record of Grave Doubts)*, Ekken demonstrated strong support for Zhu Xi while taking issue with Zhu Xi's apparent dualism of principle and material force.

Ekken's position came to advocate a dynamic naturalism that was both the basis of his studies of nature and medicine as well as of moral self-cultivation. His monism of *qi* in his philosophical writings was paralleled by a position of filiality in his popular treatises. Ekken suggested that just as reverence and respect are due to parents for their care and support, so also because humans are 'children of heaven and earth' they should encourage a filial response to heaven's mandate and care for nature's abundance.

See also: Confucian philosophy, Japanese; Education, philosophy of; Li; Neo-Confucian philosophy; Qi; Zhu Xi

MARY EVELYN TUCKER

List of works

Kaibara Ekken (1665-1714) *Ekken zenshu (Collected Works of Kaibara Ekken)*, Tokyo: Ekken zenshūkankōbu, 1910-11, 8 vols.(This is the definitive body of Ekken's collected works, which includes his philosophical essays, his educational treatises, his botanical classification work, his travelogues and his records of local geography.)

Kaibara Ekken (1665-1714) *Ekken jikkun (Ekken's Ten Moral Treatises)*, ed. Tsukamoto Tetsuzō, Tokyo: Yūhōdōshoten, 1927.(A selection of ten significant moral and educational treatises which represent Ekken's efforts to make Confucian ideas understood by ordinary Japanese.)

Kaibara Ekken (1708) *Yamato zokkun (Precepts for Daily Life in Japan)*, Tokyo: Kiyomizu Kakujirō, 1967; trans. M.E. Tucker in *Moral and Spiritual Cultivation in Japanese Neo-Confucianism: The Life and Thought of*

Kaibara Ekken (1630-1714), Albany, NY: State University of New York Press.(One of Ekken's more important educational treatises.)

Kaibara Ekken (1714) *Taigiroku (Record of Grave Doubts)*, in *Kaibara Ekken, Muro Kyūsō*, Nihon shisōtakei vol. 34, Tokyo: Iwanami shoten, 1970, 10-64; German trans. O. Graf in *Kaibara Ekken*, Leiden: Brill, 1942. (Supports Zhu Xi, but takes issue with the latter's apparent dualism of principle and material force.)

References and further reading

Graf, O. (1942) *Kaibara Ekken*, Leiden: Brill.(This lengthy work in German consists of a biography of Ekken, a discussion of his philosophy, and a translation of the *Taigiroku* into German.)

Inoue Tadashi (1963) *Kaibara Ekken*, Tokyo: Yoshikawa kōbunkan.(Extensive biographical account (in Japanese) of Ekken's life, emphasizing in particular his role as advisor to the Kuroda *daimyō*.)

Inoue Tadashi and Araki Kengo (eds) (1970) *Kaibara Ekken, Muro Kyōsō*, Nihon shisōtaikei vol. 34, Tokyo: Iwanami shoten.(In Japanese, this book provides a strong introductory essay on Ekken's thought and gives annotated selections from some of his major works.)

Irizawa Sōju (1954) *Kaibara Ekken*, Nihon kyōiku sentetsu soho, Tokyo: Bunkyōshoin.(Written in Japanese, this book emphasizes Ekken's contributions to the educational needs of his time, in particular through discussion of his educational treatises written in a more simplified Japanese.)

Ishikawa Matsutaro (ed.) (1968) *Kaibara Ekken, Muro Kyusōshū*, Sekai kyōiku hōten, Nihon kyōiku hen vol. 3, Tokyo: Tamagawa Daigaku shuppanbu.(Written in Japanese as part of a series on educational issues in Japanese history, this book discusses Ekken's educational philosophy and gives some selections from his moral treatises.)

Komoguchi Isao and Okada Takehiko (1985) *AndōSeian; Kaibara Ekken*, Tokyo: Meitoku shuppansha.(In Japanese, this work provides a general discussion of Ekken's philosophy, his moral teachings and his contributions to the practical learning of his day.)

Matsuda Michio (ed.) (1969) *Kaibara Ekken*, Nihon no meicho vol. 14, Tokyo: Chūōkoronsha.(Part of a series on major figures in Japanese history, this work provides a biographical introduction to Ekken's life and thought and gives selections in modern Japanese from some of his major works.)

Najita Tetsuo (1975) 'Intellectual Change in Early Eighteenth-Century Tokugawa Confucianism', *Journal of Asian Studies* 34 (4): 932-44.(A short essay in English which provides a penetrating analysis of Ekken's thought in relation to other trends in Tokugawa Confucianism.)

Okada Takehiko (1979) 'Practical Learning in the Chu Hsi School; Yamazaki Ansai and Kaibara Ekken', in W.T. de Bary and I. Bloom (eds) *Principle and Practicality*, New York: Columbia University Press, 231-305.(This articles provides an insightful overview of some of the key elements of Ekken's thought, especially his contributions to 'practical learning'.)

Okada Takehiko (1992) *Kaibara Ekken*, Fukuoka: Nishi nihon shinbunsha.(In Japanese, this is a helpful treatment of Ekken's life and thought by a leading Confucian scholar still living in the area of Ekken's birthplace.)

Tucker, M.E. (1989) *Moral and Spiritual Cultivation in Japanese Neo-Confucianism: The Life and Thought of Kaibara Ekken (1630-1714)*, Albany, NY: State University of New York Press.(The only book-length work in English on Ekken, this monograph provides an introduction to Ekken's life, thought and times as well as a translation and discussion of one of his major ethical treatises.)

Kant, Immanuel (1724-1804)

Immanuel Kant was the paradigmatic philosopher of the European Enlightenment. He eradicated the last traces of the medieval worldview from modern philosophy, joined the key ideas of earlier rationalism and empiricism into a powerful model of the subjective origins of the fundamental principles of both science and morality, and laid the ground for much in the philosophy of the nineteenth and twentieth centuries. Above all, Kant was the philosopher of human autonomy, the view that by the use of our own reason in its broadest sense human beings can discover and live up to the basic principles of knowledge and action without outside assistance, above all without divine support or intervention.

Kant laid the foundations of his theory of knowledge in his monumental Critique of Pure Reason (1781). He described the fundamental principle of morality in the Groundwork of the Metaphysics of Morals (1785) and the Critique of Practical Reason (1788), in the conclusion of which he famously wrote:

> *Two things fill the mind with ever new and increasing admiration and awe, the more often and steadily reflection is occupied with them:* the starry heaven above me and the moral law within me. *Neither of them need I seek and merely suspect as if shrouded in obscurity or rapture beyond my own horizon; I see them before me and connect them immediately with my existence.*
>
> *(5: 161-2; see List of works for method of citation)*

Kant tried to show that both the laws of nature and the laws of morality are grounded in human reason itself. By these two forms of law, however, he is often thought to have defined two incommensurable realms, nature and freedom, the realm of what is and that of what ought to be, the former of which must be limited to leave adequate room for the latter. Kant certainly did devote much space and effort to distinguishing between nature and freedom. But as he also says, in the Critique of Judgment (1790), it is equally important 'to throw a bridge from one territory to the other'. Ultimately, Kant held that both the laws of nature and the laws of free human conduct must be compatible because they are both products of human thought imposed by us on the data of our experience by the exercise of our own powers. This was clearly stated in his last book, The Conflict of the Faculties (1798):

> *Philosophy is not some sort of science of representations, concepts, and ideas, or a science of all sciences, or anything else of this sort; rather, it is a science of the human being, of its representing, thinking, and acting - it should present the human being in all of its components, as it is and ought to be, that is, in accordance with its natural determinations as well as its relationship of morality and freedom. Ancient philosophy adopted an entirely inappropriate standpoint towards the human being in the world, for it made it into a machine in it, which as such had to be entirely dependent on the world or on external things and circumstances; it thus made the human being into an all but merely passive part of the world. Now the critique of reason has appeared and determined the human being to a thoroughly* active *place in the world. The human being itself is the original creator of all its representations and concepts and ought to be the sole author of all its actions.*
>
> *(7: 69-70)*

Thus, Kant derived the fundamental principles of human thought and action from human sensibility, understanding, and reason, all as sources of our autonomy; he balanced the contributions of these principles against the ineliminable inputs of external sensation and internal inclination beyond our own control; and he strove both to demarcate these principles from each other and yet to integrate them into a single system with human autonomy as both its foundation and its ultimate value and goal. These were the tasks of Kant's three great critiques. In the Critique of Pure Reason, the essential forms of space, time and conceptual thought arise in the nature of human sensibility and understanding and ground the indispensable principles of human experience. He then argued that reason, in the narrow sense manifest in logical inference, plays a key role in systematizing human experience, but that it is a mistake to think that reason offers metaphysical insight into the existence and nature of the human soul, an independent world, and God. In the Critique of Practical Reason and Groundwork, however, he argued that reason as the source of the ideal of systematicity is the source of the fundamental law of morality and our consciousness of our own freedom, which is the source of all value, and that we can postulate the truth of the fundamental dogmas of Christianity, our own immortality and the existence of God, as practical presuppositions of our moral conduct but not as theoretical truths of metaphysics. In the Critique of Judgment,

Kant argued that the unanimity of taste and the systematic organization of both individual organisms and nature as a whole could be postulated, again not as metaphysical dogmas but rather as regulative ideals of our aesthetic and scientific pursuits; he then went on to argue that it is through these ideals that we can tie together the realms of nature and freedom, because aesthetic experience offers us a palpable image of our moral freedom, and a scientific conception of the world as a system of interrelated beings makes sense only as an image of the world as the sphere of our own moral efforts. In many of his last writings, from Religion within the Limits of Reason Alone (1793) to his final manuscripts, the Opus postumum, Kant refined and radicalized his view that our religious conceptions can be understood only as analogies for the nature of human reason itself.

The Enlightenment began by attempting to bring even God before the bench of human reason - at the turn of the eighteenth century, both Shaftesbury in Great Britain and Wolff in Germany rejected voluntarism, the theory that God makes eternal truths and moral laws by fiat, and argued instead that we ourselves must know what is right and wrong before we could even recognize supposedly divine commands as divine. Kant completed their argument, concluding that the human being 'creates the elements of knowledge of the world himself, a priori, from which he, as, at the same time, an inhabitant of the world, constructs a world-vision in the idea' (Opus postumum, 21: 31).

1 Life and works

Immanuel Kant was born on 22 April 1724 in Königsberg, the capital of East Prussia. He was the child of poor but devout followers of Pietism, a Lutheran revival movement stressing love and good works, simplicity of worship, and individual access to God. Kant's promise was recognized by the Pietist minister Franz Albert Schultz, and he received a free education at the Pietist gymnasium. At sixteen, Kant entered the University of Königsberg, where he studied mathematics, physics, philosophy, theology, and classical Latin literature. His leading teacher was Martin Knutzen (1713-51), who introduced him to both Wolffian philosophy and Newtonian physics, and who inspired some of Kant's own later views and philosophical independence by his advocacy of physical influx against the pre-established harmony of Leibniz and Wolff. Kant left university in 1746, just as the major works of the anti-Wolffian Pietist philosopher Christian August Crusius were appearing. Kant's upbringing would have made him receptive to Crusius, and thus he left university imbued with the Enlightenment aims of Wolffian philosophy but already familiar with technical criticisms of it, especially with Crusius's critique of Wolff's attempt to derive substantive conclusions from a single and merely formal first principle such as the logical principle of non-contradiction (see Wolff, C.).

On leaving university, Kant completed his first work, *Thoughts on the True Estimation of Living Forces* (1746, published 1749), an unsuccessful attempt to mediate between Cartesian and Leibnizian theories of physical forces. Kant then worked as a tutor, serving in households near Königsberg for the next eight years. When he returned to the university in 1755, however, he had several works ready for publication. The first of these was *Universal Natural History and Theory of the Heavens*, a much more successful scientific work than his first in which Kant argued for the nebular hypothesis, or origin of the solar system out of a nebular mass by purely mechanical means. The book was scarcely known during Kant's lifetime, however, so the French astronomer Pierre Laplace (1749-1827) developed his version of the nebular hypothesis (published 1796) independently, and the theory became known as the Kant-Laplace hypothesis only later. In 1755, Kant also published two Latin works, his MA thesis *A brief presentation of some thoughts concerning fire*, and his first philosophical work, *A new elucidation of the first principles of metaphysical cognition*, which earned him the right to offer lectures at the university as a *Privatdozent* paid directly by his students. The following year Kant published *The employment in natural philosophy of metaphysics combined with geometry, of which sample I contains the physical monadology*, which made him eligible for a salaried professorship, although he was not to receive one until 1770. In these years, Kant also published four essays on earthquakes and winds.

Kant began lecturing in the autumn of 1755, and to earn a living lectured more than twenty hours a week. His topics included logic, metaphysics, ethics, and physics, and he subsequently added physical geography, anthropology (Germany's first lectures so entitled), pedagogy, natural right and even the theory of fortifications. Except for one small essay on optimism (1759), he did not publish again until 1762, when another burst of publications began. He then published, all in German: *The False Subtlety of the Four Syllogistic Figures* (1762); *The Only Possible Argument in support of a Demonstration of the Existence of God* and *Attempt to Introduce the*

Concept of Negative Magnitudes into Philosophy (1763); *Observations on the Beautiful and Sublime* and *Inquiry concerning the Distinctness of the Principles of Natural Theology and Morality* (1764), the latter of which was his second-place entry in a competition won by Moses Mendelssohn; *Dreams of a Spirit-Seer, elucidated by Dreams of Metaphysics* (1766); and *Concerning the Ultimate Ground of the Differentiation of Directions in Space* (1768). These publications earned Kant widespread recognition in Germany. During this period, Kant was deeply struck by the work of Jean-Jacques Rousseau, especially by his *Social Contract* and the paean to freedom in *Émile* (both 1762). By this time Kant was also acquainted with the philosophy of David Hume, whose two *Enquiries* and other essays, but not *A Treatise of Human Nature*, were published in German as early as 1755.

Having unsuccessfully applied for several chairs at home while declining offers elsewhere, Kant was finally appointed Professor of Logic and Metaphysics in Königsberg in 1770. This event occasioned his inaugural dissertation, and last Latin work, *On the form and principles of the sensible and intelligible world*. Following correspondence about this work with Johann Heinrich Lambert, Johann Georg Sulzer, and Mendelssohn, however, Kant fell into another decade-long silence, broken only by a few progress reports to his recent student Marcus Herz and a few minor essays. Yet during this ‘silent decade’, Kant was preparing for his enormous body of subsequent works. Beginning in 1781, with the first edition of the *Critique of Pure Reason*, Kant unleashed a steady torrent of books. These include: *Prolegomena to Any Future Metaphysics that shall come forth as Scientific*, an attempted popularization of the first *Critique*, in 1783; two essays, ‘Idea for a Universal History from a Cosmopolitan Point of View’ and ‘What is Enlightenment?’ in 1784; *The Groundwork of the Metaphysics of Morals* and four other essays in 1785; *The Metaphysical Foundations of Natural Science*, essays on ‘The Conjectural Beginnings of Human History’ and ‘What Does it mean to Orient Oneself in Thinking?’ and two other pieces in 1786; a substantially revised second edition of the *Critique of Pure Reason* in 1787; in 1788, the *Critique of Practical Reason* and an essay on ‘The Use of Teleological Principles in Philosophy’; the *Critique of Judgment* as well as an important polemic ‘On a discovery according to which any new Critique of Pure Reason is rendered dispensable by an older one’ in 1790; the political essay ‘On the Common Saying: "That may be right in theory but does not work in practice"’ and the controversial *Religion within the Limits of Reason Alone* in 1793; *Towards Perpetual Peace* in 1795; the *Metaphysics of Morals*, comprising the ’Doctrine of Right’ and the ’Doctrine of Virtue’, in 1797, as well as the essay ‘On a putative Right to Lie from Love of Mankind’; and his last major works in 1798, a handbook on *Anthropology from a Pragmatic Point of View* and his defence of the intellectual freedom of the philosophical faculty from religious and legal censorship in the restrictive atmosphere of Prussia after Frederick the Great, *The Conflict of the Faculties*. (With Kant’s approval, some of his other lecture courses were also published, including *Logic* in 1800 and *Physical Geography* and *Pedagogy* in 1804.) Kant retired from lecturing in 1797, at the age of seventy-three, and devoted his remaining years to a work which was to be entitled ‘The Transition from the Metaphysical First Principles of Natural Science to Physics’, but which was far from complete when Kant ceased working on it in 1803. (Selections from his drafts were first published in 1882-4, and they were first fully published as *Opus postumum* in 1936-8). After a lifetime of hypochondria without any serious illness, Kant gradually lost his eyesight and strength and died 12 February 1804.

2 Kant’s work to 1770

In his first work, *Living Forces*, Kant tried to mediate a dispute about the measurement of forces between Descartes and Leibniz by employing a distinction between ’living’ or intrinsic forces and ’dead’ or impressed ones to argue that Leibniz’s measure was correct for the former and Descartes’s for the latter. This distinction could not be maintained in a uniform mechanics, and the young Kant remained ignorant of the mathematically correct solution, which had been published by D’Alembert in 1743. Nevertheless, the work already showed Kant’s lifelong preoccupation with the relation between scientific laws and metaphysical foundations. It also included the observation that the three-dimensionality of physical space is a product of actually existing forces, not the only geometry that is logically possible (10, 1: 24).

Kant’s works of 1755 reveal more of his originality and his enduring themes. *Universal Natural History*, deriving the present state of the planets from postulated initial conditions by reiterated applications of the laws of Newtonian mechanics, manifests not only Kant’s commitment to those laws, for which he was subsequently to seek philosophical foundations, but also his commitment to thoroughly naturalistic explanations in science, in which God can be the initial source of natural laws but never intervenes within the sequence of physical causes. *New Elucidation*, while not yet a methodological break from the rationalism of Leibniz, Wolff and Alexander

Gottlieb Baumgarten (1714-62) (whose textbooks on metaphysics, ethics and aesthetics Kant used for decades), breaks with them on several substantive issues. Kant begins by rejecting Wolff's supposition that the principle of non-contradiction is a single yet sufficient principle of truth, arguing instead that there must be separate first principles of positive and negative truths; following Crusius, Kant was always to remain suspicious of programmes to reduce all truth to a single principle. Kant then criticized previous proofs of the Principle of Sufficient Reason, although his own proof was also a failure. More importantly, he argued that the principle of sufficient reason does not entail the theory of pre-established harmony drawn from it by the Leibnizians: the need for a sufficient reason for any change in a substance proves the necessity rather than impossibility of real interaction among a plurality of substances. Transposed into an epistemological key, this argument was to become central in the first *Critique*. The work is also noteworthy for the first suggestion of Kant's critique of Descartes's ontological argument for the existence of God (see God, arguments for the existence of), and for a first treatment of the problem of free will as well. Here Kant defended against the indeterminism of Crusius the determinism of Leibniz (see Determinism and indeterminism), although he was later to criticize this as the 'freedom of a turnspit' (5: 97). Kant's later theory of free will, however, attempts to reconcile Crusius and Leibniz.

In the *Physical Monadology* (1756), Kant tries to reconcile the infinite divisibility of space in geometry with the need for simple, indivisible substances in metaphysics - the subsequent theme of the first *Critique*'s second Antinomy (see §8). Kant does not yet appeal to a metaphysical distinction between appearance and reality, but instead argues that because bodies in space are not ultimately composed of particles but of attractive and repulsive forces (1: 484), they may be physically indivisible even when space itself is still mathematically divisible.

Kant's works of the 1760s introduce some of the methodological as well as substantive assumptions of his mature philosophy. *The Only Possible Argument* details Kant's attack upon the ontological argument, the paradigmatic rationalistic argument because of its presupposition that an existence-statement can be derived from the analysis of a concept. Kant argues that 'existence is not a predicate or a determination of a thing' (2: 72), but rather the 'absolute positing of a thing' (2: 73); that is, the existence of its subject is *presupposed* by the assertion of any proposition, not inferred from the concepts employed in it. Kant also maintains that the other rationalist argument for theism, the argument from the contingency of the world to a necessary cause of it, as well as the empiricists' favourite, the argument from design, fail to prove the existence of a necessary being with all the attributes of God. However, Kant still holds that the existence of God can be proved as a condition of the *possibility* of any reality. Finally, Kant further develops his argument that scientific explanation cannot allow divine intervention in the sequence of events, and that God must only be seen as the original ground of the laws of nature.

Negative Magnitudes announces a fundamental methodological break from rationalism. Inspired by both Crusius and Hume, Kant argues that *real opposition* (as when two velocities in opposite directions or a pleasure and a pain cancel each other out) is fundamentally different from *logical contradiction* (as between a proposition and its negation); he then applies this to causation, arguing that the *real ground* of a state does not entail its existence logically, but is connected to it in an entirely different way. This precludes any proof of the principle of sufficient reason from merely logical considerations alone (2: 202).

The *Inquiry into the Distinctness of the Principles of Natural Theology and Ethics* continues Kant's attack upon rationalism. The question for this essay was whether metaphysics could use the same method as mathematics, which Kant firmly denied: mathematics, he argues, can prove its theorems by constructing its objects from their very definitions, but metaphysics can only use analysis to tease out the definitions of its objects from given concepts, and cannot construct the objects themselves (2: 276). The claim that the method of philosophy is analysis may sound like rationalism; however, Kant insists that in both metaphysics and ethics philosophy needs *material* as well as *formal* first principles, again precluding any purely logical derivation of philosophical theses. Kant does not yet have a clear account of material first principles - he is sympathetic to Crusius's account of indemonstrable cognitions and to the suggestion of the moral sense theorists Shaftesbury and Hutcheson that the first principles of ethics arise from *feeling*, but not satisfied with either. Without yet naming it, Kant also introduces his distinction between hypothetical and categorical imperatives (2: 298).

Still in 1764, however, the book *Observations on the Beautiful and Sublime* already announces Kant's departure from moral sense theory and introduces the most fundamental theme of Kant's ethics. Virtue cannot depend merely

on benevolent inclination, but only on general principles, which in turn express ‘a feeling that lives in every human breast and extends itself much further than over the particular grounds of compassion and complaisance…the *feeling of the beauty and dignity of human nature*’ (2: 217). In notes in his own copy of this work, Kant went even further, and first clearly stated his enduring belief that ‘freedom properly understood (moral, not metaphysical) is the supreme principle of all virtue as well as of all happiness’ (20: 31).

In *Dreams of a Spirit-Seer*, Kant ridicules traditional metaphysics by comparing it to the fantasies of the Swedish theosophist Emmanuel Swedenborg; Kant argues instead that metaphysical concepts cannot be used without empirical verification, and that therefore metaphysics can at most be ‘a science of the *boundaries of human reason*’ (2: 368). The work also contains further thoughts on morality, suggesting that the two forces of egoism and altruism define the structure of the moral world in much the way that the forces of repulsion and attraction define that of the physical world (2: 334). But Kant does not yet argue that postulates of practical reason may be a valid alternative to the delusions of metaphysics.

Finally, the brief essay on *Directions in Space* argues that incongruent counterparts, such as right- and left-handed gloves, which have identical descriptions but cannot occupy the same space, prove that the qualities of objects are not determined by concepts alone but also by their relation to absolute space. Kant did not yet raise metaphysical questions about the nature of absolute space or epistemological questions about how we could know it, but this essay can be seen as introducing the distinction between intuitions and concepts which was to be a cornerstone of Kant’s subsequent thought (see §5).

3 The *Inaugural Dissertation* of 1770 and the problem of metaphysics

Kant’s *Inaugural Dissertation* of 1770 consolidated many of the gains he had made during the 1760s and introduced a fundamentally new theory about the metaphysics and epistemology of space and time which was to remain a constant in his subsequent thought, but also left open crucial questions about the source of our most fundamental concepts. Although Kant hoped to proceed quickly to his projects in the philosophy of science and in moral and political theory, it was to take him all of the next decade to answer these preliminary questions.

Taking up where *Directions in Space* left off, Kant begins the dissertation with the distinction between intuitions (singular and immediate representations of objects) and concepts (general and abstract representations of them) as distinct but equally important elements in the ‘*two-fold genesis* of the concept [of a world] out of the nature of the mind’. The intellect (Kant does not yet divide this into understanding and reason) provides abstract concepts, under which instances are subordinated; the ‘sensitive faculty of cognition’ provides ‘distinct intuition[s]’ which represent concepts ‘in the concrete’ and within which different parts may be coordinated (2: 387). Kant goes on to claim that ‘whatever in cognition is sensitive is dependent upon the special character of the subject’, that is, the knower, so that sensation, through intuitions, represents things ‘*as they appear*’ (*phenomena*), while the intellect, through concepts, represents things ‘*as they are*’ (*noumena*) (2: 392). Kant then presents the ‘principles of the form of the sensible world’: time and space are the forms of the intuition of all objects (time is the form for all representation of objects, inner or outer, while space is the form for the representation of all outer objects) which do not arise from but are presupposed by all particular perceptions; they are singular rather than general, that is, particular times or spaces are parts of a single whole rather than instances of a general kind; and they must each be ‘the subjective condition which is necessary, in virtue of the nature of the human mind, for the co-ordinating of all things in accordance with a fixed law’, or a ‘*pure intuition*’ rather than ‘*something objective and real*’ (2: 398-400, 402-4). Only thus can we explain our knowledge of both these general claims about space and time as well as particular claims about their structure, such as the theorems of geometry (2: 404). In other words, we can explain the certainty of knowledge about space and time only by supposing that it is knowledge of the structure of our own minds, and thus of how objects appear to us, rather than knowledge about how things are in themselves. This necessarily subjective origin and significance of certainty, which Kant was later to name ‘transcendental idealism’, is the foundation for the active role of the human mind in knowledge of the world.

Kant has little to say about the source of intellectual concepts, but continues to believe that they give us knowledge of how things are independently of the structure of our own minds. His main claim, still Leibnizian, is that in order to conceive of things as genuinely distinct substances, yet as collectively interacting in a single world, we must conceive of them as contingent beings all depending upon a single necessary being (2: 407-8). Kant then argues that metaphysical error arises when the principles of sensitive and intellectual cognition are confused, but more

particularly when 'the principles which are native to sensitive cognition transgress their limits, and affect what belongs to the intellect' (2: 411) - the opposite of what he will argue later when he claims that metaphysical illusion arises from thinking that human reason can reach beyond the limits of the senses (see §8). Finally, Kant introduces as mere 'principles of convenience' the principles of universal causation and of the conservation of substance as well as a more general 'canon' of rationality, that '*principles are not to be multiplied beyond what is absolutely necessary*' (30, 2: 418). A better account of these principles will occupy much of Kant's later work (see §7).

Early readers of Kant's dissertation objected to the merely subjective significance of space and especially time, but Kant was never to surrender this theory. What came to bother him instead was his inadequate treatment of metaphysical concepts such as 'possibility, existence, necessity, substance, cause, etc.' (2: 395). In a famous letter of 21 February 1772 to Marcus Herz (10: 129-35), Kant claimed that the 'whole secret' of metaphysics is to explain how intellectual concepts which neither literally produce their objects (as God's concepts might) nor are merely produced by them (as empirical concepts are) nevertheless necessarily apply to them. But Kant did not yet know how to answer this question.

His first progress on this issue is found in fragments from 1774-5 (Reflections 4674-84, 17: 643-73). Two key ideas are found here. First, Kant finally formulates the problem of metaphysics as that of 'synthetic' rather than 'analytic' propositions: how can we know the truth of propositions in which the predicates clearly go beyond anything contained in their subject-concepts but yet enjoy the same universality and necessity as propositions which are mere tautologies, whose predicates are contained in their subject concepts (17: 643-4, 653-5)? Second, Kant here first states that the answer to this question lies in recognizing that certain fundamental concepts, not just the intuitions of space and time, are 'conditions of the concrete representation [of objects] in the subject' (17: 644) or of the unity of '*experience* in general' (17: 658). Kant's idea is that in order to ground any determinate ordering of either subjective or objective states in temporal succession, we must use the concepts of substance, causation, and interaction, and that these must therefore be categories which originate in the understanding just as the pure forms of space and time originate in the sensibility.

4 The project of the *Critique of Pure Reason*

In spite of this progress in 1775, six more years passed before the *Critique of Pure Reason* finally appeared in 1781. In an umistakeable reference to Locke's *Essay concerning Human Understanding* (see Locke, J.), Kant began the work with the promise to submit reason to a critique in order to obtain a 'decision about the possibility or impossibility of metaphysics in general and the determination of its sources, its scope and its boundaries' (A xii). The 'chief question' would be 'what and how much can understanding and reason know apart from all *experience*?' (A xvii). Answering this question would require discovering the fundamental principles that human understanding contributes to human experience and exposing the metaphysical illusions that arise when human reason tries to extend those principles beyond the limits of human experience.

But Kant's project was even more ambitious than that, as he was to make clear in the revised edition of the *Critique* six years later. There, in addition to more explicitly describing his strategy for explaining the certainty of the first principles of human knowledge as one of supposing that 'objects must conform to our knowledge' rather than vice versa (B xvi), Kant described his whole project in broader terms: 'I therefore had to offset *knowledge* in order to make room for *belief*' (B xxx). Kant did not mean to return to the sceptical fideism of earlier thinkers such as Pierre Bayle, who simply substituted religious belief for theoretical ignorance. Instead, Kant argues first that the human mind supplies necessary principles of sensibility and understanding, or perception and conception; next, that if human reason tries to extend the fundamental concepts and principles of thought beyond the limits of perception for purposes of theoretical knowledge, it yields only illusion; but finally that there is another use of reason, a practical use in which it constructs universal laws and ideals of human conduct and postulates the fulfilment of the conditions necessary to make such conduct rational, including the freedom of the will, the existence of God, and the immortality of the soul. This use of reason does not challenge the limits of theoretical reason but is legitimate and necessary in its own right.

In the Introduction, Kant defines his first task as that of explaining the possibility of synthetic a priori judgments. This notion is grounded in two distinctions. First, there is a logical distinction between analytic and synthetic propositions: in analytic propositions, the predicate-concept is implicitly or explicitly contained in the

subject-concept (for example, ‘A bachelor is unmarried’ or ‘An unmarried male is male’), so the proposition conveys no new information and is true by identity alone; in synthetic propositions, the content of the predicate is clearly not contained in the subject-concept (for example, ‘Bachelors are unhappy’) (A 6-7/B 10-11), so the proposition conveys new information and cannot be true by identity alone. Second, there is an epistemological distinction between propositions which are a posteriori, or can be known to be true only on the basis of antecedent experience and observation, and those which are a priori, or known to be true independently of experience, or at least any particular experience (A 1-2/B 1-3). Kant maintains that anything which is known to be universally and necessarily true must be known a priori, because, following Hume, he assumes that experience only tells us how what has actually been observed is, not how everything must be (A 1-2/B 3-4). Combining these two distinctions yields four possible kinds of judgments. Two of these obviously obtain: analytic a priori judgments, in which we know a proposition to be true by analysis of its subject-concept and without observation; and synthetic a posteriori judgments, in which we know factual statements going beyond subject-concepts to be true through observation. Equally clearly, a third possibility is excluded: there are no analytic a posteriori judgments, for we need not go to experience to discover what we can know from analysis alone. What is controversial is whether there are synthetic a priori judgments, propositions that are universally and necessarily true, and thus must go beyond experience, but which cannot be reached by the mere analysis of concepts. Both rationalists and empiricists had denied such a possibility, but for Kant only it could ground an informative science of metaphysics at all.

Kant’s notion of synthetic a priori judgment raises various problems. Critics have long complained that Kant provides no unequivocal criterion for deciding when a predicate is contained in a subject, and twentieth-century philosophers such as W. Quine argued that there are no analytic truths because not even definitions can be held entirely immune from revision in the face of empirical facts. Lewis White Beck showed, however, that this did not affect Kant’s project, for Kant himself, in a polemic with the Wolffian Johann August Eberhard, argued that analysis always presupposes synthesis, and that the adoption of any definition itself has to be justified, either by construction or observation; so even conceding that all judgments are ultimately synthetic, Kant’s question remains whether any of these are synthetic a priori.

Another issue is just what synthetic a priori judgments Kant intended to justify. In the ’Prolegomena’ and the ’Introduction’ to the second edition of the *Critique*, Kant suggests that it is obvious that synthetic a priori judgments exist in what he calls ’pure mathematics’ and ’pure physics’, and that his project is to show that what explains these also explains other such propositions, in metaphysics. Elsewhere, however, Kant suggests that metaphysics must show that there are any synthetic a priori judgments, even in mathematics and physics. While much of the content of the *Critique* suggests that Kant’s considered view must be the latter, he is far from clear about this.

5 Space, time and transcendental idealism

The first part of the *Critique*, the ‘Transcendental Aesthetic’, has two objectives: to show that we have synthetic a priori knowledge of the spatial and temporal forms of outer and inner experience, grounded in our own pure intuitions of space and time; and to argue that transcendental idealism, the theory that spatiality and temporality are only forms in which objects appear to us and not properties of objects as they are in themselves, is the necessary condition for this a priori knowledge of space and time (see Space; Time).

Much of the section refines arguments from the inaugural dissertation of 1770. First, in what the second edition labels the ‘Metaphysical Exposition’, Kant argues that space and time are both *pure forms of intuition* and *pure intuitions*. They are pure *forms* of intuition because they must precede and structure all experience of individual outer objects and inner states; Kant tries to prove this by arguing that our conceptions of space and time cannot be derived from experience of objects, because any such experience presupposes the individuation of objects in space and/or time, and that although we can represent space or time as devoid of objects, we cannot represent any objects without representing space and/or time (A 23-4/B 38-9; A 30-1/B 46). They are pure *intuitions* because they represent single individuals rather than classes of things; Kant tries to prove this by arguing that particular spaces and times are always represented by introducing boundaries into a single, unlimited space or time, rather than the latter being composed out of the former as parts, and that space and time do not have an indefinite number of instances, like general concepts, but an infinite number of parts (A 24/B 39-40; A 31-2/B 47-8).

Next, in the ‘Transcendental Exposition’, Kant argues that we must have an a priori intuition of space because

'geometry is a science which determines the properties of space synthetically and yet a priori' (B 40). That is, the propositions of geometry describe objects in space, go beyond the mere concepts of any of the objects involved - thus geometric theorems cannot be proved without actually constructing the figures - and yet are known a priori. (Kant offers an analogous but less plausible argument about time, where the propositions he adduces seem analytic (B 48).) Both our a priori knowledge about space and time in general and our synthetic a priori knowledge of geometrical propositions in particular can be explained only by supposing that space and time are of subjective origin, and thus knowable independently of the experience of particular objects.

Finally, Kant holds that these results prove transcendental idealism, or that space and time represent properties of things as they appear to us but not properties or relations of things as they are in themselves, let alone real entities like Newtonian absolute space; thus his position of 1768 is now revised to mean that space is epistemologically but not ontologically absolute (A 26/B 42; A 32-3/B 49-50; A 39-40/B 56-7). Kant's argument is that 'determinations' which belong to things independently of us 'cannot be intuited prior to the things to which they belong', and so could not be intuited a priori, while space and time and their properties *are* intuited a priori. Since they therefore cannot be properties of things in themselves, there is no alternative but that space and time are merely the forms in which objects appear to us.

Much in Kant's theory has been questioned by later philosophy of mathematics. Kant's claim that geometrical theorems are synthetic because they can only be proven by construction has been rendered doubtful by more complete axiomatizations of mathematics than Kant knew, and his claim that such propositions describe objects in physical space yet are known a priori has been questioned on the basis of the distinction between purely formal systems and their physical realization.

Philosophical debate, however, has centred on Kant's inference of transcendental idealism from his philosophy of mathematics. One issue is the very meaning of Kant's distinction between appearances and things in themselves. Gerold Prauss and Henry Allison have ascribed to Kant a distinction between two kinds of *concepts* of objects, one including reference to the necessary conditions for the perception of those objects and the other merely leaving them out, with no ontological consequences. Another view holds that Kant denies that things in themselves, and not merely their concepts, are spatial and temporal, and that spatial and temporal properties are, literally, properties only of our own representations of things. Kant makes statements that can support each of these interpretations; but proponents of the second view, including the present author, have argued that it is entailed by both Kant's argument for and his use of his distinction, the latter especially in his theory of free will (see §8).

The debate about Kant's argument for transcendental idealism, already begun in the nineteenth century, concerns whether Kant has omitted a 'neglected alternative' in assuming that space and time must be *either* properties of things as they are in themselves or of representations, but not both, namely that we might have a priori knowledge of space and time because we have an a priori subjective representation of them while they are also objective properties of things. Some argue that there is no neglected alternative, because although the *concepts* of appearances and things in themselves are necessarily different, Kant postulates only one set of *objects*. This author has argued that the 'neglected alternative' is a genuine possibility that Kant intends to exclude by arguing from his premise that propositions about space and time are *necessarily* true: if those propositions were true *both* of our own representations and of their ontologically distinct objects, they might be necessarily true of the former but only contingently true of the latter, and thus not necessarily true throughout their domain (A 47-8/B 65-6). In this case, however, Kant's transcendental idealism depends upon a dubious claim about necessary truth.

6 Pure concepts of the understanding

The 'Transcendental Analytic' of the *Critique* breaks new ground, arguing that the most fundamental categories of thought as well as the forms of perception are themselves human products which are necessary conditions of the possibility of experience. Like the 'Transcendental Aesthetic', its first section, the 'Analytic of Concepts', is also divided into a 'metaphysical' and a 'transcendental deduction' (B 159).

In the metaphysical deduction Kant intends to provide a principle to identify the most fundamental concepts of thought, the categories of the understanding, and then to show that our knowledge of any object always involves these categories. The key to his argument is the claim that knowledge is always expressed in a *judgment* (A 68-9/B 93-4); he then argues that there are certain characteristic forms or 'logical functions' of judgment, and that in order

for our judgments to be about objects, these logical functions of judgments must also provide the basic concepts for conceiving of objects. Thus Kant first produces a table of the logical functions of judgment, based on the premise that every judgment has a *quantity*, *quality*, *relation* and *modality*, and then produces a table of categories, under the same four headings, showing how objects of such judgments must be conceived. Thus, judgments may be universal, particular, or singular, and then their objects must be unities, pluralities, or totalities; judgments may be affirmative, negative, or infinite, and objects manifest either reality, negation, or limitation; judgments may relate a predicate to a subject (categorical judgment), or else relate one predicate-subject judgment to another as antecedent and consequent (hypothetical judgment) or as alternatives (disjunctive judgment), and objects may correspondingly manifest the relations of inherence and subsistence, causality and dependence, or community or reciprocity; finally, judgments may be problematic, assertoric, or apodeictic, thus their objects either possible or impossible, existent or non-existent, or necessary or contingent (A 70/B 95; A 80/B 106).

Kant's scheme is intuitively plausible, and he makes use of it throughout his works. But philosophers as diverse as Hegel and Quine have questioned its coherence and necessity. What is troubling for Kant's own project, however, is that he does not show why we must use all the logical functions of judgment, hence why we must use all the categories. In particular, he does not show why we must make not only categorical but also hypothetical and disjunctive judgments. Without such a premise, Kant's arguments for causation, against Hume, and for interaction, against Leibniz, are not advanced.

Kant's aim and his strategy in the transcendental deduction remain debatable, despite his complete revision of this section in the second edition of the *Critique*. Some view the transcendental deduction as a 'regressive argument' aimed at empiricism, meant to show only that *if* we make judgments about objects then we must use a priori concepts. But if Kant already established this in the metaphysical deduction, the transcendental deduction becomes redundant. It seems more natural to see the latter as intended to fix the *scope* of our use of the categories by showing that we can have *no* experience which is immune from conceptualization under them, thus that the categories enjoy universal objective validity. Because these categories originate in the logical structure of our own thought, Kant holds, we must conceive of ourselves as the autonomous lawgivers for all of nature (A 127-8, B 164).

There are many differences between the two versions of the transcendental deduction, but both employ the fundamental idea that we cannot have some form of *self*-consciousness, or 'transcendental apperception', without also having consciousness of *objects*, which in turn requires the application of the categories; then, since Kant holds that we can have no experience at all without being able to be conscious that we have it, he can argue that we can have no experience to which we cannot apply the categories. The success of this strategy is unclear. The first-edition deduction begins with a debatable analysis of the necessary conditions for knowledge of an object, which slides from the conditional necessity that we must use rules if we are to have knowledge of objects to an absolute necessity that we must have knowledge of objects, and then introduces transcendental apperception as the 'transcendental ground' of the latter necessity (A 106). In the second edition, Kant begins directly with the claim that self-consciousness of our experience is always possible, which has not met with much resistance, but then makes the inference to the necessity of knowledge of objects conceived of through the categories by equating transcendental apperception with a notion of 'objective apperception' that is equivalent to judgment about objects (B 139-40). This makes the connection between self-consciousness and the categorial judgment of objects true by definition, and undermines Kant's claim to provide a synthetic rather than analytic proof of the objective validity of the categories.

In spite of these problems, the idea that self-consciousness depends upon knowledge of objects and thus on the use of the categories to conceive of objects has remained attractive; and some of the most interesting recent work on Kant has been reconstructions of the transcendental deduction, such as those by Peter Strawson, Jonathan Bennett and Dieter Henrich. Others have concluded that Kant only establishes a convincing connection between self-consciousness and categorial thought of objects once he shows that making judgments about objects, using the categories, is a necessary condition for making judgments about the *temporal* order of our experience. This is Kant's project in the next section of the *Critique*.

7 The principles of judgment and the foundations of science

Kant proceeds from the categories to the foundations of natural science in several steps. First, he argues that the

categories, which thus far have merely logical content, must be made 'homogeneous' with experience, or be recast in forms we can actually experience. Since time, as the form of both outer and inner sense, is the most general feature of our sensible experience, Kant argues that the categories must be made homogeneous with experience by being associated with certain determinate temporal relations or 'schemata' (A 138-9/B 177-8). For example, the pure category of ground and consequence, thus far understood only abstractly as the relation of the states of objects that makes them fit to be objects of hypothetical ('if-then') judgments, is associated with the schema of rule-governed temporal succession, something closer to what we can actually experience. Focused as he is on the universality of time, Kant seems to de-emphasize spatiality unduly in the 'Schematism': for example, it would seem more natural to say that the schema of causality is the rule-governed temporal succession of states of objects within an appropriate degree of spatial contiguity.

Next, in the 'System of all Principles of Pure Understanding', Kant argues for the necessity of certain fundamental principles of all natural laws. Following the division of the categories, this chapter is divided into four parts. In the first, the 'Axioms of Intuition', Kant argues that 'All intuitions are extensive magnitudes' (B 202), and thus that all objects of experience are subject to the mathematics of discrete quantities. In the second, the 'Anticipations of Perception', Kant proves that 'In all appearances, the real that is an object of sensation has intensive magnitude, that is, a degree' (B 207), and is thus subject to the mathematics of continuous quantities; here he argues that because our sensations manifest varying degrees of intensity we must also conceive of the objects they represent as manifesting a reality that varies in degree. The first of these two 'mathematical' principles (A 162/B 201) does not add to results already established in the Transcendental Aesthetic, however, and the second depends upon an empirical assumption.

In the next section, the 'Analogies of Experience', dealing with the first of two kinds of 'dynamical' principles, Kant offers some of the most compelling and important arguments in the *Critique*. In the First Analogy, Kant argues that we can determine that there has been a change in the objects of our perception, not merely a change in our perceptions themselves, only by conceiving of what we perceive as successive states of enduring substances (see Substance). Because we can never perceive the origination or cessation of substances themselves, but only changes in their states, Kant argues, the sum-total of substances in nature is permanent (B 224). In the Second Analogy, Kant argues for a further condition for making judgments about change in objects: because even when we undergo a sequence of perceptions, there is nothing in their immediate sensory content to tell us that there is an objective change, let alone what particular sequence of change there is, we can only distinguish a '*subjective sequence* of apprehension from the *objective sequence* of appearances' (A 193/B 238) by judging that a particular sequence of objective states of affairs, *a fortiori* the sequence of our perceptions of those states, has been determined in accordance with a rule that states of the second type can only follow states of the first type - precisely what we mean by a causal law. Finally, the Third Analogy argues that because we always perceive states of objects successively, we cannot immediately perceive states of two or more objects to be simultaneous, and can therefore only judge that two such states simultaneously exist in different regions of space if they are governed by laws of interaction dictating that neither state can exist without the other (A 213/B 260).

Kant's arguments have been assailed on the basis of relativity theory and quantum mechanics. But since they are epistemological arguments that our ability to make temporal judgments about the succession or simultaneity of states of affairs depends upon our judgments about substance, causation and interaction, it is not clear that they are open to objection from this quarter. If relativity tells us that the succession or simultaneity of states of affairs may depend upon the choice of inertial frame, then Kant's theory is not refuted, but just predicts that in that case our own judgments about temporal sequence must also vary. If quantum mechanics tells us that causal laws are merely probabilistic, then Kant's theory is again not refuted but just predicts that in that case our temporal judgments cannot be entirely determinate.

In the last section of the 'Principles', Kant assigns empirical criteria to the modal concepts of possibility, actuality and necessity. The main interest of this section lies in the 'Refutation of Idealism' which Kant inserted into it in the second edition. Here Kant argues that temporal judgments about one's own states require reference to objects which endure in a way that mental representations themselves do not, and therefore that consciousness of oneself also implies consciousness of objects external to oneself (B 275-6; also B xxxix-xli). There has been controversy not only about the precise steps of the proof, but also about whether it is supposed to prove that we have knowledge of the existence of things ontologically distinct from our own representations, which seems to undercut

Kant's transcendental idealism. However, the argument of 1787 was actually just the first of many drafts Kant wrote (Reflections 6311-16, 18: 606-23), and these suggest that he did mean to prove that we know of the existence of objects ontologically distinct from ourselves and our states, although we cannot attribute to them as they are in themselves the very spatiality by means of which we represent this ontological distinctness.

Finally, in the *Metaphysical Foundations of Natural Science*, published between the two editions of the *Critique* (1786), Kant carried his a priori investigation of the laws of nature one step further by introducing not only the empirical notion of change itself but also the further empirical concept of matter as the movable in space (4: 480). With this one empirical addition, he claims, he can deduce the laws of phoronomy, the vectorial composition of motions in space; of dynamics, the attractive and repulsive forces by which space is actually filled; of mechanics, the communication of moving forces; and of phenomenology, which in Kant's sense - derived from J.H. Lambert, and very different from its later senses in Hegel or Husserl - means the laws for distinguishing apparent from real motions. This work is not an essay in empirical physics but rather an exploration of the conceptual framework into which the empirical results of physics must be fitted.

8 The illusions of theoretical reason

In the 'Transcendental Dialectic', Kant argues that the doctrines of traditional metaphysics are illusions arising from the attempt to use the categories of understanding to gain information about objects beyond the horizon of our forms of intuition. What makes such illusions inevitable is the tendency of human reason to seek the unconditioned, that is, to carry a chain of ideas to its assumed completion even when that lies beyond the bounds of sense. For example, understanding may tell us that wholes consist of parts, and sensibility may allow us to find a smaller part for any given whole; but only reason suggests that decomposition into parts must come to an end in something absolutely simple, something we could never perceive by sense. In its practical use, reason may produce ideas of the unconditioned, such as the idea of the universal acceptability of maxims of action, which do not tell us anything misleading about the world because they do not tell us anything about how the world is at all, only how it ought to be; but in its theoretical use reason appears to tell us things about the world that cannot be confirmed by our senses or are even incompatible with the forms of our perception.

This diagnosis of metaphysical error makes good sense of Kant's procedure in the 'Antinomy of Pure Reason', where he presents a series of conflicts between the form and limits of sensibility as structured by the understanding, on the one hand, and the pretensions of unconditioned reason, on the other. In early sketches of the *Dialectic* (Reflections 4756-60, 1775-7, 17: 698-713), Kant diagnosed all of the illusions of traditional metaphysics in this form. In the *Critique*, however, Kant singled out some metaphysical beliefs about the self and about God for separate treatment in the 'Paralogisms of Pure Reason' and 'Ideal of Pure Reason'. These sections offer powerful criticisms of traditional metaphysical doctrines, but require a more complex explanation of metaphysical illusion than the single idea of reason's search for the unconditioned.

In the 'Paralogisms', Kant diagnoses the doctrines of 'rational psychology' that the soul is a *substance* which is *simple* and therefore incorruptible, *numerically identical* throughout the experience of any person, and necessarily *distinct* from any external object (this is how he reformulates the Fourth Paralogism in the second edition (B 409)), as a tissue of ungrounded assertions mistaking the logical properties of the *representation* 'I' or the *concept* of the self for the properties of whatever it is in us which actually thinks (A 355/B 409). Kant's criticism of the traditional metaphysics of the soul is convincing, but does not depend on reason's postulation of the unconditioned; instead, Kant's demonstration that these doctrines arise from confusion between properties of a representation and what is represented showed that they were not inevitable illusions by destroying their credibility once and for all.

The four metaphysical disputes that Kant presents in the 'Antinomy of Pure Reason' are often read as straightforward conflicts between reason and sensibility; but Kant characterizes them as disputes engendered by pure reason itself, so a more complex reading is required. In fact, both sides in each dispute - what Kant calls the 'thesis' and 'antithesis' - reflect different forms of reason's demand for something unconditioned, and what conflicts with the limits of sensibility is the assumption that these demands give rise to a genuine dispute at all. Kant again uses the contrast between 'mathematical' and 'dynamical' to divide the four disputes into two groups, and resolves the disputes in two different ways.

In the first antinomy the dispute is between the thesis that the world has a beginning in time and a limit in space and the antithesis that it is infinite in temporal duration and spatial extension (A 426-7/B 454-5). In the second antinomy, the dispute is between the thesis that substances in the world are ultimately composed of simple parts and the antithesis that nothing simple is ever to be found in the world, thus that everything is infinitely divisible (A 434-5/B 462-3). In each case, thesis and antithesis reflect reason's search for the unconditioned, but in two different forms: in the thesis, reason postulates an ultimate *termination* of a series, and in the antithesis, an unconditional *extension* of the series. In these 'mathematical antinomies', however, Kant argues that *neither* side is true, because reason is attempting to apply its demand for something unconditioned to space and time, which are always *indefinite* in extent because they are finite yet always extendible products of our own cognitive activity (A 504-5/B 532-3).

In the two 'dynamical antinomies' Kant's solution is different. In the third antinomy, the thesis is that 'causality in accordance with laws of nature' is not the only kind of causality, but there must also be a 'causality of freedom' underlying the whole series of natural causes and effects, while the antithesis is that everything in nature takes place in accord with deterministic laws alone (A 444-5/B 462-3). In the fourth antinomy, the thesis is that there must be a necessary being as the cause of the whole sequence of contingent beings, either as its first member or underlying it, while the antithesis is that no such being exists inside or outside the world (A 452-3/B 480-1). Again, the theses result from reason's desire for closure and the antitheses result from reason's desire for infinite extension. But now the theses do not necessarily refer solely to spatio-temporal entities, so the claims that there must be a non-natural causality of freedom and a necessary being can apply to things in themselves while the claims that there are only contingent existents linked by laws of nature apply to appearances. In this case both thesis and antithesis may be true (A 531/B 559). This result is crucial to Kant, because it means that although theoretical reason cannot prove that either freedom or God exist, neither can it disprove them, and room is left for the existence of freedom and God to gain credibility in some other way.

The last main part of the 'Dialectic' is Kant's critique of rational theology. Here Kant reiterates his earlier critique of the ontological argument as well as his claim that the arguments for the existence of God from contingency and from design - the 'cosmological' and 'physico-theological' proofs - can only get from their ideas of a first cause or architect to the idea of a *perfect* being by the supposition of the ontological argument, and thus fall along with that. But he now precedes this argument with a critique of the argument for God as the ground of all possibility that he had earlier accepted: the very idea that there is an *ens realissimum*, an individual being containing in itself the ground of 'the sum-total of all possibility' (A 573/B 602), is another of the natural but illusory ideas of reason.

Kant does not, however, conclude the first *Critique* with an entirely negative assessment of pure reason. In an appendix to the 'Transcendental Dialectic', he argues that even though reason in its theoretical use cannot yield metaphysical insight, it does supply us with indispensable 'regulative' principles, of both the maximal simplicity of natural laws and the maximal variety of natural forms, for the conduct of empirical research; and in the 'Canon of Pure Reason', he argues that practical reason supplies an ideal of the highest good, the union of virtue and happiness and ultimately the union of freedom and nature, which is indispensable for moral conduct, not as its direct object but as a necessary condition of its rationality - which in turn gives ground for the practical postulation if not theoretical proof of the existence of God. Kant expands on both of these ideas in subsequent works (see §11 and §13).

9 The value of autonomy and the foundations of ethics

In his theoretical philosophy, Kant argued that we can be certain of the principles that arise from the combination of the forms of our sensibility and understanding, as products of our own intellectual autonomy; but he also argued that any attempt to see human reason as an autonomous source of metaphysical insight valid beyond the bounds of human sensibility leads to illusion. But in his practical philosophy, Kant argues that human reason is an autonomous source of principles of conduct, immune from the blandishments of sensual inclination in both its determinations of value and its decisions to act, and indeed that human autonomy is the highest value and the limiting condition of all other values.

Traditionally, Kant has been seen as an ethical formalist, according to whom all judgments on the values of ends must be subordinated to the obligatory universality of a moral law derived from the very concept of rationality itself. This interpretation has drawn support from Kant's own characterization of his 'paradoxical' method in the

Critique of Practical Reason, where he holds that the moral law must be derived prior to any determination of good or evil, rather than vice versa (5: 62-3). But this passage does not do justice to the larger argument of Kant's practical philosophy, which is that rationality itself is so valuable precisely because it is the means to freedom or autonomy. Kant expressed this in his classroom lectures on ethics, when he said that 'the inherent value of the world, the *summum bonum*, is freedom in accordance with a will which is not necessitated to action' (27: 1482), and even more clearly in lectures on natural right given in the autumn of 1784, the very time he was writing the *Groundwork of the Metaphysics of Morals*, where he said that 'If only rational beings can be ends in themselves, that is not because they have reason, but because they have freedom. Reason is merely a means' (27: 1321). Kant makes the same point in the *Groundwork* when he says that the incomparable dignity of human beings derives from the fact that they are 'free with regard to all laws of nature, obeying only those laws which' they make themselves (4: 435).

The strategy of the *Groundwork* is by no means obvious, and the real character of Kant's view emerges only gradually. In Section I, Kant tries to derive the fundamental principle of morality from an analysis of 'ordinary rational knowledge of morality.' The key steps in his analysis are: virtue lies in the good will of an agent rather than any natural inclination or any particular end to be achieved; good will is manifested in the performance of an action for the sake of fulfilling duty rather than for any other end; and what duty requires is the performance of an action not for the sake of its consequences but because of its conformity to law as such; thus the maxim, or subjective principle, of virtuous action can only be that 'I ought never to act except in such a way *that I can also will that my maxim should become a universal law*' (4: 402). In Section II, Kant apparently tries to reach the same conclusion from more philosophical considerations: by arguing on the one hand that a moral or practical law must be a categorical rather than hypothetical imperative, that is, one commanding unconditionally rather than depending upon the adoption of some antecedent and optional end, and on the other hand that happiness is too indeterminate an end to give rise to such an imperative, Kant concludes that a categorical imperative can contain 'only the necessity that our maxim should conform to this law', thus that 'there remains nothing to which the maxim has to conform except the universality of a law as such' (4: 421). This version of the categorical imperative is known as the Formula of Universal Law.

Kant then furnishes further formulations of the categorical imperative, especially the Formula of Humanity as an End in Itself - '*Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end*' (4: 429), which at the very least requires the possibility of rational consent to your action from any agent affected by it - and the formula of the kingdom of ends, the requirement that any proposed course of action be compatible with 'a whole of all ends in systematic conjunction (a whole both of rational beings as ends in themselves and also of the personal ends which each may set before himself' (4: 433). The usual interpretation is that these two formulations are supposed to follow from the Formula of Universal Law. However, several factors suggest that Kant did not mean the derivation of that formula from either common sense or 'popular moral philosophy' to be self-sufficient, and it is only with the introduction of the notion that humanity is an end in itself because of its potential for freedom, that the real 'ground of a possible categorical imperative' is discovered (4: 428). If so, then this is Kant's theory: the ultimate source of value is human freedom as an end in itself, manifested in interpersonal contexts in the possibility of freely given consent to the actions of others; conformity to the requirement of universal law is the way to ensure that this value is preserved and fostered; and the ideal outcome of the observation of such a law would be a kingdom of ends as a system of freedom, in which all agents freely pursue their freely chosen ends to the extent compatible with a like freedom for all.

10 Duties of right and duties of virtue

In the *Groundwork*, Kant's principle of morality gives rise to a fourfold classification of duties, resulting from the intersection of two divisions: between duties to oneself and to others, and between perfect and imperfect duties. Perfect duties are proscriptions of specific kinds of actions, and violating them is morally blameworthy; imperfect duties are prescriptions of general ends, and fulfilling them is praiseworthy. The four classes of duty are thus: perfect duties to oneself, such as the prohibition of suicide; perfect duties to others, such as the prohibition of deceitful promises; imperfect duties to oneself, such as the prescription to cultivate one's talents; and imperfect duties to others, such as the prescription of benevolence (4: 422-3, 429-30). It is straightforward what a perfect duty prohibits one from doing; it requires judgment to determine when and how the general ends prescribed by

imperfect duties should be realized.

In the later *Metaphysics of Morals*, Kant works out a detailed budget of duties that is generally based on this scheme, but with one key distinction: duties of justice (*Recht*) are those of the above duties that can appropriately be enforced by means of coercion, and the remainder are duties of virtue, which are fit subjects for moral assessment but not coercion (6: 213, 219). Since freedom is Kant's chief value, coercion is permitted only where it is both necessary to preserve freedom and possible for it to do so. This means that only a small subset of our duties, namely some but not all of our perfect duties to others, are duties of justice, thus proper subjects for public legislation; the majority of our moral duties are duties of virtue which are not appropriate subjects for coercive legal enforcement.

Kant's treatment of the duties of virtue is less complicated than that of the duties of justice, and will be considered first. Kant does not explicitly characterize these as duties to preserve and promote the freedom of oneself and others, as he does in the *Groundwork*, but instead characterizes them as duties to promote one's own perfection and the happiness of others: while one can directly perfect one's own freedom, one can avoid injuring but not directly perfect the freedom of another. On close inspection, however, Kant's duties of virtue require precisely that one perfect both the internal and external conditions for the exercise of one's own freedom and at least the external conditions for the exercise of the freedom of others. Thus, ethical duties to oneself include the prohibition of injury to the physical and mental bases of one's free agency, as by suicide or drunkenness, and the prescription of efforts to improve both the physical and mental conditions for the exercise of one's freedom, as by the cultivation of talents and of one's spiritual and moral faculties themselves; and ethical duties to others include both the prohibition of injuries to the dignity of others as free agents, for example by insulting or ridiculing them ('duties of respect'), and the prescription of efforts to improve the conditions for others' exercise of their own freedom, as by beneficience and sympathy ('duties of love').

Kant's foundation of his political philosophy on the duties of justice is more complicated. From the ultimate value of freedom, Kant derives the universal principle of justice, that an action is right only if 'on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law' (6: 230). Kant then argues that coercion is justified when it can prevent a hindrance to freedom, since a hindrance to a hindrance to freedom is itself a means to freedom (6: 231). This is too simple, since coercion might only compound the injury to freedom. Kant needs to add that coercive enforcement of the law is not itself a hindrance to freedom, since the threat of juridical sanction does not deprive a would-be criminal of freedom in the way that his crime would deprive its victim of freedom: the criminal exercises the choice to risk sanction, but deprives his victim of a like freedom of choice.

Kant goes on to argue that the only proper aim of coercive juridical legislation is the prevention of injury to the person and property of others; this is 'Private Law', while 'Public Law' concerns the proper form of the state, whose function is the enforcement of private law. Kant takes the prevention of injury to persons to be an obvious requirement of duty, needing no special discussion, but the right to property receives extended discussion.

Kant recognizes three classes of property: property in things, property in contracts, and contract-like property in other persons, such as marital rights. His discussion of property in things is the most important for his political theory. The gist of Kant's account is that it would be irrational to deprive ourselves of the right to place physical objects, above all land, at our own long-term disposal, since we are rational agents who may need to use such things to realize our freely chosen ends, while the things themselves are not free agents and have no rights. But since the earth is initially undivided, specific property rights are not innate but must be acquired. Since the claim to any particular thing would limit the freedom of others who might also be able to use it, however, property rights cannot be claimed unilaterally, but can only be claimed with the multilateral consent of those others, which they can reasonably give only if they too are accorded similar rights necessary for the successful exercise of their own agency (6: 255-6). For Kant, the right to property is thus not a natural right of isolated individuals, but a social creation depending upon mutual acceptability of claims. The state, finally, exists primarily to make claims to property rights both determinate and secure, and anyone claiming property rights thus has both the right and the obligation to join in a state with others (6: 256-7, 306-8). Since property exists only by mutual consent, and the state exists to secure that consent, the state necessarily has the power to permit only those distributions of property rights sufficiently equitable to gain general consent.

Both claims to property and expressions of philosophical and religious opinions, for example, are expressions of human autonomy. But while one person's property claims may directly limit the freedom of others, his beliefs do not, and thus do not require the consent of any other. The state therefore has no right to intervene in these matters. This fundamental difference between the state's proper concern with property and its improper concern with personal belief defines Kant's liberalism. It is only implicit in the *Metaphysics of Morals*, but becomes explicit in more purely political writings.

11 Freedom of the will and the highest good

Having considered some practical implications of Kant's conception of autonomy, we now turn to its metaphysical consequences.

In Section III of the *Groundwork*, Kant attempts to prove that the categorical imperative, derived in Section II by the analysis of the concept of free and rational beings, in general actually puts us under an obligation by proving that we are indeed free and rational beings. In his terminology, he wants to show that it is not merely an analytic but a synthetic a priori proposition that our wills are constrained by this imperative. Both the interpretation and the assessment of the arguments by which he proposes to accomplish this remain controversial.

The first claim that Kant makes is that 'every being who cannot act except under the idea of freedom is just on that account really free in a practical respect, that is, all laws that are inseparably bound up with freedom are valid for it just as if its will were really declared to be free in itself and in theoretical philosophy', and that every being with a will must indeed act under the idea of freedom (4: 448) (see Will, the). This seems to mean that agents who conceive of themselves as choosing their own actions, whether or not they conceive of themselves as subject to determinism, do not or perhaps even cannot consider any antecedent determinants of their actions in deciding what to do, but only what now seems most rational to do; thus they must govern their actions by rational and therefore moral laws. This seems right for agents considering their own future actions, but leaves unclear how we are to assess the freedom of the actions of others or even our own past actions.

However, Kant goes on to offer what seems to be a theoretical and therefore general proof of the existence of human freedom. He argues that theoretical philosophy has shown that we must distinguish between considering ourselves as phenomena and noumena, or members of the sensible and the intelligible worlds. From the first point of view, we must consider our actions to be governed by the causality of nature, while in the second, since we cannot consider our actions there to be governed by no law at all, we must consider them to be governed by another kind of causality, namely causality in accord with laws of reason (4: 451-3). Thus while our actions appear to be determined by natural causes, in reality they not only can but in fact must accord with laws of reason, hence with the categorical imperative.

There are two problems with this argument. First, it flouts transcendental idealism by assuming positive knowledge about things in themselves. Second, as Henry Sidgwick was to object a century later, it precludes moral responsibility for wrong-doing: if the real laws of our behaviour are necessarily rational and hence moral, any wrong-doing could only show that an agent is not rational, and therefore not responsible, at all.

Whether consciously aware of these objections or not, Kant radically altered his argument for freedom of the will in the *Critique of Practical Reason*. Here he does not argue from a theoretical proof of our freedom to the fact of our obligation under the moral law, but conversely from our consciousness of that obligation - the 'fact of reason' - to our freedom as the necessary condition of our ability and responsibility to fulfil it (5: 29-31). This argument first assumes that transcendental idealism has left open at least the theoretical possibility of freedom of the will, and then depends upon the famous principle 'ought implies can' ('Theory and Practice', 8: 287). Transcendental idealism, of course, seems problematic to many; and although the 'ought implies can' principle seems an intuitive principle of fairness, Kant does not actually argue for it. Nevertheless, since this argument assumes only that 'ought' implies '*can*', it does not imply that any agent who is obliged under the moral law necessarily *will* act in accordance with it, and thus avoids Sidgwick's problem about the very possibility of wrong-doing.

Kant depends upon this result in his next major treatment of freedom, in *Religion within the Limits of Reason Alone*, although there he seems to go too far by assuming that evil-doing is not just possible but even necessary. Kant begins this discussion with an elegant account of wrong-doing, arguing that because no human being is simply unaware of the demand of morality - that is implied by the 'fact of reason' - acting immorally never comes

from mere ignorance of it, but rather from deciding to exempt oneself from this obligation. This position is compatible with the argument for freedom in the second *Critique*, although not with that of the *Groundwork*. However, Kant goes on to argue that an evil rather than virtuous choice of fundamental maxim, or 'radical evil', is not only *possible* but inevitable, to be escaped from only by a moral conversion. This doctrine hardly follows from Kant's previous argument, and seems instead to rest on an odd mixture of empirical evidence and the lingering grip of the Christian doctrine of original sin.

The reality of freedom is only the first of Kant's three 'postulates of pure practical reason'; the other two are the existence of God and the immortality of the soul. Again Kant's argument is that, as the first *Critique* showed, neither of these can be proven by theoretical metaphysics, but they can nevertheless be postulated as necessary conditions of something essential to morality. In this case, however, they are conditions not of our obligation under the categorical imperative but for the realization of the 'highest good.' This is another complex and controversial concept. Kant typically defines it as happiness in proportion to virtue, which is worthiness to be happy (5: 110), but suggests different grounds for the necessity of this conjunction. In the *Critique of Practical Reason*, Kant sometimes treats happiness and virtue as two separate ends of human beings, one our natural end and the other our moral end, which we simply seek to combine (5: 110). In other places, however, beginning with the 'Canon of Pure Reason' in the first *Critique*, he holds that since what virtue does is precisely to coordinate our mutual pursuit of ends, and happiness arises from the realization of ends, maximal happiness would inevitably follow maximal virtue under ideal circumstances (A 809/B 837). Of course, circumstances are not always ideal for morality: as far as we can see, no one achieves perfect virtue in a normal lifespan, and such virtue as is attained is hardly always rewarded with happiness. To counter this, Kant holds that we may postulate immortality, in which to perfect our virtue, and the existence of God, who can legislate a nature in which the ends of virtue are achieved.

This theory has seemed to many to be Kant's vain attempt to save his personal faith from his own scathing critique of metaphysics. Before such a claim could even be discussed, we would have to know what Kant really means by a postulate of practical reason. Kant gives several hints about this which have not been adequately explored. In the first *Critique*, he discusses the practical postulates in a section where he considers readiness to bet as a measure of belief, thus suggesting that what he actually has in mind is Pascal's wager (see Pascal, B.): since there is no theoretical disproof of these postulates, and nothing to lose if they are false, but their value to happiness is great, it is rational to act as if they were true. In a later essay, a draft on the 'Real Progress of Metaphysics from the Time of Leibniz and Wolff' from the early 1790s (posthumously published), Kant makes an even more striking suggestion. There he says that in the assumption of the practical postulates 'the human being is authorized to grant influence on his actions to an idea which he, in accord with moral principles, has made himself, just as if he had derived it from a given object' (20: 305). Here the suggestion is that the practical postulates are nothing less than another expression of human autonomy: not theoretical beliefs at all, let alone religious dogmas, but ideas which we construct for ourselves solely to increase our own efforts at virtue. This idea, that God is in fact nothing but an idea of our own making for use within our moral practice, is a thought Kant repeatedly expressed in his very last years (see §14).

12 Taste and autonomy

Under the rubric of 'reflective judgment', defined as that use of judgment in which we seek to find unknown universals for given particulars rather than to apply given universals to particulars (5: 179-80), the *Critique of Judgment* deals with three apparently disparate subjects: systematicity in scientific concepts generally, natural and artistic beauty, and teleology or purposiveness in particular organisms and in nature as a whole (see Teleology). Even more than the idea of reflective judgment, however, what ties these subjects together is again the idea of autonomy.

In the the first *Critique*, Kant had suggested, with but few exceptions, that the search for systematicity in scientific concepts and laws - the subordination of maximally varied specific concepts and laws under maximally unified general ones - is an ideal of reason, not necessary for empirical knowledge but still intrinsically desirable. In the third *Critique*, he reassigns this search to reflective judgment, and argues that we must adopt as a transcendental but indemonstrable principle that nature is adapted to our cognitive needs (5: 185; 20: 209-10). By this reassignment Kant indicates that systematicity is a necessary condition for the acceptance of empirical laws after all, and thus a necessary condition for experience itself. Kant thereby suggests that our empirical knowledge is

neither passively received nor simply guaranteed, but dependent on our active projection of the unity of nature.

Kant next turns to judgments of taste as both a further expression of human autonomy and further evidence that the adaptation of nature to our own cognitive needs is both contingent yet reasonably assumed. Judgments of taste, beginning with the simplest such as ‘This flower is beautiful’ and progressing to more complex ones such as ‘This poem is beautiful’ and ‘This landscape is sublime’, are connected to autonomy in two ways: while they claim universal agreement, they must always be based on individual feeling and judgment; and while they must be made free of all constraint by theoretical or moral concepts, they are ultimately symbols of moral freedom itself.

Kant begins from an analysis of the very idea of an ‘aesthetic judgment’. As aesthetic, judgments of taste must both concern and be made on the basis of the most subjective of human responses, feelings of pleasure, but, as judgments, they must still claim interpersonal agreement (5: 203, 212-16). To retain their link to feelings, judgments of taste can never simply report how others respond, but must be based on one’s own free response to the object itself; in this way they express individual autonomy (5: 216, 282-5). But to claim universal agreement, they must be based on cognitive capacities shared by all, yet by a condition of those faculties that is pleasureable because it is not constrained by rules (5: 187). Such a state is one of ’free play’ between imagination and understanding, in which the imagination satisfies understanding’s need for unity by presenting a form that seems unitary and coherent without any concept, or, even where a concept of human use or artistic intention is inescapable, that seems to have a unity going beyond any such concept - artistic genius lies precisely in such transcendence of concepts (5: 317-18). With debatable success, Kant argues that this ‘free play’ must occur under the same circumstances in all human beings (5: 238-9, 290), and thus that judgments of taste can have the ‘quantity’ of universality and the ‘modality’ of necessity while retaining the ‘quality’ of independence from direct moral interest and ‘relation’ to merely subjective, cognitive interests rather than objective, practical ones.

How does aesthetic judgment so understood both express autonomy in a moral sense and also give further evidence of the contingent adaptation of nature to our own needs? Kant answers the latter question with his idea of ‘intellectual interest’: the very fact that beauty exists, he argues, although it cannot be derived from any scientific laws, can be taken by us as evidence that nature is receptive not only to our cognitive needs but even to our need to see a possibility for success in our moral undertakings (5: 300). Kant’s answer to the first question, how taste expresses autonomy in its moral sense, is more complex but also more compelling than this.

Like other eighteenth-century authors such as Edmund Burke, Kant draws a fundamental distinction between the beautiful and the sublime (see Sublime, the). Beauty pleases us through the free play of imagination and understanding. In our response to the sublime, however - which for Kant is not paradigmatically a response to art, but to the vastness of nature - we enjoy not a direct harmony between imagination and understanding, which are rather frustrated by their inability to grasp such immensities, but a grasp of them which reveals the power of reason within us (5: 257). And this, although it would seem to involve theoretical reason, symbolizes the power of practical reason, and thus the foundation of our autonomy, in two ways: our power to grasp a truly universal law, such as the moral law, and our power to resist the threats of mere nature, and thus the blandishments of inclination (5: 261-2).

In this way, the sublime symbolizes the sterner side of moral autonomy. But the experience of beauty is also a symbol of morality, precisely because the freedom of the imagination that is its essence is the only experience in which any form of freedom, including the freedom of the will itself, can become palpable to us (5: 353-4). Kant thus concludes his critique of aesthetic judgment with the remarkable suggestion that it is in our enjoyment of beauty that our vocation as autonomous agents becomes not just a ‘fact of reason’ but a matter of experience as well.

13 Design and autonomy

Kant’s critique of teleological judgment in the second half of the *Critique of Judgment* has an even more complicated agenda than his aesthetic theory. The work has roots in both eighteenth-century biology - which began the debate, lasting until the twentieth century, whether organisms could be understood on purely mechanical principles - and natural theology - that is, the great debate over the argument from design that culminated in Hume’s *Dialogues concerning Natural Religion*. Yet again Kant’s motive is to show that even our understanding of nature ultimately drives us to a recognition of our own autonomy.

The work is divided into three main sections: an examination of the necessary conditions for our comprehension of individual organisms; an examination of the conditions under which we can see nature as a whole as a single system; and a restatement of Kant's moral theology. First, Kant argues that an organism is a system of whole and parts manifesting both 'regressive' and 'progressive' causality: the whole is the product of the parts, but the parts in turn depend upon the whole for their own proper functioning and existence (5: 372, 376). But our conception of mechanical efficient causation includes only progressive causation, in which the state of any system depends upon the prior state of its parts (see Causation). The only way we can understand the regressive causation of the whole with respect to its parts is by analogy to intelligent design, in which an antecedent *conception* of the object as a whole determines the production of the parts which in turn determine the character of the resultant whole. However, Kant insists, we have absolutely no justification for adopting a 'constitutive concept' of natural organisms as a product of actual design; we are only entitled to use an analogy between natural organisms and products of design as 'a regulative concept for reflective judgment to conduct research into objects in a remote analogy with our own causality in accordance with purposes' (5: 375). In other words, seeing organisms as products of intelligent design is a purely heuristic strategy.

However, Kant next argues that if it is natural for us to investigate organisms as if they were products of intelligent design, then it will also be natural for us to try to see nature as a whole as manifesting a purposive design (5: 380-1); and only by seeing the whole of nature as a product of intelligent design - of course, only regulatively - can we satisfy our craving to transform every particularity of nature, which must always be left contingent by our own general concepts, into something that seems necessary (5: 405-7). However, from a merely naturalistic viewpoint the ultimate purpose of nature as a system must remain indeterminate - grass might exist to feed cows, or cows exist to fertilize the grass (4: 426). Nature can be seen as a determinate system only if it can be seen as collectively serving an ultimate end that is itself an intrinsic end. This can only be humanity itself (4: 427) - but not humanity merely as a part of nature, seeking happiness, which is neither a determinate end nor one particularly favoured by nature (4: 430), but only humanity as the subject of morality, able to cultivate its freedom (5: 435-6). Thus the urge to see nature as a systematic whole, an inevitable concomitant of our research into the complexities of organic life, can only be satisfied from the moral point of view in which human autonomy is the ultimate value.

Kant is still careful to remind us that this doctrine is regulative, furnishing us with a principle for our own cognitive and practical activity, not constitutive, pretending to metaphysical insight into the nature of reality independent from us. It is therefore particularly noteworthy that the last part of the critique of teleological judgment is a restatement of Kant's moral theology, the argument for belief in the existence of God as a postulate of practical reason. This restatement in a general theory of reflective judgment, the principles of which are meant above all else to guide our own activity, confirms the view that in the end the theory of practical postulates is not meant to support any form of dogma but only to serve as another expression of our own autonomy.

14 The final decade of Kant's public and private career

German intellectuals were drawn to political issues after the French revolution in 1789, and Kant was no exception. Key elements of his political philosophy were presented in essays such as 'Theory and Practice' (1793) and *Perpetual Peace* (1795) before its formal exposition in the *Metaphysics of Morals* (1797). As was argued above (§10), the foundation of Kantian liberalism is the idea that coercion is justifiable only to prevent hindrances to freedom, and thus to protect personal freedom and regulate property, every claim to which represents a potential constraint of the freedom of others unless they can reasonably agree to that claim as part of a system of property rights; but individual beliefs and conceptions of the good, whether religious or philosophical, do not directly interfere with the freedom of others and are therefore not a proper object of political regulation. Kant's development of this basic principle into a political philosophy, however, is complex and controversial.

On the one hand, Kant argued from this premise to a firm rejection of any paternalistic government, even benevolent paternalism. Government exists for the protection of the freedom individuals have to determine and pursue their own ends to the extent compatible with the like freedom of others; so a '*paternalistic government*, where the subjects, as minors, cannot decide what is truly beneficial or detrimental to them, but are obliged to wait passively for the head of state to judge how they *ought* to be happy…would be the greatest conceivable *despotism*' (8: 290-1). Further, Kant held that the sovereignty of any government derives solely from the possibility of those who are governed rationally consenting to it, and thus that it is a necessary test of the legitimacy of all laws 'that

they can have arisen from the united will of an entire people' (8: 297). These constraints could best be met in a republic, without a hereditary monarchy or aristocracy pitting proprietary privilege against public right. Finally, Kant argued in *Perpetual Peace*, only in a world federation of republics, where no proprietary rulers could identify the forcible extension of their domains with the aggrandizement of their personal property, could a cessation of warfare ever be expected.

On the other hand, Kant accompanied these liberal doctrines with a denial of any right to violent revolution, which has seemed surprising to many. But Kant's thought here is complex. Underlying his position as a whole is his view that in any situation in which different persons are bound to come into contact with each other we have not merely a moral right but a moral obligation to found or uphold a state. But one could easily argue that a tyranny is a state in name only, and that our moral obligation with regard to a tyranny is precisely to replace it at any cost with a legitimate state. Kant offers several reasons why this is not so. One claim is that violent revolution does not leave time for genuine reform in principles (8: 36), and another argument is that people revolt for the sake of greater happiness, which is an illegitimate reason for the overthrow of a state (8: 298). But these are empirical claims, and do not prove that people cannot revolt solely to remove illegitimate constraints to their freedom. Another argument Kant makes is that a constitution granting a legal right to rebel against the highest authority it creates would thereby not create a single highest authority after all, and would thus be self-contradictory (6: 319). This has seemed to many to be a sophism; but it may have been Kant's attempt to get his liberalism past the Prussian censorship, denying a *legal* right to rebel without ever explicitly denying a *moral* right to rebel.

Kant had been battling censorship even before the death of Frederick the Great in 1786. In 'What is Enlightenment?' (1785), he argued that while persons in an official capacity have to obey orders (in what he confusingly calls the 'private use of reason'), no official, not a professor or even a military officer, has to surrender his right to address his views to 'the entire *reading public*' (the 'public use of reason') (8: 37). But Kant's attack on the necessity of an established church in *Religion within the Limits of Reason Alone* (1793), even though legally published with the imprimatur of a non-Prussian university (Jena), outraged the conservative Frederick William II and his minister Wöllner, and Kant was threatened with punishment if he published further on religion. With an oath of loyalty to his sovereign, Kant promised to desist, but after the death of this king in 1797 he regarded himself as freed from this promise, and the next year issued his most spirited defence of intellectual freedom yet, *The Conflict of the Faculties*. Here Kant argued that while the theological faculty might have the obligation to advance certain dogmas approved by the state, it was nothing less than the official function of the philosophical faculty to subject all views to rational scrutiny; and in any case, a government genuinely concerned with its people's welfare would not want them to base their morality on fear or dogma but only on the free exercise of their own reason. The new government had no stomach for further suppression of the aged philosopher, and Kant was able to publish this defence of intellectual freedom without incident.

Privately, Kant's last years were devoted to the project of closing the gap between the metaphysical foundations of natural science and actual physics, begun about 1796. He never published the work, leaving behind only the notes later published as the *Opus postumum*. Here Kant tried to show that by using the categorial framework and the concept of force we can derive not only the most general laws of mechanics, as he had argued in 1786, but a much more detailed categorization of the forms of matter and its forces. Kant also argued that an imperceptible, self-moving ether or 'caloric' is a condition of the possibility of experience. In the latest stages of this work, however, Kant returned to the broadest themes of his philosophy, and tried to develop a final statement of transcendental idealism. Here he argued that 'The highest standpoint of transcendental philosophy is that which unites God and the world synthetically, under one principle' (21: 23) - where that principle is nothing other than human autonomy itself. God and the world are 'not substances outside my thought, but rather the thought through which we ourselves make these objects' (21: 21): the world is our experience organized by categories and laws of our own making, and God is the representation of our own capacity to give ourselves the moral law through reason. The moral law 'emerges from freedom…which the subject prescribes to himself, and yet as if another and higher person had made it a rule for him. The subject feels himself necessitated through his own reason…' (22: 129). This is a fitting conclusion to Kant's philosophy of autonomy.

See also: A priori; Analyticity; Autonomy, ethical; Empiricism; Free will; Kantian ethics; Neo-Kantianism; Practical reason and ethics; Rationalism; Transcendental arguments

PAUL GUYER

List of works

Citations to Kant, with the exception of the *Critique of Pure Reason*, are standardly located by the appropriate volume and page number in Kant (1900-). This practice has been followed in this entry by giving the arabic volume number followed by the arabic page number. Citations to the *Critique of Pure Reason* are given with the pagination of the first (A) and/or second (B) editions, according to whether the passage occurred in one or both editions.

1 German editions

Kant, I. (1900-) *Kant's gesammelte Schriften (Kant's Collected Works)*, ed. Royal Prussian (subsequently German) Academy of Sciences, Berlin: Georg Reimer, subsequently Walter de Gruyter, 29 vols, in 34 parts.(27 volumes are published so far. The edition is divided into four parts: *Werke (Works)* (vols 1-9), *Briefe (Letters)* (vols 10-13), *Handschriftlicher Nachlaß (Handwritten remains)* (vols 14-23), and *Vorlesungen (Transcriptions of lectures by other hands)* (vols 24-29, no volumes 25 and 26). This edition, referred to as the 'Akademie' edition, is the most complete collection of Kant's works. While some of its texts of Kant's published works have been superseded, it remains the only source for most of Kant's unpublished notes and lectures.)

Kant, I. (1956-62) *Immanuel Kant: Werke in sechs Bänden (Immanuel Kant: Works in Six Volumes)*, ed. W. Weischedel, Wiesbaden: Insel; repr. (in 12 vols but with original pagination) Frankfurt am Main: Suhrkamp Verlag, 1968.(A more modern edition of Kant's published works only, also including German translations of Kant's Latin works. Orthography is modernized, but the texts are more reliable than in the older Akademie edition.)

Kant, I. (1990) *Kritik der reinen Vernunft (Critique of Pure Reason)*, ed. R. Schmidt, with bibliography by Heinrich Klemme, Hamburg: Felix Meiner, 3rd edn.(Published in the Philosophische Bibliothek series. This edition was the basis for N. Kemp Smith's translation)

Kant, I. (1996) *Kritik der reinen Vernunft Reclam (Critique of Pure Reason)*, ed. I. Heidemann, Stuttgart: Philipp Reclam Jun.(Published in the Universal-Bibliothek series. The best modern edition of the work)

2 English translations

The Cambridge Edition of the *Works of Immanuel Kant*, edited by P. Guyer and A.W. Wood, will provide new or revised translations of all of the works Kant published in his lifetime and extensive selections from his letters, his surviving notes, drafts and fragments and surviving transcriptions of his lectures. As of 1998, the following volumes had appeared:

Kant, I. (1992) *Theoretical Philosophy, 1755-1770*, trans. and ed. D. Walford, in collaboration with Ralf Meerbote, Cambridge: Cambridge University Press.(Contains all of Kant's writings up to 1770 except for scientific works and *Observations on the Feeling of the Beautiful and Sublime*. Important works contained in this volume include his first metaphysical treatise of 1755, the *Only Possible Basis for a Demonstration of the Existence of God* (1763), the *Dreams of a Spirit-Seer* (1766), and the inaugural dissertation of 1770.)

Kant, I. (1992) *Lectures on Logic*, trans. and ed. J.M. Young, Cambridge: Cambridge University Press.(Contains the authorized handbook edited by G.B. Jäsche in 1800, and three complete sets of lecture notes from the early 1770s, 1780s and 1790s.)

Kant, I. (1993) *Opus postumum*, ed. E. Förster, trans. E. Förster and M. Rosen, Cambridge: Cambridge University Press.(Contains extensive selections from the surviving manuscripts of Kant's attempt to write a final book that would both restate the principles of his transcendental idealism and also complete the application of his fundamental principles to natural science.)

Kant, I. (1996) *Practical Philosophy*, ed. and trans. M.J. Gregor with introduction by A.W. Wood, Cambridge: Cambridge University Press.(Includes new translations of the *Groundwork of the Metaphysics of Morals* and *Critique of Pure Reason*, a revised edition of Gregor's previous translation of the *Metaphysics of Morals*, and Kant's occasional essays on moral and political philosophy.)

Kant, I. (1996) *Religion and Rational Theology*, ed. and trans. G. Di Giovanni and A.W. Wood, Cambridge: Cambridge University Press.(Includes a new translation of *Religion within the Boundaries of Mere Reason*; the *Conflict of the Faculties*, which presents Kant's main argument for the intellectual freedom of philosophers; and Kant' lectures on philosophical theology.)

Kant, I. (1997) *Lectures on Ethics*, ed. P. Heath and J.B. Schneewind, trans. P. Heath, Cambridge: Cambridge University Press.(Contains complete sets of notes from 1775-84, 1784-5 and 1793-4.)

Kant, I. (1997) *Lectures on Metaphysics*, ed. and trans. K. Ameriks and S. Naragon, Cambridge: Cambridge

University Press.(Contains complete lecture series from the mid-1770s, 1782-3, 1790-1 and 1794-5, as well as selections from other sets of notes as early as 1762-4. Kant's views on both empirical and rational psychology are particularly well represented.)

Kant, I. (1998) *Critique of Pure Reason*, ed. and trans. P. Gruyer and A.W. Wood, Cambridge: Cambridge University Press.(In addition to an extensive introduction and cross-references to many relevant notes among Kant's literary remains, this is the first edition of the *Critique* to include all of Kant's notes in his own copy of the first edition of 1781.)

3 Individual works

Kant, I. (1911) *Critique of Aesthetic Judgement*, trans., with analytical indexes, J.C. Meredith, Oxford: Clarendon Press.

Kant, I. (1928) *Critique of Teleological Judgement*, trans., with analytical indexes, J.C. Meredith, Oxford: Clarendon Press.

Kant, I. (1997) *Prolegomena to Any Future Metaphysics, with selections from the Critique of Pure Reason*, ed. and trans. G. Hatfield, Cambridge: Cambridge University Press.(A reliable revision of the Paul Carus translation of 1902.)

Kant, I. (1952) *Kant's Critique of Judgement*, trans. J.C. Meredith, Oxford: Clarendon Press.(Reprint of Kant (1911) and (1928), without indices. While loose, Meredith's remains the most felicitous translation of *Critique of Judgment*.)

Kant, I. (1960) *Observations on the Feeling of the Beautiful and Sublime*, trans. J.T. Goldthwait, Berkeley and Los Angeles, CA: University of California Press.

Kant, I. (1965) *First Introduction to the Critique of Judgment*, trans. J. Haden, Indianapolis and New York: Bobbs-Merrill.(Better than Pluhar's translation, but now hard to find.)

Kant, I. (1967) *Kant: Philosophical Correspondence 1759-99*, trans. A. Zweig, Chicago, IL: University of Chicago Press.(A selection of the most philosophically interesting letters with useful annotations.)

Kant, I. (1968) *Selected Pre-Critical Writings and Correspondence with Beck*, trans. G.B. Kerferd and D.E. Walford, with a contribution by P.G. Lucas, Manchester and New York: Manchester University Press and Barnes & Noble.(Includes Kant's correspondence with Marcus Herz in the 1770s and J.S. Beck in the 1790s, not included in Kant's *Theoretical Philosophy, 1755-1770*.)

Kant, I. (1969) *Universal Natural History and Theory of the Heavens*, trans. W. Hastie, with new introduction by M.K. Munitz, Ann Arbor, MI: University of Michigan Press.(Antiquated translation.)

Kant, I. (1970) *Metaphysical Foundations of Natural Science*, trans. J. Ellington, Indianapolis and New York: Bobbs-Merrill; repr., with *Prolegomena*, in *Philosophy of Material Nature*, Indianapolis, IN: Hackett, 1985. (The only translation currently available.)

Kant, I. (1973) *The Kant-Eberhard Controversy: An English translation together with supplementary materials and a historical-analytical introduction of Immanuel Kant's On a New Discovery According to which Any New Critique of Pure Reason Has Been Made Superfluous by an Earlier One*, trans. H.E. Allison, Baltimore, MD: The Johns Hopkins University Press.(Kant's polemical defence of the idea of the synthetic a priori from the Wolffians, with valuable supplementary material.)

Kant, I. (1974) *Anthropology from a Pragmatic Point of View*, trans. M.J. Gregor, The Hague: Martinus Nijhoff.

Kant, I. (1983) *What Real Progress Has Metaphysics Made in Germany since the Time of Leibniz and Wolff?*, trans. T. Humphrey, New York: Abaris Books.(Only English translation of this unfinished competition entry, posthumously published by Rink in 1804.)

Kant, I. (1986) *Kant's Latin Writings: Translations, Commentaries, and Notes*, trans. L.W. Beck, M.J. Gregor, R. Meerbote and J.A. Reuscher, New York, Bern, Frankfurt and Paris: Peter Lang.(Includes some fragments not included in Kant's *Theoretical Philosophy, 1755-1770*.)

Kant, I. (1987) *Critique of Judgment: Including the First Introduction*, trans. W.S. Pluhar, Indianapolis, IN: Hackett Publishing Company.(While the English is infelicitous and the editorial intervention excessive, this does include the only translation of the 'First Introduction' still in print.)

Kant, I. (1991) *The Metaphysics of Morals*, trans. M.J. Gregor, Cambridge: Cambridge University Press.(The first single-volume translation of both the Rechtslehre and Tugendlehre, supersedes all earlier separate translations.)

Kant, I.(1991) *Kant's Political Writings*, ed. H. Reiss, trans. H.B. Nisbet, Cambridge: Cambridge University Press, 2nd edn.(Includes most of what is in Kant on history as well as 'Theory and Practice' and 'Public Right' from the Rechtslehre.)

References and further reading

1 Multi-author anthologies and general background

Beck, L.W. (1965) *Studies in the Philosophy of Kant*, Indianapolis, IN: Bobbs-Merrill Company.(Contains Beck's essays on analytic-synthetic distinction and on political philosophy.)

Beck, L.W. (1969) *Early German Philosophy: Kant and his Predecessors*, Cambridge, MA: Harvard University Press.(Detailed history of German philosophy from the middle ages to Kant, the only work of its kind.)

Beck, L.W. (1978) *Essays on Kant and Hume*, New Haven and London: Yale University Press.(Contains noteworthy essays on causation.)

Beiser, F.C. (1987) *The Fate of Reason: German Philosophy from Kant to Fichte*, Cambridge, MA: Harvard University Press.(A detailed study of the reception of Kant; contains extensive bibliography of primary sources.)

Chadwick, R. (ed.) (1992) *Immanuel Kant: Critical Assessments*, London: Routledge, 4 vols.(One volume of classical critiques from Friedrich Schiller to Strawson and Bennett, followed by three volumes of recent articles.)

De Vleeschauwer, H.J. (1934-7) *La Déduction Transcendentale dans l'œvre de Kant (The Transcendental Deduction in the Work of Kant)*, Antwerp, Paris, and the Hague: De Sikkel, Champion, and Martinus Nijhoff, 3 vols.(More general than its title suggests, it is still the most detailed study of Kant's philosophical development available.)

De Vleeschauwer, H.J. (1962) *The Development of Kantian Thought: The History of a Doctrine*, trans. A.R.C. Duncan, London: Thomas Nelson.(An abridgement of De Vleeschauwer (1934-7).)

Förster, E. (ed.) (1989) *Kant's Transcendental Deductions: The Three 'Critiques' and the 'Opus postumum'*, Stanford: Stanford University Press.(Includes noteworthy essays by Dieter Henrich and John Rawls.)

Guyer, P. (ed.) (1992) *The Cambridge Companion to Kant*, Cambridge: Cambridge University Press.(Specially written chapters surveying the whole of Kant's philosophy and an extensive bibliography.)

Guyer, P. (ed.) (1998) *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, Lanham, MD: Rowman & Littlefield.(Includes essays by the leading writers on the *Groundwork* from the last two decades, and an extensive bibliography.)

Henrich, D. (1994) *The Unity of Reason: Essays on Kant's Philosophy*, ed. R. Velkley, Cambridge, MA: Harvard University Press.(Translations of Henrich's important monograph on the transcendental deduction from 1976 and several of his essays on Kant's ethics.)

Höffe, O. (1994) *Immanuel Kant*, trans. M. Farrier, Albany: State University of New York Press. (The best recent German survey of Kant's philosophy.)

Walker, R.C.S. (1978) *Kant*, London: Routledge & Kegan Paul.(A well-informed, argumentative overview.)

Wolff, R.P. (ed.) (1967) *Kant: A Collection of Critical Essays*, Garden City: Doubleday Anchor.(Includes Lewis White Beck's papers on the analytic-synthetic distinction.)

2 Individual topics

Allison, H.E. (1983) *Kant's Transcendental Idealism*, New Haven and London: Yale University Press.(Influential presentation of the 'two-aspect' theory, includes extensive bibliography.)

Allison, H.E. (1990) *Kant's Theory of Freedom*, Cambridge: Cambridge University Press.(Controversial defence of a theoretical proof of freedom.)

Ameriks, K. (1982) *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*, Oxford: Clarendon Press.(Using Kant's lectures on metaphysics, argues that Kant's departure from traditional rational psychology is not as radical as is usually assumed.)

Aune, B. (1979) *Kant's Theory of Morals*, Princeton, NJ: Princeton University Press.(A careful analytical study.)

Beck, L.W. (1960) *A Commentary on Kant's Critique of Practical Reason*, Chicago, IL: University of Chicago Press.(Only detailed commentary on the second *Critique*, defends a perspectival approach to freedom.)

Beiser, F.C. (1992) *Enlightenment, Revolution and Romanticism: The Genesis of Modern German Political Thought, 1790-1800*, Cambridge, MA: Harvard University Press.(Places Kant's political thought in detailed historical context.)

Bennett, J. (1966) *Kant's Analytic*, Cambridge: Cambridge University Press.(Together with Bennett (1974) constitutes one of the most influential analytical studies of Kant; the 1966 work remains more successful.)

Bennett, J. (1974) *Kant's Dialectic*, Cambridge: Cambridge University Press. (Together with Bennett (1966) constitutes one of the most influential analytical studies of Kant; the 1966 work remains more successful.)

Brittan, G.G., Jr. (1978) *Kant's Theory of Science*, Princeton, NJ: Princeton University Press.(A contemporary assessment of Kant's philosophy of mathematics and physics.)
Brook, A. (1994) *Kant and the Mind*, Cambridge: Cambridge University Press.(A study of Kant's philosophy of mind aimed at contemporary philosophers.)
Buchdahl, G. (1992) *Kant and the Dynamics of Reason: Essays on the Structure of Kant's Philosophy*, Oxford: Basil Blackwell.(Contains many of Kant's influential papers, especially on 'regulative principles'.)
Butts, R.E. (ed.) (1986) *Kant's Philosophy of Physical Science: Metaphysische Anfangsgründe der Naturwissenschaft 1786-1986*, Dordrecht: Reidel.(Contains important papers by G. Brittan, G. Buchdahl, M. Friedman and P. Kitcher, among others).
Cohen, T. and Guyer P. (eds) (1982) *Essays in Kant's Aesthetics*, Chicago, IL: University of Chicago Press.(Includes important papers by R. Aquila, D. Crawford and D. Henrich as well as the editors; includes extensive bibliography.)
Crawford, D.W. (1974) *Kant's Aesthetic Theory*, Madison: University of Wisconsin Press.(The first important analytical study of Kant's aesthetics, it argues for a moral completion of the deduction of taste.)
Crowther, P. (1989) *The Kantian Sublime: From Morality to Art*, Oxford: Clarendon Press.(Tries to broaden the Kantian theory of the sublime beyond morality.)
Dryer, D.P. (1966) *Kant's Solution for Verification in Metaphysics*, London: George Allen & Unwin.(Overshadowed by Strawson (1966) and Bennett (1966), remains a wordy but valuable study of Kant's theory of the synthetic a priori.)
England, F.E. (1930) *Kant's Conception of God: A Critical Exposition of its Metaphysical Development together with a Translation of the Nova Delucidatio*, New York: Dial Press.(An account of Kant's early as well as mature treatment of God.)
Ewing, A.C. (1924) *Kant's Treatment of Causality*, London: Routledge & Kegan Paul.(Old but still insightful.)
Friedman, M. (1992) *Kant and the Exact Sciences*, Cambridge, MA: Harvard University Press.(Influential papers on Kant's philosophy of geometry and physics, and a detailed study of the *Opus postumum*.)
Galston, W.A. (1975) *Kant and the Problem of History*, Chicago, IL: The University of Chicago Press.(A wide-ranging survey focusing on history and political progress.)
Gregor, M. (1963) *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik der Sitten*, Oxford: Basil Blackwell.(The standard English commentary on the *Metaphysics of Morals*.)
Guyer, P. (1979) *Kant and the Claims of Taste*, Cambridge, MA: Harvard University Press; 2nd edn, Cambridge: Cambridge University Press, 1997.(A critical scrutiny of Kant's analysis and deduction of judgments of taste, now amplified with a chapter on Kant's conception of art.)
Guyer, P. (1987) *Kant and the Claims of Knowledge*, Cambridge: Cambridge University Press.(After a study of Kant's development in the 1770s, separates Kant's theory of categories from transcendental idealism.)
Guyer, P. (1993) *Kant and the Experience of Freedom: Essays on Aesthetics and Morality*, Cambridge: Cambridge University Press.(A collection of essays placing Kant's aesthetics in its historical context and exploring connections with Kant's moral philosophy.)
Herman, B. (1992) *The Practice of Moral Judgment*, Cambridge, MA: Harvard University Press.(A vigorous defence of a non-formalistic approach to Kant's ethics.)
Hill, T.E., Jr. (1992) *Dignity and Practical Reason in Kant's Moral Theory*, Ithaca, NY: Cornell University Press.(Important papers, primarily on Kant's normative ethics.)
Howell, R. (1992) *Kant's Transcendental Deduction: An Analysis of Main Themes in his Critical Philosophy*, Dordrecht and Boston: Kluwer.(A detailed study influenced by Hintikka's philosophy of language.)
Kemp Smith, N. (1923) *A Commentary to Kant's Critique of Pure Reason*, London: Macmillan, 2nd edn.(The classical English presentation of the 'patchwork theory' of the composition of the first *Critique*, it remains stimulating and the only important English commentary on every part of the *Critique*.)
Kersting, W. (1993) *Wohlgeordnete Freiheit: Immanuel Kants Rechts- und Staatsphilosophie*, Frankfurt am Main: Suhrkamp, 2nd edn.(The most important German study of Kant's political philosophy, it argues for tension between liberal and conservative elements; 2nd edition adds introduction on Kant's contemporary impact and includes extensive bibliography.)
Kitcher, P.(1990) *Kant's Transcendental Psychology*, Oxford: Oxford University Press.(A defence of Kant's cognitive psychology.)
Korsgaard, C.M. (1996) *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.(Includes all of the author's important papers on Kant's ethics, originally published between 1983 and 1993.)
Linden, van der, H. (1988) *Kantian Ethics and Socialism*, Indianapolis, IN, and Cambridge, MA: Hackett.(Finds

grounds for socialism in Kant.)
Makkreel, R. (1990) *Imagination and Interpretation in Kant*, Chicago, IL: University of Chicago Press.(A hermeneutic approach.)
Longuenesse, B. (1998) *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*, trans. C.T. Wolfe, Princeton, NJ: Princeton University Press.
McFarland, J.D. (1970) *Kant's Concept of Teleology*, Edinburgh: Edinburgh University Press.(A useful general survey in English.)
McLaughlin, P. (1990) *Kant's Critique of Teleology in Biological Explanation*, Lewiston, NY: Edwin Mellen Press.(Focuses on Kant's theory of organisms.)
Melnick, A. (1973) *Kant's Analogies of Experience*, Chicago, IL: University of Chicago Press.(Indispensable interpretation of the Analogies.)
(1989) *Space, Time, and Thought in Kant*, Dordrecht: Kluwer.(Immensely complicated reconstruction of the genesis and content of Kant's theory of empirical judgment.)
Mulholland, L.A. (1990) *Kant's System of Rights*, New York: Columbia University Press.(An interesting successor to Gregor's commentary on the *Metaphysics of Morals*.)
Nell, O. (O'Neill) (1975) *Acting on Principle: An Essay on Kantian Ethics*, New York: Columbia University Press.(The most important interpretation of the categorical imperative.)
O'Neill, O. (1989) *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge: Cambridge University Press.(Papers on both the interpretation and application of Kantian ethics.)
Paton, H.J. (1936) *Kant's Metaphysics of Experience: A Commentary on the First Half of the Kritik der reinen Vernunft*, London: George Allen & Unwin, 2 vols.(Extensive rebuttal of Kemp Smith's 'patchwork theory'; like the contemporary work by De Vleeschauwer, makes much use of Kant's notes.)
(1947) *The Categorical Imperative: A Study in Kant's Moral Philosophy*, London: Hutchinson.(The classical English commentary on the *Groundwork*.)
Prauss, G. (1974) *Kant und das Problem der Dinge an sich (Kant and the Problem of the Ding an sich)*, Bonn: Bouvier.(An influential presentation of the 'two-aspect' interpretation of the distinction between appearances and things in themselves.)
Reich, K. (1992) *The Completeness of Kant's Table of Judgments*, trans. J. Kneller and M. Losoncy, Stanford: Stanford University Press.(Influential defense of the metaphysical deduction.)
Riley, P. (1983) *Kant's Political Philosophy*, Totowa: Rowman and Littlefield.(Examines tensions between Kant's liberalism and his teleology.)
Rosen, A.D. (1993) *Kant's Theory of Justice*, Ithaca, NY: Cornell University Press.(Defends Kant's position on revolution.)
Savile, A. (1987) *Aesthetic Reconstructions: The Seminal Writings of Lessing, Kant, and Schiller*, Aristotelian Society Series 8, Oxford: Basil Blackwell.(A useful comparative study.)
Savile, A. (1993) *Kantian Aesthetics Pursued*, Edinburgh: Edinburgh University Press.(A variety of studies, including applications of Kantian theory to music and architecture.)
Strawson, P.F. (1966) *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*, London: Methuen.(An elegant reconstruction, based on Strawson's theory of meaning, that stimulated much of the best work on Kant in the 1960s and 1970s.)
Vaihinger, H. (1881-92) *Commentar zu Kants Kritik der reinen Vernunft (Commentary on Kant's Critique of Pure Reason)*, Stuttgart: W. Spemann and Union Deutsche Verlagsgesellschaft, 2 vols.(The most detailed study of the Introduction and Transcendental Aesthetic ever produced.)
Ward, K. (1972) *The Development of Kant's View of Ethics*, Oxford: Basil Blackwell.(The only modern study of this topic in English.)
Williams, H. (1983) *Kant's Political Philosophy*, New York: St. Martin's.(A general survey.)
(ed.) (1992) *Essays on Kant's Political Philosophy*, Chicago, IL: University of Chicago Press.(Contains essays by O. Höffe, W. Kersting and O. O'Neill among others.)
Wolff, R.P. (1963) *Kant's Theory of Mental Activity: A Commentary on the Transcendental Analytic of the Critique of Pure Reason*, Cambridge, MA: Harvard University Press.(Valuable commentary on the first-edition deduction and the second Analogy.)
Wood, A.W. (1970) *Kant's Moral Religion*, Ithaca and London: Cornell University Press.(A valuable study of Kant's moral theology.)
(1978) *Kant's Rational Theology*, Ithaca and London: Cornell University Press.(A study of Kant's critique of the traditional arguments for the existence of God, drawing heavily on Kant's lectures.)
Yovel, Y. (1980) *Kant and the Philosophy of History*, Princeton, NJ: Princeton University Press.(Argues for a

proto-Hegelian account of Kant's philosophy of historical progress.)
Zumbach, C. (1984) *The Transcendent Science: Kant's Conception of Biological Methodology*, The Hague: Martinus Nijhoff.(Focuses on Kant's theory of organisms.)

Kantian ethics

Kantian ethics originates in the ethical writings of Immanuel Kant (1724-1804), which remain the most influential attempt to vindicate universal ethical principles that respect the dignity and equality of human beings without presupposing theological claims or a metaphysical conception of the good. Kant's systematic, critical philosophy centres on an account of reasoning about action, which he uses to justify principles of duty and virtue, a liberal and republican conception of justice with cosmopolitan scope, and an account of the relationship between morality and hope.

Numerous contemporary writers also advance views of ethics which they, and their critics, think of as Kantian. However, some contemporary work is remote from Kant's philosophy on fundamental matters such as human freedom and reasoning about action. It converges with Kant's ethics in claiming that we lack a substantive account of the good (so that teleological or consequentialist ethics are impossible), in taking a strong view of the equality of moral agents and the importance of universal principles of duty which spell out what it is to respect them, and in stressing an account of justice and rights with cosmopolitan scope.

Both Kant's ethics and contemporary Kantian ethics have been widely criticized for preoccupation with rules and duties, and for lack of concern with virtues, happiness or personal relationships. However, these criticisms may apply more to recent Kantian ethics than to Kant's own ethics.

1 Kant's ethics

Kant's main writing on ethics and politics can be found in *Grundlegung zur Metaphysik der Sitten (Groundwork of the Metaphysics of Morals)* (1785), *Kritik der practischen Vernunft (Critique of Practical Reason)* (1788), *Die Metaphysik der Sitten (The Metaphysics of Morals)* (1797) and numerous sections of other works and free-standing essays. Throughout these writings he insists that we cannot derive ethical conclusions from metaphysical or theological knowledge of the good (which we lack) or from a claim that human happiness is the sole good (which we cannot establish). We lack the basis for a teleological or consequentialist account of ethical reasoning, which therefore cannot be simply a matter of means-ends reasoning towards some fixed and knowable good (see Consequentialism; Teleological ethics).

Yet if reasoning about action, that is practical reasoning, is not means-end reasoning, what can it be? Kant's alternative account proposes simply that reasons for action must be reasons for all. He insists that we can have reasons for recommending only those principles of action which could be adopted by all concerned, whatever their particular desires, social identities, roles or relationships. Correspondingly, practical reasoning must reject any principles which cannot be principles for all concerned, which Kant characterizes as *non-universalizable* principles (see Universalism in ethics §5).

Kant gives this rather limited modal conception of practical reasoning some grand names. He calls it the 'supreme principle of morality' and the 'categorical imperative'. He formulates this fundamental principle of ethics in various ways. The formulation most discussed in the philosophical literature runs 'act only on that maxim [principle] through which you can at the same time will that it become a universal law' ([1785] 1903: 421). The formulation that has had and still has the greatest cultural resonance requires us to treat others with impartial respect. It runs 'treat humanity…never simply as a means, but always at the same time as an end' ([1785] 1903: 429) (see Respect for persons §2). The equivalence of these two formulations of the categorical imperative is far from obvious. One way of glimpsing why Kant thought they were equivalent is to note that if we treat others as persons rather than as things then we must not destroy or impair their abilities to act, indeed must leave it open to them to act on the same principles that we act on; hence we must act on universalizable principles. On Kant's view, one of the worst features of consequentialist ethics is that it not merely permits but requires that persons be used as mere means if this will produce good results.

Kant claims that the categorical imperative can be used to justify the underlying principles of human duties (see Duty §2). For example, we can show by a *reductio ad absurdum* argument that promising falsely is not universalizable. Suppose that everyone were to adopt the principle of promising falsely: since there would then be much false promising, trust would be destroyed and many would find that they could not get their false promises

accepted, contrary to the hypothesis of universal adoption of the principle of false promising. A maxim of promising falsely is not universalizable, so the categorical imperative requires us to reject it. Parallel arguments can be used to show that principles such as those of coercing or doing violence are not universalizable, and so that it is a duty to reject these principles.

Kant calls duties such as these *perfect* (namely, complete) duties. These are duties which can observed by each towards all others. He also provides arguments to establish the principles of certain *imperfect* (namely, incomplete) duties, such as those of helping others in need or developing one's own talents. One way in which imperfect duties are unavoidably incomplete is that they cannot be observed towards all others: nobody can help all others, or develop all possible talents. Kant calls these imperfect duties 'duties of virtue' (see Virtues and vices §§2-3).

The derivation of principles of duty from his conception of practical reason is the core of Kant's ethics, and provides the context for his discussion of many other themes. These include: the difference between internalizing principles and merely conforming to them in outward respects ('acting out of duty' versus 'acting according to duty'); the place of happiness in a good life; the need for judgment in moving from principle to act (see Moral Judgment §2); the justification of state power; and the justification of a cosmopolitan account of justice. He also develops the connections between his distinctive conceptions of practical reason and of freedom and his equally distinctive view of religion, which he sees as a matter not of knowledge but of reasoned hope for a future in which morality can be fully realized. In some works Kant articulates reasoned hope in religious terms; in others he articulates it in political and historical terms as a hope for a better this-worldly human future (see Hope §3).

2 Contemporary Kantian ethics

Much contemporary work on ethics is labelled Kantian, in the main because it does not derive an account of right action from one of good results, but rather sees the right as prior to the good (Right and good). In contemporary Kantian work obligations and rights are the fundamental ethical notions. Such work is often called *deontological ethics* (the term derives from the Greek word for *ought*) (see Deontological ethics). Deontological ethical theories are contrasted with teleological or consequentialist theories, which treat the good as prior to the right. Deontological theories are concerned with ethically required action, hence with principles, rules or norms, with obligations, prohibitions and permissions, and with justice and injustice, but not with virtues, good lives, moral ideals and personal relationships (see Fried 1978; Gewirth 1978).

Deontological ethics has many distinct forms. Many versions endorse one or another interpretation of the Kantian demand to respect persons, and think that moral principles should be universal; few mention Kant's minimalist strategy for justifying certain universally binding principles as those we must live by if we reject non-universalizable principles. Indeed, many deontological ethical theories rely on conceptions of freedom, reason and action which are unlike Kant's, and resemble those typically used by consequentialists.

One prominent range of deontological positions seeks to justify principles of justice by showing that they would be agreed to by all concerned under certain hypothetical conditions. They draw on the thought that agreements and contracts are good reasons for action, and suggest that all ethical claims are to be justified by showing that they are based if not on actual then on hypothetical agreements or contracts. These sorts of deontological theories are often called *contractarian* or *contractualist*; they are contemporary versions of social contract theories (see Contractarianism).

Some contractualists take a Hobbesian rather than a Kantian approach. They argue that principles of justice are those on which instrumentally rational persons, guided by their individual preferences, would agree (see Hobbes, T. §§6-7). Other contractualists take a more Kantian approach. They argue that principles of justice are those which would be accepted or agreed to by persons who are not merely instrumentally rational but can use certain reasonable procedures.

The best known exponent of Kantian contractualism is John Rawls, whose *A Theory of Justice* (1971) identifies principles of justice as those that would be agreed by rational and self interested beings in circumstances which ensure that their choosing will be reasonable as well as rational (see Rawls 1980, 1993). He argues that principles of justice would emerge if they were chosen by all concerned in a hypothetical situation devised to ensure impartiality and hence agreement. Rawls calls this hypothetical situation 'the original position', and represents it as one in which persons are ignorant of their own social position and personal attributes, hence of their own

advantage, hence cannot but be impartial.

Rawls claims that rational persons in this hypothetical situation would choose principles of justice that prescribe equal rights for all and the highest attainable level of wellbeing for the worst off. Since everything that differentiates individuals, and could thus provide a basis for disagreement, for bargaining, or for a need to seek agreement, is carefully excluded from the original position, it is not obvious why principles chosen in it should be thought of as matters of *agreement*, or why the parties to the original position should be thought of as *contracting* with one another. Nor is it clear why the fact that certain principles would be agreed to under these conditions justifies those principles to those in other situations. Why are principles which would be agreed to under conditions that do not obtain binding under conditions that actually obtain?

In *A Theory of Justice* Rawls argues that principles that would be so agreed are binding in other situations because they cohere, or form a 'reflective equilibrium' with 'our considered judgments' (see Moral justification §2). Principles are justified not merely because the instrumental reasoning of the hypothetically ignorant would select them, but because we would reasonably judge them congruent with our most carefully considered moral views. In his later *Political Liberalism* (1993) Rawls depicts these principles as the outcome not of hypothetical agreement in an original position, but as the hypothetical agreement of persons who are not only rational but reasonable, in the sense that they are willing to abide by principles given assurance that others will do so too (1993: 49). Principles and institutions are just if they are the focus of reasonable agreement by all concerned.

Jürgen Habermas has also advocated versions of Kantian ethics which stress agreement between agents. In earlier work he argued that the test of justification or legitimation is that a proposal would be agreed in a hypothetical 'ideal speech situation', in which communication was undistorted. In more recent work (1993), he has argued that legitimation of norms is achieved through processes of public discourse, to which each can contribute and in which all can agree.

3 Criticisms of Kantian ethics

Both Kant's ethics and contemporary Kantian ethics have been criticized from many quarters. The critics evidently include those who advocate one or another form of teleological or consequentialist theory, who believe that it is possible to establish an account of the good, from which a convincing account of the right, and specifically of justice, can be derived. However, they also include a variety of writers who reject consequentialist thinking, including communitarians, virtue ethicists, Wittgensteinians and feminist thinkers (see Community and communitarianism; Virtue ethics; Wittgensteinian ethics; Feminist ethics).

The most common and general criticisms are that, because it concentrates on principles or rules, Kantian ethics is doomed to be either empty and formalistic or rigidly uniform in its prescriptions (the complaints cannot both be true). The charge of empty formalism is based on the correct observation that principles underdetermine action; it is usually countered with the equally correct observation that quite indeterminate principles (such as 'Stay within the budget' or 'All religions are to be tolerated') may set significant constraints on action, so are not empty. The charge of rigidly uniform prescriptivity is based on the thought that rules prescribe, so must regiment. It is usually countered by the reminder that since rules can be indeterminate, they need not regiment: universal principles need not be uniformly prescriptive. An ethical theory that applies to principles can be more than empty and less than rigid.

Other critics object that since Kantian ethics focuses on obligations and rights, and in good measure on justice, it either must or does neglect other ethical categories, and in particular the virtues, good character or good lives; that 'natural and human rights…are fictions' (MacIntyre 1981: 67); and that obligations inevitably conflict in ways that render all deontological ethics incoherent. Some critics have laid particular stress on the point that in requiring impartial respect for all, Kantian ethics wholly ignores the place of happiness, of the emotions, of personal integrity and above all of personal relationships in the good life (see Morality and emotions §§2, 4). They have claimed that we must choose between an ethics of justice and one of care, an ethic of rules and one of relationships, an ethic of duty and one of virtue, and that the latter term of each pair is to be preferred.

4 Back to Kant?

Some of these criticisms are accurately aimed at significant features of various forms of contemporary Kantian

writing in ethics; many of them are less apt as criticisms of Kant's ethics. Several recent writers have suggested that Kant's ethics is the most convincing form of Kantian ethics, and that its distinctive features are strengths rather than weaknesses. Many of these writers accept much of the critique of deontological ethics, but think that not all the criticisms apply to Kant's ethics, of which they offer detailed interpretations. Part of their effort has gone into work on Kant's conceptions of action, reason and freedom, and part into work on his ethics. They have pointed out that Kant's account of practical reason and of its vindication does not assume either that all reasoning about action is instrumentally rational pursuit of preferred ends, or that ethical vindication is located in hypothetical agreements or contracts reached by reasonable procedures. They have stressed that Kant's conception of practical reason is based on universalizability rather than impartiality or reciprocity and that he views obligations rather than rights as basic to ethics. They have insisted that impartial respect for persons and a cosmopolitan approach to justice are not morally negligible matters, and have criticized communitarians, virtue ethicists and some feminist thinkers for not taking justice seriously. They have also pointed out that Kant offers accounts of the virtues, of the role of happiness in the good life, and of judgment, and argued that his position is not damagingly individualistic and that he acknowledges the importance of institutions and of social and personal relationships in human life (see Hill 1992; Korsgaard 1996; Herman 1993; O'Neill 1989).

See also: Autonomy, ethical; Kant, I. §§9-11; Practical reason and ethics

ONORA O'NEILL

References and further reading

Fried, C. (1978) *Right and Wrong*, Cambridge, MA: Harvard University Press. (Forthright defence of deontological ethics.)

Gewirth, A. (1978) *Reason and Morality*, Chicago, IL: University of Chicago Press. (Rational justification of deontological ethics.)

Habermas, J. (1993) *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, trans. W. Rehg, Cambridge: Polity Press.(Habermas' most recent extended discussion of questions of normative justification.)

Herman B. (1993) *The Practice of Moral Judgement*, Cambridge, MA: Harvard University Press.(Essays on the use of Kantian ethical reasoning; argues that it can be sensitive to context.)

Hill, T.E., Jr (1992) *Dignity and Practical Reason in Kant's Moral Theory*, Ithaca, NY: Cornell University Press. (Essays on Kant's ethics, and especially on respect for persons.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, in *Kants gesammelte Schriften*, ed. Königlichen Preußischen Akademie der Wissenschaften, Berlin: Reimer, vol. 4, 1903; trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(References made to this work in the entry give the page number from the 1903 Berlin Akademie volume; these page numbers are included in the Paton translation. This is Kant's classic, short, if difficult, introduction to ethics.)

Kant, I. (1788) *Kritik der practischen Vernunft*, trans. L.W. Beck, *Critique of Practical Reason*, New York: Macmillan, Library of Liberal Arts, 1993.(Kant's most abstract account of his ethics; particular stress on reason and the highest good.)

Kant, I. (1797) *Die Metaphysik der Sitten*, trans. M.J. Gregor, *The Metaphysics of Morals*, Cambridge: Cambridge University Press, 1991.(Kant's accounts of justice and of the virtues.)

Korsgaard, C.M. (1996) *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.(Essays on the categorical imperative and its implications.)

MacIntyre, A. (1981) *After Virtue: A Study in Moral Theory*, London: Duckworth. (Vigorous criticism of Kantian ethics from a distinctive neo-Aristotelian viewpoint.)

O'Neill, O. (1989) *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge: Cambridge University Press.(Essays on Kant's vindication of reason and its implications for ethics.)

O'Neill, O. (1996) 'Kant's Virtues', in R. Crisp (ed.) *How Should One Live?: Essays on the Virtues*, Oxford: Oxford University Press, 77-97.(Critical account of Kant on virtue.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Very influential work on liberal political philosophy; many Kantian aspects.)

Rawls, J. (1980) 'Kantian Constructivism in Moral Theory' (The Dewey Lectures), *Journal of Philosophy* 77: 515-72.(Rawls' reconsideration of his theory of justice; discusses the importance of pluralism and the impossibility of justifying a conception of the good to all.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(Reworks Rawls' theory of justice, grounding it in a distinctive conception of public reason.)
Williams, B. (1985) *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press and London: Fontana.(Varied and thoughtful criticism of Kantian ethics.)

Kaplan, Mordecai (1881-1983)

Kaplan argued that Judaism was the evolving religious civilization of the Jewish people. He attempted to recast an inherited faith in rational and natural terms. His advocacy of Jewish communalism, Jewish cultural expression in literature and art, and creativity and experimentation in liturgy had a pronounced impact on wide circles of US Jews. He took seriously the challenge to the traditional interpretation of religious values from science and modernity, and sought to establish a form of religion that was both intellectually respectable and communally responsible. Kaplan's distinctive analysis of Jewish peoplehood and the Jewish religion responsed to a wide variety of philosophical and theological perspectives and challenges.

1 Crisis and challenge

Reconstructionism, the movement in Judaism founded by Mordecai Kaplan, grew out of his perception that the modern era posed a unique threat to the continuity of Judaism and the Jewish people. The corporate political character of the medieval European Jewish community had ceased to exist and Jews no longer constituted an *imperium in imperio*. Instead, with the rise of modern political structures, Jews, like others, were enfranchised as individual citizens of a modern nation-state. Judaism had to adapt to the demands of a democratic national order where the status of the Jews as a discrete political entity was lost.

Kaplan contended that the rise of naturalism undermined a number of foundational beliefs that had sustained Jewish solidarity during the Middle Ages (see Solidarity). Modern men and women, he held, could no longer entertain belief in a supernaturalistic God who would miraculously intervene in the affairs of persons and the processes of nature. Nor could they affirm the idea of otherworldly reward and punishment. The dogma that God had revealed eternal written and oral laws to Moses at Sinai, as well as the belief that the rabbis had received a divine mandate as the legitimate expositors of those laws, were in his view no longer tenable. As a result of the collapse of these beliefs, Kaplan observed, most Jews no longer asserted that God had chosen the Jews from among all peoples for a special relationship, nor did they contend that the observance of Jewish law would allow the individual to attain otherworldly salvation after death.

All this led the Jewish masses to abandon their fidelity to Jewish law. As the strength of the rabbis diminished and the observance of Jewish law faltered, the bond that tied Jews together became increasingly looser. Judaism was endangered because no convincing rationale for justifying the existence of the Jews as a distinct and vital community obtained in the present day. The forces of modern nationalism and naturalism had combined to destroy the structure and rationale that had promoted Jewish group survival for centuries, forcing a crisis of continuity upon the Jewish people.

As Kaplan acknowledged, others had recognized and attempted to respond to this crisis. Yet, in his opinion, all previous attempts had failed. Kaplan condemned both the Reform movement and neo-orthodoxy for overly intellectualizing and spiritualizing Judaism. Each had forsaken Jewish peoplehood and each - in its own way - had consigned Judaism to the status of a religious confession. Nor did the political Zionism of Herzl fare better in Kaplan's analysis (see Zionism). Its truncated vision of the Jewish past was inauthentic, for it made no room for the religious elements of Jewish civilization. Kaplan maintained that only a reconstructionist conception of Judaism, and the course of action that flowed from it, could preserve Judaism and the Jewish people from the impairments brought on by the modern age.

2 The response of reconstructionism

Kaplan's reconstructionism was a mixture of philosophical, sociological and religious elements. He drew eclectically from thinkers as diverse as Matthew Arnold, John Dewey, Emile Durkheim, William James and George Herbert Mead. He also drew upon the ancient rabbis and the contemporary cultural Zionist Ahad Ha'am. His aim was to formulate an ideology that would preserve the identity, unity and continuity of the Jewish people and its religion in the modern world. At the core of his thought lay a social existentialism: the existence of the Jewish people must be set prior to any attempt to define its essence (see Existentialism).

Kaplan's reconstructionism focused on the proposition that Judaism was the creation of the Jewish people and, as such, existed for the people's sake. Underpinning all of his teaching was his conviction that religion, like all

human life, was subject to change. Reflecting his sociological bent, Kaplan contended that the Jewish people, as the authors of their own religious forms, had engaged throughout history in a process of definition and redefinition of these forms. Judaism was conceived as the evolving religious civilization of the Jewish people. The meaning attached to the various rites of Judaism, as well as the concepts of God and Torah, had to be understood and evaluated only in relation to the reality of a living Jewish people. Theology and abstract ideas had a place in Judaism, but that place was a pragmatic one, designed to enhance the life of the group and the individuals who constituted it. In keeping with this agenda, Kaplan set out to redefine Jewish beliefs and rituals in a manner consonant with what he identified as the modern spirit.

This led him to foster a ‘Copernican Revolution’ in Jewish thought, placing the people of Israel, not its God, at the centre of Jewish religious reflection. Kaplan refused to countenance any notion of God as a supernatural being and advanced instead a naturalistic conception of God as that process in the universe which makes for human salvation. He contended that God was that power in the universe, not ourselves, which makes for order, justice and goodness, and which helps men and women to achieve self-fulfilment in this world (see God, concepts of).

Kaplan rejected the idea that God had chosen the people of Israel. For him, the idea of divine election was unacceptable both in its traditional supernatural form and in various modern reformulations. The concept of a chosen people served no functional purpose in contemporary society, for it failed to promote Jewish survival in any meaningful way. Indeed, the exclusivity of the concept raised serious questions about the morality of a God who would display such divine favouritism and was incompatible with modern democratic universalism. Indeed, the idea of chosenness entailed a claim to superiority that could only inflame rivalry and heighten tensions between Jews and their non-Jewish neighbours. Accordingly, Kaplan replaced the idea of a chosen people with that of ‘vocation’, the calling by which every individual and nation seeks to develop and use its powers for the benefit of humanity.

Kaplan affirmed the importance of religious ritual in Jewish life. He recognized that rites and rituals provided for structure and stability in Jewish life (see Ritual). However, he would not countenance rituals as part of a supernaturally revealed code of Jewish law, nor would he envisage them as divine commandments. Instead, he identified Jewish rituals and rites as ‘folkways’ and ‘sancta’, and asserted that they embodied the sacred values of Judaism and served as the cement which bound the Jewish people together across time and space. Nevertheless, in a growing and evolving civilization such as Judaism, old forms at times become obsolete and new forms of religious expression must be developed to take their place. Kaplan, in a famous aphorism, asserted ‘The past has a vote, but not a veto’. Kaplan ceded Zionism and the State of Israel a major place in his thought. Like Ahad Ha’am, he believed that Zionism would serve to unify and awaken the Jewish people to its particular destiny, and he contended that the State of Israel would serve to regenerate the Jewish people in its contemporary state of disintegration and dissolution.

Kaplan believed that his thought would provide a basis for a modern-day expression of Judaism which would foster the establishment of an ‘organic community’ where all the elements that constitute a civilization could play a role in the life of the Jew. He viewed the synagogue as central to this process of community-building, and asserted that it had to function as more than a house of prayer and study. Kaplan therefore set out to ‘reconstruct’ the synagogue to meet the challenge of the day. He proposed that this institution be converted into an all-embracing centre of Jewish cultural and social activity. The synagogue building would become a Jewish centre, complete with swimming pool, gym, library and other facilities, the primary purpose of which would no longer be worship, but ‘social togetherness’. The large number of contemporary synagogues which follow this pattern, either wholly or in part, provides ample testimony to the influence of Kaplan’s ideals on the US Jewish landscape. His programme of Jewish reconstructionism transcended denominational affiliations. While Kaplan had more than his share of theological and philosophical critics, there is no doubt that his death at the age of 102 brought to a close the life of one of the seminal figures in US Jewish religious history.

See also: Ahad Ha’am; Jewish philosophy, contemporary

DAVID ELLENSON

List of works

Kaplan, M.M. (1934) *Judaism as a Civilization: Toward a Reconstruction of American Jewish Life*, New York:

Schocken, 1967.(This book is Kaplan's magnum opus and has been reissued many times. It represents his classical synthesis between Jewish tradition and modern civilization.)

Kaplan, M.M. (1956) *Questions Jews Ask: Reconstructionist Answers*, New York: Reconstructionist Press. (Kaplan here responds to basic questions concerning Judaism from his own reconstructionist perspective.)

Kaplan, M.M. (1958) *Judaism Without Supernaturalism: The Only Alternative to Orthodoxy and Secularism*, New York: Reconstructionist Press.(Indicates how Judaism can be revitalized and enhanced by being freed from the shackles of a supernaturalistic theology.)

Kaplan, M.M. (1960) *The Greater Judaism in the Making: A Study of the Modern Evolution of Judaism*, New York: Reconstructionist Press.(Kaplan explains how Judaism is to respond to the twin challenges of naturalism and nationalism.)

Kaplan, M.M. (1962) *The Meaning of God in Modern Jewish Religion*, New York: Reconstructionist Press. (Argues that only a Judaism devoted to what Kaplan terms a 'this-worldly salvation' can be relevant to the modern Jew.)

Kaplan, M.M. (1964) *The Meaning and Purpose of Jewish Existence*, New York: Reconstructionist Press.(Novel in its devotion to, and presentation of, the thought of the famed German Neo-Kantian philosopher Hermann Cohen.)

Kaplan, M.M. (1970) *The Religion of Ethical Nationhood*, New York: Macmillan.(Sets forth the idea that the only hope for the survival of humanity resides in the creation and ascendancy of a spiritual political leadership devoted to the eradication of humanly created evils.)

References and further reading

Eisen, A.M. (1983) 'Mordecai Kaplan and the New Jewish "Vocation"', in *The Chosen People in America*, Bloomington, IN: Indiana University Press, ch. 4.(A particularly insightful study of Kaplan's thought.)

Eisenstein, I. and Kohn, E. (eds) (1952) *Mordecai M. Kaplan: An Evaluation*, New York: Reconstructionist Press.(A collection of essays by critics and admirers of Kaplan's thought.)

Gurock, J.S. and Schachter, J.J. (1997) *A Modern Heretic and a Traditional Community*, New York: Columbia University Press.(A fascinating study of Kaplan's life and thought from the perspective of his involvement in, and ultimate estrangement from, the US Orthodox community.)

Liebman, C. (1970) 'Reconstructionism in American Jewish Life', *American Jewish Year Book* 71: 3-99. (Remains the best summary overview of Kaplan's thought and the movement it spawned.)

Scult, M. (1994) *Judaism Faces the Twentieth Century: A Biography of Mordecai M. Kaplan*, Detroit, MI: Wayne State University Press.(The definitive and comprehensive intellectual portrait of Kaplan - his life, times and thought.)

Karaism

*The Karaites (*qara'im*, or* benei miqra*) take their name from the Hebrew word for Scripture. The sect's scripturalism originated in its rejection of the 'Oral Law' embodied in rabbinic literature. Like earlier scripturalist groups - notably the Sadducees - Karaites sought to derive their practices directly from the biblical text. While Karaism is usually traced to mid-eighth-century Iraq, the early history remains murky. The sect crystallized in the Islamic East during the late ninth and early tenth centuries, calling forth stern reactions from the leaders of mainstream rabbinic Judaism. Although harsh at times, the ensuing polemics stimulated both Karaite and Rabbanite scholarship in the fields of biblical exegesis, Hebrew grammar and lexicography, jurisprudence and religious philosophy. The two groups differed sharply over points of law and practice - the calendar, dietary laws, Sabbath regulations - but typically concurred on questions of theology.*

1 The Mu'tazilite phase

Early medieval Jewish thinkers of both Rabbanite and Karaite persuasion found the *kalam* (speculative theology) of the Muslim rationalistic school known as the Mu'tazila congenial to their outlook and adopted many Mu'tazilite ideas (see Ash'ariyya and Mu'tazila). The Mu'tazilites' uncompromising definitions of God's unity and justice inform the writings of leading Rabbanites like Saadiah Gaon (d. 942) and Samuel ben Hofni Gaon (d. 1013). During the mid-tenth to mid-twelfth centuries, their Karaite contemporaries also produced works closely modelled upon Mu'tazilite patterns.

Widespread knowledge of Arabic facilitated the appropriation of Islamic theology by Eastern Jewish thinkers. Paralleling the doxographic interests of his Muslim contemporary, the *mutakallim* al-Ash'ari (d. 935), Ya'qub al-Qirqisani (d. after 938) betrays a lively interest in the history of Jewish sects. His *Kitab al-anwar wa'l-marakib (Book of Lights and Watchtowers)*, a comprehensive Karaite code, incorporates discussions of epistemology, philosophy of law, and theology. In this, it resembles the codes of later thinkers active in the Islamic world such as Al-Ghazali (d. 1111) and Maimonides (d. 1204).

While al-Qirqisani lived in Iraq, the main centre of Karaite spiritual and intellectual activity during the tenth and eleventh centuries was Jerusalem. Ascetic and apocalyptic in outlook, the Karaite Mourners for Zion preached a strict regime of repentance, prayer and Bible study. Their writings evince a certain hostility to 'alien wisdom' but do endorse some Mu'tazilite teachings. The commentaries of Japheth ben Eli - originally covering the entire Bible and largely extant in manuscript - include long speculative excursuses and strikingly apply *kalam* doctrines to the explication of such passages as Genesis 1-2, Genesis 22 and Job.

Karaism became truly scholastic with the assimilation of works emanating from the Mu'tazilite school of Basra. The writings of the Mu'tazilite theologian 'Abd al-Jabbar (d. 1025) powerfully influenced such leading Karaites as Yusuf al-Basir (Joseph ben Abraham, d. 1040) and Abu'l-Furqan ben al-Asad (Yeshu'ah ben Judah, d. after 1065). Structured around the doctrines of divine unity and justice, Yusuf's *al-Kitab al-muhtawi (Comprehensive Treatise)* is a compendium of Karaite *kalam*. Its forty chapters grapple with most of the salient issues confronting Jewish (and Muslim) *mutakallimun*. Under the rubric of divine unity we find discussions of epistemology and the obligation to engage in speculative theology (chapters 1-2), physical theories (3), existence of the Creator (4), the divine attributes - omnipotence, omniscience, existence and eternity - and their essential nature (5-11), divine incorporeality and unity (12-14) and the createdness of God's speech (15). Under the second major heading of Mu'tazilite theology, divine justice, we find discussions of God's will and self-sufficiency (chapters 16-17), theodicy (18-26), human free will (27-31), revelation and divine obligation (32-36) and reward, punishment, merits and repentance (37-40).

Like 'Abd al-Jabbar, Yusuf was a jurist. Theology naturally impinged on his juridical writings. In discussing circumcision, for example, he considers such questions as intention, compensation for suffering and the nature of obligation. His disciple Yeshu'ah ben Judah prefaces a treatise on marriage law with an elaborate epistemology. His Bible commentaries show a much greater interest in *kalam* problems than is evident in those of his predecessor Japheth ben Eli.

Like the Mu'tazilites - but unlike the Rabbanites - both Yusuf and Yeshu'ah believed that all bodies are composed

of atoms, in which accidents reside (see Atomism, ancient; Occasionalism). Among Muslims, atomism was typically linked with occasionalism, continuous creation and the rejection of natural causality. The Karaites, however, seem to have affirmed a form of natural causality as a corollary of their theory of creation.

Karaite teachings found a receptive audience among eleventh-century Byzantine Jews, some of whom studied in Jerusalem. They produced the earliest Hebrew translations of Arabic speculative writings over a century before the famous translations of the Ibn Tibbon family in Provence. Despite their odd renderings of technical terms and slavish adherence to Arabic syntax, these Hebrew versions, epitomes and anthologies linked later Byzantine Karaites to the earlier tradition. The transmission process culminated in 1148-9 with the completion of Judah Hadassi's *Eshkol ha-Kofer (Cluster of Henna)*, an encyclopedic code in rhymed Hebrew acrostics. Although more original in form than in content, the work notably includes a creed - a mark of Islamic influence - anticipating Maimonides' Thirteen Principles by a decade. Hadassi's ten articles address: (1) God's existence, unity and wisdom; (2) divine eternity and uniqueness; (3) the world's createdness; (4) the revelations to Moses and other Hebrew prophets; (5) the truth of the Torah; (6) the distinctiveness of the Hebrew language; (7) the unique and eternal sanctity of the Temple site; (8) the resurrection of the dead; (9) divine judgment; and (10) reward and punishment. Most of these articles align with positions developed in Kalam treatises like the *al-Kitab al-muhtawi (Comprehensive Treatise)*.

Numerous harsh, anti-sectarian remarks in the writings of twelfth-century Andalusian Rabbanites indicate the spread of Karaism to Islamic Spain. In an autograph letter, Judah Halevi (d. 1141) reveals that he began his *Kuzari* in response to questions from a Karaite. The third part contains an informed attack on Karaite practice and a staunch defence of rabbinic tradition. In *Sefer ha-Kabbalah (The Book of Tradition)*, Abraham Ibn Daud (d. *circa* 1180) chronicles the history of Rabbanite scholarship in order to prove the validity of the Oral Law. In documenting the heresy of Spanish Karaites, he even mentions how Rabbanite Jews were given license to persecute them. The great exegete Abraham Ibn Ezra polemicizes against numerous Karaite legal positions and records his disputations with the sectarians (for example, in commenting on Leviticus 7: 20). While a vigorous Karaite community clearly existed in twelfth-century Spain, neither the names nor works of Spanish Karaites have survived. And although the Andalusian Rabbanites seem to be implacable foes of the sectarians, Ibn Ezra, at least, regularly cites the commentaries of Japheth ben Eli and Yeshu'ah ben Judah in non-polemical contexts; their philological, rationalistic approach clearly appealed to him.

2 The post-Maimonidean phase

Later Byzantine Karaites similarly admired the clarity, erudition and rationalism of Ibn Ezra and Maimonides. Writing in fluent Rabbinic Hebrew, Aaron ben Joseph (d. *circa* 1320) and Aaron ben Elijah of Nicomedia (d. 1369) engage, criticize and frequently borrow from them and other Rabbanite authors. The latter Aaron is remarkable for his trilogy: *'Ets Hayyim (The Tree of Life)*, a theological *summa*; *Gan Eden (The Garden of Eden)*, a code; and *Keter Torah (The Crown of the Law)*, a commentary on the Pentateuch. Modelled on the *Guide to the Perplexed*, *'Ets Hayyim* is the first substantial Karaite work of religious philosophy composed in Hebrew. Its technical vocabulary is largely Tibbonid, although the older Karaite terminology also persists. Like Maimonides, Aaron clearly deems religious philosophy an authentic and original component of Judaism. But while Maimonides hails Aristotelianism as the soundest speculative system, Aaron clings to the *kalam* of his ancestors. Where Maimonides subjects *kalam* theories to rigorous criticism (*Guide to the Perplexed* 1: 73-6), Aaron defends Mu'tazilite teachings or seeks to harmonize them with Aristotelianism - sometimes through terminological sleights of hand. Consequently, *'Ets Hayyim* is more a *kalam* response to Aristotelianism than a Karaite critique of Rabbanism.

During the fifteenth century, Byzantine Karaites and Rabbanites reached an intellectual rapprochement. Increasingly, Karaite disciples incorporated the scholarship of their Rabbanite teachers into their own compositions. However, while Rabbinic writing began to take new turns, Karaite literary activity became increasingly derivative, and though small sectarian communities still flourished (notably in Poland, Lithuania, the Crimea and Egypt), the great age of Karaite intellectual achievement and innovation had effectively come to an end. Today, Karaites in Israel, Europe and the USA face the challenge of establishing their own communal identities as minorities within minorities. The old philosophical problems that exercised their ancestors are all but forgotten.

See also: Ash‘ariyya and Mu‘tazila; Islamic theology; Theology, Rabbinic

DANIEL FRANK

References and further reading

Ankori, Z. (1959) *Karaites in Byzantium*, New York: Columbia University Press, 1959.(Focuses on the activities of eleventh- and twelfth-century sectarians; comprehensive bibliography on pages 461-84.)

al-Basir, Y. (before 1040) *al-Kitab al-muhtawi (Comprehensive Treatise)*, ed. and trans. G. Vajda, *Al-Kitab al-Muhtawi de Yusuf al-Basir*, Leiden: Brill, 1985.(Reprints articles containing a virtually complete French translation of the text and extensive analyses adducing Jewish and Islamic parallels.)

Ben-Shammai, H. (1985) ‘Studies in Karaite Atomism’, *Jerusalem Studies in Arabic and Islam* 6: 243-97.(Full presentation of this central Kalam problem.)

Frank, D. (1991) ‘The Religious Philosophy of the Karaite Aaron ben Elijah: The Problem of Divine Justice’, Ph.D. dissertation, Harvard University, Ann Arbor, MI: University Microfilms International.(Contains an annotated translation of *‘Ets Hayyim*, chapters 79-90.)

Husik, I. (1940) *A History of Mediaeval Jewish Philosophy*, Philadelphia, PA: Jewish Publication Society of America.(See especially Chapter 4 on Joseph al-Basir and Jeshua ben Judah, and Chapter 16 on Aaron ben Elijah of Nicomedia.)

Lasker, D.J. (1992) ‘Karaism in Twelfth-Century Spain’, *Journal of Jewish Thought and Philosophy* 1: 179-95. (Overview with reference to virtually all known sources.)

Nemoy, L. (1952) *Karaite Anthology*, New Haven, CN: Yale University Press.(Well-chosen collection illustrating seven hundred years of Karaite thought.)

Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Cambridge: Cambridge University Press, 37-56; repr. 1990.(A useful survey with bibliography: see pages 418-21, 459-60.)

Sklare, D.E. (1995) ‘Yusuf al-Basir: Theological Aspects of His Halakhic Works’, in D. Frank (ed.) *The Jews of Medieval Islam: Community, Society, and Identity*, New York and Cologne: Brill, 249-70.(Comprehensive survey of his works and discussion of the interplay between law and *kalam*.)

Wieder, N. (1962) *The Judean Scrolls and Karaism*, London: East and West Library. (Excellent monograph on apocalyptic aspects of sectarian exegesis.)

Wolfson, H.A. (1979) *Repercussions of the Kalam in Jewish Philosophy*, Cambridge, MA: Harvard University Press.(Includes concise analyses of such topics as divine attributes, divine speech and causality as treated by Karaite scholars.)

Karma and rebirth, Indian conceptions of

The combined beliefs in karma and rebirth, that is, the retributive power of actions and decisions and a beginningless, though not necessarily endless, succession of births and deaths for living beings, constitute a fundamental premise of the great majority of India's religious and philosophical traditions. The suggestion first made by the great Muslim scholar al-Biruni (973-1048) that they are the *fundamental creed of Indian religious thought in general may be questionable, but it is certainly understandable. Although such notions are by no means exclusively Indian, they have played a far more central and pervasive role in India than in any other cultural domain.*

In a sense, the idea of karmic retribution postulates that the act itself will hold its originator responsible and accountable. Acts of moral or ritual significance will bring about their own reward or punishment, that is, favourable or unfavourable experiences. On the other hand, favourable or unfavourable experiences and conditions are forms of reward or punishment for past actions and decisions. Karmic retribution takes place through a sequence of countless existences and may involve a movement through a vast variety of forms of life. More specifically, this implies that birth into a particular species, physiological and psychological features, sex, social status, life span, exposure to pleasant or unpleasant experiences, and so on, appear as results of previous actions (usually acts committed in previous lives), and that current actions are expected to have a corresponding influence on future existences. In Sanskrit, the realm of rebirth and karmic retribution is known as saṃsāra. *Its precise scope has been subject to some debate. The most common assumption is that it coincides with sentient existence and includes the entire hierarchy of living organisms from the gods down to the plants. While later Buddhism tends to exclude the plants from this domain, Jainism finds forms of life and sentience even in the elements water, earth, and so forth. Most schools of philosophy view being in* saṃsāra *as a condition of bondage, suffering and alienation; even karmic ascent is ultimately undesirable. The ability to transcend this condition by transforming and eventually eliminating the power of karma is often associated with human existence and considered a rare privilege. Most forms of life are just forms of karmic retribution, without any capacity for karmic initiative.*

The historical origins of the doctrine of karma and rebirth cannot be determined with certainty and precision. While the Vedas and Brāhmaṇas provide significant antecedents, they do not show any clear recognition of the doctrine as such. Even in the older Upaniṣads (prior to 500 BC), its formulations are still tentative, partial and more or less isolated. It seems that the teachings of the Buddha added a new and stricter notion of causality and a far more explicit sense of moral responsibility and universal applicability to the older versions. The other important reform movement of this period, Jainism, showed an early commitment to a systematic elaboration of karmic factors and processes. Unlike the Buddhists, the Jainas developed a reified, even substantialist notion of karma. In Hindu literature, such texts as the great epic the Mahābhārata (beginning around 400 BC) give clear evidence of a fully developed and generally recognized doctrine of karma and rebirth. Subsequently, the doctrine was adopted and variously interpreted by most schools of philosophical and religious thought. It served, moreover, as a basic premise of law texts, popular narratives and mythologies, and a wide array of traditional 'sciences', such as medicine, embryology and astrology. Significant disagreements and debates occurred with regard to the status and character of the karmic agent and the subject of transmigration and rebirth (most conspicuously in connection with the Buddhist denial of a durable 'self' or ātman*). The moral relevance and metaphysical qualities of acts and decisions, the nature of karmic causality and the mechanism of rebirth, the possibility of a transfer of karma, the compatibility of knowledge and action, and the prospects of and problems concerning the elimination of karma and the ultimate transcendence of rebirth provided further topics of debate.*

In its various contexts and applications, the doctrine of karma and saṃsāra *has at least three different yet interrelated functions and dimensions: it is used to provide causal explanations (especially in the realm of life); it serves as a framework for ethical discipline and religious orientation; and it provides the rationale for a fundamental dissatisfaction with worldly existence and a commitment to final liberation from such existence. The ways in which these functions have been balanced or correlated with one another reflect fundamental trends and tensions in the Indian tradition in general.*

1 Terminology

The word ‘karma’, one of the most familiar Indian loan words in colloquial English and other Western languages, is the nominative of the Sanskrit noun *karman*, which is a derivative of the verbal root *kṛ*, ‘to do’, ‘to make’. The literal and primary meanings of *karman* are ‘deed’, ‘work’, ‘action’, ‘act’ (often with ritual connotations), but also ‘object’ in a grammatical sense. The semantic linkage with an inherent retributive power of acts and decisions, or with an accumulation and preservation of their effects, occurs in some Upaniṣads and, much more clearly and explicitly, in early Buddhism (Pāli: *kamma*). A familiar alternative for *karman* in this sense, especially in Hindu literature, is *adṛṣṭa*, ‘unseen’, that is, the ‘invisible’ results of our actions. In closer association with ritual acts, we also find the term *apūrva*. Good karma is often referred to as *dharma* or *puṇya* (‘merit’), bad karma as *adharma* or *pāpa* (‘demerit’, ‘evil’, ‘guilt’). The term *karmavipāka* refers, primarily in Buddhism, to the ‘ripening’ or ‘fruition’ of the karmic potential, while *karmāśaya* refers to the ‘karmic residue’ or the ‘accumulation of karma’.

The English term ‘rebirth’ has its literal equivalent in *punarjanma*, but this does not carry much terminological weight. In the Brāhmaṇas, it is preceded by *punarmṛtyu*, ‘redeath’, ‘recurrent death’, and *punarāvṛtti*, ‘return’ from an otherworldly realm to life on this earth. In Buddhism, which does not recognize a durable self, we find *punabbhava*/*punarbhava*, ‘repeated existence’; more specifically, the Pāli term *paṭisandhi* (Sanskrit: *praṭisaṃdhi*) refers to the causal linkage between successive lives. In the middle Upaniṣads as well as in old Buddhism and Jainism, *saṃsāra*, ‘roaming through’ (successive states or existences), ‘transmigration’, emerges as a fundamental notion. It refers to worldly existence as an aimless recurrence of life and death, and to the world itself as a karmic stage. Karma and rebirth are inseparable in this concept.

As far as English terminology is concerned, it should be noted that ‘rebirth’ is ontologically more neutral than ‘transmigration’, ‘metempsychosis’ or even ‘reincarnation’, which may suggest the existence of a durable entity and would thus seem inappropriate in Buddhism. The common phrase ‘law of karma’ has no terminological equivalent in traditional Indian thought, where there is no use for the modern concept of ‘natural laws’, nor any commitment to their systematic empirical verification. ‘Law of karma’ is not only a terminological innovation, but, in many of its usages, also a conceptual reinterpretation.

2 Sources

References to karma and rebirth are to be found in Indian texts of the most diverse types and genres, from the late Vedic to the modern period. The doctrine appears in popular mythology and philosophical thought as an unquestioned presupposition and a topic of explicit theoretical efforts. But except for Jainism, comprehensive and systematic presentations are rare.

Various suggestions and tentative statements are found in the old Upaniṣads, for instance in Bṛhadāraṇyaka Upaniṣad 6.1- and 4.4, and Chāndogya Upaniṣad 5.1-. The great epic the *Mahābhārata* contains a large number of casual references, but also some relatively systematic and coherent presentations. In the ‘law books’ (Dharmaśāstras), the most coherent and significant survey is provided by the twelfth chapter of the *Manusmṛti*. The Purāṇas and other more popular religious texts add much relevant material (see, for instance, the *Pretakalpa* of the Garuḍa Purāṇa). Among the traditional sciences, the medical texts (Āyurveda) provide the most extensive discussions. Various works compile and summarize older statements on karma and its effects, especially karmic punishments, in the form of digests. Among these, the popular and anonymous *Karmavipākasaṃhitā* and the more scholarly *Madanamahārṇava* (fourteenth century) by Viśveśvara Bhaṭṭa are available in printed editions.

More or less explicitly, the doctrine appears throughout the literature of the classical systems (*darśanas*) of Hindu philosophy; one of the more coherent treatments is found in the Yoga system (see *Yogasūtra* 2.13 with commentaries). Nyāya and Vaiśeṣika add more basic arguments concerning rebirth and the existence of the self. Advaita Vedānta has produced one of the rare philosophical monographs on karma, the *Vijñānadīpikā*, falsely attributed to Śaōkara’s disciple Padmapāda. In a more theological fashion, the numerous schools of ‘sectarian’ (Śaivite and Vaiṣṇava) thought continue the debate.

The basic acceptance of karma and rebirth in Buddhism is obvious in its most ancient documents; its full realization is associated with the enlightenment of the Buddha. However, relatively few texts in the oldest parts of the Theravāda canon, for instance the *Mahā*- and *Cūḷakammavibhaōgasutta* of the Majjhimanikāya, deal with it in a thematic fashion. The Jātaka texts describe the future Buddha’s path through more than five hundred existences. The Abhidhamma/Abhidharma part of the canon produces elaborate conceptual schemes in its analysis of karma

and rebirth. Numerous details were debated among the Buddhist schools and sects; from a Theravāda perspective, these debates are documented in the *Kathāvatthu*. The popular *Milindapañha (Questions of Menander)* pays special attention to reconciling the postulate of personal accountability with the Buddhist 'no-self' doctrine.

For the Sarvāstivāda views on karma, the *Mahāvibhāṣā* and the *Abhidharmakośa* (which also includes Sauntrāntika ideas) are exemplary documents. Vasubandhu's *Karmasiddhiprakaraṇa* combines Sautrāntika and Yogācāra ideas on karma. The spread of Buddhism beyond South Asia also led to a dissemination of the ideas of karma and rebirth. The fact that the basic premises of the doctrine were less familiar in the non-Indian traditions which adopted Buddhism sometimes led to the production of auxiliary or explanatory materials, for instance, the *paralokasiddhi* texts in Tibet, which are supposed to establish the basic facts of rebirth and life after death.

The most elaborate treatises on karma and rebirth are found in Jainism. The oldest extant sources already indicate an extraordinary scholastic commitment to this doctrine. A concise summary of the oldest developments is given in Umāsvāti's *Tattvārthasūtra*, produced around the fourth or fifth centuries CE and accepted as authoritative by the Digambara and Śvetāmbara schools of Jainism. The systematic elaboration of the karma theory continues through a large number of works, including Śivaśarman's *Karmaprakṛti*, the anonymous *Saptatikā*, Candrarṣi's *Pañcasaṃgraha* and a group of nine texts known as Karmagranthas. In addition to the Prakrit and Sanskrit originals of these works, we also find adaptations and commentaries in vernaculars, especially Gujarati.

3 Beginnings and early developments

The absence of any clear evidence for karma and rebirth in the older Vedic documents has led to extensive speculations concerning potential extra-Vedic or 'non-Aryan' origins of the doctrine. The possibility of such origins cannot be excluded. On the other hand, we have no historical documents to support and substantiate it. For the time being, the philologist and historian of ideas will have to work with those materials that *are* available, even if they do not answer the question of the origin, but provide only more or less relevant antecedents and approximations. For this endeavour, it is essential not to presuppose and impose the standards of a unified theory of *saṃsāra*, or of any other kind of systematic theory.

Different ideas about life after death coexist in Vedic thought. Among them we find such beliefs as a reappearance of the dead in the form of animals (birds or snakes), but also as their own descendants or relatives. While this may pave the way for the later concept of rebirth, it does not imply that reappearance or survival after death was taken for granted in Vedic times. Other, less popular, notions suggest that the continuation of existence after death had to be accomplished through specific, primarily ritual, activities and techniques and was by no means accessible to all living beings. Side by side with such ideas we find concepts of cosmic balance and justice (most significantly *ṛta*) which may foreshadow certain aspects of the karma theory.

Against this general Vedic background, some of the oldest Upaniṣads provide us with more specific approximations and anticipations of the later concept of *saṃsāra*. The cyclical explanation of life and death found in the combined doctrines of the 'two paths' and the 'five fires' (Bṛhadāraṇyaka and Chāndogya Upaniṣads) is among the most significant examples. This theory does not assume a direct transition from one living organism to another, but an intermediate residence in a heavenly realm, as determined by the amount of ritual merit accumulated during the preceding lifetime. The cyclical 'way of the fathers' is contrasted with the 'way of the gods', which transcends all cycles, and an enigmatic 'third condition' for creatures without access to the 'two paths' and possibly without any capability of rebirth or permanent survival. Other ancient texts speak of a transfer of ritual merit or 'works' (*karman*) from one person to another, for instance, from a dying father to his son (Kauṣītaki Upaniṣad 2.15). Bṛhadāraṇyaka Upaniṣad 3.2.13 proclaims, in a brief and elusive reference, *karman* as the decisive factor for the continuation of existence after death, but does not give any further details. Other relevant passages suggest various possibilities concerning the nature of acts, their inherent retributive power and the modalities of rebirth. The ritual dimension is supplemented by notions of moral behaviour (such as *caraṇa*; see Chāndogya Upaniṣad 5.10.7) in a wider sense. Concepts of 'desire' (*kāma*) and 'decision' or 'commitment' (*kratu*) appear in some competition with *karman* itself. The identity and nature of the agent, the mechanism of retribution, and the scope and limits of the field of transmigration remain ambiguous. The process is not governed by strict causality; there is always room for chance.

Clearer formulations of what may be called (in the sense of a convenient label) a 'standard version' of karma and

rebirth emerge in the *Mahābhārata* and other texts of the period beginning around 400 BC (for instance, the 'law texts', with their elaborate catalogues of different acts and their karmic compensation). The *Mahābhārata* itself recognizes the two basic postulates which define the 'classical' karma theory: 'deeds' require the appropriate retribution as their effect; and positive or negative experiences and modes of being require a karmic cause. We may assume that the anti-Vedic reform movements of Buddhism and Jainism played a significant role in this consolidation of the doctrine of karma and rebirth.

4 Applications and interpretations in Hindu thought

During the centuries preceding the beginning of the Common Era, the concepts of karma and *saṃsāra* became pervasive, almost universally accepted premises of Indian thought and literature, from religion, philosophy and the traditional sciences to poetry, mythology and popular narratives. Only the materialistic Cārvākas and a few other groups that denied life after death and the moral and retributive implications of actions rejected these concepts (see Materialism, Indian school of). Karma was, moreover, projected into the most ancient layers of the tradition and appeared as the indispensable background and presupposition of the most ancient texts, including the Vedic hymns. Older doctrines and ideas, such as the conceptions of fate (*daiva*) or time (*kāla*) as cosmic powers, were reinterpreted in the light of the karma theory and included in the comprehensive notion of *saṃsāra*. Early developments in some of the philosophical systems (*darśanas*) exemplify and illustrate these general observations.

There is little evidence that karma played any significant role in the cosmology of ancient Sāõkhya, probably the oldest of the classical systems of Hindu philosophy. It was absent from the original conception of the periodic manifestation and unfolding of the basic matter and matrix (*prakṛti*). With the consolidation of the karma theory, the power of deeds and the need for retributive experiences were invoked as a catalyst for the actualization of the cosmic energy of the 'matrix' and thus for the initiation of a new world cycle. However, karma remains insignificant in so far as its ethical and soteriological role is concerned. Actions as such are relegated to the realm of *prakṛti*, that is, to objective 'material' processes. They do not affect the essence of the pure subject or spirit (*puruṣa*). The acting and transmigrating entity itself is a mere product and configuration of *prakṛti* and will ultimately be dissolved in it (see Sāõkhya §§2, 6).

The allied Yoga system of Patañjali offers a more integrated, coherent and explicit explanation of karma and rebirth, which suggests an early interaction with Buddhism. Karma functions within the 'wheel of *saṃsāra*' (*saṃsāracakra*), which is kept in motion by pleasure and pain, desire and aversion, merit and demerit. The production and operation of karma depends on the presence of the five 'afflictions' (*kleśa*); the most fundamental item in this group, which includes desire and aversion, is 'nescience' (*avidyā*), a radical misconception of oneself and the world. As long as this nescience has not been eliminated, the other afflictions will remain in place, and new karma will be produced. As a rule, the karma accumulated during one lifetime will determine the conditions of the immediately following existence, that is, birth in a particular species (*jāti*), life span or 'life-quantum' (*āyus*), and affective potential, or the balance of favourable and unfavourable experiences (*bhoga*). However, this basic rule of 'karmic ripening' (*karmavipāka*) is subject to innumerable qualifications and exceptions, so that the incompatibility with such texts as the *Manusmṛti*, which do not recognize the 'one life' rule at all, may be less severe than it appears.

The treatment of karma and rebirth is somewhat different in Nyāya and Vaiśeṣika. Unlike Sāõkhya and Yoga, both Vaiśeṣika and Nyāya consider the soul or self (*ātman*, *puruṣa*) to be the originator of acts and decisions and, accordingly, as the substrate of karma. In their assessment of the moral quality of acts, they are more committed to the orthodox 'law books' than Sāõkhya and Yoga. Also, they do not advocate the 'one life' rule of karmic ripening.

According to Nyāya, the effects of actions are stored as 'dispositions' (*saṃskāra*) in the soul. Sooner or later, they will initiate the formation of a new organism, which will provide the appropriate karmic retribution. Elaborate arguments are devised to demonstrate the distinctness and immortality of the soul and the indispensability of its stored karma for the explanation of biological and physiological processes. With the general alliance and amalgamation of the two systems, the Nyāya teachings on karma and rebirth were combined with those of Vaiśeṣika and became virtually indistinguishable. However, there are obvious differences in the oldest texts. While old Nyāya shares its use of the term 'karma' with Yoga and Buddhism, Vaiśeṣika prefers *adṛṣṭa* (and assigns a different terminological role to *karman*). *Adṛṣṭa*, the 'unseen' result of our actions, includes 'merit' and 'demerit'

(*dharma* and *adharma*), two of the twenty-four 'qualities' (*guṇa*) that constitute the second ontological category (*padārtha*) of the system. It inheres in the 'souls' or 'selves' (*ātman*), which are conceived as omnipresent spiritual substances. Accordingly, its retributive potential is all-pervasive, and it can influence or direct events anywhere in the world. The preceding exposition follows Praśastapāda's presentation of the Vaiśeṣika system (*c*.500 CE). There are significant and historically symptomatic differences in the older *Vaiśeṣikasūtra*. Here, *adṛṣṭa* is not even listed among the 'qualities' of the soul, and the cosmological role of karma remains generally more ambiguous (see Nyāya-Vaiśeṣika §§4-5, 7).

Because of its commitment to the proto-karmic ritualism of the Brāhmaṇas, the contributions of classical Mīmāṃsāto the karma theory are more limited. However, in its complex and highly technical theory of the Vedic ritual, it offers significant general perspectives concerning the motivation to act, the capacity to act ritually and produce ritual merit (which it restricts, in an exemplary and influential manner, to upper-caste Hindus), the connection between actions and their results, and the formation and structure of complex acts or clusters of action. Kumārila's theory of *apūrva*, the latent power produced by rituals, deserves special attention in this context.

In the school of Advaita Vedānta, as represented by Śaōkara and his followers, the role of karma and rebirth is very significant, but also ambiguous and elusive. On the one hand, Upaniṣadic notions of the recurrence of life and death live on in Advaita Vedānta. On the other hand, later systematic and scholastic views concerning the accumulation and elimination of karma are adopted and developed in a somewhat eclectic fashion. For instance, the division of karma into that which is being accumulated (*sancīyamāna*), that which has been accumulated but may still be neutralized (*sañcita*), and that which has started producing its results (*prārabdha*) is a familiar item. But although karma is a basic premise in Advaita Vedānta, it is ultimately irrelevant. It is confined to the level of conventional, provisional truth (*vyavahāra*). In the end, karma and rebirth have only one meaning and function: to expose the ontological and soteriological deficiency of our world of time and space. Good karma is as vacuous as bad karma. The entire domain of karma and causality is a realm of ignorance and illusion (*avidyā*, *māyā*) and needs to be transcended.

The radical monism and illusionism of Advaita Vedānta was criticized by numerous schools of theistic Vedānta, especially those that flourished among the devotees of Viṣṇu (for instance, Rāmānuja's Viśiṣṭādvaita and Madhva's Dvaita Vedānta). According to these groups, karma was real; but it could be neutralized and superseded through loving devotion to God (*bhakti*) and divine grace, and was thus relegated to a lower level of relevance (see Vedānta §3; Rāmānuja; Madhva). Corresponding materials are found in the rich philosophical and theological traditions of the devotees of Śiva, such as the South Indian Śaiva Siddhānta and the Pratyabhijñāschool of Kashmir.

5 Buddhism

In the oldest extant sources of Buddhism, there is far less ambiguity about karma and rebirth than in the Upaniṣads, not to mention the older Vedic texts. In the legendary accounts of his enlightenment, the Buddha himself (*c*.500 BC) is credited with the discovery of a pervasive causality and with the realization that living beings are causal series of acts and results which extend over countless successive existences. Although the number of older canonical texts which deal with the doctrine in greater detail is relatively small, its central significance cannot be questioned. It is, in fact, inseparable from the quintessential Buddhist formula of 'dependent origination' (*pratītyasamutpāda*; Pāli: *paṭiccasamuppāda*). In the emerging sectarianism, debates on karma play an important role; different conceptions and interpretations of karma are among those factors that define the doctrinal identity of the various schools.

The basic contributions of ancient Buddhism to the development of the karma theory relate, above all, to the following five areas:

1 a stricter notion of causality which postulates a pervasive coherence of karmic events, but insists on the feasibility of choice and responsibility, that is, of acts which are not themselves karmic effects (thus avoiding the karmic determinism associated with the Ājīvika school);
2 a notion of agency which defines the act as rooted in, or even as essentially identical with, volition and decision (*cetanā*) and interprets its vocal or physical implementation as a secondary phenomenon;
3 a process ontology which interprets the connection between act and experienced result strictly and exclusively as

causal continuity and tries to avoid the assumption of an identical subject or substrate of act and result;
4 a comprehensive ethicization which replaces Vedic-Brahmanic notions of ritual correctness and caste-bound aptitude with more open and universal ideas of moral obligation and value;
5 a more radical notion of final liberation (*nirvāṇa*/*nibbāna*) and the commitment to achieve it by eliminating the roots of karmic existence, that is, selfish desire and the illusion of the self.

All these areas leave room for further analysis and debate. What, more precisely, is the nature of the mental act or intention, and how does it relate to speech and physical action? How is karmic causality transmitted, and how does it interact with other causes? Is there an intermediate existence between death and rebirth? How does karma influence the external, material world, and how does it determine the physiological or psychological constitution of sentient beings? What is the special karmic situation of a *buddha* or an *arhat*, whose selfish desire has been eliminated? Such and similar issues are discussed in detail in the Abhidharma texts and the commentarial literature of the Theravāda tradition, and analysed in the context of their complex theories of causal conditions (*pratyaya*/*paccaya*) and the ultimate constituents of reality (*dharma*/*dhamma*). The *Kathāvatthu* defends the Theravāda views against numerous alternatives.

Among the other great schools of Buddhist thought, Sarvāstivāda made significant and influential contributions to the *dharma* theory, the classification of causal factors, and the analysis of the person as a 'series' (*santāna*) of *dharmas*. Its reifying conception of 'acquisition' (*prāpti*) is supposed to explain the way the series is affected by the moral and karmic implications of decisions and actions. It also deals with the relation between mental acts and their vocal and physical manifestations and in general with the question of how latent actions (*avijñapti*) relate to their manifest counterparts (*vijñapti*). To provide a basis for causality and especially for the durable retributive power of acts in an impermanent world, Sarvāstivāda developed its distinctive theory of time, which postulates an irreducible state of non-actual being or subsistence for past and future entities, side by side with the actual existence of present entities (see Buddhism, Ābhidharmika schools of §3).

In the teachings of the Vātsīputrīya-Sāṃmatīya school, we find the peculiar and seemingly 'heterodox' concept of a durable 'person' (*pudgala*), that is, a common basis of act and result. However, this *pudgala* is described, in a somewhat elusive fashion, as neither different from nor identical with the impersonal factors that are supposed to constitute the person. It should not be identified with the Hindu notion of the 'soul' or 'self' (*ātman*).

The Sautrāntika school adds various qualifications to the Sarvāstivāda analysis of volition, action and the transition from the initial impetus to the intended act itself (*cetayitvākarman*). But most significant is its explanation of the process that leads from the act to the retributive result: any intentional act or decision affects the subsequent series of mental events (*cittasantāna*) in a particular fashion and impregnates it with a certain potential. It initiates, in collaboration or competition with other acts, a process of evolution or transformation (*pariṇāma*) within the series which will lead to a mental state of fruition or retribution. It is a kind of mental seed or germ (*bīja*) which is transformed into its appropriate fruit.

The Yogācāra-Vijñānavāda school adopts these and other Sautrāntika ideas and develops them in the context of its 'consciousness only' theory, a subjective idealism which reduces reality to states of awareness or mental occurrences. Reality in this sense, as mental and experienced reality, is *eo ipso* a medium of retribution. As a basis and matrix for the generation of retributive - favourable or unfavourable - experiences, the school postulates a 'warehouse consciousness' (*ālaya-vijñāna*), which contains the potential, or the stored 'seeds', for all actual awareness (*pravṛttivijñāna*) (see Buddhism, Yogācāra school of §§5-8).

The most radical treatment or rather transcendence of karma and rebirth is found in the Mādhyamika school of Nāgārjuna. Here (as centuries later in Śaōkara's Advaita Vedānta), karma and rebirth are relegated to the level of provisional, pragmatic and conventional truth (*saṃvṛti* or *vyavahāra*), to an understanding of oneself and the world which needs to be transcended. Neither the accumulation of good karma nor the attempt to eliminate all karma will lead to *nirvāṇa*. Karma is inseparable from a false commitment to means and ends, to acquisition, ownership and selfhood. From the standpoint of absolute truth, it is as essenceless and 'void' (*śūnya*) as the self or soul (*ātman*) itself. Realizing this voidness is liberation.

In general, the transfer of karmic merit plays a much greater role in the Mahāyāna schools than in old Buddhism; it is a defining characteristic of the *bodhisattva* ideal.

6 Jainism

The ancient movement of Jainism, founded or perhaps reorganized by the Jina Mahāvīra, an older contemporary of the Buddha, is small as far as the number of its followers (3-4 million) is concerned. But in quantity and speculative technicality, its literary contribution to the doctrine of karma and rebirth surpasses those of both Hinduism and Buddhism.

There can be no doubt that some form of the doctrine, probably with a certain degree of technical elaboration, was part of the teachings of the founder himself. However, because of problems of transmission and documentation, it is difficult to determine with precision their nature and contents. We may assume that the following features were present already in the most ancient period of Jaina thought: the conception of karma as a material or quasi-material substance that is attached to the soul (*jīva*) and suppresses its inherent potential; an extremely wide-ranging notion of life and *saṃsāra* which finds souls and the potential for rebirth and retribution not only among humans, animals and plants, but also in such substances as earth and water (though not all souls may have the capacity for final liberation); a strong emphasis on ethical practice and asceticism; a peculiar technical terminology concerning the attraction, accumulation, exclusion and elimination of karmic matter (which was later supplemented, but not replaced by, terms borrowed from the Hindu and Buddhist schools).

A very concise, highly authoritative presentation of the Jaina theory of karma in its developed form is found in Umāsvāti's *Tattvārthasūtra* (perhaps fourth century or even earlier). Numerous later commentaries, digests and systematic treatises, produced by both the Digambara and Śvetāmbara sects, add a vast amount of technical and speculative detail to this version of the doctrine, but they do not question or alter its basic premises. It can be summarized as follows. The basic framework is provided by the division of reality into living and lifeless entities (*jīva* and *ajīva*) and the karmic categories of influx (*āsrava*) of karmic matter into the soul, subsequent bondage (*bandha*), stoppage (*saṃvara*) and expulsion (*nirjarā*) of karma, and final liberation (*mokṣa*). A soul defiled by confusion, passion and selfish desire attracts particles of potentially karmic matter floating in space; and through the 'vibrations' produced by volitional activities, it binds or glues these particles to itself and converts them into actual karma. The activities relate to body, speech or mind, and the 'vibrations' vary accordingly. A soul which is thus associated with karmic matter will inevitably be an embodied soul, and will be unable to extract itself from the cycle of transmigration.

The Jainas pay special attention to preventing the influx of new karmic particles and to expelling karma that has already been accumulated. To achieve these objectives, they invoke a complex set of ethical rules and regulations, which are supplemented by practices of asceticism and meditation. The principle of nonviolence (*ahiṃsā*) constitutes the core of Jaina ethics. Classifications and subdivisions of the types and functions of karma are highly elaborate. Basically, they distinguish 'ruinous' (*ghātin*; Prakrit: *ghāi*), from 'nonruinous' (*aghātin*/*aghāi*) karmas. The first category, which impedes or obscures the internal cognitive and soteriological potential of a soul, comprises 'perception-obscuring', 'knowledge-obscuring', 'affective' (*vedanīya*) and 'confusing' karmas. The second group relates to the conditions and circumstances of a soul's embodiment; it includes those karmas which account for certain basic limitations of the soul's potential, membership of a particular species and physiological setup (*nāman*), social position (*gotra*/*goya*) and life span (*āyus*/*āuya*). Numerous subdivisions and specifications supplement these general categories.

The proliferation of speculative and scholastic details is a most conspicuous feature of the Jaina literature on karma and rebirth. But there are also other developments, such as the tradition of more basic philosophical arguments concerning the existence of the soul (as the condition of the possibility of rebirth) which is affiliated with Jinabhadra's *Viśeṣāvaśyakabhāṣya* (about sixth century). Even more significantly, we also find an orientation which tends to relegate the entire domain of karma and rebirth, and of causal interaction between soul and matter, to a level of conventional and provisional truth which needs to be transcended. The *Samayasāra* of Kundakunda, possibly an older contemporary of Umāsvāti and one of the most authoritative teachers in the Digambara tradition, exemplifies this orientation, which combines a traditional Jaina perspectivism with the Mādhyamika Buddhist notion of 'two truths'.

7 Applications and reinterpretations in modern Indian thought

With the dissemination of Indian ideas beyond South Asia, chiefly through the spread of Buddhism, the doctrine of

karma and rebirth was adopted by such traditions as China, Japan and Tibet and adjusted to various cultural contexts and presuppositions. These developments had no visible repercussions in India itself. Likewise, the encounter with those foreign religions that had their own presence in India and were generally opposed to karma and rebirth, above all Islam and later on Christianity, did not lead to any explicit defence or reconsideration of the doctrine. However, the Indian attitude to the foreign, especially Western, world began to change during the early nineteenth century, which can be described as the beginning of modern Indian thought. Western ideas gained an increasing significance for the Indian self-understanding. Traditional concepts and doctrines were reinterpreted in the light of modern Western thought; quite naturally, this included revisions and transformations in the theory of karma and rebirth. Western interpretations and reinterpretations of karma, for instance those by the theosophists and anthroposophists, contributed to and interacted with these revisions (see Theosophy §2).

Exemplary modernist statements on karma and rebirth may be found in the works of such Neo-Hindu thinkers as Vivekananda (1863-1902), Aurobindo Ghose (1872-1950) and Radhakrishnan (1888-1975). Numerous authors continue the debate in contemporary India, both in English and in Indian vernaculars. The most distinctive features of the modernist interpretation can be described as follows:

1 Karma is, more or less radically, dissociated from the traditional mythological implications of *saṃsāra*, which include heavens, hells and other transempirical realms of existence; rebirth itself is treated as a less essential adjunct of karma.
2 Karma is presented as a fundamentally scientific notion, a comprehensive 'law' and principle of explanation, which supersedes all merely physical causality and regularity.
3 There is a stronger commitment to empirical evidence, to 'case studies', to the collection and analysis of 'reports' and personal claims concerning rebirth; research in this sense is foreign to the traditional treatment of karma and rebirth.
4 The doctrine is associated with modern Western concepts of evolution and progress; the world of karma and rebirth appears not so much as the realm of aimless wandering (*saṃsāra*) which calls for transcendence and ultimate liberation (*mokṣa*), but rather as a sphere of potential self-perfection and spiritual growth.
5 In response to European criticism, any fatalistic implications of karma are strongly and passionately rejected, and its compatibility with action, initiative and social responsibility is emphasized.
6 The notions of 'collective karma', 'group karma' or even 'national karma', which have no place in traditional thought, but seem to be taken for granted in theosophy, emerge in Neo-Hindu thought and discourse, although their uses are somewhat elusive and in some cases merely rhetorical.

Such concepts appear also in modern Buddhism, together with other reinterpretations and transformations of the traditional concepts. An important trend is exemplified by the work of the modern Theravādin M.W.P. de Silva, for whom the theory of karma is primarily a theory of psychological and characterological development, and who de-emphasizes its 'judicial' or retributive implications. Even more radical forms of 'demythologizing' karma and rebirth may be found in modern Japanese Buddhism. In both Buddhism and Hinduism, reinterpretations coexist with more traditional and traditionalistic versions. There has even been a certain resurgence of mythical and esoteric ideas concerning intermediate stages and the like, most conspicuously among Western proponents of Hindu or Buddhist traditions.

8 General and concluding observations

Karma and rebirth are among the most important regulative ideas in the history of Indian thought. But there has never been one identical theory (and certainly no scientific theory) of karma to which a majority of thinkers could have agreed. The expression 'the doctrine of karma and rebirth' is only a convenient label. To be sure, there are certain shared premises - above all, the twin postulates that there should be no undeserved suffering or wellbeing and that no effect of a past deed should be lost. But apart from such rather formal premises (which may be derived from the even more fundamental postulate that justice should be inherent in the universe), there has been a wide variety of interpretations and manifestations and much room for debate. The twin postulates themselves may be applied more or less strictly. In fact, their rigid implementation would hardly be compatible with the element of freedom and initiative which is implied in the notion of karmic action.

Furthermore, the familiar association and correlation of karma and rebirth is itself problematic. While the two

concepts have become virtually inseparable in the idea of *saṃsāra*, their historical roots may, in fact, be quite different. There are also significant conceptual differences between karma and rebirth, and their relation is not a symmetrical one. While karma requires rebirth as the condition of its own inescapability, rebirth does not require karma. It could just be a merely factual continuation of existence, without subjection to retributive justice. In this sense, and in spite of its wide-ranging mythical ramifications, the concept of rebirth would, in principle at least, be more easily compatible with a 'scientific' worldview which understands the universe in terms of facts, not values or judicial standards. A world in which rebirth had a place could still be our 'natural' world, the domain of science, while a world in which karma operated would have to be a structurally different universe. Explanation in such a universe would involve justification in the sense of 'theodicy' or rather 'cosmodicy'.

However, karmic causality does not always imply a judicial order of deeds and corresponding rewards or punishments. There is also an ancient and recurrent tendency to explain the functioning of karma as a natural sequence of mental states and events, a psychological, not retributory or quasi-legal, process which links acts and decisions with the formation of dispositions, instincts, character attributes or other internal modes of being. This may, indeed, be the case in some of the oldest statements on karma, which appear in the Bṛhadāraṇyaka Upaniṣad (3.2.13; 4.4.5). Here, Yājñavalkya does not propose reward and punishment (that is, pleasant and painful experience) as the result of karma, but the formation of a good or bad person. Similar ideas are documented in ancient Buddhism. Most later schools distinguish the formation of dispositions, especially of mental 'defilements' (*kleśa*, *doṣa*, *anuśaya*) such as attachment and hate, from the accumulation of retributive karmic potential (*karmāśaya*). The 'defilements' are often presented as a condition for the perpetuation of karma; in some texts (for instance, Vasubandhu's *Abhidharmakośabhāṣya*), the fundamental disposition of existential craving (*tṛṣṇā*) appears as the cause of rebirth *per se*, while karma is said to regulate its specific circumstances. But there has also been much overlap and ambiguity (for instance, in the Jaina notion of 'ruinous' karmas).

What is an act in a karmically relevant sense? What is the significance of intention? What distinguishes good and bad deeds? How does an act produce its 'unseen' power (*adṛṣṭa*), and what is the ontological status of this power? Who or what acts? What connects the karmic agent and the subject of retribution? What is the scope and nature of karmic causality? How does it affect the mental and the physical realm? How do we know about karma and rebirth? Do we have valid arguments for their existence, or is it a matter of authoritative tradition and superhuman modes of awareness (such as 'yogic perception')? Are there basically different types of karma, and how can they be classified?

Our preceding survey has shown a wide variety of answers to these and similar questions. But this multitude of theoretical answers does not describe the breadth and complexity of the phenomenon in its entirety. In the history of Indian thought and life, karma has functioned at various levels of understanding and interpretation, as an unquestioned presupposition and a topic of theoretical inquiry, in popular mythology and in philosophical thought. In its various contexts and applications, it has at least three clearly separable, but interrelated, functions and dimensions:

1 It provides causal explanations of factual occurrences and correlates the present with the past (for instance, in traditional medical literature).
2 It provides perspectives on and incentives for actions and decisions and correlates the present with the future (most conspicuously in the normative Dharmaśāstra literature).
3 It provides a soteriological point of departure, a view of the causal and temporal world which calls for detachment, transcendence and final liberation.

In the philosophical traditions of Hinduism, Buddhism and Jainism we find all of these meanings of karma. But the third one clearly overshadows the other two, and it accounts for the pervasive sense of soteriological commitment in Indian philosophical thought.

See also: Buddhist philosophy, Chinese; Buddhist philosophy, Japanese §6; Duty and virtue, Indian conceptions of; Hindu philosophy §5; Jaina philosophy §§1, 3-4

WILHELM HALBFASS

References and further reading

Collins, S. (1982) *Selfless Persons*, Cambridge: Cambridge University Press.(A clear and attractive exposition of

the Buddhist ‘no-self’ doctrine and its implications for karma and rebirth.)

Glasenapp, H. von (1915) *The Doctrine of Karman in Jain Philosophy*, trans. G.B. Gifford, Bombay: Bai Vijibai Jivanlal Panalal Charity Fund, 1942.(A good survey of texts and teachings, but no philosophical analysis; translated from a German original.)

Horsch, P. (1971) ‘Vorstufen der indischen Seelenwanderungslehre’ (Early Stages of the Indian Theory of the Transmigration of Souls), *Asiatische Studien* 25: 99-157.(A survey of Vedic antecedents of karma and rebirth.)

Jaini, P.S. (1979) *The Jaina Path of Purification*, Berkeley, CA: University of California Press.(Discusses karma and rebirth in the general context of Jainism.)

Krishan, Y. (1983) ‘Karma Vipāka’, *Numen* 30: 199-214.(A useful collection of quotes and references concerning the notion of ‘karmic ripening’.)

Lamotte, É. (1936) ‘Le traité de l’acte de Vasubandhu: *Karmasiddhiprakaraṇa*’, *Mélanges chinois et bouddhiques* 4: 151-288; trans. L.M. Pruden, *Karmasiddhiprakaraṇa: The Treatise on Action by Vasubandhu*, Berkeley, CA: Asian Humanities Press, 1988.(Contains a valuable introduction concerning different Buddhist theories of karma.)

McDermott, J.P. (1984) *Development in the Early Buddhist Concept of Kamma/Karma*, New Delhi: Munshiram Manoharlal.(A survey and analysis of relevant debates in Theravāda literature and Vasubandhu’s *Abhidharmakośa*.)

Neufeldt, R.W. (ed.) (1986) *Karma and Rebirth: Post-Classical Developments*, Albany, NY: State University of New York Press.(A collection of articles discussing reinterpretations of the doctrine in modern Hindu, Buddhist and Western thought.)

Oetke, C. (1988) *‘Ich’ und das Ich: Analytische Untersuchungen zur buddhistisch-brahmanischen Ātmankontroverse (‘I’ and the Self: Analytical Investigations into the Buddhist-Brahmanic Controversy about Ātman)*, Stuttgart: Franz Steiner Verlag.(A pioneering analytical study of the Hindu and Buddhist arguments concerning personal identity, the existence of the self, and the meaning of rebirth.)

O’Flaherty, W.D. (ed.) (1980) *Karma and Rebirth in Classical Indian Traditions*, Berkeley, CA: University of California Press.(Twelve representative essays on karma in traditional Indian thought; a revised and updated version of the contribution by W. Halbfass was published in his *Tradition and Reflection*, Albany, NY: State University of New York Press, 1991, 291-345.)

Patañjali (*c*.2nd century BC) *Yoga Sūtras* with the commentary of Vyāsa and the gloss of Vachaspati, trans. R. Prasāda, New Delhi: Oriental Books, 1978.(A standard and readily available translation.)

Prasad, R. (1989) *Karma, Causality and Retributive Morality*, New Delhi: Indian Council of Philosophical Research.(A discussion of ethical and meta-ethical issues; see especially chapter 13.)

Reichenbach, B.R. (1990) *The Law of Karma: A Philosophical Study*, Honolulu, HI: University of Hawaii Press.(A clear presentation of major philosophical issues, but on a limited textual and philological basis.)

Schmithausen, L. (1991) *The Problem of the Sentience of Plants in Earliest Buddhism*, Tokyo: The International Institute for Buddhist Studies.(A penetrating study of early Buddhist ideas concerning life and *saṃsāra*.)

Schmithausen, L. (1995) ‘Mensch, Tier und Pflanze und der Tod in den älteren Upaniṣaden’ (Man, Animal, Plant and Death in the Earlier Upaniṣads), in G. Oberhammer (ed.) *Im Tod gewinnt der Mensch sein Selbst (In Death Man Wins Himself)*, Vienna: Akademie der Wissenschaften, 43-74.(A careful examination of the beginnings and early developments of the idea of *saṃsāra*; on pages 75-96 of the same volume, W. Halbfass discusses the relation between karma and death.)

Steinkellner, E. (1985) ‘Paralokasiddhi Texts’, *Buddhism and its Relation to Other Religions: Essays in Honour of Dr Shozen Kumoi on His Seventieth Birthday*, Kyoto: 215-22.(On the Indo-Tibetan tradition of proofs for rebirth.)

Tull, H.W. (1989) *The Vedic Origins of Karma*, Albany, NY: State University of New York Press.(This book contains some interesting suggestions, but has to be used with caution; see the review by L. Schmithausen, *Indo-Iranian Journal* 37: 151-58 (1994).)

Viśveśvara Bhaṭṭa (14th century) *Madanamahārṇava*, E. Krishnamacharya and N.M. Nambiyar (eds), Baroda: Oriental Institute, 1953.(A detailed presentation of karmic punishments, especially diseases.)

Katharsis

One of the central concepts of Aristotle's Poetics, katharsis *('purgation' or 'purification'; often spelled* catharsis*) defines the goal of the tragic poet: by depiction of human vicissitudes so to provoke the spectators' feelings of pity and fear that such emotions in them are finally purged.*

Aristotle's doctrine of *katharsis* is a response to a familiar problem in aesthetics (see Aristotle §29): why do audiences take pleasure in watching tragedies, with their depictions of terrible crimes and sufferings? Plato, who divided the human psychological system into mutually antagonistic rational and irrational components, had a ready answer: he argued that the irrational impulses enjoyed receiving stimulation and nurture from the opportunity tragedy provided them to indulge themselves in such emotions as pity and fear, which were normally kept under tight rein but which thereby became harder to control. This was one reason Plato felt obliged to exclude the tragic poets from his ideal city (see Plato §14).

Against such an accusation, Aristotle merely implied in a single sentence a defence of poetry in terms of *katharsis*: his definition of tragedy in chapter 6 of his *Poetics* culminates in the words 'by means of pity and fear performing the *katharsis* of emotions of this sort'. Because Aristotle never fully explains or justifies this phrase, the term *katharsis* has proven to be endlessly controversial - and also extremely fruitful for all European literary theory since the Renaissance. Scholars are divided on whether *katharsis* means 'purgation' or 'purification', whether the term is derived from medicine or religion, whose emotions undergo the process described, and whether it is moral or psychological, temporary or permanent. The following reconstruction is necessarily tentative.

According to Aristotle, the ultimate goal of every poet is to produce in the audience a kind of pleasure which is peculiar to the kind of poetry in question; all of the constituents of the poem are means towards that end, though not all contribute to the same degree. The production of the kind of pleasure peculiar to tragedy, involving pity, fear and the *katharsis* of such emotions, is the goal to which all the constitutive elements of a tragedy - above all the structure of its plot - are subordinated. Pity is our reasonable response to the sight of someone's undeserved suffering and fear implies our judgment of a similarity between the sufferer and us such that the same kind of suffering could befall us. To provoke these responses, the tragic poet must deploy the appropriate means: a central hero whose fate interests us strongly, who suffers (hence the better tragedies move from happiness to suffering), who is neither too evil (for then his suffering would not seem undeserved) nor too good (for then we would not consider him similar to us), and whose suffering is the result not of his culpability, nor of the gods or fate, which Aristotle ignores, but of some error of judgment he makes.

For Aristotle, pity and fear are not pleasures, but kinds of pain. If tragedy is to procure pleasure, it cannot do so by 'purifying' or 'refining' emotions such as pity and fear - a purer pain is not a greater pleasure - but only by 'purging' or 'discharging' them. In medicine, normally healthy fluids such as bile or blood can put the body at risk when they exceed a certain level; when this happens, the body restores its endangered equilibrium by a purging of these substances. The sufferings in any tragedy provoke pity and fear. But the best kind of tragedy astonishes its viewers by what Aristotle calls *peripeteia*, a sudden change in the expected outcome such that events turn out otherwise than anticipated but still retain a logical concatenation which we recognize afterwards. This shock enhances the viewers' pity and fear so strongly that these emotions and similar states of disturbance end up discharging themselves; as a result, the viewers are left free of them and are restored to their normal state after having been temporarily upset. They enjoy both the restorative pleasure that derives from liberation from pain and the intellectual pleasure that accompanies their delayed recognition of the logical concatenation of the events of the plot.

On this view of *katharsis*, tragedy does not permanently make us better people, refine our emotions or provide a profound insight into the tragic nature of the human condition. Aristotle is simply countering Plato's attack upon tragedy by agreeing with him that tragedy does provoke pity and fear and that such emotions are indeed unpleasant, but objecting that they are not entirely irrational after all and that the effect of the best tragedies is first to provoke them, then to enhance them, but finally to free us of them.

See also: Emotion in response to art; Mimēsis; Tragedy

GLENN W. MOST

References and further reading

Aristotle (*c*. mid 4th century BC) *Poetics*, trans. S. Halliwell, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1995.(Greek text with an English prose translation on facing pages.)

Belfiore, E.S. (1992) *Tragic Pleasures: Aristotle on Plot and Emotion*, Princeton, NJ: Princeton University Press, esp. 257-360.(A detailed defence of the medical interpretation of *katharsis*.)

Bernays, J. (1857) *Grundzüge der verlorenen Abhandlung des Aristoteles über Wirkung der Tragödie (Features of Aristotle's lost treatise on th eeffect of tragedy)*, Breslau: Trewendt; repr. Hildesheim: Olms, 1970; partially trans. in J. Barnes, M. Schofield and R. Sorabji (eds) *Articles on Aristotle*, vol. 4, *Psychology and Aesthetics*, London: Duckworth, 1979, 154-65.(The single most important and influential study of *katharsis* ever written; the first to argue cogently for a medical interpretation of the term.)

Else, G.F. (1957) *Aristotle's Poetics: The Argument*, Cambridge, MA: Harvard University Press.(A provocative and original critique of views still widely shared.)

Else, G.F. (1986) *Plato and Aristotle on Poetry*, ed. P. Burian, Chapel Hill, NC, and London: University of North Carolina Press.(A more generally accessible presentation of some of the arguments in Else (1957).)

Golden, L. (1992) *Aristotle on Tragic and Comic Mimēsis*, Atlanta, GA: Scholars Press.(An eloquent defence of a more cognitive view of *katharsis* than the one presented here.)

Halliwell, S. (1986) *Aristotle's Poetics*, London: Duckworth.(A general introduction to the treatise with a full treatment of *katharsis*, pages 168-201.)

Lucas, D.W. (ed.) (1968) *Aristotle: Poetics*, Oxford: Clarendon Press, 273-90.(The standard English commentary, with a good discussion of *katharsis*.)

Plato (*c*.370s BC) *Republic*, trans. G.M.A. Grube, Indianapolis, IN: Hackett, 1974.(See especially books III and X for Plato's treatment of tragedy.)

Rorty, A.O. (ed.) (1992) *Essays on Aristotle's Poetics*, Princeton, NJ: Princeton University Press.(A stimulating collection of recent articles on *katharsis* and many related issues; good bibliography.)

Kauṭilya (*fl. c.*321-*c.*296 BC)

Kauṭilya is famous as the author of the Arthaśāstra, a political treatise often compared with Machiavelli's The Prince. Although its influence on subsequent political and literary writers is noteworthy, tradition has remained somewhat ambivalent about it, especially because of its seemingly ruthless prescriptions for efficacious government. On a closer reading, however, Kauṭilya is assiduously concerned to secure the welfare and wealth of the citizens of a state under a just government, of which the king, although the sovereign, is just one among seven institutes. Upon the king falls the duty of safeguarding the good of the people in as dharma*-sanctioned a way as possible; the 'rule of the rod', intrigues and stratagems are reserved for combatting internal and external threats.*

Kauṭilya did not leave behind any information about himself, but it is generally believed that he was a minister in the kingdom of Maurya Candragupta, who ascended to power after defeating the Nandas in Magadha country. Since Candragupta promulgated the Gupta Empire *circa* 320 BC, Kauṭilya (which is a nickname; his real name was Cāṇakya) is believed to have lived and written the *Arthaśāstra*, the work for which he is famed, around 321-296 BC. A number of Western scholars have, however, proffered a much later date, closer to the early years of the Common Era or even later.

The *Arthaśāstra* is a comprehensive treatise on statecraft that gives cohesion to the political thinking of Kauṭilya's predecessors. It is tersely written, and organized into fifteen books comprising 150 chapters. Like most classical Indian writings on politics, it is mostly instructional in content and lacks theoretical depth on questions such as the ideal state, the origin of the state, and so on. Kauṭilya gives no arguments to ground his political precepts, or to explain why he makes the pursuit of wealth the principal end. He emphasizes the significance of sensual pleasure (for both king and people), and upholds *dharma* (law) as the beacon of socially ordered conduct; but *mokṣa* (spiritual enlightenment and escape from the material world) warrants no place in the *Arthaśāstra*. Kauṭilya's apparent advocacy of the 'rule of the rod' has traditionally been controversial; nevertheless certain of his important teachings were appropriated by the Dharmaśāstras (for centuries the basic legal/moral canons deferred to in India) and, on a close reading, he is mainly concerned to guarantee the citizens' wellbeing. Kauṭilya's sovereign king is just one of seven institutes of the ideal body politic; the others are the ministers, the territory and its subjects (*janapada*), the fort, the treasury, the rod-bearing army, and strategic allies.

The *Arthaśāstra* creates a secular ethos, but religious life and wisdom are given prominence by the appointment of the *purohita*, a Brahmanical priest-chancellor who is to be consulted by the monarch at every decisive step in the affairs of the state. Kauṭilya also describes laws restricting forms of marriage, the urban location of the major caste-groups, the duties of caste-groups, and inter-caste mixing. But beyond this, he cannot be said to be an advocate for the sort of inward-looking Hindu state that has been a feature of post-independence Indian political thought. To be sure, the idea of a nation-state, loosely federated or unified, or even of a representative democratic state, never took root in pre-modern Indian political thinking.

Kauṭilya is quite conscious of the diversity from ancient days of the Indian regions and accordingly allows for a degree of flexibility in matters of law and justice. The king is expected to attend each morning to pleas and petitions from subjects who may come from all walks of life and from different castes or regions. When meting out justice, the king or the state is not in a position to make laws; rather, the sovereign court's jurisdiction is to negotiate between (1) *dharma*, (2) custom or settled community law, and (3) commercial and personal transactions and written edicts. The king may overrule the latter two sources of law, but he cannot put himself above *dharma*, in accordance with which all disputes and contradictory judgments are to be decided (*Arthaśāstra* 3.1.40-4). This precept entails that the king should maintain detailed codes of law, judge each case on its legal merit and mete out punishment proportionate to the offence, not in whimsical excess. The king's ministers, the *purohita*, the ascetics, the queen and prince, the gods and, above all, *dharma* are a check to any possible deviation. (Note that these are broadly ancillary to the principal institutes: the queen is married to the king; the *purohita*, independently of the ministers, counsels the king; ascetics and leading citizens echo the territory's subjects; the gods are the heavenly guards of the fort and the treasury; and *dharma* is the impersonal rule and codes of transcendental ethics symbolized in the rod.) Kauṭilya is credited with having been among the first to set down codes of law, as distinct from listing desirable prescriptions and customary rules regardless of their moral or philosophical merits.

Kauṭilya accorded an unusually important role to philosophy in the training of the king. He called this art *anvīksīki*, understood as the art of reasoning and learning in traditional disciplines (notably Sāōkhya, Yoga and Lokāyata), alongside the Vedas, economics and politics. The king is expected to maintain a regime of study of this material throughout his life. The study of reasoning is undertaken more for its applications in matters of law than for its employment in speculative philosophy (*tarka*). Despite this positive register for philosophy, Kauṭilya does not employ philosophy *per se* in thinking through moral issues and problems in politics. Perhaps the idealistic tendencies of philosophy in the Upaniṣadic tradition made him cautious about philosophy's true worth in the more realistic pursuits of state-making. A life of detached asceticism is not worthy of the philosopher, whose life of practical activity is accomplished if it culminates in providing further instruments by which the state can ensure the protection, prosperity and sovereign stability of the people; there is no state without happy people.

Following a comprehensive treatment of topics connected with internal administration, Kauṭilya devotes a substantial discussion to foreign policy and diplomacy. The king, through his emissaries and envoys, should maintain harmonious relations with neighbouring kingdoms; however, he should be vigilant and alert to possible threats to his kingdom. To this end he is advised to deploy spies and secret agents disguised as ascetics, who, through magical spells and propaganda can create dissension among neighbouring populations, and eventually insurgency. In case an enemy should prove to be powerful, the king should enlist the help and sanction of other neighbours, in a roulette of diplomacy common in international relations anywhere.

See also: Political philosophy, Indian

PURUSHOTTAMA BILIMORIA

List of works

Kauṭilya (*c*.320 BC) *Arthaśāstra*, ed. and trans. R. Shamasastry, *Arthaśāstra of Kauṭilya*, Mysore: University of Mysore Oriental Library, 1915; 3rd edn, 1929; repr. as *Kauṭilya's Arthaśāstra*, Mysore: Mysore Printing and Publishing House, 1967.(The *Arthaśāstra* first came to the notice of Dr R. Shamasastry, a curator with the Government Oriental Library in Mysore, when he unexpectedly received a palm-leaf manuscript; this he began painstakingly to edit, translate and publish in fragments from 1905, culminating in the first ever English work on Kauṭilya in 1915.)

Kauṭilya (*c*.320 BC) *Arthaśāstra*, ed. R.P. Kangle, Bombay: University of Bombay Press, 1960, 1965; repr. Delhi: Motilal Banarsidass, 1986.(This critical edition of the Sanskrit text takes cognizance of commentaries in various vernaculars and editions issued by European Indologists that have appeared since the work's first discovery.)

Kauṭilya (*c*.320 BC) *Arthaśāstra*, ed. and trans. R.P. Kangle, *The Kauṭiliya Arthaśāstra*, Delhi: Motilal Banarsidass, 3 vols, 1986, 1988, 1992.(This erudite English translation in the idiom of Enlightened Western political philosophy is known for utilizing the text of a translation by the Russian Indologist Kalyanov (Moscow 1959), lending at once further clarity but also some confusion over the difficult passages in Kauṭilya's treatise. Part I presents the Sanskrit text, Part II is an English translation, and Part III is a study of the work.)

Kauṭilya (*c*.320 BC) *Arthaśāstra*, ed. and trans. L.N. Rangarajan, *The Arthaśāstras*, New Delhi: Penguin Books India, 1992. (This is a substantial reworking of Kangle's and Shamasastry's editions and translations in the style of a modern treatise on political economy, and has an informative introduction.)

References and further reading

Breloer, B. (1927, 1928) *Kauṭaliya Studien*, Bonn, 2 vols.(This voluminous study is a good example of early European/German interest in the discovery of the *Arthaśāstra*, which was seen as one of the earliest systematic Indo-Iranian treatises on the state.)

Krishna Rao, M.V. (1953) *Studies in Kauṭilya*, Delhi: Munshiram Manoharlal, 1979.(An early attempt at a philosophical treatment of Kauṭilya, which seeks parallels with Aristotle on conceptions of the republic and statecraft in ancient political thought.)

Radhakrishna, C. (1971) *Kauṭilya's Political Ideas and Institutions*, Varanasi: Chowkhamba Sanskrit Series Office.(This Marxian rendering of the *Arthaśāstra* locates Kauṭilya in the history of Indian political thought, and is inspired by the Indian Marxist critic D.D. Kosambi.)

Winternitz, M. (1920) *Some Problems of Indian Literature*, Delhi: Motilal Banarsidass, 1969.(Useful for discussions of dating and linkages with other texts on ancient Indian polity. Although Winternitz is an unusually reliable and balanced Indologist with no Aryan axe to grind, this work is limited by the fact that the

definitive work on the Dharmaśāstras had not been done at the time of writing.)

Kautsky, Karl Johann (1854-1938)

Karl Johann Kautsky became the leading German socialist theoretician at the time of the Second International and the authoritative exponent of 'orthodox Marxism'. He was a close associate of Engels, editor of Die Neue Zeit from 1883 until 1917, and author of numerous political, historical and theoretical works. He clashed with Bernstein in the 'revisionist debate' in the late 1890s and with Lenin during the Russian Revolution.

1 Beginnings

Karl Johann Kautsky was born in Prague. His family moved to Vienna where he went to school (the Benedictines at Melk) and then to university (although he dropped out before finishing his degree). The rise and fall of the Paris Commune helped steer his early romantic radicalism in the direction of socialism. At this time, he also began to take an interest in Darwinist evolutionary theory and in the application of science to the study of human societies (Haeckel and Büchner being particularly influential). In 1875 he joined the nascent Austrian Social Democratic Workers' Party and, over the next few years, he wrote extensively for various socialist papers. His first theoretical work, *Der Einfluss der Volksvermehrung auf den Fortschritt der Geschellschaft untersucht (The Influence of Population Growth on Social Progress)* appeared in 1880.

In 1880 he moved to Zurich where he established contact with Karl Höchberg and Eduard Bernstein and, a year later, with Marx and Engels in London. Indeed, it is from around this time that it would be fair to describe Kautsky as a Marxist. In 1882 he returned to Vienna, where he collaborated with the publisher, Dietz, in founding *Die Neue Zeit*, a journal which established itself as the most influential theoretical organ of the Second International. He was to be its editor until 1917. His editorship, together with his many other published works, established him as the leading Marxist theoretician of his time.

2 The consolidation of Kautsky's position

Broadly speaking, Kautsky endorsed Engels' analysis that, while capitalism was undoubtedly heading for a terminal economic crisis, this crisis would be forestalled by a political crisis in the course of which social democracy would be called upon to seize political power and use it to transform the failing capitalist economy into a socialist order of society. This meant, according to Engels and Kautsky, that social democracy would have to prepare for its forthcoming task by employing all legal means to improve its organizational strength and increase its popular support. Kautsky underpinned this view by emphasizing the determinist aspects of Marx's materialist conception of history. This emphasis was reinforced in the course of a long and bruising controversy (largely in the pages of *Die Neue Zeit*) with the English socialist, Ernest Belfort Bax, during the mid-1890s.

In 1885 Kautsky moved to London where he worked closely with Engels. Marx's materialist conception of history provided the perspective, and the British Museum the resources, for a number of historical studies; notably, *Thomas More und seine Utopie* (*Thomas More and His Utopia*) (1888) and *Die Klassengegensätze von 1789: Zum hundertjähren Gedenktag der Grossen Revolution* (Class Struggles in the Age of the French Revolution) (1889). He was to pursue his historical interests in later years in a series of studies in which he attempted to identify the socialist element in Western civilization, particularly in early Christianity.

In 1890 he moved to Stuttgart and, in the following year, he played a major part in drafting the new programme for the German Social Democratic Party (SPD), subsequently known as the Erfurt Programme. It established the SPD as a party fundamentally Marxist in orientation. Indeed, Kautsky's own analysis and justification of the programme, *Das Erfurter Programm in seinem grundsätzlichen Teil erläutert* (*The Class Struggle: Erfurt Programme*) (1892), played a major role in the propagation of Marxism within the German socialist movement. During the 1890's he continued his activities as editor and publicist with characteristic vigour and growing popular success. His study of 'parliamentarianism' was published in 1893, and his thoughtful critique of the policy pursued (mainly) by Georg von Vollmar and his supporters, *Die Agrarfrage (The Agrarian Question)*, appeared in 1899. However, most of his energy in the later years of the decade were devoted to the 'revisionist debate'.

3 Later controversies

Shortly after the death of Engels in 1895, Eduard Bernstein had launched a campaign urging social democrats to

abandon their revolutionary aspirations and set about achieving socialism by piecemeal parliamentary reform. Kautsky was slow to react. However, at the party conference at Stuttgart in 1898, he came out against Bernstein and, in the following year, engaged Bernstein in a lengthy controversy which culminated with the publication of his *Bernstein und das sozialdemokratische Programm: Ein Anti-Kritik (Bernstein and the Programme of Social Democracy)* (1899b). The effect of all this was to consolidate his position as chief guardian and exponent of what since came to be known as 'orthodox Marxism'. He won the admiration of Lenin and even gained the grudging respect of Rosa Luxemburg. His position at this stage in his career is exemplified in his *Die soziale Revolution (The Social Revolution)* (1902), described by the then Chancellor, Prince Bernhard von Bülow, as 'a Baedeker for the state to come'.

During the turbulent years preceding the First World War, Kautsky developed an increasingly cautious approach to the question of violent revolutionary activity. His ambitious *Ethik und materialische Geschichtsauffasung: Ein Versuch (Ethics and the Materialist Conception of History)* (1906) was an attempt to vindicate a materialist approach to ethics against the arguments of the neo-Kantians. And in his *Der politische Massenstreik (The Political Mass Strike)* (1914) he sought the middle ground between the revolutionary left and the reformist right within the SPD. When war broke out, he maintained his 'centrist' position and tried unsuccessfully to preserve party unity.

The Bolshevik seizure of power in 1917 brought Kautsky once more into the forefront of debate. In his view, a socialist society was one in which all exploitation and oppression had been abolished, and this meant that its political form had to be a full democracy. The Bolshevik regime, he argued, could not be described as democratic and was therefore not, properly speaking, a socialist regime. Furthermore, Russia was a relatively backward agricultural economy with only a small industrial sector. The social and economic preconditions for the establishment of socialism were therefore absent, and the Bolshevik 'revolution' was bound to fail, with disastrous consequences for the socialist movement as a whole. Kautsky embodied these and other thoughts in his *Die Diktatur des Proletariats* (*The Dictatorship of the Proletariat*) (1918). It elicited a furious response from Lenin in his book *Proletarskaia revoliutsiia i renegat Kautskii* (*The Proletarian Revolution and the Renegade Kautsky*) (1918) and provoked a controversy which lasted for several years.

In 1927 Kautsky published his *Die Materialistische Geschichtauffassung (The Materialist Conception of History)* in which he gathered together his thoughts on Marx's basic theory. Many regard it as his major work.

See also: Social democracy; Socialism; Marxism, Western

H. TUDOR

List of works

Kautsky, K. (1880) *Der Einfluss der Volksvermehrung auf den Fortschritt der Geschellschaft untersucht (The Influence of Population Growth on Social Progress)*, Vienna: Bloch und Hasbach.(His first significant socio-economic work.)

Kautsky, K. (1888) *Thomas More und seine Utopie*; trans. *Thomas More and His Utopia*, London and New York: International Publishers, 1927.(His most enduring contribution to the history of social and political thought.)

Kautsky, K. (1889) *Die Klassengegensätze von 1789: Zum hundertjähren Gedenktag der Grossen Revolution (Class Struggles in the Age of the French Revolution)*, Stuttgart: Dietz.(An interpretation of the French Revolution from the standpoint of historical materialism.)

Kautsky, K. (1892) *Das Erfurter Programm in seinem grundsätzlichen Teil erläutert*; trans. *The Class Struggle: Erfurt Program*, Chicago, IL: Kerr, 1910.(An exposition and defence of the Erfurt Program of 1981)

Kautsky, K. (1899a) *Die Agrarfrage (The Agrarian Question)*, Stuttgart: Dietz.(His main contribution to the debate on whether the SDP should revise its programme to accommodate the agrarian or 'peasant' interest.)

Kautsky, K. (1899b) *Bernstein und das sozialdemokratische Programm: Eine Anti-Kritik (Bernstein and the Programme of Social Democracy)*, Stuttgart: Dietz.(A systematic statement of his objections to Bernstein's revisionist views.)

Kautsky, K. (1902) *Die soziale Revolution*; trans. *The Social Revolution*, Chicago, IL: Kerr, 1903.(A good statement of Kautsky's socialist vision of the future at the high point of his career.)

Kautsky, K. (1906) *Ethik und materialische Geschictsauffassung: Ein Versuch*; trans. *Ethics and the Materialist Conception of History*, Chicago, IL: Kerr, 1907.(Kautsky's main attempt to deal with a fundamental problem in the materialist conception of history.)

Kautsky, K. (1914) *Der politische Massenstreik (The Political Mass Strike)*, Berlin: Vörwarts.(His main contribution as to how the SDP should react to the anticipated outbreak of war in Europe.)

Kautsky, K. (1918) *Die Diktatur des Proletariats*; trans. *The Dictatorship of the Proletariat*, Ann Arbor, MI: University of Michigan Press, 1964.(A critique of the Bolshevik revolution and the direction in which it was going.)

Kautsky, K. (1927) *Die Materialistische Geschichtsauffassung*; trans. *The Materialist Conception of History*, New Haven, CT: Yale University Press, 1988.(His final and definitive statement of what was called orthodox.)

References and further reading

Geary, D. (1987), *Karl Kautsky*, Manchester: Manchester University Press.(A detailed analysis of Kautsky's theoretical position.)

Lenin, V.I. (1918) *Proletarskaia revoliutsiia i renegat Kautskii*; trans. *The Proletarian Revolution and the Renegade Kautsky*, New York: Contemporary Publishers Association, 1920.(Lenin's counterblast to Kautsky's *The Dictatorship of the Proletariat*.)

Salvadori, M. (1976) *Kautsky e la revoluzione socialista 1880-1938*, Milan: Feltrinelli; trans. J. Rothschild, *Karl Kautsky and the Socialist Revolution 1880-1938*, London: New Left Books, 1979.(An analytical study of Kautsky's contribution by an eminent Italian political scientist.)

Steenson, G.P. (1978) *Karl Kautsky 1854-1938: Marxism in the Classical Years*, Pittsburgh, PA: University of Pittsburgh Press.(A good intellectual biography of Kautsky.)

Keckermann, Bartholomew (1571/3-1609)

Calvinist philosopher and theologian, Bartholomew Keckermann wrote textbooks in logic, ethics and metaphysics which were widely read and in which he advanced his notion of a system of knowledge. Like so many of his contemporaries in the late sixteenth and early seventeenth centuries, Keckermann was interested in methodological matters. As professor of philosophy in Danzig, Poland, he implemented a new curriculum intended to give students an encyclopedic education within three years. His proposals had considerable influence on subsequent educators and philosophers, especially in northern Europe.

Born and brought up in Danzig (now Gdansk), Poland in 1571/3, Keckermann was given a strict Calvinist education before studying in Wittenberg, Leipzig and Heidelberg. After receiving his MA in Heidelberg in 1594, he taught philosophy first at Studentenburse Paedagogium and then at Collegium Sapientiae where he acquired the chair of Hebrew in 1600 and received his doctorate of divinity in 1602. His growing reputation resulted in an invitation from the Danzig senate to return to the city's Gymnasium, which he did in 1602. As professor of philosophy in Danzig, Keckermann developed a new educational programme intended to give students an encyclopedic education within three years: the first year was devoted to logic and physics, the second to mathematics and metaphysics, and the third to ethics, economics and politics. Several of his textbooks, which were widely used throughout northern Europe, were published posthumously.

In the sixteenth century many philosophers were engaged in a re-evaluation of the methodological and educational assumptions inherent in the scholastic universities of Europe. Progressive Aristotelians like Jacopo Zabarella re-examined the structure of knowledge and science (see Zabarella, J.), and anti-Aristotelians like Peter Ramus proposed a reformation of scholastic logic (see Ramus, P.). Keckermann drew upon Aristotelian and Ramist ideas in his own methodological proposals. With Zabarella, Keckermann believed that the scholastics' intense occupation with the texts of Aristotle should be diverted to the development of new methods and analytic systems. With Ramus, he thought that the teaching of the liberal arts should be made more systematic so that each 'art' would have a precisely delimited field and method of exposition.

Like many progressive Aristotelians of his day, Keckermann was a critical rather than dogmatic follower of Aristotle. As opposed to what he called the 'textual' Aristotelians, who laboured over every word of the ancient texts without introducing any thought of their own, Keckermann placed himself among the 'methodical' Peripetatics, who were concerned 'to develop absolute Methods and Systems of disciplines from Aristotle's font'. According to Keckermann, methodological Aristotelians like himself (and Zabarella) would first delineate a discipline and then accommodate the text of Aristotle to it.

Keckermann was probably the first philosopher to develop a theory about systems of knowledge, and it was in his writings that the term 'system' acquired its technical sense of a body of knowledge unified by a single idea or principle. His theory emphasized the orderly presentation of doctrine. For each genuine discipline or liberal art, there was a set of precepts and rules which characterized it and through which, if properly ordered, one could gain knowledge. Therefore, when the precepts and rules of a discipline were correctly presented, any student with the requisite skill and practice could master that body of knowledge. To become learned one only needed to master all the disciplines.

Keckermann adapted his theory about systems to all the liberal arts. In his lectures and textbooks, he applied his method to logic, physics, politics, metaphysics, ethics, medicine, jurisprudence, mathematics and theology. As a good 'methodological' Aristotelian, his philosophical proposals relied heavily on the texts of Aristotle and yet went beyond them. For example, *Systema ethicae* (1607), his important and large-scale work on ethics, provides a comprehensive treatment of the major issues in the field and uses Aristotle as its primary source; it also presents its topics in a framework and order different from Aristotle's.

One of the important implications of Keckermann's theory of systems is that, when taken together, they would form a unified whole which would constitute all of knowledge and which could be imparted to students. In the conception of this encyclopedia of the 'arts' Keckermann differed significantly from Aristotle in the role he assigned to metaphysics. For Keckermann metaphysics was no longer an Aristotelian science, but a liberal art

which had the task of attributing to each of the other arts or disciplines its proper domain and place in the encyclopedia. It is also noteworthy that Keckermann placed theology among the liberal arts so that it too became part of the encyclopedia of knowledge.

See also: Aristotelianism in the 17th Century; Aristotelianism, Renaissance; Calvin, J.; Medieval philosophy

CHRISTIA MERCER

Lists of works

Keckermann, B. (1614) *Opera Omnia*, Geneva, 2 vols.(The standard edition of his work.)
Keckermann, B. (1600) *Systema logicae*, Hanau.
Keckermann, B. (1607) *Systema ethicae*, London.
Keckermann, B. (1607) *Praecognita philosophica*, Hanau.
Keckermann, B. (1609) *Scientiae metaphysicae compendiosum systema*, Hanau.(Posthumously published.)
Keckermann, B. (1610) *Systema physicum*, Danzig.(Posthumously published.)
Keckermann, B. (1612) *Praecognitorum philosophicorum libri duo; naturam philosophiae explicantes et rationem eius tum docendae tum discendae monstrantes*, Hanover.(Posthumously published.)

References and further reading

Ayers, M. and Garber, D. (eds) (forthcoming) *The Cambridge History of Seventeenth-Century Philosophy*, Cambridge: Cambridge University Press.(Includes various remarks on Keckermann's place in seventeenth-century logic and a more complete bibliography.)
Gilbert, N. (1960) *Renaissance Concepts of Method*, New York: Columbia University Press.(Chapter 10 provides a brief account of Keckermann's concept of method.)
Schmitt, C. and Skinner, Q. (eds) (1988) *The Cambridge History of Renaissance Philosophy*, Cambridge: Cambridge University Press, 632-7.(A brief discussion of Keckermann's conception of system and its relationship to metaphysics.)

Kelsen, Hans (1881-1973)

Hans Kelsen was one of the foremost (positivist) legal theorists of the twentieth century. He taught in Vienna, Cologne, Geneva and Paris, and finished his life in America, teaching in Chicago, Harvard and Berkeley. He wrote widely, on legal philosophy, constitutional and international law, and political philosophy. Kelsen is best known for his Pure Theory of Law (Reine Rechtslehre) (1934). This is the basis of a theory which, with many changes, he espoused till he died.

Kelsen wanted to purify the study of law by ridding it of all unnecessary distortion or bias, to produce a science of law that would study what the lawyer did as legal scientist without asking sociological, political, or moral and value questions. He first distinguished legal science from the sociology of law (which asked questions as to what were the causes of law and how it operated in society; see Social theory and law §2). Here Kelsen makes a distinction between normative and causal science. Legal science is a branch of normative science whereas the sociology of law is a branch of natural science concerned with the study of causes. The mode of explanation in natural science, the way in which we make sense of the world, takes the form ‘if A then B will follow’. Normative science tries to make sense of the world in terms of the principle of imputation, ‘if A, then B ought to follow’ (see Norms, legal §3).

What does Kelsen think a legal norm is? Its form is that of a direction to officials to apply a sanction if someone does something that is forbidden. Thus, ‘Thou shalt not kill’ would be a prescription addressed to a judge (a direction to an official) to sentence someone to life imprisonment (to apply a sanction) if they unlawfully kill someone (if they commit a wrong). For Kelsen the norm is the meaning of a natural act and not the natural act itself - it is an act of will. The norm is the product of an interpretation of the act. But this interpretation itself implies a norm and that is why Kelsen calls the norm a scheme of interpretation. What does this mean? I see someone standing in a large hall talking to themselves. However I can also interpret this physical act in the normative world as praying. This is the normative meaning of that particular natural act of talking. That is what that person meant by the act and what I understand by it. We both use ‘a scheme of interpretation’ instantiated here by the norm ‘one ought to ask God for help in living in a dangerous world’ to make sense of the natural act. But this raises a problem. Is the fact that the people standing in the hall interpret the person’s natural act as praying definitive?

For legal science, says Kelsen, it is not. This would be merely the subjective meaning of the act and legal science is interested in the objective meaning of the act. That meaning is one from the point of view both of the individual who receives the direction and of third parties, not only from that of the individual issuing direction. When the policeman waves a car down the meaning of that act is objective because a police official is recognized as having that right by the individuals concerned and third parties. This is so because there is a valid norm, ‘drivers ought to obey policemen’, and the mode of existence of a norm is its validity. The validity or otherwise of a norm of the positive law is to be found by reference to another superior norm. This superior norm will likewise be validated by another, more general, norm. But we do not go up an infinite ladder. We finally get to the historical starting point of norm-creation, the historically first constitution. To know this is valid we have to ask whether the natural acts of, say the first Parliament, can have the objective meaning of law-making behaviour (whether their meaning is that coercive acts ought to be carried out under conditions determined by them). All the original founders are dead so we have to suppose that their activity was some form of law-making behaviour. It is only if we assume this basic norm (the *Grundnorm*) that we can say that the norms created in the system had an objective validity. It is this *Grundnorm* that enables us to recognize all the laws belonging to a particular system. So legal science is concerned with those norms that can be traced back to the *Grundnorm* that is valid because presupposed. These form a legal system that is coterminous with the state.

It is the concept of system that enables Kelsen to differentiate law from other branches of normative science, especially morals. Validity in law depends upon the pedigree of the norm, on being able to trace it back to the *Grundnorm*. Any content whatsoever can be a legal norm as long as it has been constituted in the correct fashion and can be traced up the chain of validation. The content of law is not relevant to its validity for the purpose of legal science. And any content can be a legal norm as long as it is valid. This is not the case with morals (see Law

and morality).

Also the law comprises a ‘dynamic system’ as opposed to the static one of morals. In a moral system the norms are deduced from some *Grundnorm* such as ‘Love thy neighbour’. In a dynamic system the derivation does not quite work like that. The norm is concretized as it goes down the hierarchy. Thus, the Road Traffic Act will end up as the judge fining a particular person for a particular act. The judge’s action will have the meaning of a norm which will be the concretization of the norm made by Parliament and called the Road Traffic Act. The judge creates here an individual norm. Within limits the judge has a choice and the judge’s action is valid because it has been validly authorized by the norm above. This expresses Kelsen’s view that there is always the possibility of choice in the law because the system ultimately depends upon authorized choice. Thus morals and law differ in that they are separate normative systems and that one is static and deducible and the other dynamic and undeducible.

For Kelsen validity comes from the presupposition of the *Grundnorm*. But this presupposition is not made by reason of the efficacy or moral rightness of the *Grundnorm*. The former is a sociological question and the latter a moral question; both are distinct from legal science. According to Kelsen we can presuppose the validity of the *Grundnorm* because the system is by and large effective. But this does not mean that validity is equal to effectiveness. Rather it is saying that the condition for making judgments about validity is effectiveness. This is a fine distinction, as can be seen in revolutionary cases where the *Grundnorm* appears to have been changed. Which legal system is valid? Kelsen’s attempt to postulate this *Grundnorm* as a pure hypothesis, the content of an act of thought, detached from both morals and sociology, has been very controversial and must probably be deemed a creative and interesting failure.

See also: Bobbio, N.; Hart, H.L.A.; Law, philosophy of; Legal positivism; Weyr, F.

ZENON BAŃKOWSKI

List of works

Kelsen, H. (1925) *Allgemeine Staatslehre*, Berlin: Springer; trans. A. Wedberg, *The General Theory of Law and State*, Cambridge, MA: Harvard University Press, 1945. (One of his two major works.)
Kelsen, H. (1934) *Reine Rechtslehre*, Leipzig and Vienna: Deuticke; trans. B. Litchewski Paulson and S. Paulson, *Introduction to the Problems of Legal Theory*, Oxford: Oxford University Press, 1995.(One of his two major works.)
Kelsen, H. (1957) *What is Justice?*, Berkeley, CA: University of California Press.(A translation of essays on law, politics and justice.)
Kelsen, H. (1960) *Reine Rechtslehre*, completely rewritten and expanded 2nd edn, Vienna: Deuticke; trans. M. Knight, *Pure Theory of Law*, Berkeley, CA: University of California Press, 1967.(Rewritten and expanded edition of one of his two major works.)
Kelsen, H. (1973) *Essays in Legal and Moral Philosophy*, Dordrecht: Reidel.(Translated essays by P. Heath.)
Kelsen, H. (1979) *Allgemeine Theorie der Normen*, ed. R. Ringhofer and R. Walter, Vienna: Manz, trans. M. Hartney, *General Theory of Norms*, Oxford: Oxford University Press, 1991.(A translation of his last work, which was left in manuscript when he died.)

References and further reading

California Law Review (1971) Essays in Honour of Hans Kelsen 59 (3).(Special issue on Kelsen.)
Ebenstein, W. (1969) *The Pure Theory of Law*, New York: Kelley.(A clear introduction.)
Tur, R. and Twining, W. (eds) (1986) *Essays on Kelsen*, Oxford: Clarendon Press.(A good collection of essays.)

Kemp Smith, Norman (1872-1958)

Norman Kemp Smith is now most widely known for his translation of Kant's Critique of Pure Reason. This was begun in 1913 while Kemp Smith was completing his Commentary on Kant's Critique of Pure Reason which, together with his classic studies on Descartes and Hume, established his reputation as the greatest British philosophical scholar of his day. But he was also an outstanding member of the now forgotten British 'Critical Realist' movement, much respected by A.N. Whitehead, which also included among others Kemp Smith's mentor Robert Adamson, Adamson's English pupil G. Dawes Hicks, James Ward, and his pupil G.F. Stout. Science and scientistic fallacies, psychology, including developmental psychology and the histories of science and philosophy were alike concerns of a group of independent thinkers. Their work was obscured by subsequent English philosophers' lack of attention, and the prevailing false assumption that the work of those antecedent thinkers was dominated comprehensively by views mistakenly attributed to 'Hegel' or 'Idealism'.

1 Life

Born in Dundee in 1872, Norman Smith grew up in Cupar Angus (only in 1910 did he add his wife's maiden name 'Kemp' to his own). He studied at the University of St Andrews, and on graduation in 1894 won a Ferguson scholarship which enabled him to undertake further studies in Germany, Switzerland and France. He became a Teaching Assistant at Glasgow University in 1894, where he was acting Professor of Logic in 1901-2 after Robert Adamson's untimely death. He became Professor of Psychology at Princeton in 1906 and, in 1914, McCosh Professor of Philosophy there. He returned to London in 1915 for war work until 1918, and in 1919 was appointed to the Edinburgh chair of Logic and Metaphysics. He held this position until 1945, the Second World War delaying his retirement and adding the duties of the Moral Philosophy Chair for the period 1941-4.

2 Philosophy and the history of philosophy

Kemp Smith believed that much twentieth-century British philosophy simply rehearses long-refuted views. His view was that advance in philosophy is dependent on dialogue with earlier thinkers, which requires study of their cultural contexts, including the history of science, as well as of philosophy.

While accurate learning and historical insight distinguish his early work, *Studies in the Cartesian Philosophy* (1902), this book was also, for thinkers as diverse as Ryle, Lovejoy and Macmurray, both influential and modernizing. Although Kemp Smith was no Hegelian, he insisted that the criterion of truth is the systematic coherence of mutually related viewpoints *in the light of all appearances*. Hence axioms are illicit in philosophy, and Kemp Smith criticizes Descartes' *Cogito* by showing how it implies untenable doctrines concerning perception and mind-body dualism.

Dedicated as the book Adamson did not live to write, Kemp Smith's *Commentary on Kant's Critique of Pure Reason* (1918) has become a classic. Kemp Smith goes beyond exegesis to critical dialogue. Giving new attention to the first edition of Kant's *Critique*, Kemp Smith's account of this as a patchwork, with parts from the pre-critical phases of Kant's thinking, is much more than a simple completion of Vaihinger's work. Against subsequent tendencies of English analytic philosophy Kemp Smith insists that Kant's terminology retains a provisional character due to its having originated in different contexts. Hence one cannot assume that the same word always bears the same signification, and Kant's work is further contextualized by viewing it as an attempt not at finality but at making an advance towards greater coherence and adequacy.

3 Nature contra 'Naturalism'

Although described in the 1930s as a leading figure in the new 'Realist' philosophy, Kemp Smith characterized his thought as 'Idealist' in his 1919 Inaugural address ('*The Present Situation in Philosophy*'), by contrast with 'Naturalism', not Realism. He later described this position as one which regards ethical and aesthetic (spiritual) valuations as real determining factors in the physical world.

Kemp Smith's 1905 critique of Avenarius' misleadingly named doctrine of *reine Erfahrung* ('*Pure Experience*') relates not to any contemporary prominence of Avenarius' work, but to his having given the most thorough working out of the view that perception and the attainment of truth consist in a direct physical conforming of the

brain to the external world's particularity. In opposition to this view, Kemp Smith argues against all doctrines, including that implied by Bradley's conception of 'feeling', which allege that the brain or mind has access to sheer particularity. According to Kemp Smith, the sciences of anthropology and psychology show that the development of mind, including perception, requires the initial operation of animistic notions, theologies, myths, all of which involve the identification of general patterns; and any development of the mind towards a greater objectivity proceeds by criticism of the mutually contradictory beliefs to which these early operations give rise. Furthermore, Kemp Smith argues, the origin and function of these early operations cannot be accounted for in a physicalist theory such as that of Avenarius; the hypothesis would be of a species programmed with a tendency to self-deception while also capable of desiring and possessing truth. In fact, theologies, myths and so forth are not the sheer obstructions to knowledge Avenarius supposes; they arise from confrontations with practical needs and remain subject to challenge by further practical needs. Avenarius has obscured the truth by imposing a methodology which, by parroting categories from a currently fashionable science, mistakenly fancies itself to be scientific.

Kemp Smith (1923) further develops his position in a response to Whitehead's conception of nature (see Whitehead, A.N. §§3-4). 'Nature' is no mere scheme of abstract physical laws; instead it is of such a complexity that the richness of sense-experience, far from being a mere subjective colouring, manifests only the structures that human practical needs have revealed. Thus 'subjectivism' is rejected as missing the real complexity of reality, and it is affirmed that the routes to insight into reality may be various, with none that can be presumed exhaustive. In his *Prolegomena to an Idealist Theory of Knowledge* (1924) Kemp Smith proceeds by way of a phenomenology of mind as a 'patchwork' which takes account of the problem of illusion and its recognition. 'Intuition' is the capacity which results from the developing mind's transactions with reality and the variously mythical and generalizing schemata that have come into being within the mind's development. It comes into being through the lifetime operation of practical needs and of that gradual getting beyond them which constitutes the capacity for knowledge. Biological evolution and the development of the individual organism have equipped it to deal with 'sensa'. These sensa are the data of the senses, and also elements of the presentations witnessed when, for instance, thinking or dreaming. That actual things are seen, says Kemp Smith, is the result of the operation of 'intuition', as understood above. Thus the particular is not the basis of knowledge but the objective of a complex, lengthy pursuit.

In 'The Nature of Universals' (1927) Kemp Smith criticizes the positions of Bradley and Bosanquet, and also that of his friend G.F. Stout, and advances an account of universals which echoes that of Reid. It is not the white of any existing thing which is the universal (for that cannot be whiteness of another thing). The universal 'white' comes into being by way of contrasts made between different characteristics of things, of characteristics identified by contrast as 'colours'. Universals access reality by representing kinds and characters as these have been identified by way of contrasts, both in one's own experience and in that of others from whom one learns them. Linguistic terms are general too, but for Kemp Smith they differ from universals *qua* logical terms as a result of the different context of investigation in which logical terms arise and are applied. This difference between language and logic is crucial for Kemp Smith's friend and former assistant, John Anderson (see Universals §2).

4 Critique of 'Enlightenment'

Kemp Smith conceives of original sin as human limitedness compounded by delusions of self-sufficiency which ignore the unconscious and unwilled aspects of the development of mind and knowledge. Calvin, he held, was a 'Copernicus of psychology', his theology a pioneering scientific criticism of naive voluntarism. Yet Kemp Smith is also a critic of sophistical relativism. He argues that different viewpoints, such as those of religion, manifest different bases in the development of mind, the different perspectives consequent on different minds addressing a complex reality. In the 1920s Kemp Smith contemplated a book-length study of Romanticism, but this was not written; the critique of 'Enlightenment' in his protégé A.A. Bowman's *Studies in the Philosophy of Religion* (1938), however, follows Kemp Smith in emphasizing the reliance of shallow, optimistic naturalism in ethics on a decontextualized, ahistorical, frame of reference (see Enlightenment, Continental §2; Naturalism in ethics §2).

5 Hume, Descartes and rediscovery

While editing his late close friend F. von Hugel's *The Reality of God* (1927), Kemp Smith seems to have reconceived the project of a critique of Enlightenment, von Hugel's clichéd depiction of Hume as smiling villain being so much at odds with the seriousness and reverence for nature that Kemp Smith traced in 'The Naturalism of

Hume' (1905). Kemp Smith's introduction to what has on occasion been referred to as his 'flawless edition of a flawless work', Hume's *Dialogues on Natural Religion* (1935), identifies the 'delusion' and 'sham' of a famous quote with the ratiocinatory 'Federal' perversion of neo-Calvinist religion on which Hume was raised. Opening a new era in study of that towering figure, *The Philosophy of David Hume* (1941) shows Hume rejecting a Hobbesian self-interest doctrine but also directly criticizing Francis Hutcheson's moral intuitionism and the doctrine of design on which it is founded. Deepening his perspective by reference to continental Hume studies, Kemp Smith refutes the conception of an 'associationist' Hume beloved of the Mills and their followers, and on the same misreading anathema to their adversaries. According to Kemp Smith, Hume's later work is no falling-away from his *Treatise*; Hume had rather abandoned his false hopes for what might be accomplished by associationist theory. His writings from the *Enquiries* onwards represent Hume's initial and abiding project of a comprehensive scientific account of human beings founded on the Newtonian observational approach to 'moral questions'. His abiding merit is in his application of the approach of natural history to the study of man and society, making observations inconsistent with that very crude associationism which has often mistaken him for an ally.

This commitment to comprehensive critical-historical exegesis was for Kemp Smith not a retirement into the 'history of philosophy', but a way of advancing philosophy. In his *New Studies in the Philosophy of Descartes* (1954) Kemp Smith expounds the range of Descartes' interests and the development of his doctrines within the historical contexts which shaped their formulation. Half a century after his early challenge to Descartes' *Cogito*, Kemp Smith reconsiders the fruitful complexities of his views on mind-body dualism and perception, especially in relation to later scientific developments.

In returning to Descartes, Kemp Smith addresses the 'two cultures' problem raised by Alois Riehl, whom Princeton had honoured during Kemp Smith's time there. In his *Der Philosophische Kriticismus* (1876-87), Riehl followed the one-sided emphasis of Kant's immediate successors on literary and humanistic studies and suggested a loss of that close relation to the physical sciences which Kant shared with Descartes, a loss which imperilled subsequent philosophy. Furthermore, Riehl thought that philosophy, by way of reaction, mightequally lose itself in an exclusive pursuit of a scientistic methodology. In response, Kemp Smith maintained that in order to avoid these errors, manifest time and again in the history of philosophy, philosophers need to remain open to the history and context of philosophy, both as a warning and as a source of still unappreciated insights. The history of philosophy is an essential element of the dialogue which is philosophy's continuing project.

ROBERT R. CALDER
GEORGE DAVIE

List of Works

Smith, N. (1902) *Studies in the Cartesian Philosophy*, London: Macmillan.(Decisive for its critique of the *Cogito*, this discusses Descartes' medieval inheritance and examines his influence on Malebranche, Leibniz, Berkeley and Hume.)

Smith, N. (1905) 'The Naturalism of Hume', *Mind* 14: 149-73, 335-47.(Counters notions of Hume as an impious mischief-maker by reference to an account of Nature which echoes Calvin and foreshadows Wordsworth.)

Smith, N. (1906) 'Avenarius's Philosophy of Pure Experience', *Mind* 15: 13-31, 149-60.(William James declared this paper to have dealt a death-blow to Avenarius' theory.)

Kemp Smith, N. (1918) *A Commentary on Kant's Critique of Pure Reason*, London: Macmillan; 2nd enlarged edn, 1923.(Includes a thorough account of previous commentaries; invaluable resource for explication of Kant's terminology in general.)

Kemp Smith, N. (1919) *The Present Situation in Philosophy*, Edinburgh: Thin; repr. in *The Correspondence of Baron Friedrich von Hugel with Professor Norman Kemp Smith*, New York: Fordham University Press, 1981. (Incidentally valuable in exposing implications of doctrines pretending to be all-comprehendingly scientific. The *Correspondence* sheds light on wider references not represented in Kemp Smith's published work.)

Kemp Smith, N. (1923) 'Whitehead's Philosophy of Nature', *Issues and Tendencies in Contemporary Philosophy*, Berkeley, CA: University of California Publications in Philosophy, 4, 197-224.(Emphasizes the limitation of any theory in relation to the complexity of subject-matter.)

Kemp Smith, N. (1924) *Prolegomena to an Idealist Theory of Knowledge*, London: Macmillan.(Emphasizes the complexity of mind and seriously criticizes pretences of theory to exclusiveness. Apparent simplicities of relations between mind and reality are founded on mental processes of intense complexity.)

Kemp Smith, N. (1927) ‘The Nature of Universals’ *Mind* 36: 137-57, 265-80, 393-422.(Criticizes Bosanquet and Stout, maintaining that universals are abstract terms in debate rather than psychological representations.)

Kemp Smith, N. (1935) *‘Introduction’ to David Hume: Dialogues on Natural Religion*, Edinburgh: Nelson; 2nd augmented edn, 1947.(Draws particular attention to Hume’s restricted definition and limited experience of religion.)

Kemp Smith, N. (1941) *The Philosophy of David Hume*, London: Macmillan.(Notes contemporary references not identified in Hume’s arguments and maintains that Hume’s *Treatise* and *Enquiries* were preparatories to the later historical and political studies.)

(1954) *New Studies in the Philosophy of Descartes: Descartes as Pioneer*, London: Macmillan. (Presents the immense range of Descartes’ scientific studies, insisting on the importance of mind-body dualism in identifying questions liable to be raised by future scientific developments.)

Kemp Smith, N. (1967) *The Credibility of Divine Existence*, ed. A.J.D. Porteous, R.D. MacLennan and G.E. Davie, London: Macmillan.(This is Kemp Smith’s own selection from his papers - including 1905, 1906, 1923, 1927 above - which covers topics such as Locke, Bergson, progressivism in ethics and fear.)

References and further reading

Bowman, A.A. (1938) *Studies in the Philosophy of Religion*, ed. N. Kemp Smith, London: Macmillan.(Criticism of the Enlightenment project by a friend, much influenced by Kemp Smith.)

Calder, R.R. (1997) *A School of Thinking*, Edinburgh: Polygon.(Devotes a chapter to Kemp Smith in an account of his Scottish contemporaries as well as of Robert Adamson and the Critical Realist background.)

Ewing, A.C. (1959) ‘Norman Kemp Smith’, obituary in the *Proceedings of the British Academy* 45: 297-306. (Biographical account by a contemporary, mostly in sympathy and usually tentative regarding any disagreement.)

Kant, I. (1781/1787) *Immanuel Kant’s Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1929; 2nd corrected impression, 1933.(The second edition of the first *Critique*, supplemented with translation of the passages from the first edition which were altered or replaced in the second. Numerous reprints.)

Merz, J.T. (1896) *History of European Thought in the Nineteenth Century*, Edinburgh: Blackwood, vols 3 and 4; repr. New York: Dover, 1965.(An authoritative work by an Anglo-German, with a full appreciation of the very different context in which Kemp Smith’s thought developed.)

Metz, R. (1938) *A Hundred Years of British Philosophy*, London: Routledge.(Includes accounts of Kemp Smith, Adamson, *et al.*, with an informative depiction of the variety of British philosophy *c*.1930.)

Passmore, J. (1968) *A Hundred Years of Philosophy*, Harmondsworth: Penguin.(A useful supplement to Metz.)

Riehl, A. (1876-87) *Der Philosophische Kriticismus und seine Bedeutung für die positive Wissenschaft*, Leipzig: W. Engelmann, 1924-5.(Presentation of Kantian method which emphasizes Kant’s continuity with preceding philosophy and physical science.)

Kepler, Johannes (1571-1630)

Kepler's mathematical analysis of Brahe's observations of the motions of Mars enabled him to formulate the descriptive 'laws' of planetary motion, thus giving heliocentric astronomy an empirical basis far more accurate than it had before. He insisted that astronomy had to discover the causes *of the motions that the laws described, in this way becoming a 'physics of the sky'. In the pursuit of this goal, he formulated the notion of distance-dependent forces between sun and planet, and guessed that gravity could be explained as an attraction between heavy bodies and their home planets, analogous to magnetic action, thus pointing the way for Newton's theory of gravity.*

Kepler was introduced to the work of Copernicus while at the University of Tübingen. In 1596 he completed the *Mysterium cosmographicum*, explaining the spacing and the periods of the planets in the Copernican system along broadly Neoplatonic lines. The orbits of the six planets were to be separated by the five regular polyhedra, and the increase in the planetary periods with distance from the sun would be explained by the diminution of a driving force from the sun. Since all of this was contingent upon the Copernican planetary ordering, the treatise constituted an explicit (if unconventional) defence of that ordering, the first since the *De revolutionibus orbium coelestium* in 1543.

In 1600, the astronomer Tycho Brahe asked Kepler to defend Brahe's work against the criticisms of Nicolaus Ursus (Baer), who dismissed mathematical astronomy generally as no more than convenient fiction. In a short work, the *Apologia pro Tychone contra Ursum* (1600), Kepler drew on the history of astronomy to construct a perceptive analysis of hypothetical reasoning in astronomy. He argued that astronomers must not be content with geometrical description, since two such 'hypotheses' might very well save the phenomena equally well. One could not in that case decide between them on geometrical-empirical grounds. What is needed is a causal analysis, a physics of the motions that explains why they occur as they do. Not published until 1858, the *Apologia* has come to be recognized as one of the more acute contributions to philosophy of science from the seventeenth century.

Kepler succeeded Brahe as imperial mathematician in 1601 and began intensive work aimed at establishing the shape of Mars' orbit. He could not avail himself of the convenient device that both Ptolemy and Copernicus had relied on of using different orbits for the calculation of the planet's longitudes and latitudes. The seemingly irreducible error of $8'$ that resulted would not have been significant were it not for the unprecedented accuracy of Brahe's data. Kepler asked himself what physical causes could make the planet move in its evidently non-circular path. Gilbert's theory of the earth as a vast magnet led him to propose that magnetic forces between the two great magnets, sun and planet, could account for both the varying distance between planet and sun, and the variations in orbital velocity of the planet. Finally, in chapter 58 of the *Astronomia nova* (1609), he announced triumphantly: ' Reasoning from physical principles agreeing with experience, there is no figure left for the orbit of the planet except a perfect ellipse'. His second law, that the planet's radius vector sweeps out equal areas in equal times, is tucked away in the massive book also, but even Kepler himself did not recognize its significance until later.

In 1618, he published the *Harmonice mundi*. Applying the principles of harmony to the universe, he tried to explain the distances and periods of the planets and discovered what would come to be known later as his third law (the squares of the periods of any two planets are proportional to the cubes of their mean distances from the sun). In this work he also attempted to formulate an elaborate theoretical basis for astrology. The zodiac, he asserts, is a projection of the human soul; the positions of the planets at a person's birth mark the soul for the remainder of life. Kepler constructed horoscopes and prognostications throughout his professional career; astrology was an important source of income for him. Yet he also expresses doubt about its claims, warning that it may be a matter of luck when astrological prognostications turn out to be correct.

His next work, the *Epitome astronomiae Copernicanae* (1618-21), was widely influential. It presented the case for Copernican astronomy in textbook fashion, laying out the evidence for the earth's motions and discussing, though not as fully as Galileo would later do, the imperceptibility of shared uniform motion. Much of the work was devoted to the technical geometrical detail required for practical computation of the planetary orbits. But the emphasis of the argument was, as always, on causal reasoning; this time he proposed a swirl of magnetic 'species' that would move planets along in their paths. (Lacking the principle of inertia, he had to find a cause for

continuing onward motion.) But as in his earlier works, he did not confine himself to efficient causes only, and found a motive for Copernicanism in the symbolism of God's Trinitarian relation with the universe, the centre of the universe representing God the Father. He also sought final causes for every major feature of the planetary system, even down to the relative sizes and densities of the planets.

Kepler sought order and harmony in the universe, a conviction that guided him in his successful search for the true orbits of the planets. And it prompted him to see astronomy itself as a causal science and not merely a saving of the phenomena. His assurance that harmonies discoverable by the human mind were embodied in the cosmos derived in the first instance from his theology of creation and Neoplatonic metaphysics. Yet his achievements in astronomy, particularly his unconventional notion of a sun-centred force controlling the movement of the planets, showed the way to a self-contained quasi-mechanistic worldview.

See also: Causation; Cosmology; Explanation; Galilei, G.; Newton, I.

ERNAN McMULLIN

List of works

Kepler, J. (1600) *Apologia pro Tychone contra Ursum (A Defence of Tycho against Ursus)*; in N. Jardine (ed.), *The Birth of History and Philosophy of Science*, Cambridge: Cambridge University Press, 1984, 134-207.(The *Apologia* in English, with the addition of the Latin text and eight chapters detailing the circumstances of the work's composition and its significance in terms of contemporary philosophical concerns.)

Kepler, J. (1609) *Astronomia nova (New Astronomy)*; trans. W.H. Donahue, Cambridge: Cambridge University Press, 1992.(Kepler's most important work, available for the first time in English translation. A challenge to the translator because of its convoluted style as well as its author's habit of retracing every blind alley taken.)

Kepler, J. (1618) *Harmonice mundi (The Harmony of the World)*; trans. E.J. Aiton, A.M. Duncan and J.V. Field, Philadelphia, PA: American Philosophical Society, 1995. (With an introduction and notes.)

Kepler, J. (1937-) *Gesammelte Werke*, eds M. Caspar *et al.*, Munich: Beck.(The definitive edition, of twenty-two volumes, under the guidance of a series of devoted editors.)

References and further reading

Beer, A. and Beer, P. (eds) (1975) *Kepler: Four Hundred Years*, New York: Pergamon; vol. 18 of the series *Vistas of Astronomy*.(Large set of scholarly essays on every aspect of Kepler's life and legacy.)

Caspar, M. (1993) *Kepler*, trans. C.E. Hellman, New York: Dover.(The standard biography of Kepler, with a new introduction and references by O. Gingerich.)

Koyré, A. (1961) *The Astronomical Revolution: Copernicus, Kepler, Borelli*; trans. R.E.W. Maddison, Ithaca, NY: Cornell University Press, 1973.(The central section of this work, *Kepler and the New Astronomy*, 117-464, gives an admirably clear account of Kepler's unique contribution: a new ideal of astronomy that would require not only accurate prediction of planetary motions but an explanatory account of the forces responsible for them.)

Stephenson, B. (1987) *Kepler's Physical Astronomy*, Princeton, NJ: Princeton University Press.(A detailed account of the physical theories with which Kepler complemented his mathematical analysis of the planetary motions.)

Stephenson, B. (1994) *The Music of The Heavens: Kepler's Harmonic Astronomy*, Princeton, NJ: Princeton University Press.(A sympathetic analysis of Kepler's use of harmonies and archetypes to construct a planetary cosmology.)

Keynes, John Maynard (1883-1946)

Keynes is best known as an economist but, in the tradition of John Stuart Mill and William Stanley Jevons, he also made significant contributions to inductive logic and the philosophy of science. Keynes' only book explicitly on philosophy, A Treatise on Probability (1921), remains an important classic on the subject. It develops a non-frequentist interpretation of probability as the key to sound judgment and scientific reasoning. His General Theory of Employment, Interest and Money (1936) is the watershed of twentieth-century macroeconomics. While not, strictly speaking, a philosophical work, it nonetheless advances distinct readings of rationality, uncertainty and social justice.

1 Probability and induction

Born in Cambridge, England, Keynes was educated at Eton and Cambridge University where his father Neville lectured on logic and economics. After studying mathematics and economics and working briefly in the India Office, Keynes was elected, in 1908, a fellow of King's College Cambridge. During his twenties and thirties, Keynes undertook an informal study of philosophy in the company of the Cambridge 'Apostles' - G.E. Moore, John McTaggart, Bertrand Russell, Alfred North Whitehead, Frank Ramsey and Ludwig Wittgenstein - from whom he forged his own unique matrix of rationalist and empiricist beliefs. Perhaps the most colourful episode of Keynes' fascinating life was his founding association with the Bloomsbury Group of artists and intellectuals. He also served as an advisor to the British Treasury, most notably on the Treaty of Versailles (1919), which he repudiated, and on the early negotiations for the Bretton Woods Agreement (1946). Keynes' prophetic grasp of the political and economic transformations of his world will undoubtedly be valued for generations to come.

In his *Treatise on Probability* (1921), Keynes sought to draw out the logical foundations of probability and to graft these onto the mathematical calculus of probability. This complemented the logicist programme of Russell and Whitehead, particularly the common aim of grounding knowledge in a limited set of intuitions. Keynes viewed probability as a species of logical inference, as the degree of rational belief about a given proposition which extant knowledge authorizes us to entertain. Propositions are known by direct acquaintance with the world and may be true or false, but probability, as a property of an argument, has no bearing on the question of truth. As more evidence comes to light, the probability might increase in value. It does not, however, approximate truth in the straightforward correspondence sense. For Keynes, 'probability begins and ends with probability'.

Probability is objectively determined in the sense that it is transindividual and thus not a function of personal history. A rational person, such as a magistrate or actuary, will arrive at the correct probability judgment and use this as the guide to right action. Nevertheless, in most areas of life we must be content with a qualitative estimate. A numerical measure is only possible in the case of a set of exclusive and exhaustive equiprobable alternatives, and such cases are the exception rather than the rule. Keynes thereby contracts rather than expands the domain of numerical probability assertions. There are for him four possible options when confronted with a non-deductive argument:

(1) there is no probability;
(2) the probabilities do not all belong to a single set of magnitudes measurable in terms of a common unit;
(3) there are existing measures but they must remain unknown; or
(4) the probabilities can be measured and belong to the same set, and are thus capable of comparison.

As for constructing a scale of probabilities, Keynes grants that the end points of the series, certainty and impossibility $(1, 0)$, always exist and that, if there is a given probability, it will lie between them. But in the majority of cases it is not possible to provide a measure of that probability, let alone compare it to others.

Keynes appreciated statistics for its fact-gathering function but was sceptical about its potential for hypothesis testing. Accordingly, he did not extend the rich linkages between probability and statistics. The theorems of Bernoulli and Laplace were discredited, as was any statistical inference that presupposed a probability distribution.John Venn's frequency interpretation of probability was also misguided, in part because of its ascription of numerical measures, in part because of its contingent character and in part because large and stable samples are difficult to obtain (see Venn, J.). Moreover, the frequency interpretation entails a commitment to

natural kinds and to the belief that probabilities are features of the external world rather than of the human mind. For Keynes, probability is a purely logical relationship between the propositions of an argument. In response to criticisms from Ramsey, Keynes later equivocated on the degree of consensus required for objective probability assertions, but he remained steadfast to his rationalist interpretation.

Laplace and a number of followers endorsed the principle of indifference, which stipulated that, in the case of complete ignorance, the a priori probability of a given proposition would be 1/2. Keynes found this suspect because of its ready ascription of a number to what is, at best, known qualitatively. Furthermore, he recognized that one could always break the class of objects denoted by the predicate term into further subdivisions and thus grant that what were treated as equivalent classes before may now contain a different or an indefinite number of more elementary groupings.

One of Keynes' innovations is a property that he calls the weight of an argument, which measures the absolute amount of knowledge and ignorance respectively. Hence, in the case of gathering more relevant yet unfavourable evidence, one would increase the weight while lowering its probability. One might also have a proposition with a high probability and a low weight. The question Keynes leaves us with is whether or not one should always set out to acquire a high weight. Clearly, there comes a point where it is no longer worthwhile to acquire further information, but what principle will determine this threshold is a matter left unresolved. In his economic analysis of investment, however, Keynes made use of this property in order to highlight the ever-present uncertainty of business forecasting.

A hue of uncertainty is also cast over science in general. Inductive arguments are fundamentally probabilistic and thus do not serve to approximate the ideal state of complete certainty. Moreover, pure induction is for Keynes relatively pointless if the confirming instances do not differ except with respect to their spatiotemporal properties. It is analogical reasoning that matters the most. In order to strengthen a scientific hypothesis, it is important to vary the instances, particularly to increase the negative analogies. We are impressed with Newton's law of universal gravitation because of the variety of circumstances it fulfils rather than the mere number of them. Furthermore, prediction in itself is of no consequence since it is essentially a case of pure induction and plays into a dispositional weakness for the temporal ordering of scientific discoveries. Proper grounds for accepting a hypothesis are built up from objective logical relations as elucidated by the theory of probability.

Previous solutions to the problem of induction, such as J.S. Mill's appeal to the uniformity of nature, are discarded as spurious. Keynes concentrates rather on a probabilistic justification of analogical inference. Our convictions about the existence of other people and about the causal links between beliefs and actions are both ample evidence of the support we hold for analogizing. More importantly, we would only be in a position to draw analogies if the universe was not infinitely varied. A priori, we require a finite probability in favour of the inductive hypothesis that there is an upper limit to the independent variety of the objects that comprise our generalizations. Once we grant this so-called principle of the limitation of independent variety, we find that our experience of the world, rather than diminishing its probability, has in fact increased it. In this respect, Keynes maintains that induction is non-circular, even if it is not fully justified. He has provided a probabilistic argument for a principle that he assumes to have an initial probability.

2 Political economy

Keynes viewed economics as a moral science and unfolded one of the most detailed pictures of human agency since David Hume and Adam Smith. By distinguishing different motives for investment and different capacities to discern the flux of the business cycle, and by recognizing that individual motives and actions are rarely synchronized due to ignorance and uncertainty, he broke with the classical conviction that economies are self-equilibrating. Claiming as his forebears the mercantilist economists of the seventeenth century, he argued that the state alone had the monetary and fiscal means to achieve economic stability. But he stopped short of state socialism. With an eye to preserving individual liberty, Keynes proposed a nation of semi-autonomous bodies serving the public good, on the scale of corporations, universities or the central bank.

Keynes was sanguine that the solution to economic scarcity lay in capital accumulation guided by science, technology and the state. But capitalism as a system, while the most efficient means to this end, is grounded in cupidity and thus is morally reprehensible. Moreover, it tolerates high levels of unemployment and unwarranted

inequalities of wealth. Genuine social justice would be possible only if and when humankind freed itself from its animal passion for money and expanded its ambit of altruism. This, he hoped, would come about gradually, with the euthanasia of the *rentier* class and the redistribution of wealth through sound fiscal policy. It was conceivable to Keynes that within approximately 100 years, citizens of the developed economies would no longer be bound by economic necessity and that such abundance would unleash a new form of human flourishing and morality. Keynes' biographical essays on great minds (1933), and his lifelong devotion to the arts leaves little doubt as to what form these higher goods would take.

See also: Economics, philosophy of; Induction, epistemic issues in; Inductive inference; Probability, interpretations of

MARGARET SCHABAS

List of works

Keynes' early philosophical essays are unpublished, and are housed in the Keynes Papers, King's College, Cambridge University.

Keynes, J.M. (1921) *A Treatise on Probability*, in D. Moggridge and A. Robinson (eds) *The Collected Writings of John Maynard Keynes*, vol. 8, London: Macmillan, and New York: Cambridge University Press, 1971-89. (Keynes' one philosophical treatise. Developed an analysis of probability that complements the logistic programme of Russell and Whitehead.)

Keynes, J.M. (1931) *Essays in Persuasion*, in D. Moggridge and A. Robinson (eds) *The Collected Writings of John Maynard Keynes*, vol. 9, London: Macmillan, and New York: Cambridge University Press, 1971-89. (Keynes as policy advisor on the Treaty of Versailles, the gold standard, free trade and liberal ideals. Also offers a vision of the future in 'Economic Possibilities for our Grandchildren'.)

Keynes, J.M. (1933) *Essays in Biography*, in D. Moggridge and A. Robinson (eds) *The Collected Writings of John Maynard Keynes*, vol. 10, London: Macmillan, and New York: Cambridge University Press, 1971-89.(Keynes the master stylist captures the personalities of Isaac Newton, Stanley Jevons, Frank Ramsey and Winston Churchill, among others.)

Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*, in D. Moggridge and A. Robinson (eds) *The Collected Writings of John Maynard Keynes*, vol. 7, London: Macmillan, and New York: Cambridge University Press, 1971-89. (The most important treatise on macroeconomics of the twentieth century. Maintained that economies are not self-regulating and require government intervention in periods of high unemployment.)

References and further reading

Bateman, B.W. (1987) 'Keynes's Changing Conception of Probability', *Economics and Philosophy* 3: 97-119. (Argues that Ramsey's criticisms prompted Keynes to shift to a frequentist interpretation of probability prior to the publication of the *General Theory*.)

Carabelli, A.M. (1988) *On Keynes's Method*, New York: St Martin's Press.(Detailed study of Keynes' probability and its links to economics.)

Davis, J.B. (1994) *Keynes's Philosophical Development*, Cambridge: Cambridge University Press.(Places Keynes' thought within the context of the Cambridge philosophers G.E. Moore, L. Wittgenstein and F. Ramsey.)

Moggridge, D.E. (1992) *Maynard Keynes, An Economist's Biography*, London: Routledge.(A fine biography with detailed chapters on his probability and economic philosophy.)

O'Donnell, R.M. (1989) *Keynes: Philosophy, Economics and Politics*, London: Macmillan.(Highlights the unity of Keynes' philosophy and political economy.)

Runde, J. (1994) 'Keynes After Ramsey: In Defense of *A Treatise on Probability*', *Studies in History and Philosophy of Science* 25: 81-96.(Argues, contra Bateman, that Keynes did not significantly alter his views on probability after Ramsey's criticisms. Nice summary of the debate.)

Kierkegaard, Søren Aabye (1813-55)

Although Kierkegaard's name has come to be chiefly associated with writings on philosophical themes, his various publications covered a wide range that included contributions to literary criticism, discourses on specifically religious topics and forays into polemical journalism. Born in Copenhagen in 1813, he led an outwardly uneventful existence there until his death in 1855. None the less much that he wrote drew upon crises and turning points in his personal life; even his theoretical works often had an autobiographical flavour.

Kierkegaard held that the philosophy of his time, largely owing to the influence of Hegelian idealism, tended to misconstrue the relation of thought to reality, wrongly assimilating the second to the first; in doing so, moreover, it reflected an age in which habits of abstract reflection and passive response had blinded people to their true concerns as self-determining agents ultimately accountable for their own characters and destinies. He sought to counter such trends, exploring different approaches to life with a view to opening his reader's eyes both to where they themselves stood and to possibilities of opting for radical change. He implied that decisions on the latter score lay beyond the scope of general rules, each being essentially a problem for the individual alone; even so, his portrayal of the religious mode of existence presented it as transcending limitations experienced in alternative forms of life. Kierkegaard, himself an impassioned believer, was at the same time crucially concerned to articulate the Christian standpoint in a fashion that salvaged it from recurrent misconceptions. Rejecting all attempts to provide objective justifications or proofs of religious claims, he endorsed a conception of faith that eschewed rational considerations and consisted instead of subjective self-commitment maintained in the face of intellectual uncertainty or paradox. His account was set within a psychological perspective that laid stress upon freedom as an inescapable condition of action and experience. The complex implications he believed this to possess for the interpretation of pervasive human emotions and attitudes were discussed in works that later proved highly influential, particularly for the growth of twentieth-century existentialism. Here, as in other areas of his writing, Kierkegaard made a significant, though delayed, impact upon the course of subsequent thought.

1 Life

Kierkegaard was the youngest son of a prosperous and largely self-made Danish businessman. The father was a deeply religious but exacting and guilt-ridden individual who communicated his feelings of melancholy and anxiety to other members of his family; they certainly left a lasting impression on Kierkegaard's own character and development, causing him later to describe his upbringing as having been 'insane'. It was perhaps largely from a desire to please his father, towards whom he tended to exhibit an ambivalent mixture of love and fear, that at the age of seventeen he enrolled at the University of Copenhagen with the object of taking a degree in theology. Nevertheless, after passing his preliminary examinations he found himself increasingly attracted to other spheres of intellectual interest, particularly those involving developments in contemporary philosophy and literature; at the same time he cultivated a fashionably sophisticated lifestyle, following pursuits sharply at variance with the austere precepts that had been inculcated upon him at home. But in his journals, which he began during his protracted period as a student and continued to keep for the rest of his life, he is already to be found recording a growing dissatisfaction with the wayward mode of existence he had adopted, and the death of his father in 1838 appears finally to have prompted him to return to his academic studies with a view to settling down to a professional career. Thus by July 1840 he had been awarded his degree, and two months later he announced his engagement to marry Regine Olsen, the daughter of a highly placed civil servant. This, however, was not to be.

The story of Kierkegaard's abortive engagement is familiar from his journals, where he provided a detailed account of how he eventually broke off the relationship after an uneasy year during which he harboured regrets about his proposal. While his actual motives for making the final breach are left somewhat obscure, there can be no doubt as to its significance for his later thought and writings, allusions to it - often only lightly disguised - occurring in a variety of his works. In any case it certainly constituted a turning point. Henceforward he withdrew into a bachelor existence; moreover, although by now firmly committed to Christianity, he effectively abandoned any further thought of a clerical career and devoted himself instead to living as a writer on the very comfortable income he had inherited from his father's estate. The initial period of his authorship was in fact remarkably productive. He took less than a year over his master's dissertation *Om Begrebet Ironi (The Concept of Irony)*, successfully submitting it to the university faculty in 1841, and he followed it with a series of books, all issued

under pseudonyms, which were largely concerned with philosophical or psychological aspects of ethical and religious belief. The first, entitled *Enten-Eller (Either/Or)*, came out in two substantial volumes in 1843 and was succeeded later in the same year by *Frygt og Baeven (Fear and Trembling)* and *Gjentagelsen (Repetition)*; in 1844 *Philosophiske Smuler (Philosophical Fragments)* and *Begrebet Angest (The Concept of Anxiety)* appeared, and these in turn were followed by *Stadier paa Livets Vej (Stages on Life's Way)* in 1845 and by *Afsluttende uvidenskabelig Efterskrift (Concluding Unscientific Postscript)* in 1846. Two further pseudonymous works on connected themes, *Sygdommen til Døden (The Sickness unto Death)* and *Indøvelse i Christendom (Training in Christianity)*, were published in 1849 and 1850 respectively.

Although it is the writings listed above that have chiefly attracted the attention of subsequent philosophers and commentators, they by no means exhaust the total of Kierkegaard's literary output during the 1840s. Apart from some critical pieces, he also produced - this time under his own name - a number of directly religious discourses in which he aimed to present the essentials of Christian teaching; thus such works as his *Opbyggelige Taler i forskjellig Aand (Edifying Discourses in Various Spirits)* of 1847 were expressly designed to communicate and illustrate the true nature of the Christian message and the demands it imposed upon the individual. In their uncompromising emphasis on the severity of these requirements, and in their manner of stigmatizing the complacency and 'double-mindedness' imputable to contemporary representatives of the faith they professed to serve, the latter books can be said to foreshadow the standpoint from which, in the culminating phase of his career, he launched a violent assault upon the established Church of Denmark. The occasion for this was the death of the Danish primate, Bishop Mynster, in 1854. Kierkegaard had increasingly come to regard Mynster as exemplifying in his own person many of the shortcomings of the church as a whole, and he was therefore incensed by hearing the dead prelate pronounced instead to have been a 'witness to the truth'. As a result he set out in the following months to denounce the covert worldliness and hypocrisy that permeated the clerical establishment, first through articles in the public press and subsequently in a broadsheet printed at his own expense. The ferocity of his attacks, appearing after a spell when he had published relatively little, caused surprise and some consternation. The controversy they stirred up was, however, abruptly interrupted by Kierkegaard's sudden collapse in October 1855 and his death a few weeks later.

2 The limits of objectivity

In an early entry in his journals, written when he was still a student, Kierkegaard gave vent to the dissatisfaction he felt at the prospect of a life purely devoted to the dispassionate pursuit of knowledge and understanding. 'What good would it do me', he then asked himself, 'if truth stood before me, cold and naked, not caring whether I recognized her or not?' Implicit in this question was an outlook which was to receive mature articulation in much of his subsequent work, being particularly prominent in his criticisms of detached speculation of the kind attributable to those he called 'systematists and objective philosophers'. To be sure, and notwithstanding what has sometimes been supposed, he had no wish to be understood as casting aspersions on the role played by impersonal or disinterested thinking in studies comprising scholarly research or the scientific investigation of nature: such an approach was quite in order when adopted within the limits set by determinate fields of enquiry. But matters were different when philosophical attempts were made to extend it in a manner that purported to transcend all particular viewpoints and interests, this conception of the philosopher's task leading to the construction of metaphysical theories which sought to comprehend every aspect of human thought and experience within the disengaged perspective of objective contemplation. Kierkegaard considered Hegel to be the foremost contemporary representative of the latter ambition, the famous system to which it had given rise being in his opinion fundamentally misconceived.

Kierkegaard's general reaction to what he found unacceptable in the Hegelian theory is in fact crucial to an understanding of his own philosophical position. On his interpretation Hegel's philosophy ultimately rested upon a central error, one that involved the illicit identification of essence and existence, thought and reality. The German writer had endeavoured to exhibit the world, and the place of humanity within the world, in terms of an evolving sequence of logical categories that rendered its overall structure fully intelligible from the impersonal standpoint of pure reason (see Hegel, G.W.F. §§4-8). Kierkegaard disclaimed any desire to dispute the considerable ingenuity of the Hegelian metaphysic when this was regarded simply as an 'experiment in thought'. He insisted, however, that thought was not the same as reality, nor could anything real be validly deduced from it; in particular, it was altogether mistaken to suggest that changes and developments in the sphere of actual existence were assimilable to

dialectical transitions between timeless concepts - it was one thing to construct a self-contained logical or formal system, quite another to entertain the project of producing an existential one. In raising such objections, moreover, he was especially concerned to stress their relevance to Hegel's treatment of specifically human existence. The Hegelian world-picture presupposed the possibility of adopting an absolute, God-like point of view from which everything was seen as contributing to an interlocking and rationally determined totality; as a result, human nature tended to be reduced to a philosophical abstraction, the individual to a representative of the species, and the significance of a particular person's life and actions to their role in forwarding an all-encompassing historical process that overshadowed and transcended them. At the same time, Kierkegaard suggested that the notion of an impersonal 'knowing subject' of the type postulated by thinkers of the Hegelian school was symptomatic of a corresponding inclination to forget that the speculative philosopher was himself an 'existing human being' whose status and situation imposed necessary limits upon his outlook and cognitive credentials. Far from his viewpoint on the world being from nowhere within it, such a philosopher inescapably belongs to it in his capacity as a finite empirical individual who 'sleeps, eats, blows his nose' and who has 'to face the future'.

Although Kierkegaard's attitude to Hegel is most extensively displayed in the polemical references that enliven the pages of his *Concluding Unscientific Postscript*, scattered allusions to the faults and weaknesses of 'the System' also appear in many of his other writings. The number and variety of the contexts in which they occur indicate that he regarded the current vogue of Hegelianism as having more than a purely academic significance, the popularity it enjoyed at once reflecting and helping to promote a contemporary ethos in which what he termed the 'illusions of objectivity' exercised a pervasive and corrupting influence. Thus he conceived the age to be one wherein people had lost a clear sense of their identity as individuals ultimately responsible for their own characters, outlooks and modes of living. Instead it was customary for them to take refuge in the anonymity provided by membership of collective movements or trends and to envisage themselves as being inevitably circumscribed by the social roles they occupied in a manner that absolved them from personal accountability for their pronouncements or actions. In Kierkegaard's view, they had largely forgotten what 'it means for you and me and him, each for himself, to be human beings', succumbing to a 'quantitative dialectic' in which a bemused preoccupation with large-scale historical events and a passive submission to the levelling influence of 'the crowd' took precedence over the vital constituents of human life and experience - 'the inner spirit, the ethical, freedom'.

Confronted by such tendencies, Kierkegaard considered it to be a primary part of his task as a writer to challenge habits of thought that smothered spontaneous feeling and obstructed active commitment. He held that these had had a particularly deleterious effect in the religious sphere; the widespread belief that the fundamental tenets of Christianity could be rationally interpreted and objectively justified within the framework of the Hegelian system was symptomatic of a more general disposition to treat both religion and morality alike in a blandly contemplative spirit that detached them from the contexts of inward conviction and practical engagement to which they essentially belonged. With this in mind it was necessary in the first instance to 'make people aware', bringing home to them the limitations of their present condition and awakening them to the possibility of subjective self-determination and change.

3 Aestheticism and the ethical

Kierkegaard maintained that in his early writings he had aimed to arouse and enlarge the self-understanding of his readers by eschewing abstract instruction and by employing in its place an avowedly therapeutic method he referred to as 'indirect communication'. This meant delineating particular ways of life in a fashion that enabled people to grasp concretely and from within the distinct types of outlook and motivation involved, such a procedure being a characteristically literary or 'poetic' one. Not only were alternative positions imaginatively presented as if in a novel or play; the books in which this was done were attributed to different personages in the shape of pseudonymous authors or editors. He intended thereby to avoid the kind of *ex cathedra* didacticism he associated with standard philosophical texts of his time. Instead he favoured an undogmatic approach in which competing views and attitudes were 'allowed to speak for themselves', it being left to his readers to decide where they stood in relation to these and to make up their own minds about the practical conclusions to be drawn.

Either/Or was the first of Kierkegaard's works to be published under a pseudonym and was a book he later alluded to as clearly exemplifying his use of the above method. It purports to portray two radically dissimilar modes of existence, one characterized as 'aesthetic' and the other as 'ethical'. Both are presented through the medium of

allegedly edited papers or letters, the first set being ascribed to an individual referred to as ‘A’ and the second to an older man who is said to be by profession a judge. Aestheticism as exhibited in A’s loosely related assortment of papers is seen to take on a lively variety of forms and guises; among other things, it is held to find expression in the characters of legendary figures like Don Juan and Faust, and it is also illustrated by an account in diary form of a step-by-step seduction. By contrast, the position of the ethicist is set out in two somewhat prosaic letters which are addressed by the Judge to A and which include detailed critical analyses of the younger man’s motives and psychological prospects.

What did Kierkegaard understand by the categories he distinguished? From the text the aesthetic life emerges as one in which the individual is essentially concerned with exploring means to his own satisfaction and where there is a consequent absence of overall continuity in the course he follows. As has been indicated, however, the picture drawn is complex and multi-faceted. While in general outline it is suggestive of a person in pursuit of transient pleasures rather than following any long-term aim, there are passages where attention is chiefly focused on the aesthetic individual’s dependence upon unpredictable vicissitudes of mood or circumstance, and others again where emphasis is laid on his need to guard against the threats posed by ennui or melancholy. Not unexpectedly, it is the problematic possibilities inherent in A’s lifestyle that the Judge singles out for criticism in his comprehensive survey of the aesthetic position. Whereas the aestheticist typically allows himself to be swayed by what he conceives to be the unalterable constituents of his natural disposition, the ethically orientated individual is prone to look at himself in an altogether different light. Both his motivation and behaviour are responsive to a self-image ‘in likeness to which he has to form himself’, his particular aptitudes and propensities being seen as subject to the control of his will and as capable of being directed to the realization of demanding projects that reflect what he truly aspires to become. It is commitment to such projects which endows the ethical life with a coherence and self-sufficiency that its aesthetic counterpart conspicuously lacks (see Existentialist ethics).

Kierkegaard’s treatment in *Either/Or* of the aesthetic/ethical contrast is frequently thought to echo the Kantian distinction between inclination and duty (see Kantian ethics). But although there may be discernible affinities, there are also significant differences. Thus Kant’s predominantly schematic accounts of sensuous motivation are devoid of both the psychological penetration and the literary sophistication that characterize Kierkegaard’s wide-ranging portrayals of the aesthetic stance. And comparable divergences are apparent in the case of the ethical. Kierkegaard’s judge may be said to follow the German philosopher in highlighting the role of the will, underlining its independence of contingent circumstances and stressing its capacity to manage the sphere of natural inclination in a way that is conformable to the ethical individual’s paramount concerns. Yet while he shares Kant’s belief in and respect for the latter’s autonomy, he differs in not presenting moral requirements in terms of the purely formal prescriptions of practical reason. The self which it is the task of each individual to choose and develop is not an ‘abstract’ but a ‘concrete’ self; it stands in ‘reciprocal relations’ with its actual social and cultural surroundings, things like marriage, having a job and undertaking civic and institutional responsibilities being intrinsic to personal fulfilment in the requisite sense. It is implied, moreover, that such active participation in communal affairs, involving an unconstrained and inward adherence to standards presupposed by a shared form of life, reinforces the contrast already drawn with the unreflective or wayward ‘experimentalism’ typified by certain manifestations of the aesthetic outlook. Thus the Judge insists upon the conceptual exclusion from the ethical of whatever savours of the arbitrary or the merely capricious. At the same time, however, he indicates that this should not be thought of as circumscribing in any fundamental fashion the subjective freedom and independence of the individual. For although moral requirements must of necessity be treated as authoritative, they are not apprehended as deriving from a source ‘foreign to the personality’ but are instead experienced as springing or ‘breaking forth’ from the latter’s essential nature.

Even so, it is arguable that the internal tensions between individualistic and socially conformist strains discernible in the Judge’s representation of the ethical sphere cannot always be easily or satisfactorily resolved. Kierkegaard discussed one context in which they might be said to arise in a critical form when he went on to consider a way of life that constituted an alternative to the possibilities so far portrayed. This stage of existence, transcending the other two, was the religious.

4 The religious consciousness

Central to Kierkegaard’s account of religion is his treatment of the concept of faith, a treatment that throws into

relief the most distinctive features of his philosophical standpoint. There are two main areas in which these manifest themselves and in which it is the crucial inadequacies of human reason, practical as well as theoretical, that are emphasized.

The first concerns limitations in the outlook of accepted morality that make themselves felt at certain levels or junctures of experience and are held to call for what is termed a ‘teleological suspension of the ethical’. The implications of this *prima facie* puzzling notion are explored in *Fear and Trembling*, an intricately wrought study in which Kierkegaard’s pseudonymous author - Johannes *de silentio* - treats as his central theme the biblical story of Abraham and Isaac. Johannes portrays Abraham as being ostensibly called upon to set aside ethical concerns in deference to a higher *telos* or end that altogether transcends them. Such a situation is contrasted with the predicament of what he terms the ‘tragic hero’, the latter being someone who is forced to make a choice between conflicting moral requirements but who in doing so still remains within the bounds of the ethical domain. Thus although the decisions taken there may be at an agonizing cost, the fact that they can none the less be seen to conform to universally recognized norms renders them rationally acceptable to others and capable of gaining their respect. This, however, is not so in the case of Abraham, who, as a solitary ‘knight of faith’, responds to a divine command supposedly addressed to himself alone and having a content - the killing of his own son - that must inevitably strike ordinary thought as being both outrageous and incomprehensible. No attempt is made to soften the paradoxical character of such points. On the contrary, Kierkegaard’s pseudonym sets out to underline, indeed to dramatize, the disturbing nature of the demands which religious faith can impose on the life and conduct of an individual. At the same time, he takes practising churchmen severely to task for paying lip service to a phenomenon whose awesome significance they fail to appreciate, and he also criticizes contemporary theorists of religion for construing an intrinsically transcendent category in terms drawn from social and essentially secular conceptions of ethics. This was not to suggest that from a religious point of view moral standards and principles could in general be abrogated or overruled. It did mean, on the other hand, that within that perspective they took on a radically different aspect, one where they possessed a relative rather than an absolute status and where it was the individual’s own relation to God that was paramount, assuming precedence over all other considerations.

The claim that faith in the religious sense pertains to what exceeds the limits of human rationality and understanding recurs in the two subsequent writings that Kierkegaard referred to as his ‘philosophical works’ - *Philosophical Fragments* and *Concluding Unscientific Postscript*. Here, however, it is discussed within a wider setting and in connection with theoretical questions concerning the proper interpretation of religious assertions. Although once again ascribed to a pseudonym, albeit a different one, both books appeared under Kierkegaard’s imprint as their ‘editor’ and in any case may be taken to have expressed views that were basically his own. Thus in each of them it is made apparent that the author totally rejects the feasibility of trying to provide religious tenets with an objective foundation. The belief that the existence and nature of God could be conclusively established from resources supplied by pure reason might have enjoyed a long philosophical career; none the less it was demonstrably unacceptable, Kierkegaard largely echoing - though in a summary form and without attribution - some of the objections that Kant had levelled against arguments traditionally advanced by theologians and metaphysicians. Nor was he any more receptive to the suggestion that religious claims of a specifically historical character, such as those relating to the doctrines of Christianity, were susceptible to justification on straightforwardly empirical grounds; it was impossible to regard them as representing ordinary historical facts of the sort to which standard appeals to inductive inference and evidence would normally be considered appropriate. As he acknowledged, Lessing and Hamann were thinkers who in different ways had already underlined the problematic issues raised by the latter. But it was perhaps Hume’s contention in his first *Enquiry* that only a ‘miracle in his own person’, subverting all the principles of his understanding, could bring a reasonable individual to embrace the Christian religion which most strikingly foreshadowed Kierkegaard’s approach to the subject. No doubt Hume himself had intended his words to be taken in a strictly ironical sense. Even so, Kierkegaard implied that it was open to believers to look at sceptical asides of the type cited in a different light. For by exposing the vanity of attempts to encompass within its grasp matters that lay beyond the scope of reason, such remarks could be said to provide salutary reminders of what was really at stake. It was not to the spheres of impersonal judgment and dispassionate assent that the religious consciousness rightfully belonged, but on the contrary to those of individual choice and inner commitment.

5 Faith and subjectivity

Kierkegaard was certainly not alone in suggesting that writers who tried to justify religious belief on cognitive grounds were more confused about its true nature than some of their sceptically minded critics and to that extent posed a greater threat to it; indeed, Kant himself had virtually implied as much when he spoke of denying knowledge to make room for faith, as opposed to seeking to give religious convictions a theoretical foundation that could only prove illusory (see Kant, I. §4). The question arose, however, of what positive account should be given of such faith, and here Kierkegaard's position set him apart from many thinkers who shared his negative attitude towards the feasibility of providing objective demonstrations. As he made amply clear, the religion that crucially concerned him was Christianity, and far from playing down the intellectual obstacles this ostensibly presented he went out of his way to stress the particular problems it raised. Both its official representatives and its academic apologists might have entertained the hope of making it rationally acceptable to a believer, but in doing so they showed themselves to be the victims of a fundamental misapprehension. From an objective point of view, neither knowledge nor even understanding was possible here, the proper path of the Christian follower lying in the direction, not of objectivity, but of its opposite. It was only by 'becoming subjective' that the import of Christianity could be grasped and meaningfully appropriated by the individual. Faith, Kierkegaard insisted, 'inheres in subjectivity'; as such it was in essence a matter of single-minded resolve and inward dedication rather than of spectatorial or contemplative detachment, of passion rather than of reflection. That was not to say, though, that it amounted to a primitive or easy option. On the contrary, faith in the sense in question could only be achieved or realized in the course of a person's life at great cost and with the utmost difficulty (see Faith).

To understand what lay behind this claim it is important to recognize that Kierkegaard broadly distinguished between two levels or stages of development at which religious belief manifested itself. In his account of the first of these, in which he specified the criteria that any standpoint must conform to if it was to count as a religious one, he was at pains to emphasize the element of 'objective uncertainty' surrounding assertions about the transcendent, such uncertainty deriving from the absence of rational support previously alluded to. So construed, faith essentially involved personal venture or risk, preserving it being figuratively compared to 'remaining out upon the deep, over seventy thousand fathoms of water'. But to hold fast to a conviction in the face of a lack of objective justification or grounds was not the same as giving assent to something that appeared to be intrinsically contrary to reason, an 'offence' to the understanding itself. And it was in the latter terms that Kierkegaard referred to the Christian conception of the incarnation, this being an 'absolute paradox' that required the believer to 'risk his thought' in embracing its reality. Moreover, it was in the light of such a requirement that the level of faith aspired to in Christianity could be said to constitute 'the highest passion in the sphere of human subjectivity', exceeding other forms of religious belief in virtue of the unique nature of the demands it made upon an individual's mind and outlook.

According to Kierkegaard, the paradox of the incarnation lay in the notion that the eternal or timeless had entered the sphere of finite and temporal existence: this amounted to uniting contradictories in a fashion that meant a 'breach with all thinking'. Such a feature precluded treating it as if it could be vouchsafed by ordinary historical enquiry, and he set aside the scholarly pursuit of biblical research and criticism as altogether irrelevant to what was here at issue; quite apart from the specifically 'approximative' status he assigned to history as a branch of knowledge, the content of the particular 'hypothesis' under consideration defied logic in a way that contravened the principles governing any kind of accredited cognitive discipline. Furthermore, he regarded its paradoxical character as having another crucial consequence, namely, that there was no basis for the common assumption that the contemporary witnesses of what was recorded in the Gospels were in a better position to authenticate the reality of the incarnation than subsequent generations who had only the testimony of others to rely on. To suppose that in the present case the evidence of direct observation was superior to testimony was to fail to see that neither could ever function as more than an 'occasion' for belief of the sort of question. With both, a volitional leap of faith was necessary, one that involved a 'qualitative transition' from the realm of rational thought into that of the intellectually inaccessible or 'absurd'.

Kierkegaard's stress on the gap separating faith from reason, which it could need divine assistance to surmount, was reflected in the controversial account he offered of religious truth; this likewise received a subjective interpretation. Thus in a well-known passage in the *Postscript* he contrasted two distinct ways of conceiving of truth, one treating it as a matter of a belief's corresponding to what it purported to be about and the other as essentially pertaining to the particular manner or spirit in which a belief was held. And it was to the second of

these conceptions that he ostensibly referred when he declared that 'subjectivity is the truth', genuineness of feeling and depth of inner conviction being the decisive criterion from a religious point of view. Admittedly he has sometimes been criticized here for a tendency to shift from construing religious truth along the above lines to doing so in terms of the objective alternative, with the questionable implication that sheer intensity of subjective acceptance was sufficient to authenticate the independent reality of what was believed. But however that may be, it is arguable that in this context - as is often the case elsewhere - his prime concerns were conceptual and phenomenological in character, rather than epistemic or justificatory. Kierkegaard's central aim was to assign Christianity to its proper sphere, freeing it from what he considered to be traditional misconceptions as well as from the falsifying metaphysical theories to which there had more recently been attempts to assimilate it. If that meant confronting what he himself called 'a crucifixion of the understanding', the only appropriate response from the standpoint in question lay in a passionate commitment to the necessarily paradoxical and mysterious content of the Christian religion, together with a complementary resolve to emulate in practice the paradigmatic life of its founder (see Existentialist theology §§1, 2).

6 Psychological themes and influence

Kierkegaard's preoccupation with the category of subjectivity that ran like a continuous thread through his theoretical writings was integrally linked to his conception of human beings as individual and self-determining participants in the 'existential process'. The view that freedom and the possibility of change constituted fundamental conditions of human life and fulfilment was delineated in his so-called 'psychological works', *The Concept of Anxiety* and *The Sickness Unto Death*. In both books the structure of human personality is portrayed in developmental and volitional terms; individuals exist in the mode, not of being, but of becoming, and what they become is something for which they themselves are ultimately responsible. In this connection certain pervasive attitudes and emotions can be seen to possess a special significance, Kierkegaard giving priority of place to a form of anxiety or dread (*Angst*) which differed from sentiments like fear in lacking any determinate object and in being directed instead to 'something that is nothing'. Such a state of mind might manifest itself in a variety of ways, but he made it clear that his fundamental concern was with its relation to the consciousness of freedom. Thus he referred to the particular kind of dizziness, or vertiginous ambivalence between attraction and repulsion, that was liable to afflict us when, in certain circumstances, the realization dawned that there was nothing objective that compelled us to opt for one course of action rather than another; in the last analysis what we did was up to ourselves alone, freedom being said to 'look down into its own possibility' as though into a yawning abyss or void. Kierkegaard believed that the psychological phenomenon so identified had momentous consequences, not least for its bearing on the religious alternatives of sin and salvation. On the one hand, the story of Adam represented a mythical illustration of how the awakened consciousness of freedom could arouse an anxiety whose occurrence in this case was the precursor of sin. On the other hand, however, such an emotion might also arise when there was a possibility of making a qualitative leap, not into sin and alienation from God, but towards the opposite of this, namely, faith and the promise offered by Christianity. But here Kierkegaard reiterates the point that a presentiment of the difficulties and sacrifices entailed made the latter a course which there were strong temptations to resist; it followed that people were only too prone to conceal from themselves their potentialities as free beings, such self-induced obscurity serving as a convenient screen for inaction and a failure to change. Self-deception of this sort in fact formed a component of many of the varieties of spiritual despair which Kierkegaard picked out for analysis, as well as underpinning his diagnosis of some of the broader types of malaise he detected in the social and cultural climate of his time.

In his insistence upon the ultimacy of human freedom and his correlative attention to the devices and strategies whereby people may seek to protect themselves from a recognition of some of its disturbing implications, Kierkegaard anticipated themes that were taken up, albeit much later and often in an explicitly secular setting, by a number of leading twentieth-century writers (see Existentialism). Subjectivity and the primacy of the individual, the 'burden' of freedom, the contrast between authentic and inauthentic modes of existence - these and associated topics became familiar through the works of existentialist philosophers such as Jean-Paul Sartre and Martin Heidegger as well as figuring in the wider field of imaginative literature. Nor were those the only areas in which his ideas eventually made an impact. In the sphere of ethics his emphasis on radical choice indirectly contributed to the growth of non-cognitivist theories of value, while in religion his conception of faith had a profound influence on the development of modern Protestant theology, notwithstanding understandable reservations about

some of his more extreme claims regarding its paradoxical character.

See also: Hegelianism §4; Religion and epistemology; Religion and science

PATRICK GARDINER

List of works

Kierkegaard, S.A. (1962-4) *Samlede Vaerker*, ed. A.B. Drachmann, J.L. Heiberg and H.O. Lange, Copenhagen: Gyldendalske Boghandel, 20 vols.(The definitive edition of Kierkegaard's complete works.)

Kierkegaard, S.A. (1841) *Om Begrebet Ironi*, trans. H.V. Hong and E.H. Hong, *The Concept of Irony*, Princeton, NJ: Princeton University Press, 1989.(Kierkegaard's university dissertation. Though principally concerned with Socratic irony, he also discusses the concept in relation to writers associated with German romanticism.)

Kierkegaard, S.A. (1843a) *Enten-Eller*, trans. H.V. Hong and E.H. Hong, *Either/Or*, Princeton, NJ: Princeton University Press, 1987, 2 vols.(Presents contrasting life-views, one aesthetic and the other ethical, in the form of papers and letters attributed to two imaginary characters.)

Kierkegaard, S.A. (1843b) *Frygt og Baeven* and *Gjentagelsen*, trans. H.V. Hong and E.H. Hong, *Fear and Trembling* and *Repetition*, Princeton, NJ: Princeton University Press, 1983.(Two pseudonymous writings: the first an extended meditation on the implications of the biblical story of Abraham and Isaac; the second a portrayal in fictional form of psychological, ethical and religious issues raised by a problematic love-affair.)

Kierkegaard, S.A. (1844) *Philosophiske Smuler*, trans. H.V. Hong and E.H. Hong, *Philosophical Fragments*, Princeton, NJ: Princeton University Press, 1985.(A concise and complex study in which idealist and rationalist approaches to truth are contrasted with ones fundamental to Christian belief.)

Kierkegaard, S.A. (1844) *Begrebet Angest*, trans. R. Thomte and A.B. Anderson, *The Concept of Anxiety*, Princeton, NJ: Princeton University Press, 1980.(An analysis of a pervasive attitude of mind (*Angst*) chiefly evoked by an awareness of human freedom and of possibilities involving radical choice.)

Kierkegaard, S.A. (1845) *Stadier paa Livets Vej*, trans. H.V. Hong and E.H. Hong, *Stages on Life's Way*, Princeton, NJ: Princeton University Press, 1988.(The stages in the title are identified as aesthetic, ethical and religious, Kierkegaard's treatment of the last of these drawing indirectly but graphically on his own life and experience.)

Kierkegaard, S.A. (1846) *Afsluttende uvidenskabelig Efterskrift*, trans. D.F. Swenson, L.M. Swenson and W. Lowrie, *Concluding Unscientific Postscript*, Princeton, NJ: Princeton University Press, 1941.(The book most central to an understanding of Kierkegaard's philosophical position, both in its critique of objective or speculative approaches to religion and in its correlative stress on the irreducibly subjective character of authentic religious faith.)

Kierkegaard, S.A. (1849) *Sygdommen til Døden*, trans. H.V. Hong and E.H. Hong, *The Sickness Unto Death*, Princeton, NJ: Princeton University Press, 1980.(Contains intricate accounts of various forms of spiritual malaise seen as reflecting specific failures in individual self-development.)

Kierkegaard, S.A. (1850) *Indøvelse i Christendom*, trans. W. Lowrie, *Training in Christianity*, Princeton, NJ: Princeton University Press, 1944.(Kierkegaard's final pseudonymous work. Strongly critical of contemporary interpretations of Christianity, it anticipates the uncompromising virulence of his open attacks on the clerical establishment in the last year of his life.)

Kierkegaard, S.A. (1967-78) *Journals and Papers*, trans. H.V. Hong and E.H. Hong, Bloomington, IN: Indiana University Press, 7 vols.(A comprehensive and scholarly edition in English of assorted writings unpublished in Kierkegaard's lifetime.)

References and further reading

Arbaugh, G.B. and G.H. (1967) *Kierkegaard's Authorship*, Rock Island, IL: Augustana College Library.(Contains summaries and concise discussions of the main publications.)

Collins, J. (1983) *The Mind of Kierkegaard*, Princeton, NJ: Princeton University Press.(A clear and informative account of Kierkegaard's thought.)

Edwards, P. (1973) 'Kierkegaard and the "Truth" of Christianity', inP. Edwards and A. Pap (eds) *A Modern Introduction to Philosophy*, New York: The Free Press, Macmillan.(Criticizes a central aspect of Kierkegaard's account of religious belief.)

Gardiner, P. (1988) *Kierkegaard*, Oxford: Oxford University Press.(Introduction to Kierkegaard's philosophical writings and their historical background.)

Green, R.M. (1992) *Kierkegaard and Kant: The Hidden Debt*, New York: State University of New York Press.(Investigates the extent of Kantian influence on Kierkegaard's treatment of ethical and religious themes.)

Hannay, A. (1982) *Kierkegaard*, London: Routledge & Kegan Paul.(Comprehensive study of the structure and contemporary implications of Kierkegaard's thought.)
Law, D.R. (1993) *Kierkegaard as Negative Theologian*, Oxford: Clarendon Press. (Discusses Kierkegaard's place in the tradition of thinkers who have stressed the unknowability of God.)
Mackey, L. (1971) *Kierkegaard: A Kind of Poet*, Philadelphia, PA: University of Pennsylvania Press.(A subtle appraisal stressing the literary aspects of Kierkegaard's work.)
Malantschuk, G. (1971) *Kierkegaard's Thought*, ed. and trans. H.V. and E.H. Hong, Princeton, NJ: Princeton University Press.(Wide-ranging exploration of the presuppositions and interrelations of Kierkegaard's leading ideas.)
Michalson, G.E. (1985) *Lessing's 'Ugly Ditch': A Study of Theology and History*, University Park, PA: Pennsylvania State University Press.(Includes lengthy discussion of Kierkegaard's relation to Lessing.)
Rudd, A. (1993) *Kierkegaard and the Limits of the Ethical*, Oxford: Oxford University Press.(Relates Kierkegaard's contributions to problems in moral and religious philosophy to modern analytical work in these fields.)
Thompson, J. (1974) *Kierkegaard*, London: Gollancz.(A penetrating critical biography.)
Thulstrup, N. (1980) *Kierkegaard's Relation to Hegel*, trans. G.L. Stengren, Princeton, NJ: Princeton University Press.(Provides detailed account of the historical connections.)
Weston, M. (1994) *Kierkegaard and Modern Continental Philosophy*, London and New York: Routledge.(Relates Kierkegaard's critique of traditional forms of philosophy to the work of such later thinkers as Nietzsche, Heidegger and Derrida.)

Kilvington, Richard (d. 1361)

Richard Kilvington, an English philosopher and theologian, was born near the beginning of the fourteenth cantury and died in 1361. His academic career in Oxford (1320-38) was followed by diplomatic service and an ecclesiastical career that culminated in his serving as dean of St Paul's Cathedral, London. His known works (besides a couple of sermons) are commentaries or 'questions' (philosophical inquiries) regarding three works by Aristotle, a commentary on the Sentences of Peter Lombard (a standard academic requirement for theologians in the later Middle Ages) and the Sophismata. Only his Sophismata has been edited, translated and studied. An ordered collection of philosophical puzzles designed to raise and settle issues in natural philosophy and epistemology, it is one of the earliest and subtlest contributions to the literature associated with the Oxford Calculators.

Judging from his *Sophismata*, Kilvington's other works are very likely to be ingenious, with some importance for the history of philosophy and of science. However, because they survive only in medieval manuscripts and have not been studied since the Middle Ages, our present appraisal of him as a philosopher is confined to that one work.

Kilvington's *Sophismata* was written in Oxford in the early 1320s. Like other medieval works with that title, it grew out of disputations *de sophismatibus*, which had an established place in the fourteenth-century curricula of the universities of Oxford and Paris. A sophisma is a sentence puzzling simply as it stands or on the basis of some hypothesis, designed to bring some abstract issue into sharper focus - the medieval ancestor of 'The present king of France is bald', or 'The morning star is the evening star'. Medieval university disputations based on sophismata were designed to advance a student's training in logic even when the sophisma sentences presented issues in other subject matter. Thus although the first forty-four of Kilvington's forty-eight sophismata are especially concerned with problems regarding change, motive power, velocity or strength, his analysis of the puzzles is always conceptual rather than mathematical; and his solutions, though they often seem to be preparing the ground for a mathematical approach, are always found by applications of logic rather than by calculations. His interest is in *the logic of* change, velocity, knowing and doubting, or any other of the topics investigated in this treatise.

In dealing with his topics chosen from natural philosophy and epistemology, Kilvington employs almost all the devices of fourteenth-century logic, occasionally criticizing their standard use. In each of his sophismata he generates an apparent paradox by presenting what he takes to be at least plausible arguments for both sides of a contradiction, after which he undertakes to resolve the paradox. Opposition and resolution are familiar features of the standard scholastic method, in which the opposition is patently substantive, presenting immediately understandable disagreement over recognizable issues. In Kilvington's use of the method, however, the substantive issues are almost always masked by the (frequently fantastic) examples around which he develops them; as in Sophisma 16, which focuses on the Aristotelian analysis of a sphere's traversal of a proportionally divided plane in order to ask whether the dominant continuist account of space and time violates the principle of non-contradiction. Kilvington clearly intends to make quite general theoretical points about important issues via this oblique approach, although he often leaves it to the reader to recognize that those points have been made.

Kilvington seems to have arrived in Oxford soon after William of Ockham left in 1317. Ockham's teachings had generated strong opposition in the university even before its chancellor, John Lutterell, had charged him with heresy, and it seems very unlikely that Kilvington could have been unaffected by Ockham's views and the official reaction to them. In the *Sophismata* there is no explicit reference to any philosopher besides Aristotle, but the orthodox Aristotelian positions Kilvington takes there regarding physical and epistemological problems suggest that he may be counted among those who rejected Ockhamism.

Kilvington was personally associated with Thomas Bradwardine and Walter Burley. Bradwardine and Kilvington (and perhaps Burley in some of his later writings) constituted the first academic generation of the Oxford Calculators (see Oxford Calculators). William Heytesbury, the most prominent member of the second generation, seems to have been one of Kilvington's students and to have written his own *Regulae solvendi sophismata (Rules for Solving Sophsimata)* with Kilvington's *Sophismata* uppermost in his mind. Toward the end of his life Kilvington, with his friend Richard FitzRalph, played a prominent role in the anti-mendicant controversy.

See also: Aristotelianism, medieval; Language, medieval theories of; Logic, medieval

NORMAN KRETZMANN

List of works

Of Kilvington's works, only the *Sophismata* has been edited; the other works exist in medieval manuscripts, identified in the introduction of the translation and commentary on the edition. Like most other unedited medieval treatises, these last four works are variously titled in the manuscripts.

Richard Kilvington (1321-6) *Sophismata*, ed. N. Kretzmann and B.E. Kretzmann, *The Sophismata of Richard Kilvington*, Oxford: Oxford University Press for the British Academy, 1990.(The critical edition of the Latin text, with an index. A translation is also available in N. Kretzmann and B.E. Kretzmann, *The Sophismata of Richard Kilvington: Introduction, Translation and Commentary*, Cambridge: Cambridge University Press, 1990.)

Richard Kilvington (1324-6) Questions on Aristotle's *On Generation and Corruption*. (Questions on Aristotle; exists in manuscript form only.)

Richard Kilvington (1324-6) Commentary on Aristotle's *Physics*. (Questions on Aristotle; exists in manuscript form only.)

Richard Kilvington (1324-6) Questions on Aristotle's *Nicomachean Ethics*. (Questions on Aristotle; exists in manuscript form only.)

Richard Kilvington (*c*.1333) Commentary on Peter Lombard's *Sentences*. (Sometimes titled *Quaestiones theologicae*, this work is confined to just a few of Lombard's topics.)

References and further reading

d'Ors, A. (1991) '"Tu scis regem sedere" (Kilvington, S47 [48])', *Anuario Filosofico* 24: 49-74.(A careful examination of Kilvington's Sophisma 47 in the context of the doctrine of obligations, with attention to recent literature.)

Kretzmann, N. and Kretzmann, B.E. (1990) *The Sophismata of Richard Kilvington: Introduction, Translation, and Commentary*, Cambridge: Cambridge University Press. (Historical introduction, complete translation and philosophical commentary, with extensive bibliography and indexes.)

Stump, E. (1981) 'Roger Swineshead's Theory of Obligations', *Medioevo* 7: 135-74.(Shows the importance of Kilvington's work for understanding later developments in obligational disputation.)

Stump, E. (1982) 'Obligations: From the Beginning to the Early Fourteenth Century', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 315-41.(See especially 329-32, 'Richard Kilvington'. This chapter puts Kilvington's work on obligational disputation into its historical context.)

Kilwardby, Robert (d. 1279)

Robert Kilwardby is one of the most remarkable thinkers of the thirteenth century. He is the champion of the traditional approach to philosophy and theology, which developed the body of doctrines worked out by Augustine. His activity is set in the very crucial period of middle scholasticism, when the diffusion of Aristotle's philosophical system and its utilization for Christian theology caused a sharp conflict between the followers of the Patristic tradition, such as Kilwardby or the members of the Franciscan school, and the new theologians, such as Thomas Aquinas, who tried to express the contents of divine revelation within the Aristotelian paradigm. Kilwardby used all of his intellectual resources and ecclesiastical authority in fighting against this new trend and in defending Augustinianism, whose main theses (for example, a plurality of substantial forms in composite substances, the presence of seminal reasons in matter, universal hylomorphism, individuation by matter and form, a conceptual distinction between the soul and its faculties, and the necessity of divine illumination in order to grasp the eternal truths) he supports in his writings.

1 Life and works

The English Dominican Robert Kilwardby was a master of arts in the University of Paris during the years 1237-45, and later a student and master of theology at Blackfriars, Oxford. Prior Provincial of the English Dominicans in 1261, he became Archbishop of Canterbury in 1272 and a cardinal in 1278. He died in the papal service at Viterbo on 10 September 1279.

No listing of Kilwardby's writings and the surviving manuscripts is wholly satisfactory, since in many instances works have been attributed to him on insufficient grounds and their attribution has not yet been the subject of critical study. It is known that while he was teaching in the Faculty of Arts in Paris, Kilwardby wrote a set of *Sophismata grammaticalia* and *Sophismata logicalia*, and commentaries on *Priscianus minor* (Books 17 and 18 of Priscian's *Institutiones grammaticae*; the commentary on *Priscianus maior* is spurious), *De accentibus* and *Barbarismus Donati*, covering syntax, prosody, solecism and figurative speech (see Language, medieval theories of). In addition, he commented on seven works of Aristotle: *Categories*, *Perihermeneias*, *Prior* and *Posterior Analytics*, *Topics*, *Sophistical Refutations* and the first three books of the *Nicomachean Ethics* (the only ones available in Latin around 1240; Kilwardby is perhaps the earliest Latin commentator whose name can be firmly attached to an exposition of that text). His impressive series of commentaries concludes with works on Porphyry's *Isagōgē*, Boethius' *Liber divisionum* and the anonymous *Liber sex principiorum* (see Porphyry; Boethius, A.M.S.).

At Oxford around 1250, Kilwardby wrote the *De ortu scientiarum (On the Origin of Sciences)*, a classification of sciences, which is intended to be a general introduction to philosophy. His questions on the *Sentences* of Peter Lombard, probably connected with teaching at Blackfriars, was written perhaps around 1256. Around 1256-61 Kilwardby produced the treatises *De natura relationis (On Relation)*, *De tempore (On Time)* and *De spiritu fantastico (On Imagination)*.

To these works must be added two *epistulae*, the *Responsio de 43 quaestionibus Iohannis Vercellensis (Response to the 43 Questions of John of Vercelli)* and the *Epistula ad Petrum de Confleto (Letter to Peter Conflans)*. In these, Kilwardby states his position on many important issues concerning physics and ontology and clarifies his attitude towards Aristotle. The *Responsio* is Kilwardby's reply to the questions asked by John of Vercelli, Master of the Dominicans, of himself, Thomas Aquinas and Albert the Great on a number of physical and metaphysical topics that were then widely discussed. The letter to Peter Conflans is connected with the Oxford condemnation of 18 March 1277, which was begun at the instigation of Kilwardby himself. Many of the thirty condemned theses were peculiar to Thomistic thought, such as those concerning the merely potential nature of matter and the unicity of the substantial form. Peter Conflans, Archbishop of Corinth, had written a letter to Kilwardby in defence of the Thomistic theories. Kilwardby's response enables us to measure the distance and appreciate the conflict between the Augustinianism in Kilwardby's system and Aquinas's Aristotelianism, the two main philosophical trends of middle scholasticism (see Aristotelianism, medieval; Augustinianism).

2 Grammar and logic

Kilwardby's logico-linguistic theories are among the most important pre-modistic medieval accomplishments. According to him the disciplines of the *trivium* (grammar, logic and rhetoric) fall outside the Aristotelian division of speculative philosophy into physics, mathematics and metaphysics. They are concerned with signs rather than realities, and they constitute the most convenient introduction to philosophy, teaching the skills of correct speech, writing and reasoning. Grammar comes first, since it deals with the understanding and expression of ideas; then logic, which provides all sciences with a method and is corrective of the reasoning of other disciplines; and finally rhetoric, which serves ethics but is not subordinate to it. Rhetoric is a part of civic studies, because the rhetorician's aim is to devise reasoning appropriate for settling questions arising in civil disputes.

In his commentary on the *Categories*, Kilwardby states that the main difference between grammar and logic consists in the different points of view from which they look at language. The grammarian starts with the utterance and goes on to the thought; the logician starts with the thought and ends with the utterance. So the intentions or concepts that characterize grammar are applied to signs, while those characteristic of logic are applied to the essences of things. As a consequence, logic is concerned not only with speech and reasoning, but also with reality itself. However, the logician's treatment of reality is not that of the metaphysician, who studies the primary classes of being as parts of being. The logician, on the other hand, studies them in regard to their function as predicates or subjects in sentences. Consequently, logic is intended to be the theory of mental discourse concerning being and is therefore ontologically grounded in a correspondence between the structural connections in thought and the framework of reality (see Logic, medieval).

In the commentary on the *Perihermeneias*, Kilwardby specifies his theory of meaning. Like every other physical entity, a word (*dictio*) is composed of matter (the utterance, *vox*) and form. What gives a word its form is the act of signifying, which makes it a *significant* utterance, with its meanings as accidental qualifications. Meaning is impressed on the word as a whole. It depends on the mind, which can abstract the common features of individuals and thus produce a conceptual likeness of what is thought about. It is just this conceptual likeness which is conveyed to another person through utterance, while the sign itself is what affects the hearer's perception and evokes thought (see Language, Medieval Theories of).

3 Metaphysics

Kilwardby's metaphysical doctrine is very different from those that many other authors of his time drew from Aristotle's system. He is faithful to Augustine's thought, even if some understanding of the Aristotelian metaphysics of being is reflected in his theory when, for example, he distinguishes between *ens* (being) and *esse* (existing) and considers *esse* as the act of *ens* (see Existence). Furthermore, he maintains that from a metaphysical point of view the term 'being' is analogically predicated of everything. Therefore, he can regard the categories as classes of being, excluding a narrowly linguistic study of them.

Regarding universals and individuals, Kilwardby adopts a moderate realism in which the Platonist view is preserved for the divine ideas (exemplars of universals in the created order) and individuals are conceived of as their instantiations. This reconciliation of Plato and Aristotle includes a continuing existence of the likeness of the universal in the divine mind even when all the instantiating individuals have perished. However, in addition to this primary existence in the divine mind, universals have a substantial reality in their individuals. Kilwardby argues that if universals were not real (*in re*), our knowledge of the world would be empty, since knowledge is of the universal (see Realism and antirealism; Universals).

Kilwardby denies that the universal *in re* is numerically one in all its individuals. In his commentary on the *Sentences*, he proposes a modal unity based on an agreement in essence of all individuals of one type. (Thus, utilizing Boethian terminology, he maintains that the universal is that *by* which something is (*quo est*), rather than that which is (*quod est*)). This agreement in essence is considered either according to its concreteness - the existence it has in the many - or according to its essence in abstraction, as a construct of our mind. Taken in the first way, the universal form, despite its essential unity, has an existence that differs according to the different matter of the numerically distinct individuals, like the many images of one thing in the fragments of a splintered mirror. Taken in the second way, the form most properly has the status of a universal, that is, something common shared as a whole by a multiplicity of objects.

Since real universals rather than singulars properly correspond to the ideas in the divine mind, how can individuals

be obtained from such forms? In his commentary on Porphyry's *Isagōgē*, Kilwardby answers that it is the material aspect of substance that is the cause of individuation, so that the individual is constituted by its substantial components - form and matter - rather than by its properties, as Porphyry had thought. In his later questions on the *Sentences*, Kilwardby claims that the intrinsic causes of individuation are the form - which plays the active role in designating the matter as that of this or that individual and consequently designating itself as the form of the individual - and the matter, which plays the passive role of being designated as the individual's matter by receiving the form. The language of 'designation' (*signatio*), which had not been used in Kilwardby's earlier treatment of this topic, stems from Avicenna (see Ibn Sina); the influence of Bonaventure's *Sentences* commentary has also been suggested (see Bonaventure).

Between the *Isagōgē* commentary and the questions on the *Sentences*, Kilwardby's thought underwent a development in which more stress is placed on the role of form in individuation, and in which the aspects under which matter and form are to be considered as belonging to a particular individual are specified more exactly. Even in Kilwardby's later treatment of individuation, however, there is still no preoccupation with quantity and dimensions, which already feature in Aquinas' *Sentences* commentary.

Nor is this the only important difference between Kilwardby's ontology and that of Aquinas. It is clear even from an early work such as the *Isagōgē* commentary that he holds that there is some kind of materiality in angelic intelligences, thus adopting the doctrine of universal hylomorphism that he defends in his *epistulae*.

Another important difference lies in the notion of matter. According to Aquinas, who follows Aristotle, prime matter is pure potentiality; Kilwardby, following Augustine, maintains on the contrary that there are active powers in matter. Matter 'strives' for form, such a powerful appetite is an action, and there is no action without a form which performs it. Therefore active powers, or seminal reasons, must be present in matter as a kind of germinal existence of the fully actualized creatures, containing the principles of their subsequent development.

As a further consequence of the rule that there is no action without a form which performs it, there must be a plurality of substantial forms in a composite substance. Thus Kilwardby denies the unity of the substantial form and admits that even the human soul is not simple but is compounded of essentially different parts. On the other hand, he recognizes only a conceptual distinction between the rational soul and its faculties (see Soul, nature and immortality of the).

4 Ethics

Kilwardby's moral system is based on an accurate definition of the proper domain of ethics - human goods which belong to this life and are subject to rational inquiry and civic studies - and is aimed at a reconciliation of Aristotle and Augustine.

The achievement of the highest earthly human good constitutes the first step towards reaching blessedness. There are not two ends of human existence in keeping with the created/uncreated distinction, but a single highest uncreated good in God, to which human happiness is subordinated. He defines the happiness which is the end of civic studies as an act of the rational practical soul consisting in perfect studious activities or the best and most pleasurable perfections of the soul. This is the only kind of happiness Aristotle knew. Whether the soul or the whole human being is happy after death is not settled by Aristotle. In the *De ortu scientiarum*, Kilwardby draws a distinction between the happiness which most philosophers call *felicitas* and the one which Catholics call *beatitudo*. He draws a corresponding distinction between Aristotle's saying that there is happiness in completely virtuous activity and Augustine's claim that there is no possibility of being fully happy in this life without the beatific vision. According to Kilwardby, Aristotle is speaking of habitual (moral) and not theological virtues, and the philosophers who follow him give a diminished but not false account of virtue, since habitual virtue is a large part of virtue in this life, contributing further to the disposition to a more perfect life.

ALESSANDRO D. CONTI

List of works

Robert Kilwardby (*c*.1237-45) *In Donati Artem maiorem III (Commentary on Donatus)*, ed. L. Schmücker, Brixen/Bressanone: Typographia A. Weger Fund, 1984.(Commentary on the grammatical work of Donatus.)

Robert Kilwardby (*c*.1237-45) *Notule libri Prisciani de accentibus (Commentary on De accentibus)*, ed. P.O.

Lewry, 'Thirteenth-Century Teaching on Speech and Accentuation: Robert Kilwardby's Commentary on *De accentibus* of Pseudo-Priscian', *Mediaeval Studies* 50, 1988: 96-185.(Commentary on Priscian.)

Robert Kilwardby (*c*. 1250) *De ortu scientiarum (On the Origin of Sciences)*, ed. A.G. Judy, Auctores Britannici Medii Aevi 4, London: British Academy, 1976.(Kilwardby on natural philosophy.)

Robert Kilwardby (*c*.1256) *Quaestiones in librum primum Sententiarum (Questions on the First Book of the Sentences)*, ed. J. Schneider, Veröffentlichungen der Kommission für die Herausgabe ungedruckter Texte aus der mittelalterlichen Geisteswelt, Band 13, Munich: Verlag der Bayerischen Akademie der Wissenschaften, 1986.(Questions on the *Sentences* of Peter Lombard.)

Robert Kilwardby (*c*.1256) *Quaestiones in librum secundum Sententiarum (Questions on the Second Book of the Sentences)*, ed. G. Leibold, Veröffentlichungen der Kommission für die Herausgabe ungedruckter Texte aus der mittelalterlichen Geisteswelt, Band 16, Munich: Verlag der Bayerischen Akademie der Wissenschlaften, 1992.(Questions on the *Sentences* of Peter Lombard.)

Robert Kilwardby (*c*.1256) *Quaestiones in librum tertium Sententiarum (Questions on the Third Book of the Sentences)*, Part I, *Christologie (Christology)*, ed. E. Gössmann, Veröffentlichungen der Kommission für die Herausgabe ungedruckter Texte aus der mittelalterlichen Geisteswelt, Band 10, Munich: Verlag der Bayerischen Akademie der Wissenschlaften, 1982; Part II, *Tugendlehre (Virtue)*, ed. G. Leibold, Veröffentlichungen der Kommission für die Herausgabe ungedruckter Texte aus der mittelalterlichen Geisteswelt, Band 12, Munich: Verlag der Bayerischen Akademie der Wissenschlaften, 1985.(Questions on the *Sentences* of Peter Lombard.)

Robert Kilwardby (*c*.1256) *Quaestiones in librum quartum Sententiarum (Questions on the Fourth Book of the Sentences)*, ed. R. Schenk, Veröffentlichungen der Kommission für die Herausgabe ungedruckter Texte aus der mittelalterlichen Geisteswelt, Band 17, Munich: Verlag der Bayerischen Akademie der Wissenschlaften, 1993.(Questions on the *Sentences* of Peter Lombard.)

Robert Kilwardby (*c*.1256-61) *De natura relationis (On Relation)*, ed. L. Schmücker, Lenggries: L. Schmücker.(One of Kilwardby's later works.)

Robert Kilwardby (*c*.1256-61) *De tempore (On Time)*, ed. P.O. Lewry, *On Time and Imagination: De tempore, De spiritu fantastico*, Auctores Britannici Medii Aevi 9, Oxford: Oxford University Press for the British Academy, 1987.(Later work, on the nature of time.)

Robert Kilwardby (*c*.1256-61) *De spiritu fantastico (On Imagination)*, ed. P.O. Lewry, *On Time and Imagination: De tempore, De spiritu fantastico*, Auctores Britannici Medii Aevi 9, Oxford: Oxford University Press for the British Academy, 1987.(Later work, on the imagination.)

Robert Kilwardby (1271) *Responsio de 43 quaestionibus Iohannis Vercellensis (Responses to the 43 Questions of John of Vercelli)*, ed. H.-F. Dondaine, 'Le *De 43 questionibus* de Robert Kilwardby', *Archivum Fratrum Praedicatorum* 47, 1977: 5-50.(Reply to questions asked by John of Vercelli on a number of physical and metaphysical topics.)

Robert Kilwardby (1277-8) *Epistula ad Petrum de Confleto (Letter to Peter Conflans)*, ed. A. Birkenmajer as 'Der Brief Robert Kilwardbys an Peter von Konflans und die Streitschrift des Aegidius von Lessines', Beiträge zur Geschichte der Philosophie 20.5, *Vermischte Untersuchungen zur Geschichte der mittelalterlichen Philosophie*, Münster: Aschendorff, 1922, 36-69; an edition can also be found in F. Ehrle, 'Der Augustinismus und der Aristotelismus in der Scholastik gegen Ende des 13. Jahrhunderts', *Archiv für Literatur und Kirchengeschichte des Mittelalters* 5, 1889: 614-32.(Response to a letter from Conflans defending Thomistic theories.)

References and further reading

Braakhuis, H.A.G. (1985) 'Kilwardby versus Bacon? The Contribution to the Discussion on Univocal Signification of Beings and Non-Beings found in a Sophism attributed to Robert Kilwardby', in E.P. Bos (ed.) *Medieval Semantics and Metaphysics*, Artistarium supplementa 2, Nijmegen: Ingenium, 111-42.(An introduction to Kilwardby's theory of signification.)

Chenu, M.D. (1930) 'Les réponses de s. Thomas et de Kilwardby à la consultation de Jean de Verceil (1271)', in *Mélanges Mandonnet*, Paris: Vrin, vol. I, 191-222.(A comparison between Kilwardby's answers to John of Vercelli and those of Aquinas, with *excerpta* from Kilwardby's *epistula*.)

Gál, G. (1953) 'Robert Kilwardby's Questions on the *Metaphysics* and *Physics* of Aristotle', *Franciscan Studies* XIII: 7-28.(On the attribution to Kilwardby of a set of metaphysical and physical questions extant in ms. 509 of Gonville and Caius College, Cambridge.)

Lewry, P.O. (1975) 'The Commentary on *Priscianus Maior* Ascribed to Robert Kilwardby: The Problem of the Authorship', *Cahiers de l'Institut du Moyen-âge Grec et Latin* XV: 12*-17*.(Demonstrates the spuriousness of

the commentary on books 1-16 of Priscian’s *Institutiones grammaticae*.)

Lewry, P.O. (1978) ‘Robert Kilwardby’s Writings on the *Logica Vetus* Studied with Regard to Their Teaching and Method’, D.Phil. thesis, Oxford University.(The standard exegesis of Kilwardby’s contribution to the logic of terms and proposition, with a careful discussion of the authorship of his commentary on the *Logica vetus*.)

Lewry, P.O. (1981a) ‘Robert Kilwardby on Meaning: a Parisian Course on the *Logica Vetus*’, in J.P. Beckmann *et al*. (eds) *Sprache und Erkenntnis im Mittelalter*, Miscellanea mediaevalia 13.1, Berlin: de Gruyter, 376-84.(A survey of Kilwardby’s theory of meaning.)

Lewry, P.O. (1981b) ‘The Oxford Condemnation of 1277 in Grammar and Logic’, in H.A.G. Braakhuis, C.H. Kneepkens and L.M. de Rijk (eds) *English Logic and Semantics From the End of the Twelfth Century to the Time of Ockham and Burleigh*, Artistarium supplementa 1, Nijmegen: Ingenium, 235-78.(An account of the grammatical and logical implications of the Oxford condemnation of 1277.)

Lewry, P.O. (1983) ‘Robert Kilwardby on Imagination: the Reconciliation of Aristotle and Augustine’, *Medioevo* IX: 1-42.(A clear account of Kilwardby’s theories of soul and knowledge.)

Lewry, P.O. (1986) ‘Robert Kilwardby’s Commentary on the *Ethica nova* and *Vetus*’, in C. Wenin (ed.) *L’homme et son univers au moyen-âge*, Philosophes médiévaux 27, Louvain-la-Neuve: Editions de l’Institut Superieur de Philosophie, 799-807.(A summary of Kilwardby’s ethical system, with a discussion of the authorship of the *Nicomachean Ethics* commentary.)

Knowledge and justification, coherence theory of

Coherence theories of justification represent one main alternative to foundationalist *theories of justification. If, as has usually been thought, possessing epistemic justification is one necessary condition (along with truth and perhaps others) for a belief to constitute knowledge, then a coherence theory of justification would also provide the basis for a coherence theory of knowledge. While some proponents of coherence theories have restricted the scope of the theory to empirical justification, others have applied it to all varieties of epistemic justification. (There are also coherence theories of meaning and of truth, as well as coherence theories of ethical or moral justification.)*

The initial contrast between coherence theories and foundationalist theories arises in the context of the epistemic regress problem. *It is obvious that the justification of some beliefs derives from their inferential relations to other, putatively justified beliefs, and that the justification of these other beliefs may depend on inferential relations to still further beliefs, and so on, so that a potential regress of epistemic justification looms, with scepticism as the threatened outcome. The foundationalist solution to this problem is that one arrives sooner or later at basic or foundational beliefs: beliefs that are epistemically justified, but whose justification does not derive from inferential relations to any further beliefs and so brings the regress to an end. The defining tenet of a coherence theory of justification is the rejection of this foundationalist solution, the coherentist insisting that any belief (of the kinds to which the theory is applied) depends for its justification on inferential relations to other beliefs and eventually to the overall system of beliefs held by the believer in question. According to the coherentist, the justification of this system of beliefs is logically prior to that of its component beliefs and derives ultimately from the* coherence *of the system, where coherence is a matter of how tightly unified or interconnected the system is by virtue of inferential connections (including explanatory connections) between its members.*

Contrary to what this might seem to suggest, coherence theories do not deny that sensory observation or perception plays an important role in justification. What they deny is that this role should be construed in a foundationalist way, insisting instead that the justification of observational beliefs ultimately derives also from considerations of coherence. Specific coherence theories may also add other requirements for justification, thereby departing from a pure coherentism, while still avoiding foundationalism.

While the idea of a coherence theory has often played the role of a dialectical foil, developed theories of this kind are relatively rare and are often in serious disagreement among themselves. In this way, coherentism is much less a unified view with standard, generally accepted features, than is foundationalism.

1 History

In contrast to foundationalism, the coherence theory of justification is a relatively recent innovation in the history of philosophy. Although it is possible, albeit with some strain, to construe Spinoza and Kant as advocating versions of coherentism, the first relatively clear-cut coherentist positions are those of the late nineteenth and early twentieth century absolute idealists, especially F.H. Bradley (1914) and Bernard Bosanquet (1920) (see Bradley, F.H.; Bosanquet, B.). Unfortunately, however, the views of these philosophers are marked by a pervasive failure to distinguish epistemological and metaphysical issues, making it hard to separate their coherence theories of justification from their distinct, though related, advocacy of coherence theories of the nature of truth. (A more recent version of essentially the same position, in which this distinction is clearly drawn, is found in the work of Brand Blanshard (1939).)

Coherentism was also advocated in the 1930s by some of the logical positivists, mainly Otto Neurath (1933) and Carl Hempel (1934), in response to the foundationalist views of Moritz Schlick (1934) (see Neurath, O.; Hempel, C.G.; Schlick, F.A.M.). Neurath identifies coherence with mere logical consistency; he also, while retaining something like a justificatory appeal to observation, in effect identifies observational beliefs solely by reference to their content. He thus has no apparent response to what is perhaps the most central and obvious objection to coherence theories: that there will always be indefinitely many different coherent systems between which a coherence theory will provide no basis for a reasoned choice (see further below). Hempel avoids this problem to some extent by simply identifying observational beliefs as those beliefs with the right sort of content that are accepted by ‘the scientists of our culture circle’, but is able to offer no real rationale for such an identification. He

also, like the idealists, fails to distinguish in any clear way between a coherence theory of justification and a coherentist account of the nature of truth.

More recent coherentist positions, in contrast, generally repudiate the coherence theory of truth entirely and are more explicitly and narrowly epistemological in their character and motivation. The main arguments offered in their favour almost always derive from perceived objections to foundationalism, perhaps the central one being the charge that the foundationalist can account for the status of the allegedly basic or foundational beliefs as genuinely justified (in the sense of there being some reason or basis for thinking them to be true) only by appealing to justificatory premises of some sort and so destroying the status of such beliefs as foundational (see Foundationalism). Thus coherentists insist that there is no way to appeal for justification to anything outside of one’s system of beliefs because any such supposed source of justification would have to be apprehended by the person in question in a belief or belief-like state before it could play any justificatory role, and then it would be the belief rather than the external item that was the immediate source of justification.

As this suggests, coherentist positions are virtually always internalist rather than externalist in character, in that they insist that the basis for epistemic justification must be cognitively accessible to the believer in question; while an externalist version of coherentism is theoretically possible, it would have little philosophical point, since a foundationalist view would be vastly more straightforward if externalism were otherwise acceptable (see Internalism and externalism in epistemology).

These recent coherentist views differ from each other in a wide variety of ways, and often seem to have little in common beyond their rejection of foundationalism and their invocation in some fashion of the idea of coherence (and indeed there is often room for doubt in a particular case about how thoroughgoing the former of these two aspects really is). Coherentism is one ingredient, though never developed in a fully systematic way, in the comprehensive and difficult philosophical system of Wilfrid Sellars (1963) (see Sellars, W.S.). The epistemological position of W.V. Quine is also frequently described as coherentist in character, though other features of Quine’s position, especially his claim that epistemology should be naturalized (reduced to psychology), make it difficult to decide whether his view is genuinely a version of coherentism, as opposed to a qualified version of foundationalism (see Naturalized epistemology). More overtly coherentist positions have been advocated by Gilbert Harman (1973) (influenced especially by Quine), Nicholas Rescher (1973, 1974, 1977), Keith Lehrer (1974, 1990), and Laurence BonJour (1985) (influenced especially by Sellars).

As the foregoing suggests, coherence theories first arise as dialectical alternatives to foundationalism, rather than as views that are claimed to be initially plausible on their own. Their defence and elaboration must confront a number of standard problems and objections, with which any such view must seemingly deal in some fashion, and it is around these that the present entry is organized.

2 The regress problem and non-linear justification

The first standard problem arises from the epistemic regress problem itself. If foundationalism is repudiated (and if a genuinely infinite regress of justification is also rejected as psychologically impossible and in any case tantamount to scepticism), then the only remaining possibility for the outcome of the initial regress of epistemic justification seems to be a circle in which the chains of justification eventually loop back upon themselves. Incautious advocates of coherentism have sometimes seemed to endorse the idea that such a result is acceptable if only the circles are ‘large enough’. But the obvious objection to circular chains of justification, to which the size of the circle seems irrelevant, is that they involve circular reasoning and hence have no genuine justificatory force. This is essentially the reason that foundationalists give for rejecting the coherentist alternative and taking the regress problem to constitute a decisive argument for foundationalism.

Perhaps the most standard coherentist response to this issue, stemming originally from Bosanquet (1920), is to reject the idea, implicit in most presentations of the regress problem, that relations of justification must involve a linear, asymmetrical order of dependence among the beliefs in question. They insist instead that justification, when properly understood, is ultimately holistic and non-linear in character, with all of the beliefs in the system standing in relations of mutual support, but none being epistemically prior to the others. In this way, it is alleged, any true circularity is avoided. Such a view amounts to making the system itself the primary unit of justification, with its component beliefs being justified only derivatively, by virtue of their membership in an appropriate sort of system.

And the property of the system, in virtue of which it is justified, is of course specified as coherence.

3 The concept of coherence

But what exactly is coherence? A second obvious problem for a coherence theory is to explicate and clarify this central concept. Intuitively, coherence is a matter of how the beliefs in a system of beliefs fit together or dovetail with each other, so as to constitute one unified, organized, and tightly structured whole. And it is clear that this fitting together depends on a wide variety of logical, inferential and explanatory relations among the components of the system. But spelling out the details of this idea, particularly in a way that would allow unproblematic assessments of comparative coherence, turns out to be extremely difficult, in part at least because of its obvious dependence on more specific and still unsettled topics, such as induction, confirmation, probability and explanation.

The strongest and most demanding conception of coherence, advocated by the idealists, specifies a coherent system of beliefs as one in which each member entails and is entailed by all of the others. It seems clear, however, that this strong conception is both unrealizable by any actual system of beliefs imaginable and also of dubious cognitive value, since it would seem to make all of the beliefs but one redundant and dispensable. (These problems may be mitigated somewhat, though certainly not eliminated, by remembering that the idealists have a quite broad conception of entailment, one in which, for example, relations of nomological necessity are regarded as a kind of entailment).

At the opposite extreme, it seems equally mistaken to follow Schlick (1934) and some others in identifying coherence with mere logical consistency, since the beliefs of a logically consistent system might be entirely unrelated to each other, thus yielding no real degree of mutual support and no apparent reason for thinking that any of them are true. Somewhat more surprisingly, it also seems to be a mistake to make complete logical consistency even an absolutely *necessary* condition for any degree of coherence, as many coherentists have done. In light of such things as the preface paradox and general human fallibility, this would probably mean that few if any actual systems of belief are coherent to any degree at all, a result that seems unacceptably paradoxical (see Paradoxes, epistemic).

If there is a tenable conception of coherence along these general lines, it must seemingly fall somewhere between the two extremes just discussed. Coherence will be a matter of degree, with logical consistency being a highly relevant but not absolute criterion. Coherence will also require a high degree of inferential interconnectedness in the system, involving relations of necessitation, both strictly logical and otherwise, together with probabilistic connections of various kinds. An important aspect of this is what might be called probabilistic consistency, that is, the absence of relations between beliefs in the system in virtue of which some are highly unlikely to be true in relation to others. A further important ingredient of coherence that is much emphasized in recent discussions is the presence of explanatory relations among the components of the system, thus reducing the degree to which the beliefs of the system portray unexplained anomalies. (If ‘inference to the best explanation’ is accepted as one species of inference, then such explanatory relations can be viewed as a kind of inferential relation - see Inference to the best explanation.) Indeed, some positions such as that of Harman (1973), and perhaps also Sellars (1963), go so far as to virtually identify coherence with the presence of such explanatory relations.

The foregoing is an approximate account of the historically standard conception of coherence. While some proponents of coherentism have employed essentially this conception, others have in effect devised more idiosyncratic conceptions of coherence, conceptions whose connection to the historical concept is often quite tenuous. In particular, Rescher (1973, 1977) in fact employs both the standard conception of coherence, for certain purposes, and also a quite different concept that involves forming maximally consistent subsets of initially conflicting ‘data’ or ‘truth-candidates’, and then choosing among these subsets in a variety of ways that involve no appeal to standard coherence. And Lehrer (1974, 1990) has offered two subtly different versions of a general view that defines coherence in relation to the believer’s own subjective conception of probability or relative likelihood of truth; for a belief to cohere with the person’s system of beliefs is roughly for it to be judged to be more probable or more reasonable than any relevant competitor.

The precise nature of coherence remains a largely unsolved problem. It is important to see, however, that difficulties in this area cannot yield anything like a decisive argument against coherence theories and in favour of

their foundationalist rivals. This is so because the concept of coherence, or something so similar to it as to be capable of playing essentially the same role, is also an indispensable ingredient in virtually all foundationalist theories: coherence must seemingly be invoked to account for the relation between the basic or foundational beliefs and other non-foundational or 'superstructure' beliefs, in virtue of which the latter are justified in relation to the former. For this reason, giving an adequate account of coherence should not be regarded as exclusively or even primarily the responsibility of coherentists, despite the central role that the concept plays in their position.

4 Coherence and observation

As mentioned above, few if any coherentists have wished to deny the seemingly obvious fact that sensory observation or perception plays a crucial role in justification (although they have not always been fully explicit on this point). It is thus incumbent on a coherence theory to explain how such observation can be construed in a non-foundationalist way. The central idea is that a belief that is produced by the senses, rather than being arrived at inferentially, might still depend for its justification on coherence with the background system of beliefs. But it is crucial here that the justification in question should still depend also in some way on the fact that the belief was a result of perception, since justification that depended only on the coherence of the belief's propositional content with the rest of the cognitive system would make the observational status of the belief irrelevant.

One way to develop this idea is to focus on the fact that observational or perceptual beliefs are *cognitively spontaneous*; they simply strike the observer in an involuntary, coercive, non-inferential way, rather than as a product of any sort of inference or other discursive process, whether explicit or implicit. That a belief has this status, however, says nothing so far, according to the coherentist, about whether or how it is justified. Indeed, there is no reason to think that all cognitively spontaneous beliefs *are* justified, or even necessarily that most of them are, since the category would include hunches and irrational spontaneous convictions, as well as beliefs resulting from perception. But suppose that, as seems to be the case with most ordinary systems of belief, the system includes a belief to the effect that specific kinds of cognitively spontaneous beliefs (identified by their general subject matter, by their apparent mode of sensory production as reflected in the content of the belief, and by concomitant factors of various kinds) are, under specified (or perhaps 'normal') conditions, highly likely to be true. It then becomes possible to give a justifying reason for such a belief that appeals to its status as cognitively spontaneous and putatively observational, but still does so in a way that depends on the coherence with the background system of beliefs of the claim that a belief of this kind and produced in this way is true. Such a belief would be arrived at non-inferentially, but still justified by appeal to inference relations and coherence. (This view of observation is most explicit in BonJour (1985), but something like it seems implicit in Blanshard's (1939) talk of 'beliefs about the technique of acquiring beliefs', in Sellars's (1963) talk of 'language-entry transitions', in Quine's talk of the 'observational periphery' of the 'web of belief', in Rescher's (1973, 1974, 1977) idea of 'data', and in Lehrer's (1974, 1990) discussion of a person's trustworthiness in acquiring certain kinds of information).

The foregoing provides at best only the beginning of a coherentist account of observation, leaving various problems to be solved that can only be touched on here. First, the other beliefs needed to give a justifying reason for a particular observational belief must themselves be justified in some fashion, without relapsing at this point into foundationalism. These beliefs will include at least (1) beliefs about the conditions; (2) the general belief about the reliability of the kind of cognitively spontaneous belief in question; and (3) beliefs about the occurrence of that particular belief, including the belief that it was indeed cognitively spontaneous. The justification for (1) will presumably have to include other observational beliefs, themselves justified in the same general fashion, so that any case of justified observation will normally or perhaps always involve a set of mutually supporting observations. The justification for (2) will appeal inductively to other cases of correct observation, as judged from within the system, as well as to more theoretical reasons for thinking that beliefs of the kind in question are generally produced in a reliable way. The justification for (3) will appeal to introspective beliefs, themselves constituting a species of observation, and ultimately to the believer's comprehensive grasp of their overall system of beliefs - a grasp whose status poses one of the main problems to be considered below.

Second, it is not enough for the justification of an observational belief that a reason of the foregoing sort should merely be present in a person's system of beliefs, since such an individual might completely fail to notice that this was so, and might hold the belief on some other basis or for no reason at all. Thus even though the observational belief is not arrived at by inference, the availability of the inferential justification in question, even if never

explicitly rehearsed, must be the reason why the believer continues to accept the belief and to appeal to it for further purposes. A full account of coherentist observation would have to spell out exactly what this requirement amounts to and how it can be satisfied.

Third, the bare possibility of coherentist observation seems insufficient to accommodate the role that observation plays in our cognitive lives. Given the convictions that observation is not only possible but pervasive and that an appeal to observational evidence, whether direct or indirect, is essential for the justification of at least contingent beliefs about the world, an intuitively adequate coherence theory must somehow *require* and not just allow that a substantial observational element should be present in any justified system that includes such contingent beliefs. A view that insisted on such a requirement would thereby depart from a pure coherence theory, but might still avoid foundationalism if the coherentist account of observation is otherwise successful. (Such a requirement is relevant to several of the objections examined below.)

5 The standard objections

In considering objections to coherence theories, we may begin with the three that are historically most standard and familiar. The first of these is what is commonly referred to as 'the isolation problem' or 'the input objection': an account of justification that appeals entirely to coherence within a system of beliefs seems to have the consequence that the justificatory status of the beliefs in the system will not depend in any way on the relation of the system to the world that it purports to describe, or on any sort of information derived from that world. This would seem to mean in turn that the truth of the component beliefs, if they happened to be true, could only be an accident, and thus that there is no reason to think that they are true and so no epistemic justification. The coherentist account of observation sketched above, if it can be successfully fleshed out, provides the beginning of an answer to this objection by showing how observational beliefs that are apparently generated by the world might none the less be given a coherentist justification. In this way, a coherent system that involves a putatively observational component will at least seem from the inside to have input from the world and thus not to be isolated. Whether this seeming is likely to be veridical will depend, however, on the more general issue, discussed below, of whether and why coherentist justification should be regarded as conducive to finding the truth.

The second familiar objection, already briefly alluded to earlier, is what may be called the alternative coherent systems objection: even given a relatively strong account of coherence, there will still be indefinitely many different possible systems of beliefs, each of which is as internally coherent as the others, and so all of which will be equally justified on a coherentist view - surely an absurd result. The response to this objection also depends crucially on the idea of observation. If, as suggested earlier, it will be a requirement for justification in an adequate coherence theory that there be a substantial observational component (that is, a substantial proportion of cognitively spontaneous beliefs that the system itself certifies as likely to be true and hence worthy of being accepted), then such alternative systems can no longer be freely invented and it is no longer obvious why they should be thought to exist. Only a system that is actually accepted and employed in cognitive practice can contain cognitively spontaneous beliefs and thus satisfy the requirement for observation. There is no way to guarantee that the acceptance of such beliefs will not lead quickly to incoherence in an arbitrarily devised system, even if it is initially coherent. (As this suggests, it is coherence over a period of time and not just at a moment that is ultimately the basis for justification in all coherence theories that have been seriously advocated.)

The third of the standard objections is in effect a challenge to the coherentist to give a reason for thinking that adopting beliefs on the basis of coherentist justification is likely to lead to believing the truth. Different coherentists give very different responses to this crucial question, each problematic in its own way. These can only be briefly sketched here: (1) The absolute idealists in effect solve the problem by adopting a coherence theory of truth as well, thus reducing the gap between coherentist justification and truth (though only Blanshard is very explicit about this strategy). On such a view, truth is essentially identified with long-run justification, making it relatively easy to argue that seeking justified beliefs is likely to lead eventually to finding true ones - but at the significant cost of adopting an extremely implausible conception of truth. (2) Rescher (1973, 1974, 1977) attempts to give a pragmatic argument to the effect that the practical success which results from the employment of the coherent system makes it likely that the beliefs of the system are at least approximately true (in the sense of corresponding to independent reality). Unfortunately, however, the need for justification for the claims of practical success, which must presumably also be coherentist in character, threatens the project with vicious circularity. (3)

BonJour (1985) attempts to give an a priori 'metajustificatory argument', relying on a rationalist and foundationalist conception of a priori knowledge, for the conclusion that a system of beliefs that remains coherent over the relatively long run, while receiving apparent observational input, is likely to be approximately true (again in the correspondence sense of truth). The main reason offered is that only approximate truth could explain continued coherence in the face of new observations. In addition to defending the general account of a priori justification presupposed, such an approach must also claim that sceptical explanations of one's beliefs (for example their being produced by a Cartesian demon), are a priori less likely than the preferred explanation of correspondence with reality, a claim that many have found highly implausible. (4) Lehrer's approach is to construct alternative conceptions of justification that involve the hypothetical replacement of erroneous beliefs in a person's system of beliefs with their corrected alternatives, and then require that the person's initially justified beliefs remain justified after such replacements in order for such beliefs to constitute knowledge. The main difficulty here is that such an approach seems to concede that 'personal justification' - the sort of justification which exists before the hypothetical replacements - is not in itself conducive to finding the truth, even though such personal justification is the only sort that the believer is in general ever actually aware of.

6 The problem of access

In addition to the foregoing objections, there are a number of further problems with which an adequate coherence theory would have to deal. Perhaps the most urgent of these is that of whether coherentist justification is accessible to the believer in the way that it must be if an internalist position is to result. Assuming for the moment, as is the case with all the positions discussed here, that coherentist justification is taken to require coherence with the believer's entire system of beliefs, then there are three aspects to this problem: (1) whether the believer has adequate access to their system of beliefs; (2) whether the believer has an adequate grasp of the concept of coherence; and (3) whether the believer is able to apply the concept of coherence to their system of beliefs in a way that will yield a definite assessment. All these aspects pose serious problems, and (3) in particular is anything but trivial, even given satisfactory solutions to (1) and (2). But (1) is the most difficult and will accordingly be the main focus here.

A believer's access to their own system of beliefs is in fact seriously problematic in two quite different ways. First, there is the problem of the epistemological status of the result of such access if it were achieved, which we may think of as a reflective meta-belief describing the entire contents of the system. Such a meta-belief would be clearly contingent and empirical and hence one that on any coherence theory of the sort under consideration here ought to be itself justified by appeal to coherence. But since any coherentist justification that is to be accessible to the believer must appeal to such a meta-belief in order to characterize the system of beliefs with which the belief to be justified must cohere, a coherentist justification of that meta-belief itself appears to be totally and irrevocably circular (and no appeal to non-linear justification will help here, since what is being explained is how the very sort of non-linear justification advocated by the coherentist is possible).

The most explicit discussion of this issue is given by BonJour (1985), who responds by invoking what he calls 'the Doxastic Presumption'. The idea is that coherentist justification must *presume* that the believer's grasp of their overall system of beliefs is at least approximately correct (small corrections being possible by appeal to coherence). This has the consequence that the resulting justification is contingent upon the presumed correctness of this grasp, and hence that there is no possible answer on the part of a coherence theory to the specific variety of scepticism that questions whether this presumption is indeed correct. This has seemed to many to be a very drastic result, but it is unclear what the alternative might be, so long as foundationalism is eschewed.

Even if the foregoing issue were resolved in a satisfactory way, there is still the second aspect of the present problem: the quite sticky issue of whether ordinary believers ever in fact possess or could possess anything like the reflective grasp of the entire contents of their systems of beliefs that a coherence theory seems to require. On this issue, it seems likely that a coherence theory will have to concede that ordinary cases of justification are at best only an approximation, and perhaps a fairly distant one, to the ideal justification that a coherence theory portrays.

One further problem is worth a brief mention. If, as is almost always the case, a coherence theory appeals to coherence over the relatively long run, then the issue arises of how the memory beliefs upon which any access to the fact of continued coherence must seemingly rely, are themselves to be justified. Many philosophers have

offered coherence theories of the justification of memory beliefs, but such a view is again threatened with vicious circularity if the only reason for thinking that coherentist justification is conductive to truth - and so that the memory beliefs in particular are true - relies on coherence over time, and so on those very memory beliefs themselves.

See also: Justification, epistemic; Knowledge, concept of; Truth, coherence theory of; Truth, correspondence theory of

LAURENCE BONJOUR

References and further reading

Bender, J. (ed.) (1989) *The Current State of the Coherence Theory*, Dordrecht: Kluwer. (Critical essays on the coherence theories of Lehrer and BonJour, with replies.)

Blanshard, B. (1939) *The Nature of Thought*, London: Allen & Unwin.(An updated version of the absolute idealist coherence theory, referred to in §§1 and 4.)

BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.(The author's version of coherentism, discussed especially in §§4 and 6 above but containing elaborations of material in other sections as well; also contains an appendix discussing the views of the positivists, the absolute idealists, Lehrer, and Rescher.)

Bosanquet, B. (1920) *Implication and Linear Inference*, London: Macmillan.(Bosanquet's critique of the linear view of justification and defence of coherentism, referred to above in §§1 and 2.)

Bradley, F.H. (1914) *Essays on Truth and Reality*, Oxford: Oxford University Press. (Essays criticizing the correspondence theory of truth and defending coherentism with respect to both justification and truth; referred to in §1 above.)

Harman, G. (1967) 'Quine on Meaning and Existence 1', *Review of Metaphysics* 21 (1): 124-51.(An account of Quine's theory of knowledge, including a discussion of his coherentism, as discussed in §§1 and 4 above: Quine's own works discuss epistemological themes in passing, but there is no single unified discussion that can be cited.)

Harman, G. (1973) *Thought*, Princeton, NJ: Princeton University Press.(A coherentist position influenced by Quine, referred to in §1 above.)

Hempel, C. (1934) 'On the Logical Positivist's Theory of Truth', *Analysis* 2 (4): 49-59.(A reply to Schlick's 1934 work; referred to in §1 above.)

Lehrer, K. (1974) *Knowledge*, Oxford: Oxford University Press.(The first statement of Lehrer's coherentism, somewhat superseded by his 1990 work, but still worth comparing.)

Lehrer, K. (1990) *Theory of Knowledge*, Boulder, CO: Westview Press.(A revision of Lehrer (1974); referred to in §§1, 3 and 4 above.)

Neurath, O. (1933) 'Protokollsätze', *Erkenntnis* 3; trans. as 'Protocol Sentences', in A.J. Ayer (ed.) *Logical Positivism*, New York: Free Press, 1959, 199-208.(Neurath's early, rather simplistic version of coherentism; referred to in §1 above.)

Rescher, N. (1973) *The Coherence Theory of Truth*, Oxford: Oxford University Press.(Should be read together with Rescher (1977); referred to in §§1, 3 and 4 above.)

Rescher, N. (1974) 'Foundationalism, Coherentism, and the Idea of Cognitive Systematization', *Journal of Philosophy* 71 (19): 695-708.(A useful summary of Rescher's coherence theory.)

Rescher, N. (1977) *Methodological Pragmatism*, New York: New York University Press. (Restates and completes the view expressed in Rescher's 1973 work.)

Schlick, M. (1934) 'Uber das Fundament der Erkenntnis', *Erkenntnis* 4; trans. as 'The Foundation of Knowledge', in A.J. Ayer (ed.) *Logical Positivism*, New York: The Free Press, 1959, 209-27.(A succinct statement of Schlick's foundationalism, which was the target of both Neurath and Hempel; referred to in §§1 and 3 above.)

Sellars, W. (1963) *Science, Perception and Reality*, London: Routledge & Kegan Paul.(A collection of papers. Chapters 5 and 11 present some of the coherentist ideas discussed in §§1 and 4 above.)

Knowledge by acquaintance and description

The attempt to distinguish knowledge by acquaintance from knowledge by description is most closely associated with Bertrand Russell. The distinction is also crucial to one way of trying to develop a plausible foundationalist theory of justification and knowledge. According to Russell one can distinguish the two kinds of knowledge in terms of their respective objects. *Put crudely, one has knowledge by acquaintance of* things*, and one has knowledge by description of* propositions *(representations of reality that are either true or false). But this crude characterization of the two kinds of knowledge is misleading. Russell also seemed to believe that one can have knowledge by acquaintance of properties and even facts (where a fact is a complex consisting of a thing's exemplifying a quality or standing in a relation to another thing). The distinction, then, might be better put in terms of a kind of knowledge which has as its object something that is neither true nor false (knowledge by acquaintance) and a kind of knowledge which has as its object a bearer of truth value (knowledge by description).*

According to Russell, all knowledge of truths ultimately rests on knowledge by acquaintance. The traditional foundationalist in epistemology holds that although I can know one truth by inferring it from something else I know, not everything I know can be inferred in this way. We can avoid a regress of knowledge by holding that at least some truths are known as a result of direct awareness of or acquaintance with those aspects of the world that make the corresponding propositions true.

1 Acquaintance

To distinguish between two *objects* of knowledge is not by itself to say anything very informative about the nature of the respective knowing relations. What is it to know something (a thing, a property or a fact) by acquaintance? Unfortunately, a proponent of the view that such knowledge exists is likely to take the relation of acquaintance to be a primitive. To be sure one can invoke metaphors. When one knows a particular shade of colour by acquaintance, for example, the colour is directly and immediately 'before' one's consciousness. There is nothing 'between' the colour and oneself. By contrast, one might know truths about Julius Caesar but one's access to such truths is only through inference from other things one knows about the contents of history books and the like. A huge expanse of time stands 'between' us and the man, Caesar. The spatial metaphors of 'before' and 'between' are, however, just that - metaphors. As such they are as misleading as they are helpful. More often than not, philosophers who think that there is such a thing as acquaintance do not even take the objects of acquaintance to be in physical space. Classical acquaintance theorists were usually either dualists who took the objects of acquaintance to be properties of a non-spatial mind, or sense-datum theorists for whom sense-data, if in space at all, were in a two-dimensional space. The relational concepts of 'before' and 'between' are clearly most at home in the conceptual framework of three-dimensional space.

If one cannot define the relation, one might attempt to make the concept of acquaintance clear by giving examples of things with which one can be acquainted. Even proponents of the view that there is such a relation as acquaintance do not, however, agree on what one should take the objects of acquaintance to be. The radical empiricists held that the prime candidates for objects of acquaintance include the contents of consciousness (see Empiricism §2). Thus when I am in severe pain, the pain I feel is said to be directly and immediately 'before' my mind. The contents of my visual field, the sensory character of tactile, olfactory, auditory and gustatory sensations, my thoughts and emotions are all held by most radical empiricists to be items with which I can be directly acquainted. In addition, one might hold that the mind is capable of directly encountering such abstract entities as numbers and universals (properties which can be exemplified in more than one thing and are often thought to exist even when not exemplified by anything). The radical empiricists typically denied that one can be directly acquainted with physical objects, items in the past or items in the future. To return to the metaphor, between us and the physical world is the 'veil' of subjective fleeting sensation. Our access to the world of physical objects is always through belief in propositions which can be known only as a result of inference from what we can know about the character of the sensations with which we are acquainted. Likewise, the past (at least the distant past) can be known only through present memory experience, and the future can be known only through our knowledge of the past and present.

2 Additional criteria of acquaintance, and an initial problem

How is one to know whether or not one is acquainted with something? Here the plot gets difficult to follow. According to at least some philosophers, one is directly acquainted with something only if one can know without inference that the thing in question exists or exemplifies some contingent property. The test of whether or not one can know without inference that something exists is itself sometimes described in terms of the conceivability of error. On this view, one sign that I cannot be directly acquainted with the physical table before me now is that my evidence for believing that the table exists does not logically guarantee its existence. My evidence is said to consist of what I know about my sensations, knowledge that would be no different were I dreaming or hallucinating the table's existence. I am directly acquainted with my severe pain because the justification I now have for believing that I have this pain precludes the possibility of my being wrong - I cannot hallucinate the existence of severe pain.

Some critics complain that making the impossibility of error the mark of direct acquaintance is in tension with other things philosophers want to say about the relation of acquaintance. Specifically, acquaintance is not even supposed to involve bearers of truth value. Let us explore this alleged tension.

In addition to giving examples of things with which we can be acquainted, Russell and others also gave negative characterizations of knowledge by acquaintance. When one knows a thing by being acquainted with it, that knowledge does not involve the application of concepts to the thing - one does not categorize the thing. Presumably, knowledge by acquaintance does not even presuppose linguistic or conceptual capacity on the part of the knower. To know a *truth*, on the other hand, one must explicitly or implicitly apply a concept to a thing or property. To think that this colour with which I am acquainted is red is to apply the concept red to the thing with which I am acquainted. According to some philosophers it is only with the application of concepts that it makes sense to talk of truth and error. I can correctly or incorrectly categorize something with which I am acquainted, but the prior act of my being acquainted with the thing does not involve the possibility of error, because acquaintance does not by itself involve an attempt to categorize the thing in question.

But if knowledge by acquaintance does not even involve representations or characterizations which introduce the elements of truth and falsehood, then what has knowledge by acquaintance got to do with knowledge by description? According to Russell, all knowledge by description ultimately depends upon knowledge by acquaintance. But if knowledge by acquaintance does not even involve a bearer of truth value, how can knowledge by acquaintance yield knowledge by description (knowledge of truths)? In a famous argument against those philosophers who seek to find foundations for knowledge in elements of the world that are simply *given* to one, Wilfred Sellars (1963) claimed that proponents of the view face a dilemma (see Sellars, W.S. §3). On the one hand, they want to hold that one needs no concepts to have something given to one. Presumably, even the lower animals have qualities given to them in sensory experience. On the other hand, the given is supposed to end a regress of justification. The foundationalist argues that one cannot infer everything one knows from something else one knows, and thus that there must be a kind of knowledge that does not depend on any other knowledge and which can give us the 'first truths', the 'foundational' premises from which we can infer all other truths (see Foundationalism). But if knowledge by acquaintance does not involve the possibility of error because it does not have as its object something that can be true or false, how can it give us first truths? How can it give us *premises* (which by their very nature must be true or false) from which to infer other truths? Either knowledge by acquaintance does not involve the application of concepts and cannot therefore furnish premises for inference, or it does involve the application of concepts and cannot be distinguished from knowledge by description.

Although Russell is not as clear as he might be about the relationship between knowledge by acquaintance and knowledge by description, it seems that one might offer the following characterization of the way in which the latter depends on the former. The view itself depends on a controversial conception of truth as 'correspondence'. Roughly, the idea is that a bearer of truth value is true when it corresponds to some aspect of reality. On one view, the world contains facts (things having relational and non-relational properties) whether or not those facts are *represented* by conscious beings. Without representations of facts there are neither truths nor falsehoods. But true propositions are made true by virtue of their accurately 'picturing' facts (see Truth, correspondence theory of §1). One can be directly and immediately acquainted with the nonlinguistic (representation-independent) complexes we have called facts. We can call such acquaintance knowledge, but we must be careful to emphasize that having a fact before one in this way does not by itself give one knowledge of a truth. One can be acquainted with a fact without having the capacity to represent the fact, without having a 'picture' of the world that can be made true or false by the fact. But when one does have a representation of the world and one is directly acquainted with the

picture, the fact that makes the picture true and the correspondence between the picture and the fact, one has a kind of propositional knowledge that does not depend on any other propositional knowledge. One has foundational propositional knowledge created by direct acquaintance with features of the world. Furthermore, if my belief that I am in pain, for example, is justified by my standing in a relation of acquaintance to the very fact that makes what I believe true, it seems that my justification would (trivially) guarantee the truth of what I believe (see Russell, B.A.W. §§11-14).

3 Criticisms

This view of the way in which knowledge by acquaintance provides foundations for knowledge seems to rely on a highly controversial correspondence conception of truth. Many philosophers reject the basic idea behind the view that truth consists in representations of the world corresponding to facts. Some of these philosophers will reject the metaphysical category of fact (construed as truth-maker) as an illusion. There are no facts that are independent of conceptual frameworks, some philosophers argue. The world is not divided into things, their properties and relations. Indeed the only distinctions that exist are distinctions that we ‘carve out’ of the world with our concepts and categories. An extreme version of this view would hold that referring to a fact is just another way of talking about a proposition’s being true. To say that the world contains the fact, snow’s being white, is just another way of saying that it is true that snow is white. If such a view were plausible, then it would be a serious, almost pathetic, mistake to suppose that we could ground knowledge of truths in some relation we bear to the facts that make propositions true.

Notice that the attempt (discussed above) to understand foundational propositional knowledge as resting on knowledge by acquaintance, must make intelligible the concept of being acquainted with a *fact*. It is only if we can stand in some direct unproblematic relation to a *truth-maker* that such a relation might be thought of as yielding knowledge of a *truth-bearer*. Consequently, any philosophical misgivings one has about the metaphysical category of fact, and with it the concept of being acquainted with a fact, might translate into misgivings about the epistemic utility of the concept of knowledge by acquaintance. That we might stand in some kind of direct relation to a thing or property seems neither here nor there when it comes to gaining non-inferential, direct knowledge of truths. Only a *structured* reality could make propositions true and only acquaintance with such structure would be a plausible candidate for the source of foundational knowledge.

Even if one embraces the metaphysical category of fact construed as something that is distinct from representations of it, one might still have qualms about the intelligibility of this relation of direct acquaintance that we are supposed to bear to some, but not all, facts. Many contemporary philosophers argue that the very nature of justification precludes the possibility of having justification for believing empirical propositions that eliminates the possibility of error. Some of these arguments rely on thought experiments which describe sources of error even when one has the best imaginable evidence. Others invoke theoretical models about the nature of judgment which require us to recognize always the possibility of error with respect to empirical judgment. If, for example, judgment always involves relating something present to something past, then it seems doubtful that our justification for accepting such judgments could ever preclude the possibility of error.

Even if the acquaintance theorist can respond to criticisms, what positive reason could one give for supposing that the relation of direct acquaintance exists? Here, the dialogue is likely to reach an impasse. Proponents of the view that there is such a thing as direct acquaintance are quite likely to claim that they are directly acquainted with the fact that they are directly acquainted with certain facts. Such a claim is unlikely to make their critics happy, but then a proponent of the view probably does not care much about making critics happy. After all, if the concept of direct acquaintance with a fact is intelligible, why should one not use the concept in the course of justifying a belief that the relation exists?

See also: Certainty; Perception; Sense perception, Indian views of; Sense-data

RICHARD FUMERTON

References and further reading

BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.(Chapter 4 is most relevant. Heavily influenced by Sellars, Bonjour presents a clear argument against the doctrine of the given, an argument that includes an attack on the idea that empirical knowledge can be grounded in direct

acquaintance.)
Fumerton, R. (1985) *Metaphysical and Epistemological Problems of Perception*, Lincoln, NB: University of Nebraska.(Chapter 2 is an attempt to combine a specific correspondence conception of truth with the concept of acquaintance to present a plausible version of foundationalism.)
Russell, B. (1912) *The Problems of Philosophy*, London: Oxford University Press, 1959.(Chapter 5 is probably the clearest of Russell’s attempts to explain the distinction between knowledge by acquaintance and knowledge by description. Accessible to readers with widely varying philosophical backgrounds.)
Russell, B. (1914) ‘On the Nature of Acquaintance’, *Monist*, 24: 1-16, 161-87, 435-53.(This three-part article contains a more extensive treatment of the distinction discussed in *The Problems of Philosophy*.)
Sellars, W. (1963) ‘Empiricism and the Philosophy of Mind’, in *Science, Perception and Reality*, London: Routledge & Kegan Paul, 127-96.(Wide-ranging and in parts very difficult. The argument referred to in the text is contained in section 1 of the article.)

Knowledge, causal theory of

Epistemologists have always recognized the importance of causal processes in accounting for our knowledge of things. In discussions of perception, memory and reasoning, for example, it is commonly assumed that these ways of coming to know are fundamentally causal. We perceive things and thus come to have knowledge about them via complex causal processes; memory is, at least in part, the retention of previously gained knowledge through some sort of causal process; and reasoning is a causal process that takes beliefs as inputs and generates beliefs as outputs.

A causal theory of knowledge is a form of externalism and is based on the fundamental idea that a person knows some proposition, p, *only if there is an appropriate causal connection between the state of affairs that makes* p *true and the* person's *belief in* p. *Although this kind of theory has roots that extend to ancient times, contemporary versions attempt to make more precise the nature of the causal connections required for knowledge. The causal theory is closely related to other forms of externalist theories, such as the conclusive reasons theory, information-theoretic views and the various forms of reliabilism.*

1 Formulation of the causal theory of knowing

In the internalist-oriented environment which dominated epistemology from the time of Descartes until the middle of the twentieth century, it was not considered appropriate to refer to the causal history of a belief in providing an analysis, or definition, of the positive epistemic status of that belief (see Internalism and externalism in epistemology). Rather, the epistemologist's job was to provide definitions of concepts such as justification and knowledge independently of any assumed causal connections with the external world, and then to show how, from such analyses, one could argue ('internally') that such causal connections exist. To do otherwise would be to beg the question against scepticism about the external world.

Since the 1960s, there has been a shift away from the internalist position in epistemology. The causal theory of knowing is one of the early versions of externalism, introduced by Alvin Goldman (1967) and conceived primarily as a response to the Gettier problem which had appeared only a few years earlier (see Gettier problems). In Gettier examples, a person, *S*, has a justified belief in something that is only coincidentally true. This element of coincidence, which is perhaps the most salient feature of Gettier cases, is very difficult to explain without introducing some element of external connection between the individual's belief, the justification for that belief and the state of affairs which is the object of the belief. Goldman's original proposal was to focus on the causal connections that typically obtain between these various epistemically relevant items when a person has knowledge.

Goldman expressed his proposal as a set of truth-conditions for knowledge, in the following schematic form:

[A person] *S* knows that *p* if and only if the fact *p* is causally connected in an 'appropriate' way with *S*'s believing *p*.

The ways that are 'appropriate' include perception, memory, various other kinds of causal chains and forks, and combinations of these. Goldman adds to this the further condition that the relevant causal connections obtaining between the state of affairs *p* and one's belief must be 'correctly reconstructed by inferences, each of which is warranted'. This further condition is designed to accommodate the fact that causal chains can sometimes take very unexpected routes. If the route is unusual enough, then even though it results in a true belief that *p*, it does not provide knowledge. For example, suppose that Sally is perceiving an object only through a complex network of mirrors, or via some holographic imaging device which produces a very realistic image of the object. Then, she might come to believe that the object is in front of her when in fact it is, but she might fail to know that it is there, as she knows nothing about the unusual causal mechanism. If she were in a position to correctly reconstruct the causal chain, however, then she would have knowledge.

One example developed by Goldman that illustrates the intuitive appeal of his proposal involves a person, Smith, who perceives solidified lava lying around a mountain. On the basis of beliefs about this lava, and background beliefs, Smith inferentially comes to believe, correctly, that the mountain erupted many centuries ago. Assuming that Smith's inferences are warranted, does he know? To answer this, we must ask what sort of causal ancestry obtains between the eruption of the mountain and Smith's belief that it has erupted. If the lava that he sees resulted

from the eruption, as he imagines, then Smith does have knowledge. However, suppose that unknown to Smith the lava has been placed there by promoters who wish to make it look as though the area was once volcanic. The lava actually produced by the eruption has been completely covered by years of sedimentation, and cannot be seen. Then, there is no appropriate causal connection between eruption and perceived lava, and Smith does not know. The naturally intuitive appeal of such an example is confirmation that Goldman's causal theory captures at least part of what we require for knowledge.

2 The Gettier examples

In Gettier examples, the requirement that there be an appropriate causal connection between the fact *p*, and *S*'s (warranted) belief, is not satisfied, very much along the lines illustrated in the lava case. This can be seen by considering a specific (Gettier-style) example, introduced by Keith Lehrer (1965). Suppose Smith correctly infers that someone in his class owns a Ford, from some true evidence that justifies the false belief that a student, Mr. Nogot, owns a Ford. It so happens that another student in Smith's class, Mr. Havit, does own a Ford, but Smith has no evidence one way or the other for this proposition. The causal theory clearly explains the lack of knowledge in this example, since the state of affairs making it true that someone in Smith's class owns a Ford (namely, Havit's owning a Ford) is not causally related in any of the appropriate ways with Smith's belief. Rather, Smith's belief is caused by states of affairs which make true the evidence on which his belief that Mr. Nogot owns a Ford is based. And, these states of affairs, whatever they are, have nothing to do with Mr. Havit's ownership.

It should also be noticed that if this case were redescribed so that Smith does have evidence that Havit owns a Ford, then it would be clear (all else being equal) that he does have knowledge that someone in his class owns a Ford, for then the required causal connection would obtain. It can be concluded that in Gettier examples, as well as 'ordinary' cases in which we would tend to ascribe knowledge to individuals, the causal theory provides a clear and intuitively appealing account of knowledge.

3 Problems for the causal theory

Despite its merits, there are a number of examples that raise serious difficulties for the causal theory. One class of examples is generated by the phenomenon of causal overdetermination. Suppose Alfred comes to believe that there is a sheep in the field by hearing a recording of the sounds normally produced by sheep while also looking at a distant boulder that looks like a sheep in the field. Even if there is a sheep somewhere in the field not seen by Alfred, he does not know that one is there. Later, should he come to perceive the real sheep, he would come to know that there is a sheep in the field. The problem for the causal theory is that Alfred's later perception of the real sheep would not cause his belief in an appropriate way, because his belief already exists. The causal theory would incorrectly deny that Alfred has knowledge. This kind of case can be accommodated within the general spirit of the causal theory by allowing as 'appropriate' the relation of causal sustaining, a very weak form of causal connection.

More difficult to accommodate are overdetermination cases in which an individual is acquainted with some state of affairs which is causally sufficient for the state of affairs *p*, but which is not in fact any part of the cause of *p*. Abigail might, for example, be aware that Jones has taken a fatal dose of poison, with no antidote, and thereby come to know that Jones is dead even though in fact (but unknown to her) Jones died of other causes. The only way to accommodate such an example is to further extend the scope of 'appropriate causal connections' to include cases in which one infers *p* from something which is causally sufficient for *p* even though not the cause or any part of the cause of *p*. It is not obvious, however, that this kind of modification is in the spirit of the causal theory, for it abandons the idea that knowledge requires a causal *connection* between belief and known fact.

Equally troubling for the causal theory are situations involving logical and/or mathematical facts. One kind of situation is very similar to the example just given. Suppose Mark observes that an owl is on top of a flagpole, which he already knows to be 15 feet high. He also observes a mouse to be 12 feet from the bottom of the flagpole. Mark correctly infers that the mouse is a little over 19 feet from the owl. But, the mouse's being a little over 19 feet from the owl is no part of the cause of Mark's belief in that fact, nor are the facts from which he makes the inference themselves causes of the mouse's distance from the owl. Rather, this is a case in which the fact concluded follows logically from the observed premises. To take another example, suppose that the object of belief is itself a logical or mathematical truth, such as '2 + 2 = 4'. Surely a theory of knowledge should allow for

knowledge of such truths, but the causal theory faces serious obstacles in attempting to provide an account. Whatever the nature of the facts which make true logical and mathematical truths, they do not seem to be the sorts of facts that are parts of causal chains. If not, then there can be no appropriate causal connection of any kind between such facts and one's beliefs. It should be noted, however, that there have been some serious efforts, particularly by Mark Steiner (1973) and Philip Kitcher (1984), to accommodate mathematical and logical knowledge within the framework of a causal theory.

In his own effort to accommodate examples involving inferences based on logical connections, such as the one concerning the owl and mouse, as well as knowledge of mathematical and logical truths, Goldman (1967) proposed that the scope of 'appropriate causal connection' be extended even further to include logical and mathematical connections. Even those philosophers who are inclined to be sympathetic with the previously mentioned extensions of the notion of a causal connection have baulked at this idea. Peter Klein (1976) argues that there is no adequate way of formulating a version of the causal theory which allows logical and mathematical connections to be counted as causal.

The problematic examples discussed thus far have raised questions about the necessity of the 'appropriate causal connection' requirement. There are also difficult examples that raise doubts about the sufficiency of the 'appropriate causal connection' requirement, even in its most refined forms. Perhaps the most famous of these examples is one introduced by Goldman himself (1976), leading to his own ultimate abandonment of the early version of his causal theory.

In this example, we are to suppose that Henry is driving through the countryside looking at the scenery. One of the things he sees is a barn, and on the basis of this perfectly ordinary perceptual evidence, along with his background beliefs about barns, Henry comes to believe that there is a barn there. Unknown to him, however, the region is populated by papier-mâché barn facsimiles, of the sort found on movie-studio lots. These facsimiles, which are propped up by sticks from behind, would be mistaken by Henry, or anyone else, for real barns when sighted casually from the road. It is only by good fortune, or coincidence, that Henry has perceived a real barn rather than a facsimile. Given this, Henry cannot be said to have knowledge that the object he sees is a barn, for he could just as well have been perceiving a facsimile. The problem is that Goldman's causal theory of knowledge is satisfied: Henry's belief that there is a barn there is caused in a perfectly straightforward way by the presence of the barn, and we may suppose that Henry has correctly reconstructed through warranted inferences the causal chain leading to this belief. So, if Henry does not have knowledge, the conditions of the causal theory do not appear to be sufficient for knowledge.

In light of these and other difficult counterexamples to both the necessity and the sufficiency of the causal theory, it has largely been abandoned, at least in the form originally suggested by Goldman in which a causal connection between state of affairs and belief is required.

4 Successors to the causal theory of knowing

The causal theory of knowing still survives in many of the externalist theories which have arisen since Goldman's early proposal. Among the more prominent theories in this category are the 'conclusive reasons' approach, exemplified in the works of Dretske and Armstrong, and various versions of Reliabilism, exemplified in the later works of Goldman, and in works by Swain, Alston and Plantinga.

Goldman (1976) proposed a modification of the early causal theory in which it is required not only that the individual's belief that *p* be caused in an appropriate way by the state of affairs *p*, but also that the individual be a reliable discriminator with respect to *p* and other, alternative states of affairs that might causally substitute for *p*. When this discriminatory capacity is lacking, or is not functioning properly, then the individual fails to know even if the causal connection requirement is met. In this way, examples such as the one involving Henry can be accounted for.

But even with this modification, the causal theory cannot handle examples in which there is no causal connection of the required sort, such as the mathematical and logical cases illustrated above. Ultimately, the requirement that there be an actual causal chain linking the state of affairs that makes *p* true with the belief that *p* must be abandoned. Taking its place is the notion of an 'appropriate causal *ancestry*' of a belief. A belief is an instance of knowledge, according to this kind of causal theory, provided it is produced in an appropriate manner, where the

manner of production is appropriate just in case it is *reliable*. A theory which holds this is called a ‘reliability theory’, and is the main kind of successor to the early casual theories (see Reliabilism).

See also: Causality and necessity in Islamic thought; Information theory and epistemology

MARSHALL SWAIN

References and further reading

Alston, W. (1989) *Epistemic Justification: Essays in the Theory of Knowledge*, Ithaca, NY: Cornell University Press.(In this collection of essays, items 8, 9 and 12 present the author’s externalistic position.)

Armstrong, D. (1973) *Belief, Truth, and Knowledge*, Cambridge: Cambridge University Press.(Presentation of a form of reliabilism which requires for knowledge a ‘lawlike’ connection between belief and state of affairs; very closely related to the causal theory.)

Carrier, L.S. (1976) ‘The Causal Theory of Knowledge’, *The Southern Journal of Philosophy* 9: 3-11.(Critique of Goldman’s causal theory and presentation of an alternative view.)

Dretske, F. (1971) ‘Conclusive Reasons’, *The Australasian Journal of Philosophy* 49: 1-22.(Classic presentation of the conclusive reasons view, which is closely related to the causal and reliability views.)

Foley, R. (1987) *The Theory of Epistemic Rationality*, Cambridge, MA: Harvard University Press.(Chapter 4 has interesting arguments and examples which challenge the fundamental premise of the causal theory, namely, that at least some appropriate causal history is necessary for knowledge.)

Goldman, A. (1967) ‘A Causal Theory of Knowing’, *The Journal of Philosophy* 64 (12): 355-72; repr. in G.S. Pappas and M. Swain (eds) *Essays on Knowledge and Justification*, Ithaca, NY and New York: Cornell University Press, 1978.(The original article in which Goldman presents the causal connection view.)

Goldman, A. (1976) ‘Discrimination and Perceptual Knowledge’, *The Journal of Philosophy* 73: 771-91; repr. in G.S. Pappas and M. Swain (eds) *Essays on Knowledge and Justification*, Ithaca, NY and New York: Cornell University Press, 1978.(Presentation of the Barn-facsimile example, and development of an early version of a reliabilist view as a successor to the early causal theory.)

Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press. (Some historical tracing of the causal theories, and presentation of the author’s reliabilist views.)

Kitcher, P. (1984) *Mathematical Knowledge*, Oxford and New York: Oxford University Press.(A thorough discussion of mathematical and logical knowledge, including an effort to accommodate such knowledge within a causal/reliability theory.)

Klein, P. (1976) ‘Knowledge, Causality, and Defeasibility’, *The Journal of Philosophy* 73 (20): 792-812.(A detailed critique of the causal theory of knowing, in which it is argued that no theory of this kind can succeed in the face of examples involving logical and mathematical connections.)

Lehrer, K. (1965) ‘Knowledge, Truth, and Evidence’, *Analysis* 25: 168-75.(The ‘Nogot-Havit’ example can be found in this early discussion of Gettier problems.)

Plantinga, A. (1994a) *Warrant: The Current Debate*, Oxford: Oxford University Press.(See chapter 9 for a critical discussion of reliabilism and related theories.)

Plantinga, A. (1994b) *Warrant and Proper Function*, Oxford: Oxford University Press. (Development of the author’s version of externalism.)

Shope, R. (1983) *The Analysis of Knowing*, Princeton, NJ: Princeton University Press. (See especially chapter 5 for a good survey of causal theories of knowing and related theories.)

Skyrms, B. (1967) ‘The Explication of "X knows that p"’, *The Journal of Philosophy* 64: 373-89.(An early discussion of Goldman’s view, raising problems having to do with overdetermination.)

Steiner, M. (1973) ‘Platonism and the Causal Theory of Knowledge’, *The Journal of Philosophy* 70 (3): 57-66. (An attempt to provide for mathematical knowledge within the framework of a causal theory of knowing.)

Swain, M. (1972) ‘Knowledge, Causality, and Justification’, *The Journal of Philosophy* 69 (11): 291-300. (Discussion and revision of Goldman’s early theory, and of examples raised by Skyrms and others.)

Swain, M. (1979) *Reasons and Knowledge*, Ithaca, NY: Cornell University Press.(See especially chapters 2 and 6 for a detailed development of a causal theory of knowing designed to avoid many of the problems with Goldman’s view.)

Knowledge, concept of

The branch of philosophy concerned with the nature and extent of human knowledge is called epistemology (from the Greek epistēmē *meaning knowledge, and* logos *meaning theory). Knowledge seems to come in many varieties: we know people, places and things; we know how to perform tasks; we know facts. Factual knowledge has been the central focus of epistemology.*

We can know a fact only if we have a true belief about it. However, since only some true beliefs are knowledge (consider, for example, a lucky guess), the central question asked by epistemologists is 'What converts mere true belief into knowledge?'. There are many, and often conflicting, answers to this question. The primary traditional answer has been that our true beliefs must be based upon sufficiently good reasons in order to be certifiable as knowledge. Foundationalists have held that the structure of reasons is such that our reasons ultimately rest upon basic reasons that have no further reasons supporting them. Coherentists have argued that there are no foundational reasons. Rather, they argue that our beliefs are mutually supporting.

In addition to the constraints upon the overall structure of reasons, epistemologists have proposed various general principles governing reasons. For example, it seems that if my reasons are adequate to affirm some fact, those reasons should be adequate to eliminate other incompatible hypotheses. This initially plausible principle appears to lead directly to some deep puzzles and, perhaps, even to scepticism. Indeed, many of the principles that seem initially plausible lead to various unexpected and unwelcome conclusions.

Alternatives to the primary traditional answer to the central epistemic question have been developed, in part because of the supposed failures of traditional epistemology. These alternative views claim that it is something other than good reasons which distinguishes (mere) true beliefs from knowledge. Reliabilists claim that a true belief produced by a sufficiently reliable process is knowledge. Good reasoning is but one of the many ways in which beliefs can be reliably produced. The issue of whether the objections to traditional epistemology are valid or whether the proposed substitutes are better remains unresolved.

1 The varieties of knowledge

Knowledge comes in many varieties. I can know *how* to adjust a carburettor. I can know a person. I can know *that* mixing bleach and ammonia is dangerous. In the first case, I possess a skill. In the second, I am acquainted with someone. In the third, I know a fact. Epistemologists have differed on the relationships between these types of knowledge. On the one hand, it could be held that knowing a person (place or thing) should be construed as nothing more (or less) than knowing certain facts about that someone and possessing the skill of being able to distinguish that person from other objects. On the other hand, it has been held that knowing facts depends upon being acquainted with particular objects. Whether the reduction of one form of knowledge to another is ultimately successful is an area of contention among epistemologists (see Knowledge by acquaintance and description).

Nevertheless, it is knowledge of facts, so-called *propositional knowledge*, as opposed to knowledge by acquaintance or the possession of skills, that has been the central concern of epistemologists. The central question can be put this way: which beliefs of mine are to be counted as knowledge? This question presupposes that knowledge is a species of belief, but some might think that knowledge and belief are mutually exclusive: for example, we say such things as 'I do not *believe* that; I *know* it'. But we also say such things as 'I am not *happy*; I am *ecstatic*'. A suggested paraphrase of this expression seems to capture what is meant without denying the obviously true claim that ecstasy is a form of happiness. The paraphrase is: I am not *merely* happy, I am ecstatic. The parallel is: I do not merely believe it; I know it. Thus, this type of linguistic evidence does not support the suggestion that belief and knowledge are mutually exclusive. In general, epistemologists have held that propositional knowledge is a species of belief.

2 Propositional knowledge is not mere true belief

Propositional knowledge is a species of belief; but which beliefs are knowledge? The first thing to note is that a belief must be true in order for it to count as knowledge. But that is obviously not enough. First, true beliefs can be based upon faulty reasoning. Suppose that I believe that smoking is a leading cause of fatal lung cancer because I infer it from the fact that I know two smokers who died of lung cancer. The generalization is true, but my evidence

is too meagre for my belief to count as knowledge. Second, true beliefs can be based on false beliefs. Modifying an example used by Bertrand Russell, suppose that I believe truly that the last name of the President of the United States in 1996 begins with a ‘C’. Also suppose this belief is based upon the false belief that the President is Winston Churchill. My true belief that the President’s name begins with a ‘C’ is not knowledge because it is based on a false belief.

Third, even some true beliefs resulting from good reasoning based upon true beliefs are not knowledge. Suppose that I believe (truly) that my neighbours are at home. My belief is based upon good reasoning from my true belief that I see lights on and that, in the past, the lights have been on only when they were at home. But suppose further that this time the lights were turned on by a guest and that my neighbours had just entered the house and would not have had time to turn on the lights. In this case, I fail to know that my neighbours are home. So, the central question becomes: what must be added to true belief to convert it into knowledge?

3 Warrant

The property, whatever it is, that, if added to true belief converts it into knowledge, we may refer to as ‘warrant’. Knowledge, then, is true, warranted, belief. But simply to name the missing property does not bring us closer to understanding it and we must be careful not to think of ‘warrant’ as a sophisticated synonym for ‘justified’. Let us say that a belief is *justified* just in case we are entitled to hold it on the basis of suitable reasons available to us. In the neighbour/lights case mentioned above, we have already seen that justification is not sufficient for warrant. Whether it is even necessary will be important in the discussion that follows, especially in §6.

Given the great variety of approaches to an account of warrant, is there any common, underlying starting point embraced by epistemologists? Yes: a warranted belief is one that is not held on the basis of mere cognitive luck. Plato appeals to that intuition in the *Theaetetus*; Aristotle’s account of the transition from ignorance of the first principles in science to knowledge of them in the *Posterior Analytics* is designed to demonstrate that there are reliable cognitive mechanisms whose output is not the result of chance; Descartes (1641) proposes methods for acquiring beliefs that would (he thinks) necessarily lead to truth; Locke (1689) suggests that even if persons arrive at a true belief by accident, they are not thereby free from criticism.

Let us start with the assumption that a proposition is known just in case it is not an accident, from the cognitive point of view, that it is both believed and true. Hence the task becomes one of developing an account of warrant that accurately portrays what it is that makes a belief non-accidentally true from the cognitive point of view.

4 Foundationalism and coherentism

There are two main, traditional approaches to the account of justification: foundationalism and coherentism. Both are *normative* views about rules in virtue of which propositions *ought* to be accepted or *ought* to be rejected or *ought* to be suspended (see Knowledge and justification, coherence theory of; Foundationalism; Normative epistemology). In order to characterize these approaches, recall how the ancient Pyrrhonian Sceptics divided the possible structures of reasons that provide a basis for accepting a belief (see Epistemology, history of; Pyrrhonism). Suppose you hold a belief and offer another belief as the reason for the first - for example, suppose you believe that Ford cars are generally less expensive than BMWs. Your reason could be your belief that you were told so by a reliable person. An obvious question arises: what is your basis for believing that the person is reliable? You could answer with another reason and that reason could, itself, be supported by a further reason, and so on.

This process of providing reasons for your beliefs can have only three possible structures:

Foundationalism: The process of giving reasons could be such that not every reason is supported by another reason because there are *basic reasons* which have no need of further reasons supporting them.
Coherentism: The process of giving reasons could have no reason that is not supported by another reason, but there is not an infinite number of reasons. Thus, beliefs are mutually supporting.
Infinitism: The process of giving reasons could have no reason that is not supported by another reason, but there is an infinite number of reasons.

Foundationalism and coherentism have both been developed and defended, and there are well-known objections to

each view. In contrast, the *prima facie* objections to infinitism have seemed so overwhelming that it has not been investigated carefully. Infinitism *seems* to require that a person should have an infinite number of beliefs (which *seems* on its face to be false). In addition, it *seems* to lead inevitably to the conclusion that no belief could ever be justified, since the process of justification would never come to an end.

The standard objections to foundationalism are several. First, as the Pyrrhonians would point out, there must be a distinction between what makes a belief *properly* basic and what makes it simply one for which no other reason is, in fact, given. Otherwise, the offered 'basic' reason is arbitrary. But if there is some further reason for thinking that an offered reason is not arbitrary, then there is a reason for accepting it, and the offered reason is, thereby, not basic. Hence, there can be no foundational propositions.

Second, some preferred candidates for properly basic reasons seem not to be properly basic on closer inspection. Consider perceptual judgments - the source of most of our knowledge of the external world according to many philosophers (see Empiricism; A Posteriori). A reason for believing that there is a tree before me is that I see a tree before me. But the latter proposition does not appear to be properly basic because one could be required to explain what it is about what is seen that leads one to believe that it is a *tree* that one sees (as opposed to an illusion). Thus, some foundationalists have retreated to sensation-beliefs (so-called sense-data propositions) as their candidates for properly basic beliefs: for example, 'I seem to see a green, brown, tallish object' (see Ayer, A.J.; Broad, C.D.; Moore, G.E.). But although these propositions might seem to be properly basic, there are notorious problems with the sense-data view (see Perception, epistemic issues in; Sense-data). First, the proffered basic beliefs seem to be too meagre to provide a sufficient basis for the rich scope of things we seem to know. For example, how can my knowledge that objects persist when not being perceived be traced to particular sense-data? Second, it appears that our knowledge of the way in which to characterize our sensations (private sensations accessible only to the individual having them) depends upon our knowledge of public objects (see Criteria; Wittgenstein, L.J.J.). How could we know, for example, that we have a throbbing pain without first recognizing what it is for a public object (say, a muscle) to be throbbing?

Foundationalists have developed answers to these objections in part by liberalizing the requirements either for being properly basic or for being an acceptable pattern of inference from the foundational propositions to the non-foundational ones (see Inference to the best explanation). For example, *contextualist* accounts of knowledge have been developed that hold that a proposition is properly basic just in case it is accepted by the relevant community of putative knowers. In a discussion with a friend I could offer as my reason for believing another moon of Jupiter had been discovered that 'I read it in the newspaper'. I would not need further reasons for believing that I read it. In contrast, at a convention of astronomers that reason would not be accepted. Hence, contextualists claim, what counts as a basic reason is context-dependent.

There are two obvious responses to contextualism. The first is that it might be an accurate description of some aspects of our epistemic practices, but the fundamental Pyrrhonian question remains: what distinguishes a properly basic proposition from one that is merely offered and accepted by a community of putative knowers? The issue concerns what beliefs, if any, *ought* to be offered and accepted without further reasons. The question is not what beliefs *are* offered and accepted without further reason. The second response is a corollary of the first. Knowledge seems to be a highly prized state of belief (as Plato put it). But, if the contextualists were right, I would gain knowledge by joining a community of rather epistemically gullible and permissive folk. That hardly seems right! (See Contextualism, epistemological.)

In sum, it remains a subject of dispute among epistemologists whether the stock of purported foundational propositions can be made sufficiently rich and abundant without including too many that clearly require evidential support, or whether the patterns of inferences can be liberalized sufficiently without allowing patterns that are not sufficiently truth-conducive.

The historical rival of foundationalism is coherentism. Coherentists deny that there are basic reasons and claim that all propositions derive their warrant, at least in part, from other propositions. The fundamental objection is this: Typically, we recognize that arguing in a circle is not an acceptable pattern of inference, so what makes it acceptable in some cases? Suppose I believe that apples contain vitamin C, at least in part because I believe that fruits contain vitamin C. I would surely be appropriately accused of circular reasoning if I believed, in part, that fruits contain vitamin C because apples do.

Coherentists would be quick to point out that they are not really suggesting that one should argue in a circle. Rather, they would point to the fact that our beliefs come in bunches with a web-like structure (see Quine, W.V.). They are ‘mutually supporting’ just as the poles in a tepee are mutually supporting. A belief is warranted just in case it is a member of a set of coherent beliefs.

But whether these colourful analogies answer the basic objection is not clear. Presumably, circular reasoning is not acceptable because although it might be the case that if you believe b_1 it might be reasonable to believe b_2, and if you believe b_2 it might be reasonable to also believe b_1, their mutual support gives you no reason for believing them both. Thus, the fundamental question is this: What makes one total set of coherent beliefs, say T_1, any more acceptable than an alternative total set of coherent beliefs, say T_2?

The Pyrrhonian Sceptics would point out that coherentists either have an answer for that question or they do not. If they do, then they seem to have abandoned their central view, since there now seems to be a reason for adopting the set of beliefs, T_1, that is not one of the beliefs in T_1. Indeed, if they provide an answer, they have embraced foundationalism. If they do not have an answer, then it seems that adopting T_1 is arbitrary. Coherentists have attempted to answer this objection by giving a ‘meta-justification’ for thinking that certain kinds of coherent belief systems are likely to contain true members. Indeed, some have argued that coherent beliefs are, by their very nature, likely to be true (see Davidson, D.). But whether that strategy will suffice to answer the objections remains an open question in epistemology.

5 Defeasibility theories

A basic objection to the foundationalist’s and coherentist’s accounts of justification is that neither seems to be able to show that a true belief which satisfied their accounts would be non-accidentally true. First, as the neighbour/lights case showed, a true belief could be fully justified on their accounts, but not be knowledge. Second, as the Pyrrhonians pointed out, either the beliefs seem to rest upon arbitrary foundations or they seem to be only one of many, equally coherent sets of beliefs. The defeasibility theory was developed, in part, to address these issues. It holds, roughly, that it is not only the evidence that one possesses that makes a belief warranted; it is equally important that there is no defeating evidence that one does not possess. That is, in order for a belief to be warranted it must not only be justified (in the sense required by either the foundationalists or the coherentists) but its justification must be such that there is no truth which, if added to the reasons that justify the belief, is such that the belief would no longer be justified (see Knowledge, defeasibility theory of).

The defeasibility theory can explain why it is not a cognitive accident that the warranted belief is true. If any of the important supporting reasons (those that if removed would destroy the justification) were false, then adding the denial of those reasons (in other words, adding the truth) to one’s beliefs would undermine the justification. In addition, if there is evidence that one does not possess such that it makes it an accident that the belief is true, the propositions describing that evidence would undercut the justification.

A well-known case will help to illustrate this (see Gettier problems). Suppose that I know Tom Grabit well and I see what appears to be Tom stealing a library book: I come to believe that Tom stole a library book. And, let us suppose that Tom did indeed steal the book. Foundationalists and coherentists could deploy their accounts in order to show that the belief is justified. Nevertheless, suppose that, unknown to me, Tom has an identical twin, John, who is a kleptomaniac and was in the library on the day in question and stole a copy of the same book. Even though I arrived at a true belief as a result of good reasoning based upon true propositions, I do not know that Tom stole the book since it is accidental, from the cognitive point of view, that I arrived at the truth. I could just as easily have based my belief on having seen John stealing the book.

The defeasibility theorists would point out that the belief that Tom stole the book is defeated; if the true proposition describing John were added to my beliefs, I would no longer be justified in believing Tom stole the book. In general, the defeasibility theory can rule out accidentally true beliefs as warranted because those beliefs would not be able to stand up to the truth.

Nevertheless, the defeasibility theory has its problems. The primary one is that it seems to exclude too much from what we know. Returning to the Grabit Case, suppose that everything is as it was except that Tom does not have a

twin but that Tom's mother sincerely avows the claims about John. Now, there is a true proposition (Tom's mother has said sincerely that Tom has an identical twin, John) that defeats the original justification. Hence the belief that Tom stole the book would be defeated. But if Tom's mother were demented and there never was a twin, it seems that I knew all along that Tom stole the book.

Defeasibility theorists have tried to answer this objection by suggesting ways to distinguish between so-called misleading defeaters (for example, Mrs Grabit sincerely avows that Tom has an identical twin, John) and genuine ones (for example, Tom has an identical twin, John), but there is no agreement among epistemologists that any of these suggestions has succeeded in correctly capturing the distinction between genuine and misleading defeaters.

6 Externalism

Partly in response to the difficulties with foundationalism and coherentism even as supplemented by the defeasibility theory, epistemologists have developed a variety of alternative accounts of warrant. They have been called 'externalistic' because their accounts of warrant focus on features of the world other than the knower's reasons for belief. Two important ones are the causal theory and the reliabilist theory (see Knowledge, causal theory of; Reliabilism).

In their purest forms, these accounts begin with the view that knowledge, and hence warrant, does not require justification. The foundationalists had already conceded that there are no reasons for properly basic beliefs. This seemed to create a problem for foundationalism only because it was assumed that all beliefs needed to be justified and the 'basic' reasons appeared to be arbitrary. But drop the requirement that beliefs need to be justified in order to be warranted, and this problem immediately disappears.

Roughly, the causal theory of warrant holds that a belief is warranted if and only if the state of affairs represented in the belief is appropriately causally related to the belief. For example, suppose I come to believe that there is a bird in a tree as a causal consequence of seeing the bird in the tree. Sometimes the causal connection is more complex; but this direct type of causal connection between the belief and what it represents will suffice for our purposes.

This theory is initially appealing because it appears to satisfy the basic requirement that a warranted belief be non-accidentally true since the state of affairs represented in the belief is a cause of my belief. However, it is easily seen to be both too weak and too strong; and there seem to be some deep problems with it as a general account of warrant. It is too weak because it would count some true beliefs as warranted that clearly are not known. Recall the Grabit case. My belief that I see Tom stealing the book is caused by Tom's stealing the book, but if he has an identical twin, I do not know that Tom stole the book. It is too strong because there seem to be many beliefs that count as knowledge which can not be appropriately *causally* related to what they represent. Suppose I know that there is no elephant smaller than a kitten: what possible *causal* connection could there be between there being no elephant smaller than a kitten and my belief? In addition, potential difficulties arise about knowledge of a priori propositions (such as 2 + 2 = 4) and counterfactuals (such as, if it were raining today, we would have called off the picnic). It looks as though there is no possible way to produce a causal connection between my belief and what is represented in the belief - at least as 'cause' is usually understood (see A priori).

Nevertheless, a basic tenet of the causal theory might still be correct: Not all beliefs need to be based on reasons in order to count as knowledge. The reliabilist theory of warrant can be seen as the successor of the causal theory. Instead of requiring an appropriate causal connection between the states of affairs represented in the belief and the belief itself, a typical form of reliabilism holds that a belief is warranted just in case the process resulting in the belief produces true beliefs sufficiently often.

Thus, the non-accidental nature of the true belief receives a very straightforward analysis. The belief is non-accidentally true because the process that produces the belief produces true beliefs sufficiently often. This view has many advantages over the causal theory. My belief that elephants are larger than kittens need not be caused by that state of affairs. All that is required is that the process by which I come to believe that proposition typically (often enough) results in true beliefs. A priori or counterfactual propositions present no problem since there could be reliable processes that produce those beliefs.

Nevertheless, there are problems confronting this view. Suppose that you require that the process should produce

true beliefs on 100 per cent of the occasions on which it arises. That is a very stringent condition; but it is not stringent enough! For if the belief that Tom Grabit stole a book arises only once in the history of the world - the time I saw him stealing the book - the actual process produced a true belief 100 per cent of the times it arose; but it is not knowledge. The obvious move for the reliabilist is not only to include the actual occasions when the particular belief is produced but rather to consider whether the type of process that produced this belief would produce true beliefs of this type sufficiently often. But correctly characterizing those types has not proved easy. Is the type of belief one in which Tom is involved? Or identical twins? Or libraries? That seems too narrow. Is the type of process one in which there is first a perception and then some inferences? That seems too broad. It remains an open question whether reliabilism can produce an acceptable account of the types of processes and the types of beliefs.

Finally, there is one further objection that some epistemologists have brought against reliabilism. Perhaps it is best illustrated in a case presented by Keith Lehrer (1990: 163) that can be summarized as follows: a certain Mr Truetemp has a thermometer-with-temperature-belief-generator implanted in his head so that within certain ranges of temperatures he has perfectly reliable temperature beliefs. When it is 50 degrees, he comes to believe that it is 50 degrees. When it is not 50 degrees, he does not come to believe that it is 50 degrees. He holds these beliefs without knowing why he does.

Such beliefs would satisfy all of the requirements suggested by the reliabilists, but many epistemologists would hold that although Mr Truetemp has true beliefs and they are not accidentally true because his thermometer-with-temperature-belief-generator is reliable, they are accidentally true *from the cognitive point of view*, as he has no reasons at all for his beliefs. Indeed, some would say that what Mr Truetemp possesses is a skill (of telling the temperature) and not propositional knowledge at all.

Here we can detect a fundamental clash of intuitions. The reliabilists would hold that Mr Truetemp does know; the traditional normativists would hold that he does not. There appears to be no way to satisfy both. But this much seems clear: There are some situations in which the steps in the process that brings about a belief include the holding of reasons. In those cases in which there is no automatic true-belief-generator (as in the Truetemp case) and in which we must rely upon our reasoning to arrive at a belief, the questions asked by the traditional normativists are crucial: what must the structure of our reasons be so as to make a true belief acceptable? Are there foundational reasons? Can mutually supporting reasons be offered without begging the question? (Could reasons be infinite in number?) And need those reasons be such that they are not undermined by the truth, as the defeasibility theorists would hold? At least in some cases, it seems that normative standards for belief-acquisition apply and their satisfaction will determine whether a belief ought to be accepted. Thus, it appears that an evaluation of the conditions under which beliefs ought to be accepted, denied or suspended is inescapable (see Internalism and externalism in epistemology; Justification, epistemic).

7 Epistemic principles

Epistemic principles describe the normative epistemic status of propositions under varying conditions (see Epistemic logic). It is generally agreed that if a person, *S*, is justified in believing any proposition, *x*, then *S* is not at the same time justified in believing that not-*x*. Foundationalists and coherentists alike can, and typically do, accept this principle. Other principles are more controversial. They are intuitively plausible but they seem to provide a basis for scepticism and for some deep epistemic puzzles. Here are three of the more interesting principles.

Conjunction Principle (CON-P): If *S* is justified in believing that *x*, and *S* is justified in believing that *y*, then *S* is justified in believing that (*x* and *y*).
Closure Principle (CLO-P): If *S* is justified in believing *x*, and *x* entails *y*, then *S* is justified in believing that *y*.
Evidence Transfer Principle (ET-P): If there is some evidence, *e*, that justifies *S* in believing that *x*, and *x* entails *y*, then *e* justifies *S* in believing that *y*.

In each principle and with suitable grammatical modifications ‘justified’ could be replaced by other epistemic terms, such as ‘reasonable’, ‘plausible’, ‘evident’, ‘certain’. Furthermore, each principle is designed to capture a basis upon which a positive normative epistemic status of a proposition can be transferred to another proposition. As a corollary, ‘*S* is justified in believing *x*’ is not taken to entail ‘*S* does believe that *x*, justifiably’. For *S* may not

form the belief because of a failure to see the connection between the propositions. Finally, with regard to CLO-P and ET-P, since a tautology is entailed by every proposition, the entailment must be restricted to some form of relevant entailment and/or the range of propositions must be restricted to contingent ones (see Relevance logic and entailment). Other restrictions are no doubt necessary; but these three seemingly intuitive principles have been challenged at their core.

It is important to see some of the relationships between these principles. CLO-P does not entail CON-P since the CLO-P is about *one* proposition that *S* is justified in believing, not *sets* of propositions. In addition, CLO-P does not entail ET-P because CLO-P does not require that it is the *very same evidence*, *e*, that *S* has for *x* that justifies *y* for *S*. Thus, one can accept CLO-P without accepting either of the other principles.

8 The epistemic principles and scepticism

Scepticism - the view that we lack knowledge in those areas commonly thought to be within our ken - comes in many varieties. The most extreme view is global scepticism. It holds that we have very little, if any, knowledge. That view seems preposterous at first glance. Indeed, some epistemologists think that any theory that leads to global scepticism should, *ipso facto*, be rejected (see Commonsensism; Scepticism). Yet there are many arguments for global scepticism that are difficult to answer. In addition, more modest forms of scepticism about particular subject matters (for example, other minds or the future) have been developed. But since the more modest sceptics employ strategies similar to those employed by the global sceptics, I here consider only the most extreme form of scepticism - global scepticism (see Other minds).

We have already seen the basis for one such argument for global scepticism that can be gleaned from the Pyrrhonians, namely:

(1) All knowledge requires having reasons that are neither arbitrary nor question-begging nor infinitely many.
(2) The only structures for reasons are such that reasons are either arbitrary (foundationalism), question-begging (coherentism) or infinitely many (infinitism).
Therefore, there is no knowledge.

There are at least four possible responses to this argument: (1) the foundational, basic propositions are not arbitrary; (2) coherentism does not necessarily lead to question-begging arguments; (3) requiring infinitely many reasons for a belief does not entail that a belief cannot be justified; (4) not all knowledge entails having reasons. All but (3) have been systematically developed by epistemologists.

Pyrrhonism does not rely directly upon the epistemic principles discussed in the preceding section. But there are other important forms of scepticism that do. Consider this argument that can be traced to Descartes (see Descartes, R. §4):

(1) If I am justified in believing that there is a table before me, then I am justified in believing that I am not in one of the sceptical scenarios (evil demon worlds, for example) in which there is no table but it appears just as though there were one.
(2) I am never justified in believing that I am not in one of the sceptical scenarios in which there is no table but it appears just as though there were one.
Therefore, I am never justified in believing that there is a table before me.

Premise 1 is a clear instance of CLO-P. Since the argument is valid (if the premises are true, the conclusion must be true), there are only three plausible responses: (1) CLO-P is false; (2) the second premise is false; (3) the argument begs the question. Responses (1) and (2) are relatively easy to envisage; the third is not so obvious. Roughly, the argument goes as follows: since one of the potentially available grounds for my being justified in believing that I am not in a sceptical scenario is any proposition that entails that I am not in such a scenario, every good argument for the second premise would have to establish that I am not justified in believing that there is a table before me. But that, of course, is the very conclusion.

It is important to note that there is an apparently similar argument for scepticism employing the stronger epistemic principle, ET-P:

(1) If the evidence, *e*, that I have for believing that there is a table before me is adequate to justify that belief, then

it is adequate to justify the belief that I am not in one of the sceptical scenarios.
(2) The evidence, *e*, is not adequate to justify that I am not in one of the sceptical scenarios.
Therefore, the evidence, *e*, is not adequate to justify that there is a table before me.

Of course, it is open to epistemologists to deny ET-P. Since one can deny ET-P without abandoning CLO-P (because CLO-P does not entail ET-P), that certainly seems to be a strategy worth considering. The discussion in the next section provides additional reasons for considering that strategy.

9 The epistemic principles and some paradoxes

There are many epistemic paradoxes (see Paradoxes, epistemic). I here consider two in order to show how they depend upon some of the epistemic principles considered earlier.

The Lottery Paradox: Suppose that enough tickets (say *n* tickets) have been sold in a fair lottery for you to be justified in believing that the one ticket you bought will not win. In fact, you are justified in believing about each ticket that it will not win. Thus, you are justified in believing the following individual propositions: t_1 will not win. t_2 will not win. t_3 will not win.… t_n will not win.

Now if the conjunction principle is correct, you can conjoin them, ending up with the obviously false but apparently justified proposition that no ticket will win. So, it seems that you are in the awkward position of being justified in believing each of a series of propositions individually, but not being justified in believing that they are all true. Some philosophers have thought that this seemingly awkward position is not so bad after all, since there is no outright contradiction among any of our beliefs as long as the conjunction principle is rejected. But others have thought that making it rational to hold, knowingly, a set of inconsistent beliefs is too high a price to pay.

Others have suggested that we are not actually justified in believing of any ticket that it will lose; rather what we are justified in believing is only that it is highly likely that it will lose. But the lottery can be made as large as one wants, so that any level of probability (below 1) is reached. Thus, this suggestion seems to rule out our being justified in believing any proposition with a probability of less than 1. That is a very high price to pay! There is no generally agreed-upon solution for handling the Lottery Paradox (see Probability theory and epistemology; Confirmation theory).

The Grue Paradox: The so-called ‘Grue Paradox’ was developed by Nelson Goodman and has been recast in many ways (see Goodman, N.). Here is a way that emphasizes the role of ET-P:

All of the very many emeralds examined up to the present moment, t_{now}, have been green. In fact, one would think that since we have examined so many of them, we are justified in believing that (G): all emeralds are green. But consider another proposition, namely that all emeralds examined up to t_{now} are green, but otherwise they are blue. Let us use ‘grue’ to stand for the property of being examined and green up to t_{now} but otherwise blue. It appears that the evidence which justifies us in believing that all emeralds are green does not justify us in believing that (N): no emerald is grue.

What are we to make of this version of the paradox? First, note that it depends upon ET-P. Although (1) our inductive evidence (the many examined green emeralds) justifies (G), and although (2) (G) does entail (N), the inductive evidence does not justify (N). In other words, this version of the paradox arises because the evidence does not transfer as the principle would require. Second, note that CLO-P is not threatened by this paradox since it is the evidence for (G) that is inadequate for (N). (The issue is not whether we are justified in believing (N) whenever we are justified in believing that (G).)

But if ET-P were not valid, then the sting of this version of the paradox can be pulled. Recall the original Grabit case. In that case, I had adequate evidence for being justified in believing that Tom stole the book, that is, the person stealing the book looked just like Tom. It seems clear that this evidence is not adequate to justify the proposition that it was not Tom’s identical twin who stole the book. If it were the twin, things would appear to be just as they did appear to be. But this tends to show that we do not typically impose ET-P on our evidence.

There are other versions of the Grue Paradox that do not make explicit use of ET-P. For example, since ‘all

emeralds are green' and 'all emeralds are grue' are alternative hypotheses, it seems paradoxical that the very same evidence that justifies believing the first alternative also seems to support the second. But perhaps, like the version considered above, this apparent paradox rests on a mistaken intuition. Consider the Grabit Case once again. Here, the evidence which justifies the belief that Tom is the thief would also support the claim that Tom's identical twin stole the book. To generalize further, consider any hypothesis, say *h*, that is justified by some evidence that does not entail *h*. It is always possible to formulate an alternative hypothesis that is supported by that very evidence, namely (not-*h*, but it appears just as though *h* because of…). Thus, an intuitively plausible epistemic principle similar to ET-P might be invalid. That principle is: if there is some evidence, *e*, that justifies *S* in believing that *x*, and *x* is an alternative hypothesis to *y*, then *e* does not support *y*.

To sum up, if ET-P and similar epistemic principles do not accurately capture our normative epistemic practices and if the argument for scepticism that depends upon CLO-P begs the question, then the sting of Cartesian scepticism (considered in the previous section) is numbed and the Grue Paradox can be addressed. But those are big 'ifs', and the issue remains open (see Induction, epistemic issues in).

10 Some challenges to traditional epistemology

A traditional question asked by epistemologists is 'what ought we to believe?' Typically, the answer is given by (1) describing the types of reasons that contribute to warranting a belief, and (2) developing a set of necessary and sufficient conditions for knowledge in which the types of reasons depicted in (1) play a prominent role. But there are many challenges to this answer.

We have already seen the challenge developed by the causal theorists and the reliabilists. Roughly, they hold that our beliefs need not be the result of proper reasoning to be counted as knowledge. Sufficiently reliable belief-acquisition methods are all that is required. Indeed, some have held that epistemology, when done correctly, is a branch of psychology because the primary issue is the study of reliable belief-acquisition methods. This programme has often been referred to as 'naturalized epistemology' and, in one form, its basic tenet is there are no a priori knowable epistemic principles (see Naturalized epistemology; Quine, W.V.).

Another challenge to traditional epistemology comes from 'virtue epistemology', which makes the primary object of epistemic evaluation traits of persons rather than properties of beliefs or belief-forming processes. The virture approach has been taken farthest by Linda Zagzebski (1996) who proposes an epistemic theory modelled on virtue ethics and argues that such a theory permits the recovering of such neglected epistemic values as understanding and wisdom (see Virtue epistemology).

A further type of challenge is that of Edward Craig (1990). While allowing that the debate has been shaped by real features of the concept of knowledge, he rejects the project of analysing it in necessary and sufficient conditions. Instead, he tries to 'synthesize' the concept by deriving these features from a pragmatic hypothesis about its purpose, thus explaining the debate rather than joining it.

Even more radical challenges have been developed. First, some have argued that there is no unique method of acquiring and revising beliefs that ought to be employed by all people (see Cognitive pluralism). Second, it has been argued that the proposed conditions of good reasoning (for example, objectivity and neutrality) tacitly aim at something other than truth. They are developed to prolong entrenched power (see Feminist epistemology). Finally, it has been argued that successful belief acquisition occurs when the future can be adequately anticipated and controlled (see Pragmatism).

The defenders of traditional epistemology have two basic types of reply. First, they can examine the particular arguments developed by the critics to determine whether any one of them is sound. Second, they can point out that the critics will have to defend the reasonableness of their views by at least tacitly employing the very principles of good reasoning investigated by traditional epistemologists. Of course, this would not show that the critic's position is false, but it does at least illustrate the universality of the question 'what ought we to believe?'.

See also: Daoist philosophy §5; Knowledge, Indian views of; Kūkai; Linji; Wang Yangming

PETER D. KLEIN

References and further reading

Alston, W. (1989) *Epistemic Justification*, Ithaca, NY: Cornell University Press. (Develops and defends a theory of knowledge containing elements of internalism and externalism.)

Annis, D. (1978) ‘A contextualist theory of epistemic justification’, *American Philosophical Quarterly* 15 (3): 213-19.(A non-technical contextualist account of justification.)

Aristotle (3rd century BC) *Posterior Analytics*, ed. and trans. J. Barnes, Oxford: Clarendon Press, 1975.(Aristotle proposes and defends an empirical foundationalist account of knowledge. It can be viewed as containing a basis for a reliabilist account of knowledge of first principles in science.)

Armstrong, D. (1973) *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.(One of the first carefully developed reliabilist accounts of knowledge.)

Audi, R. (1993) *The Structure of Justification*, Cambridge: Cambridge University Press.(Develops and defends a version of foundationalism.)

Ayer, A.J. (1940) *Foundations of Empirical Knowledge*, London: Macmillan. (Develops and defends a foundationalist account of the structure of reasons.)

BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.(Contains a sophisticated defence of coherentism.)

Broad, C.D. (1965) ‘The Theory of Sensa’ in R. Swartz (ed.) *Perceiving, Sensing and Knowing*, Berkeley, CA: University of California Press.(Develops the sense-data theory of knowledge. The collection contains many of the most important papers on perception written in the early- and mid-twentieth century.)

Chisholm, R. (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall; 2nd edn, 1977; 3rd edn, 1989. (The successive editions contain increasingly complex foundationalist accounts of knowledge along with versions of the defeasibility account.)

Code, L. (1991) *What Can She Know? Feminist Theory and the Construction of Knowledge*, Ithaca, NY: Cornell University Press.(Contains a sophisticated feminist challenge to traditional epistemology.)

Craig, E. (1990) *Knowledge and the State of Nature*, Oxford: Oxford University Press.(The concept of knowledge approached by asking ‘why do we have it?’ Assumes some familiarity with the current debate. Mentioned in §10 above.)

Davidson, D. (1983) ‘A coherence theory of truth and knowledge’, in E. LePore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell, 1986.(Contains an account of the coherence theory of knowledge, as well as arguments for the claim that coherent beliefs must be true in the main.)

DeRose, K. (1995) ‘Solving the Skeptical Problem’, *Philosophical Review* 104 (January): 1-52.(Develops a contextualist theory of knowledge and uses it to address the problem of scepticism.)

Descartes, R. (1641) *Meditations on First Philosophy*, in E. Haldane and G.R.T. Ross (eds) *The Philosophical Works of Descartes*, vol. 1, Mineola, NY: Dover Publications, 1955.(Contains a classic formulation of rationalistic foundationalism. Meditation I contains the ‘Cartesian’ argument for scepticism which he rejects in the following five meditations; Mediation IV employs the notion of warrant requiring non-accidentally true beliefs - see especially paragraphs 11-13 on page 176 of this edition.)

Dewey, J. (1929) *The Quest for Certainty: Gifford Lectures 1929*, New York: Capricorn, 1960.(Contains a contextualist account of doubt and justification.)

Dretske F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(Contains a reliabilist account of knowledge employing information theory.)

Foley, R. (1987) *A Theory of Epistemic Rationality*, Cambridge, MA: Harvard University Press.(Develops and defends a sophisticated version of subjective justification.)

Gettier, E. (1963) ‘Is justified true belief knowledge?’, *Analysis* 23 (6): 121-3.(This article was responsible for focusing attention on the inadequacies of characterizing warrant in terms of justification alone.)

Goldman, A. (1967) ‘A causal theory of knowing’, *Journal of Philosophy* 64 (12) 357-72.(The first careful statement of the causal theory of warrant.)

Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press. (Contains a sophisticated development and defence of reliabilism.)

Goodman, N. (1965) *Fact, Fiction and Forecast*, New York: Bobbs-Merrill, 2nd edn.(Contains the formulation of the Grue Paradox discussed in §9 above.)

Klein, P. (1981) *Certainty: A Refutation of Scepticism*, Minneapolis, MN: University of Minnesota Press. (Examines various forms of scepticism and develops the defeasibility theory of knowledge as a response to scepticism.)

Klein, P. (1996) ‘Skepticism and closure: why the evil genius argument fails’, *Philosophical Topics* 23 (1) spring 1995: 215-38.(Develops the ‘question-begging’ reply to scepticism briefly discussed in §8 above.)

Lehrer, K. (1990) *Theory of Knowledge*, Boulder, CO: Westview Press.(An accessible introduction to the fundamental questions in epistemology that defends a version of coherentism and contains arguments against externalism including the TrueTemp example cited in §6 above; see especially pages 163-75.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. A.C. Fraser, Mineola, New York: Dover Publications, 1959.(Contains the classic defence of empirical foundationalism conforming to the constraint that knowledge cannot be accidentally true belief. See especially Book XI, chapter 23, section 28 - pages 413- 14 of this edition.)

Lucey, K. (1996) *On Knowing and the Known*, Buffalo, NY: Prometheus Books.(A comprehensive and accessible collection of essays on the concept of knowledge.)

Moore, G.E. (1953) *Some Main Problems of Philosophy*, New York: Collier Books, 1962.(The text of lectures given in 1910-11. See especially chapter 2 which develops the sense-data foundationalist theory of knowledge.)

Moser, P. (1989) *Knowledge and Evidence*, Cambridge: Cambridge University Press. (Contains a sophisticated development of foundationalism.)

Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press.(Develops and defends a reliabilist account of knowledge.)

Plantinga, A. (1993) *Warrant: The Current Debate*, Oxford: Oxford University Press. (A good source for discussions of various accounts of warrant.)

Plato (4th century BC) *Theaetetus*, in *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1961.(Suggests that knowledge cannot be mere true belief even with a justification; but Plato does not suggest what the missing feature is.)

Pollock, J. (1986) *Contemporary Theories of Knowledge*, Totowa, NJ: Rowman & Littlefield.(Examines various contemporary accounts of knowledge and justification and develops a sophisticated version of the defeasibility theory.)

Prichard, H.A. (1950) *Knowledge and Perception*, Oxford: Clarendon Press.(Defends the view that knowledge is not a species of belief. See page 88 and following.)

Quine, W.V.O and Ullian, J. (1978) *The Web of Belief*, 2nd edn, New York: Random House.(A very accessible defence of coherentism.)

Radford, C. (1966) ‘Knowledge - by examples’ *Analysis* 27 (1): 1-11.(Defends the view that belief is not a necessary condition of knowledge.)

Sextus Empiricus (*c.*200) *Outlines of Pyrrhonism*, trans. R.G. Bury, Cambridge, MA: Harvard University Press and London: Heinemann, 1976.(See especially book I, chapter 15, for the argument that the three logically possible theories of justification lead to scepticism.)

Shope, R. (1983) *The Analysis of Knowing*, Princeton, NJ: Princeton University Press. (A thorough discussion of the Gettier Problem and the various approaches to solving it.)

Sosa, E. (1991) *Knowledge in Perspective*, Cambridge: Cambridge University Press. (Contains an interesting version of reliabilism that is designed to address issues generated by traditional normative epistemology.)

Sosa, E. (1994) *Knowledge and Justification*, vols 1 and 2, Brookfield, VT: Ashgate Publishing Company. (Contains a comprehensive set of essays on knowledge and justification.)

Zagzebski, L. (1996) *Virtues of the Mind*, Cambridge: Cambridge University Press.(Develops an account of virtue epistemology based on virtue ethics.)

Knowledge, defeasibility theory of

Based upon an analogy with the legal and ethical concept of a defeasible, or prima facie*, obligation, epistemic defeasibility was introduced into epistemology as an ingredient in one of the main strategies for dealing with Gettier cases. In these cases, an individual's justified true belief fails to count as knowledge because the justification is defective as a source of knowledge. According to the defeasibility theory of knowledge, the defect involved can be characterized in terms of evidence that the subject does not possess which overrides, or defeats, the subject's* prima facie *justification for belief. This account holds that knowledge is indefeasibly justified true belief. It has significant advantages over other attempts to modify the traditional analysis of knowledge in response to the Gettier examples. Care must be taken, however, in the definition of defeasibility.*

1 Chisholm's account

The concept of epistemic defeasibility was introduced into epistemology as an ingredient in one of the main strategies for dealing with Gettier cases and other instances in which an epistemic justification for a belief fails to provide the believer with knowledge, even though the resulting belief is true (see Gettier problems). An early discussion by Roderick Chisholm noted that the problems raised by Gettier cases are analogous to problems that arise in ethical contexts. In the ethical case, we have the concept of a *prima facie* obligation, one that might be overridden. When there is a conflict between circumstances that create a *prima facie* obligation and special additional circumstances that override them, the original obligation is defeated and is no longer actual.

Building upon this notion, Chisholm defines an epistemic version of defeasibility, a concept that applies by analogy to justification for a belief. When a person's belief that *h* is *prima facie* justified by the evidence they possess, it is possible that additional evidence would undermine that justification. If this were to happen, the original justification would be defeated. Chisholm (1964: 148-9) defines epistemic defeasibility in two steps:

(1) There is a justification for *h* which has been overridden =df. There is a body of evidence *e* and a body of evidence *e′* such that *e* is true and *e* justifies *h*, and *e′* is true and the conjunction of *e* and *e′* does not justify *h*.
(2) A justification for *h* is defeasible =df. There is a body of evidence *e* such that *e* is true and *e* justifies *h* and this justification may be overridden.

The opposite of defeasibility is *indefeasibility*, which is then defined as justification which cannot be overridden.

In these definitions, it is assumed that epistemic justification is of the 'linear' variety normally associated with foundational theories of justification. This assumption is not necessary for an account of defeasibility. Defeasibility conditions can be incorporated into a coherence theory of justification (as is done by Lehrer 1990) or into a reliabilist account (as is done by Swain 1981) (see Justification, epistemic §§2-3). For ease of exposition, the linear notion of justification will be assumed here.

2 The traditional analysis and Gettier cases

The traditional analysis of knowledge maintains that a person has knowledge provided they have a justified true belief (see Knowledge, concept of). In the kind of counterexample introduced by Gettier, an individual has a justification for some true belief, but this justified true belief is not knowledge because the justification is defective as a source of knowledge. In all such cases, the individual has arrived at the justified true belief by reasoning which depends upon an essential intermediate premise which, though justified, is false. Were the falsity of this essential intermediate premise brought forth as the additional evidence *e′* in the definition (1) above, the resulting conjunction of *e* with *e′* would not justify the proposition *h*. Hence, the justification that the individual has in a Gettier case is defeasible, and this is taken to explain why the individual does not have knowledge. Specifically, the proposal is to modify the traditional definition of knowledge as justified true belief to include an additional requirement: knowledge is *indefeasibly* justified true belief. On this proposal, known as the defeasibility theory of knowledge, a justified belief will count as knowledge only if the justification cannot be overridden by the unpossessed evidence.

Consider a specific example, introduced by Keith Lehrer (1965). Suppose Smith correctly infers that (*h*) someone in his class owns a Ford from some true evidence (*e*) that justifies the false belief (*q*) that a student, Mr Nogot,

owns a Ford. It so happens that another student in Smith's class, Mr. Havit, does own a Ford, but Smith has no evidence one way or the other for this proposition. Smith's justification for *h*, which consists of *e* and his inference to *h* through the false but justified proposition (*q*) that Mr. Nogot owns a Ford, is nevertheless rendered defeasible by the true proposition *e*′, which asserts that 'Nogot does not own a Ford'. Because Smith's justification is defeasible, he does not know that someone in his class owns a Ford. All Gettier cases have essentially this same structure, and so the defeasibility theory provides a general solution to the problem they pose for the traditional analysis of knowledge.

3 Advantages of the defeasibility approach

The defeasibility theory has advantages over other, simpler, proposals for modifying the traditional analysis of knowledge in response to Gettier cases. For example, the traditional definition can be revised to require that one's justification not involve or depend essentially upon any false intermediate premises. Or, the definition of epistemic justification can be strengthened to require that only true beliefs can be epistemically justified, blocking the move through (*q*) in the above example. However, these alternative solutions, which are aimed at specific features exemplified by Gettier cases, do not result in an acceptable general account of knowledge. There are examples which show that the traditional analysis fails even when all the evidence, essential intermediate lemmas and final conclusions involved in the justification of a belief, are true. The defeasibility analysis of knowledge has the merit of providing an adequate account even in these additional examples.

One illustration of the merits of the defeasibility account of knowledge over others is the well-known 'Barn Facsimile' example introduced by Alvin Goldman (1976). A subject, *S*, is looking at a barn under normal conditions of perception. *S* comes to believe, correctly, that there is a barn there on the basis of the perceptual evidence gained by looking at the barn. It may be assumed that all of *S*'s essential evidence is true, *S* correctly infers that there is a barn present in a manner that does not depend upon any false propositions, and *S* arrives at a true belief. Even so, *S* does not know that there is a barn there because of unusual surrounding circumstances. Unknown to *S*, the neighbourhood is populated by a number of structures which consist only of barn facades propped up by sticks. The only real barn in the neighbourhood is the one that *S* happens to see, and it is just by chance that *S* is perceiving that structure rather than one of the façades. Had *S* perceived a façade rather than a barn, *S* would have mistaken it for a real barn. All efforts to modify the traditional analysis of knowledge by disallowing justifications that depend essentially upon false beliefs will fail to rule out this example, since there are no false beliefs involved. The defeasibility account gives the right answer, however, for in this case the subject's justification is rendered defeasible by the availability of the true proposition *e*′, which asserts that 'Virtually all the things in this neighbourhood that look like barns are actually barn facades'.

4 Problems facing the defeasibility approach

Perhaps the most serious problem facing the defeasibility account, as it is expressed in the definitions (1) and (2) above, is the existence of examples in which there are special circumstances that would override a subject's justification even though that justification is not defeasible. As formulated, the account of defeasibility is too strong. This phenomenon is illustrated in the well-known example of Tom Grabit, presented by Lehrer and Paxson (1969). Suppose you observe your friend, Tom Grabit, in the act of stealing a book from the library. You know Tom well, and have no doubt whatever that it is Tom who stole the book. Your belief is true, your evidence is true, and your justification does not essentially involve any false beliefs. It would seem obvious that you have knowledge. Unknown to you, however, Tom's mother, upon hearing that her son has been arrested, fabricates a lie: she tells the police that it could not have been Tom, for Tom is in some other city. Instead, she says, it was almost certainly Tom's twin brother Tim, who is a kleptomaniac. But there is no twin brother; it was in fact Tom who stole the book. Given this scenario, your justification will be rendered defeasible by the availability of the true proposition *e*′, which asserts that 'Tom's mother says Tom was out of town at the time of the theft'.

Several modifications of the defeasibility theory have been suggested that would make it sensitive to such examples of apparent or pseudo-defeasibility. Gilbert Harman (1973) suggests that we make a distinction within the class of potentially undermining counter-evidence - evidence that would qualify as overriding by (1) above - between that which is genuinely defeating and that which is not. Then, undermining evidence will prevent knowledge only if it is genuinely defeating, as it is not in the case of Tom Grabit. Unfortunately, Harman left the

distinction between genuine and non-genuine defeating counter-evidence undefined.

Another proposal, made by David Annis (1973), suggests a distinction between genuinely defeating and 'merely misleading' counter-evidence. In the Grabit example, the mother's testimony is merely misleading and not genuinely defeating. Although this proposal has considerable intuitive appeal, Annis was not able to provide a generally accepted account of the key notion of 'merely misleading' counter-evidence. An attempt to make this distinction precise was made by Peter Klein (1980 and 1981), who suggests that the misleading effect of a defeater such as the one in the Grabit example is explained by the fact that it depends essentially on a false proposition. In genuine defeat, the defeater does not so depend. This proposal may prove to be adequate, although there is some debate about the notion of essential dependence.

Yet another proposal, suggested by Barker (1976), Swain (1974, 1981), Lehrer (1990) and Pollock (1990), attempts to take into account the defeasible nature of the defeasibility relation itself. In judging whether a subject's justification is defeasible, it is a mistake to consider isolated members of the total body of unpossessed evidence which is available to the subject, as is encouraged by (1) and (2). In many cases, such as the Grabit example, a justification's apparent defeat is itself only *prima facie*. It is itself defeated by further unpossessed evidence or special circumstances. The apparently defeating effect of the mother's testimony is itself overridden by the additional circumstance that she is lying. The final answer to the question of defeasibility of a justification in a given situation can only be arrived at when all of the relevant available counter-evidence, countervailing counter-evidence and so on, is taken into account. If a justification still holds when all the special circumstances are taken into account, it is said to be 'ultimately undefeated'. In the Gettier cases, the barn-facsimile example and other cases in which the subject genuinely lacks knowledge, their justification is ultimately defeated. In genuine cases of knowledge, even if some isolated undermining counter-evidence is available, the subject's justification will be ultimately undefeated.

This final proposal has also come under critical attack. Klein has argued (1980, 1981), for example, that even in cases of genuine defeat it will turn out that the justification is ultimately undefeated, since the 'original', or target proposition is true. Given these various disagreements, it seems fair to say that the defeasibility theory remains one of the most viable alternative approaches to the analysis of knowledge, but that there is no unanimity concerning the best way to handle the misleading defeater cases.

MARSHALL SWAIN

References and further reading

All these items involve intricate argument and moderate use of technical terminology.

Annis, D.B. (1973) 'Knowledge and Defeasibility', *Philosophical Studies* 24: 199-203; repr. in G.S. Pappas and M. Swain (eds) *Essays on Knowledge and Justification*, Ithaca, NY and New York: Cornell University Press, 1978.(Introduces the distinction between genuinely defeating and 'merely misleading' counter-evidence, and suggests an account of both notions.)

Barker, J. (1976) 'What You Don't Know Won't Hurt You?', *American Philosophical Quarterly* 13: 303-8.(Early discussion of the complications involved in defeasibility.)

Chisholm, R. (1964) 'The Ethics of Requirement', *American Philosophical Quarterly* 1: 147-53.(Introduces the analogy between ethical and epistemological defeasibility.)

Goldman, A. (1976) 'Discrimination and Perceptual Knowledge', *Journal of Philosophy* 73: 771-91; repr. in G.S. Pappas and M. Swain (eds) *Essays on Knowledge and Justification*, Ithaca, NY and New York: Cornell University Press, 1978.(Presentation of the 'barn facade' example and other well-known considerations.)

Harman, G. (1973) *Thought*, Princeton, NJ: Princeton University Press.(Chapter 9 provides a sustained discussion and defence of the suggestion that reasoning through false premises cannot provide one with knowledge.)

Klein, P.D. (1980) 'Misleading Evidence and the Restoration of Justification', *Philosophical Studies* 37 (1): 81-99.(A self-contained discussion of the 'merely misleading defeaters' solution.)

Klein, P.D. (1981) *Certainty: A Refutation of Scepticism*, Minneapolis, MN: University of Minnesota Press.(See especially pages 137-70 for a sustained discussion of defeasibility.)

Lehrer, K. (1965) 'Knowledge, Truth, and Evidence', *Analysis* 25: 168-75.(The 'Nogot-Havit' example can be found in this early discussion of Gettier problems.)

Lehrer, K. (1990) *Theory of Knowledge*, Boulder, CO: Westview Press.(Chapters 1 and 7 provide a detailed discussion of defeasibility and knowledge. Excellent bibliography to 1990.)

Lehrer, K. and Paxson, T. (1969) ‘Knowledge: Undefeated Justified True Belief’, *Journal of Philosophy*, 66: 225-37; repr. in M. Roth, and L. Galis (eds) *Knowing: Essays in the Analysis of Knowing*, New York: Random House; also repr. in G.S. Pappas and M. Swain (eds) *Essays on Knowledge and Justification*, Ithaca, NY and New York: Cornell University Press, 1978.(Early discussion of defeasibility theory, containing several important examples.

Pollock, J. (1986) *Contemporary Theories of Knowledge*, Lanham, MD: Rowman & Littlefield, 1991.(The appendix contains a discussion of defeasibility theory.)

Shope, R.K. (1983) *The Analysis of Knowing: A Decade of Research*, Princeton, NJ: Princeton University Press. (Chapters 1 and 2 provide a thorough survey of defeasibility theories. The remainder of the book is a survey of attempts to analyse knowledge, and there is an excellent bibliography to 1983.)

Swain, M. (1974) ‘Epistemic Defeasibility’, *American Philosophical Quarterly* 11: 15-25; repr. in G.S. Pappas and M. Swain (eds) *Essays on Knowledge and Justification*, Ithaca, NY and New York: Cornell University Press, 1978.(Early discussion of defeasibility, with a detailed proposal and several important examples.)

Swain, M. (1981) *Reasons and Knowledge*, Ithaca, NY and New York: Cornell University Press.(See especially chapter 5 for an extended discussion of defeasible justification.)

Knowledge, Indian views of

Classical Indian epistemology centres on a complex of terms for knowledge, knower and the known or knowable, including pramāṇa*, 'means to knowledge' or 'source of knowledge'. Views about perception, inference, testimony and a few additional candidate sources are the topics of core proposals of competing epistemological theories. Certain types of scepticism are also addressed, but explaining how it is possible that we know anything has been less central than other issues. Debates about knowledge - and doubt as well - are often caught up in larger war plans concerning the nature of awareness.*

The various classical schools typically bring views about awareness with them to the epistemological arena, but a neutral, common touchstone for and important constraint on all pramāṇa *theorizing is what is called speech behaviour,* vyavahāra*, reflecting, it is presumed, bits of everyday knowledge. Verbalizations of perception, for example, 'That is a pot', of inference, for example, 'There is fire on yonder mountain' (made on the basis of the sight of smoke and an understanding of the general rule that wherever there is smoke, there is fire), of information acquired through testimony, and so on, are the givens for which a successful theory has to account.*

Principal candidate sources proposed in addition to perception and inference are testimony, analogy, circumstantial implication and negative perception. Mystical experience as a pramāṇa *for spiritual matters is viewed as a variety of perception by its advocates, and scripture as a variety of testimony. With stock examples of bits of knowledge agreed upon, disagreement typically centres on what the source is for a particular example and whether admission of any source in addition to perception and inference is ever required. Or, in some cases, a stock example is slightly modified, better to align with a stance taken on a putatively additional* pramāṇa*.*

With regard to what the sources make known, some argue that each pramāṇa *works within a range of possibilities unique to itself, with no overlap. Thus what is known by perception cannot be known by inference. Others dispute such contentions, although at least a few such restrictions on individual knowledge sources are usually recognized. Buddhists and some others appear to be motivated to deny* pramāṇa *status to testimony because appeal to testimony is used to justify what they see as objectionable religious theses. Similarly, the Cārvāka materialist denies inference, apparently out of fear of its power to prove the existence of spiritual entities such as God or the soul.*

The Buddhist Nāgārjuna and others challenge the pramāṇa *programmes proffered by epistemologists of all stripes, and provoke what may be called meta-epistemological responses that bring out connections between* pramāṇa *proposals and a logic of presumption. In particular, the Nyāya response to Nāgārjuna and company is by any light an admirable effort of philosophy.*

1 Awareness, knowledge and doubt

In the classical Indian context, knowledge is treated as a species of awareness or cognition (*jñāna*), not of belief. A common opinion is that only a verbalizable or propositional awareness can be a bit of knowledge, and the examples of bits of knowledge talked about by classical epistemologists are expressed as sentences. For instance, the sensory awareness 'That is a pot', when a cognizer is perceptually presented with a pot, would, on most views, count as a bit of knowledge, although some theorists would object to this way of framing the matter. Awarenesses normally form beliefs in that there is - again, most would agree - a causal relation between an awareness and a memory or, more precisely, a 'disposition', *saṃskāra*. This reverberates with contemporary dispositional analyses (roughly: *A* is *S*'s belief if and only if *S* would answer certain questions in certain ways or behave in certain ways). Still, the bits of knowledge focused upon by classical Indian epistemologists are not beliefs but cognitions or awarenesses - right cognitions and veridical awarenesses.

Inferential knowledge is also an awareness, as too is knowledge acquired from any of the more exotic (and disputed) sources. Those who have become accustomed to talking about arguments in terms of propositions or sentences have to guard against imposing a nonpersonal understanding upon Indian discussions of logic in particular. Unlike in Western treatments, inference is not abstracted wholesale out of the context of persons who make inferences. The result of a good inference is an inferential awareness, a proposition-laden cognitive event arrived at by a process of understanding that such-and-such is the case because such-and-such else is the case. In

practice - that is to say, in the context of philosophical discussion and debate - there is, to be sure, abstraction from actual situations of knowing; otherwise, epistemology would be impossible since it is, at least normally, a discourse about cognitive events other than those provoked or reflected by the discourse itself (although self-referential, or meta-, epistemological discourse does occur: see §4). But the personal context is not lost sight of.

Another presupposition of classical Indian epistemology that is often missed or considered strange by those weaned on Western materials is the assumption that awareness, and thus knowledge, is episodic in character. In Plato’s *Meno*, Socrates argues that part of the value of knowledge is its steadfastness: *justifiable* true belief cannot be easily unpegged. One must keep in mind that ‘knowledge’ as used in the present entry renders a Sanskrit term, *pramā*, whose meaning has been the focus of extensive probing and controversy, comparable to the amount of philosophical attention knowledge has received in the West. There is no single English term or phrase that captures precisely the meaning of any of the family of Sanskrit words rendered here as ‘knowledge’, ‘knower’, ‘source of knowledge’, and so on. Many contemporary philosophers have detailed opinions about just what knowledge is, and so to render *pramā* as ‘knowledge’, for example, or *pramāṇa* as ‘source of knowledge’ may be especially to invite misunderstanding. The classical Indian approach to epistemology and translations such as ‘knowledge’, ‘knower’, and so on, have, however, been ably defended by several scholars and philosophers, most notably by Bimal Matilal (1986), who argues that to have knowledge in the episodic sense is a necessary (though not a sufficient) condition for knowledge in the sense of (roughly) a justified true belief. A belief not acquired through a reliable means will not count as knowledge. Classical Indian epistemologists focus on the event that is, one might say, the first formation of a bit of knowledge. With many, *pramāṇa* is to be understood both as evidence and as proximate cause of the first formation of any bit of knowledge - to use the term ‘knowledge’ in the nonepisodic sense. The dispositions required for reactivation or later use of a knowledge episode are not ignored by classical epistemologists. But again, the primary focus is on knowledge as an episode.

Doubt is also a cognitive episode or awareness, and it arises in specific situations whose general features we may become apprised of - this is another task of epistemology as practised in classical India. The *Nyāyasūtra* (*c.* AD 150; although it did not reach its present form until about 400), for example, says that doubt arises when, about the same object or topic, (a) there is insufficient information to determine what something is among things sharing common characteristics (for example, a post or a person in the distance), (b) there is insufficient information to determine what something is among things fundamentally distinct (for example, whether sound is a substance, quality or motion), (c) contradictory assertions are made and (d) there is inconstant information (for example, there seems to be water in the distance and then there does not). In this way, the *Nyāyasūtra* embraces a project of articulating conditions of meaningful doubt, a project that becomes much advanced during the following centuries. Philosophers of the Nyāya school take it to be a knockdown argument against a position that it provokes irresolvable doubt, and no school, not even Nāgārjuna’s Mādhyamika, which is famed for its scepticism about philosophy (see §4), holds doubt to be anything like a foundational, or natural, philosophical position. Doubt is a cognitive event that arises under certain conditions.

2 Candidate sources

Perception, inference and testimony are treated elsewhere (see Sense perception, Indian views of; Inference, Indian theories of; Testimony in Indian philosophy). So also is the issue of how negative facts are known (for example, Devadatta’s not being at home or the absence of my glasses on the table), and whether they are known through a special *pramāṇa* (see Negative facts in classical Indian philosophy). The present entry will air arguments surrounding the candidate sources of analogy and circumstantial implication. The question of mystical experience as a possible source of knowledge about spiritual matters is assimilated in classical discussions to the epistemology of perception.

An example commonly used of knowledge acquired through *upamāna*, ‘analogy’, is, according to Nyāya, as follows. A knower *S* has been informed by a forester that a *gavaya* (a kind of ox, *Bos gavaeus*), a species with which *S* has had no prior acquaintance, is similar to a cow except for certain differences that the forester spells out. *S* proceeds into the forest and encounters a *gavaya*, recognizes it as the animal the forester said was similar to a cow and shows the recognition by saying ‘Hey, a *gavaya*!’ Philosophers of the Nyāya school hold that *S* acquired a new piece of vocabulary (a bit of knowledge) through the analogy drawn by the forester, as shown by *S*’s correct

use of the word *gavaya* to designate the animal commonly referred to by that name.

Two principal contrary positions are, first, that what is known in such cases is not a new piece of vocabulary but a similarity, which, according to some but not all those who so endorse the *pramāṇa* status of *upamāna*, is a distinct category of the real. A second camp denies *pramāṇa* status to analogy no matter how understood; that is to say, analogy is no additional source of knowledge and the cited examples of knowledge from analogy are to be understood as examples of knowledge from testimony (in so far as testimony is itself an irreducible source) or perception or inference or a combination of these.

According to the Mīmāṃsāschool, which is intent on explaining and defending Brahmanical rituals, *upamāna* is the source for knowing the similarity of two things - a particularly useful bit of information when a substitute has to be found to complete a sacrifice or other ritual performance. For example, wild rice (*nīvāra*) can be substituted for domestic rice (*vrīhi*) because of their similarity. Also, when scripture does not elaborate the details of the performance of a particular rite *X* that is similar to rite *Y*, then rite *Y* can be used as a model for the performance of *X*. In the Mīmāṃsaka version of the stock example (which probably predates that in Nyāya), a knower *S* sees a *gavaya* and is reminded of a cow: 'Similar to this is my cow.' Analogy is said to be the proximate cause of knowledge of such a similarity. Against the arguments of Buddhists and others who hold that no special source need be postulated, Mīmāṃsakas respond by pointing out peculiar features of what is made known, along with strictures (some recognized by their opponents) on the range of operation of perception and inference.

It is difficult to see what motivates the Nyāya endorsement of analogy, since, as some of that very persuasion (such as the tenth-century philosopher Bhāsarvajña) insist, testimony and perception appear sufficient to account for vocabulary acquisition, at least in the stock example. Advaita Vedāntins, in contrast, have clear and substantial motives for endorsing analogy as understood roughly in the Mīmāṃsaka fashion: the true Self (*ātman*), as declared in scripture, is known as eternal, all-pervasive and unattached *like the sky*, and not as perishable, limited and prone to attachments - that is, it is known to be *unlike the body*. Such comparisons are grounded in facts that only the special source of analogy is capable of conveying. The Self is best or directly known in a meditative or mystical experience; no knowledge source, not even the testimony of scripture, operates except within the conditions of spiritual ignorance, *avidyā*, that are transcended in direct Self-experience. Nevertheless, scripture does refer to similarities and dissimilarities to help the mystic aspirant. So convinced, some late Advaitins speculate at length on the peculiar power of analogy.

Circumstantial implication, *arthāpatti* (sometimes rendered 'presumption'), is held to be another special and irreducible source according to Mīmāṃsakas and Vedāntins, and a type of inference according to most other disputants. Two stock examples are: (a) from discovery that Devadatta is not in his house it is known by implication that he is outside, and (b) from testimony about fat Devadatta that he does not eat during the day it is known by implication that he eats at night.

Kumārila (*c.*650), a Mīmāṃsaka with an extraordinary breadth of interest and the founder of one of two principal Mīmāṃsāsubschools, is often cited for his views on *arthāpatti* by philosophers of all persuasions. He lists six types of circumstantial implication according to how the critical circumstances are known, that is, according to the six sources identified by his school: perception, inference, analogy, circumstantial implication (making known the crucial circumstances in the present case, although the circumstances leading to it may be known differently), negative perception (the source for knowledge of absences) and testimony.

As comes to be common practice for wide-ranging philosophers of almost all points of view, Kumārila finds use for *arthāpatti* in establishing positions in ontology and metaphysics. Some modern commentators have seen *arthāpatti* as a pattern of reasoning commonly employed in the system-building typical of both Western and Eastern metaphysics, namely, reasoning to a vaunted best explanation. The best explanation of the information that fat Devadatta fasts by day is that he is a glutton at night, though there are other possibilities. Similarly, a Nyāya philosopher will argue against a Buddhist nominalist that postulation of real universals does a better job of explaining certain generalities or recurrences of experience than the Buddhist theory does. However, Kumārila and others seem to hold (it is probably Prabhākara, Kumārila's pupil and rival, and founder of the second principal Mīmāṃsāsubschool, who had the insight originally) that what makes *arthāpatti* effective is the absence of all other (live?) explanatory possibilities. (Prabhākara adds a proviso, disputed by the Kumārila camp: the absence of other explanatory possibilities in a given context of inquiry.) In reasoning to the best explanation, on the other hand,

alternatives are presumably admitted. Of course, there is no modal logic in classical Indian thought informing such reflections (until in a sense in very late New Logic, according to Sibajiban Bhattacharyya (1987)); we should avoid reading too much into classical discussions of albeit interesting examples. In any case, controversies in other domains, such as the basic categories of reality, come to bear on philosophers' understanding of *arthāpatti* - as they do too in particular with perception among other candidate sources.

Nyāya philosophers and others argue that circumstantial implication is a form of inference: roughly, from the premises (1) Devadatta is fat, (2) whoever is fat eats either at night or during the day, and (3) Devadatta does not eat during the day, one infers (4) Devadatta eats at night. In terms closer to the Sanskrit locutions used to express inferences, the invariable concomitance or pervasion, *vyāpti*, between occurrences of properties *F* and *G* that would underpin inference in this case would be: any person not eating at night or during the day exhibits absence of fatness (whatever is not-*G* is not-*F*).

3 Object-wise restrictions

Another important feature of classical Indian epistemology is the tendency to find restrictions governing the operations of individual knowledge sources, restrictions tied to the nature of the objects the sources reveal. Although objects are made known only by sources, knowing them we are apparently led to a view of epistemic limitations. And the individual sources themselves, it should be pointed out, can be objects of knowledge, at least according to the most prominent views.

Mīmāṃsāsūtra 1.1.4 presents an argument against perception's being the source of ethical knowledge: perception apprehends only what *already exists*, whereas ethics - *dharma* in Sanskrit, 'right practice' - is comprised of what one *should do*. The nature of *dharma* precludes perception as the knowledge source. Thus in so far as ethical knowledge is presumed, its source must be other than perception. It is, according to Mīmāṃsā, scripture, the Veda, interpreted as a list of injunctions prescribing what we should do. In this way, a limitation of perception concerning something presumed knowable (*prameya*, 'object of knowledge', in this case *dharma*) is used as a way to defend the authority, the *pramāṇa* status, of scripture.

Disputes between Nyāya realists and Buddhist Yogācāra idealists centre on conflicts over perception and inference. Object-wise distinctions shore up a Yogācāra partitioning of the two *pramāṇa*: there are unconceptualizable individuals on the one hand, and concept-fused generalities on the other. The Yogācārin Dignāga (§1) says:

> Now the means of cognition are [immediate and mediate, namely] perception and inference. They are only two, because the object to be cognized has [only] two aspects. Apart from the particular and the universal there is no other object to be cognized, and we shall prove that perception has only the particular for its object and inference only the universal.
>
> (*Pramāṇasamuccaya* 1.B)

And he goes on to present several considerations that bear on - or that elaborate - a rather trenchant division (see Buddhism, Yogācāra school of §§1-4). Nyāya, in contrast, finds universals such as cowhood as well as individuals such as Bessie the cow to be known perceptually. Inference is seen as dependent on perception in specified ways but also as filling out what is perceptually known, and a single object, such as a fire, can be known by perception or by inference or by both (or, indeed, by testimony) (see Nyāya-Vaiśeṣika §§5-6). Inference is dependent on perception both in that it is perception that establishes a required general thesis (paradigmatically, wherever there is *F*, there is *G*) and in that every inference, in providing knowledge of the world, is said to have to have - to get started, so to speak - a perceptually warranted premise. For example, from the sight of smoke rising from a hill, it is known by inference that there is fire there, too. (Several Nyāya philosophers point out that testimony can substitute for perception here, since, for example, we can be told that there is smoke on a hill and infer fire.) Dignāga and his school view inference as operating in a world of concepts or generalities all of its own, with the role of perception in concept formation as negative, as excluding (*apoha*) or ruling out what proves unpragmatic. To label Bessie a cow is to impose a generality of the mind on her unique particularity, a generality that, all things considered, is nevertheless better than its negation, a non-cow, with the opposition of cow and non-cow understood as a law of logic, not, as in Nyāya, a natural opposition, an incompatibility in the world (see Epistemology, Indian schools of §1).

The topic of object-wise restrictions on individual sources looms large in still other schools. Saōkara (*c*.700) presents an argument (probably inspired by the Mīmāṃsaka move respecting *dharma*) to shore up an Advaita absolutist metaphysics: proclamations of scripture (the Upaniṣads) have exclusive authority within their own province, namely, the reality and nature of Brahman, the Absolute (*Brahmasūtrabhāṣya* 2.1.1). Thus no evidence derived from perception, for example, could conflict with what the Upaniṣads say of Brahman. Similarly, the Upaniṣads cannot dispute perceptual evidence about objects in the world. Indeed, Saōkara holds that 'It is not the case that one source of knowledge (ever) opposes another source, for the one source makes something known that is not at all an object of the other' (*Bṛhadāraṇyakabhāṣya* 2.120). One wonders, however, whether Saōkara intends a position, like that of Yogācāra, according to which the perceptible and the inferable are distinct realities, or whether he intends only to stress (by exaggeration) the separate domain of scriptural authority.

Later Advaitins follow Saōkara's apparent endorsement of partition, at least in part. Dharmarāja (*c*.1600), the author of a popular Advaita textbook with an epistemological slant, holds for instance that the inferential cognition 'The hill is fiery' is perceptual and immediate with respect to the 'hill portion' but conceptual and mediate with respect to the 'fiery portion' (*Vedāntaparibhāṣā* 1.31). In Nyāya, it is insisted that we can touch what we see, and that even a cognition such as 'I see fragrant sandalwood' is legitimately deemed perceptual with respect to the property of being fragrant as well as with respect to the sandalwood. The relation between the grasping organ and the grasped property is said to be 'extraordinary'. The relation, or process, involves memory, as in the case of illusion, but the property is none the less perceived. Dharmarāja says, in contrast, that the sense of sight is not capable of grasping what is smelled. Nyāya philosophers agree with their adversaries that the organs of hearing, taste and smell are confined to sounds, tastes and odours respectively, and that depending on the precise nature of what is to be known there are restrictions on the other faculties. But they maintain that sight and touch range rather unrestrictedly over most properties most things exhibit, as do also the knowledge sources of inference and testimony.

4 Scepticism and epistemology

Focused, as explained, on the nature and number of knowledge sources, classical Indian epistemology met with scepticism practically from its inception. The Buddhist Nāgārjuna (*c.* AD 150) asks how - supposing that the known becomes known through *pramāṇas* - the *pramāṇas* themselves are established: 'If the *pramāṇas* are established through other *pramāṇas*, then there is an infinite series [and] neither the beginning nor the middle nor the end can then be established' (*Vigrahavyāvartani* 32-3). Next he explores two metapositions, finding neither satisfactory: (a) a foundationalist stance and (b) a view of a reciprocal making-known, with objects taken to be sources for knowing the usual sources as well as the reverse. Against the former view, which is expressed by an analogy to fire's illuminating itself along with other things, Nāgārjuna argues that, first, fire does not illuminate itself - as though it existed in darkness - and, second, if *pramāṇas* are self-established, without relation to objects made known (*prameyas*), then they are *pramāṇas* of nothing. Against the latter view, which specifies that the *pramāṇas* are established as *pramāṇas* in relation to the objects they make known (*prameyas*), he says that without being first established itself, nothing could establish another. A son is incapable of begetting his father. Thus all *pramāṇa* programmes are bankrupt.

In the *Nyāyasūtra* and commentaries, a response to such epistemological scepticism is forged whose linchpin is a view about the burden of proof and the relevance of citing *pramāṇas* varying according to specific contexts of inquiry. Consider what you would do if you were worried about the correctness of a scale used to weigh gold, says Vātsyāyana (5th century) in commenting on *Nyāyasūtra* 2.1.16, 'as a scale can be an object of knowledge as well'. According to this earliest *Nyāyasūtra* commentator, you would take a piece of gold to another scale, or to two or three other scales, determine its precise weight and bring it back to calibrate the scale in question. What was formerly an instrument for determining the weight of gold - thus a *pramāṇa*, in the sense that the scale is what is consulted about weight - becomes what is to be made known, *prameya*, in this special instance. Similarly, a piece of gold, which was the object of knowledge, becomes a source of knowledge. There is no rule that a specific instance of operation of a knowledge source - whether perception or inference or whatever - may not itself become the object of inquiry. In that case, some other instrument or knowledge source would be employed.

The complete Nyāya response to Nāgārjuna includes four further points. First, the threat of an infinite series is faced by a purported willingness to try to provide answers as long as there are meaningful questions. Second,

infinite series, it is pointed out, do occur, for example, a seed, a sprout, a seed, and so on. If you want us to answer the question as to what stands at the end of the series, say Nyāya philosophers, then we say that your question is framed in confused terms, that there is no possible answer: infinite series have no ends. We will consider issues of justification as far back as you wish, but - and here is the first point again - doubt has to be meaningful (see §1). Thus Nyāya philosophers assume that the balance of *reasonable* doubt is not tipped against them. Third, if the question is what justifies the entire *pramāṇa* programme, the answer is inference based on an invariable concomitance which itself is known from wide experience of the success of action guided by awarenesses whose sources are perception, testimony, and so on. If the basis of this and all inference is challenged, it is pointed out - fourth and finally - that to argue against *pramāṇas* is self-defeating, for what could be the source of one's argument? In particular, Gaōgeśa (*c*.1325) counters that the sceptic who opens his mouth to attack the foundation of inference, namely invariable concomitance (or *vyāpti*, 'pervasion'), commits a pragmatic contradiction, for his behaviour shows that he assumes that there is such a concomitance between speaking and communicating to another. Similarly, the sceptic who presents an argument against inference as a *pramāṇa* cannot be in good faith.

See also: Awareness in Indian thought; Error and illusion, Indian conceptions of

STEPHEN H. PHILLIPS

References and further reading

Bhattacharyya, S. (1987) *Doubt, Belief and Knowledge*, New Delhi: Indian Council of Philosophical Research and Allied Publishers.(Contains excellent analyses of a range of cognitive phenomena, drawing on classical views of awareness as an episode in a temporal series; also introduces some of the reflections of the latest New Logicians, including their development of modal and cognitive logics.)

Datta, D.M. (1932) *Six Ways of Knowing*, London: Allen & Unwin.(Lucid exposition of the late Advaita Vedānta understanding of knowledge sources.)

Dharmarāja (*c*.1600) *Vedāntaparibhāṣā*, ed. and trans. S.S. Suryanarayana, *Vedāntaparibhāṣā by Dharmarāja Adhvarin*, Madras: Adyar Library and Research Centre, 1971.(An excellent translation of an important late classical Advaita text.)

Dignāga (*c*.500) *Pramāṇasamuccaya (Collected Writings on the Acquisition of Knowledge)*, chap. 1 trans. Masaaki Hattori, *Dignāga, On Perception*, Cambridge, MA: Harvard University Press, 1968.(This important Yogācāra Buddhist philosopher was eclipsed in his own school by the great Dharmakīrti; Hattori's translation is quoted in §3.)

Kumar, S. (1980) *Upamāna in Indian Philosophy*, Delhi: Eastern Book Linkers. (Comprehensive concerning the classical debate about analogy as a *pramāṇa*.)

Kumārila (*c*.650) *Ślokavārttika*, trans. G. Jha, Delhi: Sri Satguru, 1983.(This is a reprint of the 1908 edition; although difficult to read, it is the only available English translation of the work of the important Mīmāṃsaka philosopher.)

Matilal, B.K. (1986) *Perception: An Essay on Classical Indian Theories of Knowledge*, Oxford: Oxford University Press.(A modern classic by a premier scholar and philosopher; referred to in §1 for a defence of knowledge as episodic.)

Mīmāṃsāsūtra (*c*. AD 100) trans. Mohan Lal Sandal, *MīmāṃsāSūtras of Jaimini*, Delhi: Motilal Banarsidass, 1923-5, repr. 1980, 2 vols.(An introduction to Mīmāṃsāphilosophy accounts for a large portion of the first volume; the author's running commentary is, he says, based on the oldest extant, Śabara's.)

Nāgārjuna (*c*. AD 150) *Vigrahavyāvartani*, trans. K. Bhattacharya, *The Dialectical Method of Nāgārjuna*, Delhi: Motilal Banarsidass, 1978.(An annotated translation, quoted in §4, of a work in which the Buddhist philosopher attacks *pramāṇa* programmes.)

Nyāyasūtra (*c*. AD 150), trans. M. Gangopadhyay, *Nyāya-Sūtra with Vātsyāyana's Commentary*, Calcutta: Indian Studies, 1982.(A readable and accurate translation.)

Śaōkara (*c*.700) *Bṛhadāraṇyakabhāṣya*, ed. E. Roer, Osnabrück: Biblio Verlag, 1980.(Reprint of the 1849-56 edition; the translation of the quote in §3 is mine.)

Śaōkara (*c*.700) *Brahmasūtrabhāṣya*, ed. J.L. Shastri, Delhi: Motilal Banarsidass, 1980. (Referred to in §3.)

Sundaram, P.K. (1968) *Advaita Epistemology*, Madras University Philosophical Series 11, Madras: University of Madras.(A continuation and advance of Datta's exposition with a special focus on the *Iṣṭa-siddhi* of Vimuktātman.)

Knowledge, tacit

Tacit knowledge is a form of implicit knowledge we rely on for both learning and acting. The term derives from the work of Michael Polanyi (1891-1976) whose critique of positivistic philosophy of science grew into a fully developed theory of knowledge. Polanyi believed that the 'scientific' account of knowledge as a fully explicit formalizable body of statements did not allow for an adequate account of discovery and growth. In his account of tacit knowledge, knowledge has an ineliminable subjective dimension: we know much more than we can tell. This notion of tacit knowing in science has been developed by Thomas Kuhn, has figured prominently in theoretical linguistics and has also been studied in psychology.

Polanyi maintained that at each stage of scientific inquiry, from discovery to confirmation, a crucial role is played by experience, skill and expertise, each of which involves a kind of knowing on the part of the practitioner which is implicit and unformalizable. His distinction between tacit and explicit 'knowledge' is grounded in his distinction between subsidiary and focal 'awareness'. Drawing on Gestalt psychology's account of perceptual integration, whereby we are perceptually aware of the coherent whole to which we are attending on the basis of a subsidiary awareness of details or clues to which we are not attending, Polanyi develops a general account of awareness wherein our focused awareness of *B* is a function of our subsidiary awareness of *A* (see Gestalt psychology). The tacit integration of the subsidiary elements that makes possible our explicit recognition of the perceptual object becomes a model of a kind of implicit knowing that lies at the heart of our cognitive orientation to the world. This generalization is driven by the conviction that such tacit knowledge is basic to our understanding of language and to comprehending any complex situation. For Polanyi, tacit knowledge is not just one kind of knowing we are capable of; it is absolutely fundamental: 'all knowledge is either tacit or rooted in tacit knowledge' (1969: 144).

The feature of tacit knowledge that bears on knowing more than we can tell is both crucial and ambiguous. The distinction between subsidiary awareness and focal awareness seems to be merely functional in that items of subsidiary awareness in this experience can be objects of focal awareness in the next. Is the same true of the tacit-explicit distinction? Is it simply the case that all explicit knowing is rendered possible by a shifting background of implicit knowing, or are there some kinds of knowledge which are (at least for us) in principle non-explicit? On the weak interpretation 'knowing more than we can tell' is time-indexed and seems uncontroversial, whereas the strong interpretation invokes a species of knowledge shrouded in mystery. Recent attempts to construct computer programmes that would replicate the skills that involve tacit knowledge could shed light on the '*de facto*' versus 'in principle' character of the tacit-explicit distinction.

The basic notion that there is a dimension of scientific knowledge that is acquired only through practice and that cannot be (or is not) articulated explicitly has been developed in Thomas Kuhn's account of both normal and revolutionary science (see Kuhn, T.S. §§2-3). The recognition of what is significant in theoretical and experimental situations so fundamental to the conduct of science is learned in actually doing science rather than by the acquisition of rules for doing it. In these basic procedures the scientist has no direct access to what is known, nor generalizations in which to express it, but inasmuch as it is an ability that is transmitted through education it seems appropriate to categorize it as knowledge, albeit of the tacit variety. The distinction invoked here encompasses but is not exhausted by that articulated by Gilbert Ryle (1949) in terms of 'knowing how' and 'knowing that'. This tacit knowing which functions so pervasively in normal science also plays a critical role in those creative advances characterizable as scientific revolutions (see Knowledge, concept of §1).

The notion of tacit knowledge also figures prominently in theoretical linguistics. The debate in this area has been between the empiricist tradition (which construes language as an adventitious construct built up in a basically inductivist manner by elementary data-processing procedures) and the rationalist tradition (which maintains that behind these processing mechanisms there must be innate formal and substantive linguistic universals built into the language user to account for both acquisition and competence). The principal figure in the rationalist tradition has been Noam Chomsky and his theory invokes the notion of tacit knowledge. He argues that since the knowledge of a language involves the implicit ability to understand indefinitely many sentences, it seems reasonable to attribute to the competent speaker the grasp of a system of linguistic rules that can generate an indefinitely large number of structures. These rules are obviously not explicitly known by the speaker, but the tacit knowledge of them seems

necessary to explain not only competent performance but, in particular, the ability to learn any natural language (see Language, innateness of).

Much work has also been done in experimental psychology bearing on the notion of tacit knowledge. This research usually goes under the rubric ‘implicit versus explicit cognitive processing’ and is normally construed as mapping on to the unconscious-conscious distinction. Implicit learning and implicit memory have been the subjects of numerous studies and there seems to be considerable evidence, particularly from the study of amnesia patients, that implicit cognitive processes are more robust than their explicit counterparts. Given the emergence of cognitive science as a basic research programme, both in experimental psychology and philosophy, there has been a recent cross-fertilization of the inquiry into notion of tacit knowledge.

See also: Innate knowledge

C.F. DELANEY

References and further reading

Chomsky, N. (1965) *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press. (Chapter 1 contains a general discussion of the differences between the empiricist and the rationalist accounts of language learning.)

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press, enlarged 2nd edn, 1970.(The postscript employs the concept of tacit knowledge in an account of science.)

Polanyi, M. (1958) *Personal Knowledge: Towards a Post-Critical Philosophy*, Chicago, IL: University of Chicago Press.(The preface to this very expansive work contains brief but important observations on tacit knowledge.)

Polanyi, M. (1966) *The Tacit Dimension*, New York: Doubleday.(The first two lectures in this short book present an overview of Polanyi’s account of tacit knowledge.)

Polanyi, M. (1969) *Knowing and Being*, Chicago, IL: University of Chicago Press.(Part 3 of this collection contains some of Polanyi’s most important essays on tacit knowing.)

Reber, A. (1993) *Implicit Learning and Tacit Knowledge*, Oxford: Oxford University Press.(A good overview of recent empirical research bearing on implicit learning and implicit knowledge.)

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson’s University Library.(Chapter 2 of this classic work articulates the distinction between ‘knowing how’ and ‘knowing that’, and also provides a positive account of what is involved in ‘knowing how’.)

Knutzen, Martin (1713-51)

Martin Knutzen was a follower of Christian von Wolff. His work is the result of an effort to reconcile Wolff's system, more persuasively than Wolff himself had, with common sense, Christian faith and the latest results in the natural sciences. Because Wolff had come under fire from Christian Pietists for apparently trying to resurrect Leibniz' system of pre-established harmony, thereby possibly flirting with Spinozism and therefore atheism, Knutzen argued from Leibnizian premises that real interaction is at work in the world at large and is constitutive of the mind-body union. Knutzen was Kant's teacher.

Knutzen was born in Königsberg, East Prussia, the only child of a humble Danish merchant. Orphaned at the age of six, he was cared for by a relation who saw to his education. At fifteen, Knutzen entered the University of Königsberg where he developed an interest in the natural sciences and made two lifelong and apparently incompatible commitments: to Pietism, and to the philosophy of Christian von Wolff. Much of Knutzen's later philosophical work can be understood as an effort to reconcile these two commitments. At twenty one, he became professor extraordinary of logic and metaphysics at Königsberg. In spite of this brilliant debut, he never rose to the higher rank of professor ordinary. Knutzen was said to be an engaging lecturer. At any rate, he was a very busy one, teaching as many as six hours a day. In addition, he wrote much on philosophy, the natural sciences and theology.

Wolff had made the philosophy of Leibniz fashionable again by packaging it as an easy-to-digest system. The system was appealing, because it offered a comprehensive picture of God, the world, the human soul and things in general by elaborating the consequences of two principles: the law of contradiction and the principle of sufficient reason. As the system won followers, it came under critical scrutiny. Pietists attacked it for blasphemy, arguing that it reduced to Spinozism and thus to atheism (see Spinoza, B. de). Friends of common sense within and without the school reacted against the eccentric aspects of Leibniz' thought. Wolff's followers set out to reconcile the system more persuasively with Christian faith, common sense, and the latest results in the natural sciences. Knutzen was one of the most eminent of these innovating Wolffians.

A great stumbling block in Leibniz philosophy for common sense, and, in the eyes of some, for Christian faith, is the doctrine of pre-established harmony. According to this, no real interaction occurs among created substances. Every monad is itself the author of every change it ever undergoes. The appearance of interaction is the effect of God's having arranged things so that change in one substance would always arise in concert with change in every other. The doctrine also denies real interaction between body and soul (see Dualism). Their intimate union is the effect of selfproduced change in each, harmoniously pre-established by God. This doctrine was attacked by pietists, who complained that it represents the soul as a spiritual automaton, fatalistically spinning out its future under the principle of sufficient reason, and impiously absolves us of responsibility for our sins. But the doctrine met with hesitation even within the school, notably from Knutzen who tried to show, from Leibnizian premises, that interaction must be at work in the world, not pre-established harmony.

Knutzen's argument for real interaction borrows the following ideas from Leibniz as understood by Wolff: every body is composed of monads; every monad produces change in itself; by reason of their changing inner states, monads externally relate to one another. Since selfproduced change in one monad alters its external relations with other monads, they are also affected by such change; real interaction must therefore take place among the monads. There must also be real interaction between body and soul, argues Knutzen, because the soul is a simple substance. We may therefore expect the soul to interact with the monads, constitutive of the human body, just as these monads interact with one another. Real interaction of soul and body is properly understood as real interaction of soul and monads.

Leibniz had also argued that body-soul interaction would disturb the order of nature determined by the laws of motion, given that the total quantities of motion and 'living forces' in the world are always conserved. Knutzen agrees that we must reject any system of mind-body union which can be shown to disturb the order of nature. But he denies that, properly understood, this would ever be disturbed by real interaction. The order of nature, he says, is established by the set of rules that govern different forces in the world. But any given rule operates only under certain conditions. Knutzen says that the law of conservation of living forces holds only for bodies mechanically

interacting amongst themselves. If a body interacts with a soul, neither it nor the soul, which is immaterial, is subject to this conservation law. Interaction may therefore take place between them without disturbing the order of nature. Furthermore, Knutzen draws a parallel between the law of conservation of living forces and the law of conservation of motion. Like the former, the latter operates only under certain conditions, namely in the ideal case of collision between perfectly elastic bodies, which never obtains in the world. Since both conservation laws are restricted in important ways, their jurisdiction over nature is curtailed, leaving open the possibility of real interaction.

Perhaps due to the isolation of Königsberg, Knutzen was not renowned even in his day, though he was well known to figures more familiar to us than he himself is, such as Hamann and Kant. Knutzen's philosophy and lectures apparently influenced the latter, to whom they may have suggested the idea of reconciling Leibnizian metaphysics and Newtonian natural philosophy, as, for example, in Kant's dissertations of 1755-6.

ALISON LAYWINE

List of works

Knutzen, M. (1744) *Philosophische Abhandlung von der immateriellen Natur der Seele (Philosophical Treatise of the Immaterial Nature of the Soul)*, Königsberg: Hartung.(An attempt to refute materialism. Copies can be found in the libraries of Harvard University and the University of Chicago. They are bound with a German translation of William Warburton's *Divine Legation of Moses*, Frankfurt and Leipzig, 1751.)

Knutzen, M. (1745) *Systema causarum efficientium seu commentatio philosophica de commercio mentis et corporis per influxum physicum explicando (System of Efficient Causes or Philosophical Treatise on the Interaction of Mind and Body Explained through Physical Influence)*, Leipzig: Christian Langhem.(An attempt to refute pre-established harmony and establish real interaction. Copies can be found in the British Museum and the library of the Georg-August-Universität, Göttingen.)

Knutzen, M. (1747) *Elementa philosophiae rationalis logicae cum generalis tum specialis mathematica methodo in usum auditorum suorum demonstrata a Martino Knutzen*, Königsberg and Leipzig: Hartung.(A work of philosophical logic. Copies can be found in the libraries of Harvard University and the University of Chicago. A reprint is now available from Olms.)

References and further reading

Erdmann, B. (1876) *Martin Knutzen und seine Zeit*, Leipzig: Voss.(A very informative book about Knutzen, his work and the philosophical and historical context of his work.)

Euler, L. (1746) 'Enodatio quaestionis utrum materiae facultas cogitandi tribui possit necne ex principiis mechanicis petita', repr. in A. Speiser *et al.* (eds) *Opera omnia*, Zurich: Orell Füssli, 3rd series, vol. 2, 1911-. (A discussion of the question whether we can attribute a faculty of thought to matter. Euler explicitly refers to Knutzen.)

Kant, I. (1755) *Principiorum primorum cognitionis metaphysicae nova dilucidatio*, repr. in *Gesammelte Schriften*, vol. 1, Berlin: De Gruyter, 1968.(One of Kant's earliest writings. May show Knutzen's influence, inasmuch as Kant can be understood as trying to reconcile what he takes to be fundamental commitments of Leibniz and fundamental commitments of Newton.)

Kant, I. (1756) *Metaphysicae cum geometria iunctae usus in philosophia naturali, cuius specimen I. continet monadologiam physicam*, repr. in *Gesammelte Schriften*, vol. 1, Berlin: De Gruyter, 1968.(One of Kant's earliest writings. May show Knutzen's influence, inasmuch as Kant can be understood as trying to reconcile fundamental commitments of Leibniz and his followers in metaphysics with a fundamental commitment held by Newton and his followers concerning the nature of space.)

Wolff, C. von (1720) *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt (Sensible Thoughts on God, the World and the Human Soul, and Indeed on Everything)*, Hildesheim: Olms, 1983.(A comprehensive statement of Wolff's metaphysics.)

Kojève, Alexandre (1902-68)

Alexandre Kojève developed an idiosyncratic and widely influential reading of G.W.F. Hegel in a seminar in Paris from 1933 to 1939. Kojève read Hegel as having discovered that truth was the product of history, and that history was the product of the human desire and struggle for recognition. Kojève emphasized that once this desire was satisfied, history, properly so-called, was over. He claimed that for all essential purposes this human desire had been satisfied in the modern period, and thus that we had experienced (and Hegel had come to know) the end of history. The notes from this seminar were published in 1947 and continued to have an important impact on French philosophy throughout the post-war period.

1 Life

Kojève (born Kojèvnikoff) left Moscow at the age of 18, and spent most of the 1920s in Berlin and Heidelberg. During his years in Germany, he was strongly affected by the teaching of Martin Heidegger and by his fellow students in philosophy, Jacob Klein and Leo Strauss. By 1927 he had come to Paris, where he would give his Hegel seminar as a replacement for his friend, Alexandre Koyré. After the Second World War, Kojève worked as a negotiator and advisor for the French Ministry of Economic Affairs (especially on world trade matters). He continued to write on philosophy in his spare time.

2 The end of history

Kojève's early writings, especially those on the mystical philosophical texts of Vladimir Solov'ëv, show a concern with finding legitimate criteria for making sense of historical change. If these criteria were themselves the product of historical change, how could stable philosophical knowledge be possible? Kojève's idea of the 'end of history' (which he attributed to Hegel) was meant to answer this question.

Kojève acknowledged his departure from Hegel in his sharp separation of the natural from the historical. Kojève was interested only in the latter, and used the master-slave dialectic as a schema for understanding the stages of work and struggle in history. In this schema the human is defined by the desire for recognition - a desire that could be satisfied only with the conservation of its object (the person who provides love, honour, respect) - and the will to risk one's life in order to satisfy that desire. The realm of the human and of history is defined specifically in contradistinction to the realm of the animal and of nature.

The desire for recognition is the motor of history, and Kojève used the master-slave dialectic as an allegory of human development: bloody battles for recognition were followed by the rule of the masters over the working slaves. Slaves were those whose animal desires for self-preservation led them to surrender and to recognize the masters. The masters, however, could not satisfy their human desire, because they were recognized by *mere* slaves. Eventually the slaves deposed the masters through revolution but they remained in servitude in relation to their work. This is the bourgeois condition. Real freedom would come only through the universal recognition of each and every one as a citizen.

The final synthesis of the master-slave dialectic is the freedom of universal recognition because it provides full (human) satisfaction. It also represents the end of history, since once the desire for recognition is satisfied there will be no force to create historical (human) change. There would be merely the activity of beings seeking to satisfy animal (nonhuman) desires. In other words, rather than engaging in action to achieve freedom and recognition, after the end of history 'people' would find ways of acquiring toys of various sorts. For Kojève, the end of history was the 'definitive reality' that served as a transhistorical standard of judgment.

3 Later works

In the 1930s, Kojève situated his reading of Hegel just before a revolution that would be a final confirmation of the interpretation itself. History was ending in universal recognition, and Kojève was contributing to the self-consciousness that would be part of the final state. Perhaps in response to some important criticism of his views by Leo Strauss, and certainly in response to the increasingly congealed political situation of the early 1950s, I have argued that Kojève abandoned his 'heroic Hegelianism'. He continued to believe that the culmination of world history would define the truth of all previous events, but in his later work claimed that the end of history had

occurred at the beginning of the nineteenth century. This left the mid-twentieth-century philosopher in the ironic position of having nothing left to say, except to repeat that everything of importance had already been said in order to make sense of what had already been done.

The idea of the 'end of history' in Kojève's later work no longer refers to the triumphant ascension of humanity, but instead, for those who see that we live on the other side of it, signals a final decadence in which humans are distinguished from other animals only by their snobbery. Kojève's early combination of Marx, Hegel and Heidegger in the service of claiming the future of the world is thus incorporated into a perspective reminiscent of Weber on the routinization of life. The closure of the end of history is an iron cage in which human animals can engage in a variety of activities without struggle because their essential drives have been satisfied. Kojève takes an ironic stance towards this condition in which the struggle for recognition is replaced by conspicuous and endless consumption; historical change is replaced by animalistic or snobbish repetition.

Kojève's idea of the end of history spoke to the dissatisfaction of those on the left and right who thought that liberalism and modernity left no possibility either of genuine human life or of greatness. The philosopher continued his work among the bureaucrats until his death, helping to work out trade policies that might be suitable for the end of history.

See also: Hegelianism §4; History, philosophy of

MICHAEL S. ROTH

List of works

Kojève, A. (1946) 'Hegel, Marx et le christianisme', *Critique* 1: 339-66; trans. H. Gildin, 'Hegel, Marx and Christianity', *Interpretation* (1970) 1: 21-42; repr. in Stern (ed.), vol. 2: 359-82.(Argues for what he describes as an atheistic reading of Hegel.)

Kojève, A. (1947) *Introduction à la lecture de Hegel*, Paris: Gallimard; partly trans. J.H. Nichols, Jr as *Introduction to the Reading of Hegel*, ed. A. Bloom, New York: Basic Books, 1969.(Text of Kojève's influential lectures at l'École des Hautes-Études from 1933 to 1939.)

Kojève, A. (1968-73) *Essai d'une histoire raisonnée de la philosophie païenne (Attempt at a Systematic History of Pagan Philosophy)*, Paris: Gallimard, 3 vols.(An attempt to provide access to Hegelian wisdom by re-collecting the history leading up to it, from the Presocratics through Aristotle and the Neoplatonists.)

Kojève, A. (1973) *Kant*, Paris: Gallimard. (Kant after Hegel. In light of the latter's completion of philosophy in absolute knowing, the former's important contribution to bringing philosophy into relation with time is discussed.)

Kojève, A. (1981) *Esquisse d'une phenomenologie du droit (Sketch of a Phenomenology of Right)*, Paris: Gallimard.(Originally written during the Second World War.)

Kojève, A. (1991) 'The Political Action of Philosophers', in *Leo Strauss, On Tyranny*, ed. V. Gourevitch and M.S. Roth, New York: The Free Press.(Originally published in 1950; this edition also includes the correspondence between Strauss and Kojève.)

References and further reading

Auffret, D. (1990) *Alexandre Kojève: La philosophie, l'état, la fin de l'histoire*, Paris: Bernard Grasset.(A sketchy biography with some philosophical interpretation.)

Butler, J.P. (1987) *Subjects of Desire: Hegelian Reflections in Twentieth-Century France*, New York: Columbia University Press.(A contemporary perspective on how Hegel resonates throughout French philosophy from existentialism onwards.)

Cooper, B. (1984) *The End of History: An Essay on Modern Hegelianism*, Toronto, Ont.: University of Toronto Press.(A thoughtful account of Kojève's relevance to modernity.)

Drury, S. (1994) *Alexandre Kojève and the Roots of Postmodern Politics*, New York: St Martin's Press.(An attempt to link Kojève to a nihilistic, desperate brand of contemporary thinking about politics.)

Fukuyama, F. (1992) *The End of History and the Last Man*, New York: The Free Press.(Uses Kojève and Leo Strauss to proclaim the end of the cold war as the end of history.)

Gourevitch, V. (1968) 'Philosophy and Politics I', 'Philosophy and Politics II', *Review of Metaphysics* 22 (1): 58-84; 22 (2): 281-328.(A philosophical account of what is at stake in Kojève's thinking about politics, history and truth.)

Roth, M.S. (1988) *Knowing and History: Appropriations of Hegel in Twentieth-Century France*, Ithaca, NY:

Cornell University Press.(A discussion of Kojève as part of the Hegelian investment in the historical from the 1930s to the 1950s.)

Roth, M.S. (1995) *The Ironist's Cage: Memory, Trauma and the Construction of History*, New York: Columbia University Press, chaps 5, 8 and 9.(Discussions of Kojève, Strauss, Fukuyama and the idea of the end of history in relation to contemporary cultural criticism.)

Kokoro

Kokoro *is a comprehensive term in Japanese religion, philosophy and aesthetics often translated as 'heart', whose range of meanings includes mind, wisdom, aspiration, essence, attention, sincerity and sensibility. In Buddhist texts and in philosophy,* kokoro *(or* shin *in its Sino-Japanese reading) denotes mind, heart or inner nature, the site of human sentience or delusion. By extension, in pre-modern theories of art,* kokoro *signifies simultaneously the emotional capacity of the artist to respond to the natural world, which ideally catalyzes the act of creation; the parallel ability of an audience to respond to such a work of art and thus indirectly to the experience of the artist; and finally the evaluation of such a work as possessing the 'right conception',* kokoro ari *or alternatively* ushin.

While the specific implications of *kokoro* changed in the context of cultural development, its far-ranging meanings, which additionally include intention, vitality, knowledge, sentiment, wholeheartedness, reason, state, appearance, spirit, will and duty, imply both a dynamic potentiality and a fundamental concern with the authenticity of human being as well as its integrity. This integrity precludes any dualistic mind-body distinctions. Ultimately, *kokoro* serves as the foundation for almost all pre-modern Japanese theories of art, which may be described as expressive-affective in their orientation (see Aesthetics, Japanese).

In Buddhist writings we often encounter *kokoro*, especially in its compound form *shin*, as in the well-known Buddhist dictum *ishin denshin* (transmission from mind to mind), alluding to the ideal transmission of spiritual truth that is intuitive in nature and does not rely on verbal and other explicit means. In Buddhism, *kokoro* or *shin* is not simply the faculty for apprehending truth, but can also be the site of delusion and desire within human being, as suggested in the phrase *kokoro no yami* (the darkness of the heart), in other words, the spirit of the unawakened being still wallowing in the mire of the phenomenal world. In seeming contrast to the usage of the terms *ushin* and *mushin* in aesthetics, in Buddhism, *ushin* (having heart or mind) and *mushin* (lacking heart or mind) represent respectively the deluded state of being attached to the world (that is, possessing desires) and the liberating state of awareness that has eschewed the heart and hence all desire.

In the realm of pre-modern artistic theories, *kokoro* receives not only extensive theoretical treatment as a crucial artistic component, but becomes a term of aesthetic approbation as well. In the often quoted Japanese preface to the *Kokinwakashū(The Anthology of Japanese Poetry Ancient and Modern)*, *circa* 905, Ki no Tsurayuki conjoins the concept of *kokoro* with a parallel concept *kotoba* (words), and begins 'Japanese poetry has the human heart (*kokoro*) as its seed and burgeons forth into myriad words (*kotoba*) as its leaves.' He traces the genesis of poetry to the fact that all living beings respond to the natural world around them in the form of song, like the cries of birds in the midst of blossoms or frogs croaking in the water. Hence, *kokoro* is that faculty within us which hearkens to the world outside and poetry is what results when we give utterance to the feelings and thoughts in our hearts in the form of words. However, he also offers a secondary understanding of *kokoro* when he goes on to imply that *kokoro* (heart, conception, treatment) and *kotoba* (words, diction, materials) are the two constituent elements of any poem, which may be used to evaluate the success of any poetic endeavour. He criticizes, for example, a verse by a well-known poet of a former age Ariwara no Narihira with the comment *kokoro amarite, kotoba tarazu* (too much heart and too few words). Ideally, we infer, the poem must have a balance of *kokoro* and *kotoba*.

This interest in *kokoro* and *kotoba* is echoed and developed by the twelfth-century poet-critic Fujiwara Shunzei, who in surveying what he perceived as the second-rate nature of the poetry of his age, issues the challenge, *kotoba furuku, kokoro atarashi* (old words, new heart). He argues for the need to retain the traditional diction and materials, but to infuse them as well with a new spirit that would revitalize them. Apart from regarding *kokoro* as merely the resonant capacity within all living beings to feel and respond as well as one of the constituent components of poetry, Shunzei goes further and elevates *kokoro* into an aesthetic ideal known as *ushin*. His son Fujiwara Teika develops his father's theory still further, describing *ushin*, which may be translated as 'having heart', as signifying an intensity or conviction of feeling which is paramount. *Ushintei* (the style of *ushin*), as he comments, is the overarching style of the so-called ten styles of poetry; all great poetry for him must possess this intensity of feeling, whose articulation is influenced by the Tendai Buddhist contemplative practice of *shikan*, or concentrated meditation.

The Buddhist notion of *ushin* as still being mired in the desires of the phenomenal world, and *mushin* as a release

from such delusion, stands in sharp contrast to the role of *kokoro* in poetry, which is composed in response to the beauties of the natural world, and thus itself is illusory and deluded. Hence, a conflict emerges between poetry and religion in which poetry is seen to represent as well as encourage attachment to the phenomenal world, and doctrinally is both deceived and deceptive. The medieval poet Saigyōillustrates this predicament in his poem in the *Shinkokinwakashū(The New Anthology of Japanese Poetry Ancient and Modern)* (IV, 362), when he mentions that though a priest, he 'lacks a heart' (*kokoro naki*); in other words he has renounced the world. Nevertheless, as a human and as a poet, he remains beguiled by the natural realm and hence is trapped. This dilemma is mediated by a new understanding of poetry itself as a religious path or *michi* in the form of *shikan*, disciplined contemplation, as *kadō* (the way of poetry). Paradoxically, delusion can be an instrument of awakening as in the Buddhist notion of *hōben*, or expediency, as illustrated by the use of parables.

During this period *kokoro ari* (having heart), apart from its Buddhist implications, came to mean also decorous, or worthy of the canonical stamp of approval, in contradistinction to *mushin*. Poetry that adhered to the conventions was labelled *ushin*, and aberrant poetry that deviated too much from the norms was termed *mushin* or substandard. These terms subsequently came to be applied not only to *tanka* (the classical form of poetry) but also to the newer form of linked verse known as *renga*, which was subdivided into two kinds, serious and comic (or careless).

The *Nō* playwright and critic Zeami (*c*.1363-1443), whose theatre was grounded in the beliefs and practices of Zen Buddhism, further expands the significance of *kokoro* when he remarks cryptically in his treatise known as the *Fūshi kaden (Teachings on Style and the Flower)*, 'The flower is the heart (*kokoro*), the seed is the performance', an inversion of Tsurayuki's original assertion some centuries before. Now the heart no longer functions as the seed of poetry, but becomes instead the culmination of *Nō*, aesthetically and spiritually, as emblemized by Zeami in the metaphor of the flower. Ironically, the performance, while seminal, is secondary ultimately to the epiphanic awareness blossoming in the hearts of those expressing their feelings and those affected by it.

Kokoro also proved central to perspectives outside Buddhism, as in the commentaries of the classics written by the *kokugakusha* (scholars of national learning) in the eighteenth century, who sought to elevate indigenous traditions in opposition to imported philosophies such as Confucianism and Buddhism. Motoori Norinaga, for example, rather than merely seeing *kokoro* as either a capacity of the perceiver or the place of moral conflict within an individual, stresses the idea of *kokoro* as the intrinsic or essential nature of objects themselves, in the form of *koto no kokoro* 'the essence of abstract things', such as the sadness of an experience, and *mono no kokoro*, 'the essence of natural things', such as the beauty of blossoms.

See also: Aesthetics; Aesthetics, Chinese; Aesthetics, Japanese; Buddhist philosophy, Japanese; Zeami

MEERA VISWANATHAN

References and further reading

Hisamatsu Sen'ichi (1970) *The Vocabulary of Japanese Literary Aesthetics*, Tokyo: Center for East Asian Cultural Studies.(Useful and comprehensive summary of various terms and their interrelations.)

Ki no Tsurayuki (*c*.905) 'Kanajo' (Kana Preface), in *Kokinwakashū*, ed. Umetomo Saeki, Tokyo: Iwanami Shoten, 1958; trans. L.R. Rodd and M. Henkenius, *Kokinshū*, Princeton, NJ: Princeton University Press, 1984. (First imperially commissioned anthology of Japanese poetry whose chief compiler was Tsurayuki; his well-known preface raises crucial concerns alluded to by later theorists.)

Suzuki, D.T. (1959) *Zen and Japanese Culture*, Princeton, NJ: Princeton University Press. (An imaginative and insightful explanation of how Zen permeates Japanese art by one of the great proponents of our century; the section on 'Zen and Swordsmanship' (89-117) is especially pertinent.)

Ueda Makoto (1967) *Literary and Art Theories in Japan*, Cleveland, OH: The Press of Western Reserve University.(A rich survey of the ideas of thirteen pre-modern theorists of art from such perspectives as literature, theatre, tea ceremony, painting and calligraphy.)

Zeami (1400-02) *Fūshi kaden (Teachings on Style and the Flower)* [alternatively known as *Kadensho*, (Writings on the Transmission of the Flower)], in *Nōgakuronshū(Collection of Treatises on the Art of Nō)*, ed. Minoru Nishio, Tokyo: Iwanami, 1961; trans. J.T. Rimer and Yamazaki Masakazu, *On the Art of the NōDrama*, Princeton, NJ: Princeton University Press, 1984.(Zeami's first treatise in which he discusses elements of theatre in terms of the underlying perspectives and values of classical poetry.)

Kotarbiński, Tadeusz (1886-1981)

Kotarbiński was one of the founders and main representatives of the Polish philosophical school known as the Lwów-Warsaw School and akin to, though independent of (and less radical than), the Vienna Circle; an anti-metaphysical, pro-scientific, rationalistic school of philosophy, which was very active and influential between the First and the Second World Wars.

Kotarbiński's programme for philosophy was a minimalistic and a practical one: he stressed the need to purify the field of philosophy of questions and concepts that lack factual content or logical coherence. According to him, the term 'philosophy' should be used, if at all, to denote only logic (understood as the philosophy of cognitive thought) and the philosophy of action, including moral philosophy. His numerous (more than 500) works are devoted to logic and philosophy of action in this broad sense. One of his main original ideas is the doctrine of reism or concretism, a special version of nominalism.

Kotarbiński was admired by several generations of his pupils for his unusual pedagogical gifts, his integrity and his moral courage.

1 Ontological reism

Tadeusz Kotarbiński studied philosophy and classics at the University of Lwów. From 1918 he was a professor of philosophy at the University of Warsaw. After the Second World War he was the president of the University of Lódż(and from 1951 to 1956 a chairman at the University of Warsaw. When he retired from the University he was elected president of the Polish Academy of Science and from 1960 to 1963 he was the chairman of the International Institute of Philosophy.

His main subject of inquiry and teaching was logic in a broad sense of the term, comprising formal logic as well as semiotics and methodology of science. He wished to reconstruct the conceptual apparatus of logic in this sense to make it an exact discipline which would fulfil the rigorous criteria of intelligibility and soundness. One of the results of this project was the reduction of all categories to the category of things, known as his doctrine of *reism* or *concretism*.

In its original formulation (Kotarbińsky 1929) the doctrine of reism was the ontological thesis, answering the question 'What is there?' According to this doctrine, all objects are things, that is, only concrete, physical objects (including sentient bodies) exist. There are no so-called abstract objects: properties, relations, states of affairs, and so on. Things can be white or black, but there is no such being as whiteness or blackness; two things can be similar to each other or - if they are sentient bodies - can be friends with each other, but there is no such being as similarity or friendship. The apparent existence of beings of these kinds is an illusion which has its origin in our use of the corresponding abstract terms in the same syntactic role as names of things. To avoid the illusion one has to realize that this usage is a metaphorical one: when translated into the literal mode of speaking, sentences containing abstract terms (apparent names or *onomatoids*) turn out to refer to concrete, physical individuals.

In its old ontological version the doctrine of reism can be (and was) objected to as one which cannot be consistently formulated. For to deny the existence of properties, relations and so on, one has to use apparent names as subjects in sentences (for example, 'Properties do not exist') which - according to reism itself - cannot be literally interpreted. On the other hand, the thesis 'Whatever exists is a thing' is nothing more than a decision concerning the choice of language: namely, the decision to use a language allowing only things as values of variables which can be bound by the existential operator (see Nominalism).

2 Semantic reism

For this reason Kotarbiński reinterpreted his reism as the semantic thesis according to which any sentence having cognitive value is (or can be translated into) a sentence referring only to things; in other words, apparent names of properties, relations and so on can always be eliminated from sentences without any loss of content.

To prove this thesis one should be able to formulate methods of transforming all kinds of sentences which are generally accepted as meaningful - especially all scientific theorems - into sentences which do not contain apparent names. Yet lack of such a method does not falsify the thesis: one can argue that if, for example, in physics there is

no method of translating sentences which contain the term 'field' (which does not seem to be the name of a thing) into sentences containing names of things only, this is because the concept of field is not sufficiently clear; physicists should seek its explication.

However, there is a discipline which does have a clear and exact conceptual apparatus while including abstract terms which seem to be irreducible: mathematics. The most troublesome case is the concept of class (in the distributive, set-theoretical sense). The simplest contexts containing this term can be replaced by contexts which do not contain it (for example, the sentence '*x* is a member of the class of dogs' by the sentence '*x* is a dog'); but there seems to be no method of eliminating the term 'class' from more complex contexts ('*x* is a member of the class of the classes of…' and so on).

Thus, as a semantic thesis, reism is not a theorem but a postulate or a suggestion. The reason for accepting this postulate is that speaking - if possible - only in terms of things is the best way of avoiding Bacon's *idola fori*: illusions and confusion in thinking caused by the traps of language. Kotarbiński had his predecessors (such as, Leibniz and Brentano), but his reism was more radical as concerns the mind-body problem in the light of his thesis that only physical objects exist (pansomatism).

3 Independent ethics

As a young student, Kotarbiński began to look for a replacement for the religious justification of morality, considering the suggestions of Stoicism, Epicureanism and utilitarianism. After many unsatisfactory trials he came to the conclusion that

> ethics…, just as medicine or administration, does not need a justification derived from an outlook on life. Its rules are invariable no matter whether a rational person is a materialist, an idealist or a spiritualist…. What, on the other hand is needed is to avoid using fantasy in justifications.
>
> (Kotarbiński 1952)

In the light of this he formed an ethical system which he called independent ethics

As the main parameters for the moral evaluation of actions he distinguished the following antitheses of what he called motives: kindness - cruelty; honesty - dishonesty; courage - cowardice; competence - indolence; composure - succumbing to temptation. Further, he described the moral ideal as a trustworthy guardian: a person on whom one can count, in a difficult situation, 'not an egoist but a kind person, a reliable, courageous, competent, and composed man'. In striving for this ideal one has to establish a hierarchy of persons and needs, and to complement the evidence of one's consciousness with suitable extra-ethical assumptions.

Kotarbiński's system of ethics is independent in a double sense: first, because it is free of any religious or philosophical assumptions, and second, because according to it, the conscience of a human being is the supreme judge of their actions, and cannot be replaced by that of anyone else.

4 Praxiology

As early as his first work Kotarbiński stressed the necessity of elaborating the conceptual apparatus for practical philosophy in a broad sense; first he called it general practice, then general methodology or praxiology. The programme of general methodology was outlined by Kotarbiński at the IX International Congress of Philosophy held in Paris, in 1937. He defined it as the investigation of 'rules of intentionality in action which are of value regardless of what material is being worked upon and to which a division of labour of arts or skills belongs'.

The programme was carried out in his main exposition of praxiology, *Traktat o dobrej robocie* (1955). The book contains numerous conceptual distinctions, definitions and rules which afford the basis for the utilitarian evaluation of actions according to their effectiveness and efficiency. Kotarbiński considered it to be a closed canon of praxiology.

Some ideas of praxiology can be found in writings of earlier thinkers (A. Bogdanov, G. Hostelet), but Kotarbiński's work in this field was both pioneering and systematic.

See also: Poland, philosophy in; Vienna Circle

B. STANOSZ

List of works

Kotarbiński, T. (1929) *Elementy teorii poznania, logiki formalnej i metodologii nauk*, Lwów: Ossolineum; Wrocław, Warsaw and Cracow: Ossolineum, 1961, 2nd edn; trans. *Gnosiology: The Scientific Approach to the Theory of Knowledge*, Oxford: Pergamon Press, 1966.(Extensive handbook of logic, semantics and theory of knowledge.)

Kotarbiński, T. (1952) 'Zagadminia etyki mizateznij' ('Problems of Independent Ethics'), *Kronika* (Lódż) No. 21.(Kotarbiński's statement of an ethics that is to be independent of religion, metaphysics and politics.)

Kotarbiński, T. (1955) *Traktat o dobrej robocie*, Lódż: Ossolineum; trans. O. Woitasiewicz, *Praxiology: An Introduction to the Sciences of Efficient Action*, Oxford: Pergamon Press, and Warsaw: PWN, 1965.(The fullest exposition of praxiology.)

Kotarbiński, T. (1957a) *Wykłady z dziejów logiki (Lectures on the History of Logic)*, Lódż: Ossolineum; French trans. *Leçons sur l'histoire de la logique*, Warsaw 1964.(A survey of the history of logic from a nominalist point of view.)

Kotarbiński, T. (1957b) *Myśli o działaniu (Reflections on Action)*, vol. I of *Wybór pism (Selected Writings)*, Warsaw: PWN.(Includes the most important essays and articles on different topics of practical philosophy, written between 1913 and 1954.)

Kotarbiński, T. (1958) *Myśli o myśleniu (Reflections on Thinking)*, vol. II of *Wybór pism (Selected Writings)*, Warsaw: PWN.(Includes the most important essays and articles on the theory of knowledge, 1913-54.)

Kotarbiński, T.(1987) *Pisma etyczne (Ethical Writings)*, Wrocław: Ossolineum.(A collection of Kotarbiński's ethical writings.)

References and further reading

Ajdukiewicz, K. (1935) 'Der Logistische Antiirrationalismus in Polen', *Erkenntnis* 5: 151-61.(Historically the first account of the contribution of the Polish School in philosophy - the 'Lwów-Warsaw School' - of which Kotarbiński was co-founder and one of the main representatives. Ajdukiewicz was another eminent representative of this school.)

Gasparski, W. (1993) 'A Philosophy of Practicality: A Treatise on the Philosophy of Tadeusz Kotarbiński', *Acta Philosophica Fennica*, 53 Helsinki. (A monograph devoted to Kotarbiński's philosophy seen as the philosophy of practicality in all its parts - in praxiology as well as in ethics, metaphilosophy, philosophy of language and philosophy of science.)

Gasparski, W. and Pszczołowski, T. (eds) (1983) *Praxiological Studies: Polish Contribution to the Science of Efficient Action*, Dordrecht.(An account of the Polish contribution to praxiology as conceived by Kotarbiński, written by his closest co-workers in this field.)

Kotarbińska, J. (1989) 'Droga Tadeusza Kotarbińskiego do prakseologii' (Tadeusz Kotarbiński's Route to Praxiology), *Praxiology* 4; trans. in *Praxiology: The International Annual of Practical Philosophy and Methodology*, vol. I, New Brunswick, NJ, 1992.(A short account of the evolution of Kotarbiński's views and interest in practical philosophy, given by his wife, who was his successor as the chairman of the Department of Logic at the University of Warsaw.)

Szaniawski, K. (1984) 'Philosophical Ideas of Tadeusz Kotarbiński', *Reports on Philosophy* 8: 25-32, Kraków: Uniwersytet Jagielloński.(A summary of Kotarbiński's philosophical views, written by one of his most eminent students and followers.)

Woleński, J. (1985) *Filozoficzna szkoła lwówsko-warszawska*, Warsaw: PWN; trans. *Logic and Philosophy in the Lvov-Warsaw School*, Dordrecht: Kluwer, 1989.(Extensive monograph on the contribution to logic and analytic philosophy of the Polish school between the First and Second World Wars.)

Koyré, Alexandre (1892-1964)

The scope of his research and his effort to give civilization meaning make Alexandre Koyré one of the boldest and most influential of twentieth-century historians of scientific thought. He was Russian of origin, German by philosophical training, French by adoption, and chose the USA as his second intellectual homeland. From the mysticism of the Renaissance to Romantic philosophy, from Copernican theory to Newtonian synthesis, he interpreted modern cosmology in the light of 'the unity of human thought' and of mathematical realism, as both effect and origin of a 'spiritual' revolution.

After the Russian Revolution of 1905 Koyré was arrested and is said to have read Husserl's *Logical Investigations* in prison. From 1908 to 1914 he studied in Paris with Bergson, Lalande and Brunschvicg and in Göttingen with Husserl, Hilbert and Reinach. His made his debut with an essay on Zeno's paradoxes, centred on the concept of movement, leitmotif of his later work as a historian of science. Realist phenomenology led him to study scholastic theology with Gilson and Picavet. In 1914 he enrolled as an officer in the Foreign Legion and was posted on the Russian front, taking part in the February Revolution and fighting in the October Revolution.

Having emigrated to Paris in 1919, he taught philosophy in Montpellier and from 1931 was in charge of the *direction d'études* of 'History of Religious Ideas in Modern Europe' at the École Pratique. In 1929, an important doctoral dissertation on Boehme and an essay on Russian Romantic philosophy showed his capacity to unite the 'profound intuitions' of metaphysics with cultural contexts like Renaissance spirituality or Russian nationalism by means of 'conceptual analysis'. He studied the terminology of Russian mysticism, of Spinoza and reinterpreted Hegel phenomenologically, in a way decisively influential for the Hegelianism of Kojève and Sartre. It was largely thanks to the journal *Recherches philosophiques* (1931-7), edited by Koyré, Puech and Spaier, that phenomenology and existentialism were promulgated in French philosophy between the wars. At the conference on phenomenology of the Société Thomiste at Juvisy in 1932, Koyré criticized Husserl's evolution towards the transcendental idealism of the *Cartesian Meditations*, of which he had reviewed the French edition in 1931. He believed in phenomenology, in its ontological meaning, as a method for studying essences, not as a metaphysic, as he observed with regard to Heidegger.

At the same time Koyré began a history of science with studies of Nicolas of Cusa, Paracelsus and Copernicus. For Paracelsus he adopted Lévy-Bruhl's notion of 'mentalité': ' what is most difficult and most necessary when we tackle the study of thought that is no longer ours, is to forget what we know' (1955b: 46). It was not philosophy of history that interested Koyré, but the uncertain adventures of Western philosophy before its issues had become stable: *itinerarium mentis in veritatem* (the mind's journey toward the truth). Scientific doctrines, like theological ones, belonged to 'structures of thought', which is what he called Dilthey's *Weltanschauungen* and Hoffding's 'concepts of the world'. Koyré was moved towards science by the epistemological circle of Meyerson, Lévy-Bruhl and Hélène Metzger: ' Have I remained faithful to Meyerson? Not completely, because in my studies I have striven to show not the identical basis of human thought, but the differences of its structures in different historical periods. I have remained faithful to his precept of studying the reasons for mistakes with the same attention accorded to successes' (1986: 139).

In his masterpiece *Études galiléennes* (1939), dedicated to Meyerson and inspired by Burtt (1925), Koyré studied the difficulties and failures with regard to the concept of movement as a state and to the principle of inertia, 'paradoxes' contrary to experience, conceivable only through ideal experiments on nature made mathematical. Making space geometrical and the universe infinite defined the 'scientific revolution' as a 'revenge of Plato upon Aristotle'.

This polemically antipositivist representation of modern science became known when his essay 'Galileo and Plato' (1943) was published by the History of Ideas Club, while Koyré was in New York to found the prestigious École Libre des Hautes Études with Maritain, directing the review *Renaissance* and lecturing on Plato at Columbia University. After the war, his project for a chair in 'History of Scientific Thought' at the College de France having failed, Koyré became a member of the Institute for Advanced Study in Princeton in 1955. He continued teaching in Paris with the history of religious ideas and of scientific thought at the Department of Economic and Social Sciences at the École Pratique.

Acceleration, heliocentricity, attraction and the Newton-Leibniz controversy are the themes of the ‘scientific revolution’ developed by Koyré. In the course of his research, this category turned out to be unstable and the length of the work problematic. In *From the Closed World to Infinite Universe* (1957), his main work of synthesis, he went as far as Laplace, though he intended to extend it to Maxwell and Einstein.

‘Scientific thought does not develop in a vacuum but always within a frame of ideas, of fundamental principles, of self-evident axioms which have usually been considered as belonging properly to philosophy’ ([1961b] 1971-2: 256). By opening science to the interaction of other collective representations and recreating the language of the past with the coherence of his terminology, Koyré put the history of science on a footing with anthropological history and established a historical methodology as a new way of analysing science. His influence on Kuhn and Foucault renewed epistemological controversy, though in ways often distant from and even opposed to his own work.

See also: Anthropology, philosophy of; Galilei, G.; Mechanics, classical; Newton, I.

PIETRO REDONDI

List of works

Koyré, A. (1922) *Essai sur l’idée de Dieu et les preuves de son existence chez Descartes (The Idea of God and the Proofs of His Being in Descartes)*, Paris: Leroux; repr. New York: Garland, 1987.(The foundations of Cartesian theology through the notions of infinity and perfection as discussed in the *Meditations*. Submitted for the certificate of the École pratique des hautes études, it earned him election of a lecturership at the Section of Religious Sciences at the École. See 1938.)

Koyré, A. (1923) *L’idée de Dieu dans la philosophie de Saint Anselme (The idea of God in Saint Anselm’s Philosophy)*, Paris: Leroux; repr. ivi, Vrin, 1984.(Prepared earlier under the direction of François Picavet, it satisfied the requirements for his degree at the Sorbonne.)

Koyré, A. (1929a) *La philosophie de Jacob Boehme (Jacob Boheme’s Philosophy)*, Paris: Vrin; 2nd edn, ivi, 1971; repr. New York: B. Franklin, 1968.(His doctoral thesis dedicated to Brunschvicg and Gilson, on the mystical cosmology of a contemporary of Galileo and Descartes who was later considered a source of the Hegelian metaphysics. See 1955a.)

Koyré, A. (1929b) *La philosophie et le problème national en Russie au début du XIX^e siècle (The Philosophy and the National Problem in Russia)*, Paris: Champion; 2nd edn, ivi, Gallimard, 1976.(A critical reflection on the cultural identity of his mother country through Russian nineteenth-century followers of Hegel and *Naturphilosophie*. See also his *Études sur l’histoire de la pensée philosophique en Russie*, Paris: Vrin, 1950.)

Koyré, A. (ed.) (1934) *Copernicus, Des révolutions des orbes célestes, livre ler*, Paris, Alcan; repr. Blanchard, 1970.(A translation and commentary of the cosmological part of Copernicus’ work with a historical and interpretative introduction. A seminal contribution to an intellectual history of science proper, this edition also marked Koyré’s transition to the metaphysical commitments of late Renaissance science too. See 1961.)

Koyré, A. (1938) *Trois leçons sur Descartes (Three Lectures on Descartes)*, Le Caire: Éd. de l’Universite du Caire; repr. in *Entretiens sur Descartes*, New York: Brentano, 1944.(Published in French and Arabic while a visiting professor in Egypt, these lectures on the *Discourse on Method* introduced his notion of ’revolution’ through the geometrical and infinite universe of Cartesian physics.)

Koyré, A. (1939) *Études galiléennes*, Paris: Hermann; trans. J. Mepham, *Galileo Studies*, Atlantic Highlands, NJ: Humanities Press, 1978; Hassocks: Harvester Press, 1978.(Three essays on the problem of the law of falling bodies and the principle of inertia in Galileo and Descartes’ physics. Arguing polemically, Koyré introduced Galileo’s experiments and the discovery of inertia as belonging to Descartes.)

Koyré, A. (1943) ’Galileo and Plato’, *Journal of the History of Ideas* 4: 400-28.(Galileo’s mathematical physics as a ’Platonic revenge’ against the traditional positivist view of the history of science. Republished in 1968.)

Koyré, A. (1945) *Discovering Plato*, New York: Brentano.(Presented at Columbia University and also published in a French edition, these lectures applauding democracy through a commentary of Plato’s *Republic* had been composed in Beirut after the fall of France in 1940.)

Koyré, A. (1948) ’Du monde de l’à-peu-près à l’univers de la precision’ (From the World of Approximation to the Universe of Precision), *Critique* 28: 806-23.(An important essay applying his notion of the intellectual revolution of classical physics to the problem of the origins of Western technology. Republished in 1961b.)

Koyré, A. (1955a) *Mystiques, spirituals, alchymistes du du XVI^e siècles allemand (Mystics, Spiritualists, Alchemists of Sixteenth-Century Germany)*, Paris: A. Colin; repr. Gallimard, 1971.(The republication of these earlier essays on Boehme’s most important Renaissance sources - Schwenkfeld, Franck, Paraceisus and Weigel

- was related to the author's collaboration with the journal Annales, especially with Febvre.)
Koyré, A. (1955b) 'A Documentary History of the Problem of Fall from Kepler to Newton', *Transactions of the American Philosophical Society* 45: 329-95.(An exemplary case study of the ramifications of a physical problem through an anthropological approach.)
Koyré, A. (1957) *From the Closed World to the Infinite Universe*, Baltimore, MD: Johns Hopkins University Press.(In this best-known work he traced the relation of God to the world from Nicolas of Cusa's cosmology to Newtonian natural theology.)
Koyré, A. (1961a) *La révolution astronomique*, Paris: Hermann; repr. ivi, 1974; *The Astronomical Revolution*, London, Methuen, 1980.(His last book on Kepler's astronomical metaphysics situated between Copernicus' geometrical cosmos and Borelli's celestial mechanics.)
Koyré, A. (1961b) *Études d'histoire de la pensée philosophique* (Studies in History of Philosophical Thought), Paris: A. Colin; repr. ivi, Gallimard, 1970.(Contains the three most important essays by Koyré on Hegel: 'Note sur la langue et la terminologie hégélienne' (Note on Hegelian Language and Terminology) 1931, 'Rapport sur l'état des études hégéliennes en France' (Report on the State of Hegelian Studies in France) 1931 and 'Hegel à Jena' (Hegel in Jena) 1934. Instead of the prevailing image of the young Hegel diffused in France by Jean Wahl, these essays dealt with the problems inherent to Hegelian language, philosophy of time and *Logik*.)
Koyré, A. (1965) *Newtonian Studies*, Cambridge, MA: Harvard University Press, London: Chapman and Hall.(A collection of essays on the synthesis of terrestrial and celestial mechanics by Newton and the theological differences between Newtonian and Cartesian doctrines of space).
Koyré, A. (1966) *Études d'histoire de la pensée scientifique (Studies in History of Scientific Thougth)*, Paris: PUF; repr. ivi, Gallimard, 1973; English trans. of part of the latter: *Methaphysics and Measurement: Essays on the Scientific Revolution*, Cambridge, MA: Harvard University Press, and London, Chapman and Hall, 1968.(An important collection of his papers concerning the methodology of history of science and critical studies from the Middle Age to Pascal.)
***** (1986) *De la mystique à la science. Cours, conférences et documents, 1922-1962 (From Mystical Theology to Science. Courses, lectures and documents 1922-62)*, Paris: Éditions de l'EHESS.(This collects Koyré's course abstracts, unpublished lectures on theology and science and documents on his teaching and biography.)
Koyré, A. and Cohen I.B. (eds) (1971-2) *Isaac Newton's Philosophiae Naturalis Principia Mathematica*, Cambridge, MA: Harvard University Press, 2 vols.

References and further reading

Burtt, E.A. (1925) *The Metaphysical Foundations of Modern Physical Science*, New York: Harcourt Brace, and London: Routledge & Kegan Paul.(A seminal book on the relationships between philosophical and religious ideas underlying the early modern science.)
Clagett, M. and Cohen, I.B. (1966) 'Alexandre Koyré', *Isis* 57: 157-66.(An obituary essay attesting the role played by Koyré in the growth of the history of science as an academic discipline in the USA.)
Gillispie, Ch.C. (ed.) (1973) 'Koyré', *Dictionary of Scientific Biography*, New York: Charles Scribner's Sons, vol. 7, 482-90.(A relevant biographical source.)
Hering, J. (1964-5) 'In memoriam', *Philosophy and Phenomenological Research* 25: 453-4.(A useful source on the spread of phenomenology in France through Koyré.)
Kuhn, T. (1970) 'A. Koyré and the History of Science', *Encounter* 34: 67-9.(The document of the intellectual encounter between the main historian of the 'révolution scientifique' and the author of *The Structure of Scientific Revolutions*.)
Redondi, P. (ed.) (1987) 'Science: The Renaissance of a History, Proceedings of the International Conference Alexandre Koyré, Paris 10-14 June 1986', *History and Technology* 4.(Among the contributors: G. Canguilhem, 'Preface', 7-10; K. Schuhmann, 'Koyré et les phénoménologues allemands', 149-68; Y. Elkana, 'Alexandre Koyré: Between the History of Ideas and Sociology of Disembodied Knowledge', 115-48; E. Couumet, 'Alexandre Koyré: la révolution scientifique introuvable?', 497-530.)
Salvadori, R. (ed.) (1980) '"Introduzione", J. Hyppolite, A. Kojève, A. Koyré, J. Wahl', *Interpretazioni hegeliane*, Florence: La Nuova Italia, VII-XVIII.(A presentation of the main texts of the Parisian debate on Hegel in the 1930s.)
Speilberg, H. (1960) *The Phenomenological Movement. A Historical Introduction*, The Hague: Nijhoff, 2 vols.(A classical historical presentation of Husserl and his disciples' work.)
Vinti, C. (ed.) (1994) *A. Koyré. L'avventura intellettuale*, Naples: ESI.(Among the contributors to this conference held at the University of Perugia in 1992; F. Barone, E. Berti, M. Biagioli, P. Galluzzi, G. Jorland, E. Mirri, P. Redondi, P. Rossi, P. Zambelli.)

Wahl, J. (1965) ‘Le rôle d’Alexandre Koyré dans le développement des études hégéliennes en France’, *Archives de philosophie* 23: 323-36.(The scope and impact of Koyré’s phenomenological reading of Hegel in the 1930s.)

Zambelli, P. (ed.) (1967) ‘"Introduzione", A. Koyré’, *Dal mondo del pressappoco all’universo della precisione*, Turin: Einaudi.(An overview on the sociological and cultural insight of Koyré’s history of scientific thought.)

Zambelli, P. (1995) ‘Alexandre Koyré versus Lucien Lévy-Bruhl. From Collective Representations to Paradigms of Scientific Thought’, *Science in Context* 13: 531-55. (A contextual analysis of the influence of Lévy-Bruhl’s notion of mentalité on Koyré’s historical method.)

Krause, Karl Christian Friedrich (1781-1832)

Krause sought an overall explanation of reality in the manner of the post-Kantian idealists; the key elements of his thoughts are the concepts of organism and harmony, involving the incorporation of opposing elements rather than their annihilation. Three main themes characterize his philosophy: the doctrine of science, together with the equivalence of knowledge and being, or rational realism; the religious doctrine of panentheism which proclaims, 'everything in God'; and the social doctrine of the League of Humanity.

1 Life

Krause was born in Eisenberg, in the region of Thuringia. He studied theology, mathematics and philosophy at the University of Jena, with Fichte and Schelling. There he habilitated as a *Privatdozent* in 1802, having written his dissertation, *Dissertatio philosophico-matematica de philosophiae et matheseos notione et earum intima coniunctione (Philosophico-Mathematical Dissertation on Philosophical and Mathematical Ideas and Their Close Relation)*. With increasing success, he taught logic, natural law, the philosophy of nature and his system of philosophy, as well as music. In 1803, the Napoleonic wars forced him to move twice: first to Rudoldstadt, then to Dresden. During this period he wrote *Grundlage des Naturrechts (Basis of Natural Law)* and continued working on *System der Philosophie (System of Philosophy)*. In 1805 he joined the Archimedes Lodge of the Brotherhood of Freemasons, and studied the history of Freemasonry. Five years later he published *Die drei ältesten Kunsturkunden der Freimauererbruderschaft (The Three Oldest Documents of the Brotherhood of Freemasons)*, in which he tried not only to prove that the masonic law of secrecy contradicted two of the Brotherhood's most ancient documents, but also called for its abolition. His expulsion from the Brotherhood of Freemasons was immediately decreed, and he would always attribute his subsequent penury to his Masonic enemies. In 1811 he published *Das Urbild der Menschheit (The Ideal of Humanity)*, in which he set out his doctrine concerning the social structure of humanity. In 1813 the war reached Dresden and he had to move to Tharand; six months later he went to Berlin, where he habilitated again with *Oration de scientia humana (Oration on Human Science)*. He set up the *Berlinische Gesellschaft für Deutsche Gesprache* (Berlin Society of the History and Grammar of the German Language), although he did not succeed in obtaining a post as a teacher. In 1817 he went on a study visit to Italy, France and Germany, and wrote *Aesthetik (Aesthetics)*. In 1825 he moved to Göttingen, where Heinrich Ahrens became his student. It was Ahrens who made his master's philosophy more widely known in Spain. In 1831, student riots forced Krause to move again, this time to Munich, where Schelling opposed his appointment as a professor. Krause died the following year.

2 The System

The first part of Krause's philosophical system is a process of 'analysis' which starts from a subjective perspective, both from knowledge and feeling, and which reveals the essential self (*Ur-Ich*) through intuition (*Anschauung*). This is an expression of individual substantiality, which is still not the thinking self - although this substantiality is one of its main activities - but the real self, unified, alive, different from any other entity. The doctrine of science deals with the cognitive self. However, knowledge does not just involve the logical knowledge of the intellect, but also comprises the areas of feeling and will as parts of the concept of Reason; moreover, the three of them are essential for each other and for their complete development.

Contrary to the Fichtean view, while Krause's 'self' is a foundation of the subjective realm, it is not enough to explain either knowledge of the other or that of the self in relation to the other. Also, the self is considered as a human being, made up of body and soul, finite essences which therefore belong to two universal ones: Nature and Spirit, both of which are contained in the concept of Humanity. And these three beings, infinite in themselves, are limited by each other, that is, in the Krausean language, they are 'relative infinites' and they require a 'Foundation' which comprehends and explains them: a basic and original essence, the absolute Being or metaphysical God (*Ur-Wesen*). From this point Krause moves on to the second part of the System, the Synthesis. He begins by studying the absolute, necessary Foundation and in reverse order reconstructs the whole process of knowledge, objectively and not only in the subjective aspect of analysis.

The 'Real Intuition' of the Absolute is the culmination of the process of analysis witnessed by the self: logical knowledge finds the absolute, necessary Foundation for its own intellectual knowledge. This 'intuition' -

originating from the Latin word '*tuitio*' - has two modes of access to the absolute Foundation through the two aforementioned elements of Reason: the path of feeling, which produces religion's Real Intuition of the Supreme Being, God; and the path of will to the absolute God, which is the foundation of goodness and morality. These three paths integrate the panentheist doctrine, or 'everything in God', which is opposed to pantheism, in which all idealisms come together.

With this discovery, two important problems which idealistic pantheism did not solve are now avoided: human freedom and determinism in history. God is the absolute principle of the world, life and thought: he creates what is finite together with Man. However, this does not constitute a foundational and causal relation in time, but only in a metaphysical sense. The act of creation happened once and for all: Man, who was created like God, has, like him, infinite freedom, which is only limited by the condition that he is a finite being, but not by his essential, vital capacity. Likewise, God is the lord of history from its beginning but he does not intervene either in its temporal evolution or in its end. Contrary to the Hegelian view of history, Krause does not conclude with a fusion of all beings in the Absolute, but in a 'League of Humanity' with God. This alliance is brought about by a free and autonomous decision by both parties.

This explains the importance of the process of analysis and synthesis of the fundamental science from which all individual sciences evolve: psychology, or the analysis of the self's faculties; physics, or the science of nature; and anthropology, or the science of Man. They are all contained in history, which is the expression of Man's action and God's revelation.

The most interesting discovery of Krause's philosophy from the perspective of nineteenth-century idealistic thought is the focus on Man as a living being, that is, Man and environment temporally determined as philosophy's authentic objective. The formal process of science's doctrine of analysis and synthesis, together with the discovery of the self, do not adequately explain the reality of Man considered as a living being, determined in time. There are two essential levels of reality: the permanent and the changeable. Time distinguishes these levels, while Man contains both, inseparably. With this, Krause overcomes the dichotomy between Man and History, ideas and facts. Nature and Spirit are entwined as a supreme synthesis, and this should be their role in the science of life: to establish unions and to propitiate them, since Man is the being of union *par excellence*.

3 Influence

This extremely optimistic thought, which emphasizes the essential goodness of human beings as well as of values such as freedom and equality, made a significant impression in emerging bourgeois societies with new constitutional governments, such as Spain and the Latin American countries. For them, Krausean pedagogy, in conjunction with the ideas of F.W.A. Froebel, was especially attractive and influential as an agent of social change beginning from the full education of the individual, and not from a purely scientific education. Also influential was the philosophy of law, which establishes principles for the union of human beings. It regulates mutual relationships among individuals and social groups, not only in order to achieve their freedom but also for each individual to attain moral perfection. Societies are determined by organic functions (scientific, artistic, moral and religious society) and also by personal ones (individual, family, nation, country).

These disciplines - pedagogy and the philosophy of law - together with the Theory of Science, were extensively elaborated and developed by Krause's immediate followers. They were seen as a part of a continuous process of development which individuals elaborate throughout life when they enter into relationship with all the realities around them, and are indissolubly linked to the concepts of ethics, freedom and autonomy in every social group and in every individual. They both enabled this idealism to connect with the multiplicity of scientific disciplines and research areas which positivism generated in the later nineteenth century. Thus, in an apparent paradox, the German spiritualist current of thought became united with French positivism, combining the main concerns of ethics with the strict determinism of science. In the same way, Krause's emphasis on the concept of harmony between Nature and Spirit confirms him and all his followers as pioneers of ecological thought in its main postulate of harmony and balance between science and the natural world.

See also: Absolute, the; German idealism; Spain, philosophy in §4

Translated from the Spanish by Isabel Venceslá
TERESA RODRÍGUEZ DE LECEA

List of works

Krause, K.C.F. (1802) *Dissertatio philosophico-matematica de philosophiae et matheseos notione et earum intima coniunctione (Philosophico-mathematical Dissertation on Philosophical and Mathematical Ideas and Their Close Relation),* Jena: Voigtium.(Krause’s Ph.D. thesis, which puts forward the philosophy and mathematics he uses to construct a system which explains reality.)

Krause, K.C.F. (1803) *Grundlage des Naturrechts (Basis of Natural Law),* Jena and Leipzig: Christian Ernst Gabler.(Krause’s first attempt to formulate a world state as a social organization upon which a federation of nations could be established.)

Krause, K.C.F. (1810) *Die drei ältesten Kunsturkunden der Freimauererbruderschaft (The Three Oldest Documents of the Brotherhood of Freemasons)*, Dresden: Arnoldischen Buchhandlung.(Krause’s attack on the Freemasons, in which he called for their abolition.)

Krause, K.C.F. (1811a) *Das Urbild der Menschheit (The Ideal of Humanity)*, Dresden: Arnoldischen Buchhandlung.(Krause’s most translated work, which contains his concept of society determined by organic functions - religious, scientific, the state and art - and by personal functions, such as family, friendship and nationality.)

Krause, K.C.F. (1811b) *Versuch einer wissenschaftlichen Begründung der Sittenlehre (Attempt at a Scientific Foundation of Ethics)*, Leipzig: Reclam.(This work concerns practical philosophy, considered as an essential component in the construction of a philosophical system.)

Krause, K.C.F. (1825) *Abriss des Systemes der Philosophie (Outline of his System of Philosophy)*, Göttingen: Dieterischen Buchhandlung.(Sets out the development of the analysis and synthesis process in Krause’s philosophical system.)

Krause, K.C.F. (1828) *Abriss des Systemes der Rechtes, oder der Naturrechtes (Outline of the System of Right, or of Natural Law)*, Göttingen: Dieterischen Buchhandlung.(The philosophy of law is based on the individual’s freedom and participation in society through associations, one of them being the estate, which is equal to the rest.)

Krause, K.C.F. (1829) *Vorlesungen über die Grundwahrheiten der Wissenschaft (Lectures on the Basic Principles of Science)*, Göttingen: Dieterischen Buchhandlung.(Sets out the Krausean theory of knowledge.)

Krause, K.C.F. (1837) *Abriss der Aesthetik oder Philosophie des Schönen und der schönen Kunst (Outline of Aesthetic and the Philosophy of Beauty and Fine Arts)*, Göttingen: Dieterischen Buchhandlung.(Contains the definition of beauty as the perfect union between Nature and Spirit. This idea is central to Krause’s pedagogic project.)

Krause, K.C.F. (1848) *Vorlesungen über die psychische Anthropologie (Lectures on Psychical Anthropology),* Göttingen: Dieterischen Buchhandlung.(Presents the doctrine of Man and life, based on the Krausean concept of the union of both the material and the spiritual part of being, as well as the need of a harmonious development for being.)

Krause, K.C.F. (1900) *Der Menschheitbund (On the League of Humanity)*, Berlin: Felber.(Apart from developing the theory of a world state, this work offers a federal account of the government of nations.)

References and further reading

Abrahamson, M.R. (1991) ‘Krausism, Pepita Jiménez and the Divinization of Life’, *Letras Peninsulares* 4 (1): 225-43.(Emphasizes the importance of Krause’s philosophy of life in the Spanish novel at the end of the nineteenth century.)

Davies, H.E. (1984) ‘Alejandro Deustúa: His critique of the Aesthetic of the Vasconcelos’, *International Philosophy Quarterly* 24: 69-78.(A discussion of the aesthetic theory of the Krausean, Deustúa.)

Dendle, B.J. (1991) ‘Armando Palacio Valdés, the Revista Europea and the Krausist movement’, *Letras Peninsulares* 4 (1): 25-34.(The writer Armando Palacio Valdés sarcastically criticizes Krausean thinkers, especially Sanz de Río and Francisco de Paula Canalejas, from a philosophical as well as from a literary viewpoint.)

Diaz, E. (1989) *La filosofía social del krausismo español (The Social Philosophy of Spanish Krausism)*, Madrid: Debate, 3rd edn. (A comparative study of three Spanish Krausean thinkers: Sanz del Río, Giner and Azcárate.)

Garcia Mateo, R. (1982) *Das deutsche Denken und das modernen Spanien (German thought and modern Spain),* Frankfurt: Lang.(The importance of the Krausean social doctrine, which was promulgated in Paris and Brussels by Krause’s follower H. Ahrens.)

Kodalle, K. (ed.) (1985) *K.Ch.F. Krause (1781-1832) Studien zu seiner Philosophie uns zum Krausismo (Studies of Krause and Krausism)*, Hamburg: Meiner.(Contains various studies on Krause by specialists on Krausism

and on idealist philosophy in general: the theory of knowledge, philosophy of religion and practical philosophy.)

Lipp, S. (1985) *Francisco Giner: A Spanish Socrates*, Waterloo, Ont.: Laurier University Press.(An account of the Krausean thinker F. Giner de los Ríos' pedagogical theory and its effects on modernization in contemporary Spain.)

Lopez Morillas, J. (1981) *The Krausist Movement and Ideological Change in Spain: 1854-1874*, trans. J. López Morillas, Cambridge: Cambridge University Press.(Offers an excellent description of the various areas of influence of the Krausean movement in Spain.)

Ríos Urruti, F. de los (1916) *La Filosofía del Derecho en don Francisco Giner y su relación con el pensamiento contemporáneo (Francisco Giner's Philosophy of Laws and its Relation to Contemporary Thought)*, Madrid: Corona.(An important comparative study of Krause's philosophy of law and his contribution to this discipline.)

Rodriguez de Lecea, T. (ed.) (1983) *Reivindicacion de Krause*, Madrid: Fundación F. Ebert.(Contains various contributions about Krause and Spanish Krausism during the second half of the nineteenth century, such as the theory of knowledge, philosophy of religion, philosophy of law and social philosophy.)

Rodriguez de Lecea, T. (1987) *El Krausismo y su influencia en América Latina (Krausism and its Influence in Latin America)*, Madrid: Fundación F. Ebert.(Various contributions on the influence of Krausean books and doctrines in countries such as Argentina, Brazil, Cuba, Mexico and Puerto Rico.)

Rodriguez de Lecea, T. (1991) *Antropología y filosofía de la historia en Julián Sanz del Río (Anthropology and Philosophy of History in Julián Sanz del Río)*, Madrid: Centro de Estudios Constitucionales.(A study of the Krausean process of analysis considered as a theory of man and history.)

Ureña, E.M. (1991) *Krause, educador de la humanidad (Krause: Educator of Humanity)*, Madrid: Unión Ed.(A detailed intellectual biography of Krause.)

Willm, J. (1849) *Histoire de la philosophie allemande depuis Kant jusqu'à Hegel (History of German Philosophy from Kant to Hegel)*, Paris: Lagrange.(An important study on German philosophers contemporary to Krause, and his situation among them. Despite its age, it is still useful, though difficult to find.)

Kripke, Saul Aaron (1940-)

Saul Kripke is one of the most important and influential philosophers of the late twentieth century. He is also one of the leading mathematical logicians, having done seminal work in areas including modal logic, intuitionistic logic and set theory. Although much of his work in logic has philosophical significance, it will not be discussed here.

Kripke's main contributions fall in the areas of metaphysics, philosophy of language, epistemology, philosophy of mind and philosophy of logic and mathematics. He is particularly well known for his views on and discussions of the following topics: the concepts of necessity, identity and 'possible worlds'; 'essentialism' - the idea that things have significant essential properties; the question of what determines the referent of an ordinary proper name and the related question of whether such names have meanings; the relations among the concepts of necessity, analyticity, and the a priori; the concept of belief and its problems; the concept of truth and its problems; and scepticism, the idea of following a rule, and Ludwig Wittgenstein's 'private language argument'. This entry will be confined to the topics of identity, proper names, necessity and essentialism.

1 Life

Saul Kripke was born in Bay Shore, New York, the son of a rabbi and a teacher. The family soon moved to Omaha, Nebraska, where Kripke spent the rest of his childhood. At a very early age, he began to display prodigious mathematical ability and intense curiosity about philosophical questions. At the age of 15 he developed a semantics for quantified modal logic and his proof of the completeness of the semantics appeared in *The Journal of Symbolic Logic* when he was 18.

Kripke received his bachelor's degree in mathematics from Harvard in 1962 and was appointed to the Harvard Society of Fellows in 1963. There followed appointments as Lecturer at Princeton (1965 and 1966) and Harvard (1966-8). He was appointed Associate Professor at The Rockefeller University in 1968 (then Professor in 1972), and McCosh Professor of Philosophy at Princeton in 1977, which is the position he currently holds. Kripke was the youngest person ever chosen to give the John Locke Lectures at Oxford (1973) and he has been the recipient of many prestigious honours and accolades.

2 Identity, proper names and possible worlds

Kripke began working on these topics in the early 1960s and developed a number of positions that initially met with surprise and suspicion, but eventually became widely accepted. The most comprehensive presentation may be found in *Naming and Necessity* (1980), which is an edited transcript of three lectures given at Princeton in January 1970, together with a preface and addenda.

Kripke maintains that identity is a relation that holds between each thing and itself, never holds between any two things, and always holds (or fails to hold) of necessity (see Identity). The 'necessity of identity' is a theorem of quantified modal logic (with identity), but some philosophers nevertheless thought certain examples cast doubt upon the doctrine (see Modal logic). For example, the identity of the inventor of bifocals with the first Postmaster General can easily seem contingent. But Kripke observed that what is contingent here is that one person attained both of these (separately contingent) distinctions, and that this contingency in no way undermines the necessity of that person's identity with himself - Benjamin Franklin would have been identical with himself even if he had never been Postmaster General or worked with lenses.

This last point reflects a fundamental Kripkean doctrine about the behaviour of ordinary proper names such as 'Franklin'. It is that when we utter a sentence such as 'Franklin might never have became Postmaster General', we are describing a way in which things might have gone differently for the very person who, in actual fact, was the first Postmaster General. We are describing a 'counterfactual situation' or 'possible world' involving Franklin himself, but in which he is not Postmaster General.

Kripke does not offer a detailed ontological account of possible worlds (though he does explicitly reject the 'counterpart' account favoured by David Lewis (1968)). He does not see worlds as offering a useful analysis of our modal notions, but rather as reflecting pre-existing modal intuitions, and thinks that for most philosophical

purposes we need no precise notion of possible worlds. It is normally sufficient to think in an everyday way of counterfactual situations or alternative possibilities, and this can help relieve any worry that we are trading on a technical notion of possible worlds without first defining it. Now invoking the terminology of worlds, the idea of the previous paragraph generalizes to the claim that an ordinary proper name always designates the same thing 'in other possible worlds' that it designates in the actual world. Kripke introduced the term 'rigid designator' for singular terms that behave in this way, allowing his thesis to be compactly expressed as the claim that proper names are rigid designators. But this rather technical-sounding formulation is intended as little more than a generalization of the hardly controversial idea that 'Franklin invented bifocals' and 'Franklin might not have invented bifocals' both concern the same person.

Now consider the sentence 'Cicero is Tully', which seemingly asserts a true identity and which involves two proper names. If these names are rigid designators, then each designates the same entity in every possible world - the entity that each designates in the actual world. But in the actual world both names designate the same person. It follows that they both designate this person in every possible world, and so it must be a necessary truth that Cicero is Tully. But it has often been thought that because such a fact is not knowable a priori, but rather requires empirical investigation, it could not be necessary. Kripke sees this as a fundamental error, stemming from a confusion of the epistemic notion of a prioricity with the metaphysical notion of necessity. For Kripke, 'Cicero is Tully' expresses an a posteriori, necessary truth (see A posteriori; A priori).

The idea that 'Cicero is Tully' is merely contingent is also fostered by theories of proper names in the tradition of Frege and Russell, according to which the referent of a name is determined by a definite description (or, in more recent versions, a 'cluster' of descriptions) associated with the name. Such theories often hold that the description gives the meaning (or sense) of the name. Because the description associated with 'Cicero' might be utterly different from the one associated with 'Tully', and because these descriptions might have distinct referents in some other world, it would be natural to think that 'Cicero is Tully' is contingent in much the same way that 'The first Postmaster General is the inventor of bifocals' is contingent. But Kripke argued that such theories of proper names are incorrect. His view is more in the spirit of J.S. Mill's and Ruth Barcan Marcus', according to which names are mere 'tags' which lack senses and serve only to pick out their referents. Kripke held that a name is explicitly linked to its referent when it is first introduced, and that subsequent uses of the name refer to that entity because they trace backwards through an appropriate causal chain of uses to that initial introduction of the name. Subsequent uses do not achieve their reference to that entity as a result of its satisfying any specific descriptions associated with the name.

3 Necessity and essentialism

Essentialism is the view that things have certain of their significant properties necessarily - that they could not have existed without possessing those properties (see Essentialism). In the early 1960s the very notion of necessity was often thought to be incoherent. But even philosophers who granted the meaningfulness of the notion and conceded the existence of certain necessary truths typically found essentialism unacceptable.

Much of the antipathy towards necessity and essentialism may be traced to a confusion of necessity with analyticity (see Analyticity). Just as Kripke urged that a prioricity and necessity are distinct notions, so did he insist that analyticity is a third notion, not to be confused with the others. But such a confusion appears to underlie one of the main arguments against essentialism. It was claimed that the idea that a thing has a property necessarily makes no sense absolutely, but only relative to a specific way of designating the thing. For example, if the number nine is designated by 'the number of the planets', then it has the property of numbering the planets necessarily, but if it is designated by 'nine', it does not. So the idea of attributing a necessary property to the object independently of any designation makes no sense (see Quine, W.V.).

This may seem correct if we have not carefully distinguished necessity from analyticity. For 'The number of the planets numbers the planets' seems analytic while 'Nine numbers the planets' does not. But when we set aside analyticity and focus clearly on necessity, the question no longer concerns sentences of our language. Instead we are asking whether a certain object had to have a certain property. If the object is the number nine and the property is numbering the planets, the answer must be 'no', for we can imagine a world in which the solar system developed differently. And if the property is being a square, the answer must be yes. There is no world in which nine is not a square. (Of course it could be that the name 'nine' did not refer to nine. But, in such a world, nine - whatever it

might be called there - would still be a square.)

Kripke argued that ordinary entities also have nontrivial essential properties. Suppose a specific table, T, is made of wood. According to Kripke, although it is possible for tables to be made of other materials, no such possibility exists for T. Of course it is an empirical matter whether T has this property at all. But this merely means the matter is a posteriori, not that it is contingent. We know that such properties as T's location or finish are contingent because we can imagine a world in which T is somewhere else or differently finished. According to Kripke, we cannot imagine a world in which T is not made of wood. When we think we have imagined such a world, closer consideration reveals it is really a world in which some other table has been switched with T (or made instead of T, and so on).

Kripke relies heavily on 'intuition' here and in his other philosophical work. But he is not invoking a special power or sense. Rather, he is insisting that the concepts that concern us in philosophy are, after all, our own concepts, and accordingly that the best evidence for how they fit together can only come from thinking as carefully as we can about them. In this entry we have managed a glimpse at some of the results of this approach in Kripke's work on reference and modality. Comparably powerful and compelling results are to be found throughout the full range of his remarkable work.

See also: Semantic paradoxes and theories of truth; Semantics

MICHAEL JUBIEN

List of works

Kripke, S.A. (1959) 'A Completeness Theorem in Modal Logic', *Journal of Symbolic Logic* 24 (1): 1-14.(Proves that a formula is a theorem of quantified modal logic if and only if it is valid in Kripke's semantics. Very technical.)

Kripke, S.A. (1963) 'Semantical Considerations on Modal Logic', *Acta Philosophica Fennica* 16: 83-94; 53-355. (Presents Kripke's semantics for modal logic. Partly informal, partly technical.)

Kripke, S.A. (1971) 'Identity and Necessity', in M.K. Munitz (ed.) *Identity and Individuation*, New York: New York University Press, 135-64.(An early presentation of key ideas treated in Kripke 1980.)

Kripke, S.A. (1972) 'Naming and Necessity', in D. Davidson (ed.) *Semantics of Natural Language*, Dordrecht: Reidel, 2nd edn, 253-355.(Treats the topics of Kripke 1980.)

Kripke, S.A. (1975) 'Outline of a Theory of Truth', *Journal of Philosophy* 72 (19): 690-716.(Sketches a theory of truth designed to avoid paradox while minimizing truth-value gaps. Very technical.)

Kripke, S.A. (1976) 'Is There a Problem about Substitutional Quantification?', in G. Evans and J. McDowell (eds) *Truth and Meaning*, London: Oxford University Press, 325-419.(Discussions of the substitutional interpretation of quantification and evaluation of its philosophical significance. Difficult, often technical.)

Kripke, S.A. (1977) 'Speaker's Reference and Semantic Reference', *Midwest Studies in Philosophy* 2: 255-76.(On the potential divergence of what a speaker intends to refer to from what the speaker's words refer to, and whether the words are ambiguous. Largely nontechnical.)

Kripke, S.A. (1979) 'A Puzzle About Belief', in A. Margalit (ed.) *Meaning and Use*, Dordrecht: Reidel, 239-83. (Discusses a vital problem concerning the interplay of reference and belief. Difficult, not very technical.)

Kripke, S.A. (1980) *Naming and Necessity*, Oxford: Blackwell, and Cambridge, MA: Harvard University Press. (The key source for Kripke's views on proper names, reference, identity, necessity, essentialism and related topics. Difficult but largely nontechnical.)

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell, and Cambridge, MA: Harvard University Press.(Discusses scepticism, rule-following and Wittgenstein's 'private language argument'. Nontechnical.)

Kripke, S.A. (1986) 'A Problem in the Theory of Reference: The Linguistic Division of Labor and the Social Character of Naming', in *Philosophy and Culture (Proceedings of the 17th World Congress of Philosophy)*, Montreal: Editions du Beffroi, Editions Montmorency, 241-7.(Discusses a problem about how proper names refer.)

References and further reading

Branch, T. (1977) 'New Frontiers in American Philosophy', *The New York Times Magazine*, August 14.(Evidence of Kripke's popular fame, with discussion of his early work.)

French, P.A., Uehling, T.E., Jr. and Wettstein, H.K. (eds) (1986) *Studies in Essentialism, Midwest Studies in*

Philosophy 11.(A valuable collection of twenty-six essays on essentialism and related topics by noted authors. Includes an abundance of useful references. The essays vary in techicality and difficulty.)

Jubien, M. (1993) *Ontology, Modality, and the Fallacy of Reference*, Cambridge: Cambridge University Press. (Offers a theory of proper names and an account of essentialism that differ sharply from Kripke's, despite a fundamental agreement at the intuitive level. Difficult but not overly technical.)

Lewis, D.K. (1968) 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* 65: 113-26.(A rather technical presentation of an account of modality criticized in Kripke 1980. Requires an understanding of quantified modal logic.)

Lewis, D.K. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.(A philosophical elaboration of the key ideas of Lewis 1968, and related topics. Difficult but not overly technical.)

Marcus, R.B. (1993) *Modalities*, New York: Oxford University Press.(Collects Marcus's essays on modality, modal logic, and other topics. Of special interest are essays 1, 14, and appendix 1A, a discussion among the author, Quine, Kripke, T. McCarthy, and D. Follesdal, which followed a 1962 presentation of essay 1. Essay 1 is quite technical, essay 14 and appendix 1A are less so.)

Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Oxford University Press.(Contains an influential account of possible worlds and a general theory of modality. Difficult and fairly technical.)

Quine, W.V.O. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Offers criticisms of modal notions that are responded to in Kripke 1980. Difficult and sometimes technical.)

Kristeva, Julia (1941-)

Born in Bulgaria, Kristeva entered the Parisian scene of avant-garde intellectuals in the 1960s. Her earliest work in linguistics was shaped by the post-Stalinist communism of eastern Europe, a political climate that exerts its influence on her entire corpus, even as she distanced herself from it, to embrace an increasingly psychoanalytic perspective. Dissatisfied with scientific models of language, conceived as a mere means of communicating preconceived ideas, where words simply function as isolated symbols that represent discrete concepts, Kristeva analyses language as a signifying process. As such, language is not a static and closed system of signs, but a mobile, fluid process that implicates bodily and vocal rhythms in the generation of symbolic meanings. In La Révolution du language poétique (1974) (Revolution in Poetic Language, 1984) Kristeva fuses linguistic insights with psychoanalytic inquiry as she presents two distinct yet interrelated aspects of the signifying process, the semiotic and the symbolic. The semiotic aspect of language is vocal, pre-verbal, rhythmic, kinetic and bodily. The symbolic aspect of language is social, cultural, and rule-governed. Focusing on the interplay between the semiotic and the symbolic, Kristeva is able to analyse literary and historical texts, works of art and cultural phenomena in a way that thematizes the complex relationship between materiality and representation.

1 Life and works

Although Kristeva has come to be known to a large portion of her English readership as a French feminist, the label is misleading. Bulgarian by birth, Kristeva moved to Paris in 1966, and her engagement with feminism has been nothing if not polemical. After an early training in the sciences, she worked as a journalist while pursuing her literary studies. In France, having received a doctoral fellowship, she followed the advice of Tzvetan Todorov, and attended the seminar of Lucien Goldmann. She embarked on a career that is distinguished by its breadth and interdisciplinary approach, the main strands of which - in addition to philosophy - can be identified as linguistics, literature, and psychoanalysis. A research assistant at Claude Lévi-Strauss's Laboratory of Social Anthropology, she also took advantage of the resources offered by the Centre National de Recherche Scientifique, and the École Pratique des Haute Études. As an associate of the *Tel Quel* review, edited by Philippe Sollers, Kristeva quickly took her place at the centre of French intellectual life (see Tel Quel School). She joined the faculty of the University of Paris VII, and became a practising psychoanalyst.

Among the early influences on her work, in addition to Marxism, Mikhail Bakhtin stands out, as does her engagement with structuralism. Semiotics, or the science of signs, provided the focus of her first published works, *Semeiotiké: Recherches pour une sémanalyse* (1969) and *Le Texte du roman* (1970). Adapting and building upon the work of Roland Barthes, Ferdinand de Saussure and Charles S. Peirce, Kristeva began to develop a theoretical framework that would be more fully elaborated in the landmark text *La Révolution du langage poétique* (1974) (*Revolution in Poetic Language*, 1984).

By 1980, with the publication of *Pouvoirs de l'horreur: essai sur l'abjection* (*Powers of Horror: An Essay on Abjection*, 1982) Kristeva's growing interest in psychoanalysis had become decisive. The impact of Jacques Lacan's *Écrits* (1971), and the psychoanalytic training that Kristeva had undergone continued to inform her approach in *Histoires d'amour* (1983) (*Tales of Love*, 1983), *Soleil noir* (1987) (*Black Sun*, 1989), *Étrangers à nous-même* (1988) (*Strangers to Ourselves*, 1991) and *Les Nouvelles Maladies de l'âme* (1993) (*New Maladies of the Soul*, 1995). Her work also includes novels.

2 Linguistics

Language, for Kristeva, has transformative capacities. In order to unearth the heterogeneous character of language, considered as in process, Kristeva departs from the 'generative grammar' of Chomsky that gained popularity in the 1970s (see Chomsky, N.). The Chomskian view that 'surface' structures derive from 'deep' structures appeared to reduce the 'speaking subject' to a series of translinguistic generalities that privilege systematic structures. In its search for truth, its emphasis on logic, and its adoption of scientific procedures Saussurean linguistics fared no better than Chomsky's innatist belief in linguistic universals. Although Kristeva would rehabilitate de Saussure's interest in semiology, she found his implementation of it inadequate to the subject of enunciation, or to what she called the 'speaking subject'. Kristeva's desire to attend to the exigencies of the speaking subject is not a matter of returning to a Cartesian self, or a transcendental ego, but rather takes account of the disruptive and disturbing

qualities that invade the equilibrium of linguistic frameworks that reduce language to a series of rules, or contain it within a formal system of signs.

Kristeva calls for linguistics to change its object of study. It is no longer the theoretical rules governing language, whether these are conceived as grammatical, or semiological, that should be studied, but rather - and it is here that the influence of Roman Jakobson (see Russian literary Formalism) makes itself felt - ‘poetic language’. Far from harnessing or fixing language by establishing its foundational structures, or stabilizing it within a system, Kristeva focuses on a ‘speech practice’ that involves a dialectic between its ‘signified structure (sign, syntax, signification)’, and a ‘semiotic rhythm’. Although Kristeva uses the term ‘dialectic’ to describe the struggle that takes place between ‘language and its rhythm’, she depends on post-Hegelian and post-Marxist resonances to give meaning to the term. It is Heidegger’s vision of the strife of world and earth as the origin of the work of art, and the work of Mallarmé and Artaud, Lautréamont and Bataille, as well as that of the Russian poets Mayakovsky and Khlebnikov, that Kristeva draws on when she observes, in ‘The Ethics of Linguistics’ (in *Desire in Language*, 1977), that the poet ‘wants to make language perceive what it doesn’t want to say’.

Wary of neutralizing the specificity of the speaking subject, Kristeva wants to find a way to recapture the ‘rhythm of the body’, to reclaim the ‘semiotic materials’ that other linguistic models tend to obliterate in their fascination with the technical elements of language.

3 *Revolution in Poetic Language*

By refusing to restrict the significance of language to its meaning - to see in it only a system of representation - Kristeva insists on the materiality of language, on its emergent conditions. In *Revolution in Poetic Language*, she thematizes the divergent modalities of language under the heading ‘the semiotic and the symbolic’. Kristeva’s analysis of the semiotic is informed by Freud’s notion of instinctual drives or impulses, the unconscious, and the pre-Oedipal. The symbolic is associated with the Freudo-Lacanian notion of post-Oedipal relations, with the function of representation, and with language as a sign-system.

The semiotic drives articulate what Kristeva calls the *chora*, a term she inherits from Plato’s account of the creation of the universe in his dialogue the *Timaeus*. The *chora* is a maternal receptacle, a generative matrix, an eternal place that, Plato tells us, is neither visible nor partakes of form, but is in some way intelligible. Amorphous and formless, Plato calls it the nurse of becoming, a kind of wet-nurse. Kristeva retains the paradoxical quality she sees in Plato’s account of the *chora* as formless and undetermined, yet capable of receiving form and determination. The *chora* is neither sign nor signifier, neither model nor copy. It is pre-symbolic, not yet posited, and yet it can be named and spoken of - a process that converts the semiotic into the symbolic, conferring on the semiotic precisely the order, constraint, or law of culture that it resists. Just as Plato’s *chora* resists definition, although it can be spoken of as if it were identifiable as an entity, so Kristeva constitutes the semiotic by naming it, even as its mobile forces elude conceptualization. In both cases the very utterance involves a loss, a betrayal of what language attempts to say. Yet this is a necessary betrayal, since the semiotic relies upon the symbolic for its articulation - one might say for its very existence - even if it suffers a transformation in the process of coming to representation.

If the semiotic needs the symbolic to represent it, the register of the symbolic requires the irruption and influx of the semiotic if it is to remain capable of change. If, in the interests of the very integrity of the semiotic, the subject cannot repudiate the symbolic, neither can it do without the semiotic. The need that the symbolic has for the excesses of poetic motion, or for the otherness of musical rhythm that marks the semiotic rupturing of the symbolic might be described as an ethical and political exigency.

4 Kristeva and feminism

A visit to China which resulted in the book *Des Chinoises* (1974) (*About Chinese Women*, 1977) set the tone for Kristeva’s uneasy relationship with feminism. The book caused controversy, and Kristeva was accused of being ‘romantic’ and ‘utopian’ in her figuring of the Orient as other. Given to dramatic statements, such as ‘a woman cannot "be"’, Kristeva’s often dismissive remarks about other versions of feminism - what she calls in an interview ‘sociological protest’ (Marks and de Courtivron 1981) - did little to curry favour with her detractors. In her defence, it should be pointed out that Kristeva’s work has shared the same fate as that of Luce Irigaray and Hélène

Cixous, with whom she is so often compared. The reception of their work has been plagued by a failure to take account of their intellectual heritage, political contexts and personal biographies. The misshapen versions of their work that critics have handed down to their English-speaking readers made it easy for readers of the 1980s to accuse these three so-called French feminists (Irigaray is Belgian, and Cixous is Algerian) of 'essentialism'. Afraid that Kristeva's caution about totalizing politics, her attempt to reintroduce the body into analyses of language, and the Lacanian influence on her work amounted to a male-identified conservative and reactionary elitism, feminist critics shunned her work. Instead of understanding her insistence upon sexual difference as a refusal to neutralize the specific material conditions of the production of language, conditions which include the sexed body, her attempt to reintroduce the body into feminism was hailed as detrimental to the progress of women's quest for equality.

In the 1990s, Kristeva began to receive a fairer hearing in English-speaking feminist circles, and this was in large part due to the fact that critics began to equip themselves with the conceptual tools that Kristeva's diverse intellectual endeavours required. Not least among these is the importance of phenomenology, psychoanalysis and post-structuralism - specifically Ferdinand de Saussure's structural linguistics and its post-structuralist modifications by figures such as Roland Barthes.

5 Psychoanalysis

When Kristeva denies being to a woman, her assertion takes account of Heidegger's ontological distinction - the difference between a being and Being (or existence) as such (see Heidegger, M.) - and Lacan's notorious statement in *Seminar* XX (1972-3) that 'There is no such thing as *The* woman.' Understood in this light, rather than as a merely polemical statement, it can be read as a way of insisting upon the relevance of sexual difference, the limitations of ontological language, and the need to interrogate the possibilities of situating a woman's subjectivity on a level that does not simply conform to the position of male subjectivity. The position adopted by the female subject is thus one in excess of the boundaries of language, outside the order of being - at least in so far as language and existence are sanctioned as male.

It was Kristeva's abiding interest in language that drew her to psychoanalysis. Interested in the underside of language, Kristeva's work of the late 1980s and 1990s is preoccupied with the exploration of borderlines, marginal existence, outlawed subjectivities, and what it means to be a foreigner, an outsider. While the significance of these themes is by no means limited to sexual difference - the question of nationality, the issue of psychosis, and the desire for abjection are a few of the concerns that Kristeva pursues by focusing upon the limits and borders of subjectivity and consciousness - the question of feminine and masculine identity is a central one.

The symbolic, a Lacanian term that Kristeva invests with a new significance in counterposing it to the semiotic, interprets the father of Freud's Oedipal drama in terms of language. Since the semiotic *chora* has maternal connotations, there is a sense in which the distinction between the semiotic and the symbolic is sexually marked. This does not mean that Kristeva discerns semiotic energies only in the work of women artists. On the contrary - and this has provoked consternation to some feminists - she points more often to the works of Bellini and Giotto, Baudelaire and Proust for their capacity to evoke the sexual pleasure that she calls corporeal '*jouissance*'. This term, like so much of her work, reflects the influence of Lacan, whose impact on Kristeva's corpus is decisive, despite the critical distance she inserts between her own writing and Lacan's *oeuvre*.

TINA CHANTER

List of works

Kristeva, J. (1969) *Semeiotiké: Recherches pour une semanalyse*, Paris: Éditions du Seuil.(Early work on semiotics.)

Kristeva, J. (1970) *Le Texte du roman*, The Hague: Mouton.(Semiotic theory and linguistic literary analysis.)

Kristeva, J. (1974a) *La Révolution du language poétique*, Paris: Éditions du Seuil; trans. M. Waller, *Revolution in Poetic Language*, introduction L.S. Roudiez, New York: Columbia University Press, 1984.(Fusing insights from the diverse sources of linguistics, Bahktin, Hegel, Lacan and others, this central text firmly established Kristeva's reputation as an original theorist, mobilizing the distinction between the semiotic chora as a motile rhythmic sensible aspect of language and the symbolic realm of meaning that has become the hallmark of her work.)

Kristeva, J. (1974b) *Des Chinoises*, Paris: Éditions des Femmes; trans. A. Barrows, *About Chinese Women*,

London: Marion Boyars, 1977.
Kristeva, J. (1977) *Polylogue*, Paris: Éditions du Seuil.(Of the twenty essays comprising this collection, eight are translated in *Desire in Language: A Semiotic Approach to Literature and Art*, ed. L.S. Roudiez, trans. T. Gora, A. Jardine and L.S. Roudiez, 1980.)
Kristeva, J. (1980) *Pouvoirs de l'horreur: essai sur l'abjection*, Paris: Éditions du Seuil; trans. L.S. Roudiez, *Powers of Horror. An Essay on Abjection*, New York: Columbia University Press, 1982.(Psychoanalytic work exploring abjection, perversion and defilement in the context of Céline's novels, the Oedipal myth and biblical semiotics.)
Kristeva, J. (1983) *Histoires d'amour*, Paris: Denoël; trans. L.S. Roudiez, *Tales of Love*, New York: Columbia University Press, 1987.(Freudian inspired theoretical and historical theory of love drawing on religious and literary conceptions of agapē and eros, from Plato to Bataille.)
Kristeva, J. (1987) *Soleil noir: dépression et mélancolie*, Paris: Gallimard; trans. L.S. Roudiez, *Black Sun: Depression and Melancholia*, New York: Columbia University Press, 1989.(Analysis of melancholia and depression as different ways of mourning the lost maternal object.)
Kristeva, J. (1988) *Étrangers à nous-mêmes*, Paris: Fayard; trans. L.S. Roudiez, *Strangers to Ourselves*, New York: Columbia University Press, 1991.(Cultural, political and psychoanalytic analysis of the phenomenon of foreignness, nationality and individuality.)
Kristeva, J. (1993) *Les Nouvelles maladies de l'âme*, Paris: Fayard; trans. R. Guberman, *New Maladies of the Soul*, New York: Columbia University Press, 1995.(A meditation on the function of psychoanalysis in relation to the soul, approached as the 'psyche' by the ancients.)

References and further reading

Baruch, E. Hoffamn and Serrano, S.J. (ed.) (1988) *Women Analyze Women: In France, England, and the United States*, New York: Columbia University Press.(Contains an interview with Kristeva.)
Fletcher, J. and Benjamin, A. (ed.) (1990) *Abjection, Melancholia and Love: The Work of Julia Kristeva*, London and New York: Routledge.(A useful collection, focusing on the literary and feminist aspects of Kristeva's corpus.)
Lacan, J. (1971) *Écrits*, Paris: Éditions du Seuil; trans. A. Sheridan, *Écrits: A Selection*, New York: W.W. Norton. (Lacan's major collection of essays containing seminal pieces, such as the essay on the 'mirror stage'.)
, J. (1972-3) 'God *and the* Jouissance*of The* Woman', in *Seminar XX, Encore*, Paris: Éditions du Seuil; trans. J. Rose, in J. Mitchell and J. Rose (eds) *Feminine Sexuality: Jacques Lacan and the école freudienne*, London: Macmillan, 1982.(The translation is in a collection of essays focusing on the question of woman, with two very helpful introductory essays by the editors.)
Lechte, J. (1990) *Julia Kristeva*, London and New York: Routledge.(A general introduction, helpful for situating Kristeva's semiotics in its linguistic context.)
Marks, E. and de Courtivron, I. (eds) (1981) *New French Feminisms: An Anthology*, Brighton, Sussex: Harvester Press.(The first collection of essays to introduce Kristeva to an English speaking feminist audience. Includes an extract from an interview, 'Woman Can Never be Defined', first published in 1974 by *Tel Quel*, conducted by the radical feminist group known as 'Psych & Po'.)
Oliver, K. (ed.) (1993) *Ethics, Politics, and Difference in Julia Kristeva's Writing*, New York: Routledge.(To date the best collection of essays, representing the diverse elements of Kristeva's work, including an article by Lisa Lowe concerning Kristeva's work on Chinese women.)
Oliver, K. (1993) *Reading Kristeva: Unraveling the Double-bind*, Bloomington and Indianapolis, IN: Indiana University Press.(Particularly useful for readers interested in psychoanalysis and feminism.)
Plato (*c*.366-360 BC) *Timaeus*, in *Timaeus, Critias, Cleitophon, Menexenus, Epistles*, trans. R.G. Bury, Loeb Classical Library, vol. IX, London: Heinemann, 1975.(Plato's *Timaeus* is a classic ancient text that presents a myth of creation and concerns the themes of space, place and time. Kristeva borrows the term which plays an organizing role in the crucial distinction between the semiotic and the symbolic from Plato's discussion of place: chora, translated as matrix or receptacle, and associated with maternal giving.)

Krochmal, Nachman (1785-1840)

Nachman Krochmal was leader of the Jewish Enlightenment or Haskalah in Galicia, eastern Europe. An astute observer of the German philosophical environment, Krochmal provided one of the first Jewish responses to, and adaptations of, elements of the philosophical work of Spinoza, Kant, Herder, Schelling and Hegel. His posthumously published Moreh Nevukhei ha-Zeman (Guide to the Perplexed of Today) (1851) adapted Kantian epistemological methods to interpret the Jewish religious sources with an eye to discovering their inner, philosophical meaning. Krochmal argued that at that their deeper level these sources, both ancient and medieval, anticipated the discoveries of the German Idealist philosophers. Thus, for example, the Jewish belief in a personal God who created the world and revealed a desired way of life could remain philosophically fruitful. For, when properly interpreted, such ideas were concrete representations of the truths laid bare in the metaphysics of Hegel and Schelling. Krochmal answered the regnant philosophy of history, which considered Jewish culture to have been sublated, by arguing that Jewish religion, by its apprehension of the absolute, stood outside the historical `laws' that mandate the eventual cultural demise of all nations and states.

Although he was an important model for aspiring Jewish intellectuals in eastern Europe, Krochmal's work was of limited philosophical influence. His lasting contribution may be his implicit exposure of the unstated cultural biases of modern idealist philosophy. For his work shows that alternative cultural assumptions would turn idealist philosophy towards philosophical and religious conclusions significantly different from those of the German Idealists themselves.

1 Life and times

Nachman Krochmal was born in the city of Brody in Galicia, eastern Europe, then under Austrian rule and lived in Brody and the Galician cities of Zollkiew, Lemberg (Lwów) and Tarnopol. His birth came in the middle of the decade-long reign of the Austrian Emperor Joseph II, whose ideals committed him to modernizing the educational institutions of the Jews and instilling in them German language skills. Krochmal himself remained largely untouched by the Emperor's educational programme, but the penetration of German language and culture profoundly challenged his Galician-Jewish environment. The movement to bring `enlightenment' to the Jews, known in Hebrew as the Haskalah (see Enlightenment, Jewish), was strongly centred in Galicia and deeply affected Krochmal, who emerged there as its leader. But its programme of cultural modernization was of extremely limited reach and was not accompanied, in his lifetime, by significant political or social change. The small coterie of modernist Jewish intellectuals in Galicia experienced `modernity' mainly in the form of cultural challenges that came, in part, from an ongoing philosophical discussion in a distant land, whose intellectuals they did not know and could not engage with in any immediate way. The medieval Jewish philosophical tradition was of limited help in facing the new challenges, for the medieval Jewish philosophers typically addressed the challenge of Aristotelian thought. In Krochmal's view, Jewish intellectuals needed a new guide, to replace Maimonides' *Guide to the Perplexed* (see Maimonides, M.), if they were to find continued philosophical vitality in the Jewish tradition.

2 *Moreh Nevukhei ha-Zeman*

Roughly one-third of Krochmal's *Moreh Nevukhei ha-Zeman* addresses philosophical themes. He begins his philosophical discussion by noting the incoherence of both philosophical scepticism and religious fideism. Staking out a middle ground between these two extremes required an epistemological foundation that would explain how and why religious thinkers formulate their ideas. Drawing on Kant - as he would have to do to be convincing - Krochmal notes that the path to clear philosophical ideas moves from sensory stimuli to preliminary concepts or representations (*Vorstellungen*), and on to concepts of the understanding (*Begriffe*), and the purely intellectual ideas of reason (*Ideen*). Most people never rise beyond the level of the *Vorstellung*, or representational thought. But any successful religious community must address a broad base. So religious communication is naturally formulated representationally. Diverging from Kant, Krochmal insists that religious representations differ from the more abstract *Ideen* descriptively and linguistically, but are no less apprehensions of truth. Religions always communicate representationally, but they focus on speculative concerns and convey philosophical truths, albeit less clearly than abstract philosophical discourse. Krochmal, then, sees religions as suffused with esoteric

speculation. The task of the interpreter is to uncover the hidden speculative content. Accordingly, traditional Jewish sources and Western philosophical thinking are drawn into a lengthy philosophical dialogue in Krochmal's work.

The Jewish sources traditionally saw a purpose in created existence. Their teleological conception sustained their affirmation of the existence of God. Modern philosophy formidably challenged this view. Spinoza for one dismissed such teleology as illusory (see Spinoza, B. de §4). Kant reaffirmed the logic of teleological judgments, drawing on the interdependent and organic character of nature; but his reaffirmation allowed us to use teleology only as a regulative, not as a constitutive, idea. Hegel developed Kant's argument into a metaphysical claim, affirming the organic nature of existence. Krochmal, too, adapted Kant's argument and, like Hegel, moved the organic characterization of nature beyond the realm of regulative ideas and into metaphysics. He went on to assert that teleology presupposes a willing and thus a personal God as creator. The traditional Jewish sources, then, with their affirmation of divine teleology, communicated, by way of representations, the speculative truth of organic nature and the reality of its purposing creator.

Reliance on a teleological proof that led to a creator God raised a second of the philosophical challenges that faced modern Jews. Hegel, in his `Lectures on the Philosophy of Religion', had identified creation - *creatio ex nihilo* - as the source of the fundamental defect of Jewish religion, its sublimity. In Hegel's view the vision of a world created from nothingness by a transcendent being symbolized the absolute and unbridgeable gulf that the Jewish religion opened up between the infinite and the finite. Christianity, the consummate religion, by rejecting the transcendent otherness of God and relying on the symbolism of the incarnation, overcame that gap, sublated the finite/infinite antinomy, and so reached a more advanced stage of religion and a closer approximation to the truth (see Hegel, G.W.F. §7).

Turning to the work of the medieval Bible commentator and thinker Abraham Ibn Ezra among others, Krochmal argued that Hegel's was not an accurate portrayal of Jewish theology. Creatively drawing on Ibn Ezra and other Jewish Neoplatonists, he reconstructed a vision of a protracted divine drama, in which God begins the process of coming to self-consciousness through a purposeful act of self-limitation. The world is created not from nothingness but from God's very being. Thus it is suffused with divinity or spirit (*Geist*); the gulf between God and the world is illusory. Traditional Jewish images and representations anticipate the fundamental insights of German Idealism. The modern idealist philosophers contribute conceptual vocabulary and conceptual clarity; but it was the ancient Jewish texts that first communicated, however imperfectly, the fundamental insights into the relationship of the finite and infinite.

The awareness that the world and all that it contains, including human intellect, is suffused with *Geist* meant for Krochmal that Judaism was the first, and as yet the only, religion to recognize that nothing exists apart from God. God represents the ultimate unity of all existence, and the God of traditional Judaism corresponds to the absolute spirit of Hegelian philosophy. All other religions and cultures have only partial appreciations of the spirituality represented in the idea of God. Their fragmentary visions of spirit may elevate aesthetics or politics as a bearer of spirit, but only Jewish culture is based on a recognition of the absolute nature of spirit.

This claim of Krochmal's is no mere religious triumphalism. It would, if accepted, contribute to Judaism's continued viability in its confrontation with modern philosophy. But its implications affect the philosophy of history profoundly. Despite the substantial differences between Herder and Hegel (and some would insist that Vico is relevant here as well), each views history as a process in which cultures - represented concretely in the *Volk* or the state - emerge, mature and decay. Whether this process is organic or dialectical, the cultures perish, passing on to the next world-historical culture some central legacy. Modern Europe is the culmination (or at least the latest culmination) of this process. For Hegel, this is because the modern European state stands closer than any of its predecessors to the realization of the absolute in its historical self-consciousness (see History, philosophy of).

Krochmal accepted the idea that history is a process moving towards the realization of the absolute within human consciousness and culture. But his reading of the Jewish sources led him to insist that Judaism has already achieved self-conscious awareness of the absolute. Thus the bearers of the Jewish religion stand clear of the central `law' of the regnant philosophy of history. Unlike other ancient and medieval cultures, the Jews do not wither and perish. To be sure, Jews have an external history like any other people; but their absolute God is not limited by place or historical circumstance. Thus, when Jews are vanquished militarily, their culture and religion retain their

self-evidence; the Jews undergo cultural rebirth and begin a new cycle of growth and maturity. The laws of history proposed by Herder and Hegel apply to other cultures but misrepresent the place of the Jews in history by radically misunderstanding Jewish metaphysics and its historical repercussions.

Krochmal's original contributions to philosophy are limited. Almost all of the central philosophical ideas of *Moreh Nevukhei ha-Zeman* originated elsewhere. And its philosophical sections, written in a difficult Hebrew idiom, were beyond the reach of most of his intended audience. But his implicit exposure of the unstated cultural biases of modern idealist philosophy are an enduring achievement. Readers are quickly brought to the realization that idealist philosophy, as Hegel himself understood, represents its own age, apprehended in thought. A comparison of Krochmal with Schelling or Hegel, among others, brings the Christian, Lutheran roots of idealist thought into sharp relief. Krochmal's reversal of many of the cultural and historical judgments familiar in idealist philosophy, and his elicitation of the central metaphysical claims of idealism from Jewish sources show that a philosopher using different assumptions could develop and apply idealist metaphysical concepts to reach philosophical and religious conclusions significantly different from those of the German Idealists themselves. Krochmal's efforts to reconcile Judaism and idealism provided new insight into Judaism and demonstrated new possibilities within idealism.

See also: Cultural identity; Enlightenment, Jewish; Hegel, G.W.F.; History, philosophy of

JAY M. HARRIS

List of works

Krochmal, N. (1851) *Moreh Nevukhei ha-Zeman (Guide to the Perplexed of Today)*, ed. with Hebrew introduction by S. Rawidowicz, Berlin: Einot, 1924; repr. Waltham, MA: Ararat, 1961.(Adapts Kantian epistemological methods to interpret the Jewish religious sources with an eye to discovering their inner, philosophical meaning. Krochmal argues that at that their deeper level these sources, both ancient and medieval, anticipated the discoveries of the German Idealist philosophers.)

References and further reading

Guttmann, J. (1964) *Philosophies of Judaism*, New York: Schocken, 365-90.(A learned treatment of Krochmal's thought in a classic one-volume history of Jewish philosophy.)

Harris, J.M. (1991) *Nachman Krochmal: Guiding the Perplexed of the Modern Age*, New York: New York University Press.(The only book-length treatment of Krochmal's work; it treats Krochmal's thought as a response to idealism.)

Hegel, G.W.F. (1832-45) `Lectures on the Philosophy of Religion', in *Werke. Vollständige Ausgabe durch einen Verein von Freunden des Verewigten*, vols 11-12, Berlin: Duncker & Humblot; trans. C.P. Hodgson and R.F. Brown, *Lectures on the Philosophy of Religion*, Los Angeles, CA: University of California Press, 1984-7. (Translation of manuscripts based on student notes of Hegel's lecture courses on the philosophy of religion.)

Rotenstreich, N. (1984) *Jews and German Philosophy: The Polemics of Emancipation*, New York: Schocken, 138-59.(A discussion of Krochmal's philosophy of history and its philosophical difficulties.)

Kronecker, Leopold (1823-91)

Leopold Kronecker was one of the most influential German mathematicians of the late nineteenth century. He exercised a strong sociopolitical influence on the development of mathematics as an academic institution. From a philosophical point of view, his main significance lies in his anticipation of a new and rigorous epistemological perspective with regard to the foundations of mathematics: Kronecker became the father of intuitionism or constructivism, which stands in strict opposition to the methods of classical mathematics and their canonization by set theory.

One of the most influential mathematicians in Germany of his time, Leopold Kronecker exercised a strong sociopolitical influence on the development of mathematics as an academic institution. He was a member of the Academy of Science in Berlin and editor of *Crelle's Journal* (the *Journal für die reine und angewandte Mathematik*), and tried to control the future development of pure mathematics as a supreme science by his superb and original contributions to arithmetic, algebra and the theory of functions.

Kronecker became the father of intuitionism or constructivism, anticipating a new and rigorous epistemological perspective with regard to the foundations of mathematics. In his 'Neubegründung der Mathematik' ('New Foundations of Mathematics', 1922), Hilbert compares the views of the two outstanding intuitionists of his own time, Brouwer and Weyl, with those of Kronecker and remarks:

> I believe that as little as Kronecker succeeded in abolishing irrational numbers in his day, just as little will Weyl and Brouwer succeed today. No, Brouwer is not, as Weyl suggests, the revolution, but only the repetition of an attempted Putsch by old means, which in former times, were applied much more resolutely, but wholly without success.
>
> (1922: 160)

Why is Hilbert, the head of the mathematical establishment in the twentieth century, so hostile - and at the same time so respectful - towards Kronecker? One must remember that in 1886 - at a time when Kronecker's influence was still dominant - Hilbert had given a 'solution' to Gordan's problem (to prove the existence of a finite basis for the infinite system of invariant forms) which had perplexed the mathematical world. Instead of constructing a finite basis for the infinite system of invariant forms, Hilbert proved its existence by proving that the assumption of its nonexistence leads to a contradiction.

Such an indirect existence proof was against the basic methodological principles of mathematics that Kronecker had tried to establish. Hence, in Hilbert's opinion, Kronecker's sin was his attempt to restrict mathematical reasoning by setting up arbitrary prohibitions regarding the admissible means of developing and proving certain propositions. This, in turn, leads to an intolerable restriction of mathematical knowledge. But is Hilbert's assessment of Kronecker fair and accurate?

There are very few places in Kronecker's published work where he expresses his philosophical attitude with respect to foundational issues. But in his late essay 'Über den Zahlbegriff' ('On the Concept of Number'), 1887 he is quite explicit. Elaborating a remark by Gauss that just as mathematics is the queen of sciences, so, too, arithmetic is the queen of mathematics, he writes:

> Indeed, arithmetic stands in a relation to the other two mathematical disciplines, geometry and mechanics, that is similar to that in which mathematics as a whole stands to astronomy and the other sciences of nature….Here, however, the word 'arithmetic' should not be taken in its usual restricted sense, but in a broader sense according to which all mathematical disciplines, with the exception of geometry and mechanics, have to be included - in particular algebra and analysis. And I also believe that we will succeed some day in 'arithmetizing' the entire content of all these mathematical disciplines, so that they will be based purely and simply on the conception of number taken in its narrowest sense. Hence, all modifications and extensions of this concept [Footnote: 'I mean, in particular, the inclusion of the irrational as well as the continuous quantities.'], which in most cases were prompted by its application to geometry and mechanics, will once again be stripped off.
>
> (1887: 252-3)

What Kronecker envisages here is in essence nothing other than basing the whole of mathematics, including analysis, on the concept of positive integers without any further existential assumption; in particular without assuming the existence of irrational numbers and continuous quantities. Consequently, he tries to eliminate all numbers 'alien to proper arithmetic' by giving definitions in terms of positive integers and functions thereof. For example, negative numbers can be eliminated by replacing the factor -1 with the indeterminate *x* and replacing the equality sign with the Gaussian symbol for congruence modulo $(x+1)$. The equation $7-9=3-5$ then becomes transformed into the congruence $7+9x=3+5x(\mathrm{mod.}\ x+1)$. Although no one took this particular proposal any more seriously than other proposals to eliminate the rational and irrational numbers, not only Weyl and Brouwer but also Hilbert shared the basic epistemological convictions behind Kronecker's proposals (at least after 1920, when Hilbert developed his finitism). These convictions can be summarized in two statements.

First, the infinite exists nowhere in actuality; neither in nature nor in our thinking. It is a mere potentiality. Therefore it has to be eliminated as an actual object from mathematics and be treated as a mere instrument (Kronecker) or an idea (Hilbert) to facilitate mathematical reasoning. Second, objects or functions, such as irrational numbers or continuous quantities, which presuppose infinite sequences of objects or operations have to be replaced by their proper meaning in arithmetic (Kronecker) or be treated as 'ideal elements' (Hilbert).

See also: Intuitionism; Mathematics, foundations of; Logical and mathematical terms, glossary of; Proof theory §2

ULRICH MAJER

List of works

Kronecker, L. (1895-1930) *Werke*, ed. K. Hensel, Leipzig: Teubner, 5 vols.(The most complete edition of Kronecker's work.)

Kronecker, L. (1887) 'Über den Zahlbegriff' (On the Concept of Number), *Journal für die reine und angewandte Mathematik* 101: 337-55; repr. in *Werke*, ed. K. Hensel, Leipzig: Teubner, 1895-1930, vol. 3, part 1, 249-74. (Perhaps Kronecker's most important contribution to the foundations of arithmetic.)

References and further reading

Biermann, K.-R. (1981) 'Kronecker, Leopold', in *Dictionary of Scientific Biography*, vol. 7, ed. C.C. Gillispie, New York: Charles Scribner's Sons, 1984.(A good, readable biography.)

Edwards, H. (1987) 'An Appreciation of Kronecker', in *The Mathematical Intelligencer*, New York: Springer, vol. 9, 28-35. (A sympathetic reading of Kronecker's work.)

Edwards, H. (1989) 'Kronecker's View on the Foundations of Mathematics', in D. Rowe and J. McCleary (eds) *The History of Modern Mathematics*, vol. 1, *Ideas and Reception*, San Diego, CA: Academic Press.(A proposal to make Kronecker's algebraic point of view intelligible.)

Hilbert, D. (1922) 'Neubegründung der Mathematik' ('New Foundations of Mathematics'), *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 1: 157-77; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer, 1935.(The first paper in which Hilbert formulated his programme to prove the consistency of arithmetic by finite means.)

Hilbert, D. (1926) 'Über das Unendliche', *Mathematische Annalen* 95: 161-90; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer; trans. 'On the Infinite', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics*, Cambridge: Cambridge University Press, 2nd edn, 1983.(The most important paper in defence of finitism.)

Kneser, A. (1925) 'Leopold Kronecker', *Jahresbericht der Deutsche Mathematiker-Vereinigung* 33: 210-28.(An excellent biography.)

Majer, U. (1993) 'Das Unendliche - eine bloße Idee?', *Revue internationale de Philosophie* 47: 319-41. (A critical analysis of Hilbert's finitist point of view.)

Weyl, H. (1921) 'Über die neue Grundlagenkrise der Mathematik', *Mathematische Zeitschrift* 10: 39-79; repr. in *Gesammelte Abhandlungen*, vol. 2, ed. K. Chandrasekharan, Berlin: Springer, 1968.(The work which triggered Hilbert's response.)

Weyl, H. (1925) *Die heutige Erkenntnislage in der Mathematik*, Erlangen: Weltkreis; repr. in *Gesammelte Abhandlungen*, vol. 2, ed. K. Chandrasekharan, Berlin: Springer, 1968.(The best defence of intuitionism/constructivism in Kronecker's sense.)

Kropotkin, Pëtr Alekseevich (1842-1921)

A founder of anarchist communism and guiding spirit of the international anarchist movement after the death of Bakunin, Kropotkin was also a distinguished geographer, a scientist and a positivist. He saw the development of anarchism as one aspect of the whole movement of modern science towards an integrated philosophy. He believed that the dominant phenomenon in nature was harmony, arrived at by a continuous process of adjustment between contending forces. In human, as in animal societies, the dominant phenomenon was mutual aid: thus once metaphysics, law and state authority had been shaken off, harmony could be realized.

Born into an aristocratic Moscow family close to the Russian Imperial throne, Pëtr Kropotkin was educated at an exclusive military academy, but at 20, filled with the desire to be useful, he renounced a brilliant career to serve for five years as a military administrator in Eastern Siberia.

His hopes for liberal reform by Alexander II, the tsar who had abolished serfdom, were soon disappointed. He also lost any faith in the virtues of state discipline in society and began to move slowly towards an anarchist position (see Anarchism). He now turned to scientific exploration, leading geographical expeditions and achieving distinction as a geographer. His observations laid the foundations of his theory of mutual aid among animal species.

Abandoning military service in 1867, he joined the Russian revolutionary movement in the early 1870s, and after a sensational escape from prison to the West in 1876 became a leading figure in the International, participating in the congress debates between the Marxist and anarchist wings of the movement, which led to the theory of anarchist communism.

After the death of Bakunin in 1876 he became, through his writings and propagandist journalism in *Le Révolté* and *La Révolte*, the guiding spirit of the movement. At the same time, during his years in the West, in England, France and Switzerland, he came to enjoy high esteem as a geographer, publishing in learned journals. He returned to Russia in 1917.

Kropotkin’s anarchist communism aimed at a society without government, where harmony would be obtained

> not by submission to law, or by obedience to any authority, but by free agreements between the various groups, territorial and professional, instituted for the sake of production and consumption as also for the satisfaction of the infinite variety of needs and aspirations of a civilized society.
>
> (1927: 284)

In such a society, as in organic life, Kropotkin believed harmony would result from ’an ever-changing adjustment and readjustment of equilibrium between a multitude of forces and influences’.

The individual would not be limited in the free expression of his powers in production by a capitalist monopoly, or by obedience, which only led to the sapping of initiative. On the contrary, he would be able to obtain the complete development of all his faculties: the fullest individuation.

Like socialists, anarchist communists combated the monopolization of land and capital. But they also combated the state, which, Kropotkin argued, as the instrument for establishing monopolies in favour of ruling minorities, was the main support of the capitalist system. Anarchist communists finally believed, said Kropotkin, that the era of revolutions was not closed. Periods of revolution must be taken advantage of to reduce the powers of the state.

Kropotkin’s anarchist communism was also inspired by his scientific studies: indeed he saw the development of anarchist social theories and the growth of nineteenth-century philosophies deriving from scientific study as simply two branches of the same movement. Anarchism itself was for him a world concept based on a mechanical explanation of all phenomena and embracing the whole of nature, including human societies. Its aim was actually ‘to construct a synthetic philosophy comprehending in one generalization all the phenomena of nature’ (1927: 150). Kropotkin shared the optimism of the positivists of his time with regard to the limitless possibilities of the inductive-deductive method of scientific inquiry. He went perhaps further than Proudhon, or even Elisée Reclus, in rejecting as unscientific all metaphysics and the justification it offered for the power of Church and state, whether

that power emanated from the Christian belief in an all-powerful God or from the Hegelian concept of the Universal Spirit. He went so far, in 1913, as to attack Bergson savagely for denigrating science by arguing that intuition played an important part in scientific discovery.

Kropotkin's own scientific work had convinced him that, in human as in animal societies, the instinct of mutual aid aimed at the preservation of the species was stronger than the instinct for the self-preservation of the individual, the notion of a struggle for life between competing individuals being based on an incorrect interpretation of the work of Darwin. Instead, the fundamental struggle for each species was that of the species as a whole against external forces.

Within human societies therefore, once the monopolization of capital and rule by minorities through their instrument, the state, had been shaken off, harmony could be achieved by a continuous series of adjustments among contending forces.

The central concept in Kropotkin's thinking is that of harmony in nature, a harmony which he himself had observed both in animal species and throughout human history, and which he believed would become possible in the new anarchist communist society. Such harmony is not the result of anything preconceived, but a

> temporary adjustment established among all forces acting upon a given spot - a provisional adaptation. And that adjustment will only last on the condition that it is continually modified; that it represents at every moment the resultant of all conflicting actions. If one force is hampered in its actions, harmony will disappear, force will accumulate its effect and a new equilibrium will emerge.

(1927: 121)

Kropotkin drew an analogy here between the intermittent disruption of harmony in nature by earthquakes and that brought about in societies by revolutions.

He finally believed that the fundamental phenomenon of mutual aid in human societies provided a foundation for a new scientific morality independent of religion. In his *Ėtika (Ethics)* (1922), which remained unfinished, he surveyed the historical development of ethics, pointing to the fact that Darwin, after Bacon, had provided the basis for a new morality. However, that new morality had not yet been worked out. Kropotkin did not live long enough to develop a non-religious morality himself.

CAROLINE CAHM

List of works

Kropotkin P.A. (1892) *La Conquête du pain (The Conquest of Bread)*, Paris: Stock.(Includes a chapter on anarchist communism.)

Kropotkin P.A. (1899) *Fields, Factories and Workshops*, London: Hutchinson.(Foreshadows economic tendencies in industry and agriculture.)

Kropotkin P.A. (1902) *Mutual Aid: A Factor of Evolution*, London: Heinemann.(Collects Kropotkin's articles of the 1890s on mutual aid in the animal kingdom and human society.)

Kropotkin P.A. (1922) *Ėtika (Ethics)*, ed. N. Lebedev, Petrograd and Moscow: Golos Truda.(Historical examination of the development of ethics.)

Kropotkin P.A. (1927) *Kropotkin's Revolutionary Pamphlets*, ed. R.N. Baldwin, New York: Dover.(Includes *La Morale anarchiste (Anarchist Morality)* (1891), *L'Anarchie: sa philosophie, son idéal (Anarchism: Its Philosophy and Ideal)*; *Modern Science and Anarchism*; and 'Anarchism', from the *Encyclopaedia Britannica* (11th edn). These all provide a clear outline of Kropotkin's anarchist communism and its relation to contemporary science and philosophy.)

References and further reading

Cahm, C. (1989) *Peter Kropotkin and the Rise of Revolutionary Anarchism 1872-1886*, Cambridge and New York: Cambridge University Press.(Includes introduction on Kropotkin's political and philosophical ideas.)

Marshall, P. (1993) *Demanding the Impossible: A History of Anarchism*, London.(Contains a very clear and readable chapter on Kropotkin the Revolutionary Evolutionist.)

Miller, M.A. (1976) *Kropotkin*, Chicago, IL and London: University of Chicago Press.(Well-documented account of life and work.)

Woodcock, G. and Avakumovic, I. (1950) *The Anarchist Prince*, London: Boardman.(Good introduction to life and ideas but no documentation.)

Kuhn, Thomas Samuel (1922-96)

The early 1960s saw substantial turmoil in the philosophy of science, then dominated by logical empiricism. Most important was the confrontation of the prevailing philosophical tradition with the history of science. Whereas the philosophy of science was mainly normatively oriented, that is it tried to delineate what good science should look like, historical studies seemed to indicate that the practice of science both past and present did not follow those prescriptions.

Thomas S. Kuhn was educated as a theoretical physicist but soon turned to the history and philosophy of science. In 1962, he published The Structure of Scientific Revolutions (SSR). This book was the single most important publication advancing the confrontation between the history and the philosophy of science; it is now a classic in science studies. SSR was most influential not only in the discussion within philosophy but also in various other fields, especially the social sciences. The central concepts of SSR, like scientific revolution, paradigm shift and incommensurability, have been in the focus of philosophical discussion for many years, and the term 'paradigm' has even become a household word (although mostly not in Kuhn's intended sense). After SSR, Kuhn continued to develop his theory; apart from minor modifications it is mainly the explication of SSR's more intricate philosophical topics, especially of incommensurability, which is characteristic of his later work.

1 Life and works

Kuhn received his Ph.D. in theoretical physics from Harvard University in 1949. While working on his dissertation he became interested in the history of science. His first book-length publication in this field was *The Copernican Revolution* (1957). From the late 1950s, Kuhn started to publish in the philosophy of science, leading to the appearance of *The Structure of Scientific Revolutions* (*SSR*) (1962). This work is now a classic; it has been controversially discussed in extremely diverse fields, and translated into some twenty languages. Kuhn's historical work after *SSR* concerned mainly the genesis of quantum mechanics. *Sources for the History of Quantum Physics* (1967) contains an inventory of relevant material including interviews with many of the contributors to quantum mechanics; *Black Body Theory and the Quantum Discontinuity, 1894-1912* (1978) is a controversial account of the introduction of the quantum hypothesis. Some of Kuhn's papers both on the history and the philosophy of science are contained in *The Essential Tension* (1977).

Kuhn taught at Harvard University until 1956 when he joined the University of California at Berkeley to teach history of science. In 1964, he left for Princeton University. Finally, he became professor for philosophy and history of science at MIT, from where he retired in 1992. He died in 1996.

2 *The Structure of Scientific Revolutions*: survey

Kuhn's *SSR* was probably the most important single work in bringing out the turmoil in philosophy of science in the early 1960s. The table of contents of *SSR* displays a developmental scheme for scientific fields in the basic sciences. This sort of structuring makes its philosophical reading difficult since, for a particular topic, passages scattered around the whole book must be considered.

The first chapter opens with a much-quoted sentence that describes concisely what *SSR* is all about: ' History, if viewed as a repository for more than anecdote or chronology, could produce a decisive transformation in the image of science by which we are now possessed' ([1962] 1970: 1).

Kuhn refers here to two modes of doing history of science. For the older historiographic tradition, the predominant goal was to reach a deeper understanding of *contemporary science* by displaying its historical evolution. This was to be done by discovering the elements of today's science, for example, its concepts, theories, experimental methods and so on, in older texts and by arranging them in a chronological narrative. The resulting picture of scientific development was necessarily a *cumulative* one: science grows step by step by adding new pieces of knowledge to those already in place. But this form of historiography invariably introduces distortions into the presentation of the older science by projecting today's science into the past. Instead, following especially Alexandre Koyré's model, the 'new historiography of science' attempts a display of the historical integrity of a science in its own time. More specifically, the concepts, the research problems and the standards of evaluation of an older science must be reconstructed in a historically adequate way. The aim of *SSR* is to delineate a new image

of science ‘ by making explicit some of the new historiography’s implications’ ([1962] 1970: 3).

The picture of science that emerges from this historiography contains a developmental scheme for scientific disciplines. Before reaching maturity, nascent scientific fields are typically characterized by controversies between competing schools; there is no consensus among the practitioners of the emerging field. This competition may eventually end when one group produces an exemplary solution to a pre-eminent research problem with two characteristics: it is sufficiently unprecedented to attract the members of the other schools, and it is sufficiently open-ended to leave enough interesting problems for further scientific work. These model solutions are called paradigms; they serve to guide research in an implicit way in the following period. This period is called ‘normal science’. It is characterized by a broad consensus of the practitioners of the field about fundamental questions, and, consequently, by a particular mode of research. This mode of research can be described by a five-dimensional analogy to puzzle-solving where exemplars of puzzles include chess problems and crossword puzzles. The analogy concerns:

- the existence of regulations constraining acceptable approaches to and solutions of problems;
- the expectation of the solubility of appropriately chosen problems;
- no fundamental changes in the guiding regulations;
- absence of test or confirmation of the guiding regulations;
- individual motivations: to prove oneself an expert puzzle solver.

Normal science is always confronted with anomalies, that is, with phenomena or problems that behave contrary to the expectations supplied by the paradigm. Usually, anomalies do not call into question the validity of the guiding regulations of normal research. But under special circumstances they may, and then they become ‘significant anomalies’. Then, the practice of science changes again into ‘extraordinary science’ or ‘science in crisis’. Extraordinary science aims at amending or even overthrowing the binding regulations. Its research focuses on the significant anomalies and their context. If this research leads to a new theory that is accepted by the scientific community because it can lead to a new phase of normal science, a scientific revolution has occurred. In Kuhn’s sense scientific revolutions are ‘ the tradition-shattering complements to the tradition-bound activity of normal science’ ([1962] 1970: 6). More specifically, in a scientific revolution an accepted theory is rejected in favour of a new one. This rejection is accompanied by a change of the problem-field and its related standards of solution, and by a sometimes subtle change in basic scientific concepts. Revolutions can even be described as transformations of the world in which scientific work is done. Kuhn compresses these features of revolutions into the concept of ‘incommensurability’: this relation holds between successive traditions of normal science. In *SSR*, the concept of incommensurability was not entirely clear; it was therefore subject to much criticism and also misunderstanding, and most of Kuhn’s philosophical work after *SSR* aims at a clarification and further explication of the concept of incommensurability (see §4) (see Incommensurability).

Because of incommensurability, we must, according to Kuhn, rethink the concept of scientific progress. First, progress is not cumulative, because there are conceptual changes during revolutions. Furthermore, Kuhn denies that scientific progress is an approach to truth. Instead of conceiving of scientific progress as a teleological process, that is, one that is goal-directed, we should think of progress as Darwinian evolutionary theory does. Accordingly, there is no ‘ set goal, a permanent fixed scientific truth’, but ‘an increase in articulation and specialization’ ([1962] 1970: 172, 173).

3 *The Structure of Scientific Revolutions*: philosophical significance

Why is Kuhn’s theory of scientific development philosophically significant? Why did philosophers take notice of a theory that appears essentially historical? The reason is that Kuhn’s theory ran counter to many philosophical convictions about science in the early 1960s. I have already mentioned Kuhn’s altered view of scientific progress (§2). This view implies the untenability of those forms of realism that assert that science at least approximately describes what is really ‘out there’, independently of any observer (see Scientific realism and antirealism §1). Instead, theories describe the world in terms of concepts which are historically contingent and which may change in the future. Furthermore, because of this change of basic scientific concepts, the classical conception of reductionism is hardly tenable. According to this conception, theories may be reduced to more fundamental theories by redefining their concepts by means of the concepts of the reducing theory, and then deriving their laws

from the laws of the reducing theory, supplemented by the redefinitions and, possibly, boundary conditions (see Reduction, problems of §2). But if incommensurability prevails between the pair of theories in question, the reduction relation cannot hold, since some of the required redefinitions are blocked by meaning shifts. In fact, as has become clearer in Kuhn's work after *SSR*, mutual untranslatability of some of the key terms is the hallmark of incommensurability between theories.

Many of *SSR*'s assertions stood in marked opposition to the Popperian philosophy of critical rationalism (see Popper, K.R.). For instance, the practice of normal science as described by Kuhn may look like bad science, because it is not directed at a critical test of the guiding assumptions, but at a quasi-dogmatic exploitation of their potential. Yet, in Kuhn's view, the task of critical evaluation of fundamental theories is restricted to the period of extraordinary science, and even then scientific practice is not simply an attempt to falsify theories by confrontation with basic statements about nature, as Popper would have it. Rather, theory evaluation is always a comparative procedure in which two (or more) theories are assessed with respect to their cognitive abilities, especially, whether they can cope with the significant anomalies from which the crisis state originated (see §2). From this view, theory falsification as described by Popper is a stereotype not found in the actual history of science.

Another consequence of Kuhn's theory is to abolish the idea that science is guided by a scientific method, a set of rules rigorously to be followed (see Scientific method §2). Due mainly to Bacon and Descartes' initial influence, this idea has dominated the understanding of modern science from its very beginning. But according to Kuhn, what chiefly guides scientific research in its normal phase are exemplary problem solutions. Their cognitive potential for research is exploited not by explicit (or fully explicable) rules but rather by implicit analogies; new problems are identified in the light of achieved ones, and new solutions are judged as legitimate in a like manner.

4 Further developments

It is not only that Kuhn opposed many doctrines of established philosophy of science; an additional factor for the sometimes vehement reactions to *SSR* was a deep and widespread misreading of its main theses. Kuhn has thus constantly been challenged to articulate his main theses more fully and to refine them, in the course of which his position has also shifted in some respects.

During the early reception of *SSR*, the main target of criticism was the paradigm concept, which was seen as equivocal. Kuhn responded by distinguishing a narrow sense of paradigm, meaning exemplary problem solutions, from a wide sense which comprises all the components of scientific consensus, the 'disciplinary matrix'. The latter includes, among other elements, scientific values (like accuracy, consistency, fruitfulness, scope and simplicity) which operate constantly but which are especially visible during theory choice. Since these values only guide and do not dictate theory choice, individual scientists can disagree in matters of theory evaluation without any of them being irrational. The latter would be a consequence if theory choice were determined by some fixed set of rules which defined the canon of scientific rationality.

The most important change in Kuhn's position has been a shift from the predominant description of scientific revolutions in *SSR* in terms of visual metaphors to a description by means of a linguistic framework. In *SSR*, revolutions were depicted as something like visual gestalt switches, and their outcome as a changed way of seeing the world. Though meaning shifts played some role in *SSR*, meaning becomes increasingly the dominant theme of Kuhn's later work. But what is the meaning of a term? How is a term connected with its referents? Kuhn does not deal with these profound questions in complete generality but only with kind terms, for example, mass nouns like 'gold' or count nouns like 'cat'. In the sciences, the meanings of kind terms are usually not specified by explicit definitions; rather, they are established by processes involving ostension of typical examples of the kind in question. When objects of the same kind are ostended, they are declared similar, and those not belonging to the same kind are declared dissimilar. By this procedure, a taxonomy of objects can be built up and transmitted. Kind terms involved in the taxonomy can stand in two relations only, namely, either exclusion or inclusion. Partial overlap between contrasting kind terms is forbidden: something is either a planet or a star but not both. Kuhn refers to the net of relations in such a taxonomy as the 'structure of the [respective] lexicon'. It is now characteristic for scientific revolutions, and hence for incommensurability, that the structure of the lexicon changes, resulting in partial untranslatability between propositions articulated by means of different lexicons. Changes of this kind may be initiated by the discovery of objects violating the no-overlap principle, that is, objects that seem to belong to mutually disjoint kinds. Objects of this sort cannot be described with the given lexicon nor

can the lexicon simply be extended or refined in order to make description possible; rather a partial reordering of the categories involved is inescapable.

See also: Feyerabend, P.K.; Incommensurability

PAUL HOYNINGEN-HUENE

List of works

Kuhn, T.S. (1957) *The Copernican Revolution. Planetary Astronomy in the Development of Western Thought*, Cambridge, MA: Harvard University Press.(A widely read historical textbook foreshadowing some of *SSR*'s theses.)

Kuhn, T.S. (1962, 1970) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press. (Kuhn's classic work; the second edition, to which citations refer, contains an important Postscript.)

Kuhn, T.S. (1977) *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago, IL: University of Chicago Press.(Contains historical and philosophical articles.)

Kuhn, T.S. (1978) *Black Body Theory and the Quantum Discontinuity, 1894-1912*, Oxford: Clarendon Press. (Kuhn's controversial narrative about the introduction of the quantum into physics.)

Kuhn, T.S. (1991) 'The Road Since Structure', in A. Fine, M. Forbes and L. Wessels (eds) *PSA 1990*, East Lansing, MI: Philosophy of Science Association, vol. 2, 2-13.(Summarizes Kuhn's development since *Structure of Scientific Revolutions*.)

Kuhn, T.S. (1993) 'Afterwords', in P. Horwich (ed.) *World Changes. Thomas Kuhn and the Nature of Science*, Cambridge, MA: MIT Press, 1993, 311-41.(Kuhn's reaction to papers presented at a conference on his work in 1990.)

Kuhn, T.S., Heilbron, J.L., Forman, P. and Allen, L. (1967) *Sources for the History of Quantum Physics. An Inventory and Report*, Philadelphia, PA: American Philosophical Society.

References and further reading

Barnes, B. (1982) *T.S. Kuhn and Social Science*, London: Macmillan.(A book about the sociology of scientific knowledge taking Kuhn's work as its point of departure.)

Gutting, G. (ed.) (1980) *Paradigms and Revolutions. Applications and Appraisals of Thomas Kuhn's Philosophy of Science*, Notre Dame, IN: University of Notre Dame Press.(A collection of articles, mainly from the 1970s, documenting Kuhn's impact in various fields.)

Hacking, I. (ed.) (1981) *Scientific Revolutions*, Oxford: Oxford University Press.(Essays by influential writers on scientific revolutions.)

Horwich, P. (ed.) (1993) *World Changes. Thomas Kuhn and the Nature of Science*, Cambridge, MA: MIT Press. (Proceedings of a conference about Kuhn's work, held in 1990 at MIT; includes Kuhn's reaction to the papers presented.)

Hoyningen-Huene, P. (1993) *Reconstructing Scientific Revolutions. Thomas S. Kuhn's Philosophy of Science*, trans. A.T. Levine, foreword by T.S. Kuhn, Chicago, IL: University of Chicago Press.(A critical reconstruction of Kuhn's theory and its development; contains the most complete bibliography of his writings.)

Lakatos, I. and Musgrave, A. (eds) (1970) *Criticism and the Growth of Knowledge*, London: Cambridge University Press.(Contains some of the most influential papers discussing Kuhn's work in the 1960s.)

Popper, K. (1975) 'The Rationality of Scientific Revolutions', in R. Harré (ed.) *Problems of Scientific Revolutions*, Oxford: Clarendon Press, 72-101.(Popper's answer to Kuhn's supposedly irrationalist view of scientific revolutions.)

Scheffler, I. (1967) *Science and Subjectivity*, Indianapolis, IN: Hackett; enlarged 2nd edn, 1982. (Representative of the reception and criticism of Kuhn in the 1960s.)

Shapere, D. (1984) *Reason and the Search for Knowledge. Investigations in the Philosophy of Science*, Dordrecht: Reidel.(Contains several highly influential papers criticizing Kuhn.)

Kūkai (774-835)

Kūkai, also known by his posthumous honorific title Kōbō Daishi, was the founder of Japanese Shingon ('truth word' or 'mantra') Buddhism and is often considered the first comprehensive philosophical thinker in Japanese history. Building on the Buddhist esoteric tradition first developed in India and then in China, where Kūkai encountered it, he maintained that reality is a cosmic person, the Buddha Dainichi. Dainichi's cosmic thoughts, words and deeds form microcosmic configurations, resonances and patterns of change. By performing Shingon rituals, one can supposedly accord with the microcosmic constituents and know the foundational structures of reality that compose the sensory world in which we ordinarily live.

Kūkai was an aristocrat born on the island of Shikoku, far from the major cultural centres. Trained in the Chinese Classics, he went to the capital area of Nara-Nagaoka at the age of eighteen to study for qualification in the court bureaucracy. Within a few years, however, he left the college and isolated himself in the mountains, practising austerities. In 804 he left for China to study the texts and practices of esoteric Buddhism, which had been only unsystematically and partially introduced to Japan. In China Kūkai studied under Huiguo, the seventh patriarch of Zhenyan (Shingon) Buddhism. He returned to Japan in 806 as the eighth patriarch of the tradition and the first in Japan.

Kūkai devoted his efforts in Japan to establishing Shingon as a major school or sect. He founded a monastic community and temple complex on Mt. Kōya and a major temple, Tōji, in the new capital of Kyoto. He wrote poems, copied sacred texts, painted *mandalas* and undertook civil works such as establishing a public school for boys and girls of all classes, engineering irrigation projects and performing rituals for the relief of plagues and famines. Popular culture reveres Kūkai as a great thaumaturgist of ancient Japan.

Kūkai's philosophical contributions were no less impressive. A prolific writer, he wrote many treatises to explain and justify the basic Shingon position. His philosophical works generally concentrated on two themes, his critique of other philosophical schools and his elaboration of the relation between the nature of reality and the participation in that reality through Shingon ritualistic practice.

On the first point, he developed his theory of the 'ten mindsets' as explained in his *Jūjūshinron (Treatise on the Ten Mindsets)* and *Hizō hōyaku (Precious Key to the Secret Treasury)*. In those works he analysed the limitations of the philosophies known in Japan at the time, including Confucianism and Daoism as well as exoteric forms of Buddhism. Although it was then common in Chinese Buddhism to classify the teachings of the various schools, Kūkai's innovation was to study the teachings as an expression of specific mindsets; he was as interested in the experiential basis of the ideas as in what they said about reality and to whom they said it. This allowed him simultaneously to see their appropriateness and to criticize their incompleteness.

Kūkai's typical critique of the mindsets outside Shingon Buddhism was that they could not adequately justify the basis for their worldviews. For example, the lowest mindset understands its own human existence to be controlled by animal desires. Yet no animal desire could itself lead one to that conclusion; therefore, the mindset is based in something it cannot explain within its own terms. Towards the other end of Kūkai's scale, the ninth mindset, that associated with Kegon (Huayan) Buddhism, understands reality as a web of interdependence in which each thing reflects every other thing. Kūkai basically agreed with that metaphysical model, but questioned whether Kegon could adequately explain how that fact is known. Indeed, what would 'knowledge' even mean in such a context? Kūkai insisted that only the tenth mindset, that of Shingon's esoteric Buddhism, could adequately assert a metaphysical theory and ground that theory within its own parameters. This claim points to Kūkai's theory about knowledge as the self-aware participation in reality.

According to Shingon the cosmos as a whole is a person, the Buddha Dainichi. Like any other person, Dainichi's behaviour can be analysed in terms of thought, word and deed, the respective functions of mind, speech and body. On the cosmic level these functions give reality its configuration, resonance and patterns of change. Specifically, Dainichi's geometric contemplative visualisations (*mandala*) shape reality, Dainichi's intoning of sacred syllables (*mantra*) fill the universe with resonant states of matter-energy, and Dainichi's ritualistic gestures (*mudra*) give the universe its patterns of movement. Therefore, according to Kūkai's analysis, each thing in the universe is

ultimately an expression of Dainichi's enlightened activity.

Of course to our human senses, the universe does not ordinarily appear that way. Because the senses cannot experience the cosmos as a whole, how could Kūkai ground his metaphysical position? His argument was that, like every other thing in the universe, human beings are the product of Dainichi's self-expression. Unlike most other expressions of Dainichi, however, human beings are also persons, beings who express themselves through the activities of thought, word and deed. Hence, the basic functioning of the cosmos is recursively manifest or reflected in each human being. If we delude ourselves into thinking that our actions are strictly our own, separate from all the rest of reality, we cannot justify our metaphysics. What knows metaphysics must itself be metaphysically explainable.

On the other hand, Kūkai argued, if our everyday actions mirror the basic functions of the universe, we can recognize that our own actions are also expressions of Dainichi's activity. The knower and the known are part of the same metaphysical system. Yet how can one become fully aware of this mirroring? The answer is, by enacting the same three activities as Dainichi: contemplating the *mandalas*, intoning the *mantras* and gesturing the *mudras*. Through ritualization, one's own self-expression becomes consciously the same as the cosmos' self-expression. To know reality is to self-consciously participate in it as its own self-expression. Religious praxis and religious knowledge become, in the end, identical. According to Kūkai, that insight was the unique contribution of esoteric Buddhism.

See also: Buddhist philosophy, Japanese; Japanese philosophy; Knowledge, concept of; Metaphysics

THOMAS P. KASULIS

List of works

Kūkai (774-835) *Kōbō Daishi zenshū (Complete Works of Kōbō Daishi)*, ed. Inabe Yoshitake *et al.*, Koya-san: Mikkyō bunka kenkyū-sho, 3rd edn, 1965; trans. Hakeda Yoshito, *Kūkai: Major Works*, New York: Columbia University Press, 1972.(Inabe is the standard critical edition of the complete works in the original Chinese. Hakeda is an abridged translation of the eight treatises generally considered to be the core of Kūkai's philosophy, including a helpful introduction about Kūkai's life and thought.)

Kūkai (830) *Jūjūshinron (Treatise on the Ten Mindsets)*, in *Kōbō Daishi zenshū*, Koya-san: Mikkyō bunka kenkyū-sho, 3rd edn, 1965, vol. I.(First develops his theory of the ten mindsets.)

Kūkai (830) *Hizōhōyaku (Precious Key to the Secret Treasury)*, in *Kōbō Daishi zenshū*, Koya-san: Mikkyō bunka kenkyū-sho, 3rd edn, 1965, vol. I; abridged trans. in Hakeda Yoshito, *Kūkai: Major Works*, New York: Columbia University Press, 1972.(Compares various other philosophies to Buddhism.)

References and further reading

Kasulis, T.P. (1982) 'Reference and Symbol in Plato's *Cratylus* and Kūkai's *Shōjijissōgi*', *Philosophy East and West* 32 (4): 393-405.(Cross-cultural comparison of two early works on language and philosophy in ancient Greece and Japan.)

Kasulis, T.P. (1988) 'Truth Words: The Basis of Kūkai's Theory of Interpretation', in D.S. Lopez, Jr (ed.) *Buddhist Hermeneutics*, Kuroda Institute Studies in East Asian Buddhism 6, Honolulu, HI: University of Hawaii Press, 257-72.(A general study of Kūkai's theory of truth and its application to his evaluation of the ten mindsets.)

Kasulis, T.P. (1990) 'Kūkai: Philosophizing in the Archaic', in F. Reynolds and D. Tracy (eds) *Myth and Philosophy*, Albany, NY: State University of New York Press, 131- 50.(A general overview of Kūkai's philosophy within its historical and cultural setting.)

Kiyota Minoru (1978) *Shingon Buddhism: Theory and Practice*, Los Angeles, CA: Buddhist Books International. (A detailed, scholarly and highly technical explanation of the system of Shingon doctrine, mainly as developed by Kūkai.)

Shaner, D.E. (1985) *The Bodymind Experience in Japanese Buddhism: A Phenomenological Study of Kūkai and Dōgen*, Albany, NY: State University of New York Press.(A good comparative analysis of Kūkai's theory of the oneness of body and mind using Western phenomenological categories.)

Yamasaki Taikō (1988) *Shingon: Japanese Esoteric Buddhism*,Yasuyoshi Morimoto and D. Kidd (eds), Boston and London: Shambhala Publications.(Excellent introduction to Shingon doctrine and practise in a readable format for a general audience.)

Kuki Shūzō (1888-1941)

Kuki's philosophical project was focused on the issues arising from dualistic thinking. He incorporated into his work a cross-cultural, historical perspective, while applying Heidegger's hermeneutical ontology and exhibiting bold, systematic, speculative acumen.

Kuki Shūzō was born in Tokyo as the fourth son of Baron Kuki Ryūichi. He graduated in 1921 from Tokyo Imperial University, where he majored in Western philosophy and found Raphael von Köbel and Okakura Tenshin particularly inspirational. His graduation thesis dealt with mind-body correlativity. In 1922 he went to Europe to study under Rickert, Oskar Becker, Husserl and Heidegger in Germany, and had an association with Bergson in France. (It was Kuki who introduced Sartre to the philosophies of Husserl and Heidegger.) While in Europe, Kuki acquired extensive and detailed knowledge of Western philosophy. He returned to Japan in 1929, and in 1935 became a professor at Kyoto Imperial University, where he lectured on the history of Western philosophy (see Kyoto School).

Of his complete works - almost all of which attempt to go beyond modern European philosophy - the three best known in Japan are *Iki no Kōzō (The Structure of Iki)* (1930), *Gūzensei no mondai (The Problem of Contingency)* (1935), and *Ningen to jitsuzon (The Human and Human Existence)* (1939). In Heidegger's *Unterwegs zur Sprache (On the Way to Language)*, Kuki's *iki* is thematized in the dialogue between *iki* and a Japanese interlocutor.

Kuki's concern for dualism is expressed in *Iki no kōzō*, wherein he thematized an aesthetic mode of living [*iki+ru*] uniquely found in the intersexual dealings between man and woman that were cultivated in the 'pleasure quarter' as an expression of 'being-of-the-people', in contrast to Heidegger's *Dasein*. He analyzed this mode of living as a phenomenon of consciousness embracing three constitutive momenta: coquetry, liveliness and resignation. It is through the spiritualization of coquetry via the idealism of *bushidō* (see Bushi philosophy) and the appropriation of destiny via Buddhism (in other words, resignation) that *iki* is lived aesthetically as 'liveliness'. Kuki's general thesis is that aesthetics must be sought in the phenomenon of consciousness as it is historically nurtured and articulated in language *vis-à-vis* the 'being-of-the-people'.

In *Gūzensei no mondai*, his *magnum opus*, Kuki brought his concern for dualism to a metaphysical height. He scrutinized in 'metaphysical solitude' the problem of contingency, focusing on: (1) propositional (that is, logical) contingency, which arises from the encounter between particulars ('eidetic singularity'); (2) hypothetical contingency (which encompasses the empirical causal and the teleological); and (3) metaphysical contingency, arising from the playful discrepancies within the one-many relationship. Contingency, or lack of necessity, arises where there is a duality between the necessary one (being or Being) and the contingent other *qua* the many (non-being). Only when there is a concrete interiorization of the other within the one can the logic of life be understood. Herein can be discerned Kuki's effort to resurrect life in the otherwise empty, abstract universals devoid of the humanly constitutive elements.

See also: Contingency; Heidegger, M.; Kyoto School; Logic in Japan

SHIGENORI NAGATOMO

List of works

Kuki Shūzō (1930-39) *Kuki Shūzō zenshū(The Complete Works of Kuku Shūzō)*, Tokyo: Iwanami shoten, 1980-82, 12 vols.(Posthumous collection of Kuki's philosophical and literary works.)

Kuki Shūzō (1930) *Iki no Kōzō (The Structure of Iki)*, Tokyo: Iwanami shoten; trans. Maeno Toshikuni, *Structure de l'Iki*, Tokyo: Maison Franco-Japonaise, 1984.(An aesthetic evaluation of the phenomenon of *iki* in the being of Japanese people, relative to the other Japanese aesthetic values.)

Kuki Shūzō (1935) *Gūzensei no mondai (The Problem of Contingency)*, Tokyo: Iwanami shoten; trans. Omodaka Hisayuki, *Le problème de la contingence*, Tokyo: L'Université de Tokyo, 1966.(An attempt to resurrect a logic of life through a philosophical reflection on the problem of contingency.)

Kuki Shūzō (1939) *Ningen to jitsuzon (The Human and Human Existence)*, Tokyo: Iwanami shoten.(A philosophical probe into the 'primitive contingency' in regard to the 'here and now' of human existence in an attempt to gain a primordial understanding of both being and nothingness.)

References and further reading

Light, S. (ed.) (1987) *Shūzō Kuki and Jean-Paul Sartre: Influence and Counter-Influence in the Early History of Existential Phenomenology, Including the Notebook 'Monsieur Sartre' and Other Parisian Writings of Shūzō Kuki*, Carbondale, IL: Southern Illinois University Press.(A recounting of Kuki's stay in Europe including his association with Sartre, and a commentary on two of his essays in French, 'La notion du temps et la repreise sur le temps en Orient' and 'L'expression de l'infini dans l'art japonais'.)

Pincus, L. (1996) *Authenticating Culture in Imperial Japan*, Berkeley, CA: University of California.(A recasting of Kuki's cultural specificism surrounding the *Iki no Kōzō vis-à-vis* a postmodern, cultural and literary criticism.)

Sakabe Megumi (1990) *Fuzaino uta (The Songs of Absence)*, Tokyo: TBS Buritanika.(An in-depth and comprehensive summary of Kuki's major works.)

Viswanathan, M. (1989) 'An Investigation into Essence: Kuki Shūzō's 'Iki' no kōzō', *Transactions of the Asiatic Society of Japan* 4 (4): 1-22.(A balanced summary of *Iki no kōzō*.)

Yuasa Yasuo (1987) 'The Encounter of Modern Japanese Philosophy with Heidegger', in G. Parkes (ed.) *Heidegger and Asian Thought*, Honolulu, HI: University of Hawaii Press,155-74.(A brief introduction to Kuki, along with other prominent Japanese philosophers who studied in France and Germany in the 1920s and 1930s.)

Kumazawa Banzan (1610-91)

Trained under Nakae Tōju, the founder of the Wang Yangming school (Yōmeigaku) of Confucian idealism during the Tokugawa era in Japan, Kumazawa Banzan is known for his eclectic and pragmatic philosophy emphasizing political and economic reforms. For example, he recommended that the shogunate take greater responsibility for promoting economic equity and prosperity by supporting rice production and storage as well as the use of rice as the medium of exchange. He advocated a form of government dedicated to the Confucian ideal of benevolence.

Kumazawa Banzan is generally regarded as the second leading figure in the Japanese Wang Yangming school of Confucian idealism, following his teacher Nakae Tōju (1608-48), founder of the movement during the Tokugawa era (see Confucian philosophy, Japanese). Whereas Nakae stressed Wang's ideal of the unity of knowledge and action from the standpoint of personal discipline and cultivation, thereby setting an example that greatly inspired his student, Kumazawa, a leaderless samurai (*rōnin*), emphasized the application of this theory to the political and economic realms. He achieved great success and fame as a reformer in the service of Ikeda Mitsumasa, the lord (*daimyō*) of Bizen Province from 1645-56. However, his unorthodox views and outspoken criticism of what he considered a somewhat oppressive and ineffective government led to his resignation and a life of semi-exile in official disfavour.

His main works are two collections of dialogues on moral philosophy and social problems, commentaries on the Confucian classics, a commentary on the *Tale of Genji*, and the *Daigaku Wakumon (Dialogues on Learning)*, a series of dialogues on contemporary social, political and economic issues which was long suppressed by the Tokugawa authorities. The latter text, the title of which refers to questions concerning the Confucian classic *Daxue (Great Learning)*, actually consists of Kumazawa's recommendations for restoring vitality to government, prestige to the ruling *samurai* class, self-identity and fairness to the peasant farmers and protection for the isolated and vulnerable Japanese nation at risk from possible attacks by the Mongols.

Although a follower of the idealist school, Kumazawa was also influenced by the rival Zhu Xi school of Confucian realism (seeZhu Xi), and he combined a belief in the innate capacity of humanity to grasp truth and choose the correct conduct with an emphasis on the value of empirical knowledge of external reality. In particular, he opposed the shogunate's strict adherence to Confucian ritualism and instead advocated a government dedicated to the genuine pursuit of the Confucian philosophical ideal of *jinsei* (benevolence) and the recognition of individual merit based on performance rather than rank in the conventional class hierarchy. Kumazawa sought to eliminate class discrimination which he considered counter-productive for the shogunate, and he proposed the resettlement of *samurai* on the land, the relaxing of restrictions on the *daimyō* (especially the requirement that they divide their time between their home town and the capital in Tokyo), and the combining of warriors into one class with farmers. In addition, Kumazawa recommended for the proposed *samurai*-farmer class a partial return to rice as the medium of exchange, the storing of rice against drought or invasion, and the construction of flood basins and the afforesting of mountains.

A critic on religious issues as well as a political dissenter, he stressed the abolition of methods for suppressing Christianity and the institution of strict tests for men aspiring to become Buddhist priests. Kumazawa's views exerted a great influence on thinkers and reformers near the end of the Tokugawa shogunate in the mid-nineteenth century.

See also: Confucian philosophy, Japanese; Daxue; Political philosophy, history of; Wang Yangming; Zhu Xi

STEVEN HEINE

List of works

Kumazawa Banzan (1610-91) *Kumazawa Banzan zenshū(Collected Works of Kumazawa Banzan)*, Tokyo: Banzan zenshūkankōkai, 1940-3, 6 vols.(The collected works of Kumazawa Banzan.)

Kumazawa Banzan (1610-91) *Daigaku Wakumon (Dialogues on Learning)* trans. G.M. Fisher, 'Dai Gaku Wakumon: A Discussion of Public Questions in the Light of the Great Learning', *Transactions of the Asiatic Society of Japan* 16, 1938: 259-356.(Translation of the main text, emphasizing social, political and economic reform.)

References and further reading

McMullen, J. (1983) ‘Kumazawa Banzan’, *Kodansha Encyclopedia of Japan*, Tokyo: Kodansha, vol. 4, 307.(A concise, insightful account of his life and thought.)

Ryūsaku Tsunoda, de Bary, W.T. and Keene, D. (eds) (1958) ‘Kumazawa Banzan, A Samurai Reformer’, in *Sources of Japanese Tradition*, New York: Columbia University Press, vol. 1, 375-88.(An historical overview with translations of selected passages.)

Kyoto School

The Kyoto school of philosophy pivots around three twentieth-century Japanese thinkers who held chairs of philosophy or religion at Kyoto University: Nishida Kitarō (1870-1945), Tanabe Hajime (1885-1962) and Nishitani Keiji (1900-91). Its principal living representatives, who also held chairs at Kyoto until their retirement, are Takeuchi Yoshinori (1913-) and Ueda Shizuteru (1926-). The keynote of the school was struck by Nishida in his attempt, on the one hand, to offer a distinctively Eastern contribution to the Western philosophical tradition by bringing key Buddhist concepts to bear on traditional philosophical questions, and on the other, to enrich Buddhist self-understanding by submitting it to the rigours of European philosophy.

The name 'Kyoto school' was coined in 1932 by the Marxist philosopher Tosaka Jun (1900-45) to denounce what he saw as a bourgeois ideology - which he characterized as 'hermeneutical, transhistorical, formalistic, romantic, and phenomenological' - that had grown up around Nishida, Tanabe and their immediate disciples at the time. These latter included Miki Kiyoshi (1897-1945), Kosaka Masaaki (1900-69) and Koyama Iwao (1905-93) as well as the young Nishitani. At the time the Japanese state had taken its first definitive steps in the direction of a militaristic nationalism that would involve it in the 'fifteen-year war' with Asia and finally the West over the period 1930-45. As the leading philosophical movement in Japan, the Kyoto school was caught up in this history, although there was little unanimity among the responses of the principal figures.

Postwar criticisms and purges of the Japanese intelligentsia attached a certain stigma to the school's name, but later and more studied examination of those events, as well as the enthusiastic reception of translations of their works into Western languages, has done much to ensure a more balanced appraisal. Today, the philosophy of the Kyoto-school thinkers is recognized as an important contribution to the history of world philosophy whose 'nationalistic' elements are best recognized as secondary, or at least as an unnecessary trivialization of its fundamental inspirations.

As a school of thought, the common defining characteristics of the Kyoto school may be seen in an overlap of four nodal concerns: self-awareness, the logic of affirmation-in-negation, absolute nothingness and historicity.

1 Survey

The story of Western philosophy in Japan began just over one hundred years ago with the general opening up of Japan after more than two centuries of isolation from the rest of the world. As young scholars went to Europe to study, and philosophical literature became available, the currents of neo-Kantian and neo-Hegelian thought (see Neo-Kantianism; Hegelianism) opened a gateway to Greek, medieval and modern philosophy, almost all of which was read in the Western languages. These ideas were taken over by Buddhist thinkers anxious to update their self-understanding. While most of the work of these early pioneers was premature and uncritical, they charted a course which within a generation would produce its first original thinker in the person of Nishida Kitarō.

Like others of his age, Nishida's initial interests were in British empiricism and positivism, but by the time he began at Tokyo University, German philosophy was in vogue (even though the medium of instruction for foreign professors was English). After graduating in 1894 and before taking up employment, he had the idea of writing a book to introduce the ethical theory of the British philosopher T.H. Green to the general Japanese reading public. Although much of the subtlety of Green's language and thought was lost on the young Nishida, leading him to abandon the project in mid-stream, the association of theories of the good and the Hegelian dialectic struck a chord that resonated in him for the next fifteen years, culminating in the publication of his maiden work, *Zen no kenkyū (A Study of the Good)*, in 1911.

In general, the most important intellectual influences during these years were Bergson, William James and Hegel. However, the critical catalyst that convinced Nishida to strike out creatively in his own direction came from quite another quarter. For at least ten years, beginning in 1896, he practised meditation under a Zen master and was assigned the koan *mu*, or 'nothingness'. While there is little evidence that the variety of Zen systems of thoughts as such influenced the maturation of his philosophical ideas, the discipline of meditation was decisive. On the one hand, it convinced him that philosophy does not begin in knowing *about* things through the medium of objective logic, but in the knowing *of* things through the medium of direct experience. On the other, it showed him that

philosophy does not end there, in ineffability, but only begins its work of expressing as clearly as possible the structure of reality and the place of human meaning and action within it. For if the pure ideas transplanted from Western philosophy needed to sink roots in the lived experience of Japan, so did the kind of pure experiencing cultivated in Buddhist meditation need to mature into self-expression under the critical eye of philosophical thinking (see Buddhist philosophy, Japanese).

This ideal of bringing pure experience and clear understanding to bear on one another was the hallmark of the *Zen no kenkyū* and runs throughout all Nishida's philosophy. If Nishida's thinking straddles East and West, it does not do so on the naive supposition that the one can be made to supply what the other lacks. The fact that he chose the term 'pure experience' from James to express the direct experience he had come to know through Zen, and the Buddhist idea of 'nothingness' to clarify Western ideas about the ultimate ground of understanding, shows that the standpoint he sought was finally philosophical in the proper sense of the term. It was a standpoint that transcended East and West, not in the sense that it bleared distinctions of culture and tradition, but in the sense that it pointed to the aim of clear thinking in all culture and tradition as a lived 'religious' expression, in word and deed, of universal ideas. Through his disciples Tanabe Hajime and Nishitani Keiji, this standpoint distanced itself from Nishida's person and mode of expression to become a forum for thought of great variety, but never lost the religious-philosophical stamp that Nishida had given it.

Although Nishida himself never left Japan, he encouraged Tanabe to go to Germany. There the latter studied Husserl, Cohen and Natorp, and became close to Heidegger. The religious character of Tanabe's philosophic concerns was not associated with any particular Buddhist sect, though he favoured the Sōtō Zen of Dōgen more than the koan-centred Zen of the Rinzai sect, and later turned enthusiastically to Pure Land Buddhism and Christianity for philosophical inspiration. The unifying factor for him was Hegelian logic and dialectic. Nishitani, who also later studied in Germany, took an early interest in Nietzsche, whose thought had an enduring impact on him. Like Nishida, Nishitani's primary connection with Buddhism was Zen, both Rinzai and Sōtō; but like Tanabe, Nishitani wrestled with Christian ideas through most of his mature years, mainly through his reading of the mystics and the European existentialists.

Takeuchi's religious frame of reference has consistently been that of Pure Land Buddhism (it was through his influence that his teacher, Tanabe, took to reading Shinran), seasoned by a studied reading of modern Christian theologians such as Bultmann and Tillich. With encouragement from Nishitani, Ueda studied in Germany and wrote a doctoral thesis on Meister Eckhart. During his teaching career he wrote mainly on Zen and Christian mysticism, but since retirement has concentrated on invigorating interest in Nishida's philosophy as such. In both Takeuchi and Ueda we find less of the kind of direct engagement with Western philosophical ideas that we find in Nishida and Tanabe, and practically none of the sort of textual analysis we see in Nishida's struggles with neo-Kantian thought or Nishitani's studies on Aristotle and Descartes. In their case, the influence of Western philosophy has been filtered through their reading of Nishida and Tanabe, who are now recognized as world-class philosophers in the Western sense of the term. Far from indicating a dilution of Nishida's original inspiration, their relative distance from the West shows the sense in which the Kyoto school has turned in the direction of thought that can be approached philosophically on its own merits, without the need to establish credibility through the study of major Western thinkers.

Despite the constant presence of Western intellectual history in the writings of the Kyoto school philosophers and their insistence on the role that Japanese thought has to play in the world philosophical forum, dialogue with their philosophical contemporaries outside Japan played a surprisingly insignificant role in the development of their thought. Differences between their respective positions and debate over those differences neither solicited nor attracted the attention of the philosophical community of Europe and the United States, let alone of other Asian countries. In part this was due to the inaccessibility, until quite recently, of their works in Western languages. In part it was also due to the largely tacit assumption that only the Buddhist East, and perhaps only Japan, as the sole country to engage in modernization freely and without the pressures of colonial rule, was equipped to produce a philosophy that can bridge East and West.

The launching of the Kyoto school, and hence Japanese philosophy, into the world philosophical forum was not the result of determined efforts by the philosophers themselves, but of initiatives from their Western counterparts in philosophy and religious studies. A flurry of translations, as well as a wave of young scholars from Western

countries trained to read the original texts and converse in Japanese, not only brought their philosophy to the attention of the world community but prompted a revival of interest in the Kyoto school philosophers in Japan. Ironically, it has only been after the passing of Nishida, Tanabe and Nishitani that the philosophy of the Kyoto school has passed from the stage of being a Japanese event witnessed by a privileged few who knew the language, to a common forum of ideas in which philosophers from around the world can discuss and debate.

Given the relative isolation of the writings of the Kyoto school philosophers from direct criticism by their contemporaries abroad, their thought was free to take new directions that might not have occurred to them under the monitoring of foreign tutors. Similarly, the lack of affinities between their own language and ancient Greek and Latin, and the lack of popular currency to the vocabulary created to translate Western philosophical terminology, gave an entirely different flavour to the linguistic frame within which they worked. The borrowed vocabulary was generally translated taking only the flat, surface meaning, with no thought to associations of feeling or history or literary allusion. Despite its claims to indigenous thinking, their language is barely less alien to those untrained in its meanings than was the rest of Western philosophy.

On the positive side, the etymology and historical resonances of the Chinese characters suggested different associations, and led them in directions that the retranslation of their works back into Western languages all but entirely gloss over. Well before Heidegger’s works were translated into English, and even while the battles over how to render his wordplays into acceptable idiom were raging in the West, the Japanese showed their natural affinity for his attempts to disclose the wonderful world within worlds of language.

At the same time, their inheritance of modern philosophical questions - from epistemology and the philosophy of science to existentialism and the overcoming of metaphysics - allowed certain questions to pass by unnoticed. In this regard, one thinks particularly of the question of criteria for distinguishing literal truth from symbolic truth. Their failure to concern themselves with symbolic theory in general, and with demythologizing religious truth or distinguishing faith from rationalism in particular, made it possible for them to work from a standpoint of ‘philosophy-religion’ that is not so easily taken by theologians or philosophers in the West.

2 Self-awareness

As noted above, Nishida’s concern with Western philosophy stemmed from the need for an objective standpoint from which to rethink, systematically and rigorously, the experience of being rational within an Eastern, and in particular a Japanese, culture. Though his writings make few direct references to Zen, in his lectures he is reported to have defined his philosophical aim as seeking a rational foundation for the Zen experience of enlightenment. The term he chose for this, ‘self-awareness’ (*jijaku*), is an ordinary word in Japanese which had loose ties to Buddhist and Western philosophical thought but had not been preempted as technical jargon by either of them. On this term he laid the burden of resisting the self-understanding being urged on Japan from the outside world and at the same time of submitting vernacular patterns of self-understanding to as serious a critique as possible. In this sense, it was less a clearly defined concept than a kind of bridge that allowed him to move back and forth freely between two worlds of ideas and forge a philosophy of his own. Moreover, because of its inherent ambiguity, the term provides a kind of symbolic connection between Nishida and his disciples, even when other differences of opinion seem to pull them far apart.

The designation of awareness as *self*-awareness is more than the result of literal translation. It draws attention to two important traits of enlightened consciousness for Nishida. First, it was an awareness that comes about *of itself*, without the interposition of judgment or discursive knowing on the part of the individual. It is an event, not the result of any systematic technique of reflection; it is the foundation of the philosophic enterprise, not the aim. Second, the awareness is without any ‘object’ that distinguishes it from the ‘subject’ of the awareness. It is, one might say, an awareness of awareness itself, not of any one that is aware or about any thing that is the focus of consciousness. In other words, it is the state of pure experience prior to the distinctions and judgments that begin with the distinction of any experiencing subject and an experienced object.

In his early writings, Nishida spoke of this refinement of everyday subjectivity into self-awareness as the discovery of the ‘true self’, a no less ambiguous yet symbolically important term that appears commonly in the work of both Tanabe and Nishitani as well. The more he threw himself into the Western intellectual tradition, the more he recognized cognates there for the notion of a self that loses itself in being aware of itself, and thus

gradually gave up allusions to its distinctively Buddhist quality. His focus shifted rather to seeing the knowing, feeling and experiencing self of ordinary consciousness as the maidservant of self-awareness, in effect inverting Western philosophy as he knew it.

Nishitani took up this concern with the true self and the attainment of an awakened subjectivity in serious confrontation with Western philosophical texts. At the same time, he delved deeper into Zen Buddhist literature in search of connections between self-awareness of one's true self and the enlightenment experience. For Nishitani, self-awareness consisted of a 'realization' that entailed both *really* coming to one's senses and a coming to *reality* just as it is. Awakening to this self means that consciousness 'sees' itself as an event in reality even as it 'sees through' any attempt to set itself over against the incessant change and interdependence of all things that are in the world of being. It is not ordinary consciousness, which sets itself up as a knower of the world to be known and hence puts itself in a position to change it, but a sort of consciousness of being conscious in the world. In this sense, as Nishitani points out, self-awareness is not the awareness of a self set up in opposition to another, but of a true self in which self and other are no longer two. He contrasted this self as a 'letting go of ego', by which he meant not simply ridding oneself of selfishness or uncluttering the everyday mind, but of 'breaking through' of ego in its highest form (for Nishitani, this is the Cartesian ego) (see Descartes, R.). The latter term, adopted from Meister Eckhart, shows how Nishitani did not hesitate to point out equivalent critiques of the ego in the Christian tradition and to apply them to clarifying classical Buddhist notions.

In Tanabe, the use of the terms 'self-awareness' and 'true self' was less self-conscious than it was in Nishida and Nishitani until he came to his project of revitalizing philosophy as a 'metanoetics'. There self-awareness is seen less as an immediately accessible, primordial state than as the disciplined, deliberate letting go of the 'self that is not a self' through reason exhausting itself on its own principles. In his later writings he used the notion deliberately to stress his differences from Nishida, but the basic meaning, and the fact that it was the lodestone of true philosophical thought, were never questioned.

3 The logic of affirmation-in-negation

In giving self-awareness a central place and all but identifying it with religious experience, Nishida distanced himself from the classical Western distinction between faith and reason. At the same time, he rejected the hasty appeals to the transrational or paradox in which Zen and other Eastern philosophies have often found a convenient retreat from criticism. Rather, he recognized the need to introduce a new logic to complement the two-valued logic of affirmation-or-negation that he found in his Western counterparts. His aim was, as he said, to give 'philosophical moorings' to the centuries-old preoccupation of Eastern cultures with such things as the 'form of the formless' or the 'voice of the voiceless'. This new logic was one of affirmation-in-negation and negation-in-affirmation (see Logic in Japan).

The grammatical index of this logical relationship was the Sino-Japanese copula *soku* (rendered here as 'a-*in*-b') which Nishida used freely but did not assign an explicit place in formal logic. The most succinct statement of the *soku* logic is given by D.T. Suzuki, Nishida's friend from high school days: '*A* is not *A*, therefore *A* is *A*'. The point is not to deny the principle of contradiction by claiming that '*A* is *A*' and '*A* is not *A*' mean the same thing; nor is it simply to assert that ultimately there is nothing in the complexity of the real world that corresponds to the logical form '*A* is *A*'. Nishida did not mean to reject the need for abstract and logically regulated thinking. His point was rather to insist that the 'identity' of a thing is not contained within a thing itself, as a kind of self-subsistent fact waiting for the mind to discover it, but is always a function of its interaction with what it is not. This identity is therefore not something that can be grasped by a logically thinking subject passing judgment on a perceived object, but only in an act of self-awareness in which the logical affirmation of the thinking subject is taken up into an experience that erases the distinction between subject and object, between affirmation and negation. It is only after one has seen that '*A* is not *A*' that one is able to return to subjective judgment to understand what is meant by the claim that '*A* is *A*'.

This act of seeing that '*A* is not *A*' preoccupied Nishida for several years during which he concentrated his attentions on a process he called 'acting intuition'. Briefly put, his idea was that pure intuition needs to be deliberately *cultivated* as a way of acting on the world, participating in the world's dynamic by expressing it in creative form, without interposing the subject-object dichotomy on it. Nishida describes this as a 'seeing without a seer' and of 'knowing a thing by becoming it'. Later he used the term 'self-identity of absolute contradictories' as

a general index of the *soku* logic and its corresponding epistemological theory.

Tanabe took up the *soku* logic as a way for him to appropriate the Hegelian dialectic without succumbing to its assumptions about an unfolding of Being and consciousness in which the opposites are subsumed into a higher unity. (In fact, one of his points of disagreement with Nishida was what he saw as the reintroduction of a quasi-Plotinean One in the idea of the self-identity of absolute contradictories (see Plotinus §3).) Tanabe set the *soku* relationship in the context of a broader logic of what he called 'absolute mediation', his own alternative to the logic of non-contradiction. Like Nishida, he insisted that logic is a way of seeing that only makes sense when it is engaged in the seen, but he went further to see affirmation-in-negation as a function of a reality that is always in the making. The mutual mediation among things, their *soku* relationship, propels them through time and gives the lie to simple noncontradiction, but is also itself a part of the process. It is 'absolute' precisely in the sense that it does not stand above the process it enables. Hence while it governs contradictories like death-in-life, the *soku* relationship also applies to identities that do not involve contradiction, such as God-in-love.

While Nishitani was never persuaded that Tanabe's idea of absolute mediation corrected a fault in Nishida's logic of *soku* and tended to use it as a copula between absolute contradictories, neither did he adopt the latter's terminology in his own philosophical writings. Moreover, like Nishida, he tended to see the *soku* statement as an index of a realization which precludes the working of the objectifying ego and its principle of noncontradiction.

4 Absolute nothingness

If there is one idea that has been associated with the Kyoto school of philosophy more than any other, it is that of absolute nothingness. Nishida's introduction of the notion into philosophy seems to have been motivated by his wish to remove traces of psychologism or idealism that were present in his early writings (see Nishida Kitarō). Rather than have recourse to the Western notion of Being as a universal that would embrace both mind and the reality that it perceives or expresses, he turned to the Buddhist idea of nothingness and moulded it into an absolute that could answer philosophical questions generated outside of the Buddhist tradition (see Buddhist concept of emptiness).

In the Western philosophical heritage, nothingness has been defined in terms of Being and for the most part made subservient to Being as a privative (see Being). Its positive side is most often caught by notions of nonexistent potentialities, ideals or aims. Wherever becoming is stressed as the correlative of being, at least a minimal idea of nothingness is present. None of this contradicts Nishida's use, but neither does it come close to exhausting it.

As an idea, the designation of nothingness or emptiness (*kū* in Japanese; in Sanskrit, *śūnyatā*) as 'absolute' is nothing new to Zen or to Mahāyāna Buddhism in general, and related ideas can be found in Daoist classics as well. However, the particular term (*zettai*) that Nishida used was a translation of the Hegelian 'Absolute', and is not found in the Chinese Buddhist canon. The term means literally 'without an opposing other', while the classical Chinese term found in Sanlun meant 'without dependency'. As a pure logical relation, Nishida's absolute does not contradict its use in Schelling and Hegel, through whom the notion of the Absolute was introduced into philosophy, although he uses it differently and does not arrive at the ontological conclusions of these latter (see Absolute, the; Hegelianism). Indeed, the notion is not ontological at all, since that would reintroduce it as a subsidiary of being; rather, it is a principle that belongs to self-awareness. It is not a being or state of being or even the absence of them; it is a transcendence of the perspective of being that is meant to forfeit none of the possibility of being found in Being.

In the Kyoto philosophers, then, the idea of nothingness is cut off - that is, 'absolutized' - from its dependency on the idea of Being and set up in its place as the supreme ontological principle. The roles are thus reversed: nothingness takes over the centrality of Being, which is always dependent on an 'other' defined in terms of nothingness. To carry out this reversal in the context of philosophy, the thinkers of the Kyoto school were obliged to turn to Eastern thought, particularly Buddhism, for guidance. In this way it developed variously in each of them, suggesting that absolute nothingness can produce the same plurality of philosophies as being is able to.

For Nishida, absolute nothing belonged to his overall logic of place, a model of ever widening concentric contexts aimed at 'locating' self-awareness in reality. The first context inverts the fundamental bias of Western logic, which focuses judgment on the substantial individual, the particular (grammatical subject) that cannot become a universal (predicate) (see Universals). For Nishida, self-awareness begins in focusing on the universal that deabsolutizes the

existence of the individual. This horizon widens when self-awareness turns attention to the basic locus of universals, which is no longer in the world of individuals but only in the nothingness of consciousness. The only ‘being’ universals retain is that of events in consciousness. Thus the need to open the perspective still further, to deabsolutize consciousness by locating it in a wider context. This final, circumferenceless circle transcends the contents of consciousness altogether and opens into a universal of all universals. This is absolute nothingness. It is not any relative thing or any relative state of consciousness, but the ultimate homeground of reality itself, beyond being and our conscious judgments about it. In the logic of place, absolute nothingness is the crowning context in which everything else is ‘located’. It is none other than the locus of enlightenment itself, and in that sense becomes part of a single philosophy the religious-experiential and the philosophical-reflective.

Like Nishida, Tanabe accepted absolute nothingness as the final ground of reality, but for him this ground was essentially dynamic. That is, it is the actual ‘working’ of absolute mediation, known not by direct intuition but through its effects, namely the interdependency of the haphazard items that make up reality as we experience it. In other words, without the world of being, absolute nothingness would be static; and the world of being can only function because it is *not* nothingness. This was for Tanabe the principle behind all transformation, both personal and historical, which is always a form of being-in-nothingness.

The absoluteness of nothingness is not merely a logical quality but, for Tanabe, points to the religious dimension of the fundamentally social nature of all reality. This religious dimension was decidedly not mystical, but always practical. Positively, this meant that for the dynamism of absolute nothingness to be truly salvific required the cooperation of the individual conscience in the social world. Negatively, it meant that the progress of virtue in human society was ultimately not the work of the members who make it up or of particular structures that govern it, but of an absolute nothingness functioning as what Pure Land Buddhism calls ‘Other-power’.

Nishitani’s work represents a major advance in the idea of absolute nothingness (see Nishitani Keiji). This appears in two principal forms. On the one hand, he drew the idea closer to its Buddhist background by relating it directly to the Buddhist idea of emptiness, particularly in its Zen formulations. Throughout his mature writing he speaks of the ‘standpoint of emptiness’ as the locus from which self-awareness comes to realization. On the other, he was at pains to locate absolute nothingness squarely in the existential struggle of contemporary consciousness. The course of this struggle is laid out most clearly in his major work, *Shūkyō to wa nani ka (Religion and Nothingness)*. There the ordinary, pre-awakened self is described as the ego of self-consciousness that sets itself up outside the world of things as a knowing subject. This is the standpoint of ‘egoity’. Driven by death and an awareness of the impermanence of all things to see the empty abyss that yawns underfoot of ordinary egoity, one awakens next to an initial sense of the vanity of self and world. This is what Nishitani calls the standpoint of ‘nihility’, whose philosopher-saint was Nietzsche (see Nihilism). It represents a conversion to a standpoint of nothingness, but only a *relative* nothingness. By facing this abyss of nihility squarely and yet not clinging to it as ultimate, we open up a final standpoint, that of ‘emptiness’ as such in which things appear just as they are, in their ‘suchness’, and in which the true self is seen to reside not in the workings of egoity but in a letting go of ego. The awareness of the relative nothingness of nihility is converted spontaneously and of its own - it is, as noted above, a *self*-awareness - to an awareness of absolute nothingness as the ground of reality.

Using this model of religious conversion, Nishitani took up key notions of self and self-nature in Zen Buddhism, as well as their conceptual counterparts in Western philosophy and theology. His reflections on the impersonality of God, the kenosis of God in Jesus, and even the immaculate conception of the Virgin Mary draw freely on the Western mystical tradition, but in the end rest on his Buddhist strategy of emptying language into self-awareness of the true self.

5 Historicity

At the time that Nishida was producing his major work, the notion of historicity most dominant in Japanese philosophy was that of Marxism (see Marx, K.; History, philosophy of). The influence of Marxist thought was perhaps mostly immediately associated with the person of Miki Kiyoshi, who enjoyed amicable relations with Nishida. Nishida did not employ Marxist-style analysis in his own thought, and seems to have considered it something of a distraction. His own position offered an alternative on two fronts. First, on the abstract level, he saw the actual production of history as the self-identity of the temporal opposites of past and future and the spatial opposites of the bodily self and the natural environment, both of which are a kind of continuity-in-discontinuity.

Second, on the concrete level, he saw the need for something like an Eternal Now or absolute present to synthesize the contradictions of history as a self-expression of reality in the making (not yet) on the one hand, and as a self-negation of reality (has been) on the other. This he gathered under the rubric of ‘tradition’, in which the specific *ethnos* of particular peoples would be made manifest, and the combination of these specific peoples would constitute a global world. In Japan’s case, this tradition centred on the Imperial Throne.

Whatever consolation the subtlety of his language afforded Nishida personally during the heavy censorship of the war years (the emperor’s place in Japanese self-consciousness was referred to as a ‘self-identity of contradictories’, omitting the qualifier ‘absolute’), it was lost on nearly everyone else, particularly on his postwar critics. At the height of the war his prestige gave him access to leading political figures, some of whom had studied under him, and although they protected him, in the end he despaired of their ability to act objectively.

Tanabe was still less drawn to struggle with Marxist thought, but more than Nishida - or at least earlier than Nishida - he struggled with bringing the element of historicity into philosophy through his notion of ‘the specific’. In the context of history, this meant that the universal (humanity) and the particular (the human individual) are related to one another concretely in the form of particular cultures, languages and ethnic particularity. For Tanabe, this specific substratum, insofar as it was functional, was nonrational; and insofar as it was unreflected, was susceptible to manipulation by rational ideologies. At the same time, even though he made it clear that the ultimate foundation of the specific was not any particular historical relative situation, but absolute nothingness, he also saw that this latter needs expression in the specific in order to be known.

More than anything Nishida had written, this idea of Tanabe’s seemed to suit Japan’s growing imperialistic ideology insofar as it stressed the specific as the most real experience of history and even regarded the emperor as a unifying symbol of its foundations in an absolute, with the result that concern for the wider human community and broader humanitarian questions tended to be relegated to a more abstract and ahistorical plane. Tanabe himself yielded to this tendency in the early years of the Pacific war. Later, as the war was drawing to a close, he publicly repented in his *Zangedō to shite no tetsugaku (Philosophy as Metanoetics)*, in which he spoke of the need to historicize philosophy by first negating it as a powerless, weak-willed enterprise that has nothing to say in time of crisis, and then reaffirming it as a letting go of self. That step having been taken, he later returned to his notion of the specific, but was never quite able to remove the shadow that the war experience had cast over this important idea for the history of philosophy.

Nishitani’s ideas on the philosophy of history developed in the opposite direction from those of Nishida. As a young professor thrust into the thick of the debate over Japan’s philosophy of ‘all-out war’, recorded in the pages of the journal *Chūōkōron* in 1941 and 1942, he had not yet worked out a mature position of his own and lapsed into remarks that seem to smack of the very nationalism that was at the heart of the militarist ideology. After the war, having suffered censorship and a temporary purge from his post in Kyoto University, he wrote a number of essays on nationalism, adopting a position close to the globalism of Nishida and the critique of the specific found in Tanabe. It was only when this was behind him, however, that he was able to forge a view of history more properly in line with his standpoint of emptiness, and without the spectre of nationalism lingering in the background. Nishitani offered Zen as a corrective to what he called the ‘optical illusion’ of Western linear notions of history, which look backwards to an absolute beginning or forward to an absolute end of history. It is only in the self-awareness of the ground of history as resting underfoot in the direct immediacy of the present that we can shake off the anthropocentric myths of scientific progress, special election and irreligious secularism that infect modern consciousness.

Whether or not the approach of the Kyoto school philosophers to Western philosophy will have the lasting, positive effect on Buddhist self-understanding they intended it to, and whether or not their revision of classical philosophical themes from a Buddhist perspective will have any lasting impact on philosophical thought outside Japan, are questions too large to pronounce on summarily. No doubt many of their ideas need to be broken down for scholars to assess and examine in further detail. However, their vision remains an important contribution, indeed the first important contribution, of Japan to world philosophy.

See also: Buddhist philosophy, Japanese; Hegelianism; History, philosophy of; Japanese philosophy; Logic in Japan; Miki Kiyoshi; Nishida Kitarō; Nishitani Keiji; Tanabe Hajime

J.W. HEISIG

References and further reading

Heisig, J. and Maraldo, J. (eds) (1994) *Rude Awakenings: Zen, the Kyoto School, and the Question of Nationalism*, Honolulu, HI: University of Hawaii Press.(A multi-faceted look at the political dimensions of the Kyoto-school philosophers.)

Nishida Kitarō (1911) *Zen no kenkyū (An Inquiry into the Good)*, trans. Masao Abe and C. Ives, New Haven, CN: Yale University Press, 1990.(First work by Nishida, the founder of the Kyoto School.)

Nishida Kitarō (1917) *Jikaku ni okeru chokkan to hansei (Intuition and Reflection in Self-Consciousness)*, trans. V. Viglielmo with Takeuchi Yoshinori and J. O'Leary, Albany, NY: State University of New York Press, 1987.(Later important work by Nishida.)

Nishitani Keiji (1985) *Shūkyō to wa nani ka (Religion and Nothingness)*, trans. J. Van Bragt, Berkeley, CA: University of California Press, 1985.(Nishitani's major work on religion.)

Nishitani Keiji (1949) *Nihirizumu (The Self-Overcoming of Nihilism)*, trans. G. Parkes and Setsuko Aihara, Albany, NY: State University of New York Press, 1990.(Nishitani's work on nihilism.)

Nishitani Keiji (1985) *Nishida Kitarō*, trans. Yamamoto Seisaku and J.W. Heisig, Berkeley, CA: University of California Press, 1991.(Nishitani's biography of his teacher.)

Ohashi Ryosuke (ed.) (1990) *Die Philosophie der Kyoto-Schule: Texte und Einführung (The Philosophy of the Kyoto School: Texts and Introduction)*, Munich: Karl Alber.(A selection of writings translated into German with brief introductions and a solid bibliography.)

Takeuchi Yoshinori (1983) *The Heart of Buddhism: In Search of the Timeless Spirit of Primitive Buddhism*, New York: Crossroad.(Exploration of Buddhist philosophy.)

Tanabe Hajime (1986) *Zangedō to shite no Tetsugaku (Philosophy as Metanoetics)*, trans. Takeuchi Yoshinori and J.W. Heisig, Los Angeles, CA: University of California Press, 1986.(Tanabe's best-known work.)

Ueda Shizuteru (1965) 'Die Gottesgeburt in der Seele und der Durchbruch zu Gott: Die mystische Anthropologie Meister Eckharts und ihre Konfrontation mit der Mystik des Zen-Buddhismus' (The Birth of God in the Soul and Breakthrough to God: The Mystical Anthropology of Meister Eckhart and its Confrontation with the Mysticism of Zen Buddhism), Ph.D. dissertation, Gütersloh.(Ueda's disseration on Meister Eckhart.)

Unno Taitetsu (ed.) (1990) *The Religious Philosophy of Nishitani Keiji*, Berkeley, CA: Asian Humanities Press. (A series of essays devoted to a critical appraisal of Nishitani's thinking.)

Unno Taitetsu and Heisig, J.W. (eds) (1990) *The Religious Philosophy of Tanabe Hajime*, Berkeley, CA: Asian Humanities Press.(A series of essays devoted to a critical appraisal of Tanabe's philosophy.)

La Forge, Louis de (1632-66)

Louis de la Forge, a medical doctor by profession, was an important champion of Cartesian philosophy in mid-seventeenth century France. Through his work on the first published edition of Descartes' Traité de l'homme, as well as in his own Traité de l'esprit de l'homme, La Forge sought to complete Descartes' project of giving a full and detailed account of the human being as a union of two essentially distinct substances: mind and body. His analysis of causation introduced occasionalist elements into his otherwise orthodox Cartesian system, and he is credited with being one of the originators of occasionalism.

Louis de la Forge was born in November 1632 in La Flèche, France. As a young man he was captivated by Descartes' philosophy. When, in 1653, he moved to the intellectually lively environment of Saumur (with its schools and academies) to set up his medical practice, he quickly became known primarily as a fervent champion of Cartesian thought. In his published works he assumed the task of faithfully completing Descartes' unfinished project of giving a full and detailed account of the human being, considered particularly as a union of two radically distinct substances: mind and body (see Descartes, R. §8). At the request of Claude Clerselier, Descartes' friend, he helped to edit and draw the figures for the first published edition of Descartes' *Traité de l'homme*, which appeared in 1664 with La Forge's own critical notes. To supplement this work (which considers only the physical functions and processes of the human body, treated in purely mechanistic terms) La Forge composed his own *Traité de l'esprit de l'homme* (1665), in which he considers the soul itself, its union with the body, and all the phenomena that derive from their mutual interaction. This gained him an international reputation as one of the originators of occasionalism (see Occasionalism). His philosophical career, however, was cut short by his death in 1666.

While La Forge was not a particularly original thinker, he was not, on the other hand, an uncritical or unimaginative disciple of Descartes. His figures and notes to the *Traité de l'homme* occasionally correct his mentor. His most important philosophical contribution lies in his discussion of causation, both as it relates to bodies and in connection with the interaction of mind and body. While modest about his accomplishments and eager to remain as faithful as possible to Descartes' doctrines, La Forge went deeper than did Descartes himself into certain metaphysical problems facing Cartesian mechanism and dualism.

La Forge begins the *Traité de l'esprit de l'homme* by considering the soul itself, including its immateriality, its cognitive and volitional functions and its immortality. He then turns to the soul's union with the body. Any union between two substances consists in a certain mutual dependence of their states, a relative correlation of action and passion. In the case of bodies, it is their local contact that establishes this dependence. In the case of minds, love accounts for the fact that one mind wills what another mind wills. A mind and a body are united when thoughts in the former depend upon motions in the latter, and vice versa. The immediate explanation of these mutual dependencies is causal. In the case of bodies and of minds, where the effect resembles the cause, it is a matter of what (employing a medieval distinction) he calls 'univocal causation'. In the mind-body case, since the effect does not resemble the cause, it is 'equivocal causation'. But La Forge then considers the nature of causality itself, and here he turns to God. *Prima facie*, causal relations between bodies appear to have an intelligibility that is lacking in the body-mind case, since we think that we can understand how one body can communicate motion to another body. In fact, genuine body-body causation is no more intelligible or possible than body-mind causation, and all species of causal relationship have their ultimate ground in the divine will.

With respect to body-body relations, La Forge adopts an occasionalist solution. A body, according to the orthodox Cartesian schema that La Forge follows, is nothing more than pure extension, devoid of any spiritual properties (thoughts, desires, will). Thus, a body is, by nature, inert and passive, and has absolutely no active motive force, no power to put itself in motion. This is particularly clear when we consider that all the properties of body must be understood in terms of its underlying essence: extension. And while this is possible for such properties as size, shape, internal figure and mobility, it is clearly not possible for active force. But if bodies cannot move themselves, then they cannot move other bodies. Moreover, the motion of a body is simply a modification of its substance, and (on the Cartesian ontology) modes cannot pass from the substance to which they belong to some other substance. Thus, real transitive causation - whereby one body would communicate its motion to another - is

ruled out.

La Forge's strongest argument on this issue is based on principles concerning divine sustenance that he inherits from Descartes (see Descartes, R. §6). God, he insists, is required not just to create the world *ex nihilo*, but to sustain it in existence from moment to moment by a kind of continuous production. This applies, as well, to all the particular substances in the world. When God creates or sustains a body at a given time, he must do so in some relative place or another. But it follows, then, that the motion of a body is just its being sustained or recreated by God from one moment to the next in different successive relative places. The force that moves a body can only be the will of God, and the local contact of a body in motion with another that is at rest is only an occasion for God to modify the motion of the first body and to put the second body into motion, as dictated by the laws of nature.

While La Forge does bring God into the picture in the other two causal contexts, it is not in such an occasionalist manner. The body plays a causal role in the generation of sensations in the mind, but only in so far as the motions communicated to the brain 'occasion' the mind to bring about a sensory idea. Thus, while the body is the occasional cause of the idea, the mind is the real efficient cause. God is required in order to institute this causal body-mind relationship in the first place (and, subsequently, to sustain it), but not to be the direct efficient cause of the mental event on the occasion of the bodily event. God has also given to the soul - an active substance - the power to move the body by its volitions, without the need of constant divine involvement. In each of these cases, La Forge appears to be entirely faithful to Descartes.

La Forge's analysis of causation was of great importance in the development of occasionalism, and some of his arguments reappear without acknowledgement in Malebranche's deeper and more systematic occasionalist doctrine (see Malebranche, N. §4). Leibniz, for one, recognized La Forge's originality here, and ascribed to him (along with Cordemoy) significant modifications to the Cartesian system.

STEVEN NADLER

List of works

La Forge, L. de (1665) *Traité de l'esprit de l'homme (Treatise on the mind of man)*, in P. Clair (ed.) *Oeuvres philosophiques*, Paris: Presses Universitaires de France, 1974.(La Forge's completion of Descartes' account of the human being.)

References and further reading

Balz, A.G.A. (1951) *Cartesian Studies*, New York: Columbia University Press. (Contains a chapter on La Forge's critique of substantial forms.)

Clair, P. (1974) 'Étude Bio-Bibliographique' (Bio-bibliographical study), in La Forge *Oeuvres philosophiques*, Paris: Presses Universitaires de France.(The most complete account available of La Forge's life and writings.)

Clair, P. (1976) 'Louis de la Forge et les origines de l'occasionalisme' (Louis de la Forge and the origins of occasionalism), *Recherches sur le XVIIème siècle* 1: 63-72.(A brief study of La Forge's role in the development of occasionalism.)

Descartes, R. (1664) *Traité de l'homme (Treatise on man)*, Paris.(A mechanist account of the human body.)

Nadler, S. (1993) 'The Occasionalism of Louis de la Forge', in S. Nadler (ed.) *Causation in Early Modern Philosophy: Cartesianism, Occasionalism, and Preestablished Harmony*, University Park, PA: The Pennsylvania State University Press.(An examination of the nature and extent of La Forge's occasionalism.)

Prost, J. (1907) *Essai sur l'atomisme et l'occasionalisme dans la philosophie cartésienne (Essay on atomism and occasionalism in Cartesian philosophy)*, Paris: Henry Paulin.(One of the best studies of occasionalism, with several chapters on La Forge's arguments and his role in the doctrine's development.)

La Mettrie, Julien Offroy de (1709-51)

La Mettrie is best known as the author of the eighteenth-century materialist manifesto, L'Homme machine (1747). His interest in philosophical issues grew out of his preoccupation with medicine, and he developed a tradition of medical materialism within the French Enlightenment. Born in St Malo, into the family of a prosperous textile merchant, La Mettrie pursued a medical career in Paris. He also studied for two years with the renowned Hermann Boerhaave in Leiden. After a brief period of medical practice, La Mettrie devoted his efforts to his translations and commentaries on Boerhaave's medical works. He also began to publish the works that made him a pariah to both the Faculty of Medicine of Paris and to the orthodox - that is, his medical satires and his first work of materialist philosophy, L'Histoire naturelle de l'âme *(1745). Because of the outrage provoked by these works, he was exiled to Holland in 1745. But L'Homme machine, the text in which he applied his materialism thoroughly and explicitly to human beings, was too radical even for the unusually tolerant Dutch, and La Mettrie was forced to seek asylum at the court of Frederick the Great where he later died. His willingness to publish ideas his contemporaries considered too dangerous led the* philosophes *to repudiate him.*

1 Medical roots of materialism

La Mettrie's best known work, *L'Homme machine* (1747), has sometimes been construed, largely because of its title, as a simple application to man of Descartes' *bête machine* hypothesis (see Descartes, R. §12). But his materialism is richer, less mechanistic, and more embedded in the scientific and medical traditions of the eighteenth century than this narrow appreciation has suggested.

La Mettrie's interest in philosophical issues grew out of his preoccupation with medicine. His medical works include translations and commentaries on Boerhaave's principal works, five medical treatises on specific diseases and public health, and seven volumes of medical satires lampooning the ignorance and venality of the medical profession. His work manifests the hostility to metaphysics and commitment to an empiricism typical of many medical writers in the eighteenth century. His awareness of the divergent manifestations of disease in different individuals led him to emphasize the importance of the physiological constitution in his philosophical works. La Mettrie built upon Boerhaave's tentative correlations between Lockean epistemology and the physiology of mental process to develop a materialist philosophy. That is to say, he not only adopted Lockean psychology as the best framework within which to discuss brain functions but also made the easy transition from Locke to materialism; in all of his philosophical writings, he emphasized the physiological evidence for taking this step.

2 Materialist philosophy

La Mettrie's first philosophical work, *L'Histoire naturelle de l'âme* (1745), appears to be a conventional metaphysical treatise in style and vocabulary. But he deliberately subverted conventional philosophical arguments to argue for materialism, perhaps attempting to legitimate materialism by placing it within the established philosophical canon (see Materialism in the philosophy of mind). For example, he claimed that Aristotelian substantial forms - defined as the power of matter to acquire form, force, and the faculty of sensation - prefigured his materialism. He used the Aristotelian sensitive soul as a fruitful way to discuss mental processes and cited, as an epistemological rallying cry, the Aristotelian maxim, 'nothing in the intellect which is not first in the senses'. He wielded his materialist reading of Locke as a weapon against seventeenth-century metaphysicians, especially Leibniz and Descartes, to argue that the soul could be completely identified with the physical functions of the body and that any claims about its existence or function must be substantiated through physiology. Without the impediment to research and understanding that the notion of the immortal soul posed, one could come to a more realistic assessment of human nature by studying all available scientific data. Writing as a physician committed to the practical application of empirical physiology to investigations of human nature and the soul, La Mettrie reappraised philosophical issues with the evidence and methods of medicine and physiology.

In *L'Homme machine* (1747), La Mettrie adopted a freer and more polemical style to canvass all the scientific issues of his day for empirical evidence for materialism. He provided lengthy and elaborate discussion of the physical basis of human behaviour in order to demonstrate that the effects of the body on the soul are so striking that one cannot reasonably assume that a soul controls the body; thus one must ultimately conclude that the soul, if the term has any meaning, must be considered part of the body. He used evidence drawn from anatomy,

physiology, and psychology to assert the complete dependence of the soul on the body. Albrecht von Haller's findings on muscular irritability were particularly attractive to La Mettrie as the most conclusive evidence against those who refused to admit that matter was capable of selfmovement, particularly the Cartesians. Making no qualitative distinction between conscious and voluntary and involuntary or instinctual, La Mettrie posited an active, organic, and selfmoving 'man machine'. With his materialist theory of man substantiated by physiological experiments, La Mettrie thoroughly compared human beings to animals, despite the implicit disparagement of conventional notions of man's place in the universe which those comparisons entailed. Furthermore, he suggested that atheism was the logical outcome of his notion of an active, selfcreating and sustaining nature.

La Mettrie's other works on the philosophy of nature, *L'Homme plante* (1748) and *Système d'Epicure* (1750), further integrate human beings in nature by comparing them to lower creatures, like plants, and by placing them in the context of the unfolding of matter and motion in an evolutionary process.

In these works, La Mettrie empowered the physician, at the expense of the theologian and metaphysician, because he maintained that the physician's physiological understanding of human beings was the most sound and most likely to yield productive results. The close connections he was able to draw between physical and mental processes demonstrated the utility of knowledge gleaned by his method, which was unsystematic, incautious and unconstrained by standards of orthodoxy, bound instead by stringent empirical standards - the method, he suggested, of the reform-minded, empiricist physician. La Mettrie then turned his attention to an area where the authority of the metaphysician or the theologian traditionally prevailed over that of the physician, ethics.

3 Moral philosophy

La Mettrie's *Discours sur le bonheur* (1750) used his materialist notion of man and his place in nature to examine the implications of materialism for moral systems and for the individual in society. He raised the question of the effect society can have on the individual or the degree to which education can countervail physiological determinants.

He sought to determine whether our notions of virtue and vice correspond to human nature as revealed by physiology. The evidence of comparative anatomy led him to conclude that society was not only unnatural but also arbitrary, and that its notions of virtue and vice, while socially useful, were fundamentally at odds with nature and simply the result of socialization. Therefore, just as the physician must acknowledge the effects of the individual constitution on health and disease, La Mettrie insists, so too must the moral reformer recognize the limits that the individual constitution imposes on one's ability to behave in the ways society has defined as virtuous. The brunt of La Mettrie's moral argument is the hope that society, recognizing that its notions of virtue and vice are relative and designed merely to further its interests, will be persuaded to reward a greater range of human behaviour; thus, more individuals will be able to aspire to social virtues. In light of his understanding of human nature and morality, La Mettrie critically appraised other moral systems; he indicted Christianity and Stoicism, in particular, for their distorted views of human nature (see Virtues and vices).

4 Role in the Enlightenment

La Mettrie was more pessimistic than other *philosophes* about the possibility of social reform; for him, reform efforts were severely circumscribed by the sway that the physiological constitution exercises over the individual. His specific claim that virtue and vice are completely relative put him at odds with the aspirations of other Enlightenment moralists writing within the natural law tradition, and his hedonistic ethic put him beyond the pale as far as many of his contemporaries were concerned. But despite the fact that virtually all of the*philosophes*, even those like Diderot and d'Holbach who were indebted to him, found it dangerous to be associated with his radical materialism, La Mettrie himself was eager to proclaim his adherence to them. In his last philosophical work, the *Discours préliminaire*, written in 1751 just as the philosophic movement was beginning to coalesce around the *Encyclopedia*, La Mettrie claimed to speak for the *philosophes*. He explicitly identified his work in both medicine and philosophy with their concerns for reform. Where medicine offered hope for a more naturalistic understanding of human nature, the *médecin-philosophe* (a term he coined) might be able to reform social institutions in accord with that understanding. La Mettrie's medical materialism is his distinctive contribution to the French Enlightenment and the history of philosophy.

See also: Enlightenment, Continental; Medicine, philosophy of

KATHLEEN WELLMAN

List of works

La Mettrie, J.O. de (1747-51) *Oeuvres philosophiques (Philosophical Works)*, Berlin: Etienne de Bourdeau, 1750, 3 vols; repr. Paris: Fayard, 1987.(Contains all of the major philosophical works.)

La Mettrie, J.O. de (1747) *L'Homme machine*, trans. and ed. A. Thomson, *Man Machine and other writings*, Cambridge: Cambridge University Press, 1996.

References and further reading

Thomson, A. (1981) *Materialism and Society in the Mid-Eighteenth Century: La Mettrie's 'Discours preliminaire'*, Geneva: Librairie Droz.(Critical edition of text with monographic introduction. Advanced introduction to philosophical context.)

Vartainian, A. (1960) *La Mettrie's L'Homme Machine: A Study in the Origins of an Idea*, Princeton, NJ: Princeton University Press.(Critical edition of text with monographic introduction.)

Wellman, K. (1992) *La Mettrie: Medicine, Philosophy, and Enlightenment*, Durham, NC: Duke University Press. (Intellectual biography.)

Labriola, Antonio (1843-1904)

Antonio Labriola was the founder of Italian theoretical Marxism. Generally situated in the Marxism of the Second International, he was more questioning than others in that movement. He profoundly influenced the development of Italian thought, constantly challenging the influential idealism of Benedetto Croce and Giovanni Gentile. His attempt to maintain a place for human creativity within a deterministic Marxist view of history influenced Antonio Gramsci and helped give Italian Eurocommunism its distinctive flexibility. His concepts of 'genetic method', 'social morphology', 'philosophy of praxis' and 'social pedagogy' are indications of this attempt.

Labriola was born in Cassino (near Naples) and studied at the University of Naples. He became Professor of Moral Philosophy and Pedagogy at Rome University in 1874 and remained there until his death. In 1887 he gave lectures in the philosophy of history and two years later began a series of lectures on the French Revolution which had to be suspended when right-wing students demonstrated.

Labriola was taught by Bertrando Spaventa, leader of the Neapolitan Hegelians, who was, however, critical of Hegel's deterministic interpretations (see Hegelianism §4). Labriola adopted a similar position and began to borrow the social psychology of Johann Friedrich Herbart while eschewing his metaphysics. Later, under the impact of reading Karl Marx, Herbart's social psychology was reconceived in terms of class consciousness.

Once he perceived that the Italian middle class in the 1870s were incapable of creating an Italian state which could provide unifying cultural leadership, Labriola rejected the Hegelian notion of the 'Ethical State' (see Hegel, G.W.F. §8) and came to regard class conflict as necessary to social advancement. He leaned towards socialism from 1879, and in 1889 discovered Marx's writings, which were almost unknown in Italy at the time. Everything he had previously learned was now reshaped in Marxist categories.

He corresponded extensively with Friedrich Engels, as well as with Eduard Bernstein and Karl Kautsky attempting to engage Italy with German social democracy. He expounded Marxism in four highly original essays. His third and most insightful essay is composed from correspondence with Georges Sorel, and his final essay was incomplete at his death. Outside Italy his work was ignored, except in Russia, where his essays were applauded by Plekhanov, Lenin and Trotsky. Labriola's work generally re-entered histories of philosophy only as a result of widespread interest in Gramsci's prison notebooks.

Labriola introduced the concepts of 'genetic method', 'social morphology' and 'philosophy of praxis' in an attempt to characterize Marx's method in such a way as to avoid, on the one hand, a rigid materialist determinism and, on the other, the risk of appropriation by Italian Hegelianism. The three concepts are not really new, however, since they correspond to 'materialist method', 'social structure' and 'Marx's dialectic' in the works of Engels and others (see Engels, F. §3).

He saw Marx as having developed a systematic study of specific societies and of their modes of thought as distinctive and coherent wholes. Such study was carried out with the genetic method, which involved laying bare the complex of particular conditions (social morphology) which gives rise to the whole. Although Labriola wanted an empirical Herbartian approach, and was reluctant to speak of a dialectical method, his genetic method closely approximates the method promoted by Engels (see Dialectical materialism).

Labriola took the philosophical kernel of Marxism to be 'the philosophy of praxis', that is, 'the reversal of the Hegelian dialectic', so that thought must be understood in terms of the practice in which it is embedded. For him this implied an entirely novel way of looking at things. Although he understood that thought could shape practice, he failed to avoid the determinism inherent in a metaphysics of history which insisted on the inevitability of the proletarian revolution. It was this same naturalistic evolutionism (albeit distinct from social Darwinism, which he criticized) which led him to advocate Italian colonialism in Africa as a means of raising Italy to the stage necessary for revolution (see Evolutionary theory and social science).

Prior to his adoption of Marxism, Labriola for some years held a government position with responsibilities for the direction of public education. He came to prefer lecturing to workers. He always emphasized progressive popular education ('social pedagogy') and the creation of the cultural conditions for a society more civilized than

capitalism, regarding Marxism as having a central role as a new conception of the world.

Gramsci took his Marxism in large part from Labriola, but rejected his evolutionism. He tried to resolve the methodological difficulty Labriola had encountered with deeper explorations of ‘praxis’ and ‘social pedagogy’.

See also: Marxism, Western

GEOFFREY HUNT

List of works

Labriola, A. (1959) *Opere complete di Antonio Labriola (Complete Works)*, ed. L. Dal Pane, Milan: Feltrinelli. (The definitive collection of Labriola’s writings.)

Labriola, A. (1895) *In memoria del Manifesto dei comunisti* (In Memory of *The Communist Manifesto*), Rome. (This first essay presents the theory of historical materialism as a product of its time.)

Labriola, A. (1896) *Del materialismo storico: Dilucidazione preliminare (Historical Materialism: Preliminary Exposition)*, Rome.(Labriola’s analysis of ideologies and the structure-superstructure relation.)

Labriola, A. (1898) *Discorrendo di socialismo e di filosofia (Discourse on Socialism and Philosophy)*, Rome; 3rd essay, trans. in *Socialism and Philosophy*, trans. and with intro. by P. Piccone, St Louis, MO: Telos Press, 1980.(This innovative essay examining class consciousness comprises letters to Georges Sorel. The translation contains a useful long introduction.)

Labriola, A. (1903) *Essays on the Materialist Conception of History*, trans. C.H. Kerr, Chicago, IL.(The first English translation of Labriola’s first three essays.)

Labriola, A. (1904) *Da un secolo a un’altro (From One Century to the Next)*, Rome. (Unfinished essay, published posthumously, mainly about the ‘liberal age’ and imperialism.)

Labriola, A. (1976) *Scritti filosofici e politici (Philosophical and Political Writings)*, ed. F. Sbarberi, Turin: Einaudi.(Contains all four Marxist essays, plus many articles and letters in political philosophy. Good bibliography of Italian editions.)

References and further reading

Hunt, G. (1987) ‘Antonio Labriola, Evolutionist Marxism and Italian Colonialism’, *Praxis International* 7 (3, 4): 340-59.(Draws out the connections between Labriola’s method and his espousal of Italian colonialism.)

Mastroianni, G. (1976) *Antonio Labriola e la filosofia in italia (Antonio Labriola and Philosophy in Italy)*, Urbino: Argalia.(A short survey of the entire trajectory and influence of Labriola’s thought.)

Ragionieri, E. (1976) *Socialdemocrazia tedesca e socialisti italiani: 1875-1895 (German Social Democracy and Italian Socialists)*, Feltrinelli: Milan.(The best work for relating Labriola’s thought to the Second International.)

Lacan, Jacques (1901-81)

Jacques Lacan was a French psychoanalyst and philosopher whose contribution to philosophy derives from his consistent and thoroughgoing reinterpretation of Freud's writings in the light of Heidegger and Hegel as well as structuralist linguistics and anthropology. Whereas Freud himself had disparaged philosophical speculation, claiming for himself the mantle of the natural scientist, Lacan demonstrates psychoanalysis to be a rigorous philosophical position. Specifically, Lacan suggests that the Freudian unconscious is best understood as the effect of language (what he calls, 'the symbolic') upon human behaviour.

Lacan gives the name, 'unconscious' to what necessarily disrupts conscious life. In order to remain true to this disruptive effect, Lacan rejects the falsely reassuring clichés with which the idea of the unconscious has been popularized. He develops an understanding of the unconscious (doubtless strongly influenced by Hegel) that is based upon the difference implied in every identity - a difference conceivable as the 'gap' between the subject and any object that provides an identity to it. Since, according to Lacan, any such space or gap in some sense *belongs to* the object, identification is always imperfect. The effects that Freud associates with the unconscious derive from this imperfection.

Lacan developed his theory of the unconscious in three historical stages and corresponding to three 'fields' that define subjectivity - 'imaginary', 'symbolic' and 'real'. Lacan's major exposition of the idea of the imaginary came (while he was still in training as a psychoanalyst) in the 1936 paper later published as 'The Mirror Stage as Formative of the Function of the I' (in Lacan 1966). In that paper, Lacan posited an imaginary field dramatized by events of mirror identification (with 'images') in pre-verbal infants. The prematurity of human infants - their incapacity, compared with the young of other species, to control their own actions - creates the need for an external model for intentional unity; such is Lacan's explanation of the 'jubilation' of pre-verbal children from 6 months old onwards upon discovering their image either in a mirror or, more typically, in the mimicking gestures of another person.

But the same image that provides the site for jubilation also becomes the object of what Lacan calls 'aggressivity' - a proto-agression directed at the object of identification. For Lacan, such aggressivity results from the 'alienation' of that locus - the ego - where imaginary identification forms intentional unity. That is, in Lacanian theory the ego is first of all an object, the object in which the subject finds a 'self.' Because the ego can only appear as an object - as an other - its emergence also, according to Lacan, occasions the formation of a primordial violence directed against it.

Lacan's work in the 1950s, starting with the so-called, 'Rome Discourse' - the essay later published under the title of 'Function and Field of Speech and Language in Psychoanalysis' (in Lacan 1966) - elaborates a new dimension in his thought, the symbolic. Lacan's immersion in the structuralist linguistics of Ferdinand de Saussure and anthropology of Claude Lévi-Strauss helps him to reveal the alternative to the imaginary. Whereas the imaginary offers the exclusive either/or of obsessive identity through representation or aggressive destruction of the object of identity, the symbolic partially satisfies both impulses.

The symbolic is able to have such success based upon the structure of the sign. If, for Saussure, the sign combines a material element (the signifier) and the 'idea' which it indicates (the signified), then Lacan sees this structure as representing precisely the split condition of all human representation. That is, because the sign can only make present a signified through the mediation of a signifier, there is an absence in every presence; the idea can only be present through what represents it but is not it. Moreover, to speak a language is implicitly to accept this limitation. It is to partially master absence - the absent signified is always to some extent present - but at the price of the implicit admission that there can be no transparent presence. Language is a representation, then, that always represents the necessity of alienation (in the 'bar' dividing signifier and signified) along with a specific object or state of affairs. Lacan's reinterpretation of the Freudian Oedipal and castration dramas as the acceptance of this prohibition implicit in language (for example, the 'Name-of-the-Father' for the father) allows him to divide off the Freudian developmental schema from any culturally specific view of the family.

To address the symbolic more concretely demands one more observation. If the bar between signifier and signified

represents to the subject its own ‘gap’ or ‘want-to-be’ (‘*manque-à-être*)’, then it does so in a movement, the process that Lacan identifies with ‘desire’. Within desire, language concretely demonstrates the impossibility of complete, unmediated presence. It does so by producing an ongoing process of ‘identifications’ mediated by language and always presenting themselves, therefore, as inadequate. Desire involves an ongoing movement which destroys identity (the ego) in favour of a metonymic ‘sliding’ from one signifier to another - but always by producing a new identity in the place of the old one.

It is this sliding that Lacan valorizes over and against the obsessive identification or aggressive demolition allowed by the imaginary. Lacan thus finds the efficacy of psychoanalysis in its ability to release desire from the limitations of the imaginary. It is possible to see how far a Freudianism based upon such an ‘ethics’ of pure desire, departs from almost all psychoanalytic orthodoxy. For most Freudians the purpose of analysis is at least partly to strengthen the ego. For Lacan the purpose of analysis is to allow the subject’s desire to overcome the imaginary grip of the ego - to make identity fluid (see Alterity and identity, postmodern theories of §2; Subject, postmodern critique of the).

In the 1960s and 1970s, Lacan’s work concentrates on Freud’s least accepted thesis. As Boothby (1991) shows, Lacan takes the Freudian hypothesis of a ‘death drive’, strips it of its biological trappings and then places it at the centre of psychoanalysis: what the dynamism of symbolic desire sustains within the subject’s world is precisely the release of those energies which, within the imaginary, could only gain expression through aggression - energies that necessarily evade the stasis of imaginary representation. The symbolic sublimates these energies, the energies of the death drive, by putting them to work in the process by which language erodes identity. With these ideas, Lacan devotes his latest work to what he calls the ‘impossible real’, to that which is forbidden entry into representation.

See also: Freud, S.; Psychoanalysis, post-Freudian §3; Structuralism; Structuralism in social science §3

THOMAS BROCKELMAN

List of works

Lacan, J. (1966) *Écrits*, Paris: éditions du Seuil; selections trans. A. Sheridan, *Écrits: A Selection*, New York: W.W. Norton, 1977.(Lacan’s major collection of writings, while containing such seminal pieces as the essay on the ‘mirror stage’ and the ‘Rome Discourse’, is extraordinarily difficult. Beginners are advised to start with ‘Le Séminaire’.)

Lacan, J. (1974) *Télévision*, Paris: Éditions du Seuil; trans. D. Hollier, R. Krauss, A. Michelson and J. Mehlman, *‘Television’ and ‘A Challenge to the Psychoanalytic Establishment’*, ed. J. Copjec, New York: W.W. Norton, 1990. (Texts relating to the institutionalization of psychoanalysis including docments from Lacan’s secession from the Societé Psychoanalytique de Paris and his foundation of the École Freudienne de Paris.)

Lacan, J. (1975-94) *Le Séminaire de Jacques Lacan*, texte établi par Jacques-Lacan à Jacques-Alain Miller, vol. 1, *Les Écrits techniques de Freud*, vol. 2. *Le Moi dans la théorie de Freud et dans la technique de la psychanalyse*, vol. 3, *Les Psychoses*, 1955-6, vol. 4, *La Relation d’objet*, 1956-7, vol. 7, *L’éthique de la psychanalyse*, 1959-60, vol. 8, *Le Transfert*, 1960-1, vol. 11, *Les Quatre Concepts fondamentaux de la psychanalyse*, vol. 17, *L’Envers de la psychanalyse*, 1969-70, vol. 20, *Encore*, Paris: Éditions du Seuil; vol. 1 trans. J. Forrester, *Freud’s Papers on Technique*, vol. 2 trans. S. Tomaselli, *The Ego in Freud’s Theory and in the Technique of Psychoanalysis*, 1954-5, vol. 3 trans. R. Grigg, *The Psychoses*, 1955-6, vol. 7 trans. D. Porter, *The Ethics of Psychoanalysis*, all ed. J.-A. Miller, New York: W.W. Norton.(Transciptions of Lacan’s weekly public teaching seminar which ran, with a few breaks, from 1953 until 1979-80. Far more accessible for the beginner than *Écrits*.)

Lacan, J. (1980) *De la psychose paranoïque dans ses rapports avec la personnalité (On Paranoid Psychosis in its Relationship with the Personality)*, Paris: Éditions du Seuil.(Lacan’s 1932 doctoral thesis in psychiatry.)

Lacan, J. (1981) *The Four Fundamental Concepts of Psycho-Analysis*, trans. A Sheridan, New York: W.W. Norton.(Lacan’s seminar for 1963 in an early translation.)

Lacan, J. and the École Freudienne (1982) *Feminine Sexuality*, New York: W.W. Norton.(Late writings on feminine sexuality by Lacan and various disciples.)

References and further reading

Boothby, R. (1991) *Death and Desire: Psychoanalytic Theory in Lacan’s Return to Freud*, New York: Routledge. (Boothby’s approach provides an excellent introduction to Lacan; by using the Freudian death drive, he ties

together all of Lacan’s work in an extremely convincing and clear way.)
Weber, S.M. (1991) *Return to Freud*, Cambridge and New York: Cambridge University Press.(Weber’s introduction to Lacan, translated from lectures originally delivered in German, locates his work within the context of French post-structuralism.)
Zizek, S. (1991) *For They Know not What They Do: Enjoyment as a Political Factor*, London and New York: Verso.(Zizek, perhaps the most brilliant interpreter of Lacan today, uses Lacan as a tool for interpreting political ideology.)

Lachelier, Jules (1832-1918)

Lachelier is, along with Octave Hamelin, one of the foremost French metaphysicians of the nineteenth century. An idealist, he was a major figure in the neo-spiritualist movement, which opposed the materialism dominating scientific thought at the time. For all that, Lachelier did not set up a division between consciousness and life, but, following the example of Maine de Biran and Ravaisson, he rather saw in mind the inwardness of life, at work in the whole of nature.

As a student, Lachelier graduated top in the *agrégation* in philosophy, and in 1871 he defended his doctoral theses, 'Du fondement de l'induction' ('On the Basis of Induction') (1871) and 'De natura syllogismi' ('On the Nature of the Syllogism'). In the former, he showed that the real inductive process assumes not only the principle of causality but also a principle of persistence for essences, which are perceived through the individual, and which orientate themselves in relation to one another in a hierarchy. 'Harmony is in some way the supreme interest of nature' ([1902] 1871: 70). Nature cannot be completely external to thought, since then it would be as if it did not exist for us. It is therefore necessary to find some way of both 'making thought real, and making nature intelligible' ([1902] 1871: 74).

The article 'Psychologie et métaphysique' ('Psychology and Metaphysics') (1885) is perhaps Lachelier's greatest work, in which he goes far beyond the dogmatism of Victor Cousin. He begins by remarking that to be extended and to perceive extension are two distinct things, and neither of them is sufficient to define a consciousness. Extension appears to be outside us, but this externality is our work. Extension cannot exist in itself, because its parts, which are infinitely divisible, would be without continuity if separated from thought. Extension is 'a whole, whose parts divide but do not constitute it' (1902 [1885]: 188). Its continuity comes to it from the mind, and the perception of depth, for example, is nothing more than 'a spontaneous product of our thought'; it represents a beyond.

Consciousness is therefore primary. Further, instead of saying with the materialists that it progresses from perception to will, the contrary is true, in that it descends from will to perception. In sensation, there is always something affective. Sensibility senses itself; at the same time, it feels some quality. Thus, the subject is given to itself, and distinguishes itself from the external world.

However, we cannot say that our affections are ourselves. We ourselves *make* our feelings through our own love or hatred. The awareness of each feeling includes the operation of some tendency and ultimately the will to live. And 'if we do not feel our will, this is not because it is nothing, but because it is ourselves' (1902 [1885]: 194). Will is the principle and the hidden source of everything. Man is the only being which fixes qualities in extension and composes 'this permanent mirage which he calls the external world' (1902 [1885]: 195). But the pure will - which is closer to that envisaged by Schopenhauer than Maine de Biran (§3) - goes beyond individual existence. The self is the will refracted in the affective state whose form expresses one's character. It is found only to be lost.

From psychological analysis, Lachelier moves towards synthesis, where the primary place is given to the idea of truth. Even at the level of physical pain, it is necessary to say that 'consciousness of pain is not painful, but true' ([1902] 1885: 201). This is an elegant idea, with a hint of stoicism, for which Lachelier has often been criticized, since if the feeling of pain is not painful, how can it be true, that is to say, conform to its object? Moreover, it is above all the property of pain to be essentially present, so that no memory or idea can affect it; it is an existential shock that one also finds at the height of physical joy. But Lachelier wants to say that psychological consciousness is based on an *intellectual* consciousness, which gives it the seal of objectivity. Therefore consciousness resides in the final analysis in the truth of being, and the idea of being, being itself of being, can construct itself, from the sphere of logical necessity right up to that of living things.

The final word is given to a dialectic of liberty, through which the being that supports us produces itself. The subject is not an act, but a form, and Lachelier eloquently puts away 'any idea of a special and mysterious subject, the transcendent I in the sensible consciousness. Such a subject would in fact be only one more object', 'To know something one must in some manner *be* that very thing, and to accomplish this one must not in the first instance be other than oneself' ([1902] 1885: 206). Besides, if by some supernatural act the self of another person was put in

place of one's own, it would be absolutely impossible for us to see this. In short, individuality is based on the simple connection of memories. On the other hand, the 'light of thought' is infinitely greater than the individual self and escapes its grasp, because it is of the order of having to be rather than being. The progress of thought ends when, after seeking for itself in necessity as in its shadow, then in the will as in its body, it finally finds itself in freedom. There is not a fourth idea of being, because there is no fourth dimension of extension.

Translated by Robert Stern

MICHEL PICLIN

List of works

Lachelier, J. (1902) *Oeuvres*, vol. 1, Paris: Alcan, 4th edn.(The standard complete works, containing 'Du fondement de l'induction' ('On the Basis of Induction') (1871), 'Psychologie et Métaphysique' ('Psychology and Metaphysics') (1885) and 'Note sur le pari de Pascal' ('Note on Pascal's Wager') (1907).)

Lachelier, J. (1907) *Oeuvres*, vol. 2, Paris: Alcan, 4th edn.(This second volume of the complete works contains 'Études sur le syllogisme' ('Studies on the Syllogism') (1876, 1906), 'L'observation de Platner' ('The Observation of Platner') (1903) and 'Note sur le Philèbe' ('Note on Philebus') (1902).)

Lachelier, J. (1960) *The Philosophy of Jules Lachelier*, The Hague: Martinus Nijhoff.(An English translation of Lachelier's main works.)

References and further reading

Bouglé, C. (1937) *Les maîtres de la philosophie universitaire en France (Masters of Academic Philosophy in France)*, Paris: Maloine.(A penetrating book in the mainstream university tradition.)

Duméry, H. (1957) *Philosophie de la religion (Philosophy of Religion)*, Paris: Presses Universitaires de France.(An interesting and accessible work, written by a specialist in religious thought.)

Jolivet, R. (1934) 'Lachelier et l'idéalisme contemporain' ('Lachelier and Contemporary Idealism'), *Revue thomiste* (July-October).(A thorough article, but more ready to criticize than to understand.)

Millet, L. (1959) *Le Symbolisme dans la philosophie de Lachelier (Symbolism in the Philosophy of Jules Lachelier)*, Paris: Presses Universitaires de France.(A relatively recent and important study, which locates Lachelier in what is called 'reflexive' French philosophy.)

Séailles, G. (1920) *La philosophie de Jules Lachelier (The Philosophy of Jules Lachelier)*, Paris: Alcan.(A basic study by a thinker under the influence of Ravaisson.)

Lacoue-Labarthe, Philippe (1940-)

Philippe Lacoue-Labarthe is Professor of Philosophy at the universities of Strasbourg and Berkeley. At the centre of his thought is philosophy's ostracism of literature, which in his view characterizes the foundational scene of philosophy itself. Lacoue-Labarthe demonstrates how all Western thought, including Heidegger (perhaps its most faithful deconstructor) lies within a conception of mimesis which is still metaphysical in that it remains bound to the opposition between truth and mimesis: an 'imitation' alters (or falsifies) its original, thus contrasting with that which is 'true'. The non-metaphysical thought of mimesis proposed by Lacoue-Labarthe, by contrast, shows how a mimetic aspect structures the concept of truth itself, making it impossible to distinguish between truth and verisimilitude, model and copy. Thus a new area of investigation is laid open concerning fictional modes of thought: the manner in which myths and figures are produced.

Lacoue-Labarthe's thought is largely a rereading of the history of philosophy, from the Greeks to the present day. In Plato's *Republic*, the affirmation of philosophy's superiority runs parallel to a devaluing of literature (*mythopoïesis*, telling of myths). Mythopoetical activity is devalued because *mimēsis* (see Mimēsis), imitation, pertains, not to truth, but to verisimilitude (myths are fiction, falsifications of the original). Such activity, Lacoue-Labarthe argues, necessitates a passive, and so unfree, attitude on the part of the subject.

A similar bias against literature characterizes the whole Western tradition, at least up to what Lacoue-Labarthe calls the 'Romantic moment', which continues down to the present. With early German Romanticism, literature (poetry, in particular) starts to be re-evaluated as a place of truth. Traces of this re-evaluation can be found in Hölderlin, Heidegger and Celan (see Hölderlin, J.C.F.; Romanticism, German). Lacoue-Labarthe's own thought is situated, in a rather peculiar way, inside this 'anti-Platonic' tradition: hence his interest in early German Romanticism (*L'Absolu littéraire*, 1978, with Jean-Luc Nancy), idealism, Nietzsche and Heidegger.

In this, Lacoue-Labarthe attributes a particularly important role to Heidegger, whose thought constitutes the most extreme and complete attempt at deconstruction of the metaphysical tradition. Such thought was also responsible, however, for Heidegger's calamitous association with National Socialism in 1933. Many of Lacoue-Labarthe's works seek to explain this unsettling combination of totally different aspects in Heidegger's thought.

First, in the key work *Typographie* (1975) (*Typography*, 1989), Lacoue-Labarthe demonstrates how in Heidegger a Platonic conception of *mimēsis* - thus truth - still persists and is to a large extent undisputed. Truth is, in fact, conceived as a static determination of Being, the erection of a 'pedestal', or 'stele' (translation of *Ge-stell*, a key term in Heidegger's later works). For this reason Lacoue-Labarthe proposes setting the term *onto-theo-logy* (used by Heidegger to indicate Western metaphysics) side by side with the term *onto-typo-logy*, which would characterize the tradition of Western metaphysics *and*, at least partially, Heidegger's own thought. The concept of *onto-typo-logy* restates Heidegger's 'ontological difference' in light of the question about *mimēsis*: Being is considered in the fixed quality of a *typos* (type, seal, stamp) from which beings originate as traces or imprints which resemble the *typos* itself. This means that Heidegger thinks, as is the case in the Platonic model, that the mimetic relationship between Being and beings corrupts the truth of Being.

Second, in *La Fiction du politique* (1988) (*Heidegger, Art and Politics: The Fiction of the Political*, 1989), *L'Imitation des modernes* (1986a) as well as, in other, new ways, L*a Poésie comme expérience* (1986b, dedicated to Paul Celan), and *Musica ficta (figures de Wagner)* (1991) (*Musica Ficta: Figures of Wagner*, 1995), Lacoue-Labarthe demonstrates how the figural immutability characterizing thought in general is again found, with notable force, in Heidegger's reflection on art after 1933, in which the work of art is thought of above all as shape and support (*Gestalt* and *Ge-stell*) of truth. This way of thinking reveals, according to Lacoue-Labarthe, the 'mythical' nature of National Socialism. For, as he argues in 'The Nazi Myth' (1989, with Jean-Luc Nancy), Nazi ideology repeats the constitutive traits of the mythic process. By means of the creation of great ideal figures (leader as figure *par excellence*), it attempts to respond to the demand for identity on the part of the German people: the community identifies itself with the figure and fashions itself as would a work of art. In reality, National Socialism is, for Lacoue-Labarthe, a National *Aestheticism*. Entirely apart from his adhesion to the Nazi Party in 1933, then, Heidegger's thought would appear to demonstrate both a pronounced proximity *and* a total distancing from National Socialism.

Lacoue-Labarthe's complex relationship with Heidegger places him in a position within post-Romantic and anti-Platonic thought which is unique in three ways. First, he opens a new conception of *mimēsis*, according to which it does not corrupt truth but rather is co-originating. The two conceptions of truth pointed out by Heidegger are articulated together: *aletheia*, truth as 'revealing', always tends towards *adaequatio*, truth as adequation, but each time finds itself only to be inadequate to itself. Second, he presents a new image of the subject. In essays such as 'L'Écho du sujet' (in *Le Sujet de la Philosophie*, 1979) and 'Le Titre de la lettre' (1973, with Jean-Luc Nancy), Lacoue-Labarthe confronts psychoanalysis, which itself is based on a mimetological mythography (for example, Oedipus). Against the speculative dream of a subject capable of theorizing about its own conception, Lacoue-Labarthe sets up an image of a subject which is constitutively incomplete. The truth of the subject is captured by its verisimilitude: by the inexhaustible tension of resembling itself, which must always surrender and fail. The subject can therefore never become a figure, because every figure of the origin of the subject falls back to an engendering of the figure itself. This deconstruction of the concepts of origin and dichotomous conceptuality draws Lacoue-Labarthe's thought close to that of Jacques Derrida. Third, he produces an esthetic and ethical-political position which is unique, though in certain aspects close to that of Maurice Blanchot. In Lacoue-Labarthe's view, the incompleteness of the subject tends towards a state of impotence which must not be thought of as passivity. It is rather more a *désistance*, a stoic, active existence/resistance. The human being, by actively protecting its own passivity and the lack of a figure capable of identifying it, can thus oppose itself to any form of totalization and totalitarianism.

See also: Heidegger, M.; Nancy, J.-L.

Translated by William Finley Green
GIOVANNI SCIBILIA

List of works

Lacoue-Labarthe, P. (1975) *Typographie*, Paris: Aubier-Flammarion; trans. *Typography: Mimesis, Philosophy, Politics*, Cambridge, MA: Harvard University Press, 1989.(Collection of essays on mimesis, art and politics, Nietzsche, Heidegger, Diderot, Hölderin and others.)

Lacoue-Labarthe, P. with J.-L. Nancy (1978) *L'Absolu littéraire: théorie de la littérature du romantisme allemand*, Paris: Éditions du Seuil; trans. *The Literary Absolute: The Theory of Literature in German Romanticism*, Albany, NY: State University of New York Press, 1988.(Anthology of writings on literary theory in German Romanticism - in particular from the *Atheneum* - with full introdcution and comments.)

Lacoue-Labarthe, P. (1986a) *L'Imitation des modernes (Typographies 2)*, Paris: Galilée; trans. in *Typography: Mimesis, Philosophy, Politics*, Cambridge, MA: Harvard University Press, 1989.

Lacoue-Labarthe, P. (1986b) *La Poésie comme expérience*, Paris: Christian Bourgois.

Lacoue-Labarthe, P. (1987) *La Fiction du politique*, Paris: Christian Bourgois; trans. C. Turner, *Heidegger, Art and Politics: The Fiction of the Political*, Oxford: Blackwell, 1989.(Heideggerian allegiance to national socialism and what it shows about modern politics; modern politics is based on the imitation of the classical ideal ('the Greek model') and on art and myth insofar as they are the means of a people's cultural identity.)

Lacoue-Labarthe, P. with Nancy, J.-L. (1989) 'The Nazi Myth', *Critical Inquiry* 16 (2): 291-312.; trans. into French as *Le Mythe nazi*, La Tour d'Aigues: Éditions de l'Aube, 1991.(Nazi ideology as a process of the constitution of the myth, an extreme symptom of the treatment of the political as aesthetics.)

Lacoue-Labarthe, P. (1973) *La titre de la lettre: une lecture de Lacan*, Paris: Galilée; trans. *The Title of the Letter: A Reading of Lacan*, Albany, NY: State University of New York Press, 1992.(A deconstructive reading of Lacan.)

Lacoue-Labarthe, P. (1979) *Le Sujet de la philosophie*, Paris: Aubier-Flammarion; trans. *The Subject of Philosophy*, Minneapolis, MN: University of Minnesota Press, 1993.(Collection of essays on literature and philosophy, music and philosophy, and Nietzsche, Heidegger and Reick.)

Lacoue-Labarthe, P. (1991) *Musica ficta (figures de Wagner)*, Paris: Christian Bourgois; trans. F. McCarren, *Musica Ficta: Figures of Wagner*, Stanford, CA: Stanford University Press, 1995.(Four readings of Wagner (Baudelaire, Mallarmé, Heidegger and Adorno) show how the political can be treated as aesthetics and explain the split between Nietzsche and Wagner.)

References and further reading

Derrida, J. (1989) 'Désistance', in P. Lacoue-Labarthe, *Typography: Mimesis, Philosophy, Politics*, Cambridge, MA: Harvard University Press.(A rereading of Lacoue-Labarthe's thought starting from a term (désister) which

alludes to a condition of dispossession of the subject.)

Lakatos, Imre (1922-74)

Imre Lakatos made important contributions to the philosophy of mathematics and of science. His 'Proofs and Refutations' *(1963-4) develops a novel account of mathematical discovery. It shows that counterexamples ('refutations') play an important role in mathematics as well as in science and argues that both proofs and theorems are gradually improved by searching for counterexamples and by systematic 'proof analysis'. His 'methodology of scientific research programmes' (which he presented as a 'synthesis' of the accounts of science given by Popper and by Kuhn) is based on the idea that science is best analysed, not in terms of single theories, but in terms of broader units called research programmes. Such programmes issue in particular theories, but in a way again governed by clear-cut heuristic principles. Lakatos claimed that his account supplies the sharp criteria of 'progress' and 'degeneration' missing from Kuhn's account, and hence captures the 'rationality' of scientific development. Lakatos also articulated a 'meta-methodology' for appraising rival methodologies of science in terms of the 'rational reconstructions' of history they provide.*

1 Life

Imre Lakatos was born Imre Lipschitz in Hungary in 1922. His early life was turbulent even by the remarkable standards of the time. He was a member of the resistance during the Second World War, fortunately evading arrest (unlike his mother and grandmother, both of whom were murdered in Auschwitz). After the war he pursued a political career and by 1947 had become a powerful figure within the Hungarian Ministry of Education; in 1950, however, he was arrested and spent over three years in a Stalinist jail. Informed of the likelihood of rearrest, he fled in 1956 to Vienna, and eventually to Cambridge, where he studied for a (second) doctorate under the supervision of R.B. Braithwaite.

In the course of these studies, he became a regular attender of Karl Popper's seminar at the London School of Economics, and Popper's thought and approach had a major influence on him. Lakatos was appointed to a Lectureship at the LSE in 1960, and spent the rest of his life there, being awarded a personal chair in 1970. For unrevealed reasons, the British Home Office rebuffed two impressively supported attempts to obtain British citizenship and Lakatos remained officially 'stateless'. He died suddenly, from a heart attack, in February 1974. He had a vivid personality, strong political views and a sharp wit: he inspired intense loyalty and opposition in roughly equal measures.

2 'Proofs and Refutations': contributions to philosophy of mathematics

Lakatos published some philosophical articles (mostly book reviews) before coming to the West, and some commentators discern the influence of his Marxist and Hegelian education throughout his career, but his significant contributions to philosophy were all made after arriving in the West. His doctoral studies at Cambridge eventually formed the basis for his 'Proofs and Refutations' (originally published as a series of four journal articles in 1963-4, and in book form with further material only posthumously in 1976). Perhaps his major work, it consists of an imaginary discussion between a teacher and a group of his (frighteningly bright) students. The first part reconstructs the history of the attempts to prove the Descartes-Euler conjecture about polyhedra (that the number of vertices minus the number of edges plus the number of faces is equal to two for any polyhedron). The real history is told in the many footnotes. A second part of the discussion reconstructs the discovery of uniform convergence as the result of the 'refutation' of one of Cauchy's results. Aside from its philosophical and historical interest, the dialogue is a literary *tour de force*.

Lakatos argued that the standard picture of the development of mathematics is seriously faulty. On that picture, either an assertion is conjectured to hold and after a time a proof of it is produced, or mathematicians simply set out to 'prove' from some agreed axiomatic basis, recording as a 'theorem' any result which they hit upon that happens to interest them. Lakatos suggested that in fact theorems are invariably conjectured ahead of proof, and the proof process is a protracted affair in which initial attempts are criticized and gradually improved, along with the 'theorem' itself. At first this is trial-and-error, involving searches for counterexamples both to the original 'theorem' and to the 'hidden assumptions' that are articulated in the course of initial attempted proofs.

This trial-and-error phase of conjecture followed by undirected search for proof and/or counterexample is

eventually superseded, however, by a more systematic phase - that of ‘proof analysis’. Lakatos - inspired by his countryman Polya - argued that the process of mathematical discovery is not a ‘merely’ psychological affair to be studied by trying to delve into the minds of the great mathematicians, but can be shown instead to be governed by articulable heuristic principles. Hans Reichenbach, Popper and the Logical Positivists all saw an unbridgeable divide between - in Reichenbach’s terminology - the ‘context of justification’ and the ‘context of discovery’; and all asserted that philosophy or logic of science is concerned only with the former ‘context’. Questions of discovery were alleged neither to call for logical analysis nor to be susceptible of it. Lakatos argued that, in the case of mathematics at least, this view is importantly mistaken, and that there is a realm of logically analysable, mathematical heuristic outside the two traditional ‘contexts’. This claim forms an important link with Lakatos’ later contributions to the philosophy of science (see Discovery, logic of §2).

One of the intriguing (but often frustrating) aspects of ‘Proofs and Refutations’ is that its dialogue form, although used to brilliant effect, sometimes makes it difficult to discern what thesis is actually being propounded: not even ‘Teacher’ is always right. This has resulted in some obscurity about the underlying view of the nature of mathematical knowledge. Lakatos himself believed that his message was fundamentally Popperian - that he was extending Popper’s fallibilism into the area of mathematics. Another view - indicated in the book’s editorial footnotes - is that Lakatos simply described (in fascinating detail) the fallible *process* by which essentially infallible, logically true, knowledge is created in mathematics. These editorial footnotes have themselves been attacked as a ‘betrayal’ of the real ‘anti-formalist’ message of the book (Davis and Hersch 1981).

3 The ‘methodology of scientific research programmes’: contributions to philosophy of science

After ‘Proofs and Refutations’, Lakatos turned his attention to the philosophy of science. Although his 1968 paper ‘Changes in the Problem of Inductive Logic’ defends a broadly Popperian line against Carnapian ‘inductive logic’, Lakatos soon came to see major defects in the Popperian approach. ‘Popper on Demarcation and Induction’ (1974) contains a detailed criticism of Popper’s claim to have solved Hume’s problem. Popperian corroboration appraisals are simply summaries of how the available theories have stood up to testing *so far*. Those appraisals, therefore, can have no consequences for the comparative future reliabilities of the various theories, nor for the reasonableness of relying on one theory rather than a rival in future applications, unless some assumption is made that - in Popperian terms - links corroboration appraisals to claims about overall ‘verisimilitude’. Such an assumption amounts to a reformulation in Popperian terms of a uniformity-of-nature assumption, linking past test results to an overall judgment of the theory’s truth-likeness and hence its reliability in past *and* future tests. That (merely posited) principle has, therefore, seemed to most commentators much stronger than a mere ‘whiff’ of induction, as Lakatos represented it.

Although argued in a new way, Lakatos’ point here is similar to arguments against Popper already produced by Reichenbach, Ayer, Salmon and others. Lakatos’ views on ‘demarcation’ were altogether more innovative. Thomas Kuhn had pointed out (1962 and elsewhere) that many aspects of the development of science seem at direct odds with Popper’s falsificationist account. For example, the typical response of a theoretician to an experimental ‘refutation’ of his favoured theory is not, as Popper seemed to suggest, to reject that theory and look for an alternative, but instead to treat the experimental result as an ‘anomaly’ which could and should be accommodated within the theory. Lakatos saw his ‘methodology of scientific research programmes’ (further developed and defended in his 1970 paper) as a synthesis of the views of Popper and of Kuhn. He agreed with Kuhn that the correct unit of analysis in science is much broader than that of a single theory. A Lakatosian research programme is characterized by a ‘negative heuristic’ principle specifying its ‘hard core’ (the set of basic propositions that will be implied by every theory that issues from it), and by a ‘positive heuristic’ (a set of directives, possibly deriving from some broad metaphysical principle, governing the construction of specific theories within the programme, and governing their modification in the light of experimental difficulties that may arise). Specific theories produced by a programme *are* experimentally refutable (or would be if fully articulated), but the standard response to an actual refutation will be to look for a further specific theory within the same programme. Since this successor theory will also entail the same central (‘hard core’) claims, this process will seem like ‘holding onto’ a theory by ‘modifying’ it in the light of experimental difficulties, rather than rejecting it.

This sounds like Kuhn’s idea of ‘articulating the paradigm’ and treating experimental difficulties as ‘anomalies’. According to Lakatos, the main problem with Kuhn’s account is that it seems not to be able to explain scientific

change as a rational process. What distinguishes a proper, scientific further articulation of a paradigm or programme from a defensive nonscientific one? If it was good scientific practice for Newtonians to defend their basic ‘paradigm’ or ‘hard core’ against their failure to explain Uranus’s orbit by postulating a hitherto unknown planet, was it not equally good scientific practice for phlogistonists to defend their theory - that combustion always involves release of phlogiston - against the fact that burning mercury in air produces a heavier ‘ash’ by postulating negative weight for phlogiston (or by postulating that burning mercury involves both the release and the absorption of material)? A second, related problem is that the distinction between real science and pseudoscience seems to be endangered by Kuhn’s account. Kuhn seems committed to the view that there are no articulable ‘logical’ rules of good science or of correct response to evidence. Priestley was not ‘wrong or unscientific’ to hold on to the phlogiston theory, he was simply outvoted.

Lakatos argued that there is in fact a clear-cut distinction here - one based on the old idea of independent evidence, but given a new slant. The difference between, for example, the Newtonian shift and that involved in defending phlogistonism is the difference between a ‘progressive’ and a ‘degenerating problem shift’. The postulation of a further planet to explain the anomalous motion of Uranus within Newtonian theory not only solved the problem of Uranus’s motion but also led to further independently testable predictions - the new planet could after all be observed. On the other hand, the shift in phlogistonist assumptions simply *at best* resolved the known anomaly, while making no further testable predictions.

Lakatos was always interested in the relationship between the philosophy of science (and mathematics) and its history and felt that philosophy of science in the post-war period had suffered from its paying too little attention to actual scientific practice. In his 1971 paper ‘History of Science and its Rational Reconstructions’, he proposed a general method for the evaluation of rival philosophies or methodologies of science in terms of the ‘rational reconstructions’ they provide of the history of science and especially of historical episodes of major changes in accepted theories. His basic idea was that there is a range of historical cases in which, speaking intuitively, the ‘scientifically correct’ decisions were clearly made (for example, the acceptance of Newtonian theory or of Maxwell’s theory). A methodology should endorse (and so give a general explanation of) these cases; and should, in cases where it implies that the ‘wrong’ decision was made, be able to point to independent evidence for the operation of ‘external factors’ (political or religious interference, for example). He argued that his own methodology did better on these terms than other accounts, such as inductivism, conventionalism or Popperian falsificationism.

See also: Experiment §2; Feyerabend, P.K.; Scientific method §2; Theories, scientific §8

JOHN WORRALL

List of works

Lakatos, I. (1963-4) ‘Proofs and Refutations’, *The British Journal for the Philosophy of Science* 14: 1-25, 120-39, 221-43, 296-342; repr. with editorial footnotes in Lakatos 1976.(The original published version of Lakatos’ major work in the philosophy of mathematics.)

Lakatos, I. (1968) ‘Changes in the Problem of Inductive Logic’, in I. Lakatos (ed.) *The Problem of Inductive Logic*, Amsterdam: North Holland; repr. in Lakatos 1978. (Lakatos’ account of the dispute between Carnap and Popper over probabilistic, inductive logic.)

Lakatos, I. (1970) ‘Falsificationism and the Methodology of Scientific Research Programmes’, in I. Lakatos and A.E. Musgrave (eds) *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press, 1970; repr. in Lakatos 1978. (The definitive account of Lakatos’ views on scientific research programmes and their development.)

Lakatos, I. (1971) ‘History of Science and its Rational Reconstructions’, in R.C. Buck and R.S. Cohen (eds) *PSA 1970, Boston Studies in the Philosophy of Science 8*, Dordrecht: Reidel; repr. in Lakatos 1978.(This paper contains Lakatos’ views about ‘meta-methodology’: how to evaluate rival theory-appraisal criteria.)

Lakatos, I. (1974) ‘Popper on Demarcation and Induction’, in P.A. Schillp (ed.) *The Philosophy of Karl Popper*, Library of Living Philosophers, La Salle, IL: Open Court; repr. in Lakatos 1978.(Lakatos’ appeal to Popper for a ‘whiff of induction’ and a systematic comparison of his views on science with those of Popper.)

Lakatos, I. (1976) *Proofs and Refutations: The Logic of Mathematical Discovery*, ed. J. Worrall and E. Zahar, Cambridge: Cambridge University Press.(This book contains extra material on Cauchy and uniform convergence as well as the original 1963-4 papers.)

Lakatos, I. (1978) *The Methodology of Scientific Research Programmes: Philosophical Papers vol. 1*, ed. J. Worrall and G. Currie, Cambridge: Cambridge University Press. (Volume 1 of Lakatos' collected papers.)
Lakatos, I. (1978) *Mathematics, Science and Epistemology: Philosophical Papers vol. 2*, ed. J. Worrall and G. Currie, Cambridge: Cambridge University Press.(Volume 2 of Lakatos' collected papers, including some hitherto unpublished material.)

References and further reading

Davis, P.J. and Hersh, R. (1981) *The Mathematical Experience*, Brighton: Harvester.(Contains an endorsement of some of Lakatos' views on mathematics and criticism of the editorial footnotes in Lakatos 1976.)
Hacking, I. (1983) *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge: Cambridge University Press.(Contains an important chapter on Lakatos' methodology of research programmes.)
Howson, C. (ed.) (1976) *Method and Appraisal in the Physical Sciences*, Cambridge: Cambridge University Press. (Contains several 'case studies' of important episodes of theory-change in the history of science analysed from the point of view of the methodology of scientific research programmes.)
Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press, 2nd edn, 1970.(The most influential contemporary account of theory-change in science.)
Lakatos, I. and Musgrave, A.E. (eds) (1970) *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press.(Contains an outline of Kuhn's account of science, together with a range of commentaries and Kuhn's replies.)
Newton-Smith, W.H. (1981) *The Rationality of Science*, London: Routledge. (Contains a sympathetic, but critical analysis of the views of Lakatos and of related authors on scientific progress.)
Zahar, E.G. (1989) *Einstein's Revolution: A Study in Heuristic*, La Salle, IL: Open Court.(A critical account of the methodology of scientific research programmes, including a significant elaboration of the idea of rationally analysable heuristic, together with an extended case study of the relativistic revolution.)

Lambda calculus

The lambda calculus presents a delimited formal setting for expressing and studying properties of mathematical operations and computer programs. Its syntactical unit is the term. Application terms tu represent supplying argument u to routine t. Abstraction terms $\lambda x.t$ *(in which* λx *binds x in t) represent the explicit definition of an operation from a routine t. Term computation is by reduction: (* $\lambda x.t$*)(u) reduces to t[x/u], the result of substituting u for x in t; this corresponds to operating with t on u. Terms are equal either when they differ only in bound variables or when one converts to the other via a finite chain of reductions and counter-reductions. There are two main varieties of lambda calculi: typed and untyped (or type-free). Any untyped term is applicable to any other and equality between such terms is undecidable. In the typed case, terms are indexed by types and application is permitted only when types conform. Here, equality is decidable. A. Church proposed formalisms for both varieties during the 1930s. He and his students S. Kleene and J. Rosser proved the early syntactical metatheorems linking the untyped calculus with computability theory. In 1969, D. Scott gave the first set-theoretic construction of a functional model of the untyped calculus; since then, many other models have been devised. As for the typed calculus, Gödel's* Dialectica *interpretation inspired many of its applications to proof theory. Also, the idea, due to H. Curry and W. Howard, of using logical formulas as types reveals a close correspondence between typed terms and derivations in intuitionistic formal systems. In the last decades of the twentieth century, the lambda calculus and its models have been put to many uses in computer science, with functional programming and denotational semantics deserving special mention.*

1 Abstraction, application and currying

Gottlob Frege's doctrine of concepts and objects (1893) put abstract mathematical functions, including higher-order functions taking functions as inputs or yielding functions as outputs, at the centre of logic (see Frege, G. §3). He insisted that adequate notation for functions must distinguish them from their values. For example, the expression $x^2 + 3x + 7$, for x a natural number, is ambiguous, denoting both the value of the expression at fixed input x and the function that yields $n^2 + 3n + 7$ for every value n of x. Frege realized that, to disambiguate, it suffices to bind x with a prefix operator, such as λx, to yield

$$\lambda x.x^2 + 3x + 7.(1)$$

(Frege's own notation did not employ λ but the Greek symbol for smooth breathing.) These expressions, conveniently displaying both independent variable and a recipe for computation, Frege took to denote functions (in *extension*). Such terms as (1), now called 'abstraction terms', reflect a Fregean notion of function as grasped by abstraction from its values: $0^2 + 3\cdot 0 + 7$, $1^2 + 3\cdot 1 + 7$ and so on. Those values are retrievable from (1) by application: the term

$$(\lambda x.x^2 + 3x + 7)(0)$$

denotes the value of the function on the input 0, which we compute by dropping λx and substituting 0 for x in $x^2 + 3x + 7$ to obtain $0^2 + 3\cdot 0 + 7 = 7$.

In effect, Frege also recognized that lambda notation renders binary or multiary functions as combinations of unary ones, a conversion known as 'currying' after the logician H. Curry. For instance, a lambda term for addition, $\lambda x\, \lambda y.x + y$, denotes a function taking its arguments one at a time: first x and then y. Its evaluation at 2 yields a unary function

$$(\lambda x\, \lambda y.x + y)(2) = \lambda y.2 + y,$$

which can be evaluated at 3:

$$(\lambda y.2 + y)(3) = 2 + 3 = 5$$

to yield the desired result.

Alonzo Church was the first to devise formal systems, which he called 'calculi of lambda-conversion' (1941), devoted primarily to codifying such properties of functions. He introduced lambda notation as a variant of the caret (ˆ) employed in Whitehead and Russell's *Principia Mathematica* (1910-13) as abstraction operator. Church

intended his calculi to capture an operational conception of function distinct from the extensional notion enshrined in set theory. Set theory conceives of functions as collections of ordered pairs with domain and range so delimited that application of arbitrary function to arbitrary input is prohibited. Hence, set theory does some injustice to operations which apply universally, such as identity, which maps anything to itself, or that operation taking each entity a to the constantly a-valued function. By Church's lights, expressions of the lambda calculus were to denote operations or 'functions in intension': mathematical rules, routines or programs for computing outputs from inputs which may apply universally, even applying to themselves.

2 A formal calculus of terms

There are two general varieties of lambda calculi: 'typed' and 'untyped' (or 'type-free'), both devised by Church. In either case, the basic syntactic category is that of the 'term' (or λ-term). Atomic terms are either variables or constants. Compound terms are of two kinds. There are 'application' terms: if t and u are both terms, so is tu, denoting the input of u to routine t. And, for explicit functions, there are 'λ-abstractions': if t is a term, so is $\lambda x.t$, where x is a variable and λ binds all occurrences of x. In typed systems, each term is also assigned a type and applications *tu* are well-formed only if the types of t and of u are suitable. (Further discussion of typed systems is postponed to §6). The untyped lambda calculus suffers no such restriction: self-applications such as xx and $(\lambda x.xx)(\lambda x.xx)$ are permitted. Sample untyped terms include $\lambda x.x,\ \lambda x.a,\ \lambda x\,\lambda y.x,\ \lambda y.y$. The first gives the identity function, the second (with a constant) the constant a-valued function and the third the first projection function, mapping each item a to the constantly a-valued function $\lambda y.a$. The last yields the same function as the first: terms differing only in bound variables, known as 'α-congruent terms', are identified.

Closed abstraction terms without constants are called 'combinators'. Conventionally, combinators $\lambda x.x,\ \lambda x\,\lambda y.x$ and $\lambda x\,\lambda y\,\lambda z.xz(yz)$ are abbreviated $I,\ K$ and S, respectively, where I stands for 'identity', K for the *Konstanzfunktion* yielding constant functions, and S for a general composition function or *Verschmelzungsfunktion*. This nomenclature derives from M. Schönfinkel's combinatory logic and signals that these lambda terms perform the same operations as their combinatory namesakes (see Combinatory logic §1).

3 Computations, reductions and normal terms

It is in no way inaccurate to see lambda calculi as rudimentary programming languages, with terms representing computing recipes and their values. As a programming language, unrestricted application in the untyped lambda calculus realizes von Neumann's idea that programs serve also as data (see Neumann, J. von). There is a notion of lambda calculus computation, which may, like real-life computation, lead to non-terminating 'infinite loops'. As will be apparent (see below) the lambda calculus programming language is rich enough to express all possible algorithms on the natural numbers and is illuminated by respectable concepts of program meaning or denotation.

The notion of 'reduction' (properly, β-reduction) is the principal analogue within the lambda calculus of finitary computation. It affords an analysis of a process of stepwise evaluation obeying a series of atomic, rule-governed steps. Every such step takes its start from the idea of abstraction term exploited above: $\lambda x.t$ names the recipe which begins computation when λx is removed and a name for an input is substituted where x appears free in t. In the lambda calculus, every atomic reduction step begins with a 'redex', an application term of the form $(\lambda x.t)u$ - setting us the project of evaluating recipe $\lambda x.t$ at input u. One replaces the redex with its 'contractum', term $t[x/u]$, the result of substituting u for all free occurrences of x in t (with bound variable changes to avoid clashes). 'Computing' in the lambda calculus is iterated reduction: term v reduces to w if w results from v via a finite number of one-step reductions or replacements of contractum for redex.

Terms are equal when they are interconvertible via computation. If u is derivable from t via a finite sequence of reductions and counter-reductions (perhaps with changes in bound variables) t β-equals (or β-converts to) u; symbolically, $t = u$. A lambda theory T is a set of formal equations between terms that extends β-equality: whenever t β-equals u, then $t = u$ belongs to T. Intuitively, correct reductions preserve the meanings of the terms involved; hence, β-equal terms share the same meaning. Lambda theories, as sets of equations, respect those meanings. (Lambda models make rigorous sense of these intuitions about meaning. See §5.)

A series of reductions terminates when no further reduction is possible. Irreducible terms are said to be in 'normal form'. It follows that any variable or constant is in normal form, as is any term $\lambda x.t$ in which t itself is normal.

But, just as some computer programs loop infinitely, some untyped terms never reduce to a normal form: $(\lambda x.xxy)(\lambda x.xxy)$ reduces to $((\lambda x.xxy)(\lambda x.xxy))y$ and further reductions always reproduce the original redex. Though some untyped terms lack a normal form, when a term does so reduce, its normal form is unique, at least up to bound variables. This is a corollary of the crucial ‘Church-Rosser theorem’, which states that, if t reduces to u and to v, then there is a term w to which u and v both reduce.

Notions of equality alternative to β-equality are studied, among them $\beta\eta$-equality. This relation respects the ‘ η-rule’: that $\lambda x.tx$ reduces to t when t does not contain x free.

4 Arithmetic and undecidability

Church and his students S. Kleene and J. Rosser showed, in effect, that, as a programming system, the untyped lambda calculus is comprehensive in that one can define any computable function within it, once natural numbers are represented. Let 0 be $\lambda x\,\lambda y.y$, 1 be $\lambda x\,\lambda y.xy$, 2 be $\lambda x\,\lambda y.x(xy)$, 3 be $\lambda x\,\lambda yx(x(xy))$, and so on. These ‘Church numerals’ possess the attractive property, exemplified by 3, that $3ab$ reduces to $a(a(ab))$. Also, all Church numerals are in normal form. Let $[[n]]$ stand for the Church numeral of n. A natural number function f, which may be partial, is λ-defined by t just in case, whenever $f(n)$ exists, $t[[n]]$ reduces to $[[f(n)]]$, and, if $f(n)$ is undefined, then $t[[n]]$ does not reduce to a normal form. A function is λ-definable when it has a λ-defining term.

During 1933, Church speculated that the λ-definable functions coincide with the algorithmically computable functions. This speculation came to be known as ‘Church’s thesis’. If Church’s thesis is true, then every possible recipe for computing with the integers is λ-defined by a term in the untyped lambda calculus and every computation with that recipe appears as a reduction sequence involving that term. Church and Kleene were able to confirm Church’s thesis by proving that every function which is Gödel-Herbrand general recursive is λ-definable. One proof of this requires the construction of fixed-point combinators: terms which, given t, yield u such that tu β-equals u. Y, the ‘paradoxical combinator’ discovered by Curry,

$$\lambda x.(\lambda y.x(yy))(\lambda y.x(yy))$$

is one such; applied to any t, Yt β-equals $t(Yt)$.

As befits a programming system, the untyped lambda calculus exhibits undecidability. In 1936, Church used the calculus to prove the undecidability of first-order logic, resolving Hilbert’s *Entscheidungsproblem* (‘decision problem’; see Hilbert’s programme and formalism). A general undecidability theorem for the lambda calculus, first proved by D. Scott in 1963 and rediscovered by Curry in 1969, can be stated in terms of recursive separability. Two sets of terms are ‘recursively separable’ if there is a decidable set containing one and disjoint from the other. The ‘Scott-Curry theorem’ asserts that no two non-empty sets of terms closed under β-equality are recursively separable. A corollary is that untyped β-equality and normalizability are both undecidable.

5 Models

Until the late 1960s, syntactic study dominated research on the untyped lambda calculus. Term models were constructed but their elements were syntactic: equivalence classes of terms under such relations as β-equality. An obstacle to building models in set theory, models which capture Church’s vision of a calculus of functions, was self-application. In standard set theory, self-application of functions is not permitted, since no set-theoretic function lies in its own domain.

In 1969, Scott constructed D_∞, the first recognizably functional model for the untyped case. Scott realized that terms can be interpreted successfully as continuous (rather than arbitrary) functions on complete partial orders. Here, a complete partial order is a partially ordered set with a least element and with least upper bounds for all directed subsets. A continuous function is one whose output values are determinable by finite approximations. D_∞ is a complete partial order enabling self-application, since it is isomorphic to the space of continuous functions from D_∞ into itself. Each element of D_∞ then plays two roles: as element of the model and as function on the model’s domain.

Other models, some simpler in design, were soon forthcoming. G. Plotkin, in 1972, discovered the fundamental idea behind (what is now called) the graph model $P\omega$. The elements of $P\omega$ are subsets of the natural numbers.

$P\omega$ is not isomorphic to its space of continuous functions but, since that space is embedded within $P\omega$, it interprets self-application. A general notion of lambda model, due to A. Meyer (1982) and Scott (1976), which both D_∞ and $P\omega$ exemplify, is tripartite, comprising non-empty domain D, binary application $\circ$ on D and a unary operation Λ. $\circ$ creates on D a model of combinatory logic while Λ serves as a 'functionalizer' mapping each d in D to the function taking a into $d \circ a$ (see Combinatory logic §2). Soundness and completeness theorems confirm the adequacy of this general notion: every lambda model satisfies all basic principles of β-equality and every lambda theory has a lambda model. Intuitively, then, the meanings preserved by reductions are respected in every lambda model.

6 Typed lambda calculi

In the typed lambda calculus, inputs to a function given by a term, as well as its outputs, are restricted to a particular domain or type. For example, to say that term t has type $A \Rightarrow B$ will mean that its inputs are restricted to members of type A and its outputs to B. Hence, if $\mathbb{N}$ is the domain of natural numbers, the successor function on $\mathbb{N}$ is of type $\mathbb{N} \Rightarrow \mathbb{N}$. Typing renders formal two informal ideas. The first is the notion that every well-defined mathematical function bears a specific domain of inputs and range of outputs. The second is that underlying programming languages in which allowable inputs to a program are restricted to members of a particular type, say, natural numbers or integers. With respect to decidability, typing lambda terms makes for notable differences.

Types are defined inductively: there are atomic types and, whenever α and β are types, so is $\alpha \Rightarrow \beta$. One may interpret atomic types to include the natural numbers or the truth-values and compound types $\alpha \Rightarrow \beta$ as containing functions from type α to type β. Types are then used to index terms. Variables and constants, if any, are assumed typed. Application terms must accord with typing: if t is of type $\alpha \Rightarrow \beta$ and u is of type α, then (tu) is of type β. For abstraction terms, when x is of type α and t of type β, $\lambda x.t$ is of type $\alpha \Rightarrow \beta$.

It is straightforward to define natural concepts of reduction, equality and normality for typed terms. There are no self-applications or paradoxical combinators in typed systems. As A. Turing first proved, every typed term reduces to a normal form. Typed reduction also satisfies a Church-Rosser theorem. In fact, a strong normalization theorem obtains: given a typed term, every legitimate way of reducing it eventually ends in the same normal term. It follows that, in contrast with the untyped calculus, typed equality is decidable. Suitable Church numerals are constructible; the natural number functions definable by typed terms are, as discerned independently by H. Schwichtenberg and R. Statman, precisely the extended polynomials. These include integer polynomials such as $\lambda n.3n^2 + 6n + 4$, and functions defined from them using definitions by polynomial cases, for example, $3n^2 + 6n + 4$ if $n = 3$ and n^4 otherwise.

Proof-theoretic applications of typed lambda calculi are prominent. The calculus is extendible to a formalism T for describing the constructive functionals Gödel used in his *Dialectica* interpretation and proof of arithmetic's consistency (see Intuitionism §6). Terms in T admit a natural reduction relation which is Church-Rosser and terms are strongly normalizable.

In the typed calculus, if A, B and C are types, then compound types such as $A \Rightarrow (B \Rightarrow A)$ and

$$[A \Rightarrow (B \Rightarrow C)] \Rightarrow [(A \Rightarrow B) \Rightarrow (A \Rightarrow C)]$$

resemble formulas in implicational logic. Curry and R. Feys (1958) noted that these formulas, which are intuitionistically valid, are types suitable for the combinators K and S, respectively. Indeed, the correspondence is general: typed terms accord precisely with correct derivations in the implicational part of intuitionistic propositional logic with application mirroring *modus ponens* and abstraction conditional proof. Moreover, reductive simplifications in logical derivations coincide with reductions in the lambda calculus. This correlation of terms with proofs, known as the 'Curry-Howard isomorphism' or 'formulas as types', was successfully extended to first-order logic by W. Howard (1980) and to second-order logic by J.-Y. Girard.

7 Computers and the lambda calculus

Research on the lambda calculus is now closely allied with machine computation. For one thing, lambda calculi exhibit, in pure form, issues of decidability, non-termination and typing which face computer theorists. Also, the lambda calculus formalism has inspired a range of functional programming languages, including J. McCarthy's

LISP and J. Reynold’s GEDANKEN. Third, model theory for the lambda calculus gave rise to both operational and denotational semantics for programming languages. A semantics for a programming language is ‘operational’ when it explains the meaning of a program in terms of its effect upon the state of an abstract machine; it is ‘denotational’ when the meaning of a program is a mathematical function denoted by that program. P. Landin, in 1965, employed translation into the lambda calculus to give an operational semantics for the language ALGOL. In the early 1970s, Scott and C. Strachey used the ideas behind Scott’s lambda calculus models to give denotational semantics for programming languages exhibiting such phenomena as recursion and non-termination.

See also: Logical and mathematical terms, glossary of

DAVID CHARLES McCARTY

References and further reading

Barendregt, H. (1981) *The Lambda Calculus: Its Syntax and Semantics*, Amsterdam: North Holland, 2nd edn, 1984.(A bible of lambda calculus research; authoritative and detailed.)

Church, A. (1941) *The Calculi of Lambda Conversion*, Princeton, NJ: Princeton University Press.(The original monograph on the lambda calculus.)

Curry, H. and Feys, R. (1958) *Combinatory Logic*, vol. 1, Amsterdam: North Holland.(Once the premier work on the subject; it remains an excellent reference.)

Frege, G. (1893) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, vol. 1, Jena: Pohle; part 1 trans. and ed. M. Furth, *Basic Laws of Arithmetic: An Exposition of the System*, Berkeley, CA: University of California Press, 1964.(Furth’s introduction is extremely helpful.)

Hindley, J. and Seldin, J. (1986) *Introduction to Combinators and λ-Calculus*, London Mathematical Society Student Texts no. 1, Cambridge: Cambridge University Press.(An excellent brief introduction to the subject. Highly recommended for beginners.)

Howard, W. (1980) ‘The Formulas-As-Types Notion of Construction’, in J. Hindley and J. Seldin (eds) *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, New York: Academic Press, 479-90. (The first publication of notes circulated informally in 1969.)

Kleene, S. (1981) ‘Origins of Recursive Function Theory’, *Annals of the History of Computing* 3: 52- 67.(A fascinating eyewitness account of the birth of the lambda calculus and computability theory at Princeton in the early 1930s.)

Martin-Löf, P. (1984) *Intuitionistic Type Theory*, Naples: Bibliopolis.(Among the most influential contemporary developments of typed lambda calculus and formulas-as-types.)

Meyer, A. (1982) ‘What is a Model of the Lambda Calculus?’, *Information and Control* 52: 87-122.(A brilliant and highly readable article on a unitary general semantics for the lambda calculus.)

Rosser, J.B. (1984) ‘Highlights of the History of the Lambda Calculus’, *Annals of the History of Computing* 6: 337-49.(Another fine memoir on the early days of lambda calculus research at Princeton and later developments.)

Scott, D.S. (1976) ‘Data Types as Lattices’, *SIAM Journal of Computing* 5: 522-87.(A characteristically charming exposition of results by Scott, Plotkin and others on the graph model.)

Stenlund, S. (1972) *Combinators, λ-Terms and Proof Theory*, Dordrecht: Reidel.(Remains one of the best and most lucid introductions to the proof-theoretical applications of the lambda calculus; highly recommended.)

Stoy, J. (1977) *Denotational Semantics: The Scott-Strachey Approach to Programming Language Theory*, Cambridge, MA: MIT Press.(The first book-length introduction to Scott-Strachey semantics for programs.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7; repr. 1994.(The section ‘Propositional Functions’ in the introduction’s chapter 1 describes functional abstraction.)

Lambert, Johann Heinrich (1728-77)

Lambert was a German mathematician, physicist, astronomer and philosopher, who was among the leading figures of German intellectual life in the late eighteenth century. As a practising scientist, who made important discoveries in many areas, Lambert was interested in philosophical questions regarding the methods of scientific knowledge. In his philosophical works he sought to reform metaphysics by subjecting it to the procedures and standards of mathematics, advocating a combination of conceptual analysis and deductive construction in philosophy. With Lambert the tradition of German rationalist thought reaches directly into the time of Kant, who had great esteem for his analytic skills.

Lambert was born in Mühlhausen, Alsace (then under Swiss dominion). Having received no regular schooling beyond the age of twelve, he was almost entirely self-taught. He undertook his studies and research activities in mathematics and the natural sciences while working in various auxiliary positions in Switzerland and Southern Germany. From 1748 to 1758 he served as tutor to a Swiss family. During the final two years of that appointment he accompanied two of his pupils on their grand tour of Western Europe, which led through the major university towns and allowed him to meet some of the leading scientists of his time. In 1759 he became a member of the newly founded Munich Academy, an association that was terminated in 1762 because Lambert failed to take up residency in Munich. In 1764 he travelled to Berlin, ostensibly on his way to St Petersburg, where he hoped to gain an appointment as member of the Academy. Rather than travelling on, he stayed in Berlin for the rest of his life, becoming a member of the Berlin Academy in 1765 and assuming the post of surveyor of public works in Prussia in the same year.

Lambert was one of the leading mathematicians and natural scientists of his time. In mathematics he proved the irrationality of pi, introduced hyperbolic functions into trigonometry and made contributions to probability theory and the development of non-Euclidean geometry. In physics, Lambert mainly worked on light, heat and humidity of the air. He founded the science of photometry and also made contributions to acoustics and meteorology. In his *Cosmologische Briefe über die Einrichtung des Weltbaues (Cosmological Letters about the Structure of the Universe)* (1761) he developed a hypothesis about the nebular origin of the universe akin to the one developed by Kant in his *Allgemeine Naturgeschichte und Theorie des Himmels (Universal Natural History and Theory of the Heavens)* of 1755, which was not known to Lambert at the time.

Lambert was first and foremost a mathematician and scientist and only second a philosopher. His philosophical works all date from the 1760s and are mainly concerned with questions of method. The major influences on Lambert's philosophical views were Malebranche, Wolff and Locke. Lambert sought to improve the scientific standing of philosophy in general and metaphysics in particular by emulating mathematical procedures in philosophical reasoning. Like Leibniz before him, he was intrigued by the idea of a universal characteristic that would allow the formal presentation and definite solution to problems of all kinds. Unfortunately Lambert did not pursue this project in the last ten years of his life, effectively abandoning philosophy in favour of mathematics and the physical sciences (see Universal language).

The first of Lambert's two main philosophical works is the *Neues Organon oder Gedanken über die Erforschung und Bezeichnung des Wahren und dessen Unterscheidung vom Irrthum und Schein (New Organon or Thoughts on the Investigation and Designation of the True and its Distinction from Error and Appearance)* which appeared in 1764. The work is a treatise on logic in the Pietist tradition of Hoffmann and Crusius, but also reflects the influence of Leibniz, Wolff and Locke. In line with contemporary practice, the more strictly logical core of the *New Organon* is enriched by psychological and methodological considerations. The work is organized in four parts. Part One, which is by far the most extensive, is the 'Dianoiology or Doctrine of the Laws of Thinking' and contains the traditional logical doctrines concerning concepts, judgments and syllogistic inferences. Important innovative topics include the discussion of the methodological status of questions and problems, the role of ordinary as well as scientific, experimental experience, and the distinction between historical and scientific knowledge. Moreover, Lambert introduces a linear symbolism for the representation of judgments, which is then employed in his syllogistic logic.

Part Two of the *Neues Organon* is the 'Alethiology or Doctrine of Truth'. Lambert's key concern here is with the

nature and function of the simple concepts that serve as the building blocks for the logical construction of true propositions. Part Three, the 'Semiotics or Doctrine of the Designation of Thoughts and Things', features the traditional distinction between word, concept and thing but also includes interesting considerations on linguistic confusion. The work concludes with the 'Phenomenology or Doctrine of Appearance' (*Schein*), in which Lambert distinguishes between true and false appearance, discusses probable knowledge and investigates the role of appearance in various forms of cognition.

Lambert went on to apply the logical canon set forth in the *New Organon* in his second main contribution to philosophy, *Anlage zur Architectonic, oder Theorie des Einfachen und des Ersten in der philosophischen Erkenntniß (Plan for the Architectonic, or Theory of What Is Simple and First in Philosophical Knowledge)*, which was published in 1771 but had been finished in 1764. The work provides the propaedeutic for a foundational philosophical science in the tradition of Wolff's ontology or general theory of being. Lambert analyses the basic philosophical concepts and principles with an eye to their eventual use in the edifice of philosophical knowledge. Most of the work consists of detailed, often very subtle examinations of philosophical terms and distinctions. In the procedure of philosophy, as envisioned by Lambert, the analysis of the basic concepts is followed by the axiomatic construction of a body of propositions, which is then applied to experience in the manner of applying mathematics to physical reality. Lambert's own work did not extend significantly beyond the initial, analytic phase of the project.

While Lambert's analyses of philosophical concepts and principles exercised some influence on Tetens and Kant, it was not until some two hundred years after his death that his contributions to the theory of scientific reasoning were properly recognized.

See also: Scientific method

GÜNTER ZÖLLER

List of works

Lambert, J.H. (1965-9) *Philosophische Schriften (Philosophical Writings)*, ed. H.-W. Arndt, Hildesheim: Olms, 10 vols.

Lambert, J.H. (1761) *Cosmologische Briefe über die Einrichtung des Weltbaues (Cosmological Letters on the Structure of the Universe)*, Augsburg; repr. in G. Jakisch (ed.) *J.H. Lamberts 'Cosmologische Briefe' mit Beiträgen zur Frühgeschichte der Kosmologie*, Berlin: Akademie-Verlag, 1979.

Lambert, J.H. (1761) *Abhandlung vom Criterium veritatis (Treatise on the Criterion of Truth)*, ed. K. Bopp, Berlin: Reuther & Reinhard, 1915.(Published posthumously.)

Lambert, J.H. (1762) *Über die Methode die Metaphysik, Theologie und Moral richtiger zu beweisen (On the Method of Proving Metaphysics, Theology and Morals More Correctly)*, ed. K. Bopp, Berlin: Reuther & Reinhard.(Published posthumously.)

Lambert, J.H. (1764) *Neues Organon oder Gedanken über die Erforschung und Bezeichnung des Wahren und dessen Unterscheidung vom Irrthum und Schein (New Organon or Thoughts on the Investigation and Designation of the True and its Distinction from Error and Appearance)*, Leipzig, 2 vols; repr. in *Philosophische Schriften*, vols 1-2; repr. and ed. G. Schenk, Berlin: Akademie-Verlag, 1990, 3 vols.

Lambert, J.H. (1771) *Anlage zur Architectonic, oder Theorie des Einfachen und des Ersten in der philosophischen Erkenntniß (Plan for the Architectonic, or Theory of What Is Simple and First in Philosophical Knowledge)*, Riga; repr. in *Philosophische Schriften*, vols 3-4.

Lambert, J.H. (1781-7) *Johann Heinrich Lamberts deutscher gelehrter Briefwechsel*, ed. J. Bernoulli, Berlin, 5 vols; vol. 1 repr. in *Philosophische Schriften*, vol. 9.(Lambert's scientific correspondence conducted in German.)

Lambert, J.H. (1782-7) *Logische und philosophische Abhandlungen (Logical and Philosophical Essays)*, Berlin, 2 vols; repr. in *Philosophische Schriften*, vols 6-7.

Lambert, J.H. (1988) *Texte zur Systematologie und zur Theorie der wisssenschaftlichen Erkenntnis (Writings on the Theory of Systems and the Theory of Scientific Knowledge)*, ed. G. Sigwart, Hamburg: Meiner.(Includes selections from Neues Organon and Anlage zurArchitectonic, with bibliography.)

References and further reading

Baensch, O. (1902) *Johann Heinrich Lamberts Philosophie und seine Stellung zu Kant*, Tübingen and Leipzig: Mohr; repr. Hildesheim: Gerstenberg, 1978.(An account of Lambert's philosophy and his relation to Kant;

Baensch rejects the then current view of Kant’s indebtedness to Lambert.)

Beck, L.W. (1969) *Early German Philosophy. Kant and his Predecessors*, Cambridge, MA: Harvard University Press, 402-12.(Places Lambert in the context of German rationalist tradition in metaphysics.)

Debru, C. (1977) *Analyse et représentation. De la méthodologie à la théorie de l’espace: Kant et Lambert*, Paris: Vrin.(A study of Lambert’s contribution to the formation of Kant’s philosophy.)

Hinske, N. (ed.) (1983) *Lambert-Index. Stellenindex zu Johann Heinrich Lamberts ‘Neues Organon’*, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 2 vols.(Word index to Lambert’s *New Organon*.)

Hinske, N. (ed.) (1987) *Lambert-Index. Stellenindex zu Johann Heinrich Lamberts ‘Anlage zur Architectonic’*, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 2 vols. (Word index to Lambert’s *Plan for an Architectonic*.)

Peters, W.S. (1968) ‘I. Kants Verhältnis zu J.H. Lambert’, *Kant-Studien* 59: 448-53.(A documentation of the references to Lambert in Kant’s writings and correspondence.)

Steck, M. (1970) *Bibliographia Lambertiana. Ein Führer durch das gedruckte und ungedruckte Schrifttum und den wissenschaftlichen Briefwechsel von Johann Heinrich Lambert. 1728-1777*, Hildesheim: Olms.(A guide through Lambert’s published and unpublished writings and his scientific correspondence, with bibliography of literature on Lambert.)

Université de Haute-Alsace (ed.) (1979) *Colloque international et interdisciplinaire Jean-Henri Lambert Mulhouse, 26-30 septembre 1977*, Paris: Ophrys.(Conference proceedings with eight contributions on Lambert’s philosophy.)

Wolters, G. (1980) *Basis und Deduktion. Studien zur Entstehung und Bedeutung der Theorie der axiomatischen Methode bei J.H. Lambert (1728-1777)*, Berlin and New York: De Gruyter.(Focuses on Lambert’s importance as a philosopher of science; includes bibliography.)

Lange, Friedrich Albert (1828-75)

A German philosopher, social scientist and political activist, Lange was best known for his study of the history of materialism. He was a leading proponent of Neo-Kantianism, a critic of speculative metaphysics, and a defender of the view that philosophy should incorporate the findings of the exact sciences. As a social scientist, Lange described the emergence of a social Darwinian 'struggle for existence' in modern times due to the rapid advancement of industrialization and a growing conflict of interest among social classes.

Cognizant of the scientific trends of his time, Lange anticipated some of the central ideas of pragmatism and adopted a form of conventionalism in regard to scientific principles and concepts. Although sympathetic to materialism, Lange also saw the inevitability of an idealist element in interpretations of natural phenomena and insisted on the importance of projecting ethical, social and aesthetic ideals.

1 Life

Lange was born at Wald and died near Zurich. He was educated at the University of Bonn and later became a university instructor there in 1851. In 1870, he was appointed professor of inductive logic at Zurich. Accepting a professorship at Marburg in 1873, Lange established an influential Neo-Kantian orientation there. By 1861 he became actively involved in politics, working for constitutional reform along democratic socialist lines.

2 Philosophical writings

Lange's most important philosophical work is his impressive *Geschichte des Materialismus und Kritik seiner Bedeutung in der Gegenwart (History of Materialism)*, published in a one-volume edition in 1866 and a two-volume edition in 1873-5. The two-volume edition included references to contemporary scientific theories. A supplementary study, *Neue Beiträge zur Geschichte des Materialismus (New Contributions to the History of Materialism)*, was published in 1867. Two years earlier, he had presented *Die Grundlegung der mathematischen Psychologie (The Foundation of Mathematical Psychology)*. His last work, *Logische Studien (Logical Studies)*, appeared posthumously in 1877.

The *History of Materialism* is a detailed and informative study of the development of materialism from ancient Greece to the scientifically supported modes of materialism emerging in the 1870s. More than a scholarly history, Lange's major work presents perceptive commentaries on physical and biological theories, political economy and ethical naturalism. There is also an extensive discussion of the critical philosophy of Kant in relation to nineteenth-century scientific thought. Lange adopts a restrained positivism in regard to scientific theory and a conventionalist orientation towards scientific principles and posits. He also advocates the construction of aesthetic conceptions of 'the All' from what he called 'the standpoint of the ideal'.

Lange avoids a reductive materialism and argues that a critical analysis of materialism reveals ideal-theoretical principles, concepts and assumptions that cannot be interpreted materialistically. Hence, his considered philosophical position is a form of materio-idealism. Anticipations of pragmatism, scientific conventionalism, instrumental fictionalism and a phenomenalism in regard to scientific concepts can all be discerned in his *History of Materialism*. In terms of his critical commentaries on a variety of scientific theories and their conceptual foundations, Lange anticipated the approach later adopted in the philosophy of science.

Under the influence of Kant, Lange criticized speculative metaphysical claims to a knowledge that transcends experience. He viewed traditional metaphysics as akin to religion and art, as a form of conceptual poetry. The projection of ideals beyond the 'fragments of truth' discovered by the exact sciences was, however, construed as a basic human need. Space, time and causality, as well as the fundamental posits of science, were conceived of as human inventions or, in effect, useful fictions. Lange insisted that the discoveries in the exact sciences must be incorporated into philosophical thought in so far as they are directly relevant to our concept of actuality and theory of knowledge.

3 Social thought and influence

Though less well-known than his study of materialism, Lange's writings on social science were insightful. In

History of Materialism he already predicted that there would be ‘earthquakes’ in the sociopolitical realm due to the decline of the power of religion, the advance of science and the tensions generated by the ‘social problem’. In a neglected study of 1865, *Die Arbeiterfrage in ihrer Bedeutung für Gegenwart und Zukunft (The Problem of the Worker and its Meaning for the Present and Future)*, he analysed the implications of the increasingly rapid industrialization in Europe and its need for an army of workers and skilled technicians. He foresaw a coming social (Darwinian) struggle for existence that would generate radical sociopolitical change. A champion of workers’ rights and sympathetic to democratic socialism, Lange envisaged the emergence of a new class of technically proficient workers who would soon attain a position of leadership among European nations.

In another work of 1865, *J. St. Mills Ansichten über die sociale Frage und die angebliche Umwälzung der Socialwissenschaften durch Carey (J.S. Mill’s Views on the Social Problem and Carey’s Supposed Social-Scientific Revolution)*, Lange displayed his knowledge of J.S. Mill’s liberal-democratic ideas, as well as of Henry Carey’s defence of a capitalist national-economic theory (see Mill, J.S. §12). He was a perceptive analyst, aware of the ‘egoism’ and social atomism that the drive to accumulate capital can produce. Lange held that the rapid economic changes in Europe would lead both to social conflict and radical political oppositions. These views helped shape his socialist sentiments (see Socialism).

The influence of Lange on subsequent philosophers and philosophical trends, though generally indirect, is traceable in some instances. The Neo-Kantianism he inaugurated had a long-range effect on German thought and influenced a number of philosophical scientists (see Neo-Kantianism §§1-2). In social thought he was a forerunner of the conception of social Darwinism, not as an advocate but as a predictor of it. The fictionalist and pragmatic theory of Hans Vaihinger was shaped by Lange’s post-Kantian conventionalist theory of the use and meaning of categories, scientific postulates and theory-construction. Friedrich Nietzsche considered Lange’s *History of Materialism* ‘a real treasure-house, to be looked into and read repeatedly’. He was strongly influenced by him in regard to his own general philosophical goal of uniting philosophy, science and art, as well as in his critical analysis of knowledge.

GEORGE J. STACK

List of works

Lange, F.A. (1863) *Die Leibesübungen (Physical Training)*, Gotha: Besser.(An early work, perhaps the earliest, on the importance of ‘body-training’ or physical education for students.)

Lange, F.A. (1865a) *Die Arbeiterfrage in ihrer Bedeutung für Gegenwart und Zukunft (The Problem of the Worker and its Meaning for the Present and Future)*, Duisburg: Falk & Volmer.(An analysis of the rising problem of a Darwinian struggle for survival in European society as a consequence of the new social classes emerging out of increasing industrialization with emphasis upon the impact of the class of skilled workers and technicians on the balance of power in society.)

Lange, F.A. (1865b) *J.St. Mill’s Ansichten über die sociale Frage (J.S. Mill’s Views on the Social Problem and Carey’s Supposed Social-Scientific Revolution)*, Duisburg: Falk & Volmer.(A sympathetic analysis of Mill’s democratic liberal ideals and socialist programmes, and a critique of Henry Carey’s nationalist political economic theory.)

Lange, F.A. (1865c) *Die Grundlegung der mathematischen Psychologie (The Foundation of Mathematical Psychology)*, Duisburg: Falk & Volmer.(An attempt to place psychology on an experimental, empirical basis in terms of a mathematical model.)

Lange, F.A. (1866) *Geschichte des Materialismus und Kritik seiner Bedeutung in der Gegenwart*, Iserlohn: Baedeker; 2nd edn, Iserlohn and Leipzig: J. Baedeker, 1873- 5, 2 vols; trans. E.C. Thomas, *History of Materialism*, Boston, MA: Osgood, 1877, 3 vols.(A detailed, then-comprehensive study of the history of materialism from the ancient Greeks to 1872 (in the 2nd edition) with special emphasis on recent developments in physics, biological theory and social theory, and their impact on a Neo-Kantian epistemology, as well as ethics. A pioneering work in what was later designated philosophy of science.)

Lange, F.A. (1867) *Neue Beiträge zur Geschichte des Materialismus (New Contributions to the History of Materialism)*, Iserlohn: Baedeker.(An updating of his work of 1866, primarily in the areas of the sciences. Most of this material was incorporated into his 2nd edition of the *History of Materialism*.)

Lange, F.A. (1877) *Logische Studien (Logical Studies)*, Iserlohn: Baedeker.(An analysis of logic in terms of a nominalism that conceives of language as comprised of ‘arbitrary signs’ that refer to particulars. Logic should focus on the logical aspects of language, as well as its diachronic features.)

References and further reading

Ellissen, O.A. (1891) *F.A. Lange*, Leipzig: Baedeker.(Interesting biographical material in the form of letters.)

Salaquarda, J. (1978) 'Nietzsche und Lange', *Nietzsche-Studien* 7: 236-60, Berlin and New York: De Gruyter.(A concise discussion of some aspects of Lange's impact on Nietzsche.)

Stack, G.J. (1983) *Lange and Nietzsche*, Berlin and New York: De Gruyter.(A comprehensive study of the influence of Lange on Nietzsche's philosophical project, as well as on his critique of knowledge.)

Vaihinger, H. (1902) *Die Philosophie des 'Als-Ob'*, Berlin: Reuther & Reichard; trans. C.K. Ogden, *The Philosophy of 'As-if'*, London: Routledge & Kegan Paul, 1924. (Contains a substantial discussion of Lange's thought with special emphasis on his idea of the use of fictions in philosophy and science, as well as the value of edifying cultural ideals.)

Vaihinger, H. (1876) *Hartmann, Dühring and Lange*, Iserlohn: Baedeker.(A summary of the central conceptions of Hartmann's philosophy of the unconscious, Dühring's social and scientific theories, and Lange's phenomenalism.)

Langer, Susanne Katherina Knauth (1895-1985)

With roots in logic, philosophy of language and philosophy of mind, Susanne Langer sought to explicate the meaning and cognitive import of art works by developing a theory of symbolism that located works of art at the centre of a network of relations based firmly on semantic theory. Art works were non-discursive, presentational symbols that expressed an artist's 'life of feeling', by which observers, through a process of immediate apprehension (or intuition) came to acquire knowledge.

Langer was educated at Radcliffe College and briefly attended the University of Vienna. She held the post of tutor at Radcliffe from 1927 to 1942, followed by positions at the University of Delaware (1943), Columbia University (1945-50), the Connecticut College for Women (1954-62), and a number of visiting positions.

Langer's early writings prefigure her later theorizing on art. *The Practice of Philosophy* was influenced by work on symbolism by Whitehead and the early Wittgenstein. For Langer, the relation between a symbol and the symbolized object depended solely upon analogy of form; only by studying the structure of an entity and its comparable analogue could the symbolic relationship be established. The logical analysis of symbols - as analogues to one's concrete experiences - admitted of meanings that were incommunicable by ordinary, public discourse. Symbols, through non-discursive form, achieved cognitive weight and force equal to the original experience. As such, they transmitted knowledge of the ineffable. Her second text, *An Introduction to Symbolic Logic*, moved beyond a presentation of standard techniques in symbolic logic to an exploration of its conceptual foundations.

Her explicit theory of art began with *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite and Art*. Combining her interest in the nature of knowledge with a long-standing wish to actualize a complete theory of mind (realized decades later with the publication of *Mind: An Essay on Human Feeling*), she expanded her discussion of symbols by focusing on the role of human reason in ritual acts and the creation of art objects. Challenging traditional philosophical approaches, which had previously located the focus of attention in theories of taste, beauty or notions of aesthetic value, she sought to alter radically the most basic questions in the philosophy of art. Symbolism was the 'new key' to understanding how the human mind transformed the primal need to express oneself. All forms of human activity, including 'speech and gesture, song and sacrifice' were seen as expressive. The mental work of symbolic transformation resided in abstracting a gesture or an object from reality. Symbols, in contrast to signs (later called signals), gave expression to thoughts that went beyond what could be expressed in language. Therefore, a wide range of symbols evolved over time: simple ones became complex; ritual (or convention) became art. Art became the conveyor of inner life: artists expressed 'ideas of feeling' - 'formulation and representation of emotions, moods, mental tensions and resolutions' - not their own actual or personal feelings. The artistic significance of music, for instance, consisted in pure form - 'the sensuous percept' - apart from its literal content (what, if anything, it represented).

In *Feeling and Form*, written eleven years later as a sequel to *Philosophy in a New Key*, art was characterized as the creation of forms symbolic of human feeling. Every work of art involved (1) abstraction from actuality, thereby becoming mere semblance, a created realm of illusion, (2) plasticity (the capacity of being manipulated in the interests of expression), and (3) expressiveness whereby the symbol became transparent. A focus on the meaning of art works was replaced by a discussion of their import or significance. Intuition became the link between the qualities of the art work that constituted it a symbol and the import the work of art held for the observer. Through intuition, we perceive the 'felt life' of the artist's expression.

In *Problems of Art*, a collection of essays originally delivered as lectures, Langer refined her views: a work of art was a form expressive of human feeling, created for our aesthetic perception through sense or imagination. Emphasizing the role that artistic intention played in creative activity, she traced the unity of the arts to their semblance of organic form. Insight (understanding of the essential life of feeling) was designated the aim of art. *Reflections on Art*, a collection of twenty-six essays ranging over music, art, dance, poetry, film and architecture, focused on two main issues: expressiveness and semblance. Her list of contributors included artists and 'lay aestheticians', as well as professional philosophers. Her final work, *Mind*, ambitiously sought to explicate the role feelings play as the mind functions uniquely in humans, and in particular how an artist projects an idea of feeling

by means of art.

See also: Anthropology, philosophy of §1; Artistic expression; Music, aesthetics of §§6, 9

PEG BRAND

List of works

Langer, S.K.K. (1930) *The Practice of Philosophy*, New York: Holt.(An early general analysis of symbols, not related to art.)

(1937) *An Introduction to Symbolic Logic*, Boston, MA: Houghton Mifflin.(Conceptual foundations of how logic is needed to clarify semantics.)

(1942) *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite and Art*, Cambridge, MA: Harvard University Press.(Art is a nondiscursive symbol that conveys an artist's expressed 'ideas of feeling'.)

(1953) *Feeling and Form: A Theory of Art*, London: Routledge.(Extends *Philosophy in a New Key* to a characterization of art as the creation of forms symbolic of human feeling.)

(1957) *Problems of Art: Ten Philosophical Lectures*, New York: Charles Scribner's Sons.(Published lectures extending Langer's theory to discussions of artistic intention, the unity of the arts, and insight.)

(1958) *Reflections on Art: A Source Book of Writings by Artists, Critics, and Philosophers*, Baltimore, MD: Johns Hopkins University Press.(An edited interdisciplinary collection of essays focusing on expressiveness and semblance.)

(1962) *Philosophical Sketches*, Baltimore, MD: Johns Hopkins University Press.(A collection of essays that are studies towards the philosophy of mind presented in *Mind: An Essay on Human Feeling*.)

(1967, 1972, 1982) *Mind: An Essay on Human Feeling*, 3 vols, Baltimore, MD: Johns Hopkins University Press. (An exploration of the biological foundations of human mentality as the site of source of the artistic expression of 'living' form.)

References and further reading

Ahlberg, L.-O. (1994) 'Susanne Langer on Representation and Emotion in Music', *British Journal of Aesthetics* 34 (1): 69-80.(Compares Langer's views to those of Peter Kivy and Jerrold Levinson.)

Bertocci, P.A. (1970) 'Susanne K. Langer's Theory of Mind and Feeling', *Review of Metaphysics* 23: 527-51.(A critique of Langer's form of naturalism, her attempt to escape dualism, and her account of the unity and continuity of mind.)

Binkley, T. (1970) 'Langer's Logical and Ontological Modes', *Journal of Aesthetics and Art Criticism* 28: 455-64. (Argues that music is not an essentially significant form because music has an ontological, not a logical, function.)

Bufford, S. (1972) 'Susanne Langer's Two Philosophies of Art', *Journal of Aesthetics and Art Criticism* 31: 9-20. (Detects a 'perceivability theory' in Langer's writings in addition to her well-known expression theory.)

Davies, S. (1983) 'Is Music a Language of the Emotions?', *British Journal of Aesthetics* 23: 222-33.(Rejects Langer's theory in favour of his own, in which music is naturally meaningful.)

Hagberg, G. (1984) 'Art and the Unsayable: Langer's Tractarian Aesthetics', *British Journal of Aesthetics* 24 (4): 325-40.(Reconsidering Langer's theory on a Wittgensteinian model of meaning.)

Hansen, F. (1968) 'Langer's Expressive Form: An Interpretation', *Journal of Aesthetics and Art Criticism* 27: 165-70.(Interprets and extends Langer's original concept of 'symbol' beyond her own presentation.)

Lang, B. (1962) 'Langer's Arabesque and the Collapse of the Symbol', *Review of Metaphysics* 16: 349-65.(The failure of Langer's theory based on the cognitive content of a symbol.)

Price, K. (1993) 'Philosophy in a New Key: An Interpretation', *The Philosophy of Music Education Review* 1 (1): 34-43.(Reasons for Langer's immediate popularity upon publishing *Philosophy in a New Key* and her lasting importance; some questions her works raise.)

Reichling, M.J. (1993) 'Susanne Langer's Theory of Symbolism: An Analysis and Extension', *The Philosophy of Music Education Review* 1 (1): 3-17.(A concise overview of Langer's theory and its critics, and an analysis and extension of the theory into music.)

Reichling, M.J. (1995) 'Susanne Langer's Concept of Secondary Illusion', *The Journal of Aesthetic Education* 29 (4): 39.(Short note on 'secondary Illusion'.)

Reimer, B. (1993) 'Langer on the Arts as Cognitive', *The Philosophy of Music Education Review* 1 (1): 44-60. (Seeks to clarify Langer's position as to how the arts are cognitive, and why education in the arts is essential (and often ignored by scholars in cognitive science).)

Yob, I. (1993) 'The Form of Feeling', *The Philosophy of Music Education Review* 1 (1): 18-33.(Focuses on the

metaphorical nature of a non-discursive symbol.)

Language and gender

How do language and gender interact? This can be interpreted as asking about sexual difference in relation to language-use. How do the sexes speak, how do we speak of the sexes? And could or should these patterns change?

Not surprisingly, understanding language-gender interactions solely in terms of sexual difference yields a static and polarized picture. Men insult and swear, women flatter and wheedle, women draw others out while men monopolize conversations, men are direct and women beat around the bush, women gossip whereas men lecture. Linguistic conventions and familiar vocabulary equate humanity with males (note, for example, so-called generic uses of 'he') and sexuality with females ('hussy', for instance, once meant 'housewife'). Men are linguistically represented as actors and women as acted upon, passive. Men control the institutions controlling language - such as schools, churches, publications, legislatures. Children of both sexes, however, learn a 'mother tongue' at a mother's knee.

Such generalizations contain a few grains of truth, at least if restricted to so-called mainstream contemporary America or England. But they completely obscure the differences among women and among men and the varied forms of social relations so important to gender. One is never just a woman or a man: sexual classifications are inflected by age, class, race and much else. And gender involves not only women in relation to men as a group but also more specific cross-sex and same-sex relations ranging from egalitarian heterosexual marriages and same-sex partnerships through intense friendships and enmities among adolescent schoolgirls to camaraderie among boys on a football team. All such relations are partly constituted by people using language to and of one another; all are informed by and inform larger social arrangements. On the more linguistic side, these include dictionaries, the language arts curriculum and editorial guidelines; arrangements with a gender focus include marriage, high-school dances and gay rights legislation.

Emphasizing large-scale sex difference ignores cross-cultural and historical variation and makes change in language, in gender, or in their interaction appear mysterious. And such an emphasis erases the linguistic dynamics of a particular society's construction of gender. Yet it is in such dynamics that, for example, language shapes and is shaped by sexual polarization and male dominance. This entry highlights approaches to language and gender that root each in historically situated social practice. Linguistic change and gender change then become inseparable.

1 Recent background

In the early 1970s, feminist-inspired interest in sex and gender revived venerable philosophical and linguistic questions about how language, culture and thought interact. During the 1980s and 1990s, feminist scholarship in the USA and Europe took an increasingly linguistic turn, not only in philosophy and linguistics but also in history, literary theory, psychology, politics and sociology. From many different viewpoints, scholars argue that language - and more generally, discourse (see §3) - shapes gender identities and relations and supports male and heterosexual privilege. Outside the academy as well, language figures prominently in gender discussions. Mass circulation publications report that the sexes speak 'different languages'; feature writers complain about feminist 'word police' (and often two sentences later about 'misuses' or 'abuses' of words like *rape*); anti-sexist and gay rights activists invent new terminology (*sexism* or *homophobia*) or renovate the old (using *queer* in positive self-reference or *she* as a generic pronoun).

Language-gender debates, however, include participants with different conceptions both of language and of gender. This section touches on three recent kinds of work: the Anglo-American empirically oriented tradition, psychoanalytic theorizing from French philosophers and linguists, and work on discourse and gender construction from feminist philosophers and other theorists in literary and cultural studies.

Linguist Robin Lakoff's widely read *Language and Woman's Place* (1975) framed much subsequent discussion by American linguists and social scientists and some philosophers. Lakoff argued that women face a double-bind. To sound feminine they must speak indirectly and euphemistically; that style, however, is derided as ineffective in public arenas. She argued also that women were systematically derogated when spoken of, that language used of women conventionally implied that they were less worthy and important than men. Although she did not join in

the widespread critique of masculine generics (for example, *man* for humans generally or *he* with a general antecedent; see the introduction to Frank and Treichler (1989) for a discussion of this topic), she noted instructive examples of asymmetries in linguistic resources for speaking of men and of women: *cleaning lady*, for instance, versus the non-occurring *garbage gentleman* or the metaphorical extension of animal or food terms to refer to women, such as *cow* or *tart*.

Others challenged Lakoff's ideas. Linguists such as Janet Holmes (1995) pointed to the multiple functions of linguistic forms. For example, rising intonation on a statement, a form Lakoff interpreted as signalling women's uncertainty or insecurity, can facilitate effective communicative interaction, inviting others to speak. Tag questions ('he's pathetic, *isn't he*?') have a similar range of functions. Lakoff's critics also noted the cultural specificity of speech styles (for example, many African-American women reported Lakoff's characterizations inapplicable to their everyday speech). Furthermore, some (most notably, Penelope Brown (1980) and Marjorie Harness Goodwin (1990)) argued that sex-differentiated rhetorical strategies are not arbitrary cultural conventions. They arise as strategic responses to general social constraints and the specific demands of particular communicative contexts. Women's linguistic agency began to be explored, not just their victimization.

During the same period, some French thinkers began to articulate feminist perspectives on post-structuralist or postmodern views of language. The post-structuralist stance does not take language as a closed system for representing a pre-existing reality (see Postmodernism §2). To speak or write is never a neutral act of encoding, as dominant Anglo-American views of language seem to suggest, but always enacts a self; language is constitutive of subjectivity.

Psychoanalysis, Freud's 'talking cure', has always given language a prominent place, and French psychoanalyst Jacques Lacan made it central in his interpretation of Freud. Linguistic communication, he proposed, is an attempt to erase the first and most traumatic psychic pain, the wrenching separation of birth. To speak is to try to reconnect with the mother, or, more generally, to connect to another; this attempt at (re)connection is inherently phallic. Moreover, to speak and be understood requires one to submit to patriarchal social laws, to place oneself under the Law of the Father. None of this is a matter of being male or female: women can and do speak and write but linguistic communication remains phallic, not feminine, at a deep psychosexual level (see Feminism and psychoanalysis).

Then is silence the only feminine linguistic move? The linguistic and social *status quo*, though powerful, are not stable closed systems. Breaks in the discourse, slips of the tongue, puns, parody and repetition, all bespeak psychic difference and potential resistance, and offer a glimpse beyond the Law of the Father. Such moves disturb and destabilize the patriarchal *status quo*; they can render audible a feminine voice outside the phallic linguistic order. Again it is not sex that is at issue: a man's speech or writing may inscribe a de-centring feminine voice, exposing the instability of the ruling order and threatening its hegemony. Although the French postmodern turn does not necessarily promote the dismantling of male privilege, there is a strong feminist strain in such thinking. The feminine and masculine within each person - a kind of bisexuality - is the fulcrum on which Luce Irigaray and Hélène Cixous, for example, balance femin*ist* politics. Although their specific projects differ, both make femin*ine* outside voices more prominent, and they subvert in varied ways established patriarchal assumptions and values.

Both early American emphasis on sex-differentiated modes of speaking and sexist linguistic resources, and the French focus on the psychosexual significance of language, seemed to take sexual difference and male dominance as given, as prior to linguistic practices. Women and men, the feminine and the masculine - these were unequal poles existing outside language. During the mid-1980s and early 1990s, however, a number of English-speaking feminist philosophers and other cultural and social theorists explicitly questioned the 'naturalness' of heterosexuality and of polarized and hierarchized gender oppositions. They reread thinkers like Irigaray (1977) and emphasized the role of discourse in constructing gender identities and relations, noting that linguistic abstractions and other symbolic practices are central to constructing categories of sex and sexuality. Queer theory especially emphasized the instability of those categories, their emergence from performance, and thus the possibility of alternative forms of gender and sexuality (see, for example, Butler 1990). In a related move, thinkers with a more empirical bent began to conceptualize gender as emerging from social practice (see, for example, Connell 1987).

2 Gender and language in social practice

Much early empirical work took gender and language each as independent structured systems. Researchers asked whether these systems were correlated or connected but not how such connections might arise or why they might matter. Social practice theory suggested a different strategy. Do not look at language and gender as macrosystems; look instead at the microlevel of social activities and conventions for understanding and regulating them. Both gender and language systems can be seen as rooted in social practice; because a single activity often has both gender and language implications, these roots intertwine. Even shared roots do not, however, guarantee that the separate structures supported will connect at higher levels.

What does it mean to say that language is rooted in social practice? Natural language grammars have been argued (most notably by Chomsky and followers) to be tightly constrained by the nature of human minds, seen in this respect as sexless - and likewise unaffected by race, class and geographical origins (see Chomsky, N.; Language, innateness of). From this perspective, fruitful for much linguistic inquiry, social life seems inconsequential to the forms of language. Specifying grammatical possibilities, however, does not tell us everything about language in human history.

At the simplest level, people aim at coordinating their grammars so that communication works reliably within a community. They also talk like those around them simply in order to mark themselves as a social collectivity. Thus controversial questions arise: which grammar(s) or language(s)/dialect(s) should a particular community use? Who adjusts to whom? How do male-dominated institutions affect choice among linguistic possibilities? And (nearer the periphery from a grammatical viewpoint but central for language-users) which words with which meanings are part of shared resources? What difference can vocabulary make to thought and social interaction?

This latter question moves us beyond the form of a community's linguistic resources to the functions those resources serve. What are norms for using language? Are women supposed to listen to men more than men to women? Is it mostly men's linguistic formulations of ideas that enter common currency and serve as background when others speak and write? Who jokes and when? Who swears and when? Who is polite to whom and why? Who suggests and who orders? In which situations and to which addressees? Which meanings are contested? When, where and for whom is silence or talk prescribed? In most communities, doing things with words is highly inflected by gender relations - and also by class, race, age and other dimensions of social difference and hierarchy.

Even more than language, gender is assumed to be built on a biological foundation. Virtually all cultures sort people at birth into one of two sex categories - female and male. External genitalia, which generally predict potential reproductive roles, mark sex at birth. Thereafter, however, other socially constructed markers take precedence. In most social groups sexual classification is associated with power relations, division of labour and regulation of many other areas of life including erotic activity (heterosexism and racism, for example, dictate that desire should be directed only towards someone of the other sex and the same race). Even what we think of as most 'biological' - the roles played in species reproduction - can be affected by social activity. Consider such historical developments as bottle-feeding, frozen sperm and test-tube babies. Just as strikingly, cultures differ dramatically in the range of gendered identities they offer (not all, for instance, stop with two sexes) and the kinds of gender relations that prevail.

The substantive content of gender derives from social practices and the attitudes and expectations that drive and support them. Some areas of gender practice seem independent of language: for example, styles of dress and other kinds of bodily adornment and demeanour, conventions regulating who touches whom when and where and what the reaction should be, participation in competitive sports, food preparation and serving. Even these, however, connect to linguistic practices. A person's clothes and visual style may get them dubbed 'slutty' or 'elegant', silence is deemed consent to sexually charged touch, locker-room talk is judged unsuitable for female ears, a group of women talks in the kitchen while their male partners trade comments about the football game on the living room television.

Writing is also important. Teen magazines have articles on lip gloss and diet, self-help books advise on female orgasm, newspapers devote much of their space to sports stories and use sports metaphors even in articles about business or politics, women's but not men's magazines include many pieces on food for family and for entertaining. Of course neither speech nor writing is just words. Teen magazines, for example, link to huge cosmetics and fashion industries and to the myriad of institutions and practices pushing teenage girls towards

(exclusively heterosexual) romance. But the bottom line is that social practices of all kinds have rich linguistic texture: language is central to social life.

3 Connecting two different conceptions of discourse

The emphasis of much recent feminist thinking has been on discourse as a global feature of culture, what I will call cultural discourse. We might, for instance, talk about the (cultural) discourse of romance in which talk, writing, clothing, photographs, film and much else present romance as normatively heterosexual, females' primary aspiration in life, the only important form of cross-sex relations, and so on. Cultural discourse covers background assumptions, favoured rhetorical strategies, vocabulary, and nonlinguistic activities and representations that highlight some perspectives on a particular domain (for example, intimate interpersonal relations) and obscure others.

Linguists and many philosophers of language use a notion of discourse more narrowly linguistic and less global. What I will call a situated discourse consists of a historically located series of connected utterances or inscriptions - for example, a conversation or a written narrative. For analytic purposes we can identify a (situated) discourse with a string of sentences together with relevant aspects of the contexts in which the sentences are produced and interpreted. Cultural discourse can be seen as grounded in situated discourses, whose detailed analysis may shed light on the microstructure of larger discursive constructions of gender.

Studies of language in use give concrete linguistic content to claims about the discursive production of sex, sexuality and gender. For example, names and styles of address start forming sexed identities early on (nurses in one hospital were heard calling baby girls 'sweetheart' or 'beautiful', but using 'Jones' for the baby boy with that surname). Adolescents draw on categorizing practices that link sex and sexual decorum with social class when they label someone 'slut'. Engagement in same-sex verbal tussles (direct insults, often overlaid with humour, or 'he-said/she-said' tales about absent parties), heterosexual harassment (street comments on a woman's body) or the discourse of romance (identifying four-year-olds' opposite-sex playmates as their 'sweethearts') - such everyday language-use helps shape and sustain particular forms of gender identity and relations.

Less everyday language-use is also, of course, critically important. Sermons or scriptures may enjoin women's silence in places of worship or other forms of sex-differentiation and hierarchy. And social scientists write about gender, often equating it with statistically significant sex differences (and ignoring its inseparability from matters of class, race, gender and even age). We find, for instance, sweeping and dichotomizing claims about how women and men speak. Such work is popularized and becomes enormously influential. It gets used not only for interpreting but for regulating gender-language connections: not only is this 'on average' how women and men speak but this is how 'real' or 'nondeviant' women and men speak or ought to speak. (Although Tannen (1990) herself avoids such normativizing, her many readers have not.)

Social life depends on social interaction, whether face-to-face or more diffuse and larger in scale. Interactions are made up of more than situated discourses, of course, but language is central to most of them and virtually essential to large-scale and long-distance social exchange. Situated discourses are the primary ingredient of social life: cultural discourses derive much of their substance from situated discourses.

All situated discourses are constrained by conventions and institutions predating them. All, however, have potential effects on subsequent discourses and, more generally, on social structural constraints affecting talk and understanding, on cultural discourses and other macro-level social arrangements. For example, when talking about people in general, saying 'Each must do what she thinks best' will startle many listeners: English classes have prescribed 'Each must do what he thinks best' in such contexts. The feminine generic goes against familiar conventions, yet its increasing use (notably in American philosophical writing) begins to establish an alternative convention. The more it is used the less it shocks.

4 Changing language, changing gender

'Sticks and stones may break my bones but words will never hurt me', chant children trying to defuse the sting of labels like 'sissy', 'faggot' or 'bitch'. They seek to convince themselves and their tormentors of the view, widely endorsed in one form or another, that language has no force, is causally inert. As formal objects on their own, of course, words (more generally, languages) do not hurt people or do anything else to them. But words in use do

indeed affect people in many ways: they convince, persuade, enlighten, frighten, humiliate, amuse, disgust, titillate. Words link to social arrangements on the one hand (see Speech acts), and to causally relevant features of courses of events on the other (see Reference). These links give them many kinds of power in human affairs and thinking, both individual and collective. Thus it is not surprising that language change and social change might go hand in hand.

Take a simple example. Labelling people affects how those people and their labellers then enter into a host of social practices. Until recently most communities using American-English conventionally assigned adult women social titles on the basis of marital status (‘Mrs’ or ‘Miss’), whereas men’s social titles were not so differentiated (‘Mr’ being the only option other than occupational and professional titles). The introduction of ‘Ms’ offered women a title option supposedly neutral as to marital status. This option was adopted by a diverse group including advertisers who did not want to offend by making mistaken assumptions about marital status, young unmarried women wishing not to advertise themselves as single, and self-described feminist women claiming a status equivalent to that of male peers (whether husbands or not). There have been other changes in linguistic practice connected to marriage, sexuality and family: for example, more women retain a pre-marriage surname or hyphenate surnames, some couples create a new shared surname, and some use the mother’s surname for a child; even women who have adopted a husband’s surname frequently favour ‘Ms/Mrs Jane Doe’ over the formerly dominant ‘Mrs John Doe’ form; employers and others now inquire about a ‘spouse’ (or even a ‘partner’, finessing both marital status and sexual preference) rather than a ‘wife’; marriage ceremonies much less often ask a woman to ‘obey’ a husband, and those officiating frequently pronounce the couple ‘husband and wife’ rather than ‘man and wife’; same-sex couples go through marriage-like rituals and adopt common surnames; ‘parent’ has acquired a use as a verb (which is semantically far closer to the verb ‘to mother’ than to the verb ‘to father’).

Such changes in sociolinguistic practice have accompanied nonlinguistic changes in the institutions of marriage and family and practices associated with them. Middle-class women are less likely than they used to be to assume their economic welfare and social position will derive from husbands’ income and status. Not only do young women see divorce as a real possibility, they also see themselves as capable of significant earnings and professional achievement, whether or not they opt to marry or to have children. Some men see caring for children - parenting - as a central role for part of their lives. Heterosexual marriage of the traditional hierarchical and strongly sex-differentiated kind, though still in many ways the default option for middle-class Americans, is increasingly seen as not the only choice open, not even for women who want to bear children. Adding ‘Ms’ as a title option for women and changing laws about surnames of married women and their children could not have increased women’s participation in high-level careers if all else had remained the same, just as having available the sex-neutral verb ‘to parent’ does not suffice to get men more actively engaged in the activity it denotes. But changes in linguistic practice have been (and continue to be) part and parcel of changes in gender practices - the linguistic and nonlinguistic developments reinforce one another.

Linguistic practices change all the time. For example, new words or terms are introduced. ‘Surrogate mother’ designates a kind of relation of woman and child not earlier envisaged (and carries with it certain assumptions about the social weightiness of that relation); ‘sexual harassment’ groups together, on the basis of similar effects, kinds of situations previously either ignored or seen as very different in kind (‘sexual teasing’, for instance, and ‘seduction’). Of course, new expressions draw on associations with existing ones: ‘sexism’, for example, developed meaning in part through the implied analogy with the word ‘racism’. Sometimes an existing form is altered in its uses, as when ‘partner’ comes to designate the person with whom one lives in a sexually intimate relationship, whether that relationship is sanctioned by the state or religious authorities, whether the person so designated is a woman or a man, and whether of the same sex as, or opposite sex to, oneself. But the implication of equality remains from other uses of the term ‘partner’.

Customary usage may change while reference stays fixed. Some have begun to say ‘my child’ or ‘my kid’ in contexts where most people still say ‘my daughter’ or ‘my son’, to say ‘kids’ rather than ‘girls and boys’, and in other ways to resist identifying everyone always in sex-specific terms. The linguistic system and its interpretation are not thereby changed, but patterns of language-use are and with them what is implicated when people speak, what we take them to mean above and beyond what their words literally say (see Implicature). There are no explicit norms that say adults must identify a person to children by using ‘that woman’ or ‘that man’ or similar sex-specific forms, yet this is standard practice: ‘Say thank you to the nice lady’ rather than ‘Say thank you to the

nice person who gave you that sweet'. Changes in such practices might ultimately help effect major shifts in gender polarization and emphasis on sexual difference, an emphasis important for enforcing heterosexuality.

Could some group impose what might seem desirable linguistic and related social changes? Should they? History shows that some regulation is possible: for more than a century, schools and editors, for example, have proscribed singular 'they' (used by Shakespeare, Jane Austen and many others, including me and the authors of other *Routledge Encyclopedia of Philosophy* entries) and prescribed a supposedly sex-indefinite 'he'. Simply removing institutional sanctions (poor grades, having an article rejected) from the use of singular 'they' would almost certainly greatly reduce use of supposedly generic 'he'. But of course some people might continue its use, and even those who have dropped it might nonetheless speak in other ways indicating a view of humans as normatively male: 'During the night the villagers left in canoes, leaving us behind with the women and children.'

Linguistic conventions, both those that narrow the range of linguistic systems on which community members can draw and expect to be understood and those that promote certain patterns of usage (for instance, when to say 'thank you'), only constrain and never completely determine what community members will say. People can in various ways challenge and resist such conventions or exploit indeterminacy in them as to form and meaning. And other people can make countermoves as in the charges of a silly 'political correctness' hurled at those who have drawn attention to social biases implicit in existing linguistic conventions and have proposed alternatives. No social standpoint monopolizes all moves; subordinated interests can find expression. This does not mean that social advantage confers no linguistic advantage (it does) or that linguistically aided social change is not possible (it is). But total control of language use (and therefore of the patterns of thought and action it might facilitate) is only an Orwellian nightmare, not a real possibility. And, as many feminist thinkers have reminded us, quick linguistic fixes for sexism and heterosexism are just pipedreams (see the guidelines in Frank and Treichler 1989).

See also: Feminism; Linguistic discrimination

SALLY McCONNELL-GINET

References and further reading

Brown, P. (1980) 'How and Why are Women More Polite: Some Evidence from a Mayan Community', in S. McConnell-Ginet, R.A. Borker and N. Furman (eds) *Women and Language in Literature and Society*, New York: Praeger, 1980, 111-36.(Draws on Grice-style pragmatic theory and her own ethnographic work to develop a strategic view of politeness and, more generally, speaker style. Reprinted in Coates 1997.)

Butler, J. (1990) *Gender Trouble*, London and New York: Routledge.(Develops her view of the performativity of gender and its base in discourse.)

Cameron, D. (1985) *Feminism and Linguistic Theory*, London: Macmillan, 2nd edn, 1992.(Argues that linguists' limited conceptions of language and their inattention to feminist theory have impeded their understanding of language and gender.)

Cameron, D. (ed.) (1990) *The Feminist Critique of Language: A Reader*, London and New York: Routledge. (Extracts from a variety of sources, both literary and social scientific, including Lakoff's early writing and an interview with Irigaray.)

Coates, J. (1996) *Women Friends Talking*, Oxford: Blackwell.(Situated discourses and their detailed analysis.)

Coates, J. (ed.) (1997) *Language and Gender: A Reader*, Oxford: Blackwell.(Reprints many of the hard-to-get 'classics' on the topic.)

Connell, R.W. (1987) *Gender and Power*, Stanford, CA: Stanford University Press.(A social practice theory of gender; see especially chapter 4, 'The Body and Social Practice'.)

Eckert, P. and McConnell-Ginet, S. (1992) 'Think Practically and Look Locally: Language and Gender as Community-based Practice', *Annual Review of Anthropology* 21: 461-90; repr. in C. Roman, S. Juhasz and C. Miller (eds) *The Women and Language Debate: A Sourcebook*, New Brunswick, NJ: Rutgers University Press, 1994.(Develops social practice view of language and of gender discussed in §2.)

Frank, F.W. and Treichler, P.A. (eds) (1989) *Language, Gender and Professional Writing: Theoretical Approaches and Guidelines for Nonsexist Usage*, New York: Modern Language Association.(Includes several articles on how meanings develop in discourse and are sanctioned institutionally, as well as intelligent and detailed guidelines accompanied by discussion of the complexities of language-use.)

Goodwin, M.H. (1990) *He-Said-She-Said: Talk as Social Organization among Black Children*, Bloomington, IN: Indiana University Press.(Draws on ethnomethodological theory to argue that social structure is constituted in large part by the conversational structures community members produce in their daily interactions, examples of

which are analysed in detail.)

Hall, K. and Bucholtz, M. (eds) (1995) *Gender Articulated: Arrangements of Language and the Socially Constructed Self*, London and New York: Routledge. (Papers are all written for this volume and include Lakoff's recent work, an excellent anthropological essay by S. Gal, an essay by Eckert and McConnell-Ginet developing and applying the theoretical framework introduced in their 1992 article, and many other diverse and detailed studies of language-use.)

Holmes, J. (1995) *Women, Men and Politeness*, London: Longman.(Draws on empirical studies she and others have done of politeness and gender in language-use, noting especially multiple functions of linguistic forms.)

Irigaray, L. (1977) *Ce Sexe Qui N'En Est Pas Un*, Paris: Éditions de Minuit; trans. C. Porter with C. Burke, *This Sex Which is Not One*, Ithaca, NY: Cornell University Press, 1985.(The essay by the same name gives a short and relatively accessible introduction to her thinking.)

Lakoff, R. (1975) *Language and Woman's Place*, New York: Harper & Row.(An influential argument that women are victimized by language.)

McConnell-Ginet, S., Borker, R.A. and Furman, N. (eds) (1980) *Women and Language in Literature and Society*, New York: Praeger; repr. New York: Greenwood, 1994.(First collection to bring together empirical and theoretical work from social sciences, literature and philosophy: see especially article on defining 'anger' by N. Scheman.)

Roman, C., Juhasz, S. and Miller, C. (eds) (1994) *The Women and Language Debate: A Sourcebook*, New Brunswick, NJ: Rutgers University Press.(Collects a number of papers from literary, psychoanalytic and social scientific authors.)

Tannen, D. (1990) *You Just Don't Understand: Women and Men in Conversation*, New York: William Morrow. (An anecdotal best-seller focusing on supposed sex differences in communicative styles.)

Vetterling-Braggin, M. (ed.) (1981) *Sexist Language: A Modern Philosophical Analysis*, Totowa, NJ: Littlefield, Adams & Co.(This collection contains many philosophical discussions of language-gender issues, including a debate on the ethics of 'Ms', along with more linguistic contributions, such as V. Valian's critique of Lakoff's failure to distinguish a language system from its uses.)

Wright, E. (ed.) (1992) *Feminism and Psychoanalysis: A Critical Dictionary*, Oxford: Blackwell.(See especially the entry 'Language', by J.F. MacCannell, and its accompanying bibliography.)

Language of thought

The 'language of thought' is a formal language that is postulated to be encoded in the brains of intelligent creatures as a vehicle for their thought. It is an open question whether it resembles any 'natural' language spoken *by anyone. Indeed, it could well be encoded in the brains of people who claim not to 'think in words', or even by intelligent creatures (for example, chimpanzees) incapable of* speaking *any language at all. Its chief function is to be a medium of* representation *over which the* computations *posited by cognitive psychologists are defined. Its language-like structure is thought to afford the best explanation of such facts about animals as the productivity, systematicity and (hyper-)intensionality of their thought, the promiscuity of their attitudes, and their ability to reason in familiar deductive, inductive and practical ways.*

1 The language of thought hypothesis (LOTH)

One of Descartes' greatest challenges to the materialist theory of the mind was that it was impossible to imagine a physical device capable of the diverse intelligent behaviour often exhibited by human beings. Frege's development of formal logic, and then Turing's conception of a Turing machine offered the promise of a serious reply to this challenge, since they showed how at least deductive reasoning could be realized as a form of mechanically realizable computation (see Turing machines; Mind, computational theories of). Their proposals were an inspiration for the development of computational formalisms for various forms of non-deductive reasoning: induction, abduction, practical reason and decision theory.

But computations presuppose representations (Fodor 1975): the computations of Turing machines and many computers are, for example, often defined over *numerals* that represent *numbers*. A suggestion that has been advanced by a variety of philosophers and cognitive scientists (Harman 1973; Newell and Simon 1972; Fodor 1975, 1987) is that thinking requires precisely the sort of representations that are standardly used to *identify* thoughts, namely *sentences*: representations with logico-syntactic structure, of the sort defined recursively in logic texts in terms of names, predicates ('is bald'), variables ('*x*', '*y*'), connectives ('and', 'only if'), quantifiers ('all', 'some') and various operators ('probably', 'necessarily'). Thus, the thought that 'Necessarily God does or does not exist' is normally identified by using that sentence (or a translation of it).

The language of thought hypothesis (LOTH) is then the hypothesis that thinking consists in performing computations on sentences whose logico-syntactic parts are causally efficacious, for example, by being encoded in a creature's brain. The process is akin to theorem-proving in logic, except that, where the rules of logic are ordinarily applied by us *consciously following* the rules, according to LOTH the rules are applied by virtue of the causal structure of the brain. For example, where, in elementary logic, we follow the rule *modus ponens* - 'From "*p*" and "If *p* then *q*" derive "*q*"' - for LOTH the brain is so constructed that, if it is in a state that represents the premises, then it is (sometimes) caused to enter a state that represents the conclusion.

'Thinking' here is a generic word for processes involving propositional attitudes, and these are distinguishable in at least two ways that are often blurred in ordinary talk: by their *contents* (the sentence complement, for example, the that-clause, that follows the attitude verb) and by the different *relations* the agent may have to the same such contents, for example, *believing* or *hoping* that God exists (see Propositional attitudes). According to LOTH, different propositional attitudes of an agent involve different computational relations borne by the agent to sentences in a language of thought (LOT) that express the various thought contents of the agent. To a first approximation (see also Field 1981):

For any agent, *x*, and propositional attitude, *A* that *p*, there exists some computationally definable relation C_A such that:

$$x\ A\text{'s that } p \quad \text{iff} \quad \text{for some } \sigma\text{: } (xC_A\sigma\ \&\ \sigma) \text{ means that } p$$

For example, actively judging or desiring that God exists might be defined in terms of different computational relations, *J* or *D*, that an agent, *x*, might bear to a sentence, '$(\exists x)Gx$', that expresses the proposition 'God exists'.

LOTH must, of course, specify the computational relations for specific attitudes, as well as provide an account of how a symbol in the brain can have a specific meaning. For the first task it appeals to the familiar flow-charts of cognitive psychology. Judgment, for example, might be the output of perceptual and reasoning systems that is the

input to a decision-making one. For a theory of meaning, LOTH has turned to (combinations of) 'informational' approaches (Fodor 1987), teleological approaches (Neander 1995) and 'computational role' approaches (Field 1981).

Note that, *pace* Searle (1984; see also Chinese room argument), LOTH does not entail that syntactically defined sentences in a LOT do not have many real and important semantic properties; all that is claimed is that the relational clause and the causal processes of thinking are specifiable syntactically. However, there are some LOTH proponents (for example, Jackendoff 1987) who do seem to think that syntax would be enough to determine semantics, and others (for example, Stich 1983) who argue that the semantic clauses are not in principle determinate.

2 Arguments for LOTH

Although some philosophers (for example, Davies 1991) have ventured a priori arguments for LOTH, the chief arguments for it are that it provides the best explanation of at least the following phenomena (see Rey 1997: chaps 8-9 for a longer list).

The productivity of attitudes. People seem to be able to think a potential infinitude of thoughts, that is, to a first approximation, people can (in principle) think all permissible combinations of the primitive syntactic elements. For example, they can understand a conjunction of $n + 1$ sentences if they can understand *n* of them. Some theorists have baulked at the substantial idealization (from memory, mortality) that this involves, and so Fodor (1987) proposed a related, but more modest claim:

The systematicity of attitudes. Anyone who can think *p* can think any logical permutation of *p*: for example, if someone can think that 'Ann hates Bob only if Charles loves Di', they can also think that 'Charles loves Di only if Ann hates Bob', 'Di loves Charles only if Bob hates Ann' and so on for all permissible logical permutations. LOTH captures both productivity and systematicity by presuming that any system in which the logico-syntactic elements of sentences are causally efficacious is one in which they are readily available for recombination.

Rational and irrational relations among attitudes. Both deductive reasoning and many common fallacies are 'structure-sensitive', involving the scopes of operators, for example, negations, conditionals and quantifiers (see Scope). Quantifier scope is what distinguishes, for example, 'Everyone loves someone' form 'There is someone whom everyone loves'. Standard treatments of logic capture such structural facts in the terms of logical syntax. By insisting that that syntax is causally efficacious, LOTH is able to explain people's ability to reason, and why they are prone to certain errors (for example, misrepresenting a scope).

The (hyper-)intensionality of attitudes. Propositional-attitude ascriptions are 'intensional': terms that (even necessarily) refer to the very same thing cannot be substituted for one another without risking a change in the truth-value of the whole. There is a difference between thinking 'Water is wet', thinking 'H_2O is wet', and thinking 'The stuff of rain is wet', despite the fact that water = H_2O = the stuff of rain. LOTH distinguishes these attitudes by distinguishing syntactically between different symbolic structures to which an agent can be related (it can even distinguish them when the structures have the same 'meaning', as in *hyper*-intensional cases such as remembering that a fortnight is a fortnight as compared with remembering that a fortnight is two weeks).

The multiple roles of attitudes. Different attitudes can be directed at the same thoughts. People often *wish* for the *very same thing* that they *believe* does not presently obtain; they often come to *think* what they previously only *feared*. LOTH captures this by positing different computational relations to the same internal representation (or ones with the same content).

3 Some common objections to LOTH

Many people claim introspectively not to 'think in words', but, rather, for example, in mental images (see Imagery). Now, in the first place, LOTH is not meant to be establishable by introspection; it is purely an *explanatory* hypothesis. Moreover, by careful delineation of the different computational roles of specific representations, some (for example, Pylyshyn 1981) have argued that it can capture imagistic experiences, and even subjectivity and sensation (for example, Lycan 1990; Rey 1992).

But, second, it is difficult to think of a non-linguistic representational system with anything like the expressive power of a linguistic one. *Purely* imagistic systems, for example, do not seem adequate to represent logically complex thoughts - for example, negations, conditionals, nested quantifications - nor to distinguish thoughts about a general category (for example, cow) from ones about a particular instance (Elsie), for which the same image might serve.

Note that LOTH is not committed to the language of thought being confined only to creatures that speak a natural language, much less to the language of thought actually *being* a natural language. According to LOTH, any creature that*thinks* (as, for example, chimpanzees seem to do) will need a LOT, whether it *speaks* or not.

Wittgenstinians often object that LOTH presupposes the very sorts of processes it purports to explain: if there is a LOT, do we not need an ‘homunculus’ in the brain to read it? LOTH answers this objection by emphasizing Turing’s proposal for computation in general, whereby brute causation replaces human calculation.

But did Wittgenstein’s ‘private language’ argument not show that languages are necessarily public? (See *Philosophical Investigations* §§243-; Private language argument; Wittgenstein, L. §13.) Wittgenstein was concerned, however, only with languages whose references could not be publicly ascertained (for example, because they referred to ‘private sensations’). LOTH makes no such commitment.

Dennett (1987: ch. 3) argues that many attitude ascriptions are merely ‘interpretations’ of behaviour that do not commit us to a corresponding sentence in the head. For example, we might say of a chess-playing computer that ‘It likes to get its queen out early’, even though there might be no corresponding sentence manipulated by the computer’s program. However, LOTH is not committed to the literal truth of *every* ordinary ascription; only ones that figure in a causal explanation of such phenomena as those mentioned above.

The chief rival to LOTH is the so-called ‘radical connectionist’ models of cognition (in contrast to connectionist models that are merely implementations of ‘classical’ ones such as LOTH). Defenders of radical connectionism complain that LOTH models are too rigid, inefficient and not as neurophysiologically realistic as connectionist networks (see Smolensky 1988). However, the cognitively relevant physical properties of the brain have yet to be sufficiently identified to pass judgment on such a claim, and it has yet to be shown that radical connectionist theories can explain all the above phenomena as well as a language of thought does (see Connectionism).

See also: Mind, computational theories of

GEORGES REY

References and further reading

Davies, M. (1991) ‘Concepts, Connectionism, and the Language of Thought’, in W. Ramsey, S. Stich and W. Rumelhart (eds) *Philosophy and Connectionist Theory*, Hillsdale, NJ: Erlbaum.(An a priori argument for the language of thought.)

Dennett, D.C. (1987) *The Intentional Stance*, Cambridge, MA: MIT Press.(Essays critical of LOTH, presenting a less realistic, more ‘instrumental’ theory of mental states instead.)

Field, H. (1981) ‘Mental Representation’, in N. Block (ed.) *Readings in Philosophy of Psychology*, vol. 2, Cambridge, MA: Harvard University Press.(An extremely lucid presentation of the essential claims of LOTH, as well as of a conceptual role semantics for it.)

Fodor, J. (1975) *The Language of Thought*, New York: Crowell.(The most influential defence of LOTH as a psychological hypothesis.)

Fodor, J. (1987) *Psychosemantics*, Cambridge, MA: MIT Press.(A defence of LOTH on the basis of systematicity, as well as a development of a specific version of informational semantics for it.)

Harman, G. (1973) *Thought*, Princeton, NJ: Princeton University Press.(An early statement of a language of thought view, more from a philosophical than a psychological point of view.)

Jackendoff, R. (1987) *Consciousness and the Computational Theory of Mind*, Cambridge, MA: MIT Press.(A linguist’s attempt to work out in some detail a LOTH view, incorporating research in vision and language.)

Loewer, B. and Rey, G. (1991) *Meaning in Mind: Fodor and his Critics*, Oxford: Blackwell.(Collection of papers reacting to Fodor’s defence of LOTH, with replies by Fodor and an extended introduction by the editors.)

Lycan, W. (1990) ‘What is the ‘Subjectivity of the Mental’?’, in J.E. Tomberlin (ed.) *Philosophical Perspectives*, vol. 4, *Action Theory and Philosophy of Mind*, Atascadero, CA: Ridgeview, 109-30.(An attempt to account for subjective experiences by means of LOTH.)

Maloney, J. (1989) *The Mundane Matter of the Mental Language*, Cambridge: Cambridge University Press.(A clear and very readable defence of LOTH, with replies to many critics and consideration of problems outside its normal range, such as sensations and moods.)

Neander, K. (1995) ‘Misrepresenting and Malfunctioning’, *Philosophical Studies* 79: 109-41.(An evolutionary (teleosemantic) approach to the problem of assigning content to expressions in a language of thought.)

Newell, A. and Simon, H. (1972) *Human Problem Solving*, Englewood Cliffs, NJ: Prentice Hall.(An influential statement of the ‘classical’ LOTH approach in artificial intelligence.)

Pylyshyn, Z. (1981) ‘The Imagery Debate: Analog Media vs. Tacit Knowledge’, in N. Block (ed.) *Imagery*, Cambridge, MA: MIT Press.(Argues that LOTH can capture imagistic experiences.)

Rey, G. (1992) ‘Sensational Sentences Switched’, *Philosophical Studies* 67: 77-103.(An attempt to account for sensory experience, and the puzzle of inverted qualia, by means of LOTH.)

Rey, G. (1997) *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Oxford: Blackwell.(An introduction to recent work and an attempt to show that LOTH handles traditional problems in philosophy of mind.)

Searle, J. (1984) *Minds, Brains and Science*, Cambridge, MA: Harvard University Press.(A well-known attack on LOTH.)

Smolensky, P. (1988) ‘A Proper Treatment of Connectionism’, *Behavioral and Brain Sciences* 11: 1-23. (Influential critique of LOTH and suggestion of a connectionist alternative.)

Stich, S. (1983) *From Folk Psychology to Cognitive Science: The Case Against Belief*, Cambridge, MA: MIT Press.(A clear discussion of LOTH, and a defence of it as a purely syntactic account of thought.)

Language, ancient philosophy of

The earliest interest in language during the ancient Greek period was largely instrumental: presumed facts about language and its features were pressed into service for the purpose of philosophical argumentation. Perhaps inevitably, this activity gave way to the analysis of language for its own sake. Claims, for example, about the relation between the semantic values of general terms and the existence of universals invited independent inquiry into the nature of the meanings of those general terms themselves. Language thus became an object of philosophical inquiry in its own right. Accordingly, philosophers at least from the time of Plato conducted inquiries proper to philosophy of language. They investigated:

(1) how words acquire their semantic values;
(2) how proper names and other singular terms refer;
(3) how words combine to form larger semantic units;
(4) the compositional principles necessary for language understanding;
(5) how sentences, statements, or propositions come to be truth-evaluable;

and, among later figures of the classical period,

(6) how propositions, as abstract, mind- and language-independent entities, are to be (a) characterized in terms of their constituents, (b) related to minds and the natural languages used to express them, and (c) related to the language-independent world.

1 Pre-Platonic figures

Before Plato, evidence for inquiries into the semantic and syntactic features of language is sketchy, if suggestive. Even so, there is clear interest in language among pre-Platonic philosophers, rhetoricians and Sophists. In some cases this interest is motivated by a concern for developing persuasive speech, primarily for forensic or political use (see Rhetoric §1). In other cases, it has a more philosophical orientation. Thus, quite early, Parmenides (early to mid-5th century BC) sought to exploit a putative fact about language when defending an improbable form of monism: he argued that 'what is for saying and thinking must be' (fr. 6) and concluded that it is impossible to speak or think of what does not exist, with the consequence that it is equally impossible to speak or think of generation, change, or plurality, since each of these is implicated, directly or indirectly, in what is not. With the additional thesis that it is possible to speak and think of everything which exists (fr. 3), Parmenides concludes that the universe is an eternal undifferentiated unity, altogether remote from the appearances of plurality, change and generation delivered by the senses (see Parmenides §2). When Plato and Aristotle later exposed the fallacy in this argument, it was mainly by distinguishing syntactic and semantic features of the verb *to be* which Parmenides had failed to mark.

The Sophist Gorgias (late 5th century BC) similarly appealed to features of language when attempting to establish the nihilistic theses that:

(1) nothing exists;
(2) even if something did exist, it would be unknowable;
(3) even if something did exist and were knowable, it would be incommunicable.

On behalf of the odd counterfactual (3), in particular, he appealed to a subjectivist form of meaning: if something were perceived, it could be communicated only by means of a sign, which, even on the untenable assumption that it could be made to represent the supposed object of perception perfectly, could not be shared by more than one mind. Hence, communication about objects perceived is impossible (see Gorgias §3). However implausible this argument may sound, it is worth remarking that it appeals to a thesis about meaning that many find natural and attractive; indeed, many centuries later, Frege still felt the need to mount positive arguments against just its kind of semantic subjectivism (see Frege, G. §4).

The atomist Democritus (mid 5th-4th century BC) reportedly composed a treatise *On Words*. In this or another work he inaugurated an important discussion, taken up with great earnestness by later philosophers, concerning the signification of names. Democritus thus stakes out what appears to be the first case for a dedicated semantic

theory. He thus moves beyond the primarily instrumental interest in the semantic features of language displayed by his predecessors.

Democritus evidently argued for a form of *conventionalism*, according to which the relation between names and things named is determined by agreement or legislation. He argues that there cannot be a natural relation which ties a word to the world, since:

(1) there are homonyms, that is one name for different entities (for example, ’bank’);
(2) there are synonyms, distinct names for the same thing;
(3) entities can change names in the course of their existence.

Consequently, it would be wrong to suppose that a name is determined by anything other than an arbitrary designation.

In arguing this way, Democritus means to reject *naturalism*, a view according to which names and things named are related by some non-conventional relation. A first approximation of naturalism might be the view that all naming is at root onomatopoetic or somehow pictographic. Put thus crudely, naturalism may seem a non-starter. But other forms of naturalism, loose by comparison with these simpler versions, appeal instead to natural descriptions, with the result that names, especially proper names, all turn out to be disguised definite descriptions. Making naturalism more tenuous still, a fair number of naturalists were in the habit of appealing to the origin of names and other words, thereby confusing genetic and semantic questions.

Although in these varying contexts, the varieties of ’naturalism’ were unhelpfully conflated, the contrast between conventionalism and naturalism is important for the history of classical semantics for three related reasons. First, the ensuing debate, much of it quite subtle and sophisticated, is initially couched in terms of a contrast between conventionalism and naturalism, and even in late antiquity data to which naturalists appeal appear in a surprisingly forceful way. Second, part of the motivation for naturalism turns out to be reasonable: naturalism, unlike some varieties of conventionalism, seemed to capture the normativity of language. Some names are correct and fitting, while others are incorrect or otherwise inappropriate. It is hard to specify precisely how this can be the case if spoken sounds are applied to things in a purely conventional manner, especially when conventions may be generated quite locally. Third, whatever its ultimate faults, naturalism at least attempts to explain how reference occurs. In some of its ancient forms, conventionalism had difficulty even broaching a suggestion on this point.

2 Plato

The opposition between conventionalism and naturalism dominates Plato’s most explicit and extended treatment of language, the *Cratylus* (see Plato §15). The dialogue opposes equally untenable naturalist and conventionalist theories of naming, while at the same time promoting the virtues of each. Plato focuses initially on proper names, but it is clear that his interest extends to other singular terms and to general terms as well. Perhaps Plato means to explore facts about reference by laying out extreme forms of views which, though incompatible, each have some merit. At any rate, in the course of the dialogue he sets aside the most implausible features of conventionalism and naturalism without detailed comment.

These implausible features are, however, fully present in the initial presentations of each theory. Cratylus opens the dialogue, arguing on behalf of naturalism that it is possible for someone not to be named *n* even if ’all people were to call’ him *n* (383b6-7, cf. 429b12-c2). This is because, as we later learn, ’*n*’ is the name of something only if it bears a natural relation *R* to that thing, where *R* is specified by Cratylus only incompletely as a kind of mimetic or imitative relation (423b-428a, 430a10-b1). Thus, something qualifies as a name *n* only if it is correct, that is, only if it imitates the thing named and indeed indicates what that thing is (428e1-2).

Hermogenes responds on behalf of conventionalism:

> I cannot be persuaded that there is some correctness (*orthotēs*) of names other than convention and agreement. For it seems to me that whatever name someone gives to something, this is the correct one…. For no name is suited by nature to anything, but rather by the custom and habit of those in the habit of using it and calling things by that name.
>
> (*Cratylus*, (384c10-d9)

Hermogenes' conventionalism may seem initially more plausible than any variety of naturalism. Yet he is taken to task for failing to consider external constraints on convention. More importantly, names have the dual function of conveying information and distinguishing real natures in the world 388b7-11); if a name misses or obscures these real natures, then it will not succeed as a name, or at any rate will be at best only a sub-optimal name. Plato suggests that even this degree of normativity is incompatible with the version of conventionalism Hermogenes espouses.

In opposing this kind of conventionalism, Plato does not adopt Cratylus' naturalism. On the contrary, the Socrates of the dialogue, who evidently represents Plato's point of view, is equally critical of a naturalism which fails to distinguish between the correct name of an entity and the name which people actually employ when designating that entity. Socrates thus twice embarrasses Cratylus by forcing him to allow that his theory cannot distinguish between successfully referring to an object and referring to it correctly (429c3-5 and 429e8-430a5). The point, updating Plato's example a bit, is this: if a child points to a whale, exclaiming, 'That fish is bigger than our house!', the child has evidently successfully referred to the whale, but only incorrectly. Cratylus must allow that not only has the child failed to designate the whale, but also may have produced a false sentence by referring to a minnow in the whale's neighbourhood.

Plato, in the end, agrees with the naturalist that there must be some relation *R* which in a direct or indirect way relates a name to a thing named; but he denies that *R* can be specified in a way which wholly ignores convention. So, he equally agrees with the conventionalist that convention is relevant to determining names; but he denies that convention by itself can determine *R*. For he believes that names are information-bearers which, if correct, reflect the structure of the extra-linguistic world. Thus, the argument of the *Cratylus* has the effect of distilling what is right from both conventionalism and naturalism, and of laying down the constraints for a semantically and metaphysically adequate account of reference. It also advances the debate by distinguishing different ways in which a name may be said to be correct or incorrect: a name may be correct in a pragmatic way by fulfilling its reference-fixing function, or correct in a descriptive way by carrying information which accurately reflects the world. Plato appreciates that these functions may come apart in ways which make the job of specifying *R* exceedingly difficult. He himself does not, however, offer an articulated account of the no doubt complex relation *R*. Instead, the dialogue ends in perplexity.

Although it contains his most self-conscious and sustained treatment, the *Cratylus* hardly exhausts Plato's views on naming or language more generally. Of special note are his complex discussions of the various syntactic and semantic functions of the verb *to be*, conducted most thoroughly in the *Parmenides* and *Sophist*, both late dialogues. In the (*Parmenides* (142a, 161e-162b), raising a puzzle about saying truly '*a* is not' (we cannot predicate something of *a* unless *a* is; but then we cannot do what we plainly can do, namely deny the existence of something), Plato takes up a discussion about negative existentials which continues even today (see Existence). When, in the *Sophist*, he responds to the Parmenidean argument for monism, Plato quite rightly focuses on distinct uses of 'is'. The distinction remains an important philosophical tool for recognizing and diagnosing fallacies.

3 Aristotle

Aristotle approaches the study of language with a logician's eye. He seeks to determine how terms relate in syntactic structures of various sorts, mainly to determine and regiment the correct and incorrect forms of inference-drawing. Accordingly, he investigates the nature of meaning, the difficulties of reference, and the errors which result from failing to attend to what he calls the *homonymous* uses of words, that is the uses of words with a plurality of distinct but connected meanings, including especially core philosophical terms whose non-univocity may initially elude us. He thinks that when they seek unified definitions of these sorts of terms - at any rate, non-disjunctive definitions given as necessary and sufficient conditions - philosophers ignore the complexity of the concepts or properties they are used to express. Natural language sometimes reflects this complexity; at other times, it obscures it.

Aristotle investigates linguistic phenomena primarily in the *Categories* and *De Interpretatione*, but also in the *Metaphysics*, the *Topics*, the *Sophistical Refutations*, and the *Prior* and *Posterior Analytics*. He does not always explicitly connect these investigations, but there are several bridge passages that indicate how he takes his views in these areas to be related. Perhaps the single most important such passage opens the *De Interpretatione*, where Aristotle draws a fairly complete analogy between words and sentences on the one hand and types of affections

(*pathēmata*) in the soul on the other.

According to this analogy, there is a reasonably straightforward relation between thoughts and spoken sounds: individual words are to assertoric sentences as individual thoughts are to compound thoughts. The first members of these pairs are without truth value, being in some sense semantically atomic. The second members are by contrast necessarily either true or false. That is, Aristotle claims that bivalence obtains for all simple assertoric sentences and their mental analogues, compound thoughts.

Partly in virtue of this analogy, the opening of the *De Interpetatione* contains the seeds of four related Aristotelian semantic theses, each of which is developed in various places in the corpus:

(1) *Compositionality.* The semantic value of assertoric sentences is a function of their sub-sentential semantically relevant parts (semantically relevant, because, as he points out *De Interpretatione* 16a20-22), neither 'ton' nor 'dent' contributes anything to the meaning of 'Clinton was unusually well educated for an American president.');
(2) *Conventionality.* The written and spoken symbols used to stand for thoughts are conventional, whereas the thoughts themselves and that for which they stand are not;
(3) *Relationalism.* Conventional semantic units (written marks and spoken sounds) receive their semantic significance from those things of which they are symbols;
(4) *Signification.* The relationship in virtue of which they receive their semantic significance consists in or involves what Aristotle calls signification (*sēmainein*).

Of these, (2) and (4) merit special attention.

The second thesis, conventionalism, indicates where Aristotle stands on the question explored in the *Cratylus*. He explicates his contention thus:

> I say <a name is a significant sound> according to convention (*kata sunthēkēn*) because no name is by nature <significant>, but only when it has become a symbol. Even though inarticulate noises, e.g. those belonging to wild beasts, do reveal something (*dēlousi ti*), none of them is a name.
>
> (*De Interpretatione* 16a26-29)

He rejects naturalism, but without endorsing any simple version of conventionalism. Instead, he lays down the constraint that a sound becomes a name only when it has become a symbol, where something is a symbol only when it stands in an appropriate relation to a non-conventional mental representation.

With respect to Aristotle's approach to language and meaning, the fourth thesis, regarding signification (*sēmainein*), is easily the most central and important. It is also the most complex and difficult. Aristotle often writes as if signification were a simple meaning-relation: words signify things, even when they lack referents. Thus, for example, after distinguishing sharply between assertoric expressions, which alone have truth-value, and their constituents, which are not yet truth-evaluable, he imagines someone objecting that the word 'goatstag' (*tragelaphos*) is already false, since there are no such creatures. He responds that although 'goatstag' signifies something, it is not yet true or false, precisely because it is not yet part of an assertoric sentence (*De Interpretatione* 16a16-18). Aristotle's response correctly drives a wedge between vacuous reference and falsity; it also seems to treat signification as closely akin to sense expression: 'goatstag' has sense, but lacks reference (see Sense and reference).

Still, a complication arises for two related reasons:

(1) Aristotle sometimes appeals to signification where it seems unlikely that it can be understood in terms of sense expression;
(2) Aristotle sometimes denies signification to vacuous singular terms.

In the first case, Aristotle surely maintains that not only words signify: the word 'man' signifies rational animal, but the entity man signifies rational animal as well (*Categories* 3b10-23; *Topics* 122b16-17, 142b27-29; *Posterior Analytics* 85b18-21; *Metaphysics* 1017a22-27, 1028a10-16); further, clouds signify rain and smoke signifies fire (*Posterior Analytics* 70a10-38).

Elsewhere, he denies that a single word made to mean 'manandhorse' has signification (*De Interpretatione*

18a19-27). The probable resolution is that the signification relation is for Aristotle in some contexts semantic and in some contexts not; further, in some contexts, signification is more akin to reference and in others more akin to sense expression. In some more technical contexts, Aristotle introduces signification as a sort of essence-specification: man signifies rational animal because humans are essentially rational. Perhaps in this sense, though, 'signification' behaves all the more like the English word 'meaning': smoke means fire; 'ghoul' means 'a spirit that robs graves and devours the corpses'; and we say that although he speaks of piety, Euthyphro does not even know the meaning of the term. In some philosophical contexts, we are appropriately technical about how we understand the concept of meaning; in some other contexts, we are comparatively relaxed. This practice mirrors Aristotle's treatment of signification (see Semiotics).

4 The Stoics and other Hellenistic movements

The Hellenistic schools which flourished after the death of Aristotle became increasingly technical and specialized in their treatments of language. This is especially true of the Stoics, whose interests in logic, grammar and syntax led them to offer deeply subtle theories replete with technical vocabularies capable of an unprecedented richness and precision. The most intricate and important innovations were most likely introduced by Chrysippus.

It is difficult to recapitulate elements of the Stoic system briefly. To begin, the evidence is fragmentary, deriving from many different sources, some of them hostile and most of them secondhand. Moreover, there is a fair amount of divergence within the Stoic camp itself; it is worth remembering that we are dealing with the data of several centuries of philosophy. That said, the doxographer Diogenes Laertius provides a serviceable overview of the main features of Stoic semantic theory:

> Utterance (*phōnē*) and speech (*lexis*) differ, because while vocal sound is also an utterance only articulate sound is speech (*lexis*). And speech differs from language (*logos*), because language is always significant (*sēmantikos*) whereas speech <can> also lack significance, for example, 'blityri'; language can in no way <lack significance>. Moreover, saying (*to legein*) differs from voicing (*to propheresthai*). For while utterances are voiced, what is said are states of affairs - which turn out to be things said (or things which can be said, *lekta*).
>
> (Diogenes Laertius 7.57)

The initial view here is straightforward. Animals, including humans, make noises; but only some of these noises are meaningful. Those noises which express *lekta* are meaningful, or *sēmantikos*, and those which do not are mere sounds. If I speak German and you do not, the utterance '*verkehrt*' will be meaningful for me, but not for you; but even if we both speak German, the nonsense utterance 'blityri' will be meaningless for us both. The Stoics suggest, then, because some noises are significant and others are not, we must suppose that some noises have meanings, namely express *lekta*, while others do not. Hence, it is necessary to postulate the existence of *lekta*.

The primary motivation for the introduction of *lekta* is, then, semantic. For better or worse, with the Stoics we have the first self-conscious reification of meanings as such. Unsurprisingly, the Stoics were quick to exploit a second semantic function of *lekta*. Some *lekta*, called by the Stoics *axiōmata*, are the principal bearers of truth and falsity. According to the Stoics, then, one species of *axiōmata* can be true or false, evidently in a primary way. Our speaking truly or falsely, that is our uttering sentences with truth values, depends upon our uttering a sentence which expresses a complex of meanings, an *axiōma*. Hence, the sentence 'Dion is walking' is true just in case it expresses the proposition, held by the Stoics to be noncorporeal, *that Dion is walking* and that proposition is true. Just as words are correlated with meanings or significates, which the Stoics regard as incomplete *lekta*, so the complete sentence 'Dion is walking' gains its meaning and truth-evaluability by expressing the complete *lekton*, the *axiōma* or proposition which is true independent of its being expressed. Just how determinate Stoic thinking about propositions is remains disputed, since some reports have the Stoics introducing propositions which change truth-values and which, though noncorporeal, can perish, or go out of existence. Often enough, though, in these cases the Stoics turn out to be grappling with quite subtle problems generated by indexicals and demonstratives. Indeed, they often show more sensitivity to these problems than their critics and later expositors display when discussing their views (see Stoicism §8).

The Stoics are by no means alone in offering important advances in thought about language during the Hellenistic period. Sextus Empiricus certainly develops semantic themes. Additionally, the Epicureans, and to a lesser extent the Academics, offer treatments of meaning and language understanding, often reverting to earlier debates about

natural versus conventional signification. Later, the eclectic physician and philosopher Galen (AD 129-*c*.210) returns to the naturalism-conventionalism debate, ridiculing the naturalist propensity for etymology as 'fine friend' which is also an 'impostor'. No other Hellenistic school rivals the Stoics, however, in their genuinely innovative and impressively technical handling of semantic and syntactic matters.

See also: Dialectical school; Language, Indian theories of; Language, philosophy of; Nature and convention; Proper names; Propositions, sentences and statements; Reference

CHRISTOPHER SHIELDS

References and further reading

Annas, J. (1982) 'Knowledge and Language: the *Theaetetus* and the *Cratylus*', in M. Schofield and M. Nussbaum (eds) *Language and Logos*, Cambridge: Cambridge University Press.(An exploration of some overlooked connections between Plato's approach to naming in the *Cratylus* and the difficult dream passage of the *Theaetetus*.)

Aristotle (*c.* mid 4th century BC) *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984, 2 vols.(Although the focus of several works, Aristotle's attention to semantic matters pervades much of the corpus.)

Atherton, C. (1993) *The Stoics on Ambiguity*, Cambridge: Cambridge University Press.(A comprehensive treatment of Stoic approaches to ambiguity, with a special emphasis on Stoic classifications of forms of ambiguity and their role in fallacious argumentation.)

Bolton, R. (1976) 'Essentialism and Semantic Theory in Aristotle', *Philosophical Review* 85: 514-44.(An investigation into Aristotle's approach to nominal definition and the semantics of natural-kind terms.)

Brown, L. (1986) 'Being in the *Sophist*, a Syntactical Enquiry', *Oxford Studies in Ancient Philosophy* 4: 49-70.(A clear and accessible investigation into Plato's treatment of being in the *Sophist* which challenges some orthodox interpretations.)

Everson, S. (ed.) (1994) *Companions to Ancient Thought 3: Language*, Cambridge: Cambridge University Press. (An extremely useful anthology of papers on many aspects of ancient philosophical thought about language, with a special emphasis on continuities between ancient and contemporary concerns. This work should be consulted for additional bibliography.)

Fine, G. (1977) 'Plato on Naming', *Philosophical Quarterly* 27: 289-301.(An exceptionally clear and forthright account of Plato's approach to naming.)

Irwin, T. (1982) 'Aristotle's Concept of Signification', in M. Schofield and M. Nussbaum (eds) *Language and Logos*, Cambridge: Cambridge University Press.(A challenge to the widely held view that signification in Aristotle is a kind of meaning-relation; instead, Irwin argues, signification is a technical relation of essence-specification.)

Kretzmann, N. (1971) 'Plato on the Correctness of Names', *American Philosophical Quarterly* 8: 126-38.(A detailed and highly nuanced investigation of Plato's *Cratylus*.)

Long, A. and Sedley, D. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press, 2 vols. (The first volume contains translations of the principal fragments of Hellenistic philosophy, together with commentary. The second volume contains original texts and a very useful bibliography.)

Nuchelmans, G. (1973) *Theories of the Proposition: Ancient and Medieval Conceptions of the Bearers of Truth and Falsity*, Amsterdam: North Holland.(A full assessment of the evidence for the introduction of propositions in ancient thought.)

Plato (390s-347 BC) *Plato: Complete Works*, ed. J.M. Cooper, Indianopolis, IN: Hackett Publishing Company, 1997.(This edition contains the best translation of the *Cratylus*, together with the rest of Plato's works, many of which contain sub-sections relevant to his views on language.)

Shields, C. (1998) *Order in Multiplicity: Homonymy in the Philosophy of Aristotle*, Oxford: Oxford University Press.(A full account of Aristotle's approach to homonymy, discussed within the broader context of his semantic theory.)

Language, conventionality of

When we say that smoke means fire or that those spots mean measles, we are noting how the presence of one thing indicates the presence of another. For these natural relationships to continue, it is enough that the laws of nature remain the same. The connection between the two states is strictly causal. By contrast when we say, 'In English, "gold" means this stuff', pointing at some metal, we are insisting on an arbitrary connection between a piece of language and part of the world. We might have used another word, as other languages do, or have used this word for something else. But, for a word to have the literal meaning it does in a language, this arbitrary connection must be sustained on subsequent occasions of use. What is needed to sustain the connection is an intention on our part, not just the continued operation of natural laws.

Of course, some connections between words and things are based on natural relations; there is, for example, onomatopoeia. However, few words have this feature. For the majority of words it is quite arbitrary that they have the meanings they do, and this has led many to suppose that the regularities needed to sustain the connections between words and what they stand for are conventional rather than causal. But there are also those who deny that convention is an essential feature of language.

1 Language as governed by conventions

Words do not have intrinsic meanings. They mean what they do because speakers have given them this meaning (see Meaning and communication). But how do they do this? At first we might focus on what a speaker means by their use of words on a given occasion. This may be fixed by their communicative intentions. But for words to have a literal meaning in a public language they must have the same meaning from one occasion to the next, and many have supposed that this can be established only by conventional practice. Conventions, it is claimed, establish the regularities that give words the meanings they have for groups of speakers. So while a speaker may succeed in communicating something by their utterance of an expression on a given occasion, its literal meaning among a population of speakers is what it is conventionally used to communicate on each occasion of use.

Among those who think the appeal to conventions is essential in the study of language, there is room for disagreement about the precise explanatory role they play. Conventions are arbitrary practices that can be defined nonlinguistically; and for some this provides a reason to think they can serve in an analysis of the concept of literal meaning. The analysis proceeds in two stages. If we can explain how a speaker succeeds in meaning something by their use of words on a particular occasion then we can try to explain how this is established as the conventional meaning of those words in the common language of a given population.

An example of this programme is provided by Stephen Schiffer (1972). Schiffer begins with Grice's analysis of speaker-meaning in terms of the speaker's beliefs and intentions to explain what an expression means on a single occasion of use (see Grice, H.P.; Communication and intention). He then argues that we must appeal to conventions to show how speaker-meaning can give rise to the regular and literal meaning of expressions in a public language. A sequence of sounds or marks will have the literal meaning it does among a group of speakers when there is a convention to use those sounds or marks with a particular speaker-meaning. It will no longer be necessary to work out the speaker's meaning each time; the prevailing conventions among a population of speakers will make certain sounds and marks meaningful to them but not necessarily to others, and languages will be patterns of conventional regularities sustained by the practices of particular groups.

A different explanatory role for conventions is proposed by David Lewis (1975). Conventions governing use do not constitute language or determine meaning but they do establish which language is the actual language used by a population. For Lewis, *languages* are well-defined abstract objects - sets of expression-meaning pairs - which exist independently of speakers. We can characterize a language as a function which maps strings of signs to meanings. Among populations of humans there will exist conventions of *language* (in a second and more general sense of the term). These will be regularities in verbal behaviour depending on the beliefs and actions of people in those populations. These conventional regularities determine which of these abstract objects is the actual language used by that population. A population of speakers can be said to use a given language *L* just in case the meaning an expression has in *L* is the meaning it is conventionally taken to have in that population. The account is not reductive, however, for it assumes that speakers already know the meaning of sentences in a given language and

choose to conform to these. But it does offer an analysis of the conventions needed for members of a population to speak the same language. This goes via Lewis' general analysis of conventions which runs as follows.

A regularity *R* is a convention in a population *P* if and only if:

(1) Everyone in *P* conforms to *R*.
(2) Everyone in *P* believes the others are conforming to *R*.
(3) The belief that the others are conforming to *R* gives each person in *P* a good reason to conform to *R*.
(4) People in *P* prefer general conformity to *R* to less than general conformity to it.
(5) *R* is not the only regularity it would be possible for everyone in *P* to conform to.
(6) Conditions (1) to (5) are mutually known to members of *P*.

On this model it is in our common interest to conform to the regularities and abide by the conventions of using a given language. This facilitates cooperation and secures communication. We can coordinate our activities by communicating reliable information to one another by using the same words. This gives us a motive for wanting the regularities to be upheld.

The actual language relation is specified as follows. For a language *L* to be the actual language used in *P* is for there to be a convention of truthfulness and trust in *L*; that is, a convention of using particular sentences in *L* to make statements only if members of *P* believe them to be true, and to take statements other people make in *L* to be a basis for forming beliefs about what is true. To do this speakers must use sentences in accordance with their meanings in *L* just so long as other members of *P* do so. In this way Lewis seeks to reconcile the formal study of languages with activities of particular populations of human language-users.

Followers of Wittgenstein would reject both of the above accounts, criticizing as mysterious the idea of conventions correlating expressions with meanings *qua* entities, and rejecting the idea that meaning in a speech community can be explained in terms of a prior account of what individuals mean by their words on particular occasions. Once we accept that the meaning of expressions is determined by communal use, no appeal to Platonic entities is necessary. But since the rules that govern linguistic practice are determined by the community as a whole, an individual's understanding of language must be parasitic on communal practice. In contrast to the attempted reduction of literal meaning to conventions and speaker-meaning, an individual can only use language meaningfully when participating in the shared practices of a linguistic community. Thus Michael Dummett (1991) thinks that in using words individual language-users must hold themselves responsible to the standards of use of the language to which those words belong. On this view, a language is constituted by the conventional practices and agreed standards of usage. In addition, Dummett believes there are overarching conventions governing the overall aim of language-users to aim at the truth, as well as conventions governing the force (for example, assertoric, interrogative, imperative, and so on) with which they make their utterances.

Dummett offers an argument for the view that language must be governed by communal conventions. It concerns use of terms where we rely on experts to determine their precise meaning. The claim is that I can use the words 'fullback' or 'leveraged buyout' meaningfully while knowing little about football or finance because I rely on there being others who can use these words with greater precision. This is the phenomenon Hilary Putnam (1975) calls the division of linguistic labour (see PUTNAM, H. §3). It depends on a speaker making use of a word that has both an everyday and a more technical meaning, deferring to experts to fix the precise application of the term. However, this feature of language, although pervasive, is not essential to the existence of meaningful speech and we could get by without it. It does not show that there must be conventions of use among, or deference to, a set of experts for me to use words meaningfully.

2 Problems with convention-based accounts

In a revision to his earlier thinking, Schiffer (1987) argues that the reductive analysis of literal meaning is unlikely to succeed since every account of the notion of speaker-meaning we have devised fails. He makes powerful criticisms of Lewis' account of the actual language relation. Speakers can understand sentences of their language they have never heard before: there are potentially infinitely many of these. Clearly, there are no conventions governing the 'still to be used' part of a language. So unless the part we use fixes the unused part there will be nothing to determine which language is the actual language of a group of speakers. It is implausible that speakers

know a function which maps an infinity of sentences onto meanings. But perhaps what they know about the conventionally governed part of the language provides the resources to use and understand the rest. This would require a convention-based account of the expression-meaning of subsentential items, and the rules for their combination. Yet it is equally unlikely that speakers knowingly subscribe to conventions governing each word and grammatical construction ever used. A more plausible suggestion is that speakers have internalised a grammar of the language and this provides the semantic and syntactic resources to comprehend any sentence belonging to it. If this is correct, then why should we think it is conventions rather than psychological mechanisms that relate speakers to their languages? Schiffer argues that just as many difficulties confront us in attempting to relate speakers to grammars as to particular languages.

3 Language without conventions

Objections in principle have been raised to the idea that conventions have an essential role to play in the analysis of language. Donald Davidson and Noam Chomsky have both denied that language requires the existence of conventions. Davidson (1986) rejects the idea that conventions are necessary for linguistic competence. Chomsky (*1986*) rejects the idea that they are sufficient. For Davidson, Lewis-style conventions are too restrictive to explain successful communication. For Chomsky, they are insufficiently systematic to explain a speaker’s knowledge of language. (This was the basis of the criticism in §2.)

Both Chomsky and Davidson have opposed the idea of language as a social practice bound by a shared set of rules or conventions. Instead they argue that the fundamental notion of language is that of an individual’s language, or idiolect. For Chomsky, language has no existence outside the mind of the individual language-user. It is a body of knowledge represented in the mind/brain of the speaker. Davidson, on the other hand, views language as necessarily social, requiring the existence of others with whom we communicate. Despite their differences, neither accepts that we need to share a language in order to communicate.

For Davidson, meaning something by one’s words depends on one’s intention to be understood in a particular way and whether one succeeds. What we mean depends on what we can be understood to mean. This is possible even when speaker and hearer do not share the same meaning for a word, as the case of malapropism makes clear. Davidson goes further in conjecturing that there is no reason in principle why speakers who understand each other have to share any aspects of their language. What they would rely upon instead are general principles for interpreting one another’s behaviour: principles for ascribing meanings to each other’s words, so as to make rational sense of someone’s utterance given the beliefs and desires that best explain that person’s behaviour. This is not to deny that there are conventional aspects to language use, or that in actual speech situations people will make use of similar syntactic devices and enjoy overlapping vocabularies. But none of this is essential for communication. Even where conventions do exist in a language, it is possible to depart from them and still be understood. How far we can depart cannot be determined in advance nor settled by another set of conventions. Thus, Davidson argues, we need neither appeal to conventions nor stick to them to be understood, so they cannot play an essential role in linguistic understanding.

Chomsky agrees with Davidson that for a speaker and hearer to understand one another they do not need to agree on the meaning of words or abide by shared conventions. He departs from Davidson in rejecting the idea that success in communication is a condition of significant language use. Communication is never guaranteed and it does not enter into the conditions for something’s being a language. For Chomsky, language is determined by the cognitive psychological facts internal to the speaker’s language faculty. No two speakers will share the same idiolect; each will have their own internalized grammar and lexicon. If anything is shared it is the initial state of the language faculty characterized by the principles of universal grammar. For Chomsky, universal grammar, which comprises a set of structural principles true of all human languages, is a genetic endowment (see Language, innateness of). Speakers have no conscious or explicit knowledge of it, so cannot choose to conform to it. Nor, if Chomsky is right, is there any alternative to relying on it. Its operation cannot count, therefore, as a convention of language according to Lewis’ definition. For Chomsky it is universal grammar, and not the existence of conventions, that explains the possibility of human languages.

Whether one accepts Chomsky’s view of languages, his empirical findings point to difficulties for Lewis’ proposal. For if no two speakers ever share the very same vocabulary items or observe exactly the same rules of grammar there will be no formally precise, syntactic and semantic characterization of the language spoken by a

given community. On the other hand, what the regularity account of conventions correctly stresses is the social and public aspect of linguistic use and meaning, although in insisting on shared practices, or conformity to communal standards, it goes beyond what is necessary to ensure these. Davidson’s arguments show that meanings can be established publicly without having to be conventional. Conventions of language may abound but they have not been shown to play an essential role in language.

See also: Language, social nature of

BARRY C. SMITH

References and further reading

Chomsky, N. (1986) *Knowledge of Language*, New York: Praeger.(Chapters 1 and 2 offer a clear exposition of his psychological conception of an individual’s language.)

Davidson, D. (1986) *A Nice Derangement of Epitaphs*, in E. Lepore (ed.) *Truth and Interpretation*, Oxford: Blackwell.(Presents the case against conventions and shared public languages.)

Davies, M. (1981) *Meaning, Quantification, Necessity*, London: Routledge.(Chapter 1 presents a clear and detailed account of the analytical programme discussed in §1.)

Dummett, M.A.E. (1991) *The Logical Basis of Metaphysics*, Cambridge, MA: Harvard University Press.(Chapter 4 summarizes Dummett’s views about language, arguing against Davidson and Chomsky and in favour of a communal conception of language.)

Lewis, D.K. (1975) *Languages and Language*, in A. Martinich (ed.) *The Philosophy of Language*, Oxford: Oxford University Press, 2nd edn, 1990.(Key paper expanding on material discussed in §1.)

Loar, B. (1981) *Mind and Meaning*, Cambridge: Cambridge University Press.(A well-argued account of the project to define semantic notions in psychological terms with good observations of the difficulties noted in §2 for convention-based accounts of meaning.)

Putnam, H. (1975) ’The Meaning of "Meaning"’, *Collected Papers*, vol. 2, Cambridge: Cambridge University Press.(An important paper covering many themes beyond the scope of this entry. Of use here for the account of the division of linguistic labour.)

Schiffer, S. (1972) *Meaning*, Oxford: Oxford University Press.(A densely argued book with elaborate discussion of the analytic programme described in §1. Not elementary.)

Schiffer, S. (1987) *Remnants of Meaning*, Cambridge, MA: MIT Press.(A closely argued rebuttal of his earlier views and of any attempt to provide a theory of meaning. Challenging but not for beginners.)

Schiffer, S. (1993) ‘Actual Language Relations’, *Philosophical Perspectives* 7: 231-58.(A thorough assessment of attempts to define these relations.)

Language, early modern philosophy of

Philosophical interest in language during the seventeenth and eighteenth centuries was strong but largely derivative. Most thinkers shared Leibniz's view 'that languages are the best mirror of the human mind, and that a precise analysis of the significations of words would tell us more than anything else about the operations of the understanding'.

The three most important areas of philosophical discussion about language in the modern period were the nature of signification, the origin of human language and the possibility of animal language. Signification was generally viewed as a relation between linguistic signs and ideas. There was no agreement whether signification is entirely conventional or contains a natural element, but the view is that it is entirely natural virtually disappeared. Even those who retained the belief in the possibility of a philosophically perfect language insisted that such a language should be constructed anew, rather than rediscovered as the lost language of Adam. The traditional biblical account of the origin of language was more and more contested but, as more naturalistic theories emerged, the problem of why other animals cannot talk became especially pressing.

Debates about language in the seventeenth and eighteenth centuries were highly speculative; participants in these debates often relied on simplistic biological theories, inadequate grammars or anecdotal evidence from travellers. What makes these discussions important is less their scientific contribution than their engagement with the philosophical problems concerning the relationship between the human mind and the natural world.

1 The biblical and Lucretian accounts

Two classical accounts provided the intellectual background for modern philosophical debates about language. The first is the biblical story according to which the original names of beasts and birds were given by Adam. Renaissance speculation contends that these names were inspired by God, bore a natural relation to the creatures, and captured their essences. After the confusion at Babel, Adam's universal language was scattered into mutually unintelligible forms of speech, each of which began to decay. The second influential ancient story comes from Lucretius' *On the Nature of Things* (V.1028-90). According to Lucretius, language is an institution created by a community of humans, rather than a single name-giver. People came to form articulate sounds out of a need to coordinate their actions. At the beginning, the ability to communicate by words differed little from children's ability to indicate the object they desire by means of gestures, or animals' ability to indicate fear or pain by cries. As language developed, its structure became more complicated and its uses more diverse.

Perhaps the most criticized point of the biblical account is the claim that language originates with the naming act of a single person. Ideas from Hobbes' *Leviathan* (1651), Locke's *Second Treatise on Government* (1690) and Rousseau's *Du contrat social (The Social Contract)* (1762) provide the paradigm of how a social account of the origin of language might be framed. According to the contractarian view, social norms - whether they are laws of property or laws of meaning - are the result of explicit or tacit agreements among human beings. There is a striking parallel between Locke's criticism of the natural origin of the title to sovereignty in the *First Treatise on Government* (1690) and his opposition to the natural foundation of categorization in the *An Essay Concerning Human Understanding* (1689). Who is to be the heir is not to be decided through the study of the line of inheritance from Adam; what is to be gold is not to be decided through the study of how the word 'gold' descended from the original naming act of Adam. It is rationally grounded convention that settles these matters. Philosophers of the Enlightenment followed Locke's analysis and brushed away the biblical account. Condillac, Rousseau and Herder argued that even if language happens to be a divine creation, human beings had to reinvent it after the Flood or after they scattered over large uninhabited areas, leaving the biblical story at best irrelevant.

The most controversial point of the Lucretian account is the claim that the difference between the primitive signalling of animals and human languages is only a matter of degree. Descartes argued that the ability to understand an unlimited variety of expressions is distinctive of a mind, something that the smartest animal could never do, but which is no challenge for even the dullest human being. This shows 'not merely that beasts have less reason than men, but that they have no reason at all' (Descartes 1637: 58). In his objections, Gassendi protested against this conclusion, and insisted that animals do in fact speak. The Cartesian response - analysing animal signalling and the bodily aspects of human communication in physiological terms - was accomplished in great

detail by Géraud de Cordemoy. The conclusion was that since such an analysis provides a full account of animal signalling, but not of human language, the latter but not the former presupposes a mind (see Animal language and thought).

2 Empiricism: Hobbes

Although not primarily concerned with philosophy of language, Thomas Hobbes presented an influential and distinctively empiricist account of language in Chapter 4 of *Leviathan* (1651) and Part 1 of his *De corpore (On the Body)* (1655).

According to Hobbes, the purpose of speech is to 'transferre our Mentall Discourse, into Verbal; or the Trayne of our Thoughts, into a Trayne of Words' (1651: I.4). There are two reasons we make such a transfer: to record our thoughts for ourselves, and to communicate our thoughts to others. Words employed as mnemonic devices are *notae* (marks); words employed as means of communication are *signa* (signs) (1655: II.1-2). Hobbes holds that the first use of words is primary: if a person were alone in the world, he could create a private language and use words as marks, but not as signs ((1655: II.3).

Hobbes' philosophy of language has two main parts. The first is the account of the relationship between words and what they stand for, or the theory of *names* 1655: II); the second is the account of the connections names bear to each other, or the theory of *propositions* (1655: III).

Names signify conventionally: Hobbes thinks that even if the original names of animals come from God himself, he nevertheless chose those names arbitrarily (1655: II.4). What a name signifies is a *cogitatio* (thought), for the claim that a word like 'stone' is a sign for stones can only be understood as saying that 'he that hears it collects that he that pronounces it thinks of a stone' (1655: II.5). (Nonetheless, Hobbes often speaks of names as signifying things without the mind.) Hobbes is committed to the empiricist thesis that all human thinking is based on imagination. Since imagination itself is nothing but decaying sense, we cannot meaningfully speak of things that cannot be thought of as combinations of sensory images we had previously. Hence, for example, we cannot think of actual infinity: saying that something is infinite means merely that we are 'not able to conceive the ends, and bounds of the thing named' (1651: III.12).

Hobbes' semantics is purely nominalistic (see NOMINALISM §3). General terms do not pick out properties, but signify disjointedly every individual to which they apply (1655: II.9). In a proposition, both terms 'raise in our mind the thought of one and the same thing', and the copula 'makes us think of the cause for which those names were imposed on that thing' (1655: III.3). The proposition '*man is a living creature* raises but one idea in us, though in that idea we consider that first, for which he is called man, and next that, for which he is called living creature' (1655: V.9).

Conventionalism about signification and nominalism in semantics prepare the ground for conventionalism about truth. Truth is ascribed only to propositions, and it consists in the fact that the predicate is a name of everything that the subject is a name of. According to Hobbes, 'the first truths were arbitrarily made by those that first of all imposed names upon things, or received them from the imposition of others. For it is true (for example) that man is a living creature, but it is for this reason, that it pleased men to impose both those names on the same thing' (1655: III.8). It is controversial whether Hobbes consistently adhered to this radical view (see Hobbes, T. §3).

3 Rationalism: Port-Royal

Grammaire générale et raisonnée (General and Rational Grammar: The Port-Royal Grammar) (1660) and *La logique ou l'art de penser (Logic or the Art of Thinking)* (1662) articulate a rationalist conception of language sharply opposed to Hobbes' empiricism. Following Descartes, Arnauld and his collaborators (Claude Lancelot in *Grammar* and Pierre Nicole in *Logic*) reject not only empiricism, but conventionalism and nominalism as well.

The empiricist claim that all our thoughts originate in perception leaves unexplained how we attain ideas like those of being and thinking. Since - the authors of *Logic* insist - we do have such ideas, it follows that 'the soul has the faculty to form them from itself, although often it is prompted to do so by something striking the senses' (1662: I.1). The element of truth in conventionalism is that the association between a word and an idea is arbitrary. However, since the connections among ideas and between ideas and things are natural, we reason not about words

but ‘about the nature of things by considering ideas of the mind that people chose to mark by certain names’ (1662: I.1). Finally, while nominalists are right to claim that all existent things are singular, some entities are not things, but rather modes of things. Modes exist only in virtue of being instantiated in things, but they can be thought of independently of the things in which they inhere 1662: I.2, I.6, II.1).

The semantic terminology of *Grammar* and *Logic* is often loose. The authors talk about words signifying (*signifier*), expressing (*exprimer*) or marking (*marquer*) both things and ideas. (The relationship between ideas and things, however, is consistently called *représentation* (representation).) The principal meaning (*signification principale*) of a word or a sentence is the idea or proposition it expresses, but in order to capture its full meaning connotations must also be taken into account. These connotations are called *auxiliary ideas* (*idées accessoires*). For example, the principal meaning of the sentence ‘You are lying’ is simply ‘You know that the contrary of what you say is true’, but in addition to the principal meaning, these words also convey the idea of contempt (1662: I.14).

The principal meaning of a word can often be presented by an explicit definition. Following Aristotle’s distinction and scholastic terminology, Arnauld and Nicole distinguish between *nominal definition* and *real definition* (1662: I.12). A nominal definition captures a connection between a word and an idea; a real definition captures a connection between ideas, which in turn corresponds to a connection between real entities. Therefore, nominal definitions - at least in their purest form - are arbitrary and are not subject to rational criticism. On the other hand, real definitions in their purest form ‘do not depend on us at all, but on what is contained in the true idea of a thing’ (1662: II.16). But there are impure definitions as well. *Lexical* definitions are intended to capture the ordinary meaning of the word and are consequently discovered and not stipulated by lexicographers (1662: I.14). As lexical definitions are atypical nominal definitions, *descriptions* are atypical real definitions. Descriptions list a number of accidents which are sufficient to specify the extension of an idea, but not the common nature of things within the extension (1662: II.16).

According to Arnauld and Lancelot there are two major categories of words: those that signify the objects of thoughts and those that signify the manner of thoughts (1660: II.1, II.13). The most important subcategory within the first is that of nouns; within the second, verbs. Concrete substantival nouns (such as ‘sun’) signify substances, abstract substantival nouns (such as ‘whiteness’) signify modes (1662: II.1). An adjectival noun has two significata: ‘white’ signifies distinctly the mode of whiteness, and confusedly signifies white things. A verb is a word whose principal function is to indicate assertion, and the copulative verb ‘to be’ has only this principal function (1660: II.13). Other verbs also have a secondary role: they express ideas and thereby refer to entities, and they indicate the time with respect to which the assertion is made. For instance, the verb ‘lives’ in the sentence ‘Peter lives’ expresses the mode of being alive and indicates the assertion that this mode belongs to Peter at the time of the assertion (1660: II.13).

The mind judges by uniting or separating two ideas. The product of judging is a proposition. Propositions are *simple* if they have a single subject and a single predicate, and *compound* otherwise. Compound propositions include among others conjunctions, disjunctions and conditionals. Not all complexity in the subject or predicate terms indicates that the proposition is compound. For example, ‘The invisible God has created the visible world’ - or equivalently, ‘God who is invisible created the world which is visible’ - is a *simple but complex* proposition. For although it contains three propositions - ‘God is invisible’, ‘God created the world’ and ‘The world is visible’ - only the second of these is asserted, while the others are assumed or taken for granted ((1662: II.5; 1660: II.9). By contrast, the proposition ‘God is invisible and he created the visible world’ is compound, since it ascribes two different predicates to the same subject (see Arnauld, A.; Port-Royal).

4 Locke’s *Essay*

Although remaining firmly in the empiricist tradition, John Locke’s *An Essay Concerning Human Understanding* (1689) incorporates some of the insights of Cartesian rationalism. This synthesis makes the *Essay* the essential point of reference for all philosophy of language in the eighteenth century.

According to Locke, the primary purpose of language is communication. God or nature made our organs fit to produce articulate sounds, and through arbitrary imposition we are capable of making these sounds ‘Signs of internal Conceptions’ (Locke 1689: III.i.2; see also III.ii.1). At first glance, this view resembles Hobbes’, but - as

Leibniz remarks at the beginning of Book III of his *Nouveaux essais sur l'entendement humain (New Essays on Human Understanding)* (posthumously published in 1765) - there is an important difference: for Hobbes the private use of words is fundamental, whereas for Locke it is their use as signs for others.

The main thesis of Locke's theory of signification is that 'Words in their primary or immediate signification, stand for nothing, but the Ideas in the Mind of him that uses them' (1689: III.ii.2). Since the purpose of words is to invoke in the hearer's mind an idea identical (or at least sufficiently similar) to the idea the speaker has in mind, a word is bound to be intimately connected to an idea of the speaker. To say that primary signification is nothing but the speaker's idea does not mean that the word refers to that idea, or that in using a certain word the speaker is talking about the idea his word primarily signifies. Besides the primary signification of words, men often 'in their Thoughts give them a secret reference to two other things': the ideas in the minds of other speakers, and the reality of things (1689: III.ii.4). This supposition of *secondary signification* is indispensable for human beings, since 'without this double Conformity of their Ideas, they find, they should both think amiss of Things in themselves, and talk of them unintelligibly to others' (1689: II.xxxii.8). Locke apparently condemns only the assumption that words can directly signify the ideas of others and the reality of extramental things, without the mediation of their primary signification.

In accordance with the Cartesian tradition, Locke emphasizes that a genuine language cannot contain only singular names. We need signs which stand for more than one thing, since without these there would be too many (perhaps infinitely many) words to learn 1689: III.i.3; III.iii.2-4). With regard to the signification of these general terms Locke was a conventionalist. His primary target is the doctrine that general terms signify substantial forms, genuine universals in the things themselves that provide the ground for an objective classification of things. Since, he argues, our ideas are made up exclusively from the materials provided by experience (1689: II.xii.1), and since we cannot form ideas of substantial forms in this manner (1689: III.vi.10), substantial forms cannot be the significata of general words. The general term 'horse' signifies the *abstract general idea* of a horse, an idea that represents actual horses in virtue of representing certain sensible qualities they have in common. According to Locke, when we categorize things, we use the significata of our general terms. Fixing the significata of these terms is not a matter of discovering determinate species that exist independently of us, but rather a matter of reasonable choice from among the innumerable objective similarities of things (1689: III.iii.13).

In the case of ideas of substance, which represent natural unions of qualities, our forming of the appropriate abstract idea is independent of our linguistic abilities. The case is more complicated in the case of ideas of mixed modes, which represent collections of qualities that are not necessarily united in the same subject. For such complex ideas, although it is 'the Mind that makes the Collection, 'tis the Name which is, as it were the Knot, that ties them fast together' (1689: III.v.10). In the case of simple modes, like those of number, the dependence on naming is even more explicit: according to Locke, we could not have ideas of large numbers for which we have no names (1689: II.xvi.5).

Locke subscribes to the Port-Royal view that syncategorematic words - terms like 'is' and 'but', which he calls *particles* - stand not for ideas but for operations of the mind by which the ideas signified by other words are put together into a proposition. The truth of a proposition consists in the 'joining or separating of Signs, as the Things signified by them, do agree or disagree with one another' (1689: IV.v.2). Whether for Locke truth consists always in agreement and disagreement between ideas, or whether he allows for agreement and disagreement between an idea and what it represents, is a controversial question of Locke interpretation (see LOCKE, J. §§3-5).

5 Critique of Locke: Berkeley and Leibniz

The main tenets of Lockean philosophy of language were widely accepted among philosophers in the eighteenth century, but they did not remain unchallenged. The most important objections to the Lockean theory were made by George Berkeley and Gottfried Wilhelm Leibniz. Berkeley's critique was sketched in his *Philosophical Commentaries* (written between 1707 and 1708), and elaborated in *A Treatise Concerning the Principles of Human Knowledge* (1710), and his dialogue *Alciphron* (1732). Leibniz's objections take the form of a line-by-line critical commentary in his *New Essays on Human Understanding* (1765).

Berkeley's critique focuses on three major areas: the purpose of language, the semantics of general terms, and the nature of signification. He denies that communication is the chief end of discourse. Besides conveying our

thoughts to others, language is also used in ‘the raising of some passion, the exciting to, or deterring from an action, the putting the mind in some particular disposition’ (Berkeley 1710: Introduction 20). These are direct effects of speech, possibly achieved without the mediation of ideas. Berkeley’s remark concerning the use of the expression ‘good thing’ prefigures later emotivist treatments: the function of such expressions is to excite the appropriate feelings in us (1710: Introduction 20) (see Emotivism). Even expressions that - unlike ‘good thing’ - do signify ideas, can do so in the way the letters of algebra signify, where ‘though a particular quantity be marked by each letter, yet to proceed right it is not requisite that in every step each letter suggest to your thoughts, that particular quantity it was appointed to stand for’ (1710: Introduction 19).

Berkeley harshly criticizes a view he associates with Locke concerning the generality of words and ideas. For Berkeley, a general word stands for a class of particulars by signifying ‘an idea, which considered in itself is particular, [but] becomes general, by being made to represent or stand for all other particular ideas of the same sort’ (1710: Introduction 12). There is no need to assume that a single abstract idea is signified by each general term: by the mechanism of *selective attention* we can consider one or another feature of a particular idea, and then use it as a sign for all particular ideas that share the given feature. Berkeley’s criticism may miss its target. When Locke talks about abstraction, he also uses the language of considering certain features within a particular idea (for example, (Locke 1689: II.xi.9). Still, it remains unclear whether Locke thinks that abstract ideas are made from particulars via selective attention, or whether he in fact agrees with Berkeley that abstract ideas simply are particular ideas considered selectively.

Berkeley rejects the thesis that certain words have a secondary signification without the mind; things without the mind could be either material or immaterial, and Berkeley denies the existence of the former and the imaginability of the latter. Words like ‘horse’ or ‘white’ signify an idea without indicating anything beyond the idea; words like ‘soul’ or ‘God’ signify a substance without the mediation of an idea. A language is a network of signification relations between ideas which obtain in virtue of some customary connection. If the connections are the result of human imposition (for example, associating the idea of the letter ‘a’ with the idea of the corresponding sound, or associating the idea of the word ‘apple’ with some idea of an apple), the system of significations is an *artificial language*. If the connections result from divine imposition (for example, the connection between the ideas of fire and smoke, or between the visual idea of a square and a tactile idea of a square), the system is a *natural language* (Berkeley 1733: 40; see also Berkeley 1709: 140, 147). For Berkeley, the study of nature is the study of grammar for the divine language. Natural phenomena - which, of course, are merely collections of ideas - can signify one another, but they cannot stand in genuine causal relations to one another: it is God’s creating and sustaining act that causes everything to exist and be the way it is (1733: 13; 1710: 65, 108) (see Berkeley, G. §§5-6, 11).

Leibniz’s two main targets in the Lockean theory are claims wholeheartedly endorsed by Berkeley: that signification of words is arbitrary and that categorization is the work of the mind. According to Leibniz, signification is ‘settled by reasons - sometimes natural ones in which chance plays some part, sometimes moral ones which involve choice’ (Leibniz 1765: III.2). To support this thesis, Leibniz presents etymological links connecting contemporary words with earlier forms, and sometimes ultimately with onomatopoeia. What grounds these explanations is the principle of sufficient reason: brute accidents are as unacceptable in semantics as in physics.

Leibniz defends the semantic relevance of real definitions against Locke. Although ‘gold’ can be given a number of different nominal definitions, only a few of these will be in agreement with the genuine conceptual analysis. A real definition is grounded in a real species, and has the characteristic mark that it ‘displays the possibility of the definiendum’ (1765: III.3). Since he rejects the view that ideas we have must be completely transparent to us, he can resist Locke’s claim that our words cannot signify unknown features: ‘the name "gold"… signifies not merely what the speaker knows of gold, e.g. something yellow and very heavy, but also what he does not know, which may be known about gold by someone else, namely: a body endowed with an inner constitution from which flow its color and weight, and which also generates other properties which he acknowledges to be better known by the experts’ (1765: III.11). If pressed, we would have to come up with nominal definitions for substances; but these do not fix the meaning of the corresponding terms, since we recognize that ‘our definitions are all merely provisional’. We know that something that passes all our current tests for being gold may nevertheless fail to be gold, since ‘one might… discover a new assaying method which would provide a way of distinguishing natural gold from this artificial gold’ (1765: III.4, 6).

What underlies this insistence on real definitions is Leibniz's commitment to the ideal of a *universal characteristic*, a written language which could be understood by any rational being and which could perfect human reasoning itself. Corresponding to each complex concept or proposition, there is a complex symbol in the universal characteristic whose structure is the same as the structure of the concept or proposition signified. Occasionally, Leibniz talks about the primitive symbols of the universal characteristic as geometrical figures - like Egyptian hieroglyphs or Chinese characters - which bear some similarity to what they represent Liebniz 1875-90: IV.72-3). He also made attempts to assign numerical values to various concepts and to represent conceptual analysis as analogous to prime factorization (Couturat 1903: 245-6). But - as Leibniz himself realized - our concepts are so complex that we have no reason to believe that the process of resolving them into primitive constituents is a finite one. This does not mean that the system of numerical characteristic is useless. Since logic imposes constraints on possible assignments of numerical values to concepts, and since these constraints are expressible in the form of Diophantine equations, the system of numerical characteristic without the analysis of concepts can be used as a *formal calculus* to check the validity of inferences (1875-90: VII.205) (see Leibniz, G.W. §10).

6 The origin of language: Condillac and Rousseau

Questions about the origin of language were widely discussed in the seventeenth century but the views proposed did not go much beyond the biblical or the Lucretian accounts. This changed radically in 1746 with the publication of Étienne Bonnot de Condillac's *Essai sur l'origine des connaissances humaines (An Essay on the Origin of Human Knowledge)*.

Condillac was an enthusiastic but critical follower of Locke. His disagreement with Locke concerns two issues related to language. Condillac's first criticism is that Locke's theory about the mind is merely descriptive and fails to explain how understanding originates from sensation. The second is that Locke underestimates the role language plays in a theory of ideas. The two points are connected, since the 'the consideration of words, and of our manner of using them, might give some light into the principle of our ideas' and could yield 'a far better explication of the springs of the human understanding' (introduction to Condillac 1746).

Condillac claims that human beings can attain control over their thoughts solely in virtue of the use of *instituted signs*. Animals lack the ability to establish a conventional relation between an object and its sign, they cannot recall ideas at will, and therefore cannot prudently direct their behaviour. Animals are driven by instinct, which is 'no more than the imagination, as independent of our command, though by its activity, completely concurring to the preservation of our being' (1746: I.ii.95). In *Traité des sensations (Treatise on Sensation)* (1754), Condillac imagines a statue gradually brought to life by obtaining its five senses one by one. According to Condillac, the fully sensitized statue will be able to recognize its basic needs, remember things and learn from experience but, because it lacks speech, it will have only 'the habit of directing itself by ideas for which it does not know how to account' (1754: II.viii.35).

But it seems that in order to make use of instituted signs, we must already have sufficient reflection to choose them and to establish the conventional link between the signs and their significata (1746: I.ii.49). This is the problem that Condillac's account of the origin of language is supposed to answer. The explanation begins with natural signs - cries and gestures: these constitute the *language of action* (1746: II.i.1). The development of reflection and the development of language are intertwined: since both require the other, they evolve through a process of bootstrapping. Communicative intentions first arise due to the recognition of the accidental success of the language of action. From that point on, reflection and language interact to the mutual expansion and sophistication of both (see Condillac, É. B. de §§2, 3).

In the years following the appearance of Condillac's *Essay*, Diderot, Maupertuis and Turgot in France, Adam Smith and Lord Monboddo in England, Moses Mendelssohn and Johann Gottfried Herder in Germany each expressed their views on the origin of language. But perhaps the most significant contributions to the debate initiated by Condillac were Jean-Jacques Rousseau's *Discours sur l'origine et les fondements de l'inégalité parmi les hommes (Discourse on the Origin and Foundations of Inequality among Mankind)* (1755) and *Essai sur l'origine des langues (On the Origin of Languages)* (published posthumously in 1781). Rousseau replaces Condillac's relatively harmonious conception of the course of development with his emphasis on the conflict between human nature and social existence. He agrees that languages arise because of human needs, but he rejects the idea that these are physical needs. Physical needs would never have forced us to abandon the language of

gesture (Rousseau 1781: 1). What stimulated the first words were *moral needs*, passions. 'As man's first motives for speaking were of the passions, his first expressions were tropes' (1781: 3). That is, language and art were not separated at the beginning of history. As savage man turned into barbarian man and finally into civilized man, human needs changed, and as a consequence human languages lost their connection to music, becoming more and more rational and less and less passionate (1781: 19). And as societies assumed their final form, where 'no longer is anything changed except by arms and cash', languages have also become less favourable to liberty (1781: 20) (see Rousseau, J.-J. §2).

See also: Language, philosophy of; Language, Renaissance philosophy of; Universal language

ZOLTAŃ GENDLER SZABÓ

List of works

Arnauld, A. and Lancelot, C. (1660) *Grammaire générale et raisonnée de Port-Royal*, repr. Geneva: Slatkine, 1968; trans. J. Rieux and B.E. Rollin, *General and Rational Grammar: The Port-Royal Grammar*, The Hague: Mouton, 1975.(The Port-Royal account of the principles of grammar.)

Arnauld, A. and Nicole, P. (1662) *La logique ou l'art de penser*, ed. P. Clair and F. Girbal, Paris: Vrin, 1981; trans. J.V. Buroker, *Logic or the Art of Thinking*, Cambridge: Cambridge University Press, 1996. (The Port-Royal account of the principles of logic.)

Berkeley, G. (1707-8) *Philosophical Commentaries*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, London: Nelson & Sons, 1948-57, 9 vols.(Berkeley's previously unpublished early notebooks.)

Berkeley, G. (1709) *An Essay towards a New Theory of Vision*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, London: Nelson & Sons, 1948-57, 9 vols.(Berkeley's first published work: its subject matter is visual perception.)

Berkeley, G. (1710) *A Treatise Concerning the Principles of Human Knowledge*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, London: Nelson & Sons, 1948-57, 9 vols.(Berkeley's main philosophical work.)

Berkeley, G. (1732) *Alciphron, or the Minute Philosopher*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, London: Nelson & Sons, 1948-57, 9 vols.(Dialogue of philosophical, moral and theological content.)

Berkeley, G. (1733) *The Theory of Vision, Vindicated and Explained*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, London: Nelson & Sons, 1948-57, 9 vols.(Reply to criticism of Berkeley 1709.)

Condillac, É.B. de (1746) *Essai sur l'origine des connaissances humaines*, in G. Le Roy (ed.) *Oeuvres philosophiques*, Paris: Presses Universitaires de France, 1947-51, 3 vols; trans. T. Nugent, *An Essay on the Origin of Human Knowledge*, Gainesville, FA: Scholar's Facsimile and Reprints, 1971.(Condillac's main philosophical work.)

Condillac, É.B. de (1754) *Traité des sensations*, in G. Le Roy (ed.) *Oeuvres philosophiques*, Paris: Presses Universitaires de France, 1947-51, 3 vols; trans. G. Carr, *Treatise on Sensations*, Los Angeles: University of Southern California Press, 1930.(A treatise on the nature of sense perception.)

Descartes, R. (1637) *Discours de la méthode*, in C. Adam and P. Tannery (eds) *Oeuvres de Descartes*, Paris: Vrin/C.N.R.S., 1964-76, 11 vols.(The standard edition of Descartes' writings.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.(Hobbes' main work in political philosophy.)

Hobbes, T. (1655) *De corpore*, in W. Molesworth (ed.) *The Latin works of Thomas Hobbes of Malmesbury*, London: John Bohn, 1845, 5 vols; trans. W. Molesworth, *The English Works of Thomas Hobbes of Malmesbury*, London: John Bohn, 1839-45, 11 vols. (Hobbes' main work in natural philosophy.)

Leibniz, G.W. (1765) *Nouveaux essais sur l'entendement humain*, in German Academy of Sciences (ed.) *G.W. Leibniz: Sämtliche Schriften und Briefe*, Berlin: Akademie, 1923-; trans. P. Remnant and J. Bennett, *New Essays on Human Understanding*, Cambridge: Cambridge University Press, 1981.(A paragraph-by-paragraph commentary and critical discussion of Locke 1689.)

Leibniz, G.W.(1875-90) *Die Philosophische Schriften von Leibniz*, ed. C.I. Gerhardt, Berlin: Weidmann, 7 vols. (One of the standard editions of Leibniz's writings.)

Leibniz, G.W. (1903) *Opuscules et fragments inédits de Leibniz*, ed. L. Couturat, Paris: Alcan.(A collection of smaller, previously unpublished pieces by Leibniz.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1975.

(Locke’s main philosophical work.)

Locke, J. (1690) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1960. (Locke’s main work in political theory.)

Lucretius (*c*.55 BC) *On the Nature of Things*, ed. C. Bailey, *Titi Lucreti Cari De rerum natura libri sex*, Oxford: Oxford University Press, 3 vols, 1947.(Bailey is the monumental standard edition, with text, translation and commentary, still highly informative and not fully superseded.)

Rousseau, J.-J. (1755) *Discours sur l’origine et les fondements de l’inégalité parmi les hommes (Discourse on the Origin and Foundations of Inequality among Mankind)*, in B. Gagnebin and M. Raymond (eds) *Oeuvres Complètes de Jean-Jacques Rousseau*, vol. 3, Paris: Pléiade, 1964; trans. D.A. Cress, *Jean-Jacques Rousseau: Basic Political Writings*, Indianapolis, IN: Hackett Publishing Company, 1987.(Rousseau’s prize-winning essay on the origins of inequality.)

Rousseau, J.-J. (1762) *Du contrat social (The Social Contract)*, in B. Gagnebin and M. Raymond (eds) *Oeuvres Complètes de Jean-Jacques Rousseau*, vol. 3, Paris: Pléiade, 1964; trans. D.A. Cress, *Jean-Jacques Rousseau: Basic Political Writings*, Indianapolis, IN: Hackett Publishing Company, 1987.(Rousseau’s most famous political writing.)

Rousseau, J.-J. (1781) *Essai sur l’origine des langues*, ed. C. Porset, Bordeaux: Ducros, 1968; trans. in J.H. Moran and A. Gode (eds) *On the Origin of Languages*, Chicago, IL: University of Chicago Press, 1966. (Rousseau’s main contribution to the debate about the origin of language in the eighteenth century.)

References and further reading

Aarsleff, H. (1982) *From Locke to Saussure*, Minneapolis, MN: University of Minnesota Press.(Collection of essays on Wilkins, Locke, Leibniz, Condillac, Herder and lesser figures.)

Bertman, M. (1978) ‘Hobbes on Language and Reality’, *Revue Internationale de Philosophie* 32: 536-51.

Buroker, J.V. (1993) ‘The Port-Royal Semantics of Terms’, *Synthèse* 96: 455-75.

Chomsky, N. (1966) *Cartesian Linguistics*, New York: Lanham.(Controversial account about the connections between modern generative grammar and the Port Royal.)

Ducret, M. and Lannay, M. (1967) ‘Synchronie et diachronie: *L’essai sur l’origine des langues* et le second *Discours*’, *Revue Internationale de Philosophie* 21: 421-42.(On Rousseau’s views about the origin of language. The entire issue of the journal is devoted to philosophy of language in the eighteenth century.)

Flage, D. (1987) *Berkeley’s Doctrine of Notions: A Reconstruction Based on his Theory of Meaning*, London: Croom Helm.

Guyer, P. (1994) ‘Locke’s Philosophy of Language’, in V. Chappell (ed.) *The Cambridge Companion to Locke*, Cambridge: Cambridge University Press.(Survey article. The volume contains a detailed bibliography.)

Heinekamp, A. (1975) ‘Natürliche Sprache und Allgemeine Charakteristik bei Leibniz’, *Studia Leibnitian Supplementa* 15: 257-86. (Looks at natural language and the *universal characteristic* in Liebniz.)

Hungerland, I. and Vick, G.R. (1973) ‘Hobbes’s Theory of Signification’, *Journal of History of Philosophy* 11: 459-82.

Ishiguro, H. (1972) *Leibniz’s Philosophy of Logic and Language*, Cambridge: Cambridge University Press. (Includes discussion of the connections between Leibniz’s views on language and his metaphysics.)

Juliard, P. (1970) *Philosophies of Language in Eighteenth Century France*, The Hague: Mouton.(Discusses the philosophy of language of Condillac, Maupertuis, La Mettrie, Turgot and Rousseau.)

Kremer, E.J. (ed.) (1994) *The Great Arnauld and Some of His Philosophical Correspondents,* Toronto, Ont.: University of Toronto Press.(A collection of papers with detailed bibliography.)

Kretzmann, N. (1968) ‘The Main Thesis of Locke’s Semantic Theory’, *Philosophical Review* 77: 175-96.

Land, S. (1986) *The Philosophy of Language in Britain*, New York: AMS Press.(Discussion of the philosophy of language of Hobbes, Locke, Berkeley, Harris, Reid, Monboddo and Smith.)

Nuchelmans, G. (1983) *Judgment and Proposition from Descartes to Kant*, Amsterdam: North Holland.

Porset, C. (1977) ‘Grammatista philosophans. Les sciences du langage de Port-Royal aux Idéologues (1660-1818)’, in A. Joly and J. Stefanini (eds) *La Grammaire générale. Des Modistes aux Idéologues*, Lille: Presses Universitaires de France, 11-95.(Bibliography of works in the philosophy of language and grammar between 1660 and 1818. Also covers the secondary literature before 1976.)

Ricken, U. (1984) *Linguistics, Anthropology and Philosophy in the French Enlightenment*, London: Routledge. (General survey, includes a detailed bibliography.)

Rousseau, N. (1986) *Connaissance et langage chez Condillac*, Geneva: Libraire Droz.(Monograph on Condillac with a detailed bibliography.)

Rutherford, D. (1995) ‘Philosophy and Language in Leibniz’, in N. Jolley (ed.) *The Cambridge Companion to*

Leibniz, Cambridge: Cambridge University Press, 224-69.(Survey article. The volume contains a detailed bibliography.)

Sgard, J. (ed.) (1982) *Condillac et les problèmes du langage*, Geneva: Slatkine.(Collection of critical essays.)

Slaughter, M. (1982) *Universal Languages and Scientific Taxonomy in the Seventeenth Century*, Cambridge: Cambridge University Press.

Winkler, K.P. (1989) *Berkeley: An Introduction*, Oxford: Clarendon Press.(Chapters 1 and 2 deal with questions of language and abstraction in Berkeley.)

Language, Indian theories of

*Language is a much debated topic in Indian philosophy. There is a clear concern with it in the Vedic texts, where efforts are made to describe links between earthly and divine reality in terms of etymological links between words. The earliest surviving Sanskrit grammar, Pāṇini's intricate Aṣṭādhyāyī(Eight Chapters), dates from about 350 BC, although arguably the first explicitly philosophical reflections on language that have survived are found in Patañjali's 'Great Commentary' on Pāṇini's work, the Mahābhāṣya (*c.*150 BC). Both these thinkers predate the classical systems of Indian philosophy. This is not true of the great fifth-century grammarian Bhartṛhari, however, who in his Vākyapadīya (Treatise on Sentences and Words) draws on these systems in developing his theory of the* sphoṭa, *a linguistic entity distinct from a word's sounds that Bhartṛhari takes to convey its meaning.*

Among the issues debated by these philosophers (although not exclusively by them, and not exclusively with reference to Sanskrit) were what can be described as (i) the search for minimal meaningful units, and (ii) the ontological status of composite linguistic units. With some approximation, the first of these two issues attracted more attention during the early period of linguistic reflection, whereas the subsequent period emphasized the second one.

1 Historical sketch

Linguistic science in India started soon after the Vedic period. The earliest grammarian whose work has survived is Pāṇini (*c.*350 BC), author of the *Aṣṭādhyāyī(Eight Chapters)*. This work consists of some 4,000 aphoristic statements (*sūtras*) which describe the Sanskrit language in considerable detail, but leave no space for explicit reflections about the nature of language. Such reflections make their appearance in the voluminous *Mahābhāṣya (Great Commentary)* of Patañjali (*c.*150 BC). The *Mahābhāṣya* is a commentary on the *Aṣṭādhyāyī* (but not on all of its *sūtras*), and on the aphoristic *vārttikas* of Kātyāyana, which comment upon Pāṇini's *sūtras*. Another linguist whose work has been preserved and who, like Kātyāyana, appears to belong to the period between Pāṇini and Patañjali, is Yāska, author of the *Nirukta (Etymological Explanation)*.

All these authors precede the formation of the classical systems of Indian philosophy; their reflections on language are therefore largely unaffected by them. This changes with Bhartṛhari (*c.* fifth century), perhaps the first commentator on Patañjali's *Mahābhāṣya*, and the author of the *Vākyapadīya (Treatise on Sentences and Words)*. Bhartṛhari is well aware of the philosophies of his time, and makes ample use of them to construct his own system, which he presents as the philosophy of grammar. The subsequent Pāṇinian tradition accepts this philosophy (or what it preserves of it) as its own, but there are remarkably few grammarians who write treatises on it. Apart from the three principal commentators on the *Vākyapadīya* - Helārāja (tenth century), Puṇyarāja and Vṛṣabhadeva (dates unknown) - by far the most important among them are Kauṇḍa Bhaṭṭa and Nāgeśa Bhaṭṭa, both belonging to the most recent period of grammatical studies (after 1600). Some authors belonging to different schools of thought, however, adopt and defend some of the points of view of the grammarians. The ontological status of composite linguistic units is a subject that evokes special interest.

2 The search for minimal meaningful units

The different linguistic sciences of ancient India - and in particular grammar and etymological explanation - have to be understood against the background of the practice, common in the Vedic Brāhmaṇas (before Pāṇini), of giving etymological explanations of names of gods and of other terms, usually related to the sacrifice and often occurring in sacred formulas (*mantras*). Unlike those of modern linguistics, the etymologies of ancient India have nothing to do with the origin or the history of the words concerned. They cannot, because language, and the Sanskrit language in particular, was looked upon as stable in time; from the subsequent period we know that many even believed Sanskrit to be eternal, that is, without beginning. These etymologies establish links between things and the mythological reality that hides behind them. The god Agni ('fire'), for example, is thus called because he was created first (*agre*). There are countless etymologies of this kind in the Brāhmaṇas. These texts frequently add that the gods have obscured a number of the etymological links. The god Indra, for example, is 'really' called Indha ('the kindler') because he kindled the vital airs. However, people call him Indra because the gods are fond of the cryptic, and dislike the evident. Knowledge of the links revealed by etymologies is important in reaching one's religious goals.

Yāska's *Nirukta* takes the validity of such etymologies for granted, but secularizes their use. It presents etymologizing as a way to arrive at the meaning of unknown words. Moreover, only nouns and adjectives can be etymologically explained, and then only in terms of verbal forms: verbs explain nominal words, and show the (or an) activity that characterizes the object named. Yāska illustrates his method with the help of known words. One might expect that this procedure would lead to the identification and isolation of the common parts found in different words (such as the common part *ag* of *agni* and *agre* in the above example), and would determine their meanings. But Yāska's demands with regard to the semantic adequacy of etymological explanations are so stringent that this turns out to be impossible. He insists, for example, on two different etymological explanations for words that have two meanings. This rigour forces Yāska to be very undemanding with respect to the phonetic similarity required in etymological explanations. With the help of a number of examples from grammar, he shows that phonemes may disappear, be modified, change position, and so on. The same applies, *a fortiori*, to etymology. Similarities between words in etymological explanations may, as a consequence, be minimal: one single phoneme in common may have to do. The main thing is that one should not be discouraged; one should not stop looking for an etymological explanation simply because one does not find similar words.

It is interesting to observe that Patañjali's *Mahābhāṣya* contains a passage which, like Yāska's *Nirukta*, shows how phonemes may undergo change of position, elision and modification in grammatical derivations. Unlike Yāska, Patañjali concludes from this quite explicitly that phonemes by themselves cannot have meaning, although it seems likely that Yāska, even though he does not state it in his *Nirukta*, drew the same conclusion as Patañjali. It appears that Yāska's semantic rigour prevented him from trying to identify the ultimate meaningful constituents of Sanskrit (as we see for ancient Greek in Plato's *Cratylus*, for example). This task, but on a far less ambitious scale, is left to the grammarians, among whom Pāṇini is the most famous. His grammar is not an analysis of Sanskrit but a synthesis: it produces the words and sentences of the language, starting from their ultimate meaning-bearing constituents, essentially stems and affixes. To be precise, Pāṇini's grammar first furnishes stems and affixes on the basis of a semantic input, and these stems and affixes are subsequently joined together, and modified where necessary, so as to yield words and sentences (see Bronkhorst 1980).

In view of the background of Vedic etymologies, discussed above, one may legitimately conclude that Pāṇini considers these ultimate meaningful constituents to be really meaningful, more so perhaps than the 'surface forms' they help produce. The search for 'really meaningful' ultimate constituents of language is clearly present in the efforts of the grammarians. Pāṇini's *sūtra* 1.2.45, which recognizes but three meaningful entities, namely verbal roots, nominal stems and suffixes, indicates that words and sentences are considered to have at best a composite meaning. This search manifests itself later in the attribution by a number of Tantric thinkers of metaphysical meanings to individual phonemes. They can afford to go further than Pāṇini and Yāska in their analysis, continuing all the way to the individual phonemes, owing to the fact that they are less limited by semantic considerations. This in its turn is no doubt linked to the circumstance that sacred formulas (*mantras*) in Tantrism (unlike the Vedic ones) have shed their connection with ordinary language and its semantic constraints.

Returning to Pāṇini and grammatical analysis, later grammarians, mainly under the influence of Patañjali and Bhartṛhari, reject the position according to which the ultimate meaningful constituents presented by grammar are somehow more real than the words they produce. For them, stems and affixes are conventions, or rather inventions of grammarians. This reaction is to be understood in the light of the ontological concerns to be discussed below. The semantic analysis underlying Pāṇini's procedure, on the other hand, came to be generally accepted (albeit sometimes with slight modifications). Later thinkers use this analysis as the basis for deliberations on the relative importance of the various 'semantic elements' that Pāṇini assigns to a sentence in the understanding obtained by a hearer (*śābdabodha*). In a sentence like *caitraḥ pacati* ('Caitra cooks'), to take a simple example, the grammatical elements are: *caitra-s pac-a-ti*. Of these, the following are expressive: *caitra*, *pac* and *ti*. Thinkers of the new Nyāya school (Navya-Nyāya) consider the grammatical subject (in this case *caitra*) most important, and give (approximately) the following semantic analysis of the sentence: 'Caitra characterized by the activity of cooking'. The grammarians look upon the meaning of the verbal root (*pac*) as central, and paraphrase the sentence (again approximately) as: 'The activity of cooking whose agent is Caitra'. The Mīmāṃsakas, finally, put emphasis on the verbal suffix (here *ti*); since they are primarily interested in Vedic injunctions, and consequently in imperative and optative verbal forms, we shall not enter into the details of their analyses.

3 The ontological status of composite linguistic units

To appreciate the importance of the debate on the ontological status of composite linguistic units, one has to be aware of the great interest in ontological questions that characterizes much of Indian philosophy. In the realm of language this leads to questions like: Do words and sentences really exist? If so, how can they, given that the phonemes that constitute them do not occur simultaneously? Since, moreover, simultaneous occurrence is a condition for the existence of collective entities, do individual phonemes exist? They, too, have a certain duration, and consist therefore of parts that do not occur simultaneously.

Perhaps the first to address these questions were Buddhists of the Sarvāstivāda school. These Buddhists were active in the first centuries BC in drawing up lists of elements - the so-called *dharmas* - which were considered to constitute all there is (see Buddhism, Ābhidharmika schools of). The list accepted by the Sarvāstivādins contains three elements which correspond to phonemes (*vyañjanakāyas*), words (*nāmakāya*) and sentences (*padakāya*) respectively. This means that these Buddhists postulated phonemes, words and sentences as existing entities which, like virtually everything else in their ontology, are momentary. Little is known about the way they visualized the mutual relationship between these entities, or how they would answer the questions formulated above.

The grammarian Patañjali may have been influenced by these ideas. He certainly knew the notion of an individual phoneme and of a word conceived of as a single entity. For Patañjali, these phonemes and words are not momentary; they are, on the contrary, eternal. One should not, however, attach too much importance to this difference: for the Buddhists, everything is momentary; for many Brahmans, the Veda, and therefore also its language, is eternal. It is more important to observe that these notions play a relatively minor role in Patañjali's expositions. They acquire major significance in Bhartṛhari's *Vākyapadīya*, where they are made to fit his general philosophy that more comprehensive totalities are more real than their constituent parts.

It appears that in the period between Patañjali and Bhartṛhari a major shift of emphasis took place in the discussion of linguistic units. The discussion became centred on the linguistic unit as meaning-bearer. The problem of individual phonemes, which have no meaning, came to be separated from that of words, grammatical elements (stems and affixes) and sentences, which do. In the context of Bhartṛhari's philosophy this is understandable, for here linguistic units and the 'objects' they refer to are treated in a parallel fashion. But this shift of emphasis was not confined to the grammatical tradition. Śabarasvāmin, the author of the oldest and most important surviving commentary on the *Mīmāṃsāsūtra*, and who may be an approximate contemporary of Bhartṛhari, cites (1.1.5; see Frauwallner 1968: 38-) an earlier commentator who rejects the notion of a word as different from its constituent phonemes. This does not, however, prevent him from proclaiming that phonemes are single and eternal. In other words, phonemes and words undergo a different treatment altogether. Moreover, the author of the *Yogabhāṣya* (whose name was probably Patañjali, like the author of the *Mahābhāṣya*, although he was certainly different from the latter; recent tradition calls him Vyāsa) speaks about the single word which is without parts, without sequence, without constituent phonemes, and which is mental (on *Yogasūtra* 3.17). This Patañjali may have lived around 400 AD, and therefore perhaps before Bhartṛhari.

4 Early *sphoṭa* theory

Patañjali (the grammarian) and Bhartṛhari use the word *sphoṭa* to refer to linguistic entities conceived of as different from the sounds that reveal them. For Patañjali, the *sphoṭa* does not necessarily convey meaning; he uses the term also in connection with individual phonemes. For Bhartṛhari, the *sphoṭa* is a meaning-bearer. The *sphoṭa*, he points out, is different from the sounds which manifest it, and he makes several suggestions as to what constitutes it. It might be a mental entity. Or one might take it to be the universal residing in the manifesting sounds. One could even look upon the material basis of words, for example, wind, as being the *sphoṭa*. Bhartṛhari presents these options, but his perspectivism allows him to avoid choosing between them.

Arguments claiming to prove the existence of the *sphoṭa*, as well as arguments which try to refute it, henceforth concentrate heavily, even exclusively, on the *sphoṭa* as meaning-bearing unit. The primary question is not 'What exactly is the *sphoṭa*?' but rather 'How can a sequence of phonemes, each without meaning and not even occurring simultaneously, express meaning?' According to some, a sequence of sounds can express meaning; they have to show how it does so. Others hold that this is not possible; they solve the problem by postulating the existence of

the *sphoṭa*. These two positions find their classical expositions in Kumārila Bhaṭṭa's critique of the *sphoṭa* doctrine in his *Ślokavārttika (Commentary in Verse)* (seventh century), and in Maṇḍana Miśra's (*c.*700?) defence against these attacks in his *Sphoṭasiddhi (Demonstration of the Sphoṭa)*. Neither Kumārila nor Maṇḍana were grammarians: the former belonged to the school of Vedic hermeneutics called Mīmāṃsā; the latter, too, had links with this school.

Kumārila, elaborating the opinions of Śabarasvāmin (see §3), on whose *Mīmāṃsābhāṣya* he comments, accepts the eternal existence of individual phonemes. But he combats the notion that more than phonemes are required to understand the meaning-bearing function of language. It is true that the phonemes constituting a word are not pronounced simultaneously. But there are situations where everyone agrees that a series of activities that succeed each other in time can none the less jointly produce an effect. He gives the example of a Vedic sacrifice, whose constituent activities are performed at different times, but which produces a single result, namely heaven. Another example concerns counting: we can count objects in sequence, one after the other, and arrive at one result, their number. Furthermore, the fact that individual phonemes are without meaning does not exclude the possibility that they can express a meaning when pronounced in sequence. The parts of a cart, too, cannot fulfil the functions that a cart can fulfil. Last but not least, though the constituent phonemes of a word are not pronounced simultaneously, they are remembered together the moment the last phoneme is (or has just been) uttered.

Maṇḍana answers Kumārila's arguments one by one. He protests against the idea of the combined memory of the phonemes that constitute a word. First of all, one does not remember phonemes, but the word as a whole. Second, memory impressions can only present to us their contents, in this case phonemes, not something else, such as the meaning of the word. And third, two words may consist of the same phonemes, say 'pit' and 'tip' (a Sanskrit example is the pair *saraḥ*/*rasaḥ*, 'lake'/'taste'), so that the memories that combine their phonemes should be the same, yet they are recognized as different. Perhaps Maṇḍana's most interesting contribution to the discussion is his claim that the *sphoṭa* is directly perceived: it is gradually revealed by the phonemes.

The *sphoṭa* constitutes the central element of what came to be called the philosophy of the grammarians. All thinkers who deal with the issue, including Maṇḍana Miśra, refer in this connection to Bhartṛhari's *Vākyapadīya*. But the more encompassing ideas in the context of which Bhartṛhari worked out his ideas on the nature of linguistic entities largely escaped the attention of those who so faithfully cited him.

5 Later *sphoṭa* theory

After a lull, a revival of interest in the *sphoṭa* and related issues took place from the sixteenth century onward. Of the various authors who wrote treatises on the philosophy of grammar, Nāgeśa Bhaṭṭa (*c.*1700) was the most famous. He wrote a large number of treatises on various subjects, among them the *Sphoṭavāda (Exposition on Sphoṭa)*, the *Laghumañjūṣā(Small Casket)* and the *Paramalaghumañjūṣā(Extremely Small Casket)* (written in this order), which deal with the philosophy of grammar. These books show that Nāgeśa changed his mind several times with regard to the *sphoṭa* doctrine.

The *Sphoṭavāda* enumerates eight types of *sphoṭa*: (1) phoneme, (2) word, (3) sentence, (4) indivisible word, (5) indivisible sentence, (6) phoneme-universal, (7) word-universal, (8) sentence-universal. These *sphoṭas* are primarily meaning-bearers. The first and sixth ones, in spite of their misleading names, refer to grammatical elements (stems and affixes) rather than to phonemes. Nāgeśa's reasons for postulating these eight types are not always clear. This early work gives the impression that he collected various ideas without being able to combine them into one overarching vision.

This changes with the *Laghumañjūṣā*, which opens with the words: 'In this [work] the sentence-*sphoṭa* is most important.' Other parts of the work make it clear that Nāgeśa has been converted - no doubt under the influence of Bhartṛhari, whose *Vākyapadīya* he frequently cites - to the idea that only sentences really exist, that words and grammatical elements are no more than imaginary. He is particularly fierce with regard to grammatical stems and affixes.

Surprisingly, the *Paramalaghumañjūṣā,* meant to be an abbreviated version of the *Laghumañjūṣā*, begins, like the *Sphoṭavāda*, with an enumeration of the eight kinds of *sphoṭa*. Immediately following this it repeats the opening statement of the *Laghumañjūṣā*, according to which the sentence-*sphoṭa* is most important. Closer study reveals that Nāgeśa has been confronted with cases where the sentence-*sphoṭa* view comes into conflict with grammatical

derivations. There is a grammatical meta-rule which states that a grammatical derivation evolves in the order in which the expressive elements arise. Expressive elements acquire in this way importance, and it will not do to say that their expressiveness is merely imaginary. The issue is all the more important in view of the fact that there were different opinions as to whether the final substitutes of the grammatical elements - which appear in the 'surface forms' - are expressive, or whether the substituends are. This disagreement can have an effect on the correct derivation of words and sentences. Nāgeśa's final position is chosen in awareness of these complications. He still maintains that the sentence-*sphoṭa* is most important. But he no longer treats the other, 'imaginary', entities as lightly as he did earlier.

Nāgeśa is often thought of as the last great author in the Pāṇinian tradition. His vacillations where the *sphoṭa*-doctrine is concerned illustrate the conflict that exists between the two major issues of grammar distinguished in this entry: the search for minimal meaningful units on the one hand, and the ontological status of composite linguistic units on the other. His final position tries to give both their due: the idea inherited from Bhartṛhari that only the sentence is 'real', rather than words and smaller grammatical elements; and the idea inherited from Pāṇini that grammar is concerned with the smallest identifiable meaningful elements and the way they combine to form larger units.

See also: Interpretation, Indian theories of; Meaning, Indian theories of; MīmāṃsāČsect;3

JOHANNES BRONKHORST

References and further reading

Most of the ancient works listed here are highly technical and voluminous, and apt to be unrewarding for those who are not specialists in this field. Reliable translations are nonexistent in most cases.

Bhartṛhari (5th century) *Vākyapadīya (Treatise on Sentences and Words)*, ed. W. Rau, Wiesbaden: Franz Steiner, 1977.(A very difficult work that has occasioned a lot of controversy among scholars.)

Bronkhorst, J. (1980) 'The Role of Meanings in Pāṇini's Grammar', *Indian Linguistics* 40 (1979): 146-57.(Shows that meanings are the input of Pāṇini's grammar.)

Bronkhorst, J. (1981) '*Nirukta* and *Aṣṭādhyāyī*: Their Shared Presuppositions', *Indo-Iranian Journal* 23: 1-14.(Analyses the aims and procedures of Yāska's *Nirukta*.)

Bronkhorst, J. (1984) 'Patañjali and the Yoga Sūtras', *Studien zur Indologie und Iranistik* 10: 191-212.(§2 deals with the authorship of the *Yogabhāṣya*.)

Bronkhorst, J. (1986) *Tradition and Argument in Classical Indian Linguistics*, Dordrecht: Reidel.(On Nāgeśa Bhaṭṭa; see especially chapters 10 and 11, and appendix 3.)

Coward, H.G. and Raja, K.K. (eds) (1990) *Encyclopedia of Indian Philosophies*, vol. 5, *The Philosophy of the Grammarians*, Delhi: Motilal Banarsidass.(Contains bibliography up to 1983.)

Deshpande, M.M. (1992) *The Meaning of Nouns. Semantic Theory in Classical and Medieval India. Nāmārtha-nirṇaya of Kauṇḍabhaṭṭa*, Dordrecht: Kluwer.(Richly annotated translation of a chapter of Kauṇḍa Bhaṭṭa's *Vaiyākaraṇabhūṣaṇa*; the long introduction contains useful discussions of concepts of linguistic philosophy.)

Frauwallner, E. (1968) *Materialien zur ältesten Erkenntnislehre der Karmamīmāṃsā(Materials on the Oldest KarmamīmāṃsāEpistemology)*, Vienna: Hermann Böhlaus.(Contains text and translation of an important portion of Śabarasvāmin's *Mīmāṃsābhāṣya*.)

Gaurinath Sastri (1980) *A Study in the Dialectics of Sphoṭa*, Delhi: Motilal Banarsidass, revised edn.(Presents various arguments for and against the *sphoṭa* that were used in classical India.)

Houben, J. (1997) 'The Sanskrit Tradition', in W. van Bekkum *et al.*, *The Emergence of Semantics in Four Linguistic Traditions: Hebrew, Sanskrit, Greek, Arabic*, Amsterdam/Philadelphia, PA: John Benjamins, 49-145.(Survey of semantics in classical India.)

Kumārila Bhaṭṭa (7th century) *Ślokavārttika (Commentary in Verse)*, ed. Dvārikādāsa Śāstrī, Varanasi: Ratna Publications, 1978.(The most important commentary by the founder of the Bhaṭṭa school of Mīmāṃsā.)

Maṇḍana Miśra (*c*.700?) *Sphoṭasiddhi (Demonstration of the Sphoṭa)*, ed. and trans. M. Biardeau, *La Démonstration du Sphoṭa*, Pondicherry: Institut Français d'Indologie, 1958.(A response to Kumārila's *Ślokavārttika*.)

Matilal, B.K. (1988) '*Śābdabodha* and the Problem of Knowledge-Representation in Sanskrit', *Journal of Indian Philosophy* 16: 107-22.(On *śābdabodha*, mentioned in §2 of the present entry.)

Matilal, B.K. (1990) *The Word and the World: India's Contribution to the Study of Language*, Delhi: Oxford University Press.(Discusses various issues in Indian linguistic philosophy.)

Padoux, A. (1990) *Vāc: The Concept of the Word in Selected Hindu Tantras*, Albany, NY: State University of New York Press.(See especially chapter 5, on the meaning attributed to individual sounds.)

Pāṇini (4th century BC) *Aṣṭādhyāyī(Eight Chapters)*, ed. and trans. O. Böhtlink, Leipzig, 1887; repr. Hildesheim/New York: George Olms, 1997.(The earliest surviving Sanskrit grammar.)

Patañjali (2nd century BC) *(Vyākaraṇa-)Mahābhāṣya (Great Commentary)*, ed. F. Kielhorn, 1880-5, 3 vols; 3rd edn, ed. K.V. Abhyankar, Poona: Bhandarkar Oriental Research Institute, 1972.(A commentary on Pāṇini's grammar and Kātyāyana's *vārttikas*.)

Raja, K.K. (1969) *Indian Theories of Meaning*, Madras: The Adyar Library and Research Centre; 2nd edn, 1977. (Useful but generally nonhistorical survey of fundamental concepts.)

Scharfe, H. (1977) *A History of Indian Literature*, vol. 5, fasc. 2, *Grammatical Literature*, Wiesbaden: Harrassowitz.(Historical survey of technical linguistic literature; pays little attention to philosophical issues.)

Language, innateness of

Is there any innate knowledge? What is it to speak and understand a language? These are old questions, but it was the twentieth-century linguist, Noam Chomsky, who forged a connection between them, arguing that mastery of a language is, in part, a matter of knowing its grammar, and that much of our knowledge of grammar is inborn.

Rejecting the empiricism that had dominated Anglo-American philosophy, psychology and linguistics for the first half of this century, Chomsky argued that the task of learning a language is so difficult, and the linguistic evidence available to the learner so meagre, that language acquisition would be impossible unless some of the knowledge eventually attained were innate. He proposed that learners bring to their task knowledge of a 'Universal Grammar', describing structural features common to all natural languages, and that it is this knowledge that enables us to master our native tongues.

Chomsky's position is nativist *because it proposes that the inborn knowledge facilitating learning is* domain-specific. *On an empiricist view, our innate ability to learn from experience (for example, to form associations among ideas) applies equally in any task domain. On the nativist view, by contrast, we are equipped with special-purpose learning strategies, each suited to its own peculiar subject-matter.*

Chomsky's nativism spurred a flurry of interest as theorists leaped to explore its conceptual and empirical implications. As a consequence of his work, language acquisition is today a major focus of cognitive science research.

1 The development of Chomsky's nativism

In a review of Skinner's *Verbal Behavior*, Chomsky ((1959) rejected the behaviourist view that mastery of a language, or 'linguistic competence', consists in complexes of 'dispositions to verbal behaviour' instilled in our minds by 'operant conditioning' during childhood (see Skinner, B.F.; Behaviourism, methodological and scientific). First, children neither need nor receive the careful linguistic training that the behaviourist acquisition-theory requires. Second, our use of language is both *stimulus-independent* (not determined by the inputs we receive) and *productive* (we can utter and understand indefinitely many novel sentences). Hence competence must be more than a congeries of 'verbal habits'. Both language acquisition and linguistic competence involve complex systems of psychological states and processes.

But how to characterize these states and processes? In the early 1960s, Chomsky's 'generative' approach to linguistics challenged the 'taxonomic' approach of American structuralists such as Bloomfield and Harris (see Structuralism in linguistics §4). He urged that linguists abandon the structuralist aim of elucidating 'discovery procedures' (that is, mechanical methods which, when applied to a corpus of sentences in a language, will deliver a grammar descriptive of that language) and focus instead on the development of rigorous, formalizable syntactic theories for particular languages. The linguist's goal should be to develop *generative* grammars for natural languages, that is, sets of rules that will produce all and only the sentences in a language, together with their syntactic-structural descriptions (see Syntax §3).

Much of Chomsky's work in linguistics has aimed at refining the generative programme. Of particular interest here, however, is his contention that a grammar's usefulness outruns its ability to characterize a language. Its real importance, he argued, is psychological. Since speakers' grammatical intuitions (their judgments as to the well-formedness of sentences) constitute the data for a grammatical theory, that theory is in some sense descriptive of what speakers know about their language. Thus, Chomsky inferred, grammars supply (partial) accounts of the psychological structures underpinning language use, so providing a timely replacement for the ousted behaviourist account of linguistic competence. (Chomsky distinguishes theories of linguistic competence from theories of linguistic performance. Competence theories describe our capacity for language in a way that abstracts from the question of how that capacity results in the production and comprehension of sentences. Performance theories explain how the system of knowledge postulated in a competence theory interacts with other psychological systems such as memory, perception and the articulatory systems, in order to produce actual linguistic behaviour.)

But if grammars are competence theories, the question of how language is *learned* becomes acute. Normal children master language by about the age of eight. On Chomsky's view, this mastery entails, among other things, the

‘internalization’, ‘cognizance’ or ‘tacit knowledge’ of their language’s grammar. But a grammar is a highly complex body of syntactic, semantic, morphological and phonological rules, employing concepts and constructs far removed from experience. How do children internalize so complex a grammar at so tender an age, with apparently no conscious effort or instruction?

Chomsky argues that children’s acquisition of grammar is explicable only on the assumption that a substantial portion of the knowledge eventually attained is innate. On his view, a ‘Universal Grammar’ (UG), which specifies information about ‘linguistic universals’ (features common to all natural languages), is embodied in the language-learning mechanism itself. Thus, children need not *learn*, for instance, that declarative sentences must have a noun phrase as their subject: that fact they bring with them to the learning task. All they need do during learning is determine those facts (such as that, in English, the subject precedes the verb) that are *specific* to their language.

Chomsky is notable in that he does not rest content with the claim that we have innate knowledge of UG. Unlike his nativist predecessors, such as Plato, Descartes and Leibniz (see Nativism §2), he has offered two concrete proposals as to how that knowledge is realized.

Up to the late 1970s, he defended a ‘hypothesis-testing’ model of learning according to which children unconsciously project hypotheses about the grammar of their language, testing them against the data provided by experience. Their innate knowledge of UG, on this view, takes the form of constraints on the kinds of hypotheses they can entertain: they must be formulated in a particular language; they must have a certain form; and, in cases where the evidence is insufficient to arbitrate between two competing hypotheses, the language-learning mechanism itself will supply a decision.

More recently, Chomsky has embraced a picture of language learning as involving ‘parameter-setting’. On this view, which is as yet only partially developed, the principles of UG are encoded in the mind of the neonate, each principle containing one or more variables or ‘parameters’. A parameter’s possible values are extremely limited and are fixed by experience, different parameter-settings resulting in competencies in different languages. The ‘null subject parameter’, for example, governs whether one may omit the grammatical subject of a spoken sentence. Exposure to a language, like Spanish, where subjects may be omitted, sets the parameter one way; exposure to a language, like English, where subjects must be phonologically realised, will cause it to be set differently. Thus, children’s linguistic experience does not function as *evidence* for or against their grammatical hypotheses. Rather, and in much the same way as their hormones trigger the bodily changes that take place at puberty, exposure to the language of their community ‘triggers’ the mental changes that eventually result in linguistic competence.

Evaluating the merits of these models is in large part an empirical task and is outside the scope of this article. We will consider instead the case for linguistic nativism itself; that is, for the view that language-learning requires a task- or domain-specific mechanism of some kind.

2 The poverty of the stimulus

The commonest motivation for nativism is the argument from the poverty of the stimulus (POS). It contends that there is a ‘gap’ between the rich stock of linguistic knowledge that competent speakers acquire and the meagre supply of information about language that their experience during learning provides. This gap is so wide, the nativist contends, that no child endowed only with an empiricist-style, general-purpose learning mechanism could hope to bridge it. Language learning, therefore, requires a special-purpose device, one that embodies at the start some of the information about language that is ultimately attained.

This basic argument is elaborated in two ways. There are several a posteriori versions, due mainly to Chomsky; and a rather more a priori version with roots, seemingly, in Formal Learning Theory, a branch of mathematical linguistics.

A posteriori POS arguments are based on specific empirical claims about the inputs to and outputs of learning. In *Reflections on Language* 1975: 31 and following), for example, Chomsky claims that (1) competent speakers know the auxiliary fronting rule for question-formation; and that (2) the information available to children does not reliably distinguish the correct rule from incorrect but equally plausible alternatives. Hence, he infers, the

learner's choice of the correct rule is driven not by any general principles of theory-construction (such as the empiricist envisages), but is due rather to the fact that the acquisition device itself is biased towards grammatical rules of certain kinds, namely those conforming to UG.

Note that a posteriori POS arguments rely on specific empirical assumptions about the nature of linguistic competence and the data accessible to children. If it were discovered, for example, that speakers had *not* in any interesting sense internalized the rule in question, or that data disconfirmatory of incorrect rules were readily available, Chomsky would have failed to locate a gap - and hence to have supported his nativism. Hence, the success of a posteriori arguments turns on the degree to which their factual premises can be empirically sustained. But while some steps have recently been taken towards furnishing the supporting (for example, developmental) data required by arguments like this, their overall bearing on the nativism-empiricism controversy remains undecided.

In contrast to the preceding, the a priori POS argument (often called the Logical Problem of Language Acquisition) seeks to infer the existence of a gap from a consideration of the logical structure of the acquisition-task itself. Take some state in the learning process at which a child has acquired a grammatical hypothesis. Call the set of sentences (the language) generated by their hypothesised grammar *H*. Call that generated by the correct grammar (the target language) *L*. Then the learner is in one of five situations: (1) *H* is disjoint from *L*; (2) *H* overlaps *L*; (3) *H* is a subset of *L*; (4) *H* is a superset of *L*; (5) *H* is equivalent to *L*.

Situation (5) represents the end point of the learning process: the child has learned the target language. Situations (1), (2) and (3) are relatively unproblematic: the learner will be forced to revise their hypothesis if they hear a sentence in *L* that is not in *H*. Thus, they can converge on the correct grammar by exposure to 'positive evidence', evidence (as is provided by someone's uttering a sentence) that a given string of words is a sentence of the target language. But, crucially, if they are in situation (4) - if their grammar 'overgenerates', that is, generates all of *L* and also some sentences that are *not* in *L* - they cannot discover *this* by exposure to further sentences of the target language. For, every sentence of *L* is equally a sentence of *H*. What this learner needs is the information that a sentence that is generated by their grammar - a sentence in *H* - is *not* a sentence of *L*; only then can they 'shrink' their hypothesis so that it converges on the correct one.

This latter kind of evidence, that a sentence is *not* in the target language, is called 'negative evidence'. And case (4) is of interest because negative evidence is in general unavailable to children. *Explicit* negative evidence is scarce: children are not provided with lists of ungrammatical sentences; their errors are rarely corrected; nor do parents typically notice their own ungrammatical utterances. And *indirect* negative evidence is hard to find too. The non-occurrence of a sentence in the data, for example, is not negative evidence. For, there are infinitely many sentences of *L* that *de facto* will never be uttered. Hence from the mere fact that they have not heard a certain string of words in the data to date, the child cannot infer that it is not a sentence of the target language.

Thus learners who find themselves in situation (4) have no systematic access to evidence that would enable them to correct their overgeneral hypotheses. Which means, since children *do* eventually converge on the correct grammar, that they must never be in situation (4); the learning mechanism must be constrained such that they *never* hypothesize a grammar that is 'too large'.

Nativists contend that the general-purpose constraints on hypothesis-formulation envisaged by empiricists are insufficient to ensure that overgeneration will not occur. Hence, they conclude, the space of possible grammars must be constrained by further, linguistically-specific principles such as are provided by UG: the poverty of negative evidence in the data requires a nativist approach to language acquisition.

The a priori POS argument is at first sight much more compelling than its empirically undersupported a posteriori relatives. It is not, however, without points of weakness. First, one may question whether negative evidence is really as scarce as it seems. For example, while the non-occurrence of a particular sentence in the data may not be evidence that it is not grammatical, perhaps the repeated non-occurrence of certain syntactic forms could be evidence that sentences of that form are not grammatical. (That there must be at least *some* negative evidence available is evident from the fact that children do, as a matter of fact, recover from overgeneral grammatical hypotheses: they learn, for example, that you do not *always* add /d/ to the stem to form the past tense of a verb).

Second, one may question the nativist's contention that general-purpose constraints are insufficient to prevent a

learner's making incorrigible overgeneralizations. Not only is this contention to my knowledge unargued, there is reason to believe that in at least *some* cases, general-purpose learning strategies *are* efficacious in the face of what seem to be identical 'negative evidence' problems. For, negative evidence - evidence as to what things are *not* - is *quite generally* unavailable. When we learn what cars are, for example, we are not systematically informed about the countlessly many things that are not members of that class. Yet, and despite this paucity of negative evidence, we do manage to converge on the correct hypothesis about 'car-bonnet'. So, assuming that it is implausible to postulate a task-specific 'automotive faculty' that facilitates our learning in this domain, it must be the case that whatever general-purpose learning strategies we possess are in at least some contexts able to function successfully in the (near) absence of negative evidence.

In sum, arguments from the POS alone cannot compel nativism about language. We need additional reasons to think that there is something *special* about language acquisition - features it does not share with learning about, say, cars - such that nativism is plausible in the former case, while being clearly implausible in the latter.

3 Other arguments for nativism

Language acquisition has been held, by Chomsky and others, to be special in a number of ways that support linguistic nativism. Examples of these claims, and of the empiricist responses to them, are:

(1) *Species specificity.* Only human beings, so far as we know, speak a language. We thus possess a special biological affinity for language, an affinity which may amount to an innate language-learning faculty. Response: only human beings build cars, too, but that does not make for an innate automotive faculty. What our species possesses that others do not is general intelligence - an ability to learn from experience in a wide variety of contexts.
(2) *Critical period.* Language acquisition displays 'critical period effects': the ability to learn language declines drastically after puberty. This resemblance to other biologically-controlled developmental processes suggests that the acquisition of language too is determined more by our genes than our environment. Response: perhaps *all* learning declines after puberty. There are no data supporting the view that language-learning is special in this regard.
(3) *Facility of acquisition.* Language is acquired with remarkable speed and ease at a very early age. Processes distinct from those underpinning other kinds of learning are therefore at work. Response: given that we lack anything to compare language learning with, it is unclear that it is either notably fast, or especially easy.
(4) *Linguistic universals.* There are deep syntactic and semantic similarities among all known natural languages. Since there is no evidence to support the view that now-current languages descended from a common ancestor, these commonalities must be consequences of the way the language faculty is structured. Response: of course languages are shaped by the minds that use them. But linguistic universals may be due to constraints on what can be learned via a general-purpose learning strategy, rather than arising out of the structure of an innate language faculty.
(5) *Linguistic complexity.* Linguists, using *exactly* the 'general principles of theory-construction' beloved of the empiricist, spend entire careers laboriously constructing grammars for tiny corners of a language. But children can learn *all* of the grammar of *any* language simply by hearing it spoken around them. Surely some special faculty is at work here? Response: this argument assumes that there *is* an interesting sense in which speakers know - hence must learn - the grammar of their language. This assumption, however, may be challenged. First, there are conceptual difficulties involved in understanding what it is for a speaker to know a grammar. Thus, it is unclear whether children's ability to learn one is remarkable or not. Second, it has proved difficult to find empirical confirmation for the claim that grammars are implicated in any straightforward way in on-line linguistic processing. Thus it is unclear in what sense, if any, grammars need to be learned. And finally, one might question whether the linguist's methodology is appropriate for investigating the psychology of language use and understanding. Thus, while it is remarkable that the linguist's task in characterizing the language the child knows is dauntingly complex, this is of unclear relevance to the child's task in learning it.

In conclusion, Chomsky's claim that knowledge of language is largely innate has renewed the interest of linguists, philosophers and psychologists in the problem of explaining language acquisition. It remains, however, an open empirical question whether that explanation requires the postulation of a special-purpose learning mechanism embodying innate knowledge of language; and, if it does, what the nature of that mechanism and that knowledge

is.

See also: Chomsky, N.; Innate knowledge; Language, philosophy of; Semantics §1

FIONA COWIE

References and further reading

Chomsky, N. (1959) 'Review of Skinner's *Verbal Behavior*', *Language* 35: 26-58.(Chomsky's critique of behaviourism; discussed in §1.)

Chomsky, N. (1975) *Reflections on Language*, Glasgow: Fontana/Collins.(Seminal presentation of *de facto* POS argument (in §2) and the hypothesis-testing model of learning discussed in §1.)

Chomsky, N. (1988) *Language and Problems of Knowledge: The Managua Lectures*, Cambridge, MA: MIT Press.(Chomsky's current views about the innateness of language. Discussion of parameter-setting (in §1). References to earlier works.)

Hornstein, N. and Lightfoot, D. (eds) (1981) *Explanation in Linguistics: The Logical Problem of Language Acquisition*, London: Longman.(Sympathetic exploration of a priori POS argument; discussed in §2.)

Katz, J.J. (ed.) (1985) *The Philosophy of Linguistics*, Oxford: Oxford University Press.(Chomsky's nativism is a major focus of this collection.)

Matthews, R.J. and Demopoulos, W. (eds) (1989) *Learnability and Linguistic Theory*, Amsterdam: Kluwer.(Implications of the a priori POS argument (discussed in §2) for the theories of language learning.)

Newmeyer, O. (ed.) (1988) *Linguistics: the Cambridge Survey*, vol. 3, *Language: Psychological and Biological Aspects*, Cambridge: Cambridge University Press.(Empirical approaches to language acquisition. More up-to-date than Wanner and Gleitman (1982), but less focused on the nativism issue.)

Piattelli-Palmarini, M. (1989) 'Evolution, Selection and Cognition: From "Learning" to Parameter Setting in Biology and in the Study of Language', *Cognition* 31: 1-44.(Introduction to parameter-setting discussed in §1, with references.)

Pinker, S. (1979) 'Formal Models of Language Learning', *Cognition* 7: 217-82.(Survey of formal learning theory literature discussed in §2.)

Pinker, S. (1994) *The Language Instinct*, New York: William Morrow.(An excellent and highly readable summary of the case for linguistic nativism.)

Sells, P. (1985) *Lectures on Contemporary Syntactic Theories: An Introduction to Government-Binding Theory, Generalized Phrase Structure Grammar and Lexical-Functional Grammar*, Stanford, CA: Center for the Study of Language and Information.(Introduction to formal syntax, including a chapter on generative grammar relevant to §1.)

Wanner, E. and Gleitman, L.R. (eds) (1982) *Language Acquisition: The State of the Art*, Cambridge: Cambridge University Press.(Collection of psychologists' writings on language acquisition, with special emphasis on nativism.)

Language, medieval theories of

*A great deal of theorizing about language took place in western Europe between 1100 and 1400. The usual social context of this theorizing was the teaching of grammar, logic or theology. Rhetoric was traditionally counted as one of the language disciplines (*scientiae sermocinales*) together with grammar and logic, but in practice it received little attention. Medieval thinkers produced a vast literature on aspects of linguistic theory, but they did not write books with such titles as 'A Theory of Language'. The theories that have come down to us today have been reconstructed from a large number of sources, even when they are attributed to a single person.*

Although the medieval writers on language were very innovative, they owed some key ideas to ancient Greek and Latin authors, for example: (1) words acquire their meaning by an act of 'imposition' when a sound is chosen as the label of some thing; (2) there are three key ingredients in signification: the word, a concept and the thing signified; (3) concepts can be thought of as mental words; (4) the grammaticality of a sentence cannot be explained purely in terms of morphology; and (5) words have different contents when used as predicates of creatures and when predicated of the Creator.

The medieval thinkers disagreed as to whether words signify things directly or only through concepts. The latter view ran the danger of making concepts a screen between language and reality, but it had the advantage of being able to explain why different words can signify the same object without being quite synonymous. Many thought that general terms signify universal things, but there were also nominalist schools which held that the general terms themselves are the only universal things. Fourteenth-century nominalists located universality in concepts, also called mental terms, and only secondarily in spoken words. The language of thought had gained priority over that of speech.

The theory of grammar known today as modism was developed in the thirteenth century. This theory assumed that there are only two contributors to the sense of an expression, lexical meaning and grammatical features, and that only the latter belong to the province of grammar, which thus became a purely formal science that could claim applicability to all languages irrespective of their surface differences. The problem with modism was that it had no tools for dealing with even slightly deviant, yet intelligible expressions. Thus it would have to reject a statement such as 'the crowd are *rushing' because of the lack of concord of number.*

A number of medieval scholars focused on the various forms of metaphorical language and ambiguity, which were more appropriate for dealing with deviant expressions of the type mentioned above. It was realized that speaker, listener and context must be taken into account in order to explain how words can communicate something different from their primary sense, and how it is possible for a listener to grasp the intended sense of an ambiguous message. One motivation for this study was a need to understand how theological discourse functions. Theology also lay behind a heated debate about the ontological status of the meanings of propositions, and sacramental theology joined grammar in developing a notion of performative locutions.

The study of syntax yielded many new concepts, including those of government and dependence. Much less work was done on the evolution of languages, but Roger Bacon and Dante did offer some perceptive observations. Though never creating fully fledged artificial languages, logicians did develop a semi-artificial Latin.

1 Sources

The most important source for grammar in the Middle Ages was Priscian's *Institutiones grammaticae (Institutions of Grammar)*, a comprehensive Latin morphology and syntax composed in the early sixth century AD. Priscian not only describes and prescribes but also reasons about linguistic problems, using a theoretical framework borrowed from the second-century Greek grammarian Apollonius Dyscolus. Priscian transmitted to posterity the notion that the grammaticality (*congruitas*) of a sentence depends on a proper fit among the conceptual elements signified by its constituent words. Also important were his classifications of words, including the eight-fold partition of constituents of sentences as *partes orationis* (parts of speech). These eight parts were known from other sources, but Priscian provided the best discussion of the subject. Matters left untouched by Priscian, such as tropes and figures, were treated in Chapter 3 of Donatus' *Ars Maior*, composed in the fourth century. Under the name of *Barbarismus*, this chapter became a focus around which investigations of deviant expressions developed.

In logic, Aristotle was the main source of inspiration. His *Sophistical Refutations* made it clear that an argument does not always carry its validity on its surface, and his explanations of why a bad argument may look good gave impetus to the development of studies in semantics, and to a lesser degree in the pragmatics of linguistic communication. In *Peri hermeneias (On Interpretation)* 1, Aristotle briefly sketches which relations hold between corresponding items in writing, speech, thought and external reality. According to the standard interpretation of this text, Aristotle says that written words signify spoken ones, which in turn signify concepts, while concepts signify things. This reading can support the notion that truth, validity and their opposites are not primarily properties of spoken or written sentences and arguments, but of their mental counterparts. If the Aristotelian level of concepts is equated with that of Priscian, grammatical and logical correctness (congruity and truth) have the same primary bearers, that is, mental propositions whose 'words' are concepts. Aristotle's definitions of noun and verb in *Peri hermeneias* 2-3 differ from those of the Latin grammatical tradition, and this tension between grammar and logic created fruitful discussion. In *Categories* 1, Aristotle introduces the notion of paronymy or denomination: for example, the just (person) is denominated from justice. From this seed, theories of primary and secondary signification developed (to continue with the example, 'just' primarily signifies justice and secondarily signifies a person).

The logical works of Boethius provided the notion of imposition (*impositio*). Following Porphyry, Boethius taught that words acquire their meaning by an act by which a human 'impositor' decides to use a certain sound to name some thing. He distinguished a first imposition, by which features of the physical world acquire names ('tree', 'green', 'three cubits long' and so on), from a second one by which names acquire names ('noun', 'verb' and so on). He further suggests that a word such as 'species' is a name of names, so that 'man is a species' is a statement not about man but about the word 'man'. This suggestion helped to fuel early nominalism.

In theology, some essays by Boethius were influential until about 1200, but the principal authority was always Augustine. These two authorities inculcated in medieval writers the conviction that ordinary words do not have their ordinary meanings when used of God. For God to be high and just is not the same as his having big quantity and the quality of justice. God in his absolute simplicity cannot be analysed in terms of the Aristotelian categories. This doctrine was particularly important in the twelfth century when, according to a widely accepted analysis, ordinary predication was thought to consist in the ascription of a certain form to some bearer or bearers. Any such form was supposed to belong to some Aristotelian category. Thus, 'Socrates is a man' ascribes the substantial form 'humanity' to Socrates, 'Socrates is just' ascribes the quality of justice to him, and so on. Since the truth of 'God is just' was uncontroversial, the Augustinian-Boethian doctrine forced medieval writers to deny one or both of the claims that (a) a predicate term always signifies some categorial form, and (b) in propositions of the type '*X* is *Y*', the form signified by *Y* is always attributed to the referent(s) of *X*. Usually, some combination of the two solutions was advanced. In the first place it was claimed that 'divine terms' ('God' in particular) never signify categorial forms; in the second place it was claimed that true propositions of the type 'God is *Y*' never ascribe categorial forms to God, but either indicate a sort of total identity between God and *Y* or actually ascribe a categorial form to his creation, so that 'God is just' may be paraphrased as 'there is justice in the created world, and it derives from God'.

Augustine (*De trinitate* XV) also contributed to the theory of mental speech his notion of a soundless 'speech in the mind' (*verbum mentis*) prior to any language in the phonic medium. He further gave a definition of a sign as that which, besides presenting itself to the senses, also makes the mind think of something other than the sign itself. In somewhat different wordings, this definition also occurs in Augustine's *De doctrina christiana* 2.1.1 and *De dialectica* 5. The last-named work had few readers; one, Roger Bacon, developed from Augustine's work an unusual awareness that signification is a three-term relation involving not only sign and *significatum* but also someone to whom the sign signifies the *significatum*.

The whole patristic tradition of Biblical exegesis provided another rich source for medieval speculation about linguistic matters, not least about metaphorical language.

2 Words, concepts and things: introduction

Words were said in the Middle Ages to signify (*significare* or *designare*). A special terminology was developed for reference, but it remained a matter for debate what sort of entities significates actually were.

Few writers of the period were very concerned about written words, which they considered as mere ‘stand-ins’ for spoken words, although Pierre d’Ailly pointed out that a written message can be understood directly without recourse to sound. Concerning spoken words, late ancient sources suggested that they signify things through concepts. Typological interpretation of Biblical passages introduced the further notion of words signifying things which in turn signify other things, but this model was relevant only in Biblical studies. Some, such as John Buridan, took the word-concept-thing model to mean that words signify concepts which in turn signify things, so that words signify things only thanks to the transitivity of signification. Others, notably several late-thirteenth-century authors as well as William of Ockham in the fourteenth century, took it in the sense that for words to have the relation of signification to things, they must also be related in a certain way (subordinated) to concepts.

Those who took words to signify things directly usually did so because they felt that otherwise the concepts would become an impenetrable screen between language and external reality. If concepts were what our words signify, then it seemed we would always be talking about mere psychological entities. Buridan preferred the alternative model, *inter alia* because it can explain how two words can have the same extramental referents without being strictly synonymous. It suffices that the two words have different concepts for their direct significates.

The link between word and significate was generally held to be conventional (*ad placitum*), as Aristotle had said, but this was understood more in the sense that the link was the result of a free choice than in the sense that it depended on agreement between the users of the language. This link was supposed to have been established by ‘the impositor’, a figure who can trace his ancestry through Boethius and Porphyry back to the ‘giver of names’ in Plato’s *Cratylus*. Some authors seem to have understood the ‘impositor’ very literally as some wise man who at some time had decided what things were to be called. Others took ‘the impositor’ to be a collective designation of the people who had introduced new words in course of time, so that in practice ‘word *X* was imposed to mean *Y*’ equalled ‘the traditional meaning of *X* is *Y*’. Still others distinguished between the original wise impositor(s) who devised the phonology and other grammatical features of the language and laid down a basic vocabulary, and later users of the language who added to the vocabulary through new acts of imposition while not introducing major systematic changes.

Boethius operated with a first imposition to produce words for natural phenomena, and a second one to introduce words predicable of words (such as ‘noun’ or ‘trisyllabic’). Early nominalists included terms such as ‘species’ among those of the second imposition, but after the thirteenth century the vocabulary of logic was more often described as words of second intention, ‘second imposition’ being reserved for the vocabulary of grammar. Neither the distinction according to imposition nor that according to intention was designed to cover the whole of a language’s vocabulary, but only such items as are predicable (*categore(u)mata*). The understanding of what sort of entities first and second ‘intentions’ are changed over time. To Ockham, a word of the second intention is one that signifies mental *intentiones* (concepts), which are natural signs of other things (the individuals falling under the concept). By contrast, words of the second imposition signify things which are conventional signs, namely words.

While most medieval writers ascribed freedom of choice to ‘the impositor’, they also assumed that sounds had not been selected at random. Thus there was a justification for etymology, the art of revealing which properties of some thing the impositor paid special attention to when selecting a sound to name it. For example, many found it probable that the Latin word for ‘stone’, the accusative form of which is *lapidem,* had been chosen because it is characteristic of stones to hurt one’s foot, and in Latin ‘hurting the foot’ is *laedens pedem,* which is somewhat similar to *lapidem*. Moreover, this aggressive feature of stones could be alleged as a reason for the masculine gender of *lapidem*. Theologians commenting on Genesis 2: 19-20, where Adam appears as the first giver of names, stressed that he selected appropriate names on the basis of his understanding of the nature of the things to be named. This conventionality of language was particularly emphasized by a number of fourteenth-century thinkers, who felt that the previous age had tended to produce overly rich ontologies by assuming some real distinction corresponding to every linguistic distinction they saw.

The act of imposition could be understood as one of uniting the material aspect of a word (sound) with a formal aspect (informational content). Thirteenth-century scholastics worried about how this could be possible. One proposal was that just as the formal features of a material thing can be transferred to the mind as an immaterial likeness (species), so the intelligible species can be communicated to other people by means of a material vehicle,

the mind impressing a sort of copy of the relevant species on any particular token of a type of the sound conventionally associated with some sort of thing. Roger Bacon took this mechanism to be the explanation of magic and, apparently, of ordinary persuasion. He thought that if the speaker's mind and body cooperate to emit sound carrying some species, the species may act as a force (*virtus*) capable in favourable conditions of producing changes in both material and spiritual things. However, many thinkers of the next century were to deny material words the ability to carry information; instead, they located the information in the concepts to which words correspond.

3 Words, concepts and things: nominalism

Shortly before 1100, some philosophers began to define universals as words (*voces*) or names (*nomina*), and to consider predication a linguistic phenomenon: words are predicated of words, not things of things. Logic thus became a genuinely linguistic discipline (see Logic, medieval). Porphyry's five predicables (genus, species and so on), Aristotle's ten categories (substance, quantity and so on), and Boethius' topical differences (genus, opposite and so on) could all be viewed as names applicable to certain classes of words, that is, as words of the second imposition.

Early representatives of this 'vocalism' apparently did not make it sufficiently clear that the voces that are the bearers of universality are not mere articulated sounds without regard to their meaning, and thus they laid themselves open to the objection that, in their theories, the Greek *anthropos* and the Latin *homo* would be two different universals even though they both mean 'man'. Likewise, *homo* uttered at one time would not be identical with the equiform sound uttered at another time. The supposed universal would thus dissolve into a host of individual sounds.

This problem in early 'vocalism' was solved when Peter Abelard, the father of a school of *nominales*, made it clear that the criterion of identity for a universal was its signification: several voces with a shared signification make up one *sermo*, he held, and *sermones* are the bearers of universality. The signification of a universal term consists first in naming (picking out) individual things, and second in a relation to mental acts (*intellectus*). The individuals that a universal (such as 'man') picks out share a circumstance (*status*), for example *esse hominem* (being a man), but a *status*, Abelard held, is no thing. Some of his followers seem to have taken a *status* to be what a universal signifies; thus terms acquired significates of a type similar to those of whole propositions, for 'Socrates is a man' was supposed to signify the *dictum* that-Socrates-is-a-man (*Socratem esse hominem*) (see Nominalism).

About 1200, nominalism almost disappeared and only returned after 1310 with William of Ockham and John Buridan as its leading figures. Both the latter operated with three types of language, written, spoken and mental, and of these the mental language was considered logically most important. According to Ockham, terms of the written language are 'subordinated' to terms of the spoken language, and they in turn are subordinated to terms of the mental language. A mental term is a natural sign of some set of individuals. Its subordinate spoken and written terms are conventional signs of exactly the same individuals, so signification always links a term to individual things. Buridan sometimes uses Ockham's terminology, but more often he assumes that written terms signify spoken terms which in turn signify mental terms, which finally signify individual things, the 'ultimate significates'.

To both Ockham and Buridan, mental terms are concepts, and they are the primary bearers of universality. Oral and written terms are so derivatively. Similarly, 'true' and 'false' are primarily predicates of mental propositions. In Ockham's mature theory, a concept is identical with an act of intellection, and its ontological status is therefore that of an accidental characteristic of a soul. The ability of a concept C to signify every member of some set S derives from a combination of (1) a causal history reaching back from C to some member x of S encountered in direct cognition (intuition), (2) an isomorphism between C and x, and (3) a maximal similarity between x and any other member of S.

Ockham's mental language was almost a duplication of spoken language, but sometimes, at least, he thought of it as a language purified of the ambiguities and redundancies of its spoken counterpart and held that spoken language has two strictly synonymous terms, mental language has only one. Similarly, he claimed that grammatical features of spoken language lack mental counterparts if they are irrelevant to the truth of a proposition. Thus mental language does not have grammatical gender, but it does have tense.

Ockham and Buridan assumed multiple denotation for general terms whether they be in the subject or the predicate position, and co-reference of subject and predicate term was their fundamental criterion of truth for a categorical proposition. According to them, an affirmative categorical proposition is true if and only if the subject and the predicate term each refer to (*supponit pro*) at least one individual, and any individual to which the subject term refers is also referred to by the predicate term. The occurrence of an empty term like 'chimera' renders any affirmative proposition false and any negative proposition true (see Propositions, sentences and statements).

Thirteenth-century thinkers typically took signification to be a relation linking words with universal concepts, common natures or the like. The technical term for the relation between word and individual things was 'supposition' (see §12). For fourteenth-century nominalists, however, both signification and supposition terminate in individuals, but a word supposits only when occurring as a term in a proposition whereas it signifies even when uttered in isolation. Generally, a word's *supposita* will be a subset of its significates. The exception is the case when a word supposits for itself (or equiform words), whether in the mental, the spoken or the written variant, such as 'man' in 'man is an absolute concept', 'man is a monosyllabic word' or 'there are three letters in man'. In such cases, the term was said to be used non-significatively.

Whereas the verb's tense and modality define the relevant set of *supposita*, quantifiers affect the way a term supposits and thus the way to spell out the truth conditions of the proposition by means of the technique known as exposition. Exposition replaces the original proposition by an equivalent conjunction or disjunction of propositions, or by one proposition with a conjunct or disjunct subject or predicate. Thus 'every man runs' → 'this man runs, and that man runs, and so on', 'some man runs' → 'this man runs, or that man runs, and so on'.

In the early fourteenth century, the dominant grammatical theory was based on the notion that words possess 'modes of signifying' from which they derive their grammatical properties. It took some time for nominalist philosophers to work out an alternative, but by 1400 a new grammar had developed which took mental language to be the primary bearer of grammatical as well as of logical predicates. Mental terms were said to signify 'nominally' or 'verbally' the grammatical categories built into the concepts. Well-formedness was claimed to belong to spoken strings of words only thanks to their 'subordination' to grammatical strings of mental terms, not thanks to any grammatical properties inherent in the spoken words themselves, for how could sounds bear such properties, which obviously have an origin in the mind? Proponents of such views on grammar included Albert of Saxony, Pierre d'Ailly and one Marsilius (possibly Marsilius of Inghen). Some fifteenth-century thinkers also accepted modes of signifying, but held that they resided in concepts, not in words.

One strange result of late medieval nominalism was that thought was often considered to be more genuinely language than speech.

4 Words, concepts and things: realism

Taking universality to be a predicate only of concepts and/or words was not the prevailing view in the Middle Ages; more often, universality was held to reside somehow in things, though requiring a mental operation to be brought out into the open. This is not the place for a history of realist theories (see Realism and antirealism), but two points need to be mentioned. First, the naïve Platonism in which universal Forms, residing in a world of their own, are the principles of being and of understanding for perishable things, while being also the significates of general terms, never had currency during this period. Second, probably the most influential variant of realism was one that arose in the thirteenth century and owed much of its inspiration to the Persian philosopher Avicenna (see Ibn Sina). While accepting the reality of universals, this theory did not make them the significates of general terms. Instead, it held that such words signify 'common natures' transcending the distinction between universal and particular, existence and non-existence. As the common nature was generally identified with quiddity and essence, it was considered a principle of the understanding and being of things, but not one that would bestow either universality, particularity or existence on them. Particularity and universality were considered different manifestations, as it were, of the same underlying common nature or *res* (thing). Such 'things', the theory held, may have being in many alternative 'ways' or *modi essendi* (modes of being), in a static and in a dynamic way, in a universal or a particular way and so on, and this is reflected in human understanding and language. Thus the distinction between nouns and verbs brings out the distinction between the static and the dynamic modes of being of things.

5 Words, concepts and things: modism

About 1260-80 a grammatical theory of high generality was developed, known as *grammatica speculativa* (theoretical grammar); today the standard term for this theory is ‘modism’. The first generation of ‘theoretical grammarians’ included Boethius of Dacia, Martin of Dacia and John of Dacia. Radulphus Brito and Thomas of Erfurt, both active around 1300, belong to the second generation. Few theoretical advances seem to have been achieved after about 1310, though the works of Martin and Thomas continued to be used for about two centuries.

The term *modus significandi* (way/mode of signifying) had been in use since the twelfth century, typically to explain why an abstract and a concrete word for the same thing are not interchangeable *salva veritate*. Stephen Langton, for example, claimed that *deus* (God) and *deitas* (godhood) have the same signification and the same referent, and yet the truth of *deus est in lapide* (God is in a stone) does not entail the truth of *deitas est in lapide* (godhood is in a stone) because *deus* and *deitas* do not signify in the same way. The first proposition is true because of God’s ubiquity, the second is false because it amounts to a claim that the stone is divine. By the mid-thirteenth century, it had become common to call grammatical properties of words ‘modes of signifying’, and a distinction had been introduced between a word’s ‘general’ (lexical) signification and the ‘special signification’ due to its grammatical form.

In a further development, the notion, inspired by Avicenna, of common natures and their modes of being provided a suitable metaphysics for a linguistic theory with ‘modes of signifying’ as the key concept. It was assumed that the structure of human language reflects a correct understanding of the structure of external reality. Each of the traditional grammatical categories reflects one way of thinking, a *modus intelligendi* (way/mode of understanding) of items of reality - *res* or *natura communis* - and each way of understanding corresponds to one of the ways in which such items are, a *modus essendi* (way/mode of being). For a word thus to belong to some grammatical category means that it signifies some thing in a certain way, as having a certain mode of being: for a word to be a noun means that it signifies some thing in the way things are signified under their static aspect (mode), and for a word to be a verb means that it signifies some thing in the way things are signified under their dynamic mode. The noun *dolor* (pain) and the verb *dolere* (to ache) signify the same *res*, but in different ways.

While the modes of signifying are derived from correct understanding of the modes of real things, once they have been found and have found morphological expression, nothing prevents humans from inventing a word such as ‘chimera’ and furnishing it with a noun’s mode of signifying. That mode of signifying indicates that if reality has a match for the word ‘chimera’, that thing must have static being. No particular facts about the external world can be deduced from grammar, only some general structural features.

Modes of signifying were divided into two types, *essential* and *accidental*. The essential modes establish the eight parts of speech, while the accidental modes establish moods, tenses, numbers, genders, cases and so on. Like the essential modes, accidental modes are derived from features of reality. Thus in reality things can have an active, passive or neutral way of being; this is the origin of the modes of signifying traditionally called masculine, feminine and neuter gender. A word must not be marked for masculine gender if the intended significate can have only feminine being, for that would make it the sign of another sort of thing than the one intended. However, for a word to be comprehensible it is only required that the mode of being indicated by the mode of signifying is compatible with the actual modes of being of the thing in question. Hence for things neutral with respect to action and passion, the choice of gender is arbitrary.

According to several modists, any particular word-form may be analysed as follows. First, it is a vocal sound (*vox*), endowed with a significative function (*ratio significandi*), one essential mode of signifying and some accidental modes of signifying. Having a significative function makes it into a sign, and having articulated sound (*vox*) for the vehicle of its significative function makes it into the sort of sign called a *dictio*. Having an essential mode of signifying makes it a *pars orationis* (part of speech, or in other words, a constituent of a sentence).

A fundamental, and controversial, thesis of modism was that the significative function linking a particular type of sound to some specific significate (*significatum*, *res*) is irrelevant for the grammaticality of constructions. Grammaticality is the result of licit combinations of modes of signifying. Thus {[+ noun + substantive + plural + nominative] + [+ verb + plural + third person]} is a licit combination, and any two words having those modes of signifying make a grammatical construction, no matter what their lexical meanings are. Thus ‘stones speak’ is as

correct as ‘men speak’. (In the schema shown above, the terms ‘verb’ and so on replace the cumbersome names of the corresponding modes of signifying.)

Modists assumed that external reality is accessible to human understanding, and that whatever can be understood can be expressed. Different peoples may choose different ways of encoding their understanding of reality in sound, but reality is the same for all and is understood in the same way by all peoples. Consequently, all languages must be structurally identical; they must all have the same grammar, surface differences notwithstanding. Greek (the modists wrongly assumed) has lexicalized ‘number’ as far as nouns are concerned and indicates it by means of special words called articles instead of using endings as in Latin, but what matters is that both languages can express ‘number’. If the underlying grammatical system is the same for all languages, it follows that there is total translatability between them, and modists thought as much. Indeed, they found empirical evidence for the thesis in the well-known fact that logic was first developed in a Greek context and written down in Greek, then successfully translated into Latin. If logic is international and translatable, then surely nothing is untranslatable.

Among medieval thinkers, the modists presented the most articulate theory of the universality of grammar. Fourteenth-century opponents of modism, such as John Aurifaber, sometimes explicitly denied the universality thesis.

6 Multiple signification: proper and improper locutions

A pervasive theme in the thought of Anselm of Canterbury is that people often use improper locutions, and that these must be recognized as such and replaced by the proper ones if truth is to be found. A typical example of his use of this technique concerns the argument that for any thing (*A*) existing at t_1, it was true at $t_{<1}$ that *A* could come into being, and so any actual thing has a potential predecessor. For Anselm, potency and power were not distinct concepts, and the argument thus became a claim that at $t_{<1}$, *A* had the power to come into being. It follows that the actual *A* may have generated itself; thus the created world needs no creator, as it has its own principle of existence. Anselm’s way of countering this line of thought is to claim that ‘Before *A* came into being, *A* could come into being’ is a colloquial way of saying ‘Before *A* came into being, God had the power to bring *A* into being.’ The improper locution leaves the false impression that at $t_{<1}$ *A* is the holder of the power to produce *A*, when in fact that power is held by God. This distinction between a more and a less proper meaning of words recurs in many guises in medieval thought.

7 Multiple signification: metaphor, allegory and author’s intention

Language with one surface meaning and another one hiding beneath the surface was of particular concern to the theologians. For one thing, it was agreed that ‘God is just’ cannot be interpreted in the same way as ‘Peter is just’. For another, it was commonly assumed that not only does the Bible sometimes use expresions that are only meant to be understood metaphorically (for example, ‘sprout from the root of Jesse’ as a metaphor for Christ), but often a Biblical text carries several senses besides the one apparent on the surface (*sensus historicus*). Thus ‘heaven and earth’ in Genesis 1: 1 could be said to signify heaven and earth on one level of interpretation, angel and man on another, spirit and flesh on a third. In such cases we are dealing with *allegoria in verbis*, that is, words having another meaning than or besides their *prima facie* meaning.

However, following a suggestion from Augustine in *De trinitate*, medieval writers also assumed an *allegoria in factis* (allegory in deeds), in which the words do not directly have any but their surface meaning, but the thing meant in turn means something else. This occurs in the Bible when words directly signify some event (*factum*) from the time of the Old Covenant, but indirectly also signify some fact belonging to the time of the New Covenant. Thus if the text says that the people of Israel were delivered from Egypt, the words primarily refer to the liberation of the Israelites from the Egyptian captivity; but this event in turn was a sign of the future liberation from evil of the members of the Church, which is then indirectly signified by the text. Interpreting in terms of *allegoria in factis* was later called ‘typological exegesis’.

The assumption of simple *allegoria in verbis* was a way to make sense out of apparent nonsense, and was just one of many strategies used in the interpretation of authoritative texts, secular and sacred alike. After the thirteenth century, a distinction was often made between what a text means *de virtute sermonis* (in virtue of its wording), that is, the meaning the text conveys to anyone knowing the vocabulary and grammar of the language in question, and

what it means *ex intentione auctoris* (in accordance with the intention of the author). By identifying the latter meaning with his own favourite view, a medieval author could avoid declaring his disagreement with an old authority.

In 1340, the Faculty of Arts at the University of Paris forbade its masters to call an authoritative utterance false in virtue of its wording if they thought the author had intended it in a true sense. Instead, teachers were required to say that the text had two senses, one of which was false and the other true, though both senses were 'in virtue of the wording' of the text. According to the reasoning behind the decree, it constitutes no violation of the 'virtue' (force) of the wording if an author intends the words to carry an unusual sense, for words have no semantic 'virtue' independent of their users. The explicit motive for the 1340 prohibition was fear of the consequences of accepting the judgment 'false by virtue of the wording' passed on a Biblical passage, as there might then appear to be a mismatch between the Bible's wording and truth.

8 Multiple signification: systematic use of metaphors (Gilbert and the Porretans)

Gilbert de la Porrée, Bishop of Poitiers from 1142-54 and father of a school of *Porretani*, barely escaped a condemnation for heresy (see Gilbert of Poitiers) for his claim that the proposition *Deus est Deitas* (God is Godhood) is false, while *Deus est Deus Deitate* (God is God in virtue of Godhood) is true. He appeared to be violating the doctrine of divine simplicity (see Simplicity, divine). The Porretans were also ridiculed for tripling the Aristotelian categories. These two pieces of unusual doctrine were in fact part of one theory.

As Augustine and Boethius had made clear, words cannot function the same way when used about creatures and about God. Christian theologians know that 'God is just' is a true proposition because it has the support of Scripture, but they do not fully know what its terms mean. What God is cannot be fully grasped by a human mind, at least not in this life. Nor is it clear what it is for God to be; used about creatures 'is' indicates temporal being, but this cannot be so with the Creator (see God, concepts of). For a creature to be just is to have a certain property, but the doctrine of divine simplicity forbids this interpretation of 'just' in the case of 'God is just'. To speak about God implies transferring terms from the known realm of creatures into the unknown realm of the divine.

Starting from this problem, Gilbert and his followers developed a general theory about transferred language. In their theory, language was originally instituted to deal with the direct results of creation. The terms of this primary language describe the members of natural classes with their substantial and accidental forms. The ten Aristotelian categories are ten classes of words which may be used to describe a subject by signifying a predicate; a predicate in turn is a form that constitutes the subject's being something or in-some-way. In other words, a categorial term signifies a substantial or an accidental form. Only natural things have genuine forms. Hence, properly speaking, the ten categories pertain only to this primary language, that of the *naturalis facultas* (natural domain).

Artifacts, social institutions, values and other consequences of human activity constitute an ontologically secondary group of things to be named, but the *moralis* (behavioural) language used for the purpose is not structurally different from the one used for natural entities. People speak as if houses, prices and moral values were constituted by substantial and accidental properties in the same way as trees and colours. Propositions still have the shape '*x* is *F*', although '*F*' does not signify a genuine form. Most of the words of behavioural language are also loans from the primary language: thus 'high' has been borrowed from the category of quantity to be used in a transferred sense about prices. The semantic contents of the words are of different sorts, but the vocabulary still falls into ten groups, ten quasi-categories, and quite generally the structure of the language remains the same. The behavioural language owes its intelligibility to its retention of the structural features of the natural language.

The second-order language of logic and grammar (*rationalis facultas*) similarly retains the structure of the primary language. There is no real form of equivocation, yet '*x* is equivocal', which has the structure of a genuine predication, may be true; and if it is so, we are justified in talking about the equivocation of *x*.

Finally, there is theological discourse. Words are transferred to new and fundamentally unknown senses. The only way to retain some sort of intelligibility is by obeying the same structural laws, the same syntax, as our primary language. One of the relevant laws might be stated as follows: if '*X*' and '*Y*' are concrete terms, and '*Y*-hood' the abstract term corresponding to '*Y*', then (1) '*X* is *Y*' implies '*X* is *Y* by virtue of *Y*-hood', and (2) '*X* is *Y*-hood' is nonsensical or false. Applied to the case of 'God is… ' the rule yields the result that 'God is God' implies 'God is God by virtue of Godhood', while 'God is Godhood' is an unacceptable proposition. Gilbert did not intend to

undermine the doctrine of divine simplicity. His discovery was that no intelligibility at all is left if one changes syntax and the meaning of the vocabulary at the same time.

9 Multiple signification: metaphor as a key tool in ordinary communication (Roger Bacon)

The mainstream tendency in the thirteenth century was to treat a word's meaning as an unchangeable companion of its sound once the word had been 'imposed' to mean something. This view implied, *inter alia*, that a word does not lose its meaning if such things as it signifies cease to exist; 'donkeys are animals' would be a meaningful (and, most would say, true) proposition even if donkeys had become extinct, and 'donkeys' would signify the same as when donkeys did exist. Similarly, while recognizing a difference between 'the donkey' used to refer to the asinine species and to a particular donkey, most philosophers would deny that a switch from one use to the other amounts to a change of signification.

Around the 1260s, Roger Bacon developed a radically different view. According to Bacon, there is no way in which a word can be imposed simultaneously as the name of an existing and of a non-existing thing, of a universal and of a particular, and so on. Words are instituted and learned with some definite meaning, but it is in every speaker's power to change that meaning. In fact, this is done all the time, and generally tacitly. People are continually confronted with the task of expressing something they have not thought of or talked about before. They habitually solve the problem by means of metaphors, using a term whose meaning is somehow similar to what they now need to express. For example, to someone 'John' may signify a certain animate being. Standing in front of John's dead body, this person may identify the body with the words 'This is John', without intending to say that the corpse is an animate being. Tacitly, 'John' has been given a new meaning (body that used to be the material component of a human being called John). The speaker has performed a new imposition and has rendered the term equivocal.

The correct interpretation of the term in each particular case depends on the speaker's intention. The hearer may not grasp it, but generally the ability of speakers to create metaphors is matched by the ability of listeners to grasp them. To Bacon, metaphor was the individual person's fundamental tool for communicating new thoughts, and we have new thoughts all the time.

10 Multiple signification: the intentionalist current in grammar

Modistic grammar was intended to be an Aristotelian science, with principles of its own. In this the modists had been anticipated by others, but they were particularly insistent on the point that grammar cannot allow the use of information obtainable only from other sources than its own principles. According to the modists, grammar is concerned with grammaticality, not with truth or reality, and a string of words is grammatically well-formed (*congrua*) if it exhibits no mismatched modes of signifying. For this string to be a complete (*perfecta*) sentence, it must fulfil the further requirements of containing a subject (*suppositum*) and a predicate (*appositum*), and of leaving no relations of grammatical dependence between the ingredient words unsaturated. In modism, the lexical content of the sentence's constituent words is grammatically irrelevant, and so are non-linguistic circumstances that might influence the interpretation of an utterance. It follows that *turba ruunt* (the crowd are rushing) is ill-formed, whereas *capa categorica* (a categorical cloak) and *Socrates curret heri* (Socrates will run yesterday) are well-formed expressions.

By contrast, many twelfth-century thinkers had held that such expressions are ill-formed because they are unintelligible. Congruity was required both on the vocal level and on the level of sense. This meant that many putative propositions could be dismissed as not even being grammatically correct sentences. Even lack of referents for an ingredient term could rob an otherwise good sentence of sentencehood and propositionhood.

Constructions such as *turba ruunt* violate rules of concord (in this case, the rule that a verb agrees in number with its subject). However, such constructions could be found in authoritative writers and were generally classed as 'figurative'. For an 'error' to qualify as a figure instead of a mere error, it must have some reasonable excuse: *figura est vitium ratione excusatum*, as the saying went. Moreover, the reasons excusing the deviant use of words were held to be twofold: there are reasons why the deviation is possible, and there are reasons why it is demanded, *ratio qua potest fieri* and *qua oportet fieri*, respectively. Reasons of the first type explain why the deviant expression is intelligible to an audience (for example, *turba ruunt* strongly resembles the correct *turba ruit* and

turba, though of singular number, signifies a plurality of people, so somehow the plurality required by *ruunt* is present to the hearer's mind). Reasons of the second type explain why a speaker would choose the deviant formulation rather than a regular one (for example, the use of a plural verb is needed to give a vivid impression of the many individuals that compose the crowd); that is, such reasons appeal to the speaker's intentions in formulating the message.

Figurative and incomplete constructions invited grammarians to take the situation of communication into account. Further, as in the case of metaphors, appeal could be made to the speakers' intentions as a major factor in deciding the meaning attributable to an utterance. This ran counter to modist principles, but since modism did not have any good tools for dealing with deviant yet intelligible expressions, even modists would sometimes adopt 'intentionalism', as this context-sensitive grammar has been called. Robert Kilwardby and Roger Bacon were among the foremost intentionalists.

In theology, there was an ongoing debate whether the verbal components of the sacraments were in some sense bearers of a power to sanctify along with their power to signify. The less inclined a theologian was to accept this, the more it would be stressed that the sacrament's efficacy depends on the intention of its original institutor and of its administrator (the priest).

11 Multiple signification: primary and secondary signification

Anselm, in his *De grammatico*, introduced a distinction between what a word signifies by itself (*per se*) and what it signifies via something else (*per aliud*). By itself *grammaticus* (literate) signifies literacy plus having (or having literacy); indirectly it signifies a human being to whom having literacy applies. The human being is not directly signified by the word, but anyone who understands it can supply the information that literacy requires a human possessor. Anselm also uses the expression *appellare* as a synonym of *significare per aliud*.

In the twelfth century, it became common to distinguish a word's signification from its appellation or nomination. The things 'called' or 'named' (*appellata*, *nominata*) were whatever things the term applies to considered in their own right, whereas the term was said to 'signify' also or exclusively a concept, a form or the things qua carriers of some form. Thus 'white' might be said to name white things but to signify whiteness. In the thirteenth century, 'nomination' disappeared and 'supposition' replaced 'appellation' as the general term in descriptions of the varieties of reference.

Thirteenth-century writers worked not only with the distinction between signification and supposition, but also with a distinction between primary and secondary (*secund(ari)a* or *ex consequenti*) signification. A concrete accidental term (white) was said to signify primarily a form (whiteness) and secondarily whatever has that form (the thing that is white). Similarly, a substance-term like 'man' might be said to signify primarily human form (humanity) and secondarily the 'aggregate' of matter and form (man). A relative (such as 'parent') could be said to signify one relative thing (the parent) primarily and its correlative (the child) secondarily. Excessive use of the distinction led some to hold that a proposition could be true or false *per se* on account of the primary significates of its terms and have the opposite truth value *per accidens* on account of their secondary significates.

In the next century, William of Ockham used the distinction as a means to keep his ontology lean. He operated with (a) *absolute* terms such as 'dog', which signifies individual dogs and nothing else, and (b) *connotative* terms such as 'white' or 'parent', which directly signify whatever individual things are white or parents but connote (that is, secondarily signify) their whitenesses and children. By treating all terms in categories other than substance and quality as connotative, Ockham could avoid positing special quantitative entities, relative entities and the like.

Buridan made very similar use of the concept of connotation although, confusingly, he stood the twelfth-century terminology on its head, saying that 'white' directly signifies and supposits for something white but connotes or *appellat* its whiteness. Similarly, 'rich' supposits for a man and *appellat* his possessions. Names of fictitious entities (for example, 'chimera') signify nothing directly, but they connote the real entities such monsters are mistakenly supposed to consist of (a snake's tail, a lion's head and so on). Oblique cases of nouns are limited to connoting; thus in 'Plato's donkey is running', Plato is only connoted whereas a donkey is directly signified. Buridan calls what a connotative term supposits its 'matter' and what it connotes its 'form', but the latter does not imply that his *connotata/appellata* are supra-individual things.

12 Multiple signification: properties of terms (supposition, ampliation and restriction)

Handbooks of logic from the later Middle Ages generally devote much space to the so-called ‘properties of terms’. The most notable of these are signification, supposition, ampliation, restriction and distribution.

The twelfth-century notion of supposition may have part of its background in Trinitarian theology, which distinguished between *suppositio personalis* and *suppositio essentialis* according to whether a term stands for one of the divine persons or the divine essence. However, the further development of supposition theory took place in the context of logic. The core idea of the theory is that a term does not always represent the same objects, even though it retains its meaning. In different contexts, the term refers to different subsets of the totality of objects of which it may be truly predicated. The proposition ‘all the doctors will dispute next year’ is ambiguous as to whether it claims that all presently existing doctors will dispute next year or that all the doctors existing next year will be disputing then. This was expressed in the rule that a general term acting as subject for a verb of future tense supposits for things present or future. Sometimes this rule comes in the form ‘things present and future’. The first formulation shows an understanding of the term’s *supposita* as the minimal set of objects that will verify the proposition; the second formulation takes the term’s *supposita* to be the conjunction of those sets.

Another core idea was that a term’s *supposita* are the individuals falling under it in some particular sentential context, but from the very beginning this idea was somewhat obfuscated by attempts to allow the notion of supposition also to cover other variations in the way a term may be understood. A distinction between personal and material supposition was commonly accepted. In personal supposition, the word ‘donkey’ stands for members of the asinine species; in material supposition, ‘donkey’ stands ‘for itself’, that is, for the word ‘donkey’, as in ‘donkey is a noun’. Counting material use as a sort of supposition was a complication of the theory, and it was often unclear what exactly was meant by ‘itself’ in the formula ‘standing for itself’. Another complication was the acceptance in the thirteenth century of a ‘natural supposition’: lifted out of any sentential context, a term was said to have natural supposition for the totality of objects falling under its signification. This could only make sense in an age when signification was not taken as a relation between a word and extramental particulars but between a word and some non-particular entity. Many authors further accepted a *suppositio pro significato* or *suppositio simplex*, claiming that in ‘the donkey is a species’, ‘donkey’ stands for the universal signified by ‘donkey’.

The notion of supposition was most successful as a tool for dealing with tense, quantification and the effects of intentional verbs (such as ‘know’ and ‘promise’). In the fourteenth century, Ockham and Buridan removed many of the inconsistencies in the knowledge inherited from the previous century. The notion of supposition was inextricably linked to those of ampliation (widening) and restriction. The idea is this: the set of objects for which a term would otherwise stand can be made bigger or smaller by the occurrence of some other word. Typically the neutral situation was taken to be that of a subject of a present-tense proposition containing neither modal nor intentional verbs - for example, ‘every man is running’ - in which ‘man’ was said to supposit for presently existing men. The addition of an adjective (‘every young man is running’) will restrict the set of verifiers by excluding old men. Saying ‘may be running’ instead of ‘is running’ extends the range of the term by adding an alternative set of verifiers, that of men who are yet to be (according to some authors) or that of possible men (according to others).

Fourteenth-century logic usually operated with four domains (sometimes called ‘tenses’): the past, the present, the future and the possible. Marsilius of Inghen proposed to add a fifth ‘tense’ or domain, that of the imaginable, in order to provide verifiers for propositions of the type ‘*X* is imagined’, in which ‘*X*’ refers to an object that is indeed imaginable even though it may not be physically possible. According to his proposal, the verb ‘is imagined’ ampliates its subject so that it supposits not only for actually existing things but also for imaginable ones. This was a considerable step in the direction of developing a theory of the language of fiction, a subject to which scholastic philosophers had paid little attention save for one much-discussed example: propositions about chimeras. Chimeras were supposed to be impossible entities, though it was often unclear what sort of impossibility was involved. Marsilius’ master, John Buridan, had sided clearly with logical impossibility, and to underline this he had ended up changing the medieval standard definition of a chimera, ‘animal composed of lion, cow and young woman’, into ‘animal composed of parts that cannot possibly be put together’. However this invited the question of what to do about logically possible, and hence imaginable, things that nature does not allow. This was the background of Marsilius’ proposal of an ‘imaginable tense’.

13 Multiple signification: ambiguity

After about 1150, Aristotle's *Sophistici Elenchi* provided the general framework for most discussions of ambiguity. Aristotle used six fallacies 'in speech'. Following Galen, medieval writers usually took these six fallacies to be due to three types of ambiguity (*duplicitas*), occurring in two types of expressions, single words (*dictiones*) or complexes of words (*orationes*). The resulting system is described in Table 1:

Table 1

Expression	**Duplicity** actual	potential	imaginary (*fantastica*)
Simple	equivocation	accent	figure of speech
Complex	amphiboly	composition and division	

An actual ambiguity was supposed to be present any time the expression is uttered. The standard example of equivocation was *canis*, which can refer to (1) a dog, (2) a dogfish, or (3) the Dog Star. According to late thirteenth-century doctrine, such a term could be disambiguated only by the immediate addition of a determination suitable to only one of the significates. Thus 'dog' was supposed to have only sense (1) in 'a barking dog' while keeping all three senses in 'a dog is barking'. Aristotelian equivocation was designed primarily for nouns, but medieval writers sometimes used equivocation in all parts of speech, even pronouns and prepositions. For example, some held that 'this' is equivocal with respect to all things that may be pointed at, and the preposition *de* was considered equivocal because it can indicate local origin or material composition, as in *Socrates est de terra*, which can mean (1) 'Socrates is *from* a country' or (2) 'Socrates is *composed of* earth'.

In genuine equivocation, one sense cannot be derived from another. Each sense derives from an independent imposition when a sound was chosen, without concern for whether it had already been selected to label some other thing. According to the principles of modist grammar, this would mean that there could not strictly be such a thing as an equivocal noun. There could be one phonetic entity common to three significates, but then there would be as many nouns as there are significates. When this was realized in the 1270s, an attempt was made to save the essentials of the theory while preserving the intuition that *canis* is exactly one noun with the same grammatical properties no matter which sense is intended by the speaker or needed to make the proposition true in which the word occurs. It was proposed that there are both active and passive modes of signifying, with the active ones residing in the word and their passive counterparts (indistinguishable from modes of being) in the things signified. One active mode may correspond to three different passive ones, it was claimed, and syntactic properties depend on the active modes. This was not a happy solution, since it introduced two levels of grammatical structure with the surface structure determining syntax. As so often in the history of linguistics, ambiguity had proved a hard test for a theory (see Ambiguity).

In the principal type of equivocation, the different senses of the word are not ordered. However, standard doctrine also accepted types of equivocation in which one sense could be described as primary or proper and the other(s) as secondary, improper or transferred, as when 'running' primarily describes fast foot-propulsion, but metaphorically describes the movement of a river. Thirteenth-century writers adopted the name 'analogy' for equivocation in which one thing is signified primarily and one or more others secondarily. If a term signifies all its significates, equally it always needs a determination in order to be disambiguated. An analogical term unaccompanied by any determination was held to represent only its primary significate. An Aristotelian example was often used: 'healthy' primarily signifies a property of animate beings, while secondarily it signifies (a) a property of animal excretions, namely that of indicating that the animal they come from enjoys good health, and (b) a property of food

and drugs, that of being conducive to the health of some animate being. To actually signify (a) or (b), ‘healthy’ needs to be joined with words like ‘urine’ or ‘potion’. This particular type of analogy was called *analogia attributionis*. A considerable debate arose concerning another example, ‘man’, which some held to signify primarily a compound of body and rational soul and secondarily a mere body. In this view, the primary significate is represented by ‘man’ in ‘this is a man’ and the secondary one by the ‘man’ of ‘this is a dead man’, and so the latter proposition does not entail the former.

Some modists assumed that the significative function (*ratio significandi*) assigned to an analogical term on its imposition was a complex one selecting different significates in different contexts. Radulphus Brito gave up considering the multiple meaning a property of the word itself prior to its actual use. The secondary signification comes about, he claimed, when a speaker uses a word improperly and a listener catches his intention thanks to a charitable interpretation of an utterance that would make little sense if taken literally. In so claiming, however, Brito sacrificed the modist axiom that only two factors contribute to the sense of an utterance: lexical (imposition-given) meaning and modes of signifying.

In the thirteenth century, it became standard doctrine that *ens* (being) is analogical as to substance and accidents. Attempts were made to use this distinction as a model for the relation between the use of terms for both creatures and the creator, and a long debate about the analogy or univocity of being ensued (see Being).

The notion of potential ambiguity led to a theory according to which the *litterae* (letters, phonemes) are the matter of an expression, which can be actualized only by the addition of a form, the ‘way of pronouncing’ (*modus pronuntiandi*), consisting in stress, pauses and the like. For example, in English the letters ‘i-n-v-a-l-i-d’ are the common matter of two distinct actual words, ‘*in*valid’ and ‘in*val*id’. The fallacy of composition and division was used, *inter alia*, to explain cases in which a string of words can be interpreted either as an atomic or as a molecular proposition; thus *falsum est Socratem esse hominem si est animal* can be read as an atomic proposition (‘that-Socrates-is-a-man-if-he-is-an-animal is false’) or a molecular proposition (‘that-Socrates-is-a-man is false, if he (or: there) is an animal’). The distinction between a ‘composite’ and a ‘divided’ sense also played a central role in the analysis of modal propositions. Often, *possibile est sedentem stare* was said to mean ‘that-a-man-stands-while-sitting is possible’ in the composite (or *de dicto*) sense, and ‘a sitting man has the possibility of standing’ in the divided (or *de re*) sense.

The notion of imaginary ambiguity was always somewhat unclear, but ‘figure of speech’ became a head under which the scholastics could examine the relation between (onto)logical and grammatical categories, for example, action versus active voice, substance versus substantive and individual versus singular number.

Use-mention ambiguity was usually dealt with in terms of supposition. In ‘man is an animal’, ‘man’ supposits personally (that is, for things signified by the word ‘man’), while in ‘man is monosyllabic’, ‘man’ supposits materially (for phonetic entities equiform with the word pronounced). On the other hand, ambiguous supposition was often classified as a mode of either equivocation or figure of speech.

Logicians classified words as either categorematic (roughly = nouns + verbs) or syncategorematic (mostly quantifiers, modal terms, prepositions and conjunctions). There were few ancient precedents for inquiry into the contribution of syncategoremata to the sense of a proposition, but a major research effort was made in the later Middle Ages, especially with regard to ambiguity traceable to the occurrence of such logical operators, including scope ambiguity.

14 The significate of propositions

Any syntagm could be called an *oratio*, whether a sentence fragment such as ‘a white horse’ or a whole sentence (see Propositions, sentences and statements). Among the different types of sentences, the declarative was both the grammarians’ preferred and the logicians’ sole object of study. In logic, the technical term was *propositio*, and the standard definition (from Boethius) was *propositio est oratio verum vel falsum significans* (a proposition is a syntagm that signifies a truth (or some true thing) or a falsehood (or some false thing)). Propositions were said to be either categorical (that is, atomic) or hypothetical (that is, molecular; in medieval terminology, conditionals are only one among several types of hypothetical propositions). The Boethian definition suggests that there are entities - truths and falsehoods - signified by propositions. In the twelfth century, the significate of a proposition was usually called *enuntiabile* (statable) or *dictum (propositionis)*, (what is said (by the proposition)). Thus ‘Socrates is

running' (*Socrates currit*) signifies the *dictum* that-Socrates-is-running (*Socratem currere*), and the string of words 'that Socrates is running' is the name of the *dictum* (*nomen dicti*). It was commonly, though not universally, assumed that the objects of knowledge and belief are *dicta*.

The ontological status of *dicta* was controversial. Nominalists could not give them full reality, for that would be a repetition of the error of hypostatizing the significates of single words. On the other hand, a *dictum* is either a truth or a falsehood, and some truths at least might seem to be eternal; but if *dicta* were eternal, there would be other eternal things than God, which was scarcely an acceptable thought. The majority view seems to have been that *dicta* are a sort of quasi-thing.

Controversy also surrounded the nominalist thesis *quod semel est verum semper est verum* (what is true at one time is always true). This thesis was theologically motivated. The pre-Christian patriarchs were assumed to have believed that Christ would be born, and that was a truth; Christians believe that Christ has been born, and that is a truth. Do Christians believe in the same truth as did the patriarchs? Or did the truth in which the patriarchs believed cease to be a truth when Christ was born? In the nominalist view, 'Christ will be born' and 'Christ has been born' express the same truth supposing that they are pronounced before and after his birth, respectively. The same true *dictum* has two differently tensed names, 'that Christ will be born' and 'that Christ has been born'. The nominalist thesis amounts to a claim that the class of linguistic expressions that are verified by the same event or events express the same *dictum*. Non-nominalists replied that the differently tensed 'that' clauses are or signify different statables; some tried to escape unfortunate theological consequences by denying that statables are what knowledge and belief have for their objects.

After the twelfth century, interest in statables abated somewhat, but it came back strongly in the fourteenth. Most authors then shared the belief that propositions, whether mental or spoken, are individual entities existing in time. Previously there had been a tendency to neglect the difference between proposition token and proposition type, but now the distinction came into the foreground, especially thanks to John Buridan. He also contributed to the popularity of a three-part analysis of propositions into subject, copula and predicate, with the copula treated as a term in its own right. The copula of a mental proposition could be described as a 'complexive concept' uniting the extremes. Pierre d'Ailly, while accepting the three-part analysis, insisted that ontologically a mental proposition is a unity; it is tripartite only in the sense that it is a cognition equivalent to the conjunction of three other cognitions.

Philosophers influenced by Ockham often held that the objects of knowledge and belief are mental propositions. According to Ockham, a proposition's truth is nothing other than the proposition itself, and it has no other significate than the individual things to which its terms refer. Some found this implausible, *inter alia* because truth would then seem to depend on the existence of propositions in human minds rather than on external reality. Instead, they (most famously Gregory of Rimini), took the object of knowledge and bearer of truth to be a state of affairs, *complexe significabile (tantum)* (something that can (only) be signified in a complex way, for example by a proposition, not by a term). The same quasi-entity could also be called *totale* (or *adaequatum*) *significatum propositionis* (the total (or: adequate) significate of a proposition), as opposed to the significates of each of its ingredient terms. Gregory vigorously denied that total significates are existing entities, whether mental or extramental. Consequently, he thought, God's uniqueness as an eternal being is not compromised by claiming that even before the beginning of the world 'complexly signifiables' were true or not true in virtue of the sign and source of all truth, God himself.

Walter Burley held the unusual view that just as vocal terms have real things as their ultimate significata, so vocal propositions ultimately signify propositions in external reality. However, he seems never to have reached a satisfactory answer to the question of what sort of thing such a *propositio in re* is. In one of his formulations, a real proposition has the things signified by its subject and predicate for its matter and a composition or division of that matter (in affirmation or negation) contributed by the mind for its form.

15 Syntax

The explicit aim of all medieval grammar was to spell out the rules of grammaticality (*congruitas*), or in other words, syntactical rules. The key term in grammatical syntax was *constructio*, standardly defined as *congrua dictionum ordinatio* (congruous linking of words), and most attention was paid to the pairwise linking of words

rather than to more complex linguistic entities. The elementary rules of concord were easily describable in terms of grammatical accidence; thus substantive nouns and adjectives were both said to have the accidents of gender, number and case, and the rule ‘a substantive and its adjective must agree in gender, number and case’ was thus a rule on how to match the accidents of one word with those of another. In modistic grammar, such concord could be described as a ‘proportionality of modes of signifying’, with the idea being the same, that certain secondary semantic features of one of the words to be construed require or forbid certain semantic features of the other word.

The relationship between the words to be construed was most often described in terms of determination, government (*regimen*), demand (*exigentia*), or dependence with the last term dominating in modistic grammar. In about 1300, a construction was taken to be the union of a *dependens* (dependent, unsaturated constituent) and a *dependentiam terminans* (a constituent that saturates the dependence). Just as in logic a proposition is a union or composition of subject and predicate, most simply expressed in the form ‘noun + verb’, so in grammar a sentence is a union of *suppositum* and *appositum*, most simply expressed as ‘noun + verb’, with the verb being the dependent constituent requiring saturation such as can be provided by the noun that signifies *per modum per se stantis* (purports that its significate is a self-sufficient entity). In the complex sentence ‘Socrates goes to the church’ the verb ‘goes’ has two relations of dependence, one being saturated by ‘Socrates’, another by ‘(to) the church’. As far as the verbs are concerned, the medieval notion of dependence had approximately the same function as the modern notion of valency.

16 Informative and performative locutions

While most medieval linguistic thought concentrated on descriptive propositions, a conceptual apparatus for talking about performative locutions did exist, at least since the early thirteenth century. This was provided by a distinction between *actus significatus* (or *conceptus*) and *actus exercitus*, act(ion) signified or conceived versus act(ion) performed. The distinction was used in several ways.

First, if *A* says ‘Bravo!’ when hearing *B* say ‘They sacrificed themselves’, the exclamation may either comment on the act of self-sacrifice signified by *B*’s statement or on the act performed by *B* in uttering the statement. This raises the question of whether in the first case ‘Bravo’ is construed as an adverb with the word ‘sacrificed’ but in the second case with an event, not with a word.

Second, some words signify what others perform. The words ‘I address you’ signify an action that is performed by saying ‘Hello’; ‘or’ and ‘every’ perform - or are means of performing - what ‘disjoins’ and ‘distributes’ signify. Some logicians held that syncategoremata (that is, logical operators) were purely performative, having no signification besides their function (*officium*). According to Ockham, ‘*B* is predicated of *A*’ should not be confused with ‘*A* is *B*’: in the former proposition predication is signified, in the latter it is performed.

Third, in some situations words which are usually informative (signify an action) become performative: when a priest says ‘I baptize you’, his words are instrumental in performing what they describe. Finally, some thought a confusion of signification with performance was the source of the Liar paradox, the act of saying signified by ‘I am saying a falsehood’ being confused with the act of saying performed by uttering those words.

17 The diversity of languages

Medieval theoreticians were aware of the possibility of non-vocal sign systems fulfilling the same role as ordinary languages, but spent little time on the investigation of any other non-vocal communication than that between angels, who were supposed to use no material signs because of their immateriality. The theoreticians were, of course, aware of the diversity of human languages but usually merely acknowledged the phenomenon, noting that different peoples can use different conventional signs for the same things. The modists of the late thirteenth century produced an account of how the shared human rationality and shared external reality resulted in one common linguistic structure (grammar), capable of accidental diversification through the arbitrary choice of material (sounds) to carry the sign function. Modists did not typically raise the question of a possible common origin of the different languages’ phonetic manifestations. Medieval writers who did raise the question usually agreed that the first human language, that of Adam, was Hebrew, and that there was some sort of connection linking Hebrew to the many languages spoken after the Babylonian confusion, but few apart from Roger Bacon and Dante had any clear ideas of linguistic development.

Dante Alighieri presented a remarkably well-developed theory of linguistic change in his *De vulgari eloquentia*. Dante operates with two types of language: (a) such as is the maternal tongue of certain people, (b) such as is no-one's primary language. Languages of type (a), the vernaculars, are subject to change, while type (b) languages, the 'grammatical' ones such as Latin and classical Greek, are exempt from change, their form having been determined once and for all by codified rules. The 'grammatical' languages are artificial; they have been created to avoid the breakdown of communication that spatial and temporal distances cause in natural languages. Contrary to the humanists of later periods, Dante did not see Latin as Cicero's maternal tongue, though he was clearly aware that it was based on ancient Roman vernacular.

To Dante the natural languages, the vernaculars, are the nobler ones. All derive from the language that Adam acquired on acquiring his soul. Thus far it might appear as if Dante followed modist teaching, for if Adam's acquisition of language was inseparable from his acquisition of soul, language is innate in humans. However, whereas the modists only claimed that all languages share an abstract structure, Dante saw them all as historical developments of one original language, that of Adam, which was not just an abstract structure but a fully fledged language with phonetics, vocabulary and syntax. He tried to make sense of the Babylonian confusion by taking it to represent the emergence of sociolects, each craft developing its own way of speaking. However, his finest observations concern the way language changes with time and how geographical isolation can make two branches of a language develop in different ways. Applying his theory to European languages, he assumed that all derive from three languages that must have arrived in very ancient times. One of these was the ancestor of Northern and Eastern European languages (about which Dante obviously knew little; Roger Bacon had been better informed), the second of the Romance languages, and the last of Greek.

Since Dante's purpose was to persuade his readers that Italian could serve as a language of refined literature, he had to face the argument that is always brought up when a new literary language is being formed, namely that there is no such common language with a describable form of its own but only a large number of local dialects. His counterclaim was that it makes sense to have a supra-regional Italian language, identical with no particular dialect but acceptable to educated people everywhere.

18 An artificial language

No medieval author envisaged an artificial, uninterpreted language. Nevertheless, some developments in logic pointed in that direction. The use of variables was known from Aristotelian syllogistics and was employed by medieval writers in the same way as Aristotle had done. Moreover, a number of standard examples developed into quasi-variables, the name 'Socrates' having any human proper name for its value, 'Caesar' representing any proper name of a human being from the past, 'currit' any present-tensed active verb, and so on. Thus, if a logician asked whether the proposition 'Caesar is a man' is true, he was asking about propositions with the proper name of a no-longer-existing individual for their subject term. Buridan explicitly recognized that the standard examples in jurisprudence, 'Titius' and 'Bertha', were not genuine proper names; they do not elicit singular concepts, but common ones.

Logicians also came close to creating an artificial language when they stated rules of supposition, often reflecting tendencies rather than hard and fast rules of Latin. Thus according to many logicians, *asinus cuiuslibet hominis currit* equals 'there is a donkey such that it belongs to just any man, and this donkey is running' while *cuiuslibet hominis asinus currit* means 'for just any man it is true that his donkey runs'. Though not quite aware of the fact, these logicians had created fragments of an artificial disambiguated language. This was to earn them much scorn from humanists, who could see no good in unnatural Latin.

See also: Aristotelianism, medieval; Language, philosophy of; Language, Renaissance philosophy of; Logic, medieval; Natural philosophy, medieval §8; Nicholas of Autrecourt; Nominalism; Semantics; William of Ockham §6

STEN EBBESEN

References and further reading

Biard, J. (1989) *Logique et théorie du signe au XIV^e^ siécle*, Paris: Vrin.(Especially valuable as an introduction to some lesser-known thinkers from the fourteenth century.)

Bursill-Hall, G.L., Ebbesen, S. and Koerner, K. (eds) (1990) *De ortu Grammaticae: Studies in medieval*

grammar and linguistic theory in memory of Jan Pinborg, Studies in the History of the Language Sciences 43, Amsterdam: Benjamins.(Articles on medieval logic and grammar.)

Ebbesen, S. (1987) 'The Way Fallacies Were Treated in Scholastic Logic', *Université de Copenhague. Cahiers de l'Institut du Moyen Age Grec et Latin* 55: 107-34.(Survey article with extensive bibliography.)

Ebbesen, S. (ed.) (1995) *Sprachtheorien in Spätantike und Mittelalter*, Geschichte der Sprachtheorie 3, Tübingen: Narr.(Articles in English, German and French on a broad variety of themes. Also useful as a bibliographical guide.)

Kneepkens, C.H. (1987) *Het iudicium constructionis. Het leerstuk van de constructio in de 2de helft van de 12de eeuw*, 4 vols, Nijmegen: Ingenium.(On twelfth-century grammarians' theories of syntax. Discussion in vol. I followed by text editions in vols II-IV.)

Kneepkens, C.H. (1990) 'On Medieval Syntactic Thought with Special Reference to the Notion of Construction', *Histoire Épistémologie Langage* 12: 139-76.(Can be used as a guide to the scarce literature on medieval syntactic theory.)

Kretzmann, N. (ed.) (1988) *Meaning and Inference in Medieval Philosophy: Studies in Memory of Jan Pinborg*, Synthese Historical Library 32, Dordrecht: Kluwer.(Contains several relevant papers.)

Kretzmann, N., Kenny, A. and Pinborg, J. (eds.) (1982) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(Several relevant articles.)

Libera, A. de (1996) *La querelle des universaux. De Platon à la fin du Moyen Age*, Paris: Éditions du Seuil. (Provides an up-to-date account of medieval theories of universals.)

Marmo, C. (1994) *Semiotica e linguaggio nella scolastica: Parigi, Bologna, Erfurt 1270-1330*, Roma: Istituto Storico Italiano per il Medio Evo.(A thorough study of modistic theory; uses both grammatical and logical sources.)

Nuchelmans, G. (1973) *Theories of the Proposition: Ancient and Medieval Conceptions of the Bearers of Truth and Falsity*, North-Holland Linguistic Series 8, Amsterdam: North Holland.(A classic with strong impact on later research.)

Panaccio, C. (1991) *Les mots, les concepts et les choses*, Montreal, Que.: Bellarmin, and Paris: Vrin.(A penetrating analysis of Ockham's theory of language.)

Pinborg, J. (1967) *Die Entwicklung der Sprachtheorie im Mittelalter*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters, Texte und Untersuchungen 42.2, Münster: Aschendorff, and Copenhagen: Frost-Hansen.(The classic on modistic grammar.)

Pinborg, J. (1972) *Logik und Semantik im Mittelalter. Ein Überblick*, Stuttgart: Frommann-Holzboog.(Survey with extensive quotations of original texts.)

Pinborg, J. (ed.) (1976) *The Logic of John Buridan*, Copenhagen: Museum Tusculanum.(Somewhat outdated, but still useful as an introduction to Buridan's semantics.)

Pinborg, J. (1984) *Medieval Semantics, Selected Studies on Medieval Logic and Grammar*, ed. S. Ebbesen, London: Variorum.(Several relevant papers.)

Rijk, L.M. de (1962-7) *Logica Modernorum*, 2 vols in 3 parts, Assen: Van Gorcum.(The richest single collection of twelfth-century logic texts with extensive discussion by the editor; partly outdated, but still invaluable.)

Rijk, L.M. de (1989) *Through Language to Reality: Studies in Medieval Semantics and Metaphysics*, ed. E.P. Bos, Northampton: Variorum.(Several relevant papers.)

Rijk, L.M. de and Braakhuis, H.A.G. (eds.) (1987) *Logos and Pragma: Essays on the Philosophy of Language in Honour of Professor Gabriel Nuchelmans*, Nijmegen: Ingenium.(Several papers about the meanings of terms and propositions.)

Rosier, I. (1983) *La Grammaire spéculative des Modistes*, Lille: Presses Universitaires de Lille.(Valuable survey of modist doctrine.)

Rosier, I. (1994) *La parole comme acte. Sur la grammaire et la sémantique au XIII^e^ siècle*, Paris: Vrin.(On the 'intentionalist' current in grammar and on a wide range of linguistic topics in medieval grammar, philosophy and theology, with special attention paid to Roger Bacon; rich bibliography.)

Vineis, E. and Maierù, A. (1990) 'La linguistica medioevale', in G.C. Lepschy (ed.) *Storia della linguistica* II, Bologna: Il Mulino, 11-168.(A good introduction to the grammatical tradition in particular, with rich bibliography.)

Language, philosophy of

Philosophical interest in language, while ancient and enduring (see Language, ancient philosophy of; Language, medieval theories of; Language, Renaissance philosophy of; Language, early modern philosophy of), has blossomed anew in the past century. There are three key historical sources of the current interest, and three intellectual concerns which sustain it.

Philosophers nowadays often aspire to systematic and even mathematically rigorous accounts of language; these philosophers are in one way or another heirs to Gottlob Frege, Bertrand Russell, Ludwig Wittgenstein and the logical positivists, who strove to employ rigorous accounts of logic and of meaning in attempts to penetrate, and in some cases to dispel, traditional philosophical questions (see Logical positivism). Contemporary philosophers, too, are often attentive to the roles that philosophically interesting words (like ‘know’, ‘true’, ‘good’ and ‘free’) play in ordinary linguistic usage; these philosophers inherit from ‘ordinary language philosophers’, including G.E. Moore, J.L. Austin and again Wittgenstein, the strategy of finding clues to deep philosophical questions through scrutiny of the workaday usage of the words in which the philosophical questions are framed (see Ordinary language philosophy).

Philosophical interest in language is maintained by foundational and conceptual questions in linguistics, quintessentially philosophical problems about the connections between mind, language and the world, and issues about philosophical methodology. These springs sustain a rich and fascinating field of philosophy concerned with representation, communication, meaning and truth.

1 Philosophy of linguistics

Language is an impressive and fascinating human capacity, and human languages are strikingly powerful and complex systems. The science of this capacity and of these systems is linguistics. Like other sciences, and perhaps to an unusual degree, linguistics confronts difficult foundational, methodological and conceptual issues.

When studying a human language, linguists seek systematic explanations of its *syntax* (the organization of the language’s properly constructed expressions, such as phrases and sentences; see Syntax), its *semantics* (the ways expressions exhibit and contribute to meaning; see Semantics), and its *pragmatics* (the practices of communication in which the expressions find use; see Pragmatics).

The study of syntax has been guided since the 1960s by the work of Noam Chomsky, who, in reaction to earlier behaviourist and structuralist movements in linguistics (see Behaviourism, analytic; Behaviourism, methodological and scientific; Structuralism in linguistics; Saussure, F. de), takes an unapologetically cognitivist approach. Human linguistic capacities, he holds, issue from a dedicated cognitive faculty whose structure is the proper topic of linguistics. Indeed, Chomsky construes at least the study of syntax and (large parts of) semantics as attempts to uncover cognitive structures. Finding impressive commonalties among all known natural languages, and noting the paucity of evidence and instruction available to children learning a language, Chomsky suggests that surprisingly many features of natural languages stem from innate characteristics of the language faculty (see Chomsky, N.; Language, innateness of).

Whereas contemporary philosophers have tended to stay at a remove from work in syntax, discussing rather than doing it, semantics is another matter entirely. Here many of the great strides have been made by philosophers, including Gottlob Frege, Bertrand Russell, Ludwig Wittgenstein, Rudolf Carnap, Richard Montague and Saul Kripke. (However, quite a number of linguists and logicians who do not call themselves philosophers also have contributed heavily to semantics.) One major strand in semantics in the past century has consisted in the development and careful application of formal, mathematical models for characterizing linguistic form and meaning (see Semantics, game-theoretic; Semantics, possible worlds; Semantics, situation).

Pragmatics, at least as much as semantics, has benefited from the contributions of philosophers. Philosophical interest in pragmatics typically has had its source in a prior interest in semantics - in a desire to understand how meaning and truth are situated in the concrete practices of linguistic communication. The later Wittgenstein, for instance, reminds us of the vast variety of uses in which linguistic expressions participate, and warns of the danger of assuming that there is something aptly called their *meanings* which we might uncover through philosophy. J.L.

Austin seeks in subtleties of usage clues to the meanings of philosophically interesting terms like 'intentional' and 'true'. Austin keeps a careful eye to the several different things one does all at once when one performs a 'speech act' (for instance: uttering a sound, voicing the sentence 'J'ai faim', saying that one is hungry, hinting that one's companion might share their meal, and causing them to do so). His taxonomy has provided the basis of much subsequent work (see Speech acts; Performatives). H.P. Grice, while critical of some of Austin's methods, shared the aim of distilling meaning from the murky waters of use. Grice portrays conversation as a rational, cooperative enterprise, and in his account a number of conceptions of meaning figure as central strategies and tools for achieving communicative purposes. Grice's main concern was philosophical methodology (see §3), but his proposals have proven extremely popular among linguists interested in pragmatics (see Communication and intention; Meaning and communication). Recently, philosophers and linguists have become increasingly persuaded that pragmatic concerns, far from being mere addenda to semantics, are crucial to the questions of where meaning comes from, in what it consists, and how the many incompletenesses and flexibilities in linguistic meaning are overcome and exploited in fixing what speakers mean by their words on particular occasions (see Pragmatics; Implicature; Metaphor).

Our focus on language should not omit a field of study with a rather broader scope, namely *semiotics*, which is the study of signs and signification in general, whether linguistic or not. In the view of the scholars in this field, the study of linguistic meaning should be situated in a more general project which encompasses gestural communication, artistic expression, animal signalling, and other varieties of information transfer (see Semiotics; Animal language and thought).

2 Meaning: language, mind and world

Philosophy aims at intellectually responsible accounts of the most basic and general aspects of reality. Part of what it is to provide an intellectually responsible account, clearly, is for us to make sense of our own place in reality - as, among other things, beings who conceive and formulate descriptions and explanations of it.

In framing issues about our roles as describers and explainers, philosophers commonly draw a triangle in which lines connect 'Language', 'Mind' and 'World'. The three lines represent relations that are keys to understanding our place in reality. These relations in one or another way constitute the *meaningfulness* of language.

Mind ↔ *World.* Between Mind and World there are a number of crucial relations studied by philosophers of mind. Among these are perception, action, the mind's bodily constitution and intentionality (the mind's ability to think *about* what is in the world) (see Mind, philosophy of).

Mind → *Language.* Using and understanding language is a heavily mental activity. Further, this activity seems to be what the real existence of meaningful language consists in. In short, mind invests *meaning* in language.

Theorists of language focus on the Mind/Language connection when they consider *understanding* to be the cornerstone concept, holding, for instance, that an account of meaning for a given language is simply an account of what constitutes the ability to understand it (see Meaning and understanding). Philosophy has seen a variety of accounts of wherein understanding consists. Many have been attracted to the view that understanding is a matter of associating the correct ideas or concepts with words (see, for instance, Locke, J.; Frege, G.; Language of thought). Others have equated understanding with knowing the requirements for accurate or apt use of words and sentences (see, for instance, Davidson, D.; Dummett, M.A.E.). Still others find the key to understanding in one's ability to discern the communicative goals of speakers and writers (see, for instance, Grice, H.P.), or more directly in one's ability to 'pass' linguistically, without censure (see, for instance, Wittgenstein, L.). Certainly, these approaches do not exclude one another.

Some philosophers focus more on production than consumption - on the speaker's side of things - analysing linguistic meaning in terms of the goals and practices of speakers, and in terms of relations among communities of speakers (see Grice, H.P.; Communication and intention; Language, conventionality of; Language, social nature of).

Many of the philosophers who see understanding and use as the keys to linguistic meaning have held that the meaningfulness of language in some sense derives from mental content, perhaps including the contents of beliefs, thoughts and concepts. This enhances the interest of *cognitive semantics*, which is a thriving field of study (see

Semantics; Semantics, conceptual role; Semantics, informational; Semantics, teleological; Concepts).

It has not gone unquestioned that mind indeed can assign meaning to language, and in fact scepticism about this has figured quite prominently in philosophical discussions of language. Wittgenstein has been read as at least flirting with scepticism that there is anything our minds *can* do that would constitute meaning one thing rather than another (see Wittgenstein, L. §§10-12; Meaning and rule-following; Private states and language). W.V. Quine, starting from the thought that meaning is whatever good translation captures, and on arguments that good translation is not squarely dictated by any real facts, concludes that meaning is highly indeterminate. Quine is not alone in the view that linguistic and mental meaning are best seen not as 'out there' to be discovered, but rather as partly constituted or constructed by our practices of interpreting and translating (see Quine, W.V.; Davidson, D.; Dennett, D.C.; Lewis, D.K.; Radical translation and radical interpretation).

Language → Mind. If mind assigns meaning to language, so also language *enables* and *channels* mind. Acquiring and trafficking in a language brings one concepts, thoughts and habits of thought, with all sorts of consequences (see Sapir-Whorf Hypothesis; Linguistic discrimination; Language and gender). Indeed, having language is so crucial to our ability to frame the sophisticated thoughts that appear essential to language-use and understanding, that many doubt whether mind is 'prior' to language in any interesting sense (see Meaning and communication; Davidson, D.).

Language ↔ World. Since language is the vehicle of our descriptions and explanations of reality, philosophers are concerned about what if anything makes for a *true* or *apt* characterization of reality. Philosophers have these concerns for reasons of philosophical methodology (which we will come to in a moment), but also owing to the naturalness and plausibility of a certain picture of meaning.

According to this picture, the key to meaning is the notion of a *truth-condition*. A statement's meaning determines a condition that must be met if it is to be true. For example, my statement 'Ireland is larger than Manhattan', given what it means, is true just in case a certain state of affairs obtains (namely, a certain island's being larger than a certain other island). According to the truth-conditional picture of meaning, the core of what a statement *means* is its truth-condition - which helps determine the way reality is *said to be* in it - and the core of what a word means is the contribution it makes to this (perhaps, in the case of certain sorts of word, this would be what the word *refers* to) (see Semantics; Meaning and truth; Reference).

While the truth-conditional picture of meaning has dominated semantics, a serious challenge has been presented by philosophers, including Michael Dummett, who urge that the key to meaning is a notion of *correct use*. According to this alternative picture, the core of a sentence's meaning is the rule for its appropriate utterance. Of course, the two pictures converge if sentences are correctly used exactly when they are true. The interest of the distinction emerges only when (a 'realist' conception of) truth is dislodged from this role, whether because of scepticism about truth itself, or because truth is seen as too remote from the crucible of social practice to be the meaning-relevant criterion for correct use (see Realism and antirealism; Intuitionistic logic and antirealism; Meaning and verification; Dummett, M.A.E.; Truth, pragmatic theory of; Truth, deflationary theories of; Truth, coherence theory of; Truth, correspondence theory of). The challenge illustrates a sense in which the Mind/Language and Language/World connections can seem to place a tension on the notion of meaning (meaning is whatever we cognitively *grasp*, but the meaning of language just is its bearing on the world).

3 Linguistic philosophy

Apart from language's interest as a target of science and its centrality to our self-conception as describers of reality, language plays a key methodological role in philosophy. It is this role perhaps more than anything else that has explained the continued close attention paid to language in the past century by philosophers working in such varied areas as epistemology, aesthetics, ethics, metaphysics, the philosophy of science and the philosophy of mind.

The methodological role of language in philosophy is most easily explained by example. A philosopher is interested in the nature of value; they want to know *what goodness is*. Language enters when they observe that goodness is what is attributed when we *say* of a thing that it 'is good'. So the philosopher focuses on certain statements, and seeks an understanding of what such statements *mean* and in general of how they work. They explore whether such statements are ever objectively true or false, whether their truth or aptness varies from

speaker to speaker, whether a satisfying explanation of them entails that the word 'good' refers to or expresses a genuine characteristic (of actions, states of affairs, persons, and so on), and how their meaning relates to the distinctive sorts of *endorsement* that such statements commonly convey (see Analytic ethics; Emotive meaning).

The pattern exhibited in the example of value is apparent throughout philosophy. We are interested in knowledge, fiction, necessity, causation, or sensation, so we find ourselves studying statements *about* what interests us: statements attributing knowledge, describing fictions, asserting necessities, assigning causes and reporting sensations. Tools from the philosophy of language make available quite a number of views about what these statements mean and in general about how they do their expressive and communicative work; and these views inform and support philosophical positions on the real objects of philosophical interest. There have been dramatic and no doubt exaggerated claims about such techniques - for instance, that philosophy should simply *consist* in this sort of study of language. But it is if anything an understatement to say that linguistic sophistication has deepened philosophical understanding and has advanced debate in nearly all areas of philosophy (see Conceptual analysis).

See also: Adverbs; Ambiguity; Analyticity; Anaphora; Compositionality; Counterfactual conditionals; Deconstruction; Demonstratives and indexicals; Descriptions; Discourse semantics; Fiction, semantics of; Indicative conditionals; Indirect discourse; Language, Indian theories of; Logic in China; Logic, philosophy of; Logical form; Mass terms; Meaning in Islamic philosophy; Meaning, Indian theories of; Moscow-Tartu School; Proper names; Propositional attitude statements; Propositions, sentences and statements; Questions; Religious language; Scope; Semantics; Semiotics; Sense and reference; Vagueness

MARK CRIMMINS

References and further reading

Devitt, M. and Sterelny, K. (1987) *Language and Reality*, Oxford: Blackwell.(A brief and readable introduction to some central issues, with useful references and bibliography.)

Martinich, A. (1990) *The Philosophy of Language*, Oxford: Oxford University Press, 2nd edn.(A collection of many key writings - often quite difficult - in philosophy of language.)

Language, Renaissance philosophy of

Renaissance philosophy of language is in its essentials a continuation of medieval philosophy of language as it developed in the fourteenth century. However, there were three big changes in the fifteenth and sixteenth centuries. First, humanism led to a much greater interest in the practical study of languages, including Greek, Hebrew and vernacular languages, as well as classical Latin. Literary analysis and eloquent discourse were emphasized. Second, there was a loss of interest in such medieval developments as supposition theory, which meant that there was little discussion in logic texts of how words relate to each other in propositional contexts, and how sense and reference are affected by the presence of such logical terms as 'all', 'none', 'only', 'except' and so on. Only in early sixteenth-century Paris were these issues pursued with any enthusiasm. Third, the fourteenth-century insistence that both words and concepts were signs had several effects. There was a new interest in the classification of different sorts of signs, both linguistic and non-linguistic, particularly in the work of some early sixteenth-century Spaniards. Naturally significant mental *language was emphasized in a way that diverted the attention of logicians from spoken languages and their imperfections. Finally, concepts themselves came in for more attention, so that many of the topics discussed by logicians overlapped with what would now count as philosophy of mind, as well as with metaphysics. For instance, philosophers in the late scholastic tradition made much use of an early fourteenth-century distinction between the formal concept, which is a representative act of mind, and the so-called objective concept, which is whatever it is that is represented by a formal concept. The discussion of these issues by such writers as Pedro da Fonseca and Francisco Suárez has an obvious bearing on developments in early modern philosophy.*

1 Signs and signification

The central semantic claim of Renaissance thinkers was that the word or term is a sign, and the central semantic notion was that of signification. Frequent reference was made to Aristotle's remark (*De interpretatione* 16a3-4), 'Spoken words are signs of concepts' (see Aristotle §4); and there was widespread use of Pierre d'Ailly's definition, 'To signify is to represent some thing or some things or in some way (*aliqualiter*) to a cognitive power' (see Ailly, P. d'). Since categorematic terms such as 'cow' (unlike the categorematic terms 'concept' or 'mermaid') are intended to point to things in the world, a three-place relation of word, concept and thing was established. This account raises several problems: how syncategorematic terms (for example, 'all', 'none') can be signs; how non-referring categorematic terms are to be explained (see §3); and the precise nature of the word-concept-thing relation.

The problem of syncategorematic terms was much discussed by the nominalists of early sixteenth-century Paris (see Major, J. §2), and, following d'Ailly's definition, it was agreed that such terms signified in some way. That is, a syncategorematic term performs a function rather than pointing to an object, and this function is not itself signified by the term, though it can be in a separate locution. If I say 'No dogs are running', the act of negation is exercised; if I say 'That sentence is negative', the act of negation is signified.

The notion of signifying in some way was also extended to propositional signification by some authors, notably Fernando de Enzinas, who argued that a proposition (in the then-standard sense of an occurrent declarative sentence) does not signify things, whether ordinary things in the world or propositional complexes enjoying a special kind of existence, but performs a function. This discussion of propositional signification was linked with discussion of such issues as whether the proposition is the object of knowledge and judgment (assent and dissent) (see Toletus, F. §5), and whether it is the bearer of truth-values. These topics were still of interest to early seventeenth-century scholastics.

The precise nature of the word-concept-thing relation was the focus of the long-standing debate whether spoken words signify concepts or things. This question first became popular in the late thirteenth century, and it was still discussed by such late sixteenth- and early seventeenth-century authors as Franciscus Toletus, Sebastian de Couto (see Collegium Conimbricense) and John of St Thomas (§2). All the participants in the debate agreed on certain points. They agreed that concepts play an essential role in the significative process, for we cannot refer to objects we have no notion of. They also agreed that words are typically used to pick out things in the world rather than our own concepts. The debate concerned the way in which the role of concepts in the significative process was to be

described. Some authors, following such disparate figures as Aquinas and John Buridan, held that words primarily signified concepts and only secondarily signified things. Others, following William of Ockham, held that words signified things alone in virtue of being subordinated to concepts. Sebastian de Couto preferred to say that while the signification of things had a certain primacy, both concepts and things were made known by words, which thus enjoy a double signification. John of St Thomas held that words signify concepts more immediately and things more principally.

What has been said so far applies equally well to medieval and Renaissance treatments of the word as sign. There was, however, an important development in the fourteenth century which helps differentiate the later period from the earlier, and which stems from Ockham's insistence that the concept itself must be regarded as a sign. This notion (while foreign to Augustine) was not new, but Ockham made it central. In doing so, he made mental language (see §2) rather than spoken language the paradigm of signification.

Once the concept had come to be regarded as a sign, we find in some early sixteenth-century authors, especially Domingo de Soto (§2), a careful classification of signs, both linguistic and non-linguistic. In relation to the speaker, spoken words were said to be instrumental signs, because they were used as instruments of communication, and mental terms were said to be formal signs, because they represented by their very nature. In relation to the things signified, spoken words were said to be conventional signs, and mental terms were said to be natural signs, since their signification did not depend on choice or convention. In the various editions of his *Summulae*, Soto asked whether one should add a category of customary signs, such as napkins on the table as a sign of lunch. In the end he concluded that these signs were natural, albeit founded on a convention. Soto's classifications were elaborated by many later authors, including Sebastian de Couto and John of St Thomas; though at least one late sixteenth-century Spaniard, Domingo Báñez, did explicitly reject the view that the concept was a formal sign.

2 Mental, spoken, natural and conventional languages

The doctrine that the concept is a formal sign went hand-in-hand with the notion of mental language, a language of thought that is naturally significant and common to all human beings. Such a notion is present at least from Augustine on, but it was fully developed in the fourteenth century, first by William of Ockham and then by Pierre d'Ailly (see William of Ockham; Ailly, P. d'). Mental propositions were thought of as having syntactic structure and mental terms were thought of as having supposition, so that the notion of a language system was internalized. As a corollary, the place of grammar in the study of logic and philosophy of language was devalued. Some of the problems raised by the notion of mental language, such as the difficulty of identifying the mental correlates of demonstratives and pronouns, or of impersonal verbs such as *pluit* ('It is raining'), were still the object of lively debate in early sixteenth-century Paris. Later, the discussion of the structure of mental language disappeared, but the notion of an inner language remained. For instance, in his *Scholae (Lectures)* Petrus Ramus (§2) argued that logic or dialectic deals directly with thought, with *ratio* (reason) rather than *oratio* (discourse).

Whether mental language was seen as some kind of ideal language is unclear. Mental language certainly provided some kind of universality in that all spoken language was subordinated to it. However, it did not necessarily provide a common syntactic structure, for logically equivalent spoken propositions could turn out to be subordinated to different mental propositions. Nor was mental language necessarily thought of as being ideal in the sense of reproducing the structure of the world. On the other hand, it was supposedly ideal in the sense of being a language which was free of ambiguous terms and which contained only as much structure as was required for the formulation of judgments.

The discussion of spoken language was often related to two key episodes in the Bible: Adam's naming of the animals, and God's inflicting different languages on human beings as a punishment for building the Tower of Babel. One issue had to do with the nature of the original language and its relation to later languages. If the first language was Hebrew, as Isidore of Seville had held, in what form did it survive? If it was not Hebrew, did it completely disappear, or did it contribute at least some structure to the post-Babel tongues? A second issue had to do with the original institution of language: did Adam form it, or did God give him his language (as Sebastian de Couto, among others, held)? This in turn raised a third issue. Was the language instituted by Adam, or by God through Adam, a natural language in the sense of one that enabled users to grasp essences by virtue of a natural relationship between spoken words and the things named?

While this issue became more important in the seventeenth century with the work of such authors as Jacob Boehme and Francis Mercurius van Helmont, there was considerable discussion in the sixteenth century. The interest was due not only to biblical studies but also to the rediscovery of Plato's *Cratylus* and other classical sources, as well as to the strong Renaissance interest in magic and the Kabbalah, with the concomitant hope that a knowledge of natural language would enable one to exercise some control over the objects signified. The consensus, particularly among logicians, was that Aristotle was right, and spoken language is indeed conventional, whether it is directly God-given or not. On the other hand, the belief that language was conventional was seen as compatible with the belief that the institution of language is guided by reason (as Pseudo-Kilwardby had argued in the thirteenth century). To say that spoken words have signification *ad placitum* (by convention or agreement) does not mean that their signification is random and unmotivated. This point was emphasized by the Spanish grammarian Sanctius (Francisco Sánchez de la Brozas) in his *Minerva*, published in 1587. He felt that Aristotle, Plato and the Bible could be reconciled; and he also argued that the rational origin of language provided an underlying system of regularities in conventional languages.

3 Intentional contexts and beings of reason

Given the focus on spoken and mental terms as signs, one must ask what is referred to in intentional and modal contexts, or by fictional terms such as 'chimera'. Authors who worked within the medieval logical tradition, particularly the nominalists of early sixteenth-century Paris, attempted to treat both problems within the framework of supposition theory. There are elaborate discussions of the reference of 'horse' in such contexts as 'I promise you a horse', 'A horse is necessary for riding', 'A horse is imaginarily a chimera'. In particular, people asked whether it was legitimate to postulate reference to imaginary and impossible objects, or whether the reference of terms in special contexts could be explained solely in terms of reference to ordinary past, present, future and possible objects. In the later sixteenth century, this type of discussion disappeared from most logic texts, though there are still references in Pedro da Fonseca.

What did survive, at least in the scholastic tradition, was the discussion of beings of reason (*entia rationis*). This investigation, which can be traced back to the thirteenth century, did not focus on the problem of reference in special propositional contexts, but on the reference of terms in general. There were said to be two types of being, real beings which (with the exception of God) fall under Aristotle's ten categories, and beings of reason (*entia rationis*), which include negations, privations and relations of reason. The latter included the logical second intentions (that is, higher-order concepts used to organize first-order concepts) such as 'genus' and 'species'. Beings of reason are non-categorial and exist only because of the work of the human mind, though they may have some foundation in reality. 'Nonbeing', for instance, is a term which is significant as referring to a pure negation; and 'blindness' is a term which is significant as referring to a privation, the absence of a capacity that animals normally have; yet neither pure negations nor privations can be literally parts of the world. They must be conceptual entities. Genus and species terms (as opposed to the terms 'genus' and 'species' themselves) were often included among the terms whose significance involves relations of reason. As Domingo de Soto (§2) indicated, the point here is that when we say 'A human being is an animal' our general terms do not refer to special universal things, humanity and animality, but they imply a well-founded relation between particular individuals in the world and the concepts we use to classify them into groups. There was general agreement that chimeras should be classified as negations, and some people held that they were the most proper examples of beings of reason since, as impossible, they must be completely mind-dependent (see John of St Thomas §§2, 4; Suárez, F. §2).

4 Analogy

Analogy as a theory about a certain sort of linguistic usage has thirteenth-century roots, and involves three problems. First, there is the problem of equivocal terms as Aristotle introduces it at the beginning of his *Categories*. Equivocal terms (for example, 'bank') are those which can be used in two quite different senses, and it seems natural to extend the notion of an equivocal term to cover terms that are used in different but related senses. Second, there is the metaphysical problem of how to discuss being (*ens*) in general when the being of a substance is so different from the being of an accident. Third, there is the problem of religious language. How can words normally used of humans, such as 'just' or 'good', be meaningfully used of God, when God is so different from human beings? In answer to these problems, logicians and theologians developed a theory which divided words into three sorts, independently of context. Some were univocal (always used with the same sense), some were

purely equivocal (used with totally different senses), and some were analogical (used with related senses).

The term ‘analogy’ itself had two senses. In the original (Greek) sense, it involved a comparison of two proportions or relations. Thus ‘principle’ was said to be an analogical term when said of a point and of a spring of water because a point is related to a line as a spring is related to a river. This type of analogy came to be called the analogy of proportion, proportionality, or (by Cajetan) proper proportionality. In the second sense, analogy involved a relation between two things (or one pair of things and a third), of which one is secondary and the other primary. Thus ‘healthy’ was said to be an analogical term when said of a dog and its food because while the dog has health directly, its food is healthy only as contributing to or causing the health of the dog. This second type of analogy became known as the analogy of attribution.

One of the main subjects of debate was how to classify types of analogy, and how to apply the various types to the different metaphysical and theological cases mentioned above. Although in *De veritate (On Truth)* Aquinas said that religious language must be interpreted by means of the analogy of proportionality, in other writings he appealed to the analogy of attribution. During the fourteenth and fifteenth centuries most logicians and theologians, including Capreolus, appealed to the analogy of attribution (if they discussed the topic at all). At the end of the fifteenth century, however, Thomas de Vio, Cardinal Cajetan argued that the analogy of proportionality was the only true analogy (see Cajetan §2). Cajetan’s view was not always accepted, but there was much discussion of the issue by the later scholastics, including Soto, Fonseca (§3) and Suárez (§2) (see Silvestri, F.).

The other main subject of debate is very closely related to philosophy of mind, and springs mainly from the work of early fourteenth-century philosophers, particularly Duns Scotus. Latin translations of Aristotle maintained that the difference between a univocal term and an equivocal term was that the latter was subordinated to more than one *ratio*. The *ratio* soon came to be identified with a concept, and the question then was how many concepts are involved when an analogical term is used. There were three views among logicians and theologians. Analogical terms could be seen as straightforwardly equivocal terms subordinated to two distinct concepts; they could be seen as subordinated to an ordered cluster of concepts (possibly but not necessarily described as a disjunction of concepts), or they could be subordinated to a single concept which represents in both a prior and a posterior manner (*per prius et posterius*). Scotus and his followers rejected all these possibilities, and argued that the word *ens* was univocal, but a lively discussion continued (especially among Thomists) until the time of Suárez.

Many of these thinkers made a distinction between formal and objective concepts. The formal concept was the act of mind or conception that represented an object, and the objective concept was the object represented. If the spoken word ‘being’ corresponds to just one formal concept (a point on which there were some differences of opinion), the focus of discussion shifts to the status of the objective concept. Is it the actual thing in the world which is thought about; is it a common nature or some other kind of intermediary entity which is distinct from the external object without being mind-dependent; or is it a special kind of mind-dependent object which has only objective being, the being of ‘being thought’ (*esse objective*, *esse cognitum*)? It is at this point that philosophy of mind and ontology take over from philosophy of language.

5 Humanism

Humanists, including Valla, Erasmus and Vives, were more concerned with the description of language as a literary phenomenon and with the cultivation of eloquence than with philosophy of language proper. The late fifteenth and sixteenth centuries saw an explosion of practical language studies. New humanist grammars appeared which prepared students for literary analysis rather than philosophy and theology; Greek and Hebrew studies expanded; there was an increasing use of and interest in vernacular languages; biblical studies and translations flourished; the widespread use of printing focused attention on problems of spelling and grammatical norms; and there was new interest in forms of writing, including Hebrew characters and Egyptian hieroglyphs. This last interest was often combined with a belief in magic. Agrippa, for instance, discussed how one might manipulate the numerical values of the letters of the Hebrew alphabet in order to make contact with the spirit world.

See also: Humanism, Renaissance; Language, medieval theories of; Logic, medieval; Logic, Renaissance; Universal language

E.J. ASHWORTH

References and further reading

Ashworth, E.J. (1985) *Studies in Post-Medieval Semantics*, London: Variorum.(A collection of articles including studies of reference in intentional contexts, propositions and mental language.)

Ashworth, E.J. (1988) 'Traditional Logic', in C.B. Schmitt, Q. Skinner and E. Kessler (eds) *The Cambridge History of Renaissance Philosophy*, Cambridge: Cambridge University Press, esp. 153-62.(A useful general account with plenty of references.)

Ashworth, E.J. (1990) 'Domingo de Soto (1494-1560) and the Doctrine of Signs', in G.L. Bursill-Hall, S. Ebbesen and K. Koerner (eds) *De Ortu Grammaticae: Studies in Medieval Grammar and Linguistic Theory in Memory of Jan Pinborg*, Amsterdam and Philadelphia, PA: John Benjamins, 35-48.(Discussion of how signs were classified.)

Ashworth, E.J. (1995) 'Suárez on the Analogy of Being: Some Historical Background', *Vivarium* 33: 50-75.(How analogical terms were classified; formal and objective concepts; bibliography.)

Coudert, A. (1978) 'Some Theories of a Natural Language from the Renaissance to the Seventeenth Century', in *Studia Leibnitiana, Sonderheft 7*, Wiesbaden: Franz Steiner, 56-114.(A clear overview.)

Demonet, M.-L. (1992) *Les Voix du signe: Nature et origine du langage à la Renaissance (1480-1580) (Voices of the Sign: the Nature and Origins of Language in the Renaissance)*, Paris and Geneva: Champion-Slatkine.(A full, fascinating account of all aspects of language based on both philosophical and literary sources.)

Jardine, L. (1988) 'Humanistic Logic', in C.B. Schmitt, Q. Skinner and E. Kessler (eds) *The Cambridge History of Renaissance Philosophy*, Cambridge: Cambridge University Press, 173-98.(A useful general account, though controversial on Lorenzo Valla; shows how humanist logic was used in textual exegesis.)

Nuchelmans, G. (1980) *Late-Scholastic and Humanist Theories of the Proposition*, Amsterdam, Oxford and New York: North Holland.(Discusses mental language, the significate of the proposition, and the terminological innovations of Melanchthon and Ramus.)

Nuchelmans, G. (1983) *Judgment and Proposition from Descartes to Kant*, Amsterdam, Oxford and New York: North Holland.(First chapter discusses Renaissance Thomism, including material on formal and objective concepts.)

Padley, G.A. (1976) *Grammatical Theory in Western Europe, 1500-1700. The Latin Tradition*, Cambridge: Cambridge University Press.(Generally good, but to be used with caution where logic is concerned.)

Percival, W.K. (1982) 'Changes in the approach to language', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 808-17. (Discussion of humanist grammar teaching.)

Roncaglia, G. (1995) 'Smiglecius on *entia rationis*', *Vivarium* 33: 27-49.(Useful introduction to the notion of beings of reason, with good bibliography.)

Language, social nature of

Language is mostly used in a social setting. We use it to communicate with others. We depend on others when learning language, and we constantly borrow one another's uses of expression. Language helps us perform various social functions, and many of its uses have become institutionalized. But none of these reflections settle the question of whether language is an essentially social phenomenon. To address this we must consider the nature of language itself, and then ask which social elements, if any, make an essential contribution to its nature.

While many would accept that language is an activity that must take place in a social setting, others have gone further by arguing that language is a social practice. This view commits one to the claim that the meanings of an individual's words are the meanings they have in the common language. The former view need not accept so strong a claim: meaning depends on social interaction because it is a matter of what one can communicate to others but this does not require the existence of communal languages. A competing conception which rejects the social character of language in either of these versions is the thesis that language is mentally represented in the mind of an individual.

1 Linguistic Platonism

Languages can be used in many ways and for many purposes, but a dispute about the nature of language concerns language in its most fundamental sense. Is it fundamentally social or do individual speakers have their own languages?

One view of language, which will not be discussed in any detail in this entry, is linguistic Platonism. This is the view that languages are formally characterized abstract entities existing independently of all speakers. Platonists then seek to define a relation between these abstract objects and speakers to determine which is the actual language of an individual or a population. The actual language relation may be defined either in terms of the conventional practices adopted by a population, or the psychological make-up or linguistic intuitions of the individual language-user. Thus Platonism does not settle the issue of whether languages are properties of individuals or social groups.

2 Communal languages and idiolects

A broad division exists between those philosophers and linguists who think the fundamental conception of language is that of a shared public language, and those who consider the fundamental notion to be that of an individual's language or idiolect. Within each of these schools there is room for disagreement about the precise nature of languages and idiolects. These differences will be reviewed below.

Michael Dummett (1989), following the ideas of Ludwig Wittgenstein (1953), offers a number of arguments for the communal conception of language. One of these goes via his rejection of Platonism. For Dummett, the objectivity of meaning requires the meaning of a word to be independent of any given speaker's judgment of it. But if the meaning of words were independent of all speakers, as the Platonist suggests, this would open up the sceptical possibility that all speakers could be in ignorance about the actual meaning of words in their language. Rejecting this possibility, Dummett insists that what an expression means among a community of language-users must be determined by the publicly observable conventions governing its use.

Dummett also argues that the publicity of meaning is required to ensure successful communication. Because speakers hold themselves responsible to the conventional meanings of words in taking themselves to be speaking a given language, what they say when uttering certain words is what those words mean in the communal language to which they belong. It is only because there are such conventional practices that meaning is shared and communicable; and for Dummett, were it not for such standards of use, speakers could not know they attached the same significance to their words. Meanings would be private and incommunicable, hidden in the minds of each individual speaker. But if this were the case, we could never learn a language, or know if someone else understood it. But we can work out what people mean by the observable use they make of their expressions; and to learn or understand a language is simply to participate in these shared linguistic practices. However, no speaker will be fully competent in the conventions governing the use of words and rules for grammatically compounding them; at most speakers will have a partial mastery of the public language. Nevertheless, for Dummett communication

depends on the standards of the communal practice.

Similar considerations have been advanced by John McDowell (1988) who regards the possibility of thought and of knowing the minds of others as depending on our speaking a communal language. We arrive at our thoughts by finding the words to express them, and so in virtue of the language in which we do this we make our minds immediately available to one another. The words you use to express a thought will express the same thought for me and so, just in virtue of hearing you speak sincerely, I can know what you think. This is possible because the meanings of a speaker's words lie open to view on the surface of his practice. But they are recognizable only to those who share those practices. These ideas hail from the later work of Wittgenstein, who saw language as a social practice governed by rules of use (see Meaning and rule-following). Along with Wittgenstein, Dummett and McDowell assume that in the absence of shared languages, or observable conventions of use, our knowledge of others' meanings would be a matter of psychological speculation about what goes on in a private realm concealed behind behaviour. And, given the doubtful coherence of a private language, McDowell thinks the absence of shared languages would de-stabilize communication, meaning and thought.

Not every defender of communal languages accepts the publicity of meaning. It is possible to argue that private states are involved in the meanings of public language items so long as speakers have sufficiently similar responses to the situations in which those words are used. On this view, it would be enough for the purposes of communication that speakers attached the same significance to terms in the language: they would not also have to know that they did. Success in communication would be a contingent matter and could never be guaranteed (see Private states and language).

The strongest opposition to the communal view of language comes from those advocating the idiolectic conception of language, although there is deep disagreement amongst the protagonists over the extent to which an idiolect is social or psychological in nature. However, all proponents of the idiolectic view resist the assumption made by Dummett and McDowell that if a language is not shared then the meaning of an individual's words would have to be subjective and private. It does not follow from the fact that meanings are publicly discoverable that they are shared. I can work out the meaning someone attaches to a word without attaching that meaning to it myself. Donald Davidson (1986) stresses malapropisms as providing cases of this kind (see §4). The denial of shared language does not entail the privacy of meaning and the publicity of meaning does not require a communal language. The arguments of McDowell and Dummett designed to show that the language of an individual would be necessarily private remain inconclusive. Facts about my idiolect may be knowable by those I interact with, and linguists studying me scientifically may know more about my idiolect than I do.

3 Language as a psychological or a social phenomenon

Noam Chomsky (1986) is a vigorous opponent of the notion of shared public languages. He argues that there is no empirically respectable way to define such a notion. According to Chomsky there is no such thing, for example, as English, or Dutch, or Chinese. In each case there is no single set of grammatical rules, shared vocabulary items, or rules for pronunciation which characterize the language in question. Whether we include every idiosyncratic use of every word and rule of grammar by a particular population of speakers, or select only the intersection of those uses, what is specified is a language nobody knows or speaks. The same problems beset attempts to delimit a dialect. For this reason, Chomsky abandons the everyday use of 'language', which he calls E-language: this being an ill-assorted set of overlapping practices, patterns of deference, power relations among speakers, and so on. He replaces it with the notion of a speaker's I-language: an internal component of the individual's mind/brain which assigns meaning and structure to the sounds and signs the speaker encounters. The I-language is a cognitive psychological entity - a state of the language faculty - which together with a lexicon of word-like items determines the extent of the speaker's idiolect.

According to Chomsky, language is essentially defined by its meanings and structures, but these exist only as mental representations in the mind of the speaker - as part of the speaker's knowledge of language - and have no reality in the external or social world. Linguistically, the world contains only sounds and marks: it is creatures equipped with a language faculty who assign them meaning and structure. At first, this is achieved without reference to others. Our earliest interactions will of course influence our acquisition of vocabulary and the particular grammar we attain although, in the case of the latter, Chomsky regards this as environmental triggering: a prompt from one's surroundings to set the parameters of a system that is already in place. Thus the development

of one rather than another I-language is the result of interaction between the linguistic environment and the innate linguistic system (see Language, innateness of).

After we attain mature linguistic competence, we often defer to others in positions of authority in their use of words and assessments of grammar. But Chomsky points out that this is social behaviour undertaken after we have acquired a language, and so it does not constitute an essentially social contribution to it. On the notion of language Chomsky favours, the study of language becomes a branch of individual cognitive psychology. The study of group behaviour has no impact on it. Communication is not, on this picture, essential to language. Language is primarily a means of expressing thought. Speakers and hearers do have the means to try to make sense of one another's talk and the closer their I-languages and vocabularies are to one another, the more chance they will have of success. However, for Chomsky, communication is risky and success is never guaranteed so it cannot provide the basis for linguistic competence.

Where he is willing to acknowledge a pervasive social aspect of language is in the case of the division of linguistic labour: a phenomenon Hilary Putnam (1975) drew attention to. This occurs where there are words having an everyday sense and a more technical meaning without being ambiguous. The words 'gold' and 'carburettor' provide examples for those with little knowledge of chemistry or cars. Speakers can use these words without knowing their precise meanings because they defer to experts in their community who know how to fix the meaning and reference of those terms precisely (see Putnam, H. §3). Chomsky points out that the entries for such words in a speaker's mental lexicon will specify the meaning of the terms as far as the speaker's (partial) knowledge carries him, but they would also include an indication that further details are to be filled out by others. The meanings specified in this way do not go beyond what can be studied in an individual's knowledge of language. It is not to concede that language includes any essentially social elements. The phenomenon Putnam alludes to, while extensive, is not fundamental to the use of language. People could cease to use terms they did not fully understand without ceasing to speak or understand a language. What this shows is that the division of linguistic labour provides the weakest claim for the social character of language. Other reasons are needed to conclude that language is necessarily social in character.

4 Idiolects as social phenomena

Arguments for this are provided by Donald Davidson and Tyler Burge, although they differ in their conceptions of what makes an idiolect social.

Davidson rejects the idea of language as 'a clearly defined shared structure' which language-users learn and deploy in communicative exchanges. The appeal to conventions is neither necessary nor sufficient for linguistic communication, since speakers and hearers need not attach the same meanings to their words nor stick to the conventions they have established in order to make sense of one another's remarks. All that is required is that the hearer know what the speaker means; it is not additionally required that the hearer attach the same significance to their own use of those words. The cases of malapropism and slips of the tongue serve to make the point. I can know that someone uses 'debated' as I use 'abated' when they say 'The rain has debated'. But there is no reason for me to adopt their use of that word. So long as I can work out, by means of interpretation, what significance the speaker intends their words to have, I can understand them without having to share their language. Communication does not require speaker and hearer to belong to the same linguistic community, or to share a set of linguistic practices; it requires them to share an ability to interpret one another's utterances as part of a larger project of making sense of one another's behaviour in rational terms through the attribution of beliefs, desires and intentions. Part of this project will consist in assigning meanings to people's utterances that make sense of what they say against the background of the mental states we take them to be in. Clues of all kinds can figure in the evidence an interpreter draws upon to work out the meaning of someone's remarks, and no single set of shared conventions will be necessary or sufficient for this task.

While it is unlikely that any two speakers will ever speak the same language - minor differences in their vocabularies or grammar will always distinguish their idiolects - what they must share, according to Davidson, is the method of interpreting one another's speech. It is this condition which ensures the notion of idiolect has a social dimension since speakers can only mean what others can recognize or interpret them as meaning. The meanings of the words and sentences of their idiolects are fixed by others' publicly determined interpretations. For Davidson, the possibility of meaning anything depends on successful communication with another: having one's

words interpreted as one intends them to be interpreted. In this way, meaningful speech, although it demands nothing more than a variety of idiolects, is necessarily social (see Radical translation and radical interpretation §§7-10).

Burge (1989), in contrast, argues that social factors beyond the scope of local communicative interactions play an essential role in the meaning of words in the individual’s idiolect. He takes language to be a partly social phenomenon not just because it is learned and used in a social setting but because the meanings of a person’s words depend on the way others in their linguistic community use them. He agrees with Chomsky that the study of language is a project for individual psychology, but he argues the semantic facts about an individual’s idiolect depend essentially on their relations to other language-users. To establish that meaning must be individuated nonindividualistically and that idiolects are to some extent social, Burge (1982) constructed a famous thought-experiment in which there is an individual who uses the word ‘arthritis’ to express a number of thoughts about pain in their joints. One day they declare ‘I have arthritis in my thigh’. Since the community the individual belongs to uses the word to apply only to inflammation of the joints, what the individual says is false. Burge then asks us to consider a counterfactual situation in which the community the subject belongs to uses ‘arthritis’ to apply both to inflammations of the joints and to other rheumatoid ailments. In this counterfactual situation there is no change in the individual’s physical history or nonintentionally characterized experiences, and yet the statement they make there is true not false. Burge invites us to share the intuition that the truth-value of what is said differs in the two situations because something different is meant in each context. He concludes that meaning of the word ‘arthritis’ on the speaker’s lips is different in each of these situations because of the different linguistic communities to which they belong. By this argument Burge hopes to show that the meanings of someone’s words are individuated not just by facts about the individual but also partly by facts about the uses of words in the wider linguistic community (see Content: wide and narrow). In this way, the meanings of many words in a person’s language are not just up to him but depend essentially on the linguistic practices of others around him. But what is it for a speaker to belong to a particular linguistic community? For Burge it is a matter of his deferring to members of that community to explicate the meanings and determine the reference of his words on what are cognitive and not just pragmatic grounds. If Burge is right, facts about the meaning of our words supervene not only on facts about our use of words but also on facts about other people’s usage. This amounts to a strong claim that social factors are constitutive of meaning and hence of language and that, although language is a matter of individual psychology, it is also partly a social phenomenon.

See also: Language, conventionality of; Methodological individualism; Language and gender; Meaning and communication

BARRY C. SMITH

References and further reading

Burge, T. (1982) ‘Individualism and the Mental’, *Midwest Studies* 4: 73-121.(The source of the famous thought experiment discussed in §4, containing objections to it and responses.)

Burge, T. (1989) ’Wherein is Language Social’, in A. George (ed.) *Reflections on Chomsky*, Oxford: Blackwell, 1989.(A demanding paper which contains his exposition of the partly social, partly psychological conception Burge has of idiolects.)

Chomsky, N. (1986) *Knowledge of Language*, New York: Praeger.(Chapters 1 and 2 contain a clear statement of Chomsky’s conception of I-language discussed in §3.)

Davidson, D. (1986) ’A Nice Derangement of Epitaphs’, in E. Lepore (ed.) *Truth and Interpretation*, Oxford: Blackwell, 1986.(An important paper challenging the need for conventions and rules of language, but defending the social nature of communicative exchanges.)

Dummett, M.A.E. (1989) ’Language and Communication’, in A.George (ed.) *Reflections on Chomsky*, Oxford: Blackwell, 1989.(Careful discussion of many of the issues raised in §§2-3, including critical discussions of Davidson and Chomsky.)

McDowell, J. (1988) ’In Defence of Modesty’, in B. Taylor (ed.) *Michael Dummett: Contributions to Philosophy*, Dordrecht: Martinus Nijhoff, 1988.(A subtle and penetrating account of McDowell’s views on language and mind.)

Putnam, H. (1975) ’The Meaning of "Meaning"’, in *Collected Papers*, vol. 2, Cambridge: Cambridge University Press.(A seminal paper containing the notion of the division of linguistic labour among many other influential ideas.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(A difficult read but taken to be the origin, and the most powerful sustained defence of, the philosophical conception of language as a social practice.)

Lassalle, Ferdinand (1825-64)

Ferdinand Lassalle was one of the principal founders of German social democracy and a strong advocate of state socialism. He was associated with Marx, although he was not himself a Marxist but rather a radical philosophical idealist in the Hegelian tradition.

Ferdinand Lassalle was born in Breslau, the son of a wealthy Jewish merchant. He was involved in the abortive revolution of 1848 and was imprisoned for six months. A restlessly active man, he devoted much of his life to political agitation and to prosecuting the case of his friend, the Countess Sophie Hatzfeld, against her estranged husband. In 1863 he founded the General German Workingmen's Association, a political party unmistakably socialist in orientation. In 1864 he was killed in a duel.

Lassalle was acquainted with Marx and Engels and was influenced by their work, although not to the exclusion of other thinkers. Indeed, the doctrine he finally advocated was a collection of points drawn from a variety of sources and welded together to produce what many regarded as a strikingly original result. The centrepiece of his economic theory was 'the iron law of wages': market forces would inexorably depress wages to the 'physiological minimum' necessary to keep the workers alive. The solution to this unsatisfactory situation was, he argued, to establish producers' cooperatives through which the workers, by becoming their own employers, could remove themselves from the labour market, thus nullifying the effect of the 'iron law'. For this, and for his other economic doctrines, Lassalle was indebted chiefly to Malthus and Ricardo. His political views, however, he drew from the German philosophical idealists. From Fichte he took the idea of a German nation, now divided, but destined to become a single state with a unified territory; and from Hegel he took his view of the state as the necessary instrument for 'the elevation and development of the human race into freedom'. It followed from this that the real enemy of the industrial proletariat was not the state but the liberal bourgeoisie, together with all those who, by endorsing the capitalist market economy, had ranged themselves against the proletariat as 'one reactionary mass'. The party of the proletariat should, therefore, form a tactical alliance with the Prussian state. The basis of the alliance would be that the party would support the state against liberal opposition and help promote a 'Prussian' solution to the problem of German unification, in return for which the state would implement democratic reforms, in particular universal franchise, and provide financial support for the establishment of producers' cooperatives. In short, the ultimate objective would be a united German state with a democratic constitution and a socialist economy. It was an ingenious strategy, much of which commended itself to Bismarck, and most of which infuriated Marx.

Lassalle's earliest significant work was a study of Heraclitus (1858). It was basically a Hegelian interpretation and caused something of a stir at the time. However, it is no longer referred to by classical scholars. His other major works include his *Das System der erworbenen Rechte (System of Acquired Rights)* (1861), which features an attempt to show that, in certain cases, retrospective legislation is compatible with the idea of right as such, and his *Bastiat-Schulze* (1864), a polemic against the liberal economist Franz Hermann Schulze-Delitzsch.

See also: Socialism

H. TUDOR

List of works

Lassalle, F. (1858) *Die Philosophie Herakleitos des Dunklen von Ephesos (The Philosophy of Heraclitus, the Dark One from Ephesus)*, Berlin.(Lassalle's most ambitious philosophical and scholarly work; although a sensation at its time, it failed to make a lasting impression.)

Lassalle, F. (1861) *Das System der erworbenen Rechte (System of Acquired Rights)*, Leipzig: J.A. Brockhaus. (An essay in jurisprudence with particular reference to rights, the most notable feature of which is Lassalle's attempt to show the legitimacy of retroactive legislation.)

Lassalle, F. (1864) *Herr Bastiat-Schulze von Delitzsch*, Berlin: R. Schlingmann.(Lassalle's most celebrated work on social and economic policy.)

Lassalle, F. (1962) *Ausgewählte Texte (Selected Texts)*, trans. and ed. T. Ramm, Stuttgart: Koehler.(A useful selection from Lassalle's principal writings.)

References and further reading

Bernstein, E. (1893) *Ferdinand Lassalle as a Social Reformer*, New York: Greenwood Press, 1969.(The most important study of Lassalle available in English.)

Footman, D. (1946) *The Primrose Path: A Life of Ferdinand Lassalle*, London: Cresset Press.(Includes extensive quotations, in English, from Lassalle’s writings.)

Knapp, V.J. <(1971) ‘Ferdinand Lassalle on the State and Society: A Legacy to Welfare Statism’, *The Australian Journal of Politics and History* XVII: 377-85.(A helpful study of the main point in Lassalle’s political programme.)

Ramm, T. (1956) *Ferdinand Lassalle als Rechts- und Sozialphilosoph (Ferdinand Lassalle as Legal and Social Philosopher)*, Meisenheim am Glan, 2nd edn.(A good and searching analysis by an eminent Lassalle scholar.)

Latin America, colonial thought in

Colonial refers to Spanish and Portuguese sovereignty in America from the arrival of Columbus in 1492 up to the emergence of modern Latin American states in the nineteenth century. The intellectual life of the colonies and their mother countries at that time falls into two phases: traditional and modern. The traditional phase includes the siglo de oro*, or the Golden Age of the sixteenth and seventeenth centuries. This was a time when literature and the arts flourished, along with Scholastic philosophy, jurisprudence and theology. During the eighteenth century, traditional thought gradually gave way to modern movements, particularly from France.*

The universities founded in the mid-sixteenth century, notably those of Mexico and Peru, as well as colleges and seminaries, were impressively productive in the area of philosophy. The pressure of events such as the clash between European and Native American cultures in the sixteenth century and the struggle for independence from Spain and Portugal in the nineteenth century brought about numerous nonacademic works with philosophic content. Authors wrote in both Latin and Spanish or Portuguese and often knew native languages, such as Nahuatl and Quechua as well. Many operated in several different areas, such as the nun, Sor Juana Inés de la Cruz, one of the greatest poets in the Spanish language, who wrote a book on logic in Latin, which has since been lost.

Students studied philosophy first, then specialized in medicine, law, or theology. The core philosophy curriculum was logic, natural philosophy or physics and metaphysics. In the eighteenth century Scholastic logic, similar to what has come to be known as formal logic, was weakened and natural philosophy began to incorporate experimental science. The bulk of philosophy was affected by modern thinkers such as René Descartes.

Eighteenth-century savants were critical of Scholasticism and later Latin American intellectuals tended to disavow the entire colonial past. However, historians since the 1940s have stressed the currency of modern scholarship, especially in science and since the 1960s have been rediscovering the sophisticated philosophy of the Golden Age.

1 Issues of the conquest

Contact between Spain and the Indies gave rise to a vast body of literature, including accounts of the conquest from both the Spanish and Indian viewpoints, studies of native culture, language and history and plans for new types of communities. Mexican bishop Vasco de Quiroga, for example, used ideas from Thomas More's *Utopia* (1516) to design settlements, or *congregaciones* for Indians dispersed or decimated by disease or mistreatment. These settlements became a model for later communities in Peru, known as *reducciones* and Brazil, known as *aldeias*.

An extraordinary body of writing on ethics, jurisprudence, politics and anthropology developed around the bitter controversies over relations between Spaniards and Indians. The most important figure was Dominican friar Bartolomé de Las Casas. Early in the settlement period a concession system, or *encomienda*, developed allowing Spaniards to force Indians to work on farms and in mines in return for fair treatment, which they usually did not receive. Las Casas, a Spanish landholder, or *encomendero*, in Cuba, heard a sermon denouncing this injustice which led him to free his Indians. He spent the next half of the century defending the rights of both native and African people. In carefully reasoned writings in Latin and Spanish he stressed that cultural differences do not suppress human rights. He questioned whether natural law supported the claims of the Crown, which caused him to advocate Spain's withdrawal from the Indies. Along with his fellow Dominican Francisco de Vitoria in Spain, Las Casas made an important contribution to international law.

2 Philosophy in the colonies

The bulk of colonial philosophy was done by members of religious orders. Each of these orders pursued its own traditions. Franciscans followed John Duns Scotus, while Dominicans, Augustinians and Jesuits developed their interpretations of Thomas Aquinas. In the eighteenth century, Jesuits and Oratorians were key figures in bringing modern philosophy and science to America. American Scholastics remained in close contact with their European colleagues, although seventeenth-century Peruvians complained that Europeans paid their works scant attention.

The most important philosophical sources in the colonies were textbooks published by professors and manuscript classnotes taken down by their students. Other materials included academic programmes which listed the *theses*

participants defended publicly. These are useful in tracing the progress of modern ideas. (A thesis on methodic doubt, for example, showed that Descartes was being discussed.) Theological works also contained philosophical material. A work on theology by Chilean Franciscan Alfonso Briceño included a series of excursuses on metaphysics.

Philosophy textbooks were known as commentaries or (from the seventeenth century) *cursus philosophici* (philosophy courses). Authors mostly used Aristotle's works as a framework to cover the assigned subject and to contribute to debates current at that time. In the eighteenth century, headings were added in modern works to accommodate new interests in methodology, science, mathematics and ethics.

Logic was taught in two parts. Formal, or the 'lesser' logic was called *summulae*, since it surveyed the *Summulae logicales (Summary of Logic)* of thirteenth-century logician Peter of Spain. The 'greater' logic followed the structure of Porphyry's *Isagōgē* on universals and the 'predicables', of Aristotle's *Categories* and *Posterior Analytics*, and often of his *Topics* and *Sophisms* (see Aristotle §§1, 7; Porphyry).

3 Philosophers

Augustinian friar Alonso de la Vera Cruz began teaching philosophy in Mexico in 1540 and published the first work (1554) on the subject in the western hemisphere, a two-volume logic book. He also wrote on natural philosophy and current issues and studied the Tarascan language. Dominican Tomás Mercado included in his logic his own translation of Greek texts of Aristotle (see Mexico, philosophy in).

After teaching in Mexico for sixteen years, Jesuit Antonio Rubio wrote what he hoped would become standard university textbooks in the three core areas of philosophy. His greater work of logic appeared in eighteen editions, between 1603 and 1641, seven of which were dubbed the *Logica mexicana* and became the official textbook in the Spanish University of Alcalá. Four of his works on natural philosophy went through over thirty-five editions. His metaphysics was never published (and is lost). It was, perhaps, eclipsed by Francisco Suárez's *Disputationes metaphysicae (Metaphysical Disputations)* (1597) (see Suárez, F.). Rubio's works appeared in over fifty editions in six European countries. He also found time to study the language of the Aztecs.

The logic of Franciscan Jerónimo de Valera, published in Peru in 1610, was the first philosophy book printed in South America. Peruvian Jesuits Alonso Peñafiel, Nicolás de Olea and José de Aguilar published full courses on logic, natural philosophy and metaphysics.

Juan de Espinosa Medrano (*c*.1633-88), known as El Lunarejo, was professor of philosophy and theology in Cuzco, Peru. In his day, he was active in literary and philosophical controversies. His defence of poet Luis de Góngora holds a place in Spanish literary criticism. He was also famous for his eloquence not only in Spanish and Latin, but Quechua, the language of the Incas, in which he composed a biblical play set in Peru. He wrote his (lesser and greater) logic to defend traditional positions against a philosophical rebellion in seventeenth-century Spain which reflected modern impatience with conventional thought. His logic contains an unusual defence of Platonic ideas (see Plato).

Although eighteenth-century philosophers were critical of previous Scholasticism, they usually blended it with modernism. Franciscan José Elías del Carmen Pereira in Argentina is an example of this eclecticism. The Jesuits had founded the University of Córdoba early in the seventeenth century and taught traditional Scholasticism using Rubio's logic as a textbook. By the mid-eighteenth century, conservatives complained that 'anti-Scholastic' doctrines were being taught. After the Jesuits were expelled from the region in 1767, the Franciscans, including Pereira continued their modern tendencies. The manuscripts of his course on natural philosophy dating from 1784 are extant.

Other representatives of the modern movement are José Agustín Caballero in Cuba, Isidoro de Celis in Peru, Jesuit Diego José Abad in Mexico and Oratorian Juan Benito Díaz de Gamarra y Dávalos. Well-known scientists, mathematicians and historians included Carlos de Sigüenza y Góngora in Mexico and Pedro Peralta Barnuevo in both Mexico and Peru.

4 Formal logic

Sixteenth-century Scholastic logicians at the time of Alonso de la Vera Cruz analysed sentences in several ways.

One way, still used in logic, would be to see a sentence like ‘Bossy is a cow’ as attributing the feature of being a cow, signified by the general term ‘cow’, to the thing denoted by the name ’Bossy’. However, they often took what logicians of the twentieth century would term a ’many-sorted’ approach, wherein they reduced sentences containing general terms to sentences having only terms for things. ’Bossy is a cow’, therefore, would be equivalent to a disjunction (a sentence made up of smaller ones joined by ‘or’): ‘Bossy is this cow *or* Bossy is that cow’. Deictic terms like ’this cow’ they called ’wander terms’. Alonso took the disjunction to be true when at least one of the sentences it comprised was true, so that ’Bossy is a cow’ is true when at least one cow is Bossy. On the other hand, ‘Bossy is not a dog’ would be analysed as a conjunction (a sentence made up of smaller ones joined by ‘and’): ‘Bossy is not this dog *and* Bossy is not that dog’. A conjunction is true when every sentence it is made up of is true. A universal sentence like ‘every cow is an animal’ is analysed as a conjunction of disjunctions: ‘this cow is this animal *or* this cow is that animal …*and* that cow is this animal *or* that cow is that animal …*and*’.

Alonso (1554) and his colleagues formulated rules to expand general sentences into disjunctions and/or conjunctions (’descent’) and from these to derive general sentences (’ascent’ or ’induction’). In fact, they defined quantification (words like ’every’, ’any’, ’some’) by means of these rules. They used such procedures to study many kinds of sentences: nonstandard quantification (’Fido is not every dog’), relations (’Peter and Paul are discussing’, ’every cow belongs to some man’), modality, the notions of possibility and necessity (’Fido need not wag his tail’) and many others.

Alonso’s definition of conjunctions and disjunctions is truth-functional: the truth of a compound sentence depends on the truth or falsehood of its parts. He takes the normal case of a conditional sentence like ’if Bossy is a cow, she is an animal’ as ’strict implication’, to use a twentieth-century expression, that is, it implies necessity: ’Fido cannot be a dog without being an animal’.

In general, Golden Age logic was quite similar to logic today and is often especially interesting where it differs (see Logical and mathematical terms, glossary of).

5 A theory of science

In his *Commentarii in universam Aristotelis dialecticam (Logica mexicana)* (1603), Rubio worked out a partially original theory of the relation of logic to science. Ideally, a science is an axiom system, a structured network of inferences resting ultimately on axioms. Science is general: it is not merely about *this* animal but about *animal* as such. Science is necessary in the sense that it seeks to describe physical law. A science aims at identifying the properties, or *passiones*, of its subject matter. In biology, for example, a statement would attribute the property, say, of *having DNA* to animals. Scientists must also be logicians, who use logic in two complementary ways. From a stock of terms, they construct sentences, link them together to form inferences and finally shape the inferences into a system. This is the practical or constructive (*compositorius*) side of logic. Its theoretical or analytic (*resolutorius*) function comes into play when they trace the paths of inference back to the axioms and break up statements into their constituent terms.

Logic itself is a science, hence an axiom system comprising terms, sentences, inferences, as well as a necessary structure. But to Rubio, it is unique because its subject matter is ultimately the same as that of all the sciences and is constructed without any subject matter. (Later logicians use variables to bracket out what is not of logical interest.)

Rubio defined the entities logic deals with in terms of relations and converse relations. The above construction and analysis, for example, are the two sides of a *making up* relation. In construction, terms make up statements, statements inferences and inferences are bound together into a system. Analysis is the converse relation of *being made up of*: the network is made up of inferences, these of statements and statements of terms. Another example of relations and converse relations is the predicate and subject of a sentence. A predicate is the relation of *being stated of* and a subject is the converse relation of *being what something is stated of*. These relations are the properties that logic as a science studies and seeks to identify in scientific discourse. The scientist says ’animals have DNA’, but the logician says ’*having DNA* is a predicate’, that is, ’*having DNA* is stated of something (animal)’. In other words, the scientist attributes the logical property of *being a predicate* to the ’thing’ *having DNA*. However, science and logic deal with different sorts of properties. *Having DNA* is a real property since it belongs to things in the world like Bossy and Fido. But a logical property like *being stated of* applies only to

’things’ in the mind, like *having DNA*. Logical properties are ’of second intention’ (’higher level’ would be a twentieth-century expression), since they apply to things as known and they are ‘mental constructs’ (*entia rationis*) in the sense that they do not really apply to them.

Rubio’s theory of logic can be formulated in the following way: logic studies and in some sense constructs higher-level relations and attributes them to entities studied by other sciences. In turn, these entities apply to concrete things.

6 A defence of Plato

Juan de Espinosa Medrano (1688) developed an original defence of Plato with regard to his treatment of the problem of universals. The issue, then as now, has to do with explaining the nature of the general features of things. Logicians say that Fido and Bossy have the universal feature of *having DNA*. Fido and Bossy are concrete things, but what is *having DNA* the universal quality that they share? Thomists (followers of Thomas Aquinas) and the Scotists (followers of John Duns Scotus), grouped together by Espinosa as *peripatetics*, usually rejected nominalist and psychologistic solutions that universals were simply the words or the thoughts of a person. Peripatetics held several varieties of realism, according to which universals, or essences, are somehow more than mere words or private thoughts. However, all peripatetics routinely rejected what they took to be Plato’s theory, which claimed that universals, or ideas, exist *apart* from the concrete, singular things which embody them.

Espinosa wanted to know what peripatetics meant by realism. He carefully isolated the elements found in both Thomist and Scotist theories and came up with a common peripatetic definition of essence. Both theoretical branches claimed, for example, that universals have essential being, that is, they exist as essences, that they do not change, are independent of time and space and, most importantly, that they differ both from concrete things and from God. Espinosa went on to analyse Plato’s position. First, he questioned the usual Neoplatonic and Aristotelian interpretations of ideas: that they are (an aspect of) God or they exist by themselves apart from things. Espinosa believed that, if we judge by Plato’s aim, which was to identify the object of knowledge, then he had only to describe his ideas as the peripatetics described their essences; in other words, what Plato said about his ideas was already *included in* what peripatetics said about their essences. Therefore, if peripatetic essences do not exist apart from things, neither do Platonic ideas. Espinosa did not infer from Plato’s words that ideas do or do not exist apart. What is clear is that ‘the peripatetics, willing or not, have platonized and wounded themselves as they attacked Plato’ (1688: 67) (see Neoplatonism). Espinosa stressed that his interesting claim ’has not been put forward until now, as far as I know, although (fifteenth-century Dominic of) Flanders and (contemporary Francisco) Araujo have favoured it in part’ (1688: 62). The essences he described are not unlike twentieth-century abstract entities. Although Espinosa was aware of their ambiguous ontological status, he thought the objections against them could be resolved.

Scholastic rebels of his time, especially a group of Spanish Jesuits, had nominalistic tendencies which Espinosa attacked fiercely. His arguments, he believed, finally put nominalism to rest, after outbreaks of the doctrine during the Greek period (Heraclitus (*c*.540-480 BC)), in the twelfth and fourteenth centuries (Roscelin (*c*.1050-after 1120) and Ockham (*c*.1287-1347)) and during the seventeenth century, in Espinosa’s own time (see Heraclitus; William of Ockham; Roscelin of Compiègne). Indeed, it was he who wrote nominalism’s epitaph:

> Thrice slain and then revived, I, Ockham’s sect, A fourth time now follow Euridice; Lo, here I helpless lie among the shades: A word, a concept, sans reality.
>
> (1688: 128)

7 Natural philosophy in the eighteenth century

In his course on natural philosophy or physics Elías del Carmen Pereira (1784) blended modern science and traditional philosophy. He took his ideas from Scholastics such as Anselm of Canterbury, Aquinas, Duns Scotus, Suárez, modern philosophers, such as, Francis Bacon, René Descartes, Pierre Gassendi, Baruch Spinoza and G.W. Leibniz, modernizing Scholastics, such as, Benito Jerónomo Feijoo, Teodoro Almeida and Tomás Vicente Tosca and scientists, such as, Robert Boyle and Torricelli. Pereira’s favourite topics were scientific: the vacuum, capillary action, motion, gravity and light. However, he continued to work within the traditional conception of science, as can be gathered from his definition of general physics as the natural body and its properties.

Pereira's use of the traditional categories of matter and form was an example of an old theory disguised as a new one. He claimed to be following the example of the more cultivated nations in his refusal to quibble over prime matter, defining matter as extended particles, which, 'whatever they are', are physically, but not mathematically indivisible, and are formed in various sizes and shapes. The form of a material thing became the manner in which material particles combined, with their relationship giving rise to the specific properties of the material thing.

Pereira framed his theory of light in traditional terms of substance and accident, matter and form. Light, he believed, was not an accident of matter as the Scholastics thought, but a substance and that the effluence theory of Newton and Gassendi was also wrong (see Newton, I.). Taking his lead from Nicolas Malebranche, Pereira claimed that the 'matter' of light is the medium wherein it is propagated, that is, the ether which consists of tiny contiguous elastic globules. Its 'form' is the rapid vibratory motion transmitted instantaneously from one particle to another in straight lines. In this particle-wave theory, the globules themselves do not move away from the light source and their vibration has frequency, but little or no amplitude.

Despite his Scholastic vocabulary, Franciscan Friar Elías del Carmen Pereira's world was no longer that of the Scholastics.

See also: Argentina, philosophy in; Brazil, philosophy in; Latin America, pre-Columbian and indigenous thought in

WALTER B. REDMOND

References and further reading

Aristotle (*c*.mid 4th century BC) *Categories*, trans. J.L. Ackrill, Oxford: Oxford University Press, 1963.(One of Aristotle's works on logic.)

Aristotle *Posterior Analytics*, ed. W.D. Ross, trans. J. Barnes, Oxford: Oxford University Press, 2nd edn, 1993. (Formulates an account of rigorous scientific knowledge.)

Espinosa Medrano, J. de (1688) *Philosophia Thomistica seu Cursus Philosophicus*, Rome.(His defence of Plato on pages 52-73 appeared in Spanish translation in W. Redmond, 'Juan de Espinosa Medrano: Sobre la naturaleza de los universales', *Humanidades* 3, Catholic University of Peru, 1969: 131-85.)

Gracia, J.J.E. (1993) 'Hispanic Philosophy: Its Beginning and Golden age', *Review of Metaphysics* 46: 475-502.

Hanke, L. (1965) *The Spanish Struggle for Justice in the Conquest of America*, Boston, MA: Little, Brown & Co. (A readable account of the controversies in which Bartolomé de Las Casas was involved.)

Irving, L. (1929) *Don Carlos de Sigüenza y Góngora, A Mexican Savant of the Seventeenth Century*, Berkeley, CA: University of California Press.(The most complete work on Sigüenza in English.)

Lanning, J.T. (1940) *Academic Culture in the Spanish Colonies*, Oxford: Oxford University Press.(A classic study of philosophy theses tracing the prompt arrival of modern ideas in Latin America.)

Navarro, B. (1948) *La Introducción de la filosofía moderna en México (The Introduction of Modern Philosophy in Mexico)*, Mexico: El Colegio de México.(A study of the influence of modern philosophy in eighteenth-century Mexico.)

Pereira, E. del Carmen (1784) *Physica generalis (General Physics)*, Facultad de Ciencias Jurídicas y Sociales, La Plata University; trans. in J. Chiabra (ed.) *La enseñanza de la filosofía en la Época Colonial (Philosophy Teaching in the Colonial Period)*, Buenos Aires: Coni Hermanos, 1911: 173-333.(This book blends Scholasticism with modernism.)

Porphyry (mid 3rd century AD) *Introduction Isagōgē*, ed. A. Busse, *Isagoge et in Aristotelis categorias commentarium*, CAG 4.1, Berlin: Reimer, 1887; trans. E. Warren, *Porphyry the Phoenician: Isagōgē*, Toronto, Ont: Pontifical Institute of Medieval Studies, 1975.(English translation.)

Redmond, W. (1972) *Bibliography of the Philosophy in the Iberian Colonies of America*, The Hague: Nijhoff. (The standard bibliography of source material and secondary literature.)

Redmond, W. (1974) 'Latin American Colonial Philosophy: The Logic of Espinosa Medrano', *The Americas* 30 (4): 192-212, Washington, DC: Academy of American Franciscan History.(The logic of Espinosa in its historical setting.)

Redmond, W. (1979) 'Formal Logic in New Spain: The Work of Fray Alonso', *International Philosophical Quarterly* 19 (3): 331-51.(A history of colonial logic and a nontechnical description of its content.)

Redmond, W. (1984) 'Lógica y ciencia en la *Logica mexicana* de Rubio' (Logic and Science in Rubio's *Mexican Logic*), *Quipu: Revista Latinoamericana de Historia de las Ciencias y la Tecnología (Quipu: Latin American Journal of the History of Science and Technology)* 1 (1): 55-82.(A full account of Rubio's concept of logic and

science and their interrelation.)
Redmond, W. (1991) 'El Lunarejo on Abstract Entities', *Concordia: Internationale Zeitschrift für Philosophie (Concordia: International Journal of Philosophy)* 20: 91-8, Aachen, Germany: Augustinus-Buchhandlung-Verlag.(Espinosa's defence of Plato.)
Redmond, W. (1993) 'Logik, Wissenschaft und Literatur' (Logic, Science and Literature), *Concordia: Internationale Zeitschrift für Philosophie (Concordia: International Journal of Philosophy)* 24: 24-33, Aachen, Germany: Augustinus-Buchhandlung-Verlag.(Interesting discussion on logic and literature.)
Redmond, W. (1998, forthcoming) *Manual de la lógica extensional de identidad del Siglo de Oro*, Mexico: National University of Mexico.(A programmed textbook on late scholastic identity logic with exercises.)
Redmond, W. and Beuchot, M. (1985) *La lógica mexicana del Siglo de Oro (Mexican Logic in the Golden Age)*, Mexico: National University of Mexico.(Historical and technical aspects of the logic of de la Vera Cruz, Mercado and Rubio.)
Redmond, W. and Beuchot, M. (1995) *La teoría de la argumentación en el México colonial (Argumentation Theory in Colonial Mexico)*, Mexico: National University of Mexico.(Analysis of logical argumentation in Mexico from the sixteenth to eighteenth centuries.)
Rubio, A. de la Rueda (1603) *Commentarii in universam Aristotelis dialecticam (Logica mexicana)*, Alcalá; Cologne, 1605; Valencia, abridged edn, 1606.(The author's original theory of the relation of logic to science.)
Suárez, F. (1597) *Disputationes metaphysicae (Metaphysical Disputations)*, in *Opera omnia*, vols 25-6, Paris: Louis Vivès, 1856-66; repr. Hildesheim: Olms, 1965.(Suárez's systematic presentation of metaphysics.)
Vera Cruz, A. de la (1554) *Recognitio Summularum (Examination of the Summulae (of Peter of Spain))*, vol. 1, *Dialectica Resolutio (Analysis of Logic)*, vol. 2, Mexico City.(Suárez's work on metaphysics perhaps eclipsed that of Rubio.)
Vera Cruz, A. de la (1557) *Physica speculatio (Speculation on Physics)*, Mexico City.(The author speculates on physics.)

Latin America, philosophy in

Geographically, Latin America extends from the Mexican-US border to those regions of Antarctica to which various Latin American countries have laid claim. It includes the Spanish-speaking Caribbean. Philosophy in Latin America dates from pre-Columbian (before 1492 in Hispanic America) and precabralian times (before 1500 in Brazil). Autochthonous cultures, particularly the Aztecs, Mayas, Incas and Tupi-Guarani, produced sophisticated thought systems centuries before the arrival of Europeans in America.

Academic philosophy began in the sixteenth century when the Catholic church began to establish schools, monasteries, convents and seminaries in Latin America. The seventeenth century saw little philosophical activity as effort was made to use academic thought to maintain the status quo, which reinforced a basically medieval worldview. Intellectually, the eighteenth century perpetuated this calm traditionalism until mid-century when a generation of Jesuits tried to break with the thought of Aristotle in order to modernize it. Political turmoil prevented academic philosophy from broadening in the early part of the nineteenth century. Later in the nineteenth century and into the early twentieth, positivism eventually became entrenched in most Latin American countries. In the early twentieth century new intellectual movements began as a backlash against anti-positivism.

1 Latin American philosophy up to the nineteenth century

Indigenous cultures, particularly the Aztecs, Mayas, Incas and Tupi-Guarani, produced interesting and sophisticated thought systems centuries before the arrival of Europeans in America. Many cultural artefacts were lost or destroyed so that study of this period involves many challenges in deciphering the subtleties and complexities of the earliest thought in Latin America. Indigenous cosmologies were often linked to phenomena in the natural world (see Latin America, pre-Columbian and indigenous thought in; Brazil, philosophy in).

Academic philosophy grew up in the sixteenth century when the Catholic church began to establish schools, monasteries, convents and seminaries in Latin America. If the encounter with the New World had significant impact on the European mind, this was not initially reflected in the philosophy being taught and written in the sixteenth and seventeenth centuries, which tended to restate and reinforce medieval values. However, intriguing writings on ethics and jurisprudence grew out of the contact between Spain and Latin America. Essentially, these writings analysed the relationship between cultural differences and human rights. The Dominican friar Bartolomé de las Casas was a pivotal figure who defended the rights of native and African peoples living in the Indies in the sixteenth century (see Latin America, colonial thought in; Mexico, philosophy in).

With a few notable exceptions, the seventeenth century was largely moribund philosophically because most efforts were directed towards using academic thought to maintain the status quo, which reinforced a fundamentally medieval worldview. The main philosophical task involved justifying and protecting the Catholic faith against Protestantism and science. Scholasticism was the dominant trend. However, there were some exceptions to the dominant practices in the form of several remarkable historical and philosophical figures. Antonio Rubio’s studies on logic are remarkably advanced. Juana Inés de la Cruz had a brilliant philosophical mind and is usually considered one of the earliest feminist thinkers in America (see Latin America, colonial thought in; Feminist thought in Latin America).

Intellectually, the eighteenth century continued this calm traditionalism until mid-century when a generation of Jesuits tried to break with the thought of Aristotle and bring philosophy into ’modernity’. They were primarily influenced by post-Renaissance Italian and French philosophy. However, the Jesuit order was expelled from the Spanish-speaking world in 1767. This delayed the introduction of proto-modern European philosophy in Latin America. The eighteenth century has become the subject of much revisionist philosophical study, particularly in Mexico (see Mexico, philosophy in).

Academic philosophy still did not broaden in the early nineteenth century because of political turmoil both in various Latin American countries and in Europe. Universities occasionally closed. This inhibited academic philosophical progress as universities were the locus of much philosophical activity. A more productive forum for philosophy was often the political arena in which thoughtful essays of ideas were written by nonacademics on themes such as constitutional government, progress and autonomy (see Literature, philosophy in Latin American;

Argentina, philosophy in).

Later in the nineteenth century and into the early twentieth, positivism eventually became entrenched in most Latin American countries. This movement claimed to be an objective methodology of the sciences. It was widely believed that scientific doctrines could provide the most efficient management of society through educational and political reforms. Auguste Comte and Herbert Spencer were the primary positivist influences in Latin America (see Positivist thought in Latin America; Analytical philosophy in Latin America; Brazil, philosophy in).

2 Latin American philosophy in the twentieth century

In the early twentieth century new intellectual movements began. Arising from these was a strong, thoroughgoing anti-positivist backlash. Ideas that positivists had promoted as ’scientific’ were rejected by anti-positivists for being scientistic (see Anti-positivist thought in Latin America; Argentina, philosophy in). Philosophers entertained idealism, vitalism, pragmatism and various political and social philosophies (see Phenomenology in Latin America; Existentialist thought in Latin America). Neo-Thomist thought continued to be widely studied, primarily in the Catholic universities.

A focus on regional thought in Latin America was an outgrowth of anti-positivist thought and a consequence of the arrival of Spanish philosophers who were exiled after the fall of Republican Spain. The writings of the Spaniard, José Ortega y Gasset, were widely influential in shaping Latin American philosophical reflections. Philosophers addressed the question of authenticity as they explored whether Latin Americans were simply adopting European philosophies, or whether they themselves had any authentic philosophy to offer. Many concluded that Latin Americans were adapting, rather than adopting European philosophies to their own reality (see Existentialist thought in Latin America; Phenomenolgy in Latin America; Anti-positivist thought in Latin America).

This process of critical self-examination, or ’autognosis’, was twofold. First, philosophers in individual countries and regions of Latin America sought to identify what was unique or distinctive about their thought or being. Later, philosophical contributions made by Latin America as a whole were compared and contrasted with those of other regions in the world (see Argentina, philosophy in; Brazil, philosophy in; Mexico, philosophy in). Studying Latin American thought in comparative perspective engendered a debate of considerable longevity over whether ‘Latin American philosophy’ exists or whether ’philosophy in Latin America’ is a more accurate denotation. Every Latin American country, including Puerto Rico, can be argued to possess unique philosophical traditions. At the same time there exists an extensive body of argument and commentary on what kind of philosophy, if any, can claim to be ’universal’.

Since analytical philosophy presents perspectives, methods and projects which claim to have universal appeal and applicability, it is often embraced in academic circles and is most frequently entrenched institutionally in Mexico, Brazil and Argentina. Analytical philosophy in these countries, while not obviously a response to immediate regional social, political or economic circumstances, serves to include and validate its adherents in international circles by adopting a style widely practised and accepted by mainstream Anglo-American academic philosophy. Attracted to the linguistic ’rigour’ of analytical philosophy, some adherents claim that it is the only way to do ’real philosophy’.

The late twentieth century reveals that it is possible to speak of both ’Latin American philosophy’ and ’philosophy in Latin America’. Some areas of philosophical research imbued with regional and cosmopolitan appeal are cultural identity, feminist thought, liberation philosophy, marginality and Marxist thought in Latin America (see Cultural identity; Feminist thought in Latin America; Liberation philosophy; Marginality; Marxist thought in Latin America). Many of these areas are profoundly engaged with Latin American realities in historical context. Rather than blindly adopting canonical Western philosophical paradigms, writers in these traditions seek to broaden the definition of what is human by convincingly articulating and incorporating Latin American experience and values into both the crucial discourses of philosophy and the pressing themes of the modern world (see Marginality).

Marxist philosophy has been and most likely will continue to be significant in Latin America partly because of continuing problems of economic disparities. Concerns with retributive justice, human rights and issues of power and truth, as well as the belief that Marxist theory more accurately describes reality, contribute to the vitality of this thought. Despite the collapse of the Soviet Union and the passing of Maoism in China, for many the Cuban

Revolution of 1959 is still idealized because it continues to threaten the US ’monster’ to the north, while advancing the notion of a supportive, egalitarian and responsible community. The Peruvian José Carlos Mariátegui was an original Latin American Marxist thinker whose thought has generated interest and respect internationally (see Marxist thought in Latin America).

One of the best known and most interesting contributions of modern Latin American intellectual life is liberation philosophy. The philosophical movement originated in Argentina, although many of its practitioners reside in other Latin American countries (see Argentina, philosophy in; Mexico, philosophy in). Philosophy of liberation should not be confused with liberation theology (see Liberation theology). Philosophy of liberation attempts to explain philosophically the theoretical underpinnings of social and political phenomena, such as dependency, and reinforces theology of liberation. These movements are responses to significant events in twentieth-century Latin America such as the Cuban Revolution (1959), the Argentine ‘Dirty war’ (1976-1983) and repressive regimes which began in Guatemala in 1954, Brazil in 1964 and Chile in 1973. Other political topics for these writers included populism, Marxism and Peronism. Philosophy of liberation differs from theology of liberation, Latin Americanist philosophy and Marxist philosophy especially in terms of its more limited accessibility. Philosophers of liberation employ a complex and specialized vocabulary which requires initiation on the part of readers. In addition, philosophy of liberation is not a unified movement: it is more appropriate to speak of philosophies of liberation. Such fragmentation in this field can be partly explained by the political orientations of thinkers whose views range from the extreme left to the extreme right. Their philosophical influences vary widely and include Francophone, German and other Latin American thinkers.

Philosophical activity in Latin America is characterized by a tremendous diversity of focus and methodologies. Latin Americans are keenly aware of philosophical developments in the rest of the world and thus entertain a variety of philosophical stances: progressive and conservative, pragmatist and idealist, materialist and spiritualist. Numerous philosophical interests and projects exist in Latin America because of a diversified and active philosophical profession, an interested public, some government support, a cultural awareness of other continents among the educated and noneducated alike and a widespread faith in education as a key to development.

AMY A. OLIVER

Latin America, pre-Columbian and indigenous thought in

The term 'pre-Columbian thought' refers to the set of beliefs and ideas held by the civilizations existing in Latin America prior to the arrival of Columbus in 1492. Research in pre-Columbian thought poses several questions linked to language, interpretation, chronology and cultural diversity. They can roughly be organized according to the three main regions in which the indigenous cultures flourished upon the arrival of the Spanish invaders: Nahuatl/Aztec *in central and southern Mexico,* Quiché/Maya *from Yucatan in southern Mexico to Honduras in central America and* Quechua/Inca*, from Ecuador to northern Chile. They each correspond to an empire into which previously many diverse, distinguishable peoples were assimilated. Since the Spanish invaders had destroyed most of their 'heretical' cultural objects by 1550, the question arises whether an accurate knowledge of their thought can be obtained.*

Some ethnohistorians believe that each of the aforementioned cultures developed a hieroglyphic system of codification and documentation, called Codices*, to preserve their theocosmogony, history and wisdom. Since the sixteenth century, however, it is known with certainty that only the Aztec and Mayan cultures developed such a system. According to some historians, the Incan culture did not use any kind of writing. They probably created 'paintings', as the Spaniards called their hieroglyphs, but these were totally destroyed.*

It is known that all pre-Columbian religions worshipped the events and forces of nature. The term used to name them was translated into Spanish as 'gods' when they were acceptable, or 'demons' when they seemed heretical - the indigenous peoples were polytheistic. The gods did not dwell in a region beyond our world, but rather populated it and were actively intertwined with it. All pre-Columbian cultures believed the sun to be the highest deity. The universe was conceived as a holistic structure in which human life, society and the gods were parts of an interrelated universe. Beyond these, however, the three cultures believed in an intangible, abstract deity, or principle, which ruled above all others.

The sun, being the highest tangible deity, led the priests of these cultures to observe the skies. Based on a highly developed knowledge of astronomy and mathematics, they established an accurate solar calendar.

1 Sources

In opposition to the colonizers of the USA and Canada, the Spanish invaders of the first generation (up to 1550 AD) described in great detail the cities, temples and well-organized commerce and transport systems they encountered. They were awed by the magnificence of the civilizations they discovered in Mexico, as expressed by Hernán Cortés (*c*.1485- *c*.1547) and Bernal Díaz del Castillo (d. *circa* 1560). Since they did not know the pre-Columbian languages and were both biased by the beliefs of their own age and culture and blinded by their greed for gold and silver, they did not collect direct information about the beliefs and thought of the original inhabitants. The recording of these beliefs became the task of some of the first Franciscan missionaries in Mexico and Guatemala. Bernard Sahagún (d. *circa* 1590) termed this set of beliefs 'philosophy' and Diego de Landa (1524-79 [1978]) called them 'heresy'.

Contemporary research on the pre-Colombian cultures relies on a variety of sources: first, archaeological excavations of ancient cities, temples and reliefs, sculpture, figurines and pottery; second: with regard to the Aztec and Mayan culture, codices, which are figurative, polychromatic manuscripts drawn by Indian priests; third: manuscripts dictated by converted Indian nobility, which then were translated into Latin, or else they were transliterated by Indians from the indigenous languages with the use of the Latin alphabet. (The Aztecs spoke, among other languages, Nahuatl or Mexica; the Mayas spoke, among other languages, Maya-Quiché.) All the codices depict themes inherited by the oral tradition: theocosmogony, history, religious ceremonies and social mores. Other sources include chronicles and histories written in Spanish by missionaries.

Research on Mayan thought differs from that on the Aztecs and Incas in that many of the codices depicting texts have not been completely deciphered. However, their numerical system (which was not decimal but vigesimal) has been well understood. Mayan priests, who were also mathematicians, conceived of the number zero prior to the arrival of the Spaniards. Remaining texts, such as the *Popol Vuh* (*c*.1600 [1996]) and *The Book of Chilam Balam of Chumayel* (*c*.1600 [1967]), were written in the Maya/Quiché language using the Latin alphabet. They have since

been translated into Spanish and other languages. They record Mayan cosmogony, the history of their cities, astronomy and astrology.

The only Inca codex to be found was the *quipu*, a sort of rosary or abacus made from hemp and beads. This was designed for purposes of numerical information. Research on Inca culture has mainly been obtained from massive monuments, such as the old cities of Macchu Picchu and Cuzco, with their observatories for recording the movement of the sun, moon and stars. Archaeological digs have led to the retrieval of pottery, jewellery and mummies that escaped the Spanish destruction. The oral tradition has also helped to maintain Inca history. In its technical capability, Incan culture has been compared to that of the Roman Empire. The Incas built an elaborate system of channels to capture the melting snow from the slopes of the Andes and to transport it to the valleys where they had settled, some of which are still in use in the 1990s. Similarly, they built a road system extending from the sacred city of Cuzco, which for the Incas was the centre of the universe, to the four cardinal points - north, south, east and west. In the coastal plains huge geometrical figures have been found, whose shape and size can be appreciated only from the air and no one has been able to interpret their function and meaning. In addition and most important, are several manuscripts written in Spanish by converted Inca noblemen, for example Diego de Castro Tito Cussi Yapanqui (*c*.1529-71) and Felipe Guamán Poma de Ayala (*c*.1550-*c*.1615 [1977]). They tell us that the last Incan emperor Atahualpa, ruler at the time of the arrival of the Spaniards and killed in 1533 by the conqueror Pizarro, established a ‘clock’ by which he determined the period of preparing the soil, seeding and harvesting.

2 Worldviews

As to the thoughts and worldviews of these cultures, it is difficult to form a clear and distinct conception due to the problem of language. The first translators, the Spanish Catholic friars, came to the New World to convert the original peoples to the ‘true religion’ during the century when Spain was dominated by the Counter-Reformation. These friars had been trained in the medieval Aristotelian view of both the true God and the world, and their descriptions mirrored this background. For instance, Bernabé Cobo, a missionary travelling in Peru, appalled by the Indians’ idolatry, referred to the sacred sculptures and figurines as ‘demons’. Newly converted Indians and several friars used the term ‘god’ as a translation of the names given to the various images representing the forces and events of nature.

According to Miguel León-Portilla (1963), the Aztec peoples had a precise conception of a metaphysical God despite their belief in a variety of deities corresponding to natural recurrent events, such as the recurrence of the moon and the sun and the cycles of planting and harvesting. All the deities, both feared and revered, were thought to be manifestations of a single supreme intangible, immaterial principle endowed with diverse attributes and present in all phenomena. Its name was Ometeotl, the god of duality and wisdom. As duality, he was partially male and female. His male name was Ometecutli, ‘he… who is activity on earth’ and his wife Omecihuatl, she who ‘gives stability to (our) earth’, as translated by León-Portilla (1963). These two notions are conceived as a unity which comprises everything by virtue of being dual: dark and light, masculine and feminine, life and death. One of Ometeotl’s attributes is *tloque nahuaque*, or owner of proximity in time and space, who rules over the earth. As a representation of activity and stability, Ometeotl can be interpreted as a dialectical principle which, as activity, is the principle of destruction and recreation, but which, by giving stability to the earth, maintains the world’s continuity. One of the main scholars of Nahuatl culture, Angel María Garibay (1892-1967), translates Ometeotl as ‘two-god’ (*c*.1930 [1965]).

A similar notion of duality can be found in the Mayan religion, where duality is called Hunab-Ku. In its masculine form, it is Itzamna, whose wife is Ix-Chel, or mother earth. They are the ancestors of all other gods and of humans. Reliefs show Itzamna as an old man representing wisdom, life-giving forces and writing. Since Itzamna is wisdom, he ‘knows something about time’. Therefore, he also is the patron of astronomical and mathematical knowledge. Similar to Ometeotl for the Aztecs, Itzamna and Ix-Chel, were conceived as a unity, or single metaphysical principle.

The Incan religion was different. Ticci Viracocha, also called Pachayachachic, was thought to be the creator and ruler of all things. According to Garcilaso Inca de la Vega (1539-1616) in *Royal Commenatries of the Incas and General History* (*c*.1600 [1966]), Pachayachachic is the world’s origin and soul, the giver of life and the mover of the universe. Since this deity is incorporeal, the mind cannot have an image of its being. A person who makes

contact with this god does so through the heart. Therefore it was forbidden to pronounce its name, to build temples dedicated to it and to offer up sacrifices to it. In a sense, Pachayachachic was a hidden god. Below this deity were the diverse Viracochas: the sun, representing agriculture, the earth, representing our mother and minor deities, such as rocks and earthquakes. The Incan emperor and his dynasty were also considered to be divine, but below the higher ranking Viracochas. According to Bernabé Cobo (*c*.1550 [1990]), one of the early missionaries in Peru, Ticci Viracocha was the highest deity, or as Cobo states, the first cause.

There is the question of whether the pre-Columbian peoples were polytheistic. Some interpreters claim that the Aztec and Mayan gods are images and should not to be understood as personifications of natural events and forces, but rather as ideas of them. The Aztecs called the regular appearance of the sun *tonatiuh*, meaning the main star in the sky; the Quechuas called it *apu inti*. The idea of the sun as the giver of life and the deity which keeps the sun, understood as the soul of the world moving, amounts to the same unique first metaphysical principle. Therefore, the first bishop of Guatemala and Chiapas, Fray Bartolemé de las Casas called these religions monotheistic.

With regard to their view of the universe, the three cultures have astonishing similarities. None of them knew of the existence of other continents. They viewed the universe as divided into four regions, each corresponding to the four cardinal directions. They believed in the existence of a region below our earth, where the dead are buried and survive in the dark. The Aztecs and Mayas, but not the Incas, believed in a recurrent cycle of creation and destruction. The Aztecs measured the universe's time in cycles, known as suns, equal to fifty-two years after which, due to a cosmic cataclysm, the universe would be destroyed, and accordingly recreated.

3 Three sages

León-Portilla (1963) translated works from Nahuatl into Spanish. He interpreted the poems of the most representative sages and priests, or *tlamatimine*, meaning 'he who knows something'. Counted among these sages were Quetzalcoatl, or the 'plumed serpent' (ninth century AD), probably king of Tula, Tlacaelel, advisor to the king (*c*.1398) and Netzahualcoyotl, king of Texcoco (1402-72).

From Quetzalcoatl we know how the Nahuas, and later the Aztecs, viewed the universe. It was imagined as an enormous is land divided horizontally into four regions surrounded by enormous waters. The four regions converge at the world's centre. East is the region of light, of fertility and of life, symbolized by the colour white; north is the universe's black region, where the dead are buried; west is the sun's house, its colour red; and south is the region of seeds, its colour blue. Vertically, the universe is built in layers, some above, others beneath the earth. Above the earth the sky forms a blue dome covered by a network of paths on which the sun, stars and moon move. At the highest level is Ometeotl, other gods are on lower levels. Beneath the surface of earth there are other levels, the deepest being the region of death, or Mictlan.

Netzahualcoyotl's thought is the best known. Some texts thought to be by Netzahualcoyotl remain, such as *Anales de Cuauhtitlan* (*c*.1402-72). One of the recurrent themes refers to the world's constant flow of time and the never ending change and futility of being. The often used expression *cachuitl*, or 'what is left behind', also refers to death and the sorrow of human mortality. This concern explicitly leads Netzahualcoyotl to reflect on the unique value of human life. Being aware of the inevitability of death, he reflects upon the possibility of overcoming it. He searches for a fixed point which does not change or perish. He sees it in the mystery of *tloque nahuaque*, or 'owner of proximity in time and space', but the mystery can only be invoked, not resolved. This fixed point can only become real if it grows roots in our hearts and the sorrow about death can only be silenced in one's heart, through *flor y canto*, or poetry. As king of Texcoco, Netzahualcoyotl unsuccessfully tried to oppose the practice of human sacrifice, a ritual brought to the land by the Aztecs, also known as Mexicas, the last indigenous group who had conquered the whole region by 1325 AD.

Tlacaelel's thought was the expression of an ideology that justified Aztec political practices. He was an outspoken advocate of the glorification of Huitzilopochtli, the god of war. In order to create the new cycle of life on earth following the last cataclysm, the gods gave their own blood as the elixir of life and strength. To correspond to this sacrifice it was necessary to offer them human blood. In order to have access to this elixir, Tlacaelel supported the practice of regular warfare, or *guerras floridas* against their neighbours for the explicit purpose of obtaining prisoners. This ideology became acceptable because human sacrifices made it impossible for the next cataclysm to happen. Hence, the Aztecs saw themselves as the chosen people who could overcome the inevitable destruction of

the universe.

When the missionaries arrived in the New World, they were appalled by the immorality of the indigenous peoples. Not only were they shocked by the practice of human sacrifices, but also disgusted by the purported custom of eating the flesh of victims. However, cannibalism has been a controversial issue among historians who think it was alleged by the conquerors to justify the appropriation and plundering of the whole territory. In addition, most sixteenth-century chronicles report a variety of ‘sins’ attributed to biblical Babylon from prostitution and incest to homosexuality. It is true that the Incas worshipped fertility and many remaining *huacas*, or sacred figures in clay, are endowed with phallic forms.

Most recently, the opposite view has been taken. As the great empires required the rule of law and discipline, all three cultures were based on similar moral values. The main virtues were courage, fortitude, sobriety and obedience. Drunkenness was regarded as the source of all vices. Monogamy seems to have been the correct relationship between man and woman, since under the Incas adultery, like theft, was punished with death. Greed was condemned and riches in gold, silver and precious stones (and among the Aztecs and Mayas, the quetzal’s colourful feathers), belonged to the gods, their temples and the courts of emperors and priests.

Sahagún (*c*.1550 [1970]) outlined the many virtues an Aztec father taught his children. Both sons and daughters had to worship the gods daily, be diligent in their tasks so that time would not be wasted, be meticulous when carrying out their craft, act friendly and respectfully towards others, keep peace with them, be honest and never steal from them. He taught his daughter to practise chastity and modesty before marriage, to be obedient to her father or husband and to do the work of women, such as cooking, weaving and painting.

In addition, he taught his sons not to act arrogantly, not to sow disharmony, to cultivate the land, to learn an honourable trade or art, to be always modest and enduring and always be responsible for his wife and family.

See also: Latin America, colonial thought in

LAURA MUES DE SCHRENK

References and further reading

Anonymous (*c*.1600) *Popol Vuh*, trans. D. Tedlock, New York: Simon & Schuster, 1996.(An excellent translation of the myths and history of the Quiché/Mayan peoples.)

Anonymous (1600) *The Book of Chilam Balam of Chumayel*, Oklahoma, OK: University of Oklahoma Press, 1967.(Another translation of an anonymous text written in the Maya/Quiché language using the Latin alphabet.)

Cobo, B. (*c*.1550) *Inca Religion and Customs*, Austin, TX: University of Texas Press, 1990.(Part of a more extensive work and shows the animosity felt by the many missionaries for the pagan Incan religion and mores.)

Coe, D.M. (1992) *Breaking the Maya Code*, New York: Thames & Hudson.(An important book with regard to deciphering Mayan hieroglyphs.)

Coe, D.M. (1993) *The Maya*, New York: Thames & Hudson.(One of the best books in English discussing Mayan culture).

Cortés, H. (*c*.1523-27) *Letters From Mexico*, New York: Grossman Publishers, 1971.(Letters to Charles I of Spain. A historical document in which Cortés tries to prove the glory of his conquest.)

Díaz del Castillo, B. (*c*.1570) *The Conquest of New Spain*, New York: Penguin, 1963.(Memoirs by Cortés’s captain. Probably one of the most interesting introductions to the adventures of the conquest.)

Garibay, A.M. (*c*.1930) *Teogonía e historía de los mexicanos (The Family Tree of Pagan Gods),* Mexico: Universidad Nacional Autónoma de México, 1965.(The translation of Nahuatl texts and commentaries on their origins.)

Guamán Poma de Ayala, F. (*c*.1550- *c*.1615) *Letter to a King: A Picture History of the Inca Civilization*, trans. C. Dilke, London: Allen & Unwin, 1977.(A good translation from the Spanish. The translation reproducesGuamán Poma de Ayala’s irony and his profound moral sense. It includes his drawings.)

Inca de la Vega, G. (*c*.1600) *Royal Commentaries of the Incas and General History*, Austin, TX: University of Texas Press, 2 vols, 1966.(An excellent source of Inca history and mores.)

Landa, D. de (1524-79) *Yucatan Before and After the Conquest*, trans. W. Gates, New York: Dover Publications, 1978.(Description of religion, mores and bleeding practices of the Mayan peoples. Source for deciphering the written Maya code.)

León-Portilla, M. (1963) *Aztec Thought and Culture: A Study of the Ancient Nahuatl Mind*, Oklahoma City, OK:

Oklahoma University Press.(The main source for philosophers who do not read Nahuatl and who are interested in interpreting indigenous thought.)

Morley, S. (1946) *The Ancient Maya*, Stanford, CA: Stanford University Press, 1983.(The story of the author's research into the Mayan past.)

Netzahualcoyotl (*c*.1402-72) *Anales de Cuauhtitlan (Review of Cuauhtitlan)*, in *Chimalpopoca Codex: Noticias históricas de México y sus contornos (Chimalpopoca Codex: Historical Notices of Mexico and it Surroundings)* Mexico: Imprenta de Escalante, 1885; trans. P.F. Velázquez, Mexico: Imprenta Universitaria, 1945.(The paleography of this text was also published in German in 1938.)

Sahagún, B. de (*c*.1550) *General History of the Things of New Spain*, trans. C.E. Dibble and A.J.O. Anderson, Santa Fe, NM: School of American Research, 1970.(The complete translation of the Aztec text recording the pre-Colombian history of Mexico, written with the Latin alphabet by indigenous sages known as Sahagún's informers.)

Latitudinarianism

The term 'Latitudinarianism' designated, initially abusively, the attitudes of a group of late seventeenth-century Anglican clergy who advocated ecclesiastical moderation, voiced broad if heavily qualified support for religious toleration, and emphasized an undogmatic probabilism, 'moral certainty', a reasoned faith and moral performance over against infallibility, dogma, ritual performance and 'unreasoning' faith. They attempted to construct a 'reasonable' faith, with some emphasizing belief in carefully evaluated miracles to attest to the central truths of Christianity. The Latitudinarians had considerable influence on the thought of John Locke, among others, although Locke's anti-clericalism, tolerationism and reticence on the Trinity went beyond their positions. The most important of the Latitudinarians, listed from the most eirenic to the least, were Edward Fowler, Benjamin Whichcote, John Wilkins, John Tillotson, Gilbert Burnet, Joseph Glanvill and Edward Stillingfleet; they were particularly influenced by the thought of William Chillingworth, the Cambridge Platonists and Hugo Grotius.

1 Lives, fallibility and attitudes towards toleration

The Latitudinarians were mainly educated during controversies between Catholics and Protestants over the 'Rule of Faith', during the attempts by many puritans in the English civil wars to impose their own brands of religion on others, backed by the certainties of 'conscience', and during a time when there was a proliferation of sects emphasizing the roles of illumination, grace and private revelation. The Latitudinarians' heroes of this period were eirenic Anglicans such as William Chillingworth and Viscount Falkland who had stressed, in contrast to both Puritans and Roman Catholic controversialists, humankind's fallibility, the need to proportion assent to the evidence, the limited number of fundamentals of Christianity and the legitimacy of differing opinions. Several of the Latitudinarians were, furthermore, taught by the leading Cambridge Platonists and influenced especially by their emphases on reason and morality (see Cambridge Platonism).

The Latitudinarians reiterated many of the positions of Chillingworth, Falkland and the Cambridge Platonists, stressing that the fundamentals of Christianity were clear but should not be specified 'too forwardly', that differing opinions on other issues within Christianity were unavoidable, and that no particular form of church government was required by divine or natural law. They supported accommodating 'a latitude' of different practices and beliefs in a broad church, which they saw as a form of religious toleration, combined - with varying degrees of reluctance - with toleration for those few to be left outside. They supported the toleration provided by the Revolution of 1689 and rose at that point to occupy the highest offices in the Church of England. As Whichcote argued, 'our fallibility and the Shortness of our Knowledge should make us peaceable and gentle: because I may be mistaken I must not be dogmatical' (Ashcraft *et al*. 1992: 236). For Burnet, it was from their allowance of 'great freedom both in philosophy and divinity' that they received the name of Latitudinarians. Their argument that no particular form of church was required by Christianity, their confidence that in fact they reasoned correctly and others were 'enthusiastic' and unreasonable, and their emphasis on peace and the authority of the civil magistrate, were, however, often combined before 1689 with a preference for episcopacy and with an emphasis on the duty to unite in the civilly established church, limiting their tolerationism. The most important works of the Latitudinarians enunciating these principles were Stillingfleet's *Irenicum* (1660) and, far less eirenic, his *Unreasonableness of Separation* (1681), the *Brief Account of the New Sect of Latitude-Men* by 'S.P.' (1662), Tillotson's sermons and especially Edward Fowler's *Principles and Practices of Certain Moderate Divines* (1670).

2 Epistemology: certainty, 'moral certainty' and miracles

In a series of works and sermons John Tillotson, Edward Stillingfleet and John Wilkins refined and extended the arguments about knowledge made by Chillingworth earlier in the century. 'Absolute' infallibility was declared possible only for God. For Tillotson, however, a kind of 'conditional' infallibility was possible for humans. Immediate sense perception gave the highest assurance attainable by man alone. A similarly high level of certainty attached to demonstration, most notably possible in the field of mathematics. For demonstration, the denial of the conclusion necessarily involved a contradiction. Wilkins similarly described conditional 'infallible certainty' as associated both with the physical certainty derived from sense-data, the 'highest kind of Evidence, of which humane nature is capable' and with the mathematical certainty of matters of self-evidence, basic principles of logic, and deductions from these principles (Shapiro 1983: 85).

In most matters of religion, however, only the 'moral certainty' of belief and no higher level of certainty was possible. It was none the less sufficient. For Tillotson, such belief excluded 'reasonable cause' for a 'considerate Man' to doubt, but not 'all possibility of Mistake'. It thus involved a probability sufficiently high to call for 'a firm and undoubted assent' on grounds 'fit fully to satisfy a prudent man' (Van Leeusen 1970: 37). As Wilkins put it, there was an 'indubitable certainty' the objects of which did not compel assent but which were so plain, 'that every man whose judgment is free from prejudice will consent to them' (Shapiro 1983: 85-6). God was declared to be good, and so had created humans in such a way that they assented to evidence properly; only an unreasonable person would require more than the appropriate level of certainty. As Stillingfleet put it, God 'doth not require us to believe any thing without sufficient grounds for our believing it, and those grounds do bear a proportionable evidence to the nature of that assent which he requires' (Carroll 1975: 68).

These arguments were designed to counter those who were too dogmatic and those who were too sceptical. Dogmatists were unreasonable because many beliefs needed a far more careful assessment of testimony and evidence and a far more careful interpretation of the words of Scripture than they produced. Many beliefs could not reach the level of certainty that dogmatists claimed. Human knowledge was limited, and humans were subject to error. Sceptics were unreasonable because they wished for more certainty than was required, focused on 'subtilties', and ended with more doubt than was appropriate. Atheists were particularly unreasonable because they required more evidence for belief in the existence of God than was necessary, and neglected the testimony of the intelligent creation in the 'book of the world'. According to Tillotson, atheists were also unreasonable because, in quasi-Pascalian terms, they were imprudent in ignoring the 'danger' of choosing a few base pleasures when God might exist and if so would punish them exceedingly (see Pascal, B. §6). Tillotson and Wilkins focused on the pursuit of happiness as a motivation to be religious, with man's dominant interest being happiness in the next life. As Wilkins put it, the 'rational and prudent man' should 'order his actions in favour of that way which appears to be most safe and advantageous for his own interest' (Shapiro 1983: 89). Emphasizing the morality taught in Revelation and by natural reason, and the role of God's rewards and punishments to motivate moral performance, the Latitudinarians effectively identified evangelical righteousness with natural righteousness.

For several of the Latitudinarians miracles had been crucial to generating appropriate belief: 'to confirm a Divine Testimony to the World, and to make that appear credible which otherwise would have seemed incredible' (Stillingfleet, cited in Carroll 1975: 68). Belief in such miracles depended on the credibility of their witnesses. Such credibility therefore needed to be assessed carefully in order to yield appropriate moral certainty. Writing many works against Roman Catholicism in his early years, Stillingfleet particularly sought to distinguish true miracles from false wonders, and to repudiate Catholic claims about the continued performance of miracles. He therefore set out a series of criteria to distinguish miracles from frauds: that miracles were wrought to confirm a divine testimony; that miracles did not contradict revelation; that they left divine effects on believers; that they tended to overthrow the devil's power; that the circumstances and manner of operation of the true differed from the false; and that they evidently exceeded created power. For Stillingfleet, once Christianity was established, further miracles were unnecessary. Catholic claims were therefore false (see Miracles).

The arguments of Tillotson, Stillingfleet and Wilkins on proportioning assent to the evidence, the limitations of knowledge, the dependence of men upon appearances and the rejection as 'unreasonable' of scepticism, atheism, and dogmatism in favour of a reliance on 'moral certainty' were paralleled and extended in the fields of scientific enquiry and defences of natural religion in many other works by Latitudinarians and their associates, including Joseph Glanvill's *Vanity of Dogmatizing* (1661) and *Essays on Several Important Subjects in Philosophy and Religion* (1676), and many of the scientific works of Robert Boyle (see Glanvill, J.; Boyle, R.).

3 Locke and Latitudinarianism

Locke knew several Latitudinarians well, especially Tillotson and Fowler, and he indicated support for 'latitudinism' in his *Essay on Toleration*. The *Essay Concerning Human Understanding* (1689) presented a generalized philosophic statement significantly influenced by many elements of the Latitudinarians' account of the limitations of the human mind and proportioning of assent. Many of Locke's emphases on moralism, on God's rewards and punishments, and on the limited number of fundamentals of Christianity, were shared with the Latitudinarians. However, Locke restricted the term 'certainty' to that which is known to be true, and refused to apply it to that which was probable or 'morally certain'. Moreover, he argued that the idea of substance could not

come from sensation or reflection, and made the idea of substance an idea of ‘I know not what’, effectively undercutting some of the Latitudinarians’ attempts in the 1690s to defend the Trinity against the claims of deists and unitarians that the application of reason to religion supported their own more radical rejections of the Trinity or of Christianity itself (see Deism). In reply to Locke, Stillingfleet utilized scholastic metaphysics in defence of the mysteries of religion, holding that substance was a ‘rational idea’ emanating from the soul itself. Other Latitudinarians, such as Tillotson, in general preferred to stress that the Trinity was to be believed as ‘above’ but not contrary to reason; Tillotson defended the Trinity in the 1690s. For their emphases on reason, morality and toleration, the Latitudinarians were often accused of Socinianism; Locke was closer to it (see Socinianism).

See also: Religion, history of philosophy of

JOHN MARSHALL

References and further reading

Ashcraft, R., Kroll, R. and Zagorin, P. (eds) (1992) *Philosophy, Science and Religion in England 1640-1700*, Cambridge: Cambridge University Press.(Especially useful are essays by Rogers on ignorance and toleration; Ashcraft on Latitudinarians and intolerance; Hutton on Stillingfleet and More; and Marshall on Locke and Latitudinarianism.)

Carroll, R.T. (1975) *The Commonsense Philosophy of Religion of Bishop Edward Stillingfleet*, The Hague: Martinus Nijhoff.(Especially good on Stillingfleet’s analysis of ‘moral certainty’, criteria for miracles, scholastic rationalism, and relative intolerance. See §2 above.)

Fowler, E. (1670) *Principles and Practices of Certain Moderate Divines*, London.(A general statement of the eirenic principles of Latitudinarians.)

Glanvill, J. (1661) *The Vanity of Dogmatizing: or Confidence in Opinions. Manifested in a Discourse of the Shortness and Incertainty of our Knowledge, and its Causes, with some reflections on Peripateticism; and an Apology for Philosophy*, London; repr. New York: Facsimile Text Society, Columbia University Press, 1931.(A work supporting proportioning assent to the evidence in the comprehension of nature.)

Glanvill, J. (1676) *Essays on Several Important Subjects in Philosophy and Religion*, London.(Essays further articulating the ideas of *The Vanity of Dogmatizing*.)

Locke, J. (1667) *Essay on Toleration*, in *John Locke: Writings on Politics, Law and Religious Toleration 1667-1683*, ed. J.R. Milton and P. Milton, Oxford: Oxford University Press, forthcoming.(An early tolerationist essay.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Oxford University Press, 1975.(Locke’s *magnum opus*, significantly influenced by latitudinarians’ ideas.)

Shapiro, B. (1983) *Probability and Certainty in Seventeenth- Century England*, Princeton, NJ: Princeton University Press.(A very useful survey of the relationships between issues of probability and certainty in natural science, religion, history, law and literature in seventeenth-century England.)

S.P. (1662) *Brief Account of the New Sect of Latitude-Men*, London; repr. William Andrews Clark Library, University of California, 1963.(A defence and definition of ecclesiastical moderation against Roman Catholicism, fanaticism, atheism, enthusiasm and superstition.)

Stillingfleet, E. (1660) *Irenicum*, London.(An eirenic early work emphasizing civil authority)

Stillingfleet, E. (1662) *Origines Sacrae*, London.(A defence of the primacy and authenticity of Scripture.)

Stillingfleet, E. (1681) *The Unreasonableness of Separation*, London.(A uniformitarian attack on nonconformist separation.)

Stillingfleet, E. (1696) *A Vindication of the Doctrine of the Trinity*, London.(A scholastic metaphysical defence of the Trinity against deism, unitarianism and, less directly, against Locke’s account of the limits of the understanding.)

Tillotson, J. (1664) *The Wisdom of Being Religious*, London.(A quasi-utilitarian defence of religiosity.)

Tillotson, J. (1666) *The Rule of Faith*, London.(Against Roman Catholics on standards of judgment in religious controversy.)

Tulloch, J. (1966) *Rational Theology and Christian Philosophy*, Germany: Georg Olms Verlagsbuchhandlung, 2 vols.(Reprint of a nineteenth-century general survey of philosophy and theology; dated but still useful.)

Wilkins, J. (1672) *Of the Principles and Duties of Natural Religion*, London.(A defence of natural religion; identifying natural morality and evangelical morality and emphasizing rewards.)

Van Leeuwen, H. (1970) *The Problem of Certainty in English Thought 1630-90*, The Hague: Martinus Nijhoff. (A precise study of epistemology and certainty in the period leading up to and including the latitudinarians; excellent on the similarities and divergences of Locke and the latitudinarians.)

Lavrov, Pëtr Lavrovich (1823-1900)

*Pëtr Lavrov was one of the main theorists of Russian populism (*narodnichestvo*) - a trend of thought and a movement which crystallized after the abolition of serfdom in Russia in 1861.*

There were many different currents within this broad trend but all of them concentrated on the possibility and desirability of securing a non-capitalist way of development for Russia. This was so because the populists, having perceived (often with the help of Marx's Capital) the contradictions of capitalist development, lost their confidence in 'European' progress, recognized in capitalism only a regression and chose therefore an adamantly anti-capitalist stand, combined, as a rule, with backward-looking idealization of the peasant commune.

The dominant Western theories of social evolution - from Spencerian liberalism to Marxism - strongly emphasized the 'objective' character of the laws of social development and defined capitalism as a necessary phase of progress. Russian populist socialists deeply felt that in Russia's backward condition such theories offered a convenient tool for apologists of capitalist progress, who sanctioned and justified the suffering of the masses by referring to the 'objective laws of history' or 'the iron laws of political economy'. This led them to a demonstrative rejection of the 'objectivist' conception of progress. Lavrov was the first populist theorist who set against 'objectivism' a conscious and systematic vindication of 'subjectivism'. He was supported in this by another populist thinker, Nikolai Mikhailovskii. The common features of their views have been labelled 'subjective sociology' or the 'subjective method'.

1 Anthropologism

Pëtr Lavrovich Lavrov, the son of a wealthy Russian landowner, embarked in his youth on a military career. By 1858 he had already attained the rank of colonel. He was also interested in philosophy and sociology, and in 1860 published his first book, *Ocherki voprosov prakticheskoi filosofii (Sketches in the Domain of Practical Philosophy)*, in which he defined his position as 'anthropologism'. In 1862 he became a member of the underground revolutionary organization Land and Freedom. In 1866 he was arrested and sentenced to exile in Vologda province. In February 1870 he managed to flee abroad, made contacts with the International Workingmen's Association and, in the next year, took part in the Paris Commune. From 1873 to 1876 he published, in Zurich and in London, an influential revolutionary periodical *Vperëd (Forward)*.

Lavrov's initial philosophical position, as presented in *Ocherki voprosov prakticheskoi filosofii* and in his lectures 'On the Contemporary Significance of Philosophy' (1860), was directly connected with the antitheological and anti-Hegelian anthropologism of Ludwig Feuerbach: it was conceived as the radical overcoming of speculative idealism, which reduced individuals to mere tools of Absolute Spirit, and as a philosophical vindication of really existing human beings, creatures of flesh and blood. The human person as an indivisible physical-psychological individual was for Lavrov the only indisputable datum and the necessary starting-point for a scientific philosophy. This entailed commitment to a sort of anthropocentric scientism and anthropological relativism. Philosophical questions were interpreted as pertaining ultimately to man, to the human experience, while 'things-in-themselves' were excluded from the sphere of legitimate scientific knowledge. The relativistic component, inherent in this view, effectively undermined metaphysics but did not apply to ethics. Lavrov saw man as an ethical being and, following Kant, argued for the possibility of non-relative judgments in the sphere of 'practical reason'.

Lavrov took care to distinguish his position from positivistic scientism, based on the natural sciences. In the essay 'Zadachi positivisma i ikh reshenie' (The Tasks of Positivism and their Solution) (1868) he discussed different variants of positivist thought, recognized the centrality of positivism in contemporary intellectual life and fully endorsed the positivist commitment to strictly scientific methods. At the same time he criticized the one-sided 'objectivism' of contemporary positivism, setting against it the positivism of Auguste Comte, which recognized the legitimacy of the 'subjective method'. His conclusion was that positivism proved incapable of solving its own problems because it lacked a unifying philosophical principle. This principle was man as a feeling and thinking being, a symbol of the true unity of mind and body. Thus, the historical role of positivism was only to pose problems - their solution would be provided by an anthropological philosophy whose germ could be found in Feuerbach, Proudhon and J.S. Mill.

2 'Subjective method' and theory of history

Lavrov's preoccupation with legitimizing the 'subjective method' reached a new level in his *Istoricheskie pis'ma (Historical Letters)* (first published, under the pseudonym 'Mirtov', in a Russian periodical, in 1868-9). This short book gained him immense popularity among young radicals, tormented by feelings of social guilt and eager to sacrifice their personal interests for the good of the masses.

Historical Letters owed this enormous success mainly to one chapter, 'The Price of Progress'. The possibility of talking about progress, argued Lavrov, has been dearly bought by the human race. The personal development of some members of the privileged classes has been purchased with the labour and suffering of many generations of heavily exploited people. The 'conscious minority' should never forget its debt and should make every effort to discharge it.

Lavrov's conviction that the 'debt to the people' must be paid off now was incompatible with the theory of progress as an objective and inevitable law of development. The 'objectivist' conception of progress had to be replaced by a 'subjectivist', ethical conception, based upon 'subjective sociology'.

The basic assumptions of Lavrov's 'subjectivism' can be embraced in three points. First, it was a defence of ethicism, a standpoint derived from a strong conviction that values are independent of facts, that moral evil cannot be 'scientifically' explained away, and that moral protest against suffering is valuable and obligatory irrespective of any 'objective' condition. Second, it was an epistemological and methodological position, which denied or disputed the possibility of 'objective' methods in the social sciences: it claimed that historical and sociological knowledge can never be value-free because it always depends upon the unconscious emotions or (much better) on the consciously chosen ideals of scholars. Third, it was a philosophy of history which maintained that the 'subjective factor' - human thought and will - can effectively oppose the spontaneous tendencies of development and play a decisive part in historical process. Lavrov based upon it his 'practical philosophy' which proclaimed that 'critically thinking individuals', having united in a party, could become a social force and change the existing state of affairs in the direction indicated by their 'subjective' aims.

In its application to methodological problems Lavrov's 'subjectivism' yielded an interesting epistemology of history, anticipating, to a certain extent, the revolt against positivism in the humanities. The so-called 'laws of history', he claimed, do not exist; historical events are always unique and unrepeatable. Historians cannot limit their task to the description of facts, since 'facts' are not something given; they have to be selected and endowed with meaning, which presupposes a creative activity on the part of the knower. In looking at history, therefore, the main problem is one of selection, of finding a criterion that will make it possible to pick out 'what is important and meaningful' from the amorphous mass of historical data. Such a criterion must be subjective because it depends on the social ideal adopted by a particular scholar. All facts are classified and all historical events interpreted according to how they relate to this ideal. Progress is a category required to impose order on the raw material of history. In itself, history has no meaning: there are many meanings to be found in it, but all of them are imparted to it by human individuals. Imposing a meaning on history presupposes an ideal not only in the sphere of historical understanding but also in the sphere of historical action. Human history begins therefore with the emergence of critically thinking individuals trying to shape society by means of 'criticism' and 'idealization'.

Lavrov formulated his own ideal as follows: 'The physical, intellectual, and moral development of the individual; the incorporation of truth and justice in social institutions'. Progressive approximation to this ideal was defined by him as a process that is being accomplished by means of the critical thought of individuals who aim at the transformation of their culture. By 'culture' Lavrov meant a static social structure based on religion, tradition and folk ways; by 'civilization' he meant a dynamic social structure in which religion was replaced by science, and custom by law. These two concepts were further developed in his sociological theory.

3 Sociological conceptions

With the decline of revolutionary populism in the 1880s Lavrov returned to scholarly work and produced several books in the field of historical sociology, such as *Opyt istorii mysli novogo vremeni (Essay on the History of Modern Thought)* (1888-94), *Zadachi ponimaniia istorii (Problems in the Interpretation of History)* (1898), and the posthumously published *Vazhneishie momenty istorii mysli (Stages of Great Importance in the History of*

Thought) (1903). All these works reflect an influence of positivist evolutionism and thus appear to make considerable concessions to ‘objectivism’. On closer examination, however, it turns out that it was only a change in emphasis. In *Historical Letters* Lavrov made a distinction between history, which deals with what is unique, and sociology, which aims at discovering certain overall laws of development. Later, in an article, ’O metode v sotsiologii’ (On Method in Sociology), he clearly stated that in sociology, in contrast to history, both methods - the subjective and the objective - were justified and applicable. Recognizing the legitimacy of the ’objectivist’ approach did not change this view that conscious, subjectively motivated actions were axiologically infinitely superior to ’natural’ processes, assuming the form of ’objective laws’.

Lavrov defined sociology as a science dealing with the forms of social solidarity, which he divided into three main types. The first was the unconscious solidarity of custom; the second was a purely emotional solidarity, based on impulses not controlled by critical reflection; the third was ’conscious historical solidarity’ resulting from a common effort to attain a consciously selected and rationally justified goal. This third type represented the highest and most important form of human solidarity. It evolved later than the first two types and heralded the process of the transformation of static ‘culture’ into dynamic ‘civilization’.

In his conception of the motive force of history Lavrov consciously avoided reductionism. Social evolution, he thought, was stimulated by the individual’s diverse needs. Among biological needs the most important was the need for food, which stimulated economic development. But apart from biological needs, characterizing humanity as an animal species, there were also specifically historical needs, constituting the historical dimension of human existence. The most important of these was the disinterested ‘need for development’ typical of ‘critically thinking individuals’. Lavrov believed that this need was becoming more and more important.

Lavrov’s overall conception of social evolution was tied to the Saint-Simonian and Comtian notion of history as a succession of ‘organic’ and ’critical’ phases. Historical progress moved along a spiral in an accelerating rhythm - its successive stages were growing shorter, and the difference between organic and critical periods was constantly diminishing. This was because historical evolution offered growing opportunities to achieve a harmonious fusion of solidarity and development, order and progress.

Regardless of their scholarly value, Lavrov’s sociological conceptions are of great interest as a historical document. They reveal Lavrov as an ideologist of the intelligentsia, the author of an extremely influential and truly classical conception of the intelligentsia’s historical mission. Some scholars have suggested that his writings were an expression of a specific ’intellectual aristocratism’, or even of certain characteristic aspects of the gentry mentality (the view of the masses as an inert herd, combined with a sense of guilt). Such views were being formulated even within the populist movement; their most extreme exponent was Pëtr Tkachëv (1844-86), a ’Jacobin’ populist, who accused Lavrov of betraying egalitarianism and paving the way for the transformation of the intelligentsia into a new ruling class. In fact, however, Lavrov consciously distanced himself from the theories of intelligentsia as a ’new class’, such as the Comtian theory of an intellectual elite that would govern the hierarchically stratified society of the future. For him the intelligentsia was first and foremost the *moral* elite, the conscience of society, not an aristocracy of intellect; intellectually bright and educated people who were selfishly indifferent to the fate of the suffering masses were defined by him as ’cultural savages’, not members of the intelligentsia. He was therefore a leading theorist and spokesman of the intelligentsia as an ethical category, that is, in the nineteenth-century Russian meaning of the term.

See also: Mikhailovskii, N.K.; Positivism in the social sciences; Positivism, Russian

ANDRZEJ WALICKI

List of works

(1860) *Ocherki voprosov prakticheskoi filosofii (Sketches in the Domain of Practical Philosophy)*, St Petersburg: Glazunov.(Lavrov’s first book, in which he defined his position as ‘anthropologism’.)
(1870) *Istoricheskie pis’ma*; trans. and introduced by J.P. Scanlan, *Historical Letters*, Berkeley, CA, 1967. (Lavrov’s best-known work, first published under the pseudonym P.L. Mirtov.)
(1965) *Filosofiia i sotsiologiia: izbrannye proizvedeniia*, Moscow: Mysl’, 2 vols.(Provides a representative selection of Lavrov’s philosophical and sociological works in Russian.)

References and further reading

Copleston, F.C. (1986) *Philosophy in Russia. From Herzen to Lenin and Berdyaev*, Kent: Search Press and Notre Dame, IN: University of Notre Dame, ch. 6.

Hecker, J.F. (1915) *Russian Sociology: A Contribution to the History of Sociological Thought and Theory*, New York.(A good outline of Lavrov's philosophical sociology.)

Kareev, N.I. (1907) *Teoriia lichnosti P.L. Lavrova (P.L. Lavrov's Theory of Personality)*, St Petersburg.(The classic work on Lavrov's philosophy of man from pre-Revolutionary Russia.)

Ladokha, G. (1927) 'Istoricheskie i sotsiologicheskie vozzreniia P.L. Lavrova', in M.N. Pokrovskii (ed.), *Russkaia istoricheskaia literatura v klassovom osveshchenii*, Moscow.(An interesting Marxist analysis of Lavrov's sociology and philosophy of history.)

Pomper, P. (1972) *Peter Lavrov and the Russian Revolutionary Movement*, Chicago.

Sorokin, P.A. (1922) 'Osnovnye problemy sotsiologii P.L. Lavrova' (Main Problems in Lavrov's Sociology), in *P.L. Lavrov. Stat'i, vospominaniia, materialy*, Petrograd: Kolos.(A detailed analysis.)

Walicki, A. (1979) *A History of Russian Thought From the Enlightenment to Marxism*, Stanford, CA: Stanford University Press, ch. 12; Oxford: Clarendon Press, 1980.(Provides a general outline of Lavrov's philosophy and sociological theory.)

Law and morality

Within the tradition of natural law thinking which finds its roots in the philosophies of Aristotle and Aquinas, the political community has generally been understood in terms of a fundamental goal: that of fostering the ethical good of citizens. Law, on this conception, should seek to inculcate habits of good conduct, and should support a social environment which will encourage citizens to pursue worthy goals, and to lead valuable lives. Pragmatic considerations may sometimes suggest the wisdom of restraint in the pursuit of these goals, and citizens may therefore, on appropriate occasions, be left free to indulge depraved tastes or otherwise fall short of acceptable standards. Such pragmatic arguments for the freedom to engage in vice, however, do not call into question the legitimacy of the state's concern with individual morality.

By contrast the liberal tradition has tended to place constraints of principle upon the scope and aims of the law. The most influential such attempt was J.S. Mill's advocacy of 'the harm principle': that the law may forbid only such behaviour as is liable to cause harm to persons other than the agent. Many difficulties surround this and other, more recent, attempts to formulate and defend constraining principles. For instance, should one take into account only the immediate effects of behaviour, or more remote and diffuse effects as well? Thus it is argued that immoral behaviour which in the short term 'harms nobody' may, in the long run, lead to a decline in morality in society at large and thereby to diffuse harmful effects.

1 Two traditions

The attempt to delimit principles which confine and constrain the proper scope of legal and governmental interference is inseparable from the liberal tradition: one need only think of Locke's attempt to define such a sphere of legitimacy by reference to a body of natural rights. Yet the attempt to delineate such principles has not always focused in a clear and circumscribed way upon the particular problem of an allegedly private sphere of morality. Thus when Locke, in his *Letter Concerning Toleration* (1689), seeks to defend religious toleration, he seems to do so not by reliance on his theory of natural rights and social contract, but by means of a pragmatic argument which would have sat comfortably within the assumptions of the Thomist natural law tradition: intolerance is, for Locke, pointlessly ineffectual rather than a violation of a distinct constraining political principle (see Locke, J. §1). Debate began to focus clearly on the idea of constraining principles precluding the state from interference with private immorality only in consequence of essays such as Wilhelm von Humboldt's *The Sphere and Duties of Government* (published posthumously in 1852) and J.S. Mill's *On Liberty* (1859) (see §2 below). The latter work has exerted the most fundamental influence in English-speaking countries.

Before we examine Mill's claims, however, it is worth making a couple of preliminary points. In particular, we should reflect upon the diverse motivations which may underpin rejection of the Aristotelian tradition of political thought, and may lead to the search for principles constraining the community's entitlement to enforce 'private' moral standards. It is often thought that the liberal is committed to a 'subjectivist' or noncognitivist understanding of moral judgment, and that the liberal concern for tolerance is simply a consequence of such subjectivism (see Moral judgment §1). Such an interpretation must, however, be a mistake: for, if moral judgments are simply expressions of subjective emotion or attitude, this must apply to moral judgments about the importance of tolerance along with all other moral judgments. A better reading of liberalism would see it as flowing from the inability to reach *agreement* on moral questions, rather than from the supposedly 'subjective' character of moral judgments. This reading does not succumb to the self-subverting character of an appeal to subjectivism, but it does have its problems. In some versions, for example, it can seem to imply that the need for consensus among those who disagree is the rationale for restricting the state's role; but this seems to make the argument for tolerance dependent upon the existence of a balance of power between contending factions.

Finally, one popular line of defence claims that liberalism is reliant neither upon moral subjectivism nor upon the inconclusive nature of moral argument, but upon the value of autonomous choice: it is good that people choose their own projects and lifestyles, even when they choose degrading or unworthy options (see Autonomy, ethical). The problem posed by this approach is that it is not wholly clear that the approach can be contrasted in any very fruitful way with the concerns of the Aristotelian tradition. After all, the Aristotelian might claim that *being freely chosen* is an essential condition of a good life, so that the state's concern to encourage good lives itself dictates a

concern to protect autonomy. Viewed from this perspective, the debate with liberalism would seem to be a pragmatic argument within the parameters of the Aristotelian tradition.

Such a reconceptualization of the terrain might well be viewed as damaging by many liberals. For many of the critics of Mill and of more recent liberals have directed their fire at the very idea of abstract principles constraining legitimate interference with a realm of supposedly ‘private’ conduct. They have been less concerned to oppose the concrete applications of these principles advocated by the liberals themselves. A characteristic example is James Fitzjames Stephen, who, in his book *Liberty, Equality, Fraternity*, wrote as follows:

> I object rather to Mr. Mill’s theory than to his practical conclusions…. The objection which I make to most of his statements on the subject is, that in order to justify in practice what might be justified on narrow and special grounds, he lays down a theory incorrect in itself and tending to confirm views which might become practically mischievous.
>
> (1873: 74)

2 Mill’s ‘harm principle’

Mill proposes what he describes as a ‘very simple principle’ as being ‘entitled to govern absolutely the dealings of society with the individual in the way of compulsion and control’. The principle asserted that ‘The only purpose for which power can rightfully be exercised over any member of a civilised community, against his will, is to prevent harm to others’ ([1859] 1956: 13).

Mill’s ’harm principle’ has sometimes been taken to suggest an atomistic view of society such that there is an area of private conduct which does not impinge on others at all, and which falls outside the law’s proper domain for that reason. As his critics were quick to point out, only the most trivial actions are devoid of effects upon others. In spite of the encouragement that Mill gives to this interpretation in some of his language, however, he explicitly conceded that ‘self-regarding’ actions may nevertheless affect other people. His principle is not intended to demarcate an area of conduct that is beyond the law’s remit, so much as a type of reason to which the community should restrict itself: in considering whether an action should be prohibited, only *some* of the effects of an action should be taken into account. In particular, the action’s effects in ‘harming’ others may be taken account of, while the effects upon the actor himself must be disregarded, as must the disapproval of unaffected third parties who consider the act immoral (see Paternalism §2).

In other ways, however, the suggestion that Mill adopts an unduly ‘atomistic’ view might seem to have some validity. For, if the harm principle is construed as prescribing a focus on fairly immediate and individuated harms, it may lead us to neglect the importance of sustaining social institutions that may be undermined in more oblique ways by the law’s failure to uphold conventional moral standards. One reply to this line of attack might be that nothing in the harm principle compels such a narrowly individuated focus: if the erosion of some social institution *is* a probable consequence of liberalization in an area of conduct, and if that erosion would be harmful, the harm principle legitimates our taking such considerations into account. The reply is problematic, however, in so far as a harm principle which invites us to take account of quite remote and diffuse effects threatens to prove empty and insubstantial in application. Mill himself is quite ready to accept that highly diffuse effects upon others can constitute ‘harm’ (he is, for example, critical on this ground of the unrestricted right to procreate) but the concession does represent an erosion of the integrity of the principle.

A number of other serious problems are associated with the harm principle. One question concerns the compatibility of the principle with Mill’s commitment to utilitarianism (see Utilitarianism). For should not a utilitarian legislator take account of *all* of the effects of an action, before deciding upon its prohibition? If some people suffer deep unhappiness at the very thought of acts of which they disapprove being performed in private, how can this unhappiness be justifiably *ignored* from the viewpoint of utility? It has been suggested that a utilitarian legislator should disregard such unhappiness in so far as it flows from the adoption of nonutilitarian moral views. But even if this argument could be sustained, it would not go far enough: I may be upset by the thought of acts which I consider simply disgusting rather than immoral; and, in any case, the exclusion of harm to the actor himself remains unexplained by this approach.

It is likely that Mill saw the harm principle as an intermediate maxim, by which utilitarians should regulate their

conduct, with the ultimate objective of advancing overall happiness in the long run. Such a view avoids inconsistency, however, only at the price of rendering the entire case dependent on largely empirical claims about the long-term effects of individual liberty. Mill's defence of the harm principle could in this way be thought to manifest an undue and ungrounded optimism. Given the aggregative conception of the common good espoused by utilitarianism, it would be remarkably fortunate if the best way of advancing the collective welfare was invariably to protect individual freedoms.

Quite apart from the compatibility or otherwise of the harm principle with Mill's utilitarianism, however, is the question of the meaning of 'harm' itself. For it has been repeatedly and justifiably pointed out that the question of what constitutes harm is an evaluative question which cannot in the end be separated from our wider ethical beliefs. Attempts, such as that of Joel Feinberg (1984), to analyse harm in terms of setbacks to interests simply reformulate the problem as one concerning the nature and content of our interests. Yet, without some morally neutral account of harm, Mill's principle seems to do little to exclude criminal prohibitions which are simply based upon moral disapproval of the actor's conduct. Consider, for example, the harm principle in its application to the debate about pornography. One issue concerns the suggestion that pornography fosters a social environment within which violent crimes against women are more likely: this argument appeals to an uncontroversial instance of harm, but makes the issue depend upon highly contestable empirical theses about the effects of pornography. What then if someone suggests that women are harmed *intrinsically* by pornography, in the sense that all women are harmed when some women are depicted as being objects available for sexual gratification? This suggestion employs a notion of harm which might not be universally accepted, but it challenges us to articulate criteria for what is to be regarded as harm (see Pornography).

A relevant suggestion has been made (in a slightly different context) by Brian Barry. He suggests (1995) that 'harm' might be defined in terms of what the great majority of people, having divergent conceptions of the good, would nevertheless agree upon as 'bad'. The suggestion builds upon the sound insight that people may have divergent ethical conceptions which nevertheless have an extensive area of overlap; but in requiring something short of complete universality and unanimity, Barry's approach seems in danger of making the scope of the harm principle itself depend upon majoritarian politics of a kind that Mill was concerned to constrain.

The above arguments may or may not be fatal to Mill's position. They do not render the harm principle wholly vacuous, because that principle serves, at a minimum, to direct our attention to an action's effects on others, rather than on the actor himself; but, if sound, the arguments dramatically reduce the value of the principle as a limitation upon the proper scope of the state's coercive power.

Even the requirement that we should have regard only to effects upon others may be a less substantial constraint than it seems, for we must remember Mill's acceptance of diffuse effects as 'harm'. Acts which are widely considered to be immoral may, in the first instance, affect only the actor himself; but, by creating a communal environment in which such acts are tolerated, they may make it harder (for example) to educate children into moral standards of conduct. When combined with the impossibility of offering a morally neutral account of the nature of 'harm', such diffuse effects seem to reduce the significance of Mill's principle to vanishing point.

3 The Hart/Devlin debate

A concern for such diffuse effects of immoral conduct lies at the bottom of an argument presented by Lord Devlin in 1959, in a British Academy Maccabean Lecture in Jurisprudence. Lord Devlin was then a judge of the Queen's Bench Division of the High Court, becoming a Lord of Appeal in 1961. His Maccabean Lecture arose out of the Wolfenden Report on Homosexual Offences and Prostitution, which had been published in 1957, and which had articulated the principle that 'private immorality' should not be the concern of the criminal law. Devlin criticized this approach, on the ground that society depends for its survival upon the existence of a shared morality, and that it was therefore in principle possible for grossly immoral conduct to threaten the survival of the society. Consequently, society has a right to enforce its shared morality as a measure of self-protection. Devlin was concerned, not with the truth or falsehood of the relevant moral views, but solely with the fact that they were widely shared and fundamental to the society's stability (Devlin (1965).

H.L.A. Hart responded to Devlin's assault by pointing to the absence of empirical support for Devlin's claims that immorality could lead to social breakdown (Hart 1963). But it has more recently been suggested (George 1993)

that Devlin had in mind not a breakdown in social order so much as a loss of social cohesion, where people relate to each other purely on a basis of self-interest rather than on a basis of moral principle. If this interpretation is correct, it would make Devlin's claim both more plausible and less open to empirical refutation or confirmation.

Devlin acknowledged a debt to the Victorian judge and jurist James Fitzjames Stephen, whose book *Liberty, Equality, Fraternity* 1873) was a forceful attack upon Mill's theory. Stephen's approach was utilitarian in character, but he regarded Mill as having departed from the consistent utilitarianism of Bentham and James Mill. Hart described the works of Stephen and Devlin as revealing 'the outlook characteristic of the English judiciary' (Hart 1963).

4 Dworkin on 'external preferences'

The 'goal-based' approach of utilitarianism is rejected by Ronald Dworkin in favour of what he calls a 'rights-based' approach; but, in reality, there are good grounds for considering his position to be a modified form of utilitarianism, rather than a radical rejection of that approach. Dworkin takes the view that, in so far as utilitarianism possesses any genuine moral appeal, that appeal is a consequence of the extent to which it expresses our belief in a 'right to equal concern and respect'. He argues, however, that the notion of equal concern and respect receives inadequate expression within utilitarianism to the extent that the utilitarian takes account not only of 'personal preferences' (preferences for what I myself do and receive) but also of 'external preferences' (my preferences about what others do or receive). A proper concern for equality would exclude the influence of external preferences by recognizing a 'right to moral independence': the right not to suffer disadvantage simply on the ground that others consider one's conceptions of a good life to be ignoble or wrong (Dworkin 1985).

Dworkin believes that rights are best understood as 'trumps' over collective goals, and that, consequently, their philosophical justification is in part relative to the goals which have been adopted. The problem with taking account of external preferences (if a utilitarian goal has been adopted) is that the external preferences 'purport to occupy the same space' as the utilitarian theory itself: utilitarianism must be neutral between pushpin and poetry, but it cannot be neutral between itself and Nazism (for example) (Dworkin 1985: 363).

Dworkin insists that this argument holds even when the external preferences are not based on any general moral or political theory: they still 'invade the space claimed by neutral utilitarianism'. It is difficult to see how this argument can be sustained, however. One cannot argue, for example, that a preference 'invades the space of utilitarianism' simply because, if fully satisfied, it would necessitate a distribution not recommended by utilitarianism. Such an approach would prescribe a decision procedure which was viciously circular: for one would have to decide upon the requirements of utility before one could say whether any particular preference should be taken into account in calculating those requirements.

In any case, could it not be said that Dworkin's approach fails to show equal concern for those whose wellbeing or utility is a function of their external preferences? Suppose that my sole concern is your welfare, and my life will be a failure if your happiness is not secure. How can equality of concern dictate that my wellbeing should be disregarded?

Many will consider that the proposed 'right to moral independence' is more attractive than is the justification which Dworkin proposes for it; but then, few if any theorists have argued that people should be disadvantaged simply in consequence of a widely held disapproval of their lifestyles, so perhaps the 'right' is not very substantial after all.

See also: Law, limits of §4; Liberalism; Natural law; Privacy; Rule of law (Rechtsstaat); Toleration

N.E. SIMMONDS

References and further reading

Barry, B. (1995) *Justice as Impartiality*, Oxford: Oxford University Press.(An intelligent development of some recent liberal political theory.)

Devlin, P. (1965) *The Enforcement of Morals*, Oxford: Oxford University Press.(A vigorously argued and provocative work.)

Dworkin, R. (1985) *A Matter of Principle*, Cambridge, MA: Harvard University Press.(Parts Three and Six are particularly helpful.)

Feinberg, J. (1984) *Harm to Others*, Oxford: Oxford University Press.(Feinberg attempts to explain 'harm' in terms of setbacks to 'interests'.)

(1990) *Harmless Wrongdoing*, Oxford: Oxford University Press.(A clear and helpful survey of many of the basic issues.)

George, R.P. (1993) *Making Men Moral*, Oxford: Oxford University Press.(A clear and accessible discussion, which could form a good starting point for the beginner, in spite of being written from a standpoint which is quite critical of liberalism.)

Hart, H.L.A. (1963) *Law, Liberty, Morality*, Oxford: Oxford University Press.(A work of elegance and clarity that repays detailed study.)

Locke, J. (1689) *A Letter Concerning Toleration*, J. Horton and S. Mendys (eds), London: Routledge, 1991.(A classic text in the early development of liberal thought.)

Mill, J.S. (1859) *On Liberty*, Indianapolis, IN: Bobbs-Merrill, 1956.(A classic work which has shaped the character of modern liberalism.)

Stephen, J.F. (1873) *Liberty, Equality, Fraternity*, intro. by R. Posner, Chicago, IL: University of Chicago Press, 1991.(A robust and colourful period piece.)

Law and ritual in Chinese philosophy

The contrast between li, conventionally translated as 'rites' or 'rituals', and fa, conventionally translated as 'law', marks a distinction in Chinese political theory as to the nature of political order and the preferred means of achieving such order. Lizhi, traditionally associated with Confucianism, refers to political order predicated on and achieved primarily by reference to the li or 'rites', that is, traditional customs, mores and norms. In contrast, fazhi, associated with Legalism, refers to political order attained primarily through reliance on fa or 'laws', that is, publicly promulgated, codified standards of general applicability backed up by the coercive power of the state.

The tension between these two dominant strategies for achieving social and political order - lizhi and fazhi - is a theme that began in the classical tradition and has persisted down to the present day, even though the understanding of li and fa and the relation between them has changed and evolved over the years. For example, some thinkers saw li as context-specific, flexible norms or standards of a particular culture. Others saw li as more permanent, general standards. Some objected to the codification of li and fa; others favoured it. Some saw li and fa as interdependent, equally important and mutually reinforcing; others relied predominantly on one to the detriment, if not total exclusion, of the other.

Accordingly, it is not possible to speak of a single conception of li (lizhi) or fa (fazhi). Nevertheless, one must extract or highlight certain general or dominant features that tend to characterize lizhi and fazhi and to distinguish the one from the other, if one is to make use of this important contrast as a hermeneutical tool for interpreting, explaining and understanding a central debate in Chinese political philosophy. In reflecting on the importance of the lizhi versus fazhi distinction, five salient points emerge.

First, advocates of lizhi have tended to favour less formal means of conflict resolution than advocates of fazhi. The former believe that informal methods foster more particularized justice; the latter believe such methods provide excessive discretionary authority to those in power and thus foster abuse. Accordingly, publicly promulgated rules or laws are necessary.

Second, and a corollary to the first, fa or 'laws' are more formal norms of greater general applicability than the li, which in most instances consist of the web of informal context- and culture-specific rules that provide guidance for appropriate civil interaction in daily life among the various members of a certain community. Fazhi, therefore, is more of an externally imposed order that requires of the individual compliance more than participation. In contrast, lizhi proponents view order as emerging out of a particular context, a particular community of people. Society is what those who comprise it choose and make it to be. Participation rather than mere compliance is central.

Third, and again a corollary, lizhi and fazhi mark a contrast in the goals and aspirations of society. Advocates of lizhi tend to be more optimistic about human kind and the possibility of achieving a harmonious social order in which each person is able to find their place and play their chosen role. They tend to see society, social order, humanity, as an achievement. If human beings are willing to put aside narrowly selfish and provincial concerns, if they are willing to defer to the excellence of others and to cooperate in the project of creating a harmonious social order in which each has a place, then such an order is possible. In the process, one becomes a better person. Put differently, one's potential for individual and personal growth is inextricably bound up with the fate of society.

In contrast, many of the fazhi persuasion take a dimmer view of human nature. Humans are by and large self-interested beings. Left to their own devices, the strong will exploit the weak; the powerful will abuse the powerless. Impartial rules are necessary to limit what one person can do to another. Laws provide this minimum layer of protection, this floor below which society cannot sink and remain a peaceful, and arguably just, society. Of course law may play a more positive role as well; recourse to general laws may simply be an efficient way to ensure an equitable distribution of social resources, for example. At the end of the day, however, advocates of fazhi have been primarily concerned with the use of fa as a corrective to the abuses and weaknesses of li. Thus law's role in ensuring a minimal floor of protection, a base equality, has been underscored, explaining in part the rather negative image of law in China through the ages. In contrast, the positive side of law, the empowering aspect of rights, the ability to use law as a vehicle for social change, has been central to many Westerners' conception of law.

Fourth, as might perhaps be expected, there has been no clear victor in the lizhi *versus* fazhi *controversy. No system is perfect; no one has yet achieved the perfect means for realizing social order. Many of the same issues confront every society: how to achieve fairness and justice for all while at the same time recognizing the uniqueness of each person and each situation, and thereby achieve justice for each; how to provide those in power the necessary discretionary authority to mete out a particularized justice without falling prey to personal prejudice, bias, corruption and abuse; how to ensure that individuals have sufficient rights to shield themselves against each other and the state, and yet simultaneously encourage people not to wield those rights as weapons in endless litigation aimed at maximizing one's own personal interests to the detriment of society's interests; how to ensure that law will be a tool of the disempowered to effect social change and not simply a tool for exploitation and legitimation by those in power; how to implement a formal rule of law with institutional integrity and independence, overseen by professionals, and yet prevent the institution from becoming overly bureaucratic and law the esoteric province of the professional lawyer. The* li *versus* fa *distinction, although not a perfect fit, is useful as a way to organize Chinese thinking on, and approaches to, these and similar issues.*

A fifth and final introductory point is that both lizhi *and* fazhi*, as they have developed in China, have failed in two important respects. First, they have failed to provide effective restraints on the power of the ruler, especially institutional restraints. Second, they have failed to adequately address the need to protect the individual against the state. Again, this failure is largely institutional, although like the failure to adequately restrain the power of the ruler, it can also be traced back to certain underlying philosophical assumptions common to both the* lizhi *and* fazhi *traditions. These assumptions include the rejection of three key assumptions of the western liberal tradition: (1) that someone to treat someone with respect and as one's equal requires that one refrain from imposing one's view on them (the normative equality premise); (2) that each person knows what is best for themselves and/or people reasonably disagree about what constitutes the good (the epistemic equality premise); and (3) that the interests of the individual and state may not be, and arguably are not, reconcilable.*

1 Context

Chinese philosophical thinking has always been elitist. Few Chinese thinkers would accept the liberal assumption (the epistemic equality premise) that no person or group possesses superior moral insight. From the pre-historical mythical ancestors of China, Yao and Shun, to the Confucian sage-rulers of the Warring States period, to Mao Zedong and Deng Xiaoping in the twentieth century, Chinese leaders have been credited with an uncanny ability to fathom what is in the best interests of society and its individual members. Indeed, much of their authority to rule is predicated on their claim to special ethical insight and unique political knowledge of the way of rulership.

The rejection of the epistemic equality premise calls into question the normative equality premise (for liberals, the latter also follows as the conclusion of the former): in other words, that to treat someone with respect is to refrain from imposing one's normative views on them. In China, the government has pursued a substantive moral agenda defined in large part by the particular normative vision of the ruler. Chinese governments have been and continue to be paternalistic (much to the dismay of Americans who find, for example, Singapore's public flogging of juvenile delinquents and prohibitions on chewing gum objectionable). The image of the father dominates the political rhetoric of China. To be sure, the specifics of the image vary by school. The Confucian (*lizhi*) father-ruler is kind and compassionate and more of a facilitator of order than a dictator, whereas the Legalist father-ruler (*fazhi*) is the tough disciplinarian who well understands that to spare the rod is to spoil the child. Nevertheless, the image remains the same: the father, knowing what is best, takes care of his children.

Paternalism is on the whole antithetical to individual rights (see Paternalism). Not surprisingly, individual rights have not been and are not a prominent feature of the political landscape of China. There are of course many reasons for China's lack of a strong tradition of individual rights - which, we do well to recall, is a singular phenomenon, until this century virtually unique to the Enlightenment and post-Enlightenment West. One of the reasons for the lack of a strong rights tradition in China is the assumption that the interests of the state and the individual can be brought into harmony, an assumption particularly prevalent among *lizhi* thinkers but also present in many *fazhi* proponents. Given this assumption, it is suggested there is no need for rights to protect the individual against the state.

Ironically, even among contemporary Chinese rights advocates, few have taken note of, much less celebrated, the anti-majoritarian role of rights as 'trumps' on the collective good and the interests of society as a whole. Indeed, it

is not at all clear that even the most ardent Chinese rights advocates would agree with Ronald Dworkin that individual rights trump the overall social good or with John Rawls that ‘each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override.’ In fact, it is not at all clear that Chinese rights thinkers conceive of rights as anything other than a kind of interest, to be weighed against other interests, including the interests of the majority, the collective, society and the nation. In contrast, most Western rights proponents conceive of rights as moral ends in themselves, deontic in character, and thus different in kind from interests, which are consequentialist or utilitarian in character (see Rights). The difference between rights understood instrumentally as interests versus rights as trumps, moral principles or ends in themselves is not only theoretically fundamental but of utmost practical importance. Translating rights into the language of interests generally has the effect of decreasing the likelihood that rights will be taken seriously. When weighed against the interests of society as a whole, or against the state, the interests of society and the state will usually prevail. This is probably all the more likely in China, given the historical tendency to privilege the interests of the group and the state over the interests of the individual when those interests conflict.

The *li* and *fa* systems as developed to date in China have yet to adequately address the twin issues of restraints on the ruler and individual rights (which set restraints on the state and the majority *vis-à-vis* the individual). Arguably, *fazhi*, with its emphasis on impartial laws of general applicability, provides a more likely source for both effective restraints on the those in power and effective individual rights. Thus it is not surprising that one increasingly sees raised the banner of *fazhi*, which increasingly represents a rule of law along Western lines, with an institutionally independent and professionalized judiciary assigned the role of checking the legislative and executive branches and safeguarding individual rights. Nevertheless, the weight of the past continues to be felt. It is important therefore to understand the *li-fa* debate, and to understand the *li-fa* debate, one must return to the source.

2 Confucius’ classical conception of *lizhi*

The *locus classicus* for *lizhi* is the *Analects* of Confucius:

> Lead the people with government regulations and organize them with penal law (*xing*), and they will avoid punishments but will be without shame. Lead them with virtue and organize them through the *li*, and the people will have a sense of shame and moreover will become humane people of good character.
>
> (*Analects*)

Confucius maintained that to rely solely or even predominantly on law to achieve social order was folly. Laws, backed up by punishments, may induce compliance in the external behaviour of individuals, but they are powerless to transform the inner character of the members of society.

Confucius’ goal was not simply a stable political order in which everyone coexists in relative harmony and isolation, with each afraid to interfere with the other for fear of legal punishment. Rather, Confucius set his sights considerably higher. He sought to achieve a harmonious social order in which each person is able to realize their full potential as a human being through mutually beneficial relations with others.

For Confucius, society is the medium through which one becomes a human being. That is, one becomes a human being, a humane person, by virtue of participation in society. Personhood and humanity are functions of socialization. At birth, before the process of enculturation, of becoming humane, humans are not different from the other beasts. If one does not or is not willing to participate in society and enter into harmonious relations with others, one remains at the level of a beast; one is human *qua* member of a biological species. By turning their back on society, the bestial person (literally *xiaoren*, ‘small person’) fails to utilize the cognitive, aesthetic and spiritual powers which distinguish humans from other species. It is just the engagement of these capacities in joining with others to shape a new world, to create a significantly different and better society and to overcome one’s natural conditions in achieving innovative resolutions to conflicts, that is distinctive about humans. If one cannot overcome the passions, instincts and desires that one shares with other beasts, if one wars against all and resorts to violence and brute strength to fulfil one’s narcissistic wants, then one fails to become truly human, to achieve humanity.

Of course there will always be those who turn their back on society, who refuse to participate in creating a harmonious order and insist on pursuing their narrow self-interest in any manner possible. For such people, law

and punishments are necessary. Confucius, ever practical, did not advocate the complete abandonment of laws. Nevertheless, the goal is to foster an environment in which laws need to be imposed as little as possible: 'In hearing litigation I am much the same as anyone. If you insist on a difference, it is perhaps that I try to get the parties not to resort to litigation.' For Confucius, the ethical and political challenge is to inspire in the many members of society the desire to achieve a humane society and to encourage them to direct their energies toward the attainment of a harmonious social order. What is being required is a willingness to participate in collective living, to search for a cooperative solution, to become humane (*ren*) (see Confucian philosophy, Chinese §5).

For Confucius, the codification and public dissemination of laws sends the wrong kind of message. Laws are designed to protect the minimum interests of the members of society and to provide a mechanism for dealing with and removing those individuals who are not only unwilling to participate in fostering a harmonious social order, but whose behaviour threatens the well-being of others and the ability of society to function. Making the laws public focuses attention, not on the achievement of the highest quality of social harmony possible, but on the lowest level of participation required by society. Consequently, it may encourage some persons to look only to manipulate the system for their own advantage. Thus we find Confucius criticizing the state of Jin for publicly promulgating laws: 'Jin is going to ruin. It has lost its proper rules…. But now when those rules are abandoned and tripods with penal laws on them are cast instead, the people will study the tripods, and not care to honour men of rank.' The proto-Confucian *Shu Xiang* adds:

> In antiquity, the former kings considered the particular circumstances in regulating affairs. They did not make public general laws of punishments and penalties, fearing that this would foster a contentious attitude among the people that could not be stopped or controlled. For this reason, they used their discretionary judgment (*yi*) to keep the people in bounds…and guided them in their behaviour through the rites (*li*)…
>
> (*Shu Xiang*)

Confucius rejects law as a means for attaining social order because law focuses on external compliance. Since one is merely expected to conform one's behaviour to the given legal norm, one is denied the opportunity to fully participate in the creation of a social order more reflective of one's individual character. Laws, as standards of general applicability, do not allow for sufficient individual expression or particularity.

The formal character of a legal proceeding further diminishes the opportunity for a more contextualized justice able to account for the particular circumstances of the individuals involved. By its very nature, a formal legal system elevates procedural justice relative to, if not at the expense of, substantive justice. In fact, one of the motivations for the development of a formal legal system is to provide a procedural means of resolving interpersonal conflicts that cannot be resolved on a more informal, personal level. While substantive justice has always remained the primary goal of the legal system in China, the formal character of the process and the emphasis on predetermined procedures for resolving conflict have often been seen as obstacles to a more personalized and creative approach, to interpersonal conflict.

Confucius' politics of harmony require the voluntary participation of the individuals who collectively comprise society. If conflicts arise, as they inevitably will, each person must evidence a willingness to look for a mutually acceptable solution. Of course a willingness to cooperate is not enough to overcome all conflicts. There must also be sufficient common ground to provide a basis for discussion, understanding and potential mutually agreeable solutions. The *li* provide this essential communally owned repository of shared meaning and value on which to draw in times of conflict. One is inextricably a part of one's tradition. However different one may be from one's neighbour, there are still deep chords of affinity that bind one to other members of one's community. By tapping the areas of commonality, one may be able to find the ground upon which to build a consensus to forge a new harmony. It is the *li* that provide this common foundation for Confucius, and it is in this sense that the *li* are, in the apt description of the contemporary philosopher A.S. Cua, mutual accommodations of differing attitudes, beliefs and values in social intercourse (see Confucian philosophy, Chinese).

The *li* have often been construed as universal ethical principles. As a result, Confucius' *lizhi* has been depicted as a kind of natural law. However, the *li* are better understood as customary norms that gain favour within a particular historical tradition at a particular time and that constitute not unchanging, determinant rules of behaviour but culturally valued, though negotiable, guidelines for achieving harmony in a particular context. There is nothing sacred about the *li* in the manner of the Ten Commandments; the *li* are merely culturally and historically

contingent norms. As such, to depict them as eternal, universal norms is to render them unduly static and determinate.

Although the *li* derive their normative force in part from the fact that they have withstood the test of time and as such represent the amassed wisdom of the ages, they nevertheless must be interpreted in the light of present circumstances to retain their currency and relevance in shaping behaviour. If the *li* are not invested with new meaning and value, reinterpreted in the light of current circumstances, they will degenerate into irrelevant and trivial formal rules of etiquette.

Traditionally, Confucians were members of the *shi* (literati) class, charged with the responsibility of interpreting the *li* and making it applicable to the times. When called upon to resolve conflicts, they attempted to interpret and apply the *li* in such a way as to give effect to a particularized justice that was amenable to all parties and thus restored harmony. Accordingly, they had a vested interest in the *li* and the informal system of mediating conflict that relied heavily on their individual wisdom and judgment. In a system without formal means of appeal, each individual charged with mediation wielded considerable power, a power that could be abused in the absence of the kinds of procedural protections afforded by a formal legal system. On a grander sociopolitical level, the ruler was charged with generating and maintaining social order. Living as he did during the Warring State period, Confucius realized that society was comprised of persons with diverse interests. To bring diverse interests into harmony requires a unifying agent. The ruler was the unifying agent - the Pole Star - responsible for providing the pivotal note.

Yet the Confucian ruler was more of a facilitator of order than a dictator - at least in theory. Using persuasion rather than force, he was to inspire in others a willingness to become humane, to put aside narrow self-interest and form a harmonious society. The ruler was to lead by example. His virtue was to sweep over the people and transform them just as the wind blowing over long stalks of grass bends them as it passes. Ideally, the ruler would not have to impose his way; rather, the people would naturally defer to his example, to his superior moral cultivation. With everyone participating willingly in the collective project of creating a humane society, social order was sure to result. There was no need for the heavy hand of the law. The ruler merely set himself aright, assumed his position at the centre, and let society follow his lead. This was the ideal of the *wuwei* ruler: ‘The Master said: "If there ever was a ruler who could be said to have achieved proper order while refraining from actively imposing his way on others (*wuwei*), it was Shun. What was there for him to do? He simply made himself respectful and took up his position facing due south."’

While the ideal Confucian ruler at least in theory sought not to impose his way, he nevertheless played a pivotal role in setting a course for society, particularly in times of conflict or trouble. It was his ability to perceive creative solutions to the problems facing society that enabled him to lead (see Daoist philosophy §6).

In the end, the Confucian ruler ruled by virtue of his moral vision. His authority to lead was a function of his ability to know the way (*dao*). *Dao* refers to the patterns in the world, the ways in which things are related and the possible ways in which they might be related (see Dao). To ‘know *dao*’ is to perceive these patterns and relations. As a moral achievement, knowing the way requires that one overcome the limited perspective of one’s narrow self-interest to see oneself in relation to others. The more truly cultivated one is, the more relations and possible relations one sees. For Confucius, knowing *dao* also meant understanding tradition and how the past informed and shaped the present. Accordingly, knowledge of the *li* was an essential component of the education of the ruling class.

Confucianism is, as noted above, decidedly elitist and, as a result, paternalistic. In theory, the ruler is the one able to see things other people do not, to see a way (*dao*) to bring harmony out of diversity, to turn disorder into order and to persuade others to join in the realization of that harmonious social order by virtue of his moral vision, character and the example he set. But vision is at bottom personal. As a consequence, political order was (and has remained) largely dependent of the quality of those in power.

The absence of effective institutional checks on the power of the ruler charged with achieving sociopolitical order on a macro level mirrored the absence of procedural constraints of the Confucian literati charged with resolving conflict on an interpersonal level. The combination of the lack of effective restraints and the elitist fallacy - that the Confucian literati/sage/ruler possessed superior normative insights - left the disempowered largely at the mercy of

those in power.

3 The evolution of *li*: increased codification and stratification

The hierarchical nature of society was further reinforced by the *li*, which reflected differences in personal cultivation and social accomplishment and in so doing inevitably fostered such differentiation. Confucius maintained that humans are by nature similar. Each person has the potential to become a sage. At the same time, however, he realized that not all people will become sages. As a practical matter, people will achieve different levels of cultivation and accomplishment. A properly functioning society recognizes excellence and harmoniously integrates the varying roles of its members. Thus it was essential for Confucius that each person fulfilled their role, as delineated by the *li*: a ruler was to act as a ruler, a minister as a minister, a father as a father, a son as a son, a friend as a friend and so on.

To be sure, neither one's role nor the rules (*li*) delineating one's role were predetermined for Confucius. One could escape one's class through personal cultivation. Indeed, as one moved through life, one's role naturally changed. First a son, one later becomes the father; first the subordinate, one later assumes positions of increasing authority. Further, as noted above, Confucius insisted that the *li* required personal investment, interpretation and adaptation, lest they become ossified and outdated rules of petty etiquette.

Confucius' words, however, were not heeded. Over time, the *li* became increasingly codified and rigid. One finds in many classic texts, the *Liji (Book of Rites)*, the *Xunzi*, even in the *Analects* of Confucius, painstakingly detailed expositions of the proper behaviour for the various strata: the ruler was to use a *jian* cup, the ministers a *jia* cup; a ruler used a decorated bow, a feudal lord a red bow, a high officer a black bow, and so on. More important than the exhaustive detail of these works, however, is the underlying theory. There is a marked change in emphasis from Confucius' own view of the *li* as flexible and adaptable and of society as a continually evolving, open-ended project. Increasingly, the primary function of the *li* is taken to be class differentiation. Thus we find the Confucian Xunzi asserting:

> The ancient kings invented the *li* and standards of rightness (*yi*) for humans in order to divide them; causing them to have the classes of noble and base, the disparity between elderly and young, the distinction between wise and stupid, the able and the incompetent; all to cause people to assume their duties and each to get his proper position.
>
> (*Xunzi*)

Similar sentiments were echoed in a wide variety of classical texts, from the *Liji* to the eclectic Daoist text the *Huainanzi* to the dynastic history of the Han, the *Hanshu (History of the Former Han Dynasty)* (see Huainanzi; Chinese Classics).

As the *li* became more rigid, so did the social classifications. Confucians were privileged in the social hierarchy; they wielded considerable power in their role as the mediators of conflict and as keepers and interpreters of the *li*. The establishment of the Imperial University in the Han, with an examination system based on the Confucian classics as the core curriculum, further solidified the role and position of the Confucian literati in the social hierarchy.

4 The Legalist response

Not surprisingly, there were those who objected to the power granted the literati (and the ruler) in a Confucian *li*-based order. To the Legalists, the Confucian system of *lizhi* was nothing more than 'rule of man' (*renzhi*). The Confucian sage determined what was best in a given situation based on his own judgment and interpretation of the *li* rather than by appeal to fixed standards or laws of general applicability. Accordingly, *lizhi* strengthened the hand of the elitist class by ceding to the literati the discretionary authority to interpret and apply the *li* as they saw fit. In response, the Legalists advocated clearly codified, publicly promulgated laws applicable to commoner and noble alike as a means of undermining the dual class system in which 'the *li* do not reach down to the common people; penal law does not reach up to the great official'.

While advocating the impartial application of publicly codified laws, Legalism was hardly a 'rule of law', which, as applied to contemporary legal systems, at minimum refers to the existence of an institutionally independent and

professionalized judiciary that interprets and applies laws construed as general commands, publicly known, clearly stated, legitimately enacted and altered, and backed up by the coercive power of the state (see Law, philosophy of). Accordingly, the Legalist *fazhi* is better characterized as rule by law. Law was simply a pragmatic tool for obtaining and maintaining political control and social order. In the Legalist view, humans are self-interested. To avoid conflict and achieve order, they must be manipulated through a reliable and impartial system of rewards and punishments. Clear, codified, public law lets every person know what is expected and what the consequences of their actions will be.

Law was also a pragmatic tool in the sense that the Legalists understood that reliance on the personal qualities and judgment of the ruler would in many instances, if not most, lead to disarray. According to Han Feizi, most rulers are simply average in their abilities: the truly exemplary ruler such as Yao or Shun is as rare as the truly evil and incompetent ruler such as Jie or Zhou. Thus the Legalists sought to design a system that would work even with - or despite - a ruler who is neither exceptionally bright, morally good or politically adept. To do this, they had to conceal the weaknesses of the ruler by erecting a screen of institutional mechanisms, political strategies and techniques (*shu*). Law was one such mechanism. Once the system of laws with their attendant punishments was made known, the laws were to be impartially applied. Neither the ruler nor the bureaucrat who applied the law on a daily basis was to allow personal bias or relationships to sway the outcome. In fact, Han Feizi encouraged the ruler to remain behind the scenes and allow his ministers to carry out the day to day functions of governance. This was the Legalist understanding of *wuwei* - literally, non-action - that is so central to Laozi's Daoist philosophy (see Daoist philosophy §6).

Nevertheless, the ruler remained the ultimate authority, both in theory and practice. In the final analysis, law was what pleased the ruler. Accordingly, the ruler retained the authority to promulgate and change laws. Further, although the Legalists advised the ruler against lightly changing the laws, they did allow that laws should be changed when their costs outweighed their benefits. In a direct attack on the importance invested by the Confucians in the *li*, the Legalists cautioned against undue reverence toward outdated norms.

In sum, Legalist law was positive law, not natural law. It was a pragmatic tool for effecting social order. While law was intended as a means to limit the arbitrary power of the ruling class and the ruler, Legalist *fazhi* or rule by law ultimately failed, as did Confucius' *lizhi* or rites-based order, to impose effective checks on the ruler. By allowing the ruler the final word, Legalism lent theoretical legitimacy to the excesses of the first Emperor of the Qin, the first emperor to unite China whose reign, marked by strict laws and draconian punishments, was short-lived.

5 Huang-Lao: a natural law alternative

There arose in response to the deficiencies of both Confucianism and Legalism a new school, the Huang-Lao school, which flourished during the early Han period (*circa* 200 BC). Huang-Lao attempted to limit the power of the ruler and provide a theoretical and moral foundation for a law-based rule (*fazhi*) by grounding the sociopolitical order in a normatively predetermined natural order. Ultimate authority lies not with the ruler but with the Way (*dao*). The Way/*dao* gives rise to or determines the laws (*dao sheng fa*); the ruler is merely the medium who by overcoming personal, subjective biases is able to apprehend the objectively given Way. Laws are therefore determinate normative standards to be discovered by the sage-ruler: 'There is a distinction between right and wrong: use the law to adjudicate between them. Being empty, tranquil and listening attentively, take the law as the tally.'

Just as the sage is responsible for discovering the law but not for creating it, so is he responsible for impartially applying the law but not for interpreting it. In the legal empire of Huang-Lao, the scales of justice are finely calibrated objective standards. Discretion is eliminated. It is neither the sage's duty nor his role to balance the arguments pro and con in light of the particular circumstances or the culturally and historically contingent norms, attitudes and beliefs of a particular community (in other words, the *li*). Unlike his Confucian (and Daoist) counterpart, the Huang-Lao sage is not called upon to build a consensus out of dissension, to realize a harmony amenable to all concerned parties. He is a judge, not a mediator. His task is to decide who is right and who is wrong, nothing more, nothing less. On one side of the scales of justice goes the deed, on the other the Way. The burden on the sage is to eliminate subjective bias so that he is able to apprehend the deed as it actually is and the Way as it should be.

Thus, like the Legalists and contra Confucius, Huang-Lao advocates an impartial application of publicly promulgated, codified laws of general applicability. However, unlike the Legalists, Huang-Lao attempts to constrain the power of the ruler. Whereas the Legalist ruler was the ultimate authority as to what the law is and how it should be interpreted and thus 'above the law', in the Huang-Lao universe the Way is the ultimate authority and thus the ruler, like all others, must abide by the laws. Accordingly, law for Huang-Lao is not merely a political tool to be used by the ruler to further his own ends. The ruler cannot change the law at will, nor can he circumvent it by issuing pardons.

In sum, whereas law for the Legalists was positive law, a pragmatic means of attaining social order, law for Huang-Lao is natural law, grounded in the normatively predetermined natural order or Way. By grounding law in the Way, Huang-Lao attempted to circumscribe the power of the ruler and thereby avoid the evils of the despotic first emperor of China. However, while there may have been limits on the ruler in theory, there were no such limits in practice. Given that there is no way to verify the ruler's claim to have discovered the Way and hence the correct laws, the ruler's power remains as unchecked in practice as that of either the Confucian or Legalist ruler.

6 The evolution of *li* and *fa*: Xunzi's synthesis

In the absence of a clear victor, the *lizhi-fazhi* debate evolved over time, surfacing in slightly altered forms, in different combinations. On a theoretical level, there have been attempts to synthesize *li* and *fa*, as in the writings of Xunzi.

Xunzi, like Hobbes, portrays the human condition as a struggle to overcome self-interest. In antiquity, people lived in a state of nature. Dominated by their inherent tendencies toward self-interestedness, they competed one with the other for the scarce goods available, giving rise to strife and social disorder. To rectify the chaos and set the social order aright required some system of laws or normative guidelines. In explaining the origin of this system, Xunzi turned to the mythical Ancient Kings (whereas Hobbes turned to the myth of the social contract):

> What is the origin of the *li*? I reply, man is born with desires. If his desires are not satisfied for him, he cannot but seek means to satisfy them himself. If there are no limits and degrees to his seeking, then he will inevitably fall to wrangling with other men. From wrangling comes disorder and from disorder comes exhaustion. The Ancient Kings hated such disorder, and therefore they established the *li* in order to curb it, to train men's desires and to provide for their satisfaction…. This is the origin of the *li*.
>
> (*Xunzi*)

Faced with divergent interests and conflicting claims among the members of society, Xunzi turns to the *li* as a mechanism for adjudication and social control. But the *li* do not function alone; they form part of a larger complex of *li yi fa du* (rites, normative judgment, law and statutes). To rule effectively requires use of both *li* and *fa*.

Xunzi differed from Confucius not only in the importance he attached to *fa* but in his conception of *li*. Xunzi saw the *li* as a means of differentiating the various classes or society not just socially but economically. That is, the *li* provided a principle for determining an equitable distribution of the scarce resources available to society: the nobility (including Confucian literati) received more than those at the bottom of the hierarchy. In contrast, Confucius emphasized the integrative, social aspect of the *li* rather than the divisive, economic aspect. The *li* made possible harmonious interaction among the various social groups by providing a common network of behavioural norms. Civil interaction was possible because all shared the *li*.

Furthermore, in synthesizing *li* and *fa*, Xunzi tended to remake the *li* in the image of *fa*. Xunzi's *li* are similar to *fa* in that they require less participation and more compliance than in the case of Confucius. *Li* for Xunzi are norms created and imposed by those in power on those without power, albeit with the interests of the disempowered in mind. The lower classes are not required to participate in generating the *li* or investing the *li* with personal meaning so much as they are required to comply. Hence one finds Xunzi placing great emphasis on moral training and habituation. The goal is to drill on the *li* until one can comply without reflection. To be sure, once one has thoroughly internalized the *li*, one may experiment with them and be flexible in their application. But for most mere compliance is all that is expected.

Another difference is that whereas Confucius' *li* were context-specific, Xunzi's *li* are less so. For Xunzi, the *li* arose as a response to the unchanging nature of humans as self-interested beings. In his view, the ancient sages

were able to take the measure of human nature and design the network of *li* accordingly.

Because human nature is relatively stable, so also are the *li*. This is not to say, however, that the *li* are universal, eternal truths. For Xunzi, the particular laws and *li* by which one adjudicates conflicts and achieves social order are human products, human artifacts. The process by which the sage generates the *li* is one of pragmatic experimentation: ‘The sage gathers together his thoughts and ideas, experiments with various forms of conscious activity, and so produces *li* and sets forth laws and regulations.’

Xunzi realized that circumstances change and that the *li* and laws must change as well. Nevertheless, he was concerned that there be some basic guidelines, a single agreed set of standards, laws and norms for guiding behaviour and resolving conflicts. To change the *li* and laws without regard to whether change is desirable would be both inefficient and unfair in that it would make it impossible for people to plan their futures. When essential, the *li* and laws could be changed. But given the constancy of human nature, there is as a rule little call to do so.

Although differing from Confucius in some respects, Xunzi remains true to his Confucian roots in others. Perhaps most importantly, he allows the sage the ultimate discretion to determine and apply *li* and *fa* based on his own best judgment:

> Just as there are disorderly rulers, but no such thing as a disorderly state, so there are exemplary persons who effect proper order but no such thing as laws that can effect it…. The laws cannot establish themselves and cases cannot determine themselves. Where there is the right person, they exist; where the right person is lost, they perish. Law is the beginnings of proper order, but the exemplary person is the source of laws.
>
> (*Xunzi*)

Xunzi’s goal was clear: to synthesize Legalism’s impartial application of generally applicable laws with Confucius’ concern for a particularized justice achieved through the discretionary judgment of a mediator who looks to the nuanced norms, attitudes and beliefs of the particular parties and community. Problems remained, however. The emphasis on informal means of resolving conflict ceded considerable power to the mediator. On the formal side, the ruler remained largely unchecked. There were no effective institutional impediments to limit the ruler’s power. Finally, like his predecessors, Xunzi offered no protection to the individual against the state.

7 The Confucianization of law and Legalization of Confucianism

Xunzi’s attempt to synthesize *li* and *fa* on a theoretical level was mirrored in practice by the so-called Confucianization of law and, conversely, the Legalization of Confucianism. The Confucianization of law has been discussed and documented by Ch’ü T’ung-Tsu (1961). It refers to the way in which the legal system came to adopt and reflect the important values of Confucianism. No doubt the formal legal system did incorporate and reflect many Confucian values. For example, the legal system paid particular attention to familial relations and filial piety. One of the unpardonable ‘ten offences’ was unfilial behaviour. Another was the murder of one’s parents. But while it was unpardonable for a son to murder his father, the father was permitted to kill or mutilate his unfilial son, as long as he first petitioned the court. Similarly, in keeping with the deference owed to parents, a son could be punished for simply accusing his father. As a result, the law officially permitted a son to conceal the crimes of the parent. One also had a legal duty to support one’s parents (as is still the case). In fact, the legal codes allowed for the non-execution of a criminal so that the criminal could look after his parents.

The influence of Confucianism and the *li* was also reflected in the hierarchical nature of the legal system and the concern for particularized justice. Punishments were meted out in accordance with one’s status and the status of the victim. Officials were treated more favourably than commoners. They could not be arrested, investigated or sentenced without permission of the emperor. Some were exempt from torture for certain crimes. All benefitted from sentence reductions, and could redeem certain punishments either by paying a fine or accepting a demotion. Status was also important for the official *qua* victim. Generally speaking, the higher the status of the victim and the lower the status of the offender, the more severe the punishment.

The law also drew a distinction between free persons and slaves. Slaves and members of the inferior class were punished more severely for killing, injuring, marrying or engaging in illicit sex with a free person or member of superior class than the contrary. Gender, age and moral character were also factors expressly incorporated into the legal codes’ scheme of punishment.

The influence of Confucianism on the legal system is equally apparent in the seriousness with which the system treated homicide and personal injury. Such crimes were punished with particular severity because they undermined the Confucian *lizhi* system. The Confucian approach to conflict resolution is one of persuasion, not force. Disputes are handled outside the courts through informal mediation. To be successful, the parties must be amenable to persuasion, willing to engage in discussion and compromise if necessary. If the situation has deteriorated to the point of violence, the parties will most likely be hostile to each other and less willing to listen to reason and compromise. Violence and killing, therefore, constitute a rejection of the Confucian mediation process. Once they occur, the state is forced to intervene and to mete out a severe punishment as a deterrent.

Of course, the continued reliance on extrajudicial means such as mediation to settle disputes is itself a product of the Confucian *lizhi* tradition. Another product of that tradition is the legal system's enduring tendency to pursue substantive justice at the cost of procedural justice. Perhaps the best indicator of the concern for substantive justice is the ready availability of appeal. Litigants could continue to press their case all the way up to the emperor if they so desired.

More generally, the influence of Confucianism and *lizhi* continued to be felt in the people's attitudes toward law. Most people felt it was a disgrace to be involved in a lawsuit. At trial, one would be forced to kneel and often subjected to torture. Litigation was also expensive, as captured by the Chinese proverb 'win your lawsuit but lose your money'. Thus most people attempted to resolve conflicts through extrajudicial channels if at all possible.

Although much has been made of the Confucianization of the law, equally important, though less discussed, is the 'Legalization' of Confucianism. Perhaps this phenomenon has drawn less attention because law has always played some role in a Confucian world. As noted above, Confucius saw a role for penal law. By the time of Xunzi, the role of law had expanded considerably. The importance of law is even more evident when one turns away from theory and the ideal society envisaged by Confucius and Mencius toward the real world. Recent archeological discoveries reveal, for instance, that contract law existed as early as the Han. In addition, each dynasty had its own legal code, which was often quite developed and intricate, and provided not just a finely calibrated system of punishments, but procedures for appeal and review of cases. The importance of law in pre-modern China is better appreciated if one considers not just the formal legal system, which was predominantly penal, but the full range of legal norms, including clan regulations and the rules of a particular guild or profession, which served at least to some extent as a form of civil law.

It would, therefore, be a mistake to portray pre-modern China according to some idealized Confucian stereotype that discounts the importance of law and minimizes the role of the formal legal system. At the same time, it would be a mistake to present the pre-modern Chinese state as a rule of law in a contemporary Western sense, at least if that is understood to entail equality before the law and the impartial administration of justice by an independent and professional judiciary. The pre-modern legal system continued to reflect the tension between *li* and *fa*, between the goal of a particularized justice that fully accounts for each person's circumstances and role in society and the goal of impartial justice meted out in accordance with rules of general applicability. It was, however, no more successful in its attempt to combine the best of *li* and *fa* in practice than Xunzi was in his attempt to synthesize the two in theory. The problems that threatened to undermine Xunzi's theoretical synthesis surfaced in practice. Corruption was rampant; the wealthy and powerful used their wealth and connections to circumvent justice. The ruler remained unchecked. Individuals were left at the mercy of the powerful state. As a result, the legal system was often feared more than respected, and reformers continued to press for change.

8 *Li* and *fa* in the modern era

The *li*-*fa* debate has continued in the modern era. With modernization as their slogan, turn-of-the-century reformers such as Liang Qichao declared ideological war on Confucianism, arguing for *fazhi* over and against *lizhi* and *renzhi* (rule by man). In the early years after the communist victory, the new regime was forced to turn to legal specialists, most of whom were Nationalist supporters, to design and implement a legal system. At the same time, however, many important positions within the legal system were filled by 'new cadres' appointed not for their legal skills but for ideological dependability. The specialists favoured a rule of law; the new cadres, supported by the central authorities, a more *li*-based 'rule of man'.

Ironically, the politics of the Cultural Revolution provided an interesting twist to the debate, with the post-Mao

government siding with 'revolutionary' Legalism and rule of law and opposing 'revisionist' Confucianism and rule of man. After the lawlessness of the Cultural Revolution and Mao's death, Deng Xiaoping then made reform of the legal system a top priority. For the time being, advocates of rule of law and a formalized legal system continue to have the upper hand.

The ongoing theoretical battle between the supporters of *li* and *fa* has meant in practice wide oscillation between formal and informal means of conflict resolution. During the Mao era, China lacked a formal criminal procedure law. Although the courts remained responsible for punishing major criminals, less informal channels for dispensing justice were equally if not more important. Work and residential units handled minor offences, and the police imposed administrative sanctions ranging from warnings to fines to re-education through labour. Mediation remained popular.

It is important to realize, however, that during the socialist period the state has controlled and continues to control both the formal and informal means of resolving conflict. Mediation committees, even at the neigbourhood level, are part of the state. The goal is to impart socialist ideology, raise the level of political consciousness of the disputants and resolve disputes in a way consistent with the policies of the Communist Party. Thus in many instances, mediation is far from the voluntary process envisaged by many Western advocates of alternative dispute resolution. Participation is often mandatory; there is often little choice but to accept the solution proposed by the mediator.

Furthermore, contemporary mediation and other informal or extrajudicial means of resolving conflict remain subject to abuse. Before, the Confucian literati were responsible for mediating an informal, particularized justice; now it is the *ganbu* (political cadres) that wield authority. When these cadres put their own interests first, the process breaks down and corruption becomes rampant as the parties seek to exercise their connections (*guanxi*) to influence the 'neutral' arbitrator or judge. Accordingly, many reformers have suggested that less emphasis be placed on informal techniques such as mediation and, conversely, the role of the formal legal system expanded. Their belief is that a formal legal process affords one greater procedural rights and protections, whereas more informal approaches too often subject one to abuse by those in power.

On the formal side, however, law and the judiciary remain firmly under the control of the state. The official view continues to be that law is to serve the Party. Ren Jianxin, President of the Supreme Court, recently stated: 'The independent exercise of judicial power by the people's courts occurs under the leadership of the Party. In order to adhere to [that principle] the independent exercise of judicial power by the people's courts must be supervised and supported…. Courts at all levels must consciously keep aligned with the central Party line, obey orders, listen to commands.'

Ironically, recourse to law has often meant harsh punishments in accordance with the latest Party dictates. Normally, the first official reaction to deviant or criminal behaviour is to attempt to informally persuade the violator to reform. Only if the deviant behaviour persists is the violator subject to the formal legal system and increasingly harsh punishments. However, during periods of political instability when the Party feels the need to address a particular pressing issue such as corruption or the outbreak of political demonstrations, the Party will insist that any violators be subject to the full force of the law. Accordingly, law, as for the Legalists, has remained a pragmatic tool to be exploited by those in power.

Despite the turn toward formal law, the influence of the past and *lizhi* thinking remains strong. One indicator is the ongoing concern for mediation and informal, extrajudicial means of conflict resolution. Perhaps most fundamentally, however, *li*-consciousness continues to shape the very understanding of the nature of law. Law, for most, is not understood as a universal norm. Rather, laws are culturally contingent, albeit generally applicable, norms. This understanding of law explains in part the difficulties in US-China discussions of 'human rights'. The US side considers such rights, if not self-evident, at least universally applicable regardless of time and place, historical traditions or level of economic development. In contrast, the Chinese view human rights, and particularly the civil and political human rights championed by the US, as the contingent product of Western Europe's Enlightenment traditions and the current economic and political circumstances of the US itself (see Rights). Although such rights may be worthy goals, they are predicated on assumptions about the nature of the individual, the relationship of the individual to the state, and the normative value of the individual and individual freedom *vis-à-vis* the good of the community and society as a whole. Accordingly, when China complains that the US is

interfering in China's domestic affairs, it is reflecting an understanding of laws, including human rights, as culturally contingent. When China conditions discussion of rights on the other's willingness to 'seek common ground while reserving differences, maintain mutual self-respect and promote understanding and cooperation', as is China's official policy, it is merely reflecting a traditional understanding of the nature of ethical norms and laws and the proper way to resolve conflict, whether interpersonal or international.

Conversely, China's human rights problems are emblematic of long-existing internal tensions in both China's political theory and practice. The human rights problem results in part from a lack of institutional restraints on those in power, which today means the Communist Party. The lack of effective institutional constraints - a professionalized and truly independent judiciary, for example - has been a characteristic of *lizhi* and *fazhi* throughout the ages.

The human rights problem also results from the lack of a constitutional tradition that looks to the constitution as a source of legally enforceable rights. China has had many constitutions: twelve since 1908. Many of these constitutions have set forth an impressive array of individual rights. However, Chinese citizens do not enjoy such rights in practice because the constitution is not justiciable. Rather, the constitution is hortatory: it sets forth the aspirations and goals of society. In that sense, the constitution reflects the Confucian tradition of emphasizing the development of the moral character of people and the role of the government as a provider of moral guidance and a vision of the good life.

Further, the fact that constitutions come and go is consistent with the traditional *lizhi* concern for context-specificity. Chinese constitutions do not set forth universal principles, eternal truths. Rather, they provide a snapshot of the cultural, political, economic and social values important at that particular time to that particular society. For instance, China has recently amended its constitutions to reflect significant changes in the economic system: socialist commodity economy is out, socialist market economy is in. If the political situation changes, and the conservatives and hardliners regain power, the constitution would probably be amended once again: socialist market economy would be out, socialist commodity economy back in.

9 Conclusion

Socialist China has continued to debate the merits of an informal *li*-based type of political order and a more formal *fa*-based political order. Increasingly, the institutional components of a formal legal system are being developed and implemented: a professional judiciary is being trained, new laws are being codified, programs aimed at educating people about the legal system and their rights are beginning to have effect, as evidenced by the dramatic rise in the number of lawsuits.

Nevertheless, critics abound, and many of these continue to raise the banner of *fazhi*. Increasingly, however, *fazhi* is understood as 'rule of law' in the sense of an professionalized and independent judiciary with the institutional integrity and authority to constrain the Party and safeguard individual rights. Whether such a system can be achieved, and how the traditional concern with *li* and informal justice will affect such a system, remains to be seen.

See also: Chinese Classics; Chinese philosophy; Confucian philosophy, Chinese; Daoist philosophy; Fa; Law, philosophy of; Legalist philosophy, Chinese; Marxism, Chinese; Paternalism; Sovereignty; Xunzi

R.P. PEERENBOOM

References and further reading

Ames, R.T. (1988) 'Rites as Rights: The Confucian Alternative', in L.S. Rouner (ed.) *Human Rights and the World's Religions*, Notre Dame, IN: University of Notre Dame Press.(Contrasts the two concepts of rites and rights.)

Bodde, D. and Morris, C. (1967) *Law in Imperial China*, Cambridge, MA: Harvard University Press.(Survey of the history of law in China.)

Ch'ü T'ung-tsu (1961) *Law and Society in Traditional China*, Paris: Mouton.(Survey of the history of law in China.)

Cohen, J.A. (1966) 'Chinese Mediation on the Eve of Modernization', *California Law Review* 54: 1201-26. (Discusses the Chinese legal practice of mediation.)

Edwards, R.R. ***et al.*** (1986) *Human Rights in Contemporary China*, New York: Columbia University Press. (Looks at the issue of human rights in a Chinese context.)

Fu Hualing (1992) ‘Understanding People’s Mediation in Post-Mao China’, *Journal of Chinese Law* 6: 211-46. (Another examination of the concept of mediation.)

Hall, D. and Ames, R.T. (1987) *Thinking Through Confucius*, Albany, NY: State University of New York Press. (Study of the impact of Confucius on Chinese thought.)

Hulsewe, A.F.P. (1955) *Remnants of Han Law*, Leiden: Brill.(Discussion of law in the Han period.)

Leng Shao-chuan and Chiu Hungdah (1985) *Criminal Justice in Post-Mao China*, Albany, NY: State University of New York Press.(A look at recent and current law in China.)

Li, V.H. (1978) *Law Without Lawyers*, Boulder, CO: Westview Press.(A look at Chinese attitudes to law and legal practice.)

MacCormack, G. (1990) *Traditional Chinese Penal Law*, Edinburgh: Edinburgh University Press.(Study of traditional Chinese law.)

Nathan, A. (1986) *Chinese Democracy*, Berkeley, CA: University of California Press.(Looks at the concept of democracy as understood in China.)

Peerenboom, R.P. (1993) *Law and Morality in Ancient China: The Silk Manuscripts of Huang-Lao*, Albany, NY: State University of New York Press.(Detailed study of law in the ancient period.)

Peerenboom, R.P. (1993) ‘What’s Wrong with Chinese Rights? Toward a Theory of Rights with Chinese Characteristics’, *Harvard Human Rights Journal* 6: 29-57.(Looks at Chinese approaches to the concept of human rights.)

Tay, A. (1987) ‘The Struggle for Law in China’, *University of British Columbia Law Review* 561.(Study of changes in the law in China following the end of the Maoist period.)

Law, economic approach to

The development of an economic approach to legal practice has been the most important jurisprudential development in the last third of the twentieth century. Economic analysis has been offered as both a positive and a normative jurisprudence: as an analysis of important features of existing legal practices and as an ideal against which these practices ought to be evaluated. For some, economic analysis has a narrow explanatory range (in various fields of private law, corporations and taxation, and anti-trust law, for example), while others make broader claims for its ability to illuminate any area of law. Finally, there is a difference between those who focus on one explanation and those who focus on prediction, but all offer positive economic analysis of law based on the concept of economic efficiency as defined in welfare economics and applied to law by Coase, Posner, Calabresi and others.

1 The concept of economic efficiency

Economics, and social sciences generally, in the early part of this century was under the influence of both utilitarianism and logical positivism. The correct course of conduct would be that which maximized utility (utilitarianism) (see Utilitarianism); the correct social science methodology required that claims be empirically verifiable or testable (logical positivism) (see Positivism in the social sciences §2). This was not an altogether happy marriage, largely because judgments about which course of action are utility-maximizing required judgments of interpersonal utility which were viewed as unverifiable. Because there exists no common metric against which the gains to one person can be compared with the losses others might suffer, social policy assessment of utilitarian terms in particular is thus called into question.

The sociologist Vilfredo Pareto introduced a mechanism for ordering social states that did not require interpersonal comparisons of utility but did allow policy-makers to favour some policies over others as utility-maximizing. This ordering relationship is known as ‘Pareto superiority’: one state of the world, W, is preferable to another, S, provided no one prefers S to W and at least one person prefers W to S (see Pareto principle). Put slightly differently: W is Pareto superior to S if and only if no one is better off in S than in W and at least one person is better off in W than in S. (These two characterizations are sometimes mistakenly equated on the grounds that the only reason why a person would have a preference for one state to another is that they would be better off in one or the other.)

The concept of Pareto optimality is then defined in terms of Pareto superiority. A state, W, is Pareto optimal if there exists no state, S_i, that is Pareto superior to it. The world is Pareto optimal when there are no ways of rearranging holdings (or anything else redistributable) to make at least one person better off without making someone else worse off.

There are obvious limitations to the Pareto rankings. Very few interesting real-world policy choices produce only winners (or, more weakly, produce no losers). Second, Pareto optimality can be extremely indiscriminate. Assuming a starting point in which A has ten units of some resource and B has none, any reallocation can improve B’s standing by harming A’s. Pareto superiority is almost never satisfied in practice, while Pareto optimality is too easily satisfied.

To help alleviate the first problem, several economists have introduced related ranking relationships, the most famous of which is the Kaldor-Hicks criterion: a state of the world, W, is Kaldor-Hicks superior to another state, S, provided that those who gain in going from S to W could compensate those who lose in such a way that, were compensation to occur, W would be Pareto superior to S. In principle, Kaldor-Hicks allows us to compare states that create winners and losers without making interpersonal comparisons, thus extending the Pareto rankings.

Kaldor-Hicks employs a notion of ‘hypothetical compensation’. In contrast, Pareto superiority requires actual compensation (so that losers, when compensated, are no longer losers). Actual compensation may be costly, and in some cases these costs might suffice to eliminate the gains associated with reallocation. From a utilitarian point of view compensation may be self-defeating. But, if the point of the ranking relationship is simply to have a reliable indicator that a policy is utility-increasing, then hypothetical compensation may be all that is required.

Kaldor-Hicks is not, however, a transitive ordering relationship. Two social states may be Kaldor-Hicks superior

to one another. In that case, it is impossible to infer that either has more total utility than the other. Thus Kaldor-Hicks cannot be used to judge the utilitarian merits of various policy options. Numerous variants on the Kaldor-Hicks criterion have been introduced to overcome the intransitivity problem. Tibor Scitovsky, who first established the intransitivity paradox that beset Kaldor-Hicks, developed his own variant, which was shown also to be intransitive by Kenneth Arrow. Of course, one way of obviating the problem would be to make interpersonal comparisons of relative utility, but doubts as to the meaningfulness of such comparisons were precisely what led to the alternative ranking systems.

2 The 'Coase Theorem'

Ronald Coase's seminal piece 'The Problem of Social Cost' (see Coleman and Lange 1992) established the so-called 'Coase Theorem'. In a well-functioning market, individuals trade with one another in ways that benefit all concerned. These exchanges are Pareto superior or Pareto improving (barring adverse third-party effects). In a perfect or idealized market, such exchanges occur until no further mutually advantageous exchanges can be made. At that point the outcome would be Pareto optimal. There are many barriers to successful exchange, however. If property rights are uncertain, then exchange will be difficult. Take the familiar case of the rancher and the farmer. The rancher believes he can raise cattle and the farmer believes he has the right to grow corn. The relevant uncertainty concerns whether, if the rancher's cows eat the farmer's corn, the rancher will be liable in damages to the farmer.

In his essay Coase argues that, under certain conditions, it will not matter from the point of view of efficiency whether the rancher is free to raise as many cows as he would like without being liable to the farmer or whether, instead, the farmer can prohibit the rancher from raising cows altogether. The argument has two parts. First, what would be the efficient allocation of cows and corn?

Suppose the rancher and the farmer were one person, rancher-farmer, then the answer is simple. Each cow imposes costs in terms of corn that will be destroyed by roaming cows. The rational rancher-farmer will continue raising cows until the marginal costs of doing so exceed the marginal benefit. When the rancher-farmer can earn more from the corn their cows would destroy than they would from the cows, then they will stop raising cows. That is the efficient allocation of cows and corn. Any additional cows will be more costly than beneficial.

The second part of the argument is that when rancher and farmer are two individuals they will raise the same allocation of cows and corn they would have as rancher-farmer. This is so no matter which party has relevant right. Suppose that the rancher and the farmer are fully rational, informed about each other's preferences and, finally, that the transaction costs between them are trivial. Then suppose that the farmer is entitled to prohibit the rancher from raising cows altogether. We know from the above discussion that the value of the last bit of corn to the farmer is less valuable than is the first cow to the rancher. Thus, the rancher can afford to offer the farmer a sum of money equal to, or somewhat greater than, the value of the lost corn to the farmer (and less than the value of the first cow to the rancher). Accepting that sum will make the farmer better off and the rancher will have his first cow. This will apply right until that point at which the rancher will be unwilling to pay the farmer an amount of money equal to the value of the corn the farmer will lose. This is just another way of saying that the value of the next cow will fall below the value of the forgone corn. When that occurs, no further transactions will occur. The net effect will be an efficient allocation of cows and corn.

Conversely, it is easy to see that this same line of argument would work were we to begin with the assumption that the rancher had a right to raise as many cows as he would like free from liability to the farmer. The last cow is less valuable to him than is the first bushel of corn to the farmer. The farmer would be in a position to offer the rancher a mutually advantageous exchange: cows for corn. This will go on until it is no longer rational for the farmer to purchase rights from the rancher. And that will be the efficient point at which no more mutually advantageous trades can be made.

One way of looking at the Coase Theorem is that when there are barriers to market exchange that would otherwise be mutually advantageous or efficient, then the best first option is to see if a secondary market can be set up that allows individuals to contract around their problem. In this case, the barrier to efficiency is an *externality*, that is, the social costs cows impose on corn; the solution is transactions (themselves made possible by the absence of transaction costs). Always look to the contract or the market solution. Why? Simply because such a process

assures Pareto improvements: rational, fully informed agents make mutually advantageous exchanges.

But when the market fails and it is too costly to create secondary markets, then look to the law to establish legal rights and duties that mimic the outcome the market process would have come to had the transaction costs been less troublesome. When it is not possible completely to mimic the efficient outcome, the law should do what it can to reduce transaction costs and encourage private solutions.

3 The Coase Theorem in action: contracts and torts

In a fully specified contract, all the parties would anticipate and respond to all possible contingencies. Nothing could happen in the future that the parties had not contemplated and responded to. For a variety of reasons, no contract is fully specified. Some contingencies are not accounted for in a normal contract. That does not mean, unfortunately, that none of those unlikely events will come about. How should the costs of such an event be allocated among parties who have not determined in advance how they would allocate the risks between or among themselves?

There is always a possibility that the parties could work it out privately. Suppose they cannot, and the matter comes to a court. What should the court do? What rule should the court apply in *default* of the parties having an explicit arrangement? One answer, suggested by the Coase Theorem, is that the court should impose the allocation of rights and responsibilities the parties would have come to themselves had transaction costs not made it impossible for them to contract such an arrangement. There is another alternative that also has links to the original Coase article. The court should forget about the two parties currently before it. Instead it should focus on future contracting parties. In doing so it should set default provisions that will encourage future contractors to reveal information that it would be socially desirable for them to reveal but which neither party would normally have a rational self-interest in revealing. So we can use the default rule to promote efficiency in at least two different ways: either by minimizing transaction costs so that the parties can spend the resources they have in areas that are more important to them, or by promoting *socially* efficient revelation of information by future contracting parties.

Now to torts. A injures B. Should A be made to pay B's costs or should they be required to shoulder their own costs? Should a rule of fault or of strict liability be applied? If A has to bear B's costs whether or not A is at fault (strict liability), that would be just like making A be the 'owner' of both activities, by making A the owner of B's costs. Even if A has to bear B's costs, it does not mean that A will no longer injure B. Rather, A will decide whether it is worth it to injure B and bear the costs or whether it is better to forego the injurious activity. That, in turn, depends on the relative costs and benefits. In economics, this practice of forcing A to bear the costs he imposes on B is referred to as *internalizing externalities*. A rule of strict liability appears to lead to efficient outcomes in just the way the rancher-farmer example does.

The goal of tort law is not to eliminate accidents altogether. After all, the costs of preventing an accident may exceed the costs of the accident itself. It may cost $100 to prevent a $50 accident, and that would not be rational. If it would have cost A that much to prevent a similar injury to B, he would not have prevented the injury. Instead he would have risked the injury and saved $50.

The goal is not to avoid accidents but to minimize the sum of the costs of accidents and the costs of avoiding them. Sometimes minimizing costs means reducing the probability of an accident's occurrence. Sometimes reducing the probability requires that precautions be undertaken by both injurers and victims. A rule of strict liability will not give a victim any incentive to take precautions, because the victim is fully compensated for his loss. Why would he accept any additional costs? A rule of strict liability will not be efficient under such circumstances.

Consider the alternative. Under a rule of negligence, an injurer will not be liable unless he is at fault or negligent in causing another harm. What do we mean by negligence? In economic analysis, negligence is the failure to take cost-justified precautions. Precautions are cost-justified when incurring them is less costly than the injury they are designed to prevent, discounted by the probability of its occurrence. So failure to take precautions would be negligent if it would cost $50 to prevent a $100 accident - but not if it would cost $150 to prevent the same accident.

It will always be rational for an injurer to take all cost-justified precautions under a negligence rule. If he does, he escapes liability; if he does not, he will bear the full costs of the accident which, *ex hypothesi*, exceed his

prevention costs. Thus, rational agents will not be negligent. Because the injurer always takes rational precautions, the accidents that occur are not his fault. That means the victim will have to bear his own costs under a negligence rule, and hence will always have a similar incentive to take all cost-justified precautions. If he can reduce the probability of his incurring costs by taking precautions that are less costly than the harm, then he will do so; and so on. Thus, there will be cases when it would make sense to have a rule of negligence, and other times when a rule of strict liability will be in order.

See also: Economics and ethics §4; Law, philosophy of; Market, ethics of the §2

JULES L. COLEMAN

References and further reading

Coleman, J. (1992) *Risks and Wrongs*, Cambridge: Cambridge University Press. (Expansion of the material in §4 of this entry.)

Coleman, J. and Lange, J. (1992) *Law and Economics*, Aldershot: Dartmouth. (Wide-ranging collection of essays, including R. Coase’s ‘The Problem of Social Cost’, with an introduction by the editors.)

Posner, R. (1972) *Economic Analysis of Law*, Boston, MA: Little, Brown.(Seminal work for modern Chicago-based economic analysis, particularly influential in US legal thought.)

Law, Islamic philosophy of

*One of the principles of Islam which precedes juristic discussion proper is that God, the creator and lord of the world, has commissioned humanity to believe, confess and act in particular ways. The details of this commission (*taklif*) were handed down through a sequence of prophets, culminating in Muhammad, and were then embedded in two literary structures which together constitute revelation (*wahy*): the Qur'an, which is the word of God, and the* hadith*, short narratives of the prophet's life and sayings which give expression to his (and his community's) ideal practice or* sunna*. The totality of beliefs and rules that can be derived from these sources constitutes God's law or* shari'a.

Juristic literature has generated two major literary genres. One, known as usul al-fiqh *(roots of jurisprudence), deals with hermeneutical principles that can be used for deriving rules from revelation; it represents, in part, something like a philosophy of law. The other, dominant genre,* furu' al-fiqh *(branches of jurisprudence), is an elaboration of rules which govern ritual and social activities. An overall philosophy of law in Islam, not fully articulated in the pre-modern tradition, can only be discovered through consideration of both genres.*

1 Revelation

In Islamic belief there are a number of principles, derived from the exercise of the intellect or from history, which precede juristic discussion proper. These include, for example, that God exists, that he is creator and lord of the world, and that he has commissioned humanity to believe, confess and act in particular ways. The details of God's commission (*taklif*) have been mediated through a sequence of prophets culminating in the seal of all the prophets, Muhammad, whose message abrogates previous messages and is for all peoples. With the death of the Prophet, the divine command has been embedded in two literary structures which together constitute revelation (*wahy*): the Qur'an, which is the word of God, a miracle, and the *hadith*, short narratives of the Prophet's life and sayings which give expression to his (and his community's) ideal practice or *sunna*. The totality of beliefs and rules that can be derived from these sources constitute God's law or *shari'a* (see Islamic theology).

The Qur'an is usually deemed to contain no more than about 500 verses of legal import. The body of *hadith* was immensely larger. It was contained in a number of admired collections, the core of which included, for the mainstream Sunni community, the pre-eminent collections of Bukhari (d. AH 256/AD 870) and Muslim ibn al-Hajjaj (d. AH 261/AD 875). Beyond these, there were a number of collections that might be brought into the play of juristic discussion; the major sectarian group, the Shi'is, also had their own collections. The vast bulk of *hadith* material, as contrasted with the modest quantity of juristic material in the Qur'an, ensured that *hadith* was in practice the dominant element of revelation in hermeneutical discussions. The Book is more in need of the *sunna* than the *sunna* of the Book, said the Syrian jurist Awza'i (d. AH 157/AD 774), echoing the efforts of other scholars to articulate the controlling effects (judging, abrogating and explaining) of *sunna* on the Qur'an.

2 *Usul al-fiqh*

The literary tradition of *usul al-fiqh* (roots of jurisprudence) is usually thought to begin with the *Risala* of Muhammad ibn Idris al-Shafi'i in the third century AH (ninth century AD). However, there is then a hiatus between this work and the emergence, some two centuries later, of other works of the same kind. Modern scholars have put this down to pseudepigraphy, or to the community's engagement with theological and intellectual problems. Once established, this tradition, although it developed in a variety of ways, showed remarkable structural and conceptual unity over the centuries, with individual books always including a presentation of hermeneutical principles and an elaboration of the theory of *ijtihad* (independent judgment).

Since revelation was constituted by written texts, a primary bundle of hermeneutical techniques related to linguistic and rhetorical structures. These were usually presented under simple antithetical headings (the general and the particular, commands and prohibitions, the clear and the ambiguous, the absolute and the qualified, truth and metaphor and so on), which might or might not be integrated into a general theory of language and rhetoric. Consideration was given to the principle of abrogation (*naskh*), a result of diachronic revelation, and (with reference to *hadith* only) to the mode of transmission, which was either general report (*tawatur*, giving rise to certain knowledge) or isolated report (*ahad*, giving rise to uncertain knowledge or opinion). Consensus (*ijma'*),

whether of the community or of scholars, on the meanings of revelation was discussed, as was the operation of analogy (*qiyas*) as a means (variously qualified) to permit extrapolation of rules from a finite body of revealed texts. These, together with a limited number of extra items, either of substance (such as the opinions of the Prophet’s companions) or of judgment (such as the relevance of *maslaha*, or social welfare) represented the major focuses of analysis and discussion within a single and more or less unified literary tradition for about a thousand years. The hermeneutical loyalties of the mainstream Sunni community were summarized by reference to the four principles (*usul*) of Qur’an, *sunna*, consensus and analogy.

The whole bundle of interpretative devices and principles of judgment was acknowledged to lead to conflicting possibilities (*ta‘arud*) and to the necessity for rational and justified preference (*tarjih*). The context and the significance of juristic preference depended on the theory of *ijtihad*, the expression of which was a culmination and a kind of resolution for all other arguments in a work of *usul*. *Ijtihad* literally means effort; technically, it means the exertion of the utmost possible effort by a trained jurist, taking into account all the relevant texts of revelation and principles of interpretation, in order to discover, for a particular human situation, a rule of law. Underlying this definition there is an important epistemological principle. It concedes that most of the details of the law are not known (not certain) but are a matter of skilled (and preferably pious) deduction on the basis of principles that are themselves subject to debate and incapable of providing certainty.

Within this area, the jurists were committed to acknowledging the views of other jurists, if adequately defended, and to the elaboration of systematic arguments to defend their own views. Committed in this respect to debate and uncertainty, the jurists (in this context *mujtahids*, those who undertake *ijtihad*) also acknowledged a need for final decisions in particular cases. This was provided by asserting that the result of an act of *ijtihad* was binding both on the *mujtahid* himself and, where relevant, on those who were not experts in the law and could not participate in juristic debate (*muqallids*). These, by an exercise of choice (which was itself an act of *ijtihad*), were required to commit themselves to a particular *mujtahid* and to accept his rulings. The theory of *ijtihad* thus provides both an epistemology (permitting and encouraging debate and intellectual play) and a structure of authority. In its former aspect it accounts (in part) for the vitality of the tradition of *furu‘ al-fiqh*, and in its latter aspect it justifies the participation of the jurists in positions of authority, notably as judges (*qadis*) and the jurisconsults (*muftis*).

The structures set out above, which were capable of considerable and diverse development, represent the main features of *usul* literature for both the Sunni community and for the Shi‘is. The latter differed from the Sunnis in rejecting most forms of analogical argument. Summarizing their *usul*, they substituted for the Sunni principle of analogy that of intellect (*‘aql*) deemed by them (but not by the Sunnis) to be capable of independent moral judgment. The Shi‘i community was initially suspicious of the theory of *ijtihad* (perhaps because it too easily acknowledged plurality and uncertainty in the law), but it was integrated into their works of *usul* from the time of ‘Allamah al-Hili in the eighth century AH (fourteenth century AD). Amongst the standard classics of the Sunni tradition are the *Mustasfa* of al-Ghazali (§2) and the *Muwafaqat* of Shatibi (d. AH 790/AD 1388), showing an original foregrounding of the principle of *maslaha*. One of the outstanding *usul* writers of the Shi‘i tradition after ‘Allamah is the nineteenth-century Shaykh Murtada ibn Muhammad Ansari.

3 *Furu‘ al-fiqh*

The other major genre of juristic literature, *furu‘ al-fiqh* (branches of jurisprudence), is constituted primarily by rules (positive law). It might be expected that individual writers in this genre were, from generation to generation, engaged in the process of *ijtihad*. However, this is not quite the case. In the course of the ninth and tenth centuries AD, the Islamic community became committed to a pattern of juristic loyalties whereby, in the end, all Muslims identified themselves with particular schools (*madhahib*) of the law. Within the Sunni tradition there were four dominant schools, the Maliki, Hanafi, Shafi‘i and Hanbali schools, each named after its founder. These acknowledged each other and also gave qualified acknowledgement to a number of minor schools, and to the Shi‘is. The vast majority of significant jurists belonged to one of the major schools, usually by virtue of birth and geography, only rarely by choice and adoption. When they wrote a work of *furu‘ al-fiqh*, they gave expression to the rules (with the attendant patterns of dispute and debate) that they had inherited within their school. The fundamental acts of *ijtihad* were thus projected back to the founder and to the early masters. By an ongoing act of loyalty, commitment and preservation, successive generations of jurists rediscovered and restated the rules of the tradition to which they belonged.

Works of *furu‘* therefore show a dominant hermeneutical orientation towards earlier works in their own tradition, and not towards revelation. This is reflected in the characteristic patterns of citation, which invariably recall the opinions and judgments of earlier masters within the school, and the literary forms of such works (epitome, commentary, supercommentary), all marks of hermeneutical commitment to a particular school. In so far as writers in this tradition actually deployed arguments of the type described in works of *usul*, they did so in order to demonstrate that the inherited structure of rules could be aligned with revelation and not for the purposes of *ab initio* deduction of the law. Developments in the law, manipulation of its concepts and their application to new cases were always carried out in the light of the inherited structure. The inconcinnity between the principles of jurisprudence as set out in a work of *usul* and the practice of a writing jurist was eventually acknowledged with the recognition that, in relation to the school founders, these were principles of discovery while, in relation to later jurists, they were principles of justification. Hermeneutical thinking within the tradition (based on juristic texts, not texts of revelation) was known as *ijtihad fi ’l-madhhab* or school-*ijtihad* and distinguished from the independent *ijtihad* of the founding figures.

4 Contemporary trends

Muslim thinkers of the twentieth century, committed to various programmes of legal reform or political action, have developed a number of theoretical props which take them away from the traditional modes of juristic expression. They have often abandoned the particularity of school loyalties; instead, they have adopted law-drafting techniques that reflect the realities of modern nation-states, borrowed legal and social principles from a variety of sources, and argued strenuously that the door of independent *ijtihad* is open, meaning that they can again make independent legal judgments based on direct confrontation with revelation (often using a definition of revelation, at least in relation to *hadith*, which is more limited than that of the past). The major, and certainly the most obvious, modern responses to the juristic tradition have been practical, either in the service of state law or in the service of political opposition. There has been a corresponding lack of interest in the philosophy that is articulated in the traditional forms of juristic discourse, especially those of *furu‘*. In different ways, both of the traditional genres, *usul* and *furu‘*, acknowledge the exploratory nature of the effort of defining God’s law and situate themselves in a flexible and pluralist system of rules. They may have more relevance to contemporary problems than is generally conceded (see Islamic philosophy, modern).

The modern linguistic calque *falsafat al-tashri‘* (philosophy of legislation) or its equivalent is currently used in several Islamic countries to designate a variety of academic activities. These range from conservative scholarship, drawing heavily on the tradition of *usul al-fiqh*, to analytic descriptions of legal developments in modern Islamic states and something like the Western discipline of philosophy of law.

See also: Epistemology in Islamic philosophy; Ethics in Islamic philosophy; Islamic philosophy, modern; Islamic theology; Law and ritual in Chinese philosophy; Law, philosophy of

NORMAN CALDER

References and further reading

Calder, N. (1993) *Studies in Early Muslim Jurisprudence*, Oxford: Clarendon Press. (The most recent substantial effort to describe the early stages of Islamic juristic thinking.)

Calder, N. (1996a) ‘Islamic Law’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 979-98.(Offers a general account of Muslim juristic literature, initiating a literary description of the tradition of *furu‘ al-fiqh*.)

Calder, N. (1996b) ‘Al-Nawawi’s Typology of *mufti*s and its Significance for a General Theory of Islamic Law’, *Islamic Law and Society* 3 (2): 137-64.(Demonstrates the emergence and elaboration of a clear theoretical distinction between independent *ijtihad* and *ijtihad fi ’l-madhhab*.)

Hallaq, W. (1994) *Law and Legal Theory in Classical and Medieval Islam*, London: Variorum.(An important reassessment of scholars’ approaches to the history of *ijtihad*.)

Hallaq, W. (1992) ‘*Usul al-fiqh*: Beyond Tradition’, *Journal of Islamic Studies* 3 (2): 172-202.(The bibliography of this article contains an up-to-date list of Hallaq’s substantial and important body of studies related to *usul al-fiqh*.)

Masud, M.K., Messick, B. and Powers, D. (1996) *Islamic Legal Interpretation: Muftis and their Fatwas*, Cambridge, MA: Harvard University Press.(An important and wide-ranging collection of articles representing some of the most recent thinking about the theory and practice of Islamic law.)

Schacht, J. (1950) *Introduction to Islamic Law*, Oxford: Clarendon Press.(Still the classic introduction to the subject, with an excellent but now dated bibliography.)

Law, limits of

Questions concerning the proper limits of law are of particular interest to thinkers in the Western political tradition of individualism. In this tradition the law is regarded primarily as an instrument of coercion and the problem is to define the scope of law in such a way that it fulfils its necessary purposes at minimum cost to individual liberty. Debate therefore centres on the proper ends of legal coercion. Two law-limiting strategies are commonly adopted; the practical and the moral. As the most important ends of human life (salvation of the soul, or its secular equivalent, moral integrity) are taken to require the uncoerced, 'inward', assent of the individual, the effective scope of the law is significantly limited on practical grounds to the regulation of 'outward' behaviour. On the moral question concerning which behaviour ought to fall within the purview of the law, conservatives contend that society has a right to enforce its moral values by criminalizing whatever behaviour its members regard as 'sinful'. The characteristic strategy of liberals is to respond by arguing that unorthodox or unpopular activities must clearly be shown to be 'harmful' before they may properly be outlawed. Much debate focuses upon the interpretation of this principle (the 'harm principle'), particularly what is to count as 'harm' for purposes of legislation, and whose harm is properly in question. Dissatisfaction on these crucial issues has led some liberals to reject what they consider to be the fruitless attempt to draw the line between liberty and law by balancing individual against social harm in favour of various theories of individual rights, while it has led others (communitarians and some feminists) to question the individualist assumptions in terms of which the problem of the limits of law gains its especial urgency.

1 Individualism and the limits of law

Should the law be used to promote religion, regulate trade and 'enforce morals'? Or, should it be narrowly confined to the bare essentials of preserving the peace and preventing violence? The question of the proper limits of law was at best a marginal one for the ancient Greeks and the medievals, and remains so today for many non-Western thinkers. It is, however, an issue which has been of particular concern within the modern (post-sixteenth century) Western tradition of political thought, a preoccupation reflecting the importance accorded to individual liberty in this tradition, together with the marked cultural diversity of modern Western societies and the unprecedented growth in the power of the state. The conceptual background against which the question of the proper limits of law is most naturally raised is one of 'individualism', according to which society is composed of independent individuals who associate politically on the basis of a mutual sacrifice of liberty in return for the various benefits of social cooperation. This view encourages a search for reasons or principles which may be used to assess the appropriateness both of existing laws and of proposed legislation by defining the proper balance between individual liberty and state power. 'Appropriateness' may, however, mean either practical efficiency or moral rightness (see Liberalism §3).

2 Practical limits to law

It is manifestly desirable that laws be efficient; the proposed ends of the law should be achievable by the means available to the law. Now the law, broadly understood as a system of rules established by a legal authority requiring us to do or forbear on pain of punishment, ultimately relies upon coercion for its enforcement (see Law, philosophy of). Consequently, the scope of the law is limited to those ends which can be attained by regulating *behaviour* by playing upon *fear*. Locke deployed the efficiency argument against the state enforcement of religion. Beginning from the proposition that the main purpose of religion is the salvation of souls, which requires an unforced act of faith on the part of the individual believer, he argued that legislators who set out to enforce religion can at best compel 'outward conformity' rather than genuine conviction of belief (see Toleration §1). Hence the legal enforcement of religious orthodoxy commits a practical error - it mismatches means and ends. The efficiency argument thus sets limits to the scope of law by dividing society into separate spheres, in Locke's case the religious and the secular, legal coercion being confined strictly to the latter. Later liberals confine the scope of state action even within the secular sphere itself, especially in respect of morality and economics (see Smith, A.). In regard to morality the argument focuses upon the conditions for the attainment of moral virtue. Thus, Kant argued that a morally good person is someone who not only does the right thing, but does it for morally the right reasons (see Kant, I. §10). As the freely adopted intentions of the agent are of paramount ethical concern, the (positive) promotion of virtue is assigned to education and placed permanently beyond the reach of the rough

coercive engine of the law, which is restricted to the more modest role of the (negative) prevention of vice.

The efficiency argument well reveals the twin poles around which liberal individualism revolves: reverence of individual freedom of choice and a corresponding distaste for legal coercion. The position, however, has its weaknesses. In particular, it takes a debatably narrow view of the strategies open to legislators. Thus, although 'true faith' might not be directly enforceable, it might still be produced by indirect means, for example, by banning religious competition, censorship and by controlling education. Likewise, the Kantian position on moral goodness perhaps underestimates the scope for indirect coercion by way of using the law to sustain a framework of broadly educative institutions and practices within which morally desirable motives might be elicited and encouraged. Even so, these possibilities raise in their turn difficult questions concerning the 'authenticity' of beliefs and the 'genuineness' of motives produced by such methods of 'social engineering'. At this point the efficiency argument leads on to issues connected with freedom of the will and human agency, and to questions concerning the nature of law and the educative role of the state (see Free will; Autonomy, ethical).

3 Moral limits to law

What the law *can* do is one thing; what it *ought* to do is another. A useful way of approaching the moral question is to ask how far law should go in pursuit of the essentially negative and in general perfectly feasible goal of preventing vice. Clearly, no morally defensible legal system can permit murder and rape, ignore violence and intimidation, or countenance gross dishonesty and theft. Peace and security, the basic prerequisites of any minimally tolerable social life, are universally essential moral ends. But ought the law to be narrowly confined to securing this 'moral minimum', or should it be extended to regulate those less immediately vital aspects of our relations with others which nonetheless often express passionately held convictions about right and wrong, serve to distinguish the characteristic ethos of one community from that of another, and even to define our own sense of moral identity and personal worth?

On this question two contending schools of thought may be distinguished: conservatism and liberalism. The case for legal conservatism goes as follows. Murder, rape, theft and the like are evil acts. By outlawing evil the law expresses our deep and abiding moral convictions and, by convicting and punishing the criminal, the judge and jury give emphatic vent to our community's shared sense of moral outrage and disgust. The reach of the law is thus virtually unlimited: it stretches as far as the community's sense of sin. The chief difficulty with the conservative doctrine is that, murder and the like aside, it is far from obvious that the criminal law can any longer act (if it ever did) as the legal vehicle of a commonly shared morality. Indeed, the problem of the limits of law is an acute one precisely because little general agreement is to be found on such questions as euthanasia, abortion, genetic engineering, pornography, homosexuality, the selling and consumption of addictive drugs, 'politically correct' language, or the killing or mistreatment of animals. Thus, although legal conservatism often sails under the colours of moral populism, it sanctions the dogmatic imposition of a morality which might have little support beyond the accidental prejudices of the notoriously introverted judicial mind. Hence the need, in the liberal view, for a clear and rationally defensible moral principle which affords a critical perspective on the ends of law.

4 Law-limiting principles and the crisis of liberalism

The fruits of the liberal quest for a morally justifiable law-limiting principle have taken two main forms. The first is the Principle of Equal Freedom: 'People should be legally free to do as they please providing only that they do not interfere with the like liberty of others'. This principle reflects the liberal's concern that the state should act merely as a neutral referee of people's freely chosen activities rather than determining their ends and directing their lives (see Neutrality, political). But because it fails to specify concretely what people should be free to do, a law permitting bodily assault would satisfy the rule just as well as one prohibiting it. By contrast its rival, the harm principle, has the immediate advantage of connecting substantively with the most intuitively arresting feature of our moral minimum. The advocates of the harm principle readily concede that we are properly morally outraged by murder and the like but, *pace* the conservative, it is the objective harm done, and not the subjective feelings of the community, that justifies the criminalization of these activities. In addition, the harm principle is advanced as furnishing an empirically applicable decision procedure for rationally resolving disagreements upon cases beyond the moral minimum, and especially upon the more controversial ones such as 'victimless crimes' of the kind alluded to above.

The harm principle was classically formulated by the utilitarian philosopher J.S. Mill (§12). Mill argued that the general happiness is maximized only when coercive interference with the individual is rigorously restricted to that unavoidably necessary to 'prevent harm to others'. A great deal of contemporary discussion of the limits of law is concerned with the interpretation of Mill's version of the principle. Two questions are particularly pressing: what is to count as 'harm' for the purposes of legislation? and whose harm is properly at issue?

Physical injury clearly counts as harm for the purposes of legislation, but what about injuries sustained by voluntary participants in competitive sports? We clearly need to qualify the principle with a sub-rule of non-consent to cover cases of this kind; and this in turn raises the question of what is to count as consent - explicit agreement? tacit consent? passive acquiescence? (see Consent). Moreover, does harm extend to psychological harm, and does that in turn stretch to emotional hurt? If so, should offensive behaviour fall under a legal ban too? And is behaviour offensive if any one person is liable to be upset by it? Clearly, if we concede too much in this direction the protected sphere of liberty is in danger of shrinking to a nullity; a line must be drawn, yet drawing it at any particular point smacks of arbitrariness. Mill himself is by no means clear on these questions but it has been argued that we can sort out the legally relevant harms in terms of 'set-backs' to *interests* (see Needs and interests §1). Physical injuries, psychological traumas and the like set us back in the sense of preventing us from effectively pursuing our chosen ends; they thus harm our interests and may properly be forbidden. Lesser hurts do not set our interests back and must therefore be permitted. However, serious set-backs can occur to an individual's economic and professional interests in competitive settings which are not themselves undesirable. What is required therefore is a tenable conception of a wrongful harm or set-back, something that is unlikely to be supplied outside the context of a general and inclusive moral theory. Mill believes that he has such a thing in utilitarianism. The weakness of utilitarianism in the eyes of many liberals, however, lies precisely in its tendency to subordinate the individual to the general interest (see Utilitarianism).

Similar difficulties attend the question of whose harm is at issue. Mill himself takes a radical line here, rejecting all laws against self-inflicted harm on the grounds that they constitute a paternalistic interference with the freedom of adults to live their own lives and make their own mistakes; a position on the face of it scarcely consistent with the maximization of general utility and one which also implies highly controversial conclusions about drug taking, euthanasia, sadomasochistic sex, and so on (see Paternalism). Equally problematical is the question of whether the principle should be restricted to harm to other *individuals* (including foetuses?) or should encompass harm to *society*, including its political institutions and practices, together with basic social forms such as the family and associated values and ideals. This was the issue of a celebrated debate between the eminent English judge Lord Devlin and the influential legal philosopher H.L.A. Hart concerning the legalization of homosexuality. Devlin argues that if the purpose of the criminal law were simply to protect the individual from harm the consent of the victim of an assault or a rape would be a complete defence, but quite properly the law pursues the assailant, irrespective of consent; and it does so because it represents the profound interest society has in the integrity of the rules which protect the welfare of all its members. Given that the purpose of the criminal law is thus primarily to protect society rather than the individual, Devlin infers that homosexuals enjoy no unconditional right to pursue their consensual activities, for the law may always legitimately ask whether, in so doing, they harm or undermine institutions (such as the family) or values (such as heterosexual love) essential to society's survival or wellbeing (see Devlin 1965). And the conclusion may be generalized: no socially unorthodox activity can claim to be, as a matter of principle, 'not the law's business'.

The import of Devlin's quasi-conservative argument is thus seriously to challenge the grounds of the distinction, so vital to the liberal campaign to limit the scope of law, between the private and the public spheres of society (see Privacy). What is particularly significant about Hart's response (1963) to Devlin is that it reveals the vulnerability of utility-based liberal individualism. Thus, *qua* utilitarian, Hart absorbs the force of Devlin's argument concerning the social purpose of the law, extends the harm principle to society, and concentrates his defence of individual liberty upon Devlin's claim as to what in fact constitutes social danger - in his opinion, Devlin's fear that homosexuality threatens the collapse of society is no more credible than the emperor Justinian's conviction that it is the cause of earthquakes. It is not clear, however, that Devlin does, or needs to, go further than the proposition that heterodox activities might constitute a significant danger to society to make his point. In any event, the effect of the Hart-Devlin confrontation is so to shift the debate that the outcome depends on hard-to-verify conjectures about the conditions of social stability and breakdown.

Under pressures of this kind other liberals have tended to fall back upon another Millian idea, that of 'individuality', interpreted as implying that our most vital interest as human beings lies not simply in happiness *per se* but rather in the specific satisfactions associated with developing our autonomy and displaying our individual uniqueness. As individuality essentially involves the outward expression of inward convictions and needs, it entails a presumption in favour of individual liberty in respect of unpopular or unorthodox behaviour sufficiently powerful to counter all but the most overwhelming reasons for legal interference. This raises the objection from socialists, communitarians and some feminists that liberals are prepared to erode the shared practices and values which constitute our identities as social beings in favour of an atomized association of socially disconnected 'abstract' individualists (see Community and communitarianism §3; Feminist political philosophy). Other thinkers, critical of the whole utilitarian idea of aggregating and balancing individual harm against social good, maintain that the scope of law must be limited by a theory of individual rights grounded independently of utility (see Rights; for prominent examples of this position, see Dworkin, R. and Rawls, J.). The question as to the proper limits of law therefore appears increasingly to be embraced within a broader debate as to the philosophical credentials of individualistic liberalism.

See also: Freedom and liberty; Justice

G.W. SMITH

References and further reading

Devlin, P. (1965) *The Enforcement of Morals*, Oxford: Oxford University Press.(Chapter 1 formulated an influential conservative view concerning the problems facing attempts to specify moral limits to law.)

Feinberg, J. (1984-6, 1988) *The Moral Limits of the Criminal Law*, vol. 1, *Harm to Others*, vol. 2, *Offense to Others*, vol. 3, *Harm to Self*, vol. 4, *Harmless Wrongdoing*, Oxford: Oxford University Press.(A magisterial defence of a modified version of Mill's harm principle.)

Hart, H.L.A. (1963) *Law, Liberty and Morality*, Oxford: Oxford University Press. (Response to Devlin (1965), arguing that he misrepresents the judicial process and misunderstands the relationship of the individual to society.)

Locke, J. (1689) *Locke's Letter Concerning Toleration in Focus*, ed. J. Horton and S. Mendus, London: Routledge, 1991.(Contains Locke's *Letter* and includes a number of articles discussing the tenability of his efficiency argument.)

Martin, M. (1987) *The Legal Philosophy of H.L.A. Hart*, Philadelphia, PA: Temple University Press.(Chapter 8 is a lucid discussion of the Hart-Devlin debate.)

Mill, J.S. (1859) *Mill's Essay On Liberty in Focus*, ed. J. Gray and G.W. Smith, London: Routledge, 1991. (Includes Mill's *On Liberty* together with a number of articles on the interpretation of the harm principle.)

Oakeshott, M. (1993) *Morality and Politics in Modern Europe*, ed. S.R. Letwin, London: Yale University Press. (An incisive and authoritative account of the limited 'office of government' as understood in the tradition of individualism.)

Law, philosophy of

Law has been a significant topic for philosophical discussion since its beginnings. Attempts to discover the principles of cosmic order, and to discover or secure the principles of order in human communities, have been the wellsprings of inquiry into law. Such inquiry has probed the nature and being of law, and its virtues, whether those that it is considered as intrinsically possessing, or those that ought to be cultivated by lawgivers, judges or engaged citizens. A dialectic of reason and will is to be found in philosophical speculation about the underpinning principles of law. On the one side, there is the idea that the cosmos itself, and human society too, contain immanent principles of rational or reasonable order, and this order must be capable of discovery or apprehension by rational (or 'reasonable') beings. On the other side, there is the view that order, especially in society and in human conduct, is not found but made, not disclosed to reason but asserted by acts of will. Either there is a 'law of reason - and nature' or there is a 'law by command of the sovereign - or of God'. A third possible element in the discussion may then enter, that of custom as the foundation of law.

Implicit in the opposition of reason and will is the question of practical reason: does reason have a truly practical role concerning ultimate ends and nonderivative principles of action, or is it only ancillary to pursuit of ends or fulfilment of norms set by will? Alternatively, does reason already presuppose custom and usage, and enter the lists only by way of critique of current custom and usage? In either case, what is at issue is the very existence of such a thing as 'practical reason' (see Practical reason and ethics). For law is about human practice, about societal order enforced and upheld. If there can be a law of reason, it must be that reason is a practical as well as a speculative faculty. The radically opposed alternative sets will above reason, will oriented to the ends human beings happen to have. Norms and normative order depend then on what is willed in the way of patterns for conduct; reason plays only an ancillary part in the adjustment of means to ends.

A further fundamental set of questions concerns the linkage of the legal with the political. If law concerns good order, and if politics aims at good order in a polity, law must be a crucial part of politics; but in this case a subordinate part, for politics determines law, but not law politics. On the other hand, politics may be considered at least as much a matter of actual power-structures as a matter of speculation about their beneficial use for some postulated common good. In the latter case, we may see law as that which can in principle set limits on and control abuses of power. Politics is about power, law about the shaping and the limiting of power-structures. The issue then is how to make law a master of politics rather than its servant.

1 Law as reason

In the *Republic*, Plato depicts Thrasymachus, proponent of the thesis that justice is the will of the powerful, as being refuted comprehensively by Socrates (see Plato §14). The refutation postulates a human capability to discern principles of right societal conduct independently of any formal enactment or legislative decision made by somebody with power. These principles in their very nature are normative, not descriptive. In Aristotle, the same general idea emerges in the form of noticing that whereas much that is observed as law is locally variable and arbitrary, there appear to be fundamental common principles across different polities. Some principles may then be legal simply 'by enactment', but others seem to be so 'by nature'. Explorations of the nature of humans as rational and political animals may then help to underpin the idea of that which is right by nature, but that exploration is more the achievement of Aristotle's successors in the Stoic tradition than of himself (see Stoicism §18). Roman jurists adapted some of the Stoic ideas of natural law in their expositions of the civil law, and subsequently, for medieval and early modern Europe, the existence of the Justinianic (see Justinian) compilation of the whole body of Roman law was held by many thinkers to embody in large measure the promise of law as 'written reason' (see Roman law; compare Gaius; Bartolus of Sassoferrato; Pothier, R.J.).

In any event, the greatest flowering of the Aristotelian idea came with its fusion into the Christian tradition in the work of Thomas Aquinas (§13), hugely influential as this has been in the developing of Catholic moral theology in the succeeding centuries. After at least a century of relative neglect among legal scholars, especially in the English-speaking world, the last quarter of the twentieth century has seen a strong revival of the Thomistic approach in the philosophy of law (see Natural law), with contemporary thinkers developing the idea of the basic goods implicit in human nature, and showing both how these can lead to the elaboration of moral principles, and

then how positively enacted laws can be understood as concretizations of fundamental principles.

In the seventeenth century, other strands of essentially the same idea had led to the belief, for example, of Hugo Grotius, that basic principles of right conduct and hence of human rights are themselves ascertainable by intuition and reason (compare Pufendorf, S.; Stair, J.D.). Kant's representation of the principles of practical reason is the classical restatement of this position in its most philosophically rigorous form (see Kant, I. §§9-11; Kantian ethics §1).

In a wide sense, all these approaches may be ascribed to rationalism, as contrasted with voluntarism (see Rationalism; Voluntarism). For they treat law, or its fundamental principles, as discoverable by rational and discursive means, independently of the intervention of any legislative will. They do not, of course, deny the need for legislative, or adjudicative or executive, will. Even if fundamental principles stand to reason, their detailed operationalization in actual societies requires processes of law-stating, law-applying and law-enforcement. But the issue is whether these are fundamentally answerable at the bar of reason and practical wisdom (*prudentia*), or not. To the extent that they are so answerable, we have a concept of some 'higher law', some law of reason, by which to justify, to measure and to criticize the actual practice of human legal institutions. If the rational derivation of this depends in some way on a teleological understanding of human nature and its relation to the creator and the rest of the created universe, we may reasonably enough call this a 'law of nature' or 'natural law'.

2 Law as will

But there is another possible account of higher law. It can be thought of as a law laid down by God for his creation. The divine will, not the divine reason, must be the source of law. It cannot be for created reason to presume to judge of the creator's wisdom. The omnipotence of the creator entails that the law will be whatever the creator wills it to be, and to be law by virtue of that will, not by any independent reason and nature of things. Indeed, the nature of things will be just what the creator wills it to be, and the names of things will be matters of convention derived from human linguistic usage. Concepts are not essences that guide us to essential meanings. Nominalism and voluntarism are inevitable bedfellows (see Nominalism).

It is therefore inaccurate to suppose that the theory of natural law as a kind of higher law presupposes rationalism. There can indeed be a voluntaristic species of 'natural law', though the voluntaristic tradition will more likely speak of 'divine law' or 'God's law' than of natural law *simpliciter* (see Austin, J.). Moreover, one element in the religious upheavals associated with the Reformation was an insistence on the need for unmediated regard to the (scripturally revealed) divine law, rather than to the custom or tradition of sinful human institutions such as the Church. It is not for fallen human reason to set itself above or even beside the revealed will of God. But that revealed will must be received as a law binding above all others.

In this state of things it becomes questionable whether to accept any human law at all; and, on the voluntarist hypothesis, to see how law other than God's law can have any obligatory force at all. To the saving of human law there are only two possible moves: either it must be shown that God in fact wills our obedience to the very kings and other superiors we actually have (as in the theory of 'the divine right of kings'), or it must be the case that the binding will arises from the consent of human beings themselves, expressed through some original social contract. The divine will then enters the picture only to the extent of making obligatory the fulfilment of compacts voluntarily agreed, a point to which may be added a grimly Hobbesian acknowledgement that covenants without swords are but words, so the true binding force of the obligation of the law will derive from the effective might of the very ruler whom the social compact institutes in that office (see Hobbes, T. §§6-7). In this Hobbesian form, natural law has practically reached a vanishing point (though Locke's response envisages the state of nature as governed by reason in the form of a law of nature, grounding presocietal rights of human beings to life, liberty and estate (see Locke, J. §§9-10). The greatest legal expression of the Lockean vision of law, applied to expounding the English common law, is in the work of Sir William Blackstone. The *coup de grâce* was administered by Hume and Bentham, the latter having as his particular target Blackstone's work. They argue that the social contract is a fifth wheel on the carriage in either Hobbesian or Lockean form, since all the reasons that there are for obeying the law that we have supposedly agreed to apply with equal force even if we did not agree to it, and there is no evidence anywhere of any such agreement as a historical phenomenon (see Bentham, J. §6).

3 Law as custom

Whence then comes the law? Hume ascribes it to convention and custom primarily, coupled with reflection upon the pleasing quality (the utility) of rigorous observance of customary norms (see Hume, D. §5). Bentham and Austin restrict the role of custom or 'habit' to the issue of obedience. Whoever is habitually obeyed by the many in a numerous society is in a position to enforce their commands by effectively coercive sanctions up to and including death. Thus do they differentiate the positive law from other forms of so-called law such as scientific law, laws of honour, or personal moral codes. Law is such by command of a sovereign, the one habitually obeyed who habitually obeys no other (see Sovereignty).

Legal positivism of this stamp is an easy bedfellow with political utilitarianism, and programmes of legal reform. Codification of law is an associated ambition, justified on utilitarian grounds (see Utilitarianism). Codification is also a distinctive phenomenon of the early nineteenth century, product of the Enlightenment critique of the old customs of the *ancien régime*, though also of spadework in the exposition of civil law partly achieved under the aegis of late legal rationalism. After the Code Napoléon, promulgated in France in 1804, there followed a century of codification and legislative modernization of law in many places, and with this characteristically went approaches in legal philosophy that stress the essential emergence of law from a sovereign's will, or the will of the state as a rational association (in Hegelian vein; see Hegelianism). Nevertheless, this movement produced its own counter-movements, stressing the importance of the spirit of the people as the basis of law (see Savigny, F.K. von; Bryce, J.; Jurisprudence, historical), or more prosaically locating it primarily in custom, a view particularly popular in the context of the common law (see Common law; Selden, J.).

Twentieth-century critics of classical positivism accuse its authors of confusing 'commands' with 'binding commands' (see Kelsen, H.; compare Weyr, F.) or of mislocating the roots of legislative authority in mere 'habit', rather than in the 'internal point of view' of those for whom the system within which authority is exercised has normative force (see Hart, H.L.A.). The Kelsenian version of positivism rests it on the necessary presuppositions for a value-free science of law, and other thinkers have pursued further the question of 'legal science' (see Bobbio, N.); the Hartian version rests it on the customs of at least the official and political classes in a state, whose practices concerning the recognition of certain criteria for the validity of legal rules define the ultimate 'living constitution' of a state, its 'rule of recognition' (see Legal positivism).

A notable offshoot of or development from positivistic legal study has been the development of ever-more rigorous approaches to conceptual analysis (see Legal concepts) and categorization, seeking to account for the use of concepts like 'duty', 'right', 'ownership' and others in the framework of general legal norms (see Norms, legal). Hohfeld's analysis of 'fundamental legal conceptions' (see Hohfeld, W.N.) has had many followers and critics, and contemporaries in other traditions have taken a somewhat more psychologistic approach to the task (compare Petrażycki, L.). Reflection on legal concepts as institutions or 'institutional facts' has led to developing an 'institutional' theory of law that transforms what was originally a naturalistic conception into a positivistic one (see Institutionalism in law; Weinberger, O.).

4 Laws and values

One way or another, whether in voluntaristic versions or in those that place more weight on customary or institutional aspects of law, nearly all forms of or approaches to legal positivism have insisted on the strong value-relevance of positive law. The matter of doubt has not been 'ought laws to be just?', but whether their being just is a condition of their being genuinely legal. The 'scientific' character of pure legal analysis has indeed been contrasted with the exercise of moral judgment or moral sentiment, or the engaging in ideological argumentation, that is involved in the critique of law as unjust or otherwise unsatisfactory from the viewpoint of human needs and aspirations. Some, however, have thought that critique itself can have a scientific or at least an objective basis, grounded in the fundamentals of human nature. Classical utilitarianism and nineteenth-century law reform are a case already noted; they had successors in the 'jurisprudence of interests' (see Jhering, R. von; Pound, R.), and, albeit with certain qualifications, in the later twentieth-century 'economic analysis of law' (see Law, economic approach to).

The need to subject law to critique is obvious from many points of view, none more urgently than that which takes note of the burdensome impact of legal sanctions on human happiness and liberty. If laws characteristically carry punishments for their infraction, some theory to justify penal institutions is called for (see Crime and punishment). Whether there are any abstractly stateable limits to the legitimacy of interference with liberty through legal

intervention has been another heated debate (see Law and morality).

Nevertheless, the positivists' claim that they can combine an a-moralistic conceptual analysis of law and its institutions with a readiness for critique of actual laws on moral and political grounds, and with a last-resort readiness to disobey or defy the law when it is unjust to an extreme, has been doubted by some. Gustav Radbruch felt himself driven by his experience of the Nazi years (and also, perhaps, by the implications of the radical voluntarism of Carl Schmitt) to abandon such a claim and to insist on a conceptually necessary minimum of basic justice in anything we can recognize as 'law' at all. The interpenetration of equity with law, and the interweaving of ideas of justice, equity and law, can be taken to point to a similar moral (see Justice, equity and law), and idealistic approaches to legal theory give a deeper grounding for such an approach (see Legal idealism).

5 Law as politics

However one takes one's stand on will against reason, or on natural law against legal positivism, most of the theoretical approaches so far considered give some way of accounting for the independent existence of law as a distinct social phenomenon. Law's independence, at least when underpinned by an independent judiciary, has been held to promise the possibility of effective control over arbitrary state action while at the same time guaranteeing at least the justice of formal equality to citizens and the degree of predictability allegedly desired by modern rational subjects. Here we have the 'rule of law' ideal that demands government under the forms of law and law in the form of clearly identifiable rules (see Rule of law (Rechtsstaat); compare Dicey, A.V.; Fuller, L.L.). Yet the mere existence of some body of sacred or secular texts embodying rules of law is not enough for any socially realistic account of law, or for any politically persuasive vision of the rule of law (see Social theory and law; compare Millar, J; Renner, K.). The statute book is not self-applying or self-interpreting (compare Wróblewski, J.). To secure the rule of law it is necessary to have prospective rules published to all. But, as L.L. Fuller points out, it is necessary that they be interpreted in a reasonable and purposive way, and faithfully carried into action by the officials of the state whose rules they are. How is this to be secured?

Many schools of thought, chief among them the realists (see Legal realism) in Europe (see Olivecrona, K.; Ross, A.) and in the USA (see Holmes, O.W., Jr; Llewellyn, K.N.; Frank, J.), have stressed the widely discretionary character of legal interpretation, both in relation to the general rules of the law, and in relation to the categorization of fact-situations as subsumable under the law for one purpose or another. On inspection, 'facts' can turn out as elusive as 'laws', and the study of legal processes of proof assumes a certain urgency (see Legal evidence and inference). All in all, it is a serious and difficult question to discern what, if anything, can render decisions reasonably 'reckonable' given the broad discretion vested in those who interpret the law.

One form of response has been to find that law is reckonable not on the basis of the official rules and standard doctrine, but rather on the basis of the 'situation sense' of a judiciary with a common understanding of political and policy objectives underlying law. These insights of the 'realists' have been carried forward more boldly by contemporary feminist jurisprudence, one version of which finds social prejudice directing law through the biases of judges. Another version locates an inner masculinity in the legal rules themselves, even and especially at their most abstract; the asserted values of objectivity and impersonality ultimately come under question as presumptions of doubtful desirability (see Feminist jurisprudence).

Within more mainstream jurisprudence the developed response to realism has been to work out extended theories of the rule of law, acknowledging that law is more than positive rules but arguing for the existence of other mechanisms within law controlling the role of substantive elements in decision-making (see Legal reasoning and interpretation). Such responses find a certain coherence within law, but by contrast the more developed critical (including critical feminist) approaches argue that there are central fractures and fault lines within the law, reflecting ultimately competing political visions of human association, often summed up as individualism versus community-values (see Critical Legal Studies). Ronald Dworkin's argument for coherence and integrity in law evokes the idea of an interpretive community, but seems too readily to assume that for any actual legal order there can be found a single consensual interpretive project, even in principle (see Dworkin, R.; compare Legal hermeneutics).

Taking an overall view, the project of establishing the rule of law as an independent base for the critique and control of state action is put in serious doubt, since interpretation is through-and-through political; and appeals to

the rule of law can themselves be moves in a political game, expressions of ideology rather than of higher values. It may be that in the end legal philosophy is faced, today as at its beginnings, with this dilemma: either legal reasoning and moral reasoning have that kind of in-principle objectivity proposed by natural law theory in its rationalist versions, or the theatre of law is simply a theatre presenting endlessly the power-play of rival wills and visions of the good. Many have sought a third way, not yet with acknowledged success.

See also: Legalist philosophy, Chinese; Fa; Halakhah; Law and ritual in Chinese philosophy; Law, limits of; Legal discourse; Law, Islamic philosophy of; Villey, M.

BEVERLEY BROWN
NEIL MacCORMICK

References and further reading

Harris, J.W. (1980) *Legal Philosophies*, London: Butterworth.(Straightforward and well-written introduction to issues and schools of thought in philosophy of law.)

Hayek, F.A. (1973, 1976, 1979) *Law, Legislation and Liberty*, London: Routledge & Kegan Paul.(Three-volume critique of the pretensions of constructivist rationalism, whether of utilitarian positivists or of rationalistic naturalists, in favour of a 'critical rationalism' reflecting on the accumulated societal wisdom implicit in an evolved and essentially customary law, and developing an account of the rule of law on this basis.)

Kelman, M. (1987) *A Guide to Critical Legal Studies*, Cambridge, MA: Harvard University Press.(Readable and sympathetic account of, and contribution to, the 'critical' approach that regards all legal activity as intrinsically political - and ideological.)

Kingdom, E. (1991) *What's Wrong with Rights?*, Edinburgh: Edinburgh University Press.(An interesting collection of essays putting a moderate feminist case against the biases inherent in received legal categories.)

Lloyd of Hampstead and Freeman, M. (1985) *Lloyd's Introduction to Jurisprudence*, London: Stevens, 5th edn. (Another useful introduction, supported by many selected texts for reading.)

Rommen, H. (1947) *The Natural Law: a Study in Legal and Social History and Philosophy*, St Louis, MO and London: Herder Book Company.(Full and careful statement of implications and applications of natural law theory from a Catholic point of view.)

Shiner, R. (1992) *Norm and Nature: the Movements of Legal Thought*, Oxford: Clarendon Press.(A more challenging text that deals with the tensions in legal thought between positivist - or voluntarist - and anti-positivist approaches, concluding that the dialectic between them contains a truth available from neither on its own; advanced reading.)

Law, William (1686-1761)

William Law is popularly known for a single devotional work, A Serious Call to a Devout and Holy Life (1728), but was the author of seventeen other treatises concerned with many of the philosophical and theological issues of the earlier eighteenth century in England. Early on he embraced the thought of Nicolas Malebranche, and later the mystical philosophy of Jakob Boehme. Law's methodology is an awkward mixture of these influences, for he wanted to find reason in the 'creaturely Spirit' of the natural life (The Spirit of Love, 1752). He seems finally to have abandoned reason in favour of Boehme's 'metaphysical scheme', which enables him to discover meaning and coherence in a post-Enlightenment world.

William Law lived mostly in King's Cliffe, Northamptonshire, where he was born and died. He studied at Emmanuel College, Cambridge, where he was elected a fellow in 1711, but resigned this position in 1716, being unable to take the oath of allegiance to the Hanoverian George I. He was thus one of the 'nonjurors', who believed that the Stuarts were by Divine Right the legitimate monarchs of England and consequently true heads of the Church. Accordingly deprived of his college fellowship, he passed the next years variously in study and writing and as a tutor in the Gibbon household at Putney, Surrey, where he instructed the father of the great historian. About 1740, he retired to King's Cliffe at the invitation of Hester Gibbon, the historian's aunt, and one Mrs Hutcheson. Together they formed a kind of religious community, and in these circumstances Law, who never married, continued his reflective and orderly life to the end.

Law's writing is often polemical. *Three Letters to the Bishop of Bangor* (Benjamin Hoadly) his first work, written between 1717 and 1719, was a defence of the apostolic succession of bishops, in which he asserts the spiritual authority of the Church against an Erastian prelate who urged the ecclesiastical supremacy of the state (1762, 1:15). Later Law attacked Bernard Mandeville's *Fable of the Bees* (1714), lamenting 'the miserable Fruits of *Free-thinking*, which…tend not only to set us loose from the Regards of Religion, but to destroy whatever is reasonable, decent, or comely in human Nature' ([1723] 1762: vol. 2, 49) (see Mandeville, B.). Rather belatedly, Law joined the attack on the Restoration theatre in *Absolute Unlawfulness of the Stage-Entertainment fully Demonstrated* (1726). In his work Law argues persistently for right action. Laid out in terms of lively vignettes of persons who live frivolous or arid or else pious lives, his *Serious Call to a Devout and Holy Life* (1728) provides a systematic plan for ordering one's moral and devotional life.

Given Law's devotion to the reasonable spiritual life, his development of a mystical theology may seem curious, yet he believed he was recovering deeply felt, if elusive, truths. Never systematic, Law's philosophical ideas are expressed in generalities and his theological views often lead to untraditional conclusions. From 1740 onwards, he gave himself over to the study of Jakob Boehme, and his later writing, especially *The Spirit of Prayer* (1749) and *The Spirit of Love* (1752), is filled with a sense of cosmic goodness. The later style may be well suited to its subject; but it is the plain and straightforward Law of *A Serious Call* that Dr Samuel Johnson pronounced 'the finest piece of hortatory theology in any language' (Boswell 1791: vol. 1, 390).

See also: Malebranche, N.; Tindal, M.

PAUL G. STANWOOD

List of works

Law, W. (1762) *Complete Works*, London: Richardson; privately repr. for G. Moreton, Brockenhurst, Hants and Canterbury, Kent, 1892-3, 9 vols.(The standard edition, which collects the individual works, all but the last of them published by Richardson of London. In addition to those works mentioned in the text and below, the *Complete Works* includes *A Practical Treatise upon Christian Perfection* (1726), *The Case of Reason…In Answer to a Book, entitled 'Christianity as old as the Creation'* [by Matthew Tindal] (1731), *Letters to a Lady Inclined to Enter into the Communion of the Church of Rome* (1731-2), *A Demonstration of the Gross and Fundamental Errors of a Late Book, Called 'A Plain Account of the Nature and End of the Sacrament of the Lord's Supper'*[probably by Bishop Hoadly] (1737), *The Grounds and Reasons of Christian Regeneration* (1739), *An Earnest and Serious Answer to Dr Trapp's Discourse of the Folly, Sin and Danger of being Righteous Over-Much* (1740), *An Appeal to All that Doubt, or Disbelieve the Truths of the Gospel…To which are Added, Some Animadversions upon Dr. Trapp's Late Reply* (1740), *The Way to Divine Knowledge* (1752),

A Short but Sufficient Confutation of the Reverend Dr. Warburton's Projected Defence (1757), *Of Justification by Faith and Works* (1760), *A Collection of Letters* (1760) and *An Humble, Earnest, and Affectionate Address to the Clergy* (1761).)

Law, W. (1723) *Remarks upon a Late Book, Entituled 'The Fable of the Bees'*, Bristol: Thoemmes Press, 1992. (An attack on Bernard Mandeville's book.)

Law, W. (1728) *A Serious Call to a Devout and Holy Life*, repr. in *A Serious Call to a Devout and Holy Life and The Spirit of Love*, ed. P.G. Stanwood, New York: Paulist Press, 1978.(Law's best-known work.)

Law, W. (1752) *The Spirit of Love*, repr. in *A Serious Call to a Devout and Holy Life and The Spirit of Love*, ed. P.G. Stanwood, New York: Paulist Press, 1978.(Law sought to find reason in the 'creaturely Spirit' of the natural life.)

References and further reading

Boswell, J. (1791) *The Life of Samuel Johnson, LL.D.*, London: Dent, 1949, 2 vols.(Quotes Samuel Johnson on Law.)

Clarkson, G.E. (1992) *The Mysticism of William Law*, New York: Lang.(Law's indebtedness to Boehme and to mystical theology.)

Cornwall, R.D. (1993) *Visible and Apostolic: The Constitution of the Church in High Church Anglican and Non-Juror Thought*, Newark, NJ: University of Delaware Press.(Situates Law within the High Church and Nonjuring movement of the eighteenth century; includes bibliographical references.)

Grant, P. (1985) 'William Law's *Spirit of Love*: Rationalist argument and Behemist Myth', in *Literature and the Discovery of Method in the English Renaissance*, Athens, GA: University of Georgia Press.(Surveys the influences on Law's thought.)

Hopkinson, A.W. (1948) *About William Law: A Running Commentary on his Works*, London: SPCK.(Chapters on Law the man, the controversialist, the moralist, the mystic, the theologian.)

Rudolph, E.P. (1980) *William Law*, Boston, MA: Twayne.(Introductory guide to Law's life and work.)

Talon, H. (1948) *William Law: A Study in Literary Craftsmanship*, New York: Harper.(Concerned principally with Law's literary style.)

Walker, A.K. (1973) *William Law: His Life and Thought*, London: SPCK.(Comprehensive survey of Law's life, with summaries of his eighteen volumes.)

Warren, A. (1978) 'William Law: Ascetic and Mystic', intro. to *A Serious Call to a Devout and Holy Life and The Spirit of Love*, ed. P.G. Stanwood, New York: Paulist Press; repr. in *In Continuity: The Last Essays of Austin Warren*, ed. G.A. Panichas, Macon, GA: Mercer University Press, 1996.(A general survey of the life and work.).

Young, B.W. (1994) 'William Law and the Christian Economy of Salvation', *English Historical Review* 109: 308-22.(Studies Law's soteriology.)

Laws, natural

It is widely supposed that science aims to identify 'natural laws'. But what are laws of nature? How, if at all, do statements of laws differ from 'mere' general truths which include generalizations true only 'accidentally'? Suppose, for example, it happens to be true that all iron spheres (past, present and future) are less than 1 km in diameter. Contrast this with the truth of 'all electrons are negatively charged'. There seems to be a clear intuitive distinction between these two truths, but is there any principled distinction between them that can be drawn and defended?

This has been the traditional focus of philosophical attention concerning laws of nature, and basically two mutually opposed philosophical accounts have been developed. According to the first account, there are real necessities in nature, over and above the regularities that they allegedly produce (whether or not these regularities are held to be observable), and law-statements are descriptions of these necessities. According to the second account, there are no necessities but only regularities (correlations, patterns) and laws are descriptions of regularities (though perhaps not of any regularity but only of the most basic or most general ones). There are significantly different variants of each account; and also positions that altogether deny the existence of general laws (or deny that science should aim to describe them).

Any one of these accounts, if it is ultimately to be coherent and defensible, has to successfully address four interrelated issues: the meaning of a law statement - the semantic issue; the fact to which a law statement refers and which makes it true - the metaphysical issue; the basis on which claims to know a law are justified - the epistemological issue; the capacity to explain adequately the variety and roles of scientific laws - the explanatory issue.

In attempting this task, each of the available accounts faces its own distinct difficulties. For example, if there are necessities in nature, as the first account claims, how exactly do we identify them: how can we tell which of the inductively confirmed regularities are laws? On the other hand, if there are only regularities, as the second account claims, does this mean that our intuitions and scientific practices are awry and that there really is no distinction between laws and accidental generalizations?

The difficulties facing all extant accounts become even more marked when we face up squarely to the surprisingly wide variety of (putative) laws supplied by current science and to the complexity of the relations between those putative laws and regularities and causes.

1 Laws in science

Science exhibits a diversity of kinds of law (law statements). Causal laws specify succession: B causally follows A in time; sunlight causes sunburn (see Causation). Laws of synchronic coexistence specify atemporal relationships: for a gas at equilibrium, temperature equals pressure times volume (in suitable measurement units). Basic structural laws specify the fundamental principles defining a theory framework: spacetime world lines of individuals neither intersect nor branch (that is, no two things can be at the same place at the same time and nothing can be at two places at once). Derived structural laws are laws of diachronic coexistence, A at time $t1 = A$ at time t2: energy is conserved in an isolated system. Functional-analysis laws connect functional capacities of complex systems: teleological systems require negative feedback to maintain goal orientation.

Some laws are inclusionary: all electrons have charge e; some are exclusionary: nothing travels faster than the speed of light *in vacuo*. Some laws are deterministic: $f = ma$, Newton's second law of motion, while others are statistical: given an incident UV photon, a skin cell has probability P of dying (see Determinism and indeterminism; Probability, interpretations of). Some laws interrelate localized individuals (for example, the gas coexistence law above, $PV = T$), others concern field states (for example, the superposition principle, which has no equivalent among individuals). Symmetries, that is, invariance constraints, ground derived structural laws and may lie behind causal laws (quantum field theories) (see Conservation principles; Field theory, quantum).

There is no single or simple relationship of laws to causes (of law statements to causal statements). The causal gravitational action of the sun directly produces the planetary motions, but while causal collision processes are ultimately responsible for gas behaviour only their statistically average behaviour at equilibrium is captured in the

constraint $PV = T$. Many laws, for example that all ravens are black, are the end product of many interacting causes (constrained by specific initial conditions). Some macroscopic succession laws cannot, at present, be clearly understood theoretically as being the result of interacting microscopic causes (for example the second law of thermodynamics, because in current theory the thermodynamic, but not the microscopic dynamic, law is time-irreversible - see Thermodynamics). Some laws derive from the inoperativeness of causes (the conservation laws); others have no simple relation to causes (for example, structural laws, the quantum Pauli exclusion principle).

Nor is there any single or simple relationship between laws and regularities (between law statements and descriptions of regularities). Kepler’s laws of planetary motion specify simple, if empirically approximate, regularities. Newton’s first law of motion, given the gravitation law, is empirically never instantiated and so has no directly associated regularities, but does specify a higher-order regularity of convergence on first-law behaviour as system forces decrease. There is evidently no analytic solution to the Newtonian gravitational three-body problem, so there is no easily specified associated empirical regularity; yet the earth-moon-sun system exhibits a solution. Statistical relationships do not directly issue in any particular empirical regularities, only in approximate higher-order ones (laws of large numbers). Further, a nonlinear system showing chaos will never repeat its sequence of states, yet its dynamics is law-governed, indeed deterministic (see Chaos theory). Here again there is no useful logical formula that describes the dynamic pattern (as a pattern); we can only explore it through computational approximation from the generating equations. Thus evolution and history might show no repeated patterns, yet unfold according to some very complex (largely unknown) nonlinear dynamic laws.

Conversely, empirical regularities have diverse relations to laws. Consider two clocks exhibiting a regular connection between the readings on their two faces. This regularity may arise from (a) a direct physical connection between them whereby the face reading of one sets (causes) the face reading of the other; (b) an immediate common cause, a single human setting both clocks to the same time; (c) a mediate common cause, two separate people separately setting their clocks to the same socially shared time, ultimately generated from a common standard clock; (d) two mutually independent random quantum events initially coincidentally setting the clocks to the same time; (e) at each second the two clocks coincidentally resetting to the same times as in (d).

There are laws (law statements) referring to (i) real systems (for example, Newton’s second law of motion), (ii) idealized systems which could not be real but which approximate real systems in some respect (for example ’ $PV = T$’, which assumes a gas of point particles - see Fictionalism; Idealizations; Models), (iii) systems which could be real but are never instantiated (for example, Newton’s first law, given gravitation), or only approximately instantiated (Kepler’s laws in our solar system). Groups (ii) and (iii) nonetheless both render the mathematical structure of the corresponding natural laws simpler and exhibit the structure of component causes (for example, forces).

Within the explanatory issue, then, call achieving a coherent account of the diversity of kinds of law the ’diversity problem’, understanding the complex relations of laws to regularities the ’pattern problem’ and understanding the diverse roles of idealization in laws the ’idealization problem’. Philosophical theories of science have often ignored or simplified the distinctions among, and rich diversity of, scientific laws; a good philosophical account of natural laws would itself bring order into these complexities.

Scientists use laws to explain both other (’less basic’) laws and particular events (see Explanation §1), to deduce what will happen (predict) and what would have happened had (counter-factually) circumstances been different, to separate complex causes into their components and to distinguish between accidental and law-based empirical regularities, to decide when experimental and inductive inference procedures are justified (see Confirmation theory §1; Experiment; Inductive inference §4), to decide what basic properties there are, and so on. An account of the role of laws is thus central to a philosophical account of science. Call providing such an account the ’roles problem’.

In sum, the explanatory issue comprises the diversity, pattern, idealization and roles problems. Responses are based on coordinated proposals for the semantic, metaphysical and epistemological issues.

2 Two philosophical accounts

Necessitarian theories. We make sense of our experience both through introducing causes which produce or

necessitate their effects, and through counter-factual reasoning about possible though not actual cases governed by laws which still continue to constrain what is possible. These are the intuitive sources of necessitarian accounts, which introduce a metaphysics of necessary connections as truthmaking conditions for law statements.

Proponents of nonlogical necessitarian theories - developed earlier this century by Johnson, Kneale and von Wright, among others - hold that both laws and true accidental regularities are confirmed by factual observation, so are in that sense contingent, but the law 'all A are B' also rules out the possibility of a non-B A, in analogy to the laws of logic specifying the logically possible. The view that laws do possess logical necessity - noncontingent, logical necessitarianism - was defended earlier this century within the idealist rationalist tradition, including Blanshard,Ewing and Stout. Recently versions have been revived of both nonlogical necessitarianism, for example Armstrong (1983), and logical necessitarianism, Swoyer (1982).

The central metaphysical problem is to characterize necessity. Is it (i) an object or (ii) a relation? Does it connect (a) objects or (b) properties, powers dispositions? There is a traditional objection to (i): how can adding an object create a necessary connection? (Notice, however, that quantum theory represents forces as something like exchange of special objects, bosons.) There is a traditional objection to (ii): it involves developing a metaphysics of real universals (properties, powers, or dispositions) but this is implausible/incoherent. Will statistical laws require the postulation of special propensities as their truthmakers, and what are these? Would basic negative laws require real negative universals? Would uninstantiated and idealized laws require unactualized and idealized universals? Arbitrary positive answers trivialize this approach; but what principled basis for response is there? (These cases are also troubling for version (i).)

The necessitarian semantic problem is to provide a meaning for law statements that renders them logically stronger than statements of regularity. Though there are many modal logics, it is controversial whether the truth conditions of their statements identify appropriate natural necessities. It has proved similarly difficult to construct a satisfying logic of counterfactual conditionals, which natural necessities distinctively support, and it is controversial whether agreement on a satisfactory formal theory could do more than describe proper inferential practice, assuming natural necessities somehow otherwise identified (see Counterfactual conditionals).

The central necessitarian epistemological problem is that of identifying necessities: how to tell which among inductively confirmed regularities are laws? Any principle that would render inferences from evidence to laws deductively valid (for example that nature is finite and causally uniform) must itself be justified in the same way (vicious circularity) and arguably there are no observable necessitating connections. David Hume drew the sceptical conclusion that no claim to know which evidentially confirmed regularities are laws is ever justified, our scientific knowledge being confined to affirming past regularities and relying on them out of 'habit', without rational warrant. Appeal to some scientific methodology *per se* does not alleviate this problem.

Eliminative necessitarians respond by retaining necessitarianism while dismissing laws as part of the metaphysics of science. Cartwright (1983) argues that laws are so idealized and indirectly related to empirical phenomena that they cannot be regarded as real, while individual causes are real and the basis for scientific reasoning. Eliminationism faces its own difficulties with the metaphysical and semantic issues, but especially with the epistemic and explanatory issues, since it must show how scientific practice can be properly understood without effective recourse to laws, while simultaneously identifying and using causes.

Regularity theories. Historically, objections to necessitarianism, especially epistemological objections, have motivated the development of regularity theories of laws. According to these theories, laws in nature are nothing more than matter-of-fact correlated event patterns. There is no metaphysical distinction between accidental and lawful regularities. The law-likeness of accepted scientific generalizations describing these patterns is at best organizational, a matter of our attitude to, or use of, these statements as 'inference licenses' (Lange 1993), which support habitual, because pragmatically useful, practices of induction, prediction and counter-factual reasoning. Often law-like status is held to be determined by depth of entrenchment in science - (roughly) the wider the generalization, the deeper its entrenchment and the more law-like its accorded status.

The contingent, nonlogical regularity version is the simplest, most common one and its supporters include Braithwaite (1953) and Hesse (1980) among many. It has no need for special necessities in nature, and avoids special semantics for law statements by making the simple logical form ' $(x)(Ax \rightarrow Bx)$' the universal paradigm.

Conversely, however, it is difficult to provide an account of our practices of counterfactual inference since, given '-A*a*', ' $(x)(Ax \rightarrow Bx)$' is vacuously instantiated independently of whether or not 'B*a*' holds. It is also difficult to explain how there can be uninstantiated laws, whether or not pertaining to idealized entities, since all uninstantiated extensional conditional statements are vacuously true. More generally, it is difficult to see how a simple regularitiy account could explain the complex relationship between laws and empirical regularities; regularities are neither necessary nor sufficient for laws.

In the light of the difficulties facing nonlogical regularity theories, some philosophers sought to retain their metaphysical simplicity by adopting instead the view that laws are contentless conventions, linguistic meaning definitions, and necessarily true on that account. They were held to be pragmatically useful in the organization of empirical information, in roughly the way that any other linguistic practice is, but no more. Earlier Poincaré, Mach and Reichenbach defended versions of this noncontingent, logical regularity view and later conventionalist debates echo them (see Conventionalism). This theory provides a sense of necessity to laws compatible with a regularity metaphysics, but only at the cost of eliminating their empirical content. It is then hard to explain the content of scientific theories and hence also to make sense of scientific method. How can we explain why laws are so useful, and when they are useful, if they are only conventional definitions? Why should they ever be changed? Why are there so often no viable scientific alternatives to a theory, if alternatives are created merely by linguistic redefinition?

Another, opposite, reaction to the same difficulties is to retain empiricism while dismissing laws as of no significance to an account of science; this is eliminative regularity theory. Van Fraassen (1989) argues that scientific theories are concerned with symmetry and continuity, not universality or necessity. His view is that laws are anachronisms, remains of an early modern necessitarian essentialist metaphysics, and not something at which science can, or should, aim. He recognizes the importance of theory in science but, pursuing empiricism, construes it largely pragmatically rather than realistically, rejecting all postulated modal connections as unwarranted metaphysics. While this view faces its own problems in responding to the metaphysical, semantic and epistemological issues, special interest attaches to the explanatory issue, since eliminativism needs to show how scientific practice can be properly understood without effective recourse to laws.

3 A note on laws and logical analysis

Philosophy of science in the twentieth century has been dominated by the analytic tradition deriving from logical empiricism and Frege-Hilbert-Russell logical analysis (see Analytical philosophy §2). This tradition emphasizes syntactical logical form, and analysis as the basic tool of philosophical inquiry. In these terms, the traditional issues surrounding natural law have been: (A) What is the syntactical logical form of law statements? (B) What distinction in logical form demarcates (statements of) laws from accidental regularities? These questions have not received satisfactory answers. For example, it has generally been held that law statements have at least to be unrestricted universal claims of the form 'all A are B'. But this claim is either substantive and false, for example statistical laws are not perspicuously of this form, having instead such forms as 'probability (B, given A) = P', or it is true but uninformatively weak, since nothing is excluded (even statistical, idiosyncratic and accidental relationships can still be forced into this form). Many philosophers, representing otherwise opposing positions about laws, now reject this syntactic search as the vehicle for addressing the semantic issue, treating the semantic structure as primary; for example Suppe (1989).

Formal analysis, whether syntactic and/or semantic, continues to be applied to other related problems. The idea of accidentality, for example, may be explicated (Popper 1990) as that of logical dependence on initial conditions (in some theory): the more condition-dependent a statement is, the relatively more accidental it will be (Sklar 1990). And the idea of entrenchment has been formally elaborated in various ways. For example, according to Piaget one does not have a natural law until it specifies a true pattern that is demonstrably (that is, mathematico-logically) complete, the ultimate laws of science being those of logic and mathematics themselves. Others attribute law-likeness only to science satisfying certain conditions, for example to the system of scientific truths having the best combination of simplicity and informativeness (Lewis 1973). Galileo's law of free fall by itself would not count in either case, but Newton's laws of motion plus gravitation plausibly would.

Here also belong analyses of law-likeness in terms of probability, subjective (Skyrms 1980; Urbach 1988) and objective (Mellor 1990), along with other formal analyses aimed at transcending various distinctions (Mormann

1994).

All of these, and the many other, formal analyses are mutually conflicting and highly controversial. Their relations to the competing regularity and necessitarian positions depends on exactly how they are developed. (Lewis and Popper, for example, claim to combine possible worlds semantics for necessity with regularity theory.)

4 Conclusion

The existence and nature of laws is a complex problem. The diversity of philosophical theories in response is matched only by the diversity of laws themselves. The reader will find many further themes recurring in the literature - for example: the analysis and status of possibility and possible worlds; laws and initial/boundary conditions; causes, capacities and laws; concepts, laws and generality - which cannot be pursued here. There is no uncontroversial theory of laws: all face difficulties. As fast as we scientifically unravel nature’s mysteries, so fast does the nature of that understanding become mysterious.

See also: Positivism in the social sciences; Pragmatism; Scientific method; Theories, scientific

C.A. HOOKER

References and further reading

Armstrong, D.M. (1983) *What is a Law of Nature?*, Cambridge: Cambridge University Press.(Good, readable review and critique of regularity, followed by defence of property nonlogical contingent necessitarian.)

Armstrong, D.M., Cartwright, N.D., Earman, J. and Fraassen, B.C. van (1993) ’Book Symposium: Laws and Symmetry’, *Philosophy and Phenomenological Research* LIII, 411-44.(Critique and defence of van Fraassen’s eliminativist regularity.)

Bigelow, J.C., Ellis, B.D. and Lierse, C. (1992) ’The World as One of a Kind: Natural Necessity and the Laws of Nature’, *British Journal for Philosophy of Science* 43: 371-88.(Logical, noncontingent necessitarian.)

Braithwaite, R.B. (1953) *Explanation*, Cambridge: Cambridge University Press.(Nonlogical, contingent, regularity.)

Byerly, H. (1990) ’Causes and Laws: The Asymmetry Puzzle’, in A. Fine, M. Forbes and L. Wessels (eds) *PSA 1990*, vol. 1, East Lansing, MI: Philosophy of Science Association.(Complexity of relations between causes and laws.)

Cartwright, N.D. (1983) *How the Laws of Physics Lie*, New York: Oxford University Press.(Eliminativist nonlogical, contingent necessitarian.)

Daly, C. (1994) ’Laws and Coincidences Contrasted’, *Analysis* 54: 98-104.(Accidental versus law-like patterns.)

Dretske, F.I. (1977) ’Laws of Nature’, *Philosophy of Science* 44: 248-68.(Nonlogical, contingent necessitarian.)

Earman, J. (1978) ’The Universality of Laws’, *Philosophy of Science* 45: 173-81.(Difficulty of defining generality for laws.)

Earman, J. (1984) ’Laws of Nature: The Empiricist Challenge’, in R.J. Bogdan (ed.) *D.H. Armstrong*, Dordrecht: Reidel.(Critique of nonlogical, contingent necessitarian.)

Fraassen, B.C. van (1989) *Laws and Symmetry*, Oxford: Clarendon Press.(Good critique of necessitarian and regularity, followed by development of eliminativist regularity; model theoretic framework for semantics of law statements.)

Harré, R. and Madden, E.H. (1975) *Causal Powers: A Theory of Natural Necessity*, Oxford: Blackwell.(Powers nonlogical, contingent necessitarian.)

Hesse, M. (1967) ’Laws and Theories’, in P.M. Edwards (ed.) *Encyclopaedia of Philosophy*, London: Collier Macmillan.(Contains discussion of empirical versus theoretical laws.)

Hesse, M. (1980) ’A Revised Regularity View of Scientific Laws’, in D.H. Mellor (ed.) *Science, Belief and Behaviour*, Cambridge: Cambridge University Press.(Nonlogical, contingent regularity.)

Hooker, C.A. (1992) ’Physical Intelligibility, Projection, Objectivity and Completeness: the Divergent Ideals of Bohr and Einstein’, *British Journal for the Philosophy of Science* 42: 491-511.(Field versus particle laws; invariance, objectivity, and laws.)

Jones, R. (1990) ’Determinism in Deterministic Chaos’, in A. Fine, M. Forbes and L. Wessels (eds) *PSA 1990*, vol. 2, East Lansing, MI: Philosophy of Science Association.(Laws and regularities in nonlinear systems.)

Lange, M. (1993) ’Lawlikeness’, *Nous* 27: 1-21.(Laws as inference licences; review of traditional formulations.)

Lewis, D. (1973) *Counterfactuals*, Oxford: Blackwell.(Nonlogical, contingent regularity; counterfactual reasoning in possible worlds context.)

Mellor, D.H. (1990) ’Laws, Chances and Properties’, *International Studies in the Philosophy of Science* 4,

159-70; repr. and revised in *Matters of Metaphysics*, Cambridge: Cambridge University Press, 1991. (Probabilistic nonlogical, contingent regularity.)

Mormann, T. (1994) 'Accessibility, Kinds and Laws: A Structural Explication', *Philosophy of Science* 61: 389-406.(Structural logical, noncontingent necessitarian.)

Popper, K.R. (1990) *A World of Propensities*, Bristol: Thoemmes.(Propensity nonlogical, contingent necessitarian.)

Ruby, J.E. (1986) 'The Origins of Scientific "Law"', *Journal of the History of Ideas* 47: 341-59.(Useful historical discussion.)

Sklar, L. (1990) 'How Free are Initial Conditions?', in A. Fine, M. Forbes and L. Wessels (eds) *PSA 1990*, vol. 2, East Lansing, MI: Philosophy of Science Association.(Laws and initial conditions.)

Skyrms, B. (1980) *Causal Necessity*, New Haven, CT: Yale University Press.(Probabilistic laws and invariance.)

Suppe, F. (1989) *The Semantic Conception of Theories and Scientific Realism*, Chicago, IL: University of Illinois Press.(Model theoretic framework for semantics of law statements.)

Swoyer, C. (1982) 'The Nature of Natural Laws', *Australasian Journal of Philosophy* 60: 203-23.(Logical, noncontingent necessitarian.)

Tooley, M. (1977) 'The Nature of Laws', *Canadian Journal of Philosophy* 7: 67-98.(Property nonlogical, contingent necessitarian.)

Urbach, P. (1988) 'What is a Law of Nature? A Humean Answer', *British Journal for Philosophy of Science* 39: 193-210.(Subjective probabilistic nonlogical, contingent regularity.)

Vallentyne, P. (1988) 'Explaining Lawhood', *Philosophy of Science* 55: 598-613.(Concept nonlogical, contingent necessitarian with structure-of-histories law structure and counterfactual reasoning.)

Walters, R.S. (1967) 'Laws of Science and Lawlike Statements', in P.M. Edwards (ed.) *Encyclopaedia of Philosophy*, London: Collier Macmillan.(General historical discussion of philosophical issues.)

Wilson, M. (1990) 'Law Along the Frontier: Differential Equations and Their Boundary Conditions', in A. Fine, M. Forbes and L. Wessels (eds) *PSA 1990*, vol. 2, East Lansing, MI: Philosophy of Science Association.(Laws and boundary conditions.)

Le Clerc, Jean (1657-1737)

Le Clerc was not a particularly original philosopher - his position was somewhat eclectic - but his journals and textbooks make him an important historical figure. He acted as an intermediary between English and Continental traditions. In religion he advocated an attitude of toleration.

The Swiss thinker Jean Le Clerc (Johannes Clericus) was born in Geneva. After his theological studies there he travelled extensively in Europe, read Spinoza and started a correspondence with the Remonstrant theologian Philippus van Limborch. In 1683 he settled permanently in Amsterdam where he became a minister of the Remonstrant Brotherhood and was given a chair in philosophy at their seminary. In 1686 he started the journal *Bibliothèque Universelle* (25 vols), which in 1703 was followed by the *Bibliothèque Choisie* (26 vols) and in 1714 by the *Bibliothèque ancienne et moderne* (29 vols). These journals were instrumental in spawning debates which involved, among others, Leibniz and Bayle. They also were a channel for the spread of English ideas to the Continent, where people learned about Locke, the Cambridge Platonists, Berkeley and others by reading Le Clerc's abstracts (see Cambridge Platonism). Le Clerc edited the collected works of Erasmus and a work of Hugo Grotius (*De veritate religionis christianae*, 1709). His correspondence testifies to his friendship with numerous famous contemporaries.

Le Clerc was familiar with the philosophy of Descartes from his studies in Geneva with Robert Chouet. But although in physics and in metaphysics he remained basically Cartesian all his life, Cartesianism was never for him a closed system of thought. He not only modified and corrected Descartes' ideas in the light of Locke, Boyle and Newton, but was also influenced by the Platonism of Cudworth. Thus, we see him taking issue with Descartes on topics such as the existence of innate ideas (which he rejects) and animal machines (he attributed to animals some power of thinking). He was fascinated by Cudworth's theory of 'plastic natures' - which he publicized in his journals - because he believed that they provide a better explanation of the formation of animal organisms than the one-dimensional mechanism he found in Descartes. He remained a metaphysical dualist, however, and also believed that a consistent application of Descartes' principles (more consistent at any rate than we find in Descartes himself) is of very great use in establishing the truth of the Christian religion.

According to Le Clerc, religion does not require us to abandon our reason: we can never deceive reason, nor can reason ever deceive us. Accordingly, there is fundamental agreement between reason and revelation, not in the sense that revelation can ever be replaced by reason, but rather that the right use of reason leads us to accept revelation and that revelation in turn is never contrary to reason. In religion, Le Clerc's attitude was formed by Erasmus and Grotius. He believed in the value of philosophical and critical methods in theology and was an advocate of tolerance. His opposition to 'enthusiasm' (his reason for rejecting Malebranche) and 'fanaticism' had much influence on English Latitudinarians.

THEO VERBEEK

List of works

Le Clerc, J. (1685) *Sentiments de quelques théologiens de Hollande sur l'Histoire critique du Vieux Testament composé par le P. Richard Simon*, Amsterdam.

Le Clerc, J. (1686) *Défense des Sentiments*, Amsterdam.

Le Clerc, J. (1692) *Logica sive ars ratiocinandi*, Amsterdam; many reprints.

Le Clerc, J. (1695) *Physica sive de rebus corporeis*, Amsterdam; many reprints.

Le Clerc, J. (1696a) *De l'incrédulité ou l'on examine les motifs et les raisons générales qui portent les incrédules rejeter la religion chrétienne*, Amsterdam; trans. as *A Treatise on the causes of incredulity*, London, 1697, 1720.

Le Clerc, J. (1696b) *Ars critica in qua ad studia linguarum latinae, graecae et hebraicae via munitur*, 2 vols, Amsterdam; 6th edn, 1778.

Le Clerc, J. (1698) *Opera philosophica*, 4 vols, Amsterdam.

Le Clerc, J. (1959) *Lettres inédites de Le Clerc à Locke*, ed. G. Bonno, Berkeley and Los Angeles, CA.

Le Clerc, J. (1984) *Arminianesimo e tolleranza nel Seicento olandese: Il carteggio Ph. van Limborch/Jean Le Clerc, a cura di Luisa Simonutti*, Florence: Olschki.

Le Clerc, J. (1987-) *Epistolario*, ed. M. Sina, vol. 1 (1679-89), vol. 2 (1690-1705), Florence, 1991.(In progress.)

References and further reading

Barnes, A. (1939) *Jean Le Clerc et la République des Lettres*, Paris.

Bots, H. (1977) 'L'esprit de la République des Lettres et la tolèrance dans les trois premiers périodiques savants hollandais', *XVIIe siècle* 116 (1977): 43-5.

Colie, R.L. (1957) *Light and Enlightenment: A Study of the Cambridge Platonists and the Dutch Arminians*, Cambridge: Cambridge University Press.

Haag, E. and Haag, E. (1846-58) *La France protestante*, Paris, 10 vols.(Volume 6, pages 464-70 are relevant.)

Pitassi, M.C. (1987) *Entre croire et savoir: Le problème de la méthode critique chez Jean Le Clerc*, *Kerkhistorische Bijdragen*, vol. 14, Leiden: Brill.

Reesink, H.J. (1931) *L'Angleterre et la Littérature anglaise dans les trois plus anciens Périodiques de Hollande de 1684 à 1709*, Paris.

Verniére, P. (1954) *Spinoza et la pensée française avant la Révolution*, Paris: PUF, 2 vols.(Volume 1, pages 73 and following are relevant.)

Le Doeuff, Michèle (1948-)

Michèle le Doeuff has created new possibilities for philosophical writing. By working between philosophy and Shakespearean drama, social history and personal letters, she demonstrates how philosophy's concepts gather meaning by circulating between different forms of discourse. In so doing she takes up many of the main problems of philosophy, including the nature of the self, the possibility of philosophical ethics, and the place of women in society and in philosophy. She shows, too, how philosophy's proper commitment to critical reasoning represses the role of imagery in its own creative thought; philosophy seems dedicated to achieving theoretical results beyond its means, and imagery is misused to disguise its inevitable incompleteness. The spectre of women, or of some other group typified as 'irrational' is used to reassure philosophy of its own integral rationality.

Michèle le Doeuff has been employed as a philosopher at the École Normale Supérieure de Fontenay, the Centre National de la Recherche Scientifique in Paris and (1995) the University of Geneva. Her work generated discussion internationally after the publication in English in 1977 of 'Women and Philosophy' and, in 1979, of 'Operative Philosophy'. A study of the role of imagery in philosophy and science, *Recherches sur l'imaginaire philosophique* (1981) (*The Philosophical Imaginary*, 1989) was followed by translations and studies, first of Bacon's *New Atlantis* (1983b) and then of Shakespeare's *Venus and Adonis* (1986). Her most extended work, *L'Étude et le rouet* (1989) (*Hipparchia's Choice*, 1991), written as four 'Notebooks', interweaves semi-fictional and historical anecdote, comment on the nature of philosophy, critical analysis of existentialism, and reflection upon philosophical biography and the philosophical 'subject'.

Le Doeuff exposes the male bias in the images, anecdotes and theory of Sartrean existentialism, and expresses critical astonishment at de Beauvoir's *tour de force* in adapting this system for her analysis of the position of women in *The Second Sex* (see Beauvoir, S. de). She claims that de Beauvoir placed her feminism within a structure of ideas which is itself an epitome of the male fear of women. Le Doeuff emphasizes the imagery which Sartre uses in *Being and Nothingness* to connect the female body with ontological 'lack' and the threat posed by matter generally, to 'masculine' free consciousness. In *The Philosophical Imaginary* le Doeuff develops a strategy of analysing philosophy at the 'neuralgic points' signalled by its imagery. Kant, for example, writes of 'Truth' in terms of the image of a northern island separated by a stormy ocean from the southern islands of illusion where duty and pleasure can coexist. Le Doeuff characterizes Kant's rhetorical image as a classic form of parental threat and seduction. His efforts to set limits to metaphysical speculation only indicate how much lies beyond that limit to be cordonned off.

Le Doeuff's theme of the role of utopias in philosophical enterprise connects with her use of imagery in analysing the partial blindness any theory has to its own level of operation. More's *Utopia*, for instance, is not only a political ideal, but a trope of the ideal life of the individual scholar who desires a world in which nothing disturbs his inner peace. The figure of Utopus, the island whose life is reflected in its internal harbour of still water, represents the triumph of tranquil private enjoyment as an ideal life for its citizens. It is this dream of a political system in which narcissists could cooperate enough to meet their material needs which lies behind More's criticism of existing political institutions. More's political vision is argued to spring from a too limited idea of pleasure, and of what people can gain from and with each other. Le Doeuff pays similar attention to 'voice' in philosophy, showing how voice, too, operates vitally even in writing which affects to be free of affect. While being critical of the operation of voice, philosophers must accept and use it, rather than merely deploring it as an 'irrational' element in discourse.

Le Doeuff's work stresses this need for a complex attitude towards elements which help create philosophy and yet threaten its attempt at being reasonable. 'Masculinist' and 'logocentric' assumptions have been endemic in the production of philosophy and yet it is no use, she maintains, for women to attempt to redress the situation by expressing a special 'femininity' in philosophy. That femininity is a fantasy conjured up by a 'masculinity' bent on arrogating the field of reason to itself. At the same time, le Doeuff attacks the pretensions of 'masculinism' by all available means of argument and satire. Her critique of philosophy is not that it practises some excess of reason. It exhibits too *little* reason. In the name of a higher, or benignly legislative devotion to Reason, it has defined its own integrity by excluding women, or people of 'inferior' cultures, or children, as constitutionally on the other side of

Reason. It is not philosophy's ideal of being logical which does the harm. Rather it is 'logocentrism', by which le Doeuff means the hardy tendency of philosophers to 'form projects beyond philosophy's means' by aiming at one 'logos' - one account of things to legitimate all other forms of discourse. A philosophy prepared to interact with the verbal interjections of unregularized 12-year-olds, she observes, 'may be a form…that no longer considers its incompleteness a tragedy' (1989: 126).

Le Doeuff's study of the operation of imagery in philosophy leads on to a concern with various images of the natural and the philosophical 'subject'. Though scathing of Husserl's inflated agenda for re-establishing philosophy on the basis of a 'transcendental ego', she sees the trick of transcendence - the illusion of being master of everything which is consciously intimate to us - as natural to human thought. Le Doeuff does not present a *theory* of the subject in one of the standard styles of idealism, dualism or materialism. Rather, borrowing from the tradition of phenomenology whose conclusions she repudiates, she looks into the natural origins of the fantasies such excessive theories articulate. Analysis of philosophical theory is set beside studies of the speeches of Shakespearean characters in which, humorously or tragically, they engage in the fantasy of transcending their 'empirical specificities'. She shows how the narrative voice constructed by a Descartes, a Husserl or a Sartre, becomes transformed into a 'mind distinct from a body', a 'transcendental ego', or a 'totally free consciousness'.

See also: Existentialism; Feminism; Postmodernism; Subject, postmodern critique of the; Consciousness

MAX DEUTSCHER

List of works

Le Doeuff, M. (1977) 'Women and Philosophy', *Radical Philosophy* 17; also in T. Moi (ed.) *French Feminist Thought*, Oxford: Blackwell, 1987.(Influential discussion of the significance of women's exclusion from philosophy.)

Le Doeuff, M. (1979) 'Operative Philosophy: Simone de Beauvoir and Existentialism', *Ideology & Consciousness* 6.(Opens new discussion of de Beauvoir's transformation of existentialism in *The Second Sex*.)

Le Doeuff, M. (1980a) *Recherches sur l'imaginaire philosophique*, Paris: Payot; trans. C. Gordon, *The Philosophical Imaginary*, London: Athlone, 1989.(Argues for philosophy 'in the mode of the essay, which does not foreclose issues'; pursues the problematic role of imagery in upholding the image of philosophy as essentially free of imagery.)

Le Doeuff, M. (1980b) 'La Philosophie renseignée', in C. Delaceaupagne and R. Maggiosi (eds) *Philosopher*, Paris: Fayard.(On teaching philosophy in the lycées.)

Le Doeuff, M. (1981-2) 'Pierre Roussell's Chiasmas', *Ideology & Consciousness* 9.(Analyses 'chiasmas' in historical and contemporary socio-biology - the theorist protests scepticism about what they know best, while asserting authoritative knowledge about the moral and social scene which lies beyond their competence.)

Le Doeuff, M. (1982a) 'Spécial-femmes, un grumeau sur la langue', *Alidades*.(The perils of being published under the rubric of 'women's writing'.)

Le Doeuff, M. (1982b) 'Quelle modernité philosophique?', *La Revue d'en Face* 12.(On the displacement of women writers by 'progressive' deconstructionist and psychoanalytically oriented male reviewers.)

Le Doeuff, M. (1983a) 'Utopies scolaires', *Revue de Métaphysique et de Morale* 88 (2); trans. 'Utopias: Scholarly', *Social Research* 49 (2).(Analysis of the 'utopia' as a projection of a form of society ideal for a vision of intellectual life.)

Le Doeuff, M. (1983b) *'La Nouvelle Atlantide' suivi de 'Voyage dans la pensée baroque'*, Paris: Payot. (Translation, with Margaret Llasera, of Francis Bacon's *New Atlantis*, followed by a monograph analysing the significance of the imagery, rhetoric and narrative of Bacon's utopia.)

Le Doeuff, M. (1984a) 'L'Unique Sujet parlant', *Esprit* 89.(Studies Sartre's letters to 'Castor and some others', sketching the collapse of the ideal of freedom into that of being unique master of discourse.)

Le Doeuff, M. (1984b) 'The Public Employer', *m/f* 9.

Le Doeuff, M. (1986a) *'Venus et Adonis' suivi de 'Genèse d'une catastrophe'*, Paris: Alidades-Distique. (Translation of Shakespeare's *Venus and Adonis*, followed by analysis of a 'modern narcissism' as at the heart of the 'catastrophe' of Venus' failed love for Adonis.)

Le Doeuff, M. (1986b) 'Du sujet', in *Cross-references*, Society for French Studies, supp. pub. 8.(Analysis of the philosophical 'subject', relating Hegel, Lacan, Descartes and Freud to a project to be found in many of Shakespeare's characters, both tragic and comic. A moral for the contemporary scene.)

Le Doeuff, M. (1987) 'Ants and Women, or Philosophy without Borders', in A. Phillips Griffiths (ed.) *Contemporary French Philosophy*, Cambridge: Cambridge University Press.(Caustic analysis of how

philosophers, so careful about the lives of ‘ants’, feel free in their pronouncement on women. Philosophy needs borders open to all forms of knowledge, rather than a special autonomy.)

Le Doeuff, M. (1989) *L’Étude et le rouet*, Paris: Éditions du Seuil; trans. T. Selous, *Hipparchia’s Choice*, Oxford: Blackwell, 1991.(Comprising four ‘notebooks’, deals with the problems for women within philosophy, existentialism as a theory of consciousness and freedom, the possibility of a philosophical politics and ethics, and a role for utopianism.)

Le Doeuff, M. (1990) ‘Woman, Reason, etc.’, *Différences* 2 (3).(Discusses the risk of anti-intellectualism in the idea of special relation of women to reason.)

Le Doeuff, M. (1991) *Francis Bacon: du progrès et de la promotion des savoirs*, trans. and introductory essay, Paris: Gallimard.

Le Doeuff, M. (1992) ‘Gens des science: essai sur le déni de "mixité"’, *Nouvelles questions féministes* 13 (1).(Studies the way groups of women and men are treated as if essentially male.)

Le Doeuff, M. (1993) ‘Le chromosome du crime: a propos de XY’, *Futur antérieur: féminismes au présent*, Paris: L’Harmattan.(Discusses the phenomenon of the publication of books whose politics is not to be read, but to create a transient shock, thus to be quoted as exemplifying a new wave.)

Le Doeuff, M. (1995) ‘Problème d’investiture (de la parité etc.)’, *Nouvelles questions féministes* 16 (2).(Argues that a legal requirement of equal numbers of women and men in parliament, as sometimes suggested, could have extremely conservative effects. Suggests alternative strategies.)

References and further reading

Beauvoir, S. de (1949) *Le deuxième sexe*, Paris: Gallimard; trans. H.M. Parshely, *The Second Sex*, Harmondsworth: Penguin, 1972.

Deutscher, M. (ed.) (forthcoming) *Imagery, Self and Feminism in the Philosophy of Michèle le Doeuff*, New York: Humanities Press.

Deutscher, M. (1987) ‘Stories, Pictures, Arguments’, *Philosophy* 62 (240).

Gatens, M. (1986) ‘Feminism, Philosophy and Riddles without Answers’, in C. Pateman and E. Grosz (eds) *Feminist Challenges*, Boston, MA: Northeastern University Press.

Gatens, M. (1991) *Feminism and Philosophy*, Cambridge: Polity Press.

Griffiths, M. (1995) ‘(Auto)biography and Epistemology’, *Educational Review* 47 (1).

Grimshaw, J. (1996) ‘Philosophy, Feminism and Universalism’, *Radical Philosophy* 76.

Grosz, E. (1987) ‘Feminist Theory and the Challenge to Knowledges’, *Women’s Studies International Forum* 10 (5).

Grosz, E. (1989) *Sexual Subversions*, Sydney: Allen & Unwin.

Kraus, S. (1989) ‘Simone de Beauvoir: Between Sartre and Merleau-Ponty’, *Temps Modernes* 45 (520).

La Caze, M. (1994) ‘De Beauvoir and Female Bodies’, *Australian Feminist Studies* 20.

Lagree, J. (1984) ‘Utopian Fantasizing and Philosophical Conceptualization’, *Études Philosophiques* 4. Lloyd, G. (1983) ‘Masters, Slaves and Others’, Radical Philosophy 34.

Mackenzie, C. (1986) ‘Simone de Beauvoir: Philosophy and/or the Female Body’, in Pateman and Gross (eds) Feminist Challenges, Boston, MA: Northeastern University Press.

Marks, E. and de Courtivron, I. (eds) (1980) *New French Feminisms*, New York: Schocken Books.

Moi, T. (1986) ‘Existentialism and Feminism: The Rhetoric of Biology in *The Second Sex*’, *Oxford Literary Review* 8 (1-2).

Morris, M. (1981-2) ‘Operative Reasoning: Michèle le Doeuff, Philosophy and Feminism’, *Ideology & Consciousness* 9.

Morris, M. (1988) *The Pirate’s Fiancée*, London: Verso.

Nelson, C. (1987) ‘Feminism, Language and Philosophy’, *New Literary History* 19 (1).

Spivak, G. (1981) ‘French Feminism in an International Frame’, *Yale French Studies* 62.

Le Grand, Antoine (1629-99)

Le Grand was the foremost expositor and popularizer of Cartesian philosophy in England during the seventeenth century. He wrote on ethics, politics, logic and numerous scientific topics. His Cartesian system is one of the most complete, emphasizing mind-body dualism and interaction, innate ideas, mechanism and method. His theory of signification is important in the Cartesian debate on ideas.

Born in France at Douai, Antoine Le Grand was sent as a Franciscan missionary to England where he spent most of his life, spreading Cartesian philosophy. He was well received by the Neoplatonists at Cambridge, but not so at Oxford where Cartesian philosophy was banned.

Le Grand's early writings are not principally Cartesian. In *Le sage des Stoiques ou l'homme sans passions* (*Man without Passion*) (1662) he expounds the doctrines of the late Stoic, Seneca: the principles of the passions are useless to virtue, and thus a wise man ought to live without passion. However, Le Grand later rejected this position because, like Descartes, he came to view the passions as conducive to life in their preservation of the mind-body union. In his main political work, *Scydromedia* (1669), he depicts the ideal state as one whose dominant features include autocracy limited by law, open trade, private property, education of the youth and religious unity.

Le Grand's comprehensive *Institutio philosophiae* (*An Entire Body of Philosophy*) (1672) includes applications of Cartesian principles of the sort that Descartes had only envisaged, with discourses on insects, plants, natural theology and ethics, plus expositions of many traditional Cartesian topics such as the *Cogito*, innate ideas and proofs of God's existence. Consistently anti-Aristotelian, Le Grand substitutes Cartesian principles of method for those found in ancient and scholastic logic. He rejects the law of non-contradiction as a first principle because it presupposes that something exists, and thus is dubitable. In the Cartesian order of philosophizing, first truths must be indubitable propositions; since the *Cogito* presupposes nothing, it is indubitable and hence a first principle. Similarly, Le Grand replaces Aristotelian form and matter with Cartesian substance and mode.

Le Grand defends Descartes' account of ideas as representational modes of the mind. The nature of ideas was a topic greatly debated in the latter half of the seventeenth century. A crux of the debate was to explain how ideas which do not resemble their objects can represent them. Le Grand's contribution was to explicate the subject (idea)-object relation in terms of signification: ideas are signs that acquire signification by a causal process ultimately dependent on God's will.

Like a minority of his Cartesian contemporaries, such as Desgabets and Régis, Le Grand retains the interaction of mind and body while rejecting any occasionalist solution to the problem of how two essentially different substances causally interact. Additionally, Le Grand defends Descartes' doctrine of the free creation of eternal truths - that God created all things in their essence and their existence. Malebranche argued that this doctrine undermines the a priori basis for knowledge, although Le Grand never drew this consequence.

Against the traditional Aristotelian account of soul, Le Grand defends the Cartesian 'beast-machine' doctrine that non-human animals lack souls and hence are subjects for purely mechanical explanations. The animal soul is essentially corporeal, unlike the human soul which is essentially thinking and incorporeal. Thus the mechanizing of the sensitive soul is achieved without abandoning a spiritual account of the rational soul.

Le Grand was not an original thinker but he was instrumental in bringing Cartesian philosophy to a wide audience in England. He gave Cartesianism a scholastic form to aid its acceptance, and he was thoroughly loyal to Descartes' ideas, making him, uniquely, a complete Cartesian apologist.

See also: Descartes, R.; Sergeant, J.

PATRICIA A. EASTON

List of works

Le Grand, A. (1662) *Le sage des Stoiques ou l'homme sans passions, selon les sentiments de Sénèque*, The Hague; repr. *Les caractères de l'homme sans passions, selon les sentiments de Sénèque*, Paris, 1663; Lyons, 1665; trans. G. Richard as *Man without Passion: Or, the wise Stoick, according to the Sentiments of Seneca*,

London, 1667.(Advances the Stoic doctrine that the passions are not useful to man, a view Le Grand later rejected.)

Le Grand, A. (1669) *Scydromedia seu sermo quem Alphonsus de la Vida habuit corram comite de Falmouth de monarchia liber primus (Conversations on Scydromedia, Alphonsus de la Vida to the Earl of Falmouth)*, London; Nuremberg, 1680; German trans. U. Greiff, Bern: Lang, 1991.(Semi-fictional work which depicts the ideal state, which Le Grand called 'Scydromedia'. This work is available only in Latin or German, both which are found under the Latin title *Scydromedia*.)

Le Grand, A. (1671) *Philosophia veterum, e mente Renati Descartes more scholastico breviter digesta*, London; repr. and augmented as *Institutio philosophiae secundum principia d. Renati Des-Cartes, nova methodo adornata & explicata, cumque indice locupletissimo actua (Institution of philosophy according to the principles of René Descartes: The new method praised and explained, with expanded index)*, London, 1672; Nuremberg, 1679; Geneva, 1694.(Le Grand later edited, expanded and appended two smaller works to this work, the result of which was translated into English by R. Blome and published as *An Entire Body of Philosophy*.)

Le Grand, A. (1672) *Institutio philosophiae secundum principia Domini Renati Des-Cartes, nova methodo adornata & explicata in usum juventutis academicae*, trans. corrected and expanded R. Blome as *An Entire Body of Philosophy, According to the Principles of the Famous Renate des Cartes, in Three Books, I The Institution; II The History of Nature; III Dissertation on Brutes*, London: Roycroft, 1694; repr. with intro. R.A. Watson, New York: Johnson Reprint Corporation, 1972.(The first comprehensive presentation of the Cartesian system, translated from the Latin and published in English to reach a wide audience, probably intended as a textbook.)

Le Grand, A. (1673) *Historia naturae variis experimentis & ratiociniis elucidata (The history of nature elucidated by various experiments and reasons)*, London, 1680; Nuremberg, 1678.(This work, dedicated to Robert Boyle, treats of the nature of body in general and of various kinds of bodies in particular in order to demonstrate the truth of Cartesian principles.)

Le Grand, A. (1675) *Dissertatio de carentia sensus et cognitionis in brutis (Dissertation on the absence of sense and knowledge in brutes)*, London; Lyons, 1675; Nuremberg, 1679; English trans. appended to *An Entire Body of Philosophy*, London, 1694.(A careful account of animal souls as essentially corporeal, consisting in the movement of blood through the body, and human souls as essentially incorporeal. Le Grand aligns the Cartesian account with that of Gassendi and Fabri against that of the Aristotelians. The work is available in Latin and English; the English version, *Dissertation of the Want of Sense and Knowledge in Brutes, is appended to An Entire Body of Philosophy*, 1964.)

Le Grand, A. (1679) *Apologia pro Renato Des-Cartes contra Samuelem Parkerum, S.T.P. archidiaconum cantuariensem, instituta & adornata (A defence of Descartes against Samuel Parker)*, London, 1681; Nuremberg, 1682.(Samuel Parker, an Oxford theologian who condemned Descartes' philosophy on the grounds that it leads to atheism, had effected a ban on its public entrance to Oxford university. This work is only available in the original Latin editions.)

Le Grand, A. (1698) *Dissertatio de ratione cognoscendi et appendix de mutatione formali, contra J.S. methodum sciendi (Dissertation on recognizing reason, with an appendix of rule exchange, against the method of knowing of J. Sergeant)*, London.(An impassioned defence of Descartes and Malebranche on ideas and other metaphysical matters against his English opponent, John Sergeant. This work is available only in its Latin, 1698 edition.)

References and further reading

Bouillier, F. (1854) *Histoire de la philosophie cartésienne (A history of Cartesian philosophy)*, Paris, 2 vols. (General introduction to Le Grand's life and work especially with regard to his role in the popularization of Cartesianism.)

Rosenfield, L.C. (1968) *From beast-machine to man-machine; animal soul in French letters from Descartes to La Mettrie*, enlarged edn, New York: Oxford University Press.(Very readable with an excellent discussion of the 'beast-machine' doctrine found in Descartes and his successors.)

Ryan, J.K. (1935) 'Anthony Legrand, 1629-99: Franciscan and Cartesian', *The New Scholasticism* 9: 226-50. (Examines Le Grand's life and work as a Franciscan and a Cartesian, especially the dangerous political circumstances of his mission in England; thorough bibliography.)

Ryan, J.K. (1936) 'Scydromedia: Anthony Legrand's Ideal Commonwealth', *The New Scholasticism* 10: 39-55. (Describes in detail the ideal commonwealth that Le Grand creates in his 1669 moral-political work *Scydromedia*.)

Watson, R.A. (1966) *The Downfall of Cartesianism: 1673-1712, a study of epistemological issues in late 17th century Cartesianism*, The Hague: Martinus Nijhoff.(Clear and excellent discussion of the Cartesian controversies regarding the nature of ideas, including Le Grand's theory of signification.)

Le Roy, Édouard Louis Emmanuel Julien (1870-1954)

Le Roy was a French mathematical physicist and Catholic modernist philosopher. Starting from a philosophy of life similar to Henri Bergson's philosophy of creative evolution, he argued that the capacity for invention was fundamental to human existence. This led him to develop a radical form of conventionalism according to which scientific facts are created rather than discovered. Henri Poincaré attacked this view, arguing that the scientist creates only the language in which facts are expressed.

Le Roy was born in Paris. His father was in the transatlantic shipping business, with his own outfitting company in Le Havre. Le Roy entered the École Normale Supérieure in 1892, passing the *agrégation* in 1895 and earning a doctorate of science in 1898. For many years, he taught mathematics at a Paris *lycée*. An assistant to Henri Bergson at the Collège de France from 1914 to 1920, he was named Bergson's successor in the chair of modern philosophy in 1921, and remained in that position until his retirement in 1941. In 1919 Le Roy was elected a member of the *Académie des Sciences Morales et Politiques*; in 1945 he succeeded Bergson as a member of the Académie Française.

Le Roy was a controversial representative of the Catholic modernist movement in France. He held a pragmatic view of religious dogma, whereby dogma has practical value as a source of moral directives in spite of its seeming inscrutability and incompatibility with positive, rational knowledge. These ideas, developed in such works as *Dogme et critique* (1906), earned Le Roy condemnation by Pope Pius X in an encyclical of 1907, and his *Le problème de Dieu* (1929) was put on the Index in 1931. At the centre of Le Roy's thinking is a vitalist philosophy of life and a spiritualist philosophy of freedom closely allied to Bergson's doctrine of creative evolution and similar in crucial respects to the ideas of Le Roy's friend, Teilhard de Chardin (see Bergson, H.-L.; Teilhard de Chardin, P.; Vitalism). The human capacity for invention is fundamental and points us back to the critical phenomenon of 'hominization', when the noosphere emerges from the biosphere. A free, self-creative human agent becomes the basis of Le Roy's mature moral philosophy and philosophy of religion, in which central roles are played by transcendence, a quest for self-fulfilment and respect for a generalized moral obligation and spiritual need arising out of the human vital spirit.

As with Bergson, intuition is for Le Roy the way to true knowledge, which consists in a non-discursive immediacy characteristic of the primitive, authentic relationship of instinctive life to its surroundings. All discursive knowledge, however, is invention. Such invention may be necessary in the form of practical rules as a guide for human action, but, as invention, discursive knowledge can make no more claim to truth than the arbitrary rules of a game. This idea, with its emphasis on the mediating role of the senses and conceptual forms, owes an acknowledged debt to the Neo-Kantianism of Émile Boutroux, but Le Roy made it the basis for a radical form of conventionalism that he first presented in a widely discussed article, 'Science et philosophie' (1899-1900) (see Conventionalism). Le Roy's point of view gained considerable notoriety when it became the subject of an extended critical discussion by Henri Poincaré in 'Sur la valeur objective des théories de physiques' (1902) (see Poincaré, J.H. §4). The challenge of Le Roy's more radical conventionalism led Poincaré to some valuable and essential clarifications of his own, more moderate, conventionalism. Two major points were at issue. First, Le Roy had argued that scientific facts are merely the artificial creations of the scientist, citing as examples the concept of the atom, the phenomenon of the eclipse and the rotation of the earth. Poincaré responded by doubting that one could distinguish crude facts and invented scientific facts as sharply as Le Roy might think possible and by insisting that crude facts themselves play a role in science by imposing scientific facts upon us. Poincaré grants the presence of a creative moment in science, but he confines it to the conventional choice of disguised definitions, as in the choice of a unit metre stick or the choice of Euclid's parallel postulate, choices analogous to the choice of a language. Once those conventions are accepted, the remaining propositions of science have a determinate truth-value ascertainable by empirical means. Moreover, the role of such arbitrary conventions diminishes steadily as we move from the basic levels of science, such as geometry and kinematics, to mechanics and physics. The second point of difference between Poincaré and Le Roy concerned the question whether, among the conventional variants of a given scientific theory, there remained any 'universal invariant', a content common to the different formulations. Poincaré interpreted Le Roy as denying the existence of such an invariant content, but himself insisted upon it. Arguing again by analogy with translation between two languages, such as French and German,

Poincaré held that one and the same proposition could be expressed equivalently in either language, with determinate rules of translation allowing us to go back and forth between the two.

In his *La théorie physique* (1906), Pierre Duhem declared his sympathy for Le Roy's conventionalism in opposition to Poincaré's critique (see Duhem, P.M.M. §4). Two years later, in *Identité et réalité* (1908), Émile Meyerson expressed sympathy for Le Roy's views, but identified himself more closely with what he regarded as the less radical position of Duhem. Le Roy's radical conventionalism had perhaps its most profound influence on Kazimierz Ajdukiewicz. By contrast, his work in the philosophy of mathematics, and his writings on microphysics and relativity theory, were of less lasting import.

DON HOWARD

List of works

Le Roy, É. (1899-1900) 'Science et philosophie' ('Science and Philosophy'), *Revue de métaphysique et de morale* 7 (1899): 375-425, 503-62, 706-31; 8 (1900): 37-72.(First statement of Le Roy's radical conventionalism.)

Le Roy, É. (1900) 'La science positive et les philosophies de la liberté' ('Positive Science and the Philosophies of Liberty'), in *Congrès internationale de philosophie, I Paris 1900, Philosophie générale*, Paris: Colin, vol. 1, 313-41.(A second widely read exposition of Le Roy's conventionalism.)

Le Roy, É. (1901) 'Un positivisme nouveau' ('A New Positivism'), *Revue de métaphysique et de morale* 9: 143-4. (Influential proclamation of a new anti-metaphysical movement.)

Le Roy, É. (1906) *Dogme et critique (Dogma and Criticism)*, Paris: Bloud.(Pragmatic view of dogma, condemned by Pius X.)

Le Roy, É. (1912) *Une philosophie nouvelle, Henri Bergson*, Paris: F. Alcan; trans. V. Benson, *The New Philosophy of Henri Bergson*, London: Williams & Norgate, and New York: Holt, 1913.(Study of Bergson.)

Le Roy, É. (1927) *L'Exigence idéaliste et la fait de l'évolution (The Exigency of Idealism and the Fact of Evolution)*, Paris: Boivin & Cie.(Le Roy's philosophy of life.)

Le Roy, É. (1928) *Les origines humaines et l'évolution de l'intelligence (Human Origins and the Evolution of Intelligence)*, Paris: Boivin & Cie.(On hominization and the noosphere.)

Le Roy, É. (1929) *Le problème de Dieu (The Problem of God)*, Paris: L'Artisan du livre.(Put on the Index in 1931.)

Le Roy, É. (1929-30) *La pensée intuitive (Intuitive Thought)*, Paris: Boivin & Cie, 2 vols.(On intuition as the source of genuine knowledge.)

References and further reading

Abelé, J. (1955) 'Le Roy et la philosophie des sciences' ('Le Roy and the Philosophy of Science') *Études* 284: 106-12.(On the debate between Le Roy and Poincaré.)

Duhem, P. (1906) *La théorie physique. Son objet et sa structure*, Paris: Chevalier et Rivière; 2nd edn, 1914; trans. P.P. Wiener, *The Aim and Structure of Physical Theory*, Princeton, NJ: Princeton University Press, 1954. (Claims priority over Le Roy; otherwise expresses general agreement.)

Giedymin, J. (1978) 'Radical Conventionalism, its Background and Evolution: Poincaré, Le Roy and Ajdukiewicz', in J. Giedymin (ed.) *Kazimierz Ajdukiewicz: The Scientific World-Perspective and Other Essays, 1931-1963*, Dordrecht: Reidel, xix-lii; repr. in J. Giedymin, *Science and Convention: Essays on Henri Poincaré's Philosophy of Science and the Conventionalist Tradition*, Oxford: Pergamon, 1982, 109-48.(An informative study of influence of Le Roy's conventionalism.)

Meyerson, É. (1908) *Identité et réalité*, Paris: Alcan; 2nd edn, 1912; trans. K. Loewenberg, *Identity and Reality*, London: Allen & Unwin, 1930; repr. New York: Dover, 1962.(Expresses sympathy for Le Roy's conventionalism while claiming a position closer to Duhem.)

Miranda, M. do C.T. de (1957) *Théorie de la verité chez Édouard Le Roy* (The Theory of Truth in Édouard Le Roy), Paris: Lecoffre.(The only extended study of Le Roy's theory of truth.)

Poincaré, H. (1902) 'Sur la valeur objective des théories de physiques', *Revue de métaphysique et de morale* 10; repr. in *La valeur de la science*, Paris: Flammarion, 1905; trans. G.B. Halstead, *The Foundations of Science*, New York: Science Press, 1913; repr. separately as *The Value of Science*, New York: Dover, 1958.(Critical attack on Le Roy's claim that scientific facts are created.)

Learning

Learning is the acquisition of some true belief or skill through experience. Rationalist/idealist philosophers held that the very constitution of thought guarantees that fundamental laws hold of the world we experience, and that our understanding of these laws was therefore innate, not learned. The empiricist tradition, doubtful of these Rationalist claims, denied that much was innate, and held that learning occurred through associations of mental representations. This view was lent support by the nineteenth-century development of physiological psychology, which led to a view of learning as a system of adjustments in a network without any intervening representations, a perspective that led in turn, in the twentieth century, to behaviourist studies of stimulus-response associations, and eventually to contemporary neural net computational models.

Empiricism, however, had also invited, especially with Hume, doubts that the correspondence between mental representations and the world could be known. Hume believed people learn, or at least form new habits, but he did not think there could be any normative theory of learning - any way of making it 'rational'. His scepticism led to the development by Bayes and other statisticians of formal theories of how learning from evidence ought to be done. However, the standards that developed in the form of the theory of subjective probability proved impossible to apply until very fast digital computers became available.

The digital computer in turn prompted both novel normative theories of learning not considered by the statistical tradition, and also attempts to describe human learning by computational procedures. At the same time, a revolution in linguistics held that humans have an innate, specialized algorithm for learning language. Applications of computation theory to learning led to an understanding of what computational systems - possibly including people - can and cannot reliably learn. Major issues remain concerning how people acquire the system of distinctions they use to describe the world, and how - and how well - they learn the causal structure of the everyday world.

1 Eighteenth- and nineteenth-century empiricism

The empiricist tradition, represented in the eighteenth and nineteenth centuries by Hume and Mill, among philosophers, and by Maudsley among psychologists, developed an associationist psychology in which mental dynamics is driven by 'associations' among 'ideas', roughly bits of images or propositional content. Associations were thought to be constituted both by internal relations among ideas - similarity, for example - and external physical relations - near simultaneity of acquisition, for example. The doctrine of ideas distinguished simples and compounds. Experience presents compound ideas (or 'impressions' from which ideas are formed) which are separated into simples by a kind of internal analysis (see Empiricism; Hume, D. §2; Mill, J.S. §§2, 3).

Empiricist psychology produced foundational problems: how such can such processes lead to faithful representations of the external world, and how can we know that they do? Hume argued that they could not be known to do so. The dominant alternative answers in the eighteenth and nineteenth (and, one might argue, twentieth) centuries were forms of idealism, from the phenomenalism of Mill and Mach, which held that all reality is appearance, to the transcendental idealism of the Kantians, who agreed with Hume that we cannot know the world 'in itself', but held that the very conditions for the possibility of thought and experience guarantee the fundamental geometric, temporal and causal structure of experience (see Idealism; Kant, I. §§4-6; Mach, E.; Phenomenalism).

Helmholtz read Kant as offering a (false) psychological theory rather than a logical solution to sceptical doubts about knowledge and the apparent phenomena of a prioricity. Helmholtz offered empirical analyses of the visual and auditory systems, emphasizing both their physiological mechanisms and the external conditions required for their reliability. His work represents perhaps the first really 'naturalized' epistemology, rejecting all a priori justifications of the reliability of human belief.

The network structure of the cortex, established by Cajal in the 1890s, gave new impetus to associationist theories of learning. In psychologists as varied as William James and Sigmund Freud, 'ideas' were vaguely and variously associated with states of particular cells or collections of cells, or with distributions of energy among cells. Learning was thought to be some process by which these systems were adjusted in response to inputs from sensory

transducers. The network idea offered at least the skeleton of a unified account of the acquisition of both behavioural and inferential skills. Under James' influence, Edward Thorndike called himself a 'connectionist' and network talk became (and remains) both a metaphor for associations, a detailed computational model, and a theory of brain function. In the first two roles, the strengths of the connections between network nodes have commonly been assumed to be modified either by their temporal pairing or by a feedback process sensitive to a discrepancy between the output of the system and a 'teaching' signal. This discrepancy (the error in the output) is fed back into the system that maps inputs to outputs, altering the connection strengths so as to reduce the discrepancy. In the behaviourist tradition (see Behaviourism, methodological and scientific), the first process for modifying the strengths of connections is called classical or Pavlovian conditioning, while the second is called instrumental or operant conditioning. In more recent 'neural net' computer simulations (see Connectionism), the first process is called 'unsupervised learning', while the second is called 'supervised learning', although no real neural process is known that corresponds to propagating errors back through a network.

2 Cognitive nature of the associative process

Thorndike is often regarded as one of the fathers of (scientific) behaviourism, the psychological movement that holds that psychology should focus exclusively on causal relations between 'stimuli' and behavioural 'responses'. One stimulus became associated with another, or with a response, as a result of 'reward' or 'reinforcement'. Behaviourism permitted associationism to be put to a fairly rigorous scientific test. However, experimental results from the study of putatively associative learning in a small number of basic experimental paradigms indicate that the process is considerably more complex than was originally thought. For example, the strengths of associations do not develop independently of one another. Whether or not the temporal pairing of two stimuli - say, the appearance of a diagonal red stripe on a round plastic button followed a few seconds later by the presentation of food to a pigeon - does or does not produce an association (in this case, between the stripe and the food) depends on the strengths of other associations that have developed prior to the current training or are developing during this training. If, for example, the average interval between food presentations is as high in the absence of the diagonal stripe as it is in its presence, the experimental chamber itself becomes strongly associated with food, and this association between the chamber and food blocks the formation of an association between the diagonal stripe and the food. The dependence of changes in one associative bond on the strengths of other associative bonds reflects the ability of the associative process to pick out the best and simplest predictors from among several competing stimuli that predict the occurrence of something important to the animal (hereafter called a 'reinforcer' because it strengthens the animal's reaction to the predicting stimulus).

A second complexity is that animals 'associate' two stimuli regardless of the temporal relation between them. 'Associate' is placed in quotes because it here means something much more vague like 'learns something about the temporal relation between'. If a stimulus reliably predicts the omission of a reinforcer, the animal learns an inhibitory association, that is, it learns that one stimulus predicts the omission of the other. If the two stimuli are made to occur truly randomly, so that knowledge of the time of occurrence of one provides no information about the time of occurrence of the other, then the animal learns that one does not predict the other. This is shown by a subsequent test in which the stimuli are temporally paired, so that now one stimulus does reliably predict the other. The exposure to the early truly random phase retards the animal's 'associating' the two stimuli in this second phase.

Note the difficulty of conceptualizing the learning in the truly random phase as the establishment of a conductive link (a true association) between the two stimuli. The one stimulus neither excites nor inhibits an expectation of the other - the only possibilities in a purely associative theory of learning - and yet the animal has learned something. Thus, it is difficult to conceptualize this learning in associative terms, because it is difficult to say what has become associated with what.

The discovery of the subtleties of the associative process in simple animal learning paradigms has led many modern theorists to adopt a more cognitive interpretation of the associative process. The process seems to depend on abstract relations between stimuli such as 'contingency' rather than on simple temporal pairing. This more cognitive view has less neurobiological transparency, and seems to require something like the resources of a computational theory of mind (see Mind, computational theories of).

3 Probability and learning

In the eighteenth and nineteenth centuries, statistics developed normative theories of inference unconnected with human psychology, but bounded by human computational limitations. A typical statistical problem concerned resolving recorded observations into a function of quantities satisfying a deterministic regularity and a quantity, representing the sum of a multitude of unknown quantities influencing the observed values, satisfying a probability distribution. After Legendre's development of least squares early in the nineteenth century, that method soon dominated practice in astronomy and geodesy. Laplace and Bayes advocated methods of learning involving forming new probabilities by conditioning old (or 'prior') probabilities on new evidence. George Boole attempted to develop more general probabilistic methods for inferring causes, but his proposals had little success.

Late in the nineteenth century Pierce introduced the idea of randomization in experiments, and Fisher's elaboration of that idea early in the twentieth century led directly to almost all of modern experimental design in the non-physical sciences, establishing a norm about how causes are to be learned in subjects from biology to economics (see Statistics; Statistics and social science). Roughly in the same period, the work of Galton, Pearson and Yule on regression, correlation and contingency tables led to the application of regression and association for causal inferences from non-experimental observations. Yule himself carefully investigated the limitations of such methods, which are still widely used but regarded as unreliable by experts. Beginning at the turn of the century, a number of psychologists developed the methods of factor analysis, which amounted to algorithms for generating theories of how unobserved features influenced observed quantities.

There are two approaches to probability theory, 'objective' (for example, 'frequentist') ones, which emphasize how the probability that something is *G*, given that it is *F*, is a function of the number of *F*s that have been *G* in the past, and 'subjectivist' ones, which emphasize the degree of a person's confidence in a hypothesis (see Probability, interpretations of). Introduced by Ramsey and De Finetti early in the twentieth century, the modern formulation of subjective probability treats probability as a measure of degree of belief, utility as a measure of desire, maximum expected utility as a decision principle, and conditional probability as a (normative) mechanism of learning. This 'Bayesian' perspective offered at least the framework of a general theory of inference and decision, but its practical influence was very limited both because of the dominance of frequentist ideas in statistics, and because the calculations required by Bayesian norms in even simple problems were beyond unaided human capacities. Kahneman, Tversky, Dawes and many others have shown regularly reproducible forms of human judgment that seem to violate Bayesian norms, but others have argued that such experiments fail to capture the range and frequency of decision-making tasks in which human competence evolved, and for that range of tasks human psychology satisfies Bayesian norms (see Rationality of belief). Others have offered speculations about how the nervous system may implement (unconsciously, of course) Bayesian computations (see Probability theory and epistemology; Rationality, practical).

Using either the calculus of probability or logical 'partial entailment' relations, or simply *modus ponens*, Keynes, Carnap, Reichenbach, Hempel, Popper and others proposed theories of 'confirmation' or 'corroboration' of hypotheses (see Confirmation theory; Inductive inference). These proposals concerned themselves largely with the a priori justification of hypotheses; no attempt was made to show that the confirmation relations they defined coincided with procedures for reliably discovering the truth, that is, for learning. These were regarded as issues exclusively for psychology. Concern with capturing reliable methods has come to dominate recent theories of learning, and 'reliabilist' approaches in epistemology generally (see Reliabilism).

4 Learning and computation

Using logical methods derived from Frege, Carnap and Russell sketched the 'construction' of concepts of properties, objects, and spatial and temporal relations from categories held to be more immediate: 'sense-data' in Russell's case, recollections of similarity between two 'elementary experiences' - Gestalt episodes - in Carnap's case (see Sense-data; Logical positivism). These theories of concept formation were not offered as mathematical psychology but as analyses (explications) of ordinary concepts (see Concepts §§3, 5-6; Conceptual analysis). Carnap's theory had the distinction of being algorithmic, and might arguably be considered the first attempt at an artificial intelligence programme (see Artificial intelligence).

The advent of the theory of computation and the digital computer prompted a number of related developments in

the middle of the twentieth century. Psychologists began exploring algorithmic procedures for learning. Under the guise of ‘concept learning’, Bruner and his colleagues, and later Hunt, developed effective versions of Francis Bacon’s methods. From instances of the predicates involved, these programs learned universal biconditionals with a single predicate on one side and any logical formula not containing that predicate on the other side. The psychologists gave some rather sophisticated thought to how the representation of propositions influences the computational complexity of their application in inference. The subject of artificial intelligence was born at the same time, and it immediately focused on how to make computers simulate learning. Although none of these algorithms was an accurate description of human learning, they were used to promote the proposition that various aspects of human learning may be algorithmic or computational.

Chomsky revolutionized linguistics by supposing that language speakers recognize grammatical sentences, and distinguish them from non-grammatical sentences by using an unconscious algorithm. He proposed that humans acquire the algorithm for their native language by applying a highly specific learning procedure specific to language and shared by all people. Chomsky claims that the procedure (‘Universal Grammar’) is not, itself, learned along ‘empiricist’ lines, but is innate along lines he associates with the early Rationalists. The idea naturally raised many questions of a kind that still trouble cognitive science and linguistics: how can we know whether there is a special learning facility for language? If there is such an algorithm, any human language must be of a kind the algorithm can learn from the evidence available to children (see Language, innateness of). How does that restrict the languages that are possible for humans?

The last question has been addressed in part by computational learning theory, a subject that has one of its roots in the philosophy of Hans Reichenbach. Reichenbach had argued that the aim of inductive inference is to judge limiting frequencies, and that the simple rule of guessing that limiting frequencies equal observed frequencies converges to the correct limit if a limit exists. His student, Hilary Putnam, combined Reichenbach’s long-run, reliabilist perspective with recursion theory to create one branch of the modern subject of computational learning theory. The framework for a discovery problem assumes that a set of alternative hypotheses are to be assessed, and any member of a set of (infinitely long) data streams is possible; each hypothesis is consistent with some data streams and inconsistent with other data streams. The investigator is idealized as a function from initial segments of data streams to hypotheses. The investigator succeeds on a data stream provided that the function representing the investigator has a false (for that data stream) hypothesis as its value on at most a finite number of initial segments of the data stream. The investigator solves the discovery problem provided it succeeds on every possible data stream allowed in the formulation of the problem.

The framework captures many reliabilist intuitions, and it has the further virtue of flexibility: success criteria can be altered; the relation between data streams and hypotheses can be altered; and constraints of many kinds can be imposed on the function representing the investigator. It does not represent inductive success as getting the correct answer in the infinite future, but rather as getting the correct answer after a finite amount of evidence is acquired, but without any fixed guarantee on how large ‘finite’ will be.

Putnam, and independently at about the same time, E. Mark Gold, characterized when a problem can be solved by an effective learner - that is, one whose learning function is a Turing machine (see Turing machines). Gold subsequently adapted the framework to model language learning, where a language is represented as a set of numbers that can be listed by a computer. For some collections of possible languages, there exist algorithms that can learn any language in the collection; for other collections, no such language is possible. On the Chomskian view, the collection of possible human languages must be of the first kind, a learnable class. So, by studying algorithms, something can be learned about the structure of human language. Computer scientists investigated both the language framework and a related one in which the data streams are graphs of recursive functions.

Kelly developed the framework into topological, measure-free alternatives to statistical analyses of reliable inference, with a whole class of counter examples to standard philosophical morals about methods of inquiry. For example, there are solvable (by computable learners) problems that cannot be solved by computable learners that output only hypotheses consistent with data received and background assumptions; other solvable (again, by computable learners) problems that cannot be solved by computable learners who change hypotheses only when they are contradicted by the evidence; and still others that cannot be solved by computable Bayesian learners. For learners that do not have to be computable, traditional confirmation strategies fail to solve discovery problems that

can be solved by other means. Kelly showed there is a universal strategy that succeeds whenever success is possible but is neither Reichenbach's nor Popper's. Virtually every methodological principle advanced by philosophers of science turns out to be sub-optimal or to entail some sacrifice of reliability.

Social aspects of learning have been undertaken in the context of computational learning theory, usually as studies of 'team learning' in which a number of parallel investigators process data and success requires that only one of them succeed. Recently, using very simplified reliability frameworks, Kitcher has provided an economic analysis of social gains in reliability when investigators compete and cooperate with the selfish aim of being the first to learn the truth about a scientific question.

We are still far from understanding how humans develop so that they distinguish relevant features of the world and know the causal relations among everyday circumstances. A theory about algorithms for learning causes has recently developed in philosophy and computer science, but psychologists have not yet compared it with human performance, and statistically sophisticated psychological models of human judgment of causality are only now appearing.

See also: Innate knowledge

C.R. GALLISTEL
CLARK GLYMOUR

References and further reading

Amsel, A. (1989) *Behaviorism, Neobehaviorism, and Cognitivism in Learning Theory: Historical and Contemporary Perspectives*, Hillsdale, NJ: Erlbaum.(A neo-behaviourist critique.)

Blumberg, M.S. and Wasserman, E.A. (1995) 'Animal Mind and the Argument from Design', *American Psychologist* 50: 133-44.(A critique of theories that posit animal consciousness from a moderately behaviourist perspective.)

Chomsky, N. (1959) 'Review of Verbal Behavior (by B.F. Skinner)', *Language* 35: 26-58.(An influential critique of behaviourist learning theory.)

Gallistel, C.R. (1990) *The Organization of Learning*, Cambridge, MA: Bradford Books/MIT Press. (Rationalist/nativist treatments of learning.)

Gould, J.L. and Marler, P. (1987) 'Learning by Instinct', *Scientific American* 256: 74-85.(Examination of ethological evidence for innate cognitive faculties.)

Hull, C.L. (1930) 'Knowledge and Purpose as Habit Mechanisms', *Psychological Review* 37: 511-25.(Classic statement of the neo-behaviourist position.)

Klein, S.B. and Mowrer, R.R. (eds) (1989) *Contemporary Learning Theories*, 2 vols, *Instrumental Conditioning Theory and the Impact of Biological Constraints on Learning* and *Pavlovian Conditioning and the Status of Traditional Learning Theory*, Hillsdale, NJ: Erlbaum.(Modern associative learning theory.)

Osherson, D.N., Stob, M. and Weinstein, S. (1986) *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists*, Cambridge, MA: MIT Press.(A contemporary rationalist approach to learning.)

Pavlov, I.V. (1928) *Lectures on Conditioned Reflexes: The Higher Nervous Activity of Animals*, London: Lawrence & Wishart.(The central work on 'classical' behaviourist conditioning.)

Pinker, S. (1994) *The Language Instinct*, New York: William Morrow.(An excellent and highly accessible summary of the Chomskian revolution in linguistics.)

Rumelhart, D.E. and McClelland, J.L. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, *Foundations*, vol. 2, *Psychological and Biological Models*, Cambridge, MA: MIT Press.(Neural net approaches to associative learning by computers.)

Skinner, B.F. (1950) 'Are Theories of Learning Necessary?', *Psychological Review* 57: 193-216.(Classic statement of the radical behaviourist position.)

Skinner, B.F. (1990) 'Can Psychology be a Science of Mind?', *American Psychologist* 45: 1,206-10.(A radical behaviourist critique of cognitive approaches.)

Watson, J.B. (1919) *Psychology from the Standpoint of a Behaviorist*, Philadelphia, PA: Lippincott.(The early, classic statement of the behaviourist programme.)

Lebensphilosophie

In its most general sense Lebensphilosophie *denotes a philosophy which asks after the meaning, value and purpose of life, turning away from purely theoretical knowledge towards the undistorted fullness of lived experience. In the second half of the nineteenth century and the early twentieth century the concept of 'life' assumed a central role in German philosophy.* Lebensphilosophie *typically opposes rigid abstractions with a philosophy based on feeling and intuition, and seeks to establish the priority of 'life' as an all-encompassing whole. The central claim underlying its various manifestations is that life can only be understood from within.*

The first recorded use of the term *Lebensphilosophie* is in a 1772 collection of essays by G.B. von Schirach, *Über die moralische Schönheit und Philosophie des Lebens (On moral beauty and the philosophy of life)*. In the late eighteenth and early nineteenth centuries the expression was employed to characterize a philosophy for practical life and was used synonymously with terms such as *Lebensweisheit* (wisdom of life) and *Lebenskunst* (the art of living) in order to provide general ethical and practical maxims for correct living.

While there are obvious points of continuity with this popular usage, the roots of the modern conception of *Lebensphilosophie* should, more generally, be traced back to the mid-eighteenth-century reaction against Enlightenment rationalism. The emphasis on feeling and immediacy, and the lived experience of truth in the work of figures such as Hamann and Herder, together with the search for a unifying principle prior to the abstractions of reason, leads to a prioritization of 'life' over 'mere understanding'. The *Sturm und Drang* mobilized a set of fundamental oppositions - between the living and the dead, the concrete and the abstract, the organic and the mechanical, the dynamic and the static - whose influence is carried through to modern *Lebensphilosophie* in the Idealism of Schelling and the early Hegel, as well as the Romantic attempt to establish the identity of philosophy and life, poetry and thought (see Romanticism, German).

However, the centrality of the concept of 'life' to German philosophy in the decades between 1870 and 1920 cannot be understood without reference to the work of Schopenhauer and Nietzsche. Schopenhauer's concept of the will as a single, unifying principle is both a precursor and model for subsequent attempts to grasp life as an all-encompassing metaphysical category: here the critique of rationalism first takes on the form of a metaphysics of the irrational. In contrast, Nietzsche's importance resides in his attempt to explicate truth in terms of its function for life, and his insistence on the dynamic, historical and conflictual character of human claims to knowledge. A third and later figure of comparable importance is the French philosopher Henri Bergson, whose conception of time as lived experience exercised an enormous influence on German philosophy in the late nineteenth and early twentieth centuries. Bergson's ideas were taken up as part of the vitalist opposition to scientific materialism and the attempt to show that life cannot adequately be explained in mechanical and physical terms: biology rather than physics provides the central categories for understanding life in its own terms (see Vitalism).

Lebensphilosophie must also be seen as part of a broader cultural movement which extended beyond the confines of philosophy. This movement is marked, on the one hand, by a striving to overcome, through a vitalistic affirmation of the fullness of life, the limits placed on the isolated individual, and on the other hand by a mood of pessimism and decadence which received its most powerful expression in the arts. In the work of figures as diverse as Georg Simmel, Ludwig Klages and Oswald Spengler, the concept of life is taken up as a critical category by means of which the conflicts and contradictions of modern society could be identified and diagnosed.

There are two principal reasons for the relative obscurity into which *Lebensphilosophie* has fallen today, despite the dominance which it exercised prior to and immediately after the First World War. The first is the extent to which some of its principal insights were taken up in a methodologically more rigorous and productive way in Husserlian phenomenology and Heidegger's 'philosophy of existence' (see Heidegger, M.). The second is the extent to which the irrationalism and cultural criticism of *Lebensphilosophie* fed into the doctrines of National Socialism. Perhaps the most enduring achievement of *Lebensphilosophie* is to be found in the work of Wilhelm Dilthey. His emphasis on the concrete fullness of lived experience and the diversity of the manifestations of social and cultural life prior to the abstractions of reason results in a new attempt to ground philosophically the human sciences and the important methodological distinction between *Erklärung* (explanation) and *Verstehen* (understanding) (see Explanation in history and social science).

See also: Enlightenment, Continental

JASON GAIGER

References and further reading

Bollnow, O.F. (1958) *Die Lebensphilosophie*, Berlin/Göttingen/Heidelberg: Springer Verlag.(Comprehensive survey.)

Kühne-Bertram, G. (1987) *Aus dem Leben - zum Leben. Entstehung, Wesen und Bedeutung populärer Lebensphilosophen in der Geistesgeschichte des 19. Jahrhunderts (From life - to life. The origin, nature and significance of popular philosophers of life in the history of ideas in the 19th century)*, Frankfurt: Peter Lang. (An attempt to trace the influence of popular, pre-scientific writings on later, academic *Lebensphilosophie*.)

Lieber, H.-J. (1974) *Kulturkritik und Lebensphilosophie. Studien zur Deutschen Philosophie der Jahrhundertwende (Cultural criticism and Lebensphilosophie. Studies in German philosophy at the turn of the century)*, Darmstadt: Wissenschaftliche Buchgesellschaft.(Contains essays on Dilthey and Simmel and on the relationship of *Lebensphilosophie* to National Socialism.)

Misch, G. (1930) *Lebensphilosophie und Phänomenologie. Eine Auseinandersetzung der Dilthey'schen Richtung mit Heidegger und Husserl (Lebensphilosophie and phenomenology. A critical discussion of the relationship between Dilthey's approach and that of Husserl and Heidegger)*, Bonn: Friedrich Cohen.(Early analysis of the relationship between *Lebensphilosophie* and phenomenology which seeks to defend the superiority of Dilthey's approach.)

Rickert, H. (1920) *Die Philosophie des Lebens. Darstellung und Kritik der philosophischen Modeströmungen unserer Zeit (The philosophy of life. A presentation and critique of the current trends in philosophy)*, Tübingen: J.C.B. Mohr.(Contemporary critique of *Lebensphilosophie* from the standpoint of Neo-Kantianism.)

Scheler, M. (1915) 'Versuch einer Philosophie des Lebens' ('An essay on the philosophy of life'), in *Vom Umsturz der Werte. Abhandlungen und Aufsätze (On the overturning of values. Treatises and essays)*, Bern: Francke, 1955.(Identification of Nietzsche, Dilthey and Bergson as the founding fathers of contemporary *Lebensphilosophie*.)

Schirach, G.B. von (1772) *Über die moralische Schönheit und Philosophie des Lebens. Reden und Versuche (On moral beauty and the philosophy of life. Lectures and essays)*, Altenberg.(First publication to use the term *Lebensphilosophie*.)

Schnädelbach, H. (1984) *Philosophy in Germany, 1831-1933*, trans. E. Matthews, Cambridge: Cambridge University Press.(Focused history of German philosophy written for an English-speaking audience. Contains a chapter devoted to the concept of 'life'.)

Sendlinger, A. (1994) *Lebenspathos und Décadence um 1900. Studien zur Dialektik der Décadence und der Lebensphilosophie am Beispiel Eduard von Keyserlings und Georg Simmels (The pathos of life and decadence around 1900. Studies in the dialectic of decadence and Lebensphilosophie through the example of Eduard von Keyserling and Georg Simmel)*, Frankfurt: Peter Lang.(Investigation into the broader cultural context of *Lebensphilosophie*.)

Simmel, G. (1921) *Der Konflikt der modernen Kultur*, trans. K.P. Etzkorn in G. Simmel (ed.) *The Conflict in Modern Culture and Other Essays*, New York: Teachers College Press.(Account of the centrality of the concept of life by one of the key thinkers of the period.)

Lefebvre, Henri (1901-91)

Henri Lefebvre was a Marxist and existential philosopher, a sociologist and a theorist of the state. His humanistic neo-Marxism has been influential throughout Europe. In the English-speaking world he is best known for his analyses of 'everyday life', his work in the sociology of urban and rural life, and his theory of social space. Lefebvre was one of the most prominent early critics of structuralism, and is considered by some to be the first post-structuralist. He was a relentless critic of academic philosophy's metaphysical tendencies.

During the 1920s, Lefebvre was associated with the surrealist movement in Paris, and his work from this period has been described as proto-existentialist. In 1929, however, he joined the Parti Communiste Français (PCF) and by the 1940s had become the PCF's leading intellectual spokesperson and a Marxist critic of Sartrean existentialism (see Sartre, J.-P.; Marxism, Western). With Norbert Guterman, Lefebvre was the first to translate the works of the young Karl Marx into French, and his *Le Matérialisme dialectique* (*Dialectical Materialism*) (1939) earned him the title 'Father of the Dialectic' in France. He was also the first interwar writer to attempt an anti-fascist reading of Nietzsche, and one of the few to publish rigorous critiques of nationalism and National Socialism.

Lefebvre's main philosophical work, *Logique Formelle, Logique Dialectique* (Formal logical, dialectical logic) (1947a), provides a critique of formal logic and categorical reasoning. In 1957 he was excluded from the PCF for his attacks on Stalinism and his humanist unorthodoxy, and was subsequently reconciled with Sartre. In the 1960s he inspired figures in the French avant garde, including his teaching assistant Jean Baudrillard and members of the Situationniste Internationale, many of whom were his students, including Michele Bernstein, Guy Debord and Daniel Cohn-Bendt, the student leader of the May 1968 demonstrations in Paris. The failure of those demonstrations to lead to substantial changes in the French way of life was seen, however, as an indictment of Lefebvre's method of revolution-through-festival, derived from the surrealists; in subsequent years his humanist Marxism came under direct attack from Althusser's 'Scientific Marxism' (see Althusser, L.P.).

Lefebvre's influential interpretation of Marxism was similar in tone to that of the young Georg Lukács, emphasizing alienation as the central concept of Marx's thought. Alienation in the economic sphere was explored using the tools of historical dialectical materialism, a methodology on which Lefebvre placed great store; but he also argued that Marxism would remain incomplete so long as it was restricted to the economic sphere rather than incorporating all aspects of life. He presented three-part dialectics in order to undermine the tendency to simplistic dualisms apparent in much Marxist thought. Lefebvre's 'retro-jective' method, by which he attempted to operationalize historical materialism for research purposes, provided the basis for Sartre's regressive-progressive method.

In his *Critique de la vie quotidienne* (*Critique of Everyday Life*) (1947b) Lefebvre critically analyses the banality of life in the modern world in terms of the 'mystification' of its alienated, boring quality. His theory is a reinterpretion of Lukács' and Heidegger's notion of 'everydayness' or banality (*Alltäglichkeit*) in the more sociological terms of alienation and everyday social routines. In this, he is a precursor of the work of Bourdieu on habitual routines ('habitus'). A 'romantic revolutionary', Lefebvre sought a 'revolution in everyday life' in which spontaneous, ludic expression would be liberated from the restrictions of authoritarian rationality and alienated routine. He was one of the first authors to theorize the malaise of everyday modern life in terms of the 'globalization' of production.

Lefebvre was a founder of rural sociology and of applied sociology in France, where he established the first Institute of Sociology at Strasbourg. In the 1960s and 1970s, as a professor at Nanterre, he concentrated on urban problems, presenting his theory in the 1974 work *La production de l'espace* (*The Production of Space*). He argued that notions of spatiality, the interaction of bodily action and the constructed environment, and theories of geography are all aspects of a social space created by people, not something which occurs naturally. This 'space' is a dialectical product of social activity which reacts back on social action to channel it in a continuous dialectic of ongoing 'spatialization'. In the English-speaking world, where Lefebvre's Marxist and existentialist theories are less well-known, his most enduring influence has been in the fields of geography, urban planning and cultural theory.

Lefebvre served as a conduit for ideas from generation to generation, including surrealists (1920s), Marxists (1930s), existentialists (late 1940s) and situationists (early 1960s), to American neo-Marxists and analysts of late-twentieth-century postmodernity, such as Fredric Jameson and David Harvey. His key ideas of autonomous self-organization and spontaneity influenced the German Green Party and East German artists' movements, and British punk-anarchist subcultures in the 1980s.

See also: Existentialism; Marx, K.; Post-structuralism

ROB SHIELDS

List of works

Lefebvre, H. (1938) *Hitler au pouvoir, bilan de cinq années de fascisme en Allemagne (Hitler into power, a report of five years of fascism in Germany)*, Paris: Bureau d'éditions.(Critical essay on fascism and the working class.)

Lefebvre, H. (1939a) *Le Matérialisme dialectique*, Paris: Alcan; trans. J. Sturrock as *Dialectical Materialism*, London: Cape, 1968.(The most widely available introduction to the subject, and to Marxism in general.)

Lefebvre, H. (1939b) *Nietzsche*, Paris: Editions sociales internationales.(Anti-fascist study of Nietzsche.)

Lefebvre, H. (1947a) *Logique formelle, logique dialectique (Formal logical, dialectical logic)*, Paris: Editions sociales.(Volume 1 of *A la lumière du matérialisme dialectique*, 2nd volume censored and destroyed during the Second World War. Mimeo of author's proofs available at Carleton University Library, Ottawa, Canada).

Lefebvre, H. (1947b) *Critique de la vie quotidienne, I, Introduction*, Paris: Grasset; 2nd edn trans. with preface by J. Moore, *The Critique of Everyday Life, Volume 1*, London: Verso, 1991.(First of four volumes published between 1947 and 1981.)

Lefebvre, H. (1959) *La Somme et le reste (Total and dividend)*, Paris: La Nef de Paris, 2 vols; repr. as one vol., Paris: Bélibaste, 1973.(Dialectical autobiography.)

Lefebvre, H. (1963) *La vallée de Campan - Étude de sociologie rurale (The Campan valley, a study in rural sociology)*, Paris: Presses Universitaires de France, Collection 'Bibliothèque de sociologie contemporaine'; 2nd edn, 1991.(Pioneering work on peasant life.)

Lefebvre, H. (1968) *La vie quotidienne dans le monde moderne* Paris: Gallimard, Collection Idées; trans. S. Rabinovitch as *Everyday Life in the Modern World*, Harmondsworth: Penguin, 1971.(Introduction to the concepts of everyday life and alienation.)

Lefebvre, H. (1971) *Au-delà du structuralisme (Beyond structuralism)*, Paris: Anthropos. (Critique of structuralism.)

Lefebvre, H. (1974) *La production de l'espace*, Paris: Anthropos; trans. D. Nicholson-Smith as *The Production of Space*, Oxford: Blackwell, 1991.(Dialectical theory of social space.)

Lefebvre, H. (1975a) *Hegel, Marx, Nietzsche, ou le royaume des ombres (Hegel, Marx, Nietzsche, or the kingdom of shadows)*, Paris: Tournai, Casterman. Collection 'Synthèse contemporaines'.(Synthesis of dialectics and Nietzschean philosophy.)

Lefebvre, H. (1975b) *Le temps des méprises: Entretiens avec Claude Glayman (The era of errors - interview with Claude Glayman)*, Paris: Stock.(Autobiographical interviews detailing the life of Paris intellectuals in the 1920s and 1930s.)

Lefebvre, H. (1975c) *L'Idéologie structuraliste (The Structuralist ideology)*, Paris: Le Seuil, Collection Points 6. (Critique of structuralism.)

Lefebvre, H. (1996) *Henri Lefebvre Writings on Cities*, ed. and trans. E. Kofman and E. Lebas, Oxford: Blackwell.(Introduction and selected readings.)

References and further reading

Ajzenberg, A. (ed.) (1994) *Traces de futurs. Henri Lefebvre, le possible et le quotidien (Traces of futures. Henri Lefebvre, the possible and the banal)*, Paris: La Société Française.(Retrospective colloquium on Lefebvre's influence.)

Burkhard, F. (1986) *Priests and Jesters: Henri Lefebvre, the 'Philosophies' Gang and French Marxism between the wars*, Ph.D. thesis, Department of History, University of Wisconsin - Madison.(Detailed historical research on Paris intellectuals during the mid-1920s.)

Gottdiener, M. (1985) *Social Production of Urban Space*, Austin, TX: University of Texas Press.(Application of Lefebvre's theory of space to urban planning.)

Hess, R. (1988) *Henri Lefebvre: L'Aventure d'un siècle (Henri Lefebvre: The adventure of a century)*, Paris.

(Standard French biography.)
Jameson, F. (1991) *Postmodernism: Or, the cultural logic of late capitalism*, London: Verso.(Analysis of the late twentieth century, partly inspired by Lefebvre.)
Jay, M. (1984) *Marxism and Totality - the adventures of a concept from Lukács to Habermas*, Berkeley, CA: University of California Press.(Historical study including a summary of Lefebvre's status and context of his thought.)
Martins, M. (1983) 'The theory of social space in the work of Henri Lefebvre', in R. Forrest, J. Henderson and P. Williams (eds) *Urban Political Economy and Social Theory: critical essays in urban studies*, London: Gower, 160-85.(Systematic discussion of Lefebvre's relevance to political economy.)
Meyer, K. (1973) *Henri Lefebvre: ein romantischer Revolutionnär (Henri Lefebvre: a romantic revolutionary)*, Vienna: Europa Verlag.(Standard German biography.)
Poster, M. (1975) *Existential Marxism in Postwar France: From Sartre to Althusser*, Princeton, NJ: Princeton University Press.(Historical context of Lefebvre's proto-existential contributions.)
Ross, K. (1988) *The Emergence of Social Space: Rimbaud and the Paris Commune*, London: Macmillan.(Assesses and expands Lefebvre's studies of the (1848) Paris Commune (PAC) and theories of 'festive revolution'.)
Sartre, J.-P. (1956) *Being and Nothingness*, trans. H.E. Barnes, New York: Philosophical Library.(Discussion and modification of Lefebvre's dialectical methodology.)
Shields, R. (1997) *Henri Lefebvre: a critical introduction*, London: Routledge. (Intellectual biography with standard bibliography.)
Soja, E. (1985) 'The spatiality of social life: towards a transformative retheorisation', in D. Gregory and J. Urry (eds) *Social Relations and Spatial Structures*, London: Macmillan.(Exposition of Lefebvre's theory of space.)

Legal concepts

Concept-formation is an important component of law-formation. Well-developed legal orders are profoundly conceptual in nature. Throughout Western legal history, legislators have aimed at basing their law-making on concepts of a general scope (such as 'property', 'possession', 'usufruct', 'criminal intent' and many others) - and even more so legal scholars in their reconstruction and development of law. Legal thinking makes use of concepts with many different functions and varying logical status. A distinction can be made between concepts that are an integral part of law themselves (here called L-concepts) and concepts that belong to the professional vocabulary of lawyers and jurists in their handling of the law (J-concepts). Among the L-concepts there are on the one hand concepts whose meaning is totally determined by the rules of one single legal system and on the other hand concepts that pertain to two or more legal systems. The latter concepts have a comparative function.

J-concepts provide lawyers with a language enabling them to give an intellectual structure to the legal material, to characterize and discuss the professional-juridical handling of law and the methods used for performing that task, to specify the functions of law and to formulate the underlying values of (the handling of) the legal system.

There was a tendency in earlier legal philosophy to hypostasize legal concepts, for example, the concept of 'right' in classical natural-law doctrine: that is, to postulate real entities to which our concepts/terms refer. The legal philosophy of the twentieth century has to a large extent been a reaction against this tendency. This reaction has taken three different directions: (1) to reduce the abstract legal concepts to factual phenomena such as certain human behaviour or socio-psychological factors (mainly within US and Scandinavian realism); (2) to assign to legal concepts a normative ontological status, placing them in a world of norms, distinct from the world of facts; and (3) to analyse legal concepts in a contextual setting, that is, to find out how they function in actual legal discourse.

1 Legal concept-formation: L-concepts and J-concepts

Legal orders have, historically, emerged from a need to restrain unhampered human violence. In more developed legal orders, rational argumentation has to a large extent superseded violence as a means of solving social problems. Western legal orders, based on or at least influenced by Roman law and the dogmatic study of it since the end of the eleventh century, have a strong intellectual stamp: disputes shall be solved according to general principles and vital social relationships defined with accuracy by precise legal concepts. 'Normative order…is…a hard won production of organizing intelligence' (MacCormick (1990: 557). Just as theory-formation and concept-formation in science are 'so closely interrelated as to constitute virtually two different aspects of the same procedure' (Hempel 1952: 1), there is a similarly close connection that verges on coalescence between law-formation and concept-formation. Western science and Western law equally rest on a common base - rationalism - which has its roots in the world of ideas of ancient Greece and Rome. Well-developed legal orders are profoundly conceptual in character.

Legal thinking, in all branches and on all levels, makes use of concepts with very different functions and of varying logical status. Let us, roughly, compare the conceptual world of, for example, a zoologist and that of a lawyer. The zoologist, like all other professionals, needs in his work a professional (technical) language. This language refers in different respects to the subject matter of zoology - foxes, halibuts, wings, cells - which are mainly entities of a physical nature (things). In contrast to the zoologist's vocabulary, the lawyer's vocabulary refers to the subject matter of the law, which is a language (a product of human culture) too, with an abundance of terms signifying different concepts. Thus, in the legal sphere of life we have to allow for two, in principle distinct, languages: the legal one (in a proper sense) - the object language - and the juridical one - the meta-language. We refer to the concepts of the first category (concepts of law) as legal concepts proper (L-concepts) and to the concepts of the latter category (concepts about law) as juridical concepts (J-concepts). L-concepts have a law-stating function, while J-concepts have a juridical-operative function.

Let us start with the L-concepts. Within this category we can distinguish between L-concepts pertaining *uniquely* to one single legal system, for example, French law (UL-concepts, where 'U' indicates this kind of 'uniqueness'), and L-concepts pertaining to two or more legal systems, comparative L-concepts (CL-concepts). (A legal system is here regarded as a set of legal norms, or rules.)

A UL-concept is an L-concept that is wholly dependent, with regard to its meaning, on the content of a given legal system at a given point of time (for example, the law in force in France on 1 January 1994). If a given concept, for example, ‘ownership’, is a UL-concept, then the meaning of ‘ownership’ is completely determined by the prerequisites (descriptions of legal, or operative, facts) and the legal consequences which, in a given legal system at a given point of time, decide how ownership arises and expires, as well as what the legal effects (the legal significance) of ownership are. From this it follows, for example, that the Swedish concept of ownership differs somewhat from the Norwegian one, and also that it is changed as soon as the legal rules (‘the ownership rules’) are changed with respect to the above-named descriptions.

A CL-concept is an L-concept whose meaning is dependent on the content of two or more legal systems. Suppose that a jurist is interested in the concept ‘marriage’, not with respect to some single legal system but to all, or some family of, legal systems. This amounts to finding the ‘lowest common denominators’ of all rules concerning marriage - the gist, as it were, of the phenomenon of marriage. This is, in fact, not a concept of any single legal system and hence, it might be argued, not a concept of law but a concept about law. But it is, no doubt, a concept of law in the respect that it is totally dependent on the content of legal systems - it is not a purely sociological concept. For the formation of CL-concepts the logical technique of ideal-type definitions is appropriate. CL-concepts might be of both scientific (within the field of comparative legal science) and legislative value (for example, in international legislative work of the kind that aims to harmonize different national legal systems with each other or to work out international rules in some particular sphere). As we have seen, one and the same term, for example, ‘marriage’, might signify both a UL-concept and a CL-concept.

2 UL-concepts: further distinctions

Within the category of UL-concepts some further distinctions can be made.

Antecedent and consequent UL-concepts

Take the legal rules

(r_1) If *a* owns *o*, then *a* is allowed to use (sell, pawn,…) *o*.

and

(r_2) If *a* has bought or inherited or received as a gift or… (and not sold, or…) *o*, then *a* owns *o*.

In r_1 ‘ownership’ refers to legal (or operative) facts; the phrase ‘if *a* owns *o*’ is shorthand for ‘if *a* has bought or inherited or received as a gift or… (and not sold or given away or destroyed or…)’; ‘ownership’ is a prerequisite expressed in the antecedent of r_1.

In r_2 ‘*a* owns *o*’ is shorthand for ‘*a* is allowed to use and sale and pawn and lease and…*o*’; ‘ownership’ here refers to the legal effects (or legal significance) expressed in the consequent of r_2.

In the Scandinavian discussion on rights in the 1940s and 1950s (initiated by Per Olof Ekelöf; see Wedberg 1951) the function of right-concepts in legal discourse (legislative, dogmatic and so on) was regarded as a shorthand-function. If we combine the antecedent of r_2 and the consequent of r_1 we get

(r_3) If *a* has bought or inherited or received as a gift or… (and not sold, or…) *o*, then *a* is allowed to use (sale, pawn,…) *o*.

where, apparently, the concept ‘ownership’ is not to be found - an observation which induced Alf Ross to the idea that right-concepts are semantically meaningless vehicles of deduction (Ross 1957).

Genuine and non-genuine UL-concepts

A UL-concept has a genuine or non-genuine function depending on whether it appears in genuine or non-genuine legal statements in the sense indicated by Ingemar Hedenius (1941). Let us explain this - now internationally accepted - distinction with the help of an example:

(r_4) Anyone who travels by car in Sweden must drive on the right-hand side of the road.

This statement can be understood in two different ways. If r_4 appears in an official statute it is natural to understand r_4 as a normative statement, a decree, which the legislator has issued with the intention of directing a certain kind of human behaviour. Understood in this way r_4 is lacking in truth-value; it is neither true nor false. Statements with this function are genuine legal statements. But suppose that a Swede visits England and is asked by an English person what the rules of the road are in Sweden, and that in reply they utter the (somewhat magisterial) statement r_4. When they make this reply it is not their intention - at least not in the first place - to steer the questioner's behaviour; their intention is merely to inform the questioner of the content of valid Swedish law in this respect. And when r_4 has this function it is a non-genuine legal statement, and has as such truth-value; it is either true or false. Statements of legal dogmatics are usually of this kind. It is to be noted that r_4 has the same linguistic formulation irrespective of whether it is used in the genuine way or the non-genuine way, and that genuine and non-genuine legal statements are often formulated alike. But the non-genuine legal statement always contains, logically speaking, an explicit or understood clause 'according to valid law in S at t', where S is some society and t is a point of time. Thus, in its non-genuine function r_4 is equivalent - so long as Swedish law is intended - to the statement

(r_5) Anyone who travels by car in Sweden must drive on the right-hand side of the road according to (now) valid Swedish law.

Official and dogmatic UL-concepts

Official UL-concepts are found in statutes, preparatory materials, judgments and other official texts. Dogmatic UL-concepts are found in legal-dogmatic works (for example, 'causality', 'proximate cause', 'right *in rem*'). Dogmatic concepts can, but need not, have equivalents among the official concepts. Legal dogmatists can quite well create their own concepts when they consider that the official battery of concepts is insufficient, and this happens now and then. Such concepts can later be incorporated in the law. Historically speaking there are also a number of concepts which, as products of academic law, have found their way into municipal law or are implicitly contained within it. An official concept and a dogmatic one need not have the same meaning, even if they are expressed by the same term (for example, 'possession' in Swedish statutory language is not equivalent to 'possession' as it is used in Swedish doctrinal literature).

3 J-concepts

Turning to the J-concepts, the technical tools used in the legal-professional handling of legal orders (that is, of L-concepts), we may list some important functions of juridical language and distinguish between different kinds of J-concepts on the basis of those functions (the list is by no means exhaustive). We need:

1 a set of concepts helping us to structure the law in a logical and functional respect (morphological concepts);
2 a set of concepts helping us to describe in a precise manner the professional-juridical handling of the law (praxeological concepts);
3 a set of concepts helping us to speak clearly and articulately about the relations between the law and the social reality in which the law 'operates' and which it is designed to influence - in short, about the functions of law (teleological concepts);
4 a set of concepts helping us to describe in a precise manner the programmes (methods, principles, techniques) for the juridical handling of law, be it in legislation, administration of law, legal dogmatics or otherwise (methodological concepts);
5 a set of concepts helping us to articulate our ideas about the values pertaining to the juridical handling of law (ideological concepts).

As for morphological concepts, we can differentiate between two levels: micro-morphological and macro-morphological. At the former level it is the component parts of individual (types of) legal rules that are studied. Here we meet concepts such as 'prerequisite', 'legal fact', 'legal consequence', the various concepts for (Hohfeldian) legal positions (see Hohfeld, W.N.), the various normative modalities 'shall', 'may', 'must not'

(expressing obligation, permission, prohibition) and so on. At the macro-morphological level we study the component parts of legal systems, that is, different types of legal rules and different kinds of relations between such rules. As an example of such a classification can be mentioned Hart's distinction 1961) between primary rules and secondary rules (rules of recognition, rules of change and rules of adjudication) (see Hart, H.L.A.). As examples of relations between rules can be mentioned the concepts (*r* and *r′* are legal rules) '*r′* is analogous to *r*', '*r′* is in an e-contrario relation to *r*', '*r′* is an extensive interpretation of *r*' and so on.

As examples of praxeological concepts can be mentioned 'complying with the law', 'breaking the law', 'administration of law', 'application of law', 'interpretation', 'evaluation of evidence', 'pleading', 'statutory drafting' and so on - all concepts concerning different kinds of legal activities.

Teleological concepts (concepts about the functions of law) are, for example, those introduced by Raz (1979) - namely, 'primary', 'secondary' and 'indirect functions' - and also concepts describing the relationship between a legal rule, *r*, and some social aim or purpose, *p*, for example, '*r* is expedient with respect to *p*' or '*r* is effective with respect to *p*'.

The flora of methodological terms is rich and varies from country to country. The underlying concepts, however, are very much the same. Some examples are: 'objective', 'subjective', 'historical', 'intentional', 'linguistic', 'logical-grammatical', 'systematic', 'teleological' and 'pragmatical interpretation', 'reasoning by example', 'reasoning by analogy'. To the methodological arsenal of concepts also belong the concepts dealing with the factual basis of legal argumentation - the legal sources - such as 'statute', 'act', 'precedent', 'case law', 'custom', 'preparatory materials', and the concept 'legal source' itself.

Among the ideological concepts we find concepts such as 'legal certainty' (or 'predictability'), 'legality', 'equality before the law', 'the rule of law' and '*Rechtsstaat*' ('law-state').

4 Substantialism and its opponents

In the classical natural-law doctrine of the seventeenth and eighteenth centuries there is a tendency towards substantialism, that is, to Platonic hypostasizing of legal concepts (Hobbes is an exception, but he was after all not a full-fledged natural law philosopher anyway). During the last two centuries the development of sciences and logic has contributed to a sharpened awareness among legal philosophers as to ontological matters. The reaction against substantialism has taken three, in principle different, directions.

One is reductionism. This is an idea of venerable age. It is found as early as the time of the Presocratics, in their efforts to bring the apparent variety of entities back to one single substance. Its programme is clearly stated by Ockham (*non est ponenda pluralitas sine necessitate*; a multiplicity shall not be assumed unless it is necessary). In the legal philosophy of the first half of the twentieth century reductionistic tendencies are strong, especially within the US and Scandinavian realist movements. Here the reduction consists in reducing abstract ('fictional') legal entities to a certain kind of human behaviour or to socio-psychological pressures and reactions. (We assume that abstract concepts such as 'legal order' have the same ontological status as the abstract entities to which these concepts refer, for example, legal orders (see Legal realism).)

Another one is separatism. This consists in locating legal concepts in a separate ontological realm of their own. Kelsen's *Pure Theory of Law* is an endeavour to separate normative entities (which at least must include the L-concepts) from factual entities by differentiating the world of Is (*Sein*) from the world of Ought (*Sollen*) (see Kelsen, H.).

A third reaction against substantialism is functionalism. Instead of asking about some legal concept, *c*, 'What does *c* refer to in reality?', legal scholars with a functional approach formulate the analytical question like this: 'How does *c* function in legal thinking?' Hohfeld belongs to the functionalist jurists, his classic work on jural relations is indeed entitled *Fundamental Legal Conceptions as Applied in Judicial Reasoning*. Functionalism was also the explicit starting point for the fruitful discussion on rights that took place in Scandinavia during the 1940s and 1950s (Wedberg 1951). Hart's idea (based on Bentham; Hart 1954) that legal concepts shall be defined in a contextual setting (by contextual definitions) is in the same vein.

Apart from ontological differences concerning legal concepts, we can discern different attitudes as to their

importance in legal life, oscillating between extreme absolutism and extreme relativism. Absolutism is here taken to mean a tendency to regard legal concepts as perpetual (static) and also precise with regard to their meaning, and relativism as the opposite of this. The *Begriffsjurisprudenz* (conceptual jurisprudence), flourishing in Germany during the second half of the nineteenth century, is characterized by a high degree of absolutism: when confronted with some legal problem, for example, whether some clause in a contract is about an easement or a usufruct, the task of the lawyer was to ascertain whether the right in question fell under the concept ‘easement’ or that of ‘usufruct’, and when this was done the problem was solved, since the legal effects of, for example, easement were fixed beforehand and implicit in the concept itself. This kind of legal technique is, in fact, to a certain degree indispensable in any well-developed legal order, but what gave *Begriffsjurisprudenz* its bad reputation was that it grossly disregarded the social consequences of the legal decisions and overestimated the precision of the legal concepts.

See also: Concepts; Conceptual analysis; Law, philosophy of; Legal discourse

ÅKE FRÄNDBERG

References and further reading

Frändberg, Å (1987) ‘An Essay on the Systematics of Legal Concepts: A Study of Legal Concept Formation’, *Scandinavian Studies in Law* 31: 83-115.(Expansion of the material of §§1-3 of this entry.)

Hart, H.L.A. (1954) ‘Definition and Theory in Jurisprudence’, *Law Quarterly Review* 70: 37-60.(Develops the idea that legal concepts shall be defined in a contextual setting.)

Hart, H.L.A. (1961) *The Concept of Law*, Oxford: Clarendon Press.(Contributes to concept-formation within legal morphology by the distinction between primary and (three kinds of) secondary rules.)

Hedenius, I. (1941) *Om rätt och moral (On Law and Morals)*, Stockholm: Tidens förlag.(Author of the well-known distinction between sentences expressing legal norms and sentences stating the existence (validity) of such norms.)

Hempel, C.G. (1952) *Fundamentals of Concept Formation in Empirical Science*, Chicago, IL: University of Chicago Press.(Deals with concept-formation in empirical, mainly natural, sciences in general.)

Hohfeld, W.N. (1923) *Fundamental Legal Conceptions as Applied in Juridical Reasoning and Other Legal Essays*, ed. W.W. Cook, New Haven, CT: Yale University Press.(A classic work on different right-concepts with a contextual approach.)

Kelsen, H. (1960) *Reine Rechtslehre*, completely rewritten and expanded 2nd edn, Vienna: Deuticke; trans. M. Knight, *Pure Theory of Law*, Berkeley, CA: University of California Press, 1967.(Rewritten and expanded edition of one of Kelsen’s two major works.)

MacCormick, N. (1990) ‘Reconstruction after Deconstruction: A Response to CLS’, *Oxford Journal of Legal Studies* 10: 539-58.(Argumentation for the value of rational reconstruction of legal ideas within legal dogmatics, which to a large degree is a matter of concept-formation; see especially pp. 555-8.)

Raz, J. (1979) ‘The Functions of Law’, in *The Authority of Law*, Oxford: Clarendon Press, 163-79.(A contribution to the elaboration of a general classification of the functions of law, a task which also to a large degree is a matter of concept-formation.)

Ross, A. (1957) ‘Tû-tû’, *Harvard Law Review* 70: 812-25.(Argues that right- concepts are semantically meaningless vehicles of deduction.)

Wedberg, A. (1951) ‘Some Problems in the Logical Analysis of Legal Science’, *Theoria* 17: 246-75.(Epitomizes excellently the Scandinavian discussion in the 1940s and the early 1950s about the function of right-concepts in legal discourse.)

Legal discourse

'Legal discourse' signifies a strong interplay between law and language, linking together law as like language and law as itself language. However, unlike other linguistically modelled accounts, this approach involves a strong opposition to formalisms and their mirror-image realisms. Language as used cannot be 'deduced' from any pregiven matrix or set of propositions but must be studied in terms of its own modalities. The theory of law-as-discourse takes inspiration from the study of legal rhetoric and from socio-legal analyses of the courtroom, but was developed in its own right in the post-structuralist turn in linguistics. Law-as-discourse requires an understanding of the operation of legal talk in different registers, and gestures towards an intertwining of the social, the legal and the linguistic by focusing on the speaker-hearer situation, locution and action.

1 Law as rhetoric

While the currency of the term 'discourse' derives from post-structuralist developments in linguistics, this type of analysis was inaugurated in law by two important precursors, the New Rhetoric movement and socio-legal courtroom studies. Both, like post-structuralism, sought to break with logicist-formalist models of language and parallel models of law as a code or system of ideally transparent propositions of universal import held together by a formalizable syntax. This rethinking, in turn, posed the problem of how to reconceptualize 'the social' as both impossible to differentiate from 'the legal' yet as constitutively law's 'outside'.

The New Rhetoric movement, led by Chaim Perelman and Lucy Olbrechts-Tyteca beginning in the mid-1940s, sought to examine the process of argumentation as a way of using language, with its own prescriptions and proprieties to be understood through a return to Aristotle and the Roman jurists, discursive techniques employed to move an audience to a decision to act. But this return required the dethronement of demonstrative reasoning. 'Any theory of argumentation must break with the notion of reason and reasoning which comes from Descartes and has left its stamp on western philosophy for 300 years' (Perelman (1963: 134). For Descartes, the very existence of dispute was the sign of error, since truth was marked by transparency and intersubjective agreement. In this tradition, rhetoric had thus languished, cast at worst as a mode of duplicity, sophism and deliberate audience-manipulation involving pseudo-logic and base, populist, appeals to the emotions; or at best a source of literary stylistic flourish.

Perelman had himself begun in the logical positivist tradition, arguing that justice was an ultimately meaningless concept. Although formalizable as 'treat like persons alike', the values of what would count as like persons or as like treatment were always contested and hence arbitrary and socially contingent. Yet where did this leave law, at all levels functioning through competing arguments, in the two-party structure of the appeal court as much as the courtroom?

More radically, justice turned out to be the rule rather than the exception. *All* formalizations suffered the same fatal flaw. 'The insufficiency of the strictly formalist attitude becomes clear as soon as the question of interpreting a logistic system arises. But without attribution of meaning, axioms are not assertions' (Perelman 1963: 148) but merely arbitrary computations. Frege's conceptual Platonism, Tarski's factual correspondences, and the rise of conventionalism all testified to formalism's need for an extra-systemic semantic supplement - which, thus beyond control of the system, undermines its fundamental claim to autonomy and closure. Meaning/value was in turn (as with justice) always contestable. Thus, as it would be put subsequently, if law operates around a core of settled meaning, 'settled' does not mean pregiven and linguistically fixed but merely not currently in dispute (Jackson 1985: 156).

Shifting direction, Perelman came to see that disputation was a more orderly process that grew out of the engagement between speakers and hearers rooted in a common frame of reference, beginning with a prescribed understanding of their correlative roles and associated forms of diction (thus, teacher-pupil, fellow citizens). Assuming that all parties have a common concern to do justice, but that the audience's conceptions of this will begin in the familiar, the effective orator will tailor argument to the local audience. For every proverb there is a counter-proverb. The speakers thus lodge their recommended courses of action within contrary realms of meaning - different construals of what it is right to do - available via one or another 'locus of the preferable'. In a key example, Perelman 1979) contrasts the loci of quantity and quality respectively as asserting the durable, the useful,

the accessible, constant, symmetrically ordered, that which is useful to the greater number, the eternal and the abstract, being and form, as against the unique, the original and the difficult, the natural, the irreplaceability of the moment and the individual, the instant as the figure of totality, the poetic mode that recognizes the uniqueness of the other and the particularities of love and obligation. There are also prescriptions for the temporal ordering of argument.

Rhetoric is sometimes glossed as describing 'who can make arguments and where (and in what terms)', as does Perelman (1979) when commenting that rhetoric has always been a matter of political regulation. But here a great danger lies. In emphasizing the (indubitably) political nature of the designation of who can speak and hear, the social nature of discourse tends to be cast as an externality. Correspondingly, to speak of 'what' can be said loses the fact that the 'what' of speaking consists in the 'how'.

This example also holds together a fundamental tension in Perelman's work. The drive on the side of the locus of quantity takes him to develop a normative account of argumentation as justification in relation to a universal audience and a democratic politics, in keeping with rhetoric's birth in the agora of open political debate and a broad conception of logic that encompasses probabilistic reasoning. This is what links his work to rationalist theories of legal discourse (see Legal reasoning and interpretation §3; Habermas, J. §4).

On the other hand, Perelman also recognizes rhetoric as having a certain generic affiliation with the locus of quality, being ever about materialities rather than abstractions, audiences being always local. Rhetoric and common law reasoning share the need to cast the new in relation to tradition and customary practice and their 'crabwise' form of connection by way of analogy. Hence too law's generic conservatism and endorsement of any prevailing 'sense of normality, or likelihood' (Goodrich (1987: 181). More hopefully, others, such as Boyd White 1985), would emphasize the heterogeneity of community with its contesting hierarchies of value: what is common as merely the starting point of cultural transformation of meaning.

To follow the materialities of rhetoric is perhaps most promising. Goodrich's *Languages of Law* (1990) thus associates together rhetoric's use of place (topos/locus) and sequence with the bodily, the way that memory techniques are situated in imageries of dwelling places, or streets, the idea of patterns linking cases, with the idea of habit as images or emblems of how to live. Similarly, the figurative dimensions of texts may speak a different and more revealing story than their propositional mode, as Douzinas, following de Man, has shown most effectively by looking at Dworkin's Hercules metaphor (Douzinas *et al*. (1991). The dimension of time and sequence is equally important: moving an audience means moving them through a series of positions and differential engagements or 'identifications' in which each stage is dependent upon the former. Legal discourse becomes a kind of memory garden. Rhetoric, as Perelman constantly emphasized, is about effecting a *change* in the hearer, by contrast to the atemporal quality of demonstrative reasoning. In this mode Bernard Jackson 1988) has argued that the 'real time' of the trial process cannot be reduced to its retrospective view, in which the end is summed in the eternal non-time of legal or logical justification.

2 Courtroom studies

Socio-legal studies of the courtroom are the natural locus of legal rhetoric as forensic argumentation and, in turn, are ever in danger of lapsing into seeing rhetoric as low persuasion. However, in contributing to the study of law as discourse, they deal with two key issues: differential audience address and the relation between legal and non-legal discourse.

Law's address is multiple. Perelman considered this divided locution in terms glossed by Hart (in the introduction to Perelman (1963): the 'rhythms' of the 'natural' 'common' language are quite different from the 'technical language' common to members of a discipline or profession, language that is the 'private preserve of initiates' 1963: 156). As Pat Carlen (1976) insisted, this split address raises serious problems of legitimization, law's continuing need, internally and publicly, to be seen to be acting according to principles of justice and fairness. This is constantly under strain because of law's existence as both formal-constitutive abstract rules and situational rules determining the use of law in practice, which she links to the socio-linguistic categorization of elaborated and restricted codes. The elaborated code being characterized in terms of flexibility, the availability of many alternative expressions, abstraction, context-independence, enabling talk between strangers, and characterizes law's universalistic diction, its general language of justification. But, in its situational usage, law operates

according to a restricted code, highly concrete and local, a kind of argot, highly context-dependent, indexical, non-transportable and poor in substitute expressions, a use of language typical of closed communities with fixed structures - in this case primarily the closed community of lawyers. The lay 'audience' - which includes parties to the proceedings - thus suffers a double cognitive dissonance. (Carlen then goes on to describe the 'remedial routines' by which any resulting protests are discounted.)

Bennett and Feldman's focus (1981) is closer to the concern with argumentation, now inflected in terms of the question how juries make sense of legal arguments in trial courts in reaching their verdict. They find the bridging feature to be not the classical tropes of rhetoric but, rather, the literary mode of narrative. The story-form allows legal categories to be filled in with everyday meanings. Argumentation here means, precisely, the ability of prosecution or defence to produce a properly constructed story, potentially outweighing a more 'scientific' assessment of the evidence (as well as importing social prejudice). While providing an important counter to more abstract and philosophical understandings of narrative as coherence (MacCormick (1978), they barely glimpse the significance of narration - the way of telling the story: that very I-you relationship central to rhetoric - and law as discourse.

3 Legal discourse

The official theorization of law as discourse draws directly on linguistics, in which the term 'discourse' exists as a paradox. It begins as a stated definition: any stretch of language longer than a minimal unit of meaning (conventionally, a sentence). But filling out that definition immediately puts in question the ability of any 'minimal' unit to function autonomously, to have meaning in isolation from the discursive context in which it is bedded. Any smaller unit - a sign - does not have meaning, yet a sentence is already a miniature 'environment'. This is a more sophisticated statement of Perelman's critique of formalism as always in need of a semantic supplement.

Linguistic formalism is here construed broadly to encompass both Anglophone reference-based logical positivism and Francophone code-based structuralisms. Formalism's failure is that it cannot account for the meaning or function of any particular instance of speaking, carving it up, rather, between a 'logical deduction' from the system - the actualization of one of a matrix of given possibilities - and the residue, cast as unanalysable social and psychological contingency that can be registered only diachronically when such usage builds up sufficient critical weight to feed back into redefining the system itself. Discourse theory begins, rather, from speaking, the speaker-hearer relationship and how that is linked to its discursive context. The meaning of terms cannot be taken as transparent but is determined by its existence in different discursive formations and use in different registers.

The most ambitious attempt to theorize law in these terms is found in Peter Goodrich 1987). Drawing together a barrage of discourse theorists, but centrally Pêcheux (1982), the book offers a double attack on conventional jurisprudence (see Legal hermeneutics). On the one hand, most of modern jurisprudence is cast into the realm of formalism with its symptomatic failings. On the other hand, those things which jurisprudence takes to be law's essential features - its unity and coherence, its singular monolithic voice of authority, can themselves be shown to be effects produced by identifiable discursive techniques and social violences. (There is a perhaps surprisingly *marxisant* tone reminiscent of exposés of the tricks of ideology.)

Law's self-representation is disposed under the category of 'intra-discourse', designated the surface level, that describes the actual discursive techniques of law's utterances. The image of law as a coherent, auto-authorizing body is thus effected by law's continuing construction of its own history, as maintained by closure mechanisms of exclusion and discursive techniques of linkage, co-reference and internal continuity yielding the sense of a location in a continuous time: all context-dependent mechanisms that found law's claims to be universal and context-independent. Unlike Perelman, who strongly contrasted law as rhetoric (as a continuous address to an audience) to dialectic and other forms of dialogue, Goodrich claims that law pretends to be a dialogue but is really a monologue.

The level of institutionalization is, similarly, presented in terms of what is excluded from visibility and, further, posited very strongly here as a social externality that determines membership conditions of law's internal point of view. Less crudely, these cultural acquisitions evoke something of Bourdieu's account of 'knowing how' as a set of dispositions or postures, or Duncan Kennedy's 'training for hierarchy' that might locate the lawyer's

'belonging' in a more properly inarticulate and generally cultural form.

Third, and finally, there is the level of inter-discourse, law's relation to other discourses. Law's univocal mode suppresses the multi-vocal heteroglossalia of the social and law's self-referentiality means that it is an impoverished language cut off from denotation. 'Far from being expressly bound by any extra-linguistic content or denotative meaning, law's terms are connotative and symbolic: they lend themselves to an obfuscating, rhetorical or symbolic usage, to figurative or analogical manipulations within which their actual menage or referent is obscure' (Goodrich 1978: 179). This is a rather puzzlingly traditional characterization.

Although this is a highly reduced account of the book, in many ways, this is not the best of Goodrich's texts to study in order to understand what it means to look at law as discourse. What he provides elsewhere is a rich and dense series of accounts of law as compacting together many different specific genres of writing, traditions of interpretation and inscriptions of memory. It is in these works, rather than theorization, that the fruitfulness of approaching law as discourse can be seen.

See also: Law, philosophy of

BEVERLEY BROWN

References and further reading

Bennett, W.L. and Feldman, M.S. (1981) *Reconstructing Reality in the Courtroom*, New Brunswick, NJ: Rutgers; London and New York: Tavistock.(A much cited attempt to consider law as narrative that, unfortunately, has a rather undeveloped account of narrative structure.)

Bourdieu, P. (1984) *Distinction: A Social Critique of the Judgment of Taste*, London: Routledge & Kegan Paul. (Seminal work on the anthropology of culture, delineating the class-differentiated formation of aesthetic taste.)

Boyd White, J. (1985) *Heracles' Bow: Essays on the Rhetoric and Poetics of the Law*, Madison, WI: University of Wisconsin Press.(A rather literary approach to law as rhetoric that provides a good sense of moving the audience through a series of moral positions.)

Carlen, P. (1976) *Magistrates' Justice*, Manchester: Martin Robertson.(A classic text in socio-legal studies of the courtroom that draws together a wide range of theoretical approaches applied to empirical material.)

Douzinas, C. and Warrington, R. with McVeigh, S. (1991) *Postmodern Jurisprudence: The Law of Text in the Texts of Law*, London: Routledge.(A critical analysis of contemporary mainstream jurisprudence that gives full play to analysing texts by their metaphors. Key text for laws as discourse.)

Goodrich, P. (1987) *Legal Discourse: Studies in Linguistics, Rhetoric and Legal Analysis*, London: Macmillan. (Theorizes legal discourse by reference to a wide range of theories of linguistics. Fundamental for exposition.)

(1990) *Languages of Law: From Logics of Memory to Nomadic Masks*, London: Weidenfeld and Nicolson. (Exploration of different discursive presences in law, especially aspects of the 'symbolic' - visual, mnemonic, Christian, with an argument that the masks of contemporary law have become dislocated from their spatial locus.)

Jackson, B. (1985) *Semiotics and Legal Theory*, London and New York: Routledge & Kegan Paul.(A serious and scholarly analysis of law in terms of structuralist (Greimasian) analysis and which also looks into questions of pragmatics and semantics.)

Jackson, B. (1988) *Law, Fact and Narrative Coherence*, Merseyside: Deborah Charles. (Beginning from courtroom studies, seeks to argue a general account of legal norms as narrative.)

Kennedy, D. (1986) 'Freedom and Constraint in Adjudication: A Critical Phenomenology', *Journal of Legal Education* 36.(Original statement of Kennedy's famous notion that the methods of legal education replicate law's ordering as a hierarchy of values and sources; hence legal education as a 'training for hierarchy'.)

MacCormick, N. (1978) *Legal Reasoning and Legal Theory*, Oxford: Clarendon Press.(Classic modern text on jurisprudence comparable in stature to the work of Dworkin and, similarly seeking to deal with the residue of law that does not easily fit formalism-positivism.)

Pêcheux, M. (1982) *Language, Semiotics and Ideology*, London: Macmillan.(Key example of French development of discourse theory.)

Perelman, C. (1963) *The Idea of Justice and the Problem of Argument*, London: Routledge & Kegan Paul; Atlantic Highlands, NJ: Humanities Press.(Includes the early logical positivist article on justice and subsequent developments of the argument.)

Perelman, C. (1979) *The New Rhetoric and the Humanities: Essays on Rhetoric and its Applications*, Dordrecht,

Boston and London: Reidel Synthese Library, vol. 140.(Classic text available in English outlining not only law as rhetoric but the wider project.)

Legal evidence and inference

In the field of law there is a rich legacy of scholarship and experience regarding the properties, uses and discovery of evidence in inferential reasoning tasks. Over the centuries our courts have been concerned about characteristics of evidence that seem necessary in order to draw valid and persuasive conclusions from it. Thus, they have been led to consider such matters as the relevance of evidence, the credibility of the sources from which it comes, and the probative or inferential force of evidence. Court trials usually involve inferences about events in the past. The past can never be completely recovered. In addition, evidence about past events is frequently inconclusive, conflicting or contradictory, and often vague or ambiguous. The result is that inferences about past events are necessarily probabilistic in nature. Our courts have also been concerned about whether the interests of fairness require that, on occasion, evidence might be inadmissible, even though relevant and credible. Evidential and inferential issues such as the ones just mentioned are also of concern to philosophers and persons in other disciplines. This entry concerns several evidential issues of particular interest in legal contexts.

1 Proof and admissibility

Evidence in legal contexts has some distinctive features. Two issues are central in modern scholarship on evidence and inference in legal contexts. The first concerns study of the properties and uses of evidence in the process of judicial proof. To be persuasive in the process of proof, evidence must have certain ‘credentials’ established by defensible arguments (see §2 below). The second issue concerns study of rules regarding the admissibility of evidence. Not all legal systems have such rules. In ones that do, evidence with justified proof-related credentials is still not necessarily admissible at trial. Concern about the proof credentials and admissibility of evidence has not always been as great as it is today. This concern intensified very slowly as adversarial processes for settling disputes and the concept of jury trials came into existence.

In England in the eleventh and twelfth centuries, persons accused of crown or criminal offences had the burden of proving their innocence. This burden was discharged in three stages: judgment, trial, sentence. At the judgment stage the accused was obliged to select one of three methods of ‘proof’: by the swearing of oaths by witnesses, by ordeals of various kinds, or by combat. The trial thus simply consisted of the implementation of one of these methods. The trial outcome depended not on any evidence but on the judgment of God. It was then believed that, if the accused was truly innocent, God would strike down anyone who gave a false oath against the accused, would heal any injury inflicted on the accused during an ordeal, and would side with the accused in a trial by battle. Sentence was then delivered in accordance with God’s verdict. With the decline of proof methods based upon oaths, ordeals and combat, jury trials gradually came into existence. But in their earliest forms these juries consisted of witnesses against the accused as well as other persons having vested interests in the case. They were not seen as objective triers of facts.

Interest in evidence emerged as the jury came to be identified not as witnesses but as a body of persons who deliberate impartially upon evidence given by external witnesses for both sides of a dispute. Thus, advocates came to be concerned about the evidential merits of their own cases and how effectively they could counter the evidence and arguments given by an opponent. A landmark case occurred in 1670 regarding the role of jurors in the process of proof. Prior to this time, jurors could be attainted for giving verdicts contrary to what the court regarded as manifest evidence. In the case of Edward Bushell, Chief Justice John Vaughan ruled that jurors could not answer questions of law but courts could not answer questions of fact. Thus, in Anglo-American courts today jurors and not the court (unless it is a bench trial) assess the credibility and force of evidence during the process of proof. As the right to cross-examine opposing witnesses became established, Sir Matthew Hale could claim (in 1739) that the adversary system, involving the parties in contention, their counsels, and the court, was unmatched for ‘beating and boulting the truth’. (As against this, the twentieth-century legal realist Jerome Frank was to compare adversarial methods to throwing pepper in the eyes of a surgeon.)

Interest in rules governing the admissibility of evidence also emerged as the adversarial system developed. Concern about the admissibility of hearsay evidence, for example, arose following the trial of Sir Walter Ralegh in 1603. Ralegh was convicted of treason partly on the basis of evidence having no better status than rumour or gossip, since the source of it was never identified. Rules concerning the admissibility of hearsay date from the

acknowledged injustice in Ralegh's case. Since those times, what are now termed the laws of evidence began to emerge. In one sense these laws of evidence consist of assorted exceptions to a principle of free proof (virtually all evidence admissible), as advocated by Jeremy Bentham and now practised, for example, in Sweden. These laws of evidence have hardly been stationary. They have changed over time and will almost certainly continue to do so in response to continuing debates about their fairness, changes in political climate, and changes in patterns of litigation.

Even if all rules of admissibility were abolished, there would still be concern about what constitutes proof. In discussing evidence and its inferential roles in proof, we encounter matters that are continuing sources of debate among legal scholars, philosophers, probabilists and others. Settlement of any dispute, however it is accomplished, rests at some stage upon arguments based on evidence. Few episodes of litigation, whether criminal or civil, actually come to trial. Thus it would be a mistake to consider the trial as paradigmatic in any study of the properties and uses of evidence. This acknowledged, bodies of evidence encountered in any litigation have certain well-known properties. The evidence is necessarily incomplete, often on important matters. Items of evidence are usually inconclusive to some degree, meaning that the evidence is consistent with more than one probandum, hypothesis or fact in issue. Evidence is often imprecise or vague; we cannot tell exactly what it means. Evidence comes to us from sources which are not perfectly credible. Finally, bodies of evidence usually reveal interesting patterns of dissonance; some items seem to favour one conclusion, others a different conclusion.

These characteristics of bodies of evidence combine to make any conclusion drawn from them probabilistic in nature. Thus, forensic standards of proof involve hedges or qualifiers (discussed in §§4-5 below) that acknowledge these omnipresent characteristics of evidence. Judicial proof, like proof based on incomplete and fallible evidence in other contexts, is probabilistic rather than demonstrative or necessary. There is no forensic standard that requires proof with certainty or beyond the possibility of doubt.

2 The credentials of legal evidence

In the Anglo-American judicial system, a rationalist-empiricist view of the nature of proof in adversarial proceedings has been predominant in the past two centuries. On this view, matters at issue in a dispute are to be proved at specified levels of probability on the basis of evidence that is both relevant and credible or believable. Such a process seems rational, at least with respect to earlier proof methods involving oaths, ordeals or combat, as well as juries composed of adversary witnesses. The empiricism stems from the requirement that proof be based upon evidence observable in some way to those who must render verdicts in disputes. Two of the most prominent rationalist-empiricists have been Jeremy Bentham and, in the USA, John Henry Wigmore.

There is considerable dispute about what constitutes rationality in all of human inference and choice. Regarding evidence and inference, part of this controversy concerns the means by which the probative or inferential force of evidence ought to be assessed and then combined in reaching a conclusion (see §§4-5 below). Choices made by those who participate in settling disputes involve more than evidence, inference and probabilities; they also involve value-related ingredients. In any dispute there are multiple stakeholders including the parties in contention, their advocates, court officials and society at large. Settling disputes is difficult given the competing and often internally inconsistent value systems of these stakeholders. Any pretension to rationality in legal matters appears to involve a seamless integration of matters concerning both the probability and value ingredients of choice as well as the process by which they are combined.

It appears that an item of evidence must have three credentials in order for it to be in any way persuasive during the process of proof. The necessary credentials concern the relevance, credibility and probative (inferential) force of evidence. Of course, no evidence comes with these three credentials already established; they have to be justified by cogent arguments, often having many stages, and by careful judgments (see §3 below). An item of evidence is commonly said to be relevant in a legal dispute if it makes the existence of some material fact at issue 'more probable' or 'less probable' than it would be without the evidence. This is often put subjectively: evidence is relevant if it would cause anyone to change their beliefs about the likeliness of some material fact at issue. We might also say that evidence is relevant if it has some probative force in belief revision. There are no rules in law saying how much force any item of relevant evidence ought to have.

Evidence can be relevant in different ways. It is said to be directly relevant if a chain of reasoning can be formed

from the evidence to a major issue in the case. There are two species of directly relevant evidence: direct and circumstantial. Direct evidence goes in one reasoning step to some fact at issue and, if credible, would settle the matter. Circumstantial evidence, even if perfectly credible, supplies only some but not complete grounds for belief in some matter. But some evidence has only indirect relevance; such evidence is said to be ancillary or auxiliary in nature. Ancillary evidence either strengthens or weakens links in chains of reasoning set up by directly relevant evidence. Generalizations are asserted to license a reasoning step or link. Ancillary evidence bears upon whether or not this asserted generalization holds in the present case in which it is being invoked. For example, we might assert the generalization: 'if a person testifies under oath to the occurrence of an event, then this event probably did occur'. Tests of this generalization, applied to a particular witness, involve ancillary evidence concerning this witness's veracity, objectivity and observational sensitivity. We might say that ancillary evidence is evidence about the strength of other evidence. The relevance-status of evidence obviously depends upon the case at hand; a datum can be relevant evidence in one situation and quite irrelevant as evidence in another.

Suppose an item of evidence whose relevance can be established. The question then is: can we believe what this evidence says? Credibility considerations form the very foundation for all arguments based on evidence; that is, they form the first links in a chain of reasoning. There are different discernible forms of evidence (see §3 below); judgments about the credibility of evidence depend upon what form evidence takes. For tangible evidence, open to direct inspection by fact-finders, such matters as the authenticity and chain of custody are of interest as well as the accuracy of the evidence if it exists in the form of some sensor record such as a photograph or radar image. For testimonial evidence involving the assertions of human witnesses, credibility assessment is quite different. We first need to establish how a person obtained the information they report: (1) by direct observation; (2) at second-hand from another source (possibly hearsay); or (3) by inference from other information (opinion evidence). Like other human characteristics, the credibility of a witness has many attributes. Among the important attributes of witness credibility are veracity, objectivity and observational sensitivity or accuracy.

Assessing the probative force of an item of evidence is a judgmental matter for which there are no legal rules. Many factors enter into such judgments such as the credibility of the source from which the evidence comes, the probabilistic strength of links in chains of reasoning from the evidence, the number of links in the chain, the rareness or improbability of the events reported in the evidence, and the possible influence of other evidence (see also §4 below).

Courts may rule on whether an item of evidence is relevant, but fact-finders must judge the credibility and probative force of evidence. Judgments about any of these three credentials of evidence will naturally depend upon a person's standpoint or frame of reference. Disagreements about the relevance, credibility and probative force of evidence are to be expected in legal and in other contexts. One major reason is that persons having different standpoints, backgrounds, objectives and values will effectively view the same evidence through different perceptual 'lenses'.

3 Structural issues

Analyses of the relevance, credibility and probative force of evidence involve both structural and probabilistic issues. Structural issues arise during the construction of arguments or chains of reasoning to justify these three credentials of evidence. There are several major structural difficulties, the first of which concerns the fact that chains of reasoning we construct often have many links. Each link inserted in a chain of reasoning from an item of evidence to some major probandum or hypothesis identifies a source of uncertainty recognized by the person constructing the chain of reasoning. Both Bentham and Wigmore recognized that there are very few inferences in which a chain of reasoning has just one link. Wigmore applied the name 'catenated' to those inferences involving two or more links. Today such inferences are said to be cascaded, multi-stage or hierarchical; the term 'inference-upon-inference' is also employed. Until recently, there has been scant reference outside the field of law to catenated or cascaded inference.

The second structural difficulty concerns arbitrariness in the construction of a reasoning chain. Two persons may justifiably construct different chains of reasoning from the same evidence to the same probanda; in other words, they take different reasoning routes. They may of course each take reasoning routes from the same evidence to different probanda. The logician John Venn noted that both the number and the labelling of stages of reasoning is often quite arbitrary. Wigmore emphasized that the form an argument takes depends upon who constructed it. In

short, there are no normatively ‘correct’ arguments. This has a distinct bearing on the issue of rationality in inference since more than one argument may be justifiable from the same evidence. Plausibility and the avoidance of *non sequitur*s seem to be our major guides in argument construction. The construction of any argument is in fact a creative act, as is the discovery or generation of evidence and possible probanda in the first place.

In evaluating the credentials of evidence a major structural difficulty in legal and other contexts is that we usually have masses of evidence to contend with. Arguments constructed from masses of evidence can be stunningly complex and, when depicted graphically, they resemble complex networks; they are now in fact called inference networks. Wigmore (in 1913) was the first to study complex inference networks based upon masses of evidence. He devised a graphic scheme for charting arguments from evidence to the ultimate probandum in a case. Modern analysis of Wigmore’s evidence charts shows that his inference networks can be characterized mathematically as directed acyclic graphs. A network is a kind of graph structure containing two basic elements: nodes and arcs. In an inference network nodes represent evidence and interim or final probanda (propositions to be proved); arcs represent probabilistic linkages among nodes. Inference networks are directed because reasoning proceeds inductively from evidence to major or ultimate probanda. They are acyclic because there is no path of arcs from any node that brings you back to this same node; this would amount to being in an inferential loop.

Inferences involving legal disputes tend to be well-structured from the top down. To prove a major fact at issue, at some specified level of probability, certain elements must be established. These elements, called penultimate probanda, stem from substantive law concerning the legal matter at hand. For example, a case of first-degree (premeditated) murder may involve the penultimate probanda: (1) Y was killed; (2) it was defendant X who killed Y; (3) X intended to kill Y (it was not an accident); and (4) X formed this intent beforehand. In essence, each of these penultimate probanda serves as a touchstone for determining the relevance of evidence.

Attention to structural matters provides a basis for categorizing evidence without regard to its substance or content. Jurists, and perhaps historians, must be prepared to evaluate evidence having virtually any conceivable content or substance. The relevance and credibility credentials of evidence discussed above assist first in the identification of distinguishable forms of evidence. We ask two questions: (1) how does the evidence stand in relation to the person evaluating it? and (2) how does the evidence stand in relation to the major facts at issue?

Answers to the first question allow us to identify properties of evidence that govern how its credibility is to be assessed. Discussed in §2 above are two of the basic forms of evidence: tangible and testimonial; there are others. In some cases evidence expected but missing can itself be taken as evidence. Missing evidence is not the same as negative evidence, the report of the nonoccurrence of an event. It has been recognized in law for quite some time that the non-production of evidence invites the inference that the evidence is unfavourable to the interests of the person who refuses to provide it. Other evidence comes from various authoritative sources such as almanacs and various tables of physical or chemical phenomena. In addition, there are certain facts that can be accepted without further proof, such as: strychnine in sufficient doses is lethal. Information from authoritative sources and certain accepted facts may be judicially noticed by courts and accepted without further proof. Answers to the second (relevance) question above allow us to identify evidence as being direct, circumstantial or ancillary (as noted in §2 above). Combinations of answers to these two questions allow us to categorize any evidence. For example, we may have tangible evidence that is circumstantial on some matter at issue, or have testimonial evidence that is indirectly relevant or ancillary.

In purely structural terms, we may also identify recurrent combinations of evidence. Dissonant evidence comes in two basic forms: contradictory evidence and conflicting evidence. Contradictory evidence involves events that cannot happen together, in which case we naturally look to the credibility of the sources from which it comes. Conflicting evidence, however, concerns events which seem to favour different probanda or hypotheses but which could possibly occur together. In resolving such conflicts we also look to the credibility of our sources; but we must also consider the reasons why we say the evidence is conflicting. Perhaps we could explain away this apparent conflict with better understanding of the situation at hand. Harmonious evidence also has two basic forms: corroboration and convergence. Two forms of corroboration can be discerned. In one case, we have successive reports of the same event. In another case we have ancillary evidence that favours the credibility of a source reporting a certain event. Convergent evidence involves different events we believe to favour the same probandum or hypothesis. Finally, some combinations of evidence exhibit synergism in the sense that one item of

evidence seems to enhance the probative force of another. In other situations, one item of evidence may seem redundant, and thus have less inferential force, when considered in light of other evidence.

4 Probability and the probative force of evidence

Evidence is relevant if it has some probative force on material facts at issue. How much probative force an item or body of evidence has can only be graded in probabilistic terms; this is a consequence of evidence being incomplete, inconclusive, imprecise and not completely credible (see §2). There are no legal rules that prescribe how the probative force or weight of evidence should be assessed; such assessment calls for judgments on the part of jurors or judges (in bench trials). The task of weighing evidence rests upon the ordinary or common-sense reasoning ability that fact-finders are credited with possessing. Since the early 1600s, when interest in probability calculations emerged, there have been intermittent attempts to determine the extent to which mathematical probability theories can offer any guidance in the task of weighing and combining evidence in legal affairs. Some of the very earliest probabilistic analyses involved study of the relation between the probative force of testimony and the credibility of its sources.

Probabilistic analyses of the task of assessing the probative force of evidence have drawn a mixed reaction on the part of evidence scholars in law. One reason is that the theory of probability stemming from the work of Blaise Pascal (§5) in the seventeenth century is rooted in enumerative processes such as those encountered in determining probabilities in games of chance or in replicable statistical analyses. Inferences in law, however, usually concern singular or unique events that are not replicable or subject to enumerative analysis. In the conventional or Pascalian theory of probability, the one most persons learn about on first exposure to the subject, probabilities are taken to be numbers between zero and one (inclusive) that are additive across mutually exclusive events. A process called conditioning allows for the revision of a probability in light of new evidence. A consequence known as Bayes' rule, consistent with the properties of Pascalian probability and the manner in which a conditional probability is defined, emerges as a canon for probabilistic reasoning based on evidence. This rule has been the subject of controversy among philosophers, probabilists and statisticians. Some, but not all, believe Bayes' rule to be the canon for inductive and probabilistic inference (see Probability theory and epistemology).

Pascalian probability is subject to a variety of interpretations in addition to those involving aleatory (chance) or relative frequency (statistical) situations. One alternative, congenial to the application of Bayes' rule, involves viewing Pascalian probability in an epistemic sense to indicate a person's degree of subjective or personal belief, based on evidence, that a certain proposition is true or that some event has occurred or will occur. This interpretation allows singular or unique events to be graded probabilistically. Terms in Bayes' rule called likelihoods and likelihood ratios form one very useful way of grading the probative force of evidence. Study of these ingredients of Bayes' rule allows the capture of a wide assortment of evidential subtleties that arise in the task of assessing the probative force of the various forms and combinations of evidence mentioned in §3 above. At the same time, however, various difficulties with a Pascalian interpretation of probability, and the use of Bayes's rule, have led to some well-articulated alternatives to the Pascalian system of probability. Study of the relative merits of these alternative formal systems forms one basis for the 'probability debates' now taking place among scholars of evidence in jurisprudence.

One natural question concerns why we should expect the Pascalian system, rooted in enumeration, to be necessarily adequate in capturing all observed attributes of the epistemic probability judgments so often made regarding singular or unique events? The additivity of probabilities in the Pascalian system requires a complete commitment of belief to an event and to its complement. For example, if on evidence we believe probandum H has probability 0.8, we must also say we believe that not-H has probability 0.2. In recent years a formal non-additive system of probabilistic reasoning called 'belief functions' has been developed that allows a person to be uncommited in various ways in making epistemic probability judgments. This system, also rooted in much earlier studies, allows capture of a variety of credal states that are elusive to Pascalian probabilities. For instance, belief functions allow us to distinguish between lack of belief and disbelief, to capture the concept of ignorance as lack of evidence, and to give a more precise meaning to the term 'doubt'.

Since the time of Francis Bacon (§4), it has been recognized that inductive reasoning can involve the eliminative rather than the enumerative use of evidence. In recent years, a system of probabilities called 'Baconian probabilities' has been developed that is expressly congenial to induction by elimination. In eliminative induction

alternative probanda or hypotheses are subjected to a variety of different evidential tests, each one designed to eliminate at least one of the hypotheses. A hypothesis surviving our best attempts at variative elimination has the highest Baconian probability. A distinguishing feature of Baconian probability is the importance it places on the completeness of coverage (or the sufficiency) of evidence. The weight of evidence in Baconian terms depends not only upon questions we have answered by evidence but also upon recognized questions we have not yet answered with evidence. One very important element of the current probability debates concerns the extent to which judicial proof may be eliminative in nature.

In the process of judicial proof, probability statements are made in linguistic terms rather than in numbers. Thus we have such forensic proof standards as 'beyond reasonable doubt', 'clear and convincing evidence', and 'balance of probabilities'. Attempts to cast these standards in precise numerical terms inevitably fail; these forensic standards acknowledge the imprecision latent in so much of legal reasoning. The common-sense generalizations we assert to license stages of reasoning are stated imprecisely. For example, we assert: 'persons who utter threats *often* intend to carry them out'. Since no one compiles any relevant statistics on such matters, we use imprecise linguistic qualifiers such as 'usually', 'often', 'frequently' or 'probably'. Evidence can also be imprecise; we may have evidence reporting that a witness has not very good eyesight or that it was very dark when the witness made an alleged observation. There is now in existence a well-developed formal system of fuzzy inference and fuzzy probability that permits some very precise ways of coping with the imprecise or fuzzy linguistic variables so often encountered in judicial proof. This system rests upon a multivalued logic instead of the two-valued logic forming the basis for other formal systems of probability (see Fuzzy logic §3).

It does appear that the task of assessing the probative force of evidence is far too rich an intellectual activity to expect that any single formal system of probability can capture all of this richness. Each one of the formal systems just discussed is informative and resonates to some but not all of the important attributes of the task of weighing evidence.

5 Inference from a mass of evidence

Inference from evidence in judicial proof is embedded in the further process of choice. Following a trial, courts and jurors render verdicts and decide upon consequences. In other forms of settlement choices are made by both sides during the process of negotiation and bargaining. The inferential element in all such decisions involves the very difficult task of drawing conclusions from masses of evidence in which all the forms and combinations mentioned in §3 above are likely to be represented. There are no laws of evidence telling fact-finders how to combine all the evidence in reaching a final conclusion. Although the various probability theories discussed in §4 offer suggestions about how such complex inference tasks might be decomposed to simplify them, the actual inferential judgments made by fact-finders are holistic and are neither decomposed nor otherwise assisted by means of any formal theory concerning the combination of evidence. There is now an abundance of empirical research on strategies that fact-finders employ in drawing conclusions from masses of evidence.

The only inferential guidelines offered in the laws of evidence concern belief intensity thresholds required for verdicts in criminal and civil cases. These belief thresholds exist in the form of the imprecise or fuzzy forensic standards of proof such as 'beyond reasonable doubt' (in criminal cases) and 'balance of probabilities' or 'preponderance of evidence' (in civil cases). One element of the current 'probability debates' concerns whether or not contemporary theories of probability offer ways of adding precision in prescribing these forensic standards.

See also: Law, philosophy of; Legal reasoning and interpretation; Probability, interpretations of

DAVID A. SCHUM

References and further reading

Anderson, T. and Twining, W. (1991) *Analysis of Evidence: How to Do Things with Facts Based upon Wigmore's Science of Judicial Proof*, Boston, MA: Little, Brown. (Expansion of the structural matters discussed in §3 of this entry.)

Cohen, L.J. (1977) *The Probable and the Provable*, Oxford: Clarendon Press. (Comments on the necessity of alternative formal systems of probability to capture requisites of inference in legal affairs.)

Eggleston, R.E. (1978) *Evidence, Proof, and Probability*, London: Weidenfeld and Nicolson.(Comments from a legal scholar on matters discussed in §4 of this entry.)

Hastie, R., Penrod, S. and Pennington, N. (1983) *Inside the Jury*, Cambridge, MA: Harvard University Press. (Empirical studies of the processes jurors employ in drawing conclusions from masses of evidence; expansion of matters in §5 of this entry.)

Plucknett, T. (1956) *A Concise History of the Common Law*, Boston, MA: Little, Brown.(Chapter 4 concerns the emergence of the jury system as discussed in §1 of this entry.)

Schum, D. (1994) *The Evidential Foundations of Probabilistic Reasoning*, New York: John Wiley & Sons.(A study of the properties, uses, and discovery of evidence in probabilistic reasoning; particular emphasis on alternative views of the probative force of evidence. Expansion of material in §§2-4 of this entry.)

Tillers, P. and Green, E. (1986) 'Symposium: Probability and Inference in the Law of Evidence', *Boston University Law Review*, 66 (2-3): entire issue.(A record of debates concerning the application of probability theories in legal inference. Expands on §4 of this entry.)

Twining, W. (1985) *Theories of Evidence: Bentham and Wigmore*, Stanford, CA: Stanford University Press. (Expansion of the material in §§1-2 of this entry.)

Legal hermeneutics

Legal hermeneutics studies the interpretation and meaning of written law. By way of contrast with the related discipline of forensic rhetoric, the study of oral argumentation and persuasion in court, legal hermeneutics is best defined as a textual discipline which dates back to the classical codifications of Greek and Roman law. While the term hermeneutics is derived from the messenger god Hermes and is used by Aristotle as the title of a work concerned with the logic of interpretation, Peri hermeneias, legal hermeneutics really only developed with the growth and medieval reception of Roman law. The Roman legal tradition was pre-eminently a tradition of ius scriptum *or written law, of codes and codifications. The custody, interpretation and transmission of the great texts of Roman law gave rise to an exemplary discipline of hermeneutics. Its task was that of preserving, collating, translating and applying the archaic, foreign (Latin) and fragmentary texts of the legal tradition to contemporary and vernacular circumstances far removed in time and space from the original contexts of the written law. In later eras and particularly in periods of crisis or renewal of the legal tradition, the problems of text, interpretation and meaning in law have led to the revival of the concerns and traditions of legal hermeneutics.*

1 1 Law, philology and hermeneutics

The first meaning of the term 'hermeneutics' is one which derives from the function of the god Hermes, who was said to have delivered the messages of the gods to the mortal realm. Hermes translated the divine idiom into human speech and thus became, together with his Latin form Mercury, the emblem of the task of translation between different orders, times and places. In so far as Hermes carried and communicated the words, the eloquence and the authority of the gods, his task was always tied to the question of law. In its broadest definition, hermeneutics identified the problem of meaning and interpretation with that of access to the divine source and spirit of all human law. Hermeneutics thus identified the art of interpretation with that of unveiling and translating the alien or hidden message of the divine or distant author of the laws.

While the dominant model of post-classical literary hermeneutics was theological and concerned the exegesis and interpretation of the Scriptures and the canons or laws of the Church, a parallel tradition of legal hermeneutics existed. The secular model for legal hermeneutics developed initially and somewhat paradoxically around the two main codifications of Roman law, namely the Twelve Tables from the fifth century BC and the sixth-century *Corpus iuris civilis*. These codifications, although vastly different in style, both endeavoured to state a divinely inspired and absolute body of law in a written form which would preclude the need for interpretation. The Twelve Tables dealt with disputed points of law and aimed, through the use of writing, to take the resolution of such issues out of the hands of the pontiffs, the interpreters of the law. So too the Eastern Roman emperor Justinian promulgated his massive codification of Roman law with the accompanying injunction that 'no-one - neither those who presently practise jurisprudence, nor those who will practise it in future - may presume to compose commentaries on these laws', a ruling which is explained slightly later by means of the equation of interpretation with perversions (*perversiones*) of law. The jurist or practitioner of legal hermeneutics was thus originally to know and obey the law rather than to construe it, precisely because its authority was far greater than that of any interpretation.

Hermes was also, however, a thief: he stole from the gods by virtue of taking their words and presenting them in secular and novel contexts. The art of legal interpretation also borrowed from this metaphor of theft of meaning in allowing that passage of time brought with it the decline or decay of law and change in the circumstances of its application. Where circumstances had changed or time had rendered the written law less appropriate to novel contexts, the jurist had to use first fiction and then later analogy to apply it. In both circumstances, whether through the fiction of treating the litigants 'as if' they fell within the written rule, or through analogy with the clear meaning of the text, the interpretation was subordinate to the text of law, the unwritten meaning an implication or logical inference from the authoritative text. The art of hermeneutics thus acted covertly to supplement the written law and to legislate new meanings in the guise of textual reconstruction. It stole the power of the lawmaker but it did so under the veil of observance or in the form of the self-effacement of the juristic interpreter.

In its classical periods, and particularly during the medieval reception of Roman law or 'twelfth-century revolution' in interpretation, legal hermeneutics took the form of a practical exercise in the philological

reconstruction of the 'true' or literal meaning of the texts of law. Constrained by the prohibition upon commentary, the interpreter of the law was restricted to the task of providing grammatical glosses and word-by-word explanations of the rules of law. Their task was to harmonize and systematize the archaic and fragmentary textual inheritance of law and through understanding it in ever more profound detail find ways of applying it to the circumstances and contexts of a later age. The glossator was the ideal type of the legal interpreter, an erudite but faceless technician of meaning, a philologist of law who purported to reconstruct and harmonize the terms and rules of written law without ever interfering or interposing between the letter of the law and its destination. In this form, hermeneutics was a discipline which guarded over both tradition and text, an art of custody and of an esoteric and exclusory knowledge of ancient rules or laws.

2 2 Hermeneutics and jurisprudence

The figures of Hermes and Mercury are also suggestive of passage and of turnings, of sleight of hand and the flight of words: Hermes and Mercury both guard and admit, conserve and create. The other face of legal hermeneutics is associated with the speed and slippage or transition that is associated with a law that was always conceived as Janus-faced or split between its divine and human sources. Hermeneutics was thus also associated with the magic or art of legal representation and with words of law, with Coke's *vocabula artis*, which hid behind their vulgar surface a thesaurus of artistic meanings and erudite truths. Hermeneutics was in this sense also always potentially a political and interventionist practice in which the interpreter of law would create new meanings and invent new applications for the cold prose, tired words or dead letters - Bacon's *litera mortua* - of written law (see Common law §2).

This stronger and more modern sense of legal hermeneutics developed as a reaction against the glossatorial tradition and came about primarily through the rediscovery of the classical tradition and law in the Renaissance. Legal humanism pilloried the glossatorial tradition of unthinking or 'imbecilic' reverence for the post-classical compilation and interpolation of Roman law and returned rather to the surviving sources of Roman law itself. In place of the glossators' slavish observance of textual letter and rule, humanism sought to comprehend the law in terms of the context of its historical enunciation. Roman law, according to François Hotman, was not an adequate model for contemporary law; rather it represented the property interests and privileges of the Roman bourgeoisie and should be understood in terms of (and restricted to) that social and political context. The rules of legal hermeneutic method devised by the humanists and associated particularly with Jacques Cujas, Andrea Aliciato and François Hotman on the Continent and with Christopher St German and later Sir Francis Bacon and Sir Edward Coke in England, entailed a much stronger philological sensibility and rhetorical awareness.

The exemplary form of Renaissance hermeneutics was the treatise on *Digest* 50.16, *De verborum significatione* (on the meaning of words) (see Roman law §2). While such treatises were still commentaries upon the classical Roman sources of law, and upon the terms and rules of the textual tradition, they went far beyond the singular text of the earlier tradition and expounded theories of interpretation which greatly exceeded the style of the gloss both in method and in scope of disciplinary reference. The first feature of such hermeneutics lay in the conscious comprehension of law as tradition and interpretation, as both written and unwritten, as sense and reference. The key interpretive maxim in relation to this approach was *Digest* 1.3.17 (Celsus): '[t]o know the law is not to know the words of the law, but its force and power [*vim ac potestatem*]'. In other comparable formulations the words of the law were contrasted to its 'root reason', 'intention', 'cause', 'rationale', 'purpose' and 'true sense'. The radical alterity of the past and the fragmentary character of its textual remains both forced the jurist towards a critical hermeneutics which could add to the literal meaning a covert sense or spirit of law, a meaning which could only be discovered by linguistic inference or supplementation, by reading between the lines *subauditio* or *subintellectio*. In short, the function of such a hermeneutics was to supplement the paucity of the textual remains of law and to recreate a fuller meaning by use of a wide array of philological, historical and linguistic techniques of reconstruction.

The second striking feature of hermeneutic jurisprudence lay in the methodologies developed to accommodate and expound the unwritten meaning or '*labile ratio*' of the legal text. Once it was explicitly recognized that the terms and rules laid down in antique texts could never apply directly, or without interpretation or 'interposition', to the linguistic context and social circumstances of contemporary disputes, the task of hermeneutics became that of finding something more in the law than was merely written. To this end the treatises on *De verborum*

significatione developed an extensive array of techniques of interpretation drawn from a wide variety of other disciplines. History, philology, logic and rhetoric were the principal sources for extension and innovation in relation to written rules and allowed for the systematic development of techniques of contextual interpretation and supplementation. While there was no one standard set of rules for the construction of legal meaning from between the lines of the law, a series of methods and contexts were developed which still play a role in the contemporary jurisprudence of interpretation. Most notably, construction of textual meaning would start from philological criticism of the text itself and would move to reconstrue it by reference to what may be classified synoptically as linguistic, legal and rhetorical contexts. The linguistic context would refer to etymological analysis as well as to grammatical rules of construction. The legal context refers not simply to the use of analogy in the extension of rules but also to the need to relate disparate rules and maxims, customs and local laws, to the specific text subjected to interpretation. The rhetorical context, finally, referred not only to the various forms of figuration by which meaning changes according to context and use but also to specifically legal senses of the intention of the author of the laws (*mens legis*) and the force or reason of their utterance.

3 3 Philosophy of language, precedent and hermeneutics

The various post-Renaissance revivals of legal hermeneutics have tended to coincide with crises and renewals of the legal tradition, associated most particularly with movements for codification or other reforms of law. The nineteenth century in particular witnessed a pan-European resurgence of hermeneutic method in the arts in general and in jurisprudence in particular. The peculiar concern of the new legal hermeneutics was that of the place of language and of meaning within the efflorescence of legislation and administrative regulation that accompanied the growth of the modern industrial state. While the method and models of textual understanding were still predicated upon the earlier theological and juristic conceptions of construction and interpretation, the more secular concern of legal philosophy was that of the place of meaning within the development of a science of law.

One model of hermeneutics returned to the glossator's exegetical conception of absolute and authoritative texts. The highly restrictive hermeneutics which developed around the *Code Napoléon* (1804), for example, asserted that all meaning was to be drawn from the legal text and from nowhere but the text. In a similar vein, Jeremy Bentham in England believed that codification of the law could take the problem of legal interpretation out of the arbitrary jurisdiction of the common law judges and return it to the people in the form of simple, clear and comprehensive textual expressions of the rules of law. Codification, however, was only ever a temporary solution to the hermeneutic problem of interpretation and passage of time, change of context and of circumstance. The indeterminacy of language inevitably raised again the problems of finding unwritten meanings which exceeded or displaced the surface sense or letter of the law. What was new about the post-Enlightenment hermeneutics was that philosophy of language and particularly phenomenology, social science, linguistics and literary criticism replaced the earlier traditions of philology and rhetoric as the principal disciplines to come to the aid of jurisprudence. It is significant, however, that when Hans-Georg Gadamer (§3) or Emilio Betti, arguably the two great hermeneutic philosophers of the twentieth century, turn to the definition of hermeneutics they place law at the heart of the discipline. In Gadamer‘s terms, law is exemplary of the historical problem of writing and it is as central to the proper formulation of the problem of meaning as that of the relation between historical sense and contemporary application, between text and judgment (Gadamer 1976).

The problem of hermeneutics, according to Gadamer, is the problem of law. It is that of creating contemporary meanings out of the antique and alien materials of tradition. The lawyer faces this problem in an acute form, and therefore is forced to resolve it in an exemplary way by reconciling the spirit of the tradition with the needs of contemporary communities. For Gadamer, legal hermeneutics could resolve the problem of the meaning of tradition by understanding the prejudices or pre-judgments by means of which the tradition already governed the contemporary instance. Gadamer's notion of 'the pre-schematisation of our European attitude' draws heavily upon Heidegger's conception of language, and particularly the philosophical language of the Greeks, as a code already imprinted with the metaphysics and the meanings which the subsequent tradition has lived out. For Gadamer and in a somewhat different form for Betti, the prior existence of linguistic meaning did not wholly resolve the legal problem of law-applying acts in the sense that neither linguistic tradition nor precedent could wholly determine the contemporary instance. The hermeneutic conception of understanding (*Verstehen*) also entailed an active sense of the inevitability of the creation and extension of meanings by recourse, in Gadamer's theory, to the spirit of the legal tradition through the hermeneutic unlocking of 'original' meanings and, in Betti's theory, through the use of

reasoning by analogy.

4 4 Critical legal hermeneutics

The problem of precedent and of the frequently disjunctive relation of norms or rules to decisions and to practices has consistently troubled the various modernist attempts to construct a science of law (see Legal reasoning and interpretation). Hermeneutics, and the problems of language and of writing more broadly, were a consistent threat to the jurisprudential systematizers of the common law tradition in particular. Predicated not upon a code but upon a disparate series of statutes, reported decisions, customs, maxims and other borrowings, the claim that law formed a coherent and unitary body of determinate rules was always hermeneutically precarious. While the positivist tradition of jurisprudence on occasion has recognized a certain latitude, *Verstehen* or 'internal attitude', in the interpretation of law as a system of rules, the concept of a legal system has most generally been presented, as for example in the work of the Austrian legal theorist Hans Kelsen, in such a manner as to exclude any strong conception of hermeneutics from the theory of law (see Legal positivism §5; Kelsen, H.).

Historicist and more properly hermeneutic reactions against positivistic and formalistic theories of law were contemporary with their original elaboration. Such reactions to legal science could be broadly speaking conservative, as for example was the German Friedrich Karl von Savigny's historicist legal hermeneutics, which endeavoured at the beginning of the nineteenth century to counter codification with a historical science of law. The other tendency of hermeneutics, however, was more critical and iconoclastic. What may be termed critical legal hermeneutics, drawing upon disciplines as diverse as sociology, linguistics and psychoanalysis, denies that there is any unitary body of law at the level of social systems and equally disputes the possibility of 'true', objective or singular legal meanings at the level of textual interpretation. The earliest political expression of a species of critical legal hermeneutics came with the US legal realist movement in the elite law schools of the USA in the early part of the twentieth century (see Legal realism §2). Concerned primarily with issues of statutory interpretation and the decisional practices of the courts, realist legal hermeneutics asserted the priority of judgment or court practice over written norms and so either implied or directly asserted the indeterminacy of legal meaning and the creative role of all law-applying acts.

Where legal realism was concerned primarily with drawing attention to the oral and contingent context of the courtroom and thereby challenged the naïve belief in a text-based science of law, subsequent developments of critical legal hermeneutics have focused almost exclusively on the indeterminacies or undecidability of written law. Variously located within critical race theory, feminist jurisprudence and critical legal scholarship, the contemporary hermeneutic challenge to the objectivity of the legal text is formulated in terms, respectively, of ethnicity, gender and politics (see Critical Legal Studies). Basing itself most broadly upon theories of textual indeterminacy, one of the key goals of critical legal hermeneutics has been to politicize law by evidencing the malleable and discretionary character of its textual practice. The legal text is a species of writing, a performative use of language or *écriture*, and it should be analysed as such before or as well as being consigned to the genre or tradition of legal science. In this respect, critical legal hermeneutics is concerned to treat the legal text as a form of literature and to subject it to the disciplines and dismemberments that the arts now throw in the way of textual meaning. Following the tradition of humanistic hermeneutics, but using contemporary disciplinary idioms, the critical legal endeavour has aimed to link legal language both to the inequity and to the literal and the metaphorical violence of law. Critical hermeneutics is aimed thus at producing an ethics of legal reading and of writing law, an ethics which ideally links text to judgment, word to act and, in its broadest philosophical expression, written law to the constraints of justice.

See also: Hermeneutics; Law, philosophy of

PETER GOODRICH

References and further reading

Berman, H. (1983) *Law and Revolution: The Formation of the Western Legal Tradition*, Cambridge, MA: Harvard University Press.(Accessible guide to the disciplinary development of the modern legal tradition.)

Betti, E. (1955) *Teoria della legge e degli atti giuridici (teoria generale e dogmatica)*, Milan: Guiffrà.(General statement of legal hermeneutic method.)

Douzinas, C., Warrington, R. and McVeigh, S. (1990) *Postmodern Jurisprudence: The Law of Texts in the Text*

of Law, London: Routledge.(Contains a critical discussion of contemporary hermeneutics as well as an introduction to postmodern theories of legal textuality.)

Gadamer, H.-G. (1976) *Truth and Method*, London: Sheed and Ward.(Crucial work on the history and philosophy of hermeneutics.)

Goodrich, P. (1995) *Oedipus Lex: Psychoanalysis, History, Law*, Berkeley and Los Angeles, CA: University of California Press.(Historical and critical account of common law hermeneutics, with particular attention paid to the relevance of a psychoanalytic hermeneutics of law.)

Goodrich, P. (ed.) (1997) *Law and the Unconscious: A Legendre Reader*, Basingstoke: Macmillan.(Translation of a selection of works by a French legal philosopher who has developed a unique theory of the historical sources of Continental legal hermeneutics.)

Justinian (533) *The Digest of Justinian*, T. Mommsen, P. Krueger and A. Watson (eds), Philadelphia, PA: University of Pennsylvania Press, 1985.(The Latin text by Mommsen and Krueger with a readable translation by various writers, edited by A. Watson.)

Kelley, D. (1970) *Foundations of Modern Historical Scholarship: Language, Law, and History in the French Renaissance*, New York: Columbia University Press.(Classic study of Renaissance legal humanism and its role in the development of social theory.)

Levinson, S. and Mailloux, S. (eds) (1988) *Intepreting Law and Literature: A Hermeneutic Reader*, Evanston, IL: Northwestern University Press.(Critical legal reader in literature and law.)

Leyh, G. (ed.) (1992) *Legal Hermeneutics: History, Theory, and Practice*, Los Angeles and Berkeley, CA: University of California Press.(Collection of contemporary essays on legal hermeneutics from a political perspective.)

Maclean, I. (1993) *Interpretation and Meaning in the Renaissance: The Case of Law*, Cambridge: Cambridge University Press.(Excellent expansion of material in §2 of this entry.)

Schlag, P. (1996) *Laying Down the Law: Mysticism, Fetishism, and the American Legal Mind*, New York: New York University Press.(Critical Legal Studies account of the paucity of common law hermeneutics.)

Stein, P. (1966) *Regulae Iuris: From Juristic Rules To Legal Maxims*, Edinburgh: Edinburgh University Press. (Historical account of the Roman origins and subsequent developments of the concept of legal meaning.)

Legal idealism

The term 'legal idealism' has various meanings. These include: the notion that laws, and the rights and duties they confer, genuinely exist, in which legal idealism is opposed to legal realism; the notion that law is intimately connected with moral or social values, in which legal idealism is opposed to legal positivism; the notion that one can move from certain premises about human reason and will to systematic principles for legal development and decision-making; and the notion that one can derive systematic principles of a similar kind from the requirements of social life. It is also sometimes used to imply too much faith in the capacity of law to solve problems. The enduring issues of legal idealism concern establishing systematic principles for legal development and decision-making.

1 1 The varieties of legal idealism

Legal idealism is a particularly confusing term as it is used of different, and sometimes incompatible, approaches to law. First, and at its broadest, it covers all those theories which regard law as normative or standard-setting, as expressing what ought to be done. Such theories typically treat law as consisting of norms or rules or principles and as conferring rights and duties (see Legal concepts; Norms, legal). Legal idealism, here, is tied to a certain kind of philosophical idealism, since law is seen as the product of some act of mind or will. As such its norms genuinely form part of our world. For most thinkers, this is also the case with the rights and duties to which it gives rise. So understood, legal idealism embraces most, if not all, natural law thinking (see Natural law) and even certain types of legal positivism, such as that of Hans Kelsen. It stands in opposition, however, to other types of legal positivism, for which matters such as rules and duties are at best imaginary entities, with talk about them to be explained away in terms of attempts to predict or promote certain forms of human behaviour (see Legal realism §1).

Second, the term 'legal idealism' is used for theories that regard law as intimately connected with certain specified moral or social values, such as freedom, justice, democracy, cultural evolution or group solidarity. Here idealism is understood in a sense closer to that of ordinary language, where it is tied to the having of ideals. In this context, legal idealism once again covers natural law thinking and also certain other approaches that resemble it. Here legal idealism stands in opposition to legal positivism, where it is claimed that law is to be identified solely by reference to the form it takes or the source from which it comes, with no built-in restrictions as to content or purpose (see Legal positivism).

Third, there is what is known as transcendental legal idealism. A transcendental inquiry is concerned with investigating the conditions under which certain kinds of experience or knowledge are possible. This kind of approach, using the tradition of Kant (§5), was adopted within a group of theories of largely, although not exclusively, German origin. Such theories begin from certain premises about the nature of human reason and will and develop a legal philosophy from them. One of the objects of such a philosophy is to provide a 'scientific', largely in the sense of 'systematic', basis for the development of law and for legal decision-making. The main departure from Kant is in terms of the emphasis that is placed on the social purposes of law.

Fourth, there is what is sometimes called absolute legal idealism, following the work primarily of Hegel (§8). This starts from the premise that there is some absolute idea and argues that history works inexorably towards its fulfilment. Law, in both form and content, reflects stages in the development of this idea.

Fifth, there is historical/social legal idealism. This differs from transcendental and absolute legal idealism mainly by taking the actual nature of social life as a given and attempting to find within it the ideas or ideals to which law-making and legal decision-making should conform if they are to serve it properly.

Sixth, and finally, legal idealism is sometimes used pejoratively for those theories that are thought to express a naïve faith in the capacity of law to solve social problems.

Various points are worth noting about these different kinds of idealism. A theorist may be an idealist of the first kind without being one of the second kind. Kelsen, for example, subscribed to the objective existence of norms but actively sought a pure, that is value-free, theory of law. On the other hand, the third, fourth and fifth kinds of idealism are all value-laden and so are really just particular versions of the second kind. They are, however,

sometimes presented as if they did nothing more than restate the factual requirements of social life. Finally, some thinkers are difficult to allocate, since their theories seem to incorporate elements from more than one of the particular versions.

2 2 Transcendental and absolute legal idealism

The starting point is Kant's distinction between theoretical or speculative and practical uses of human reason. Both are concerned with formulating principles or laws, theoretical reason with what is and practical reason with what ought to be. Both can be pure, in the sense that the principles involved can be arrived at, not by looking to experience, but by looking for the conditions that make such experience intelligible. Pure practical reason is directed to discovering the conditions for moral action. These are found in the laws that humans should set themselves if they are to be free, that is to determine their own wills on a fully rational basis. The actions that ought to be performed under such laws are duties. When it comes to actual human behaviour, however, duties may be performed under two different sorts of motive. The first sort is internal, where the very existence of the duty is the motive for the action, in which case the law is an ethical one and the duty a matter of virtue. The second sort is external, deriving from law in the juridical or legal sense, where conformity may be brought about by other types of compulsion and the duty is a matter of right. For Kant, however, juridical law, in the form of positive or statutory right proceeding from the will of a legislator, should be confined to enforcing perfect duties. Such duties are primarily concerned with the avoidance of wrongdoing to others, in the form of interfering with their freedom. Justice is thus seen as the sum of the conditions under which individual wills can be conjoined in accordance with a universal law of freedom.

Fichte (§4), though a disciple of Kant, departed from certain of his ideas. One such idea was that there are 'things-in-themselves' which are unknowable, our knowledge being limited to the structure we impose on what appears to us in the course of our experiences. Instead Fichte sought to construct experience from the activity of the conscious self. By discarding the possibility of any reality beyond mind and the objects of experience, he paved the way for the development of Hegelian absolute idealism. His legal philosophy, too, started from the way in which self-conscious beings 'think themselves' and, in parallel with Kant, Fichte saw this as involving them ascribing freedom both to themselves and to others. In so far as the individual respects the freedom of others as if it were absolute, the individual acts ethically and achieves spiritual freedom. Law, however, only requires individuals to limit their freedom to the extent necessary to secure the equal freedom of others. It provides the compulsion by which the duties so generated can be enforced by others as rights. Fichte went further than Kant, however, in terms of the types of rights that should be afforded, extending these into the economic sphere.

Hegel's legal philosophy concentrated on the way in which the free and rational will embodies or objectifies itself in the world in the form of institutions, so proceeding towards a reconciliation with the universal will. It does so at the level of abstract right, involving the rights and duties of human beings as such and covering property, contract and wrong (both civil and criminal); morality, which is the rational will directed towards its own development; and the ethical life, which is a product of the dialectic between abstract right and morality. The ethical life itself begins with the family. To this is counterposed civil society, involving law and the administration of justice. Both are brought together in the state, in which individuals realize their true universal wills, finding, in their duty, their substantive freedom. Much of Hegel's importance for legal theory, however, lies in his philosophy of history, according to which there is a continuing progression towards absolute freedom, stages in this progression being represented by particular states at particular times.

These thinkers, for whom legal philosophy was simply an aspect of their more general philosophy, had considerable influence on the ideas of those concerned more directly with law. For example, Stammler developed a theory of justice obviously indebted to the ideas of Kant. He distinguished between the notion of law as a particular kind of human volition, as an inviolable, sovereign, combining will, and the idea of law, which is concerned with achieving the perfect harmony of individual purposes, with attaining the social ideal of a community whose members will this freely as an end. From this Stammler derived his four principles of just law: two of respect, limiting the manner and extent to which members of the community can impose their own will on others; and two of participation, limiting the manner and extent to which any member may be excluded from the benefits of community life. The first principle in each set is to do with the avoidance of arbitrariness; the second is to do with treating others in a neighbourly way. In considering how these principles might be applied in practice,

he constructed a model for just law, involving a special community whose actual regulations and the conflicting claims under them are ‘laid before the judgment of the principles of just law’. Natural law thinking, he argued, erred not so much in its method as in claiming universal validity for the rules produced by that method, since this ignores the fact that everything to do with human wants and the way they are satisfied relies on experience and so is subject to constant change.

Del Vecchio also drew a distinction between the notion or concept of law and the idea of law, but developed it in a different manner. In effect he equated the concepts of law and justice and treated them, like Fichte, as derived from the need for reciprocity. He regarded the parallel ideas, however, as more than merely an expression of the preconditions for fulfilling some social end. In his view, the idea of justice has a content which springs from the human conscience. Law is to be seen, in the manner of Hegel, as the progressive realization of justice, the recognition of the absolute value of the personality and the equal freedom of all. Del Vecchio’s combination of Kantianism and Hegelianism was not particularly new. It is, for example, to be found in the writings of the Scottish jurist W. Galbraith Miller in 1884. Del Vecchio’s later works showed a greater move both towards the ideas of Hegel and away from Kantian individualism into corporatism and nationalism. A similar change is to be found in the ideas of Binder, both he and Del Vecchio becoming apologists for political developments in Germany and Italy respectively.

3 3 Historical/social legal idealism

One theorist who attempted to address both the actual nature of social life and the way in which it changed through history was Kohler. In common with Jhering he saw law as a means to a social end but differed from him by rejecting social utility as a criterion. Like Hegel, he saw history as the unfolding of an idea but argued against him that its progress was neither necessarily visible nor logical. Kohler identified the idea as that of civilization, by which he meant the increasing capacity of humans to control both external, and their own, nature. Law has the dual function of serving the needs thrown up by the current state of civilization within a society and of fostering its development. In consequence, what he calls ‘jural postulates’, allowing the law to cope both with differing rates of change and its own fixity, must be developed (see Pound, R.). Kohler arguably had absolutist tendencies to the extent that he identified a single ideal, civilization, and treated law as something that should foster it. (For a different view of the relation between law and cultural values, see Radbruch, G.)

Other thinkers concentrated directly on social life. For example, Duguit saw social solidarity as both a fact and a requirement of social life and law as both an aspect and a requirement of social solidarity. Gény was critical of Stammler for failing to show how his principles of just law might be applied to concrete problems. In a type of natural law thinking, he turned to identifying what will affect any body of positive law - namely factors of a physical or psychological and of a historical kind, general considerations of rationality, and the values embraced by the civilization concerned. The lawyer must take account of what is given in this way, applying his technique to fashion law according to the needs of social life. Interpretation is thus seen as a creative activity and the judge should be informed by *libre recherche scientifique* (free scientific research).

A parallel development in Germany was that of *Interessenjurisprudenz*, which, in the writings, for example, of Heck, argued for the application of a proper sociological method to legal problems. Following Jhering, this was held to involve the identification, classification and balancing of interests, of both a material and an intellectual kind. A much more radical approach, sponsored by theorists such as Ehrlich and Kantorowicz, was that of the *Freierechtslehre* or ‘free law’ school, which advocated the maximum possible recourse to justice and equity in legal decision-making.

4 4 An overview

Legal idealism, in all its guises, runs up against the much broader philosophical question of the relationship between fact and value. Idealist theories are criticized, no matter which way they appear to move, for failing either to bridge the gap between the two or to show that there is actually no such gap to bridge. In the legal context, idealist theories follow different paths in trying to overcome these criticisms. While admitting that facts and values are of a different order, they may variously try to show either how legal norms can, none the less, form part of the world of fact or else how they are somehow required by or derivable from facts. Alternatively, they may effectively deny the distinction, arguing that legal norms, at least those comprising the natural law, exist as a

matter of moral fact. Among the most interesting more recent attempts to root norms in social fact is the 'institutional theory of law' of MacCormick and Weinberger (see Weinberger, O.; Institutionalism in law §4). Habermas' 'theory of communicative action', offering a transcendental argument anchored in the intuitive knowledge of competent members of modern societies, provides a different approach to the same problem and one of equal relevance for legal theory.

See also: Justice, equity and law; Law, philosophy of

ELSPETH ATTWOOLL

References and further reading

Del Vecchio, G. (1905-08) *Il presupposti filosofici della nozione del diritto; Il concetto del diritto; Il concetto della natura e il principio del diritto*, 3 separate vols published together, trans. J. Lisle, preface J.H. Drake, *The Formal Bases of Law*, New York: Macmillan, 1921.(For further of his ideas on justice, in English, see his *Justice,* ed. A.H. Campbell, Edinburgh: Edinburgh University Press.)

Duguit, L. (1913) *Les transformations de droit public*, Paris: A. Colin; trans. F. Laski and H.J. Laski, with intro. by H.J. Laski, *Law in the Modern State*, London: Allen & Unwin, 1921.(The introduction gives a helpful summary of his ideas.)

Ehrlich, E. (1903) *Freierechtsfindung und Freierechtswissenschaft (Free Law Finding and the Science of Free Law)*, Leipzig: C.L. Hirschfield.(For an account in English, see Gény 1917. For his ideas more generally, in English, see *Fundamental Principles of the Sociology of Law*, trans. W.L. Moll, Cambridge, MA: Harvard University Press, 1936.)

Fichte, J.G. (1796-7) *Grundlage des Naturrechts nach Prinzipien der Wissenschaftslehre*, Jena and Leipzig: C.E. Gabler; trans. A.E. Kroeger, preface W.T. Harris, *The Science of Rights*, London: Trubner & Co., 1889.(Of particular importance for those interested in the origins of the idea of the *Rechtsstaat*.)

Friedmann, W. (1967) *Legal Theory*, 5th edn, London: Stevens & Sons.(Especially chapters 15, 16 and 27. These cover most of the theorists mentioned in the entry.)

Gény, F. (1917) 'Judicial Freedom of Decision, its Necessity and Method', in *Science of Legal Method*, trans. E. Bruncken and L.B. Register, intro. by H.N. Sheldon and J.W. Salmond, Boston: Boston Book Company.(Also containing essays by Ehrlich, Kohler and Pound.)

Heck, P. von (1932) *Begriffsbildung und Interessenjursiprudenz (The Formation of Concepts and the Jurisprudence of Interests)*, Tübingen: Mohr.(For a brief account of his ideas in English, see Friedmann 1967.)

Hegel, G.W.F. (1821) *Grundlinien der Philosophie des Rechts*, Berlin: Drucker & Humboldt; trans. T.M. Knox, *Philosophy of Right*, Oxford: Clarendon Press, 1942.(A closely argued text which can be difficult without some prior knowledge of Hegel's philosophy more generally.)

Kant, I. (1797) *Metaphysische Anfangsgründe der Rechtslehre*, Köningsberg: F. Nicoloumo; trans. W. Hastie, *Kant's Philosophy of Law*, Edinburgh: T. & T. Clark, 1887.(Although aspects of Kant's more general philosophy are incorporated in the text, the concentration is on its detailed implications for law.)

Kantorowicz, H. (1925) *Aus der Vorgeschichte der Freierechtslehre (On the Antecedents of the Free Law Doctrine)*, Mannheim: J. Bensheimer.(For a wider view of his ideas, in English, see *The Definition of Law*, ed. A.H. Campbell, Cambridge: Cambridge University Press, 1958.)

Kohler, J. (1909) *Lehrbuch der Rechtsphilosophie*, Berlin and Leipzig: W. Rothschild; trans. A. Albrecht, preface A. Kocourek, *Kohler's Philosophy of Law*, Boston: Boston Book Company, 1914.(The preface is by a significant figure in analytical jurisprudence.)

Stammler, R. (1902) *Die Lehre von dem richtigen Rechte*, Berlin: J. Guttentag; trans. I. Husik, appendices by J.C. Wu and F. Gény, *The Theory of Justice*, New York: Macmillan, 1925; repr. New York: A. M. Kelley, 1969. (Gény's appendix is highly critical of the theory.)

Stone, J. (1965) *Human Law and Human Justice*, London: Stevens & Sons. (Especially pages 167-92. Covering Stammler and Kohler and, also, Nelson, who has not been included in this entry.)

Legal positivism

Legal positivism is the approach in the philosophy of law which treats 'positive law' - law laid down in human societies through human decisions - as a distinct phenomenon, susceptible of analysis and description independently of morality, divine law or mere natural reality. It shares with philosophical positivism the aim of dealing in facts, but these are facts about legality and legal systems. Insistence on the distinctness of positive law has been integral to the 'rule of law ideal' because of the aim of clear law applied by neutral legal officials. However, debates about positivism have been marred by a degree of conceptual confusion: positivism often appears to mean something different to its supporters and to its enemies, and many attacks are launched against straw men. Consequently, much depends on the definition of legal positivism that is used.

Attempts have been made to put some order into the discussion. Consider, for instance, H.L.A. Hart's list of meanings of legal positivism (which cumulatively count as features of positivism): (1) law as human commands; (2) absence of any necessary connection between law and morals; (3) the study of law as meaning, as distinct from sociology, history and evaluation; (4) the contention that a legal system is a closed system, sufficient in itself to justify legal decisions; (5) non-cognitivism in ethics (Hart 1958). Norberto Bobbio's list is shorter and more orderly, but at first sight not too different (Bobbio 1960): legal positivism has been conceived as: (1) a neutral, scientific approach to law; (2) a set of theories depicting the law as the product of the modern state, claiming that the law is a set of positive rules of human origin, and ultimately amounting to a set of statutes, collected in legal systems or orders; (3) an ideology of law that gives a value to positive law as such, implying that it should always be obeyed. However, in this list, unlike Hart's, the 'meanings' cannot be added together, the first and last being incompatible. The connection between the three points is as follows: for positivists the theories of Bobbio's second point (law is made up of rules produced by the state) yield a scientific and value-free approach to law; for the adversaries of legal positivism they yield only ideology, that is hidden value judgments in favour of the power of the State.

The shortest way to understand what is at issue in these abstract discussions is to proceed by contrasting legal positivism with its main critics' approach to law. It is noteworthy that on this point legal realists and natural law theorists, although starting from different and even opposite points of view, agree in concluding that legal positivism is an ideological, covertly evaluative, thesis.

1 Positivism versus natural law theory

One of the basic ideas of legal positivism is the separation of law and morals: according to positivism defining what the law is should be kept separate from deciding whether it is just or unjust, good or evil. This is the basic conflict with the iusnaturalists (natural law theorists) who argue that describing the law necessarily requires moral value judgments and choices (see Natural law §§1, 4). Yet, positivism's other main opponents, legal realists, who accept the positivistic aim of a value-free description of law, none the less argue that the positivistic 'formalist' approach is itself contaminated by hidden unrecognized value choices, and therefore that legal naturalism and positivism are really the same. Natural law theorists, taking positivists at their word with a vengeance, often group all their adversaries, including realists, together under the common label of positivism (meaning a value-free approach to describing the law), the lack of what they believe to be a proper understanding of the vital role played by morality in the life of law (Weinreb 1978; Fuller 1958).

Therefore, a little time spent on conceptual clarification will not be wasted. Glanville Williams proffers the most radical answer to this kind of conceptual question about competing definitions: 'Everyone is entitled for his own part to use words in any meaning he pleases: there is no such thing as an intrinsically "proper" or "improper" meaning of a word' (Williams [1945] 1956: 134-6). However, while Williams' thesis is a good antidote to philosophical verbosity, it is a little too simplistic. Not all conceptual discussions can be solved merely by arbitrary definition. In fact, concepts are fragments of theories; when we examine the concept of law in legal positivism, we are looking not just for the meaning of a word but also for a theory which will offer the best approach to legal problems and eventually to the final practical problem of obedience.

So, what is at issue in positing that the 'is' of law can be identified separately from any 'ought'? It means that nearly all positivists replace the moral commitment to justice (however this is defined), which is at the basis of the

iusnaturalistic account of the law, with the aim of a value-free description of effective legal systems (however this is gauged). In other words, effectiveness thus replaces justice in the positivistic conception of law (Hart 1961; Kelsen 1945). Critics of positivism may be forgiven for thinking this to be a rather perverse attitude, making it all too easy for positivistic lawyers to neglect their (moral) duty to take a moral stance towards law. This is sometimes called the '*reductio ad Hitlerum*' of positivism, which is accused of preaching acquiescence to every (effective) law, even when it is as morally evil as Nazi law. This charge is, in turn, firmly rejected by positivists, who observe that something can only be rationally judged and valued if we first know what it is. This is precisely Bentham's basic distinction (see Bentham, J.), which is at the origins of legal positivism, between censorial and expository jurisprudence (Bentham 1977). Many positivists will also add that we gain nothing but confusion in insisting, as natural law theorists do, that unjust laws are not really laws (*lex iniusta non est lex*).

However, natural law theory is saying more than that; otherwise its more dismissive critics would be right in saying that naturalists are merely engaged in a vain battle of words. Iusnaturalists maintain that the law cannot be individuated, described, interpreted or applied unless we continuously make use of moral values and moral choices (Finnis 1973): a normative element is part of the basic choices in the very description of the law. Hence, some naturalists would argue that this should also be a defensible ethical choice, rather than the morally inadequate positivistic standard of effectiveness. In any case, the positivistic attitude, so its critics would argue, leads not to empirical science but, rather, to the moral choice of conformism, an attitude to which lawyers are all too easily prone and which should not be encouraged further.

The positivist might reply that this kind of conformism is indeed an essential part of the rule of law, the ideal embodied in many contemporary legal and political institutions, that power be conferred, controlled and exercised through general rules only (Scarpelli 1965; Fuller 1958). This ideal is a component of all the basic institutions of contemporary Western states, where a legislative body is supposed to make general laws (statutes), while different bodies (the judiciary, the administration) apply them, within the limits of powers granted to them by other legal rules (rules of competence, procedural rules) (see Fuller, L.L.). It is obvious that this ideal is based on, at least, judicial compliance to law and, more widely, obedience of the subjects who fall under law or seek to use it. It is equally obvious that only in so far as it is possible to have an objective knowledge of the meaning of such laws does the whole enterprise make sense; only then will law's reality match the appearances. Legal positivism purports to provide a model of such an objective knowledge, of a legal science conceived as the study of whatever rules come from the appropriate sources and authorities. The next question then must be, what exactly is a (legal) rule?

2 Varieties of legal positivism

All positivists think that the law is a set of legal rules, or norms, a belief that is strictly connected to the fundamental value of the rule of law (see §1). Law is conceived as a device for transferring authority and (political) legitimacy by means of rules; legal rules connect individual choices, such as judicial sentencing, to general decisions such as legislation. But, what is a rule? The main versions of legal positivism can be distinguished by their different accounts of what a legal rule is.

The most immediate and common answer is that a (legal) rule is a command. This thesis is typical of early positivism, as represented by the theories of John Austin (Austin 1832). The second answer is that (legal) rules are judgments as in Kelsen's mature legal positivism (see Kelsen, H.). According to Kelsen, legal rules or norms are conditional, hypothetical judgments of the form: if you committed such and such action, then you (legally) ought to be punished (Kelsen 1945). This idea also avoids the main objection to the older ('naïve') positivism of Austin's kind: that law cannot be reduced entirely to commands, which exist only as part of a direct relation between real persons (see Austin, J.).

The Kelsenian version of legal positivism sees a legal rule as a conditional ought-judgment. However, this renders problematic positivism's claims to legal empiricism. The 'existence' of the sentence/proposition 'if you kill you ought to be punished' is totally unaffected both by the fact that a killing has happened and by the fact that the killer is actually punished. This loose connection with facts is a frailty of Kelsen's brand of positivism. The norms or rules in the Kelsenian sense need to be somehow anchored to social facts, so that the Kelsenian positivist may be enabled to distinguish imaginary or ideal law from real positive law. As we shall see below (§4), Kelsen and all modern positivists attempt to solve this problem with the theory of the existence of a legal system.

The later, ‘analytical’, versions of legal positivism, while preserving Kelsen’s idea that the law is of an impersonal and conditional character, try to translate the whole into more empirical terms. Law is seen as a social phenomenon, essentially as meaning and language. These versions of legal positivism claim that a neutral description of the law, which is vital to the ideal of the rule of law, can be attained by treating the law as a language; in fact this later positivism is usually influenced by some form of linguistic philosophy. Typically, Kelsen’s ‘pure theory of law’ is set on these new philosophical foundations; but the rest of the building is left substantially unchanged, and especially so his theory of a legal system. Legal rules and legal systems are now conceived mainly as (sets of) sentences and meanings, which can be neutrally described without having to be accepted or followed. Their empirical existence is thought to be guaranteed by their being an actually spoken language, something provided, therefore, with a social dimension.

However, there is a significant divide between the Continental and the Anglo-Saxon versions of this linguistic positivism. On the Continent, Bobbio (Bobbio 1950) claimed that jurisprudence should follow the ‘modern philosophy of science’ (meaning the neopositivistic version) and translate the language of the law, especially the language of statutes and codes, into a rigorous and coherent and complete language (see Bobbio, N.). By contrast, some years later in England, H.L.A. Hart used the analytical tools of ordinary language philosophy which roots language ‘rules’ in social use. He claimed that legal rules are essentially an abstraction from effective social practices, characterized by a fundamental linguistic element and certain normative uses of language (the ‘internal discourses’) (Hart 1961).

Bobbio’s theory and Continental positivism was also perhaps influenced by a tradition of statutory law (law as command from above). Its major difficulty was, accordingly, to distinguish between positive law elaborated and made rigorous by the efforts of linguistic jurisprudence, and rigorous but non-positive law (which would have meant falling into the arms of legal naturalism). For this task, subsequent Continental positivism has relied heavily on Kelsen’s theory of the basic norm, which is conceived as the linchpin connecting the whole of a legal system, a huge body of words, to social effectiveness and reality. Hart’s positivism is rooted in reality through its concept of rule as social practice, and its version of the basic norm theory has a much less daunting job to do: simply to recognize when rules are part of a legal system, rather than being the sole warranty of their existence.

3 Positivism versus legal realism

Discussion about the nature of legal ‘reality’, the role of language in legal analysis, and the issues of value-freedom are illuminated by considering the legal realist perspective (see Legal realism). This discussion shows another essential aspect of legal positivism. As the name suggests, realists strive to attain the underlying reality of law which, they say, is hidden and distorted by other approaches, such as positivism or naturalism. The European legal realists share with analytical legal positivists the idea that law is language (see §2). Both also accept that, if it is true that legal rules are language, it is also a language that is often ‘misspoken’, so far as official legal rules (especially statutory law) are concerned. However, they draw different conclusions from this fact based on different ideas as to how a language should be described.

The situation here is similar to that of a natural language where the official grammar and lexicon are significantly detached from the living spoken tongue. Positivists could then be compared to the members of a purist academy who, while well aware of how the language is being used, nevertheless maintain that this is not how the language is, but only the way it is (mis)spoken, and that the task of the linguists - by analogy, the positivist legal philosopher - is to tell how a language is correctly spoken and to describe the rules which make up such a language. By contrast, the legal realists are like those linguists who maintain that the only reality of a language is the actual practice of speech; that linguistic rules are, at most, a useful device for summing up these practices and instances of actual use in so far as they are inferred or induced from actual practice. As for the British positivists, while H.L.A. Hart may, at first sight, seem immune from this criticism, most realists are unhappy with his notion that the single legal rule need not be socially effective, provided it belongs to a legal system (see below) which on the whole is effective (see Hart, H.L.A.). Hart’s critics say that this means taking into account lawyers’ image of the law as an ordered whole, rather than the often disordered reality of the law itself. They would also add that all forms of normative grammar are prescription rather than empirical description. Hence the charge that contemporary legal positivists are engaged in surreptitiously upholding official state law.

In the empiricist, language-conscious framework in which this debate is taking place today, the realist attitude looks more attractive. Before reaching this conclusion, however, a key question needs to be considered. The discussion between realists and positivists can be whittled down to the following: how far is knowing what the law is, different from knowing what the law is for its ‘speakers’? To make sense of the ideal of the rule of law, how far should we be able to distinguish what the law says from what (some) people believe it says? In other words, can we say that the intellectual practices of lawyers are a set of cognitive activities with an objective content rather than superstitious beliefs with a hidden social function?

The European realists thus also take effectiveness as the criterion of law but define that in terms of actual use. They maintain that a value-free and scientific approach to law should take into account all and only the effective legal rules, be they in the official book or not, as extracted by means of the punctual observation of legal practice. While no positivist will deny that such an enterprise is legitimate, albeit very ambitious, they will, however, call it the sociology of law. Moreover positivists would insist that jurists, legal officials, as such, need a different kind of knowledge of legal rules, which is what Kelsen called ‘normative science’ and Hart an approach ‘from the internal point of view’. They will need to know what certain rules prescribe not in order to predict legal behaviour, but in order to know how they should behave, in case they accept them. In fact, it could be argued that neither the realistic nor the positivistic accounts of law offer a genuinely sociological empirical explanation, even though both contain some essential reference to social reality. The realist will describe all and only the legal rules which are effective with judges, the positivist will take into account only those legal systems which are effective as a whole, disregarding the fact that some rules exist only ‘on paper’.

4 Legal systems

The debate between positivism and realism brings to the forefront a key concept of legal positivism, that of a legal system. To positivistic jurisprudence, legal systems are both semantic entities and social realities. Legal rules are first of all words and meanings, prescriptions which can be described and understood, interpreted and used to give a normative legal meaning to facts and actions. Legal rules, being language, are also part of the world of social facts. Legal matters are to be decided by means of arguments or reasons taken mainly from within a legal system; these arguments or reasons are, of course, the legal rules. A legal system contains not only the rules which directly govern actions (called ‘primary rules’ by Hart), but also the ‘secondary’ rules, rules about rules, by which the former are ‘serviced’ (recognized, made, changed, applied). Most positivists assume that every legal system implies one supreme rule, containing or summing up the criteria by which all the other rules of the system are recognized as part of the system (basic norm in Kelsen’s theory, rule of recognition in Hart’s) (see §3).

The notion of a legal system explains the ‘formalist’ aspects of legal practice, the fact that law is conceived by positivistic jurists as being both ‘closed’ and ‘self-sufficient’ or, in a sense, ‘complete’. This does not mean that all legal issues must always be argued without referring to non-legal arguments, that is, moral and sociological facts or reasons. It means, rather, that only those reasons recognized by legal systems are compelling and that they should prevail over all the others, which can be used only to fill the gaps in the system (they amount to using discretionary powers).

According to this point of view, a legal system, with its enormous complexity, is not qualitatively different from a single rule. When we accept a single rule, we can use it as a reason for justifying our practical choices. However, this only postpones the problem, as now we have to justify the acceptance of the rule itself. The same happens with a legal system, but on a much wider scale, according to the positivistic conception (Raz 1975). By accepting a legal system we can put off our basic problem for quite a while, but at the end, we shall have to justify our acceptance or rejection of any particular legal system as a guide to our actions. This is sometimes neglected by people engaged in legal practice, who become, so to speak, moral hostages of the internal complexity of a legal system, thus embracing positivism in the third of Bobbio’s senses mentioned above. On the contrary, as we have seen, positivistic philosophers can well claim to be innocent of this particular sin, because of the thesis of the separation of law and morals.

In fact, having substituted effectiveness for justice as a criterion for ascertaining which normative system is law in a certain society, legal positivists will argue that such a choice is made for descriptive purposes only, and is wholly compatible with rejecting such a legal system on moral grounds. Critics of positivism, to be credible, thus have to delve deeper. To be effective, an attack has to be directed against the positivistic model of legal description; thus

some critics argue that a moral bias is already present whenever we try to distinguish a legal system from other kinds of moral and social rules and, equally, when we assume that in any society there is only one such legal system. Others argue that legal systems do not even exist as a social fact, being rather the product of a normative value-laden interpretation of social reality.

This is not to say that the positivist conception, if correct, is without advantages. The process consisting in the use of the mental tool of the concept of a normative system transforms, at least up to a point, the process of reaching a legal decision into a calculus based on known rules. This is precisely what is called ‘the rule of law’ by its supporters and ‘formalism’ by its critics (see Rule of law (rechtsstaat)). But here uncertainties and difficulties as to this representation of legal activity will crop up at many levels. First, we might doubt that the meaning of the words which make up such rules can be ascertained in any objective way at all. This is the classic problem of interpretation. Such objections usually come from non-positivists (see §5). Second, a host of problems is raised by the ‘dynamic’ or ‘formal’ nature of legal systems, and by the connections between legal rules. If a legal system has to be a calculus, then it is a strange kind of calculus which has to refer continuously to external facts - such as the fact that a statute has been enacted by the legislature or that a court has actually issued a decision - in order to complete its own calculations. This difficulty is well recognized by legal positivists themselves (Kelsen 1979). Positivism thus emphasizes that the systematicity of legal systems is different from purely deductive normative systems where the content of individual concrete rules can be deduced from first principles. On the contrary, a ‘dynamic’ legal system requires a continuous input from external facts and decisions; laws are made and abrogated by legislators, not deduced from first principles. Even judicial sentencing is, obviously, not the result of pure deduction from general rules; sentences have to be pronounced by a judge in the course of a trial, and if they happen to be incorrect deductions, so much the worse for logic and legal calculus. The dynamic or formal aspect of law, so it might be argued, is a matter of power and authority rather than of logic and calculus.

The point is not just that laws can be broken. This would not worry the legal positivist much. Laws are made precisely because they are capable of being violated. The point is that the law itself appears to be able to break its own rules, so to speak, as when an unconstitutional statute is considered valid in a legal system unless it is repealed through a special legal procedure, or a legally wrong sentence is legally binding unless it is reversed on appeal (Paulson 1980). That is why the positivistic autonomous ‘calculation’ of legal rules and decisions can conflict with facts and arguments which are not purely external, as they are considered relevant by the calculus itself (as, for instance, the wrong decision by an authority of the system). This is a powerful argument in the hands of the critics of positivism, who believe that the theory of the legal system and therefore the rule of law idea are a disguise for the exercise of naked power; while positivists try and rebut it with the theory of delegated power (Austin) and of discretion (Hart), claiming that discretionary powers are still controlled by the legal calculus, in so far as the external input is allowed and limited by the system itself (Austin 1832; Hart 1961).

5 Interpretation and rights

Uncertainty as to the real ‘closure’ of a legal system will crop up also at the level of words which make up such rules. This of course is the problem of interpretation. Some realists advance an entirely sceptical theory of interpretation (rule-scepticism) in which no certainty is possible. While many will dismiss such extreme arguments as bad semantics, even the most hardened of positivists will feel much less confident when it comes to certain operations within legal systems, or about certain words of legal discourse that express concepts which have a fundamental role in legal arguments, especially words designating (legal) rights.

Rights are easily identifiable as the Trojan Horse by which external material slips into an otherwise apparently closed legal system. Opponents of positivism will then say that this is because such material is not really external to the law to begin with, but is only seen as such by an overly narrow (legalistic, formalist) conception of law and the rule of law. The legal naturalists will claim that this clearly shows that certain moral ideas are a necessary part of the construction of a right (and hence that morality is necessary to the task of describing law). This argument draws formidable support from the Constitution of the USA as interpreted by their courts, which includes several moral principles and makes them justiciable (Hart 1977).

To legal positivists, wholly open-ended interpretation would make the rule of law an empty travesty, for it depends upon the end of certainty. They point to the difference between mere legality and strict legality. The rule of law requires legality in the second sense, which implies not only that legal decisions be taken according to a law, but

also that the words of the law be formulated by positivistic-minded legislators in such a way as to have a relatively determined meaning; and that interpreters must make the maximum effort in good faith to extract such a meaning before resorting to discretion (Ferrajoli 1990).

All this explains the apparently perverse positivistic attempts to 'sterilize' legal rights. If the rule of law in this strict sense has to survive, then rights must be reducible to the words of positive legal rules. Rights must derive from rules and not the other way around.

Most positivist accounts of the theory of legal rights, such as Kelsen's, stress that law must be entirely reduced to legal norms, and that therefore rights and legal institutions must also be so reduced. Ross' legal realist theory of a legal right (Ross 1951) follows the same reductionist line as Kelsen's (see Ross, A.). Legal rights amount to no more than a conceptual facility which merely makes it easier to sum up the underlying legal rules. The difference is that the positivist will say that the content of a right is a set of duties and powers attributed by legal rules, usually statutes, while a realist will say that rights and duties are empty words without any reference to facts, which at most (Ross) are a useful shorthand for describing and predicting the probable behaviour of the courts, and at worst (Olivecrona 1971) are powerful irrational social factors such as people wrongly believing in the existence of rights and being conditioned by such illusions (see Olivecrona, K.).

The opposite position claims that rights contain more than the words of the legal rules which express them; that they contain a normative surfeit, which can be extracted out of them without making new law. In this case rights could be a source of 'new' law which is not new but already there, although sometimes previously unstated and unsuspected. In the last twenty years Ronald Dworkin has moved a brilliant attack on the tenets of legal positivism in its Hartian form. Precisely in the name of legal rights, Dworkin has formulated in a renovated form the traditional theory of normative surfeit, that law cannot or should not be reduced to a set of explicit rules (Dworkin 1978).

The forward line of defence of positivism on this specific front is today held by Neil MacCormick (MacCormick and Weinberger 1986). As all good defences, MacCormick's stance is composed of a counterattack and a retrenchment. Against Dworkin, the basic tenet of positivism, the distinction between describing and evaluating the law, is reaffirmed (in the name of Bentham); the retrenchment is MacCormick's institutional theory of law, where a theory of right still based on rules is advanced, which attempts to make sense, through the concept of institution, of the aspects of legal rights not strictly reducible to already explicit statutory words (see Institutionalism in law §5).

See also: Law, philosophy of; Legal concepts; Legal reasoning and interpretation; Norms, legal

MARIO JORI

References and further reading

Austin, J. (1832) *The Province of Jurisprudence Determined*, London: Weidenfeld & Nicolson, 1955.(This classic text of legal positivism is from Austin's lectures: it is the basic statement of the theory of law as commands.)

Bentham, J. (1977) A Fragment on Government, in. J.H. Burns and H.L.A. Hart (eds) *A Comment on the Commentaries and a Fragment on Government*, London: Athlone Press.(This is another great classic, preceding Austin's work, though mainly unpublished till later. In many respects, a more subtle account than Austin's.)

Beyleveld, D. and Brownsword, R. (1989) 'Normative Positivism: The Mirage of the Middle-Way', *Oxford Journal of Legal Studies* 9: 463-512.(For advanced readers: this clearly depicts relations between positivism, realism and naturalism.)

Bobbio, N. (1950) 'Scienza del diritto e analisi del linguaggio', in U. Scarpelli (ed.), *Diritto e analisi del linguaggio (Legal Science and Analysis of Language)*, Milan: Edizioni Di Comunita, 1976.(Advanced reading on linguistic analysis and the idea of law as science - a principled and structured body of knowledge; this attempts to combine neopositivistic ideas and Kelsen's legal theory, through reflections on the analytical approach to the philosophy of law.)

(1960) *Giusnaturalismo e positivismo giuridico (Legal Positivism and Legal Naturalism)*, Milan: Edizioni Di Comunita.(A collection of essays: unfortunately still unavailable in English, but far clearer than anything written in English on the controversy between legal positivism and naturalism.)

(1961) *Il positivismo giuridico (Legal Positivism)*, Turin: Giappichelli.(Adds a historical survey to the subject

matter of the previous book. Unfortunately also unavailable in English, but by far the best introduction for university students to legal positivism and contemporary legal thinking.)

Dworkin, R. (1978) *Taking Rights Seriously*, London: Duckworth.(This is a collection of essays; intriguing and currently very fashionable in English-speaking legal philosophy; brilliant and difficult writing; advanced reading.)

Ferrajoli, L. (1990) *Diritto e ragione: teoria del garantismo penale (Law and Reason: A Theory of Guarantees in Criminal Law)*, Bari: Laterza.(The general part of this excellent book on individual rights in penal law makes a powerful contribution on the theme of legal certainty. Of average difficulty, but only available in Italian.)

Finnis, J. (1973) 'Revolutions and Continuity of Law', in A.W.B. Simpson (ed.), *Oxford Essays in Jurisprudence*, Oxford: Oxford University Press, 2nd series. (Extremely intelligent essay by a leading representative of naturalistic thought in post-Hartian philosophy of law: Hart's analytical lesson has been well learned and put to use to criticize some positivistic ideas.)

Fryer, B., Hunt, A., McBarnet, D. and Moorhouse, B. (eds) (1981) *Law, State and Society*, London: Croom Helm.(Like the following item, this collects ideas and debates of Anglo-Saxon legal philosophy after Hart; more critical of Hartian positivism than the following. Advanced reading.)

Fuller, L.L. (1958) 'Positivism and Fidelity to Law', *Harvard Law Review* 71: 630-67.(This debate with Hart is a widely known classic, widely reprinted, constantly quoted; Fuller is passionate but confused; his later *The Morality of Law* (revised edn 1969) is a clearer and more convincing statement of the position sketched in 1958.)

Gavison, R. (ed.) (1987) *Issues in Contemporary Legal Philosophy: The Influence of H.L.A. Hart*, Oxford: Clarendon Press.(Like the previous item, this collects ideas and debates of Anglo-Saxon legal philosophy after Hart; nearer to Hartian positivism than the preceding. Advanced reading.)

Jori, M. (ed.) (1992) *Legal Positivism*, Aldershot: Dartmouth Publishing Co.(A collection of essays, adding up to a more extensive formulation of the ideas condensed in this entry. Introductory reading.)

Hacker, P. and Raz, J. (eds) (1977) *Law, Morality and Society: Essays in Honour of H.L.A. Hart*, Oxford: Clarendon Press.(The essays in this collection are largely comments on Hart's legal philosophy. Most of the positivistic ideas subsequently debated in the Anglo-Saxon world can be found here.)

Harris, J.W. (1979) *Law and Legal Science: An Inquiry into the Concepts Legal Rule and Legal System*, Oxford: Clarendon Press.(This is a very clear presentation of and critical essay on positivistic theses. The author knows his Kelsen, and writes from that point of view. This might be used as an introduction, but perhaps better as more advanced reading.)

Harris, J.W. (1958) 'Positivism and the Separation of Law and Morals', in *Essays in Jurisprudence and Philosophy*, Oxford: Clarendon Press, 1983.(This is the manifesto of Hartian positivism. A well-argued classic, of medium difficulty; the other side of the debate with Fuller noted above.)

(1961) *The Concept of Law*, Oxford: Clarendon Press.(This is a classic and a must; conceived as an introduction to jurisprudence for Oxford students, and apparently plain common-sense reading, it is much more complex than it looks, and has been an outstandingly influential work.)

(1977) 'American Jurisprudence through English Eyes: The Nightmare and the Noble Dream', in *Essays in Jurisprudence and Philosophy*, Oxford: Clarendon Press, 1983.(The title here speaks for itself; a brilliant and witty essay, clearly positivistically biased. A must for US jurists dealing with Old England and vice versa.)

(1945) *General Theory of Law and State*, trans. A. Wedberg, Cambridge, MA: Harvard University Press. (A mature, and excellently translated, version of Kelsen's 'pure theory', with an adaptation by the author of his thoughts to address an Anglo-Saxon audience; but still very German.)

(1979) *Allgemeine Theorie der Normen* (*General Theory of Norms*), trans. M. Hartney, Oxford: Clarendon Press, 1991.(This is the final, posthumous, statement of the legal positivism of the most important legal thinker of the twentieth century; an excellent introduction by the translator relates the text to Kelsen's earlier work.)

MacCormick N. (1981) *H.L.A. Hart*, London: Edward Arnold.(The best commentary on Hart's thinking in the English language. Also useful for university students.)

MacCormick, N. and Weinberger, O. (1986) *An Institutional Theory of Law: New Approaches to Legal Positivism*, Dordrecht: Reidel.(An attempt to overcome some of the problems of positivism as a legal theory, not limited to discussion of the Anglo-Saxon theorists - hence particularly useful reading for Anglo-Saxon students.)

Olivecrona, K. (1971) *Law as Fact*, London: Stevens & Co., 2nd edn.(A good presentation of the European version of legal realism. Useful introductory reading.)

Paulson, S.L. (1980) 'Material and Formal Authorization in Kelsen's Pure Theory', *Cambridge Law Journal* 39: 172-93.(Advanced reading. It points out with brilliant clarity some weak points of Kelsen's conception of

‘normative theory’.)
Postema, G.J. (1987) ‘The Normativity of Law’, in R. Gavison (ed.), *Issues in Contemporary Legal Philosophy: The Influence of H.L.A. Hart*, Oxford: Clarendon Press.(A view of the complex relation between law and social reality.)
Raz, J. (1975) *Practical Reason and Norms*, London: Hutchinson.(Goes beyond legal theory to discuss and clarify some basic aspects of normative speech and thinking. Not as well known as it deserves. Intermediate reading.)
Ross, A. (1951) ‘Tû-tû’, *Harvard Law Review* 70: 1456-67.(Caustic and amusing, the basic idea, that a legal ‘world’ does not exist, should be considered seriously. Necessary reading for all legal students.)
Sampford, C. (1984) ‘Legal System and Legal Theory’, in D. Galligan (ed.), *Essays in Legal Theory*, Melbourne: Melbourne University Press.(Impassioned piece of post-Hartian legal theory, challenging the supposed systematicity of law. Interesting symptom of what is going on among Anglo-Saxon legal philosophers.)
Scarpelli, U. (1965) *Cos’il positivismo giuridico*, Milan: Edizioni Di Comunita.(Turning the traditional defence of legal positivism as a factual approach to law upside down, the author defends positivism as morally and politically indispensable to maintaining the rule of law. Advanced reading.)
Scarpelli, U. (ed.) (1976) *Diritto e analisi del linguaggio (Law and Linguistic Analysis)*, Milan: Edizioni Di Comunita.(This is a collection of essays belonging to the analytical school of legal philosophy, from all countries, dealing with legal theory and linguistic analysis.)
Stewart, I. (1987) ‘Closure and the Legal Norm: An Essay in Critique of Law’, *Modern Law Review* 50: 908-33. (A critical assessment of legal normativism. Advanced reading.)
Tur, R. and Twining, W. (eds) (1986) *Essays on Kelsen*, Oxford: Clarendon Press. (Excellent comments on various aspects of Kelsen’s legal theory by mainly Anglo-Saxon authors, but containing also contributions by R. Vernengo and O. Weinberger.)
Weinreb, L.L. (1978) ‘Law as Order’, *Harvard Law Review* 91: 909-59.(Balanced exposition of the natural law point of view on those points which have been particularly stressed by positivists.)
Williams, G. (1945) ‘The Controversy Concerning the Word "Law"’, repr. in P. Laslett (ed.), *Philosophy, Politics and Society*, 1st Series, Oxford: Blackwell, 1956.(The earliest application of linguistic-analytical ideas to legal philosophy. Strangely neglected today.)

Legal realism

'Legal realism' is the term commonly used to characterize various currents of twentieth-century legal thought which stand opposed to idealism. (Hence, 'realism' in this context ought to be understood not as a body of thought which opposes nominalism, but as an instance of nominalism.) In the Scandinavian countries, legal realism was modelled on Axel Hägerström's critique of idealist metaphysics, and sought ways to account for legal rights and duties without presupposing or postulating the existence of ideal objects or entities. In the USA, legal realism evolved as a critique of the idealism implicit in the vision of the common law which was promoted by C.C. Langdell, first Dean of the Harvard Law School, and in the laissez-faire *ideology of the late nineteenth- and early twentieth-century Supreme Court. Realist jurisprudential sentiments - primarily as articulated in terms of the so-called indeterminacy critique - continue to bear an influence on late twentieth-century critical legal thought.*

1 Scandinavian legal realism

Within twentieth-century legal thought, the term 'legal realism' is used very broadly to describe two branches of jurisprudence which stand opposed to legal idealism (see Legal idealism §1). These two branches of jurisprudence - Scandinavian and US legal realism - are not connected, nor do they conceive of idealism in the same fashion. In the Scandinavian countries, legal realism represents a reaction against the influence of German idealist metaphysics on jurisprudential thought. In the USA, realism constitutes a critique of idealism implicit in late nineteenth- and early twentieth-century common law teaching and doctrine.

The Scandinavian version of realist jurisprudence is both inspired and epitomized by Axel Hägerström's writings on the philosophy of law and morals (Hägerström (1953). Legal rights and duties, he contends, cannot be identified as factually existent and therefore cannot be considered 'real'. Despite their unreality, however, rights and duties clearly do exist in some sense. The question which Hägerström thus poses is that of how, if these entities do not exist empirically in the physical world, they can be known and their contents ascertained. His answer to this question is that rights and duties constitute personal feelings of entitlement and obligation (see Emotivism). There exist 'feelings of duty in regard to the restrictions which affect oneself', just as there exist 'feelings of power in regard to the acquisition of advantages… which are regarded as rights'. Thus it is that 'an individual who infringes the right of another lays himself open to reactions from the latter which gain a special strength through feelings of power of a special kind, which are connected with the idea of rights' (Hägerström 1953: 15).

Whereas Hägerström acknowledges that the concepts of right and duty possess psychological force (even if they cannot be empirically ascertained), his friend Vilhelm Lundstedt went so far as to argue that rights and duties are merely chimeras, discussion of which is inappropriate within a science of law (Lundstedt 1957). Indeed, Lundstedt would happily have seen analysis of rights and duties banished from jurisprudence. Other writers within the Scandinavian realist tradition none the less recognized that the concepts of right and duty possess considerable intuitive appeal. While these concepts may be essentially hollow, Alf Ross noted, the very fact that legal language is so dependent on the words 'right' and 'duty' suggests that legal statements about rights and duties - as opposed to rights and duties themselves - deserve serious philosophical attention (Ross 1958 §35). Karl Olivecrona offered a still stronger argument in favour of the study of rights and duties. '[T]he proposition that there is a law without rights and duties', he observed, 'seems to be contradictory since it belongs to the very essence of the law that it regulates people's rights and duties' (Olivecrona 1971: 184). Furthermore, even if the argument that rights and duties are chimeras was correct, 'the practical conclusion should not be to throw the terms overboard but to study their functions' (Olivecrona 1971: 178).

2 US legal realism

In the context of US jurisprudence, 'legal realism' is a remarkably ambiguous concept. At the beginning of the 1930s - responding to Roscoe Pound's cursory dismissal of 'the realists' (Pound named no one specifically) as a group of legal philosophers who exaggerated the jurisprudential significance of social science while underestimating the importance of legal rules, doctrines and principles (Pound 1931) - Karl Llewellyn endeavoured to put flesh on the bones of realist jurisprudence by identifying a 'sample' of realist legal philosophers (see Pound, R.). Since then, various excellent studies have been produced, focusing on, among other

things, the political (Purcell 1969) and social-scientific (Schlegel 1979) dimensions of realist legal thought, as well as on particular realist figures (Twining 1985).

US legal realism might be described very basically as a mood of discontentment with the idealism implicit in late nineteenth- and early twentieth-century US common law thought. This discontentment was directed towards two distinct juridical tendencies.

First, legal realism represented an expression of dissatisfaction with Langdellian idealism. The appointment of Christopher Columbus Langdell as Dean of the Harvard Law School in 1870 signalled the emergence of modern legal education in the USA (LaPiana 1994). Langdell's primary pedagogic initiative was to dispense with the traditional lecture method of teaching and put in its place the so-called 'case method' of instruction. The case method is a style of teaching which emphasizes the acquisition of knowledge through Socratic dialogue. The student subjected to this method is required to read judicial decisions in advance of class so as to be prepared to answer questions about their content and their reasoning. By teaching their subject in this fashion, Langdell and his followers were able to promote a highly abstract, idealistic - what they considered to be a scientific - image of the common law as a set of logically interconnected core doctrines and principles.

Legal realists tended to object to the scientific pretensions of Langdellianism. The idea that teaching law by the case method was tantamount to teaching law as a science was considered by various realists to constitute a form of erroneous idealism. In short, Langdellianism depicted law as mere logic rather than as experience, as something to be found in books rather than in society. Certain legal realists believed that law could be conceived as a science, but not in the Langdellian sense. For these realists, social science was the key to depicting reality.

Different legal realists took the concept of social science to denote different things. Predictivism and empiricism feature prominently as social-scientific themes in realist jurisprudence. Through 'fact research', certain realists believed, it may be possible to assess the quality of particular laws as instruments of social control and to predict the likelihood of those laws being infringed or ignored in the future (see Moore and Callahan 1943). It may well be that certain laws are difficult to enforce, or, when successfully enforced, entail an inordinately weak sanction. Someone may decide to break a particular law if they believe that their chances of being caught in breach of that law are minimal, or if the profit that they will acquire by breaking that law outweighs the fine which they might incur. Legal realists, in presenting this argument in social-scientific terms, were echoing the theory of Oliver Wendell Holmes of the 'bad man' (Holmes 1897). Other realists, in resorting to the social sciences, attempted to move realist legal thought beyond the shadow of Holmes. In this context, the work of Jerome Frank is exemplary.

Second, just as US legal realism may be understood as a reaction against Langdellian idealism, so too it may be understood as a reaction against the idealism embodied in the economic outlook of the late nineteenth- and early twentieth-century US Supreme Court. In a number of decisions - *Lochner* v. *New York* ([1905] 198 US 45) and *Coppage* v. *Kansas* ([1915] 236 US 1) are the most illustrious - the Court basically combined *laissez-faire* with social Darwinism. All citizens, the Court argued, possess an equal right to compete in the marketplace. This does not mean, however, that all citizens in fact possess an equal capacity to do so. Some will compete more, some more successfully than others, so that the strongest economic actors will prosper at the cost of the economically weak. In other words, an inevitable corollary of a formal equality of bargaining rights is a real inequality of bargaining power. Guided by this philosophy, the Supreme Court felt justified in constitutionally invalidating state legislation that attempted to redress inequalities of bargaining power. Freedom of contract was thus constitutionally enshrined.

For certain legal realists, the constitutional philosophy of the Supreme Court served simply to highlight the impossibility of real freedom of contract. Owing to the fact that there exists inequality of bargaining power among economic agents, they argued, markets must be essentially coercive rather than free. According to Robert Hale, one's capacity to make choices in the marketplace is determined by the extent of one's bargaining power (Hale 1923). Morris Cohen expressed the point pithily: 'The freedom to make a million dollars is not worth a cent to one who is out of work. Nor is the freedom to starve, or to work for wages less than the minimum of subsistence, one that any rational being can prize - whatever learned courts may say to the contrary' (Cohen 1933: 560). While the USA was suffering the effects of the Great Depression, the Supreme Court continued to invalidate economic regulatory statutes on the ground that they offended against the constitutional sanctity of free exchange.

This ‘economic’ strand of realist legal thought has sometimes been interpreted as an outright rejection of the possibility of economic freedom. But realists such as Hale and Cohen were not denying the possibility of economic freedom *per se*. They were arguing, rather, that such freedom cannot flourish in an unregulated market. Hale in particular believed that regulation may serve to facilitate rather than to inhibit market freedom. His argument was by no means as radical as has often been assumed (as, for example, by Kennedy 1993), who interprets the argument as a rejection of the possibility of market freedom).

3 The legacy of legal realism and the problem of indeterminacy

The general assumption that US legal realism was somehow radical - that it was the rebellious offspring of the modern US law school - ought to be questioned. Certainly many legal realists took issue with the Langdellian idealist belief that rules are generally certain and capable of mechanical application to legal disputes. What we do not find in the literature of US legal realism, however, is a celebration of legal indeterminacy (see Critical Legal Studies §2). One of the classic realist insights - that judges may sometimes ignore rules and principles and decide cases instead on the basis of personal intuition or ‘hunch’ (Hutcheson 1929) - was regarded by many realists as a cause for lament rather than celebration. Very simply, indeterminacy in adjudication was regarded as a problem.

It was a problem, furthermore, to which realism offered no persuasive answer. Some legal realists believed that remedying the problem of indeterminacy in adjudication demanded the development of a new kind of legal certainty: Holmesian, predictive certainty as opposed to Langdellian, deductive certainty. This line of argument finds its most complete expression in Llewellyn’s masterful study of the ‘steadying factors’ in the common law tradition (Llewellyn 1960). Still other realists, however, insisted that the belief that there exist determinate patterns of regularity in adjudication was no less ill-founded than Langdellian legal science itself (see Frank 1930). These realists - figures such as Jerome Frank, Wesley Sturges and Thurman Arnold - identified, but did not know what to do about, the existence of indeterminacy in adjudication. Thus it was that the problem of indeterminacy revealed the fundamental limitation of realist jurisprudence in the USA. Harold Berman recounts an illustrative story in this context concerning Thurman Arnold delivering a seminar at the Yale Law School in 1947. In this seminar, Arnold was apparently spinning the familiar realist line about how instinct rather than rule-following is the dominant feature of adjudication when he was interrupted and asked if that is how he had decided cases while a judge on the Columbia District Court of Appeals in the early 1940s. Berman takes up the story:

> Arnold paused before answering; one had the impression that he was transforming himself from Mr Hyde to Dr Jekyll as the professor in him yielded to the judge. He replied, ‘Well, we can sit here in the classroom and dissect the conduct of judges, but when you put on those black robes and you sit on a raised platform, and you are addressed as "Your Honor", you *have* to believe that you are acting according to some objective standard.’
>
> (Berman 1974: 30)

Arnold is certain that there is something more to the adjudicative process than what lawyers have traditionally identified. But he is unable to pinpoint what that ‘something more’ might be. Other realists, in confronting this problem, placed their faith in the notion of expertise: indeterminacy need not be feared, they argued, so long as legal officials are experts in what they are entrusted to do (see Frank 1949). Many legal philosophers who searched for safeguards to US democracy in the face of growing tyranny abroad were far from appeased by this argument (Duxbury 1992; Purcell 1969). Yet this is about as near as US legal realism came to offering an original theory of adjudication. One of the primary legacies of realism rests in the fact that it impressed upon US academic lawyers the importance of developing such a theory. It is to this task that US jurisprudence in the post-realist era has been overwhelmingly devoted.

See also: Law, philosophy of; Nominalism; Realism and antirealism; Social theory and law

NEIL DUXBURY

References and further reading

Berman, H.J. (1974) *The Interaction of Law and Religion*, London: SCM Press. (Study of the impact of religious ideas on legal thinking.)

Cohen, M.R. (1933) ‘The Basis of Contract’, *Harvard Law Review* 46: 553-92.(Classic realist critique of the notion of freedom of contract.)

Duxbury, N. (1992) ‘The Reinvention of American Legal Realism’, *Legal Studies* 12: 137-77.(Analysis of

post-Second World War criticisms and misinterpretations of realist legal thought.)
Duxbury, N. (1995) *Patterns of American Jurisprudence*, Oxford: Oxford University Press.(Study of US jurisprudence from the 1870s to the present.)
Frank, J. (1930) *Law and the Modern Mind*, Gloucester, MA: Peter Smith, repr. 1970. (Classic realist account of judicial decision-making.)
Frank, J. (1949) *Courts on Trial: Myth and Reality in American Justice*, Princeton, NJ: Princeton University Press.(On the uncertainties surrounding the jury system and judicial decision-making.)
Hägerström, A. (1953) *Inquiries Into the Nature of Law and Morals*, English trans. C.D. Broad, Stockholm: Almqvist and Wiksell.(Scandinavian realist critique of rights.)
Hale, R.L. (1923) 'Coercion and Distribution in a Supposedly Non-coercive State', *Political Science Quarterly* 36: 470-94.(Realist critique of *laissez-faire* ideology.)
Holmes, O.W. (1897) 'The Path of the Law', *Harvard Law Review* 10: 457-78.(The classic presentation of US legal realism as the jurisprudence of prediction.)
Hutcheson, J.C. (1929) 'The Judgment Intuitive: the Function of the "Hunch" in Judicial Decision', *Cornell Law Quarterly* 14: 274-88.(Confession by a realist judge that judicial decision-making is often nothing more than reliance on intuitions.)
Kalman, L. (1986) *Legal Realism at Yale: 1927-1960*, Chapel Hill, NC: University of North Carolina Press. (Institutional history of US legal realism at the law school with which it is most famously associated.)
Kennedy, D. (1993) *Sexy Dressing Etc.: Essays on the Power and Politics of Cultural Identity*, Cambridge, MA: Harvard University Press.(Collection of essays on critical legal studies, post-structuralism and critical race theory.)
LaPiana, W.P. (1994) *Logic and Experience: The Origin of Modern American Legal Education*, New York: Oxford University Press.(Study of the origin of modern US legal education at Harvard in the 1870s.)
Llewellyn, K.N. (1931) 'Some Realism about Realism - Responding to Dean Pound', *Harvard Law Review* 44: 1222-64.(Classic attempt to identify realist jurisprudence and its prime representatives.)
Llewellyn, K.N. (1960) *The Common Law Tradition: Deciding Appeals*, Boston, MA: Little, Brown.(Late-period US legal realist study of the adjudicative process.)
Lundstedt, V. (1957) *Legal Thinking Revised*, Stockholm: Almqvist and Wiksell. (Scandinavian realist critique of the concepts of right and duty.)
Moore, U. and Callahan, C. (1943) 'Law and Learning Theory: a Study in Legal Control', *Yale Law Journal* 53: 1-136.(Classic example of the social-scientific dimension of US legal realism.)
Olivecrona, K. (1971) *Law as Fact*, London: Stevens and Sons.(Study of major jurisprudential traditions by a Scandinavian jurist.)
Pound, R. (1931) 'The Call for a Realist Jurisprudence', *Harvard Law Review* 44: 697-711.(Critique of US legal realism as an anti-legalistic jurisprudence.)
Purcell, E.A. (1969) 'American Jurisprudence between the Wars: Legal Realism and the Crisis of Democratic Theory', *American Historical Review* 75: 424-46.(Study of how US legal realism came to be perceived in certain academic quarters as antithetical to democratic values.)
Ross, A. (1958) *On Law and Justice*, London: Stevens and Sons.(Scandinavian realist analysis of legal rights.)
Schlegel, J.H. (1979) 'American Legal Realism and Empirical Social Science: from the Yale Experience', *Buffalo Law Review* 28: 459-586.(Account of the empiricist dimension of US legal realism.)
Twining, W. (1985) *Karl Llewellyn and the Realist Movement*, London: Weidenfeld and Nicolson, revised edn. (Study of a major US legal realist.)

Legal reasoning and interpretation

Legal reasoning is the process of devising, reflecting on, or giving reasons for legal acts and decisions or justifications for speculative opinions about the meaning of law and its relevance to action. Many contemporary writers, such as Aulis Aarnio (1987), Robert Alexy (1988), Manuel Atienza (1991) and Aleksander Peczenik (1989), propound the view that legal reasoning is a particular instance of general practical reasoning. They suppose, that is to say, that reasoning can link up with action, guiding one what to do, or showing whether or not there are good reasons for a proposed course of action or for something already done. They suppose also that in law reason links up to legal decisions in this way. Both suppositions are well founded. Law regulates what to do and how to respond to what has been done, doing so within an institutional framework of legislatures, lawcourts, enforcement agencies and the like. It is a feature of legal institutions that they are expected to have, and usually do give, good reasons for what they do, and to do this in public. Legal reasoning is therefore not only a special case of practical reasoning, but a specially public one.

Rationality in action has at least two requirements: first, attention to facts, to the true state of affairs in relation to which one acts; second, attention to reasons for action relevant to the facts ascertained. The former aspect concerns reasoning about evidence; the latter, reasoning about rules or norms as reasons for action. In law, such rules and other norms have an institutional character. But how are these applied - by some kind of deductive reasoning, or nondeductively? Behind the rules of the law, there presumably lie other reasons, reasons for having these rules. What kind of reasons are these, developed through what modes of discourse? A discourse of principles, perhaps - but then how do reasons of principle themselves differ from rules? Reasoning from either rules or principles must always involve some process of interpretation, so how does interpretive reasoning enter into the practical reason of law? Answering such questions is the business of a theory of legal reasoning. Legal reasoning is to be understood as a form of practical reasoning concerning these very issues.

1 Legal reasoning and questions of fact

A general requirement of sound practical reasoning is that there be due consideration of the facts relevant to a decision. A decision must be appropriate or fitting to its context. Doubts about and evidence concerning questions or opinions of fact have to be weighed and conclusions reached about the facts at issue. In law, decisions have to be made in relation to procedurally formalized claims or accusations, usually concerning particular past events. Claims and accusations have to be proved by admissible evidence. But what constitutes proof? The narrative coherence of an account of events connected to reliable present testimony seems to be an essential test, but proof of this sort can never amount to absolute certainty. Hence all legal proceedings have to address some standard of proof: ‘proof beyond reasonable doubt’, or ‘proof on the balance of probabilities’, or that proof which is found ‘convincing’ by the trier of fact. The burden of proof has to be determined.

So far as proof demands a judgment of probabilities, it is a question of whether or how far probability should be interpreted according to a Pascalian or mathematical/statistical model, or should be interpreted as a form of ‘Baconian’ probability, as argued by Jonathan Cohen (1977). It is submitted that the overall probability of a case should be interpreted in Cohen’s terms, but with due allowance made for the legitimacy of statistical methods when forensic-scientific analysis can cast light on some key evidential fact or facts (see Legal evidence and inference).

Important though careful factual reasoning is in processes of decision-making, on its own it is insufficient to bring practical reasoning to any conclusion. Some transformation of factual reasoning into normative conclusion is called for. That is, there has to be some further reason why to act or decide in a certain way given that the situation is held to be such-and-such. Such reasons must be right-making, showing the act or decision as right for those facts.

2 Rules as reasons: deductively applied?

What kind of such reasons are there? One possible answer is ‘rules’. If there were a rule that act A ought to be performed in circumstances C, then, if it were established that C, the appropriate act in the circumstances would be A. It seems straightforward to turn this line of reasoning around and express it as a *modus ponens* hypothetical

argument. ‘If ever C, then A is right; C, so in this case A is right.’ This assimilates practical syllogisms to ordinary deductive arguments, save that the major premise is a normative rather than a factual generalization. But the view that practical reasoning can be straightforwardly assimilated to ordinary deductive reasoning is extremely controversial. And lawyers have frequently warned against the confusion of law with (deductive) logic.

There are difficulties. First, what is presupposed in appealing to a rule in such a setting? By virtue of what could a governing rule be said to exist? Two possibilities may be suggested. The first is the establishment of a rule by some form of authority, for example as a rule of a game, by the governing body of the sport in question; or as a rule of a school or university, by its governing body; or, finally, a rule of a legal system, established by the legislature. For a rule to count as a reason in this setting, it has to be presupposed that it is right to defer to that authority, and the problem of justification thus seems to lead out beyond any established rule or rules. We therefore need to look at the second possibility: that there is some reason to follow the rule other than the fact of its being an authoritatively established rule in some rule-system. What could such reason be?

3 Grounds for rules: appropriateness and rhetorical argument

At this point, there seem to be three possibilities: that the rule is endorsed as appropriate in itself; that it derives from a principle which is appropriate in itself; or that it is a rule justified by the value of its consequences, that is, by the consequences of following it as a rule - but in that case, some principle confirming the positive value of such consequences is presupposed, and the appropriateness of the principle is also presupposed. Returning to the question of authority, it can now also be said that the reason for accepting authority must evidently be the appropriateness of following the rules and/or principles that constitute the authority.

We are forced into considering the notion of that which is ‘appropriate’. One account could be in terms of some faculty of direct apprehension or intuition (see Intuitionism in ethics), whereby we are able immediately to recognize such appropriateness or other moral qualities. But if such an appeal to intuition is sound, why is there need for appeal to a rule at all? If we are capable of intuitively grasping the appropriateness of a rule in itself, or of the following of a rule established by an authority whose appropriateness is intuitively clear, can we not then simply apply that faculty of intuition directly to the case itself? May it not be that the appropriateness of the act to its circumstances is itself the justifying reason for the act? M.J. Detmold (1984) has argued that if one were incapable of seeing the force of such reasoning from the appropriateness of a particular act to its particular circumstances, one would be equally incapable of seeing the appropriateness of a rule, and hence incapable of arguing from rules. The foundation of legal (and moral) reasoning in this view lies in our capacity to judge what is right in individual cases, including a capacity to perceive when it is right to treat a particular case as a case covered by a rule, and decide accordingly.

Whatever may be the ultimate inevitability of some appeal to intuition, it does seem that recourse to this ‘black box’ ought to be deferred as long as possible in the analysis of any process of reasoning. Especially in practical reasoning, and most spectacularly in the case of legal reasoning, there are possibilities of argumentation, reasoning and discussion of a quite public kind over the issue of what is here called ‘appropriateness’.

Wherever there is a process of public argumentation, there is rhetoric. The modern rediscovery of rhetoric as a discipline owes much to reflection on legal reasoning. Theodor Viehweg, drawing on Aristotle, has pointed out the significance of *topoi*, or ‘commonplaces’, in rhetorical arguments. An argument for a particular rule or proposition can be supported by reference to some accepted *topos*, and arguments progress by working towards, or from, such commonplace positions. In law, there are maxims and long-standing principles and presumptions, such as ‘no one can give a better right than they have themselves’ or ‘a later law derogates from an earlier one’ and suchlike. Likewise, there are well-established argument forms such as *argumentum a fortiori*, *argumentum a maiori ad minus*, *argumentum per analogiam* and the like. An argument in such a recognized form starting from or working towards a recognized *topos* is well calculated to be persuasive in its given context.

In a not dissimilar way, Josef Esser (1970) drew attention to the importance of *Vorverständnisse*, ‘pre-understandings’, the taken-for-granted assumptions that enter any judgment of what is acceptable in the setting of legal argumentation - and in the preference of one method of arguing over another in a particular case. Once premises and mode of argument are settled, it is simple to produce an argument that satisfactorily justifies the conclusion reached. But the problem then becomes one about the reasonable choice of premises and method, so

there must be inquiry into pre-understandings. Aulis Aarnio (1987) has suggested that in the end these may simply have to be assessed as the ‘form of life’ that they constitute.

Chaim Perelman’s *La Nouvelle Rhétorique* (1958) emphasizes that arguments are necessarily addressed to an audience, and that persuasiveness is audience-relative. This is especially obvious in legal practice, where trained advocates put cases before courts as persuasively as possible, and judges decide after weighing their rival arguments on points of law. But persuasiveness is not the same as soundness. The issue for a theory of reasoning-as-justification is not what argument actually persuades a particular judge or jury, but what ought to convince any rational decision-maker. In this connection, Perelman postulates a ‘universal audience’ as providing the ultimate test; whatever argument would convince an audience of all intelligent and interested beings is a sound one. At the same time, Stephen Toulmin (1964) proposed a reinterpretation of traditional logic as regulating argumentative moves, and supplying warrants for moves from premises to conclusions, rather than as capturing timeless truths about rational thought.

The weakness of the rhetorical turn in analysis of practical reasoning is to reduce appropriateness to persuasiveness. Appeals to the universal audience or to any actual consensus seem unhelpful, since apparently we work out what would persuade the universal audience by reference to what is sound, not vice versa. Yet again, certain ‘critical’ approaches to legal thought urge that the claim to an objective soundness of legal reasons is the grandest rhetorical turn of all, enabling ideology or class interest to be misrepresented as some kind of ultimately sound reason or sound reasoning (see Critical Legal Studies; Legal discourse).

4 Procedural aspects: rational discourse

A procedural approach to practical reasoning may provide a solution to the problems posed by theorists of rhetoric and their critics. The question is what constraints upon reasoning it is rational to accept in order to come to acceptable conclusions. Here, the rhetoricians’ emphasis on the interpersonal context of argumentation is reiterated. In its light, the concept of universality has two uses: first, in stipulating universalizability of reasons - for the present instantiation of C to count as a reason now for doing A, it would have to be acceptable to hold A appropriate whenever an instance of C occurs; second, in testing the acceptability of any such reason - the universalized reason has to be acceptable to any interested party, and be challengeable by reference to anyone’s feelings, interests or opinions; but any opinion or feeling or interest is challengeable as to its relevancy, and so on. As Jürgen Habermas (§4) and followers like Robert Alexy (1988) argue, it may be possible to test practical propositions by reference, at least in principle, to the interests and views of the totality of persons in any way affected by or concerned with them. Habermas proposes a test by reference to dialogue in an ‘ideal speech situation’, envisaged as one in which all forms of coercion or interpersonal power or domination are put aside for the purposes of conducting (or imagining the conduct of) interpersonal discourse (see Communicative rationality §1). Analysis of the necessary constraints on such a discourse yields a procedural approach to testing the kinds of principles that rational discourse-partners could accept, acknowledging the types of desires and interests they actually have. Important in this is the idea that accepted principles or *topoi* should be subject to challenge, but are considered acceptable until successfully challenged, for example, on the ground that they cannot pass the test of universalizability or on the ground that they owe their origins to inequalities of power irrelevant as reasons in the ideal speech situation.

It is doubtful whether this procedural approach wholly disposes of recourse to intuition. It remains a question why some principles of argument must be accepted as ‘rational’ after all. But the merit of the approach is that it both postpones and narrows appeals to intuition, and scrutinizes the appropriateness of practical principles and rules in the light of acknowledged constraints of rational discourse. Commonplace principles are still needed as starting points, but they are challengeable within the argumentation. One requirement of rationality is that principles and rules be shown both to be mutually consistent and also to have overall coherence, working in priority rankings and weighing procedures to resolve *prima facie* conflicts.

This discourse-theoretical evolution of the rhetorical approach largely substitutes for without finally supplanting intuition in accounting for the idea of the appropriateness or rightness of a rule as a reason for action, and the appropriateness of other reasons in justifying either the content of a rule or the authoritative status ascribed to it. Insistence on universality in both the senses deployed here is also of value in blocking appeal to some particular ineffable intuition as a sufficient justifying reason for a particular decision in a particular case. The reason must at

least be statable as ‘whenever C, then A’, and some reason given for that, including a testing of it for consistency and coherence with other accepted rules, principles and unchallenged commonplaces.

5 Rules again: exclusionary reasons

Hitherto, it has been taken for granted that rules can be justifying reasons for actions. Now we are in a position to pose the question: ‘Why rules?’ Why should we have started from thinking of rules as reasons for decisions, rather than principles of rational decision-making such as might emerge from a procedurally articulated practical rationality? Rules are one type of normative entity. Their distinctiveness lies in what has been called their character as ‘exclusionary reasons’ (Raz (1975) or ‘entrenched generalisations’ (Schauer 1991). These terms draw attention to the special way one appeals to a rule in reasoning, namely as a factor that excludes consideration of otherwise relevant features of the situation. Appealing to a rule is appealing to an ‘exclusionary reason’.

Consider this: as a matter of principle, one ought to drive with due regard for the safety of others; it is a rule that one stop at red traffic lights. To respect the rule as a rule means abstaining from deliberation in particular cases whether due regard for the safety of others requires stopping at a given red light on a given occasion. The rule is a reason to stop and a reason not to reflect further on the wisdom of stopping (which does not mean that there may not be emergencies that cancel the rule). One can, by contrast, respect the principle as a principle while allowing that in a given context it can be outweighed by considerations arising from rival principles.

How then can rules belong in a theory of practical reason? How can reason mandate the exclusion of reasons? A common answer is that human coordination problems may give special utility to fixed rules for application by all parties whose actions have to be coordinated. In any interpersonal context in which pure principled reasoning guided by procedures for rational practical discourse would or could lead to conflict of soundly justified opinions (and actions) on points of conduct in shared social space, recourse to a rule is the best way to achieve the least unreasonable results on the whole, even though there are occasions when application of the rule yields non-optimal results. (Traffic-light rules exemplify this clearly, and indeed in Britain the Highway Code largely comprises rules for solving coordination problems in areas where reliance on pure principle would be obviously hazardous.)

Pure recourse to unaided practical reason without recourse to rules and authority would not yield satisfactory grounds for interpersonal decision where coordination is called for. It is therefore reasonable for humans to have or to establish authorities exercising powers of rule-making and of rule-applying. That is, it is reasonable to have law (in various forms), and reasonable that law comprise in large measure relatively detailed rules established through legislation (and possibly through precedent also). This is what establishes the special sphere of legal reasoning as a separate branch of general practical reasoning, with appropriate special principles, rules and forms of argument. Rules duly established through law are appropriate grounds for decision-making both by citizens in relation to their own conduct and by judges and other officials in decision-making in cases involving citizen-citizen or citizen-state conflict or dispute.

Earlier, it was suggested that reasoning where rules are applied is essentially deductive. To show that a rule is being applied, one states the rule ‘if C then A’, and states one’s finding that ‘C’ is satisfied in this case. ‘A’ can then be stated to follow as the appropriate action in the circumstances that obtain. In very straightforward cases, this indeed seems to be a complete account. But cases are frequently harder, in obvious ways. First comes the problem of possible conflicting interpretations of the rule. What does the rule properly mean? How should it be interpreted for this case? Second comes the problem known to lawyers as that of ‘qualification’ or ‘characterization’ of the facts and circumstances. For example, suppose there is an acknowledged rule not to commit adultery. If a married woman wishes to have a child, but cannot become pregnant by her husband, would it count as adultery if she participated in artificial insemination by donor (AID)? Would it make a difference if the donor were unknown or known? Or if her husband objected or consented? How should we qualify these facts having regard to the assumed rule? If a wife is pregnant through AID and if adultery is a ground for divorce, is her husband entitled to a divorce on the ground of her adultery? Whenever a rule is in view, it is crucial whether or not given facts can be legitimately so characterized as to fit the rule, or be distinguished from it. A similar question can be posed in a broader way, asking which of a range of possible interpretations of the general rule should prevail, given the wording of the rule, the general legal context of its enactment, and any values at stake in its implementation.

6 Legal reasoning beyond deduction

Even in cases where rules are *prima facie* applicable in a direct way, there can be problems of qualification and of interpretation, and a part of the rhetorical training of lawyers develops their attentiveness to such problems whenever helpful to the case they must advocate. Moreover, there may be many situations which, although not covered directly and unambiguously by an established rule, do nevertheless bear some closer or remoter analogy to the situation the rule expressly covers, and to which acceptable arguments of principle favour a resolution founded in this analogy. Decisions taken to deal with such problems or to uphold or reject the analogy may themselves constitute precedents offering grounds for future resolution of like problems. There is an obvious ground of justice for paying regard to such precedents in later cases, and the more the regard that is in fact paid on that ground, the stronger becomes the justice of holding to precedent, in view of expectations founded on the practice of following them.

This shows why the (essentially deductive) rule-application paradigm of argumentation, justifying a decision simply as an application of a rule to a particular case considered to be an uncomplicated instance of the generic case stipulated in the rule, is rarely if ever sufficient in itself fully to justify a decision. There have to be reasons for a favoured interpretation of the rule, reasons for a favoured qualification of a fact-situation, reasons in favour of pursuing or rejecting an analogy. Theories of legal reasoning have tried to elucidate its character in these areas that lie outside the simple deductive model.

Some of its features derive directly from constraints of practical reason itself. Thus reasons for interpretations or classifications have to be universalizable, as must be grounds for decision supported by analogy. There must be some explicit or implicit ‘holding’ or ‘ruling’ or ‘maxim’ of the case that accounts for the decision favoured, and that must be universalizable if not explicitly stated in universalistic terms: ‘given that adultery requires an actual act of sexual intercourse, pregnancy by AID is not to be considered proof of adultery’ would be an example. Any claim to justification through such a holding entails a commitment to its universal applicability - to every instance of AID complained of as adulterous *per se*, not only this one. This pushes the problem back a stage, but does not solve it, for the question at once arises how to justify the ruling or holding conceived in universalistic terms. There are but two possibilities. There must be deontological or teleological reasons in play (‘rightness reasons’ or ‘goal reasons’, according to R.S. Summers 1978). Either it must be shown to be in principle right to restrict ‘adultery’ in this way, or it must be shown to be preferable to the alternative upon consequentialist grounds.

Reasons of principle have a dual aspect here. On the one hand, since the context of decision is a legal one, and since the concepts at issue are legal, bedded in fundamental doctrines of family law, there is a necessary interpretive element: one has to establish a good account of the values that underpin the particular view of marital fidelity (or whatever, depending on the area of law in view) enshrined in the whole relevant body of law. On the other hand, there is the test of general practical reasonableness (here interpreted in the procedural model of practical rationality, though others are available). Teleological or consequentialist reasoning places the stress on the legal acceptability or unacceptability of the state of affairs represented by one or another of the rulings that come under view. Evaluation of the ‘acceptable’ or ‘unacceptable’ in the legal setting has regard to the legal-institutional framework, and to values that make sense in this framework. So the difference between deontological and teleological in this setting may be more one of presentation than one of fundamentals. In either event, an effort is made to interpret the broad sweep of the law in terms of its point or its underlying value or values, and on this account Ronald Dworkin 1986) has characterized ‘law’ itself as an essentially ‘interpretive concept’.

However that may be, there is another aspect of interpretation rather narrower and more exact in its reference. As noted, there are frequently issues of the interpretation properly to be put upon the terms and provisions of enacted law. While it is undoubtedly the case that this proceeds under the above-noted constraint of universalizability, nevertheless reasons for interpretations are also often strongly (and legitimately) relativized to the particular statute, its particular wording, its particular legal context, and (sometimes) particular purposes imputed to it either on objective grounds or having regard to evidence of the particular aims of the legislature that enacted it. Universalizably justifiable practices of interpretation require attention to the integrity of linguistic communication through the medium of enacted law, to the overall coherence of statute law, both internally to the particular statute and externally in testing its consistency and fit with the rest of the legal system in which it takes its place. A

similar principle in favour of the overall coherence of law is also in play where argument by analogy, in particular argument by analogy with settled precedents, is advanced in justification of decisions. Here too, the case for or against extending by analogy from the law as already settled to some novel formulation is subject to testing and critique both deontologically and teleologically. Interpretation of statutes and of precedents, and arguments by analogy in both contexts, have accordingly a particular part to play in legal reasoning.

Some commentators treat this as too restricted a view of the place of inductive or analogical reasoning. E.H. Levi's classic *An Introduction to Legal Reasoning* (1948) portrayed analogical development as the essential moving force of the common law, and its most characteristic style of reasoning. More recently, Steven J. Burton (1985) and Bernard S. Jackson (1988) have argued respectively that analogical reasoning has either coequal status with the deductive rule-based paradigm suggested above or (in Jackson's case) clear primacy over it, on the ground that all our reasoning has most fundamentally a narrative structure, such that the human capacity for pattern-matching among narratives accounts both for the human skill of analogical reasoning and the human capacity to make sense of rules. The present attempt to postpone recourse to and narrow the scope of appeals to intuition applies also to this essentially intuitionistic conception of pattern-matching; but it is not hostile to the claim that analogical reasoning is omnipresent in the adjustment of legal norms to social values. It is submitted that although argument by analogy has indeed a special role in the development of law and in processes of rational 'discovery' in developing new legal doctrine, nevertheless it has to be viewed as one part of a wider conception of legal reasoning as a special case of practical reasoning.

See also: Discovery, logic of; Dworkin, R.; Law, philosophy of; Legal discourse

NEIL MacCORMICK

References and further reading

Aarnio, A. (1987) *The Rational as Reasonable*, Dordrecht/Boston/London: Kluwer.(Gives an account of the relationship of rationality and reasonableness to forms of life.)

Alexy, R. (1988) *Theory of Legal Argumentation*, trans. R. Adler and N. MacCormick, Oxford: Clarendon Press. (The best statement of the case for the 'special case thesis', namely that legal reasoning is a special case of general practical reasoning; and the best English account of Habermas' theory of rational discourse as a guide to legal discourse.)

Atienza, M. (1991) *Las Razones del Derecho (The Law's Reasons)*, Cuadernos y Debates, 31, Madrid: Centro de Estudios Constitucionales.(An extremely clear and helpful discussion of most of the approaches discussed in the present article.)

Burton, S.J. (1985) *An Introduction to Law and Legal Reasoning*, Boston/Toronto: Little, Brown.(A good differentiation of the deductive and the narrative styles of legal reasoning; also a good, reader-friendly general account of the issues covered in this article.)

Cohen, L.J. (1977) *The Probable and the Provable*, Oxford: Clarendon Press. (Argues for a non-mathematical interpretation of probability in legal proof, and advances a theory of 'Baconian probability'.)

Detmold, M.J. (1984) *The Unity of Law and Morality*, London: Routledge. (Contains the best version of an essentially intuitionist account of what makes a decision right.)

Dworkin, R. (1986) *Law's Empire*, London: Fontana.(A discussion of law as an interpretive concept, and of the forms of reasoning involved in legal interpretation.)

Eggleston, R. (1983) *Evidence, Proof, and Probability*, 2nd edn, London: Weidenfeld & Nicolson.(A powerful defence of the applicability of Pascalian - or mathematical - probability in legal proof.)

Esser, J. (1970) *Vorverständnis und Methodenwahl in der Rechtsfindung (Pre-understanding and Choice of Method in the Search for Law)*, Frankfurt: Suhrkamp. (Gives an interesting account of the 'pre-understandings' always involved in acts of legal interpretation, and discusses the place of the 'acceptable' as a standard of valuation of arguments.)

Jackson, B.S. (1988) *Law, Fact, and Narrative Coherence*, Roby: Deborah Charles.(Contains an interesting argument for reasoning as essentially narrative in structure, and a strong critique of deductivism.)

Levi, E.H. (1948) *An Introduction to Legal Reasoning*, Chicago, IL: Chicago University Press.(A classic account of the role of analogy and induction in the development of law - how the exception can grow to become the rule.)

MacCormick, N. (1978) *Legal Reasoning and Legal Theory*, Oxford: Clarendon Press.(An introduction to many of the ideas stated above.)

MacCormick, N. and Summers, R.S. (1991) *Interpreting Statutes: A Comparative Study*, Aldershot: Dartmouth

Publishing Co.(Essays on interpretation and the theory of interpretation taking account of nine major legal systems.)

Peczenik, A. (1989) *On Law and Reason*, Dordrecht/Boston/London: Kluwer.(An excellent account of the unity of practical reason and the differences of factual, moral and legal reasoning.)

Perelman, C. (1958) *La Nouvelle Rhétorique (The New Rhetoric)*, Paris: Presses Universitaires de France.(An outstanding statement of the modern significance of rhetoric in the Aristotelian tradition, commented on, for example, in Alexy above.)

Raz, J. (1975) *Practical Reason and Norms*, London: Hutchinson.(The classical account of rules as ‘exclusionary reasons’.)

Schauer, F. (1991) *Playing by the Rules*, Oxford: Clarendon Press.(Gives an account partly comparable with Raz (above) of the role of rules in practical reasoning.)

Summers, R.S. (1978) ‘Two Types of Substantive Reasons: The Core of a Theory of Common-law Justification’, *Cornell Law Review* 63: 707-88.

Toulmin, S.E. (1964) *The Uses of Argument*, Cambridge: Cambridge University Press.(Challenges traditional interpretations of logic and explores structures of practical arguments; hence influential in discussions of practical reasoning in law.)

Legalist philosophy, Chinese

Legalist philosophy constitutes one of the three dominant streams of Chinese philosophy along with Confucian and Daoist philosophies. It aims to establish objective, impartial and impersonal standards for human conduct. It sets forth prescriptive models using such metaphors as the builder's plumb line and carpenter's L-square. 'Modelling after' implies reshaping and remoulding of human behaviour, not by moral suasion but by the application of fa, *a term designating both prescriptive standards and promulgated penal law, designed to achieve public interest.*

*The idea of remoulding by punishment and reward is predicated on the Legalist conception of human nature. Innately self-interested human nature underlies human behaviour of liking reward and disliking punishment. Hence, penal law is both natural and an objective prescriptive technique for behavioural control that seeks to harmonize both the individual and the public interest. Penal law is efficacious in so far as it is issued from an authoritative power (*shi*) based on impersonal, institutionalized position of rulership and borne up, however tacitly, by the support of the people.* Shi *cannot govern effectively, Legalists argue, without the organizational power of bureaucracy under the centralized control of the ruler. For the ruler, controlling bureaucracy means mastering the technique (*shu*) of comparing 'word' (*ming*) and actual 'performance' (*xing*), not only through objective mechanisms of empirical verification but also by means of 'the two handles' of power over life and death. Hence, the technique holds bureaucracy accountable.*

The Legalist philosophy of governing by fa, shi *and* shu *was in effect a new model for sociopolitical reorganization. It became increasingly popular during the Warring States period, a time of incessant political struggle and of irreversible systemic disintegration of the Zhou feudal order. Legalists called for a radical systemic transformation through this new model in the name of historical relativism: 'There are as many situations as there are generations… and situations change, so the measures change' (*Han Feizi 49*).*

Historical relativism notwithstanding, Legalist philosophy envisages a 'natural' and 'automatic' polity that, once established, accords with dao *(that is, the way the natural world operates spontaneously). The ruler practices 'non-action', 'emptiness' and 'quiescence' so as to embody* dao, *and thereby personifies objective, impersonal standards over subjective, personal preferences. Once this 'natural' polity is established, the ruler does not act while his subordinates act according to* xingming *accountability as noted above. The ruler does not act as the centre point of the scale does not move, and yet he knows which side is heavy and which side light. In the end, the ruler does not act so that he can act, that is, so he can employ 'the two handles' to control his subordinates. This seemingly 'natural' and 'automatic' polity still requires a 'sage' ruler extraordinarily adept at covert statecraft. Only such a 'sage' ruler can hope to achieve order, wealth and power for himself and for his people.*

1 *Fa* as prescriptive standard and penal law

The term *fa* means not only 'law' but also means 'model', 'standard' or 'method'; therefore, it needs to be understood contextually (see Fa; Law and ritual in Chinese philosophy). Hence, the conventional translation of *fajia* (school) as the 'Legalist School of Thought' is misleading in that it overemphasizes the term 'law'. Moreover, the term *fajia* is retrospectively attributed by the Han doxographers, particularly Sima Tan (d. 110 BC), to a group of thinkers and statesmen who shared a sufficient measure of common philosophy of governing but certainly did not constitute a 'school' in the sense of the Confucian and Mohist schools. They did not have master or canonical texts and they did not live a communal life.

The third of the 'Seven Standards' of *Guanzi*, attributed to the proto-Legalist Guan Zhong (d. 645 BC), defines *fa* as 'standard' and refers to the carpenter's ink and line, compasses and L-square, the scale and the volume measures, metaphors that suggest mechanical precision and impartiality. The 'Seven Standards' are set forth as the knowledge required by the ruler for effective government. The fifth is specifically designed for transforming or reforming the people: 'Giving to or taking from, endangering or securing, benefiting or harming…killing or giving life, are called "incentives" and "deterrents"' (*Guanzi* 35). The idea of 'standard' is succinctly stated by the pre-eminent Legalist Han Feizi: 'The [*fa*] no more makes exception for men of high station than the plumb line bends to accommodate a crooked place in the wood' (*Han Feizi* 6).

Fa, understood as 'model' or 'standard' of behavioural pattern, is not exclusively Legalist; it is clearly present in

Mozi (canons A70, A94) and *Mencius* (4B1). These texts placed primary emphasis on 'government by virtuous example' and held that even the benevolent heart of a virtuous ruler would not achieve good government without the aid of the carpenter's compass and L-square. While the punitive connotation of the word *fa* became more pronounced in the Warring States period, for Legalists, *fa* continued to mean a whole network of 'standardized' patterns of behaviour. Thus to the Legalist Shang Yang (d. 338 BC), 'changing *fa*' meant a call for a new 'model' of sociopolitical institutions with corollary implications for behavioural change according to the Legalist standards.

Fa is most commonly understood as penal law. Contrary to conventional wisdom, however, the Legalists did not invent 'severe punishments and strict laws'. In the proto-Confucian canonical texts such as the *Shujing (Book of Documents)*, King Wu, the founder of the Zhou dynasty (*c*.1040-56 BC), and his brother Duke of Zhou emphasized the rational role of coercive force for a righteous cause. The Duke of Zhou advises Prince Kang to administer severe punishments according to the laws and not to bend them to agree with his 'inclinations'. Those who commit '[serious] crimes…are greatly abhorred, and how much more detestable are the unfilial and unbrotherly.… You must punish them severely' (*Shujing* IX). As Schwartz (1985) observes, the enduring legacy of meting out 'the most horrendous penal sanctions' for violation of family morality, which was supposedly held together by kinship affection, was legitimized by canonical texts. The Legalist doctrine of 'severe punishments and strict laws' likewise harks back to canonical texts. In these texts, however, there exists a judicious balance between civil and coercive administration, whereas Shang Yang and Han Feizi stress Draconian penal codes under the compelling exigencies of Warring States power politics.

2 *Fa* and human nature

How do Legalists justify their claim that *fa* is both 'objective' and 'natural'? How is *fa* related to their conception of human nature? In fact, the Legalist approach to human nature is sociological. Of the goodness and badness of human nature, Guanzi observes: 'It is not a matter of man's nature, but of poverty' *Guanzi* 35). Han Feizi speaks of changing values and behavioural patterns when the population increases in geometric progression while provision for it grows in only arithmetic progression (*Han Feizi* 48). As a consequence, people compete so fiercely for survival that disorder becomes rampant even if rewards are doubled and punishments multiplied. Han Feizi contends: 'The generosity with resources in ancient times was not benevolence, it was because resources were ample; the competition and robbery is not dishonesty, it is because resources are sparse.' Thus the Legalist philosophy of human nature is significantly different from that of Hobbes.

Legalists argued moreover that self-interestedness does not necessarily imply harming others. Han Feizi observed that the physician sucks patients' wounds and holds their blood in his mouth not because he loves them but because he expects profit from them. It does not mean the cartwright is 'benevolent' for wishing people to be wealthy enough to buy his carriages, nor does it mean the carpenter is 'cruel' in desiring to sell as many coffins as he can. Self-interestedness is clearly differentiated from selfishness, for it is considered a natural human condition that is ethically neutral.

People are more powerfully motivated by self-interest than the sentiments of concern for each other. This is also recognized in Western philosophy. In his *The Theory of Moral Sentiments*, Adam Smith says: 'Every man is, no doubt, by nature, first and principally recommended to his own care.' Thus, when one hears of millions of Chinese being swallowed up by a stunning and hellish earthquake, one might momentarily feel badly but would then go about one's business and sleep soundly. 'If he was to lose his little finger tomorrow, [however], he would not sleep tonight' (Smith 1759 [1976]: III.3.4). Legalists assert that *fa* must accord with human feelings of self-interest for it to be 'objective' and 'natural'. Because human feelings have likes and dislikes, reward and punishment can be effectively applied, and when prohibitions and order prevail, good government can be achieved. The critical function of *fa* is to effectively regulate and channel the general tendency of human nature toward personal welfare in order to realize the principal goals of good government.

Given the natural tendency to be self-interested, how many people would spontaneously harmonize their concern for private welfare with public interest? How is the convergence of private and public interest to be effected? Han Feizi replies that one can neither count on arrows that are straight of themselves nor count on pieces of wood that are round of themselves in a thousand generations, and 'yet how is it that people of every generation ride carts and shoot birds? It is because the tools for straightening and bending are used…' (*Han Feizi* 50). The 'tools' are, of

course, *fa*. *Fa* to one who governs is like the carpenter's L-square and compass - the objective, impersonal and prescriptive instruments for 'straightening and bending' - in short, for remoulding and channelling the general tendency of self-interested human nature toward the realization of public purposes.

What is new about Legalist view of *fa* is that it is consciously designed to institute an automatic polity (see §6). Individuals are compelled to grasp the utilitarian logic of automatically and habitually moulding their behaviour in conformity with *fa*, not only as objective standards but also as prescriptive 'tools' of penal codes, so as to achieve their self-interested goals. Given self-interested human nature, once the prescriptive mechanism of penal codes is in place, human behaviour becomes automatic and predictable, just as Adam Smith's 'economic man' becomes predictable once a free market is established. Legalists envisage a habitual convergence of private and public interest which will become automatic in due course. It is clearly more in one's own interest to be rewarded for hard labour in farming and for valour in battle than to face the terrible consequence of behaving otherwise.

3 *Shi*: authoritative power base

Shen Dao (*fl. c.*310 BC), the Legalist best known for his theory of *shi* or authoritative power base, regarded *shi* as a 'natural position of advantage' *vis-à-vis* others. *Shi* is a charismatic and mysterious 'potency' (*de*) that inspires awe and obedience among the masses. It is *shi* that enables the flying dragon to ride the clouds and the soaring serpent to float in the mist. Without the *shi* of clouds and mist, the dragon and serpent might as well be earthworms (fr. 10). 'As a simple yeoman', without *shi*, even Yao, a pre-dynastic sage ruler, 'could not command his neighbours' (fr. 12), while Jie, the depraved last emperor of the Xia dynasty, because of *shi*, could wreak havoc on the whole world. In essence, Shen Dao was expounding the very phenomenon of authority, *conditio sine qua non* for an orderly and civilized society.

None the less, Han Feizi's critique of Shen Dao centres on the theory of 'natural position of advantage', obviously analogous to the Confucian view of 'superior moral qualities' of virtuous rulers (see Han Feizi). Confucians argue that it is the superior qualities of the dragon that enable it to soar, and it is the superior moral virtues of Yao that brought order and benefits to the world. The argument curiously ends with an observation: 'It belongs to man's essential nature that the worthy are fewer than the unworthy' (*Han Feizi* 40). If good government depends on the worthy, Han Feizi argues, the world is condemned to almost continuous disorder and misgovernment inasmuch as Yao only appears once in 'a thousand generations'. His concern is with the rulers in between the exceptionally virtuous Yao and the notoriously depraved Jie, that is, those who regularly occupy the position of *shi*.

For Han Feizi, then, good government depends not so much on 'superior moral qualities' of one who holds authority as on *shi* being 'properly ordered'; that is, *shi* securely based on clearly defined and promulgated laws that are rigorously enforced. In short, there exists a contradiction between *shi* based on 'positive' law and the imperatives of a 'higher' moral norm. Han Fei asserts that *shi* based on law and a moral code are mutually exclusive: 'A man selling spears and shields was praising his shields as so hard that nothing could penetrate them. Immediately afterwards, he said in praise of his spears "My spears are so sharp they will penetrate anything." When some one asked "What if I penetrate your shield with your spear?" he was at a loss to answer.' Han Feizi concludes: 'The two names [power and morality] cannot be made to stand [simultaneously]. The Way of the Worthy may not be forbidden, yet it is the Way of Power to forbid anything' (*Han Feizi* 40).

In Han Feizi's conception of *shi*, there is an unmistakable shift toward an institutionalized and impersonal authoritative power exclusively based on law, consciously designed by statecraft and bureaucratically administered under the ultimate sanction of 'the two handles'. Shen Dao too is aware of the human fallibility of those who wield authority and the need for objective standards: 'When one who rules men casts aside the laws and governs on his own initiative, the punishments and rewards, seizures and grants, will be measured out according to the discretion of the ruler' (fr. 61).

Shen Dao moreover affirms an important attribute of authority, namely, 'it has weight' because it 'receives aid from many' (fr. 14). 'That the power of the Three Kings [three dynasties of Xia, Shang and Zhou] and the Five Protectors [of the Spring and Autumn period] was the same order as that of Heaven and the Earth… was due to the fact that they obtained aid from all' (fr. 16). Han Feizi likewise affirms: 'When the ruler wins the hearts of the people', he elevates himself to the position of authority 'without being raised' because the ruler is 'upheld by the masses of people with united hearts' (*Han Feizi* 28). However, he asserts that the ruler's *shi* is upheld by the

masses not because of ‘natural position of advantage’ but rather because the well-ordered system of *fa* underpinning the ruler’s *shi* will in the long run bring order and benefits to the people. In the final analysis, Legalists placed greater emphasis on the impersonal and institutionalized *shi* of rulership firmly based on *fa* than on authority borne up by ‘natural position of advantage’ or ‘virtuous qualities’ of the ruler. The concept of the ‘consent of the governed’ as the basis of legitimate authority, of course, was never a part of Legalist elitism.

4 *Shu*: the art of rulership

As is true of any complex sociopolitical entity, the Legalist model of sociopolitical organization cannot function without authority, nor can a ‘properly ordered’ authority based on law govern well without an efficient bureaucracy. The ruler who wields authority must therefore master *shu*, the ‘method’ or ‘technique’ of controlling bureaucracy and its personnel.

Shen Buhai (d. 337 BC) is best known for his theory of *shu*, although *shu* as the arcanum of government had been an essential part of covert statecraft long before his theorizing about it. What is significant about Shen’s theory of *shu* is that it specifically sets forth a rational paradigm for bureaucratic accountability. Furthermore, it advances a theory of bureaucracy based exclusively on merit. Shen says: ‘"Method" is to scrutinize achievement and [on that ground alone] to give rewards, and to bestow office [solely] on the basis of ability’ (fr. 23).

Shen Buhai’s theory, elaborated by Han Feizi, argues that *shu* consists in holding officials accountable by impersonal application of objective and verifiable standards, so that an official’s actual ‘performance’ accords with the ‘name’ of the office held. He uses the term ‘tally’ to connote a method of precise calibration and verification. For instance, an official holds the debtor’s portion of a tally according to his title (‘name’), and he is accountable for actual performance that tallies with the creditor’s portion held by the ruler. Holding firm to the ‘main cord’, the ruler controls subordinates by means of the tallies. Both the metaphor of ‘main cord’ of a tightly-knit net (fr. 1(4)) controlling the tallies and the concept of ‘cutting deeply’ (fr. 27) as a means of holding subordinates accountable correspond to the Legalist concept of centralized control by holding fast to ‘the two handles’. Whenever a minister utters a ‘word’ (proposals), the ruler measures it against actual performance. If the actual performance indeed ‘tallies’ (*Han Feizi* 5) with the ‘word’, the minister is rewarded; if the performance does not ‘tally’ with the ‘word’, the minister is punished. Han Feizi’s exacting theory of accountability derives from his own experience as a prince of a ruling house. The critical observations and remonstrations throughout his writings are, in effect, diagnoses of the political pathology of his time and his prescriptions for remedying them. In particular, he addresses the problem of sovereigns being deluded by irresponsible eloquence that promotes partisan interests and weakens the state.

Han Feizi also drives home the harsh consequences of violating the requirements of functional specificity:

> Once in the past Marquis Chao [Zhao] of Han got drunk and fell asleep. The keeper of the royal hat, seeing that the marquis was cold, laid a robe over him. When the marquis awoke, he was pleased and asked his attendants, ‘Who covered me with the robe?’ ‘The keeper of the hat’, they replied. The marquis thereupon punished both the keeper of the royal hat and the keeper of the royal robe. He punished the keeper of the robe for failing to do his duty, and the keeper of the hat for overstepping his office. It was not that he liked the cold, but he considered the trespass of one office upon the duties of another to be a greater danger than cold.
>
> (*Han Feizi* 7)

How does Han Feizi or Shen Buhai know that the tallies match? The ruler’s *shu* is twofold: it consists of (1) empiricism and (2) purposive Daoism (see §6). While Shen Buhai speaks of ‘looking into affairs by means of their names’ (fr. 1(9)) and ‘careful supervision’ of subordinates (fr. 17(1)), he relies more heavily on the *shu* of purposive Daoism. Han Feizi emphasizes both and provides the most elaborated philosophical arguments for purposive Daoism. He stresses the need for the ruler to gather ‘causes of different affairs for comparison and observation’ (*Han Feizi* 17) and to follow ‘the procedures of comparison and verification to judge [actual performances]’ (*Han Feizi* 14), for ‘to be sure of anything that has no corroborating evidence is stupid’ (*Han Feizi* 50).

Fa and *shu* are different in that *fa* is overtly promulgated and primarily intended for regulating the behaviour of general public, whereas *shu* is the covert arcanum of government and primarily intended for controlling the

behaviour of bureaucracy. *Fa* and *shu* are similar in that they are impersonal and objective 'tools' for controlling human behaviour. Han Feizi criticized Shang Yang for having neglected *shu* and found Shen Buhai wanting for not having unified *fa*. *Fa*, *shi* and *shu* constitute the three interdependent pillars of Legalist philosophy.

5 Historical relativism

Shang Yang relates three different historical periods and their corollary behavioural and value patterns (*Shangjunshu* 1). Since the former generations did not share the same doctrines, he asks rhetorically, 'Which antiquity shall one imitate? The emperors and kings did not copy one another, so what rites should one follow?' (*Shangjunshu* 7). He saw more than one way to govern the world and found no need to imitate antiquity. He was quite prepared to institute new *fa* to remould the people (see §1).

Han Feizi elaborates on Shang Yang's historical relativism. Radical in thought and perceptive of what was to emerge in the wake of the systemic disintegration of the Zhou feudal order, he advances a new philosophy of history that directly challenges the traditional and characteristically Confucian 'Return to the Time of the Ancient Sages and Early Kings' philosophy of history (see History, Chinese theories of). His historiography divides the history of China into three eras: (1) remote antiquity, (2) the middle age and (3) the present age, each era having its distinctive character: 'Men of remote antiquity strove to be known as moral and virtuous; those of the middle age struggled to be known as wise and resourceful; and now men fight for the reputation of being vigorous and powerful' (*Han Feizi* 49). The difference between antiquity and 'the present age' is that in antiquity 'events were few' and 'measures were simple'. Because there were relatively few people, it was easier to make light of profit and to yield to others, 'But the present age is one of numerous affairs and great struggle' (*Han Feizi* 47). He succinctly summarizes his historical relativism: 'There are as many situations as there are generations' and 'situations change, so the measures change' (*Han Feizi* 49).

Han Feizi's historiography is iconoclastic in that he unmasks the irrelevance of the sacred precepts of ancient sages and early kings in and of themselves. He is primarily concerned with the effective ways of dealing with the reality of his own time according to the principle: 'situations change, so the measures change'. He confidently concludes that *fa*, understood as a new model of sociopolitical institutions with the corollary implications for behavioural change, is the most effective instrument for dealing with the reality of 'the present age' (see §6). In his historical relativism, the main emphasis is on the 'situations change', not 'measures change':

> Change or no change, the sage does not mind. For he aims only at the rectification of government. Whether or not the ancient traditions should be changed, whether or not the existing institutions should be removed, all depends upon the question whether or not such institutions and such traditions are still useful for present-day political purposes.
>
> (*Han Feizi* 18)

Therefore, he was judicious in avoiding indiscriminate rejection of the sacred precepts of the past merely for the sake of change. Rather, his criticism was directed against those who, while oblivious as to the 'condition of the wicked and the villainous' and clueless as to the effective means of rectifying the course of the state, blindly venerate the past and bewilder the ruler. Han Feizi's overriding concern was no less than 'to save the present age'; hence, he placed particular emphasis on 'politics of consequence', not unlike Max Weber's 'ethics of responsibility'.

6 The Legalist vision and Daoism

Despite historical relativism, the new model of Legalist institutions represents an ahistorical 'ideal type'. This paradigm of Legalist polity is consciously designed to incorporate certain aspects of Daoism so that, once established, it operates 'naturally' and 'automatically'. In short, this paradigm may be equated with the Legalist vision and response to the perennial question of 'what is the good society?' In contrast, Machiavelli was primarily preoccupied with the techniques of manipulating power and was less concerned with devising a new paradigm.

The Legalist search for objective, impersonal and impartial standards goes beyond a simple rationalization to justify political power. Han Feizi presents the most sophisticated philosophical inquiry into the relationship between Legalism and Daoism (see Daoist philosophy). He comments extensively on various sections of Laozi's *Daodejing* (see Daodejing) in chapters 20, 'Interpreting Laozi', and 21, 'Illustrating Laozi'. He writes: '*Dao* is

that from which all things are as they are, and it is the collecting point of all principles [*li*]' (*Han Feizi* 20). Things are essentially what they are because of their respective principles (*li*) (see Li).

Li etymologically refers to 'venation in pieces of jade'; hence, it symbolizes natural pattern and order that are independent of human thinking, willing and behaving. Rocks are hard; dogs bark; the four seasons run their course; day follows night; people are born and they die.

Water has its own nature: one who is drowning dies as one drinks too much of it, and one who is thirsty lives as one drinks a proper amount of it. Han Feizi's *li* represents the constants or 'objective regularities' in the natural order of things. Human beings may choose to accord with *li* or go against it at their own risk.

Shen Buhai says: 'Heaven's Way has no private [concerns] therefore, it is always correct' (fr. 8). It is the humans who must discern and conform to the constant *dao* to be 'correct', for *dao* 'embraces all opposites' and is not partial to human concerns (see Dao). It is *dao* that patterns the universe and is the source of *li*: '*dao* is the beginning of myriad things and the standard of right and wrong' (*Han Feizi* 5). Han Feizi sees the natural world and the human world in one continuum. The very same *dao* and *li* pervade both worlds, and it is by being informed by *dao* and *li* that the ruler is able to conform to the truly objective and natural order of things and realize the fullness of one's 'potency':

> The ancients who completed the principal features of Legalism, looked upon Heaven and earth, surveyed rivers and oceans, and followed mountains and ravines; wherefore, they ruled as the sun and the moon shine, worked as the four seasons rotate, and benefited the world in the way clouds spread and winds move.
>
> (*Han Feizi* 29)

Thus, the ruler should model after *dao* and govern according to *dao*'s objective standards by not acting (*wuwei*) contrary to the natural order of things and by cultivating 'emptiness' and 'quiescence' so as to embody *dao*. In so doing, the ruler acquires the 'potency' (*de*) for governing (see De). Han Feizi advises the ruler to be 'humble'; this means not only refraining from overindulgence in seeing and hearing but also moderating mental activity. In short, it means bringing one's sensual and heart-mind activities under control so that one knows how to 'conserve'. By knowing how to 'conserve' through restraining his subjective inclinations, the ruler is better able to accord with *dao* and *li*.

Having instituted the new model of Legalist polity according to the objective and impersonal standards of *dao* and the natural feelings of 'likes' and 'dislikes' of human nature, the sage ruler reposes in 'non-action' (*wuwei*): 'He remains empty and waits for [the subordinates'] service, and they will exert [*yuwei*] their abilities by themselves' (*Han Feizi* 8). Thus, a *wuwei*-*yuwei* dialectical relationship marks the functional differentiation of the ruler and the subordinates. The ruler, uniquely personifying *dao* and grasping 'the two handles', reposes in *wuwei*, while the subordinates actively engage in fulfilling their *xingming* accountability. The sage ruler, by his gnostic grasp of *dao*, knows the source of everything and, by being in accord with objective standards, knows the 'origin of good and evil'. Such a ruler, by practising *shu* of 'emptiness' and 'quiescence', waits for all 'names' to name themselves and all 'affairs' to settle themselves as though automatically (*Han Feizi* 5). Han Feizi resorts to metaphors to clarify the *shu* of 'non-action': like the stationary mirror, the ruler can see and compare the beautiful and the ugly; like the centre point of the scale, he can weigh the light and heavy without moving (*Han Feizi* 19).

Is this comparing and weighing truly automatic? However 'natural' and 'automatic' the new paradigm of Legalist polity may seem, in the final analysis, the ruler does not act *so that* he can act, and Legalist polity requires a ruler who is exceptionally adept at covert techniques of statecraft. Like the Smithean theory, the Legalist idea of 'automatic society' proved to be illusory.

7 Enduring significance of Legalist philosophy

First, Han Feizi's drawing a parallel between Heaven and the ruler is highly symbolic, for it implies the ruler's uniqueness beyond the ordinary social hierarchy. To accentuate this uniqueness, he pushes the comparison to its ultimate level, drawing an analogy between *dao* and the ruler: '*Dao* is never a pair. Hence it is called one. Therefore, the intelligent ruler esteems singleness, the characteristic of *dao*' (*Han Feizi* 8). Just as *dao* does not identify with myriad things, accordingly the ruler sets himself apart from his ministers and people. It is in this conscious personification of *dao* that the ruler transcends the rest and achieves uniqueness. This uniqueness in turn

endows him with an extraordinary power to bring all affairs under control (*Han Feizi* 29). Han Feizi may well be basing his theory of the ruler's uniqueness on Chapter 25 of the *Daodejing*: 'Hence the way [*dao*] is great; heaven is great; earth is great. Within the realm there are four things that are great, and the king counts as one.' He may also base his philosophical argument for the ruler's extraordinary power to control all affairs on Chapter 14 of the *Daodejing*: 'Hold onto the *dao* of old in order to master the things of the present' (see Daodejing).

Thus, Han Feizi provided the most sophisticated philosophical justification for autocracy. The enduring significance of Han Feizi's Legalist philosophy is the legacy of two millennia of autocracy in China. Despite the dominance of Confucian orthodoxy, the political institutional framework of autocratic, centralized power has been essentially Legalist. Further, China is and has been an eminently bureaucratic society. Shen Buhai and Han Feizi's rational paradigm for bureaucratic accountability and merit-based functional specificity - for better or worse - has shaped the Chinese pattern and rhythm of life up to the present day. Of course, there were also strong departures from the Legalist model. However, after the eighteenth-century *philosophes* took a fancy to the Legalist model of bureaucracy based on merit, Shen Buhai may be considered 'the ancestor' of a worldwide institution of bureaucracy (Graham 1989). Lastly, the covert tradition of political technique (*shu*), inspired by purposive Daoism, namely 'non-action' as well as decisive action according to 'timeliness', continues to remain a significant aspect of Chinese philosophy of governing and political culture.

See also: Authority; Confucian philosophy, Chinese; Marxism, Chinese; Chinese philosophy; History, Chinese theories of; Dao; Daoist philosophy; Fa; Han Feizi; Law and ritual in Chinese philosophy; Li; Political philosophy, history of; Political philosophy, nature of

LEO S. CHANG

References and further reading

Chang, L. (1979) 'Metamorphosis of Han Fei's Thought in the Early Han', in H. Rosemont and B. Schwartz (eds) *Studies in Classical Chinese Thought*: *Journal of the American Academy of Religion* 47 (3): 503-47.(An exploration into the continuing influence of Han Feizi's thought in the Early Han.)

Chang, L. and Wang, H.P. (1983) *Han Fei sixiangdi lishi yanjiu (A Research into the History of Han Fei's Thought)*, Taipei: Linking (Lianjing) Publishing Co.; 2nd edn, Beijing: Zhonghua Book Co., 1986.(A detailed exposition of Han Feizi's main ideas.)

Fung Yu-lan (1952) *A History of Chinese Philosophy*, trans. D. Bodde, Princeton, NJ: Princeton University Press. (Vol. 1, chapter 13 contains a clear exposition of Legalist philosophy for upper level undergraduates.)

Graham, A.C. (1989) *Disputers of the Tao: Philosophical Argument in Ancient China*, La Salle, IL: Open Court. (Erudite and evenhanded treatment of Legalist philosophy; for advanced students.)

Guan Zhong (d. 645 BC) *Guanzi*, trans. of twelve chapters and notes by W. Allyn Rickett, *Kuan-tzu: A Repository of Early Chinese Thought*, Hong Kong: University of Hong Kong Press, 1965; trans. of thirty-five chapters and notes by W. Allyn Rickett, *Guanzi*, Princeton, NJ: Princeton University Press, 1985.(The work attributed to Guan Zhong is a miscellaneous collection of writings probably from the fourth to the second century BC. This is the most authoritative translation, intended for scholars in the field.)

Han Feizi (*c*.280-233 BC) *Han Feizi*, trans. W.K. Liao, The *Complete Works of Han Fei Tzu*, London: Arthur Probsthain, vol. 1, 1938; vol. 2, 1959; trans. of twelve chapters and notes by B. Watson, *Han Fei Tzu: Basic Writings*, New York: Columbia University Press, 1964.(A good introduction to Legalist philosophy for upper level undergraduates. Liao is the only complete translation.)

Han Feizi (*c*.280-233 BC) *Han Feizi jishi (Collected Explanations of Han Feizi)*, ed. Chen Qiyu, Beijing: Zhonghua Book Co., 1958.(One of the exhaustively annotated texts on Han Feizi.)

Han Feizi (*c*.280-233 BC) *Zengding Han Feizi jiaoshi (Further Revisions of Collations and Explanations of Han Feizi)*, ed. Chen Qitian, Taipei: Taiwan Commercial Press, 1972.(Like Chen Qiyu's work, it is one of the thoroughly annotated texts on Han Feizi. This work has an unconventional textual sequence.)

Schwartz, B.I. (1985) *The World of Thought in Ancient China*, Cambridge, MA: Harvard University Press. (Sophisticated and insightful exposition of Legalist philosophy that transcends the stereotypical views; for advanced students.)

Shang Yang (390-338 BC) *Shangjunshu*, trans. J.J.L. Duyvendak, *The Book of Lord Shang*, London: Arthur Probsthain, 1928; repr. Chicago: University of Chicago Press, 1963.(Attributed to Shang Yang, this work was in all probability compiled over several decades at the beginning of the third century BC. This is the only translation available.)

Shen Buhai (d. 337 BC) Fragments, in *Shen Pu-hai: A Chinese Political Philosopher of the Fourth Century B.C.*,

Appendix C: The Shen Pu-hai Fragments, trans. H.G. Creel, Chicago: University of Chicago Press, 1974. (Thoroughly researched but controversial work; advanced reading. The actual dates of the Shen Buhai fragments are unknown.)

Shen Dao (*c*.310 BC) Fragments, in P.M. Thompson (trans.) 'The Shen Tzu Fragments', University of Washington thesis, 1970, 512-75.(Painstakingly researched syntactical and philological exegesis; the only available translation.)

Smith, A. (1759) *The Theory of Moral Sentiments*, ed. D.D. Raphael and A.L. Macfie, Oxford: Clarendon Press, 1976.(Smith's major work on moral philosophy.)

Vandermeersch, L. (1965) *La formation de Légisme: recherche sur la constitution d'une philosophie politique caractéristique de la Chine ancienne (The Foundation of Legalism: Researches on a Political Philosophy unique to Ancient China)*, Paris: Publications de École Française d'Extrême-orient.(A comprehensive and imaginative treatment of the central figures and their development of Legalist 'mentality' in their respective periods.)

Wang, H.P. (1991) *Xianqin fajia sixiangshi lun (Discourses on Pre-Qin Legalist Thought)*, Taipei: Linking (Lienjing) Publishing Co.(The most comprehensive and lucid account of Legalist philosophy.)

Wang, H.P. and Chang, L. (1986) *The Philosophical Foundations of Han Fei's Political Theory*, Honolulu, HI: University of Hawaii Press.(An elaboration of the Legalist vision and Daoism.)

Legitimacy

Legitimacy refers to the rightfulness of a powerholder or system of rule. The term originated in controversies over property and succession, and was used to differentiate children born of a lawful marriage from those who were 'illegitimate'. From thence the term entered political discourse via controversies over the rightful succession to the restored French throne after the Napoleonic period. However, questions about what makes government rightful have been a central issue of philosophical debate since the ancient Greeks, and in this sense the concept, if not the term, 'legitimacy' is as old as political philosophy itself. Its significance lies in the moral, as opposed to merely prudential, grounds for obedience which follow for subjects where power is rightfully acquired and exercised, and in the depth of allegiance which such political authorities can call upon in times of difficulty.

What, then, makes government legitimate? Most thinkers agree that a necessary condition is that power should be acquired and exercised according to established rules, whether these are conventionally or legally defined. However, legal validity cannot be a sufficient condition of legitimacy, since both the rules and the power exercised under them also have to be morally justifiable. Two broad criteria for moral justifiability can be distinguished: (1) political power should derive from a rightful source *of authority; (2) it should satisfy the rightful* ends *or purposes of government. Most philosophical disputes about legitimacy take place either within or between these two broad positions; any adequate account of it must embrace both however.*

1 The source and ends of political authority

Historically, claims made for the rightful source of political authority have been enormously diverse, including: divine authority, mediated via religious doctrine and its interpreters; tradition and the principle of continuity with a society's past; some infallible truth, such as Marxism-Leninism, about society's future destiny and its chosen agents; and the people as a whole, and the principle of popular sovereignty. Although the first of these is still defended within parts of the Islamic world, most Western political philosophy of the modern era has sought to establish a secular source for political authority. With the authority of tradition relegated to at most a secondary plane, and Marxism-Leninism now discredited, the principle of popular sovereignty is virtually uncontested in the contemporary world. Much room for disagreement remains, however, about which constitutional arrangements best realize this principle; and - between competing conceptions of nation and nationalism - about who precisely constitutes the people that is the source of political authority (see Nation and nationalism).

The alternative idea - that legitimate authority is one which realizes the rightful ends or purposes of government - has historically produced an equally varied set of criteria, according to how these ends or purposes have been defined: whether government exists to guarantee basic societal needs for security and subsistence; to realize some ideal conception of the good society and virtuous citizenship; or to protect specified individual rights. However, a common feature of any 'performance' criterion of legitimacy, even more clearly than for the first definition, is to establish a standard against which individual rulers or regimes can be assessed, and, where need be, found wanting. That is to say, theorizing about legitimacy typically has a critical, rather than a merely apologetic purpose; and it tends to flourish when the authority of government is already uncertain or contested (see Authority).

2 Social contract theories

The two broad types of legitimating criteria considered above - the source and the ends of political authority - were brought together in the early modern period by theorists of the social contract, such as Hobbes and Locke. According to social contract theory, government was a product of conscious human artifice; its legitimacy derived from its authorization by the people themselves. At the same time the deficiencies of the pre-governmental 'state of nature', which led people to enter into the social contract, both defined the proper purposes of government and also set limits to its legitimacy: according to Hobbes (1651), through the provision of physical security; according to Locke (1689), through the protection of natural rights and the rule of law. Where these purposes were manifestly broken, the obligation of subjects to government was at an end. For both writers, then, where popular authorization established the *grounds* of legitimate rule, the true purpose of government defined its *limits* (see Contractarianism §2, 5).

Subsequent critics of the two writers pointed to their arbitrary characterization of the state of nature from which the ends of government were to be derived. They also questioned whether (given that one took place) a once-for-all authorization of government through the social contract could conceivably serve to legitimate political authority for subsequent generations. Locke rightly rejected Hobbes' contention that contracts made under the duress of conquest were valid. However, his own notion of 'tacit consent', whereby those who chose not to leave the country or disown their inheritance were presumed to have agreed to the existing governmental arrangements, hardly constituted a robust condition for popular consent to, or authorization of, government (see Consent §2).

In his work *The Social Contract*, Rousseau §3 claimed to have remedied the deficiencies in previous contract theorizing (Rousseau 1762). Political authority could be legitimate, he argued, only where citizens had the right to vote in person on legislation and where the executive was directly accountable to the sovereign legislative assembly. Whereas previous consent theory offered citizens no effective choice about accepting existing governmental arrangements, the unlimited legislative authority Rousseau gave his citizen body provided it with the continuous opportunity to modify received social and political rules. It could therefore meet the most strenuous conditions for popular consent to government. However, these conditions were bought at a price: a radical inequality of citizenship status between the sexes and the impracticality of direct legislation outside small-scale societies. If people could never autonomously consent to a representative system, then where could legitimate government realistically be found? Moreover, Rousseau's implausible contention that those who voted against any legislative proposal in the sovereign assembly had in some sense really willed it enabled him to evade the need to set constitutional limits to the scope of legislation, or to guarantee protection for individual or minority rights (see General will §§1-2).

3 'Performance' criteria of legitimacy

The cul-de-sac which social contract theory seemed to have reached with Rousseau led most subsequent thinkers to develop the second 'performance' dimension of legitimacy. For the utilitarians, the sole test of a rightful system of government lay in its capacity to promote the 'greatest happiness of the greatest number' (Bentham 1776, 1778) (see Bentham, J.). Democratic arrangements were to be defended, not on the principle of popular sovereignty, but only in so far as they served to protect the interests of the large majority against the sectional interests of the few (Mill 1821), or to develop those active qualities of citizenship which would enable people to relate their own interests to those of society as a whole (Mill 1861) (see Mill, J.S.). In contrast, idealist thinkers from Hegel to Green defined the purpose of government in more exclusively ethical terms: to develop the highest moral form of life for society. The rightful form of government was that which would be most conducive to this end.

'Performance' criteria of legitimacy have also been at the fore in recent political philosophy, albeit in a modified form of social contract theory. In the work of Rawls, J. and his followers, the idea of the social contract is purely hypothetical: it is a device for establishing what social and political arrangements people *would* agree to, if they could imagine themselves ignorant of the social positions they would occupy or the individual characteristics and preferences they might possess (Rawls 1971). From this imagined pre-political condition we can then derive a list of basic rights and freedoms and criteria for fair decisional procedures which should command universal assent. What makes government legitimate on this view is not its *actual* authorization by the people, but its conformity to standards which they *would* authorize if they were truly impartial.

4 Legitimacy and liberal democracy

Rawls' work comes close to characterizing the underlying principles of a liberal-democratic order. As a theory of legitimacy, however, it is inadequate, being too one-dimensional. To satisfy the principle of popular sovereignty, a liberal-democratic regime has to meet criteria for the *actual* authorization of government by the people: not only through the regular, free and fair election of public officials in conditions of electoral equality, but also through the realistic opportunity to modify received constitutional arrangements by means of referendum and citizens' initiative. Only the combination of these conditions could plausibly meet the criteria for popular consent and authorization which social contract theory was aiming for, but failed to achieve. At the same time, the liberal-democratic state has also to meet continually evolving standards for the protection of individual rights and guarantee of social welfare if it is to satisfy defensible performance criteria of legitimacy (see Democracy).

Any adequate philosophical account of legitimacy, therefore, has to consider both the rightful source and ends of

political authority; legitimate government must meet criteria of authorization and performance standards together. To recognize this is to identify one of the characteristic dilemmas of liberal democracy: the two dimensions of legitimacy may conflict with, as well as reinforce, one another. The governments and political arrangements that people actually choose to authorize may fail to meet defensible standards of fairness or rights protection for all citizens. This is because empirical persons with definable beliefs and interests differ in practice from the disembodied agents of a hypothetical social contract; and the circumstances of actual decision making can be far removed from the impartiality or Habermas' 'ideal-speech situation' of a pure legitimacy (see Habermas, J.). This tension is at its most acute in the conflict between forms of ethnic and particularistic self-definition of a people as a self-governing nation, on the one hand, and the universalistic standards implicit in any justifiable defence of individual rights or democratic citizenship, on the other.

5 Social science and legitimacy

Whereas a tendency of philosophical accounts of legitimacy is to dissolve the tensions between different legitimating principles by positing a transcendental world of impartial citizens inhabiting the good society, the tendency of social science accounts lies in a different, reductionist direction: to reduce legitimacy to whatever a given people *believes* to be legitimate. Social scientists are interested in legitimacy as an explanatory rather than a normative concept: to help explain why people actually obey authority, rather than why or when they should; why regimes actually persist, rather than whether they deserve to. From this perspective legitimacy tends to become, not a quality of regimes or governments, but an attribute of subjects: their internalized 'belief in legitimacy' (Weber 1914). Regimes are legitimate if they are able to generate and maintain the appropriate state of mind in their populations (Lipset 1960). Such an account produces a radical disjuncture between the projects of political science and political philosophy; analysing the empirical processes of *legitimation* comes to replace the construction of defensible criteria for *legitimacy*.

Social scientists can only escape from this reductionist dead-end if they acknowledge the same conceptual structure to legitimate power as do philosophers - as a matter of conformity to rules which are themselves justifiable in terms of specific beliefs about the rightful source and ends of political authority - and if they see the process of legitimation to lie in the actions of relevant subjects confirming their support for the regime or its powerholders (Beetham 1991). Equipped with this conceptual structure, social science can then make its own distinctive contribution to the study of legitimacy: through a comparative analysis of different legitimating principles and processes in different societies and epochs, and of the way these are embedded in institutional structures and practices; through an identification of the gaps that may develop between the underlying principles of a regime and their supporting beliefs, on the one hand, and its actual practice on the other; and through an examination of the processes of de-legitimation which may herald its collapse.

A comparative analysis of this kind can also serve to complement a philosophical account of liberal-democratic legitimacy by exploring why it tends to win out in competition with other legitimating principles, especially once a population attains a developed level of education and literacy. At this point the explanation of why societies come to have the legitimating principles they do cannot be divorced from the defensibility of their grounds for holding them; and such principles must be recognized as legitimate in a context of international, not merely national, scrutiny and debate. In the explanation of legitimacy, in other words, the projects of normative philosophy and a non-behaviouristic social science can be seen to become one and the same.

DAVID BEETHAM

References and further reading

Beetham, D. (1991) *The Legitimation of Power*, Basingstoke: Macmillan.(Comprehensive review of legitimacy from a social science perspective; see §5.)

Bentham, J. (1776, 1778) *A Fragment on Government and An Introduction to the Principles of Morals and Legislation*, ed. W. Harrison, Oxford: Clarendon Press, 1948.(Classic statement of the utilitarian criterion for rightful government; see §3.)

Connolly, W. (ed.) (1984) *Legitimacy and the State*, Oxford: Blackwell.(Useful introductory collection.)

Green, T.H. (1885) *Lectures on the Principles of Political Obligation*, ed. P. Harris and J. Morrow, Cambridge: Cambridge University Press, 1986.(Classic review of political obligation and its relation to state legitimacy; see §3.)

Habermas, J. (1976) *Legitimation Crisis*, London: Heinemann.(Influential account of the criteria for political legitimacy and the conditions for legitimacy crisis in contemporary capitalist democracies; see §4.)
Hegel, G.W.F. (1821) *Philosophy of Right*, ed. T.M. Knox, Oxford: Clarendon Press, 1952. (Classic account of legitimacy from the standpoint of philosophical idealism; see §3.)
Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991. (Controversial but highly influential early modern theory which derives political legitimacy from the authorization of an original social contract; see §2.)
Lipset, S.M. (1960) *Political Man*, New York: Doubleday.(Source for a subjective account of legitimacy in popular acceptability; see §5.)
Locke, J. (1689) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 3rd edn, 1988.(Important for a consent theory of political legitimacy, which also emphasizes the rule of law and rights protection as conditions; see §2.)
Mill, J. (1821) ‘Essay on Government’, in T. Ball (ed.) *James Mill: Political Writings*, Cambridge: Cambridge University Press, 1992.(Classic utilitarian statement of the criteria for good government; see §3.)
Mill, J.S. (1861) ‘On Representative Government’, in A.D. Lindsay (ed.) *J.S. Mill: Utilitarianism, Liberty, Representative Government*, London: Dent, 1964.(Important development of utilitarian theory, incorporating a more complex account of human interests; see §3.)
Rawls, J. (1971) *A Theory of Justice*, Oxford: Clarendon Press.(This famous work on justice also offers a theory of political legitimacy and of the limits of obligation; see §3.)
Rousseau, J.-J. (1762) *The Social Contract*, ed. M. Cranston, Harmondsworth: Penguin, 1965.(Important for the stringent criteria it sets for the legitimacy of, and popular consent to, law and government; see §2.)
Weber, M. (1914) *Economy and Society*, ed. G. Roth and C. Wittich, New York: Bedminster Press, 1968.(The source for much discussion of legitimacy within twentieth-century social science; see §5.)

Leibniz, Gottfried Wilhelm (1646-1716)

Leibniz was one of the central figures of seventeenth-century philosophy, indeed, one of the central intellectual figures of his age. Born and educated in Germany, he travelled to Paris in 1672 and quickly entered into its lively intellectual and scientific life, acquainting himself with the most advanced ideas then in circulation. It was there that he invented the infinitesimal calculus, and laid the foundations for the philosophical and scientific programmes that were to occupy him for the rest of his life. He returned to Germany in 1676, entering the service of the House of Hanover where, except for brief absences, he remained until his death. There, along with his court duties, he had time for a wide variety of intellectual activities that eventually gained him an international reputation.

Leibniz's philosophy, particularly his metaphysics, can appear otherworldly and complex. But there are a few simple themes and basic commitments that run through his thought. At root is his philosophical optimism, the commitment that this is the best of all possible worlds, freely created by a rational God who always chooses the best for a good reason. This best of all possible worlds, Leibniz held, is 'the one which is at the same time the simplest in hypotheses and the richest in phenomena' (Discourse on Metaphysics §6). For this reason, the world must be governed by a variety of general principles to which Leibniz appealed in his philosophy: there must be a sufficient reason for everything in the world; there are no jumps in nature; there must be exactly the same power in the full cause as there is in the complete effect, among many others. While such principles do not deductively determine the rest of Leibniz's philosophy, they do play a major role in shaping it; they constitute a kind of lens through which he viewed the major philosophical issues of his age.

One such issue concerns the ultimate make-up of the world. Like many of his contemporaries, Leibniz adopted a mechanistic view, according to which everything in the physical world is explicable in terms of the size, shape and motion of the tiny bodies that make up the grosser bodies of experience. But he rejected the idea that this could be the ultimate explanation for things. Behind the mechanistic world of inanimate bodies in motion, Leibniz saw a world of living things and souls - active, genuinely individual, genuinely different from one another, the true atoms of nature, the true reality - which he eventually called monads. At the deepest level, Leibniz's world was made up of an infinity of mind-like entities, each with its own perceptions that change from moment to moment according to an internal programme by way of the faculty of appetition, all in harmony with one another so that they all reflect the same world. While the world of physics is mechanistic, it is merely phenomenal, the confused appearance of a deeper reality. A consequence of this was Leibniz's famous doctrine of pre-established harmony. In contrast to Descartes, for whom mind and body interact, and in contrast to the occasionalists, for whom God is the true cause who brings about motion in the body on the occasion of a volition and a sensation in the mind on the occasion of a stimulation of the appropriate nerves in the body, for Leibniz God created the mind (a single monad) and the body (itself a collection of monads) in perfect harmony with one another so that their mental and physical states would always correspond in the appropriate way.

A second set of metaphysical issues of central concern to Leibniz involves the interlocking questions of necessity, contingency and freedom. In response to contemporaries such as Hobbes and Spinoza, Leibniz tried to find room for contingency and freedom in his world. He argued that even though God is, in a sense, constrained to choose the best, he does so freely. Consequently, the world he created, the best of all possible worlds, exists contingently, and at least some features of it are contingent, those whose contraries are not in themselves impossible. So for example, 2 + 3 = 5, true in every possible world, is necessary, while 'Adam sins', whose contrary is not impossible, is contingent. But over and above contingency and divine freedom, Leibniz also wanted to make room for human freedom. According to Leibniz, when God created Adam as a part of this best of all possible worlds, he knew that Adam would sin; it is part of the concept of Adam that he sins, part of his internal 'programme' that he will eat the apple, and part of the internal 'programmes' of the monads that make up his body that he will actually eat the apple. But, Leibniz argued, what God builds in is that Adam freely *chose to sin. God builds into the world the reasons that incline Adam's will without necessitating it, correctly predicting what Adam will do, and building the rest of the world around the consequences of Adam's free actions.*

Important as they are, these two concerns constitute only a small portion of Leibniz's thought, even within the domain of philosophy. In psychology, he introduced a distinction between conscious and unconscious perceptions

and tried to understand the way in which unconscious perceptions (‘petites perceptions’) in part determine conscious perceptions (‘apperceptions’). In epistemology, he is important for his sophisticated version of the innatist hypothesis, and for appreciating the role that a mathematical theory of probability can play in understanding the world. In logic, Leibniz advanced programmes for a new formal logic more powerful than Aristotle’s, and for a universal language. In ethics and political thought, he contributed to the seventeenth-century natural law tradition. In natural philosophy, he emphasized the importance of the notion of force and advanced the broadly Cartesian programme of a physics grounded in conservation laws. Outside philosophy he is well known for his work on the calculus. Though he co-discovered it with Newton, it is his notation that is still used, and his version probably had the greater influence in his day. But he was a major contributor to many other fields, including geology, natural history, linguistics and European history. Though he left no real school of followers, he deeply influenced philosophy after his death, particularly in eighteenth-century Germany.

1 Life

Gottfried Wilhelm Leibniz was born in Leipzig on 1 July 1646. He later recalled how his father, who died when he was only six years old, had instilled in him a love of learning. Leibniz started school when he was seven, but more important than his formal education in those years was his reading. He taught himself Latin at an early age in order to be able to read Livy and Calvisius, and because of that was admitted into his late father’s extensive library, where he read widely. At fifteen Leibniz entered University, first the University of Leipzig (1661-66), and then the University of Altdorf (1666-67), graduating with degrees in law and in philosophy. The education he received there was conservative, a mixture of traditional Aristotelian school philosophy and Renaissance humanism. Though invited to join the faculty at Altdorf, he chose instead to enter the service of the Elector of Mainz, where he stayed until he was sent to Paris in the spring of 1672 on diplomatic business.

While he had done significant work in a number of areas before going to Paris, including law, theology, mathematics and physics, the trip was crucial to Leibniz’s intellectual development. In the later part of the seventeenth century, learned Europe was in the midst of a great intellectual revolution; the older Aristotelian philosophy of the schools was being challenged by a new mechanist philosophy which rejected the form, matter and qualities of the Aristotelian world, replacing them with a world in which everything was to be explained in terms of size, shape and motion. In this new world there was a special emphasis on mathematics, which was increasingly applied to problems in physics in a way quite foreign to the Aristotelian philosophy.

Though he had taken an interest in the moderns while in Germany (Hobbes was particularly influential on his early thought), it was only after he reached Paris that Leibniz was able to enter the mainstream of European intellectual life. There he came to know the important mathematician and physicist Christiaan Huygens, who introduced him to new ideas which Leibniz absorbed quickly. In those years, Leibniz laid the foundations of his calculus, his later physics and his philosophy. While there were no publications at the time, many unpublished notes survive, important for understanding the emergence of his mature thought.

Leibniz returned to Germany in December 1676, passing through Holland, where he discussed philosophy with the reclusive Spinoza. It was then that he first entered the service of the House of Hanover. He served under Duke Johann Friedrich until his death in 1679, under Duke Ernst August from 1680 to 1698, and then, finally, under the Elector Georg Ludwig, who ascended the throne of Great Britain as King George I in 1714. Except for his travels, he remained at Hanover for the rest of his life. There Leibniz undertook a very wide variety of tasks. He served as a mining engineer (unsuccessfully supervising the draining of the silver mines in the Harz Mountains), as head librarian of a large collection of books, as a general advisor and a diplomat, and was particularly interested in finding ways for the Catholics and the Protestants to reunite. Leibniz was also given the responsibility for writing a history of the House of Hanover. While he collected and published many previously unknown historical documents and published a number of other historical writings, this project barely got off the ground. All that he seems to have completed was a geological history of the region of Lower Saxony, the *Protogaea*. While it proved to be an important work in the history of geology when it was finally published in 1749, it seems not to have pleased Leibniz’s employers who had hoped for a history of somewhat more recent times.

Through the rest of his life, Leibniz continued to explore the philosophical, scientific and mathematical questions that interested him from his earliest years. The 1680s and 1690s saw some of his most important writings. In these years, he published his new infinitesimal calculus and a variety of papers outlining his new approach to physics,

particularly his new science of dynamics, the science of force and its laws. The *Brevis demonstratio* of 1686 presents for the first time a refutation of Descartes' conservation law, and hints at the foundations of a more adequate physics. The details are developed in his unpublished *Dynamica* (1690), some material from which is published in the *Specimen dynamicum* in 1695, as well as in the numerous answers to attempted refutations of his argument from tenacious Cartesians. In philosophy, Leibniz published his *Meditationes de cognitione, veritate et ideis (Meditations on knowledge, truth and ideas)* in 1684, and in 1686 composed the *Discours de métaphysique (Discourse on metaphysics)*, eventually published in 1846; the main arguments from the latter are discussed in a series of letters with the Catholic theologian Antoine Arnauld, letters Leibniz contemplated publishing in later years. These same themes are found, somewhat transformed, in two important publications in the 1690s, the *Système nouveau de la nature et de la communication des substances (New system of the nature and the communication of substances)* (1695b) and the *De ipsa natura (On nature itself)* (1698). In the first decades of the next century, Leibniz continued to be very active. Important in these years were the *Nouveaux essais (New essays)* (1704), a close examination of Locke's *Essay Concerning Human Understanding*, abandoned at Locke's death and unpublished until 1765. But he did publish his *Théodicée (Theodicy)* (1710), a compendium of philosophical and theological ideas involving further development of themes that go far back in his thought. His final philosophical works were short summaries, intended only as brief guides to his work, the *Monadologie (Monadology)* and the *Principes de la nature et de la grâce (Principles of nature and grace)*, both of which probably date from 1714.

Throughout these years Leibniz kept up a vast correspondence, including exchanges with Huygens, Johann Bernoulli, Burchardus de Volder and Bartholomaeus Des Bosses, among many others. One exchange is particularly important. Leibniz had been at war with his English counterpart, Sir Isaac Newton, for many years; their rivalry went back to at least the early 1690s, and probably to their first contact in the mid-1670s. The affair was ugly, with accusations of plagiarism regarding the calculus from both sides, and bitter disagreements over the foundations of physics. The rivalry finally resulted, in 1715-16, in a correspondence between Leibniz and Samuel Clarke, the latter standing in for Newton himself (see Clarke, S. §3). The exchange was published by Clarke in 1717.

When his employer Georg Ludwig went to London in 1714 to take the throne of Great Britain, Leibniz did not follow. He was out of favour for his failure to make progress on the history of the House of Hanover, as well as for his generally old-fashioned manner. Furthermore, it is likely that Georg feared that the dispute with Newton and the British intellectual establishment would cause difficulties. Whatever the reason, Leibniz remained in Hanover, where he died on 14 November 1716. Though celebrated in his life and considered a universal genius for the breadth of his interests and activities, in death he was virtually ignored, buried with little ceremony in a grave that was to remain unmarked for many years.

2 The programme

Leibniz never wrote a single work, book or article, that constitutes a canonical exposition of his thought, preferring the short article or letter where he presents his thought from one or another point of view, often in response to the thought of another (Descartes was a favourite target), or in response to questions from a correspondent. Indeed, Leibniz's complex thought seems to resist the kind of comprehensive treatment found in works like Descartes' *Meditations* or Spinoza's *Ethics*. Furthermore, it is only to be expected that Leibniz's beliefs changed over his long career, and from one presentation of his philosophy to another.

Despite its complexity, there are some themes and characteristics that run throughout Leibniz's thought, at least in the mature period that starts after his return from Paris in the late 1670s, the period on which this entry concentrates. (While there was not a radical break from the early years to the later, there is certainly a marked development.) Basic to his thought was his philosophical optimism: this is the best of all possible worlds, freely created by a rational God, who always chooses the best for a good reason, without any arbitrariness. It is because of our limited understanding that we cannot determine a priori all the general or particular features of this world. This conception of God and his creation shaped Leibniz's philosophy: the world is ultimately both rational and in every way perfect. Furthermore, though Leibniz's philosophical intelligence ranged widely, certain problems were particularly important to him. In an untitled note from the late 1680s he wrote: 'there are two labyrinths of the human mind, one concerning the composition of the continuum, and the other concerning the nature of freedom, and they arise from the same source, infinity' (Leibniz 1989: 95). The labyrinth of the composition of the

continuum concerns the ultimate make-up of the world; the labyrinth of freedom concerns how freedom and contingency are possible in the world. The solution to both involves understanding the literally infinite complexity found in the world God created. Leibniz had an opinion about virtually every philosophical and scientific issue of his day, but these two issues consistently drew his attention.

3 God: creation and theodicy

Like many of his contemporaries, Leibniz thought that the existence of God could be proved, and he was particularly attracted by the so-called Ontological Argument, invented by Anselm and revised by Descartes (see God, arguments for the existence of §§2-3). According to the Ontological Argument, as given by Descartes and paraphrased by Leibniz in *Meditations on knowledge, truth and ideas* (([1684a] 1989: 25), 'whatever follows from the idea or definition of anything can be predicated of that thing. Since the most perfect being includes all perfections, among which is existence, existence follows from the idea of God…Therefore existence can be predicated of God'. Leibniz's contribution to the argument is the observation that, as it stands, the argument is not valid: 'from this argument we can conclude only that, if God is possible, then it follows that he exists'. For the argument to work, we must establish the self-consistency of the definition of God. But the consistency of the definition of God follows directly from the fact that God 'is without limits, without negation, and consequently without contradiction' (*Monadology* §45). In addition to this version of the ontological argument, Leibniz also used a cosmological argument for the existence of God, arguing from the existence of contingent things in the world, things whose reason lies outside of themselves, to the existence of a necessary being (*De rerum originatione radicali (On the ultimate origination of things)* (1697); *Monadology* §45). Finally, Leibniz argued from the existence of eternal truths: 'Without [God] there would be nothing real in possibles, and not only would nothing exist, but also nothing would be possible' (*Monadology* §43).

In the opening sections of the *Discourse on Metaphysics* (1686b: §6), Leibniz argued that 'God has chosen the most perfect world, that is, the one which is at the same time the simplest in hypotheses and the richest in phenomena', a formula that recurs often in his writings. While this is the main account of creation, in other texts, particularly the essay *On the ultimate origination of things*, he argued that 'there is a certain urge for existence or (so to speak) a straining toward existence in possible things or in possibility or essence itself; in a word, essence in and of itself strives for existence' Leibniz [1697] 1989: 150). Leibniz continued: 'From this it is obvious that of the infinite combinations of possibilities and possible series, the one that exists is the one through which the most essence or possibility is brought into existence'. Such an account of creation has the apparent implication that God is not necessary for it, and that creation results from a quasi-mechanistic weighing of possibilities with respect to one another. But Leibniz emphasized that God is the ground of all possibles, and that it is God who ultimately actualizes the possibles that 'win' the 'contest'. The 'striving possibles' account of creation would seem to be a metaphorical way of expressing Leibniz's usual account in terms of God's choice of the best of all possible worlds.

Leibniz's account of creation had a number of important implications. First, against Descartes and Spinoza, it entailed that there is a standard of goodness and perfection that exists independently of God; God creates the world because it is good, a world which is good not just because it is the creation of God (*Discourse* §2). Furthermore, unlike Malebranche, Leibniz held that the world could not have been created better than it is (*Discourse* §§3-4). Leibniz's doctrine of creation can also be read as a direct attack against a conception of God argued by Spinoza. Central to Spinoza's enterprise in the *Ethics* is an attack on the view that God is like us, that he has aims and goals, that he chooses things for a reason, and that he is bound by standards of goodness that exist independently of his will. This anthropomorphic view of God, Spinoza argued, is an illusion, a projection of our own nature onto nature at large (see Spinoza, B. de §4). Against Spinoza, Leibniz presented his own God, who deliberately chooses to create this world for a particular reason, because it is the best of all possible worlds, a reason intelligible to us. It is on this basis that Leibniz argued against both Descartes and Spinoza for the importance of final causes in nature.

Leibniz's account of creation also addressed the problem of understanding divine justice, in particular, how sin, evil and suffering are possible in a world created by God - the 'theodicy' problem, to use the word coined by Leibniz. His answer was complex, filling many pages in *Theodicy*, the only philosophical book he published in his lifetime. Briefly, his argument was that evil is a necessary and unavoidable consequence of God's having chosen to create the best of all possible worlds. However bad we might think things are in our world, they would be worse

in any other.

Leibniz’s account of creation was closely connected with a number of his key principles, most prominently the Principle of Sufficient Reason. As he wrote later in the *Monadology* (§53), ‘since there is an infinity of possible universes in God’s ideas, and since only one of them can exist, there must be a sufficient reason for God’s choice, a reason which determines Him towards one thing rather than another’. The Principle of Sufficient Reason entails that the universe is in principle rational and intelligible: God must always act for a reason, and as a consequence, there must be a reason for everything. But the account of creation was also connected with a number of other principles in Leibniz’s philosophy (discussed below). In this way one can say that the doctrine of creation underlies all of Leibniz’s philosophy. Had we God’s intellect, we would be able to derive all of the features of this world directly from its being the best of all possible worlds. As it is, our understanding of God’s creation will enable us to fix certain general truths about this world, and set certain bounds on our hypotheses about the way things are.

Leibniz’s interest in philosophical theology was not just the interest of a philosopher. He believed that his understanding of truths about God and nature would greatly assist the undertaking of uniting the Catholic and Protestant Churches under the umbrella of the true philosophy.

4 Metaphysics: substance, monad and the problem of the continuum

Leibniz is famous for his claim that he solved the problem of the composition of the continuum. In so far as the continuum (length, area, volume) is divisible, it would seem to be made up out of parts. But what parts could make it up? If the parts are extended (like atoms), then they too are divisible, and we require an account of their composition as well. On the other hand, if the parts are non-extended (like points), then it is difficult to see how they could make up an extended magnitude. Leibniz’s solution was this: the mathematical continuum should not be thought of as being *composed* of parts at all; while it has parts, those parts are the result of the division of the whole, and thus are posterior to it. On the other hand, Leibniz claimed, while real physical extensions have parts, there are no physical continua. Physical extended things are at root discrete multitudes whose constituents are substances (‘Remarques sur les Objections de M. Foucher’ (Remarks on the objections of M. Foucher) (1696); (Leibniz to de Volder, 19 January 1706). This raises one of the central problems for Leibniz’s philosophy: what are these substances that constitute the metaphysically ultimate constituents of the world?

While there are many paths into his views on substance, Leibniz’s critique of Descartes’ notion of corporeal substance is a convenient starting place. Descartes held that the essence of body is extension. What this meant is that bodies are geometrical objects made concrete, entities that have no properties that are not grounded in extension. Colour, taste, sound and so on are not themselves in bodies, but are only sensations in minds caused by our interaction with extended substances. While Leibniz as a mechanist agreed with this last claim, he rejected the Cartesian conception of body on which it is based (see Descartes, R. §§8, 11).

Leibniz offered a number of arguments against the Cartesian conception of bodily substance: (1) The notion of extension presupposes some quality that is extended, like whiteness in milk or resistance to new motion in every body, and so is not the kind of thing that by itself could constitute the essence of anything (Leibniz to de Volder, 30 June 1704; ‘Note on Cartesian natural philosophy’, 1702). (2) In so far as extended things are divisible, they are aggregates made up out of parts. But the reality of the aggregate presupposes some genuine individuals of which the aggregate is composed; no such individuals can be found in Cartesian bodies (Leibniz to Arnauld, 30 April 1687; *Monadology* §§1-3). (3) If the world is full and there are no vacua, and if the world is filled with Cartesian extended substance, then there can be no change in the world. For any supposed change would consist in one portion of body replacing another, identical in every way (‘On nature itself’). (4) If body were just extension, then it would be perfectly inert, and would have to be moved by God. If so, then God’s creation would be imperfect for lacking creatures which cannot themselves carry out any of God’s commands. Indeed, such a world would reduce to Spinoza’s world in which finite things are just modes of God (‘On nature itself’). Because of arguments like these, Leibniz wanted to take the Cartesian mechanist analysis of body back one step further, and resolve even extension into something more basic still, a world of substances that are genuinely individual, genuinely active, and which contain properties that distinguish individual substances from one another.

While there are a number of important discussions of the nature of substance in Leibniz’s writings, two are

especially noteworthy: the one he gave in the *Discourse on Metaphysics* at the start of his mature period, and the one he gave at the very end of his life in the *Monadology*. (There is a third important conception of substance that arises in the dynamical writings, discussed below in connection with his physics.)

Leibniz begins Section 8 of the *Discourse on Metaphysics* by noting that ‘it is evident that all true predication has some basis in the nature of things, and that, when a proposition is not an identity, that is, when the predicate is not explicitly contained in the subject, it must be contained in it virtually’. (This principle, which probably derived from Leibniz’s logical studies a few years earlier, was closely connected with the Principle of Sufficient Reason in Leibniz’s mind; the containment of the concept of the predicate in the concept of the subject constitutes the ‘sufficient reason’ for the truth of a proposition. This connection with his logic has caused some commentators to see Leibniz’s metaphysics as fundamentally logical in its inspiration.) And so, Leibniz claims, ‘the subject term must always contain the predicate term, so that one who understands perfectly the notion of the subject would also know that the predicate belongs to it’. He concludes that ‘the nature of an individual substance or of a complete being is to have a notion so complete that it is sufficient to contain and to allow us to deduce from it all the predicates of the subject to which this notion is to be attributed’. Since he held that there must be something in the substance itself in virtue of which this complete notion holds of it, he also concludes that at any given time, a substance must contain marks and traces of everything that is true of it, past, present and future - though only God could see them all. (It is not clear whether this committed Leibniz to holding that all properties of a given individual are essential to that individual, making him a kind of ‘superessentialist’, or whether he takes the weaker position that they are merely internal to the individual, making him a ‘superintrinsicalist’. Opinions differ among the commentators.)

In the *Monadology* Leibniz offers a somewhat different characterization of substance. Using the term ‘monad’ that he adopted to express the notion of an individual substance in the late 1690s, he expounds: ‘The monad…is nothing but a simple substance that enters into composites - simple, that is, without parts. And there must be simple substances, since there are composites; for the composite is nothing more than a collection, or aggregate, of simples. But where there are no parts, neither extension, nor shape, nor divisibility is possible. These monads are the true atoms of nature and, in brief, the elements of things’ (*Monadology* §§1-3). So understood, the Leibnizian world is grounded in non-extended simple substances, whose principal property is non-divisibility and thus, Leibniz inferred, non-extension.

From these basic characterizations of the individual or simple substance (what Leibniz called a ‘monad’ after the mid-1690s), he inferred a number of important properties. The individual substance or monad is a genuine unity that cannot be split, something explicit in the *Monadology* account, less so on the earlier account in the *Discourse*. Consequently, it can begin only by divine creation, and can end only with divine annihilation; it is naturally ungenerable and incorruptible. On both accounts, individual substances or monads are the sources of all their activity, and cannot be altered or changed by the direct action of others; it is in this sense that Leibniz said that ‘monads have no windows through which something can enter or leave’ (*Monadology* §7). In the *Discourse* he derives this from the fact that a substance contains within itself all of the grounds of all its properties; there is no need - and no room - for any external causality. In the *Monadology* it is derived directly from the fact that monads are non-extended. The apparent action of one substance on another must be analysed in terms of the relations between the internal states of the one and the internal states of the other (as discussed below). Finally, because of the relations that hold between one substance and another, Leibniz argued that each individual substance or monad reflects the entire world of which it is a part, a thesis closely connected with the hypothesis of pre-established harmony (also discussed below). Though all the individual substances reflect the same one world, they each reflect it from a different point of view, adding the perfection of variety to God’s creation (*Discourse* §§9, 15; *Monadology* §§4-7). This conception of harmony can be traced back to the Paris period and, perhaps, to Leibniz’s earliest writings on physics.

On Leibniz’s view, substances are distinguished from one another by their momentary perceptions, and by the appetitions, the internal source of a substance’s activity that lead from one perceptual state to another. In so far as a substance has such appetitions, ‘the present is pregnant with the future’ (*Monadology* §22). Since there can be no external influences, each monad is created by God with a kind of internal programme, as it were, which determines all of the states that it will take and the order in which it will take them. Although the Cartesian soul is an important model for the individual substance (Leibniz to de Volder, *c.*1699), there are significant differences.

While the momentary states are called perceptions, not all such perceptions are conscious. (Conscious perceptions are said to be 'apperceptions' in Leibniz's terminology, though because nature makes no leaps in this best of all possible worlds, there must be a continuous gradation between the unconscious and the conscious.) In scholastic thought, appetition is the general faculty that leads to change in a substance, of which will (or rational appetite) is a special case in rational souls. For Leibniz, too, not all appetition is rational. For these reasons, he distinguished carefully between rational souls, like ours, and monads with lesser degrees of consciousness and rationality - what he sometimes calls 'bare monads' (*Monadology* §§8-24).

5 Metaphysics: monad, body and corporeal substance

Much of Leibniz's attention was focused on the level of the individual substance or monad, the atom of nature and the building-block of his world, that which in some sense underlies the world of bodies. But in addition to the simple substances, Leibniz often also recognized complex substances, corporeal substances, particularly in the 1680s and 1690s. Corporeal substances are understood on analogy with the human being, a soul (itself an individual substance) united with an organic body. Leibniz often used Aristotelian language to characterize the corporeal substance, calling the soul its form, and the organic body its matter (see Aristotle §§8, 11). The organic body of a corporeal substance is itself made up of corporeal substances, each of which is a soul united to another, smaller organic body, in a sequence of tinier and tinier organisms that goes to infinity, a manifestation of the infinite variety in this best of all possible worlds that God created. Leibniz distinguished corporeal substances from corporeal aggregates, aggregates of animate corporeal substances whose unity is only mental, imposed by the mind, which perceives a group of substances together. While these corporeal substances are ultimately made up of non-extended individual substances, Leibniz's position (at least before 1704) seems to have been that these corporeal substances, as substances, are the genuine individuals whose reality grounds the aggregates that constitute inanimate bodies.

As discussed below, the soul of a corporeal substance is united to its body by virtue of pre-established harmony. However, by 1704, in response to criticism from René-Joseph de Tournemine, Leibniz came to think that this link does not produce genuine unity, and the notion of a corporeal substance becomes problematic for him. While he continued to assert that the physical world is made up of an infinite hierarchy of organisms, after this date he was not so sure that these organisms constitute genuine substances. (Nevertheless, Leibniz always thought that every monad has a body, and cannot exist without one, even if the monad together with its body does not constitute a genuine substance. Even in death the monad has a body, just a body radically smaller than the one it had had in 'life'.) The problem of constructing complex substances from monads led Leibniz in his correspondence with Des Bosses to explore the idea of a *vinculum substantiale*, or a substantial bond. While it is not clear that he ever really endorsed this idea, he does seem to have taken the problem of corporeal substance seriously in that dialogue.

However the issue of corporeal substance is treated, body had a kind of subordinate status for Leibniz. While corporeal substances may be genuine substances, genuinely individual and genuinely active, and thus genuinely real, they are still grounded in non-extended individual substances or monads. And inanimate bodies are inevitably phenomenal, whether the appearance resulting from a multitude of organic corporeal substances, or simply the appearance presented by an infinite multitude of non-extended substances. In this way, one can see Leibniz's philosophy as an inspiration for the distinction between the noumenal and the phenomenal worlds in Kant's philosophy. But in contrast to Kant, who claimed that we cannot know the noumenal world of the thing-in-itself, Leibniz is quite confident that he knows exactly how things are in themselves: they are monads (see Kant, I. §3).

6 Metaphysics: mind, body and harmony

A basic feature of Leibniz's metaphysics was his doctrine that everything reflects the entire world in which it exists. This harmony among things derives from God at creation, who adjusts the perceptions of individual substances or monads to one another in creating a world more perfect by virtue of its variety. And so, despite the fact that individual substances cannot communicate directly with one another, and thus have no real metaphysical causal relations with one another, yet there is an extended sense in which what happens in one substance can be considered the cause of what happens in another. Leibniz wrote: 'The action of one finite substance on another consists only in the increase of the degree of expression together with the diminution of the expression of the other, insofar as God requires them to accommodate themselves to one another' (*Discourse* §15; compare *Monadology* §52). God, in creating a given substance to perform a particular action at a given time, creates all

other substances in such a way as to reflect that action at that time. This is what might be called *physical* causality, as distinct from *metaphysical* causality which Leibniz denied among finite things.

While every monad or substance is related in some way to every other, there is a special relationship between the mind and the body of a living thing, such as the human being: ‘Although each created monad represents the whole universe, it more distinctly represents the body which is particularly affected by it, and whose entelechy it constitutes. And just as this body expresses the whole universe through the interconnection of all matter in the plenum (that is, space without empty place), the soul also represents the whole universe by representing this body, which belongs to it in a particular way’ (*Monadology* §62; compare *Discourse* §33). In this way, the mind is connected with the world by virtue of the special connection it has with the body; on Leibniz’s understanding of causality, mind and body can be the ‘physical’ causes of changes in one another.

So Leibniz solved to his satisfaction one of the central problems in seventeenth-century metaphysics: the interaction between mind and body. Because of the special harmony between mind and body, just when my body is in the state it would be in if it were pricked by a pin, my mind is programmed to have a sensation of pain. And just when my mind is in the state of willing my arm to raise, my body is in the physical state that would result in the raising of my arm, again not because of any direct causal connection (Leibniz to Arnauld, 28 November/8 December 1686 and 30 April 1687). For that reason, Leibniz wrote: ‘According to this system, bodies act as if there were no souls (though this is impossible); and souls act as if there were no bodies; and both act as if each influenced the other’ (*Monadology* §81). This is what he originally called the hypothesis of concomitance, but called the hypothesis of pre-established harmony when he published it for the first time in the *New system* (1695b).

The view is summarized in an analogy he often used. The mind and the body can be compared to two clocks that keep perfect agreement. One hypothesis to explain their agreement is that of natural influence, the hypothesis that there is some physical connection between the one clock and the other. This corresponds to Descartes’ view of mind-body interaction, where there is real causal influence. The second hypothesis is that someone watches over the two clocks and, by tinkering with them, always keeps them in agreement. This corresponds to the occasionalism of many of Descartes’ followers, in which mind-body causality is mediated by God who causes sensations in the mind on the occasion of an appropriate bodily state, and actions in the body on the occasion of the appropriate volition in the mind (see Occasionalism). Finally there is the hypothesis that the clocks are so well made that they will always remain in perfect agreement with one another. This corresponds to the hypothesis of pre-established harmony, which Leibniz thought to be the most defensible (Leibniz to Basnage de Beauval 3/13 January 1696).

Leibniz offered a number of arguments directly against occasionalism. He argued, for example, that there must be genuine activity in things themselves because a world of genuinely active things is more perfect than a world of things manipulated by God; indeed, Leibniz claimed, a world of inert things is just the Spinozistic world in which God is the only substance of which other things are modes (‘On nature itself’). He also argued that occasionalism posits perpetual miracles, in so far as God is called in to do that which goes beyond the power of things to do by their own nature (Leibniz to Arnauld, 30 April 1687). As noted below, the conception of the physical world that informs Leibniz’s dynamics is itself a direct challenge to occasionalism. Nevertheless, Leibniz did share at least one important doctrine with occasionalism: that finite substances have no real causal relations with one another. This doctrine may strike a modern reader as eccentric, but it would have been rather less so for a seventeenth-century reader.

Leibniz often presented the hypothesis of pre-established harmony as a solution to the problem of mind-body interaction. But, at the same time, it allowed Leibniz to reconcile the mechanistic conception of the world with a conception grounded in final causes. He wrote: ‘The soul follows its own laws and the body also follows its own; and they agree in virtue of the harmony pre-established between all substances.…Souls act according to the laws of final causes, through appetitions, ends, and means. Bodies act according to the laws of efficient causes or of motions. And these two kingdoms, that of efficient causes and that of final causes, are in harmony with each other’ (*Monadology* §§78-9). In more concrete terms, behaviour (raising one’s hand, for example) can be explained either in terms of a volition and the harmony God established between mind and body, or purely in terms of the laws of motion, as applied to the physical body. By pre-established harmony, these two explanations will always agree. In this way Leibniz managed to reconcile the dualism of Descartes with the stricter mechanism

of Hobbes; everything in the body can be explained in purely mechanistic terms, while, at the same time, Leibniz could also hold that human beings (and other living organisms) have souls which are the causes of much of their behaviour.

In addition to explaining the interaction between mind and body, when first introduced, Leibniz held that pre-established harmony also explains the union of mind and body, that which makes a single substance out of a mind and the collection of individual substances that constitutes its body (*Discourse* §33). In this way, pre-established harmony provided a central support for Leibniz's account of corporeal substance. Unfortunately, however, it proved inadequate to the task. In May 1703, René-Joseph de Tournemine pointed out that whatever resemblance one might suppose between two clocks, however justly their relations might be considered perfect, one can never say that the clocks are united just because the movements correspond with perfect symmetry. While it does not challenge pre-established harmony as an account of mind-body interaction, the argument is as simple as it is devastating against the somewhat different claim that pre-established harmony accounts for mind-body unity. In consequence, Leibniz came to question the place of complex corporeal substance in his philosophy, as discussed above (see Malebranche, N. §3; Occasionalism).

7 Metaphysics: necessity, contingency and freedom

Central to Leibniz's philosophy were a variety of problems concerning necessity, contingency and freedom, problems which arise in a variety of ways from a variety of sources. Spinoza stood behind many of Leibniz's worries. According to Spinoza, everything in the world is necessary and nothing is contingent, so that things could not be other than they are. Indeed, everything that is genuinely possible is actual and if something does not actually exist, it is because it could not. Everything follows from the divine nature, not by choice but by blind necessity. Furthermore, Spinoza argued, everything in the world is determined and what we take to be human freedom is just an illusion. We think that we are free because we are ignorant of the causes outside us that determine us to do what we do.

Other problems came from Leibniz's own views. Some came from Leibniz's principle in accordance with which 'when a proposition is not an identity, that is, when the predicate is not explicitly contained in the subject, it must be contained in it virtually' (*Discourse* §8). If every predicate true of an individual was part of its very concept, how could it fail to be necessary? A closely related problem followed from Leibniz's claim that every individual substance contains everything that can happen to it, past, present and future, which seems to entail that everything was determined from the beginning, and there is no room for the freedom of a creature. Here the problem concerns not necessity and contingency, but determinism and human freedom. Even if it were contingent that a certain creature has a certain built-in history, given that history, there does not appear to be room for freedom.

Leibniz offered a number of approaches to this problem in his writings. His basic response to the Spinozistic attacks on contingency is the claim that God *freely* chose the best of all possible worlds. He wrote in the early 1680s in an essay entitled *De libertate (On freedom)* 'God produces the best not by necessity but because he wills it' Leibniz [1680-2] 1989: 20). Yet, since God is perfect, it would seem that his nature necessarily determines his will to choose the best.

This led Leibniz directly to another account of contingency. In that same document, he continued by noting that 'things remain possible, even if God does not choose them'. That is, even if God *necessarily* created the best of all possible worlds (a concession Leibniz does not always make), unactualized possibles are still, in and of themselves, possible. The recognition of such unactualized possibles is what brought him back from the precipice of necessitarianism, so Leibniz wrote in another essay from the late 1680s (Leibniz 1989: 21). Elsewhere, he characterized those possibles that God chooses to create as necessary, but only *ex hypothesi*, on the hypothesis that God chose to create them. Though necessary in this limited sense, they are contingent in so far as their contraries are not self-contradictory (*Discourse* §13).

From time to time Leibniz used the kindred notion of compossibility. Two individuals are said to be compossible when they can be actualized at the same time, and are said not to be compossible when they cannot. In this way one can say that a possible world is a maximal set of compossible individuals. The notions of compossibility and incompossibility are not, however, logical notions, taken narrowly. Two individuals may fail to fit in the same possible world because they are logically in contradiction with one another (in a sense that must be specified), or

because they fail to harmonize with one another.

Leibniz sometimes also suggested that it is contingent that this particular world is the best of all possible worlds. So, even if God necessarily created the best of all possible worlds, it is still contingent that he creates *this* world. These arguments address the worries that derive from Spinoza's view that God necessarily gave rise to this world (see Spinoza, B. de §4). But, as noted above, there are other more Leibnizian worries to address as well. If in any true proposition the concept of the predicate must be contained in the concept of the subject, how can any truth fail to be necessary? Leibniz gave one kind of answer in the *Discourse on Metaphysics* (§13) where he simply asserts that there are two kinds of conceptual containment. While all predicates are contained in the concept of the subject, some are contained necessarily, and some contingently. But in some documents, probably from the late 1680s, he attempted a different solution. He noted first that in some cases we can demonstrate that the predicate is contained in the subject in a finite number of steps. However, in other cases this cannot be done. 'In contingent truths, even though the predicate is in the subject, this can never be demonstrated, nor can a proposition ever be reduced to an equality or to an identity, but the resolution proceeds to infinity' (Leibniz 1989: 96). To demonstrate a contingent truth, one must show that a given individual with a given property is one among an infinity of individuals in a possible world that is the best among an infinity of other possible worlds, something that cannot be shown in a finite number of steps.

Beyond the question of necessity is the issue of human freedom. Take an individual substance, which contains everything that has happened, is happening and will happen to it. Even if one can establish that the sequence of 'happenings' it contains is contingent, yet by virtue of containing all these happenings, it would seem not to be free to do anything other than what it does. Contingency is thus compatible with strict determinism, which is incompatible with human freedom.

Leibniz's solution was that while God may build certain actions into a given individual, he can build them in as *free* actions: 'God sees for all time that there will be a certain Judas whose notion or idea…contains this free and future action' (*Discourse* §30). God does make us with free will, and the ability to choose one thing over another. So, when he chooses to create a given individual with a given life-history, he will include the conditions that will lead that individual to choose one thing over another. But the actual choice is ours, and it is free, Leibniz argued. In this way, 'God inclines our soul without necessitating it' (*Discourse* §30). Furthermore, while we can choose other than the way we do, God in his omniscience can predict what we will actually choose, and build its consequences into our future programme. This divine foreknowledge does not change the character of the events themselves: 'God foresees things as they are and does not change their nature.…Thus they are assured but they are not necessary' (*Dialogue effectif sur la liberté de l'homme et sur l'origine du mal (An actual dialogue on human freedom and on the origin of evil)* [1695c] 1989: 112). Thus Leibniz had no worse problems on this score than does anyone who believes in divine omniscience.

Leibniz's doctrine did raise a knotty problem about the identity conditions for individuals, however. If all properties of a given individual are programmed in from the beginning, then though some may be contingent, and though some may be free, still, they define the individual as the particular individual that it is; were they different, then we would be dealing with another individual altogether, it would seem. From time to time Leibniz acknowledged that we might want to talk about what might have happened if Judas (*our* Judas, the Judas in *this* possible world) had not renounced Christ (*Leibniz-Arnauld Correspondence*, May 1686, (Leibniz 1875-90 vol. 2: 41-2); the specific example at issue there is not Judas, as in the *Discourse*, but Adam). But often Leibniz seemed quite willing to embrace a different view: 'But someone…will say, why is it that this man will assuredly commit this sin? The reply is easy: otherwise he would not be this man' (*Discourse* §30). In this way, given that every substance mirrors the entire world in which it finds itself, Leibniz often committed himself to the thesis that a person can belong to only *one* possible world.

8 Epistemology: ideas and sensation

Despite the fact that Leibniz is usually categorized as a continental rationalist, his main interest was not epistemological. At the same time, he did contribute to the discussions of his day on questions relating to ideas and knowledge.

In the *New Essays* (II.1.1), Leibniz defines an idea as follows: 'an idea is an immediate inner object [which]

expresses the nature or qualities of things'. He emphasizes that we can think that we have an idea when we do not really have one. So, for example, there can be no idea of a fastest motion because the notion is incoherent. But, he notes, 'At first glance we might seem to have the idea of a fastest motion, for we certainly understand what we say; but yet we certainly have no idea of impossible things'. Mistaking our comprehension of the phrase 'fastest motion' for having a genuine idea can lead us into contradiction in this case. But in other cases, for example in mathematics, where we often use symbols without fixing ideas to them, we often must work symbolically because of the complexity of working directly with ideas themselves. In this sense one can have thought and even reasoning when we do not have ideas in the proper sense. This observation is connected with a distinction Leibniz drew between real and nominal definitions. A nominal definition is a definition in which one can doubt whether or not the notion defined is genuinely possible; a real definition is one in which the possibility of the notion defined has been established. One can thus say that it is only of real definitions that one can be sure that they correspond to a genuine idea (*Meditations* [1684a] 1989: 25-6; *Discourse* §24).

Leibniz was a supporter of innate ideas in a number of senses. First of all, he argued that there are certain particular ideas that are innate to the mind, and do not or cannot come through the senses: 'The ideas of *being*, *possible*, and *same* are so thoroughly innate that they enter into all our thoughts and reasoning, and I regard them as essential to our minds' (*New Essays* I.3.3). He made a similar claim for other notions, such as infinity (*New Essays* II.17.3). In this connection he used his celebrated marble analogy in the preface to the *New Essays*. Ideas and truths are in the mind, he argued, just as the shape of Hercules might already be in the veins of a block of marble, making that shape more likely to emerge when the sculptor begins to hammer on it, even though considerable effort may be required to expose the shape: 'This is how ideas and truths are innate in us - as inclinations, dispositions, tendencies, or natural potentialities'.

Leibniz's metaphysics, however, committed him to a stronger position still, that *every* idea is innate, strictly speaking, since nothing can enter a mind from the outside. He wrote: 'The mind always expresses all its future thoughts and already thinks confusedly about everything it will ever think about distinctly. And nothing can be taught to us whose idea we do not already have in our mind' (*Discourse* §26). But even though all ideas are strictly innate, Leibniz could distinguish between the ideas of sensation that in a certain sense come to us from outside, and the ideas that do not and cannot do so. As with the explication of physical causality in the context of a view in which there can be no real metaphysical causality between finite things, Leibniz could say that 'we receive knowledge from the outside by way of the senses, because some external things contain or express more particularly the reasons that determine our soul to certain thoughts' (*Discourse* §27).

Sensations are distinguished from other notions not only by their causal origin (in Leibniz's somewhat extended sense), but also by the fact that they are confused, in contrast to the distinct notions one uses, say, in mathematics. A notion is distinct when one has 'marks and tests sufficient to distinguish a thing from all other similar' things; distinct notions include number, magnitude, shape and so on. A notion is confused 'when I cannot enumerate one by one marks sufficient for differentiating a thing from others, even though the thing does indeed have such marks and requisites into which its notion can be resolved'. In this sense 'colours, smells, tastes, and other particular objects of the senses' are confused (*Meditations* [1684a] 1989: 24). Indeed, they are the confused perception of the geometrical properties of bodies that, on the mechanist programme, ground the perception of sensible qualities. 'When we perceive colours or smells, we certainly have no perception other than that of shapes and of motions, though so very numerous and so very small that our mind cannot distinctly consider each individual one in this, its present state, and thus does not notice that its perception is composed of perceptions of minute shapes and motions alone' (*Meditations* [1684a] 1989: 27). Elsewhere Leibniz used the analogy of a wave to understand this phenomenon. When we hear the roar of the ocean, we are actually hearing just a large number of individual waves, lapping on the shore. But since we cannot distinguish the sounds each individual wave makes, we hear it as an undifferentiated roar. This is just the way the confused perception of the corpuscular microstructure of bodies results in our sensation of colour, taste and so on (*New Essays* 1704: preface). In this way Leibniz rejected the claim that the connection between a particular sensation and its mechanical cause is the result of a perfectly arbitrary divine decree; by the Principle of Sufficient Reason, there can be no such arbitrariness in the world (*New Essays* II.8.13 and following, IV.6.7). Thus, it would seem, the distinction between sensations and ideas of the intellect is not a matter of kind, but a matter of degree, degree of distinctness and confusion.

An important part of this account of sensation was Leibniz's doctrine of *petites perceptions* (minute perceptions).

Like Descartes, Leibniz believed that we think all the time. However, unlike Descartes, he denied that we are always conscious of what we think. He held that ‘at every moment there is in us an infinity of perceptions, unaccompanied by awareness or reflection; that is, of alterations in the soul itself, of which we are unaware because these impressions are either too minute and too numerous, or else too unvarying, so that they are not sufficiently distinctive on their own’ (*New Essays* preface). Though we do not apperceive (that is, consciously perceive) each of them individually, these unconscious perceptions have their effects on us. They are what underlie and explain sensation, as suggested earlier. Furthermore, they also have their effect on the conscious choices that we make (*New Essays* II.20.6).

Finally, Leibniz also had a clear position in the debate then raging in the intellectual world over Malebranche’s view that we see all things in God, that is, that ideas do not exist in finite minds, but only in the mind of God, where they are seen by finite intellects without actually being in them (see Malebranche, N. §2). Leibniz quite clearly rejected Malebranche’s view: ‘Even if we were to see everything in God, it would nevertheless be necessary that we also have our own ideas, that is, not little copies of God’s, as it were, but affections or modifications of our mind corresponding to that very thing we perceived in God’ (*Meditations* [1684a] 1989: 27; compare *Discourse* §29).

9 Epistemology: knowledge and probability

In a famous passage of the *Monadology* (§§31-2) Leibniz writes: ‘Our reasonings are based on *two great principles*, *that of contradiction*, in virtue of which we judge that which involves a contradiction to be false, and that which is opposed or contradictory to the false to be true, and *that of sufficient reason*, by virtue of which we consider that we can find no true or existent fact, no true assertion, without there being a sufficient reason why it is thus and not otherwise’. These two principles correspond to two different kinds of truths, ‘those of *reasoning* and those of *fact*’ *Monadology* §33).

A truth of reason can be known with certainty by a finite demonstration consisting of a finite number of steps containing simple ideas, definitions, axioms and postulates; these truths are necessary and can be known a priori. Sensation can give us particular instances of these truths, but can never attain the kind of universality one finds in necessary truths. As Leibniz wrote in the preface to the *New Essays*: ‘necessary truths, such as we find in pure mathematics and particularly in arithmetic and geometry, must have principles whose proof does not depend on instances nor, consequently, on the testimony of the senses, even though without the senses it would never occur to us to think of them’.

While Leibniz agreed with Descartes that such truths are innate, he distanced himself from Descartes’ appeal to clear and distinct perception. Against those who appeal to Descartes’ axiom that ‘whatever I clearly and distinctly perceive about a thing is true or is assertable of the thing in question’, Leibniz objected that ‘this axiom is useless unless we use criteria for the clear and distinct, criteria which we have made explicit’ (*Meditations* [1684a] 1989: 26-7). While Leibniz agreed with Descartes that we have an innate capacity to recognize these innate truths, as a practical matter, he preferred to constrain the mind by formal rules of logic, unlike Descartes, who rejected formal logic (see §10 below).

Since in all predications, the concept of the predicate is contained in the concept of the subject, all knowledge is in principle a priori; if we only had sufficient knowledge of the subject, we could see everything that is true of it, contained in its complete concept. But this is only possible for God. Humans, incapable of performing the analysis that will reveal the truth a priori must make appeal to the senses in order to discover truths of fact. In fact, Leibniz thought, ‘we are all mere Empirics in three fourths of our actions’ (*Monadology* §28).

Because of the importance of empirical knowledge, Leibniz called for a genuine logic of probability. The modern theory of probability was born in the 1650s with the correspondence between Pascal and Fermat, and then with Christiaan Huygens’ little treatise, *Tractus de ratiociniis in aleae ludo (Treatise on reasoning in games of chance)* (1657). The theory very quickly developed in the seventeenth century, as new practical applications were quickly found. But Leibniz was not satisfied that it had yet been applied to the most general question of all, the kind of reasoning we do about matters of fact on the basis of sensation when demonstration is impossible. And so, in the *New Essays* (IV.2.14) he called for a new science: ‘I maintain that *the study of the degrees of probability* would be very valuable and is still lacking, and that this is a serious shortcoming in our treatises on logic. For when one

cannot absolutely settle a question one could still establish the degree of likelihood on the evidence, and so one can judge rationally which side is the most plausible.…I suspect that the establishment of an *art of estimating likelihoods* would be more useful than a good proportion of our demonstrative sciences, and I have more than once contemplated it'. But even though Leibniz may have contemplated it, he himself never made a serious attempt to develop the logic of probability that he called for here. However, his call was heard by David Hume, who saw his *Treatise* as, in part, answering Leibniz's challenge.

10 Logic and language

From his youth, Leibniz dreamed of constructing a perfect, logical language, 'a certain alphabet of human thoughts that, through the combination of the letters of this alphabet and through the analysis of the words produced from them, all things can both be discovered and judged'. This programme, which Leibniz called the 'universal characteristic', gets its first expression in the very early work, *Dissertatio de arte combinatoria (Dissertation on the art of combinations)* (1666). But it is most fully developed later, from the mid-1670s into the 1680s.

Leibniz's programme had two parts. First, one must assign characteristic numbers to all concepts that show how they are built up out of simpler concepts. Leibniz tried a number of schemes for this, but one strategy was to assign simple concepts prime numbers, and then assign complex concepts the product of the characteristic numbers of its constituent simple concepts. The second part of the programme was then to find simple mechanical rules for the truth of propositions in terms of the characteristic numbers of their constituent concepts. Leibniz's fundamental rule in his Universal Characteristic was the principle discussed above in connection with his metaphysics: a predicate is true of a subject if and only if its concept is contained in the concept of the subject. If the concepts in question can be expressed numerically, then Leibniz thought that the rule can be given a mathematical form as well, and the truth of a proposition could be established by a simple arithmetical calculation. Leibniz's project in these writings was to show how this basic intuition about truth could be extended to propositions that are not in simple subject-predicate form. He also sought to extend the programme to formalize the validity of the standard inferences in Aristotelian logic. Even if he could not assign definite characteristic numbers to particular concepts, Leibniz tried to show that for certain configurations of premises and conclusions, if the premises are true (on his definition of truth), then so too must be the conclusion.

The programme was very ambitious; it if were successful, it would allow the truth or falsity of any proposition, necessary or contingent, to be determined by calculation alone. However, it soon dawned on Leibniz that the idea of finding all the conceptual dependencies necessary to express the contents of notions numerically was utopian in the extreme, particularly given the doctrine of infinite analysis of contingent truths Leibniz came to in the late 1680s. This realization still left in place the more modest programme of validating patterns of inference. But even this more modest programme turned out to be beyond Leibniz's ability to bring to completion, and after the early 1690s he seems to have given up trying to make it work, although he returned to it from time to time.

But even though this particular programme collapsed, the idea of formalism was quite basic to Leibniz's thought. Part of the reaction against the Aristotelian philosophy of the schools was an attack on formal logic. Descartes, Locke and others in the seventeenth century argued that we all have an innate ability to recognize truth, what was often called intuition, and that we should cultivate that capacity, and not waste our time learning formal rules. While Leibniz certainly agreed that we do have the innate capacity to grasp certain truths, he still thought that formalism is very important ((Leibniz to Elisabeth of Bohemia, 1678). Much of our reasoning is 'blind' or symbolic, Leibniz thought, conducted through the manipulation of symbols without having a direct hold on the ideas that underlie the symbols. For that reason we must have clear and unambiguous symbol systems, and strict rules for manipulating them (*Meditations*).

This view is evident in the papers on the Universal Characteristic. But it also underlies another project of the same period, the differential and integral calculus, one of Leibniz's greatest accomplishments, worked out by 1676 and made public from 1684. Though others before him had solved many of the particular problems his calculus could solve, problems relating to tangents, areas, volumes and so on, Leibniz invented a simple notation, still used in the calculus ('d' to represent the operation of differentiation, and ' ' to represent the operation of infinite summation (integration)), and worked out a collection of simple rules for applying these operations to equations of different kinds. In this way, Leibniz was able to produce simple algorithms for solving difficult geometrical problems 'blindly', by manipulating certain symbols in accordance with simple rules.

Another issue closely connected with Leibniz's logic is that of relations. In the *Primae veritates (First truths)* [1689] 1989: 32) Leibniz wrote: '*There are no purely extrinsic denominations* [that is, purely relational properties], denominations which have absolutely no foundation in the very thing denominated.…And consequently, whenever the denomination of a thing is changed, there must be a variation in the thing itself'. In this way, all relations must be, in some sense, grounded in the non-relational properties of things. But it is not clear that Leibniz held that relations had to be reducible to non-relational predicates of things. In one example he gives, he paraphrased 'Paris is the lover of Helen' by the following proposition: 'Paris loves, and by that very fact [*eo ipso*] Helen is loved'. While this certainly relates the relation '*A* loves *B*' to two propositions that have the form of simple subject-predicate propositions ('*A* loves' and '*B* is loved'), it should be noted that the predicates in question ('loves' and 'is loved') would seem to be implicitly relational; whether this is an accidental feature of the example Leibniz chose or a clue to Leibniz's views is a question of some dispute. Furthermore, it is important not to ignore that which connects the two propositions ('and by that very fact'), without which one cannot say that the two non-relational propositions capture the relation '*A* loves *B*' (Leibniz 1966: 14). Other texts suggest that individuals properly speaking have non-relational properties, and that the relations between things are something imposed by the mind onto the world: 'My judgement about relations is that paternity in David is one thing, sonship in Solomon another, but that the relation common to both is a merely mental thing whose basis is the modifications of the individuals' (Leibniz to Des Bosses, 21 April 1714). But in saying that the relations between individuals are 'merely mental', Leibniz does not necessarily mean to dismiss them. He wrote: 'God not only sees individual monads and the modifications of every monad whatsoever, but he also sees their relations, and in this consists the reality of relations and of truth' (Letter to Des Bosses, 5 February 1712).

In addition to formal languages, Leibniz was also keenly interested in the study of natural languages. Like many of his contemporaries, he was interested in the controversies over the question of the Adamic language, the language spoken in Eden and from which all modern languages supposedly derive. This, among other motivations, led him to the empirical study of different languages and the etymology of words (see Universal language).

11 Natural philosophy

Leibniz is read today largely for his philosophical writings. But in his day, he was, if anything, better known for his work in mathematics and natural philosophy. Like many of his contemporaries, Leibniz was a mechanist. Indeed, he was in a sense a much stricter mechanist than the Cartesians. Because of his doctrine of pre-established harmony (see §6 above), one can always give a purely mechanistic explanation of any physical phenomenon, even in humans, unlike in the Cartesian system, where causal interaction between mind and body, direct or occasional, can disrupt the laws governing the body. However, Leibniz's version of the mechanist programme departed significantly from other main versions of the programme of his day, particularly the Cartesian version.

Leibniz rejected the Cartesian analysis of body as extended substance (see §4 above). Instead, he argued that we must go to a deeper level of analysis, behind the extension of bodies to the substances that are the ultimate constituents of reality. Below the level of inanimate extension there are tiny organisms, souls joined to organic bodies which Leibniz, in at least one period of his thought, considered genuine corporeal substances. At a deeper level still there are the non-extended simple substances or monads that ground the reality of corporeal substances. On this view, the extended bodies of the Cartesian world are phenomena, aggregates of substances that are unified by virtue of being confusedly perceived together.

Leibniz also rejected Descartes' central law of nature. For Descartes, God conserves the same quantity of motion in the world, the size times the speed of bodies taken together (see Descartes, R. §11). But Leibniz argued that what is conserved is not bulk times speed, but bulk times the *square* of speed, mv^2, a quantity associated with what he called *vis viva* or living force. To defend this view, he used a cluster of a posteriori arguments which assumed the Galilean law of free-fall (the distance fallen is proportional to the square of the speed acquired in free-fall) together with the Principle of the Equality of Cause and Effect, in accordance with which there is always as much ability to do work in the cause as there is in the full effect. Leibniz showed that, on these assumptions, the Cartesian conservation law entails that the ability to do work can either be gained or lost in certain circumstances, whereas on the assumption of the conservation of mv^2, this does not happen. Leibniz used this strategy in the *Brevis demonstratio (Brief Demonstration of a Notable Error of Descartes)* (1686a), where he first published this result. In addition, he offered an a priori argument in which, arguing from certain abstract notions of motion, action

and effect, together with an intuitive principle of the conservation of effect, he reached the same conclusion (*Discourse* §17; *Dynamics* preliminary specimen). This challenge to Descartes' conservation law elicited numerous responses from the Cartesian community in what came to be called the *vis viva* controversy.

Leibniz saw the replacement of the conservation of the quantity of motion by the conservation of mv^2 as leading us to introduce into the world of physics something over and above the purely geometrical qualities of size, shape and motion that pertain to the extended substance of the Cartesians. This something is what he called force, the new science of which he named dynamics. While force can cause motion and is sometimes manifested in motion, Leibniz carefully distinguished the two. In emphasizing the distinction between force and motion, Leibniz was rejecting not only the Cartesian tradition, but his own early physics where, following Hobbes, he identified force with motion.

Leibniz recognized a variety of different kinds of forces in nature. At the most fundamental level, he distinguished between primitive and derivative forces, and between active and passive forces. Thus, in all, there are four basic kinds of force: primitive and derivative active force, and primitive and derivative passive force. Active force is of two sorts, living force (*vis viva*), which is associated with bodies actually in motion (a ball moving with a definite velocity), and dead force, which is associated with the instantaneous push from which actual motion results, as in gravitation or elasticity. Passive force, on the other hand, is the force that arises in reaction to the active force of another body. It also has two varieties, impenetrability (the force that prevents two bodies from occupying the same place at the same time) and resistance (the force that opposes new motion). The distinction between primitive and derivative force is quite different. Primitive force, active and passive, is the metaphysical ground of activity and passivity, that in a body by virtue of which it is capable of acting (doing work) or resisting. Derivative forces, for Leibniz, were particular states of activity and passivity that exist in a body at a particular time. In this way, primitive force is not a measurable quantity, but something in body that grounds the reality of the derivative forces, which are measurable quantities.

This notion of force was linked directly to Leibniz's notion of corporeal substance: 'Primitive active force, which Aristotle calls first entelechy and one commonly calls the form of a substance, is another natural principle which, together with matter or passive force, completes a corporeal substance' ('Note on Cartesian natural philosophy' [1702] 1989: 252). At least in the 1680s and 1690s, when Leibniz recognized corporeal substances, the primitive forces seem to have been the form and matter of the corporeal substances that ground the reality of the physical world. Derivative forces would then be interpreted as the momentary states of these corporeal substances. The position is somewhat different after Leibniz began to doubt the reality of corporeal substance (see §5 above). Then, he wrote, 'I relegate derivative forces to the phenomena, but I think that it is obvious that primitive forces can be nothing but the internal strivings of simple substances, strivings by means of which they pass from perception to perception in accordance with a certain law of their nature' (Leibniz to de Volder, 1704 or 1705). In this way, the dynamics can be regarded as another perspective on the same entities discussed in Leibniz's more metaphysical writings.

Leibniz held that these forces (or better, the motion that they cause) obey rigorous mathematical laws. These laws include the conservation of living force, mv^2, virtually equivalent to the modern law of the conservation of kinetic energy, and the conservation of bulk times the velocity (a vector quantity), mv, identical to the modern law of the conservation of momentum. (Because Leibniz's conservation of mv involved the directionality of the motion, it is distinct from the Cartesian conservation of quantity of motion, which Leibniz rejected.) While he disagreed with Descartes about the specific contents of the laws, he can be seen as advancing the Cartesian programme of building a physics grounded in mathematically expressible conservation laws. But even though Leibniz's laws are expressible in mathematical terms, they - like the forces that they govern - are grounded in certain metaphysical principles that are imposed on the world by the wisdom of God: 'Although the particular phenomena of nature can be explained mathematically or mechanically…nevertheless the general principles of corporeal nature and of mechanics itself are more metaphysical than geometrical' (*Discourse* §18).

One such general metaphysical principle was noted in connection with the establishment of Leibniz's conservation law, the Principle of the Equality of Cause and Effect. But there were others as well. Leibniz made frequent use of the Principle of Continuity, according to which nothing happens through a leap. Leibniz used this principle to refute Descartes' laws of impact, where small changes in the initial conditions (say the comparative sizes of the

bodies in question, or their motion) can result in radically different results. This principle was also used to refute atomism. If there are perfectly hard atoms, not made up of smaller separable parts, then in collision their motion would change instantaneously at the moment of impact. So, Leibniz concluded, there cannot be any such atoms in nature. Indeed, he used this argument to conclude that every body, no matter how small, is elastic. Leibniz also made appeal to the Principle of Plenitude to argue that there can be no vacuum or empty space in the world, since if God *can* create something consistent with his other creations, he *must* do so. Finally, as seen below, Leibniz used the Principle of Sufficient Reason in connection with his relativistic account of space and time.

The very fact that the world is the product of divine wisdom allowed Leibniz to appeal to final causes in his physics. This differentiates him from both Descartes and Spinoza, both of whom rejected final causes. Leibniz agreed with both that everything in nature can be explained through efficient cause alone - that is, through the laws of motion alone. But often, particularly in optics, it is much easier to solve problems by appealing to God's wisdom, and discovering the way in which a most perfect being would have created his universe (*Discourse* §22; *Specimen of dynamics* 1695a: part I). However, the appeal to final cause only supplements the understanding of nature by efficient causes, and does not replace it. It is another manifestation of divine harmony that the explanations by efficient causes and by final causes always coincides: 'In general we must hold that everything in the world can be explained in two ways: through the kingdom of power, that is, through efficient causes, and through the kingdom of wisdom, that is, through, final causes.…These two kingdoms everywhere interpenetrate each other…so that the greatest obtains in the kingdom of power at the same time as the best in the kingdom of wisdom' ((*Specimen of dynamics* [1695a: part I] 1989: 126-7).

So far we have been discussing Leibniz's work in relation to that of other mechanists, particularly those of the Cartesian school. But it is also important to understand Leibniz's relations with another contemporary and often bitter rival, Isaac Newton.

In opposition to Newton, who held an absolutist conception of place and space, Leibniz argued that space is 'only relations or order or orders of coexistence, both for the actually existing thing and for the possible thing one can put in its place' (*Remarks on Foucher* [1696] 1989: 146). If Newton were right, Leibniz argued, and there was absolute space, then God could create a world in which what is currently east and west are exactly reversed, for example. But if so, by the Principle of Sufficient Reason, then God could have no reason to create one such world over another. Given that he did, he cannot have been faced with such a choice. Leibniz concludes that the two purported Newtonian worlds are really just one world, a world in which space is just constituted by the relations between things (Leibniz to Clarke, 3rd paper §5). Newton's absolutist account of space was supposed to ground an absolutist account of motion as well. For Newton, motion was the change of place of a body with respect to absolute space. Leibniz rejected this too, arguing that motion is a completely relativistic notion, a matter of the relation between bodies over time and that alone (*Specimen of dynamics* part I; Leibniz to Huygens, 12/22 June 1694).

Leibniz also rejected Newton's theory of universal gravitation. He read Newton as holding that gravity is an essential property of matter as such, and he was appalled. For Leibniz, all change in body had to happen through the intermediary of contact and collision; the idea of action at a distance that seemed to underlie Newton's theory of universal gravitation was an intellectual disaster, a treasonable abandonment of the new mechanical philosophy and a return to the worst abuses of the schoolmen. Leibniz, whose early mechanism seemed so radical at the time, could not adjust to the new Newtonian philosophy, soon to take over the intellectual world (see Clarke, S.; Newton, I.).

While the emphasis here has been on the aspects of Leibniz's work in physics most relevant to his philosophical programme, he was much more widely interested in the natural world. He left notes on engineering problems, on chemistry, on geology and on curious observations in natural history including the report of a talking dog, and a goat with an odd hairstyle.

12 Ethics and political thought

Although Leibniz's ethical and political writings are not widely read today, they constitute an important part of his corpus, unsurprising, given Leibniz's own involvement in politics. Leibniz's ethical and political thought, squarely within the natural law tradition, was based on the notions of justice, charity and virtue (see Natural law). Leibniz

wrote: ‘Charity is a universal benevolence, and benevolence the habit of loving or of willing the good. Love then signifies rejoicing in the happiness of another, or, what is the same thing, converting the happiness of another into one’s own’ *Codex Iuris Gentium Diplomaticus (The diplomatic code of the law of nations)* [1693: introduction] 1988: 171). In a note on felicity (Leibniz [*c*.1694-8] 1988: 83-4), he connected justice, wisdom, and virtue to charity: ‘Virtue is the habit of acting according to wisdom.…Wisdom is the science of felicity, [and] is what must be studied above all things.…To love is to find pleasure in the perfection of another. Justice is charity or a habit of loving conformed to wisdom. Thus when one is inclined to justice, one tries to procure good for everybody, so far as one can, reasonably, but in proportion to the needs and merits of each’.

For Leibniz, human justice is the same as God’s justice, though, of course, less perfect. Leibniz wrote in the *Monita quaedam ad S. Puffendorfii principia (Observations on the Principles of Pufendorf)* ([1706] 1988: 69): ‘In the science of law…it is best to derive human justice, as from a spring, from the divine, to make it complete. Surely the idea of the just, no less than that of the true and the good, relates to God, and above all to God, who is the measure of all things’. Similarly, Leibniz wrote in *Méditation sur la notion commune de la justice (Meditation on the common concept of justice)* ([1702-3] 1988: 60) that ‘as soon as [the concept of justice] is founded on God or on the imitation of God, it becomes universal justice, and contains all the virtues’.

In so far as charity is defined in terms of universal love and benevolence, justice is something quite distinct from power. This is true even for God. ‘Justice, indeed, would not be an essential attribute of God, if He himself established justice and law by His free will’. In this sense, God is as bound by the eternal laws of justice as he is bound by truths of reason: ‘Justice follows certain rules of equality and of proportion [which are] no less founded in the immutable nature of things, and in the divine ideas, than are the principles of arithmetic and of geometry’ (*Observations on Pufendorf* [1706] 1988: 69). (Here, perhaps is the origin of the theodicy problem for Leibniz: if God is bound by the same ideal of justice that binds us, then we must show how the works of the all-perfect creator can be seen to conform to that ideal.) So, too, are we bound by a standard of justice that exists independently of our wills.

Leibniz recognized three degrees of justice. The lowest, a minimal sort of justice, is simply not to harm others. The second degree is to give each their due, what it is that is owed to them. The highest, though, is to behave with genuine beneficence toward others, and to do that which will promote their happiness; this is what Leibniz calls piety (Leibniz to Coste, 4 July 1706: appendix).

Leibniz’s conception of justice as the charity of the wise also placed virtue and obligation outside of the scope of a contract. For Hobbes, for example, the notion of justice arises from a contract that we make with one another in forming a society, and the notion of justice has no applicability outside that framework. Commenting on Shaftesbury in 1712, Leibniz wrote: ‘Our illustrious author refutes with reason…those who believe that there is no obligation at all in the state of nature, and outside government; for obligations by pacts having to form the right of government itself, according to the author of these principles, it is manifest that the obligation is anterior to the government which it must form’ (Leibniz 1988: 196). Indeed, he noted, there are societies, among the native Americans for example, in which the sovereign thought necessary by Hobbes is altogether absent: ‘entire peoples can be without magistrates and without quarrels, and…as a result men are neither taken far enough by their natural goodness nor forced by their wickedness to provide themselves with a government and to renounce their liberty’. In people sufficiently wise, then, justice and charity are sufficient to hold society together, without the need of a contract.

But Leibniz was a practical politician, as well as a theorist of politics. He generally worked for a Europe unified under the leadership of a unified church, a Christian Europe in which there are no conflicts between different Christian states. This, in part, is what was behind his plan for the reunification of the Catholics and the Protestants. It was also behind his attempt, as early as 1671, to persuade the French to attack Egypt, a non-Christian country, rather than to invade the Netherlands. In practice, however, Leibniz was an opponent of French expansionism under Louis XIV (as much as he was an admirer of French culture), and a supporter of a union of Protestant countries in Northern Europe (his *Mars Christianissimus* (1684b) was a brilliant satire directed against Louis XIV’s foreign policy). He was also an active participant in the successful campaign in support of the claim of the House of Hanover for the throne of England.

13 The Leibnizian tradition

It is important to remember when considering Leibniz's influence that much of what we now know of Leibniz's writings was unknown to his readers for many years after his death. The full dimensions of Leibniz's thought emerged only slowly, as new texts came to light. Indeed, there is still no complete edition of his work.

At the time of his death, and in the decade afterwards, only a small selection of Leibniz's texts was available. There were a fair number of publications in mathematics and physics, some legal writings and some documents collected in connection with his unfinished history of the house of Hanover. In philosophy, however, there were only a few essays. During his lifetime, Leibniz had published *Meditations on Knowledge, Truth, and Ideas* (1684a), the *New System* (1695b), *On Nature Itself* (1698) and the *Theodicy* (1710). The Leibniz-Clarke correspondence was published soon after his death, and a Latin version of the *Monadology* appeared in 1721. On the other hand, the *New Essays* did not appear until 1765, and works that we now consider central, such as the *Discourse on metaphysics*, did not appear until 1846. Many of his philosophical writings and correspondence had to await the monumental edition of C.I. Gerhardt, which appeared between 1875 and 1890. Many texts have yet to appear.

Despite the relative paucity of his available writings, Leibniz was much read and debated in the eighteenth century. One of his early supporters was the German professor Christian Wolff who had corresponded with Leibniz during his life. He composed numerous volumes expounding a Leibnizian philosophy in an ordered and orderly way. Wolff's systematic philosophy made it ideal for the academy, and his ideas were widely influential. But there were opponents, particularly a group of pietist theologians at the University of Halle, but others as well, including Maupertuis, Crusius, Condillac and, most famously, Voltaire, who made Leibniz into the comical Dr Pangloss of his *Candide*. Kant received his philosophical education in the atmosphere of this debate between the Leibnizians and the anti-Leibnizians in the German intellectual world. His philosophy, both pre-critical and critical, shows the marks of his knowledge of Leibniz's writings.

See also: Atomism, ancient; Contingency; Freedom, divine; Infinity; Mendelssohn, M.; Probability theory and epistemology; Substance; Will, the

DANIEL GARBER

List of works

Many of Leibniz's writings, including some of the most important of them, remained unpublished during his lifetime. As a consequence, some of them can be dated only approximately.

Leibniz, G.W. (1768) *Leibnitii opera omnia (The complete works of Leibniz)*, ed. L. Dutens, Geneva, 6 vols; repr. Hildesheim: Olms, 1989.(Contains a wide range of Leibniz's papers, both inside and outside philosophy, many of which have not been reprinted since the eighteenth century. It gives the best sense of what Leibniz meant to his contemporaries.)

Leibniz, G.W. (1849-63) *Mathematische Schriften (Mathematical writings)*, ed. C.I. Gerhardt, Berlin and Halle: A. Asher & comp. and H.W. Schmidt, 7 vols; repr. Hildesheim: Olms, 1962.(Still the most complete collection of Leibniz's papers and letters in mathematics and physics, in the original languages, many of which are directly connected to his more philosophical interests.)

Leibniz, G.W. (1875-90) *Die philosophischen Schriften (Philosophical writings)*, ed. C.I. Gerhardt, Berlin: Weidmannsche Buchhandlung, 7 vols; repr. Hildesheim: Olms, 1978.(Still the most complete collection of Leibniz's philosophical papers and letters in the original languages.)

Leibniz, G.W. (1923-) *Sämtliche Schriften und Briefe (Collected writings and letters)*, ed. Deutsche Akademie der Wissenschaften (before 1945, Preussische), Berlin: Akademie Verlag.(This is to be the new complete critical edition of Leibniz's writings in the original languages, edited to the highest standards. Currently still in its early stages, it must be supplemented by earlier editions. In recent years it has been supplemented by a 'Vorausedition', giving pre-prints of editorial work in progress.)

Leibniz, G.W. (1666) *Dissertatio de arte combinatoria (Dissertation on the art of combinations)*, Leipzig; repr. in *Die philosophischen Schriften*, vol. 4; partial trans. L.E. Loemker in *Philosophical Papers and Letters*, Dordrecht: Reidel, 1969. (Leibniz's first important publication, concerning the theory of mathematical combinations, together with various philosophical digressions. It also contains a suggestion of his later concern with a universal language.)

Leibniz, G.W. (*c*.1680-2) *De libertate (On freedom)*, in *Textes inédites*, vol. 1, ed. G. Grua, Paris: Presses Universitaires de France, 1948; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis,

IN and Cambridge, MA: Hackett Publishing Company, 1989.(Interesting essay on human and divine freedom and contingency, unpublished in Leibniz's lifetime.)

Leibniz, G.W. (1684a) 'Meditationes de cognitione, veritate et ideis' (Meditations on knowledge, truth and ideas), *Acta Eruditorum* (November 1684): 537-42; repr. in *Leibnitii opera omnia*, vol. 4; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989. (Important explanation of Leibniz's views on concepts, truth and knowledge.)

Leibniz, G.W. (1684b) *Mars Christianissimus (Most Christian War-God)*, Cologne; trans. P. Riley in *Leibniz: Political Writings*, Cambridge: Cambridge University Press, 1988. (A satire on the diplomatic policies of Louis XIV.)

Leibniz, G.W. (1686a) 'Brevis Demonstratio erroris memorabilis Cartesii' (Brief demonstration of a notable error of Descartes), *Acta Eruditorum* (March 1686): 161-3; repr. in *Mathematische Schriften*, vol. 6; trans. L.E. Loemker in *Philosophical Papers and Letters*, Dordrecht: Reidel, 1969.(A refutation of Descartes' law of the conservation of quantity of motion in physics.)

Leibniz, G.W. (1686b) *Discours de métaphysique (Discourse on metaphysics)*, 1846; ed. H. Lestienne, Paris: Alcan, 1907; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(Unpublished in Leibniz's time, this is a central text in which Leibniz gives an exposition of some central elements of his metaphysics as of 1686.)

Leibniz, G.W. (1686-90) *The Leibniz-Arnauld Correspondence; repr. with related letters and documents in Die philosophischen Schriften*, vol. 2: Weidmannsche Buchhandlung, 1875-90, 1956; trans. and ed. H.T. Mason, Manchester: Manchester University Press, 1967.(Very important series of letters exchanged between Leibniz and the Cartesian philosopher Antoine Arnauld just as Leibniz was setting out his mature philosophy, providing a kind of commentary on themes developed in the *Discourse*. Although unpublished in his lifetime, Leibniz probably intended it for publication.)

Leibniz, G.W. (*c.*1689a) *De libertate (On Freedom)*, in *Textes inédites*, vol. 1, ed. G. Grua, Paris: Presses Universitaires de France, 1948; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(Another essay on contingency and freedom, unpublished in Leibniz's lifetime, which makes use of his infinite analysis account of contingency.)

Leibniz, G.W. (1689b) *Primae veritates (First truths)*; repr. in *Opuscules et fragments inédits*, 1903; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989. (An important, unpublished summary of Leibniz's metaphysics as of the 1680s. Originally thought to have preceded the 1686 composition of the *Discourse*, it is now firmly dated at 1689.)

Leibniz, G.W. (1690) *Dynamica (Dynamics)*, in *Mathematische Schriften*, vol. 6; partial trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(Unpublished in Leibniz's time, this is a systematic exposition of Leibniz's physics. Apart from the translation of the preliminary discourse in Ariew and Garber, the bulk of this work has not been translated from the original Latin.)

Leibniz, G.W. (1692) *Protogaea*, 1749; French trans, ed. B. de Saint-Germain and J.-M. Barrande, Toulouse: Presses Universitaires du Mirail, 1993.(A treatise on geology and the early days of the earth, written as the first part of Leibniz's history of the House of Hanover.)

Leibniz, G.W. (1693) *Codex Iuris Gentium Diplomaticus (The diplomatic code of the law of nations)*, Hanover; partial trans. P. Riley in *Leibniz: Political Writings*, Cambridge: Cambridge University Press, 1988.(Collection of diplomatic papers, with a long introduction on political philosophy.)

Leibniz, G.W. (*c.*1694-8) *La félicité (Felicity)*; repr. in *Textes inédites*, vol. 1; trans. P. Riley in *Leibniz: Political Writings*, Cambridge: Cambridge University Press, 1988. (Important but unpublished note from the 1690s in which Leibniz discusses his theory of justice as the charity of the wise man.)

Leibniz, G.W. (1695a) *Specimen dynamicum (A specimen of dynamics)*, Part I, *Acta Eruditorum* (April 1695); repr. (Parts I and II) in *Mathematische Schriften*, vol. 4; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(Important work that links the technical physics of the *Dynamica* with more philosophical themes. Only Part I was published during Leibniz's life.)

Leibniz, G.W. (1695b) 'Système nouveau de la nature et de la communication des substances' (New system of the nature and the communication of substances), *Journal des Sçavans* (27 June 1695): 294-300; repr. in *Die philosophischen Schriften*, vol. 4; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(A popular presentation of Leibniz's metaphysics, featuring the first public presentation of the hypothesis of pre-established harmony.)

Leibniz, G.W. (1695c) *Dialogue effectif sur la liberté de l'homme et sur l'origine du mal (An actual dialogue on*

human freedom and on the origin of evil); repr. in *Textes inédites*, vol. 1; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(This seems to be a record of a dialogue that actually took place between Leibniz and Baron Dobrzensky, counsellor of state and war of Brandenburg.)

Leibniz, G.W. (1696) 'Remarques sur les Objections de M. Foucher' (Remarks on the objections of M. Foucher), *Histoire des ouvrages des Savans* (February 1696): 274-6; repr. in *Die philosophischen Schriften*, vol. 4; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(An excellent but brief account of Leibniz's view on the problem of the continuum.)

Leibniz, G.W. (1697) *De rerum originatione radicali (On the ultimate origination of things)*; repr. in *Die philosophischen Schriften*, vol. 7; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(This important brief essay on creation and contingency remained unpublished until the nineteenth century.)

Leibniz, G.W. (1698) 'De ipsa natura' (On nature itself), *Acta Eruditorum* (September 1698): 427-40; repr. in *Die philosophischen Schriften*, vol. 4; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(Significant essay on the importance of introducing genuinely active individuals into the world, against the Cartesian position that bodies are bare, extended substances.)

Leibniz, G.W. (1702) 'Note on Cartesian natural philosophy'; repr. in *Die philosophischen Schriften*, vol. 4; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(Untitled by Leibniz and not published in his lifetime, this is an important summary of the philosophical aspects of Leibniz's dynamics.)

Leibniz, G.W. (1702-3) *Méditation sur la notion commune de la justice (Meditation on the common concept of justice)*; trans. P. Riley in *Leibniz: Political Writings*, Cambridge: Cambridge University Press, 1988. (Unpublished in Leibniz's lifetime, this is an important source for understanding his political philosophy.)

Leibniz, G.W. (1704) *Nouveau essais sur l'entendement humain (New essays in human understanding)*, 1765; repr. in *Sämtliche Schriften und Briefe*, series 6, vol. 6; trans. P. Remnant and J. Bennett as *New Essays on Human Understanding*, Cambridge: Cambridge University Press, 1981.(Unpublished in Leibniz's lifetime, this is a point-by-point discussion of Locke's *Essay Concerning Human Understanding*.)

Leibniz, G.W. (1706) *Monita quaedam ad S. Pufendorfii principia (Observations on the Principles of Pufendorf)*; repr. in *Leibnitii opera omnia*, vol. 4, part 3; trans. P. Riley in *Leibniz: Political Writings*, Cambridge: Cambridge University Press, 1988. (Unpublished in his lifetime, these are Leibniz's comments on the political thought of Samuel Pufendorf, the seventeenth-century jurist.)

Leibniz, G.W. (1710) *Essais de Théodicée (Essays on Theodicy)*, Amsterdam; trans. E.M. Huggard as *Theodicy*, La Salle, IL: Open Court, 1985.(A treatise in which Leibniz attempts to justify the ways of God to man. Much of the work is a response to the writings of Pierre Bayle.)

Leibniz, G.W. (1712) *Jugement sur les oeuvres de Mylord Shaftesbury (Judgment of the works of the Earl of Shaftesbury)*; repr. in *Leibnitii opera omnia*, vol. 5; trans. P. Riley in *Leibniz: Political Writings*, Cambridge: Cambridge University Press, 1988. (Unpublished in Leibniz's lifetime, this is a free-ranging discussion of Shaftesbury's work that includes much of interest for Leibniz's political thought.)

Leibniz, G.W. (1714a) *Monadologie*, 1721; repr. in *Die philosophischen Schriften*, vol. 6; critical edn, ed. A. Robinet, *Principes de la nature et de la grâce…et Principes de la philosophie ou monadologie*, Paris: PUF, 1986; trans. R. Ariew and D. Garber as *Monadology* in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(Central text, not published during his lifetime, in which Leibniz gives a summary of his metaphysics at the end of his life.)

Leibniz, G.W. (1714b) 'Principes de la nature et de la grâce' (Principles of nature and grace), *Acta Eruditorum* (September 1698): 427-40; repr. in *Die philosophischen Schriften*, vol. 6; critical edn, ed. A. Robinet, *Principes de la nature et de la grâce…et Principes de la philosophie ou monadologie*, Paris: PUF, 1986; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(Another late summary of Leibniz's metaphysics, a companion to the *Monadology*, also not published during his lifetime.)

Leibniz, G.W. (1714c) 'De ipsa natura' (On nature itself), *Acta Eruditorum* (September 1698): 427-40; repr. in *Die philosophischen Schriften*, vol. 4; trans. R. Ariew and D. Garber in *Leibniz: Philosophical Essays*, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company, 1989.(Significant essay on the importance of introducing genuinely active individuals into the world, against the Cartesian position that bodies are bare, extended substances.)

Leibniz, G.W. (1715-16) *The Leibniz-Clarke Correspondence*, 1717; trans. and ed. H.G. Alexander, Manchester: Manchester University Press, 1956.(The Alexander edition has a useful introduction and notes to the translation of the exchange.)

Leibniz, G.W. (1903) *Opuscules et fragments inédits (Unpublished short works and fragments)*, ed. L. Couturat, Paris: Alcan; repr. Hildesheim: Olms, 1966.(The first publication of many of Leibniz's papers on logic, language, and related areas of metaphysics, published in the original languages; this collection shaped earlier twentieth-century views of Leibniz's programme as driven by his logic.)

Leibniz, G.W. (1948) *Textes inédites d'après les manuscrits de la bibliothèque provinciale de Hanovre (Unpublished texts, following the manuscripts in the provincial library at Hanover)*, ed. G. Grua, Paris: Presses Universitaires de France, 2 vols; repr. New York: Garland, 1985.(The first publication, in the original languages, of a valuable selection of writings concentrating on ethical, political and theological subjects.)

Leibniz, G.W. (1966) *Logical Papers*, ed. and trans. G.H.R. Parkinson, Oxford: Oxford University Press.(A collection of Leibniz's logical papers, translated into English, with an extensive introduction.)

Leibniz, G.W. (1969) *Philosophical Papers and Letters*, ed. and trans. L.E. Loemker, Dordrecht: Reidel.(The most extensive collection of Leibniz's writings in English, with a long introduction and useful notes.)

Leibniz, G.W. (1973) *Philosophical Writings*, ed. G.H.R. Parkinson, trans. M. Morris and G.H.R. Parkinson, London: Dent; Totowa, NJ: Rowman & Littlefield.(A useful collection in English translation.)

Leibniz, G.W. (1988) *Leibniz: Political Writings*, ed. and trans. P. Riley, Cambridge: Cambridge University Press, 2nd edn.(An excellent collection of Leibniz's moral and political writings, with a useful introduction.)

Leibniz, G.W. (1989) *Leibniz: Philosophical Essays*, ed. and trans. R. Ariew and D. Garber, Indianapolis, IN and Cambridge, MA: Hackett Publishing Company.(A widely available translation of a selection of Leibniz's most important philosophical texts.)

References and further reading

Adams, R.M. (1994) *Leibniz: Determinist, Theist, Idealist*, Oxford: Oxford University Press.(An important recent monograph that focuses on questions of contingency, natural theology, substance and body in Leibniz. Highly recommended for the serious student.)

Aiton, E.J. (1985) *Leibniz: A Biography*, Bristol: Hilger.(A recent biography in English.)

Broad, C.D. (1975) *Leibniz: an Introduction*, Cambridge: Cambridge University Press.(Though somewhat dated, still a good philosophical introduction to Leibniz's thought.)

Couturat, L. (1901) *La logique de Leibniz d'après des documents inédits (The logic of Leibniz from unpublished documents)*, Paris: Alcan; repr. Hildesheim: Olms, 1961.(Still the best study of Leibniz's programme for logic and related areas.)

Duchesneau, F. (1994) La *Dynamique de Leibniz (Leibniz's Dynamics)*, Paris: Vrin. (A good recent study of Leibniz's programme for physics.)

Frankfurt, H. (ed.) (1972) *Leibniz: A Collection of Critical Essays*, New York: Doubleday Anchor.(Contains many classic essays, including those of Russell and Couturat.)

Gueroult, M. (1967) *Leibniz: dynamique et métaphysique (Leibniz: Dynamics and Metaphysics)*, Paris: Aubier.(A classic study of the connection between Leibniz's dynamics and his metaphysics.)

Ishiguro, H. (1972) *Leibniz's Philosophy of Logic and Language*, Ithaca, NY: Cornell University Press, 2nd edn, 1990.(A good study focusing on questions relating to logic and language.)

Jolley, N. (ed.) (1995) *The Cambridge Companion to Leibniz*, Cambridge: Cambridge University Press.(A recent collection of articles surveying the various aspects of Leibniz's thought, but focusing on his philosophy.)

Mercer, C. (1998, forthcoming) *Leibniz's Metaphysics: Its Origins and Development*, Cambridge: Cambridge University Press.(A lively study of Leibniz's early philosophy and how it evolved into the mature thought.)

Müller, K. and Krönert, G. (1969) *Leben und Werk von G.W. Leibniz. Eine Chronik (The life and work of G.W. Leibniz: a chronology)*, Frankfurt: Klostermann.(An exhaustive summary of what is known about Leibniz's life, whereabouts, and when he was working on what, with documentation, arranged chronologically.)

Ravier, E. (1937) *Bibliographie des Oeuvres de Leibniz (Bibliography of the works of Leibniz)*, Paris: Alcan; repr. Hildesheim: Olms, 1966.(Despite some inaccuracies, the best guide to the publication of Leibniz's writings, from his lifetime to the 1930s.)

Russell, B. (1937) *A Critical Exposition of the Philosophy of Leibniz*, London: Allen & Unwin, 2nd edn. (Advances the view that Leibniz's metaphysics is grounded in his formal logic. While the main thesis is now generally rejected, it was highly influential, and the book contains many still-valuable discussions.)

Rutherford, D. (1995) *Leibniz and the Rational Order of Nature*, Cambridge and New York: Cambridge University Press.(Ascribes the systematic unity of Leibniz's thought to his vision of the best of all possible

worlds.)

Sleigh, R.C. (1990) *Leibniz & Arnauld. A Commentary on their Correspondence*, New Haven, CT, and London: Yale University Press.(While it focuses on what Leibniz was thinking in the crucial mid-1680s, this is also an excellent commentary on some of the most important philosophical themes in Leibniz's thought. Highly recommended for the serious student.)

Studia Leibnitiana (1969-), Weisbaden: Steiner Verlag.(A journal that focuses on studies of Leibniz and his age. In addition to its regular issues, it also publishes numerous supplementary volumes containing collections of essays, conference proceedings, and short monographs that pertain to Leibniz and related issues in the history of philosophy.)

Voltaire, F.M. de (1759) *Candide, ou l'optimisme (Candide, or optimism)*, Paris.(A caricature of Leibniz appears in this popular tale in the person of Dr Pangloss. It is available in numerous modern editions, both in French and in English translation.)

Leibowitz, Yeshayahu (1903-94)

Unlike the major intellectual currents that shaped religious thought in the modern world, Leibowitz's thought is deeply anchored in the Israeli context. Both as philosopher and activist, Leibowitz lived and articulated the paradoxes of modern Israel where he lived and was best known. His reputation as a Socratic gadfly to the establishment reflected his ongoing critique of both Israeli society in the light of Judaism, and Judaism in the light of the revolutionary implications of the creation of the State of Israel.

On the one hand, he was a Jewish patriot, a fighter for Jewish independence from all forms of foreign rule; on the other hand, he was a harsh, relentless critic of national and political expressions of chauvinism in the Israeli establishment. A strictly observant Jew, Leibowitz had less impact on traditional religious Jews than on secular Israelis. His central message is that what makes Jews distinctive as a group is neither their theology nor their Bible, but the system of law with which they regulate their lives. Judaism is a communal concept, and there is no point in religious Jews ignoring the State of Israel, or expecting others to bear their civil burdens for them. Religious law has to be reconciled with life in the political reality of the state, and this necessitates changing those attitudes to the law which reflect the historical conditions of life in exile.

1 The challenge to Judaism

According to Leibowitz, the essential subject of Judaism is not the individual but community. Its primary concern is not with personal sin and finitude, but with community and its commitment to serve God. He argued that the essential factor which distinguishes Jews from other nations is not theology or the Bible, but *halakhah* Jewish law (see Halakhah). The laws governing what Jews eat, when they work and how they worship, constitute the uniqueness of the Jewish nation. The primacy of *halakhah* in Judaism is less a value judgment than an empirical fact of history. The relative tolerance of Jewish communities for a variety of theological opinions stands in sharp contrast to their intolerance to forms of practice that deviate from *halakhah*. Jews share a distinctive form of religious life, regardless of their divergent or even contradictory conceptions of God.

Jewish life remained broadly characterizable in terms such as these until the beginning of the Emancipation at the end of the eighteenth century. But emancipation led to the growth of movements to modify or abandon parts of *halakhah* while preserving the synagogue as a place of worship. It also led to the emergence of secular Zionism. Leibowitz argues that the secular option, which claims continuity with the Jewish people yet abandons *halakhah* and the worship of God as defining features of Jewish identity is a falsification of the facts of Jewish history.

The question of the relationship between religion and state is connected to the larger theological issue of the religious significance of historical events. To Leibowitz, God cannot be understood in personalistic theistic terms. All attempts at describing God from a human perspective are tantamount to idolatry. In this respect, Leibowitz's position is similar to Maimonides' negative theology (see Maimonides, M.). One might characterize his analysis of religious language as prescriptive rather than descriptive. One cannot talk to or about God; one can only act in the presence of God. Religious language is informed by and must be limited to the worship of God. God is not to be invoked to justify ethical or political judgments; for such usage makes God subservient to human needs. Paradigmatic for Leibowitz's understanding of Judaism is the binding of Isaac, where Abraham is ready to sacrifice his only son (representing history, the self, the human) in unconditional obedience to God's will. The Jewish model for Leibowitz's religious orientation, however, is not the personalistic theistic framework of the Bible, but the religious ethos and communal framework of the talmudic tradition. The revelatory movement of God in history, as represented in the Bible, is replaced by the worshipping movement of the Jews towards God, which the talmudic tradition embodies. Leibowitz distinguished Judaism from Christianity in terms of the latter's promise of personal salvation and liberation from finitude and sin. In Judaism, the world as it is - with the promise of redemption - expresses the will and wisdom of God. All events mirror God's power, just as all nature manifests God's wisdom. No particular event can be singularly endowed with religious meaning, so there is no theology of history. Neither the rebirth of Israel nor the tragic events of the twentieth century have instrinsic religious meaning. Similarly, holiness is not something intrinsic to things, either animate or inanimate. Holiness results solely from human action in the service of God; it is a *halakhic* and not an ontological category. Judaism as represented concretely in *halakhah* is a constant effort and striving to realize God's commands in the world.

2 Zionism and the critique of religious Zionists

Unlike the religious conservatism of the ultra-Orthodox who relegate the legitimacy of Jewish political autonomy to an eschatological future, Leibowitz participated actively in the political struggle to re-establish an autonomous, self-governing Jewish community. While believing that nothing in the past fifteen hundred years could be compared to this heroic decision of Jews to alter their political destiny, he realized that the majority of these Jews were not motivated by a religious purpose.

Leibowitz defined Zionism as the natural expression of a national community seeking to cast off political subjugation and subservience to others. This collective impulse is in itself legitimate and healthy; it has no need of religious justification (see Zionism).

Nevertheless, religious Zionists adopted, or rather, co-opted this nationalist impulse by redefining the State of Israel's natural - albeit remarkable - political renewal in redemptive, Messianic terms. Yet, despite their extension of religious consciousness to include this new political reality, the religious Zionists refused to acknowledge the need to change practices based on *halakhah* which reflected very different historical conditions. How can one participate religiously in statehood, asks Leibowitz, without recognizing the incongruity of practising a *halakhah* wherein Jews are not fully responsible for maintaining the overarching socio-political reality of the state?

Leibowitz presents a Jewish categorical imperative for the religious community: act in such a way that you could wish all Jews to act in a similar fashion. Leibowitz thus rejected the granting of military exemptions to yeshivah students and religious girls. He similarly rejected all other concessions to the 'special interests' of observant Jews.

3 Israel and *halakhah*

Leibowitz is a *halakhic* existentialist who believes that the decision to participate in statehood cannot find legitimization in a framework that mirrors the historical conditions of exile. Unlike *halakhic* reform movements in the diaspora which, he believed, were usually motivated by a desire for social accommodation and convenience, Leibowitz was uncompromising in his demand for individual discipline and conformity to the standards of *halakhah*.

Modernity and the desire for social integration had no normative weight in his eyes. Leibowitz believed that the impulse for change that grew out of the new Jewish political reality was legitimate and urgent, since it was indigenous to Judaism and reflected the Jewish situation itself. Given the dependence of Judaism on the existence and the life of a Jewish society and people, it is contradictory and inauthentic for *halakhists* to ignore the need to revise Jewish law in the light of the expanded scope of Jewish social and political responsibilities. There is, then, a striking - but logically consistent - dichotomy in Leibowitz's attitude towards *halakhah*. He is conservative and authoritarian regarding the personal dimension of *halakhic* practice (for example, *kashrut* and the laws of sexual purity). But at the same time he is boldly innovative in his call for reform in areas related to public life in the State of Israel.

This dichotomy highlights the unique nature of Leibowitz's philosophic agenda in contrast to that of most other Jewish thinkers in the twentieth century. Franz Rosenzweig and Hermann Cohen, for example, sought ways of legitimizing Judaism in a Christian society. Rosenzweig claimed that Judaism bore witness to the eschatological moment in history; Hermann Cohen focused on the messianic impulse of universality. Mordecai Kaplan tried to free Judaism from the problematic notions of supernaturalism and divine election and thus to facilitate the integration of Judaism into modern pluralistic societies. Joseph Soloveitchik and Abraham Heschel aimed at renewing Judaism as a serious option for the individual in modern society. Soloveitchik tried to enliven the *halakhic* option by presenting Judaism as a compelling existential drama; Heschel wanted to rehabilitate the theological and spiritual sensibilities of the individual Jew. For Leibowitz, by contrast, the problem was not Christianity or the philosophical critiques of religion, but whether Judaism could be a viable and authentic possibility for an autonomous political community. While in the Diaspora, Judaism has to deal with the problems of the individual and individual relations with the surrounding culture. But in Israel the community has precedence over the individual. The individual now needs to be convinced that Judaism is an acceptable framework for the whole of society. The unique character of Leibowitz's existentialism is that it is directed towards providing for the authenticity of life as a member of the community.

See also: Halakhah; Jewish philosophy, contemporary; Zionism

D. HARTMAN

List of works

Leibowitz, Y. (1975) *Yahudut, Am Yehudi u-Medinat Yisrael (Judaism, The Jewish People and the State of Israel)*, Tel Aviv: Schocken.(An account of his views on what is politically distinctive about Judaism.)

Leibowitz, Y. (1979) *Sichot al Pirkei Avot (Talks on The Ethics of the Fathers)*, Tel Aviv: Schocken.(Reflections on Talmudic themes and the philosophy of Moses Maimonides.)

Leibowitz, Y. (1980) *Emunato shel ha-Rambam*, trans. J. Gluker, *The Faith of Maimonides*, New York: Adama, 1987.(Brilliant and concise analysis of Moses Maimonides.)

Leibowitz, Y. (1992) *Yahdut, Am Yehudi v'Medinat Yisrael*, ed. E. Goldman, trans. E. Goldman, Y. Navon, Z. Jacobson, G. Levi and R. Levy, *Judaism, Human Values and the Jewish State*, Cambridge, MA: Harvard University Press.(Compilation of Leibowitz's essays on the meaning of Judaism in the modern worlds and a critique of the political structure in Israel, where religion has been coopted by the state.)

References and further reading

Hartman, D. (1985) *A Living Covenant*, New York: Free Press, 109-30.(A critique of Leibowitz's anthropology.)

Hartman, D. (1990) *Conflicting Visions*, New York: Schocken, 57-107.(A sympathetic introduction to Leibowitz's philosophy; a dialogue between the author and Leibowitz on the significance of Israel.)

Hartman, D. (1993) 'Yeshyahu Leibowitz', in S. Katz (ed.) *Interpreters of Judaism in the Late Twentieth Century*, Washington, DC: B'nai Brith Books, 189-204.(Short biography of Leibowitz.)

Sagie, A. (ed.) (1995) *Yeshyahu Leibowitz: Olamo Ve'haguto (Yeshayahu Leibowitz: His World and Philosophy)*, Jerusalem: Keter.(A broad collection of critical essays in Hebrew.)

Lenin, Vladimir Il'ich (1870-1924)

Lenin, leader of the October 1917 Revolution in Russia, wrote mainly about politics and economics, but as a Marxist of his generation he assumed that ideas about society needed to rest on sound philosophical premises. He was a militant atheist. He also regarded any other version of 'materialism' than his own as being a perversion of Marxism. Initially his works proposed an epistemology based on a crude analogy with photography. But in the First World War he revised his ideas after studying Hegel, and began to emphasize provisionality in the pursuit of scientific knowledge. Nevertheless he never disowned his earlier writings. And after his death his confused philosophical oeuvre retained axiomatic status in Marxism-Leninism.

1 Marxism and the defence of 'materialism'

Lenin had an influence on philosophical debate and education in his country and abroad entirely out of proportion with his expertise. Born in Simbirsk in the Russian Empire, he studied Marx's writings in his adolescence and became a follower of Georgii Plekhanov, who was the leading intellectual influence upon the clandestine Marxist groups in the Russian Empire in the 1880s and 1890s. Lenin's early articles took economic development and political struggle as their main themes. But he never lacked interest in questions of philosophy. On most fundamental questions he was Plekhanov's pupil, and proud of the fact. In so far as he intervened in the Russian Marxist discussions on epistemology, he contented himself with two activities: his repudiation of all religious belief and his castigation of the growing influence of Kant's ideas upon several prominent Russian Marxist thinkers.

He was equally opposed to the attempt by a group within his own section to inject a voluntarist element into Russian Marxist theory by linking it with the empiriocriticism of Mach (see Russian Empiriocriticism). Its main proponent, Aleksandr Bogdanov, was for some years an ally of Lenin in the Bolshevik faction which was formed after the Second Congress of the Russian Social-Democratic Labour Party in 1903. Bogdanov had always felt uneasy about central aspects of Plekhanov's philosophical work; and as Bogdanov and Lenin fell out about general political and cultural policy, so they were drawn into exposing their private dispute on epistemology. Lenin to a great extent stuck by Plekhanov's tenets. Above all, he followed Plekhanov in asserting that the 'external world' exists independently of human cognition. In treating the mind-matter dichotomy, he propounded that 'matter' imposed itself upon 'mind'. Indeed Lenin felt that Plekhanov described the process too weakly. Whereas Plekhanov imagined the mind as registering 'external reality' through a series of 'hieroglyphs', Lenin declared the process to be unmediated by any such process: instead the mind was supposedly a mechanism akin to a camera - and Lenin believed that cameras had the capacity for the exact registration of 'external reality'.

The extraordinary optimism of Lenin's epistemology is displayed in the conclusions he drew from this argument. In particular, he claimed that the Marxist mode of understanding society enabled its proponents to attain 'absolute truth'. To the counter-argument that such a claim would render obsolete any further research on economics, politics and culture, Lenin declared that *Das Kapital (Capital)* had yielded up only a portion of the available absolute truth. Nothing in Marx's *magnum opus* was going to be disproven; on the contrary, it would stand as a monument of unassailable verity. According to Lenin, *Capital* also laid down the necessary guidelines whereby to contribute further portions of absolute truth. Thus Marxism was 'scientific'; it was militantly 'materialistic', atheistic and began from the premise that the material and social world were in constant flux and that the study of changing phenomena was permanently required.

These opinions were stridently asserted in *Materializm i ėmpiriokrititsizm (Materialism and Empiriocriticism)*, the tract Lenin issued against Bogdanov in 1909. His style was overtly political. He affirmed that there were two 'parties' in philosophy: materialism (which he advocated) and idealism (which he opposed) (see Partiinost'). By entering into contention against Lenin, Bogdanov had removed himself from the only acceptable party - and Lenin the politician argued that by the same action Bogdanov had left the path of revolution and had abandoned organized Marxism. Lenin treated any refusal to espouse his own narrow version of Plekhanovite Marxism as a lurch into idealism that could eventually result in a resurgence of religious belief. As it happened, some among Bogdanov's political associates - notably Anatolii Lunacharskii and Maksim Gorkii - were contemporaneously developing a form of Marxism which proposed a kind of collective deification of the working class. This had

become known as 'god-building', and gave Lenin his polemical opportunity to write off all his opponents among Bolsheviks as aiming to effect a *rapprochement* with religion.

Bogdanov, who was a far subtler student of contemporary philosophy than his antagonist, did not take this lying down. He pointed out that Lenin's notion of absolute truth was antiscientific and metaphysical. Lenin made obeisance to Marx and Engels in a style of ritualist religion. Bogdanov argued that Marx had never made many of the claims made by Lenin in his name, and that the basic premise of any reasonable interpretation of Marxism should be that all analyses - including Marx's own - were merely provisional. Nor did Bogdanov fail to poke fun at Lenin's primitive understanding of how a camera works.

2 The exploration of 'dialectics'

Yet it would be wrong to give the impression that Lenin was merely exploiting philosophy for political advantage. At a basic personal level, his concern about epistemology was genuinely felt. Lenin made the assumption that any 'correct' (a favourite adjective!) statement about politics or economics had to be rooted in a 'correct' philosophical position. This, for him, was essential to Marxism. It consequently seemed to Lenin that if he disagreed with rival Marxists about their politics or economics, then a philosophical discrepancy must somehow lie at the foundations of the disagreement. Never was this clearer than in the First World War, when Lenin furiously objected to the policies pursued by the German Marxist theoretician Karl Kautsky. Kautsky had refused to make a clean break with the German Social Democratic Party over its decision to vote war credits to the German government. Lenin denounced this as a betrayal of socialist internationalism, and his reaction was not only to polemicize against Kautsky politically but also to investigate the philosophical premises implicit in Kautsky's wartime writings.

And so Lenin studied not only Marx and Engels but also those earlier philosophers who might help him understand the origins of Marxist epistemology. This brought him especially to Hegel and Aristotle; for his intuition was that Kautsky had an excessively deterministic approach to Marxism and underestimated the importance of 'dialectical' thought - of which both Hegel and Aristotle were notable exponents. Lenin was pretty thorough: he read Aristotle in a bilingual text which included the original Greek. He also re-read Marx's *Theses on Feuerbach*. Lenin's studies surprised and delighted him. With more than a touch of immodesty he decided that 'not one Marxist has completely understood Marx in the past half-century'.

His wartime notebooks on philosophy were meant to provide the material for a large, convincing exposition of the need for 'dialectics' to be bedded in the core of Marxism. His various jottings, which were not published at the time, exhibit several changes of stance from *Materialism and Empiriocriticism*. He abandoned his simplistic reflectional theory of cognition. Instead he stressed that human perception was a process of interaction not only between mind and matter but also between mind and concepts. Furthermore, he urged that the validity of theory could be judged only by 'practice' (*praktika*). In this fashion he was unwittingly moving close to the philosophical stance of his old adversary Bogdanov, who held that knowledge had to be obtained experimentally and could never be more than provisional in character. Lenin came to his new conclusions independently of Bogdanov. For whereas Bogdanov's ideas were a fusion of Marx and Mach, Lenin had refreshed his ideas by looking at Hegel - and continued to revile empiriocriticism as a reactionary doctrine.

Hegel's writings, especially those which dwelt on dialectical processes, by contrast had a strong attraction for Lenin, who contended that Hegelianism provided Marx and Engels with the key understanding that 'interruptions of gradualness', 'leaps' and 'breaks', were basic to the functioning of both the material and the social universe. Dialectics thereby became a philosophical justification for political revolution; and Lenin contended that it was precisely on this point that Kautsky had refrained from coming to terms with authentic Marxism. Lenin also argued that Kautsky had ignored the need for philosopher-revolutionaries to take chances, to speculate, to dream. Lenin had said something along these lines in his famous political treatise of 1902, *Chto delat' (What Is To Be Done?)*; but in the First World War such an argument became virtually a credo: 'The approach of mind (man) to a particular thing is…complex, divided, zigzaggish, *including within itself* the possibility of a flight of fantasy from life.…It is stupid to deny the role of fantasy even in the strictest science'.

Yet simultaneously Lenin, who was an intelligent but confused amateur in philosophy, retained much of the views expressed in *Materialism and Empiriocriticism* in 1909. As ever, he insisted that man's abstract conceptions about the world derived from 'a knowledge of the pattern of the objective links of the world'. His belief about there

being two entirely separate camps in philosophy, materialism and idealism, was unshaken. And he jotted down in his notebooks that there was no reason to deny the attainability of 'living, fruitful, true, powerful, omnipotent, objective, absolute human knowledge'. In 1920, furthermore, when he fell into public dispute again with Bogdanov (and indeed with fellow-Bolshevik Party leader Nikolai Bukharin), Lenin ordered the republication of *Materialism and Empiriocriticism*.

3 Lenin's philosophical legacy

Consequently Lenin left a legacy of great contradiction. He had changed much of his thought, but had done this by adding bits without excising earlier ones. As a philosopher he would hardly deserve a footnote in the history of twentieth-century thought if it had not been for his political career as the founder of the USSR. In everything he wrote he showed greater confidence than expertise. Even in most of his apparently original interpretations of Marx, he had largely been anticipated by Bogdanov. In his analysis of Hegel and Kant he misrepresented the differences between the two. Nevertheless in death he became the object of a secular state. Lenin himself had never claimed to be more than a follower of Marx. Nor had he attempted to assemble all his various political, economic and philosophical works into a single, authoritative set of tenets. This was done for him by his successors, especially Stalin; and his writings were codified and designated as Marxism-Leninism.

For most of the ensuing decades it was *Materialism and Empiriocriticism* which was used as the supreme philosophical text of the regime (although the *Philosophical Notebooks* became more widely discussed after Stalin's death in 1953). Even so, there was enough confusion and incompleteness in Lenin's published thought for Soviet philosophers to have plenty to argue about. The practice of philosophy in the USSR, cramped and abused as it was, would have had an even harsher environment if Lenin had been a more expert epistemologist.

See also: Bogdanov, A.A.; Plekhanov, G.L.V.; Russian empiriocriticism

ROBERT SERVICE

List of works

Lenin, V.I. (1909) *Materializm i ėmpiriokrititsizm*, in *Polnoe sobranie sochinenii*, vol. 18, Moscow: Gospolitizdat, 1958-69, 5th edn.(Many republications and translations in the Soviet era.)

Lenin, V.I. (1895-1916) *Filosofskie tetradi*, in *Polnoe sobranie sochinenii*, Moscow: Gospolitizdat, 1958-69, 5th edn, vol. 29; trans. C. Dutt, *Philosophical Notebooks*, in *Collected Works*, vol. 38, Moscow: Progress and London: Lawrence & Wishart, 1968-9.(First published in Russian in the *Leninskii sbornik* series, Moscow.)

Lenin, V.I. (1902) *Chto delat'*, in *Polnoe sobranie sochinenii*, Moscow: Gospolitizdat, 1958-69, 5th edn, vol. 5; trans. J. Fineberg, *What Is To Be Done?* in *Collected Works*, vol. 5, Moscow: Progress and London: Lawrence & Wishart, 1968-9.

References and further reading

Kolakowski, L. (1978) *The Main Currents of Marxism*, vol. 2, *The Golden Age*, Oxford. (Magisterial, hostile account of Leninist philosophy which situates Lenin's ideas in the Marxist tradition and provides a well-constructed bibliography).

Read, C. (1979) *Religion, Revolution and the Russian Intelligentsia, 1900-1912*, London.(Useful general account of the intellectual milieu of prewar Russian intellectual discussions.)

Scanlan, J.P. (1985) *Marxism in the USSR. A Critical Survey of Current Soviet Thought*, London.(Outstanding survey of the scope of Marxism-Leninism, including its philosophy, after Lenin's death; impressive and useful bibliography.)

Service, R. (1985, 1991, 1995) *Lenin: A Political Life*, vols 1-3, London.(General analysis of Lenin's career, involving an examination of the connections between his politics and his philosophy.)

Wetter, G. (1958) *Dialectical Materialism. A Historical and Systematic Survey of Philosophy in the Soviet Union*, London.(Pioneering study of the Marxism of Lenin and Stalin; useful despite the further development of scholarship since the book's publication.)

Leont’ev, Konstantin Nikolaevich (1831-91)

One of the more original and provocative nineteenth-century Russian thinkers, Leont’ev directed a powerful intellectual attack at the dominant historical movement of his time: the process of modernization that was sweeping across Western Europe and making inroads into Russia. Leont’ev resisted the sociopolitical aspects of this process: the spread of democratic, egalitarian and constitutional principles. He also resisted its cultural and psychological aspects: the emergence of a standardized, homogenized mode of life and set of values.

Leont’ev defended the traditional values and institutions of monarchy, established Church, aristocracy and - especially - the distinctiveness and variety of national cultures. It was this distinctiveness, based on the isolation of nation states, that he saw as increasingly threatened by the advancing technologies of transport and communication.

1 Social philosophy and philosophy of culture

Konstantin Nikolaevich Leont’ev was born into a gentry family in the village of Kudinovo. He studied medicine at Moscow University and served as a military surgeon during the Crimean War, 1853-6. During his first two decades of literary activity (1854-76) he wrote mostly short stories and novels. His talent was recognized by Ivan Turgenev, who helped and encouraged him. Over the next two decades (roughly, 1872-91) Leont’ev produced a flood of brilliant and disturbing philosophical and political essays. He served for a decade as a Russian diplomat in various parts of the Ottoman Empire, 1863-73; he also worked in the Censor’s Office in Moscow, 1880-7. In 1871 he experienced a spiritual crisis and lived for a year in a Russian monastery on Mount Athos. His efforts to become a monk were gently, and wisely, rebuffed by his spiritual advisors, who felt that he was not yet ready to give up the ‘world’ of writing and publishing. He in fact took secret monastic vows a few months before his death, at the Trinity Monastery near Moscow.

Leont’ev has often been called a ‘Russian Nietzsche’, not in the sense of a Russian thinker influenced by Nietzsche, since he died before Nietzsche’s ideas became known in Russia (see Nietzsche: impact on Russian thought). Rather, it is sometimes claimed that Leont’ev anticipated Nietzsche’s main doctrines. This claim is seriously misleading. Leont’ev did share, in advance, Nietzsche’s stress on the priority of aesthetic over moral and practical values; his ‘aesthetic immoralism’ dates from the mid-1860s, a time when the dominant Russian Nihilists and Realists were deriding aesthetic values (see Russian Materialism: ‘the 1860s’; Nihilism, Russian). For example, a leading character in an early Leont’ev novel made the shocking assertion: ‘*A single century-old magnificent tree is worth more than twenty faceless men*; and I will not cut it down in order to be able to buy medicine [to treat] the peasants’ cholera!’ (1912-14, vol. 1: 306; emphases added).

Leont’ev also agreed with Nietzsche in celebrating the ‘poetry of war’, regarding security and comfort as deadening to cultural creativity; and of course he was entirely at one with Nietzsche in opposing the levelling and standardizing tendencies of the time. But there were a number of key differences. For Nietzsche, Christianity and democracy (and indeed socialism) were on the same side; in contrast Leont’ev spoke of the ‘Antichrist of democracy’. Leont’ev embraced strong nation states and state churches, both of which Nietzsche opposed. There is no counterpart in Nietzsche’s work to Leont’ev’s sweeping condemnation of modern technology. Leont’ev preached what he called the ‘poetry of *life*’ - the social and historical life of nations and ethnic groups in all its colourful diversity. Nietzsche preached the ‘poetry’ or ‘aesthetics’ of a universal *high culture*. Nietzsche accepted, and Leont’ev rejected, a powerful orientation towards the remote historical future. Leont’ev remained a Christian, preaching ‘love of one’s neighbour’; Nietzsche, the ‘anti-Christian’, preached ‘love of the far-off’, that is, ‘love of the high culture of the remote historical future’ and endorsed ‘instrumental cruelty’ towards one’s neighbours who are weak, sickly and uncreative. In a word, Leont’ev embraced, while Nietzsche repudiated, ‘good Samaritanism’. At the same time, Leont’ev introduced a highly original corrective, one which might well be considered ‘Nietzschean’: he insisted that Christian charity not be one-sidedly ‘democratic’. It should not be directed only at suffering workers or wounded soldiers (of whom Leont’ev had had direct experience during the Crimean War); it should include the powerful and privileged in their hour of need, for instance, defeated generals and aristocrats abused by the mob.

Leont’ev’s rejection of modernity is more sweeping than Nietzsche’s. In fact, the only aspect of the many-faceted

process of modernization that Leont’ev accepted was its stress on the subjectivity and - with qualifications - freedom of the individual person. On this point Leont’ev went further in Hegel’s direction than Nietzsche was prepared to do.

Leont’ev offers many impressive catalogues of the glories of sociopolitical and cultural diversity and vividness in ancient Greece and Rome and in medieval Europe, along with catalogues of the horrors of sociopolitical and cultural monotony and drabness in present-day Western Europe and - increasingly - Russia. The former he celebrated as ‘the poetry of life’; the latter he castigated as its ‘prose’ or ‘prosiness’. Here is one example, which sets out four kinds of differences that are being eliminated by the hated process of modernization: (1) social (the difference between a peasant, a nobleman and ‘a marquis in velvet and plumes’); (2) occupational (that between a soldier and a priest); (3) national-ethnic (that between a Basque and a man from Brittany, or between a Tyrolian and a Circassian); and (4) ‘intrareligious’ (that between ‘a Trappist in a hair shirt’ and a prelate in brocade). In this fourth difference Leont’ev would appear to have overshot his mark: even in the secularizing Europe of the late nineteenth century there was still a clear distinction between a Trappist in his hair shirt and a prelate in brocade.

The violence of Leont’ev’s repudiation of technology - both the principal forms already familiar in his own day (for instance railroad and telegraph) and those being introduced during his final years (telephone, electric light and so on) - is surprising. So is his claim that such ‘peaceful’ technology is much more destructive than the newer technologies of war (like the machine gun). But Leont’ev’s point is that these faster, more efficient, more pervasive forms of transport and communication are undermining the ‘isolation of states’ and national and ethnic cultural traditions, thus facilitating the general process of standardization and homogenization of culture. Like other aspects of the modernization process, technological advance is reducing the vivid poetry of life to colourless prose.

2 Philosophy of history

In 1869-71 Nikolai Danilevskii, in his influential book *Rossiia i Evropa (Russia and Europe)*, set out a large-scale theory of ‘cultural-historical types’ intended to show that the Slavic-Russian type possessed hitherto unrecognized cultural vitality and a special promise of future cultural creativity. Leont’ev’s philosophy of history was similar to, but in important ways distinct from, Danilevskii’s. Like the latter, he used an essentially biological model of historical development. Historical cultures, like living organisms, develop from a stage of ‘initial simplicity’ to a second stage of ‘flourishing complexity’ and then sink, through a third process of ‘levelling interfusion’, to organic decay and death. Western Europe, which had achieved its stage of flourishing complexity in the Renaissance, was already far advanced in the terminal process of ‘levelling interfusion’. Leont’ev estimated the life span of historical cultures at about 1,200 years and sometimes asserted - with the Slavophiles - that Russia was a younger culture than Western Europe and therefore had more of its period of flourishing complexity still ahead of it. But at other times, and more typically, he claimed that nineteenth-century Russia possessed a unique combination of institutional elements - autocracy, Russian Orthodoxy and a system of deep and stable class divisions - which would make it possible to ‘freeze Russia over’ and thus save it from the creeping Western European ‘rot’. Note that he claimed no special superiority - as Danilevskii had - for the uniquely Slavic elements of Russian culture; indeed, he stressed that all were derived from a non-Slavic source, namely Byzantium, that was *Greek* in language and culture.

Leont’ev sometimes spoke of his own position as ‘anti-European’, and one of his critics accused him of pervasive ‘miso-Europeanism’. Both terms are misleading. Leont’ev’s animus was directed against the process of modernization that was coming to Russia from Western Europe, not against Western Europe as such. As we have seen, he admired much in Western European history up to the Renaissance, specifically including the Roman Catholic Church and the institution of the Germanic Knights. At the same time the unifications of Italy and Germany that were under way in Leont’ev’s lifetime struck him as destructive of the diversity of ‘ethnic’ or ‘regional’ cultures of the various parts of the newly unified national wholes (say, Venice and Naples, or Bavaria and Prussia). It is true that Leont’ev was enchanted by the rich cultural diversity that he encountered during his stays in Greece, Turkey and the Balkans, and credited this to their more complete isolation (than Russia’s) from Western European influences. But Leont’ev’s fundamental protest was against modernization, not against Europe as such.

Looking sombrely into the future, he saw the triumph of a worldwide socialist movement - not because the

socialists were right but because their principal adversaries, the liberals, were wrong. Leont'ev even envisaged the possibility that such a socialist movement - 'the feudalism of the future', as he called it - might be headed by a Russian tsar!

3 Philosophy of religion

From the beginning Leont'ev's approach to religion, like his approach to secular matters, had a strong aesthetic colouring. He was charmed by the beauty of the Russian Orthodox service - the singing, the icons, the incense. But later his religious convictions took a severe 'Byzantine' turn and he came to stress the fear of God, ascetic discipline and a profound disillusionment with all things earthly as merely 'a passing dream'. In his final letters to Vasilii Rozanov Leont'ev commented bitterly that 'Christian preaching' was joining forces with the hated modernization process to 'kill the aesthetics of life on earth'. He added, in what appears to have been a tone of quiet desperation: 'We must help Christianity even at the cost of our beloved aesthetics' (letter of 27 May 1891).

Turning to the main proponents of a 'renewed' Christianity in Russia in his time, Leont'ev found both Tolstoi (he was right in this) and Dostoevskii (here he was wrong) guilty of falling into a humanistic 'rose-coloured' Christianity that verged on 'anthropolatry' - 'a new faith in *earthly man* and in earthly mankind - in the ideal *self-sufficient, autonomous* worth and dignity of the *individual person*' (1912-14, vol. 8: 160; original emphases). He was especially infuriated by Tolstoi's presumption in editing and rewriting the New Testament and by Tolstoi's reference, in a conversation with Leont'ev, to '*my* Gospel'.

The originality and depth of Leont'ev's ideas about history, society, politics, culture and religion led to his being 'heard but not heeded, read but not understood' in his own time and for many years after his death. It is only in the twentieth century, especially during its final decade (and mostly in post-communist Russia) that Leont'ev's provocative and unsettling works have been widely reprinted and seriously discussed.

GEORGE L. KLINE

List of works

Leont'ev, K.N. (1912-14) *Sobranie sochinenii (Collected Works)*, 9 vols, St Petersburg: V. Sablin (vol. 9 was published by the Russkoe knizhnoe t-vo Deiatel'); xerographic repr., Ann Arbor, MI: University Microfilms, 1967.(Includes Leont'ev's early fiction, later reminiscences and essays on social, political and religious themes. Does not include his correspondence or his autobiographical essay 'Moia literaturnaia sud'ba' (My Literary Fate).)

Leont'ev, K.N. (1890) 'Analiz, stil' i veianie. O romanakh Gr. L.N.Tolstogo'; trans. S.E. Roberts, 'The Novels of Count L.N. Tolstoy: Analysis, Style, and Atmosphere - a Critical Study', in *Essays on Russian Literature: The Conservative View. Leontiev, Rozanov, Shestov*, Athens, OH: Ohio University Press, 1968, 225-356.(This is the only complete English translation of a Leont'ev work in philosophy or literary criticism.)

Leont'ev, K.N. (1969) *Against the Current: Selections from the Novels, Essays, Notes, and Letters of Konstantin Leontiev*, ed. G. Ivask, trans. G. Reavey, New York: Waybright & Talley.(See especially the section entitled 'Philosophy', 133-229.)

Leont'ev, K.N. (1993) *Izbrannye pis' ma 1854-1891 (Selected Letters 1854-1891)*, ed. D.V. Solov'ëv, St Petersburg: Pushkinskii fond.(This important edition contains 700 of Leont'ev's letters, of which 478 are published here for the first time.)

Leont'ev, K.N. (1996) *Vostok, Rossiia i slavianstvo: Filosofskaia i politicheskaia publitsistika. Dukhovnaia proza (The East, Russia and Slavdom: Essays on Philosophical and Political Themes. Spiritual Prose)*, ed. G.B. Kremnev, introduced by V.I. Kosik, Moscow: Respublika.(This lengthy collection, which includes previously unpublished and uncollected pieces, contains such philosophically important essays as 'Vizantizm i slavianstvo' (Byzantium and Slavdom) (1875), 'Srednii evropeets kak ideal i orudie vsemirnogo razrusheniia' (The Average European as an Ideal and Instrument of Universal Destruction) (1872-84), and 'Plemennaia politika kak orudie vsemirnoi revoliutsii' (The Politics of the Tribe as an Instrument of Universal Revolution) (1888). There are also essays on the religious views of Dostoevskii and Tolstoi.)

References and further reading

Berdiaev, N.A. (1926) *Konstantin Leont'ev: Ocherk iz istorii russkoi religioznoi mysli (Konstantin Leont'ev: An Essay in the History of Russian Religious Thought)*, Paris: YMCA-Press, trans. G. Reavey, *Leontiev*, London: Bles, 1940; repr. Orono, ME: Academic International, 1968.(A pioneering study, which stresses Leont'ev's

critique of religion and culture.)

Dolgov, K.M. (1995) *Voskhozhdenie na Afon: Zhizn’ i mirosozertsanie Konstantina Leont’eva (The Ascent of Mount Athos: Konstantin Leont’ev’s Life and World View)*, Moscow: Luch.(This full and sympathetic account makes judicious use of Leont’ev’s correspondence.)

Kline, G.L. (1968) *Religious and Anti-Religious Thought in Russia*, Chicago, IL: University of Chicago Press, 35-54.(Offers a fuller treatment of some of the topics discussed in the present essay.)

Kologriwof, I. von (1948) *Von Hellas zum Mönchtum: Leben und Denken Konstantin Leontjews (From Greece to Monasticism: Konstantin Leont’ev’s Life and Thought)*, Regensburg: Gregorius Verlag.(A careful and comprehensive study by a Russian Orthodox priest.)

Korol’kov, A. and Kozyrev, A. (eds) (1995) *K.N. Leont’ev. Pro et contra: Lichnost’ i tvorchestvo Konstantina Leont’eva v otsenke russkikh myslitelei i issledovatelei, 1891-1917 (K.N. Leont’ev: Pro and Contra: Konstantin Leont’ev’s Personality and Work as Appraised by Russian Thinkers and Scholars, 1891-1917)*, St Petersburg: Izdatel’stvo Russkogo Khristianskogo gumanitarnogo instituta.(Among the thinkers represented are Berdiaev, Rozanov and Vladimir Solov’ëv. Among the scholars are A.A. Aleksandrov, Father I. Fudel’ and B.Z. Griftsov.)

Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, vol. 1, Paris: YMCA-Press, 433-57; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, vol. 1, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, 434-53.(A sympathetic but not uncritical account by a Russian Orthodox priest who was the leading émigré historian of Russian philosophy.)

Leśniewski, Stanisław (1886-1939)

Leśniewski was one of the most distinguished members of the Warsaw School of Logic. His scientific development can be divided into two periods. In the first 'philosophical' period (1911-16), he worked on problems on the borderline of logic and philosophy. In the second period (1916-39), Leśniewski concentrated on mathematical logic. Together with Łukasiewicz, he established the Warsaw School of Logic. Leśniewski intended to build a comprehensive system of logic which might be the basis for all knowledge. His system, unorthodox in many points, consists of three parts: protothetic (a generalized sentential calculus), ontology (a calculus of names) and mereology (a theory of the whole/part relation).

1 Life

Leśniewski was born on 30 March 1886 in Serpukhova, Russia. He studied philosophy in Leipzig and Munich (where he attended courses by Cornelius, Geiger and Pfänder). In 1910 he went to Lwów to prepare his Ph.D. thesis on existential sentences under the supervision of Twardowski; he obtained his Ph.D. in 1919. In 1915-18 he lived in Moscow, where he taught mathematics in Polish schools. At that time, he was a 'leftist' in his political views and delivered several lectures on Marxist philosophy; he entirely abandoned this ideology after the October Revolution. In 1919 Leśniewski became the professor of philosophy of mathematics at the University of Warsaw. He died of cancer on 13 May 1939.

2 Early views

In his Ph.D. dissertation, Leśniewski argued that all negative existential sentences (for example, sentences of the type '*x* does not exist') are self-contradictory. This was a direct consequence of the thesis that every sentence refers to an object. Leśniewski's doctoral dissertation was written from the point of view of a very traditional grammar and logic; he was at that time very strongly influenced by Cornelius, J.S. Mill, Marty and Husserl.

Leśniewski's first contact with new ideas in logic took place in 1911, when he read *On the Principle of Contradiction in Aristotle* by Łukasiewicz (PAU, Kraków, 1910). Leśniewski reported his contact with this book in the following way:

> In the year 1911…I came across a book by Jan Łukasiewicz about the principle of contradiction in Aristotle.…This book became a revelation…for the first time in my life I learned of the existence of the 'symbolic logic' of Bertrand Russell as well as his 'antinomy' regarding the 'class of all classes, which are not elements of themselves'.
>
> (Leśniewski 1992: 181)

Leśniewski undertook to resolve Russell's antinomy (see Russell, B.A.W. §§4-8). After some unsuccessful attempts within standard set theory, he came to the view that the concept of a class must be reinterpreted in terms of the mereological concept of an aggregate. Given this interpretation, the expression of 'the class of all classes which are not elements of themselves' has no reference and the antinomy disappears. Leśniewski's analysis of the concept of class was the starting-point for his mereology.

Leśniewski's other investigations in the first period concerned the principle of contradiction, the principle of the excluded middle, and the concept of truth. He tried to prove the principle of contradiction via a proof of the sentence 'no object contains contradictions'. But in the case of the principle of the excluded middle, Leśniewski denied its universal validity, because he maintained that sentences of the type '*a* is *b*' with empty names in the place of '*a*' are false; thus, both '*a* is *b*' and '*a* is not *b*' are false. Leśniewski also defended the absolutism of truth: every truth is eternal and sempiternal.

Leśniewski's own evaluation of his early works was decisively negative:

> I struggled with a number of problems which were beyond my powers at that time.…I have mentioned those works desiring to point out that I regret that they have appeared in print, and formally 'repudiate' them herewith…affirming the bankruptcy of the 'philosophical'-grammatical work of the initial period of my work.
>
> (Leśniewski 1992: 197-8)

However, a closer analysis shows clearly that Leśniewski did not reject all of his views from the early writings. As well as the mereological conception of class, he retained the absolutist conception of truth and the view that all sentences of the type '*a* is *b*' are false if '*a*' is an empty name.

3 General remarks on Leśniewski's logical systems

Leśniewski's system of logic comprises three systems: protothetic (an extended sentential calculus), ontology (a calculus of names) and mereology (a theory of mereological classes). This succession indicates the logical order of those systems. However, Leśniewski invented his systems in just the opposite succession. Mereology appeared as the first in 1916. Then, at the beginning of the 1920s, he completed his ontology and protothetic.

There is a considerable difference in style between Leśniewski's early work and his later work on protothetic and ontology (see Leśniewski 1929 for example). A booklet of 1916 on the foundations of the general theory of classes is written in an informal manner. Then, Leśniewski, under Chwistek's influence, began to employ formal languages as tools in doing logic. His formalism was very extreme in the sense that he required a complete codification of the languages of formal systems. Hence, he regarded Russell and Whitehead's *Principia Mathematica* as a very obscure work, full of linguistic confusions: for example, it conflates object-language and metalanguage. Indeed, Leśniewski was probably the first person to point out that Frege achieved much more precision than Russell did. On the other hand, Leśniewski rejected formalism understood as a view that logic and mathematics are games with symbols devoid of meaning. He called his view 'intuitionistic formalism', although this view has nothing to do with Intuitionism as a position in the foundations of mathematics.

Leśniewski was a radical nominalist. He took formal systems to consist of concrete expressions (symbol-tokens, to use a popular label). Thus, on his view, formal deductive systems are always finite in the number of formulas, although they are 'unfinished' or 'unbounded', because it is always possible to add further formulas. Hence, his systems are equipped with very precise directives for introducing new expressions into the body of a system. A special role is played by the rules of definition for new signs. Leśniewski was the first logician who explicitly stated the principles of correct definition in formal systems. His systems are axiomatic, but he also used techniques of natural deduction. Certainly, his logic is peculiar, if by a deductive system we understand what is typically meant. However, it is commonly recognized that his formalization is the most precise in the entire history of logic (see Nominalism).

4 Protothetic

Protothetic is a generalized sentential calculus. It is a sentential logic with quantifiers which bind variables of several semantic categories: sentential variables, functorial variables which range over functors forming sentences from sentential arguments, functorial variables which range over functors of functors, and so on. In general terms, quantifiers in the theorems of protothetic bind variables of an arbitrary semantic category which are definable when we start with a category of sentences.

The full system of protothetic is based on equivalence as the sole primitive term. An important result for the formalization of protothetic on the basis of equivalence was obtained by Tarski (in 1923) who showed that equivalence and the universal quantifier suffice to define negation and conjunction. The functions of protothetic are not limited to giving rules for sentential connectives, because this system also contains rules for the quantifiers which are used in ontology and mereology. Protothetic is, in a sense, an absolute sentential logic, because the principle of bivalence and the principle of extensionality are among its theorems. Thus, protothetic can be considered as an adequate representation of the classical idea of logic. Protothetic is consistent, but the problem of its completeness is not yet fully solved. What is known is that elementary protothetic (that is, protothetic with quantifiers restricted only to sentential variables) is complete.

5 Ontology

Ontology is the system which arises when the functor 'is' is added to protothetic. This functor forms sentences from names. Ontology considers the grammatical structure '*a* is *b*'. Such a sentence is false if its subject is empty or refers to more than one object. This meaning is captured by the axiom of ontology which says: for any *a* and *b*, the sentence '*a* is *b*' is equivalent to the conjunction of sentences (i) for some *c*, *c* is *b*, (ii) for any *c* and *d*, if *c* is *a* and *d* is *a*, then *c* is *d*, (iii) for every *c*, if *c* is *a*, then *c* is *b*. The intuitive content of the axiom is this: '*a* is *b*' is true

if and only if '*a*' is non-empty, there is only one object to which '*a*' refers ('*a*' is a singular name), and whatever falls under '*a*' also falls under '*b*'.

Ontology is supplemented by several definitions of sentence-forming functors operating on names. Two examples illustrate this matter: (i) for any *a*, *a* exists if and only if for some *x*, *x* is *a*; (ii) for any *a*, *a* is an object if and only if for some *x*, *a* is. (i) gives a definition of existence, and (ii) of being an object. These examples show how 'is' is understood in Leśniewski's ontology. This meaning of 'is' closely corresponds to the sense of *est* in Latin as the copula in 'Socrates est homo'. 'Is' in the considered meaning has no spatiotemporal connotations and it does not indicate the membership-relation from set theory. Also, Leśniewski's 'is' is different from 'is' in 'there is justice' or 'every man is mortal'. In general, one must be very careful in translating the basic functor of ontology (symbolically denoted by 'Čepsiv;') by the English 'is', because the latter is modified by articles which do not occur in Latin (or Polish).

Leśniewski's ontology is consistent, but the question of its completeness is still open. Ontology performs the role of predicate logic, but there are also important differences. For example, identity is definable in first-order ontology, whereas it is not definable in the standard first-order logic, and the theorems of ontology are true in any domain, including the empty one. Why was this system called 'ontology'? Leśniewski chose this term because he conceived ontology as a logical theory which offers a general theory of objects in the sense of Aristotle and his followers.

6 Mereology

Mereology is a theory of classes, but classes in their mereological (collective) sense, contrary to the standard set theory which considers classes in their distributive meaning. The crucial differences between these conceptions of class is that elementhood in mereology is transitive, but membership in set theory is not: if *x* is a mereological element of *y* and *y* is a mereological element of *z*, then *x* is also a mereological element of *z*, but the comparable rule does not hold for membership. Another peculiarity of mereology is that there is no empty set.

For Leśniewski, the mereological concept of class is more intuitive than the concept of set in the distributive sense. He argued that intuitions concerning collective classes are perfectly consistent with Cantor's famous statement that any manifold which can be considered as a unity is a set (see Cantor, G.; Cantor's theorem). Naturally, a problem arises whether mereology is as strong as set theory. The answer is negative: mereology is weaker (see Mereology).

7 Concluding remarks

Leśniewski influenced the development of logic in Poland very strongly. Together with Łukasiewicz, he trained many logicians, including Tarski (who obtained his Ph.D. under Leśniewski), Lindenbaum, Wajsberg, Sobocinski, Słupecki and Lejewski. Sobocinski and Lejewski became his main followers. In spite of his unorthodoxy in logic, Leśniewski invented several ideas which belong to the standard logical canon. For example, he initiated the theory of syntactic categories developed by Ajdukiewicz in the 1930s. Leśniewski revived Frege's object-language/metalanguage distinction and used it to suggest a solution to the Liar paradox (this solution was later developed byTarski); he also stated the conditions for correct formal systems commonly accepted by Polish logicians.

His own logical ideas are on the margin of the mainstream of logic. However, an interest in Leśniewski's logical systems is an important ingredient of logical investigations, and studies on his logical systems have been undertaken in all parts of the world, particularly in Poland, in England (among Lejewski's students - the Manchester School), and in the USA (among Sobocinski's students - the Notre Dame School). These studies have concentrated on the properties of Leśniewski's systems, and their relations to standard theories, like first-order logic, Boolean algebra and set theory. There are also attempts to apply his ideas to the analysis of natural language; for example, ontology has been interpreted as a theory of plural terms. Leśniewski's logical ideas also influenced philosophy. His ontology is nominalistic, because it admits only particulars as values of variables, and this idea was an important influence on Kotarbiński. Thus, it can be said that there is a distinctive Leśniewskian paradigm in logic.

See also: Poland, philosophy in

JAN WOLEŃSKI

List of works

Leśniewski, S. (1992) *Collected Works*, vols I and II, eds S.J. Surma, J.T. Srzednicki, D. I. Barnett and V.F. Rickey, Dordrecht: Kluwer.(Contains translations of Leśniewski's papers published in 1911-38 as well as a bibliography of Leśniewski's and writings on him until 1977.)

Leśniewski, S. (1911) 'Przyczynek do analizy zdan egzystencjalnych' ('A Contribution to the Analysis of Existential Propositions'), *Przeglad Filozoficzny* 14: 329-45.(Leśniewski's doctoral dissertation.)

Leśniewski, S. (1912) 'Proba dowodu ontologicznej zasady sprzecznosci' ('An attempt at a Proof of the Ontological Principle of Contradiction'), *Przeglad Filozoficzny* 15: 202-26.(An attempt to derive the principle of contradiction from the thesis that no objects contain contradictions.)

Leśniewski, S. (1913a) 'Krytyka logicznej zasady wylaczonego srodku' ('The critique of the Logical Principle of the Excluded Middle'), *Przeglad Filzoficzny* 16: 315-22.(Leśniewski argues that the principle of excluded middle fails for sentences with empty names.)

Leśniewski, S. (1913b) 'Czy prawda jest tylko wieczna czy tez wieczna i odwieczna?' ('Is All Truth Only True Eternally or is It Also True without a Beginning?'), *Nowe Tory* 18: 493-523.(A defence of an 'absolutist' conception of truth.)

Leśniewski, S. (1916) *Podstawy ogolnej teorii mnogosci I (Foundations of the General Theory of Sets I)*, Moscow: Poplawski. (An informal presentation of Leśniewski's mereological set theory.)

Leśniewski, S. (1927-31) 'O podstawach matematyki' ('On the Foundations of Mathematics'), *Przeglad Filozoficzny* 30: 164-206; 31: 261-91; 32: 60-101; 33: 77-105; 34: 142-70.(Contains a general overview of Leśniewski's systems, a detailed exposition of mereology, and remarks on the meaning of an expression '*a* is *b*').

Leśniewski, S. (1929) 'Grundzüge eines neuen Systems der Grundlagen der Mathematik', *Fundamenta Mathematicae* 14: 1-81.(Contains a detailed exposition of protothetic.)

Leśniewski, S. (1988) *Lecture Notes in Logic*, eds J.T.J. Srzednicki and Z. Stachniak, Dordrecht: Kluwer.(Lecture notes on protothetic, ontology, arithmetic and Whitehead's theory of events, composed from students' notes.)

References and further reading

Luschei, E. (1962) *The Logical Systems of Leśniewski*, Amsterdam: North Holland.(A systematic exposition of Leśniewski's systems.)

Mieville, D. (1984) *Un developpement de systèmes logiques de Stanislaw Leśniewski*, Berne: Lang.(A recent presentation.)

Simons, P. (1992) *Logic and Philosophy in Central Europe: From Bolzano to Tarski*, Dordrecht: Kluwer. (Contains several essays on Leśniewski's ideas.)

Srzednicki, J.T.J, Rickey, V.F. and Czelakowski, J. (eds) (1984) *Leśniewski's Systems Ontology Mereology*, The Hague: Nijhoff, and Wrocław: Ossolineum.(A collection of papers on ontology and mereology.)

Stachniak, Z. (1981) *Introduction to Model Theory for Leśniewski's Ontology*, Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.(An attempt to build a model theory in the standard sense for Leśniewski's ontology.)

Surma, S.J. (ed.) (1977) *On Leśniewski's Systems*, *Studia Logica*, vol. XXXVI, 4. (A collection of papers on Leśniewski's systems.)

Wolenski, J. (1989) *Logic and Philosophy in the Lvov-Warsaw School*, Dordrecht: Kluwer.(Chapter VII contains an overview of Leśniewski's ideas.)

Lessing, Gotthold Ephraim (1729-81)

Gotthold Ephraim Lessing occupies a central place in eighteenth-century European belles-lettres. He was a significant religious and theological thinker whose work puzzled his contemporaries and still provokes debate. He has been variously called a deist, a concealed theist, a Spinozist-pantheist, a panentheist, and an atheist. He was a significant dramatist whose major works include Minna von Barnhelm, known as the first modern German comedy, and Nathan the Wise, which places Lessing in the tradition of eighteenth-century toleration and humanism. He was an active promoter of the contemporary German theatre and an influential drama critic and theorist. He had broad classical and antiquarian interests. And he has some claims to being one of the early developers, if not a founding father, of the discipline of philosophical aesthetics.

Philosophically, Lessing belongs to the tradition of G.W. von Leibniz and Christian Wolff and was familiar with the post-Wolffian aesthetics being developed by Alexander Baumgarten and his follower Georg Friedrich Meier. Most importantly, perhaps, Lessing was acquainted with Moses Mendelssohn, to whose work his own philosophical writings bear many similarities and who read and commented on Lessing's aesthetic writings. But Lessing cannot be identified with any of these philosophical sources and influences. His work retains many rationalist presuppositions, but Lessing also consciously sought a more inductive approach. He adhered to neoclassical standards with respect to beauty and the application of rules of art, but severely qualified those standards by justifying them empirically and appealing to emotional effects rather than to ideal forms or Cartesian clarity. Lessing's aesthetics must be inferred from his work, particularly from his Laocoon, some of the numbers of the Hamburg Dramaturgy, and to a lesser extent from short works such as 'How the Ancients Represented Death' and the letter of 26 May 1769 to Friedrich Nikolai. What emerges is a sometimes inconsistent and fragmentary aesthetic, which one might describe as a critical rationalism.

1 The interpretation of the 'Laocoon' group

Lessing's *Laocoon* is a complex work that embeds an even more complex aesthetic. The published text, which was projected as the first part of a longer work that Lessing never completed, interweaves two principal themes. The first is a dispute with Johann Joachim Winckelmann over the interpretation of the statue group *Laocoon*. Winckelmann had claimed that the expression on the face of the priest Laocoon when he is fatally entangled with the serpents is more restrained than the scream described by Virgil in the *Aeneid* when dealing with the same incident. Winckelmann attributed this restraint to the classical Greek virtues of control and dignity in adversity, and ultimately to a superior Greek sense of ideal form. Lessing had seen only engravings of the *Laocoon* and not the statue itself, even as a copy, and he therefore conceded to Winckelmann the interpretation of the group itself. (This would not be granted by many contemporary art historians who find the *Laocoon* extremely emotionally wrought compared to classical models.) But Lessing disputed the significance of Winckelmann's interpretation by pointing out that classical painting and sculpture were frequently described as depicting extreme suffering; thus it was not Greek virtue as such that restrained the artists in this case. Rather, Lessing argued, artists are constrained by their medium. The restraint shown by the Laocoon sculptors is a product of their awareness of the rules of visual media. Virgil, as a poet, is bound by different rules.

2 The comparison of painting and poetry

Comparisons between painting and poetry were a staple of eighteenth-century critical theory, and Lessing entered these debates with vigour. Fundamentally, he distinguished between the spatial properties of painting and the temporal properties of poetry. Painting and sculpture are grasped at once by visual perception. But poetry is understood sequentially. Lessing based rules about the appropriate means of imitation for each art form on this comparison. Painters and sculptors are limited to a single moment that they can capture completely. This, in turn, dictates the specific rules that give painting and sculpture access to beauty. If the *Laocoon* sculptors had tried to follow Virgil by depicting the distorted face of a screaming Laocoon, the viewer would have been overwhelmed with disgust by the immediate effect, and they would have failed in their attempt to show what was happening because sculpture could not hope to capture the scene's movement and action. It is limited to a 'pregnant moment', and in order to capture that moment, the sculptors had to abstract from the visual scene those parts of it that would convey the ideal beauty and restraint that was their true object. Virgil, on the other hand, could show change and

development. Poetry imitates action, not objects. The sequence of words corresponds to a sequence of events. Thus the scream and suffering are immediately qualified by what goes before and comes after. Instead of evoking disgust, the emotions are changed and qualified. In each case, Lessing insisted, it is the aesthetic whole that must be considered.

In spite of Lessing's emphasis on the difference between poetry as an imitation of action and painting as a spatial presentation, and his insistence that each has a different set of rules which cannot be reduced to a common set, he did not completely separate the art forms. They have a common object, beauty, and a common effect, pleasure. Lessing's argument is that neither can attain this common objective without restricting itself to its own rules. In principle, he granted relative equality to the different art forms. In practice, however, poetry is clearly the superior form. Painting and sculpture are limited in two ways. The doctrine of the pregnant moment limits painting and sculpture not just to certain techniques but also to certain 'objects'. Painting has a limited number of subjects available to it. And, while painting and sculpture gain a certain immediacy from their presentational forms, that very immediacy restricts them to a concrete, visual presence that makes it difficult for them to escape the present moment. When painting becomes allegorical in order to try to enhance its significance, it violates its own spatial rules. Poetry, on the other hand, is not subject to such limitations. It appeals to the imagination, not actual vision. It can depict emotions as well as objects. And it can use its extensive, moving forms to go beyond the mundane to the significance of beauty itself. The temporal forms of poetry enhance its effects while the spatial forms of painting act as limits.

3 Natural and artificial signs

In distinguishing painting and poetry, Lessing makes use of a more fundamental distinction between natural and artificial signs. This distinction was set out clearly by Mendelssohn, who distinguished natural signs from arbitrary signs. He described natural signs as those that are connected to their objects by properties of the object. Arbitrary signs have no such natural connection and are linked to their objects only by convention (1771: 1, 437). For example, a blush is a natural sign of heightened emotion, and a picture is a natural sign of its object by virtue of its resemblance. But the choice of sounds and letters in language, while it may have natural origins, is essentially arbitrary. Thus painting uses natural signs, while poetry, which depends on language, uses arbitrary signs. Lessing's dramatic theory explores ways in which arbitrary signs for the imitation of action can be converted to natural signs when they form dramatic wholes. The theory is essentially rationalist; its goal is to account for how the intuition available in art and nature can be adequate to concepts and ideas. But Lessing, by focusing on the effects, turned the theory in the direction of the critical object.

4 Lessing and rationalist aesthetics

Lessing's aesthetics is best understood in the context of the rationalist aesthetic that was being developed at that time in Germany, particularly by Alexander Baumgarten and Georg Friedrich Meier, under the influence of Christian Wolff. Lessing himself implicitly located his own position in the preface to *Laocoon*. The amateur is in the position of the audience while the philosopher formulates general rules, and the critic applies them - sometimes correctly, sometimes erroneously. Lessing occupied that critical position, and he took from the philosophers - Spinoza, Leibniz, Wolff, Baumgarten, Mendelssohn and others - general aesthetic rules. But in applying those rules, he modified them to an extent of which he himself may not have been aware.

In many respects, Lessing presumes an aesthetic that can be derived more or less explicitly from Leibniz. Its object is beauty, which is understood as the perfection of a whole. Ideal beauty is found in distinct ideas whose clarity comes from their uniformity. Their reflection is found in the corporeal beauty which is a harmony of parts. Moses Mendelssohn defined beauty as the intuitive cognition of perfection. A cognition is intuitive, he wrote, when its object is immediately present to our senses or when we attend to the object rather than the sign for the object (1757: 1, 170). Lessing's formulation was similar except that Lessing was more concerned with concrete, critical applications to poetry and painting. This led him to distinguish the different possibilities provided by the differences in media. Painting is adapted to material beauty (1766: 104). Poetry, on the other hand, is adapted to a different kind of description: that of the presentation of emotions (1766: 111). Ultimately, painting and poetry have the same object: beauty. Poetry, however, has the advantage of presenting effects and thus of giving an apprehension of beauty itself, while painting can only show material beauty in its outward forms.

All rationalist aesthetics implicitly denigrates both art and nature. Beauty is grasped intuitively, but there are two forms of intuition: the sensory and the rational. Rational intuition is capable of apprehending ideas distinctly without the confusion of images and parts. Thus, it is superior to the confused sensory intuition of perception and the imagination. No rationalist aesthetic completely escapes this limitation on the products of art and nature. However, Alexander Baumgarten argued that sensate intuition was the superior form if one considered its effectiveness in guiding people to beauty, and Lessing followed Baumgarten in this respect. But Baumgarten considered sensate intuition essentially quantitative. He reasoned that since description adds to the completion of an idea, the more descriptive a painting or poetic passage, the more aesthetically perfect it is, and hence the more 'poetic' and beautiful. Lessing explicitly rejected this quantitative argument in favour of a distinction based on the effects that can be achieved. Poetry, on Lessing's view, should not strive for extended descriptions; Homer limits himself to a single trait (1766: 79). Poetry's ability to focus selectively and thus enhance its effects gives it the advantage over painting.

Thus, while Lessing remains clearly within the rationalist circle if one compares him to such sentimentalists and sense-theorists as Lord Shaftesbury or Francis Hutcheson, he departs in significant ways from the syllogistic and axiomatic descriptiveness of more orthodox rationalist formulations. Art is basically imitation, but for Lessing, imitation is directed towards action and emotion as well as objects. Lessing was thus led to consider both rules and means in a more concrete, inductive way than the axiomatic, deductive procedures of rationalist metaphysics had dictated. Rules are justified not by their logical form but by their effects. Since it is pleasure which is aimed at, one must be prepared to adjust the rules to the observed results. At this point, Lessing is very pragmatic. Empiricist aesthetics in England and France developed along psychological lines; it attempted to predict how certain emotional effects followed from artistic practice and natural beauty. Rationalist aesthetics, on the other hand, began with principles of unity and harmony and attempted to derive their application as a set of rules for the production of art and aesthetic effects. Lessing trusts neither. Emotions are not really subject to rules (1766: 28). On the contrary, rules are judged by their effects. At the same time, Lessing acknowledges the superiority of science to art. The object of science is knowledge; it is not subject to control by anything other than truth. But the object of art is pleasure. Since pleasure is not absolute, it may be judged by other standards and subjected to control (1766: 14).

5 Lessing's critical practice and aesthetic principles

In a number of ways, then, Lessing, as a critic who employs the philosophical 'rules' of his rationalist contemporaries, complicated and modified those rules and practices in ways that enhanced their critical application. In the process, he sometimes exhibited contradictory impulses. His definition of art included both his aesthetic tendencies. In the course of discussing ancient artefacts, Lessing wrote: 'I should prefer that only those be called works of art in which the artist had occasion to show himself as such and in which beauty was his first and ultimate aim' (1766: 55). The goal of art is beauty, and beauty can be understood in terms of harmony and perfection. But the other specific difference cited is that the artist had occasion to show himself, and that expressionist aesthetic is potentially at odds with the imitation implied in the reference to beauty. Lessing did not believe that imitation is pleasurable in itself. But he did hold that it produces pleasure as a result of the mental activity it stimulates. If feelings and emotions are the proper objects of imitation, Lessing was looking in a place far removed from his rationalist sources. Given his own distinction between critic and philosopher, Lessing is a critical rationalist for whom the role of the critic is central.

His aesthetic remains critically grounded, but it also universalizes:

> The purpose of art is to save us this abstraction in the realms of the beautiful, and to render the fixing of our attention easy to us. All in nature that we might wish to abstract in our thoughts from an object or a combination of various objects, be it in time or in place, art really abstracts for us, and accords us this object or this combination of various objects as purely and tersely as the sensations they are to provoke allow.
>
> (1767-8: 226)

Art is neither pure sensate representation nor absolute truth. The former is too chaotic, too confused. The latter is beyond our finite limits. Lessing's use of the theory of natural and arbitrary signs, his combination of imitation and expression, of rules and genius, of feeling and absolute form may not always be consistent, but it has as its object a

middle ground which produces a critical aesthetic of considerable practical power.

See also: Poetry §2; Tragedy §5

DABNEY TOWNSEND

List of works

Lessing, G.E. (1886-1924) *Sämtliche Schriften*, Lachmann; 3rd edn, trans. and ed. F. Muncker, Stuttgart, Berlin and Leipzig: Goschen, later De Gruyter, 23 vols. (The standard scholarly edition of Lessing's writings.)

Lessing, G.E. (1766) *Laocoon: An Essay on the Limits of Painting and Poetry*, trans. E.A. McCormick, Indianapolis, IN: Bobbs-Merrill, Library of Liberal Arts, 1962.(A central aesthetic text; the first part of an uncompleted three-part work.)

Lessing, G.E. (1767-8) *Hamburg Dramaturgy*, trans. E.C. Beardsley and H. Zimmern, in *European Theories of the Drama*, ed. B.H. Clark, rev. H. Popkin, New York: Crown, 1965.(Central critical texts. A series of periodical essays and a primary source for Lessing's dramatic theory and critique of Aristotle's Poetics in relation to the dramatic unities.)

Lessing, G.E. (1769) 'How the Ancients Represented Death', in *Death and the Visual Arts*, trans. E.C. Beasley and H. Zimmern, New York: Arno Press, 1977.(Reprint of 1879 translation. Polemical reply to an attack on *Laocoon*.)

Lessing, G.E. (1785) *The Spinoza Conversations Between Lessing and Jacobi*, trans. G. Vallee, J.B. Lawson and C.G. Chapple, Lanham, MD: University Press of America, 1988.(Includes translations and extended introductory discussion of the controversial report by F.H. Jacobi of his 1780 conversation with Lessing.)

References and further reading

Cassirer, E. (1951) *The Philosophy of the Enlightenment*, trans. F. Koelln and J.P. Pettegrove, Princeton, NJ: Princeton University Press.(Useful if somewhat exaggerated estimate of Lessing's influence on the development of aesthetics.)

McClain, J. (1986) 'Time in the Visual Arts: Lessing and Modern Criticism', *Journal of Aesthetics and Art Criticism* 44 (1): 41-58.(Discussion of Lessing's influence on critical treatment of time.)

Mendelssohn, M. (1757) 'Reflections on the Sources and Combinations of the Fine Arts and Letters', in *Gesammelte Schriften*, Stuttgart and Bad Cannstatt: Frommann, 1971.(A primary source for Hendelssoton's distinction between natural and conventional symbols and for his attempts to classify the arts.)

Mendelssohn, M. (1771) 'On the Chief Principles of the Fine Arts and Letters', in *Gesammelte Schriften*, Stuttgart and Bad Cannstatt: Frommann, 1971.(A revised version of the 1757 essay.)

Robertson, J.G. (1965) *Lessing's Dramatic Theory*, New York: Benjamin Blom.(Extended technical discussion of Lessing's stage theory.)

Rudowski, V.A. (1986) 'Lessing *contra* Winckelmann', *Journal of Aesthetics and Art Criticism* 44 (3): 235-44. (Lessing's relation to Winckelmann.)

Rudowski, V.A. (1971) *Lessing's Aesthetica in Nuce*, Chapel Hill, NC: University of North Carolina Press. (Includes text of 1769 letter to Nicolai.)

Wellbery, D.E. (1984) *Lessing's Laocoon: Semiotics and Aesthetics in the Age of Reason*, Cambridge: Cambridge University Press.(Theory is heavily semiotic, but also includes very good discussion of philosophical background. Citations from Mendelssohn above are Wellbery's translations.)

Leucippus (5th century BC)

The early Greek philosopher Leucippus was the founder of atomism. Virtually nothing is known of his life, and his very existence was disputed in antiquity, but his role as the originator of atomism is firmly attested by Aristotle and Theophrastus, although the evidence does not allow any distinction between his doctrines and those of his more celebrated successor Democritus. He wrote a comprehensive account of the universe, the Great World-System*. The single surviving quotation from his work asserts universal determinism.*

Leucippus was a shadowy figure even in antiquity. Virtually nothing is known of his life. His birthplace was disputed; some sources associated him with one or other of the two principal centres of early Greek philosophy, Elea and Miletus, others with Abdera, the birthplace of Democritus. The statement in a number of sources that he was a pupil of Zeno of Elea may be merely an inference from the Eleatic influence on the foundations of atomism (see Democritus §2). Of his dates, all that can be said is that he lived during the fifth century BC: he was almost certainly older than Democritus, who is described by Aristotle and Simplicius as his 'associate' (*hetairos*), implying that Democritus was his pupil (as is also attested by Diogenes Laertius). In the post-Aristotelian period he was overshadowed by Democritus to such an extent that atomism came generally to be regarded as the work of the latter, while Epicurus and his successor Hermarchus are reported to have denied that Leucippus ever existed. His historicity is, however, firmly attested by Aristotle, who is the primary source for atomism. No list of Leucippus' works exists. The catalogue of works of Democritus produced by Thrasyllus in the first century AD contains two books entitled *World-System* (*diakosmos*), respectively, the *Great World-System* and the *Small World-System*, with a note that the school of Theophrastus attributed the former to Leucippus. It is probable that that attribution is correct, the attribution to Democritus being due to the overshadowing process mentioned above. The single quotation surviving from Leucippus is said by its source, Stobaeus, to come from a work *On Mind*; it is uncertain whether that title refers to a section of the *Great World-System* or to a separate work. The pseudo-Aristotelian treatise *On Melissus, Xenophanes and Gorgias* refers to 'the so-called arguments of Leucippus', presumably a reference to a work purporting to be by Leucippus.

Aristotle, followed by Simplicius, sometimes attributes the fundamentals of atomism to Leucippus (see, for example, Aristotle, *On Generation and Corruption* 325a23, b7). Elsewhere, however, and sometimes even in the same context (for example, *On Generation and Corruption* 325a1-2) Aristotle refers to Leucippus and Democritus jointly, and not infrequently to Democritus alone (as was the standard practice of later writers). That Leucippus was in some sense the pioneer is undoubted, but it is impossible to determine what elements of the theory of atomism are to be attributed to either, or to what extent their enterprise was collaborative. (For discussion of the principles of atomism, see Democritus (§2).) Of the two most detailed reports of the atomists' account of the formation of worlds, one is attributed by Diogenes Laertius to Leucippus (IX 30-3), and is presumably a summary of the account in the *Great World-System*. It ascribes the formation of worlds to agglomerations of atoms which, formed by chance collisions, develop a rotation in which atoms of different sizes are sifted out, the smaller being extruded into the infinity of space, the larger forming a spherical structure which gradually differentiates itself into a cosmos as the larger atoms sink to the centre and the smaller are forced to the outside. The rotation of the cosmic aggregates was apparently unexplained, and was according to some sources attributed to chance. This raises the question whether Leucippus believed that some events, such as the cosmic rotation, are objectively random, or are merely ascribed to chance in the weaker sense that their cause is undiscoverable. The single quotation, 'Nothing happens in vain, but everything from reason and of necessity' (fr. 2, Stobaeus, I 4.4), suggests the latter. While it could in isolation be understood as asserting universal purposiveness in nature, the secondary sources, including Aristotle, are unanimous that the atomists denied all purposiveness in natural events. Hence the quotation is best understood as asserting universal determinism; 'in vain' is understood as 'for no reason', and 'of necessity' specifies the reason for which everything happens - namely, the irresistible force of causation. On that interpretation, the cosmic rotation must have had some determinate cause, but a cause undiscoverable to us.

C.C.W. TAYLOR

References and further reading

Aristotle (*c.*mid 4th century BC) *On Generation and Corruption*, ed. H.H. Joachim, Oxford: Oxford University

Press, 1922; trans. C.J.F. Williams, Oxford: Oxford University Press, 1982. (This work contains extensive critical discussions of the views of earlier philosophers on the basics of physical theory, for example, the elements, change and so on.)

Bailey, C. (1928) *The Greek Atomists and Epicurus*, Oxford: Clarendon Press, 64-108.(Seeks to individuate the contribution of Leucippus.)

Barnes, J. (1984) ‘Reason and Necessity in Leucippus’, *Proceedings of the First International Congress on Democritus*, Xanthi: International Democritean Foundation, 141-58.(An incisive discussion of the single remaining fragment.)

Barnes, J. (1987) *Early Greek Philosophy*, Harmondsworth: Penguin, 242-3.(Translates the fragment and an important text from Simplicius referring to Leucippus.)

Diels, H. and Kranz, W. (1951) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, vol. 2, 70-81. (The standard collection of the ancient sources; includes Greek texts with German translations.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols.(The most detailed and comprehensive English-language history of early Greek thought; a review of the evidence for Leucippus can be found in volume 2, pages 383-6, and a comprehensive discussion of atomism in volume 2, pages 389-507.)

Levinas, Emmanuel (1906-95)

In the 1930s Levinas helped to introduce the phenomenological philosophy of Husserl and Heidegger to the French. Subsequently his work attained classic status in its own right for his attempt to explore the meaning of ethics from a phenomenological starting-point. In Totalité et infini (1961) (Totality and Infinity, 1969) Levinas locates the basis of ethics in the face-to-face relation where the Other puts me in question. My obligations to the Other are not contracted by me. They not only precede any debts I incur, but also go beyond anything I could possibly satisfy. In later works, most notably Autrement qu'être (1974) (Otherwise than Being, 1981) Levinas explores further the preconditions of this account, most especially by investigating the I that was said to be put in question in the encounter with the Other. In analyses that stretch phenomenology to its limits and beyond, Levinas finds alterity within the self.

1 Life and philosophical development

Emmanuel Levinas was born in Lithuania. In 1923 he went to Strasbourg University where his studies included the philosophy of Henri Bergson. From 1928 to 1929 Levinas was at Freiburg University, where he studied first with Husserl and then with Heidegger. This led him in 1930 to publish *La Théorie de l'intuition dans la phénoménologie de Husserl* (*The Theory of Intuition in Husserl's Phenomenology*, 1973) a reading of Husserl that was informed by Heidegger's criticisms of Husserl's intellectualism. Since that time almost all of Levinas' philosophical works have taken their point of departure from either Husserl or Heidegger. If the practice of a philosopher like Levinas can be summarized in a phrase, one would have to say that he works at the limits of phenomenology by investigating the enigma behind the phenomenon. Whereas what lies hidden is, for Heidegger, primarily the being of beings, for Levinas, it is the excess of being, or in Plato's phrase, the good beyond being.

Levinas first presented his own thought in his essay on 'Evasion' (1935) and two short studies published immediately after the Second World War, *De l'existence à l'existant* (1947a) (*Existence and Existents*, 1978) and 'Le temps et l'autre' (1947b) (*Time and the Other*, 1987). These works offer analyses of nausea, fatigue and insomnia, as well as of the alterity found in death, fecundity and eros. They were already part of an attempt to go beyond the thought of being, although it was not until 'L'ontologie, est-elle fondamentale?' (1951) ('Is Ontology Fundamental?'), which was largely a criticism of Heidegger, that it became apparent that Levinas intended to issue his challenge to the dominance of ontology within Western philosophy specifically in the name of ethics and the alterity of the other human being. During the war, Levinas had been imprisoned in Germany in a camp for French soldiers, but on his release and in response to the Nazi persecution of the Jews, which none of the members of his family left in eastern Europe survived, he renewed his involvement with the Judaism of his childhood. From that time his studies developed along two trajectories. His main philosophical works, *Totalité et infini* (1961) (*Totality and Infinity*, 1969) and *Autrement qu'être* (1974) (*Otherwise than Being or Beyond Essence*, 1981), which increasingly stretch the language of Western philosophy to its limits, were accompanied by confessional writings, such as *Nine Talmudic Readings* (1990) and *À l'heure des nations* (1994) (*In the Time of the Nations*, 1988) which translate Hebraic wisdom into the language of Greece, by which he meant the universal language of the West. Although Levinas insisted, especially at first, that the two kinds of writing were entirely distinct and appealed to different kinds of evidence, their separation is far from straightforward. Terms like 'election', 'persecution' and 'hostage' that Levinas explored in his writings on the Talmud have found their way into his philosophical studies. Although Levinas has refused to describe himself as a Jewish philosopher, the unity of his thought lies in his attempt to address the nonsectarian question of ethics after the Holocaust by employing the resources of both phenomenology and Judaism. In this entry I focus on his contribution to ethics as the portion of his work that has attracted most attention.

2 *Totality and Infinity*

The guiding insight of *Totality and Infinity* is that the Other human being in their separation calls me into question. In the face-to-face relation my self-assurance disappears and I find myself in bad conscience. Levinas does not look to ethics to restore good conscience. On the contrary, I never have the luxury of knowing that I did the right thing in any given situation. Nor is it simply a question of knowledge. I can never do enough. My obligations to the Other are unlimited, as well as being asymmetrical in the sense that I have no right to ask of the Other what the

Other asks of me. In society, furthermore, every effort to meet my obligations to one person comes at the expense of my efforts on behalf of the others. Levinas refuses all the standard ways of diminishing the impact of these impossible demands, such as conceiving ethics in terms of intention or employing the model of legal responsibility where I am only answerable for what is in my power. Levinas has readily conceded that he does not provide an ethics but only an account of the condition of ethics. It should be added that what he describes as ethics is very different from what has hitherto gone under that name among philosophers.

The reciprocal system of obligations that usually passes for ethics most often goes under the name of justice in Levinas' thought. Levinas finds the system of justice to be in accord with reason, but he finds that, just as the I is put in question by the Other within the face-to-face relation, so justice is put in question by the face of the Other. The ethical relation with the Other exceeds the third-party perspective of neutral reason. As a result, Levinas is quite explicit that his account of the ethical is not constrained by what he calls 'formal logic'. Since Derrida's 1964 essay on Levinas, 'Violence et métaphysique' (1967) ('Violence and Metaphysics', 1978) questions have been raised about whether such an account is possible, thinkable and sayable. These enquiries, far from being uniformly negative, have led to some of the most far-reaching developments within continental philosophy, not least in the works of Derrida.

The way in which ethics questions justice also serves as an answer to the frequently heard criticism that Levinas' thought has little to offer to political theory. It is true that he has not said much about society, but his challenge to the good conscience of ethics applies equally well to politics. Levinas refers in one place to the tears invisible to the bureaucrat. In this very simple way he draws attention to what grand schemes overlook. Nor is the relation between ethics and justice one-sided in favour of the former. Just as ethics in Levinas' sense keeps justice from being satisfied with itself, so justice can also serve as a corrective to ethics. If there was only one other person in the world, I could devote myself to that person, but because I live in society, I can address my obligations to the Other only by neglecting the others. It is here that the place of reason and calculation becomes important also for ethics. Levinas found a formula that encapsulated the interruption of ethics by justice when he described how the third party, in the sense of the whole of humanity, looks at me in the eyes of the Other.

3 *Otherwise than Being*

In *Otherwise than Being* Levinas returned to the way that the Other questions the identity of the I. To be for the Other is to be without identity. This led Levinas to introduce a new structure that he calls 'substitution' and which he summarizes with Rimbaud's formula 'the I is an other'. Levinas also says that 'subjectivity is being hostage'. He means by this that subjectivity is not the isolated ego of modern philosophy, but the restlessness of being disturbed by the other. For Levinas, questions which would seem to many to be the most fundamental of ethical questions, such as 'Why does the other concern me?', 'Am I my brother's keeper?' or, with reference to Shakespeare's *Hamlet*, 'What's Hecuba to me?', have already come too late for ethics. These questions presuppose a concept of the ego which is derivative from my proximity to the Other. In this way Levinas makes more precise the goal, announced in 1951, of having ethics supplant ontology as first philosophy. In *Otherwise than Being* it is pursued largely by exploring the realm of selfhood that previous philosophy has left unmined.

There is some controversy as to whether the analyses of *Otherwise than Being* supplant or only supplement those found in *Totality and Infinity*. Certainly there are differences. The emphasis on separation in the earlier works gives way to an account of proximity which locates at the heart of selfhood a certain restlessness unconcerned with reciprocity. The foundational narratives of *Totality and Infinity*, reminiscent of social contract theory, that describe the first encounter that an I has with an Other, give way in *Otherwise than Being* to structures that are said to be 'older even than the a priori'. There is no longer an attempt to coordinate the ethical and the ontological: the former interrupts the latter. If *Totality and Infinity* describes how the ego is put in question by the Other, *Otherwise than Being* puts the concept of the ego in question, not by denying the substantiality of the subject as other continental philosophers have done, but by locating alterity within selfhood so that I can be said to be 'in myself' only through the others. Levinas himself has indicated that it was his rejection of the ontological language of *Totality and Infinity* that led to *Otherwise than Being*, but he did not renounce the earlier analyses, which he has often repeated in essays that postdate *Otherwise than Being*.

According to Levinas, the needs of the Other make demands on me that are impossible for me to fulfil. Most traditional ethical systems reject any such multiplication of my responsibilities on the grounds that it is destructive

of good conscience. Traditional ethical philosophies also have no place for Levinas' insistence that one is responsible even for what took place before one was born. They would see Levinas as extending the concept of responsibility to the point that one's sense of responsibility for what is within one's power is diminished. However, Levinas regards all such attempts on my part to limit specifically my responsibility as a reduction of ethics to justice. It amounts to limiting the demands I make on myself to the same standards that it is reasonable and proper for me to apply to others. From Levinas' perspective, what ordinarily passes for ethics is an evasion of the ethical, but he is well aware that it would be impossible to live constantly according to his conception of ethics. His account of the face-to-face is a description of being obsessed by the Other that amounts to a psychosis. Even though in interviews Levinas has questioned whether his work is properly called ethics, his questioning of the ontological basis of Western philosophy has succeeded in raising the question of whether what passes for ethics in that tradition has yet got to the bottom of the call to be ethical.

See also: Alterity and identity, postmodern theories of §3; Phenomenology, epistemic issues in

ROBERT BERNASCONI

List of works

Levinas, E. (1930) *La Théorie de l'intuition dans la phénoménologie de Husserl*, Paris: Felix Alcan; trans. A. Orianne, *The Theory of Intuition in Husserl's Phenomenology*, Evanston, IL: Northwestern University Press, 1973.(The first major study of Husserl in French, still valued as an introduction to his work.)

Levinas, E. (1935) 'De l'evasion' ('Evasion'), *Recherches Philosophiques* 5: 373-92.(Reissued under the same title in 1982 by Fata Morgana with an introductory essay and notes that extended it to a 121-page book.)

Levinas, E. (1947a) *De l'existence à l'existant*, Paris: Vrin; trans. A. Lingis, *Existence and Existents*, The Hague: Martinus Nijhoff, 1978.(Studies of fatigue, indolence, effort and horror that mark the beginning of his attempt to leave the climate of Heidegger's thought without regressing to a pre-Heideggerian position.)

Levinas, E. (1947b) 'Le temps et l'autre', in *Le Choix - le monde - l'existence*, ed. J. Wahl, Grenoble and Paris: Arthaud; trans. R.A. Cohen, *Time and the Other*, Pittsburgh, PA: Duquesne University Press, 1987.(The translation adds two more recent essays, including 'Diachrony and Representation' (1982), which is especially important.)

Levinas, E. (1951) 'L'ontologie, est-elle fondamentale?', *Revue de Métaphysique et de Morale* 56: 88-98.(A translation can be found in *Levinas: Basic Philosophical Writings* (1996).)

Levinas, E. (1961) *Totalité et infini*, The Hague: Martinus Nijhoff; trans. A. Lingis, *Totality and Infinity*, Pittsburgh, PA: Duquesne University Press, 1969.(The classic work that lays out Levinas' basic positions on ethics and alterity.)

Levinas, E. (1974) *Autrement qu'être ou au-delà de l'essence*, The Hague: Martinus Nijhoff; trans. A. Lingis, *Otherwise than Being, or Beyond Essence*, The Hague: Martinus Nijhoff, 1981.(Levinas' most important and most sophisticated work but also the most difficult.)

Levinas, E. (1982a) *De Dieu qui vient à l'idée*, Paris: Vrin.(The most important of over a dozen collections of essays. It includes the classic 'God and Philosophy'.)

Levinas, E. (1982b) *Éthique et infini*, Paris: Fayard; trans. R. Cohen, *Ethics and Infinity*, Pittsburgh, PA: Duquesne University Press, 1985.(Interviews with Philippe Nemo that serve as an excellent introduction.)

Levinas, E. (1988) *À l'heure des nations*, Paris: Minuit; trans. G.D. Mole, *In the Time of the Nations*, London: Athlone Press, 1994.

Levinas, E. (1990) *Nine Talmudic Writings*, trans. A. Aronowicz, Bloomington, IN: Indiana University Press. (Contains translations of *Quatres lectures talmudiques* (1968) and *Du sacré au saint* (1977).)

Levinas, E. (1996) *Levinas: Basic Philosophical Writings*, ed. A. Peperzak, R. Bernasconi and S. Critchley, Bloomington, IN: Indiana University Press.(A collection of Levinas' most important philosophical essays with introductory essays and notes added by the editors.)

References and further reading

Bernasconi, R. and Critchley, S. (eds) (1991) *Re-reading Levinas*, Bloomington, IN: Indiana University Press. (Includes essays by Derrida, Irigaray and many of the major commentators on Levinas focusing especially on Levinas' later thought and his account of the feminine.)

Derrida, J. (1967) 'Violence et métaphysique', *L'Écriture et la différence*, Paris: Éditions du Seuil; trans. A. Bass, 'Violence and Metaphysics', *Writing and Difference*, Chicago, IL: University of Chicago Press, 1978.(Revised version of the classic study from 1964 that has set the terms of subsequent discussion of Levinas and that may

have influenced Levinas himself.)
Llewelyn, J. (1995) *Levinas: The Genealogy of Ethics*, New York: Routledge.(Close readings of some of Levinas' major texts that is instructive for beginners and scholars alike.)
Peperzak, A. (1993) *To the Other*, West Lafayette, IN: Purdue University Press.(Best guide to Levinas for beginners.)

Lévi-Strauss, Claude (1908-)

Lévi-Strauss is one of the outstanding figures of mid-twentieth century intellectual life, influential far beyond the boundaries of France or the French language, and he has continued to write new work well into his eighties. His name is linked above all to the structuralist movement, of which he has been probably the most single-minded and unwavering exponent, and he was one of the key figures in the experiment of applying the insights of linguistics to the material of the social sciences. Through his work, public recognition of the discipline of anthropology grew dramatically to become an important element in discussion of philosophical issues.

The philosophical environment in France, in which structuralism developed and against which it reacted, was that of Sartrean humanism. In contrast to the existentialist emphasis on individual subjectivity, structuralism expected to find objective solutions to problems in the study of human beings. It was a form of intellectual modernism, a radical break with previous theoretical models and philosophical traditions, symptomatic of post-war optimism for the global applicability of science. It was hostile to metaphysics, bracketed the search for truth and was indifferent to the human subject. And, in contrast to Bergson's emphasis on continuity and flux, structuralism took discontinuity as its founding principle.

1 Life

Claude Lévi-Strauss was born in 1908. His grandfather was the rabbi of Versailles and his father was a painter. After studying philosophy in Paris he went to the new university of São Paulo in 1934 as professor of sociology, and during his five years in Brazil spent some time travelling in the interior. Returning for military service in 1939 he escaped to New York after the fall of France, and spent most of the Second World War lecturing at the New School for Social Research. Two decisive influences on his intellectual development derive from this period: first, contact with Roman Jakobson and second, exposure to the empirical approach of US anthropology. Since then he has lived and worked in Paris, mostly at the Collège de France, and became a member of the Académie française in 1973.

2 The impact of structural linguistics

Attending Jakobson's lectures on linguistics and phonology in 1942 seems to have been a revelation for Lévi-Strauss. Structural linguistics taught him that 'instead of being led astray by a multiplicity of terms, one should consider the simplest and most intelligible relationships uniting them' (1983, ch. 9: 139). The idea that the phoneme has no intrinsic meaning, but only acquires significance in relation to other similar elements, became the heart of his new approach to anthropological debates. The structural study of phonology emphasized two further principles that were to become fundamental to his work: the importance of unconscious mental activity in the production of language, and the fact that language is organized through discontinuous elements linked in particular ways.

Many commentators have linked Lévi-Strauss' structuralism directly to Saussure. However, it is clear from more recent writings that Lévi-Strauss' source of inspiration was Jakobson's interpretation of Saussure. His own use of Saussure was highly selective, emphasizing above all the arbitrary nature of the linguistic sign, the synchronic basis to linguistic analysis, and the syntagmatic and paradigmatic relations that link linguistic signs. These involve relations of contiguity and similarity, analogous to Jakobson's distinction between metonymy and metaphor (see Structuralism in linguistics).

3 The study of kinship

During this same period, Lévi-Strauss engaged in an extensive study of the ethnological literature, seeking to uncover the 'simplest and most intelligible relationships' that united all the wealth and diversity of the ethnographic record. His first area of interest was kin relations, a topic that linked him directly to the central concerns of anthropology and to Durkheim (§3) and Mauss' theories of social solidarity. His first book, *The Elementary Structures of Kinship* (1949), brought these different strands together in a vast synthetic work that transformed the way that anthropologists understood kinship.

Lévi-Strauss proposed that the incest taboo is analogous to the phoneme, in that it has no meaning of its own but

serves to link two domains. This is achieved by the exchange of women between kin-groups, which constitutes the origin of culture and is its ongoing pre-condition. Like all his work, his theory of kinship is controversial, but many of its insights have become canonical.

4 Totemism and *The Savage Mind*

Kinship led Lévi-Strauss to the theme of totemism and systems of classification, and the application of his theories to more obviously mental phenomena. He took issue with the argument, associated particularly with Lucien Lévy-Bruhl, that the thought processes of primitives are based on entirely different principles to our own. In *The Savage Mind* (1962) he showed how all human cultures employ the same principles to define and order their environment. In fact, 'the savage mind' is an unfortunate - if by now well-established - translation of *La pensee sauvage*, since the English expression implies that he is referring to some distinctive primitive mentality. All humans are alike, in his view, in the way that we rely fundamentally on sensual perception for our systems of classification and knowledge - the 'science of the concrete'. At the same time, primitives are just as capable of abstract thought as we are.

However, Lévi-Strauss does specify the significance of different levels of abstraction, expressed through his well-known contrast between the *bricoleur* and the engineer. The French *bricoleur* is a sort of jack of all trades, who puts things together with whatever materials lie to hand. Most of us for most of the time are *bricoleurs*, close to percepts, using the science of the concrete. By contrast, the 'engineer' operates at 'one remove' from observable properties, constructing concepts appropriate for the use to which they will be put.

In practice, Lévi-Strauss has devoted his life's work to studying the thought principles of *bricoleurs* rather than engineers. But implicitly his own work stands as perhaps a key example of what he means by the engineer. He believes that he has identified the abstract principles that underlie all everyday human thought processes, and that anthropology is particularly suited to the attainment of knowledge because it seeks the 'view from afar'. No civilization can think of itself unless it uses others as a basis for comparison. The importance which he attaches to detachment and distance is also at the heart of his critique of Sartre published as the final chapter of *The Savage Mind*.

5 The science of mythology

The most extensive, and probably the best-known, part of Lévi-Strauss' work is dedicated to the understanding of myth, another manifestation of the 'science of the concrete'. While myth uses the material world as means of thought, it is free from the constraints of material reality. Lévi-Strauss contrasts his own approach to myth with that of Jung, who in Lévi-Strauss' view erred by assuming that myths have stable meanings and a referential function. For Lévi-Strauss, the search for the manifest meanings of myths misses their true nature, which lies beneath the multifarious and often fantastic surface phenomena, where the human mind (*l'esprit humain*) rehearses unceasingly the opposition between nature and culture and the contradictions of human life, without offering a clear explanation or meaning.

Lévi-Strauss' approach to the analysis of myths discounts their narrative form in favour of identifying their component parts, in the spirit of the Russian Formalism of his mentor Jakobson, and others such as Vladimir Propp, author of the hugely influential *Morphology of the Russian Folk Tale*. To identify these discontinuous elements Lévi-Srauss in *The View From Afar* coined the concept of the 'mytheme' analogous to the phoneme - 'a purely differential and contentless sign' (1983: 145). These elements are linked by the principles of metaphor and metonymy, of syntagmatic and paradigmatic relations. The language of myths is, then, an unconscious one, rather similar in his view to that of the genetic code: far from being the consciously-created stories of individual narrators, 'myths operate in men's minds without them being aware of the fact' (*The Raw and the Cooked* (1964: 12).

As such, each myth is connected to others, not only within the same language and cultural group, but beyond that to myths recorded for far-away groups speaking different languages. In his four volume *Mythologiques* 1964-71) he traces mythical themes and structures from one end of the American continent to the other. Crucial to this enterprise is the concept of transformation: in Lévi-Strauss' hands this is a relatively static concept, referring typically to the way that myths are related through inversion, that is, the underlying structure of two apparently

divergent and unrelated myths contains the same elements, but arranged in a different way.

His concept of transformation - a formal rather than a dynamic one - typifies his temporal concepts. Synchrony, the suspension of temporal movement, was a guiding principle for structuralism, and for Lévi-Strauss synchrony expresses a core value. Music has provided a significant analogy, for example in *The Raw and the Cooked* where the temporal unfolding of music (and by analogy the narrative unfolding of myths) is subordinate to the structural principles which unite a piece of music, and link it to other pieces.

6 Lévi-Strauss' theory of knowledge

True knowledge for Lévi-Strauss arises from contemplation. In this respect pre-literate societies, where mythical structures are found in their purest form, come closest to his ideal. They typically strive to negate the effects of history and change, in contrast to the modern world where these are celebrated as good in themselves. In many of his writings Lévi-Strauss emphasizes the unity of the human mind across time and culture, stressing the similarity of purpose and method between our distant ancestors who domesticated animals and plants and our own scientific concerns (especially *Structural Anthropology II* (1973, ch.18). In others he implicitly attributes cognitive superiority to the 'engineer' or indeed to the anthropologist. And yet in other respects it is clear that he admires the disappearing world of the primitives who understood the significance of synchrony.

His passionately argued belief that the primitives were as much intellectuals as we are represents a grand if flawed attempt to create a unified theory of the human mind based on detailed comparative work. It derives from a model of human nature which is grounded in intelligence (in contrast to other models which privilege the emotions and instincts, or self-interest and strategy). However for all his intellectualism, many of Lévi-Strauss' ideas have stood the test of time more for their aesthetic insights and intuitions, than for his theoretical elaborations. As a number of commentators have pointed out, the whole elaborate edifice of structural linguistics conceals as much as it elucidates his aims and methods. He is closer to literature, to music, and above all to painting, and in a late work comments that 'the suppression by chance of ten or twenty centuries would not appreciably affect our knowledge of human nature. The only irreplaceable loss would be the works of art' 1993: 176).

In his analysis of actual myths, Lévi-Strauss' rigorous formalism and implicit reductionism is complemented by an extraordinary grasp of empirical detail and a sensitive attention to the distinctive properties - visual, tactile, olfactory, chemical, aural, functional - of whatever it is that the myth purports to be about. This attention to detail partially undermines the importance which he himself attaches to the Saussurean principle of the arbitrariness of the sign. It was clearly an important notion which enabled him to get away from the referential functions of language, but it should not blind us to the fact that most of the mythical signs which he has devoted the last thirty years of his life to analysing are not arbitrary in the Saussurean sense at all.

See also: Anthropology, philosophy of; Structuralism; Structuralism in literary theory; Structuralism in social science

OLIVIA HARRIS

List of works

Lévi-Strauss, C. (1949) *Les structures elementaires de la parente*, Paris: Presses Universitaires de France; *The Elementary Structures of Kinship*, Boston, MA: Beacon Press, 1969.(A vast synthetic work which transformed the way that anthropologists understood kinship.)

Lévi-Strauss, C. (1955) *Tristes Tropiques*, Paris: Plon; *Tristes Tropiques*, London: Penguin, 1976.

Lévi-Strauss, C. (1958) *Anthropologie Structurale*, Paris: Plon; *Structural Anthropology*, New York: Basic Books, 1963.

Lévi-Strauss, C. (1962) *Le Totemisme Aujourd'hui*, Paris: Presses Universitaires de France; *Totemism*, Boston, MA: Beacon Press, 1963.

Lévi-Strauss, C. (1962) *La pensee sauvage*, Paris: Plon; *The Savage Mind*, Chicago, IL: Chicago University Press, 1966.(Shows how all human cultures employ the same principles to define and order their environment.)

Lévi-Strauss, C. (1964-71) *Mythologiques: Introduction a la Science de la Mythologie (Mythologiques: Introduction to the Science of Mythology)*, 4 vols; vol. 1 *Le cru et le cuit*, Paris: Plon, 1964, *The Raw and the Cooked*, New York: Harper & Row, 1969; vol. 4 *L'homme nu*, Paris: Plon, 1971, *The Naked Man*, London: Cape, 1981.

Lévi-Strauss, C. (1973) *Anthropologie Structurale II*, Paris: Plon; *Structural Anthropology II*, New York: Basic Books, 1976.(Examines the unity of the human mind across time and culture, stressing the similarity of purpose between our distant ancestors who domesticated animals and plants and our own scientific concerns.)
Lévi-Strauss, C. (1978) *Myth and Meaning*, London: Routledge.
Lévi-Strauss, C. (1983) *Le regard eloigne*, Paris: Plon; *The View from Afar*, New York: Basic Books, 1985.
Lévi-Strauss, C. (1993) *Regarder, Ecouter, Lire (Looking, Hearing, Seeing)*, Paris: Plon.

References and further reading

Leach, E. (1970) *Lévi-Strauss*, Modern Masters Series, London: Fontana.(Sympathetic but not reverential account of the main principles of Lévi-Strauss' structuralism from an anthropological perspective. Early but still worth reading.)
Merquior, J.G. (1986) *From Prague to Paris. A Critique of Structuralist and Post-structuralist Thought*, London: Verso.(Locates the philosophical and intellectual environment of Lévi-Strauss' work clearly and engagingly.)
Pace, D. (1983) *Claude Lévi-Strauss: The Bearer of Ashes*, London: Routledge. (Well-written comprehensive overview of Lévi-Strauss' work and its philosophical implications.)
Pavel, T. (1989) *The Feud of Language: A History of Structuralist Thought*, Oxford: Blackwell.(An excellent highly critical account of the whole structuralist enterprise, dense and requiring some previous understanding.)
Sperber, D. (1979) 'Claude Lévi-Strauss', in J. Sturrock (ed.) *Structuralism and Since*, Oxford: Oxford University Press.(A very useful critical but sympathetic synthesis of his *oeuvre*, by one of his most gifted students.)
Todorov, T. (1988) 'Knowledge in Social Anthropology: Distancing and Universality', *Anthropology Today* 4 (2): 2-5.(A subtle and complex exploration of Lévi-Strauss' theory of knowledge.)

Lewis, Clarence Irving (1883-1964)

The American philosopher C.I. Lewis held that in all knowledge there are two elements: that which is presented to sense and the construction or interpretation which represents the creative activity of the mind. Contrary to Kant, Lewis claimed that what is fixed and unalterable is not the structure that we bring to the sensibly presented, but rather the sensibly presented itself. The categories that mind imposes do not limit experience; they determine the interpretation we place upon experience, and if too much of experience eludes our categorizations, new ones should be established. It is pragmatically necessary that we create interpretive structures which will work in getting us around in sensory experience. This important and novel doctrine, Lewis' 'pragmatic a priori', emerged through the development of ideas which took root during his study of logic. The problems of choosing among alternative logics led him to assert the need for pragmatic criteria. The way we conceptually structure or categorize experience answers to pragmatic criteria of purposes, intents and interests. Only within a context defined by a priori categorizations can empirical judgments be made. These empirical judgments proceed from apprehensions of the sensibly presented to assertions of objectivities. Moral judgments require both judgments of good and decisions of right. Judgments of value are tied to qualitative satisfactions disclosed in experience and are empirical claims. Decisions about the morally right are based on imperatives of reason.

1 Life

Clarence Irving Lewis, an American philosopher who focused on epistemology, logic and ethics, was born in 1883 in Stoneham, Massachusetts, and died in Menlo Park, California, in 1964. He was educated at Harvard, receiving his BA in 1905 and his Ph.D. in 1910. After receiving his Doctorate, he taught at the University of California from 1911 to 1920, and then at Harvard, where from 1930 he was the Edward Peirce Professor of Philosophy until his retirement in 1953.

2 Logic and the a priori

Lewis' important and novel doctrine of the pragmatic a priori emerged through the development of ideas which took root during his study of logic. This work in logic, combined with a healthy respect for Kantian epistemology, a long exposure to Roycean idealism and an appreciation of certain basic tenets of classical American pragmatism, produced the context from which the pragmatic a priori, the vital core of Lewis' conceptual pragmatism, took shape (see Royce, J. §3; Pragmatism §2).

Lewis spent many years in the study of logic, disturbed mainly by the conception of implication developed in the extensional logic of Whitehead and Russell's *Principia Mathematica* (1910-13) (see Russell, B.A.W. §3; Whitehead, A.N. §2). This 'material implication' deviated strikingly from the ordinary sense of implication. According to it, a false proposition implies any proposition, while a true proposition is implied by any proposition. The problem, Lewis held, lay in the fact that the logic of propositions formulated in *Principia* is an extensional one, while ordinary deductive inference depends upon the meanings of the propositions used, and hence is rooted in intensional relations. The solution to this problem led Lewis to the development of the system of strict implication in symbolic logic and carried him beyond logic into the field of epistemology and the development of a detailed theory of meaning and analyticity.

A second type of issue, arising out of the entertainment of the possibility of an alternative to the logic of material implication, led to his interest in the existence of various types of alternative logics such as many-valued logics (see Many-valued logic §1; Modal logic §1). It also carried him beyond logic to the development of a theory of knowledge asserting the free creation of and pragmatic selection among various possible a priori conceptual schemes as tools of interpretation of experience.

Lewis held that the behaviour of symbolic systems operates in the same way as the behaviour of the human mind, that there is nothing in them which we have not put in ourselves, but that they teach us the meaning of our commitments. Thus, a priori truth is independent of experience because it is purely analytic of our conceptual meanings. The line between the a priori and the a posteriori, then, coincides with the divisions between the conceptual and the empirical, between the contributions of mind and what is given in experience, between the analytic and the synthetic.

3 Meaning and analyticity

According to Lewis, the meaning or deductive significance of a proposition, which can be thought of as a kind of term, coincides with the intension of the proposition, in the sense that whatever is deducible from the proposition is contained in its intension. A proposition signifies a state of affairs. A state of affairs is not a chunk of reality to be denoted, but rather includes all that the assertion of the state of affairs as actual implies, and only what it thus implies.

All real definitions specify intensional modes of meaning. What Lewis here means by a real definition is neither symbolic convention nor dictionary definition, but rather an explicative statement which relates a meaning to a meaning. There are two types of meanings related in such real definitions: linguistic meaning and sense meaning. Linguistic meaning is the pattern formed by the relation of a term or proposition to other terms or propositions. Sense meaning is the criterion in mind which determines the application of the term or proposition; it is the criterion by which what is meant is to be recognized.

Analytic truths state relations between sense meanings and not merely between linguistic meanings. Lewis sets out to undercut the conventionalist position that analytic truth expresses nothing beyond what is or can be determined by the language system which embodies it. Although he holds that there are conventional elements in the choice of symbols, in the assignment of the symbols to the meanings and in the choice of the meanings to be considered, he insists that the interrelation of the meanings is neither linguistic nor arbitrary. The analyticity of the linguistic meaning is determined by the fixed intensional relationships of sense meanings.

Lewis' position reveals both a sympathy towards, and yet sharp criticism of, Kant's epistemology. This critical appreciation of the insights of Kant is nowhere more in evidence than in his rejection of Kant's synthetic a priori, a rejection based directly on Lewis' own use of the Kantian concept of schematism (see Kant, I. §7). A sense meaning as a schema is a rule or prescribed routine and an imagined result of it which will determine the applicability of the word. By this means Lewis proposes to solve the epistemological problem of the nature of the necessity of analytic truth and how we know such necessity. We perform an experiment in imagination. We know that 'All squares are rectangles' because in envisaging the test which a thing must satisfy if 'square' is to apply to it, we observe that the test it must satisfy if 'rectangle' is to apply is already included. Both Lewis' theory of meaning and his notion of the test schema reveal that his theory of analyticity is basically one of inclusion or containment.

Empirical meaning, then - meaning as a criterion in mind by reference to which the applicability of a term or the believability of an empirical statement is to be attested in experience - is a mode of intensional meaning. Lewis emphatically rejects as epistemologically untenable the nominalist conception that individuals are the first knowables and are primitively determinable by ostensive reference. It is only by reference to intensional meaning as criterion in mind by which one applies or refuses to apply a term that denotation is possible.

4 The a priori and pragmatic considerations

The truth of any empirical generalization is determined by the relationship of a conceptual scheme to given experience, and remains true, relative to that scheme. Empirical truth does not literally change, nor are contexts of interpretation proven false. Rather, conceptual schemes which do not work adequately as interpretive contexts are replaced by others. The sense in which truth is 'made by the mind' and is 'relative to human interest and purpose' is the sense in which we choose the interpretive structures by which we prescribe the outlines of realities of particular types and in so doing set the context for the discovery of empirical truth. However, the analytic or a priori truth contained in the interrelationships of the interpretive structures is subject neither to the contingencies of experience nor to the whims of convention; rather, it is eternally true. Such truth is not material truth but formal truth or validity.

Lewis holds not only that our choice of an analytic, a priori conceptual scheme is conditioned by experience in that it is based on pragmatic considerations operative in the light of past experience, but also that the logic which the conceptual schemes apply and by which they are interrelated is itself based on pragmatic considerations and hence is, in the last analysis, conditioned by experience. Lewis' pragmatism runs quite deep. Humans, for Lewis, are essentially acting beings. Meanings which mind entertains, the logic which explicates such meanings, and mind itself emerge from behavioural responses to the environment in which humans find themselves. Our ways of

behaving towards the world around us which are made explicit in our accepted logic are those which have lasted because they work. The final ground, even of the validity of the principle of consistency as well as the validity of ordinary inference, is firmly rooted in a pragmatic, evolutionary-based direction of successful activity. The principle of consistency, and that ordinary inference which explicates our meanings in accordance with the principle of consistency, is rooted in a ‘pragmatic imperative’: if it is rejected, then thought and action themselves become stultified.

Lewis’ concept of the a priori occupies a unique position in the debate concerning the nature of a priori knowledge and the very possibility of an analytic-synthetic distinction. Drawing on a fundamentally Kantian scheme made responsive to the insights of American pragmatism and adapted to fit the needs of contemporary logic, Lewis has established an a priori which is coextensive with the analytic, yet which cannot be said to be empirically vacuous. It both arises from experience and has possible reference to experience.

5 Empirical knowledge

Lewis distinguishes three levels of empirical statement which enter into the formulation of empirical knowledge: expressive utterances, terminating judgments and objective beliefs.

Expressive statements, or the expressive use of language, signify appearances or indubitable content. They neither assert nor deny any objective reality of what appears; neither do they predict anything. Because they are confined to description of the immediate content of presentation, they cannot be in error. The immediate content is not judged, it is ‘had’.

Terminating judgments state the prediction of a particular passage of experience. They find their cue in what is given, but state something taken to be verifiable by some test which involves a way of acting. Since terminating judgments are expressed in expressive language, if the judgment is true, its truth is known with certitude, since there is no further means to bring it into question.

Finally, there are non-terminating judgments or judgments of objective fact. Such judgments express objective beliefs which can never be completely verified but are always further verifiable. Because such beliefs can never be completely verified they are probable, not certain. Non-terminating judgments are expressed in the form, ‘If this is a physical object O, then if S appearance and action A occur, then in all probability E appearance will occur’. The major, or first, implication relation of the non-terminating judgment expresses the logical entailment of strict implication. The second implication relation of the non-terminating judgment, or in other terms the implication relation holding within the terminating judgment, expresses what Lewis calls ‘real relations’. These real relations are causal relations involving contrary-to-fact conditionals and probability relations.

Our physical-object concepts embody habits of belief, and a habit is not just a collection of possible appearances and responses to these appearances; it is a rule of organization of possible appearances and possible responses which, in its very being as a rule, contains the possibility of generating more explicit relationships than can ever be enumerated. The meaning of objectivity is something over and above a collection of related appearances, and this meaning is supplied by the fixed rule of organization. Thus, though a physical-object statement is confirmable by verifying experiences, it is not reducible to or totally translatable into the verifying experiences.

Perhaps no single aspect of Lewis’ philosophy has been subject to more frequent and diverse attacks than his concept of the given element in experience. He expressed an ongoing frustration that his understanding of the given had been so misinterpreted over the years, stressing that the point he is trying to make is so obvious that he wonders how anyone could contest it. If there is nothing given, there would be no content for thought, nor could there be success or failure in action, and prediction would be incomprehensible. His emphasis on the given, as he himself stresses, is not a foundationalist or phenomenalist concern about data from which we build up a world of objects, but a pragmatic concern with the way in which we proceed to verify our beliefs. If nothing is to be counted as certain, then probable knowledge is not possible, for knowledge would be merely ‘probably probable’. The probability involved in our predictive claims must be founded in something that is not subject to probability claims. The sensibly presented, as apprehended, is a ‘taken’, but is pragmatically certain in that it serves as the test for our various claims, but cannot itself be tested since it makes no reference to future experience.

6 Ethical writings

Lewis' moral theory is an extension of the view developed in his philosophy of knowledge. There are two dimensions involved, a given element and a priori categorization. Lewis rejects the sceptical view that evaluations are emotive and unverifiable, and asserts instead that they are a species of empirical knowledge.

What is intrinsically valuable or valuable for its own sake is the quality of an experience. An experience further has contributory value through the contribution it makes to the full quality of the life of the person having it. These are reported in expressive terms, leading to decisively verifiable terminating judgments, and moving on to objective probability statements which are judgments of extrinsic value and attribute a property of value to an objectivity of some kind. Extrinsic values may be inherent or instrumental. The former lead directly to experiences which are intrinsically valuable, while the latter arise as means to an end that is inherently valuable. Judgments about right and wrong are not empirical, but rather are determined by principles which in their prescriptive character refer to values. The fundamental rational imperative is to think and act in a way that one will not later regret. Rationality requires that actions must be in accordance with objective knowledge of relationships productive of value experience in the future for all, rather than in accordance merely with the felt quality of one's immediate experience.

See also: Intentionality

SANDRA B. ROSENTHAL

List of works

Lewis, C.I. (1970) *Collected Papers*, ed. J.D. Goheen and J.L. Mothershead, Jr, Stanford, CA: Stanford University Press.(All of Lewis' published essays and his most important unpublished papers.)

Lewis, C.I. (1918) *A Survey of Symbolic Logic*, Berkeley, CA: University of California Press.(As far as possible, deals with diverse systems and important developments of logic within a common notation.)

Lewis, C.I. (1929) *Mind and the World Order*, New York: Charles Scribner's Sons; repr. New York: Dover, 1956. (Conceptions presented here grew out of his interest in exact logic; includes detailed presentation of his pragmatic a priori.)

Lewis, C.I. (1946) *An Analysis of Knowledge and Valuation*, La Salle, IL: Open Court.(A study of meaning and analytic truth, empirical knowledge and an application of these to issues of valuation.)

Lewis, C.I. and Langford, C.H. (1932) *Symbolic Logic*, New York: Appleton Century. (A technical study of symbolic or mathematical logic.)

Lewis, C.I. (1955) *The Ground and Nature of the Right*, New York: Columbia University Press.(Fundamental issue of 'the right' in various modes as these bear upon ethics.)

Lewis, C.I. (1957) *Our Social Inheritance*, Bloomington, IN: Indiana University Press. (Public lectures delivered at Indiana University in 1956.)

Lewis, C.I. (1969) *Values and Imperatives (Studies in Ethics by C.I. Lewis)*, ed. and with intro. by J. Lange. (Consists of four interrelated lectures delivered in 1959, plus six others, four of which were never before published.)

References and further reading

Flower, E. and Murphey, M. (1977) *A History of Philosophy in America*, New York: Capricorn Books and G.P. Putnam's Sons, 2 vols, 891-958.(A broad overview of Lewis' position, including his moral theory.)

Kuklick, B. (1977) *The Rise of American Philosophy: Cambridge, Massachusetts, 1860-1930*, New Haven, CT, and London: Yale University Press, 533-62.(An overview of Lewis' position within the context of American philosophy and philosophy at Harvard University.)

Rosenthal, S.B. (1976) *The Pragmatic A Priori: A Study in the Epistemology of C.I. Lewis*, St Louis, MO: Warren H. Green.(Main focus on Lewis' epistemology, but broadens to include metaphysical implications; pragmatic dimension stressed throughout.)

Saydah, J.R. (1969) *The Ethical Theory of Clarence Irving Lewis*, Athens, OH: Ohio University Press.(A detailed study of Lewis' ethical position.)

Schilpp, P.A. (ed.) (1968) *The Philosophy of C.I. Lewis*, The Library of Living Philosophers, vol. 13, La Salle, IL: Open Court.(Various papers discussing Lewis' work, with a response by Lewis.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press; 2nd edn, 3 vols, 1927.(Development based on extensional logic and the relation of material implication.)

Lewis, Clive Staples (1898-1963)

C.S. Lewis was a British religious writer. Originally trained as a philosopher at Oxford, he combined literary scholarship and the writing of fiction with clear and persuasive argumentation for traditional Christianity. His religious works continue to be best sellers, and much of his writing is directly or indirectly of philosophical interest. For the non-philosophical public, he probably remains the best representative of the position that religion is more rational than any alternative. Lewis is regarded by his admirers as more than a philosopher: in the words of Walsh (1949), he is the twentieth century's 'apostle to the skeptics'.

The most interesting of Lewis' arguments for the existence of God is one sometimes called 'The Mental Proof'. One way of getting a preliminary insight into this argument is to ask whether nature is a product of mind, or mind is a product of nature. If God created nature, as Christians believe, then nature is understandable by reason because it is a product of reason. It sounds superficially plausible to say that if mind were the product of nature, nature would be understandable by mind as well. But in fact, there is no acceptable general rule that a product can understand what produces it.

Furthermore, if nature is seen as non-rational, as without intelligence or purpose of its own, a special difficulty arises. One thing may produce another by design, by chance or by the unfolding or development of some process already inherent in the producing agent. But if nature is without intelligence or purpose, it cannot produce mind by design, nor can it produce it as the unfolding or development of something already there. So a mindless nature can produce mind only by chance. But if mind is only a chance product of nature, how can we trust our reasoning powers, how can we expect our minds to give us the truth about anything? In fact, Lewis slightly overstates his point in the first edition of *Miracles* (1947): he says that when we find that a position has been caused by non-rational factors, such as chance, we should immediately dismiss it.

In the same work, Lewis also makes some incisive criticisms of Hume's arguments against miracles, that 'there must be a uniform experience against every miraculous event'. First, says Lewis, Hume begs the question by assuming at the outset that there is 'uniform experience' against miracles. Second, the question of whether the universe is such as to allow miracles is not a question that can be settled just on the basis of experience. The most compelling reason for the uniformity of nature is that nature was the result of design by an intelligent being. And if we admit that, then the being who designed the laws of nature can suspend them (see Miracles).

A number of Lewis' arguments for God exhibit the same pattern: a fact of experience is vividly described, then explained by Christian theism. For example, our experience of recognizing the claims of morality and at the same time failing to meet those demands is explained by a God who is the source of morality and a human race estranged from that God (see Religion, history of philosophy of §5). Or our experience of longings which can be satisfied by no finite object, but which seem to indicate a real need calling out for a real object can be best explained by a God who is our final good. On Lewis' view, attempts to explain these facts of our experience without bringing in God lead to our explaining them away.

Lewis also gives an argument which has been called the 'Lewis trilemma' (see 'The Shocking Alternative' in *Mere Christianity*). He argues that the claim by Jesus to be God cannot be separated from the New Testament accounts of his life and work; for example, his claim to forgive sins on his own authority could only be made by God. Lewis then poses three alternatives: Liar, Lunatic or Lord. But Jesus is recognized by almost everyone as a great moral teacher: could such a teacher lie or be subject to an insane delusion? Lewis concludes that seeing Jesus *only* as a great moral teacher is not a possibility left open by the evidence; he must be more or less than this.

See also: God, arguments for the existence of

RICHARD L. PURTILL

List of works

Lewis, C.S. (1940) *The Problem of Pain*, London: Geoffrey Bles; New York: Macmillan, 1943.(An examination of the problem of evil.)

Lewis, C.S. (1942) *The Screwtape Letters*, London: Geoffrey Bles; New York: Macmillan, 1943.(An insightful analysis of moral psychology.)

Lewis, C.S. (1943) *The Abolition of Man*, Oxford: Oxford University Press; New York: Macmillan, 1947.(Revives a classical approach to moral philosophy.)
Lewis, C.S. (1945) *The Great Divorce*, London: Geoffrey Bles; New York: Macmillan, 1945.(A fresh approach to traditional religious doctrines.)
Lewis, C.S. (1947) *Miracles*, London: Geoffrey Bles; New York: Macmillan.(Re-examines important issues in the philosophy of religion.)
Lewis, C.S. (1952) *Mere Christianity*, London: Geoffrey Bles; New York: Macmillan.(A very popular argument for Christianity.)
Lewis, C.S. (1955) *Surprised by Joy*, London: Geoffrey Bles; New York: Harcourt, Brace & World, 1956.(A religious autobiography exploring one kind of religious experience.)
Lewis, C.S. (1960) *The Four Loves*, London: Geoffrey Bles; New York: Harcourt, Brace & World, 1960.(A discussion of moral psychology.)

References and further reading

Beversluis, J. (1985) *C.S. Lewis and the Search for Rational Religion*, Grand Rapids, MI: Eerdmans.(An attack on both Lewis and rational arguments for religion.)
Cunningham, R.B. (1967) *C.S. Lewis: Defender of the Faith*, Philadelphia, PA: Westminster Press.(An analysis of Lewis' apologetics by a theologian.)
Kreeft, P. (1969) *C.S. Lewis: A Critical Essay*, Grand Rapids, MI: Eerdmans.(The first analysis of Lewis' arguments by a philosopher.)
Purtill, R.L. (1981) *C.S. Lewis's Case for the Christian Faith*, New York: Harper & Row.(A favourable assessment of Lewis' arguments.)
Walsh, C. (1949) *C.S. Lewis: Apostle to the Skeptics*, New York: Macmillan.(The first full-length study of Lewis' works.)

Lewis, David Kellogg (1941-)

David Lewis has made extremely important and influential contributions to many topics in metaphysics, philosophical logic, the philosophy of science, the philosophy of mind, the philosophy of language, the philosophy of probability, rational decision theory, and ethics and social philosophy. His work on counterfactuals and the philosophy of modality has been especially influential.

1 Life

David Kellogg Lewis was born on 28 September 1941 in Oberlin, Ohio, where his father taught government at Oberlin College. His mother, a historian, did some teaching for the college, but did not hold a regular appointment. Lewis attended Swarthmore College, from which he received a BA in philosophy (high honours) in 1962. His postgraduate studies were at Harvard, where he received an MA in 1964 and a Ph.D. in 1967. He served as Assistant Professor of Philosophy at UCLA from 1966 to 1970, and has since taught at Princeton. Lewis has for many years had a special relationship with Australia, where he has spent two or three months of every year, usually without formal academic appointment, since 1979. In September 1995 he was awarded an honorary doctorate by the University of Melbourne.

2 Counterfactual conditionals

In 'A Theory of Conditionals' (1968), Robert Stalnaker proposed the following semantical analysis of counterfactual conditionals: the counterfactual having antecedent *p* and consequent *q* (' $p \rightarrow q$') is true if and only if either *p* is impossible or *q* is true in the possible world closest to actuality in which *p* is true (or '*q* is true in the *p*-world closest to actuality'). (Stalnaker's analysis assumes that for any possible world *w* and any possible proposition, there is a unique world that is the closest world to *w* in which that proposition is true.) One important consequence of this analysis is that any sentence of the form '($p \rightarrow q$) or ($p \rightarrow \mathrm{not}\text{-}q$)' - for example, 'If Lincoln had not been assassinated, Hitler would (still) have invaded Poland in 1939, or else if Lincoln had not been assassinated, it would not have been the case that Hitler invaded Poland in 1939' - must be true. Lewis proposed an alternative analysis (1973a), which replaces the second disjunct of Stalnaker's analysis with the condition that there is some world *w* in which both *p* and *q* are true, such that every *p*-world at least as close to actuality as *w* is also a *q*-world.

On reflection, it seems that a counterfactual with a possible antecedent can be true even if there is no one closest world in which that antecedent is true. Suppose, for example, that there are two or more *p*-worlds that are equally close to the actual world, that all worlds closer to actuality than these are not-*p*-worlds, and that *q* is true in all these 'tied' *p*-worlds. It seems evident that $p \rightarrow q$ is true in this case. Lewis' analysis entails that a counterfactual is true if its antecedent and consequent are so related, and it avoids the counterintuitive consequence that '($p \rightarrow q$) or ($p \rightarrow \mathrm{not}\text{-}q$)' is a valid sentence-schema. For these reasons, it would be generally conceded that the logic of counterfactual conditionals that is generated by Lewis' semantics is much more intuitive than the logic that is generated by Stalnaker's semantics; most philosophers of logic, in fact, would probably be willing to say that Lewis' logic is the correct logic of counterfactuals.

There would be less agreement about the value of Lewis' informal, philosophical account of the basic items in his semantics: possible worlds and the 'closeness' relation. Lewis' idiosyncratic account of 'possible worlds' will be examined below. His account of 'closeness' is given in terms of comparative similarity - the concept illustrated by the comparison 'San Francisco is more like Boston than it is like Calcutta'. According to Lewis, the worlds closest to *w* are the worlds most similar to *w*. There has been considerable controversy about whether Lewis' truth-conditions for counterfactual conditionals are correct if closeness is understood in terms of similarity (see Counterfactual conditionals).

3 Modality: logic and ontology

Possible worlds are at the centre of Lewis' philosophy of modality. By the late 1960s it had become common to appeal to 'possible worlds' in discussions of modality. But many philosophers doubted whether any coherent philosophical sense could be made of the notion, particularly if each 'possible world' was to be associated with a well-defined set of individuals, those existing in it. By what criterion, it was asked, could an individual existing in

one world - say, Quine's 'as he is in the actual world' - be judged identical with an individual existing in some other world and having a set of properties very different from Quine's?

Some philosophers think of possible worlds as abstractions of some kind, perhaps as internally consistent and 'complete' stories, the actual world being the one that is true. Such philosophers reply to sceptics about the intelligibility of 'cross-world identity' in this way: Quine can 'turn up in' various stories, stories both true and false, even if the false stories ascribe to him properties very different from the properties ascribed to him by the true stories. According to Lewis, however, a world is no mere story. A world is a spatiotemporally connected and closed object (or at least an object closed under some relations that, like spatiotemporal relations, are 'external'). For Lewis, a world is a very substantial thing indeed: our cosmos is one of them, and the others differ from it 'not in kind but only in what goes on at them'. A 'possible object' like Quine is 'in' a world if it is a part of it, and no possible object is a part of more than one world. Lewis' answer to the sceptics is therefore that Quine is simply not identical with anything in any other world.

The totality of worlds, Lewis maintains, is a 'logical plenum'; there is a world for every possible arrangement of things. (He is able to state this requirement without the use of any modal terms.) Although we rightly single out one member of the logical plenum as 'the actual world', there is in an important sense nothing special about it, for 'actual' is simply an indexical term like 'now' and 'here'.

Most philosophers have found the ontology that underlies Lewis' account of modality incredible. Many doubt whether there are any 'worlds' besides our own cosmos. Many wonder how Lewis knows that, even if there are other worlds, they are numerous and various enough to form a logical plenum. Many wonder why Lewis thinks that, given that there are all these 'worlds', they have anything to do with modality, or why each of them would be, as Lewis says each of them is, a 'way things could have been'. Another sort of worry, one that can perhaps be more profitably discussed, is closely related to Lewis' answer to the sceptics about cross-world identity: how can Lewis accommodate our convictions about 'modality *de re*'? No doubt Quine could have been a geographer. In the language of possible worlds, there is a world in which Quine is a geographer. But, the worrier points out, if Lewis is right, Quine is not in any world but our own (see De re/de dicto).

Lewis replies that, although Quine is not in any other world, there are people in some other worlds who play the same role in their worlds that Quine plays in his (our) world - who are the people in those worlds who are the most similar to Quine. If there are people in the world *w* who enjoy this status, they are the counterparts of Quine in *w*. (But since the question, 'Which things are most similar to *x*?' is a pragmatic question, one that can be answered in various ways depending on one's interests, there is no unique 'counterpart relation' - no one relation that is expressed by '*x* is a counterpart of *y*'.) Lewis analyses modal statements *de re* by reference to the counterparts of the *res* they concern: to say that Quine could have been a geographer is to say that he has a counterpart who is a geographer. This analysis has been disputed on the ground that Quine's counterparts are not, after all, Quine; therefore, it is urged, no statement about their features could be equivalent to any statement about what he could have been. Lewis replies that the property 'could have been a geographer' is the property 'has a counterpart who is a geographer'. If, therefore, Quine has a counterpart who is a geographer, Quine has the property 'could have been a geographer'. But this reply requires a qualification. Since there is no one counterpart relation, there is no one property that is expressed by 'could have been a geographer'. Whether we are willing to say that Quine could have been a geographer is, in consequence, a function of the interests that we bring to our discussions of how Quine could have been different.

In 'Counterpart Theory and Quantified Modal Logic' (1968), Lewis presents his thoughts about counterparts and modality in a very general form. He presents an extensional first-order language whose primitive nonlogical vocabulary comprises only predicates - for example, '*x* is a possible world', '*x* is a counterpart of *y*' and so on - and he embodies his theory of counterparts ('Counterpart Theory') in eight axioms stated in this language. He presents an algorithm for translating sentences of quantified modal logic into this language. He goes on to make some illuminating observations about quantified modal logic by showing that certain 'famous' problematical sentences in the language of quantified modal logic correspond under the translation algorithm to theorems, and certain others to non-theorems, of Counterpart Theory; he shows that if various formal restrictions were placed on the counterpart relation - restrictions that would in his view be arbitrary - the theorems that would thereby be added to Counterpart Theory would be translations of certain of the 'famous' sentences of quantified modal logic.

He shows, finally, that, although all sentences in the language of quantified modal logic can be translated into his extensional language, the converse does not hold. The latter is therefore a richer language: it has all the expressive power of the language of quantified modal logic and more besides.

An important feature of Lewis' account of modality is that, unlike most other accounts of modality, it is reductive. If Lewis is right, modal terms can be eliminated from our discourse in favour of spatiotemporal terms, 'part', and an indexical term like 'we'. (It is, moreover, arguable that if one accepts Lewis' modal ontology, one need include in one's ontology no non-individuals, but sets ultimately grounded on individuals. Propositions, for example, can be identified with sets of worlds, and properties with sets of the things in them.) Whether or not those who find Lewis' modal ontology incredible are right so to regard it, it is undeniable that a lot more can be done with it than with any other modal ontology. We have seen that it provides a reductive analysis of modality. We have not been able to discuss its astounding successes - granting only its truth - in dealing with a very wide range of philosophical problems. (The philosopherJohn Perry has remarked that it 'goes through philosophical problems the way a McCormick reaper goes through wheat'.) Perhaps the same feature of the ontology - the vast range of spatiotemporal objects it assumes - accounts both for its extraordinary power and the 'incredulous stares' that have been so often directed at its inventor. Lewis does not dispute the validity of the maxim 'Entities are not to be multiplied beyond necessity', but he insists that, once the content and consequences of a theory have been made as clear as they can be, whether the entities assumed by that theory lie inside or outside the bounds of 'necessity' is a matter of individual, subjective judgment.

See also: Modal logic; Modal logic, philosophical issues in

PETER VAN INWAGEN

List of works

Lewis, D.K. (1966) 'An Argument for the Identity Theory', *Journal of Philosophy* 63: 17-25.(An influential argument for the mind-body identity thesis.)

Lewis, D.K. (1968) 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* 65: 113-26.(Lewis' first statement of his distinctive treatment of modality.)

Lewis, D.K. (1969) *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.(A classic account of conventions as regularities which are recognized as solutions to coordination problems.)

Lewis, D.K. (1970a) 'Anselm and Actuality',*Nous* 4: 175-88.(Includes Lewis' indexical conception of actuality.)

Lewis, D.K. (1970b) 'How to Define Theoretical Terms', *Journal of Philosophy* 67: 427-46.(An elegant treatment of the role of theoretical terms in scientific theories, drawing on the work of Ramsey and Carnap.)

***** (1973a) *Counterfactuals*, Cambridge, MA: Harvard University Press.(A full statement of Lewis' 'possible-world' treatment of counterfactual conditions.)

(1973b) 'Causation', *Journal of Philosophy* 70: 535-67.(Causation as counterfactual dependence.)

(1976) 'Survival and Identity', in *The Identities of Persons*, ed. A.O. Rorty, Berkeley, CA: University of California Press.(Argues that puzzles concerning personal identity are best solved by regarding persons as suitably related aggregates of person-stages.)

(1979a) 'Counterfactual Dependence and Time's Arrow', *Nous* 13: 455-76.(Uses his treatment of counterfactuals to explain asymmetries in time.)

(1979b) 'Attitudes De Dicto and De Se', *Philosophical Review* 88: 513-43.(Uses the theory of possible worlds to provide a sophisticated treatment of propositional attitudes.)

(1980) 'A Subjectivist's Guide to Objective Chance', in *Studies in Inductive Logic and Probability*, vol. II, ed. R.C. Jeffrey, Berkeley, CA: University of California Press.(An attempt to provide an account of chance in terms of degrees of belief and frequencies; subsequently Lewis has acknowledged that the account is flawed - see relevent Postscript in 1986b and Introduction xiv-xvii.)

(1983a) *Philosophical Papers*, New York: Oxford University Press, vol. 1. (Contains Lewis' most important pre-1981 papers on ontology, the philosophy of mind, and the philosophy of language, including 1966, 1968, 1970a, 1970b, 1976 and 1979b. Also includes Postscripts to various of the papers, and a brief but illuminating Introduction in which Lewis discusses his way of doing philosophy.)

(1983b) 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* 61: 343-77.(Discusses theories of universals.)

(1986b) *Philosophical Papers*, New York: Oxford University Press, vol. 2.(Contains two previously unpublished papers and Lewis' most important pre-1982 papers on the relation of counterfactual conditionals to time, probability, causation and causal explanation, causal dependence, and decision theory, including 1973b, 1979a,

1980, and 1986a. Also includes Postscripts to various of the papers.)
(1986c) *On the Plurality of Worlds*, Oxford: Blackwell.(A sustained exposition and defence of Lewis' ontology of possible worlds.)
(1991) *Parts of Classes*, Oxford: Blackwell.(A brilliant and highly original discussion of set theory, conducted from the point of view provided by Lewis' ontology.)
(1994) 'Reduction of Mind', in *A Companion to Philosophy of Mind*, ed. S. Guttenplan, Oxford: Blackwell.(A reformulation of his early argument for the mind-body identity thesis, drawing on 1970b, plus an extended critical discussion of theories of mental content.)

References and further reading

Armstrong D. (1989) *A Combinatorial Theory of Possibility*, Cambridge: Cambridge University Press.(In Chapters 1 and 2 Armstrong sets out his case for a 'naturalist' alternative to Lewis' theory of possibility.)
Plantinga, A. (1987) 'Two Concepts of Modality: Modal Realism and Modal Reductionism', in *Philosophical Perspectives*, ed. J. Tomberlin, Atascadero, CA: Ridgeview, vol. 1, 189-231.(Argues that because Lewis proposes to reduce possibilities to worlds, he is not the modal realist he is generally taken to be.)
Stalnaker, R. (1968) 'A Theory of Conditionals', in *Studies in Logical Theory*, ed. N. Rescher, Oxford: Blackwell.(The paper which started current debates on counterfactual conditionals.)
Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press.(In Chapter 3, Stalnaker gives an alternative to Lewis' account of possible worlds, and in Chapter 7 he responds to Lewis' criticisms of his theory of conditionals.)
van Inwagen, P. (1986) 'Two Concepts of Possible Worlds', *Midwest Studies in Philosophy* 11: 185-213. (Extended critical exposition and discussion of Lewis' position.)

Li

Li *means 'pattern' or 'principle', and as a verb can also refer to the creation of orderly pattern. Mencius believed that the human heart-mind had an inherent taste for and attraction to such 'good order'. His contemporary, Zhuangzi, was the first to use* li *to refer to an underlying normative 'pattern' structuring and giving order to the entire world. This sense in turn influenced the Confucian thinker Xunzi, who employed and developed the concept to express his understanding of Confucianism. Certain Buddhist thinkers used the notion of* li *to describe first 'emptiness' and later 'Buddha-nature', the common underlying characteristic of all phenomena. A version of this idea was adopted by neo-Confucian thinkers, who believed that while each thing manifested its own particular* li*, it also contained within it the* li *of all other things. Thus there is a profound metaphysical identity between self and world.*

The earliest and most basic sense of the character *li* ('pattern' or 'principle') is derived from the simpler graph which serves as its central component: a rectangular field divided into equal quadrants to form an 'orderly pattern'. As a verb, *li* has the related meaning of creating such a pattern, thus 'to set in good order'.

The Mohists were the first Chinese thinkers to use *li* with a distinctly philosophical sense (see Mohist philosophy). In the *Mozi*, *li* is used to mean 'good order' (for example, in Chapter 25 it is opposed to 'chaos'). In later Mohist writings, *li* means the principles governing such order and orderly principles of thought (Graham 1978: 192). These two senses of 'good order' are both present in *Mencius* 6A7, which claims that all human minds agree in approving of *li* (good order) and *yi* (what is right).

The most important and extensive early uses of *li* occur in the *Zhuangzi*. There, *li* appears both in the sense of the grand pattern underlying all phenomena and the individual instantiations of this pattern in discrete things. The *Zhuangzi* contains the first occurrences of 'heavenly principles', 'principles of the Way' and 'great principles', terms which connect the notion of 'pattern' to a greater cosmic scheme, lending it a wider metaphysical role and greater normative force. Being explicitly linked to 'Heaven' and the 'Way', *li* describes not only how things are but also how they should be. We also see several references to 'the principles of the myriad things' and in one case to 'the various principles of the myriad things', ideas that prove important in neo-Confucian thought (see Zhuangzi).

Xunzi was deeply influenced by these new senses of the term and incorporated them into his Confucian philosophy. Formally, his views are remarkably similar to those in the *Zhuangzi*, but their content - what he claimed to be the underlying pattern of the cosmos - was significantly different. Xunzi's ideal was a consciously anthropocentric harmony between human needs, desires and capacities, and Nature. Zhuangzi's vision lacked this human centre; he focused on 'heavenly principles', a term which does not appear in Xunzi's writing.

The view that *li* provides both a descriptive and normative pattern for individual things, while also being part of a larger cosmic pattern, gave rise to the idea that this larger pattern 'united' the various phenomena of the world. This idea can be found in Han dynasty texts such as the Huainanzi. It was most developed by thinkers such as Wang Bi, who speaks of 'ultimate principle' as an object of mystical contemplation and religious veneration. His contemporary Guo Xiang, who held similar views, also believed that the 'pattern' of the world was not the result of some predetermined plan but rather of 'spontaneous' processes. He insisted that the ultimate nature of the world was *wu* (non-existence) (see You-wu).

Chinese Buddhists adapted these later senses of *li* to express their own views. In the *Sutra in Forty-two Sections*, *li* is used to mean 'the true nature of reality', a full appreciation of which will result in complete enlightenment. Zhi Dun developed the equation of *li* with *wu* 'non-existence', understanding both in the Buddhist sense of 'emptiness'. Huayan Buddhists understood the 'emptiness' of reality - the lack of an independent, individual nature - to mean that all things are manifestations of a shared 'Buddha-nature'. Huayan works, such as the *Treatise on the Golden Lion*, argue that every 'phenomenon' contains within it all the various *li* (see Buddhist philosophy, Chinese; Buddhist concept of emptiness).

Like Huayan Buddhists, neo-Confucians believed that *li* describes the fundamental character of reality and that the spiritual life lies in grasping this shared, unifying principle. Cheng Yi was famous for the idea - derived from the

earlier Buddhist thinker Daosheng (d. AD 434) - that ‘principle is one; its manifestations many’. However, neo-Confucians believed that principle defined a classical Confucian world. Not unlike the earlier difference between Zhuangzi and Xunzi, their disagreement with the Buddhists turned largely on the content - not the formal structure - of their metaphysical views.

Zhu Xi believed *li* was logically prior to, though not actually separable from existing things. It lodged in *qi* (ether) to form actual things, riding it ‘as a person rides a horse’ (see Qi). While each thing contained all the myriad *li*, a given thing’s endowment of *qi* allowed only certain *li* to be manifested. Thus while in some deep sense things were the same, in the actual world each had its own particular (manifested) principle.

Neo-Confucians such as Lu Xiangshan and Wang Yangming took issue with this view. They feared that Zhu’s teachings, which described *li* as ‘above actual things and events’, would lead students away from their moral heart-minds and into quietistic contemplation. They argued that the mind did not contain but was itself principle. Later neo-Confucians such as Wang Fuzhi, Yan Yuan and Dai Zhen criticized these earlier thinkers for being infected with Daoist and Buddhist ideas. They argued for a return to the original sense of *li*: the ‘good and proper order’ inherent in the world (see Neo-Confucian philosophy §2).

See also: Cosmology; Dao; Law and ritual in Chinese philosophy; Neo-Confucian philosophy; Mohist philosophy; Qi; Xunzi; Zhuangzi

PHILIP J. IVANHOE

References and further reading

Chan Wing-tsit (1964) ‘The Evolution of the Neo-Confucian concept of *li* as principle’, *Tsing-Hua Journal of Chinese Studies* n.s. 4.2: 123-49.(Thorough collection of examples of *li* from throughout Chinese history.)

Graham, A.C. (1978) *Later Mohist Logic, Ethics and Science*, Hong Kong: Chinese University Press, and London: School of Oriental and Asian Studies.(Remarkable reconstruction and analysis of the later Mohist sections of the *Mozi*, which concerns the systematic study of argumentation, ethics and science.)

(1992) *Two Chinese Philosophers*, LaSalle, IL: Open Court, 2nd edn. (Insightful study of the two neo-Confucian thinkers, Cheng Hao and Cheng Yi; helpful discussions of *li* especially on pages 8-22.)

Liber de causis

The Liber de causis (Book of Causes) is a short treatise on Neoplatonist metaphysics, composed in Arabic by an unknown author probably in the ninth century in Baghdad. Through its twelfth-century Latin translation, it greatly influenced mature medieval philosophy in the West.

Drawing heavily on the Greek Neoplatonist Proclus, the Liber de causis represents a development of late Neoplatonism along two lines. On the one hand, the author modifies and simplifies Proclus' theory of causes to accord more closely with the three-part division of ultimate causes advanced by the founder of Neoplatonism, Plotinus. On the other hand, the author introduces some of the metaphysical principles of Qu'ranic or biblical monotheism. The result is a metaphysically provocative reinterpretation of Neoplatonist thought which, because it seemed to accommodate Platonist philosophy to the medieval worldview, made the Liber de causis a natural source text for medieval philosophers.

1 Form, content and origin

The *Liber de causis (Book of Causes)* is presented *more geometrico*, that is, it proceeds by way of thirty-one discrete propositions, or theses (often amplified by corollaries), each accompanied by passages of interconnected supporting argumentation. Despite the systematic presentation, however, the work as a whole appears somewhat fragmentary, no doubt because the author was overly ambitious. It appears that he was trying to set forth, in outline, an entire theory of how the universe has been originated by a first cause (an entity consisting in pure goodness and perfect being) and how this cause's creative activity comports with the agency of other fundamental causes such as the Neoplatonist supramundane 'mind' and universal 'soul'. In the known Arabic manuscripts, the treatise is entitled either 'The Book of Aristotle's Exposition on the Pure Good' or 'Discourse on the Pure Good' (*Kalam fi mahd al-khair*); its designation as a *kalam* (discourse, discussion or spoken account) might be taken to indicate that the work is not actually intended to be a systematic treatise but, rather, a collection of teachings or meditations. Clearly, however, the work's shortcomings as a systematic treatise do not preclude a systematic philosophical intention on the part of its author or even a coherent underlying structure. Aquinas, in his *Super Librum de causis expositio (Commentary on the Liber de causis)* proposed an underlying structure in which every proposition has an identifiable expository function.

The *Liber de causis* was long thought, by Islamic as well as Western medieval philosophers, to be a work of Aristotle (see Neoplatonism in Islamic philosophy §1). In these philosophers' eyes, it filled out Aristotle's otherwise deficient account of the ultimate causes of the universe as presented in *Metaphysics* XII. Even before Aquinas demonstrated that the work was un-Aristotelian and indebted to Proclus' *Elements of Theology*, speculation had arisen about its authorship. Albert the Great's idea that the author might have been a twelfth-century Jewish writer in Spain, Abraham Ibn Daud, has until recently been considered a plausible hypothesis. Discoveries in the 1970s and 1980s, however, have given the controversy new direction. Very likely, the author of the *Liber de causis* was a Muslim or Christian thinker writing in the Near East, probably at the intellectual centre in Baghdad, in the ninth century (in any case, no later than the last quarter of the tenth century).

2 Philosophical context

Many historians of philosophy have seen the *Liber de causis* as merely an anonymous epitome of Proclus' *Elements of Theology* (see Proclus). Such views, however, miss the work's recourse to certain ideas of Plotinus, from which Proclus had departed. The author of the *Liber de causis* rejects Proclus' version of the doctrine of 'hypostases', the levels of reality posited by Neoplatonism as beyond the physical world, or causally prior to it (see Neoplatonism). Proclus argues that the fundamental causes of the universe consist in a complex system of hypostases arranged in a causal series, beginning with the first cause, the absolute One, and continuing to its ultimate effects in the material world through the intermediary hypostases of Finitude/infinity, Being, Eternity, Mind, Forms, Time, Soul and so on. By contrast, the *Liber de causis* re-affirms the simpler Plotinian causal series of three hypostases, the perfectly unified and simple first cause (the One), Mind and Soul.

What is at stake philosophically in the *Liber de causis*' simplification of the Neoplatonist series of hypostases and

the denial that there are intermediary hypostases between the first cause and the hypostasis of finite being as Mind? First, the *Liber de causis* breaks with Proclus' (and perhaps ultimately Plato's) principle that any common feature of things requires a special cause, or hypostasis, in which all the things possessing the feature participate. Second, the *Liber de causis* responds to one of the most basic problems of Neoplatonist metaphysics: if causal dependence involves a certain sort of similarity of effect to cause (participation of effect in cause), how can causes give rise to effects quite different from themselves? Locating a mediating factor between cause and effect in intermediate levels of reality, or hypostases, is a tendency that can be seen already in Plotinus. The *Liber de causis* represents a development of Neoplatonist metaphysics since it attempts to redress the extreme multiplication of hypostases found in later Neoplatonism.

The *Liber de causis* also exemplifies the creative interaction of Islamic and Christian thought with the legacy of Greek philosophy (especially Neoplatonism). This is illustrated by its presentation of the first cause as truly existent and as pure being (in contrast to the Neoplatonist characterization of it as in some sense beyond being), its philosophical refinement of the idea of creative causality, and its derivation from creative causality of a doctrine of universal providence. The treatise is apparently a product of the early, seminal stages of interaction between monotheistic speculation and Greek philosophy in the Islamic milieu, which took place in the latter half of the eighth century and through the ninth century (see Greek Philosophy: impact on Islamic philosophy). Since this formative period was a time in which Christian scholars and translators strongly influenced Islamic thought, the *Liber de causis* can also be said to have arisen out of early contact between Christianity and Islam.

3 A source for later medieval philosophy

Although further research will no doubt identify more echoes of the *Liber de causis* in the writings of later Islamic and Jewish medieval philosophers, its direct and definite influence on their work appears to have been relatively slight (exceptions are Ibn Sab'in, Al-Baghdadi and Ibn Ezra). Its substantial influence is to be seen more in its contribution to a general store of philosophical ideas than in its shaping of particular doctrines. In the Latin West, by contrast, the *Liber de causis* served as one of the most frequently used sources of Neoplatonist metaphysics and aided Christian medieval philosophers, even those strongly influenced by Aristotle, in building Platonist principles and ideas into the very foundations of their philosophical systems (see Platonism, medieval). Translated into Latin, probably around 1180 in Toledo by Gerard of Cremona, and known as *Liber Aristotelis de expositione bonitatis purae* or more simply as *Liber de causis*, the work became available to Western philosophers along with the bulk of Aristotle's works (see Translators). It made up part of the massive influx of Greek and Arabic learning into western Europe in the latter half of the twelfth century, quickly becoming a standard text; 237 extant Latin manuscripts have been identified. While its widespread acceptance was encouraged by the attribution to Aristotle, the *Liber de causis* was studied for its intrinsic interest. Its influence continued after Aquinas established its connection with the thought of Proclus.

The influence of the *Liber de causis* can be gauged in part by the fact that a high proportion of the major philosophers of the thirteenth and fourteenth centuries, as well as many minor figures, wrote commentaries on it. Twenty-seven commentaries from before 1500 have been identified. Written in a wide variety of genres, these commentaries include works by Roger Bacon, Albert the Great, Thomas Aquinas, Siger of Brabant, Henry of Ghent (attribution doubtful), Giles of Rome and others. Even more importantly, the direct influence of the *Liber de causis* shows in the frequent citations of it and the manifold adaptations of its doctrines found in major philosophical writings from the 1230s through the Renaissance.

For medieval philosophers, the *Liber de causis* represented not only a source of Neoplatonist ideas but also an adaptation of Neoplatonist metaphysics of the sort that they themselves were intent upon. Thus Aquinas, in his detailed and masterful commentary, is concerned to show how the *Liber de causis* differs from and is superior to its source in Proclus. Using a translation of the *Elements of Theology* made in 1268 by William of Moerbeke, Aquinas draws attention both to *Liber de causis*' affinities with Proclus' systematization of Neoplatonist metaphysics, and its divergences from it. In his analysis, he employs the metaphysical writings of the fifth-century Christian Neoplatonist Pseudo-Dionysius as a key to understanding the *Liber de causis* and also as a corrective when he thinks that the author has not gone far enough in re-working classical Neoplatonism. Affecting medieval philosophers' accounts of creation, divine nature and ontology, the *Liber de causis* provides them with particular ideas about creative and natural causation, the individuation of pure being and the constitution of finite being.

See also: Albert the Great; Gerard of Cremona; Greek philosophy: impact on Islamic philosophy; Islamic philosophy: transmission into Western Europe; Neoplatonism; Platonism, medieval; Plotinus; Proclus; Ulrich of Strasbourg

HANNES JARKA-SELLERS

References and further reading

1 Editions and translations

Badawi, A. (ed.) (1955) 'Procli: Liber (Pseudo-Aristotelis) de expositione bonitatis purae (Liber de causis)', *Neoplatonici apud Arabes*, *Islamica* 19: 1-33.(A critical edition of the Arabic text, to be superseded by the edition being prepared by R.C. Taylor.)

Bardenhewer, O. (1882) *Die Pseudo-aristotelische Schrift über das reine Gute bekannt unter dem Namen 'Liber de causis'*, Freiburg: Herder'sche Verlagsbuchhandlung; repr. Frankfurt: Minerva, 1961.(The pioneering work of modern scholarship, containing an Arabic text, a Latin text and a German translation.)

Brand, D.J. (1984) *The Liber de Causis: Translated From the Latin with an Introduction*, Marquette, WI: Marquette University Press.(The best approach to the *Liber de causis* for the English reader, with a helpful introduction and select bibliography. Aquinas' analysis of the contents of the *Liber de causis* is provided in the introduction.)

Pattin, A. (1966) 'Le Liber de causis. Édition établie à l'aide de 90 manuscrits avec introduction et notes', *Tijdschrift voor Filosophie* 28: 90-203.(An edition of the Latin translation, to be shortly superseded by the nearly critical edition being prepared by R.C. Taylor.)

Zerachja B. Izsak (1916) *Pseudo-Aristotelis Liber de causis*, ed. I. Schreiber, Budapest. (An edition of a medieval Hebrew translation of the *Liber de causis*.)

2 Medieval commentaries

Albert the Great (1265-75) *Opera omnia [...] Tomus XVII pars II, De causis et processu universitatis a prima causa*, ed. W. Fauser, Münster: Aschendorff, 1993.(The critical edition of Albert the Great's commentary, which is a massive and philosophically rich work.)

Aquinas, Thomas (1272) *Super Librum de causis expositio*, ed. H.D. Saffrey, Louvain: B. Nauwelaerts, Fribourg: Société philosophique, 1954; *In Librum de causis expositio*, ed. C. Pera, Torino: Marietti, 1955; trans. V. Guagliardo, C. Hess and R.C. Taylor, *Commentary on the Liber de Causis*, *Thomas Aquinas in Translation* vol. 1, Washington, DC: Catholic University of America Press, 1996.(Saffrey's critical edition of Aquinas' commentary is complemented by Pera's study-edition, containing a wealth of helpful material. Guagliardo, Hess and Taylor provide an accessible translation with introduction and notes.)

Bacon, Roger (1241-5) *Opera hactenus inedita Rogeri Baconi, XII, Quaestiones supra Librum de causis*, ed. R. Steele and F.M. Delorme, Oxford: Clarendon Press, 1935.(An excellent edition.)

Giles of Rome (1280) *Opus super authorem de causis Alpharabium*, Venice: Iacob Zoppinus, 1550; repr. Frankfurt: Minerva, 1968.(A Renaissance edition, to be superseded shortly by the critical edition being prepared by F. del Punta *et al*.)

Henry of Ghent (1245-5) *Les Quaestiones in Librum de causis attribuées à Henri de Gand*, ed. J. P. Zwaenepoel, Louvain: Publications Universitaires, 1974.(A critical edition with introduction.)

Siger of Brabant (1255-85) *Les Quaestiones super Librum de causis de Siger de Brabant*, ed. A. Marlasca, Louvain/Paris: Publications Universitaires - B. Nauwelaerts, 1972.(A critical edition with introduction.)

3 Further reading

Beierwaltes, W. (1963) 'Der Kommentar zum *Liber de causis* als neuplatonisches Element in der Philosophie des Thomas von Aquin' (The Commentary on the *Liber de causis* as a Neoplatonic Element in the Philosophy of Thomas Aquinas), *Philosophische Rundschau* 11: 192-215.(A philosophically perceptive overview of the influence of the *Liber de causis* on Aquinas and other philosophers of his time.)

D'Ancona Costa, C. (1995) *Recherches sur le Liber de causis (Research on the Liber de causis),* Paris: Vrin.(The most comprehensive and thorough general study, with very extensive bibliography.)

Endress, G. (1973) *Proclus Arabus. Zwanzig Abschnitte aus der 'Institutio theologica' in arabischer Übersetzung (Proclus Arabus: Twenty Sections of the Instituto theologica in Arabic Translation)*, Wiesbaden: F. Steiner Verlag.(A foundational study for our understanding of the sources and origin of the *Liber de causis*.)

Rowson, E. (1984) ‘An Unpublished Work by Al-Amiri and the Date of the *Liber de causis*’, *Journal of the American Oriental Society* 104: 193-9.(An important contribution to our understanding of the provenance of the *Liber de causis*.)

Sweeney, L. (1959) ‘Research Difficulties in the *Liber de causis*’, *The Modern Schoolman* 36: 109-16; repr. in L. Sweeney (ed.) *Divine Infinity in Greek and Medieval Thought*, New York: Peter Lang, 1992.(An authoritative and influential treatment of the philosophical character of the *Liber de causis*.)

Taylor, R.C. (1983) ‘The *Liber de causis*: A Preliminary List of Extant Manuscripts’, *Bulletin de philosophie médiévale* 25: 63-84.(A treasury of information for further research into the role of the *Liber de Causis* in medieval philosophy.)

Taylor, R.C. (1986) ‘The *Kalam fi mahd al-khair* (*Liber de causis*) in the Islamic Philosophical Milieu’, in J. Kraye, W.F. Ryan and C.B. Schmitt (eds) *Pseudo-Aristotle in the Middle Ages: The ‘Theology’ and other Texts*, London: The Warburg Institute, 1986.(A summary discussion of the place of the *Liber de causis* in Islamic and later Jewish philosophy.)

Liberalism

Liberal political philosophy explores the foundations of the principles most commonly associated with liberal politics: freedom, toleration, individual rights, constitutional democracy and the rule of law. Liberals hold that political organizations are justified by the contribution they make to the interests of individuals, interests which can be understood apart from the idea of society and politics. They reject both the view that cultures, communities and states are ends in themselves, and the view that social and political organizations should aim to transform or perfect human nature. People have purposes of their own to pursue, either economic or spiritual (or both). Since those purposes do not naturally harmonize with one another, a framework of rules may be necessary so that individuals know what they can count on for their own purposes and what they must concede to the purposes of others. The challenge for political philosophy, then, is to design a social framework that provides this security and predictability, but represents at the same time a safe and reasonable compromise among the disparate demands of individuals.

1 Liberal politics

In politics, the term ’liberalism’ denotes a family of positions centred around constitutional democracy, the rule of law, political and intellectual freedom, toleration in religion, morals and lifestyle, opposition to racial and sexual discrimination, and respect for the rights of the individual.

Often these positions are associated with a suspicion of state authority, with a view that the powers of government should be constrained if not minimized, and with a confidence in the ability of individuals to organize themselves on the basis of the market, the free interplay of ideas and the loose and informal associations of civil society. Liberal support for democracy is therefore sometimes qualified by fear of ’the tyranny of the majority’ and by apprehensions about the extent and intrusiveness of the power that a populist state is capable of exercising.

These attitudes are not, however, characteristic of all forms of liberalism. In Britain, in the late nineteenth and early twentieth century, a group of thinkers known as the New Liberals made a case against *laissez-faire* and in favour of state intervention in social, economic and cultural life. The New Liberals, who included T.H. Green and L.T. Hobhouse, saw individual liberty as something to be achieved under favourable social circumstances. The poverty, squalor and ignorance in which most people lived made it impossible in their view for freedom and individuality to flourish, and the New Liberals believed that these conditions could only be ameliorated through collective action coordinated by a strong welfare-oriented interventionist state.

In the USA, since the early part of the twentieth century, the term ’liberalism’ has been associated with ‘progressive’ economic reform, a commitment to the modest redistribution of income that takes place in a welfare state, a suspicion of business and an abiding faith in the legal regulation of economic affairs. The more *laissez-faire* version of liberalism is called ’conservatism’ in the USA, and Europeans are often disconcerted to hear ’liberal’ used there as label for positions that they themselves would describe as left-wing or moderately socialist.

This is not just terminological confusion. Those in the USA who call themselves ’liberals’ do also hold the positions outlined at the beginning of this article, and their disagreement with ’conservative’ opponents is partly a live and unresolved issue about the implications of traditional liberal premises in so far as social and economic policy is concerned. Does individual freedom require private ownership? Is poverty compatible with liberty? Can civil and political rights be equal if economic power is not? Liberalism is a family of positions, and these remain important family disputes.

2 Political philosophy

In philosophy, ’liberalism’ is not just the name of a loosely organized and quarrelsome family of substantive political opinions. It refers also to a heritage of abstract thought about human nature, agency, freedom, and value, and their bearing on the functions and origins of political and legal institutions.

That heritage takes its rise in early modern English political philosophy - most notably in the work of Thomas Hobbes and John Locke §10. It is also the political philosophy of the European Enlightenment, represented in its most philosophically articulate form in the writings of Jean-Jacques Rousseau, François-Marie Arouet de Voltaire,

Henri-Benjamin Constant de Rebecque and, a little later, Immanuel Kant. In the nineteenth century, philosophical liberalism is represented, first, in the utilitarian theories of Bentham and J.S. Mill, and later in the ’Idealism’ of T.H. Green.

Inevitably, because of our proximity, it is harder to identify canonical works of twentieth-century liberalism. There was a long period in the twentieth century in which liberal philosophers seemed to lose their taste (or their nerve) for grand theory on the scale of Hobbes or Kant, a period during which they seemed to pride themselves on the piecemeal, analytic and unsystematic character of their thought. In a Cold War context, these were regarded as healthy signs of being ‘non-ideological’. That phase seems to have passed, and more confident versions of philosophical liberalism have re-emerged in the work of late-twentieth-century writers like F.A. von Hayek §3, Robert Nozick §2, Ronald Dworkin, Joseph Raz and, most importantly, John Rawls §4.

Some will quibble about one or two of the names on this list. Was Hobbes really a liberal? Was Rousseau? We should remember, however, that ‘liberalism’ has never been a label over which any group has exercised collective control. As a result, the term is at the mercy of its most casual users, and indeed the attempt to define ’liberalism’ is undertaken most commonly not by its practitioners but by its opponents, with predictable caricatural results.

Even so, the challenge is not just to correct misrepresentations. The philosophical positions that we most plausibly identify as liberal often represent distinctive expressions of ambivalence about human nature and political life, rather than dogmatic formulae in a liberal catechism. We have seen this already in the values and principles which constitute liberalism in the political sense: liberals disagree about property, economic equality and the role of the state. At the more philosophical level, liberals disagree about the nature of value, the meaning of freedom and the connection between individual and social purposes.

What follows is an attempt to lay out some of those positions and controversies. But defining liberalism is, on the whole, a frustrating pastime. There are many ways of mapping this philosophical landscape, and there is no substitute for grappling with the disparate detail of the theories propounded by particular liberal philosophers.

3 Individualism

Let us begin with some basic ethical premises. The deepest commitment of liberal political philosophy is to *individualism*, as a fundamental proposition about value. Liberal individualism has four parts to it.

First, liberals believe that the individual person is what matters for the purposes of social and political evaluation. We may be interested in the fate of a culture, a language, a community or a nation, but for a liberal such interest is always secondary or derivative. Ultimate value has to do with how things are for ordinary men and women, considered one by one: their pains and pleasures, their preferences and aspirations, their survival, development and flourishing. Of course, people do care about each other: individualism is not the same as egoism. But individualism excludes social and collective entities from the realm of ultimate goods.

There is less agreement about the grounds for this individualization of value. John Locke, writing in the seventeenth century, based it on each person being the workmanship and property of God, which meant that we were ‘made to last during His, not one anothers Pleasure’ (1690). This relation to God was direct, unmediated and unconditional in the case of each individual. It therefore established a basis for our rights with respect to one another that did not presuppose validation by larger social structures.

Modern liberalism, however, is a secular tradition, and its history since Locke’s time is largely a history of the attempt to establish this individualism without appealing to the idea of God. Utilitarian thinkers linked the notion of value analytically to desire or preference, and they inferred, from the fact that desiring and preferring were attributes of individuals, that the fundamentals of value must be individualistic too (see Utilitarianism). Those following in the tradition of Kant, on the other hand, linked value analytically to the lonely individualism of will, conscience and the sense of duty, and drew the conclusion that each person, *qua* moral agent, was entitled to be regarded as an end in themselves, not just a means to broader social ends. The Kantian view has perhaps fared better in modern political philosophy, although its underlying argument - that because moral thinking takes place at the level of individual minds and wills, individual minds and wills must also be the fundamental objects of moral concern - has yet to be rendered in a compelling form.

Second, liberals believe that there is something particularly important in the capacity of individuals to direct their actions and live their lives, each on their own terms. They believe in the importance of freedom - although what that belief amounts to is one of the controversies referred to earlier. Some define freedom in negative terms: their libertarianism amounts simply to a condemnation of force, coercion and interference in human life. Freedom, they say, is what flourishes when these constraints are taken away, and there is nothing apart from the removal of constraints that needs to be done politically in order for freedom to flourish. Positive conceptions of liberty allow the state a much greater role than this: they may see freedom or autonomy as something to be achieved, rather than taken for granted, in the life of an unrestrained individual, something that requires educated individual capacities and favourable social conditions (see Freedom and liberty).

Some conceptions of positive liberty go well beyond this, moving out of the liberal realm altogether. If freedom is identified with the performance of social duty, or attributed to individuals only by virtue of their participation in some social whole, then the resulting theory can hardly be described as liberal.

Also, if freedom is presented as the achievement of a happy few, something of which the ordinary mass of humanity is incapable, then again we are not dealing with a liberal conception. Although liberal freedom is sometimes a developmental concept, it is not an aristocratic or utopian one. The free direction of a human life is seen as something which ordinary people are capable of, under decent social and political circumstances. When Colonel Rainsborough exclaimed in the Putney Debates of 1647, ‘Really, I think that the poorest he that is in England has a life to live as the greatest he’, he gave voice to an egalitarianism that lies at the foundations of the liberal tradition (Wootton 1986: 286).

The third aspect of liberal individualism, then, is a commitment to equality. We have to be careful how we formulate this. Liberal philosophers are not necessarily egalitarians in the economic sense. But they are committed as a matter of the basic logic of their position to a principle of underlying equality of basic worth. People are entitled to equal concern for their interests in the design and operation of their society’s institutions; and they have the right to be equally respected in their desire to lead their lives on their own terms (see Equality).

Feminists have sometimes questioned whether this liberal commitment to equality extends across boundaries of gender. In the writings of Locke, Rousseau and Kant it is easy to find throwaway lines that would be described today as sexist or misogynistic. No writer in the liberal canon committed himself explicitly and at length to the emancipation of women much before J.S. Mill’s essay *The Subjection of Women* in 1868. Nevertheless, the legacy of liberal carelessness on this issue does not pose any major theoretical difficulties for the position that men and women are equal in their moral and political capacities and in the respect to be accorded those capacities. Indeed, the more challenging feminist critique is that liberals exaggerate (rather than deny) the similarities between men and women - that they fail to either acknowledge or accommodate crucial elements of ‘difference’ in moral reasoning and ethical demeanour (see Feminist political philosophy).

A fourth element of liberal individualism involves an insistence on the rights of individual reason. This involves not just freedom of thought, conscience or discussion, but a deeper demand about justification in politics: the demand that rules and institutions of social life must be justified at the tribunal of each individual’s reason.

We see here an important connection between liberal thought and the philosophical legacy of the Enlightenment. The Enlightenment was characterized by a burgeoning confidence in the human ability to make sense of the world, to grasp its regularities and fundamental principles and to manipulate its powers for the benefit of mankind. That drive to understand nature is matched in Enlightenment thought by an optimism at least as strong about the possibility of understanding society and human nature. In one aspect, this optimism is the basis of modern sociology, history and economics. But it is also the source of certain normative attitudes towards social and political justification - an impatience with tradition, mystery, awe and superstition as the basis of order, and a determination to make authority answer at the tribunal of reason and convince us that it is entitled to respect. The social world, even more than the natural world, must be thought of as a world *for us* (for *each* of us) - a world whose workings are to be understood by the active enquiries of the individual mind, not by religious dogma, mindless tradition or the hysterics of communal solidarity.

4 The economic side of human nature

Liberals accord intrinsic value to people as individuals, and attach particular importance to each individual’s

capacity to organize a life on their own terms. What terms are these likely to be? What nature of beings are these whose individual freedom we value? And what are the uses to which their freedom is likely to be put?

Critics commonly associate liberal individualism with an egoistic and acquisitive view of human nature. They say the classic liberals all gave pride of place, among human motivations, to the desire for power, pleasure and material possessions. Humanity, they argue, is reduced in liberal theory to nothing more than a competitive mass of market individuals - voracious consumers with unlimited appetites, hostile or indifferent to the well-being of others, and requiring no more of their political and legal institutions than that they secure the conditions for market activity.

The picture is not entirely a distortion. Liberal individualism does recognize that individuals' interests do not necessarily or naturally harmonize with one another. Each individual has a life of their own to lead, and there is no guarantee that one person's desires will not conflict with another's. Sometimes, as in Hobbes's theory, this is represented as an inherent hostility, a competitive diffidence and a 'mutuall will of hurting', issuing in a 'war of all against all'. Sometimes, as in John Rawls' work, it is seen simply as a postulate of mutual disinterest (rather than hostility). Mostly it is seen in Immanuel Kant's words as a matter of the 'unsociable sociability of men' - that there are things we share in common, things that drive us to society, things we can only accomplish together, as well as aspects of our nature that make us prickly, adversarial and wilfully isolated individuals.

Moreover, although - as we shall see - liberals believe that there are terms on which individuals with diverse or even opposed interests can live in peace with one another, it has never been part of their political philosophy that reason, enlightenment or socialization would put an end to this basic diversity or competitiveness. (To the extent that Rousseau suggests that the social contract might produce 'a remarkable change in man', his speculations take him outside the liberal tradition.) In the nature of things, humans will inevitably come up with diverse and opposed views of what makes life worth living, while the exigencies of our situation in the world - the moderate scarcity of material resources and our vulnerability to one another - will always furnish the raw materials for anxiety, competition and conflict (see Human nature).

A related objection is that liberals subordinate politics to economics: they see political structures merely as instruments for securing economic peace and market interaction, and they ignore the higher calling for the state outlined, for example, in the theories of Aristotle, Hegel or Hannah Arendt.

The image is accurate, but it is not clear why it should be regarded as an objection. Certainly, liberals do not regard participation in politics as an end in itself; unlike the civic humanists, they do not think that the most important virtues and activities are those oriented towards politics and the formal exercise of power over others (see Republicanism). It does not follow that they think of political participation as a narrow self-interested enterprise. In political science, the term 'liberalism' is commonly associated with interest group politics, but philosophical liberals are about equally divided on the question of whether voters in a democracy should orient their decisions to the common good or to their own interests (with the common good emerging as some sort of resultant from the political process). The point, however, is that even those who believe we should vote on our views about the common good still maintain that politics is, in the end, a means to promote the interests of individuals (*all* individuals), not an end in itself. They may believe in Rousseauian democracy, they may even hope that democratic participation can bring out the best in people (although many are dubious about that), but their firmest conviction is that individuals have interests and purposes of their own to pursue which have nothing intrinsically to do with politics or the state, and that the function of government is to facilitate those individual purposes not judge them or replace them with political or social ones.

To say that these purposes are individual is not necessarily to say that they are economic or materialistic in their content. It is surprising, in fact, how few liberal theorists have actually held the economically acquisitive picture of human nature. Hobbes did, certainly, and so did some of the eighteenth-century political economists. But many others in the liberal tradition see material motives as means to individual ends that may well be ethical, even spiritual in their content. John Locke, notorious in some circles as the apostle of possessive individualism, insisted that our primary mission in life is to ascertain what our creator requires of us in the way of conduct and worship: 'the observance of these things is the highest obligation that lies upon mankind, and…our utmost care, application and diligence, ought to be exercised in the search and performance of them because there is nothing in this world that is of any consideration in comparison with eternity' (1689 (1991): 42). Modern liberals, too, tend to stress the

ethical and cultural character of individual pursuits. We each have our own conception of happiness or the good life - a view about what makes a life worth living - and it is the diversity of individual ideals of this kind that political structures must accommodate.

In general, there is an intriguing ambivalence in the liberal tradition about whether this shift from economic to ethical individualism presents the social problem as more or less intractable. On the one hand, it seems to make the situation look better. Economic conflict is a zero-sum game: what you have I cannot have, or, worse still, what you have puts you in a better position to take what I have away from me; I therefore have an excellent reason of self-protection to deprive you of as many resources as I can. Ethical and spiritual individualism, by contrast, seems less intrinsically competitive: 'one man does not violate the right of another', wrote Locke, 'by his erroneous opinions…nor is his perdition any prejudice to another man's affairs' (1689 (1991): 42). The appropriate social posture for religious or ethical individualists seems to be the mutual indifference of Rawls' theory rather than the competition or conflict of Hobbes.

In fact, of course, that has not been our experience. Wars of religion have been at least as deadly as wars for territory or resources. We may think of commercial life as bland or shallow, but there is a certain sense of relief in Voltaire's comment about the London Stock Exchange: 'Here Jew, Mohammedan and Christian deal with each other as though they were all of the same faith, and only apply the word infidel to people who go bankrupt' (1734 (1980): 41). Even Hobbes, whose *Leviathan* is the *locus classicus* for the economic war of all against all, was adamant that the problem of the struggle for resources could be solved, since people would be willing to make concessions to a strong state that could keep the peace. But sectarian religious fervour he thought of as a form of madness, and he doubted whether the partisans of rival religious ideals could ever come to terms with one another (see Toleration).

Critics of liberalism will no doubt persevere in their charge that the tradition flatters the materialistic side of human nature at the expense of cultural and spiritual aspirations. They will say that liberals have paid too much attention to the ways in which political and legal structures can foster market economies and too little to the contribution they can make to the quality of ethical choice. A number of liberal writers have taken this criticism seriously. Joseph Raz (1986), for example, has argued that there is an aspiration towards value in the very concept of autonomy, so that a liberal commitment to freedom should not be thought of as incompatible with a social commitment to ethical perfectionism (see Perfectionism).

There comes a point, however, when liberal philosophers simply have to stand up and defend the channelling of political energies towards the real (but apparently soluble) problems of famine, plague and poverty, and away from moral, cultural and religious disputes, which promise little more in the way of progress than war, sectarianism and cults of ethical correctness. The preoccupation with economics is not based on scepticism about the ethical or spiritual dimensions of human life. It is based rather on a moderate sense of what politics can and cannot accomplish in a world where people disagree about God, value and the meaning of life, but largely converge in their desire to avoid hunger and disease and to better their material conditions.

5 The social contract

To the extent that liberal philosophy emphasizes diversity and conflict among individual purposes, it seems to steer us in the direction of anarchism. For what set of actually existing or realistically practicable institutions could possibly accommodate the individualism we have outlined?

In fact, liberal theorists have always held that something like the modern state, with its familiar institutions of law and representative government, *can* be made legitimate. They reject both the anarchist view that freedom is vindicated only in the absence of state authority and obligation and the utopian premise that a just society presupposes a radical change in human nature. This puts them in a rather ambivalent position in so far as existing 'liberal' societies are concerned - meaning here the constitutional democracies of North America, western Europe, Australasia and Japan. Liberalism has been a remarkably successful political ideology, inasmuch as its leading principles - freedom, toleration and equality before the law - have been accepted as part of the self-image or public relations of the world's most powerful and prosperous societies. Its proponents are uneasy, however, with the common inference that the social, economic and political reality of these societies is what liberal principles amount to in practice (just as Marxists were uneasy about the presentation of the Soviet Union and its satellites as

‘actually existing socialism’). They insist, quite properly, that liberalism is a set of critical principles, not an ideology or rationalization, and that it provides a basis for condemning things like deepening poverty, secretive and oligarchical government, legal abuses and the continuing legacy of racism and sexism in modern democracies. Even so, the existence of the self-styled ’liberal’ democracies is important. It helps sustain the sense that liberalism is a reformist rather than a revolutionary creed and that we already know, at least in outline, what a truly liberal society would be like. That sense of moderate reformism is not just a strategic ideological advantage. Liberalism claims to respect men and women as they are. It does not require generation after generation to undergo sacrifice for the sake of an endlessly postponed utopia. It suggests instead that political structures can be set up in now a way which represents a safe and reasonable compromise among individuals’ disparate demands.

In its classical form - in the writings of Hobbes, Locke, Rousseau, and to a lesser extent Kant - the argument from liberal premises to the legitimacy of something like the modern state was presented in terms of the social contract (see Contractarianism). The argument goes something like this.

Imagine people living outside any framework of political authority, exercising the right to direct their own lives and their own dealings with one another, in what liberal philosophers have called ‘the state of nature’. Using this as a baseline, try to model the development of political institutions as a way in which individuals exercise their freedom not as a way in which their freedom is abrogated.

The social contract model represents the functions of government in terms of a set of difficulties that such people would face in a state of nature. Conceptions of these difficulties vary. There may be an internecine struggle for resources (Hobbes and Hume); there may be an appreciation of others’ rights, but no reliable mechanism for enforcing them (Locke); or there may be disagreement about justice, with each person seeking to do what seems right or good in their opinion (Kant). The common element is that people in the state of nature would lack a reliable sense of what they could count on in social and economic life. What they most need is a secure set of rules, impartially administered and enforced, to provide a framework in which peaceful cooperation and long-term production is possible.

Government, then, is represented in terms of an agreement by each person with all of the others (in a given territory) to cooperate in the institution and maintenance of permanent rule-making and rule-enforcing agencies. The contract is not an agreement between the government and the individual. Instead, it presents legal and governmental institutions as structures of cooperation among individuals, and it uses that idea as a basis for deriving limits on governmental powers and restraints on particular individuals’ or factions’ exploitation of those powers. Tyrannical exploitation and arbitrary government are ruled out on this conception, inasmuch as they cannot be represented as any sort of improvement over the situation individuals would face if they tried to live without any political institutions at all.

Some theorists, Rousseau and Rawls for example, use the social contract idea also as a way of thinking about the content of legal rules: we can discuss what rights we have and the just distribution of resources by asking what assurances would have to be given to each individual to secure their consent to the basic structure of social arrangements. In the hands of these theorists, the social contract is a test of substantive political justification. Others see it in more procedural terms: the social contract models the construction of political and constitutional mechanisms, which will then work out substantive solutions on a basis that is relatively independent of the contract idea. Hobbes’s theory is the most extreme example of this: the Hobbesian contract is simply an agreement to authorize an individual or organization (a ’sovereign’) to solve the problems that generate conflict in the state of nature in any way that promises improvement. The absolute authority with which the sovereign is thus endowed has led some to deny that Hobbes is a liberal. Certainly, he is not wedded to the positions identified with liberalism at the beginning of this article. But his underlying political philosophy is liberal: his value premises are individualistic, and he is unyielding in his view that political institutions (with the powers he accords them) must be justified in relation to the interests of each individual, as well as in his optimism that such justification is possible.

Not all liberal philosophers appeal to the idea of the social contract. Many prefer to develop their theories without the mediation of this model. The arguments of J.S. Mill’s 1859 essay *On Liberty*, for example - which many regard as the quintessential statement of liberal principles - are presented directly as claims that individuals are entitled to make, against their society and their government, without any historical pretence that governments were set up by

individuals to validate those claims (see Law, limits of). Others use the social contract to justify some but not all of the political constraints they propose. In the theories of John Locke and Robert Nozick, the social contract argument presupposes a distribution of 'natural' property rights. For them, the function of the social contract is to support and police these rights, not reconceive or redistribute them. In other words, Locke and Nozick propose that property rights should be justified directly in moral argument, without appealing to the social contract idea. I suspect that something like this is true of all liberal theories. The social contract is an intermediate rather than a fundamental idea: one that presupposes that individuals are free, equal and rational, and that political power requires a justification which connects with the interests of each of them.

When it is understood in this way, as a method of modelling the force of certain deeper assumptions or theorems about justification, the social contract can be used as a purely hypothetical device in normative argument. As Kant and Rawls have pointed out, we need not be embarrassed by the fact that no such contract ever took place. It is still a useful test to apply to a constitution or to a set of laws. For if we conclude, even hypothetically, that our laws or our constitution would not have commanded the agreement of all those who are constrained by them, we will have discovered a significant dissonance between our political arrangements and the fundamental (pre-contract) notion of respect for each individual - a dissonance that ought to be of concern to liberal philosophers whether they are interested in the niceties of contract theory or not.

JEREMY WALDRON

References and further reading

Arblaster, A. (1984) *The Rise and Decline of Western Liberalism*, Oxford: Blackwell.(A critique of liberal political thought from the sixteenth to the twentieth centuries.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, eds J.H. Burns and H.L.A. Hart, London: Athlone Press, 1970.(Classic statement of utilitarian morality.)

Berlin, I. (1969) *Four Essays on Liberty*, Oxford: Oxford University Press, 118-72.(Develops the distinction between negative and positive freedom, discussed in §3.)

Bramstead, E.K. and Melhuish, K.J. (1978) *Western Liberalism: A History in Documents from Locke to Croce*, London: Longman.(A comprehensive collection of statements by liberal politicians and statesmen in the European tradition, as well as liberal political philosophers, together with a commentary by the editors.)

Dworkin, R. (1977) *Taking Rights Seriously*, London: Duckworth.(A theory of law and political morality centred on the idea that individual rights sometimes 'trump' utilitarian justifications.)

Dworkin, R. (1978) 'Liberalism', in S. Hampshire (ed.) *Public and Private Morality*, Cambridge: Cambridge University Press.(An argument that moral equality lies at the heart of liberalism.)

Freeden, M. (1978) *The New Liberalism; An Ideology of Social Reform*, Oxford: Oxford University Press.(A useful discussion of 'the new liberalism' referred to in §1.)

Green, T.H. (1886) *Lectures on the Principles of Political Obligation*, ed. B. Bosanquet, London: Longman, 1941.(An influential work combining liberal and Hegelian themes.)

Hayek, F.A. (1960) *The Constitution of Liberty*, London: Routledge & Kegan Paul.(A statement of the connection between liberty and the rule of law, and a critique of the modern welfare state.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.(Classic statement of liberal premises of economic individualism and the war of all against all, leading to the contractual institution of an absolute sovereign.)

Hobhouse, L.T. (1964) *Liberalism*, New York: Oxford University Press.(An example of 'the new liberalism' referred to in §1.)

Holmes, S. (1993) *The Anatomy of Antiliberalism*, Cambridge, MA: Harvard University Press.(A vigorous defence of liberal political theory and source of some of the arguments about economic liberalism in §4.)

Hume, D. (1739) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, rev. P.H. Nidditch, Oxford: Clarendon Press, 1978, Book III.(A classic account of the emergence of property and justice.)

Kant, I. (1991) *Political Writings*, ed. H. Reiss, Cambridge: Cambridge University Press, pp. 73-87 and 131-64. (Various writings in this volume insist both on the hypothetical nature of social contract reasoning, and - within the contract model - on the duty of individuals to enter and remain in political society with those with whom they find themselves disagreeing about justice.)

Locke, J. (1689) *A Letter Concerning Toleration*, ed. J. Horton and S. Mendus, London: Routledge, 1991. (Referred to in the discussion of economic versus spiritual versions of liberalism in §4.)

Locke, J. (1690) *Two Treatises of Government*, ed. R. Tuck, Cambridge: Cambridge University Press, 1988.(The

paradigmatic statement of liberal contract theory.)
Manning, D. (1976) *Liberalism*, London: Dent.(A brief overview of liberal political philosophy.)
Mill, J.S. (1859) *On Liberty*, ed. C.V. Shields, Indianapolis, IN: Bobbs-Merrill, 1956.(A defence of individuality and freedom of thought and discussion, regarded by many as the most direct statement of liberal principle.)
Mill, J.S. (1868) 'The Subjection of Women', in *J. Mill and H. Taylor, Essays on Sex Equality*, ed. A. Rossi, Chicago, IL: University of Chicago Press, 1970.(The first sustained statement of gender equality by a philosopher in the classic liberal tradition.)
Nozick, R. (1974) *Anarchy, State and Utopia*, Oxford: Blackwell.(A vigorous modern defence of private property and the minimal state, in the tradition of John Locke.)
Rawls, J. (1971) *A Theory of Justice*, Oxford: Oxford University Press.(Perhaps the most famous construction of liberal theory in modern times, using the idea of a hypothetical contract to explore issues of justice and fairness.)
Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(A defence of the claim that liberal principles of justice must command support among a wide variety of ethical and philosophical conceptions in a modern pluralist society.)
Raz, J. (1986) *The Morality of Freedom*, Oxford: Clarendon Press, 165-216 and 369-430.(An exploration of the connection between autonomy and perfectionism, mentioned in §4.)
Rousseau, J.-J. (1762) *The Social Contract and Discourses*, trans. G.D.H. Cole, London: Dent, 1955.(A version of contractarian theory that teeters on the brink between liberal and non-liberal political thought.)
Voltaire, F.-M.A. de (1734) *Letters on England*, trans. L.Tancock, Harmondsworth: Penguin, 1980.
Waldron, J. (1987) 'Theoretical Foundations of Liberalism', *Philosophical Quarterly* 37: 127-50.(A discussion of the difficulty of defining 'liberalism' and of the connection between liberalism and Enlightenment thought, mentioned at the end of §3.)
Wootton, D. (1986) *Divine Right and Democracy: An Anthology of Political Writing in Stuart England*, Harmondsworth: Penguin, 285-317.(Includes a transcript of the Putney Debates of 1647 in which Colonel Rainsborough gave voice to the principle of liberal equality, discussed in §3.)

Liberalism, Russian

Unlike early English liberalism which stressed individual freedom from state control and from the 'tyranny of the majority', Russian liberalism generally emphasized the importance of legality in government, the state's positive role as guarantor of civil liberty, and the gradual achievement of social justice through reform. In the century between Peter the Great's death in 1725 and the Decembrist uprising of 1825 various politicians and thinkers proposed the introduction of representative institutions into the Russian government and recommended that serfdom be abolished in the Empire. These proposals reflected admiration for Western European models of government and the impact of the Enlightenment on Russia's ruling elite. Because the autocracy ultimately rejected these proposals and made public discussion of them all but impossible, liberalism did not take root in this period. The genesis of Russian liberalism as a philosophically elaborated, politically coherent movement occurred after the death of Tsar Nicholas I in 1855, when the government eased censorship and announced its commitment to peasant emancipation. Mid-nineteenth-century Russian liberalism owed its intellectual inspiration to Hegelianism and French juste-milieu *liberalism. Russian liberals argued that social progress in the empire had almost always come about at the state's initiative; they could scarcely imagine building a just society without the cooperation of a strong state.*

At the beginning of the twentieth century Russian liberals confronted what proved to be an insuperable challenge: how to establish a viable constitutional order in an empire riven by social and ethnic strife where neither government nor the powerful socialist movement favoured a rule-of-law state. Wrestling with this difficulty, the liberals split into two factions: a 'left' or 'radical' wing that valued social justice over the sanctity of property rights; and a 'right' or 'conservative' wing that valued legal equality over social equality, and therefore interpreted socialism as a species of utopianism. Between 1905 and 1917 the liberal movement reached its political zenith, as liberal politicians exercised their influence in the State Duma and opposition movement. Following the October 1917 Bolshevik Revolution liberalism was banned in the Soviet Union. It re-emerged, albeit in altered form, during communism's collapse when neo-liberals claimed for liberalism a prominent position in the 'normal' constellation of Russian political forces. The main goals of post-Soviet liberalism are the defence of civil and political rights, the establishment of the rule of law, the assertion of individual property rights and the gradual construction of a market economy.

1 Absolutism, Enlightenment and proto-liberalism

The post-Petrine Russian government was a unitary absolutist state that, in theory, allowed neither individual nor corporate autonomy. Until the mid-nineteenth century the state depended on a tacit alliance between the Crown and the landed nobility dedicated to the continuation of serfdom. Between 1725 and the Decembrist uprising in 1825, however, various politicians and thinkers - among them Count Dmitrii Golitsyn (1663-1738), Count Nikita Panin (1718-83), Denis Fonvizin (1744-92), Nikita Murav'ëv (1796-1843), Mikhail Speranskii (1722-1839) and Nikolai Novosil'tsov (1761-1836) - proposed to establish new agencies to limit the Tsar's personal authority or to subject the Crown to the higher authority of a fundamental law; in certain instances, these proposals were accompanied by recommendations that serfdom be abolished. Historians have sometimes interpreted these reformist efforts as the origins of Russian liberalism.

However, only two eighteenth-century Russian thinkers elaborated philosophical justifications for liberal government, Semën Desnitskii and Aleksandr Radishchev. Desnitskii (?-1789), a professor of law at Moscow University, was influenced by the Scottish Enlightenment, particularly by Adam Smith whose lectures he heard at Glasgow and whose *Theory of Moral Sentiments* he often quoted. Desnitskii divided social evolution into four ascending stages - hunting, pastoral, agricultural and commercial society - each having a legal system peculiar to it. He thought that representative government was appropriate to commercial society; popular elections expressed the electorate's historically conditioned deference towards the wealthy; and that the duty of the wealthy was to serve the less fortunate, as the dictates of virtue require. Desnitskii criticized Russian absolutism as lawless and arbitrary. He proposed that the imperial government establish an elective Senate, with representatives from the commercial classes as well as the nobility, to advise the crown on legislation. He was one of the first Russian advocates of civil rights for women.

Radishchev (1749-1802) has usually been classified as a revolutionary because he rejected nobiliary privileges, noted that tyrannicide has sometimes been the penalty for despotism and predicted a peasant revolution. However, he was far from consistent and, with equal justice, may be considered a forerunner of liberalism. Radishchev's principal commitment was to equality before the law. He followed Mably in asserting that good government requires every citizen, including the sovereign, to submit to the law; like Mably and Rousseau, he thought that government should depend upon citizens' mutual consent without which the social compact dissolves. He admired the English and American systems for their devotion to legality and encouragement of an active citizenry. His famous tract, *A Journey from St Petersburg to Moscow* (1790), can be read as an explicit warning to the Russian government and nobility to mend their ways, lest the peasants rise in bloody revolt. Although he was a republican, Radishchev did not support the Jacobin terror, because the terrorists violated civil rights. Unfortunately for Russia, Desnitskii's work was quickly forgotten and Radishchev's tract was burned by the imperial authorities (see Enlightenment, Russian).

2 Mid-nineteenth century liberalism

Russian liberalism's crystallization in the 1850s owed much to external circumstances, especially the Crimean debacle, the death of Tsar Nicholas I and the accession to the throne in 1855 of the reform-minded Alexander II. However, these events only punctuated the evolutionary process in which a portion of educated society committed itself to liberal values. During the 1840s, at the height of the Westernizer-Slavophile debate over Russian national identity, the Westernizers defined Russia as a member of the European family of nations. The historian Timofei Granovskii (1813-55) claimed that, under the aegis of the modern state, European societies had been gradually advancing towards political liberty. He hinted that even in Russia this movement would culminate with a *Rechtsstaat* wherein all individuals would have the possibility to develop their potential freely. In 1847 the jurist Konstantin Kavelin (1818-85) published a Westernizer manifesto, 'A Brief Survey of Juridical Relations in Ancient Russia', asserting that Peter I's reforms represented a victory for the principle of free individuality. Kavelin hoped that the contemporary Russian state and educated society would build upon this victory by inaugurating broad new reforms. In St Petersburg during the late 1840s and early 1850s he studied the peasant question in collaboration with other junior officials interested in serfdom's abolition; his 1855 'Memorandum on the Emancipation of Peasants in Russia' was probably the most compelling justification for ending serfdom written by a Russian intellectual. By the early 1840s, Granovskii, Kavelin, the literary critics Pavel Annenkov (1813-77) and Aleksandr Druzhinin (1824-64), the historian Sergei Solov'ëv (1820-79) and the oft-derided 'philosophical tea-merchant' Vasilii Botkin (1811-69) had arrived at a consensus on the values they thought essential to free society: respect for law, individual dignity, freedom of conscience, toleration of intellectual diversity and hostility towards serfdom. All agreed that, under Russian conditions, the state would have to assume primary responsibility for realizing these values in practice. The Westernizers' worldview was underpinned by Hegelianism, a current that dominated Russian thought during the second quarter of the century (see Hegelianism, Russian). Hegel's evolutionary conception of history, his efforts to reconcile the thinking individual's autonomy with the collective life of the *polis*, his antipathy to unmediated state power and his support for constitutional monarchy implicitly informed the Westernizers' approach to Russia's problems.

The credit for transforming the amorphous Westernism of the 1840s into a recognizable liberal programme belonged to Boris Chicherin (1828-1904), arguably the leading Russian liberal theoretician of the century. He thought that liberals should concentrate on achieving seven objectives: freedom of conscience; peasant emancipation; freedom of speech; freedom of the press; academic freedom; publicity of all government activities; and public legal proceedings. In his classic *On Popular Representation* (1866), he sharply distinguished between the civil rights mentioned above and political rights such as voting and representation in a national assembly. He criticized Western liberals such as John Stuart Mill for mistakenly identifying liberalism with the achievement of political rights. He maintained that Russians should accustom themselves to civil rights and the legal culture entailed by such rights before establishing a constitutional government. His faith in monarchy as an engine of legal reform, his hesitations about the advisability of constitutional government in Russia, his bitter opposition to socialism, and his allegedly doctrinaire Hegelianism led many contemporaries to consider him a conservative or even a reactionary. He advertised himself as a 'conservative-liberal', professing faith in the platform 'strong government and liberal measures'. Nevertheless, late in his life Chicherin laid greater stress on individual freedom and the need for a constitution. His last major work, *Philosophy of Law* (1900), was a sweeping defence of

individual freedom in the spirit of what Friedrich von Hayek later called ‘classical liberalism’. Chicherin’s individualism profoundly affected the philosophers Nikolai Berdiaev (1874-1948), Pavel Novgorodtsev (1866-1924) and Pëtr Struve (1870-1944) who came to share his aversion to socialism and collectivism.

Westernism was also the main source of Aleksandr Herzen’s ‘Russian socialism’, a doctrine that contributed to the birth of Russian populism and anarchism but that has sometimes been seen as affiliated with the liberal tradition. Herzen (1812-70) was a lifelong believer in individual freedom, an opponent of tyrannous abstractions and an occasional advocate of republican government. His greatest champion, Sir Isaiah Berlin, regards him as a harbinger of contemporary liberal pluralism; others have associated him with libertarianism and the Russian federalist movement. However, Herzen was famously dismissive of European *juste-milieu* liberalism. He hated the modern bureaucratic state, wished for the triumph of a ‘social republic’ based on the peasant commune and anticipated a new socialist age in history. His revolutionary and millenarian side had little or nothing in common with the statist liberalism of Chicherin (see Herzen §3).

3 Late imperial liberalism

Between 1890 and 1917 Russian liberals redefined their political agenda. They pressed the government for a constitution and national elections, a concession reluctantly granted by Nicholas II in October 1905. In the State Duma from 1906 to 1917 liberal parties played a signal role, but they overcame neither the Tsar’s scepticism about representative government nor socialist opposition to the legal order. During this turbulent period liberal thinkers were divided over which objective deserved priority - legality or social justice. Left-wing liberals sought the rule of law as well as social justice but generally valued the latter over the former. Their practical aims - land redistribution from the nobility to the peasantry, an eight-hour day for workers, and a social-insurance state - overlapped with the moderate socialist agenda so extensively that left liberals tended to think of socialists as their natural allies. Indeed, Pavel Miliukov (1859-1943), Russia’s best-known liberal politician, boasted in 1905 that ‘our programme is undoubtedly the most leftist of all those advanced by similar political groups in Western Europe’. The philosophical foundations of left liberalism derived from disparate religious and secular sources. The religious philosopher Vladimir Solov’ëv (1853-1900) combined social Christianity with a Kantian insistence on individual dignity and the objectivity of ethics when he argued that a good society must guarantee to each member that minimum of welfare necessary for a dignified existence. Although Solov’ëv himself cannot be counted a left liberal, his ideas developed a wide currency in the younger generation which sought in Christian egalitarianism a solution to Russia’s social problems. The Ukrainian liberal Bogdan Kistiakovskii (1868-1920), influenced by the Austrian *Kathedersozialist* Anton Menger, discovered a historical continuity between liberalism and socialism: he predicted that civil rights and the liberal *Rechtsstaat* would prepare the way for social rights and a socialist rule-of-law state. After the 1917 revolutions, the liberal-socialist synthesis reached its culmination with Sergius Hessen (1887-1950) who predicted a new type of liberal state - one that would foster conditions in which citizens might choose the good without being hindered by economic privation.

The ‘right’ or ‘conservative’ liberals followed Chicherin in stressing that cultivation of individual freedoms and respect for law must take priority over redistributing wealth and the social insurance state. In 1909 authors of the *Vekhi (Signposts)* symposium accused Russian radicals, including left liberals, of supporting social equality at the expense of legality and individual freedom (see Signposts movement). The historian Mikhail Gershenzon (1869-1925), an outspoken defender of individualism, described egoism as the ‘great force’ that had made of the Western bourgeoisie ‘God’s unconscious instrument on earth’. Struve’s essay in *Signposts* attacked the radicals’ proclivity to ‘stand aside’ from the government rather than cooperate with it, to preach revolutionary destruction rather than to seek self-perfection as true lovers of liberty ought to do. Connecting all the *Signposts* essays was the assumption that the inner, especially religious life of the individual ‘is the only sound foundation for any social structure’. A devastating critique of socialism from a conservative liberal perspective came from Novgorodtsev, whose *On the Social Ideal* (1911-16) interpreted Marxism as a doomed millenarian religious movement aimed at achieving the Kingdom of God on earth; Novgorodtsev juxtaposed to Marxism belief in a transcendent God, in natural law, in a *Rechtsstaat* and in individual dignity. During the political crises of 1917-18 several conservative liberals, among them Struve and Novgorodtsev, supported a temporary military dictatorship as an alternative to socialism. This policy showed the degree to which right liberals feared socialism and ‘revolutionary democracy’, as well as the unpalatable choices that confronted all Russian liberals at the end of the old regime.

Of course, the lines between left and right liberals were far from absolute. Struve moved from left to right between 1902 and 1909, whereas Miliukov tried to bridge the distance between the two wings of the party. Some thinkers - Novgorodtsev is a good example - at times straddled the left-right philosophical divide. In any event, the two groups shared a central conviction, - namely that a strong government is a necessary precondition of liberty, however defined. Thus, in the main Russian liberalism was state-centred rather than federalist, pluralist or libertarian.

4 The revival of liberalism in post-Soviet Russia

Soon after the Bolshevik Revolution, the Soviet government banned liberal political parties. From 1917 to 1991 the Soviets treated liberalism as an ideological deviation punishable by public obloquy, arrest, hard labour or even death. Public advocacy of liberal values was therefore out of the question; sadly, the academic study of liberalism's legacy was circumscribed. Even before 1991, however, dissident intellectuals supported certain elements of the liberal worldview, especially guarantees of civil rights and government legality. During the 1970s, the most remarkable of these thinkers, the physicist Andrei Sakharov (1921-90), sought to combine socialist welfare with civil and political rights of the 'Western' type. Between the mid-1970s and early 1980s he jettisoned his support for Soviet power; before his death in 1990 he was preparing a draft constitution the central tenets of which were guarantees of human rights and the division of government authority. Philosophically, Sakharov was a *naïf*, but he succeeded in endowing his humane vision with the force of common sense. After the collapse of the USSR in 1991, self-described liberals devoted themselves to two short-term objectives: destroying the Communist Party's hold over society and demolishing the command economy. Their long-term goals were the defence of civil and political rights, the gradual construction of a market economy, and creation of a rule-of-law state. Apparently, none of these goals was perceived as inherently incompatible with revitalization of a powerful centralized Russian state. Thus, while many liberals enthusiastically supported ending the Soviet empire and permitting devolution of certain prerogatives to local or regional levels of government, they also wanted a powerful Russian presidency in Moscow. Whether post-Soviet liberalism can contribute to the diminution of economic and political strains generated by the disappearance of the communist system, whether it can bridge the gulf between a strong state and individual liberty, and whether it can manage to balance property rights with social equality remains to be seen. At the close of the twentieth century, liberalism no longer exists as it did for so long in Russia's past, clandestinely, 'behind seven seals, beneath seven locks', as an untried, mysterious and therefore badly understood alternative to absolutist government.

See also: Liberalism

G.M. HAMBURG

References and further reading

Berdiaev, N.A., Gershenzon, M.O., Kistiakovskii, B.A., Struve, P.B. *et al.* (1909) *Vekhi. Sbornik statei o russkoi intelligentsii*, Moscow; B. Shragin and A. Todd (eds) *Landmarks: A Collection of Essays on the Russian Intelligentsia*, New York: Karz Howard, 1977.(An able translation of the controversial right liberal anthology.)

Berlin, I. (1978) 'Alexander Herzen', in H. Hardy and A. Kelly (eds), *Russian Thinkers*, New York: Viking Press, 186-209.(A wonderful, sympathetic portrait of Herzen as pluralist.)

Fischer, G. (1958) *Russian Liberalism, from Gentry to Intelligentsia*, Cambridge, MA: Harvard University Press.(The only book-length overview of post-emancipation liberalism in English; argues that the 'new liberalism' was championed by liberal intelligentsia.)

Gooding, J. (1986) 'The Liberalism of Michael Speransky', in Slavonic and East European Review 64 (3): 401-24.(Against Raeff's portrayal of Speranskii's conservatism, this article makes a strong case for Speranskii as progenitor of Decembrism and moderate liberalism.)

Hamburg, G.M. (1992) *Boris Chicherin and Early Russian Liberalism 1828-1866*, Stanford, CA: Stanford University Press.(Examines the connection between Westernism and early liberalism.)

Karpovich, M. (1955) 'Two Types of Russian Liberalism: Maklakov and Miliukov' in E.J. Simmons (ed.), *Continuity and Change in Russian and Soviet Thought*, Cambridge, MA: Harvard University Press.(A penetrating analysis of the left-right split in liberalism, *c.* 1905-7.)

Leontowitsch, V. (1957) *Geschichte des Liberalismus in Rußland*, Frankfurt-am-Main: Vittorio Klostermann.(Controversial survey of liberalism from the eighteenth to twentieth centuries, concentrates on right liberalism

in government and educated society.)

Madariaga, I. de (1984) 'Portrait of an Eighteenth-Century Russian Statesman: Prince Dmitry Mikhaiylovich Golitsyn', in Slavonic and East European Review 62 (1): 36-60.(A model article on Golitsyn's intellectual/political milieu.)

Mironenko, S.V. (1990) *Stranitsy tainoi istorii samoderzhaviia. Politicheskaia istoriia Rossii pervoi poloviny XIX stoletiia*, Moscow: Mysl'.(A shrewd assessment of Alexander I's and Novosiltsov's liberalism.)

Offord, D. (1985) *Portraits of Early Russian Liberals*, Cambridge: Cambridge University Press.(A vividly written but subtle study of liberalism's genesis.)

Pipes, R. (1970) *Struve. Liberal on the Left, 1870-1905*, Cambridge, MA: Harvard University Press.(A compelling portrait of Struve's intellectual odyssey from Marxism to left liberalism.)

Pipes, R. (1980) *Struve. Liberal on the Right, 1905-1944*, Cambridge, MA: Harvard University Press.(A magisterial work on the leading conservative liberal of the twentieth century.)

Radishchev, A.N. (1790) *Puteshestvie iz Peterburga v Moskvu*; trans. and ed. R.P. Thaler, *A Journey from St. Petersburg to Moscow*, Cambridge, MA: Harvard University Press, 1958.(As discussed in §1.)

Raeff, M. (1957) *Michael Speransky: Statesman of Imperial Russia, 1772-1839*, The Hague: Martinus Nijhoff. (The classic biography reading Speranskii as a conservative.)

Raeff, M. (ed.) (1966) *The Decembrist Movement*, Englewood Cliffs, NJ: Prentice Hall.(Includes Muraviev's constitutional project.)

Ransel, D. (1970) 'Nikita Panin's Imperial Council Project and the Struggle of Hierarchy Groups at the Court of Catherine II', in *Canadian Slavic Studies* 4 (3): 443-63.(A sceptical view of Panin's constitutionalism.)

Read, C. (1979) *Religion, Revolution and the Russian Intelligentsia 1900-1912: The Vekhi Debate and its Intellectual Background*, New York: Barnes & Noble.(A monograph on the *Signposts* anthology and its religious dimension.)

Reddaway, W.F. (ed.) (1931) *Documents of Catherine the Great*, Cambridge: Cambridge University Press.(A useful collection of documents on Catherine's attitude towards the Enlightenment.)

Sakharov, A. (1968) *Progress, Coexistence and Intellectual Freedom*, New York: Norton.(Sakharov's early reflections on reconciling intellectual liberty with socialism.)

Sakharov, A. (1975) *My Country and the World*, New York: Alfred A. Knopf. (A trenchant criticism of Soviet socialism from a liberal humanist perspective.)

Schapiro, L. (1986) 'Liberalism and the Law', in E. Dahrendorf (ed.), *Russian Studies*, New York: Viking Press, 29-130. (Sympathetic essays on conservative liberalism.)

Walicki, A. (1979) *A History of Russian Thought from the Enlightenment to Marxism*, Stanford, CA: Stanford University Press.(The finest survey of Russian intellectual history with extended treatment of liberalism.)

Walicki, A. (1987) *Legal Philosophies of Russian Liberalism*, Oxford: Clarendon Press.(A brilliant exploration of six liberal legal philosophies in late imperial Russia and thereafter; very good on social liberalism.)

Liberation philosophy

Philosophy of liberation emerged in Argentina early in the 1970s with the explicit intention of proposing a liberating alternative to the diagnosis of structural dependence offered by the social sciences (particularly the so-called 'theory of dependence'). Some of the original intentions of liberation philosophy were to make poor and marginalized people the subjects, or authors, of philosophy and to collaborate in the process of distancing philosophy from academia and exclusively professional settings. Social conflict and pressing national needs were topics of debate at that time. All thought started with the recognition and assessment of the experience of alterity. Horacio Cerutti-Guldberg has proposed the phrase 'philosophies for liberation' as this kind of reflection deals with multiple philosophical positions and privileges the historical process over philosophy.

1 Initial postulations

In 1973 some theses were formulated outlining the goals of an incipient philosophical programme shared by some Argentine philosophers who were the driving force behind liberation philosophy. One goal was to create a universally valid, although historically situated Latin American philosophy. The intention was also to confront the situation of structural dependency legitimized through a self-serving and academicist philosophy as well as to explain philosophically the needs of the great exploited majority and the urgencies of the poor and oppressed Latin American people. These poor people would appear as the bearers of a historical innovation to be theorized through liberation philosophy.

Shortly after this programme was enunciated, it was divided into at least four discursive variants. These products of diverse theoretical foci and political and ideological positions had an impact on the philosophy. One could no longer justifiably speak of a single liberation philosophy as a movement or common programme. These variants presented in a very coherent way their theses regarding their point of departure, subject, methodology and conception of philosophy.

2 Four discursive variants

Ontologicist philosophy postulated as its point of departure a historical zero, as if one could begin with a *tabula rasa* - the mind in its uninformed state - regarding world philosophical tradition. The task of elaborating a new nondialectical rationality was undertaken. The task was to establish the most characteristic way of being in the world for the Latin American. The contrast between the Spanish verbs, *ser*, meaning the permanent state of *being*, as in 'I am Bill, or Jane' and *estar*, the changing state of *being*, as in 'I am happy, or sad' was emphasized as the key to understanding reality. It was the 'people' who were spoken about within the area of ontologicist philosophy as opposed to the 'liberal individual' or the 'proletarian class'. The philosopher would translate everyday events into philosophical language. In the face of Western rationalism, this variant contrarily valued and recognized the irrational dimensions of life. Gunther Rodolfo Kusch and Mario Casalla are representatives of this position.

Analectic philosophy was rooted in a criticism of Eurocentric or North Atlantic modernity. It rejected Latin American historical memory as purely imitative and therefore justifying its own domination. The imploring face of the poor would be the point of departure which would bring about analectic behaviour capable of opening up to all that is new and beyond the dialectic totalities. The philosopher would be a prophetic teacher capable of giving form to the message of those without a voice. The otherness and exteriority of the marginalized would be secured by divinity, the absolute Other. Subscribers to this viewpoint include Juan Carlos Scannone and Enrique Dussel.

Historicism warned that the origins of a philosophy which began with the experience of alterity should recognize its foundations in Latin American history. A historiographic reconstruction of Latin American philosophical memory constitutes a valuable aid in this effort. Contrary to the Hegelian model, which reduces political freedom to freedom of thought, historicism proposed concentrating on those subjects who reformulate popular demands and the way in which they reform them starting with their own interests. It defined itself as a comprehensive philosophy of the marginal and a new way of thinking which anticipated the future. Arturo Roig has applied himself to the theory of historicism.

Problematization of philosophy was the reconceptualization of philosophy, questioning the unfounded claims of originality and the attempt to begin from a historical zero. This version favoured the liberation process over

philosophy and reclaimed the express mediation of the social sciences and ideological self-criticism. It strove for the deprofessionalization of philosophy and articulated the historical dimension in conjunction with the systematic. Practitioners of this theory are Manuel Ignacio Santos, Severino Croatto and Horacio Cerutti-Guldberg.

3 Exterior and interior exile

In under three years, many people were working on the philosophy of liberation in each of its four variants. After the Argentinian military coup in 1976, a good portion of the philosophers who were adherents to the last three previously mentioned variants had already been removed from their teaching posts in the national universities. This was owing to various types of sophistry which led to expulsion or to contracts not being renewed. Some academics had already left the country; others had to abandon academic activity in order to survive, else they altered the scope of academic production. Practitioners of ontologicist philosophy yielded to the new repressive politics imposed by the dictatorship and there was no shortage of people who softened their positions or made them more flexible in order to coexist with the new regime.

Those who left the country, particularly Roig, Dussel and Cerutti-Guldberg, continued to advance and revise their positions, bringing about a wide and generalized debate which streamlined many theses, redefined foci and enriched initial arguments. Nevertheless, it seemed as though the basic positions were maintained to the point that works which can be situated in any of the four modalities mentioned have continued to be written.

4 Latin Americanization and globalization of the debate

The tireless work of those in exile promoted debate on the diverse modalities of this philosophy throughout the Latin American region and continued adding contingents of philosophers of diverse nationalities to the production of a body of thought which acquired regional nuances. In different parts of the world, especially in the USA and Europe, the debate on these issues was prolonged. The following points have been discussed extensively: the cognitive scope of the notion of the people as a collective of exploited social sectors united against all forms of dependency, attitudes towards historical and philosophical memory, the role of the philosopher as theorist or preacher, the ideological dimension of argumentation, the political consequences of the philosophical consideration of alterity, the insufficiencies of philosophy in the face of feminist or indigenous demands, conceptual ambiguities, interpretative difficulties of the same discursive modalities, the vague boundaries of the phenomenon of liberation and the relationships between the social sciences and philosophy.

The convergence of Latin American philosophy and liberation philosophy strengthened both traditions but also led to much confusion. An example of one type of confusion was the idea that Latin American philosophy has always been a liberation philosophy. From a historiographic viewpoint, this gives the term an excessive range.

5 Work in progress

After more than two decades of development of liberation philosophy, the branches of current research can be characterized as follows: ontologicism remains stable and does not have many variants, although studies of Kusch’s work are under way which would incorporate revised arguments. The progress of analectic philosophy is a little more convoluted. On the one hand, it bears a resemblance to ontologicist culturalism, while on the other hand, it is influenced by Marx whose work reveals antecedents and confirmations of the analectic version of liberation philosophy. An international and intercultural debate has arisen and has been advanced between analectic ethics of liberation and the ethics of German moral philosopher, Apel. From historicist and problematizing positions, works on the history of ideas in Latin America have increased and their methodological and epistemological connotations have been under debate. Several research projects are under way. The debate over the idea of utopia and its heuristic power has intensified and an improved theoretical formulation of the utopian is being researched. Concern grows for the rethinking of democracy, the state and connected phenomena, and an interesting debate between the theses of the philosophies of liberation, modernity and postmodernism has been outlined. After the assassination in 1989 of Jesuit Ignacio Ellacuría, a leading liberation scholar, and his companions in El Salvador, research has burgeoned to continue his effort to establish a tradition of Latin American liberation philosophy after Xavier Zuribi.

6 Perspectives

In its primary international arenas, philosophy of liberation has been identified sometimes exclusively with analectic philosophy. In reality, 'liberation philosophy' has gone beyond analectic philosophy, which is a long way from exhausting the concerns and emergent theories of the 1970s. The confusion between liberation philosophies and other variants of Latin American philosophy has not helped. A careful and nuanced division of positions and arguments should make feasible a qualitative advance in reflection.

Trapped in its formulations with an ambiguous or salvationist language without sufficient capacity to settle internal struggles, it is barely worthwhile advancing the idea of a homogeneous and unified philosophy of liberation. It would seem more fruitful to advance the progress of and research in the promotion of dialogue based on the following facts relating to Latin American and global reality: the fruitfulness of opening up to conceptual consequences of the experience of alterity, the growing increase in poverty as an almost uncontrollable phenomenon of the world economy, the intolerable injustice of the social structures of the world in the late twentieth century, the scientific and technological revolutions that modify the very guidelines of the political agenda and the demands for participation and the dissatisfaction of the masses. Many of the questions formulated through the philosophies for liberation directly address serious economic, social, cultural, religious, ethnic and political problems of the world in the late 1990s. The challenge consists of building better conceptual weapons to confront these problems theoretically and practically.

See also: Alterity and identity, postmodern theories of; Liberation theology

HORACIO CERUTTI-GULDBERG

References and further reading

Castro-Gómez, S. (1996) *Crítica de la razón latinoamericana (Criticism of the Latin American Intellect)*, Barcelona: Puvill Libros.(A critical analysis of the analectic and historicist variants of liberation philosophy in terms of postmodernism and postcolonialism.)

Cerutti-Guldberg, H. (1983) *Filosofía de la liberación latinoamericana (Philosophy of the Latin American Liberation)*, Mexico: Fondo de Cultural Económica, 2nd edn, 1992.(The first book to examine the antecedents, foundations and initial theses of liberation philosophy from an epistemological perspective. Four discursive variants of liberation philosophy are identified.)

Cerutti-Guldberg, H. (1988-9) 'Actual Situation and Perspectives on Latin American Philosophy for Liberation', *The Philosophical Forum* 20 (1-2): 43-62.(Assessment of liberation philosophy in the late 1980s.)

Cerutti-Guldberg, H. (1995) *Filosofías para la liberación: Čiquest;Liberación del filosofar?*, Toluca: Universidad Autónoma del Estado de México.

Dussel, E. (1977) *Introducción a una filosofía de la liberación (Introduction to Liberation Philosophy)*, Mexico: Extemporáneos.(The best introduction available to the analectic version of liberation philosophy.)

Dussel, E. (1985) *La producción teórica de Marx. Un comentario a los Grundrisse*, Mexico: Siglo XXI

Islas (1991), special issue 99 (May-August), Havana.(Special issue on philosophy of liberation.)

Roig, A.A. (1981) *Teoría y crítica del pensamiento latinoamericano*, Mexico: Fondo de Cultura Económica.

Roig, A.A. (1983) 'De la historia de las ideas a la filosofía de la liberación' (From the History of Ideas to the Philosophy of Liberation), *Latinoamérica* 10: 45-72.(An article tracing the history of Latin American philosophy.)

Roig, A.A. (1984) 'Cuatro tomas de posición a esta altura de los tiempos' (Four Philosophical Positions at the Height of our Times), *Nuestra América* 11 (May-August): 55-9.(Four different approaches to Latin American philosophy.)

Samour, H. (1994) 'Introducción a la filosofía de la liberación de Ignacio Ellacuría' (Introduction to the Liberation Philosophy of Ignacio Ellacuría), in *El compromiso político de la filosofía en América Latina (Political Compromise of Latin American Philosophy)*, Bogotá: El Búho, 11-37.(A discussion of Ellacuría's work in the area of liberation philosophy.)

Scannone, J.C. (1981) 'La cuestión del método de una filosofía a latinoamericana' (The Question of the Method of Latin American Philosophy), *Stromata* 1-2 (January-June): 75-192.(Discusses the otherness and exteriority of the marginalized.)

Schutte, O. (1993) *Cultural Identity and Social Liberation in Latin American Thought*, Albany, NY: State University of New York Press.(The best analysis in English of the tradition which spans from Mariátegui to Latin American philosophy and theology of liberation to philosophy of liberation.)

Liberation theology

Also known as theology of liberation, liberation theology is simultaneously a social movement within the Christian Church and a school of thought, both of which react against human suffering due to poverty and various forms of oppression. The essence of liberation theology consists in an interpretation of Christian salvation that retains its transcendent eschatological content and draws out its historical dimensions and their implications for personal life, the social sphere and the public action of the Church. Salvation contains various levels of liberation.

Liberation theology is most commonly associated with Latin America, where it emerged during the 1960s. As both movement and theology, it is at present a worldwide phenomenon, taking on different characteristics according to culture, situation, the kind of oppression that predominates, and concrete political and social exigencies. Although some liberation theologians have employed Marxist language as a tool for social analysis, the underpinnings of liberation theology lie in Christian faith.

Liberation theology is predominately Roman Catholic in Latin America because of the Catholic majority; but as a movement and a school of thought it unites Catholic and mainstream Protestant Churches. Evangelical Christians are often antipathetic to liberation theology because of their individualism and other-worldliness.

1 Historical development

The conditions against which liberation theology is a reaction have their prehistory in colonialism and neo-colonialism. Liberation theologians and historians chart the history of colonialism and analyse the social, economic and political dependencies of developing countries upon developed nations. In varying degrees, the relation of Latin American nations to the developed nations is mirrored in each developing country by a two-class system of educated, resourceful people and the poor or destitute majority.

The most immediate explanation of liberation theology, however, lies in the actual conditions in which poor people live and the natural human reaction to them. Christian theologians, among others, came to recognize the massive poverty of the majority of people in Latin America as radically unjust and systemically supported by social institutions. As various organizations of students and workers mobilized for social justice, the moral perception of the deadly character of destitution became a condition for appreciating the liberating potential of theology. Theologians began to interpret Christian faith as responsive to the cause of justice and the poor. By the early 1960s, essays in what would come to be called liberation theology were already being written.

Two public events in the life of the Roman Catholic Church in the 1960s served as a catalyst for liberation theology. One was the Second Vatican Council (1962-5). Its *Pastoral Constitution on the Church in the Modern World* (*Gaudium et Spes*) described the Church as deeply concerned with human life in history and with the problems of poor peoples. It encouraged regional churches to internalize this message and adapt it to each local situation. The other was the Second General Conference of Latin American Bishops in 1968, in Medellín, Colombia, which produced documents describing the social situation of the poor majority, characterizing it as institutionalized violence and sin. The bishops promised that the Church would address these inhuman conditions. The corporate commitment of the hierarchy of the Latin American Church encouraged the movement of liberation theology on all levels.

The 1970s witnessed an outpouring of work in liberation theology, Gustavo Gutiérrez's *A Theology of Liberation* (1971) being the 'classic text'. The themes that dominated the discussion during this decade were the distinctive method of this theology, by contrast with theology in Europe, and the fundamental nature and role of the Church in society. This period ended with the Third General Conference of Latin American Bishops at Puebla, Mexico, in early 1979. Marked by heavy internal Church politics and closely watched by people outside Latin America, its documents roundly supported liberation theology while balancing it with more traditional themes.

The next period of development, through the 1980s to 1992, included developments in theology, Christian life, and ethics, as well as an effort by the Vatican to control liberation theology. Two documents were issued in 1984 and 1986, the first condemning specific extremes perceived within liberation theology, the other developing the Vatican's version of the anthropological grounds of social liberation and emancipation. Leonardo Boff, a Brazilian liberation theologian, was forbidden to lecture for one year in 1985. By the time of the fourth meeting of Latin

American bishops, in Santo Domingo in 1992, which also commemorated the five-hundredth anniversary of the coming of Christianity to the Americas, the Vatican had quieted episcopal support for liberation theology.

The development of liberation theology extended well beyond Latin America. After Medellín, other parts of the developing world immediately resonated with a theological method and language which addressed their own problems of poverty and social marginalization. But in each case, liberationist themes were received into a different situation with a distinctive modality. In south Asia, liberation theology combined with the need for interreligious dialogue and cooperation; in Africa, liberation theology merged with movements of inculturation. Liberation theology is recognizably pluralistic; common formal themes take different concrete shapes around the world.

The Vatican is not hostile to the essence of liberation theology; Vatican teaching is strongly influenced by it. At the Synod of Bishops meeting in Rome (1971), in Paul VI’s influential encyclical *On Evangelization in the Modern World* (1976), in the Vatican’s *Instruction on Christian Freedom and Liberation* (1986), in John Paul II’s encyclical *On Social Concern* (1988), the leadership of the Roman Catholic Church has endorsed the centrality of the commitment to the welfare of the poor and marginalized which is the core of liberation theology. The agenda and language of liberation theology has thus become part of Catholic theology and social teaching from the centre, so that the internalization of liberation theology has spread not only laterally but also back through the centre to the whole Church.

Liberation theology extends beyond the confines of Roman Catholicism and Latin America, and cuts across the confessional lines of the mainstream Churches; it is ecumenical. Also, the social captivity and oppression to which it responds transcend sheer poverty. The negative reality underlying the Black liberation theology in South Africa and the United States is racism. Black liberation theology in the USA developed autonomously out of the remnants of slavery in US society, the civil rights movement, and the language and spirituality of the Black Churches. This liberation theology became articulate in the 1960s through the impetus of Martin Luther King, Jr, and, more formally, the writings of James Cone, its leading theologian. Pervasive androcentrism and patriarchalism are addressed by feminist liberation theologians such as Rosemary Radford Ruether, and this theology too has its own prehistory (see Feminist theology).

Although the term ‘liberation theology’ refers to current movements and modes of thought, these have a tradition that extends back through the whole history of the Christian Church and finds deep roots in the Jewish and Christian Scriptures. For example, the language of liberation theology is closely analogous to the discourse of the prophets and Jesus’ own concern for those marginalized in his society. The term ‘liberation’ is thus a present-day symbol for something essential to the Christian message. A form of liberation theology appears wherever there is systematic destruction of human freedom.

2 Liberation theology as a movement

Liberation theology is generated by a social movement broader than the Church but also reflected in the Church. In Latin America, this was a people’s movement based on an insight into the sheer negativity of structural poverty and injustice and a committed effort to change society. By definition, the movement that liberation theology represents is located on the left of the political spectrum. Its constituencies are rural peasants, urban squatters, the unemployed, and exploited workers. Its allies are often students, middle-class and educated people open to change, and various organizations dedicated to social amelioration. What is true in Latin America finds its analogies in other places.

Various mechanisms structure and sustain the movement. The principle-turned-slogan that supports the liberationist agenda from within the Church at large is ‘the preferential option for the poor’. This principle developed in the wake of Medellín and was formally accepted at Puebla in 1979 and endorsed by Vatican documents and papal statements. The principle is not meant to be divisive but to affirm that the whole Church in all of its members should reflect the love of God that reaches out especially to those who suffer most and are marginalized. In fact, however, the principle divides because it reflects actual social divisions. In most places where liberation theology has taken hold, clear ideological differences distinguish the Churches of the wealthy and those of the poor. The parish whose constituency is the poor is the main locus of the liberation movement. This is where poor people come together, where Christian ministry is mobilized, where various organizations address

specific needs, and where Christian liberation language is heard in worship and public devotion.

In some respects even more fundamental are ‘basic ecclesial communities’. These are small groups of Christians, usually homogeneous so that everyone knows each other, which meet regularly for prayer and reflection. They developed simultaneously with the liberationist movement in Brazil in the 1960s and now exist worldwide. Many combine faith and the practical concerns of survival in daily life. These communities are subgroups within a parish and, when faced with clear social objectives, are animated with a liberationist spirit.

The liberationist movement is sustained by organizations that draw people together from across the lines of basic ecclesial communities, parishes and dioceses. Regional and national agencies address issues of justice. Workshops, conferences, institutes and summer schools focus reflection on liberation theology and animate the growth and social coherence of the movement. Different organizations address specific objectives.

Finally, at the core of the liberationist movement is a distinctive piety and manner of living the Christian life. This spirituality includes responsible and committed service to the poor. In its ideal form it thrives on the idea of the human solidarity of all. It acknowledges God present within the poor, animating them to take control of their lives in the face of oppression.

3 Characteristics and significance

The theology of liberation as a form of thought and a body of literature grew up out of the movement. Some representative theological leaders are: the late Juan Luis Segundo, a Jesuit from Uruguay who studied in Paris and began writing in the early 1960s; Gustavo Gutiérrez, a diocesan priest from Lima, Peru, who studied in Louvain, Lyons and Rome, and began writing in the late 1960s; José Míguez Bonino, a Methodist theologian from Argentina who after studying in the USA undertook seminary teaching in his home country; Leonardo Boff, a Brazilian and former Franciscan who completed his doctorate in Germany in the early 1970s and began teaching in a seminary in Petropolis; and Jon Sobrino, a Spanish-born Jesuit teaching in San Salvador who also studied in Germany and began writing in the mid-1970s. This first generation of liberation theologians, still active in the 1990s, is joined by a new generation of theologians who grew up as liberationists.

Five typical characteristics of liberation theology seen in connection with some of the major topics of Christian theology define its distinctiveness; these are described below.

First, broadly conceived, liberation theology’s method is hermeneutical: it interprets the past symbols of Christian scripture and history on the basis of the questions raised by the world of the suffering poor. The general goal is to make sense of the Christian message for the innocent victim of human oppression. Theological questions thus arise out of the actual situation of the people for whom liberation theology offers a voice. The option for the poor is a premise of the theology and links it with the movement of liberation praxis. Social sciences are employed to gain a critical appreciation of the situation. Some Marxist analysis is used by some liberation theologians as a tool to understand society and history, but without the philosophical premises of Marxist doctrine. The goal of the theology is a historically critical understanding of the Christian message that will engender further emancipatory praxis.

Second, liberation theology is historically conscious in dealing with Jesus Christ. Its Christological reflection begins with the human life of Jesus and not with doctrines about him. Jesus’ historical mission, particularly his prophetic judgments, displayed concern for people who either suffered physically or were excluded by social censure. Liberation theologians find humanization, compassion, healing, and empowerment of freedom dramatized by Jesus. Understanding the doctrines about Jesus begins with his ministry. That Jesus was raised from death, that Jesus is divine - these classic doctrines mean that at its deepest level salvation is liberation by God from sin and final death. But they also give divine sanction to Jesus’ historical ministry as representing the will and values of God for human life. Jesus embodies God’s salvation in the end and God’s salvific will and power within history.

Third, liberation theology proposes a doctrine of the Church that engages society. Following the lead given by Vatican II’s decree on the Church in the modern world, liberation theology regards the Church as an agent of God’s intentions for human life in history. It continues the mission represented in Jesus’ ministry. The Church is to be a sign of God’s love liberating people. Membership of the Church does not imply passivity, but a dependence on God that generates creative action on behalf of the poor. Sacraments are interpreted as vehicles for nurturing

and empowering the Christian life of service. Publicly, the Church cannot avoid being involved in social issues, as distinct from party politics, and should always be concerned for society's weakest members. The role of the Church in society is to preserve a tradition of humane Christian values as a critical context for public policy.

Fourth, liberation theology promotes a this-worldly religious life. Christian faith itself is closely connected with praxis, with a reflective, committed way of living. Christian life consists in an imitation of Christ according to the demands of society. It consists in being a member of a mission Church, which entails some active assumption of responsibility for that mission. Liberation spirituality combines contemplation and action, mysticism and politics. It relies on a notion of God's grace that does not stymie but empowers human freedom in history.

Fifth, liberation theology is intrinsically ethical. Liberation theologians have criticized the absolute right of private property, various forms of capitalism, and the national security state. The option for the poor, a bias towards socialism, and the involvement of the Church in political and social issues raise ethical questions. How can a theology that is partisan for one group be ethical? What kind of ethical discernment is needed for a Christian praxis in this or that situation? Are there general norms for an ethics of Christian liberation? These questions were addressed from the very beginning in liberation theology, but in a variety of different ways and usually within the context of broader theological analysis. Gradually a more systematically focused liberation ethics appeared, one which examines the foundations of Christian liberation ethics and its principles and norms.

Liberation theology consists in different theologies that share a number of basic principles and commitments. These principles are foundational. Liberation theology has always considered itself less a body of content, more a fundamental way of going about the discipline of theology. Within the Roman Catholic Church, liberation theology is the first developed theology that is fully historically conscious and reflectively engaged in the process of history. From the beginning, liberation theology has spontaneously bound together theology and ethics, which for various reasons were separated from each other in the modern period. Theology emerges out of praxis and leads back to it. Because of these primordial principles, liberation theology has profoundly influenced Christian theology in the last third of the twentieth century. No Christian theology today can bypass the question of its social situation and the implied beneficiaries of its understandings. The questions and themes raised by liberation theology are now part of the discipline itself.

See also: Religion and morality §§3-4; Religion and political philosophy §3

ROGER HAIGHT

References and further reading

Boff, L. (1985) *Church, Charism and Power: Liberation Theology and the Institutional Church*, New York: Crossroad.(A collection of essays at once critical of Church institutions and optimistic about the power of basic ecclesial communities.)

Cone, J.H. (1969) *Black Theology and Black Power*, New York: Seabury.(A seminal work reflecting the emergence of Black theology out of the civil rights movement in the USA.)

Gutiérrez, G. (1971) *A Theology of Liberation*, Maryknoll, NY: Orbis Books, 1988. (Latin America's classic statement of liberation theology and still the best introduction.)

Haight, R. (1985) *An Alternative Vision: An Interpretation of Liberation Theology*, Mahwah, NJ: Paulist Press.(A synthetic essay interpreting liberation theology in a more universal theological idiom.)

Hennelly, A.T. (ed.) (1989) *Liberation Theology: A Documentary History*, Maryknoll, NY: Orbis Books.(An indispensable collection of texts charting the development of liberation theology in Latin America.)

King, M.L., Jr (1986) *A Testament of Hope: The Essential Writings of Martin Luther King, Jr*, ed. J.M. Washington, San Francisco, CA: Harper & Row.(A compendium of King's most important sermons and speeches, with selections from his books.)

McAuliffe, P. (1993) *Fundamental Ethics: A Liberationist Approach*, Washington, DC: Georgetown University Press.(A critical and constructive analysis of foundations for a liberationist ethics.)

McGovern, A.F. (1989) *Liberation Theology and Its Critics: Toward an Assessment*, Maryknoll, NY: Orbis Books.(A fair analysis of the positions of liberation theologians and an appraisal of various critiques.)

Míguez Bonino, J. (1975) *Doing Theology in a Revolutionary Situation*, Philadelphia, PA: Fortress Press.(An early work that reflects the growing self-awareness of liberation theology in terms of social context, theological method and constructive content.)

Ruether, R.R. (1983) *Sexism and God-Talk: Toward a Feminist Theology*, Boston, MA: Beacon Press.(A

treatment of the method of a feminist theology and an application of it to various theological problems.)

Schubeck, T.L. (1993) *Liberation Ethics: Sources, Models, and Norms*, Minneapolis, MN: Fortress Press.(A synthetic portrayal of the ethical teaching of Latin American liberation theologians.)

Segundo, J.L. (1971-5) *Theology for Artisans of a New Humanity*, Maryknoll, NY: Orbis Books, 5 vols.(A unified series dealing with central Christian doctrines apart from Christology.)

Sobrino, J. (1978) *Christology at the Crossroads*, Maryknoll, NY: Orbis Books.(The most influential liberation Christology of the 1980s.)

Libertarianism

In political philosophy 'libertarianism' is a name given to a range of views which take as their central value liberty or freedom. Although occasionally the term is applied to versions of anti-authoritarian Marxist theory (the 'libertarian left'), more commonly it is associated with a view which champions particularly pure forms of capitalism. Libertarians endorse the free market and unfettered free exchange, and oppose paternalistic or moralistic legislation (for example, laws regulating sexual behaviour or the consumption of alcohol or drugs). Liberty, on such a view, is identified with the absence of interference by the state or by others. The legitimate state exists purely to guard individual rights, protecting people and their property from force, theft and fraud. This is the 'minimal state' or 'night-watchman state' of classical liberalism. The state has no authority to engage in the redistribution of property (except to rectify the effects of theft, and so on) or, in certain versions at least, to pursue policies designed to further the common good. Such activities are viewed by the libertarian as illegitimate interferences with an individual's right to do what they wish with their own person or property.

1 Key features of libertarianism

The central issue for libertarianism is the proper nature and justification of the state. Libertarians view the state with great suspicion and are particularly opposed to its attempts to supervise and control people's lives, or to carry out policies in the name of social justice (see Justice). At the same time they assign to individuals strong rights to non-interference and reject the 'collectivist' idea that people can justly be coerced into carrying out society's goals or purposes (see Rights §3). However, in most versions of the theory it is recognized that some central authority - a minimal state - is required to protect citizens from each other and from the intrusions of other states.

All forms of libertarianism assume a 'protected sphere' of strong personal rights to life and liberty. Opinions differ about individual rights to property. The main libertarian camp suggests that just as we have absolute rights of self-ownership, we can also form equally strong rights over external objects. However, left-wing libertarians argue that gross inequalities of property diminish the liberty of the poor. Those who possess no property will live or die at the mercy of the owners of private property. Left-wing libertarians argue that property should be distributed much less unequally or that rights to property should be qualified to achieve this effect (see Property §3).

In response, mainstream (right-wing) libertarians have suggested that the market, insurance and, as a last resort, charity, provide alternatives to state support for the poor. But of greater weight is the libertarian claim that there are certain detrimental consequences of allowing governments to regulate and redistribute property. First, any such system is very intrusive, restricting people's freedom of choice and requiring regular investigation of individual property holdings. Second, it interrupts the efficient running of the market, distorting incentives and price mechanisms. Third, it is counterproductive: government intervention often fails to achieve its aims and has unpredicted and unwanted effects. These are all reasons to avoid state intervention.

Nevertheless a response is owed to the challenge that the poor have very little liberty. Here a distinction between liberty and ability is crucial. There are many things I am unable to do - run a mile in four minutes, for example - which cannot be said to be limitations of my liberty. For a libertarian an individual's liberty is restricted only if some person has reduced the options available, and has done so in a way which violates that individual's rights. So if a person has few options as an unfortunate side effect of the free and legitimate actions of others, then, while their ability to do certain things is restricted, their liberty is unimpeded. But if that person is forced to act at the point of a knife, then their options have been shaped by the rights-violating actions of another; thus their liberty has been violated. On this view, then, economic disadvantage affects ability, not liberty (see Freedom and liberty).

2 Antecedents of libertarianism

Libertarianism, while a relatively recent development, was formed out of several currents. The first is the issue of 'the individual versus the state': libertarians of all ilks clearly wish to give the state only a very narrow authority. The second is the question of what should properly count as the individual's own concern: libertarianism strongly resists the idea that the state should have a role in the regulation of possession and exchange or in the manner in which an individual chooses to live their own life. A third impetus to libertarian thinking, underlining the other two, is the classical liberal tradition, emphasizing the economic advantages of a market free from government

intervention.

In tracing the development of libertarianism, it is important to see the first of these issues - the individual versus the state - as historically prior. Many of the thinkers cited as sources for libertarianism were prepared to accept great restrictions on property holdings. It is the antipathy to a powerful, centralized state that gave libertarianism its initial spur; its transformation into a defence of absolute property rights and extensive market society has been a relatively recent development.

Given these somewhat confusing origins, it is not surprising that various forebears have been claimed for libertarianism. Clearly there are affinities between Locke’s defence of individual property rights and the doctrine of the inviolability of property that forms the heart of libertarianism (see Locke, J. §10). Equally, the classical liberal notion of the *laissez-faire* state, derived from Adam Smith (§4), has had great impact on libertarian thinking. On similar grounds, even David Hume (§5) has, with some justice, been cited as an influence on libertarian thought.

We can also see an overtly anarchist strand in libertarianism, springing from Proudhon and Max Stirner, and reaching twentieth-century individualist thought via the US libertarian Benjamin Tucker (see Anarchism §2). Thus, part of the foundation of libertarianism is an emancipatory politics, formed in radical opposition to Marxism and opposed in particular to Marx’s advocation of a highly structured, authoritarian period of politics prior to the withering of the state (see Marx, K. §12). The anarchist observation that such means would inevitably lead away from the intended end inspired reflection on how a broadly egalitarian society could be organized on truly voluntary grounds. It is something of an irony, then, that such ideas have given rise to a tradition which sometimes, and with good reason, is called anarcho-capitalism.

3 Philosophical foundations

Although it is possible to state a ’core’ libertarian view, many different philosophical foundations for it have been offered. Consequently the details of the views defended have significant differences too: for example, opinions differ about whether libertarians can allow people to opt out of market society into, say, socialism if that is what they prefer. The following, used alone or in combination, are perhaps the most important lines of defence of the doctrine.

The best-known approach, that associated with Robert Nozick (§2), is to try to justify libertarianism in terms of a doctrine of natural rights to self-ownership: as I have the right to determine the course of my own life (provided I respect the similar rights of others), I have no enforceable obligations to others unless I have brought these on myself by my own voluntary action. If a government makes me contribute to the alleviation of the plight of the needy, or forces me to act against my will, then it behaves as if it is part-owner of me, and so violates my right to self-ownership. An important variant of rights-based libertarianism justifies rights not in terms of self-ownership, but in terms of human flourishing. Here the central claim is that only in a libertarian society can a life fully worthy of human beings be achieved.

A second style of justification has greater affinities with consequentialist reasoning. Libertarian society is, at least in part, justified by the economic efficiency of market society. Interventions in the free market are criticized on the grounds that they will impede enterprise, distort incentives and thus diminish the effectiveness of the market. This type of approach does, however, allow some room for guarded government intervention to ensure the smooth running of the market and to prevent the formation of monopolies. The state also has the role of supplying ‘public goods’, in the technical sense, such as street lighting or a pollution-free atmosphere, which the market is notoriously poor at producing. This type of consequentialist libertarianism thus has a reason for permitting certain government activities which would be prohibited by pure rights-based libertarianism.

A third approach tries to justify libertarianism by means of a hypothetical contract. First we posit a hypothetical ’state of nature’ in which individuals, considered as holders of private property, subsist alone or in family groups. It is recognized, however, that social cooperation would produce benefits for all. Accordingly, individuals come together to draw up terms of association. All have a reason to accept legally enforceable rights and duties of non-interference, and, perhaps, duties of emergency aid. However, the rich and powerful, who are in the stronger bargaining position, have reason to resist any more extensive duties, and so this limits the agreement that can be achieved. Nevertheless, certain further functions of governments could be seen as agreed to in this context. In

particular, governments may tax individuals to supply public goods.

A fourth approach takes as its starting point subjectivism about values. In this way libertarianism is represented as the only morally neutral society. A society which redistributes income appears to subscribe to the value of 'need-satisfaction'. But if the wealthy do not accept this value, then to engage in redistribution is to force people to live according to values they do not hold, and so is a form of tyranny. Those who do value the alleviation of others' distress are at perfect liberty to do whatever they want with their own property. But there should be no compulsion to subscribe to this value or to any other.

A fifth approach, less common, is inspired by a point from Kant's moral philosophy. It begins with Kant's distinction between actions performed *for the sake of duty* and actions performed *in accordance with duty* (see Kant, I. §10). If we believe that only the former have moral worth, then an extensive set of legally enforceable duties removes the scope of true virtue: if I am forced to transfer some money to you by law, then, while it may be an action performed in accordance with duty, it has not been done for the sake of duty, but out of coercion, and hence there is no moral worth in my action. This form of libertarianism, then, recommends a slimmed-down state to make virtue possible. (It should be noted that Kant himself did not draw these political conclusions from his moral philosophy.)

4 Problems and criticisms

Libertarians have found themselves struggling with two major problems. First, how can the state be justified? If we cannot show that everyone consents to the state, its existence seems an illegitimate interference with individual liberty. Furthermore, if people make different levels of contribution to the upkeep of the state - some may be unable to contribute, for example - then the state appears redistributive in its nature. Second, libertarianism presupposes that individuals have strong rights to property. But everything that is now owned is constructed out of materials which, once, no individual owned. How can initial appropriation be justified? While strenuous attempts have been made to solve these problems, no consensus has developed. It is worth noting that consequentialist libertarians find themselves in less difficulty on these issues.

In addition to these problems critics have claimed that right-wing libertarians are not entitled to adopt the language of liberty in defence of their view, given the lack of liberty of the poor. This goes hand in hand with a denial that such libertarians have framed the distinction between liberty and ability correctly. Another influential criticism is that one cannot defend powerful individual rights without also accepting a duty to support the extensive social institutions that not only sustain those rights but also give them their point and value. Hence, it is claimed libertarianism is self-defeating. Suffice to say that libertarians strongly contest these points, while pointing out the disadvantages they see with all other systems.

JONATHAN WOLFF

References and further reading

Barry, N. (1986) *On Classical Liberalism and Libertarianism*, London: Macmillan.(Useful survey of a wide range of libertarian positions and some liberal antecedents.)

Buchanan, J. (1975) *The Limits of Liberty*, Chicago, IL: University of Chicago Press.(An attempt to provide contractual foundations for libertarianism.)

Cohen, G.A. (1979) 'Capitalism, Freedom, and the Proletariat', in A. Ryan (ed.), *The Idea of Freedom*, Oxford: Oxford University Press, 9-25; revised and repr. in D. Miller (ed.) *Liberty*, Oxford: Oxford University Press, 1991.(Criticizes libertarians for abusing the language of liberty.)

Friedman, D. (1973) *The Machinery of Freedom*, New York: Harper & Row.(Presents a consequentialist version of libertarianism.)

Nozick, R. (1974) *Anarchy, State, and Utopia*, Oxford: Blackwell.(Philosophically powerful and entertaining presentation of rights-based libertarianism.)

Steiner, H. (1994) *An Essay on Rights*, Oxford: Blackwell.(A defence of left-wing libertarianism.)

Taylor, C. (1979) 'Atomism', in A. Kontos (ed.) *Powers, Possessions and Freedom*, Toronto, Ont: University of Toronto Press, 39-61; repr. in *Philosophy and the Human Sciences*, Cambridge: Cambridge University Press, 1985.(Argues that the 'atomism' or individualism of libertarianism - or liberalism - cannot be sustained.)

Libertins

The term ‘libertin’ was first used in France in the late sixteenth century as a term of abuse directed against alleged free-thinkers and atheists who were linked with radical Italian philosophers of the previous century. It subsequently came to be associated with a sceptical literary tradition and a group of scholars, philosophers and antiquarians who discreetly ensured the circulation of such doctrines as Epicureanism, Pyrrhonian scepticism, mechanical philosophy, Baconian empiricism and the ‘new’ astronomy. After the disappearance of this group in the middle years of the century, the term came to connote only debauchery and irreverence.

1 ‘Libertins’

The term ‘libertin’ has been used to describe a number of disparate literary, social and philosophical groups in France from the mid-sixteenth to the late seventeenth centuries. It was at one time claimed that libertins were rationalistic atheists who derived their philosophical doctrines from the Averroists of the School of Padua, notably Pietro Pomponazzi and, later, Cesare Cremonini (1550-1630). But this view was subsequently discredited both by those such as Charles Schmitt (1983) who argue that Paduan Aristotelianism was not Averroist in character, and those cultural historians who, like Lucien Febvre (1982), claim that the sixteenth century was an age which found the notion of God indispensable to its systems of thought (see Aristotelianism, Renaissance §5). More recently, libertins have been linked to a tradition of mitigated scepticism which is said to be an essential ingredient of the scientific revolution: but even this thesis, best known through the work of Richard Popkin (1992), has now been challenged.

The word ‘libertin’ emerged in the late sixteenth century as a term of abuse directed at those who were thought to have rejected traditional authority and were indifferent or irreverent in matters of religion. By the 1620s, it was associated with a broad group of sixteenth-century radical thinkers including Pomponazzi, Girolamo Cardano, Giordano Bruno and Tommaso Campanella, many of whom were later to be linked in some way with the authorship of the *Traité des trois imposteurs (Treatise on the three impostors, namely Moses, Jesus and Muhammed)*. In France, the association with scandalous religious opinions was extended to include the imputation of a debauched, hedonistic way of life, taken to be the outward expression of an indifference towards, or denial of, the afterlife; this is the version of libertinage against which the intemperate jesuit François Garasse fulminated in his *Doctrine curieuse des beaux-esprits de ce temps (The doctrine of the over-inquisitive wits of our time)* (1623), and which the mathematician-priest Marin Mersenne impugns in his tract *L’impieté des déistes, athées et libertins de ce temps (The impiety of the deists, atheists and libertins of our time)* (1624). These attacks were stimulated not only by the writings of Giulio Cesare Vanini, a priest who was executed in Toulouse in 1619 for spreading allegedly materialistic doctrines, but also by the behaviour of a free-living group of noblemen in Paris whose acknowledged leader was the protestant poet Théophile de Viau. In his writings, there are hints of Lucretian atomism (see Lucretius), a celebration of sensuality, and, aesthetically, a linking of freedom of thought with freedom of poetic form. Théophile suffered imprisonment and trial in 1623-5; his fate heralded a greater circumspection in libertin circles, which was accentuated further by the stricter regime of the Cardinal de Richelieu (1630-42).

2 The philosopher-libertins

The more philosophical manifestation of libertinage flourished, however, under Richelieu; this is the so-called ‘libertinage érudit’ of the group of scientists, antiquarians, scholars and writers which included Pierre Gassendi, François La Mothe Le Vayer, Gabriel Naudé, Marin Mersenne and Guy Patin. It developed out of the meetings of the unofficial learned academy which met at the house of the Du Puy brothers in Paris. Many of this group shared a passionate commitment to the free exchange of scientific and cultural information at a time when, outside France, the effects of the Roman index of forbidden books and the Inquisition were felt strongly. For this reason, and also because a number of the group were themselves in religious orders, great discretion was shown in their discussion of the new empirical theories of Bacon, Copernican and Galilean astronomy, Epicurean materialism, mechanical philosophy and the ideas of radical Italian thinkers of the previous century (see Bacon, F.; Copernicus, N.; Epicureanism; Galileo, G.). It may well be that certain members of this group were free thinkers. However, it now seems implausible that the group as a whole shared these convictions, even if their discreet behaviour is taken to

indicate that they knew that their discussions were potentially subversive to traditional beliefs.

A number of literary figures whose works have a significant philosophical aspect are associated with this coterie, notably La Mothe Le Vayer. His *Dialogues faits à l'imitation des anciens (Dialogues in imitation of the ancients)* (1630-1) are strongly influenced by Montaigne's brand of Pyrrhonism, and draw inspiration, as does Montaigne, from the writings of the ancient sceptic Sextus Empiricus. Orasius Tubero, the wise sceptical voice in these dialogues, comprehensively discredits religious dogmatism, and even may imply support for atheism: one by one he patiently destroys the traditional arguments for the existence of God, exposes the prejudices in various religious systems, shows these to be inconsistent with each other even on fundamental points (such as whether religious practice is a matter of conscience or outward observance), and deploys Lucretian arguments against the notion of providence. Like Sextus Empiricus and Montaigne, he seems to favour discreet conformism allied with suspension of judgment, which he himself was to practise subsequently when under the patronage of Richelieu and, later still, as tutor to Louis XIV between 1652 and 1660. Most of the other publications of the so-called libertins of this circle appeared in Latin: Naudé's editions of the works of Agostino Nifo and Cardano, Gassendi's life of Epicurus and account of Epicurean philosophy, Mersenne's mathematical works. Between 1648 and 1655, the principal members of the group died, bringing to an end an early but important episode in the development of the *République des lettres*.

After 1661, discussion of the issues which preoccupied these 'libertins' became much more circumspect in the new, increasingly repressive atmosphere of Louis XIV's absolutism; the revocation of the Edict of Nantes of 1685, which took away the religious privileges of Protestants in France, had a yet more drastic effect, as did the persecution of other allegedly heretical religious practices such as Jansenism and Quietism. By the end of the century, the term 'libertin' had become detached from the nexus of connotations which it possessed in 1620; atheism in itself was no longer shocking, provided that discretion was shown in its expression; and it was no longer scandalous to make public one's adherence to the 'new' astronomy. It is therefore not surprising that the term lost much of its colour, and finally slipped, by the middle of the eighteenth century, so that it was a term used only to designate rakes and voluptuaries, whose lifestyle testified to their indifference to religion.

See also: Clandestine literature; Renaissance philosophy

IAN MacLEAN

References and further reading

All these texts are historical and expository, and do not presuppose specialist knowledge.

Adam, A. (1964) *Les libertins au XVIIe siècle (Seventeenth-century libertins)*, Paris: Buchet/Chastel. (Well-selected collection of seventeenth-century texts in French, with helpful historical and literary introductions.)

Febvre, L. (1962) *Le problème de l'incroyance au XVIe siècle: la religion de Rabelais*, Paris: Michel, 1982; trans. B. Gottlieb as *The problem of unbelief in the sixteenth century: the religion of Rabelais*, Cambridge, MA: Harvard University Press.(An entertaining but contentious general study of the 'mentality' of the sixteenth century with respect to the question of atheism by the co-founder of the Annales School of history.)

Garasse, F. (1623) *Doctrine curieuse des beaux-esprits de ce temps (The doctrine of the over-inquisitive wits of our time)*, Farnborough: Gregg, 1971.(Mentioned in §1 above.)

Godard de Donville, L. (1989) *Le libertin des origines à 1665: un produit des apologètes (The libertin from his origins to 1665: a product of apologists)*, Paris: Biblio 17.(A refutation of Pintard (1943), demonstrating that the concept of the 'libertin' emerges from Counter-Reformation theologians and controversialists.)

Kristeller, P.O. (1968) 'The myth of Renaissance atheism and the French tradition of free thought', *Journal of the History of Philosophy* 6: 233-43.(Attack on the view that a strand of atheistic thought passed from sixteenth-century Italian thinkers into seventeenth-century France.)

La Mothe Le Vayer (1630-1) *Dialogues faits à l'imitation des anciens (Dialogues in imitation of the ancients)*, Paris: Fayard, 1988.(Mentioned in §2 above.)

Mersenne, M. (1624) *L'impieté des déistes, athées et libertins de ce temps (The impiety of the deists, atheists and libertins of our time)*, Paris: Billaine.(Mentioned in §1 above.)

Pintard, R. (1943) *Le libertinage érudit dans la première moitié du XVIIe siècle (Learned libertinage in the first half of the seventeenth century)*, Paris: Boivin.(Still the best general study of philosophical libertinage, but to be read in the context of Godard de Donville's 1989 refutation.)

Popkin, R.H. (1992) *The third force in seventeenth-century thought*, Leiden: Brill.(A recent study of the influence

of scepticism on seventeenth-century scientific, religious and philosophical thought.)

Schmitt, C.B. (1983) *Aristotle and the Renaissance*, Cambridge, MA: Harvard University Press.(Includes a reassessment of Paduan Aristotelianism.)

Spink, J.S. (1959) *French free thought from Gassendi to Voltaire*, London: Athlone Press.(Clear and sensible account of its subject.)

Traité des trois imposteurs (Treatise on the three impostors, namely Moses, Jesus and Muhammad); repr. Florence: Einaudi, 1991.(Attributed to Spinoza by Silvia Berti. There is no certain evidence that the work existed in the sixteenth century.)

Lichtenberg, Georg Christoph (1742-99)

Lichtenberg was a German mathematician, physicist and astronomer, a highly successful university teacher in the field of experimental physics and a prolific writer of essays on scientific and cultural issues. He was one of the leading public figures of the German late Enlightenment, much admired for his satirical and witty style. Today he is known chiefly for his posthumously published notebooks. In thousands of entries, ranging from a single word to several pages, Lichtenberg recorded observations and thoughts about himself and the world. Only a fraction of those entries belong to the tradition of aphoristic writing developed after Lichtenberg and under his influence. A good proportion of the entries are of philosophical interest. Through the radical form as well as content of the notebooks Lichtenberg has become a contemporary to generations of his philosophical readers.

1 Life

Lichtenberg was born in Oberramstadt near Darmstadt as the seventeenth and last child of a Lutheran pastor. From 1763 until 1767 he studied mathematics, astronomy and natural history at Georgia-Augusta University in Göttingen. From 1767 until 1770 he was a private tutor to three English students, whom he accompanied on their journey home to England in 1770. That same year he was appointed to an extraordinary professorship in the philosophy faculty at Göttingen, charged with the teaching of mathematics and physics. In 1772-3 he undertook astronomical work to determine the longitudes and latitudes of three North German cities, a task carried out in the service of the English king, who was also the elector of the German principality of Hannover. During 1774-5 he spent another sixteen months in England. In 1775 Lichtenberg was promoted to an ordinary professorship at Göttingen, where he remained, with the exception of minor travels in Northern Germany, until his death from lung and heart ailments that had afflicted him for a long time as a result of a rachitic deformation of the thorax.

2 Works

Lichtenberg was not a professional philosopher but exercised considerable influence on later thinkers both within philosophy proper and within the wider sphere of intellectual life. In order to understand Lichtenberg's philosophical significance, a more detailed presentation of the form and substance of his works is indicated.

During his lifetime Lichtenberg's early notoriety and eventual fame rested on his dual activities as a practitioner and teacher of experimental physics and as a popular writer on scientific and broader cultural matters. His main contribution to physics was the discovery of electrostatic phenomena involving the geometrical arrangement of particles on electrically charged plates (known as Lichtenberg figures), a discovery that underlies the twentieth-century development of xerography. Lichtenberg was the first to install a functioning lightning rod in Germany. He took part in scientific debates in probability theory, the physics and chemistry of gases, and meteorology.

In the field of letters Lichtenberg was a prolific and widely recognized essayist. Many of his contributions were satirical treatments of contemporary fashion and foibles. He engaged in several literary polemics, one of them against the widely influential physiognomics of J.K. Lavater, who sought to decipher character traits from facial features as represented in silhouette drawings. Lichtenberg's most successful publications for the general public were the detailed explications accompanying the German and French publication of several series of copper engravings by the English painter and graphic artist, William Hogarth, which appeared between 1794 and 1799.

Lichtenberg would be just another, once highly regarded figure of the German late Enlightenment, were it not for the posthumously published notebooks, which he had kept over some thirty years and which are filled with thousands of entries of many kinds and subject matters, ranging from book excerpts and scientific notes to witty remarks and linguistically and intellectually daring pronouncements on himself and his times. Originally intended as material for future scientific and literary projects, the 'waste books' (*Sudelbücher*), as Lichtenberg ironically called them, provide the reader with a wealth of material that defies systematization and reduction to simple positions. Through their highly original form as well as content, the waste books have exercised considerable influence on the thought and writing of generations of posthumous readers, among them Schopenhauer, Nietzsche, Freud and Wittgenstein.

3 Philosophical outlook

Many of the waste book entries are philosophically interesting as scattered documents of a sustained reflection on the peculiarities, weaknesses and limitations of the human mind revealed in Lichtenberg's own experience. Unlike the writings of ancient and modern moralists, Lichtenberg's observations are not gained or formulated with reference to an overall account of human nature. Rather they should be seen in the contemporary context of the emergence of empirical anthropology and psychology and autobiographical writing with their cultivation of individuality in character and manner of thinking. Lichtenberg's highly experimental, witty as well as daring use of the German language can be regarded as a device designed to approach more successfully the ineffable character of human experience.

A considerable subset of the waste book entries contains reflections on philosophical problems and positions. Yet it would not do justice to the fragmentary nature of this material to reduce it to a body of straightforward views or doctrines. Lichtenberg's philosophical reflections do not add up to a comprehensive or even consistent picture of the world. Compared to the systematic work of academic philosophers of his time, Lichtenberg's thoughts appear eclectic and syncretistic. What is specifically philosophical about Lichtenberg's thinking is not so much its propositional content as its modality: the sceptical attitude, the avowal of a subjective perspective and the tentative character of his reflections, typically expressed by such devices as first-person voice, subjunctive mood and laconic formulation.

Several areas of philosophical inquiry attracted Lichtenberg's thinking, chiefly among those methodological issues in the scientific study of nature and the relation between mind and body. The heterogeneous philosophical remarks are linked by a theoretical as well as practical concern with the nature and limits of human knowledge. Main philosophical influences on Lichtenberg were the German school philosophy, which was part of his university training; British associationist psychology and psychological materialism (David Hartley, Joseph Priestley), with which he became acquainted during his second visit to England; Spinoza's pantheistic metaphysics, which underwent a considerable revival in late eighteenth-century German thought (see Spinoza, B. de); and Kant's critical philosophy, whose principled limitation of human knowledge resonated with Lichtenberg's growing agnosticism about the ultimate nature of soul, world and God in the final decade of his life (see Kant, I.).

In his theory and practice of the sciences Lichtenberg worked under the assumption of the overall unity of nature. He regarded the separation of fields of inquiry and sets of phenomena as a reflection of the limits of human knowledge. He often suggested possible relations between distinct areas of inquiry and their respective domains of objects and cultivated thinking in 'paradigms', whereby the theory of one set of phenomena is used to aid thinking about a different, apparently unrelated set of phenomena. He stressed the heuristic and hypothetical character of the paradigmatic method. For Lichtenberg a systematically unified science of nature was not so much a future possibility as an idea or device for conceptual orientation in the manifold pursuit of knowledge of nature. Especially noteworthy is his sceptical attitude toward the applicability of pure mathematics to empirical reality. Lichtenberg favoured the experimental method as a way to control the unnoticed influence of subjective factors in the observation of nature. He sided with Bacon's contemporary followers (for example, A.V. Haller) and against Newton in supporting the use of hypotheses in the study of nature.

During the 1790s Lichtenberg's thinking was considerably influenced by Kant's theoretical philosophy. Lichtenberg's understanding of Kant focused on the role of human subjectivity in the constitution of knowledge and its objects. He took Kant's restriction of human knowledge to appearances as support of theoretical egoism or the view that all we can really know are our own feelings and thoughts. Lichtenberg did not deny the existence of an independent reality but insisted on its unknowability. In line with his Spinozistic leanings he favoured the thought that the apparent distinction of mind and matter might have an unknowable substratum that is neither of the two.

In further expanding upon Kant's transcendental idealism, Lichtenberg employed different Latin propositions to articulate the distinction between the existential independence of things from us (*praeter nos*) and their spatial externality with respect to us (*extra nos*), arguing that naïve realism consists in the illicit shift from the former to the latter. He inferred that something need not be spatially outside of us in order to be independent from us, thus allowing for the possibility that something might be in us which is yet independent from us. Lichtenberg encapsulated this line of thought in his most famous and influential philosophical aphorism, in which he proposes

to replace the Kantian ‘I think’ with the phrase, ‘it thinks’, to be construed along the lines of the locution ‘it lightens’, thus indicating that the I is not the producer but merely the observer of its own thoughts - an un-Kantian move that was not lost on critics of the autonomy and identity of the thinking subject in nineteenth- and twentieth-century philosophy.

GÜNTER ZÖLLER

List of works

Lichtenberg, L.C. and Kries, F. (eds) (1800-6) *Vermischte Schriften (Various Writings)*, Göttingen: Dieterich, 9 vols.(Vols 1-2 contain a first selection from the waste books; vols 3-5 contain the essays; vols 6-9 contain the physical and mathematical writings, a good part of which have never been reprinted.)

Lichtenberg, L.C. and Kries, F. (1844-53) *Vermischte Schriften. Neue vermehrte, von den Söhnen veranstaltete Ausgabe*, Göttingen: Dieterich, 14 vols.(A substantially revised edition of *Various Writings*, edited by Lichtenberg’s sons, with larger selections from the waste books and the essays in vols 1-2 and 3-6, respectively, and a first selection of Lichtenberg’s letters in vols 7-8 and the Hogarth explications in vols 9-14.)

Lichtenberg, L.C. and Kries, F. (1902-8) *Aphorismen. Nach den Handschriften herausgegeben von A. Leitzmann*, Berlin, 5 vols; repr. Nendeln and Liechtenstein: Kraus Reprint, 1968, 3 vols.(The only critical edition of the waste books, omits the entries from physics.)

Lichtenberg, L.C. and Kries, F. (1968-92) *Schriften und Briefe (Writings and Letters)*, ed. W. Promies, Munich: Hanser, 4 vols with 2 vols commentary.(Vols 1-2 contain the waste books, including the entries from physics, vol. 3 contains most of the works published during Lichtenberg’s lifetime, vol. 4 offers a substantial selection of the letters; detailed commentary volume to the waste books.)

Lichtenberg, L.C. and Kries, F. (1983-92) *Briefwechsel (Correspondence)*, ed. U. Joost and A. Schöne, Munich: Hanser, 4 vols.(Most comprehensive edition of letters from and to Lichtenberg; vol. 5 with supplements and index in preparation.)

Lichtenberg, L.C. and Kries, F. (1990) *Aphorisms*, trans. R.J. Hollingdale, London and New York: Penguin.(A good selection from the waste books in an excellent translation.)

References and further readings

Baasner, R. (1992) *Georg Christoph Lichtenberg*, Darmstadt: Wissenschafliche Buchgesellschaft.(A systematic survey of work on Lichtenberg; includes bibliography.)

Gockel, H. (1973) *Individualisiertes Sprechen. Lichtenbergs Bemerkungen im Zusammenhang von Erkenntnistheorie und Sprachkritik*, Berlin and New York: de Gruyter.(Relates Lichtenberg to the rhetorical tradition and earlier as well as contemporary theories of knowledge and language.)

Jung, R. (1972) *Lichtenberg-Bibliographie*, Heidelberg: Stiehm.(Bibliography of works by and on Lichtenberg published up to 1971; 15 entries on Lichtenberg and philosophy.)

Kauther, R. (1992) ‘Lichtenberg und Kant’, *Lichtenberg-Jahrbuch* 5: 56-77.(Stresses the difference between Lichtenberg’s pragmatic anthropology and Kant’s theory of reason.)

Koehne, R. (1963) ‘Gedanken und Exzerpte zur Bestimmung der philosophiegeschichtlichen Stellung Lichtenbergs’, in M. Horkheimer (ed.) *Zeugnisse. Theodor W. Adorno zum sechszigsten Geburtstag*, Frankfurt: Europäische Verlagsanstalt, 133-51.(Attempts to establish the systematic, proto-Hegelian nature of Lichtenberg’s thinking.)

Mautner, F.H. (1968) *Lichtenberg. Geschichte seines Geistes*, Berlin: de Gruyter. (The standard comprehensive intellectual biography.)

Patzig, G. (1992) ‘Über den Philosophen Lichtenberg’, *Text und Kritik* 114: 23-6.(A succinct treatment of Lichtenberg as an anti-systematic philosopher.)

Promies, W. and Joost, U. (1988-) *Lichtenberg-Jahrbuch (Yearbook of the Lichtenberg Society)*, 1 vol. to date. (Includes bibliography; supplants its precursor, ed. W. Promies, *Photorin. Mitteilungen der Lichtenberg-Gesellschaft*, 12 vols, 1979-87.)

Schöne, A. (1982) *Aufklärung aus dem Geist der Experimentalphysik. Lichtenbergische Konjunktive*, 2nd edn, 1983.(An examination of Lichtenberg’s usage of the subjunctive mood as an expression of his experimental spirit in physics and literature.)

Stern, J.P. (1959) *Lichtenberg. A Doctrine of Scattered Occasions*, Bloomington, IN: Indiana University Press. (The most detailed general treatment of Lichtenberg in English.)

Von Wright, G.H. (1942) ‘Georg Christoph Lichtenberg als Philosoph’, *Theoria* 8: 201-17.(A detailed

presentation of Lichtenberg as a philosopher; claims him as a precursor of logical empiricism.)
Zöller, G. (1992) 'Lichtenberg and Kant on the Subject of Thinking', *Journal of the History of Philosophy* 30: 417-41.(Places Lichtenberg's aphorism on the 'it thinks' in the context of his selective appropriation of Kant's idealist theory of self and world.)

Life and death

Problems concerning life and death are among the most dramatic and intractable in philosophy and they feature in all fundamental areas of philosophical inquiry, especially ethics. Most basic is the problem of what account to give of the value of life itself. This problem has had two main dimensions. One has been the controversy over what precise account to give of death; this has revolved around the issue of whether death is, as it is commonly perceived, an evil, and premature death a tragedy. The other has been the equally puzzling question of how to explain the positive value of life, and to resolve the problem that the more rich we make our account of the value of life, the more the value of life, and hence the nature of the wrong done by killing someone, seems to vary with the quality of the life of the person concerned.

A second set of problems concerns the definition of death and appropriate criteria for death. Death, as the most extreme consequence of violence, also leads one into psychological discussions of aggression and into issues of political violence, terrorism, war and capital punishment in political philosophy. Third, there has been concern with a number of practical moral issues, including abortion and euthanasia. Finally, issues have arisen concerning the relation of the value of the life of persons to other sorts of lives, those of animals, for example, or the life and survival of the ecosystem itself.

This discussion will concentrate on the central themes of the value of life and the harm and wrong represented by death.

1 The value of life

It is not only the evil of death that presupposes that life has value and directs us to account for this value, but also everyday discriminations between lives of different sorts. Even the food of vegetarians involves the premature death of living things, and vegetarians usually accept priorities between different animals and between human individuals at different stages of development. If the hospital is on fire, should we attempt to rescue the patients before the hospital cat, and some patients before others? Should the terminally ill, the very old or those in persistent vegetative state be rescued before or after those with radically different life expectancy and degrees of richness and variety in their lives? Should those responsible for their own poor health (heavy smokers, for example) be preferred to the more prudent? Only an account of the value of life will tell us both why lives should be saved and whether and to what extent it is legitimate to choose between lives.

The chief recent attempts to provide a theory of the value of life have sought to identify those features of the most valuable creatures (humans) which might explain their peculiar value. Most theories combine autonomy, self-consciousness and intelligence as the relevant features (see Autonomy, ethical). Creatures with such capacities have often been termed ‘persons’ (see Persons). Radically different accounts of how to apply such criteria of personhood have emerged. Philosophers of broadly consequentialist orientation have claimed that only creatures who actually possess the relevant characteristics count as persons (see Consequentialism). A major difficulty for such accounts is their counter-intuitive conclusion that creatures which most people do regard as valuable (foetuses and neonates, for example) either are valuable not in virtue of any intrinsic properties that they possess, but only in so far as they are valued by persons properly so called (their parents, perhaps), or will be valuable only in terms of future expected utility.

Others, accepting broadly the same criteria for personhood, have argued that creatures structured to possess such capacities or members of a natural kind that typically possesses such capacities are valuable whether or not particular individuals (foetuses, for instance) actually possess them. Another approach rests content with stipulating that humans are more valuable than others simply in virtue of their species membership. These ‘natural kind’ theorists cannot account for the discriminations people make between the moral importance of, say, foetuses, and other members of the same natural kind. If the life of a mother and her foetus are in danger and both cannot be saved, most would believe it right to prefer the mother.

Philosophers faced with the sorts of problems considered so far often produce *ad hoc*, ‘common-sense’ modifications to their general theories to overcome difficulties with hard cases (see Common-sense ethics). For example, natural kind theorists often admit of grades of natural kind membership, using terms like ‘fully-fledged’

humans to account for differences in attitudes to foetuses and adults.

Ronald Dworkin, in an original account of the value of life, has argued that the sanctity of life must be understood in terms of the waste of investment in life represented by death. Dworkin (1993) distinguishes two dimensions of investment that might be wasted by death: the natural and the human. Natural investment implies that nature itself makes an investment in terms of time, trouble and natural resources when life is created and that investment increases in a linear way as the life continues. In the case of human investment, there is both the investment of the human whose life it is (in terms of self-creation both conscious and unconscious) and that of the other people who invest time, effort and resources in creating and sustaining that life. On this view the wrong of causing premature death is that of squandering this natural and human investment. A conservative view of the wrong of euthanasia and abortion, for example, prioritizes natural investment, while a more liberal view will prioritize a particular interpretation of the human contribution to a life.

This and other accounts suggesting that a life is valuable in proportion to its *richness* give importance to factors which differ across lives. On these views, lives will be more valuable the more the investment in them or the more rich and varied they are. This gives rise to the major problem that no two lives will be equally valuable and huge problems of discrimination between people are inevitable.

2 Criteria for death

Since Epicurus there has been persistent philosophical interest in the problem of accounting for the evil of death (see Epicurus §2). We have already presupposed that this problem can be solved by discussing the value of life. Epicurus' problem is in a sense paradoxical. He made substantially the same point as Wittgenstein: 'death is not an event in life'. If it is not, then there is no one to whom death happens, no one for whom death is an evil. But, of course, most people fear death for the disaster it represents. Epicurus' point turns on the necessity of a harm's being experienced, yet we are all familiar with things we rationally regard as harms whether or not we experience them. I have a rational preference not to fall into a persistent vegetative state even if I will not be aware of this happening and thereafter will never be aware of anything again.

The more interesting and important problems concern the nature of the evil death represents and the importance of defining death, or rather of identifying appropriate criteria for the occurrence of death.

If, as most believe, death is an evil, and premature death a tragedy, and if we can say what makes this so, we will also be answering our first question, about what it is that makes life valuable. Here it is important to distinguish the question, 'What makes life valuable - for you?' from the question, 'In virtue of what is life the sort of thing that can be valuable?' The first question is likely to have as many answers as there are persons to whom it is put, the second question rather fewer.

If we move to the issue of criteria for death we can perhaps see why this is so. Death is as old as life, and people have seldom been at a loss as to when grief is appropriate. In other words, death is not a concept which required elucidation. Traditionally, permanent cessation of breath and/or heartbeat was accepted as a reliable indicator of death. While there may have been some uncertainty as to when cessation could be regarded as permanent, the onset of *rigor mortis* and the decomposition of the body could be relied upon to settle the matter in due course.

Problems arose when technology enabled the heartbeat and breathing of individuals to be maintained almost indefinitely. This was so even when the individuals concerned were otherwise so badly injured as to make it certain both that they would never regain consciousness, and that they would die if mechanical support were withdrawn. But are such individuals *dead*? Why is the question important? Why does it even arise?

Individuals on what is popularly known as 'life support' do not appear dead. They breathe, they are supple and perfused with blood, not cold and stiff like a corpse. In order to justify the cessation of life support (I continue with the popular term because it highlights the paradox we are discussing) with the inevitable consequence that the individual would die, it first had to be clear why it was appropriate to let this individual die, why their life had ceased to be valuable, in the sense of worth saving. Second, the technology of efficient life support narrowly preceded the development of organ transplants. If the individual on life support was to be eligible as an organ donor, their organs had to be in good condition. The condition of the organs was optimized by the maintenance of life support. Finally, pressure created by scarce resources meant that the intensive care beds necessary for life

support were in demand and their occupation by one individual rather than another had to be justified.

The practical way out of the problem was to invent a new set of criteria for the occurrence of death. The idea was not to take individuals off life support and wait for them to die, but to declare them dead while still having their life systems sustained. This could be done if death of the brain were to be accepted as a necessary and sufficient condition for death of the organism as a whole. It is now generally accepted in many different societies and cultures that brain death is the criterion for death of the organism as a whole even though the rest of the organism can be kept ‘alive’ (breathing, blood circulating) after brain death. The major dispute has been over whether whole brain death is necessary for death to be declared to have occurred, or death of the brain stem, the conduit through which all electrical activity in the brain has to pass.

This agreement about brain death is significant, for it surely contains an acknowledgement that it is mental activity, and the things that mental activity supports, that are relevant to the value of life - that when the capacity for consciousness has departed permanently, all that matters has gone. The point of declaring individuals ‘dead’ was to mark the fact that all that matters about an individual had disappeared and that other things of importance could now be permitted. For example, organs could be made available for donation, intensive care beds released for other urgent cases, friends and relatives released from the often considerable burden of care and support, and grief could begin.

The acceptance of brain death marks a change in understanding of what matters about life and at the same time a reassertion of a traditional conception of respect for the sanctity of life and a correlated insistence that only death takes individuals beyond our moral concern. Brain death is such an attractive notion precisely because it permits the preservation of the concept of death as the crucial moral divide and at the same time allows us to think differently of human individuals who yet breathe. However, the artificiality of brain death as a criterion for the death of the entire organism should be borne in mind. There is an important sense in which brain death is at best a new conception of what it is to be dead and at worst an uneasy compromise between facing squarely the issue of what matters about life and harnessing the massive unreflective consensus about the significance of ‘death’.

3 Persistent vegetative state

That this is so can be seen more clearly if we consider the condition of persistent vegetative state, and the landmark judgment by Britain’s House of Lords in the case of Tony Bland. Bland sustained brain damage after being crushed in a crowd of spectators at a football match in 1989, at Hillsborough football stadium in Sheffield, England. The brain damage left him permanently and irrevocably unconscious, in what is now termed a ‘persistent vegetative state’ (PVS). PVS is not fatal; people like Bland can remain alive for thirty or more years. They are not ‘brain dead’. In this they are akin to infants born with anencephaly (absence of a brain) or with their cerebral cortex destroyed.

Bland’s parents, who accepted that their son had ceased to exist in any real - biographical - sense, although his body remained alive, were prevented from obtaining the solace of grief. In desperation, they asked the English courts to declare that it would be lawful for medical staff to withdraw feeding and other life sustaining measures so that their son would die. It is not clear why there was any necessity to take the Bland case to the courts, since it was already well established that there was no obligation to sustain a baby by feeding (*Re C* [1989] 2 All ER 782 and *Re J* [1990] 3 All ER 930).

Eventually the House of Lords ruled unanimously that such a course of action would be lawful (*Airedale NHS Trust* v. *Bland* [1993] 1 All ER 821 H.L.). The problem was, of course, that although Tony Bland had permanently ceased to have ‘a life’ in any meaningful sense, he was not dead and would not die unless the courts permitted doctors to take steps to that end.

A slightly later case was concluded in the Court of Appeal in January 1994. The Master of the Rolls, Sir Thomas Bingham, held, in a bizarre judgment with which the other two lord justices of appeal concurred, that it was permissible for doctors to end the life of a patient by refusing life-prolonging treatment when the consultant and ‘a number of other doctors’ agreed that such a course was in the patient’s best interests and ‘no medical opinion contradicted it’ (*Frenchay Healthcare NHS Trust* v. *S*, Court of Appeal, Judgment, 14 January 1994). Tony Bland’s condition resembled those with brain death in that he had irrevocably lost the capacity for consciousness. The difference is that those in PVS still have electrical activity in the brain and through the brain stem. Does this

difference amount to a morally relevant difference between those in PVS and those who are brain dead? Although the House of Lords was reluctant to change the definition of death, or even address that issue, it is clear from its decision that it thought Bland's life, because he had lost all capacity for consciousness, did not retain the sort of value that required it to be sustained. In the words of Lord Keith of Kinkel in his judgment in that case, 'It is, however, perhaps permissible to say that to an individual with no cognitive capacity whatever, and no prospect of ever recovering any such capacity in this world, it must be a matter of complete indifference whether he lives or dies'.

There was no question in Bland's case of competing claims on the resources required to sustain him, so that the decision to permit a course of action designed to achieve the death his parents sought was a deliberate, conscious decision to end his life. A hotly debated question is whether such a decision constitutes a form of euthanasia. Although the House of Lords strongly denied that this is what it was doing, its decision in the Bland case was thought by many to legalize, for the first time in the United Kingdom, a form (albeit very restricted) of euthanasia. That the case of Tony Bland establishes a precedent for legally sanctioned euthanasia in the United Kingdom is confirmed by the words of Lord Mustill in his judgment in that case:

> The conclusion…depends crucially on a distinction drawn by the criminal Law between acts and omissions, and carries with it inescapably a distinction between, on the one hand what is often called "mercy killing", where active steps are taken in a medical context to terminate the life of a suffering patient, and a situation such as the present where the proposed conduct has the aim for equally humane reasons of terminating the life of Anthony Bland by withholding from him the basic necessities of life. The acute unease which I feel about adopting this way through the legal and ethical maze is I believe due in an important part to the sensation that however much the terminologies may differ the ethical status of the two courses of action is for all relevant purposes indistinguishable.

The key features of Lord Mustill's judgment are, first, the acknowledgement that the course of action requested of, and approved by, the courts 'has the aim…of terminating the life of Anthony Bland'; and, second, that the supposed difference between acts and omissions relied on by the common law tradition to make moral and legal distinctions, characterizes two courses of action that are ethically 'for all relevant purposes indistinguishable'. This decision made the United Kingdom the second country in Europe to have judicially recognized the necessity of bringing to an end the lives of at least some innocent individuals who have not requested death. The Netherlands legalized euthanasia under certain conditions in a High Court case decided in 1984 and later formally enshrined euthanasia in its legal system.

It is important to emphasize the proviso 'who have not requested death', for other instances of courts defending the right to die have turned on precisely this issue. The landmark United States case concerning PVS, that of Nancy Cruzan, depended crucially on whether Cruzan had expressed a wish to die prior to falling into PVS, and indeed it is often described as a case establishing the right to die (*Cruzan* v. *Director, Missouri Department of Health* [1990] 497 US 261).

4 The ethics of euthanasia

Arguments about the ethics of euthanasia are essentially the same as, and have been coloured by, arguments about the ethics of suicide. The wrong of suicide has, since the death of Socrates, often been seen in terms of either a violation of some idea of the sanctity of life, or the wrong of depriving a sovereign or a god of the use of a body which was theirs to dispose of. Euthanasia, as essentially assisting suicide, while of contemporary relevance as we have seen, has reawakened the centuries old debate about suicide (see Suicide, ethics of).

Those who defend the legitimacy of euthanasia have three main approaches to defending the ethics of what they propose. First, some see euthanasia (like suicide) as a dimension of human freedom and argue that the value represented by respect for autonomy is incomplete unless it encompasses the limiting case of suicide or assisted suicide. On this view, no further justification is required. The second view is based on compassion and tends to undermine this purist approach to the ethics of euthanasia. It argues that suicide and euthanasia are legitimate ways of bringing to an end suffering which cannot be adequately controlled or ended in any other way. This approach can undermine autonomy because it lapses if there is an equally effective way of controlling the pain and suffering. The third type of defence of euthanasia is exemplified by Ronald Dworkin's account (see §1), which argues that

respect for the intrinsic value of life, properly understood, sees life essentially as meaningful and valuable because of the shape given to it by the individual whose life it is, and that this shaping power must include control over life's end. Unlike the first defence of euthanasia which appeals simply to autonomy, Dworkin's approach places autonomy at the service of, and hence subordinate to, a conception of the intrinsic value of life. On this view it is not all autonomous decisions to end life that are justified, but only those which conduce to the agents own conception of what it is that makes their life make sense.

Arguments against the legitimacy of euthanasia take two forms: they take a stand on principle or they attempt to undermine the cogency of the arguments in favour of euthanasia. The principled approach either harkens back to the idea that an individual's life is not theirs to dispose of, belonging to the sovereign or the deity or both (one via the other), or takes a stand on the sanctity of life. The more pragmatic approach tends to suggest that the freedom to end one's own life is not part of autonomy properly understood, or that there are other compassionate and effective ways of controlling pain, both physical and psychological.

It is difficult to resolve the differences over euthanasia when the issue is one of principle. Perhaps the most obvious reconciling strategy is to seek cases in which those opposed to euthanasia would concede the legitimacy of killing, and ask if related justifications might not hold good in the case of euthanasia. For example, not all opponents of euthanasia are pacifists, and even pacifists might understand extreme exceptions to the rule against killing the innocent.

One test might be how people feel about the following case. A lorry driver is trapped in the blazing cab of his vehicle following an accident. A policeman is on the scene and sees that the driver cannot be extracted before the flames get to him and he is burned alive. The policeman can let him be burned alive or can give him a quick and relatively painless end by shooting him in the head. The driver says, 'Please shoot me; do not let me be burned alive!' Those opposed to euthanasia in all circumstances must give one answer to the policeman's dilemma, those in favour will give the alternative.

There is one final sort of objection to euthanasia. It avoids the policeman's dilemma but has its own problems. Some people are not opposed to euthanasia on principle and would permit it in exceptional cases, like the policeman's dilemma. However, they regard it as constituting a 'slippery slope' which if permitted would lead to unacceptable forms of killing. They therefore object to the *legalization* of euthanasia, but can cope with isolated and exceptional instances, by forgiveness rather than by justification. The question advocates of the slippery slope objection must answer is whether it is reasonable and rational to criminalize behaviour they admit to be both moral and defensible, and whether or not the unacceptable levels of the slope can be guarded against in another way.

5 Contraception and infanticide

Contraception raises issues of life and death analogous to those we have discussed. First, there are methods of contraception which operate to effect early abortion by, for example, preventing implantation of a fertilized egg. Second, those who regard the potentiality argument as giving moral status to the foetus will, if they are consistent, see contraception as one way in which potential human beings have their potential frustrated. There is, of course, another dimension to the ethics of contraception which is not related straightforwardly to issues of life and death. That is where one sexual partner conceals from the other the nature, existence and/or reliability of the methods of contraception used or leads the other partner to believe a method of contraception is being used when it is not (see Truthfulness).

Since the advent of HIV/AIDS, a popular method of contraception has become more significant as a barrier to infection, and this has added to the moral responsibility of using one particular method of contraception. Questions are often raised as to whether someone who uses no method of contraception, or fails to use a condom, is willing to conceive and bear the responsibility of a child or is willing to run the risk of HIV infection.

It is sometimes suggested, particularly by Catholic thinkers, that contraception subverts the purpose of sex, which supposedly was designed by the deity for procreation. This is a curious argument, however, because if sexual relations are wrong except when they could conceivably result in procreation, then sex between infertile people or during pregnancy is wrong. If on the other hand it is practices which weaken the prospect of new people coming into existence which are to be avoided, then it looks as though it is a celibate priesthood, or the existence of nunneries, which are an affront to God's purpose.

Infanticide raises special moral problems only for those who see a morally relevant difference between the foetus and the neonate. If abortion is permissible, infanticide will surely be permissible on the same terms, unless the newborn differs in some relevant way from those foetuses the abortion of which is permissible. Attempts have been made to identify such differences in three main ways: in terms of either some capacities possessed by the newborn and not the foetus, or the newborn's supposedly greater potential for personhood, or the social relations it forms for the first time on consciously encountering other beings. All of these alleged differences are controversial, and we should note that the last, socialization, leaves unprotected any and all unloved, unwanted and unclaimed infants (see Reproduction and ethics).

See also: Bioethics, Jewish; Death; Life, meaning of; Medical ethics; Nursing ethics; Suicide, ethics of

JOHN HARRIS

References and further reading

Beecher, H. (1971) 'The New Definition of Death, Some Opposing Viewpoints', *International Journal of Clinical Pharmacology* 5.(Early source on the emerging debate over criteria for death.)

Committee of the Harvard Medical School (1968) 'A Definition of Irreversible Coma', *Journal of the American Medical Association* 205: 337-40.(Traces the rationale for changes in the definition of death.)

Dworkin, R. (1993) *Life's Dominion*, London: HarperCollins, 213-14. (Sophisticated and sustained discussion of abortion and euthanasia.)

Glover, J. (1977) *Causing Death and Saving Lives*, Harmondsworth: Penguin. (Classic discussion of killing and letting die.)

Harris, J. (1980) *Violence and Responsibility*, London: Routledge & Kegan Paul. (Discussion of the moral difference between acts and omissions and our responsibility for both.)

Harris, J. (1985) *The Value of Life*, London: Routledge & Kegan Paul.(Introduction to medical ethics which discusses the value of life.)

Keown, J. (ed.) (1995) *Euthanasia Examined: Ethical Clinical and Legal Perspectives*, Cambridge: Cambridge University Press.(Authors from various disciplines consider euthanasia.)

Kleinig, J. (1991) *Valuing Life*, Princeton, NJ: Princeton University Press.(General account of life and death issues.)

Lamb, D. (1985) *Death, Brain Death and Ethics*, London: Croom Helm.(Collection of essays on euthanasia.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press.(Interesting in this context for its discussion of the possibility of a human being's instantiating a number of different persons and hence 'containing' more than one life.)

Rachels, J. (1987) *The End of Life*, Oxford: Oxford University Press.(A defence of voluntary euthanasia.)

Steinbock, B. (ed.) (1980) *Killing and Letting Die*, Englewood Cliffs, NJ: Prentice Hall.(Excellent collection of essays on life and death.)

Tooley, M. (1983) *Abortion and Infanticide*, Oxford: Oxford University Press.(A controversial argument to the effect that abortion and infanticide are morally indistinguishable.)

Life, meaning of

This is an obscure yet central topic in philosophy. Often associated with the question whether human beings are part of a larger or divine purpose, the question, 'What is the meaning of life?' seems to invite a religious answer. Much philosophical discussion, however, questions the necessity of this association. Attention to the inevitability of death has often seemed to make life's meaning problematic, but it is not obvious how immortality could make the difference between meaning and its absence. The theme of absurdity runs through much discussion of those who believe the universe to be indifferent. Though our lives have no significance, they argue, we must live as if they do. In the face of this absurdity, some advocate suicide, others defiance, others irony. One may also turn away from the issue of cosmic significance, and look for meaning elsewhere.

1 The meaning of 'the meaning of life'

The question, 'What is the meaning of life?', probably arouses both more contempt and more respect for philosophy than any other. On the one hand, the question is notoriously vague and has encouraged much pompous nonsense. On the other, the urge to understand the point of our existence is deep and pervasive, and is indicative of qualities of mind that are arguably central to being human.

A major difficulty besetting the topic is lack of clarity about the subject itself. Drawing comparisons with other contexts in which we ask for meanings tends to increase confusion. When we ask for the meanings of words or phrases, we ask what they are typically used to communicate. Life, however, is not an item in a system of communication. It does not seem to be used or intended to represent something beyond itself. In certain circumstances we also talk about the meaning of nonlinguistic items: footprints mean someone else has been here; a rash on the skin means the child has measles. Analogies with these uses of 'meaning', however, have not proved helpful.

Religion, particularly Judaeo-Christianity, provides a natural context for the question of the meaning of life (see Religion and morality). If one believes that a supernatural being created the world with some grand design, the question asks for the purpose of that design or the place of life within it. However, the philosophical topic of the meaning of life - or the set of overlapping topics that have come over time to be associated with the phrase - cannot be restricted to issues that make sense only on religious assumptions.

Central concerns that come under the topic include questions about whether life has a purpose, whether life is worthwhile, and whether people have any reason to live, independently of their specific circumstances and interests. Any of these questions may be asked about life, or, more commonly, about human life, but they can also be asked about individual lives, particularly about one's own life. We can search for purposes, reasons, values that are acceptable from points of view external to ourselves, or we can restrict our attention to the realm of desires and goals found in our psyches or our communities, indifferent to possible perspectives beyond the human. Although the phrase 'the meaning of life' seems to assume only one meaning to life, we may be led to reject this assumption without concluding that life is meaningless. Often, the focus of the question shifts in the very process of trying to answer it.

To inquire into the meaning of life, then, is like engaging in a search in which you are not sure quite what you are looking for until you find it. Any attempt to give an unambiguous paraphrase of 'the meaning of life' is bound, like the phrase itself, to foreclose certain options and cut off paths of inquiry that should not be ruled out in advance.

2 The relevance of death

The sense that there is a problem about the meaning of life is frequently brought on by the contemplation of death (see Death §8). Indeed, it is often thought, as it was by Schopenhauer (1851) and by Tolstoi (1886), that the question arises precisely because our lives will end in death (see Schopenhauer, A. §6). However, as some philosophers have noticed, the connection between our finitude and life's meaning is puzzling. If the assumption that we will all die makes life seem meaningless, how would the opposite assumption - that we will live forever - make the situation any better?

A possible explanation for the connection between the thought of death and the fear that life is meaningless is that

facing up to one's own mortality destroys one's prospects for happiness. If ultimate happiness were likely or even possible, we might not feel the need for meaning - one does not need a reason to live as long as living is fun, and the aim of achieving ultimate happiness, were it attainable, might be purpose enough. For some, however, the knowledge that they will die makes happiness impossible. Somewhat differently, recognition of the inevitability of the death of one's culture and one's species as well as of oneself may make the interests and goals one formerly had appear worthless or silly.

Again, belief in a God can provide relief from these concerns. The promise of an afterlife, in which at least some achieve eternal bliss, renews the possibility of working towards ultimate happiness. Independently, the existence of an eternal and superior being who cares about us and about what we make of our lives relieves the worry that our goals and our conduct are insignificant.

3 Absurdity

If God does not exist, many argue, then human life is absurd. The human condition, these philosophers claim, would then contain a fundamental and unchangeable disharmony. Albert Camus focused on the clash between our demand that the world be reasonable, orderly and caring and the reality of its being silent, blank and indifferent. Thomas Nagel emphasizes the discrepancy between the objective insignificance of our lives and projects and the seriousness and energy we devote to them. How are we to respond?

Because coming to recognize the indifference of the universe can be a shattering experience, the thought of suicide arises naturally. If all your goals are founded on the assumption that your existence or your actions matter to some entity or process larger and less in need of validation than yourself, the discovery that there is no such entity leaves you stripped of all direction. If, moreover, you think that any direction you take will necessarily reintroduce the assumption you now know to be false, then it may appear that the only option that avoids contradiction is suicide. Camus (1955), however, believed that there is a noncontradictory way of life available. He described 'absurd man' as living 'without appeal', in defiance of the world's indifference to him. Such a person embraces life as fully as possible but without ever forgetting or denying the absence of any rational foundation for it.

Nagel offers a milder response (1971): the recognition of our insignificance is a function of our distinctively human ability to adopt an external view upon ourselves; as such there is no reason to try to deny it or escape from it. At the same time, if our lives are cosmically insignificant, so is the matter of how we respond to this fact. In the light of this argument, Nagel suggests, defiance seems too overblown and dramatic, and irony is more appropriate.

Richard Taylor (1970) draws a different moral from the silence of the universe: the recognition that life is, as it were, objectively meaningless, should convince us to turn our search for meaning inward. The kind of meaning in life that it makes sense to care about is meaning to us. Life has meaning if we are able to engage in activities that we find meaningful. Otherwise, not.

These philosophers all share the view that if there is nothing larger and more intrinsically valuable than ourselves to whom we may see ourselves as positively attached, then life is meaningless in at least one important sense. In this they agree with those who rest a positive view of life's meaning on the existence of a benevolent God. Because they also believe that the condition for meaning is not met and that we none the less must live as if life had meaning, they conclude that human life is absurd. As pointed out by Joel Feinberg (1992), however, there is a difference between a situation's being absurd and a person's being so. By taking a proper attitude towards our predicament, whether that be defiance or irony or some third alternative, we can at least save ourselves from being ridiculous.

It is not clear, however, that we are rationally required to make even this relatively unpessimistic concession to the view that human life is absurd. As we have seen, this view rests on the idea that there is an inescapable clash between what we demand or inevitably presume about our place in the universe and the reality of our situation. But the tendency to want or insist on our cosmic importance may be less deep and inevitable than these philosophers think. Taking life by the horns, pursuing one's projects with energy and devotion, need not rest on delusions of grandeur. It is at least not obvious that when an Olympic athlete strains to the limit in an effort to break a world record, or that when a mother forgoes sleep and comfort to nurse her child back to health, she must believe that her achievement will be of cosmic significance.

4 Subjective and objective meaning

Although discussions of the meaning of life are often associated with considerations about our place in the universe, there are also contexts in which the intelligibility of the contrast between meaningful and meaningless lives appears to be wholly independent of the cosmic issue.

We have already mentioned the view that the kind of meaning it makes sense to care about is subjective meaning. Some, like David Wiggins (1976), believe that a wholly subjective account of meaning cannot do justice to the ordinary use of the term. As Wiggins points out, the idea of a distinction between a meaningful life and a meaningless one is not equivalent to the more obvious and uncontroversial difference between a life that is subjectively satisfying or fulfilling and one that is not. When we wonder whether our lives have meaning, we are not engaged in a wholly introspective enterprise, and when we search for a way to give meaning to our lives, we are not looking for a pill that will make us happy. The life of Sisyphus, condemned by the gods perpetually to roll a stone up a hill only to see it roll down again, has been offered, at least since Camus' writings, as a paradigm of meaninglessness. If we imagine Sisyphus as perversely fulfilled by this repetitious and futile activity, it is not clear whether we would evaluate his life as more meaningful or more dreadful.

Accounts of meaning in life need not be restricted to purely subjective and purely objective alternatives, however. The most natural paradigms of meaningful lives are both abundantly fulfilling subjectively and admirable or worthwhile as judged from points of view external to the agents themselves. The kind of life most comfortably described as meaningful appears to be one in which there is a happy connection between a subject's lively interests and the range of things that are worthy of interest. Meaning seems to arise when subjective attraction meshes with objective attractiveness.

Whether and how this kind of meaningfulness relates to the concern that seems most naturally to call for a connection to some divine or cosmic purpose are difficult issues. Moreover, the notion of 'objective attractiveness' (or objective worth or value), to which this conception of meaningfulness makes reference, is notoriously controversial. Whether such a notion is ultimately intelligible, particularly in the absence of a religious metaphysics, constitutes a major philosophical question on its own (see Moral realism). That the issue of the meaning of life should open onto and connect with other major philosophical issues, however, should come as no surprise. It is, after all, one of the deepest and most basic topics in all of philosophy.

See also: Buddhist concept of emptiness; Existentialism; Existentialist ethics; Good, theories of the; Nihilism

SUSAN WOLF

References and further reading

Baier, K. (1957) 'The Meaning of Life', in E.D. Klemke (ed.) *The Meaning of Life*, New York: Oxford University Press, 1981, 81-117.(Delivered originally as the inaugural lecture at the Australian National University in Canberra, this article defends the view that meaning in life is compatible with a secular worldview.)

Camus, A. (1943) *Le Mythe de Sisyphe*, Paris: Gallimard; expanded edition, 1945; extended edition trans. J. O'Brien, 'The Myth of Sisyphus', in *The Myth of Sisyphus and Other Essays*, New York: Alfred A. Knopf, 1955, 1-102.(Classic discussion of the absurdity of the human condition and the proper response to it - defiance.)

Feinberg, J. (1992) 'Absurd Self-Fulfillment', in *Freedom and Fulfillment*, Princeton, NJ: Princeton University Press.(Especially clear and accessible discussion of the idea of absurdity and of the kind of fulfilment that is possible in spite of it.)

Klemke, E.D. (ed.) (1981) *The Meaning of Life*, New York: Oxford University Press. (Anthology containing religious and secular accounts of the meaning of life, including portions of the Baier, Camus, Nagel, Taylor and Tolstoi pieces listed here.)

Nagel, T. (1971) 'The Absurd', *Journal of Philosophy* 68 (20): 716-27; repr. in E.D. Klemke (ed.) *The Meaning of Life*, New York: Oxford University Press, 1981, 151-61.(Analyses absurdity as the clash between pretension and reality and advocates irony as the proper response.)

Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press, ch. 6.(Wide-ranging exploration of the idea of a meaning to life, analysing meaning as the transcending of limits in a wider context of value.)

Schopenhauer, A. (1851) 'On the Sufferings of the World', trans. T.B. Saunders, in R. Taylor (ed.) *The Will to Live: Selected Writings of Arthur Schopenhauer*, New York: Ungar, 1967.(Presents a deeply pessimistic view

about the misery and meaninglessness of human life, advocating suicide as a proper response. See also 'On the Vanity and Suffering of Life' and 'The Vanity of Existence' in this volume.)

Taylor, R. (1970) *Good and Evil*, New York: Macmillan, ch. 18.(Argues that what makes life meaningful is subjects' attachments to the activities that occupy them.)

Tolstoi, L. (1886) *Smert' Ivana Il'icha*, trans. A. Maude, *The Death of Ivan Il'ich*, New York: New American Library, 1960.(Vivid description of the sense that, in the face of death, one's life is meaningless.)

Tolstoi, L. (1884) *Ispoved*, trans. A. Maude, *A Confession*, in *A Confession, The Gospel in Brief and What I Believe*, London: Oxford University Press, 1971. (Autobiographical account of being struck by the need to understand the meaning of life, and of finding the only acceptable answer in faith in God.)

Wiggins, D. (1976) 'Truth, Invention, and the Meaning of Life', *Proceedings of the British Academy*, 62: 331-78; repr. in G. Sayre-McCord (ed.) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press, 1988, 127-65. (Argues against a purely subjective account of meaning, in favour of a nonsubjective but anthropocentric analysis of value - difficult reading.)

Wolf, S. (1997) 'Happiness and Meaning: Two Aspects of the Good Life', *Social Philosophy & Policy* 14 (1): 207-25.(Expounds the view that meaningfulness in life arises from active and subjectively fulfilling engagement in projects of objective worth.)

Life, origin of

The appearance of maggots on meat or of intestinal tapeworms supported an ancient belief in the spontaneous generation of life. This idea was challenged in the seventeenth century but not abandoned before Pasteur's experiments. Scientists now agree that terrestrial life had a single origin, but differ in explanations. Some believe that life began with the onset of protein-based metabolism, supported by evidence of spontaneous abiotic amino acid synthesis and theoretical models of self-sustaining and evolving systems of enzymes. Others believe life began with the appearance of nucleic acid-based molecular replicators and have organized their research efforts around the vision of a primordial 'RNA world'.

Aristotle's observation that some life forms 'arise not from living animals but from putrescent matter' went largely unquestioned until the beginning of the seventeenth century. Descartes had no difficulty conceiving of spontaneous generation occurring simply on the basis of matter in motion, though subsequent Cartesian mechanists could not countenance the possibility of life arising spontaneously from nature - construed as passive and without inherent purpose. Their deistic alternative, theories of pre-existence and preformation, endeavoured to attribute all life forms to original acts of creativity. Subsequent generations were deemed to be held in miniature in the germs of their parents and new generations could 'evolve' from the old through the unfolding and elaboration of parts already in place. Committed to an extended regress of miniatures nested inside miniatures the doctrine of pre-existent germs became increasingly vulnerable. Where 'Cartesian matter' lacked the wherewithal to become self-organized, eighteenth-century scientists conceived of a 'Newtonian matter' upon which some uncharacterized force could operate (see Matter). Natural self-organizing capacities were attributed to the existence of 'penetrating forces' shaped by an internal mould (see Buffon, G.L.L. Comte de §3) or by the expansion and resistance of opposing 'vegetative' forces (Needham). Set in motion, the new theories of spontaneous generation were carried into the nineteenth century by both the French *philosophes* and the German *Naturphilosophen* (see Naturphilosophie §3). Where the former opposed a deistic Cartesianism that had become conservative with an uncompromised materialism that could do it all, the latter sought a romantic *rapprochement* of nature and culture, organic and inorganic, within the unity of *Geist*.

Nineteenth-century cell theory resulted in debates over whether cells could arise out of acellular protoplasm as postulated by Schwann (allowing for the possibility of spontaneous generation), or whether cells arise only from other cells (*omnis cellula e cellula*) as proclaimed by Virchow. Scientific opinion became grouped around the poles of mechanical materialism, whose adherents were committed to some form of abiogenesis (life from nonlife), and vitalism (including vital materialism) whose adherents could countenance heterogenesis (life from another form of life) but never abiogenesis (see Vitalism §2). The elucidation of the alternation of life cycles of trematodes led to the demise of any remaining belief in the heterogenesis of parasites from degenerating tissue, and Pasteur's experiments undermined belief in the abiogenetic origins of infusoria. By the end of the nineteenth century, German cytologists too had resolved that not only do cells arise by division from other cells, but even nuclei are never created *de novo*, arising only from previous nuclei. What remained of spontaneous generation was only the conviction of dedicated materialists that there *had* to have been an abiogenetic *origin of life*.

Buttressed by advances in astro-chemistry and cosmogony and a nascent biochemistry of organic catalysis, interest in theories of the origins of life re-emerged in the twentieth century with Oparin and Haldane's visions of a pre-biotic primordial soup from which self-sustaining aggregates of colloidal reactants emerged. Subsequent work proceeded along the lines of an opposition that had already begun to take shape during the 1920s. Whereas Oparin (1924, 1938) and Haldane (1929) looked to the onset of metabolism with an interpretation of life as a *process*, Muller (1929) identified the origins of life with the appearance of an entity, an as yet uncharacterized gene, distinguished by its ability to mutate and self-replicate. With metabolism associated with the catalytic activity of protein-based enzymes, and with genetic replicators (after Watson and Crick (1953)) identified with nucleic acids, that is, deoxyribonucleic acid and ribonucleic acid (DNA and RNA), which serve as templates for their own synthesis, debates about the origins of life have also been construed in terms of 'protein-first' versus 'nucleic acid-first' theories.

Replicator-first theories gained credence from studies by Orgel and Eigen that demonstrated RNA synthesis in

vitro in the absence of either an enzyme or a *prior* template, respectively, and further with the discovery by Cech (1986a, 1986b) of RNA-based catalytic activity. However strong evidence for the abiotic synthesis of nucleic acid *precursors*, or of the ability of RNA catalysts to synthesize arbitrary nucleic acid sequences, has been lacking. Metabolism-first theories gained credence from the studies by Muller that showed that amino acids, the precursors of proteins, may well have been present in an abiotic primitive soup, and from progressively more sophisticated mathematical modelling of the possible dynamics of complex autocatalytic systems.

Aspects of *both* the metabolism first and the replicator first theories can be found in a two-stage model proposed by Dyson (1985) which takes its cue from the dual activity of one small molecule. Adenosine triphosphate (ATP) functions universally in living cells as both the principal intermediary in energy metabolism and as one of the four precursors of nucleic acid synthesis. Postulating the emergence of life as protein-based metabolism, ATP (and perhaps other) nucleic acid precursors may thus be produced biosynthetically as metabolic intermediates. The first nucleic acids would then arise by the chance polymerization of accumulated ATPs. Owing to their ability to serve as templates for further nucleic acid synthesis, these polymers then emerged as intracellular parasites akin to viruses. Dyson hypothesizes a progressive symbiosis between these parasitic replicators, which become stable repositories of polymer information, and the ‘host’ metabolic system which continues to provide nucleic acid precursors.

This two-stage model preserves ontological implications of both points of view. Life appears, first of all, as the holistic cooperativity of complex autocatalytic dynamics, yet remains contingently susceptible to ‘selfish’ short circuiting by ‘born again’ parasitic replicators.

See also: Chemistry, philosophical aspects of; Evolution, theory of; Genetics; Molecular biology; Unity of science

LENNY MOSS

References and further reading

Aristotle (*c*.mid 4th century BC) *Generation of Animals (De Generatione Animalium)*, trans. A.L. Peck, Cambridge, MA: Harvard University Press, 1942.(Spontaneous generation referred to in book one.)

Bernal, J.D. (ed.) (1967) *The Origin of Life*, London: Weidenfeld & Nicolson.(Interesting, albeit dated, discussion by the noted crystollographer and historian; includes reprints of the 1920s articles by Oparin and Haldane.)

Cairns-Smith, A.G. (1982) *Genetic Takeover and the Mineral Origins of Life*, Cambridge: Cambridge University Press.(Good summary of the crystal-as-template theory.)

Cech, T.R. (1986a) ’A model for the RNA-catalyzed replication of RNA’, *Proceedings of the National Academy of Sciences* 83, 4,360-3.(Cech received the Nobel prize for discovery of RNA catalysis.)

Cech, T.R. (1986b) ‘RNA as an enzyme’, *Scientific American* 255: 64-75.(Accessible summary of the RNA as enzyme work which gave rise to the ‘RNA world’ model of the origins of life.)

Crick, F.H.C. (1981) *Life Itself*, New York: Simon & Schuster.(Crick’s version of the panspermia hypothesis of life on earth derived from extraterrestrial seeds.)

Dawkins, R. (1976) *The Selfish Gene*, Oxford: Oxford University Press.(An uninhibited explication of the replicator-first ontology.)

De Duve, C. (1991) *Blueprint for a Cell: The Nature and Origin of Life*, Burlington, NC: Neil Patterson.(Contains excellent annotated bibliography.)

Dyson, F. (1985) *Origins of Life*, Cambridge: Cambridge University Press.(Terse and clever, Dyson’s book unites the protein-first and nucleic acid-first accounts into a two-stage model.)

Eigen, M. (1973) ’The Origin of Biological Information’, in J. Mehra (ed.) *The Physicist’s Conception of Nature*, Dordrecht: Reidel.(Detailed treatment of the hypercycle model which attempted to identify the origins of biological information in self-sustaining interactions between proteins and RNA.)

Eigen, M., Gardiner, W., Schuster, P. and Winkler-Oswatitsch, R. (1981) ’The Origin of Genetic Information’, *Scientific American* 244: 88-118.(Accessible discussion of the RNA quasi-species and hypercycle ideas.)

Farley, J. (1974) *The Spontaneous Generation Controversy: From Descartes to Oparin*, Baltimore, MD: Johns Hopkins University Press.(Excellent survey of the antecedents, within modern science, to the contemporary debate about the origins of life.)

Haldane, J.B.S. (1929) ’The Origin of Life’, in *The Rationalist Annual*; repr. as an appendix in J.D. Bernal *The Origin of Life*, London: Weidenfeld & Nicolson, 1967.(Along with Oparin, one of the founding articles offering a ‘protein-first’ model.)

Joyce, G.F. (1993) ’Prospects for Understanding the Origin of the RNA World’, in *The RNA World*, Cold Spring

Harbor, NY: Cold Spring Harbor Laboratory Press.(Detailed discussion of experimental evidence in support of the primordial RNA world hypothesis.)

Kauffman, S.A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*, New York: Oxford University Press.(Ambitious work by dynamic-systems theorist attempting to bring self-ordering phenomena into a neo-Darwinian framework.)

Lenoir, T. (1982) *The Strategy of Life: Teleology and Mechanics in Nineteenth Century German Biology*, Chicago, IL: University of Chicago Press.(An excellent account of the role of Kantian teleology in the formulation of modern embryology, physiology, cell theory and pathology.)

Margulis, L. (1970) *Origin of Eukaryotic Cells*, New Haven, CT: Yale University Press.(Fascinating account of the evolution of the eukaryotic cell through the symbiotic engulfment and association of prokaryotic cells.)

Muller, H.J. (1929) 'The Gene as the Basis of Life', *Proceedings of the International Congress of Plant Sciences* 1: 897-921; repr. in *Studies in Genetics: The Selected Papers of H.J. Muller*, Bloomington, IN: Indiana University Press, 1962.(Earliest exposition of a gene-first ontology.)

Oparin, A.I. (1924) *Proiskhozhdenie zhizni*, Moscow: Moskovskii Rabochi; trans. A. Synge, repr. as an appendix in J.D. Bernal (ed.) *The Origin of Life*, London: Weidenfeld & Nicolson, 1967.(The article which initiated the contemporary debate on the origins of life.)

Oparin, A.I. (1938) *Origin of Life*, New York: Dover.(Oparin here takes dialectical materialism into account in his considerations of the origins of life.)

Roe, S.A. (1981) *Matter, Life, and Generation: Eighteenth-Century Embryology and the Haller-Wolff Debate*, Cambridge: Cambridge University Press.(Most accessible account of the eighteenth-century debate between preformationism and epigenesis.)

Watson, J.D. and Crick, F.H.C. (1953) 'A Structure for Deoxyribose Nuleic Acid', *Nature* 4,356: 737-.(The classic paper on the structure of DNA.)

Limbo

According to traditional Roman Catholic teaching, limbo is the postmortem destination of those who have not been baptized, but are not guilty of sin. Lack of baptism bars such people from salvation, but their innocence means that they do not deserve the punishment of hell. They were thought to fall into two groups: the righteous of the Old Covenant, prior to the redemption of Christ, and unbaptized children. The former were supposed to have gone to heaven after Christ's death, but the latter had to stay in limbo forever. The existence of limbo was never dogmatically defined, and it was never given as much attention as heaven, hell or even purgatory, each of which represented a fate which human beings earned in part through personal choice. Nowadays, the possibility that unbaptized babies might be consigned to hell is not widely entertained, and some thinkers hold that the requirement of baptism for salvation is open to interpretation. Consequently, the idea of limbo is not as widely discussed as it once was.

1 History of the idea of limbo

In traditional Roman Catholic theology, limbo is a place for certain of the dead who are deprived of salvation because they have not been baptized, but who do not suffer the punishment of hell. It was usually thought to consist of two distinct parts: a limbo of the fathers (*limbus patrum*), or 'bosom of Abraham', which was a place for the righteous of the Old Covenant who could not enter heaven before the redemption of Jesus Christ; and a children's limbo (*limbus infantium*), which was a place for unbaptized babies and young children prior to the age at which they could commit sin. Such children would be in a state of original sin, but innocent of personal guilt. The limbo of the fathers was said to be empty ever since Christ's descent into hell (limbo) after the crucifixion, when all its inhabitants were transported to heaven, but the babies in the children's limbo were supposed to remain there for all eternity.

There is no mention of limbo in the Bible, and the problem of the fate of unbaptized babies was not addressed until the fifth century in response to the Pelagians, who asserted that God does not deny deceased unbaptized infants eternal life, although they are not in heaven either. The Pelagians maintained that they are in a state of innocence, not damnation, where they experience no pain and have a measure of happiness. Augustine replied that these infants are condemned to hell, but their punishment is much milder than that of adults, and it is preferable to annihilation. Augustine's position seems harsh, but his primary interest in this matter was to use it to get the Pelagians to reveal the heresy at the heart of their doctrine - the denial of original sin (see Pelagianism; Sin §2). The Pelagians' view on infants who die without being baptized was thus a casualty of the debate over original sin. Ironically, their position is very close to the idea of limbo that developed seven centuries later.

In the Greek Church, Augustine's contemporary, Gregory of Nyssa, had written that unbaptized infants neither suffer the torments of hell nor enter the life of the blessed in heaven. Otherwise the Augustinian position was accepted for centuries and was not broken until Peter Abelard argued that unbaptized infants do not suffer the pains of hellfire, but only grief at separation from God. Peter Lombard agrees with Abelard in his *Sentences* (*c.*1145-51), saying that unbaptized infants suffer only 'darkness' in the world to come (*Sententiarum libri quatuor*, lib. II, dist. 33, cap. 2). This development was important because Lombard's *Sentences* became a kind of university textbook, with tremendous influence on subsequent theology. Bonaventure clearly distinguishes between the children's limbo and hell, saying that divine justice establishes children in a state of unchanging love and knowledge, but which is neither joyful nor sad (*Il Sent.*, d.33, a.3, q.2). On the other hand, Albert the Great, who invented the term 'limbo', combines limbo and hell (*De resurrectione*, part 3). Thomas Aquinas took Bonaventure's position, arguing that original sin is not a positive fault, but a privation. Since the punishment must fit the crime it would be unreasonable for God to give a positive punishment for something that is not a positive fault. In arguing against punishment through sensory pain, Aquinas says that if a human being's powers are to be harmed through such pain, the punishment must bear some relation to a fault in the person who used those powers. But there can be no such fault in the case of an infant, so punishment through sensory pain would be unreasonable (*De malo*, q.5, a.3).

The problem of limbo was not handled explicitly by the Council of Trent (1545-63) and subsequently it was treated as an open question. The existence of limbo subsequently became the subject of controversy, with the

Augustinians and the Jansenists denying its existence and the Jesuits defending it. In this century it has been on the decline. It is not mentioned in the documents of Vatican II, the Ecumenical Council called by Pope John XXIII in the 1960s; while *Lumen gentium*, the ‘Dogmatic Constitution on the Church’, mentions briefly the necessity of baptism for salvation, it does not elaborate.

2 The status of limbo in contemporary thought

The limbo of the fathers has never received much attention, presumably because it is now supposed to be empty. The children’s limbo, however, not only poses an obvious theological problem, but also a pastoral one: what should priests tell the grieving parents of a child who dies shortly after birth or, for that matter, in the womb? The proposal that there is a children’s limbo arises out of two suppositions: that there is no salvation without baptism; and that unbaptized babies have made no personal choice whereby they would merit eternal punishment. The limbo of the fathers is predicated on the first assumption together with the assumption that the patriarchs of the Old Covenant would have gone to heaven had they lived after the time of Christ. The compromise position on the souls in these two limbos is that they are deprived of heavenly bliss (temporarily in the case of the patriarchs, and forever in the case of the unbaptized babies), but that they do not suffer. The justification for the limbo of the fathers has obvious extensions to other classes of persons who would have gone to heaven had they been in more favourable circumstances (such as the unbaptized in non-Christian countries), but these individuals were almost always placed in a different category. However, it is not at all clear why their status should not be parallel to that of the inhabitants of the *limbus patrum*. The latter were simply born at the wrong time; the former were simply born in the wrong place. An appreciation of this difficulty and admiration for the virtuous sages of antiquity led Dante, in the *Divina Commedia*, to propose a limbo of the sages, in which Socrates, Plato and Virgil were found. Dante’s poetic power reflects an enduring sensibility, one which has probably outlasted the more traditional theological views on the fate of these people.

It seems odd that the tradition never questioned the assumption that there is no salvation without baptism, but only the assumption that those in limbo do not suffer the pains of hell. However, the concept of baptism was sometimes interpreted broadly enough to include some persons who had not actually gone through the sacramental rite. Aquinas proposed two categories, Baptism of Blood and Baptism of Desire, to handle some of the problematic cases, and there were frequent appeals to these categories in Catholic theology until Vatican II. It is even possible to interpret baptism so widely that it applies to everyone. For example, Bruno Webb (1953) has suggested that death itself might be a kind of sacrament from which the saving power of baptism arises. It has become increasingly common, however, simply to dispute the requirement of baptism for salvation. Among Protestant philosophers, Jonathan Kvanvig (1993) has argued that there is no good reason for the doctrine of limbo, since it arises from a combination of a very severe conception of hell and a very restrictive conception of how it is possible to reach heaven.

The idea of limbo has therefore fallen into disuse, not because it is generally thought that the inhabitants of limbo are really in hell, but because it is often thought that they are in heaven. Moreover, the coherence of the concept has also been questioned. Ladislaus Boros argues that the notion of an exclusion from the human being’s proper and intended end that is nevertheless not experienced as suffering is self-contradictory (see Hick 1976). More and more, there is doubt that baptism is essential for salvation, so there is a growing tendency to reject the primary assumption of the dilemma which the existence of limbo was intended to resolve.

See also: Evil, problem of; Grace §3; Heaven; Hell; Purgatory; Salvation

LINDA ZAGZEBSKI

References and further reading

Aquinas, T. (1266-72) *Quaestiones disputatae de malo (Disputed Questions on Evil)*, trans. J.T. Oesterle, Notre Dame, IN: University of Notre Dame Press, 1981, q.5, a.3.(Aquinas does not have a great deal to say about limbo, but his work should be consulted because of his importance in the history of Christian doctrine. See also *Summa theologiae* IIIa, q.69.)

Dyer, G.J. (1964) *Limbo: Unsettled Questions*, New York: Sheed & Ward.(A good summary of the history of the idea of limbo.)

Hick, J. (1976) *Death and Eternal Life*, New York: Harper & Row.(An influential book, which includes a discussion of both Western and Eastern eschatologies; also discusses the position of Ladislaus Boros mentioned

in the text.)
Hill, P.J. (1962) *The Existence of a Children's Limbo According to Post-Tridentine Theologians*, dissertation, Vatican City: Pontificia Universitas Gregoriana.(A good treatment of the history of the idea of the children's limbo, the limbo most discussed in theological history. Written prior to Vatican II.)
Kvanvig, J.L. (1993) *The Problem of Hell*, New York: Oxford University Press. (This book is a philosophical treatment of the Christian idea of hell, but includes an argument against the idea of limbo on page 57.)
Lombard, P. (*c*.1145-51) *Sententiae in IV libris distinctae (Sentences in 4 Separate Books)*, 2 vols, ed. I.C. Brady, Grottaferrata: Collegi S. Bonaventura ad Claras Aquas, 3rd rev. edn, 1971-81, Libri quatuor, lib. II, dist. 33, cap. 2.(A discussion of the fate of unbaptized infants which had a strong influence on subsequent medieval theology.)
Webb, B. (1953) 'Unbaptized Infants and the Quasi-Sacrament of Death', *The Downside Review* (Summer): 243-57.(An example of the attempt to defend the salvation of unbaptized infants by broadening the concept of baptism.)

Linear logic

Linear logic was introduced by Jean-Yves Girard in 1987. Like classical logic it satisfies the law of the excluded middle and the principle of double negation, but, unlike classical logic, it has non-degenerate models. Models of logics are often given only at the level of provability, in that they provide denotations of formulas. However, we are also interested in models which provide denotations of deductions, or proofs. Given such a model two proofs are said to be equivalent if their denotations are equal. A model is said to be 'degenerate' if there are no formulas for which there exist at least two non-equivalent proofs. It is easy to see that models of classical logic are essentially degenerate because any formula is either true or false and so all proofs of a formula are considered equivalent. The intuitionist approach to this problem involves altering the meaning of the logical connectives but linear logic attacks the very connectives themselves, replacing them with more refined ones. Despite this there are simple translations between classical and linear logic.

One can see the need for such a refinement in another way. Both classical and intuitionistic logics could be said to deal with static truths; both validate the rule of modus ponens*: if* $A \to B$ *and A, then B; but both also validate the rule if* $A \to B$ *and A, then* $A \wedge B$*. In mathematics this is correct since a proposition, once verified, remains true - it persists. Many situations do not reflect such persistence but rather have an additional notion of causality. An implication* $A \to B$ *should reflect that a state B is accessible from a state A and, moreover, that state A is no longer available once the transition has been made. An example of this phenomenon is in chemistry where an implication* $A \to B$ *represents a reaction of components A to yield B. Thus if two hydrogen and one oxygen atoms bond to form a water molecule, they are consumed in the process and are no longer part of the current state. Linear logic provides logical connectives to describe such refined interpretations.*

1 Syntax

Given these new interpretations of implication and conjunction, they are given new symbols: '$\multimap$' and '$\otimes$', respectively. Thus the chemical reaction mentioned above would be given symbolically as $(\mathrm{H} \otimes \mathrm{H} \otimes \mathrm{O}) \multimap \mathrm{W}$.

Despite the appeal of these refined connectives it is clear that our logic is very weak: a formula can be used once only (hence the term 'linear'). To describe situations of static truths, where formulas can be reused (that is, allowed to persist) or ignored, we introduce a new logical connective '!' where $!A$ means that the formula *A* may be used as many times as needed (including zero). Thus the traditional notion of implication can be simulated by $(!A) \multimap B$. This idea leads to simple translations of both classical and intuitionistic logics into linear logic.

Given two transitions $A \multimap B$ and $A \multimap C$, we clearly have the transition $(A \otimes A) \multimap (B \otimes C)$, but not $A \multimap (B \otimes C)$. However, we shall define a new transition $A \multimap (B \mathbin{\&} C)$, which represents the superimposition of the two transitions. It represents a choice of the resulting state *B* or *C*. Despite its disjunctive feel the connective '$\&$' is really a conjunction - both states are being offered but only one may be chosen, via one of the axioms $(A \mathbin{\&} B) \multimap A$, $(A \mathbin{\&} B) \multimap B$. Consequently linear logic has two forms of conjunction, one ($\otimes$) where the initial states are disjoint and one ($\&$) where they are identical. The former is known as a multiplicative connective and the latter as additive. Symmetrically linear logic has two forms of disjunction, the multiplicative form written '$\star$' and the additive form written '$\oplus$'.

Linear formulas are built up from a collection of (positive and negative) atomic formulas, written p and $p^{\perp}$, by means of the multiplicative connectives, '$\otimes$' and '$\star$', additive connectives, '$\&$' and '$\oplus$' and the so-called exponentials, '!' and its dual '?'. The connectives have units (in the sense that in classical logic true (T) is a unit for conjunction ($\wedge$), as $A \wedge \mathrm{T} \equiv A$): two multiplicative units, 1 and $\perp$, and two additive units, $\top$ and 0; such that $A \otimes 1 \equiv A$, $A \star \perp \equiv A$, $A \mathbin{\&} \top \equiv A$ and $A \oplus 0 \equiv A$. (The logical symbols used here match Girard's rather idiosyncratic selection, apart from '$\star$' which Girard writes as an upside-down ampersand.)

As with classical logic, negation, written $A^{\perp}$, is a defined notion given inductively over a composite formula. This is given by:

$$(p^\perp)^\perp \text{ is } p$$
$$(A \otimes B)^\perp \text{ is } A^\perp \star B^\perp$$
$$(A \star B)^\perp \text{ is } A^\perp \otimes B^\perp$$
$$(A \mathbin{\&} B)^\perp \text{ is } A^\perp \oplus B^\perp$$
$$(A \oplus B)^\perp \text{ is } A^\perp \mathbin{\&} B^\perp$$
$$1^\perp \text{ is } \perp$$
$$\perp^\perp \text{ is } 1$$
$$\top^\perp \text{ is } 0$$
$$0^\perp \text{ is } \top$$
$$(!A)^\perp \text{ is } ?(A^\perp)$$
$$(?A)^\perp \text{ is } !(A^\perp).$$

Also as with classical logic, we are able to define an implication using negation and disjunction: thus $A \multimap B$ is $A^\perp \star B$.

Gentzen's sequent calculus (see Natural deduction, tableau and sequent systems §2) is a tool for studying various rules of logics. Sequents are written $\Gamma \vdash \Delta$, where Γ $(= A_1, \ldots, A_n)$ and Δ $(= B_1, \ldots, B_m)$ are finite sets of formulas. The traditional reading of such sequents is that $A_1 \wedge \ldots \wedge A_n$ implies $B_1 \vee \ldots \vee B_m$. In the linear setting it says that $A_1 \otimes \ldots \otimes A_n$ implies $B_1 \star \ldots \star B_m$. However, we can employ a shorthand by considering only sequents of the form $\vdash \Delta$, where general sequents of the form $\Gamma \vdash \Delta$ are simulated using $\vdash \Gamma^\perp, \Delta$. The main rules are as follows.

$$\frac{}{\vdash p, p^\perp} \quad \text{(identity)}$$

$$\frac{\vdash \Gamma, A \qquad \vdash A^\perp, \Delta}{\vdash \Gamma, \Delta} \quad \text{(cut)}$$

$$\frac{}{\vdash 1} \quad (1) \qquad \frac{\vdash \Gamma}{\vdash \Gamma, \perp} \quad (\perp)$$

$$\frac{\vdash \Gamma, A \qquad \vdash \Delta, B}{\vdash \Gamma, \Delta, A \otimes B} \quad \text{(tensor)}$$

$$\frac{\vdash \Gamma, A, B}{\vdash \Gamma, A \star B} \quad \text{(par)}$$

$$\frac{\vdash \Gamma, A \qquad \vdash \Gamma, B}{\vdash \Gamma, A \mathbin{\&} B} \quad \text{(with)}$$

$$\frac{\vdash \Gamma, A}{\vdash \Gamma, A \oplus B} \quad (\text{sum}_1)$$

$$\frac{\vdash \Gamma, B}{\vdash \Gamma, A \oplus B} \quad (\text{sum}_2)$$

The difference between the multiplicative and additive connectives can be seen quite clearly by comparing the 'tensor' and 'with' rules. As is the case normally, writing the sequents in a two-sided fashion, and restricting the right-hand side so as to contain at most one formula (for example, $\Gamma \vdash A$), results in the so-called intuitionistic fragment of linear logic. It should be noted that some of the connectives are not available in the intuitionistic fragment. The sequent calculus formulation enjoys the important metatheoretic property of cut elimination (Gentzen's *Hauptsatz*), that is, there is an algorithm which rewrites any proof into one which does not make use of the 'cut' rule.

Girard also showed how one may refine the sequent calculus formulation further by replacing the sequents with graphs. Thus formulas become nodes and applications of the rules become methods for transforming graphs. For

example, the ‘identity’ rule is replaced with two nodes (p and $p^{\perp}$) and an arc between them, whereas the tensor rule takes the two nodes *A* and *B* and places arcs between them and a new node $A \otimes B$. These graphs are known as ‘proof nets’. Several sequent proofs are represented by a single proof net and so one may think of proof nets as revealing the essential inner structure of a linear logic proof.

2 Models

An important feature of linear logic is that its models need not be degenerate, in the sense given earlier. This is in contrast to classical logic. Using the mathematical language of category theory it is relatively easy to describe abstractly which structures are necessary for a model of linear logic: essentially they form particular categories first identified in the 1970s by Barr (see Barr 1991; Category theory, introduction to). A further problem is to construct concrete models, which are ideally complete (that is, they contain only denotations of provable formulas and valid deductions).

Girard offered two models of linear logic in his seminal 1987 paper. The first, ‘phase spaces’, is a model based on a certain sort of algebraic structure. The second, ‘coherent spaces’, is a graph-theoretic model. A large number of different models have since been identified and again it should be stressed that all of these models are non-degenerate.

An appealing model was given by Blass (1992), based on earlier work on constructive logic by Lorenzen. Here the meaning of a proposition *A* is given by specifying a game, or dialogue, between two players, one of whom is trying to assert *A* and the other refute it. Thus a proposition is interpreted as a game and validity by a winning strategy. The connectives can then be understood as operations on games. For example, the connective ‘ $\otimes$’ is interpreted as an operation which interleaves two disjoint games or, in other words, plays them side by side. These refined notions of games have been the subject of some intensive work. Despite this effort, and the intuitive appeal of the approach, these models suffer from being incomplete - they validate additional axioms, mostly ‘weakening’: $A \multimap 1$. Some models validate an additional axiom, known as ‘mix’: $(A \otimes B) \multimap (A \star B)$. The status of this axiom (which can also be given as a sequent calculus rule) is uncertain but may be decided by a convincing (game) model. Indeed providing a convincing complete model still remains an open problem.

3 Decidability

One can gain a useful insight into a logic by considering how difficult it is to decide whether a given formula is provable in that logic. Most first-order logics are undecidable but there is considerable variety when considering propositional logics. Both propositional classical logic and propositional intuitionistic logic are decidable albeit with differing complexity ($\mathcal{NP}$-complete and $\mathcal{PSPACE}$-complete, respectively; see Complexity, computational §4). Given that we can encode both these logics into linear logic one might expect linear logic to be decidable. Surprisingly it has been shown that propositional linear logic is undecidable. Indeed even the intuitionistic fragment of linear logic is undecidable. A considerable amount of work exists considering the decidability and complexity of various fragments of linear logic obtained by selecting a subset of the connectives and units.

4 Applications

As linear logic is a refined version of both classical and intuitionistic logic, it is not surprising that it has been applied to many areas of mathematical logic and theoretical computer science. Linear logic’s refined connectives and notion of proof offer a new understanding of classical logic’s connectives and proof deductions. In particular the notion of proof dynamics - the process of cut elimination mentioned in §1 - can be studied with improved accuracy using proof nets. As there are simple translations from classical to linear logic, it is possible to use the non-degenerate models of linear logic to tease out non-degenerate models for classical logic. Many researchers refer to this process as the extraction of the ‘constructive content of classical logic’.

Logic is used by computer scientists in many ways to represent computation. One method is to represent a program as a proof and computation as the process of simplifying the proof or, in other words, bringing it to a normal form (see Computability theory §3). Another method is to represent a program as a set of rules and a goal, and computation as the process of searching for a proof of the goal given the rules. Given its refined nature, linear logic sheds new light on both these methods.

Importantly, deriving from some of the ideas of the game models described in §2, linear logic appears naturally to encode a process of computation which is inherently parallel in nature. Realizing this process and relating it to existing models of parallel computation is another important open area.

See also: Logical and mathematical terms, glossary of

G.M. BIERMAN

References and further reading

Barr, M. (1991) '⋆-Autonomous Categories and Linear Logic', *Mathematical Structures in Computer Science* 1: 159-78.(A summary of earlier work by Barr which is re-considered in the light of linear logic.)

Blass, A. (1992) 'A Game Semantics for Linear Logic', *Annals of Pure and Applied Logic* 56: 183-220.(A detailed presentation of the game model discussed in §2.)

Girard, J.-Y. (1987) 'Linear Logic', *Theoretical Computer Science* 50: 1-101.(The original paper and still essential reading.)

Girard, J.-Y., Lafont, Y. and Regnier, L. (eds) (1995) *Advances in Linear Logic*, Cambridge: Cambridge University Press, esp. 1-42, 109-22.(Essentially a collection of research papers but includes an introduction to linear logic by Girard and a survey of decidability results by Lincoln.)

Hyland, J.M.E. (1997) 'Game Semantics', in A.M. Pitts and P. Dybjer (eds) *Semantics and Logics of Computation*, Cambridge: Cambridge University Press.(A clear tutorial on how to construct particular game models of linear logic.)

Troelstra, A.S. (1992) *Lectures on Linear Logic*, CSLI Lecture Notes 29, Stanford, CA: Center for the Study of Language and Information.(A broad but rapid treatment of linear logic.)

Linguistic discrimination

'Linguistic discrimination' is a redundancy. Discriminating is at the heart of what languages do. The question, of course, is when they can be said to do it invidiously, or rather when we, in our use of language, can be said to be discriminating invidiously. In Aristotelian terms, the proper use of linguistic discriminations is to make the right sort and number of discriminations in the right ways and at the right times - that is, not to discriminate between those things that, for the legitimate purposes at hand, ought to be seen as the same; to discriminate between those things that, for the legitimate purposes at hand, ought to be seen as different; and to discriminate in ways that advance legitimate and not illegitimate purposes. Disputes about what constitutes linguistic discrimination (in the invidious sense) revolve around both the legitimacy of our purposes and, in the light of those purposes, the aptness of particular discriminations. Such disputes presume both that our language shapes our actions (that linguistic discrimination plays a role in maintaining unjust inequalities) and that our actions can shape our language (that acknowledging such discrimination can and should lead to linguistic change).

1 Arguments for causality and remediability

Claims for the causal efficacy of linguistic discrimination range from the claim that some thoughts are made either unavoidable or impossible because of the language in which thinking occurs, to the claim that language shapes, without determining, how we think. The stronger claim, often appealing to the Sapir-Whorf hypothesis, sees language as providing the conceptual tools without which the experienced world is inchoate and only in terms of which can sense be made (see Sapir-Whorf hypothesis). In its strongest form, this hypothesis is surely false: language-users are inventive, and all languages have room for invention.

It is harder to refute the view that the languages we use shape our thoughts, making some ways of thinking more 'natural', more readily comprehended by our interlocutors, harder to avoid if they come to seem problematic. Richly nuanced vocabularies exist to describe some parts of the experienced world and not others. Nor are all the differences semantic: Black English is less well suited than is Standard English for expressing the agentless passive, that is, the idea that 'it happened' without indicating that someone or other did something (Jordan 1985).

That the language we use does not absolutely constrain us is no reason not to attend to the ways in which it shapes our thinking, leading us to imply what we do not intend, or making it difficult for us to say what we mean, or what we would mean if we had the words to mean it with. It is also remediable. Languages constantly change in response to an indefinite range of social forces; and while there can be no guarantee that a suggested reform will catch on, there is no reason a priori to argue that it could not do so. One may choose to resist it, but that choice is as political as is the argument for change.

2 Taxonomic schemes

Metaphysicians have long debated the nature of kinds. Are some of them 'natural' (whatever that means)? Are only such kinds, if there are any, 'real'? What about kinds created by our taxonomizing practices? Inspired largely by the work of Michel Foucault, theorists are examining the social and political contexts within which taxonomies operate and the value judgments they both depend on and reinforce (see Root 1993). Case studies explore the social construction of race (Goldberg 1990) and gender (Lorber 1994); as well as of kinds of practices such as sexual harassment (MacKinnon 1979) and child abuse (Hacking 1992), conditions such as premenstrual syndrome (Zita 1988), life stages such as childhood (Aries 1962), and persons such as homosexuals (Stein 1992) and heterosexuals (Katz 1990). Questions about invidious discrimination arise with respect to kinds that arguably rest on dubious science or on politically questionable social practices.

Thus racial classifications have been argued to be discriminatory because they are biologically unfounded, on the assumption that races, if 'real', are so because of discoverable biological differences between different racial populations. Since there appear not to be such differences, racial classifications distinguish as different those who, on the alleged basis, are in fact the same. Alternatively, it can be argued that race is best understood as a social classification, in which case the issue is whether as such race can be made coherent and, if so, whether the social practices that make it coherent are just. If racial classification cannot be made coherent, then it is discriminatory for distinguishing between people who, on its own terms, are not distinguishable; if it is coherent but grounded on

unjust social practices, then racial classification would be invidiously discriminatory because it was wrong to engage in the practices that make it descriptively correct. Others argue that, although founded on racist practices, racial classification has acquired a conceptually adequate grounding in practices that include those of anti-racist resistance and positive group identification (see Race, theories of). Similar questions are raised about the classifying of persons as homosexual or heterosexual: apart from social practices of stigmatizing homosexuality as unnatural, perverse or sick, what reasons are there for classifying persons according to whether they desire those of the same or the opposite gender (see Sexuality, philosophy of)?

The situation is somewhat different regarding discrimination by gender (see Language and gender). There is better reason than in the case of race to regard the distinction between males and females as grounded in biology, although it is a matter of dispute whether the social practices that at the very least build upon, extend and ‘clean up’ the biological differences are morally and politically justifiable. But there is a further question: gender ascription, in English and in many other languages, is uniquely non-optional, in particular because of the gender inflection of pronouns. This ubiquitous marking of gender is arguably discriminatory, implying that, unlike all the other equally real differences between people, gender is ubiquitously relevant. (Similarly, the argument for the use of ’Ms’ rests not on the wrongness of distinguishing between unmarried and married women but on the wrongness of requiring that such a distinction be marked whenever a woman is referred to by title.) Ubiquitous gender-marking is part of the social construction of gender, both in making gender neutrality difficult to express even when it is agreed to be appropriate and in making gender ambiguity socially stigmatizing even when it is acknowledged to be possible.

3 The paradigmatic and the generic

Semantic spaces are complex: not all unquestionable members of a class are equally paradigmatic. It seems to be a feature of human cognition to learn categories by grouping members around paradigmatic exemplars, and variation in paradigms is not for the most part idiosyncratic. It is a matter of dispute how much of the contouring of semantic space is ’hard-wired’, but it is clear that for at least many categories, paradigmatic exemplars owe that status to culturally variable, frequently unjust, social arrangements.

Thus, it has been pointed out (Spelman 1988) that within racist cultures, white people are paradigmatic and people of colour are variously ’different’, so that even anti-racist arguments claim that for all relevant purposes, people of colour are ’just like’ white people. Similarly for other forms of privilege: those who are, for example, male or middle-class or heterosexual are paradigmatic in the sense that their achievements, interests and needs are taken as the measures against which those of others are evaluated. When, however, the category is itself a subordinate one, paradigmatic exemplars are likely to be members of a relevantly subordinated group. In either case, those who conform to the paradigm are typically ’unmarked’, while those who are not paradigmatic are ’marked’: ’male nurse’, ’gay author’.

It is a matter of controversy the extent to which such language is itself discriminatory, rather than just accurately reflecting reality. The arguments for claiming specifically linguistic discrimination point out that being in the majority and being paradigmatic are not always the same thing and, even when they are, linguistic discrimination helps to maintain, rather than innocently reflecting, social discrimination.

A limiting case of the paradigmatic is the false generic. False generics are terms for the paradigmatic members of a group improperly used to refer to all members. It has been a matter of controversy whether in English masculine nouns and pronouns are also, ambiguously, genuinely generic, or whether they should be regarded as false generics. The arguments for the latter position are both linguistic and political. Linguistic arguments (see Mercier 1995) point to the often irresolvable ambiguities that result from the employment of the masculine as generic, as well as to the anomalous and illogical rules that govern such employment. Political arguments point to the advantages that accrue to those who are taken to be generically human, rather than ’different’, and to the alienation increasingly experienced by female listeners and readers uncertain of the extent to which, if at all, they are included among the ’men’ under discussion.

4 Stereotyping and stigmatizing

Another type of linguistic discrimination concerns the vocabulary used to refer to members of privileged and

subordinated groups and, by analogy, parts of the nonhuman world that are variously associated with them. Tropes of light and darkness, for example, are central to European discussions of reason, especially those that were articulated concurrently with the colonization of Africa - the ’dark continent’ - to which Europeans had the right and duty to bring ’enlightenment’. Gender stereotypes permeate talk about ’mother’ nature, the ’hard’ versus the ’soft’ sciences, ’active’ reason versus the ’passions’. Such vocabulary relies on and reinforces stereotypes that are inextricably bound up with socially discriminatory practices and, for reasons similar to those given in §3, can be argued to constitute linguistic discrimination.

Stigmatizing vocabulary, whether used to refer directly to members of stigmatized groups or derivatively to things associated with such groups, exemplifies linguistic discrimination not so much in how the group in question is being distinguished but in how language is being used in reference to them. The clearest examples are derogatory terms applied on the basis of race, gender, sexuality, disability, national origin, religion, and so forth. A related form of linguistic discrimination is the use of terms referring to stigmatized groups as negative characterizations: ’Indian-giver’ or ’Jewing someone down’. Implicitly or explicitly gendered vocabulary typically encodes problematically stereotypical views concerning, for example, the different valuations of male versus female sexuality and assertiveness. Language referring to disabilities is often problematically generalized in ways that imply, for example, that blindness stands in the way of ’seeing’ what follows from a set of premises.

Even when such forms of linguistic discrimination are not blatantly bigoted, there is good reason to avoid them, given the usual desirability of communicating as clearly as possible what is intended to the intended audience. Increasingly, audiences are reading discriminatory intent in the use of discriminatory language (including the supposedly generic masculine), and such reading habits are not in the power of authors to circumvent. To the extent that it is no part of authorial intent either to communicate negative or marginalizing attitudes towards members of particular groups, or to exclude members of such groups (and their allies) from one’s audience, it is wise to eschew vocabulary that will predictably have those effects.

5 Value-laden terminology

Whether or not it is in theory possible for the discriminations made by some particular language (or part of a language) to reflect only morally and politically non-tendentious judgments of similarity and difference, it is clear that such neutrality does not characterize a large number of the distinctions made in the languages we actually use. Public discussions are significantly shaped, in ways that frequently elude explicit debate, by the language in which the issues are framed. A central tenet of moral and political thinking is that like cases ought to be treated alike; consequently the crux of many arguments lies in judgments of similarity and difference, carried out in language that can make some of those judgments nearly automatic and others virtually unthinkable.

Examples abound, from military euphemisms that draw attention towards the achievement of strategic objectives and away from killing people, to abortion debates that pit those who favour choice against those who would coerce child-bearing or those who are pro-life against those who would kill babies. Mainstream press coverage in the USA distinguishes between freedom-fighters and terrorists in ways that map not the activities in which those individuals and groups engage so much as their congruence with or opposition to US foreign policy. Large-scale violence in inner city communities is thought of in one way when called a ’riot’, in another when called an ‘uprising’.

The notion of an ’essentially contested concept’ (Gallie 1956) raises similar questions. Such concepts are deeply value-laden, carrying strongly positive or negative connotations: ’democracy’ or ’repression’. Disputants typically share a small core of paradigmatic exemplars and differ over which additional examples are relevantly similar. A discriminatory use of an essentially contested concept would be one that applied and withheld the term in ways that are argued to be either undermotivated or motivated in problematic ways.

The point is not that we ought to eschew language that encodes controversial judgments of similarity and difference. Rather, we should be aware of what is encoded in our language and ready to argue for the judgments expressed therein and to challenge those of others with which we disagree. To call a particular use of language discriminatory goes beyond the claim that it implies a disputable judgment: rather, it commits one to arguing for why and how that judgment ought to be disputed.

6 Languages and dialects

The historical development of languages has gone on in tandem with the histories of the formation of nations, the spread of and resistance to colonization and imperialism, the globalization of culture and the influence of nationalist movements, and the development and spread of communication technologies. Language is typically bound up with national identity, often in complex and contradictory ways. For example, movements of African nationalism sometimes embrace the language of the colonizer as the language of national unity, rather than divisively choosing among a number of indigenous languages.

Questions about whether something is a language, rather than a dialect, or a dialect, rather than a degraded use of language, will always raise contentious issues concerning present usage and users, as well as about their histories. Charges of discrimination will be appropriate when such judgments appear to be ethnocentric or otherwise biased, as, for example, in the controversy over whether American Black English is a legitimate dialect. The argument that it is rests on pointing out the rule-governed nature of its syntactic, semantic and phonological differences from Standard English, and on the history of its development from the interaction of West African languages with English (Smitherman 1977).

Languages and dialects are attached to particular groups of people, who are related to other groups in frequently hierarchical ways, and those hierarchies are reflected in the ways in which the languages and dialects are thought of and treated. Attempts are sometimes made to maintain the ’purity’ of a language by codifying a version of it that carries both historical and present day prestige. Educational systems typically single out certain dialects as ’standard’ or ’received’: as the paradigmatic exemplars of the language they tend to be regarded not as dialects, but as the language itself, unmarked. Similarly, some speech patterns count as ’accents’, and different accents lead to different stereotypical ascriptions of intelligence and social standing.

Reflecting the dominance of the USA following the Second World War, English is increasingly the global language - used, for example, in commerce, electronic communications, aviation and science - a development that privileges those who speak it fluently and marginalizes or threatens the survival of other languages. To regard the globalization of English as discriminatory is to regard that situation as other than natural, inevitable or equitably beneficial. Discrimination is also an issue when it comes to the semantic, syntactic and phonological diversity of English as spoken in different countries. Britain and the USA each have large amounts of rather different sorts of clout when it comes to determining whether or not something counts as ’proper’ English, while even native English speakers from elsewhere are generally regarded as speaking either incorrectly or in a substandard dialect.

See also: Discrimination

NAOMI SCHEMAN

References and further reading

Aries, P. (1962) *Centuries of Childhood*, London: Jonathan Cape.(On the ‘invention’ of childhood. One of the earliest discussions of how an apparently natural kind - children - can be socially constructed.)

Christensen, F.M. (1994) ‘"Sexual Harassment" Must Be Eliminated’, *Public Affairs Quarterly* 8: 1-17.(Argues against the legitimacy of the concept of sexual harassment.)

Gallie, W.B. (1956) ’Essentially Contested Concepts’, *Proceedings of the Aristotelian Society* 56: 167-98. (Explanation of the concept referred to in §5.)

Goldberg, D.T. (ed.) (1990) *Anatomy of Racism*, Minneapolis, MN: University of Minnesota Press.(Includes essays on the nature of racial classification and a bibliography.)

Hacking, I. (1992) ’World-making by Kind-making: Child abuse for example’, in M. Douglas and D. Hull (eds) *How Classification Works: Nelson Goodman among the Social Sciences*, Edinburgh: Edinburgh University Press.(This volume includes other relevant essays on the nature of kinds in relation to Goodman’s work on similarity.)

Jordan, J. (1985) ’Nobody Mean More to Me than You and the Future Life of Willie Jordan’, in *On Call: Political Essays*, Boston, MA: South End Press.(Discussion of the expressive resources of Black English and of the practical choices about its use.)

Katz, J.N. (1990) ’The Invention of Heterosexuality’, *Socialist Review* 20: 7-34.(Argues that heterosexuality as an identity or ’sexual orientation’ - no less than homosexuality - is distinctively modern.)

Lorber, J. (1994) *Paradoxes of Gender*, New Haven, CT: Yale University Press.(Critical overview of the social

construction of gender; includes extensive bibliography.)

MacKinnon, C.A. (1979) *Sexual Harassment of Working Women*, New Haven, CT: Yale University Press. (Germinal text for the recognition of sexual harassment as a distinctive type of activity.)

Matsuda, M.J., Lawrence III, C.R., Delgado, R. and Crenshaw, K.W. (1993) *Words that Wound: Critical Race Theory, Assaultive Speech, and the First Amendment*, Boulder, CO: Westview Press.(Arguments for the assaultive power of language and about the legal consequences of recognizing that power, especially in light of the US constitutional protection of the freedom of speech.)

Mercier, A. (1995) 'A Perverse Case of the Contingent A Priori: On the logic of emasculating language (A reply to Dawkins and Dummett)', *Philosophical Topics* 23: 221-59.(Arguments, some linguistically technical, against the allegedly generic use of masculine language.)

Root, M. (1993) *Philosophy of Social Science*, Oxford: Blackwell.(Chapter 10, 'Sorting data into kinds', discusses the ways in which kinds can be socially constructed.)

Shapiro, M. (ed.) (1984) *Language and Politics*, New York: New York University Press.(Includes Foucault's 'The Order of Discourse' as well as a number of other, methodologically diverse essays on the language of politics and the politics of language.)

Smitherman, G. (1977) *Talkin and Testifyin: The Language of Black America*, Detroit, MI: Wayne State University Press.(Arguments for the legitimacy of Black English, an account of its history, and outlines of its syntax, semantics and pragmatics.)

Spelman, E.V. (1988) *Inessential Woman: Problems of Exclusion in Feminist Thought*, Boston, MA: Beacon Press.(Includes a discussion of the ways in which dominant (in this case white) members of a group (women) come to be the paradigmatic or even the sole understood referent of the generic term.)

Stein, E. (ed.) (1992) *Forms of Desire: Sexual Orientation and the Social Constructionist Controversy*, New York and London: Routledge.(Philosophically framed collection of essays on what it means, and whether or not it is true to say that, in particular, homosexuality as an identity or 'sexual orientation' is distinctively modern. Includes a bibliography.)

Young, A. (1995) *The Harmony of Illusions: Inventing Post-traumatic Stress Disorder*, Princeton, NJ: Princeton University Press.

Zita, J.N. (1988) 'The Premenstrual Syndrome: Dis-easing the female cycle', *Hypatia* 3: 77-99.(Arguments for the role of masculine bias in the characterization of PMS and in the social construction of the 'clinical body'.)

Linji (810/15-67)

Linji was one of the most reputed, and influential Chinese Chan masters in the history of East Asian Buddhism. He belonged to a school which advocates sudden enlightenment without dependence on words: there is an extralinguistic reality that can be intuitively apprehended through the rigorous meditative training. A person with this intuition escapes dualistic thinking and has grasped the freedom to act decisively, utilizing creatively whatever is presented before him/her. Linji's method of teaching is often characterized as 'thundering shouts and showering sticks', actions which are used to effect an awakening in his disciples. His reputation rests primarily on his ability to seize on this opportunity.

Linji Yixuan, a native of Nanhua county, became a monk at an early age, and studied *vinaya*, the *sūtras* and their commentaries extensively. Unable to gain peace of mind through textual studies, however, he changed to Chan (in Japanese, Zen) Buddhism under the tutelage of Huangbo (?-850). The *Linjilu* describes his dedication to Chan meditation as pure and single-minded. He attained great enlightenment through the assistance of both Huangbo and Dayu, prior to which event he said he was 'totally in the dark'. After twenty years with Huangbo, Linji moved in 849-50 to a temple in Zhengding county of Hebei province, where he taught his disciples for some ten years until his death at the Xinghua Temple. Posthumously, he was given the title 'Chan Master of Illuminating Wisdom'. The *Linjilu*, a record of his sermons, critical examinations and a biography, was compiled in 1120 by Yuanjue Zongyan.

Linji's philosophy focuses on the affirmation of the 'here and now' in the everyday world by bringing at a stroke the transcendental world (that is, the world of *buddhas* and patriarchs) into this world by collapsing the distinction between them. For example, we find in the *Linjilu* that 'there is a true person of no rank in a red chunk of your flesh, and it always comes and goes out of your face'. The 'true person of no rank' refers to the transparent light which, though invisible to ordinary eyes, shines forth in one's daily activities ('your face'). Because the light permeates the person as they originally are, the person lacks nothing, and in order to be emancipated from the sufferings and struggles encountered while living in the world and beyond, all one has to do is to come to this awakening. However, this awakening is paradoxical in that 'if you try to seek it, it retreats into the distance, and if you don't seek it, it appears before your eyes'.

The paradox arises, Linji observed, because the ordinary person suffers from a 'dis-ease' in the pursuit of the Buddha Way. He attributed the cause mainly to the lack of self-confidence. Doubts created in the mind cause the mind to fluctuate, chasing thoughts or images one after another: the fluctuating mind operates on the basis of dependence on people, language and things, through which the mind enters a conditioning process of transformation with the external world. Unable to discern this transformational process and mistaking it to be real, the ordinary person accepts the phenomena thus presented. For Linji, however, these are simply metamorphosing illusions and delusions, and are 'self-binding without any rope'. True reality discloses itself in a clear illumination beyond any dependence upon people, language and things.

Linji's cure for the 'dis-ease' was to discover the mind that does not fluctuate by achieving the stillness of meditation, and thereby to disengage it from the dependent, conditioning process of transformation so that the mind becomes immovable. He called it the one (unified and unifying) mind. If the mind is immovable, it becomes nondependent. It may be pointed out that the mind of dependence objectifies the things of experience and substantializes them through its discursive and discriminatory activity. This renders the mind incapable of letting things disclose themselves in total transparency because it creates the distinction between the appearance of a thing, and its nature. By contrast, the nondependent mind performs none of these activities, for it does not engage an object by discriminating and objectifying it.

Linji's position of nondependent mind leads to the thesis that anything existent lacks a substantial nature that persists through time. Everything is empty of substance, including the mind which experiences its world. In the experience of emptiness where all *dharmas* are experienced as one in the transparency of pure light, the person realizes that nothing whatsoever is obtainable, even the goal of becoming a *buddha*, for if everything is empty, nothing can be obtained. The mind which realizes this is without form, that is, 'no-mind', in virtue of which its activity is capable of penetrating the whole world (see Buddhist concept of emptiness).

In summary, the nondependent mind frees the person from the adverse, delusory influences of the everyday world, for the conditioning process is incapable of affecting the immovable, no-mind. The person with nondependent mind is also free from the karmic destiny of birth and death, because the experience of emancipation *vis-à-vis* the emptiness of things unties the person from his/her self-binding (see Karma and rebirth, Indian conceptions of). Moreover, based on this realization, the action performed by the person of nondependence does not leave any karmic residue, and therefore the freedom from karmic determination enables the person to freely engage, by way of verbal or bodily behaviour, the matters of the triple world - the experiential worlds of desire, form and formlessness. Freedom for Linji is grounded in the power of discerning 'the *buddha* from the devil, the true from the false, and the sacred from the profane'. Linji encapsulated the characteristics of the person of nondependence in his statement: 'If one becomes a master in any place, wherever he stands is true'. He called such a person the 'person of no-event'. He exhorted people to make use of the nondependent, immovable no-mind by embodying it.

See also: Buddhist philosophy, Chinese; Buddhist concept of emptiness

SHIGENORI NAGATOMO

References and further reading

App, U. (ed.) (1993) *Rinzairoku ichijisakuhin (A Concordance to the Record of Linji (Rinzai))*, Kyoto: International Research Institute for Zen Buddhism.(Locates any phrase or character in the *Linjilu* in order of their occurrence.)

Dumoulin, H. (1988) 'Lin-chi', in *Zen Buddhism: India and China*, New York: Macmillan, ch.10.(A general historical introduction.)

Iriya Yoshitaka (ed. and trans.) (1989) *Rinzairoku (The Record of Linji)*, Tokyo: Iwanami Shoten.(A text with accompanying translations incorporating Tang colloquialisms.)

Sasaki, R.F. (1975) *The Recorded Sayings of Ch'an Master Lin-chi Hui-chao of Chen Prefecture*, Kyoto: Institute of Zen Studies, Nanazno College.(A reliable translation of the *Linjilu*.)

Yanagita Seizan (1972) 'The Life of Lin-chi I-hsuan', *Eastern Buddhist* 5: 70-94.(A detailed account of Linji's life.)

Linnaeus, Carl von (1707-78)

Linnaeus was educated in Sweden, and became a doctor of medicine in Harderwijk, Holland, in 1735. He visited other European countries then, but he never left Sweden after his return in 1738. After practising as a physician in Stockholm, he moved to Uppsala University as professor of medicine and botany in 1741. He articulated four different but complementary ways of understanding nature - through two kinds of classification, and through what can be called developmental and functional/ecological interactions. Linnaeus is best known for his classificatory work, for which he received material from all over the world. His classificatory precepts are elaborated in the Philosophia botanica of 1751, an enlarged version of the 365 aphorisms of his Fundamenta botanica of 1735; the other aspects of his work are diffused through his writings. His artificial classification system, initially very popular, was replaced by the 'natural' system, more slowly in botany than in zoology, and more slowly in England than in some other countries. Current biological nomenclature is based on his Species plantarum, edition 1 (for plants), and Systema naturae, edition 10 (for animals). His codification of botanical terms remains influential. Almost 200 dissertations, most written by Linnaeus, were defended by his students. In these and other less well-known works, including the unpublished Nemesis divina (Stories of Divine Retribution), he covered a wide range of subjects. Quinarian thinking is noticeable in Linnaeus' work - there are five ranks in systems, five years' growth in flowers - and in some of the occult works that he knew. He also shows a strong combinatorial bent and a tendency to draw close analogies between the parts of animals and plants.

1 Formal classification

Linnaeus was educated at Lund and Uppsala universities (he was originally destined for the priesthood); his field knowledge was largely restricted to Sweden. During Linnaeus' three years in Holland, mostly in the employ of the wealthy banker George Clifford, he published extensively, including the first editions of *Systema naturae* and *Genera plantarum*. Later, his work mostly represents elaborations of his early ideas.

Linnaeus defined natural history as the classification and naming of plants, animals (including humans) and minerals according to the number, shape, position and proportion of their parts. Classification (*dipositio*) was based on system, Ariadne's thread, and was applicable to all branches of knowledge. There were five ranks, class, order, genus, species and variety, members of each rank (except the last) including several members of the subordinate rank. Successive divisions were based as far as possible on essences, 'lubricious' distinctions, such as those between herb and tree, being avoided. Classes and orders in the sexual system of plants ('such loathsome harlotry', according to Johann Siegesbeck) were distinguished by stamen number, whether or not the stamens were joined together, and their position relative to the pistil; they were the work of nature and art.

Genera and species were the work of nature. Genera, the focus of Linnaeus' work, were defined by features of the fructification (flower and fruit), largely those of external form. Linnaeus calculated that the different parts of the fructification, when combined in all possible ways, would yield sufficient features to distinguish all plant genera. These features were often taken from Linnaeus' examination of one species per genus. Other species were added because of their agreement in *habitus*, features other than those of the fructification. Hence Linnaeus' species of *Vaccinium* (cranberries, blueberries), although similar in appearance, had different numbers of stamens in the flower - and so should have been placed in different classes. Linnaeus' contemporaries were dismayed by such inconsistencies; Linnaeus was not.

Species were the basic unit of nature, and he first thought they were all created. They were distinguished by differences in the number, shape, proportion and position of parts like the stem, leaf and inflorescence, but not by odour, length of life, and so on. Varieties, variation produced by sowing the seed of one individual, were often the result of cultivation and were rarely of importance.

The idea of 'character' anchors the system. The essential character was the essence of a genus in a natural system, while the factitious character defined the genus in an artificial system. Both were based on features of the fructification and, when combined, made a natural character. The description was the natural character of the whole plant, so it included the habitus. For species, natural characters were descriptions, essential characters were differentiae. Finally, a *naturalissima* character contained features common to many plants.

The correct naming (*denominatio*) of plants was critical. The name of a genus was a single word ideally reflecting either its habitus or an essential feature, and Linnaeus proposed extensive guidelines for naming, relegating unacceptable names to synonymy. The specific name was a polynomial, the essential character of the species. A mere twelve words could provide differentiae for all species in a genus of 100 species when the differentiae were qualified by adjectives and used synoptically (that is, dividing the genus successively in half: 50, 25, 13, 7, 3, 2, 1). Since essential characters were not known a priori, as new species were added to a genus the polynomials (differentiae) of the other species might have to be changed. Moreover, when a species was moved from one genus to another, its name would change if the differentiae in the new genus differed. Names were not linked to specimens, all individuals of a species sharing the same essence; Linnaeus treated specimens as being by and large interchangeable.

The trivial name was a single, usually adjectival, word which, with the generic noun, formed the binomial. (Linnaeus abbreviated book titles in a comparable fashion, and binomials were initially provided for convenience.) Few nomenclatural rules applied to trivial names, which did not reflect the essence, but their value was such that they pervaded biological literature after 1753 and are often described as ‘Linnaean names’. Linnaeus’ reform of genera and species and his nomenclature were widely influential.

2 The natural method

The natural method was ’the ultimate goal of botanists’. There genera were grouped into orders; these were named, but not described (exceptions are the early *Musa Cliffortiana* and Paul Dietrich Giseke’s descriptions in the *Praelectiones* of 1792, presumably those given by Linnaeus in lectures in 1771, but never published). Linnaeus realized that natural orders could not be defined. Even the most ’natural’, such as the *Umbelliferae*, the carrot family, lacked features that were unique to and constant within them. Until these were found, natural groups were ’like a bell without a clapper’; in modern parlance, they were polythetic.

Plants showed relationships on all sides, like territories on a map. (The diagram of relationships drawn by Giseke showed ’genealogical-geographical affinities’; too much emphasis should not be placed on the genealogical element, since trees of knowledge were described in similar terms.) However, plants still undiscovered caused problems in establishing both relationships and groups. When discussing such problems, Linnaeus made the oft-quoted statement, *natura non facit saltus* (nature does not make leaps); continuity made definition difficult. Thus although there is no simple *scala naturae* in Linnaeus’ work, relationships form a continuum. Characters do not define groups, indeed, in the chain of variation there is no real discontinuity even between animals and plants: plants have flowers, as do animals - but only those animals like *Hydra*, with its corolla-like circlet of tentacles, immediately adjacent to plants in the series. Such continuity pervades Linnaeus’ work, *Homo troglodytes*, which linked *Homo sapiens* to other animals, being another striking example.

3 Developmental interactions

Interactions between two basic tissues, the cortex and medulla, explain the development of organs, the relationship between the inside and the outside of organisms, why different characters were used at different levels of the systematic hierarchy, the general nature of organic diversity, and even ideas of disease and its cure. Organisms were made up of cortex (the male principle; responsible for form) and medulla (female; life). Form was manifest in the cortex only after a basically epigenetic interaction with the medulla. In most animals, determinate in growth, the cortex enclosed the medulla, while in most plants the medulla pushed through the cortex and growth was indeterminate; both plants and animals had definite form. Fungi, however, had only a thin layer of cortex, so their form was notably labile and their classification, which depended on constant form, consequently chaotic. Finally, in organisms like the genus *Chaos* (amoeba) itself there was neither cortex nor constant form. Linnaeus toyed with the idea of establishing a separate kingdom for such organisms.

The vegetative organs of the plant, and most of the fructification, were cortical in origin. Each organ was formed from a particular subtype of cortical tissue, thus stamens represented the innermost part of the cortex (lignum, or wood); the pistil, however, was almost pure medulla. Fertilization thus brought together male (cortical) and female (medullary) elements. If for any reason the medulla died or was used up, the plant itself died; thus in annual plants all the ’growing points’ of the medulla entered the ovaries, none remaining to continue growth.

Linnaeus equated flowering in plants with metamorphosis in insects. The young stages of the organism, the outside parts, were discarded, and the adult (insect, flower), representing the insides, emerged. Hence the different parts of the plant fructification were equated with elytra (wings) and genitalia, and the butterfly with 'a flying flower'. Metamorphosis confirmed Linnaeus' use of characters of the fructification in defining genera; they were characters of the adult. Vegetative characters, external and juvenile, were useful only as specific differentiae. (He also discussed the relationship between the animal and plant kingdoms as a form of metamorphosis in which cortex was sloughed off.)

The cortex-medulla theory had numerous implications. Linnaeus drew a parallel between the growth of the vegetative bud and that of the flower: the flower, with its five main organs, representing the equivalent of five years' vegetative growth in one. Moreover, by incorporating ideas of plant nutrition, curiosities such as flowers with only a single covering that was green on the outside and coloured on the inside, and the association of doubled flowers with varying degrees of sterility, could be explained. From observations of such phenomena, including *terata* (monstrosities), Linnaeus concluded that plant structures were interconvertible - independently of and prior to similar ideas expressed by Goethe (see Goethe, J.W. von §3). Leaves could change into sepals, sepals into petals, and petals into stamens. (Note that most observations Linnaeus used here were valueless for his classification; doubled flowers were at most varieties.)

Finally, the cortex-medulla theory is integral to ideas Linnaeus developed on the generation of organic diversity. His species concept - that all species were created by God - was soon amended. Hybridization, he thought, produced the new genus *Peloria* (in fact, a mutant of toadflax) and also occurred between other genera (in combinations that are patently impossible), yet it would not destroy the overall taxonomic framework. Since the cortex of the male determined vegetative form and hence the specific character, and the medulla of the female effectively determined features of the fructification and hence the generic character, hybrids were simply new species in established genera. Thus hybridization could account for variation in taxonomically difficult genera like *Rosa*. Later, Linnaeus came to think that God had created a small number of cortex-medulla combinations representing orders or even the three main groupings of plants (based on cotyledon number). Successive hybridizations then yielded the diversity of form as manifest in the classificatory hierarchy.

4 Functional/ecological relationships

Linnaeus held physico-theological ideas of balance and compensation in nature, with a series of controls, dependencies and connections existing within and between organisms. Bird colours were important in courtship; the flowers, fruits and leaves of plants were arranged to facilitate pollination and seed dispersal. Most insect larvae ate one species of plant, and often one part of that plant; some ate other insects; plant galls housed yet other larvae. Water in the pitchers of *Tillandsia* was for the use of thirsty humans, birds and beasts alike; evergreen trees grew in barren places to provide shelter for animals in winter. Food chains, from the minute *Monoculus* to gnat to frog to pike to seal, and ecological successions, from crustaceous lichens on a rock by the sea to foliose lichens to mosses to herbs and shrubs, formed other connections, as did geomorphological changes.

Ultimately, reproduction had to be checked, since simple calculation showed that otherwise single species would come to dominate nature. Hence small and useful, that is, edible, animals reproduced more than large and useless animals. After caterpillar larvae devastated grasslands, more plant species could grow there. Death and life were inseparably connected; nature was in balance.

Organisms were for human use, too, although in general Linnaeus did not think that some organisms existed for the simple and unrestricted use of others. However, he argued that the study of natural history allowed humans to use natural productions more effectively (his strong cameralist inclinations are relevant here), and about one in six of Linnaeus' theses deal with this theme. The final link in the theodicy was the *Nemesis divina*; God kept humans in check.

5 Conclusions

Linnaeus' ideas owe much to those of Cesalpino, and ultimately to Aristotelian thought. Logical division yields natural groupings only in a taxonomy of analysed entities, and Linnaeus realized that he did not have such knowledge of animals and plants; his claims for the results of the a priori reasoning he applied were modest.

Stamens, although important in reproduction, were not a true *fundamentum divisionis*, hence classes and orders were the work of nature and art and were only partly natural. Indeed, Aristotle had argued that logical division was an inappropriate tool for the classification of organized beings. It is difficult to relate Linnaeus' ideas of development directly to those of Aristotle, in which the male is usually the active element, shaping a passive female substance. However, Aristotle's general teleological framework bears comparison with that of Linnaeus (see Aristotle §17).

Linnaeus came to believe that species were the work of time; they were not simultaneously created - although by God's combinatorial, they were foreordained. Such developments of his ideas - let alone his interest in the origin of form and the functional interconnections within and between organisms - escaped many of his contemporaries and successors, and he came to exemplify essentialism, stasis and classification.

See also: Species; Taxonomy

P.F. STEVENS

List of works

Linnaeus, C. von (1735) *Systema naturae, sive regna tria naturae systematicae proposita per classes, ordines, genera, species* […], Leiden: T. Haak.(Broadsheet; Linnaeus' first classification of plants, animals and minerals.)

Linnaeus, C. von (1735) *Fundamenta botanica quae majorem operum prodromi instar theoriam scientiae botanica per breves aphorismos tradunt* […], Amsterdam: Schouten.(365 aphorisms explaining Linnaeus' classificatory philosophy and practice.)

Linnaeus, C. von (1751) *Philosophia botanica in qua expicantur fundamenta botanica cum definitionibus partium, exemplis terminorum, observationibus raiorum, adjectis figuris aeneis*, Stockholm: Kiesewetter.(An elaboration of Linnaeus' classificatory philosophy and practice.)

Linnaeus, C. von (1753) *Species plantarum, exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specific, nominbus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digesta*, Stockholm: Salvius.(Full descriptions and synonymy of all species of plants. Facsimile reprint, with important essays by W.T. Stearn and J.L. Heller, London: The Ray Society, 1957.)

Linnaeus, C. von (1758-9) *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*, 10th edn, Stockholm: Salvius, 2 vols.

Linnaeus, C. von (1766-88) *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*, 12th edn, Stockholm: Salvius, 3 vols.

Linnaeus, C. von (1785-90) *Amoenitates academiae seu dissertationes variae physicae, medicae botanicae anteac seorsim editae nunc collectae et auctae cum tabulis aeneis*, ed. J.C.D. von Schreber, Erlangen: J.J. Palm, 10 vols.(Theses, most of which were written by Linnaeus and defended (and the publication paid for) by his students. The first edition, Stockholm and Leipzig: G. Kiesewetter, 1749-69, 7 volumes, is less complete. The original theses should always be consulted. Some theses have been translated; L'équilibre de la nature, translated by B. Jasmin, introduction and notes by C. Limoges. The 1972 edition, Paris: Vrin, contains translations of five important theses dealing with relationships in nature.)

References and further reading

Blund, W. (1971) *The Compleat Naturalist: A Life of Linnaeus*, New York: Viking Press.(A general biography.)

Cain, A.J. (1995) 'Linnaeus's natural and artificial arrangements of plants', *Botanical Journal of the Linnean Society* 117: 73-133.(See also references; clarification of the structure of Linnaeus' classifications.)

Frängsmyr, T. (ed.) (1983) *Linnaeus, the Man and His Work*, Berkeley, CA: University of California Press.(See especially Gunnar Brobert, '*Homo sapiens*: Linnaeus' classification of man'.)

Giseke, P.D. (1792) *Praelectiones in ordines naturales plantarum*, Hamburg: Hoffman.(Notes made by Giseke and Fabricius from lectures given by Linnaeus.)

Koerner, L. (1993) 'Nature and Nation in Linnaean Travels', Harvard: Ph.D. Thesis, Department of History.(The importance of Linnaeus' cameralist ideas).

Larson, J.L. (1971) *Reason and Experience: The Representation of Natural Order in the Work of Carl von Linné*, Berkeley, CA: University of California Press.(The relationship between Linnaeus' theory and his practice.)

Stafleu, F.A. (1971) *Linnaeus and the Linnaeans: The Spreading of their Ideas in Systematic Botany, 1735-1789*, Utrecht: A. Oosthoek.(Linnaeus' ideas and their reception.)

Stafleu, F.A. and Cowan, R. (1981) *Taxonomic Literature: A Selective Guide to Botanical Publications and*

Collections, with Dates, Commentaries, and Types, vol. III, Lh-O, Utrecht: Bohn, Scheltema & Holkema. (Pages 71-111 provides an entry to the bibliographic and biographic literature on Linnaeus.)

Stevens, P.F. and Cullen, S.P. (1990) 'Linnaeus, the Cortex-Medulla Theory, and the Key to his Understanding of Plant Form and Natural Relationships', *Journal of the Arnold Arboretum* 71: 179-220.(How the cortex-medulla theory underpins much of Linnaeus' work.)

Weinstock, J. (ed.) (1985) *Contemporary Perspectives on Linnaeus*, Lanham, MD: University Press of America. (See especially articles by Wolf Lepenies and Lars Gustaffson on *Nemesis divina*.)

Lipsius, Justus (1547-1606)

Justus Lipsius was a Flemish humanist and classical philologist whose work on Tacitus and Seneca led him to give the first full, formal account of Stoicism as a philosophical system, and also to #develop Neostoicism, an influential political and moral theory. His most popular book was De constantia (On Constancy), an account of how to maintain steadfastness in the face of public evils. He loved gardens and dogs.

Justus Lipsius (Joest Lips) was born near Brussels, and lived through the Protestant revolt against Spanish Catholic rule over the Netherlands. Lipsius came from a Catholic family, and attended the Jesuit college at Cologne before going to the University of Louvain. Here he studied law, but soon turned to classical philology (though he never perfected his Greek). His first published work, *Variarum lectionum libri IV (Four Books of Mixed Readings)* (1569), was dedicated to Cardinal Granvelle, who appointed him Secretary for Latin Letters. Granvelle introduced Lipsius to scholarly circles in Rome, where he spent the years 1568-70. In 1571 he went to Vienna, and then to Jena, where he held the chair of rhetoric and history from 1572 to 1574. To hold this chair, he had to profess Lutheranism. On leaving Jena in 1574, he went to Cologne, where he married a Catholic wife, then to Louvain, where he completed his law degree in 1576. Fearing difficulties with the Spanish authorities, he next went to Leiden where he was professor of history from 1579 to 1590. Here, he had to profess Calvinism. In 1591 he made his peace with the Jesuits at Mainz and returned to the Spanish Netherlands. In 1592 he was appointed to the chair of history and Latin literature at Louvain, a post he still held at his death. He was also named historiographer to the King of Spain in 1595. Lipsius, along with his dog Mopsus, appears in a famous painting, *The Four Philosophers,* by Peter Paul Rubens, whose brother had been a pupil of Lipsius at Louvain.

Lipsius' first notable achievement was the edition (1574) of the Roman historian Tacitus, whose works had been rediscovered in the late fourteenth and fifteenth centuries. Lipsius was among those thinkers who interpreted Tacitus as counselling political prudence and non-involvement in affairs of state. In 1589, Lipsius published his own work on politics, *Politicorum sive Civilis Doctrinae libri sex (Six Books of Politics or Civil Doctrine)*. His treatment of military affairs stimulated Dutch army reforms, but the chief importance of the work lay in Lipsius' appeal to the political and moral values of Ancient Rome, and his acceptance of the central role of the state. The importance that Lipsius attached to political stability led him to advocate the enforcement of a state religion, but in the first edition he also counselled toleration of private belief. As a result, his work was placed on the Index.

Lipsius' advocacy of Neostoicism in political theory was supplemented by his Neostoicism in ethics and his work on Stoicism as a philosophical system. In his very popular book *De constantia (On Constancy)*, first published in 1584, and republished many times, both in Latin and in translation, Lipsius discussed the individual's attitude towards public troubles, advocating steadfastness guided by reason, freedom from the emotions, patience in adversity, and cheerful subjection to God. He recognized that Stoicism had to be modified for, contrary to Stoic views, God is not subject to destiny, but rather destiny is subject to God. Divine providence is not an impersonal fate, and it can be reconciled with both chance and human free choice.

In 1604 Lipsius published two books on Stoic doctrine in which he argued that Stoicism, while not wholly acceptable, was the philosophy best suited to Christianity. The first, *Manuductio ad Stoicam philosophiam*, is a handbook dealing with ethical questions, including friendship and suicide (which Lipsius was against), and with the sources for Stoicism. The second (*Physiologia stoicorum*) is a study of Stoic physics as providing an essential foundation for the study of Stoic ethics, given the injunction to lead one's life in accordance with nature. Lipsius was more positive about the compatibility of Stoic views on fate with Christian views about divine providence than he had been in *On Constancy*. He also took up Stoic views on God as a divine fire infused throughout nature, on eternal matter, and on soul as material. The book presents the first systematic modern account of Stoicism (see Stoicism).

Lipsius' last contribution to the knowledge of Stoicism was his 1605 edition of Seneca's *Opera omnia* which includes an unfinished commentary and essays. Seneca had always been important to Lipsius, not just as a thinker but as a stylistic model.

In a letter of 1603, Lipsius summed up his work in these words: 'I was the first, or the only one, of my time to turn

my scholarship to Wisdom; out of Philology I made Philosophy' (Morford: 137).

See also: Political philosophy, history of §7; Seneca; Stoicism

E.J. ASHWORTH

List of works

Lipsius, J. (1569-1606) *Opera omnia (Complete Works)*, Wesel, 1675, 4 vols.

(1569) *Variarum lectionum libri IV (Four Books of Mixed Readings)*, in *Opera omnia*, Wesel, 1675, 4 vols.

(1584) *De constantia*, trans. J. Stradling, *Two Bookes of Constancie*, ed. R. Kirk, New Brunswick, NJ: Rutgers University Press, 1939.

(1589) *Politicorum sive Civilis Doctrinae libri sex (Six Books of Politics or Civil Doctrine)*, in *Opera omnia*, Wesel, 1675, 4 vols.

(1604) *Manuductio ad Stoicam philosophiam (Handbook for Stoical Philosophy)* and *Physiologia stoicorum (Stoic Physics)*, trans. J. Lagrée, *Juste Lipse et la restauration du stoïcisme. Étude et traduction des traités stoïciens: De la constance, Manuel de philosophie stoïcienne, Physique des stoïciens (Extraits)*, Paris: Vrin, 1994. (Provides extracts of *Manuductio* and *Physiologia* in Latin with French translation.)

(1978-87) *Justi Lipsi Epistolae (Letters of Justus Lipsius)*, ed. A. Gerlo, M.A. Nauwelaerts, H.D.L. Vervliet, S. Sué and H. Peeters, Lettern en Schone Kunsten van België, Brussels: Koninklijke Akademie voor Wetenschappen.

(1996) *Principles of Letter-Writing. A Bilingual Text of Justi Lipsi Epistolica Institutio*, ed. and trans. R.V. Young and M. Thomas Hester, Carbondale, IL: Southern Illinois University Press.

References and further reading

Abel, G. (1978) *Stoizismus und frühe Neuzeit. Zur Enstehungsgeschichte modernen Denkens im Felde von Ethik und Politik (Stoicism and the Early Modern Period. On the Genesis of Modern Thought in the Areas of Ethics and Politics)*, Berlin and New York: de Gruyter.(A recent survey of Stoicism in the early modern period, covering Lipsius as well as other major figures.)

Lagrée, J. (1994) *Juste Lipse et la restauration du stoïcisme. Étude et traduction des traités stoïciens: De la constance, Manuel de philosophie stoïcienne, Physique des stoïciens (Extraits) (Justus Lipsius and the Restoration of Stoicism…)*, Paris: Vrin.(A good general account, with Latin texts and French translations of extracts from Lipsius' philosophical works.)

Morford, M. (1991) *Stoics and Neostoics. Rubens and the Circle of Lipsius*, Princeton, NJ: Princeton University Press.(Full of detailed information, especially about Lipsius' friends and students. Useful bibliography.)

Mouchel, C. (1990) *Cicéron et Sénèque dans la rhétorique de la Renaissance (Cicero and Seneca in Renaissance Rhetoric)*, Marburg: Hitzeroth. (The opposition between Cicero and Seneca as models for style.)

Oestreich, G. (1982) *Neostoicism and the Early Modern State*, ed. B. Oestreich and H.G. Koenigsberger, trans. D. McLintock, Cambridge: Cambridge University Press. (Deals with Lipsius' political thought and its influence in part 1, pages 13-131.)

Saunders, J.L. (1955) *Justus Lipsius. The Philosophy of Renaissance Stoicism*, New York: The Liberal Arts Press. (Still the standard work on Lipsius.)

Literature, Japanese philosophy in modern

Since the last quarter of the nineteenth century, virtually all major lines of Western thought and the works of both major and minor Western philosophers have been explored and used by Japanese writers in an effort to forge a modern Japanese literature. The history of translation alone reveals a concern to bring over synoptic summaries of Western philosophy, as well as the primary works of specific thinkers. Academic philosophy as a discipline of advanced study was established in the 1880s, the decade which corresponds to the beginnings of widespread literary reform and the often-cited creation of the first modern Japanese novel, Futabatei's Ukigumo (Floating Cloud) in 1889.

However, Japanese novelists, dramatists, poets and critics did not assimilate philosophical influences naïvely or passively, nor was Japanese literature made over in the shape of specific Western ideas regarding the nature and function of the self, society or literary aesthetics. Indeed, the avid translation and discussion of Western ideas frequently provoked a nativist reaction or modification. The revival of traditional tropes, the language of Confucian ethics, Buddhist practice and Shintō legends), itself often reflects the pervasive presence of Western ideas on the modern literary scene.

1 Modern literature

Few Japanese literati since the Meiji period (1868-1912) have been philosophers in an academic sense. It can be difficult to isolate specific Cartesian, Kantian, Bergsonian or Heideggerian elements in a work allegedly written under their influence. As Japanese literary critics themselves have pointed out, time and again writers responded less to an idea as such than to particular imagery, or even to the biography or the historical predicament of certain thinkers. This is especially true where Modern, Enlightenment and post-Enlightenment philosophers are concerned, that is, cultural figures who were struggling to fathom intellectual, historical and social problems analogous to those confronting Japanese writers in the late nineteenth and twentieth centuries.

The beginnings of modern literature in Japan are usually traced to the 1880s, when the literary critic Tsubouchi Shōyōand the novelist Futabatei Shimei gave voice to a new style of realism and pioneered a literary language that placed greater emphasis on speech and the expression of ordinary life. Their efforts reflect a conception of literature shaped by Western modes of literary realism (in Futabatei's case, by his study of Russian and his translations of Turgenev) and by their awareness of the ascendancy of the realist novel on the Western scale of literary values. This shift in emphasis, from literature traditionally viewed in Japan as a rhetorical medium for lyrical expression to that of a transparent reflector of things as they are, paralleled shifts and reforms being undertaken in Japanese politics, law and education. Most such reforms occurred within the prevailing atmosphere of utilitarian, Enlightenment and social Darwinist thought, as it was understood through the writings of John Stuart Mill, August Comte, Jean-Jacques Rousseau, Herbert Spencer and others.

This mid-Meiji reform of literature, making way for a realism that would transcribe modern life as it is, met with a strong idealist reaction, itself mediated by a knowledge of similar counter-Enlightenment tendencies in Western (especially German) thought and writing. Hegel and Eduard von Hartmann were known to Mori Ōgai, who had studied in Germany and based his earliest writings on that foreign experience. Contrary to the brand of literary realism propounded by Shōyō, Ōgai argued that literature exists in a world of ideas and ideals apart from the everyday, where writers and works of genius from anywhere in the world exist on an equal, if lofty, plane. Whether realist or idealist, these mid-Meiji reformulations of literature were calculated to raise Japanese literature to the perceived level of superiority achieved by Western writers. They also reflect a general anxiety over Japan's vulnerability *vis-à-vis* the West and ultimately, over Japanese identity in the modern world.

2 The self

By the turn of the century, when a confessional strain of 'I-novel' writing took shape as the dominant genre of modern Japanese literature - a dominance that prevails to our own day - Christianity as a system of Western thought found its way into literary expression. Anti-institutional, introspective, Calvinist and Quaker-inflected brands of Protestantism, most clearly exemplified in the life and work of the Christian convert and educator Uchimura Kanzo and his 'no-Church' movement, both justified and provided a rich vocabulary for the literary

emphasis on love and the self. Training at Christian mission schools, a conversion experience or some significant exposure to Christianity, was shared by many of the most significant Japanese writers of the twentieth century, including Kitamura Tōkoku, Tayama Katai, Shimazaki Tōson, Kunikida Doppo, Masamune Hakuchō, Shiga Naoya and Arishima Takeo. Apostasy often followed a declaration of faith, which only deepened the need to confess, or the theme of betrayal, within the genre of the I-novel.

The ‘discovery’ of this interior self, necessitated by Christianity, has been described in Japanese criticism as either a symptom of, or a solution to, several modern problems. The self has been seen to serve a variety of cultural purposes: to fill the void left by the collapse of an orderly, neo-Confucian world view, to have satisfied psychological demands made of writers within a rapidly modernizing nation-state, and to have provided a system of ‘transcendent’ values to rival the predominant stress on immanence within the traditional modes of Japanese thought, Confucianism, Buddhism and Shintoism. Still, in specific works of literature, the emphasis on the self or on the value of romantic love often gives way to its loss or rejection. For example, in Shiga Naoya’s famous *Anya Korō(A Dark Night’s Passing)*, the afflictions of self-consciousness yield to a new found sense of harmony within nature and the dissolution of the self within a comforting, spirit-filled landscape.

The literary expression of personal experience was further reinforced by prevailing trends in academic philosophy. Neo-Kantian thought, as well as the pragmatism of William James, was especially dominant in the Taishōperiod (1912-25), serving as a spur to Nishida Kitarō, Japan’s first and perhaps greatest modern philosopher. Nishida’s stress on ‘pure experience’ at once represents a reflection of Western idealist ethics and consciousness of self, joined to an Eastern sense of meditation praxis, as in Zen Buddhism. This conjunction of foreign and native ideas is a widely observed characteristic of the literature of this period, as seen in the eclectic use of both Western and Asian imagery in the fiction of Nagai Kafūand Tanizaki Jun’ichirō. Both exhibit traces of Western decadence, which leads to their discovery of the exotic in old Japan or in tales of the disappearance and partial recovery of tradition. This tendency was further exhibited in the work of Watsuji Tetsurō and Kuki Shūzō, influential cultural critics in the 1920s and 1930s. Close familiarity with German phenomenology in the work of Husserl and especially Heidegger led them to explore the distinctive climate or structure of Japanese culture.

The prevailing emphasis in modern Japanese literature on expressions of self, wherein general questions of social structure or cultural identity were engaged subjectively if at all, did not go unchallenged. The great critic of the time, Kobayashi Hideo, focused on the peculiar narrowness of the Japanese I-novel and of modern writers held in sway by a welter of Western ideas and influences. To recover both greater breadth and a sense of cultural balance, Kobayashi argued for the creation of fiction whose protagonists were bound to a recognizable history and society.

3 Marxism

Marxism had considerable influence on Japanese scholarship, education and literature, and dominated literary and cultural criticism in the late 1920s, giving rise to proletarian realism and to social problem plays in the ‘new theatre’ movement. Notable Marxist critics include Miyamoto Kenji, whose essay ‘Haiboku no bungaku’ (Literature of Defeat) in 1929 represented a sharp if ultimately sympathetic critique of the subjectivist fiction and suicide of Akutagawa Ryūnosuke, and Nakano Shigeharu, whose practice as a socially conscious poet and critic spanned several decades including the crisis of recantation or *tenkō*, the phenomenon of ideological reversal that was so conspicuous a feature of Japanese intellectual life in the 1930s. Socialist realism in fiction was practised by Kobayashi Takiji, who became a martyr to the proletarian cause when he was killed in prison in 1933. Nevertheless, at this time, as well as in the immediate postwar years and later amid the political turmoil of the 1960s, Marxism offered Japanese writers both a way to criticize modernity, and a coherent account of the interconnectedness of individuals and groups within society. Confucianism had formerly provided such an account, though with an emphasis on stability and continuity that seemed increasingly remote from a modern world of unceasing change (see Confucian philosophy, Japanese).

Amid the growing forces of nationalism and militarism in the 1930s and the state of total war in the 1940s, many Marxist ideologues either fell silent or subtly exchanged their leftist critique of the West and of modernity for a rightist critique of Western cultural imperialism. A renewed emphasis on the cultural differences between Japan and the West, ironically mediated by a conscious use of German romantic thought from Herder and Schiller through to Heidegger, led to a revival of things Japanese at this time, and to the literary nationalism of such critics as Yasuda Yojūrōand Hayashi Fusao, and of the poets ItōShizuo and Hagiwara Sakutarō. Their tone of lament, and

their imagery of homelessness, wandering and loss, was balanced by their advocacy of Japanese tradition. In poems or impressionistic essays, they sketched provincial journeys and encounters with simple folk as they made their ‘return’ to the country and home, Japan.

4 Postwar period

After the Second World War, during the early period of American occupation, the intellectual atmosphere was marked by the re-emergence of suppressed or apostate Marxists, and by the philosophy of existentialism (see Existentialism). Occupation censors took over from war-time military censors, and writers sought to account for their loss of personal autonomy and authentic experience, as well as for the zero-degree environment of incinerated cities and a society in confusion and collapse. This atmosphere of Sartrean nausea surrounded several leading postwar Japanese authors, contributing to the desert and labyrinth allegories of Kobo Abe, the descriptions of war and moral atrocity in Ōoka Shohei, the tales of deformed historical consciousness in Ōe Kenzaburō, the wasteland poetry of Tamura Ryūichi, the experimental mythology and apocalyptic tone of the dramatists Kinnoshita Junji, Terayama Shūji and Kara Jurō, and the literary criticism of EtōJun and Yoshimoto Takaaki. These authors were concerned, for different political motives, with issues of false consciousness, bad faith and, in the 1960s, agendas for cultural and social reform. They looked for greater freedom from the influence of Western, especially American, culture.

From the 1970s through to the mid-1990s, Japanese literature registered the changes and rifts that marked Western thought over this same period, actively contributing to that critique, uncovering yet again pockets of Japanese or Asian ‘difference’ from the West, and highlighting characteristics that identify contemporary Japan as a model of post-modernity. Continental philosophy and cultural critique, with renewed attention to the Frankfurt school (see Frankfurt School), has been widely translated and discussed in literary journals. So too have Jacques Derrida and Michel Foucault, severe critics of western metaphysics and institutional structures, whose ideas have been used to frame analogous interrogations of Japanese phonocentrism and of Japan’s own coercive social structures. Some see the underlying critique of language and of meaning within post-structuralism as sharing insights with Buddhist or Daoist thought. It has also been argued that deconstruction, for example, is everywhere on display in the Japanese literary tendency toward indeterminate structures, such as elliptical verse, open-ended narratives and non-linear, polysemic theatre (see Deconstruction). There is a sense that these recent turns in Western thought, directed as they are against stable notions of the self and of linguistic and literary meaning, are oddly congruent with the most traditional assumptions behind Japanese literary practice. In this way, critics such as Karatani Kōjin and Asada Akira have sought to situate writing in Japan within a universal crisis of meaning, while taking account of the historical pressures that have provoked both this general crisis and its inflection in Japan.

Finally, it should be said that modern Japanese literature, preoccupied as it has been with the representation of the real or the expression of the personal, reveals too an anxiety that the real may be a tissue of ideas and ideologies, and that the personal may be an illusion in an environment of incessant makeover and change. Beyond any specific influence, this may be the main legacy of Western thought to modern Japanese writing: a quickened sense for the difference between material reality and our ideas about it, as well as for the power of ideas to cross local boundaries and to define human experience in universal terms.

See also: Aesthetics; Aesthetics, Japanese; Poetry

PAUL ANDERER

References and further reading

Anderer, P. (1995) *Literature of the Lost Home: Kobayashi Hideo - Literary Criticism, 1924-39*, Stanford, CA: Stanford University Press.(The early writings of modern Japan’s foremost literary critic.)

Doak, K.M. (1994) *Dreams of Difference: The Japan Romantic School and the Crisis of Modernity*, Berkeley, CA: University of California Press.(A study of nationalist discourse in the 1930s, focusing on critics and thinkers influenced by the German Romantics and Heidegger.)

Karatani Kōjin (1994) *Origins of Modern Japanese Literature*, ed. and trans. B. de Bary, Chapel Hill, NC: Duke University Press.(A study of the impact of Christianity, Marxism and other systems of Western thought on Japanese writing since the late nineteenth century.)

Keene, D. (1984) *Dawn to the West: Japanese Literature in the Modern Era*, New York: Holt Rinehart & Winston.(A comprehensive guide to all forms of modern Japanese literature; see especially the entries under

‘Criticism’.)
Miyoshi Masao and Harootunian, H.D. (eds) (1988) ‘Postmodernism and Japan’, *South Atlantic Quarterly* 87 (3).(A special volume dealing with post-structuralist thought, theories of national and transnational identity, and contemporary debates in Japan.)
Nishida Kitarō (1911) *Zen no kenkyū(An Inquiry into the Good)*, trans. Masao Abe and C. Ives, New Haven, CN: Yale University Press, 1987.(The seminal work of modern Japan’s most distinguished philosopher.)
Rimer, J.T. (ed.) (1990) *Culture and Identity: Japanese Intellectuals during the Interwar Years*, Princeton, NJ: Princeton University Press.(Useful collection of essays on cultural criticism.)
Silverberg, M. (1990) *Changing Song: The Marxist Manifestos of Nakano Shigeharu*, Princeton, NJ: Princeton University Press.(A study of a prominent Marxist critic and the impact of Marxism on Japanese cultural life.)

Literature, philosophy in Latin American

Within the Latin American intellectual community, the relationship between philosophy and literature constitutes one of the most interesting chapters in its development. Much Latin American literature is characterized by profound philosophical concerns, focusing on the question of identity. From the time of the conquest and colonization of the American continent in the 1500s, a debate regarding the humanity of the recently discovered inhabitants began in Spain. This debate would prove to be one of the most revealing controversies of sixteenth-century Europe. At the point of colonial expansion, Europe projected a logocentric vision which would incite a unique Latin Americanist philosophical discourse relating to the question of identity.

During the nineteenth century, philosophical discourse was formulated principally through literary expression. At first the quest for a cultural identity was the philosophical focus, although two conflicting positions were evident: the desire to achieve cultural independence from Europe and a yearning for Latin America to become European. This latter position inspired the urge to identify with European culture and from the mid-1900s, with the political and economic success of the USA.

In the twentieth century, from the time of the University Reform of 1918, an academic philosophy emerged close to that of Europe and began to diversify the Latin American philosophical panorama. From the various philosophical stances which arose at that time, one that dominated the cultural arena, despite its occasional relegation to a secondary position in academia, was the urge to articulate a Latin Americanist philosophical discourse which would succeed in transcending its own frontiers through liberation philosophy, beginning in the 1960s.

1 Philosophy as literature

In the philosophical discourses of the colonial and nineteenth-century periods, two trends appear in Latin America: an academic one inspired by the European movements and practised by the Scholastics, such as Alonso de la Vera Cruz (1504-84) and José de Acosta (1540-1600), Scotist Alfonso Bricño (1590-1667), Neoplatonist José Manuel Peramás (1732-93), Enlightenment thinker Eugenio de Santa Cruz y Espejo (1747-95) and positivist Carlos Octavio Bunge (1875-1918). The other Latin Americanist trend articulated its philosophical discourse through literary expression, particularly the essay and the novel. This group of philosophers comprised Mexicans José Joaquín Fernández de Lizardi (1776-1827) and Justo Sierra (1848-1912), Venezuelan-Chilean Andrés Bello (1781-1865), Argentines Juan Bautista Alberdi (1810-84) and Domingo Faustino Sarmiento (1811-88), Chileans José Victorino Lastarria (1817-88) andFrancisco Bilbao (1823-65), Cubans José de la Luz y Caballero (1800-62) and José Martí (1853-95), Ecuadorian Juan Montalvo (1832-89) and Puerto Rican Eugenio María Hostos (1839-1903). Through the literature of this second trend of the nineteenth century emerge the first expressions of Latin American thought.

In 1815 during the fight for political independence from Spain, leader Simón Bolívar (1783-1830) presented the plight of Latin America in the following terms:

> We are neither Indian nor European, but a species midway between the legitimate proprietors of this country and the Spanish usurpers. In short, although Americans by birth we derive our rights from Europe, and we have to assert these rights against the rights of the natives, and at the same time we must defend ourselves against the invaders.
>
> ([1815] 1981: 63-4)

Bolívar only succeeded in defining Latin Americans by decribing what they were not: ‘We are neither Indian nor European’. His Americanism makes reference to his birth only in as much as, culturally, he believed himself to have inherited ‘Our rights from Europe’. It is in this way that the identity plight became a Latin American obsession.

During the 1800s Latin America was viewed by its people as a scene of confrontation between civilization and barbarism. Argentine essayist Domingo Faustino Sarmiento (1811-88) interpreted Latin America’s social conflict as two opposing forces: ‘one civilized, constitutional, European; the other barbaric, arbitrary, American’. Sarmiento felt he was dealing with ‘two distinct societies that were rivals and incompatible’ (1845: 58). The significance of the thought exemplified in the work of Sarmiento is manifested more precisely in novels by such

authors as Ecuadorian Juan León Mera (1832-94). In Mera's *Cumandá o un drama entre salvajes (Cumandá or Drama Among Savages)* (1879) Ecuador's inhabitants separate themselves into two large groups: the foreigners and the natives. The first comprises the 'whites' (of European descent and identified as Christians) and the other group is made up of Indians or Mestizos (of mixed blood). In Mera's work Ecuador does not have a social identity. It is seen as a border of conflict which is defined, in Sarmiento's terms, as the place of confrontation between civilization (European culture) and barbarism (what is perceived as the autochthonous, native culture).

Towards the end of the 1800s the optimism which characterized the intellectuals of independence turned to pessimism. Like the initial liberalism, the conservative positions under the positivist motto *Ordem e Progresso*, or order and progress, also had failed politically, socially and economically. However, among the pessimistic voices of, for example, Bolivian Alcides Arguedas (1879-1946), a group of essayists emerged and looked to the past in order to present a new stance on the concept of identity. Peruvian Manuel González Prada (1848-1918), Uruguayan José Enrique Rodó (1871-1917), Mexican Justo Sierra (1848-1912), Cuban José Martí (1853-95) and Puerto Rican Eugenio María Hostos (1839-1903) are notable figures in the development of this new philosophy. During this period, it was recognized that large sections of the population had been and were being ignored and that, if Latin America was going to save itself, it must save its Indians. In his essay 'Our America' (*c*.1870), Martí declared, 'To know is to resolve. To know one's country and to govern it with that knowledge is the only way of liberating it' (*c*.1870).

2 Philosophy in literature

The independence in 1898 of the last Spanish colonies in America (Cuba, Puerto Rico), the construction of the Panama Canal, the intervention of the USA in the affairs of several Latin American countries and the First World War motivated Latin American philosophers to project their national concerns across the continent. The Latin Americanist philosophers emerged with the intention of reflecting on the problem of identity on a continental scale. Counted among these were Colombian Baldomero Sanín Cano (1861-1957), Argentines Alejandro Korn (1860-1936) and Francisco Romero (1891-1962), Uruguayan Carlos Vaz Ferreira (1873-1958), Mexicans Antonio Caso (1883-1946) andSamuel Ramos (1897-1959), Bolivian Guillermo Francovich (1901-90) and Peruvian José Carlos Mariátegui (1895-1930). The University Reform initiated in Argentina in 1918 extended rapidly throughout Latin America and gave rise to the development of an academic philosophy. From the beginning of the 1930s these philosophers felt the need to recover the Latin American cultural past as a preliminary step to the formulation of an original philosophy.

In the area of literature the concern was more personal and immediate. The problems which arose from an awareness of one's own circumstances were raised in poetry, theatre, the novel and the essay. Latin American reality was analysed and deconstructed and solutions were proposed. In short, this literature dealt with thoughts which were artistically expressed and parallel with, and sometimes even crossing over into, academic philosophy. However, this literature did not concern itself with contextualizing philosophical discourse. Hence, visions of a possible utopian development in Latin America were expressed by authors such as the Uruguayan Rodó (1871-1917) and Mexican Vasconcelos. In his work Rodó defended the integral personality of the person as well as the concept of select minorities. He criticized what he considered to be the levelling propensity of democracies and the excessive concern for materialism and mercantilism in the USA. He proposed that the Latin American 'spirit' be their objective. Against the degradation levelled at the Mestizo people by the positivists, Vasconcelos recognized the Mestizo essentiality, that is, the interracial and intercultural nature of Latin American people, which he perceived to be the future of humanity. According to Vasconcelos (1925), in Latin America, a fifth race, the cosmic race (a mixed race incorporating the creative power of the other races) would emerge, called to guide the destinies of humankind.

The novel during the period up to the 1950s surpassed the philosophical debates surrounding the concepts of civilization and barbarism which prevailed during the nineteenth century. The works by Mexican Mariano Azuela (1873-1952), Argentine Ricardo Güiraldes (1886-1969) and Venezuelan Rómulo Gallegos (1884-1969) were powerful artistic creations which contributed to a profound philosophical reflection in which the cultural past and the problem of Latin American identity were embraced and deconstructed. Gallegos in *Doña Bárbara* (1929) studied the intuitions of Martí and the vision of Vasconcelos. In this novel the problems of Latin America continued to be defined as a confrontation between civilization and barbarism. Imitation was considered barbarism

as were the imposition of alien laws upon one's own reality and egoism in search of rapid enrichment. For Gallegos, the civilizing force is that which trusts in the potential of national powers and which recognizes racial and cultural miscegenation as the basis of national identity.

The novel, particularly the *indigenista* novel, became an instrument of denunciation concerning the state of the pre-Columbian descendants' population. The indigenous population gradually recognized through its exclusion, one of the basic causes of the economic, social and political underdevelopment in which some of the Latin American countries found themselves. Peruvian César Vallejo (1892-1938) and Ecuadorian Jorge Icaza (1906-78) crudely developed the consequences of a culture of oppression: one which paralysed the internal development of a country, accentuated the weakness of its oligarchy and accepted as natural imperialist impositions from the exterior. Peruvians Ciro Alegría (1909-67) and José María Arguedas (1911-69) sought to recover the excluded populations.

The question of identity continued to be the impetus of philosophical reflection within literature. The work of Argentines Eduardo Mallea (1903-82) and Ezequiel Martínez Estrada (1895-64) contained profound philosophical reflections which had repercussions that transcended national boundaries. In countries such as Mexico, where the identity crisis was precipitated by internal circumstances, authors and philosophers, such as Samuel Ramos (1897-1959), Rodolfo Usigli (1905-79), Octavio Paz (1914-) and Carlos Fuentes (1929-), shared a mutual concern.

3 Contextualization of philosophy in literature

In the early 1950s Latin American philosophers began to formulate original thought. *América como conciencia (America as a Collective Awareness)* (1953) by Mexican Leopoldo Zea, exemplified this initial stage which reached maturity in the 1960s. For the first time Latin American thought broke its own boundaries by means of 'liberation thought'. In the 1960s such thought split into three complementary fields: pedagogy, exemplified by Brazilian Paulo Freire (1921-), theology, demonstrated in the work of Peruvian Gustavo Gutiérrez (1928-) and philosophy, reflected in the work of Leopoldo Zea (1912-). All these thinkers analysed and deconstructed Western modernity. Their works emerged at the same time as the postmodern discourses of central Europe, although contrary to postmodernism, they were formulated in an attempt to attain a new, global utopian view of humanity.

Liberation thought in literature had immediate repercussions. Its theoretical originality and rigour stimulated writers to deconstruct and contextualize philosophical discourse as well as think about its theological, pedagogical and socioeconomic consequences. In terms of artistic expression, in addition to the liberation thought which guided them, Latin American writers and central European postmodern theorists found inspiration in the deconstruction of modernity by Argentinian Jorge Luis Borges (1899-1986) and Colombian Gabriel García Márquez (1928-), or the liberation works of Salvadorean Manlio Argueta (1936-) and Nicaraguan Claribel Alegría (1924-).

The context of economic thought was also formulated in Chile by Brazilians Fernando Henrique Cardoso and Enzo Faletto and analysed and deconstructed by Colombian Manuel Mejía Vallejo. Liberation theology is given some context in works such as *El ocaso de Orión (The Setting of Orion)* (1972) by Bolivian Oscar Uzín Ferández and *La cruz invertida (The Cross Upside Down)* (1970) by Argentine Marcos Aguinis. Aguinis's novel, written at the same time as Gustavo Gutiérrez's seminal work on liberation theology (1971), deals with the theological implications of the concept of the 'new man', providing for the poor and tackling the issue of celibacy, as well as analysing the Eurocentrist discourse of the Catholic church and its power structures. Its content prophesied the Vatican's subsequent confrontation with liberation theologians.

Theorists practising liberation thought did not concern themselves directly with the condition of women in Latin America, however, the deconstruction of oppression and the plan for liberation implicit in the new man concept provoked the creation of powerful femininist movements. These movements saw the struggle for women's liberation as intimately linked to that struggle for the liberation of humankind. A new literary model known as the testimonial narrative grew up at this time, pertaining to works by such authors as Mexican Elena Poniatowska (1933-) and Elizabeth Burgos's discussion of Rigoberta Menchú (1983).

The process of contextualizing liberation thought in literature is exemplified, along with more traditional themes, in the novel *Porqué se fueron las garzas (Why the Herons Left)* (1980) by Ecuadorian Gustavo Alfredo Jácome. Jácome's novel also deconstructs the trends of the so-called intercultural philosophy. This philosophy assumes

intercultural relations can be established through the media of literature and philosophy to create a dialogue among equals.The author breaks with the traditional indigenous novel by presenting an Indian with a doctoral education as rector of a school. Jácome goes beyond the search for identity by showing that it is more likely obtained when issues of racial and cultural discrimination are not constantly under examination. The protagonist creatively transcends the oppression that is assumed when one's discourse is expressed in a European language (Spanish). The novel exposes the concealed oppression that carries with it the aspiration in the 1990s of establishing an intercultural dialogue. *Porqué se fueron las garzas (Why the Herons Left)* shows that the most recent state of globalization has established an intercultural dialogue which has a hierarchical diagram and a tacit schema of oppression at its base.

See also: Cultural identity; Feminist thought in Latin America; Liberation philosophy; Liberation theology

JOSÉ LUIS GÓMEZ-MARTÍNEZ

References and further readings

Aguinis, M. (1970) *La cruz invertida (The Cross Upside Down)*, Barcelona: Editorial Planeta, 1983.(The best Argentine novel on the conflict between the priests promoting theology of liberation and the Church and state.)

Alegría, C. (1941) *El Mundo es ancho y ajeno (Broad and Alien is the World)*, Madrid: Alianza Editorial, 1983. (Indigenist novel that seeks to recover the excluded Indian population of Peru.)

Arguedas, A. (1909) *Pueblo enfermo: contibución a la psicología de los pueblos hispano-americanos (Ailing People)*, Barcelona: Vda. de Luis Tasso, 1958.(An essay on the psychology of the Bolivian people from a positivistic perspective.)

Arguedas, J.M. (1958) *Deep Rivers*, trans. F.H. Barraclough, intro. J.V. Murra, Austin, TX: University of Texas Press, 1978.(A classic indigenist novel of Peru from the point of view of a child.)

Arguedas, J.M. (1964) *Todas las sangres (All the Races)*, Madrid: Alianza Editorial, 1980.(An indigenist novel depicting the emerging voices of the indigenous population.)

Azuela, M. (1916) *Los de abajo (The Underdogs)*, Mexico: Siglo Veintiuno Editores, 1970.(A first major novel of the Mexican Revolution.)

Berryman, P. (1987) *Liberation Theology*, New York: Pantheon.(A clearly written introduction to theology of liberation; it traces the origins, spread and impact.)

Bolívar, S. (1815) *The Jamaica Letter*, in M. Pérez Vila (ed.) *Simon Bolívar: His Basic Thoughts*, Caracas: Presidency of the Republic of Venezuela, 1981.(An essay reflecting on the fight for independence and on the future of the Spanish colonies.)

Burgos, E. (1983) *Me Llamo Rigoberta Menchúy así me nació la conciencia (Rigoberta Menchú: an Indian Woman in Guatemala)*, Mexico: Siglo Veintiuno Editores, 1993.(The most widely known work of the so-called testimonial literature.)

Cardoso, F.H. and Faletto, E. (1969) *Dependencia y desarrollo en América Latina (Dependency and Development in Latin America)*, Mexico: Siglo Veintiuno Editores, 1969.(The need for a social, cultural and political contextualization of the different theoretical systems of economics.)

Freire, P. (1971) *Pedagogía del oprimido (Pedagogy of the Oppressed)*, Mexico: Siglo Veintiuno Editores. (Examines the contemporary structure of education as a structure of oppression.)

Fuentes, C. (1962) *La muerte de Artemio Cruz (Death of Artemio Cruz)*, Mexico: Fondo de Cultura Económica, 1962.(A novel of the Mexican Revolution that examines its social ramifications.)

Gallegos, R. (1929) *Doña Bárbara*, Madrid: Espasa-Calpe, 1990.(A twentieth-century novel discussing the Venezuelan view of the nineteenth-century conflict betweeen civilization and barbarism.)

Gómez-Martínez, J.L. (1993) 'Discurso narrativo y pensamiento de la liberación: *La cruz invertida* en la contextualización de una época' (Literary Discourse and Liberation Thought: *The Cross Upside Down* in the Contextualization of a Period), *El ensayo en nuestra América (The Essay in Our America)*, Mexico: Universidad Nacional Autónoma de México.(A study of the novels cited in §3.)

Gracia, J.J.E. and Camurati, M. (eds) (1989) *Philosophy and Literature in Latin America: A Critical Assessment of the Current Situation*, Albany, NY: State University of New York Press.(An introduction to the philosophy in Latin American literature, written for the nonspecialist.)

Güiraldes, R. (1926) *Don Segundo Sombra*, Buenos Aires: Losada, 1967.(A novel that idealizes the Argentine Gaucho tradition.)

Gutiérrez, G. (1971) *Teología de la liberación: Perspectivas (A Theology of Liberation: History, Politics and Salvation)*, Salamanca: Sígueme, 1990.(A theological work in the social, economic and political context of Latin America.)

Icaza, J. (1934) *Huasipungo (The Villagers)*, Buenos Aires: Losada, 1973.(Indigenist novel that explores the consequences of cultural oppression in Ecuador.)

Jácome, G.A. (1980) *Porqué se fueron las garzas (Why the Herons Left)*, Barcelona: Seix Barral.(A novel that deconstructs the trends of the so-called intercultural philosophy.)

Mallea, E. (1937) *Historia de una pasión argentina (History of an Argentine Passion)*, Madrid: Espasa-Calpe, 1969.(A literary blend of the rhetoric of the essay and that of the novel, in search of an Argentine identity.)

Martí, J. (*c*.1870) 'Our America', in *The America of José Martí: Selected Writings*, New York: Minerva Press, 1968.(An essay reflecting on the intellectual history of Latin America and its social and political consequences.)

Martínez Estrada, E. (1933) *Radiografía de la pampa (X-ray of the Pampa)*, Buenos Aires: Losada, 1976.(An essay on Argentine identity through one of its regions.)

Mejía Vallejo, M. (1972) *Al pie de la ciudad (At the Foot of the City)*, Barcelona: Destino.(A novel that contextualizes the impact of the economic development of the 1960s in the cities of Colombia.)

Mera, J.L. (1879) *Cumandá o un drama entre salvajes (Cumandá or Drama Among Savages)*, Madrid: Espasa-Calpe, 1967.(A novel that views the indigenous people of Ecuador from the European perspective of Romanticism.)

Muñoz, B. (1982) *Sons of the Wind: The Search for Identity in Spanish American Indian Literature*, New Brunswick, NJ: Rutgers University Press.(The *indigenista* novel as a social indicator, which interprets key issues of Latin American cultural development.)

Oliver, A.A. (1993) 'Values in Modern Mexican Thought', *The Journal of Value Inquiry* 27: 215-30.(A solid introduction to the philosophy of Zea in the context of Mexican culture.)

Paz, O. (1950) *El laberinto de la soledad (The Labyrinth of Solitude: Life and Thought in Mexico)*, Mexico: Fondo de la Cultura Económica, 1963.(A classic essay.)

Poniatowska, E. (1969) *Hasta no verte, Jesús mío (See You Never my Jesus)*, Mexico: Ediciones Era.(A testimonial novel depicting the social problems of the 1960s in Mexico.)

Ramos, S. (1934) *El perfil del hombre y la cultura en México (Profile of Man and Culture in Mexico)*, Madrid: Espasa-Calpe, 1951.(A philosophical essay on Mexican culture and its people.)

Sarmiento, D.F. (1845) *Facundo: Civilización y barbarie*, Madrid: Espasa-Calpe, 1967; *Civilization and Barbarism: Life in the Argentine Republic in the Days of the Tyrants*, New York: Collier-Macmillan, 1961.(A classic positivist approach to the social situation in Argentina.)

Schutte, O. (1988-9) 'Philosophy and Feminism in Latin America: Perspectives on Gender Identity and Culture', *The Philosophical Forum* 20: 62-84.(An informative essay.)

Schutte, O. (1993) *Cultural Identity and Social Liberation in Latin American Thought*, Albany, NY: State University of New York Press.(A philosophical approach with an interdisciplinary context. It focuses on cultural identity, liberation and feminist thought and includes a bibliography.)

Stabb, M.S. (1967) *In Quest of Identity: Patterns in the Spanish American Essay of Ideas, 1890-1960*, Chapel Hill, NC: University of North Carolina Press.(An expression of Latin American thought. An extensive bibliography is provided.)

Uzín Ferández, O. (1972) *El ocaso de Orión (The Setting of Orion)*, La Paz: Los Amigos del Libro, 1982.(A novel reflecting on the new social awareness of the priests and traditions of the Catholic church.)

Vallejo, C. (1931) *El tungsteno (Tungsten: a Novel)*, Barcelona: Laia, 1981.(Indigenist novel that explores the consequences of cultural oppression in Peru.)

Vasconcelos, J. (1925) *La raza cósmica (The Cosmic Race)*, Madrid: Aguilar, 1966.(An essay on the importance of the intercultural and interracial nature of Latin American people.)

Zea, L. (1953) *América como conciencia (America as a Collective Awareness)*, Mexico: Universidad Nacional Autónoma de México, 1972.(A seminal work of Latin American philosophy.)

Zea, L. (1969) *La filosofía latinoamericana como filosofía sin más (Latin American Philosophy as 'Philosophy')*, Mexico: Siglo Veintiuno Editores, 1980.(Zea argues for a Latin American philosophy independent from European philosophy.)

Llewellyn, Karl Nickerson (1893-1962)

Karl Llewellyn was philosophically the most original of the 'American Realist' jurists. His line of argument, sometimes misleadingly called 'rule-scepticism', casts doubt on received approaches to the formulation of legal rules and traditional assumptions about the part they play in law. His approach to law is an essentially functionalist one, owing much to philosophical pragmatism.

Llewellyn was an outstanding US legal thinker, innovative both in legal theory and in approach to legal education and practice. He played a major part in drafting the Uniform Commercial Code (UCC) for the USA. He emphasized the need to appreciate the fact-situations of legal cases in their full lawyerly detail, and the need to reflect on them from sociological and anthropological standpoints, rather than simply tackling them in terms of an oversimplifying prescriptivist jurisprudence.

The target of Llewellyn's jurisprudential critique was the tradition of legal study that takes the task to be that of ascertaining, ordering and transcribing a rather abstract body of rules and principles; this he saw as connecting with a model of legal processes that deems them a straightforward application of ascertained rules and principles to fact-situations. These approaches are characteristic of 'legal formalism', which the realists, a turbulent and critical younger generation of US lawyers in the 1920s, took to be the dominant view in law schools and indeed to be something of an official legal ideology in the country at large (see Legal realism §2).

Llewellyn's intellectual mission was to show that the actual practice of the law, especially in common law systems attaching great weight to precedent, cannot be matched to or helpfully explained by the formalist model. Much more goes into constructing a case as a lawyer than adding together a formulaic rule-statement and an abstract statement of the facts of the case. Precedents do not yield simple rules; rather they have to be subjected to painstaking analysis and synthesis, and looked at in groups by judges, advocates and scholars for their several purposes. There are contexts in which it may be helpful to test the results of this process by viewing them as if they were predictions of judicial conduct rather than prescriptions. The point is that a judge's role is to act 'reasonably' about disputes. This by no means licenses disregard for statutes and precedents, but also brings into play more intuitionistic elements described by Llewellyn as 'situation sense', which requires sensitivity to political and economic realities (see Legal reasoning and interpretation §§1-3).

Behind this lies a functionalist stance in the philosophy and sociology of law, well expressed in his 'law-jobs' theory, where he analyses legal roles and legal rules in terms of jobs to be performed in society. With E.A. Hoebel, he applied this theory to anthropological study of the traditional dispute-settlement techniques of the Cheyenne. It underlies all his later work, and yet was never fully developed as a comprehensive philosophy of law, partly because of the great demands on his time taken by first drafting and then (even harder) securing acceptance for the UCC, in which his approach to drafting exemplifies his theory. His last great work, *The Common Law Tradition: Deciding Appeals* (1960), extols what is in effect a self-conscious functionalism in adjudication, the 'grand style' of judging versus the rival 'formal style'. This is an eloquent plea for acknowledgement of the rule of law as a reconciliation between legal certainty, or 'reckonability', and legal adaptiveness to changing social and economic circumstances.

See also: Critical Legal Studies §2; Frank, J.; Fuller, L.L.; Law, philosophy of; Pound, R.; Social theory and law

NEIL MacCORMICK WILLIAM TWINING

List of works

Llewellyn, K.N. (1933) *The Case-Law System in America*, trans. M. Ansaldi, ed. P. Gewirtz, Chicago, IL: Chicago University Press, 1989.(An early work that reminds us of Llewellyn's strong pre-Nazi German connections; the 1989 edition is a translation of his explanation of common law methods to a German audience first published in Germany in 1933 on the basis of lectures given in Leipzig in 1928-9.)

Llewellyn, K.N. (1940) 'The Normative, the Legal, and the Law-Jobs', *Yale Law Journal* 49: 1355-1400.(The original and best exposition of the law-jobs theory.)

Llewellyn, K.N. (1960) *The Bramble Bush*, New York: Oceana Publications.(An exuberant guide to legal study for first-year law students, containing some of the notorious overstatements often used out of context to

discredit legal realism.)
Llewellyn, K.N. (1960) *The Common Law Tradition: Deciding Appeals*, Boston, MA and Toronto, Ont.: Little, Brown.(The work in which Llewellyn responded to the unease about law generated by some responses to realism, and expounded his view of the 'Grand Style'.)
(1962) *Jurisprudence: Realism in Theory and Practice*, Chicago, IL: Chicago University Press.(A valuable, though not comprehensive, collection of Llewellyn's essays and papers on legal theory.)
Llewellyn, K.N. and Hoebel, E.A (1941) *The Cheyenne Way*, Norman, OK: Oklahoma University Press.(The application of the law-jobs theory to legal-anthropological investigations among the Cheyenne.)

References and further reading

Twining, W. (1985) *Karl Llewellyn and the Realist Movement*, London: Weidenfeld and Nicolson.(An outstanding work of intellectual biography, covering all aspects of Llewellyn's life and work in a sympathetically critical way, and with comprehensive references and bibliography.)

Llull, Ramon (1232-1316)

One of the most extraordinary figures of thirteenth-century Europe, Llull was a self-taught lay theologian and philosopher, chiefly concerned with reforming Christian society and converting unbelievers. Details of his life remain obscure, but over 200 of his writings survive. Most of these expound his personal dialectical system, the Great Universal Art of Finding Truth, an encyclopedic collation of commonplace doctrines that attempts to show how all human knowledge conforms to divine truth. Largely ignored during Llull's lifetime and denounced as heretical in the later Middle Ages, the Great Art became very popular in the Renaissance as a programme of universal knowledge.

1 Career

Ramon Llull came from a prosperous Catalan merchant family that had helped conquer his native Majorca from the Arabs in 1230. Little information exists about his life: several letters document his contacts with civil or ecclesiastical authorities, and his writings include some autobiographical details. Most information comes from a brief *Vita (Life)* composed in 1311 by admirers at Paris. Llull pursued secular affairs (perhaps as a royal official) until about 1263, when visions of Christ moved him to serve God instead. He dedicated himself to three goals: writing a book to refute 'the errors of the infidels', establishing schools to train Christian evangelists in Oriental languages and converting all unbelievers. He spent ten years in private study and contemplation, acquiring a Muslim slave to teach him Arabic and perhaps receiving some formal education from local religious houses. Eventually he conceived the plan for a Great Universal Art of Finding Truth, which he attributed to divine inspiration. Around 1275 Prince James of Majorca summoned Llull to Montpellier so that church authorities could examine his writings. After this he travelled constantly, presenting his plans at the courts of Aragon, France, and the papacy, visiting the University of Paris, attending meetings of religious orders and church councils, and making private missionary trips to North Africa and the Eastern Mediterranean. Llull's greatest success came apparently in 1311 at the Council of Vienne, which created chairs of Oriental languages at several major universities. However, he often complains that his contemporaries ignore him, and depicts himself as a lone 'fool'. Llull wrote some 250 works in Catalan, Latin and Arabic. None of the Arabic works has survived.

2 The Great Universal Art of Finding Truth

Nearly all Llull's writings expound his Great Universal Art of Finding Truth, which broadly constitutes a comprehensive scheme for achieving the ancient Christian ideal that Bonaventure called 'retracing the arts and sciences to theology'. Llull's Art synthesizes the methods of contemplation and missionary argument that he evidently absorbed during his decade of private study. Though often compared to scholastic works such as Aquinas' *Summa theologiae*, Llull's system is in fact more popularizing. He often recommends it as a simpler alternative to the complex university curricula of his day, which he characterizes as abandoning the quest for divine truth in favour of quibbling over human errors. Llull especially criticized the schoolmen of Paris as 'Averroists' (see Averroism). His spiritual and intellectual values were thus basically conservative: he sought to maintain the subordination of philosophy to theology and the primacy of cloister over classroom in Western Christian culture.

Llull's Great Art tries to show how the many truths of human learning all depend on one divine truth. It accomplishes this by tracing the manifestation of certain 'divine dignities' (attributes of the Godhead) in every entity. The mature versions of his system specify nine divine dignities (goodness, greatness, eternity, power, wisdom, will, virtue, truth and glory) as 'absolute principles' of all being and knowledge. Nine more 'relative principles' (difference, concord, contrariety, beginning, middle, end, superiority, equality and inferiority) explain the diffusion and operation of the absolute principles in various 'subjects' of existence (God, angels, heavens, humans, imagination, senses, elements, vegetal power, and skills and arts). Finally, nine 'rules' (whether? what? why? from what? what kind? how much? where? when? and how?) function as heuristic questions for guiding inquiry about the principles and subjects. Llull usually creates lists of other categories necessary for applying the Great Art to particular fields of inquiry.

The most remarkable feature of his system is its use of the letters B, C, D, E, F, G, H, I and K to symbolize each set of nine principles or rules (the letter A is reserved to symbolize the coincidence of all absolute principles in

God). Llull’s handling of these symbolic terms owes little to the propositional or syllogistic structures of Aristotelian logic. Instead, he employs circular or tabular charts (called ‘figures’) to show all possible combinations of two or three letters, in the manner of mathematical multiplication tables. These dual and triple letter combinations are intended to comprehend all possible philosophical and theological propositions. Llull develops the combinations into discursive statements and analyses their ‘truth’ by applying a wide range of analogical, allegorical, proportional and figural arguments. He insists that these arguments constitute ‘necessary reasons’ capable of ‘demonstrating’ the truth of Christian belief. Successfully using the combinatory figures and necessary reasons of Llull’s Great Art depends almost completely on a correct understanding of the meanings assigned to the letters B through K. This understanding in turn invariably requires extensive interpretative work, guided by unswerving fidelity to Christian dogma.

Llull almost never cites any authorities but his own writings and routinely recasts all theological or philosophical questions in the vocabulary of his Great Art. Consequently, the sources for his work remain very difficult to identify. Llull’s pre-eminent concern for evangelizing probably motivated his preference for commonplace doctrines (such as the divine attributes) that would be comprehensible to Christians, Jews and Muslims alike. Some scholars have sought global precedents for Llull’s arguments in a single Hebrew, Latin or Arab authority, such as the *Zohar*, Bonaventure or Algazel (see Al-Ghazali). However, his erudition rarely exceeds the level found in widely-circulated Latin or Arab theological and philosophical compendia.

His combinatory methods and necessary reasons do generate some novel propositions. For example, his metaphysics relies heavily on broadly Neoplatonic principles of the emanation, participation and resemblance of all beings (see Neoplatonism). He advocates the reality of all universals and attributes both matter and form to all things, except God. Yet he does not functionally distinguish substantial from accidental forms, and rarely appeals to Aristotelian concepts of causality. Instead he refers all real or mental relationships to the ‘coessential implication’ of the divine dignities in the Godhead, and ascribes to every being three ‘innate correlatives’ of ‘act’, ‘activity’ and ‘passivity’ as a reflection of the divine Trinity. In psychology, he follows well-known Augustinian, Aristotelian or Avicennan doctrines, but also treats imagination as a separate level in the soul and classifies speech as a sixth sense (because it serves the exalted purpose of comprehending God). In logic, his doctrines largely correspond to the Scholastic ‘old logic’. However, he simplifies sophistics by proposing a ‘master fallacy of equivocation’, arguing that this explains all the other fallacies identified by Aristotle as failures to distinguish the apparent sense from the true sense (that is, the Christian interpretation) of disputed propositions, such as ‘The world is eternal’ (see Eternity of the world, medieval views of). He also recommends a new method of demonstration *per aequiparantiam* that will show ‘coessential’ relationships. In ethics, he extends a concept of moral finality to all beings, insisting that every being possesses the dual ‘intention’ of either serving God or contributing to this service. Ultimately, Llull’s arguments make most sense if regarded as a comprehensive natural theology: his Great Art is arguably the Christian Middle Ages’ most ambitious attempt to explain how all creation manifests its Creator (see Natural theology).

3 Llull’s legacy

In the fourteenth century, Llull’s work came to attract non-academic lay people and clergy who valued its spiritual ideals, although the Aragonese Inquisition challenged their orthodoxy. Nicholas of Cusa was the first of several important thinkers whose interest in Llull made the Great Art widely known in the Renaissance and early modern era. Authorities from Giordano Bruno to Leibniz praised its pansophistic method, and schools of Lullism flourished on Majorca into the eighteenth century. An extensive body of modern non-scholarly literature on Llull exists; only a few of the best critical or historical studies are in English.

See also: Bonaventure; Natural theology; Nicholas of Cusa

MARK D. JOHNSTON

List of works

Llull, Ramon (1263-1316) *Opera Latina*, Palma de Mallorca: Maioricensis Schola Lullistica del CSIC, 1959-67 and Turnhout: Brepols, 1978-, 18 vols to date; *Obres de Ramon Lull*, ed. M. Obrador y Benassar, Salvador Galmés *et al.*, Palma de Mallorca: Comissió Editora Lulliana, 1906-17 and Diputació Provincial de Balears and Institut d’Estudis Catalans, 1923-50, 21 vols.(The former is the standard modern critical edition of Llull’s Latin

texts, while the latter is the most comprehensive edition of the Catalan texts. Many Latin texts not printed elsewhere can be found in *Opera omnia*, ed. I. Salzinger, Mainz, 1721-40, 9 vols; repr. Frankfurt: Minerva, 1965. Other Catalan editions can be found in *Nova edició de les obres de Ramon Llull*, Palma de Mallorca: Patronat Ramon Llull, 1990-, 2 vols to date, and *Obres essencials*, ed. M. Batllori *et al*., Barcelona: Selecta, 1957-61, 2 vols, the latter being a well-prepared omnibus of Catalan texts. For English editions, see *Selected Works of Ramon Llull*, trans. A. Bonner, Princeton, NJ: Princeton University Press, 1985, an excellent English translation of selected texts with good introduction. *Doctor Illuminatus: A Ramon Llull Reader*, trans. A. Bonner and E. Bonner, Princeton, NJ: Princeton University Press, 1993, is an abridgement of *Selected Works*, but with updated introductory material.)

References and further reading

Anonymous (*c*.1311) *Vida de Ramon Llull: les fonts escrites i la iconografia coetànies*, ed. M. Batllori and J.N. Hillgarth, Barcelona: Associació de Bibliòfils de Barcelona, 1982.(Latin and Catalan texts, with extensive notes in Catalan.)

Brummer, R. (1976) *Bibliographia Lulliana: Ramon-Llull-Schrifttum 1870-1973*, Hildesheim: H. Gerstenberg. (Complete bibliography of scholarship on Llull.)

Garcías Palou, S. (1981) *Ramon Llull y el Islam*, Palma de Mallorca: Gráficas Planisi.(Comprehensive analysis of Llull's relations with Islamic world.)

Hillgarth, J.N. (1971) *Ramon Lull and Lullism in Fourteenth-Century France*, Oxford: Clarendon Press.(Best critical study in English of Llull and his times.)

Johnston, M.D. (1990) '*Affatus*: Natural Science as Moral Theology', *Estudios Lulianos* 30: 3-30, 139-59. (Explains Llull's proposal of speech as a sixth sense and theories of spiritual psychology.)

Johnston, M.D. (1995) *The Evangelical Rhetoric of Ramon Llull*, New York: Oxford University Press. (Comprehensive study of Llull's preaching and literary doctrines.)

Johnston, M.D. (1981) 'The Reception of the Lullian *Art*, 1450-1530', *Sixteenth Century Journal* 12: 31-48. (Reviews rise of Renaissance enthusiasm for Llull's work.)

Johnston, M.D. (1987) *The Spiritual Logic of Ramon Llull*, Oxford: Clarendon Press.(Detailed analysis of Llull's logical doctrines.)

Moreno Rodríguez, F. (1982) *La lucha de Ramon Llull contra el averroismo entre 1.309 y 1.311*, Madrid: Universidad Computense de Madrid.(Reviews Llull's polemic against the 'Averroist' scholars of Paris.)

Oliver, A. (1965-9) 'El Beato Ramón Llull en sus relaciones con la escuela franciscana de los siglos XIII-XIV', *Estudios Lulianos* 9: 55-70, 145-65; 10: 49-56; 11: 89-119; 13: 51-65.(Best account of Llull's relations with Franciscan thought.)

Platzeck, E.-W. (1962-4) *Raimund Lull. Sein Leben-Seine Werke. Die Grundlagen seines Denkens (Prinzipienlehre)*, 2 vols, Rome: Editiones Franciscanae and Düsseldorf: Verlag L. Schwann.(Major comprehensive study of Llull's life and work.)

Pring-Mill, R.D.F. (1973) 'The Analogical Structure of the Lullian Art', in *Islamic Philosophy and the Classical Tradition: Essays Presented to Richard Walzer*, Columbia, SC: University of South Carolina Press, 315-26. (Excellent brief explanation of Llull's methods of argumentation.)

Salleras i Carolà, M. (1986) 'Bibliografia lul.liana (1974-1984)', *Randa* 19: 153-98.(Useful update to Brummer's *Bibliographia Lulliana*.)

Studia Lulliana (1957-).(Formerly titled *Estudios Lulianos*, this journal publishes and reviews the latest specialized scholarship on Llull and his milieu.)

Urvoy, D. (1980) *Penser l'islam: les présupposés islamiques de l''Art' de Lull*, Paris: Vrin.(Most detailed and reliable account of Llull's knowledge of Islam.)

Yates, F.A. (1954) 'The Art of Ramon Lull', *Journal of the Warburg and Courtauld Institute* 17: 115-73.(Remains a seminal study of Llull's system.)

Locke, John (1632-1704)

John Locke was the first of the empiricist opponents of Descartes to achieve comparable authority among his European contemporaries. Together with Newton's physics, the philosophy of An Essay concerning Human Understanding gradually eclipsed Cartesianism, decisively redirecting European thought. Neoplatonic innatism was replaced with a modest, naturalistic conception of our cognitive capacities, making careful observation and systematic description the primary task of natural inquiry. Locke saw himself as carrying out just such a descriptive project with respect to the mind itself. Theorizing is the construction of hypotheses on the basis of analogies, not penetration to the essences of things by super-sensory means. In religion Locke took a similarly anti-dogmatic line, advocating toleration and minimal doctrinal requirements, notably in Epistola de tolerantia (A Letter concerning Toleration) and The Reasonableness of Christianity. Through his association with the Earl of Shaftesbury he became involved in government, and then in revolutionary politics against Charles II and James II. The latter involvement led to exile, and to Two Treatises of Government, a rejection of patriarchalism and an argument from first principles for constitutional government in the interests of the governed, and for the right of the misgoverned to rebel. Locke published his main works only after the 'Glorious Revolution' of 1688. He undertook important governmental duties for a time, and continued to write on many topics, including economics and biblical criticism, until his death. The Essay, Epistola and Second Treatise remain centrally canonical texts.

Locke held that all our ideas are either given in experience, or are complex ideas formed from simple ideas so given, but not that all our knowledge is based on experience. He accepted that geometry, for example, is an a priori science, but denied that the ideas which are the objects of geometrical reasoning are innate. 'Experience' includes 'reflection', that is reflexive awareness of our own mental operations, which Cartesians treated as a way of accessing innate ideas, but which Locke calls 'internal sense'. To have ideas before the mind is to be perceiving given or constructed sensory or quasi-sensory images - things as perceived by sense. In abstraction, however, we consider only aspects of what is presented: for example, a geometrical proof may consider only aspects of a drawn figure, allowing generalization to all figures similar in just those respects. Universal knowledge is thus perception of a relation between abstract ideas, but we also have immediate knowledge, in sensation, that particular external things are causing ideas in us. This awareness allows us to use the idea as a sign of its external cause: for example, the sensation of white signifies whatever feature of objects causes that sensation. Representation is thus fundamentally causal: causality bridges the gap between reality and ideas. Consequently we have sensitive knowledge of things only through their powers, knowledge of their existence without knowledge of their essence. Each way in which things act on the senses gives rise to a phenomenally simple idea signifying a quality, or power to affect us, in the object. Some simple ideas, those of the 'primary qualities', solidity, extension, figure, motion or rest, and number (the list can vary) can be supposed to resemble their causes. Others, ideas of 'secondary qualities', colour, smell, taste and so forth, do not. We also form ideas of the powers of objects to interact.

Our idea of any sort of substantial thing is therefore complex, including ideas of all the qualities and powers by which we know and define that 'substance'. Additionally, the idea includes the 'general idea of substance', or possessor of the qualities, a placemarker signifying the unknown underlying cause of their union. Locke distinguishes between the general substance, matter, and the 'particular constitution' of matter from which flow the observable properties by which we define each sort of substance - gold, horse, iron and so on. This 'real constitution' or 'real essence' is distinguishable only relatively to our definition or 'nominal essence' of the species. Locke extends this conceptualist view of classification to individuation in a famous, still influential argument that a person is individuated, not by an immaterial soul, but by unifying and continuous consciousness.

Because their real essences are unknown to us, we are capable only of probable belief about substances, not of 'science'. In mathematics, however, real essences are known, since they are abstract ideas constructible without reference to reality. So too with ideas of 'mixed modes' and 'relations', including the ideas of social actions, roles and relationships which supply the subject-matter of a priori sciences concerned with law, natural, social and positive. The three legislators are God, public opinion and government. God's authority derives from his status as creator, and natural or moral law is his benevolent will for us. Locke's political theory concerns the authority of governments, which he takes to be, at bottom, the right of all individuals to uphold natural law transferred to a central agency for the sake of its power and impartiality. Economic change, he argues, renders this transfer imperative. In a state of nature, individuals own whatever they have worked for, if they can use it and enough is

left for others. But with land-enclosure (which benefits everyone by increasing productivity) and the institution of money (which makes it both possible and morally justifiable to enjoy the product of enclosure) this primitive property-right is transcended, and there is need for an authority to ordain and uphold rules of justice for the benefit of all. Any government, therefore, has a specific trust to fulfil, and should be organized so as best to safeguard this role. A ruler who rules in his own interest forfeits all rights, as a criminal at war with his subjects. Then rebellion is justified self-defence.

1 Life and main works

John Locke was born at Wrington, Somerset in England on 28 August 1632. His father was a small property owner, lawyer and minor official, who served on the side of Parliament in the civil war under the more influential Alexander Popham. Through Popham, Locke became a pupil at Westminster School, then the leading school in England. From Westminster he was elected in May 1652 to a Studentship at Christ Church, Oxford, conditionally tenable for life.

During the next fifteen years at Oxford Locke took his degrees (B.A. 1656, M.A. 1658) and fulfilled various college offices, becoming Tutor in 1661. Between 1660 and 1662 he wrote three manuscripts on issues of Church and State, individual conscience and religious authority, two now published together and known as *Two Tracts on Government*, and *An necesse sit dari in Ecclesia infallibilem Sacro Sanctae Scripturae interpretem? (Is it necessary to have in the church an infallible interpreter of holy scripture?)*. Although his answer to the last question was predictably negative, in the *Tracts* he expressed a less-than-tolerant view of conscientious religious unorthodoxy, assigning to rulers the right to determine details of religious observance for the sake of public peace. While Censor of Moral Philosophy at Christ Church in 1664 he completed the Latin manuscript now known as *Essays [or Questions] on the Law of Nature*, which presaged his mature views - both his general empiricism and his conception of moral obligation as an obligation to God to obey natural law. This work also rejects wayward and dogmatic appeals to conscience, in favour of reason based on experience.

The politics of religion, at the time a large part of politics, was not Locke’s only extracurricular interest. His reading-notes (‘commonplace books’) of this time indicate an interest in Anglican theology, and by 1658 he was reading and taking lecture notes in medicine with the assiduity appropriate to a chosen career. This interest extended to chemistry and, in the 1660s, to the new mechanical philosophy as expounded, for example, by Robert Boyle, whom Locke had met by 1660. Locke also read the main philosophical works of Descartes, and some Gassendi, but his record focuses on their versions of corpuscularianism, bypassing metaphysical and epistemological underpinnings (see Descartes, R. §§11-12; Gassendi, P. §§2, 4). On the evidence, natural philosophy attracted Locke more at this time than metaphysics, although the coarse empiricism of *Essays on the Law of Nature* is close to that of Gassendi. Yet Locke could hardly have remained ignorant of the battle among the new philosophers between ‘gods’ and ‘giants’ - between those, led by Descartes, in the Platonic-Augustinian metaphysical tradition and those, headed by Gassendi and Hobbes, who developed ancient empiricist and materialist theory.

In 1665 Locke’s university life was interrupted by a diplomatic mission to Brandenburg as secretary to Sir Walter Vane. About this time he decided against entering the church, but took the one way of nevertheless keeping his Studentship (without obligation to reside in Oxford) by transferring formally to medicine. In 1666 came a momentous meeting with Lord Ashley (Anthony Ashley Cooper, who became Earl of Shaftesbury in 1672), whose London household Locke subsequently joined in 1667. Here his medical and political interests alike received a more practical edge than they had previously possessed. He began collaborating closely with the pre-eminent physician, Thomas Sydenham, and in 1668 successfully supervised an operation on Lord Ashley to drain an abscess on the liver. In the years following he continued to act as medical advisor within Ashley’s circle, supervising the birth of Ashley’s grandson, later the philosophical Third Earl of Shaftesbury. A manuscript of this time in Locke’s handwriting (but perhaps wholly or partly by Sydenham), ‘De Arte Medica’, is strongly sceptical of the value of hypotheses, as opposed to experience, in medicine.

During this same period, presumably influenced by his patron, Locke wrote the manuscript *Essay concerning Toleration* (1667), departing from his earlier, nervously illiberal justification of constraint and advocating toleration of any religious persuasion not constituting a positive moral or political danger - provisos excluding, respectively, atheists and Roman Catholics. In 1667 Ashley became a member of the governing ‘cabal’ which

followed Clarendon's period as Lord Chancellor, and in 1672 became Lord Chancellor himself. Under Ashley, and for a while after Ashley's fall from office in 1673, Locke was involved in government. He began to work on economic questions, and for some years helped in the organization of the newly founded colony of Carolina. He was registrar to the commissioners of excise (perhaps a sinecure) from 1670 to 1675, secretary for presentations (in charge of ecclesiastical patronage) in 1672-3, and secretary and treasurer to the Council for Trade and Plantations (no sinecure) in 1673-4.

Nevertheless he found time for new intellectual interests. Not later than 1671 he put down for discussion by a group of friends what he later claimed (inaccurately, given *Essays on the Law of Nature*) to have been his first thoughts on the powers of the understanding. He found the topic sufficiently gripping for a more extensive treatment than such an occasion would have demanded in '*Intellectus humanus cum cognitionis certitudine, et assensus firmitate*' (The human intellect, the certainty of knowledge and the confirmation of belief), dated 1671, with a longer (and as strongly empiricist and imagist) redrafting in the same year entitled 'An Essay concerning the Understanding, Knowledge, Opinion and Assent' - the manuscripts now known as Drafts A and B of *An Essay concerning Human Understanding*.

In 1675 Locke moved to France, beginning at the same time to write his journal. He met physicians and philosophers, undertook a programme of reading in French philosophy and continued working on his 'Essay'. On returning to England in 1678, after the fabricated 'Popish plot', he was again caught up in politics and in attempts to exclude Charles' brother James from the succession. Charles dissolved Parliament in 1681, and Shaftesbury led a group of Whigs planning insurrection. During this period Locke probably wrote the bulk of the *Two Treatises of Government*; the first, at least, to support moves for James' exclusion, the second possibly later to advocate actual rebellion. He also wrote, with James Tyrrell, a long response (still unpublished, 1997) to Edward Stillingfleet's *Unreasonableness of Separation*, defending the position of nonconformists against Stillingfleet's criticisms. In 1682 Shaftesbury went into exile, dying soon after. When the Rye House plot to assassinate Charles and James was uncovered in 1683, Locke himself prudently moved to Holland, where he contacted other, more overtly active exiles. His connections provoked expulsion from his Christ Church Studentship in 1684, and at the time of Monmouth's rebellion he went into hiding to escape arrest. His intellectual activities continued unabated, the *Essay* being largely written by 1686. In 1685-6 he wrote *Epistola de Tolerantia (Letter concerning Toleration)*, perhaps in response to the revocation of the edict of Nantes. He made friends, and discussed theological questions, with the remonstrant Philippus van Limborch and Jean Le Clerc, publishing various items in the latter's journal, *Bibliothèque universelle et historique*, including a review of Newton's *Principia* (1686) and a ninety-two page abridgement of the *Essay* (1688).

In 1688 the 'Glorious Revolution' brought the deposition of James, and Locke returned to England the following year. He declined the post of ambassador to Brandenburg, accepted an undemanding post as commissioner of appeals (annual salary, £200) and set about publishing his writings. *Epistola de Tolerantia* was published pseudonymously in Holland in May 1689, and Popple's English translation followed within months. The *Two Treatises* were revised and published anonymously, and the *Essay* followed in December (with authorship acknowledged), although both books were dated 1690. *A Second Letter concerning Toleration* (1690) and *A Third Letter for Toleration* (1692) were in response to attacks by an Anglican clergyman, Jonas Proast. *Some Considerations of the Consequences of the Lowering of Interest and Raising the Value of Money*, partly based on the manuscript of 1668, was published in 1691 (dated 1692) against Parliamentary measures of the time. In 1691, Locke accepted the invitation of an old friend, Damaris Masham and her husband to live with them, as far as his concerns permitted, at Oates in Essex. Country life seems to have ameliorated the asthma which dogged his last years. *Some Thoughts concerning Education* (1693, revised 1695), a significant work in the history of educational theory, was based on a number of letters of advice to his friend, Edward Clarke. In 1694 came the second edition of the *Essay*, with important additions including a controversial chapter on identity. In 1695 he published a new work, once more anonymously, *The Reasonableness of Christianity*. John Edwards' attacks on its liberal, minimalist interpretation of Christian faith were rebutted in two *Vindications* (1695b, 1697a) (see Latitudinarianism).

Locke continued to be engaged on economic questions, and in 1695 he joined a committee to advise the Chancellor of the Exchequer on monetary policy. His recommendations, supported by further papers, were accepted. In 1696 came an important government appointment to the Council for Trade and Plantations, and for

four years he fulfilled fairly onerous duties on the Board of Trade for the considerable annual salary of £1,000. At the same time he engaged in an extended controversy with Edward Stillingfleet, who found the *Essay* theologically suspect. *A Letter to the Right Reverend Edward, Lord Bishop of Worcester* was followed by two further Letters in reply to Stillingfleet's Answers. Despite its controversial style, Locke's argument is often a cogent clarification of his position. The exchange prompted significant alterations to the fourth edition of the *Essay* (1700) and long passages were included as footnotes in the posthumously published fifth edition. In June 1700 Locke resigned from the Board of Trade, a sick man, and thereafter lived mostly at Oates. Pursuing a long-standing interest in biblical criticism, he set about the work which was posthumously published as *A Paraphrase and Notes on the Epistles of St. Paul*, an important contribution to hermeneutics (see Hermeneutics, biblical). In 1702 he wrote the reductive *Discourse of Miracles*, and in 1704 began a *Fourth Letter on Toleration*. On October 28th 1704 he died as Damaris Masham read to him from the Book of Psalms. For the last years of his life he was generally respected as, with Newton, one of Britain's two intellectual giants, a reputation undiminished by death.

2 The structure of Locke's empiricism

Locke's mature philosophy is 'concept-empiricist', but not 'knowledge-empiricist': he held that all our concepts are drawn from experience, but not that all our knowledge is based on experience. Yet his early position, in *Essays on the Law of Nature* and the first part of *Draft A*, was 'knowledge-empiricist' in just this sense - even the axioms of geometry gain assent 'only by the testimony and assurance of our senses' (*Draft A* I: 22-3). However, according to *Draft A*, when we find that certain relations hold without exception, we assume that they hold universally and come to employ them as 'standards' of measurement embodied in the meaning of our terms. Locke sees this as implying a choice: an axiom can *either* be interpreted as an 'instructive', but uncertain summary of experience, *or* as a quasi-definition, founded on experience but 'only verbal…and not instructive'. But later in *Draft A* he discards the notion that geometrical axioms can be interpreted empirically, taking them only in a sense in which they can be known by 'demonstration' or 'the bare shewing of things or proposing them to our senses or understandings' (*Draft A* I: 50) - that is, by intuitions with perceived or imagined instances (for example, diagrams) as their objects. At the same time he recognizes that mathematical propositions are not plausibly regarded as merely verbal. The possibility of alternative interpretations of universal propositions, *either* as certain, but verbal, *or* as instructive, but uncertain, is now restricted to propositions about substances, such as 'Man is rational'. Locke has shifted, in effect, from knowledge-empiricism towards a concept-empiricism which allows 'instructive' a priori knowledge (the last being the acknowledged ancestor of Kant's synthetic a priori - see Kant, I. §4).

Locke's intuitionism shapes his attack on the innatism characteristic of the Platonic-Augustinian-Cartesian tradition (see Innate knowledge). Starting with propositions, Locke rebuts the argument from alleged universal assent, or assent by all who have come to the use of reason. But ideas are what is before the mind in thought, and propositions are ideas in relation. Locke's underlying thesis is that to take either knowledge or ideas to be innately 'imprinted on the mind' in a merely dispositional sense (and they are clearly not actual in all human beings from birth) would be contrary to any intelligible notion of being 'in the mind': 'Whatever idea was never perceived by the mind, was never in the mind' (*Essay* I.iv.20). Locke concedes dispositional knowledge and ideas, retained by the memory and capable of being revived, but he understands both intentionality and knowledge in terms of perception, and finds nonsensical the notion of perception which never has been conscious and actual. This strongly intuitionist model rules out dispositional innatism as an intelligible possibility. Rationalist intuitionism, from Locke's point of view, is simply incoherent. And since the only dispositional ideas and knowledge are what is retained in the memory, what is before the mind as the object of intuition or demonstration must be experiential or sensory.

Locke also argues that there are no general maxims of logic or mathematics to which all assent when they come to the use of reason, since many rational but illiterate people never consider such abstract principles. He does not accept that reasoning merely consonant with logical principles is equivalent to assent to them, or, for example, that distinguishing two things is tacit employment of the idea of identity. Explicit abstract principles and ideas come late and with so much difficulty that people cannot agree on ideas of impossibility, identity, duty, substance, God and the like - just the ideas most supposed innately luminous. That rational people assent to certain propositions on first proposal is beside the point, since such people will only have understood the terms of the proposition in question by abstraction from experience. Then they will assent, not because the proposition is innate, but because

it is evident. To describe the bare capacity to perceive such truths as the possession of innate principles and ideas will make all universal knowledge innate, however specific or derived. Turning to practical principles and the idea of God, Locke appeals to anthropology to rebut the claim that any of these are universally recognized. The main thrust of his argument, however, is conceptual.

Locke’s empiricism has another central feature. Like Gassendi and Hobbes, he expressly accords independent authority to the particular deliverances of the senses (see Gassendi, P. §3). Descartes had argued that sensation requires interpretation employing innate, purely intellectual ideas even in order for us to conceive of its objects as independent bodies. For Descartes, moreover, natural sensory belief is defenceless in the face of sceptical argument - secure knowledge of the existence of bodies can only be achieved through a rational proof involving reflection on the role and mechanisms of sense (see Descartes, R. §9). This emphatic subordination of sense to reason Locke rejects just as firmly: the senses are ‘the proper and sole judges’ of the existence of bodies. He sees the senses as knowledge-delivering faculties in their own right, prior to any understanding of their mechanisms: ‘the actual receiving of ideas from without…makes us know, that something doth exist at that time without us, which causes that idea in us, though perhaps we neither know nor consider how it does it’ (*Essay* IX.xi.2). The sceptic’s doubt about the external world is a mere pretence, not to be taken seriously: ‘no body can, in earnest, be so sceptical, as to be uncertain of the existence of those things which he sees and feels’. Echoing Lucretius, Locke sees the reason employed in sceptical argument itself as standing or falling with the senses: ‘For we cannot act any thing, but by our faculties; nor talk of knowledge it self, but by the help of those faculties which are fitted to apprehend even what knowledge is’ (*Essay* IX.xi.3). Locke does identify features of sense-experience which militate against scepticism: for example, sensory ideas depend on physical sense-organs, and are systematically and unavoidably consequent on our situation; the deliverances of different senses cohere; there is a ‘manifest difference’ between ideas of sense and ideas of memory and imagination (most dramatically with respect to pain), as there is between acting in the world and imagining ourselves acting; and so on (*Essay* IX.xi.4-8). Yet all these considerations are simply ‘concurrent reasons’ which further, but unnecessarily confirm ‘the assurance we have from our senses themselves’ - ‘an assurance that *deserves* the name of knowledge’ (*Essay* IX.xi.3).

Locke’s explanation of the certainty and extent of ‘sensitive knowledge of existence’ hinges on his view that in sensation we are immediately aware, not only of sensations or ‘ideas’, but of their being caused by things outside us. We are thus able to think of the unknown cause through its effect in us: ‘whilst I write this, I have, by the paper affecting my eyes, that idea produced in my mind, which, whatever object causes, I call *white*; by which I know, that that quality or accident (i.e. whose appearance before my eyes, always causes that idea) doth really exist, and hath a being without me’ (*Essay* IX.xi.2). This claim ties in with another, that ideas of simple sensory qualities are always ‘true’, ‘real’ and ‘adequate’: ‘their truth consists in nothing else, but in such appearances, as are produced in us, and must be suitable to those powers, [God] has placed in external objects, or else they could not be produced in us’ (*Essay* II.xxxii.14). Simple ideas are ‘distinguishing marks’ which fulfil their function well enough whatever unknown difference lies behind the sensible distinction. But this function fits them for another, as terms in the natural language of thought. The idea of white signifies, that is *indicates*, its unknown cause, and also signifies, that is *stands for*, that feature of things in thought. So the limited causal knowledge that sensation supplies allows us to have contentful thought and knowledge of the external world. The idea of power extends such pretheoretical knowledge: our idea of the melting of wax, joined to the idea of active or passive power, can be employed as a sign of whatever in the sun melts wax, or of whatever in wax causes its melting. Consequently Locke decides to treat ideas of powers as simple ideas, and knowledge of powers as observational. The senses do not give knowledge of the essence or nature of bodies, but they do give knowledge of their existence, and enable us to distinguish between them.

3 Ideas of sensation and reflection: their retention and abstraction

Locke’s employment of the word ‘idea’ responds to a variety of antecedents. Like Descartes, he uses it ambiguously both for representative states (acts, modifications) of mind and, more frequently, for the represented objects as they are represented or conceived of, the so-called ‘immediate’ objects of perception and thought. To have an idea before the mind is generally, for Locke, to be contemplating something under a certain conception rather than contemplating a psychological state. To ‘perceive a relation between ideas’ is to perceive a relation between things-as-conceived-of. But Locke’s account also looks back to the Epicurean view of sensations as signs of their unknown causes in the motion of atoms or ‘corpuscules’ (see §2), a view which points away from the

Cartesian and scholastic presumption of intrinsically representative elements in thought towards a purely causal understanding of representation, treating ideas as blank sensory effects in the mind. Locke never resolves the tension between these different conceptions of an idea, although each of them is necessary to his theory.

Locke strongly opposes the Augustinian-Cartesian view that knowledge and truth consist in the conformity of human conceptions with God's conceptions, the divine ideas or archetypes employed in creation and revealed to us in our active use of reason. For Descartes, human reason is only accidentally involved with the senses, whereas for Locke there are no purely intellectual ideas. The task traditionally assigned to intellect - universal thought - Locke assigns to 'abstraction', taken to be the mind's in some sense separating out elements of raw experience and employing them as 'standards and representatives' of a class. What this means will be considered.

Although Locke sometimes writes that all words stand for ideas, ideas are the mental correlates of terms or names: that is, words that can stand in subject or predicate place. He adheres to the traditional view that 'particles', such as prepositions, conjunctions, the copula and the negative, signify, not ideas, but 'the connection that the mind gives to ideas, or propositions, one with another' (*Essay* III.vii.1). They do not *name*, but *express* 'actions of the mind in discoursing': for example, 'but' expresses various mental operations together named 'discretive conjunction'. The mental actions or operations expressed by 'is' and 'is not' are either the 'perception of the agreement or disagreement of ideas', which is Locke's definition of (at least, general) knowledge, or the 'presumption' of such a relation, which is Locke's account of belief or judgment. As commonly in earlier logic, merely considering a proposition is not distinguished from knowing or judging it to be true.

The aim of Book II of the *Essay* is to establish that all our ideas derive from experience: that is, that the way we conceive of the world (including ourselves) is ultimately determined by the way we experience the world. 'Experience' includes not only 'sense', but reflection ('reflexion') - not reflection in today's sense but reflexive awareness of our own mental operations. Platonists, Aristotelians and Cartesians all assigned the reflexive awareness of thinking to intellect rather than to sense. For Descartes, the innateness of such ideas as *substance*, *thought* and even *God* consists in the potentiality of their becoming explicit through the mind's reflecting on itself, and Leibniz argues accordingly that, simply by admitting reflection as well as sense, Locke admits innate ideas (see Leibniz, G.W. §8). Locke, however, claims that reflection, 'though it be not sense, as having nothing to do with external objects, yet it is very like it, and might properly enough be call'd internal sense' (*Essay* II.i.4). Thereafter he treats sense and reflection as theoretically equivalent (although reflexive knowledge of one's own existence is 'intuitive' rather than 'sensitive' - *Essay* IV.ix.3). This move not only extends the empiricist principle to such non-sensory notions as *willing*, *perceiving*, *contemplation* or *hope*, but also contradicts the Cartesian model of thought as transparent to itself, propounding a gap between how thinking appears to the subject and what it really is in itself - the latter being unknown. Locke also insists that reflection is second-order awareness, presupposing sense-perception as the first mental operation. And though 'ideas in the intellect are coeval with sensation' (*Essay* II.i.23), it seems that the mind must 'retain and distinguish' ideas before it can be said to 'have ideas' dispositionally, stored in the memory for employment as signifiers in thought. Ideas of reflection in particular are achieved only 'in time' - and here 'reflection' acquires some of its modern affinity with 'contemplation'. Children, Locke's accounts of both reflection and particles imply, can discern or compound ideas without having the ideas of discerning or compounding, and few of those who employ particles to *express* various mental actions ever pay them enough attention to be able to *name* them. Locke does assert that in the reception of ideas 'the understanding is merely passive', but he also allows that attention, as well as repetition, helps 'much to the fixing any ideas in the memory' (*Essay* II.x.3).

The 'retention' of ideas in the memory, therefore, is a necessary condition of discursive thought, and its description significantly echoes Hobbes' account of memory as 'decaying sense'. What decay are - 'it may seem probable' - images in the brain, and hostility to the separation of intellect from imagination pervades the *Essay*. Descartes' famous argument for such a separation - that we can accurately reason about a chiliagon although we cannot form a distinct image of it - is directly rebutted: the reasoning is made possible by our precise idea of the number of the sides (itself dependent on the technique of counting), not by a clear and distinct idea of the shape. 'Clear' ideas are, by definition, such as we receive 'in a well-ordered sensation or perception'. Locke's treatment of abstraction accords with such express sensationism. 'Abstract ideas' are particulars, universal only in 'the capacity, they are put into…of signifying or representing many particular things' (*Essay* III.iii.11). Locke means that in abstract thought the mind relates to, and employs, sensory images in a certain way, not that it manufactures

sense-transcendent objects of intellect. Abstract ideas are what we have *distinctly* before the mind in general thought, but distinctness may be achieved by 'partial consideration', not absolute separation: 'Many ideas require others as necessary to their existence or conception, which are yet very distinct ideas. Motion can neither be, nor be conceived without space' (*Essay* II.xiii.11-13). The very abstract ideas of *being* and *unity* are ideas of anything whatsoever considered as existing, or as one. Geometry gave Locke his paradigm of 'perception of the relation between ideas'. But where Cartesians saw the role of geometrical diagrams to be the stimulation of intellectual ideas, for Locke, as for Hobbes, the object of reasoning and source of 'evidence' is the diagram itself, whether actual or imagined. (Kant's 'intuition' owes something to Locke.) Given these structural features of his theory, it seems undeniable, as some have denied, that Locke's ideas are essentially sensory (or reflexive) images (see Hobbes, T. §4; Kant, I. §5).

4 Five sorts of idea

Book II of the *Essay* presents an alternative to Aristotle's doctrine of categories, the traditional typology of entities capable of being named or predicated (see Aristotle §7). That Locke's classification is of *ideas* rather than of *things* stresses that the categories are purely conceptual. He identifies five broad types: simple ideas, ideas of simple modes, ideas of mixed modes, ideas of substances and ideas of relations. Simple ideas come first in the Lockean order of knowledge, as substances come first in the Aristotelian order of being. Simple ideas are necessarily given in experience, whereas complex ideas can be constructed by 'enlarging' ('repeating') or 'compounding' simple ideas. Ideas of relations result from 'comparing' ideas. 'Abstracting' is more a matter of focusing on an idea or, better, an aspect of an idea, whether given or constructed, than of creating a new one (see §3). Locke sometimes acknowledges that the overarching compositional model is problematic in its application, but it is put into doubt even by his formal introduction of the notion of a simple idea. The ideas of the sensible qualities of a body, Locke claims, though produced by the same body, in some cases by the same sense, are evidently distinct from one another, each being 'nothing but *one uniform appearance*, or conception in the mind' (*Essay* II.ii.1). Yet to ascribe the conceptual distinctions between, for example, a thing's shape, its motion and its colour to a primitive articulation of *appearance* is to beg a crucial question.

Under the topic of simple ideas Locke expounds his famous distinction between primary and secondary qualities (*Essay* II.viii). Since the cause of a simple idea may be quite different in character from the idea itself, we should distinguish the idea in the mind from the corresponding quality (that is, the power to cause the idea) in bodies. Certain qualities, however, are necessary to our conception of bodies as such. These are the primary qualities, 'solidity, extension, figure, motion, or rest, and number', just those which figured in corpuscularian speculations. Locke's proposal (displaying the tension, described in §3 of this entry, between two conceptions of representation) is that in the perception of a primary quality the represented cause, the basis of the power in the object, is qualitatively *like* the idea caused: 'A circle and square are the same, whether in idea or existence' (*Essay* II.viii.18). Only this will allow that the action of external bodies on the senses is 'by *impulse*, the only way which we can conceive bodies operate in' (*Essay* II.viii.12) - an appeal to the seventeenth-century commonplace that mechanical explanations are peculiarly intelligible. But then ideas of 'colours, sounds, tastes, etc.', Locke's 'secondary qualities', must also be mechanically stimulated. Hence secondary qualities 'are nothing in the objects themselves, but powers to produce various sensations in us by their primary qualities, i.e. by the bulk, figure, texture, and motion of their insensible parts' (*Essay* II.viii.10). Ontologically they are in the same boat as the power of fire to cause pain or, indeed, its power to melt wax.

The idea of power itself Locke attributes to experience of regular patterns of change, giving rise first to expectations that 'like changes will for the future be made in the same things, by like agents, and by the like ways', and then to the thought that in one thing exists the possibility of being changed and in another 'the possibility of making that change' (*Essay* II.xxi.1). So we form the idea of power, active and passive: the power of fire to melt wax and the power of wax to be melted are aspects of fire and wax known and identified only through their joint effect. The idea of power is thus a place-marker for attributes which could in principle be known more directly.

The ideas or experiences of pleasure and pain are important simple ideas, since they are responsible for our ideas of good and evil, and are 'the hinges on which our passions turn' (*Essay* II.xx.3). (This hedonistic theory of motivation and value is examined in §9 of this entry.)

‘Simple modes’ constitute another problematic category. Locke starts with modes of extension, the subject-matter of geometry, with which he compares modes of duration. Here his thesis is that we acquire ideas of particular modes of extension (that is, determinate lengths and figures) or duration (that is, periods) in experience, and can then repeat (or divide) them so as to construct ideas of possible lengths, figures or periods not previously experienced. Roughly, ‘modification’ here is compounding like with like. The same model supplies Locke’s account of ideas of numbers, achieved by the repetition or addition of units, aided and ordered by the linguistic technique of counting. Yet he also recognizes qualitative simple modes, effectively conceding that ideas of different ‘shades of the same [experienced] colour’ are constructible. Even with quantitative ‘modes’, where the ‘repetition’ model has some plausibility, it is problematic what is a simple idea. The idea of determinable *extension* is a plausible candidate, with its determinates as ‘modes’, but the repetition model presupposes simple units. Locke impatiently responds that the smallest sensible point ‘may perhaps be the fittest to be consider’d by us as a *simple idea* of that kind’ (*Essay* (5th edn) II.xv.5n.), but he was evidently more concerned to argue that ideas of novel determinate figures are *somehow* constructible from what has been given, and so to subvert a Platonic-Cartesian argument for innateness, than to insist on the adequacy of a rigid compositional model.

Another target in Locke’s account of simple modes is Descartes’ conceptual identification of space and matter in the thesis that the essence of matter is extension. For Locke, both the essence of matter and the nature of space are unknown. He argues that our idea of a vacuum is not contradictory, since our ordinary idea of body includes solidity as well as extension, but he declines to choose between relational and realist theories of space. Yet comparison of the *Essay* with earlier notes and drafts indicates that, having first held a Hobbesian relational view, Locke came gradually to favour a realism close to that of Newton (see Descartes, R. §§8, 11; Newton, I. §4).

Ideas of mixed modes arise with the combination of unlike simple ideas, as in the idea of a rainbow. But Locke’s paradigms are ideas of human actions and institutions, the materials of demonstrative moral and political theory. Like ideas of geometrical figures, ideas of mixed modes can properly be formed without regard to what exists. Ethical thought is none the worse for being about a virtue or motive or political constitution which is nowhere actually instantiated. Ideas of substances are different, for they concern the real rather than the ideal: ‘When we speak of *justice*, or *gratitude*…our thoughts terminate in the abstract ideas of those virtues, and look not further; as they do, when we speak of a *horse*, or *iron*, whose specific ideas we consider not, as barely in the mind, but as in things themselves, which afford the original patterns of those ideas’ (*Essay* III.v.12). Moreover, whereas ideas of substances are formed on the presumption that the complex idea represents a really or naturally united thing, the unity of mixed modes is essentially conceptual. Indeed, ‘Though…it be the mind that makes the collection, tis the name which is, as it were the knot, that ties them fast together’. Different languages slice up the field of human life and action in different ways, determined by the practices and priorities of the communities that speak them. This thesis can be extended to natural modes such as freezing, since even here it is the term tied to a striking appearance, not a natural boundary, which slices out the particular process from the general process of nature. That, Locke plausibly assumes, is not how it is with horses.

The chief thought behind Locke’s somewhat confusing account of ideas of substances is that our idea of a thing or stuff is a compound of ideas of its qualities, but the thing itself is not a compound of qualities (*Essay* II.xxiii). The substance-accident structure is a feature of our ideas and language, not a structure in reality. It is a feature which marks our ignorance of the underlying nature of things, since we always conceive and talk of a substance as a *thing* possessing certain qualities, that is, as a ‘*substratum*, in which [the qualities] do subsist, and from which they do result, which therefore we call substance’. The mistake of dogmatic philosophers is to think that they can form *simple* conceptions of substances matching their unitary natures. Aristotelians are so misled by language that, just because, ‘for quick despatch’, we employ one name, ‘gold’ or ‘swan’, they think it a ‘simple term’ corresponding to a ‘simple apprehension’. Cartesians take the simple essences of matter and spirit to be extension and thought. Yet so far are we from catching the nature of any thing in our *complex* idea of it that, if it is asked what the subject is of the qualities by which we define it (the colour and weight of gold, for example), the best answer we can give is ‘the solid extended parts’, that is, the mechanistic ‘corpuscularian’ hypothesis as advanced by Boyle. If it is asked in turn ‘what it is, that solidity and extension inhere in,’ we can only say, ‘we know not what’. Our idea of the substance is of ‘nothing, but the supposed, but unknown support of those qualities, we find existing, which we imagine cannot subsist, *sine re substante*, without something to support them’. Such an idea is ‘obscure and relative’. Ideas of specific substances are ‘nothing but several combinations of simple ideas,

co-existing in such, though unknown, cause of their union as makes the whole subsist of itself' (*Essay* II.xxiii.6). Locke's point is that no theory, not even the corpuscular hypothesis, gives an account of the ultimate nature of things.

Finally come ideas of relations - father, son-in-law, enemy, young, blacker, lawful and so on (*Essay* II.xxv-xxviii). Like ideas of modes, ideas of relations can properly be constructed without regard to reality, in particular if they are conventional relations. Adequate ideas even of natural relations, Locke claims, are possible without adequate ideas of the things related: we can grasp the essence of fatherhood without knowing the essence of man or even the mechanisms of reproduction. His point is that the biological details are irrelevant to the rights and duties of a father - a question rationally determined in his own attack on patriarchalism in *Two Treatises*. From this point of view, relations are theoretically close to modes. Yet Locke does allow certain relations to have peculiar ontological significance. Causal, spatial and temporal relations are universal relations which pertain to all finite beings. Identity and diversity are so too: a thing is diverse from anything existing in a different place at the same time, 'how like and indistinguishable soever it may be in all other respects', and the continuity of individual substances is spatio-temporal. The last important type of relations to be picked out for special discussion is that of moral relations, or the relations of actions to some law 'whereby good or evil is drawn on us, from the will and power of the law-maker' (*Essay* II.viii.5).

5 Substances, mixed modes and the improvement of language

On Locke's account of communication (*Essay* III.i-ii), names should, by common convention or special agreement, excite in the hearer's mind just the same ideas as they are associated with in the speaker's mind. Collaborative progress in the sciences depends on 'clear and distinct' or 'determined and determinate' ideas - that is, on consistent and agreed association of ideas and words (*Essay* II.xxix; compare 'Epistle to the Reader'). Locke's discussion of language is shaped by his belief that these conditions of the transference of knowledge were in his time commonly unsatisfied, especially in two domains. First, there was no agreed classification of 'substances' (living things and chemicals) based on careful observation and experiment. Second, the ideas associated with the names of mixed modes often varied both in the usage of different people and in that of the same person at different times. Two mistakes in particular disguise these shortcomings of language. The first is the assumption that a common set of words ensures a common language in the full sense, with a shared set of meanings. So people may argue about 'honour' and 'courage' without realizing that they mean different things, or nothing at all, by the words. The second mistaken assumption is that words have meaning by standing for things directly, as if the meaning of 'salt', 'gold' or 'fish' were fixed demonstratively, by what is named. The first assumption chiefly corrupts our thought about mixed modes, the second relates 'more particularly, to substances, and their names' (*Essay* III.ii.5). Locke's radical and influential views about the latter will be considered first.

The 'idea of substance in general' employed in ideas of specific substances is the idea of something unknown underlying the attributes known by experience (see §4 of this entry). Many have objected, following Leibniz, that here Locke confusedly postulates ignorance of the subject of attributes which is not ignorance of attributes of the subject. Yet he holds that our ignorance of 'the substance of body' and 'the substance of spirit' is an ignorance of the *natures* of these things - ignorance manifested in our inability to understand the internal cohesion or (he adds in later writings) mutual attraction of bodies, or to explain what thinks in us, and how it does so. His approval of the corpuscularian hypothesis and Newton's mechanics is qualified - the best available physical theory leaves too much unexplained to be the whole truth (Newton did not disagree). The idea of substance is a place-marker for essences which are unknown, but knowable, if possibly not by human beings (see Newton, I.).

One feature of Locke's theory which has made difficulties for the present interpretation is the distinction he makes between substance and 'real essence'. The real essence of a thing, Locke says, may be taken for 'the very being of a thing, whereby it is what it is', the 'real internal, but generally in substances, unknown constitution of things, whereon their discoverable qualities depend' (*Essay* III.iii.15). Nevertheless, '*essence*, in the ordinary use of the word, relates to sorts' (*Essay* III.vi.4). Species and genera, or sorts of things, Locke asserts, are creatures of the understanding, with membership determined by abstract ideas made on the basis of experienced resemblances, not by the presence in each of a specific form, or by a common derivation from a divine archetype. Ultimately it is a matter of arbitrary definition which observable resemblances we count as necessary for membership of this or that named sort. It is not just that specific real essences are unknown, since (Locke argues) even if we did know the real

constitution of things as well as clock-makers know the works of clocks, it would still be up to us where to draw the boundaries between species, and what to include in our abstract ideas or ‘nominal essences’. The real essence of a species can therefore only be ‘that real constitution…which is the foundation of all those properties, that are combined in, and are constantly found to co-exist with, the *nominal essence*’ of the species (*Essay* III.vi.6). Here, the model is that of a universal matter determinately modified as a variety of particles interacting mechanically so as to constitute the material things of ordinary experience. Since at the fundamental level these observable quasi-machines differ from one another merely quantitatively, and can do so by indefinite degrees, there are no absolute boundaries among them. There are only the discernible resemblances and differences consequent on their underlying mechanical differences - ‘the wheels, or springs…within’. Even more certainly our actual classification is not based on knowledge of any such boundaries. Talk of the real essence of a species, and the distinction between its ‘properties’ and its ‘accidents’ (’properties’ flowing from the essence), are therefore, contrary to Aristotelian assumptions, *de dicto* and relative to the nominal essence defining the name of that species (see Aristotle §8; De re/de dicto; Essentialism).

This conception of a real essence assigns it a role closely related to that of substance. What, after all, is the ‘unknown cause of the union’ of any of the ‘combinations of simple ideas’ by which ‘we represent particular sorts of substances to ourselves’, if not the real essence underlying the nominal essence in question? Yet Locke sometimes distinguishes both the notion and knowledge of real essence from the notion and knowledge of substance. That is not, however, because the ‘substance’ is an irremediably unknown subject underlying even essence, but because it is the common stuff of a variety of species of things, ‘as a tree and a pebble, being in the same sense body, and agreeing in the common nature of body, differ only in a bare modification of that common matter’ (*Essay* II.xiii.18). The unknown modification is the specific ‘real essence’, and the equally unknown general nature of matter is the ‘substance’. Locke also envisages deeper differences of kind between substances: ‘God, spirits and body’ are all ‘substances’ only because we think of each of them indeterminately as *something*, not because of a shared nature. But by the same token we distinguish spirit and body only because we cannot understand how matter could think, not because we can grasp their separate essences, as Descartes had supposed (see Descartes, R. §8). Indeed, since we are equally unable to understand how spirit and body might interact, or how spirits could occupy places, the issue between materialist and immaterialist accounts of minds is for Locke undecidable, and at best a matter for speculation.

Locke’s corpuscularian conception of a world of machines, resembling and differing from one another by continuous degrees, is consonant with his independent epistemological conviction that names have meaning only through association with ‘ideas’, rather than directly with ‘things as really they are’. Together they motivate his programme for improving natural classification which advocates, not the allegedly impossible Aristotelian ideal of identifying the natural hierarchy of genera and species, but general agreement on a practically useful way of gathering and ordering the things in the world, taking into account such dependable concomitances of qualities and powers as appear to careful observation and experiment. Locke saw the future of biology and chemistry - and even of mechanics - in descriptive ‘natural history’, justifiable as a useful, orderly record of dependable means to ends but falling short of systematic ‘science’. Despite its apparent pessimism, his view has survived in biological taxonomy as a continuous tradition of scepticism as to the reality of our taxonomical divisions. In semantic theory, Locke’s broad conception of how the names of substances have meaning has only recently been eclipsed by a quasi-Aristotelian view (see Kripke, S.A.; Putnam, H.).

Locke saw equal need for a programme of agreeing definitions in ethics, where his target is less the notion that moral and political terms name independent realities, than the assumption that the very existence of a word in a language ensures that it has a fixed, common meaning. ‘Common use’, Locke concedes, ‘regulates the meaning of words pretty well for common conversation’ (*Essay* III.ix.8) - for the ‘civil’ rather than ‘philosophical’ use of words. But where precision is required, as in the establishment or interpretation of a law or moral rule, reliance on ordinary usage leaves us vulnerable to the trickery of rhetoricians who prove bad qualities good by shifting the meaning of terms; or to the subtleties of interpretors, whether of civil or revealed law, who render unintelligible what started off plain. The remedy is to give the names of virtues and vices, and of social actions, roles and relations the fixed and unequivocal definitions necessary for a clear and unwavering view of right and wrong, ‘the conformity or disagreement of our actions to some law’.

6 Knowledge and belief

Like Platonists and Cartesians, Locke drew a strong distinction between knowledge and belief (also called opinion, judgment, or assent), but the ground and placing of this division between two forms of propositional ‘affirmation’ differed from theirs fundamentally. As the distinction is expounded in Book IV of the *Essay*, universal knowledge or ‘science’ does not have special objects, whether a transcendent intelligible world in the mind of God or innate intellectual ideas. Its difference from general belief lies in the way in which ideas are related in the mind. In universal knowledge, the ‘connection and agreement, or disagreement and repugnancy’ of ideas is ‘perceived’, whereas in belief it is ‘presumed’ on the basis of ‘something extraneous to the thing I believe’. To follow a proof gives knowledge of the conclusion, whereas to accept the conclusion on the authority of a mathematician constitutes only belief. Similarly in the case of ‘sensitive knowledge’ of particular existence, what we ourselves perceive we know to be so, but what we infer, or accept on testimony, we merely believe.

Knowledge, as well as assent, is subject to ‘degrees’: there are degrees, not only of probability, but of ‘evidence’. The first degree of knowledge is intuitive knowledge, in which the mind ‘perceives the truth, as the eye doth light, only by being directed toward it’. Intuition ‘leaves no room for hesitation, doubt, or examination’. The second degree is demonstrative knowledge, where the truth is perceived by the aid of one or a chain of ‘intermediate ideas’. Doubt or mistake is possible at any point in the sequence with respect to connections not currently in view. Hence, ‘Men embrace often falsehoods for demonstration’. Locke’s chief model for ‘intermediate ideas’ is geometrical: for example, the lines employed in the intermediate steps of the Euclidean proof which allow us to see that the angles of a triangle are equal to the angle on a straight line. Although his conception of intuition can seem Cartesian, the profound difference is that, for Locke, ideas which are objects of intuition are essentially a product of sense (including reflection) and imagination. As *Draft B* puts it, the angles and figures I contemplate may be ‘drawn upon paper, carved in marble, or only fancied in my understanding’ (*Drafts* vol. 1: 152). Consequently Locke often talks as if we can *literally* perceive a necessary relation between ideas. Another difference from Descartes, as also from Hobbes, is that he rejects the pretensions of proposed analytical methods to uncover self-evident principles from which the phenomena can be deduced.

The third degree of knowledge is sensitive knowledge of the existence or ‘co-existence’ of qualities in external things. Locke’s first introduction of this category seems tentative, even an afterthought, as if it is called knowledge only by courtesy. In order to fit his main definition of knowledge it has to be interpreted as the perception of the agreement of ideas of sensible qualities with the idea of existence, an analysis Locke unsurprisingly declines to develop. Yet ‘sensitive knowledge of existence’ does straightforwardly satisfy his other definition of knowledge: what is known in sensitive knowledge (that is, that something external is causing an idea of sense) is known directly, ‘perceived’ and not inferred (see §2 of this entry). Locke was writing in a context in which, despite Gassendi’s Epicurean claim that sensory knowledge is the most evident of all, it was widely assumed that knowledge in the full sense comprises only knowledge of necessary first principles, demonstrated ‘science’, and perhaps reflexive knowledge. Locke wanted both to concede to orthodoxy that the evidence and certainty of our sensory knowledge is not as high as that of intuition and demonstration, and to insist that, nevertheless, ‘sensitive knowledge of existence’ does give a degree of immediate certainty and ‘deserves the name of knowledge’.

Knowledge is also categorized in terms of four propositional relations (forms of ‘agreement’) between ideas, namely ‘identity (or diversity)’, ‘relation’, ‘necessary connection or coexistence’ and ‘existence’ (*Essay* IV.i). By ‘identity’ Locke intends tautologies such as ‘Gold is gold’ and ‘Red is not blue’. Intuitive knowledge of such identities is achieved simply by discerning ideas. The category also includes such truths as ‘Gold is a metal’ or ‘Gold is malleable’, when the property predicated is included in the thinker’s definition of gold. Thus ‘identity’ covers all and only ‘trifling’ or ‘verbal’ propositions (see §2 of this entry).

The categories, ‘relation’, ‘necessary connection or coexistence’, and ‘existence’, on the other hand, together include all ‘instructive’ propositions. The category, ‘relation’, in part a response to Locke’s earlier difficulty over the informativeness of mathematics (see §2 of this entry), also marks his rejection of analytical methods in science. As well as geometrical axioms and theorems, ‘relation’ presumably includes more exciting Lockean principles: as that, if anything changes, something must have a power to make it change; that, if anything exists, something must have existed from eternity; and that a maker has rights over his artefact. Categorical propositions about natural things, however, fall either under ‘existence’ or under ‘necessary connection or coexistence’. Our own existence is known intuitively, God’s existence demonstratively (Locke employs an idiosyncratic hybrid of the cosmological

and teleological proofs), and, as discussed, the existence of bodies by sense. The category ‘necessary connection or coexistence’ owes its disjunctive name to a rather complicated relation between particular and universal propositions. Particular *coexistences* are perceived by sense, for example, when we simply observe that yellowness, heaviness and the metallic qualities coexist in a particular subject together with malleability (that is, that this gold is malleable), without perceiving *necessary connections* between them. Locke assumes, however, with most mechanists, that necessary connections do hold between universally coexistent properties even if we cannot perceive or grasp them. Since he contends that no natural science based on the essences of substances has been achieved, he offers only very limited examples of perceived necessary connections, as ‘whatever is solid is impenetrable’ and ‘a body struck by another will move’. (According to the short, posthumously published *Elements of Natural Philosophy*, the laws of inertia are evidently necessary, but the law of gravity is based only on experience.) In the absence of knowledge, beliefs in universal coexistences (for example, that all gold dissolves in aqua regia), when we *presume* unperceived connections, may be inductively based on sensitive knowledge of particular coexistences. That is descriptive natural ‘history’, not ‘science’. In general, if the idea of a particular quality is deducible from the idea of a substance, that is only because the predication of that quality is an identity: that is, universal propositions about substances, if certain, are ‘trifling’ and, if ‘instructive’, are uncertain (see §2 of this entry). In contrast, instructive a priori sciences are possible just because their objects are constructed by us: our ideas of simple or mixed modes, formed without essential reference to actuality, themselves constitute the subject-matter of mathematics and ethics. In other words, these demonstrative sciences are possible, as natural science is not, just because they deal, hypothetically, with abstractions.

The degrees of assent are ‘Belief, Conjecture, Guess, Doubt, Wavering, Distrust, Disbelief, etc.’ (*Essay* IV.xvi). Probability is ‘the measure whereby [the] several degrees [of assent] are, or ought to be regulated’. When assent is unreasonable, it constitutes ‘error’. Reasonable assent is regulated according to the proposition’s conformity with the thinker’s own experience or the testimony of others. The proposition may concern ‘matters of fact’ falling within human experience, or else unobservables lying ‘beyond the discovery of our senses’. Locke identifies four broad degrees of probability with respect to ‘matters of fact’: (1) when the general consent of others concurs with the subject’s constant experience; (2) when experience and testimony suggest that something is so for the most part; (3) when unsuspected witnesses report what experience allows might as well be so as not; and (4) when ‘the reports of history and witnesses clash with the ordinary course of nature, or with one another’ - a situation in which there are no ‘precise rules’ for assessing probability. Finally, with respect to unobservables, ‘a wary reasoning from analogy’ with what falls within our experience ‘is the only [natural] help we have’ and the only ground of probability (see Descartes, R. §4). Although Locke, in striking contrast to Descartes, brings probability into the centre of epistemology, ‘belief’ is always treated as a practical surrogate for ‘knowledge’, and he takes induction itself to be grounded on the assumption of underlying, unknown necessary connections: ‘For what our own and other men’s constant observation has found always to be after the same manner, that we with reason conclude to be the effects of steady and regular causes, though they come not within the reach of our knowledge’ (*Essay* IV.xvi.6).

Another deliberate and radical difference from Descartes relates to the role of will in cognition. For Locke, knowledge is like sense perception: we may choose where and how hard to look, but we cannot then choose what we see. Belief is similar: ‘Assent is no more in our power than knowledge…. And what upon full examination I find the most probable, I cannot deny my assent to’ (*Essay* IV.xx.16). Yet we are morally responsible for both error and ignorance in so far as it results from our not employing our faculties as we should. In a number of chapters of the *Essay*, Locke examines the causes of error, finding them, with many writers of his time, in the same appetites, interests, passions, wayward imaginings and associations of ideas as may motivate voluntary actions. Linguistic confusion, and its deliberate exploitation (see §5 of this entry), sometimes plays a role, and sometimes, like Malebranche and others, Locke has direct recourse to physiology, explicitly merging his explanations of error with explanations of madness. In contrast to Hobbes, he places the merely habitual ‘association of ideas’ in the pathology of ‘extravagent’ thought and action: ‘all which seems to be but trains of motion in the animal spirits, which…continue on in the same steps…which by often treading are worn into a smooth path, and the motion in it becomes easy and as it were natural’ (*Essay* II.xxxiii.16). But culpable error arises, on Locke’s official view, when we ‘hinder both knowledge and assent, by stopping our enquiry, and not employing our faculties in the search of any truth’ (*Essay* IV.xx.16). It is the failure to use our power to pause for ‘full examination’ which leaves a space for beliefs motivated by interests and passions. But this two-stage model - the first stage voluntary, the second

involuntary - proves too difficult to maintain, and sometimes passions and interests are taken to act on the will *between* enquiry and judgment, by distorting our ‘measures of probability’ themselves. Locke’s approach is more common-sensical than that of Descartes, but the psychology of motivated error is a hard nut which he also failed to crack.

7 Faith, reason and toleration

Locke’s views on belief, probability and error owed much to traditional philosophy of religious belief, and to the great debate of his century about the relationship between faith and reason. He was strongly influenced by writers in the Anglican ‘probabilist’ tradition, who argued for toleration within the Church with respect to all but an essential core of Christian dogma. William Chillingworth had rejected as absurd the traditional conception of a moral requirement to have ‘faith’ in the sense of a conviction equal to that of knowledge but beyond what is rationally justified. To recognize a proposition as probable to a certain degree is to believe it just to that degree. Revelation therefore cannot be a basis for belief distinct from probability, but is something the significance of which has to be rationally assessed, capable at best of increasing the probability of certain propositions. Similarly for Locke, when revelation grounds belief that would otherwise be improbable, that is just one natural reason outweighing another: ‘it still belongs to reason to judge of the truth of its being a revelation, and of the signification of the words, wherein it is delivered’ (*Essay* IV.xviii.8). For if ‘reason must not examine their truth by something extrinsical to the perswasions themselves; inspirations and delusions, truth and falshood will have the same measure’ (*Essay* IV.xix.14).

Accordingly, like Chillingworth, More and others, Locke combined a purportedly reasonable acceptance of the Bible as revelation with a critical approach to its interpretation, taking into account that it was written by men in particular circumstances. An alleged revelation which conflicts with what is naturally evident loses its claim to be revelation. Certain revealed truths (such as the Resurrection) lie ‘beyond the discovery of our natural faculties, and above reason’, but Locke had little time for mysteries: ‘to this crying up of faith, in opposition to reason, we may, I think, in good measure, ascribe those absurdities that fill almost all the religions which possess and divide mankind’ (*Essay* IV.xviii.7, 11). Locke took the existence of God and the content of moral law to be demonstrable by reason, and, according to *The Reasonableness of Christianity* and its *Vindications*, the only essentially revealed truth of the New Testament is that Christ is the Messiah, promising forgiveness of sins to those who sincerely repent and do their imperfect best to keep the law of nature. The Bible also makes that law plain to those without the leisure or capacity to reason it out - a difficult enough task for anyone, as Locke ruefully acknowledges. The meaning of scripture is thus for Locke primarily moral, and the ‘truth, simplicity, and reasonableness’ of Christ’s teaching is itself a main reason for accepting it as revelation. Saving faith involves works, not acceptance of ‘every sentence’ of the New Testament under this or that preferred interpretation.

Much the same goes for immediate revelation. Even the genuinely inspired would need proofs that they really were inspired, and the errors of commonplace ‘enthusiasm’ are ascribed, as by More, to physiology, ‘the conceits of a warmed or overweening brain’. The advocate of immediate personal revelation over reason ‘does much what the same, as if he would persuade a man to put out his eyes the better to receive the remote light of an invisible star by a telescope’ (*Essay* IV.xix.4). Divine illumination necessarily depends on, and is not separable from, the natural light - ‘reason must be our last judge and guide in everything’. Locke echoes Chillingworth’s basic principle: the lover of truth, unbiased by interest or passion, will not entertain ‘any proposition with greater assurance than the proofs it is built upon will warrant’. The implication of this standard, in the actual circumstances of life, is toleration: ‘For where is the man, that has incontestable evidence of the truth of all that he holds, or of the falshood of all he condemns; or can say, that he has examined, to the bottom, all his own, or other men’s opinions?’ (*Essay* IV.xvi.4).

Locke’s *Letter on Toleration*, the mature fruit of considerably more unpublished writing directly on the issue, links his epistemology with his political thinking. Belief is not something that can be commanded or submitted to the authority of the government, whose concern is not with saving souls but the preservation of property. Necessarily each individual must judge as they see fit, and the truth needs no help, having its own efficacy. But the right to toleration is nevertheless viewed in the context of the right and duty to seek salvation and true doctrine without harm to others, harm which is at least threatened by all who deny the authority either of moral law or of the established government. Atheists therefore forfeit the right in principle, and Roman Catholics as a matter of

political fact. (See Latitudinarianism; Socinianism.)

8 Personal identity

The main aim of the chapter of the *Essay* entitled ('Of Identity and Diversity' (II.xxvii) is to explain how immortality is compatible with materialism. In order to maintain an agnostic neutrality on the question of the immateriality of the soul, Locke had, first, to rebut the Cartesian claim that self-awareness supplies a clear and distinct idea of a simple, continuously existing substance; and, second, to show that the metaphysical issue is irrelevant to 'the great ends of morality and religion' (*Essay* IV.iii.6). He argues that, although the moral agent is indeed the continuously existing, rational, self-aware subject of consciousness, the 'person', the identity of this subject over time is determined by the continuity of unitary consciousness itself, not the continuity of an immaterial soul. Locke can therefore accept the Resurrection and Last Judgment as tenets of his 'reasonable' Christianity, without commitment to dualism, on the supposition that the consciousness of the resurrected person is continuous, through memory, with that of the person who died. This conclusion avoids an objection to his concept of demonstrative ethics as a science of modes, that morality relates to 'man', a substance, not a mode. His response is that morality concerns, not 'man' as a biological species, but 'man' as rational, the 'moral man', indeed all rational beings. 'Person', as he puts it, is a 'forensic term'.

Locke's argument starts from the claim that questions of identity over time are always questions as to the continuous existence in space of something of a certain kind, and that difficulties may be avoided by 'having precise notions of the things to which it is attributed'. The identity of non-substances is parasitic on that of substances: 'All other things being but modes or relations ultimately terminated in substances', their identity will be determined 'by the same way' (*Essay* II.xxvii.2). Locke holds that events and processes ('actions') are not strictly identical from moment to moment, each part of what we consider one process being distinct from every other part. Substances, however, genuinely continue to exist from moment to moment. The identity of 'simple substances' - material atoms and the presumed simple 'intelligences' - is straightforward. Each excludes others of the same kind out of its place by its very existence - a principle definitive of identity. But difficulties arise with compound substances. Strictly, a body composed of many atoms is the same just as long as the same atoms compose it - yet 'an oak, growing from a plant to a great tree, and then lopped, is still the same oak'. Locke's explanation is that 'in these two cases of a mass of matter, and a living body, *identity* is not applied to the same thing' (*Essay* II.xxvii.3). Although he does not clearly distinguish the two views, he seems to hold individuation, rather than the identity-relation, to be kind-relative. A plant or animal is not just 'a cohesion of particles anyhow united', but such an organization of parts as enables the continuation of its characteristic life, for example, as an oak. In fact the *species* of the living thing is irrelevant to Locke's theory (fortunately, given his view that the definition of 'oak' will differ from speaker to speaker). The essential claim is that life is a principle of unity and continuity distinct from simple cohesion, thus allowing a living thing and the mass of matter that momentarily composes it so to differ in kind as to be capable of occupying the same place at the same time.

Locke defines *person* as 'a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing at different times and places' (*Essay* II.xxvii.9). His thesis is that, just as life constitutes a distinct individuative principle of unity and continuity, so does reflexive consciousness. He argues for the logical independence of the continuity of consciousness from both the continuity of substance (whether supposed material or immaterial, simple or complex) and the continuity of animal life by a series of imagined cases: for example, for someone now to possess Socrates' soul would not make him the same person as Socrates, unless he remembered Socrates' actions as his own; whereas if souls are the seat of consciousness, and the soul of a prince could migrate to the body of a cobbler, 'everyone sees he would be the same person with the prince, accountable only for the prince's actions. But who would say it was the same man?' (*Essay* II.xxvii.15). Locke viewed such cases, not necessarily as real possibilities, but as compatible with our partial understanding of things, our *ideas*: 'for such as is the idea belonging to that name [namely 'person', 'man' or 'substance'], such must be the identity'. Yet in the crucial case of the Resurrection, we are left wondering how continuous existence through time - not to speak of space - is achieved simply by a fit between present consciousness and past experience and actions. Indeed, as Berkeley and Reid argued, memory-links seem both too little and too much for the continuity of a substantial thing. Yet, despite these and other difficulties for Locke's theory, it set the agenda for subsequent discussion and versions of it still have adherents (see Personal identity).

9 Ethics, motivation and free will

With Locke's conviction that a demonstrative ethics is possible went a belief that what stood in its way was the deplorable slipperiness, openly encouraged by the practice of rhetoric, of a moral language in which terms are not consistently tied to ideas (see §5 of this entry). Both were consonant with his apparently early conviction that Natural Law theory, as pursued by such as Hooker and Grotius, is capable of development into a full account of our duties to God and our fellows - even though he had first seen Natural Law as empirically based (see §3 of this entry). But Natural Law theory also gave him what could not be supplied by the conception of a quasi-geometrical system of rights and duties flowing from the definitions of mixed modes and relations: the conception of an unconditional obligation to act in accordance with moral principle against what we might otherwise desire (see Natural Law).

In the *Essay* the argument starts, as might be expected, with the question of how our basic concepts of value are derived from experience. Locke has no doubt about what it is in experience that makes anything matter to us. Like other empiricists of his time, he is both a psychological and an ethical hedonist. Pleasure and pain supply not only our sole motives but also our ideas of good and evil: 'That we call *good* which is *apt to cause or increase pleasure*, or *diminish pain in us; or else to procure, or preserve us the possession of any other good, or absence of any evil*' (*Essay* II.xx.1). The passions are 'modes' of pleasure and pain arising from, or involving, value-judgments: thus hope is 'that pleasure in the mind, which every one finds in himself, upon the thought of a probable future enjoyment of a thing, which is apt to delight him'; fear is 'an uneasiness of the mind, upon the thought of a future evil likely to befall us' (*Essay* II.xx.9-10). Desire is the 'uneasiness a man finds in himself upon the absence of any thing, whose present enjoyment carries the idea of delight with it' (*Essay* II.xx.6). This theory of motivation faces certain problems. First, how do we get from judgments of good and evil, of what conduces to pleasure and pain, to judgments of right and wrong, of what we morally ought or ought not to do? Second, having got there, if the passions, as modes of pleasure and pain, constitute our only motives, what passion could motivate us to do what is right? Third, in what, if anything, does choice and free will - moral agency - consist?

Locke's answer to the first question, already given in *Essays on the Law of Nature*, is that the concept of obligation comes with the relational concept of law: '*Morally good and evil* then, is only the conformity or disagreement of our voluntary actions to some law, whereby good or evil is drawn on us, from the will and power of the lawmaker; which good and evil, pleasure or pain, …is that we call *reward* and *punishment*' (*Essay* II.xxviii.5). Locke makes it clear that the notion of obligation presupposes the right, as well as the power of the lawgiver to legislate and punish - in *Essays* Locke's 'power' is explicitly *potestas*, authority, rather than *potentia*, mere force. There are, he says in the *Essay*, three kinds of law: divine law, the measure of sin and moral duty; civil law, the measure of crimes and innocence; and the law of opinion or reputation, the measure of virtue and vice. God legislates by *ius creationis*, the maker's right over what is made, and divine law is binding on all rational creatures capable of pleasure and pain. God's law accords with his wisdom and benevolence, so that we can know it by reflecting on what a wise and benevolent Deity would require of us. Unsurprisingly, Locke's ethics is heavily utilitarian.

The relationship between divine law and civil law, and the standing of the civil magistrate under divine law, is the subject-matter of Locke's political theory as expounded in *Two Treatises*. The notion of a 'law of reputation', sometimes called 'the philosophical law', has a more complex role in his thought. It is Locke's explanation of popular secular morality, but it also represents his view of the possibility of non-theistic philosophical systems of ethics. Roughly, the thought is that ordinary morality, sanctioned by public approval and disapproval, exists as a means to the preservation of society, itself a condition of the happiness of individuals. As arrangements differ between societies, so do their moral concepts and what counts as virtue and vice in each, although naturally there will be overlap given their shared aim of self-preservation. Since the divine law too is concerned with the good of human beings, and with self-preservation as a duty, the law of reputation will tend to coincide with divine law. In the aborted fragment of the *Essay*, 'Of Ethick in General', Locke suggests that philosophers may have some inkling of the divine law, but they confuse it with the law of reputation. Consequently their systems reduce either (like that of Hobbes) to an advocacy of what tends to the preservation of society, or (like that of Aristotle) to the elaboration of a set of definitions of the behaviour of which a particular society approves or disapproves (King [1829] 1864: 308-13). Locke does not deny the social importance of the law of reputation, however, and in *Some Thoughts concerning Education* he assigns a necessary role in a child's moral education to public esteem and

shame. His complaint is that an explanation of moral obligation in terms of the value of certain actions to society, and the value of society to the individual, cannot explain how we may be morally obliged to do something contrary to our own felt interest - our interest, at least, in this world. Self-interest may commonly prescribe adherence to social rules, but it may not always do so. As Locke says in the *Essays on the Law of Nature* ([1664] 1954: 204) 'a great number of virtues, and the best of them, consist only in this: that we do good to others at our own loss'.

Locke's position is, then, that in order to explain both moral obligation and moral motivation (conflated in the usual seventeenth-century notion of obligation), we need to see morality as a system of laws prescribed by a supremely rational, just and benevolent creator to whom we owe the duty of obedience as creatures, and whose power to reward and punish in the next life is capable of motivating anyone who duly considers it (see Voluntarism). Like any theistic explanation of morality's binding force, this proposal is incoherent, and in its case the incoherence lies in the combination of the view that obligation is created by law with the claim that we have a natural obligation to obey the law of our creator. Locke, however, was more exercised by the problem of why consideration of the afterlife so often fails to move theists to do their duty. Indeed, he accorded the problem wider scope, since he followed Pascal in the thought that the bare possibility of there being an afterlife, given the infinite good at stake, ought in reason to motivate the Christian life (see Pascal, B. §6). Locke's explanation of the human capacity to know the better and choose the worse involved a refinement of his theory of motivation which echoes his theory of error. In the first edition of the *Essay*, in the long chapter 'Of Power', he held that 'the choice of the will is everywhere determined by the greater apparent good' (*Essay* II.xxi.70). By the second, he believed that mere consideration of future benefit will not move us to action unless it gives rise to 'an *uneasiness* in the want of it' - that is, to desire. Only a present passion - and, it seems, a kind of pain - can move to present action. It may require some reflection on the situation, over and above the simple recognition of probable or possible consequences for good or ill, to bring desire up to scratch, and to 'suit the relish of our minds to the true intrinsic good or ill, that is in things'. Someone who sees the good but does not pursue it has not reflected enough: 'Morality, established upon its true foundations, cannot but determine the choice of any one that will but consider' (*Essay* II.xxi.70).

Locke's increased emphasis on the role of deliberation in his hedonistic theory of moral motivation complicates his much revised account of liberty. He adopted a self-determinist view of free will - a free action is not one that is causally undetermined, but one determined by the agent's 'own desire guided by his own judgement'. He defines 'liberty' as 'the power to act or not to act according as the mind directs' (*Essay* II.xxi.71). But another power became increasingly important to him, the power 'to stand still, open the eyes, look about, and take a view of the consequence of what we are going to do, as much as the weight of the matter requires' (*Essay* II.xxi.67), and it is in this power, he often suggests, that the liberty of rational agents really consists. The tension is unresolved, for Locke never retracts the rhetorical question to which he himself seems to have given an answer: 'For how can we think anyone freer than to have the power to do what he will?' (*Essay* II.xxi.21) (see Free will).

10 Political theory

Locke's mature political theory is set out in 'An Essay concerning the True Original, Extent, and End of Civil Government', the second of *Two Treatises of Government*, the first being a point-by-point rebuttal of Robert Filmer's biblically based patriarchalism (see Filmer, R.). Locke's primary contention is that the right to govern comes with a duty to govern in the interest of the governed. Failure by the government to recognize or observe this duty creates the right to rebel. Like the Natural Law theories of Hooker, Grotius and Pufendorf on which he draws, Locke's argument moves from first principles, in effect a fragment of his proposed demonstrative ethics; but much of its richness derives from links with his practical political concerns and interests. It presents attitudes and actions attributable to Charles II and James II as a betrayal of trust, hostile to those features of the British constitution most adapted to the essential purposes of government; but he also states principles relating to property, money, social conventions, taxation, punishment, family relations, inheritance, the rights of the poor, enclosure of land, the practice and justification of colonial settlement, and more.

Filmer had argued that both political authority and property rights exist only by divine institution - by God's giving Adam dominion over the creatures, by the subjection of Eve, and by Adam's natural paternal rights over his children. Monarchs are deemed natural inheritors of Adam's rights. A part of Locke's strategy, pursued in both Treatises, against this doctrine was to drive wedges between the possession and inheritance of property and the

possession and conveyance of authority, and between paternal (or, as Locke prefers, parental) authority and political authority. For example, the right of children to inherit their parents' property stems from their natural right (not just that of the eldest child) to sustenance by their parents, a right which cannot be supposed to embrace either patriarchal authority or political power. The analogy of power and property in Filmer's argument, however, was not only in relation to inheritance, for it entailed that individual ownership is simply a grant of use by the king, making taxation - its partial withdrawal - his personal right. Locke was therefore concerned to give property a quite different role in his explanation of political society.

For Locke, government is a human invention, to which personal property is prior. In a state of nature, he argues in the Second Treatise, human beings have an obligation, in accordance with divine or natural moral law, 'to preserve the rest of mankind', their equals as creatures and servants of God, by a rational extension of their duty to preserve themselves. More specifically, 'no one ought to harm another in his life, health, liberty or possessions' (*Two Treatises* II.6). Yet, before government, 'everyone has a right to punish the transgressors of that law to such a degree, as may hinder its violation'. The 'state of nature' is not, for Locke, a merely ideal abstraction, but a historical situation in which members of simple societies have lived and still live, unless in time of war, and in which independent national governments always necessarily exist. For international relations are not governed by positive law prescribed and sanctioned by constituted authority. In this situation the victim of aggression - or indeed any onlooker, for the violation is of the natural law which maintains the welfare of all - has the right to destroy the aggressor until offered peace, reparation and security for the future. Within civil society itself this 'right of war' or self-defence exists whenever the law cannot be effectively exercised, whether in the immediate circumstances of threatened harm, or when the administration of the law is manifestly corrupt, and itself employed to commit violence and injury.

'Liberty' in the state of nature is freedom from any constraint but the moral law of nature. Under government, it is freedom from the 'arbitrary will of another man', and from any human rule but the 'standing rule…common to everyone of that society'*Two Treatises* II.22). (Locke sees slavery as continuation of war - it is just if the war is just, when it is in lieu of capital punishment, the justly enslaved, like criminals, being 'outside civil society'. Yet this hardly stands as an endorsement of contemporary colonial slavery - indeed Locke denies that the children of aggressors can be justly enslaved, or even disinherited.) 'Possessions' arise in a state of nature with the act of appropriation which is a necessary condition of the use of any of the comestibles naturally available to all: 'this law of reason makes the deer, that Indian's who hath killed it' (*Two Treatises* II.30). Such appropriation is an extension of the principle that 'every man has a property in his own person', and therefore in 'the labour of his body'. Consequently whatever someone has 'mixed his labour with' is his, provided that it is for use, and 'there is enough, and as good left in common for others' (*Two Treatises* II.27). This principle applies also to the enclosure of land for agriculture, which vastly increases its productivity. With land, as in all else, 'labour makes the far greatest part of the value of things, we enjoy in this world' (*Two Treatises* II.42). Nevertheless, before the conventional use of money, no one would have either motive or right to produce more than they could use, give to others or exchange before it spoils. To take something from the common store and let it spoil is against natural law. Money, however, is an artifice which modifies the whole nature of property-rights, since it can be stored indefinitely without spoiling. Money makes it worthwhile to exploit land fully, and supplies a just means of keeping the product. So far from wronging others, enclosure and improvement greatly increase 'the common stock of mankind', making 'a day labourer in England' better off than a king among the (Native) Americans (*Two Treatises* II.41). Significant disparity of wealth becomes both possible and morally justified, on the assumption that none will suffer absolutely (in a Board of Trade paper Locke simply asumed that everyone should have 'meat, drink, clothing and firing…out of the stock of the kingdom, whether they work or no' - Bourne 1876 vol. II: 382). But the effect is greatly to complicate the administration of the law of nature, and to render its application uncertain, as well as to encourage its breach through greed.

All this, on top of the standing need for both impartiality and sufficient force to punish malefactors, necessitates government. The chief role of such government is to determine rules to order and preserve property. Common defence is another imperative. Government with such legislative and executive powers comes into existence when people, by consent, resign their 'executive power of the law of nature… to the public' (*Two Treatises* II.89). Each individual member gives consent, but is thereafter bound to move with the majority. To the objection that no such agreement has ever taken place, Locke argues that, although 'government is everywhere antecedent to records',

cases abound of new or primitive societies with elected leaders. In the first instance, this may be 'some one good and excellent man' or effective general, or indeed the father of a familial group, but experience of unrestrained monarchy encourages legislatures of 'collective bodies of men', with none above the law. In any case, consent is normally tacit, and given in the active enjoyment of the benefit of the law, whether by possession of land or 'barely travelling freely on the highway'. Such tacit consent obliges obedience to the law, although the obligation lasts only as long as the enjoyment, leaving the individual free to give up the benefit and 'incorporate himself into any other commonwealth'. Express consent, however, binds the individual to obey and assist a particular government until its dissolution (or breach of trust).

A subject's ultimate obligation is to the supreme power, which is the legislative, itself bound by the law of nature in its choice of means, 'established and promulgated laws', for the preservation of its subjects and their property. Given this role, a government has no right to tax its subjects without their consent 'either by themselves, or their representatives chosen for them'. In order to minimize the risk of the legislative acting in its own, rather than in the public interest, it is best that it be an assembly which meets from time to time, separate from the continuously acting executive. A third, 'federative' power of war, peace and alliances is less easily directed by antecedent laws than the executive power, but falls naturally into the same hands, since both depend on public force. Locke allows some qualification of the absolute separation of powers, and subordination of the executive to the legislative, in recognition of the 'prerogative' power of the English king to dissolve and convene Parliament as circumstances require, and to employ discretion in the execution of the laws (Locke notes without express approval the power to veto legislation). Yet Locke sees prerogative as justified only as falling under 'the power of doing public good without a rule' in the face of unforeseen circumstances, and as dangerously capable of abuse. Its continuous employment contrary to the public good, for example by refusing to convene the legislative or by tampering with the rules for its election, makes the king himself a rebel and destroyer of the government, at war with his own subjects, returning them to a state of nature with a right to set up a new government.

11 Influence

Perhaps no modern philosopher has had a wider influence than Locke. His immediate achievement was, with Newton, to bring to an end the dominance within Europe of Cartesian science and philosophy, unseating the broadly Neoplatonic notion that mind and world share a common, divinely imposed structure, in favour of a modest, naturalistic conception of human capacities. Careful observation and systematic description are more valuable than the construction of hypotheses purportedly achieved by super-experiential means. Locke's own 'historical' treatment of the mind as a familiar, describable but deeply mysterious part of nature had considerable influence on European thought. His theory of classification influenced later taxonomy, and his brilliantly original theory of personal identity is still a standard text for philosophical discussion. His philosophy was one of the chief influences on Kant, but can still suggest an alternative to Neo-Kantian conceptualism. If his ethical theory appears to be the last throes of early modern natural law theory rather than a new beginning, within that structure he enunciated a classic justification of responsible, tolerant and broadly democratic political society which has remained a major resource for political theorists ever since.

See also: Empiricism; Rationalism

MICHAEL AYERS

List of works

Locke, J. (1823) *The Works of John Locke*, London: T. Tegg *et al.*, 11th edn, 10 vols; repr. Aalen: Scientia, 1963. (Still, in 1997, the most complete edition of Locke's works, together with some of the more philosophically interesting correspondence.)

Locke, J. (1975-) *The Clarendon Edition of the Works of John Locke*, general eds P.H. Nidditch (to 1983), J.W. Yolton (1983-92), M.A. Stewart (1992-), Oxford: Oxford University Press.(A critical edition, planned to include all the published works and significant manuscript material, the bulk of which is in the Bodleian Library, Oxford. Volumes are available or in active preparation in 1997 as indicated under individual titles below.)

Locke, J. (*c*.1660-2a) *Two Tracts on Government*, ed. P. Abrams, Cambridge: Cambridge University Press, 1967. (Useful introduction and notes. The 'tracts', the first in English, the second in Latin, debate 'Whether the Civil Magistrate may lawfully impose and determine the use of indifferent things in reference to Religious

Worship?' Contrary to his later stance, Locke argues in these early manuscripts for the right of the magistrate to regulate religious observance for the sake of public peace.)

(*c*.1660-2b) 'An necesse sit dari in Ecclesia infallibilem Sacro Sanctae Scripturae interpretem?' (Is it necessary to have in the Church an infallible interpreter of Holy Scripture?), ed. J.C. Biddle in 'John Locke's Essay on Infallibility: Introduction, Text, and Translation', *Journal of Church and State* 19 (1977): 301-27.(A critique of the notion of infallibility.)

Locke, J. (1664) *Essays on the Law of Nature*, ed. and trans. W. von Leyden, Oxford: Oxford University Press, 1954; ed. and trans. R. Horwitz, J. Strauss Clay and D. Clay as *Questions concerning the Law of Nature*, Ithaca, NY: Cornell University Press, 1990.(An early manuscript in Latin in scholastic form which, despite later change in Locke's conception of moral knowledge, throws light on his mature moral theory as well as the development of his thought. Von Leyden's edition includes related manuscript material.)

Locke, J. (1667) *An Essay concerning Toleration, in John Locke: Writings on Politics, Law and Religious Toleration 1667-1683*, ed. J.R. Milton and P. Milton, Clarendon Edition, Oxford: Oxford University Press, forthcoming.(This manuscript signals Locke's change of mind to a more liberal view of religious toleration.)

(?) (*c*.1668) 'De Arte Medica', in H.R.F. Bourne *The Life of John Locke*, 2 vols, London: Henry S. King, 1876, vol.1, 222-7; repr. in K. Dewhurst *Dr Thomas Sydenham (1624-1689): His Life and Original Writings*, London: Wellcome Historical Medical Library, 1966.(A manuscript in Locke's handwriting, but possibly by Sydenham, expounding a strongly empirical method in medicine.)

(?) (1671) *Drafts for the Essay concerning Human Understanding, and Other Philosophical Writings*, vol. 1, ed. P.H. Nidditch and G.A.J. Rogers, Clarendon Edition, Oxford: Oxford University Press, 1990-.(Volume I contains *Drafts A and B* (1671); the forthcoming volumes II and III contain *Draft C* (1685) and associated manuscript material. Together they cast considerable light on the development and significance of Locke's general philosophy, including projected, but unfinished chapters of the *Essay*.)

Locke, J. and Tyrrell, J. (1681-3) *Locke on Separation*, ed. J. Marshall, Clarendon Edition, Oxford: Oxford University Press, forthcoming.(A jointly written, hitherto unpublished response to Edward Stillingfleet's *The Mischief of Separation* (1680) and *The Unreasonableness of Separation* (1681) in which Locke and Tyrrell defend toleration of religious nonconformity.)

Locke, J. (1686) 'Newton, Is. Philosophiae naturalis principia mathematica…' (Mathematical principles of natural philosophy), *Bibliothèque universelle et historique*.(Review, by an admiring layman, of Newton's *Principia*.)

Locke, J. (1688) 'Abrégé d'un ouvrage intitulé Essai philosophique touchant l'entendement', *Bibliothèque universelle et historique*; separately published as *Extrait d'un livre anglois… intitulé Essai philosophique concernant l'entendement…*, Amsterdam, 1688; trans. as 'An Extract of a Book, Entituled, A Philosophical Essay upon Human Understanding', in *The young-students-library, containing extracts and abridgements…*, London, 1692; in *Drafts for the Essay concerning Human Understanding, and Other Philosophical Writings*, vol. 3, ed. P.H. Nidditch and G.A.J. Rogers, Clarendon Edition, Oxford: Oxford University Press, 3 vols, 1990-.(Locke's effective abstract of *An Essay concerning Human Understanding*.)

Locke, J. (1689) *Epistola de Tolerantia*, Gouda; ed. R. Klibansky *Epistola de Tolerantia: A Letter on Toleration*, trans., intro. and notes J.W. Gough, Oxford: Oxford University Press, 1968; trans W. Popple as *A Letter concerning Toleration*, London, 1689, 2nd edn, 1690, ed. J. Tully, Indianapolis, IN: Hackett Publishing Company, 1983.(Locke's classic argument for religious toleration. Popple's vigorous translation is the one through which the work has chiefly been known to English-speakers.)

Locke, J. (1689/90a) *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Clarendon Edition, Oxford: Oxford University Press, 1975.(Originally published in December 1689 but carrying the date 1690. Locke's chief and greatest work, arguing comprehensively that what we can think and know is limited by the way we experience the world, attacking dogmatic pretensions to grasp the essences of things, and affirming that 'reason must be our last judge and guide in everything', including morals and religion.)

Locke, J. (1689/90b) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1960, 2nd edn, 1967; ed. I. Harris, with associated manuscript material, Clarendon Edition, Oxford: Oxford University Press, forthcoming.(In the First Treatise 'the False Principles and Foundation of Sir Robert Filmer and His Followers are Detected and Overthrown' - that is, the patriarchal theory of monarchy. The Second Treatise, 'an Essay concerning the True Original, Extent, and End of Civil-Government', is a major classic of political theory, arguing that government is morally, and should be constitutionally, answerable to the governed. Laslett's influential introduction stimulated continuing debate as to the immediate context and purposes of *Two Treatises*, and its relation to the rest of Locke's thought.)

Locke, J. (1690) *A Second Letter concerning Toleration*, London; repr. in *The Works of John Locke*, vol. 6, London, 11th edn, 1823.(Locke defends religious toleration, and his argument of *Epistola de Tolerantia*, in

response to Jonas Proast's *The Argument of the Letter concerning Toleration, Briefly Consider'd and Answer'd*.)

Locke, J. (1691/2) *Some Considerations of the Consequences of the Lowering of Interest, and raising the Value of Money*, in *Locke on Money*, ed. P.H. Kelly, Clarendon Edition, Oxford: Oxford University Press, 1991.(Advice on the coinage which became Government policy.)

Locke, J. (1692) *A Third Letter for Toleration, to the Author of the Third Letter concerning Toleration*, London; repr. in *The Works of John Locke*, vol. 6, London, 11th edn, 1823.(Continues the argument of *Epistola de Tolerantia* and *A Second Letter concerning Toleration*, against Jonas Proast).

Locke, J. (1693) *Some Thoughts concerning Education*, London; 3rd enlarged edn, London, 1695; ed. J.W. Yolton and J.S. Yolton, Clarendon Edition, Oxford: Oxford University Press, 1989.(Consciously modest in scope, but an important and interesting work in the history of educational theory.)

Locke, J. (1695a) *The Reasonableness of Christianity*, London; 2nd edn, 1696; ed. J. Higgins-Biddle, Clarendon Edition, Oxford: Oxford University Press, 1998.(Locke's latitudinarian credo, cutting articles of faith to a minimum and emphasising the moral dimension of Christianity.)

Locke, J. (1695b) *A Vindication of the Reasonableness of Christianity, etc. from Mr Edwards's Reflections*, London; repr. in *The Works of John Locke*, vol. 7, London, 11th edn, 1823.(The first of Locke's responses to John Edwards' attack on *The Reasonableness* in *Some Thoughts concerning the Several Causes and Occasions of Atheism, Especially in the Present Age*, 1695.)

Locke, J. (1697a) *A Second Vindication of the Reasonableness of Christianity,…*, London; in *The Works of John Locke*, vol. 7, London, 11th edn, 1823.(A reply to John Edwards' response to *A Vindication of the Reasonableness of Christianity*.)

Locke, J. (1697b) *A Letter to the Right Reverend Edward, Lord Bishop of Worcester, Concerning some Passages relating to Mr. Locke's Essay of humane Understanding: In a late Discourse of his Lordship's, in Vindication of the Trinity*, London; in *The Works of John Locke*, vol. 4, London, 11th edn, 1823; repr. in *The Locke-Stillingfleet Debate*, ed. M.A. Stewart, Clarendon Edition, Oxford: Oxford University Press, forthcoming.(Stimulated by the theological objections of Edward Stillingfleet, Bishop of Worcester, Locke explains in particular, in this *Letter* and two further *Replies* (1697/1699), his theories of substance, real and nominal essence, and personal identity, and his agnostic attitude towards the issue between dualism and materialism. M.A. Stewart's edition will contain Stillingfleet's contributions to the debate.)

Locke, J. (1697c) *Mr Locke's Reply to the… Bishop of Worcester's Answer to his Letter…*, London; in *The Works of John Locke*, vol. 4, London, 11th edn, 1823; repr. in *The Locke-Stillingfleet Debate*, ed. M.A. Stewart, Clarendon Edition, Oxford: Oxford University press, forthcoming.(Continues the argument of *A Letter to the…Bishop of Worcester*.)

Locke, J. (*c*.1699) *The Elements of Natural Philosophy*, in *The Works of John Locke*, vol. 3, London, 11th edn, 1823.(An introductory survey of natural philosophy as Locke saw it, written for the son of his hosts, the Mashams.)

Locke, J. (1699) *Mr. Locke's Reply to the…Bishop of Worcester's Answer to his Second Letter…*, in *The Works of John Locke*, vol. 4, London, 11th edn, 1823; in *The Locke-Stillingfleet Debate*, ed. M.A. Stewart, Clarendon Edition, Oxford: Oxford University press, forthcoming.(Continues the argument of *A Letter to the…Bishop of Worcester, and Mr Locke's Reply….*)

Locke, J. (1700-4) *A Paraphrase and Notes on the Epistles of St Paul*, London, 6 vols, 1705-7; ed. A.W. Wainwright, Clarendon Edition, Oxford: Oxford University Press, 2 vols, 1987.(Locke's chief contribution to biblical hermeneutics, now published with associated manuscript material.)

Locke, J. (1702) *A Discourse of Miracles*, in *The Works of John Locke*, vol. 9, London, 11th edn, 1823.(Defines miracles as divinely purposive and contrary to common experience, but not as contrary to the unknown laws of nature).

Locke, J. (1704) *Fourth Letter on Toleration*, in *The Works of John Locke*, vol. 6, London, 11th edn, 1823. (Locke's last, unfinished shot, following those of 1690 and 1692, in defence of the position taken in *Epistola de Tolerantia* against Jonas Proast.)

Locke, J. (1976-) *The Correspondence of John Locke*, ed. E.S. de Beer, Clarendon Edition, Oxford: Oxford University Press, 8 vols, plus forthcoming index.(Around 3,650 letters from and to Locke in the period 1652-1704 supply important evidence of his life and times. A number of exchanges, for example, those with Molyneux and Limborch, are philosophically important.)

Locke, J. (1991) *Locke on Money*, ed. P.H. Kelly, Clarendon Edition, Oxford: Oxford University Press, 2 vols. (Contains all Locke's writings on economics (1668-95), with a long introduction explaining their significance.)

Locke, J. (forthcoming) *John Locke: Writings on Politics, Law and Religious Toleration 1667-1683*, ed. J.R.

Milton and P. Milton, Clarendon Edition, Oxford: Oxford University Press.(Contains the 1667 ‘Essay concerning Toleration’ and other manuscript writings from this crucial period of Locke’s association with Shaftesbury.)

Locke, J. (forthcoming) *The Journals of John Locke*, ed. H.A.S. Schankula, Clarendon Edition, Oxford: Oxford University Press, 4 vols.(Important evidence of Locke’s life, interests, reading and the development of his thought, 1675-1704. Some philosophically significant entries were previously published in *An Early Draft of Locke’s Essay*, ed. J.I. Aaron and J. Gibb, Oxford: Oxford University Press, 1936.)

References and further reading

Aaron, R. (1937) *John Locke*, Oxford: Oxford University Press, 3rd edn, 1971.(For long one of the two standard commentaries on the *Essay*, differing from Gibson (1917) in several important respects, some, but not all, improvements.)

Alexander, P. (1985) *Ideas, Qualities and Corpuscles: Locke and Boyle on the External World*, Cambridge: Cambridge University Press.(An extended consideration of the version of corpuscularianism favoured by Locke, and of its bearing on his epistemology and metaphysics.)

Ashcraft, R. (1986) *Revolutionary Politics and Locke’s Two Treatises of Government*, Princeton, NJ: Princeton University Press.(Makes strong, controversial claims about Locke’s active engagement in revolutionary politics during the writing of the *Two Treatises*.)

Ashcraft, R. (1987) *Locke’s Two Treatises of Government*, London: Unwin & Hyman, 1987.(An important analysis of the arguments of *Two Treatises*.)

Ashcraft, R. (ed.) (1991) *John Locke: Critical Assessments*, London: Routledge, 4 vols.(A large selective collection of some of the more significant articles written on Locke’s thought, a volume each on political theory, general philosophy, education and economics.)

Attig, J.C. (1985) *The Works of John Locke: a comprehensive bibliography from the seventeenth century to the present*, Westport, CT and London: Greenwood.(A list of editions and translations, including abridgements and selections, of Locke’s writings, together with other works, antecedent and precedent, constituting the immediate controversial context. Secondary works are cited in relation to publication details.)

Ayers, M. (1991) *Locke*, London: Routledge, 2 vols; repr. in 1 vol. as *Locke: Epistemology and Ontology*, London: Routledge, 1993.(The most comprehensive commentary on the *Essay*, interpreting and assessing Locke’s arguments in their intellectual context, but also offering detailed argument with respect to their continuing philosophical significance.)

Bourne, H.R. Fox (1876) *The Life of John Locke*, London: Henry S. King, 2 vols.(An important biography, but now largely outdated. Bourne made use of available documents, but lacked direct access to the manuscripts now in the Lovelace Collection in the Bodleian Library.)

Chappell, V. (ed.) (1994) *The Cambridge Companion to Locke*, Cambridge: Cambridge University Press.(A well-organized and useful collection of specially written articles covering Locke’s main philosophical concerns, with a good selective bibliography.)

Christophersen, H.O. (1930) *A Bibliographical Introduction to the Study of John Locke*, Oslo: Jacob Dybwad; repr. New York: Burt Franklin, 1968.(A pioneering, incomplete bibliography.)

Colman, J. (1983) *John Locke’s Moral Philosophy*, Edinburgh: Edinburgh University Press.(A pioneering analysis of Locke’s ethical theory which is careful, judicious and comprehensive in its analysis of Locke’s arguments about ethics, but Colman’s view of their relation to Locke’s theology is open to question.)

Cranston, M. (1957) *John Locke: A Biography*, London: Longmans; repr. Oxford: Oxford University Press, 1985.(Based rather narrowly on Locke’s then-recently rediscovered correspondence and journals, this is a readable account of his life, but is inaccurate on many details and unhelpful on his intellectual development.)

Dunn, J. (1969) *The Political Thought of John Locke: An Historical Account of the Argument of The Two Treatises of Government*, Cambridge: Cambridge University Press.(A classic study, placing Locke’s political theory in its historical and intellectual context and revealing some of the coherence of his thought that had escaped earlier commentators.)

Edwards, J. (1695) *Some Thoughts concerning the Several Causes and Occasions of Atheism, Especially in the Present Age*, London.(Contains an attack on Locke’s *Reasonableness* which, together with further attacks in *Socinianism unmask’d* (1696), *The Socinian Creed* (1697), and *A brief vindication of the fundamental articles of the Christian faith…from Mr. Lock’s reflections upon them* (1697), stimulated Locke’s own *Vindication* (1695b), and *Second Vindication* (1697c).)

Gibson, J. (1917) *Locke’s Theory of Knowledge and its Historical Relations*, Cambridge: Cambridge University Press.(An intelligent, still useful commentary, despite important mistakes.)

Hall, R. and Woolhouse, R.S. (1983) *80 years of Locke Scholarship: a bibliographical guide*, Edinburgh: Edinburgh University Press.(A useful bibliography covering 1900-80.)

Harris, I. (1994) *The Mind of John Locke*, Cambridge: Cambridge University Press.(Making wide use of unpublished manuscripts, this is a complex but rewarding attempt to integrate a wide range of Locke's concerns, and to show how they informed the writing of *Two Treatises*.)

Horton, J. and Mendus, S. (eds) (1991) *John Locke: A Letter Concerning Toleration in Focus*, London: Routledge.(A useful collection of papers on this work.)

King, P., Lord (1829) *The life and letters of John Locke*, London; new edn, 1964.(King had access to the manuscripts Locke left to his cousin, an earlier Peter King - Lord Chancellor and first Baron King - most of which are now in the Lovelace collection in the Bodleian Library, Oxford. This 'biography' is an ill-organized selection from these journals, papers and letters, but for many years was a valuable source of information otherwise inaccessible, and records some materials since lost.)

Leibniz, G.W. (*c*.1704) *New Essays on Human Understanding*, 1765; trans. and ed. P. Remnant and J. Bennett, Cambridge: Cambridge University Press, 1981.(Translation of Leibniz's *Nouveaux essais*, a great, if somewhat unsympathetic point-by-point commentary on Locke's *Essay* from a rationalist point of view.)

Locke Newsletter (1970-), ed. R. Hall, Department of Philosophy, University of York.(An unpretentious annual newsletter, with articles, notes, reviews and, very usefully, ongoing bibliography. Both articles and reviews have been of uneven quality, but it gives an overview of Locke studies.)

Long, P. (1959) *A Summary Catalogue of the Lovelace collection of the papers of John Locke in the Bodleian Library*, Oxford: Bodleian Library.(An essential tool for Locke scholars.)

Lowe, E.J. (1995) *Locke on Human Understanding*, London: Routledge.(A clear philosophical introduction to the *Essay* for students, making sensible use of recent scholarship in interpreting Locke's arguments.)

Mackie, J.L. (1976) *Problems from Locke*, Oxford: Oxford University Press.(A selective philosophical commentary which, although lacking the historical dimension which would often clarify the arguments discussed, has been influential in rekindling interest in Locke's general philosophy.)

Marshall, J. (1994) *John Locke: Resistance, Religion and Responsibility*, Cambridge: Cambridge University Press.(A contextual account of the development of Locke's political, religious, social and moral ideas, making wide use of unpublished writings. Perhaps the best study to date of Locke's religious views and their place in his thought.)

Milton, J.R. (1994) 'Locke's Life and Times', in V. Chappell (ed.) *The Cambridge Companion to Locke*, Cambridge: Cambridge University Press, 1994.(A useful summary intellectual biography, with an account of other available biographical sources.)

Parry, G. (1978) *John Locke*, London: Allen & Unwin.(A judicious introduction to Locke's political philosophy.)

Proast, J. (1690) *The Argument of the Letter concerning Toleration, Briefly Consider'd and Answer'd*, Oxford; repr. New York: Garland, 1984.(The work which, with its sequel, *A Third Letter concerning Toleration*, stimulated, respectively, Locke's *Second Letter concerning Toleration* (1690), and *Third Letter for Toleration* (1692).)

Rogers, G.A.J. (ed.) (1994) *Locke's Philosophy: Content and Context*, Oxford: Oxford University Press.(A lively collection of papers by leading Locke scholars on subjects including Locke's life in Oxford, substance, perception, freedom of will, meaning, atomism, aboriginal rights, sin and Locke's influence abroad.)

Stillingfleet, E. (1680) *The Mischief of Separation*, London.(Stillingfleet's argument against toleration of nonconformity in this and his subsequent *The Unreasonableness of Separation* (1681) stimulated Locke and Tyrrell to compose a point-by-point rebuttal (1681-3).)

Stillingfleet, E. (1697) *A Discourse in Vindication of the Doctrine of the Trinity: with an Answer to the Late Socinian Objections against it from Scripture, Antiquity and Reason*, London.(Stillingfleet accused Locke of sympathy with Socinianism, in this work and in two sequential *Answers*, stimulating Locke's *Letter to the…Bishop of Worcester* (1697), and two *Replies*, (1697/1699).)

Tully, J. (1980) *A Discourse of Property: John Locke and his Adversaries*, Cambridge: Cambridge University Press.(An illuminating study of ideas about ownership, in particular about the relation between making and owning, and their role in Locke's thought.)

Tully, J. (1993) *An Approach to Political philosophy: Locke in Contexts*, Cambridge: Cambridge University Press.(A selection of papers on Locke and his context by a leading commentator on Locke's political philosophy.)

Wolterstorff, N. (1996) *John Locke and the ethics of belief*, Cambridge: Cambridge University Press.(A vigorous argument as to the philosophical and cultural significance of Locke's principle that reason should be our guide in everything, and in religion in particular.)

Woolhouse, R.S. (1983) *Locke*, Brighton: Harvester Press.(A useful short introduction to Locke's general

philosophy.)

Yolton J.S. and Yolton J.W. (1985) *John Locke: a reference guide*, Boston, MA: G.K. Hall.(A useful bibliography.)

Yolton, J.W. (1956) *John Locke and the Way of Ideas*, Oxford: Oxford University Press.(This short but informative book, locating Locke's thought in its English context, was a landmark for historical study of his general philosophy.)

Yolton, J.W. (1970) *Locke and the Compass of the Human Understanding*, London: Cambridge University Press. (An important contribution to the interpretation of Locke's philosophy of science and moral theory.)

Logic in China

Technically, classical China had semantic theory but no logic. Western historians, confusing logic and theory of language, used the term 'logicians' to describe those philosophers whom the Chinese called the 'name school'. The best known of these were Hui Shi (380-305 BC) and Gongsun Lung (b. 380 BC?). This group now also includes the Later Mohists and the term 'distinction school' (translated as 'dialecticians') has become common.

The importance of the more detailed Mohist work came to light in modern times. The Confucian tradition had lost access to it. Rescuing that text rekindled a long-lost interest in Chinese theories of language. The restored Mohist texts give us a general theory of how words work. A term picks out part of reality. Some terms are more general than others; terms like 'dobbin' or 'horse' or 'object' might pick out the same thing. When we use a term to pick something out, we commit ourselves to using the name to pick out similar things and 'stopping' with the dissimilar. Thus, for each term we learn an 'is this' and an 'is not'. 'Is not' generates an opposite for each name and marks the point of distinction or discrimination.

Chinese doctrine portrays disagreements as arising from different ways of making the distinctions that give rise to opposites. The word bian *(distinction/dispute) thus came to stand for a philosophical dispute. The Mohists argued that, in a 'distinction/dispute', one party will always be right. For any descriptive term, the thing in question will either be an 'is this' or an 'is not'.*

Mohists were realistic about descriptions and the world. Real similarities and differences underlie our language. They rejected the claim that words distort reality; to regard all language as 'perverse', they noted, was 'perverse'. The Mohists failed, however, to give a good account of what similarities and differences should count in making a distinction. Mohists also found that combining terms was semantically fickle. In the simplest case, the compound picked out the sum of what the individual terms did. Classical Chinese lacked pluralization so 'cat-dog' works like 'cats and dogs'. Other compound terms (such as 'white horse') worked as they do in English. The confusion led Gongsun Long to argue, on Confucian grounds, that we could say 'white horse is not *horse'.*

Confucius' linguistics centred on his proposal to 'rectify names'. Confucius used a code with fixed formulations, and therefore tended to treat moral problems as turning on which terms we use in stating them. The abortion dispute illustrates this well. Both sides agree to the rule 'do not kill an innocent *person': the dispute becomes one of whether to use the term 'person' or 'foetus'. In contrast, Mohists argued that we should not alter normal term use to get moral results. We simply accept that guiding compounds may not follow normal use. A thief is a person, but killing a thief (executing) is not killing a person (murdering).*

These results bolstered Daoist scepticism about words. We never will fashion a 'constant' dao. According to Zhuangzi, even a realistic theory of language (like that of the Mohists) will not give constant guidance. He drew from Hui Shi's approach to language, which emphasized relative terms such as 'large' and 'small'. We may talk of a large horse (relative to other horses) or a large horsefly (relative to other flies), but 'large' itself has no constant standard of comparison. From the premise, 'all such distinctions are relative', Hui Shi fallaciously concluded that 'reality has no distinctions in itself'. Zhuangzi rejected this conclusion and ridiculed Hui Shi's monism. If we say 'everything is one', then our language attempts to 'point to' everything. If it succeeds, then in addition to the 'one-everything' there is the reference to it. That makes two. The whole consisting of everything and saying so then makes three. Referring to that whole makes four, and the fact that we have referred to it makes five, and so on.

Zhuangzi shifted Hui Shi's focus slightly, and concentrated on 'this' and 'that'. These do refer to things, but each use is different. Language, he argued, is not fixed on the world but on our relationships with it. Each existing language (different ways of making guiding distinctions) is equally natural. Human debate is as natural as the chirping of birds. We cannot appeal to nature to settle our disputes about ethics. The standards are not constant; they are historical, variable and diverse in different moral communities.

Distinctions are real, but we can never know if we have found the right ones. Zhuangzi accepts a real world in which language works. Thus, he celebrates the endless possible ways of distinguishing 'this' from 'not-this'. Some alternatives will certainly work better (assuming our present values) than the one we have now. The problem is

that any standard we could use to decide about that would itself be controversial.

The final word came from Xunzi and his student Han Feizi. The former, a Confucian, understood Zhuangzi's arguments to show that the only standard of correct usage must be convention itself. Thus he renewed Confucian tradition and promoted it politically as the only viable and valid conventional system. He advocated government suppression of dissenting voices who 'confuse language' and 'create new terms'. In the end, only the ruler may change language (and then only the 'descriptive' terms). The standards of social assent and dissent come from the Confucian 'sage-kings'. We must adhere to these as the only acceptable ideals; the alternative is anarchy in moral discourse and, consequently, in society.

Han Feizi, seized on Xunzi's attitude toward coercion while discarding the appeal to ancient tradition. Han Feizi had considerable influence on the draconian Qin emperor who ruthlessly carried out his injunction to stamp out philosophical disputes about ethics. This brought the rich tradition of creative philosophy to an abrupt end; religious thought and scholasticism dominated the rest of Chinese intellectual history.

1 Confucius: rectifying names

To understand the development of theory of language, we must place it in the context of Chinese ethical thought. The central focus of ethical dispute was about *dao*, or 'guiding discourse' (see Confucian philosophy, Chinese §2). Confucius championed the historical 'guiding discourse' of the sage-kings, and purportedly studied it in the ancient documents that formed the curriculum of his school. The *Liji (Book of Rites)* is the paradigm for Confucius' conception of ethics.

Confucius' semantic framework was the relation between language and action, not between language and objects or reality. The implicit role of language was prescriptive rather than descriptive. Confucius, oblivious to ethical criticism of his discourse-based guide, addressed mainly problems in practical interpretation. His theory implied a pragmatic relation between language and objects. In order for a discourse to guide us, we must correctly pin names on the world's 'stuff'.

Confucius argued for imitation as the way to achieve this. Social leaders model the proper use of names in publicly practising the 'ritual'; the people learn from these examples and are then able to follow the code as it applies to them. Confucius called this 'rectifying names', and treated it as the key to good government:

> Zilu said, 'The ruler of Wei awaits your taking on administration. What would be Master's priority?' The Master replied, 'Certainly - rectifying names!' … If names are not rectified then language will not flow. If language does not flow, then affairs cannot be completed. If affairs are not completed, ritual and music will not flourish. If ritual and music do not flourish, punishments and penalties will miss their mark. When punishments and penalties miss their mark, people lack the wherewithal to control hand and foot. Hence a gentleman's words must be acceptable to vocalize and his language must be acceptable as action. A gentleman's language lacks anything that misleads - period.
>
> (*Analects* 13:3)

The strategy of setting examples threatens a regress unless someone in the chain of models knows what example to set in some other way. Stopping the regress pushes Confucianism toward intuitionism. Confucius seemed to regard a mysterious quality, *ren* (humanity), as the key to correct practical interpretation of the *li* (ritual). 'Humanity' is a moral insight that guides the attribution of terms in following circumstances.

One does not rectify by consulting definitions: no Chinese accounts of language generate or point to anything like the concept 'meaning'. The model of the language-world relation is political. Social authorities tag things for ethical purposes. This tagging facilitates guiding discourse. Two theories of Confucian tagging emerge: one determined by traditional training and one by innate moral intuition.

Classical Chinese grammar reflects this model of language-world relations in its topic-comment structure. In the place of sentences, literary Chinese makes 'comments' on contextually indexed 'stuff'. Classical sentences required no subject. Language appeared as far more context dependent than in modern Western thought, with its focus on the subject-predicate sentence and 'complete thought'.

Implicit cognitive theory also mirrored this topic-comment structure. Western philosophy focused on propositional

belief, a mental state that represents a fact, and knowledge. However, ancient Chinese grammar had neither propositional belief nor knowledge structures. Chinese grammar suggests a person tending to deem some real object (*X*) as *P* (the comment). Knowing is correctly deeming or assigning *P* to *X* (or knowing *X*'s being deemed *P*). Literary Chinese used adjectives, one-place verbs and even nouns as two-place predicates in situations where we would use propositional belief structures: one *P*'s some *X*. An alternative structure uses the Daoist concept of *wei* ('deem-act', the *wei* of *wuwei*, 'lack deeming action'). One, with regard to *X*, 'deems-acts' it *P*.

For these reasons (along with the absence of functional inflection, an ideographic conception of writing and the close grammatical similarity of proper and common nouns, adjectives and even verbs), Chinese linguistic theory focused on the question of what term to assign to things rather than on the propositional units so central to Western theory of language and logic (see Language, philosophy of). The dominant conception was that a word had a scope or range of application, rather than that of referring to individuals or objects. This tendency reflects the fact that Chinese nouns resemble mass-nouns in having cumulative reference, in lacking both grammatical number and articles, and in being associated with various ways of individuating. The use of sortals for individuation became regularized for Chinese nouns at the end of the classical period.

2 Mozi: language utilitarianism

The natural development of this model in Mozi's early work (and the subsequent elaboration in later Mohism) focuses on *bian* (distinction/dispute) (see Mohist philosophy; Mozi). A term's use involves an 'is this/right' (*shi*) and an 'is not this/wrong' (*fei*). To learn the term is to learn to 'is this/is not this' appropriately with it. Mozi argued that society should use the pre-conventional or natural 'will' toward benefit (and against harm). This contrast guides the assignment of 'is this/is not this' to words used in social discourse. This interpretive proposal flowed imperceptibly into a proposal to order guiding dialogues differently, to change Confucius' traditional 'guiding discourse'.

The clearest example is Mozi's argument about spirits and fate. General utility dictates that social discourse should include the string 'lack spirits' and 'have spirits'. He represents this conclusion as an example of knowing the *dao* of *you-wu* (have-lack) (see You-wu). That means making a 'is this/is not this' (*li-hai*) distinction for each of these terms using the standard of 'benefit-harm'. We use either 'have' or 'lack' of things when so doing will lead to general utility.

The implication was initially anti-realistic. Mozi advocated three standards of language use. The first recognized the historical, conventional aspect of language. Language should conform to the guidance intentions of the ancient sage-kings. Second, language standards should be applicable by ordinary people using their 'eyes and ears'. Mozi's favourite examples of *fa* (standards) are measurements: a plumb line, a compass, a square and stakes for plotting where the sun rises and sets (see Fa). Finally, we should use words in ways that maximize general utility. For Mozi, these standards pull in essentially the same direction. He assumed the people's good motivated the sage kings and he repeatedly likened the 'will of nature' to an objective utility measurement.

These standards govern the content and practice of discourse, regulations, injunctions, maxims and slogans. Including any string in the approved discourse was 'making it constant'. The ideal was a discourse *dao* that could consistently (reliably) and correctly (objectively) guide society. Mozi identified that *dao* as the one that resulted in the greatest utility for the country and its people. Thus, description or assignment of names is a handmaiden to ethics.

To count as the constant *dao*, Mozi's 'benefit harm' standard must itself be 'constant'; it should be a reliable, unambiguous and objectively correct, unchanging standard. He argued that in fact it was, since it came from *tian* (nature) rather than from society, convention or contingent history (see Tian).

Mozi's attack on conventional guiding discourse led Mencius to defend Confucianism by postulating an innate moral intuition. Mencius argued that language should not manipulate or guide human action. Guidance should only come from the innate patterns or dispositions in the heart-mind; these include an innate ability to 'is this/is not this' (see Xin (heart-and-mind)). The heart-mind selects the appropriate 'is this/is not this' assignment and thus the appropriate action in real contexts. Society should not distort or reshape those natural moral inclinations. Mencius believed that, left to itself, everyone's heart-mind would innately select the 'correct' action for them.

Laozi pointed in a similar direction but undermined Mencius' optimism. Indeed, we should resist the conventional socialization that comes with language. Learning names both constrains natural spontaneity and creates new and disruptive competitive desires. However, Laozi portrayed natural behaviour as being much more 'primitive' than did Mencius. The realm of social concern would extend no further than the local agrarian village. Laozi also emphasized the anti-language aspect of intuitive guidance more than did Mencius. He hinted that Mencius' idea that some particular moral values were intuitive or innate was a result of confusing the unconscious result of learning a guiding language with innate intuition. Learning names involves training in how to make distinctions and how to 'desire' with them. The names, distinctions, desires and actions are linked distortions of natural spontaneity. Laozi's conclusion is his opening line: no guidance in discourse form is constant (see Daoist philosophy §2).

3 Rediscovery of later Mohist theory

Before discussing the later Mohists, let us glance at the textual explanation of the loss and recovery of the Mohist Canons. The current 'best' textual theory says the Mohists wrote two 'Canons' (I and II). Each consisted of a series of short maxims; the first half of each Canon was written vertically across the top of a standard-sized book of bamboo strips, and the second half on the bottom. The terse analytic theorems were then keyed to another bamboo book containing longer explanations, examples or arguments for them. The second set of bamboo books was indexed by putting the first character of the claim alongside the first character of the explanation.

When scribes later copied each strip of the canons straight through, they turned this clever system into a puzzle. With no understanding of linguistic theory, they treated the whole corpus as a set of consecutive essays. Since classical Chinese had no punctuation or grammatical inflection, this textual disaster obscured the slogans, jumbled the order and shrouded the indexing principle. The scribes absorbed the indexing character as part of the text of the now orphaned explications.

Given the philosophical sophistication and difficulty of the text, the Mohist school's obliteration at the beginning of China's philosophical Dark Age (roughly 200 BC-1000 AD) and the placement of the Canons in the middle of the most vociferous anti-Confucian classical text, the medieval Confucian orthodoxy had little motivation to tackle the puzzle until the late Qing dynasty. Sun Yirang (1848-1908) found the essential clue to unzipping and analysing the content in a phrase in the middle which read, 'read these horizontally'. Other Chinese scholars tried different reconstructions and this work still goes on. Angus Graham's *Later Mohist Logic, Ethics and Science* first delivered a version of the reconstructed text to Western sinologists in 1978 (Graham 1978).

As his title suggests, Graham shared the common view that the subject matter was logic. He assumed Mohists intended the text as a deductively connected set of definitions and propositions. The statements, however, do not resemble definitions. Genus-species form is infrequent and the slogans fit into a theory of use, not of meaning. They are far from forming a deductive set. Still, Graham's assumption guided his reconstruction in ways that made the content accessible to analysis. Many problems and obscurities remain, but Graham's reconstruction reveals a systematic and reasonably coherent theory of language.

The maxims do deal with central philosophical concepts and, like Chinese dictionaries, frequently give lists of substitution characters or a range of examples. Some slogans are metaphors which the explications exploit; others are helpful ways of re-thinking and reflecting on a familiar concept. In addition to theory of language, intelligible sections of the Canon present fragments of epistemological, geometrical, optical and economic theory.

4 Later Mohists: names and distinctions

The School of Names was technically not a school (see Mohist philosophy §5). Traditional scholars gave the name to individual thinkers who analysed names in conflicting ways; their motivations reflect the differing trends in pre-analytic political thinking. The later Mohists developed Mozi's pragmatic, reality-based approach to naming. Gongsun Long defended the Confucian ideal of an unambiguous guiding scheme of names (the one produced by rectifying names). This principle, he presumed, dictated a goal of one-name-one-thing. Hui Shi raised sceptical problems designed to show that language 'constancy' across all situations was impossible (a Daoist conclusion).

The later Mohists spoke of four 'objects' of *zhi* (knowing): names, 'stuff', union and deeming-action (see Zhi). Mozi used an example that helps us fix the relevant concept of 'knowing'. Blind people can 'know-to' produce

utterances like ‘black like coal’ or ‘white as snow’ but cannot distinguish things when placed in front of them. They know names, but not ‘stuff’ or union. Thus they cannot use the names to guide their actions.

The Mohists, as we noted above, lacked a doctrine of ‘belief’. ‘Knowing union’ meant competently assigning terms and descriptions to objects. The Mohists accepted that ‘knowing names’ was conventional knowledge. They stressed, however, that we apply conventions to an external reality, known independently of language. The goal of knowledge remained practical guidance, not representation or picturing. Once conventionally attached to some reality, the inherent similarities and differences determined a term’s application. Conventions presuppose a world-guided way to mark distinctions. Mohists portray name-object relations mereologically: a name applies to a scattered reality determined by *tong-yi* (same-different).

The Mohist terms for ‘reference’ - *ju* (pick out) and *qu* (choose) - had a practical tone. A name picks out some ‘stuff’ from the background. Convention determines which similarities and differences mark the boundary between *shi* (‘is this’, what a name picks out) and *fei* (‘is not this’, what it excludes). In using a name, we commit ourselves to go to some real limit and then stop.

The Mohists argued against the one-name-one-reality principle of rectifying names. Names, the Mohists argued, could be very general (like ‘thing’ itself), or based on similarity classes (like ‘horse’) or applied to only one thing (like ‘John’). They saw no objection in principle to having overlapping scopes and even two names for the same thing (like ‘puppy’ and ‘dog’).

Elaboration and defence of this account of *bian* (distinction) led to complications. How should we expand the account to explain what a string of two *ming* (‘names’, characters) picks out? The model used was the string *niu-ma* (ox-horse). The Mohist took this to pick out a compound stuff, the sum of the range of the two component terms (‘draft animals’). They called the unit a *jian* (whole) and its parts *ti* (substantive parts). This analysis of compounds made them analogous to general terms. However, this treatment raised several questions. In what sense is a compound really two things? Could we not view anything as a compound of more basic stuffs? Is there any fundamental kind of ‘substantive part’?

These questions lead to another: are there any basic *tong-yi* (same-different)? The Mohists gave no clear answer. They noticed many senses in which things can be ‘same - different’; some realities might be different only in being called by different names. Being ‘two’ was necessarily differentiated even though called by one name. Realities could also be the same in the sense of being included in some compound object. Conversely, they could be different in not being included in some ‘substantial part’: they could be the same or different in being in the same place or not, and finally they could be same or different in belonging to the same *lei* (kind) (see Mohist philosophy §5).

However, the Mohists analysed *lei* in a loose way. Having that with which *tong* (same) was the criteria of being *tonglei* (same kind). Not having ‘same’ was the criteria of not being of a ‘kind’. Although they might initially have intended to limit ‘kind’ to natural kinds, the account generalized it to almost any similarity based grouping of stuffs. Thus the Mohists could refer to oxen and horses as the same ‘kind’. The only clear examples of not-‘kind’ are things so unlike they are not comparable. ‘Which is longer, wood or night?’ they suggest, is an unintelligible question because it compares two different ‘kinds’.

The looseness of this account of classifying buttressed the sceptic’s position (see §§8-9) that the world offered no reliable basis for fundamental distinctions. The Mohists seemed vaguely aware of the difficulty and did condemn *kuangju* (‘wild picking out’, where we use irrelevant ‘same’ or ‘different’ to classify or distinguish). They gave no account, however, of what marks a ‘wild’ or irrelevant way of grouping ‘stuff’.

5 Semantic paradoxes and compound terms

The central term of assessment in the Mohist study was *ke* (assertible). They used it in several related ways. An expression may be described as ‘assertible’ of some object. If this object is introduced by another term, then ‘assertible’ became a way of exploring semantic relations between terms (see Semantics). Mohists asked whether we can sometimes, always or never describe things picked out by term *X* as *Y*.

The analysis, although in terms of assertibility rather than truth, yielded a familiar and important conclusion against certain forms of relativism (see Relativism). The Mohists argued that in any dispute involving

'distinctions', there will be a 'winner'. If one disputant claimed it was ox and the other that it was not ox, one would be correct. When one disputant claimed the object was ox and the other that it was horse, they did not count that as a distinction dispute. This was merely a formal result, but the Mohists took it as confirming that the world, not conventions, determined the right designation. The winner is the one whose description *dang* (hit on) it.

Another result was a distant cousin of 'All sentences are false'. Unlike the classic liar paradox (see Semantic paradoxes and theories of truth), the universal sentence does have a consistent truth value: it is always false. The Mohist conceptual tools, however, lacked two concepts, namely 'sentence' and 'truth'. Instead they construct the following replacement: with regard to language, deeming it exhaustively inadmissible is inadmissible. They explain this result as arising from one's own language and suggest the proposer try harder to find acceptable words.

This result undermines the anti-language positions, notably those of Mencius and Laozi. It invalidates any claim that language distorts 'guidance'. Two similar results undermine different formulations of similar points. One version decries distinctions. To make a distinction is to regard something as 'is not this', and to say 'distinctions are wrong' is to make a distinction. The Mohists say, 'to *fei* "*fei*" is perverse' (that is, to say 'saying "wrong" is wrong' is wrong). To recommend making no distinctions is to make a distinction (between making and not-making) and thus the recommender violates his own recommendation. Similarly, the Mohists observe, to teach that teaching lacks use, lacks use.

The analysis of compound terms produced some of the most striking evidence of the different conceptual structure. The Mohists ask what was assertible of the things picked out by compound terms. One standard case was 'ox-horse'. The Mohists note that 'not-ox' is assertable of 'ox-horse' on the same grounds that 'ox' is. The explanation goes that part of ox-horse is non-ox, so 'non-ox' is 'assertible'. We can understand the idea by reflecting on another example. We may ask someone how many children they have by asking how many 'boys girls' they have. Suppose the answer is 'three'. Now we may ask how many are boys. The answer may come back, 'none'. This would be a case in which we could say 'their boy-girls are not boys'.

However, the Mohists seems to have something stronger in mind. Even if the answer was 'two girls and a boy', the Mohists would argue that it would be right to say that (some of) their boy-girls were non-boy. Thus the Mohist concludes that, although we cannot say ox is non-ox or horse is non-horse, we can say intelligibly that ox-horse is non-ox-non-horse. The Mohists explain this result as arising from the fact that ox-horse 'does not interpenetrate'. In any compound that is ox-horse, the minimal parts will be either ox or horse, but not both. They term it a 'separable' compound.

The contrast is hard-white. This is also a compound term but, in this case, the components are inseparable. Wherever you go in that which is appropriately called hard-white, you get both. This is the kind of compound term more familiar in Western languages, the intersection compound (see Language, philosophy of). The scope of the combined term is an intersection of the scopes of the two component terms. Ox-horse, by contrast, is a sum compound. The scope of the combined term is the union of the scopes of the two component terms.

The Mohists do not give us any rule for distinguishing intersection compounds from union compounds beyond the metaphysical feature of 'penetrating' or 'excluding'. They do not explicitly use the language of scope. They also characterize compounds as 'inseparable' and 'separable'. This curious result arises partly because they treat both nouns and adjectives as 'names'. Both pick out or distinguish one part of reality from the rest. This leaves us with the impression of arbitrariness in how names form compounds, which Mohists explain via metaphysical 'penetrability' or its opposite.

6 Phrase matching

Although the Mohists proposed a realist theory of reference, they embedded this in a further theory that we are guided by language. Thus the fourth object is knowing how to 'deem act'. The Mohists viewed the object of combining names as guiding action. Thus names pick out 'stuff', while strings or *ci* (phrases) convey intentions. This led them to analyse mainly ethical compounds. The other study, of 'ox-horse' and 'hard-white' compounds (see §5), construed them merely as complex names (though it gave no consistent principle for determining their conventional scope). In the long intact section of this text, the Mohists pursued a different analysis, one that can take a superficially logical form:

Premise: *X* is *Y* (white horse is horse)
Conclusion: *KX* is *KY* (ride white horse is ride horse)

They called this ‘matching phrases’, and argued that it was not reliable. Success would be as follows: whenever the version with simple terms was positive (an ‘is this’ phrase), then the parallel with compound terms should be positive (a *ran* (‘so’) phrase). Conversely, a negative base (*X* is not *Y*: a ‘not this’ phrase) should yield a negative result (*KX* is not *KY*: a ‘not so’ phrase).

The Mohists list the different kinds breakdowns:

- Sometimes an ‘is this’ yields a not-‘so’;
- Sometimes a ‘not this’ yields a ‘so’;
- Sometimes a reference is comprehensive and sometimes not;
- Sometimes one reference is ‘is this’ and another one is ‘not this’

The rest of the essay consists of examples that illustrate the respective outcomes.

Graham, drawing on a part-of-speech analysis of ‘is this’ (subject) and ‘so’ (verb), treated this chapter as evidence that the Mohists discovered the subject-predicate sentence. He translated the procedure as ‘matching sentences’ and treated it as a discussion of logical form. The first two models do rely on syntactic complexes (*X Y ke*) which resemble syllogistic premises, but the latter two do not.

Classical Chinese uses no articles and has no ‘is’ verb. Expressions ending with the particle *ye* (assert) mark descriptive uses of referring expressions (noun-phrases). This signals that one is applying a descriptive term to a contextually selected object, not using the term to pick out the object. Translators typically render such structures in English as ‘(*X*) is *Y*’. In Chinese comments, the topic term (the *X*) is optional. The comment functions the same when the string is simply ‘*Y* assert’ and context supplies the topic. The ‘assert’, in other words, does not link two terms, it links a term to an object.

The *shi-ran* (is this-so) analysis given by the Mohists does fit the examples in a way consistent with the topic-comment analysis. If, on the other hand, we focus on the examples as sentences and treat the pattern as a form of inference, then the Mohist analysis will resemble a kind of algebraic logic. However, consistent with a topic-comment analysis, it may be better to regard it as extending the analysis of the conventional semantic effects of combining ‘names’ to form ‘phrases’.

That the analysis rests on conventional semantics of terms rather than logic of sentences sheds a new light on how the examples work. The Mohists do not use the model to correct conventional reasoning errors; rather, they use what we would conventionally say to determine whether a result is a ‘so’ or a not-‘so’. The most thoroughly illustrated breakdowns are those where an ‘is this’ base produces a not-‘so’. The examples are:

- Someone’s parents are people; one’s serving one’s parents is not one’s serving the people;
- One’s younger brother is a handsome man; one’s loving one’s younger brother is not one’s loving a handsome man;
- A carriage is wood; riding a carriage is not riding wood;
- A boat is wood; entering a boat is not entering wood;
- Robbers are people; abounding in robbers is not abounding in people; lacking robbers is not lacking people.

The Mohists expand on the last example in a way that signals both the ethical importance of the analysis and the nature of the alleged breakdown in parallelism:

- Disliking the abundance of robbers is not disliking the abundance of people.
- Desiring to be without robbers is not desiring to be without people.
- Everyone would agree with these so they should not object if we say ‘robbers are people but killing robbers is not killing people’.

We guess, following Graham, that the Mohists are defending their inherited doctrine of universal love by arguing that it is consistent with the (presumed) practice in Mohist communities of executing thieves.

What the denial amounts to is that, even if naming is objectively constant and reliable, the use of names in descriptions of actions or intentions does not reliably take one from an ‘is this’ to a ‘so’. An execution is not murder. Loving a brother is morally required, loving a handsome man is (presumably) shameful. Serving one’s parents is one kind of duty and serving the people another. One does not fulfil the latter in merely doing the former.

What emerges is an alternative strategy for dealing with the problem that Confucius addressed via rectifying names (see §1). The Mohists resist the implication that in executing thieves they must deny that thieves are people. They deny instead that executing thieves is murdering people. Rectifying takes place at the ‘phrase’ or ‘so’ level rather than at the ‘name’ level.

The next set of examples illustrate the converse case, those where we start with a ‘not this’ base and the result is a ‘so’:

- To read books is not books; to like reading book is to like book.
- Cockfights are not cocks; to like cockfights is to like cocks.
- About to fall in a well is not falling in a well. To stop one about to fall in a well is to stop one falling in a well.

This time the Mohists are expanding on fatalism: ‘That there is fate is not fated; to deny that there is fate is to deny fate’. It is harder to reconstruct a problem that is plausibly solved by this analysis.

The algebraic form is abandoned in illustrating the next two breakdowns in parallelism. The first is ‘part comprehensive; part not’:

- ‘Loving people’ depends on comprehensively loving people. ‘Not loving people’ does not.
- ‘Rides horses’ does not depend on comprehensively riding horses. To have ridden on horses is enough to count as riding horses.

These examples highlight the mass-substantive structure of reasoning about reference. In one phrase the term-reference is implicitly comprehensive, in the other it is not.

Finally, we come to the examples of one ‘is this’ and one is ‘not this’:

- Fruit of a peach is a peach, fruit of a bramble is not a bramble.
- Asking about a person’s illness is asking about the person. Disliking the person’s illness is not disliking the person.
- A person’s ghost is not the person. Your brother’s ghost is your brother. Offering to a ghost is not offering to a person. Offering to your brother’s ghost is offering to your brother.
- If the indicated horse’s eyes are blind, then we call the horse blind. The horses eyes are large yet we do not call the horse large.
- If the indicated oxen’s hairs are brown, then we call the oxen brown. The indicated oxen’s hairs are many yet we do not call the oxen many.
- One horse is horse; two horses are horse. Saying ‘horses are four footed things’ is a case of one horse and four feet, not a pair of horses and four feet. Saying ‘horses are partly white’ is two horses and some white, not one horse and partly white.

There is no further analysis or summing up. The moral, we assume, is a negative one. Had we treated the first algebraic model as a kind of logic, the Mohists’ argument by example would still show that it was invalid; it is not a reliable form in the sense that a true premise formally guarantees a true conclusion. However, the essay is not about sentences and truth at all. It is about whether there are reliable parallels developing from terms to longer (guiding) phrases.

The implicit answer is ‘no’. The Mohists offer no constant or consistent principles guiding the construction of longer ‘phrases’ out of terms even when the terms are consistently applied to external realities. They offer no way to systematically rationalize the conventional patterns of use. They retreat implicitly from Mozi’s goal of replacing convention with a constant *dao* (guide). If there is such a ‘guide’, then it is not a product of any simple projection from term reference. Moral guidance cannot derive from knowledge of natural kinds; it requires conventional,

creative human social activity. Dialectically, the negative result gives ammunition to the Daoists who argue that no constant ‘guide’ exists (see Daoist philosophy §2).

7 Gongsun Long: one name, one thing

Against the Mohist background, we can now make some sense of the previously obscure writings of Gongsun Long. The text that bears his name consists of a few dialogues. Each is introduced by a counterintuitive paradox followed by a discussion or chain of reasoning. Graham argued that at least two of these were forged out of misunderstood fragments of the Mohist writings. They apparently copied the phrases after the Canons had been zipped together and the indexing characters mixed into the text.

Two of the remaining dialogues: ‘White Horse’ and ‘Referring and Things’, serve as useful examples. They pose difficult puzzles, to which scholars have offered speculative, controversial and mutually inconsistent interpretations. The various interpretations flow partly from different principles used in selecting interpretive theories or ‘translation manuals’. A couple of interpretations are given here, serving both to illustrate that point and to allow us to locate and discuss alternative views of Gongsun Long’s theory.

In a (perhaps spurious) preface to the dialogues, Gongsun Long explains his motivation. He cites an example of Confucius rectifying names (see §1) and alleges that he is defending that view. Confucians mostly reject the affiliation, but it makes formal sense. If rectifying is to remove ambiguity from guidance, then it requires that only one name in a guidance situation can refer to the object. I either regard the male before me as ‘father’ or as ‘ruler’ or as ‘person’. Supposing he is all three, if I am to extract guidance from codified rules, I must decide which rule to use. That means deciding which term is relevant to this action situation. The Mohist attack on that policy rejected any one-name-one-thing principle.

The Mohist account of compounding had negative implications from this point of view. Separable or sum compounds, such as ‘ox-horse’, conform to the one-name-one-thing principle. The scope of each term remains constant when we combine the terms. The combination names a sum of the two. Hard-white compounding, by contrast, violates the principle of strict clarity and consistency in naming. The scopes change when we compound the terms.

Other sources record Gongsun Long as defending two theses: ‘separating the inseparable’ and ‘separating hard-white’. Since Graham argued that the dialogue on ‘Hard-White’ was forged, we cannot rely on it for an explanation. Still, we can confidently interpret the slogan since ‘hard-white’ is the Mohist example of an ‘inseparable’ or ‘interpenetrating’ compound. To ‘separate’ them would be to treat them as ‘excluding each other’ and the compound as a sum compound. Gongsun Long thus objects to the hard-white model.

The phrase ‘white horse’ selects one term from each type of compound. The White Horse dialogue begins with a question in the canonical analytic form: ‘Is "white horse not horse" assertible?’ followed by the answer, ‘assertible’. The rest is generally agreed to be a discussion between the sophist defending the answer and an ‘other’ who raises objections. The first defence is that ‘white’ names a colour and ‘horse’ names a shape. Shape and colour are different, so a combination of shape and colour is not merely a shape.

The other most cited argument is ‘if you ask for a horse, both a black or yellow horse can arrive. If you ask for a white horse, a black horse or yellow horse will not arrive.’ This illustrates an argument thread that ‘*X* is not *Y*’ follows from ‘*X* is different or distinguishable from *Y*’. The linking theme is that ‘white horse’ is a combination of two things and this requires that ‘white horse not horse’ be ‘assertible’. A Mohist might respond that ‘asking for a white horse’ is indeed different from ‘asking for a horse’, but a white horse is still a horse.

One line of interpretation treats the term ‘horse’ as referring to the abstract object ‘horseness’, and thus ‘white horse’ to ‘white-horseness’. The opening sentence thus states the *true* proposition that the two abstract concepts are distinct entities. Since the connected terms are logically singular, the ‘not this’ represents ‘non-identity’. This line of interpretation is motivated by the principle of charity and undermined by the principle of humanity. It makes the puzzling sentence true (by Western lights) but does not explain how the sophist would have had access to the concepts involved. It does not apply to the two terms when used in the supporting arguments, all of which refer to concrete horses.

An alternative interpretation, that ‘horse’ refers to horse-stuff, can exploit the ‘distinct-hence-different’ line of

argument and still consistently interpret the concrete references in the rest of the dialogue. If we similarly regard white as the mass-substantive - white stuff - rather than the abstract 'whiteness', we can see a connection to the Mohist theory of compound names. Gongsun Long regards 'intersection' or 'interpenetrating' compounds as contrary to the one-name-one-thing principle. If white horse consists of two names, each should consistently name (scattered) things. Used in combination, their 'naming' should remain consistent. Thus, they should name the sum of the two stuffs and, as in the case of 'ox-horse', 'not-horse' would be assertable of it.

Alternatively, we may either deny that 'white horse' consists of two names or that they are the same names as when used separately. 'White horse' must be thought of as having no essential relation to 'horse' but as a *sui generis* term for a new stuff. We might then say 'white horse is horse' is not analytically true. Its truth is an accident of usage which *could have been otherwise*. Thus, 'white horse not horse' is assertible.

Gongsun Long's argument then becomes a dilemma. Either we regard 'white-horse' as a sum-compound term - in which case the 'ox-horse' result follows - or we regard it as a *sui generis* non-compound name - in which case the conventions of its use could tie it to anything at all; it need not necessarily be horse. The assumption must be that a name is the same only if it has the same scope (names the same thing). Since 'horse' in 'white horse' does not have that scope it is not the same name and its use in the compound constitutes an arbitrary new term.

The other dialogue poses, if possible, even more daunting barriers to interpretation. The first sentence seems to be an explicit contradiction: everything under heaven is *zhi* (pointing), and yet 'pointing' is not 'pointing'. The rest of the dialogue is content-thin and teeters repeatedly on the edge of pure syntactic contradiction. The only nouns are the puzzling *zhi* (pointing) along with *wu* (thing-kind) and *tianxia* (the world). Most interpreters take the issue to be the meaning of 'pointing' (a rare, mild consensus), and most treat it as semantic reference or meaning.

There are reasons for worry: as we saw, the Mohists focused on 'picking out' 'stuff' from its surroundings. There is no evidence of the semantic concept of 'meaning'; 'reference' is a more plausible evolution of that theory. Using 'meaning,' however, makes this dialogue mesh better with the abstract interpretation of the 'White Horse' thesis. Otherwise, it is hard to find any evidence of a sense-reference distinction or find any indication that it is individual objects rather than mereological wholes or types that we 'point to' (see Sense and reference; Mereology).

One speculative interpretation exploiting the 'reference' interpretation is both philosophically interesting and relevant to issues that emerge in theory of language and metaphysics. Graham treats the crucial first phrase as meaning that although one can refer to each thing, one cannot refer to 'everything' because in that reference one does not refer to one's own act of referring. Zhuangzi later makes a similar argument against assertions of absolute monism: to say everything is one is to have the one and the saying, which makes two. Tempting as it is, the interpretation has little theoretical connection to the White-Horse thesis. Graham treats both dialogues as dealing with the principle that whole is different from the part. The principle in question needs both careful formulation and plausible motivation.

8 Hui Shi: relativism and monism

We have even less direct textual evidence of the philosophical views of Hui Shi. All that remains are ten paradoxical sayings recorded in the *Zhuangzi* and some stories of his playing as Zhuangzi's 'debating companion'. Still, we can be confident of his position because plausible motivations for the sayings are more intuitive and are plausibly reflected in Zhuangzi's Daoism (see Zhuangzi).

Hui Shi focuses on such distinctions as large/small, thick/thin, high/low, south/north and today/yesterday. The common feature is the variability in which term from each pair we can assert of some object. Most of his paradoxes make sense as contrary comments about the same thing, made from different points of view, such as these examples from *Zhuangzi* 33:

- Heaven is as low as the earth; mountains are level with marshes.
- The sun from one perspective is in the middle from another declining.
- Natural kinds are from one perspective living and from one dying.
- I go to Yue today and arrive yesterday.

The most important one for theory of language purpose strikes at the Achilles heel of Mohist realism, the

construction of similarity classes:

• The ten thousand thing-kinds are ultimately alike and ultimately different. This is called the great similarity-difference.

As the *Zhuangzi* develops this insight, it suggests that we can find a difference between any two things no matter how alike, and a similarity between any two things no matter how different. So, even if there are objective similarities and differences, they do not justify any particular way of distinguishing between thing kinds. For every category and name, we could have had conventions that just as consistently and with equal 'world-guidedness' divide things differently.

However, the list of Hui Shi's sayings begins and ends with seeming absolutes. He seems to take a classic, so-called Daoist metaphysical view of an undifferentiated single totality:

• The ultimately great which has nothing outside it - call it the Great One!
• The ultimately small has nothing inside it - call it the Small One!
• Universally love the ten thousand thing-kinds; the cosmos is one 'substantive part'.

The concluding statement echoes the Mohists' ethical doctrine and one of their technical terms. We do not have Hui Shi's reasoning, but a plausible explanation is that normally taken by interpreters of Zhuangzi. It is the familiar inference of absolutism from relativism. If all distinctions are relative to some perspective, then we can conclude that reality has no distinctions. All distinctions are false: reality is an undifferentiated total one.

The *Zhuangzi* presentation of Hui Shi's views notes: 'He had many perspectives and his library would fill five carts, but his doctrine was self-contradictory and his language did not hit the target: the intent to make sense of things' (*Zhuangzi* 33). Zhuangzi presumably understood the incoherence of denying distinctions and, if we accept Graham's speculation about pointing and things, also the notion of an ultimate one - an 'everything' concept. Whether or not Gongsun Long rejected the inference, Zhuangzi clearly did. Zhuangzi almost paraphrases Hui Shi:

> 'The cosmos and I were born together, the ten-thousand things and I are one.' Now, having already constructed a 'one' is it possible to say something about it? Having already called it a 'one' can we fail to say something about it? 'One' and saying it make two. Two and one make three and going from here, even a skilled calculator can't keep up with us, let alone an ordinary man.
>
> (*Zhuangzi* 2)

Zhuangzi displays an immense fondness for Hui Shi alongside a dismissive, almost belittling attitude toward the result of his 'distinction-dispute'. Traditional accounts have reckoned this as a mystic's haughty disdain for logic. However, when we grasp that Hui Shi's doctrines have nothing to do with logic and everything to do with theory of language, a different view of the dynamic emerges. It pairs an erudite, enthusiastic and loquacious but somewhat wooly-minded semantic dilettante (Hui Shi) and a language theorist *par excellence* (Zhuangzi). This is not a case of mysticism versus logic, it is a case of clear versus befuddled theory of language. Zhuangzi enjoyed debating with Hui Shi because he was one of the few with enough learning to be worth refuting, even though he was a relatively easy target for a dialectician of Zhuangzi's calibre.

9 Zhuangzi: Sceptical perspectivalism

Zhuangzi develops perspectivalism in a more consistent direction. He does not reject language (as perhaps Laozi did). Naturalism, taking the point of view of 'nature', does not require abandoning language. Human language, from the empty greetings and small talk to the disputes of philosophers, is another natural 'noise'. We are all 'pipes of nature' (*Zhuangzi* 2).

Zhuangzi grants the Mohist point that language has 'aboutness': 'Language is not blowing breath; there is language for language. That which it languages, however, is radically underdetermined' (*Zhuangzi* 2). He develops this claim with the aid of Hui Shi's relativism and his own analysis of the indexicality of all distinctions, and starts by asking to what does 'this' or 'that' refer? Is there anything that cannot be a 'this' or a 'that'? These terms do not have a rigid, naming relation to an external reality; they trace our changing relations to the realities. Zhuangzi's perspectival pluralism is not a version of Western subjectivity. He does not assign any special

perspective to the individual consciousness or internal representations. The kinds of perspective Zhuangzi discusses range from the ways people who speak different languages or accept different moral theories constitute a perspective. At the other extreme, each of us takes different perspectives at different times in our lives - or for that matter, different times of the day.

The Mohists' *ke* (assertible) is also relative to perspective, the changing conventions of usage and principles within disputing factions and schools. Any language that actually is spoken is 'assertible'. Zhuangzi says the appearance of right and wrong in language is a function of elaboration and embellishment of a way of speaking. So the disputes between Mohists and Confucians amount to different elaborated ways of assigning 'is this' and 'not this'.

Zhuangzi calls his 'perspective' on the relativity of language *ming* (clarity). It is a perspective from which we can project backward to an 'axis' of linguistic guidance systems. From that 'axis' there would be no limit to what could be treated as 'is this' or 'not this'; a *dao* (guide) is merely the path one takes from that axis. Thing-kinds are made so by our classifications.

Lacking any limit on possible systems of naming and guiding, we lack any limit on know-how. No matter how much we advance and promote a way of dealing with things, there are things at which we will be deficient. To have any developed perspective is to leave something out. This, however, is not a reason to avoid language and a perspective; it is simply the inevitable result of limitless knowledge and limited lives.

'Sages' project their perspectives and prejudices on 'nature' and 'those who have arrived' know to deem everything as one. Zhuangzi does not recommend we use that attitude. Instead of trying to transcend and abandon the usual or conventional ways of speaking, we should treat them as useful. They enable us to communicate and get things done. That is all it is intelligible to ask of them.

Beyond the usefulness of our language, we don't know the way things are in themselves. We may dub our lack of that metaphysical knowledge as *dao*. To tax our minds by trying to treat everything as 'one' differs from that admission of ignorance only in the emotion that accompanies it. In the end, neither scepticism nor monistic mysticism allows us to say anything about ultimate reality. They merely exhibit different attitudes towards saying… nothing.

10 Xunzi: Confucian conventionalism

The final chapter in Chinese language theory comes in the 'Rectifying Names' chapter of the *Xunzi*. Xunzi focuses on language because he wants to reassert that 'ritual' is the only standard of correct behaviour. He rejected Mencian intuitionism and gleaned insights from Zhuangzi and the dialecticians. The apparent moral he drew was that, since reality cannot be a standard of language correctness, the default standard must be convention.

Appeal to the usage of the sage-kings determines correct name use. The correct account of that usage is a historical tradition, by which Xunzi means the judgment of Confucian scholar-gentlemen. Thus Confucianism is vindicated by the weakness Mozi had exposed. Xunzi then goes on to construct an explicitly conventionalist theory of language which carries political implications.

Xunzi introduced an important clarification. He distinguished between two kinds of 'distinctions': *gui-jian* (noble-base) and *tong-yi* (same-different). The former correspond roughly to value distinctions and the latter to empirical or descriptive distinctions. The latter are the basis for interaction with other cultures, and a king is entitled to change these 'miscellaneous' terms. Even a king, however, cannot change conventional evaluative distinctions (ranks, titles, punishments or anything in the *li* (ritual)). For these we rely on the sage-kings' *dao* (guide) via the scholar-gentlemen's interpretation. Xunzi regards moral terms as conventional 'artifice' arising from thought, not from nature.

Political authorities rectify names for the original Confucian purposes (order and obedience). Xunzi treats the positions and paradoxes of the dialecticians solely in political terms. Philosophy of language causes social instability by undermining the public guiding language. Philosophers confuse the conventional relations of names and make 'is this/not this' unclear. This we must stop: we must have but one standard of terminology. The king, not disputing philosophers and warring schools, will govern introduction of new descriptive terms.

The king should keep three things in mind as he creates names:

(1) the reason for having names: the reason for having names is coordinating social behaviour and achieving social order. Hence, value terms govern how we assign descriptive terms.
(2) the basis of classifying as similar and different: we classify by taking the distinctions delivered by sense organs and using them according to the dictates of a heart imbued with the correct evaluative distinctions.
(3) the essentials of regulating names: the basis of regulating names is social order and the preservation of a stable, traditional scheme of language.

Xunzi's account of classifying similar and different takes a markedly empirical (epistemological) turn. Unlike the Mohists, Xunzi did not rely on claims that reality presented objective similarities and differences. Zhuangzi had argued that human standards of 'is this/not this' were no more natural than the opposing ones of other animals. Taking Zhuangzi's hint, Xunzi focused on human sense reactions to reality. Indeed, no neutral, inter-species ways of distinguishing things as similar and different exist. Still, while the senses of one species work differently from those of another, those of any one species makes similar distinctions.

All humans sense and respond to approximately the same range of natural distinctions (see Sense-data). The eyes of humans distinguish the same range and bands of colours, the mouth the same classifications of taste, the ears the same range and discriminations of pitch and so on. The shared nature of intra-species distinction-making underwrites the possibility of community and language. Thus, we abandon any appeal to cosmic nature and rely on what is pragmatically possible for humans in achieving natural human goals.

Our language conventionally clusters some sensible differences and ignores others. Historical, conventional standards dictate how it does this. These norms are transmitted into the cultured gentleman's heart when he masters the transmitted, sage-king's scheme of values. The heart rules the sense organs (as it did for Mencius). It determines what range of sensible discriminants counts as categories for moral purposes. Thus the categories mesh with the moral system of the sage - kings and match the clustering they originated.

It is clear that Xunzi absorbed a good deal of his contemporary theory of language. It is less clear if he understood the arguments and motivation. He deals with the problem of compound terms by ignoring it: 'If a single term is sufficient to convey the intent then use that and otherwise use a compound term.' The intent, presumably, is the conventionally understood intent. Xunzi does accept the Mohist view of names with varying scopes, and disowns the one-name-one-thing ideal. The only important kind of clarity or consistency is the constancy of convention.

Xunzi treats a number of related problems about names in sensitive fashion. He saw that spatial separation was a basis of describing two things of the same kind as two 'stuffs'. He then defines 'change' as being when a thing's spatial position does not change (exhibits characteristic continuity) and its type does. We then treat the thing as the same thing that has changed characteristics. This discussion of metamorphosis is the closest approximation of the classical Western problem of change (see Change).

Whether or not Xunzi understood the theories behind the paradoxes he criticizes, he clearly did not appreciate them. He exhibits no philosophical fascination with solving conceptual puzzles for their own sake or using them to drive theory. He criticizes paradoxical statements on primarily political grounds, namely the deleterious social effects of asserting their conclusions. Each upsets conventional ways of using terms. Xunzi's solution is political rather intellectual: ban them.

Xunzi classifies the paradoxes into three groups, which vaguely suggests the line of thought leading to them. Each, he argues, violates one of the three insights into names. The reason for naming is coordinating behaviour, so paradoxes which 'use names to confuse names' include the Mohist's claim that 'killing thieves is not killing men'. This uses a theory of names to yield a conclusion that sounds unconventional. So, we forbid saying them.

The second set 'use reality to confuse names', and the central examples are like Hui Shi's relativity paradoxes. These ignore the shared human empirical basis for assigning similarity and difference and use the fact that having different perspectives on reality might lead us to saying unconventional things about size and shape. So, the king will forbid saying them.

The final group use names to confuse reality and includes 'white horse not horse'. Xunzi's analysis does not help resolve the interpretive puzzles about the line of reasoning since it addresses only the pragmatic consequences of allowing such theorizing. His solution, once again, is for the king to avoid and prevent such distracting sophistry.

11 The aftermath: death of philosophy

One of Xunzi’s students, Han Feizi, was a minor royal in one of the Warring States, who became a central figure of the Legalist school. He had learned a smattering of Chinese theory of language, and he exaggerated the threat of interpretative anarchy to justify repressing philosophy and language creativity. He followed Xunzi’s argument that the ruler should enforce uniformity in language but rejected using a scholarly tradition as the norm. His theory of regulation and punishment was based on a crude argument about shape and name which takes us back to the unexplained Confucian notion that names by themselves guide action. An official post is a capsule description of functions (duties) the holder should perform. In the light of recent discoveries, this doctrine appears to be an application of the doctrine of a cult of ruler worship (Huang-Lao) which taught that the ‘guide’ was in nature and names were embedded in natural shapes.

Legalism became the official doctrine of the repressive Qin empire, which brought the classical period of Chinese philosophy to an abrupt halt (see Legalist philosophy, Chinese). In the aftermath, the insights of Chinese theory of language slipped into obscurity. Huang-Lao became the dominant theory surviving during China’s philosophical dark age until the importation of Buddhist theory. The early medieval Daoist interpreters argued that we can have names only for things we see. Suppression had worked its magic.

See also: Chinese philosophy; Confucian philosophy, Chinese; Daoist philosophy; Language, philosophy of; Mohist philosophy; Mozi; Xunzi; Zhuangzi

CHAD HANSEN

References and further reading

Fung Yu-lan (1952) *History of Chinese Philosophy*, trans. D. Bodde, Princeton, NJ: Princeton University Press.(A classic account that highlights the abstract interpretation of the ‘White Horse’ dialogue. Good for general purposes.)

Graham, A. (1978) *Later Mohist Logic, Ethics and Science*, Hong Kong: The Chinese University Press, and London: School of Oriental and African Studies.(The only source in English for the Later Mohist text. Difficult: understanding Graham virtually requires knowledge of classical Chinese.)

Graham, A. (1981) *Chuang-tzu: The Inner Chapters*, London: Allen & Unwin.(A translation of the *Zhuangzi*, including the summary of Hui Shi’s paradoxes.)

Graham, A. (1989) *Disputers of the Tao: Philosophical Argument in Ancient China*, La Salle, IL: Open Court.(An easier but less detailed treatment in the context of Graham’s account of ancient Chinese thought.)

Hansen, C. (1983) *Language and Logic in Ancient China*, Ann Arbor, MI: University of Michigan Press.(Difficult, based on Graham’s reconstruction. A philosophical argument for a radically different interpretation of the linguistic doctrines.)

Hansen, C. (1992) *A Daoist Theory of Chinese Thought*, New York: Oxford University Press.(An easier and more extended treatment in an account of ancient Chinese philosophy that emphasizes language.)

Thompson, K.O. (1995) ‘When a "White Horse" is not a "Horse"’, *Philosophy East and West* 45 (4): 481-99.(A definitive view of Gongsun Long’s most important dialogue.)

Logic in Islamic philosophy

Islamic logic was inspired primarily by Aristotle's logical corpus, the Organon (which according to a late Greek taxonomy also included the Rhetoric and Poetics). Islamic authors were also familiar with some elements in Stoic logic and linguistic theory, and their logical sources included not only Aristotle's own works but also the works of the late Greek Aristotelian commentators, the Isagōgēof Porphyry and the logical writings of Galen. However, most of the logical work of the Islamic philosophers remained squarely within the tradition of Aristotelian logic, and most of their writings in this area were in the form of commentaries on Aristotle.

For the Islamic philosophers, logic included not only the study of formal patterns of inference and their validity but also elements of the philosophy of language and even of epistemology and metaphysics. Because of territorial disputes with the Arabic grammarians, Islamic philosophers were very interested in working out the relationship between logic and language, and they devoted much discussion to the question of the subject matter and aims of logic in relation to reasoning and speech. In the area of formal logical analysis, they elaborated upon the theory of terms, propositions and syllogisms as formulated in Aristotle's Categories, De interpretatione and Prior Analytics. In the spirit of Aristotle, they considered the syllogism to be the form to which all rational argumentation could be reduced, and they regarded syllogistic theory as the focal point of logic. Even poetics was considered as a syllogistic art in some fashion by most of the major Islamic Aristotelians.

Since logic was viewed as an organon *or instrument by which to acquire knowledge, logic in the Islamic world also incorporated a general theory of argumentation focused upon epistemological aims. This element of Islamic logic centred upon the theory of demonstration found in Aristotle's Posterior Analytics, since demonstration was considered the ultimate goal sought by logic. Other elements of the theory of argumentation, such as dialectics and rhetoric, were viewed as secondary to demonstration, since it was held that these argument forms produced cognitive states inferior in certitude and stability to demonstration. The philosopher's aim was ultimately to demonstrate necessary and certain truth; the use of dialectical and rhetorical arguments was accounted for as preparatory to demonstration, as defensive of its conclusions, or as aimed at communicating its results to a broader audience.*

1 The subject matter and aims of logic

As was their custom in discussing all the branches of philosophy, the Islamic philosophers devoted considerable attention to identifying the subject matter studied by logic and the aims at which logical studies are directed. Al-Farabi, whose logical and linguistic writings comprise the majority of his philosophical output, epitomizes the approach to logic that is characteristic of Islamic Aristotelianism. In his *Ihsa' al-'ulum (Enumeration of the Sciences)*, he defines logic as an instrumental, rule-based science aimed at directing the intellect towards the truth and safeguarding it from error in its acts of reasoning. He defends the need for such a science of reasoning on the grounds that it is possible for the mind to err in at least some of its acts, for example, in those in which the intelligibles sought are not innate, but are rather attained discursively and empirically 'through reflection and contemplation'. Al-Farabi compares logic to tools such as rulers and compasses, which are used to ensure exactness when we measure physical objects subject to the errors of sensation. Like these tools, logical measures can be employed by their users to verify both their own acts of reasoning and the arguments of others. Indeed, logic is especially useful and important to guide the intellect when it is faced with the need to adjudicate between opposed and conflicting opinions and authorities.

Al-Farabi's view of logic as a rule-based science which governs the mind's operations over intelligibles is repeated in many of his introductory logical works, and it formed the foundation for Ibn Sina's later refinements (see Ibn Sina). In the opening chapters of his *al-Madkhal (Introduction)*, the first logical book of his encyclopaedic work *al-Shifa' (Healing)*, Ibn Sina describes the purpose of logic as one of enabling the intellect to acquire 'knowledge of the unknown from the known'. Like al-Farabi, he defends the need for logic by arguing that the innate capacities of reasoning are insufficient to ensure the attainment of this purpose, and thus they require the aid of an art. While there may be some cases in which innate intelligence is sufficient to ensure the attainment of true knowledge, such cases are haphazard at best; he compares them to someone who manages to hit a target on occasion without being a true marksman. The most important and influential innovation that Ibn Sina introduces

into the characterization of logic is his identification of its subject matter as 'second intentions' or 'secondary concepts', in contrast to 'first intentions'. This distinction is closely linked in Ibn Sina's philosophy to his important metaphysical claim that essence or quiddity can be distinguished from existence, and that existence in turn can be considered in either of its two modes: existence in concrete, singular things in the external world; or conceptual existence in one of the soul's sensible or intellectual faculties (see Existence).

In *al-Madkhal*, Ibn Sina argues that logic differs from the other sciences because it considers not conceptual existence as such (this would be psychology), but rather the accidents or properties that belong to any quiddity by virtue of its being conceptualized by the mind. These properties, according to Ibn Sina, include such things as essential and accidental predication, being a subject or being a predicate, and being a premise or a syllogism (see Logical form §1). It is these properties that allow the mind to connect concepts together in order to acquire knowledge of the unknown; they provide the foundation for the rules of reasoning and inference that logic studies. They are moreover formal properties in the sense that, as properties belonging to all concepts in virtue of their mental mode of existence, they are entirely independent of the content of the thought itself; they are indifferent to the intrinsic natures of the quiddities which they serve to link together.

In the *Ilahiyyat (Metaphysics)* of *al-Shifa'*, Ibn Sina introduces the terminology of first and second 'intentions' or concepts in order to express the relation between the concepts of these quiddities themselves - which are studied in the theoretical sciences - and the concepts of the states and accidents of their mental existence which logic studies: 'As you know, the subject matter of logical science is second, intelligible intentions (*al-ma'ani al-ma'qula al-thaniyya*) which are dependent upon the primary intelligible intentions with respect to some property by which they lead from the known to the unknown' (*Ilahiyyat* Book 1, ch. 2, in Anawati and Zayed 1960: 10-11). For example, the second intentions of 'being a subject' and 'being a predicate' are studied in logic independently of whatever first intentions function as the subject and predicate terms in a given proposition, for example, 'human being' and 'rational animal' in the proposition 'a human being is a rational animal'. The logical second intentions depend upon the first intentions because the first intentions are the conceptual building blocks of the new knowledge which second intentions link together: but logic studies the second intentions in abstraction from whatever particular first intentions the logical relations depend upon in any given case.

2 Logic, language and grammar

The attention that Ibn Sina and al-Farabi devote to the proper characterization of the subject matter of logic stems in part from a concern to distinguish logic from grammar. In the ancient and medieval traditions, the study of logic was closely tied to the philosophical consideration of language (see Language, medieval theories of; Logic, ancient; Logic, medieval), and for this reason many Arabic grammarians - whose linguistic theories were developed to a high degree of complexity and sophistication - were contemptuous of the philosophers for importing Greek logic, which they saw as a foreign linguistic tradition, into the Arabic milieu. This attitude toward Greek logic is epitomized in a famous debate reported to have taken place in Baghdad in 932 between the grammarian Abu Sa'id al-Sirafi and Abu Bishr Matta, a Syriac Christian who translated some of Aristotle's works into Arabic and is purported to have been one of al-Farabi's teachers. The extant account of the debate is heavily biased towards al-Sirafi, who attacks logical formalism and denies the ability of logic to act as a measure of reasoning over and above the innate capacities of the intellect itself. His principal claims are that philosophical logic is nothing but Greek grammar warmed over, that it is inextricably tied to the idiom of the Greek language and that it has nothing to offer speakers of another language such as Arabic.

It is against the background of such attacks that the discussions of the relations between logic, language and grammar by al-Farabi, Ibn Sina and al-Farabi's pupil Yahya ibn 'Adi (also a Syriac Christian and translator), are to be understood. Al-Farabi and Yahya both present essentially the same perspective on the relations between logic and language, a moderate perspective which Ibn Sina later rejects. In the *Ihsa' al-'ulum*, al-Farabi argues that logic and grammar both have some legitimate interest in language, but whereas grammatical rules primarily govern the use of language, logical rules primarily govern the use of intelligibles:

> And this art [of logic] is analogous to the art of grammar, in that the relation of the art of logic to the intellect and the intelligibles is like the relation of the art of grammar to language and expressions. That is, to every rule for expressions which the science of grammar provides us, there is a corresponding [rule] for intelligibles which the science of logic provides us.

(*Ihsa' al-'ulum*, in Amin 1968: 68)

More precisely, al-Farabi explains that although grammar and logic share a mutual concern with expressions, grammar provides rules that govern the correct use of expressions in a given language, but logic provides rules that govern the use of any language whatsoever in so far as it signifies intelligibles. Thus, logic will have some of the characteristics of a universal grammar, attending to the common features of all languages that reflect their underlying intelligible content. Some linguistic features will be studied in both logic and grammar, but logic will study them as they are common, and grammar in so far as they are idiomatic. On the basis of this comparison with grammar, then, al-Farabi is able to complete his characterization of the subject matter of logic as follows: 'The subject-matters of logic are the things for which [logic] provides the rules, namely, intelligibles in so far as they are signified by expressions, and expressions in so far as they signify intelligibles' (*Ihsa' al-'ulum*, in Amin 1968: 74).

Like al-Farabi, Yahya ibn 'Adi, in a treatise entitled *Maqala fi tabyin al-fasl bayna sina'atay al-mantiq al-falsafi wa-al-nahw al-'arab (On the Difference Between Philosophical Logic and Arabic Grammar)*, makes his case for the independence of logic from grammar based upon the differences between the grammar of a particular nation and the universal science of logic. He argues that the subject matter of grammar is mere expressions (*al-alfaz*), which it studies from the limited perspective of their correct articulation and vocalization according to Arabic conventions. The grammarian is especially concerned with language as an oral phenomenon; the logician alone is properly concerned with 'expressions in so far as they signify meanings' (*al-alfaz al-dalla 'ala al-ma'ani*) (*Maqala fi tabyin*, in Endress 1978: 188). To support this claim, Yahya points out that changing grammatical inflections do not affect the basic signification of a word: if in one sentence a word occurs in the nominative case, with the appropriate vocalization, its signification remains unchanged when it is used in another sentence in the accusative case and with a different vocal ending.

In Ibn Sina's view, however, such accounts of the logician's interest in language and its differences from that of the grammarian did not go far enough. In keeping with his own understanding of logic as the science which studies second intentions, Ibn Sina criticized such earlier attempts to introduce linguistic concerns into the subject matter of logic. In *al-Madkhal*, Ibn Sina labels as 'stupid' those who say that 'the subject matter of logic is speculation concerning expressions in so far as they signify meanings (*ma'ani*)'. However, Ibn Sina does not deny that the logician is sometimes or even often required to consider linguistic matters; his objection is to the inclusion of language as an essential constituent of the subject matter of logic. The logician is only incidentally concerned with language because of the constraints of human thought and the practical exigencies of learning and communication. Ibn Sina goes so far as to claim that, 'if logic could be learned through pure thought so that meanings alone could be attended to in it, then it would dispense entirely with expressions'; but since this is not in fact possible, 'the art of logic is compelled to have some of its parts come to consider the states of expressions' (*al-Madhkal*, in Anawati *et al.* 1952: 22-3). For Ibn Sina, then, logic is a purely rational art whose purpose is entirely captured by its goal of leading the mind from the known to the unknown; only accidentally and secondarily can it be considered a linguistic art.

3 Conceptualization and assent

While the close links between logic and linguistic studies emerge in the Islamic philosophers' consideration of the subject matter of logic, the links between logic and epistemology come to the fore in the consideration of the divisions within logic and the order of the books within Aristotle's Organon. All the principal Islamic Aristotelians organize their understanding of the divisions of logic around the epistemological couplet of *tasawwur* (conceptualization), and *tasdiq* (assent), which constitute for them the two states of knowledge that logic aims to produce in the intellect.

Conceptualization is the act of the mind by which it grasps singular (though not necessarily simple) essences or quiddities, such as the concept of 'human being'. Assent, by contrast, is the act of the intellect whereby it makes a determinate judgment to which a truth-value can be assigned; in fact, conceptualization is defined in Islamic philosophy principally by contrast with assent. Thus, any act of knowledge that does not entail the assignment of a truth-value to the proposition that corresponds to it will be an act of conceptualization alone, not assent. More specifically, the Islamic philosophers link assent to the affirmation or denial of the existence of the thing

conceived, or to the judgment that it exists in a certain state, with certain properties. Thus, assent presupposes some prior act of conceptualization, although conceptualization does not presuppose assent.

One of the purposes of including a consideration of the *tasawwur-tasdiq* dichotomy in introductory discussions of the purpose of logic is to provide an epistemological foundation for the two focal points of Aristotelian logic, the definition and the syllogism (see Logical form §1). The purpose of the definition is identified as the production of an act of conceptualization, and the purpose of the syllogism is identified as causing assent to the truth of a proposition. However, since the definition and the syllogism are both considered in the *Prior* and *Posterior Analytics* and the works that come after them in the Organon, the study of the ways of producing conceptualization and assent presupposes as its foundation the study of single terms and propositions in the *Categories* and *De interpretatione*.

4 Predicables, categories and propositions

In keeping with the ancient Greek tradition, the Islamic philosophers considered the books of the Organon to be an ordered series which begins with the study of the signification of simple terms in the *Categories* and then proceeds to the study of propositions in the *De interpretatione*. In addition to these two Aristotelian texts, a work of the Neoplatonist Porphyry, known as the *Isagōgē(Introduction)*, was appended to the beginning of this series as an introduction to the study of the *Categories* (see Aristotle §7). It was concerned with the five predicables: genus, species, difference, property and accident. While all of the Islamic Aristotelians wrote commentaries on the *Isagōgē* and utilized its grouping of the predicables, not all were convinced of its utility as an introduction to Aristotle. Ibn Rushd openly expresses such doubts in the introduction to his *Talkhis kitab al-maqulat (Middle Commentary on the Categories)*, where he indicates that his original intention was to omit the *Isagōgē* entirely from his series of middle commentaries on the Organon. At the end of his work on the *Isagōgē* itself, he explains bluntly that he does not believe that Porphyry's text is a helpful introduction to the study of logic and questions whether it is really a logical text at all. His sole reason for completing the commentary, he tells us, was to comply with a request made by his friends.

The logical character of the *Categories* presented a related problem for other Islamic philosophers. In the introduction to his *Sharh al-'ibarah (Great Commentary on the De interpretatione)*, al-Farabi rehearses some of the controversies inherited from the Greek tradition over the relations between the *Categories* and the *De interpretatione*. As al-Farabi points out, the *De interpretatione* can be understood quite well without a prior knowledge of the *Categories*, and the former work makes no explicit references to the latter. Moreover, the *De interpretatione* is principally concerned with the formal relations amongst propositions, such as contradiction and contrariety, whereas the *Categories* is concerned with the signification or meaning of terms as such. Furthermore, in its opening chapters, the *De interpretatione* considers in formal terms the simple parts of which propositions are composed, that is, the noun and the verb. Despite these concerns, however, al-Farabi opts for the traditional ordering of these books on the grounds that the *Categories* is relevant to the whole of logic, since it studies 'the simplest of the subject matters in which logic actualizes itself'. In his *Falsafa Aristutalis (Philosophy of Aristotle)*, al-Farabi opts for a similar solution to the logical status of the Categories, explaining that it comprises an investigation and classification of 'the instances of being from which the first premises are compounded', which are 'the primary significations of the expressions generally accepted by all' (*Falsafa Aristutalis*, in Mahdi 1969: 82-3).

Al-Farabi's misgivings in both of these texts stem from the largely ontological focus of the Aristotelian *Categories*, which calls into question its placement within the Organon. This concern was echoed later by Ibn Sina, who points out that many of the discussions in the *Categories* would be better placed in metaphysics or psychology, since they pertain to the study of expressions as directly signifying external or mental beings, in other words, to first rather than to second intentions. But since the *Categories* is useful in instructing us how to formulate definitions - which is one of the principal goals of logic - its placement in the Organon can be justified on practical grounds.

Islamic philosophers viewed *De interpretatione* as a study of the composition and truth-values of categorical propositions. Thus al-Farabi, in his great commentary on this text, explains that the term 'interpretation' used in the title of the work means 'complete statement' (*al-qawl al-tamm*). A complete statement, according to al-Farabi, must be one which causes a complete understanding in the mind; in other words, one in which assent occurs along

with conceptualization. This is achieved principally by a simple, predicative, categorical statement (*al-qawl al-jazim al-hamli al-basit*) which affirms or denies a predicate of its subject.

5 Theory of argumentation

For the entire Islamic tradition, the crowning glory of Aristotelian logic is the syllogistic theory outlined in the *Prior* and *Posterior Analytics*, especially the latter. The purpose of logic is to provide the means whereby knowledge is to be acquired, and the most valuable type of knowledge is that which is certain and necessary, that is, knowledge gained according to the paradigm of demonstrative science laid out in the *Posterior Analytics*. This part of logic, in the words of al-Farabi's *Ihsa' al-'ulum*, is 'the strongest and pre-eminent in dignity and authority. Logic seeks its primary aim in this part alone, and the rest of its parts are only for its sake' (*Ihsa' al-'ulum*, in Amin 1968: 89). Even the formal study of the syllogism itself is primarily undertaken for the sake of its employment in demonstrations.

In their formal syllogistic theory, the Islamic Aristotelians mainly follow Aristotle's *Prior Analytics*. While they are aware of the fourth figure traditionally ascribed to Galen, the tendency is to dismiss this figure as superfluous and intuitively implausible, as Ibn Sina does in the seventh method of his *al-Isharat wa-'l-tanbihat (Remarks and Admonitions)*; or to ignore it entirely, as al-Farabi does in his *Kitab al-qiyas (Book on the Syllogism)*. Similarly, the Arabic philosophers knew of the alternative propositional logic of the Stoics and incorporated elements of it in their discussions of conditional or hypothetical (*shartiyyah*) syllogisms (see Logic, ancient). However, they did not accept the Stoic inference schemata, nor did they treat conditional connectives as truth-functional, since they did not consider the parts of conditional statements to be complete propositions in their own right. Moreover, for the Islamic logicians 'conditional' was a generic term which included both 'conjunctive' (*al-muttasilat*) conditionals (of the form, 'if… then') and 'disjunctive' (*al-munfasilat*) conditionals (of the form, 'either… or'). Conditional syllogisms of both sorts were viewed as relying upon a process of 'reiteration' or 'repetition' (*istithna'*), a term which referred to the repetition of the antecedent or the consequent, or one of the two disjuncts, in so far as it formed the second premise of a syllogism. Thus in the conjunctive conditional syllogism, 'If it is daytime, then it is light; but it is daytime, therefore it is light', 'it is daytime' would be labelled the *mustathna'* or reiterated premise, since it is by its restatement that the syllogism reaches its conclusion.

When we turn to the specific application of syllogistic theory to particular types of argumentation, the epistemological concerns of Islamic logic surface once more. In particular, the Islamic philosophers explained the primacy of demonstration, and the ancillary role of dialectical, rhetorical, poetic and sophistical syllogisms, by reference to the epistemic status of the premises used in each type of syllogism, and the type of assent they could produce to the conclusion of the syllogisms in which they were employed. The classification of syllogisms and their premises according to the nature of their assent is found in the logical writings of all the major Islamic philosophers, but the most complete and systematic classification of premises occurs in three of Ibn Sina's works, *al-Burhan (Demonstration)*, in *al-Shifa'*, *al-Najah (Deliverance)* and *al-Isharat wa-'l-tanbihat (Remarks and Admonitions)*. Although these three accounts differ somewhat in the number and variety of the premises listed in each, generally they present a single and consistent theory. Demonstrative syllogisms are composed of premises which necessitate assent and include self-evident first principles as well as sensible, empirically evident propositions. Dialectical syllogisms are based upon generally accepted beliefs (*al-mashhurat*), which are equivalent to the *endoxa* of Aristotle's Topics; on premises granted for the purposes of dialectical debate; and in general, on all premises assented to because they are universally accepted by all people, or by people deemed authoritative. Rhetorical syllogisms are similar to dialectical ones, except that they are accepted unreflectively and on the basis of a more limited authority, relative, for example, to a particular group or sect; as such, they are only supposed or presumed to be 'generally-accepted beliefs'. Sophistical premises are those accepted because of some misleading resemblance to another type of premise, and poetic premises are those that produce a motion in the faculty of imagination (*al-takhyil*), not an act of intellectual assent.

The inclusion of rhetorical and poetical syllogisms in this enumeration reflects a common assumption among Islamic philosophers that Aristotle's *Rhetoric* and *Poetics* are parts of his logical Organon. This assumption was inherited by the Islamic tradition from the Greek commentators, and it was used by them in part to account for the differences between philosophical and popular modes of discourse and argumentation, particularly in the context of discussions of the relations between philosophy and religion. The Islamic philosophers held that whereas

philosophers rely principally upon demonstrative and dialectical syllogisms, religious leaders and theologians generally use rhetorical and poetical syllogisms to persuade the general populace. Religion is thus viewed as an image or reflection of philosophical, demonstrative truth propounded in language and argument-forms that can be easily understood by the mass of humanity.

The place of dialectic within the theory of argumentation is perhaps the most ambivalent in Islamic logic. While dialectic is seen as inferior to demonstration, its importance for philosophy is none the less recognized. A good example of this is found in al-Farabi's enumeration in his *Kitab al-jadal (Book on Dialectic)* of the ways in which dialectic serves philosophers. According to al-Farabi, dialectic hones argumentative skills, introduces the principles of the special demonstrative sciences, alerts the mind to the self-evident principles of demonstration, helps to develop communicative skills and provides the means for refuting sophistry. Of these five uses, only the fourth is external to the proper aims of philosophy and closer to the tasks usually reserved to theology and religion. The other four pertain to the learning or acquisition of truly philosophical skills, even if they lie outside the strictly demonstrative aims that are the ultimate end of philosophy.

In the case of the theory of demonstration itself, Islamic logicians organized their commentaries on the *Posterior Analytics* around the definition and the demonstrative syllogism as the means by which both conceptualization and assent are most perfectly attained. Al-Farabi's *Kitab al-burhan (Book on Demonstration)* offers an excellent summary of the standard approach taken by Islamic philosophers to theory of demonstration and its epistemological aims. Just as he identified the categorical statement as the embodiment of perfect assent on the propositional level, here al-Farabi identifies demonstrative certitude as complete or perfect assent on the level of syllogistic inference. Moreover, certitude is defined by al-Farabi in terms of what we would now label 'second-order' knowledge:

> Certitude is for us to believe, concerning the truth to which we have assented, that it is not possible at all for what we believe about this matter to be different from what we believe it to be; and in addition to this for us to believe, concerning our belief, that another belief is not possible - in the sense that whenever some belief about the first belief is formed, it is impossible for it to be otherwise, and so on *ad infinitum*.
>
> (*Kitab al-burhan*, in al-'Ajam and Fakhry 1986-7, 4: 20)

Certitude requires not just knowledge of a conclusion, *p*, but knowing that we know *p*. This sort of certitude al-Farabi calls 'necessary certitude'. However, he also allows for non-necessary certitude, which holds 'only at a particular time', and thus can be applied to propositions about merely contingent beings: 'Necessary certitude and necessary existence are convertible in entailment, for what is verified as necessarily certain is necessarily existent' (*Kitab al-burhan*, in al-'Ajam and Fakhry 1986-7, 4: 22). While al-Farabi recognizes both of these varieties of certitude to be forms of perfect assent, in his view necessary certitude alone fulfils the strict conditions of Aristotelian demonstration, since it alone will pertain to objects which cannot be other than they are.

Al-Farabi's remarks on the utility of dialectic, combined with his extension of the notion of perfect assent beyond the confines of strict and necessary demonstration, illustrate the overall breadth of the Islamic philosophers' theories of argumentation. Despite their professions of the primacy of the demonstrative paradigm within philosophy, the Islamic Aristotelians recognized a broad range of legitimate and useful argument forms and acknowledged their importance as philosophical tools leading to knowledge of the unknown.

See also: Aristotelianism in Islamic philosophy; Aristotle; al-Farabi; Ibn Sina; Logical form; Logic, ancient; Logic, medieval; Logic, philosophy of; Meaning in Islamic philosophy; Syntax

DEBORAH L. BLACK

References and further reading

Abed, S.B. (1991) *Aristotelian Logic and the Arabic Language in Al-Farabi*, Albany, NY: State University of New York Press.(An excellent consideration of the central issues in al-Farabi's linguistic philosophy.)

Black, D.L. (1990) *Logic and Aristotle's 'Rhetoric' and 'Poetics' in Medieval Arabic Philosophy*, Leiden: Brill. (Discusses the interpretation of these Aristotelian texts as works of logic.)

al-Farabi (*c*.870-950) *Ihsa' al-'ulum (Enumeration of the Sciences)*, ed. U. Amin, Cairo: Librairie Anglo-Égyptienne, 3rd edn, 1968.(Al-Farabi's major work, in which he investigates in detail different forms of knowledge.)

al-Farabi (*c*.870-950) *Kitab al-burhan (Book on Demonstration)*, in R. al-'Ajam and M. Fakhry (eds) *al-Mantiq 'inda al-Farabi*, Beirut: Dar el-Mashreq, 1986-7, 4 vols.(*Al-Mantiq* is a collection of al-Farabi's writings on logic.)

al-Farabi (*c*.870-950) *Falsafah Aristutalis (Philosophy of Aristotle)*, trans. M. Mahdi in *Alfarabi's Philosophy of Plato and Aristotle*, Ithaca, NY: Cornell University Press, 1969.(The *Falsafah Aristutalis* contains al-Farabi's account of the Organon.)

al-Farabi (*c*.870-950) *Sharh al-'ibarah (Great Commentary on De interpretatione)*, trans. F.W. Zimmerman, *Al-Farabi's Commentary and Short Treatise on Aristotle's 'De Interpretatione'*, Oxford: Oxford University Press, 1981.(Very learned introduction which sets al-Farabi's text against the background of the Greek and Arabic traditions in logic.)

al-Farabi (*c*.870-950) Commentary on the *Isagōgē*, ed. and trans. D.M. Dunlop, 'Al-Farabi's *Eisagoge*', *Islamic Quarterly* 3, 1956: 117-38.(Excellent translation of this crucial logical text, given the significance of the *Isagōgē* for this period of logic.)

al-Farabi (*c*.870-950) *Risalah* on Logic, ed. and trans. D.M. Dunlop, 'Al-Farabi's Introductory *Risalah* on Logic', *Islamic Quarterly* 3, 1957: 224-35.(A brief but influential summary of his views on logic.)

al-Farabi (*c*.870-950) Introductory Sections on Logic, ed. and trans. D.M. Dunlop, 'Al-Farabi's Introductory Sections on Logic', *Islamic Quarterly* 2, 1955: 264-82.(Another translation of a section on logic, designed by al-Farabi as a propadeutic to the study of philosophy itself.)

al-Farabi (*c*.870-950) Paraphrase on the *Categories*, ed. and trans. D.M. Dunlop, 'Al-Farabi's Paraphrase of the *Categories* of Aristotle', *Islamic Quarterly* 4, 1958-9: 168-97; 5, 1958-9: 21-54.(A short summary of what al-Farabi took the main points of the *Categories* to be.)

al-Farabi (*c*.870-950) Commentary on the *Prior Analytics*, trans. N. Rescher, *Al-Farabi's Short Commentary on Aristotle's 'Prior Analytics'*, Pittsburgh, PA: Pittsburgh University Press, 1963.(Another of al-Farabi's Aristotelian commentaries.)

Galston, M. (1981) 'Al-Farabi on Aristotle's Theory of Demonstration', in P. Morewedge (ed.) *Islamic Philosophy and Mysticism*, Delmar, NY: Caravan Books, 23-34.(Clear account of how al-Farabi developed Aristotle's notion of demonstrative reasoning, the best form of argumentation.)

Gyekye, D. (1971) 'The Terms *Prima intentio* and *Secunda intentio* in Arabic Logic', *Speculum* 46: 32-48. (Explanation of two key logical terms, linking them to their Greek and Latin equivalents and explaining how they are used in al-Farabi's logic.)

Ibn 'Adi, Yahya (893-974) *Maqalah fi tabyin al-fasl bayna sina'atay al-mantiq al-falsafi wa-al-nahw al-'arab (Treatise on the Difference between the Arts of Philosophical Logic and of Arabic Grammar)*, ed. G. Endress, *Journal of the History of Arabic Science* 2, 1978: 192-81.(Influential work from the early years when the divisions between logic and language were still highly controversial.)

Ibn Rushd (*c*.1170) Short Commentaries on Aristotle, ed. and trans. C.E. Butterworth, *Averroes' Three Short Commentaries on Aristotle's 'Topics', 'Rhetoric', and 'Poetics'*, Albany, NY: State University of New York Press, 1977.(Ibn Rushd's short commentaries on areas of thought which he thought were less strongly logical than demonstrative and dialectical reasoning, yet which still embody logical techniques in some form.)

Ibn Rushd (*c*.1174) Middle Commentary on the *Isagōgē*, trans. H.A. Davidson, *Averroes' Middle Commentary on Porphyry's 'Isagoge' and on Aristotle's 'Categoriae'*, Cambridge, MA: Mediaeval Academy of America, 1969.(A middle commentary on a crucial logical text by Porphyry, which was important in linking Ibn Rushd's use of logical language with that of his Islamic predecessors.)

Ibn Rushd (*c*.1174) Middle Commentaries on *Categories* and *De interpretatione*, trans. C.E. Butterworth, *Averroes' Middle Commentary on Aristotle's 'Categories' and 'De interpretatione'*, Princeton, NJ: Princeton University Press, 1983.(Ibn Rushd's middle commentaries on two Aristotelian texts which are important for what they show of how he developed some of the key terms of his philosophical logic.)

Ibn Sina (980-1037) *al-Isharat wa-'l-tanbihat (Remarks and Admonitions)*, Part One trans. S.C. Inati, *Remarks and Admonitions, Part One: Logic*, Toronto: Pontifical Institute of Mediaeval Studies, 1984.(The logic portion of *al-Isharat wa'l-tanbihat*.)

Ibn Sina (*c*.1014-20) *al-Shifa' (Healing), al-Ilahiyyat (Theology)*, vol. 1 (Books 1-5) ed. G. Anawati and S. Zayed; vol. 2 (Books 6-10) ed. S. Dunya, M.Y. Moussa, and S. Zayed, Cairo: Organisation Générale des Imprimerie Gouvernementales, 1960.(The part of *al-Shifa'* which deals with theological issues.)

Ibn Sina (*c*.1014-20) *al-Shifa' (Healing), al-Madkhal (Isagōgē)*, ed. G. Anawati, M. El-Khodeiri and F. al-Ahwani, rev. I. Madkour, Cairo: Al-Matba'ah al-Amiriyyah, 1952.(Ibn Sina's account of Porphyry's *Isagōgē*, giving him the opportunity to specify the meaning he gives to logical language.)

Inati, S. (1996) 'Logic', in S.H. Nasr and O. Leaman, *History of Islamic Philosophy*, London: Routledge, ch. 48,

802-23.(Analysis of the main concepts in Islamic logic, and the role of logic itself in philosophy.)

Lameer, J. (1994) *Al-Farabi and Aristotelian Syllogistics: Greek Theory and Islamic Practice*, Leiden: Brill.(A thorough study of al-Farabi's logical writings and their ancient sources.)

Madkour, I. (1969) *L'organon d'Aristotle dans le monde arabe (The Organon of Aristotle in the Arab World)*, 2nd edn, Paris: Vrin. (A useful overview of the logic portions of Ibn Sina's Shifa'.)

Mahdi, M. (1970) 'Language and Logic in Classical Islam', in G.E. von Grunebaum (ed.) *Logic in Classical Islamic Culture*, Wiesbaden: Harrasowitz, 51-83.(An account of the debate between al-Sirafi and Abu Bishr Matta.)

Margoliouth, D.S. (1905) 'The Discussion Between Abu Bishr Matta and Abu Sa'id al-Sirafi on the Merits of Logic and Grammar', *Journal of the Royal Asiatic Society*: 79-129(A translation of the famous debate; for a fuller account see also Mahdi 1970.)

Marmura, M.E. (1963) *Studies in the History of Arabic Logic*, Pittsburgh, PA: University of Pittsburgh Press.(An uneven collection of studies and translations but with some useful items.)

Marmura, M.E. (1975) 'Ghazali's Attitude to the Secular Sciences and Logic', in G.F. Hourani (ed.) *Essays on Islamic Philosophy*, Albany, NY: State University of New York Press.

Rescher, N. (1964) *The Development of Arabic Logic*, Pittsburgh, PA: University of Pittsburgh Press. (Demonstration that in spite of his opposition to philosophy, al-Ghazali was an enthusiastic supporter of logic. Somewhat dated but still useful.)

Rescher, N. (1980) 'Avicenna on the Division of the Sciences in the *Isagoge* of his *Shifa*'', *Journal for the History of Arabic Science* 4: 239-50.(Very clear analysis of the text.)

Sabra, A.I. (1980) 'Avicenna on the Subject Matter of Logic', *Journal of Philosophy* 77: 757-64.(Excellent analysis of Ibn Sina's approach to logic.)

Wolfson, H.A. (1973) 'The Terms *Tasawwur* and *Tasdiq* in Arabic Philosophy and Their Greek, Latin and Hebrew Equivalents', in I. Twersky and G.H. Williams (eds) *Studies in the History and Philosophy of Religion*, Cambridge, MA: Harvard University Press, vol. 1, 478-92.(Important analysis of this critical distinction in Islamic logic, with discussions of its origins in wider philosophy.)

Zimmermann, F.W. (1972) 'Some Observations on al-Farabi and the Logical Tradition', in S.M. Stern *et al.* (eds) *Islamic Philosophy and the Classical Tradition*, Oxford: Oxford University Press, 517-46.(Discussion of the role which al-Farabi played in establishing the notion of logic as a separate theoretical inquiry.)

Logic in Japan

'Logic' became an explicit topic in Japanese philosophy only in the twentieth century. Most effort has been directed to developing a dialectical logic in a Hegelian mode rather than a symbolic system. The Japanese term coined for logic in this sense is ronri *(the 'principles of discourse or argument'). The term* ronrigaku *(*ronri *+ -ology) is the more common term for formal, symbolic logic. In the twentieth century two Japanese philosophers, Nishida Kitarō(1870-1945) and Tanabe Hajime (1885-1962), developed distinctive dialectical logics that drew on key assumptions from traditional Japanese thought, but followed a Western style of analysis and articulation.*

Nishida's logic of basho *(place,* topos, *field) located the contextual premises of empirical and idealist judgments within a trans-judgmental domain called the 'acting-intuiting'. He related this to his logic of the predicate, which rejected the Aristotelian priority of the grammatical subject as signifier of substance, instead making the subject the qualifier of the predicate, the signifier of the event. Tanabe criticized Nishida's system as ahistorical and transcultural. His logic of species gave priority to the specificity of cultural and historical embededness, the middle ground between the universal ought and individual freedom.*

1 Dialectical logic

To appreciate the development of logic in modern Japan, it is useful to bear in mind three tendencies from traditional Japanese philosophy. First, Japanese thought has tended to emphasize internal rather than external relations. That is, in a relationship the two relatents, *a* and *b*, are generally understood to be not separate entities that have been connected by a relating principle, but to be two overlapping entities: part of *a* is *b* and part of *b* is *a*. Traditional Buddhist philosophy assumes no entity ever exists independently as a discrete substance; it is always in flux and depends for its existence on some other entity (see Buddhist philosophy, Japanese).

The second point in traditional Japanese thought is the application of internal relations to the part/whole model (see Mereology). In esoteric, Tendai and Kegon Buddhism, the whole has typically not been understood as simply the sum of the parts. Because the whole and parts are internally related, the part (as in a recursive set) always reflects in some way the whole.

The third point about traditional Japanese thought is that even opposites are often understood as being internally related. If, therefore, *a* and not-*a* overlap in some way, that suggests that *a* can only be fully *a* insofar as it contains something of not-*a*. At the least, in defining an entity, it is typically distinguished from and therefore related to its opposite. In East Asian texts, such as the translation of the Indian Buddhist *Perfection of Wisdom* texts, the Sino-Japanese term *soku* is commonly used to express this identity within difference: *a soku* not-*a*. The term *soku* can be used as a special kind of copula or can be used as a phrase meaning something like 'that is to say'. English translations (not completely successfully) tend to capture this sense by using the Latinate phrases '*a qua* not-*a*' or '*a sive* not-*a*' (see Kyoto School §3).

When Western philosophy entered Japan in the latter half of the nineteenth century, Japanese philosophers became aware that the dominant Western modes of logical expression could not adequately express the premodern tendencies just explained. Among Western systems, Hegel's dialectical logic (see Hegel, G.W.F.) seemed to most resemble the *soku* way of relating opposites and many Japanese philosophers studied it intensely in the early twentieth century. In the end, however, most Japanese philosophers found it unacceptable. In fact, it seemed almost the mirror image of the kind of dialectical logic sought by the Japanese. First, Hegelian logic is grounded in being, whereas traditional Asian (especially Buddhist and Daoist) philosophy gave more primacy to nonbeing or nothingness. Second, the Hegelian system of dialectic expresses how static beings are transformed into events of becoming. In Japanese thought, however, events are the starting point and logical analysis is used to abstract from it discrete beings within the becoming. Third, the Hegelian dialectic is progressive. It shows the development of thought through time into what it is becoming. In Japanese thought, however, the dialectical process is more commonly used regressively to understand how thought comes about and out of what it is formed. Therefore, although modern Japanese philosophers have continued to use the term 'dialectic' (*benshōhō*), except for Marxists and certain other neo-Hegelians, the term usually refers to a distinctively Japanese logic of dialectic.

Two of the best known dialectical Japanese logics were developed by Nishida Kitarō (1870-1945) and Tanabe

Hajime (1885-1962). Both figures were trained in mathematics (in fact, in his younger days, Tanabe was Japan's most renowned philosopher of mathematics and science), well versed in Hegelianism and steeped in Buddhism (Nishida was a Zen Buddhist, Tanabe a Pure Land Buddhist). Both belonged to what is often called the Kyoto School of Japanese philosophy.

2 Nishida's logic of *basho* and logic of the predicate

Nishida attacked Western paradigms of logic from two directions, epistemological and metaphysical/syntactical. When taking the former approach he generally spoke of his logic as the 'logic of *basho*'. *Basho* is an ordinary Japanese word meaning 'place' that can serve as an equivalent for the technical terms *topos* or 'field'. Nishida understood the philosophical construction of reality to take place within three experiential domains or *basho*: empiricism, idealism and what he called the 'acting-intuiting'. The empiricist speaks from out of the '*basho* of being'; that is, the empiricist judges the world of physical entities (see Empiricism). The idealist, on the other hand, judges the world of psychological entities, showing their role in the construction of our experience of physical entities (see Idealism). This is the realm of consciousness, experience, self and the transcendent ego. Sometimes the Kyoto School refers to this as the '*basho* of relative nothingness', because from the standpoint of the empiricist this psychological world is considered so transparent as to be omitted from the discussion of what is. Relative to empirical reality, it is a place of nothingness.

The third *basho* is supposedly less commonly articulated in the West, although Nishida found suggestive references to it in William James, Bergson, some neo-Kantians and some religiously or aesthetically oriented philosophers. The third *basho* is where knowing and doing are virtually a single act. It occurs when a person responds spontaneously without deliberation to the information being intuited, or when their intake of information is the same act as going out and gathering information. Nishida calls that immediacy the single event of 'acting-intuiting'. This realm is unknown to the categories of both the empiricist's material world and the idealist's inner world. In fact, in its immediacy and oneness, it cannot be analysed in any systematic way at all. Therefore, it can be called the '*basho* of absolute nothingness' (see Buddhist concept of emptiness).

In themselves, these three realms may constitute the contents of kinds of knowing, but Nishida argues there is a formal connection among them: there is a hierarchical 'logic of *basho*'. His analysis of the hierarchy is based in his evaluation of how the *basho* are implicated in each other through the logical analysis of the conditions of judgment. Consider an empirical judgment such as 'The cat is black'. Nishida asks in what place, in what *basho*, one is standing when making that judgment. He argues that for such an empirical judgment to be made, one must have an experience of perceiving that black cat. In other words, one assumes there is a perceiver and a perception that makes the judgment possible. Once we recognize this fact, we realize that judgments within the *basho* of being are logically dependent on experience within the *basho* of relative being, the idealist's world of consciousness and perception. In turn, I may ask about the judgment of 'I see that the cat is black'. That judgment of the *basho* of relative nothingness is itself only possible as a reflection and abstraction out of an experience of the acting-intuiting. So, the *basho* of relative nothingness has as its logical precondition experience within the *basho* of absolute nothingness. The acting-intuiting, being immediate and unitary, is not itself capable of making judgments; yet at the same time, it is the totality out of which reflection can generate the judgments of either idealism or empiricism. For Nishida, it is the universal that becomes concrete through the mediation of the other two *basho*. In this way Nishida's dialectic takes us from the simplest of truths, the empirical judgment, to show regressively its logical basis in the immediacy of absolute nothingness.

Nishida also attacked the Western bias of the ontology of being over nonbeing, of substance over event and of the grammatical subject over the grammatical predicate (see Being; Substance; Events; Syntax). Instead of the Aristotelian model in which the sentential subject is correlated with the metaphysical substance and the predicate with the attribute, Nishida proposed a 'logic of the predicate'. In his logic, the primary unit of reality is the event, and only through abstraction do we separate out of that event the sentential subject or the ontological substance. The predicate is a universal whose function becomes concrete and specified through the sentential subject. In doing so, Nishida has in effect made the substance into an attribute of the event. By changing the metaphysical priority from substance to event and the logical or syntactic priority from subject to predicate, Nishida's Buddhism-inspired logic is grounded in change or process rather than stasis or essence. When merged with the threefold model of the logic of *basho*, Nishida attempted to develop a comprehensive philosophical system

including a metaphysics, a phenomenology, an epistemology and an analysis of judgment. His system was at once intelligible to Western analysis, yet simultaneously continuous with the premodern Japanese intellectual tradition.

3 Tanabe's logic of species

Tanabe criticized Nishida's logic of *basho*, proposing two qualifications or fundamental modifications. First, he questioned the implicit assumption that a totally unmediated experience (the acting-intuiting) is possible. Second, he believed Nishida's system was too simplistic in postulating a direct relation between the individual and the universal without the mediation of culture or society. He believed that without taking account of ethnicity and nationality, we cannot truly understand the dialectical development of thought and value through history. The arena of human activity is the social realm, not the unmediated realm of the acting-intuiting, nor the abstracted realm of intellectual philosophical positions like idealism and empiricism.

Tanabe referred to the level of the transcendent universal (that is, not conceptually particularized) totality as the level of 'genus'. Below it is the level of 'species', and below that the level of the 'individual'. In social terms, the logic of genus applies to humankind as a whole, the source of absolute values for ethics and the spiritual basis for the state. The logic of the individual, on the other hand, applies to epistemological judgment and the basis of personal creativity. Between the two lies the logic of species that applies to particularized ethnic, cultural and national values. Without that mediated middle realm, the actual lived experience of human action cannot be explained.

Tanabe began to develop his logic of species during the 1930s, but as he observed developments in Japan during the militarist period, he began to build into the system a more explicit mechanism of self-criticism. He did not feel free to publish his revisions until after the Second World War, however, at which point he emphasized the self-critical moment in his system by developing the philosophical concept of *range*. He himself translated this term as 'metanoia', a philosophical and personal sense of repentance and conversion that leads one to question the absoluteness of any philosophical system one has devised. Essentially, the dynamic of the logic of species is to be the dialectical mediation between the totalitarianism of the genus and the liberalism of individual. From the genus, the species derives the sense of absolute transcendent values applicable to all mankind; from the individual, the species derives the pragmatic concerns of working toward the particular benefit of one's fellow individuals. The nation, indeed the collective or personal activity of any human endeavour, can therefore never speak with the authority of the absolute; on the other hand, no such endeavour can ever be truly individual either, because all expression is enmeshed in the social. Once again, the realm of human activity is not that of fixed being, but that of Buddhist mediation, interdependence and a dialectical tension that deconstructs any system as soon as it makes it claim to be either absolute or individualistic.

See also: Buddhist philosophy, Japanese; Japanese philosophy; Logic, philosophy of; Nishida Kitarō; Tanabe Hajime

THOMAS P. KASULIS

References and further reading

Carter, R.E. (1989) *The Nothingness Beyond God: An Introduction to the Philosophy of Nishida Kitarō*, New York: Paragon House.(An introductory text emphasizing religious and ethical issues; Chapter 2 is a discussion of the logic of *basho*.)

Dilworth, D. (trans.) (1970) *Fundamental Problems of Philosophy*, Tokyo: Sophia University Press.(Translation of six of Nishida Kitarō's essays related to his logic of *basho*.)

Dilworth, D. (1987) *Last Writings: Nothingness and the Religious Worldview*, Honolulu, HI: University of Hawaii Press.(Translation of Nishida Kitarō's last essays with Dilworth's analysis of Nishida's logic in a comparative context.)

Dilworth, D. and SatōTaira (trans.) (1969) 'The Logic of the Species as Dialectics', *Monumenta Nipponica* 24 (3): 273-88.(Translation of Chapter 1 of Tanabe Hajime's *Shu no ronri no benshōhō(Dialectic of the Logic of Species)*.)

Nishida Kitarō (1911-45) *Nishida Kitarōzenshū(Collected works of Nishida Kitarō)*, ed. Watsuji Tetsurō*et al.*, Tokyo: Iwanami Shoten, 1978, 19 vols.(Standard edition of original works.)

Piovesana, G.K. (1968) *Contemporary Japanese Philosophical Thought*, New York: St. John's University Press.(Good survey of modern Japanese philosophy including a chapter on Nishida Kitaro and a section on Tanabe

Hajime’s logic of species.)

Tanabe Hajime (1910-61) *Tanabe Hajime zenshū(Collected Works of Tanabe Hajime)*, ed. Nishitani Keiji *et al.*, Tokyo: Chikuma Shobō, 1963, 15 vols.(Standard edition of original works with vols 6 and 7 being the locus of many of Tanabe’s central essays on the logic of species.)

Tanabe Hajime (1946) *Zangedōto shite no tetsugaku*, trans. Takeuchi Yoshinori, *Philosophy as Metanoetics*, Berkeley, CA: University of California Press, 1986.(Translation of one of Tanabe Hajime’s major work dealing with metanoia.)

Unno Taitetsu and Heisig, J.W. (eds) (1990) *The Religious Philosophy of Tanabe Hajime*, Berkeley, CA: Asian Humanities Press.(Good collection of essays on Tanabe’s philosophy, several of which refer to the logic of species and its differences from Nishida’s logic of *basho*.)

Wargo, R.J.J. (1972) *The Logic of Basho and the Concept of Nothingness in the Philosophy of Nishida Kitarō*, Ann Arbor, MI: University Microfilms.(A dissertation, this work has never been surpassed as the best Western treatment of Nishida’s logic. It includes a translation of the ‘General Summary’ from Nishida’s *Ippansha no jikakuteki taikei (System of Self-consciousness of the Universal)*.)

Logic in the 17th and 18th centuries

Logic in the seventeenth century was characterized by attempts to reconcile older viewpoints, such as those of Ramus and Melanchthon, and by criticism of the nature and scope of traditional logic. F. Bacon indicated induction, rather than deduction, as the object of logic, thus opening the way for a logic of the empirical sciences. Descartes proposed to replace the complicated precepts of old logic by simple rules of method. However, even the authors of the Port-Royal Logic, who were influenced by Descartes, could not follow him all the way and continued to teach traditional doctrines, albeit with a new attention to the doctrine of ideas. Other logicians, following Locke, tried to modernize logic by concentrating on an analysis of human cognitive faculties, of the idea-word relation and of other than certain knowledge, thus broadening the scope of logic so as to account for probability. Another suggestion for the improvement of logic came from those who thought that logic should assume mathematics as an example either for its axiomatic-deductive method or for the inventive techniques of algebra. The last of these suggestions prompted research in the area of logical calculi. But this kind of research benefited from the doctrines devised by non-mathematically oriented authors who thus provided the logical framework in which algebraic techniques would be tried. This general background accounts not only for the exceptional logic of Leibniz, but also for some logical calculi worked out in the eighteenth century.

1 Seventeenth-century eclecticism

Seventeenth-century logic was conditioned by the debate of the previous two centuries for and against Aristotelianism and scholasticism, both of which had a quasi-autonomous life until well into the eighteenth century, so that differences of opinion persisted.

However, followers of Aristotle's defender Philip Melanchthon and of the anti-Aristotelian Petrus Ramus merged to form a Philippo-Ramist or Semi-Ramist school (see Logic, Renaissance §5). They restored the parts of Aristotelian logic left out by Ramus (conversion, the square of opposition, categories, demonstration, fallacies), but also attended to Ramist doctrines (for example, definition and division). The Philippo-Ramist B. Keckermann published a*Systema Logicum* (1600) in which he assigned to logic the aim of being systematic, that is, of being a first science directing the human mind; hence the additional name of Systematics given to his followers, such as J.H. Alsted, author of *Clavis Artis Lullianae (A Key to Lull's Art)* (1609) and *Logicae Systema Harmonicum* (1614), and R. Sanderson, who wrote a *Logicae Artis Compendium* (1618).

Eclecticism also characterized followers of the sixteenth-century logician J. Zabarella, according to whom Aristotle had provided an analytic method of discovery ('*regressus*') for mathematical demonstrations. Many seventeenth-century Zabarellists endorsed Ramist and scholastic doctrines, including F. Burgersdijck, whose *Institutionum logicarum libri duo (Two Books of Logical Institutions, 1626)* became the standard handbook in the Netherlands.

Another eclectic, J. Jungius, was one of the outstanding logicians of the century. He was acquainted with Philippo-Ramist dialectic and with the then recent scholasticism of Fonseca and Suarez. He also had Systematic and Zabarellist teachers and admired the *logistica* of Viète. In his *Logica Hamburgensis* (1635) he dealt with a logic of relations, non-syllogistic inferences, contraposition and special syllogisms.

2 Criticism of the syllogism

However, some of the major philosophers of the seventeenth century persisted in criticizing traditional logic, in particular the syllogism. In his *Instauratio magna* (1620) Francis Bacon, while acknowledging the cogency of the syllogism, observed that it cannot guarantee that the ideas occurring in it are abstracted from reality and defined correctly. Since induction presides over the abstraction and definition of ideas, he indicated induction, rather than deduction, as the object of logical research, thus opening the way for a logic of the empirical sciences.

Descartes repeated the sceptical argument that the syllogism, apart from being sterile, begs the question and has no more claim to certainty than non-inferential truths. As is clear from his *Regulae* (1628), for him the problem of inference was not how to infer (since we do this naturally), but how to infer with certainty. Descartes' solution to this problem rests on an analysis of our scientific knowledge which singles out two cognitive operations: intuition and deduction. Intuition is the basic operation because it makes us certain of simple propositions, whereas

deduction is just a chain of successive intuitions. Having found this solution, Descartes saw little use for a complicated logic: indeed, he considered it detrimental to the natural light of reason. As a substitute he listed, in the *Discours de la méthode* (1637), some rules of method meant as mere guidelines for the cognitive operations whose certainty he claimed to have explained.

3 Logic after Descartes

So limited a scope for logic was by no means universally accepted. For example, H. Fabri [Mosnerius] in the first tome of his *Philosophia* (1646) joined late medieval doctrines to mathematics and devised a combinatory calculus of 576 syllogistic moods in all figures. He also introduced truth tables based on three truth-values: truth, falsity and partial falsity. Later in the century independent research was pursued at a higher level in the *Logica demonstrativa* (1697) of G. Saccheri. He emphasized the demonstrative role of the so-called '*consequentia mirabilis*' showing that 'If not *p* then *p*' entails '*p*'.

But even scholars who accepted Descartes' rules of method as the core of logic had to account for traditional doctrines. This is clear from the *Logica vetus et nova (Old and New Logic)* (1654) of Clauberg; and even more from the *Logica restituta* (1662) of the independent Cartesian A. Geulincx, who revived medieval doctrines, related disjunction, conjunction and negation by means of their truth-conditions and devised a 'logical cube' whose faces represented logical axioms and forms of argumentation.

Old and new doctrines were also adopted in *La Logique ou l'art de penser* (1662), known as 'The Port-Royal Logic', by A. Arnauld and P. Nicole. The authors referred to Descartes' views (the 1664 edition mentions the manuscript of Descartes' *Regulae* (1628), and showed appreciation of Pascal's theory of definition, probability and method (they included a special section on method, imitated by many subsequent logics), and of Augustine's theory of language, from which their attentiveness to the idea-word relation was derived. As for traditional instruction, their syllogistic was more accurate than might be expected of a Cartesian logic. For they devised a combinatory calculus of possible moods with rules to choose the valid ones, and correctly warned against confusing the fourth figure and the first figure with transposed premises. None the less, the Port-Royal logicians agreed with Descartes that deduction has a derivative importance as compared with judgment. And given that they defined judgment as a comparison of ideas, they paid great attention to the ideas themselves. The Port-Royal doctrine of ideas revolves around the notions of comprehension (later, 'intension') and extension. The comprehension of an idea consists of the attributes it comprises in itself that cannot be removed from it without destroying it; the extension of an idea consists of the subjects with which the idea agrees. Comprehension and extension stand in inverse relation: the greater the comprehension of an idea, the smaller its extension, and vice versa. This relation regulated all other relations between ideas and operations (for example, abstraction and composition) performable on them.

The influence of Port-Royal doctrines upon French logic was evident right up to the end of the eighteenth century, for example from the *Logique* (1780) and the *Langue des calculs (Language of Calculi, 1798)* of É.B. de Condillac. In England the Port-Royal Logic was acknowledged even by scholars of a declared Aristotelian disposition, such as J. Wallis, who, in his *Institutio logicae* (1687), rejected Ramist syllogisms with singular terms and classified syllogistic figures competently. H. Aldrich, too, though mostly concerned with his valuable presentation of syllogistic, mentioned Port-Royal doctrines when discussing methodology and the history of logic in his *Artis Logicae Compendium* (1691).

4 The emergence of a logic of cognitive faculties

Towards the end of the seventeenth century, logic began to centre increasingly on an analysis of cognitive faculties. Such an analysis, introduced by Descartes, was carried further in the *Essay concerning Human Understanding* (1690) of John Locke and was the stepping-stone to what has been called 'facultative logic'. In this kind of logic the content of ideas is as important as their form, and equally important are the signs designating ideas. This not only explains why Locke equated logic with semiotics, but also gives an insight into the origins of the importance of language for later philosophy and logic.

The logic of cognitive faculties was promoted by the success of Locke's philosophy which, especially in Germany (for example, in the 1752 *Vernunftlehre (Doctrine of Reason)* of G.F. Meier), was often preferred to Descartes'

because of its avoidance of innate ideas and its applicability to other than certain knowledge (which later brought the study of probability within its scope). The influence of Locke's philosophy is evident in England in both the *Logick* (1725) and the *Improvement of the Mind* (1741) of I. Watts, who proposed a mixture of Locke with Port-Royal, as well as in the *Elements of Logick* (1748) of W. Duncan and the *Logicae compendium* (1759) of F. Hutcheson. Locke also influenced the *Système de réflexions* (1712) of the Swiss J.P. de Crousaz.

5 Logic and mathematics

At almost the same time as the birth of the logic of cognitive faculties, another image of the discipline began to emerge. It was born of the conviction that logic compared unfavourably with mathematics.

Some ascribed the superiority of mathematics to its axiomatic-deductive method. Indeed, Descartes' rules of method were interpreted as recommendations to start with a few simple and acknowledged notions (axioms) and thence to proceed to unknown ones (theorems). The supporters of such an interpretation maintained that the axiomatic-deductive method should replace logic and be the model for the right conduct of understanding. A prime example of this view is the *Medicina mentis* (1687) of von Tschirnhaus, who indicated the mathematical method as a therapy in all fields of knowledge and practical life.

Some attributed the superiority of mathematics to the problem-solving and inventive techniques of algebra. The search for equations relating unknown elements with given elements, exemplified in Descartes' *Géométrie* (1637), was considered the true Cartesian logic and was absorbed into the tradition which viewed mathematics as a universal science of invention (*mathesis universalis*).

This assessment of the merits of algebra outlived the imitation of the mathematical method: the latter was either reduced to a mere synthetic order of exposition, or restricted to mathematics because its wider use was held responsible for the degeneration of Cartesianism into Spinozism. Algebraic tools, on the other hand, were fitted for use in the construction of logical calculi. Despite this, the calculi which were actually or tentatively constructed at the time are not comparable to the later algebra of logic, mainly because they were still bound to the quantitative interpretation of algebra.

6 Leibniz and the emergence of logical calculi

Leibniz's logical calculus is connected with his project of an '*ars characteristica combinatoria*', whose accomplishment required (1) an alphabet of thoughts and appropriate symbols for it, (2) a rational grammar common to all languages, (3) an encyclopedia of the sciences and (4) a universal science by which the contents of all sciences would be ordered systematically and combined so as to discover new truths. This last requirement, in its turn, demanded a logical calculus.

Leibniz could not achieve his entire project, but none the less he wrote several incomplete versions of a logical calculus in papers dating from 1679 to the early 1690s (see Leibniz 1966). Common to all versions - which were related to Leibniz's many logical interests (syllogistic, logic of relations, modality and so on), as well as to his monadology - is the idea that a '*calculus ratiocinator*' is a mechanical procedure for drawing conclusions. It is based on: a universal mathematics similar to algebra; symbols standing for concepts or for propositions; and substitution rules governed by the principle that what is concluded with certain variable letters can be concluded equally with other letters satisfying the same conditions.

For all its exceptionality and complexity Leibniz's calculus represents a development of logic that can also be found in others' work which came soon after. This did not depend on Leibniz's direct influence because his calculus remained unpublished until the nineteenth century. The likely explanation is the existence of certain favourable conditions: the reputation of algebra as a universal means of discovery (see §5); and the existence of doctrines of ideas such as that of Port-Royal and the emerging logic of cognitive faculties (see §§3, 4 above). In this respect Cartesian logic and the logic of cognitive faculties were only apparently disconnected from logical calculi: the former provided the latter with a logical field of application. Indeed Locke, who belittled syllogistic, hoped for an algebra of logic. Admittedly, he did not construct a logical calculus and nor did Hobbes, despite his statement in *De Corpore* (1655) that ratiocination is computation. But an early attempt made by Jakob Bernoulli to establish a *Parallelismus ratiocinii logici et algebraici* (1685), while ending in admitted failure, was demonstrably based on a comparison of algebra with the Port-Royal doctrine of ideas.

7 Logical calculi in the eighteenth century

Between the 1740s and the 1760s progress both in the study of ideas and in algebra prompted a number of logical calculi by German-speaking authors.

The theory of ideas had been much developed in Germany due to the influence of A. Rüdiger. In his *De sensu veri et falsi* (1709) he maintained that the syllogism can be a means of discovery provided that, given a premise, we connect one of its terms with a new concept that stands in one of a set of well-defined idea-relations with the other term. For example, by connecting the predicate-term of a universal affirmative premise with any concept subordinated to the subject-term, we establish enthymematically the conclusion of a 'synthetic syllogism'. In general, Rüdiger's point was that if the syllogism can be used to discover truths, then logic does not have to be given a mathematical formulation in order for it to be inventive.

Taking the contrary viewpoint, Christian Wolff supported a mathematical view of logic maintaining that all mathematical demonstrations consist of chains of syllogisms in the first figure. Therefore in his *Vernünfftige Gedanken von den Kräfften des menschlichen Verstandes (Rational Thoughts on the Powers of the Human Understanding)* (1713) and his *Philosophia rationalis* (1728) he considered it futile to advocate the mathematical method as a substitute for logic or to charge traditional logic with being sterile and nugatory.

In the 1730s a number of authors, strengthened by the confidence in the value of logic characterizing the then dominant Wolffism, resumed logical work even along lines unforeseen by Wolff. In particular, J.P. Reusch in his *Systema logicum* (1734) gave a description of judgments and inferences as based on a set of idea-relations (reminiscent of Rüdiger) and on a principle of substitutivity for ideas.

But Reusch did not take the further step of building a calculus of idea-relations. This step was taken by J.A. Segner in his *Specimen logicae universaliter demonstratae (An Example of a Universally Demonstrated Logic, 1740)*. By using the notion of containment in intension and by an accurate employment of negative terms, he classified, assigned symbols to, and subjected to a set of rules the five idea-relations later known as the Gergonne relations. This produced a rare and beautiful example of a calculus applicable to syllogistic and to some non-syllogistic inferences.

An extensional perspective was chosen by G. Ploucquet in his *Fundamenta philosophiae speculativae* (1759). In his syllogistic calculus all concepts, including predicates, are quantified and there is a single general rule stating that concepts must retain in the conclusion of an argument the quantity they have in the premises.

Many topics were treated by J.H. Lambert in his *Neues Organon* (1764), ranging from probabilistic inference to symbolic knowledge to a theory of truth and error. In it he also represented the A, E, I, O propositions and syllogisms extensionally by means of straight lines. Such diagrams were independent of but served the same purpose as the circles used by Euler in his *Lettres à une Princesse d'Allemagne sur divers sujets de physique et de philosophie (Letters to a German Princess on Various Subjects of Physics and Philosophy, 1760-2)*. Circles, lines and other figures had also been used by J.C. Sturm (*Universalia Euclidea*, 1661), J.C. Lange (*Nucleus logicae Weisianae*, 1712), Ploucquet and the then unpublished Leibniz. But Lambert's most interesting contribution to logic is a calculus he worked out in letters and in some of his *Logische und philosophische Abhandlungen (Logical and Philosophical Essays, 1782- 7)* which were published after his death. He based it on concepts, their mutual interrelationships, and on the similarities between logical and algebraic operations. This calculus - which allowed for a tentative logic of relations and was capable of expressing composite propositions - was intensional. This feature was criticized by the Ploucquetian G.J. Holland in a letter (1765) to Lambert in which he sketched his own, extensional calculus. (For a comparison of Lambert's and Ploucquet's calculi see Bök (1766).)

Interest in logical calculi faded in the later part of the eighteenth century due, in all likelihood, to the influence of Kant's ideas on logic. In his *Nova dilucidatio (New Exposition, 1755)* Kant had maintained that Leibniz's *combinatoria* was far from being a genuine '*ars inveniendi*' but apparently did not deny that an *ars inveniendi* based on an algebraic calculus of concepts could be achieved. However, in his *Critique of Pure Reason* (1781) and in his *Logik* (1800), he developed a conception according to which an application of algebra to non-mathematical concepts would have no inventive power. Algebra is synthetic a priori, hence inventive and absolutely certain only on condition that its symbols stand for mathematical concepts. The latter are the only concepts whose objects can

be exhibited a priori in intuition, which gives algebraic proofs heuristic advantages and freedom from error. Kant's separation of logic from mathematics was counterbalanced by his rejection of the conception of a logic of the cognitive faculties that he had originally favoured. His mature view was that pure logic, as opposed to transcendental logic, is a formal science - a view not lightly to be dismissed.

See also: Logic machines and diagrams; Logical and mathematical terms, glossary of; Port-Royal; Universal language

MIRELLA CAPOZZI

References and further reading

Aldrich, H. (1691) *Artis Logicae Compendium*, Oxford.(Aldrich's clear presentation of logic made this book popular until the middle of the nineteenth century.)

Alsted, J.H. (1609) *Clavis Artis Lullianae (A Key to Lull's Art)*, Strasbourg; repr. Hildesheim: Olms, 1983.(As well as being a follower of Keckermann, Alsted was interested in Lull's art.)

Alsted, J.H. (1614) *Logicae Systema Harmonicum*, Herborn.(As with many Systematics, Alsted also showed some sympathy towards the doctrines of Ramus.)

Arnauld, A. and Nicole, P. (1662) *La Logique ou l'art de penser*; trans. and ed. J.V. Buroker, *Logic or the Art of Thinking*, Cambridge: Cambridge University Press, 1996.(One of the most influential books in the history of logic.)

Arndt, H.W. (1971) *Methodo scientifica pertractatum: Mos geometricus und Kalkülbegriff in der philosophischen Theorienbildung des 17. und 18. Jahrhunderts*, Berlin and New York: De Gruyter.(A good guide to appreciate the importance of mathematics for philosophy and logic in the seventeenth and eighteenth centuries.)

Ashworth, E.J. (1967) 'Joachim Jungius (1587-1657) and the Logic of Relations', *Archiv für Geschichte der Philosophie*, 49: 72-85.(A brief outline of Jungius' views on relations.)

Ashworth, E.J. (1974) *Language and Logic in the Post-Medieval Period*, Dordrecht and Boston, MA: Reidel.(A classic guide to logic 1500-1650.)

Auroux, S. (1993) *La logique des idées*, Montreal, Que., and Paris: Bellarmin-Vrin.(A discussion of the Port-Royal Logic. It also contains a good bibliography of logical works in French 1700-1800.)

Bacon, F. (1620) *Instauratio magna (Distributio operis)*, repr. in *The Works of Francis Bacon*, ed. J. Spedding, R.L. Ellis and D.D. Heath, London: Longman, 7 vols, 1857-74; repr. Stuttgart: Frommann (Holzboog), 1962. (This work was a source not only for the logic of empirical sciences but also for later logic textbooks' theory of errors.)

Bernoulli, J. (1685) *Parallelismus ratiocinii logici et algebraici*, Basilea; trans. T. Boswell, 'The Brothers James and John Bernoulli on the Parallelism Between Logic and Algebra', *History and Philosophy of Logic* 11: 211-24, 1990.(An early attempt at a logical calculus. This work, which was demonstrably written by *Jakob* Bernoulli, was an academic dissertation in which Jakob Bernoulli acted as president and his younger brother Johann as respondent, hence the usual but incorrect attribution to 'the brothers Bernoulli'.)

Blanché, R. (1970) *La logique et son histoire, d'Aristote à Russell*, Paris: Colin.(An introductory book on the history of logic.)

Bochenski, I.M. (1956) *History of Formal Logic*, trans. and ed. I. Thomas, Notre Dame, IN: University of Notre Dame Press, 1961.(An introductory book on the history of logic with several excerpts from original texts.)

Bök, A.F. (ed.) (1766) *Sammlung der Schriften, welche den logischen Calcul Herrn Prof. Ploucquets betreffen, mit neuen Zusätsen (A Collection of Writings Concerning Professor Ploucquet's Calculus, With New Supplements)*; repr. and ed. A. Menne, Stuttgart and Bad Cannstatt: Frommann (Holzboog), 1970.(Apart from some texts by Ploucquet giving an improved version of his calculus, this collection includes reprints of texts concerning the debate raised by Ploucquet's work and by its comparison with the work of Lambert; also includes essays by Ploucquet and Lambert stressing the differences between their respective work.)

Buickerood, J.G. (1985) 'The Natural History of the Understanding: Locke and the Rise of Facultative Logic in the Eighteenth Century', *History and Philosophy of Logic* 6: 157-90.(This work introduced the term 'facultative logic' in the history of logic.)

Burgersdijck, F. (1626) *Institutionum logicarum libri duo (Two Books of Logical Institutions)*, Leyden.(This textbook became the standard logical handbook in the Netherlands after 1635.)

Capozzi, M. (1987) 'Kant on Logic, Language and Thought', in D. Buzzetti and M. Ferriani (eds) *Speculative Grammar, Universal Grammar and Philosophical Analysis of Language*, Amsterdam and Philadelphia, PA: Benjamin, 97-147.(Gives an insight in the evolution of Kant's conception of logic.)

Capozzi, M. (1994) 'Algebra e logica in Jakob Bernoulli', in *Atti del Congresso 'Logica e filosofia della scienza:*

problemi e prospettive', Lucca 7-10 gennaio 1993, Pisa: ETS, 55-74.(Considers one of the first attempts to establish the relations between algebra and logic.)

Clauberg, J. (1654) *Logica vetus et nova (Old and New Logic)*, repr. in *Opera omnia philosophica*, ed. J.T. Schalbruch, Amsterdam, 2 vols, 1691; repr. Hildesheim: Olms, 1968.(The work of a German Cartesian who tries to reconcile old and new doctrines for the benefit of students.)

Condillac, É.B. de (1780) *Logique ou les premiers developpements de l'art de penser*, Paris; repr. in *Oeuvres philosophiques*, vol. 2, ed. G. Le Roy, Paris: Presses Universitaires de France, 1947-51; repr. and trans. W.R. Albury, *La Logique (Logic)*, New York: Abaris Books, 1979.(De Condillac was among French logicians influenced by Port-Royal doctrines.)

Condillac, É.B. de (1798) *La langue des calculs (The Language of Calculi)*, ed. S. Auroux and A.M. Chouillet, Villeneuve d'Ascq: Presses Universitaires de Lille.(Published posthumously.)

Crousaz, J.P. de (1712) *Système de réflexions qui peuvent contribuer à la netteté et à l'étendue de nos connaissances, ou Nouvel essai de logique (System of Reflections which Can Contribute to the Neatness and to the Extension of our Knowledge, or New Essay of Logic)*, Amsterdam, 2 vols.(An example of Locke's influence on continental logicians.)

Daston, L. (1988) *Classical Probability in the Enlightenment*, Princeton, NJ: Princeton University Press.(An important contribution to the subject.)

Descartes, R. (1628) *Regulae ad directionem ingenii*, 1701; trans. in *The Philosophical Writings of Descartes*, vol. 1, ed. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1984-5.

Descartes, R. (1637a) *Discours de la méthode*, trans. 'Discourse on the Method', in *The Philosophical Writings of Descartes*, vol. 1, ed. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1984-5.(Descartes' contention is that in the search for truth one need follow only four simple precepts, disregarding the great number of those of traditional logic.)

Descartes, R. (1637b) *La Géométrie*; repr. and trans. in *The Geometry of René Descartes*, ed. D.E. Smith and M.L. Latham, LaSalle, IL: Open Court, 1925; repr. New York: Dover, 1954.

Duncan, W. (1748) *Elements of Logick*, London; repr. Hildesheim: Olms, 1968.(States clearly that the object of logic is the study of the faculties of human understanding and of cognitive processes.)

Euler, L. (1760-2) *Lettres à une Princesse d'Allemagne sur divers sujets de physique et de philosophie (Letters to a German Princess on Various Subjects of Physics and Philosophy)*, St. Petersburg, 1768- 72; repr. in *Leonhardi Euleri opera omnia*, Leipzig: Teubner, 1911.(In a few pages the famous mathematician Euler gives a clear representation, by means of circles, of the basic categorical propositions entering simple syllogisms.)

Fabri, H. [P. Mosnerius] (1646) *Philosophiae tomus primus… cuncta excerpta ex praelectionibus R.P. Honorati Fabry (First Tome of Philosophy… together with excerpts from R.P. Honoré Fabri's lectures)*, Lyon.(Fabri's calculus of 576 syllogistic moods was discussed by Leibniz in his *De arte characteristica combinatoria*.)

Gardner, M. (1983) *Logic Machines and Diagrams*, 2nd edn with foreword by D. Michie, Brighton: Harvester. (An outline of the development of mechanically induced inference.)

Gaukroger, S. (1989) *Cartesian Logic: An Essay on Descartes's Conception of Inference*, Oxford: Clarendon Press.(A clear and well-argued introduction to the topic.)

Geulincx, A. (1662) *Logica fundamentis suis, a quibus hactenus collapsa fuerat,… restituta (Logic Restored to its Foundations… From Which it Had Fallen)*; repr. in *Opera philosophica*, ed. J.P.N. Land, Den Haag, 3 vols, 1891-3; repr. and ed. H.J. de Vleeschauwer, Stuttgart and Bad Cannstatt: Frommann (Holzboog), 5 vols, 1968. (It is noteworthy that, contrary to the example of the Port-Royal Logic, published in the same year, Geulincx denied that method is a part of logic.)

Hacking, I. (1975) *Why Does Language Matter to Philosophy?*, Cambridge: Cambridge University Press.(The origins of the importance of language for philosophy are traced back to the doctrines of ideas of Locke and Port-Royal.)

Hamilton, W. (1860-9) *Lectures on Metaphysics and Logic*, ed. H.L. Mansel and J. Veitch, Edinburgh and London: Blackwood, vol. 2, appendix.(Hamilton gives an account of the logic of many seventeenth- and eighteenth-century authors, with special attention to early attempts at a quantification of predicates.)

Hobbes, T. (1655) *Elementa philosophiae: de Corpore*; trans. in *The English Works of Thomas Hobbes*, vol. 1, ed. Sir W. Molesworth, London: John Bohn, 11 vols, 1839; repr. Aalen: Scientia, 1961.(Maintains that reasoning is the same as computation and in particular is the same as adding and subtracting.)

Holland, G.J. (1765) 'Letter to Lambert (9 April 1765)', in J.H. Lambert, *Philosophische Schriften*, vol. 9.1, Hildesheim: Olms, 1968, 17-20.(In this letter Holland is critical of Lambert's intensional calculus and sketches his own, extensional one.)

Hoorman, C.F.A., Jr (1976) 'A Further Examination of Saccheri's Use of ''Consequentia Mirabilis''', *Notre*

Dame Journal of Formal Logic 17: 239-47.(This paper takes into account modern literature on the subject.)
Howell, W.S. (1956) *Logic and Rhetoric in England, 1500-1700*, Princeton, NJ: Princeton University Press.
Howell, W.S. (1971) *Eighteenth Century British Logic and Rhetoric*, Princeton, NJ: Princeton University Press. (Both Howell's books are somehow outdated, but they are still the only comprehensive introduction to logic in England 1500-1700.)
Hutcheson, F. (1759) *Logicae compendium*, Glasgow.(Logic is here proposed as the study of cognitive faculties, but the author also gives a kind of axiomatic presentation of syllogisms.)
Jungius, J. (1635-8) *Logica Hamburgensis (Hamburg Logic)*, Hamburg, 4 vols; 2nd edn repr. Hamburg, 1681; ed. R.W. Meyer, Hamburg: Augustin, 1957.(Leibniz held Jungius in the highest esteem as a logician, especially for what he had left unpublished. But Jungius' published work is anyhow commendable for a lucid elaboration of results taken from a rich tradition on scientific proof.)
Kant, I. (1755) *Principiorum primorum cognitionis metaphysicae nova dilucidatio (A New Exposition of the First Principles of Metaphysical Knowledge)*; repr. in *Gesammelte Schriften*, Berlin: De Gruyter, 1902-10, vol. 1, 389; trans. F.E. England, in *Kant's Conception of God: A Critical Exposition of its Metaphysical Development Together with a Translation of the 'Nova Dilucidatio'*, London: Allen & Unwin, 1929.
Kant, I. (1800) *Logic*, ed. G.B. Jäsche, repr., together with lecture notes taken by Kant's logic students *c*.1770-90, in *Lectures on Logic*, trans. J.M. Young, Cambridge: Cambridge University Press, 1992. (The only logical work published under Kant's name in his lifetime.)
Keckermann, B. (1600) *Systema Logicum*, Hannover; repr. in *Opera omnia*, Genf, 2 vols, 1614; repr. Hildesheim: Olms, forthcoming.(Discusses what makes a system out of a body of doctrines. After Keckermann the term 'system' was commonly used to designate the scientific construction of logic.)
Kneale, W. and Kneale, M. (1962) *The Development of Logic*, Oxford: Clarendon Press, 1984.(The most comprehensive introduction to the history of logic.)
Lambert, J.H. (1764) *Neues Organon oder Gedanken über die Erforschung und Bezeichnung des Wahren und dessen Unterscheidung vom Irrthum und Schein (New Organon or Thoughts on the Investigation and Designation of the True and its Distinction from Error and Appearance)*, Leipzig, 2 vols; repr. in *Philosophische Schriften*, vols 1-2, ed. H.W. Arndt, Hildesheim: Olms, 1965.(This famous book is divided into four parts: Dianology (or logic proper), Alethiology, Semiotics and Phenomenology.)
Lambert, J.H. (1782-7) *Logische und philosophische Abhandlungen (Logical and Philosophical Essays)*, Berlin, 2 vols; repr. in *Philosophische Schriften*, vols 6-7, Hildesheim: Olms, 1967-9.(These essays were published after Lambert's death.)
Lange, J.C. (1712) *Nucleus logicae Weisianae… variis addimentis novis… auctus et illustratus (The Nucleus of Weise's Logic… Enlarged and Illustrated by Various New Additions)*, Giessen.(Lange refers to the logic of Christian Weise (1681), but improves on it by using circles to represent not only the usual categorical propositions and syllogisms but also disjunctive, copulative and comparative syllogisms.)
Leibniz, G.W. (1966) *Logical Papers: A Selection*, trans. and ed. G.H.R. Parkinson, Oxford: Clarendon Press. (Collected papers on logic from 1679 to the early 1690s.)
Lewis, C.I. (1918) *A Survey of Symbolic Logic*, Berkeley, CA: University of California Press; revised edn, New York: Dover, 1960.(An early attempt to give an outline of the development of logic.)
Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1975. (A philosophical classic which had a great impact on the general concept of logic in Europe in the eighteenth century.)
Meier, G.F. (1752) *Vernunftlehre (Doctrine of Reason)*, Halle; repr. Hildesheim: Olms, forthcoming; also *Auszug aus der Vernunftlehre (Excerpts from the Doctrine of Reason)*, Halle; repr. in I. Kant, *Gesammelte Schriften*, vol. 16, Berlin: De Gruyter, 1914.(Meier was the first German professor to use Locke's *Essay* as a textbook. He however advocated a form of logic strongly reminiscent of Wolff. Kant used the shorter version as a handbook for his lectures on logic.)
Nuchelmans, G. (1983) *Judgement and Proposition: from Descartes to Kant*, Amsterdam: North Holland.(A very important contribution to the subject.)
Nuchelmans, G. (1991) *Dilemmatic Arguments: Towards a History of Their Logic and Rhetoric*, Amsterdam: North Holland.(Discusses Geulincx in particular.)
Ong, W. (1958) *Ramus, Method, and the Decay of Dialogue*, Cambridge, MA, and London: Harvard University Press.(A general history of Ramism and of sixteenth century logic.)
Ploucquet, G. (1759) *Fundamenta philosophiae speculativae (Foundations of Speculative Philosophy)*, Tübingen. (A first formulation of Ploucquet's calculus, which he improved in essays included in the collection edited by Bök (1766).)

Reusch, J.P. (1734) *Systema logicum*, Jena; repr. in C. Wolff, *Gesammelte Werke*, §3, vol. 26, Hildesheim: Olms, 1990.(In the second edition of 1741 Reusch refers advanced readers to the work of Segner.)

Risse, W. (1964-70) *Die Logik der Neuzeit*, Stuttgart and Bad Cannstatt: Frommann (Holzboog), 2 vols.(The first comprehensive account of logic 1500-1780. It contains extensive quotations from almost all the authors it cites.)

Risse, W. (1965) *Bibliographia Logica: Verzeichnis der Druckschriften zur Logik mit Angabe ihre Fundorte,* Hildesheim: Olms, vol. 1, 1472-1800.(An invaluable bibliographical source.)

Roncaglia, G. (1996) *Palaestra rationis. Discussioni su natura della copula e modalità nella filosofia 'scolastica' tedesca del XVII secolo,* Florence: Olschki.(Includes a detailed exposition and evaluation of modal theories discussed in German universities 1600-70.)

Rossi, P. (1968) *Francis Bacon: from Magic to Science*, trans. S. Rabinovitch, London: Routledge. (Still to be recommended because it connects Bacon with the logical and methodological trends of his time.)

Rüdiger, A. (1709) *De sensu veri et falsi libri IV (Four Books about the Sense of True and False)*, Halle; enlarged 2nd edn, 1722; repr. Hildesheim: Olms, forthcoming.(Apart from synthetic syllogistic, the author maintains that mathematics is based on sensation and provides an extensive treatment of non-mathematical probability.)

Saccheri, G. (1697) *Logica demonstrativa, theologicis, philosophicis et mathematicis disciplinis accomodata (Demonstrative Logic Suitable for the Theological, Philosophical And Mathematical Disciplines)*, Torino; repr. and ed. W. Risse, Hildesheim: Olms, 1980.(Includes a noteworthy treatment of syllogistic and a careful study of consequences, which include the 'consequentia mirabilis'.)

Sanderson, R. (1618) *Logicae Artis Compendium*, ed. with an intro. by E.J. Ashworth, Bologna: CLUEB, 1985. (An important book in the Oxford curriculum. The editor's introduction to the 1985 edition is of particular value also as a general introduction to the logic of the sixteenth century.)

Schepers, H. (1959) *Andreas Rüdigers Methodologie und ihre Voraussetzungen*, Kant-Studien, Ergänzungsheft 78, Cologne: Kölner Universitäts-Verlag.(An exemplary account of Rüdiger's philosophy and logic.)

Scholz, H. (1931) *Concise History of Logic*, trans. K.F. Leideckeras, New York: Philosophical Library, 1961.(An early attempt to sketch the development of mathematical logic.)

Segner, J.A. (1740) *Specimen logicae universaliter demonstratae (An Example of a Universally Demonstrated Logic)*, ed. with an intro. by M. Capozzi, Bologna: CLUEB, 1990.(Segner received little attention from his contemporaries, but is praised by historians of logic for his ability in applying algebra to logic. The editor's introduction to the 1990 edition gives an overall view of logic in Germany in the first half of the eighteenth century.)

Sturm, J.C. (1661) *Universalia Euclidea*, Den Haag.(The author works out twelve new syllogistic moods.)

Styazhkin, N.I. (1964) *History of Mathematical Logic from Leibniz to Peano*, trans., Cambridge, MA, and London: MIT Press, 1969.(An outline of the development of mathematical logic.)

Tschirnhaus, E.W. von (1687) *Medicina mentis, sive tentamen genuinae logicae, in qua disseritur de methodo detegendi incognitas veritates (Medicine for the Mind, or an Essay with Genuine Logic, Investigating the Method for the Discovery of Unknown Truths)*, Amsterdam: Magnus & Rieuwerts; 2nd, revised edn, *Medicina mentis, sive artis inveniendi praecepta generalia (The Medicine of the Mind, or General Precepts of the Art of Discovery)*, Leipzig: Fritsch, 1695; repr. with an intro. by W. Risse, Hildesheim: Olms, 1964.(The author's idea of logic as a method for the discovery of truth and as a therapy against errors is supported by his conviction that logic should follow the example of mathematics but does not produce anything similar to mathematical logic.)

Venn, J. (1881) *Symbolic Logic*, London: Macmillan, 2nd edn, 1894; repr. New York: Chelsea, 1979. (A book that has created an interest in the history of logic of the seventeenth and eighteenth centuries.)

Wallis, J. (1687) *Institutio logicae,* Oxford.(Wallis was a famous mathematician and one of the most significant English logicians of his time.)

Watts, I. (1725) *Logick, or The Right Use of Reason in the Enquiry after Truth*, London.(A very popular book which was reprinted well into the nineteenth century.)

Watts, I. (1741) *The Improvement of the Mind, or A Supplement to the Art of Logick*, London.(The large number of reprints of this book was due to the author's ability to put Locke's theory of knowledge in the form of a logical handbook inspired by the Port-Royal Logic.)

Wolff, C. von (1713) *Vernünfftige Gedanken von den Kräfften des menschlichen Verstandes*, Halle; repr. in *Gesammelte Werke*, §1, vol. 1, ed. H.W. Arndt, Hildesheim: Olms, 1978; trans. L. Hawes, W. Clarke and R. Collins, *Logic, or Rational Thoughts on the Powers of the Human Understanding, with Their Use and Application in the Knowledge and Search of Truth*, London, 1770.(A little book which gives a noteworthy analysis of the proof of a mathematical theorem in terms of a series of syllogisms in the first figure.)

Wolff, C. von (1728) *Philosophia rationalis sive Logica (Rational Philosophy, or Logic)*, Frankfurt and Leipzig, 2nd edn, 1740; repr. in *Gesammelte Werke*, §2, vols 1/2-3, ed. J. École, Hildesheim: Olms, 1983; partial trans. R.J. Blackwell, *Preliminary Discourse on Philosophy in General*, Indianapolis, IN: Bobbs-Merrill, 1963.(An elaborate exposition of logic that became the standard reference book, even of oppositors, not only in Germany but throughout Europe.)
Yolton, J. (1986) 'Schoolmen, Logic and Philosophy', in T.H. Aston (ed.) *The History of the University of Oxford*, ed. L.S. Sutherland and L.G. Mitchell, Oxford: Clarendon Press, 565-91.(Especially concerns Locke and his influence on English logicians such as Watts and Duncan.)

Logic in the 19th century

The nineteenth century was one of the most active periods for logic in Western philosophy. It is regarded foremost as being the first time logic became 'symbolic' and 'mathematical'.

There was tremendous diversity and conflict in logic in this period - indeed there were substantial debates over whether there even is a subject called 'formal logic' and over what the most basic logical forms were. There was an explicit discussion in the early part of the century of whether logic should be extensional or intensional, opinion eventually settling on a purely extensional conception. By the end of the century, many of these debates had been resolved or had withered away, and the most widespread conception and practice of logic coalesced in the early twentieth century into a view we now identify with the works of Frege, Russell and Whitehead.

The nineteenth century brought, for the first time, from Boole and Frege, distinct proposals on how to symbolize logic that were both extensively developed and had widespread influence. Boole is correctly regarded as the father of modern symbolic logic. Frege shared two concerns of many nineteenth-century mathematicians - avoiding incorrect derivations and providing rigorous and clear foundations for the infinitesimal and derivative calculus - and therefore sought to develop a very clear notion of mathematical 'proof'. His notation has not been taken up but his influence, especially on the move towards symbolization, has been considerable.

De Morgan was the first logician extensively and symbolically to discuss the logic of relations. However, a systematic algebraic notation for relations was provided only later by Peirce, and developed also by Schröder.

1 Major movements and themes

Among the main topics in nineteenth-century logic were (1) the purpose or scope of logic, (2) the extent to which logic can usefully be symbolized or diagrammed, (3) the organization or codification of basic and derivative logical principles, such as in an axiomatization, (4) the relationship of logic to mathematics, (5) the relationship of types of logic, such as deductive and inductive, to each other, and to other ideal forms of reasoning, such as the scientific method, (6) how far logic should depart from the account of quantification in Aristotle's theory of the categorical syllogism, (7) the relationship of types of deductive logic, especially of categorical to propositional logic (usually called 'hypothetical logic' in the nineteenth century), and (8) the extent to which logic needs to treat relations in a special way.

There was little discussion of modal logics; of what we now call 'deviant' logic, such as discussions of nonstandard connectives (including the non-material conditional); or of the relationship of logic and natural language. There was an explicit discussion only in the early part of the nineteenth century (and in Schröder 1890-1905, vol. 1) of whether logic should be 'extensional' or 'intensional', opinion eventually settling on a purely extensional conception. Extensional logic treats terms only according to the concrete entities to which they refer; intensional logic treats terms according to their 'meaning', variously called intension, comprehension or connotation (see Intensional logics §1). C.I. Lewis (1918) advanced the thesis that the extensive symbolic development of logic required this extensional conception.

An already existing English (nominalistic) penchant for a logic of individual, concrete things was further solidified in works by Whately, Boole, Mill, De Morgan, Venn and Peirce; W.S. Jevons (1864) is unusual for his intensional theory of logic. German logic, owing to its often-noted idealist tendencies, had traditionally gravitated towards the approach that logic deals with concepts, not with 'things'. But works by Hermann Grassmann (1844) and Schröder (1877, and the monumental *Vorlesungen* of 1890-1905) had reversed this course. The works of Frege fall squarely within the older German intensional tradition, but this feature was virtually ignored by his English popularizers, such as Russell, and was later conflated with the firmly extensionalist, set-theoretic views of Dedekind and Cantor that arose contemporaneously in mathematics.

2 Purpose and scope of logic

Early modern philosophy, from the Renaissance to the late eighteenth century, had bequeathed numerous candidates for what was to be identified as the purpose or goal of 'logic'. The Renaissance Reform movement in logic, begun by Ramus and continuing through the Port-Royal logicians, modified and simplified the Aristotelian

conception of logic, and augmented it with a study of 'method' - which included inductive logic, for example. The English textbook movement, associated with works by figures such as Isaac Watts in the eighteenth, and Richard Whately and J.S. Mill in the nineteenth centuries, adopted standard Reformist topics, but were less interested in a philosophical theory of logic than in improving the reasoning of their student-readers. (Many twentieth-century textbooks unwittingly follow the Reformist topics and the pedagogical approach.)

Several issues in the very characterization of logic itself emerge. Was logic to cover all forms of reasoning, or only deductive reasoning? Was logic to be a theoretical account of the patterns of valid inference, or a practical, 'how-to' manual to help readers reason better or identify the faulty reasoning of others? The Reform and English-textbook traditions included treatments of deductive and inductive logic, together with other types of reasoning (analogy, causation, or scientific reasoning). This broad scope, together with an overriding pedagogical purpose, usually kept writing at a fairly superficial level, simply summarizing earlier views. The influential and subtle text of Richard Whately (1826) and the popular text of Mill (1843) fall squarely within this tradition. Even Augustus De Morgan (1847) treated 'the Calculus of Inference, Necessary and Probable'. The works of George Boole and Charles S. Peirce are also devoted to both deductive and inductive logics. (Only Peirce seems to have had a fully integrated and original theory of reasoning of all sorts; other authors simply followed the Reform syllabus.) Presumably because of a long-standing interest in a priori sentences and forms of inference going back to Leibniz, German logic was always focused exclusively on deductive logic. This included authors as diverse as Kant, Moritz Drobisch, Ernst Schröder and Frege.

Kant's views had considerable impact on German logic in the early nineteenth century and sharpened the tension between traditional, narrower conceptions of logic and a 'psychological' logic about the way humans must reason (see Kant 1800). Kant's conception of logic as the science of a priori judgments of all sorts, popularized by William Hamilton, led many English writers (Boole, Alexander Bain, Mill) to speak of the 'laws of thought'. In Germany, psychological theories and philosophical theories of logic were more thoroughly mixed by Wilhelm Wundt, Wilhelm Schuppe, Otto Külpe and - under the additional influence of Hegel - Christoff Sigwart and R.H. Lotze (and F.H. Bradley and B. Bosanquet) and many other psychologist-philosophers. These unholy mixtures brought fulminations from Frege that have been widely repeated in the twentieth century, often to suggest a fatal empirical taint to all of nineteenth-century logic. They are, however, rather 'laws' in the sense of how human beings 'must' (Kant) or 'should' (Peirce) reason; this is at worst a transcendental or normative theory of judgment and reasoning, rather than being simply empirical psychology.

Georg Hegel (1812) attacked what had been the often underdescribed guiding concept of most previous theories of logic, namely the idea of 'logical form', arguing that the distinction between 'form' and 'content' (of judgments or thoughts) was illegitimate. His own conception of logic defies simple summary, but vastly extended the Kantian conception. The extensions and critiques of Kant and Hegel drove early nineteenth-century German logic away from traditional logic. German logic as a philosophical discipline survived in the works of several anti-Kantian (and anti-Hegelian) logicians, among them Solomon Maimon, Moritz Drobisch (1838) and Bernard Bolzano, and was renewed through the later historical reappraisal of Leibniz and Aristotle by F.A. Trendelenburg (1847-67). Its true resurrection was, however, almost entirely due to professional mathematicians - Schröder and Frege, then Hilbert - who had little understanding of, or sympathy with, traditional philosophical logic.

3 Logic and mathematics: Boole

The idea of an 'algebra of logic' is a very old one, going back at least to observations by Jakob Bernoulli in 1685 (see Logic in the 17th and 18th centuries), with developed theories in unpublished manuscripts by Leibniz, and in published works by Lambert, Ploucquet, Euler and Grassmann (1844). George Boole then could modestly be credited only with the 'rediscovery' of the algebra of logic (1847, 1854). (Boole was apparently unaware of all these precedents, but was perhaps influenced by English views that were creeping towards a notion of an 'abstract' algebra, such as George Peacock's.) Nevertheless, his was an extensional theory of logic, was more developed than previous theories, occurred at a time when logic was ripe for a new and more mathematical approach and so had a sustained impact (direct or indirect) on all later symbolic theories (except Frege's). Boole is for these reasons correctly counted as the founder of modern symbolic logic.

However, Boole's logic had four features that were widely perceived as shortcomings. First, it had no account of relations. C.S. Peirce (1867-90) merged Boole's theory with De Morgan's work on relations to develop the

‘algebra of relations’ (see §5 below). Second, Boole treated propositional logic as a different interpretation of terms (as classes of ‘occasions’ or of ‘times’, rather than of concrete things). This reduction seemed unproblematic to Booleans but became suspect as Hugh MacColl (in a series of articles, ‘The Calculus of Equivalent Statements’, 1877-80) and Frege drew attention to what we now think of as propositional logic. Third, particular statements (using ‘Some… ’, that is, *I* or *O* propositions) were enduringly problematic, and eventually pointed the way towards the generally inadequate handling of quantification by Booleans. Boole would indicate that some *A*’s are *B*’s by writing $A = \mathrm{v}B$, where ‘v’ can be thought of as an operator that forms an arbitrary non-empty subset. Peirce and Schröder eventually introduced quantifier-like class operators, Π and Σ, to form classes with, respectively, all individuals and some individuals having a given logical attribute. The intersection of two classes, which we indicate by $A \cap B$, might then be written as $\Pi i \; Ai \times Bi$ - the class of all i’s that are both *A* and *B* - as well as by the simpler $A \times B$.

Finally, Boole’s symbol, ‘ $+$’, as described unambiguously in his more influential work of 1854, which can mean class union or propositional disjunction, was notoriously non-inclusive: ‘ $A + B$’ is interpretable only when *A* and *B* are mutually exclusive. Boole was not without his reasons for this troublesome requirement - it greatly simplifies the calculation of joint probabilities in inductive logic, permits an inverse operation to ‘ $+$’, and gets rids of the odd ‘ $1 + 1 = 1$’ (Lambert’s reasoning as well) - but as a feature of deductive logic, it was widely regarded as merely cumbersome. An inclusive interpretation was proposed by Jevons and Peirce in 1867, and this was widely adopted among Booleans. John Venn (1881) held on to the exclusive interpretation - presumably because of his joint concern for inductive logic.

Boole’s earlier work (1847) is in some ways a remarkable, subtle and little-understood work: it does not discuss ‘classes’ and instead gives a theory of what Boole calls ‘elective symbols’. These behave like subset-forming operators on the universe, 1.

Although Boole was paid homage by Peirce and Schröder as the founder (for Schröder, with Grassmann) of the algebra of logic, there are nevertheless respects in which they moved considerably beyond him. They thoroughly revised his handling of quantification, increasingly using quantifier-like operators. They were more interested in the algebra of relations, regarding the simple algebra of logic as somewhat trivial. And finally, they eventually moved away from Boole’s ‘equational’ style of presentation. That is, rather than using a primary symbol of class identity, ‘ $=$’, they used a symbol of class subsumption (‘subset’): Peirce’s ‘ $\prec$’ and Schröder’s ‘ $\subseteq$’.

A deeper difficulty was with Boole’s apparent reasoning why we should want to turn logic into algebra at all. He seems to have had the idea that doing so would make logic-learning easier and, since many people had algebraic symbol-manipulating abilities, logical reasoning could then rely upon these. He did not articulate deep or philosophical reasons why logic really ‘is’ an algebra; it was convenient and helpful to think this way. (In largely unpublished works of the early 1880s) Frege extensively criticized the underlying motivations for this connection between algebra and logic as superficial and philosophically casual.

The works of Peirce and Schröder in the late nineteenth century moved in a direction compatible with Frege’s critiques. Namely, they became less and less algebraic: no longer equational, and with the introduction of quantifiers, they contained variable-binding operators that disallow context-free algebraic-style substitutions. They also ceased pointing out parallels between logic and numerical algebra, and substantive uses of ‘l’ and ‘0’ disappeared.

4 Logic and mathematics: Frege

Frege speaks only briefly of his conception of the purpose of logic: it is to use a ‘*Begriffsschrift*’ (‘conceptual notation’ - Trendelenburg’s translation of a concept used in discussing Leibniz’s twin goals of a ‘*calculus ratiocinator*’ and a ‘*characteristica universalis*’). That is, Frege’s goal in his notation is to display in a perspicuous way the relationships between concepts and propositions (and presumably not merely those that are usefully similar to relationships in mathematics). The goal of the whole of logic is to demonstrate the correctness of deductions without gaps between premises and conclusions, using acknowledged formal and precise rules of inference. Frege’s focus seems initially to have been on determining the correctness (and hence non-synthetic, a priori nature) of mathematical proofs, rather than examining reasoning of all sorts, as had traditionally been the subject of logic. Logic before Frege had not devoted itself only, or even extensively, to reasoning in mathematics;

mathematicians were thought to be those least in need of help. Frege seems to have shared two concerns of many nineteenth-century mathematicians: avoiding incorrect derivations, like many faulty 'proofs' of the parallel postulate from other axioms and postulates in Euclidean geometry; and providing rigorous and clear foundations for the infinitesimal and derivative calculus ('analysis' - see Analysis, philosophical issues in §§1-2). His solution was, in essence, to develop a very clear notion of what counts as a mathematical 'proof'.

One might then guess that Frege was importantly involved in the axiomatic movements of the late nineteenth century, which first occurred in geometry (Hilbert and Veblen), then extended to the axiomatization of number theory (Peirce, Dedekind and Peano), set theory (Zermelo) and logic itself (R.A. Bernstein and Hilbert). This is not precisely correct, for Frege does not have a theory of axiomatics; certainly he had a general desire for care and precision, and developed reasons for this care and clarity, but his own demonstrations are neither more nor less likely to be flawed than those of Boole, Peirce, or Schröder, and are not axiomatic in any definite sense. Furthermore, Frege's notational system was so unusual and cumbersome that even Peano insisted that Frege's formulas be translated into his own notation before he would consider them. Our present notation for quantificational logic derives primarily from that of Whitehead and Russell's *Principia Mathematica*, which they in turn borrowed from Peano (see Peano 1973) - and these symbolisms were taken from original works by Peirce and Schröder's notational devices.

Frege's contribution lay elsewhere. He developed the first theory of quantification. The Boolean quantifiers of Peirce and Schröder were typically only class abstraction operators, with implicit rules given for their use. Frege also gave a very careful account of propositional logic (using notational devices for the material conditional and for negation) and made it the core of his theory - whereas for Aristotelians and Booleans, propositional logic was dependent on the logic of categorical statements, understood as a theory of classes. Furthermore, in later works, Frege was to use this theory to develop an account of the nature of numbers (1893, 1903) that was to have an enormous impact on Russell, and on the philosophy of mathematics. Finally, Frege contributed a great many concepts that have become part of the philosophy of logic: what we now call predicates and propositional functions; the use of 'functions' in logic; the distinction between sense and reference (see Sense and reference §§1-2); and many others.

Logic can be 'mathematical' in several distinct senses. It can simply use mathematical symbols and rules governing them; this is the method of Boole. It can be developed in the methodical, non-circular, clear way that was epitomized in Euclid's axioms and postulates for geometry, but had rarely been applied to other branches of mathematics, or to logic until the very late nineteenth century. It can be an independent discipline that is applied to mathematics, to give an analysis of concepts such as number or proof; this is closer to Frege's methodology. Finally, it can use mathematical techniques - of diagrams, demonstrations and proofs, and careful descriptions of mathematical structures. This is the modern and popular sense in which logic is sometimes said to be 'mathematical'. Peirce and Schröder might have understood and appreciated a claim, for example, that propositional logic defines a distributive lattice - a certain abstract algebraic structure - but Boole probably would not have understood such a claim and Frege would not have thought it insightful.

5 Relations

De Morgan is famous in logic for an often-repeated sentence he never wrote, roughly to the effect that all of Aristotelian logic was helpless to show the validity of the inference 'Since every horse is an animal, every head of a horse is the head of an animal'. The spirit is De Morgan's (except that he was surprisingly sympathetic to Aristotelian logic), but the example was Jevons'. The greatest logical and philosophical minds had long been aware of difficulties with relations, although they had often treated relational arguments as a marginal phenomenon. Plato and Aristotle were keenly aware of issues with regard to relations, and the medievals developed many theories of them (in supposition theory and in the theory of 'oblique' syllogisms - see Logic, medieval §§5, 7). Joachim Jung, the Port-Royal logicians and Leibniz were all aware of this literature - as were De Morgan and Peirce.

With the exception of the remarkable treatment of functions by Lambert in the mid-eighteenth century, De Morgan (1860) deserves credit for being the first logician extensively and symbolically to discuss the logic of relations. He briefly surveys the history of relations (starting with Aristotle), argues for the importance of relations in common thought, and for the centrality of relations in mathematics. He argues that traditional logic had only treated the

relations ‘identity’ and ‘non-identity’. De Morgan emphasizes the formal properties of identity that form the basis of inferential operations, namely transitivity and ‘convertibility’ (what we would now call symmetry or commutativity). He then proposes a notation for relations (or at least for identity), but demurs from a systematic algebraic notation, later to be provided by Peirce.

Peirce first made his proposals in a rambling essay of 1870. They were presented in a more succinct manner in papers of 1882 and 1883. He published his most successful axiomatic-like presentation of the non-relational and relational algebra of logic (extensively using the quantifier-like Σ and Π operators) in 1885 (see Peirce 1867-90). His logic of relations, unlike many logical proposals, was woven into his more general philosophical system, into his philosophy of mathematics, and into his theory of infinity. (Peirce, following a hint from De Morgan, defines infinite classes as those in which certain relational arguments are invalid, although they are valid for finite ones: this is the theory of the ‘transposed’ syllogism.) Peirce used his system extensively, proved a number of results in the logic of relations, invented a number of technical tools (such as ‘prenex’ form, in which all quantifiers occur first and with global scope), and suggested Church’s theorem: that there is no general mechanical routine for determining the validity or invalidity of an argument once relations and quantifiers are permitted.

In 1895, in the third volume of his *Vorlesungen*, Schröder presented the algebraic theory of relations in a lengthy but highly organized fashion, including the proofs of hundreds of theorems. He uses Peirce’s notation and theory, adding some notational variants - he is notably careful about the distinct universes of discourse (Boole’s ‘1’) for two- and higher-place relations, and, more fastidious still, proposes distinct empty classes (Boole’s ‘0’). Schröder uses Peirce’s quantifier-like Σ and Π expressions, but presents both relational logic (1895: vol. 3) and non-relational logic (1890: vol. 1) in a less axiomatic-like way than Peirce had in 1885. He connects the logic of relations to various matrix, numerical and geometric representations, sometimes using these as what we would now describe as ‘models’. Bertrand Russell in 1910, already familiar with Frege’s and especially Peano’s notations, pronounced the algebra of logic ‘cumbersome’.

The Peirce-Schröder notation and theory with relations was used, mainly in German-language logical work, until about 1920 (and in Skolem’s work for slightly longer). An overview is presented in English in Whitehead’s work of 1897, but otherwise no translation of Schröder’s *Vorlesungen* has ever appeared in English or French. (Although reprinted in 1966, it runs to over 2,000 pages and lacks an index.) Peirce’s work was buried in American mathematics journals of the 1870s and 1880s, at least until the publication of his *Collected Papers* from 1931 on. He published no book summarizing his logic. Consequently, since the 1920s, the algebra and logic of relations as a distinct field has been known only to scholars of the history of logic and mathematics, or to those fluent in German. Interestingly, no conclusive proof has ever been printed that the algebra of logic, particularly when enriched with its quantifier-like expressions or with relations, is inadequate in any technical respects. (Tarski claimed in 1941 that there were formulas of first-order predicate logic it could not express, but did not offer a complete demonstration.)

Frege stands curiously outside of the history of attempts to deal with the logic of relations. This is because he seems to have been completely unaware of the history of the logic of relations, and because his own theory was developed so independently of the other efforts to address the problem (that is, he was unaware of the works of De Morgan and Peirce, while Schröder’s work on relations appeared after Frege’s system was already formulated). His own theory, even in the early *Begriffsschrift* (1879), can obviously express everything that the algebraic theories could. Propositional functions (predicates) can have one or more argument places (1879: §10), but Frege does not draw attention to the difference between one-place functions (monadic predicates such as ‘is a cat’) and those with two or more places relational predicates, such as ‘is smaller than’). He seems unaware of or uninterested in the different logical and philosophical behaviour of properties and relations.

6 Main discoveries of nineteenth-century logic

In terms of successful influences, systematic symbolizations comprise the most obvious logical heritage of the nineteenth century. This is a tradition that clearly began with Boole, was modified and extended by Peirce (and to a lesser extent by Schröder), and then slightly reformed by Peano. Frege had little influence on notation now in common use, but considerable influence on the impetus to use symbolization. The linking of logic with mathematics had enormous impact on twentieth-century logic, but there are many distinct such links: the exact connections seen by De Morgan, Boole, Peirce and Frege are thoroughly different, ranging from the view that both

logic and mathematics can and should use similarly behaving symbolic systems, to views that mathematics 'is' logic (Frege, and perhaps Peano), or that logic 'is' mathematics (Boole, De Morgan, certainly Peirce). A further influence was the use of variable-binding operators, like those seen in mathematics (especially in the calculus). Finally, logic of both the Boolean and Fregean sort regarded a treatment of relational predicates as necessary, and not a marginal phenomenon - although this view is only implicit in Frege.

Both Peirce and Schröder used something quite similar to truth tables, a discovery usually credited to Wittgenstein. They also occasionally used models to show the independence of logical principles, a technique that was rediscovered and applied to logic in the 1920s; models themselves were suggested by the 'interpretations' and 'universes of discourse' of Boole and De Morgan. A clear notion of the 'completeness' of a logical system was absent, although they understood notions of independence and consistency; Peirce had some idea of decidability.

The value of studying nineteenth-century logic today, for other than its historical value, might include: algebraic - or at any rate 'mathematical' - structures inherent in logic as a theory of idealized thought or of the world; a reconsideration of the special structures and features to be found in the logic of relations; a perspective of the relationship of logic and mathematics that is closer to De Morgan's and Peirce's than to Frege's or Russell's; a return to an intensional conception of logic that once dominated logic and was rediscovered by Church in the 1950s and 1960s in the works of Frege; and conceptions of collective entities, and infinite collections, that retreat from the ornate and perplexing 'sets' of the twentieth century to something closer to the 'classes' of Boole, Peirce and Schröder - or perhaps to the concepts of Frege.

See also: De Morgan, A.; Logic machines and diagrams; Logical and mathematical terms, glossary of; Mill, J.S. §§2-3

RANDALL R. DIPERT

References and further reading

Boole, G. (1847) *The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning*, London: G. Bell & Sons; repr. in *Studies in Logic and Probability*, ed. R. Rhees, London: Watts, and LaSalle, IL: Open Court, 1952; also repr. in *An Investigation of the Laws of Thought*, New York: Dover, 1973.(Boole's first and in some ways best attempt at a symbolic calculus; far less influential than the *Laws of Thought* (1854) but philosophically far more interesting and exact.)

Boole, G. (1854) *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities*, London: Walton & Maberley; repr. in *An Investigation of the Laws of Thought*, New York: Dover, 1973.(Boole's most famous work, and arguably the most influential work in the history of symbolic logic. Some, including Peirce, have nevertheless considered it to be inferior to his (1847 work.)

Boyer, C.B. and Merzbach, U.C. (1968) *A History of Mathematics*, New York and London: Wiley; 2nd edn, 1989, repr. 1991.(A standard work in the history of mathematics, with useful discussion of the intertwined histories of abstract algebra and 'Boolean' logic, and of the mathematization of logic in the late nineteenth century.)

De Morgan, A. (1847) *Formal Logic: Or, The Calculus of Inference, Necessary and Probable*, London: Open Court, 1926.(A careful work from the same year as Boole's first major logical work; it had less influence than it deserved because of De Morgan's resistance to algebraic-style symbolizations.)

De Morgan, A. (1860) 'On the Syllogism IV and on the Logic of Relations', in *On the Syllogism and Other Logical Writings*, ed. P. Heath, New Haven, CT: Yale University Press, 1966.(The first major work solely on the logic of relations, this had an enormous influence on Peirce, and through him on Schröder and others. De Morgan continued to resist an algebraic-style symbolization but was nevertheless highly rigorous.)

Dipert, R.R. (1990-1) 'The Life and Work of Ernst Schröder' and 'Individuals and Extensional Logic in Schröder's *Vorlesungen über die Algebra der Logik*', in V. Peckhaus (ed.) *Schröder*, special issue of *Modern Logic* 1 (2/3).(The only extensive works in English on Schröder's life and work.)

Drobisch, M. (1838) *Neue Darstellung der Logik nach ihren einfachsten Verhältnissen*, Leipzig: Voss; 3rd revised edn, 1863.(Perhaps the most important and influential logic book in German before Frege, and probably known by him; experiments with notations and diagrams, and shows some independence from the then-dominant Aristotelian and Kantian/Hegelian approaches.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge,

MA: Harvard University Press, 1967, 1-82.(The first major work of logic in a modern, twentieth-century, non-algebraic calculus; a remarkable and much-praised work but with little direct influence on later developments.)

Frege, G. (1893, 1903) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, Jena: Pohle, 2 vols; part 1 of vol. 1 trans. M. Furth, *Basic Laws of Arithmetic: An Exposition of the System*, Berkeley, CA: University of California Press, 1964.(A difficult work of enormous importance and influence, especially on Russell and for logicism: that mathematics was really logic.)

Grassmann, H. (1844) *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik*, Leipzig: Otto Wigand, 2nd edn, 1862; repr. in *Gesammelte mathematische und physikalische Werke*, ed. F. Engel, New York: Chelsea, 1969; trans. L.C. Kannenberg, 'The Linear Theory of Extension: A New Branch of Mathematics', in *A New Branch of Mathematics: The 'Ausdehnungslehre' of 1844, and Other Works*, La Salle, IL: Open Court, 1995.(A strikingly idiosyncratic symbolic work. Grassmann proposes a theory of all extension [one-dimensional quantities] including logic; predates Boole but was well-known only in the German-speaking world, with the exception of Peano.)

Hailperin, T. (1976) *Boole's Logic and Probability: A Critical Exposition from the Standpoint of Contemporary Algebra, Logic and Probability Theory*, Amsterdam and New York: North Holland; 2nd edn, 1986.(Perhaps the best secondary monograph on Boole's logic; unusually sensitive to the integrated aspects of his deductive and inductive logics; possibly the first edition is superior.)

Hegel, G.W.F. (1812-16) *Science of Logic*, trans. A.V. Miller, London: Allen & Unwin, 1969.(A work whose major thesis is that there is no distinction between form and matter in judgments, this work had a huge and negative impact on philosophical formal logic in nineteenth-century Germany.)

Jevons, W.S. (1864, 1869) 'Pure Logic' and 'The Substitution of Similars', in *Pure Logic and Other Minor Works*, London: Macmillan, 1890.(Jevons' early works were, with a less influential essay of Peirce's (1867), the major English symbolic works to advocate an 'exclusive' approach to the or/union operator.)

Jevons, W.S. (1873) *The Principles of Science: A Treatise on Logic and Scientific Method*, New York: Dover, 1958.(An important English logic text, and one of the few that is intensional.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, New York: St Martin's Press, and London: Macmillan, 1965.(A major work with little apparent concern for logic; its extension of a priori to include some synthetic judgments, as well as the 'Table of Categories', nevertheless altered nineteenth-century logic.)

Kant, I. (1800) *Logic*, ed. G.B. Jäsche, repr., together with lecture notes taken by Kant's logic students *c*.1770-90, in *Lectures on Logic*, trans. J.M. Young, Cambridge: Cambridge University Press, 1992.(A 'minor' work of Kant, it nevertheless shows the influence of his philosophy and had considerable impact on later German logicians.)

Kneale, W. and Kneale, M. (1962) *The Development of Logic*, Oxford: Clarendon Press, 1984.(The standard but now dated work in the history of logic; has usable discussions of the nineteenth century, adequate discussions of major figures such as Boole and Frege, but thin and unphilosophical discussions of other figures and important issues such as relations.)

Lewis, C.I. (1918) *A Survey of Symbolic Logic*, doctoral dissertation, Berkeley, CA: University of California Press; revised edn, New York: Dover, 1960.(Quite a dated work in many ways, but one of the few historical works in the field with a single, well-developed, unifying and enlightening thesis: the rise of extensional logic, and its separation from intensional logics.)

MacHale, D. (1985) *George Boole: His Life and Work*, Dublin: Boole Press.(The only extensive discussion of Boole's life, including some of his non-logical work - such as his work in calculus.)

Merrill, D.D. (1990) *Augustus De Morgan and the Logic of Relations*, Boston, MA: Kluwer.(The only major work on either De Morgan or the early history of relations.)

Mill, J.S. (1843) *System of Logic: Ratiocinative and Inductive*, in *Collected Works of John Stuart Mill*, vols 7 and 8, London: Routledge, 1991.(Sometimes said to be the most-published logic book ever perhaps since surpassed by Copi's *Introduction to Logic*); an accessible, well-organized work often lamented for its lack of rigour, poor sense of history, and undeserved influence.)

Peano, G. (1973) *Selected Works of Giuseppe Peano*, trans. and ed. H. Kennedy, Toronto, Ont.: University of Toronto Press.(A good selection in English of the works of Peano, whose notation and conception of logic were probably more influential than any other nineteenth-century logician; a very small collection.)

Peirce, C.S. (1867-90) 'On an Improvement in Boole's Calculus of Logic' and other important logical papers, in *Collected Papers*, vol. 3, ed. C. Hartshorne, P. Weiss and A.W. Burks, Cambridge, MA: Harvard University Press, 1933; and in *Writings of Charles S. Peirce: A Chronological Edition*, ed. M. Fisch *et al.*, Bloomington, IN: Indiana University Press, 1982-.(The *Collected Writings* includes the more carefully edited versions - at

least of works up to 1886.)

Schröder, E. (1877) *Der Operationskreis des Logikkalküls*, Leipzig: Teubner.(Schröder's first, brief and elegant work on logic; shows familiarity with works by Boole and Grassmann, but not yet with works by Peirce or De Morgan, or with the theory of relations.)

Schröder, E. (1890-1905) *Vorlesungen über die Algebra der Logik (Lectures on the Algebra of Logic)*, Leipzig: Teubner, 3 vols; repr. New York: Chelsea, 1966.(The monumental *locus classicus* of the algebra of logic, including relations; a highly systematic work that totals thousands of pages and hundreds of theorems; unindexed, it remains untranslated and has probably been read in its entirety by only a handful of German-speaking readers. See Dipert (1990-1) for discussion.)

Styazhkin, N.I. (1964) trans. *History of Mathematical Logic from Leibniz to Peano*, Cambridge, MA, and London: MIT Press, 1969.(A briefer treatment of nineteenth-century logic than Kneale and Kneale (1962), but often more insightful and knowledgeable in its discussions of symbolisms, of German logic in the eighteenth and nineteenth centuries, and of Russian figures such as Poretskii.)

Trendelenburg, F.A. (1847-67) *Historische Beiträge zur Philosophie*, Berlin: Bethge.(Together with his *Logische Untersuchungen* and editions of Aristotle's *Prior Analytics*, this book began interest in the logics of Leibniz and Aristotle in the German-speaking world; probably read in part by Frege.)

Venn, J. (1881) *Symbolic Logic*, London: Macmillan, 2nd edn, 1894; repr. New York: Chelsea, 1979. (By the time it appeared, this work was unoriginal, but usefully summarized Boolean logic in a book-length form more readable than Boole's own works; the first Boolean 'textbook'.)

Whately, R. (1826) *Elements of Logic*, Delmar, NY: Scholars' Facsimiles and Reprints, 1975.(A highly intelligent, succinct, non-symbolic work that represents the high-water mark of the non-symbolic English 'logic textbook' tradition.)

Logic in the early 20th century

The creation of modern logic is one of the most stunning achievements of mathematics and philosophy in the twentieth century. Modern logic - sometimes called logistic, symbolic logic or mathematical logic - makes essential use of artificial symbolic languages.

Since Aristotle, logic has been a part of philosophy. Around 1850 the mathematician Boole began the modern development of symbolic logic. During the twentieth century, logic continued in philosophy departments, but it began to be seriously investigated and taught in mathematics departments as well. The most important examples of the latter were, from 1905 on, Hilbert at Göttingen and then, during the 1920s, Church at Princeton.

As the twentieth century began, there were several distinct logical traditions. Besides Aristotelian logic, there was an active tradition in algebraic logic initiated by Boole in the UK and continued by C.S. Peirce in the USA and Schröder in Germany. In Italy, Peano began in the Boolean tradition, but soon aimed higher: to express all major mathematical theorems in his symbolic logic. Finally, from 1879 to 1903, Frege consciously deviated from the Boolean tradition by creating a logic strong enough to construct the natural and real numbers. The Boole-Schröder tradition culminated in the work of Löwenheim (1915) and Skolem (1920) on the existence of a countable model for any first-order axiom system having a model.

Meanwhile, in 1900, Russell was strongly influenced by Peano's logical symbolism. Russell used this as the basis for his own logic of relations, which led to his logicism: pure mathematics is a part of logic. But his discovery of Russell's paradox in 1901 required him to build a new basis for logic. This culminated in his masterwork, Principia Mathematica, written with Whitehead, which offered the theory of types as a solution.

Hilbert came to logic from geometry, where models were used to prove consistency and independence results. He brought a strong concern with the axiomatic method and a rejection of the metaphysical goal of determining what numbers 'really' are. In his view, any objects that satisfied the axioms for numbers were *numbers. He rejected the genetic method, favoured by Frege and Russell, which emphasized constructing numbers rather than giving axioms for them. In his 1917 lectures Hilbert was the first to introduce first-order logic as an explicit subsystem of all of logic (which, for him, was the theory of types) without the infinitely long formulas found in Löwenheim. In 1923 Skolem, directly influenced by Löwenheim, also abandoned those formulas, and argued that first-order logic is all of logic.*

Influenced by Hilbert and Ackermann (1928), Gödel proved the completeness theorem for first-order logic (1929) as well as incompleteness theorems for arithmetic in first-order and higher-order logics (1931). These results were the true beginning of modern logic.

1 What is modern logic?

Modern logic crystallized gradually during the period 1850-1950. It was not created by any one person or school, but evolved from the work of several. We now recognize that the following features were needed for its mature development:

(a) There is a formal language whose symbols separate it from ordinary language.

(b) In the formal language, *L*, the syntax is clearly separated from the semantics. The syntax is like the grammar of a natural language, specifying which sequences of symbols (called 'formulas') are permissible. Meaning does not occur in the syntax but in the semantics. Thus logical concepts such as sentence, axiom, formal proof and consistency are part of the syntax. Concepts such as truth, satisfiability and definability are part of the semantics. Formal proofs can be checked effectively and do not depend on the meaning of the terms in them.

(c) Within the syntax of *L*, the logical symbols are clearly separated from the non-logical symbols. The logical symbols include connectives such as 'and', 'not', 'if…then…'. The meaning of the non-logical symbols changes with different interpretations, while that of the logical symbols does not. The non-logical symbols may include individual constants, which, in the semantics, are interpreted as specific individuals. There may also be function constants and relation constants, which are interpreted as specific functions and relations. Besides the logical and non-logical symbols, there are variables, which may be free or bound.

(d) The syntax gives a recursive definition for the formulas of *L*.

(e) There is a clear distinction between the object language *L*, in which our formulas occur, and the metalanguage, in which we speak about *L*. Within the metalanguage, we must define what truth means in *L*. Theorems in the metalanguage are called metatheorems.

(f) Usually there will be levels within our symbolic logic. The lowest level (called propositional logic) is that of the connectives ‘and’, ‘not’, and so on. The next level will have individual variables *x* and quantifiers (‘for all *x*’, ‘there exists an *x*’) over those variables. If that is the highest level, then we have first-order logic. But there may also be function variables (or relation variables) and quantifiers over them, giving second-order logic (see Second- and higher-order logics). In the simple theory of types, there are individual variables, function variables, variables for functions of functions, and so on, and this is sometimes called ω-order logic. We must decide which order our logic will have.

(g) We must also specify other matters. Will the semantics of our logic have only two truth-values, ‘true’ and ‘false’, as in traditional logic, or will it have three or more truth-values (see Many-valued logics)? Will our logic be ‘classical’ (in which the law of the excluded middle holds) or ‘intuitionistic’ (in which that law fails; see Intuitionism)? Will our logic be finitary (like all the logics discussed above) or will it be infinitary (that is, have infinitely long formulas, or have variable-binding operators that give the same effect, or have rules of inference with infinitely many premises; see Infinitary logics)?

(h) Finally, we must study the metatheory of *L*, that is, what general results can be proved, in the metalanguage, to be true of *L*. This includes the consistency and independence of the logical axioms for *L* as well as whether *L* is ‘complete’ (that is, whether every true sentence of *L* is provable in *L*).

2 Logic in 1900

In 1900 there were five main approaches to logic. The first of these was traditional Aristotelian logic, in which the syllogism was central. It had split into several branches, encompassing the idealist logic of Bradley and the pragmatist logic of Dewey. The remaining approaches differed from the first by using symbols in a fundamental way.

The second approach was that of Boole. In 1854 Boole had introduced a logical calculus, whose symbols were borrowed from mathematics, and given his calculus three interpretations: propositions, classes and probability. Around 1900, Boole’s active successors were Peirce and Schröder. During the 1880s Peirce had modified Boole’s logic by introducing quantifiers and relations. The second approach was regarded as trivial by certain Oxford logicians who espoused the traditional one (see Passmore 1968: 156, 240).

The third approach originated in Italy. There Peano began his work on logic (1888) within the Boolean tradition, influenced by Schröder (1877). To avoid confusing Schröder’s logical symbols ‘ $\cdot$’ and ‘ $+$’ (for ‘and’ and ‘or’, respectively) with the corresponding arithmetical symbols for addition and multiplication, Peano replaced them with his own, derived from Hermann Grassmann: ‘ $\cap$’ and ‘ $\cup$’. Peano retained a Boolean ambiguity by interpreting ‘ $\cap$’ and ‘ $\cup$’ not only as ‘and’ and ‘or’ between propositions but also as ‘intersection’ and ‘union’ of classes. (In 1906 Russell removed this ambiguity by keeping ‘ $\cap$’ and ‘ $\cup$’ for ‘intersection’ and ‘union’ while introducing a new symbol ‘ $\vee$’ for ‘or’ and using a dot for ‘and’; nevertheless, like Frege, he had no non-logical symbols and so did not satisfy requirement (c) above.) Peano went beyond the Boolean tradition by symbolizing much of mathematics.

The fourth approach, that of Frege, was isolated from the others. Frege proposed a second-order symbolic language for treating mathematics - particularly the real and complex numbers - in his *Begriffsschrift* (1879). There he replaced the traditional logical notions of subject and predicate by the more mathematical ones of function and argument. In the culmination of his work (*Grundgesetze der Arithmetik*, 1893 and 1903), he attempted to establish that all truths of arithmetic (but not geometry) can be proved from logic alone (see Logicism). To do so, he used a very strong logic, which in 1902 turned out to be inconsistent. No doubt his daunting two-dimensional symbolism partly accounts for his limited influence at the time. But Frege was also isolated by his explicit refusal to work in the Boolean tradition, as well as by his many innovations and his severe criticisms of other logicians.

The fifth approach to logic, that of Hilbert, was rooted in the foundations of geometry. In 1899 he gave axioms for Euclidean geometry that were purely formal, although not yet expressed in symbols. He proposed three main tasks for such an axiom system: establishing its consistency, independence and completeness. (At that time, 'completeness' was imprecise and required only that the axioms imply all known theorems of geometry.)

So at the turn of the century there was no unified tradition in logic and no common view of what logic is and what its goals are; at the same time, it was generally agreed that there is only one correct logic. This fragmented situation continued until at least 1930 (although the Boole-Schröder tradition was strong until the 1920s), when the theory of types became increasingly important. After 1930, first-order logic gradually replaced the theory of types.

3 The algebraic tradition of Boole and Schröder

During the 1890s Schröder continued the Boolean tradition (see §2 above). Schröder's logic began with his 'identical calculus' or lattice theory (1890: 161); it was related to a calculus of classes, one of domains and one of propositions. In 1895, his calculus of relations elaborated that of Peirce.

Although a critic of Frege, Schröder too regarded arithmetic as a part of deductive logic (1890: v) and later went beyond Frege's version of logicism by stating that 'I consider pure Mathematics to be only one branch of general Logic' (1898: 46).

The Boole-Schröder tradition culminated in the work of Löwenheim and Skolem. Löwenheim's most important paper (1915), on Schröder's calculus of relations, distinguished between first-order and second-order logic while retaining elements of infinitary logic (see §10 below). Löwenheim's semantic theorem was that if a first-order formula is true in all countable domains, then it is true in all domains. Löwenheim was well aware that this result failed in his full logic, which was a kind of infinitary second-order logic.

In 1920 Skolem extended Löwenheim's theorem, and in this form it is known as the 'Löwenheim-Skolem theorem': if a countable set of first-order formulas has a model, it has a countable model. This was the first major metatheorem for first-order logic.

4 Peano's tradition

The most important of Peano's logical accomplishments was to invent a clear symbolism. Much of it - such as $\in$ for class membership and $\supset$ for material implication - continues today, thanks to its adoption by Russell. By contrast, the logical symbolisms of Frege and Schröder are all but extinct.

In 1891 Peano believed that he could express any logical relation by using seven primitive symbols: $\in$ ('is'), $=$ ('is equal'), $\supset$ ('implies'), $\cap$ ('and'), $\cup$ ('or'), $-$ ('not') and $\wedge$ ('absurd'). He was mistaken, as Frege remarked. Partly under Frege's influence, Peano used nine kinds of primitive symbol in 1897: the old signs $\in$, $\supset$, $\cap$, $-$, together with the signs *K* ('class'), (x, y) ('ordered pair'), $=_{\mathrm{Df}}$ ('definition'), parentheses and variables. Yet both Frege and Peano accepted the principle of comprehension (that is, every predicate determines the class of those things satisfying the predicate), which eventually led to the paradoxes.

Peano tried to render mathematics completely unambiguous by expressing it in his symbolism. He first used his symbolism in geometry (1888), next in his postulates for the natural numbers (1889), and then he applied it to differential equations (1890). He assembled a group of Italian collaborators to create the encyclopedic *Formulaire de mathématiques* (1894-1908), which aimed to symbolize all major mathematical theorems.

Among those collaborators was Alessandro Padoa, who investigated the undefinability of a term from other terms in a given axiom system (1901). Padoa emphasized that the notion 'definable' was not absolute but depended on the particular axiom system and its primitive symbols. This insight was overlooked when the paradoxes of definability, such as Richard's paradox, were discovered soon afterward (see Paradoxes of set and property §6).

5 Russell's response to Peano and Frege

In 1900, before he discovered the paradoxes, Russell adopted Peano's symbolism and used it to develop his own logic of relations. Russell was particularly impressed by Peano's distinction between the membership relation $\in$ and the relation $\supset$ construed as inclusion. Peano did not distinguish, however, between the implication $p \supset q$ and the inference 'From the proposition *p* infer the proposition *q*', although this distinction had already been made by

Frege (1879). The deduction theorem - a metatheorem showing the correct relation between the two - was only proved by Tarski in about 1921.

For Russell, symbolic logic had three parts: the calculi of propositions, of classes and of relations. In the *Principles of Mathematics*, he gave primacy to the propositional calculus over the calculus of classes (1903: 12), and also proposed his logicism, which went beyond Frege's by including geometry and parts of mechanics: all mathematical concepts are definable from the basic concepts of logic, and all mathematical theorems are provable from the axioms of logic.

In 1903, after finishing the *Principles*, Russell shifted from an analysis of logic in the spirit of Peano to one in the spirit of Frege. Peano had given the universal quantifier only in the form $\phi x \supset_x \psi x$, that is, 'For every *x*, ϕx implies ψx'. Russell introduced a new notation for an idea that he took from Frege: $(x)\phi x$, that is, 'For every *x*, ϕx'. Likewise, Peano had been able to express an existential quantifier only in the form $\exists a$, that is, 'The class *a* is non-empty'. Russell introduced a Fregean existential quantifier: $(\exists x)\phi x$, that is, 'There is an *x* such that ϕx'. Russell's notation later became standard. It had the virtue of disentangling quantifiers from the quantifier-free part of a formula and of disentangling logical formulas from the notion of class. Finally, Russell adopted Frege's assertion sign, '$\vdash$'. Thus, for Russell, $\vdash (x)\phi x$ meant 'It is true that $(x)\phi x$'. (Nowadays $\vdash$ does not mean 'it is true that' but rather, 'it is provable that', and indicates a theorem of logic; see Church (1956: 82).)

6 Hilbert, the geometric tradition and the American postulate theorists

Hilbert's interest in logic stemmed from his devotion to the axiomatic method, which he first exploited to give a definitive form to the axioms of Euclidean geometry in 1899. Here he continued the geometric tradition by using models to establish the consistency and independence of his axioms. Yet he distanced himself from the traditional philosophical concern with what points or numbers 'really' are by treating them as unspecified 'things', defined only by the axioms. As long as the 'things' satisfied the axioms, he insisted, it did not matter if they were points, lines and planes, or tables, chairs and mugs. In accordance with requirement (b) in §1 above, the syntax was separated clearly from the semantics.

At that time Hilbert corresponded with Frege on the foundations of geometry. Frege held the traditional view that the geometric axioms express intuitions, while Hilbert supported the modern view that an axiom system at most determines the 'things' mentioned up to isomorphism (see Nagel 1939). Frege insisted that the only way to prove the consistency of an axiom system was to give a model, while Hilbert believed that the consistency of a system yields the existence of a model in which it is 'true' (see Frege 1980: 39). Brouwer disputed Hilbert's claim in 1929. Around 1930, Gödel proved this claim for first-order logic, but, using incompleteness, refuted it for second-order logic (see §9).

From 1900 until 1940, the axiomatic method, as formulated by Hilbert and Peano, was used by a group of mathematicians sometimes called the American postulate theorists (see Scanlan 1991) to axiomatize a variety of mathematical systems - including the complex numbers, Boolean algebra, and fields - and to prove the consistency and independence of the axioms. This group included Huntington, Veblen and E.H. Moore. In 1902, while giving an axiomatization for the real numbers, Huntington isolated the idea that this system has, up to isomorphism, only one model. Such a system is 'categorical' (to use the name that Veblen gave it in 1904). A year later, Huntington conjectured that every categorical system *S* is deductively complete, that is, every sentence expressible in *S* is provable or refutable in *S*. It was not until Gödel's incompleteness theorems that Huntington was proved wrong. Moreover, in first-order logic, any system having an infinite model is not categorical (see §8).

7 *Principia Mathematica*

Between 1902 and 1913 Russell and Whitehead wrote *Principia Mathematica* (1910-13), which aimed to give strong evidence for logicism by axiomatizing logic (via the theory of types) and deducing much of pure mathematics from it. In doing so, Russell was particularly concerned to avoid the paradoxes of logic. He also shared with Peano a desire to minimize the number of axioms and primitive concepts.

Principia Mathematica strongly influenced the development of logic, but most of that influence did not come until the 1920s. When its second edition appeared in 1925, Russell considered the most important improvement to logic to be the reduction by Sheffer of the primitive concepts of propositional logic from two ('or' and 'not') to one

('neither… nor') and the reduction by Nicod of its axioms from five to one. (Ironically, later logical research hardly used such reductions.) But in that edition Russell did not mention a more important advance made by Wiener (1914) and known to Russell: the definition of ordered pair (and hence of relation) in terms of classes.

Not everyone accepted the theory of types. Veblen argued that 'formal logic has to be taken over by mathematicians…. There does not exist an adequate logic at the present time, and unless the mathematicians create one, no one else is likely to do so' (1925: 141).

Already in the *Principles* (1903: §17) Russell believed it impossible to establish the independence of one axiom of logic from the others. In *Principia Mathematica*, Whitehead and Russell retained this conviction (1910: 95). For Russell, as for Frege, one cannot stand outside of logic, and so it is not surprising that *Principia Mathematica* lacked a metalanguage. This notion of metalanguage originated with Hilbert (1922), from whom it was adopted by Tarski (1930) and Carnap (1934). It was natural when a system of logic was viewed as simply another mathematical system. (Russell (1922) independently considered levels of language.) Nowadays, the most important theorems of logic are metatheorems, that is, theorems in the metalanguage about the system of logic, and they involve the sort of notions that Hilbert used in his theorems about geometry: consistency, independence and completeness.

8 The emergence of first-order logic

By 1955, first-order logic was dominant among logicians who were mathematicians (though the theory of types remained important among philosophers). However, first-order logic was slow in coming. It first clearly emerged as a 'subsystem' of logic in Hilbert's lectures of 1917, published in Hilbert and Ackermann (1928). (Löwenheim had separated a subsystem of logic from his full system in 1915, but his subsystem allowed quantifiers to be expanded into infinitely long formulas, and so was richer than first-order logic, in which all formulas have finite length.)

Hilbert, unlike Löwenheim, was very careful to specify his syntax, and gave, for the first time, a recursive definition of first-order formula (see (d) in §1 above). Hilbert also stated axioms for first-order logic. While he had borrowed his axioms for propositional logic from *Principia Mathematica*, first-order logic was his own. Moreover, he differed from Russell by posing the metamathematical questions of the consistency, independence and completeness of his axioms. He proved the consistency of both propositional and first-order logic, but was unable to answer the other questions. Hilbert wanted to give axiomatizations for various theories within logic, whereas Russell wanted a strong logic in which to deduce those theories from his logical axioms. Hilbert's long-term goal was to prove the consistency of all of classical mathematics by finitary means (see Hilbert's programme and formalism.)

In 1917 Hilbert persuaded Bernays to work with him on logic. Bernays' first results (1918, published only in 1926) established the independence and completeness of propositional logic. Meanwhile, Post had independently found and published them (1921).

Also in 1917, Hilbert emphasized the '*Entscheidungsproblem*': to find a decision procedure for every mathematical problem. Within the Schröderian tradition, decision procedures had already been studied by Löwenheim (1915), who solved the *Entscheidungsproblem* for monadic predicates within first-order logic, a solution improved by Skolem (1919). Within Hilbert's school, Bernays gave a decision procedure for propositional logic in 1918, and in 1928 argued that the *Entscheidungsproblem* was the central problem of logic (see Proof theory §2). Work on the *Entscheidungsproblem* consisted of finding a decision procedure for more and more complicated classes of first-order sentences in prenex normal form (that is, consisting of a quantifier-free matrix preceded by a string of quantifiers), where the complexity of the class was measured by the number of quantifiers in the prefix.

Meanwhile, in 1923, Skolem abandoned the infinitely long formulas adopted from Schröder and Löwenheim, and began working explicitly in first-order logic. Skolem made the radical proposal that the membership relation '$\in$' of set theory be treated, not as a part of logic in the way that Peano and Russell had done, but like any other relation (that is, as a non-logical symbol). In logic, such relations could be given various interpretations, and so should '$\in$'. This proposal resulted in the so-called 'Skolem paradox' of first-order logic: set theory has a countable model although it contains uncountable sets (see Paradoxes of set and property §9).

In France, Herbrand (1930) made a major contribution to the relation between propositional logic and first-order logic.

9 The definitive metatheorems: Gödel on completeness and incompleteness

The most important person influenced by Hilbert's logic was Gödel, whose doctoral dissertation (1929) solved Hilbert and Ackermann's problem of showing, for countable languages, the completeness of first-order logic: every valid first-order formula is provable. In fact, Gödel proved the stronger result that every consistent set of first-order formulas has a model (1929).

When Gödel's completeness theorem was published in 1930, he included as a corollary the 'compactness theorem': a set of first-order formulas has a model if every finite subset has a model. The importance of this theorem was only appreciated later by Maltsev (1936) and Henkin (1947) when they introduced uncountable languages.

Even more important were Gödel's incompleteness theorems (1931). His first incompleteness theorem stated that if a 'decidable' set Σ of formulas implies the axioms for the natural numbers with addition and multiplication, then there is a sentence in the language of Σ that is neither provable nor refutable. (Here 'decidable' means that there is a mechanical procedure for determining whether a formula is in Σ.) The second incompleteness theorem stated that such an axiom system Σ cannot prove its own consistency (see Gödel's theorems).

The incompleteness theorems were devastating. They refuted Russell's logicism by showing that there are true mathematical sentences which cannot be proved in the theory of types. Likewise, they refuted Hilbert's programme, since the consistency of number theory cannot be established except by a stronger theory. Church soon established that there is no decision procedure for first-order logic. Finally, the incompleteness theorems led to attempts, by Carnap and Zermelo, to circumvent them by introducing an infinitary logic (see §10).

10 The prehistory of infinitary logic

Peirce and Schröder treated quantifiers in a way that differs from the modern one. If we fix a finite set with *n* elements $a_1, a_2, \ldots, a_n$ as our universe *D* of discourse and let *Px* mean that *x* has the property *P*, then in *D* we can express 'All *x* have property *P*' by ' Pa_1 and Pa_2 and… and Pa_n'. If *D* is infinite and its elements are $a_1, a_2, \ldots, a_n, \ldots$, then Peirce and Schröder still treat 'All *x* have property *P*' as equivalent to ' Pa_1 and Pa_2 and… and Pa_n and… '. (An analogous statement holds in *D* for 'Some *x* has *P*' and ' Pa_1 or Pa_2 or… or Pa_n or… '.) Thus a quantifier could be expanded as an infinite conjunction or disjunction. This was the beginning of an infinitary logic, but one that was not yet formulated as a distinct kind of logic.

Here we see two distinct ways of treating quantifiers: (1) 'For all *x*, *Px*' and (2) the infinitely long conjunction ' Pa_1 and Pa_2 and… and Pa_n and… '. Frege, Peano and Russell used quantifiers only in the first way. By contrast, Peirce and Schröder used both ways interchangeably. Hilbert (1905) was influenced by Schröder to define quantifiers in the second way when the universe *D* was the natural numbers. In his mature work (1923), Hilbert still viewed quantifiers as going beyond the finitary.

Meanwhile, within the Schröderian tradition, Löwenheim used not only such infinitely long conjunctions and disjunctions but also infinite strings of quantifiers, in his paper proving Löwenheim's theorem. Skolem continued to use this infinitary logic when in 1920 he extended Löwenheim's theorem to a countable set of first-order formulas, but by 1923 he abandoned infinitary logic in favour of first-order logic.

Infinitary logic arose, independently, in Zermelo's paper (1931) reacting against first-order logic and specifically against Skolem's argument that there is a countable model for set theory (see Paradoxes of set and property §9). Zermelo formulated a strong infinitary logic in which, he hoped, the Löwenheim-Skolem theorem was false. Soon he had a second reason to propose an infinitary logic, namely to avoid Gödel's incompleteness theorems and to have every valid sentence be provable. But Zermelo's logic was adopted by no one.

Hilbert, Tarski and Zermelo each formulated a different version of an infinitary rule of inference, now known as the ω-rule. (Influenced by Hilbert and Tarski, Carnap also used the ω-rule.) This rule requires the language of logic to have infinitely many names, one for each natural number *n*. If a formula $\phi(a_n)$ has been proved for the

name a_n of each natural number *n*, the ω-rule infers 'For every *x*, $\phi(x)$'. Without the ω-rule, an axiomatization of the natural numbers could be ω-inconsistent, that is, prove $\phi(a_n)$ for each natural number *n* but also prove 'There is some *x* such that not $\phi(x)$'. Unlike Carnap, Hilbert and Zermelo, all of whom supported this rule, Tarski was occasionally sceptical of it.

Only in 1955 were infinitely long formulas accepted as a legitimate part of mathematical logic - thanks to Henkin, Tarski and Karp - in apparent ignorance of earlier work on such formulas by Zermelo and others. By this time, first-order logic was firmly established as the logical framework for mathematics, and those infinitary formulas were considered within various extensions of first-order logic. A different kind of extension was proposed by Mostowski (1957): generalized quantifiers. One such quantifier was 'there exist uncountably many'. For that quantifier, Mostowski asked if it is possible to prove a completeness theorem analogous to that for first-order logic. Later the answer turned out to be 'yes'.

11 A glance past Gödel: recursion theory, model theory, set theory

Modern logic began with Gödel's completeness and incompleteness theorems. Today they are the goal of most introductory logic courses. From his incompleteness theorems, recursion theory emerged during the 1930s, when the notion of 'computable' was made precise by the Gödel-Herbrand notion of recursive function and, equivalently, by Turing machines (see Computability theory).

Model theory emerged as a distinct part of logic around 1950, although its first important result (Löwenheim's theorem) dates back to 1915. One fundamental notion - relational structure - was already vaguely present in Hilbert (1899). A relational structure is an ordered pair consisting of a set and a family of relations on that set; such structures include groups, fields and Boolean algebras, but not topologies. Model theory began to develop into its modern form through the discovery that the strong completeness theorem for first-order logic (that is, any consistent set of first-order sentences has a model) implies results about structures. Connected with this was Henkin's use of languages with uncountably many symbols (1947) and Tarski's classification (1952) of structures up to 'elementary equivalence' (that is, being indistinguishable in first-order logic).

Meanwhile, set theory developed much closer ties with logic. In 1938 Gödel used first-order logic to prove the important result that if the usual axioms of set theory are consistent, they remain so when the axiom of choice and the generalized continuum hypothesis are assumed (see Axiom of choice; Set theory). Gödel's idea was to consider, not all sets, but only those sets first-order definable from sets of previous levels (see Constructible universe). And in 1963 first-order logic was essential to Cohen's proof that the axiom of choice and the generalized continuum hypothesis are independent of the usual axioms of set theory (see Forcing).

12 Coda: what is modern logic?

Until about 1850, it was fairly easy to say what logic was. Today it is much harder, although classical first-order logic remains dominant. After 1920, many-valued logics (that is, those having three or more truth-values) were investigated by Post and Łukasiewicz. Likewise, modal logic was seriously studied, thanks to C.I. Lewis (1918; see Modal logic). And during the 1950s infinitary logic crystallized. Finally, there are now many specialized logics (see, for example, Quantum logic).

During the 1970s first-order logic was characterized by 'Lindström's theorem': first-order logic is the strongest logic that contains propositional logic and for which the compactness theorem and the Löwenheim-Skolem theorem hold. Thus any logic which extends first-order logic must violate the compactness theorem or the Löwenheim-Skolem theorem. An idea of the wide range of modern logic can be obtained by consulting Barwise (1977), Barwise and Feferman (1985), and Gabbay and Guenthner (1983-9).

See also: Logical and mathematical terms, glossary of

GREGORY H. MOORE

References and further reading

Barwise, J. (ed.) (1977) *Handbook of Mathematical Logic*, Amsterdam: North Holland.(An excellent survey of modern logic.)

Barwise, J. and Feferman, S. (eds) (1985) *Model-Theoretic Logics*, New York: Springer.(The best survey of

infinitary logic.)

Bernays, P. (1918) ‘Beiträge zur axiomatischen Behandlung des Logik-Kalküls’ (Contributions to the Axiomatic Treatment of the Logical Calculus), unpublished *Habilitationsschrift*, Göttingen University; published in much revised form, ‘Axiomatische Untersuchung des Aussagen-Kalküls der *Principia Mathematica*’ (Axiomatic Investigation of the Propositional Calculus of *Principia Mathematica*), *Mathematische Zeitschrift* 25: 305- 20, 1926.(First proof of the completeness of propositional logic; compare with Post (1921).)

Carnap, R. (1934) *Logische Syntax der Sprache*, Vienna: Springer; revised and trans. A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul, and New York: Harcourt Brace, 1937.(A precise version of the theory of types and the ω-rule.)

Church, A. (1956) *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press.(An erudite but useful introduction to modern logic.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. ‘*Begriffsschrift*, a Formula Language, Modelled upon That of Arithmetic, for Pure Thought’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82.(Frege’s first, ground-breaking work.)

Frege, G. (1893, 1903) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, Jena: Pohle, 2 vols; repr. as *Grundgesetze der Arithmetik*, Hildesheim: Olms, 1966; Part 1 of vol. 1 trans. M. Furth, *Basic Laws of Arithmetic: An Exposition of the System*, Berkeley, CA: University of California Press, 1964; extracts from vol. 2, including ‘Frege on Russell’s Paradox’, in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980. (Volume 1 gives a detailed development of his system of logic. Volume 2 has Frege’s reaction to Russell’s paradox.)

Frege, G. (1980) *Philosophical and Mathematical Correspondence*, trans. H. Kaal, ed. G. Gabriel *et al*, Chicago, IL: University of Chicago Press.(Translation of Frege’s letters. The correspondence with Hilbert and Russell is especially interesting.)

Gabbay, D. and Guenthner, F. (eds) (1983-9) *Handbook of Philosophical Logic*, Dordrecht: Reidel, 4 vols.(A standard modern source on classical and non-classical logics.)

Gödel, K. (1929) ‘Über die Vollständigkeit des Logikkalküls’, doctoral dissertation, University of Vienna; trans. ‘On the Completeness of the Logical Calculus’, in *Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort, New York and Oxford: Oxford University Press, 1986, 44-101.(First proof of the completeness theorem for first-order logic.)

Gödel, K. (1930) ‘Die Vollständigkeit der Axiome des logischen Funktionenkalküls’, *Monatshefte für Mathematik und Physik* 37: 349-60; trans. ‘The Completeness of the Axioms of the Functional Calculus of Logic’, in *Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort, New York and Oxford: Oxford University Press, 1986, 102-23; and in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 582-91.(First published proof of the completeness of first-order logic, together with the compactness theorem.)

Gödel, K. (1931) ‘Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I’, *Monatshefte für Mathematik und Physik* 38: 173-98; trans. ‘On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems’, in *Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort, New York and Oxford: Oxford University Press, 1986, 126-95; and in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 592-617.(First proof of Gödel’s incompleteness theorems.)

Grattan-Guinness, I. (1981) ‘On the Development of Logics between the Two World Wars’, *American Mathematical Monthly* 88: 495-509.(Overview of logic from 1920 to 1940.)

Grattan-Guinness, I. (1985) ‘Russell’s Logicism versus Oxbridge Logics, 1890-1925: A Contribution to the Real History’, *Russell: The Journal of the Bertrand Russell Archives*, new series, 5: 101-31.(Traditional logic versus the symbolic logic of Peano and Russell.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press.(Includes translations of many relevant articles.)

Henkin, L. (1947) ‘The Completeness of Formal Systems’, unpublished doctoral dissertation, Princeton University.(First proof of the completeness theorem for uncountable languages.)

Herbrand, J. (1930) ‘Recherches sur la théorie de la démonstration’, *Prace Towarzystwa Naukowego Warszawskiego*, Wydzial 3, no. 33; trans. ‘Researches on Proof Theory’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967,

525-81. (Herbrand's theorem.)

Hilbert, D. (1899) *Grundlagen der Geometrie*, Leipzig: Teubner, 7th edn, 1930; 2nd edn trans. L. Unger and P. Bernays, *Foundations of Geometry*, La Salle, IL: Open Court, 1971.(Consistency and independence results in geometry; the modern axiomatic method.)

Hilbert, D. (1905) 'Über die Grundlagen der Logik und der Arithmetik', in A. Krazer (ed.) *Verhandlungen des dritten internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, Leipzig: Teubner, 174-85; trans. 'On the Foundations of Logic and Arithmetic', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 129-38. (Hilbert's first attempt to prove the consistency of arithmetic.)

Hilbert, D. (1922) 'Neubegründung der Mathematik. Erste Mitteilung' (New Foundations of Mathematics. First Communication), *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 1: 157-77; trans. 'The New Grounding of Mathematics', in W.B. Ewald (ed.) *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, vol. 2*, Oxford: Clarendon Press, 1996, 1115-34.(His first explicit use of 'metamathematics'.)

Hilbert, D. (1923) 'Die logischen Grundlagen der Mathematik' (The Logical Foundations of Mathematics), *Mathematische Annalen* 88: 151-65; trans. 'The Logical Foundations of Mathematics', in W.B. Ewald (ed.) *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, vol. 2, Oxford: Clarendon Press, 1996, 1134-48.(His mature work on proof theory.)

Hilbert, D. and Ackermann, W. (1928) *Grundzüge der theoretischen Logik*, Berlin: Springer, 2nd edn, 1938; trans. L.M. Hammond, G.G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, ed. R.E. Luce, New York: Chelsea, 1950.(First modern textbook in mathematical logic.)

Lewis, C.I. (1918) *A Survey of Symbolic Logic*, Berkeley, CA: University of California Press; revised edn, New York: Dover, 1960.(A historical account of modern logic, together with an axiomatization of modal logic.)

Löwenheim, L. (1915) 'Über Möglichkeiten im Relativkalkül', *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, 'On Possibilities in the Calculus of Relatives', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 232-51.(First important results on the semantics of first-order and infinitary logic.)

Maltsev, A.I. (1936) 'Untersuchungen aus dem Gebiete der mathematischen Logik', *Matematicheskii Sbornik*, new series, 1: 323-36; trans. 'Investigations in the Area of Mathematical Logic', in B. Wells (ed.) *The Metamathematics of Algebraic Systems: Collected Papers, 1936-1967*, Amsterdam: North Holland, 1971.(First use of the compactness theorem for uncountable first-order languages.)

Moore, G.H. (1980) 'Beyond First-Order Logic: The Historical Interplay Between Mathematical Logic and Axiomatic Set Theory', *History and Philosophy of Logic* 1: 95-137.(The influence of logic on set theory, and vice versa.)

Moore, G.H. (1988) 'The Emergence of First-Order Logic', in W. Aspray and P. Kitcher (eds) *History and Philosophy of Modern Mathematics*, Minneapolis, MN: University of Minnesota Press, 95-135.(How first-order logic emerged from richer logics, including higher-order and infinitary logics.)

Moore, G.H. (1997) 'The Prehistory of Infinitary Logic: 1885-1955', in *Tenth International Congress of Logic, Methodology, and Philosophy of Science, Florence, August 1995*, vol. 2, M.L. dalla Chiara, K. Doets, D. Mundici and J. van Benthem (eds) *Structures and Norms in Science*, Dordrecht: Kluwer, 105-23.(The early history of infinitary logics from Peirce through Löwenheim and Skolem to Novikov and Bochvar.)

Mostowski, A. (1957) 'On a Generalization of Quantifiers', *Fundamenta Mathematicae* 44: 12-36. (Introduces 'generalized' quantifiers as an extension of first-order logic.)

Nagel, E. (1939) 'The Formation of Modern Conceptions of Formal Logic in the Development of Geometry', *Osiris* 7: 142-224; repr. in *Teleology Revisited and Other Essays in the Philosophy and History of Science*, New York: Columbia University Press, 1982, 195-259.(The influence of geometry on modern logic.)

Padoa, A. (1901) 'Essai d'une théorie algébrique des nombres entiers, précédé d'une introduction logique à une théorie déductive quelconque', in *Premier Congrès International de Philosophie*, vol. 3, *Logique et Histoire des Sciences*, Paris: Armand Colin, 309-65; partial trans. in van Heijenoort (1967), 118-23.(Method for showing a term to be indefinable.)

Passmore, J. (1968) *A Hundred Years of Philosophy*, Baltimore, MD: Penguin.(A general history emphasizing logic, especially good on Oxbridge logicians.)

Peano, G. (1888) *Calcolo geometrico*, Turin: Bocca.(Includes Peano's first contribution to logic, which was used in geometry.)

Peano, G. (1889) *Arithmetices principia, nova methodo exposita*, Turin: Bocca; partial trans. 'The Principles of Arithmetic', in J. van Heijenoort (1967), 83-97.(Includes the Peano postulates for the natural numbers.)

Peano, G. (1890) ‘Démonstration de l’intégrabilité des équations différentielles ordinaires’, *Mathematische Annalen* 37: 182-228.(First use of logical symbolism outside of arithmetic and geometry.)
Peano, G. (1891) ‘Principii di logica matematica’, *Revista di matematica* 1: 1-10; trans. ‘Principles of Mathematical Logic’, in *Selected Works of Giuseppe Peano*, Toronto, Ont.: University of Toronto Press, 1973, 153-61.(Peano’s first full logical system.)
Peano, G. (1897) ‘Logique mathématique’, in *Formulaire de mathématiques*, Turin: Bocca, vol. 2, §1. (Peano’s revised logical system.)
Post, E.L. (1921) ‘Introduction to a General Theory of Elementary Propositions’, *American Journal of Mathematics* 43: 163-85; repr. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematics, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 264-83.(First published proof of the completeness of propositional logic; compare with Bernays (1918).)
Russell, B.A.W. (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992.(First publication of Russell’s paradox and his logicism.)
Russell, B.A.W. (1906) ‘The Theory of Implication’, *American Journal of Mathematics* 28: 159-202.(His full development of propositional logic.)
Russell, B.A.W. (1922) ‘Introduction’, in L.J.J. Wittgenstein, *Tractatus Logico-Philosophicus*, trans. C.K. Ogden and F.P. Ramsey, London: Routledge; trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1961. (Russell’s introduction of a hierarchy of metalanguages.)
Scanlan, M. (1991) ‘Who Were the American Postulate Theorists?’, *Journal of Symbolic Logic* 56: 981-1,002. (Early American axiomatic theory.)
Schröder, E. (1877) *Der Operationskreis des Logikkalküls*, Leipzig: Teubner.(Schröder’s first logical system.)
Schröder, E. (1890) *Vorlesungen über die Algebra der Logik (Exakte Logik) (Lectures on the Algebra of Logic (Exact Logic))*, vol. 1, Leipzig: Teubner; repr. New York: Chelsea, 1966.(Summary of the Boolean tradition in logic.)
Schröder, E. (1898) ‘On Pasigraphy: Its Present State and the Pasigraphic Movement in Italy’, *The Monist* 9: 44-62.(Schröder’s logicism.)
Tarski, A. (1930) ‘Über einige fundamentale Begriffe der Metamathematik’, *Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie* 23: 22-29; trans. J.H. Woodger (1956), ‘On Some Fundamental Concepts of Mathematics’, in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(Tarski’s first venture into metalogic.)
Tarski, A. (1952) ‘Some Notions and Methods on the Borderline of Algebra and Metamathematics’, in *Proceedings of the International Congress of Mathematicians, Cambridge, Massachusetts*, Providence, RI: American Mathematical Society, vol. 1, 705-20.(Beginnings of modern model theory.)
Tarski, A. (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press; repr. and ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(Translations of his papers on logic.)
Veblen, O. (1925) ‘Remarks on the Foundations of Geometry’, *Bulletin of the American Mathematical Society* 31: 121-41.(Veblen’s plea for involving mathematicians with logic.)
Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press; 2nd edn, 1925; repr. London: Routledge, 1994.(Their full development of the theory of types.)
Wiener, N. (1914) ‘A Simplification of the Logic of Relations’, *Proceedings of the Cambridge Philosophical Society* 17: 387-90.(Shows that ordered pairs, and hence relations, can be defined in the theory of types.)

Logic machines and diagrams

By 'logical diagrams' we generally mean any two-dimensional representations of logical relationships, such as of class inclusion or consequence. One usually also means representations using non-typographical symbols or geometric figures. Such diagrams were first used in the seventeenth and eighteenth centuries, but gained wide currency only in the nineteenth; the best known are the Euler and Venn diagrams. It is an open question whether logical diagrams are useful only as elementary pedagogical devices, or have implications for advanced logical research.

The conception of an organism, the mind, or of the universe as 'machine' was not really attractive and useful until machines were widespread, complex and able to perform interesting tasks. This occurred first in the late Renaissance, and initiated ways of thinking that dominated the seventeenth and eighteenth centuries. It was also becoming evident that machines could be used to perform some complex, repetitive or difficult tasks more reliably or faster than human beings.

The very idea of machines that can perform 'symbolic' tasks, such as mathematical, logical or, eventually, linguistic ones required first a symbolism. For this reason, the idea of computers for mathematical or logical tasks, and systems of mathematical and logical notation, are strongly intertwined: one must have efficient ways of feeding information into a machine, and interpreting the results.

1 Logical notation

Among the higher-level purposes for any symbolism in logic are: to abbreviate or simplify the writing or the display; to highlight strictly logical relationships without non-logical distractions (roughly Leibniz's *lingua philosophica* or his *characteristica universalis*); or to facilitate insights into the ways these representations may be manipulated, such as through inference (roughly Leibniz's *calculus ratiocinator*, 'calculus of reason').

It is customary to divide notational devices between those that represent substantive (often called 'categorematic') terms and those that represent logical relations or properties (often called 'syncategorematic' terms). Symbols for categorematic terms can stand for determinate individual or abstract entities, such as '*c*' for Julius Caesar, '*a*' for the class of apples, '*A*' for the property of being an apple, or '*P*' for the sentence 'All apples are red'. Symbols may also stand for arbitrary, non-specific examples of such terms, such as in the example, 'Consider a class of objects, *c*, or a proposition, *P*'. These are usually called variables, schemata or place-holders.

Perhaps the earliest use of abbreviations for categorematic terms was by Aristotle, who occasionally used single letters to stand for non-specific categorical terms in his *Prior Analytics*. While many later Aristotelian logicians occasionally used letters to represent categorical terms, only in the late medieval and early modern period do we find any use of special symbols, other than Latin or other 'natural language' expressions, for syncategorematic (logical) notions. One of the first of these was Luis Vives' use in the first half of the sixteenth century of a V-shaped symbol to represent term inclusion:

$$(1)\quad \begin{array}{c} A \\ \mathrm{v} \\ B \\ \\ \\ \end{array} \qquad\qquad (2)\quad \begin{array}{c} A \\ \mathrm{v} \\ B \\ \mathrm{v} \\ C \end{array}$$

Diagram (1) represents that the term *A* is wholly included in *B*. ('All *A*'s are *B*'s.') Diagram (2) indicates that *A* is included in *B*, *B* is included in *C*, and, implicitly, *A* is (thus) included in *C*.

2 Philosophies of logical notation

The final goal of any logical notation is presumably to best allow one to carry out and communicate logical work. With Vives' and later notations, we begin to see four intermediate goals concerning how logical notation can achieve this final goal:

(1) The use of linear linguistic symbols, primarily deriving from natural language or other common usage (such as

mathematics) and used mainly as abbreviations. Polish notation is an extreme example of this phenomenon, using, for example, '*Kpq*' for $p \& q$. (We might also include here mnemonic aids: for example, 'K' for Konjunktion - alas, only for speakers of some languages.)
(2) The use of special symbols chosen mainly for their distinctness from natural language symbols.
(3) The use of special symbols or notations that are 'ideographic' (Styazhkin) or 'iconic' (Peirce): properties of the symbols themselves are intended to mirror attributes of the represented logical relationship.
(4) Two- (or higher) dimensional notational systems using elements of 1, 2, or 3; that is, diagrams.

Observe that (1) and (4) may clash: achieving (1) may aid writers, printers and proof-readers, while achieving (4) may aid readers of the notational system. Similarly, (1) obviously clashes with (2) and perhaps with (3), if the natural language symbols are not fortuitously iconic ('and' is a non-palindrome although conjunction is symmetric).

Leibniz was fascinated by ideographic notations (owing to a passing interest in Chinese pictograms), but it was C.S. Peirce who first seems to have consciously chosen and systematically discussed symbolisms according to their iconic characteristics. One of his points is that to represent a non-symmetric logical relationship, one should use a visually non-symmetric symbol. That is, since $A \supset B$ does not always have the same truth-conditions as $B \supset A$, then one should use an asymmetric symbol, such as $\supset$, $>$ or his own $-\!\!<$. Such considerations have been at work implicitly in the choice of virtually all logical symbols for conjunction, disjunction and so on (except those chosen in line with (1), such as '*K*' and 'Čamp;').

Various symbols that came into widespread use in algebra and calculus in the sixteenth, seventeenth and eighteenth centuries, such as '$=$' and '$+$', involved similar (implicit) considerations, and hark back to suggestions by Viète, Descartes and Leibniz. The 'algebra of logic', such as we see in works by Leibniz, Lambert and, especially, Boole, borrows symbols already in use from algebraic notation (and thus is a species of goal (1)), but capitalizes upon their original iconic character in mathematics - goal (3). Both Peirce and Schröder later gave up the use of symbols for logical constants that were simply borrowed from algebra, using considerations from both (2) and (3), while preserving the algebraic (substitutional) character of many operations. (See Logic in the 19th century.)

While the use of parentheses as indicating priority or scope of operations is of long-standing mathematical use, many of the other symbols now in use derive from the work of Giuseppe Peano. (These were sometimes inspired by earlier works in the algebra of logic, and were later injected into the mainstream of logic by Russell, Whitehead and others.) These symbols were primarily chosen by Peano according to consideration (2) - distinctiveness - and according to their suggestive, or etymological character: '$\in$' for the Greek 'esti'; '$\supset$', a backward 'C', for the converse of 'containment', and so on. He seems not to have considered the 'iconic' arguments of Leibniz and Peirce, and forcefully rejected the more original two-dimensional proposals of Frege.

3 Euler and Venn diagrams

The two most widely used styles of logical diagrams are usually referred to as 'Euler' and as 'Venn' diagrams. Both are primarily used to represent inclusion or subsumption of classes or terms in Aristotelian logic. There are two main distinctions between these diagrams. Euler's method involves placing circles of various sizes in various patterns; Venn's uses circles of uniform size in fixed positions, which are then modified by other markings. Euler's method does not have any unambiguous way of indicating the existence of individuals satisfying a certain description, other than implicitly by the spacing of the circles. Venn has such techniques.

Leonhard Euler developed his method in his *Lettres à une Princesse d'Allemagne sur divers sujets de physique et de philosophie (Letters to a German Princess on Various Subjects of Physics and Philosophy)* in 1761, published in 1772 (see Logic in the 17th and 18th centuries §5). The method was later adapted by Gergonne, Maimon and others, and was frequently given an intensional as well as the original extensional interpretation. The following diagrams exhibit the method:

Euler Diagrams

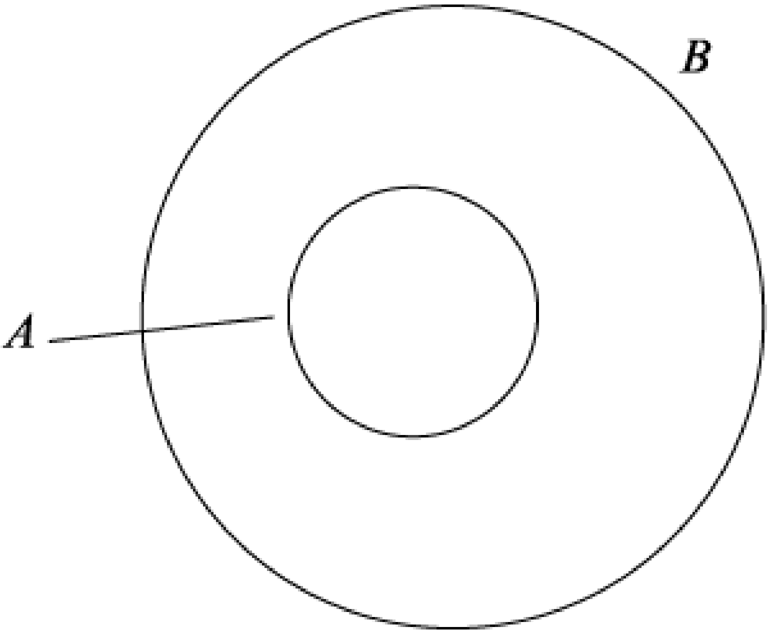

1. *A*-statement

Intensional reading: all the properties in concept *A* are included in the concept of *B*; all *A*'s are *B*'s. Extensional reading: all of the *A*-type things fall in the class of *B*-type things; all *A*'s are *B*'s.

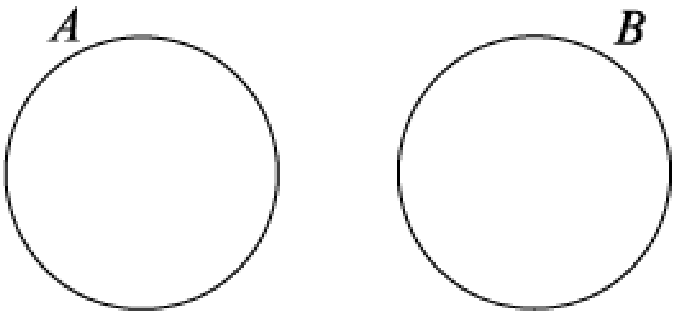

2. *E*-statement

Intensional reading: concepts *A* and *B* share no properties. Extensional reading: no individual is both *A* and *B*.

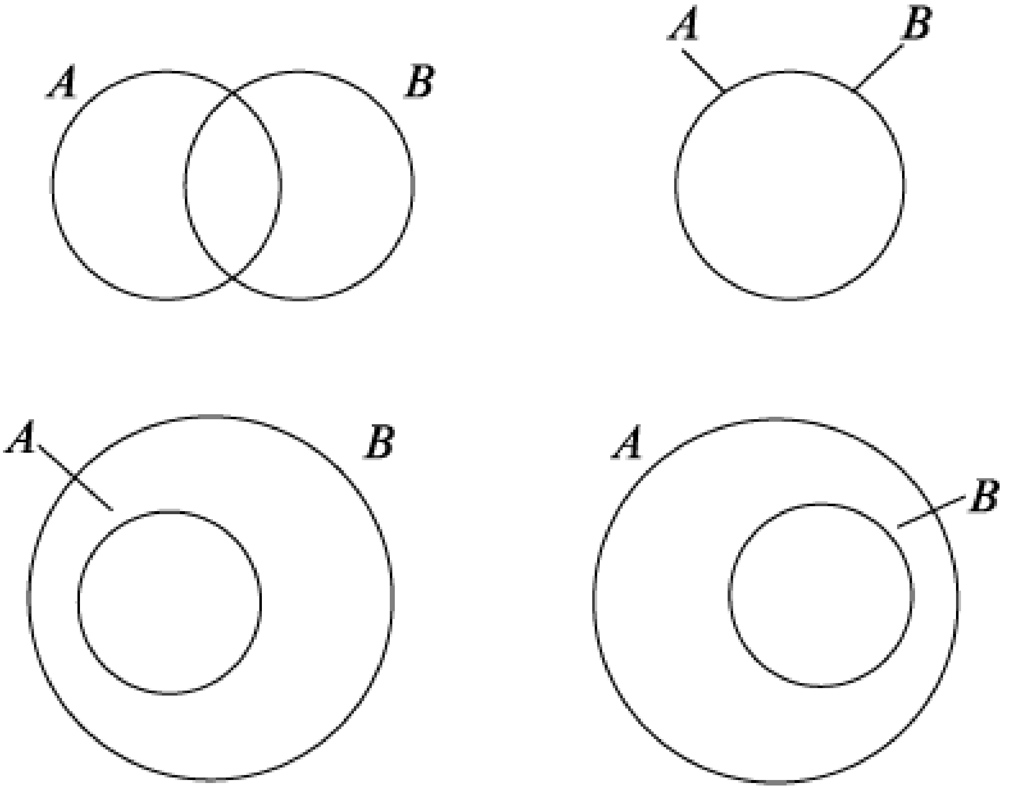

3. *I*-statement

Intensional reading: concepts *A* and *B* share some properties. Extensional reading: there are individuals that are both *A* and *B*. (Alternatives classified by Gergonne.)

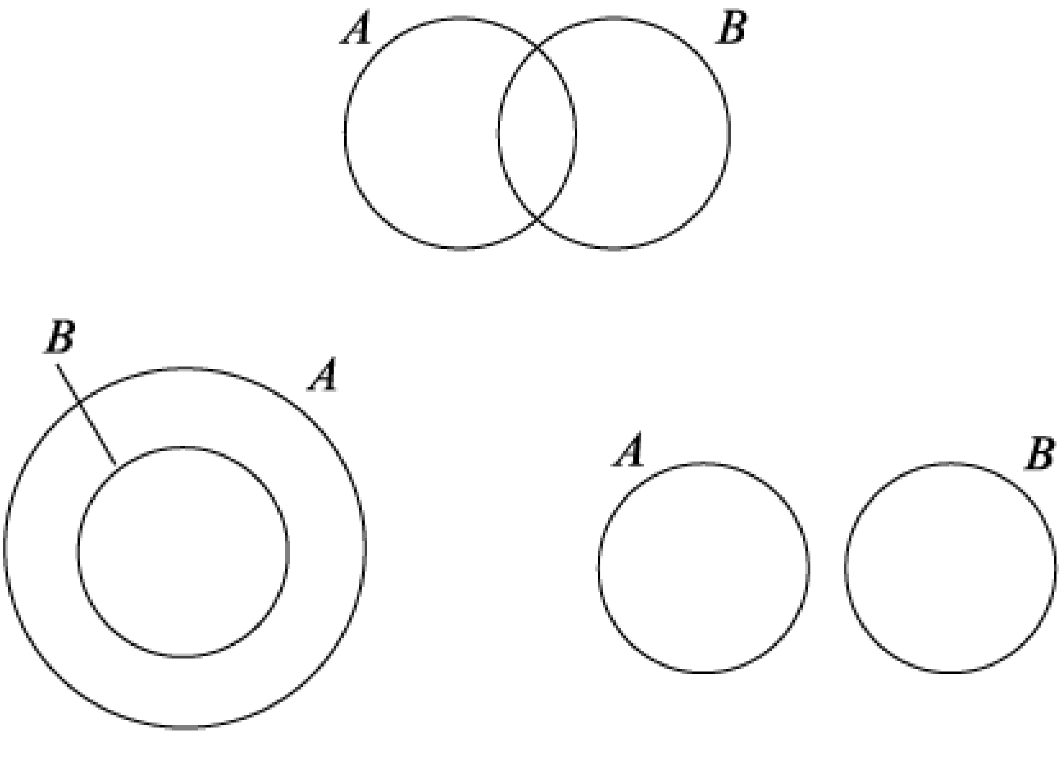

4. *O*-statement

Intensional reading: there are properties in concept *A* that are not in concept *B*. Extensional reading: there are individuals that are *A* but not *B*. (Alternatives classified by Gergonne.)

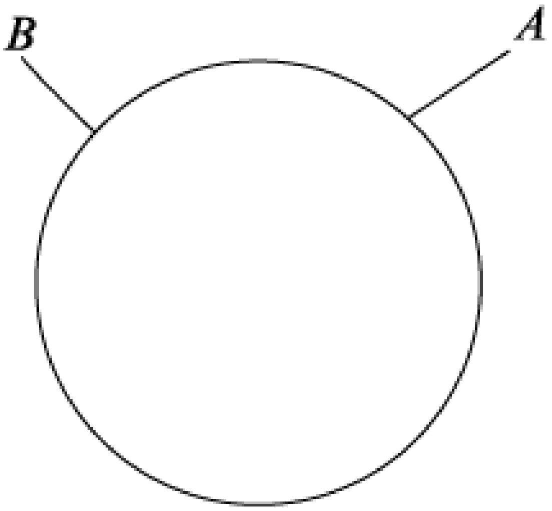

5. Identity

Intensional reading: *A* and *B* share all their properties. Extensional reading: the individuals that are *A* are the same as those that are *B*.

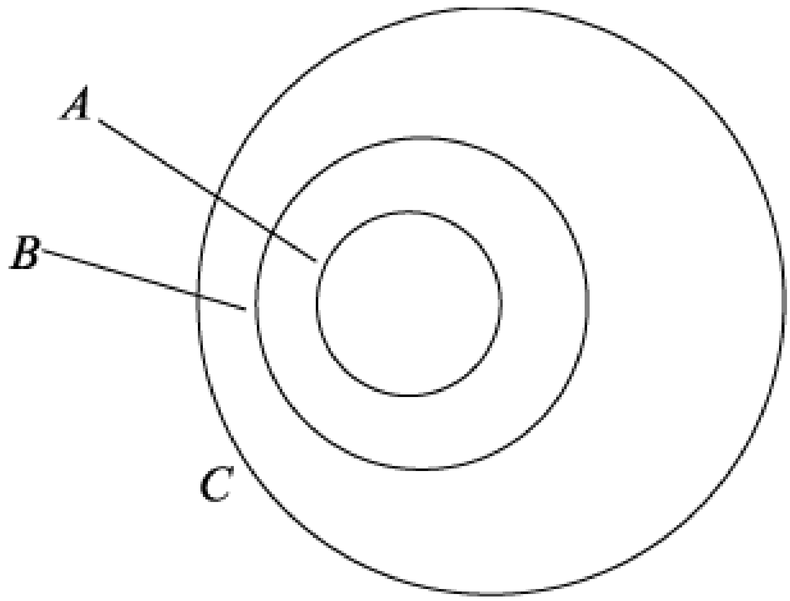

6. 'Barbara' syllogism

All *A*'s are *B*'s, all *B*'s are *C*'s, so all *A*'s are *C*'s, extensionally interpreted.

Observe that the intensional and extensional reading of the *A*-statement are 'reversed': to say that all apples are fruit is to say that all components of the concept of fruit (come from a plant, contain a seed, are edible, and so on) are included in the more numerous components of the concept of 'apple'. Observe also that there is an overlap in the alternative diagrams of the *O*- and *A*-statements (made explicit by Gergonne). Such an observation led to William Hamilton's (and George Bentham's) notorious doctrine of the 'quantification of the predicate', although similar ideas were common in the eighteenth and nineteenth centuries: we should distinguish between 'All *A*'s are some *B*'s' (1, 3, and 4) and 'All *A*'s are all *B*'s' (5).

The method of using movable, usually intensionally interpreted, circles was first suggested in published works by J.C. Sturm in 1661, and in unpublished writings of Leibniz. Contemporaneous with Euler, J.H. Lambert (1764) developed a method of using overlapping lines, also extensionally interpreted:

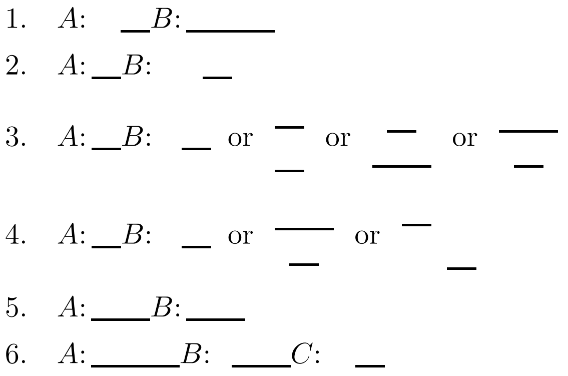

This method had also been anticipated by Leibniz in unpublished notes. In fact, the uses of circles or lines iconically to represent logical relationships reappears so often, and apparently independently, in the history of logic since the seventeenth century that one can count the idea - like that of an ‘algebra’ of logic - as more or less obvious to anyone who thought much about logic.

Venn diagrams are usually credited to John Venn in an article of 1880, and then more extensively developed in his *Symbolic Logic* (1881). This method was anticipated by similarly extensional diagrams of the German logician Moritz Drobisch in his 1838 *Neue Darstellung der Logik (New Conception of Logic)* (see Logic in the 19th century §2). After a decline in usage in the early twentieth century, Venn diagrams have reappeared in many introductory logic texts of the second half of the twentieth century.

Venn Diagrams

Statements are interpreted extensionally: points in an area indicate individuals having certain properties. What properties individuals have (or what classes they belong to) is indicated by inclusion in a circle: if they lack a property, they fall outside of that circle. The non-existence of individuals with a certain description is indicated by shading the corresponding area. The existence of an individual with certain properties (or lacking certain properties) is shown by an ‘*x*’ if it is fully determinate what properties it has, or by a line across two or more areas if it is not determinate.

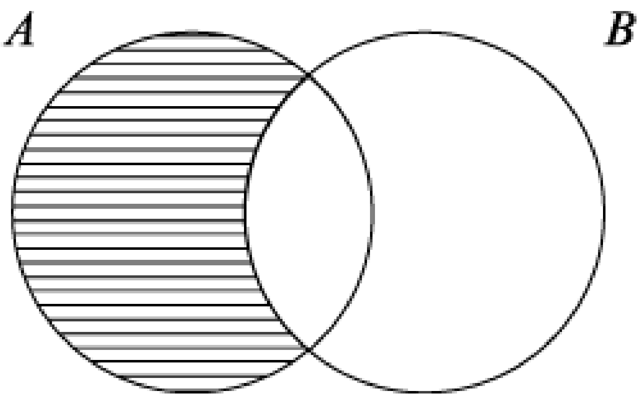

All *A*’s are *B*’s.

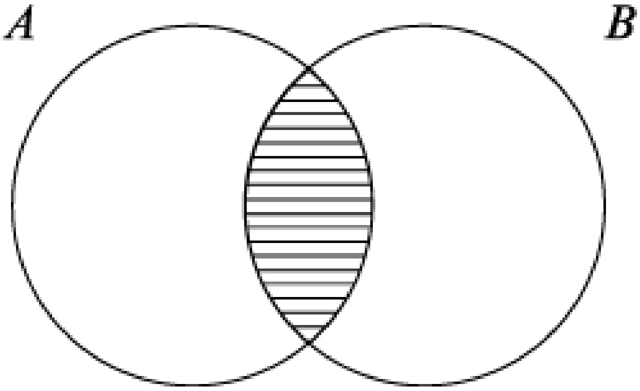

No *A*’s are *B*’s.

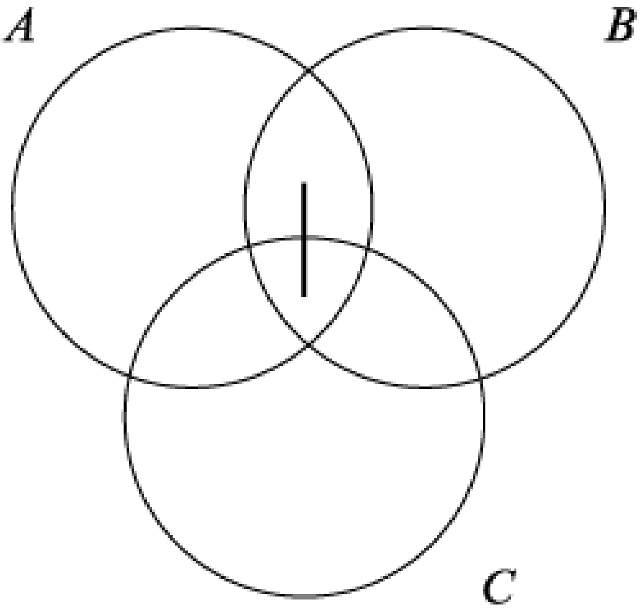

Some *A*’s are *B*’s: maybe they are *C*’s, maybe not.

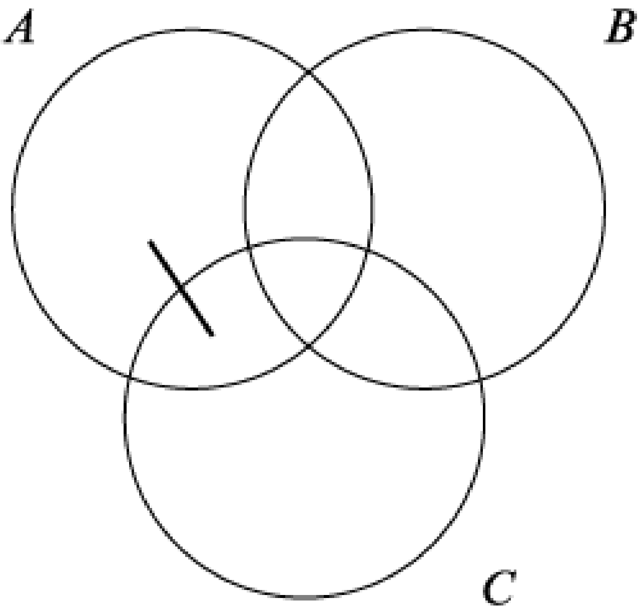

Some *A*’s are not *B*’s.

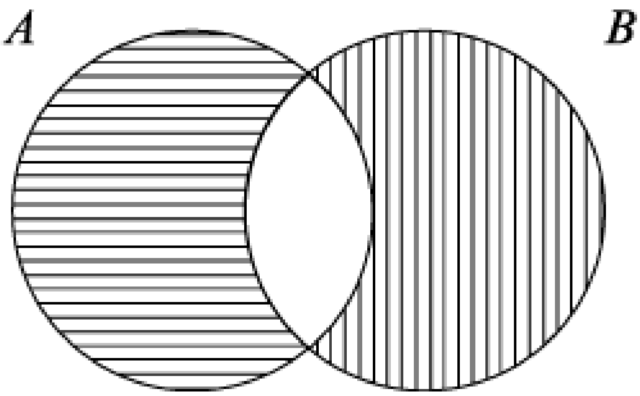

All *A*’s are *B*’s, and all *B*’s are *A*’s.

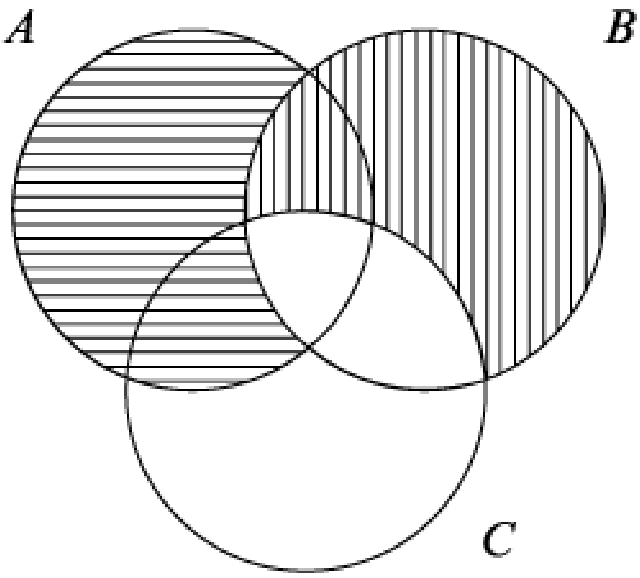

All *A*’s are *B*’s.

All *B*’s are *C*’s.

So, all *A*’s are *C*’s.

The details of this interpretation are suggested by Drobisch and Venn, but carefully stated by Schröder in his *Vorlesungen* (1890-1905). Other diagrammatic methods in the nineteenth century included a system of triangles used by William Hamilton, and the use of matrices or tables by Allan Marquand and Lewis Carroll. Still other methods in the twentieth century include Veitch diagrams (see Veitch 1952) and Karnaugh maps. All of these methods, with the exceptions of the methods of Peirce and Frege, are versatile enough only for Aristotelian (monadic predicate) or propositional logic, and not for full first-order predicate logic.

4 Frege’s *Begriffsschrift*

Gottlob Frege’s notation, the famous ‘*Begriffsschrift*’ (‘conceptual notation’; see Frege, G.), is partly diagrammatic. He draws the material conditional in a two-dimensional way, writing $P \supset Q$ as follows.

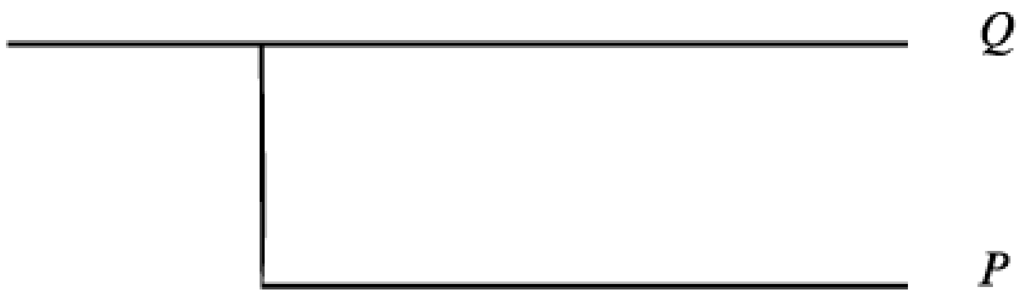

The only specifically two-dimensional feature seems to be intended to highlight a feature of *modus ponens*, the main rule of inference using the material conditional - namely, the condition (antecedent) 'drops away'. The use of the negation line and the descending vertical line (negation stroke) in the material conditional was clearly intended to suggest the interpretation of the material conditional $P \supset Q$ as $-P \vee Q$. Frege stubbornly stuck to his own notational system throughout his life, in spite of the difficulty it presented for typographers, criticism from Peano (and Schröder), its non-iconic and ahistorical character, and the fact that even admirers such as Russell and, later, Wittgenstein and Hilbert, never adopted it. Nevertheless, something of what motivated Frege, namely a concern for displaying features of *inferences* in a two-dimensional way, rather than the logical content or 'sense' of individual sentences, leads one to helpful formats for displaying mathematical or logical demonstrations ('proofs'), such as the later 'natural deduction' methods of Gerhard Gentzen. (See Natural deduction, tableau and sequent systems §1).

5 Peirce's existential graphs and conceptual graphs

There is little suggestion in any of these proposals that diagrams have any point much beyond a heuristic value, being 'suggestive', and helping to learn logic at the introductory level. No one seems to have proposed that diagrams have a serious, advanced use for professional logicians, or represent more accurately an underlying metaphysical reality. A notable exception are Charles S. Peirce's 'existential graphs' (see Peirce, C.S. §8). Peirce, a historian of logic and a careful reader of his contemporaries, knew about the diagrammatic methods of Euler, Lambert and Venn. In the 1890s he developed three sophisticated systems of logical diagrams that he called the alpha, beta and gamma existential graphs. (His choice of 'graph' for these diagrams is somewhat unfortunate, since they are not graphs in the technical sense first used by J.J. Sylvester in 1878, now in wide use, and which Peirce would have known.) Both graph theory and Peirce's graphs were inspired, interestingly, by chemical diagrams. Furthermore, other chemical metaphors occur in both the works of Frege ('saturated') and Peirce. Alpha graphs are similar to Venn diagrams, and can be used for either categorical logic or for what we would call propositional logic. Beta graphs can be used to represent any quantified relational sentence or set of sentences, and are thus a fully general method for diagramming sentences in first-order predicate logic - the first such method in the history of logic. Gamma graphs are used for the representation of modal statements. All three graph systems are 'iconic' in a sense of which Peirce was explicitly aware; he was also interested in enhancing their calculational usefulness for producing new inferences ('forward inference'), or for identifying the correctness of given conclusions from given premises ('backward inference').

Furthermore, the graphs are not just employed for their usefulness in some narrow sense, but grew out of views he had developed much earlier: his 'philosophy of notation' of 1885. This was the first statement of what became a large, developed theory of signs (his semiotics) and a philosophy of mind - of thoughts themselves, at least in all a priori disciplines, being 'diagrammatic'. (Peirce's views are reminiscent of the better-known 'picture theory' of Wittgenstein's *Tractatus*.) Peirce believed that our visual notations should match or facilitate, as much as possible, the best possible mental diagrams, and this means that the notational system should exhibit useful logical relationships. In earlier, algebraic-style writings, this was developed within a standard linear sequence of symbols; in later writings, he seems to have believed that two-dimensional representations were required for the properly 'iconic' display of some logical relationships, especially once relations are introduced.

Peirce's ideas about existential graphs were not much noticed until the 1960s, and have not had a noticeable impact, except as a curiosity, until the 1980s. In computer science and artificial intelligence, Peirce's ideas have been put forward as 'conceptual graphs' (by John Sowa 1991) and have found a wider audience. This is in part because there had existed in artificial intelligence a prior interest in using 'semantic networks' for representations of the meaning of natural language sentences. In cognitive science and learning theory, there is also a resurgence of interest in optimizing the display of information to native human cognitive processes, and thus of presenting complex relational information through visual images.

6 Early logic machines and mechanical machines for Aristotelian logic

Interestingly, the very first speculations about or design of machines that performed conceptual tasks - one hesitates to call them exactly 'logic' - were inspired not by arithmetic calculators, but by a strange mixture of medieval magic and theology. This was the fantastic *Ars Magna* ('Great Art' or 'Technique') of the Spanish

(Majorcan) mystic and tinkerer Ramon Llull (§2). Llull was interested in the way the various concepts (God, evil,…) could be combined. To this end, he designed various notations, as well as ‘machines’ that used labelled, movable, concentric wheels to show these combinations. Although his views were respected by a variety of later thinkers such as Bruno and Swift, his greatest impact was on the development of the combinatoric theory by Pascal and Leibniz in the late seventeenth century.

Not surprisingly, the thought occurred to Pascal, Leibniz and others of using the combinatoric wheels of Llull, but labelling them with numerals, and gearing them so that they could do various arithmetic calculations. There are thus two conceptual steps that were necessary to produce the idea of logic machines. First, one must be able to think of machines as being able to perform calculations - to compute. (This in turn requires a notation and machines of some complexity and accuracy, that is, many finely machined parts.) Second, one must be able to think of reasoning and inference as a type of ‘mechanical’ calculation: this was the contribution of Llull, refined by Pascal and Leibniz, and explicitly stated by Leibniz as the *calculus ratiocinator*.

Both Pascal and Leibniz designed and had constructed for them geared arithmetic calculators. Leibniz also proposed, perhaps for the first time, the idea of a logic machine, and left some suggestions about how this might be done - but never developed precise designs nor constructed an actual machine.

The first reasonably successful logic machine was probably the eighteenth-century ‘Demonstrator’ of Charles Stanhope. It used a notation system that anticipated William Hamilton’s variant of Aristotelian logic, and could also perform elementary arithmetic and probabilistic calculations. The Demonstrator was constructed from a block into which two moving slides were inserted, one of grey wood and the other of red glass; the positions of these slides served as the input premises and the extent of their overlap, visible in a small square in the block, represented the strongest possible conclusion validly derivable from these premises. That is, the Demonstrator was a ‘forward inference’ machine that actually generated a conclusion, rather than determining the validity or invalidity of a given conclusion with given premises. The use of the Demonstrator for deductive logic was limited to categorical logic with two premises.

The high-water mark of mechanical logic machines was reached by William Stanley Jevons’ machine, usually called the ‘Logical Piano’, from the appearance of its keyboard. While restricted to categorical logic, this 1870 machine represented a vast improvement over Stanhope’s Demonstrator, both logically and mechanically. Consisting of a keyboard for inputting premises and, above it, connected by pulleys and levers, a display of terms, the Jevons machine could accommodate up to four terms, and thus arguments with more than two premises (‘sorites’), with designs for machines with up to ten terms. It could also handle ‘or’ terms in either subject or predicate terms (understood inclusively), such as ‘All fruit are apples or oranges’. It used a Boolean equational system, and thus could also be considered the first algebra machine, since it essentially solved systems of equations. Like the Demonstrator, one did not input a conclusion to determine its validity given a set of input premises, but instead the machine output all valid conclusions with the input premises in the logical system.

7 Electrical machines and the twentieth century

A slight improvement in Jevons’ design was made by Allan Marquand, who had been a logic student of Peirce and later became a professor of art history at Princeton. He called his a ‘New Logical Machine’ (1885). It was not logically more sophisticated than Jevons’ machine, but had a better display, consisting of a grid of rotary indicators, looking something like a water or gas meter. Marquand also sketched designs for the first electrical logic machine, using relays and switches. (The very first electrical circuit diagram to be used for logic or interpreted logically was proposed in a letter from Peirce to Marquand.) Nothing really came of these proposals, and they, like Babbage’s ideas for a general-purpose symbolic manipulator, were historical curiosities that had no discernible influence on electrical and electronic computers of the mid- and late twentieth century.

While there were a number of authors who developed mechanical logic machines in the late nineteenth and early twentieth centuries, mostly in imitation of Jevons’ device, the collapse of the Aristotelian approach to logic made most of these techniques obsolete. Indeed, the logic of Frege, Russell and Whitehead, which permits quantifiers and relations, and which we know as predicate logic, makes the very idea of logic machines a daunting task. Expressions no longer consist of certain restricted combinations of terms. Furthermore, as Church’s theorem was later to suggest, it may be that certain logical tasks in this framework are mechanically impossible. Similarly, the

idea of allowing a machine to reason whatever it can from given premises - forward inference, imitative of human reasoning - becomes extremely problematic. Designs for electrical logic machines appeared again shortly after the Second World War, first with work by two students of W.V. Quine, William Burkhart and Theodore Kalin in 1947, a number of machines in the late 1940s and early 1950s, and culminating in the 'Logic Theory Machine' of Allen Newell and Herbert Simon in 1956. (During the Second World War, Claude Shannon had developed his influential logical interpretation of electrical circuitry, thus rediscovering the observations of Peirce and Marquand.) Significantly, all these machines were restricted to propositional logic, and thus also fell far short of the expressivity and complexity of quantificational logic. Logic machines and computer programs in the twentieth century have been largely focused on 'backward' inference: confirming that a given conclusion does, or does not, validly follow from a given set of premises.

The history of logic machines grew distinct, after the early days of Leibniz and Pascal, from the history of machines to perform arithmetic calculations. There were enormous economic pressures to develop numerical calculators (for accounting and also for navigational tables), and few incentives to develop machines that reasoned using logic. Another line of development that was initially distinct from that of logic machines was the development of general purpose, programmable computers. Most early calculating and logic machines of all sorts were one-purpose and 'hard-wired'. Charles Babbage is the originator of the conception of a general-purpose 'computer', but the approach was rediscovered in the twentieth century, and intensively developed in mechanical, electrical and electronic implementations during and after the Second World War. In this approach, we build a 'computer' so that it can be 'programmed' to perform any task of symbolic manipulation that is capable of algorithmic description. Although the idea of such a general-purpose symbol-manipulator is an old one, going back to Babbage in the early nineteenth century, no one seems to have thought of programming such machines to perform logical tasks until Turing, nor succeeded until the Logic Theory Machine. Today, virtually all computational projects in logic are performed on general-purpose computers, and thus in an important sense the idea of dedicated 'logic machines' is obsolete. It is the software (programming), not the hardware, that is designed to perform logical tasks. Starting in the 1960s and accelerating rapidly in the 1970s and 1980s, a large amount of computer software (including 'mechanical theorem provers') now exists that performs a variety of logical tasks in various logical systems, including first- and second-order predicate logic, 'deviant' logics and modal logics.

See also: Computability and information; Computability theory; Logic and mathematical terms, glossary of

RANDALL R. DIPERT

References and further reading

Cajori, F. (1952) *A History of Mathematical Notations*, La Salle, IL: Open Court; repr. New York: Dover, 1993. (A classic book - and the only one - on a monumentally important topic in intellectual history; nevertheless dated and much in need of a treatment using recent historical, logical and cognitive-science research.)

Eames, C. and Eames, R. (1990) *A Computer Perspective: Background to the Computer Age*, ed. G. Fleck, Cambridge, MA: Harvard University Press, new edn.(A pictorial and well-researched history of computers that includes pictures, diagrams and facsimiles of Stanhope's, Jevons' and Marquand's logic machines, Peirce's circuit diagrams, and some post-Second World War work.)

Gardner, M. (1982) *Logic Machines and Diagrams*, Chicago, IL: University of Chicago Press, 2nd edn.(A classic discussion of the topic, and the only monograph, but somewhat elementary and superficial.)

Goldstine, H. (1972) *The Computer from Pascal to von Neumann*, Princeton, NJ: Princeton University Press, 1993.(The history of arithmetic and general-purpose computers, including some 'logic' machines.)

Kneale, W. and Kneale, M. (1962) *The Development of Logic*, Oxford: Clarendon Press, 1984. (Standard but now dated work in the history of logic, with some brief discussions of notations and diagrams, and mention of some logic machines.)

Lambert, J.H. (1764) *Neues Organon oder Gedanken über die Erforschung und Bezeichnung des Wahren und dessen Unterscheidung vom Irrthum und Schein (New Organon or Thoughts on the Investigation and Designation of the True and its Distinction from Error and Appearance)*, Leipzig, vol. 1, part 1, 109-33; repr. in *Philosophische Schriften*, vols 1-2, ed. H.W. Arndt, Hildesheim: Olms, 1965.(Introduces the diagrams sometimes known as 'Lambert lines'; also referred to in later works by Venn and John Neville Keynes in the nineteenth century.)

Peano, G. (1973) *Selected Works of Giuseppe Peano*, trans. and ed. H. Kennedy, Toronto, Ont.: University of Toronto Press.(A small collection of Peano's huge output, this includes some of his important writings on

logical and mathematical notations, and his visionary proposals for a new universal language, ‘Latin without inflections’.)

Peirce, C.S. (1931-58) *Collected Papers*, ed. C. Hartshorne, P. Weiss and A.W. Burks, Cambridge, MA: Harvard University Press, 8 vols; and *Writings of Charles S. Peirce: A Chronological Edition*, ed. M. Fisch *et al.*, Bloomington, IN: Indiana University Press, 1982-.(Important logical papers of 1867-86 appear in volume 3 of the *Collected Papers*; other volumes include some writings on the existential graphs and iconicity.)

Roberts, D.D. (1973) *The Existential Graphs of Charles S. Peirce*, The Hague: Mouton.(The only book-length work in English on Peirce’s existential graphs; more accessible than Peirce’s own works.)

Schröder, E. (1890-1905) *Vorlesungen über die Algebra der Logik*, Leipzig: Teubner, 3 vols; repr. New York: Chelsea, 1966.(The monumental *locus classicus* of the algebra of logic, including relations; extensive discussions of diagrams, various notations, and philosophical aspects of them.)

Shin, S.J. (1990) *The Logical Status of Diagrams*, Cambridge: Cambridge University Press.(Situation theory is used to provide a precise analysis of reasoning with Venn diagrams.)

Sowa, J. (ed.) (1991) *Principles of Semantic Networks*, San Mateo, CA: Morgan Kaufmann. (Discussions of various network and graph formalisms in cognitive science, including the Peirce-inspired conceptual graphs.)

Styazhkin, N.I. (1964) *History of Mathematical Logic from Leibniz to Peano*, Cambridge, MA, and London: MIT Press, 1969.(Often more useful discussions of diagrams than Kneale and Kneale (1962), including Lambert lines.)

Veitch, E.W. (1952) ‘A Chart Method for Simplifying Truth Functions’, *Proceedings of the Association for Computing Machinery* 1952: 127-33.(An exposition of what have come to be known as ‘Veitch diagrams’.)

Venn, J. (1881) *Symbolic Logic*, London: Macmillan, 2nd edn, 1894; repr. New York: Chelsea, 1979. (By the time it appeared, this work was unoriginal, but introduced in textbook fashion the now-popular Venn diagrams.)

Logic of ethical discourse

Logic, as a discipline, is largely concerned with discovering principles and methods for evaluating the evidential strength between the premises and conclusions of arguments. Because the meanings of terms (and the concepts they express) that occur in arguments bear importantly on questions about evidential relations, much of the work on the topic of logic and ethics has been preoccupied with questions about the meanings of moral terms and concepts, and with the correct linguistic analysis of sentences that contain them. Taking logic to include issues about meaning (which has commonly been done by those who refer to the so-called 'logic of moral discourse') is to construe the subject broadly. But the field of logic is often construed quite narrowly to refer to the study of formal languages whose syntax, axioms and inference rules are sufficiently determinate to allow decisions about what counts as the theorem in such a language. On the narrower understanding of logic, the intersection of logic and ethics has mainly to do with work in deontic logic. This article takes up issues concerning the intersection of ethics and logic broadly construed. The intersection of logic and ethics concerns questions about the nature of moral reasoning. Some philosophers have attempted to deduce substantive moral conclusions from factual statements - in particular, to derive 'ought' statements from 'is' statements. If one can successfully carry out such deductions, then moral reasoning is guided properly by consideration of nonmoral facts from which moral conclusions can be derived. However, the eighteenth-century philosopher David Hume is often credited with arguing that no such deductions are correct; that there is a gap between factual 'is' statements and moral 'ought' statements. There is disagreement over whether or not Hume's negative claim is correct; but even if it is, there may still be logical features of moral concepts that impose constraints on proper moral reasoning.

One such widely discussed constraint is the thesis of universalizability, according to which relevantly similar cases must receive the same moral evaluation. One implication of this thesis is that moral judgments about particular cases entail universal moral principles and so some have argued that all correct moral reasoning must be understood in terms of subsuming particular cases under general moral principles. Although many philosophers have accepted this subsumptive model of moral reasoning, it has come under attack by philosophers who argue that proper moral reasoning is primarily a matter of sensitively discerning the morally relevant details of a case under consideration and rendering a moral judgment about it without the guidance of principles.

1 The is/ought question and Hume's law

In a famous passage (1739/40), David Hume complains that writers on morality often move almost 'imperceptibly' from factual statements about what is the case to conclusions about what ought to be the case, and do so without adequately explaining or justifying this transition from 'is' to 'ought' (see Hume, D. §4). Hume then demands that 'a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it' (1739/40: 469). What has come to be called 'Hume's law' is the claim that the sort of 'deduction' Hume calls for cannot be negotiated successfully, hence the maxim, 'No "ought" from an "is"'. (Whether Hume actually held this maxim is the subject of some scholarly debate.) The alleged is/ought gap is often used to support the claim that ethics is logically autonomous in the sense that no moral statements can be correctly inferred from any set of nonmoral premises (a thesis which, as we shall see in §4, is much stronger than Hume's law).

Although questions about the relation between 'is' statements and (moral) 'ought' statements has generated a great deal of philosophical discussion, the problems and issues they involve are part of a more general issue about the relation between factual statements and normative statements (in particular moral statements of all types, including statements about value or goodness) (see Moral realism). However, most of what follows will focus on the is/ought issue.

To understand the debate surrounding the alleged is/ought gap, more must be said about Hume's law and its philosophical presuppositions. The law presupposes that a useful distinction can be made between factual ('is') statements and moral (specifically, 'ought') statements, though, of course, many moral statements (such as 'Keeping one's promises is right' and 'Lying for personal gain is something one ought not do') contain the word 'is'. If we take for granted an understanding of specifically moral uses of such terms as 'right', 'good' and 'ought', then, intuitively, moral statements are statements in which some moral term like 'right', 'good' or 'ought' is

predicated of some object of evaluation (whether or not the statement contains the word 'is'). Nonmoral statements are all those statements in which moral terms are not predicated of some object of evaluation. Factual statements represent a proper subset of nonmoral statements which purport to assert the existence of, or describe (using vocabulary other than terms like 'right', 'good' and 'ought'), things, events, states of affairs or whatever. Now the root idea behind Hume's law is that it is not possible to infer validly an 'ought' statement from a set of factual statements, since, as Hume puts it, 'this *ought*, or *ought not*, expresses some new relation or affirmation' (1739/40: 469; original emphasis) not contained in the premises. In order, that is, for a moral 'ought' statement to be inferred validly from a set of factual statements, the latter would have to entail the former; but there is no such entailment relation since factual claims do not include any moral terms and, in particular, they do not include any use of the term 'ought'. So, for example, from the factual claim of the form (F) one cannot infer validly a moral claim of the form (M):

(F) All or most people do, or would upon reflection, approve of such and such type of action being performed.
(M) Such and such type of action ought to be performed.

Obviously, one way to refute Hume's law and overcome the alleged is/ought gap would be to include among the set of factual premises what can be called 'moral bridge principles' that would link moral terms with nonmoral terms and thus allow for the valid deduction of moral conclusions. So, for instance, if we add to (F) a bridge principle of the form (B), then statements of the form (M) can be inferred validly by canons of ordinary logic:

(B) If all or most people do, or would upon reflection, approve of such and such type of action, then actions of that type ought to be performed.

Granted, (B) together with (F) jointly entail (M), but this does not disprove Hume's law unless a case can be made for claiming that (B) is a nonmoral, factual claim. And this in turn depends on how we interpret it. Traditionally (and this is another presupposition of the philosophical debate over Hume's law) any claim (if true) is either an analytic truth, in which case it is true simply in virtue of the meanings of its constituent terms, or a synthetic truth, in which case its truth does not depend simply on the meanings of its constituent terms (see Analyticity). If construed properly as a synthetic truth, then any instance of (B) has the status of a substantive moral claim, and hence adding it to an appropriate instance of (F) to derive an instance of (M) will not count as a violation of Hume's law - in fact such a derivation tends to confirm that law. However, if we can find an appropriate moral bridge principle that is construed properly as an analytic truth, then, since in general such truths simply report conceptual or linguistic truths about language, such a principle would simply describe a nonmoral fact about proper linguistic usage. So it would appear that if the is/ought gap is to be negotiated successfully, there must be analytic bridge principles that link moral terms and concepts with nonmoral terms and concepts. Moreover, although linking moral terms and concepts analytically with any sort of nonmoral, descriptive terms and concepts would serve to bridge the gap, most serious philosophical efforts to find any such analytic connections represent versions of ethical naturalism (see Naturalism in ethics). Ethical naturalism (at least in its traditional form) involves the attempt to define basic moral terms and concepts by terms and concepts that describe or refer to the sorts of natural properties and relations that are the subject of scientific inquiry. Thus, those who wanted to defend Hume's law were particularly concerned to refute versions of ethical naturalism.

One of the most forceful defences of Hume's law was offered by G.E. Moore (1903), who claimed that all attempts to define moral terms and concepts using naturalistic terms and concepts commit the 'naturalistic fallacy' (see Moore, G.E. §1). To expose this fallacy, Moore noted that for any genuine definition of a moral term like 'good' by some naturalistic term, N, the question, 'Entity E has property N, but is it good?', should be a closed question, that is, a question the answer to which anyone would know to be affirmative simply on the basis of understanding it. But he pointed out with a great deal of plausibility that all such questions were 'open questions', indicating that no such proposed definitions were genuine.

2 Searle's challenge

In 1964, John Searle published what became a much discussed paper, 'How to Derive "Ought" from "Is"', in which he challenges Hume's law by offering a deduction of an 'ought' statement from a set of factual statements. Here are the basic steps in his alleged deduction:

(1) Jones uttered the words, ‘I hereby promise to pay you, Smith, five dollars.’
(2) Jones promised to pay Smith five dollars.
(3) Jones placed himself under (undertook) an obligation to pay Smith five dollars.
(4) Jones is under an obligation to pay Smith five dollars.
(5) Jones ought to pay Smith five dollars.

Searle argues that philosophers have generally failed to recognize that in addition to ‘brute facts’, such as ‘John is six feet tall’, there are ‘institutional facts’, facts that presuppose certain institutions like marriage, promises and private property. These institutions are defined by systems of rules - constitutive rules - and some of these rules, like the ones characteristic of the institution of promising, involve obligations, permissions and responsibilities. Searle’s deduction, then, is supposed to work by appealing to an institutional fact (Jones made a promise), which, given the constitutive rules that define what a promise is, allows one to infer something about what Jones ought to do.

Searle’s deduction works only if he is right in claiming that the full-blown argument (including the supplementary premises) does not represent or presuppose any evaluative and specifically moral claims. Searle’s critics doubt that his deduction is successful on this count.

3 Some implications of accepting Hume’s law

But what if Hume’s law is correct? What are the implications for the logic of moral argument? Here are three observations. (1) We have noted that Hume’s law prohibits any *deduction* of moral statements from factual premises, but this leaves open the possibility that moral statements can be inferred correctly from factual premises by some *nondeductive* mode of inference, perhaps inductively. (2) Moreover, even if Hume’s law is correct, the so-called ‘autonomy of ethics’ thesis does not, strictly speaking, follow. Recall that, according to this thesis, moral discourse is autonomous in the sense that moral claims cannot be deduced from a set of nonmoral premises (see Autonomy, ethical §5). But, even granting Hume’s law, it might still be possible to derive moral conclusions from nonmoral normative (and hence nonfactual) premises. For instance, some philosophers have wanted to derive substantive moral conclusions from normative claims about rationality (where there is no attempt to construe rationality claims as a type of factual claim) (see Practical reason and ethics §3). So, the truth of Hume’s law does not establish the autonomy of ethics. (3) Even if there are no bridge principles directly linking moral concepts with nonmoral concepts by means of some analytic definition, there might still be bridge principles, expressing what some have called ‘synthetic definitions’ of moral terms, patterned after the sorts of definitions that are offered in science (see Brink 1989; Boyd 1988).

These remarks make clear that Hume’s law can be strengthened in two ways: by prohibiting inductive as well as deductive inference from the factual to the moral, and by prohibiting transition from any sort of nonmoral claims, including nonmoral normative claims, to moral claims. Presumably, many defenders of Hume’s law would also want to defend a strengthened version of the law that would prohibit the possibilities just mentioned for negotiating the is/ought gap. If these philosophers are right, where does that leave things with our main question about the logic of moral argumentation?

4 Universalizability and the role of principles

Some philosophers, who accept Hume’s law in both its original and its strengthened versions, claim that the meanings of moral terms and concepts impose special consistency constraints on proper moral argumentation. One such widely discussed consistency constraint stems from the alleged *universalizability* of moral statements (see Universalism in ethics). The basic idea is that any two or more relevantly similar objects being morally evaluated must receive the same moral evaluation. Thus, if Jane’s lying to Fred on some occasion was morally permissible, then the same moral evaluation holds for any relevantly similar action. Of course, everything depends on how ‘relevantly similar’ is interpreted.

The most influential recent treatment of universalizability is owing to R.M. Hare, as presented in *Freedom and Reason* (1963). For Hare, one situation is relevantly similar to the first if it shares with the first all of those qualitative properties, P^1-P^n, that served as the basis (namely, one’s reasons) for the original judgment. For example, if Peter judges that Paul’s action was wrong because it was a case of misleading a customer about the

quality of a product in order to make a sale, then Peter is bound, on pain of logical contradiction, to judge of any other action having these features that it too is wrong, or else withdraw his original moral judgment. One immediate implication of Hare's thesis of universalizability is that moral judgments entail universal moral principles. The moral judgment that action A is wrong because it has properties P^1-P^n commits one to the moral principle, 'Any action having properties P^1-P^n is wrong'. Thus, on Hare's view, even if there are no nonmoral bridge principles that would link moral concepts to nonmoral concepts, correct moral reasoning and argumentation about the morality of specific cases nevertheless exemplify the *subsumptive* model of reasoning: particular objects of moral evaluation are subsumed under a moral principle which governs a range of relevantly similar cases.

Hare's universalism, and in particular his use of the thesis of universalizability, is apparently open to certain counterexamples. If, for example, one judges that some particular action performed on some particular occasion was wrong because in that situation properties P^1, P^2 and P^3 were present, then even if in some new situation the same properties are present, there may be some other property which, in the new situation but not in the original situation, defeats the reasons in favour of judging the action to be wrong. Of course, the defender of universalizability will reply that the two cases are not relevantly similar because in addition to all of those features that count in favour of some particular moral evaluation, we must also recognize any features that count against that same evaluation. The second case, then, is not relevantly similar to the first one and we have no violation of the principle of universalizability. Thus, the defender of universalizability is moved to broaden the universalizability base (and hence the account of relevant similarity) to include not only those properties taken as reasons in favour of a particular moral evaluation, but also those properties taken as reasons against a particular moral evaluation. But this manoeuvre is not entirely satisfactory since, in addition to both positive and negative reasons regarding some moral evaluation of a situation, there are all sorts of potentially relevant defeaters whose absence is a condition of the correct moral evaluation about the particular case. For example, if a man stops to help a woman whose car has broken down, our reasons for moral approval include obvious facts about the situation (one person needing help, another stopping to render aid, and so on). But there are all sorts of possible defeaters whose absence is required for our positive moral evaluation of the man's actions, including, for instance, that his motive is not to seduce or in any way harm the woman. So in addition to properties that figure among a person's reasons there are other relevant aspects including the absence of certain defeaters that affect moral judgments about particular cases. And so the defender of universalizability seems forced to broaden the universalizability base even further. But once one begins to broaden the universalizability base in this way so that any aspect of a situation (including both properties instantiated in the situation and those potential defeaters that are not instantiated) counts as part of what makes two situations relevantly similar, what started out as an interesting thesis yielding interesting moral principles approaches triviality. That is, broadening the universalizability base in this way tends to reduce the thesis to the claim that any two exactly similar objects of moral evaluation must be judged in the same way. But then, the sorts of moral principles yielded by the thesis of universalizability tend to become less useful in constraining moral argumentation - they do not apply to a very interesting range of cases.

But apart from problems for Hare's notion of universalizability, there remains the idea that correct moral reasoning proceeds by way of subsumption under moral principles. As we shall see, however, the subsumption model is not without its problems. To understand the problems we need to say a bit more about moral principles.

5 Basic moral principles and the universality of reasons thesis

Many moral philosophers have supposed that underlying all right and wrong actions is some nonmoral feature or property (or perhaps small set of them) whose presence makes all right actions right and whose absence makes all wrong actions wrong. One traditional aim of moral theory has been to discover such features or properties, which can then be expressed as basic moral principles - principles that tell us what, ultimately, makes an action right or wrong. Some philosophers have been *ethical monists*: they suppose that there is some single ultimate feature whose presence makes right actions right and whose absence makes wrong actions wrong. For example, according to hedonistic versions of act utilitarianism, facts about how much pleasure and pain would be experienced by persons (and perhaps other sentient creatures) were some action to be performed determine in every case whether or not the action if morally right (see Utilitarianism; Hedonism). In contrast, *ethical pluralists*, such as W.D. Ross 1930), insist that there is no such single ultimate feature, but rather a small set of irreducible ultimate features whose presence or absence tends to make actions either right or wrong (depending on the feature in question), even though the presence or absence of any one such feature may not be decisive in determining whether or not an

action is (all things considered) right (see Moral pluralism). Despite their disagreement over the number of ultimate features, both monists and pluralists are united in accepting what can be called the *universality of reasons* thesis: there is at least one ultimate right/wrong-making feature of actions such that for it (1) it is morally relevant whenever they are present and (2) it always counts in the same way - either as a right-making feature or as a wrong-making feature. And so, for both camps, correct moral reasoning proceeds by way of subsuming particular cases calling for moral evaluation under basic moral principles.

The universality of reasons thesis is challenged by so-called moral particularists who then reject the subsumptive model of moral reasoning and argumentation.

6 Particularism, reasons holism, and the rejection of moral principles

Moral particularism rejects the universality of reasons thesis in favour of what some call *reasons holism*, according to which what counts as a reason in some particular case, and how it counts, is a matter of the details of the particular situation. Reasons holism finds intuitive support by reflecting on common examples. For instance, normally the fact that I borrowed a book from you is a reason for me to return it to you. But if I learn that you have stolen the book, then the fact that I borrowed it from you is no reason at all to return it to you; in fact it seems to be a reason against doing so.

For our purposes, the important thing to notice is that reasons holism amounts to a rejection of basic moral principles (at least a rejection of the idea that such principles are required for making sense of moral reasoning and the difference between correct and incorrect moral argumentation) since such principles, as usually interpreted, presuppose that the properties featured in them are always morally relevant and relevant in the same way. But if basic moral principles are not (or need not be) involved in correct moral argumentation, what does the particularist have to say about such argumentation? Furthermore, since in ordinary moral reasoning and argumentation we often do appeal to moral generalizations, principles or rules in defending our moral views about particular cases, how can the particularist accommodate this phenomenon?

In response to the first question, the particularist insists that correct moral reasoning and argumentation, instead of being a matter of subsuming particular cases under general principles, proceeds by examining the details of particular cases and making judgments based on one’s moral discernment about the case at hand. The particularist accepts the idea that relevantly similar cases must be judged in the same way, but insists (contra universalists) that because of the potential complexity of particular cases and the fact that what counts as a reason is a contextually sensitive matter, there are no useful moral principles that force one, on pain of logical contradiction, to come to some particular moral conclusion about the case in hand. Rather, instead of relying on moral principles to guide proper moral reasoning and argumentation, the particularist insists on the use of trained moral perception which allows one to discern the morally relevant features in a particular case and then assess their bearing on the morality of the case. In response to the second question, particularists can allow that a moral principle can serve as a useful heuristic, reminding us of the sort of importance a property can have in particular circumstances. But as mere heuristic devices, moral principles are not cast in the role of premises in moral arguments guiding us to make judgments about present cases that are consistent with past moral judgments (see Moral particularism).

The particularist mounts a fairly strong case against certain allegedly ultimate properties conforming to the universality of reasons thesis, as the borrowed book example and other such examples make clear. However, it is harder to imagine cases in which the fact that someone’s action would be a case of knowingly causing the death of a normal human being is not morally relevant or indeed counts in favour of performing the act. The same can be said for actions in which one knowingly causes others to experience significant pain. If this speculation is correct, then perhaps universalism can survive the objections of the particularist.

The ethical naturalists and some anti-naturalists like Hare agree that correct moral reasoning and argumentation is subject to constraints of logical consistency imposed on such reasoning and argumentation by universal moral principles. Thus, both subscribe to a subsumptive model of moral reasoning and argumentation, according to which proper judgments about particular cases are subsumed under universal moral principles. The main difference between Hare and the naturalists is over the status of such principles. In sharp contrast to the universalists, the particularists reject the universalist assumption about moral reasons and consequently reject the subsumptive model of moral reasoning and argumentation. Instead, the particularist argues that correct moral reasoning and

argumentation can rely on trained moral perception of the morally relevant details of particular cases as a way of coming to have correct or at least justified moral beliefs about particular cases.

See also: Analytic ethics; Deontic logic; Prescriptivism

MARK TIMMONS

References and further reading

Altham, J.E.J. (1985) 'Wicked Promises', in I. Hacking (ed.) *Exercises in Analysis*, Cambridge: Cambridge University Press, 1-21.(An example of opposition to Searle's purported derivation of 'ought' from 'is'.)

Boyd, R. (1988) 'How to be a Moral Realist', in G. Sayre-McCord (ed.) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press.(A fairly advanced discussion and defence of ethical naturalism.)

Brink, D.O. (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.(A defender of ethical naturalism who construes moral principles as synthetic necessary truths.)

Dancy, J. (1993) *Moral Reasons*, Oxford: Blackwell, chaps 4-7.(A critic of ethical universalism (generalism) and a defender of particularism.)

Hare, R.M. (1963) *Freedom and Reason*, Oxford: Oxford University Press.(A defender of ethical universalism.)

Hudson, W.D. (1969) *The Is/Ought Question*, London: Macmillan.(A collection of important essays on the is/ought controversy.)

Horgan, T. and Timmons, M. (1992) 'Troubles for New Wave Moral Semantics: The ''Open Question Argument" Revised', *Philosophical Papers* 21: 153-29.(The authors are critical of the sort of semantic view needed to support the version of ethical naturalism defended by Brink and Boyd.)

Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978.(Contains the original statement of 'Hume's law' - the claim that it is not possible to derive an 'ought' statement from a set of 'is' statements.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press, chaps 1-2.(Contains Moore's attack on ethical naturalism.)

Prior, A.N. (1949) *Logic and the Basis of Ethics*, Oxford: Oxford University Press.(Contains a historical survey of various attempts to derive 'ought' statements from factual statements as well as a discussion of the naturalistic fallacy.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Oxford University Press, ch. 2.(Contains a defence of ethical pluralism.)

Searle, J. (1964) 'How to Derive "Ought" from "Is"', *Philosophical Review* 73: 43-58; repr. in W.D. Hudson (ed.) *The Is/Ought Question*, London: Macmillan, 1969.(Contains Searle's attempt to derive 'ought' statements from 'is' statements.)

Logic, ancient

Western antiquity produced two great bodies of logical theory - those of Aristotle and the Stoics. Both aim to explain what distinguishes good arguments from bad. Both see that the best arguments are valid and that an argument's validity depends on its form. For both, therefore, logic's business is to identify the valid argument forms. Both theories do this by laying down a small number of basic argument forms - Aristotle's 'perfect syllogisms', the Stoics' 'indemonstrables' - and rigorously deriving other valid forms from them. Both theories also try - though in a less systematic manner - to classify the ways in which an argument can go wrong.

Here the similarities between these two logics end. Their most significant differences can be illustrated by comparing basic argument forms from each. The argument 'Every swan is an animal and every animal is moving, so every swan is moving' has the same form as the argument 'Every musician is human and every human is a substance, so every musician is a substance'. The Aristotelian expression of this form is 'A belongs to all B and B belongs to all C, so A belongs to all C'. In this form the letters 'A', 'B' and 'C' stand for any terms whatever, and 'A belongs to all B' replaces 'Every B is an A'. This represents the Aristotelian approach. Compare it with the following. The argument 'If it is day then it is light, it is day, so it is light' has the same form as the argument 'If Dion walks then Dion moves, Dion walks, so Dion moves'. This form is expressed by the Stoics as 'If the first then the second, the first, so the second'. Here the expressions 'the first' and 'the second' stand for any declarative sentences whatever.

In both cases, the validity of the argument form is tantamount to the validity of all arguments having that form (though the Stoics, unlike Aristotle, require that the precise words used in an argument should recur in its form). But the Aristotelian argument form is different in kind from the Stoic one: while it abstracts from terms, the Stoic form abstracts from sentences. Aristotelian logic is a term logic, Stoic logic a sentential one.

1 Aristotle's logic: general

Aristotle was the first great logician (see Aristotle §§4-5, 7). His logical writings are traditionally grouped together as the Organon, comprising *Categories*, *De interpretatione*, *Prior Analytics*, *Posterior Analytics*, *Topics* and *On Sophistical Refutations*.

Categories and *Topics* share a concern with distinguishing different types of predication. In *Categories* beings are classified into substances, quantities, qualities, relatives, places, times, positions, states, actions and affections. Within any one category, beings are related to one another in tree-structures linked by the 'said of' relation; for example, in the category of substance, animal is said of swan and of horse. Between categories there are relations of 'inherence': for example, colour (a quality) inheres in body (a substance). The 'said of' and inherence relations underlie all true predications. Substances are either 'primary' (for example, this horse) or 'secondary' (for example, horse). Among all beings, primary substances are basic: all others are said of or inhere in primary substances, whereas primary substances neither are said of nor inhere in anything. Though *Categories* is concerned mainly with ontological questions, it also states a few principles of inference, including the following: 'When one thing is predicated of another as subject, everything said of the predicate will also be predicated of the subject'.

Topics divides predications into four types: (1) essential and convertible, (2) essential and non-convertible, (3) non-essential and convertible, (4) non-essential and non-convertible. Any predicate thus stands to its subject as (1) a definition (for example, 'biped land-animal' is the definition of man), (2) a 'genus' or 'differentia' (for example, colour is the genus of white, biped is the differentia of man among land-animals), (3) a 'peculiarity' (for example, 'walking in the gymnasium' would be a peculiarity of a particular man if he were the sole walker there) or (4) an 'accident' (for example, evil is an accident of chance). These four types of predication are traditionally known as the 'predicables', and *Topics* formulates a large number of principles (or, commonplaces, 'topoi') concerning them, for example, 'the genus is predicated essentially of whatever the species is predicated of' (IV 2).

The ninth book of *Topics* is known by the separate title *On Sophistical Refutations*. Aristotle there distinguishes fallacies that depend on linguistic factors (such as lexical or syntactic ambiguity) from those that do not.

Fallacies not depending on linguistic factors are seven. The fallacy of 'accident' attributes to a thing what belongs only to an accident of the thing, for example, supposing Coriscus happens to be approaching, it would be wrong to

infer that you do not know Coriscus, given the premise that you do not know who is approaching. The fallacy of '*secundum quid*' argues from a qualified to an unqualified statement, for example, supposing you have a belief about a non-being, it would be wrong to infer that a non-being *is*, given the premise that a non-being is-an-object-of-belief. The fallacy of '*ignoratio elenchi*' argues to the wrong point, for example, by showing that something is double in height when required to show that it is double in width. The fallacy of '*petitio principii*' or begging the question is committed when someone 'tries to prove by means of itself what is not known by means of itself' (*Prior Analytics* II 16). The fallacy of 'affirming the consequent' is committed when the direction of implication is confused, for example, by confusing the statement that whatever is created has a beginning with the statement that whatever has a beginning is created. The fallacy of 'false cause' arises from the inclusion of redundant material in a *reductio* argument, for example, arguing that the diagonal is incommensurable with the side of a square because if it were commensurable, *and if traversing a distance entailed first traversing half the distance*, then we would be committed to the absurd statement that motion is impossible. The fallacy of 'many questions' occurs when what in fact are several questions are treated as if they were just one. (See Fallacies.)

De interpretatione contains a semantic theory of categorical statement-forms. *Prior Analytics* contains the theory of the syllogism. *Posterior Analytics* applies the syllogistic to the special case where a syllogism embodies scientific knowledge. It deals with the orderly exposition of a body of scientific knowledge in an axiomatic deductive system: the axioms of such a system will state how things are in themselves and necessarily, and the deductions' premises will explain the truth of their conclusions.

2 Aristotle's semantics

Non-modal propositions are of four forms: '*A* belongs to all *B*', '*A* belongs to no *B*' (these two are 'universal'), '*A* belongs to some *B*' and '*A* does not belong to some *B*' (these two are 'particular'). The letters '*A*' and '*B*' stand for terms, that is, nouns or noun phrases. '*A*' is predicate and '*B*' subject in these forms. The first and third forms are 'affirmative', the second and fourth 'negative' (*Prior Analytics* I 2).

Semantics for universal propositions are given in the so-called '*dictum de omni et nullo*': 'We use the expression ''predicated of all'' when none of the subject can be taken of which the other term cannot be said, and we use ''predicated of none'' likewise' (I 1).

The universal affirmative and particular negative form a pair of 'contradictories', as do the universal negative and particular affirmative. The universal affirmative and universal negative are 'contraries'. Contradictories cannot both be true, nor can they both be false; contraries cannot both be true but may both be false (*De interpretatione* 7).

'*A* belongs to no (or some) *B*' implies - Aristotle says 'converts to' - '*B* belongs to no (or some) *A*'. '*A* belongs to all *B*' converts to '*B* belongs to some *A*' (*Prior Analytics* I 2). This can be so only if the universal affirmative has 'existential import', that is, implies that something is *A* and something is *B*.

Modal propositions are of the forms '*A* belongs necessarily (or contingently) to every *B*', '*A* belongs necessarily (or contingently) to no *B*', '*A* belongs necessarily (or contingently) to some *B*' and '*A* does not belong necessarily (or contingently) to some *B*'. Aristotle does not state truth-conditions for modal propositions, but he does define the 'contingent' (*endechomenon*) as 'that which is not necessary but, being assumed, results in nothing impossible'. He also distinguishes two senses in which *A* might be said to belong contingently to all *B*: either *A* is contingent for everything to which *B* belongs or *A* is contingent for everything for which *B* is contingent (I 13).

'*A* belongs necessarily to no (or some) *B*' converts to '*B* belongs necessarily to no (or some) *A*'. '*A* belongs contingently to some *B*' converts to '*B* belongs contingently to some *A*'; but '*A* belongs contingently to no *B*' does not convert (I 3). Contingency modals are subject to the following distinctive laws: '*A* belongs contingently to all *B*' is equivalent to '*A* belongs contingently to no *B*', and analogously for particulars (I 13).

3 Aristotle's syllogistic

Aristotle defines a 'syllogism' as 'discourse in which, certain things being proposed, something other than them results of necessity from their being so' (I 1). Syllogisms have at least two premises. In a two-premised syllogism, one term (the 'middle') occurs in both premises. Any two-premised syllogism is in one of three 'figures'

depending on whether the middle term is (1) once predicate and once subject or (2) twice predicate or (3) twice subject. The conclusion's predicate is the 'major' term, and its subject the 'minor'. The premises are designated as major or minor depending on which of the conclusion's terms occurs in them.

The non-modal syllogistic is based on four 'perfect' or complete syllogisms in the first figure. These (with their medieval names) are 'Barbara' (*A* belongs to all *B* and *B* to all *C*, so *A* belongs to all *C*), 'Celarent' (*A* belongs to no *B* and *B* to all *C*, so *A* belongs to no *C*), 'Darii' (*A* belongs to all *B* and *B* to some *C*, so *A* belongs to some *C*) and 'Ferio' (*A* belongs to no *B* and *B* to some *C*, so *A* does not belong to some *C*).

Modal syllogisms may be pure (with premises and conclusion all containing the same modality) or mixed. Pure modal syllogisms all derive from pure perfect first-figure syllogisms containing either the mode of necessity or that of contingency throughout. Mixed modal syllogisms all derive from mixed perfect first-figure syllogisms whose major premise and conclusion both contain the mode of necessity (or that of contingency), but whose minor premise is non-modal. Thus, Aristotle counts as perfect, not only the four non-modal syllogisms of figure one, but four pure necessity-syllogisms, four pure contingency-syllogisms, four mixed necessity-syllogisms and four mixed contingency-syllogisms - a total of twenty. Not perfect (or even valid) are those first-figure inferences in which the major premise is non-modal while the minor premise and conclusion contain an identical modality.

The perfect syllogisms function as axioms from which other syllogisms are derived (Aristotle says other syllogisms are 'reduced' to the perfect ones). Aristotle's syllogistic is thus the first axiomatized system in history (predating Euclid).

An incomplete syllogism may be perfected in one of three ways. First, by 'direct reduction':

> In the second figure, if the negative proposition is necessary, then the conclusion will be necessary.… Let *A* be possible for no *B*, and simply belong to all *C*. Then, since the negative converts, *B* is possible for no *A*. But *A* belongs to all *C*; so *B* is not possible for any *C*.
>
> (I 10)

Here the first premise is converted, thus producing a pair of premises from which the desired conclusion follows by virtue of a 'Celarent' with necessary major and non-modal minor. Second, by 'indirect reduction':

> If *A* belongs to all *C* and *B* does not belong to some *C*, it is necessary that *B* does not belong to some *A*. For if *B* belongs to all *A* and *A* to all *C*, *B* belongs to all *C*; but we supposed it did not belong.
>
> (I 6)

Here, by taking the conclusion's contradictory together with the first premise, we get the second premise's contradictory by virtue of 'Barbara'. Third, by 'ecthesis':

> when both *A* and *B* belong to all *C*, *A* will belong necessarily to some *B*.… If *A* and *B* both belong to all *C*, and we take one of the *C*s, say *D*, *A* and *B* will belong to this, so *A* belongs to some *B*.
>
> (I 6)

Here the inference's validity is explained by reference to an individual, *D*, which is supposed to be both an *A* and a *B*: there must be such an individual if *A* and *B* both belong to all *C*, but if there is such an individual then *A* belongs to some *B*.

In Book II Aristotle works out some metatheorems for his syllogistic, including the following:

(1) A syllogism with true premises must have a true conclusion.
(2) Any most general form of syllogism may have (a) all premises true, (b) some premises true and some false, (c) all premises false and the conclusion true.
(3) The falsity of all (or some) of the premises of a syllogism stated in its most general form does not necessitate the truth of the conclusion.
(4) No syllogism stated in its most general form commits a logical *petitio principii*, that is, none has the form '*p*, *q* so *r*' where *p* implies *r*. This result does not hold for all syllogistic forms: '*A* belongs to all *B*, *B* belongs to all *B*, so *A* belongs to all *B*' commits a logical *petitio*.
(5) No syllogism by virtue of its most general form commits an epistemological *petitio* in the sense that knowing

or believing premises of that form implies knowing or believing the corresponding conclusion.

4 Theophrastus

Aristotle's successor Theophrastus also wrote on logic. None of his logical writings survives, but we have reports in ancient sources, notably Alexander of Aphrodisias' commentary on Aristotle's *Prior Analytics*. Theophrastus further developed his master's syllogistic in five main ways. First, he pointed out that first-figure syllogisms may have indirect conclusions, for example, from the premises '*A* belongs to all *B* and *B* to all *C*' we can deduce '*C* belongs to some *A*'.

Second, in modal semantics Theophrastus rejected Aristotle's definition of the contingent as two-way possibility, and consequently rejected the equivalence between '*A* contingently belongs to all *B*' and '*A* contingently belongs to no *B*'. He also took '*A* contingently belongs to no *B*' to convert to '*B* contingently belongs to no *A*'.

Third, he considered Aristotle's perfect mixed necessity-syllogisms invalid, requiring instead that the modality of the conclusion be no stronger than the weakest modality in the premises.

Fourth, with regard to 'prosleptic' propositions such as '*A* belongs to all of that to none of which *B* belongs' (briefly discussed by Aristotle), Theophrastus considered them to be equivalent to ordinary categoricals.

Fifth, Theophrastus recognized 'wholly hypothetical' syllogisms such as 'If *A* then *B*, if *B* then *C*, so if *A* then *C*', and arranged them in three figures according to the position of their 'middle'. The letters '*A*', '*B*' and '*C*' here stand for sentences rather than terms.

5 Stoic logic and semantics

Among the Stoics and their predecessors, the most famous logicians were Diodorus Cronus, Philo (the Dialectician) and Chrysippus. We have only a few fragments of their work, plus reports in Sextus Empiricus' *Outlines of Pyrrhonism* and *Against the Logicians*, Diogenes Laertius' *Lives of the Philosophers* and other ancient sources.

The Stoics distinguished four causes of argumentative failure: 'incoherence', as in the argument 'If it is day it is light, wheat is being sold in the market, so Dion is walking'; 'redundancy', as in the argument 'If it is day it is light, it is day, Dion is walking, so it is light'; 'formal invalidity', as in 'If it is day it is light, it is not day, so it is not light'; and 'deficiency', as in 'Either wealth is good or bad, it is not good, so it is bad', which is said to be deficient because the first premise omits mention of the possibility that wealth is neither good nor bad. In the first three cases (but not the fourth) the argument as a whole does not instantiate a valid form, though only the first and third are cases of invalidity.

They also studied paradoxes, including the liar paradox ('If you say that you are lying, and speak the truth, then you are lying'), 'the hooded man' (which, like Aristotle's fallacy of accident, concerns an unrecognized friend concealed beneath a hood), the 'sorites' or heap (which asks how many grains it takes to make a heap; see Vagueness §2) and 'the horned man' ('What you haven't lost you still have, you haven't lost horns, so you still have horns' - a relative of Aristotle's fallacy of many questions).

Stoic semantics and syllogistic is a propositional logic, dealing with the ways in which one proposition may be formed from one or more propositional components. The forms of propositional composition on which the Stoics concentrated are negation, conjunction, disjunction and the conditional. They recognized that a conjunction 'the first *and* the second' is true if and only if both conjuncts are true. A disjunction 'either the first *or* the second' they took in the exclusive sense, requiring that one and only one disjunct be true.

The semantics of conditionals was hotly debated by the Stoics and other contemporary logicians. According to Philo, a conditional proposition ('If the first then the second') is true if and only if it does not have a true antecedent and a false consequent. According to Diodorus Cronus, a conditional is true if and only if it neither is nor ever was possible for the antecedent to be true and the consequent false. A third semantics for conditionals - 'by connection' - proposed that the antecedent be 'incompatible' with the denial of the consequent. A fourth - 'by suggestion' - proposed, somewhat vaguely, that the consequent be 'potentially in' the antecedent. (See Philo the Dialectician §2; Diodorus Cronus §§3-4; Consequence, conceptions of §2.)

6 Stoic syllogistic

There are five 'indemonstrables':

(1) If the first then the second, the first, so the second.
(2) If the first then the second, not the second, so not the first.
(3) Not both the first and the second, the first, so not the second.
(4) Either the first or the second, the first, so not the second.
(5) Either the first or the second, but not the second, so the first.

Further inferences were shown valid by the application of one or more rules to the indemonstrables. There were four of these rules, of which only two are now known. The first is equivalent to Aristotle's rule of indirect reduction: if two propositions imply a third, then either of these together with the negation of the third implies the negation of the remaining one. The third rule is the one which underlies Aristotle's procedure of direct reduction: if two propositions imply a third then the first, together with any propositions that imply the second, implies the third.

The third rule is used in proving the inference 'If the first and the second then the third, not the third, but the first, so not the second'. By the second indemonstrable we have 'If the first and the second then the third, not the third, so not both the first and the second'. Applying the third rule to this and the third indemonstrable we get 'If the first and the second then the third, not the third, but the first, so not the second'.

Both the first and third rules are used in the following (modern) proof of the validity of the Stoic inference 'If you know you are dead then you are not dead, if you know you are dead then you are dead, so you do not know you are dead', whose form is 'If the first then not the second, if the first then the second, so not the first'. Applying the first rule to the first indemonstrable we get (a) 'Not the second, the first, so not if the first then the second'. But, again by the first indemonstrable, we have (b) 'If the first then not the second, the first, so not the second'. Applying the third rule to (a) and (b) we get (c) 'If the first then not the second, the first, so not if the first then the second'. And applying the first rule to (c) we get 'If the first then not the second, if the first then the second, so not the first'.

The Stoics accepted the inference 'If the first then if the first then the second, but the first, so the second'. This is said to be validated by two applications of the first indemonstrable. But strictly speaking - and the Stoics were fond of speaking strictly - all that is validated by that procedure is the inference 'If the first then if the first then the second, but the first, *and the first*, so the second'. What is needed to validate the target inference is a 'rule of contraction': if an inference with a repeated premise is valid then so is the same inference without the repetition. This rule is also tacitly used in the above proof of the 'knowing you are dead' syllogism. The rule of contraction, though required by the Stoic system, is nowhere stated in the surviving sources.

The Stoics recognized that if an argument is valid then the conditional whose antecedent is the conjunction of the argument's premises and whose consequent is its conclusion is necessary. Thus, the validity of the first indemonstrable implies the necessity of the conditional 'If ''the first and if the first then the second'' then the second'.

Some parts of Aristotelian metatheory - namely metatheorems (2c), (3) and (5) in §3 above - do not carry over to the system of indemonstrables. (2c) Given Philonian semantics for the conditional, not both premises of the first indemonstrable can be false: if 'the first' is false then 'If the first then the second' is true. (3) Consequently, given Philonian semantics for the conditional, if both premises of the first indemonstrable are false, the conclusion must be true. (5) The first indemonstrable is epistemically question-begging on the Philonian reading of the conditional: the premise 'If the first then the second' on this reading could be known only if 'the first' were known to be false (which it could not seeing it is known to be true) or 'the second' were known to be true. (But none of this is the case for non-Philonian readings.)

7 Modern interpretations

The 'traditional' logic books of the nineteenth and early twentieth centuries reduced categorical and hypothetical syllogistic to rules of thumb and deprived it of its systematicity and rigour. The categorical syllogism was treated

as an inference, exemplified by ‘All men are mortal, Socrates is a man, so Socrates is mortal’, an example nowhere to be found in *Prior Analytics*. The modal syllogism, if mentioned at all, was dismissed as useless or incoherent.

Łukasiewicz (1951) was the first to interpret Aristotle’s syllogistic as a formalized axiomatic system, rigorous by the highest standards. He returned to Aristotle’s text, and interpreted the syllogism not as an inference but as a conditional statement whose antecedent conjoins two universal or particular (not singular) statements. He was thereby able to treat the syllogistic as an axiomatic system whose theses have the structure of conditional statements with conjunctive antecedents. Because of this structure, the system required an ‘auxiliary theory’ of propositional logic, specifying the logic of conditionals, conjunctions and negation. Practically none of this auxiliary theory is to be found in Aristotle. From Łukasiewicz’s point of view the Stoics therefore marked an advance on Aristotle in articulating a propositional logic.

Corcoran (1972) rejected Łukasiewicz’s interpretation of Aristotelian syllogistic as an axiomatic system whose theses are conditional statement-forms; instead he treated it as a natural-deduction system, thus avoiding the need for an auxiliary propositional logic. An Aristotelian syllogism is for him not just a set of premises and a conclusion but something with a deductive structure. This interpretation enabled him to give modern formal representations not only of Aristotle’s syllogisms but also of Aristotle’s own manner of proving them. He could view Aristotle’s system ‘from the standpoint of modern logic’ without imputing gross oversights to him.

Smiley’s independent work arrived at these same results, but went further than Corcoran in abandoning the aim that every expressible, semantically valid formula should be a thesis of the system. He did this on two grounds. First, Aristotelian syllogistic is not governed by the ‘rule of expansion’ (if certain premises imply a conclusion then any set that includes those premises implies the conclusion). Second, it does not obey the rule of contraction (see §6). The addition of either of these rules to Aristotle’s system, Smiley showed (1973), leads to un-Aristotelian results. The rule of expansion entails that from two mutually contradictory premises every conclusion follows. Thus ‘*A* belongs to every *B*, *A* does not belong to every *B*, so *C* belongs to some *D*’ should be a syllogism; but it is not. Since ‘*A* belongs to all *C*, *A* belongs to all *C*, so *A* belongs to some *A*’ is a syllogism, contraction would entitle us to drop the repetition and say that ‘*A* belongs to all *C*’ implies ‘*A* belongs to some *A*’. So ‘*A* belongs to no *A*’ would imply ‘*A* does not belong to all *C*’. Therefore the premises ‘*B* belongs to no *A*, *B* belongs to all *A*’ - which imply ‘*A* belongs to no *A*’ - would also imply ‘*A* does not belong to all *C*’. However, ‘*B* belongs to no *A*, *B* belongs to all *A*, so *A* does not belong to all *C*’ is not an Aristotelian syllogism. On Smiley’s interpretation, a syllogism is a deduction in a sense of ‘deduction’ that requires all premises to be used, and which heeds the number of times a premise is used. Smiley constructed a formal system based on such a notion of deduction, proved its completeness, and constructed a decision-procedure for it. These results allowed him to formalize some of Aristotle’s own metatheory, in particular metatheorem (3) (see §3 above).

Thom (1981, 1996) further developed the Smiley interpretation, showing that Aristotle’s syllogistic does not include proper substitutions in theses: thus, while ‘*A* belongs to all *B*, *B* belongs to all *C*, so *A* belongs to all *C*’ is Aristotelian, ‘*A* belongs to all *B*, *B* belongs to all *A*, so *A* belongs to *A*’ is not. This restriction is needed if the interpreter is to construct a formalization to which all of Aristotle’s metatheory applies, since some of that metatheory holds only for theses stated in their most general form (see (2)-(5) in §3 above). He also showed how to develop Aristotle’s sketchy remarks about ‘ecthesis’ into a natural deduction system of singular syllogisms, on which the whole syllogistic (modal as well as non-modal) can be based.

Modern interpretations of modal syllogistic take their impetus from McCall (1963), who restored faith in the rigour of Aristotle’s modal system by formalizing it axiomatically. Johnson (1989) provided a formal semantics for McCall’s formalization. Johnson’s semantics treats ‘*A* belongs necessarily to every (or no) *B*’ as true if and only if the *B*s are among (or exclude) the *essential A*s; ‘*A* belongs necessarily to some *B* (or necessarily does not belong to some *B*)’ he treats as true if and only if some essential *B* is an essential *A* (or is an essential non-*A*). If, as Johnson required, the truth-conditions of Aristotle’s modal sentences make reference to the essential *A*s, then it seems that modal syllogistic is underpinned by some doctrine of essentialism. This idea was developed by Patterson (1995), who explored philosophical links between the semantics of modal sentences and the doctrine of the predicables.

Modern interpretations of Stoic logic spring from the work of Mates (1953), and Kneale and Kneale (1962). The complex relationship between the Stoic and Aristotelian systems has been studied by Frede (1987).

See also: Logical and mathematical terms, glossary of; Stoicism §11

PAUL THOM

References and further reading

Alexander of Aphrodisias (*c*.200) *In Aristotelis Analyticorum Priorum Librum I Commentarium* (Greek text), ed. M. Wallies, Berlin: Reimer, 1883; *On Aristotle's Prior Analytics 1.1-7*, trans. J. Barnes, S. Bobzien, K. Flannery and K. Ierodiakonou, London: Duckworth, 1991.(Includes much material on the Stoics and reports of Theophrastus' logical theories as discussed in §4, but not his discussion of modal syllogisms; excellent bibliography.)

Aristotle (*c*.mid 4th century BC) *The Complete Works of Aristotle: The Revised Oxford Translation*, vol. 1, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984.(See *Categories* for the theory of categories, *Posterior Analytics* for scientific syllogisms, *On Sophistical Refutations* and *Topics* for predicables and fallacies, all discussed in §1; *De interpretatione* for semantics, as discussed in §§1-2; and *Prior Analytics* for syllogistic (Book I) and metatheory (Book II) as discussed in §§2-3.)

Barnes, J. (1984) 'Terms and Sentences: Theophrastus on Hypothetical Syllogisms, Dawes Hicks Lecture on Philosophy 1983', *Proceedings of the British Academy* 69: 279-326.(Further non-technical material on Theophrastus.)

Corcoran, J. (1972) 'Ancient Logic and its Modern Interpretations', in J. Corcoran (ed.) *Proceedings of the Buffalo Symposium on Modernist Interpretations of Ancient Logic, 21 and 22 April, 1972*, Dordrecht: Reidel. (Partly technical treatment of Aristotelian syllogistic as a natural deduction system.)

Diogenes Laertius (early 3rd century) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols; revised H.S. Long, 1972.(Greek text with facing English translation. Volume 2 includes material on the Stoics (book VII) and a life of Chrysippus.)

Frede, M. (1987) 'Stoic vs. Aristotelian Syllogistic', in *Essays in Ancient Philosophy*, Oxford: Clarendon Press. (Non-technical discussion.)

Johnson, F. (1989) 'Models for Modal Syllogisms', *Notre Dame Journal of Formal Logic* 30 (2): 271- 84. (Technical treatment; referred to in §7.)

Kneale, W. and Kneale, M. (1962) *The Development of Logic*, Oxford: Clarendon Press.(Good introduction to the subject, including an extensive bibliography.)

Lear, J. (1980) *Aristotle and Logical Theory*, Cambridge: Cambridge University Press.(Further non-technical reading on some of the topics discussed in §3.)

Lejewski, C. (1976) 'On Prosleptic Premisses', *Notre Dame Journal of Formal Logic* 17 (1): 1-18. (Technical discussion of prosleptic propositions, mentioned in §4.)

Łukasiewicz, J. (1951) *Aristotle's Syllogistic from the Standpoint of Modern Formal Logic*, Oxford: Oxford University Press; 2nd enlarged edn, 1957.(Partly technical treatment; the first to interpret Aristotle's syllogistic as a rigorous, formalized axiomatic system.)

Mates, B. (1953) *Stoic Logic*, Berkeley and Los Angeles, CA: University of California Press.(A modern interpretation with a good introduction and bibliography.)

McCall, S. (1963) *Aristotle's Modal Syllogisms*, Amsterdam: North Holland.(Axiomatic formalization of Aristotle's modal system; technical.)

Nortmann, U. (1994) 'Does Aristotle's Modal Logic Rest on Metaphysical Assumptions?', *Analyomen* 1: 115-25. (An alternative approach to those of Johnson and Patterson; partly technical.)

Patterson, R. (1995) *Aristotle's Modal Logic: Essence and Entailment in the Organon*, Cambridge: Cambridge University Press.(Non-technical treatment; referred to in §7.)

Patzig, G. (1968) *Aristotle's Theory of the Syllogism: A Logico-Philological Study of Book A of the Prior Analytics*, trans. J. Barnes, Dordrecht: Reidel.(Non-technical discussions of some of the topics mentioned in §3; good bibliography on non-modal syllogistic.)

Sextus Empiricus (*c*.200) *Outlines of Pyrrhonism*, trans. J. Annas and J. Barnes, Cambridge: Cambridge University Press, 1994.(Material on the Stoics, as discussed in §5.)

Sextus Empiricus (*c*.200) *Against the Logicians*, vol. 2, trans. R.G. Bury, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933-53.(Material on the Stoics, as discussed in §5.)

Smiley, T.J. (1973) 'What is a Syllogism?', *Journal of Philosophical Logic* 2 (1): 136-54.(Technical treatment; referred to in §6.)

Thom, P. (1981) *The Syllogism*, Munich: Philosophia.(Largely technical treatment on modern interpretations; good bibliography on non-modal syllogistic.)

Thom, P. (1996) *The Logic of Essentialism: An Interpretation of Aristotle's Modal Syllogistic*, Dordrecht: Kluwer. (Largely technical; bibliography on modal syllogistic.)

Logic, medieval

Medieval logic is crucial to the understanding of medieval philosophy, for every educated person was trained in logic, as well as in grammar, and these disciplines provided techniques of analysis and a technical vocabulary that permeate philosophical, scientific and theological writing. At the practical level, logic provided the training necessary for participation in the disputations that were a central feature of medieval instruction, and whose structure - with arguments for and against a thesis, followed by a resolution - is reflected in many written works. At the theoretical level, logic, like other subjects, involved the study of written texts through lectures and written commentaries. The core of the logic curriculum from the twelfth century onwards was provided by the logical works of Aristotle. These provided the material for the study of types of predication, the analysis of simple propositions and their relations of inference and equivalence, the analysis of modal propositions, categorical and modal syllogisms, fallacies, dialectical Topics, and scientific reasoning as captured in the demonstrative syllogism. Comprehensive as this list might seem, medieval logicians realized that other logical subjects needed to be investigated, and, again from the twelfth century onward, new techniques and new genres of writing appeared. The main new technique involved the use of 'sophismata', or puzzling cases intended to draw attention to weaknesses and difficulties in logical definitions and rules. The new genres of writing especially included works on 'supposition theory', which concerned the types of reference that the subjects and predicates of propositions have in different contexts, and works on 'syncategoremata', which concerned the effect on sense and reference produced by the presence and placing of such logical terms as 'all', 'some', 'not', 'if...then', 'except', and so on. Other important topics for investigation include 'insolubles', or semantic paradoxes, and 'consequences', or valid inference forms. These new developments were seen as providing a supplement to Aristotelian logic, rather than an alternative. The only context in which people occasionally suggested that Aristotelian logic was inapplicable was that of Trinitarian theology, and the only logician who deliberately set out to reform logic as a whole was Ramon Llull.

The study of medieval logic involves two kinds of difficulty. In the first place, few texts are available in translation, and indeed, many are not even available in printed form. In the second place, there is a problem of interpretation. For a very long time, the specifically medieval contributions to logic were ignored or despised, and when people began to take them more seriously, there was a strong tendency to look at them through the spectacles of modern formal logic. More recently, scholars have come to realize that medieval interests cannot be mapped precisely onto modern interests, and that any attempt, for example, to make a sharp distinction between propositional and quantificational logic is misleading. The first task of the modern reader is to try to understand what the medieval logician was really concerned with.

1 Background

The story of medieval logic is the story of logic in western Europe after the split of the Roman Empire into the Latin-speaking, Catholic west and the Greek-speaking, Orthodox east (see Byzantine philosophy §3). Because this split was followed by barbarian invasions and constant upheaval, very little happened between 550 and 1000, and learning was mainly kept alive in Benedictine monasteries. However, it is important to consider the Latin texts that were transmitted, for they formed part of the basis for later developments. One very important early source is the *Peri Hermeneias* by Apuleius of Madaura which dates from the second century AD. It discusses the proposition and the categorical syllogism, and it was influential until the late twelfth century. It was the main source for Martianus Capella (early fifth century), author of *De nuptiis Philologiae et Mercurii (The Marriage of Philology and Mercury)*, Cassiodorus (sixth century), author of *Institutiones*, and Isidore of Seville (seventh century), author of *Etymologies*, all of whose encyclopedic works include sections on logic (see Encyclopedists, medieval §§4, 6, 7). Other early works include Marius Victorinus on definitions and a work attributed to Augustine, the *Categoriae decem (The Ten Categories)*, which is from the fourth-century circle of Themistius and was the most intensely studied logical work in the ninth and tenth centuries.

Most important of all are the works of Boethius. He seems to have been responsible for translating all of Aristotle's Organon (that is, his six works on logic) into Latin, and all but the *Posterior Analytics* survive. He also translated Porphyry's *Isagōgē (Introduction)*, and wrote commentaries on some of Aristotle's logic, on Porphyry and on the *Topics* of Cicero. In addition he composed monographs of his own on Division, on Topics, on

categorical syllogisms and on hypothetical syllogisms, that is, on conditional propositions and arguments built up from them (see Boethius, A.M.S. §§2-3). At first, little attention was paid to this body of writing, but signs of revival first appear at Charlemagne's court, for which Alcuin (d.804) wrote a little work on dialectic (see Carolingian renaissance §2). The first outstanding figure is Gerbert of Aurillac, who taught in the cathedral school at Reims from about 972 before becoming a bishop and finally Pope. He taught the *Logica vetus (Old Logic)* (that is, Porphyry's *Isagōgē*, Aristotle's *Categories* and *On Interpretation*), Cicero's *Topics* and a good deal of Boethius.

As western Europe became more settled, so educational institutions began to flourish. Better monastic schools (for example, Bec, where Anselm taught) and cathedral schools (see Chartres, School of) developed, and in the twelfth century, especially in Paris, schools grew up around individual masters such as Abelard and Adam of Balsham [Parvipontanus]. The monographs and commentaries of Boethius had been largely recovered by the beginning of the eleventh century and, with the *Logica vetus*, became the focus of attention. The *Liber sex principiorum (Book of Six Principles)* (that is, the last six categories, about which Aristotle had had less to say), attributed to Gilbert of Poitiers, became part of the *Logica vetus*. At the same time the philosophical study of grammar became important, with a focus on the *Institutiones grammaticae (Institutions of Grammar)* of Priscian (see Language, medieval theories of §1). Anselm's important discussion of the reference of the word '*grammaticus*' ('expert in grammar' or 'literate') shows an attempt to harmonize Aristotle's and Priscian's doctrines (see Anselm §5). By far the most outstanding figure, and the first really original logician, was Peter Abelard (§§2, 4). Unfortunately, his own turbulent life and the important events of the later twelfth century combined to diminish any influence his logical work may have had.

2 After 1150: a general survey

Logic after 1150 is shaped by four interrelated developments: the recovery of the rest of Aristotle's logical works, along with other texts; the foundation of universities and the *studia* of the new religious orders; the development of 'sophismata', a new technique for handling logical problems; and the development of new areas of logic.

New translations. Aristotle's *Topics* and *Sophistical Refutations* were known by the 1130s and the entire *Logica nova*, including the *Prior* and *Posterior Analytics*, was known by 1159 when John of Salisbury referred to them in his *Metalogicon*. All circulated in translations by Boethius, with the exception of the *Posterior Analytics*, which was translated by James of Venice in the twelfth century. Both the *Topics* (see Alexander Neckham) and the *Sophistical Refutations* were used in the twelfth century, but most especially the latter, and the *Prior Analytics* gradually displaced Boethius' works on the categorical syllogism. The *Posterior Analytics* was absorbed more slowly. The first known complete commentary is that by Grosseteste (§1), dating from the early 1220s.

In the second half of the twelfth century Arabic logic began to appear. The first part of Avicenna's *Logic* (see Ibn Sina §1), dealing with the *Isagōgē*, was translated by Gundissalinus, as was Farabi's *Catalogue of Sciences*, whose second chapter dealt with logic. The *Logic* of Ghazali was also translated. In the 1230s the logic commentaries of Averroes on the *Categories*, *On Interpretation* and the *Prior* and *Posterior Analytics* were translated, though they were less successful than the Arab works translated earlier. Some Greek commentators were also translated. Themistius' *Paraphrase of the Posterior Analytics* and some extracts from Philoponus on the *Posterior Analytics* became known in the late twelfth century. Much later William of Moerbeke translated Simplicius on the *Categories* (1266) and Ammonius on *On Interpretation* (1268). The frequently cited but now lost commentary by 'Alexander' on *Sophistical Refutations* was probably the translation of a work by Michael of Ephesus (*c*.1130). Other Greek works on logic were not translated until the Renaissance. The important thing about all these texts is not just their content but the fact that they provide a full curriculum for an organized institution. (See Aristotelianism, medieval §3 for the twelfth-century reception of translations.)

New institutions. By the beginning of the thirteenth century the Universities of Oxford, Paris and Bologna were taking shape as organized institutions. Gradually other universities were founded, along with the *studia* of the new teaching orders, especially the Dominicans (at whose Cologne *studium* Albert the Great taught) and the Franciscans (at whose London *studium* Ockham taught), and all these institutions offered a training in the liberal arts (see Encyclopedists, medieval §1). Two things should be noted. First, a degree from the arts faculty was required for all advanced (graduate) study in theology, medicine and law (though especially in Italy, medicine and law, unlike theology, could be pursued at the undergraduate level). Second, logic formed a very large part of the

arts curriculum, especially during the first two years. One reason for its importance was the large place given to disputation at all levels of university teaching: logic clearly offers training in the techniques of debate, and inculcates quickness in analysing and responding to arguments. However, university teaching also involved the close study of texts, especially the texts of Aristotle, and many commentaries, as well as original textbooks, were written by medieval logicians. (See Natural philosophy, medieval §5 for teaching practices.)

Sophismata: a new technique. Medieval logic is characterized by a new technique, the analysis and solution of sophismata. The word 'sophisma' covers two phenomena. First, there is the sophisma sentence itself, which is a logical puzzle intended to introduce or illustrate a difficulty, a concept, or a general problem. The origin of these sentences seems to lie in the twelfth-century use of '*instantiae*', or counter-examples. Second, there is the technique of the sophismatic disputation, which is used to show that the very same reasoning which supported a plausible thesis could also be used to establish something implausible. By the end of the twelfth century the sophisma was established in different genres of logical and grammatical writing. These genres included treatises on syncategoremata, that is, those words which perform a function in a sentence by negating, conjoining, disjoining and so on (see §8), and treatises on sophismata themselves. The relationship between these treatises was very close. Typically, a treatise on syncategoremata would start with a syncategorematic term and then appeal to sophismata to illustrate the difficulties its use might cause. A treatise on sophismata would start with a sophisma and, using disputational techniques, appeal to facts about a syncategorematic term or to logical distinctions to solve it, or to show that different truth-values were possible according to different senses of the sophisma sentence. Various types of sophismata and treatises on sophismata survive. It seems that in Paris the sophisma was an element of teaching, and had to be debated in the schools, that is, within a fairly formal setting. Very elaborate sophismata survive from the latter part of the thirteenth century which were given solutions by the teaching masters, for example, Boethius of Dacia, and which were often used as a vehicle for the straightforward discussion of interesting logical problems. On the face of it, Oxford sophismata seem to have been quite different. They were the subject of live debate, but this debate was at a strictly undergraduate level, and was part of the practical training in logical disputation. They were not primarily a vehicle for discussing doctrine, and they seem to disappear as a genre when the curriculum changes. Fourteenth-century Oxford sophismata had two special features. First, they exhibit a new emphasis on the '*casus*', or initial hypothesis about the context of the sophisma-sentence, which might itself be identified as the source of the problems in determining the truth-value of that sentence. This notion of *casus* played a key role in treatises on insolubles (see §10) and obligations (see §11), which are both closely linked to sophismata. Second, many of them are about natural philosophy, and introduce mathematical considerations and calculatory techniques (see Oxford Calculators §§2, 4).

Among the surviving fourteenth-century treatises on sophismata are important works by Richard Kilvington, William Heytesbury, John Buridan and Albert of Saxony. In the fifteenth century, Paul of Venice wrote on sophismata, as well as using them extensively in other works. Both the technique and the genre of writing disappeared early in the sixteenth century.

3 After 1150: new logical writings

While the writings of Aristotle were central to the logic curriculum, there were matters that he did not discuss, which left room for a considerable number of new developments. The most prominent is the so-called '*logica moderna*' or '*modernorum*' (a label coined by twentieth-century authors). Also known as 'terminist logic', this includes supposition theory and its ramifications (see §7). Treatises on supposition theory deal with the reference of subject and predicate terms in propositions, and they have as a corollary the treatises on syncategoremata, which deal with all the other terms in propositions. The material dealt with in treatises on syncategoremata was also dealt with in treatises on sophismata (see §2 above), but in the fourteenth century much of it also appeared in treatises on the 'proofs of terms' (see §8). Three more important developments are found in treatises on consequences (see §9), insolubles (see §10) and obligations (see §11).

While supposition theory constitutes the most important new development, it was not always privileged. Paris in the second half of the thirteenth century, and Bologna and Erfurt at the beginning of the fourteenth century, were centres of 'modist logic', and tended not to use supposition theory. Modist logicians (for example, Peter of Auvergne) were more interested in the sense than the reference of terms, and they made much use of the relationships between the *modi essendi* (modes of being) of things, the *modi intelligendi* (modes of understanding)

and the *modi significandi* (modes of signifying) of words. They were also characterized by their lengthy discussions of the rational nature of logic and its relation to second intentions (see §4). Another rival to supposition theory as a tool of analysis was provided by 'proofs of terms', which were particularly important in the later fourteenth century.

Another new form of writing was the comprehensive textbook. At least six survive from the thirteenth century, including those by William of Sherwood, Peter of Spain and Roger Bacon. In the fourteenth century we find those by William Ockham, John Buridan and Albert of Saxony. Some universities, especially Oxford and Cambridge, preferred to use loose collections of brief treatises on various topics, and a good example of such a collection is the *Logica parva* of Paul of Venice.

There are three things to note about these new writings. First, they were not officially part of the curriculum at most universities, but that did not prevent their being used. Second, because Latin was the common language of instruction in western Europe, a text written in one place could be used anywhere else. Peter of Spain's text was generally the most popular, but Buridan was particularly important in Poland, and in late fourteenth- and fifteenth-century Italy short texts from England were heavily used. Third, there is no genuine division between the authors of Aristotle commentaries and the authors of the new texts. To take just one example, John Buridan wrote a comprehensive textbook, commentaries on parts of Aristotle's Organon, and separate treatises on consequences and sophismata, the last containing a noteworthy treatment of insolubles.

4 The nature of logic

The purpose of logic had nothing to do with the setting up of formal systems or the metalogical analysis of formal structures. Instead, it had a straightforwardly cognitive orientation. Everyone accepted the view that logic is about discriminating the true from the false by means of argument. This is why logic was essential to the speculative sciences, since it provided the instruments for finding truth and for proceeding from the known to the unknown.

The first problem with logic was what to call it. The name '*dialectica*' was most prevalent until the thirteenth century, when '*logica*' gained the upper hand. Both words have various associations. '*Dialectica*' in the broad sense just is logic, but the word also has three narrower senses: dialectic as the art of debate; dialectic as the art of finding material for arguments; and dialectic as a kind of reasoning which falls between demonstrative and sophistical reasoning. The first sense was appealed to by Roger Bacon when he said that logic as an art has to do with disputation. The second sense is associated with the discussion of Topics, the headings under which the material for arguments can be sorted. Because the study of Topics also included maxims, or self-evidently true generalizations, which could provide the warrant for different types of argument, there is a close link between Topics and argumentation. Hence, the third and most usual sense of *dialectica* had to do with topical or dialectical syllogisms as a part of logic. Medieval logicians treated Aristotle's distinction between dialectical and demonstrative syllogisms as an epistemological one concerning the status of their premises, so that dialectical syllogisms had the same formally valid structure as demonstrative syllogisms.

The problems associated with the word '*logica*' have to do with whether logic is a linguistic or a rational pursuit. Isidore of Seville noted that '*logica*' comes from the Greek word '*logos*' which can mean '*sermo*' (word) or '*ratio*' (reason). As a result logic could be called either a *scientia sermocinalis* (linguistic science) or a *scientia rationalis* (rational science). There were considerations supporting both titles. On the one hand, the Stoics had divided philosophy into natural, moral and rational, and the last was equated with logic which could then, as Boethius pointed out, be seen as both an instrument and a part of philosophy (see Stoicism). On the other hand, logic was one of the liberal arts (see Encyclopedists, medieval §1) and belonged to the *trivium*, along with rhetoric and grammar, which made it seem a linguistic science. This emphasis was intensified by the discovery of Arab logicians who included Aristotle's *Rhetoric* and *Poetics* in the Organon, or logical works, a classification accepted by Albert the Great and Aquinas, among others. Some logicians, such as William of Sherwood, preferred to call logic just a linguistic science, but many others in the thirteenth century, including Robert Kilwardby and St Bonaventure, called it both linguistic and rational.

In the late thirteenth and fourteenth centuries the notion of logic as a rational science only became predominant. This move was partly associated with the rediscovery of the *Posterior Analytics*, the new emphasis on demonstrative science, and the increased importance of the classification of science into practical and speculative,

along with the associated division of speculative science into natural philosophy, mathematics and divine science. Unlike the Stoic classification, this one offered no obvious place for logic. Some later logicians, including Ockham, Buridan and Albert of Saxony, classified logic as a practical science, but most preferred to think of it as speculative (though of course with practical applications). People then called logic a supporting science (*scientia adminiculativa*), subordinated to the three important sciences. The notion that logic was a tool directing reason in the acquisition of knowledge was one reason for calling it rational.

However, the most important reason for calling logic rational had to do with views about its subject matter, and here the influence of Avicenna is crucial, for it was Avicenna who had said that logic was about second intentions. Second intentions, often identified with beings of reason (*entia rationis*), are those higher level concepts we use to classify our concepts of things in the world, and they include such notions as genus, species, subject, predicate and syllogism. There were obvious ontological difficulties here, for one could argue that second intentions pick out special common objects including both universals and logical structures, and one could also argue that second intentions are just mental constructs reached through reflection on individual things and on actual pieces of discourse or writing. Nominalists and realists obviously disagreed on ontology, but this did not prevent such nominalists as Ockham from agreeing that logic deals with second intentions, and that the syllogism the logician considers is neither a thing in the world, nor a piece of writing or speaking. Some people preferred to say that logic was about things in the world as they fall under second intentions, and other people preferred to pick out some special second intention such as argumentation or the syllogism as the subject of logic, but there was still a strong consensus that the objects of logic are rational objects.

These remarks have to be balanced by the obvious fact that people reason in ordinary language, and that ordinary locutions are often vague, ambiguous or misleading. Medieval logicians spent an enormous amount of time on the analysis of ordinary Latin, not because they thought that logic was theoretically concerned with ordinary language, but because this was the only way to avoid fallacious reasoning. Indeed, the avoidance of fallacy is at the heart of all the new types of logical writing.

5 Syllogistic

Despite the new forms of logical writing, syllogistic remained basic, and the syllogism was generally regarded as the central form of argumentation, to which all other forms could be reduced. This view is a lot less restrictive than one might imagine, because of the general way in which 'syllogism' was defined. Following Aristotle, logicians wrote: 'A syllogism is an expression in which, when certain things have been asserted, something else must occur by means of the things which were asserted' (Peter of Spain, in Kretzmann and Stump 1988: 217). This definition encompassed modal syllogisms, syllogisms with singular terms (called 'expository syllogisms'), tensed syllogisms with past and future verbs, syllogisms with 'oblique' (or inflected) terms such as genitives (for example, 'Every man's donkeys are running, Brownie (the donkey) is a man's, therefore Brownie is running'), and syllogisms containing such exponible terms (see §8) as 'only' and 'except'.

The standard core of syllogistic included immediate inference as captured by the square of opposition, the rules of equipollence (for example, 'Not every man is running' is equivalent to 'Some man is not running'), the laws of conversion and contraposition, and the categorical syllogism itself, together with the laws of reduction which allow one to prove non-first-figure syllogisms on the basis of the first-figure syllogisms, treated as axioms, and some other rules including conversion. (For details, see Logic, ancient.) The main medieval contribution here was the provision of mnemonic devices, particularly the verse 'Barbara, Celarent,…', which accounts for nineteen valid categorical syllogisms, together with detailed rules of reduction (see Peter of Spain, in Kretzmann and Stump 1988: 224-5).

One topic of discussion was whether a fourth figure should be recognized. This question relates to the ways in which the major and minor terms can be defined. From the sixteenth century onward, it became more usual to adopt the definition of Philoponus and to say that the major term is the predicate of the conclusion, while the minor term is the subject of the conclusion. This definition gives four figures which represent four possible arrangements of the major, minor and middle terms in two differentiated premises. Alternatively, if one adopts the standard medieval definition, whereby the major term is the term that appears with the middle term in the first premise, one can still speak of four figures with differentiated premises but it turns out that each figure also has indirect modes, in which the major term is the subject of the conclusion. The normal medieval account encompassed just the direct

and indirect modes of the first figure, and the direct modes of the second and third figures. People argued that the fourth figure is just the first indirect figure with transposed premises, and they did not count this as involving a genuinely different disposition of the middle term. John Buridan seems to have been the first to realize (in rather brief remarks) that on the medieval account, one will still have four distinct figures, and that the direct modes of the first figure will count as the indirect modes of the fourth figure in just the same way as the direct modes of the fourth figure count as the indirect modes of the first.

Another subject of discussion concerned the relation of perfect and imperfect syllogisms. ‘Perfect’ syllogisms were first-figure syllogisms whose necessity was obvious, and so needed no proof in terms of simpler argument forms, though they can be said to be founded on the regulative principles *dici de omni* (which ‘occurs when nothing is to be subsumed under the subject of which the predicate is not said’) and *dici de nullo* (which ‘occurs when nothing is to be subsumed under the subject from which the predicate is not removed’: Peter of Spain, in Kretzmann and Stump 1988: 217). Links between *dici de omni* and the Topic of quantitative whole were occasionally mentioned in the thirteenth century, but the general view was that perfect syllogisms do not need to be justified by appeal to Topics. Other syllogisms, however, are imperfect, and need to be justified by the rules of reduction, which involve the use of the laws of conversion. This immediately raises the problem of circularity, for if all arguments are reducible to the syllogism, it seems to follow that the laws of conversion are themselves reducible. There was considerable uncertainty on this point, and Burley, in his early treatise on consequences, says both that everything is reducible to the syllogism and that the laws of conversion are not. One way out of the impasse was to see the syllogism itself as just one among many forms of consequence (see §9), but although this is the approach taken by the later Burley and by Buridan, followed by Albert of Saxony, it is less prevalent than has often been suggested. Nor is there any reason to suppose that Burley or Buridan regarded themselves as showing that quantificational logic is based on propositional logic when they suggested that general inferences with unanalysed parts have some kind of priority.

The laws of contraposition raised an interesting semantic issue, which was discussed by Buridan and by various other logicians up to the early sixteenth century. It was commonly held that affirmative sentences with a non-referring subject were false, and the corresponding negative sentences true. But if one accepts term-negation while holding that sentences with term-negation (and no other) are affirmative, contraposition of the sort ‘All men are beings, therefore all non-beings are non-men’ will be invalid, given that there are no non-beings. Buridan and other late medieval logicians responded by claiming that the laws of contraposition could hold only when an extra existential premise was supplied. It is important to note a corollary of their view. In classical quantificational logic, the standard interpretation of sentences with non-referring terms is that universal sentences (formulated with ‘if…then’) are true, and particular ones (formulated with ‘and’) are false. As a result, quantificational formulations of the simple inferences ‘All *A*s are *B*, therefore some *A*s are *B*’ and ‘No *A* is *B*, therefore some *A*s are not *B*’ are rejected as invalid, as are the quantificational formulations of syllogisms with universal premises and particular conclusions. When medieval logicians had to decide on a uniform way of treating sentences with non-referring terms they decided not to draw a line between universal and particular sentences, but between affirmative and negative ones. As a result, the simple inferences captured in the square of opposition, the laws of equipollence, the laws of conversion and the standard categorical syllogisms remain valid. It is only the laws of contraposition that need reformulation.

6 Compounded and divided senses; modal logic

The discussion of modal logic is in part parallel to the discussion of categorical syllogistic. In general introductions to logic the discussion of modal propositions was followed by the discussion of simple modal inferences and equivalences as captured in a square of modal opposition, modal conversion and, finally, modal syllogisms. Buridan is noteworthy for the large number of modal equivalences he considered, and Ockham for his thorough exploration of different kinds of modal syllogism. Modal logic also figures in treatises on consequences, on proofs of terms, and on compounded and divided senses. The central modal notions employed were ‘necessary’, ‘impossible’, ‘possible’ and ‘contingent’, of which the last two need some explication. In the earlier period such logicians as Peter of Spain followed Boethius in treating ‘possible’ as equivalent to ‘contingent’, but later logicians, such as Buridan, made a careful distinction between ‘possible’ as compatible with ‘necessary’ but not with ‘impossible’, and ‘contingent’ as incompatible with both ‘necessary’ and ‘impossible’.

A full understanding of modal logic requires a consideration both of the basic tool of analysis, the distinction between compounded and divided senses, and also of the general theory of modality.

The distinction between compounded and divided senses has its origin in Aristotle's discussion of the fallacy of composition and division, to which Abelard was one of the first to pay careful attention. The basic point concerns two ways of reading the sentence 'A seated man can walk'. Interpreted according to its compounded sense, this proposition is *de dicto* (about a dictum or 'that' clause - see De re/de dicto §1) and means 'That-a-seated-man-walk (that is, while seated) is possible'. Interpreted according to its divided sense, the proposition is *de re* (about a *res* or thing), and means 'A seated man has the power or ability to walk'. The proposition is false in the first sense, but true in the second. It became standard when considering modal inferences in general and modal syllogisms in particular to distinguish between the compounded and divided senses of premises and conclusion, and to work out the logical results of these different readings.

In the fourteenth century the distinction between the compounded and divided senses was very widely used, as can be seen from the work of William Heytesbury, the first to devote a complete treatise to this topic. He listed nine types of logical problems which could be solved by paying attention to the distinction. These problems were signalled by the presence of modal words but also, for example, by the presence of future tense verbs, verbs producing confused supposition (see §7), terms such as 'infinite' which could be taken both categorematically and syncategorematically (see Oxford calculators §§2, 4), and words about acts of will and intellect. This last case ties in with epistemic logic, as a branch of applied modal logic. For instance, Heytesbury cites the inference 'Your father you believe to be a donkey; therefore, you believe your father to be a donkey' (Kretzmann and Stump 1988: 426-7), and remarks that it is invalid because the antecedent is taken in the divided sense ('Your father happens to be the thing [at a distance] that you take to be a donkey': true) and the consequent is to be taken in the compounded sense ('You believe that your father is a donkey': false).

The general theory of modality underwent interesting changes during the Middle Ages. In the earlier period, there were two possible approaches. On one approach, the interpretation of modalities is straightforwardly ontological and possibility is seen as a potency, an aspect of beings and events to be discussed in the context of change and motion. A second interpretation of modalities is the so-called statistical interpretation. On this account, to say that an object or state of affairs is necessary is to say that it occurs at all times, to say that it is possible is to say that it occurs at some but not all times, and to say that it is impossible is to say that it never occurs. For both these views, to talk about modalities is just to talk about what exists or occurs at various times, and modality is not some feature of the world over and above what there is. Moreover, they are tied to the principle of the necessity of the present, that is, that if *P* is true at time *t*, then it is not possible for not-*P* to be true at time *t*. An alternative model of possibility was at first confined to purely theological contexts, but in the twelfth century such authors as Gilbert of Poitiers and Abelard began to take a new approach, and in the fourteenth century Duns Scotus worked out the full philosophical implications of the alternative model, the so-called synchronic interpretation. On this model, possibility involves reference, not just to the present state of affairs, or to the potentialities of objects, but to states of affairs or worlds which are here and now alternative to the present world. There is no necessity of the present, and the basic notion is that of conceptual consistency. This new view leads to more complete and satisfactory formulations of the relations between modal sentences, especially when they are taken in the divided sense.

7 Supposition theory

The most notable new theory that took shape in the twelfth century was supposition theory and its ramifications, particularly ampliation and restriction. In the early period, *suppositio* was taken to be a property of a substantive term, while *copulatio* was a property of an adjectival term, and *appellatio* a property of a term that referred to actual existents. Later, the term *copulatio* was dropped, and *appellatio* acquired a totally new meaning (see below). Roughly speaking, supposition theory concerned what a categorical term (the subject or predicate of a proposition) can be taken to stand for in a particular context, and it was used to diagnose fallacies and test inferences.

The three main types of supposition were material, simple and personal. A term was said to have 'material' supposition when it stood for itself or its equiforms, as in 'Man is a noun'. Material supposition took the place of twentieth-century quotation devices, and gave rise to the same sorts of problems. A term was said to have 'simple' supposition when it stood for a universal, as in 'Man is a species', but this type was controversial, and not only because the question of the nature of universals is itself highly controversial. Peter of Spain suggested that

nouns used as predicates, such as ‘animal’ in ‘Every man is an animal’, had simple supposition, and William Arnaud in the late thirteenth century suggested that singular terms, such as ‘Socrates’ in ‘Socrates is an individual’, had simple supposition. These views were not generally accepted. Finally, a term has ‘personal’ supposition when it is taken for its normal referents, as when ‘man’ is taken for Socrates, Plato and so on.

Some logicians distinguished accidental personal supposition from natural supposition which allowed a term to have pre-propositional reference to all its referents, past, present and future, while others insisted that supposition must be purely propositional and contextual. This debate in turn affects the doctrines of ‘ampliation’, whereby the reference of a term can be extended, and ‘restriction’ (the opposite). Parisian logicians, such as Jean le Page (writing *c.*1235), tended to accept natural supposition, and to say (like Buridan in the fourteenth century) that terms had natural supposition in scientific propositions - that is, universal necessary truths - so that no ampliation was necessary. As a corollary, in non-scientific propositions the supposition of terms was restricted in various ways. For English logicians in the thirteenth century, all supposition was contextual, and the notion of ampliation had to be used when the subject of a proposition was to extend beyond present existent things. Both groups agreed that modal terms were particularly apt to produce ampliation. For instance, in ‘A man can run’, ‘man’ is ampliated to supposit for present and future men (Lambert of Lagny, mistakenly called ‘of Auxerre’, in Kretzmann and Stump 1988: 117).

The three types of personal supposition most often appealed to are determinate, purely confused (*confuse tantum*), and confused and distributive. These types were normally illustrated by means of the descent to singulars. For instance, to say that the subject of a particular affirmative proposition, ‘Some *A* is *B*’, has ‘determinate’ supposition is to say that one can infer the disjunction of singular propositions, ‘This *A* is *B*, or that *A* is *B*, or the other *A* is *B*,…’. To say that the predicate of a universal affirmative proposition, ‘Every *A* is *B*’, has ‘purely confused’ supposition is to say that one can infer a proposition with a disjoint predicate, ‘Every *A* is this *B* or that *B* or the other *B*,…’. To say that the subject of a universal affirmative proposition has ‘confused and distributive’ supposition is to say that one can infer a conjunction of propositions, ‘This *A* is *B*, and that *A* is *B*, and the other *A* is *B*,…’.

Some people distinguished between mobile and immobile cases. For instance, no descent is possible from ‘Only every *A* is *B*’, and so *A* has ‘immobile’ supposition. A fourth type of supposition is ‘collective’ supposition, as in ‘Every man is hauling a boat’, given that they are doing it together. Here any descent will involve a conjoint subject, as in ‘This man and that man and the other man…are all hauling a boat’.

The theory of personal supposition was used to solve a variety of problems. Some involved tense (for example, how to convert the proposition ‘No old man will be a boy’); some involved multiple quantification (for example, how to distinguish ‘Every donkey of some man is running’ from ‘Some donkey of every man is running’, to take an early sixteenth-century example); and some involved such verbs as ‘promise’, as in ‘I promise you a horse’. How such sentences should be analysed, and what kinds of inferences they could figure in, elicited a variety of complex solutions. The positions adopted included these: Heytesbury took it that ‘horse’ had purely confused supposition; Paul of Venice added that the supposition was immobile; and Buridan appealed to the new doctrine of appellation, according to which some verbs appellate a form, that is, bring into play the notion of something coming under a certain description. Ockham preferred to replace the sentence by a more complex sentence, ‘You will have one horse by means of my gift’. This solution is closer to the type found in treatises on the proofs of terms, which focused, not on the reference of the term ‘horse’, but on the analysis of the word ‘promise’ in terms of giving a right to an object.

8 Syncategoremata; proofs of terms

Supposition theory concerned the reference of the subject and predicate terms in propositions; treatises on syncategoremata dealt with all the other terms, such as ‘all’, ‘some’, ‘not’, which appear in a proposition and exercise some logical function. The two areas of investigation were very closely associated. Not only are there terms such as ‘infinite’ which can be interpreted both as categorematic and as syncategorematic, but some terms were dealt with in each type of treatise. For instance, in the first half of the thirteenth century the word ‘all’ or ‘every’ (*omnis*) was discussed in the tracts on distribution associated with supposition theory in the textbooks of Jean le Page, Peter of Spain and Nicholas of Paris, but in the treatises on syncategoremata by the Englishmen William of Sherwood and Robert Bacon. Peter of Spain and Nicholas of Paris both wrote treatises on

syncategoremata in which '*omnis*' is not discussed. As was noted above (§2) these works were very closely associated with those on sophismata.

Treatises on syncategoremata were most prominent in the thirteenth century, and they did not altogether disappear in the fourteenth. For instance, the late-fourteenth-century English logician, Richard Lavenham, wrote one. However, two other forms of writing come to the fore. First, there are a lot of short treatises on particular syncategorematic terms, including 'know' (*scire*), 'begins' (*incipit*) and 'ceases' (*desinit*), and terms with the power of producing purely confused supposition. Second, and most important, are the treatises on the proofs of terms (or proofs of propositions), whose best-known example is the *Speculum puerorum* of Richard Billingham, an Oxford author of the mid-fourteenth century.

In this context, a proof seems to be a method of clarifying a sentence containing a particular sort of term, or of showing how one might justify that sentence. There were three groups of terms. 'Resoluble' terms are those whose presence calls for explanation or clarification through ostensive reference, as captured in an expository syllogism (that is, one with singular terms). Thus 'A man runs' is resolved into the expository syllogism 'This runs; and this is a man; therefore a man runs'. 'Exponible' terms are those whose presence calls for exposition of the sentence in terms of a set of equivalent sentences. For instance, the sentence 'Only a man is running', which contains the exclusive term 'only', is expounded as 'A man is running and nothing other than a man is running'. Other exponible terms are exceptives, such as 'except', reduplicatives, such as 'inasmuch as', 'begins' and 'ceases', 'infinite', and so on. In fact, they are the terms which had figured prominently in treatises on syncategoremata. Finally, there are 'official' (*officiales*) or 'officiable' (*officiabiles*) terms, so called because they performed a function (*officium*). These included any term that governed a whole sentence or that treated a whole sentence as modifiable, such as modal terms, and such terms as 'know', 'believe', 'promise', 'desire' and 'owe'. Analysis of sentences containing such terms shows why they are referentially opaque when taken in the compounded sense.

Treatises on proofs of terms were very popular during the late fourteenth and fifteenth centuries, but they were joined by treatises which dealt with exponible terms alone. Pierre d'Ailly wrote one such treatise, and several were written in the early sixteenth century.

9 Consequences

The notion of '*consequentia*' was already discussed in the context of conditional statements by Garlandus Compotista, who wrote in the eleventh century, and was heavily influenced by Boethius' work on hypothetical syllogisms and on Topics. However, it was not until the beginning of the fourteenth century that consequences became the subject of separate studies, one of the very earliest being written by Walter Burley around 1300. The origin of these treatises is controversial. It has been argued that consequences grew out of the study of Topics, especially topical maxims (see §4), but there is no sign of a theory of consequences in thirteenth-century works on Topics, and very little in thirteenth-century commentaries on the *Prior Analytics* or the *Sophistical Refutations*, both places where discussion might be expected. There is some evidence of relevant discussion in treatises on syncategorematic terms, especially under '*si*' (if…then), but it seems most likely that ordered treatment grew out of the English desire to produce lists of rules for use in disputations about sophismata. This ties in with the strong orientation of many fourteenth-century English treatises towards syncategorematic terms. A small group of authors, including Ockham and the later Burley, did produce treatises between about 1325 and 1340 which are more oriented towards Topics, but these constitute a minority treatment. The rare Continental treatises, such as that of John Buridan (*c.*1335), had more to do with syllogistic than with syncategoremata. English treatises, especially those by such later fourteenth-century authors as Ralph Strode and Richard Ferribrigge, were used and commented on in the fifteenth century in Germany and Italy.

One problem to do with consequences is the relationship between a *consequentia* and a conditional statement. Unlike modern logicians who make sharp distinctions between syntax (what the rules allow one to infer) and semantics (validity), and between language (statements) and metalanguage (statements about statements), medieval logicians seem to have been indifferent about whether to present consequences as rules of inference, as arguments which may be valid or invalid, or as conditional statements. Buridan, for instance, defined a consequence as a hypothetical (that is, compound) proposition, and said that a good consequence could be called true or valid. The standard remark that all true conditionals are necessary, and all false ones impossible, strengthened the tendency to equate consequences and conditionals, as did the use of 'antecedent' to refer to the conjunction of the members of

the set of premises and ‘consequent’ to refer to the conclusion of a consequence.

Another problem concerns the definition of validity. The claim that a consequence is valid if and only if it is impossible for the consequent to be false when the antecedent is true was questioned for two reasons. In the first place, the propositions involved were taken to be occurrent items, whether written, spoken or mental. They could fail to exist, in which case there is nothing to carry a truth-value. Alternatively, their meaning could be at odds with their actual expression, as in ‘Every proposition is affirmative, therefore no proposition is negative’. Such problems were discussed at length, for example, by Buridan, who solved them by substituting a definition in terms of signification, that is, that a consequence is valid if and only if it is impossible for it to be as signified by the antecedent without its being as signified by the consequent.

The presence of ‘if and only if’ raises the second problem. The truth definition (or Buridan’s substitute) may provide a necessary condition for validity, but is it sufficient? If it is sufficient, then we must accept the paradoxes of strict implication, that is, that anything follows from an impossible proposition, and a necessary proposition follows from anything. The debate about these paradoxes began in the twelfth century, and there was a series of attempts to provide a second condition which, with the first, would be sufficient for validity. Abelard had a containment principle by which the dictum of the antecedent should contain the dictum of the consequent (see Abelard, P. §4), and Robert Kilwardby in the next century, like Strode in mid-fourteenth-century Oxford, said that the consequent had to be understood in the antecedent. Some people in the thirteenth century focused on reality, and argued that a consequence must capture a causal relation, and that as a result the antecedent must be about a state of affairs that can at least be supposed to be possible. None of these people could accept the paradoxes as formally valid. On the other hand, the Parvipontani or Adamites (followers of Adam of Balsham) in the twelfth century and Buridan in the fourteenth were happy to accept the first clause as offering both necessary and sufficient conditions for validity, with all that that implied for the acceptance of the paradoxes.

One favourite division of consequences, found in such authors as William of Sherwood and Robert Kilwardby, was into natural and accidental (or non-natural) consequences. A natural consequence met the two-clause condition for validity, and an accidental one did not. The paradoxes could then be classified as accidental consequences. Later authors tended to make a division into formal and material consequences, but we have to be careful here, as there were two approaches to the notion of material consequence. For some, a material consequence involved the independence of antecedent and consequent, and just was the two paradoxes. For others, such as Buridan, a materially valid consequence is defined as one which does not hold for all terms arranged in the same way, and includes such enthymemes as ‘If all animals are running, then all humans are running’.

Another division of consequences is into those valid absolutely (*simpliciter*) and those valid as of now (*ut nunc*). This division, which appears in Ockham, has been much discussed, though it is in fact relatively rare, at least in fourteenth-century English treatises. Consequences valid *ut nunc* have been thought to correspond to the modern material conditional, but this is not so, as modal notions are definitely involved. Given that Socrates may not exist, and given that sentences with non-existent subjects are held to be false, ‘If all men are running, then Socrates is running’ is not generally valid. However, if Socrates does exist, then it is impossible for the antecedent to be true and the conclusion false, and we have a consequence which is valid as of now.

The question of the paradoxes also leads us to the question of the relationship between consequences and propositional logic, as discussion of the paradoxes provided one of the main occasions for sequences of rules to be used in a proof. The typical sequence was as follows: from the contradiction ‘*P* and not-*P*’ we can infer *P*, and from *P* we can infer ‘*P* or *Q*’. From ‘*P* and not-*P*’ we can also infer not-*P*, and from ‘*P* or *Q* and not-*P*’ we can infer *Q*. Hence, from ‘*P* and not-*P*’ we can infer *Q*. This sequence is found in the twelfth-century theologian Alexander Neckham, and it reappears through the centuries. But the explanations given differ widely. In his treatise on syncategoremata, Peter of Spain appealed to topical maxims, and claimed more generally that we need topical maxims as a foundation for consequences. On the other hand, in Pseudo-Scotus, writing *c*.1350, we get an appeal to what seem to be straightforward rules of propositional logic.

It has been claimed that later treatments of consequences showed that medieval logicians grasped that quantificational logic is founded on propositional logic, but this is anachronistic. Texts on consequences contain a mixture of rules (from the point of view of the modern logician, that is). Some are general (‘If *A* is the antecedent of *B*, and *B* is the antecedent of *C*, then *A* is the antecedent of *C*’). Some are propositional and truth-functional

('Since a conjunction is true if and only if its conjuncts are true, from ''*A* and *B*'' we may infer *A*'). *Modus ponens*, *modus tollens* and De Morgan's rules are given. Other rules are syllogistic; yet others are modal. Medieval logicians had no interest in classifying rules according to different formal systems, since they did not operate with the notion of a formal system in the first place.

10 Insolubles

'Insolubles' (*insolubilia*), so-called, it was claimed, because they are difficult to solve rather than completely insoluble, are semantic paradoxes. The simplest version is the liar paradox, 'What I am saying is false', given the *casus* or initial situation that this is all that is said, but complex versions with hypothetical propositions ('God exists, and some conjunctive proposition is false') or sequences of mutually referring propositions ('Suppose that Socrates says ''Plato says something false'', and Plato says ''Socrates says something true''' - Albert of Saxony, in Kretzmann and Stump 1988: 357, 349) were also discussed. In the twentieth century such paradoxes have been used to cast doubt on the very foundations of semantic theory, and have led to elaborate distinctions between levels of language and metalanguage. Medieval logicians, however, showed no signs of such a crisis mentality, and while they did employ certain restrictions on self-reference, and make certain distinctions between language and metalanguage, these techniques were generally limited to the problem in hand.

A very early formulation of the liar paradox is found in Adam of Balsham, writing in 1132, and later in the twelfth century Alexander Neckham gives a formulation. Full discussion seems to have started at least by the second half of the twelfth century, and continued until the sixteenth century. In the early period insolubles seem to have been discussed in the framework of Aristotle's *Sophistical Refutations*, and after that, within the general framework of a sophismatic disputation. From about 1320 to 1350 there was a period of intense and sophisticated activity, during which new solutions were given, and the framework of an obligational disputation was used (see §11). After 1350 some authors, such as John Wyclif and Pierre d'Ailly, still had original solutions to offer, but most authors, including Paul of Venice, merely elaborated previous ones.

The early solutions to the paradox include 'restriction', which forbids self-reference, either by saying that a part ('false') cannot supposit for a whole ('I say what is false'), or by saying that 'I say' cannot refer to the present moment; and '*cassatio*', according to which a paradox says nothing at all. The first major new solution was that of Thomas Bradwardine (*c.*1320-4). Bradwardine claims that propositions signify whatever follows from them and that insolubles signify that they are true. As a result, an insoluble signifies that it is true, but it is true if and only if it is both true and false. This cannot be, so it is false. John Buridan (§2) and Albert of Saxony both follow Bradwardine. Roger Swyneshed was author of another popular solution. He said that truth requires not just correspondence with reality but also absence of self-falsification, and Paul of Venice gives a version of this solution in his *Logica Magna*.

11 Obligations

Treatises on 'obligations' discussed the rules for obligational disputations, which were logical exercises for students. They got their title because of the rules obliging one to answer in a certain way. Obligational terminology is found in the late twelfth century, and the theory was already well-advanced in Paris by 1240-50. In 1302 Walter Burley wrote a treatise which summarized much previous material, and laid the foundations for later developments. Other important treatises were by Roger Swyneshed (1330-5), Ralph Strode and Marsilius of Inghen. The terminology of obligational disputation, including the use of a *casus* or description of the initial situation, permeated much fourteenth-century writing, including theology.

Various types of obligational disputation were discussed. One classification, which has thirteenth-century origins and was popularized by Marsilius of Inghen in the second half of the fourteenth century, related to the three legitimate ways of responding: '*concedo*' or 'I grant it'; '*nego*' or 'I deny it'; and '*dubito*' or 'I doubt it'. Walter Burley had six types, including '*impositio*', in which words and sentences were given new meanings for the duration of the disputation.

The most important of the different types of obligational disputation was '*positio*'. In this type of disputation, the opponent began the game by positing a proposition which the respondent had to grant, provided that it was logically possible. This *positum* was normally false, and it was usually trivial in content. Next the opponent had to

propose a series of other propositions, which could be either true or false. In replying to these, the respondent had to follow the rules of the game. If a proposition followed from the set formed by the *positum*, the propositions already granted and the negations of propositions already denied, it had to be granted. If a proposition was inconsistent with the set so determined, it had to be denied. If a proposition was such that neither it nor its negation followed from the set so determined, it was called *impertinens* or irrelevant, and the respondent had to reply to it according to his current state of knowledge. The task of the opponent in this game was to lead the respondent to the point of accepting an inconsistent set, either by granting and denying the same proposition, or by granting a proposition whose negation would follow from the set of propositions already granted. The task of the respondent, of course, was to avoid falling into these traps. In order to be successful, the respondent needed an excellent grasp of the logical relations between propositions and also an excellent memory. This is why obligational disputations could serve as valuable exercises for young logicians.

Some changes are reflected in the treatises. In the early period, impossible *positio* was very important, and still occurs in Ockham, though very rarely afterwards. In this kind of disputation, one begins with an impossible proposition which can at least be conceived to be true, a condition which excluded the use of explicit contradictions. Another change related to developments in modal logic. In the thirteenth century, the respondent had to deny that the actual 'now' of disputation was the instant at which the false but possible *positum* is treated as true, and as a result the replies to irrelevant propositions were related to actual time, while the replies to relevant propositions were related to an imaginary time. Once the notion of synchronic alternatives to the actual world had been adopted, all replies could be related to the same time. In the later period, Roger Swyneshed altered the relationship between irrelevant propositions and the *positum*. In the usual rules, once an irrelevant proposition had been admitted into the sequence, it (or its negation) became relevant. For Swyneshed, relevance was strictly a matter of a relation between the *positum* and those propositions that follow from or are incompatible with it alone. This produces two separate sets of responses, one involving relevant propositions, and one involving irrelevant propositions. As a result, one can deny a conjunction both of whose parts have been granted, and grant a disjunction both of whose parts have been denied. This view generated a good deal of discussion in the fourteenth and fifteenth centuries, and was usually rejected.

The biggest problem presented by treatises on obligations has to do with their purpose. They were certainly closely connected with consequences, insolubles and sophismata. Equally certainly they were closely connected with undergraduate training. Among other possibilities, it has been suggested that they also presented theories of counter-factual conditionals, or that they represented thought-experiments in which retrospective construction of a possible world takes place.

See also: Language, medieval theories of; Logic, Renaissance; Logical and mathematical terms, glossary of

E.J. ASHWORTH

References and further reading

Ashworth, E.J. (1978) *The Tradition of Medieval Logic and Speculative Grammar from Anselm to the End of the Seventeenth Century: A Bibliography from 1836 Onwards*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(Annotated bibliography with full indexes.)

Ashworth, E.J. and Spade, P.V. (1992) 'Logic in Late Medieval Oxford', in J.I. Catto and R. Evans (eds) *The History of the University of Oxford*, vol. 2, *Late Medieval Oxford*, Oxford: Clarendon Press, 35-64.(Detailed account of the period 1330-1500 when English logicians were influential throughout Europe.)

Broadie, A. (1993) *Introduction to Medieval Logic*, Oxford: Clarendon Press, 2nd edn.(A brief, readable introduction to some parts of medieval logic. The first edition should not be used.)

Buzzetti, D., Ferriani, M. and Tabarroni, A. (eds) (1992) *L'Insegnamento della Logica a Bologna nel XIV Secolo (Logic Teaching at Bologna in the Fourteenth Century)*, Bologna: L'Istituto per la Storia dell'Università.(A useful collection of Italian studies and Latin texts related to logic teaching at an important Italian university.)

Green-Pedersen, N.J. (1984) *The Tradition of the Topics in the Middle Ages: The Commentaries on Aristotle's and Boethius' 'Topics'*, Munich and Vienna: Philosophia.(The standard study of Topics; from Aristotle to the sixteenth century.)

Jacobi, K. (1988) 'Logic (ii): The Later Twelfth Century', in P. Dronke (ed.) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press, 227-51.(Excellent introduction to the debates in

the twelfth-century schools of logic.)

Jacobi, K. (ed.) (1993) *Argumentationstheorie: scholastische Forschungen zu den logischen und semantischen Regeln korrekten Folgerns (Theory of Argumentation: Scholastic Researches into the Logical and Semantic Rules of Correct Inference)*, Leiden, New York and Cologne: Brill.(A collection of 35 studies, 25 in English, on inference in various parts of medieval logic, with full introductions to the different sections in both German and English. An excellent source, with a full bibliography.)

Knuuttila, S. (1993) *Modalities in Medieval Philosophy*, London and New York: Routledge.(A difficult but rewarding study of modal logic and how it was employed.)

Kretzmann, N., Kenny, A. and Pinborg, J. (eds) (1982) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(Much of this volume is devoted to logic between 1200 and 1350; the best single source on this period available.)

Kretzmann, N. and Stump, E. (trans.) (1988) *The Cambridge Translations of Medieval Philosophical Texts*, vol. 1, *Logic and the Philosophy of Language*, Cambridge: Cambridge University Press.(An invaluable collection of translations of authors from Boethius to Albert of Saxony; designed to supplement the *Cambridge History* (Kretzmann, Kenny and Pinborg 1982).)

Lewry, P.O. (l984) ‘Grammar, Logic and Rhetoric 1220-1320’, in J.I. Catto (ed.) *The History of the University of Oxford*, vol. 1, *The Early Oxford Schools*, Oxford: Clarendon Press, 40l-33.(A lot of useful material about logic and logic teaching.)

Libera, A. de (1985-92) ‘Bulletin d’histoire de la logique médiévale’, *Revue des sciences philosophiques et théologiques* 69: 273-309, 71: 590-634, 76: 640-66.(A detailed, critical discussion of recent literature, with mini-introductions to a number of topics, including deontic logic and obligations.)

Maierù, A. (1994) *University Training in Medieval Europe*, trans. and ed. D.N. Pryds, Leiden: Brill. (This work, by a leading historian of logic, focuses particularly on how logic was taught, and its place in the curriculum, not just in universities, but in the *studia* of the religious orders.)

Pinborg, J. (1984) *Medieval Semantics: Selected Studies on Medieval Logic and Grammar*, London: Variorum.(A collection of seminal articles, published between 1969 and 1981.)

Read, S. (ed.) (1993) *Sophisms in Medieval Logic and Grammar: Acts of the Ninth European Symposium for Medieval Logic and Semantics, held at St Andrews, June 1990*, Dordrecht, Boston, MA, and London: Kluwer. (A very useful collection of articles.)

Rijk, L.M. de (1962-7) *Logica Modernorum: A Contribution to the History of Early Terminist Logic*, Assen: Van Gorcum, 2 vols in 3 parts.(A seminal collection of mainly twelfth-century texts with discussion. The source of much subsequent research, especially on supposition theory.)

Spade, P.V. (1988) *Lies, Language and Logic in the Late Middle Ages*, London: Variorum.(Articles on insolubles, Ockham’s semantics, and obligations, by a leading historian of logic.)

Stump, E. (1989) *Dialectic and Its Place in the Development of Medieval Logic*, Ithaca, NY, and London: Cornell University Press.(A very readable collection of studies, and an excellent introduction to dialectic. Some of the material on consequences and obligations needs updating.)

Tweedale, M. (1988) ‘Logic (i): From the Late Eleventh Century to the Time of Abelard’, in P. Dronke (ed.) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press, 196-226.(A good account of the early period, including Abelard.)

Logic, philosophy of

Philosophy of logic can be roughly characterized as those philosophical topics which have emerged either from the technical development of symbolic (mathematical) logic, or from the motivations that logicians have offered for their technical pursuits. In settling on a list of subjects to classify as philosophy of logic, therefore, there is a certain degree of arbitrariness, since the issues which emerge from the technical development of logic can equally well be assigned to such areas as semantics, philosophy of language, philosophy of mathematics, epistemology, and even ethics (see Semantics; Language, philosophy of; Mathematics, foundations of).

1 The impact of modal logic

In the broad area of mathematical logic, the biggest philosophical punch is packed by modal logic, including tense logic (see Modal logic; Modal logic, philosophical issues in; Tense and temporal logic). Modal logic has been important since Aristotle (see Logic, ancient; Logic in the 17th and 18th centuries; Logic in the 19th century; Logic in the early 20th century) but has only been put on a rigorous footing in the second half of the twentieth century, by such figures as Hintikka, Kanger, Prior, and most especially Kripke (see Semantics, possible worlds). The most important philosophical outgrowth of this mathematical work is contained in Kripke's three lectures from January 1970 published as 'Naming and Necessity', in which Kripke draws out some ways in which possible worlds semantics is in tension with then-prevailing orthodoxies in the philosophy of language and mind. Some of Kripke's views have become new orthodoxies since (see Essentialism; Proper names; Reference §§2-4; for related work by David Lewis, Robert Stalnaker, David Kaplan and others that uses the possible worlds framework, see Counterfactual conditionals; Demonstratives and indexicals; Descriptions).

To give some flavour of developments here, consider the familiar Fregean view that the relation of *reference* which holds between a name and its bearer is sustained by the relation of *presentation* which holds between the *sense* of the name and the bearer of the name: the name refers to such-and-such an object precisely because it expresses a sense which presents that object (see Frege, G. §3; Sense and reference). When pressed for an explanation of what the senses of names are like, the natural Fregean response is to specify them, as Frege himself did in some cases, using definite descriptions (see Descriptions). So, for instance, the sense of the name 'Aristotle' might be 'the pupil of Plato who taught Alexander'. However, though it may well in fact have been Aristotle who taught Alexander, there are many ways things might have gone (many 'possible worlds') in which someone other than Aristotle is taught by Plato and teaches Alexander: suppose Aristotle had got the appointment but been killed in an accident before he could take it up, and had been replaced at Philip's insistence by another pupil of Plato. The description 'the pupil of Plato who taught Alexander' is therefore 'non-rigid', in Kripke's terminology. That is, it can pick out different individuals in different possible worlds, and in some worlds may pick out no one (Philip for some reason comes to distrust Platonic pedagogy and fails to conduct an equal opportunity search). But it is clear from the formal semantics for modal logic that there is conceptual 'room' for a category of expression which is 'rigid', in the sense that it picks out the same object in every possible world, or at least in every possible world where it picks out any object at all. So the formal semantics prompts the question whether names in natural language behave as if their reference is determined by a sense which presents different individuals at different worlds, or whether they behave as if they are rigid designators. With a series of brilliant examples Kripke demonstrates that names are rigid designators and therefore do not express reference-determining senses which are non-rigid (see Proper names).

The idea that a formal semantics for a kind of logic provides an account of a possible semantics for a category of natural-language expression, opening the door to debate on whether it is the right account or not, also captures some of the philosophical bearing of kinds of logic other than modal logic. Thus free logic shows how name-like expressions can function without standard existential commitment (see Free logics, philosophical issues in); intuitionistic logic and many-valued logic show how a language can have a compositional semantics even if its sentences are not used to make statements with verification-transcendent truth-conditions which always either obtain or fail to obtain (see Compositionality; Intuitionistic logic and antirealism; Many-valued logics, philosophical issues in; Presupposition). And second-order logic offers a particular way of understanding the semantic import of a range of puzzling locutions, such as plural quantifiers (see Second-order logic, philosophical issues in). In all these cases the formal semantics for the logical system prompts debates about how well the

semantics carries over to natural language.

2 Logic and language

There is also a collection of long-established topics discussion of which can be much improved, in rigour at least, in the light of the development of modern logic. For example, a distinction between propositions (or statements, or sentential contexts) which are *de dicto* and propositions (and so on) which are *de re* originates in medieval philosophy. But only contemporary modal logic affords the tools for a precise characterization of this distinction, although it must be granted that the distinction remains a puzzle in *epistemic* contexts (see De re/de dicto; Descriptions §2; Propositional attitude statements). Other topics which can be classified in this way include Essentialism, Existence, Identity, Indicative conditionals, Modal operators, Quantifiers and Vagueness. Again, to give some of the flavour of this kind of work, consider the *de re*/*de dicto* contrast. There is an evident syntactic difference between ‘It is necessary that parents have children’ and ‘Parents are such that it is necessary that they have children’, but just because there is a syntactic difference, it does not follow that there is any interesting difference in meaning. But the difference can be brought out quite precisely in possible worlds semantics. To say that it is necessary that parents have children is to say that in every possible world, the people who are parents *in that world* have children in that world; and this is an obvious truth. On the other hand, to say that parents are such that it is necessary that they have children is to say that the people who are parents *in the actual world* are such that they have children in every possible world. This is clearly false, even putting aside contingency of existence of actual parents. For given anyone who is actually a parent, there is a way things could have gone - a possible world - in which that person is childless, hence not a parent (see Quantifiers, substitutional and objectual; Modal operators).

When a formal semantics for a system of logic is applied to a fragment of natural language, a very precise account of the literal content of sentences in that fragment is given. But there may be aspects of the meanings of those sentences which are omitted. Philosophical views may then divide over whether the formal semantics has been shown to be wanting as an account of the semantics of the fragment, or whether instead the aspects of meaning not captured have been shown not to belong to literal content (see Presupposition). In the case of indicative conditionals, for instance, the formal semantics that is relevant is the simplest possible kind, namely, the truth-functional account of ‘if…then…’. According to this account, ‘If *p* then *q*’ is true if *p* is false or if *q* is true, regardless of the actual meanings of *p* and *q*. So in particular, any indicative conditional with a true consequent is true; examples would include ‘If lead floats in water then lead sinks in water’ and ‘If the solar system has nine planets then the Conservative Party lost the British elections in 1997’. Barring an astrological justification of the latter, both these conditionals look decidedly odd. But oddness is one thing, falsity another. The idea that such conditionals are false is based on the thought that if a conditional is true, then in establishing it in the most direct manner, non-redundant use has to be made of the antecedent. Spelling this out leads to relevance logic (see Relevance logic and entailment; Indicative conditionals). On the other hand, if we say the conditionals are merely odd, we are led to some theory of communication to explain the oddness (see Grice, H.P.; Implicature).

But we should not take away the impression that the traffic is all one way, from logic to language or from pure mathematics to pure philosophy. There is a two-way street here, with the above comments on conditionals representing a common phenomenon; that of a concern in the philosophy of language giving rise to a formal development which in turn feeds back into philosophy. For example, the idea that for a conditional to be true, the most direct way of establishing it must make non-redundant use of its antecedent seems clear enough on the face of it, but familiarity with logic of conditionals literature may well lead one to reconsider. This kind of dialectical interplay should continue to be a fruitful source of philosophical research for the foreseeable future.

See also: Adverbs; Ambiguity; Anaphora; Dummett, M.A.E.; Fallacies; Formal and informal logic; Identity of indiscernibles; Imperative logic; Indirect discourse; Intensional entities; Intensionality; Kripke, S.A.; Linear logic; Logic in China; Logic in Islamic philosophy; Logic in Japan; Logical and mathematical terms, glossary of; Logical constants; Logical form; Logical laws; Mass terms; Necessary truth and convention; Ontological commitment; Predication; Prior, A.N.; Propositions, sentences and statements; Quantification and inference; Quantifiers, generalized; Questions; Quine, W.V.; Russell, B.A.W.; Scope; Type/token distinction; Use/mention distinction and quotation; Vagueness

GRAEME FORBES

References and further reading

Hughes, R.I.G. (ed.) (1993) *A Philosophical Companion to First-Order Logic*, Indianapolis, IN: Hackett Publishing Company.(New and reprinted papers of varying levels of difficulty.)

Tomberlin, J.E. (ed.) (1994) *Logic and Language*, Philosophical Perspectives 8, Atascadero, CA: Ridgeview.(A collection of new papers on relevant topics.)

Logic, Renaissance

Renaissance logic is often identified with humanist logic, which is in some ways closer to rhetoric than to the study of formal argumentation. This is a mistake, for although changes did take place, a hard core of logical teaching remained the same throughout the medieval and Renaissance periods and into the eighteenth century. Logic was embedded in the educational system as the main study of beginning undergraduates, and although institutional changes had an effect on the presentation and use of logic texts, the study of valid arguments was always central.

There are two obvious differences between medieval texts and their sixteenth-century successors. The first is a new emphasis on following the order and material of Aristotle's Organon*, with the consequent emphasis on the categorical syllogism as the central type of argument. Such medieval material as survived was strictly subordinated to this end; and even though the humanist logicians Agricola and Ramus had tried to ignore Aristotelian syllogistic and the doctrines propaedeutic to it (such as conversion and opposition), their omissions were rapidly remedied by subsequent textbook writers.*

The second difference has to do with language and style. Medieval writers treated Latin as a technical, almost artificial language. They were deeply concerned with the effects that different word-orders and the addition of extra logical particles had on both meaning and reference, and they made heavy use of sophismata*, deliberately constructed problematic or puzzling sentences. Although Latin remained the language of instruction, the approach of a Renaissance logician, whether humanist or Aristotelian, commentator or textbook writer, is totally different.* Sophismata *have completely disappeared, and so too has any attempt to treat Latin as a technical language in which different word-orders represent different logical structures. The propositions used for such operations as syllogistic conversion are presented in an already fully standardized form, and they are always relatively simple.*

Why these changes came about is a difficult question. Humanism coexisted too long with medieval logic for humanism to be the sole explanation, and the fact that Renaissance logicians returned to the commentaries of Averroes and Aquinas on the Organon shows that mere revolt against anything medieval is not a sufficient explanation either. Changes in grammar teaching, changes in the relation of logic to the study of natural science, and changes in other parts of the university curriculum presumably have a good deal to do with the appearance of a new style of logic. In particular, the humanist emphasis on logic as a tool for analysing discourse focused attention on the use of logic in literature, history and biblical studies, and demanded a combination of simplicity and literary elegance, rather than any genuine formal innovation. There was no concurrent move to relate logic to the new developments in mathematics, and logic was not to be seen as a formal system linked with other formal systems until the nineteenth century.

1 The nature and divisions of logic

One of the features distinguishing sixteenth-century textbooks, whatever their orientation, from their medieval predecessors was their approach to the nature and divisions of logic. Medieval authors had asked whether logic is more properly a linguistic discipline directed towards spoken words (*scientia sermocinalis*) or a rational discipline directed towards concepts (*scientia rationalis*), and this question did not disappear (see Logic, medieval §4). However, its presentation was to some extent altered. The influence of the humanists, especially Agricola, gave general popularity to Cicero's definition (*Topics* II 6) of logic as an art of discourse (or *ratio disserendi*) (see Cicero §2), and this notion is also found in such scholastic authors as Toletus and Fonseca. At the same time, logic was frequently said to have as its subject either conceptual entities (beings of reason: *entia rationis*) (see John of St Thomas §2; Language, renaissance philosophy of §3), or second intentions (see Zabarella, J. §2). Toletus denied a distinction between these two notions in this context, on the grounds that to talk about the kind of conceptual entity which has no existence independently of the mind is just to talk about those second-order logical concepts or intentions (genus, species, syllogism and so on) which we use to organize first-order concepts, those which apply directly to things in the world.

One issue which was the subject of lively debate concerned the very use of the word *logica* as opposed to *dialectica*. It was a medieval commonplace that *dialectica* could be used in two senses, a broad sense which equated dialectic with logic, and a narrow sense, in which dialectic was the kind of probable argumentation discussed in Aristotle's *Topics* (see §4) (see Fonseca, P. da §2). In the sixteenth century, Ramus argued in his

Scholae (Lectures) that the division of logic into demonstrative, dialectical and sophistical reasoning is baseless, and that all logic is dialectic (see §4). Zabarella, on the other hand, believed that 'dialectic' did name a distinct part of logic, and should be reserved for that part alone.

Another issue which reflects the impact of humanism concerned the way in which logic was divided. Perhaps the most important division was that into invention and judgment, invention being the part of logic which finds the subject matter for argumentation, and judgment being the part which organizes the subject matter. This division came principally from Cicero (*Topics* II 6) and was well known to medieval logicians, before it was further popularized by Agricola, Melanchthon and Ramus. A mark of Ramist textbooks was that they began with invention, but Melanchthon, followed by various authors including John Seton (see §5) put judgment first. Another way of organizing which was to become especially popular in the seventeenth century (see John of St Thomas §2; Logic in the 17th and 18th centuries §1) was a division in accordance with the three acts of mind, the apprehension of simples (terms), composition and division (propositions), and reasoning (arguments). Like the division into invention and judgment, this division has a long history.

2 The medieval inheritance

Although Aristotle served as a focal point for the study of logic during the Middle Ages and Renaissance (see §3), one of the main features of medieval logic was the production of independent texts on particular topics, some of which had no reference to Aristotelian logic. The most important of these texts were the so-called *parva logicalia*, or treatises dealing with the properties of terms, including their reference in various contexts. Here we find tracts on supposition, on relative terms, on ampliation, appellation, restriction and distribution (see Logic, medieval §7). These core treatises were supplemented in various ways, particularly by treatises about exponible terms, those logical particles such as 'except' and 'only' whose presence calls for the analysis of an apparently simple categorical proposition into several conjoined propositions.

A second major group of treatises consisted of the 'three tracts of the moderns', which concern propositions and the relations between them. Here we find treatises on consequences, obligations and insolubles. Treatises on consequences covered all types of argumentation, beginning with a general discussion of what constitutes a formally valid inference, and including a good deal of propositional logic. The syllogism often appeared as a special example of one kind of argumentation. Treatises on obligations dealt with the rules to be followed in a certain kind of disputation which was specifically designed to test the logical skills of undergraduates, and which deliberately confined itself to exploring the logical consequences of affirming an often bizarre falsehood. Treatises on insolubles dealt with semantic paradoxes, such as the standard liar paradox, 'What I am now saying to you is false', and they explored in some depth the semantic presuppositions of language (see Semantic paradoxes and theories of truth).

During the fifteenth century much of this medieval material was used throughout Europe, and Italy was particularly enthusiastic about the products of fourteenth-century Oxford logicians, such as Heytesbury and Ralph Strode. *Summulae* or general textbooks from the earlier period were also used. The best known is by Peter of Spain, and it gives a fairly complete outline of Aristotelian logic as well as the *parva logicalia*. However, it would be a mistake to think that only Peter of Spain was used, for in some places (such as England) he was never used, and in other places only parts of his work were used. Buridan (§2), Albert of Saxony and Paul of Venice were the authors of other popular textbooks. Indeed, the first medieval logic text known to have been printed is Paul's *Logica parva*, printed in 1472 (and still being reprinted in Venice as late as 1614).

At the beginning of the sixteenth century, particularly in Paris, there was a strong revival of the medieval tradition. Separate treatises on supposition theory, exponibles, consequences, insolubles and obligations continued to be written (see Major, J. §2), and to these were added new treatises on such matters as terms and opposed propositions. At the same time, new commentaries on Peter of Spain were being published in Paris, Germany and Spain. The 1529 edition of Domingo de Soto's *Summulae* included commentary on the first and fourth books of Peter of Spain along with treatises dealing with terms and supposition theory, exponibles, insolubles, and obligations (see Soto, D. de).

Outside Spain, this activity came to an abrupt halt around 1530. No new independent treatises on Peter of Spain were written; the publication of works written during the first three decades of the sixteenth century ceased, and

the publication of such medieval authors as Buridan and Heytesbury also ceased. Only in Spain and its colonies did authors continue to produce commentaries on Peter of Spain, such as that by Alonso de la Vera Cruz, first published in Mexico in 1554, or that by Tomás Mercado, published in Seville in 1571 (see Latin America, colonial thought in §3; Mexico, philosophy in §1).

By the end of the sixteenth century those parts of medieval material which could not easily be integrated with Aristotelian logic simply disappeared, at best being the subject of stray and often unflattering references. Insolubles were no longer discussed, obligations were replaced by a discussion of ordinary disputations in which truth was the main issue, and many of the topics connected with supposition theory were dropped. But exponibles, consequences and the central part of supposition theory remained in truncated and simplified form (see §6).

3 Aristotle and Aristotelianism

At least in principle, Aristotle's Organon played a central role throughout the medieval and Renaissance period (see Aristotle §4), though not all parts were studied with equal fervour. For instance, the *Topics* was sometimes omitted altogether, and sometimes only certain of its books were assigned. The text of the Organon was accompanied by explanation and commentary, and the invention of printing brought the wide dissemination of some of the medieval commentaries on Aristotle's logical works, including those by Walter Burley and Paul of Venice. Later there was a return to commentaries by the great theologians of the thirteenth and early fourteenth centuries. Thus, the commentaries of Aquinas on *De interpretatione* and *Posterior Analytics* reached a peak of popularity in the mid-sixteenth century, and the commentaries by or attributed to Albert the Great (§2), Giles of Rome and John Duns Scotus were also published (see Aristotelianism, medieval §5). The great Giunta edition of Aristotle with the commentaries of Averroes (1550-2) should also be mentioned.

The impact of humanist classical studies on Renaissance logic is particularly important in two respects: first, the edition and translation of Greek commentators on Aristotle, only some of whom had been known during the Middle Ages; and second, the edition and translation of Aristotle himself. During the fifteenth century the Byzantine humanist and philosopher Johannes Argyropoulos (*c.*1415-87) had produced new Latin translations, but as the Greek text became better known, more and more dissatisfaction was felt with former translations, including those of Argyropoulos. A good example of the new Greek-based texts is the Organon edition of Giulio Pace first published in 1584. In it we find the Greek text side-by-side with a new translation designed not only to read well but also to capture the philosophical significance of Aristotle's words. In the margins there is a commentary dealing with difficult points both of theory and of translation. However, the old translations did not disappear, because they provided much of the standard vocabulary in philosophy, and the work of surviving medieval commentators was keyed to them (see Humanism, Renaissance).

So far as content is concerned, there were few developments within Aristotelian logic. There was considerable discussion of scientific method (see §7), and there was also some lively debate about the acceptability of the fourth figure of the syllogism, particularly inspired by Zabarella. Otherwise, Aristotelianism's main achievement was its successful stand against attempts such as those by Lorenzo Valla (§4) and Petrus Ramus to replace basic Aristotelian notions. This achievement was mainly due to the felt need for formal content in logic teaching (see §5) (see Aristotelianism, Renaissance §1).

4 Humanist logic and the topics

To understand humanist logic, one must consider the relations between dialectic, the topics, and rhetoric. 'Dialectic' is a term with various meanings. For those who did not use it as a mere synonym for logic, at least three approaches were possible. One could present dialectic as the study of debate or disputation; one could present dialectic as the study of the topics (roughly speaking, headings under which material could be gathered); or one could present dialectic as the study of probable argumentation.

The last two approaches are closely related to the concerns of rhetoric, and there is a tendency in the literature to speak as if humanist logic was the result of subordinating logic to rhetoric. However, while it is true that rhetoric texts featured the invention or finding of topics, and that rhetoric, as the art of persuasion, could be construed as dealing with arguments about probabilities, the three standard works about topics had always been firmly part of the logic syllabus. These were Aristotle's *Topics*, together with Boethius' two works, *De topicis differentiis (On*

Topical Differences) and *In Ciceronis Topica (On Cicero's Topics)* (see Boethius, A.M.S. §3). Boethius' first work was more important in the medieval period, though in the fourteenth century it was largely replaced by the account of topics given by Peter of Spain in his *Summulae*. In the sixteenth century Boethius' second work became popular again, as did Cicero's original text. Agricola himself was well-acquainted with both Aristotle and Boethius.

The notion that dialectic is concerned with probable argumentation comes from Aristotle's *Topics*, which opens with a distinction between demonstrative reasoning, said to deal with certainties, and dialectical reasoning, said to deal with probabilities. However, there are two ways of reading the distinction. One way involves two types of logic, a logic of formally valid deductive arguments, and a logic of informal arguments or strategies. The other way of reading Aristotle's distinction is to assume that 'argument' or 'reasoning' in the Topics refers only to the syllogism as described in the *Prior Analytics*, namely a particular kind of formally valid deductive structure. This second reading led medieval logicians to interpret Aristotle's distinction between demonstrative reasoning and dialectical reasoning as an epistemic one. Thus, demonstrative reasoning involves formal syllogisms with premises that are certainly true, and dialectical reasoning involves the same formal syllogisms with premises that are only probably true. It was this epistemic reading that led Ramus to argue that dialectic and logic are the same, since the same patterns of argumentation appear in both. In Agricola, on the other hand, we find a somewhat greater awareness of the possibility of a properly non-formal logic.

The general failure of humanist logicians, including Agricola, to pursue dialectic as an informal logic of probabilities was closely associated with their treatment of the other aspect of dialectic, dialectic as a topics-logic, or means of finding the material for arguments. In Boethius, the notion of a topic had covered two things. First, a topic was a maxim (or 'maximal proposition'), a self-evidently true universal generalization which could either be inserted into an argument as a premise or which could be appealed to as providing a warrant for an argument. Some of these maxims were themselves turned into formal deductive arguments by medieval logicians, who had a tendency to absorb invention into judgment, but others were not easily formalized. As such, they provide the perfect nucleus for a developed informal logic. The second sense of topic for Boethius was the topic as the *differentia* of a maximal proposition, that is, the characteristic that enables us to classify maximal propositions into groups. To list the topics in this second sense is simply to list the headings under which material can be gathered: 'definition', 'genus', 'cause', 'opposite', 'similar' and so on.

What we find in Agricola (§2) is a deliberate rejection of the maxims, and of material that might belong to judgment rather than invention. He argued that the purpose of any maxim is simply to present a necessary argument, and that they are quite inappropriate for all the cases in which we are dealing with probabilities. We should note here that Agricola is clearly talking not just of the epistemological status of the premises of an argument, but of a probabilistic and informal relationship between premises and conclusion. He also claimed that maxims were too restrictive, for they represent an attempt to force a wide range of material into a narrow compass; and that they are of no use to someone who really understands the nature of topics. What we are left with, then, is the topics solely as headings under which material can be organized. Their most obvious link with argumentation, both formal and informal, has been broken.

In the hands of Agricola's later follower, Petrus Ramus (§2), we find the topics presented in an Agricolan way, without maxims, but with a new theoretical foundation. They are not merely useful headings for gathering material, but they represent the mind's natural organization of data. As such, topics function as categories; and, indeed, Ramus emphasized that his topics are more useful and more natural than Aristotle's ten categories (see Categories §1). If we are asked to discuss war, peace, or the state, it is no use to think in terms of the category of substance. Instead we must appeal to such topics as cause, event and opposites.

The result of these reforms was a subordination of rhetoric to logic (rather than the reverse), in three respects. First, the topics were now to be confined to dialectic, and rhetorical topics, which had covered material of a more particular kind, were to be reduced to dialectical topics. Second, rhetoric's *dispositio* is introduced into the dialectic text through the discussion of method as a way of ordering discourse. Third, we find that rhetoric, at least in the eyes of the humanist logician, was to be confined to the business of ornamenting discourse, and discussing only those persuasive devices which did not involve argumentation.

5 Humanist and Ramist textbooks

The needs of the university curriculum meant that humanist insights and changes were only partially reflected in logic teaching. Thus Agricola’s seminal *De inventione dialectica* (see Agricola §2) was often used in conjunction with a little Aristotelian handbook by George of Trebizond. The very popular textbooks of Melanchthon provide another good example. He enjoyed the Agricolan emphasis on clarity of style and the use of literary allusions; he accepted the importance of the topics and the part of logic called invention; and he discussed method. At the same time, he remained a convinced Aristotelian, who believed that students needed to be taught some formal logic. The formal techniques he used were those of syllogistic, and his work included a discussion of the other standard Aristotelian logical subjects, including the categories and the square of opposition for propositions.

Another example from the first part of the sixteenth century is John Seton’s *Dialectica*, first published in England in 1545, and later to appear with annotations by Peter Carter. Seton explained in his introduction that he had written the work because of the absence of a suitable text for the instruction of the young. Aristotle was too difficult, Agricola had deliberately restricted himself to the subject of invention, and Melanchthon’s style was not suited to elementary teaching. He makes frequent references to Agricola, as well as to Cicero, Quintilian and Erasmus. He included the bulk of traditional Aristotelian logic from the categories to the syllogism, which he described as the most important part of judgment, and he also retained the strictly medieval doctrine of the supposition of terms, though in a considerably truncated form.

In the second part of the century the most important textbook writer was Petrus Ramus (§2), the most notorious logician of the period. He is known both for his attacks on Aristotle and for the simplified logic presented in his *Dialectique* of 1555, and in a variety of Latin versions. The *Dialectique* had two parts. The first, on invention, covered the topics, and the second, on judgment, presented a deliberately simplified version of the syllogism followed by an account of method as a means of ordering in the arts and sciences. No reference was made to such standard material as the categories, the square of opposition, conversion, demonstration and fallacies. On the other hand, the work is rich with quotations from the poetry and prose of classical authors, which must have strengthened the impression among students that logic was both easy and fun.

The book had a remarkable publishing history, but it is not clear how much impact it actually had on the university scene. In England, Ramus certainly enjoyed some popularity at Cambridge but was less well received at Oxford. Even in Germany, where 151 editions appeared, university teachers of logic soon found serious deficiencies in Ramus’ book, and in the 1590s a new school of textbook writers known as the Philippo-Ramists appeared in Germany. Their aim was to combine what was best in Ramus with what was best in the more Aristotelian work of Philipp Melanchthon. As a result, they tended to restore those parts of Aristotelian logic that Ramus had deliberately omitted, and the syllogism was once more presented as central to formal logic (see Logic in the 17th and 18th centuries §1).

6 Late scholastic textbooks

If England and Germany saw the production of slim textbooks exemplifying the marriage between humanism and a simplified Aristotelianism, Spain and Italy were the scene of a much more substantial textbook production. Two important texts, both recommended by the Jesuit *Ratio Studiorum* of 1599, and thus possibly studied by Descartes at La Flèche, were the *Introductio in dialecticam Aristotelis (Introduction to the Dialectic of Aristotle)* of Franciscus Toletus (§3) and the *Institutionum dialecticarum (Dialectical Instructions)* of Pedro da Fonseca (§2). Toletus’ book was first published in Rome in 1560, and the last of its eighteen editions appeared in Milan in 1621. Fonseca’s work was first published in Lisbon in 1564, and the last of its fifty-two editions appeared in Lyons in 1625. The works were fairly similar in content. The Organon is very much central but medieval material is used to supplement Aristotle in various places. Exponibles are discussed, albeit in a simplified form; some non-syllogistic consequences dealing with truth and modality are listed, and material about supposition and related doctrines is presented as an aid to understanding fallacies.

The most important English text of this type was Robert Sanderson’s *Logicae Artis Compendium* which appeared at Oxford in 1615 and was to remain a standard text there until well into the eighteenth century. This, too, contains a full account of strictly Aristotelian logic, together with brief discussions of the medieval doctrines of supposition, exponibles and consequences, and a full discussion of method as a way of ordering discourse.

7 The discussion of method

A particular feature of textbooks, especially in the second part of the sixteenth century, was the discussion of method. First, and most important, is the study of scientific method, which is particularly characteristic of Italian universities. These had always been distinguished from Northern European universities by their strong emphasis on law and medicine, combined with a relatively slight emphasis on theology. The faculties of arts at such places as Padua provided studies leading to a degree in arts and medicine, and as a result the main emphasis was placed on logic and natural philosophy, as propaedeutic to medicine. In the fifteenth century works stemming from Oxford and Paris on such topics as the intension and remission of forms were particularly important (see Natural philosophy, medieval §6), but after about 1520 these works fell into sudden oblivion. By the mid-sixteenth century the focus of attention was on the Greek commentators and on Averroes (see Ibn Rushd), especially his *Physics* commentary. The main Aristotelian text studied was the *Posterior Analytics*. From Cajetan of Thiene (Gaetano da Thiene) (1387-1465) to Nifo there was a gradual refinement of writings on demonstration and scientific method which culminated in the work of Jacopo Zabarella (§4). His discussions of the methods of resolution, composition, and the so-called *regressus* (a method for uniting resolution and composition) were summarized in many later textbooks, often in uneasy alliance with two other types of method.

A second notion of method presents it as a way of organizing discourse, a method fully discussed by Agricola (§2). Agricola showed how argumentative and expository passages from classical literature could be analysed, a method which brought the logical text very close to the new humanist curriculum of many institutions (see Humanism, Renaissance §7). Agricola's approach was picked up and further popularized by Melanchthon (§2).

A third type of method is that of Ramus (§3), which is particularly directed towards the presentation of a subject, and best known for its presentation through diagrammed dichotomies. This too could be and was adapted to the analysis of literary and biblical texts.

See also: Logic, ancient; Aristotelianism, medieval; Aristotelianism, Renaissance; Aristotle; Humanism, Renaissance; Language, medieval theories of; Language, Renaissance philosophy of; Logic, medieval

E.J. ASHWORTH

References and further reading

Ashworth, E.J. (1974) *Language and Logic in the Post-Medieval Period*, Dordrecht and Boston, MA: Reidel.(A comprehensive account, thematically arranged, and making use of the techniques of modern symbolic logic.)

Ashworth, E.J. (1978) *The Tradition of Medieval Logic and Speculative Grammar from Anselm to the End of the Seventeenth Century: A Bibliography from 1836 Onwards*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(Annotated bibliography with full indexes.)

Ashworth, E.J. (1985) *Studies in Post-Medieval Semantics*, London: Variorum.(A collection including studies of supposition theory, insolubles and exponibles.)

Ashworth, E.J. (1988) 'Traditional Logic', in *The Cambridge History of Renaissance Philosophy*, ed. C.B. Schmitt, Q. Skinner and E. Kessler, Cambridge: Cambridge University Press, 143-72.(A useful general account with plenty of references.)

Jardine, L. (1988) 'Humanistic Logic', in *The Cambridge History of Renaissance Philosophy*, ed. C.B. Schmitt, Q. Skinner and E. Kessler, Cambridge: Cambridge University Press, 173-98.(A useful general account, though controversial on Valla; shows how humanist logic was used in textual exegesis.)

Lohr, C.H. (1988) *Latin Aristotle Commentaries II. Renaissance Authors*, Florence: Olschki.(Alphabetically organized list of Renaissance Aristotelians, with biographies and bibliographies of primary and secondary sources.)

Mack, P. (1993) *Renaissance Argument: Valla and Agricola in the Traditions of Rhetoric and Dialectic*, Leiden, New York and Cologne: Brill.(Excellent account of humanist logic, including Melanchthon and Ramus.)

Risse, W. (1964) *Die Logik der Neuzeit. I Band. 1500-1640 (The Logic of the Modern Period)*, Stuttgart and Bad Cannstatt: Frommann.(A dense, scholarly account, without much awareness of the medieval background, and with no awareness of twentieth-century logical developments.)

Risse, W. (1965) *Bibliographia Logica. Verzeichnis der Druckschriften zur Logik mit Angabe ihrer Fundorte. Band I. 1472-1800 (Bibliography of Logic: List of Printed Works on Logic with Details of their Location)*, Hildesheim: Olms.(The bibliography organized chronologically but it is fully indexed.)

Sanderson, R. (1618) *Logicae Artis Compendium*, ed. E.J. Ashworth, Bologna: Editrice CLUEB.(This book

contains a facsimile of the 1618 edition of Sanderson's work (referred to in §6), together with indexes and a lengthy introduction which gives full information about English logic in the sixteenth century, including a discussion of John Seton (referred to in §5).)

Vasoli, C. (1968) *La dialettica e la retorica dell'Umanesimo. 'Invenzione' e 'Metodo' nella cultura del XV e XVI secolo (Humanist Dialectic and Rhetoric: 'Invention' and 'Method' in the Culture of the 15th and 16th Centuries)*, Milan: Feltrinelli.(Dated but classic account of humanist logic.)

Wallace, W.A. (1995) 'Circularity and the Paduan *Regressus*: From Pietro d'Abano to Galileo Galilei', *Vivarium* 33: 76-97.(Outline of Paduan discussions of scientific method.)

Logical atomism

The name 'logical atomism' refers to a network of theses about the parts and structure of the world and the means by which language represents the world. Wittgenstein, in his Tractatus Logico-Philosophicus, expounds a version of logical atomism developed by him around the time of the First World War, as does Russell in works published contemporaneously. It is no accident that their work on logical atomism shares a common surface description since it resulted from their mutual influence at Cambridge. The common theme is that the meaning of our sentences is rooted in a primitive relation between simple expressions and their simple worldly bearers, the logical atoms. In a logically perfect language, atomic sentences describe configurations of these atoms, and complex sentences are combinations of the atomic sentences. But sentences of ordinary language may have a misleading surface form which is revealed as such by analysis. The common theme masks considerable differences of doctrine. In particular, there are differences in the nature of logical atoms and in the arguments for the existence of these atoms.

1 Wittgenstein's logical atomism

Wittgenstein's logical atomism is expounded in his *Tractatus Logico-Philosophicus* (1921) in which he explores the structure of the world and the structure of any language fit to represent the world (see Wittgenstein, L. §5). The unit of sentential representation is the atomic sentence, and Wittgenstein claims that every sentence can be analysed into a combination of atomic sentences which describe atomic states of affairs. Each atomic sentence has exactly one of the two truth-values, true and false, and there are no necessarily true or necessarily false atomic sentences. If an atomic sentence is true, then the state of affairs it describes exists, the existent atomic state of affairs being called 'an atomic fact'. If an atomic sentence is false, then the state of affairs it describes does not exist (see Facts §1; Truth, correspondence theory of).

An atomic sentence is structurally isomorphic to the atomic state of affairs described. It consists of an arrangement of names, and these names have simple objects as their meanings. The atomic state of affairs described by the atomic sentence consists of the named objects arranged in a way which corresponds to the arrangement of the names in the sentence. The objects named in atomic sentences and featuring in atomic states of affairs are Wittgenstein's logical atoms.

The atomic sentences form a system which is fit to represent any possible world. A possible world is represented by an assignment of truth-values to the atomic sentences. Atomic sentences are independent in the sense that there are no logical relations between them which rule out certain assignments of truth-values to the atomic sentences. The actual world is completely described by the atomic sentences which are true and false in it. So the actual world is the totality of atomic facts, the existent states of affairs described by the true atomic sentences.

Atomic sentences form the basis for Wittgenstein's analysis of the sentences of ordinary language. Any sentence of ordinary language has a unique translation into a sentence which is a truth-functional combination of atomic sentences. This is the principle of extensionality. A complex sentence is a truth-functional combination of atomic sentences just in case the truth-value of the complex sentence is entirely determined by the truth-values of its constituent atomic sentences. The sentential connectives of the propositional calculus, such as 'Čamp;', yield truth-functional combinations of atomic sentences.

Some truth-functional combinations of atomic sentences yield sentences which are true no matter what truth-values are assigned to the constituent atomic sentences, and sentences which are false no matter what truth-values are assigned to the constituent atomic sentences. These are the tautologies (such as 'P or not-P') and the contradictions (such as 'P & not-P'), respectively. Wittgenstein claims that all necessary truths are tautologies and all necessary falsehoods are contradictions.

Sentences of ordinary language may appear to be atomic, but analysis can reveal them to be complex. The aim of analysis is thus to uncover the accurate representations of the world which lie behind the misleading surface forms of ordinary language. The unit of sentential representation is the atomic sentence, but an atomic sentence is itself compounded out of names which have simple objects as their meanings. Hence, the meaning of every sentence of our language is grounded in this primitive meaning relation holding between names and the simple objects which

they name. But the more familiar entities of ordinary discourse, such as tables and chairs, are complex. So a sentence about a complex entity is bound to be analysed as a combination of atomic sentences, these sentences describing the arrangements of the simple objects which are the parts of the complex entity. Here we have the idea of analysis as the decomposition of complex entities into arrangements of their simple parts.

Why should there be simple objects at all? Why can complex objects not divide for ever and ever into less complex parts? Wittgenstein argues a priori from the possibility of representation to the existence of his simple objects using a *reductio ad absurdum*. First, he assumes that there is an atomic sentence featuring an expression standing for a complex entity. Second, he brings in a background assumption that sense be determinate, that is, every sentence has exactly one of the two truth-values in any possible world. As an illustration, suppose that the sentence ‘the broom is to the left of the cupboard’ is an atomic sentence. It features the expression ‘the broom’ which stands for a complex entity. The broom exists only if its parts exist and are arranged in the right way. Consider a possible world in which the broom does not exist - our sentence is not true in this world. But neither can it be false. For the sentence is atomic and thus does not contain as part of its meaning a condition which determines that the sentence is false when the broom does not exist. But every sentence must be either true or false in any possible world. So an atomic sentence must contain only expressions which stand for simple objects.

This shows that simple objects must exist but does not tell us what they are like. It appears that Wittgenstein thought that this was not a logician’s job. Early interpreters of the *Tractatus*, taking their cue from Russell’s logical atomism and Wittgenstein’s examples, took the simples to be the units of perceptual experience, in other words sense-data. But the independence of atomic sentences precludes such an interpretation. ‘This sense-datum is red’ and ‘this (same) sense-datum is green’ would be atomic sentences but they cannot both be true.

Wittgenstein later came to think that his model of the relationship between language and the world was radically misconceived. His most immediate criticism was to reject the independence of atomic sentences, which he did by considering sentences featuring colour predicates. This was at the same time the denial that all necessary truths and falsehoods are tautologies and contradictions, respectively. For the necessary falsehood ‘this sense-datum is red and green’ is not a contradiction. In the first part of his *Philosophical Investigations* (1953) Wittgenstein mounts a sustained attack against the pivotal claims that the meaning of sentences is rooted in the relationship between name and bearer and that the sense of sentences is determinate. Thus the justification for the existence of logical atoms is overturned.

2 Russell’s logical atomism

Russell developed his logical atomism over two decades, beginning with ‘On Denoting’ (1905) and ending with ‘Logical Atomism’(1924) (see Russell, B. §11). During this time he continually revised his ideas. (This section sketches the version of logical atomism he expounds in the eight 1918 lectures entitled *The Philosophy of Logical Atomism*.) Russell had originally coined the phrase ‘logical atomism’ to distinguish his picture of the world as containing many separate things from the opposing picture advocated by the British Idealists, such as F.H. Bradley, according to whom the world is an indivisible whole. His logical atomism justifies his method of analysis, for this method aims to reduce ordinary sentences to complex combinations of atomic sentences, sentences which feature expressions standing for Russell’s many separate things.

Wittgenstein and Russell disagree about the structure of atomic sentences. For Russell, atomic sentences consist of simple symbols which divide into proper names and predicates. Simple symbols have simple things (Russell’s logical atoms) as their meanings. The mark of a simple symbol is that understanding it consists in being acquainted with its meaning. Russell held that one cannot be acquainted with an ordinary object such as a chair, but only with the perceptual experiences, the sense-data, one has in perceiving the chair, such as the visual experience of a brown rectangular patch and the tactile experience of a rough and hard surface. He also allowed that one can be acquainted with the meanings of predicates, properties and relations. So atomic sentences feature proper names standing for sense-data, and predicates standing for properties and relations of sense-data. For example, ‘this is red’ is an atomic sentence in which the proper name ‘this’ stands for a sense-datum and the predicate ‘…is red’ stands for a property ascribed to the sense-datum.

If an atomic sentence is true, it is made true by a corresponding atomic fact which contains the meanings of the simple symbols of the atomic sentence. For example, ‘This is red’ is made true by the fact consisting of the

sense-datum named by ‘this’ having the property redness. But Russell’s inventory of facts far exceeds Wittgenstein’s totality of atomic facts. He also has negative facts which make false sentences false. ‘This is red’ is made false by the negative fact that the sense-datum is not red. Further, Russell has general facts corresponding to true general sentences such as ‘All men are mortal’ and ‘Some man is mortal’.

Whereas Wittgenstein remains silent about the nature of his simple objects, Russell claims that his simples are sense-data and their properties and relations. This entails that Russell cannot hold that atomic sentences are independent, because two atomic sentences ascribing different colours to a given sense-datum cannot both be true. Russell is more specific about his simples because his argument for their existence turns on an epistemic condition, the principle of acquaintance. This principle holds that the understanding of any sentence consists in acquaintance with the meaning units from which the meaning of the whole sentence is compounded. The principle is satisfied by an atomic sentence because understanding such a sentence consists in acquaintance with the meanings of its simple symbols. A non-atomic sentence must be completely analysed into a sentence which consists entirely of simple symbols standing for objects with which we are acquainted. Analysis thus provides a route whereby the analysed sentence can be understood. The surface form of an ordinary sentence is not a reliable guide to the underlying meaning units which make up the meaning of the sentence. For example, ‘Socrates is snub-nosed’ seems to feature a proper name, ‘Socrates’, standing for a person. But Russell claims that the analysis of such a sentence reveals that ‘Socrates’ is not really a proper name but a descriptive phrase picking out a complicated logical construction from sense-data.

Sense-data and their properties and relations are thus not only the fundamental units of meaning, but also the ultimate constituents of reality. Socrates and his properties and relations are reduced to logical constructions from these ultimate constituents. This reduction has the epistemological advantage embodied in Ockham’s razor. By taking Socrates to be a logical construction from sense-data, one reduces the risk of error because beliefs about sense-data and constructions from sense-data are supposed to be more secure than beliefs about objects which stand behind the sense-data.

Russell’s logical atomism was well-received by the logical positivists who confused it with Wittgenstein’s logical atomism (see Logical positivism §3). They fixed on Russell’s reductionist programme for translating empirical sentences into sentences about sense-data. This programme is fraught with difficulties both of principle and detail. One chief difficulty, promoted by Wittgenstein in his *Philosophical Investigations*, is the alleged incoherence of a private language consisting of reports of a single person’s sense-data (see Private language argument).

See also: Knowledge by acquaintance and description; Pluralism

ALEX OLIVER

References and further reading

Anscombe, G.E.M. (1959) *An Introduction to Wittgenstein’s Tractatus*, London: Hutchinson. (Chapter 1 argues against the view that Wittgenstein’s logical atoms are sense-data.)

Griffin, J.P. (1964) *Wittgenstein’s Logical Atomism*, Oxford: Clarendon Press.(A clear account of Wittgenstein’s logical atomism, emphasizing the influence of Hertz’s ideas about models in natural science.)

Pears, D.F. (1967) *Bertrand Russell & the British Tradition in Philosophy*, London: Fontana. (A detailed exposition of the development of Russell’s logical atomism.)

Russell, B. (1905) ‘On Denoting’, in *Logic and Knowledge*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 41-56.(Introduces the Russellian theory of definite descriptions which lies at the heart of Russell’s logical atomism.)

Russell, B. (1911) ‘Knowledge by Acquaintance and Knowledge by Description’, in *Mysticism and Logic*, London: Longman, 1918, 209-32.(States the principle of acquaintance.)

Russell, B. (1914) *Our Knowledge of the External World*, London: Allen & Unwin.(Chapters 1-4 contain the rudiments of the metaphysics of facts and logical constructions from sense-data.)

Russell, B. (1918) ‘The Philosophy of Logical Atomism’, in *Logic and Knowledge*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 177-281.(The version of Russell’s logical atomism sketched in this entry.)

Russell, B. (1924) ‘Logical Atomism’, in *Logic and Knowledge,* ed. R.C. Marsh, London: Allen & Unwin, 1956, 323-43.(A review of the principles of Russell’s logical atomism in which sense-data drop out of the picture.)

Urmson, J.O. (1956) *Philosophical Analysis,* Oxford: Clarendon Press.(An example of the conflation of Russell’s and Wittgenstein’s logical atomism, but useful for tracing the influence of logical atomism on the

logical positivists.)
Wittgenstein, L. (1921) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London: Routledge & Kegan Paul, 1974.(Wittgenstein's statement of his logical atomism.)
Wittgenstein, L. (1929) 'Some Remarks on Logical Form', supplement to *Proceedings of the Aristotelian Society* 9: 162-71.(Rejection of the independence of atomic sentences.)
Wittgenstein, L. (1953) *Philosophical Investigations*, trans. G.E.M. Anscombe, Oxford: Blackwell. (Paragraphs 1-242 undermine the fundamental principles of 'the author of the *Tractatus*'.)

Logical constants

A fundamental problem in the philosophy of logic is to characterize the concepts of 'logical consequence' and 'logical truth' in such a way as to explain what is semantically, metaphysically or epistemologically distinctive about them. One traditionally says that a sentence p is a logical consequence of a set S of sentences in a language L if and only if (1) the truth of the sentences of S in L guarantees the truth of p and (2) this guarantee is due to the 'logical form' of the sentences of S and the sentence p. A sentence is said to be logically true if its truth is guaranteed by its logical form (for example, '2 is even or 2 is not even'). There are three problems presented by this picture: to explicate the notion of logical form or structure; to explain how the logical forms of sentences give rise to the fact that the truth of certain sentences guarantees the truth of others; and to explain what such a guarantee consists in.

The logical form of a sentence may be exhibited by replacing nonlogical expressions with a schematic letter. Two sentences have the same logical form when they can be mapped onto the same schema using this procedure ('2 is even or 2 is not even' and '3 is prime or 3 is not prime' have the same logical form: 'p or not-p'). If a sentence is logically true then each sentence sharing its logical form is true. Any characterization of logical consequence, then, presupposes a conception of logical form, which in turn assumes a prior demarcation of the logical constants. Such a demarcation yields an answer to the first problem above; the goal is to generate the demarcation in such a way as to enable a solution of the remaining two.

Approaches to the characterization of logical constants and logical consequence are affected by developments in mathematical logic. One way of viewing logical constanthood is as a semantic property; a property that an expression possesses by virtue of the sort of contribution it makes to determining the truth conditions of sentences containing it. Another way is proof-theoretical: appealing to aspects of cognitive or operational role as the defining characteristics of logical expressions. Broadly, proof-theoretic accounts go naturally with the conception of logic as a theory of formal deductive inference; model-theoretic accounts complement a conception of logic as an instrument for the characterization of structure.

1 The Tarskian paradigm

Suppose we are given a theory of truth for a language *L*, a theory which explains how the truth-conditions of sentences of *L* are determined by their construction from the semantic properties of a set of primitive expressions. A semantic characterization of logical consequence and logical truth for *L* appeals to the idea of an 'interpretation' of *L*, that is, roughly, an assignment of semantic properties of the appropriate sort to the nonlogical (or 'lexical') primitive expressions of *L*, properties which, once assigned, determine a (unique) truth-value for each sentence in *L*. If *p* is a sentence and *S* a set of sentences of *L*, then *p* is said to be a logical consequence of *S* if *p* is true on each interpretation of *L* rendering every sentence of *S* true, and *p* is logically true if it is true on each interpretation of *L*.

The technical implementation of this idea for elementary logic goes back to Löwenheim (1915), but its full significance was appreciated only in the 1930s; it receives an important exposition in Tarski's 1936 paper 'Über den Begriff der logischen Folgerung' ('On the Concept of Logical Consequence'). Let *L* be an elementary first-order language. The logical constants of *L* are given by enumeration as the four connectives '$\neg$' (negation), 'Čamp;' (conjunction), '$\vee$' (disjunction), '$\rightarrow$' (material conditional) and the two quantifiers '$\forall$' (universal quantifier) and '$\exists$' (existential quantifier). In some cases, the identity symbol 'Čequals;' is also distinguished as a logical constant. The formulas of *L* arise in the following three ways: (1) by applying any primitive *n*-place predicate symbol *R* to *n* terms $t_1, \ldots, t_n$ to obtain a formula $R(t_1, \ldots, t_n)$; (2) by applying connectives to given formulas; and (3) where $x_0, x_1, \ldots$ is a fixed set of individual variables, by applying the quantifiers '$\forall$' and '$\exists$' to a variable x_n and formula p to give formulas $(\forall x_n)p$ and $(\exists x_n)p$. An interpretation of *L* consists, first, of a non-empty set *D*, its domain or universe; second, of an assignment of relations and functions of the appropriate type over *D* to the predicate and function symbols of *L*; and third, of an assignment of objects in *D* to each individual constant (or name) in *L*. If the identity predicate is treated as a logical constant, it is understood that the interpretation assigns to it the identity relation over *D*, that is, the set $\{\langle a, a\rangle : a \in D\}$.

If *J* is an interpretation of *L*, the meanings of the logical expressions determine, for each formula of *L*, when the

formula is true on *J* relative to an assignment of objects in *D* to the individual variables of *L*. Such an assignment may be represented by a sequence over *D*, that is, a function mapping the natural numbers $\{0, 1, 2, \ldots\}$ into *D*. A sequence *s* then expresses the assignment to each variable x_n of the object $s(n)$ from *D*. A formula which is true on *J* with respect to an assignment s is said to be 'satisfied' by *s* in *J*. The sequences which satisfy a given formula on *J* are determined by those which satisfy its component formulas (subformulas). Thus, for example, a sequence satisfies the conjunction of two formulas on *J* when it satisfies both conjuncts, and satisfies the negation of a formula when it fails to satisfy the formula. Finally, if $q = (\exists x_n)p$, a sequence *s* satisfies *q* on *J* if and only if there exists an assignment *s′* differing from *s* only in the value assigned to the variable x_n such that *s′* satisfies *p* in *J*. This expresses the idea that *q* should be true on an assignment if it is possible to vary the object assigned to the quantified variable in such a way as to make *p* true. An analogous clause may be given for the universal quantifier. The interpretation *J* is said to be a 'model' of a set *S* of formulas with respect to an assignment if each formula in the set is true in *J* with respect to that assignment; *S* is said to be 'consistent' if it has a model. Tarski then says that a formula *p* is a logical consequence of a set *S* of formulas if and only if for any interpretation *J* of *L* and assignment *s* in *J*, if *J* is a model of *S* with respect to *s*, then *p* is true in *J* with respect to *s*; equivalently, *p* is a logical consequence of *S* if the result of adding $\neg p$ to *S* is inconsistent. A formula is logically true if it is true in each interpretation of *L* with respect to each assignment; and two formulas are said to be logically equivalent when they are true in just the same interpretations relative to just the same assignments (see Tarski, A.; Model theory; Tarski's definition of truth).

2 Extensions of elementary logic

It is fair to say that the account of logical consequence and logical truth we have just sketched for elementary logic has set the standard for semantic accounts of these notions in other contexts. Typical of these are extensions of elementary logic that arise by treating additional quantifiers as logical constants.

A first-order quantifier *Q* may be regarded as describing a relation between a universe *D* (any non-empty set) and the subsets *E* of *D* such that the sentence $(Qx)(Ax)$ is true on an interpretation that assigns *E* to $A(x)$ over the universe *D*. Consider, for example, the quantifier symbol Q_0, read 'for infinitely many'. The sentence $(Q_0x)A(x)$ holds in an interpretation *J* if $A(x)$ holds in *J* relative to infinitely many values for the variable *x*. In effect, then, Q_0 describes the relation between a universe and its infinite subsets. Similarly, the quantifier Q_1 describes the relation between a universe and its uncountable subsets (if any), where a set is called 'countable' if it can be put into a one-to-one correspondence with a subset of the natural numbers. Q_0 and Q_1 are examples of 'generalized quantifiers' (see Quantifiers, generalized).

Elementary logic is characterized by three properties which are focal points of any evaluation of its extensions. The 'completeness theorem' for elementary logic states that if a formula *p* is a logical consequence of a set *S* of formulas in elementary logic, then *p* is in fact provable from *S* within the deductive systems for elementary logic usually considered. Combined with the soundness theorem, which states the converse property, this result has the effect of saying that the logical consequence relation for an elementary language can be represented syntactically: the consequences of any set are precisely the formulas that stand to that set in a syntactically characterized relation of deducibility. The 'compactness property' holds for a relation of logical consequence just in case a formula is a consequence of an infinite set *S* of formulas only if it is a consequence of some finite subset of *S*. Finally, the (downward) Löwenheim-Skolem theorem for elementary logic states that if a set of formulas has a model, then it has a model with a countable universe. The quantifiers Q_0 and Q_1 properly extend elementary logic in rather different ways: the consequence relation that arises by adding Q_0 to elementary logic has the Löwenheim-Skolem property but not the compactness or completeness properties; while the corresponding relation derived from Q_1 has both the completeness and compactness properties, but not the Löwenheim-Skolem property (see Keisler 1971). Other extensions of elementary logic have neither property. An example is second-order logic, which extends elementary logic by allowing quantification into predicate places (see Löwenheim-Skolem theorems and nonstandard models; Second-order logic, philosophical issues in).

3 Notions of invariance

In a model-theoretic definition of logical consequence, the logical constants are given simply by a list. The choice of one such list over another may seem, to some extent, ungrounded. Thus, Tarski wrote in 1936 that 'no objective

grounds are known to me which would permit us to draw a sharp boundary between the two groups of terms', that is, to solve the demarcation problem. However, in later years Tarski was led to a less sceptical attitude towards the demarcation problem (see Tarski 1966). The solution he suggested consisted of a very general condition - it is satisfied, in fact, by all of the generalized quantifiers usually considered - but one strong enough to exclude many of the expressions which are generally regarded as nonlogical. Let us formulate the requirement for the case of a quantifier symbol *Q* that applies to a formula $A(x)$ and variable *x* to yield a sentence $(Qx)(Ax)$. (The generalization of this constraint to other classes of expressions is straightforward and may be found in the literature - see McCarthy (1981: §3) and Sher (1991: ch. 4).) The meaning of *Q* must determine, for each chosen universe *D* and each interpretation of $A(x)$ over *D*, a truth-value for the sentence $(Qx)(Ax)$ over *D*. Now consider any one-to-one function mapping *D* on to an alternative universe D^*. Such a function fixes an interpretation for $A(x)$ over D^*, consisting precisely of the images under the function of the items in the given interpretation of $A(x)$ over *D*. The requirement is then that the two interpretations assign the same truth-value to $(Qx)(Ax)$. Such a function is said to constitute an 'isomorphism' of the one interpretation with the other; and so a succinct statement of the indicated condition is that the behaviour of a logical constant must be 'invariant under isomorphism of interpretations'. Tarski's suggestion is that the boundary between logical and nonlogical notions coincides with the boundary between the expressions which satisfy the present invariance condition and those which do not.

What is the significance of the invariance requirement? Interpretations are 'isomorphic' when they share the same structure, so the invariance condition says roughly that the semantic result of applying a logical constant should depend only on the structure of the interpretation of the expressions to which it is applied, independently of properties and relations specific to objects in any particular interpretation. This requirement may be used to confer a clear sense on one very traditional thesis about logical notions, namely that they are 'topic-neutral' concepts. Tarski noticed that this sort of topic-neutrality is related to a unifying idea behind certain generalizations of the notion of a 'geometry', an idea developed by the German mathematician Felix Klein (1872). Klein viewed a space *S* as being given by a set with a certain geometric structure. Consider a collection G of one-to-one transformations mapping this set onto itself. Geometric structures can in many cases be described by families of such transformations. For example, consider the collection of all transformations that can be defined in terms of rotations and linear motions (translations). Then the properties of (plane) Euclidean geometry are characterized by the fact that they are 'invariant under transformations of this type', in the sense that any such transformation maps a set *S* of points on to a set congruent to *S* and, conversely, any set congruent to *S* can be obtained from *S* by a transformation of this type. Similarly, topology, a generalized geometry, may be viewed as the study of those properties which are preserved by one-to-one transformations mapping a space 'continuously' (intuitively, without breaking or tearing it) onto itself. Tarski's idea was to use the limiting case of this situation to characterize logic: logic is the study of those properties of a 'space' (that is, a set or universe) which are preserved by *any* one-to-one transformation mapping that space onto itself or, more generally, onto some other set. The indicated invariance requirement for *Q* may be seen as the special case of Tarski's constraint where we focus on relations of a universe and its subsets of the sort described by quantifiers.

The invariance condition is generally regarded as necessary for logical status. The question we must now consider is whether it is also sufficient. One difficulty faced by an affirmative answer concerns the necessitation requirement: that it is not possible - in a strong, uniquely logical sense of 'possible' - that each member of a set of sentences is true and one of its logical consequences is false. The difficulty is that there appear to be expressions which are counted as logical by the invariance criterion which lead to consequence relations violating this requirement. For example, let *P* represent a fixed contingently false sentence of English, for example, 'The French Revolution failed'. Consider a quantifier symbol *Q* interpreted as follows: in any interpretation *J*, $(Qx)(Ax)$ holds in *J* if *P* and $A(x)$ holds for some member of *J*, or not-*P* and $A(x)$ fails for each member of *J*. Since *P* is in fact false, if *Q* were a logical constant then $(Qx)(Ax)$ would be logically equivalent to $(\forall x)\neg A(x)$. This is a violation of the necessitation requirement, since in a logically possible world in which *P* is true, $(Qx)(x = x)$ is true, but $(\forall x)\neg(x = x)$ is false.

However, it may be that the lesson to be learned from examples such as the one just considered is not that the invariance constraint fails to underwrite the necessitation requirement, but that the invariance requirement has been understood too narrowly (see Sher 1991; McCarthy 1987). The example would not arise if that requirement were

understood ‘rigidly’, that is, in such a way as to apply to the interpretation of *Q* ‘across all logically possible worlds’. For consider a world w_1 in which *P* is true containing the same objects as a world w_2 in which *P* is false. Then the transformation that associates each object with itself will map the interpretation of the formula x = x in w_1 on to its interpretation in w_2 (both are the whole universe), but the sentence $(Qx)(x=x)$ is true in w_1 and false in w_2, which shows that the invariance condition for *Q* breaks down if that condition is applied to the interpretation of *Q* across all logically possible worlds. Let us say that an expression satisfying this stronger invariance condition is ‘rigidly invariant’. The suggestion is then that rigid invariance be taken as the defining property of the logical constants. The problem created by this proposal lies in a threat of circularity: we would require a characterization of the notion of logical possibility, sufficient for a demarcation of the logically possible worlds, not dependent upon a prior demarcation of the logical constants.

4 Syntactic characterizations of logical notions

An alternative approach to the demarcation problem locates the distinctive feature of the logical constants in the role that they play in deductive arguments. On this view, a logical expression is one whose meaning can be characterized by a set of rules of inference of a certain sort. These rules may be formulated in different ways; for example, in terms of a sequent calculus, or a natural deduction system (see Natural deduction, tableau and sequent systems). Each formulation will incorporate in some way a distinction between rules that license the introduction of a logical expression in an inference (introduction rules), and those which describe the conditions under which it can be eliminated (elimination rules). Thus, for example, the introduction condition for conjunction may be given by the rule

$A, B/A \mathbin{\&} B,$

and its elimination conditions by the rules

$A \mathbin{\&} B/A, \quad A \mathbin{\&} B/B.$

A rule of the form

$X_1, \ldots, X_n/Z$

is to be read as an inference-licence, saying that if $X_1, \ldots, X_n$ have been inferred, then it is permissible to infer Z. The existential quantifier may be similarly characterized by the rules

$A(t)/(\exists x)A(x), \quad (\exists x)A(x)/A(c),$

where in the first rule t is any term and in the second c is a constant subject to the restriction that it not appear previously in the inference in question (the ‘non-occurrence restriction’). Analogous rules characterize the other notions of classical quantification theory.

The most straightforward reading of the syntactic view is that an expression counts as logical if and only if it is possible to write a set of introduction and elimination rules for it. Arthur Prior criticized this idea by means of a simple counterexample (1960). In the absence of substantive restrictions governing the interaction of introduction and elimination rules, it would seem that we could characterize a connective ‘tonk’ by means of the following pair of rules:

$A/A \text{ TONK } B, \quad A \text{ TONK } B/B,$

from which we obtain the schema A/B, licensing the inference of any sentence from any other, as a derived rule of inference! In order to avoid this sort of pathology, Belnap (1961) imposed the requirement that adding a logical constant to a given deductive system *S* by reference to a set of rules of inference result in a ‘conservative’ extension of *S*; one in which no formula in the notation of *S* is provable that is not already provable in *S*.

Even thus amended, however, the syntactic conception of what it is to be a logical constant would appear to be too liberal. Let *S* be a standard system of proof for elementary logic, and suppose that *P* represents a fixed nonlogical predicate. Consider the quantifier *Q* such that $(Qx)(Ax)$ holds in any interpretation *J* if and only if some object falling under *P* satisfies $A(x)$ in *J*. (Thus, for example, if *P* represents ‘…is a bird’, the quantificational expression *Q* means ‘For some bird…’.) Let S^+ result from *S* by adding the following introduction and elimination rules for *Q*:

$A(t), P(t)/(Qx)A(x) \quad (Qx)A(x)/A(c), P(c),$

where in the second rule *c* is an individual constant subject to the non-occurrence requirement. It may be shown that S^+ is a conservative extension of *S*, but the expression *Q* thus characterized is not a logical constant.

Belnap’s requirement may be seen as a by-product of a later viewpoint on syntactic characterizations of logical notions developed by Ian Hacking (1979). On this view, the fundamental rules of inference associated with the logical constants implicitly define their semantic interpretation. The three rules given above for conjunction determine its semantic interpretation in the following sense: there is precisely one truth table for ‘Čamp;’ with respect to which these rules are sound. The introduction and elimination rules for the existential quantifier can similarly be shown to characterize its semantic interpretation. It is clear that the indicated rules for *Q* do not thus characterize its semantic role, for its interpretation over any set varies with the interpretation assigned to *P*. Of course, to speak of a set of rules of inference as ‘characterizing’ the interpretation of a connective or a quantifier symbol presupposes that such an expression is to have an interpretation of the indicated (classical) sort; the justification of this assumption must be given in other terms. The idea is that in ascribing such an interpretation to an expression we look for the one - or *a* one - that makes the introduction and elimination rules associated with the expression truth-preserving; the expression counts as logical if over any universe there is only one such interpretation.

What is the relation between the semantic and syntactic solutions to the demarcation problem? One answer is that they correspond to two broad conceptions of the nature of logic (see Tharp 1975). On the first conception, logic is a theory of a certain type of deductive inference. On this view, what is distinctive about logical notions is their role in deductive inference, and logical consequence consists in a (perhaps idealized) type of deducibility. The syntactic view (on which, under very general conditions, a completeness result links a relation of logical consequence to the deductive system that characterizes the logical constants) provides a natural interpretation of this conception. An alternative conception, of more modern vintage, sees logic as an instrument for the characterization of structure. A logic (that is, a choice of logical constants) may be said to ‘characterize’ a structure if there is a set *S* of sentences in the logic such that any model of *S* exemplifies that structure. On this conception, the invariance criterion would seem to be very natural, for the behaviour of logical notions should be preserved under ‘structure-preserving’ transformations of any subject matter to which they are applied. From this point of view, there will be no particular premium on completeness, and the Löwenheim-Skolem property is an anomaly. Elementary logic is a paradigm of the first conception of logic and second-order logic typifies the second.

See also: Logical and mathematical terms, glossary of; Logical form; Logical laws; Proof theory

TIMOTHY McCARTHY

References and further reading

Barwise, J. and Feferman, S. (1985) *Model-Theoretic Logics*, Berlin: Springer.(An excellent sourcebook in generalized model theory, containing detailed treatments of all the technical issues touched on in this entry.)

Belnap, N. (1961) ‘Tonk, Plonk, and Plink’, *Analysis* 22: 130-4.(Addresses problems with syntactic criteria for interpreting logical constants in terms of introduction and elimination rules.)

Etchemendy, J. (1990) *The Concept of Logical Consequence*, Cambridge, MA: Harvard University Press.(An important critical evaluation of the Tarskian paradigm, emphasizing problems pertaining to the necessitation requirement.)

Hacking, I. (1979) ‘What is Logic?’, *Journal of Philosophy* 76 (6): 499-523.(Develops a sophisticated proof-theoretic account of the nature of logical constants.)

Harman, G. (1986) ‘The Meaning of Logical Constants’, in E. Lepore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford and New York: Blackwell, 125-34.(Explains the meaning of logical constants in terms of ‘immediate implications’ and ‘immediate incompatibilities’.)

Keisler, H.J. (1971) ‘Logic with the Quantifier "there exist uncountably many"’, *Annals of Mathematical Logic* 1: 1-93.(Gives a completeness result for logic with the ‘uncountability’ quantifier, by means of a particularly nice axiomatization.)

Klein, F. (1872) *Vergleichende Betrachtungen über neuere geometrische Forschungen*, Erlangen: Deichert; repr. in *Gesammelte mathematische Adhandlungen*, Berlin: Springer, 1921, vol. 1, 460-97; trans. M.W. Haskell, ‘A Comparative Review of Recent Geometric Researches’, *Bulletin of the New York Mathematical Society* 2:

215-49, 1993.(The inspiration for Tarski 1966; of considerable interest in itself.)
Lindström, P. (1969) 'On Extensions of Elementary Logic', *Theoria* 31: 1-11. (A seminal article in generalized model theory, establishing that elementary logic is characterized by the completeness, compactness and Löwenheim-Skolem properties.)
Löwenheim, L. (1915) 'Über Möglichkeiten im Relativkalkül', *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, 'On Possibilities in the Calculus of Relatives', in J. van Heijenoort, *From Frege to Gödel: A Sourcebook in Mathematical Logic*, Cambridge, MA: Harvard University Press, 1967. (A seminal paper in the theory of models, giving the first statement and proof of the Löwenheim-Skolem theorem for elementary logic.)
McCarthy, T. (1981) 'The Idea of a Logical Constant', *Journal of Philosophy* 78 (9): 499-523. (Discusses the invariance criterion and its problems, and includes an extension of the criterion to intensional logics.)
McCarthy, T. (1987) 'Modality, Invariance, and Logical Truth', *Journal of Philosophical Logic* 16: 423-43. (Establishes a general result saying that on weak assumptions if a set of notions is invariant across a class of possible worlds, then a logic based on those notions will satisfy a necessitation requirement with respect to that class.)
Peacocke, C. (1987) 'Understanding Logical Constants', *Proceedings of the British Academy* 83: 153-200. (Explains what is special about logical constants in terms of the way they are understood.)
Prior, A. (1960) 'The Runabout Inference-Ticket', *Analysis* 21: 38-9. (Criticizes the thesis that logical constants are generally characterized in terms of introduction and elimination rules.)
Sher, G. (1989) 'A Tarskian Conception of Logic', *Pacific Philosophical Quarterly* 70: 341-68. (An illuminating discussion of Tarski's later views on the demarcation problem.)
Sher, G. (1991) *The Bounds of Logic*, Cambridge, MA: MIT Press/Bradford. (A general discussion of problems in the philosophy of classical logic from a model-theoretic point of view.)
Tarski, A. (1936) 'Über den Begriff der logischen Folgerung', *Actes du Congres International de Philosophie Scientifique* 7: 1-11; repr. 'On the Concept of Logical Consequence', in *Logic, Semantics, Metamathematics*, trans. and ed.J.H. Woodger, Oxford: Oxford University Press, 2nd edn, 1983. (An early important informal discussion of the semantic conception of logical consequence.)
Tarski, A. (1966) 'What are Logical Notions?', ed. and with intro. by J. Corcoran, *History and Philosophy of Logic* 7: 143-54. (An illuminating discussion of notions of invariance in relation to logic.)
Tharp, L. (1975) 'Which Logic is the Right Logic?', *Synthese* 31: 1-21. (An excellent introduction to the problems treated in this entry.)

Logical form

Consider the following argument: All men are mortal; Socrates is a man; therefore, Socrates is mortal. Intuitively, what makes this a valid argument has nothing to do with Socrates, men or mortality. Rather, each sentence in the argument exhibits a certain 'logical form' and those forms, taken together, constitute a pattern that guarantees the truth of the conclusion given the truth of the premises. More generally, the logical form of a sentence of natural language is what determines both its logical properties and its logical relations to other sentences.

The logical form of a sentence of natural language is typically represented in a theory of logical form by a well-formed formula in a 'logically pure' language whose only meaningful symbols are expressions with fixed, distinctly logical meanings (for example, quantifiers). Thus, the logical forms of the sentences in the above argument would be represented in a theory based on pure predicate logic by the formulas '$\forall x(Fx \rightarrow Gx)$*', 'Fy' and 'Gy', respectively, where 'F' and 'G' are free predicate variables and 'y' a free individual variable. The argument's intuitive validity is then explained by the fact that the logical forms of the premises formally entail the logical form of the conclusion. The primary goal of a theory of logical form is to explain as broad a range of such intuitive logical phenomena as possible in terms of the logical forms that it assigns to sentences of natural language.*

1 Logical form in Aristotelian logic

Broadly speaking, the 'logical form' of a sentence of natural language is what determines both its logical properties and its logical relations to other sentences. Most notably, the logical form of a sentence *S* determines whether or not it is logically true, and the logical forms of the sentences in a set *K* together with the logical form of *S* jointly determine whether or not the latter is a logical consequence of the former.

Although particularly prominent in the twentieth century, due especially to the influence of Bertrand Russell, the idea of logical form can be traced back to Aristotle. In the so-called 'traditional logic' that stems from Aristotle's work in the *Prior Analytics*, the central form of argument is the 'syllogism', that is, an argument consisting of two premises and a conclusion. For example:

(1) Every whale is a mammal.
Some carnivore is a whale.
$\therefore$ Some carnivore is a mammal.

Aristotle was the first to recognize explicitly that the validity of a syllogism (that is, its conclusion being a logical consequence of its premises) is entirely independent of the common noun phrases, or 'terms', in its constituent sentences. Rather, each sentence in a valid syllogism such as (1) above exhibits a certain form and, taken together, those forms constitute a pattern that, of itself, guarantees the truth of the conclusion given the truth of the premises.

To capture this idea systematically, Aristotle first restricted his attention to syllogisms the constituent sentences of which exhibit any of four basic sentential forms, known traditionally as *A*, *E*, *I* and *O*, respectively: 'Every α is a β', 'No α is a β', 'Some α is a β' and 'Some α is not a β', where 'α' and 'β' represent the roles of the subject and predicate terms in such sentences. Thus, the first premise of (1) is of the sentential form *A*, and both the second premise and the conclusion are of the form *I*. These sentential forms alone, however, are not enough to characterize the pattern that (1) exhibits. A further logically relevant feature of (1) is the way in which the terms occur in its constituent sentences: 'carnivore', for instance, is the subject term of both the second premise and the conclusion but does not occur at all in the first premise. To capture this feature generally, Aristotle introduced schematic variables *F*, *G*, *H*,…. Call the result of replacing the terms in a sentence that exhibits one of the basic sentential forms with distinct schematic variables a 'schematic' form. The pattern exhibited by a given syllogism can then be represented by replacing its constituent sentences with schematic forms that preserve the arrangement of terms in the syllogism. For example, (1) exhibits the following pattern:

(2) Every *F* is a *G*.
Some *H* is an *F*.
Some *H* is a *G*.

By sole means of representations such as (2), Aristotle proved, for each possible syllogistic pattern, whether or not it is valid; that is, for Aristotle, whether or not it is impossible for the premises of any instance of the pattern to be true and the conclusion false. Thus, the schematic forms of its premises and conclusion (relative to a given choice of schematic variables) completely determine the validity or invalidity of each pattern. Aristotle's schematic forms are thus paradigms of logical forms.

The central goal of a theory of logical form is to explain the logical properties of (and relations between) the members of a broad class *K* of sentences of natural language in terms of the logical forms that the theory assigns to those sentences. With only four general types of logical form to choose from - the four basic sentential forms, the scope of Aristotle's account as it stands is far too limited to count as a fully fledged theory of logical form. However, a perusal of any modern text with a section on traditional logic yields a variety of techniques for translating sentences with entirely different grammatical forms into instances of Aristotle's four basic sentential forms. For example, a simple individual assertion such as

(3) Matthew is a politician

is obviously not an instance of any of the basic sentential forms, since it does not begin with one of the quantifiers 'Every', 'Some' and 'No'. But by introducing a term that picks out a class containing only Matthew, (3) can be translated into the intuitively equivalent, albeit somewhat stilted, sentence

(4) Every person identical to Matthew is a politician,

which exhibits the sentential form *A*. Again, a sentence such as

(5) Someone gave *The Brothers Karamazov* to Andrea,

which involves a transitive verb, can be put into the appropriate sentential form by replacing the verb with its nominal '-er' counterpart followed by the preposition 'of', to produce a term that picks out exactly the class of individuals to which the original predicate applies. (5) then gives way to

(6) Some person is a giver of *The Brothers Karamazov* to Andrea.

Supplemented by such techniques, Aristotle's logic has a considerably broader application, as a much larger class of sentences can be assigned logical forms than would initially appear to be the case. So supplemented, then, Aristotle's account is plausibly taken to be the first genuine theory of logical form (see Logic, ancient §§1-3).

2 A general characterization of logical form

Although differing considerably in content and detail, the basic components and characteristics of Aristotle's theory of logical form (in this somewhat anachronistic rendering) are essentially the same as in contemporary accounts deriving from Bertrand Russell. In this section, these components and characteristics are made explicit.

In general, the notion of logical form is always relative to a theory *T* of logical form which is directed towards some broad class *K* of sentences of natural language. Call *K* the 'target class' of the theory. The logical forms of the members of *K* are represented by formulas in a 'logically pure' canonical language; that is, a language that contains (perhaps in addition to punctuation) only variables and logical constants - expressions with fixed and, according to the theory, distinctly logical meanings (see Logical constants §1). For Aristotle, this is the language of schematic forms the logical constants of which are 'Every', 'Some', 'No', 'is a' and 'is not a'; and for Russell and his followers it is generally (some variant of) the language of *Principia Mathematica* (1910-13), with its now familiar quantifiers and propositional connectives.

The route from the sentences of the target class *K* to their logical forms according to *T* consists of two steps. First, and most important, is the translation of the members of *K* into an 'impure' hybrid language that consists of the canonical language supplemented by nonlogical constants: for Aristotle, ordinary terms such as 'carnivore' and also stilted terms such as 'person identical with Matthew'; for Russellians, individual constants such as 'Socrates' and *n*-place predicates such as 'mortal' (or perhaps abbreviations thereof such as '*s*' and '*M*'). These nonlogical constants carry the same meaning as their informal counterparts and are used to construct the hybrid language translations of the sentences in which those counterparts occur. The translation of a sentence of *K* into a hybrid

language is typically known as its ‘analysis’. (Sentences such as ‘Every boy danced with a girl’, which are logically ambiguous, should of course receive a distinct analysis for each possible reading.) The analysis of a sentence is a paraphrase of the sentence which displays its logical form overtly in its surface grammatical structure. Thus, (3) and (5) are translated as (4) and (6), respectively, while for the Russellian, (7), for example, is translated as (8):

(7) If every man is mortal and Socrates is a man, then Socrates is mortal.
(8) ($\forall y$(man(y) $\rightarrow$ mortal(y)) & man(Socrates)) $\rightarrow$ mortal(Socrates).

In the second step of the process, the logical forms for the sentences of *K* are derived simply by uniformly replacing all the nonlogical constants in the analyses of those sentences with variables of the appropriate type. So, for example (assuming a choice of replacement variables for the nonlogical constants), the sentences in (1) - which in this case are identical with their analyses - yield the logical forms in (2), and (8) likewise yields (9) as the logical form of (7):

(9) $(\forall y(Fy \rightarrow Gy)\ \&\ Fx) \rightarrow Gx.$

As noted, the primary goal of a theory *T* of logical form is to account for the logical properties of the sentences in its target class *K* in terms of their logical forms. This is accomplished by means of a logical theory for the canonical language of *T* in which formal correlates of the ordinary logical notions - logical truth, logical consequence and so on - are defined for the formulas of the language. The apparent logical properties of the members of *K* are then explained (or explained away) by virtue of the fact that the formal correlates of those properties hold (or fail to hold) among the logical forms of the members of *K*. Thus, once again, the intuitive validity of (1) is explained by the formal validity of (2). Similarly, for the Russellian, the intuitive logical truth of (7) is explained by the formal logical truth of (9) - typically understood as truth in any domain under any interpretation (relative to that domain) of the free variables ‘*F*’, ‘*G*’ and ‘*x*’, following the work of Tarski (1933; see Model theory). In either case, the actual meanings of the nonlogical expressions in the natural language sentences in question play no role in determining the logical properties of those sentences. Rather, as the theories demonstrate, the logical properties are determined by their logical forms alone.

3 Comparing theories of logical form

The central principle for comparing theories of logical form is the following:

CP *T′* is preferable (*prima facie*) to *T* if *T′* explains a greater range of logical behaviour than *T* without appeal to extralogical principles (or, ‘meaning postulates’).

To illustrate, intuitively (10) follows directly from (5):

(10) Someone gave something to Andrea.

And yet, in Aristotle’s theory, as noted, (5) must be analysed along the lines of (6) and (10) along the lines of (11):

(11) Some person is a giver of something to Andrea.

However, then the analyses (6) and (11) exhibit only the obviously invalid argument pattern

(12) Some *A* is a *B*.
Some *A* is a *C*.

To capture the original entailment, (12) must be treated as enthymematic: that is, there is a suppressed premise with the logical form ‘Every *B* is a *C*’, namely

(13) Every giver of *The Brothers Karamazov* to Andrea is a giver of something to Andrea.

Aristotle’s logic, therefore, explains the relation between (5) and (10) by introducing an extralogical principle linking the two predicate terms, in effect reducing it to an instance of the argument pattern (2).

By contrast, represented in predicate logic, the logical forms of (5) and (10) constitute the valid pattern

(14) $\exists x Hxyz$
$\exists y \exists x Hxyz$

Hence, that (10) is a consequence of (5) is explained on the basis of their logical forms alone, without appeal to any extralogical principles. So, *prima facie*, a theory based on predicate logic is preferable to Aristotle's theory.

The '*prima facie*' qualification in CP marks the influence of other factors besides extent of explanatory range that are relevant to an evaluation of competing theories of logical form. Two particularly influential principles can be observed in recent debates over the logical forms of one or another class of sentences: logical conservatism and ontological conservatism.

The shift from Aristotle to Russell exhibits a typical way in which one theory *T′* of logical form can broaden the explanatory range of another *T*; namely, by extending or otherwise subsuming the canonical language and logic of *T*. In such cases, the language of *T′* is able to provide logical forms for all sentences in the target class of *T*. In addition, the language of *T′* includes logical constants and classes of complex expressions not present in the language of *T* that enable *T′* to construct logical forms not available to *T*. Appropriate logical principles for its new forms then enable *T′* to explain logical phenomena that are left unexplained (without extralogical principles) in *T*, for example, (10)'s being a logical consequence of (5). The principle of logical conservatism stipulates that supplementation to a given canonical language should be kept to a minimum. More exactly, let *T* and *T′* be theories with roughly equal target classes and let *L* be a background logical theory (first-order logic, for example) that is common to both *T* and *T′*; then

LC *T′* is preferable (*prima facie*) to *T* if *T′* requires fewer extensions to *L* than *T*.

The increase in explanatory range of predicate logic, of course, is far too vast for LC to override CP in the choice between Aristotle and Russell. A better example is found in recent well-known work on the logical form of action sentences with adverbial modifiers. For example:

(15) April kissed Jonathan tenderly.

It is intuitively clear that (15) entails

(16) April kissed Jonathan.

Typical predicate logic analyses of these sentences represent 'kissed' and 'kissed tenderly' as distinct two-place predicates. The entailment is then explained by means of a meaning postulate to the effect that if *A* kisses *B* tenderly, then *A* kisses *B*. As with Aristotle's explanation of the relation between (5) and (10), then, in this explanation the apparent entailment is actually enthymematic. However, the entailment can be explained directly by treating an adverb such as 'tenderly' as an adverbial operator *t* that, when prefixed to an *n*-place predicate *F* (such as 'kissed'), yields a new *n*-place predicate [*tF*] ('tenderly kissed') (see Adverbs §1). The logic of these new constructions is then characterized generally by the principle that

(17) $[\alpha\Pi]x_1 \ldots x_n \rightarrow \Pi x_1 \ldots x_n$, for any adverbial operator α and *n*-place predicate Π.

The logical forms of (15) and (16), then, on this approach, are (18) and (19), respectively:

(18) $[tF]xy$
(19) *Fxy*

and so by the new logical principle (17), (16) follows from (15) directly in virtue of their logical forms.

Most philosophers would probably agree that this increase in explanatory range is significant enough to override LC and warrant the added apparatus. However, Davidson (1967) proposed a logically more conservative analysis which avoids the new apparatus. Specifically, for Davidson, the proper analysis of action sentences takes the structure of an action verb such as 'kissed' to involve an implicit parameter for an event. Thus, (15) is to be analysed as

(20) $\exists x$ (kissing-of(April, Jonathan, *x*) & tender(*x*))

(read, roughly, as ‘There is an event *x* such that *x* is a kissing of Jonathan by April and *x* is tender’), and (16) as

(21) $\exists x$ (kissing-of(April, Jonathan, *x*))

The entailment from (15) to (16) is then explained by standard logical principles governing conjunction and the existential quantifier, and the apparently hasty introduction of new constructions and unfamiliar logical principles is avoided. Since their explanatory ranges are the same, then, CP provides no support for the adverbial operator account over Davidson’s, and so, by LC, Davidson’s account is to be preferred (see Adverbs).

As Davidson’s account illustrates, however, conservatism with regard to logic is often accompanied by liberalism with regard to ontology - in this case, the postulation of events. The ‘ontological commitments’ of a theory *T* of logical form consist of the kinds of things that must exist if the analyses that *T* assigns to the sentences of its target class are to be meaningful (see Ontological commitment). The second principle - ontological conservatism - is that such commitments are to be kept to a minimum; that is, more exactly, where again *T* and *T′* have roughly equal target classes,

OC *T′* is preferable (*prima facie*) to *T* if *T′* has fewer ontological commitments than *T*.

Unlike LC, this principle favours the operator account of adverbial modification over Davidson’s more ontologically permissive account. The choice between the two thus turns on one’s preferred brand of conservatism.

LC tends to be overridden by CP when they conflict, as additional apparatus is usually viewed as a small price to pay for greater explanatory range. However, more is often at stake in conflicts between CP and OC, as theories of logical form that have great explanatory range often exact a high price in ontological commitment: possibilia (that is, possible but non-actual objects) are introduced to explain sentences involving modality (see Modal logic, philosophical issues in §§1-2), intensional entities to explain the logic of attitude verbs (see Propositional attitude statements §§1-2) and so on. One must either choose between the two or offer a competing account with the same explanatory range but fewer ontological commitments. Since Russell first formulated his theory of descriptions (see Descriptions) to counter what he saw as Meinong’s ontological excesses, the development of competing theories of logical form has been, and largely remains, the primary tool for metaphysical discovery and the central approach to metaphysical debate in the twentieth century (see Analytical philosophy).

See also: Logical and mathematical terms, glossary of; Montague, R.M.; Quine, W.V.

CHRISTOPHER MENZEL

References and further reading

Aristotle (*c*.350s-330s BC) *Prior Analytics*, trans. and with commentary by R. Smith, Indianapolis, IN: Hackett Publishing Company, 1989, esp. I.1-7.(Referred to in §1. The primary source for Aristotle’s formal logic.)

Davidson, D. (1967) ‘The Logical Form of Action Sentences’, in N. Rescher (ed.) *The Logic of Decision and Action*, Pittsburgh, PA: University of Pittsburgh Press, 81-95; repr. with replies to critics in D. Davidson, *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, esp. 137-46.(The source of Davidson’s account of action sentences with adverbial modifiers; discussed in §3.)

Harman, G. (1972) ‘Logical Form’, *Foundations of Language* 9: 38-65; repr. in D. Davidson and G. Harman (eds) *The Logic of Grammar*, Encino, CA: Dickenson, 1975, 289-307.(A fine discussion of theories of logical form on which the comparative principles in §3 are based.)

Russell, B.A.W. (1914) ‘Logic as the Essence of Philosophy’, in *Our Knowledge of the External World*, London: Allen & Unwin, 33-53; London: Routledge, 1993. (Perhaps Russell’s strongest statement of his view of the role of logical form in logic and philosophy.)

Russell, B.A.W. (1919) *An Introduction to Mathematical Philosophy*, London: Routledge, 1993. (An engaging if not always wholly clear account of the basic issues in philosophical logic and the philosophy of mathematics. Chapter 16 is a classic account of Russell’s theory of definite descriptions; chapter 18 includes a lengthy discussion of logical form.)

Tarski, A. (1933) *Pojęcie prawdy w jezykach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956), ‘On the Concept of Truth in Formalized Languages’, in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(Referred to in §2. The classic work of

model-theoretic semantics.)
Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(Referred to in §2. Probably the most influential work in logic in the twentieth century; extremely technical.)
Wittgenstein, L.J.J. (1922) *Tractatus Logico-Philosophicus*, trans. C.K. Ogden and F.P. Ramsey, London: Routledge; trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1975.(A unique, highly influential work by Russell's most famous student in which a very interesting, somewhat eccentric notion of logical form plays a prominent role.)

Logical laws

There are at least three different kinds of answer to the question 'What is a logical law?' One establishes what it means for something to be a logical law. This answers the semantic question: What is the meaning of 'logical law'? The second explains what makes something a logical law. This answers the metaphysical question: What is the ground of logical law? The third tells you what the logical laws are. This answers the question: What is the extension of 'logical law'?

Even though logic is often seen as a complete science, the answers to all three questions are disputed. For example, there are at least three different conceptions of what it means for something to be a law of logic. Different conceptions account for logic in terms of necessity, truth in all models, and proof.

There are also different answers to the metaphysical question. If truth-preservation is central to logic, then the ground of logic depends on the metaphysics of truth. If logic is a matter of the meanings of terms, then the metaphysics of meaning is important for logic. Unfortunately, there is no widespread agreement on the metaphysics of meaning or truth.

Finally, there is no widespread agreement as to what the logical laws are. There are two general disputes here. First, it is not clear what notions count as logical. Does logic contain laws about identity, second-order quantification or modality? Second, given agreement on the scope of logic, there are still questions about the logical laws in that area. Intuitionists, quantum logicians, relevance and paraconsistent logicians each reject things taken as laws by others, even in the language of 'and', 'or' and 'not'.

1 The semantic question

When we claim that something is a logical law, what do we mean? Sometimes we use 'logical law' to mean 'theorem of a formal system'. Depending on taste, we might call theorems of first-order classical logic, or some other formal system, laws of logic. This is a derivative use of the phrase, akin to calling the claims of a particular scientific theory laws of nature. The propositions of a theory may be laws of nature, but only if the theory 'gets it right'. The theory must describe the world in the right way for its claims to be laws of nature. Similarly, theorems of a formal system are only logical laws if they 'get it right' in some appropriate way. The interest consists in elaborating what it means for a theory to 'get it right'. So, by 'logical law' we do not simply mean 'theorem in a particular formal system'.

Another non-answer to this question is that logical laws are the ways we cannot help but think. Anyone familiar with the history of logic in the twentieth century will be aware that almost any principle that some take to be a logical law, others reject as invalid. If a logical law is something to which no one can help but assent, nothing counts as a logical law. No purely psychological answer, in terms of 'laws of thought', will give us logical laws.

Now to a more plausible account. The goal in the study of logic is an account of deductively valid inferences: these are the logical laws. A deductively valid inference is one in which necessarily, if the premises are true, so is the conclusion. Laws of logic are those inferences that are necessarily truth-preserving. This conception goes back to Aristotle (see Logic, ancient §§1-3).

> A syllogism is a form of words in which when certain assumptions are made, something other than what has been assumed necessarily follows from the fact that the assumptions are such.
>
> (*Prior Analytics*: 24b18)

We are interested in necessarily truth-preserving inferences because given these, and given true premises from which to infer, we will never (and in fact, *can* never) deduce falsehood. Given our concern for finding truth and avoiding falsehood, it is easy to see why logical laws are interesting, on this picture. It is also clear how logic can be normative. Given that we have the goal of gaining truth and avoiding falsehood, we *ought* to deduce validly. (Which is not to say that we ought not reason in other ways as well.)

We have one answer to our semantic question: a logical law is a necessarily truth-preserving inference. However, many have found this answer unsatisfactory. The main reason for dissatisfaction with this account is the reliance upon the intensional notion of necessity. Some have sought to give an account of logical laws that has no recourse

to necessity, or to any other intensional notion.

There are two important analyses of logical consequence which, at least at first glance, avoid using the notion of necessity. One was given its canonical exposition by Tarski in his 1936 essay ‘Über den Begriff der logischen Folgerung’ (’On the Concept of Logical Consequence’), though it was prefigured by Bolzano in 1837. Tarski wrote ‘The sentence *X follows logically* from the sentences of the class *K* if and only if every model of the class *K* is also a model of the sentence *X* ([1936] 1956: 417; original emphasis).

There are no modal notions here, provided that there is an adequate non-modal explanation of what it is to be a model. There are a number of ways to do this, and there is no need for us to consider them in detail. Suffice to say that models are structures in which each of the non-logical constants in a language is given an interpretation in terms of the objects in the model, and the logical constants have a fixed interpretation in each model (see Logical constants §1; Model theory).

Model-theoretic notions of consequence do not, as they stand, make any appeal to modality. However, they do not give a modal-free account of necessary truth-preservation.

For example, the model-theoretic account needs a collection of logical terms. If *all* terms are logical, then the interpretation of each term is fixed, and there is only one model. This would make logical truth vacuous, as Tarski realized (1936). Now suppose that the terms ‘blue’ and ‘coloured’ are not counted as logical terms. The deduction ‘My car is blue; therefore my car is coloured’ fails to be valid on the model-theoretic account, because we can assign disjoint extensions to the predicates ‘blue’ and ‘coloured’. In contrast, the inference is necessarily truth-preserving. In this way, necessary truth-preservation and model-theoretic validity come apart. (For a detailed account of how the notions differ, see Etchemendy’s *The Concept of Logical Consequence* (1990).)

These two approaches to analysing logical validity differ quite sharply from a third account. According to the proof-theoretic account of validity, an argument is valid just when there is a proof of the conclusion from the premises. This approach is championed by Prawitz (1974) with many other constructivists, and Wagner (1987) (a non-constructivist) who sees this view in the work of Gottlob Frege. The proof-theoretic account grounds validity in the meanings of the logical constants. These meanings determine the validity of simple deductions (such as conjunction elimination: from $p \,\&\, q$ to derive *q*). An argument is valid if and only if there are simple deductions from the premises to the conclusion (see Proof Theory).

On this account, logical validity is analytic, and the epistemic function of logic (as a calculus for ideal justification) is obvious. As with model theory, a proof-theoretic account of validity depends on the choice of logical particles, and the rules that govern them.

The proof-theoretic account differs from the other two approaches. If validity is ultimately proof-theoretic, then the class of logical validities is recursively enumerable. If we can list the basic rules, we can list the valid deductions (provided that proofs are finite). Similarly, validity is compact (if *K* follows from *X*, then *K* must follow from some finite subset of *X*). Neither of these properties is essential to the other conceptions of logical validity.

Without doubt, each of these conceptions has a place in the study of logical law. It is harder to discern how they are to be related. My tentative proposal is this. Validity is a matter of necessary truth-preservation. Model-theoretic validity is important, because models can represent possibilities (however ontological commitment to possibilities is to be analysed). Sometimes, models do not represent real possibilities (as when the extension of ‘blue’ is not a subset of the extension of ‘coloured’) so, model-theoretic validity can differ from actual validity. Similarly, proof-theoretic validity can coincide with actual validity. Although analytic truth and necessary truth may not coincide in all cases, simple deductions involving logical particles are instances of necessary truth-preserving inferences. Perhaps these simple deductions will capture all of the necessary truth-preserving inferences in the language. In this case, the proof procedure is complete. On the other hand, the validities in the language may not be recursively enumerable, as in the case of second-order logic. In these cases, no recursive proof-theoretic account will capture all of the validities.

2 The metaphysical question

Even if we have decided what we mean when we call something a logical law, the issue of the *ground* of logical

law remains. If validity is necessary truth-preservation, an account has to be given of necessity and of truth. Similarly for the model-theoretic approach. If models are to bear any resemblance to the world, truth-in-a-model ought to have the same structure as truth *simpliciter*. In these cases, doctrines of truth are relevant to the outcome of logical theory. For example, intuitionists have claimed that, at least in the domain of mathematics, truth comes by way of construction or verification. This claim has resulted in disagreement about logical laws (see Intuitionistic logic and antirealism). The case is similar with quantum logic. Given a particular reading of the correspondence theory of truth (in which propositions correspond to subspaces of the Hilbert spaces of quantum physics), quantum logic seems to follow naturally (see Quantum logic). On the other hand, the correspondence, coherence or pragmatist theories of truth do not seem to dictate a particular theory of logic. Metaphysical doctrines of truth are relevant to an account of logic, but they need not determine that account.

Some take it that logical validity is independent of any particular account of truth, because logical laws are purely a matter of convention. Clearly convention has a large part to play in language, but it is much more than this to say that logical laws are simply true by convention. For example, it is clear that it is a matter of convention that 'snow is white or snow is not white' is used to express a truth (and hence, a logical truth). It is also a matter of convention that 'snow is white' refers to snow. In this case it would be very strange to say that the whiteness of snow is *purely* a matter of convention. Similarly, 'snow is white or snow is not white' is not true solely by convention, but also by the way the world is (see Necessary truth and convention).

Convention in logic is more at home in the view of logic as proof theory. Here, validity is a matter of the meanings of logical particles. Terms get meanings by convention. However, this does not exempt us from the difficult matters of semantics. Logic depends on the meanings of individual particles such as 'not'.

Is the meaning of 'not' a matter of its truth-functionality, and does this yield classical negation? Or is negation to be analysed inferentially, with intuitionists? The analysis of meaning undergirds the proof-theoretic account of validity. No account of logical validity exempts the practitioner from the difficult task of giving an account of the ground of logical law. This is essentially a matter of metaphysics and semantics.

3 The question of extension

Even given answers to the semantic and metaphysical questions, we are not at an end. We have yet to give an account of which inferences qualify as logical truths. Logicians have a ready supply of formal systems designed to do just this: there is a whole wealth of different formal systems designed to give an account of valid inference in different languages. These systems differ from each other in a number of ways. They can contain the truth-functional connectives, quantifiers for objects (both the standard existential and universal quantifiers, and perhaps more exotic quantifiers), quantifiers of higher order, the identity relation, modalities of various sorts, and so on. If validity is construed as necessary truth-preservation, there is no need to make a principled choice of logical constants, because any term may be relevant in determining validities. The only choice is a pragmatic one, of which laws are worth studying. The way is open for a logic of colour, a logic of perception sentences, a logic of action, or a logic of all three. These 'logics' will be partial accounts of validity, just as a theory of electromagnetism is a partial account of physical law.

On the other accounts of validity, a distinction must be made between logical and non-logical terms, in order for there to be a univocal sense of logical validity. This is not an easy task, because there are no principled criteria that something must satisfy in order for it to be a logical constant. People disagree over whether generalized quantifiers of objects (such as 'uncountably many') and second-order quantifiers ought to feature in logic (see Quantifiers, generalized; Second-order logic, philosophical issues in).

Second, formal systems may differ in terms of their results, even if they have the same language. For example, intuitionistic logic and classical logic disagree about double negation elimination (not-not-*p*; therefore *p*) provided that you take intuitionistic logic and classical logic to have the same vocabulary. There is a subtlety here. It is obvious that two different formal systems can be reconciled by treating them as modelling different things. Then intuitionistic logic and classical logic can be seen as not disagreeing, because they have different negations. Quine said as much when he claimed that changing a logic amounted to changing the subject (1970).

However, this is not the whole story. Classical logic and intuitionistic logic can disagree as to the validity of real arguments. Given a particular argument in natural language, an intuitionist and a classical logician may disagree

about its validity. If there is a fact of the matter as to whether the argument is valid, then classical logic and intuitionistic logic can be seen to disagree about this fact. The change of logic does not involve a change of subject when the subject (in this case, natural language arguments) is fixed in advance. Given a particular instance of the natural language ‘not’, classical logic and intuitionistic logic are different accounts of the valid arguments involving this fragment of natural language. The natural language particle is *prior* to its formalizations, which are intended to capture the meaning of the particle. These formalizations can capture the meaning well, or not well, and hence these formalizations can disagree.

With intuitionistic logic, the locus of disagreement is the conception of truth or, alternatively, the meanings of the logical particles. Given a verificationist or constructivist view of truth, you have a reason to be an intuitionist with regard to the standard logical particles. However, if you disagree with constructivism, and you do not analyse the meanings of logical terms in constructivist terms, you can use intuitionistic logic as a logic of necessary provability preservation. Disagreement about intuitionism can only be resolved by agreeing on theories of truth and meaning. The case is similar with quantum logic, as we saw before.

Disagreement becomes more subtle when we consider paraconsistency (permitting contradictions in some way to hold) and relevance. According to most logics, the argument from *p* & not-*p* to *q* is valid, but according to paraconsistent logics (including all logics of relevance - see Anderson and Belnap 1975; Anderson, Belnap and Dunn 1992), it is not. In this disagreement no doctrines of truth or meaning stand out as motivating the peculiar features of the logics. Instead, there seem to be two major reasons that proponents of these systems can have for analysing validity in this way.

The first is quite simple. If a contradiction were true, or even possibly true, it would be clear that the inference from *p* & not-*p* to *q* would be invalid. While many doubt the coherence of this supposition, enough work has been done in the field of paraconsistent logic to show that the approach is coherent (see Paraconsistent logic; Priest, Routley and Norman 1989).

If we admit the need to be able to deduce in inconsistent situations (whether they are epistemic situations, possible situations or actual situations), a logic without the inference from *p* & not-*p* to *q* is necessary. Note that this does not involve any particular doctrine about the nature of truth (correspondence, pragmatic or coherence), but simply the view that it is possible that a contradiction be true. Resolving such disagreements is a matter of examining the arguments for and against the truth of contradictions.

The second motivation for the view that the argument from *p* & not-*p* to *q* is invalid is quite different. On this view you need not hold that it is possible that a contradiction be true. Instead, you maintain that for an argument to be valid, and for a conditional to be true, its antecedent must be relevant to its consequent. If this is the case, then there may be a reason to reject the inference (see Relevance logic and entailment). Again, this disagreement is not about a particular doctrine of the nature of truth. Instead, it is a disagreement about the relationship between relevance on the one hand, and validity and conditionality on the other. Opponents can trade intuitions about the validity or otherwise of individual arguments, but this kind of discussion is rarely fruitful. A saner approach might go as follows: if the relevantist can develop a coherent theory of validity, which models our valid argument at least as well as can be done otherwise, and which, in addition, gives an account of phenomena that cannot be modelled otherwise, then, clearly, the relevantist position is viable and valuable. If it cannot, it will not succeed as a theory of validity.

In all of these considerations, we have seen that logic is quite similar to other sciences. The practitioners have some idea of what the subject matter is (valid inferences) but there is debate about its exact nature. The task is to model the subject matter, and there is no wide consensus about how this is to be done. However, as in other sciences, this does not mean that the study is not worthwhile, nor that it will not enrich our knowledge of the way things are.

See also: Logical and mathematical terms, glossary of; Logical form

GREG RESTALL

References and further reading

Anderson, A.R. and Belnap, N.D., Jr (1975) *Entailment: The Logic of Relevance and Necessity*, vol. 1,

Princeton, NJ: Princeton University Press.(Together with volume 2, this gives the canonical exposition of much of the work on logics of relevance implication and entailment.)

Anderson, A.R., Belnap, N.D., Jr and Dunn, J.M. (1992) *Entailment: The Logic of Relevance and Necessity*, vol. 2, Princeton, NJ: Princeton University Press.(Both volumes together give the canonical exposition of much of the work on logics of relevance implication and entailment. Volume 2 also includes an interesting discussion of different approaches to relevance validity.)

Aristotle (*c.* mid 4th century BC) *Prior Analytics*, trans. Richard McKeon, in *The Basic Works of Aristotle*, New York: Random House, 1941.(Aristotle's work on the syllogism.)

Boolos, G. (1975) 'On Second Order Logic', *Journal of Philosophy* 72: 509-27. (Defence of second-order logic as *logic*.)

Etchemendy, J. (1990) *The Concept of Logical Consequence*, Cambridge, MA: Harvard University Press. (Critique of the model-theoretic account of validity.)

Prawitz, D. (1974) 'On the Idea of a General Proof Theory', *Synthese* 27: 63-77. (Exposition and defence of the proof-theoretic account of validity.)

Priest, G., Routley, R. and Norman, J. (1989) *Paraconsistent Logic: Essays on the Inconsistent*, Munich: Philosophia.(Essays on paraconsistent logic.)

Quine, W.V. (1970) *Philosophy of Logic*, Englewood Cliffs, NJ: Prentice Hall. (Exposition and defence of the model-theoretic account of validity, with a discussion of the scope of logic, and 'deviant' logics.)

Shapiro, S. (1991) *Foundations Without Foundationalism*, Oxford: Oxford University Press.(Defence and applications of second-order logic.)

Sher, G. (1991) *The Bounds of Logic*, Cambridge, MA: MIT Press.(Discussion of first-order quantification and the scope of logic, from the perspective of model theory.)

Tarski, A. (1936) 'Über den Begriff der logischen Folgerung', *Actes du Congrès International de Philosophie Scientifique* 7: 1-11; trans. J.H. Woodger, 'On the Concept of Logical Consequence', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Oxford: Clarendon Press, 1956.(The classic exposition of the model-theoretic account of validity.)

Wagner, S. (1987) 'The Rationalist Conception of Logic', *Notre Dame Journal of Formal Logic* 28: 3-35. (Exposition and defence of the proof-theoretic account of validity, and first-order logic.)

Logical positivism

Logical positivism (logical empiricism, neo-positivism) originated in Austria and Germany in the 1920s. Inspired by late nineteenth- and early twentieth-century revolutions in logic, mathematics and mathematical physics, it aimed to create a similarly revolutionary scientific philosophy purged of the endless controversies of traditional metaphysics. Its most important representatives were members of the Vienna Circle who gathered around Moritz Schlick at the University of Vienna (including Rudolf Carnap, Herbert Feigl, Kurt Gödel, Hans Hahn, Karl Menger, Otto Neurath and Friedrich Waismann) and those of the Society for Empirical Philosophy who gathered around Hans Reichenbach at the University of Berlin (including Walter Dubislav, Kurt Grelling and Carl Hempel). Although not officially members of either group, the Austrian philosophers Ludwig Wittgenstein and Karl Popper were, at least for a time, closely associated with logical positivism.

The logical positivist movement reached its apogee in Europe in the years 1928-34, but the rise of National Socialism in 1933 marked the effective end of this phase. Thereafter, however, many of its most important representatives emigrated to the USA. Here logical positivism found a receptive audience among such pragmatically, empirically and logically minded American philosophers as Charles Morris, Ernest Nagel and W.V. Quine. Thus transplanted to the English-speaking world of 'analytic' philosophy it exerted a tremendous influence - particularly in philosophy of science and the application of logical and mathematical techniques to philosophical problems more generally. This influence began to wane around 1960, with the rise of a pragmatic form of naturalism due to Quine and a historical-sociological approach to the philosophy of science due mainly to Thomas Kuhn. Both of these later trends, however, developed in explicit reaction to the philosophy of logical positivism and thereby attest to its enduring significance.

1 Historical background

Immanuel Kant, in the positivists' eyes, had made a lasting contribution to scientific philosophy - particularly in his rejection of the possibility of supersensible metaphysical knowledge and his reorientation of theoretical philosophy around the two questions 'how is pure mathematics possible?' and 'how is pure natural science possible?' In answering these questions Kant developed his famous defence of synthetic a priori knowledge - knowledge independent of sensible experience yet nonetheless substantively applicable to the empirical world. For Kant, the mathematical physics of Newton paradigmatically exemplified such synthetic a priori knowledge through its reliance on Euclidean geometry and fundamental laws of motion such as the law of inertia. Kant's theory of a priori faculties of the mind - the faculty of pure intuition or sensibility and the faculty of pure understanding - was then intended to explain the origin of synthetic a priori knowledge and thus make philosophically comprehensible the possibility of Newtonian mathematical physics.

After the intervening dominance of post-Kantian idealism, a number of German-speaking philosophers renewed the call for a scientific, epistemological and non-metaphysical form of philosophy. But these Neo-Kantian philosophers also had to face an important new challenge to the Kantian synthetic a priori: the nineteenth-century development of non-Euclidean geometry by Gauss, Bolyai, Lobachevskii, Riemann and Klein (see Geometry, philosophical issues in §§1, 3). Although some Neo-Kantians attempted to defend the uniqueness and apriority of Euclidean geometry nonetheless, others - especially those of the Marburg School such as Paul Natorp and Ernst Cassirer - aimed to generalize the synthetic a priori beyond its particular embodiment in classical Euclidean-Newtonian mathematical physics (see Neo-Kantianism). This latter tendency was similar in important respects to ideas the logical positivists were to elaborate.

But the most important nineteenth-century predecessors of logical positivism were Hermann von Helmholtz, Ernst Mach and Henri Poincaré. Through their efforts to comprehend the radical changes sweeping through nineteenth-century science, these three thinkers initiated a new style of scientific philosophy later taken up and systematized by the positivists. The changes in question included the rise of non-Euclidean geometry, the formulation of the conservation of energy and general thermodynamics, and the beginnings of scientific physiology and psychology. Helmholtz made fundamental contributions to all three areas. He based geometry on the postulate of 'free mobility' of rigid bodies, and, since all classical geometries of constant curvature - negative, positive or zero (Euclidean) - satisfy this postulate, he opposed the Kantian commitment to the aprioricity of

geometry: whether space is Euclidean or non-Euclidean is an empirical question about the actual behaviour of rigid bodies. In physiology, Helmholtz articulated a general principle of psycho-physical correlation whereby our sensations correspond to - but are in no way pictures or images of - processes in the external physical world. These processes consist, in the end, of microscopic atoms interacting via central forces, and, on this basis, Helmholtz developed his famous interpretation of the conservation of energy.

Mach and Poincaré can be seen as reacting, in diverse ways, to Helmholtz. Mach attacked especially atomism and the idea of a psychophysical correlation between two incommensurable realms, and he advanced a programme for the unity of science based on immediately perceptible 'elements' or 'sensations'. The task of science consists solely in seeking correlations among such elements (as in phenomenological thermodynamics), and all dualistic and atomistic tendencies are to be purged as metaphysical via historico-critical analysis. This Machian empiricism exerted a decisive influence on the logical positivists. Poincaré, on the other hand, influenced the positivists primarily through his philosophy of geometry. He agreed with Helmholtz's emphasis on the free mobility of rigid bodies but disagreed with Helmholtz's empiricism. According to Poincaré, the idea of a rigid body is an idealization that cannot be straightforwardly instantiated in the physical world. By freely choosing one of the three classical geometries as, so to speak, a definition of rigidity, we then first make it possible to carry out empirical investigations with real physical bodies. Physical geometry is thus neither synthetic a priori nor empirical: it is 'conventional' (see Conventionalism).

2 Relativistic physics

Albert Einstein's special (1905) and general (1916) theories of relativity entered this volatile intellectual situation as a revelation (see Einstein, A.; General relativity, philosophical responses to; Relativity theory, philosophical significance of). And the relativistic revolution in physics directly stimulated Schlick, Reichenbach and Carnap to initiate a parallel revolution in scientific philosophy. All three thinkers agreed that relativity - especially through the general relativistic description of gravitation via a (four-dimensional) geometry of variable curvature - definitively refutes the Kantian idea that Euclidean geometry is synthetic a priori. Moreover, relativity arises from critical reflection on the empirical significance of spatiotemporal concepts in physics (in particular, the concept of simultaneity and the concept of motion) and thus demonstrates the fruitfulness of Mach's basic point of view. At the same time, however, through its use of sophisticated abstract mathematics, relativity also illustrates the limitations of Machian empiricism (according to which even mathematical concepts have an empirical origin). All three thinkers therefore attempted to formulate an intermediate position that would do justice to both Machian empiricism and the continued importance of a priori mathematical elements in physics. Poincaré's concept of convention came to play a central role.

Schlick, Reichenbach and Carnap first pursued rather different paths. Whereas Schlick emphasized from the outset that the Kantian synthetic a priori has no place at all in the new relativistic context, Reichenbach and Carnap initially attempted to salvage important aspects of Kantianism. Reichenbach began by distinguishing the idea of necessary and unrevisable truth from the idea of necessary presupposition of a given scientific conceptualization of nature. For Reichenbach, relativity refuted the former but embodied the latter. Kant was right that the necessary presuppositions of Newtonian physics included Euclidean geometry and the laws of motion. In moving to relativistic physics, however, these are replaced by fundamentally new presuppositions. We thus end up with a relativized version of the Kantian a priori (as constituting the presuppositions of a particular theory). Carnap, by contrast, began by distinguishing metrical from topological features of physical space. The latter are indeed synthetic a priori as Kant thought (they even depend on a kind of pure intuition), but the former - as general relativity has shown - essentially involve the behaviour of empirically given bodies. We thus end up with a weakening of the Kantian a priori (from metrical to topological features).

These early attempts to salvage aspects of the synthetic a priori did not survive, however. For Schlick's view that relativity is simply incompatible with Kant eventually won the day. Although the distinction between Poincaré's conventionalism and Helmholtzian empiricism was not entirely clear (and Reichenbach, in particular, preferred to associate his later viewpoint with Helmholtz rather than Poincaré), both Reichenbach and Carnap soon came to replace the Kantian notion of the a priori with Poincaré's concept of convention. Yet this form of conventionalism (unlike Poincaré's) was forged in the crucible of a revolutionary new physics and thus demonstrated the vitality and relevance of a new philosophy.

3 Logic and the foundations of mathematics

Whereas the positivists appealed to Poincaré's concept of convention (as realized, so they thought, in relativistic physics) to give a new answer to Kant's question concerning the possibility of pure natural science, they appealed to modern developments in logic and the foundations of mathematics to give a new answer to Kant's question concerning the possibility of pure mathematics. There were in fact two distinguishable sets of developments here. The formal point of view, typified by David Hilbert's logically rigorous axiomatization of geometry, freed geometry from any reference at all to intuitively spatial forms and instead portrayed its subject matter as consisting of any things whatever that satisfy the relevant axioms (see Geometry, philosophical issues in §2; Hilbert's programme and formalism). Geometry is rigorously and a priori true, not because it reflects the structure of an intuitively given space, but rather because it 'implicitly defines' its subject matter via purely logical - but otherwise entirely undetermined - formulas. Mathematical truth, on this view, is identified with logical consistency. The 'logicism' of Gottlob Frege and Bertrand Russell, by contrast, aimed to construct particular mathematical disciplines (especially arithmetic) within an all-embracing system of logic. On this view mathematical disciplines (like arithmetic) indeed have a definite subject matter about which they express truths: namely, the subject matter of logic itself (propositions, classes, and so on). As thus purely logical, however, such pure mathematical disciplines express merely analytic truths and are not synthetic a priori (see Logicism).

Hilbert's formal point of view was pursued especially by Schlick, who in a sense made the notion of implicit definition, together with the associated distinction between undetermined form and determinate (given) content, the centrepiece of his philosophy. The logicist point of view, by contrast, was pursued especially by Carnap, who studied with Frege and then was decisively influenced by Russell. Indeed, Carnap was inspired by Russell's conception of 'logic as the essence of philosophy' to reconceive philosophy itself on the model of the logicist construction of arithmetic. He began, in *Der logische Aufbau der Welt* (1928), by developing a 'rational reconstruction' of empirical knowledge - an epistemology - within the logical framework of Russell and Whitehead's *Principia Mathematica* (1910-13). By defining or 'constituting' all concepts of empirical science within this logic from a basis of subjective 'elementary experiences', Carnap's reconstruction was to show, among other things, that the dichotomy between empirical truth and analytic/definitional truth is indeed exhaustive.

Yet the logic of *Principia Mathematica* was afflicted with serious technical difficulties: the need for special existential axioms such as the axioms of infinity and choice. Partly in response to such difficulties, Ludwig Wittgenstein asserted in his *Tractatus Logico-Philosophicus* (1922) that logic has no subject matter after all: the propositions of logic are entirely tautological or empty of content (see Wittgenstein, L.J.J. §§3-7). Carnap eagerly embraced this idea, but he also attempted to adapt it to the new, post-*Principia* technical situation - which involved the articulation of the 'intuitionist' or 'constructivist' point of view by L.E.J. Brouwer and the development of meta-mathematics by Hilbert and Kurt Gödel (see Intuitionism §§2-3; Logic in the early 20th century §§6-9; Mathematics, foundations of). In *Logische Syntax der Sprache* (1934) Carnap formulated his mature theory of formal languages and put forward his famous 'Principle of Tolerance' - according to which logic has no business at all looking for true or 'correct' principles. The task of logic is rather to investigate the structure of any and all formal languages - 'the boundless ocean of unlimited possibilities' - so as to map out and explore their infinitely diverse logical structures. Indeed, the construction and logical investigation of such formal languages became, for Carnap, the new task of philosophy. The concept of analyticity thereby took on an even more important role. For this concept characterizes logical as opposed to empirical investigation and thus now expresses the distinctive character of philosophy itself.

4 The Vienna Circle

Otto Neurath, Hans Hahn, and the physicist Philipp Frank initiated a discussion group in Vienna, beginning in 1907, in which they considered a combination of Machian empiricism with Poincaré's new insights into the conventional character of physical geometry. Deeply impressed by Schlick's work on relativity theory, they arranged (apparently with Einstein's help) to bring Schlick to the University of Vienna in 1922 to take over the Chair in Philosophy of the Inductive Sciences previously held by Mach. What we now know as the Vienna Circle quickly took shape. Reichenbach, who had become acquainted with Carnap through their common interest in relativity, introduced him to Schlick in 1924. In 1925 Carnap lectured to the Circle in Vienna on his new 'constitutional theory of experience' and became assistant professor under Schlick in 1926. The Circle then

engaged in intensive discussions of Carnap's epistemology and Wittgenstein's *Tractatus*. Wittgenstein's view that all propositions are truth-functions of 'elementary propositions' was combined with Carnap 's constitution of scientific concepts from a basis of 'elementary experiences' so as to create a new, logically rigorous form of empiricism according to which all meaningful - scientific - propositions are reducible to propositions about immediately given experience. And this was articulated as the 'official' philosophy of the Vienna Circle in the famous manifesto *Wissenschaftliche Weltauffassung* in 1929.

Neurath was the driving force in thus turning the Vienna Circle into a public philosophical movement. Trained in economics and the social sciences, Neurath was extremely active politically as a scientific neo-Marxist. In particular, he took the community of natural scientists as the model for a rationally organized human society, and, on this basis, he advocated a reorganization of both intellectual and social life from which all non-rational, 'metaphysical' elements would be definitively purged. In this sense, Neurath saw the philosophical work of the Vienna Circle as a reflection of the wider movement for a *neue Sachlichkeit* then current in Weimar culture - as typified, for example, by the Dessau Bauhaus. As in the wider culture, this movement stood in philosophy for a rejection of individualism in favour of the cooperative, piecemeal, and 'technological' approach to problems exemplified in the sciences, and it was therefore particularly hostile to what was perceived as a return to the metaphysical system-building of post-Kantian idealism by influential German philosophers such as Martin Heidegger. Carnap was especially sympathetic to Neurath's broader philosophical-political vision and clearly expresses this vision in the Preface to the *Aufbau*. Schlick, by contrast, preferred a more individualistic model of philosophy and resisted the idea of a 'movement'.

This divergence between a 'left wing' and a 'right wing' of the Circle emerged in the sphere of epistemology in a debate over 'protocol-sentences' in the years 1930-4. At issue was the status of the basic propositions or protocols in which the results of scientific observation are recorded. It had initially appeared, in Carnap's constitutional system of the *Aufbau*, that such propositions must express private, subjective sense-experience. For Neurath, however, this view was inconsistent with the publicity and intersubjectivity required by science. He therefore advocated a more naturalistic conception of protocols as sentences accepted by the scientific community as recording the results of observation at a given time (see Naturalized philosophy of science). These sentences must thus be expressible within the public and 'physicalistic' language of unified science and hence, like all other sentences, are in principle revisable. Schlick was deeply shocked by Neurath's view - which he took to represent an abandonment of empiricism in favour of the coherence theory of truth (see Truth, coherence, theory of). Carnap attempted, in typical fashion, to mediate the dispute: at issue was simply a choice between two different languages in which to formulate or rationally reconstruct the results of unified science. Although Neurath's thoroughly intersubjective 'physicalistic' language (where, as Karl Popper emphasized especially, every sentence is revisable) was clearly preferable on pragmatic grounds, Carnap held that this choice - like every other choice of formal language - is in the end conventional (see Popper, K.R. §3). Empiricism, in Carnap's hands, is itself framed by conventional and hence non-empirical choices (see Vienna Circle).

5 Emigration, influence, aftermath

The rise of the Nazi regime set off a wholesale migration of logical positivists to the English-speaking world. Carnap, who had become professor at Prague in 1931, moved in 1936 to the University of Chicago. Reichenbach, who had fled to Istanbul in 1933, moved in 1938 to the University of California at Los Angeles. (After Reichenbach's death in 1953 Carnap took over his position at UCLA, beginning in 1954.) Neurath, after leaving Vienna for The Hague in 1934, fled to England in 1940 - where he worked in Oxford until his death in 1945. Friedrich Waismann fled for England as well, where he lectured at Oxford from 1939. Philip Frank emigrated to the USA (also from Prague) in 1938 and settled at Harvard in 1939. Karl Menger took up a position at Notre Dame in 1937, and Kurt Gödel became a member of the Institute for Advanced Study at Princeton in 1940. Herbert Feigl went first to the University of Iowa in 1933 and then to the University of Minnesota in 1940, where he founded the influential Minnesota Center for the Philosophy of Science in 1953. Carl Hempel joined Carnap at the University of Chicago in 1939 and, after teaching at Queens College and Yale, settled at Princeton in 1955. (Schlick was murdered by a deranged student at the University of Vienna in 1936.)

The growth of philosophy of science in the USA was decisively shaped by the work of Carnap, Reichenbach and Hempel. Reichenbach influenced especially the development of philosophy of physics through his work on

geometry, relativity and the direction of time (see Relativity theory, philosophical significance of §3). Hempel published extraordinarily influential papers on the logical analysis of explanation and confirmation and thereby furthered the ideal of scientific philosophy first articulated by Carnap (see Explanation; Confirmation theory). Carnap himself continued the construction of formal languages in which such concepts as testability, modality and probability could be rationally reconstructed or 'explicated' and thus contributed further to the same ideal. Indeed, Carnap's explication of concepts through the construction of formal languages influenced the English-speaking world of analytic philosophy far beyond the borders of philosophy of science. Developments in formal semantics and philosophy of language, in particular, rested on Carnap's initial work on modality (see Semantics §2).

The Carnapian ideal of explication is based on a sharp distinction between logical and empirical investigation, analytic and synthetic truth. In his *Logical Syntax of Language* (1934) Carnap had attempted a general explication of the concept of analyticity itself - a general formal method for distinguishing, within the context of any given formal language, the analytic from the synthetic sentences of that language. After accepting Alfred Tarski's semantical conception of truth in 1935, however, Carnap abandoned the approach of *Logical Syntax* and frankly admitted that (although explications for various particular languages could still be constructed) he now had no generally applicable explication of the concept of analyticity. After studying with Carnap in the early 1930s, W.V. Quine then exploited this situation to attack the concept of analyticity as such and, on this basis, to attack the Carnapian ideal of logical explication as well (see Quine, W.V. §§2, 4). Philosophy, for Quine, is itself a kind of empirical science - a branch of human psychology or 'naturalized epistemology' (see Naturalized epistemology). Moreover, at the same time that Quine was articulating this new philosophical vision, Thomas Kuhn published *The Structure of Scientific Revolutions* (1962) in the *International Encyclopedia of Unified Science* edited by Carnap and Charles Morris. Whereas Carnap had relegated the (conventional) choice of scientific language to the limbo of pragmatics, Kuhn concentrated on those factors - especially social factors - which, in a scientific revolution, determine precisely this kind of choice. These ideas, in harmony with Quine's more general naturalistic point of view, then led to historical and sociological approaches to the study of science and thus, in the end, to the decline of logical analyses of scientific language in the Carnapian style.

See also: Analysis, philosophical issues in; Analyticity; Bridgman, P.W.; Emotivism; Empiricism; Logical atomism; Meaning and verification; Operationalism; Positivism in the social sciences; Theories, scientific; Unity of science

MICHAEL FRIEDMAN

References and further reading

Ayer, A. (ed.) (1959) *Logical Positivism*, New York: Free Press.(Very useful short collection. Contains, in particular, some of the most important papers from the 'protocol-sentence' debate discussed in §4.)

Carnap, R. (1922) *Der Raum: Ein Beitrag zur Wissenschaftslehre*, Berlin: Reuther & Reichard.(Carnap's early analysis of relativity theory. Discussed in §2.)

Carnap, R. (1928) *Der logische Aufbau der Welt*, Berlin: Weltkreis; trans. R. George, *The Logical Structure of the World*, Berkeley, CA: University of California Press, 1967.(Referred to in §§3, 4.)

Carnap, R. (1934) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *The Logical Syntax of Language*, London: Routledge & Kegan Paul, 1937.(Referred to in §§3, 5. English version contains important material not appearing in the original.)

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press.(Carnap's most influential post-war contribution to logic and philosophy of language. Second edition appearing in 1956 contains important supplementary material.)

Carnap, R. (1950) *Logical Foundations of Probability*, Chicago, IL: University of Chicago Press.(Carnap's fundamental analysis of 'logical probability' and the concept of degree of confirmation.)

Carnap, R. (1963) 'Intellectual Autobiography', in P. Schilpp (ed.) *The Philosophy of Rudolf Carnap*, La Salle, IL: Open Court.(Contains detailed information about the historical development of the logical positivist movement.)

Coffa, A. (1991) *The Semantic Tradition from Kant to Carnap*, Cambridge: Cambridge University Press.(Most extensive scholarly treatment of the background to and evolution of logical positivism.)

Creath, R. (ed.) (1990) *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work*, Berkeley, CA: University of California Press.(Excellent source for the Quine-Carnap debate on analyticity discussed in §5.)

Feigl, H. (1969) 'The Wiener Kreis in America', in D. Fleming and B. Bailyn (eds) *The Intellectual Migration*, Cambridge, MA: Harvard University Press.(Discussion of the positivists' migration to America by a participant.)

Frank P. (1949) *Modern Science and its Philosophy*, Cambridge, MA: Harvard University Press.(Classic discussion of the positivist movement by a participant. Contains, in particular, good discussions of the influence of Mach and Poincaré.)

Haller, R. (1993) *Neopositivism*, Darmstadt: Wissenschaftliche Buchgesellschaft.(Special attention to Neurath.)

Hempel, C. (1965) *Aspects of Scientific Explanation*, New York: Free Press.(Collection of Hempel's classic papers which decisively shaped philosophy of science in the USA.)

Kraft, V. (1950) *Der Wiener Kreis*, Vienna: Springer; trans. A. Pap, *The Vienna Circle*, New York: Philosophical Library, 1953.(Good discussion of how Carnap's *Aufbau* and Wittgenstein's *Tractatus* were combined by the Vienna Circle.)

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(Referred to in §5. Extraordinarily influential reaction to logical positivist philosophy of science.)

Neurath, O. (1981) *Gesammelte philosophische und methodologische Schriften*, eds R. Haller and H. Rutte, Vienna: Hölder-Pichler-Tempsky.(Most extensive collection of Neurath's philosophical works in the original. Contains the manifesto *Wissenschaftliche Weltauffassung* on pages 299-336.)

Neurath, O. (1983) *Philosophical Papers 1913-1946*, ed. and trans. R. Cohen and M. Neurath.(Most extensive collection of Neurath's philosophical works in English.)

Popper, K. (1935) *Logik der Forschung*, Berlin: Springer; *The Logic of Scientific Discovery*, London: Hutchinson, 1958.(Popper's extraordinarily influential discussion of scientific method - partly in the spirit of, but also partly explicitly opposed to, logical positivism.)

Reichenbach, H. (1920) *Relativitätstheorie und Erkenntnis Apriori*, Berlin: Springer; trans. M. Reichenbach, *The Theory of Relativity and A Priori Knowledge*, Berkeley, CA: University of California Press, 1965. (Reichenbach's earliest analysis of relativity theory. Discussed in §2.)

Reichenbach, H. (1928) *Philosophie der Raum-Zeit-Lehre*, Berlin: de Gruyter; trans. M. Reichenbach and J. Freund, *The Philosophy of Space and Time*, New York: Dover, 1958.(Reichenbach's most influential analysis of geometry and relativity theory.)

Reichenbach, H. (1951) *The Rise of Scientific Philosophy*, Berkeley, CA: University of California Press.(Widely read popularization.)

Reichenbach, H. (1978) *Selected Writings, 1909-1953*, ed. M. Reichenbach and R. Cohen, Dordrecht: Reidel. (English collection of Reichenbach's most important papers.)

Russell, B. and Whitehead, A. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press. (Referred to in §3. Monumental presentation of modern mathematical logic. Exerted a profound influence on logical positivism.)

Schlick, M. (1917) *Raum und Zeit in der gegenwärtigen Physik*, Berlin: Springer; trans. H. Brose, *Space and Time in Contemporary Physics*, Oxford: Oxford University Press, 1920.(Schlick's widely read and influential treatment of relativity theory. Discussed in §2.)

Schlick, M (1918) *Allgemeine Erkenntnislehre*, Berlin: Springer; trans. H. Feigl and A. Blumberg, *General Theory of Knowlege*, La Salle, IL: Open Court, 1974.(Schlick's major epistemological work. Contains, in particular, the general theory of 'implicit definition' discussed in §3. Second edition appearing in 1925 - on which the English translation is based - is significantly revised.)

Schlick, M (1938) *Gesammelte Aufsätze*, Vienna: Gerold.(Collection of papers, 1926-36.)

Schlick, M (1979) *Philosophical Papers*, eds H. Mülder and B. van de Velde-Schlick, Dordrecht: Reidel.(Most extensive collection of Schlick's works in English. Contains, in particular, the Brose translation of Schlick 1917.)

Uebel, T. (1992) *Overcoming Logical Positivism from Within*, Amsterdam: Rodopoi.(Most extensive treatment - with special emphasis on Neurath's role - of the 'protocol-sentence' debate discussed in §4.)

Waismann, F. (1967) *Wittgenstein und der Wiener Kreis*, Frankfurt: Suhrkamp; trans. J. Schulte and B. McGuiness, *Wittgenstein and the Vienna Circle*, Oxford: Blackwell, 1979.(Very useful material on Wittgenstein's special influence.)

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, trans. C. Ogden with an introduction by B. Russell, London: Routledge & Kegan Paul.(Referred to in §§3, 4. Although Wittgenstein himself was not officially a member, this work and Carnap's *Aufbau* constitute the two most central texts of the logical positivist movement. New translation D. Pears and B. McGuinness, 1961.)

Logicism

The term 'logicism' refers to the doctrine that mathematics is a part of (deductive) logic. It is often said that Gottlob Frege and Bertrand Russell were the first proponents of such a view; this is inaccurate, in that Frege did not make such a claim for all of mathematics. On the other hand, Richard Dedekind deserves to be mentioned among those who first expressed the conviction that arithmetic *is a branch of logic.*

The logicist claim has two parts: that our knowledge of mathematical theorems is grounded fully in logical demonstrations from basic truths of logic; and that the concepts involved in such theorems, and the objects whose existence they imply, are of a purely logical nature. Thus Frege maintained that arithmetic requires no assumptions besides those of logic; that the concept of number is a concept of pure logic; and that numbers themselves are, as he put it, logical objects.

This view of mathematics would not have been possible without a profound transformation of logic that occurred in the late nineteenth century - most especially through the work of Frege. Before that time, actual mathematical reasoning could not be carried out under the recognized logical forms of argument: this circumstance lent considerable plausibility to Immanuel Kant's teaching that mathematical reasoning is not 'purely discursive', but relies upon 'constructions' grounded in intuition. The new logic, however, made it possible to represent standard mathematical reasoning in the form of purely logical derivations - as Frege, on the one hand, and Russell, in collaboration with Whitehead, on the other, undertook to show in detail.

It is now generally held that logicism has been undermined by two developments: first, the discovery that the principles assumed in Frege's major work are inconsistent, and the more or less unsatisfying character (or so it is claimed) of the systems devised to remedy this defect; second, the epoch-making discovery by Kurt Gödel that the 'logic' that would be required for derivability of all *mathematical truths can in principle not be 'formalized'. Whether these considerations 'refute' logicism will be considered further below.*

1 Preliminary discussion

Although the thesis that arithmetic is a part of logic was stated quite explicitly by Frege (1884, 1893, 1903) and by Russell (1903), it is surprisingly difficult to determine the exact content of this thesis; and, indeed, to determine whether it should be understood to mean the same thing to each of them. For the obvious question presents itself: exactly what are we to understand by 'logic'? How do we tell whether a certain basic assumption, or a certain basic concept, properly belongs to logic, or not? And, since an important part of the thesis is its epistemological aspect - the claim that mathematical knowledge is purely logical in nature - one would naturally expect an account of the grounding of our knowledge of logic. On the first of these questions - how to tell what belongs to logic - Frege has little to say (although he acknowledges the importance of the issue for his philosophy, and does provide a few pregnant suggestions); Russell has some lengthy considerations, which have proved not to be very satisfactory. On the second question, Frege and Russell appear to disagree sharply. Both regard the thesis as being in opposition to Kant; but whereas Frege, in denying that arithmetic is based upon any Kantian 'pure intuition' and maintaining that it is entirely grounded in logic, concludes that Kant was wrong to consider arithmetical propositions synthetic, and holds by contrast that, as propositions of logic, they are analytic (roughly speaking: known to be true by 'analysis' of their meanings alone), Russell denies the latter, and holds that logic itself is synthetic in character:

> Kant never doubted for a moment that the propositions of logic are analytic, whereas he rightly perceived that those of mathematics are synthetic. It has since appeared that logic is just as synthetic as all other kinds of truth; but this is a purely philosophical question, which I shall here pass by.
>
> (1903: 457)

Yet even here the issue is not so clearly drawn; for one must still ask whether Frege, in affirming that logic is analytic and not synthetic, and Russell, in affirming the opposite, understood the words 'analytic' and 'synthetic' in the same way - in particular, whether their opposite formulations might none the less have expressed the same opinions. The possibility is lent some weight by the fact that in certain late writings Frege refers to a 'logical source of knowledge' as a distinct knowledge-source (alongside 'sense perception' and 'the geometrical

source of knowledge') ([1924-5] 1979: 267, 278-9). Russell's view that logical truth is synthetic may be essentially identical with Frege's view that knowledge of such truth requires a 'knowledge-source'.

2 Connection with developments in arithmetic, algebra and analysis

The thesis that mathematics is a part of logic could not have been maintained before logic had undergone the deep transformation which took place in the late nineteenth century, most especially through the work of Frege. Before that time, actual mathematical reasoning could not be carried out under the recognized logical forms of argument. The new logic, however, made it possible to represent standard mathematical reasoning in the form of purely logical derivations. It is equally true that no such claim could have been made except for a deep transformation that had taken place in mathematics itself during the nineteenth century; and a brief consideration of this may help to shed some light on a plausible way to construe the logicist thesis.

From the time of Aristotle, it had been usual to regard mathematics as concerned with 'discrete' and 'continuous' quantity: whole numbers on the one hand; continuous magnitudes on the other. In the seventeenth and early eighteenth centuries it was generally supposed that our knowledge of mathematical truths was exclusively concerned with 'relations of ideas', and that - even if these ideas were obtained by some process of abstraction from empirical things' - this character explained how it is that the knowledge in question is certain and independent of experienced 'facts' (such a view was common, for instance, to the rationalist Leibniz and the empiricist Hume). Kant had presented arguments to subvert this position: he argued, rather convincingly, not just that logic (as it then existed) was insufficient to warrant the reasoning of mathematicians, but also that such reasoning demanded attention to more than 'ideas', or (rather) concepts - that a kind of mental visualization of the objects falling under mathematical concepts was an indispensable ingredient.

Among mathematicians - although within a fairly small circle - confidence in *either* the Kantian *or* the pre-Kantian traditional view of the nature of mathematical knowledge was disturbed by the development, in the early nineteenth century, of new forms of geometry: these seemed to show that neither mere consideration of 'relations of ideas' (that is, of such concepts as 'point', 'line', 'distance' and so on), nor the appeal to 'intuition', could decide the truth of propositions of geometry (see Geometry, philosophical issues in). At the same time, attempts were begun to work out new foundations for analysis - that is, the theory of the real (and complex) numbers, functions thereof, and limiting processes - that should be independent of any 'extrinsic' appeal to an 'intuition' of continuous magnitude. Again, in parallel (and partly in connection) with the preceding, systematic procedures were developed for the introduction of 'new objects' - or 'ideal objects' - in several branches of mathematics: for example, 'infinitely distant points' and 'points with imaginary coordinates' in geometry; real and imaginary numbers (alongside the rational numbers) in analysis; and, perhaps most decisively for our subject, 'ideal factors' of algebraic integers in the newly developing subject of algebraic number theory. In this last context in particular, Dedekind achieved fundamental new results with the help of concepts that involved the treatment of *classes* of (already introduced) objects as new objects ('modules', 'ideals', 'orders', 'number fields', among others) which could themselves be made subject both to general mathematical investigation and to actual calculation.

With all of this work in the background, Dedekind was led to examine the foundations of the theory of the whole numbers themselves; and he arrived at the conviction that this theory, together with algebra and analysis which are further derived from it, is 'a part of logic' - in saying which, he tells us, 'I mean to imply that I consider the number-concept entirely independent of the notions or intuitions of space and time, that I consider it *an immediate result from the laws of thought*' ([1888] 1901: 31; emphasis added). The way he develops his theory allows us to conclude that he regards both the concept of number and the theorems of arithmetic, as well as the concepts and theorems of the other mathematical theories he mentions, as resulting from certain fundamental abilities of the mind: the ability to form classes (his word is 'systems') of objects of thought (which themselves then have the status of such objects); to correlate or map such systems onto one another; and to 'create' new objects to 'represent' ones already present ('to relate things to things, to let a thing correspond to a thing, or to represent a thing by a thing'); and he tells us that without this ability, 'no thinking is possible'.

This version of the logicist thesis - namely, that mathematics (or that part of it which is claimed to belong to logic) requires no other source, either of its concepts or of its knowledge, than what is presupposed by all systematic thought - may be taken as applicable to the doctrines of all who have made that claim, whatever epistemological or

metaphysical differences may attach to their further elaboration of what those 'presuppositions of all thought' might be.

3 The logicist foundations of arithmetic

The fundamental problem for the logicist theory of the arithmetic of the ordinary whole numbers was to show how (1) the general concept of such numbers, (2) the individual numbers themselves, and (3) the modes of reasoning that suffice for the mathematical demonstrations of their properties, can all be derived from principles of logic. It will be convenient here to consider first the solution offered by Frege (1884), with which Russell's substantially coincides; a brief comparison with Dedekind's rather different approach will be made later.

Frege set about creating an entirely systematic and universally applicable formal system codifying all logical concepts and processes: his *Begriffsschrift*, or 'concept writing', which he characterized as 'a formula language…for pure thought' (1879). He therefore faced the task of making explicit, within the framework of such a formal system, how the cardinal numbers could be defined.

Now for Frege, as for Dedekind, a central logical role was played by a notion of 'class' - Dedekind's 'system', characterized by him in the following way:

> 'Such a system *S*…is completely determined when with respect to every thing it is determined whether it is an element of *S* or not. The system *S* is hence the same as the system *T*…when every element of *S* is also an element of *T*, and every element of *T* is also an element of *S*'
>
> ([1888] 1901: 45)

(Compare with 'extensionality' in Set theory, different systems of.) However, in Frege's opinion, such notions as that of a system or class, and that of a 'correspondence' ('a thing belongs to a thing') are 'not usual in logic and are not reduced [by Dedekind] to acknowledged logical notions' ([1893] 1964: 4). Frege therefore relies, for the *logical* ground of the notions of class and correspondence, upon the 'acknowledged' logical notions of a 'concept' (which when it is unary corresponds to a property, and when binary, ternary, …, to a two-place, three-place,…relation) and of the 'extension' of such a concept (in particular, the extension of a unary concept is a class). The existence of such an extension for every concept, satisfying the principle that two concepts have the same extension if and only if every object falling under either also falls under the other (see Dedekind's characterization of 'system' above), is a fundamental principle of the logic on which Frege bases arithmetic.

To avoid certain verbal complexities, we now follow Russell in formulating the arithmetic notions in terms of classes. Two classes *A*, *B* are said to be 'equinumerous' if, for some relation *R*: (1) for every element *x* of *A* there is an element *y* of *B* such that *x* has the relation *R* to *y*, but to no other element of *B*; and (2) for every element *y* of *B* there is an element *x* of *A* such that *x*, but no other element of *A*, has the relation *R* to *y*. One says that *R* is a one-to-one relation between *A* and *B*. 'Equinumerous', as just defined, is a binary relation that holds between certain classes. Being equinumerous with a given class *A* is, then, a property which holds of certain classes. In Frege's usage, the extension of this property [or concept] - and thus, the class of all such classes - is called 'the number of the class *A*'; thus we have succeeded, according to the premises of this whole procedure, in defining the individual numbers as 'objects', in purely logical terms. As to the general concept of number, that is now easy: '*n* is a number' means that there is some class *A* such that *n* is the number of the class *A*.

A most crucial step remains to be taken. We have defined the concept of number - allegedly at least - in purely logical terms; but this is too wide a concept to serve as the basis for ordinary arithmetic. Indeed, ordinary arithmetic has to do with the non-negative integers, and these, in the present context, are numbers of *finite* sets. But no such restriction appears in our definition of number: we have defined 'the number of the set *A*' for arbitrary sets - for example, for the set of all the points on a given line. At the same time - what at first will seem a quite unrelated issue - we have yet to show how all arithmetical reasoning can be carried out by purely logical argumentation concerning the number-concept.

The two issues have in fact a deep connection; and it is one of Frege's great accomplishments (also achieved independently by Dedekind) to have perceived the connection and found in it the simultaneous solution to both of them. A characteristic feature of reasoning in arithmetic is the 'argument from *n* to $n+1$', or the principle of mathematical induction:

If a certain property *P* holds of the number 0, and if whenever *P* holds of a number *n* it holds also of the number $n+1$, then *P* holds of all the whole numbers.

(Of course, to formulate this principle at all, one must have defined the process of adding 1 to a given number; but that is a detail, easily accomplished.) The chief problem for placing the argumentation of arithmetic on a 'purely logical' footing was to find a logical justification for this principle. But, speaking 'intuitively', it is clear that the principle in question cannot apply except to finite numbers (for the property of being finite holds of 0, and holds of $n+1$ whenever it holds of *n*; but of course holds only of finite numbers). This 'intuitive' argument can of course have no systematic standing so long as the notion of finite number has not been defined. Frege's idea, however, was to use this 'intuitive' insight as the basis for a purely logical definition of the concept of finite number:

A number *x* is said to be 'finite' if and only if *x* belongs to every class *C* such that: (1) 0 belongs to *C*; and (2) whenever a number *n* belongs to *C*, so does $n+1$.

With this definition, not only are the finite numbers singled out, but all the remaining concepts of traditional number theory can be introduced, and the reasoning of traditional number theory can be carried out within the framework of Frege's extended logic.

4 Extension to other mathematical domains

The extension of these results to the arithmetic of the rational, real and complex numbers offers no new fundamental difficulties, after the pioneering work that had already been done by mathematicians (see Numbers §7).

A difficulty does arise in the theory of functions: namely, the need - first brought to prominent notice by Zermelo (1904), who characterized it as a 'logical principle' - for what has come to be called the 'axiom of choice'; this says, in effect, that if *f* is a function defined on a domain *A*, such that, for each *x* in *A*, $f(x)$ is a non-empty set, then there is a function *g*, also defined on *A*, such that, for each *x* in *A*, $g(x)$ is an element of $f(x)$.

An interesting point arises when we come to geometry. Frege never extended the thesis of logicism to geometry; but, in a rather polemical exchange with Hilbert concerning the latter's path-breaking work on the foundations of geometry (see Frege 1980), Frege actually came near to formulating the point of view from which this extension can - and, indeed, so far as logicism is viable at all, *should* - be made: namely, one regards geometric theory - or, rather, *a* geometric theory (for there are infinitely many) - as concerned with a species of structure (of any one of which there may be many exemplars), characterized - indeed, defined - by the 'axioms' of that geometry; the theorems deduced from those axioms are then demonstrated 'by pure logic' to hold of any exemplar of the species so defined. That is the view maintained by Whitehead and Russell in the projected treatment of geometry in their major work, *Principia Mathematica* (1910-13). And it is the view that had already been expressed, in effect, by Dedekind and still earlier by Riemann (see Stein 1988: 244, 251-3). What is also noteworthy is that Dedekind, in contrast to Frege and to Russell and Whitehead, adopts the same point of view concerning even the system of the whole numbers: that is, he characterizes the *structure* a system must have to serve as that of the whole numbers, in explicit preference to identifying those numbers with any 'objects' specified in other than those structural terms.

5 The setbacks; a brief evaluation

The first of the setbacks for the logicist view was the discovery of a contradiction in Frege's logic (see Paradoxes of set and property §2), which arises from the assumption, already noted, that every concept has an extension - a very strong form of what has come, in the theory of sets, to be called a 'comprehension principle'. Frege himself was eventually led to abandon his logicist view of arithmetic, and to suggest instead that arithmetic is not only - as Kant thought - synthetic a priori, but is actually to be grounded in the 'geometric knowledge-source' ([1924-5] 1979: 276-80). Others, however - notably Russell ([1903] 1937: introduction), Ramsey (1925) and Carnap (1931) - felt no need to abandon the logicist thesis, but reconciled themselves to a modified view of logic, designed to circumvent the contradiction and still allow the essential modes of reasoning. Most celebrated among logicians as such a modified logic is Russell's 'theory of types'; but the axiomatic theory of sets, which has proved most serviceable as a medium for mathematics, could perfectly well also claim to be a form of 'logic' (see the characterization quoted from Zermelo, the pioneer of axiomatic set theory, in §4 above).

The second setback is more profound. Frege's own invention of a way of rendering logic purely formal made it possible to apply modes of mathematical reasoning to the formal linguistic structures of logical systems themselves. The undertaking of such application was a fundamental contribution of Hilbert. This process has further revolutionized our way of looking at logic itself. Among the most basic results of this examination are two theorems of Gödel (1930, 1931), one of which tells us that 'elementary logic' - or first-order logic - can be formulated so that every proposition that deserves to be considered logically true can be derived by 'purely formal', or 'mechanical', means (for example, these propositions can be generated by a computer program); whereas the other tells us that no such thing can be achieved for a system - whether a higher-order logic (of which the theory of types is an example), or a first-order axiomatized theory (of which set theory is an example) - that is strong enough to generate the theorems of elementary arithmetic.

With Gödel's theorems there should be mentioned a discovery that preceded them by several years: the Löwenheim-Skolem theorem (see Löwenheim 1915 and, for clearer and more adequate proofs, Skolem 1920, 1923). This theorem implies that any theory axiomatized in first-order logic that has an infinite model (domain that satisfies the axioms) has a countably infinite model (one whose domain can be placed in one-to-one correspondence with the natural numbers). In a very remarkable paper, Skolem (1923) invoked this theorem (and other considerations) to cast grave doubt on the view that set theory could be 'a satisfactory ultimate foundation for mathematics'. Indeed (restricting attention here to the implications of the Löwenheim- Skolem theorem), it is, for example, a fundamental proposition of classical analysis that the set of all the real numbers is uncountably infinite. The existence of a model for the allegedly foundational theory containing only countably many elements seems, then, prima facie to show that the foundational attempt has failed: that, in particular, the assertion within the theory that a certain set is uncountable has not captured what ought to be the genuine meaning of the term 'uncountable'. The issue is both deep and subtle; here, it must suffice to remark that both Gödel's theorems and the Löwenheim- Skolem theorem lead to the notion of nonstandard models - the theory of which has since become an important part of logic (see Löwenheim-Skolem theorems and nonstandard models).

These results may be summed up by saying that *there is not, and in principle cannot be, a 'completely formalized' logic within which all mathematically statable questions are correctly answerable*. Whether one chooses to call this a refutation of logicism, or an unexpected result about the 'uncompletable' nature of logic itself, the result is unquestionably a deep and interesting one simultaneously concerning mathematics and logic; and, for its discovery, a great debt is owed to those who advanced the thesis of logicism.

See also: Arithmetic, philosophical issues in §4; Hilbert's programme and formalism; Intuitionism §1; Logical and mathematical terms, glossary of

HOWARD STEIN

References and further reading

Carnap, R. (1931) 'Die logizistische Grundlegung der Mathematik', *Erkenntnis* 2: 91-105; trans. E. Putnam and G. Massey, 'The Logicist Foundations of Mathematics', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 41-52.(A classic account of logicism. Referred to in §5.)

Carnap, R. (1939) *Foundations of Logic and Mathematics*, Chicago, IL: University of Chicago Press.(A lucid account of Carnap's view of the relation of logic to the analysis of language, and of how mathematics - a part of the purely logical aspect of language - acquires relevance for the formulation of propositions of empirical science.)

Dedekind, R. (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; repr. in *Gesammelte mathematische Werke (Collected Mathematical Works)*, ed. R. Fricke, E. Noether and Ö. Ore, Braunschweig: Vieweg, 1932, vol. 3, 335-90; trans. W.W. Beman (1901), 'The Nature and Meaning of Numbers', in *Essays on the Theory of Numbers*, New York: Dover, 1963.(Dedekind's theory of the natural numbers. Referred to in §§2-4.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; repr. Hildesheim: Olms, 1964; trans. S. Bauer-Mengelberg, '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in van Heijenoort (1967), 5-82.(Includes the first formulation of Frege's system of logic. Referred to in §3.)

Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der*

Zahl, Breslau: Koebner; repr. and trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Evanston, IL: Northwestern University Press, 2nd edn, 1980.(An extended philosophical essay which culminates in Frege's definitions of number and finite number, and his analysis of mathematical induction; referred to in §3.)

Frege, G. (1893, 1903) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, Jena: Pohle, 2 vols; repr. Hildesheim: Olms, 1966; part 1 of vol. 1 trans. M. Furth, *Basic Laws of Arithmetic: An Exposition of the System*, Berkeley, CA: University of California Press, 1964.(The first two volumes of Frege's projected complete systematic treatment of arithmetic, including the theory of the real numbers, on an explicitly logical foundation. In consequence of Russell's discovery of a contradiction in the system, the third volume was never written. Referred to in §§1, 3.)

Frege, G. (1924-5) 'Erkenntnisquellen der Mathematik und der mathematischen Naturwissenschaften', 'Zahlen und Arithmetik' and 'Versuch einer neuen Begründung für die Arithmetik', in *Nachgelassene Schriften und Wissenschaftlicher Briefwechsel*, vol. 1, ed. H. Hermes, F. Kambartel and F. Kaulbach, Hamburg: Meiner, 1969; trans. P. Long and R. White, 'Sources of Knowledge of Mathematics and the Mathematical Natural Sciences', 'Numbers and Arithmetic' and 'A New Attempt at a Foundation for Arithmetic', in *Posthumous Writings*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Chicago, IL: University of Chicago Press, 1979. (Three very late brief manuscripts, giving Frege's final thoughts on the foundations of mathematics. Referred to in §§1, 5.)

Frege, G. (1980) *Philosophical and Mathematical Correspondence*, ed. B. McGuinness, Chicago, IL: University of Chicago Press, 34-48.(Frege's correspondence with Hilbert on the latter's *Foundations of Geometry*. Referred to in §4.)

Gödel, K. (1930) 'Die Vollständigkeit der Axiome des logischen Funktionenkalküls', *Monatshefte für Mathematik und Physik* 37: 349-60; trans. S. Bauer-Mengelberg, 'The Completeness of the Axioms of the Functional Calculus of Logic', in van Heijenoort (1967), 582-91.(Gödel's proof of the completeness of first-order logic; referred to in §5.)

Gödel, K. (1931) 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38: 173-98; trans. J. van Heijenoort, 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems', in van Heijenoort (1967), 592-617.(Gödel's celebrated incompleteness theorems, the first of which is referred to in §5.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press.(Includes translations of several of the works discussed in this article, with introductions.)

Löwenheim, L. (1915) 'Über Möglichkeiten im Relativkalkül', *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, 'On Possibilities in the Calculus of Relatives', in van Heijenoort (1967), 232-51. (Löwenheim's original proof - difficult, and couched in now unfamiliar notation - of the Löwenheim-Skolem theorem; van Heijenoort's introduction is very helpful. Referred to in §5.)

Ramsey, F.P. (1925) The Foundations of Mathematics, *Proceedings of the London Mathematical Society*, series 2, 25(5): 338-84; repr. in *The Foundations of Mathematics and Other Logical Essays*, ed. R.B. Braithwaite, London: Routledge, 1931.(Referred to in §5. The posthumous collection, of both previously published and unpublished works, in which this essay appears includes important discussions of the theory of types, the axiom of choice and (a point not discussed in the text above) Russell and Whitehead's axiom of infinity - needed in their version of a logicist theory in order to obtain a complete system of arithmetic.)

Russell, B.A.W. (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992.(A historically important work, of substantial difficulty; quoted in §1.)

Skolem, T. (1920) 'Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen', *Videnskapsselskapets skrifter, I. Matematisk-naturvidenskabelig klasse* 4; §1 trans. S. Bauer-Mengelberg, 'Logico-Combinatorial Investigations in the Satisfiability or Provability of Mathematical Propositions: A Simplified Proof of a Theorem by L. Löwenheim and Generalizations of the Theorem', in van Heijenoort (1967), 254-63. (Skolem's more satisfactory proof, and strengthening, of Löwenheim's theorem on countable models; referred to in §5.)

Skolem, T. (1923) 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre', in *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, Helsinki: Akademiska Bokhandeln, 217-32; trans. S. Bauer-Mengelberg, 'Some Remarks on Axiomatized Set Theory', in van Heijenoort (1967), 291-301.(A brief, remarkably clear and very deep discussion of a series of issues concerning the foundations of mathematics. Referred to in §5.)

Stein, H. (1988) ‘Logos, Logic, and Logistiké: Some Philosophical Remarks on the Nineteenth-Century Transformation of Mathematics’, in W. Aspray and P. Kitcher (eds) *History and Philosophy of Modern Mathematics*, Minneapolis, MN: University of Minnesota Press, 238-59.(Includes a discussion of some of the nineteenth-century developments that form the mathematical background to the logicist thesis referred to in §2; and gives an indication of the views of Riemann and Dedekind on geometry, referred to in §4.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(Referred to in §4.)

Zermelo, E. (1904) ‘Beweis, daß jede Menge wohlgeordnet werden kann’, *Mathematische Annalen* 59: 514-16; trans. S. Bauer-Mengelberg, ‘Proof that Every Set can be Well-Ordered’, in van Heijenoort (1967), 139-41. (The article that brought prominently to the attention of mathematicians the axiom of choice - here ‘the assumption that coverings exist’; referred to in §4.)

Logos

The noun logos *derives from the Greek verb* legein*, meaning 'to say' something significant.* Logos *developed a wide variety of senses, including 'description', 'theory' (sometimes as opposed to 'fact'), 'explanation', 'reason', 'reasoning power', 'principle', 'ratio', 'prose'.*

Logos *emerges as a philosophical term with Heraclitus (*c.*540-*c.*480* BC*), for whom it provided the link between rational discourse and the world's rational structure. It was freely used by Plato and Aristotle and especially by the Stoics, who interpreted the rational world order as immanent deity. Platonist philosophers gave pre-eminence to* nous*, the intuitive intellect expressed in* logos*. To Philo of Alexandria and subsequently to Christian theologians it meant 'the Word', a derivative divine power, at first seen as subordinate but eventually coordinated with the Father.*

1 Greek philosophy

Logos became an important term in almost all philosophical schools. It emerged about 700 BC as the accepted term for discourse at any length, though seldom if ever naming a single word; it displaced the older terms *epos* and *mythos*, which survived only with specialized meanings.

Heraclitus expounded a highly original philosophy in enigmatic phrases. *Logos* apparently means his own account of the world, which is 'common' but rarely grasped (frs 1, 2); it also suggests the everlasting cosmic order whose contrasting aspects or phases disclose an underlying unity (fr. 51). All phenomena spring from a basic element, fire, and return to it in regular 'measures' (*logoi*); earth exchanges with sea in the same proportion, *logos* (frs 30, 31, 90). This could suggest that changing phenomena can be known; but Heraclitus was seldom credited with this view.

The Sophists of the fifth century BC who *inter alia* taught the art of public speaking, naturally emphasized the power of *logos*, which Gorgias in his *Encomium of Helen* personifies as 'a great potentate who brings truly godlike things to pass' and Isocrates represents as the foundation of human culture.

Plato uses the term *logos* in almost all the senses mentioned above, though without giving it exclusive emphasis as compared with other cognate words (especially *dialektikē*, 'dialectic'); moreover, the general term *nous* for intellect, thought or intelligence is basic to his philosophy (see Nous); *logos*, and sometimes *dianoia*, 'thinking' (*Republic* 533d), stand for its expression. Both are contrasted with sense-experience (for example, *Republic* 509d) which Plato associates with instability and untruth; immunity from change is a precondition of real knowledge. In *Timaeus* 27-28 he contrasts unchanging being with change: 'the one is apprehended by thought (*noēsis*) with discourse (*logos*), being ever self-same; the other grasped by belief and unreasoned sensation'. Since mathematics expounds unchanging truths, Plato gave it an important place in his educational system.

He presents Socrates as initiating discussions which lead towards a 'true and certain and comprehensible *logos*' of ethical concepts and so promote right action, ignoring the body and its sensations as well as unsound *logoi* which lead to contradictions (*Phaedo* 90). Distrust of sense-experience is not always expressed; the *Seventh Letter* (344) states that knowledge calls for a combination of nomenclature, *logos* (definition), sensation and civilized discussion. The *Theaetetus* presents some attempts to define knowledge: it cannot be mere sensation, nor mere true belief, which might arise by accident; one must be able to give an account of it (*logon dounai*); but no satisfactory definition of *logos* is found. Discussion is necessary since written words can mislead (*Phaedrus* 275); its aim is to 'ignore all the senses [!] and press on through *logos* towards the essential reality of each thing' (*Republic* 532), its Idea or Form. Discussion (*dialektikē*) seeks a classification of all such Forms, showing their dependence on the Form of the Good, though this itself is an 'unpostulated' or absolute source, *anhypothetos archē* (*Republic* 510) (see Archē), 'beyond being in dignity and power' (509b).

Aristotle defines *logos* as a composite 'significant utterance', *phonēsemantikē* (*De Interpretatione* 4 16b26), but actually uses it in a wide variety of senses. It often deputizes for more specific terms, and thus traverses useful distinctions, for example, sentence/treatise, or proposition/syllogism. It often occurs in Aristotle's logical works, but the term 'logic' is a later coinage; Aristotle speaks of 'analytics'. His ethics prescribes following reason, or 'right reason' (*orthos logos*), sometimes conceived mathematically as a mean between two extremes. In his

metaphysics *logos* indicates, and sometimes equates with, the substance, form or essence of things (*ousia*, *eidos*, *to ti ēn einai*). Its relation to *nous*, pure intelligence, is left unexplained; but the intelligence moving the cosmos is always designated by *nous*, not *logos*. (See Aristotle §16.)

The Stoics held that only material things were fully real; but they were not, in the ordinary sense, materialists (see Stoicism §§3, 5). They held that the universe embodied a *logos* as its supreme directive principle. Natural objects, plants and animals have their own increasingly complex patterns of behaviour, culminating in the human reason, itself a lower analogue of the universal *logos*. This principle could thus be named in material terms as *pneuma* (spirit, breath; see Pneuma), in functional terms as *logos*, reason, and in valuational terms as *theos*, god. They did not, however, discard the common use of *logos* to denote purely human reason, or specialized forms of it ('argument', 'experience'); they accepted the common designation of animals as unreasoning, *aloga*.

Passions were commonly described as 'impulses contrary to *logos*' (see Stoicism §19); but were occasionally ascribed to, or identified with, 'a bad and licentious *logos*', and described as 'decisions', *kriseis*. But since *logos* could also mean 'the world order', conformity to *logos* could imply 'accepting the predetermined order of events', so that moral freedom was interpreted as resignation.

2 Christian theology: antecedents and developments

The Israelites came to believe that God's name, though revealed, should not be uttered; they used periphrases, for example, 'heaven', and ascribed divine action to intermediate agencies, including God's Word (*logos*) and his Wisdom (*sophia*). These are prominent in Philo of Alexandria, who held that the best Greek thought can be found, by appropriate exegesis, in the Jewish scriptures.

Philo echoes the 'dogmatic' Platonism reconstituted by Antiochus, with some Stoic elements, and Pythagorean ideas stemming probably from fellow-Alexandrian Eudorus (see Neo-Pythagoreanism; Platonism, Early and Middle §§4,8-9). He is our first extant source for the view that the Platonic Ideas are thoughts in God's mind. He describes the Logos as the 'Idea of Ideas' (*Migration* 103), the 'intelligible world' (*Creation* 24-), the 'divider' (*tomeus*) who arranged the pattern of Ideas for creation; indeed as 'a second God' (*Allegories* 1.32-), rightly honoured by those (Stoics?) unable to grasp the supreme reality. Stoic influence appears in his description of the Logos as 'fiery' (*Cherubim* 28) and 'physically pervasive' (*The Heir* 217). In *Immutable* 31 the Logos is God's elder son, the world his younger (compare the Platonist triad of God, the Ideas and matter?). Sophia is described in comparable terms; once only (*Flight* 109) she appears as 'mother' of the Logos, associated with God in a triadic divine family.

The New Testament appears as indifferent or even hostile to philosophy (*Colossians* 2:8). Exceptions are the hints of an 'argument from design' (*Acts* 17:22-29, *Romans* 1:20) and the prologue to the Fourth Gospel, apparently aimed at readers with philosophical interests, where the Logos appears as God's associate, a source of light or enlightenment, who in due course took human form. The title Logos suggests the Jewish theology of God's Word (*memra*) as developed in the biblical *Wisdom of Solomon* and Philo; it also expresses the Christian claim that Jesus is God's accredited representative and agent, appointed as his Son (see *Romans* 1:1-4, *Hebrews* 1-), the title which became dominant in the Gospel and the New Testament in general. The Christian Trinity of Father, Son and Spirit (2 *Corinthians* 13; 14, *c.* AD 55) - with the alternatives Logos and Sophia - can hardly derive from the very differently formulated triadic theology attributed to Moderatus (*c.*50-100), though later Christians could claim the support of philosophers such as Numenius. (Plotinus, who sees *nous* as the first derivative power, was little known to Christians before Augustine.)

Ignatius (*c.* AD 35-107) describes Jesus as God's Word (*logos*) 'proceeding from silence' (*Magnesians* 8.2). Justin Martyr (*c.*100-65) inaugurated a distinctive Logos-theology, which although retaining the Trinitarian confession emphasized the Logos, with philosophical echoes, as God's agent and interpreter at the expense of the Holy Spirit. Origen (*c.*185-*c.*254) was the most important exponent of this theology, which was superseded by the Council of Nicaea in 325.

See also: Antisthenes

CHRISTOPHER STEAD

References and further reading

Armstrong, A.H. (ed.) (1967) *The Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press.(Comprehensive survey from Plato to Anselm.)

Dillon, J. (1977) *The Middle Platonists*, London: Duckworth.(Indispensable guide to Platonic tradition to *c.* AD 200.)

Kahn, C.H. (1979) *The Art and Thought of Heraclitus*, Cambridge: Cambridge University Press.(Authoritative study with texts and translations.)

Kittel, G. (1967) 'Logos', in *Theological Dictionary of the New Testament*, Grand Rapids, MI: Eerdmans, vol. 4, 71-136.(Comprehensive; fullest on biblical usage.)

Sandbach, F.H. (1975) *The Stoics*, London: Chatto & Windus.(Well-written introductory account.)

Stead, G.C. (1991) 'Logos', in *Theologische Realenzyklopädie*, Berlin: de Gruyter, vol. 21, 432-44.(Includes Hellenic, biblical and Christian usage.)

Stead, G.C. (1994) *Philosophy in Christian Antiquity*, Cambridge: Cambridge University Press.(Introductory: Thales to Augustine.)

Loisy, Alfred (1857-1940)

Loisy was a French biblical exegete who worked in the tradition of biblical criticism whose earlier members included D.F. Strauss and Ernest Renan. His critical views involved a sharp separation between the Jesus of history and the Christ of Catholic faith, and he came to regard the doctrine of the incarnation of God in Jesus Christ, a central Christian doctrine, as merely metaphorical and symbolic. He has been called the father of Catholic modernism.

Alfred Loisy was born in Ambrières, Marne, and was ordained a Catholic priest in 1879. He became Professor of Hebrew at the Catholic Institute in Paris in 1881, where he attended Ernest Renan's lectures from 1882 to 1885. He was made Professor of Biblical Exegesis in 1889, but was removed from his position in 1893 after publishing criticism of Catholic seminaries for not taking account of contemporary critical scholarship. From 1893 to 1899 he taught religion in Neuilly, and from 1901 to 1904 he was a lecturer in the École Pratique des Hautes Études in Paris. In 1909 he became Professor of the History of Religion in the Collège de France and also, after 1927, in the École des Hautes Études. Loisy came to describe his own views as pantheistic and positivistic, and was excommunicated in 1908.

Loisy's work was in the critical tradition of D.F. Strauss and Ernest Renan. In a series of works, he recorded and practised what the Church regarded as radical biblical and historical criticism. One example of this is his treatment of the fourth Gospel. (It is traditional in New Testament studies to refer to the Gospels of Matthew, Mark and Luke as the Synoptic Gospels, as distinct from that of John, the fourth Gospel.) In *Le Quatrième Évangile (The Fourth Gospel)* (1903), Loisy argued that the fourth Gospel is the coherent work of a single great theological genius, a symbolic description of Christian belief, a foundation stone of Christian truth, but not a trustworthy account of the life and teaching of Jesus Christ. The Synoptic Gospels, he held, are almost impersonal works which collect traditional reminiscences concerning Jesus, and have a greater historical value than the Gospel of John. More theologian than historian, still more an apologist or defender of the Christian faith, the author of the Gospel of John, according to Loisy, offers a brilliant history of Christianity since the resurrection of Jesus. But the Gospel does not deal with the times and events with which it ostensibly deals, namely what Jesus did and taught during his life and ministry. The author was unable to distinguish between what he received from tradition and what he found in the depths of his own imagination and conviction.

The fourth Gospel was intended, Loisy maintained, to support the legitimacy of Christianity in the context of contemporary Judaism. Its author manifests a supreme indifference to history, selecting from the traditional materials available what suits his purpose, and altering it as his purpose requires. That purpose is the construction of allegorical pictures that convey a wholly mystical and spiritual Christ, who is not subject to the conditions of historical human existence. The language used is figurative and ambiguous; it concerns a Christ who lives in the Church, and has scant connection with the Jesus who called disciples and was crucified. History belongs to the Synoptic Gospels, and the facts that they refer to receive theological and transcendental interpretation in the Gospel of John. But the difference in the worlds of thought occupied by the Synoptics and John is so great that the standard practice of weaving texts and themes from both sources into a single portrayal of the life and teaching of Jesus is undermined. This laid the foundation for a sharp distinction between the Jesus of history and the Christ of Catholic faith. Loisy saw in the author of the Gospel of John the first Christian mystic, for whom history is swallowed up in his mysticism.

Loisy's views, of course, were and are controversial, even setting aside the doctrinal implications that he drew from his critical work. Both his dismissal of the Gospel of John as a historical source and his treatment of that Gospel as theologically oriented in a manner different in kind from the other Gospels are challengeable.

KEITH E. YANDELL

List of works

Loisy, A. (1902) *L'Évangile et l'église*, Paris: E. Nourry, 5th edn, 1930; *The Gospel and the Church*, New York: Prometheus Books, 1988.(Reply to A. Harnack's *The Essence of Christianity*; a defence of Loisy's Catholicism.)

Loisy, A. (1903) *Le Quatrième Évangile (The Fourth Gospel)*, Paris: Vrin.(An important example of Loisy's radical approach to biblical texts.)
Loisy, A. (1924) *My Duel with the Vatican*, New York: Greenwood Press, 1968.(Loisy presents his side of the story.)
Loisy, A. (1933) *La Naissance du christianisme*, trans. L.P. Jacks, *The Birth of the Christian Religion*, London: Allen & Unwin, 1948.(Loisy's account of the origins of Christianity.)

References and further reading

Hoskyns, Sir E. (1957) *The Fourth Gospel*, revised edn, ed. F.N. Davey, London: Faber & Faber.(A commentary on the Gospel of John which discusses the history of its interpretation; see chapter 2.)
Kummel, W.G. (1970) *Das Neue Testament: Geschichte der Erforschung seiner Probleme*, Verlag Karl Alber; trans. S. McLean Gilmour and H.C. Kee, *The New Testament: The History of the Investigation of its Problems*, Nashville, TN: Abington Press, 1972.(This is a comprehensive study of the history of New Testament criticism, with copious quotations from critics.)

Lombard, Peter (1095/1100-1160)

Peter Lombard's philosophical views are important given the formative role his Sententiae in IV libris distinctae (Four Books of Sentences) played in the education of university theologians in the high Middle Ages, many of whom were also philosophers. Peter staunchly opposes theologies, cosmologies and anthropologies of a Platonic or Neoplatonic type. While conversant with new trends in logic in his day, he is disinclined to treat theological issues as illustrations of the rules of formal logic or natural philosophy, preferring to view them from a metaphysical perspective. In his doctrine of God he deliberately eschews terminology associated with any one philosophical school. In his anthropology and sacramental theology he shows a marked preference for Aristotelianism. The hospitability of his theology to Aristotelianism and to a philosophical treatment of a range of theological questions made his Sentences elastic enough to accommodate the reception of Greco-Arabic thought and to serve as a pedagogical framework usable by philosophers of every persuasion during the succeeding three centuries.

Born in the Novara region of Italy, Peter left no traces on the historical record until he was noticed by Bernard of Clairvaux in the mid-1130s. With Bernard's advice and help, he went to France to study theology, first at the cathedral school at Rheims and then, in 1136, at Paris, probably under Hugh of St Victor. His earliest works were commentaries on the Psalms, written before 1138, and the Pauline epistles (first rescension 1139-41). He had become a recognized master by 1142. In 1145 Peter became a canon of Notre Dame, an unusual honour for a foreigner, teaching there until 1159 when he was elected bishop of Paris, a post he held until his death in 1160. His major work, the *Sententiae in IV libris distinctae (Four Books of Sentences)*, sums up his systematic theology. The final version of this work was completed in 1155-7.

For Peter, not all branches or schools of philosophy are equally pertinent to theology. He is uninterested in natural philosophy and says less about it than do the patristic authorities on whom he relies for his account of creation. He is familiar with the logicians' debates over universals. Without taking a personal stand on the priority or posteriority of universals to concepts standing for individuals, he refers to them in glossing 'all creation' (Romans 8: 19), which, he states, is a universal collecting all aspects of the human nature that is to be saved. Peter applies the contemporary nominalist view of the unitary signification of nouns and verbs and their consignification in oblique cases or past and future tenses, using this principle against Peter Abelard's argument that God could do otherwise or better than he does, in a defence of divine omniscience and omnipotence that contributes to the development of the distinction between God's absolute and ordained power (see Abelard, P.; Nominalism; Omnipotence).

Peter rejects any kind of emanationist, exemplarist or immanentalist understanding of God and his relationship to creation, as well as the equation made in some quarters during his time of the Platonic One, Nous and World Soul with the Trinity (see Neoplatonism). In glossing Romans 1: 20, he agrees with Marius Victorinus that the Neoplatonic triad of being, life and thought is apposite to the Trinity but his appreciation of the Platonic tradition stops there. Sometimes he favours a philosophical vocabulary neutral enough to shoulder its theological duties, unconstrained by the denotations which particular schools impose on key terms. At other points he opts for Aristotelianism. Opposing the Platonic notion of the human being as a soul using a body, he joins the Aristotelians in defining it as an integral union of body and soul. This principle undergirds his treatment of the creation of mankind, the fall and its consequences, Christ's human nature, the redemption, ethics and the sacraments. While he accents inner intentionality in ethics, he thinks that good intentions should be expressed in appropriate external actions, and while he holds that it is consent that makes a marriage, he regards the physical as well as the spiritual union of spouses as sacramental. He calls the present consent of spouses the efficient cause of marriage, and the ends of marriage (fidelity, lifelong commitment and offspring) its final cause. He applies 'substance' and 'accidents' in their Aristotelian sense to the change in the eucharistic elements at the time of consecration.

In other areas, Peter's use of Aristotelianism is muted by his felt need for a more neutral lexicon or because he wants to emphasize something else. He offers four proofs for God's existence, including Aristotle's argument from motion to a prime mover. His other proofs move from physical to metaphysical arguments in accenting the idea that the deity is the ground of being of the creation, rather than the idea that he is the cause of effects in the

phenomenal world (see God, arguments for the existence of). In treating the deity, Peter prefers a generic definition of 'substance' as referring simply to an entity's basic nature, because the Aristotelian understanding of 'substance' as referring to composite created beings is inapposite to God. Similarly, he rejects the Aristotelian definition of 'relation' as an accident when he discusses the mutual relations that name the persons of the Trinity. In another case, his doctrine of conscience as the spark of reason not extinguished even in the worst of sinners, Peter draws on an idea ultimately Stoic in provenance which he is the first medieval thinker to revive (see Stoicism).

More generally, Peter's acknowledgement of natural reason as a real epistemic state and his positive use of it in theology in opposition to Neoplatonic negative theology, his resolutely metaphysical approach to the deity and to the unmanifested Trinity as the supreme objects of knowledge that reason can address, and his interest in human nature as such, before the fall, open up zones for philosophical reflection which later scholastics could and did develop, whatever their philosophical proclivities. While issuing Aristotelians a warning concerning the appositeness of 'substance' and 'relation' to the divine nature, he provides notable support for that school. At the same time, he throws down the gauntlet to defenders of a more thoroughgoing Platonism.

See also: Aristotelianism, medieval; God, concepts of; Platonism, medieval; Trinity

MARCIA L. COLISH

List of works

Peter Lombard (1139-41) *Collectanea in omnes d. Pauli Epistolas*, ed. J.P. Migne, Patrologia Latina cursus completus, vols 191-92, Paris, 1880.(Peter's commentary on the Pauline epistles. Material of philosophical interest is found primarily in gloss on Romans.)

Peter Lombard (final recension 1155-7) *Sententiae in IV libris distinctae (Four Books of Sentences)*, ed. I.C. Brady, 3rd revised edn, Grottaferrata: Collegi S. Bonaventurae ad Claras Aquas, 1971, 2 vols.(Text of the *Sentences*, Peter's major work, excellently introduced and annotated by editor.)

References and further reading

Bertola, E. (1956) 'Il problema di Dio in Pier Lombardo' (The Problem of God in Peter Lombard), *Rivista di filosofia neo-scolastica* 48: 135-50.(Places Peter's doctrine of God in its contemporary context.)

Bertola, E. (1959) 'La dottrina lombardiana dell'anima nella storia delle dottrine psicologiche del XII secolo' (Lombard's Doctrine of the Soul in the History of Psychological Doctrine in the Twelfth Century), *Pier Lombardo* 3 (1): 3-18.(Only brief study of Lombardian psychology in contemporary context.)

Colish, M.L. (1990) 'Gilbert, the Early Porretans, and Peter Lombard: Semantics and Theology', in J. Jolivet and A. de Libera (eds) *Gilbert de Poitiers et ses contemporaines: Aux origines de la logica modernorum*, Naples: Bibliopolis, 229-50. (Peter's theological language and the influence, positive and negative, of Gilbert of Poitiers and his early disciples on him.)

Colish, M.L. (1992) 'Peter Lombard and Abelard: The *Opinio Nominalium* and Divine Transcendence', *Vivarium* 30: 139-56.(Contrasts use of the theory of the unitary signification of nouns and verbs in Abelard and Lombard.)

Colish, M.L. (1994) *Peter Lombard*, Leiden: Brill, 2 vols.(The only full length modern study, covering all aspects of Peter's thought, with extensive bibliography.)

Delhaye, P. (1961) *Pierre Lombard: Sa vie, ses oeuvres, sa morale (Peter Lombard: His Life, His Works, His Ethics)*, Montreal, Ont.: Institut d'Études Médiévales.(The only monographic treatment of Peter's ethics.)

Gilson, E. (1945) 'Pierre Lombard et les théologies d'essence' (Peter Lombard and the Theologies of Essence), *Revue du moyen âge latin* 1: 173-6.(Good if brief study correctly noting Peter's emphasis on divine transcendence.)

Lonergan, Bernard Joseph Francis (1904-84)

The Canadian philosopher and theologian Bernard Lonergan approached the problems of philosophy by inviting his readers to attend to the mental acts in which they engage when they come to know anything. He claimed that these acts are of three fundamental kinds: 'experience' of the data of sensation, feeling or mental activity; 'understanding' possible explanations of that experience; and 'judgment' that one such explanation is in each case certainly or probably so. Denial that we engage in these three types of mental activity is actually self-destructive, since we have to engage in them in the very act of justifying such denial. In getting to grips with what it is to come to know, we also gain a vital clue as to the overall nature of the world which is to be known; and light is thrown on the relation between the natural and the human sciences, and on the questions of ethics and religion.

1 The nature of knowledge

Bernard Joseph Francis Lonergan was born in 1904, in Buckingham, Quebec. He entered the Jesuit Order in 1922, and studied at Heythrop College, England, and the Gregorian University in Rome. After some years as professor of theology at Jesuit seminaries in Montreal and Toronto, he was appointed to the faculty of the Gregorian University in 1953. In 1965 he returned to Canada, after a serious illness; later he taught at Harvard Divinity School, and finally at Boston College.

'Thoroughly understand what it is to understand, and not only will you understand the broad lines of all there is to be understood but you will possess a fixed base, an invariant pattern, opening upon all further developments of understanding' (Lonergan 1957: xxviii, 748). The main object of Lonergan's most substantial philosophical book, *Insight: A Study of Human Understanding*, is to get readers to take possession of their own rational self-consciousness; once this is done the philosophy which he calls 'critical realism' will show itself to be the only one which does justice both to human knowledge and to the world which is to be known by it. The true judgments of which human knowledge consists are to be arrived at by putting two kinds of question to the data of experience. I may inquire with respect to any phenomenon what it may be, or why it may occur; and having thought of an answer, I may ask another sort of question with respect to that, 'Is it so or not so?' The first kind of question, which culminates in an act of understanding or 'insight', Lonergan terms 'a question for intelligence'; the second, which issues in a judgment that something is the case or not the case, 'a question for reflection'. Questions for reflection, as opposed to those for intelligence, may properly be answered 'Yes' or 'No'. 'What sort of a rose is that?', or 'What is the cause of the unpleasant smell in this room?', cannot be answered in such a way; but 'Is it a rugosa?', or 'Is it the fish we threw into the garbage yesterday?', may certainly be so. Examples of the two kinds of question, and of the answers which may be given to them, are readily to be found in the most ordinary circumstances of life, as well as in the further reaches of theoretical science and scholarship. After *Insight*, Lonergan refers to the 'transcendental precepts' - 'be attentive, be intelligent, be reasonable' - which correspond respectively to experience, understanding and judgment; and claims that these precepts are relevant to the acquisition of knowledge of every kind.

It might be asked whether the claim that we perform such acts is not mere 'mentalism' or folk psychology, destined to disappear before the march of science. The answer is, that I cannot deny that I am at least to some extent really attentive, intelligent and reasonable, without thereby disqualifying myself from saying anything at all which is worth taking seriously. I can hardly argue soundly for a conclusion which implies that I am incapable of soundly arguing for a conclusion. And all such argument is a matter of being attentive, intelligent and reasonable - in other words, of attending to relevant evidence in experience, of envisaging possibilities which may account for that experience, and of affirming with more or less certainty the possibility which does seem best to account for it. Lonergan refers to the method underlying his philosophy as 'generalized empirical method', since it is based on our awareness not only of our sense-experience and feeling, but also of the mental acts (of questioning, hypothesizing, marshalling evidence, judging and so on) which we apply to these.

2 Science and metaphysics

With the advance of the sciences, it has become more and more clear that the possibilities which may be envisaged to explain what is observable are not themselves couched in terms of what is observable. Mass, valency and

electrical charge are not direct objects of sight or hearing; but they have been intelligently conceived and reasonably affirmed as constitutive of the best available explanation of a huge range of observable phenomena. 'Experiential conjugates' are properties of things in relation to our senses. 'Pure conjugates', as progressively discovered by scientific inquiry, are properties of things as related to one another. The moral, that knowledge can reach beyond experience, is also applicable to the notorious problem of other minds. Though I cannot have experience of your thoughts and feelings, I can certainly conceive that you may be thinking and feeling in a certain way, and verify the judgment that you are actually doing so on the evidence of my experience.

Does scientific realism commit one to determinism? (see Determinism and indeterminism; Scientific realism and antirealism). Modern physics seems to demand statistical laws to explain such phenomena as radioactive decay; there appears to be no adequate reason to suppose that these laws are a mere cloak for ignorance of underlying classical laws. The application within the universe of statistical as well as classical laws is to be grasped by what is called an 'inverse insight' to the effect that a whole familiar body of assumptions and strategy of questioning are wrong. (It is by inverse insight that one grasps that the square root of two cannot be the quotient of two whole numbers, however large, that one might have expected; or the crucial point of the special theory of relativity, that the common-sense notion of absolute rest is without application (see Relativity theory, philosophical significance of).) The world-order turns out to be constituted by 'schemes of recurrence' each characterized by a set of classical laws, and each with certain probabilities of emergence and of survival which are determined by statistical laws. These schemes of recurrence form a hierarchy corresponding to the different sciences (physics, chemistry, biology, sensitive psychology and so on).

Metaphysics, on Lonergan's account, is to be based on the theory of knowledge; the nature and structure of knowing determines the overall nature and structure of what is to be known. Reality is not the object of naïve extroverted consciousness, as we seem biased towards believing; it is nothing other than what is to be known so far as we exercise our attentiveness, intelligence and reasonableness to the full. 'Positions' in epistemology, metaphysics, ethics and so on, which are consistent with being intelligently conceived and reasonably affirmed, are to be distinguished from 'counter-positions', which are not so. Behaviourism in psychology is a fair example of a counter-position; since it follows inevitably from it that human beings, including behaviourists, never believe what they do because it is intelligent and reasonable for them to do so. Erroneous metaphysical doctrines are due largely to mistaking one part of the business of coming to know for the whole. Empiricism is 'a bundle of blunders', based on the fundamental error that 'what is obvious in knowing', that is, experience, 'is what knowing obviously is' (see Empiricism). Idealism rightly attends to the creative acts of intelligence which are essential in knowing; but does not take sufficient account of the fact that hypotheses concocted by intelligence may be verified or otherwise as obtaining in the actual world (see Idealism). The 'critical realism' which is for Lonergan the correct metaphysical stance takes account of the fact that knowledge is to be had in judgments as to what is so, in which experience and understanding both play their part. While taking full account of the 'turn to the subject' characteristic of philosophy since Descartes, critical realism reaffirms central elements in the philosophies of Aristotle and Aquinas.

3 Human science, ethics and religion

While the methods of the human sciences cannot be reduced to those of the natural sciences, it should not be inferred that a kind of divination is necessary to grasp the meanings implicit in human speech and action. In getting to know what Isaiah was about in propounding his oracles, or Churchill in pursuing his foreign policy, just as in the case of inquiries into the carbon cycle or the drift of the continents, I have to be attentive, intelligent and reasonable; the difference is that in the case of what is studied by the human sciences, the object as well as the subject of inquiry is to be understood as applying, at least to some extent, the same mental capacities. To envisage a fully 'scientific' account of human beings which would dispense with such 'primitive' and 'mentalistic' explanation is self-destructive for the reasons already given. All normally endowed human beings have experience, understand and judge; but it is one thing to do so, another to attend to the fact that one does so, another still to work out the consequences in an adequate philosophy. Throughout most of history, human worldviews have been dominated by myth, where thought on issues not immediately affected by practice is directed by emotion-laden images rather than by rational self-consciousness.

Value judgments, like (other) judgments of fact, may be due to more or less attentiveness, intelligence and

reasonableness; they have to be distinguished sharply from direct response to experienced inclination or aversion. I may want a good breakfast, and act accordingly; but I may also conceive and work for an economic order where as many people as possible get a good breakfast. Of course, it is one thing to make a value judgment, another thing to act in accordance with it. To complete our account of the transition from knowing what is true to knowing and doing what is good, a fourth kind of mental act, that of ‘decision’, has to be invoked; and a corresponding fourth ‘transcendental precept’, ‘be responsible’. Of course, it is generally taken for granted that seeking for the truth is an aspect of being good; and, sure enough, following the evidence where it leads is apt to involve a responsible decision - especially when one’s peers or paymasters have an interest in one’s failing to do so.

The arguments of Hume and Kant against rational theism are to be contested as due to an inadequate theory of knowledge. In accordance with a fully worked-out critical realism, one conceives and affirms an ‘unrestricted act of understanding’ (God) which understands all possible worlds, and wills the one which actually obtains; such a being must exist if the intelligibility of the world is to be adequately explained. Given the notorious fact of moral evil and human deficiency in general, one might have expected God to act in such a manner as to remedy the situation; attention to human history indicates that God has actually done so. However far rational self-consciousness is developed, the human emotional needs apt to give rise to myth remain; the remedy takes the form of an actual human history which provides humankind with the emotion-laden symbols of myth without its falsifications.

See also: Religion and science

HUGO MEYNELL

List of works

Lonergan, B.J.F. (1957) *Insight. A Study of Human Understanding*, Toronto, Ont.: Toronto University Press, 1992.(Lonergan’s most complete exposition of his philosophy.)

(1967) *Collection*, London: Darton, Longman & Todd.(Includes useful introductions to aspects of Lonergan’s thought.)

Lonergan, B.J.F. (1968) *Verbum. Word and Idea in Aquinas*, London: Darton, Longman & Todd.(Relates Lonergan’s principles to the philosophies of Aristotle and Aquinas.)

References and further reading

Crowe, F.C. (1992) *Lonergan*, London: Chapman.(A general introduction to Lonergan’s thought, against the background of his life.)

McShane, P. (ed.) (1972) *Language, Truth and Meaning: Papers from the International Lonergan Congress 1970*, Dublin: Gill & Macmillan.(A collection of appreciations and criticisms, and comparisons with other philosophical positions.)

Meynell, H. (1976) *An Introduction to the Philosophy of Bernard Lonergan*, London: Macmillan; 2nd edn, 1992. (An exposition and evaluation of the arguments of *Insight*.)

Sala, G.B. (1994) *Lonergan and Kant. Five Essays on Human Knowledge*, Toronto, Ont.: University of Toronto Press.(Compares Lonergan’s *Insight* with Kant’s *Critique of Pure Reason*, concentrating especially on the role of the a priori in the work of the two thinkers.)

Lorenzen, Paul (1915-95)

Paul Lorenzen, German philosopher of mathematics and sciences, programmatically set about implementing mathematical constructivism in wider philosophical contexts. Trained as a mathematician, he spent the greater part of his teaching career at the University of Erlangen, Germany. Here he assembled what came to be known as the 'Erlangen school', which included Wilhelm Kamlah, Kuno Lorenz, Jürgen Mittelstraß, Peter Janich, Oswald Schwemmer and others. In its heyday, the school also influenced work at the universities of Konstanz and Marburg, and was one of the main alternatives to 'traditional' philosophies such as hermeneutics.

The school's interests embraced mathematical logic, the major thinkers of the idealist and hermeneutic traditions (though not Heidegger), and a high level of philological expertise in classical philosophy. Regrettably, though, its encyclopedic initiative has remained incomplete, and the school has largely disbanded in the face of increasing polarization between straight analytical philosophy and a tougher response by existing traditions. To some extent, Lorenzen's own express interest in left-wing political traditions became a liability in view of the sobriety and retrenchment that characterized the 1980s.

1 Lorenzen's constructivism

Lorenzen's work articulates concerns brought on by the fate of philosophy under National Socialism. He distrusted instrumentalist refusals to engage with foundational issues in epistemology, since this had (it appeared) contributed to the nationalism of the cultural sciences, the positivism of jurisprudence, and the racism of those philosophers of science who proclaimed such fantasies as 'Aryan physics'. Lorenzen's 'constructivism', by contrast, starts from the insistence that it is possible to establish a firm bridgehead in reality by a version of operationalism: we know what we have created, and to the extent that we repeatably create external things we are by that token entering reality. So while we may be unable to describe 'reality' *tout court*, we can say as much as we want about those parts that we can reproduce at will by means of effective procedures.

Lorenzen contends that all knowledge is derived from privileged practices and 'operations', that there is no knowledge that is not in some sense practical knowledge, and that it is precisely active intervention that constitutes our rational being in the world. There are two major steps in Lorenzen's arguments. The first is the determination of knowledge-constituting activities: where do deliberate human practice and intelligible reality coincide as perfectly as operationalist epistemology envisages? The second step extends this epistemological principle to issues in moral philosophy.

2 Counting and other operations

Traditionally, counting, as represented by the natural numbers and arithmetic, was the constructivists' favoured 'operation', as shown in Brouwer's contention that mathematics 'is more an activity than a doctrine'. Despite supplying the basis of the most refined structures the mind is capable of conceiving, counting is a practice common to all rational beings, and one whose practical effectiveness is vindicated by repeated application in all the concerns of human life. The philosophical challenge was to show that complex mathematical instruments such as analysis - or, failing that, equivalent substitutes for them - could be derived, 'constructively', from basic intuitive operations like counting. Non-constructive steps in the argument (such as the *tertium non datur*) had to be excluded, or interpreted in terms of some additional constructive principle. This is the area from which Lorenzen's own earlier work grew.

But counting is only a single, restricted 'operation'. In particular, it does not suffice to ground other branches of mathematics, such as geometry or mechanics, which require some account of spatial and temporal elements. To obtain these, Lorenzen turned to the work of Hugo Dingler (Dingler was an opponent of the Vienna Circle; discussion of his ideas can be found in early issues of *Erkenntnis*). Dingler had argued that a Euclidean geometry could be reconstructed by reference to certain intuitive practical activities. In particular, the production of plane surfaces was something for which lens grinders had an eminently effective procedure (they grind three objects together until one designated surface of each is exactly congruent with the designated surfaces of both others). Once plane surfaces have been effectively derived, other elements (straight lines, points) could be derived by further intuitive manipulation (a straight line is the intersection of two plane surfaces, a point is the intersection of

two straight lines). Just as constructivist mathematics rejected the use of non-constructive elements in justifying maths (because they interrupted the chain of derivation from assured reality), so also Dingler rejected non-Euclidean approaches because they could not be intuitively grounded (see Dingler 1987).

The attempt to provide 'operative' foundations for mechanics - and hence for physics - has been extended into chronometry and stochastics in the so-called 'protophysics' of Lorenzen's pupil Janich. The central feature of this broadened constructivism is the attempt to find procedures that are repeatable under contextual invariance. Counting has this effect: whether you count apples or sheep, the results of this procedure (natural numbers) have the same properties in either case. In principle at least, Dingler's lens-grinding procedure in every case produces something (a plane surface) with the same properties. Clearly, few procedures satisfy this requirement of contextual invariance. Baking depends for the properties of its results on 'contextual' elements such as the ingredients I use; running depends on where I start from; and so on. But to the extent that we can find procedures that are contextually invariant, we have discovered operations that *are generally constitutive of intelligible reality*.

The most satisfactory examples of such procedures are in the schemata by which we conceptualize reality. Counting, as we have seen, is one of those. Lorenzen made two attempts, however, to extend this into the philosophy of language (see Constructivism; Constructivism in mathematics; Operationalism).

3 Ortholanguage and dialogical logic

One attempt consisted in the development of a so-called 'ortholanguage' (with Oswald Schwemmer). This was based on an inversion of the 'classical' priority of syntax over semantics. Although the constructivist interpretation of counting as a general procedure already abandons the set-theoretic model of the universe, a general prioritization of semantics requires more. The ortholanguage attempted to supply this by specifying absolutely general *pragmatic* features of speech. In its details this project was not very successful and seems to have been abandoned (the basic semantic 'procedure' was giving orders, which does not entirely commend itself!). None the less it formed part of a wider project which produced the *Logical Propaedeutics* (1973), which became the Erlangen school's most widely influential text. It is best described as a constructivist introduction to philosophical logic.

Logical Propaedeutics, apart from setting out in general terms a derivation of the principles of rational communication from certain foundational pragmatic interests, provides an interpretation of the logical constants. This is perhaps the most interesting and enduring of Lorenzen's contributions. He calls it 'dialogical logic'. The classical approach to philosophical logic, Lorenzen argues, rests on metaphors derived from set theory. The assumption is that intelligible statements about the world are reducible to elementary propositions about whether objects are or are not elements of sets. A classical approach can then interpret the logical constants - or at least the connectives - in terms of these depictions of states of affairs. Wittgenstein's use of truth tables in the *Tractatus* is a celebrated example (see Wittgenstein, L.). (Lorenzen also calls this approach 'semantic'.) Unless the system is to be disengaged from ontology and treated instrumentally, however, problems with 'difficult' sets (non-finite sets, self-membership and so on) have to be faced. In particular, truth tables presuppose that the truth-values of all the elementary propositions are determinate. But this raises difficulties, for example, in the case of the universal quantifier, and is by no means as intuitive as its proponents claim.

Lorenzen's constructivism supplies an answer. The logical constants are now interpreted as strategies in two-person games. They indicate a certain procedure in an argument; and the truth-values of elementary propositions are then presented no longer as pre-existing states of affairs, but as matters which might (or might not) fall for determination in the course of the strategy pursued. It might be, for example, that dialogue partner *P* puts forward a conditional sentence, $p \rightarrow q$. While the truth table interprets this as a collection of fixed assertions about *p* and *q*, the dialogue interpretation says it is no more than an offer to negotiate in certain determinate ways. If, for example, the opponent *O* challenges with the statement '*p*', then, in order to make good the initial claim, *P* will affirm '*q*'. If *P* cannot do that, *O* will win. If, on the other hand, *O* challenges with '$\neg q$', *P* will have to affirm '$\neg p$'; and so on. For certain affirmations there is always a winning strategy, and such formulations correspond to the 'logical truths' or 'tautologies' of the classical interpretation.

The advantage of this becomes more obvious in the case of the quantifiers. A universal quantifier is now not to be interpreted as a statement about some actual (and possibly infinite) totality of states of affairs, but as an offer to

negotiate on the basis that the assertor challenges their opponent to find a counterexample - or, after a completed dialogue, that the challenge was made and not successfully taken up.

This idea has a precursor in Gentzen’s ‘natural deduction’, and to a degree it has a rival in similar ‘games’ devised by J. Hintikka (1983). The core of Lorenzen’s approach, however, is its strong constructivism, which is most evident in its insistence on the concrete temporality of the dialogue. As Lorenzen points out, much depends on the dialogical rules adopted for these exchanges, for they determine which statements are to be tautologies, and it is over the question of what is to be a tautology and what is not that, for example, constructivists and classical logicians part company. These rules relate largely to the number of times each participant may respond to statements by an opponent (by attacking affirmations or defending themselves against attacks). If these rules are completely liberalized, and each side may continue attacking and defending until they have covered everything, then central classical tautologies (*tertium non datur*, *reductio ad absurdum*) become universally defensible strategies. These are excluded once the rules are tightened, for example, by specifying that *P* - the person making the original declaration - may *attack* any statement made by *O*, but may only *defend* himself or herself against the last attack.

Of course, the problem now arises over how such rule-making decisions may themselves be justified ‘constructively’. Ideally, one would wish to do this by showing that the dialogue rules were intuitively more satisfactory than, say, the ontological insinuations of truth tables. This is not easy. Lorenzen seems to offer two responses. One assumes that the tighter rules are the intuitive ones. However, we need most of classical logic (including *reductio*) to be able to reason at all. We therefore have to enrich our original dialogical intuition with liberalization. In other words (interpreting this) we need tolerance in order to do science.

The other response does not really follow from the dialogical interpretation, but recurs to constructivism’s ontological conviction that future facts (or otherwise unavailable facts) are not facts and hence do not fit into truth tables. This justifies rules excluding *tertium non datur*, and vindicates dialogues as part of the dynamic by means of which human beings construct their own future facts (see Dialogical logic).

4 Conclusion

Lorenzen is a transcendental pragmatist: that is to say, he thinks that certain activities provide direct access to reality, and that the conditions of such activities are thereby in some sense transcendental conditions of knowledge in general. The problem lies in identifying these activities. Counting is one possibility, and has done useful service for classical intuitionism. It is limited, however; and other attempts to extend the range of foundational activities within physics (Dingler, operationalism generally) have not been entirely convincing. The most widely canvassed transcendental activity - by Apel, Habermas and others as well as by Lorenzen - is language itself. But, as Wittgenstein asked in the *Philosophical Investigations*, to what extent is ‘language’ a single activity? The various transcendental-pragmatic attempts to determine precisely what this activity is have usually led to suggestions that are unconvincing or trite or both. In particular, the recurring suggestion (in Lorenzen, Habermas and elsewhere) that speaking somehow presupposes tolerance and openness, and thus has a natural telos in good behaviour, seems ultimately specious.

Despite any reservations one may have about the system eventually generated by Lorenzen and his followers, their work combines deep scholarship with breadth of interest and a properly philosophical emphasis on the moral and political aspects of reason.

JULIAN ROBERTS

List of works

Lorenzen, P. (1955) *Einführung in die operative Logik und Mathematik (Introduction to Operative Logic and Mathematics)*, Berlin: de Gruyter.(An early, and technical, account of Lorenzen’s constructivist philosophy of mathematics.)

Lorenzen, P. and Kamlah, W. (1973) *Logische Propädeutik; Vorschule des vernünftigen Redens (Logical Propaedeutics: Rudiments of Rational Discourse)*, Mannheim: Bibliographisches Institut.(Written with his colleague Wilhelm Kamlah, this ‘constructivist’ introduction to philosophical logic popularized the ‘Erlangen Circle’.)

Lorenzen, P. and Schwemmer, O. (1973) *Konstruktive Logik; Ethik und Wissenschaftstheorie (Constructive*

logic, ethics and philosophy of science), Mannheim: Bibliographisches Institut.(The mature presentation of Lorenzen's work.)

Lorenzen, P. (1974) *Konstruktive Wissenschaftstheorie (Constructive Epistemology)*, Frankfurt: Suhrkamp.(A collection of more popular papers.)

Lorenzen, P. (1978) *Theorie der technischen und politischen Vernunft (The Theory of Technical and Political Reason)*, Stuttgart: Philipp Reclam.(A collection of popular essays.)

Lorenzen, P. and Lorenz, K. (1978) *Dialogische Logik (Dialogical Logic)*, Darmstadt: Wissenschaftliche Buchgesellschaft.(A more technical collection of essays by Lorenzen and his pupil Lorenz.)

Lorenzen, P. (1980) *Methodisches Denken (Methodical Thinking)*, Frankfurt: Suhrkamp.(A collection of popular essays.)

Lorenzen, P. (1987) 'Critique of Political and Technical Reason', *Synthese* 71 (2): 127-218. (The Evert Willem Beth Lectures of 1980. Written in English, this series of lectures was Lorenzen's last exposition of his work.)

References and further reading

Dingler, H. (1987) *Aufsätze zur methodik (Essays on Method)*, ed. U. Weiss, Hamburg: Felix Meiner.(This selection includes Dingler's seminal 1936 attempt at an operationalist grounding of natural science.)

Felscher, W. (1986) 'Dialogues as a Foundation for Intuitionistic Logic', in Gabbay and Guenther (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 3.(A useful review of the Erlangen school's dialogical logic.)

Hintikka, J. (1983) 'Semantical Games and Transcendental Arguments', in J. Hintikka and J. Kulas, *The Game of Language*, Dordrecht: Reidel, 33-46. (Hintikka's semantical games are an alternative to Lorenzen's; this is an account.)

Janich, P. (ed.) (1992) *Entwicklungen der methodischen philosophie (Developments in Methodical Philosophy)*, Frankfurt: Suhrkamp.(Recent essays on the Erlangen school; includes extended discussion of Dingler's influence.)

Roberts, J. (1992) *The Logic of Reflection*, New Haven, CT: Yale University Press.(Relates Lorenzen to Frege, Wittgenstein, Husserl and Habermas.)

Losev, Aleksei Fëdorovich (1893-1988)

A leading Russian philosopher, religious thinker and classical scholar of the twentieth century, Losev made important contributions to the theory of language, myth and symbol, as well as to the understanding of ancient Greek thought and culture. He strongly resisted central aspects of modernization, in particular the spread of what he saw as 'bourgeois' and 'philistine' secularism, positivism, materialism, atheism, selfish individualism and 'machine civilization'.

A scholar of staggering erudition with a lifelong passion for the ancient world, especially the thought of Plato, Losev could be a fierce, sometimes abusive polemicist. His relation to the Soviet regime ranged from open defiance in the late 1920s (which led to prison and the gulag for almost three years) to at least pro forma *acceptance of certain key elements of Marxist ideology and philosophy in the period between 1953 and 1988.*

1 Philosophy of language, myth and symbol

Aleksei Fëdorovich Losev was born in Novocherkassk. While studying at that city's classical gymnasium, from which he graduated in 1911, he made his first acquaintance with the works of Plato and Vladimir Solov'ëv, both of whom were to remain at the centre of his intellectual interest for the rest of his long life. He graduated from Moscow University in 1915 in both philosophy and classics. Among his teachers were L.M. Lopatin, G.I. Chelpanov and G.G. Shpet. Boris Pasternak was a university classmate.

In 1914 the university authorities sent Losev to Berlin to do research in medieval philosophy in preparation for a university career in Russia. His stay was cut short by the outbreak of war, and he never again travelled outside Russia. In 1919 he was named a professor at the University of Nizhny Novgorod. From 1921 he taught the history of aesthetics at the Moscow Conservatory and then, with the rank of professor from 1923, taught a variety of subjects at various institutions of higher education. From 1942 until his death (in Moscow) he was a professor of classics at the Lenin Pedagogical Institute in Moscow. As a result of his trying experiences in the gulag in the early 1930s he was judged in 1947 to be clinically blind. His prodigious scholarly output during his final four decades (a lifetime total of some 400 scholarly works, including thirty monographs) was made possible by his second wife, A.A. Takho-Godi, a classicist and his former student, as well as by various polyglot secretaries.

When between 1927 and 1930 Losev was publishing his first eight, 'privately printed', volumes - what Khoruzhii has called his 'Octateuch' (*Vos'miknizhie*) - he used three techniques to confuse the Soviet censors: (1) misleading titles, especially for his 900-page *Ocherki antichnogo simvolizma i mifologii (Essays in Ancient Symbolism and Mythology)* (1930a), which in fact consists of an extended monograph on Plato's theory of ideas and shorter monographs on Plato's social philosophy and Aristotle's ethics and aesthetics; (2) a ponderous technical terminology with many Greek-based terms and a dense and difficult philosophical style - especially prominent in *Filosofiia imeni (The Philosophy of the Name)* (1927b); (3) careful avoidance of the words 'God' or 'Deity' in what are clearly theological discussions (again, most in evidence in Filosofiia imeni). It was not until his final book of the series, *Dialektika mifa (The Dialectic of Myth)* (1930b) that Losev ventured to use the word 'God' - which in fact occurs frequently in the final chapters of that work.

Losev is unique among Russian thinkers in combining close scholarship in several languages - including the translating of, and commenting on, key classical and medieval texts - with bold and sometimes fanciful speculation. His translations of works of Aristotle, Plotinus, Proclus, Sextus Empiricus and Nicholas of Cusa have been widely praised, as has the splendid four-volume translation of Plato (1968-72) that he co-edited with V.F. Asmus.

Although Losev's Russian Orthodox convictions can be read between the lines in several of his works published in the late 1920s, he managed to avoid public censure, except for one attack by A.M. Deborin in 1929. Perhaps partly as a result of this relative impunity, he was emboldened to make his criticisms of Soviet ideology and practice sharper and more explicit, and his religious views much more open, in two works published in 1930. In *Ocherki antichnogo simvolizma i mifologii* he declared that a proletarian state cannot permit the existence of free art or science; if it does, such science may begin to refute Darwin's theory. 'It is necessary to forbid the refutation of Darwin and to devote major resources to the refutation of Einstein' ([1930a: 822] 1993: 837). As he admitted to a

Western visitor in the 1960s, this was meant ironically; but the Soviet censors took it with deadly seriousness. And they were outraged by his claim that '''dialectical materialism'' is a scandalous absurdity, the total flaunting of all dialectic and a most typical abstract bourgeois metaphysics' (1930b: 147; 1994: 123).

The axe fell promptly, with denunciations in *Pravda* and *Izvestia*, and a diatribe delivered to the 16th Party Congress (1930) by L.M. Kaganovich. Losev was arrested in April 1930 and spent eighteen months in prison (four of them in solitary confinement) under interrogation. He was sentenced to ten years in the gulag, and his wife to five years, but both were released in 1933 on grounds of failing health. Losev was able to continue teaching but forbidden to publish (except one or two translations) until 1953, when, by joining in the universal praise of Stalin's pamphlet *Marxism and Problems of Linguistics*, he was able to resume publication, mainly in the field of classical philology.

In a letter of 30 January 1923 to his mentor and friend Fr. Pavel Florenskii, first published in 1990, Losev made clear that his interest in the philosophy of the name and the theory of language and symbolism generally, had theological roots. He sided with Florenskii in defending the controversial claims concerning *imiaslavie* ('glorification of the name [of God]'). His ontological model - widely applied in a variety of fields of thought and culture - is that of an essence (*sushchnost'*) and its 'energies' (*energii*). An 'energy' for Losev is not an Aristotelian *energeia*, in other words, the actualization (of a potentiality), but rather the 'manifestation' or 'expression' of an essence. As he wrote to Florenskii: 'The name of God is an energy of God's essence'. But, wishing to avoid the pantheism with which many defenders of *imiaslavie* had been charged, Losev added: 'The name of God is God Himself, but God Himself is not a name; God is above every name' (*Kontekst 1990*: 15). The theological implications of this position will be explored (briefly) in §3.

Losev envisages a hierarchy of essence-energy relations: *eidos* (Platonic Form) is manifested in myth, myth is manifested in symbol, and symbol is manifested in the person. Losev's dialectical phenomenology - as Zenkovsky aptly called it - is more dialectical than phenomenological, more indebted to Plato, Plotinus and Hegel than to Husserl. As both Zenkovsky and Khoruzhii (1994) have noted, Losev always enriched his Husserlian analysis of the structure of meanings with a very un-Husserlian intuition of inclusive 'total-unity' (*vseedinstvo*) inspired by Vladimir Solov'ëv. This is evident in his theory of dialectic and of language. Dialectic, he insists, 'is the sole method capable of grasping living reality as a [dynamic] whole'. And words - especially proper names - have a 'magical power'; they are instruments of 'living communion' of human beings with one another and with nature. A person for whom words were only sounds, devoid of meanings, would live in a 'deaf-and-dumb' universe, plunged in 'gloom and madness', an 'eternal prisoner of himself,… antisocial, uncommunicative, non-conciliar [*nesoboren*]' (1927b: 47; 1993: 642).

By 1927 Losev knew the first two volumes (1923, 1925) of Cassirer's seminal work, *The Philosophy of Symbolic Forms*. He found much that was congenial there, and was astonished to discover that Cassirer's view of myth and mythical time was very close to his own view of music and musical time. But he rejected Cassirer's sharp contrast between myth and science and asserted, against Cassirer, that myth has its own 'mythical truth' and 'mythical authenticity [*dostovernost'*]'. Furthermore, Losev insisted that mythical consciousness is an essential accompaniment of every culture, including our own, rather than an early, now superseded stage of cultural development, as many theorists of myth assumed (see Cassirer, E.).

2 Philosophy of history and critique of contemporary culture

Losev's views of European history exhibit an intriguing parallel with those of Konstantin Leont'ev. For both thinkers the high point of West European cultural and religious development came with the Middle Ages, permeated as it was by religious values and at the social level by *sobornost'* ('conciliarity' or 'organic spiritual togetherness'). (A significant difference is that Losev did not follow Leont'ev in looking to contemporary Greece and Turkey as models of effective resistance to the modernization that was sweeping Russia.) Since the Renaissance the cultural and spiritual movement of Europe has been retrograde: the Enlightenment, the French Revolution and the development of an influential bourgeoisie have resulted in social atomization, a self-centred individualism and the flattening and emptying of cultural and spiritual values in the movements of secularism, liberalism, technologism, positivism, atheism, a materialist ontology and an abstractly rationalistic, non-dialectical metaphysics - what Hegel had called *Verstandesmetaphysik*. When Zenkovsky said that Losev in his early works had developed a 'metaphysics of the name' he meant this as a compliment. Losev would have taken it as an insult,

insisting that Zenkovsky use the word ‘ontology’ if he meant simply a general theory of being.

Like Leont’ev, Losev was often abusive in his criticism of the various aspects of modernity just enumerated. Thus he spoke of the ‘nightmare of Kantian subjectivism and dualism’, branding Kant’s philosophy a ‘synthesis of a prison [the human self alone with its ‘hallucinations’] and nihilism’ (’Veshch’ i imia’ (1993: 860)). In praising Husserl for repudiating the naturalistic philosophy of his time, Losev went on to call it ‘a naturalistically-decadent [or ‘corrupt’ - *rastlennaia*] philosophy’ ([1930a] 1993). In 1962 he turned his abuse towards Husserl himself, accusing him of a ‘monstrous nihilism’ and of being a ’most dreadful illustration of the period of disintegration [*raspad*] of bourgeois philosophy’.

Losev’s repudiation of high technology is even more idiosyncratic than Leont’ev’s, and somewhat differently motivated. For Leont’ev trains, the telegraph and the telephone were destroying the independence of national and ethnic cultures; in Losev’s view machinery of every kind, along with electric light and power, are obtrusive, vulgar and destructive of high culture generally and spiritual values in particular. One can sympathize with his claim that ‘Gothic architecture is more beautiful than the latest barracks-like buildings’, and one can understand his statements that ‘Platonism is more beautiful than materialism’ and that the ‘ringing of [church] bells is more beautiful than the howling of automobiles’. But one is at a loss when Losev adds: ‘burning people at the stake is more beautiful than executing them by shooting’ (1930b). And his diatribe against electric light (which has a less hysterical prototype in a Berdiaev text of 1916) is simply baffling. For Losev electric light is ’dead, mechanical’, and exhibits the ’limitation and emptiness of Americanism and machine production’. It is ’lacking in Grace,… banal, and boring’. *Human* actions, like praying or making love, cannot be performed by electric light, but only inhuman and dehumanizing actions, such as examining a victim of torture or presenting a promissory note for collection. It is an act of ‘nihilism’ for a Russian Orthodox believer, Losev asserts, to replace a ‘living, trembling candle flame’ by the ‘trivial abstraction… of vulgar electric light’. He concludes: ‘Apartments in which there is no living flame - in a stove, in candles, or icon-lamps - are dreadful places’ (1930b: 63-4; 1994: 53-4).

There is a certain historical irony in the fact that Losev’s mentor and friend, Fr. Pavel Florenskii, no less devoutly Orthodox than Losev, was at this very time working conscientiously as a scientist and engineer to bring electricity to Soviet factories, offices and homes.

3 Philosophy of religion and philosophical theology

A key fact of Losev’s biography, not mentioned in §1, is directly relevant to his views on religion. For nearly six decades he had kept from the world, including his friends, associates and students, the fact that in 1929 he and his first wife had secretly taken monastic vows, that he was the monk Andronik. To his friends, even though many of them were aware that in his final two or three years Losev had taken to wearing a Russian Orthodox cross, this was a stunning revelation, first made public five years after his death at the Losev centennial conference in Moscow (October 1993). They might have been less stunned if they had paid closer attention to the several eloquent passages in *Dialektika mifa* celebrating the simple and devout lives of Russian Orthodox monks and nuns - their fasting, prayers and spiritual discipline.

Claims by such commentators as Nakhov (see ‘The Life and Thought of Aleksei Losev’ (1996)) that 1953-88 was Losev’s sincerely neo-Marxist period tend to be undercut by the fact that he was a Russian Orthodox monk throughout that period. This fact also focuses a clear new light on many of the discussions of religion in Losev’s works - both early and late - that had appeared on the surface to be either ambiguous or insufficiently explicit. And it throws a fresh, though not necessarily reassuring, light on Losev’s troubling statements (in his 1990 book on Vladimir Solov’ëv) about the ’ghastly historical satanism’ of such ‘amoral’ thinkers as Leont’ev, Rozanov and Nietzsche.

In his first published essay (1916) Losev had declared that ’Plato’s thought, which was… unilluminated by the Light of [Christian] Truth’, had nevertheless managed to attain ‘almost Christian revelations’, even though in Plato’s time the ‘world did not yet know the Word of God’. During the 1920s Losev spoke not of God (until the final chapters of *Dialektika mifa* in 1930) but of the ‘Primordial One’ (*Pervo-Edinoe*) or ‘Primordial Essence’ (*Pervo-Sushchnost’*). He reinterpreted the Platonic ideas as ‘radiant symbols’ which involve a ‘dialectical self-disclosure of the Primordial One’. His theology at this point is both apophatic (negative) and mystical: the Primordial Essence is knowable not in Itself but only through its ‘energies’. The name of that Essence ‘glows in

the immaculate splendour of sempiternal light… having overcome the darkness of the *meon* (non-being)'. For Losev the cosmos is a complex system of 'magical' and 'miraculous' words and meanings, all of which are grounded in the Primordial Essence, a Being - here Losev's language became quite biblical - from whose 'womb' there 'pours the inexhaustible stream of Its life and Its ever-new determinations' (1927b: 163; 1993: 732).

In his 1923 letter to Florenskii (see §1) Losev set out the last of his 'theses' concerning *imiaslavie* in these words:

> a person's energy is nothing more than the Divine energies as received by that person. In mystery and miracle God Himself, and He alone, acts authentically in His names, and the person becomes no more than a vessel of God's name.
>
> (*Kontekst 1990*: 17)

Losev would doubtless be willing to have these words stand as the valediction of the devout and learned, if sometimes irascible, monk Andronik.

GEORGE L. KLINE

List of works

Losev's main philosophical ideas are presented in eight works - ranging in length from fewer than 200 pages to more than 900 pages - privately printed between 1927 and 1930:

Losev, A. (1927a) *Antichnyi kosmos i sovremennaia nauka (The Ancient Cosmos and Contemporary Science)*, Moscow.(A close study of ancient Greek dialectic, including such principles as being and becoming and such categories as rest and motion, identity and difference, and space and time.)

Losev, A. (1927b) *Filosofiia imeni (The Philosophy of the Name)*, Moscow. (A phenomenological investigation, inspired by *imiaslavie* or onomatodoxy - the glorification of the name of God - of the objective and pre-objective structures of the name in their relation to *eidos*, *logos* and *meon*.)

Losev, A. (1927c) *Dialektika khudozhestvennoi formy (The Dialectic of Artistic Form)*, Moscow.(A highly theoretical discussion of such topics as the categorial structure of the *eidos*, essence and its 'energies', the antinomies of myth and meaning, and a typology of artistic forms.)

Losev, A. (1927d) *Muzyka kak predmet logiki (Music as a Subject of Logic)*, Moscow.(A phenomenology of pure or absolute music, a comparison of music and mathematics, and a close discussion of musical time, rhythm, melody, harmony and so on.)

Losev, A. (1928) *Dialektika chisla u Plotina (Plotinus' Dialectic of Number)*.(A Russian translation, with extensive commentary, of Plotinus' discussion of number in *Enneads* VI, 6.)

Losev, A. (1929) *Kritika platonizma u Aristotelia (Aristotle's Critique of Platonism)*.(A Russian translation, with detailed commentary, of Books 13 and 14 of Aristotle's *Metaphysics*.)

Losev, A. (1930a) *Ocherki antichnogo simvolizma i mifologii (Essays in Ancient Symbolism and Mythology)*, Moscow: Mysl'; 2nd edn 1993, with additional annotations.(Shorter essays on ancient symbolism and certain of the pre-Socratic philosophers, an extended monograph on Plato's theory of ideas, and shorter monographs on Aristotle's ethics and aesthetics and Plato's social philosophy.)

Losev, A. (1930b) *Dialektika mifa (The Dialectic of Myth)*, Moscow. This work is currently being translated into English by V. Marchenkov. A portion of this translation was published in *Symposion* 1, 1996. (Negative and positive definitions of myth; the 'dialectical' relation of contemporary mythology to science, religion and politics; and the distinction between 'relative' and 'absolute' mythology.)

Losev, A. (1963-94) *Istoriia antichnoi ėstetiki (A History of Ancient Aesthetics)*, Moscow: various publishers, 8 vols.(Losev's work in ancient thought and culture is perhaps best represented by this monumental work, the second volume of which (Moscow: Iskusstvo, 1969, 145-677) contains a magisterial study of Plato's views of art and aesthetics.)

Losev, A. (1990) *Vladimir Solov'ëv i ego vremia (Vladimir Solov'ëv and His Times)*, Moscow: Progress.(Losev's work in the history of Russian philosophy is best represented by this posthumously published volume.)The following posthumously published volumes, together with the reprint of 1930a, collect all of the eight volumes of 1927-30.

Losev, A. (1993) *Bytie. Imia. Kosmos (Existence. Name. Cosmos)*, Moscow: Mysl'.(Includes 1927a and 1927b, along with Losev's 1916 essay 'Eros u Platona' (Eros in Plato) and a previously unpublished article 'Veshch' i imia' (Thing and Name).)

Losev, A. (1994) *Mif. Chislo. Sushchnost' (Myth. Number. Essence)*, Moscow: Mysl'.(Includes 1928, 1929 and 1930b, and several previously unpublished texts.)

Losev, A. (1995) *Forma. Stil'. Vyrazhenie (Form. Style. Expression),* Moscow: Mysl'.(Includes 1927c and 1927d as well as several shorter works.)

Losev, A. (1996) *Mifologiia grekov i rimlian (The Mythology of the Greeks and Romans)*, Moscow: Mysl'. (Includes two extended works: *Antichnaia mifologiia v ee istoricheskom razvitii (Ancient Mythology in its Historical Development)*, first published in 1957, and a previously unpublished work, *Teogoniia i kosmogoniia (Theogony and Cosmogony).*)Several of Losev's shorter essays on language, symbolism and poetry are translated, by J. Hellie, in the special Losev issue of *Soviet Studies in Literature*, vol. 20 no. 2-3 (1984).

References and further reading

Averintsev, S. (1990) 'Pamiati uchitelia' (In Memory of my Teacher), in *Kontekst 1990*, 3-5.(A moving personal statement by a leading Russian intellectual.)

Averintsev, S. (1994) '''Mirovozzrencheskii stil''': podstupy k iavleniiu Loseva' (The Style of a Worldview: Approaches to the Losev Phenomenon), in *Nachala* no. 2-4: 76-87.(A perceptive discussion of Losev's 'love-hate' relation to such historical figures as Scriabin and Plato and such historical phenomena as the Renaissance and Soviet totalitarianism.)

Ehlen, P. (1996) 'A.F. Losevs personalistische Ontologie' (A.F. Losev's Personalistic Ontology), in *Studies in East European Thought* 48: 83-108.(Stresses the personalist, as opposed to the symbolist, elements in Losev's early ontology.)

Haardt, A. (1993) *Husserl in Rußland: Phänomenologie der Sprache und Kunst bei Gustav Shpet und Aleksej Losev (Husserl in Russia: The Phenomenology of Language and Art in Gustav Shpet and Aleksei Losev)*, Munich: Fink.(Part 2 contains a careful examination of the Husserlian influences on Losev's philosophy of language and theory of art.)

Khoruzhii, S. (1994) 'Ar'ergardnyi boi' (A Rearguard Action) in *Posle pereryva: Puti russkoi filosofii (After the Interruption: The Paths of Russian Philosophy),* St Petersburg: Aleteiia, 209-53.(A lucid and informed account which stresses Losev's position in the 1920s as an embattled Russian Orthodox thinker. Originally published as 'Ar'ergardnyi boi: Mysl' i mif Alekseia Loseva' (A Rearguard Action: The Thought and Myth of Aleksei Losev), in *Voprosy filosofii* (1992) no. 10.)

Kline, G.L. (1990) 'La Philosophie en Union Soviétique autour de 1930', in *Histoire de la littérature russe: Le XXe siècle*, *Gels et dégels*, Paris: Payard, 256-8, 259-63, 964-6.(An examination of Losev's early work in the light of Gustav Shpet's and in opposition to the position of A.M. Deborin, leading Marxist-Leninist philosopher of the period.)

Kline, G.L. (1994) 'Vospominaniia ob A.F. Loseve' (Reminiscences of A.F. Losev), in *Nachala* no. 2-4: 63-73. (Recollections of half a dozen conversations with Losev in Moscow between 1956 and 1968.)

Scanlan, J.P. (1984) 'A.F. Losev and the Rebirth of Soviet Aesthetics After Stalin', in J.J. O'Rourke *et al* (eds) *Contemporary Marxism: Essays in Honour of J.M. Bochenski,* Dordrecht: Reidel, 221-35.(An informative account of Losev's aesthetic views, which draws on such early works as *Dialektika khudozhestvennoi formy* (1927c) and such later works as *Problema simvola i realisticheskoe iskusstvo (The Problem of the Symbol and Realist Art)* (1976).)

Scanlan, J.P. (1994) 'A.F. Losev and Mysticism in Russian Philosophy', in *Studies in East European Thought*, 46: 263-86.(A detailed study of Losev's changing views concerning the place of mysticism in the history of Russian thought.)

'The Life and Thought of Aleksei Losev' (1996) A special issue of *Russian Studies in Philosophy* 35: 3-91. (Includes L. Gogotishvili, 'The Early Losev' (1989), S. Khoruzhii, 'The Idea of Total-Unity from Heraclitus to Losev' (1994), I. Nakhov, 'Losev and Marxism' (1991), and Fr. V.V. Mokhov, 'Aleksei Fëdorovich Losev and Orthodoxy' (1993). Referred to in §3.)

Zenkovsky, V.V. (1950) Istoriia russkoi filosofii, Paris: YMCA-Press, vol. 2, 372-8; trans. G.L. Kline, *A History of Russian Philosophy*, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, vol. 2, 833-9.(A compact and thoughtful analysis of Losev's early works - except for Dialektika mifa - in their relation to Plato, Hegel, Husserl and Solov'ëv. Referred to in §§1and 2.)

Lossky, Nicholas Onufrievich (1870-1965)

In 1922, the Russian neo-Leibnizian idealist Nicholas Onufrievich Lossky, one of his country's most distinguished professional philosophers, was banished from Russia along with more than a hundred other non-Marxist intellectuals whose influence the communist authorities feared. A prolific writer before his exile, Lossky continued to write and publish widely abroad, becoming not only the dean of the Russian émigré philosophical community but a thinker well known in Europe and the English-speaking world through many translations of his works.

The systematic structure and rationalistic tone of Lossky's philosophizing set him apart from most of his fellow Russian idealists, but like them he proceeded in his thinking from a strong conviction of the truth of Christianity; he wrote of his commitment to 'working out a system of metaphysics necessary for a Christian interpretation of the world' (1951: 266). He adhered to a radical form of theism according to which the created natural order has nothing in common ontologically with the divine order that created it.

Lossky is best known for a set of interrelated views in epistemology and metaphysics connected with what he considered his fundamental philosophical insight - the principle that 'everything is immanent in everything'. According to his doctrine of 'intuitivism' in epistemology, all cognition is intuitive; there is an 'epistemological co-ordination' of subject and object such that any object, whether sensory, intellectual or mystical, is immediately present in the mind of the knower. As the heir to a Leibnizian tradition in Russian metaphysics represented before him by Aleksei Kozlov and others, Lossky advanced a theory of 'hierarchical personalism' in which Leibniz's monads became interacting 'substantival agents' existing at various levels of development; the choices of these ideal beings generate the material world (henceLossky's term 'ideal realism' for his ontology) and their reconfigurations and reincarnations move the cosmic process towards the perfection of the Kingdom of God. In his 'ontological theory of values' Lossky affirms a metaphysical basis for absolute values and attributes all evil - including diseases and natural disasters - to the misuse of free will by substantival agents, both human and subhuman.

1 Life and neo-Leibnizian system

Lossky was born into a petty nobility family of Polish origin residing in the province of Vitebsk in western Russia. He received his philosophical education mainly at St Petersburg University (1891-8), studying under the noted Neo-Kantian Aleksandr Ivanovich Vvedenskii (1856-1925); subsequently (1901-3) Lossky also had periods of study in Germany with Wilhelm Windelband, Wilhelm Wundt and Georg Müller. He began teaching at St Petersburg University in 1900, and in 1916, by then a major figure in Russian philosophical circles, he was appointed to a chair of philosophy at the University.

A professed atheist and social radical in his youth, as an adult Lossky returned to his family's Russian Orthodox faith - a 'conversion' prompted in part by studying the thought of his religiously minded compatriots Tolstoi, Dostoevskii and Solov'ëv. At the time of the Russian revolution Lossky was active in the Constitutional Democratic Party. An opponent of the Bolsheviks both politically and philosophically, he was dismissed from his university position in 1921 and expelled from the country in November 1922.

Lossky's professorial career abroad was spent mostly in Czechoslovakia, where he taught at universities in Prague, Brno and Bratislava over the period 1922-45. In 1946 he moved to the United States, and in 1947 he was named professor at the St Vladimir Theological Seminary in New York. He retired from teaching in 1950. A frequent visitor to France in his later years, Lossky died near Paris on 24 January 1965.

The only Russian predecessor whose philosophical influence Lossky explicitly acknowledged was Aleksei Aleksandrovich Kozlov (1831-1901), a professor of philosophy at Kiev University from 1876 until illness forced him to retire in 1887, when he settled in St Petersburg. Thereafter developing his views in the private journal *Svoë slovo (A Personal Word)* (1888-98), Kozlov is credited with initiating a Leibnizian movement in Russian philosophy that was continued not only by Lossky but by Lev Mikhailovich Lopatin (1855-1920), who became professor of philosophy at Moscow University and president of the Moscow Psychological Society; by Kozlov's son, Sergei Alekseevich Askol'dov (1870-1945), who was a fellow student of Lossky's at St Petersburg University and is known particularly for his works in epistemology; and by a number of other less prominent thinkers.

The Russian Leibnizians all subscribed to a form of pluralistic idealism or personalism that was inspired by Leibniz's monadology (see Leibniz §§4-7). They saw in the monadology a conceptual framework for establishing the reality and independence of the individual person without abrogating the principle of the organic unity of the cosmic order. Even Leibniz's supposedly 'windowless' monads (see Leibniz, G.W. §4), these thinkers pointed out, already contained the stamp of the entire universe within each of them; the only thing needed for a fully interpenetrating, organic cosmos is to release the monads from their causal isolation and allow them to interact. This step was taken by Kozlov, who elaborated a metaphysical system called 'panpsychism', according to which individual conscious spirits are both substances and agents; Kozlov also, like Lossky after him, developed an intuitionistic epistemology. Lopatin, too, emphasized the agency of the spiritual entities that comprise reality, attributing to them the 'creative causality' from which all spatiotemporal causality is derived. A defender of free will, Lopatin was the first of this group of thinkers to devote close attention to questions of ethics. The Russian Leibnizians regarded the material world as a product of the activity of individual spirits and advanced theories of reincarnation based on Leibniz's concept of metamorphosis.

Lossky, the longest-lived and most productive member of the group, developed a worldview that was unusual in the history of Russian philosophy for its deliberately systematic character and its rationalistic elaboration. It was also an unusually comprehensive system, incorporating logic, philosophical psychology, epistemology, metaphysics, philosophy of religion, ethics and aesthetics - to each of which fields Lossky devoted at least one major book. The only traditional philosophical discipline that Lossky largely avoided was social and political philosophy - a surprising omission for a Russian philosopher.

In the sections that follow, Lossky's system is examined under the various labels that he himself employed to identify its principal elements.

2 Intuitivism

Lossky's intuitionist epistemology was sketched in his first book (1903) before its formal presentation in *Obosnovanie intuitivizma (The Intuitive Basis of Knowledge)* (1906). Its inspiration, however, dates from about 1898, as Lossky explains in an autobiographical passage in a subsequent work. He relates that, driving through St Petersburg in a cab one foggy autumn evening, sunk in solipsistic doubt,

> I was thinking that there are no sharp boundaries between things, when suddenly the thought flashed into my head: 'Everything is immanent in everything'. I sensed at once that the riddle was solved, that the elaboration of this idea would provide the answer to all the questions that disturbed me…. From that time on the idea of an all-pervading cosmic unity became my guiding thought. Developing it led me to *intuitivism* in epistemology, and to *the organic worldview* in metaphysics.
>
> (1938: 157; original emphases)

Lossky believed that this insight (dubbed the doctrine of 'absolute immanentism' by the eminent historian of Russian philosophy, V.V. Zenkovsky) pointed the way to overcoming what he called the 'epistemological individualism' of philosophers in the Cartesian tradition. These philosophers, according to Lossky, typically hold that because the only content of consciousness is subjective states, the knowing subject and the object of knowledge are separated from each other and can be connected only causally. Efforts to specify the causal relation invariably result in some form of subordination, he argues - subordination either of subject to object (materialism) or of object to subject (spiritualism). But if immanence is mutual, subject and object are immanent in each other and neither is subordinate, so that the correct conception is what Lossky calls the 'epistemological co-ordination' of subject and object. The cognized object itself is present in consciousness and is grasped as it is in itself (not as a copy, image or other representation). This immediate grasp of the object in itself is what Lossky means by 'intuition'.

Furthermore, the 'absolute' character of immanence entails that *all* objects of consciousness, of whatever sort, are immanent in consciousness, so that every object is directly 'given'; epistemologically, there is no root difference between sense perception and the spiritual contemplation of God. Hence Lossky affirms that all cognition is grounded in intuition, although he concedes that the establishment of *knowledge* on this ground requires further acts of attention, discrimination, and the like. In his book *Chuvstvennaia, intellektual'naia i misticheskaia intuitsiia (Sensory, Intellectual and Mystical Intuition)* (1938), Lossky identifies three varieties of intuition, which

he differentiates not epistemologically but according to the ontology of their objects: sensory intuition has as its object ‘real being’ - being with spatiotemporal existence; intellectual intuition is addressed to ‘ideal being’ - abstract entities conceived on the model of Platonic ideas; mystical intuition, finally, is intuition of ‘metalogical being’ - being that transcends the laws of identity, noncontradiction, and excluded middle, such as God. Lossky considered philosophical discourse a form of intellectual intuition, and thus held that rational, discursive thinking, rather than standing opposed to intuition, is one of its species.

The prominence of intuition in Lossky’s philosophy suggests comparisons with his French contemporary, Henri Bergson. Lossky knew Bergson’s work well and spoke highly of some of his views, especially the analysis of the physiology of perception and the theory of dream memory advanced in *Matter and Memory* (1896). At the same time, Lossky believed that there was a profound difference between his intuitionism and that of Bergson: for the latter, intuition discloses a conceptually unarticulated, irrational reality, whereas intellectual intuition as Lossky understood it displays the thoroughly rational and systematic structure of ideal being.

Zenkovsky criticizes Lossky’s statement of his theory of intuitivism on the grounds that it supposedly contains no actual intuitions; he finds in Lossky no ‘radiance of truth’, but only ‘ordinary theoretical *constructions* or *hypotheses*’ ([1948-50] 1953: 660-1; original emphases). But if we accept Lossky’s view of rational philosophical discourse as proceeding through the intuition of ideal being (‘intellectual intuition’), then ‘ordinary theoretical constructions’ do not exclude but rather presuppose the presence of intuitions. Moreover, Lossky’s initial consciousness of universal immanence (the sudden thought that ‘everything is immanent in everything’) seems a classic instance of the ‘radiance of truth’, on a par with the intuitive sense of ‘the oneness of being’ found in many mystical philosophies through the ages.

A more weighty criticism voiced by Zenkovsky has to do with Lossky’s apparent inability to explain the possibility of radical cognitive error. Lossky is willing to acknowledge no more than a partial deviation from truth in the epistemological process whereby intuition becomes knowledge; ‘like the other forms of intuition’, he writes, ‘mystical intuition cannot lead to ideas and concepts that are completely erroneous’ (1938: 198). In Lossky’s epistemology the fundamental coincidence of idea and reality is guaranteed by the mutual immanence of subject and object; indeed, if intuition is the ground of all cognition, and if in intuition the object *itself* is present in consciousness, it is difficult to understand how even a partial cognitive failure can occur. Yet surely, as Zenkovsky points out ([1948-50] 1953: 673), our ideas are at times radically erroneous. The reality of error requires us to assume a disjunction between subject and object that Lossky’s philosophy of ‘absolute immanentism’ cannot admit.

3 Theism

The ‘organic worldview’ to which Lossky subscribes in metaphysics is a view of the ‘world’ or cosmic order, but it does not describe the relation between God and that world. In Lossky’s metaphysics (best expounded in his book, *Mir kak organicheskoe tseloe (The World as an Organic Whole)* (1917)), the cosmic order within which organic unity prevails is a realm of being utterly different from the supernatural realm to which it owes its existence. Arguing that a world system cannot contain the ground of its own existence, Lossky contends that its ground can only be a Supracosmic Principle that transcends all systems. This is the realm of ‘metalogical being’ - a realm to which such cosmic categories as ‘reason’ and ‘person’ do not apply. Lossky holds, however, that in religious experience we recognize this Supracosmic Principle as God, as a Person, as Triune and the like, but he cautions that all such designations are based on ‘metalogical analogy’. Even the relation between the two realms requires analogical language: it can and should be called ‘creation out of nothing’, but ‘creation’ must be understood noncausally, because a causal relation requires some degree of ontological homogeneity, which does not exist in the relation between these two realms.

This dualism distinguished Lossky’s outlook from the monistic current, inspired by Vladimir Solov’ëv, that was more prevalent among Russian idealist philosophers of the early twentieth century (see Solov’ëv §§1-2, 5). Solov’ëv’s concept of ‘total-unity’ (*vseedinstvo*) incorporates Creator and created in a comprehensive organic whole with no ontological gulf between them. Whereas for this reason Solov’ëv and his followers are sometimes charged with pantheism, Lossky recommends his own view as the antithesis of pantheism, ‘the purest form of theism’ (1951: 265). Lossky does profess adherence to the doctrine of ‘Sophiology’ developed by Solov’ëv and other Russian idealists (see Solov’ëv §2), but he insists that Sophia is not an aspect or element of the Deity but is a

purely created spirit, standing at the head of all creation ('next to Christ as His closest co-worker') and receiving earthly incarnation in the form of the Virgin Mary.

4 Hierarchical personalism and ideal realism

In its treatment of the created order, Lossky's metaphysics marked the culmination of the neo-Leibnizian movement of ontological pluralism or personalism in late nineteenth- and early twentieth-century Russia. Agreeing with Kozlov that the myriad ontological units that comprise the cosmic order are not isolated or 'windowless', as Leibniz had conceived them, Lossky calls his interacting monads 'substantival agents'. He regards them as ideal (that is, non-spatiotemporal), active entities that are created by God as free agents and that ground the reality of every individual thing in the universe, from the smallest subatomic particle to complex objects at all levels of development. Even the simplest of these entities has the ability to 'choose' to subordinate itself to others to form a molecule, a cell or an organism. Indeed each substantival agent can rise to become eventually a human person, defined by Lossky as an ideal being capable of acknowledging absolute values and recognizing the duty to implement them; at still higher levels of development stand such 'persons' as tribes, states, heavenly bodies and the universe as a whole. On this basis, Lossky calls his outlook 'hierarchical personalism', with the qualification that lower, less developed substantival agents are merely potential rather than actual persons.

But what, then, of the spatiotemporal form assumed by these ideal agents in the material world? According to Lossky it is a by-product of their own free choices. To the extent that substantival agents, created by God, reject the Divine Principle and strive 'egoistically' to set themselves apart from the supraempirical Kingdom of God, their 'struggle' and 'acts of attraction and repulsion' produce the phenomenon of materiality. Thus space, time and matter are not God's creations. God creates only substantival agents, and he creates them radically free and indeterminate; hence he is not responsible for what each substantival agent makes of itself. Because on this view 'real' (spatiotemporal) being is a result of choices made by 'ideal' being in the form of substantival agents, Lossky calls his ontology 'ideal realism'.

It is not easy to square Lossky's conception of the indeterminate nature of substantival agents with the confidence he simultaneously exhibits in the progressive, teleological character of cosmic development, expressed in such statements as 'the ultimate universal victory of the good is guaranteed by the structure of the world' (1941: 79). Although on the one hand each substantival agent is said to forge its own destiny freely from the moment of its creation, on the other hand it would appear that the 'freedom' of the agent is limited by a direction (from the simple to the complex) and a goal (the production of human life) that are inherent in cosmic evolution. Thus Lossky wrote in one of his more imaginative passages:

> The human self is an agent that perhaps led the life of a proton billions of years ago. Then, gathering around itself a few electrons, it assumed the type of life of oxygen. Making its body still more complex, it then raised itself to the type of life of, say, a droplet of water. After that it made the transition to the type of life of a unicellular animal. After a series of reincarnations - or better, a series of metamorphoses, to use Leibniz's term - it became a human self.
>
> (1941: 33)

More tellingly, Lossky contended that through continued reincarnation 'all agents sooner or later overcome their selfishness' and enter into the Kingdom of God. He introduces the term 'normal [normative] evolution' for 'the line of development which leads straight to the threshold of the Kingdom of God' (1951: 264). To account for this benign tendency on the part of free agents, Lossky can only appeal to a principle of 'the inseparability of existence and value' that he says exists unconsciously as an ideal in every individual (1941: 51). But if individuals do harbour such an ideal, they are not fully indeterminate.

Questions of consistency aside, Lossky's doctrines of the voluntaristic character of evolution (even inorganic evolution) and of reincarnation as an evolutionary mechanism generated much controversy among his fellow Russian religious philosophers. Zenkovsky dismisses the doctrines as 'the most fantastic hypotheses' ([1948-50] 1953: 663).

5 The ontological theory of values

Lossky's conception (cited in §4) of a fully developed person as a substantival agent capable of acknowledging

absolute values presupposes that such values exist. In his ethical works (the most important of which are his 1931 and 1949 books), Lossky criticizes ethical relativism and elaborates what he calls his ‘ontological theory of values’, according to which absolute values are found in ‘the fullness of being’ that is God, as known through religious experience. In accordance with this conception, everything existing is valued from the point of view of its participation in and communion with the divine ‘fullness of being’: ‘Existence that brings us nearer to the absolute fullness of life is a positive value, and that which draws us away from it is a negative value’ (1951: 258).

Because spatiotemporal being results, as we have seen, from the wilful self-assertion of free substantival agents, the material world is truly a ‘fallen’ realm for Lossky; to call it an incarnation of evil would be redundant, for incarnation in itself is evil: ‘The whole [of] nature consists of entities which would have been members of the Kingdom of God had they not entered the path of egoism’ (1951: 262-3). The ’path of egoism’ or self-assertion is thus the source of all evil, natural as well as humanly produced. Unlike some other religious thinkers, who attribute humanly caused evil to the misuse of free will but have no comparable explanation of diseases and natural disasters such as earthquakes, Lossky ascribes the latter evils as well to immoral choices (presumably by subhuman agents such as molecules, cells and tectonic plates): all are ‘kinds of derivative evil necessarily connected with the relative separation of agents from one another and leading to ruptures and dissolution’ (1951: 262). Natural evils will be eliminated (along with humanly created evils) only when nature itself ceases to exist - as it indisputably will when all agents have attained the path of divine righteousness.

But if the explanation of natural evil presents Lossky with no special difficulty, on another level he is faced with the more general problem of explaining why there is evil at all, natural or human. He holds that evil at every stage of evolutionary complexity is produced by the will of substantival agents - specifically, by their will to elevate themselves above God. But what is there in the nature of these ideal beings as they come from the hands of God to induce them to make such bizarre and ultimately self-defeating choices - especially since, as we have seen, they come equipped with an inborn sense of ideal value? In the words of Natalie Duddington, Lossky’s first and principal English translator, ‘How could beings so admirably fitted by their Creator for a life of blissful intercommunion with Him suddenly conceive the impious thought of putting themselves in his place?’ (1968: 10). Duddington justifiably concludes that in Lossky’s account of the cosmic order as originally created by God there is nothing to explain the initial and continuing presence of evil.

Thus the problem of accounting for cognitive error, noted in Lossky’s epistemology in §2, is echoed in his ethics by the problem of accounting for moral error.

JAMES P. SCANLAN

List of works

Lossky, N.O. (1903) *Osnovnye ucheniia psikhologii s tochki zreniia voliuntarizma (Basic Doctrines of Psychology from the Standpoint of Voluntarism)*, St Petersburg: St Petersburg University Press.

Lossky, N.O. (1906) *Obosnovanie intuitivizma*, St Petersburg: M.M. Stasiulevich; trans. N. Duddington, *The Intuitive Basis of Knowledge: An Epistemological Inquiry*, London: Macmillan, 1919.

Lossky, N.O. (1917) *Mir kak organicheskoe tseloe*, Moscow: Leman & Sakharov; trans. N. Duddington, *The World as an Organic Whole*, London: Oxford, 1928.

Lossky, N.O. (1922) *Logika (Logic)*, Petrograd: Nauka i shkola, 2 vols.

Lossky, N.O. (1927) *Svoboda voli*, Paris: YMCA-Press; trans. N. Duddington, *Freedom of Will*, London: Williams & Norgate, 1932.

Lossky, N.O. (1931) *Tsennost’ i bytie*, Paris: YMCA-Press; trans. S. Vinokooroff as Part I of N.O. Lossky and J.S. Marshall, *Value and Existence*, London: Allen & Unwin, 1935.

Lossky, N.O. (1934) *Dialekticheskii materializm v S.S.S.R. (Dialectical Materialism in the USSR)*, Paris: YMCA-Press.

Lossky, N.O. (1938) *Chuvstvennaia, intellektual’naia i misticheskaia intuitsiia (Sensory, Intellectual and Mystical Intuition)*, Paris: YMCA-Press.

Lossky, N.O. (1941) *Bog i mirovoe zlo: Osnovy teoditsei (God and Worldly Evil: Principles of a Theodicy)*, Berlin: Za Tser’kov’.

Lossky, N.O. (1949a) *Usloviia absoliutnogo dobra (Conditions of the Absolute Good)*, Paris: YMCA-Press.(First circulated in mimeographed form in Bratislava in 1944.)

Lossky, N.O. (1949b) ‘Absolute Criterion of Truth’, in *The Review of Metaphysics* 2 (8): 47-96.(The original

Russian text of this work was first published posthumously under the title ‘Intuitivism’ in *Grani* (1970) 77: 144-70; 78: 212-40.)
Lossky, N.O. (1951) *History of Russian Philosophy*, New York: International Universities Press.(Includes Lossky’s summary of his own philosophy and that of the other Russian neo-Leibnizians, 158-62, 251-66, 381-83.)
Lossky, N.O. (1961) ‘Smysl istorii’ (The Meaning of History), in *Vol’naia mysl’* 3 (September), Buenos Aires; repr. in *Zapiski russkoi akademicheskoi gruppy v S.Sh.A.* 19: 95-105.(This article is a good late statement of Lossky’s ‘hierarchical personalism’.)
Lossky, N.O. (1968) *Vospominaniia: Zhizn’ i filosofskii put’ (Memoirs: Life and Philosophical Journey)*, Munich: Wilhelm Fink; repr. St Petersburg: St Petersburg University Press, 1994.
Lossky, N.O. (1991) *Izbrannoe (Selected Works)*, Moscow: Pravda.(Includes the complete Russian text of the 1906, 1917 and 1927 works listed above.)
Lossky, N.O. (1996) *Mir kak osushchestvlenie krasoty: Osnovy estetiki (The World as the Manifestation of Beauty: Principles of Aesthetics)*, Moscow: Progress Publishers.(The first publication of this work, long thought to have been left unfinished but actually completed by Lossky before his death.)

References and further reading

Askol’dov, S. (1912) *Aleksei Aleksandrovich Kozlov*, Moscow: A.I. Mamontov.(A study of the life and work of the originator of Russian neo-Leibnizianism by his son and follower.)
Askol’dov, S. (1914) *Mysl’ i deistvitel’nost’ (Thought and Reality)*, Moscow: Put’.(Askol’dov’s best work on problems of epistemology.)
Duddington, N. (1968) ‘Epistemology and Metaphysics: Lossky, Frank, and Others’, Conference on Idealist Philosophy in Russia, Aix-en-Provence, France, 25-29 March (mimeographed).(Referred to in §5.)
Kohanski, A.S. (1936) *Lossky’s Theory of Knowledge*, Nashville, TN: Vanderbilt University.(An exposition and analysis of Lossky’s epistemology and its metaphysical implications; contains bibliographies of works by and about Lossky.)
Kozlov, A. (1888-98) *Svoë slovo (A Personal Word)*, St Petersburg: A. Kozlov, 5 vols.(Referred to in §1. Kozlov’s mature exposition of his philosophy.)
Lopatin, L.M. (1911) *Filosofskie kharakteristiki i rechi (Philosophical Characterizations and Addresses)*, Moscow: Put’.(Essays on Leibniz and other Western philosophers as well as several Russian philosophers by a prominent neo-Leibnizian.)
Lopatin, L.M. (1886-91) *Polozhitel’nye zadachi filosofii (The Positive Tasks of Philosophy)*, Moscow: Lisner & Roman, 2 vols.(Lopatin’s basic work in metaphysics and epistemology.)
Lossky, B. and Lossky, N. (1978) *Bibliographie des oeuvres de Nicolas Lossky*, Paris: Institut d’Études Slaves. (An excellent complete bibliography of works by N.O. Lossky and translations of his works, prefaced by a detailed chronology of his life.)
Jakovenko, B., ed. (1932) *Festschrift N. O. Losskij zum 60. Geburtstage (A Festschrift for N.O. Lossky on His Sixtieth Birthday)*, Bonn: Friedrich Cohen.(Contains a bibliography of Lossky’s works up to 1932.)
Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, 2 vols; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, vol. 2, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, 630-76.(Referred to in §2 and §4. Exposition and critique of the philosophies of Lossky and the other Russian neo-Leibnizians by the authoritative historian of Russian philosophy.)

Lotze, Rudolph Hermann (1817-81)

Lotze was among the pre-eminent figures in German academic philosophy between the demise of Absolute Idealism and the rise of Neo-Kantianism proper. He sought to avoid two extremes: first, that of an idealism which seeks to deduce the world from a single, general principle; and, second, that of a realism which, by divorcing reality from the mind, splits the world into two utterly separate spheres. The search for knowledge should be tempered by a recognition of the results of natural science and sobered by the awareness that reality will, by necessity, always outstrip thought. Furthermore, our mental life cannot be reduced to purely intellectual functions: feelings and evaluations, for example, are also an integral part of human existence. While there can be no a priori deduction of a metaphysical system, a teleological interpretation, which elucidates the ultimate value of man and the world, must supplement purely naturalistic explanation. The universe has the significance of an unfolding plan, where things are subject to the general laws of order, expressing spiritual import. In this way, Lotze combined a kind of respect for the findings of scientific research with his own peculiar idealistic programme.

1 Philosophy of biology

Hermann Lotze was born in Bautzen and attended the Gymnasium at Zittau. At the University of Leipzig he studied medicine and philosophy. In 1844 he was nominated at Göttingen as the successor to Herbart in the chair of philosophy. Here he remained until the call to Berlin in 1881. His work after the ‘materialism controversy’ (from the convention at Göttingen in 1854) made clear his anti-naturalism and his affinities with the epistemological bent of early Neo-Kantianism (see Neo-Kantianism §2). A quasi-systematic philosopher, Lotze none the less eschewed the presumptions of German Idealism while preserving its ethical imperative.

As a student of Christian Hermann Weisse,Lotze fell under the sway of his teacher’s post-Hegelian view of idealism (so-called *Spätidealismus*). Yet as a medical scientist he came to reject key aspects of the idealist and Romantic philosophy of nature, most notably, its dependence upon the notion of ‘vital force’ (*Lebenskraft*) (see Naturphilosophie). After Kant there had quickly arisen an acceptance, within biological thought, of a special vital force, indicative of an organizing principle of organic bodies.Lotze presented a series of arguments designed to correct false applications of this concept. He reiterated Kant’s warning that the assumption of purposive organization could only be a regulative (not a constitutive) principle. As part of an organizing principle of merely subjective validity, the concept of vital force must not be hypostatized: that is, such force ought not to be treated as a material *cause* of organic phenomena.

Therefore, while inquiry must begin with the assumption of a purposive organization, no conclusion should be drawn which identifies vital force as an actual causal principle in the biological realm. On the other hand, this assumption does not imply that mechanistic explanation may be precluded. Teleology and mechanism are mutually compatible, not competing, investigative procedures. Lotze rejects the notion of a special force as superfluous for the scientific side of the investigation. Yet the formal constraint of ‘order’ or ‘purpose’ remains a necessary presupposition: mechanism should be understood as the way in which purpose realizes itself in our world.

As some concept of finality is required to fill the ‘metaphysical vacuum’ left by the intrusion of mechanistic procedures (not only in life science, but in the human sciences as well), an idealistic *Weltanschauung* is combined with a clear recognition of the power of scientific explanation. This requirement harks back to Fichte’s imperative to determine what *is* on the basis of what *ought* to be, but with an important difference: this ethical requirement, in Lotze’s hands, was translated into a new language of ‘values’. The terminology of ‘is’ and ‘ought’ is replaced by that of ‘existence’ and ‘meaning’: science may be allowed to determine what actually exists, but philosophy is required to expound what this existence means. These early convictions are retained in the mature expositions of Lotze’s thought, found in his *Mikrokosmus (Microcosmos)* (1858-64) and *System der Philosophie (System of Philosophy)*. Of the latter, which was intended as a three-volume work, only two volumes had been published at his death: *Logik (Logic)* (1874) and *Metaphysik (Metaphysics)* (1879).

2 Logic

The first book of the first volume of Lotze’s *System* begins with ‘pure logic’. He starts by affirming that our mental life - the ‘current of ideas’ - essentially involves ‘connection’: ideas appear in relations of succession or

simultaneity. But does this connection denote real coherence or mere coincidence? That is, are the associative connections found in the current of ideas sufficient for objective knowledge or indicative only of subjective error and illusion? Lotze concludes that coherence is distinguished from mere conjunction only by reference to a ground (or basis) for such coherence. And this investigation is the province of ‘thought’ proper.

Hence to logic, and not to psychology, belongs the discovery of the bases of justification. The dog can associate the raised stick (in the hand of its master) with pain, but only human thought can relate the identical matter in inferential terms, in virtue of logical consequence. The peculiar nature of thought thus lies in the supplement that it provides over and above the mere current of ideas: the addition of a justificatory ground (or rationale) for the connection. This, in turn, depends critically upon the capacity of thought to impose logical form. To provide for a firm structure of truth, it must arrange its building stones precisely. This foundation must, further, be of a form suitable for thought; so there must exist some process by which thought logically apprehends the sensory manifold.

Lotze believed that concept-formation belongs to thought proper and not to the psychological processes of abstraction or synthesis. This principle is to be established by a careful determination of the different stages involved in the refining of ideation into thought. The first step in the creation of logical building stones is the conversion of ‘impressions’ into ‘ideas’. This takes place by naming: in the process of designation one separates out objective ‘content’ from subjective ‘act’. The logical objectification (of what initially appears as subjective) proceeds through the categorization effected by the recognized parts of speech. In meaning (as opposed to mere venting) one calls upon the formal, grammatical categories of substantive, adjective or verb. These are mirrored in (or are shadows of) the logical categories of object, property and relation. While thought is not dependent upon language *per se*, it does depend upon some such inward articulation that respects these fixed logical categories; by contrast, the musical scale is an articulated structure, but not one which can support propositional thought.

Lotze proposed few formal innovations in logic. Among them was that found in his insistence that the coordination of part-concepts (*Merkmale*) in the intention of a given concept be expressed not in simple additive relations but rather in a functional representation. Instead of a simple equation, such as $S = a + b + c + \ldots$, a functional notation, such as $S = F(a, b, c, \ldots)$, indicates that the relations of dependence among the part-concepts must be determined according to a general rule or law. (The constant reference to rules and laws furnishes a leitmotiv of Lotze’s entire philosophical endeavour.)

3 Epistemology

Among Lotze’s best-known contributions to the theory of knowledge, found in the third book of his *Logic* (1874), is his reinterpretation of Platonism: the construal of metaphysical objectivity as epistemic validity. Lotze argued that Plato had been fundamentally misunderstood. In specifying that the Ideas were non-spatial, Plato had intended to imply that they were not ‘things’ at all. (Here ‘Idea’ invokes thought-*content*, not an *act* of ideation.) These contents are to be understood as having an existence completely different from the type of existence associated with objects or things. While physical objects can possess ‘being’ in the strict sense and events merely ‘occur’, such contents have ‘validity’. This implies a kind of epistemological realism: there are thoughts (or propositions) which are true or false and which are true or false independently of and antecedent to our judgment of them.

Such epistemic realism is sharply distinguished from any metaphysical variety. The latter view holds that these thought-contents must exist (or subsist) in some particular place. It may entail a false hypostatization that seems to follow from the correct emphasis we give to the independence of our thoughts from us. To assert the ‘objective significance’ of thought-contents is, hence, not to ascribe to them any sort of ‘real’ existence. Our apprehension of a thought seems to presuppose its existence in a preordained ‘place’; yet this can be nothing other than its setting within the inferential relations which hold between different thoughts and their meaningful components. As such, it is part of ‘a whole’: whatLotze calls the world of the ‘thinkable’. It is thus that the apprehension of a thought-content, as part of a larger system, may appear to be analogous to the perception of a physical object.

This picture affects the approach to epistemic justification that emerges in Lotze’s views on the validity of the Euclidean axioms. Even if our spatial perception depended upon some cognitive capacity such as ‘intuition’, Lotze argued, this fact alone could not account for our taking these axioms and their derivable theorems to constitute justified belief. Lotze tried to detach the concept of the a priori (as justificatory) from that of the innate (as genesis or source). Any sense in which one might speak of the innateness of Euclidean geometry only indicates causal or

psychological origin and is irrelevant to the question of justification. Instead, Lotze characterized a priori truths as those rooted in 'truths of universal validity, and thus prior to the particular instances in the sense of being rules by which they are determined' ([1874] 1887 2: 131). Yet this exhibition of truth must not be confused with the panlogicist attempt to discover truth by means of an a priori deduction. Rather, it is 'reduction' (or the regressive method) which makes possible the apprehension of the ultimate presuppositions of thought.

4 Metaphysics

Lotze also affirmed the centrality of 'law' in his conception of metaphysics. While metaphysics has a role in his system, it is a limited one: namely, to delimit not the thinkable, but the 'actual' - that is, what can be conceived of as actually existing or occurring without pain of contradiction. This constraint brings him to attack Herbart's theory of being in the first book of the *Metaphysics* (1879). The Herbartian view is indicated in the notion of 'pure being' as *Setzung*: that 'what we call the true Being should be found only in the pure "position", void of relation'. In fact, on this view, it is only because pure being lacks relations that existing things are able to enter into relations at all. Herbart concludes that reality is composed of something plural, simple and indestructible: what he called 'the Reals'. Since reality is compounded of substances that can suffer no change, and are hence immutable, the relations (into which things 'enter in') must be utterly external and accidental (see Herbart, J.F.).

Lotze's polemic stands in sharp opposition to this. He proceeds with an eye to the possible contradictions inherent in such an approach. On Lotze's account, relations 'between' things are impossible: therefore such relations must be internal. This relatedness is characteristic of being, yet not essentially so: what being *means* is something indefinable. Yet, the reality of being is ultimately exhausted by the reality of its relations. Being must be portrayed as part and parcel of a determinate and interconnected 'whole'. Hence, it is inconceivable that one might arrive at pure being through abstraction, by an attempt to negate all relations. This activity could only empty the concept of 'position' of all possible content.

Lotze also questions the view which makes the being of things (*qua* self-subsistent things) equivalent to a substantival notion of the Real. He points out that since 'real' can be used as an adjective, it may indicate a property of things. So the term 'real' cannot be used to characterize the being of things any more than the term 'position' can. Since what is real must be seen as part of an ordered system, its 'essence is only to be found in a law'. This points the way toward an interpretation of both the individual thing and the larger cosmos as systematic wholes, not as accidental juxtapositions of qualities or entities. All must be recognized as the ultimate workings of some Good, which orders and gives purpose to our reality.

WhileLotze formed no school, he influenced, most notably, W. Dilthey, R. Eucken, C. Stumpf, W. Windelband, G. Frege, E. Husserl, H. Rickert and B. Bauch. Abroad, his readers and admirers included Josiah Royce andF.H. Bradley. Lotze was a figure who marked the transition from the classical era of Goethe andHegel to a more scientifically informed philosophy. An original thinker in his own right - and despite his wide-ranging influence - todayLotze remains in relative obscurity.

DAVID SULLIVAN

List of works

Lotze, R.H. (1841) *Metaphysik (Metaphysics)*, Leipzig: Weidmann.(The 'lesser' metaphysics, published whileLotze was a *Docent* at Leipzig.)

Lotze, R.H. (1843) *Logik (Logic)*, Leipzig: Weidmann.(The 'lesser' logic.)

Lotze, R.H. (1852) *Medicinische Psychologie, oder Physiologie der Seele (Medical psychology, or physiology of the soul)*, Leipzig: Weidmann.(Lotze's physiological psychology, notable for its discussion of his theory of 'local signs'.)

Lotze, R.H. (1856-8, 1858-64) *Mikrokosmus, Ideen zur Naturgeschichte und Geschichte der Menschheit, Versuch einer Anthropologie*, Leipzig: S. Hirzel, 3rd edn, 1876-80; trans. E. Hamilton and E.E.C. Jones *Microcosmos: An Essay Concerning Man and his Relation to the World*, Edinburgh: T.&T. Clark, 1885.(Lotze's popular exposition of his views on human nature and on the meaning of human existence.)

Lotze, R.H. (1857) *Streitschriften (Polemical writings)*, Leipzig: S. Hirzel.(A reply to criticism by I.H. Fichte, a 'late-idealist'.)

Lotze, R.H. (1868) *Geschichte der Aesthetik in Deutschland (History of aesthetics in Germany)*, Munich: J.G. Cotta.(A three-part work, containing a detailed review of prior German aesthetic theory.)

Lotze, R.H. (1874) *Logik, Drei Bücher vom Denken, vom Untersuchen und vom Erkennen*, Leipzig: S. Hirzel, 2nd edn, 1880; *Logik, Erstes Buch, vom Denken*, and *Logik, Drittes Buch, vom Erkennen*, ed. G. Gabriel, Hamburg: Felix Meiner, 1989; trans. and ed. B. Bosanquet, *Logic, in Three Books: Ontology, Cosmology and Psychology*, Oxford: Clarendon Press, 1884; 2nd edn, 1887.(The first volume of Lotze's incompleted *System der Philosophie*. The new German edition, in two volumes for Books One and Three, contains both valuable introductory essays and extensive bibliographies.)

Lotze, R.H. (1879) *Metaphysik, Drei Bücher der Ontologie, Kosmologie und Psychologie*, Leipzig: S. Hirzel; trans. and ed. B. Bosanquet, *Metaphysics in Three Books: Ontology, Cosmology and Psychology*, Oxford: Clarendon Press, 1884; 2nd edn, 1887.(The second volume of Lotze's incompleted *System der Philosophie*.)

Lotze, R.H. (1885-91) *Kleine Schriften (Lesser writings)*, ed. D. Peipers, Leipzig: S. Hirzel. (A posthumous collection including, *inter alia*, short essays and book reviews.)

Lotze, R.H. (1989) *Kleine Schriften zur Psychologie (Lesser writings on psychology)*, ed. R. Pester, Berlin: Springer.(A recent collection of original source texts with additional material and commentary.)

References and further reading

Besoli, S. (1992) *Il valore della verità (The value of truth)*, Florence: Ponte Alle Grazie.(A detailed examination of Lotze's logic and of his epistemology.)

Morgan, M.J. (1977) *Molyneux's Question: Vision, Touch, And The Philosophy Of Perception*, Cambridge and New York: Cambridge University Press.(An idiosyncratic work containing an extended discussion of Lotze's 'theory of local signs'.)

Orth, E.W. (1986) 'Rudolf Hermann Lotze: Das Ganze unseres Welt-und Selbstverständnisses' (Rudolf Hermann Lotze: the Systematic Understanding of Self And World), in *Grundprobleme der grossen Philosophen. Philosophie der Neuzeit IV (Basic Problems of the Great Philosophers. Modern Philosophy IV)*, ed. J. Speck, Göttingen: Vandenhoeck & Ruprecht.(An important survey article which divides Lotze's work into three phases of development and contains detailed biographical information.)

Santayana, G. (1971) *Lotze's System of Philosophy*, ed. P.G. Kuntz, Bloomington, IN, and London: Indiana University Press.(Santayana's 1889 doctoral dissertation at Harvard, underJ. Royce, transcribed with an extensive introduction and exhaustive bibliography supplied by the editor.)

Schnädelbach, H. (1984) *Philosophy in Germany, 1831-1933*, trans. E. Matthews Cambridge: Cambridge University Press.(An important general reference work, which also contains a chapter largely devoted to Lotze's theory of value.)

Wagner, G. (1987) *Geltung und normativer Zwang (Validity and normative force)*, Freiburg and Munich: Verlag Karl Alber.(A study of the development of Neo-Kantianism which also situates Lotze's metaphysics in its historical context.)

Willard, D. (1984) *Logic and the Objectivity of Knowledge*, Athens, OH: Ohio University Press.(While devoted to the early history of Husserl and phenomenology, this work also contains many references toLotze and his influence.)

Woodward, W.R. (1999) *From Mechanism to Value. Hermann Lotze: Physician, Philosopher, Psychologist 1817-1881*, Cambridge and New York: Cambridge University Press.(This book promises to be the definitive study of Lotze in English to date.)

Love

Love is usually understood to be a powerful emotion involving an intense attachment to an object and a high evaluation of it. On some understandings, however, love does not involve emotion at all, but only an active interest in the wellbeing of the object. On other accounts, love is essentially a relationship involving mutuality and reciprocity, rather than an emotion. Moreover, there are many varieties of love, including erotic/romantic love, friendly love, and love of humanity. Different cultures also recognize different types of love. Love has, as well, a complicated archaeology: because it has strong links with early experiences of attachment, it can exist in the personality at different levels of depth and articulateness, posing special problems for self-knowledge. It is mistake to try to give too unified an account of such a complex set of phenomena.

Love has been understood by many philosophers to be a source of great richness and energy in human life. But even those who praise its contribution have seen it as a potential threat to virtuous living. Philosophers in the Western tradition have therefore been preoccupied with proposing accounts of the reform or 'ascent' of love, in order to demonstrate that there are ways of retaining the energy and beauty of this passion while removing its bad consequences.

1 Love: emotion, relationship, action

Love is frequently understood to be a powerful emotion. It seems to involve both intense attachment to an object and a high appraisal of the value of the object. Often, though not always, the object is seen as something one needs in one's own life; for this reason love is often connected with projects of possession or incorporation, and with jealous emotions towards the object seen as separate and capable of frustrating the lover's needs. Spinoza (1677) argued that love involves an awareness of the object as promoting one's own wellbeing (see Spinoza, B. de §9). Since all particular objects are, in virtue of their separateness from the self, also capable of frustrating wellbeing, all love, he concluded, is essentially ambivalent, mixed with anger and even hate. One may, however, hold that love is an emotion or emotions while insisting that these emotions can be free from jealousy and possessive desire. Thus Plato, in the *Phaedrus*, imagines love as a powerful response to beauty and value that is linked closely, in good people, to reverence and awe; it thus respects the separateness of the object and seeks its good. These accounts describe different experiences, both of which can be real (as Plato, unlike Spinoza, recognized).

Love is not just an emotion: it can also be a type of relationship. Aristotle, in *Nicomachean Ethics*, insisted that (friendly) love always involves mutual awareness and mutual good will (see Friendship; Aristotle §25). While any account of love needs to make room for loves that are unrequited, or directed at objects which cannot reciprocate (infants, some animals) or do not do so clearly (God), Aristotle's insistence on interaction and mutuality provides an important ingredient for a normative account of many types of human love, both friendly and romantic/erotic. Indeed, the refusal to conceive of love in relational terms is a central defect in many instances of erotic love, in which the loved object is indeed treated as an object, to be possessed and immobilized. Although Proust believed that such projects were essential to erotic love, one might doubt this.

Some loves may not involve strong emotion at all. Kant (1797) insisted that 'pathological love' (love involving passive emotion) was inferior to 'practical love', an active attachment to the good of others, including emotions of respect and concern. Whether or not we agree, we should acknowledge that this active practical commitment is one type of love: the love of humanity, for example, may be best understood this way. The Greek Stoics believed that even erotic love could be reimagined in a way that rendered it compatible with the *apatheia*, emotionlessness, proper to the wise (see Stoicism §19). It would be an active enthusiasm about the good of the object, without the currents of agonizing passivity that usually characterize erotic attachment.

2 Types of love

English, like Latin, has only a single term for a large family of different experiences. Other languages, such as ancient Greek and modern Japanese, disambiguate the varieties from the start by the use of different terms. But even in English and Latin, we may distinguish different species of love. Erotic/romantic love is linked closely to sexual desire, while friendly love apparently is not. The love of parents for children and children for parents is frequently thought in the modern era to have an erotic dimension; but this was not the view of most earlier

cultures, nor is it likely to be true in cultures where well-off parents rarely saw their children. Ancient Greek culture thought of *eros* as sexual, concerned with possession, and potentially destructive; *philia*, which could obtain between either friends or relations, was seen as mutual and reciprocal, concerned with benefit, and a positive cultural force. Christian *agapē* is distinguished from both of these loves by its essentially selfless character; its paradigm is Christ's gift of his life for the redemption of sinful humanity (see Charity).

We may also categorize loves by their object-type. We love other people, and it is reasonable to expect that these loves will involve some reciprocity and mutuality. People's loves of animals may be very intense; they vary greatly in the type of reciprocity they offer. People also intensely love inanimate objects such as works of art and natural beauty. Such loves cannot be reciprocal. Love may also take as its object a moral abstraction, such as social justice, or the good of humanity. This type of love is especially well explained on the Stoic/Kantian model, as involving active commitment more than emotion.

The love of God or gods has been understood in many different ways. The Stoics thought that loving God was loving the rational purpose that animates the universe; such a love was best understood as a form of active thought, without any emotional receptivity. Spinoza's *amor intellectualis dei* follows this paradigm. Augustine, criticizing Stoic *apatheia*, insisted that a strongly emotional form of love, mingled with fear, guilt and mourning, is most appropriate to a Christian life (see Morality and emotions §5). Many Christian thinkers follow his lead. Jewish conceptions of the love of God tend to stress proper action, both ritual and ethical (see Ritual §§2-3). Modern religious thought continues these debates.

3 Cultural variation

Most societies contain many different types and conceptions of love. But cross-cultural variation also complicates analysis. Societies vary (a) in the behaviour they deem fitting in a relationship of love; thus modern American lovers behave publicly in ways that would have been unthinkable in nineteenth-century India. Variation is also present (b) in the norms societies teach concerning appropriate objects of love; thus fifth-century BC Athens taught young men that it was to be expected that they would have strong erotic desires for both males and females; many modern cultures do not convey this impression. Societies also vary (c) in their normative evaluations of different species of love itself - differing, for example, about whether erotic love is something noble or shameful, good or bad. All these variations can be expected to shape not only concepts but also the experience of love itself.

Most interestingly, societies also vary (d) in the precise taxonomy of types of love their language and form of life display and perpetuate. For example, ancient Greek *eros* is imagined as a terrible power that seizes the personality and causes it to fix on an object with overwhelming intensity. Its goal is taken to be the possession of the object. Medieval courtly love, by contrast, stresses the ideal purity and remoteness of its object and associates love with a tender and refined attention to that object. Those who nowadays have lost the beliefs and forms of life that grounded courtly love cannot experience exactly that passion.

Differences in taxonomy are often revealed and further shaped by differences in terminology: thus the fact that ancient Greeks distinguished *eros* from *philia* and Romans used only the single word *amor* probably shaped thought and experience to at least some extent, although Romans clearly did recognize different varieties of *amor*. (Similarly in the modern world, the fact that Japanese has several distinct words for what English calls 'love' probably reveals some real difference in experience, although these differences should not be overestimated.) In the modern world, understanding of cultural variation is complicated by inter-cultural contact and the translation of formative texts: thus the fact that Japanese *ai* is used to translate biblical *agapē* no doubt shapes the evolution of that concept as applied to experience.

4 Love and human development

People begin to have strong emotions before they can move or speak. A human infant's combination of cognitive maturity with physical helplessness gives rise to a complex and ambivalent emotional life, as it sees that the very objects it needs for comfort and survival are also separate and ungovernable. Spinoza's profound conjecture about the relationship between love and anger has by now received much clinical and experimental confirmation. One task of human development is to manage and even surmount this ambivalence, which will exist in many different forms in different lives, as love is powerfully shaped by the individual identity of the early objects of attachment.

The early experiences that shape the pattern of a person's loves are remembered imperfectly, if at all; even to put them into words is to change them. And yet it seems likely that they shadow one's later experiences. Proust plausibly suggested that when an adult embraces a lover, he or she is at the same time embracing the shadow of an earlier object. Thus Albertine is also the mother whose goodnight kiss the small boy so eagerly awaited. And yet it is difficult to grasp these aspects of oneself; even to the extent that one does, one alters the past by making it definite and articulate. So people's self-knowledge in love is likely to be very imperfect.

5 Love and human good: the ascent of love

Love is generally acknowledged to be a source of beauty and value in life. No philosopher, therefore, has proposed its complete removal. But it is also judged to involve various difficulties for a person aiming at a good and virtuous life. One concern is that love entails partiality: by focusing intensely on the value of a single object, one loses sight of the legitimate claims of other objects and goals (see Impartiality). A second concern is with excessive neediness: by allowing a single object to become central to one's life, lovers put themselves at the mercy of events that they cannot control, thus sacrificing their dignity and agency. Finally, partly because of this passivity, love is often linked with anger and revenge, whether against the loved object or a rival, or both. A society that wants to reduce anger and violence therefore may have reasons to discourage love.

Philosophers in the Western tradition have therefore been preoccupied with the project of constructing a reform or 'ascent' of love that would enable us to retain its mystery and beauty while purifying it of deforming excesses. For Diotima in Plato's *Symposium*, the ascent centrally involves the idea of an abstract object. Once one understands that the real object of one's love is not a body or even a whole person, but the beauty that is seated in that body or person, then one can begin a process of reform, compared to climbing a ladder, through which one eventually comes to love all the beauty in the universe, and, beyond that, to contemplate the deathless form of the beautiful itself in all its unity. In this way, lovers become invulnerable to life's vicissitudes: the object of their love will never betray or disappoint them.

Christian proponents of a 'ladder' of love tend to criticize the Platonist programme for its goal of personal self-sufficiency. Proper humility requires that one retain a continual awareness of one's incompleteness and neediness. Christian writers also attempt to retain the love of particular individuals as part of purified love.

Spinoza returned to a Platonic proposal for the contemplative reform of love: by focusing on the mind's freedom from external contingencies, one ultimately comes to love the deterministic structure of the entire universe, and one's mind is freed from the passivity and ambivalence that characterize human attachments.

A remarkable modern interpretation of the Platonic tradition can be found in Proust's *À la recherche du temps perdu (Remembrance of Things Past)* (1914-27), which asserts that each of a writer's loves is like a step in a ladder that leads him upwards to general forms, in which, alone, his intellect finds comfort and delight. By using one's past experiences of pain and vulnerability as raw material for a creative work of art, one surmounts vulnerability and achieves a kind of freedom from time and death.

None of these reformers likes real human beings very much. This tradition therefore gives rise to a counter-tradition that tries to restore human beings to a greater acceptance of their loves as they are, seeing the interest in ascent as itself a disease that needs curing. Much of this tradition lies outside philosophy. A salient example is Joyce's *Ulysses* (1922), which playfully turns Diotima's ladder upside down, indicating that it is only in the inconstant and flawed emotion of daily life that real love is to be found. By linking religious idealism with anti-Semitism and Bloom's love of the body with a general love of humanity, Joyce indicates, too, that the ascent tradition may be a cause of social hatreds, rather than their cure.

See also: Abravanel, I.; Emotions, nature of; Emotions, philosophy of; Family, ethics and the; Sexuality, philosophy of

MARTHA C. NUSSBAUM

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985, bks VIII, IX.(On friendly love.)

Augustine (397-401) *Confessionum libri tredecim (Confessions)*, trans. F.J. Sheed, Indianapolis, IN: Hackett

Publishing Company, 1993.(Argues that highly emotional love is most appropriate to a Christian life.)

Augustine (413-27) *De civitate Dei (The City of God)*, trans. P. Levine, Loeb Classical Library, Cambridge, MA: Harvard University Press, 1966.(Discusses the emotions proper to a Christian life.)

Bowlby, J. (1982) *Attachment and Loss*, New York: Basic Books, 3 vols, 2nd edn.(Important psychological study of development of love in children.)

Cavell, S. (1969) 'The Avoidance of Love: a Reading of *King Lear*', in *Must We Mean What We Say?*, Cambridge: Cambridge University Press; repr. 1976.(Influential discussion of people's reasons for avoiding love.)

Dante (1313-21) *Divina Commedia*, trans. J. Ciardi, *The Divine Comedy*, New York: E.P. Dutton, 3 vols, 1989. (The most influential medieval Christian account of love, combining theoretical views of Aquinas with insights drawn from the tradition of courtly love.)

Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978.(Important discussions of love and hate.)

Joyce, J. (1922) *Ulysses*, New York: Modern Library, 1961.(Suggests that real love is to be found only in the flawed emotion of daily life.)

Kant, I. (1797) *Metaphysische Anfangsgründe der Tugendlehre*, trans. J.W. Ellington, *Metaphysical Principles of Virtue*, Indianapolis, IN: Hackett Publishing Company, 1964.(Discussion of the relationship between love as passion and love as active commitment.)

Klein, M. (1921-45) *Love, Guilt, and Reparation and Other Works, 1921-45*, London: Tavistock, 1985.(Important psychoanalytic treatment of love, envy, guilt.)

Murdoch, I. (1993) *Metaphysics as a Guide to Morals*, New York: Allen Lane, The Penguin Press. (Novelist-philosopher discusses the relationship between love and a vision of the good.)

Nussbaum, M. (1995) '*Eros* and the Wise: The Stoic Response to a Cultural Dilemma', *Oxford Studies in Ancient Philosophy* 13: 231-67.(Discusses the Stoic project of having erotic love without need and vulnerability.)

Plato (*c*.386-380 BC) *Symposium*, trans. A. Nehamas and P. Woodruff, Indianapolis, IN: Hackett Publishing Company, 1989.(Proposes the ascent of love to contemplation of ideal beauty.)

Plato (*c*.366-360 BC) *Phaedrus*, trans. A. Nehamas and P. Woodruff, Indianapolis, IN: Hackett Publishing Company, 1995.(An account of erotic passion mingled with reverence and awe.)

Price, A. (1989) *Love and Friendship in Plato and Aristotle*, Oxford: Clarendon Press.(Excellent treatment of the texts, with subtle insights about the topic.)

Proust, M. (1914-27) *À la recherche du temps perdu*, trans. C.K. Scott Moncrieff and T. Kilmartin, *Remembrance of Things Past*, New York: Random House, Vintage, 1981.(Great philosophical novel analysing love's relation to need, jealousy and artistic creativity.)

Spinoza, B. (1677) *Ethica Ordine Geometrico Demonstrata (Ethics Demonstrated in a Geometrical Manner)*, trans. E. Curley, *Ethics*, Harmondsworth: Penguin, 1996.(Analyses the ambivalent character of human emotion and proposes a process by which we may be freed from bondage to emotion, in favour of the intellectual love of God.)

Vlastos, G. (1973) 'The Individual as Object of Love in Plato's Dialogues', in *Platonic Studies*, Princeton, NJ: Princeton University Press.(Important analysis and criticism of ancient views of love.)

Löwenheim-Skolem theorems and nonstandard models

Sometimes we specify a structure by giving a description and counting anything that satisfies the description as just another model of it. But at other times we start from a conception we try to articulate, and then our articulation may fail to pin down what we had in mind. Sets seem to have had such a fate. For millennia sets lay fallow in logic, but when cultivated by mathematics in the nineteenth century, they seemed to bear both a foundation and a theory of the infinite. The paradoxes of set theory seemed to threaten this promise. With an eye to proving freedom from paradox, versions of set theory were articulated rigorously. But around 1920, Löwenheim and Skolem proved that no such formalized set theory can come out true only in the hugely infinite world it seemed to reveal, for if it is true in such a world, it will also be true in a world of the smallest infinite size. (Versions of this remain true even if we augment the standard expressive devices used to formalize set theory.) But then, Skolem inferred, we cannot articulate sets determinately enough for them to constitute a firm foundation for mathematics.

1 Background

To see how a model could fail to be standard, one first needs some idea of what a model is supposed to be. The basic distinction here is that between words and the world, signs and things. If the extension (denotation, reference) of each sign in a sentence or collection of sentences is held utterly fixed, and no re-interpretation is allowed at all, then what that sentence or collection commits us to should also be fixed. In that case there is little point in trying to distinguish various ways those sentences might come out true, that is, various models for them. If, in contrast, the extensions of none of the signs are in any way constrained, then just about anything could be gerrymandered into a way of making those sentences come out true. In this case too it seems unlikely that there will be interesting distinctions among the models. So the interest in models for a collection of sentences depends on holding the extensions for some but not all of the signs in those sentences fixed. (See Model theory.)

In modern algebra, a group is defined as a non-empty set *G* and a function $\circ$ such that:

(a) whenever *a*, *b* are in *G*, so is $a \circ b$;
(b) whenever a, b, c are in *G*, then $a \circ (b \circ c) = (a \circ b) \circ c$;
(c) there is a member *e* of *G* such that $a \circ e = e \circ a = a$ for each *a* in *G*; and
(d) for each *a* in *G* there is a member a^{-1} of *G* such that $a \circ a^{-1} = a^{-1} \circ a = e$.

The integers under addition constitute a group. The one-to-one correspondences between a set and itself also form a group under composition of those correspondences. Note that any model for the axioms is a group; there is no question of any groups being somehow standard, and others not. (On the other hand, because addition of integers commutes, that is, $a + b = b + a$, while composition of one-to-one correspondences need not, composition of correspondences illustrates groups in general better than addition of integers.)

If we put the second axiom of groups into the notation of quantification theory, we get

$$(\forall a, b, c)(a, b, c \in G \to a \circ (b \circ c) = (a \circ b) \circ c).$$

Any choice of a non-empty set for *G* and a binary function on *G* for $\circ$ is a candidate for a structure or interpretation making this axiom true. In this way, the extensions of the signs ‘*G*’ and ‘$\circ$’ are allowed a huge range of variation. But those of ‘$\forall$’, ‘$\in$’, ‘$\to$’ and ‘$=$’ are held fixed; they must be taken as universal quantification, membership, the conditional, and identity. This mixture of variation and constancy is part of what gives groups an interesting structure.

Modern algebra was a nineteenth-century development, and groups are an artefact of that growth. There was no favoured or intended group to begin with; it is instead the general group structure that proved to be of interest. So there is no distinction between standard and nonstandard models of the group axioms.

We think of groups as *given* by their axioms; they have no life prior to their characterization that would distinguish intended models of that description from unintended ones. But sets probably began life among us as extensions of predicates. On the one side there are predicates like ‘is a lion’. On the other, there are the lions. One way words and things link up is that all and only the lions are members of the extension of the predicate ‘is a lion’. We take the extension of a predicate as the collection, set or class of all and only the things of which that predicate is true.

Disguised as extensions sets led fairly inactive, boring lives in logic for ages.

But around the turn of the nineteenth century, reliance on geometric intuition led analysis, the theory growing out of differential and integral calculus, into paradox; intuition is not always trustworthy on infinitary processes. So a movement arose in nineteenth-century mathematics to replace geometry with arithmetic. We had learned in analytic geometry to individuate a point in the plane by an ordered pair of numbers, the point's abscissa and ordinate. Now we could rethink a literal curve passing (for example, continuously) through such points as some sort of arithmetical law, or even equation, relating the abscissas of such points to their ordinates. This rethinking of analysis is called its arithmetization.

The arithmetization of analysis focused attention on the kind of numbers used as coordinates of points on the plane. Nowadays we call these real numbers; they are the numbers used for lengths, temperatures and many other natural quantities, and given by decimals. Arithmetizing analysis calls for a clear view of the reals and their laws. To those with a taste for abstraction or essence, it even raises the question of what a real number is. Later on in the nineteenth century, Richard Dedekind showed how to answer this question by taking reals as (roughly) certain sets of rational numbers. (One can think of rationals as quotients of integers in lowest terms.) The details to one side, the idea is to rethink reals as certain constructions out of rationals using sets to do the constructing. Such constructions can be shown to satisfy the laws of reals needed to do analysis (see Dedekind, J.W.R.).

At that point we seem to be on to something. Reals are constructed from rationals using sets. It is a lot easier to construct rationals from integers (positive, negative and zero) using sets and to construct integers from natural numbers (non-negative whole numbers) using sets. Then Frege and Cantor seem independently to have got the idea of constructing natural numbers out of sets alone (at least that is how we in hindsight might put it). Sets seem to be becoming the basic stuff of mathematics, so if these sets are the same old extensions of logic, then perhaps, as Frege claimed, mathematics is reducible to logic (see Cantor, G.; Frege, G.).

The basic idea was to begin with the as-many-as relation. One set *A* has as many members as another set *B* precisely in case there is a binary relation *R* such that (1) for each $a \in A$ there is exactly one $b \in B$ such that *aRb*; (2) *R* matches different members of *A* with different members of *B*; and (3) for each $b \in B$ there is an $a \in A$ such that *aRb*. Beyond set-membership, only notions of logic (all, some, identity, and so on) are needed for this exposition of as-many-as. But with it, Cantor and Frege could take the number N_A of members of a set *A* as the set of all sets *B* with as many members as *A*, and a number as the number of members of some set (see Logicism).

But there is more. Mathematics had always looked to be about things like numbers, of which there are infinitely many, and triangles, of which there are infinitely many and in each of which there are infinitely many points. But mathematics lacked a good general account of the infinite *per se*. Dedekind provided one; a set is infinite precisely in case it has exactly as many members as one of its proper subsets (that is, a set included in but not exhausting the original set). Dedekind's account might have been only a curiosity had Cantor not proved that for each set, even infinite, there is a strictly larger set. At a single stroke, Cantor revealed a wealth of different infinite sizes, and thus a realm of transfinite arithmetic, which Hilbert called Cantor's paradise. Set theory looked to promise both a foundation for mathematics and a whole new mathematics of the infinite (see Set theory; Cantor's theorem; Continuum hypothesis).

But there were serpents in paradise. Most philosophers have seen Russell's paradox. Cantor's starts from a predicate like 'is a thing' which is true of everything. Every set would have to be a subset of *U*, the extension of this predicate, and so be no larger than *U*. But by Cantor's result cited in the previous paragraph, there is a set strictly larger than *U*. This contradiction together with Russell's and a third due to Burali-Forti constitute the paradoxes of set theory, and around the turn of the twentieth century some regarded these paradoxes as a crisis in the foundation of mathematics (see Paradoxes of set and property).

One response to this crisis was to formulate the mathematics at issue with unprecedented rigour. We would be utterly explicit about exactly which signs (that is, marks) may be written, how these marks may be concatenated, and how such strings may be manipulated. This idea of a formal system had been implicit in Frege, but it acquires a new prominence in Hilbert's programme for responding to the crisis in the foundation of mathematics. For once we have such a formal system we can ask whether anything with the shape of a contradiction can be cranked out by its rules of manipulation. Here we might prescind from all questions about how these marks might be

interpreted and focus instead only on which finite combinations of them the rules for their manipulation permit. This latter inquiry sounds like very elementary mathematics indeed. Hilbert's programme might look to show how to use only the most concrete and trustworthy mathematics, combinatorics, to prove the consistency of, and so to that extent legitimate trust in, the most basic (and so most powerful) set theory, at least as formalized (see Formal languages and systems; Hilbert's programme and formalism).

But now we have come back to a distinction between words and the world. On the one side, in the world, we have the sets. These are not just any old sets; they are to be a foundation for all of mathematics and to embody a new mathematics of the infinite. We have expectations of these sets; there are, for example, to be enormously many sets, many more than there are natural numbers or even real numbers. So much was promised by Cantor's theorem long before set theory got formalized, or even axiomatized. On the other side, in language, we have formalized set theory. To formalize it means to give its primitive marks and the rules for their manipulation just in terms of their shapes without a thought for how these marks might be assigned things in the world as extensions or interpretations. We want that purely formal presentation in hopes of as concrete and combinatorial a consistency proof for the system as possible. One might have other aims in formalizing, but Hilbert wanted a consistency proof as a *laissez-passer* back into Cantor's paradise. But having made this distinction between words and the world, one then can see the possibility of the words having an interpretation in the world quite different from what was intended.

2 One proof of the Löwenheim-Skolem theorem

In 1915 Leopold Löwenheim, a secondary school teacher, published a claim that a formula in the calculus of relatives that comes out true under an interpretation in an infinite domain also comes out true under an interpretation in a domain of the smallest infinite size. This size is that of the natural numbers, and domains of that size are called countably infinite. We might now describe the formulas in which Löwenheim was interested as those in which only the truth functional connectives and the quantifiers are interpreted uniformly across domains, that is, non-empty sets; the other signs, that is, constants, function letters and predicate letters, may be interpreted by any member of, or function or relation (of suitable polyadicity) on, any such domain. The consensus nowadays is that Löwenheim failed to prove his claim.

But after the First World War, the Norwegian logician Thoralf Skolem gave two rather different but equally correct proofs of Löwenheim's claim; this is the origin of the Löwenheim-Skolem theorem. Nowadays, one of these arguments is usually presented with a proof of the completeness of quantification theory (first-order logic). To show completeness, we start from a valid formula, that is, one true under all interpretations in all non-empty domains, and we must show that it has a proof by the rules of quantification theory. Equivalently by contraposition, we must show that a formula without such a proof fails in some interpretation. For this it suffices to show that any formula that yields no explicit contradiction by the rules is true in some interpretation. We can in fact show that the domain of such an interpretation need be no larger than countably infinite. But any formula true in some hugely infinite domain cannot yield a contradiction by the rules, and so by our result is true in an at most countably infinite domain.

Skolem not only proved Löwenheim's claim: he also strengthened it. For example, the claim holds not just for single formulas, but for countably infinite sets of formulas. Now remember our precisely formalized set theory. As standardly presented, this can be taken as a countably infinite set of sentences in which, over and above logic (which may be taken to include identity), there is only a single binary predicate (intended for the membership relation). Replace this binary predicate by a dyadic predicate letter allowed to be interpreted by any two-place relation on any non-empty domain. We get a set of formulas to which Skolem's strengthening of Löwenheim's result applies. Call this set of formulas the logical skeleton of set theory. It has lost its meat, the real extension of the predicate originally intended for the membership relation. It can be taken as the set of formulas that would be true were the predicate interpreted as intended, not just the theorems of some axiomatization of set theory. If set theory as standardly presented is, interpreted as intended, true, then because of Cantor's result on different infinite sizes, this set of formulas is true in a hugely infinite domain, namely, the world; and so by the Löwenheim-Skolem theorem, it is also true in an at most countably infinite domain. Indeed, since a countable model always has an isomorphic copy in the natural numbers, that countable model can be taken as one on the natural numbers. Hence the logical skeleton of set theory cannot come out true only when it is interpreted as intended in set theory. One

might begin to worry whether the laws of sets pin them down determinately enough for them to be a firm foundation for mathematics.

But raising this worry, as we have, with a spotlight on the logical skeleton of set theory suggests a way to assuage it. Surely parallel evolution could result in two very different animals whose skeletons are none the less indistinguishable. If so, an animal's skeleton need not individuate it. Similarly, the wealth of sets comes not from the logical form of (first-order) truths about membership, but rather from the nature of membership. For by the Löwenheim-Skolem theorem, the logical forms of the first-order truths of set theory come out true in a world vastly smaller than the real world of sets. So those forms alone fail to pin down the full nature of the membership relation, which is what makes the real world so big. That nature cannot be isolated just by the logical form of first order truths about it. Note in this connection that the proof of the Löwenheim-Skolem theorem sketched above makes no use of the huge model it supposes except to infer simple consistency; it constructs the small model from the merely consistent theory alone. Perhaps it is not surprising that with the meat thrown away, bones no longer carry its tang.

3 Another proof of the theorem and Skolem's paradox

But Skolem also proved the theorem in another way that makes much more intimate use of the huge original model. To get the gist, we need the idea of a Skolem function, which is another way of taking a quantifier in a model. A formula, perhaps with free variables, is in prenex form when all its truth functional connectives lie in the scopes of all its quantifiers. There are mechanical routines for finding prenex equivalents of formulas. Take prenex forms for each of the formulas in the logical skeleton *T* of set theory. Let these formulas be interpreted as intended in a huge domain *D* of sets on which *E* is the membership relation, the intended extension of the binary predicate letter that is the only extra-logical sign in *T*. (There may be a problem here. For model theory to be a theory, the domains of models should be objects we know how to handle. For that reason, they are usually required to be sets. But the standard response to paradoxes like Cantor's is nowadays to deny that there is a set of all sets. So if *D* is to be treated by the laws of all sets (as it is), then it had better not exhaust the sets, in which case *E* cannot be *the* membership relation.) Next we glance at orderings of *D*. A linear ordering of *D* is a two-place relation that arranges its elements as in a straight line; picture the relation of being-to-the-right-of. A well-ordering of *D* is a linear ordering *R* of *D* such that every non-empty subset of *D* has an *R*-least member, which then by linearity is unique. Being-to-the-right-of is not a well-ordering of the points on a line. But an equivalent of the axiom of choice of set theory says there is a well-ordering *R* of *D* (see Axiom of choice). Pick an arbitrary member *d* of *D* for don't-care cases. Let

$(\forall x)(\exists y)M(x,y,z)$

be one of our prenex forms, where *M* is a quantifier-free truth functional compound in which only *x*, *y* and *z* are variables. Skolem functions come from existential quantifiers in such prenexes. The existential quantifier binding *y* lies in the scope of one universal quantifier binding *x* and in a formula in which *z* occurs free. So the Skolem function for this quantifier is a two-place function *f* of the arguments *x* and *z*. For any members *x* and *z* of *D*, the value of *f* at *x* and *z* is the *R*-least member *a* of *D* such that $M(x,a,z)$ if there is any such *a* at all, but *d* otherwise (the don't-care case). Skolem functions are a little more complex in general, but that is the basic idea.

Take one prenex form for each formula in *T*. That is a countable infinity of forms, each of which has only finitely many existential quantifiers in its prenex. So there is only a countable infinity of Skolem functions. Pick any countable subset *A* of *D*, but toss *d* into it. Let A_0 be *A*, and for each natural number *n*, let A_{n+1} be A_n together with all the values of all the Skolem functions on members of A_n. We can show that for each *n*, A_n is countable, and thus that $D' = A_0 \cup A_1 \cup \ldots$ is also countable. D' is also closed under all Skolem functions; that is, if *f* is a *k*-adic Skolem function and $d_1, \ldots, d_k \in D'$, then $f(d_1, \ldots, d_k)$ is also in D'. This last is crucial; it comes from the fact that any prenex of any formula is finitely long, which comes from the fact that formulas are of finite size. Now let E' be *E* cut down to D'; that is, $aE'b$ if and only if aEb, and a, b are both in D'. The relation E' on the domain D' constitutes our new small model of *T*; it is called the Skolem hull of our huge old model.

A definition of satisfaction converts the extension E' of *T*'s only extra-logical sign into extensions in D' for each of its formulas. (See Tarski's definition of truth.) It is only for members of D' that there is much point in asking whether they do or do not belong to these extensions. Satisfaction also converts *E* into extensions in *D* for

formulas of *T*, and we can compare how these extensions treat objects from the overlap of their domains, namely D'. Let *F* be a formula of *T* in which the variables $x_1, \ldots, x_k$ occur free and let $d_1, \ldots, d_k$ be members of D'. Then we can show that $d_1, \ldots, d_k$ satisfy *F* as interpreted by *E* in *D* if and only if they satisfy it as interpreted by E' in D'. The two interpretations agree where both make sense. This includes all sentences of *T*, so the same sentences of *T* are true in both. All the truths of set theory were to hold in the big interpretation, so they hold in the smaller one, even the formalization of Cantor's theorem.

This time around E' has been carved directly out of *E*. All that has been pared away is whatever in *D* was unnecessary to make the formulas of *T* true when interpreted by *E*. But that turns out to be almost all of *D*. In particular, although a formula of *T* was intended under *E* to say that there are more than countably many things (in *D*), and though E' has enough of *E* for both to satisfy the same formulas in D', that formula intended to assert uncountability is true in D' even though D' is countable.

In 1922 Skolem wrote:

> *axiomatizing set theory leads to a relativity of set theoretic notions, and this relativity is inseparably bound up with every thoroughgoing axiomatization*….In order to obtain something absolutely non-denumerable, we would have to have either an absolutely non-denumerably infinite number of axioms or an axiom that could yield an absolutely non-denumerable number of first-order propositions. But this would in all cases lead to a circular introduction of the higher infinites; that is, on an axiomatic basis higher infinites exist only in a relative sense.
>
> (Skolem 1922)

These words are the origin of what has come to be called Skolem's paradox. It is pretty cut and dried what Russell's paradox is, but not Skolem's. Just what distinction between absoluteness and relativity did Skolem intend? Was it to be the same for notions as for existence? How does it line up with the countable/uncountable distinction? Do his results rule axiomatic set theory out as a foundation for mathematics?

These questions might lead us in many directions, but we have space only to gesture in one. When, then, might formulas capture, express or pin down a notion? One fairly evident idea is that they do so just in case all and only the models for those formulas exemplify the notion. In this way, the axioms for a group pin down the notion of a group because the groups are, not surprisingly, the models for those axioms. Perhaps formulas catch a notion absolutely if all and only the models for those formulas exemplify the notion. Then we note that uncountability is exemplified by some but not all models for *T*, so *T* does not capture uncountability absolutely but only, Skolem may have meant, at best relatively. (Maybe, but now there seem to be many things he might have meant.) Any first-order theory with uncountable models also has countable models, so no first-order theory guarantees that its models are countable, and thus satisfaction of a first-order theory does not guarantee the existence of an uncountable infinity. But satisfaction of the axioms of a group does guarantee the existence of a group. Some such contrast may be what Skolem meant by calling the existence of uncountable infinities relative.

4 One try at a solution to the paradox

Skolem seems to have had it in for set theory. But maybe the Löwenheim-Skolem theorem reveals not so much flaws in axiomatic set theory as limitations in the expressive powers of first-order logical notation. Skolem's words above hint that only a somehow question begging uncountable infinity of notations could guarantee uncountability in models of those notations. The circle is too faint. Is it vicious to use ink to describe ink, or to describe finitude in finitely many words? The hint seems suspect. Add to the notation a new quantifier '*Ux*' such that '$(Ux)Fx$' is true only in a domain with uncountably many members that satisfy *F*, and count as models only interpretations that verify '$(Ux)(x = x)$'. Would limiting ourselves to such models be any more circular than limiting ourselves to models without non-self-identical members?

Second-order quantification is quantification into the positions of predicates, rather than singular terms, in sentences. (See Second-order logic, philosophical issues in.) Such quantification might seem less obtrusive than uncountable individual quantification, but it also lets us say that there are uncountably many things, though framing the claim calls for a bit more ingenuity. But if the domains of models for formulas with quantifiers of either sort must be sets, then discourse with either sort of quantifier is subject to generalized versions of

Löwenheim-Skolem. Formal languages are usually distinguished by choices among countable infinities of individual constants, predicate letters and function signs. If so, there are only as many such languages as there are reals, and though there are lots of reals, they make up a set. So there are only that many theories, and thus satisfiable theories, too. To each such satisfiable theory assign the least cardinal that is the size of the domain of a model of the theory; all sets have cardinals, so if domains must be sets, this assignment works. Then (by the axiom of replacement) there is a set of all these least cardinals, and so a smallest cardinal greater than all of them. Whether we allow uncountable individual or second-order quantification, this cardinal is to those theories as the cardinal next after that of the natural numbers is to first-order theories according to the Löwenheim-Skolem theorem. For each is a size smaller than that of the intended world of set theory, yet each is a size larger than that of a world in which set theory comes out true. So long as the domains of models must be sets and languages distinguished by choices from a set of primitives, some such result seems inevitable whatever expressive devices we allow our languages; but if the domains need not be sets, it is unclear what they might be and how we might handle them.

See also: Logical and mathematical terms, glossary of

W.D. HART

References and further reading

Chang, C.C. and Keisler, H.J. (1973) *Model Theory*, Amsterdam: North Holland, 3rd edn, 1990, 141-2.(The authors present a quick but contemporary version of Skolem functions and hulls.)

Fraassen, B.C. van (1971) *Formal Semantics and Logic*, New York: Macmillan, 122-5.(The author presents a tidy exposition of Skolem hulls.)

Hart, W.D. (1970) ‘Skolem’s Promises and Paradoxes’, *Journal of Philosophy* (1970): 98-109. (An exposition and discussion of the paradox from a Platonist perspective.)

Quine, W.V. (1976) ‘Ontological Reduction and the World of Numbers’, in *The Ways of Paradox and Other Essays*, revised edn, Cambridge, MA: Harvard University Press, esp. 212-20.(Unlike Skolem, Quine worries about whether the theorem shows that Pythagoras was right and only numbers exist.)

Skolem, T. (1920) ‘Logico-Combinatorial Investigations in the Satisfiability or Provability of Mathematical Propositions’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 252-63.(Skolem proves the version of the theorem using the axiom of choice.)

(1922) ‘Some Remarks on Axiomatized Set Theory’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 290-310.(Skolem proves the version of the theorem without the axiom of choice. This paper contains his original presentation of his paradox.)

Tarski, A. and Vaught, R. (1956) ‘Arithmetical Extensions of Relational Systems’, *Compositio Mathematica* (1956): 81-102.(The authors set out the basics of the contemporary understanding of how Skolem functions and hulls work.)

Lu Xiangshan (1139-93)

A leading Chinese philosopher of the twelfth century, Lu Xiangshan was the founder of that dimension of neo-Confucian thought known as the learning of the heart-and-mind. Lu emphasized the necessity for personal responsibility and action in everyday affairs, as opposed to the search for moral understanding through classical texts.

Lu Xiangshan (named Jiuyuan, styled Zijing and popularly called Xiangshan) was a leading thinker of the Song dynasty (960-1279) philosophical movement commonly called neo-Confucianism (see Neo-Confucian philosophy). He is considered the founder of that dimension of neo-Confucianism known as the learning of the heart-and-mind (see Xin) and was an important rival of Zhu Xi, the pre-eminent twelfth-century philosopher and advocate of the learning of principle. Lu came from a large, multi-generational, elite family in south China that engaged in a medicinal drug trade, farming, local defence, social welfare projects and Confucian education. He attained the highest degree, served briefly but vigorously in office, and was famous for his teaching. His few writings, recorded sayings and biography are collected in the *Xiangshan quanji (Complete Works of Lu Xiangshan)*.

Critical of contemporary scholars' preoccupation with the civil service examinations and speculative ontological thought concerning *li* (principle(s) or pattern(s) of things) and *qi* (matter-energy) (see Li; Qi), Lu focused on actual human experience. Reflecting the importance of the *Yijing (Book of Changes)* (see Yijing) and Mencius in his thought, his primary concerns were the principles ordering the world (human society and nature) and the necessity for assuming personal responsibility for moral-social action.

Lu maintained that there is only one principle in (and one ordering source of) the universe. Shared by the 'three ultimates' of heaven, earth and humanity, this principle orders all things and affairs. It is the Way, the heart-and-mind, and human nature, and (as *qi*) does not exist apart from the world. Lu summarized his position in such claims as 'the heart-and-mind is principle' and 'my heart-and-mind and the universe are one'. This principle is public-mindedness and moral rightness, in contrast to selfishness and private profit.

Lu stressed the necessity to take responsibility for one's actions and to pursue Confucian learning in order to develop and implement one's moral-social nature. Emphasizing inner, natural capacities over outer, textual knowledge, Lu taught people to distinguish between morality and selfishness and to establish their aim. Echoing Mencius, he called this effort 'establishing one's greater part' (that is, the heart-and-mind). This kind of knowing was completed by moral-social action. In addition, a person should engage in constant self-examination and self-correction, while seeking help from reading and teachers and friends.

Scholars disagree about the interpretation of Lu's thought. Although some describe it as monism and subjective idealism, these Western labels are highly misleading because Lu was not concerned with the kinds of philosophical questions they imply, such as fundamental substances of being, objective-subjective reality, mind as consciousness and the unreliability of sensory knowledge.

See also: Neo-Confucian philosophy; Self-cultivation in Chinese philosophy

ANNE D. BIRDWHISTELL

List of works

Lu Xiangshan (*c*.1150-90) *[Lu] Xiangshan [xiansheng] quanji (Complete Works of Master Lu Jiuyuan)*. (Collected writings in 34 chapters published in 1212, recorded sayings in 2 chapters published in 1237, complete works in 36 chapters pub. in 1521 and 1559. There are three modern 'editions', the Sibu beiyao edition, the Sibu congkan edition and the Guoxue jiben congshu edition.

References and further reading

de Bary, W.T., Chan Wing-tsit and Watson, B. (1960) *Sources of Chinese Tradition*, New York: Columbia University Press, vol. 1, ch. 18, 509-14.(Selected translations with an emphasis on universal mind and intuition, in contrast to Zhu Xi.)

Cady, L. (1939) 'The Philosophy of Lu Hsiang-shan, A Neo-Confucian Monistic Idealist', unpublished Ph.D.

dissertation, Union Theological Seminary.(Numerous translations, a comprehensive discussion of life, thought and contexts, with Lu viewed as a monistic idealist.)

Chan Wing-tsit (1963) *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press, ch. 33, 572-87.(Selected translations and brief comments; Lu is viewed as an idealist, in contrast to Zhu Xi.)

Chan Wing-tsit (1989) *Chu Hsi: New Studies*, Honolulu, HI: University of Hawaii Press, ch. 26, 435-61. (Discussion of Zhu and Lu's personal relationship.)

Chang, C. (1957) *The Development of Neo-Confucian Thought*, New York: Bookman Associates, vol. 1, ch. 13, 285-307; 2nd edn, 1962.(Summary of Lu's life and thought, selected translations and the debate with Zhu Xi; Lu is viewed here as an idealist.)

Ching, J. (1974) 'The Goose Lake Monastery Debate (1175)', *Journal of Chinese Philosophy* 1 (2): 161-78. (Translations and discussion of Zhu and Lu's meeting.)

Foster, R.W. (1987) 'Differentiating Rightness from Profit: The Life and Thought of Lu Jiuyuan (1139-1193)', unpublished Ph.D. dissertation, Harvard University.(Examination of Lu's ethical philosophy and its political application.)

Fung Yu-lan (1952-3) *A History of Chinese Philosophy*, trans. D. Bodde, Princeton, NJ: Princeton University Press, vol. 2, ch. 14, 572-92.(Discussion and selected translations, giving an idealist and subjective interpretation of Lu's thought.)

Huang, S. (1944) *Lu Hsiang-shan: A Twelfth Century Chinese Idealist Philosopher*, New Haven, CN: American Oriental Society.(General introduction to Lu's thought, context and influence; he is viewed here as a monistic idealist.)

Hymes, R. (1989) 'Lu Chiu-yuan, Academies, and the Problem of the Local Community', in W.T. de Bary and J.W. Chaffee (eds) *Neo-Confucian Education: The Formative Stage*, Berkeley, CA: University of California Press, ch. 16, 432-56. (Discussion of Lu in the context of social institutions.)

Kim, O.C. (1980) 'Chu Hsi and Lu Hsiang-shan: A Study of Philosophical Achievements and Controversy in Neo-Confucianism', unpublished Ph.D. dissertation, University of Iowa.(Emphasis on the this-worldly focus of Lu's thought and contexts, rejects idealist and monistic interpretations; good bibliography.)

Tillman, H.C. (1992) *Confucian Discourse and Chu Hsi's Ascendancy*, Honolulu, HI: University of Hawaii Press, 187-234.(Discussion of life and thought from historical perspective, rejecting interpretation of extreme philosophical subjectivism, and also describes Zhu and Lu's debates; good bibliography.)

Lucian (*c.* AD 120-80)

Lucian of Samosata (in ancient Syria) was one of the most original and engaging figures of post-classical Greek culture. He produced a diverse and influential corpus comparable in size to that of Plato (consisting of seventy-six authentic libelli). Formally the dialogue (in both Platonic and Cynic forms) dominates (thirty-six of seventy-three prose works), but there are also satiric narratives, tall tales (for example, A True Story), 'Cynic' diatribes (for example, On Mourning), and multifarious lectures, or essays (for example, The Master of Rhetoric) in his singular oeuvre.

Like other Greek writers in the Empire, Lucian is difficult to classify generically or ideologically. Who was this jester from the East, who dared ape the classics and yet succeeded in persuading posterity that he was, in the words of Macaulay, 'the last great master of Attic eloquence and Attic wit' (Macaulay)? A Hellenized Syrian of the Christian era who made Greek culture come to life as if were contemporary (Duncan 1979), Lucian was, none the less, not an absolute traditionalist: he parodied the classicizing literary fashions of Greeks in the Empire (such as 'Atticism' in *Lexiphanes*) and treated Greek ethnocentrism and the easy idealization of the classical past as targets ripe for satire (see for example, *Anacharsis, and Philosophers for Sale!*). Although long associated with philosophy, he was not a philosopher but a sophist, satirist and parodist, who, following Plato, the Cynics (for example, Menippus), and the Old Comic poets (for example, Aristophanes), took philosophers and their discourse as one of his principal subjects in a lengthy career as a writer and performer giving public readings of his work throughout the Roman Empire.

Lucian's work is philosophically significant in at least three senses. First, it is a source of ostensibly 'eye-witness' accounts of contemporary philosophical and religious figures, such as, the Cynic/Socratic Demonax of Cyprus, whom Lucian claimed as a teacher (in his *Life of Demonax*), the notorious Cynic Peregrinus of Parium and the 'false-prophet' Alexander of Abonoteichus, all of whom were extremely influential figures in their day and were, as such, mercilessly anatomized by Lucian in satiric narratives (*On the Death of Peregrinus*, *Alexander or the False-Prophet*) in which Democritus and Epicurus provide the philosophical armature. Second, he provides a general commentary on and evaluation of the reception of classical philosophical traditions in the polyglot cultural landscape of the Roman Empire. Lucian is suspicious of the inflated value of tradition(s) in a classicizing culture; accordingly, two works satirize the experience of 'conversion' to classical philosophical traditions (to Platonism in the *Nigrinus*, to Stoicism in the *Hermotimus* (compare *The Cynic*)). Similarly, the hilarious *Philosophers for Sale!*, in which the founding fathers of Greek philosophy are auctioned off as slaves, uses parody and caricature to satirize *all* the idols of the philosophers' tribe. (Lucian takes no prisoners: neither Diogenes the Cynic nor Pyrrho the Sceptic is spared satiric scrutiny, despite their own credentials as critics of philosophy.) The philosophical import of texts in these first two categories is more or less overt and explicit. Third, there are the philosophical implications of his practice as a seriocomic (*spoudogeloios*) writer of dialogue, satire and parody, which are generally Cynical or Sceptical in tendency. For example, in the *Anacharsis* two legendary sages (Solon the Athenian and Anacharsis the Scythian) engage in dialogue but, instead of generating philosophical insight, seem to speak past each other, unable to inhabit a culturally alien point of view. Similarly, in his *Symposium* Lucian constructs an ironic counter-image to the myth of philosophy as an ascent to timeless wisdom memorialized by Plato in his own *Symposium*.

While Lucian's parodic treatment of philosophers and their discourse makes him indispensable reading for any student of the history of philosophy, the value of good parody resists summary by its very form, as does a literary repertoire of over 170 characters. Lucian's own philosophical stance is consequently elusive and is never stated directly (even in the *Demonax*). If we consider his *oeuvre* as a whole, however, there are clear patterns of aversion and affinity. Lucian is attracted to Epicureanism for its rational hedonism and critique of religious ideology (see Epicureanism); to Cynics for their courageous practice of *parrhēsia* ('free speech') - the moral basis of satire - and their unflinchingly satiric perspective on the *ancien régime* (see Cynics); to Pyrrhonism for its 'suspension of judgment' on such questions as the value of philosophy as an activity or an institution (see Pyrrhonism). It makes sense, therefore, that Lucian has been of most interest to satirists (for example, Rabelais, Jonson, Swift, Voltaire), sceptics (for example, Hume) and highly literary thinkers of sceptical tendency (for example, Montaigne, Diderot) who do not fit easily into the history of philosophy. Lucian was taken up as a model of the witty, erudite cultural

critic by Renaissance Humanists and is historically significant for transmitting Cynic and Sceptic tradition to Europe (see Humanism, Renaissance §6). David Hume is said to have read Lucian’s comic *Dialogues of the Gods* on his deathbed. In an age as awash in parody, cynicism and religious entrepreneurs as it is alienated from its own traditions, Lucian is finding new audiences.

See also: Cynics §4

R. BRACHT BRANHAM

List of works

Lucian (*c.* AD 140-80) Works, trans. A.M. Harmon, K. Kilburn and M.D. Macleod, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1913-17, 8 vols; trans. H.W. Fowler and F.G. Fowler, Oxford: Clarendon Press, 1905, 4 vols. (The former includes parallel Greek text and English translation; the latter is an almost complete English translation by two masters of English prose.)

References and further reading

Branham, R.B. (1989) *Unruly Eloquence: Lucian and the Comedy of Traditions*, Cambridge, MA: Harvard University Press.(Analyses Lucian’s use of literary and philosophical traditions focusing on the idea of the seriocomic or *spoudogeloios*.)

Clay, D. (1992) ’Lucian of Samosata: Four Philosophical Lives (Nigrinus, Demonax, Peregrinus, Alexander Pseudomantis)’, in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin: de Gruyter, II 36: 5, 3406-50.(Analyses Lucian’s *Nigrinus*, *Demonax*, *Peregrinus*, and *Alexander*.)

Duncan, D. (1979) *Ben Jonson and the Lucianic Tradition*, Cambridge: Cambridge University Press.(Studies Lucian’s reception in Renaissance England.)

Jones, C.P. (1986) *Culture and Society in Lucian*, Cambridge, MA: Harvard University Press. (Places Lucian in his historical context.)

Niehues-Pröbsting, H. (1979) *Der Kynismus des Diogenes und der Begriff des Zynismus (The Cynicism of Diogenes and the [Modern] Concept of Cynicism)*, Munich: Wilhelm Fink.(The best account of the whole Cynic tradition (ancient and modern) in one volume; contains a brilliant analysis of Lucian’s *Peregrinus* and of the relation of ancient Cynicism to modern cynicism.)

Relihan, J.C. (1992) *Ancient Menippean Satire*, Baltimore, MD: Johns Hopkins University Press.(The best history of the genre; includes close readings of Lucian’s Menippean pieces.)

Robinson, C. (1979) *Lucian and his Influence in Europe*, Chapel Hill, NC: University of North Carolina Press. (Surveys Lucian’s reception in Europe.)

Sloterdijk, P. (1988) *Critique of Cynical Reason*, trans. M. Eldred, London: Verso.(An influential interpretation of Cynicism focusing on the relation of ancient Cynicism to modern cynicism; includes a discussion of Lucian’s *Peregrinus*.)

Lucretius (*c*.94-*c*.55 BC)

Titus Lucretius Carus was a Roman Epicurean philosopher and poet. About his life and personality little can be said with certainty, yet his only known work, 'On the Nature of Things' (De rerum natura), is of considerable size and one of the most brilliant achievements of Latin poetry. A didactic poem in six books, it expounds Epicurean physics. Its manifesto is to abolish the fear of gods and of death by demonstrating that the soul is mortal and the world not governed by gods but by mechanical laws.

The two main biographical sources - St Jerome, and the *Life of Virgil* attributed to Donatus - are unreliable and often inconsistent. The one near-contemporary reference to Lucretius is a letter of Cicero's written in 54 BC, revealing that he and his brother had read and admired the poem. All that we can say with confidence about Lucretius' life is that he lived in the earlier half of the first century BC. He was probably a member of the aristocratic family the Lucretii. Some take the cognomen Carus as testimony that he was a freedman or a slave. However, the dedication of his poem to one Memmius is strong evidence that he was, or aspired to be, a friend of C. Memmius, the wealthy patron of the poets Cinna and Catullus. Much more suspect is Jerome's story of Lucretius' madness caused by a love potion, and of his eventual suicide. Although now rightly rejected, the story was influential in Lucretian criticism, confirming for some the poet's profound pessimism.

The strongest evidence that at Lucretius' death the poem had not received its final revision is V 155, where he makes a promise, nowhere redeemed, of a full treatment of the nature of the gods. Whether other features, such as repeated lines and passages, are also evidence of its unfinished state is more disputed. But since Lucretius states at VI 92-5 that Book VI is to conclude the work, the poem cannot be substantially incomplete.

Lucretius writes as what the Romans called a 'learned poet'. He was a highly educated writer, whose sources of inspiration lay largely in Greek literature. The poem itself belongs to the tradition of didactic poetry, which stemmed from Hesiod's *Works and Days*, although a more direct model is the physical poem of Empedocles, *On Nature* (see Hesiod; Empedocles). He also shows debts to satire, epic and tragedy. Although complaining about the 'poverty' of the Latin language, Lucretius shows skill in adapting existing words to new concepts. His poetical expressions include archaism and periphrasis, but also formal expository devices based on Epicurean methods of argumentation.

The poem is divided into three pairs of books. Books I and II analyse the infinite universe into its microscopic components, atoms and void, refute rival theories of the elements, and show how atoms combine to produce the familiar phenomenal world. The next pair moves up to the human level. Book III covers the soul and its mortality, culminating in an eloquent denunciation of the fear of death. Book IV explains cognitive and other vital functions, ending with a diatribe against sexual passion. Finally, Books V and VI seek to demystify the world as a whole. The former explains its origin and future destruction, its structure and early human history. The latter accounts naturalistically for a comprehensive series of phenomena, from earthquakes to magnets.

In a pair of famous metaphors, Lucretius describes himself as 'traversing the pathless haunts of the Pierides [that is, Muses], where no one's foot has trodden before' (I 921-50), and as planting his feet in the precise tracks which Epicurus has marked for him (III 3-4). In short, he regards himself a faithful Epicurean, but as breaking new ground in his use of poetry. A central question of Lucretian criticism has long been the relationship between poetry and philosophy. It has been doubted whether strict Epicurean philosophy approved of poetry, but the evidence for this is inconclusive at best. Lucretius himself regards his subject matter as off-puttingly obscure, yet essential to human happiness. He therefore compares his use of poetry to the honey which doctors smear on the lip of a cup of bitter medicine in order to trick children into drinking it, for their own greater good. Epicurus is the saviour of mankind, whose philosophy alone can free us from the shackles of religion and from terror at the prospect of death. Both the formal addressee Memmius and the intended wider readership are novices in Epicureanism. Lucretius attracts them in at the beginning with a traditional hymn to Venus, then proceeds to lead them systematically through his arguments about the world's nature and structure. But at the end he confronts them with a disturbing account of the great plague at Athens, which might be regarded as a final test.

Lucretius' poem concentrates on physics, but important ethical statements are made in the proems to several books

and in the concluding parts of Books III, IV and V. He makes many notable contributions to our understanding and appreciation of Epicureanism. For instance, at II 216-93 he offers the only detailed surviving argument for the ‘swerve’ of atoms, which in Epicureanism underpins free will. Book III’s concluding diatribe against the unacknowledged source of most human misery, the fear of death, is especially celebrated. And the history of early human civilization in Book V has been much admired, especially for its reconstructions of the origins of society, language and justice.

In the world of Latin Christianity, Lucretius became the leading voice of Epicureanism, much quoted and vilified as an enemy of religion. He was widely read and studied in the Renaissance, with a tendency to Christianization of his thought (see Renaissance philosophy). He remains a primary source for our knowledge of Epicurean physics.

See also: Atomism, ancient

MICHAEL ERLER

List of works

Lucretius (*c.*55 BC) *On the Nature of Things*, ed. C. Bailey, *Titi Lucreti Cari De rerum natura libri sex*, Oxford: Oxford University Press, 3 vols, 1947; trans. R.E. Latham, revised by J. Godwin, *Lucretius On the Nature of the Universe*, London: Penguin, 1994; trans. W.H.D. Rouse, revised M.F. Smith, *Lucretius, De rerum natura*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1975.(Bailey is the monumental standard edition, with text, translation and commentary, still highly informative and not fully superseded; Latham offers a serviceable English translation; Rouse/Smith a conservative text with good facing English translation.)

References and further reading

Boyancé, P. (1963) *Lucrèce et l’épicurisme*, Paris: Presses Universitaires de France. (A classic study of Lucretius as poet and philosopher.)

Clay, D. (1983) *Lucretius and Epicurus*, Ithaca, NY, and London: Cornell University Press.(One of the best-integrated philosophical studies of Lucretius’ poetry.)

Erler, M. (1994), ‘Epikur; Die Schule Epikurs; Lukrez’ (Epicurus; The School of Epicurus; Lucretius), in H. Flashar (ed.) *Die Philosophie der Antike, Band 4: Die Hellenische Philosophie*, Basle: Schwabe, 381-490. (Informative about Lucretius and his philosophy, with extensive bibliography.)

Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press, 2 vols.(Discussion of various passages of Lucretius in their philosophical context.)

Segal, C. (1990) *Lucretius on Death and Anxiety*, Princeton, NJ: Princeton University Press.(A study of Lucretius’ poetry and thought, centred on Book III.)

Łukasiewicz, Jan (1878-1956)

Before 1918, Łukasiewicz's interests centred on logic (in the broad sense) and philosophy, and he worked on induction and probability. He also wrote an important historical book on the principle of contradiction in Aristotle. After 1918, Łukasiewicz concentrated almost entirely on mathematical logic and was the main organizer of the Warsaw School of Logic. The discovery of many-valued systems of logic is perhaps the most important result he achieved. He also invented an ingenious logical symbolism in which brackets (or other punctuation signs) are not necessary (bracket-free or Polish notation). Propositional calculi became a favourite topic of Łukasiewicz's logical investigations. The history of logic was another subject in which Łukasiewicz achieved important results.

1 Life

Jan Łukasiewicz was born on 21 December 1878 in Lwów. In 1896 he began to study law at the University of Lwów, but in the next year he changed his mind and decided to study philosophy under Kazimierz Twardowski. He obtained his Ph.D. on the basis of a thesis on the theory of induction. In 1904-5 he studied in Berlin, with Karl Stumpf, and in Louvain, with the cardinal Mercier. In 1906 he obtained his habilitation degree on the basis of a dissertation on the concept of cause. In the same year he became a *Privatdozent* at the University of Lwów and began courses in logic. His course in 1906 was the first in mathematical logic in Poland. In 1909 Łukasiewicz went to Graz, where he participated in Meinong's seminars (see Meinong, A.). In 1911 he was appointed an extraordinary professor in Lwów. When the University of Warsaw reopened in 1915 (it had been closed by the Tsarist government in the nineteenth century), Łukasiewicz was invited to assume the full professorship there. In 1918 he briefly left to become Minister of Religious Denominations and Public Education in Paderewski's government, but returned in 1920 when the university established a special position for him as professor of philosophy at the Faculty of Mathematics and Science; in fact, it was a position in mathematical logic. As a professor, Łukasiewicz began a very intensive research and teaching programme with the aim of establishing a logical school in Poland. He fully succeeded and, together with Stanislaw Leśniewski and the Warsaw mathematicians (Siperpiński, Mazurkiewicz, Kuratowski), he trained a group of brilliant logicians, including Alfred Tarski (who soon after became the third leader of the Warsaw Logic Group). In the interwar period Łukasiewicz was twice elected the Rector of Warsaw University; he was also awarded an honorary degree by the University of Münster. In 1943-4 he taught in the clandestine University in Warsaw, but he left Poland in 1944 and spent the last months of the Second World War in Germany. In 1946 he became the professor of mathematical logic at the Royal Irish Academy in Dublin, which awarded him an honorary degree in 1955. He died in Dublin on 3 February 1956.

2 Early writings

Induction and probability were the favourite subjects of Łukasiewicz's research in the period 1902-13. At first, he tried to develop the inverse theory of induction proposed by the nineteenth-century logicians Jevons and Sigwart. At the same time, Łukasiewicz was always sceptical about the possibility of evaluating inductive conclusions by probability. He argued that if we have an empirical hypothesis referring to the infinite domain, then its initial (a priori) probability approaches zero and no finite empirical evidence can change this situation. This argument anticipates Popper's basic tenets of anti-inductivism (see Popper, K.R. §2). Łukasiewicz maintained this scepticism about the value of induction in his later writings (1929); he argued that induction had no scientific value, because science is basically deductive and creative.

While working on induction, Łukasiewicz encountered the problem of the logical foundations of probability (1913). In line with his scepticism about induction, he held that probability has no use in the logical evaluation of sentences which are either true or false. On the other hand, formulas with free variables ('indefinite proposition' is his term) can be characterized by their logical probability. Let *D* be a domain consisting of a finite number of objects and let *Fx* be a formula with *x* as a free variable. Assume that *n* is the number of objects in *D* and *m* is the number of those objects which satisfy the formula *Fx*. Under these assumptions, the ratio m/n expresses the probability of *Fx*. The finite case is extensible to infinite domains by means of the mathematical theory of probability.

During this period Łukasiewicz also produced important work on the liar paradox. For Łukasiewicz (1915), the sentence 'This sentence is false' cannot be a value of variables representing sentences. Thus, it does not fall under general logical rules including itself. However, Łukasiewicz did not explain why the sentence cannot be a value of a variable. Tarski carried this out later, developing Łukasiewicz's ideas and arguing that the liar paradox arises because the sentence confuses object-language and metalanguage (see Semantic paradoxes and theories of truth §1).

O zasadzie sprzeczności u Arystotelesa (On the Principle of Contradiction in Aristotle) (1910) is Łukasiewicz's most important work of the period 1902-17. This book has two objectives: a contribution to the history of philosophy by providing an original interpretation of the principle of contradiction in Aristotle, and an evaluation of arguments in its favour (see Aristotle §5). According to Łukasiewicz, we can find three meanings of the principle in Aristotle: first, the logical principle that no proposition and its negation can be true together; second, the ontological principle that no object can have and not have the same property at the same time; and third, the psychological principle that one cannot have contradictory beliefs. Now, the psychological principle of contradiction is plainly false. Since one can prove the equivalence of the logical and ontological principles of contradiction, there remains the problem of providing a justification for one of them; Łukasiewicz argues that Aristotle's justification is not convincing. His main thesis is that we can give a justification for the principle (it follows from the definition of 'being an object'), but that it is not thereby shown to be among the axioms of logic. Hence the reason that we accept it is ethical rather than logical: we need some means of distinguishing truths from falsehoods.

3 Many-valued logic

Łukasiewicz's doubts about the principle of contradiction clearly set him on the path of many-valued logic. Before leaving the university in 1918 for his post in the government, he announced his discovery of a non-Aristotelian logic, then presented the idea in two talks in Lwów in 1920. Łukasiewicz's first motivation for many-valued logic was ethical. He believed that creativity, freedom and responsibility require a special ontology in which determinism fails. For this reason, he introduced a third logical value: possibility, which can be ascribed to sentences about future events. Łukasiewicz was also guided by the ancient question of what truth-value, if any, statements about the future have. However, he soon began to describe his system as 'many-valued logic' instead of 'non-Aristotelian logic'. For it is not clear that Aristotle himself considered the law of the excluded middle to be universally valid; there are statements in Aristotle's writings which suggest that this principle does not apply to sentences about the future. Furthermore, it was Chrysippus, a Stoic, who argued very strongly for the universality of the excluded middle. Thus Łukasiewicz called many-valued logic 'non-Chrysippean' rather than non-Aristotelian. However, on his final view, the issue does not lie in the acceptability of certain principles of logic, the law of contradiction or the law of the excluded middle, for example. Since logic is based on metalogical assumptions, the difference between two-valued and many-valued logic ultimately depends on whether one accepts the principle of bivalence: that every sentence is either true or false.

From the formal point of view, Łukasiewicz first elaborated a three-valued logic. Let the fraction $\frac{1}{2}$ denote the third value. If *p* has this value, its negation not-*p* has the same value. A disjunction always has the value of the 'greater or equal' disjuncts, a conjunction has that of the 'smaller or equal' conjuncts, and an implication has the third value, if its antecedent is 'third' and its consequent false or its antecedent is true and its consequent 'third'; in all other cases an implication is true, except where it has a true antecedent and a false consequent, in which case it is false. So the truth-tables for cases in which only truth and falsity are involved remain the same as usual. Łukasiewicz then generalized many-valued logic to systems with an arbitrary, finite or infinite, number of values. The conditional 'If *p*, then *q*' is true if $p \leq q$, and has a value given by the formula $1 - p + q$ if $p > q$. The value of not-*p* is always calculated as $1 - p$. Łukasiewicz also considered the axiomatization of many-valued logic. For example, he and Wajsberg established that many-valued logic with a countable infinite number of values is axiomatized by the formulas (in Łukasiewicz's notation; see §4 below): *CpCqp*, *CCpqCCqrCpr*, *CCCpqqCCqpq*, *CCCpqCqpCqp*, *CCNpNqCqp*.

Many-valued logic became a basis for Łukasiewicz's treatment of modality. He wanted to have a modal logic with the following principles: (1) if it is not possible that *p*, then not-*p*, (2) if not-*p*, then it is not possible that *p* and (3) for some *p*, it is possible that *p* and it is possible that not-*p* (Aristotle's principle). Łukasiewicz proved that (1)-(3)

cannot be represented together in two-valued logic. Tarski suggested to him that possibility could be defined within three-valued logic by the formula ‘If not-*p*, then *p*’. However, Łukasiewicz did not complete his modal logic before the Second World War. He returned to this question after 1945 and offered several systems of modal logic (see Modal logics §1).

A historical question remains: who discovered many-valued logic? There are anticipations in antiquity and the Middle Ages. One can also find several hints in Peirce’s work and in that of the Russian philosopher Vasiliev. However, it is evident that Łukasiewicz first elaborated many-valued logic as a mature formal system along with its intuitive motivation (see Many-valued logic §1).

4 Propositional calculi

Łukasiewicz invented a special notation for logic. This notation, often called ‘Polish’ notation, ensures that punctuation signs (brackets, dots and so on) are dispensable. The principle is that functors are always written before arguments. For example, the formulas $(p \rightarrowtail \neg p) \rightarrow q$ and $p \rightarrow (\neg p \rightarrow q)$ have respectively ‘*CCpNpq*’ and ‘as their ‘Polish’ counterparts *CpCNpq*’ (‘*C*’ stands for implication, ‘*N*’ for negation). It can be proved that the succession of letters in the Polish notation uniquely determines the structure of expressions.

Łukasiewicz achieved several results in propositional calculi. For example, he axiomatized the classical propositional calculus by the following set of axioms: *CCpqCCqrCpr*, *CCNppp*, *CpCNpq*. He also investigated partial propositional calculi, that is, calculi in which there occur only some of the propositional functors, such as systems with equivalence or implication as sole functors. Giving the most economic axiomatic bases for propositional calculi was his favourite subject-matter. For example, he proved that the sole axiom for the equivalential propositional calculus must have at least ten letters. Łukasiewicz also investigated intuitionistic logic.

5 The history of logic

Łukasiewicz had a deep understanding of the history of logic and proposed a study of it from the point of view of modern mathematical logic. The result was very striking. Łukasiewicz, recognizing that there is a basic difference between the propositional calculus and syllogistic logic which had been overlooked before mathematical logic arose, discovered that the ancient Stoics had elaborated a propositional calculus and thus had a system of logic which differed fundamentally from that of Aristotle (see Logic, ancient §5; Stoicism §10). Łukasiewicz also found interesting examples of propositional logic in the Middle Ages (see Logic, medieval §9). But Aristotle remained Łukasiewicz’s favourite philosopher. Even before 1939 he began research on Aristotelian logic which resulted in an extensive reconstruction of syllogistic logic from the point of view of modern logic. In his book *Aristotle’s Syllogistic from the Standpoint of Modern Formal Logic* (1951) he brought this work together, while displaying his new methodology for the history of logic.

6 Philosophy of logic

Many-valued logic leads to the question of which logic is ‘right’. At first, Łukasiewicz advocated the ontological conception of logic: logic is a description of reality. Hence, one and only one system from the family of rivals is true in the real world. He believed that the infinite many-valued logic is right. However, after 1945 he took a more conventionalist route: logics are devices of our thinking, and when we decide which tool is to be chosen, we must take into account various factors, such as simplicity. Łukasiewicz regarded logic as an independent science which is the servant neither of philosophy nor of mathematics. However, he always maintained that logic is of the utmost significance for philosophy. In particular, he demanded that philosophy should be axiomatized.

See also: Polish logic §2

JAN WOLEŃSKI

List of works

Łukasiewicz, J. (1970) *Selected Works*, ed. L. Borkowski, Amsterdam and Warsaw: North Holland and PWN. (A selection of the most important logical and philosophical papers of Łukasiewicz.)

Łukasiewicz, J. (1903) ‘O indukcji jako inwersji dedukcji’ (On Induction as Inversion of Deduction), *Przegląd Filozoficzny* 6: 9-24; 138-52.(Łukasiewicsz’s Ph.D. thesis; an attempt to develop the inversion theory of

induction.)
Łukasiewicz, J. (1906) 'Analiza i konstrukcja pojęcia przyczyny' (An Analysis and Construction of the Concept of Cause), *Przegląd Filzoficzny* 9: 105-79.(Łukasiewicz's *Habilitation* dissertation; an attempt to define causality as a logical relation.)
Łukasiewicz, J. (1909) 'O prawdopodbieństwie wniosków indukcyjnych' (On the Probability of Inductive Conclusions), *Przegląd Filozoficzny* 12: 209-10.(Łukasiewicz's arguments against the probabilistic justification of induction)
Łukasiewicz, J. (1910) *O zasadzie sprzeczności u Arystotelesa*, Cracow: Polska Akademia Umiejętności; partly trans. V. Wedlin in 'On the Principle of Contradiction in Aristotle', *Review of Metaphysics* 24: 485-509.(An extensive treatment of Aristotle's views on the principles of contradiction, together with its interpretation in the framework of algebra of logic.)
Łukasiewicz, J. (1913) *Die logischen Grundlagen der Wahrscheinlichkeitsrechnung (The Logical Foundations of Probability)*, Cracow: Polska Akademia Umiejętności; partly trans. O. Wojtasiewicz in *Selected Works*, Borkowski, 1970.(An attempt to formulate the logical foundations of probability.)
Łukasiewicz, J. (1915) 'O nauce' (On Science), in *Poradnik dla samouków (A Guide for Autodidacts)*, Warsaw: Heflich i Michalski, XV-XXXIX.(Contains Łukasiewicz's treatment of the liar paradox.)
Łukasiewicz, J. (1918) *Wykład pożegnalny w Uniwersytetu Warszawskiego, 7 marca 1918 (Farewell Lecture in the University Hall of the University of Warsaw, 7 March 1918)*, Warsaw; trans. O. Wojtasiewicz in *Selected Works*, Borkowski, 1970.(The first public announcement of many-valued logic and motives behind it.)
Łukasiewicz, J. (1920a) 'O pojęciu możliwości' (On the Concept of Possibility), *Ruch Filzoficzny* 6: 169-70; trans. O. Wojtasiewicz in *Selected Works*, Borkowski, 1970.(Łukasiewicz's first talk on many-valued logic, given in Lwów in 1920; contains formal details of many-valued logic.)
Łukasiewicz, J. (1920b) 'O logice trójwartościowej' (On Three-Valued Logic), *Ruch Filzoficzny* 6: 170-1; trans. O. Wojtasiewicz in *Selected Works*, Borkowski, 1970.(The second talk on many-valued logic in Lwów in 1920; a continuation of 1920a.)
Łukasiewicz, J. (1929) *Elementy logiki matematycznej*, Warsaw: PWN; trans. O. Wojtasiewicz, *Elements of Mathematical Logic*, Oxford: Pergamon Press, 1963.(Lecture notes of Łukasiewicz; contain an exposition of Polish notation, a system of propositional logic, various metalogical results, information on many-valued logic, plus a brief formal treatment of syllogistic.)
Łukasiewicz, J. (1930) 'Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls', *Compt. Rend. Soc. d. Sc. Let. Varsovie, Cl. III* 23: 51-77; trans. H. Weber in *Selected Works*, Borkowski, 1970.(A systematic exposition of intuitions and formal problems of many-valued logic.)
Łukasiewicz, J. (1951) *Aristotle's Syllogistic from the Standpoint of Modern Formal Logic*, Oxford: Oxford University Press; 2nd enlarged edition, 1957.(A systematic modern treatment of Aristotelian syllogistic; the second edition contains chapters on modal syllogistic.)

References and further reading

Bolc, L. and Borowik, P. (1992) *Many-valued Logics, 1, Theoretical Foundations*, Springer-Verlag.(An extensive monograph on many-valued logics, including a survey of Łukasiewiczian logics.)
Borkowski, L. and Słupecki, J. (1958) 'The Logical Works of Jan Łukasiewicz', *Studia Logica* 8: 7-56.(Contains a presentation of all Łukasiewicz's works in logic, including his ideas in philosophy of science.)
Kwiatkowski, T. (1981) 'Jan Łukasiewicz - A Historian of Logic', *Organon* 16/17: 169-88.(A survey of Łukasiewicz's works in the history of logic.)
Malinowski, G. (1993) *Many-valued Logics*, Oxford: Oxford University Press.(A concise survey of many-valued logics from the contemporary point of view.)
Rescher, N. (1967) *Many-valued Logic*, New York: McGraw-Hill.(A extensive historical and systematic survey of many-valued logics with a very useful bibliography.)
Seddon, F.A. (1996) *Aristotle and Łukasiewicz on the the Principle of Contradiction*, Ames, IA: Modern Logic Publishing.(A critical analysis of Łukasiewicz's treatment of the principle of contradiction in Aristotle; however, Seddon bases his criticism on Łukasiewicz (1971), which is only a summary of Łukasiewicz (1910).)
Sobociński, B. (1957) '*In Memoriam* Jan Łukasiewicz', *Philosophical Studies* (Maynooth, Ireland) 6: 3-49. (A detailed presentation of Łukasiewicz's logical ideas; this paper also includes Łukasiewicz's short autobiography.)
Woleński, J. (1989) *Logic and Philosophy in the Lvov-Warsaw School*, Dordrecht: Kluwer.(Chapters 1, 3, 5, 6 and 10 contain information about Łukasiewicz's views in philosophy as well as his works in logic and philosophy of science.)

Woleński, J. (1994) ‘Jan Łukasiewicz on the Liar Paradox, Logical Consequence, Truth, and Induction’, *Modern Logic* 4: 392-400.(Contains a summary of Łukasiewicz ’s views on the liar paradox, logical consequence, truth and induction.)

Wójcicki, R. and Czelakowski, J. (eds) (1977) *Selected Papers on Łukasiewicz Sentential Calculi*, Wrocław: Ossolineum.(A collection on Łukasiewiczian logics from the contemporary perspective.)

Lukács, Georg (1885-1971)

Lukács' Geschichte und Klassenbewusstsein (History and Class Consciousness) (1923) is, for both its intrinsic merits and its enormous influence, the most important work of Marxist philosophy to have appeared in the twentieth century. It sought to render explicit the dependence of Marx's thought on Hegel's dialectic as a means of elucidating both the distinctive character of historical materialism as a form of theoretical inquiry and its revolutionary rejection of the modes of thinking prevailing in capitalist society. Lukács' general aim had been shared by the authors of the first philosophical reflections on Marx's project - Engels and Plekhanov, for example, had stressed its debt to Hegel. Lukács, however, sought to draw Marx into that broad current of twentieth-century Continental thought which has drawn a sharp distinction between the methods of the physical sciences, suitable at best for analysing inanimate nature, and those of the human sciences, whose aim is to interpret human actions in the light of the thoughts which move them. Thus Lukács sees Marx as the theorist, not of the laws of the dialectic or of inevitable social transformation, but of revolutionary subjectivity, of the proletariat as 'the identical subject-object' of history. This was a version of Marxism which suited the times, in the immediate aftermath of the Russian Revolution of October 1917. As the revolutionary tides receded, Lukács found philosophical and political reasons for retreating to a more orthodox historical materialism which laid much greater stress on objective constraints and processes than his version of the early 1920s had. Yet the force of its overall argument and the quality of its individual analyses have made History and Class Consciousness a constant reference-point in subsequent discussions of Marxist theory.

1 Life

Lukács is perhaps the only major Marxist philosopher to have had a significant intellectual career prior to his becoming a Marxist. He later described his youthful views as 'romantic anti-capitalism' - a rejection of existing society based on nostalgia for a past believed to be less fragmented and individualist than the present. The sources of this outlook are no doubt connected to the tensions in Lukács' own background. He was born in Budapest on 13 April 1885, the son of a successful self-made Jewish banker. The predominantly Jewish business class were outsiders in a Hungary still dominated by the gentry, a fact which perhaps helps to explain what Lukács was to call his 'passionate rejection of the order then existing in Hungary', expressed, for example, in an enthusiasm for the revolutionary poetry of Endre Ady. The prime philosophical influences on Lukács were, however, German. He took part in the Heidelberg Circle which Max Weber gathered around him between 1906 and 1918 (see Weber, M. §1). Lukács' pre-Marxist writings, of which the most notable was *Die Theorie des Romans (The Theory of the Novel)* (1916), are permeated by the ideas of members of the circle such as Georg Simmel. They combine a strongly critical attitude towards a modernity conceived as rationalized and soulless with a Neo-Kantian insistence on sharply distinguishing between facts and values, so that authentic values are held to be incapable of achieving practical realization.

Lukács' decision to join the new Hungarian communist party in December 1918 and his participation in its abortive attempt to take power at the beginning of 1919 reflected the belief that this 'tragic' antinomy could after all be overcome, that socialist revolution could bring into being a society which would not suffer from the disharmonies characteristic of modern capitalism. Exiled after the failure of the 1919 Revolution, Lukács was for the rest of his life a leading figure in the international communist movement, indeed he was a party functionary until he took up academic positions after his return to Hungary in 1945.

Controversy, however, dogged his career as a communist. Initially, he was a 'left communist', on the wing of the communist international which demanded an unremitting revolutionary offensive against capitalism. *Geschichte und Klassenbewusstsein (History and Class Consciousness)* (1923), Lukács' masterwork, is arguably still infused with the Messianic hopes which inspired his original adhesion to communism. Its hostile reception in the communist press was, however, one sign of the consolidation of a new 'Leninist' orthodoxy to which Lukács progressively adapted himself. Although *History and Class Consciousness* had an enormous impact on German intellectual life under the Weimar republic (to the extent that Lukács' admirer Lucien Goldmann claimed that Heidegger's *Being and Time* (1927) was, in part, a response), Lukács soon began to distance himself from his book. He always insisted that his reasons for doing so were philosophical (see §3). Certainly, he stuck by the views which he first formulated in the mid-1920s - which involved, in stark contrast to his youthful writings, an

emphasis on the positive virtues of bourgeois civilization and on the continuities between them and those of a socialist society - even when it was politically dangerous to do so. Lukács' participation in Imre Nagy's government during the Hungarian Revolution of 1956 nearly cost him his life. Some commentators, however, believe that, in the very last years before his death on 4 June 1971, Lukács, under the impact of the student revolts in the West and in reaction to the 1968 invasion of Czechoslovakia, was returning to the more revolutionary Marxism of his youth.

2 *History and Class Consciousness*

History and Class Consciousness may be seen as the most powerful philosophical expression of a general revolt against the dominant form of Marxism in the Second International (1889-1914). Antonio Gramsci called the Bolshevik Revolution of October 1917 'The Revolution against *Capital*'; the victory of socialist revolution in a backward country overturned the Marxist orthodoxy formulated by theorists such as Karl Kautsky and Georgii Plekhanov according to which history proceeded by natural necessity according to deterministic economic laws. Both Lukács and Gramsci were influenced by Georges Sorel, for whom revolution was an act of collective will.

However, rather than abandon Marxism as Sorel did, Lukács sought to reformulate it in terms consistent with this conception of revolution. This involved, in the first place, an identification of orthodoxy with method. Marxism could not be equated with any specific theoretical proposition, even the claim that production provides social life with its foundation. Famously Lukács took this to the extreme of asserting that 'if recent research had disproved once and for all every one of Marx's individual theses' (1923: 1), the truth of Marxism itself would be unaffected.

The distinctive feature of the Marxist method, according to Lukács, is that it conceives the different elements of social life, not as isolated fragments, but as aspects of an integrated whole. '*The primacy of the category of totality is the bearer of the principle of revolution in science*' (1923: 27; original emphasis). This assertion is a variation on Hegel's declaration that 'the True is the whole'. Lukács further took from Hegel the idea that it is possible to know the totality only because the subject and object of knowledge are identical: 'consciousness here is not the knowledge of an opposed object but is the self-consciousness of the object' (1923: 178). Therefore, 'reality can only be understood and penetrated as a totality, and only a subject which is itself a totality is capable of this penetration' (1923: 39).

Lukács' appropriation of Hegel provides the basis of the critique of bourgeois thought and society which forms the leitmotif of the central essay in *History and Class Consciousness*, 'Reification and the Consciousness of the Proletariat'. Capitalism is precisely a form of society which operates as if it were not a totality. Taking over and integrating into a Marxist framework the ideas of Weber and Simmel, Lukács depicts a society in which quantity increasingly drives out quality and where instrumental rationality reigns: individual aspects of social life may be modernized and subjected to bureaucratic procedures, but the relationships which bind these together into a whole are incomprehensible. But whereas Weber conceived this process of rationalization as an inescapable consequence of the triumph of modernity, Lukács treats it as a developed form of what Marx in *Capital* calls commodity fetishism (see Marx, K. §10). It is distinctive of capitalism, Marx argues, that social relationships among human beings take on the form of relationships between things. The exchange of commodities on the market rules people's lives, setting a price on their labour power and, in times of economic crisis, perhaps condemning them to unemployment and poverty. Lukács takes over this idea and develops it into the idea of a process of 'reification' permeating the whole of social life. Even the most advanced forms of modern philosophy from Descartes to Kant are scarred by commodity fetishism, and in particular by their inability to grasp society as a whole. Hegel alone recognized that the social totality could only be understood as and by a total subject, but expressed this thought in the mystified form of the concept of Absolute Spirit.

Lukács' solution to what he calls 'the antinomies of bourgeois thought' involved, once again, making strategic use of Marx's *Capital*. Capitalist exploitation depends, according to Marx, on the transformation of labour power, or the ability to work, into a commodity bought and sold for the market. For Marx the chief interest of this state of affairs lies in the way that it makes possible, he believes, the extraction of surplus value from the worker. For Lukács, however, it is of significance chiefly as the limit-point of reification: the workers - that is, the bulk of humanity - are transformed into saleable objects. But, since the exploitation that follows is the presupposition of the entire network of commodity relationships constituting capitalist society, the worker is the pivot of this society. It follows, Lukács believes, that only the working class is in a position to comprehend capitalist society as a

totality. For ‘the worker can only become conscious of his existence in society when he becomes aware of himself as a commodity’. Therefore, ‘his consciousness is *the self-consciousness of the commodity*; or in other words it is the self-knowledge, the self-revelation of the capitalist society founded upon the production and exchange of commodities’ (1923: 168; original emphasis).

The proletariat is thus the total subject required to understand society as a totality, because knowing its own position involves knowing the set of relations of which, as absolute commodity, it is the source. It is, therefore, ‘the identical subject-object of the social and historical processes of evolution’. Further, it is this peculiar status of the working class which makes historical materialism possible. Marx and his successors theoretically articulated the self-understanding of the proletariat. They were able to do so, despite their usually not being workers themselves, because this self-understanding depends on the objective position of the proletariat in capitalist society. There is, in principle, a distinction between the class consciousness that can be imputed to workers on the basis of this position and their actual beliefs and desires. Moreover, the process through which workers become aware of their real situation in society is primarily an intellectual one. None the less, ‘the consciousness of the proletariat must become deed’ (1923: 178). Lukács believes that this is partly because workers becoming self-aware will have practical consequences: ‘*the act of consciousness overthrows the objective form of its object*’ (1923: 178; original emphasis). At the same time, the very process of class struggle in which workers are compelled to engage by their economic circumstances will transform their consciousness. In this way the reified structures of capitalist society will be broken up.

History and Class Consciousness is undeniably a philosophical *tour de force*. Lukács skilfully draws on the resources of German classical idealism and a close reading of Marx’s works to offer a version of Marxism which is, primarily, a theory of revolutionary subjectivity. It is, however, easy to see why later Marxists, such as the Frankfurt School, were more influenced by Lukács’ analysis of reification than by his theory of class consciousness (see Frankfurt School §1). He himself later called the book ‘an attempt to out-Hegel Hegel’, with the proletariat as a sociological version of Absolute Spirit. If this self-criticism is hard to fault, some of the charges made by others seem less fair. The concept of imputed class consciousness, for example, has been seen as an anticipation of Stalinism, since it seems to allow the revolutionary party to tell the working class what the latter ‘objectively’ thinks. In fact, the idea of a party hardly figures in the key ‘Reification’ essay; it appears mainly in the two final chapters of *History and Class Consciousness*, which engage much more directly, and empirically, with historical and social realities than the rest of the book.

3 Lukács as an orthodox communist

For Lukács to come to terms with a communist movement which by the mid-1920s was becoming increasingly Stalinist in any case involved abandoning the main positions of *History and Class Consciousness*. The decisive step in this transition was taken in ‘Moses Hess and the Problem of the Idealist Dialectics’ (1926). Here he defends the later Hegel’s ‘reconciliation with reality’, most famously expressed in the Preface to the *Philosophy of Right* (1821). Lukács praises ‘Hegel’s magnificent realism, his rejection of all utopias, his concern to conceive philosophy *as the conceptual expression of history itself* and not as philosophy *about* history’ (1972: 188), and denounces his left Hegelian critics, such as Hess, for regressing to the subjective idealism of Kant and Fichte. The same stance - an identification with the unfolding historical process, warts and all - informs Lukács’ later philosophical writings, for example, *Der junge Hegel* (*The Young Hegel*) (1948). In Moscow in the late 1920s he was among the first to study Marx’s *Economic and Philosophic Manuscripts of 1844*. This led him to seek to reformulate the dialectic of subject and object at work in *History and Class Consciousness* as the interaction of man and nature, mediated by labour. This forms the leading idea of the posthumously published *Zur Ontologie des gesellschaftlichen Seins* (*Ontology of Social Being*) (1971).

Lukács’ later ‘reconciliation with reality’ involved a strongly positive appraisal of classical bourgeois culture, especially of nineteenth-century literary realism. Indeed Michael Löwy calls ‘Lukács’ political and intellectual career from 1928 onwards… a consistent attempt to "reconcile" Stalinism with bourgeois-democratic culture’ (1979: 204). Certainly there was a marked change from Lukács’ pre-Marxist years, when he had been influenced by Dostoevskii’s novels, works which arguably anticipate many of the formal innovations of modernism. Lukács’ voluminous aesthetic writings from the 1930s onwards wage constant war on the modernism of Joyce and Kafka, which, he argues, marks a regression from the pinnacle of bourgeois cultural achievement, the classical realist

novel. Writers such as Tolstoi and Mann, despite their bourgeois limitations, anticipate the future achievements of socialist realism. Never a time-server (he greeted Solzhenitsyn as a socialist realist novelist). Lukács certainly became a cultural conservative, at odds with Marxist champions of modernism such as Brecht and Adorno. Subsequently, however, Fredric Jameson (1981) sought to develop a version of Lukácsan Marxism in which there is a place for modernism, and even for postmodernism.

ALEX CALLINICOS

List of works

Lukács, G. (1916) *Die Theorie des Romans*, Berlin: Paul Cassirer, 1920; trans. A. Bostock, *The Theory of the Novel*, London: Merlin, 1978.(Lukács' most important pre-Marxist work, expressing what he came to regard as a romantic rebellion against, rather than a materialist critique of capitalism.)

Lukács, G. (1923) *Geschichte und Klassenbewusstsein*, Berlin: Malik-Verlag; trans. R. Livingstone, *History and Class Consciousness*, London: Merlin, 1971.(The most important single work of twentieth-century Marxist philosophy; the English edition contains a highly self-critical preface written in 1967.)

Lukács, G. (1924) *Lenin*, Berlin: Malik-Verlag; trans. N. Jacobs, *Lenin*, London: New Left Books, 1970.(Written immediately after Lenin's death, and offering an interpretation of the unity of his thought around the theme of 'the actuality of the revolution'.)

Lukács, G. (1948) *Der junge Hegel*, Zurich and Vienna: Europa Verlag, 2 vols; trans. R. Livingstone, *The Young Hegel*, London: Merlin, 1975.(A detailed discussion of the development of Hegel's thought up to *The Phenomenology of Spirit* (1807), stressing its sociopolitical content and the theme of labour, and anticipating in many respects the later *Ontology of Social Being*.)

Lukács, G. (1954) *Die Zerstorung der Vernunft*, Berlin: Aufbau-Verlag; trans. P. Palmer, *The Destruction of Reason*, London: Merlin, 1980.(A wide-ranging assault on what Lukács regarded as the German irrationalist tradition culminating in Heidegger.)

Lukács, G. (1957) *Il significato attuale del realismo critico*, Turin: Einaudi; trans. J. Mander and N. Mander, *The Meaning of Contemporary Realism*, London: Merlin, 1963.(A succinct polemical statement of Lukács' views on realism and modernism.)

Lukács, G. (1963) *Die Eigenart des Asthetischen (The Specificity of the Aesthetic)*, Neuwied: Luchterhand, 2 vols. (The magnum opus of Lukács' later years.)

Lukács, G. (1971) *Zur Ontologie des gesellschaftlichen Seins*, Neuwied: Luchterhand; selections trans. D. Fernbach as *The Ontology of Social Being*, London: Merlin, vol.1, *Hegel* (1978), vol. 2, *Marx* (1978), vol. 3, *Labour* (1978).(An attempt to restate the concept of Being starting from an analysis of labour as the mediating activity uniting subject (humankind) and object (nature).)

Lukács, G. (1972) *Political Writings, 1919-1929*, trans. M. McColgan, London: New Left Books.(Includes, along with many more directly political texts reflecting mainly Lukács' 'left communist' period, 'Moses Hess and the Problem of the Idealist Dialectics' (1926), in which he announced his conversion to a more right Hegelian 'realism'.)

References and further reading

Arato, A., and Breines, P. (1979) *The Young Lukács and the Origins of Western Marxism*, London: Pluto. (Particularly good at reconstructing Lukács' German intellectual context and exploring the debate provoked by *History and Class Consciousness*.)

Bloch, E. *et al.* (1977) *Aesthetics and Politics*, London: New Left Books.(The main texts of the controversy among Marxists over modernism, including an important exchange between Lukács and Brecht.)

Jameson, F. (1981) *The Political Unconscious*, London: Methuen.(An ambitious attempt at a new synthesis by a cultural theorist heavily influenced by Lukács.)

Kadarkay, A. (1991) *George Lukács*, Oxford: Blackwell.(An informative if philosophically weak biography.)

Lowy, M. (1979) *George Lukács - From Romanticism to Bolshevism*, London: New Left Books.(A careful, politically shrewd account of Lukács' formation as an intellectual, culminating in *History and Class Consciousness*, but also dealing briefly with his later career.)

Stedman Jones, G. (1971) 'The Marxism of the Early Lukács', *New Left Review* 70: 27-64.(Incisive and highly influential critique of *History and Class Consciousness*, written from a standpoint sympathetic to Althusser.)

Lushi chunqiu

The Lushi chunqiu (Spring and Autumn Annals of Master Lu Buwei), composed 241-238 BC, marks a firm beginning for the eclectic movement in Qin and Han philosophy. It embraces various pre-Qin philosophies such as Lao-Zhuang and Huang-Lao Daoism, Confucianism, Mozi, Legalism, the logicians, the military arts, Agriculturalists, Yang Zhu, Zou Yen and Story Tellers. As a compendium of classical knowledge, the Lushi chunqiu contains cultural and philosophical material on the art of rulership.

The *Lushi chunqiu*, or *Lü-shih Ch'un-ch'iu* (Spring and Autumn Annals of Master Lu Buwei) was completed between 241 and 238 BC by guest scholars at the estate of Lu Buwei during his appointment as Prime Minister of Qin (249-237 BC). It is divided into three sections: the Twelve Chronicles (*shi'erji*), containing twelve subsections with five chapters each and a postscript; the Eight Observations (*balan*), divided into eight subsections of eight chapters each (the first subsection is missing a chapter); and the Six Discussions (*liulun*), divided into six subsections with six chapters apiece. These numbers possibly had symbolic significance, representing the Three Powers, heaven, earth and humans.

The lead chapter of each Chronicle, arranged according to the months of the four seasons, gives instructions on the monthly court ritual, also found in the 'Monthly Commands' chapter of the *Liji (Book of Rites)*. The remaining chapters in each Chronicle discuss numerous concerns of rulership from an eclectic perspective, loosely arranged under a seasonal programme. The spring section contains Yang Zhu and Daoist-like advice on self-cultivation; the summer presents Confucian approaches to education and music; the autumn focuses on military affairs; and the winter focuses on frugal funerals and administering state personnel.

The text is eclectic, and one ought to be wary of scholars who contend that *Lushi chunqiu* as a whole advocates one particular philosophy or argues exclusively against one. Although the postscript claims that the Twelve Chronicles relate the Yellow Emperor's Daoist teachings and the spring Chronicles are Daoist-like, it would be misleading to say that the text is chiefly Daoist because other chapters are clearly Confucian, Mohist and Legalist. Some contend that the text is anti-Legalist; however, the chapter 'Examining the Present' (*chajin*) advocates the Legalist doctrine of not following the standards of the early kings, and some of this material appears in edicts written later by the notorious Legalist Prime Minister of Qin, Li Si. The *Lushi chunqiu* is in fact drawn from various pre-Qin philosophers.

Some of the Daoist ideas in the *Lushi chunqiu* concern the ruler's self-cultivation and ruling by nonaction (*wuwei*) (see Self-cultivation in Chinese philosophy). The ruler practises non-action while the ministers perform their duties. The ruler is warned against overindulgence in sensual gratification, and to carefully guard and protect life; it is the ministers who must sacrifice their lives for the king. The ruler is encouraged to govern with public spiritedness; political factions and selfish exploits are to be avoided. The ruler is expected to be obscure and unknown. His self-cultivation generates harmony with nature, which maintains political order.

The Confucian teachings in the text also deal with the ruler's self-cultivation and rulership. Court ritual and music are emphasized; the *Lushi chunqiu* contains the most extensive extant pre-Qin material on ritual music. Analogies are drawn between musical and political harmony. Following the teachings of Confucius and Mencius, some passages delineate the importance of governing by moral virtue, following the example of the early sage-kings. Other passages resemble the heavy-handed teachings of Xunzi, emphasizing rewards and punishments and the need for the ruler to heed the remonstrance of teachers and ministers.

Mohist ideas are present as well. Many of the chapters, like Mozi, advocate a practical approach to rulership, especially learning from past historical errors. The Mohist material in the *Lushi chunqiu* advocates the value of keeping the people happy and not squandering the state treasury on royal funerals; the latter counters Confucian practice.

The text of Zou Yen and his Five Phases (*wuxing*) philosophy is lost, but passages in the *Lushi chunqiu* represent those teachings. One such passage narrates how past rulers established their respective dynasties by following the cosmic changes of the Five Phases. For example, King Wen founded the Zhou dynasty by complying with the Fire Phase. It further states that water will conquer fire; the King of Qin used this cosmology to justify his unification

of the empire in 221 BC. The *Lushi chunqiu*'s Militarist chapters reject disarmament, and advocate maintaining troops to fight a just war. A virtuous ruler ought to attack an immoral king who oppresses his people.

The *Lushi chunqiu* contains most of the extant pre-Qin Agriculturalist philosophy. Details on farming practices and the importance of bountiful harvests to maintain political order and attract immigrants are discussed. The Agriculturalists, like the Confucian ritualists, emphasize that the ruler should take timely action to create and maintain cosmic and political harmony. Finally, the Story Tellers, the masters of rhetoric, contribute a vast number of interesting, sometimes humorous, illustrative stories to elucidate various philosophical points in the *Lushi chunqiu*. Some of their practical guidelines are applicable to our age. For example, the various discussions on the relationship between humans and nature are relevant to environmental ethics. Some of the ideas on governing and employment of the masses are pertinent to modern politics. The advice given for the ruler's self-cultivation can be applied by anyone.

See also: Chinese philosophy; Confucian philosophy, Chinese; Daoist philosophy; Legalist philosophy, Chinese; Mohist philosophy; Self-cultivation in Chinese philosophy; Yangzhu

JAMES D. SELLMANN

References and further readings

Bodde, D. (1967) *China's First Unifier, A Study in the Ch'in Dynasty as seen in the Life of Li Ssu*, Hong Kong: Hong Kong University Press.(A valuable source for understanding the Qin (Ch'in) dynasty.)

Carson, M.F. (1985) *A Concordance to Lü-shih ch'un-ch'iu*, Taipei: Chinese Material Center.(Contains material on the structure and history of the text; required for a comprehensive study of specific characters.)

Fung Yu-lan (1952) *A History of Chinese Philosophy*, trans. D. Bodde, Princeton, NJ: Princeton University Press. 2 vols.(Volume 1 makes extensive use of *Lushi chunqiu* to explicate the philosophy of the Hundred Schools, especially Yang Zhu and Zou Yen.)

Graham, A.C. (1989) *Disputers of the Tao*, La Salle, IL: Open Court.(Proposes that Yangism is at the core of the text, and describes its cosmology.)

Hsiao Kung-ch'üan (1979) *A History of Chinese Political Thought*, trans F. Mote, Princeton, NJ: Princeton University Press.(Part 2, Section 3 develops the idea that the text is anti-Legalist.)

Louton, J. (1984) 'Concepts of Comprehensiveness and Historical Change on the *Lü-shih ch'un-ch'iu*', in H. Rosemont, Jr (ed.) *Explorations in Early Chinese Cosmology*, Chico, CA: Scholars Press, 105-18.(Argues that its authors understood historical change to be a product of the interaction of humans and nature.)

Lushi chunqiu (241-238 BC), trans. R. Wilhelm, *Frühling und Herbst des Lü Bu We*, Düsseldorf: Eugen Diederichs Verlag, 1979.(The first complete translation of the *Lushi chunqiu* in a Western language.)

Needham, J. (1956) *Science and Civilisation in China*, vol. 2, *History of Scientific Thought*, Cambridge: Cambridge University Press.(Provides an insightful survey of pre-Qin philosophy. Many of the later volumes cite passages from the *Lushi chunqiu* to illustrate Chinese science.)

Sellman, J. (1983) 'On Mobilizing the Military: Arguments for a Just War Theory from the *Lü-shih ch'un-ch'iu*', *Asian Culture Quarterly* 11 (4): 26-43.(Contains a translation of the relevant portion of the text.)

Sellman, J. (1985) 'The *Lü-shih ch'un-ch'iu*'s Proposal of Governing by Filial Piety', *Asian Culture Quarterly* 13 (1): 43-62.(Explicates the philosophical arguments, and provides a translation of one chapter.)

Sellman, J. (1990) 'Seasonality in the Achievement of *Hsing* in the *Lü-shih ch'un-ch'iu*', *Asian Culture Quarterly* 18 (2): 42-68.(Explicates the eclectic understanding of human nature.)

Watson, B. (1962) *Early Chinese Literature*, New York: Columbia University Press, 186-9.(Provides a critical review and a translation of one chapter.)

Luther, Martin (1483-1546)

Martin Luther was an Augustinian monk who found the theology and penitential practices of his times inadequate for overcoming fears about his salvation. He turned first to a theology of humility, whereby confession of one's own utter sinfulness is all that God asks, and then to a theology of justification by faith, in which human beings are seen as incapable of any turning towards God by their own efforts. Without preparation on the part of sinners, God turns to them and destroys their trust in themselves, producing within them trust in his promises made manifest in Jesus Christ. Regarding them in unity with Christ, God treats them as if they had Christ's righteousness: he 'justifies' them. Faith is produced in the sinner by the Word of God concerning Jesus Christ in the Bible, and by the work of the Holy Spirit internally showing the sinner the true subject matter of the Bible. It is not shaped by philosophy, since faith's perspective transcends and overcomes natural reason. Faith, through the working of God's Holy Spirit within the believer, naturally produces good works, but justification is not dependent upon them - they are free expressions of faith in love. Nevertheless, secular government with its laws and coercion is still necessary in this world because there are so few true Christians. Luther's theology brought him into conflict with the Church hierarchy and was instrumental in the instigation of the Reformation, in which the Protestant churches split from Rome.

1 Life

Martin Luther was born in Eisleben, Germany. While studying at the University of Erfurt in 1505, he was nearly struck by lightning in a storm. In terror he vowed to become a monk. Against the advice of friends and family, he entered the local monastery of Observant Augustinians shortly afterwards.

Over the next few years he found himself assailed by doubts as to his salvation and the forgiveness of his sins, which the nominalist theology and penitential practices current at that time did little to assuage. On becoming a Bible professor at Wittenberg in Electoral Saxony in 1512, Luther began to develop his own theological response to these trials through his lectures on the Psalms (1513-15) and on Romans (1515-16), eventually producing a theology of justification by faith (see §§2-3). In the process, he began to criticize the selling of 'indulgences' (reductions of one's punishment and guilt by authority of the Pope), publishing his *Disputatio pro declaratione et vitutis indulgentiarum (Disputation on the Power and Efficacy of Indulgences*, the 'Ninety-five Theses') explaining this disagreement in 1517. Trouble with Rome ensued, and escalated in 1519 when Luther, in a debate in Leipzig, was forced to shift the debate to the issue of papal authority.

In 1520 Luther began working out the implications of his theology. Such works as *Von der Freiheit eines Christenmenschen (The Freedom of a Christian)* and *An den christlichen Adel deutscher Nation (To the Christian Nobility of the German Nation)* outline a programme of reform for the Church. However, the Pope soon threatened him with excommunication, and at the Diet of Worms in 1521 Emperor Charles V confirmed the sentence, forcing Luther into hiding. While there he began work on a German translation of the Bible, a task which was to occupy him until 1534. The Emperor was unable to enforce his edict, however, and Luther was able to return to Wittenberg in 1522.

Meanwhile, others had been pursuing the reformation of the Church, and on his emergence Luther became embroiled in controversy with them. He rejected iconoclastic reform, and later opposed the Peasants' Revolt, in which radical reformers played a key role. His main debate with moderate reformers was over the Eucharist. At Marburg in 1529, attempts to forge a consensus broke down, with Luther sticking fiercely to his belief in the real presence of Jesus Christ in the Lord's Supper. Another significant debate was with Erasmus (§1) over free will, leading to the publication in 1525 of Luther's *De servo arbitrio (The Bondage of the Will)*.

In 1530 Charles V made another attempt to enforce the quashing of the Reformation, but he was not in a position to clamp down on those German princes (the 'Protestants') who refused to carry out his wishes. Luther was left with the freedom to refine and disseminate his theology, pursue controversies, and influence - in Saxony and beyond - those reforms of education, monasticism, the liturgy, preaching and ecclesial organisation in which he had been engaged since 1517. He died while visiting Eisleben in 1546.

2 Theology of humility

Luther was educated in a tradition of voluntaristic theology stemming from William of Ockham (§10) and refined by such thinkers as Gabriel Biel (see Voluntarism). They taught that God in his absolute power has freely ordained a covenant with human beings by which he promises to save those who ‘do what is in them’ (*facere quod in se est*). Biel, for instance, claimed that it is possible for human beings of their own natural powers to love God supremely, at least for an instant, and that God has covenanted to respond to this with his merciful acceptance.

Luther’s earliest theology, in his 1513-15 lectures on the Psalms, retained a *quod in se est* position, albeit one which had been significantly reinterpreted. For Luther, doing *quod in se est* is not a heroic moral achievement, but the acknowledgement of one’s sin, the recognition that one has no power to do good. This self-accusation and confession is the fitting disposition for the reception of God’s grace. God, according to the covenant that he has ordained, will save those who cast themselves upon his mercy in this way, even though they have done nothing which from a human perspective merits salvation ([1513–-15] 1974-6: 267, 373; compare [1515-16] 1972: 135). It was this view which led Luther to begin criticizing the sale of indulgences, which in his view did little to inculcate this humility, perhaps even militating against it (a view found throughout the *Ninety-five Theses*).

3 Justification by faith

From 1515 to 1518 or 1519, catalysed by reading some of Augustine of Hippo’s writings, Luther moved away from the theology of humility and developed an understanding of ‘justification by faith’ which was to remain largely unchanged for the rest of his life (see Augustine §§6, 13). He came to believe that the *quod in se est* position was wrong, and that sinners are passive with respect to their justification ([1525] 1972: 267). That is, the faith to which God responds by saving the sinner is itself a work of God in the sinner, and not the sinner’s achievement. The salvation of some and not others is related not to the fact that some and not others do *quod in se est*, but to God’s ineffable predestination (1517b: thesis 29; [1525] 1972: 292).

Central to this theology is an understanding that human beings are, apart from God’s active grace working in them, not free to choose to fulfil God’s law (1517b: theses 4, 5, 7, 16; 1518: thesis 13). The law which defines righteousness and unrighteousness is not one which could be fulfilled by external obedience; it is a spiritual law which requires a right orientation of the ‘inner person’ ([1522] 1960: 367). It is not that one becomes righteous by performing righteous acts, but that one can only perform truly righteous acts if one has already become righteous (1517b: thesis 40). Since the Fall, when Adam and Eve ate the forbidden fruit in Eden, human beings have been ‘turned in on themselves’ and require God’s intervention to break their self-reliance and turn them to him.

God awakens faith in the sinner by the power of his Word given in the Bible and communicated through preaching, and by the power of his Spirit at work in the sinner (1536a: thesis 28). The Word both condemns, leading sinners to despair of their own righteousness, and inculcates trust in the promises of God made evident in the life of Jesus Christ. In response to this faith that he himself has awoken, God ‘justifies’ the sinner: he refuses to impute sin to the sinner’s account. To put it positively, God decides to impute righteousness to sinners, to treat them as if they were righteous. God does this because of the intercession of his son, Jesus Christ, who in the Incarnation identified himself with sinners, and to whom sinners are joined in faith. Luther speaks of a marriage between sinners and Jesus, with God in his mercy choosing to regard the sinners as participating in Christ’s own righteousness. Faith is both the acceptance of this relationship to Christ and the refusal to attempt to stand outside this relationship when confronted by God (1536a).

In and of itself this justification does not remove actual sin from believers, but leaves them paradoxically ‘at one and the same time righteous and sinners’ - sinners according to what is in them, righteous according to God’s decision. Nevertheless, as he justifies believers, God gives them the gift of the Holy Spirit, who, in a secondary process, works in them to remove actual sin, and to produce righteousness within them by making them conform to Christ. This process is a continual one, and is not finished until the believer is taken into heaven ([1522] 1960: 370).

4 Ethics and politics

Faith is in one sense the fulfilment of the law. Luther called it the fulfilment of the First Commandment, which is to have no other gods before God ([1520a] 1966: 23-4), but stressed that faith itself is not a work but a gift from God. Whatever is done in faith is righteous, because faith by the power of God’s Holy Spirit naturally expresses

itself in love for God and for other people, and naturally works to remove the actual sin remaining in the believer ([1520a] 1966: 27). In this sense the believer is entirely free from all constraints. No law is needed for the believer ([1520a] 1966: 34).

However, true Christians are few and far between in this world, and God has ordained coercive secular government in order to restrain the excesses of sinners. Luther thus claimed that there are two ‘kingdoms’: one of Christians, in which there is complete freedom and love is the basis of action, and to which Jesus’ injunction to ‘turn the other cheek’ is appropriate; and another ‘of the world’, in which justice is the basis of action and punishment is necessary. Luther argued that obedience to secular authority is a necessary service on the part of Christians who value peace, even when that authority is deemed unjust (unless the authority tries to coerce a rejection of faith, in which case resistance in such internal matters is proper) (1523). Spiritual authorities, set up by the community of believers for its own good order, should exercise no secular or coercive power - and Luther thus raged against the peasants when they (to his understanding) tried to coerce spiritual changes in the Peasants’ Revolt of 1525. In his antagonism, he did little justice to their genuine grievances.

Nevertheless, in the absence of an accepted ecclesial authority after the break with Rome, Luther called upon Elector Frederick the Wise - who, although one believer among many, had certain resources - to help in establishing Luther’s reforms of the Church in Saxony. This development paved the way for the later establishment of a State Church.

5 Theological method: philosophy and revelation

According to Luther, philosophy is able to identify and describe things in their worldly relations. It can even partially identify God - as the source of the world’s ordering. Yet it can know nothing of God as gracious and merciful. Philosophy is therefore incompetent to describe anything in its full relationship to God, because God can only be properly identified through his self-identification in Jesus Christ. Even with regard to things in the world, then, philosophy is only partially adequate (see, for example, [1536b] 1960: 137-8).

God’s self-identification is grasped only in faith, since it is a self-identification ‘hidden’ in weakness and suffering. Only by the Holy Spirit’s illumination of the scriptural Word can this revelation be grasped. Faith then construes everything on the basis of the revelation of God in Christ (1518: thesis 20). The perspective of faith is dual, however, incorporating both a Law perspective and a Gospel perspective. The two are dialectically related. Law is the instrument of God’s ‘alien work’, by which he judges and condemns the world, and philosophy can contribute to an understanding of Law. Law is, however, wholly in the service of God’s ‘proper work’, about which philosophy knows nothing, which is to save the world by bringing it to faith and justifying it. God’s proper work is ‘hidden’ within his alien work, and both are shown forth (to the eyes of faith) in the cross of Christ, where God’s saving intention is hidden beneath suffering and condemnation ([1525] 1972: 140).

Only faith in Jesus Christ opens up this true dual perspective of Law and Gospel, which is the complete interpretive perspective for any worldly or spiritual reality, and which incorporates within it those insights of philosophy which it does not destroy. For instance, Luther was happy that Aristotle’s *Logic* should continue to be taught in the universities as long as the inability of syllogistic logic to deal with God’s proper work was acknowledged, but he railed against the use of Aristotle as a foundation for theology, and considered the *Ethics* in particular to be antithetical to the doctrine of justification by faith (1517b: thesis 41).

6 Theological method: biblical interpretation

Rather than philosophy, Luther put biblical exegesis at the heart of his project. In his lectures on the Psalms, he used a complex system which involved a distinction between historical and spiritual interpretation (the former relating the Psalms to ancient Israel and the Jews, the latter viewing them with the eyes of faith, and treating them as prophetic of God’s work in Jesus Christ), and combined that with the medieval scheme of the four senses of Scripture (literal, allegorical = doctrinal, tropological = moral, and anagogical = relating to the end-times). Luther rejected historical interpretation in favour of spiritual interpretation, and agreed with the medieval consensus in privileging the literal sense ([1513-15] 1974-6: 3-5).

As time went on, however, he simplified this system greatly, concentrating on a combined literal-tropological sense and spiritual interpretation, according to which the whole Bible is taken to have as its subject matter Christ’s

work for us, and our response ([1513-15] 1974-6: 402). More precisely, the Bible is both Law and Gospel (the former predominating in the Old Testament, the latter in the New, but both found in each), which correspond respectively to God's alien work of condemnation and humiliation and his proper work of justification ([1520b] 1957: 348). Those with eyes of faith (that is, with the knowledge and experience of what Christ has done for them, conveyed by the Holy Spirit) can see that the message of the whole Bible is that Christ died for sinners; this message is hidden in Law and revealed in Gospel (1521).

See also: Calvin, J.; Hus, J.; Justification, religious; Melanchthon, P.; Wyclif, J.

M.A. HIGTON

List of works

Luther, M. (1483-1546) *D. Martin Luthers Werke: Kritische Gesamtausgabe (Martin Luther's Works: Complete Critical Edition)*, Weimar: Hermann Böhlhaus Nachfolger, 1883-1986, 97 vols.(The most complete and most commonly cited of several original-language editions; of the 97 volumes, 12 are devoted to the *Deutsche Bibel (German Bible)*, 18 to *Briefwechsel (Letters)*, 6 to *Tischreden (Table Talk)*, and the remaining 61 to essays, lectures, treatises and other works.)

Luther, M. (1483-1546) *Luther's Works*, ed. J. Pelikan and H.T. Lehmann, St Louis, MO: Concordia, and Philadelphia, PA: Fortress/Muhlenberg, 1955-86, 55 vols.(A thorough selection, normally the best or only available English translations, with short introductions to each piece.)

Luther, M. (1513-15) *Dictata super Psalterium*, in *D. Martin Luthers Werke*, vols 3-4 and 55; trans. H.J.A. Bouman, *First Lectures on the Psalms*, in *Luther's Works*, vols 10 and 11, ed. H.C. Oswald, St Louis, MO: Concordia, 1974-6.(Luther's earliest lectures, exploring a theology of humility. The English translation does not contain Luther's glosses.)

Luther, M. (1515-16) *Divi Pauli apostoli ad Romanos Epistola*, in *D. Martin Luthers Werke*, vol. 56; trans. W.G. Tillmanns and J.A.O. Preuss, *Lectures on Romans: Glosses and Scholia*, in *Luther's Works*, vol. 25, ed. H.C. Oswald, St. Louis, MO: Concordia, 1972.(Easier to follow than the Psalms lectures, these mark the move towards a theology of justification by faith.)

Luther, M. (1517a) *Disputatio pro declaratione et vitutis indulgentiarum*, in *D. Martin Luthers Werke*, vol. 1; trans. C.M. Jacobs and H.J. Grimm, *Ninety-five Theses, or Disputation on the Power and Efficacy of Indulgences*, in *Luther's Works*, vol. 31, *Career of the Reformer I*, ed. H.J. Grimm, Philadelphia, PA: Fortress/Muhlenberg, 1957.(Began the controversy that led to Luther's break with Rome: a seminal document of the Reformation.)

Luther, M. (1517b) *Disputatio contra scholasticam theologiam*, in *D. Martin Luthers Werke*, vol. 1; trans. H.J. Grimm, *Disputation Against Scholastic Theology*, in *Luther's Works*, vol. 31, *Career of the Reformer I*, ed. H.J. Grimm, Philadelphia, PA: Fortress/Muhlenberg, 1957.(Luther's break with the late-medieval theology in which he was educated.)

Luther, M. (1518) *Disputatio Heidelbergae habita*, in *D. Martin Luthers Werke*, vol. 1; trans. H.J. Grimm, *Heidelberg Disputation*, in *Luther's Works*, vol. 31, *Career of the Reformer I*, ed. H.J. Grimm, Philadelphia, PA: Fortress/Muhlenberg, 1957.(A first programmatic statement of Luther's theology, contrasting the perspective of the cross with human perspectives.)

Luther, M. (1520a) *Von den guten Werken*, in *D. Martin Luthers Werke*, vol. 6; trans. W.A. Lambert and J. Atkinson, *A Treatise on Good Works*, in *Luther's Works*, vol. 44, *The Christian in Society I*, ed. J. Atkinson, Philadelphia, PA: Fortress/Muhlenberg, 1966. (Luther defends himself against the charge that he has rejected the doing of good works.)

Luther, M. (1520b) *Von der Freiheit eines Christenmenschen*, in *D. Martin Luthers Werke*, vol. 7; trans. W.A. Lambert and H.J. Grimm, *The Freedom of a Christian*, in *Luther's Works*, vol. 31, *Career of the Reformer I*, ed. H.J. Grimm, Philadelphia, PA: Fortress/Muhlenberg, 1957.(The best introduction to Luther's ethics.)

Luther, M. (1520c) *An den christlichen Adel deutscher Nation von des christlichen Standes Besserung*, in *D. Martin Luthers Werke*, vol. 6; trans. C.M. Jacobs and J. Atkinson, *To the Nobility of the German Nation Concerning the Reform of the Christian Estate*, in *Luther's Works*, vol. 44, *The Christian in Society I*, ed. J. Atkinson, Philadelphia, PA: Fortress/Muhlenberg, 1966.(One of several calls to reform that Luther issued at this time, this one informative on the relationship between secular and ecclesial power.)

Luther, M. (1521) *Ein klein Unterricht, was man in den Evangeliis suchen und gewarten soll*, in *D. Martin Luthers Werke*, vol. 10.1; trans. E.T. Bachmann, *A Brief Instruction on What to Look for and Expect in the Gospels*, in *Luther's Works*, vol. 35, *Word and Sacrament I*, ed. E.T. Bachmann, Philadelphia, PA: Fortress/Muhlenberg, 1960.(A simple introduction to the structure of Luther's hermeneutics.)

Luther, M. (1522) 'Vorrede auff die Epistel S. Paul an die Römer', in *D. Martin Luthers Werke*, *Deutsche Bibel*, vol. 7; trans. E.T. Bachmann, 'Preface to the Epistle of St Paul to the Romans', in *Luther's Works*, vol. 35, *Word and Sacrament I*, ed. E.T. Bachmann, Philadelphia, PA: Fortress/Muhlenberg, 1960.(An influential summary of Luther's understanding of justification by faith.)

Luther, M. (1523) *Von weltlicher Oberkeit, wie weit man ihr Gehorsam schuldig*, in *D. Martin Luthers Werke*, vol. 11; trans. H. Höpel, *On Secular Authority, How Far it Must be Obeyed*, in *Luther and Calvin on Secular Authority*, ed. H. Höpel, Cambridge: Cambridge University Press, 1991.(Luther's 'two kingdoms' doctrine of secular and spiritual power.)

Luther, M. (1525) *De servo arbitrio*, in *D. Martin Luthers Werke*, vol. 18; trans. P.S. Watson and B. Drewery, *The Bondage of the Will*, in *Luther's Works*, vol. 33, *Career of the Reformer III: The Bondage of the Will*, ed. P.S. Watson, Philadelphia, PA: Fortress/Muhlenberg, 1972.(Luther's longest exposition of his understanding of humanity's lack of freedom.)

Luther, M. (1536a) *Die Disputation 'de iustificatione'*, in *D. Martin Luthers Werke*, vol. 39.1; trans. L.W. Spitz, *The Disputation Concerning Justification*, in *Luther's Works*, vol. 34, *Career of the Reformer IV*, ed. L.W. Spitz, Philadelphia, PA: Fortress/Muhlenberg, 1960.(A clearly defined, mature presentation of the doctrine of justification.)

Luther, M. (1536b) *Die Disputation 'de homine'*, in *D. Martin Luthers Werke*, vol. 39.1; trans. L.W. Spitz, *The Disputation Concerning Man*, in *Luther's Works*, vol. 34, *Career of the Reformer IV*, ed. L.W. Spitz, Philadelphia, PA: Fortress/Muhlenberg, 1960.(A clear presentation of Luther's theological anthropology, and his attitude towards philosophy.)

References and further reading

Aland, K., Reichert, E.O. and Jordan, G. (1970) *Hilfsbuch zum Lutherstudium (Guide to Luther Studies)*, Witten/Ruhr: Luther-Verlag.(A list of all Luther's works, and an introduction to all major editions.)

Althaus, P. (1963) *Die Theologie Martin Luthers*, Gütersloh: Gütersloher Verlag; trans. R.C. Schultz, *The Theology of Martin Luther*, Philadelphia, PA: Fortress Press, 1966.(Dated, but very clear and comprehensive.)

Bainton, R.H. (1950) *Here I Stand: A Life of Martin Luther*, New York: Abingdon-Cokesbury Press.(Still the most readable single-volume life of Luther, dated and hagiographic but basically reliable.)

Brecht, M. (1981-7) *Martin Luther*, vol. 1, *Sein Weg zur Reformation 1483-1521*, vol. 2, *Ordnung und Abgrenzung der Reformation 1521-1532*, vol. 3, *Die Erhaltung der Kirche 1532-1545*, Stuttgart: Calwer Verlag; trans. J.L. Schaaf, *Martin Luther*, vol. 1, *His Road to Reformation 1483-1521*, vol. 2, *Shaping and Defining the Reformation 1521-1532*, Minneapolis, MN: Fortress Press, 1985-90.(The best detailed biography of Luther around, although the translation makes it heavy going.)

Cargill Thompson, W.D.J. (1984) *The Political Thought of Martin Luther*, ed. P. Broadhead, Brighton: Harvester Press.(A thorough survey.)

Dalferth, I.U. (1988) 'The Law-Gospel Model', in *Theology and Philosophy*, Oxford: Blackwell.(An excellent and very clear account of the relationship of theology and reason in Luther.)

Lohse, B. (1980) *Martin Luther - Eine Einfuhrung in sein Leben und sein Werk*, Munich: Verlag C.H. Beck; trans. R.C. Schultz, *Martin Luther: An Introduction to his Life and Work*, Edinburgh: T. & T. Clark, 1987.(The place to begin in studying Luther: an introduction to contemporary and historical debates, to Luther's life and work, and a survey of the primary texts.)

McGrath, A.E. (1985) *Luther's Theology of the Cross: Martin Luther's Theological Breakthrough*, Oxford: Blackwell, 1985.(Good on the medieval background, and on the stages of Luther's development.)

McSorely, H.J. (1969) *Luther: Right or Wrong? An Ecumenical-Theological Study of Luther's Major Work, The Bondage of the Will*, New York: Newman Press, and Minneapolis, MN: Augsburg Publishing House.(A Roman Catholic defence of the orthodoxy and coherence of Luther's denial of the freedom of the will, if not of all his arguments.)

Oberman, H.A. (1986) *The Dawn of the Reformation: Essays in Late Medieval and Early Reformation Thought*, Edinburgh: T. & T. Clark.(Important and careful articles trying to sort out the precise nature and role of Luther's innovations and contributions.)

Steinmetz, D.C. (1980) *Luther and Staupitz: An Essay in the Intellectual Origins of the Protestant Reformation*, Durham, NV: Duke University Press.(Contains a sensitive account of Luther's theology of humility.)

Steinmetz, D.C. (1986) *Luther in Context*, Bloomington, IN: Indiana University Press.(A good selection of articles, including on the hidden God and free will.)

Luxemburg, Rosa (1871-1919)

Rosa Luxemburg, of Polish-Jewish origins, was for most of her life a prominent activist and theorist on the radical left of the German Social Democratic Party. She defended revolutionary Marxism against the 'revisionist' critique ofEduard Bernstein; she developed an original and controversial Marxist theory of imperialism; and she advocated direct revolutionary action by the masses, as contrasted with Lenin's insistence on 'democratic centralism' and the leading role of the Party.

1 Beginnings

Rosa Luxemburg was born in Zamość near Lublin in eastern Poland, then part of Russia. Her father was a merchant of Jewish extraction. Two years later the family moved to Warsaw where she spent her childhood. While still at school she engaged in illegal political activities and, in 1889, had to flee to Zurich, where she attended the university. When, shortly afterwards, the Social Democratic Workers Party of Poland was founded, she joined and soon became a leading member. From the beginning, she took the view that, with the emergence of the world market, nationalism had become regressive and dangerously disruptive and that, in particular, the growing integration of the Polish and Russian economies had rendered the demand for Polish independence (advanced by many Polish radicals) redundant. This launched a controversy in which she crossed swords with, among others, Kautsky and Liebknecht. In 1897 she submitted her doctoral dissertation, *Die industrielle Entwicklung Polens (The Industrial Development of Poland)*, still regarded by many as a seminal work. A year later she participated in a marriage of convenience with Gustav Lübeck, a German citizen. This gave her the right to live in Germany and enabled her to take part in the activities of the German Social Democratic Party (SPD).

She lost no time in seizing the opportunity. The 'revisionist' debate was in full spate and, in the summer of 1898, she went to Berlin where she wrote a comprehensive refutation of Bernstein's position (see Bernstein, E.). The result appeared as a series of articles in the *Leipziger Volkszeitung*. In these she argued that transnational cartels, sophisticated credit systems and improved means of communication, far from providing the 'means of adaptation' enabling the capitalist system to soften the impact of economic crises, merely enhanced the system's self-destructive tendencies and hastened the day of its inevitable economic collapse. She agreed that social reforms had a place in the party's programme, but only in so far as they raised the revolutionary consciousness and combat readiness of the working class. Finally, she suggested that Bernstein's basic position was that of philosophical idealism, and that this meant that he had defected to the camp of the reformist *petit bourgeoisie*. It was a case presented with remarkable intellectual acumen. Bernstein himself described her arguments as 'true examples of false dialectics, but handled at the same time with great talent' - a rare compliment. Her collected articles against Bernstein were published as a book, *Sozialreform oder Revolution? (Social Reform or Revolution?)*, in 1899.

In the following years, she consolidated her position in the German socialist movement as a leading theorist and agitator, becoming associated in particular with Kautsky and Franz Mehring. In 1904 she added to her credentials by spending three months in prison for slandering the crown (*Majestätsbeleidigung*); and in the same year, she wrote two substantial articles criticizing Lenin's democratic centralism as applied to party organization (see Lenin, V.I.). In these, as in all her subsequent writings on party organization and tactics, she took a strong stand on a principle Marx and Engels had always insisted upon; namely, that the liberation of the working class can be achieved only by the working class itself. In her view, this meant that the revolution had to be a matter of direct spontaneous action by the workers and that the party as such should act in an enabling rather than in a directing role. 'Historically', she declared, 'the errors committed by a truly revolutionary movement are infinitely more fruitful than the infallibility of the cleverest Central Committee' ([1903/4] 1961: 108).

This view was reinforced when, in 1905, she went to Warsaw to take part in the revolutionary upheavals which followed Russia's defeat in its war with Japan. Her experiences inspired one of her best political works, *Gewerkschaftskamp und Massenstreik (The Mass Strike, The Political Party and The Trade Unions)* (1906). Here she elaborated her 'spontaneity theory' and argued that democratic proletarian organizations are not the precondition of revolutionary action but rather the consequence of it.

2 The analysis of capitalism

Luxemburg's most important contribution to Marxist theory was her *Die Akkumulation des Kapitals (The Accumulation of Capital)*, published in 1913. In it, she pointed out that a capitalist economy can expand only if it can realize surplus value by finding purchasers for the goods it produces. Since the workers in a capitalist economy receive only the value of their labour power, but not the surplus value which they produce, they are unable to repurchase their whole product, thus leaving a surplus product. If the capitalists themselves purchase the surplus product, they gain nothing. No capital is accumulated. The surplus product must therefore find a market beyond the boundaries of capitalism, that is to say, in the non-industrial or pre-capitalist world. Hence the competition for markets in the peasant economies of Asia and Africa, and hence the struggle for political control of those markets - a phenomenon known as imperialism. When, in due course, all non-capitalist areas of the world market have been conquered and absorbed into the capitalist economy, the accumulation of capital will, of necessity, come to a halt and capitalism itself will collapse.

Rosa Luxemburg's analysis carried three important implications. First, it vindicated the claim that the terminal crisis of capitalism was inevitable. Second, it meant that imperialism was, not (as Lenin, for instance, would have it) the 'final stage of capitalism', but a structural feature of capitalism as such. Finally, it meant that war on a worldwide scale between capitalist states was unavoidable.

3 Last years

At the outbreak of war, Luxemburg was among the few who resisted the wave of patriotic war fever which swept the German SPD. In 1916 she was imprisoned and remained so for the rest of the war. Her last significant work, written while she was still in prison and published posthumously, was *Die russische Revolution (The Russian Revolution)* (1922). Once again she voiced her doubts about Lenin's notion of party organization and tactics. His 'Jacobin' style of organization might have been appropriate in the bourgeois revolution, but (she argued) it could not be employed in the very different circumstances of the proletarian revolution. The danger was that the dictatorship of the proletariat could easily be replaced by the dictatorship of a party or clique. For her, the dictatorship of the proletariat implied direct rule by the proletariat and therefore the active participation of the mass of the people. It was, she insisted, no paradox to think of the dictatorship of the proletariat as taking the form of 'unrestricted democracy'.

Upon her release from prison at the end of the war, she joined the leadership of the radical Spartacus group. Following the ill-timed and unsuccessful Spartacist insurrection in 1919, she was arrested and, together with Karl Liebknecht, murdered on her way to prison.

See also: Revolution; Marxism, Western

H. TUDOR

List of works

Luxemburg, R. (1898) *Die industrielle Entwicklung Polens (The Industrial Development of Poland)*, Leipzig: Duncker & Humblot.(Her first serious contribution to Marxist socio-economic analysis.)

Luxemburg, R. (1899) 'Organisationsfragen der russischen Sozialdemokraten' *Die Neue Zeit* 2: 484-92 and 529-35; repr. in *Iskra* 69, 10 July 1904; *Sozialreform oder Revolution*, Leipzig: G. Heinisch; repr. *Social Reform or Revolution?*, New York: Three Arrows, 1937; trans. and ed. D. Howard, *Selected Political Writings of Rosa Luxemburg*, New York and London: Monthly Review Press, 1971.(Her famous critique of Edward Bernstein's revisionism.)

Luxemburg, R. (1906) *Gewerkschaftskamp und Massenstreik (The Mass Strike, The Political Party and the Trade Unions)*, New York: Harper & Row, 1971, 52-135.(Luxemburg's reassessment of social democratic tactics following the upheavals in Russia in 1905.)

Luxemburg, R. (1913) *Die Akkumulation des Kapitals (The Accumulation of Capital)*, London: Routledge, 1951. (Her main work on imperialism and its relation to capitalist development.)

Luxemburg, R. (1903/4) trans. and ed. with intro. by B.D. Wolfe, 'Leninism or Marxism?', in *The Russian Revolution and Leninism or Marxism?*, Ann Arbor, MI: University of Michigan Press, 1961.(Her critique of Leninism and the direction the Bolshevik revolution was taking.)

Luxemburg, R. (1922) *Die russische Revolution: Eine kritische Wurdigung*, ed. with intro. by P. Levi, Berlin: Verlag Gesellschaft; trans. and ed. with intro. by B.D. Wolfe, 'Leninism or Marxism?', in *The Russian Revolution and Leninism or Marxism?*, Ann Arbor, MI: University of Michigan Press, 1961.(A critique of the

direction taken by the Bolshevik revolution in Russia: her last significant work.)
Luxemburg, R. (1971) *Selected Political Writings of Rosa Luxemburg*, ed. D. Howard, New York and London: Monthly Review Press.(A useful selection of her main writings.)

References and further reading

Bronner, S.E. (1987) *Rosa Luxemburg: A Revolutionary for our Times*, New York: Monthly Review Press.(A sympathetic but not uncritical account.)
Frohlich, P. (1972) *Rosa Luxemburg*, New York: Monthly Review Press.(Well-known work, originally published for the Left Book Club.)
Geras, N. (1976) *The Legacy of Rosa Luxemburg*, London: Humanities Press.(A subtle and perceptive analysis.)
Nettl, J.P. (1966) *Rosa Luxemburg*, London: Oxford University Press.(An excellent and, indeed, authoritative intellectual and political biography.)

Lyotard, Jean-François (1924-)

Jean-François Lyotard is a prominent French philosopher who is generally considered the leading theorist of postmodernism. His work constitutes an insistent critique of philosophical closure, historical totalization and political dogmatism and a re-evaluation of the nature of ethics, aesthetics and politics after the demise of totalizing metatheories.

In his early works, Lyotard confronts the limitations of dialectical philosophy and structuralist linguistics and analyses the disruptive, extradiscursive force of desire and the nonrepresentational or figurative dimensions of art and literature. In La Condition postmoderne (1979a) (The Postmodern Condition, 1984), he treats narrative pragmatics and language games as the bases for a critical approach to postmodern art and politics, as well as to the problem of justice. Recent texts insist on the obligation of philosophy, politics and writing to bear witness to heterogeneity and to what is repressed or forgotten in all representations of the past. His work questions the limits of philosophy, aesthetics and political theory in terms of problems linked to the irreducible complexities of art and literature and the nonrepresentational affects of historical-political events.

1 Phenomenology, structuralism and the question of art

Jean-François Lyotard was born in Versailles in 1924 and educated in Paris. After passing the 'Agrégation' in philosophy in 1950, he taught for one year in a lycée in Constantine, Algeria, and for seven years in La Flèche, France. He was a member of the radical political group 'Socialism or Barbarism' from 1954 to 1964 and regularly contributed articles to the journal it published and later to *Pouvoir Ouvrier* (*Worker Power*). He has taught at the Sorbonne, at Nanterre, where he played an important role in 'the events of May '68', and at the University of Paris at Vincennes and then Saint-Denis, until his retirement in France in 1987. He was a founding member of the International College of Philosophy in Paris and its second director. He has also been associated with various universities in the United States, including the University of California, San Diego, Johns Hopkins University, Emory University and the University of California, Irvine, where for eight years he was Professor of French and Critical Theory. He has also been a visiting professor in Germany.

Lyotard's first works offer critical perspectives on phenomenology, structuralism, Hegelian dialectics and Marxist political theory. In *La Phénoménologie* (1954) (*Phenomenology*, 1991), which reveals the influence of both Heidegger and Merleau-Ponty on his early thinking, Lyotard analyses phenomenology's radical notion of historicity and the possibilities phenomenology offers for uncovering and moving beyond the limitations of objectivism, subjectivism and idealism.

In *Discours, figure* (1971), phenomenology, Marxism and structuralism are treated as ideological restrictions or deformations of the forces and constructs that underlie ordered reality and thought. Lyotard proposes a deconstructive approach to aesthetics to open the way for new forms of political practice (see Deconstruction). He insists on the radical differences between the linguistic or philosophical determination of meaning and the pictorial, spatial displacement of meaning in painting and experimental forms of poetry. He criticizes philosophy for the way it has attempted to master the sensible by making it discursive or meaningful and thus transform art into a form of philosophy.

Lyotard attacks the structuralist model for language and especially the Lacanian interpretation of Freud as serious restrictions and repressions of desire. He insists on the transgressive effects of desire, whose traces are manifest in the productive-disruptive *force* and not the completed *form* or structure of the figure. The exteriority and resistance of the figural to the discursive are for Lyotard also signs of the resistance of historical-political events to philosophical or historical totalizations. A libidinal aesthetics points to the possibility of a libidinal politics.

Économie libidinale (1974) (*Libidinal Economy*, 1993) attacks the dogmatism of Marxist theory in an attempt to uncover the disruptive traces of libidinal drives that resist being systematized within the dialectic and that are signs of the possibility of alternate forms of political practice. What Lyotard himself calls 'the metaphysics of desire' at the base of his 'libidinal works' produces a humorous and at times outrageous satire of Marx and other political theorists and dramatically indicates the limitations of theory in general and its inability to foster or account for radical, unprogrammed actions or creations.

2 Postmodernism, narratology and critical judgment

In *La Condition postmoderne* (1979a) (*The Postmodern Condition*, 1984), Lyotard defines the postmodern as an era characterized by its incredulity concerning the ability of foundational, totalizing metanarratives - whether liberal or Marxist - to effectively legitimate the sciences and the arts. Lyotard argues that the self-legitimating pragmatics of ‘little narratives’ break with the forms of philosophical and political totalization characteristic of modernism. He defends a concept of language games he derives from Wittgenstein and the principle that the specificity of the rules governing each game must be respected and the different games not confused with or derived from each other. Postmodernism is presented as the basis for a politics rooted in heterogeneity and dissension which ‘would respect both the desire for justice and the desire for the unknown’ (1984: 67).

The central problem posed by Lyotard’s later work is how to create links among the heterogeneous realms and discourses constituting postmodernity without negating their specificity. His innovative readings of Kant bring to light the radical, postmodern implications of Kant’s refusal to derive critical judgment from the cognitive faculty (see Kant, I. §12). Lyotard argues that the form of universalization related to critical (reflective) judgment can never be more than a demand and that Kant’s notion of a universal *sensus communis* does not therefore imply consensus and a predetermined notion of community, as it does for communicational theorists such as Habermas and Rorty, but ‘dissensus’ and an open, conflictual, always undetermined community.

Lyotard argues that the problem of judgment should be considered essential to both the aesthetic and political realms, because it is in them that representation, conceptual knowledge and ethical determination are confronted with their own limitations. The derivation of either realm from any concept or predetermined system constitutes the form of aesthetic and political dogmatism against which Lyotard’s entire critical enterprise is aimed.

3 The *différend* and the unrepresentable in history and memory

In *Le Différend* (1983) (*The Differend*, 1988), Lyotard defines the political in terms of the conflict of fundamental ‘differends’ (disputes), which, unlike forms of litigation that can be judged according to a determined set of rules or laws, cannot be equitably resolved for lack of a universal rule of judgment applicable to all of the arguments. He makes the problem of the linkage between conflicting idioms, phrases and faculties the primary obligation of critical thinking and of a politics rooted in diversity and disputation. To be judicious in terms of differends is to acknowledge the heterogeneity of the discursive universe and the incommensurability of different genres and idioms. It is to bear witness to the injustice occurring whenever ‘a plaintiff is deprived of the means of arguing and by this fact becomes a victim’ (1988: 24).

In *L’Enthousiasme: la critique kantienne de l’histoire (Enthusiasm: The Kantian Critique of History)* (1986a) and *Leçons sur l’analytique du sublime* (1991) (*Lessons on the Analytic of the Sublime*, 1994), Lyotard develops further the postmodern implications of Kant’s *Critique of Judgment* and particularly his concept of the sublime. He argues that the Kantian sublime makes possible a critical approach to history based on the recognition of the limitations of presentation and a respect for the fundamental and irreducible heterogeneity of the sociopolitical realm.

In *The Differend* and other recent works, Lyotard appropriates Adorno’s analysis of the status of art and culture ‘After Auschwitz’ (in *Negative Dialektik*, 1966) and emphasizes what he calls the agony of art and its responsibilities to memory (see Adorno, T. §5). In *Heidegger et ‘les juifs’* (1988a) (*Heidegger and ‘the Jews’*, 1990), which is Lyotard’s contribution to the debate in France over the philosophical and political implications of Heidegger’s relation to National Socialism and of his silence on the subject of the Shoah, Lyotard argues that writing and philosophy have a fundamental obligation both to those who have been excluded from or marginalized in history and to ‘the Forgotten’ or what has not and cannot be represented in thought, history and memory.

It would be possible to connect different texts of Lyotard to the later works of the phenomenologist Merleau-Ponty, to Levinas’s radical reformulation of the ethical, to Adorno’s work on aesthetics and politics, and finally to the work of other contemporary French philosophers such as Deleuze and Derrida. Lyotard’s treatment of the links between aesthetics and politics, his provocative analyses of Marx, Freud, postmodernism and the Kantian notion of the sublime, as well as what Derrida has called ‘the categorical challenge’ he has made to his contemporaries to make the problem of judgment a primary concern (*La Faculté de juger*,1985), all give his work a unique place within contemporary French critical philosophy.

See also: Postmodernism; Post-structuralism in the social sciences §§4-5

DAVID CARROLL

List of works

Lyotard, J.-F. (1954) *La Phénoménologie*, Paris: Presses Universitaires de France, Collection 'Que sais-je?'; trans. B. Beakley, *Phenomenology*, Albany, NY: State University of New York Press, 1991.(A presentation and critical analysis of phenomenology's notion of historicity and its critique of objectivism, subjectivism and idealism.)

Lyotard, J.-F. (1971) *Discours, figure (Discourse, Figure)*, Paris: Klincksieck.(A radical deconstructive approach to aesthetics which criticizes the structuralist model for language and especially the Lacanian interpretation of Freud as serious restrictions and repressions of desire.)

Lyotard, J.-F. (1973a) *Dérive à partir de Marx et Freud*, Paris: Union Générale des Éditions, Collection '10/18'; trans. S. Hanson, A. Knab, R. Lockwood, J. Maierand R. McKeon, *Driftworks*, New York: Semiotext(e), 1984. (The translation contains five chapters from *Dérive* and other essays dealing with Marx, Freud and the politics of the student movement of May '68 that attack the dogmatic, systematic restrictions of different versions of Marxism and psychoanalysis in the name of disruptive libidinal drives and politics.)

Lyotard, J.-F. (1973b) *Des Dispositifs pulsionnels (Mechanisms of the Drives)*, Paris: Union Générale des Éditions, Collection '10/18'.(A collection of essays on art, music, cinema, politics, Adorno and Nietzsche which demonstrate the subversive effects of libidinal politics.)

Lyotard, J.-F. (1974) *Économie libidinale*, Paris: Éditions de Minuit; trans. I. Hamilton Grant, *Libidinal Economy*, London: Athlone, 1993.(A humorous and at times outrageous satire of the dogmatism of Marxist theory which attempts to uncover the disruptive traces of libidinal drives that resist being systematized within the dialectic and that are signs of the possibility of alternate forms of political practice.)

Lyotard, J.-F. (1979a) *La Condition postmoderne: rapport sur le savoir*, Paris: Éditions de Minuit; trans. G. Bennington and B. Massumi, *The Postmodern Condition: A Report on Knowledge*, Minneapolis, MN: University of Minnesota Press.(A critical approach to the conditions of knowledge which defends the self-legitimating pragmatics of 'little narratives' in a postmodern era when the foundational, totalizing metanarratives used to legitimate the sciences and the arts have collapsed and are no longer credible.)

Lyotard, J.-F. with Thébaud, J.-L. (1979b) *Au juste: conversations*, Paris: Christian Bourgois; trans. W. Godzich, *Just Gaming*, Minneapolis, MN: University of Minnesota Press, 1985.(A series of 'conversations' which deal with topics such as pagan philosophy, Plato, Kant and Wittgenstein and stress the fundamental importance of dissension and a critical notion of justice in an era when there exist no universally acknowledged criteria to determine judgment.)

Lyotard, J.-F. (1983) *Le Différend*, Paris: Éditions de Minuit; trans. G. Van Den Abbeele, *The Differend: Phrases in Dispute*, Minneapolis, MN: University of Minnesota Press, 1988.(In this work Lyotard defines the political in terms of the conflict of fundamental 'differends' (disputes), which cannot be equitably resolved for lack of a universal rule of judgment applicable to all of the arguments, and treats the problem of the linkage between conflicting idioms, phrases and faculties as the primary obligation of critical thinking and of a politics rooted in diversity and disputation.)

Lyotard, J.-F. (1985) *La Faculté de juger (The Faculty of Judging)*, with J. Derrida, V. Descombes, G. Kortian, P. Lacoue-Labarthe, and J.-L. Nancy, Paris: Éditions de Minuit; Lacoue-Labarthe's essay trans. as 'Talks', in *The Work of Jean-François Lyotard*, special issue of *Diacritics* 14 (3), 1984.(A collection of essays having to do with Lyotard's work and the problem of judgment which were chosen from papers given at the Cérisy-la-Salle Colloquium on Lyotard of July-August 1982 entitled 'Comment juger?' (How to Judge?).)

Lyotard, J.-F. (1986a) *L'Enthousiasme: la critique kantienne de l'histoire (Enthusiasm: The Kantian Critique of History)*, Paris: Éditions Galilée.(A critical reading of the significance of Kant's notion of the sublime in the *Critique of Judgment* and in his texts on history which propose a notion of politics and justice that cannot be derived from knowledge or any predetermined idea.)

Lyotard, J.-F. (1986b) *Le Postmoderne expliqué aux enfants: correspondance, 1982-1985*, Paris: Éditions Galilée; trans. and ed. J. Pefanis and M. Thomas, *The Postmodern Explained: Correspondence, 1982-1985*, Minneapolis, MN: University of Minnesota Press, 1992.(A collection of short essays in which Lyotard defends his notion of the postmodern in art and politics.)

Lyotard, J.-F. (1988a) *Heidegger et 'les juifs'*, Paris: Éditions Galilée; trans. A.. Michel and M.S. Roberts, *Heidegger and 'the Jews'*, Minneapolis, MN: University of Minnesota Press, 1990.(Lyotard's contribution to the debate in France over the philosophical and political implications of Heidegger's relation to National Socialism and of his silence on the subject of the Shoah in which Lyotard argues that writing and philosophy

have a fundamental obligation to ‘the Forgotten’ or what has not and cannot be represented in thought, history and memory.)

Lyotard, J.-F. (1988b) *Peregrinations: Law, Form, Event*, New York: Columbia University Press.(A series of lectures given at the University of California, Irvine in May 1986 in which Lyotard describes his own intellectual formation and the development of his approach to ethics, politics, art and history. These essays are followed by ‘A Memorial of Marxism’, which analyses his own ‘*différend*’. Includes an excellent bibliography of texts by and about Lyotard compiled by E. Yeghiayan.)

Lyotard, J.-F. (1989a) *La Guerre des Algériens: écrits, 1956-1963 (The Algerians’ War: Writings, 1956-1963),* Paris: Éditions Galilée; 11 of 13 chapters trans. in *Political Writings*, ed. B. Readings, Minneapolis, MN: University of Minnesota Press. (A collection of essays written by Lyotard on the Algerian War for the radical Marxist journal *Socialisme ou barbarie* and preceded by a short preface written in 1989.)

Lyotard, J.-F. (1989b) *The Lyotard Reader*, ed. A. Benjamin, Cambridge, MA: Blackwell.(A representative selection of chapters from Lyotard’s books and other essays on art, literature and politics.)

Lyotard, J.-F. (1991) *Leçons sur l’analytique du sublime*, Paris: Éditions Galilée; trans. E. Rottenberg, *Lessons on the Analytic of the Sublime,* Stanford, CA: Stanford University Press, 1994.(A rigorous analysis of the Analytic of the Sublime in Kant’s *Critique of Judgment* which deals with the limits of presentation in philosophy, art and politics and emphasizes the necessity of reflexive thinking.)

Lyotard, J.-F. (1993a) *Political Writings*, ed. B. Readings, Minneapolis, MN: University of Minnesota Press. (Contains essays from *Tomb of the Intellectual and Other Papers* (1984), *The Algerians’ War* (1989) and recent essays from various sources.)

Lyotard, J.-F. (1993b) *Toward the Postmodern / Lyotard*, ed. R. Harvey and M.S. Roberts, Atlantic Highlands NJ: Humanities Press.(Contains essays from *Dérive à partir de Marx et Freud* and *Des dispositifs pulsionnels,* as well as more recent essays on art and literature.)

Lyotard, J.-F. (1996) *Signé Malraux*, Paris: Grasset.(A biography of André Malraux.)

References and further reading

Adorno, T. (1966) *Negative Dialektik*, Frankfurt: Suhrkamp; trans. E.B. Ashton, *Negative Dialectics*, New York: Continuum, 1983.(The last section, ‘Meditations on Metaphysics’, which includes ‘After Auschwitz’, is an important reference for Lyotard’s recent thinking on art and culture.)

L’Arc (1976) *Lyotard*, special issue of *L’Arc* 64.(A collection of essays in French on the aesthetic and psychoanalytical implications of Lyotard’s early work.)

Benjamin, A. (ed.) (1992) *Judging Lyotard*, London and New York: Routledge.(A good collection of essays from different critical perspectives focused on the implications of Lyotard’s approach to postmodernity, of his differences with Habermas and of his approach to aesthetic and political judgment.)

Bennington, G. (1988) *Lyotard: Writing the Event,* New York: Columbia University Press.(A good introduction and insightful analysis of three of Lyotard’s most important works, *Discourse, Figure*, *Libidinal Economics* and *The Differend*. Special emphasis is given to Lyotard’s affinities with Derrida.)

Carroll, D. (1987) *Paraesthetics: Foucault, Lyotard, Derrida*, London and New York: Methuen.(A critical analysis of Lyotard’s strategic use of art and literature to indicate the limits of politics and theory. It focuses on Lyotard’s recasting of politics in terms of the problem of critical judgment and an aesthetics of the sublime.)

Carroll, D. (1990) ‘The Memory of Devastation and the Responsibilities of Thought: "And let’s not talk about that"’, foreword to *Heidegger and ‘the Jews’*.(An analysis of Lyotard’s philosophical and political critique of Heidegger and of his own paradoxical relation to deconstruction.)

Diacritics (1984) *The Work of Jean-François Lyotard*, special issue of *Diacritics* 14 (3).(A collection of essays dealing with issues raised by *The Postmodern Condition*, *Just Gaming* and *The Differend*. Includes an interview with Lyotard.)

La Faculté de juger (The Faculty of Judging). (1985) Paris: Éditions de Minuit.(Collection of essays from the Cérisy-la-Salle Colloquium of July-August 1982 which was devoted to Lyotard.)

Ingram, D. (1987/8) ‘Legitimacy and the Postmodern Condition: The Political Thought of Jean-François Lyotard’, *Praxis International* 7 3-4.(A good presentation and analysis of the implications of the differences in Lyotard’s approach to legitimation in *The Postmodern Condition* and *The Differend*.)

Jameson, F. (1984) ’Foreword’ to *The Postmodern Condition: A Report on Knowledge*.(A presentation of this work from a critical Marxist position.)

Lingis, A. (1994) ‘Some Questions About Lyotard’s Postmodern Legitimation Narrative’, *Philosophy & Social Criticism* 20 (1-2).(A critique of Lyotard’s postmodern notion of narrative and performance as the bases for legitimation.)

Nordquist, J. (1991) *Jean-François Lyotard: A Bibliography*, Santa Cruz, CA: Reference and Research Services, 'Social Theory: A Bibliographic Series', No. 21.(A good bibliography of books and articles by and about Lyotard, listed by topic.)

Pefanis, J. (1991) *Heterology and the Postmodern: Bataille, Baudrillard and Lyotard*, Durham, NC: Duke University Press.(An original presentation of *Libidinal Economy* and a defence of Lyotard's controversial treatment of Marx and his politics and ethics in general.)

Raffel, S. (1992) *Habermas, Lyotard, and the Concept of Justice*, London: Macmillan.(A critique of both Habermas and Lyotard in terms of an alternative idea of justice. The analysis of Lyotard focuses exclusively on *The Postmodern Condition*.)

Rorty, R. (1985) 'Habermas and Lyotard on Post-Modernity', *PRAXIS INTERNATIONAL* 4 (1). Republished in R. Bernstein (ed.) *Habermas and Modernity*, London: Polity Press, 1985, and in R. Rorty, *Essays on Heidegger and Others*, Cambridge: Cambridge University Press.(An analysis of the differences between Habermas and Lyotard from a liberal, communicational perspective which focuses on the problem of consensus and its implications for democracy.)

Rorty, R. (1991) 'Cosmopolitanism without Emancipation: A Response to Jean-François Lyotard', in *Objectivity, Relativism, and Truth*, Cambridge: Cambridge University Press; originally appeared in *Critique* 41 (456), 1985, trans. P. Saint-Amand.(A critique of Lyotard's reading of Wittgenstein and of his notion of differends from a liberal, pragmatist position that defends the notion of universal history.)

Sim, S. (1992) *Beyond Aesthetics: Confrontations with Poststructuralism and Postmodernism*, Toronto, Ont.: University of Toronto Press.(A critique of postmodernism and post-structuralism from a rigid socialist-materialist position. It contains three chapters on Lyotard, the first two highly critical of what is called his 'neo-conservative' anti-foundationalist position, the last one more positive about the political implications of his pragmatics and agonistics.)

Tomiche, A. (1994) 'Rephrasing the Freudian Unconscious: Lyotard's Affect-Phrase', *Diacritics* 24 (1).(A clear and insightful analysis of Lyotard's recent work on Freud and the problem of the affect.)

Weber, S. (1985) 'Literature - Just Making It', an Afterword to *Just Gaming*, trans. W. Godzich, Minneapolis, MN: University of Minnesota Press.(A critical analysis of the prescriptive dimensions of Lyotard's use of language games and the postmodern notion of justice he advocates.)

Yeghiayan, E. (1988) 'Checklist of Writings By and About Jean-François Lyotard: A Selected Bibliography', in *Peregrinations: Law, Form, Event*, New York: Columbia University Press.(Most complete bibliography of works by and on Lyotard through 1987.)

Lévi-Strauss, Claude (1908-)

Lévi-Strauss is one of the outstanding figures of mid-twentieth century intellectual life, influential far beyond the boundaries of France or the French language, and he has continued to write new work well into his eighties. His name is linked above all to the structuralist movement, of which he has been probably the most single-minded and unwavering exponent, and he was one of the key figures in the experiment of applying the insights of linguistics to the material of the social sciences. Through his work, public recognition of the discipline of anthropology grew dramatically to become an important element in discussion of philosophical issues.

The philosophical environment in France, in which structuralism developed and against which it reacted, was that of Sartrean humanism. In contrast to the existentialist emphasis on individual subjectivity, structuralism expected to find objective solutions to problems in the study of human beings. It was a form of intellectual modernism, a radical break with previous theoretical models and philosophical traditions, symptomatic of post-war optimism for the global applicability of science. It was hostile to metaphysics, bracketed the search for truth and was indifferent to the human subject. And, in contrast to Bergson's emphasis on continuity and flux, structuralism took discontinuity as its founding principle.

1 Life

Claude Lévi-Strauss was born in 1908. His grandfather was the rabbi of Versailles and his father was a painter. After studying philosophy in Paris he went to the new university of São Paulo in 1934 as professor of sociology, and during his five years in Brazil spent some time travelling in the interior. Returning for military service in 1939 he escaped to New York after the fall of France, and spent most of the Second World War lecturing at the New School for Social Research. Two decisive influences on his intellectual development derive from this period: first, contact with Roman Jakobson and second, exposure to the empirical approach of US anthropology. Since then he has lived and worked in Paris, mostly at the Collège de France, and became a member of the Académie française in 1973.

2 The impact of structural linguistics

Attending Jakobson's lectures on linguistics and phonology in 1942 seems to have been a revelation for Lévi-Strauss. Structural linguistics taught him that 'instead of being led astray by a multiplicity of terms, one should consider the simplest and most intelligible relationships uniting them' (1983, ch. 9: 139). The idea that the phoneme has no intrinsic meaning, but only acquires significance in relation to other similar elements, became the heart of his new approach to anthropological debates. The structural study of phonology emphasized two further principles that were to become fundamental to his work: the importance of unconscious mental activity in the production of language, and the fact that language is organized through discontinuous elements linked in particular ways.

Many commentators have linked Lévi-Strauss' structuralism directly to Saussure. However, it is clear from more recent writings that Lévi-Strauss' source of inspiration was Jakobson's interpretation of Saussure. His own use of Saussure was highly selective, emphasizing above all the arbitrary nature of the linguistic sign, the synchronic basis to linguistic analysis, and the syntagmatic and paradigmatic relations that link linguistic signs. These involve relations of contiguity and similarity, analogous to Jakobson's distinction between metonymy and metaphor (see Structuralism in linguistics).

3 The study of kinship

During this same period, Lévi-Strauss engaged in an extensive study of the ethnological literature, seeking to uncover the 'simplest and most intelligible relationships' that united all the wealth and diversity of the ethnographic record. His first area of interest was kin relations, a topic that linked him directly to the central concerns of anthropology and to Durkheim (§3) and Mauss' theories of social solidarity. His first book, *The Elementary Structures of Kinship* (1949), brought these different strands together in a vast synthetic work that transformed the way that anthropologists understood kinship.

Lévi-Strauss proposed that the incest taboo is analogous to the phoneme, in that it has no meaning of its own but

serves to link two domains. This is achieved by the exchange of women between kin-groups, which constitutes the origin of culture and is its ongoing pre-condition. Like all his work, his theory of kinship is controversial, but many of its insights have become canonical.

4 Totemism and *The Savage Mind*

Kinship led Lévi-Strauss to the theme of totemism and systems of classification, and the application of his theories to more obviously mental phenomena. He took issue with the argument, associated particularly with Lucien Lévy-Bruhl, that the thought processes of primitives are based on entirely different principles to our own. In *The Savage Mind* (1962) he showed how all human cultures employ the same principles to define and order their environment. In fact, 'the savage mind' is an unfortunate - if by now well-established - translation of *La pensee sauvage*, since the English expression implies that he is referring to some distinctive primitive mentality. All humans are alike, in his view, in the way that we rely fundamentally on sensual perception for our systems of classification and knowledge - the 'science of the concrete'. At the same time, primitives are just as capable of abstract thought as we are.

However, Lévi-Strauss does specify the significance of different levels of abstraction, expressed through his well-known contrast between the *bricoleur* and the engineer. The French *bricoleur* is a sort of jack of all trades, who puts things together with whatever materials lie to hand. Most of us for most of the time are *bricoleurs*, close to percepts, using the science of the concrete. By contrast, the 'engineer' operates at 'one remove' from observable properties, constructing concepts appropriate for the use to which they will be put.

In practice, Lévi-Strauss has devoted his life's work to studying the thought principles of *bricoleurs* rather than engineers. But implicitly his own work stands as perhaps a key example of what he means by the engineer. He believes that he has identified the abstract principles that underlie all everyday human thought processes, and that anthropology is particularly suited to the attainment of knowledge because it seeks the 'view from afar'. No civilization can think of itself unless it uses others as a basis for comparison. The importance which he attaches to detachment and distance is also at the heart of his critique of Sartre published as the final chapter of *The Savage Mind*.

5 The science of mythology

The most extensive, and probably the best-known, part of Lévi-Strauss' work is dedicated to the understanding of myth, another manifestation of the 'science of the concrete'. While myth uses the material world as means of thought, it is free from the constraints of material reality. Lévi-Strauss contrasts his own approach to myth with that of Jung, who in Lévi-Strauss' view erred by assuming that myths have stable meanings and a referential function. For Lévi-Strauss, the search for the manifest meanings of myths misses their true nature, which lies beneath the multifarious and often fantastic surface phenomena, where the human mind (*l'esprit humain*) rehearses unceasingly the opposition between nature and culture and the contradictions of human life, without offering a clear explanation or meaning.

Lévi-Strauss' approach to the analysis of myths discounts their narrative form in favour of identifying their component parts, in the spirit of the Russian Formalism of his mentor Jakobson, and others such as Vladimir Propp, author of the hugely influential *Morphology of the Russian Folk Tale*. To identify these discontinuous elements Lévi-Srauss in *The View From Afar* coined the concept of the 'mytheme' analogous to the phoneme - 'a purely differential and contentless sign' (1983: 145). These elements are linked by the principles of metaphor and metonymy, of syntagmatic and paradigmatic relations. The language of myths is, then, an unconscious one, rather similar in his view to that of the genetic code: far from being the consciously-created stories of individual narrators, 'myths operate in men's minds without them being aware of the fact' (*The Raw and the Cooked* (1964: 12).

As such, each myth is connected to others, not only within the same language and cultural group, but beyond that to myths recorded for far-away groups speaking different languages. In his four volume *Mythologiques* 1964-71) he traces mythical themes and structures from one end of the American continent to the other. Crucial to this enterprise is the concept of transformation: in Lévi-Strauss' hands this is a relatively static concept, referring typically to the way that myths are related through inversion, that is, the underlying structure of two apparently

divergent and unrelated myths contains the same elements, but arranged in a different way.

His concept of transformation - a formal rather than a dynamic one - typifies his temporal concepts. Synchrony, the suspension of temporal movement, was a guiding principle for structuralism, and for Lévi-Strauss synchrony expresses a core value. Music has provided a significant analogy, for example in *The Raw and the Cooked* where the temporal unfolding of music (and by analogy the narrative unfolding of myths) is subordinate to the structural principles which unite a piece of music, and link it to other pieces.

6 Lévi-Strauss' theory of knowledge

True knowledge for Lévi-Strauss arises from contemplation. In this respect pre-literate societies, where mythical structures are found in their purest form, come closest to his ideal. They typically strive to negate the effects of history and change, in contrast to the modern world where these are celebrated as good in themselves. In many of his writings Lévi-Strauss emphasizes the unity of the human mind across time and culture, stressing the similarity of purpose and method between our distant ancestors who domesticated animals and plants and our own scientific concerns (especially *Structural Anthropology II* (1973, ch.18). In others he implicitly attributes cognitive superiority to the 'engineer' or indeed to the anthropologist. And yet in other respects it is clear that he admires the disappearing world of the primitives who understood the significance of synchrony.

His passionately argued belief that the primitives were as much intellectuals as we are represents a grand if flawed attempt to create a unified theory of the human mind based on detailed comparative work. It derives from a model of human nature which is grounded in intelligence (in contrast to other models which privilege the emotions and instincts, or self-interest and strategy). However for all his intellectualism, many of Lévi-Strauss' ideas have stood the test of time more for their aesthetic insights and intuitions, than for his theoretical elaborations. As a number of commentators have pointed out, the whole elaborate edifice of structural linguistics conceals as much as it elucidates his aims and methods. He is closer to literature, to music, and above all to painting, and in a late work comments that 'the suppression by chance of ten or twenty centuries would not appreciably affect our knowledge of human nature. The only irreplaceable loss would be the works of art' 1993: 176).

In his analysis of actual myths, Lévi-Strauss' rigorous formalism and implicit reductionism is complemented by an extraordinary grasp of empirical detail and a sensitive attention to the distinctive properties - visual, tactile, olfactory, chemical, aural, functional - of whatever it is that the myth purports to be about. This attention to detail partially undermines the importance which he himself attaches to the Saussurean principle of the arbitrariness of the sign. It was clearly an important notion which enabled him to get away from the referential functions of language, but it should not blind us to the fact that most of the mythical signs which he has devoted the last thirty years of his life to analysing are not arbitrary in the Saussurean sense at all.

See also: Anthropology, philosophy of; Structuralism; Structuralism in literary theory; Structuralism in social science

OLIVIA HARRIS

List of works

Lévi-Strauss, C. (1949) *Les structures elementaires de la parente*, Paris: Presses Universitaires de France; *The Elementary Structures of Kinship*, Boston, MA: Beacon Press, 1969.(A vast synthetic work which transformed the way that anthropologists understood kinship.)

Lévi-Strauss, C. (1955) *Tristes Tropiques*, Paris: Plon; *Tristes Tropiques*, London: Penguin, 1976.

Lévi-Strauss, C. (1958) *Anthropologie Structurale*, Paris: Plon; *Structural Anthropology*, New York: Basic Books, 1963.

Lévi-Strauss, C. (1962) *Le Totemisme Aujourd'hui*, Paris: Presses Universitaires de France; *Totemism*, Boston, MA: Beacon Press, 1963.

Lévi-Strauss, C. (1962) *La pensee sauvage*, Paris: Plon; *The Savage Mind*, Chicago, IL: Chicago University Press, 1966.(Shows how all human cultures employ the same principles to define and order their environment.)

Lévi-Strauss, C. (1964-71) *Mythologiques: Introduction a la Science de la Mythologie (Mythologiques: Introduction to the Science of Mythology)*, 4 vols; vol. 1 *Le cru et le cuit*, Paris: Plon, 1964, *The Raw and the Cooked*, New York: Harper & Row, 1969; vol. 4 *L'homme nu*, Paris: Plon, 1971, *The Naked Man*, London: Cape, 1981.

Lévi-Strauss, C. (1973) *Anthropologie Structurale II*, Paris: Plon; *Structural Anthropology II*, New York: Basic Books, 1976.(Examines the unity of the human mind across time and culture, stressing the similarity of purpose between our distant ancestors who domesticated animals and plants and our own scientific concerns.)
Lévi-Strauss, C. (1978) *Myth and Meaning*, London: Routledge.
Lévi-Strauss, C. (1983) *Le regard eloigne*, Paris: Plon; *The View from Afar*, New York: Basic Books, 1985.
Lévi-Strauss, C. (1993) *Regarder, Ecouter, Lire (Looking, Hearing, Seeing)*, Paris: Plon.

References and further reading

Leach, E. (1970) *Lévi-Strauss*, Modern Masters Series, London: Fontana.(Sympathetic but not reverential account of the main principles of Lévi-Strauss' structuralism from an anthropological perspective. Early but still worth reading.)
Merquior, J.G. (1986) *From Prague to Paris. A Critique of Structuralist and Post-structuralist Thought*, London: Verso.(Locates the philosophical and intellectual environment of Lévi-Strauss' work clearly and engagingly.)
Pace, D. (1983) *Claude Lévi-Strauss: The Bearer of Ashes*, London: Routledge. (Well-written comprehensive overview of Lévi-Strauss' work and its philosophical implications.)
Pavel, T. (1989) *The Feud of Language: A History of Structuralist Thought*, Oxford: Blackwell.(An excellent highly critical account of the whole structuralist enterprise, dense and requiring some previous understanding.)
Sperber, D. (1979) 'Claude Lévi-Strauss', in J. Sturrock (ed.) *Structuralism and Since*, Oxford: Oxford University Press.(A very useful critical but sympathetic synthesis of his *oeuvre*, by one of his most gifted students.)
Todorov, T. (1988) 'Knowledge in Social Anthropology: Distancing and Universality', *Anthropology Today* 4 (2): 2-5.(A subtle and complex exploration of Lévi-Strauss' theory of knowledge.)

Mach, Ernst (1838-1916)

Mach was an Austrian physicist and philosopher. Though not one of the great philosophers, he was tremendously influential in the development of 'scientific philosophy' in the late nineteenth and early twentieth centuries. A vigorous opponent of 'metaphysics', he was celebrated as a progenitor of logical positivism. His work is regarded as a limiting case of pure empiricism; he stands between the empiricism of Hume andJ.S. Mill, and that of the Vienna Circle.

Mach's positivist conception of science saw its aims as descriptive and predictive; explanation is downgraded. Scientific laws and theories are economical means of describing phenomena. Theories that refer to unobservable entities - including atomic theory - may impede inquiry. They should be eliminated where possible in favour of theories involving 'direct descriptions' of phenomena. Mach claimed to be a scientist, not a philosopher, but the 'Machian philosophy' was 'neutral monism'. Close to phenomenalism, it saw the world as functionally related complexes of sensations, and aspired to anti-metaphysical neutrality.

1 Life

Born in Moravia, Ernst Mach studied at Vienna, and became professor of mathematics at Graz in 1864. In 1867 he was appointed professor of physics at Prague. From 1895 until 1901 he was professor of history and theory of inductive science at Vienna, before becoming a member of the upper house of the Austrian parliament. Mach's scientific interests were extraordinarily wide. He contributed to acoustics, theory of electricity, hydrodynamics, mechanics, optics, thermodynamics, and the physiology and psychology of perception.

2 Positivism, rejection of metaphysics, thought-economy

Mach was the leading positivist of the later nineteenth century, and the most important progenitor of the Vienna Circle (see Vienna Circle; Logical positivism). Positivism originates with Comte. Central to it is the celebration of scientific knowledge and corresponding hostility to 'metaphysics'; the aim of philosophy is to understand the dominant, scientific mode of thought of the age. Science is seen as descriptive and predictive; scientific knowledge is exclusively of the correlations between phenomena. Comte and Mach differed in the *extent* of their hostility to metaphysics. Comte believed that speculation about the underlying, real nature of things was unwarranted; Mach held that no such nature, the 'things-in-themselves' behind the phenomena, exists.

This rejection separates Mach from 'two-worlds' Kantianism; but he is closer to late nineteenth-century scientific Neo-Kantians who favoured a 'dual aspect' account (see Neo-Kantianism §3). Mach's professed idealist origins (see §4 below), and the labels 'empirio-criticism' and 'critical positivism' bestowed on his position, indicate some affinity to Neo-Kantianism and a corresponding distance from earlier empiricism. Positivism does not imply inductivism, the view that theories are churned out mechanically from the data (see Inductive inference; Mill, J.S. §5). None the less, for Mach all genuine knowledge is scientific knowledge. Although the logical positivists may have rejected merely the transcendental underpinnings of the synthetic a priori, Mach seems to reject it completely, regarding space and time as given, and not as forms of intuition (for which he was criticized by Schlick). Hence there is some truth, but not complete truth, in the traditional view of Mach as the purest empiricist.

Mach's work predated intellectually and for the most part chronologically the 'linguistic turn'. He was innocent of Frege's revolution in logic, central to the outlook of members of the Vienna Circle such as Carnap andReichenbach. Mach's concern with 'analysis' in *Die Analyse der Empfindungen* (*The Analysis of Sensations*) (1886a) was ontological, not the 'logical' or 'conceptual' analysis foundational to analytic philosophy (see §4). Likewise his 'verificationism' was not the developed semantic thesis it became with the Vienna Circle. It was Mach's vigorous hostility to 'metaphysics', more than his detailed views, that was influential.

Central to Mach's philosophy is 'economy of thought', influenced by the economist Emanuel Herrman. Scientific laws and theories are economical descriptions of phenomena. The order science imposes on experience is a biological process of adaptation that overcomes limitations of memory; it does not reflect anything intrinsic to nature. *Die Mechanik in ihrer Entwicklung historisch-critisch dargestellt (The Science of Mechanics)* (1883) is a history of the evolution of science; its origins in 'the manual arts' are emphasized, and false abstraction and 'misplaced rigour' are debunked. 'All so-called axioms are…instinctive knowledge', the funded experience of the

human race. The economy of thought is at the root ofMach's so-called 'sensationalism'. Talk of 'bodies' or 'objects' results from a 'partly instinctive, partly voluntary and conscious economy of mental presentation'.

3 Scientific methodology: laws, theories and explanation

Science, for Mach, is 'the compendious representation of the actual'. Known laws of nature - that is, regularities or interconnections between appearances - are explained and new laws discovered by the construction of theories. Explanation is downgraded, on the grounds that it makes a misguided appeal to metaphysical causes of things. The positivist account of scientific theorizing, with its anti-explanatory bias, is nowadays regarded as moribund. However, as Laudan has argued, Mach's own account of theory is more subtle than is often acknowledged. He was not hostile to the speculative construction of theories, as such. None the less, even if theories are required for the integration of data, they remain *about* the data; given that the correlation of the data takes one beyond them, entities should not be postulated beyond what is strictly necessary for that correlation.

In *Populärwissenschaftliche Vorlesungen (Popular Scientific Lectures)* (1894) Mach distinguishes 'direct' and 'indirect' descriptions. The former involves terms that have acquired 'a fixed significance independent of every object and connexion', taking on 'an "abstract" or "conceptual" import': for example, 'red' or 'straight'. (It has been said that direct descriptions are given in terms of sensations, but Mach is not explicit.) A 'theory' or 'theoretical idea' involves *indirect* description - the extension of analogy - and hence unfamiliar phenomena are made comprehensible. (In talking of 'auxiliary' and 'transitional theories', Mach later implies that there are theories that consist simply in direct descriptions; there is ambiguity here.)

But the picture expressed in an indirect description may impede further inquiry, Mach believes. Black's discovery of specific heat was guided by the picture of thermal processes as the flow of a substance from one body to another; which, however, impeded the realization that heat is produced by friction, for which the picture of heat as a form of motion is more appropriate. Either picture may be heuristically valuable, but neither should be regarded as constituting the reality behind the observable phenomena.

Hence a theory may be heuristically and predictively useful; but the difference between theory and observation is purely quantitative. What is essential is that the theory contain the correct quantitative relations between work done and heat generated, between heat disappearing in one place and appearing in another, etc. Once we have established the functional dependence at the level of appearances, the theoretical model has done its work and should be jettisoned: 'as the new facts grow familiar, [one should] substitute for indirect description *direct* description, which contains nothing that is unessential and restricts itself absolutely to the abstract apprehension of facts'.

Laudan terms this the 'thesis of positivistic elimination'; 'purely observational' and 'mixed theories' (which refer in part to entities beyond the reach of experimentation) should have a different methodological status, and the latter eliminated where practicable in favour of the former. Hence one reason forMach's notorious opposition to atomic theory, namely that it has outlived its usefulness; rather than, as customarily imputed, that atoms are 'unobservable'. In so far as Dalton discovered new connections between phenomena, the theory, properly interpreted as a heuristic device, had been acceptable, Mach believed. Many of Mach's opponents in the atomist debate accepted his methodology and defended atomic theory on its terms.

Mach's deflationary attitude towards theoretical entities is also expressed in his view that theory-choice is underdetermined by empirical evidence. In *Erktenntnis und Irrtum (Knowledge and Error)* (1905) he writes that concepts 'must be consistent with observation and…must be logically consistent with one another. These two requirements are satisfiable in *numerous* ways'. Thus, despite his biological determinism, there is a conventionalist element in Mach's scientific methodology. The pervasive role of the 'economy of thought' implies a more thoroughgoing instrumentalism than forMill, for whom inductive proof was still central (see Conventionalism).

4 Neutral monism

'There is above all *no* Machian philosophy but at most a scientific methodology and a psychology of knowledge…I am by no means a philosopher, but only a scientist' (1905). In one sense, Mach's disclaimer was disingenuous. His philosophical standpoint was one shared with positivists and non-positivists around the turn of the century - namely, neutral monism (a term coined by Russell, though Mach himself referred to both 'monism'

and ‘neutrality’) (see Neutral monism).

The Science of Mechanics introduced a constant theme, leading to the alternative characterization of his position as ‘sensationalism’:

> Sensations are not signs of things; but on the contrary, a thing is a thought-symbol for a compound sensation of relative fixedness. Properly speaking the world is not composed of ‘things’…but of colours, tones, pressures, spaces, times, in short what we ordinarily call individual sensations.
>
> ([1883] 1974: 579)

Mach was greatly influenced by Fechner’s ‘psychophysics’, and worked on the physiological basis of sensation, while opposing Fechner’s mechanistic reduction of mental phenomena to a concourse of physical atoms (see Fechner, G.T. §2). What is required for developing psychophysical investigation, Mach believed, is a *sensational* atomic basis. He aspired to ‘the gathering up of the sciences into a single whole’, but this early version of the ‘unity of science’ was not materialistic. In *The Analysis of Sensations*, Mach traces his idealist roots. In adolescence he was a Kantian, but then he reacted against the thing-in-itself: ‘the world…suddenly appeared to me as one coherent mass of sensations, only more strongly coherent in the ego’. He still shared Berkeley’s belief in an ego not identical with sensations, before finally rejecting that too. Mach concluded that the useful habit of regarding complexes as unities generates the notion ‘at first impressive, but subsequently recognized as monstrous, of a "thing-in-itself", different from its "appearance", and unknowable’. The idea of a transcendental, unknowable Ego is for Mach the last remnant of the detested ‘thing-in-itself’.

Mach emphasizes that the ‘elements’ are neither mental nor physical, subjective nor objective, but ‘neutral’. The title ‘neutral monism’ is therefore more appropriate than ‘sensationalism’, since it implies a rejection of two dualisms, mental-physical and subject-object. The complexes of elements,Mach claims, are functionally interdependent. The green of a leaf is a physical element in its dependence on other properties of the leaf and its environment, and a psychical element in its dependence on my perceptual processes. Which it should be regarded as depends simply on whether one’s interest is that of the physicist or the psychologist.

Mach’s account of the subject prefigures twentieth-century eliminative accounts of the self. But his analysis of objects more strongly anticipates an often complementary phenomenalism. For its originator, J.S. Mill, phenomenalism is the seemingly *non*-neutral view that physical objects are constructions from sensations (see Phenomenalism). Despite protestations of neutrality, doesMach’s view of objects as actual and possible sensations and his rejection of any full-blooded distinction between appearance and reality simply amount to phenomenalism?

Mill’s is the ontological thesis that physical objects are collections of actual and possible sensations, construed as mental, or at least subjective, items. Mach wants to resist this, but claims both that: (1) ‘another’s sensations are no more directly given to me than mine to him’ (1905), yet (2) there is no ‘essential distinction’ between my sensations and those of another person; ‘the same elements are connected at different points of attachment, namely those of the Ego’. (1) implies a subjective interpretation of the elements. Mach recognizes the pressure to go in this direction; but since the ego is, he thinks, a mere thought-economical pseudo-unity, he believes he is entitled to (2), which implies an objective interpretation. His vacillation is understandable but unsatisfactory.

But this ambivalence in neutral monism itself determined the structure of twentieth-century linguistic phenomenalism. Mach was followed by later phenomenalists when he tried to correct Mill’s subjectivism, but he was unable to take Carnap’s further ‘neutralizing’ or anti-metaphysical step of replacing (or perhaps implementing) the ontological thesis with a semantic one - that physical objects are *logical* constructions from sense data (see Carnap, R.). This became possible only after the ‘linguistic turn’ of analytic philosophy.

Neutral monism, then, is the ‘Machian philosophy’. But the scientific methodology was more important, and Mach wrote ten times as much on it as he did on sensationalist epistemology. In another sense,Mach’s disclaimer of any ‘Machian philosophy’ is an important precursor of Carnap’s view of philosophy as the ‘logic of science’.Mach’s attempt to generate a unified science on the basis of ontological neutrality was combined with a circumscribed role for philosophy - the dissolution of ‘pseudoproblems’ - as adjunct to that science. His disclaimer is echoed in Carnap’s positive claim in *Logische Syntax der Sprache (The Logical Syntax of Language)* (1937) that ‘the logic of science takes the place of the inextricable tangle of problems which is known as philosophy’.

See also: Science, nineteenth century philosophy of; Relativity theory, philosophical significance of §6

ANDY HAMILTON

List of works

Mach, E. (1883) *Die Mechanik in ihrer Entwicklung historisch-critisch dargestellt*, Prague; trans. T. McCormack (1893) *The Science of Mechanics*, La Salle, IL: Open Court, 1974.(A history of the evolution of science, including a comprehensive attempt to reconstruct classical mechanics on positivist lines.)

Mach, E. (1886a) *Beiträge zur Analyse der Empfindungen*, Jena; 5th edn, *Die Analyse der Empfindungen*, Jena, 1906; trans. C. Williams (1914) *Contributions to the Analysis of Sensations*, La Salle, IL: Open Court, 1986. (Mach's most developed statement of his neutral monism.)

Mach, E. (1886b) *Die Prinzipien der Wärmelehre*, Leipzig; trans. T. McCormack, P. Jourdain and A. Heath, *Principles of the Theory of Heat*, Dordrecht: Reidel, 1986.(Influential account of the history of the theory of heat.)

Mach, E. (1894) *Populärwissenschaftliche Vorlesungen*, Leipzig; trans. T. McCormack (1894) *Popular Scientific Lectures*, La Salle, IL: Open Court, 1986.(Exposition of familiar Machian scientific themes for a more general audience.)

Mach, E. (1905) *Erktenntnis und Irrtum*, Leipzig; trans. T. McCormack and P. Foulkes, *Knowledge and Error*, Dordrecht: Reidel, 1976.(A collection of essays which constitutesMach's most mature statement of his scientific epistemology.)

Mach, E. (1906) *Space and Geometry*, Chicago, IL; trans. and collected by T. McCormack, La Salle, IL: Open Court, 1984.(Treatment of the concepts of 'physiological' and 'geometrical' space, and of the history of geometry. Collected from three essays originally published in *The Monist*, 1901-3.)

References and further reading

Blackmore, J. (1972) *Ernst Mach - His Life, Work and Influence*, Berkeley, CA: University of California Press. (Pedestrian, but the only full-length biography.)

(ed.) (1992) 'Ernst Mach - A Deeper Look', in *Boston Studies in the Philosophy of Science* vol. 143, Dordrecht: Kluwer.(Useful correspondence, some good articles and an invaluable bibliography.)

Carnap, R. (1934) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul, Trench, Trubner & Co., 1937.(Carnap's presentation of his conventionalism.)

Cohen, R. and Seeger, R. (eds) (1970) 'Ernst Mach, Physicist and Philosopher', *Boston Studies in the Philosophy of Science* vol. 6, Dordrecht: Reidel.(Useful collection.)

Haller, R. (1982) 'Poetic Imagination and Economy: Ernst Mach as Theorist of Science', in J. Agassi and R. Cohen (eds) *Scientific Philosophy Today - Essays in Honour of Mario Bunge*, *Boston Studies in the Philosophy of Science* vol. 67, Dordrecht: Reidel; also in J. Blackmore (ed.) *Boston Studies in the Philosophy of Science* vol. 143, Dordrecht: Kluwer.(Useful corrective to the received picture of Mach.)

Hamilton, A. (1990) 'Ernst Mach and the Elimination of Subjectivity', *Ratio* (new series) 3 (2): 117-35. (Expansion of the material of §4.)

Holton, G. (1970) 'Mach, Einstein and the Search for Reality', in *Thematic Origins of Scientific Thought - Kepler to Einstein*, Cambridge, MA: Harvard University Press, 1972; also in Cohen and Seeger (eds) 1970.(Discussion ofMach's influence on Einstein.)

Laudan, L. (1981) 'Ernst Mach's Opposition to Atomism', in *Science and Hypothesis: Historical Essays on Scientific Methodology*, Dordrecht: Reidel. (Pellucidly relates Mach's anti-atomism to his central concerns. An outstanding article of a different level of sophistication to most writing on Mach and positivism.)

Mandelbaum, M. (1971) *History, Man and Reasons: A Study in Nineteenth-Century Thought*, Baltimore, MD: Johns Hopkins Press, chaps 1, 14.(A classic treatment of nineteenth-century intellectual history.)

Schlick, M. (1985) *General Theory of Knowledge*, La Salle, IL: Open Court, 194-233.(A penetrating critique ofMach's neutral monism by a leading member of the Vienna Circle; first published 1925.)

Skorupski, J. (1993) *English-Language Philosophy 1750-1945*, Oxford: Oxford University Press, ch. 3. (Introductory though not particularly easy.)

Smith, B. (1987) 'Austrian Origins of Logical Positivism', in B. Gower (ed.) *Logical Positivism in Perspective*, London: Croom Helm.(Authoritative historical enquiry.)

Machiavelli, Niccolò (1469-1527)

Florentine diplomat, dramatist and political thinker, Machiavelli's treatise, Il principe (The Prince) (1532a), has earned him notoriety as a political immoralist (or at least an amoralist) and a teacher of evil. In The Prince, Machiavelli posits a complex relationship between ethics and politics that associates princely virtù *with the capacity to know and act within the political world as it 'is', and with the beastly abilities to dispense violence and practise deception. Behind this argument dwells the distinctly Machiavellian insight that politics is a realm of appearances where the practice of moral or Christian virtues often results in a prince's ruin, while knowing 'how not to be good' may result in greater security and wellbeing for both prince and people. Machiavelli warns that the prince's possibilities for success in this matter are always mediated by fortune; hence the prudent prince is one who is prepared to resist fortune by adapting his procedure to the times and his nature to 'the necessity of the case'.*

A less notorious but equally influential text is the Discorsi sopra la prima deca di Tito Livio (Discourses on the First Ten Books of Titus Livy) (1531), in which Machiavelli offers a defence of popular liberty and republican government that takes the ancient republic of Rome as its model and emphasizes the role of the people in the 'public administration' of the city. However, Machiavelli also argues that a republic is only as successful in self-governance as its citizens are infused with civic virtù *and therefore not corrupted. Accordingly, he praises the work of political founders who craft republican laws and institutions, and religious founders who fuse God and* patria *as one in the people's hearts. The apparent tension between Machiavelli's republican sympathies in Discourses and his elitist proclivities in The Prince has helped to fuel a vast interpretive literature concerning his political attitudes, his theory of politics, and the nature and meaning of 'machiavellianism' in Western political thought.*

1 Machiavelli in politics

From the beginning Niccolò Machiavelli's life, like the city of Florence itself, was caught in the ebb and flow of power that characterized the fortunes of the great banking and ruling house of the Florentine Medici. Machiavelli was born into the Florence of Lorenzo the Magnificent, a Medici prince who ruled the city until his death in 1492. When Piero de' Medici fled Florence at the approach of the French army in 1494, the family's hold on power was temporarily broken. After the short-lived regime of the charismatic friar Savonarola, the Florentine republic was restored, and Machiavelli was positioned to begin a career in politics. As Second Chancellor and Secretary to the Ten of War, he worked as a government functionary in affairs of state, travelling widely on diplomatic missions to various secular and ecclesiastical courts, as well as to the campaign headquarters of Cesare Borgia. As a result, Machiavelli became intimately acquainted with the strategies that are a part of 'the art that is necessary to one who commands' ([1532a] 1950: 53).

The Florentine republic was relatively short-lived. With the aid of the Spanish army, the Medici recaptured the city in 1512. Machiavelli was arrested, imprisoned and tortured, under suspicion as a co-conspirator in a plot to assassinate the Medici brothers. In 1513, he was released from prison and sent to rural isolation on his farm in the Tuscan countryside. 'To-day fortunate, tomorrow ruined,' was Machiavelli's succinct summation later. He would never again participate actively in the political affairs of Florence.

Like many displaced political actors, Machiavelli turned to writing. Within the first year of his exile, he produced a short treatise on *l'arte dello stato* wherein, he reported to his friend Francesco Vettori, 'I go as deeply as I can into considerations on this subject, debating what a princedom is, of what kinds they are, how they are gained, how they are kept, why they are lost' ([1498-1527] 1961: 142). Under this seemingly prosaic description smoulders the book that ignited a firestorm of invective in the sixteenth century, a book that continues to be one of the most controversial texts in the history of Western political thought.

2 *Il principe* (The Prince)

In terms of its formal structure, *Il principe (The Prince)* consists of a dedication (to Lorenzo the Magnificent, son of Piero de' Medici), and twenty-six chapters that can be divided into five general sections: chapters I-XI: a typology of states and how they are acquired; chapters XII-XIV: observations on militia and military command;

chapters XV-XIX: advice to a prince regarding character and conduct; chapters XX-XV: advice to a prince regarding circumstances and conditions (fortresses, favours, functionaries and fortune); and chapter XVI: an exhortation to Italy that reconfigures the credo of Pope Julius II: *Fuori i Barbari!* (Put the barbarians out!). The formal structure of the treatise belies its substantive complexity, however, and the magnitude of rival interpretations that attend it. (There are now well over 3,000 commentaries.)

The notoriety of *The Prince* rests primarily upon the nature of Machiavelli's advice concerning the 'methods and rules' that a leader must follow in order to exhibit strength (*virtù*), gain and maintain his position within a particular territory (*lo stato*), and secure the state itself. On these matters, Machiavelli crafts a perspective that aligns politics with warfare, and justifies the deployment of force, the exercise of cruelties, the practice of deceit and the manipulation of appearances in the service of a political mentality that is 'disposed to adapt itself according to the wind' and 'able to do evil if constrained' ([1532a] 1950: 65). In the opening paragraph of chapter XV, in which this perspective on politics crystallizes, Machiavelli dismisses the usefulness of those writers who imagine 'republics and principalities which have never been seen or known to exist in reality', and declares that he will go 'to the real truth of the matter' (*verità effettuale*) ([1532a] 1950: 56). The real truth is that in politics, a ruler must be ready to play upon falsehoods and appearances, 'being often obliged, in order to maintain the state, to act against faith, against charity, against humanity, and against religion' ([1532a] 1950: 65). Accordingly, a prince must 'learn how not to be good,' and 'use this knowledge and not use it, according to the necessity of the case' ([1532a] 1950: 56).

In producing this perspective, *The Prince* forcefully deviates from and craftily subverts an entire genre of Renaissance advice-books that traditionally had aligned princely *virtù* (those qualities that enable a ruler to acquire honour, glory and fame) with the Platonic and Ciceronian 'cardinal' virtues of wisdom, justice, courage and moderation, and with the Christian ethic of goodness and righteousness. Machiavelli's treatise may not have 'revolutionized' the genre of advice-books, as Skinner (1978) contends, but it surely constitutes an unprecedented Renaissance challenge both to moral philosophy and to Christianity as normative bases for a theory of politics. This challenge may have earned the Florentine the admiration of 'anti-philosophers' and 'anti-Christs' of later ages (Nietzsche foremost among them), but more immediately it induced the Roman papacy to place all of Machiavelli's writings on the Index, where they remained for almost 300 years. *The Prince* also set in train its author's long-lived reputation as an atheistic, evil and satanic 'Machiavel', a man 'inspired by the Devil to lead good men to their doom' (Berlin 1982: 35). In contemporary vocabulary, the terms 'machiavellian' and 'machiavellianism' capture an understanding of politics as a domain that embraces naked self-interest, the maintenance of rulership at all costs, the utility of unethical behaviour and the centrality of power as an end that justifies any means.

3 The politics of princes

Despite the durability of these characterizations, Machiavelli posits a more complex relation between value (or ethics) and politics in *The Prince* than either an immoralist, 'teacher of evil', or an amoralist, *realpolitik*, rendering of his views allow. It is true that, unlike any other humanist writer, he openly associates princely *virtù* not with law (which he deems often insufficient to the combat that is ineliminable in politics), but with the abilities to dispense violence, or imitate the lion, and practise deception, or imitate the fox. Behind this argument dwells the distinctively Machiavellian insight that, in matters of state, 'some things which seem virtues would, if followed, lead to one's ruin, and some others which appear vices result in one's greater security and wellbeing' ([1532a] 1950: 57). Goodness is sometimes disastrous in politics, and cruelty less ruinous than clemency. Thus knowledge of beastly abilities and when and how to use them is essential, and it is acquired in two ways: by action, and by the historical study of the good and bad fortunes of 'eminent men'. The men Machiavelli admires and recommends for study are unexcelled in their capacity to exert coercive power and trick others into doing what they want, and thus bring fortune (*fortuna*) under control: the bloody Pope Alexander VI, and the warring and impetuous Pope Julius II, the ruthless Cesare Borgia, the scheming Roman emperor Severus, the more humane Marcus, the wily *condottiere* Francesco Sforza, and Scipio, Cyrus and Alexander the Great.

Yet there are also apparently effective rulers that Machiavelli introduces and condemns: Agathocles the Sicilian, Oliverotto da Fermo, and a host of nefarious Roman emperors. In proffering these examples, *The Prince* appears to formulate a standard of right, or at least prudential, conduct that is appropriate to *virtù*. Thus it throws into

question the standard view that machiavellianism celebrates politics as a sphere of unrestrained evil, and power as the exercise of unremitting violence. Machiavelli's most notable condemnation concerns Agathocles, who arrived at his position as prince in Syracuse through the cold-blooded slaughter of the senators and richest men of the city. Machiavelli remarks: 'it cannot be called virtue to kill one's fellow-citizens, betray one's friends, be without faith, without pity and without religion; by these methods one may indeed gain power but not glory' ([1532a] 1950: 32). The comment on glory (*gloria*) introduces a crucial qualification, because it suggests that the truly *virtù*oso prince is not one who uses his power simply for self-aggrandizement, but in order to construct something that will outlast him, and sanctify his name.

Even in political situations where glorious deeds are unlikely, however, and pure power reigns supreme, Machiavelli distinguishes between ferocious and contemptible conduct, and praises as *virtù*ous those leaders who are feared but not hated. At issue here is a deeper distinction that does not obviate conflict and cruelty as inescapable aspects of political action, but nevertheless differentiates cruelties that are 'well-committed' from those that are 'ill-committed', and therefore increase rather than diminish with time. It is precisely Machiavelli's refusal to disentangle conflict and cruelty from politics that has led some commentators to configure him as a teacher of evil who divorces politics from ethics (Strauss 1958). For others, however, his thought represents a powerful reconstitution of an action concept that, even as it withdraws politics from the domain of purely moral or Christian judgment, reinvests it as a value sphere with its own peculiar and demanding configurations of good and evil, and hence its own norms and standards of ethical evaluation (Berlin 1982).

4 *Virtù* and *fortuna*

Whatever their ethical implications, neither Machiavelli's action concept of politics nor his notion of princely *virtù* can be fully appreciated without considering the role he assigns to *fortuna* in human affairs. The metaphysical picture behind his account of politics is that of an all-pervading necessity that forms the fabric of the world, but wherein freedom, as the exercise of human control over circumstances, is a real possibility at least half of the time. Thus princely *virtù* also presupposes a capacity to master fortune and navigate within the inherent flux and turmoil of political affairs. It is perhaps this elusive capacity that Machiavelli struggled hardest to comprehend, in response to a phenomenological puzzle about action, character, circumstance and method. Strategically, he cast the solution to this problem as the ability to be prudent (*prudente*), and thereby read circumstances, anticipate probabilities and master events. In terms of character, he presented it as the ability to be inherently flexible, the willingness to deviate from what comes naturally and the determination to exercise a steely self-control. Thus, Machiavelli writes: '[the prince] is happy whose mode of procedure accords with the needs of the times, and similarly he is unfortunate whose mode of procedure (*modo di procedere*) is opposed to the times' ([1532a] 1950: 92).

Readers should note, following Pitkin (1984), that Machiavelli's 'metaphysics' is decisively gendered, in so far as he represents *fortuna* not only as a river, but also as a woman, or an engulfing female power that threatens to overwhelm and crush the man of *virtù*. 'Fortune varies and commands men, and holds them under her yoke'; but if the prince is impetuous and willing 'to conquer her by force', he can subdue fortune and thereby use her to his advantage (1498-1527: 99, 94). The symbolic gendering of action, chance and circumstance that Machiavelli effects on a state level in *The Prince* is recapitulated on a domestic level in his play *Mandragola* ([*c*.1518] 1957), where the oppositional erotics of instrumental violence is delivered as comedy. Both *The Prince* as politics, and *Mandragola* as domestics, allow us to see the limits of a world that is constructed as a struggle between *virtù* and *fortuna*, a world wholly devoid of human mutuality, reciprocity and genuine civic life.

5 *Discorsi* (The Discourses)

Discorsi sopra la prima deca di Tito Livio (Discourses on the First Ten Books of Titus Livy) appears to redress, if not repudiate, the uncivic orientation of *The Prince* and its comedic counterpart *Mandragola* by directing its political attentions elsewhere. In this work (probably completed in 1518 although not printed until 1531), Machiavelli distances himself from 'those who generally dedicate their works to some prince' and resolves to undertake the dangerous task of introducing 'new principles and systems' like the explorers of 'unknown seas and continents' in order to 'open a new route, which has not yet been followed by any one'. The new route is discoverable only by way of a return to the past or to 'examples of antiquity' ([1531] 1950: 102-4). Although he implies at the start that the subject matter of his treatise will encompass a vast array of historical and political

topics, *Discourses* in fact concentrates upon one singular historical example, the Roman republic, and one singular type of ruler: 'the sagacious legislator of a republic…whose object is to promote the public good, and not his private interests, and who prefers his country to his own successors' ([1531] 1950: 138). Thus *Discourses* is best read as illuminating the presence of the past, in order to assess what is necessary if modern republics are to flourish and endure like ancient Rome. More indirectly but no less decisively, *Discourses* is also a biting condemnation of the corruption of Florence, a subject that Machiavelli addressed more subtly in *Istorie Fiorentine (The History of Florence)*, a work commissioned in 1520 (although not published until 1532) by a new Medicean court that, at long last, rewarded his literary if not his diplomatic talents.

Discourses is divided into three books (of 60, 33 and 49 chapters respectively) that span an impressive array of political topics concerning the early history of Rome and the Romans: origins and founders, religion, armies, conquests and colonies, the constitution, institutions of government, customs, laws (civil and agrarian), character, warfare and diplomacy, notable leaders (including a tribute to Cincinnatus) and citizens. A careful look at the chapter titles reveals the question that is central to Machiavelli's investigation of Rome: What makes republican regimes stable and enduring? Pocock (1975) argues that in *Discourses* Machiavelli configures the relationship between the all-too-human 'republic' and the unremitting forces of 'time' as a historical confrontation that cannot be deferred. But Machiavelli also attempts to mediate the confrontation through a careful study of the conditions and civic qualities that are necessary if republics are to have even a minimal chance of effectively countering the ordinances of *fortuna* and 'achieving greatness like that of Rome'. Machiavelli's study of civic *virtù* is at the heart of his republican theory of citizenship (see Republicanism §2).

6 The politics of republics

If we approach Machiavelli's classical republican theory as a complex of elements, then four emerge as primary. The first, which can hold for either principalities or republics, is *libertà*, or the condition of being free from subjection to a foreign power and governed by one's own laws. The significance that Machiavelli attaches to liberty cannot be overstated, for he argues that only a city free from servitude is potentially capable of achieving greatness, whether in power, or wealth, or both. For this reason he cautions against attempts to establish republics in countries not suited to them lest the government 'lack proper proportions and have but little durability' ([1531] 1950: 257). Nevertheless and in general, the institutions that are essential to civic liberty are better perfected in republics than in monarchies.

If republics are better safeguards of liberty than monarchies, then republics which entrust their liberty to the people are even more likely to retain freedom than those who favour the nobility. The second element of Machiavellian republicanism involves a defence of the people (*popolo*), as the active guardians of public liberty. As usual, Machiavelli's reasoning on this matter is complex and contextual, and it does not denote an absolute rule. In Sparta and Venice for example, where power was entrusted to the nobility, liberties endured far longer than in Rome. Nevertheless, a republic that desires to extend its empire, rather than merely seek its own preservation, would do well to follow Rome and entrust liberty to the people. A republic that emulates Rome recognizes that its liberty depends upon admitting its people 'to a share of its glory', for the people are superior in maintaining the 'institutions, laws and ordinances' upon which the freedom of the republic rests.

Machiavelli recognizes, however, that a self-governing republic that is animated by ordinary citizens is only as *virtù*ous as its laws and institutions (*ordini*) allow. Thus the third element of his republican theory places great emphasis upon the talents of the founders of republics, like Romulus, and even more emphasis upon those founders who recognize God and religion 'as the most necessary and assured support of any civil society' and use them to divinize the republic's laws and institutions before an awe-struck people, as Numa did in Rome. The welfare of a republic does not consist, then, 'in having a prince who governs it wisely during his lifetime, but in having one who will give it such laws that it will maintain itself even after his death' ([1531] 1950: 146, 148). Among the specific institutions that Machiavelli recommends as most beneficial to republican liberty is a government that combines 'a prince, a nobility and the power of the people' under the same constitution. This was the case in ancient Rome, where the consuls, the senate and the tribunes maintained a tense equilibrium, and kept each other in check. The significance of Machiavelli's advice on this matter has less to do with the later 'separation of powers' doctrine than with his insistence that 'all the laws that are favourable to liberty' result from partisan struggles between the nobility and the people ([1531] 1950: 119).

The emphasis on struggle which, as we have seen, is an indelible aspect of Machiavellian politics, leads to the fourth and most audacious element of Machiavellian republicanism. This element forwards a view of civic participation (*participazione*) as the inevitable clash of interests within and among diverse competing political forces, and it embraces a view of the republic as a 'theatre of turbulence'. From this perspective, the prosperity of republics hinges not upon quelling turbulence or repressing divisiveness, but rather on finding ways of channelling the energies and agitations within the state to good effect. This is particularly true with regard to the people. Accordingly, Machiavelli insists that: 'every free state ought to afford the people the opportunity of giving vent, so to say, to their ambition' ([1531] 1950: 120). Behind this remark rests an even deeper conviction: when civic disturbances generate *ordini* that give the people 'a share in the public administration', they also establish the most assured guardians of liberty, as the tribunes were in Rome.

From the complex of elements (*libertà*, *popolo*, *ordini* and *participazione*) that constitutes the republican orientation of *Discourses* arises a host of secondary issues that are no less important to Machiavelli's analysis of how republics can survive the test of time and fortune. Primary among these is the issue of vulnerability, or what threatens a republic's stability from within and from outside. Since he is well aware of the vulnerability of republics to threats from within, Machiavelli is alert to the problem of corruption (or the disintegration of civic *virtù*), especially as it manifests itself in ambition (*ambizione*) and in factions that threaten to turn productive conflict into internecine warfare (see Corruption §1, 3). Book III, chapter 6 of *Discourses*, 'Of Conspiracies', unmasks the motives and methods that induce individuals, whether alone or in groups, to engage in such warfare, and plot against princes and republics. In true Machiavellian fashion, it also instructs potential conspirators about what works and what does not.

Machiavelli is also aware of the vulnerability of republics to threats from outside, and on this score he repeats and develops an admonition that is found in *The Prince* and elaborated at great length in *Arte della guerra (The Art of War)* (1521): if glory is to be attained, a state must rely upon its own armies and cultivate in its citizens a military *virtù*, as the Romans did. Thus: 'any republic that adopts the military organization and discipline of the Romans…will always and under all circumstances find [her soldiers] to display a courage and dignity similar to that of the Romans' ([1531] 1950: 504). Without good military organization there can be neither 'good laws nor anything else good', and a republic is left prey to the vicissitudes of fortune. Although Machiavelli is in this instance comparing the 'wretched military organization' of Venice to that of Rome, it is hard not to see Florence in this depiction. Read in this light, *Discourses* intones a dirge for Florence, even as it strikes symbols of tribute to ancient Rome.

7 The motives of Machiavelli

Almost inevitably, the comparison of the subject matter of *The Prince* and *Discourses* gives rise to the intriguing question of Machiavelli's motives and his political attitudes. For several centuries, commentators have found it difficult to overlook what seems to be a glaring contrast between the Medicean sympathies of *The Prince*, in which Machiavelli extols princes and insists that all that the people want is not to be oppressed, and the republican convictions of *Discourses*, in which the people are elevated as the guardians of liberty, and the lives of princes are short. In essence, the interpretive question seems to be this: How is it possible to hail Machiavelli as a defender of republican citizenship and liberty, self-government and civic *virtù*, when these appear to be the very regimes and values he teaches his Medici protégé to subvert?

In response, some critics (such as Hans Baron and J.R. Hale) argue that Machiavelli was in fact less a republican and more of a realist who adjusted his advice as political conditions warranted. Thus *The Prince* can be read as a clear-sighted, if opportunistic, attempt to gain favour with the ruling regime. Other critics insist upon Machiavelli's unwavering commitment to republicanism (as Spinoza did), and reconfigure the notorious advice-book as (a) a subversive handbook for republicans (an interpretation favoured by Rousseau); (b) a satire on princely governance (provided by Garrett Mattingly); or (c) a masterful act of political deception, intended for and against the Medici (offered by Mary Dietz). Still others, including Sheldon Wolin, suggest that there are no real dissimilarities between the two texts at all; thus the republicanism of the *Discourses* can be taken as the salutary result of the political acts of a well-advised, *virtù*ous prince. Whatever else, these debates over Machiavelli the man are proof that Machiavelli the theorist of politics sought and perhaps mastered the ability to be simultaneously deceptive and truthful, realistic and ethical, cruel and humane.

See also: Huainanzi; Political philosophy, history of §4

MARY G. DIETZ

List of works

Machiavelli, N. (1960-5) *Opere*, in *Bibliotheca di classici italiani*, ed. S. Bertelli, Milan: Feltrinelli, 8 vols.

Machiavelli, N. (1883) *Lettere familiare*, Florence: Edordo Alvisi G.C. Sansone.

Machiavelli, N. (1498-1527) *The Letters of Machiavelli*, trans. A. Gilbert, New York: Capricorn, 1961. (Machiavelli's wide-ranging correspondence and commentary on affairs public and private, cultural and literary, international and domestic.)

Machiavelli, N. (*c*. 1518) *Mandragola*, trans. A. Paolucci and H. Paolucci, New York: Bobbs-Merrill, 1957.(A domestic comedy of sexual intrigue, deception and gender transgression.)

Machiavelli, N. (1521) *Arte della guerra (The Art of War)*, trans. E. Farneworth, New York: Washington Square Press, 1965.(A military classic in the form of a dialogue. Subjects range from matters concerning ancient and contemporary arms and armour, battalion formations, and Greek and Roman methods of war, to giving commands, dividing spoils and avoiding battles and traps.)

Machiavelli, N. (1531) *Discorsi sopra la prima deca di Tito Livio (Discourses on the First Ten Books of Titus Livy)*, in *The Prince and The Discourses*, trans. L. Ricci, New York: Modern Library, 1950.

Machiavelli, N. (1532a) *Il principe (The Prince)*, in *The Prince and The Discourses*, trans. L. Ricci, New York: Modern Library, 1950.(The famous treaty on power, conduct, effectiveness and princely *virtù* in politics that earned Machiavelli his reputation as a political 'immoralist'.)

Machiavelli, N. (1532b) *Istorie Fiorentine (History of Florence)*, trans. H.A. Rennart, Washington, DC, and London: Walter Dunne, 1901.(A historical study of transformation and change in Florentine domestic and international politics from the Middle Ages (1010) until the late Renaissance (1492).)

References and further reading

Berlin, I. (1982) 'The Originality of Machiavelli', in *Against the Current: Essays in the History of Ideas*, New York: Penguin.(A classic survey of the existing literature on Machiavelli and his motives also makes a case for the 'pagan' morality of his politics.)

Grazia, S. de (1989) *Machiavelli in Hell*, Princeton, NJ: Princeton University Press.(Pulitzer Prize-winning biography.)

Pitkin, H.F. (1984) *Fortune is a Woman: Gender and Politics in the Thought of Niccolò Machiavelli*, Berkeley, CA: University of California Press.(Insightful analysis of the psycho- and sociological dimensions of feminine/masculine symbolics in Machiavelli's writings and their implications for his political thought.)

Pocock, J.G.A. (1975) *The Machiavellian Moment: Florentine Republic Thought and the Atlantic Republican Tradition*, Princeton, NJ: Princeton University Press.(An innovative study of Florentine thought in the era of Machiavelli, and its anglicization as 'neo-Machiavellian' in the eighteenth century.)

Schevill, F. (1936) *History of Florence from the Founding of the City Through the Renaissance*, New York: Ungar Publishing Company.(A classic study of Florence and Florentine culture.)

Skinner, Q. (1978) *The Foundations of Modern Political Thought*, vol. 1, *The Renaissance*, Cambridge: Cambridge University Press.(A pathbreaking work that situates and analyses Machiavelli's writings in the context of Florentine and Renaissance humanist traditions.)

Skinner, Q. (1981) *Machiavelli*, New York: Hill & Wang.(An accessible and instructive introduction to Machiavelli's political thought.)

Strauss, L. (1958) *Thoughts on Machiavelli*, Chicago, IL: University of Chicago Press.(A textual reading of the blasphemous and dangerously seductive nature of Machiavelli's political writings as deviations from the Platonic-Christian tradition.)

MacIntyre, Alasdair (1929-)

Alasdair MacIntyre has contributed to the diverse fields of social, moral and political philosophy. He is one of the leading proponents of a virtue ethical approach in moral philosophy, part of a wider attempt to recover an Aristotelian conception of both morality and politics. His return to ancient sources has been powered by a critical indictment of the modern moral predicament, which MacIntyre regards as theoretically confused and practically fragmented; only a return to a tradition which synthesizes Aristotelian and Augustinian themes will restore rationality and intelligibility to contemporary moral and political life.

MacIntyre's long career culminated in the trilogy of works *After Virtue*, *Whose Justice? Which Rationality?* and *Three Rival Versions of Moral Enquiry*. These works claim that contemporary moral and political philosophy analyses incoherent fragments of a Judaeo-Christian theistic ethic that has lost its point with the increasing secularization of modern culture, leading to practical fragmentation and theoretical incoherence. MacIntyre's aim is to reconstruct a purpose and context for moral thought from the fragments of coherent moral life that survive only in marginal communities alienated from the main currents of the modern world.

This project draws on both philosophy and history. MacIntyre argues that the philosophical justification of rival positions in both theoretical and practical contexts involves narrative explanation. Such narratives set out how competing traditions of inquiry develop from, and are opposed to, each other. A theory is demonstrably superior to another if it explains both the successes and the failures of the previous theory. This form of justification requires an evaluatively engaged historical inquiry, whether applied to scientific theories or moral views.

Underpinning this methodology is MacIntyre's view of rationality and agreement. Rational agreement or disagreement is only possible in the context of a fixed framework of inquiry sharing paradigm uses of concepts and rational procedures. This kind of framework is socially and historically embedded in traditions of inquiry. Thus within or across traditions that can re-interpret each other's terms, rational disagreement is possible. When this is not so, there is interminable disagreement, a failure of rationality, and 'incommensurability'.

MacIntyre's methodology sets the stage for his distinctive history of moral philosophy in the West. Moral life was once rational and unified, in the ancient Greek *polis*, as described in the ethical and political works of Aristotle. This was guaranteed by the relation Aristotle envisaged between his account of human beings as they are, the set of prescriptions directing them as to how to realize their essential nature, and the ideal of human excellence that would be the result (see Aristotle). The discrediting of Aristotle's teleological metaphysics led to the loss of a sense that a human life had a unified purpose which could ground the value of that life. Thus all subsequent moral philosophy in the West is in MacIntyre's view a fruitless search for a means to connect the original elements of Aristotle's tripartite schema, once the idea of the purpose of an individual life has been lost.

This analysis culminates in MacIntyre's apocalyptic account of the contemporary moral situation, which he views as in complete intellectual disorder. The reasons for this disorder are threefold. First, some traditions of thought are just incoherent: they fail to achieve the internal integrity and coherence required in MacIntyre's account of inquiry. Second, even if forms of inquiry are coherent, conflicts among them are pointless since without a matrix of commensurability disagreement across traditions is in principle irresoluble. Third, the contemporary world fails to recognize the intellectual authority of the Neo-Augustinian Catholicism MacIntyre views as uniquely privileged in its relation to other traditions. MacIntyre criticizes contemporary liberalism for vainly aspiring to this role of adjudicating disputes within moral and political culture from a privileged standpoint. Liberalism attempts to regulate this disagreement in the name of higher ideals or privileged procedures, but in MacIntyre's view liberalism has no such intellectual authority over other traditions.

MacIntyre proposes that we resurrect Aristotelian ethics in the modern world. The key element is a new account of the overall purpose of a human life, which permits the restoration of an ethic of virtue, and the cultivation of appropriate forms of community. Politically, MacIntyre is a utopian, seeking the development of small communities analogous to the ancient *polis*, united by a common conception of the good life.

MacIntyre's project has met with a mixed reception in contemporary philosophy; his sharply critical tone, particularly directed against liberalism, has met with robust criticism in its turn. The first line of criticism

challenges the methodological underpinnings of MacIntyre's account: it is argued that MacIntyre has no principled way of distinguishing between the disagreement to which he assigns a positive role within moral traditions and the 'incommensurable' disagreement between traditions he criticizes. In response to this charge MacIntyre has developed a distinctive account of the relation between translation and interpretation, arguing that partial failures of translatability across languages illustrate the type of incommensurability he is concerned with (see Radical translation and radical interpretation).

Second, MacIntyre's account of morality has been accused of being both reductive and foundationalist: we should not see our task in moral philosophy as constrained by MacIntyre's tripartite schema, fitting moral injunctions as a whole to a non-moral foundation in 'human nature'. MacIntyre views the relation between the three elements of the tripartite schema as internally related, unintelligible in isolation; but this internal connection would only be possible in an ideal life. Similarly, MacIntyre's argument does not rest on a prior conception of human nature that constrains possible developments of morality: reflective moral life may be a partial determinant of this 'essence'.

The third criticism is that MacIntyre is actually far more modern than he seems, and rhetoric aside he is as committed to the values of modern moral and political philosophy as any modern liberal. These critics point out that MacIntyre is committed to such values as autonomy and positive freedom, and exaggerates the difference between ancient and modern ethical outlooks.

Their case is strengthened by the final line of criticism, which challenges MacIntyre's historical narrative, and presents an alternative account of the history of moral thought in the West that offers a better explanation than MacIntyre's version. Philosophers such as Charles Larmore and J.B. Schneewind have presented a view of modern moral life more hospitable to the rise of modern liberalism, and Charles Taylor's *Sources of the Self* (1989) offers an alternative view of modern morality far more sympathetic than MacIntyre to the values of the modern world. This third line of criticism is the most powerful since it matches MacIntyre's strengths of combining philosophical insight with wide historical learning.

See also: Virtue ethics (§5)

ALAN THOMAS

List of works

MacIntyre, A. (1977) 'Epistemological Crises, Dramatic Narrative, and the Philosophy of Science', *The Monist* 60: 433-72.(An indispensable guide to MacIntyre's early methodology.)

MacIntyre, A. (1981) *After Virtue: A Study in Moral Theory*, London: Duckworth, 2nd edn, with Postscript, 1985. (Clearest statement of the whole project and the account of virtue and the good life. Postscript is a valuable clarification.)

MacIntyre, A. (1984) 'The Relationship of Philosophy to Its Past', in R. Rorty, J.B. Schneewind, Q. Skinner (eds) *Philosophy in History*, Cambridge: Cambridge University Press.(Further methodological reflections on the creative synthesis envisaged between philosophy and history.)

MacIntyre, A. (1985) 'Relativism, Power, and Philosophy', *Proceedings and Addresses of the American Philosophical Association*, Newark, DE: American Philosophical Association, 5-22.(Argues for a principled distinction between the disagreement internal to, and between, traditions. Contests the challenge to this view from Donald Davidson's views on language and translation.)

(1988) *Whose Justice? Which Rationality?*, London: Duckworth.(Extended the scope of the argument from the theory of the good to the theory of practical reasoning. Contains the clearest statement of MacIntyre's criticism of liberalism.)

MacIntyre, A. (1990) *Three Rival Versions of Moral Enquiry*, London: Duckworth.(A return to more methodological reflections, contrasting MacIntyre's methodology with that of post-Enlightenment political rationalism and Nietzschean genealogy.)

References and further reading

Horton, R. and Mendus, S. (1994) *After MacIntyre*, Cambridge: Polity Press.(Multi-author collection of critical essays on MacIntyre's work. Also contains an invaluable reply from MacIntyre clarifying several misunderstandings of his project. Valuable bibliography of MacIntyre's published work.)

Larmore, C. (1987) *Patterns of Moral Complexity*, Cambridge, Cambridge University Press.(Contains an extended discussion of MacIntyre's project, and criticizes both MacIntyre's epistemological foundationalism

and political romanticism. The latter claim should be balanced by a reading of MacIntyre's 'Reply' in the Horton and Mendus volume.)

O'Neill, O. (1989) 'Kant After Virtue', in *Constructions of Reason*, Cambridge: Cambridge University Press.(A criticism of MacIntyre's reading of Kant, typical of corrections of MacIntyre's treatment of individual figures. Also contains a valuable discussion of MacIntyre's attitude to modernity.)

Schneewind, J.B. (1982) 'Virtue, Narrative and Community: MacIntyre and Morality', *Journal of Philosophy* 79: 653-63.(Schneewind's papers are critical but constructive, and suggest the outlines of Schneewind's alternative narrative of the moral philosophy of the modern period.)

Schneewind, J.B. (1983) 'Moral Crisis and the History of Ethics', *Midwest Studies in Philosophy*, 8: 525-39.(See Schneewind (1982) above.)

Taylor, C. (1989) *Sources of the Self*, Cambridge: Cambridge University Press.(A magisterial survey of the historical development of the modern moral identity, indebted to MacIntyre for its general methodology, but very different in its assessment of modern moral philosophy and its resources. An optimistic work which is an essential counterweight to MacIntyre's views.)

Mādhava (d. 1386)

Mādhava was a minister, scholar and philosopher in India in the fourteenth century. He gave support and advice at the founding of the Vijayanagara Empire in southern India, which lasted 300 years. He is best known for his Sarvadarśanasaṃgraha (Survey of the Major Philosophical Systems). In it, he presented sixteen systems of philosophy, starting with the materialists, discussing the Buddhists, Jainas, several Hindu schools and the school of Logic (Nyāya), and ending with Advaita Vedānta. To him and to many scholars at that time, Advaita Vedānta was the most complete and sophisticated philosophy. Mādhava is often identified with other people, especially Vidyāraṇya, the saint and abbot of a monastery in Śṛōgeri. The general view now is that Mādhava may have become a monk, receiving as a religious name that of Vidyāraṇya. This Vidyāraṇya was also an accomplished philosopher. His Pañcadaśī (Fifteen Chapters), a digest of Advaita Vedānta, is still popular today.

1 Mādhava and his times

Mādhava was a statesman, philosopher and polyhistor who lived in southwest India in the fourteenth century. Four names are associated with this one person. After examination of the various hypotheses regarding these names, the most frequent solution seems to be to regard two of them, Mādhava and Vidyāraṇya, as the civil and religious names respectively of the same person, and to treat Bhāratītīrtha as the name of his teacher and Sāyaṇa as the name of his brother. This opinion ignores the indigenous tradition.

By the fourteenth century, the Muslims were well settled in the north of India and were making almost continuous efforts to occupy the south as well. The Vijayanagara empire, founded by the brother kings Bukka I and Harihara I in the fourteenth century, resisted the invasions for a long time, flourishing until the middle of the seventeenth century. It was based in a region south of the rivers Krishna and Tungabhadra. At that time the empire was a bastion of the best of Hindu culture, the intellectual strength and growth of which was reflected in architecture, a variety of arts, and learning.

Mādhava was one of the primary instigators of this cultural tradition. He spent part of his life at the court as an advisor to the brother kings. It is speculated that he may have left to become a monk at the monastery in Śṛōgeri ('The Mountain Peak'), a few miles south of the capital, Vijayanagara ('The Victorious City'). When he entered the monastery, he received the religious name Vidyāraṇya ('The Forest of Wisdom or Knowledge'). Adding a religious name to a civil one was unusual.

Traditional Indian sources do not identify the two personages. On the contrary, they insist on their difference. The abbots at Śṛōgeri entered the monastery as celibate young men. It would be exceptional that a mature person like Mādhava, who had spent his life engaged in worldly affairs, should later have become an abbot.

2 Mādhava's works

The ascription 'Mādhavīya' ('of Madhava') occurs in the titles of a number of works. We know from inscriptions that grants were given to learned men (pandits) to help produce commentaries on the Vedic tradition. Mādhava, acting at his kings' behest to propagate the Hindu tradition, appointed several capable scholars to produce works, such as commentaries on the Vedas. Among them was Sāyaṇa, who seems to have been a leader of a writing guild. Mādhava was the mastermind and 'purse' behind this voluminous body of work. His name lent authority, and he himself could also have contributed to these works. It is virtually impossible to determine their exact authorship or date.

Because of the uncertainty of authorship, it is also difficult to make a definitive list of works. It is possible that Mādhava co-authored chapters or even whole works. In the case of the *Sarvadarśanasaṃgraha (Survey of the Major Philosophical Systems)*, he may have shared a chapter with his brother, Sāyaṇa. Of the several works that are ascribed to Vidyāraṇya, the *Pañcadaśī (Fifteen Chapters)* may have been written by, or at least in collaboration with, Vidyāraṇya's teacher, the abbot Bhāratītīrtha. The *Jīvanmuktiviveka (Correct Insight into the Concept of Liberation-while-Alive)* is attributed to Vidyāraṇya.

Of some thirty works that are ascribed to Mādhava, Vidyāraṇya, Sāyaṇa and Bhāratītīrtha, several are still only in manuscript form. A number were edited either in part or in full in the nineteenth century and at the beginning of

the twentieth century. Some were translated into various Indian languages and several also exist in English today.

Some of the best-known works that have proved to be important for the development and preservation of the Hindu tradition are the *Sarvadarśanasaṃgraha*, the *Jīvanmuktiviveka*, the *Pañcadaśī* and the *Vivaraṇaprameyasaṃgraha (Compendium of Objects of Knowledge in the Vivaraṇa Tradition)*, to mention just a few. In the *Sarvadarśanasaṃgraha*, the author discusses sixteen philosophical systems, which he arranges according to their relation (or lack of it) with Vedānta philosophy. He starts with the materialists, who are the most alien to Vedānta, and ends with the Vedānta teaching. The materialists praise immediate enjoyment and have no appreciation for spiritual striving and growth, whereas Vedānta aims at spiritual accomplishment, with its final goal of liberation from repeated physical existences. Between these two systems, he presents, among others, those of the Buddhists, the Jainas and the Naiyāyikas (Logicians). The author shows himself here as an erudite scholar of philosophy, skilled in logic.

The *Jīvanmuktiviveka* deals with the state of being-liberated-while-still-alive. It is well known that the Buddha attained such a state of enlightenment and continued to teach for almost five more decades before he died. The idea that it is not only possible, but quite likely that a person will live for a long time after attaining full enlightenment is explained here. In this state, a person still functions normally, but is not subject to the usual desires and passions, and does not accumulate karmic impressions - either good or bad - that will result in future rebirth. After gaining the insightful experience, the person still lives on, just as a potter's wheel keeps turning even though the pot is already completed. The author supports his statements with a number of quotations from the most essential sources of Hindu ideology: the *Bhagavad Gītā*, the Upaniṣads, the *Yogasūtra*, and so on. Many consider this work highly eclectic.

The *Pañcadaśī* presents in an accessible manner the main teachings of Vedānta and is still well liked and much revered. It is divided into three sections consisting of five chapters each. The first set deals with correct insight (*viveka*) leading to knowledge of reality, the second with consciousness which illuminates, and the third with the bliss experienced in the state of trance.

Vidyāraṇya's *Vivaraṇaprameyasaṃgraha* is a work of the greatest importance to Advaita dialectics. Vidyāraṇya shows himself here as an erudite philosopher arguing against his opponents' views. Although his is considered a commentary on Prakāśātman's commentary on Padmapāda's work, each of these is not a commentary as such, but rather an expansion and elaboration of the previous work in the same *vivaraṇa* ('laying out bare' - in other words, 'analysis') tradition of Vedānta. Padmapāda is considered the father of the Vivaraṇa school. His exposition is based on the first four sections of Śaōkara's *Brahmasūtrabhāṣya*.

Vidyāraṇya's *Vivaraṇaprameyasaṃgraha* deals with the method by which it is possible to examine true objects and distinguish between empirical and philosophical knowledge. In order to make these distinctions, it is essential that the objects of knowledge (*prameya*) be properly established. Vidyāraṇya was also a philosopher of language and had particular interpretations of traditional terms in Vedānta. He departed from the usual Vedāntic understanding of concepts such as self (*ātman*); according to him, Vedānta does not teach that the self continues to exist after the death of the physical body.

3 Advaita Vedānta refined and restated

Although many of these works show a thorough knowledge of all the schools of Indian philosophy, their primary focus is on the nondualist Vedānta of Śaōkara. This philosophy was concerned with finding the true nature of reality, and liberation. In general, medieval Indian thinkers did not aim at originality and innovation. Only those with strong minds and personalities changed the course of the teachings, giving rise to new varieties of Vedānta, such as the dualist school (see Vedānta §3). We tend to single out these thinkers and teachers, whereas we often do not know the names of those who stayed in the fold of the old tradition. We know of Mādhava and Vidyāraṇya because of their significant activities besides teaching Śaōkara's Vedānta. Mādhava, for instance, is notable for his well-rounded knowledge and active life as a statesman. Although they did not necessarily formulate original thoughts, Mādhava, Vidyāraṇya and numerous others contributed to keeping alive the Vedānta tradition by restating and refining it. Mādhava's activities helped to preserve the Hindu traditions by building a cultural and political stronghold against attacks and foreign influences.

EDELTRAUD HARZER CLEAR

List of works

Mādhava (14th century) *Sarvadarśanasaṃgraha (Survey of the Major Philosophical Systems)*, trans. E.B. Cowell and A.E. Gough, *The Sarva-darśana-saṃgraha or Review of the Different Systems of Hindu Philosophy*, Chowkhamba Sanskrit Series 10, Varanasi: Chowkhamba, 6th edn, 1961.(An excellent overview of the main ideas of the various philosophical schools. The work exercised considerable influence in the study of Indian philosophies.)

Vidyāraṇya (14th century) *Jīvanmuktiviveka (Correct Insight into the Concept of Liberation-while-Alive)*, ed. and trans. S. Subrahmanya Sastri and T.R. Srinivasa Ayyangar, Madras: The Adyar Library and Research Centre, 1978.(The text and translation deal with the issues of attaining freedom from rebirth and place emphasis on yogic practices.)

Vidyāraṇya (14th century) *Vivaraṇaprameyasaṃgraha (Compendium of Objects of Knowledge in the Vivaraṇa Tradition)*, trans. S.S. Suryanarayana Sastri and Saileswar Sen, Kumbakonam: The Sri Vidya Press, 1941.(An accessible translation of an important philosophical work so far neglected. Among other things, the main focus is on establishing the objects of knowledge according to Vedānta and refuting the notions of the Buddhists and others.)

Vidyāraṇya (14th century) *Pañcadaśī (Fifteen Chapters)*, trans. Hari Prasad Shastri, *Panchadasi: A Treatise on Advaita Metaphysics by Swami Vidyaranya*, London: Shanti Sadan, 1956.(A popular and authoritative guide to the teachings of Advaita Vedānta.)

Vidyāraṇya (14th century) *Praṇavamīmāṃsā*, ed. and trans. P. Olivelle, '*Praṇavamīmāṃsā*: A Newly Discovered Work of Vidyāraṇya', *Annals of of the Bhandarkar Oriental Research Institute* 62, 1981.(Translation of a text on the sacred syllable OM, after the *Praṇavakalpa* of the Skandapurāṇa.)

References and further reading

Fort, A. (1996) 'Liberation While Living in the *Jīvanmuktiviveka*: Vidyāraṇya's "Yogic Advaita"', in A. Fort and P. Mumme (eds) *Living Liberation in Hindu Thought*, Albany, NY: State University of New York Press.(Points to the focus on the use of yoga and also shows the dependence on the *Laghu Yogavāsiṣṭha (The Abridged Yogavāsiṣṭha)*, almost functioning as a commentary on this text.)

Hacker, P. (1978) 'Zur Geschichte und Darstellung des Hinduismus' (On the History and Interpretation of Hinduism), in *Kleine Schriften (Minor Works)*, Wiesbaden: Franz Steiner.(The new hypothesis here is that it was Vidyāraṇya who started the 'Śaōkara' monasteries.)

Kripacharyulu, M. (1986) *Sāyaṇa and Mādhava-Vidyāraṇya: A Study of Their Lives and Letters*, Guntur: Rajyalakshmi Publications.(A challenge to the prevailing opinions. The author amasses a lot of information, especially epigraphic, but references and annotations are inadequate.)

Mahadevan, T.M.P. (1957) *The Philosophy of Advaita with Special Reference to Bhāratītīrtha-Vidyāraṇya*, Madras: Ganesh & Co.(A clear account of Vidyāraṇya's metaphysics and epistemology.)

Mishra, G. (1984) 'Vidyāraṇya on Method, Object and Limit of Philosophical Investigation', *Indian Philosophical Quarterly* 11 (3).(A study of Vidyāraṇya's philosophical method in the *Vivaraṇaprameyasaṃgraha* showing that his is a logico-linguistic analysis.)

Sprockhoff, J.F. (1964, 1970) 'Der Weg zur Erlösung bei Lebzeiten, ihr Wesen und ihr Wert, nach dem *Jīvanmuktiviveka* des Vidyāraṇya' (The Path to Liberation-while-Alive, its Nature and Significance, in the *Jīvanmuktiviveka* of Vidyāraṇya), *Wiener Zeitschrift für die Kunde Süd- und Ostasiens* 8 (1964), 14 (1970).(A summary of the text and an excellent analysis of the ideas.)

Madhva (1238?-1317?)

Madhva, Hindu theologian and ascetic, founded the philosophical school commonly called Dvaita Vedānta, but which Madhva and his followers termed tattvavāda*, or realism. The name Dvaita refers to Madhva's dualistic interpretation of the Hindu canonical texts known as the Upaniṣads, also known as Vedānta. In contrast to the monist and semi-monist systems of his two major Vedāntin predecessors, Śaōkara and Rāmānuja, Madhva asserted the absolute difference between God (*īśvara*) and human souls (*jīva*), claiming that they were uncreated, eternal principles with fundamentally distinct natures. Madhva delineated the respective natures of God and souls so as to assert God's complete transcendence of the world and to legitimate the practice of devotion as the principal means of attaining liberation from the cycle of rebirth (*saṃsāra*). Madhva's realist epistemology served as the foundation for this ontological emphasis on difference (*bheda*).*

Our knowledge of Madhva's life comes from hagiographical literature and from scattered references to personal experience in his works. While traditions differ regarding his dates, inscriptional evidence corroborated with statements by Madhva make the dates 1238-1317 most likely. Madhva was born near Udupi in the present-day Indian state of Karnataka, where he established a monastic compound which continues to function as the ritual centre of modern Mādhvaism. Madhva presented himself as an incarnation of the Hindu god Vāyu who would save humankind through correct knowledge. This unique claim for a Hindu thinker has prompted some scholars to speculate that Madhva was influenced by Christianity. In fact, Madhva's milieu was heavily Jaina and his epistemology and ontology arguably reflect this influence.

Madhva's emphasis on the ontological significance of difference and its counterpart, individuality, was supported by his realist epistemology, which asserted the inherent validity of experience wherein difference is presented as an undeniable fact. Madhva used this ontological framework to support his sectarian devotion to the Hindu god Viṣṇu and articulated in many of his writings those unique qualities which, he believed, render Viṣṇu superior to all other Hindu gods.

Madhva's emphasis on the ontological significance of difference (*bheda*) was a direct challenge to his famous predecessor Śaōkara, a monist who claimed that the individual human soul was completely identical with the Godhead at the level of ultimate truth. By contrast, Madhva divided reality into two separate, uncreated principles, God (*īśvara*) and matter (*prakṛti*), and further subdivided matter into sentient matter, or souls (*jīva*), and insentient matter (*jaḍa*). Differences exist between these three categories and between individual members of the same category, resulting in a fivefold schema of difference: souls-God, souls-souls, souls-insentient matter, insentient matter-insentient matter, insentient matter-God; God cannot differ from himself. According to this view, no two things are identical, although their difference might be extremely subtle. Madhva rejected the Nyāya-Vaiśeṣika position on universals (*sāmānya*), arguing that universalized conceptions ultimately rest on perceived similarities (*sādṛśya*) between particular objects (see Universals, Indian theories of §2).

Madhva saw this particularity as resulting from the fact that qualities are completely identical with their substances. However, Madhva wanted to avoid taking the Buddhist position, which states that since all entities are nothing more than their qualities, they are impermanent and inessential. To distinguish between qualities and substances while upholding the notion that qualities are essential to the substance, Madhva posited the concept of *viśeṣa*, often translated as distinction. In Madhva's system, these distinctions exist between the entity and its attributes in such a way that the attributes are nevertheless integrated into the entity. Thus, the concept of *viśeṣa* allows an entity to possess different, often changing attributes without compromising the object's inherent unity or ontological validity. This challenging doctrine engaged many of Madhva's commentators.

Viśeṣas play an important role in the relationship between God and the world. One of God's permanent attributes is his independence (*svatantratva*), whereas souls and insentient matter are permanently dependent (*paratantra*) upon God. While insentient matter is beginningless, God transforms it from an original chaos into an ordered reality, and finally returns it to its original form. God, however, remains unaffected due to his transcendence.

Souls, although eternal in nature and generally free in agency, depend upon God for their destinies, since God governs the law of karma and grants liberation through grace. While the recognition of one's dependence upon

God inspires one to acts of devotion (*bhakti*) which can lead to liberation, ignorance can lead to eternal damnation. Fatalistically, Madhva asserted that a given soul's capacity to recognize its dependence upon God was a result of its permanent attributes and thus predetermined. In accordance with the three traditional Hindu locales for rebirth, Madhva asserted a threefold gradation of souls: those destined for eternal liberation, those destined for eternal *saṃsāra*, and those damned to eternal hell. While God oversees this structure, he is not implicated in its apparent injustices, since it is the souls' inherent properties which bring these about. God merely metes out destinies in accordance with behaviour. Thus Madhva maintained God's supremacy both by removing him from the presence of evil and placing him in charge of retribution.

The most crucial element of Madhva's epistemology was the role he gave to the 'witnessing subject' (*sākṣin*), which coordinates the senses and illuminates their data. Madhva posited the *sākṣin* as infallible and able to pass final judgment on the validity of the data presented to it, thereby asserting the self-validation of experiential knowledge. Indeed, while Madhva advocated three means of knowledge (*pramāṇas*) - perception, inference and verbal testimony (especially sacred scripture) - he insisted that the latter two *pramāṇas* never be so construed as to contradict perception. This position was one of his principal attacks upon the Advaitins, who looked to statements in the sacred texts declaring the identity of the soul with God as proof of the unreality of ordinary experience. However, Madhva fiercely defended the sanctity of the revealed literature (Veda) due to its ability to grant access to privileged domains of knowledge such as the nature of duty (*dharma*) and liberation (*mokṣa*).

See also: Knowledge, Indian views of; Monism, Indian; Vedānta

VALERIE STOKER

List of works

Madhva (1238?-1317?) *Sarvamūlagranthāḥ (Collected Works)*, ed. Bannanje Govindacarya, Udupi: Akhila Bharata Madhva Maha Mandala Publication, 1980.(A Sanskrit edition of Madhva's collected works; translations of the most important of these are listed below.)

Madhva (1238?-1317?) *Anuvyākhyāna (Commentary [on the Brahmasūtra])*, trans. S. Siauve, *Les Noms védiques de Viṣṇu dans l'Anuvyākhyāna de Madhva, Brahma Sūtra I.1.1, adhikaraṇa 2 à 12*, Pondicherry: Institut Français d'Indologie, 1959; trans. S. Siauve, *Les Hierarchies spirituelles selon l'Anuvyākhyāna de Madhva III.2.121-167; III.3.68-123, 193-203; III.4.4-51, 75-111, 232-278*, Pondicherry: Institut Français d'Indologie, 1971.(Siauve 1959 is a translation of those portions of the *Anuvyākhyāna* that discuss Viṣṇu's Vedic names; Siauve 1971 is a translation of those portions bearing on the topic of the gradation of souls.)

Madhva (1238?-1317?) *Brahmasūtrabhāṣya (Commentary on the Brahmasūtra)*, in B.N.K. Sharma, *The Brahmasūtras and their Principal Commentaries: A Critical Exposition*, Bombay: Bharatiya Vidya Bhavan, 1978.(Provides translations of key sections of Madhva's *Brahmasūtrabhāṣya* and places these in comparative perspective with other Vedāntin commentaries.)

Madhva (1238?-1317?) *Daśaprakaraṇāni (Ten Treatises)*, ed. P.P. Lakshminarayana Upadhyaya, Madras: Dharmaprakash Press, 1969, 4 vols.(A Sanskrit edition of Madhva's ten works that are not commentaries but independent treatises; includes commentaries by Vyāsatīrtha, Rāghavendratīrtha, Srinivāsatīrtha, Jayatīrtha and Vedeṣabhikṣu, important followers of Madhva.)

Madhva (1238?-1317?) *Mahābhāratatātparyanirṇayaḥ*, trans. K.T. Pandurangi, Chirtanur: Sriman Madhva Siddhantonnahini Sabha, 1993.(Translation of portions of Madhva's commentary on the central story of the *Mahābhārata*.)

Madhva (1238?-1317?) *Srimadviṣṇutattvavinirṇaya*, trans. S.S. Raghavachar, *An Examination of the Truth of Vishnu*, Mangalore: Sharada Press, 1971.(Very clear translation of an accessible Madhva text that outlines his basic philosophy; one of the 'Ten Treatises'.)

Madhva (1238?-1317?) *Upaniṣadbhāṣyāḥ (Commentaries on the Upaniṣads)*, esp.: Rai Bahadur Sris Chandra Vasu (trans.) *Bṛhadāraṇyaka Upaniṣad with the Commentary of ŚrīMadhvāchārya, Called also Anandatīrtha*, Allahabad: The Panini Office, 1916; Major B.D. Basu (trans.) *The Aitareya Upaniṣad (with Madhva's Commentary)*, Sacred Books of the Hindus, Allahabad: The Panini Office, 1925.(Translations of Madhva's commentaries on selected Upaniṣads, the later Vedic texts that form the ideological basis for the Brahmasūtras.)

References and further reading

Sharma, B.N.K. (1961) *The History of the Dvaita School of Vedānta and its Literature*, Delhi: Motilal

Banarsidass.(Crucial survey of almost all known literature of the Dvaita school.)

Sharma, B.N.K. (1961) *Sri Madhva's Teachings in his own Words*, Bombay: Bharatiya Vidya Bhavan.(Provides clear translations of portions of Madhva's important works in explanatory contexts.)

Sharma, B.N.K. (1962) *The Philosophy of Sri Madhvācārya*, Bombay: Bharatiya Vidya Bhavan. (Survey of Madhva's teachings.)

Siauve, S. (1968) *La Doctrine de Madhva*, Pondicherry: Institut Français d'Indologie. (Clear explanation of Madhva's central teachings.)

Viṣṇudāsācārya (1390?) *Vādaratnāvali*, trans. E. Gerow, *The Jewel Necklace of Argument (The Vādaratnāvali of Viṣṇudāsācārya)*, New Haven, CT: American Oriental Society, 1990.(Thorough translation of a later Dvaita work.)

Zydenbos, R. (1991) 'On the Jaina Background of Dvaita Vedānta', *Journal of Indian Philosophy* 19: 250-71. (Well-documented discussion of the Jaina influence on Dvaita.)

Mahāvīra (6th-5th century BC)

Mahāvīra's significance for Jaina philosophy is comparable to that of his contemporary, Buddha, for Buddhist philosophy. Both are regarded as the source of ideas, concepts and categories with far-reaching implications for later philosophical activity. In their respective traditions, both Mahāvīra and Buddha are recognized as enlightened or omniscient beings because they grasped the essential nature of reality, human life and the world. The teachings ascribed to them were at first passed on orally and were compiled into their present form several centuries later.

Despite similarities between Mahāvīra and Buddha, both in their biographies and in their rules for ascetics, Jainism is clearly an independent stream which has made its own valuable contribution to the development of Indian philosophy. One significant difference between Buddha and Mahāvīra is that Mahāvīra was not the founder of a new movement, but rather a reformer of the teachings of his predecessor, Pārśva.

'Mahāvīra' (literally 'Great Hero') is an epithet used especially for the Jaina saviour Vardhamāna. His dates are a matter of considerable debate, partly because they are related to those of the Buddha's, which are also problematic; neither wrote any works, which otherwise might have helped with deciding their dates. According to the Śvetāmbara sect of Jainism, Mahāvīra died at the age of seventy-two in 527 BC, whereas the Digambara sect gives the date as 510 BC, and recent research suggests 477 BC. The source of Mahāvīra's teachings is a body of scripture which was redacted in its present canonical form by a council held in Valabhī, Gujarat, around the fifth century AD. The Digambara sect rejects this Śvetāmbara version, believing the original canon to have been irretrievably lost, although it does have texts regarded as procanonically sacred. The style of both sets of texts is very repetitive, and whatever has a direct bearing on philosophy has to be extracted from a huge volume of material on a wide range of other themes, such as the religious practices of ascetics and the conduct of monks and nuns. The teachings are generally in the form of questions put by disciples and answers given by Mahāvīra or one who has heard his teaching, with little, if any, philosophical discussion of the concepts.

The authority which Mahāvīra's teaching enjoys in the Jaina tradition is inextricably related to the position given to him and twenty-three others as universal teachers. Historicity is ascribed only to the last two, Pārśva and Mahāvīra. All have the epithet 'Jina', meaning 'Victor' or 'Conqueror' of the passions fettering an individual. The tradition acknowledges them as omniscient beings who have obtained enlightenment through extreme asceticism. Their teachings are like a ford that enables others to cross to the other side of human existence, and hence they are called 'Tīrthaṃkaras', 'Fordmakers'.

Mahāvīra's teaching and biography represent an ascetic (*śramaṇa*) tradition that was already well established during his lifetime, one to which even his predecessor Pārśva belonged 250 years earlier. The exact origins of Indian asceticism are difficult to trace, but it is possible that there were world-renouncers before the arrival of the Aryans around 1700 BC. Mahāvīra's teaching characterizes this ascetic tradition in that he too holds that life in the world intrinsically entails suffering, ignorance and pain, which may be overcome through the knowledge of reality and of one's essential nature that is attainable in an ascetic life. This basic teaching is associated with the theory of karma, whereby karma is responsible for the cycles of existence (*saṃsāra*) that living beings have to undergo until they realize how, through Mahāvīra's teachings, accumulated karma may be destroyed and future ones avoided. In its simplest form, the theory, as it applies to human beings, is presented through an agricultural metaphor: one's actions, including thoughts and words, are said to leave behind traces which, like seeds, bear fruit in an appropriate environment either in one's present existence or a future one, the former being also determined by previous karma. This theory, shared by Buddha and the Upaniṣad thinkers, is the hallmark and essential starting point of all soteriological Indian systems and seems to have been pan-Indian during Mahāvīra's time, which is perhaps why no philosophical justification is offered for its acceptance in his teachings (see Karma and rebirth, Indian conceptions of).

Mahāvīra's interpretation of karma is unique, in that it is seen as particles of subtle matter which attach themselves to the soul, hindering the development of what later thinkers called the three jewels of Jainism, namely the soul's innate powers of right belief, right knowledge and right conduct. Mahāvīra mentions killing or violence (*hiṃsā*) as the cause of karma and says that the practice of nonviolence (*ahiṃsā*) is the common doctrine of all Jaina

fordmakers: ‘The Arhats and Bhagavats of the past, present, and future, all say thus, speak thus, declare thus, explain thus: all breathing, existing, living, sentient creatures should not be slain, nor treated with violence, nor tormented, nor driven away.’

It was not until around the fourth century AD that Mahāvīra’s metaphysics, epistemology, ethics and cosmology, unsystematically scattered around in the various canonical works, were presented in a manner comparable to the basic philosophical texts of the Hindu systems. This was done by Umāsvāti (also called Umāsvāmī) in his *Tattvārthasūtra*.

See also: Jaina philosophy; Manifoldness, Jaina theory of

JAYANDRA SONI

References and further reading

Deleu, J. (1970) *Viyāhapannatti (Bhagavaī): the Fifth Aōga of the Jaina Canon*, Brugge: De Tempel.(A philological analysis of the Prakrit work also known as the *Bhagavatī*, which is an important source for early Jaina philosophy.)

Dundas, P. (1992) *The Jains*, London: Routledge.(An excellent modern survey of the history of Jainism to the present time, based on original sources.)

Glasenapp, H. von (1925) *Der Jainismus: Eine indische Erlösungsreligion (Jainism: An Indian Liberation Religion)*, Berlin: Alf Häger Verlag.(One of the earliest general surveys of Jainism in the West and still a standard basic work; not yet translated into English.)

Jacobi, H. (trans.) (1884) *Jaina Sūtras: Part I*, Sacred Books of the East, Oxford: Oxford University Press; repr. Delhi: Motilal Banarsidass, 1964.(Translation from the Prakrit of two Jaina canonical works, the *Ākārāōga Sūtra* (or *Ācārāōga Sūtra*) and the *Kalpa Sūtra*, with an introduction on the origin and development of Jainism.)

Jacobi, H. (trans.) (1895) *Jaina Sūtras: Part II*, Sacred Books of the East, Oxford: Oxford University Press; repr. Delhi: Motilal Banarsidass, 1964.(Translation from the Prakrit of two Jaina canonical works, the *Uttarādhyayana Sūtra* and the *Sūtrakritāōga Sūtra*, with an introduction on the Jaina records of the philosophical doctrines of the heretics.)

Jain, H. and Upadhye, A.N. (eds) (1939-59) *Ṣaṭkhaṇḍāgama (Scripture in Six Parts)*, Solapur: Jaina Samskṛti Samrakṣaka Saōgha, revised 3rd edn, 1992, 16 vols.(An important eighth-century Digambara work, the composition of which is based on the preaching of Mahāvīra; also includes Vīrasena’s commentary, *Dhavalā*.)

Ohira, S. (1994) *A Study of the Bhagavatīsūtra. A Chronological Analysis*, Ahmedabad: Prakrit Text Society. (The *Bhagavatīsūtra* is a Jaina canonical work; this book is a systematic and historical analysis of the first half of the text - chapters 1-20, regarded as the oldest part - and includes a survey of currents of thought in the canonical period.)

Punyavijaya, Muni. *et al*. (eds) (1971) *Paṇṇavanāsuttam* Jain Agama Series 9.2 (*Prajñāpanā Sūtra*) , Bombay: Mahāvīra Jaina Vidyālaya. (A more systematic presentation of the contents of the fifth Aōga above.)

Schubring, W. (1926) *Worte Mahāvīras. Kritische Übersetzung aus dem Kanon der Jaina (Mahāvīra’s Words. Critical Translation from the Jaina Canon)*, Göttingen: Vandenhoeck & Ruprecht.(Mahāvīra’s teachings presented on the basis of original sources.)

Schubring, W. (1935) *Die Lehre der Jainas*, Berlin and Leipzig: Walter de Gruyter & Co.; trans. W. Beurlen, *The Doctrine of the Jainas. Described after the Old Sources*, Delhi: Motilal Banarsidass, 1962.(This standard work is a survey of Jaina doctrine based on original sources.)

Schubring, W. (ed.) (1974) *Isibhāsiyāim, A Jaina Text of the Early Period*: Ahmedabad: L.D. Institute of Indology.(A relatively ignored ancient work, perhaps of the fourth century BC, the title of which literally means ‘Sayings of the Ṛṣis’; it records the views of Mahāvīra and his predecessor Pārśva, together with the views of over forty others, many of whom are legendary figures.)

Upadhye, A.N. *et al*. (eds) (1977) *Mahāvīra and his Teachings*, Bombay: Bhagavān Mahāvīra 2500th Nirvāṇa Mahotsava Samiti.(Contains a series of articles by eminent scholars, not only on Mahāvīra, but also on Jaina philosophy, religion, literature, history and art.)

Maimon, Salomon (1753/4-1800)

Educated as a rabbi in Lithuania, Shlomo (Salomon) ben Yehoshua migrated to Germany and adopted the surname Maimon in honour of Maimonides. His criticism of Kant's dualism and his monistic account of the human mind as an imperfect expression of God's infinite mind influenced Fichte, Schelling and Hegel. Kant regarded him as the critic who understood him best.

Maimon's system combines rational dogmatism with empirical scepticism. As a rational dogmatist, he argues that cognition requires the absolute unity of subject and object. Maimon therefore criticizes Kant's dualistic divisions between the mental form and extra-mental matter of knowledge, and between the faculties of sensibility and understanding. Experience in Kant's sense - empirical knowledge - is possible only if these dualisms are merely apparent. Our finite minds must be imperfect expressions of an infinite, divine mind that produces the form and matter of knowledge. Through scientific progress, our minds become more adequate expressions of the infinite mind. Kant has not refuted Hume's scepticism, which could be refuted only if science became perfect. Perfect science is an ideal for which we must strive but which we will never reach.

Maimon is deeply indebted to Maimonides, but he reformulates Maimonidean ideas in light of modern mathematical physics and deploys them within a Kantian investigation of the possibility of experience. The result is a unique encounter between medieval and modern philosophy that decisively influenced German idealism and remains philosophically interesting.

1 Life and works

Maimon was born in Sukoviborg, Lithuania, and educated in the Talmud. He studied Kabbalah, was attracted then repelled by early Hasidism (see Hasidism); and, under the influence of Maimonides, migrated to Germany in search of secular wisdom (see Maimonides, M.). Turned away from Berlin, he became a wandering beggar and then taught at Posen, where he composed several Hebrew works. Maimon finally entered Berlin *c.*1780 and was befriended by enlightened Jews associated with Mendelssohn (see Mendelssohn, M.; Enlightenment, Jewish). However, Mendelssohn advised him to leave Berlin because of the tensions arising from Maimon's reluctance to earn a livelihood and from his sensuous lifestyle. Maimon studied German and contemporary philosophy and science. After returning to Berlin, Maimon's critical notes on Kant's *Critik der Reinen Vernunft (Critique of Pure Reason)* were forwarded to Immanuel Kant by Markus Herz, Kant's student, physician and friend in 1789. Kant realized the work's excellence. The notes were published (1790), followed a year later by a Hebrew commentary on Maimonides (1791). These works laid the foundations for a decade of publication, during which Maimon contributed to almost every area of contemporary philosophy. Influenced by Maimonides's naturalistic account of prophecy, Maimon pioneered the scientific study of parapsychology (see Parapsychology). He also argued that Maimonides' ideas could solve contemporary philosophical problems. Maimon polemicized against Reinhold and Schulze, and exercised great influence upon Fichte. His financial problems were alleviated when Count Kalkreuth offered to house him, and he died in 1800 at Kalkreuth's estate in Nieder-Siegersdorf, where he was buried outside the Jewish cemetery. His autobiography (1792/3) provides important testimony about eighteenth-century Jewish life.

2 Criticism of Kant's dualism

Maimon severely criticized Kant's attempt to refute Hume's scepticism. Hume had argued that our claims to find necessary (for example, causal) connections between events are justified neither by logic (because there is no contradiction in denying such connections) nor by experience (because we experience what is, not what must be) (see Hume, D.). Kant responded that experience is impossible unless we regard such connections as necessary, which cannot be derived from logic or experience but are justified as necessary conditions for the possibility of experience. These necessary conditions are the categories of the understanding (see Kant, I.). However, Maimon argues that Kant understands experience differently from Hume. For Kant, experience is empirical knowledge - organized information about the sensible world. But for Hume, experience is an unorganized body of sensory impressions. Kant has not refuted Hume because Hume need not concede that we have experience in Kant's sense, so he need not concede that its necessary conditions are justified.

Furthermore, Maimon argues that Kant has failed to specify necessary and sufficient conditions for applying the categories to sensory impressions. Kant cannot explain how we apply the categories, whereas Hume explains that we claim necessary connections because we become accustomed to associations among events of certain types.

In Maimon's view, Kant's failure arises from the twofold dualism he attributes to the human mind. First, the mind is divided from the things in themselves that produce the matter of sensation. Second, the mind is divided into two faculties, each supplying its own form. Sensibility supplies space and time, while the understanding supplies the categories. Maimon argues that the categories can be justifiably applied only if the mind produces sensations, matter as well as form - thus, only if sensibility and understanding are not distinct. Empirical knowledge is possible if Kant's dualisms are overcome, if Kant's divergent mental faculties are only apparent features of the human mind, which is in reality an imperfect expression of the divine mind, generating both the form and the matter of experience.

To explain how experience is possible, then, Maimon rethinks Kant's central notions. He understands the thing-in-itself not as an extra-mental reality but as the reality known by the infinite mind. The thing-in-itself is the goal of scientific progress and not, as Kant thinks, the unknowable substratum of experience.

To overcome the dualism between sensibility and understanding, Maimon rejects Kant's view that space and time are non-conceptual forms of sensibility. Instead, he regards them as the minimal conceptual conditions for thinking diverse objects. Diverse objects must be the same in one respect and different in another, so two forms are required: the objects might occupy the same space at different times or different spaces at the same time. Whereas Kant takes space and time to be objects of intuition, Maimon regards them as concepts, misleadingly represented as objects of intuition by the imagination.

Maimon overcomes the dualism between sensations and the categories by drawing on the infinitesimal calculus as developed by Leibniz. He regards sensations as composed, not of passive affections, but rather of differentials or infinitesimal degrees of mental activity below the threshold of consciousness. Instead of pertaining to a distinct faculty of sensibility, these differentials are ideas of understanding: they are the elements that the understanding would ideally reach if it carried out an infinite analysis. For its part, the understanding seeks to know objects in their genesis, and its categories are rules for the production of objects. To categorize a sensation is to grasp the law-like relationship between its differentials (dy/dx) through which a sensed object is produced. There is no gulf between sensations and the understanding, so the categories can be applied to sensible objects - and experience is possible.

3 Rational dogmatism and empirical scepticism

To some of Maimon's contemporaries, these monistic revisions of Kant suggested a way to refute Hume. But Maimon thinks Hume cannot be refuted, for although empirical knowledge is possible, it is not actual. This empirical scepticism follows from Maimon's rational dogmatism about knowledge, which in turn arises from his unique combination of Maimonides's medieval Aristotelianism with an understanding of mathematics based on the modern accomplishments of Leibniz and Newton (see Maimonides, M.; Leibniz, G.W.; Newton, I.).

Maimon is dogmatic about the conditions prescribed by reason for the attainment of knowledge: knowledge requires the absolute unity of the subject, the object and the act of knowing. Thus in arithmetic, he claims, we produce mathematical objects (numbers) without any input from outside the mind; and there is no real distinction between our number-generating minds, the numbers generated, and the rules by which we generate them. Maimon combines this modern understanding of mathematics as creation with a medieval understanding of God. Maimonides attributed to the philosophers the (Aristotelian) dictum that God is the intellect as well as the intellectually cognizing subject and the intellectually cognized object. He also said that God is always engaged in actual knowing and is therefore always absolutely one, whereas we attain absolute unity only if we actualize our potential for knowledge. On Maimon's interpretation, God is the mathematical mind that generates the world, and we can know the world in so far as we understand the mathematical rules of its generation. Experience (empirical knowledge) is possible if we are made in God's image or, in Maimon's formulation, if the finite mind is the schema of the infinite mind. When we engage in mathematics, we are divine.

So Maimon rejects Kant's dualisms, not only because Kant's response to Hume is inadequate, but also because dualism is incompatible with knowledge. He believes experience can be actual only if we eliminate dualism from

science and become one with God. However, this is an ideal, for which we must strive but which we can never attain. So Hume is right to be sceptical about our claims to actual empirical knowledge.

Maimon's remarks about mathematics suggest that even mathematics is tainted by dualism and must be transformed before experience is actual. Symbolic proofs explain why premises imply a conclusion, but they may employ unreal concepts, such as 'regular dodecahedron'. In contrast, constructions in intuition, such as geometric figures, can show that a concept is real, but not why. The dualism of symbol and intuition must be eliminated, creating a mathematics that, at present, cannot even be imagined.

The moral of Maimon's scepticism is not despair but a commitment to scientific progress and intellectual honesty. We must strive for perfection without pretending we are already perfect (see Scepticism).

4 The categories and the principle of determinability

As rules of production, Maimon's categories differ from Kant's, which are empty forms derived from the forms of logical judgment. Dissatisfied with Kant's reliance on traditional logic, Maimon reconsiders Aristotle's pioneering account of the categories and logic. He undertakes to derive the categories and the forms of judgment from a first principle, rendering logic a systematic science at last.

He proposes the principle of determinability: in a judgment determining a real object, the subject can be thought without the predicate but not the predicate without the subject. The infinite mind may be conceived as a single chain of judgments, beginning with the principle of determinability, determining every subject by a unique predicate, and culminating in the thing-in-itself as the fully determinate object of experience. This monistic system is akin to that of Spinoza. But we cannot complete it - only approximate it as science progresses.

5 Practical philosophy

Maimon rejects Kant's doctrine of the primacy of the practical. This is the doctrine that reason seeks an absolute ground, but entangles itself in contradictions and illusions when it seeks that absolute ground through the scientific pursuit of knowledge, whereas reason finds an absolute ground in the practical pursuit of moral action. In Maimon's view, reason can find an absolute ground through the scientific pursuit of knowledge, if dualism is eliminated. Contradictions and illusions arise only when reason conflicts with imagination. Maimon's ethics are closer to Maimonides and Aristotle than to Kant, and he undertakes to derive morality from theoretical reason (see Maimonides, M.; Aristotle).

Maimon is committed only to religious doctrines reconcilable with reason. He finds Jewish monotheism more rational than Christianity and praises rabbinic morality, but criticizes the artificial development of Jewish law since the destruction of the Temple and does not oppose conversion to Christianity, if it is pursued for the sake of one's intellectual development. Maimon himself attempted conversion but, when he declared himself committed to rationalizing Christianity and uninterested in it for its own sake, he was rejected as too much of a philosopher to be become a Christian.

See also: Enlightenment, Jewish; German Idealism; Hume, D.; Jewish philosophy in the early 19th century; Kant, I.; Maimonides, M.

PAUL FRANKS

List of works

Maimon, S. (1965-76) *Gesammelte Werke (Collected Works)*, ed. V. Verra, Hildesheim: G. Olms, 7 vols.(A photoreproduction of all Maimon's German publications. No critical edition exists.)

Maimon, S. (1790) 'Versuch über die Transcendentalphilosophie mit einem Anhang über die symbolische Erkenntnis und Anmerkungen' (Essay on Transcendental Philosophy with an Appendix on Symbolic Cognition and Annotations), in *Gesammelte Werke (Collected Works)*, ed. V. Verra, Hildesheim: G. Olms, 1965-76, vol. 2.(Maimon's most influential work, written in admiration and sharp criticism of Kant's *Critik der Reinen Vernunft (Critique of Pure Reason)*.)

Maimon, S. (1791) *Giv'at ha-More (The Hill of the Guide)*, ed. S.H. Bergmann and N. Rotenstreich, Jerusalem: Israel Academy of Sciences and Humanities, 1965.(Introduction to the history of philosophy and an incomplete commentary on Maimonides's *Guide to the Perplexed*, the text that most influenced Maimon's life and work.

Develops Maimon’s central thoughts, such as the absolute unity of subject, act and object in knowledge, and criticizes Maimonides’s Aristotelian physics.)

Maimon, S. (1792/3) *Salomon Maimons Lebensgeschichte: Von ihm selbst geschrieben und herausgegeben von K. P. Moritz*, trans. J. Clark Murray, *The Autobiography of Solomon Maimon*, London: East and West Library, 1954.(An abridged translation, omitting ten chapters on Maimonides’s philosophy. Includes an essay on Maimon’s philosophy by S.H. Bergman, which provides a clear, short introduction.)

Maimon, S. (1794) ‘Briefe des Philaletes an Aenesidemus’, trans. and annotated by G. di Giovanni, ‘Letters of Philaletes to Aenesidemus’, in G. di Giovanni and H.S. Harris (eds and trans.) *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism*, Albany, NY: State University of New York Press, 1985.(A partial translation of Maimon’s reply to Aenesidemus-Schulze’s sceptical attack on Kant’s philosophy. Provides an excellent sketch of Maimon’s system and the distinctiveness of his scepticism.)

References and further reading

Altmann, A. (1973) *Moses Mendelssohn: A Biographical Study*, Philadelphia, PA: Jewish Publication Society, 360-4.(An excellent study of the enlightened Jewish circles in which Maimon moved in Berlin, with important details about Maimon’s relationship with Mendelssohn.)

Atlas, S. (1965) *From Critical to Speculative Idealism: The Philosophy of Salomon Maimon*, The Hague: Martinus Nijhoff.(A clear and comprehensive account of Maimon’s philosophy.)

Atlas, S. (1970) ‘Salomon Maimon’s Doctrine of Fiction and Imagination’, *Hebrew Union College Annual* 40/41: 363-89.(A clear account of a difficult but central area of Maimon’s philosophy.)

Baumgardt, D. (1963) ‘The Ethics of Salomon Maimon (1753-1800)’, *Journal of the History of Philosophy* 1(2): 199-210.(The only English article on the sadly neglected topic of Maimon’s ethics, exploring his increasingly sceptical attitude towards Kant’s critical philosophy.)

Beiser, F. (1988) *The Fate of Reason: German Philosophy from Kant to Fichte*, Cambridge, MA: Harvard University Press.(A clear and indispensable account of Maimon’s philosophy and his pivotal role in development of German idealism.)

Bergman, S.H. (1967) *The Philosophy of Salomon Maimon*, trans. N. Jacobs, Jerusalem: The Magnes Press.(The best comprehensive book on Maimon in English. An excellent introduction.)

Zac, S. (1988) *Salomon Maimon: critique de Kant*, Paris: Les Éditions du Cerf.(A useful study, particularly for its emphasis on Maimon’s conception of philosophy and his relation to Spinoza.)

Maimonides, Abraham ben Moses (1186-1237)

Jewish theologian, mystical pietist, physician, and the only son of Moses Maimonides, with whom he studied rabbinics, philosophy and medicine. Upon his father's death, Abraham became the spiritual and temporal head of Egyptian Jewry and a leading rabbinical authority. Using this position, he propagated a form of Jewish pietism, introducing ideas and ritual practices inspired by Islamic mysticism. Moving beyond defence of his father's legal and philosophical writings, Abraham's most important work, Kifayat al-'Abidin (Complete Guide for Devotees) - a monumental compendium of jurisprudence and religious philosophy - develops his own pietist interests. It attracted something of a following in its time but also met with concerted Jewish opposition to its Sufi-inspired ideals.

Born in Fustat, Egypt, Abraham showed exceptional gifts that were noted from an early age. In 1204, a time of temporal and spiritual upheavals, he was chosen, a youth of eighteen, to lead Egyptian Jewry, the first of his family to occupy the office of *rayyis al-yahud*, which would become hereditary and remain in the family for two centuries, in large part because of the prestige of their distinguished ancestor Moses Maimonides. An able administrator and a constructive activist as well as a theorist, Abraham enjoyed close relations with the Muslim authorities as court physician to the Ayyubid ruler al-Malik al-Kamil. Much of his writing, perforce, was devoted to defending his father's works against their rabbinical and philosophical detractors. His *Milhamot ha-Shem (Wars of the Lord)*, written after 1235, responded in particular to the criticisms of the rabbis of Provence.

Abraham wholeheartedly espoused his father's philosophical system and was seen as a continuator of the Arabic Aristotelian tradition. But his mature views diverged widely and creatively from such ideas. Moses Maimonides had considered the ultimate human aim to be knowledge of God, but the son stressed ethical perfection. His ascetic mysticism earned him the epithet by which he is consistently cited in later literature, Abraham *he-hasid* ('the Pious'). Identifying with a pietistic circle who were dissatisfied with the rationalism of some contemporary peripatetics, Abraham brought Sufi ideals into traditional Judaism, inspired by Abraham Abu'l-Rabi'a (d. 1223), also known as *he-hasid*, whom he calls 'our master in the Way'. He became the leader of this circle, whose adepts included his father-in-law, Hanan'el ben Samuel, and his own son 'Obadyah (1228-65), author of the mystical *al-Maqala al-Hawdiyya (The Treatise of the Pool)*.

Pressing his prerogative as *nagid* (communal leader), he tried to enforce pietistic practices on the larger community, much as his father had sought to introduce various ritual reforms. Abraham's far-reaching measures included such Islamic-influenced practices as ablutions of the feet before prayer, sitting in a dignified posture during prayer, standing in rows, raising the hands in supplication, and punctuating prayer with profuse prostrations.

Abraham defended his pietism in the *Kifayat al-'Abidin (Complete Guide for Devotees)* (written *circa* 1230). This monumental legal and ethical treatise is not extant *in toto*, but substantial manuscripts survive in Genizah collections. Regrettably, the final section, dealing with *wusul* (the mystical goal), seems to be lost.

Written in Arabic in a warm and attractive style, the work is well constructed. The early parts rehearse the legal rulings of Abraham's father, albeit in a distinctively spiritualized vein. But the fourth and final section, on the 'special way', highlights the virtues of *tariq* (the path): sincerity, mercy, generosity, gentleness, humility, faith, contentedness, abstinence, mortification and solitude. These are *maqamat* (stages or stations), well known in the classical Sufi manuals. The goal, *wusul* ('arrival'), is in both cases encounter with God and the certitude of his light.

Abraham openly admires the Muslim Sufis, whose practices, he claims, derive originally from the prophets of Israel (see Mystical philosophy in Islam). He even finds a Jewish origin for Sufi self-mortifications - combating sleep, solitary retreats in dark places, weeping, nightly vigils and daily fasts. Notable is the obligation of the novice to take as his guide an experienced teacher who has traversed all the stages of the path and will initiate the novice in the intricacies of mystical discipline before bestowing on him his mantle, as Elijah did on Elisha.

Departing from the juridical mode of his father's legal code, the *Mishneh Torah*, Abraham stresses the spiritual significance of the traditional Jewish precepts (*mitzvot*, divine commandments) and the 'mysteries' they conceal,

in much the same manner of al-Ghazali's classic Islamic summa, the *Ihya' 'ulum al-din (Revival of the Religious Sciences)* (see al-Ghazali). The idea that Abraham Maimonides had rediscovered lost mysteries of Jewish provenance in traditions preserved by the Sufis but long forgotten by the Jews in the tribulations of their exile allowed his circle to justify their adoption of Muslim customs and symbolisms. Calling themselves 'the disciples of the prophets', the Jewish pietists confidently awaited the imminent renewal of prophecy in Israel. The ancient Jewish traditions recovered from the Sufis were integral to the 'prophetic discipline'. Restoration of that discipline would hasten the return of prophecy itself, which Moses Maimonides had predicted in principle, but for a seemingly remote and certainly unspecifiable date.

Abraham had intended to compose a biblical commentary, but only the sections on Genesis and Exodus seem to have been completed, some time after 1231. Here, as in the *Kifayat*, he gives free reign to his mystical predilections, often, for polemical reasons, projecting into the patriarchal past his own innovations, depicting ancient biblical figures as pietists in much the way that Sufi literature clothes Muhammad and his companions in the garb of the early Sufis. He often alludes to an esoteric interpretation of the 'subtle mysteries' of the Pentateuch. These are not mentioned in his father's system. They typically point to pietist doctrines.

Although intended to enhance the spiritual ambience of Jewish worship with new forms of decorum like those to be observed among the Muslim neighbours of the Egyptian Jews, the novel practices Abraham Maimonides championed did not go unchallenged. Despite his office and familial prestige, which furthered pietist aims immensely, Abraham confronted fierce opponents, who went as far as to denounce him to the Muslim authorities, accusing the Jewish pietists of introducing 'false ideas', 'unlawful changes' and 'gentile customs' into the synagogue.

This opposition, and the spiritual elitism that had characterized the movement from its inception, may explain why it never gained universal approval. With the general decline of Oriental Jewry, Abraham Maimonides' construction of a Sufi-influenced Jewish pietism, gradually sank into oblivion.

See also: Maimonides, M.; Mystical philosophy in Islam

PAUL B. FENTON

List of works

Maimonides, Abraham ben Moses (*c*.1230) *Kifayat al-'Abidin (Complete Guide for Devotees)*, trans. Y. Duri, *Sefer ha-Maspiq le-'Ovdey ha-Shem*, Jerusalem, 1973; trans. N. Dana, *Kitab Kifayat al-'Abidin (Part Two, Volume Two), The Arabic Original with an Introduction and an Annotated Hebrew Translation*, Ramat-Gan: Bar-Ilan University, 1989; ed. and trans. S. Rosenblatt, *The High Ways to Perfection of Abraham Maimonides*, New York and Baltimore, MD: Johns Hopkins University Press, 1927-38.(The *Kifayat* is a code of laws with pietistic leanings. It is incompletely preserved. Dana's edition provides the Judaeo-Arabic original and Hebrew translation of the liturgical section of the work; Rosenblatt's edition comprises the Judaeo-Arabic text of the ethical section of the work, with English translation.)

Maimonides, Abraham ben Moses (after 1235) *Milhamot ha-Shem (Wars of the Lord)*, ed. R. Margaliot, Jerusalem: Mossad ha-Rav Kook, 1953.(Abraham Maimonides' defence of his father's work.)

Maimonides, Abraham ben Moses (before 1237) *Birkhat Abraham (Responsa of Abraham)*, trans. B. Goldberg, Lyck, 1859; ed. A. Freimann and S.D. Goitein, *Abraham Maimuni, Responsa*, Jerusalem: Mekize Nirdamin, 1937.(Legal correspondence of Abraham. The Freimann and Goitein edition comprises the Arabic text with Hebrew translation.)

Maimonides, Abraham ben Moses (before 1237) 'A Treatise in Defence of the Pietists', trans. S.D. Goitein, *Journal of Jewish Studies* 16, 1965: 105-114.(Polemical tract by Abraham.)

Maimonides, Abraham ben Moses (before 1237) *Ma'aseh Nissim (The Act of Miracles)*, trans. B. Goldberg, Paris: Brill, 1867.(Polemical. Abraham Maimonides' reply to Daniel ha-Babli's strictures on Moses Maimonides' code of laws, the *Mishneh Torah*.)

Maimonides, Abraham ben Moses (*c*.1237) *Abraham Maimonides Commentary on Genesis and Exodus*, trans. E. Wiesenberg, ed. S.D. Sassoon, London, 1958.(A pietistic commentary, infused with Sufi ideas and terminology. The Wiesenberg-Sassoon edition consists of a Hebrew translation of the work and the Arabic original.)

References and further reading

al-Ghazali (1096-7) *Ihya' 'ulum al-din (Revival of the Religious Sciences)*, Cairo: Matba'ah Lajnah Nashr al-Thaqafah al-Islamiyyah, 1937-8, 5 vols; partial translations can be found in E.E. Calverley, *Worship in Islam: al-Ghazali's Book of the Ihya' on the Worship*, London: Luzac, 1957; N.A. Faris, *The Book of Knowledge, Being a Translation with Notes of the Kitab al-'Ilm of al-Ghazzali's Ihya' 'ulum al-din*, Lahore: Shaykh Muhammad Ashraf, 1962.(Close in spirit to the *Kifayat*, of which it may have served as a model.)

Cohen, G. (1967-8) 'The Soteriology of Abraham Maimuni', *Proceedings of the American Academy for Jewish Research* 35: 75-98, 36: 33-56.(An account of Abraham's thought in its socio-religious context.)

Eppenstein, S. (1914) *Abraham Maimuni: sein Leben und seine Schriften*, Berlin: Louis Lamm Verlag.(The first modern account of Abraham's life and works.)

Fenton, P.B. (1981) 'Some Judaeo-Arabic fragments by Rabbi Abraham he-Hasid, the Jewish Sufi', *Journal of Semitic Studies* 26: 47-72.(Discusses the writings of Abraham's close companion.)

Fenton, P.B. (1987) *Deux Traités de mystique juive (Two Essays on Jewish Mysticism)*, Lagrasse: Verdier.(Survey of Sufi influences on Judaism and translation of two pietist texts.)

Goitein, S.D. (1967) 'Abraham Maimonides and his Pietist Circle', in A. Altmann (ed.) *Jewish Medieval and Renaissance Studies*, Cambridge, MA: Harvard University Press, 145-64.(Account of Abraham's activities based on Genizah documents.)

Goitein, S.D. (1988) *A Mediterranean Society*, Berkeley, CA: University of California Press, vol. 5, 476-96. (Portrait of Abraham Maimonides.)

Maimonides, 'Obadayah (*c*.1250) *al-Maqala al-Hawdiyya (The Treatise of the Pool)*, trans. and ed. P. Fenton, *The Treatise of the Pool, al-Maqala al-Hawdiyya by 'Obadyah Maimonides*, London: Octagon Press, 1981.(A pietist, Judaeo-Arabic vade-mecum for the mystical way, based on rabbinical teachings and Sufi asceticism, written by Abraham's son. As well as giving a translation and notes on the text, Fenton's edition includes the original Arabic text, edited from a manuscript in the Bodleian Library, Oxford and Genizah fragments.)

Wieder, N. (1948) *Islamic Influences on the Jewish Worship*, London: Phaidon.(Study of Abraham's ritual innovations.)

Maimonides, Moses (1138-1204)

*Called the Rambam in the Hebrew sources, an acronym on his name, and known in Islamic texts as Musa ibn Maimun, Rabbi Moses ben Maimon is best known in the West as Moses Maimonides and generally recognized as the greatest of the medieval Jewish philosophers. Maimonides lived his mature life in Egypt and earned his living as a physician. He was the author of ten medical works but gained fame in his own lifetime from his work on Jewish law (*halakhah*), chiefly the Kitab al-Fara'id (Sefer ha-Mitzvot, that is, the Book of the Commandments), cataloguing the traditional 613 commandments of the Pentateuch; Kitab al-Siraj (Sefer ha-Maor, Perush ha-Mishnah, Commentary on the Mishnah); and, above all, the Mishneh Torah (The Law in Review), a comprehensive and still authoritative code of rabbinic law. The clarity and definitiveness of the Mishneh Torah led to its criticism and (after Maimonides' death) even condemnation by some rabbis, who prized the ongoing dialectic of Talmudic disputation and felt suspicious of Maimonides' rationalism.*

Maimonides' philosophic masterpiece, the Dalalat al-Ha'irin or Guide to the Perplexed, was written in Arabic, with a view to helping the more intellectually inquisitive readers of the Torah, who were troubled by the apparent disparity between biblical and scientific/philosophical ideas. The work frames a powerful but not supercilious rationalism that locates and accommodates many biblical postulates and profits from the instruction of the rabbinic (Talmudic) sources and from critical appropriation of the achievements of Muslim philosophers and theologians and their Greek predecessors. It defends the doctrine of the world's creation against the eternalism of Neoplatonic Aristotelians but rejects the notion that creation (or eternity) is subject to proof. Rather, Maimonides argues, creation is preferable to its alternative, and more plausible, because it preserves the idea of divine volition as an explanation for the emergence of complexity from divine simplicity, and because it marks the difference God's act made to the existence and nature of the world.

God is pure perfection and absolute simplicity. The Torah's anthropomorphisms themselves lead us to that realization, if we follow the dialectic by which prophetic language directs us to ever higher conceptions of divine transcendence. Biblical poetry and the concrete demands of the Law are accommodations to our creaturely limitations. Such accommodations are made possible by the material side of the prophet's nature, as manifested in language and imagination, which are, no less than intellect, expressions of God. For matter in general is an expression of God, apprehensible to us through what seems wilful or arbitrary in nature. It is not a positive principle or hypostasis, but it is a necessary concomitant of the act of creation itself. For without it nothing other than God would exist. Our task as humans is to discipline our material natures - not to battle or seek to destroy them but to put them to work in behalf of our self-perfection, through which our inner, intellectual affinity to God will be realized.

Maimonides' synthetic approach, accommodating to one another the insights of reason and the teachings of Scripture and tradition, was highly valued by Aquinas, who frequently cites him, and by other European philosophers such as Jean Bodin. Leibniz warmly appreciated Maimonides' thought, as his reading notes reveal. Among subsequent Jewish thinkers, Maimonides' work became the paradigm of Jewish rationalism for his admirers and detractors alike. His philosophy was at the core of the philosophic tradition that Spinoza addressed. Even today practitioners of Jewish philosophy stake out their positions in reference to Maimonides and formulate their own views as appropriations, variants or interpretations of the elements of his thought.

1 Life and works

Called the Rambam in the Hebrew sources, an acronym on his name, and known in Islamic texts as Musa ibn Maimun, Rabbi Moses ben Maimon is best known in the West as Moses Maimonides and generally recognized as the greatest of the medieval Jewish philosophers. A physician and jurist as well as a philosopher, Maimonides was born in Cordoba, but his family fled the forced conversions of the Almohad invasion in 1148. From about 1160 they lived quietly in Fez, the heart of the Almohad movement, but came to public attention with the youthful Maimonides' publication of a defence of a rabbinic ruling his father had made promulgating full acceptance by the Jewish community of those who had undergone a nominal conversion. Journeying to Palestine in 1165, the family were nearly lost in a storm at sea but made their way to Acre. Maimonides prayed at the ruined Temple site and fulfilled a shipboard vow of pilgrimage to Hebron, burial site of the biblical patriarchs. Abraham bore a special

significance for Maimonides. In keeping with the biblical and midrashic tradition, he viewed the patriarch as a natural theologian whose religion was formed not in the tradition of his fathers but by reason confronting nature and experience.

Finding permanent residence impossible in the undeveloped land, then under the Crusader kingdom of Jerusalem, the family settled in Cairo, where Maimonides was prostrated with grief by the death of his younger brother David, a jewel merchant, in a shipwreck on the Indian Ocean. Recovering and taking over the support of the family, Maimonides turned to medicine. Like the ancient Talmudic sages, he disdained making his livelihood from his Judaic learning and judicial authority. He developed an active medical practice and became court physician to the *wazir* of Saladin, who overthrew the Fatimid dynasty in 1171. Late in life Maimonides wrote ten medical treatises, but it was his work in rabbinic law that gave him fame and made him a leader of the Egyptian Jewish community as early as 1167.

Three works secure his place among the greatest rabbinic jurists: his *Kitab al-Fara'id*, known in Hebrew as the *Sefer ha-Mitzvot*, cataloguing and classifying the traditional 613 biblical laws, lays the foundations; his Arabic *Kitab al-Siraj* or *Book of the Lamp*, known in Hebrew as the *Sefer ha-Maor, Perush ha-Mishnah* (1168), expounds the rational purposes of the ancient rabbinic code; and his *Mishneh Torah* (*c.*1185), written in rabbinic Hebrew and popularly known as the *Yad Hazakah (The Strong Hand)*, or simply the *Yad*, an allusion to its fourteen volumes, since the numerical value of the letters of the Hebrew word *yad*, or hand, is fourteen. This work organizes the vast complexity of Talmudic and post-Talmudic Jewish law in a still authoritative codification, allowing clear access to definitive legal rulings on matters of *halakhah* (see Halakhah).

Despite the prestige of these works - in part because of it - Maimonides' jurisprudence was not universally hailed. The *Mishneh Torah* in particular was criticized: its title, which might be rendered 'The Law in Review', and which echoes a title traditionally assigned to the biblical book of Deuteronomy, seemed almost to promise a second revelation. The economy of the work, which placed a premium on stating clear-cut practical decisions, demanded the omission of rabbinic citations and dispensed with the voluminous dialectic which for many was the heartsblood of the Law as an intellectual discipline, even a way of life.

Maimonides was in fact sometimes impatient with the digressive and midrashic (homiletical) methods of rabbinic discourse. His insistence on juridical clarity, determinateness and Socratic conceptual organization (as sharply contrasted with the associative exposition of the Talmud) were clearly much on his mind when he turned from writing commentaries on parts of the Talmud to his more distinctive commentary on the Mishnah. The notion that study of the law could somehow supplant the robust life that the law meant to regulate seemed especially unwholesome to him - as it had to his Andalusian predecessor Judah Halevi.

A brilliant and faithful jurist, Maimonides drew the axioms of his theology from scriptural rather than rabbinic sources; he worked eclectically and creatively with the rabbinic responses to biblical ideas to forge a coherent and appropriable framework of thought - just as his rabbinic predecessors had done. But traditionalists, legal positivists and fideists could not fail to note that where Maimonides could not find room to cite his legal authorities (and some of them are still being found, giving the lie to long-standing claims that he based some of his legal doctrines sheerly on his own opinions) he did make room, in the first volume of his code, to frame a thematic ethics and to elaborate a natural theology and cosmology - and, in the last volume, to develop a biblically grounded theory of sovereignty and public law. Similarly, he does not regard it as digressive in discussing the laws of penitence to include a defence of human freedom of will. Reacting against what they viewed as an excessive rationalism, a number of rabbis condemned the *Mishneh Torah*, although others defended it vigorously. Denounced to Church authorities by Jewish adversaries, the first and most thematic volume of Maimonides' code, the *Sefer ha-Mada (Book of Knowledge)*, was burnt at Montpellier in 1232, along with the seemingly heretical *Guide to the Perplexed*. Despite the reservations of some Jewish traditionalists to this day about the rationalism of his philosophical and jurisprudential work, Maimonides remains a towering figure in *halakhah*. His work is the cornerstone of all subsequent Jewish philosophy, and it has had a significant impact on philosophy at large. His synthetic approach, accommodating to one another the insights of reason and the teachings of Scripture and tradition, was highly valued by Aquinas, who frequently cites him, and by other European philosophers such as Jean Bodin. Leibniz warmly appreciated Maimonides' thought, as his reading notes reveal. Maimonides' work became the paradigm of Jewish rationalism for his admirers and detractors alike. It was a major philosophic

backstop to the thinking of Spinoza, providing a clear paradigm of the committed Jewish philosophy that Spinoza would address, whether by way of refutation, accommodation or revision. Even today practitioners of Jewish philosophy stake out their positions in reference to Maimonides and formulate their own views as appropriations, variants or interpretations of the elements of his thought.

2 *Guide to the Perplexed*: its literary form and intent

Maimonides' philosophic masterpiece, *Dalalat al-Ha'irin (Guide to the Perplexed)* (*c*.1190), was written in Arabic but widely circulated in Hebrew translations as the *Moreh Nevukhim*. Aiming to mediate between the scriptural and philosophic idioms, it opens by deriving a sophisticated negative theology from a subtle deconstruction of the anthropomorphisms that had long troubled thoughtful readers of the Torah. This discussion, filling the first seventy chapters of the *Guide*, is widely but erroneously read as a refutation of anthropomorphism. I say erroneously, because Maimonides assumes that the falsity of anthropomorphism is well established by argument and well known to his reader. But Maimonides does not do much to prevent such misprision. He does not state the objective of his discussions in this extended section of the first part of the *Guide* but leaves casual readers to fend for themselves. One natural outcome is the commonplace assumption that what Maimonides takes for granted is what he is seeking to prove. That, in Maimonides' view, is better than allowing his philosophic work to do more harm than good.

The difficulty that Maimonides confronts is this: the *Guide* is written to aid a reader who is confused by the seeming disparity between biblical and scientific/philosophical ideas. To mention only the most obvious issues: Scripture tells us that the world is created, but Aristotelian philosophers present numerous proofs to show that the world has always existed with the same, invariant nature that we observe today. Scripture speaks of God as the ruler of nature and the lawgiver of humanity in general and of Israel in particular. But the philosophical tradition argues that the divine transcends all knowledge of particulars, let alone concern for them. Indeed, it is readily enough argued that chance seems to rule, in this life at least, and that sufferings outweigh happiness. So where are the divine justice and judgment that loom so large in the vision of the Torah and the prophets? How is prophecy even possible, if God is so far above us as philosophy seems to teach? And how can we relate the philosophical teaching of divine transcendence to the vivid concreteness of prophetic imagination?

Maimonides unifies these problems about transcendence and immanence, providence and revelation, creation and eternity, finitude and privation under the rubrics of two rabbinic problematics: the account of creation and the account of the chariot. The first rubric alludes to the opening of Genesis, the second to that of Ezekiel, in whose epiphany, despite all his reverent periphrasis, God seems to take human form. Voicing a sense of the paradox or mystery surrounding these two passages, the rabbis of the Talmud (Haggigah 11b, 13a) forbade public instruction in the issues they involve. Only in private teaching were these to be discussed, and then only the subject headings were to be broached.

Interpreting this ruling, Maimonides understands 'the account of Genesis' as the rabbis' name for cosmology, the principles of nature or physics, understood in relation to God; 'the account of the Chariot' is, by metonymy, the entire realm of theology, that is, metaphysics understood in relation to God. The core issue in both cases: how can the Infinite be manifest in finite terms?

The cosmological side of this question is evident in the Neoplatonists' problematic of the Many and the One - why is there anything other than God? It crops up polemically in the sensitive and vexed area of creation, where philosophers of Aristotelian stamp find it incongruous to ascribe an origin to the world, in part because their naturalism balks at subjecting the cosmos to external interference but in part because they wish to avoid compromising divine transcendence in temporality.

Viewed in this light, creation itself appears to be an instance of theophany. The problem of evil and that of prophetic revelation reflect the same issue: why does divine creation content itself with a world as flawed as this one? Why would revelation single out one people or individual to hear God's word, leaving others dependent or unenlightened altogether; and why should God's expression, as epiphany or as law, take the precise form ascribed to it by individual prophets, with all their human weaknesses and differences of outlook and circumstance?

The ordinary mind, Maimonides finds, is generally content to accept what is familiar, not noticing the difficulties a tradition may contain. But such faith, if faith it may be called, is of limited value and entirely dependent on the

veracity and coherence of those who establish a tradition. In the strictest sense, such unreflective acceptance is not faith at all. For belief is commitment to what one understands; it cannot be confined to mere conformity.

To the more questioning, especially those who have read scientific and philosophical literature, theological problems are apparent. Responsible intellectual leaders must work creatively to discover for themselves and show others how these problems can be addressed. Progress here, as in other areas of intellectual inquiry, is possible. Indeed, our progress towards God can be asymptotic, Maimonides argues, invoking the geometrical image of two lines that grow ever closer together, even though they will never meet. But progress, to be cumulative, requires writing and teaching, and that presents a difficulty; it is not responsible to introduce others to problems whose solutions they might not grasp. Maimonides takes this issue very seriously, holding a teacher responsible even for the misinterpretations of followers and disciples.

Taking a hint from the language of the ancient Talmudic restriction, Maimonides adopts an ‘esoteric’ mode of writing in the *Guide*. Not that he writes obscurely or treats his subject as occult or hermetic. He simply does not state the problems to be addressed. A reader who has grappled with the same questions will readily enough discover the relevance of the discussions in the *Guide*. But that relevance is not spelled out - let alone promoted in the mercantile manner of twentieth-century pedagogy. The order of exposition, moreover, as with the texts that Maimonides studied most closely - the Hebrew Bible, the Talmud, the Aristotelian corpus - demands that one read well into the text before being able to make much of it, since every part is relevant to the rest, and no passage is wholly self-explanatory or transparent.

Maimonides’ work, like Aristotle’s, and the Jewish canonical sources, is clearly not part of the Cartesian modernist project, with its Euclidean pyramidal structure, designed to move upwards from foundational axioms, postulates and definitions to theorems, scholia and lemmas. Rather, the work is bound together like a Gordian knot - not impenetrable, but resistant to simple unravelling by the teeth and fingers of the unprepared.

Uncritical thinking is probably just as common today as it was in Maimonides’ time, but exposure to doubt and (somewhat conventional) lines of objection to biblical categories and assumptions is far more widespread. Today’s unexamined faith is often a form of scepticism. So the *Guide to the Perplexed*, although written to address specific issues at a very particular historical juncture, has the character of a classic - not only because it is synthetic, original and insightful, but also because readers today find its approach and many of its arguments as relevant as when first set down.

Following the rabbinic injunction that the problems of creation and theophany are to be taught not publicly but in a direct and idividual communication, Maimonides adapts to his purposes the Arabic literary genre of the epistolary essay (*risala*). The *Guide* is written in the form of a letter addressed to a particular disciple, whose needs and questions can be addressed in an informal, dialectical style, rather than the formal, systematic style of a treatise like the *Mishneh Torah*. Having sketched the background, skills and uncertainties of this first intended reader, Maimonides has created a persona, not in the authorial voice but in the figure of the reader, clearly alerting others to the scope and relevance to their own concerns of the arguments they will encounter.

In keeping with his expository strategy, Maimonides never calls the *Guide* a book. He moves with seemingly casual, conversational rhythms from exegetical remarks to philosophical observations, arguments, typologies and critiques, relying on sensitive readers to follow his hints, supply the unstated problematic, and piece together the larger argument. He well understands that the broader impact of the opening chapters will be fully felt only by readers who take the time and trouble to read them once again after having progressed through the work as a whole.

It might seem that Maimonides has simply employed a legal fiction, and a thin one at that, to skirt the rabbinic restriction on the sort of teaching he wants to pursue. But his strategy is far subtler than that. He has successfully kept the *Guide*’s problematic within the circle of its intended audience. Casual readers typically emerge from an encounter with the *Guide* without any knowledge of its subject matter; most casual readers do not finish it, let alone begin again, so as to be able to piece together the scattered elements of its larger argument.

3 *Guide to the Perplexed*: creation and eternity

If we begin not where Maimonides did, with the deconstruction of biblical anthropomorphism, but where Genesis

does, with the beginning, we find Maimonides in a spirited defence of the idea of creation against the eternalism of al-Farabi, Ibn Sina, Aristotle, and the whole Neoplatonic Aristotelian tradition (see Eternity of the world, medieval views of). The problem is not, Maimonides argues, that philosophical proof contradicts scriptural dogma. On the contrary, if we had proof that creation was impossible, as Aristotelian philosophers long argued, we would simply allegorize the biblical account of creation, just as we allegorize biblical anthropomorphism. But Aristotle knew that he had no such proof. That is why he speaks so emphatically when he tries to assure us of the world's eternity. After all, it was he who taught us the difference between proof and mere persuasion.

What the philosophers of the Aristotelian tradition have done is simply read into the metaphysics of being the familiar character of nature: no event without a prior event, no process without potentiality, no potentiality without matter as its substrate. Of course assumptions like these will generate an eternal cosmos. But they amount to nothing more than an aprioristic attempt to generalize the conditions of the settled order of nature and impose the assumptions we derive from study of that order on the wholly unknown conditions of the world's disputed origin - as if a brilliant young man who knew nothing of sex or procreation were to infer the impossibility of embryogenesis and foetal development from his knowledge of the biology of mature adults: how could a human live in an enclosed sac, without breathing, moving about, ingesting solid food, or excreting bodily wastes?

Rejecting efforts to demonstrate the world's eternity, Maimonides even-handedly rejects creationist attempts of the same sort. If it is really true, he argues, that the world must reflect God's timelessness, as eternalists suppose, then change itself becomes impossible, and the argument has proved too much. If the claim is that God, to be a creator, must be a constant creator, then creation becomes not impossible but necessary. Those who try to demonstrate creation, not surprisingly, produce a similar result: the world is new at every instant. The radical occasionalism of Arabic dialectical theology that results from such thinking, Maimonides argues, renders nature unintelligible and divine governance irrational (see Islamic theology §§1-2; Occasionalism). Creation, in that case, to be the outcome of a clear proof, must cease to be God's choice and become a sheer necessity and unending chore.

But, although efforts to demonstrate the world's creation or its eternity fail, there are grounds for choosing between the rival cosmological visions. Creation is more plausible conceptually and preferable theologically: more plausible, because a free creative act allows differentiation of the world's multiplicity from divine simplicity, as the seemingly mechanical necessitation of emanation, strictly construed, cannot do; preferable, because Avicennan claims that God is author of the world and determiner of its contingency are undercut by the assertion that at no time was nature other than it is now (see Ibn Sina). Without the notion of creation at a particular time, one cannot say that atheism ensues, as al-Ghazali had supposed. The view of the Neoplatonic Aristotelians is neither incoherent nor insincere. But it does unduly attenuate the world's dependence on God's act, and it does leave unexplained the emergence of complexity from simplicity - problems which the human idea of a will that is not simply reducible to intellect or understanding far more readily addresses.

4 *Guide to the Perplexed*: God's governance of nature

In discussing God's governance of nature, Maimonides adopts and adapts the emanative ideas of the Neoplatonic philosophers (see Neoplatonism). But he modifies their notion that emanation is a necessary progression, poetically compared to the radiation of light from the sun (which, after all, cannot help but shine) but more perspicuously described in terms of the necessitation of theorems by their axioms. In place of such deductivist, intellectualist models, Maimonides uses a more voluntaristic model, allowing for creativity and novelty in nature's unfolding, while rejecting the more radical voluntarism of the occasionalists, who make God the immediate author of all events, without the intervention of natural causes or human volitions (see Voluntarism). If persons and things do not act and react, Maimonides argues, was it not otiose of God to create foodstuffs that play no role in our sustenance, or to ordain commandments that we have no power to obey or disobey?

Taking the Neoplatonists to task for not exploiting adequately the resources of their own ontology for addressing the problem of evil, Maimonides couches his response to the Epicurean dilemma in a gloss on the Book of Job. The Satan (or 'adversary') responsible for Job's sufferings, he notes, is not called one of the 'sons of God' but is said only to have come along with them, 'in their midst'. One might suppose, with the Talmudic sage Resh Lakish, that Job's adversary was sin - equating sin and death with the adversary (see Sin). Maimonides likes the allegory but rejects its application here. For it is a premise of the Book of Job (which Maimonides, like Resh Lakish, quite clearly takes to be a fiction, although he regards it as one with considerable verisimilitude) that Job was without

sin (see Evil, problem of).

The Satan, then, would be Job's vulnerability, that is, his materiality, the physical nature that makes us all subject to sufferings and loss. For, as Galen explained, one should not expect bodies of the sort of matter that makes our own to last forever. Here we see the weight of Scripture's saying that the Satan came 'in the midst of' the sons of God: matter, that is, otherness, alienation from the unity of the Divine, is not a positive principle, a hypostasis, angel or Form, but a concomitant to the emergence of finite being, a requisite of creation and createdness. The existence we enjoy has vulnerability as its price; divine generosity could not be expressed and vouchsafed to lesser beings at all unless those beings were indeed made other, given a reality that is at once their own and (by the same token) inevitably deficient. Matter is that independence and deficiency. It is at once the good woman of Proverbs 31, constantly active on our behalf, and the married harlot of Proverbs 7, never contented but constantly changing forms/partners.

Our human task is to overcome the limitations of matter, but we can do so through its strengths. Thus mortification of the flesh is not a worthy but an unwholesome ideal, based on the illusion that if a little medicine is good for us a lot will be that much better. Medical arts, politics, and otherworldly endeavours are precious but instrumental efforts, since they sustain the human individual and community, allowing us the necessary time and opportunity to seek and reach the human goal.

What is that goal? Maimonides categorically agrees with Plato (*Theaetetus* 176b) and Leviticus (19:2): it is to become as like to God as humanly possible, to emulate and pursue God's holiness, to realize and fulfil what is divine and holy in ourselves. Job's reward is not in the restoration of his family and possessions but in the epiphany he is granted, a communion with God that is not the recompense of his sufferings but an emblem of the sort of human attainment that makes all sufferings if not bearable then at least outweighed.

It is not true, as Muhammad ibn Zakariyya' al-Razi and even Saadiah Gaon supposed, that sufferings overbalance good fortune in this life. But the deeper flaw in the emotively appealing notion that life is a vale of tears is that it assays human life in terms of pain and suffering, whereas the ultimate coin is not hedonic at all but intellectual. Reason is the surest guide to the good life and test of sound and unsound norms and precepts; but it is also our closest bond with God, our purchase on immortality, and the focus of divine providence within us. For providence, in the human case, is not confined to the level of the species and its generalized nature. Indeed, it was Aristotle, once again, who taught us that only individuals are real, not species as such.

Contrary to the Aristotelian teaching (expounded by Peripatetics such as Alexander of Aphrodisias in the face of the Stoic doctrine of providence - see Alexander of Aphrodisias; Stoicism §20), providence does extend to the individual, through our capacity to perfect the mind, realize our inner affinity to God, and grasp the immortality that lies so close at hand. This is symbolized in the biblical account of Adam, whose ready access to the fruit that would give him eternal life represents not a long vanished moment of apotheosis in the mythic past (and still less a damning fall) but the universal human condition, of an immortal intelligence strapped to a body that is at once its trap and its springboard to the divine.

5 *Guide to the Perplexed*: anthropomorphism and the human connection with God

Quietly celebrating the human affinity with God, Maimonides (as noted by the medieval commentator Narboni) begins his survey of biblical anthropomorphisms with 'image and likeness' (see Narboni's *Be'ur le-Sefer Moreh Nevukhim (Commentary on the Guide)*: I 1). Maimonides explains that some Hebrew words (as, for example, when we read that Joseph was fair of form) denote only physical appearance; but 'likeness' can refer to spiritual affinities. Thus in Psalms (102:6), the psalmist likens himself to a pelican in the desert: not that he looks like a pelican, but that he is as desolate as a pelican would be in such a place. As his survey proceeds, Maimonides unfolds a hierarchy of meanings, with the most physical at the bottom, social and other senses rising higher, and the spiritual/intellectual surmounting the rest: wherever the Torah applies to God a term that bears physical (or other privative) connotations, that term can always be taken in a higher, spiritual/intellectual sense, and that is the sense we must pursue.

Maimonides does not here refute anthropomorphism but assumes it to have been refuted by well-known arguments based on the incompatibility of divinity with privation. Correspondingly, he pursues the idea of perfection, the opposite of privation, as the notion that orients this hierarchy, as it orients the ladder of love in Plato's *Symposium*

- or Jacob’s ladder in the Book of Genesis. What Maimonides finds here is an ontic hierarchy that fuses higher degrees of reality with higher degrees of intellectuality. He shows that none of the Torah’s anthropomorphisms need be taken literally (‘jealousy’ refers to exclusivity, ‘coming and going’ to manifestation in a human awareness), but in the process he vindicates the legitimacy of conceptualizing the Divine in terms of the human mind (and will).

We begin to see how prophecy is possible, as a human response to the exigency of relating to others the sheer power of Perfection. The Torah, using human language, speaks of action in terms of motion, of existence in terms of matter. Only the Tetragrammaton signifies God directly, signalling nothing more (but nothing less) than the absoluteness of God’s being - ‘I am that I am’. All other epithets only point towards Perfection, assigning to God some notion humanly regarded as a perfection, while other passages systematically exclude the privations concomitant with the perfections known to us. Thus the deconstruction of biblical anthropomorphism is the work of Scripture itself; the Torah teaches of divine transcendence through the medium of ordinary human language. It uses terms intelligible to us through their generality to convey the idea of divine uniqueness. This was what the rabbis meant in saying, ‘Great is the boldness of the prophets’ to speak of the Creator in the language of his creations.

Prophets are historically situated human beings, but they are not and cannot be mere ignoramuses whose minds are magically filled with divine words; nor can they be immoralists, simply plucked up to be inspired. They are and must be wise individuals, filled with the kind of philosophic insight that can grow only within an upright character. But they are gifted also through the material side of their nature, with the language and imagination to body forth these ideas in concrete visions, laws, rituals, poetic and rhetorical symbols, to speak of them persuasively and act on them with conviction.

Prophets themselves learn nothing new from revelation, although it may bring to the surface ideas that were only latent in their minds. The ignorant remain ignorant; but the gift of imagination in the wise, if they are disciplined by the moral virtues, especially courage and contentment, gives wing to ideas, rendering them accessible to the masses and setting them into practice. The message of all true prophets is thus coherent and universal, an expression of the truth itself.

In principle, any philosopher of character and imagination might be a prophet; but in practice the legislative, ethical and mythopoetic imagination that serves philosophy finds fullest articulation in Judaism. Its highest phase, where imagination yields to pure intellectual communion, was unique to Moses, elaborated in Judaism and its daughter religions. This judgment, that the burden of prophetic teaching finds its epitome in the Torah, is no mere subjective preference: it is grounded in intensive study of the means by which the biblical commandments, read thematically, serve to inform human character and understanding. The most general summary of the outcome of that study is in the Maimonidean thesis that the Mosaic law is no mere personal code, private vision or closely held teaching like Abraham’s; nor is it a mere derivative faith. Rather, it is a systematic way of life based on sustained, intellectual contact with the Divine. When courage and contentment - read here ‘self-confidence and independence’ - return, with the restoration of Israel to her land, prophecy will resume among this people, to the great benefit of the nations of the world, who will observe the peace and prosperity of Israel living under God’s law and will flock to follow not her religion but her example. It is this that is to be understood by the lion’s lying down with the lamb: the nations will learn the ways of peace, and Israel will be the little child that leads them (see Prophecy).

God governs through nature. The forms of things are the objective manifestations of his wisdom, intelligible through their fit with the subjective rationality of our own divinely imparted intelligence. Where a human discoverer must take apart the waterclock to understand its workings, the inventor knows it first: his knowledge does not follow but anticipates the design. The same is true with God, whose knowledge is prior to the realities we know, since it is their cause.

But the matter of things, readily enough associated with chance and the irrational, is another aspect of God’s manifestation in nature, the aspect that we would link, through its seeming arbitrariness, with the human faculty of willing. In God’s ultimate and absolute unity, will and wisdom are united; their multiplicity is subjective. So in biblical parlance all the operations of nature but also the workings of sheer chance are ascribed to God. But we humans must always think in human categories, so it is not possible for us, while we live, to grasp the absoluteness

of God's unity and hold in one mental breath what we can conceptualize only as the wisdom and the will of God.

Human finitude, accordingly, denied Moses the vision he requested, of God's face; but a perfect mind allowed him to see God's 'back' - what follows from God, the effects of God in nature, where divine absoluteness becomes specific and can be apprehended as if in attributes like those of a person. God, of course, in his infinitude, transcends personality, as even little children should be taught. Attributes belong only to composite beings. But Moses' request, after all, was a practical one: he wanted to know God so that he would know how to govern; and for practical purposes, we have knowledge of what God is like by examining the attributes that should be perfected in ourselves, the very attributes named in Moses' epiphany: mercy, compassion, justice - and, of course, the human power that all of these presuppose, that of reason, our direct link with the Divine.

See also: Aristotelianism, medieval; Aristotelianism in Islamic philosophy; Bible, Hebrew; Creation and conservation, religious doctrine of; God, concepts of; Halakhah; Maimonides, A.; Midrash; Nahmanides, M.; Religion, philosophy of; Theology, Rabbinic

L.E. GOODMAN

List of works

Maimonides, Moses (1168) *Kitab al-Siraj (Commentary on the Mishnah)*, complete translation available in Hebrew only, *Sefer ha-Maor, hu Perush ha-Mishnah*, ed. M.D. Rabinowitz, Tel Aviv: Rishonim, 1948; English translation of the part of the work entitled *Shemonah Perakim (Eight Chapters)*, trans. J. Gorfinkle, *The Eight Chapters of Maimonides on Ethics*, New York, 1912; repr. New York: AMS, 1966.(Expounds the rational purposes of the ancient rabbinic code. *Shemonah Perakim* is a thematic introduction to the mishnaic collection of ethical aphorisms known as *Pirkei Avot (The Sayings of the Fathers)*. It develops a capsule virtue ethics and sketches a theory of action aiming to rebut determinism.)

Maimonides, Moses (*c*.1185) *Mishneh Torah (The Law in Review)*, Books I and II trans. M. Hyamson, *The Book of Knowledge and the Book of Adoration*, Jerusalem: Feldheim, 1974; Books III-XIV trans. as *The Code of Maimonides*, Yale Judaica Series, New Haven, CT: Yale University Press, 1949-72.(Maimonides's authoritative fourteen-volume code of rabbinic law, written in mishnaic Hebrew with a view to providing lucid and comprehensive guidance to the requirements of *halakhah*. The work opens with detailed discussions (in the first volume) of the intellectual and moral foundations of rabbinic law. It closes with an exposition of public law and a description of the Messianic age of Israel's political restoration. The intervening volumes discuss each requirement of the civil, criminal and ritual law of Israel.)

Maimonides, Moses (*c*.1190) *Dalalat al-Ha'irin (Moreh Nevukhim, Guide to the Perplexed)*, ed. S. Munk, *Le Guide des Égarés*, Arabic text, critically edited, with annotated French translation, Paris, 1856-66, 3 vols; repr. Osnabrück: Zeller, 1964; trans. S. Pines, with an introductory essay by L. Strauss, *The Guide of the Perplexed*, Chicago, IL: University of Chicago Press, 1969; *Rambam*, extensive selections from the *Guide* with commentary by L.E. Goodman, New York: Viking, 1976; reissued, Los Angeles, CA: Gee Tee Bee, 1984. (Maimonides' philosophic masterwork, written in Arabic in the form of a letter to a disciple who has confronted the apparent disparities between philosophic/scientific and biblical ideas.)

Maimonides, Moses (before 1204) *Millot ha-Higgayon (On Logic)*, ed. Y. Satnov, with commentary by M. Mendelssohn, Berlin: B. Cohen, 1927.(Maimonides' work on logic.)

Maimonides, Moses (before 1204) *Kitab al-Fara'id (Sefer ha-Mitzvot, Book of the Commandments)*, trans. with notes by C.B. Chavel, *The Commandments (Sefer ha-Mitzvot)*, London and New York: Soncino, 1967. (Maimonides' listing of the 613 commandments traditionally held to be conveyed in the Pentateuch. In keeping with his thematic approach, Maimonides lists first among the 248 positive commandments 'I am the Lord thy God' and first among the negative commandments 'Thout shalt have no other gods before me'. These are the two commandments that were rabbinically said to have been learned from God himself, as Maimonides explains, citing the Babylonian Talmud, Makkot 23 b.)

Maimonides, Moses (before 1204) *Iggerot ha-Rambam (Maimonides' Epistles)*, trans. with notes by A. Halkin and discussions by D. Hartman, *Epistles of Maimonides: Crisis and Leadership*, Philadelphia, PA: Jewish Publication Society, 1993.(Comprises 'The Epistle on Martyrdom', 'The Epistle to Yemen' and 'Essay on the Resurrection'.)

References and further reading

Baron, S.W. (ed.) (1941) *Essays on Maimonides: an Octocentennial Volume*, New York: Columbia University Press.(Papers of the historic Maimonides conference held in the 1930s in Cordoba for the 800th anniversary of

the philosopher’s birth, then dated in 1135.)

Buijs, J. (1988) *Maimonides: A Collection of Critical Essays*, Notre Dame, IN: Notre Dame University Press. (Includes studies by Harry Wolfson, Alexander Altmann, Seymour Feldman, Larry Berman, and others, including some reprinted classics.)

Hartman, D. (1976) *Maimonides, Torah and Philosophic Quest*, Philadelphia, PA: Jewish Publication Society,1976.(Seeks to overcome the polarity some find between Maimonides’ rabbinic and his philosophical work.)

Katz, S. (ed.) (1980) *Maimonides: Selected Essays*, New York: Arno Press.(Reprints a number of classic essays, some of them now fairly inaccessible.)

Kellner, M. (1990) *Maimonides on Human Perfection*, Atlanta, GA: Scholars Press. (A careful examination of Maimonides’ conception of the ultimate goal of human existence.)

Kraemer, J. (ed.) (1991) *Perspectives on Maimonides: Philosophical and Historical Studies*, Oxford: Oxford University Press.(Includes essays by Alfred Ivry, Warren Harvey, Arthur Hyman, Tzvi Langermann and others on Maimonides’ logic, theology and jurisprudence, along with historical studies of his environment.)

Leibowitz, Y. (1987) *The Faith of Maimonides*, trans. J. Glucker, New York: Adama Books.(Monograph by a well-known exponent of Jewish legal positivism and fideism.)

Narboni, Moses (1355-62) *Be’ur le-Sefer Moreh Nevukhim (Commentary on Maimonides’ Guide to the Perplexed)*, French trans. M.R. Hayoun, *Moshe Narboni*, Tübingen: J.C.B. Mohr, 1986.(The French translation is marred by an excessive attempt at literalism, but this volume provides valuable information about Narboni and a valuable text.)

Pelaez, J. (ed.) (1991) *Sobre la Vida y Obra de Maimonides*, Cordoba: Ediciones El Almendro.(Proceedings of a major international conference celebrating the 850th anniversary of Maimonides’ birth.)

Pines, S. and Yovel, Y. (eds) (1986) *Maimonides and Philosophy*, Dordrecht: Kluwer.(Proceedings of an anniversary conference with papers by a number of distinguished scholars.)

Robinson, I., Kaplan, L. and Bauer, J. (1990) *The Thought of Maimonides: Philosophical and Legal Studies*, Queenston, Ont.: Edwin Mellen.(Another anniversary symposium gathering a representative sampling of recent work on Maimonides.)

Sarachek, J. (1935) *Faith and Reason: the Conflict over the Rationalism of Maimonides*, repr. New York: Hermon Press, 1970.(Wide-ranging study of the figures and issues of the ‘Maimonidean controversy’.)

Silver, D.J. (1965) *Maimonidean Criticism and the Maimonidean Controversy, 1180-1240*, Leiden: Brill, 1965. (Reviews the same ground.)

Strauss, L. (1952) *Persecution and the Art of Writing*, Glencoe, IL: Free Press.(This brief collection of essays, including chapters on Maimonides, al-Farabi, and Spinoza, was the manifesto of Strauss’ approach to the reading of esoteric texts. It spawned a cottage industry of exegesis, the lesser examples of which link laboured and mannered readings of texts with inferences or innuendoes about the cynical insincerity of their authors, among whom Maimonides remains a focal point of interest.)

Twersky, I. (1991) *Studies in Maimonides*, Cambridge, MA: Harvard University Center for Judaic Studies.(Papers by Shlomo Pines, Aviezer Ravitzky, Jacob Levinger, Moshe Idel, and Gerald Blidstein on the relationship of Maimonides’ thought to law and mysticism and on his literary impact and communal role.)

Maine de Biran, Pierre-François (1766-1824)

Maine de Biran claimed that the starting point for our understanding of human beings lay in introspective psychology: it was the awareness of willed effort. A proper understanding of the will should be the foundation of all work in psychology, including empirical psychology, as well as in the human sciences in general, which should work together towards a coordinated 'anthropology'. Contrary to the assumptions of associationist psychology, mental facts were essentially relational, and language was a constitutive feature of them, rather than being a secondary device intended to represent them. But conscious mental life arose from and was influenced by a subconscious underlayer which could be studied only by the joint use of physiological and introspective methods. Biran rejected the view that mental states can be reduced to, or are nothing other than, physical states. There was a partial 'symbolic' correspondence between them, which meant that physical accounts and mental accounts could not be translated into each other without loss. Later in his life, though still maintaining the belief that psychology was primary, he held that it was necessary to accommodate questions of metaphysics, morality and religion. He published very little during his lifetime, and many of the 'works' found in editions are (sometimes conjectural) editorial restitutions made from a jumbled mass of much corrected manuscripts. His failure to complete a single work on what the eighteenth century had called the 'Science of Man' (which he said was the greatest interest of his entire life) reflects the times: this eighteenth-century project was fragmenting into the multiplicity of human sciences which were emerging as the nineteenth century came. But his insistence on the primacy of the will remains a major challenge for the human sciences of today.

1 Life and works

Pierre-François Gont(h)ier de Biran, French philosopher and public figure, was born at Bergerac in 1766. His father was a doctor. In 1785, after his schooling in Périgueux, he became a royal guardsman and in 1789 was wounded at the defence of Versailles. Though the Guards were dissolved in 1791, Biran (who by now was called 'Maine', a name acquired from a family estate at his majority) still envisaged a career in the military engineers, but this came to nothing. At the end of 1792, he retired to the estate of Grateloup, near Bergerac. At this time, reading was a main occupation and he wrote mainly notes or meditations, rather than works. He did begin a translation of Beccaria with a commentary and somewhat later a memoir on signs. But both these works were left incomplete, setting a pattern repeated throughout his life of intellectual work only sometimes carried to the point of completion.

Gouhier (1948) said of him that 'he was a one-book man - but he never wrote the book'. This book would have been an attempt to fulfil the eighteenth-century project of a 'Science of Man', and he worked at it constantly, once calling it 'the greatest interest of my entire life'. In earlier stages, his attention was focused on psychological topics. His memoir on habit (*Mémoire sur l'influence de l'habitude sur la faculté de penser*, 1803) received a prize from the Institut de France in 1802, and another reflecting the conception of mental science held by the *idéologues* (*Mémoire sur la décomposition de la pensée*, 1804) was awarded the prize in 1805. Publication of this work was stopped at the page-proof stage for mysterious reasons. A memoir on the data immediately present to consciousness (*De l'aperception immédiate*, 1807), replying to a question set by the Academy of Berlin, was awarded a prize in 1807. There followed a series of memoirs on much more specific topics delivered to the Medical Society of Bergerac, which Biran founded, the first on unconscious perception, the second on phrenology, and the third on sleep, dreams and somnambulism. Biran then returned to larger-scale work, with the memoir on the relations between body and mind (*Mémoire sur les rapports du physique et du moral de l'homme*, 1811), submitted to the Royal Academy of Copenhagen, which awarded him the first prize in 1811. In the same year, he signed a contract with a Paris publisher for the publication of revised versions of the memoirs of Berlin and Copenhagen. But he quickly changed his mind, speaking now of a single 'large work'. We know that he regarded this in 1811-12 as 'almost finished'. It concerned the foundations of psychology and its relations to the natural sciences (*Essai sur les fondements de la psychologie, et sur ses rapports avec l'étude de la nature*). But Biran came to think that metaphysical questions of a kind which he formerly disqualified now needed to be treated. This already emerges in a work allied in subject matter to the *Essai* (*Les rapports des sciences naturelles avec la psychologie*), which may have arisen as a reworking of the second part of the *Essai* and may be dated to around 1814. Within a year or two, Biran began to take a much stronger interest than before in questions to do with the nature of morality, and this further extension of interest put completion at a greater distance. In 1819, he

congratulated himself in his diary on not having published ‘anything definitive’ yet, since the spiritual life now needed to be taken into account. In 1820, a specific project intervened. He revised the memoir of Copenhagen for the use of a doctor, Antoine-Athanase Royer Collard, who had consulted him in connection with a course on insanity (*aliénation mentale*) which Royer Collard was to give in the asylum of Charenton in the following year. At the same time, he signed another contract with the same publisher as before for the publication of this revised work, together with the memoir of Berlin, thereby reinstating the plan of ten years earlier. But the publication did not take place, and in 1822 Biran wrote in his diary of going back to his manuscripts and making a work from them which could shortly be published. Among his last attempts at this was a work on ‘anthropology’ in the broadest sense, with the title *Nouveaux essais d’anthropologie* (1823), now making explicit in the title itself the nature of Biran’s ambition; but he died in 1824 without fulfilling his lifelong goal.

2 Problems of the manuscript tradition

The many thousands of surviving folios, often much corrected, present a fascinating spectacle of a mind at work, and a headache for subsequent editors, one of whom wrote in exasperation a generation after Biran’s death of this ‘atrocious scribble’ (‘affreux barbouillage’). The first editor was Victor Cousin (delegated by Biran’s literary executor). He was careless and unscholarly, publishing only four volumes. The second initiative came from François-Marc Louis Naville, a Protestant pastor from Geneva who had met and admired Biran, and who was helped by his son Ernest. They collected about 12,000 pages of manuscript, and Ernest wrote ‘what a task to bring light into this chaos!’. They did a considerable amount of work, recruiting various helpers, and published a version of Biran’s *Essai sur les fondements de la psychologie*. The third editor was Pierre Tisserand, who produced between 1920 and 1949 an edition of Biran’s works in fourteen volumes. Though incomplete and not very scholarly, it made available for the first time a wide and representative collection of texts. Henri Gouhier produced an edition of the philosopher’s diaries between 1954 and 1957. Finally, a more scholarly thirteen-part edition (consisting of seventeen volumes) of the collected works was started in 1984, under the direction of François Azouvi. This edition, for the first time, makes it sufficiently clear to the reader that many of the ‘works’ of Biran were never regarded by him as complete, and that texts have often had to be established by editorial conjecture.

Why Biran completed (and published) so little has been a subject of discussion. A simple answer is that he was too busy with other things. He led an active public life, holding regional and national offices during the revolutionary period, under the consulate of Bonaparte, under the Empire of Napoleon and after the Restoration of the monarchy. Yet this answer seems too simple. Voutsinas diagnosed a deep narcissism in Biran: he avoided publication since he feared the wound of his work not being esteemed. Despite the somewhat obsessive character of Biran’s constant revisions, additions, deletions and reorderings, it may be more enlightening to place his work in its broader intellectual context. For the ‘Science of Man’ was already being fragmented. Biran was thinking about these questions during the birth-pangs of neurophysiology, empirical psychology, social anthropology, economics, political science, linguistics and so on. He was well aware that his project demanded some way of coordinating all these emerging disciplines. He himself made a serious study of some of them. We also find attempts at tabulation. The core should be a reflective psychology which would recognize the will in a way which the British empiricists and their successors had failed to do, and the multiple empirical disciplines studying human nature should be related to and should depend upon this core. But he was never satisfied with his own attempts at seeing how they should fit together. In short, he was pursuing an objective which ran against the temper of the times. Arguably, he was trying to complete a task which, by its nature, could never be completed.

3 The mature position

Biran’s philosophical work stands at a time of transition from the certainties of classical thought. Its character and content permit it to be seen in the context of the meditative tradition of such as Montaigne and Pascal, a tradition which gives prominence to the intimate workings of one’s own mind. This is clear not only from the philosopher’s personal diaries, but also in the central role which he gives to introspection or reflection in his mature theory. Those who see him in this light regard him as part of a lineage which leads to late nineteenth-century French ‘spiritualism’, illustrated by authors like Ravaisson. On the other hand, Biran allies himself strongly with the eighteenth-century empiricist tradition illustrated by authors like David Hartley, Condillac and Charles Bonnet. This emerges not only from his own avowals, but also from his account of the mind which, though diverging from those of his predecessors, is nevertheless clearly a successor to them. It is also

associated with Biran's detailed interest in a variety of fields of empirical study, attested both by the contents of his library, and by the references made in his works. He was interested in physiology, psychiatry, penology, experimental psychology, hypnotism, linguistics, phrenology, ethnography, medicine, pedagogy. In the latter case, it may be noted that he corresponded with Pestalozzi, and appointed one of Pestalozzi's pupils to establish and run a school on Pestalozzian principles in Bergerac (see Education, history of philosophy of §8). It would be insufficient however to give Biran only this double lineage, from the empiricist and the meditative traditions, for he also took an interest in the rationalists, writing on Leibniz. In addition, he read and reckoned with Kant.

It is the works of the first decade of the nineteenth century, culminating in the *Essai*, which develop the position in and about psychology for which Biran is especially known. He accepted what was regarded as a commonplace, that John Locke was the pioneer of mental science. But he held that Locke had made a number of crucial errors. Locke was right to hold reflection to be one of the two sources of knowledge, but wrong to think that it consisted in awareness of 'ideas' (a term newly adapted to mean 'mental representations'). This was to give rise to an essentially passive model of the mind. Perception became the simple reception of data, and action would be either the mechanical result of reception of such data (together with any internal mechanisms at work), or an arbitrary event mysteriously outside the world of causality (see Perception). This view gave rise to a series of problems which Biran thought that no one could answer, or had answered (including Kant). He held that the primary datum of consciousness was the awareness of willed effort (and, in particular, muscular effort) (see Action §4).

The inability of classical empiricism to give an adequate account of the will was illustrated by Hume's claim that the will is 'nothing but *the internal impression we…are conscious of, when we knowingly give rise to any new motion of our body*'. The circularity of this account is frequently criticized by Biran, who insists that what is given to consciousness is essentially relational. An example of a primary fact would be the awareness of raising one's arm. This cannot properly be decomposed into two elements - awareness of a volition and awareness of the arm going up - for then the relation will escape us. 'The inner sense has no object.' What is given is a *fact*, not an object. Of course, we can express our understanding of the fact by referring to the terms of a relation; but the relation is prior to these terms. When I raise my arm, the relation already exists in the effort exerted. It is not that there are two successive components of the experience, either of which might exist without the other.

It may seem that problems arise here. Is there not a causal process capable of empirical study, which consists of successive events in the nervous system which result in the arm going up? Biran conceded this: indeed, he took an active interest in physiology. But he rejected the reductionist view of mental events (see Reductionism in the philosophy of mind). The relation between the awareness of raising one's arm and the physiological story was, he said, 'symbolic'. That is, the one story corresponded systematically but only partially to the other. Neither could be translated into the other without loss.

But if I can be aware of raising my arm, it seems that this presupposes knowledge. I must know that it is my arm that I am raising. This leads to another important difference between Biran and Locke. For Locke thought that ideas were prior to language, and that the function of language was simply to communicate them. Biran, by contrast, held that ideas could not exist without language. Thus a first deliberate raising of the arm is itself the birth of language, since it must bring into being the awareness of the arm. In so far as the movement itself expresses this awareness, it is a sign in the making.

But how can some creatures raise themselves by their bootstraps into a conscious life? Here, Biran departs both from the rationalist and from the empiricist traditions. For he maintains the existence of a passive underlayer of our experience, a great range of affective states which are capable of being drawn up into our conscious life, and which influence it. His interest in this is evidenced not only in his diaries, but also in his philosophical writings, where he discusses at length phenomena such as dreams. There was 'an inner New World to be discovered some day by a Columbus of metaphysics' (*Maine de Biran: Journal*, vol. 1: 176). Physiology and introspection would be needed to 'plunge into the underground caverns of the soul' (*Maine de Biran: Journal*, vol. 1: 240).

These, then, are the main features of Biran's mature position. The awareness of willed effort, a new distinction between passive and active, the crucial role of language, the relational nature of mental phenomena.

4 Later views

While elaborating these views, Biran usually held on to the rejection of 'metaphysics'; but he found increasingly

that questions about the 'primitive facts' could not be properly treated without approaching metaphysical, moral and religious questions. His attempts at dealing with these questions can be seen in the generally more fragmentary works of the last decade of his life.

Near the end of his life, Biran reviewed what he had been doing in his diary, saying that he had come to:

> scorn everything that had previously been my main concern, and to which I had attached some importance and some glory, and I reproached myself for spending my life in erecting a mere scaffold…. However, giving today the chief importance to man's relations with God and with the society of his fellows, I still think that a thorough knowledge of the relations of the *self*…with the concrete person must precede in order of time or study all theoretical or practical research on these two first relations; it is experimental *psychology*, or a science at first purely reflective, that should lead us in due order to the determination of our moral relations with our fellows, and our religious ones with the infinite and higher being…
>
> (*Maine de Biran: Journal*, vol. II: 376)

Biran's partial disillusionment with his own work must be taken seriously: it plays a part in understanding why he completed so little. Yet it opened many very different paths to the future. His claim that mental states are essentially relational ('rien n'est dans la conscience qu'à titre de rapport') prefigures phenomenology (see Phenomenology, epistemic issues in). His claim that there is an internal relation between language and ideas prefigures contemporary critiques by analytic philosophers of the older empiricist tradition. His work coincided with and influenced the early development of empirical psychology. His meditative style prefigures French spiritualism. His claim that there is a new world of the unconscious to be explored by some Columbus versed in physiology and introspection prefigures Freud. And these are not merely points of historical interest. For the central role which he gives to the will, and the way in which he pursues it, are of great current importance, given that late twentieth-century debates about the nature of the mind, including those concerning the development of cognitive science, still often make assumptions which Biran decisively challenged two centuries earlier.

See also: Introspection, psychology of; Will, the

F.C.T. MOORE

List of works

Most of the works of Maine de Biran were not published during his lifetime, and some were published for the first time more than a century and a half after his death.

Maine de Biran, P. F. (1766-1824) *Œuvres de Maine de Biran*, ed. F. Azouvi, Paris: Vrin, 1984-, 13 vols.(The most complete collection of Maine de Biran's works.)

(1766-1824) *Œuvres de Maine de Biran*, ed. P. Tisserand, Paris: Alcan and Presses Universitaires de France, 14 vols, 1920-49.(This was the first attempt at a comprehensive edition of the works of Biran. It is useful but unscholarly.)

(1766-1824) *Maine de Biran: Journal*, ed. H. Gouhier. Neuchâtel: Éditions de la Baconnière, 3 vols.(A valuable edition of most of Biran's surviving diaries. It does not include some 'private' materials which the Naville family in Geneva considered that they could not release, even many years after the death of Biran's last surviving blood relative.)

(1803) *Mémoire sur l'influence de l'habitude sur la faculté de penser*, Paris: Henrichs; trans. M.D. Boehm, *The influence of habit on the faculty of thinking*, Westport, CT: Greenwood Press, 1970.(A discussion of habit and thought.)

(1807) *De l'aperception immédiate (On immediate awareness)*; repr. ed. J. Echeverria, *Maine de Biran: de l'aperception immédiate (mémoire de Berlin 1807)*, Paris: Vrin, 1963.(A work which emphasizes the priority of phenomenology in the 'study of man'.)

References and further reading

Azouvi, F. (1995) *Main de Biran: la science de l'homme (Main de Biran: The Science of Man)*, Paris: Vrin.(A thorough treatment of Biran's approach to the 'science of man'.)

Baertschi, B. (1992) *Les rapports de l'âme et du corps: Descartes, Diderot et Maine de Biran (The relationships between mind and body: Descartes, Diderot et Maine de Biran)*, Paris: Vrin.(A useful work, setting Biran's treatment of the mind/body question in a wider context.)

Baertschi, B. (1982) *L'Ontologie de Maine de Biran (The Ontology of Maine de Biran)*, Fribourg: Éditions

Universitaires.(An interesting but not elementary study of the ontological questions which arise from Biran's mature position.)

Boas, G. (1925) *French Philosophies of the Romantic Period*, Baltimore, MD: Baltimore.(One of the few studies in English. Approachable but superficial.)

Cresson, A. (1950) *Maine de Biran, sa vie, son œuvre (Maine de Biran: His Life and Work)*, Paris.(An acceptable introductory work.)

Delbos, V. (1931) *Maine de Biran et son œuvre philosophique (Maine de Biran and his Philosophical Work)*, Paris: Vrin.(A standard scholarly study.)

Fessard, G. (1938) *La Méthode de réflexion chez Maine de Biran (The Reflective Method of Maine de Biran)*, Paris: Bloud.(A worthwhile exploration of the reflective method.)

Ghio, M. (1962) *Maine de Biran e la tradizione biraniana in Francia (Maine de Biran and the Biran Tradition in France)*, Turin: Edizioni di Filosofia.(A fair introductory work for readers with good Italian.)

Gouhier, Henri (1948) *Les Conversions de Maine de Biran*, Paris: Vrin.(The best overall study of Biran's work. It is not intended as introductory reading.)

Gouhier, Henri (1970) *Maine de Biran par lui-même (Maine de Biran by Himself)*, Paris: Editions du Seuil.(An excellent introduction to Biran for the interested layman.)

Hallie, P.P. (1959) *Maine de Biran, Reformer of Empiricism, 1766-1824*, Cambridge, MA: Harvard University Press.(An adequate introduction to Biran's philosophy for the reader who requires a text in English.)

Henry, M. (1965) *Philosophie et phénoménologie du corps (essai sur l'ontologie biranienne) (Philosophy and Phenomology of the body (Essay on the Ontology of Biran))*, Paris: Presses Universitaires de France.(An essay of some interest written from a phenomenological point of view.)

Lacroze, R. (1970) *Maine de Biran*, Paris: Presses Universitaires de France.(This overview need not be neglected.)

La Valette Monbrun, A. de (1914) *Maine de Biran: essai de biographie*, Paris: Fontemoing. (This is the only biography, and is therefore mentioned here in spite of the fact that it is unscholarly and unsatisfactory.)

Lemay, P. (1936) *Maine de Biran et la Société Médicale de Bergerac (Maine de Biran and the Medical Society of Bergerac)*, Paris, Liège.(A worthwhile study of Biran's Medical Society, and his contributions to it.)

Le Roy, G. (1937) *L'Expérience de l'effort et de la grâce chez Maine de Biran (The Experience of Effort and of Grace in Maine de Biran)*, Paris: Boivin.(Takes up Biran's later interest in the spiritual life as a higher passivity.)

Montebello, P. (1994) *La décomposition de la pensée: dualité et empirisme transcendantal chez Maine de Biran (The Deconstruction of Thought: Duality and Transcendental Empiricism in Maine de Biran)*, Grenoble: Millon.(Interprets Biran's position as a form of transcendental empiricism.)

Moore, F.C.T. (1970) *The Psychology of Maine de Biran*, Oxford: Clarendon Press.(A critical review of Biran's mature position, from a broadly analytic perspective. Contains a useful bibliography.)

Paliard, J. (1925) *Le Raisonnement selon Maine de Biran (Reasoning according to Maine de Biran)*, Paris: Alcan. (A useful study of Biran's view of reasoning.)

Rhomeyer-Dherbey, G. (1974) *Maine de Biran, penseur de l'immanence radicale (Maine de Biran, Thinker of Radical Immanence)*, Paris: Segers.(This study may be read with some profit.)

Truman, N.E. (1904) *Maine de Biran's Philosophy of Will*, New York: Macmillan.(This, in spite of various defects, is a quite useful and approachable work for the reader requiring a text in English.)

Vancourt, R. (1944) *La Théorie de la connaissance chez Maine de Biran (The Theory of Knowledge in Maine de Biran)*, Paris: Aubier, 2nd edn.(A good study of Biran's epistemology.)

Voutsinas, D. (1964) *La Psychologie de Maine de Biran (The Psychology of Maine de Biran)*, Paris: SIPE.(A somewhat eccentric diagnosis of Biran, which nevertheless has points of interest.)

Major, John (1467-1550)

John Major was one of the last great logicians of the Middle Ages. Scottish in origin but Parisian by training, he continued the doctrines and the mode of thinking of fourteenth-century masters like John Buridan and William of Ockham. Using a resolutely nominalist approach, he developed a logic centred on the analysis of terms and their properties, and he applied this method of analysis to discourse in physics and theology. Although he came to oppose excessive dependence on logical subtlety in theology and maintained the authority of Holy Scripture, Major's work was stubbornly independent of the growing influence of humanism in Europe. Later, he would be regarded as representative of the heavily criticized 'scholastic spirit', being referred to disparagingly by Rabelais as well as by later historians such as Villoslada (1938), but at the beginning of the sixteenth century, his teaching influenced an entire generation of students in the fields of logic, physics and theology.

1 Life

John Major (or Mair) was born in 1467 at Gleghornie near Haddington in Scotland. He began his studies at Haddington, and after a period at Cambridge, went to Paris. He received his MA in 1494, and in 1495 began to teach as a regent master at the College of Montaigu. In 1506 he became a Doctor of Theology. In 1518 he returned to Scotland where he was principal of the University of Glasgow, as well as teaching logic and theology. From 1523 to 1526 he taught logic and theology at the University of St Andrews. Returning to Paris in 1526, he re-edited his works on logic, as well as publishing commentaries on Aristotle's *Physics* (1526) and *Nicomachean Ethics* (1530), and on the four Gospels (1529). In 1533 he became Provost of St Salvator's College at St Andrews, and remained there until his death in 1550.

2 Logic

Although in the last period of his life at St Andrews Major was regarded as a great authority in theology, he first became famous as a logician. Between 1499 and 1506 he published a large number of logical works, later published as a collection. These works included commentaries on Aristotle and on Peter of Spain, treatises on terms and their properties, on exponibles, consequences, insolubles and obligations (see Logic, medieval).

John Major was trained by the masters who had taken part in the revival of nominalism in the last twenty-five years of the fifteenth century, especially Thomas Bricot, who edited the logic of John Buridan (§2). John Major himself edited John Dorp's early fifteenth-century commentary on Buridan in 1504, and he also edited the logic of the young Spaniard Jerónimo Pardo (d. 1502 or 1505). In his own logical works, his approach was that of a terminist. He accorded a central place to the study of the properties of terms in propositions, particularly their different types of reference, a subject which was thought to be indispensable for avoiding errors of reasoning. This approach goes back to the great synthesizers of the thirteenth century, especially Peter of Spain who was the favourite reference point for the Parisian masters, but it culminated at Paris with the work of John Buridan who placed the study of terms in a nominalist context. In this Major remained faithful to Buridan, often referring to and explicitly adopting 'the solution of the nominalists'.

Major's logic presupposes a theory of spoken and conceptual signs which owed a great deal to William of Ockham (§§6-7). In his treatises on terms, John Major assumed the fundamental presupposition of late medieval logico-linguistic analysis: a term is a sign, and as a logico-linguistic sign, it can only be contemplated in the context of a general semiotic theory. Having defined a sign generally as that which is taken in place of a thing, Major departed from Buridan to assume a position closer to that of Ockham or Albert of Saxony. For him, a spoken sign refers in a proper and primary way to a thing and not, as Buridan had held, to a concept. It is none the less necessary to take into account the properly conceptual components of signification, and so John Major gave a very detailed analysis of the different types of referential relation between the word, the concept, and the thing signified. In this context, he gave a long analysis of the *sophisma* or puzzle case, *debeo tibi equum* ('I owe you a horse'), explaining the different positions which had become standard, namely those of William of Ockham, William Heytesbury, and the one Major preferred, that of John Buridan. This involves distinguishing between the contexts in which a thing is owed under a certain description (the 'appellation of reason'), and those contexts in which direct reference is made to a particular object.

There were very lively debates during this period about the structure of mental language, and John Major developed a theory of mental language close to that of Pierre d'Ailly (see Ailly, P. d'), in which signifying, thinking and representing have a tendency to be assimilated to one another. Following an idea due to Gregory of Rimini, Major admitted that there were two types of mental language, one of which is the image of spoken language, while the other forms mental language properly so-called. He criticized Ockham's view that synonymy as well as some accidental characteristics, such as grammatical gender, could form no part of mental language.

John Major is a witness to the continued life and capacity for innovation of logic at the University of Paris at the beginning of the sixteenth century. He sometimes shows a tendency to wordiness, and in his hands theories become over-refined and over-complicated. However, this complication primarily shows his desire to take into account the difficulties which had become apparent in the course of the two previous centuries, while continuing to make use of terminist and nominalist principles of analysis.

3 Natural philosophy, theology and history

In his commentary on the *Physics* of Aristotle (see Aristotle §8), John Major followed the initiatives of John Buridan (§4) and Marsilius of Inghen (§3) by reformulating problems of natural philosophy on the basis of an analysis of propositions. At the same time, he showed an interest in gathering observations of natural phenomena. In his treatise *De l'infini (On the Infinite)*, added to his logical works in 1506, he took up the distinction between categorematic and syncategorematic senses of the term 'infinite' which had been popularized by Buridan, and he defended the existence of an actual infinite (see Natural philosophy, medieval).

John Major was also a theologian. His commentary on the *Sentences* of Peter Lombard, published in various versions from 1509, was considered by the humanists as representative of a theology which used logic too broadly, introduced too many properly philosophical considerations, and made too regular appeal to Aristotle and Averroes (see Ibn Rushd §§2, 5). It is true that Major, who relied heavily on the commentary by Gregory of Rimini did, like Gregory, include a great deal of logic and philosophy in his theological discussion. None the less, he recommended a moderate attitude on this point. He affirmed the primacy of Holy Scripture, and criticized the excessive use of dialectical subtleties in the study of Scripture. He himself distinguished scholastic theology from positive theology, by which he seems to have meant the more legalistic and moralistic developments of religious teaching.

In religion Major seems to have been animated by the desire to reconcile orthodoxy with the critical approach of nominalism. He regretted the fact that some of his students turned towards the Reformation, and for his part he defended fidelity to dogma and submission to the Catholic Church. In his *In Mattheum ad literam expositio (Literal Exposition of Matthew)* (1518), and then in his commentary on the four Gospels (In quator evangelia expositiones) (1529), he firmly defended the Roman Catholic Church against the theories of Wyclif, Hus and Luther.

In his politico-ecclesiastical works, however, he showed the influence of the Gallican and conciliarist traditions. He limited the power of the Pope in temporal matters, and above all he defended the superiority of church councils over the Pope on the grounds that the fullness of power resides in the church itself, as represented by a council (see Political philosophy, history of §§4, 8).

His *History of Greater Britain*, published in 1521, studies the history of Scotland as parallel to the history of England, and recommends a bringing together of the two kingdoms. The work is written in a critical manner, rejecting a number of traditional stories.

4 Influence

Major's teaching gave birth to a whole school. His pupils (the Spaniards Luis and Antonio Coronel, Gaspar Lax, the Scotsmen Robert Caubraith, David Cranston, George Lokert, the Belgian Peter of Brussels and John Dullaert, among others) carried on the teaching they had received at Montaigu. While in theology some of them, such as Patrick Hamilton (who was burned for heresy at St Andrews in 1527) and the famous reformer John Knox, embraced Protestantism, in logic most were faithful to nominalism. However, the intellectual climate was to change, especially in Paris, becoming hostile to logical subtleties. Not only did interest develop in languages and rhetoric, but polemical texts called into question the very principles of the logical and scientific developments of

the preceding centuries. The work of John Major was done just before this break in tradition, and his merit is that he assured the last hours of glory for the old forms of thought at the University of Paris, with a force that allowed it to continue for a while in Scotland and in Spain.

See also: Buridan, J. §§2-3; Language, medieval theories of; Language, Renaissance philosophy of §§1-2; Logic, medieval; Logic, Renaissance; Nominalism §2

Translated from the original French by E.J. Ashworth
JOËL BIARD

List of works

Major, J. (1506a) *Le Traité 'De l'infini' de Jean Mair (John Major's Treatise 'On the Infinite')*, ed. H. Élie, Paris: Vrin, 1938.(Modern edition of *De l'infini* with translation into French.)

Major, J. (1506b) *Libri quos in artibus in collegio Montis acuti parisius regentendo compilavit (The books which John Major put together while teaching arts at the College of Montaigu in Paris)*, Paris.(This collection of nineteen logical works was reprinted, with some additions and changes, in 1508, 1513 and 1516. There are also a great many early editions of individual works: for details see Farge 1980.)

Major, J. (1518) *In Mattheum ad literam expositio (Literal Exposition of Matthew)*, Paris: Jean Granjon.

Major, J. (1521) *A History of Greater Britain, as well England as Scotland*, trans. A. Constable, Edinburgh: Scottish History Society, 1892.(English translation.)

Major, J. (1529) *In quator evangelia expositiones (Commentary on the Four Gospels)*, Paris: Josse Bade.

References and further reading

Biard, J. (1986) 'La logique de l'infini chez Jean Mair' (The Logic of the Infinite in John Major), *Études philosophiques* 3: 329-48.(Study of the relations between logic and physics in Major's treatise *De l'infini*.)

Broadie, A. (1985) *The Circle of John Mair: Logic and Logicians in Pre-Reformation Scotland*, Oxford: Clarendon Press.(Thematic study of logical theories at the time of John Major and his pupils.)

Burns, J.H. (1954) 'New Light on John Major', *Innes Review* 5: 83-100.(Detailed bibliography of Major's works.)

Élie, H. (1951) 'Quelques Maîtres de l'université de Paris vers 1500' (Some Masters of the University of Paris Around 1500), *Archives d'histoire doctrinale et littéraire du moyen âge* 18: 193-242.(Presents the life and career of fifteenth- and sixteenth-century philosophers, with special emphasis on Major.)

Farge, J.F. (1980) *Biographical Register of Paris Doctors of Theology 1500-1536*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(A very useful reference work for the biography of many early sixteenth-century philosophers. Includes a biography and bibliography of Major on pages 304-11.)

Oakley, F. (1961-2) 'On the Road from Constance to 1688: the Political Thought of John Major and George Buchanan', *The Journal of British Studies* l (2): 1-31. (Discusses parallels between Major's conciliarism and the case of secular rulers.)

Oakley, F. (1965) 'Almain and Major: Conciliar Theory on the Eve of the Reformation', *The American Historical Review* 70: 673-90.(Study of the theory of the primacy of a general council over the pope.)

Renaudet, A. (1918) *Préréforme et humanisme à Paris (1497-1517) (The Pre-Reformation Movement and Humanism in Paris)*, Paris: Libraire d'Argences; 2nd edn, 1953; repr. Geneva: Slatkine, 1981.(Contains a chapter on John Major, in the context of a study of the general evolution of ideas at the end of the fifteenth century and the beginning of the sixteenth century.)

Villoslada, R. (1938) *La universidad de Paris durante los estudios de Francisco de Vitoria, O.P. (1507-1522) (The University of Paris during the Studies of Francisco de Vitoria)*, Rome: Gregorianum.(Referred to in the introduction. Chapter 6 is devoted to John Major, from a point of view very critical of scholastic thought.)

Malebranche, Nicolas (1638-1715)

Nicolas Malebranche (1638-1715), a French Catholic theologian, was the most important Cartesian philosopher of the second half of the seventeenth century. His philosophical system was a grand synthesis of the thought of his two intellectual mentors: Augustine and Descartes. His most important work, De la recherche de la vérité (The Search After Truth), is a wide-ranging opus that covers various topics in metaphysics, epistemology, ethics, physics, the physiology of cognition, and philosophical theology. It was both admired and criticized by many of the most celebrated thinkers of the period (including Leibniz, Arnauld and Locke), and was the focus of several fierce and time-consuming public debates. Malebranche's philosophical reputation rests mainly on three doctrines. Occasionalism - of which he is the most systematic and famous exponent - is a theory of causation according to which God is the only genuine causal agent in the universe; all physical and mental events in nature are merely 'occasions' for God to exercise his necessarily efficacious power. In the doctrine known as 'vision in God', Malebranche argues that the representational ideas that function in human knowledge and perception are, in fact, the ideas in God's understanding, the eternal archetypes or essences of things. And in his theodicy, Malebranche justifies God's ways and explains the existence of evil and sin in the world by appealing to the simplicity and universality of the laws of nature and grace that God has established and is compelled to follow. In all three doctrines, Malebranche's overwhelming concern is to demonstrate the essential and active role of God in every aspect - material, cognitive and moral - of the universe.

1 Life and works

Nicolas Malebranche was born in Paris on 6 August 1638, one of the many children of Catherine de Lauzon and Nicolas Malebranche, a royal secretary. Because of a malformation of the spine which caused lifelong pain, he was kept at home for his education, under the direction of his mother, until the age of sixteen. In 1656, he graduated from the Collège de la Marche. The education he received there, and from three years studying theology at the Sorbonne, was heavily laden with Aristotelianism, and it left Malebranche highly dissatisfied. After rejecting the offer of a canonry at Notre-Dame de Paris, Malebranche entered the Oratory in 1660 and was ordained in 1664.

His four years in the Oratory proved to be of great intellectual consequence. While studying the Bible, ecclesiastical history and Hebrew, Malebranche, like other Oratorians, immersed himself in the writings of Augustine. There were also Cartesians among his teachers, who introduced him to the doctrines of Descartes. He did not read any of Descartes' works, however, until 1664 when he happened upon a copy of Descartes' *Treatise on Man* (*L'homme*) in a bookstall. The event was life-changing: according to his biographer, Father André, the joy of becoming acquainted with so many discoveries 'caused him such palpitations of the heart that he had to stop reading in order to recover his breath'. Malebranche devoted the next ten years of his life to studying mathematics and philosophy. He was particularly taken by Descartes' critique of the Aristotelian philosophy that he had earlier found so stultifying and sterile (see Medieval philosophy).

Those ten years of study culminated in the publication in 1674-5 of *De la recherche de la vérité (The Search After Truth)*, Malebranche's first and most ambitious work and a synthesis of the systems of Augustine and Descartes. Malebranche's stated goal in the *Recherche* is to investigate the sources of human error and to direct us towards the clear and distinct perception of truth - truth about ourselves, about the world around us and about God. His motives are deeply theological, and he is ultimately concerned to demonstrate the essential and active role of God in every aspect - material, cognitive and moral - of the created world. The *Recherche*, quickly supplemented by seventeen *Eclaircissements*, contains early but solidly-argued presentations of Malebranche's three most famous doctrines: the vision in God, occasionalism and his theodicy.

While the Abbé Simon Foucher, canon of Sainte Chapelle of Dijon, was the first in a long line of critics of Malebranche's doctrines, the Jansenist theologian and Cartesian philosopher Antoine Arnauld was undoubtedly the harshest and most acute (see Foucher, S.; Arnauld, A. §3). Arnauld approved of the *Recherche* upon first reading it. But when he later learned of Malebranche's views on grace and divine providence - sketchily presented in the *Recherche* but more fully expounded in the *Traité de la Nature et de la Grace (Treatise on Nature and Grace)* in 1680 - he embarked on a detailed critique of the major theses of the *Recherche*. Arnauld's *Des vraies et des fausses idées* (1683), and Malebranche's reply, *Réponse du Père Malebranche au livre des vraies et des*

fausses idées (1684), were only the opening salvos in what would come to be a long, often bitter public battle on both philosophical and (to the participants more importantly) theological matters. Although Arnauld's allies succeeded in having the *Traité* put on the Index Librorum Prohibitorum in 1690 (the *Recherche* was added in 1709), their exchanges - public and private - continued until Arnauld's death in 1694. The debate, one of the great intellectual events of the seventeenth century, attracted the attention of Leibniz, Spinoza, Bayle, Locke, Newton and many others.

After the publication of the *Recherche*, Malebranche turned to a 'justification' of the Catholic religion and morality, presented in suitably Malebranchian terms and published as the *Conversations Chrétiennes* in 1676. This was followed in 1683 by the *Médiations Chrétiennes et Métaphysiques*, which consists of dialogues in which 'the Word' explains and defends Malebranche's system. That same year Malebranche also published his *Traité de Morale* in which he undertakes a rigorous demonstration of a true Christian ethics.

By the mid-1680s, Malebranche was widely regarded as the most important, if highly unorthodox, representative of the Cartesian philosophy. His regular correspondents included Leibniz and the physicist Pierre-Sylvain Régis. With Leibniz he debated the Cartesian account of the laws of motion (Malebranche published his *Traité des lois de la communication du mouvement* in 1692), as well as occasionalism and the nature of causal relations. With Régis, who defended a more orthodox brand of Cartesianism, he discussed natural philosophy and the nature of ideas (see Leibniz, G.W. §11; Régis, P.-S.).

Having been forced in his arguments with Arnauld, Régis and others to clarify, develop and even modify his doctrines, Malebranche decided, at the urging of friends, to compose a treatise in which he would both present an up-to-date and concise picture of his theories and defend them as a proper Augustinian (and Catholic) system. The *Entretiens sur la métaphysique (Dialogues on Metaphysics)* were published in 1688 and were supplemented in 1696 by the *Entretiens sur la mort*, which Malebranche wrote after an illness from which he did not expect to recover. In 1699, he was elected to the Académie Royale des Sciences.

During the last fifteen years of his life, Malebranche remained actively engaged in philosophical, theological, and scientific matters, publishing the *Entretien d'un philosophe chrétien et d'un philosophe chinois, sur l'existence et la nature de Dieu* in 1708 and his *Réflexions sur la prémotion physique* in 1715. He also continued to work on the *Recherche*, producing the sixth edition, the last to appear in his lifetime, in 1712. In June 1715, Malebranche became ill while visiting a friend in Villeneuve St. Georges. He was taken back to the Oratory in Paris, where he died on October 13.

2 Vision in God

Malebranche's system is fundamentally Augustinian, and he was inspired by Descartes' philosophy not just because of the scientific and mathematical discoveries he found there but especially because of its own Augustinian nature. He tried to be as faithful as possible to Augustine's thought while at the same time incorporating into his doctrines what he took to be the important metaphysical and epistemological insights of the new mechanistic science. This is particularly evident in his theory of ideas and account of human knowledge.

Nearly all philosophers in the seventeenth century agreed that perception and knowledge is mediated by immaterial representations immediately present to the mind, after Descartes generally called 'ideas'. There was much disagreement, however, over the origin and ontological status of these ideas: How do ideas become present to the mind? Are ideas 'modifications' of the mind? Are they *acts* of the mind or the *objects* of perception? Malebranche's doctrine of the vision in God was the most unorthodox and controversial theory of ideas in the period. He looked back to the Christian-Platonic and Augustinian model, according to which ideas proper are archetypes or essences in the divine understanding. Human beings have access to these ideas, which serve as the ground of all eternal truths, through a continuous process of illumination that informs their cognitive powers.

Malebranche orients his discussion of ideas and knowledge in the *Recherche* around the problem of error, particularly the errors that arise when we base our judgments on the testimony of the senses and the imagination, rather than on reason and understanding. He relies in his analysis of error on what is probably the most comprehensive and systematic account of the physiology of sense and imagination in the seventeenth century. For example, based upon a detailed investigation into the structure of the eye and the geometry of optics, he argues that our sight does not present to us the extension of external bodies as it is in itself, but rather only as it is in

relation to our body. He describes the eye as a kind of 'natural spectacles', and examines the ways in which the material images conveyed into the brain (and thus the sensory images these stimulate in the soul) are variously affected by the distance between the eyes, the different humours in the eyes, the shape of the crystalline lens and its distance from the retina, and other factors. The difference between sensation and imagination is due simply to the fact that in imagination the tiny fibres in the centre of the brain are agitated not by the impressions made by external objects on the exterior surface of the nerves, but by the flow of the animal spirits initiated by the soul itself. These animal spirits are easily affected by the changing state of the body and by external things, as well as by 'moral' causes (prejudices, 'conditions of living'). The imagination thus represents a second important source of the differences between minds and how things are represented to them. Malebranche's goal in highlighting how the senses and the imagination 'lead us into error' is to warn us not to allow their images to serve as a basis for judgment about the truth of things. We should rely, rather, on the testimony of reason and the perfect evidence of clear and distinct ideas.

Malebranche's account of the nature of ideas is grounded in the basic dualist framework of Cartesian ontology. Mind, or thinking substance, is unextended thought, and has absolutely nothing in common with matter, defined as extension. A material body has only the mathematical properties of shape, size, divisibility and mobility. All other sensible qualities - colour, heat, cold, taste, odour and so on - are really only sensations in the mind, mental modifications occasioned by external material objects. While such sensations may indicate to us the presence of bodies to our own body, they cannot provide us with clear and distinct knowledge of those bodies. Sensations can only inform us of what is presently taking place in our own minds; they have no representational value with respect to the external world. Malebranche claims, however, that in addition to our own inner sensations we have access to representative ideas. Ideas, unlike qualitative sensations, have a clear and distinct representational content - generally of a quantitative nature - and provide us with unambiguous and complete knowledge about objects and their properties. The idea of the square, for example, presents with perfect evidence all the information needed for full knowledge of the geometric properties of squares. On the other hand, the heat one feels when near a fire is simply an obscure sensation that reveals nothing about the nature of fire itself as an extended reality. Malebranche's ideas are pure concepts, and their content has no sensory component whatsoever. They just are the logical essences of things or kinds. His epistemological distinction between ideas and sensations derives from Descartes' distinction between clear and distinct ideas and obscure, confused ideas. It roughly corresponds to the distinction made famous by Boyle and Locke between primary quality ideas and secondary quality ideas (see Boyle, R.; Descartes, R. §8; Locke, J. §4).

This epistemological difference between ideas and sensations is, for Malebranche, grounded in an ontological one. While sensations are mental, ideas are not. Representative ideas are 'present to the mind', but they are not modifications of the mind's substance. (Their presence is what allows them to represent extended beings to the mind.) Rather, they are in God, and finite minds have access to them because God wills to reveal them to those minds, all of which exist in a perpetual union with God. That ideas do not belong to the mind becomes clear from the fact that some of our ideas are infinite, and it is impossible for a finite mind to have an infinite modification. Likewise, our ideas are all general - they only become particular or specific when combined with some sensory components - and a particular substance such as the mind cannot have a modification that is general.

Ideas function in all of the mind's cognitive activities; most importantly, in conception and perception. In conception, the mind apprehends a pure idea by itself, without any sensations to particularize the experience (for example, conceiving a geometric circle). In perception, ideas are present but are accompanied by various sensory elements. Thus, when we perceive the sun, the pure idea of a geometric circle is accompanied by colour (yellow) and heat sensations.

Malebranche offers two kinds of argument to demonstrate that representative ideas are required for knowledge and perception. First, he argues that there is generally an unbridgeable distance between the mind and its objects. He sometimes appears to mean this literally, and to be relying on the fact that the mind is not locally present to the external bodies it knows and sees, and that in cognition it is neither the case that objects travel to the mind or that the mind leaves the body to 'travel across the great spaces' that separate it from its objects. His more considered intention, however, is probably to draw our attention to the metaphysical, rather than the physical gulf that separates mind from matter: bodies, being extended, cannot be united with, and thus present to, the unextended mind in the way required for direct cognitive acquaintance. Thus, what is required are intermediary entities that

can both represent extended bodies and be immediately present to the mind: ideas.

Second, it is often the case that we have a perceptual experience of an object when in fact the object itself does not actually exist (for example, in hallucinations and dreams). But Malebranche insists upon the truth of the principle of intentionality, or the claim that every perception must be object-directed - ‘to perceive is to perceive something’ (see Intentionality). As Malebranche understands this principle, every perception must be the direct and immediate apprehension of some really present object. As the illusory cases illustrate, however, the intentional objects of the mind cannot be really existing material bodies; for if they were, we could never have perceptual experiences of objects that do not exist. Thus, it must be that we directly apprehend ideas, non-material representations, even though sometimes there is no external body corresponding to the idea.

3 Vision in God (cont.)

Malebranche’s position is that the ideas that function in human perception and knowledge are simply the ideas - the eternal essences and archetypes - in the divine understanding: that ‘we see all things in God’. In the *Recherche* itself, he relies mainly on an argument from elimination. He shows how all other accounts of the source of our ideas - the Scholastic and Epicurean doctrines, various Cartesian theories (innatism, self-production by the soul) - are untenable, thus leaving the vision in God as the only viable alternative (see Epicureanism). But in this work and others, he also marshals more positive considerations in support of his doctrine. His account, he insists, is simpler than any other hypothesis (hence more worthy of God’s ways), and best fosters a proper and pious sense of our ontological and epistemological dependence upon our creator: we as knowers and perceivers are not self-sufficient, no more than we as beings are self-sufficient substances.

In the *Eclaircissements* appended to the *Recherche*, Malebranche’s debt to Augustine becomes more overt. For Augustine, what we see in God are eternal and immutable truths. Truth is, by its nature, changeless, universal and uncreated. Moreover, truth is higher than, and common to, many minds. Hence, it can be nowhere but in the divine reason, in God himself (in Augustine’s words, truth *is* God) (see Augustine §§6-7). For Malebranche, too, truth is necessarily universal, immutable and infinite. But the truths we know just are relations between ideas. Thus, the ideas themselves must be universal, immutable, and infinite. And such ideas can only be those in God’s understanding. The vision in God is the only possible explanation for our common knowledge of necessary truth - we are all similarly united with one universal, infinite Reason, in which we perceive the same ideas.

Malebranche’s doctrine of the vision in God is motivated, then, not just by the problem of how we perceive bodies in the external world - a specifically Cartesian concern - but primarily by the problem of how we have knowledge of eternal truths. And behind it all lies Malebranche’s deep fear of the dangers of scepticism. The vision in God represents for Malebranche the most effective countersceptical strategy available. First, we can be sure that the ideas we apprehend in sense-perception really do represent bodies in the world because these ideas just are the archetypes that served and directed God in creating those bodies. Thus, they cannot fail to reveal the nature of extended things as they are. On the other hand, if ideas were, as many Cartesians insist, merely fleeting and subjective modifications of the soul, there would be no justification for believing that what they represent about bodies in the world really characterizes those bodies. This would seriously undermine the certainty of physics. Second, any sceptical worries about the necessity, universality and objectivity of mathematical and moral knowledge are forestalled by showing that the ideas upon which such knowledge is based are mind-independent (non-subjective) realities, accessible to all knowers in a universal Reason.

To be sure, there are many relevant questions that the doctrine of vision in God cannot, and is not intended to, answer. For example, while the ideas we apprehend in God inform us as to the essences of things, they cannot provide any evidence about the existence of things. Certainty in this regard can only come about through a combination of faith and sensory experience. This is not to say that Malebranche is a sceptic when it comes to the existence of the external world. Although he claims that we cannot have rational and demonstrative certainty about the existence of a world of things, we still have no good reasons to doubt of it, as well as a natural propensity to believe in it. Nor does the vision in God provide us with a clear and distinct idea of the soul, which would make possible a science of the soul as certain as our science of bodies (physics) based on the idea of body. Our knowledge of the soul is limited to the testimony of *sentiment intérieur*, to what we actually experience in consciousness, and Malebranche rejects Descartes’ claim that we know the soul as well as (or even better than) the body.

The vision in God is a systematic attempt to combine the doctrine of divine illumination that Malebranche found in Augustine with a somewhat deviant Cartesian metaphysics and philosophy of mind. The theory is deeply Augustinian in inspiration, but is geared also to answer certain epistemological questions that really only arise in the seventeenth century.

As Malebranche’s contemporary critics were quick to point out, the doctrine was fraught with ambiguities and inconsistencies. For example, if there is a single immutable idea in God for each object or kind of object he created, how is it that we perceive change and motion in extended things? In the *Eclaircissements* Malebranche denies that he ever meant that there is a plurality of individual, discrete ideas in God. He insists that there is in God an ‘infinite intelligible extension’, and that by applying our minds to this extension in various ways we apprehend representations of different extended bodies in different states of being.

Foucher, in one of his criticisms of Malebranche, focused on the ontological status of ideas. If ideas are spiritual, but are not substances, then they must be modifications of some spiritual substance. Malebranche denies that they are (he insists that they are neither modifications of our minds nor of the divine substance), but then what else can they be? And if they are not ‘ways of being’ of our minds, then they must be as external to our minds as anything else and not ‘present’ to the mind in the manner required for direct cognition. Foucher also asks how immaterial ideas could possibly represent material bodies that they do not resemble (see Foucher, S. §3).

Arnauld argued that the whole notion of ideas as representative beings, distinct from the mind’s perceptions and which are in fact the mind’s objects in perception, is false and even incoherent. Ideas, Arnauld insisted, are not distinguished from the mental acts (he calls them *perceptions*) which represent objects to the mind but are not themselves distinct objects. He claims that Malebranche’s theory, far from explaining how we perceive bodies external to us, actually demonstrates that we never perceive such bodies, and that the mind is surrounded by a ‘palace of ideas’ beyond which it has no cognitive access. Arnauld also focused in his attack on Malebranche’s doctrine of the infinite intelligible extension, and wonders how Malebranche can avoid the charge of materialism, of having placed extension really (that is, ‘formally’) in God and thus making God himself extended (see Arnauld, A. §3). This point is taken up in Spinozistic terms by Dortuous de Mairan, a young man whom Malebranche had once tutored. Mairan was impressed by Spinoza, and he challenged Malebranche both to refute Spinoza’s arguments and to show how the relationship between the infinite intelligible extension in God and the material extension of bodies differs from the substance/affection (or mode) relationship that for Spinoza characterizes the relationship between infinite extension and particular bodies (see Spinoza, B. §§2-3).

4 Occasionalism

Just as the doctrine of the vision in God demonstrates the epistemological dependence that we as knowers have upon God, so the causal doctrine of occasionalism demonstrates the ontological dependence that we and *all* beings have upon an omnipotent God (see Occasionalism). Nothing, it claims, exists or happens in the universe that is not a direct and immediate effect of the divine will. Although occasionalism has its ancestry in certain medieval Islamic, Jewish and Christian theories of causation and divine omnipotence, especially the voluntarist tradition (see Al-Ghazali; Damian, P.), as well as in Descartes’ metaphysics; and while there were others before Malebranche who were, to one degree or another, occasionalists (see Clauberg, J.; Cordemoy, G. de; Geulincx, A.; La Forge, L. de), Malebranche was the first to argue systematically for a thoroughgoing and rigorous version of the doctrine.

Occasionalism is the doctrine that all finite created entities are absolutely devoid of causal efficacy and that God is the only true causal agent. God is directly, immediately and solely responsible for bringing about all phenomena. When a needle pricks the skin, the physical event is merely an occasion for God to cause the appropriate mental state (pain); a volition in the soul to raise an arm or to think of something is only an occasion for God to cause the arm to rise or the idea to be present to the mind; and the impact of one billiard ball upon another is an occasion for God to move the second ball. In all three contexts - mind-body, mind alone and body-body - God’s ubiquitous causal activity proceeds in accordance with certain general laws, and (except in the case of miracles) he acts only when the requisite material or psychic conditions obtain.

Far from being an *ad hoc* solution to a Cartesian mind-body problem as it has traditionally been portrayed, occasionalism is argued for by Malebranche (and others) from general philosophical considerations of the nature

of causal relations, from an analysis of the Cartesian concept of matter and, perhaps most importantly, from theological premises about the essential ontological relationship between an omnipotent God and the created world that he sustains in existence.

Malebranche's first argument that there are no real causal powers in finite created substances and that God is the sole causal agent, focuses on the motion of bodies. He begins with the causal principle that in order for one thing, *A* (which can be a substance or a state of being of a substance), to count as the cause of another thing, *B*, there must be a necessary connection between the existence of *A* and the existence of *B*. 'A true cause as I understand it is one such that the mind perceives a necessary connection [*liaison nécessaire*] between it and its effects' (1674-5 vol. 2: 316). But we can find no such connection between any two physical events, nor between any human mental event and a corresponding physical event. For example, it is certainly conceivable that one can will to raise one's arm but the arm will not rise. 'When we examine our idea of all finite minds, we do not see any necessary connection between their will and the motion of any body whatsoever. On the contrary, we see that there is none and that there can be none' (1674-5 vol. 2: 313). When we consider God, however, as an infinitely perfect being, we see that there *is* such a necessary connection between the divine will and the motion of bodies, since it is logically impossible that an omnipotent God should will to move a body and it does not move; such is the nature of omnipotence. God, therefore, is the only true cause of the motion of bodies.

Malebranche's second argument is based on the 'inconceivability' that any natural cause, any finite mind or material body, should have 'a force, a power, an efficacy to produce anything' (1674-5 vol. 3: 204). First, an idea of body - that is, the clear and distinct idea of extension - represents it as having only one property: the entirely passive faculty of 'receiving various figures and various movements' (1688: 148). It certainly does not represent body as having any active power. Here Malebranche is drawing out the ramifications of Descartes' conception of matter: a Cartesian material body, *qua* pure extension, is essentially passive and inert, and devoid of any motive force. In fact, such a force or power is perceived as *incompatible* with the notion of extension, since it cannot be reduced to or explained in terms of relations of shape, divisibility and distance. Thus, bodies cannot *act*, whether on minds or other bodies. Second, whatever minimal knowledge I have of my soul does not involve the perception of any power, whether to move the body or even to produce its own ideas. All I perceive through inner consciousness is an actual volition to move my arm upwards, and all I notice in my body is that my arm subsequently rises. But I do not perceive, either by inner consciousness or by reason, any power on the part of the soul by means of which it might effect this motion. It is in this sense that 'those who maintain that creatures have force and power in themselves advance what they do not clearly perceive' (1674-5 vol. 3: 204). Indeed, according to Malebranche, I perceive a general incompatibility between the idea of a created finite being and such a power or productive faculty. Only in my idea of the will of an infinite being do I clearly and distinctly recognize any element of power whatsoever.

The third argument is based on a supposedly intuitive premise which (echoing an argument for occasionalism introduced by Geulincx) sets an epistemic condition on the notion of 'cause': in order to count as the cause of an effect, a thing must know how to bring about that effect. Malebranche then appeals to the evident fact that this condition is not satisfied by our minds in order to show that we do not, in fact, cause those motions that we consider voluntary: 'There is no man who knows what must be done to move one of his fingers by means of animal spirits' (1674-5 vol. 2: 315). This same condition rules out the mind's ability to produce its own ideas. It also rules out, *a fortiori*, the possession of causal efficacy by bodies.

Malebranche's most general argument against real interaction appeals to God's role as creator and sustainer of the universe. The argument - which has its roots in both medieval and Cartesian doctrines - purports to show that it is an 'absolute contradiction' that anything besides God should move a body. God's activity is required not only to create the world, but, since creatures are absolutely dependent on God, to sustain it in existence as well. Indeed, for God there is no essential difference between creating and sustaining: 'If the world subsists, it is because God continues to will that the world exist. On the part of God, the conservation of the creatures is simply their continued creation' (1688: 316). When God conserves/recreates a body, he must recreate it in some particular place and in some relation of distance to other bodies. If God conserves it in the same relative place from moment to moment, it remains at rest; if God conserves it successively in different places, it is in motion. But this means that God is and can be the *only* cause of motion: 'The moving force of a body, then, is simply the efficacy of the volition of God who conserves it successively in different places…. Hence, bodies cannot move one another, and

their encounter or impact is only an occasional cause of the distribution of their motions' (1688: 162). And finite minds are no more causes of motion than bodies are.

Thus, God is the direct and efficacious cause of every event in nature; finite beings are only 'secondary' or 'occasional' causes. Malebranche's doctrine can be seen as embedded within a voluntarist tradition that extends from certain medieval thinkers - many of whom attacked the Aristotelian theory of nature in the name of safeguarding God's omnipotence - up through Descartes (see Voluntarism).

This does not mean that for Malebranche natural philosophy has been reduced to a single theocratic claim. At the level of physics proper, the task of the scientist is still to uncover regularities in nature and formulate the laws that govern the correlations between events. The programme of the mechanical philosophy, to which Malebranche enthusiastically subscribes, remains the same: to discover the hidden mechanisms that underlie observed phenomena and to frame such explanations (referring to secondary causes) solely in terms of matter and motion. What Malebranche's occasionalism does is to give an account of the metaphysical foundations of Cartesian physics. Motion, the primary explanatory element in the new science, must ultimately be grounded in something higher than the passive inert extension of Cartesian bodies; it needs a causal ground in an active power or force. Because a body consists in extension alone, motive force cannot be an inherent property of bodies. Malebranche accordingly - and his account seems to be a logical extension of the role Descartes gives to God as the 'universal and primary cause of motion' (see Descartes, R. §§6, 11) - places the locus of force in the will of God, and bodies behave the way they do because that is how God moves them around, following the laws of nature he has established.

Malebranche did, in fact, use his metaphysics of motion to modify some details of Descartes' physics, particularly the rules governing bodily impact. This led to an extended debate with Leibniz over the laws of motion. In the *Recherche* Malebranche insists, contrary to Descartes, that bodies at rest do not have a force to remain at rest (unlike bodies in motion, which have a force to remain in motion). This conclusion follows directly from his occasionalist account of motion: although God does need to will positively to put a body in motion, he does not need to apply any force to keep it at rest; all he needs to do is will that it continue to exist, that is, recreate it in the same relative place. Thus, the tiniest body in motion will contain more force than the largest body at rest. And this means that three of Descartes' seven rules of impact are wrong. Leibniz, in his general critique of Cartesian physics, praised Malebranche for recognizing Descartes' errors. He insisted, however, that because Malebranche was still wedded to Descartes' conservation law (where what is conserved is quantity of motion, rather than the quantity of motive force that Leibniz proposed), Malebranche has failed to see that *all* except the first of Descartes' rules, along with the new rules he had substituted for the ones he rejected, were wrong (see Leibniz, G.W. §11). In 1692, Malebranche published his *Des lois de la communication des mouvements*, in which he concedes that Leibniz is right about the rules themselves, but continues to maintain the old conservation law. It was not until a letter to Leibniz in 1699 and the 1700 edition of the *Recherche* (which contains a revised version of *Des lois*) that Malebranche admits that Descartes' conservation law is false.

5 Ethics

For Malebranche, God's ubiquitous causal activity does not eliminate freedom of the will in human beings. God is the direct source of an invincible inclination or 'natural motion' in the soul towards good in general. We cannot not will to be happy, and we necessarily love what we 'clearly know and vividly feel' to be good. But it is in our power to allow or to refuse to allow this general determination towards good given to us by God to rest upon one or another of the particular things we believe to be good. All minds, he notes, love God 'by the necessity of their nature', and if they love anything else, it is by a free choice of their will. We sin when, rather than directing the will by a clear and distinct perception of the supreme good to the love of God, we allow it to be directed away from God towards the pleasing but false goods presented to us by our senses.

In the *Traité de morale* (1683b), Malebranche elaborates on just what this love of God involves, providing a fuller account of our ethical duties. Within God, there lies an immutable order or law which God consults when acting. This order is constituted by what Malebranche calls 'relations of perfection', which in turn entail a hierarchy of value among beings. Order dictates, for example, that a beast is more perfect, hence more worthy or 'estimable' than a rock, and a human being more perfect and worthy than a beast. A human being is thus to be treated with more consideration than a horse. Through our 'union with the eternal word, with universal reason' we can have

rational knowledge of order. Our duty, then, consists in ‘submitting ourselves to God’s law, and in following order’. We ought, like God, to regulate our actions and esteem by consulting it. In this way, ‘there is a true and a false, a just and an unjust’ that is binding upon *all* intelligent beings. Malebranche insists that our principal duty and our ‘fundamental and universal virtue’ is to love order and obey its precepts.

6 Philosophical theology

Malebranche’s occasionalism takes on greater importance in the context of his theodicy, or justification of God’s ways in the realms of nature and grace. In his *Traité de la nature et de la grace* (1680), Malebranche undertakes the task of explaining how God’s omnipotence, benevolence and perfection can be reconciled with the persistence of evil and imperfections in the natural world (including human suffering and sin) and with the apparent unfairness and inefficiency in the distribution of divine grace and everlasting happiness.

Our concept of God tells us that God is infinitely wise, good, powerful and perfect. And yet the world which God has created certainly appears to us to be quite imperfect in its details and full of disorders of every variety. As Theodore, Malebranche’s spokesman in the *Entretiens sur la métaphysique*, exclaims, ‘The Universe then is the most perfect that God can make? But really! So many monsters, so many disorders, the great number of impious men - does all this contribute to the perfection of the universe?’ (1688: 211). Aristes, his interlocutor, is led thereby to wonder either about the efficacy of God’s will or the benevolence of God’s intentions.

The resolution of this conundrum, as presented in both the *Entretiens* and the *Traité*, is to be found in the consideration not just of the particular, superficial and obvious details of the universe, but also of the means undertaken to achieve and sustain the whole. God, according to Malebranche, looks not only to the final result of his creative act (that is, to the goodness and perfection of the world *per se*), but also to his work or ways of operation. And the activity or means most expressive of God’s nature are of maximum simplicity, uniformity, fecundity and universality. God does not accomplish by complex means that which can be accomplished by simple means; and God does not execute with many particular volitions that which can be executed by a few general volitions. This holds true even if it means that the world created by God could be spared some imperfections were God to compromise the simplicity and generality of his operations. Thus, the perfection of the world in its details as a product is completely relative to the mode of activity that is most worthy of God. God might increase the absolute perfection of the world, perhaps by decreasing the number of defects or evils therein. But this would entail greater complexity in the divine ways and constant departures from the general laws of nature established at creation.

Thus, the world that God has created is the one of the infinitely many possible worlds that best reconciles perfection of design with simplicity and generality of means of production and conservation. By a number of ‘particular volitions’ - that is, volitions that are *ad hoc* and not occasioned by some prior event in accordance with some law of nature - God could correct deformities of birth, keep fruit from rotting on trees, prevent physical disasters about to occur by the regular course of the laws of nature, and forestall sin and wickedness. But, Malebranche insists, ‘we must be careful not to require constant miracles from God, nor to attribute them to him at every moment’ (1680: 34). God, in other words, acts only by ‘general volitions’ - that is, volitions that are in accordance with some law and whose operation is occasioned by a prior event, as dictated by that law - and the most simple ways.

Similar considerations apply to the problem of grace. A benevolent God wills, with what Malebranche calls a ‘simple volition’, that sinners convert and that all humans should be saved. But clearly not all humans are saved; many are lost. And not all those who are saved appear to be worthy of salvation or ready to receive grace. The anomaly is again explained by the generality of God’s volitions. The distribution of grace is governed by certain general laws willed by God, and the occasional causes responsible for the actual distribution of grace in accordance with those laws are the thoughts and desires in the human soul of Jesus Christ. Because Jesus *qua* human has finite cognitive capacities, he cannot at any given time attend to all the relevant facts about the agent upon whom grace is to be bestowed - for example, whether they are ready to make the best use of it - or actually think of all who deserve to be saved. Thus, as with the distribution of evil and imperfection in the natural world, God *allows* grace to be distributed unevenly and even inequitably by the laws of grace in combination with the occasional causes that activate them.

It was the *Traité*, with its claim that God wills to save *all* humans and the implication that God's volitions are not always efficacious (since not everyone is saved), that initially aroused Arnauld's ire and occasioned his attack on Malebranche's whole system. Arnauld, as a Jansenist, was committed first and foremost to a strong doctrine of predestination and to the efficacy of divine volitions, particularly in the matter of grace. Pierre Bayle rallied to Malebranche's defence on this and other issues, and the pages of Bayle's *Nouvelles de la république des lettres* in the 1680s became an important battleground for the debates instigated by Malebranche's works.

7 Influence

Malebranche's influence in the seventeenth and eighteenth centuries was significant, but subtle and often unacknowledged. There is no question that his contemporaries recognized him as *the* major representative of the Cartesian system, however unorthodox his Cartesianism may have been. Leibniz's arguments against 'the Cartesians', for example, are often directed at Malebranche. And yet, despite his criticisms of occasionalism, Leibniz was himself impressed by Malebranche's discussion of causation and critique of interaction between substances. Moreover, Leibniz's own theodicy and solution to the problem of evil were clearly influenced by what he read in Malebranche. Like Malebranche, Leibniz insists that God in creation chooses from an infinity of possible worlds, and that God pays particular attention not just to the created theatre itself, but especially to its relationship with the laws of nature and grace. Malebranche considers the laws as separate from 'the world', and gives them a higher value. He grants that the world God created may not be, absolutely speaking, the best of all possible worlds, but it is the best that can be done given the absolute simplicity of means God employs. Leibniz, on the other hand, considers the laws and the universe they govern together as 'the world', and insists that the combination *is* the best world overall. But they agree that evil and sin occur because God *allows* them to occur as a result of the ordinary course of nature as governed by the laws God has chosen. They agree that God could diminish the imperfections of the created world, but only by violating the simplicity of the divine ways (as Malebranche would put it), or by detracting from the *overall* optimality of the world (as Leibniz would say). Leibniz even goes so far as to suggest to Malebranche that, in the end, their accounts are the same, although Malebranche disagrees.

Malebranche's influence extended across the Channel (and not just to such overt followers as John Norris). Despite Berkeley's indignant claim that 'there are no principles more fundamentally opposed than his [Malebranche's] and mine', there are obvious echoes of Malebranche's doctrines in Berkeley's works (see Berkeley, G. §2). For example, Berkeley's ideas, like Malebranche's ideas, are not modifications of the finite mind but are independent of it. They are also in the mind of God, since it is 'an infinite spirit who contains and supports' the world of ideas. Similarly, with respect to causation, Berkeley denies that our ideas of bodies provide us with any notion of causal power or efficacy. And while Berkeley departs from Malebranche's doctrine and grants real causal power to the human soul (and thus is not a complete occasionalist), he insists, like Malebranche, that the ordinary course of natural phenomena, the regularities and correspondences in our ideas of things, are the result of the causal activity of the will of a governing spirit, that is, God.

Hume was more forthcoming in acknowledging his debt to Malebranche in his conclusions about causality. His arguments denying that our idea of body affords us any notion of causal power and his insistence that all that experience reveals is a constant conjunction between events seem to come right out of Malebranche's *Recherche*. Both Hume and Malebranche stress the centrality of the concept of necessary connection to our understanding of causation, and both deny that such necessity can be discovered (by reason or experience) between any things in nature. The difference is that Malebranche held that we can perceive a necessary connection between the will of God and any event willed by God, while Hume rejected any such claim.

STEVEN NADLER

List of works

Malebranche, N. (1958-67) *Oeuvres complètes de Malebranche (Complete works of Malebranche)*, ed. A. Robinet, Paris: J. Vrin, 20 vols.(The standard edition of Malebranche's writings. The main philosophical works are in the volumes listed below.)

Malebranche, N. (1674-5) *De la recherche de la vérité*, in *Oeuvres*, vols 1-3, 6th edn, 1712; trans. T. Lennon and P.J. Olscamp as *The Search After Truth/Elucidations of the Search After Truth*, Columbus, OH: Ohio State University Press, 1980. (Malebranche's most important philosophical work. Vol. 3 of the *Oeuvres* contains the

Éclaircissements.)
Malebranche, N. (1676) *Conversations chrétiennes (Christian conversations)*, in *Oeuvres*, vol. 4.(A 'justification' of the Catholic religion and morality.)
Malebranche, N. (1680) *Traité de la nature et de la grace*, in *Oeuvres*, vol. 5; trans. P. Riley as *Treatise on Nature and Grace*, Oxford: Oxford University Press, 1992. (Malebranche's theodicy and account of human freedom.)
Malebranche, N. (1683a) *Méditations chrétiennes et métaphysiques (Christian and metaphysical meditations)*, in *Oeuvres*, vol. 10.(A reworking of his metaphysical and epistemological doctrines including the vision of God.)
Malebranche, N. (1683b) *Traité de morale (Treatise on morality)*, in *Oeuvres*, vol. 11.(A treatise in Christian ethics and human happiness.)
Malebranche, N. (1684-94) *Recueil de toutes les réponses à M. Arnauld (Collection of all the responses to M. Arnauld)*, in *Oeuvres*, vols 6-9.(Contains correspondence related to the Malebranche-Arnauld debate, as well as Malebranche's *Réponse* to Arnauld's *Des vraies et des fausses idées*, and some other pieces.)
Malebranche, N. (1688) *Entretiens sur la métaphysique et sur la religion*, in *Oeuvres*, vol. 12; trans. W. Doney as *Dialogues on Metaphysics*, New York: Abaris, 1980.(A presentation, in dialogue form, of Malebranche's doctrines *c.*1688, after some development since the *Recherche* in the light of various debates.)
Malebranche, N. (1692) *Traité des lois de la communication du mouvement*, in *Oeuvres*, vol. 2. (A critique of Descartes' account of the rule of impact.)
Malebranche, N. (1696) *Entretiens sur la mort*, in *Oeuvres*, vol. 13.(A continuation of the dialogues on metaphysics concentrating on death and the soul.)
Malebranche, N. (1708) *Entretien d'un philosophe chrétien et d'un philosophe chinois (Dialogue between a Christian philosopher and a Chinese philosopher)*, in *Oeuvres*, vol. 15. (Christian apologetics.)
Malebranche, N. (1715) *Réflexions sur la prémotion physique (Reflections on physical premotion)*, in *Oeuvres*, vol. 16.(Malebranche's response to Abbé Boursier on God's action on creatures.)

References and further reading

Alquié, F. (1974) *Le Cartésianisme de Malebranche (The Cartesianism of Malebranche)*, Paris: J. Vrin.(An examination of what Malebranche inherits from orthodox Cartesianism and what he modifies or rejects.)
André, Y. (1886) *La vie du R. P. Malebranche, prêtre de l'Oratoire, avec l'histoire de ses ouvrages (The life of Father Malebranche, priest of the oratory, with the history of his works)*, Paris; Geneva: Slatkine Reprints, 1970.(Written in the first half of the eighteenth century; still the standard source on Malebranche's life.)
Arnauld, A. (1683) *Des vraies et des fausses idées*; trans. E. Kremer as *On True and False Ideas*, Lewiston, NY: Edwin Mellon Press, 1990; also trans. S. Gaukroger, Manchester: Manchester University Press, 1990. (Arnauld's attack on Malebranche's theory of ideas and doctrine of the vision in God.)
Brown, S. (ed.) (1991) *Nicolas Malebranche: His Philosophical Critics and Successors*, Assen: Van Gorcum.(A collection of essays on Malebranche, his influence, and his contemporary and later critics.)
Church, R.W. (1931) *A Study in the Philosophy of Malebranche*, London: Allen & Unwin.(A valuable general study of Malebranche's philosophy.)
Dreyfus, G. (1958) *La volonté selon Malebranche (The will according to Malebranche)*, Paris: J. Vrin.(A thorough account of Malebranche on the divine will and the human will.)
Easton, P., Lennon, T. and Sebba, G. (1991) *Bibliographia Malebranchiana: A Critical Guide to the Malebranche Literature 1638-1988*, Edwardsville, IL: Southern Illinois University Press.(A comprehensive bibliography covering both Malebranche and his friends and critics.)
Gouhier, H. (1948) *La philosophie de Malebranche et son expérience religieuse (The philosophy of Malebranche and his religious experience)*, Paris: J. Vrin.(An examination of some essential Malebranchian doctrines and their theological aspects.)
Gueroult, M. (1955) *Malebranche*, Paris: Aubier, 3 vols.(An exhaustive and masterly study of Malebranche's system.)
Jolley, N. (1990) *The Light of the Soul: Theories of Ideas in Leibniz, Malebranche, and Descartes*, Oxford: Clarendon Press.(Examines Malebranche's theory of ideas, especially in comparison to Descartes' and Leibniz's accounts.)
McCracken, C. (1983) *Malebranche and British Philosophy*, Oxford: Clarendon Press.(Looks at the way Malebranche's philosophy was variously assimilated and criticized in early modern British philosophy; chapters on Locke, Berkeley and Hume.)
Nadler, S. (1992) *Malebranche and Ideas*, New York: Oxford University Press.(An analysis of Malebranche's theory of ideas and doctrine of the vision in God.)

Prost, J. (1907) *Essai sur l'atomisme et l'occasionalisme dans la philosophie cartésienne (Essay on atomism and occasionalism in Cartesian philosophy)*, Paris: Henry Paulin.(A fine study of occasionalism and its roots in Descartes' philosophy, with several chapters on Malebranche.)

Radner, D. (1978) *Malebranche: A Study of a Cartesian System*, Assen: Van Gorcum. (A useful general and analytical presentation of Malebranche's doctrines.)

Robinet, A. (1955) *Malebranche et Leibniz: relations personelles (Malebranche and Leibniz: personal relations)*, Paris: J. Vrin.(A valuable resource that contains correspondence and other documents concerning the mutual relations between Malebranche and Leibniz.)

Robinet, A. (1965) *Système et existence dans l'oeuvre de Malebranche (System and existence in the work of Malebranche)*, Paris: J. Vrin.(One of the most important studies of Malebranche's philosophy, with emphasis on its development.)

Rodis-Lewis, G. (1963) *Nicolas Malebranche*, Paris: Presses Universitaires de France.(A general account of Malebranche's life and thought.)

Walton, C. (1972) *De la recherche du bien: A Study of Malebranche's Science of Ethics*, The Hague: Martinus Nijhoff.(An analysis of Malebranche's moral philosophy, concentrating on his account of virtue and duty.)

Mamardashvili, Merab Konstantinovich (1930-90)

Merab Mamardashvili was one of the Soviet Union's most influential thinkers in the fields of phenomenology and philosophy of consciousness. Although he preferred the Socratic genres of the dialogue, interview and philosophical meditation to the abstract rigours of more systematic philosophy, he left substantial published work on Descartes, Hegel, Kant and French literature (especially Proust).

*Mamardashvili began his career as a historian of philosophy, with a series of close readings of Karl Marx. By the 1970s he had evolved his own distinctive style of 'philosophizing out loud', addressing the foundations of European philosophy based on Descartes and Kant, at the core of which was the search for the 'free phenomenon' (*svobodnyi fenomen*) or the 'event of a thought' (*sobytie mysli*). In Kantian fashion, Mamardashvili attended to those a priori conditions of lived experience which govern that moment when reality enters the transcendental realm - but he switched the emphasis: rather than the mental problems presented by the a priori moment, Mamardashvili concentrated on what he called a 'metaphysics of the a posteriori', that is, on the actual event, or advent, of a thought. Perhaps the single motivating question of his life was: 'How is a new thought possible?' Among his many answers, developed in public lectures and interviews during the last twenty years of his life, was the notion that the very processes of thought provoke 'hearing a thought' in another. From this follows his concern with dialogic forms and his interest in the Cartesian dualism of soul and body - not as a necessary truth but as a 'productive tautology' that makes internal reason and a 'grammatical' analysis of thinking possible on a palpable basis.*

1 Life

Mamardashvili was born in Gori, Soviet Georgia (the birthplace of Stalin), to a military family. After residence in Leningrad and Kiev, the family resettled in Tbilisi in 1941, where he finished high school; he then entered the Department of Philosophy at Moscow State University, graduating in 1954 and receiving his candidate's degree (Ph.D.) in 1957. Always more interested in the nature of human consciousness as *experience* rather than as system or theory, Mamardashvili did not immediately attach to an academic institution. After working as editor for the journal *Voprosy filosofii (Questions of Philosophy)*, he spent 1961-6 in Prague as editor of *Problems of Peace and Socialism*; upon his return to Moscow he lost his rights to travel (managing to visit his beloved France only in 1988). At this time Mamardashvili began delivering lectures (or *besedy*, 'conversations', as he preferred to call them) in various institutes to students of philosophy, psychology, pedagogy, theatre and cinematography. In 1970 he defended his doctoral dissertation in Tbilisi; he worked in that city and in Moscow until his sudden death at the age of 60.

Disciples of his thought are very active in post-communist Russia, especially in the Laboratory for Post-Classical Studies (of the Institute of Philosophy, Russian Academy of Sciences), pursuing questions of symbolic consciousness, history of science and theory of creativity. Among his most prominent students are Aleksandr Piatigorskii, Mikhail Ryklin, Anatolii Akhutin andMamardashvili's archivist Iurii Senokosov.

2 Critical works

Mamardashvili's genre was the interview, the lecture series, the roundtable discussion, which enabled what he called '*filosofstvovanie vslukh*', 'philosophizing out loud'. In his meditations on Kant, Mamardashvili defined *filosofstvovanie* as a 'sensual and palpable turn of the soul, a feeling for ideas', which was always prompted by an 'element of life, not of theory' ([1982b] 1991: 150). As one of Mamardashvili's most prolific students, Mikhail Ryklin, has remarked, oral speech for Mamardashvili was like improvisation for a jazz musician: to emerge at all, his most creative ideas needed a live listener present. Mamardashvili did not write 'for the drawer' because, in his words, he 'utterly lacked any consciousness of persecution' and preferred talking to writing (1990a: 178).

3 'Philosophizing' versus 'Russian' philosophy

Mamardashvili did not believe that one could be a 'national philosopher'. He was no partisan of the Russian Religious-Philosophical Renaissance (in this he resembles Pëtr Chaadaev rather than Nikolai Berdiaev or Vladimir Solov'ëv); in his lectures on Proust, he remarked that 'one of the ideas that had ruined Russian culture at the beginning of the century was the theurgical idea, the notion that the world had been created so that life itself could

be constructed like a work of art' (Kruglikov 1994: 111). He charged the radical intelligentsia with '"dirty", littered consciousness and slovenly thinking' (1990a: 132); and even the *Vekhi (Signposts)* opposition, in his opinion, was too partial to endless acts of repentance. Mamardashvili's attitude towards religion was complex. In his 'Besedy o myshlenii' (Conversations about Thinking) (1986-7), discussing the sophisticated forms of French secular and sacred language, he remarked that 'the language of religion is necessary in order to distinguish a person striving towards the good from a [simply] good person' - a distinction that arrives very late in 'infantile cultures' such as the Russian, and when it does (Mamardashvili says, agreeing with Dostoevskii), it goes unnoticed, causes anguish and often generates evil ([1986-7] 1991: 25).

In general we do not think enough about *how we think*, nor about the consequences of an act of thought. Russia was especially deficient in this talent, eager to use words but unwilling to accept results; thus it has been in the 'hellish condition' of a country of 'eternal pregnancy' ([1987] 1990: 372). Descartes and Pushkin were exemplarily honest thinkers in this sense, Mamardashvili believed, as men who became free by accepting responsibility for control over their own words and gestures (Kruglikov 1994: 20-3); Dostoevskii was less so, for, in some of his preconceptions - his odd assumption, for example, that to be poor was *ipso facto* to be simple and honest - he challenged the sensible workings of the mind ([1986-7] 1991: 26). Ideas themselves, in the structure of consciousness, are so 'complex, delicate and multi-layered' that we should 'tremble from fear before the responsibility' of handling them (1990a: 130). As Ryklin sums up his teacher's position: on the one hand, 'in his work, all of philosophy becomes ethics'; on the other, 'ethics itself is presented exclusively in the form of a theory of cognition' (*Voprosy filosofii* (1991) 5: 21).

Mamardashvili did not believe that ideas were ever produced by the 'masses'. For culture is assimilated individually and cannot be collectivized or generalized. By 'culture', Mamardashvili understood not a set of guidelines or fixed values, but rather a 'definitiveness of *form* within which people were capable (and prepared), through their acts, to practise *complexity*' (1990a: 173). Accordingly, Mamardashvili's recurring concerns were individual autonomy; the necessity to keep separate the empirical and the non-empirical worlds (Mamardashvili resisted all utopian and apocalyptic models of history); a more modest societal role for art and the literary word ('art for art's sake' is more democratic, he felt, than prophetic or socially committed art (1990a: 127)); how to achieve 'civilization', the major binding force or grammar uniting highly independent and particularized minds; and, underlying all these concerns, the mechanism of thinking itself.

Mamardashvili's major questions and modes of thought nevertheless reflect the Russian philosophical tradition in three ways. First, like his heroes Descartes and Tolstoi, he directs his insights on consciousness and morality not to academic philosophers but to untrained minds seeking clarification in everyday language. Second, he draws on great literature as primary material for philosophical speculation - and indeed, he sees no firm boundary between art and this type of 'science'. And third, his subject matter is grounded not in formal logic or abstract systems but rather evolved under the inspiration of several exemplary creative figures in European culture: Hegel, Kant, the early Marx, Descartes, Husserl, Proust, Kafka and Mandelstam.

Mamardashvili devoted dozens of 'conversations' to the meaning of philosophy itself. It is not a science, he claimed, nor is it a picture of the world. In an interview near the end of his life (1990), published under the title 'Solitude is my Profession', Mamardashvili defines the philosophical moment as 'a shard from the broken mirror of universal harmony, fallen into the eye or the soul', 'an allegory for the passion of freedom' (Kruglikov 1994: 64). Elsewhere Mamardashvili explains that such freedom is in fact a necessity: for only the rational epiphany of a newly coalesced thought can reverse natural law and move from death to life (otherwise, the cosmos would have long ago contained only inert matter). But this epiphany is not religious; religion elicits from us a worshipful response and we can suspend its acts passively, indefinitely, within a system of faith. Philosophy, on the other hand, is an interval, a 'pause that is essential to the structure of the life of consciousness, essential to human acts'.

Thus philosophy is a place - or an attitude - that we occupy when we have the courage to say 'I do not understand'. It does not judge content (good and evil, right or wrong), and it does not 'fill up'; like moral conscience, it is best defined as a form and a dynamics, an inner reflex that, once stimulated, will *not turn away*; and thus it exists eternally and its problems repeat forever. (Following Socrates, Mamardashvili insists that to exist, philosophy - like virtue - must be accomplished again and again; whereas evil can register its effects by happening only once.) We philosophize when our consciousness is exacerbated by a sudden break or discontinuity between itself and its

surroundings; in order to mend that break we must commit to an effort to think honestly or honourably (*myslit' chestno*). This is extremely difficult to do; and for this reason, Mamardashvili considered Descartes and Kant among the greatest European philosophers.

When Mamardashvili delivered his *Kartezianskie razmyshleniia (Cartesian Meditations)* in Moscow in the winter of 1981, it marked, according to one eyewitness, the rebirth of professional philosophical thought in the Soviet Union. Through the medium of Descartes, these fifteen meditations reintroduced the themes of faith, the possibility of God within a rationalist universe, the 'miracle of individual thought', and the centrality of epistemology to any primary philosophy. In his opening lectures, Mamardashvili declared that 'for thought, the worst enemy is the *past*'; knowing that every mental image becomes more concrete, precise and manageable the less it is actually present, Descartes filled his own 'Meditations' and 'method' with the word 'now'. This was welcome balm to his Soviet audience, weary of Marxist-Leninist surrogates for the present tense. In his final lecture, Mamardashvili gently chided both the Russian tradition of the radical intelligentsia and the revived clichés of Russian religious thought through the sobering lens of Cartesian thinking - with its methodical doubt, its separation of body and spirit, its assumption (and here Mamardashvili also drew on Proust) that too many 'serious conversations' were a 'waste of time' and better worked out in solitude. On Christ's crucifixion, Mamardashvili remarked that

> 'the cross is an ironic reminder to us that we need not herd people into an image, inasmuch as this image, within our heads, is capable of killing. And there are cultures and even whole epochs when people love the dead most of all, because the dead can no longer exit from their own image and thus they speak only that which we speak ourselves'
>
> ([1991] 1993: 341)

4 Contribution to the performing arts and practical aesthetics

This fascination with the rights and possible reconstitution of a genuine *present* lies behind Mamardashvili's interest in the theatre and visual temporal art (he was a frequent lecturer at theatre training institutes); and, as he wrote in *Teatr* magazine in 1989, 'the theatrical stage is a place for the "presence of absent reality"' (1990a: 138). It also permeates his 1982-4 lectures on Proust. Among the major themes in those lectures are the 'indivisibility of presence' (Kruglikov 1994: 114), the nature and degree of energy required to awaken ourselves from habit, and the phenomena of 'aroused attention' and memory. How, indeed, does a 'thought' or a 'desire' start? What does it mean to actualize a part of the past towards this end?

'Spontaneous memory' occurs in us in response to an apparently arbitrary summons. But, Mamardashvili asks, can we *command* ourselves to experience something strongly, to love, to be aroused, to be upset or to be riveted in place? The Proustian moment - being triggered by a present triviality into re-experiencing a profound past reality - Mamardashvili calls the 'zone of helplessness' (Kruglikov 1994: 94). It is a finicky and subtle state. If we try to force it, then the memory-experience is almost always negative: it feels empty, and we feel at a loss. For any mental gesture that does not provide the potential for a free supplement from the present, that is only a forced calling-up of the past, is a void that will be filled with illusion. Mamardashvili interprets the folksaying '*Sviato mesto pusto ne byvaet*' ('A sacred space is never empty') to mean that evil always settles in emptiness; the Good demands fullness, that is, a responsive act from us in return. When this condition is not met, we have negative memory: that which summons us in the absence of an active and sentient present. Mamardashvili concludes that for Proust (and for him), our life is real to the extent that a precious event in it is *created anew by thought* (Kruglikov 1994: 97). Ending one of his lectures on 'What is Philosophy?' with a discussion of Proust's novel, Mamardashvili noted:

> (The text of the novel is the means by which feelings can be experienced.) Not those that *you* experience…but you-as-an-other, recreated, fixed in time and space, posited by a movement of decoding, which does not depend on the chance happenings of life and in which you can establish yourself as a person of fate who controls his own fate. For you there will be no other fate. But that which will be, you already control.
>
> (*Voprosii filosofii* (1991) 5: 10)

Certain general statements can now be made. In Proust - and in philosophy in general - Mamardashvili investigated 'sudden beginnings', or what he called 'surprise': those experiences, thoughts and events that seem to arise 'out of nothing'. Unlike Russian apocalyptic and messianic thinkers (who studied collective ends), and unlike

his senior colleague in the philosophy of consciousness, Bakhtin (who exclusively studied middles), Mamardashvili made a point of studying the miraculous surprise of *individualized beginnings*. Culture itself is defined as that set of specific forms within which individuals could realize increasing complexity.

All true philosophizing has a phenomenological aspect, Mamardashvili taught, although the impulse for his own questioning came not from Husserl but from Marx (in his own words, he moved from Marx to an 'intuition of "bodies", that is, to "structures of palpable agency"' in the world, and then to the 'concrete object-ness of a thought' as the 'living, extra-mental reality of the soul' (1990a: 102)). Mamardashvili was not a political philosopher, nor a social one. He was, rather, a personalist and intentionalist of the sort that has special resonance in a culture like Soviet Russia, so familiar with political abuse and collective utopia: while greatly admiring Kant and considering him not part of that 'epoch of strictly German philosophy, so repulsive to me' ([1982b] 1991: 121), Mamardashvili distrusted the power that moral norms held over an individual and worried about the proper procedures for stitching ourselves into them ('In a broad sense', Mamardashvili writes, 'for Kant, a norm is a grimace, a going to church… Grimaces of the spirit' ([1982b] 1991: 145)). Individual will is integrated into norms, he believed, but all social and spiritual creativity is dependent upon on the participation of 'a sufficient number of adult, non-infantile people who rely on their own individual minds' ([1987] 1990: 374). In a lengthy excursus on memory and its role in the poet Mandelstam's famed '*toska po mirovoi kul'ture*' (longing for world culture), Mamardashvili remarked, again in the spirit of Chaadaev, that Russians have never achieved a secure sense of continuity, of belonging, of having authentic organs of birth ([1986-7] 1991: 40-4); and for this reason, perhaps, Russians are so familiar with the terror at possible non-being which fuels philosophy.

In the late 1980s, Mamardashvili elaborated his 'Principle of the Three Ks': *Karteziia* (Cartesianism, Descartes), Kant and Kafka ([1984b] 1990: 109-12). The Cartesian 'I' asserts its own first-person-singular existence and doubts every other thing. The Kantian 'I' is subordinated to a third-person transcendental point of view; its strength lies in its attention to the real-world conditions and coordinates that shape an act. The Kafkaesque 'I' knows neither first nor third person; it is a zombie, a humanoid without agency functioning in the realm of the absurd. Curiously, Mamardashvili suggested that it was only in the *glasnost'* period - for he did not live to see the end of Soviet communism - that the third, Kafkaesque 'I' was in fact recognized as such on Russian soil: before that time, absurdity was blamed on politics, on Stalinism, and thus explained away not as a spiritual option (which it is) but as a brutal imposition by politics. In laying out his taxonomy of the three types of self, Mamardashvili reveals himself to be an unsentimental phenomenologist in the Stoic vein. 'A philosopher is not surprised at evil, destruction, chaos or disorder, although usually we, people, are surprised at that…. What is surprising and remarkable is that there is good…. Just a tiny island of order in the chaos - that's remarkable' ([1979] 1992: 31).

CARYL EMERSON

List of works

Mamardashvili, M.K. (1968) *Formy i soderzhanie myshleniia (The Forms and Content of Thinking)*, Moscow. (Survey of phenomenological and Kantian paradigms for the production of thought; academic in approach.)

Mamardashvili, M.K. (1979) 'O poniatii filosofii' (On the Concept of Philosophy), lecture at Rostov State University, in *Novyi Krug (Kiev)* 1992: 1.(Mamardashvili's discussion of 'two categories of things': those that happen, and those that do not happen, 'by themselves'; the notion of freedom as that which happens 'without choice'; the difference between philosophy and culture.)

Mamardashvili, M.K. (1982a) *Simvol i soznanie (metafizicheskie rassuzhdeniia o soznanii, simvolike i iazyke) (Symbol and Consciousness: Metaphysical Speculations on Consciousness, Symbolic Structures and Language)*, with A. Piatigorskii, Jerusalem.(Non-technical discussion of the structure of consciousness, the sign, the symbol, the natural versus the artificed in language, as part of a search for the epistemological basis for semiotics.)

Mamardashvili, M.K. (1982b) *Kantianskie variiatsii (Kantian Variations)*, lectures delivered in Moscow; fragments published in *Kvintessentsiia: Filosofskii almanakh 1991*, Moscow, 120-57.(An examination of Kant's sentiment that 'a soul overfilled with feelings, and not speech, is the highest of all perfections'; the role of words for Rousseau and of meaning for Kant.)

Mamardashvili, M.K. (1984a) *Klassicheskii i neklassicheskii idealy ratsional'nosti (Classic and Non-Classic Ideals of Rationality)*, Tbilisi. (Investigations of a relatively non-technical nature into the problem of observation, the notion of a phenomenon and its multiple dimensions; and the symbolic element in all rational thought.)

Mamardashvili, M.K. (1984b) 'Soznanie i tsivilizatsii' (Consciousness and Civilization), lecture delivered in Batumi, published in *Kak ya ponimayu filosofiiu (How I Understand Philosophy)*, 1990, 107-21. (Includes Mamardashvili's 'three Ks' as ways of organizing the world: Decartes, Kant and Kafka.)

Mamardashvili, M.K. (1986-7) 'Besedy o myshlenii' (Conversations about Thinking), lecture series delivered at Tbilisi University, in V.A. Kruglikov (ed.) *'Mysl' izrechennaia… ': sbornik nauchnykh statei*, Moscow, 1991, 13-50.(Meditations on the unexpected origins of a thought; on the difficulty of willing a desire versus willing an insight; the responsibilities that accrue to thinking.)

Mamardashvili, M.K. (1987) 'Filosofiia i svoboda' (Philosophy and Freedom) (lecture), in *Kak ya ponimayu filosofiiu (How I Understand Philosophy),* Moscow 1990, 365-74.(The Good as a productive state, not as an attribute or a product, because freedom can only produce more of itself; Russia as a land of 'eternal pregnancy'.)

Mamardashvili, M.K. (1990a) *Kak ya ponimayu filosofiiu (How I Understand Philosophy)* Moscow.(A collection of sixteen seminal lectures and interviews, and fifteen formal articles, from the 1970s to 1990s.)

Mamardachvili [Mamardashvili] (1990b) *'La pensée empêchée (Entretien avec Annie Epelboin)'*, Editions de l'Aube.(Biographical material.)

Mamardachvili [Mamardashvili] (1991) *Kartezianskie razmyshleniia (Cartesian Meditations)*, lectures delivered in Moscow, expanded and published in book form, Moscow, 1993.(On Descartes the man and his masks; the desirability of thinking in the present; thought and epiphany; the responsibilities of becoming conscious.)

References and further reading

Kruglikov, V. (1994) *Kongenial'nost' mysli: o filosofe Merabe Mamardashvili (The Congeniality of Thought: On the Philosopher Merab Mamardashvili)*, Moscow.(A posthumous collection in three parts, containing primary material (an essay and interview by Mamardashvili), appreciations by fellow philosophers, biographical information and summaries of his thought.)

Murchland, B. (1991) 'The Mind of Mamardashvili', Ohio: Kettering Foundation.(Monograph prepared after several days of conversation with Mamardashvili during the last year of his life.)

'Pamiati filosofa' (In Commemoration of the Philosopher), special forum devoted to Mamardashvili in *Voprosy filosofii* (1991) 5: 3-25.(Memoiristic and eulogistic articles by seven scholar-disciples, including Iu. Senokosov, M. Ryklin, A. Piatigorskii and A. Akhutin.)

Mandeville, Bernard (1670-1733)

Bernard Mandeville's Fable of the Bees (1714) scandalized contemporaries by arguing that the flourishing commercial society they valued depended on vices they denounced. It resulted not only from the complementary satisfaction of appetites but was also based upon pride, envy and shame, which Mandeville traced to 'self-liking'. Numerous individuals, driven by their own desires, acted independently to produce goods which required extensive, cooperative operations - an idea central to the economic concept of a market.

Mandeville initially appeared to credit 'skilful politicians' with originating morality and society. However, in defending and expounding his views, he set out 'conjectural histories' of the gradual development of many complex social activities and institutions, including language and society itself, thereby denying that they had been invented by public spirited heroes. Throughout his works, Mandeville adopted a strict criterion of virtue, repeatedly denying that he was advocating, rather than exposing, the vices he identified as inherent in human society.

1 Early writings

Bernard Mandeville studied philosophy and medicine at Leiden, qualifying in 1691. In the mid-1690s he settled in England, specializing in nervous and digestive disorders. His early verse works included *Some Fables after the Easie and Familiar Method of Monsieur de la Fontaine* (1703; enlarged into *Aesop Dress'd*, 1704), two tales from which were by Mandeville himself. In 1705 he published a longer original verse fable, *The Grumbling Hive: or, Knaves Turn'd Honest*, whose bees continually complain of the vices and faults of their society, despite its power and prosperity. Granted virtue and honesty by Jove, they decline into contented, impoverished simplicity.

Mandeville then turned to female dialogues. In *The Virgin Unmask'd* (1709) 'an elderly maiden lady' advises her niece against the servility of marriage and lectures about the danger to Europe posed by Louis XIV. In the period 1709-10, Mandeville wrote thirty-two 'Lucinda' and 'Artesia' papers for the *Female Tatler*, an imitator of Richard Steele's *Tatler*. Mandeville ridiculed Steele's encomiums on the dignity of human nature and his advocacy of quasi-aristocratic 'politeness'; his 'Oxford gentleman' maintained that both sociability and prosperity were based on vice - on pride, flattery and the invention and satisfaction of luxurious desires. His 'sisters' debated honour, the advisability of military service in Marlborough's wars and the possibility of living to make money. Eight papers contended that women are fully capable of all the traditional male virtues, giving numerous historical, biblical and mythical examples (Goldsmith 1985).

2 Human nature and morality

The Fable of the Bees: or, Private Vices, Publick Benefits (1714), Mandeville's best-known work, reprinted *The Grumbling Hive*, adding extensive remarks and essays which analysed human conduct, finding it motivated by avarice, prodigality, envy, pride, honour, shame and luxury - with contrary passions balancing each other. 'Skilful politicians', moralists and lawgivers had socialized human beings by persuading them to restrain their immediate self-interested desires in return for the praise accorded to their noble self-sacrifice, thereby gratifying their desires to think well of themselves and be admired by others.

Much of this account of human nature was derived from the French *moralistes* and Pierre Bayle (see Bayle, P. §1; Mandeville 1714; Horne 1978; Hundert 1994), who had also pointed out the worldly benefits of unregenerate human conduct. Defining virtue as requiring self-denial for the good of others or one's own rational improvement enabled Mandeville to claim that he condemned the vicious passions to which he traced all human conduct, including apparently religious or virtuous conduct. Contemporary critics (for example, William Law and George Berkeley) accused Mandeville of advocating vice and subverting the standards of right conduct; Adam Smith, in his *Theory of Moral Sentiments* (1759, part VII, sect. ii, ch. 4), calls Mandeville's a 'licentious system' because it removes the distinction between vice and virtue. Kaye (Mandeville 1714) argued that Mandeville tended towards a consequentialist acceptance of the benefits of vice thereby implying a *reductio* of his professed strict (or 'rigourist') standard of virtue. None the less some commentators have accepted Mandeville's claim, reiterated in his last work *A Letter to Dion* (1732), that Dion's book *Alciphron* so misrepresented the *Fable of the Bees* that Berkeley could not have read that work (Monro 1975). Mandeville's satiric analysis of human nature along with

his scorn for reformers and standards of 'politeness' and virtue have puzzled subsequent commentators as well as his contemporaries (Castiglione 1986).

Originally aimed at Steele (in the *Female Tatler* papers and the first edition of the *Fable*) and at the proselytizing Societies for the Reformation of Manners, 'A Search into the Nature of Society', added to the second edition of the *Fable* (1723), extended the attack to the Third Earl of Shaftesbury. Also added to that edition was the 'Essay on Charity and Charity-Schools', which resulted in the *Fable*'s notoriety. In it Mandeville distinguished true charity from natural pity or compassion, discovered pride behind magnificent charitable bequests, asserted that charity schools were promoted by parish busybodies to confirm their sense of self-importance, and denied that the schools were socially beneficial in decreasing irreligion and crime, claiming rather that they were harmful in providing poor children with education and aspirations incompatible with the life of drudgery required of them by a prosperous commercial society. This provoked the *Fable*'s being presented as a public nuisance by a Middlesex Grand Jury and incited numerous refutations, most notably William Law and Francis Hutcheson (see Hutcheson, F. §§2-3).

It seems likely that a materialist and naturalist theory underlay Mandeville's ridicule of the dignity of human nature. The *Fable* treats humans as similar to other mammals in their anatomy, feelings, motives and behaviour - a rejection of the Cartesianism Mandeville learned at university. Its discussion of the slaughter and eating of higher animals and the debate between a lion and a man about superiority support that hypothesis, as do the methodological views expressed in Mandeville's *Treatise of the Hypocondriack and Hysterick Passions* (1711, 1730). This *Treatise* is a set of medical dialogues in which the doctor (clearly Mandeville) talks his patients into cures based on simpler diets, more exercise and fewer, milder medicines; he advocates the empirical study of diseases, sceptically rejecting fashionable and fanciful medical theories.

3 Economics

The paradox of private vices creating public benefits was illustrated by showing how natural and moral evils, such as disasters like the Great Fire of London, crimes like theft, vices like prostitution and harmful activities like the liquor trade, benefited society by providing work, circulating wealth, preserving chastity and, through taxes, by supporting the state. Thus Mandeville extended his 'selfish' conception of human nature, which has reminded commentators of Hobbes (see Hobbes, T. §6), into the thesis that human needs and vices produce social wellbeing. By serving the desires of others in order to satisfy their own desires, individuals acting independently produce complex social processes and institutions (including brewing, baking and, in 1723, the worldwide activity necessary to provide the luxury of scarlet coats for common soldiers). Hayek (1966) therefore credited Mandeville with the basic insight involved in the theory of the free market; namely, the social importance of the unintended consequences of individual actions. But claims that Mandeville was a conscious precursor of classical economics are exaggerated (Goldsmith 1985). Despite some passages advocating free trade, rejecting sumptuary laws, defending luxury, accepting the export of specie (gold and silver) and insisting that a nation must buy foreign goods if it expects to sell its goods abroad, Mandeville also states typically mercantilist views without expounding a clear economic theory (Horne 1978). Government must maintain a system of justice, provide defence and also regulate internal and external trade. Only *A Modest Defence of Public Stews* (1724), usually attributed (without definite proof) to Mandeville, provides a clear example of a self-regulating system of supply and demand - a scheme for a system of public houses of prostitution.

4 Later writings

In 1720, before the *Fable*'s notoriety, Mandeville published the moderate, whiggish *Free Thoughts on Religion*. Relying on Bayle to the point of plagiarism, Mandeville sceptically advocated irenic tolerance while expressing anticlerical views (also found in the *Fable*). *Free Thoughts* supported the Protestant succession, the Hanoverian dynasty and the British mixed constitution.

In two later works, signed 'by the author of the *Fable of the Bees*', Mandeville supplemented and defended his views. Both were dialogues between the same principal interlocutors: Cleomenes, who combines Christianity with advocacy of Mandeville's doctrines, and the worldly, sceptical Horatio. *The Fable of the Bees. Part II* (1729) explicitly distinguishes between two types of self-gratification: self-love (satisfying self-preservatory desires) and self-liking (satisfying the desires for superiority, admiration and one's good opinion of oneself). (Rousseau later

used the same distinction.) *An Enquiry into the Origin of Honour and the Usefulness of Christianity in War* (1732) not only pursues Mandeville's discussion of honour and contrasts true with nominal religion, but also further refines the analysis of self-liking in showing that it causes both shame and pride. The *Fable Part II* is especially notable for elaborating the historical thesis suggested in Mandeville's earlier remarks on the development of artefacts, processes and institutions (Goldsmith 1985). The designs of ships and clocks are the result of the labour and experience of many ages. So too are the elaborate laws and the complex institutions of government which - contrary to some previous views - have not been invented by heroic individual benefactors and lawgivers. The conjectural history of the stages of human development from animal existence through savagery to commercial society, already indicated as early as the *Female Tatler*, shows that the *Fable*'s story of the civilizing of humans by skilful politicians and moralists is not to be taken literally. In holding that the development of language itself must have taken many ages, Mandeville implicitly rejected literal, biblical chronology.

Mandeville's views were widely known and extensively used and commented on in the eighteenth century, by thinkers such as Hutcheson, Voltaire and Rousseau among others; David Hume and Adam Smith clearly knew Mandeville's discussion of human nature, the utility of vice and the beneficial results of complex social interaction (Goldsmith 1988; Hundert 1994).

M.M. GOLDSMITH

List of works

Mandeville, B. (1703) *Some Fables after the Easie and Familiar Method of Monsieur de la Fontaine*, London: Richard Wellington; enlarged edn, *Aesop Dress'd; or a Collection of Fables Writ in Familiar Verse*, London: Richard Wellington, 1704; reset edn of *Aesop Dress'd*, London: 'sold at Lock's-Head adjoyning to Ludgate', 1727[?]; reprint of reset edn of *Aesop Dress'd*, intro. J.S. Shea, Augustan Reprint Society Publication no. 120, Los Angeles, CA: Augustan Reprint Society, 1966.(Verse fables, based on La Fontaine's version of *Aesop*, but containing two original fables composed by Mandeville.)

Mandeville, B. (1705) *The Grumbling Hive: or, Knaves Turn'd Honest*, London: printed for S. Ballard and sold by A. Baldwin.(The verse fable used in *The Fable of the Bees*.)

Mandeville, B. (1709) *The Virgin Unmask'd: or, Female Dialogues Bewixt an Elderly Maiden Lady, and Her Niece*, London: sold by J. Morphew and J. Woodward; reprinted, intro. S.H. Good, Delmar, NY: Scholars' Facsimiles, 1975.(Dialogues on mores, manners and political affairs.)

Mandeville, B. (1709-10) *The Female Tatler*.(Thirty-two issues between numbers 52, 2 November 1709, and 111, 29 March 1710, signed 'Lucinda' and 'Artesia', discussing topics including the origin and development of society, honour and female virtue.)

Mandeville, B. (1711) *A Treatise of the Hypocondriack and Hysterick Passions*, London: printed for the author, D. Leach, W. Taylor and J. Woodward; enlarged 3rd edn, *A Treatise of the Hypocondriack and Hysterick Diseases*, London: printed for J. Tonson, 1730; reprinted 1st edn and reprinted 3rd edn, intro. S.H. Good, Delmar, NY: Scholars' Facsimiles, 1976.(Dialogues between doctor and patients on the causes and cures of nervous and digestive diseases.)

Mandeville, B. (1714) *The Fable of the Bees: or, Private Vices, Publick Benefits*, London: printed for J. Roberts; 2nd edn, enlarged, London: printed for E. Parker, 1723; 3rd edn, enlarged, London: printed for J. Roberts, 1724; reprinted, ed. F.B. Kaye, Oxford: Clarendon Press, 1924.(Mandeville's best-known work. Kaye's classic critical edition includes life, bibliography, annotations, discussions of Mandeville's thought, its background and influence, brief summaries of contemporary critics and a chronological list of references to Mandeville's work to 1923.)

Mandeville, B. (1729) *The Fable of the Bees. Part II*, London: printed for J. Roberts; reprinted, ed. F.B. Kaye, Oxford: Clarendon Press, 1924, vol. 2; reprinted, Indianapolis, IN: Liberty Classics, 1988.(Dialogues expanding, explaining and refining the *Fable*'s doctrines; included in Kaye's classic edition.)

Mandeville, B. (1720) *Free Thoughts on Religion, the Church, and National Happiness*, London: sold by T. Jauncy and J. Roberts; reprinted, intro. S.H. Good, Delmar, NY: Scholars' Facsimiles, 1981; 2nd edn, London: printed for John Brotherton, 1729; reprinted, Stuttgart-Bad Cannstatt: Frommann, 1969.(Discussion of religion heavily indebted to Pierre Bayle, otherwise relatively straightforward advocacy of toleration, anticlericalism and political whiggism.)

Mandeville, B. (1724) *A Modest Defence of Public Stews: or, An Essay upon Whoring, as it is now practis'd in these Kingdoms*, London: printed by A. Moore; reprinted, intro. R.I. Cook, Los Angeles, CA: Augustan Reprint Society, no. 162, 1973.(Standardly attributed to Mandeville; advocates public houses of prostitution to preserve

chastity.)

Mandeville, B. (1732) *An Enquiry into the Origin of Honour and the Usefulness of Christianity in War*, London: printed for J. Brotherton, 1732; reprinted, intro. M.M. Goldsmith, London: F. Cass, 1972.(Dialogues on honour and the distinction between real and nominal Christianity.)

Mandeville, B. (1732) *A Letter to Dion, Occasion'd by his Book call'd Alciphron, or the Minute Philosopher*, London: printed for J. Roberts; reprinted, intro. J. Viner, Los Angeles, CA: Augustan Reprint Society, no. 41, 1953.(Defends the *Fable of the Bees* on the grounds that it advocates strict morality and exposes vice; Viner denies that Mandeville advocated *laissez-faire* while favouring mercantilist government intervention.)

References and further reading

Castiglione, D. (1986) 'Considering Things Minutely: Reflections on Mandeville and Eighteenth-Century Science of Man', *History of Political Thought* 7 (3): 463-88.(A subtle discussion of Mandeville's satiric critique of eighteenth-century values: his enigmatic views resisted 'moralization'.)

Castiglione, D. (1992) 'Excess, Frugality and the Spirit of Capitalism: Readings of Mandeville on Commercial Society', in J. Melling and J. Barry (eds) *Culture in History: Production, Consumption and Values in Historical Perspective*, Exeter: University of Exeter Press.(Reviews understandings of Mandeville's 'consumptionist' account of capitalism and shows how the impact of his 'paradox' was reduced as conceptions of society and capitalism changed.)

Chalk, A.F. (1966) 'Mandeville's *Fable of the Bees*: A Reappraisal', *Southern Economic Journal* 33 (1): 1-16. (Mandeville's views are transitional between mercantilism and laissez-faire.)

Colman, J. (1972) 'Bernard Mandeville and the Reality of Virtue', *Philosophy* 47: 125-39.(Mandeville's Augustinianism is satirical. He does not deny that virtue exists, only that it is innate; virtuous actions are a product of moral education.)

Dickinson, H.T. (1976) 'Bernard Mandeville: An Independent Whig', *Studies on Voltaire and the Eighteenth Century* 151-5: 559-70.(Mandeville's political views are whig, but neither unambiguously 'Court Whig' or 'Country Whig'.)

Dickey, L. (1990) 'Pride, Hypocrisy and Civility in Mandeville's Social and Historical Theory', *Critical Review* 4: 387-431.(Mandeville historicized the concept of self love - pride becomes socialized into hypocrisy at a decisive moment in the civilizing process, thereby providing intersubjective constraints in commercial societies.)

Edwards, T.R., Jr (1964) 'Mandeville's Moral Prose', *ELH* 31: 195-212.(Mandeville expresses moral indignation about the complacency of the upper classes.)

Goldsmith, M.M. (1985) *Private Vices, Public Benefits: Bernard Mandeville's Social and Political Thought*, Cambridge: Cambridge University Press.(Considers Mandeville's views, §§2-3, 4, on the operation and development of social institutions; emphasizes the importance of the *Female Tatler*; rejects the contention that Mandeville was an economic thinker; argues that he accepted modes of life rejected by the 'polite' aristocratic ethic of the early eighteenth century; extensive bibliography.)

Goldsmith, M.M. (1986) '"The Treacherous Arts of Mankind": Bernard Mandeville and Female Virtue', *History of Political Thought* 7 (1): 93-114.(Discusses Mandeville's proto-feminist views in some of the *Female Tatler* essays.)

Goldsmith, M.M. (1988) 'Regulating Anew the Moral and Political Sentiments of Mankind: Bernard Mandeville and the Scottish Enlightenment', *Journal of the History of Ideas* 49 (4): 587-606.(Examines Hutcheson's, Hume's and Smith's discussions of Mandeville and their use of his views.)

Hayek, F.A. (1966) 'Dr Bernard Mandeville', *Proceedings of the British Academy* 52: 125-41.(Classic attribution to Mandeville (§3) of *laissez-faire* individualism and the notion of 'unintended consequences'.)

Hirschman, A.O. (1977) *The Passions and the Interests: Political Arguments for Capitalism before Its Triumph*, Princeton: Princeton University Press.(The eighteenth-century transition from the doctrine that irrational 'passions' had to be controlled by reason to acceptance of their control by private 'interests' in which Mandeville is significant.)

Horne, T.A. (1978) *The Social Thought of Bernard Mandeville: Vice and Commerce in Early Eighteenth-Century England*, New York: Columbia University Press.(Emphasizes Mandeville's relations to the 'reformation of manners', the French *moralistes*, Shaftesbury and mercantilism, and sees him as justifying commercial society, §§2-3.)

Hundert, E.J. (1994) *The Enlightenment's Fable: Bernard Mandeville and the Discovery of Society*, Cambridge: Cambridge University Press.(Extensive treatment of both the British and continental Europe intellectual context and and the sources of Mandeville's views; emphasizes Mandeville's importance for many

Enlightenment thinkers and for the development of several areas of Enlightenment thought; extensive bibliography.)

Jack, M. (1989) *Corruption and Progress: the Eighteenth-Century Debate*, New York: AMS Press. (Mandeville's contention that moral decline is intrinsic to material progress is central to the eighteenth-century debate about 'corruption'.)

Lovejoy, A.O. (1961) *Reflections on Human Nature*, Baltimore: Johns Hopkins University Press.(Analysis of the self-regarding passions labelled 'pride' with discussion of Mandeville.)

Maxwell, J.C. (1951) 'Ethics and Politics in Mandeville', *Philosophy* 26: 242-52.(Mandeville's ethical views are not inconsistent; moreover he applied a rigoristic standard to individual conduct, but a consequentialist one to politicians.)

Monro, H.D. (1975) *The Ambivalence of Bernard Mandeville*, Oxford: Clarendon Press.(Discusses most aspects of Mandeville's life and works: comprehensive review of moral positions attributed to Mandeville and extensive examination of him as an ethical theorist; critically annotated select bibliography, including works on Mandeville 1924-72.)

Primer, I. (ed.) (1975) *Mandeville Studies: New Explorations in the Art and Thought of Dr Bernard Mandeville*, The Hague: Nijhoff.(Essays on many aspects of Mandeville's writings, mainly by literary scholars; useful bibliography.)

Rosenberg, N. (1963) 'Mandeville and Laissez-Faire', *Journal of the History of Ideas* 24: 183-96.(Holds that Mandeville mainly advocated *laissez-faire* economics.)

Schneider, L. (1987) *Paradox and Society*, New Brunswick: Transaction Books.(Mandeville as a sociologist.)

Scott-Taggart, M.J. (1966) 'Mandeville: Cynic or Fool', *Philosophical Quarterly* 16: 221-32.(Mandeville proposed a naturalistic explanation of the fact of morality, thereby raising the problem of grounding moral values.)

Smith, A. (1759) *The Theory of Moral Sentiments*, ed. D.D. Raphael and A.L. Macfie, Oxford: Clarendon Press, 1976.(Smith's moral philosophy, developing the theory of the 'impartial spectator' and denouncing Mandeville's 'wholly pernicious views', which eliminated the distinction between virtue and vice.)

Manicheism

Manicheism is a defunct religion, born in Mesopotamia in the third century AD *and last attested in the sixteenth century in China. Its founder, Mani (*c.*216-76), had some familiarity with Judaism, Christianity, Zoroastrianism and Buddhism, and aimed to supplant them all. He taught a form of dualism, influenced by earlier Gnostics: God is opposed by forces of darkness; they, not God, created human beings, who nevertheless contain particles of light which can be released by abstemious living. Two points of contrast with Catholic Christianity are particularly striking. First, in Manicheism, sinfulness is the natural state of human beings (because of their creators), and does not stem from Adam's Fall. Second, the Manichean God did not create and does not control the forces of darkness (although he will eventually triumph); hence the problem of evil does not arise in as stark a form as it does for the all-powerful Christian God.*

Although Mani's own missionary journeys took him eastwards, it was in the Roman Empire to the west that the main impact of his teaching was first felt; Augustine of Hippo was an adherent for nine years. The religion was eventually suppressed in the Roman Empire, and driven east by the Arab conquest of Mesopotamia. In the West, various Christian heresies were loosely called Manichean throughout the Middle Ages.

1 The background

Dualism emerges into history with the Persian prophet Zoroaster (or Zarathustra), who lived perhaps as early as the thirteenth century BC, and who taught that the High God Zervan or his son Ormuzd (Mazda) battles with Ahriman, the Spirit of Darkness, and will one day send a Saviour to redeem mankind. Here we have the seeds of Mani's religion, seeds that may have been nourished by the Zoroastrian state cult of the Sassanian dynasty which came to power in Persia during Mani's lifetime. But a stronger affinity was with Gnosticism. Gnostic tenets, perhaps originating among certain 'God-fearing' Judaic communities (and perhaps themselves owing something to much more ancient Zoroastrian influence on Judaism), become known to us around the end of the first century AD. Many of the groups they affected, such as the followers of Marcion (d. *c.*160), absorbed enough Christianity to be attacked as Christian heretics.

Within much diversity the common features of Gnosticism were these. God the First Principle emanated divine associates or 'Aeons'. Among them was one who through pride or curiosity fell. It was this one, the Demiurge, who created the visible world and human beings in it. Human beings contain a spark of divinity, which it became God's purpose to rescue by imparting knowledge, gnosis, through his messenger Jesus. Gnostics repudiated the Old Testament, sometimes identifying the Demiurge with Yahweh and honouring Yahweh's opponents such as the Serpent and the Sodomites. Jesus was not born, or was born through Mary's ear, and he stood by the cross while a surrogate phantom suffered and died in his place. Gnosis is imparted only to the initiated. It teaches that the human body is evil, and it therefore enjoins ascetic practices and deplores procreation (see Gnosticism).

2 Mani

Mani (or Manes or Manichaeus) was born around 216 and brought up near Ctesiphon on the Tigris among a sect of baptists apparently founded a century earlier by a Jew named Elchasai. The sect observed the Sabbath, washed a great deal, and expected the imminent end of the world following the appearance of a final prophet. At the age of 12, and again at 24, Mani saw visions which he came to compare with the apostle Paul's. Like Paul he felt called to renounce the legalism of his upbringing and to take a message to the world; unlike Paul's, his message included a complete cosmological story. With the favour of the Sassanian king Shapur I he preached in Persia, organized missions as far west as Alexandria, and himself travelled eastwards, perhaps to India. After the death of Shapur's successor in 273 Mani fell out with the next king and was confined to prison, where he died in 276.

Mani held that God, the father of truth, exists before all things and persists after all things, and that everything exists through his power (*Concerning the Origin of His Body*, 66.4-15; Cameron and Dewey 1979: 52-3). But he also warned against those who 'teach that good and bad come from the same source, and introduce a single principle, not distinguishing or dividing darkness from light and good from bad…as we preach' ([Hegemonius] v). In prosaic summary, the story he preached was this. When first the Lord of Darkness and his demons invaded the realm of light, God sent Primeval Man attired in bright elements to resist them. But Primeval Man was defeated

and stripped of his attire, which the demons swallowed. In his quest to retrieve this stolen light God emanated seductive Messengers, the sight of whom caused emissions and abortions among the demons, from which sprang plants and beasts. Some of the light being thus trapped anew, a demon couple engendered Adam as a repository for most of the remainder. With Jesus' help we, as Adam's descendants, can by breathing and belching release our light upwards, feeding the moon as it waxes, which as it wanes feeds the sun in turn; but both eating and procreation have the contrary effect of trapping light. In the end Jesus will come again and light will triumph.

Mani's church was organized, with grand books, music, bishops and eventually monasteries, for the purpose of assisting God in this cosmic struggle. Its membership was divided between Elect and Hearers. Even Hearers (such as Augustine was) must not eat meat, or kill, must fast on Sundays, avoid procreation, and above all sustain the Elect - men and women - who were additionally forbidden to drink wine, copulate, farm, or prepare their own food.

The Manichees were interested in natural, especially celestial, phenomena and offered explanations of them which had to be consistent with Mani's cosmology.

3 Augustine and the Manichees

Perhaps in emulation of Paul, Mani signed his letters 'Apostle of Jesus Christ'. In Augustine's day the African Manichees claimed Christianity for their own and called Catholics *semichristiani*. One thing that underlay this surprising pretension was the Manichees' refusal to compromise the austerity they discerned in Paul's primitive Christian communities. Another may have been Mani's dualism, which not only could offer comfort to sensitive and frightened people in the thought that their sins and shortcomings stemmed from the machinations of superior powers, but could easily in the third and fourth centuries find heralds of that thought among Gnosticizing Christians. Manicheism is not strictly dualistic: its God is above all, and there is assurance that he will win in the end. Besides, Christians and Manichees shared a common cosmological concern with the problem of evil, agreeing both that it is a problem (why should there be any?) and that the solution lies in the past, when an originally happy state was upset by some fall from good behaviour.

What at this high level chiefly marks off Manicheism from Catholic Christianity, as the latter emerged in the councils and apologetics of the fourth century, is Mani's conception of human evils, sin and suffering, as tracing their origin back to an act - creation by demons - which unlike the biblical Fall was not the act of any human. Thus the two religions differ in their anthropology: for Manichees, sin and suffering are our natural state, not an Adam-induced lapse from it (see Sin §2). Augustine saw that difference, and used it to frame the risky charge that Manichean metaphysics leaves no room for human responsibility - risky because he was later to find the same charge brought against his own Catholic metaphysics by the Pelagians (see Augustine (§§6, 9, 12); Pelagianism §1-2). Manichees also, of course, seemed to have a readier means than Catholics of exculpating God from complicity in evil; for their God had not created and did not control the powers of darkness.

4 Dualism in Christianity

Augustine wrote of the time when Manicheism was losing its grip on him, 'It still seemed to me that it is not we who sin, but some other nature within us' (*Confessions*, 5.10.18); here was one comfort he would lose by resuming Catholic Christianity. Another was freedom to criticize the Bible, for the early Manichees won much of their missionary success by an aggressive policy of mocking the Old Testament and exposing inconsistencies in the Catholic canon of the New (their own New Testament canon was more Gnostic). But aggressiveness faded under persecution in the West, till the religion was exterminated from Christendom by about 600. It continued in Mesopotamia until driven east by the Arab conquest in the seventh century. For a time it was the state religion of the Uighur Turks in Sinkiang, and from the T'ang period it established a lasting presence in China, being last attested on the Fukien coast in the sixteenth century.

In Christendom, 'Manichee' became an opprobrious epithet for heresies that pronounced human nature evil (so excluding redemption) and harboured a class of Elect (so rejecting infant baptism). Paulicians, who flourished in Armenia and the eastern confines of the Roman Empire in the eighth and ninth centuries, believed that the Demiurge, not God, made the world and Jesus only seemed to die. In tenth-century Bulgaria, Bogomil (Theophilus, 'God-beloved') echoed much of Mani's cosmology and ethics. His views guided the

thirteenth-century Sclavonian church of Bosnia, called Patarene by the West. During the twelfth century there were Poblicans in Flanders, perhaps inheritors of Paulicianism through itinerant weavers. And the Cathars of the same period, powerful in southern France until de Montfort's 'crusade' overthrew them, joined Manichean dietary and sexual proscriptions with the old Gnostic rejection of Christ's humanity. All of these sects, while attracting support from simple people, had leaders who knew history and tried to model their lives on past practice: not indeed on the outlandish Mani, about whom they probably knew little and cared less, but on primitive Christianity as transmitted through early Gnosticism. They succumbed to Catholic persecution in the West, and many in the Balkans converted to Islam after the Turkish conquests of the fifteenth century. But even in modern times like-minded movements arise now and again, nourished by certain books of the Bible, especially Daniel and Revelation.

See also: Evil; Evil, problem of

CHRISTOPHER KIRWAN

References and further reading

Mani is credited with seven works in Syriac, and one in Middle Persian addressed to King Shapur, but only fragments survive. Traditions were preserved by Muhammad ben Ishâk's 'Doctrines of the Manicheans' in the *Fihrist*, a history of Arabic literature to 987 compiled by an-Nadîm, and less reliably in various hostile writings from the fourth and fifth centuries by Christians and others. Further Manichean writings have come to light during the twentieth century in Chinese and other languages (Turfan 1904-15, Tun-huang 1905), Coptic (Egypt, Medinet Madi 1930), Greek (Cologne 1969) and Latin. The greater part of the Coptic collection, including letters of Mani himself, went to the Berlin Academy from where it disappeared in 1945 before publication was complete; most of the remainder, including homilies edited by Polotsky and a Psalm Book edited by Allberry, is in the Chester Beatty Collection in Dublin. Many of these texts are collected in Adam (1969). The Cologne manuscript edited by Koenen and Römer is a gospel-cum-apocalypse in Greek from a Syriac original, some of it presented as autobiography by Mani. It dates from about 400.

Adam, A. (1969) *Texte zum Manichäismus (Manichean Texts)*, Kleine Texte für Vorlesungen und Übungen, vol. 175, enlarged edn, Berlin: de Gruyter.(Contains the basic texts, apart from the Cologne manuscript, in Greek, Latin or German translation.)

Allberry, C.R.C. (ed.) (1938) *A Manichaean Psalm Book, Part II*, Manichaean Manuscripts in the Chester Beatty Collection 2, Stuttgart: W. Kohlhammer.(Sumptuous edition with English translation. From the Coptic find.)

Augustine (397-401) *Confessions*, Corpus Scriptorum Ecclesiasticorum Latinorum, vol. 33, Vienna: Tempsky, 1896, and elsewhere; trans. H. Chadwick, Oxford: Oxford University Press, 1991, and elsewhere.(One of the many works in which Augustine as a Catholic bishop responded to charges of lingering Manicheism.)

Augustine (400) *Contra Faustum Manichaeum*, Corpus Scriptorum Ecclesiasticorum Latinorum, vol. 25.1, Vienna: Tempsky, 1891; trans. R. Stothert, in P. Schaff (ed.) (1886-9) *A Select Library of the Nicene and Post-Nicene Fathers*, vol. 4, Grand Rapids, MI: Eerdmans, 1974.(A lengthy polemic against the most celebrated local Manichee.)

Cameron, R. and Dewey, A.J. (1979) *The Cologne Mani Codex 'Concerning the Origin of His Body'*, Society of Biblical Literature, Texts and Translations 15, Early Christian Literature Series 3, Missoula, MT: Scholars Press.(Incomplete text and translation of the work later edited by Koenen and Römer.)

Chadwick, H. (1989) 'The Attractions of Mani', *Compostellanum* 34: 203-22; reprinted in H. Chadwick, *Heresy and Orthodoxy in the Early Church*, Aldershot: Variorum, 1991.(Describes Manicheism as it presented itself to Augustine and his contemporaries.)

[Hegemonius] (*c*.320) *Acta disputationis Archelai cum Manete haeresiarcha*, ed. C.H. Beeson, Leipzig: Hinrichs, 1906.(A hostile biography of Mani, but using Manichean material. Forged in Syriac and preserved only in Latin. The chief early source in the West.)

Koenen, L. and Römer, C. (eds) (1985) *Der Kölner Mani-Kodex (The Cologne Mani Codex)*, diplomatic edition, Papyrologische Texte u. Abhandlungen, vol. 35, Bonn: Habelt.(Edition of the Cologne manuscript, a tiny codex measuring 4.5 by 3.5 cm.)

Lieu, S.N.C. (1992) *Manichaeism in the Later Roman Empire and Medieval China: A Historical Survey*, Tübingen: Mohr, revised 2nd edn.(An excellent study of the religion's history, West and East.)

Polotsky, H.J. (1934) *Manichäische Homilien (Manichean Homilies)*, Manichaean Manuscripts in the Chester Beatty Collection 1, Stuttgart: W. Kohlhammer. (Sumptuous edition with German translation. From the Coptic find.)

Runciman, S. (1947) *The Medieval Manichee*, Cambridge: Cambridge University Press.(Traces the Christian heresies that were lumped together as Manichean.)

Manifoldness, Jaina theory of

The Sanskrit term anekāntavāda *literally means 'not-one-sided doctrine', and refers to the Jaina epistemological theory of manifold standpoints from which an object may be considered and the manifold predications that can be made with regard to it. It evolved out of Mahāvīra's ethical emphasis on nonviolence - the multidimensional nature of objects should not be violated by single, absolutist (*ekānta*) predications about them. Respect for life is thus transformed in its philosophical application into a principle of respect for other views. The theory has come to be called the central philosophy of Jainism and was developed in a milieu of intensive debate between the various Indian philosophical schools. Though the theory was based on Mahāvīra's teaching, it implicitly presupposed, in its later highly developed form, various philosophical alternatives (representing the views of other schools of thought), which it sought to syncretize. Each standpoint and predication presents a partial truth and, according to Jainism, only the theory of manifoldness does justice to the complex nature of entities. While it can be seen as an attempt to practice intellectual nonviolence, it is evident that the Jainas adhered to it zealously and defended it as vehemently as the others did their own views.*

1 Historical background

There is an obvious similarity between the Jaina theory of manifold predication (*anekāntavāda*) and its precursor, the theory of analysis or differentiation (*vibhajyavāda*), which is more famous as Buddha's method of dealing with questions, but was also employed by Mahāvīra. Buddha is renowned for not categorically answering metaphysical questions and for seeking an appropriate answer by breaking up (*vibhajya*) a question into its component parts, often thereby demonstrating it to be a pseudo-question. B.K. Matilal (1981: 11) refers to two types of this theory: the first 'operates by dividing the subject class into sub-classes' (to the question 'Will all beings be reborn?' the answer is 'Only those with defilements'); the second 'operates by specifying or relativizing the predicate' (the answer to the question 'Is man superior or inferior?' is 'Man is inferior to gods and superior to the lower beings'). According to Matilal, it was the second type that Mahāvīra adopted and which was developed into the *anekānta* method.

The *BhagavatīSūtra*, part of the Jaina canon, has examples of this style of teaching that are comparable with the better-known ones of Buddha (for example, IX, 33, 485b; Deleu: 164). To quote Matilal again:

> The Buddha's method was one of withdrawal from philosophical disputes, for he avoided committing himself to any extreme view. But Mahāvīra's method was one of commitment, for he attempted to understand the points of view of the fighting parties (in a philosophical dispute) so that their dispute could be resolved and reconciled. Thus, the essence of the *anekānta*-*vāda* lies in exposing and making explicit the standpoints or presuppositions of the different philosophical schools.
>
> (1981: 23)

Both Jaina and Buddhist sources refer to one Gośāla, a contemporary of Mahāvīra and Buddha and founder of the Ājīvaka sect, which disappeared very soon after his death. He is reputed to have asserted that all entities have a triple character: existent, nonexistent, and both existent and nonexistent (or living, nonliving, and both living and nonliving). The Jaina theory seems to interpret these as positive, negative, and both positive and negative; when they were combined with the possibility that an entity can have an inexpressible nature, a sevenfold formula emerged (see §3).

Just as the Buddhists used the famous fourfold alternative to establish the doctrine of the void (Nāgārjuna's alternatives are: everything is true; not everything is true; both everything is true, and not everything is true; or neither everything is true nor is everything not true), so the Jaina developed the theory of standpoints and sevenfold predication to defend the doctrine of manifoldness (see Nāgārjuna §3). The key to the Jaina position is the need to qualify an assertion or to make it conditionally true in order to leave open other possibilities. This represents the Jaina respect for the views of other schools and at the same time implicitly criticizes their absolutist stands, which have the potential for dogmatism and intolerance. The way to overcome one-sided doctrines is to adopt Mahāvīra's method of accepting metaphysical theories with a qualified affirmation. The philosophic attempt to come as close as possible to capturing the true, multidimensional nature of an entity is to qualify one's assertion with the particle *syāt*.

2 The significance of *syāt*

The Jaina theory of the manifoldness of reality has two aspects: the doctrine of standpoints (*nayavāda*) and the sevenfold predication (*saptabhaōgī*) associated with it, whereby each type of predication is an epistemological assertion from a certain viewpoint (*syāt*). The theory of manifoldness (*anekāntavāda*) is often used as a synonym for the theory of viewpoints (*syādvāda*). The word *syāt* is identical with a verbal form (the third person singular of the optative mood) of the Sanskrit root 'to be' (*as*), and means things like 'perhaps, possibly, maybe it is or exists'. This has led to a description of the theory as agnostic, giving the impression of doubt, uncertainty and indeterminacy and an inability definitely to determine an object of investigation. The crucial point of the theory is missed, however, if the word is interpreted as a verbal form. Each of the seven statements (see §3) which may be explicitly made with respect to an entity has, in addition to the word *syāt*, a finite form of the same root 'to be', the third person singular present indicative (*asti*, implicit in the fourth case). The technical sense in which it is employed by the Jainas, which semantically does justice to its use, requires an analysis of *syāt* as an indeclinable word meaning 'from a point of view' or 'in a certain sense', with the obvious extension 'from another point of view' or 'in another sense'. Thus, the Jaina theory does not imply that it is epistemologically difficult to determine the nature of an object, but rather that it can be definitely determined from several particular points of view, even though some assertions have a negative structure or express indescribability.

3 The seven standpoints

The two aspects of the theory of manifoldness mentioned at the beginning of §2 relate to two different but theoretically related contexts in which it is employed. The doctrine of standpoints (*nayavāda*) is the general framework indicating seven ways of considering an object or entity, and the theory of viewpoints (*syādvāda*) represents the seven assertions which may be made regarding an object or entity. Both aspects are intended to demonstrate how partial truths about entities are presented. Jainism makes a clear distinction between knowledge of an entity as a whole and knowledge of a part or an aspect of an entity. This is the epistemological distinction implicit in the terms *pramāṇa*, which signifies complete knowledge of an entity in all its qualities and modes of appearance, and *naya*, which signifies partial knowledge with reference to an entity as it appears to be in the present time and place, having a particular quality and mode which may later change. Complete knowledge of an entity, simultaneously grasping it in all its modes and in all possible ways, occurs only in the case of beings unhindered by karma and is a sign of omniscience. In the case of human knowledge, which is restricted to the functions of the sense organs whose capacities are impeded by karma, only partial knowledge is posssible. The doctrine of standpoints thus represents the various kinds of partial knowledge accessible to human beings. It serves two functions: it facilitates analyses of objects of inquiry and, when applied to other schools of thought, it enables Jainism to classify the partial truths of those schools in terms of one or the other standpoint.

Jainism evinces what may be called a passion for enumeration, classification and categorization of practically all the major philosophical terms: the single term 'karma', for instance, is divided into eight important kinds with a further 148 others; and knowledge is deemed to be derived in two ways, each with its own specific kinds (see Jaina philosophy §2). Theoretically, an infinite number of standpoints are possible when considering an entity, but traditionally Jainism enumerates the following seven in the context of the theory of manifoldness:

(1) The common or nondistinguished standpoint (*naigamanaya*) refers to an entity without taking its general and specific characteristics or qualities into consideration. The words 'jar' and 'cloth', for example, can be used in two senses according to this standpoint. They may refer to specific objects within one's view or they may refer to the class of objects that they denote. The example of 'cooking' is also mentioned here; ambiguity can arise when no clear distinction is made between the activity of actually cooking and the preliminary preparations. The general and particular properties of an object or activity are not explicitly distinguished in this standpoint, yet statements concerning 'jar' and 'cooking' are easily understandable. This standpoint represents what is understood conventionally; it is emphasized, according to the Jainas, by the Nyāya-Vaiśeṣika school. The fifth-century Jaina philosopher Siddhasena Divākara does not recognize this as a separate standpoint.

(2) The general or collective standpoint (*saōgrahanaya*) takes into consideration the class to which an entity belongs, as, for example, when one refers to fruit trees in general. This point is indirectly related to the problem of universals and treats class properties as but one way among others of referring to entities. Class properties like

'fruitness' and 'treeness' are treated as representing a partial truth about an entity. With reference to the all-encompassing category 'being', this standpoint represents the Vedānta school, with its notion of Brahman as the only ultimately real substance. Whereas the first standpoint may be ambiguous in that it can refer to both a particular object as well as a class of objects, this standpoint represents the partial knowledge of an entity because it emphasizes the general aspect alone.

(3) The practical standpoint (*vyavahāranaya*) is a complement to the second and a corollary to the first in that it refers to something specific. When one asks for a fruit, the reference is to something specific, for fruit in general cannot be brought. It is a standpoint which implicitly particularizes an entity and, like the first standpoint, is generally understood. The practical standpoint implies classification and differentiation and is said to represent the view of the materialist Cārvāka school. Like the second standpoint, it avoids the ambiguity of the first, but is no less a partial standpoint, an emphasis on which can furnish only partial knowledge of an entity, because it emphasizes a single, particular object.

(4) The 'straight thread' standpoint (*ṛjusūtranaya*) considers an entity as it appears at the present moment, ignoring both its previous form and the one it will have in future. An emphasis on this partial truth leads to the metaphysical conclusion of existence as momentary. The Jaina theory includes the Buddhist doctrine of universal flux here as an example of the kind of partial truth furnished by this standpoint (see Momentariness, Buddhist doctrine of §1).

(5) The verbal standpoint (*śabdanaya*) represents the method of referring to an entity in terms of the syntactic relations between the words used in a sentence where the meaning is determined by grammatical inflections which can offer different semantic interpretations. The analysis of this standpoint is inextricably related to Sanskrit grammar. The meaning of a sentence is dependent on the relations between its words and these relations are grammatically governed. The partial knowledge is represented by the specific interpretation given to the relations, which then furnish a meaning relative to it.

(6) The subtle standpoint (*samabhirūḍhanaya*) makes a subtle distinction between synonyms in terms of their etymology. This standpoint represents a partial truth in which only etymological differences are emphasized, whereas the object referred to is the same. Thus the same person may be referred to as a king, regent, or sovereign; the Sanskrit equivalents of these words can produce different meanings depending on their derivations. A partial view is represented according to this standpoint when the meaning of synonymous words is explained *only* on the basis of their etymology.

(7) The 'thus-happened' standpoint (*evambhūtanaya*) is an extension of the above in that it restricts the meaning of a word to a particular occasion of use. Thus the word 'cook' should be used, according to this standpoint, only when a person is actually cooking. The restriction of the use of words to actual cases gives only a partial truth because it either ignores a person's other activities, or could imply that a person is not a cook by occupation (say) when doing something else.

The second aspect of the Jaina theory of manifoldness, which is correlated with the theory of standpoints, represents statements or predications about an object that can be made from only one specific point of view. As with the seven standpoints, the seven viewpoints - which constitute the framework of the Jaina theory of sevenfold predication - are intended to apply to any specific entity. One can make the following assertions about any entity:

(i) From one point of view it exists (*syād asti*).
(ii) From one point of view it does not exist (*syān nāsti*).
(iii) From one point of view it exists and from another point of view it does not exist (*syād asti syān nāsti*). This assertion, combining (i) and (ii) into a single unit, does not apply simultaneously, thereby avoiding an apparent contradiction. The assertion may be fully expressed as, for example: 'from one point of view a pot exists (as a pot) and from another point of view it does not exist (as a tree).'
(iv) From one point of view it is inexpressible (*syād avaktavyam*).
(v) From one point of view it both exists and is inexpressible (*syād asti cāvaktavyam*).
(vi) From one point of view it both does not exist and is inexpressible (*syān nāsti cāvaktavyam*).
(vii) From one point of view it exists, does not exist and is also inexpressible (*syād asti ca nāsti cāvaktavyam*).

Jainism ascribes validity to the need for these kinds of assertions about an entity on the basis of lists of factors mentioned in the Jaina canon which were employed when investigating the nature of an entity. The lists originally

served as tools for investigating words, especially the titles and chapter titles of canonical texts, but later took on an epistemological significance when applied to assertions about entities in general. They were ‘gateways of investigation’ (*anuyogadvāra*) which served to ‘set down’ or ‘bring forward’ (*nikṣepa*) the context in which the investigation of an object applies. Of the several lists (containing up to thirteen items) that feature in the canonical texts, the list of four is the most popular. According to this list, clarity about an entity can be achieved if there is an analysis that considers it in terms of: the name or the word used to designate the entity; the form or the way in which it can be illustrated; the substance out of which it is constituted; and its specific state or condition at the moment of investigation. Other factors, such as place, time and the means by which an object is known, were also used, with analyses taking on complex dimensions when repetitively (and often futilely) applied to the objects with which an entity is compared. However, this method of analysis directly influenced the later form of the theory of manifoldness and served, in the last resort, as its justification. The historical background and traditional application of the theory were generally ignored by its critics when assessing its contribution.

The Jaina theory of manifoldness was regarded by the other schools of Indian philosophy as generally representing the entire Jaina philosophy, so much so that it was what they attacked above all. It was asserted that the theory is self-contradictory, and that it demonstrates an uncertain and vague standpoint, with equivocality as its greatest error. Jainism was criticized for not being in a position clearly to describe reality. This explains why the majority of Jaina thinkers take every opportunity to justify the theory (which thus features in practically all Jaina philosophical studies) and to vindicate the nonabsolutism of Jainism. One of the most insightful and even independent thinkers within Jainism is Vidyānandin (ninth century), who belonged to the Digambara tradition of Jainism. Accepting three basic kinds of predicates, the positive, the negative, and the neutralized or indescribable, he concludes that only seven combinations are possible, which shows that the Jaina theory of manifoldness takes all cases into consideration when analysing an entity.

See also: Epistemology, Indian schools of; Knowledge, Indian views of; Mahāvīra

JAYANDRA SONI

References and further reading

Alsdorf, L. (1973) ‘Nikṣepa - A Jaina Contribution to Scholastic Methodology’, *Journal of the Oriental Institute (Baroda)* 22: 455-63; repr. in *Ludwig Alsdorf Kleine Schriften*, ed. A Wetzler, Wiesbaden: Franz Steiner Verlag, 1974, 257-65.(Explains the significance and practical use of the term, mentioned in the second last paragraph of the entry.)

Bhatt, B. (1978) *The canonical Nikṣepa. Studies in Jaina Dialectics*, Leiden: Brill. (Deals in detail with the term mentioned in the second last paragraph of the entry, using Jaina canonical sources.)

Malliṣeṇa (13th century) *Syād-vāda-mañjarī,* trans. F.W. Thomas, *The Flower-Spray of the Quodammodo Doctrine*, Delhi: Motilal Banarsidass, 1968.(An annotated translation of a technical work by a Jaina logician who defends the sevenfold standpoint against attacks by other schools of Indian philosophy. The work is a commentary which made famous Hemacandra’s *Anyayogavyavacchedikā*.)

Matilal, B.K. (1981) *The Central Philosophy of Jainism (Anekānta-Vāda)*, Ahmedabad: L.D. Institute of Indology.(More extended discussion of the material covered in §§1, 3. Comparison with Buddhist standpoint dealt with in detail, on the basis of original texts.)

Mookerjee, S. (1978) *The Jaina Philosophy of Non-Absolutism*, Delhi: Motilal Banarsidass, 2nd edn.(Discusses the theory of manifoldness and allied problems, like universals and particulars, to demonstrate the basic non-absolutist standpoint of the Jainas.)

Padmarajiah, Y.J. (1963) *A Comparative Study of the Jaina Theories of Reality and Knowledge*, Delhi: Motilal Banarsidass, repr. 1986.(Discusses Jaina ontology, and to a lesser degree, Jaina epistemology, in comparison with other schools of Indian thought.)

Many-valued logics

Many-valued logics may be distinguished from classical logic on purely semantic grounds. One of the simplifying assumptions on which classical logic is based is the thesis of bivalence, which states that there are only two truth-values - true and false - and every sentence must be one or the other. Many-valued logics reject the thesis of bivalence and permit more than two truth-values.

1 Motivations

Bivalence, in conjunction with some plausible assumptions about the nature of time, seems to lead to determinism. It seems plausible that no matter what I do now, I cannot change what was the case yesterday. This fact suggests that statements about the past are not contingent in the way most of us feel that statements about the future are. Suppose we want to know whether Charles will go sailing tomorrow, Saturday. We think Charles is a free agent, and his going sailing tomorrow is dependent on how he feels, his view of the weather, and other such factors. Consider the sentence ‘Yesterday it was the case that Charles will go sailing on Saturday’. Since yesterday is past, the truth-value of this sentence is fixed, even if we do not now know what it is. If the sentence is true, then it must be the case that Charles is going to go sailing tomorrow; and if the sentence is false, then it must be the case that Charles is not going to go sailing tomorrow. Either way, it seems that whether or not Charles is going sailing tomorrow is already determined, so the notion of future contingency is an illusion.

One way of escaping this argument for determinism is to reject the thesis of bivalence. The most obvious first approach would be to insist on three truth-values: ‘true’, ‘false’ and ‘indeterminate’. This problem was discussed by Aristotle (in *De Interpretatione*), although his position on the issue of bivalence remains a matter of debate. Indeterminists, such as the Epicureans, rejected the thesis of bivalence; determinists, such as the Stoics, advocated it. In the modern era, the Polish logician Łukasiewicz developed several many-valued logics as a means to escape the determinist argument (see Łukasiewicz, J. §3).

Many other motivations for many-valued logics have been advanced. In every case, there is much debate over the extent to which many-valued logic solves the problems which motivated its advocacy.

Another ancient motivation for expanding the stock of truth-values was the consideration of modalities. Some sentences, for example, ‘Two plus two is four’, seem to be not just true, but *necessarily* true. ‘The sun is shining today’, however, may well be true, but is not necessarily true since we can imagine an alternative. Such considerations suggest having at least four truth-values: ‘necessarily true’, ‘contingently true’, ‘contingently false’, ‘necessarily false’. However, it has been shown that most important systems of modal logic do not correspond to any finite-valued semantics.

Another common motivation for many-valued logics is the consideration of uncertainty. Uncertainty from many sources has been considered: from the incompleteness of human knowledge; from the lack of termination of machine computations; from the conflicting judgment of expert opinion; from true statistical fluctuations in nature; and from accumulating error. A major problem is that uncertainties do not combine in a truth-functional way, but many-valued logics are generally truth-functional.

Many attributions which we commonly make every day (for example, ‘Charles is ugly’) seem to be vague, that is, subject to degree, rather than being black or white (see Vagueness). Investigations of many-valued logics have been motivated by attempts to deal with inherently vague aspects of natural language. Considerations of vagueness have led to interest in so-called ‘fuzzy logic’ in computing science and electrical engineering (see Fuzzy logic).

Many-valued logics have been advocated as a means of avoiding various paradoxes, including the usual semantic paradoxes (see Semantic paradoxes and theories of truth), such as the liar paradox, as well as paradoxes associated with the comprehension axiom of set theory (see Paradoxes of set and property). For the semantic paradoxes, unless the syntax of the language is unnaturally restricted so that crucial aspects of the semantics cannot even be expressed, versions of the paradoxes simply reappear. For example, suppose we have three truth-values, and consider the sentence ‘This sentence is either false or has indeterminate truth-value’. If its value is false, then what it says is true; if its value is indeterminate then what it says is true; but if its value is true, then it is either false or indeterminate. Similarly, the paradoxes of adopting the unrestricted comprehension axiom in axiomatic set theory

cannot be avoided solely by adopting a many-valued logic.

2 An example

As a simple example, suppose that instead of the classical semantic range of just ‘true’ and ‘false’, we use the following.

T: logically true
t: empirically true
U: undeterminable
f: empirically false
F: logically false

The values T and t are said to be ‘designated’ values, or ways of being true. The other values are said to be ‘undesignated’. We could give semantic characterizations for a great number of connectives, but here are just a few examples.

A	$J_{\mathrm{T}}(A)$	$J_{\mathrm{t}}(A)$	$J_{\mathrm{U}}(A)$	$J_{\mathrm{f}}(A)$	$J_{\mathrm{F}}(A)$	$\sim A$	$\neg A$
T	T	F	F	F	F	F	F
t	F	T	F	F	F	f	F
U	F	F	T	F	F	U	T
f	F	F	F	T	F	t	T
F	F	F	F	F	T	T	T

$\supset$	T	t	U	f	F
T	T	t	U	f	F
t	t	t	U	f	f
U	U	U	U	U	U
f	U	U	U	U	U
F	U	U	U	U	U

<>< *eq* >

$\rightarrow$	T	t	U	f	F
T	T	t	f	f	F
t	T	t	t	f	f
U	T	t	t	t	t
f	T	t	t	t	t
F	T	T	T	T	T

The J_i connectives have the property that $J_i(A)$ has the value T if and only if *A* has the value *i*. They are known as ‘autodescriptive’ operators; they allow the expression in the syntax of the underlying semantic motivation. While $\neg$ corresponds to ‘is not designated’, the connective $\sim$ is the more usual complement negation. The conditional $\supset$ reflects the view that if the antecedent is not true, then we cannot tell the value of the conditional. However, $\rightarrow$ parallels the usual material conditional. The disjunction operator $\vee$ is the usual ‘max’.

The connectives $\neg$, $\rightarrow$ and $\vee$ are ‘pseudo-classical’ operators; if we map all the designated values to ‘true’ and the undesignated values to ‘false’, then the truth-conditions for these operators correspond to the classical connectives. The connectives $\sim$ and $\supset$ are not pseudo-classical.

Given our semantic account, we may inquire about an appropriate proof theory. Note that if we restrict our language to $\sim$ and $\supset$, then there will be no logical truths (formulas that are always designated); any formula composed with just these two connectives will take the undesignated value U when all atomic formulas take value U. Hence a proof theory for tautologies would not be appropriate were our language so restricted. However, an axiomatization of the consequence relation would still be of interest, and would include such sequents as $A, A \supset B \vdash B$.

Although often presented as restrictions of classical systems, many-valued logics can almost always be formulated as conservative extensions of classical logic. The trick is to include pseudo-classical connectives in the language.

A classical proof theory may be stated using the pseudo-classical connectives. Then using the autodescriptive operators, the additional required principles can be formulated. For example, some principles arise simply by stating the truth-conditions for each connective. The last entry on the table for $\supset$, above, corresponds to $J_{\mathrm{F}}(A) \rightarrow (J_{\mathrm{F}}(B) \rightarrow J_{\mathrm{U}}(A \supset B))$ for an axiomatic system or to $J_{\mathrm{F}}(A), J_{\mathrm{F}}(B) \vdash J_{\mathrm{U}}(A \supset B)$ for a sequent system.

3 Technical matters

All the metatheoretical concerns raised with classical logic also arise with respect to many-valued logics. The model theory of many-valued logics can be developed in much the same way as that of classical logic. Depending on the specifics of the syntax, proof theories are formulated in all the usual varieties: axiomatic, natural deduction, single- and multiple-conclusion consequence relations. Many-valued extensions of non-classical logics may be constructed, and many-valued versions of standard identity theory have been formulated.

Two frequently encountered questions warrant special mention, though we provide only partial answers here. (1) Given an arbitrary semantic theory, can we always find a corresponding proof theory? For the finite-valued case, the answer is affirmative, providing the syntax has appropriate connectives; but if the syntax is limited, it may be impossible to formulate an appropriate proof theory for a single-conclusion consequence relation. For the infinite-valued case, most interesting quantificational systems will not be axiomatizable; even for propositional logics it will be impossible to formulate a finitistic proof theory if the syntax is rich enough (for example, including autodescriptive operators). (2) Given a proof theory, can we always find a corresponding many-valued semantics? If we are concerned only with logical truth (not consequence in general) and if the proof theory admits a rule of substitution, then there will always be an infinite-valued semantics, though the designated values may turn out not to be decidable.

See also: Logical and mathematical terms, glossary of; Modal logic

CHARLES G. MORGAN

References and further reading

Belnap, N.D. (1977) ‘A Useful Four-Valued Logic’, in J.M. Dunn and G. Epstein (eds) *Modern Uses of Multiple-Valued Logic*, Boston, MA: Reidel.(Uncertainty due to conflicting opinion.)

Black, M. (1937) ‘Vagueness: An Exercise in Logical Analysis’, *Philosophy of Science* 427-55.(Logic of vagueness.)

Chang, C.C. (1965) ‘Infinite-Valued Logic as a Basis for Set Theory’, in Y. Bar-Hillel (ed.) *Logic, Methodology and Philosophy of Science*, Amsterdam: North Holland.(Paradoxes of set theory.)

Chang, C.C. and Keisler, H.J. (1966) *Continuous Model Theory*, Princeton, NJ: Princeton University Press. (Many-valued model theory.)

Dugundji, J. (1940) ‘Note on a Property of Matrices for Lewis and Langford’s Calculi of Propositions’, *Journal of Symbolic Logic* 5: 370-99.(Important modal systems do not have finite-valued semantics.)

Fitting, M.C. (1992) ‘Many-Valued Modal Logics’, parts 1 and 2, *Fundamenta Informaticae* 15: 235- 54, 17: 55-73.(Many-valued extensions of modal logics.)

Fraassen, B.C. van (1971) *Formal Semantics and Logic*, New York: Macmillan.(Finding a semantics for a given proof theory.)

Kleene, S.C. (1952) *Introduction to Metamathematics*, Princeton, NJ: Van Nostrand.(Uncertainty due to lack of termination of computation.)

Martin, R.L. (ed.) (1970) *The Paradox of the Liar*, New Haven, CT: Yale University Press.(On many-valued logics as a response to the liar paradox.)

McCall, S. (ed.) (1967) *Polish Logic 1920-1939*, London: Oxford University Press.(Source for papers of Łukasiewicz.)

Morgan, C.G. (1974) ‘A Theory of Equality for a Class of Many-Valued Logics’, *Zeitschrift für mathematische Logik* 20: 427-32.(Many-valued identify theory.)

Morgan, C.G. (1975) ‘Similarity as a Theory of Graded Equality for a Class of Many-Valued Logics’, in *Proceedings of the 1975 International Symposium of Multiple-Valued Logic*, Long Beach, CA: IEEE Society, 436-49. (Many-valued identity theory.)

Morgan, C.G. (1976) ‘Many-Valued Propositional Intuitionism’, in *Proceedings of the Sixth International Symposium on Multiple-Valued Logic*, Long Beach, CA: IEEE Society, 150-6.(Many-valued extension of

intuitionistic logic.)
Morgan, C.G. (1979) 'Local and Global Operators and Many-Valued Modal Logics', *Notre Dame Journal of Formal Logic* 20: 401-11.(Many-valued extensions of common modal logics.)
Morgan, C.G. and Pelletier, F.J. (1976) 'Some Notes concerning Fuzzy Logics', *Linguistics and Philosophy* 1: 79-97.(Finding proof theory for infinite-valued semantics.)
Putnam, H. (1957) 'Three-Valued Logic', *Philosophical Studies* 8: 73-80.(On uncertainty due to the incompleteness of human knowledge.)
Reichenbach, H. (1944) *Philosophic Foundations of Quantum Mechanics*, Berkeley, CA: University of California Press.(On uncertainty due to statistical fluctuations in nature.)
Rescher, N. (1969) *Many-Valued Logic*, New York: McGraw-Hill.(Survey of historically important systems.)
Rosser, J.B. and Turquette, A.R. (1952) *Many-Valued Logics*, Amsterdam: North Holland.(Finding proof theory for finite-valued semantics.)
Scarpellini, B. (1962) 'Die Nicht-Axiomatisierbarkeit des unendlichwertigen Prädikatenkalküls van Łukasiewicz', *Journal of Symbolic Logic* 27: 159-70.(Finding proof theory for infinite-valued semantics.)
Scott, D. (1974) 'Completeness and Axiomatizability in Many-Valued Logic', in *Proceedings of the Tarski Symposium, Proceedings of Symposia in Pure Mathematics*, vol. 25, Providence, RI: American Mathematical Society, 411-35.(On uncertainty due to accumulating error.)
Shoesmith, D.J. and Smiley, T.J. (1978) *Multiple-Conclusion Logic*, Cambridge: Cambridge University Press. (Proof theory for a given semantics.)
Thomason, S.K. (1978) 'Possible Worlds and Many Truth Values', *Studia Logica* 37: 195-204.(Many-valued extensions of indexical logics.)
Urquhart, A. (1986) 'Many-Valued Logic', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, vol. 3, Dordrecht: Reidel, 71-116.(Survey of important technical results.)

Many-valued logics, philosophical issues in

The first philosophically-motivated use of many-valued truth tables arose with Jan Łukasiewicz in the 1920s. What exercised Łukasiewicz was a worry that the principle of bivalence, 'every statement is either true or false', involves an undesirable commitment to fatalism. Should not statements about the future whose eventual truth or falsity depends on the actions of free agents be given some third status - 'indeterminate', say - as opposed to being (now) regarded as determinately true or determinately false? To implement this idea in the context of the language of sentential logic (with conjunction, disjunction, implication and negation), we need to show - if the usual style of treatment of such connectives in a bivalent setting is to be followed - how the status of a compound formula is determined by the status of its components.

Łukasiewicz's decision as to how the appropriate three-valued truth-functions should look is recorded in truth tables in which (determinate) truth and falsity are represented by '1' and '3' respectively, with '2' for indeterminacy (see tables in the main body of the entry). Consider the formula $A \vee B$ *('A or B'), for example, when A has the value 2 and B has the value 1. The value of* $A \vee B$ *is 1, reasonably enough, since if A's eventual truth or falsity depends on how people freely act, but B is determinately true already, then* $A \vee B$ *is already true independently of such free action. There are no constraints as to which values may be assigned to propositional variables. The law of excluded middle is invalidated in the case of indeterminacy: if p is assigned the value 2, then* $p \vee \neg p$ *also has the value 2. This reflects Łukasiewicz's idea that such disjunctions as 'Either I shall die in a plane crash on January 1, 2030 or I shall not die in a plane crash on January 1, 2030' should not be counted as logical truths, on pain of incurring the fatalistic commitments already alluded to.*

Together with the choice of designated elements (which play the role in determining validity played by truth in the bivalent setting), Łukasiewicz's tables constitute a (logical) matrix. An alternative three-element matrix, the 1-Kleene matrix, involves putting $2 \rightarrow 2 = 2$*, leaving everything else unchanged. And a third such matrix, the 1,2-Kleene matrix, differs from this in taking as designated the set of values* $\{1, 2\}$ *rather than* $\{1\}$*. The 1-Kleene matrix has been proposed for the semantics of vagueness. In the case of a sentence applying a vague predicate, such as 'young', to an individual, the idea is that if the individual is a borderline case of the predicate (not definitely young, and not definitely not young, to use our example) then the value 2 is appropriate, while 1 and 3 are reserved for definite truths and falsehoods, respectively. Łukasiewicz also explored, as a technical curiosity, n-valued tables constructed on the same model, for higher values of n, as well as certain infinitely many-valued tables. Variations on this theme have included acknowledging as many values as there are real numbers, with similar applications to vagueness and approximation in mind.*

1 Second thoughts on some popular applications

Motivated by a concern that the principle of bivalence ('every statement is either true or false') involves an undesirable commitment to fatalism, Łukasiewicz introduced truth tables involving the value 'indeterminate' in addition to 'true' and 'false' in the 1920s (see Łukasiewicz 1970). His decision as to how these three truth-values should be combined is shown in the following tables:

$\wedge$	1	2	3
*1	1	2	3
2	2	2	3
3	3	3	3

$\vee$	1	2	3
1	1	1	1
2	1	2	2
3	1	2	3

$\rightarrow$	1	2	3
1	1	2	3
2	1	1	2
3	1	1	1

A	$\neg A$
1	3
2	2
3	1

The asterisk flagging the first occurrence of '1' in the tables is the conventional way of indicating that this value is 'designated': in other words, that this value plays the role, in defining validity, played by truth in the bivalent setting.

The formal apparatus introduced by these tables is a special case of the notion of a matrix *M* with which are associated an algebra (a non-empty set - here, the set of matrix elements - closed under some specified operations), that we may call $A(M)$; and a set of matrix elements which are to count as the designated elements, which we will call $D(M)$. An alternative three-element matrix, the 1-Kleene matrix, differs from Łukasiewicz's only in having $2 \rightarrow 2 = 2$. A third such matrix, the 1,2-Kleene matrix, differs from this in taking as designated the set of values $\{1, 2\}$ rather than $\{1\}$.

One might expect the term 'matrix' to apply to the individual tables for the connectives depicted in such diagrams, but the standard logical usage, reflected in the definition just given, means that it is the whole set of tables that depicts a single matrix. If we call this matrix *M*, then the set of *M*-elements is $\{1, 2, 3\}$ and the algebra $A(M)$ has operations '$\wedge$', '$\vee$', '$\rightarrow$', '$\neg$', behaving as depicted in the tables. (When more care is called for, we distinguish these operations from the connectives, writing something like '$\wedge^M$', '$\vee^M$' and so on.) Finally, $D(M) = \{1\}$. A logic which is - to use terminology defined in §2 below - 'determined by' a matrix with *n* elements but by no matrix with fewer than *n* elements is called an *n*-valued logic.

From the table for '$\vee$' we can see that $1 \vee 2 = 1$. While this seems satisfactory (if *A* is true then $A \vee B$ will be true regardless of the value assigned to *B*), we can see that all is not well by looking at another case. The table for '$\wedge$' tells us that $2 \wedge 2 = 2$: a result which is suggested by the case in which the two conjuncts concerned are independent ('I shall die in a plane crash', 'Britain will become a republic'), as well as by the case in which the conjuncts are identical, but which goes obviously wrong otherwise ('I shall die in a plane crash', 'I shall not die in a plane crash' - to take an extreme case). The falsity of the conjunction 'I shall die in a plane crash and I shall not die in a plane crash' is fixed regardless of how anyone decides to act. Thus it should receive the value 3. But of course we cannot simply emend Łukasiewicz's matrix and insert a '3' in the middle box of the '$\wedge$' table, since we shall then have the wrong result for the other cases. Similar reasoning applies in the case of disjunction: the disjunction of 'I shall die in a plane crash' and its negation is not something whose eventual truth can be prevented, so it should have the value 1. So we cannot acquiesce in Łukasiewicz's $2 \vee 2 = 2$. The only conclusion possible is that treating the threefold classification Łukasiewicz had in mind as a range of values with respect to which the familiarly truth-functional connectives could indeed be taken as truth-functional is misconceived. Not every threefold classification of statements gives us a three-valued logic.

It seems appropriate to return a similarly negative verdict on the proposed application of 3- (and more) element matrices, including the 1-Kleene matrix (see Cleave 1974), to vagueness (see Vagueness §3). In these and other cases the truth-functional paradigm is simply inappropriate for the notions being modelled (Geach 1949). The problem of failed truth-functionality which afflicts these approaches arises out of the fact that there is a linear ordering of the truth-values with respect to which '$\wedge$' and '$\vee$' are interpreted as *min* and *max*. In the case of Łukasiewicz's three-valued tables - contrary to what is suggested by the labelling - we are taking 1 as the greatest and 3 as the least, with 2 intermediate. Not all many-valued tables exhibit this kind of linear ordering, however, including some that Łukasiewicz himself proposed for treating modal notions; typically, though, there are more suggestive ways of developing the semantics for such applications than by the laying down of such tables (see Prior 1955). We return in §4 to other applications; we shall also have occasion to observe that the above talk of rejecting the principle of bivalence turns out on reflection to be somewhat dubious - even apropos of certain more promising applications of matrix semantics than those touched on above.

2 Logic

Suppose we have a matrix *M* and a formal language whose primitive connectives are in one-to-one correspondence with the operations of $A(M)$. Then an assignment of elements of *M* to the formulas of that language, in which $A(M)$ is used to compute the value of compounds on the basis of their components, will be called an *M*-evaluation. This generalizes the idea of a 'valuation', which term we shall always reserve for the bivalent case (so a valuation assigns either T or F to every formula), and we wish now to adapt the familiar idea of validity as preservation of truth. The resulting notion is to apply to sequents, and we begin by distinguishing three conceptions of what a sequent is to look like - three (as we shall say) 'logical frameworks'.

The most general framework we call SET-SET, and it takes a sequent of a given language to be something of the form $\Gamma \succ \Delta$ (read 'Γ, therefore Δ' or 'Γ yields Δ'), where Γ and Δ are arbitrary sets of formulas of that language (Blamey 1986). If we impose the condition that Δ has to be $\{B\}$ for some formula *B*, we have the more

restricted logical framework SET-FMLA, and write $\Gamma \succ B$ rather than $\Gamma \succ \{B\}$. Finally, if we impose, in addition to this condition on Δ, the further condition that Γ must be empty, we have the framework we call FMLA, and when no confusion arises we identify the sequents $\succ B$ (that is, $\emptyset \succ B$) which this framework provides with the formulas *B* themselves.

Early approaches to logic (Frege, Hilbert) used the framework FMLA, syntactically codifying a logic with the axiomatic method. The provable formulas were formal analogues of the 'logical truths' of that area of natural language being formalized. The sequents of SET-FMLA are formal analogues of arguments with premises on the left of the '$\succ$', and the conclusion to the right. For SET-SET, we have a generalization in which, roughly speaking, we may say that sets of formulas on the right are understood disjunctively, just as those on the left are (as in SET-FMLA) understood conjunctively. More precisely, in the bivalent setting, we say that a sequent $\Gamma \succ \Delta$ of SET-SET holds on a valuation *v* (for the language in question) just in case if $v(A) = \mathrm{T}$ for every $A \in \Gamma$, then $v(B) = \mathrm{T}$ for some $B \in \Delta$. Since the other logical frameworks simply restrict the class of sequents, we may use this same definition in connection with them. A set of sequents which contains all substitution instances of any of its elements will be what we have in mind as a *logic* in that framework. The set of sequents determined by a class of valuations is the set comprising precisely those sequents holding on each valuation in the class. This set may or may not constitute a logic, on the definition just given, depending on what class of valuations is chosen.

It remains to provide a 'matrix semantics' version of the above. We simply make clear the intention that the designated values are those which play the role of truth. Suppose, then, that *h* is an *M*-evaluation, for some matrix *M*; we say that the sequent $\Gamma \succ \Delta$ holds on *h* just in case

If $h(\Gamma) \subseteq D(M)$ then $h(\Delta) \cap D(M) \neq \emptyset$,

where $h(\Gamma)$ denotes $\{h(A) |\ A \in \Gamma\}$. Again, this definition fixes what it is for a sequent of SET-FMLA or FMLA to hold on an evaluation. Finally, a sequent is *valid* in a matrix *M* if it holds on every *M*-evaluation. We will not use the term 'tautologous sequent' (or 'tautology', in the case of FMLA) except for the special case of validity in the familiar two-element matrix. For this special case, also, we do not need the distinction between valuations and evaluations. The set of sequents determined by a class of matrices comprises precisely those sequents valid in each matrix in the class: such a set will always be a logic (as defined above). A logic determined by the class $\{M\}$ will be said to be determined by the matrix *M*. Notice that the logic in FMLA determined by the 1-Kleene matrix (which differs only from Łukasiewicz in that $2 \to 2 = 2$) is just the empty set, the assignment of the value 2 to every formula being an evaluation for this matrix. Since Łukasiewicz worked in FMLA, this is one reason that in his tables we have $2 \to 2 = 1$ (rather than $= 2$). (The logics determined by the Kleene matrices in the other frameworks are of course non-empty.)

3 Logic (cont.)

We say a logic is 'many-valued in the narrow sense' when it is determined by some matrix, and 'many-valued in the broad sense' when it is determined by some class of matrices. In FMLA this is a distinction without a difference, an early result of Lindenbaum and Tarski having established that in this framework every logic is many-valued in the narrow sense. However, in the case of the other frameworks, the distinction is a real one.

In SET-FMLA, for example, every logic closed under certain structural rules ('overlap', 'dilution' and 'cut' (for sets) - see Shoesmith and Smiley 1978) is broadly many-valued. We call a relation '$\vdash$' between sets of formulas and individual formulas (of some given language) a consequence relation (on that language) when there is some collection Σ of sequents of SET-FMLA which is closed under the SET-FMLA structural rules and for which we have $\Gamma \vdash A$ if and only if the sequent $\Gamma \succ A$ belongs to Σ. (This means that '$\Gamma \succ A$' is treated as an alternative notation for $\langle \Gamma, A \rangle$.) Closure under these structural rules is a necessary condition for a logic to be determined by any class of valuations. These same rules must be satisfied by any logic determined by a class of matrices for the following reason. Given an *M*-evaluation *h* (for some matrix *M*), we define the (bivalent) valuation v_h induced by *h* thus: $v_h(A) = \mathrm{T}$ if and only if $h(A) \in D(M)$, for all formulas *A*. Then a sequent holds in the many-valued sense on *h* just in case it holds, in the bivalent sense, on v_h. Accordingly a set of sequents is determined (many-valued sense) by a class $\mathbb{C}$ of matrices precisely when it is determined (bivalent sense) by the class of valuations $\{v_h\colon h$ is an *M*-evaluation for some $M \in \mathbb{C}\}$. Thus the structural rules must also be satisfied by

any SET-FMLA logic which is many-valued in the broad sense. Now while a necessary and sufficient condition for a logic in SET-FMLA to be many-valued in the broad sense is that it should be a consequence relation, there is a further condition to be satisfied in the case of the narrow sense, which is called the cancellation condition in Shoesmith and Smiley (1978). (See also Wójcicki 1974.) This further condition is not satisfied in the case of the consequence relation determined by the pair comprising the 1-Kleene matrix and the 1,2-Kleene matrix, for example, which is not, therefore, a many-valued logic in the narrow sense.

Note that the narrow sense could be further narrowed to exclude determination by a two-element matrix: it is only on such a usage that Łukasiewicz could be counted a pioneer of many-valued logic, for example. A further narrowing in the 'opposite' direction comes by disallowing determination by an infinite matrix to suffice for many-valuedness. The motivation for excluding infinite matrices comes from more than one source. Philosophically, there may be a feeling that such a case involves too many distinctions for the values so distinguished to represent a plausible generalization of the idea of truth-values (Haack 1978: 214). Perhaps the assignment made by an evaluation in this case is better thought of as the assignment of a 'proposition' to a formula (see Dunn 1975). There is also a technical pressure tending towards excluding the infinite case, coming from the Lindenbaum-Tarski result mentioned above together with an early concentration on the framework FMLA: since *every* logic is many-valued in the absence of the restriction to finite determining matrices, we might impose this restriction to avoid making the term 'many-valued' vacuous. However, since (again) this threat of vacuity does not arise for other logical frameworks, we prefer not to narrow further the above narrow sense of 'many-valued'. It is better to have a separate word for the property of being determined by a finite matrix; the usual word is 'tabular'.

The first demonstration that a logic was not tabular came when Gödel (1932) addressed this question for intuitionistic sentential logic in FMLA. We can introduce the argument he gave for this conclusion by considering a well-known puzzle. Suppose you have a drawer containing (only) twenty socks, ten of which are blue and ten of which are green. You have to choose a pair of the same colour to wear, but because the room is dark, you cannot see the colours. The question is: what is the smallest number of socks you need to remove in order to be sure that among them there will be a matching pair? And the answer is: (not *eleven* but) *three*. The reason is that, since there are only two colours involved, three socks cannot all be of different colours, so you are guaranteed a matching pair. This is a special case of what is sometimes called the 'pigeonhole principle': if a number of pigeons are to be placed in a smaller number of pigeonholes, then at least one pigeonhole must have more than one pigeon in it.

With the example of the socks in mind, consider the question of whether or not the following formula is a (two-valued) tautology:

$Pigeon_3 \quad (p \leftrightarrow q) \vee (p \leftrightarrow r) \vee (q \leftrightarrow r).$

It should not be necessary to construct the eight-line truth table to answer this question. Since there are only two truth-values (T, F) to go round, no valuation can assign different truth-values to all three propositional variables *p*, *q* and *r*. (The variables here are playing the role of the pigeons - whence the labelling of the formula - with the truth-values being the pigeonholes.) So each valuation must assign the value T to at least one of the disjuncts of the above formula, thus assigning T to the whole disjunction, which is therefore a tautology.

We turn to Gödel's argument that intuitionistic logic in FMLA - which we shall denote by IL - is not determined by any finite matrix. Let CL denote classical sentential logic in FMLA (alias the set of tautologies in two-valued logic). To begin with, let us see how considerations such as those of the preceding paragraph reveal IL to be determined by no three-element matrix, when supplemented by three additional facts: (1) IL $\subseteq$ CL; (2) IL has the 'disjunction property': for all formulas *A*, *B*, if $A \vee B \in$ IL, then either $A \in$ IL or $B \in$ IL; and (3) if *C* is a disjunction one of whose disjuncts is a formula of the form $A \leftrightarrow A$, then $C \in$ IL. (We take '$\leftrightarrow$' as a primitive connective for simplicity here; in fact the whole argument goes through if this is replaced by '$\rightarrow$'.) Now consider the formula

$Pigeon_4 \quad (p \leftrightarrow q) \vee (p \leftrightarrow r) \vee (p \leftrightarrow s) \vee (q \leftrightarrow r) \vee (q \leftrightarrow s) \vee (r \leftrightarrow s).$

This formula is not provable in IL, in view of facts (1) and (2) above. Now suppose that some three-element matrix *M* determines IL. For any *M*-evaluation *h*, fact (3) and the pigeonhole principle imply that $Pigeon_4$ holds on *h*, since *h* cannot assign distinct *M*-elements to all four propositional variables. Thus, though not IL-provable, $Pigeon_4$ is *M*-valid: so IL is not determined by *M*. The same argument shows (by analogous consideration of the

formula we naturally call $Pigeon_n$) that, for any finite number *n*, IL is not determined by any matrix having *n* elements: IL is not tabular.

4 From evaluations to valuations

Our introduction, in §3, of the valuation v_h induced by a matrix evaluation *h*, serves as a reminder of the twofold division which is always present in the background of the many-valued setting, in the form of the ubiquitous designated/undesignated distinction. The notion of validity for sequents is always a matter of invariable possession or preservation of truth: of T as a value of the functions v_h. Distinctions within the classes of designated and undesignated matrix-values do not matter for validity, and enter the semantic account only because the connectives concerned cannot be seen as simply 'designation-functional': in the case of Łukasiewicz's and Kleene's three-valued tables, for example, while $h(\neg A)$ is uniquely determined by $h(A)$, there is no such unique determination of $v_h(\neg A)$ by $v_h(A)$. So we do our compositional semantics at the level of evaluations, and then modulate to valuations for talk of validity.

Here is an analogy: imagine that chemistry were concerned in the first instance with the colours of chemical compounds. We would like to know what colour a compound of substances of given colours will be. The trouble is that just knowing what colour the constituents are is not enough to tell us what colour the compound will be. We also need to know about their chemical composition, so that this extra information can be used to predict the eventual colour of the compound, via its chemical composition. Similarly, even if all we are interested in knowing about a composite formula *A* # *B* (where # is some binary connective) is whether or not it has a designated value on evaluation *h*, we cannot obtain this information - that is, the value of v_h(*A* # *B*) - just from knowing $v_h(A)$ and $v_h(B)$ (the 'colours' of *A* and *B*). Rather, we need to know $h(A)$ and $h(B)$, so that we can apply our matrix operation # to these values and obtain *h*(*A* # *B*), and then, eventually, v_h(*A* # *B*). This is precisely to acknowledge the possibility of a failure of what was called designation-functionality above.

But are the subdivisions effected by the apparatus of many-valued logic of merely instrumental interest? Such a suggestion has been elaborated over a long period by Michael Dummett, beginning with his article 'Truth' in 1959. The problem Dummett noted is that when we consider the point of the true/false dichotomy as applied to whole sentences (of an interpreted language such as English) which might be used to make assertions, there does not seem to be room for alleging any particular sentence to fall on neither side:

> We need to distinguish those states of affairs such that if the speaker envisaged them as possibilities he would be held to be either misusing the statement or misleading his hearers, and those of which this is not the case: and one way of using the words 'true' and 'false' would be to call states of affairs of the former kind those in which the statement was false, and the others those in which the statement was true.
>
> ([1959] 1978: 11-12)

Dummett went on to describe in the light of this criterion how particular cases in which statements have been described as neither true nor false should be classified as true or as false, depending on the nature of the case. The various values of many-valued matrices are then seen as subdivisions of the true (the designated values) and the false (the undesignated) - different ways of being true or of being false, the difference between which needs to be taken into account for settling how the fundamental binary true/false distinction is to apply to compounds into which they enter as components. We repeat one of Dummett's illustrations:

> I once imagined a case in which a language contained a negation operator '$-$' which functioned much like our negation save that it made '$-(A \rightarrow B)$' equivalent to '$A \rightarrow -B$', where '$\rightarrow$' is the ordinary two-valued implication. In this case, the truth or falsity of '$-(A \rightarrow B)$' would not depend solely on the truth or falsity of '$A \rightarrow B$', but on the particular way in which '$A \rightarrow B$' was true (whether by the truth of both constituents or by the falsity of the antecedent). This would involve the use of three-valued truth tables, distinguishing two kinds of truth. In the same way, it might be necessary to distinguish two kinds of falsity.
>
> ([1963] 1978: 155)

Indeed, Dummett (1959) suggested this last possibility as the appropriate gloss on the response of those who describe sentences containing non-denoting singular terms as 'neither true nor false'. The response is a natural one if one thinks of falsity of a sentence as the truth of its negation (supposing the identification of a sentence's

negation is not itself problematic), but since it would be equally erroneous ('misleading') to assert either the given afflicted sentence or its negation, in terms of the more fundamental true/false dichotomy, both sentences fall alike on the false side. The two undesignated values of Łukasiewicz's three-valued logic can be regarded in the same light: as cases in which a sentence having either value are not, the facts being what they are, fit for assertion. They differ in respect of whether the negation of the sentence in question suffers the same fate (according to Łukasiewicz's metaphysical assumptions, that is, and setting to one side the problems about truth-functionality mentioned in §1).

It is for reasons such as this that Dummett speaks of 'the legitimate but comparatively superficial character of the rejection of the law of bivalence involved in the use of a many-valued logic' (1973: 448). In the same work, Dummett draws a distinction between the 'assertoric content' and the 'ingredient sense' of a statement, introduced in terms of the distinction between knowing the meaning of a statement 'in the sense of grasping the content of an assertion of it' and 'in the sense of knowing the contribution it makes to determining the content of a complex statement in which it is a constituent' (1973: 447). Sentences (or statements) which are alike in respect of assertoric content are true (that is, have some designated value) under the same conditions, which does not imply that the replacement of one by the other in some complex sentence will secure a similar likeness for the two complex sentences produced. When the complex sentences differ in respect of designation in this way, the components in respect of which they differ themselves differ in ingredient sense, their similarity of assertoric content notwithstanding. Thus, in terms of our formal notation, and slightly oversimplifying, equality of ingredient sense, for *A* and *B*, corresponds to having $h(A) = h(B)$ for every matrix-evaluation *h*, while equality of assertoric content corresponds to having $v_h(A) = v_h(B)$ for each such evaluation *h*.

Dummett's rationale for many-valued (that is, *k*-valued for $k > 2$) logic, then, is that there is sometimes a need to record differences in ingredient sense among statements which do not differ in assertoric content. Without going into too much technical detail, though, we can see that this is not quite accurate. We call a consequence relation '$\vdash$' 'extensional' if for all formulas *A*, *B*, with $C(A)$ and $C(B)$ the results of putting formulas *A* and *B* into some context *C*, and for all sets Γ of formulas, we have

If $\Gamma,\ A \vdash B$ and $\Gamma,\ B \vdash A$, then $\Gamma,\ C(A) \vdash C(B)$;

and we call '$\vdash$' 'congruential' when the above condition is satisfied by all *A*, *B*, *C* for the special case in which $\Gamma = \emptyset$. Congruentiality is strictly weaker than extensionality (consider normal modal logics, for example) and, as the end of the last paragraph makes clear, the need to distinguish ingredient sense from assertive content arises precisely when the logic (in SET-FMLA, for expository convenience) is not congruential. Yet, as the reader may care to check, it is failures of the stronger condition, extensionality, rather than just failures of congruentiality, that force a matrix treatment to involve more than two values.

5 From evaluations to valuations (cont.)

We end by briefly taking up one further issue raised by the above exposition: the case of intuitionistic logic. A perusal of Dummett (1959) reveals that his talk (quoted above) of the 'superficial character of the rejection of the law of bivalence' in many-valued logic is intended to contrast with the 'deeper' character of such a rejection in the case of intuitionistic logic. One reason for not considering intuitionistic logic to be many-valued seems hardly to the point here: namely the fact that, as we saw in §3, intuitionistic logic in FMLA is not determined by any finite matrix. (Nobody questions the classification of Łukasiewicz's infinite-valued logic as many-valued, for example.) By the Lindenbaum-Tarski result cited in that section, this ceases to be the case if 'finite' is deleted. Turning to SET-FMLA, the consequence relation associated with intuitionistic logic satisfies the cancellation condition (alluded to in §3) and so here too there is a single determining matrix. On formal grounds, then, we have to say that intuitionistic logic in either of these frameworks is indeed many-valued in the narrow sense of the term. Cardinality of determining matrices aside, are there any good reasons for not regarding IL as many-valued? There is a sense of 'many-valued' which we have not touched on, which allows an affirmative answer to this question: but it has more to do with what its adherents say about a logic than with the logic itself. The usual sort of reason for denying that IL is many-valued is that it would be inconsistent, according to IL itself, for an intuitionist to claim of some given statement that it was neither true nor false - at least when the falsity of *A* is identified with the truth of $\neg A$. (This use of 'falsity' is of course not that focused on in the quotation from Dummett (1959).) For

such a claim would then be of the form $\neg(A \vee \neg A)$, or, IL-equivalently, of the form $\neg A \wedge \neg\neg A$: but the negation of anything of this form is itself IL-provable. So the claim is not one an intuitionist could consistently make. The trouble with this as a way of distinguishing IL from a paradigmatically many-valued logic such as Łukasiewicz's three-valued logic is that dangerously similar moves can be made in the context of the latter logic also. If, instead of adopting an external perspective and saying that *A* is a statement which is neither true nor false, we think of representing such a claim as the claim that for the given *A* (which concerns, say, someone's later free actions), $\neg(A \vee \neg A)$, then again we have the equivalence with $\neg A \wedge \neg\neg A$, and this time the further equivalence with $\neg A \wedge A$. And while the negation of the latter is not provable in Łukasiewicz's logic (so that here we have a disanalogy with the case of IL) it remains the case that nothing of that form can receive the designated value on any matrix evaluation: so it is presumably after all not the kind of claim anyone attracted to Łukasiewicz's revision of classical logic ought to be making. As in the case of IL, such a theorist is distinguished from an adherent of classical logic by refraining from making certain claims (for example, some having the form $A \vee \neg A$), rather than by trying to make contrary claims. Further work, then, is called for on the part of anyone wishing to maintain that in some philosophically significant sense Łukasiewicz's logic is many-valued while that of the intuitionists is not.

See also: Intuitionism; Logical and mathematical terms, glossary of; Łukasiewicz; Many-valued logics

LLOYD HUMBERSTONE

References and further reading

Blamey, S.R. (1986) 'Partial Logic', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, vol. 3, *Alternatives to Classical Logic*, Dordrecht: Reidel, 1-70.(The logic in SET-SET determined by the 1-Kleene and 1,2-Kleene matrices taken together is the $\{\wedge, \vee, \neg, \rightarrow\}$-fragment of the 'partial logic' presented here, although Blamey urges us to take the 'indeterminate' not as a truth-value on a par with the other two, but as representing undefinedness in respect of truth-value.)

Cleave, J.P. (1974) 'The Notion of Logical Consequence in the Logic of Inexact Predicates', *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 20: 307-24.(Proposes an application of the Kleene matrices to the logic of vagueness.)

Czelakowski, J. (1981) 'Equivalential Logics (I), (II)' *Studia Logica* 40: 227-36, 355-72.(Includes a discussion bearing on what we call 'extensionality' in §4.)

Dummett, M.A.E. (1959) 'Truth', *Proceedings of the Aristotelian Society* 59: 141-62; repr. with postscript (1972) in *Truth and Other Enigmas*, London: Duckworth, 1978.(Distinguishes deep from superficial grounds for rejecting the principle of bivalence.)

Dummett, M.A.E. (1963) 'Realism', in *Truth and Other Enigmas*, London: Duckworth, 1978, 145-65.(Further elaboration of the distinction between deep and superficial grounds for rejecting the principle of bivalence.)

Dummett, M.A.E. (1973) *Frege: Philosophy of Language*, London: Duckworth.(See the appendix to chapter 12 for a critical review of some central concepts of matrix semantics, using some traditional terminology not employed here.)

Dummett, M.A.E. (1991) *The Logical Basis of Metaphysics*, Cambridge, MA: Harvard University Press, esp. 48.(More on assertoric content versus ingredient sense.)

Dunn, J.M. (1975) 'Intensional Algebras', in A.R. Anderson and N.D. Belnap, *Entailment: The Logic of Relevance and Necessity*, vol. 1, Princeton, NJ: Princeton University Press, 1975, 180-206.(A gentle introduction to algebraic semantics.)

Fine, K. (1975) 'Vagueness, Truth, and Logic', *Synthese* 30: 265-300.(This paper, like Kamp's (1975), includes a good formulation of the 'failed truth-functionality' argument, as in §1 above, and an explanation of the appeal of a different treatment of vagueness, in terms of 'supervaluations'.)

Geach, P.T. (1949) 'If's and And's', *Analysis* 9: 58-62.(First published appearance of the 'failed truth-functionality' argument.)

Gödel, K. (1932) 'Zum intuitionistischen Aussagenkalkül', *Anzeiger der Akademie der Wissenschaften Wien, mathematisch-naturwissenschaftliche Klasse* 69: 65-6; trans. 'On the Intuitionistic Propositional Calculus', in *Kurt Gödel: Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay and J. van Heijenoort, Oxford: Oxford University Press, 1986, 223-5.(In this paper Gödel shows that intuitionistic logic in FMLA is not tabular. A by-product of the argument is a proof of the existence of an infinite sequence of distinct logics intermediate between classical and intuitionistic propositional logic.)

Haack, S. (1978) *Philosophy of Logics*, Cambridge: Cambridge University Press.(Chapter 11 provides a short

survey of some intended applications of many-valued logic.)
Kamp, H. (1975) 'Two Theories about Adjectives', in E.L. Keenan (ed.) *Formal Semantics of Natural Language*, Cambridge: Cambridge University Press, 123-55.(This paper, like Fine's (1975), includes a good formulation of the 'failed truth-functionality' argument, as in §1 above, and an explanation of the appeal of a different treatment of vagueness, in terms of 'supervaluations'.)
Łoś, J. and Suszko, R. (1958) 'Remarks on Sentential Logics', *Indagationes Mathematicae* 20: 177-83.(This paper includes an ancestor of the cancellation condition of Shoesmith and Smiley (1978).)
Łukasiewicz, J. (1970) *Jan Łukasiewicz: Selected Works*, ed. L. Borkowski, Amsterdam: North Holland.(See especially 'On Three-Valued Logic' (1920), 'On Determinism' (Łukasiewicz's inaugural address at Warsaw University; 1922) and 'Philosophical Remarks on Many-Valued Systems of Propositional Logic' (1930).)
Prior, A.N. (1955) 'Many-Valued and Modal Systems: An Intuitive Approach', *Philosophical Review* 64: 626-30.(Shows how to relate certain many-valued matrices to the possible world's semantics for modal logic later popularized by Kripke.)
Rescher, N. (1969) *Many-Valued Logic*, New York: McGraw-Hill.(Provides extensive historical information on the subject, including more on the Kleene matrices, which are versions of Kleene's 'strong' three-valued truth tables, the 'weak' ones being associated with Bochvar.)
Scott, D.S. (1974) 'Completeness and Axiomatizability in Many-Valued Logic', in L. Henkin *et al*. (eds) *Proceedings of the Tarski Symposium*, Providence, RI: American Mathematical Society.(This work provides a useful introduction to semantics in terms of valuations, and proposes a treatment of many-valued logic in which the 'many values' are replaced by 'many valuations'; the logical framework is a finitized version of SET-SET.)
Scott, D.S. (1976) 'Does Many-Valued Logic Have Any Use?', in S. Körner (ed.) *Philosophy of Logic*, Oxford: Blackwell, 64-74.(A less technical discussion of the philosophical underpinnings of Scott 1974.)
Seeskin, K.R. (1971) 'Many-Valued Logic and Future Contingencies', *Logique et Analyse* 14: 759-73.(Includes the 'failed truth-functionality argument' mentioned in §1.)
Seeskin, K.R. (1974) 'Some Remarks on Truth and Bivalence', *Logique et Analyse* 17: 101-9.(Among other things, the author notes some features of Łukasiewicz's three-valued logic which are anomalous, given his informal motivation. These reflect what Blamey (1986) calls non-monotonicity (in the table for implication).)
Shoesmith, D.J. and Smiley, T.J. (1978) *Multiple-Conclusion Logic*, Cambridge: Cambridge University Press.(Includes a general discussion of many-valued logic and the theory of consequence relations. The cancellation conditions for SET-FMLA and SET-SET appear in chapter 15.)
Smiley, T.J. (1976) Comment on Scott (1976), in S. Körner (ed.) *Philosophy of Logic*, Oxford: Blackwell, 74-88.(This deals with the 'failed truth-functionality' point from §1.)
Urquhart, A. (1986) 'Many-Valued Logic', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, vol. 3, *Alternatives to Classical Logic*, Dordrecht: Reidel, 71-116.(Good for information about certain aspects of many-valued logic, such as fundamental completeness, but makes the mistake of claiming that there is no distinction between - as we would put it - the broad and narrow senses of 'many-valued' in the framework SET-SET. See Wroński 1987 for correction.)
Wó jcicki, R. (1974) 'Note on Deducibility and Many-Valuedness', *Journal of Symbolic Logic* 39: 563-6.(Traces the history of the Shoesmith-Smiley cancellation principle.)
Wó jcicki, R. (1988) *Theory of Logical Calculi*, Dordrecht: Kluwer.(A comprehensive survey of matrix methodology. Congruentiality, mentioned in §4, is treated under the name 'self-extensionality'.)
Wroński, A. (1987) 'Remarks on a Survey Article on Many Valued Logic by A. Urquhart', *Studia Logica* 46: 275-8.(Corrects a serious error in Urquhart 1986.)
Wroński, A. (1987) 'Remarks on a Survey Article on Many Valued Logic by A. Urquhart', *Studia Logica* 46: 275-8.(Corrects a serious error in Urquhart (1986).)

Marcel, Gabriel (1889-1973)

Marcel was a distinguished French playwright and music critic as well as philosopher. It was he who coined the term 'existentialism', although he was reluctant to be pigeon-holed a 'Christian existentialist'. Born into a well-off family of civil servants, Marcel - never a healthy man - worked for the Red Cross during the First World War, an experience which shaped his view of human relationships and confirmed a religious conviction that led to conversion to Roman Catholicism in 1929. After an early flirtation with F.H. Bradley's idealism, Marcel independently developed a phenomenology of human existence and a religious conception of being similar, in several respects, to those of Karl Jaspers and Martin Buber. He was much in demand as a lecturer in his later years.

Marcel was an idiosyncratic writer. Disavowing any 'system', his early writings were ruminative diaries, while his later lectures largely consisted in 'orchestrating', as he put it, the themes tentatively announced in those diaries. The personal anecdotes, of his wartime work for example, testified to his belief that reflection on certain experiences - of 'homelessness' and 'hope', for instance - is a key to understanding our place in the universe. Indeed, they reflect his sense, shared with other existentialists, of 'the primacy of being over knowledge' ([1933] 1949: 8), and his denunciation of 'the execrable habit of considering problems…in abstraction from…the very texture of life' ([1935] 1949: 102). Marcel describes his philosophy as a 'concrete' one, and as a true empiricism distinguished by a refusal to follow most empiricists in their myopic focus on *perceptual* experience.

Marcel's ambition was 'to restore its ontological weight to human experience' ([1935] 1949: 103). Restoration is required because human beings now live in - to cite the title of his best-known play - 'a broken world', where they have been reduced to mere 'functions' by the levelling processes of modern technology and politics. As such, while people feel alienated or homeless, they have all but lost the capacity to reflect on their existence and its relation to being at large. A main reason is the predominance of scientific thinking, which reduces all questions to 'problems' with 'technical' solutions, thereby obscuring the sphere of the 'metaproblematic' or 'mysterious'. A 'mystery' is something I am so 'caught up' in that objective analysis, which presupposes a sharp distinction between myself and what is 'in front of me', is inappropriate. For being is precisely what 'withstands…exhaustive analysis..[into] elements…devoid of intrinsic significance' ([1933] 1949: 5), so scientific thought is hostile to ontology, to inquiry into being.

A main culprit behind the hegemony of problem-centred thinking has been philosophies, like Cartesianism, which portray the self as an inner, closed-in sphere 'spectating' a world outside it. This has encouraged the views that the mind is distinct from the body, that we perceive internal data which 'translate' external stimuli, and that 'other minds' need to be inferred from perceptual data. The general effect is to promote the image of myself standing to the world as a 'subject' to an array of 'objects'. Against this, Marcel insists that 'I am not watching a show', that 'ours is a being whose…essence is to be in every way *involved*' ([1935] 1949: 21, 116-17). I am not distinct from my body as a participating agent in the world. Perception is not awareness of inner data, but active 'reception' of the world, a 'power of opening oneself' to it ([1940] 1964: 29). And in relations like love, other persons are directly 'present' to me, not phantoms whose existence I infer from their behaviour. Marcel's is 'a metaphysic of *we are* as opposed to a metaphysic of *I think*' ([1950b] 1951: 9).

While Cartesianism and its relatives are philosophical errors, it is depressingly familiar in our 'broken world' to live as if they were true, like the solipsistic figures in Proust's later volumes. Marcel's task is to summon us to a 'secondary reflection' which 'reconquers [our] unity' ([1950a] 1950: 103) with the world and each other, one broken by the 'primary reflection' of analytical science, and thereby to free us from the 'sclerosis' which confines us within narrow 'concentric zones'. In particular, he summons people to be 'available' (*disponible*), a stance best manifested in absolute 'fidelity' or commitment to others. Far from sacrificing individuality and freedom through being 'available' to others, this is the only way to achieve them. Not only can I 'really know myself only when I have committed myself' to others ([1940] 1964: 163), but it is only through 'collaboration with his freedom' - through our mutual acknowledgement of one another as 'Thou' - that I myself attain freedom ([1935] 1949: 107). (Marcel, predictably, is acutely critical of Sartre's bleak and agonistic portrait of human relations in *Being and Nothingness*).

Marcel indicates, however, a potential paradox in the making of absolute commitments to others. Given the uncertainty of the future, how for example can a man, unless naïve or disingenuous, commit himself to a lasting loving relationship with his wife? The solution is that he must place hope and faith, if not precisely in God 'as someone other', then at least in a mysterious and co-operative process of being. Thus mundane fidelities only 'become possible…with that absolute fidelity which [is] faith' ([1940] 1964: 167). Indeed, in the intimations of hope and confidence for the future resides the best testimony to a divine, mysterious order. The traditional proofs for God's existence are of little interest to Marcel, since even if they cannot be logically faulted, they clearly have no power to convince people not already predisposed to belief. He remains hesitant, however, over the need for specifically Christian belief for the possibility of 'availability' and 'fidelity'. On the one hand, this does not logically 'presuppose the data of Christianity'; on the other, we are the legatees of a rich Christian tradition which may, as a matter of fact, provide us moderns with our only viable resources for full and undespairing participation in the world and the lives of our fellows ([1933] 1949: 29f).

See also: Buber, M.; Existentialism; Existentialist ethics; Existentialist theology; Jaspers, K., Sartre, J.-P.

DAVID E. COOPER

List of works

Marcel, G. (1927) *Journal métaphysique*, Paris: Gallimard; trans. B. Wall, *Metaphysical Journal*, Chicago, IL: Regnery, 1952.(A diary of early reflections on Marcel's staple themes of mystery and human relationships.)

Marcel, G. (1933) 'On the Ontological Mystery', trans. M. Harari in *The Philosophy of Existence*, London: Harvill, 1949.(Seminal essay on the distinction between the 'problems' addressed by science and the 'mysteries' explored by existential philosophy. This book also contains Marcel's critique of Sartre, 'Existence and Human Freedom'. The essays in this book are arguably the best introduction to Marcel's thinking.)

Marcel, G. (1935) *Être et avoir: philosophie de l'esprit*, Paris: Aubier; trans. K. Farrer, *Being and Having: An Existentialist Diary*, London: Dacre, 1949.(Like the *Metaphysical Journal*, a loose collection of reflections on human experience and relations.)

Marcel, G. (1940) *Du refus à l'invocation*, Paris: Gallimard; trans. R. Rosthal, *Creative Fidelity*, New York: Noonday, 1964.(Marcel's most thorough discussion of the theme of 'availability'.)

Marcel, G. (1945) *Homo Viator: prolégomènes à une métaphysique de l'espérance*, Paris: Aubier; trans. E. Crauford, *Homo Viator: Introduction to a Metaphysic of Hope*, New York: Harper & Row, 3rd edn, 1965.(Focuses on the intimations of the eternal and the divine which are to be found in our everyday lives.)

Marcel, G. (1950a) *Le Mystère de l'être*, Paris: Aubier, 2 vols; trans. G. Fraser, *The Mystery of Being*, vol. 1, *Reflection and Mystery*, London: Harvill, 1950.(The two volumes of *The Mystery of Being* are Marcel's Gifford Lectures of 1949-50 and are the most complete representation of his increasingly religious concerns.)

Marcel, G. (1950b) *The Mystery of Being*, vol. 2, *Faith and Reality*, trans. R. Hague, London: Harvill, 1951.(The two volumes of *The Mystery of Being* are Marcel's Gifford Lectures of 1949-50 and are the most complete representation of his increasingly religious concerns.)

Marcel, G. (1963) *The Existential Background of Human Dignity*, Cambridge, MA: Harvard University Press. (Marcel's William James Lectures at Harvard in 1961-2; the most socially and politically aware of Marcel's works.)

References and further reading

Gallagher, K.T. (1962) *The Philosophy of Gabriel Marcel*, with a Foreword by G. Marcel. New York: Fordham University Press, 3rd edn, 1975.(General and lucid introduction to Marcel's thought.)

McCown, J. (1978) *Availability: Gabriel Marcel and the Phenomenology of Human Openness*, Missoula, MT: Scholars Press.(Short and rewarding study of Marcel on human relationships and the divine.)

Troisfontaines, R. (1953) *De l'existence a être: la philosophie de Gabriel Marcel*, 2 vols, Paris: Vrin.(Gigantic study of Marcel by a sympathetic fellow Catholic.)

Marcus Aurelius (AD 121-80)

Marcus Aurelius Antoninus, Emperor of Rome, was the author of a book of philosophical reflections written in Greek and known as the Meditations. These reflections are based primarily on Stoicism, but also reveal the influence of other currents of thought and of his experience as emperor. Marcus was deeply influenced by Epictetus and shares his interest in the inner mental life and the psychology of moral improvement. He combines a deep commitment to the providential cosmology traditional in the Stoic school with a more pronounced religious sensibility and a frequent emphasis on the insignificance of human life in space and time. The Stoic recognition of the irreducibly social character of human nature is obviously pertinent to an emperor whose career consisted largely of self-sacrificing public service.

1 Life

Marcus was born in Rome in AD 121, was adopted by the future emperor Antoninus Pius in 138, and succeeded to the throne (jointly with Lucius Verus) in 161. In 169 Verus died, leaving Marcus sole emperor until 177 when his son Commodus joined him. Marcus died on campaign.

Marcus was the last in a series of efficient and humane emperors chosen by adoption whose reigns made the second century AD the high point of Roman imperial culture. The stability of this culture was threatened, however, by military challenges, and Marcus' reign was beset by wars on its borders. (It was during these campaigns, late in his life, that Marcus composed the notebooks which were published after his death.) The reign of his son Commodus was unsuccessful; his assassination in 192 ushered in a period of uncertainty which ended with the consolidation of the rule of Septimius Severus in 197. It is tempting to regard Marcus' reign as the acme of the Roman system and to assign Marcus much of the credit. But the seeds of decay were already present and Marcus did little to prevent the decline.

As a member of the ruling elite, Marcus was educated in Latin oratory and in law, but he also had Greek tutors, including the 'sophist' Herodes Atticus. Marcus' mentor in Latin was Marcus Cornelius Fronto, one of the most famous literary figures in Rome. Fronto's influence on Marcus was personal as well as literary and the correspondence (in Latin) between them rounds out the picture of Marcus which we get from his philosophical writings.

To the disappointment of Fronto, Marcus chose the study of philosophy (which he had begun under the Stoic Apollonius) over law and oratory. The catalyst for this choice was Quintus Junius Rusticus, who not only provided a model of philosophical principle applied in public life, but also introduced Marcus to the *Discourses* of Epictetus (*Meditations* I 7). Marcus' personal circle included philosophers of various schools, some of them active in the administration of the empire. He describes Claudius Severus, a Peripatetic, as 'brother' and credits him with inspiration in the area of political theory and practice; Severus introduced Marcus to the ideas of Thrasea Paetus and other figures of the philosophical opposition to Nero (I 14). Severus may be responsible for Marcus' acceptance of the un-Stoic denial of the equality of moral errors (II 10) (see Stoicism). Another influence was Marcus' friend Claudius Maximus, a political and military leader as well as a (possibly Peripatetic) philosopher (I 15); yet another friend, Cinna Catulus, was probably a Stoic (I 13). A certain Alexander, a Platonist, was given a key political post (Greek secretary) at a time of crisis (I 12). Perhaps the best insight into Marcus' philosophical formation comes from reflection on his relationship with Sextus of Chaeronea, a professional philosopher and nephew of Plutarch of Chaeronea. Marcus continued to attend Sextus' lectures even while emperor; in the *Meditations* (I 9.1) Marcus says that he learned from this Platonist the meaning of 'life according to nature' - the defining slogan of the Stoic school (see Stoicism §17). This same philosophical breadth is reflected in Marcus' establishment of the first ever chairs of philosophy. In AD 176 he endowed four chairs at Athens, one for each of the recognized major schools - Platonism, Aristotelianism, Stoicism and Epicureanism.

2 The *Meditations*

The *Meditations* reflects this lack of school dogmatism. While Stoicism provides the framework, Platonic and other ideas are also accommodated. Marcus' philosophical importance does not lie in the preservation or development of the Stoic system but rather in his demonstration of its adaptability to various circumstances and

philosophical temperaments.

The framework for Marcus' thought is providential cosmology. The cosmos is an organic whole; each individual is not just a part of that whole, but a limb or organ of the living universe (VII 13). Our well-being is inseparable from that of the whole; he frequently calls humans mere 'fragments' of it. We are similarly related to human society: our private benefit can never conflict with that of the collectivity. Our human nature is fundamentally social; personal fulfilment can never be achieved at the cost of the welfare of the whole. This idea is made into a test of what counts as harm to an individual: 'what is not harmful to the city does not harm the citizen either' (V 22). Civic welfare subsumes, but does not negate, that of the individual. Marcus emphasizes the Stoic claim that whatever happens by nature is in the best interests of each rational agent, that nothing genuinely bad can befall anyone except by way of their own failings. The conjunction of providential determinism with an emphasis on our role as parts of an organic whole brings Marcus close to the *amor fati* ('love of fate') of which Stoics are often accused.

Marcus' conception of the true interests of each person rests on the Socratic idea that nothing bad can happen to any good person. Also Socratic are his determination to teach, not punish, the wrong-doer (everyone should be forgiven, since all are aiming at their own good - XI 16) and the notion that a happy life is based on critical examination of one's convictions. This reveals, he thinks, that nothing unexpected or unbearable ever happens to a rational person (V 18). Like Epictetus, he emphasizes the role of mental reservation (IV 1, V 20), the analysis of impressions, and impersonal detachment; the one thing which ultimately matters is our mental life, our desires, beliefs and convictions.

If the cosmos subsumes individual well-being, then each of us is transient, important only for the rationality we embody, a rationality identified with an indwelling divinity. Hence Marcus emphasizes the smallness of human life, that we are a mere point in space and time (for example, XII 32). The inevitabilty of material change and the cyclical certainty of birth and death are themes to which Marcus returns repeatedly, often drawn from Heraclitus imagery; nothing as predictable as our own deaths can be regarded as important. Hence there are no surprises for anyone who has spent even half a lifetime in critical reflection on the world (XI 1.2); the most important thing is to structure one's beliefs around the rationally inevitable. Another link between cosmology and ethics emerges from Marcus' repeated reflection on Epicurean atomism - a system opposed to the providential cosmology shared by Platonists and Stoics (see X 6). He states the case tersely at XI 18.1: 'If not atoms, then there is a nature which organizes the universe; and if this is the case, then the inferior exist for the sake of the superior, and the superior [that is, rational beings] for the sake of each other.'

Marcus' views on the composition of human beings can be singled out as an illustration of his openness to non-Stoic influences. We are made up of three components (XII 3): body, 'spirit' or *pneumation*, and intelligence (Nous, identifiable with an internal divinity); although we must care for all three aspects of ourselves, only the third is really our own. The separation of *pneuma*, the stuff of the entire soul according to traditional Stoic theory (see Pneuma), from our rationality is a mark of Platonic influence; this *pneuma* is associated by Marcus with automatic, puppet-like responses. If our reason is separable from the rest of our soul and from our body, then our personal identity has been detached from our empirical selves and the early Stoic unity of the person has been abandoned. Hence Marcus says: 'Wipe out impression; stop your impulse; extinguish desire; keep your commanding-faculty (*hēgemonikon*) in your own control' (IX 7). The similarity to Epictetus is limited. For Marcus, to set impulse, impression and desire in opposition to the commanding-faculty appears as a step on the road to Neoplatonism (see Neoplatonism).

But Marcus never abandons the foundations of Stoic physical theory. He repeatedly claims that everything is either 'material' or 'causal' (V 13, VII 29, VIII 11, IX 37, XII 29). Though the terminology is new (it is also reflected in Seneca, *Letters* 65.2), this is the same physical dualism which Zeno of Citium and Chrysippus employed when they divided the universe into the active and passive principles and identified the former with a rational divinity (see Stoicism §§3, 5). Building on this, the founders of Stoicism had outlined a cosmology in which divine guidance and rational coherence were integral parts of the natural world. Despite the appeal of Peripatetic and especially Platonic thought, Marcus retained a Stoic commitment to a unified and physicalist conception of the natural world. Marcus the emperor clung tenaciously to the political and military stability of the second century AD, but he was the last to do so. Similarly, Marcus the philosopher clung to his Stoic tradition.

BRAD INWOOD

List of works

Marcus Aurelius (AD 121-80) Meditations in A.S.L. Farquharson, *The Meditations of the Emperor Marcus Antoninus*, Oxford: Clarendon Press, 1944; trans. *The Meditations of Marcus Aurelius Antoninus*, Oxford: Oxford University Press, 1990.(Text, translation and commentary.)

References and further reading

Arnold, E.V. (1911) *Roman Stoicism*, Cambridge: Cambridge University Press. (Standard work, somewhat dated but still useful.)

Asmis, E. (1989) 'The Stoicism of Marcus Aurelius', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36 3: 2,228-52. (A thorough survey of influences on Marcus, both personal and philosophical; emphasis on his openness to other schools and the impact of his own moral reflection.)

Birley, A. (1987) *Marcus Aurelius: A Biography*, revised edn, New Haven, CT and London: Yale University Press.(Reliable biography and history.)

Fronto (*c*. AD 95-166) *The Correspondence of Marcus Cornelius Fronto*, trans. C.R. Haines, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 2 vols, 1919-20.(Latin text with English translation of the letters between Fronto and various friends and figures at court, including Marcus both before and after he became emperor.)

Rutherford, R.B. (1989) *The Meditations of Marcus Aurelius: A Study*, Oxford: Clarendon Press.(Thoughtful literary study of the *Meditations*.)

Marcuse, Herbert (1898-1979)

Herbert Marcuse endured a brief moment of notoriety in the 1960s, when his best-known book, One-Dimensional Man (1964), was taken up by the mass media as the Bible of the student revolts which shook most Western countries in that decade. Though Marcuse's actual political influence was uneven, his public image was not wholly misleading. On the one hand, he popularized the critique of post-war capitalism that he, with the other theorists of the Frankfurt School, had helped develop: the Western liberal democracies were, they argued, 'totally administered societies' permeated by the values of consumerism, in which the manufacture and satisfaction of 'false needs' served to prevent the working class from gaining any genuine insight into their situation. On the other hand, Marcuse never fully subscribed to the highly pessimistic version of Marxism developed by the central figures of the Frankfurt School, Adorno and Horkheimer. He hoped that revolts by an underclass of 'the outcasts and the outsiders, the exploited and persecuted of other races and other colours, the unemployed and unemployable' would stimulate a broader social transformation. Underlying this affirmation of revolutionary possibilities was a conception of Being as a state of rest in which all conflicts are overcome, where rational thought and sensual gratification are no longer at war with one another, and work merges into play. Intimations of this condition - which could only be fully realized after the overthrow of capitalism (and perhaps not even then) - were, Marcuse believed, offered in art, 'the possible Form of a free society'. Imagination could thus show politics the way.

1 Life

Marcuse was born in Berlin. Like many young intellectuals of his generation, he was radicalized by his experience of the First World War, and, as a soldier stationed in Berlin, took part in the Revolution of November 1918. Though he studied Marxism during and after the war (when he went to Freiburg University), Marcuse was never involved in the Communist movement. After reading *Being and Time* soon after its publication in 1927 Marcuse returned to Freiburg to work under Martin Heidegger. His early philosophical writings represented a remarkable attempt, as he later put it, 'to combine existentialism and Marxism', though Marcuse later turned against what he came to think of as the 'false concreteness' of Heidegger's philosophy (see Existentialism). The main work of this period was Marcuse's *Habilitationsschrift*, the dissertation submitted to obtain the right to teach in German universities. Originally called *Hegel's Ontology and the Foundations of the Theory of Historicity* and completed in 1932, this text was (accounts differ) either never formally submitted or rejected by Heidegger on political grounds.

Marcuse left Germany in December 1932, a few weeks before Hitler came to power. Thanks to the help of, among others, Edmund Husserl (another of his teachers at Freiburg), he was invited to join the Institute for Social Research, which had left its base at Frankfurt University to take refuge in Geneva. Marcuse followed the Institute to Paris and then, in July 1934, to New York, and became one of its most active theorists, publishing a series of major essays in its journal, the *Zeitschrift für Sozialforschung*. Marcuse's active collaboration in the collective work of the Frankfurt School culminated in his second major study of Hegel, *Reason and Revolution* (1941).

Somewhat ironically, in the light of his later notoriety, Marcuse worked in the Office of Strategic Services, the predecessor of the Central Intelligence Agency, between 1942 and 1950. Unlike Horkheimer and Adorno, he did not return to Frankfurt after the Second World War, but became a US citizen. After holding research posts at Columbia and Harvard, Marcuse taught at Brandeis University from 1954 till his retirement in 1965. During these years he wrote *Soviet Marxism* (1958) and his two most influential books, *Eros and Civilization* (1955) and *One-Dimensional Man* (1964). Between 1965 and 1970 Marcuse held a series of temporary appointments at the University of California at San Diego. It was then that he became an international figure, thanks to his depiction in the mass media as the intellectual author of the radical student movement of the late 1960s. Though this was grossly exaggerated (two leaders of the French student revolt in May 1968 declared, 'None of us have read Marcuse'), Marcuse did make clear his sympathy with the students. In this respect he was quite different from Horkheimer and Adorno, who became increasingly conservative as they grew older. Marcuse's most important later works were perhaps *An Essay on Liberation* (1969) and *The Aesthetic Dimension* (1978). Taken ill during a lecture tour of Germany, he died at Starnberg.

2 Eros and capitalism

Marcuse's starting point, like that of the other leading figures in the Frankfurt School, was provided by the two decisive moves taken by Georg Lukács in *History and Class Consciousness* (1923). First, he developed Marx's theory of commodity fetishism into a general critique of reification (see Marx, K.). Capitalist society, Lukács claimed, seemed to its participants to be a collection of fragments; while individuals might adopt the instrumentally rational means to achieve their particular ends, they lacked any well-founded understanding of the social whole. Second, Lukács identified a unique vantage-point from which such understanding could be gained - that of the proletariat, conceived less as an empirical social class than as the 'identical subject-object' of history, able because of its reduction to the status of a commodity to grasp the meaning of the whole as it strove towards the revolutionary overthrow of capitalism.

The Frankfurt School was enormously influenced by Lukács's theory of reification, but found the idea of the proletariat as a stand-in for the Hegelian Absolute quite implausible. The result was the radical pessimism of Horkheimer's and Adorno's *Dialectic of Enlightenment* (1947), where liberal-democratic capitalism is depicted as a seamless web of reification, as implicitly totalitarian as its fascist and Stalinist opponents were openly, and the rise of instrumental reason is traced back to the first efforts of human beings to control nature. Reason is inextricably implicated in the domination of both nature and human beings. The task of critical theory (as the Frankfurt School tended to call their version of Marxism) is henceforth a negative one, endlessly to expose the unreconciled and unjust character of prevailing society, rather than to identify and promote tendencies leading to social transformation.

Marcuse managed to be part of this evolution, yet remain apart from it. On the one hand, *One-Dimensional Man* is in many ways a popularized and updated *Dialectic of Enlightenment*. Post-war capitalism, Marcuse argues, has overcome any tendency towards economic crisis. The attainment of abundance, however, leads not to liberation but to the generation of 'false needs', whose satisfaction leads to 'euphoria in happiness' and prevents individuals from making their decisions autonomously. The working class is in consequence unlikely to play the role assigned it by classical Marxism as the maker of socialist revolution, and has instead become part of the 'conservative social base' of capitalism. These changes are aspects of an emerging 'pattern of *one-dimensional behaviour* in which ideas, aspirations, and objectives that, by their content, transcend the established universe of discourse and action are either repelled or reduced to terms of this action' ([1964] 1968: 27). Critical theory is thereby forced merely to negate existing society. 'In the absence of demonstrable agents and agencies of social change, the critique is thus thrown back to a high level of abstraction' and 'it cannot be positive' ([1964] 1968: 11, 199).

None the less, as Jürgen Habermas observes, there is an 'affirmative feature of Herbert Marcuse's negative thinking' (Habermas 1988). This is perhaps most evident in Marcuse's most original book, *Eros and Civilization* (1955). He seeks to fuse Marx and Freud, drawing especially on the latter's speculative metapsychological writings. Marcuse attempts to diminish the conservative force of Freud's claim that instinctual repression is a constant of human nature dictated by the need for self-preservation. Instead, he argues that the 'performance principle' requiring the denial of sensual gratification is historically variable, changing according to the needs of different societies. Where class domination exists, surplus-repression arises - that is, repression that is required, not for self-preservation, but in the interests of the ruling class. Domination thereby roots itself deep within the unconscious. Consequently, 'from the slave revolts in the ancient world to the socialist revolution, the struggle of the oppressed has ended in establishing a new, "better" system of domination'. Nevertheless, 'the very progress of civilization under the performance principle has attained a level of productivity at which the social demands upon instinctual energy to be spent in alienated labour could be considerably reduced' ([1955] 1969: 75, 101). This offers the prospect of a social revolution which, by abolishing surplus-repression, would release considerable libidinal energies. Marcuse envisages a society in which the conflict between Freud's life and death instincts would be overcome, as they found common ground in an escape from suffering and repression, in a state of calm and repose which would represent a renunciation of what Marx called the realm of necessity, of material toil, a state of affairs in which, as Schiller had anticipated, work would be transformed into play, into 'a productivity that is sensuousness, play and song'.

3 Art and Being

Marcuse's philosophy of history is thus one in which 'the restless labour of the transcending subject terminates in the ultimate unity of subject and object: the idea of "being-in-and-for-itself", existing in its own fulfilment'

([1955] 1969: 88). The reconciling totality which Marcuse posits at the end of history is not, however, Lukács's identical subject-object. 'Being is experienced as gratification,' Marcuse claims of the future society, 'so that the fulfilment of man is at the same time the fulfilment, without violence, of nature' ([1955] 1969: 122-3). The 'ultimate unity of subject and object' is thus one that harmonizes humankind with nature. Marcuse here remains faithful to the Heideggerian interpretation of Hegel which he developed in Freiburg at the end of the 1920s. In *Hegel's Ontology* and related writings he had argued that the most basic category of the Hegelian dialectic is that of Life, conceived as the 'unifying unity of subjectivity and objectivity as a "subjective substance"' ([1932] 1987: 155) - that is, nature must be thought of, not as an aggregate of static objects governed by mechanical laws, but as a single process of continual self-transformation which manages through all these changes, and of which consciousness is simply the most developed form.

Whatever we make of this interpretation of Hegel, which shows the influence of Dilthey as well as Heidegger on the young Marcuse, its appeal to the idea of Life in order to conceptualize the unity of subject and object provides his later writings with their implicit structure and thus helps to explain why he remained, in Martin Jay's words, 'a prophet of identity and reconciliation' (1973). Marcuse believed that contemporary society offered, at its fault-lines, glimpses of 'the repressed harmony of sensuousness and reason'; the 'new sensibility' he discerned emerging in the counter-culture of the 1960s offered one such intimation, as did modern art. Marcuse was fond of quoting Stendhal's definition of beauty as 'the promise of happiness'. The radical minorities which had, in some way or other, come to see through the status quo (and, more important, to live beyond it) were, he believed, 'potential catalysts of rebellion within the majority'. Though such claims resonate with the (to most) hopelessly dated revolutionary hopes of the 1960s, many themes of Marcuse's philosophy - the totalizing power it accords to Life, for example, and the cognitive function assigned to art - recall rather German classical idealism from Schiller to Schelling. Perhaps he is best remembered as one of those (Bloch is another) who tried to continue this tradition in a Marxist idiom.

See also: Frankfurt School

ALEX CALLINICOS

List of works

Marcuse, H. (1932) *Hegels Ontologie und die Grundlegung einer Theorie der Geschichtlichkeit*, Leipzig and Frankfurt; trans. S. Benhabib, *Hegel's Ontology and the Theory of Historicity*, Cambridge, MA: MIT Press, 1987.(Marcuse's *Habilitationsschrift*, a reinterpretation of the Hegelian dialectic in the light of Dilthey's philosophy of Life and Heidegger's existentialist phenomenology.)

(1941) *Reason and Revolution*, London: Oxford University Press.(A powerful interpretation of the Hegelian tradition, culminating in Marx, in which this tradition is conceived as a 'negative philosophy', and counterposed to the positivist subordination of reason to 'the authority of established fact'.)

Marcuse, H. (1955) *Eros and Civilization*, New York: Beacon Press; repr. London: Sphere, 1969.(A highly original attempt to synthesize Marx and Freud by means of the conception of Life as the dynamic reconciliation of subject and object developed in Marcuse's early writings.)

Marcuse, H. (1958) *Soviet Marxism*, New York: Columbia University Press.(The most sustained effort within the Frankfurt School tradition to distinguish its version of Marxism from the official ideology of the Soviet Union.)

Marcuse, H. (1964) *One-Dimensional Man*, Boston, MA: Beacon Press; repr. London: Sphere, 1968.(Marcuse's best-known work; a critique of contemporary Western society and thought which draws heavily on Horkheimer's and Adorno's *Dialectic of Enlightenment* and is notable for the vigour with which it pursues the natural sciences and anlytical philosophy as instances of one-dimensional 'positive thinking'.)

Marcuse, H. (1968) *Negations*, trans. J. Shapiro, London: Allen Lane.(A collection of essays that includes some of Marcuse's contributions to the *Zeitschrift für Sozialforschung* during the 1930s.)

Marcuse, H. (1969) *An Essay on Liberation*, London: Allen Lane.(Though written before the French upheaval of May-June 1968, this is perhaps Marcuse's most politically optimistic work, in which he defies the ban imposed by both orthodox and Frankfurt Marxism on 'utopian speculation'.)

Marcuse, H. (1972) *Studies in Critical Philosophy*, trans. J. de Bres, London: New Left Books.(Collection that includes an important early essay on Marx's *1844 Manuscripts*.)

Marcuse, H. (1978) *The Aesthetic Dimension*, London: Macmillan.(A late work representing a relatively extreme version of Marcuse's tendency to see art as an intimation of a reconciled human existence.)

References and further reading

Adorno, T.W. and Horkheimer, M. (1947) *Dialectic of Enlightenment*, trans. J. Cumming, London: Allen Lane, and New York: Herder & Herder, 1972.(The Frankfurt School's most sustained critique of liberal-capitalist civilization, in which reason itself stands condemned for its role in establishing the domination of nature at the basis of social oppression.)

Habermas, J. (1988) 'Psychic Thermidor and the Rebirth Rebellious Subjectivity', in R. Pippin *et al*., *Marcuse: Critical Theory and the Promise of Utopia*, South Hadley, MA: Bergin & Garvey, 3-12.(A sympathetic discussion of the 'affirmative' quality of Marcuse's thought by the philosopher generally thought of as the Frankfurt School's main heir.)

Jay, M. (1973) *The Dialectical Imagination*, London: Heinemann.(A major intellectual history of the Frankfurt School.)

Katz, B. (1982) *Herbert Marcuse and the Art of Liberation*, London: Verso.(Intellectual biography with a good bibliography.)

Kellner, D. (1984) *Herbert Marcuse and the Crisis of Marxism*, London: Macmillan.(Concerned, as the title suggests, with situating Marcuse within the Marxist tradition.)

Lukács, G. (1923) *History and Class Consciousness*, trans. R. Livingstone, London: Merlin, 1971.(The most important single work of twentieth-century Marxist philosophy; the English edition contains a highly self-critical preface written in 1967.)

MacIntyre, A. (1970) *Marcuse*, London: Fontana/Collins.(A devastating critique concerned particularly - and somewhat ironically in the light of the author's later development - to demonstrate the defects in Marcuse's treatment of analytical philosophy and classical Marxism.)

Pippin, R., ***et al.*** (1988) *Marcuse: Critical Theory and the Promise of Utopia*, South Hadley, MA: Bergin & Garvey.(A valuable collection of essays that includes some of the best critical work on Marcuse, and an important interview in which he discusses his relationship with Heidegger.)

Marginality

Traditional definitions of marginal persons include those who live in two worlds, but do not feel well integrated into either and those who live in societies which are in the process of being assimilated and incorporated into an emerging global society. The influence of Anglo-American and European cultures has brought this situation into existence. A broader, more contemporary understanding of marginality is the condition of feeling marginal in relation to various concepts of the centre. This state produces a stigmatized identity, which either aspires to inclusion or assimilation into the centre, or demands recognition of and respect for a separate but equal existence. This condition of marginality can be experienced in varying degrees by many kinds of people.

Often gender, sexual preference, age, ethnicity, geography and religion are factors which can influence perceptions of marginality. Those who perceive themselves, or who are perceived by others to be marginal are often female, dark-skinned, very young or elderly, poor, disabled, nonheterosexual, displaced, exiled, immigrant, rural, indigenous, 'foreign', outcast, persecuted, or otherwise 'different' from those who occupy positions of privilege in the centre, or the metropolis. Critics of the term 'marginality' believe it has become overused to the point of losing descriptive precision because, they argue, almost everyone has experienced some form of marginality. In philosophy, however, the phenomenon of feeling, or being, perceived as peripheral, or on the margin, has generated critical perspectives which have enlightened discourse on social integration and stratification; personal suffering and economic, political, and cultural inequality. In addition, analyses of marginality have called into question notions of the 'universal' and the 'objective' set forth by many Western philosophers.

1 Forms of marginality

People who live in two worlds, but do not feel well integrated into either, could include, for example, *Nuyoricans*. This group comprises people of Puerto Rican origin living in New York, who are often perceived as Puerto Ricans or, more generically, as Latinos or Hispanics by other New Yorkers. Nuyoricans are thought of as 'Americanized' by those who live in Puerto Rico and they are considered as traitors to 'real' Puerto Rican culture. The Nuyorican's existence is marginal on the US mainland as well as on the island of Puerto Rico.

Chicanos choose to identify themselves by a name other than Mexican-American to indicate that they are Mexicans with a nonAnglo image of themselves living in the USA. They are not Mexicans who would feel at home in Mexico, with which many of them are unfamiliar, yet neither do they feel assimilated or choose to identify with the mainstream Anglo-American culture they experience in Los Angeles and elsewhere.

Similarly, *black Latinos* often are categorized as, or feel pressurized into identifying with either African-Americans or Latinos, although they do live in both worlds. Immigrants and exiles in general may feel that they live a dual, conflicting existence, as they attempt to maintain familiar cultural traditions which may become increasingly distant while struggling to adapt to new environs and linguistic challenges which some fail to master to their satisfaction.

With the influence of Anglo-American and European cultures those who live in societies which are in the process of being assimilated and incorporated into an emerging global society include people in developing nations, like Africa, Asia and Latin America. Their awareness of dominant or metropolitan cultures is often reinforced by television and imported music. A proliferation of icons of North American popular culture can be found, for example, in the southern Mexican states of Oaxaca and Chiapas, where indigenous peoples have protested against their marginality relative to both Mexico City and the USA. Consciousness of marginality also influences debates in Latin America over which trade agreements would be most beneficial in bringing particular countries closer to the centres of power. A corollary response to the awareness of marginality and the process of assimilation is the rejection of incorporation into a mainstream. Instead, a defence of the interests and traditions of marginal groups and the right of these groups to coexist would be favoured, respecting pieties without adopting one another's values.

The broader condition of feeling marginal in relation to various concepts of the centre, thus possessing a stigmatized identity, is often linked to concepts of otherness and alienation. It should be noted, however, that while

a marginal person may feel alienated, someone who feels alienated is not necessarily marginal. Marginal people are victims of a seemingly immutable, structural alterity. They feel or are made to feel voiceless rather than vocal, powerless rather than empowered, 'barbarous' rather than 'civilized' and unequal rather than equal. They may perceive themselves as subjects rather than as citizens as they often lack the tools, the means and the context to engage in transformational and participatory political processes (see Citizenship). Those whose marginality is not best captured by the earlier definitions of either living in but not belonging to two worlds, or living in an assimilating culture, fit into the alienated world of marginality, or subaltern otherness. Such marginal persons may be female, dark-skinned, very young or elderly, poor, disabled, lesbian, gay, bisexual, transsexual or transgendered, identified with religious or ethnic minorities, rural, or otherwise geographically or ideologically peripheral.

2 Marginality in philosophy

Philosophy produced by thinkers who feel themselves to be culturally or politically marginal is often different in significant ways from philosophy produced by those who perceive themselves as belonging to the 'centre'. This theme is present in the works of recent thinkers such as, James Baldwin, Edward Brathwaite, V.S. Naipaul, Octavio Paz, Adrienne Rich, Edward Saïd, Derek Walcott and Leopoldo Zea (see African philosophy, Anglophone; African philosophy, Francophone; Liberation philosophy; Mexico, philosophy in; Feminist thought in Latin America).

There are in existence metropolitan philosophers, who may or may not be aware of their position of privilege, and there are those who have been 'expelled from the centre of the world and…condemned to search for it through jungles and deserts or in underground mazes of the labyrinth' (1961: 209), as Octavio Paz has indicated. Aristotle maintained that the Greek man was a rational creature capable of governing those who are less rational, such as children, women and slaves. Such a stance testifies that marginality has been a time-honoured tradition throughout the history of Western philosophy. For Leopoldo Zea, marginality is a mechanism whereby one human can deny the other of humanity. One's humanity, in his view, becomes circumstantial, resting on accidents such as skin colour, gender, social class and level of education. Zea refers to the *occidentals*, as opposed to the *accidentals*; Frantz Fanon uses the image of the 'wretched' (1961) and José Ortega y Gasset writes of the existentially 'shipwrecked'. The image offered by Mexican humanist Alfonso Reyes was that Mexican intellectuals would not be invited to the banquet table of Western civilization, although they had many contributions to offer.

3 The value of marginality

Argentine philosopher Arturo Andrés Roig believes that 'The truth is not found primarily in the totality, but in determinate forms of particularity with power to create and recreate totalities from a place outside the latter, as alterity' (1983: 113). Like Roig, most philosophers whose work arises out of marginality and marginalized peoples believe they are testifying that dominant philosophical canons are uninformed or underinformed by the experience of many who are socially and culturally removed from those metropolises in which mainstream philosophers live and think. Leopoldo Zea explains, 'The great Greek, medieval, modern, and contemporary philosophers never had to worry about being original or about their cultures being strange since both their cultures and the people who created these cultures were considered universal' (1971: 15). Zea maintains that the problem of philosophy in Latin America 'is the awareness that its existence is a marginal existence' (1971: 19).

However, being marginal is not necessarily a negative intellectual phenomenon. On the contrary, there is an ironic sense in which the philosophy of marginality is a gift from the centre to the margin. Marginality is a gift to philosophers who live and think on the margin in the sense that it affords them a certain latitude to make original contributions in areas that have been ignored through the undetected or unimagined provincialism of the mainstream. Marginality becomes a methodology which challenges how and by whom values are grounded to such an extent that marginal thinkers believe that their viewpoint offers a vantage point that better illuminates and more comprehensively grounds values.

Philosophy from the margin more easily lends itself to self-criticism, in opposition to philosophy from the centre which has no pressing need to criticize itself precisely because it represents the centre, or universal. Therefore, marginal philosophy can be more thought-provoking and compelling.

See also: Alterity and identity, postmodern theories of; Cultural identity

AMY A. OLIVER

References and further reading

Baldwin, J. (1962) *The Fire Next Time*, New York: Dell.(Analysis of racism.)
Fanon, F. (1961) *The Wretched of the Earth*, London: Penguin.(Analysis of colonialism.)
Paz, O. (1961) *The Labyrinth of Solitude: Life and Thought in Mexico*, trans. L. Kemp, New York: Grove. (Discussion of solitude in Mexico.)
Roig, A.A. (1983) *Teoría y crítica pensamiento latinamericano (Theory and Criticism of Latin American Thought)*, Mexico: Fondo de Cultura Económica.(Essential book for specialists in Latin American thought.)
Stonequist, E.V. (1937) *The Marginal Man: A Study in Personality and Culture Conflict*, New York: Charles Scribner's Sons.(Discussion of marginality.)
Weinstein, M. and Weinstein D. (1981) 'The Problematic of Marginality in Mexican Philosophy', *Canadian Journal of Political and Social Theory* 4 (3).(Issues of marginality in Mexico are discussed.)
Zea, L. (1971) *La esencia de lo americano (The Essence of the American)*, Buenos Aires: Editorial Pleamar. (Discussion of certain characteristics of the inhabitants of the Americas.)
Zea, L. (1988) *Discurso desde la marginación y la barbarie (Discourse of the Marginal and the Barbarous)*, Barcelona: Editorial Anthropos.(Analysis of marginality in relation to civilization and barbarism.)
Zea, L. (1992) *The Role of the Americas in History*, ed. A.A. Oliver, trans. S. Karsen, Savage, MD: Rowman & Littlefield.(Analysis of Latin American marginality.)

Maritain, Jacques (1882-1973)

Maritain was one of the most influential twentieth-century interpreters of the thought of Thomas Aquinas. His interests spanned many aspects of philosophy, including aesthetics, political theory, philosophy of science, metaphysics, education, liturgy and ecclesiology.

His acknowledged masterpiece is The Degrees of Knowledge (1932). In this work, Maritain expands on Thomistic thought and seeks to explain the links between philosophy, science and religion as branches of wisdom. Rather than being a close study of Thomism, this work expands on Thomistic ideas and puts them into the context of the modern world. In natural science, for example, he distinguishes between empirical knowledge of nature and philosophy of nature; the latter consists in the knowledge of essence, while the former is concerned with the knowledge of form.

In moral philosophy, Maritain expands on Aquinas, holding that no true conception of the human ultimate end is philosophically possible, and that moral philosophy therefore must be subordinated to moral theology. Later in his career, Maritain concentrated more strongly on theology, but throughout his life his Roman Catholic faith informed all of his works. He continues to be read widely today, with a worldwide reputation which is especially strong in France and North America.

1 Life

Jacques Maritain was one of the two or three most prominent Thomists during the long second phase of the revival of study of St Thomas Aquinas. It is difficult to think of an area of philosophy to which he did not devote himself. His first book appeared in 1906 and his last in the year he died. He lived through the period when Thomism was established throughout the Catholic world and survived into a post-conciliar period when that effort came under acute criticism.

Maritain was born in Paris in 1882. In 1900, while a student at the Sorbonne, he met Raissa Oumansov, a fellow-student; they were true soul mates, and were married in 1902. After their conversion to Catholicism, it was Raissa who first discovered St Thomas Aquinas. Their home in Versailles became a place where artists, authors, philosophers and theologians gathered. When they moved to Meudon in 1923, these meetings were formalized as the *Cercle d'Études Thomistes*. Maritain's *Prayer and Intelligence* (1922) sets forth the complementarity of the spiritual and intellectual lives.

Maritain taught first at the Lycée Stanislas, then at the Institut Catholique, but his thought and writings were not driven by academic duties. He first visited North America in 1933, and returned there in 1938. Étienne Gilson had hoped to enlist Maritain as a permanent associate at the Pontifical Institute of Mediaeval Studies in Toronto, but Maritain proved elusive. He became a frequent lecturer at Chicago and Notre Dame, and was with his wife in the United States when France fell in 1940; they remained until 1944, when Jacques was named French Ambassador to the Vatican. Raissa's two volumes of memoirs, *We Have Been Friends Together* and *Adventures in Grace*, acquainted a generation of Americans with their dramatic story, their friends, and the burgeoning Catholic culture of France.

Maritain served as the president of the French delegation to UNESCO during the drafting of the Universal Declaration on Human Rights, but resigned as ambassador to accept an appointment at Princeton in 1948. The Maritains returned to France in 1960, where the ailing Raissa died. Jacques settled in Toulouse, where he taught philosophy and theology to the Little Brothers of Jesus. In 1965, at the close of the Second Vatican Council, Paul VI presented Maritain with a message addressed to intellectuals. Maritain divided his time between Toulouse and Kolbsheim near Strasbourg. In 1970, this quintessential layman entered the religious life as a Little Brother of Jesus. He died in 1973 and is buried with Raissa at Kolbsheim.

2 Thomism

In *Aeterni Patris* (1879), Leo XIII called for a renewal of Christian philosophy, particularly that of St Thomas Aquinas, as a way of combating modern errors. The response to the encyclical occurred in stages, eventually fanning out across the globe. Journals were founded, new institutes and associations were formed, conferences

were held, and there was a flood of interpretations and prolongations of the thought of Aquinas. At first a largely clerical phenomenon, the Thomistic revival became, with Maritain and Étienne Gilson, a lay effort which more effectively related traditional Catholic thought to the wider culture.

Maritain, like so many of his generation, came under the influence of Henri Bergson, whom he examined in comparison with Aquinas in his first book, *Bergsonian Philosophy* (1913). *Antimodern* (1922) might suggest that Maritain saw Thomism simply as antithetical to contemporary thought, but that this was not his view is clear from *Angelic Doctor* (1929), as well as earlier works. However, the most comprehensive statement of his philosophical vision is given in *The Degrees of Knowledge* (1932).

The *Degrees* is Maritain's masterwork. He adopts Aquinas' distinction between philosophy and theology, whereby the latter is discourse whose principles are provided by divine revelation, and held by faith. One who does not hold the principles to be true will not hold as true the conclusions drawn from them. Theology, considered as a truth-seeking discipline, is thus an activity of believers. Philosophy, on the other hand, is discourse which proceeds from principles in the public domain, knowable by anyone; the philosopher must link his inferences to what everybody already knows.

This is obviously an Aristotelian view. Maritain sees Aquinas as an Aristotelian, and, accordingly, the Aristotelian division of philosophy into theoretical and speculative disciplines governs his thought. One of the interesting features of his work is the way he expands and reshapes that basic division to accommodate contemporary advances in thought.

Maritain's aim in the *Degrees* is to lay before us a vast and connected and hierarchical panorama so that we see our quest for knowledge as a graded ascent to wisdom *tout court*. He first discusses the relation between philosophy and experimental science, goes on to a plea for critical realism, discusses philosophical knowledge of sensible reality and then sets forth his metaphysics. The second part treats the relation between philosophy and mystical experience, and argues that the culmination of the human quest for truth is that wisdom which is a gift of the Holy Spirit. Like Aquinas, Maritain seeks a comprehensiveness which finds unity beyond necessary distinctions.

In 1931, thanks to a taunt by Émile Bréhier (in the *Revue de Métaphysique et de Morale*), a debate began on the nature of Christian philosophy. Was it possible for Christians to engage in philosophy or must their efforts be seen as the question-begging of believers? Maritain agreed that it would be nonsense to say that the Christian faith has no influence on the thinking of the believer. But the influence of faith has its analogue in the pre-philosophical assumptions of one ignorant of or hostile to the faith. Maritain distinguished between the act of philosophizing as a moral and human deed, and the product of such activity. With respect to the latter, believer and unbeliever, despite their differing existential starting points, must meet the same criteria for success. Maritain rejected the charge that Christian faith disqualifies one from philosophy; indeed, he considered it an aid and stimulus.

Few of Maritain's writings are exegetical, that is, close studies and interpretations of texts of Aquinas. Rather, having absorbed Aquinas, Maritain was interested in making him intelligible to modern philosophers, creating a new synthesis continuous with Aquinas' historical achievement. Étienne Gilson once wrote that he himself had spent his life seeking to know exactly what Aquinas meant, while Maritain was seeking to do in the present what Aquinas had done in the past.

3 Main arguments

Aesthetics. In 1920, Maritain published *Art and Scholasticism*, his second book. His wife was a poet, their friends were artists in various media - Rouault, Cocteau, Claudel, Péguy, Julien Green - and he felt a need to ask what the relation between art and philosophy is. The little book is a fascinating *mélange* of Aristotelian and Thomistic lore, but reveals as well Maritain's wide knowledge of literature and art. Many artists welcomed the Aristotelian truism that art aims at the perfection of the thing made, not of the maker. A salient mark of Maritain's aesthetics is his likening of poetic knowledge to the judgment of connaturality that Aquinas attributed to prudence. The Mellon Lectures, *Creative Intuition in Art and Poetry* (1953), bring to fruition Maritain's thoughts over many decades. They emphasize the knowledge which is prior to and may be expressed in artistic production.

Philosophy of science. For Aquinas, philosophy of nature establishes the possibility of a science beyond itself. If

natural change ultimately requires an uncaused cause, and thus one outside the realm of things that come to be (natural things), a science which treats of all being and not just natural beings suggests itself. Change involves matter; the changeless lacks matter, so to be and to be material are not identical. The development of modern science called into question the validity of Aristotelian physics, and this led some Thomists to seek to bypass it and begin with metaphysics. Maritain developed a distinction between experimental or empiriological (Maritain’s neologism) knowledge of nature and the philosophy of nature. The former, mathematical physics, falls short of knowledge of essence, the knowledge that characterizes philosophy of nature. This suggests that the matter/form analysis of sensible substance is ontological knowledge, profounder than the knowledge gained by the sciences. Developed in a number of books, Maritain’s position on this matter is also the first discussion in *The Degrees of Knowledge*.

Metaphysics. While Maritain’s position on metaphysics was initially similar to that of Aquinas, seeing the science of being as dependent for its very possibility on achievements in philosophy of nature, his approach changed, as is evident from *Existence and the Existent* (1947). There Maritain speaks of an intuition of being as the *sine qua non* of metaphysics. This intuition is not common to all, but is possessed only by a few. This development puts Maritain in the camp of those who feel that metaphysics can be undertaken without any prior science, philosophical or otherwise, of sensible reality. The ‘intuition of being’ seems to arise from the influence of existentialism. To some degree, it points to the need for a sense of wonder as the presupposition of philosophical questioning. Furthermore, it appears to be Maritain’s response to the interpretation of Gilson, who saw Aquinas’ use of *esse* as the key to his thought; the grasp of *esse* is part and parcel of the recognition that essence and existence are distinct, a recognition Gilson felt was missing from Aristotle.

Moral philosophy. It was in the area of moral philosophy that Maritain was most innovative. In Aquinas, we encounter the view that there are certain principles of action which are part of the natural repertoire of any human agent. Aristotle had articulated the ultimate end appropriate to the nature of the human agent, which end provides such limited happiness or fulfilment as is possible for humans. From the Christian point of view, Aristotle’s sense of the inadequacy of human happiness is just what is to be expected, since humans are destined for a higher happiness after their earthly existence. Aquinas thus spoke of imperfect and perfect beatitude, with ‘imperfect’ representing Aristotle’s own sense and not simply a Christian judgment. Maritain’s view is markedly different: he holds that no true conception of the ultimate end for humans is philosophically possible. Moral philosophy ‘adequately understood’ must be subordinated to moral theology, that is, must accept as true a conception of an ultimate end beyond our natural powers to discover. Critics accused Maritain of smudging the distinction between philosophy and theology and he was soon caught up in elaborate defences of his position (in, for example, *Science and Wisdom* (1935)).

Political philosophy. From involvement with the right-wing Action Française, condemned by the Catholic Church in 1926, Maritain moved leftward to the position of *Integral Humanism* (1936). Visits to the USA altered his critique of capitalism (see, for example, *Reflections on America* (1957)). The fall of France, and its wartime occupation, deeply affected him, and he became a champion of democracy and human rights. The role he played in the Universal Declaration on Human Rights of 1948 is evident in his Walgreen Lectures, which were published as *Man and the State* (1951). Here Maritain confronts the conflict between natural law and natural rights. He argues that the duties consequent on natural law and natural rights are two ends of the same thought.

A more timely problem arose from the fact that the signatories of the Universal Declaration held such radically different views of those rights. Maritain was incapable of a cynical explanation of this paradox, and he proposed a distinction between gnoseological and ontological natural law. Perhaps it would not simplify his point too much to say that the nature of the speakers and of the things spoken of in such situations provide an objective basis for agreement, whatever current misunderstandings obtain.

Theology. In the final phase of his career, Maritain’s writings became overtly theological, doubtless because they were occasioned by teaching the Little Brothers of Jesus in Toulouse. He and Raissa had written *Liturgy and Contemplation* in 1959, and in *The Degrees of Knowledge* he had ventured into mystical theology. *On the Grace and Humanity of Jesus* (1967) and *On the Church of Christ* (1970) and much of the posthumously published *Approches sans entraves* (1973) are purely theological. Among the surprises to be found in these late writings is a questioning of the eternity of the punishment of the damned even though they are forever incapable of the beatific

vision.

4 Influence

From the time of his conversion in 1906, Maritain regarded his Catholicism as the most essential thing in his life. The Christian vocation, the call to respond to grace and to the gifts of the Holy Spirit, was from the outset the frame within which his intellectual work was carried out. His understanding of the concept of Christian philosophy, as well as the constitution he wrote for the *Cercle d'Études Thomistes* (see his *Carnet de notes* (1965)), display his sense of the profound union of the intellectual and spiritual lives. Both Raissa and Jacques strove for holiness, and the testimony of friends suggests that their efforts were not unrewarded. They were instrumental in the return to the faith or conversion of many. The published correspondence between Maritain and Julien Green gives some flavour of what will be found in other, unpublished, letters.

Maritain's estimate of what was happening in the Church in the wake of Vatican II was not cheerful. *The Peasant of the Garonne* (1966) lamented the influence of Teilhard de Chardin and of phenomenology. He saw a resurgence of the modernism that had been condemned by Pius X in 1907. Some dismissed Maritain as out of date and grumpy.

As it happens, his influence continues and, indeed, increases. The Jacques Maritain Center at the University of Notre Dame, Indiana, was founded in 1948, and soon such centres sprang up around the world. There are two international Maritain societies and many national societies. A fifteen-volume *Oeuvres complètes* was completed in 1994, and a twenty-volume English edition began to appear in 1995.

See also: Thomism

RALPH McINERNY

List of works

Maritain, J. (1975, 1978) *Jacques Maritain Oeuvres*, ed. H. Bars, 2 vols.(By and large, these are selections rather than complete works; volume 1 covers the years 1912-39, volume 2 1940-63.)

Maritain, J. (1982-94) *Oeuvres complètes*, ed. J.-M. Allion, M. Hany, D. Mougel, R. Mougel, M. Nurdin and H. Schmits, Fribourg: Éditions Universitaires, 15 vols.(The volumes are organized chronologically and cover the period 1906-60, including the writings of Raissa as well as Jacques. Each volume is well indexed, with bibliographical notes, and includes pieces from periodicals and prefaces to others' works, as well as some indication of the response to Maritain's work. The French text of this collection is definitive.)

Maritain, J. (1995-) *Complete Works*, ed. R. McInerny, B. Doering and F. Crosson, Notre Dame, IN: University of Notre Dame Press.(With some exceptions, each volume contains a number of works. Volume 7, *The Degrees of Knowledge*, was one of those that appeared in 1995. The project should be completed by 2005.)

References and further reading

Bréhier, É. (1931) 'Y-a-t-il une philosophie chrétienne?' (Is There a Christian Philosophy?) *Revue de Métaphysique et de Morale*, April-June.(Started the debate on the nature of Christian philosophy described in §2.)

Doering, B.E. (1983) *Jacques Maritain and the French Catholic Intellectuals*, Notre Dame, IN: University of Notre Dame Press.(Places Maritain within the French intellectual scheme.)

Kernan, J. (1975) *Our Friend Jacques Maritain*, New York: Doubleday.(A charming personal memoir by an American friend of the Maritains.)

Leo XIII (1879) *Aeterni Patris*, in J. Maritain, *Le Docteur angélique*, in *Oeuvres complètes*, vol. 4, Paris: Éditions Saint-Paul, 1983; trans. J.W. Evans and P. O'Reilly, *St Thomas Aquinas*, New York: Meridian Books, 1958.(Includes a discussion of the encyclical, as well as its text.)

McInerny, R. (1988) *Art and Prudence: Studies in the Thought of Jacques Maritain*, Notre Dame, IN: University of Notre Dame Press.(A collection of essays written at various times, dealing chiefly with Maritain's aesthetic and moral positions.)

Marius Victorinus (*fl.* 4th century AD)

Gaius Marius Victorinus was a rhetorician active in Rome in the fourth century AD. Classically educated and with an interest in philosophy, he converted to Christianity late in life and transferred his philosophical interests to Christian works. Strongly influenced by Neoplatonism, particularly by the works of Plotinus and Porphyry, he sought to articulate Christian concepts such as the Trinity in Neoplatonic terms. His writings on the Trinity and the soul influenced Augustine and other patristic philosophers.

Born in Africa some time in the decade AD 281-91, Victorinus acquired a classical education, including considerable knowledge of philosophy and its history. He wrote commentaries on the works of Cicero, produced treatises on grammar and logic and translated Greek philosophical works into Latin. Associated with aristocratic and senatorial circles, his success as a rhetor is evinced by the commissioning of a statue in his honour in the Forum of Trajan during his lifetime. In these activities, Victorinus represents a well-established Roman cultural type: the successful rhetorician with a learned interest in philosophy. Yet Victorinus was in the cultural vanguard of his time in one critical respect, his conversion to orthodox Christianity. We learn from St Jerome that this occurred 'in extreme old age', perhaps in the mid 350s. His late career was as an author of Christian theology, generating a series of commentaries on the letters of St Paul (Ephesians, Galatians, Philippians), as well as treatises of intra-Christian polemics. It is from these later theological works, directed primarily against the Arian heresy, and his translation efforts that Victorinus' subsequent philosophical reputation is derived.

Victorinus' interest in the philosophy of his time went well beyond the norm for a learned rhetorician. He translated some of Aristotle's logical works into Latin, as well as a large number of treatises from the *Enneads* of Plotinus and possibly some works of Porphyry, including the *Isagōgē*. This effort suggests both technical knowledge of philosophy and serious commitment to a cultural programme. His goal appears to have been the dissemination into the Latin world of contemporary Neoplatonism, particularly that of the Plotinian school (see Neoplatonism). Here Victorinus was successful not only among pagan readers, but also among his subsequent co-religionists, the Christians. These translations were essential to the assimilation of Plotinian and Porphyrian thought by Latin Christian intellectuals in the late fourth century (see Patristic philosophy). This fact is famously attested in Augustine's conversion narrative, the *Confessiones*; here, in Books VII-VIII, he describes the intellectual effect of reading - in the mid-380s - 'some books of the Platonists' (*Confessiones* VIII. 9.13) translated by Victorinus. These Neoplatonic treatises were decisive in motivating Augustine's acceptance of Christian monotheism, as they must have been at an earlier time for Victorinus himself (see Augustine).

Of Victorinus' philosophical ideas, we have only an imperfect record. His goal was to articulate the Christian Trinity in Neoplatonic terms, and to do so in a fashion that secured the orthodox conception of the consubstantiality of the Father and the Son. This theological desideratum was the source of his philosophical originality. In order to secure their monotheism, Christian thinkers were inclined either (a) to describe the Father, Son and Holy Spirit as modes or names of a single God, or (b) to treat the Son and Spirit as ontological derivatives of a primordial Father (see Trinity). The former, modalistic solution was articulated philosophically by Sabellius among others; the latter, subordinationistic approach was adopted by Origen of Alexandria, the Christian student of Plotinus' teacher Ammonius, and more forcefully by his later followers, including Victorinus' Arian opponents. Victorinus appears to have sought a way through this scholastic dichotomy by resorting to the subtle hypostatic theories of the Plotinian-Porphyrian school.

His approach was to treat the divine Father as a first principle defined in a fashion similar to the Plotinian One. The Father was above all being, beyond both the material things of the physical world and the Platonic intelligibles. In Victorinus' view, God, as the primordial Father, was best understood as 'pre-being' (*proon*), an absolute and perfect source for all finite reality. Beyond specification or determinate description, the Father was therefore beyond knowledge and the capacity of human conceptualization. As such, the Father makes his presence known through his Word or Son, the first being. The Son is the initial determination of the unlimited Father, the perfect manifestation of his hidden divine source. As the visible aspect of God, the Son is the 'form of God', the divine principle that brings out the latent nature of the Father. The Son thus gives finite form to the indefinite or formless substance of God the Father.

What is especially interesting about this Neoplatonic treatment of these Christian notions is Victorinus' recognition of the non-subordinationistic possibilities of Plotinian theology. He insists that the Father, as the unbounded source of all reality, cannot be juxtaposed as a distinct entity, superior in its own nature to that of the Son. Rather, the Father is the ontological ground of the Son, and as such, cannot be viewed either as external to the Son or as another being superior to him. If the Father is 'pre-being', then he is directly present to his first definite manifestation, his Word. Ontological subordination would seem therefore to be a misreading of what is not a hierarchical relation between separate entities. In Victorinus' model, the omnipresence of the Plotinian One thus becomes the basis for securing the consubstantiality of Father and Son, the hallmark of Nicene orthodoxy.

Victorinus is also noteworthy for his treatment of the human soul. The soul is brought into being by the creative activity of the Word, and bears its image: its capacity for rational reflection in particular reflects the order and structure of the Word. Within the soul, Victorinus discerned a triad of aspects indicative of its source in the divine: being, life and intellect. While fallen, the human soul retains this divine image, and has the capacity to restore its likeness to the Logos by exercising its intelligence and sloughing off the effects of its descent into time and change. In emphasizing the triadic structure of the descended soul and its isomorphism with the divine Trinity, Victorinus anticipated Augustine's subsequent interest in Trinitarian models of the soul and so the Western Christian articulation of the relation of the human and the divine (see Soul, nature and immortality of the).

See also: Augustine; Encyclopedists, medieval; Neoplatonism; Platonism, medieval; Patristic philosophy; Plotinus; Soul, nature and immortality of the; Trinity

JOHN PETER KENNEY

List of works

Marius Victorinus, Gaius (after 355) *Marii Victorini Opera (Works of Marius Victorinus)*, ed P. Henry, P. Hadot and F. Gori, Corpus Scriptorum Ecclesiasticorum Latinorum vol. 83, Vienna: Hoelder-Pichler-Tempsky, 1971-86.(Collected works of Victorinus.)

Victorinus, Gaius (after 355) *Ad Candidum (To Candidus)*, ed. P. Henry and P. Hadot in *Traités théologiques sur la Trinité* I-II, Paris: Éditions du Cerf, Sources Chrétiennes, 1960, vols 68-9.(A theological reply to an extant letter from Candidus, an Arian Christian.)

Victorinus, Gaius (after 355) *Adversus Arium (Against the Arians)*, ed. P. Henry and P. Hadot in Traités théologiques sur la Trinité I-II, Paris: Éditions du Cerf, Sources Chrétiennes, vols 68-9, 1960.(A major attack against Arianism in four books.)

Victorinus, Gaius (after 355) *In epistolam Pauli ad Ephesios (On the Letter of Paul to the Ephesians)*, in J.P. Migne (ed.) *Patrologia Latina*, Paris, 1844, vol. 8, 1235-94.(Expository commentary on one of the Pauline epistles.)

Victorinus, Gaius (after 355) *In epistolam Pauli ad Galatas (On the Letter of Paul to the Galatians)*, in J.P. Migne (ed.) *Patrologia Latina*, Paris, 1844, vol. 8, 1145-98.(Expository commentary on one of the Pauline epistles.)

Victorinus, Gaius (after 355) *In epistolam Pauli ad Philippenses (On the Letter of Paul to the Philippians)*, in J.P. Migne (ed.) *Patrologia Latina*, Paris, 1844, vol. 8, 1197-1236.(Expository commentary on one of the Pauline epistles.)

References and further reading

Clark, M.T. (1976) 'Marius Victorinus Afer, Porphyry, and the History of Philosophy', in R. Baine Harris (ed.) *The Significance of Neoplatonism*, Norfolk, VA: Old Dominion University, 265-73.(A short study in the metaphysics of Plotinus, Porphyry and Marius Victorinus.)

Clark, M.T. (1981) 'The Neoplatonism of Marius Victorinus the Christian', in H.J. Blumenthal and R.A. Markus (eds) *Neoplatonism and Early Christian Thought*, London: Variorum, 153-9.(Victorinus and the concept of the divine One.)

Clark, M.T. (1982) 'A Neoplatonic Commentary on the Christian Trinity: Marius Victorinus', in D.J. O'Meara (ed.) *Neoplatonism and Christian Thought*, Albany, NY: State University of New York Press, 24-33.(An assessment of Victorinus' use of Neoplatonism in his theology.)

Cooper, S.A. (1995) *Metaphysics and Morals in Marius Victorinus' Commentary on the Letter to the Ephesians*, New York: P. Lang.(A study of Neoplatonism and Christian thought in the Ephesians commentary.)

Hadot, P. (1968) *Porphyre et Victorinus*, Paris: Études Augustiniennes, 2 vols.(A major study of Victorinus and

Porphyry, including Greek and Latin texts.)
Hadot, P. (1971) *Marius Victorinus: recherches dur sa vie et ses oeuvres (Marius Victorinus: Research on His Life and Works)*, Paris: Études Augustiniennes.(A discussion of biographical evidence.)
Manchester, P. (1992) ‘The Noetic Triad in Plotinus, Marius Victorinus, and Augustine’, in R.T. Wallis and J. Bregman (eds) *Neoplatonism and Gnosticism*, Albany, NY: State University of New York Press, 207-22.(An examination of triadic schemes in the metaphysics of these figures.)
Markus, R.A. (1967) ‘Marius Victorinus and Augustine’, in A.H. Armstrong (ed.) *The Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press, 329-40.(The best overview in English of Victorinus’ thought.)
Steinmann, W. (1990) *Die Seelenmetaphysik des Marius Victorinus (The Metaphysics of the Soul of Marius Victorinus)*, Hamburger Theologische Studien 2, Hamburg: Steinmann & Steinmann.(A detailed study of Victorinus’ doctrines on the soul.)

Market, ethics of the

Markets are systems of exchange in which people with money or commodities to sell voluntarily trade these for other items which they prefer to have. Most economic transactions in advanced societies are of this kind, and any attempt to replace markets wholesale with a different form of economic coordination seems destined to fail. But questions about the ethics of markets are still of considerable practical concern, for two reasons at least. First, we need to make collective decisions about the proper scope of markets: are there goods and services which in principle should not be distributed and exchanged through market mechanisms - medical care, for instance? Second, markets work within a framework of property rights which sets the terms on which people can exchange with one another, and this too is subject to collective decision: for instance, should a person's labour be regarded as a commodity like any other, to be bought and sold on whatever terms the parties can agree, or does labour carry special rights that set limits to these terms? Are employees morally entitled to a share of the profits of the companies they work in, to take a concrete issue?

To guide such decisions, we need to apply general ethical principles to market transactions. First, are markets justified on grounds of efficiency, as is often claimed? What criterion of efficiency is being used when such claims are made? Second, can we regard the outcome of market exchanges as just, or, at the other extreme, should we see them as necessarily exploitative? Third, do market exchanges necessarily alienate people from one another and destroy their sense of community? These are very different questions, but an overall assessment of market ethics needs to address each of them, and perhaps others besides.

1 Efficiency

Informally, we describe institutions as efficient when they satisfy people's desires or preferences to the greatest possible extent. In this informal sense, it is easy to understand the efficiency claims made on behalf of markets. Viewed narrowly as a system of commodity exchange, a market enables people to exchange items they desire less for items they desire more, whereas whenever such exchanges are prevented, some people will have less of what they want than they might otherwise have. Viewed more broadly, as a mechanism for producing goods and services as well as exchanging them, a market provides incentives for people to make goods and provide services that other people want to buy. It also, as Hayek (1967) in particular has argued, economizes on information: individual people can make rational decisions about what and how much to produce on the basis of limited information about the costs and prices of a small number of goods and services, whereas a non-market system of production (such as a centrally-planned economy) would require some agency to assimilate data and coordinate behaviour across the whole economy (see Hayek, F.A. §2).

Economists have formalized the efficiency claim in the first theorem of welfare economics, which states that when a perfectly competitive market reaches equilibrium, this must be Pareto-optimal, meaning that no one's welfare can be further increased without reducing the welfare of someone else (see Economics and ethics §4; Pareto principle).

The efficiency claim is, however, subject to some large qualifications (for a fuller discussion, see Buchanan 1985). Real markets usually fall far short of the idealized markets of economic theory. Competition is imperfect - for instance rigidities are introduced when people invest in machinery or skills which cannot be easily redeployed when demand changes; individuals have been shown not to behave as rational maximizers of their utility; and, perhaps most importantly, many economic transactions generate positive or negative externalities such as spillover effects on third parties who have not consented to the exchange: Smith may open a new supermarket on the edge of town, and I and many others may voluntarily buy our groceries from him rather than from the corner store, but the effect is that the store has to close, imposing substantial costs on those who cannot easily travel to the supermarket; thus resulting in a negative externality. We cannot say therefore that real markets will be efficient, even by the Pareto criterion; there may be further resource transfers that such markets will not induce that would make everyone better off.

Next, Pareto optimality yields only a weak sense of efficiency. There are likely to be many resource allocations of which it is true that any reallocation will lower the welfare of at least one individual. The resource allocations generated by markets are characteristically quite unequal. By a stronger criterion of efficiency - for instance, a utilitarian criterion which tells us to aggregate welfare across individuals - efficiency would usually be increased

by redistribution from the better off to the worse off, at least up to the point where incentives begin to be affected (see Utilitarianism).

Finally, markets are said to be efficient in so far as they satisfy individuals' felt desires, which are expressed in their market behaviour - for instance, in their choice of consumer goods. This argument depends on interpreting individuals' welfare subjectively, in terms of existing preferences (see Welfare §1). If instead we were to measure welfare in terms of real interests or needs, then even the informal argument for market efficiency is put in question. Where markets meet desires that it is not genuinely to people's advantage to have satisfied, how can we advocate them on efficiency grounds?

In the light of these qualifications, the efficiency principle only provides a weak reason to favour markets; perhaps more importantly, it cannot adjudicate between economies in which almost everything is left to market mechanisms and economies where the state is more active in regulating the market and redistributing resources, such as social democracies (see Social democracy).

2 Justice and exploitation

Whether market outcomes are considered just or unjust depends on the theory of justice being applied (see Justice). At one extreme stands the view, associated particularly with Hayek, that the concept has no application to the results of a spontaneous process. Since no one intends to bring about the final outcome of a large set of market transactions, that outcome cannot be regarded either as just or as unjust. Against this, however, it has plausibly been argued that market outcomes are predictable, in their general shape if not their detail, and this is sufficient to bring the concept of social justice into play (Plant 1991).

Another view is that market transactions are always justice-preserving as long as they are fully voluntary - that is, neither side employs force or fraud to secure the transaction. This view appeals to a historical entitlement theory of justice of the kind favoured by Nozick and other libertarians (see Libertarianism; Nozick, R. §2). However, it is vulnerable to the problem of externalities that we have already encountered: a voluntary exchange between *A* and *B* may alter, perhaps for the worse, the position of *C*. Thus it seems that we have to look beyond individual transactions to consider the overall distribution of resources that results from the operation of market mechanisms over time.

Experience tells us that this distribution is likely to be quite unequal. Many philosophers would argue that such inequality is just in so far as it reflects the different choices that individuals may make - the choice to work rather than play, or to save rather than consume, for instance. Dworkin has developed a theory of this kind, where resources are initially to be divided equally among individuals, but where subsequent holdings may legitimately reflect such choices. On Dworkin's view, however, resource inequalities that stem from differences in native talent are unjust, and since much of the inequality in real-world markets appears to be of this kind, a just society would have to redistribute resources from the more talented to the less talented through a tax system.

This view can be challenged by invoking a stronger sense of desert, according to which people deserve the results of their productive efforts even when these depend on native abilities and dispositions that are not themselves deserved (see Desert and merit). Market outcomes could then be regarded as just in so far as they reflected the different productive achievements of individuals, as opposed to being the results of luck (Miller 1989). Such a justification will once again be vulnerable to the imperfections of real-world markets, where factors such as inherited advantage or monopoly power mean that many individuals will hold resources that by this criterion they do not deserve.

Philosophers in the socialist tradition have argued that many market transactions are exploitative. Notwithstanding their appearance of voluntariness, inequalities of resources or of bargaining power mean that one party is able to impose terms that make the exchange an unequal one. This applies particularly to labour contracts between workers and owners of capital. Marx developed the most celebrated theory of capitalist exploitation, arguing that an excess supply of labour would force workers to accept subsistence wages while their work created surplus value for their employers to appropriate (see Marx, K. §12). However, this theory relies on the one hand on a labour theory of value which now has few supporters, and on the other hand on a prediction about the future course of capitalism which has not been borne out.

More recent theories of economic exploitation have abandoned the labour theory of value, and have also attempted to widen the theory beyond the case of capital-labour exchanges. According to Roemer (1988), for instance, we can identify exploitation by asking whether there is a group of people who have to work for longer than is socially necessary to earn the bundle of goods they consume, and another group who need to work for less time. It emerges from these theories that although exploitation may emerge within markets, its source is always inequality in the assets or endowments that people bring to market exchanges. We can therefore enquire into the conditions under which markets can function without exploitation. This has led some political philosophers to develop theories of market socialism, examining the possibility of a market economy without private ownership of productive capital (Miller 1989; Schweickart 1993; Arnold 1994) (see Socialism §6).

3 Alienation and community

Even if markets can be framed in such a way that they are reasonably just and non-exploitative, they may still be criticized on the grounds that they set participants against one another and prevent them from enjoying fraternal or communal relations. Although market behaviour need not be motivated by narrow self-interest, it is in standard cases ‘non-tuistic’ (to use Wicksteed’s phrase), meaning that neither party to an exchange has any intrinsic concern for the welfare of the other. Each uses the other as an instrument for their own purposes, whatever these might be: when I buy oranges from the greengrocer, I have no interest in his wellbeing.

Marx once again gave fullest expression to this charge, when he described workers under capitalism as alienated, in part because they were prevented from enjoying those communal relations with their fellow workers which properly expressed their nature as human beings (see Marx, K. §4; Alienation §§3-5). But many others have attacked market societies for fomenting individualism and competition at the expense of social solidarity and community (see Community and communitarianism §3).

In reply to this, defenders of the market point out that, where markets are working properly, self-interested (or non-tuistic) behaviour on the part of each individual may contribute to the welfare of everyone; as Adam Smith famously argued, each is led by an ‘invisible hand’ to promote the interest of society more effectively than when they attempt to promote it directly (Smith 1776). Thus there is no conflict between producing or exchanging with a view to one’s own advantage and contributing to the welfare of others.

This reply only partly answers the original charge, however, because the existence of a community depends not just on how people act but on their intentions. We cannot describe the relationships that exist between buyers and sellers in a market as inherently communal. Yet there is no reason why communal relations should not develop alongside those of market exchange, for instance in workplaces, local communities or political associations. A society might then have a communitarian character overall, even though its economy was predominantly market-based.

What seems certain is that a world in which people interacted with one another solely through market transactions would be not only morally barren, but probably unsustainable. The question that arises is how to prevent the market from encroaching upon spheres of social life that presently act as a counterbalance to it. Recently philosophers have begun to investigate the reasons why certain goods and certain human relationships need to be kept insulated from market mechanisms: for instance, why body parts should not be bought and sold, or why marriage should not be turned purely into a negotiated contract between husband and wife (Anderson 1993; Andre 1995; Walzer 1983). The reasons are quite varied, but one concerns the way exposure to market mechanisms may change the character of relationships for the worse, by, for instance, undermining trust: if I know that doctors stand to profit from the treatment they prescribe, will I have the same confidence in them as when they have no financial stake in the outcome?

A careful examination of the ethics of markets is likely to lead neither to an unqualified endorsement nor an unqualified rejection of market mechanisms. Instead it will encourage us to ask questions about the institutional framework of the market. How can we ensure that it remains competitive, and therefore efficient? Can we guarantee that each participant has access to sufficient resources that they will not be exploited? How do we prevent markets from encroaching into spheres of human life where they have no place?

DAVID MILLER

References and further reading

Acton, H.B. (1993) *The Morals of Markets and Related Essays*, ed. D. Gordon and J. Shearmur, Indianapolis, IN: Liberty Press.(Rejects the claim that market exchange and competition must violate ethical principles.)

Anderson, E. (1993) *Value in Ethics and Economics*, Cambridge, MA: Harvard University Press, ch. 7.(Explores the kinds of value that cannot be realized through market relationships.)

Andre, J. (1995) 'Blocked Exchanges: A Taxonomy', in D. Miller and M. Walzer (eds) *Pluralism, Justice, and Equality*, Oxford: Clarendon Press.(Classifies grounds for holding that certain goods should not be privately owned or exchanged.)

Arnold, N.S. (1994) *The Philosophy and Economics of Market Socialism*, New York: Oxford University Press.(A critique of market socialism, arguing that opportunities for exploitation are greater under such a system than under capitalism.)

Buchanan, A. (1985) *Ethics, Efficiency, and the Market*, Oxford: Clarendon Press, ch. 2.(Perhaps the best general assessment of the ethical and efficiency arguments for and against the market.)

Hayek, F.A. (1967) 'The Principles of a Liberal Social Order', in *Studies in Philosophy, Politics and Economics*, London: Routledge & Kegan Paul.(A summary version of Hayek's defence of the market, and his attack on ideas of social justice.)

Miller, D. (1989) *Market, State, and Community: Theoretical Foundations of Market Socialism*, Oxford: Clarendon Press.(Principled defence of market socialism.)

Plant, R. (1991) *Modern Political Thought*, Oxford: Blackwell, ch. 3.(Argues against Hayek's claim that ideas of distributive justice cannot be applied to markets.)

Roemer, J. (1988) *Free to Lose: An Introduction to Marxist Economic Philosophy*, London: Radius.(Reworks and generalizes the Marxist theory of economic exploitation.)

Schweickart, D. (1993) *Against Capitalism*, Cambridge: Cambridge University Press.(Ethical critique of capitalism, and defence of democratic market socialism.)

Sen, A. (1985) 'The Moral Standing of the Market', *Social Philosophy and Policy* 2 (2): 1-19.(Helpful brief survey of the ethical criteria by which markets can be assessed.)

Smith, A. (1776) *An Inquiry into the Nature and Causes of The Wealth of Nations*, ed. R.H. Campbell, A.S. Skinner and W.B. Todd, Oxford: Clarendon Press, 1976, book IV, ch. 2.(Classic defence of market economy as contributing to social welfare.)

Walzer, M. (1983) *Spheres of Justice*, Oxford: Martin Robertson.(Explores reasons why some goods should not be allocated by market mechanisms.)

Marsilius of Inghen (1330-96)

The theological and philosophical works of Marsilius of Inghen are characterized by a logico-semantical approach in which he followed John Buridan, combined with an eclectic use of older theories, often dating from the thirteenth century. These were sometimes more Aristotelian and sometimes more Neoplatonist. The label 'Ockhamist', which is often applied to Marsilius, has therefore limited value. He was influential on Central European philosophy of later centuries, both through his own philosophy and by the way he stimulated reform of university programmes. In the sixteenth century there were still references to a 'Marsilian way' in logic and physics.

1 Logic

The semantic approach is pivotal in Marsilius' philosophy. In his theory on the properties of terms, *suppositio* is the acceptance of a term in a proposition for the thing(s) concerning which the term is verified by way of the copula of the proposition. For example, in the proposition 'a man is running', the term 'man' supposits for any man that exists, because concerning any of them it may be verified through the copula. The definition of supposition is in line with William of Ockham and, more strictly, with Buridan, but Marsilius' semantics has distinctive characteristics. Unlike Buridan, he accepts 'something imagined' as the significate of a fictional term such as *chimera*. Moreover, Marsilius does not adopt 'natural supposition', which in his view is a kind of atemporal signification, for example, of 'man' and 'animal' in 'man is an animal'. In his *Logic*, John Dorp (*fl.* 1400) follows Buridan in all respects except for the chapter on supposition, in which Marsilius is his model.

A commentary on Alexander of Villadei's *Doctrinale* is ascribed to Marsilius, probably correctly. In this work, he presents a 'conceptualistic grammar'; congruency in language is not relegated to the level of spoken thought (as it was by Ockham and Aurifaber), but is a property primarily of mental language. The result is that the grammar of mental language, on which spoken language is dependent, is a 'speculative science' (see Language, medieval theories of).

2 Theory of science

Both Marsilius and Buridan differ from Ockham in accepting a broader view of the formal object of a science. Whereas Ockham is more nominalistic (according to him the immediate formal object is a mental proposition), Buridan and Marsilius also accept a 'remote object' (the term) and a 'most remote object' (the thing) as significates. When trying to determine the formal object of theology, Marsilius agrees with Thomas Aquinas that a science has one single formal object, which Marsilius defines for theology as 'God as goal of man'. This formal object is in fact a predicate, which supposits for God (Marsilius' semantic approach again), and to which all other objects are related. Any other science has a formal object analogous to theology, such as syllogism in the case of logic.

3 Natural philosophy

The concept of impetus received much attention in the history of medieval physics. According to Marsilius, a projectile when thrown moves forward not because it was pushed by the air (which in its turn was moved by the thrower), which is faster than the natural downward movement of the thing thrown (this was Aristotle's opinion). Instead, the action of the thrower gives the projectile a kind of property (the 'impetus') that causes the movement and diminishes on its own. Here Marsilius follows Buridan in certain respects, rather than Ockham.

In the medieval commentaries on Aristotle's *On the Heavens* the problem of the spot on the moon occupied a prominent place. Theories about the origin of the spot had to reckon with the generally accepted theory of the simple nature of the heavenly bodies. According to Marsilius and other fourteenth-century Parisian philosophers, the moon is one of the ethereal planets (it was not earthlike, as Albert the Great thought). They denied that the moon's spot originated from outside (as Thomas Aquinas maintained). In a simple body like the moon, density and rarity follow primarily from the essence, Marsilius says, while the spot follows from it in a secondary sense, because the essence is dependent on the causality of the sun. In their different ways, these philosophers blurred the Aristotelian demarcation between heavenly bodies and the sublunary region (see Natural philosophy, medieval).

Marsilius also deals with two problems on the borderline between physics and theology. He holds that from the viewpoint of natural philosophy, there cannot be a plurality of worlds separate from each other. From the viewpoint of faith, however, God can create more worlds, in which natural kinds may differ from those in our world. Here Marsilius' view differs from that of most thirteenth-century philosophers. Marsilius believes that the world could not possibly be eternal, and in this he dissents from, for example Thomas Aquinas and Thomas of Strasbourg, who held that God's power to create an eternal world did not detract from his perfection and his own eternity (see Eternity of the world, medieval views of).

See also: Aristotelianism, medieval; Buridan, J. Logic, medieval; Natural philosophy, medieval; William of Ockham

E.P. BOS

List of works

Marsilius of Inghen (1362-7) Treatises on the Properties of Terms, ed. and trans. E.P. Bos, *Treatises on the Properties of Terms: A First Critical Edition of the Suppositiones, Ampliationes, Appellationes, Restrictiones and Alienationes*, Dordrecht: Reidel, 1983.(See this work for Marsilius' views on semantics.)

Marsilius of Inghen (1362-77) *Abbreviationes super octo libros Physicorum Aristotelis (Abbreviations on Aristotle's Physics)*, Venice: Octaviani Scoti Sociorum, 1521.(Includes Marsilius' impetus theory.)

Marsilius of Inghen (1362-77) *Quaestiones in De caelo (Questions on Aristotle's On the Heavens)*, in MS Cuyk en St Agatha, Kruisherenklooster, C 12, f. 123ra-171rb.(See for the problem of the spot on the moon and the possibility of more worlds.)

Marsilius of Inghen (1392-4) *Quaestiones super quattuor libros Sententiarum (Questions on the Four Books of Sentences)*, Strasbourg: Martin Flach, 1501; repr. Frankfurt: Martinus Flach, Jr, 1966.(A primary source for Marsilius' theory of science and his view of the eternity of the world.)

References and further reading

Braakhuis, H. and Hoenen, M. (eds) (1992) *Acts of the International Marsilius of Inghen Symposium*, Nijmegen: Ingenium.(Recent articles on various aspects of Marsilius' philosophy.)

Marsilius of Padua (1275/80-1342/3)

*Marsilius of Padua's Defensor pacis (Defender of Peace), written in 1324, is the most revolutionary political treatise of the later Middle Ages. Discourse One of the Defensor pacis can plausibly be read as a complete theory of the secular state. In a much longer second discourse, Marsilius attacks papal and priestly political power, which, especially in the claims to 'fullness of power' (*plenitudo potestatis*) sometimes made for the papacy, he presents as a major threat to civic tranquillity. The distinctive features of Marsilian theory are (a) its emphasis on broad participation in the legislative process as a guarantee of sound law, and (b) its insistence that supreme coercive power in any community must be held by a single, secular, popularly authorized 'ruling part' (*pars principans*).*

1 Life

Marsilius was born Marsilio dei Mainardini, son of a notary to the University of Padua. He lectured on natural philosophy and engaged in medical research and practice in Paris, where he was rector of the University in 1313. In 1326, when his authorship of the *Defensor pacis (Defender of Peace)* was discovered, he fled to the court of Ludwig of Bavaria, then at odds with the papacy over the need for papal approval of his election as emperor several years earlier. Marsilius defended Ludwig's authority as Roman emperor in the tract *De translatione imperii* (On the Transfer of the Empire), presumably written in 1326-7, through advice and action during Ludwig's Italian campaign of 1327-30, and in the *Defensor minor*, written by 1342. Marsilius was dead by April 1343.

2 The 'state' and its peace

The discussions of popular assemblies and civic affairs in Discourse One of the *Defensor pacis* are most easily understood in relation to the first of the two milieux in which Marsilius passed his political life, the northern Italian commune. From the beginning, however, Marsilius sought to construct a theory which could be used to resist papalist claims to power on any scale. Accordingly, the formal unit of his analysis is the *regnum*, which he defines, peculiarly, as 'something common to every species of temperate regime, whether in a single city or many' (*Defensor pacis* 1.2.2). The generic character of this definition, the dedication of the *Defensor pacis* to Ludwig of Bavaria, the Empire's history as a particular object of papal jurisdictional claims and the context of Marsilius' later life and writings give some support to an imperialist as well as a republican interpretation of his thought.

At any level, the aim of political association, according to Marsilius, is 'the sufficient life'. His exegesis of this Aristotelian concept is a comparatively modest one. Although he notes that those who live a civil life have leisure for the activities of the moral and intellectual virtues, the 'living well' of Aristotelian politics, Marsilius is less concerned than most medieval authors (for example, Thomas Aquinas) with orienting politics toward the perfect natural or supernatural fulfilment of human nature. He is more concerned with the factors producing or impeding the peaceful resolution of disputes and the cooperative functioning of those parts of a community which are devoted to meeting basic earthly needs. Thus, while he follows Aristotle (§27) in including the priesthood as one of the parts of a *regnum* (along with farmers, artisans, merchants and the like), he treats it as a part established and controlled by lay authority.

3 Legislation and government

The modest aim or final cause of Marsilian politics dictates an inclusive view of the parties qualified to participate as efficient causes in achieving that aim. To be sure, Marsilius is not a pure majoritarian. When he argues that the primary and proper efficient cause of the law is 'the people or the whole body of citizens, or the weightier part thereof', he immediately explains that by the 'weightier part' (*pars valentior*) he means to take into consideration both 'the quantity and the quality of the persons in that community over which the law is made' (*Defensor pacis* 1.12.3). Even with this qualification, however, Marsilius' conception of legislative authority is exceptionally populist for the period in which he wrote.

Marsilius follows Aristotle and medieval tradition in arguing that the rule of law is necessary for achieving civic justice, the common benefit and political stability. He departs from tradition (how sharply is controversial) in making coercive enforceability, rather than rational content, the essence of law. Hence, although the wisdom of

experts may be called on to formulate proposed legislation, it is the people's consent that gives these proposals the force of law. Citizen participation in the legislative process is important for Marsilius even on the score of knowledge, for he holds that the people can discern what is for their mutual benefit and are thus able to assess and improve the counsels of the wise few.

Popular authorization is also required for the establishment (and, if necessary, correction) of a government to enforce the law. Although Marsilius argued for elective monarchy as generally the best form of government, he left the choice of administration by one, few or many rulers to the people. Whatever its internal makeup, however, it is vital that the supreme government of a city or larger community operate as a numerical unity in relation to local jurisdictions (if any) and in establishing and regulating the other parts of the state mentioned above. If there were a plurality of governments, not reduced or ordered under one as supreme, 'the judgement, command and execution of matters of benefit and justice would fail, and because men's injuries would therefore be unavenged the result would be fighting, separation, and finally the destruction of the city or *regnum*' (*Defensor pacis* 1.17.3). Marsilius compared the cooperative functioning of the various parts of a temperate political community to the functioning of a healthy, well-ordered animal, in which 'the primary principle which commands it and moves it from place to place is one' (*Defensor pacis* 1.17.8). It is by impeding the operation of government, the political primary moving principle, that papal and other ecclesiastical interventions in civic affairs are inimical to peace.

4 Religion

Marsilius' insistence that all coercive power be concentrated in a popularly controlled unitary secular government was rightly seen as an attack on the characteristic legal dualism of medieval society, in which church courts had jurisdiction over clerics in all matters and over lay people in some, while civil courts had jurisdiction over the laity in the matters that remained. In Discourse Two of the *Defensor pacis*, Marsilius defended his own position as the authentically Christian one. He argued that, in contrast with recent popes, Christ and the apostles were not a threat to secular peace. While teaching what must be believed and done to attain blessedness in the future life, they acknowledged the authority of emperors and other secular rulers in this life. In this part of his argument, Marsilius incorporated contemporary Franciscan contentions that poverty and lack of political power were appropriate for those who wished to follow Christ most closely. However, the leading Franciscan thinker of the day, William of Ockham, sharply criticized some of Marsilius' own theses (for example, that Peter had no authority over the other apostles and that final authority for determining Christian doctrine rests with a general council). Marsilius' views on the relations between clerical and lay authority were influential in the sixteenth-century Protestant Reformation. His conception of popular participation as the source of all law and of all governmental authority has been seen as an anticipation of, if not a traceable influence on, Rousseau.

See also: Aquinas, T.; Aristotelianism, medieval; John of Jandun; Natural law; Sovereignty; William of Ockham

A.S. McGRADE

List of works

Marsilius of Padua (1324) *Defensor pacis (Defender of Peace)*, ed. C.W. Previté-Orton, Cambridge: Cambridge University Press, 1928; trans. A. Gewirth, vol. 2 of *Marsilius of Padua: The Defender of Peace*, New York: Columbia University Press, 1956; repr. with new introduction in Cambridge Texts in the History of Political Thought, Cambridge: Cambridge University Press, 1996. (Marsilius' major contribution to contemporary political debate and to political philosophy.)

Marsilius of Padua (1326-7?) *De translatione imperii (On the Transfer of the Empire)*, ed. C. Jeudy and J. Quillet in *Marsile de Padoue, Oeuvres Mineures*, Paris: Editions de Centre National de la Recherche Scientifique, 1979; trans. C.J. Nederman, *Marsiglio of Padua, Writings on the Empire*, Cambridge Texts in the History of Political Thought, Cambridge: Cambridge University Press, 1993. (A discussion of transfers of imperial authority attacking the pope's power to make them.)

Marsilius of Padua (by 1342) *Defensor minor*, ed. C. Jeudy and J. Quillet in *Marsile de Padoue, Oeuvres Mineures*, Paris: Éditions de Centre National de la Recherche Scientifique, 1979; trans. C.J. Nederman, *Marsiglio of Padua, Writings on the Empire*, Cambridge Texts in the History of Political Thought, Cambridge: Cambridge University Press, 1993.(*Defensor minor* is a summary of *Defensor pacis*, with a treatise on imperial jurisdiction in matrimonial cases.)

References and further reading

Gewirth, A. (1951) *Marsilius of Padua and Medieval Political Philosophy*, vol. 1 of *Marsilius of Padua: The Defender of Peace*, New York: Columbia University Press. (Unmatched in depth and analytic precision; the philosopher's best introduction to medieval political thought in general.)

Nederman, C.J. (1995) *Community and Consent: The Secular Political Theory of Marsiglio of Padua's Defensor Pacis*, Lanham, MD: Rowman and Littlefield. (Concise account of Discourse One, with the emphases indicated in the title. Bibliography, including references to selected studies in vols 5 and 6 (1979 and 1980) of *Medioevo: Rivista di Storia della Filosofia Medievale*, Padua: Editrice Antenore, which are entirely devoted to Marsilius.)

Quillet, J. (1970) *La Philosophie Politique de Marsile de Padoue*, Paris: Vrin. (Marsilius as primarily concerned with the Empire.)

Marston, Roger (*c*.1235-*c*.1303)

Roger Marston, an English Franciscan philosopher-theologian, was a pupil of John Pecham and a fellow student with Matthew of Aquasparta. Following closely in the footsteps of his master, Marston championed the views of Augustine in a conscious effort to counteract the growing fascination with Averroistic Aristotelianism. Of his works, three sets of Quaestiones disputatae (Disputed Questions) and four sets of quodlibetal questions, the Quodlibeta Quatuor (Four Quodlibets), survive.

Marston was born in England in the second quarter of the thirteenth century. As a Franciscan friar, he studied in Paris during the late 1260s and early 1270s, where he most likely heard the *collationes* of Bonaventure. He was a pupil of John Pecham during the latter's Paris regency (1270-1) and a fellow student with Matthew of Aquasparta. Upon returning to England and receiving the magisterium, he taught at Cambridge and Oxford between *circa* 1276 and 1284. He served as minister provincial of the English Franciscans from 1292 to 1298, when he was most instrumental in promoting the academic career of Duns Scotus. He died around 1303 and was buried at Norwich. Of Marston's works, while his commentary on the *Sentences* of Peter Lombard appears to have been lost, three sets of *Quaestiones disputatae (Disputed Questions)* - On Eternal Emanation, On the State of Fallen Nature and On the Soul - as well as four sets of quodlibetal questions, the *Quodlibeta Quatuor*, have survived.

Following closely in the footsteps of his master, John Pecham, Marston championed the views of Augustine in a conscious effort to counteract the growing Averroistic Aristotelianism (see Augustinianism; Averroism). In response to the charge of arch-conservatism, he remarked that he did not cling to tradition out of mere habit but, after a reasonable scrutiny of the evidence, formed opinions that harmonized the writings of the saints with the wisdom of the philosophers. Marston must have felt his views vindicated by the 219 propositions condemned by Etienne Tempier in 1277, since he refers to these 'articles' occasionally in his *Quaestiones disputatae* (see Aristotelianism, medieval).

Marston showed some interest in metaphysics, dismissing, for example, any real distinction between essence and existence. His concern with logic was predominantly with its application to other sciences. He was aware of the *ars obligatoria* (*Quodlibet* IV q.5: 375; *Quaestiones disputatae de anima* q.10: 453) (see Logic, medieval). Apart from purely theological questions and matters of canon law, his philosophical thinking focused on epistemology, natural philosophy and psychology (as understood in medieval times).

Regarding epistemology, Marston maintained that the material elements of cognition - for example, grasping the terms of a proposition - come from sense knowledge or the imagination and reside in the intellectual memory. The formal elements, however - in other words, the infallible and immutable evidence for the truth - come from what Augustine called the 'eternal reasons' or divine illumination. Like Roger Bacon, Marston claimed that this was the same as the agent intellect of Aristotle and Averroes (see Ibn Rushd), although in opposition to Averroes, he claimed that there was no single agent intellect for all mankind, but that each individual had his own. Marston claimed that 'species' - for example, 'images' or 'idols' remaining after sensation or intellection - were necessary for knowledge. Like Pecham and unlike Duns Scotus, Marston claims a direct intellectual knowledge of singulars.

In the realm of natural philosophy, Marston insists (contrary to Aquinas) that matter has its own essence and positive nature apart from any form. Marston is likewise an indefatigable defender of the plurality of grades of the form theory (*Quodlibet* II q.22: 232-78), in opposition to the 'uniformism' of Aquinas and the 'dimorphism' of Henry of Ghent. The vegetative and sensitive functions (grades of the form) are not supplanted by the infusion of the rational soul, as Aquinas held, but they persist as grades of the rational soul considered as the form of the human being (see Natural philosophy, medieval).

Like his teacher Pecham, Marston advanced 'proofs' for the immortality of the soul. In much the same fashion as his protégé Duns Scotus, Roger strongly defended the freedom of the will to the degree that choice could be exercised not simply regarding means to the final end, but even with regard to man's final goal as well. In the footsteps of Bonaventure and Pecham, Marston rejected even the possibility of a created world eternally coexistent with the creator.

Roger Marston is noteworthy for the clarity and well-organized presentation of his views. He is likewise an

invaluable witness to the doctrinal disputes prevalent at his time.

See also: Augustinianism; Duns Scotus, J.; Matthew of Aquasparta; Pecham, J.

GIRARD J. ETZKORN

List of works

Marston, R. (1276-84) *Quaestiones Disputatae (Disputed Questions)*, Bibliotheca Franciscana Scholastica Medii Aevi VII, Quaracchi: Collegii S. Bonaventurae, 1932.(There are three sets of questions, on eternal emanation, on the state of fallen nature, and on the soul.)

Marston, R. (before 1303) *Quodlibeta quatuor (Four Quodlibets)*, ed. G. Etzkorn and I. Brady, Bibliotheca Franciscana Scholastica Medii Aevi XXVI, 2nd edn, Grottaferrata: Colegii S. Bonaventurae, 1994. (Four sets of quodlibetal questions.)

References and further reading

Belmond, S. (1934) 'La theorie de la connaissance d'après Roger Marston' (The Theory of Knowledge According to Roger Marston), *La France Franciscaine* 17: 153-87.(Good introduction to Marston's theory of knowledge.)

Bonafede, G. (1939) 'Il problema del "lumen" in frate Ruggero di Marston' (The Problem of 'Light' According to Brother Roger Marston), *Rivista Rosminiana de Filosofia e di Cultura* 33: 16-30.(Marston on illumination.)

Cairola, J. (1951) 'L'opposizione a S.Tommaso nelle "Quaestiones disputatae" di Rugero Marston' (Opposition to St Thomas in the 'Disputed Questions' of Roger Marston), *Scholastica ratione historico-critica instauranda in Bibliotheca Pontificii Athenaei Antoniani VII*, Rome: Pontificium Athenaeum Antonianum, 447-60.(Marston's arguments against Aquinas.)

Daniels, A. (1911) 'Anselmzität bei dem Oxforder Franziskaner R. von Marston' (The Influence of St Anselm on the Oxford Franciscan R. Marston), *Theologische Quartelschrift* 93: 35-59.(A pioneering study on Anselm's influence.)

Etzkorn, G. (1962) 'The Grades of the Form according to Roger Marston, O.F.M.', *Franziskanische Studien* 44: 418-54.

Gilson, E. (1933) 'Roger Marston, un cas d'augustinisme avicennisant' (Roger Marston: A Case of Avicennizing Augustinianism), *Archives d'Histoire Doctrinale et Litteraire du Moyen Age* 8: 39-42.(An attempt to characterize historical influences on Marston.)

Glorieux, P. (1979) 'Marston (Roger) (d. 1303),' in *Catholicisme hier aujourd'hui demain VIII*, Paris: Letouzey et Ane, 724-5.(Short biography.)

Hissette, R. (1971) 'Les doctrines metaphysiques de Roger Marston' (The Metaphysical Doctrines of Roger Marston), unpublished dissertation, Louvain.(One of the few studies of Marston's metaphysics.)

Hissette, R. (1972) 'Roger Marston, a-t-il professe l'hylemorphisme universel?' (Did Roger Marston Espouse Universal Hylomorphism?), *Recherches de Théologie Ancienne et Médiévale* 29: 205-23.(On Marston's metaphysics.)

Hissette, R. (1980) 'Esse-essentia chez Roger Marston' (Existence and Essence According to Roger Marston), in *Sapientiae doctrina. Mélange de théologie et de litterature médiévales offerts à Dom Hildebrand Bascour, O.S.B.*, Louvain: Imprimerie Orientaliste, 110-8.(On Marston's views of essence and existence.)

Pelster, F. (1928) 'Roger Marston, O.F.M. (d. 1303), ein englischer Vertreter des Augustinismus' (Roger Marston, OFM (d. 1303), An English Champion of Augustinianism), *Scholastik* 3: 526-56.(An attempt to characterize Augustine's influence.)

Prezioso, F. (1950) 'L'attivita del soggetto pensante nella gnoseologia di Matteo d'Acquasparta e di Ruggiero Marston' (The Activity of the Thinking Subject in the Epistemology of Matthew of Aquasparta and Roger Marston), *Antonianum* 25: 259-326.(A good study of Marston's theory of knowledge.)

Martineau, Harriet (1802-76)

Harriet Martineau has been called the first woman sociologist and the first woman journalist in England, both better claims on the attention of posterity than her mostly derivative philosophical writings. Yet she is a revealing - and was in her own time widely influential - instance of the survival of eighteenth-century determinism and materialism. Although she eventually rejected the Unitarianism in which she had been brought up, the Necessarian philosophy she drew from it merged easily into her mature positivism. Her abridged translation (1853) of theCours de philosophie positive was the first introduction of this seminal work byAuguste Comte to English-speaking readers.

Harriet Martineau was born in Norwich on 12 June 1802. As a child, she suffered ill health and more than the usual fears and loneliness; by the age of 20 she was almost totally deaf. Her excellent, if formally haphazard, early education was extended by carefully planned reading, with strong encouragement from a family solidly established among the liberal Unitarians in Norwich and elsewhere.

Devotional and religious themes marked her early writing - mostly in the Unitarian *Monthly Repository* - but she turned increasingly to philosophical issues. Entirely dependent on her pen after the death of her father, a textile manufacturer, in 1826, and the failure of the family firm three years later, in 1832 she launched *Illustrations of Political Economy*, a monthly series of 23 tales which caught, as she did so often, a new enthusiasm at its height, bringing her a celebrity she never lost. On completion of the tales, she spent the years 1834-6 in the USA. *Society in America* (1837), the first of two accounts of her travels, shows formidable powers of observation and synthesis and contains much original sociological comment, while in *How to Observe: Morals and Manners* (1838), she set out general principles for social analysis. *Deerbrook* (1839), her only large-scale novel, offered a characteristically sympathetic portrayal of middle-class life and enforced the subordination of passion to duty and service to others. Extraordinarily generous with her own modest income, she advocated many causes, among them anti-slavery, education, and the legal, political and economic liberation of women. Her radicalism broadened as she grew older and the stringency of her early political economy lessened.

Stricken in 1840 by debilitating symptoms resulting from an ovarian cyst, she retreated to Tynemouth, where her writing continued apace and her thinking matured, with some oddly self-absorbed turns. In 1845, she pronounced herself cured by mesmerism and resumed an active life, eventually settling in the Lake District. An extended tour of the Near East, recorded in *Eastern Life, Present and Past* (1848), further distanced her from religion, and *Letters on the Laws of Man's Nature and Development*, written with Henry George Atkinson in 1851, proclaimed her loss of conventional belief. A brilliant, opinionated critic, a gifted contemporary historian and a keen observer of politics and international affairs, in 1852 she began to write regularly for the London *Daily News*, contributing more than 1,600 leading articles before her retirement in 1866. She died in Ambleside on 10 April 1876.

Harriet Martineau was no mere popularizer. Rather, insistent on candid 'publication of opinion' and determined to instruct the nation, she gathered together many of the major intellectual strands of her time, forcing them into broad public awareness at crucial moments.

Like most serious-minded English Unitarians of her generation, she was persuaded by Necessarianism, a deterministic philosophy derived from the psychological and religious speculations of David Hartley (§2), the systematizer of associationism, and transmitted through the scientist and theologian Joseph Priestley, from whom she also drew her materialism. Convinced that free will is an illusion, Necessarians distinguished their views from fatalism or Calvinist predestinarianism by insisting on the malleability of motives which, built up through association, inescapably determined all human actions. Education and personal influence could thus bring everyone to think rightly, to understand the laws of nature and even to attain perfection. This optimism and certainty survived her firm, unfairly scornful rejection of the Unitarianism that had engendered it.

Open to a wide range of current discussion, she added her own observations and experience, including mesmeric phenomena, to create a scientistic synthesis in which Baconian faith in fact-based induction was conditioned by insistence on the importance of guiding, selective principle. As religion came to seem no more than a historically conditioned phenomenon, she turned to positivism, for which she had been prepared by the Saint-Simonian

missionaries who came to London from France in the early 1830s. In 1853, she published an abridged translation of Auguste Comte's *Cours de philosophie positive*, its first introduction to the English-speaking world. Although she rejected social hierarchy, the subordination of women and the quasi-religious turn ofComte's later work, she found in the centrality he gave to laws of thought and to scientific views of society the final distillation of her Necessarianism (see Comte, A.).

She had shared that early commitment with her younger brother James, a Unitarian minister who, from 1833, was in rebellion against Priestleyanism. Influenced by German theology and philosophy and by romanticism, he insisted that apprehension of religious truth was grounded in introspection rather than in revelation and evidences from nature. Their increasing alienation, culminating in his extremely hostile review of the Atkinson letters in the *Prospective Review*, made her a willing conspirator in the successful campaign in 1866 (led by the psychologist Alexander Bain) to deny him, as a minister, the chair of philosophy at the historically non-sectarian University College London, a post many thought a merited recognition of his periodical writings and his lectures at the Unitarian Manchester College. The lectures became books in the 1880s, too late to have the influence they might have commanded twenty years earlier.

The posthumous publication of her frank and judgmental autobiography inflicted considerable damage on her reputation, which, though she never lacked admirers, only began a real recovery in the 1960s, when historians, literary critics and feminists began to take proper measure of a career of astonishing energy and accomplishment.

See also: Positivism in the social sciences §1

R.K. WEBB

List of works

Martineau, H. (1829) 'Essays on the Art of Thinking', *Monthly Repository* 3: 521-6, 599-606, 707-12, 747-57, 817-22, reprinted in 1836 in H. Martineau, *Miscellanies*, Boston, MA, 2 vols.(Early summary of philosophical views.)

Martineau, H. (1832-4) *Illustrations of Political Economy*, London: Charles Fox, 23 vols in 25.(Tales exemplifying economic principles.)

Martineau, H. (1837) *Society in America*, London: Saunders & Otley, 2 vols.(Distinguished and original analysis.)

Martineau, H. (1838) *How to Observe: Morals and Manners*, London: Charles Knight. (General principles of social analysis.)

Martineau, H. (1839) *Deerbrook*, London: Moxon, 3 vols.(Overlong, didactic, but revealing novel.)

Martineau, H. (1844) *Life in the Sickroom: Essays by an Invalid*, London: Moxon.(Self-conscious introspection suggestive of larger evolution of her thought.)

Martineau, H. (1848) *Eastern Life: Present and Past*, London: Moxon, 3 vols.(Perceptive and revealing travel account of particular importance for evolving religious views.)

Martineau, H. and Atkinson, H.G. (1851) *Letters on the Laws of Man's Nature and Development*, London: Chapman.(Extravagant avowal of mesmerism and free thought.)

Martineau, H. and Atkinson, H.G. (1877) *Harriet Martineau's Autobiography, with Memorials by Maria Weston Chapman*, London: Smith, Elder, 3 vols.(Despite its idiosyncrasies, one of the great autobiographies of the century.)

Martineau, H. and Atkinson, H.G. (1983) *Harriet Martineau's Letters to Fanny Wedgwood*, ed. E.S. Arbuckle, Stanford, CA: Stanford University Press.(Extended series of charming, revealing letters.)

Martineau, H. and Atkinson, H.G. (1985) *Harriet Martineau on Women*, ed. G.G. Yates, New Brunswick, NJ: Rutgers University Press.(Illustrates and assesses Martineau's central, somewhat ambiguous place in feminist history.)

Martineau, H. and Atkinson, H.G. (1990) *Harriet Martineau: Selected Letters*, ed. V. Sanders, Oxford: Clarendon Press.(Representative sampling of a vast output by a brilliant correspondent.)

Martineau, H. and Atkinson, H.G. (1993) *Harriet Martineau and the Daily News*, ed. E.S. Arbuckle, New York: Garland.(Selection of leading articles.)

References and further reading

Comte, A. (1853) *The Positive Philosophy of Auguste Comte, freely translated and condensed by Harriet Martineau*, London: Chapman, 2 vols.(Martineau's translation brought Comte's positivism to the

English-speaking world.)
Hoecker-Drysdale, S. (1992) *Harriet Martineau: First Woman Sociologist*, London: Berg.(Brief, biographically structured, analysis of Martineau's sociology, with excellent bibliographies.)
Sanders, V. (1986) *Reason over Passion: Harriet Martineau and the Victorian Novel*, Hassocks: Harvester Press. (Extensive and balanced discussion of Martineau as a writer of fiction.)
Webb, R.K. (1960) *Harriet Martineau, a Radical Victorian*, London; Heinemann. (Primarily an intellectual biography, stressing Unitarian and Necessarian sources.)

Marx, Karl (1818-83)

Karl Marx was the most important of all theorists of socialism. He was not a professional philosopher, although he completed a doctorate in philosophy. His life was devoted to radical political activity, journalism and theoretical studies in history and political economy.

Marx was drawn towards politics by Romantic literature, and his earliest writings embody a conception of reality as subject to turbulent change and of human beings as realizing themselves in the struggle for freedom. His identification with these elements in Hegel's thought (and his contempt for what he regarded as Hegel's apologetic attitude towards the Prussian state) brought Marx to associate himself with the Young Hegelians.

The Young Hegelians had come to believe that the implicit message of Hegel's philosophy was a radical one: that Reason could and should exist within the world, in contrast to Hegel's explicit claim that embodied Reason already did exist. Moreover, they also rejected Hegel's idea that religion and philosophy go hand in hand: that religion represents the truths of philosophy in immediate form. On the contrary, the Young Hegelians saw the central task of philosophy as the critique of religion - the struggle (as Marx himself was to put it in his doctoral dissertation) 'against the gods of heaven and of earth who do not recognize man's self-consciousness as the highest divinity'.

Marx came to be dissatisfied with the assumption that the critique of religion alone would be sufficient to produce human emancipation. He worked out the consequences of this change of view in the years 1843 to 1845, the most intellectually fertile period of his entire career. Hegel's philosophy, Marx now argued, embodied two main kinds of mistake. It incorporated, first, the illusion that reality as a whole is an expression of the Idea, the absolute rational order governing reality. Against this, Marx's position (and on this point he still agreed with the Young Hegelians) was that it is Man, not the Idea, who is the true subject. Second, he charged, Hegel believed that the political state - the organs of law and government - had priority in determining the character of a society as a whole. In fact, according to Marx, this is the reverse of the truth: political life and the ideas associated with it are themselves determined by the character of economic life.

Marx claimed that the 'species-being' of Man consists in labour, and that Man is 'alienated' to the extent that labour is performed according to a division of labour that is dictated by the market. It is only when labour recovers its collective character that men will recognize themselves as what they are - the true creators of history. At this point, the need to represent the essence of human beings in terms of their relation to an alien being - be it the Christian God or Hegelian Geist *- will no longer exist.*

In the mature writings that followed his break with the Young Hegelians, Marx presented a would-be scientific theory of history as a progress through stages. At each stage, the form taken by a society is conditioned by the society's attained level of productivity and the requirements for its increase. In pre-socialist societies this entails the division of society into antagonistic classes. Classes are differentiated by what makes them able (or unable) to appropriate for themselves the surplus produced by social labour. In general, to the extent that a class can appropriate surplus without paying for it, it is said to be an 'exploiting' class; conversely, a class that produces more than it receives is said to be 'exploited'.

Although the exploiting classes have special access to the means of violence, exploitation is not generally a matter of the use of force. In capitalism, for example, exploitation flows from the way in which the means of production are owned privately and labour is bought and sold just like any other commodity. That such arrangements are accepted without the need for coercion reflects the fact that the ruling class exercises a special influence over ideas in society. It controls the ideology *accepted by the members of society in general.*

In Das Kapital (Capital), the work to which he devoted the latter part of his life, Marx set out to identify the 'laws of motion' of capitalism. The capitalist system is presented there as a self-reproducing whole, governed by an underlying law, the 'law of value'. But this law and its consequences are not only not immediately apparent to the agents who participate in capitalism, indeed they are actually concealed from them. Thus capitalism is a 'deceptive object', one in which there is a discrepancy between its 'essence' and its 'appearance'.

In Marx's view, it is inevitable that capitalism should give way to socialism. As capitalism develops, he believed,

the increasingly 'socialized' character of the productive process will conflict more and more with the private ownership of the means of production. Thus the transition to collective ownership will be natural and inevitable. But Marx nowhere explained how this collective ownership and social control was to be exercised. Indeed, he had remarkably little to say about the nature of this society to the struggle to which he devoted his life.

The Critique of the Gotha Programme envisaged two phases of communist society. In the first, production will be carried out on a non-exploitative basis: all who contribute to production will receive back the value of what they have contributed. But this, Marx recognized, is a form of 'equal right' that leaves the natural inequalities of human beings unchecked. It is a transitional phase, although inevitable. Beyond it there lies a society in which individuals are no longer 'slaves' to the division of labour, one in which labour has become 'not only a means of life but life's prime want'. Only then, Marx thought, 'can the narrow horizon of bourgeois right be crossed in its entirety and society inscribe on its banners: from each according to his ability, to each according to his needs!' This is the final vision of communism.

1 Life and works

Marx was born on 5 May, 1818, in Trier, a small, originally Roman city on the river Moselle. Many of Marx's ancestors were rabbis, but his father, Heinrich, a lawyer of liberal political views, converted from Judaism to Christianity and Marx was baptized with the rest of his family in 1824. At school, the young Marx excelled in literary subjects (a prescient schoolteacher comments, however, that his essays were 'marred by an exaggerated striving after unusual, picturesque expression'). In 1835, he entered the University of Bonn to study law. At the end of 1836, he transferred to Berlin and became a member of the Young Hegelian *Doktorklub*, a bohemian group whose leading figure was the theologian, Bruno Bauer. The views of the *Doktorklub* became increasingly radical (to some extent, it would seem, under Marx's influence) in the late 1830s.

Marx's father died in 1838 and in the next year - perhaps not coincidentally - Marx abandoned the law in favour of a doctorate in philosophy. His thesis, *Differenz der demokritischen und epikureischen Naturphilosophie (Difference between the Democritean and Epicurean Philosophy of Nature)* was accepted by the University of Jena in 1841. Marx had hoped to use it to gain an academic position, but, after Bruno Bauer's suspension from his post at the University of Bonn, it became apparent that such hopes would have to be abandoned in the current political climate.

Marx turned instead to journalism, involving himself with the newly-founded *Rheinische Zeitung* and taking over the editorship in October 1842. However, the paper came increasingly into conflict with the Prussian government and was banned in March 1843. At this point, Marx decided to move abroad. In the summer he married Jenny von Westphalen (after an engagement of six years) and during a long honeymoon in Kreuznach worked on *Zur Kritik der Hegelschen Rechtsphilosophie (Critique of Hegel's Philosophy of Right)* and the essay 'Zur Judenfrage' ('On the Jewish Question') in which he started to formulate his disagreements with his fellow Young Hegelians. He and Jenny moved to Paris in October of that year. It was in 1844 that Marx met up again with Friedrich Engels (whom he had known slightly in Berlin) and the alliance was formed that was to last for the rest of Marx's life. Together Marx and Engels wrote *Die Heilige Familie (The Holy Family)* (1845a), a polemic against Bruno Bauer. More important, however, was the body of writing on economics and philosophy that Marx produced at this time, generally known as *The Paris Manuscripts* (1844).

Marx was expelled from France in 1845 and moved to Brussels. In the spring of 1845, he wrote for his own clarification a series of essays on Feuerbach. These 'Theses on Feuerbach' are one of the few mature statements we have of his views on questions of epistemology and ontology. In 1845-6 Marx and Engels wrote *Die deutsche Ideologie (The German Ideology)* which, although it too remained unpublished, contains an authoritative account of their theory of history and in particular of the place of ideas in society. Marx's developing economic views were given expression in a polemic against Proudhon, *La Misère de la Philosophie (The Poverty of Philosophy)*, published in 1847.

Das Kommunistische Manifest (The Communist Manifesto), written by Marx and Engels as the manifesto of the Communist League in early 1848, is the classic presentation of the revolutionary implications of Marx's views on history, politics and economics. During the revolutionary upsurge of 1848 Marx returned to Germany, but with the defeat of the revolutionary movement he was forced to leave, first for Paris, and then, in August 1849, for London,

where he would live in exile for the rest of his life.

The years of exile in Britain were difficult ones for Marx (and even more so for his loyal and devoted family). He was in constant financial difficulty and had to rely heavily on Engels and other friends and relations for support. His theoretical activities were chiefly directed to the study of political economy and the analysis of the capitalist system in particular. They culminated in the publication of the first volume of *Das Kapital (Capital)* in 1867. However, *Das Kapital* is the tip of a substantial iceberg of less important publications and unpublished writings. Among the former, the Preface to *Zur Kritik der politischen Ökonomie (A Contribution to the Critique of Political Economy)* published in 1859, contains the classic statement of Marx's materialist theory of history. The second and third volumes of *Das Kapital*, left unfinished at Marx's death, were edited and published posthumously by Engels. In addition, three volumes of *Theorien über den Mehrwert (Theories of Surplus-Value)*, a series of critical discussions of other political economists, written in 1862-3, were published in the early twentieth century. An extensive and more or less complete work, the *Grundrisse der Kritik der politischen Ökonomie* (known both in English and in German as the *Grundrisse*) was written in 1857-8 but only published in 1939. The Introduction to the *Grundrisse* is the mature Marx's most extended discussion of the method of political economy. In addition, there exist numerous notebooks and preliminary drafts, many (if not, at the time of writing, all) of which have been published.

Political economy apart, Marx wrote three works on political events in France: *Die Klassenkämpfe in Frankreich (Class Struggles in France)* (1850), *Das achtzehnte Brumaire des Louis Bonaparte (The Eighteenth Brumaire of Louis Bonaparte)* (1852) and *The Civil War in France* (1871). Among his many polemical writings, the *Kritik des Gothaer Programms (Critique of the Gotha Programme)* (1875) is particularly important for the light it throws on Marx's conception of socialism and its relation to ideas of justice.

Marx was in very poor health for the last ten years of his life, which seems to have sapped his energies for large-scale theoretical work. However, his engagement with the practical details of revolutionary politics was unceasing. He died on 14 March 1883 and is buried in Highgate Cemetery, London.

2 Marx as a Young Hegelian

Marx is relevant to philosophy in three ways: (1) as a philosopher himself, (2) as a critic of philosophy, of its aspirations and self-understanding, and (3) by the philosophical implications of work that is, in Marx's own understanding of it, not philosophical at all. Broadly speaking, these three aspects correspond to the stages of Marx's own intellectual development. This and the following section are concerned with the first stage.

The Young Hegelians, with whom Marx was associated at the beginning of his career, did not set out to be critics of Hegel. That they rapidly became so has to do with the consequences they drew from certain tensions within Hegel's thought. Hegel's central claim is that both nature and society embody the rational order of *Geist* (Spirit). Nevertheless, the Young Hegelians believed, it did not follow that all societies express rationality to the fullest degree possible. This was the case in contemporary Germany. There was, in their view, a conflict between the essential rationality of *Geist* and the empirical institutions within which *Geist* had realized itself: Germany was 'behind the times' (see Hegel, G.W.F. §§5-8; Hegelianism §§2-3).

A second source of tension lay in Hegel's attitude towards religion. Hegel had been prepared to concede a role to religion as the expression of the content of philosophy in immediate form. The Young Hegelians, however, argued that the relationship between the truths of philosophy and religious 'representation' was, in fact, antagonistic. In presenting reality not as the embodiment of reason but as the expression of the will of a personal god the Christian religion establishes a metaphysical dualism that is quite contrary to the secular 'this-worldliness' which (although Hegel himself might have been too cautious to spell it out fully) is the true significance of Hegel's philosophy.

This was the position endorsed by Marx at the time of his doctoral dissertation on Epicurus and Democritus. Its subject was taken from a period of Greek thought that displayed parallels with the Germany of Marx's own time. Just as the Young Hegelians faced the problem of how to continue philosophy after Hegel, so Epicurus wrote in the shadow of another great system, that of Aristotle. Epicurus is more successful than Democritus, Marx believes, in combining materialism with an account of human agency. Furthermore, Marx admires Epicurus for his explicit critique of religion, the chief task of philosophy, he asserts, in all ages.

In its destruction of the illusions of religion, the Young Hegelians believed that philosophy would provide both the necessary and the sufficient conditions for human emancipation and the achievement of a rational state. In the works that he wrote in Kreuznach in 1843 (the unpublished draft of the *Critique of Hegel's Philosophy of Right* and the essay 'On the Jewish Question') and shortly thereafter (the '*Critique of Hegel's Philosophy of Right*: Introduction') Marx called this position into question.

In the *Critique of Hegel's Philosophy of Right* Marx makes two main criticisms of Hegel. The first is that Hegel's real concern is to retrace in the political realm the outlines of his own metaphysics, rather than to develop an analysis of political institutions and structures in their own right. This gives his political philosophy an apologetic function, for it leads him to present the contradictions that he finds in reality as essentially reconciled in the supposedly higher unity of the 'Idea'. But they are not, says Marx. On the contrary, they are 'essential contradictions'.

Chief among such contradictions is that existing between the 'system of particular interest' (the family and civil society - that is, economic life) and the 'system of general interest', namely, the state. And this leads to Marx's second criticism. Hegel, Marx alleges, assumes that the state, because it is 'higher' from the point of view of Hegelian logic, can effectively reconcile the contradictions of economic life. In fact, in Marx's view, it is civil society that exists prior to the state. The state arises from the condition of civil society and is always subordinate to the form of the latter.

3 Philosophy and the critique of religion

Marx presents the implications of these criticisms for the critique of religion in the *Critique of Hegel's Philosophy of Right*: 'Introduction'. This short essay is a compressed masterpiece of vehement rhetoric, seething with antithesis and chiasmus. In Germany, Marx writes, 'the critique of religion is essentially completed'. Thus the problem is how to go beyond it. Marx's first step is to explain the significance of that critique, as he understands it.

The world of religion is a reflection of a particular form of society: 'This state, this society, produce religion, which is an inverted world-consciousness, because they are an inverted world'. That is to say, only an inverted, secular world would produce religion as its offshoot. In religious belief, Man finds himself reflected in the 'fantastic reality of heaven', whilst he can find only 'the semblance of himself, only a nonhuman being' in this world. Religion thus provides a realm in which individuals can realize themselves, at least partially, given that full and adequate self-realization is not possible in the profane world. In this way, religion preserves the social order of which it is a by-product, both by deflecting attention from its defects and by providing a partial escape from it. In Marx's famous words, 'Religion is the sigh of the oppressed creature, the heart of a heartless world and the soul of soulless conditions. It is the opium of the people'.

Thus religion and the form of life associated with it are open to criticism at three points. (1) There is, first, the impoverished and distorted world of which religion is a by-product. (2) There is the way in which the image of reality produced by religion is falsely transfigured. (3) Finally, there is the failure by human beings to recognize the fact that religion has its origins in mundane reality.

It is this last element towards which the critique of religion is directed. Critique of religion connects religion back to its unacknowledged origins in social existence. Yet this is not enough. The critique of religion, inasmuch as it is a call to people to abandon their illusions, is also, according to Marx, 'the call to abandon a condition that requires illusions'. By itself the critique of religion cannot remove the distortion and impoverishment of the world from which religion arises. This is of course Marx's real project, for which the criticism of religion has merely prepared the ground.

Once the criticism of religion has done its work, philosophy must move on 'to unmask human self-alienation in its secular forms'. The critique of religion ends, Marx says, 'in the doctrine that man is the supreme being for man; thus it ends with the categorical imperative to overthrow all conditions in which man is a debased, enslaved, neglected, contemptible being' (1843a: 251).

Much of this analysis represents common ground between Marx and his Young Hegelian former associates. Marx concedes that philosophy has both a critical role to play in exposing the illusions of religion and an affirmative one

in establishing an ideal of human fulfilment. Nevertheless, Marx takes the Young Hegelians to task for thinking that philosophy alone provides a sufficient condition for human emancipation. Philosophy, he maintains, must move beyond itself: ‘criticism of the speculative philosophy of right does not remain within itself, but proceeds on to tasks for whose solution there is only one means - praxis’. For this, a material force - a ‘class with radical chains’ - is required, namely, the proletariat.

At this stage, then, Marx is critical not so much of the content of philosophy, but of what we might call the metaphilosophical belief associated with it: that it is possible (as he puts it in relation to the Young Hegelians) ‘to realize philosophy without transcending it’. A truly successful critique of religion would require the transformation of the social conditions within which religion is generated and sustained.

4 Alienated labour

In Paris, Marx threw himself into the study of political economy. His objective was to amplify his critique of Hegel and the Young Hegelians with a more far-reaching account of the nature of ‘civil society’. The *Paris Manuscripts* thus provide a unique link between Marx’s economic theory and his philosophical view of human nature. The concept which brings the two together is that of alienation (*Entfremdung*) (see Alienation §§3-5). Although Marx had made little use of this term in his earlier writings, the structure of the concept is clearly anticipated in his critique of religion. The fundamental idea is that an entity or agent gives rise to a product or expression that is distinct from but at the same time essential to itself. This secondary product comes to be cut off from its origin. In consequence, the agent suffers a loss of identity in some sense. Thus, for the agent to realize itself fully, it must remove the separation that has come between itself and its own product.

In the central discussion of the *Paris Manuscripts*, Marx sets out to apply the concept of alienation to the labour process. Alienation, Marx argues, is characteristic of a situation in which (1) labour is directed towards the production of commodities (that is, goods exchangeable in the market) and (2) labour itself is such a commodity. Marx divides the alienation involved in labour into three main forms.

(1) There is, first, the separation of the worker from the product of labour. It is in the nature of the labour process that it involves ‘appropriating’ the external world. But when labour is alienated, the sensible, external world becomes an object to which the worker is bound, something that is hostile to them, instead of being the means to their self-realization.

(2) At the same time, the labour process itself becomes alien to the worker. Because the imperatives according to which labour takes place come to the worker ‘from outside’ (that is, from the market, either directly or indirectly) labour is no longer an act of self-realization. It becomes, from the worker’s point of view, ‘an activity directed against himself, which is independent of him and does not belong to him’.

(3) Finally, Marx says, the consequence of these two forms of alienation is to alienate man from what he calls his ‘species-being’ (*Gattungswesen*). The latter concept (of which Marx made frequent use in 1843-4) is adapted from Ludwig Feuerbach. Man, says Marx, is a species-being ‘because he looks upon himself as the present, living species, because he looks upon himself as a *universal* and therefore free being’.

An analogy that may help to clarify this apparently circular definition can be made with the family. In a limited sense, people can be part of a family without consciously behaving accordingly (at the limit, we can think of members of a family who do not even know that they are related). But in order to be a family in a fuller sense, people must relate to one another *as* a family, and at least a part of this is that they should be aware that they *are* a family. So it is with human species-being. While the fundamental phenomenon on which the family is based is a biological relation, in human species-being it is labour. Thus, as labour is alienated in other respects, so people become alienated from their species-being. The consequence is the alienation of members of the species from one another.

Each of these three points is, one might think, somewhat questionable. Surely, in any situation in which individuals do not produce entirely for themselves, it will be inevitable that the products of labour are ‘separated’ from the original producer. Likewise, the labour process cannot be something that is freely chosen by individuals as long as they are objectively constrained by the nature of the material world and the resources available to them in finding efficient means to given ends. Finally, it is not at all clear what is involved in human

beings ‘re-appropriating’ their ‘species-being’.

One way of making the concept of alienated labour more precise is to ask what it might be for labour to be non-alienated. Marx addresses the issue at the end of a discussion of James Mill’s *Elements of Political Economy*. ‘Let us suppose’, Marx begins, ‘that we had produced as human beings’. In that case, he claims, each of us would have ‘affirmed’ both ourselves and our fellows in the process of production. In the first place, I, the producer, would have affirmed myself in my production. At the same time, I would be gratifying a human need - that of my neighbour, for whom I am in this case producing. Thus, in meeting your need, I would have mediated between you and the species: ‘I would be acknowledged by you as the complement of your own being, as an essential part of yourself’. In this way, production and the meeting of needs involves a mutuality of self-realization and reciprocal recognition:

> In the individual expression of my own life I would have brought about the expression of your life, and so in my individual activity I would have directly *confirmed* and *realized* my authentic nature, my *human*, *communal* nature.
>
> (1844: 277-8)

These ideas help to explain Marx’s antagonism towards what he would call ‘bourgeois’ political theory. In so far as traditional political philosophy takes as its fundamental question how to reconcile competing interests, its starting point is, from Marx’s point of view, unacceptably individualistic. For what entitles us to assume that the interests of individuals are bound to be antagonistic? Rather than asking how to allocate rights and duties fairly when interests conflict, the task, Marx believes, is to move humanity towards a form of life in which conflicts of interest are no longer endemic.

5 The critique of philosophy

Although the *Paris Manuscripts* show Marx’s increasing engagement with political economy, they do not represent an abandonment of his concern with philosophy. The attitude that Marx takes towards philosophy, however, now becomes more critical than it had been in his earlier, Young Hegelian period. In part, this can be traced to Ludwig Feuerbach, whom Marx quotes approvingly at several points (see Feuerbach, L. §2). It was Feuerbach’s great achievement, Marx writes, ‘to have shown that philosophy is nothing more than religion brought into thought and developed in thought, and that it is equally to be condemned as another form and mode of existence of the alienation of human nature’. Thus Marx now regards philosophy as essentially continuous with religion, not a force directed against religion, as he had represented it at the time of his doctoral dissertation.

Marx makes a number of negative remarks regarding philosophy in general, but his more specific critical comments are directed towards Hegel. Like Feuerbach, he takes the view that Hegel has brought philosophy to a point of completion. The dynamic principle at the heart of Hegel’s philosophy, according to Marx, is that of ‘abstract mental labour’. Nevertheless, despite the genuinely critical elements contained within it, Hegel’s philosophy is vitiated by its idealist assumptions. In the end, for Hegel, alienation is merely a matter of the separation of the products of thought from thought itself, something to be overcome by a philosophical reorientation of consciousness. To go beyond Hegel, it would be necessary to make the concept of real, concrete labour fundamental. But this, Marx suggests, leads beyond philosophy itself.

Marx pursues these ideas in the ‘Theses on Feuerbach’, written in the spring of 1845. Here he makes it explicit that his disagreement is not only with idealistic philosophies, such as Hegel’s, but also with would-be materialist ones, Feuerbach’s included. In incorporating within itself an idea of ‘activity’, idealism has important advantages over materialism. It is, Marx writes,

> the chief defect of all hitherto existing materialism (that of Feuerbach included)…that the thing, reality, sensuousness, is conceived only in the form of the object or of contemplation, but not as sensuous human activity, praxis, not subjectively. Hence, in contradistinction to materialism, the *active* side was developed abstractly by idealism - which, of course, does not know real sensuous activity as such.
>
> (1845b: 421)

It should be noted that this passage is ambiguous. Is Marx envisaging a new kind of materialism (one that would not have the defects of ‘hitherto existing materialism’) or is it a call to leave philosophy - both materialism and

idealism - behind altogether? Interpreters of Marx who take the former view have ascribed an implicit philosophical position to him (often called ‘dialectical materialism’). Nevertheless, the fact remains that Marx himself never developed such a position explicitly, and the conclusion of the ‘Theses on Feuerbach’ appears to lead away from philosophy entirely: ‘The philosophers have only *interpreted* the world in various ways; the point is to *change* it.’

The German Ideology, which Marx and Engels wrote from September 1845 to the summer of 1846, continues this line of argument. As in so many of Marx’s writings, the rhetorical trope from which the criticism starts is that of an inversion of an inversion. The Young Hegelians, Marx alleges, think of themselves as engaged in a struggle with the illusions that hold the Germans in their grip. But in fact they are in the grip of an illusion themselves: the illusion that ideas are an independent, determining force in political life. Feuerbach is not excepted from this criticism. Although he purports to demystify the realm of pure ideas, he still remains, according to Marx and Engels, ‘in the realm of theory’. Feuerbach, they claim, ‘never arrives at really existing, active men, but stops at the abstraction "man"’.

The alternative that Marx and Engels propose is, of course, also a theory, but it is a theory, they claim, of a quite different kind. ‘In direct contrast to German philosophy, which descends from heaven to earth’, their purpose is to present an account which will ‘ascend from earth to heaven’. Instead of translating general ideas back into equally general anthropological categories, the aim is to give a specific account of their historical origins. In so doing, it undermines the presuppositions on which the philosophical enterprise rests and philosophy, as an independent branch of knowledge, loses its medium of existence:

> The philosophers would only have to dissolve their language into the ordinary language, from which it is abstracted, to recognize it as the distorted language of the actual world, and realize that neither thoughts nor language in themselves form a realm of their own, that they are only *manifestations* of actual life.
>
> (1845-6: 118)

6 The theory of ideology: (1) The reflection model

The German Ideology is filled with polemical assertions of the priority of material life over the world of religion, thought and speculation. But it sets out to do more than sloganize. Its aim is to develop the framework for a scientific explanation of how the material life conditions and determines thought and culture. By the time *The German Ideology* came to be written, the term ‘ideology’ had established itself in German as referring to systems of ideas detached from and out of proportion to empirical reality (Heinrich Heine, with whom Marx was on intimate terms in Paris, used it in that sense). In *The German Ideology* this is certainly part of the meaning of the term. But the concept also has a wider explanatory function (see Ideology).

Since the ancient world, political thinkers had been concerned with the role that ‘false’ or irrational forms of consciousness play in political life. To this extent, the Young Hegelian critique of religion represented the latest manifestation of a very long tradition. However, the originality of Marx’s concept of ideology lies in the way that it brings the idea of false consciousness together with a distinctively modern conception of society.

At the end of the eighteenth and the beginning of the nineteenth century, a conception of society came to the fore in Germany and France, according to which societies, like organisms, have the power of maintaining and reproducing themselves through time. Marx was very much taken with this view, which he endorsed in the *Critique of Hegel’s Philosophy of Right*. Chief among the conditions for a society to reproduce itself, according to Marx, are the ideas held by its members. Thus false consciousness, rather than being simply an accidental feature of human nature (albeit one with enormous political consequences) should be regarded as a phenomenon to be explained by the particular character of the society in which it is to be found.

If societies do not rest solely on coercion, then this is because those who are oppressed or exploited for some reason accept this. As Marx puts it bluntly: ‘the ideas of the ruling class are in every epoch the ruling ideas’. But how does this come about? What sort of connection holds between the economic structures of a society and the ideas of its members? *The German Ideology* contains two analogies that might serve as mechanisms for the explanation of the connection between material life and ideas. The first is embodied in the following famous passage:

> If in all ideology men and their circumstances appear upside-down as in a camera obscura, this phenomenon arises just as much from their historical life-process as the inversion of objects on the retina does from their physical life-process…. We set out from real, active men, and on the basis of their real life-process we demonstrate the development of the ideological reflexes and echoes of this life-process. The phantoms formed in the human brain are also, necessarily, sublimates of their material life-process, which is empirically verifiable and bound to material premises.
>
> (1845-6: 47)

Let us call this the 'reflection model' of ideology. The idea is that ideology relates to material life as images do to reality in a camera obscura or on the retina of the human eye: items in reality are reproduced accurately, but in reverse.

Yet brief consideration of the analogy shows that, as it stands, it is completely inadequate. It is indeed true that the images on the human retina are 'upside-down'. But does this mean that human beings do not perceive the world about them accurately? Of course not. The fact is that, as far as human perception is concerned, 'upside-down' is the right way up for images to be on our retinas. And this points the way towards the problem with Marx's analogy. By describing *all* consciousness as reversed or inverted the contrast between 'true' and 'false' loses its sense.

A further objection arises later in the quoted passage in which Marx continues the reflection analogy when he speaks of the ideological 'reflexes and echoes' of real life-processes. Ideological ideas are, he goes on to say, 'phantoms' and 'sublimates'. These metaphors carry with them an important implication: ideological thought is the effect of real processes, but it is itself insubstantial, without material reality or causal power. If this is Marx's considered view, then it is clearly disastrous for the theory of ideology. For the point of the theory of ideology was to explain how it was that certain forms of thought served to sustain particular societies. Thus these forms of thought are, by assumption, not ineffective, but have very important causal effects: helping to maintain a particular social and economic order.

Finally, it is not obvious that ideology relates to material life as mind relates to matter. Is the implication that ideology is immaterial and material life non-intellectual? This plainly contradicts Marx's basic position. Not only would it be odd for an avowed materialist to suggest that ideas are something basically insubstantial, but, even more importantly, it conflicts with the idea that economic life, so far from being unconscious or unreflective, is the central part of man's cognitive engagement with external reality.

7 The theory of ideology: (2) The interests model

There is, however, another model at work in *The German Ideology*. While the reflection model draws on the parallel between the ideological process and a traditional, realist account of perception (the immaterial mind passively mirrors a mind-independent reality) what we may call the 'interests model' develops from a more instrumentalist approach to epistemology. That Marx was (at this time, at least) attracted to such views is apparent from the 'Theses on Feuerbach'. In the second thesis he writes, 'The dispute over the reality or non-reality of thinking that is isolated from practice is a purely *scholastic* question.' From this point of view, the most significant aspect of ideas is not their relationship to a mind-independent reality, but that they are the products of practical activity, and that this practical activity is itself guided by interests. The materialistic view of history that this leads to, Marx and Engels say: 'does not explain practice from the Idea, but explains the formation of ideas from material practice'.

The problem with the interests model does not lie in the view that ideas are the product of interests itself, which is, of course, very plausible (although it is more difficult to determine just what proportion of our ideas are products of interests in this way - surely not all of them - and to explain just how it is that interests should assert themselves in the process by which ideas are formed). The problem is that ideological ideas are not simply ideas formed in the pursuit of interests. They are, in fact, supposed to be ideas that go *against* the interests of a large number of those who hold them (and in this way further the interests of others). How do ideas of this kind come to be accepted?

Marx and Engels's answer starts from the following claim:

> The class which has the means of material production at its disposal, has control at the same time over the

> means of mental production, so that thereby, generally speaking, the ideas of those who lack the means of mental production are subject to it.
>
> (1845-6: 64)

But this is not a satisfactory solution. Marx and Engels seem to view those who live under the domination of the ruling class as passive victims, taking their ideas like obedient chicks from those who control the 'means of mental production', with no critical reflection as to whether the ideas are either true or in their own rational interests. Yet why should one suppose that the ruling class is capable of promoting its interests effectively and forms its ideas in response to those interests, while the dominated classes simply accept whatever is served up to them?

Marx and Engels do, however, attempt to make their claim more plausible in their discussion of the nature of mental production. It is, they write, the most significant development in the division of labour that mental and manual labour become separated:

> Division of labour only becomes truly such from the moment when a division of material and manual labour appears…. From this moment onwards consciousness *can* really flatter itself that it is something other than consciousness of existing practice, that it *really* represents something without representing something real; from now on consciousness is in a position to emancipate itself from the world and to proceed to the formation of 'pure' theory, theology, philosophy, ethics, etc.
>
> (1845-6: 51-2)

The separation between mental and manual labour, Marx and Engels maintain, does not really lead to the formation of autonomous ideas; the ideologists who produce ideas are still part of the ruling class whose interests their ideas represent. Nevertheless, it offers an explanation as to why such ideas should be accepted by those, the dominated classes, whose interests they oppose: they are accepted because they are apparently disinterested. The ideologist, on this view, is like a bribed referee: able to influence the outcome of a game all the more effectively for the fact that he is falsely believed to be impartial.

Are ideologists, then, engaged in deception? Do they know the partiality of their ideas but present them none the less as if they were neutral and disinterested? On the contrary. According to Marx and Engels, ideologists are sincere - and, because they sincerely believe in the independence and objective validity of their own ideas, they are able to persuade others to accept them as such all the more effectively. Herein, however, lies the problem. How are we to suppose it to be true that the ideologists should both be constrained so that they produce ideas in the interests of the ruling class of which they are, appearances to the contrary, a part, and that they (and those who accept the ideas from them) remain sincerely unaware of the nature of this connection? Why do they *think* that they are independent when in fact they are not? And, if they are not independent, how do the class interests they share with the rest of the ruling class assert themselves?

In any case, it is clear why Marx should now become so hostile to philosophy: like any supposedly 'pure' theory, philosophy represents a deceptive abstraction from the particular circumstances and material interests that it serves. This move to detach ideas that are the products of material interests from the interests that they represent is epitomized, for Marx and Engels, in Kant (the 'whitewashing spokesman' of the German bourgeoisie, as they call him). Kant, they write:

> made the materially motivated determinations of the will of the French bourgeois into *pure* self-determinations of 'free will', of the will in and for itself, of the human will, and so converted it into purely ideological determinations and moral postulates.
>
> (1845-6: 99)

For Marx and Engels, at this stage at least, 'moral postulates' are, by their very nature, ideological.

8 Historical materialism

'Where speculation ends - in real life - there real, positive science begins', according to Marx and Engels in *The German Ideology*. The science to which they are referring is the materialist theory of history, whose classic statement is given in the Preface to *Zur Kritik der politischen Ökonomie (A Contribution to the Critique of Political Economy)* (1859).

Taken most generally, the materialist theory of history asserts that the manner in which human beings produce the necessities of life determines the form of the societies in which they live. Every society other than the most primitive produces a ‘surplus’ beyond what it immediately consumes. The manner in which this surplus is ‘appropriated’ - taken from the direct producers and redistributed - determines the class structure of the society in question. If society is divided between direct producers and those who benefit from the former’s ‘unpaid surplus labour’ (something that is true of all societies where a surplus exists, prior to the advent of socialism) the relationship between classes is antagonistic.

At any stage, the size of the surplus is an expression of the level of development of the ‘productive forces’ - the resources, physical and intellectual, upon which material production draws. Every society contains both an economic ‘base’, composed of ‘relations of production’ (the relations producers have to the means of production and to one another) and a legal and political ‘superstructure’, corresponding to the base. The relations of production favour the development of the productive forces up to a point. Beyond this they become, Marx says, ‘fetters’ upon the forces of production, and a conflict arises which leads eventually to the replacement of the existing relations of production with new and superior ones.

Presented in these terms, it is clear that the materialist theory of history is intended as an exercise in social science rather than philosophy. Thus it may seem surprising that it should have attracted such enduring attention on the part of philosophers. However, scientific theories may be of concern to philosophers if their assumptions are novel, obscure or questionable, even if the intentions behind them are in no way philosophical (examples are Darwin, Freud and Newton). In the case of Marx’s theory of history, it is not just the meaning of and evidence for the particular claims to be found in the theory that have been controversial. The more general issues of the form of explanation that Marx employs and the kind of entities such an explanation presupposes have been continuing matters of dispute.

Interpreters of Marx divide broadly into three groups on these questions. In the first are those for whom Marx’s theory of history is intended to be scientific in the way that any other scientific theory is. With some qualifications, the majority of the earliest Marxists (for example, Engels himself, Kautsky and Plekhanov) fall into this group. On the other hand, those who believe that there is a contrast between Marx’s conception of science and the natural sciences may be divided into those who see Marx’s theory as a transformation of Hegel’s theory of history and those for whom it is fundamentally anti-Hegelian. The most influential presentation of the former interpretation is to be found in Georg Lukács’ *History and Class Consciousness* (1921), while the latter is particularly associated with the French philosopher, Louis Althusser (see Althusser, L. §§2-3; Kautsky, K.; Lukács, G. §2; Plekhanov, G. §2).

In the late 1970s the first approach was revived in the English-speaking world by G.A. Cohen’s seminal *Karl Marx’s Theory of History: A Defence* (1978). According to Cohen, historical materialism can be presented in a way that contains nothing that should be unacceptable to anyone who accepts the legitimacy of Darwinian biology (see Darwin, C.).

The two theories are, in Cohen’s view, importantly parallel to one another, for both employ ‘functional explanation’ (see Functional explanation). When Marx says that the relations of production *correspond* to the forces of production, what he means, according to Cohen, is first that the relations are in some sense ‘good for’ the (development of the) forces and second that they obtain *because* they are good for the forces. (The same analysis, suitably adapted, applies to the correspondence between superstructure and base.) What is distinctive about Darwinian biology, however, is not just that it employs functional explanation, but that it provides a convincing account (what Cohen calls an ‘elaborating explanation’) of why its functional explanations are true: the process of natural selection. Does Marxism have an equivalent elaborating explanation?

All the indications are that it does not. In response to this, there have been two main lines of argument. One is that the theory should have (but lacks) such an explanation and that it is the task of a sympathetic reconstruction of Marx to provide one. On the other hand, it is also possible to argue that the search for what Jon Elster has called ‘micro-foundations’ is misguided (1985). Thus the functional explanations that Marx invokes in the theory of history rest on the fact that there really are collective agents (classes, for example). On this ‘collectivist’ reading it is sufficient simply to appreciate the nature of collective agency to see why collective agents should feature in

functional explanations: they have the power to act purposively to bring about their ends. No reductive 'elaborating explanation' is necessary.

To take this view is to align oneself with the second and third groups of Marx's interpreters and to affirm the fundamental gap between Marx's theory of history and the explanations of the natural sciences (where functional explanations are not simply left unelaborated). If so, the Marxist theory of history cannot draw on the general prestige of science for its justification.

9 Political economy

In contrast to his relatively brief and schematic statements concerning general history, Marx wrote very extensively about the economic system under which he himself lived. *Das Kapital*, which presents Marx's definitive analysis of capitalism, is a work of exceptional methodological complexity, as is already suggested by its sub-title, 'Critique of Political Economy'. The phrase is ambiguous. Is Marx's objective to criticize the bourgeois economy or bourgeois economics? In fact, Marx rejects this as a false antithesis: the subject matter of the book is both. Ten years before its publication, Marx described the work that was to become *Das Kapital* in a letter: 'The…work in question is a *critique of the economic categories*, or, if you like, the system of bourgeois economy critically presented. It is a presentation [*Darstellung*] of the system and, simultaneously, a critique of it'.

The two aspects go together in Marx's view because economic categories are not simply the means employed by an observer to classify some inert mass of data. They are themselves a part of social reality, 'abstract forms' of the social relations of production.

Bourgeois economists, Marx alleges, characteristically fail to recognize that their categories are specific to capitalism, and so they treat the capitalist mode of production as one 'eternally fixed by nature for every state of society', Marx alleges. A 'critical presentation' of economics must counteract the false eternalization of the economy that bourgeois economics carries within itself.

As it stands, this is a criticism of the limitations in the self-understanding of bourgeois economics rather than a challenge to its empirical content. Yet empirical explanation is a central part of Marx's project. 'It is', he writes in the Preface to *Das Kapital*, 'the ultimate aim of this work to lay bare the economic law of motion of modern society.' Has bourgeois economics failed to discover this law or has it simply not put its categories in historical context? At its strongest, Marx's case is that both criticisms are true and that the former failing is a result of the latter. The 'law of value' that Marx claims to have discovered could not, he says, have been discovered by economic science 'so long as it is stuck in its bourgeois skin'.

The connection that Marx sees between the categories of economic life and the categories of economic analysis is made more complicated by the structure that he ascribes to capitalism. Marx believes that an indispensable ingredient for understanding capitalism is the contrast between its 'essence' - its underlying determinants - and its 'appearance' - the way that it immediately strikes those who live in it. Corresponding to this distinction are two kinds of bourgeois economic thought: what Marx calls 'classical economy', on the one hand, and 'vulgar economy' on the other. Classical economy (the tradition whose greatest representatives were Ricardo and Adam Smith) aims towards the essence of capitalism: it 'nearly touches the true relation of things', although it is not able to formulate that relation explicitly. According to Marx, it is the mark of the 'vulgar economy' of his own time, by contrast, that it 'feels particularly at home in the alienated outward appearances of economic relations'. Yet this means that it is fundamentally unscientific, for 'all science would be superfluous if the outward appearance and the essence of things coincided'. A truly scientific political economy must go beyond the immediately received categories of economic life. This is what Marx believes that he himself has achieved (and he considers himself for this reason to be the heir of the tradition of classical political economy).

In a letter to Engels, written at the time of the publication of the first volume of *Das Kapital*, Marx singles out what he calls the 'twofold character of labour' as the most important point in his book. Labour, Marx claims, is both the source of value and, at the same time, under capitalism, a commodity itself. Yet this commodity (labour-power, as Marx calls it) is a commodity of a special kind. Its value is not the same as the value of the commodities produced by the labour that is exercised on behalf of its purchaser, the capitalist. This discrepancy, in Marx's view, explains the 'origin' of surplus-value - the fact that the capitalist appropriates the surplus-labour of the worker under the guise of a fair exchange. In discussing the manner in which, in capitalist society, labour is

sold to capitalists as a commodity, in exchange for wages, Marx writes: ‘Hence we may understand the decisive importance of the transformation of the value and price of labour-power into the form of wages, or into the value and price of labour itself. This phenomenal form, which makes the actual relation invisible, and, indeed, shows the direct opposite of that relation, forms the basis of all the juridical notions of both labourer and capitalist, of all the mystifications of the capitalist mode of production, of all its illusions as to liberty, of all the apologetic shifts of the vulgar economists’.

Thus we see Marx making three claims: (1) that we should see reality as layered, having a surface appearance governed by an underlying structure; (2) that to make such a distinction is characteristic of the scientific approach to reality in general; and (3) that the phenomenal form conceals the real relations (it ‘makes the actual relation invisible and indeed shows the opposite of that relation’).

However, claims (1) and (2) do not entail (3). According to claims (1) and (2) (in themselves extremely plausible) the way that we see the world is not, immediately, adequate for us to explain the way that the world is. But that does not make our immediate perception of the world false. It simply lacks a theory. Yet Marx’s claim (3) is much stronger: reality presents itself in a way that deceives those who immediately perceive it. Marx’s own statements to the contrary, it seems that this third claim is best understood not as a general consequence of the nature of scientific understanding but as a specific feature of capitalism. Capitalism mystifies those who live under it, Marx believes, because it is a ‘deceptive object’. To penetrate its surface scientifically it is necessary to go beyond the limitations of bourgeois political economy.

10 The fetishism of commodities

The most detailed discussion that Marx provides of a case where the surface of capitalism presents itself as ‘false’ is to be found in ‘The Fetishism of Commodities and the Secret Thereof’, in *Das Kapital*. This discussion is a recognizable reworking of the central themes to be found in the treatment of alienated labour in the *Paris Manuscripts*.

In the eighteenth-century sense of the term, fetishists were those non-European peoples whose religion involved the worship of inanimate objects. Fetishism is a fallacy attributing to objects in the world some quality (power and personality) that they, in fact, lack. Marx’s conception of commodity fetishism shares this structure, but differs in an important way. The fetishism of commodities is not a matter of subjective delusion or irrationality on the part of perceivers, but is somehow embedded in the reality that they face.

According to Marx, two separate facts or properties are distorted in the commodity-form. First, the ‘social character’ of human beings’ labour appears (falsely) as ‘objective characteristics of the products themselves’, and second (in consequence of the first fact, as Marx asserts) the producers’ own relationship to their ‘collective labour’ appears ‘as a social relationship between objects, existing externally to the producers’.

The first issue concerns what the ‘social character’ that is apparently a property of the products themselves amounts to. Is it the sheer fact that the commodity *is* a commodity? This suggestion must be rejected, for the belief that the product is a commodity is in no way a false or deceptive one. Likewise, it cannot be something concealed from the producers that commodities *do* as a matter of fact exchange for one another in certain proportions: it is hard to see how anyone could live their lives within a market society without having an adequate understanding of facts of this kind (enough, at least, to be able to buy something to eat). The best interpretation of Marx’s argument is that it is not such first-order facts about commodities but a second-order one that is the source of deception: it is not *that* commodities can be exchanged with one another in certain ratios but *the reason why* they exchange in the ratios that they do that is their hidden secret.

Marx’s account of the illusion regarding the social character of the products of labour is complemented by the account he gives of the second element in commodity fetishism. Because commodity production takes place as a process by which the producers’ activities are coordinated solely through the imperatives of a system of market exchanges, it follows, Marx says, that ‘the social relations between their private acts of labour manifest themselves as what they are - that is, not as the immediate social relationships of persons in their labour but as material relationships between persons and social relationships between things’.

Implicitly, the market commensurates the labour of each individual with the labour of every other producer -

individual labour has its value in relation to the way in which others perform the same labour. The socially useful character of the labour of the individual producers thus appears to them, according to Marx, 'only under those forms which are impressed upon that labour in everyday practice, in the exchange of products'.

Here again, Marx is indicating an illusion of the second rather than the first order. The individual producers are aware of the role of the market in determining the way in which they labour. In this they are quite correct. But they also believe (falsely) that it is the market that makes their labour useful (rather than recognizing it as a contingent fact about capitalist production that their socially useful labour takes on a market-determined form). Society generates such false beliefs spontaneously, Marx claims. The world of commodities 'veils rather than reveals', he says, the social character of private labour and of the relations between the individual producers.

That the true source of the value of commodities lies in the labour expended in their production is, Marx maintains, a matter of simple scientific truth. So, too, is the fact that the social character of private labour consists in the equalization of that labour under the auspices of the market. Nevertheless, fetishism is a matter of 'objective illusion' and knowledge of these truths does not dispel such false appearance. The discovery of the law of value 'by no means dissipates the objective illusion through which the social character of labour appears to be an objective character of the products themselves' any more than 'the discovery by science of the component gases of air' altered the atmosphere that people breathed.

The analogy that Marx chooses here is not a happy one. Admittedly, it is absurd to think that a scientist's discovery about an object should change the object itself. But that is not the issue. It is not a question of whether the atmosphere itself changes after the discovery of its component gases, but whether the way in which we think about it changes. It is only if we suppose that capitalism, unlike the atmosphere, is an object of a particular kind - a deceptive object - that it is possible to claim that it will continue to encourage such false beliefs in the face of contrary knowledge.

But it is not just that the individuals who live in a society based on commodity production are deceived by it regarding the way that it works. The way that it works is itself criticized by Marx. Above all, the 'social character of labour' is made private in actuality. This is not a misperception or false belief, but a contradiction: a discrepancy between what Marx takes to be the intrinsic nature of social labour and the way that it is in fact organized. Capitalism is not just deceptive, but also defective.

11 Morality

The question whether Marx's theory has a moral or ethical dimension is one of the most controversial of all issues surrounding the interpretation of his work, and the difficulty facing interpreters is easily seen. On the one hand, Marx has a number of uncompromisingly negative things to say about morality. Moreover, after 1845 at least, he affirms that his own theory is not a utopian or ethical one but 'real, positive science'. Yet, on the other hand, much of the language that he uses to describe capitalism is plainly condemnatory (for instance, that it is antagonistic, oppressive and exploitative). Does this not represent an inconsistency on Marx's part? Is he not moralizing and rejecting morality at the same time?

This section will present a line of interpretation according to which Marx is not inconsistent. The interpretation depends on a contrast between certain doctrines typical of moral philosophy (which, it will be argued, Marx rejects) and the rejection of ethical values as such (to which, it will be argued, he is not thereby committed). However, it should be noted that this interpretation is controversial and involves considerable reconstruction of the rather sparse evidence that we have of Marx's views.

It is helpful to start, as Marx himself did, with Hegel's critique of Kant. Both Marx and Hegel share the belief that morality, as embodied in Kant's moral philosophy, is, as they put it, 'abstract' (see Hegel, G.W.F. §8). There appear to be three interconnected elements compressed into this criticism:

(1) First, morality is alleged to be abstract in the sense that it contains principles expressed in universal form (in Kant's case, the 'categorical imperative' to 'act only according to that maxim which you can, at the same time, will to be a universal law' (see Kantian ethics)). While such principles may function as a test upon proposed actions, they do not, the argument goes, determine the content of the action to be performed. Thus, the claims of moral philosophy to the contrary, specific content is surreptitiously imported into ethics from the existing

institutions or codes of behaviour of the society in question.

(2) Second, morality is abstract to the extent that it takes the form of a mere injunction: an imperative that is addressed to people's 'moral reason', telling them to act in a certain way because that is 'good in itself'. Moral action is detached thereby from other forms of human action and, as a result, moral theory has nothing to say about the conditions under which the forms of behaviour that it commends will be realized in practice.

(3) Finally, morality may be said to be abstract in that it contains an unhistorical understanding of its own status. It presents its principles as if they were the axioms of some timeless moral geometry. Yet, in fact, every system of morality is a way of seeing the world that arises in particular circumstances and responds to definite needs within those circumstances.

Although one or more of these features may be present in the forms of moral philosophy with which we are most familiar, it is not clear that they are a necessary feature of every view that one might call 'moral'. Not all ethical positions have to express themselves as systems of universal principles that we are enjoined to follow because they are good for their own sake. Admittedly, many philosophers would argue that to combine the value commitments characteristic of morality with the meta-level doctrine that such values are, in the end, expressions of interest (Marx's version of (3) above) inevitably undermines, as Nietzsche might have put it, the value of value itself. But it is at least arguable that the two standpoints are compatible. The path from sociological determinism to moral scepticism is not as steep, slippery and remorseless as it is sometimes claimed to be.

If this is conceded, we can draw a distinction between morality in two senses: morality as a quasi-Kantian system of principles (which Marx rejects) and morality as a set of values embodying a conception of what is good for human beings (which he can consistently accept). To present things in this way, however, may seem to give insufficient weight to the vehement hostility which Marx shows towards ideas of justice and rights, in particular. On the interpretation being proposed here, Marx's animus is best understood as aimed at what he sees as the assumptions behind such values, rather than at the fact of their being values as such.

Roughly speaking, we may think of rights as things that permit individuals to act in certain ways, in given circumstances, should they wish to do so, and to be able to claim correlative duties on the part of others. A duty, correspondingly, would require individuals to act in some way, whether they wished to or not. Justice (if we do not think of it simply as a matter of rights and duties) would consist of principles on which benefits and burdens are distributed in cases where interests conflict.

What these values have in common is that they provide a framework which regulates and limits the self-seeking behaviour of individuals. They are values that assume a conflict between (to put it in Kantian terms) 'duty' and 'inclination'. Just as Marx supposes that the categories of bourgeois economics eternalize the forms of bourgeois economic life, so, he believes, discussion of rights (which he denounces in the *Critique of the Gotha Programme* as 'ideological nonsense') eternalizes a situation in which the good of each individual is independent and so can only be advanced at the expense of others. Right, moreover, can only apply a fixed and equal standard to unequal individuals, 'from outside'.

For the liberal, who is concerned to protect the individual's powers of self-direction against the intrusions of others, the attraction of the idea of rights is that it presupposes nothing about individuals' characters and personalities. For Marx, on the other hand, that is just its weakness: rights do nothing to transform human nature. Against this, it is clear that Marx, from the time of the *Paris Manuscripts*, sees social progress as characterized by a form of community in which (as he and Engels put it in the *Communist Manifesto*) 'the free development of each is the condition for the free development of all'. Marx's ethical ideal is one of solidarity in which all advance together.

Hence Marx's reluctance to use the language of justice to condemn capitalism becomes more intelligible. It is not that Marx thinks that exploitation, expropriation, oppression, slavery and misery (a few of the terms he applies to the capitalist system) are *just*. But he is reluctant to use language that would suggest that these are forms of injustice for which 'justice' (in the sense of giving 'each their due') is the final and sufficient remedy.

12 Socialism

It may seem odd, given that Marx devoted his life to the achievement of a socialist society, how brief and

unspecific his accounts of it are. One explanation that is often advanced for this apparent neglect is the following. Marx believed, it is said, that thought is limited to its own time. Thus it would have been improper for him, living under capitalism, to try to anticipate the nature of the society that would replace it and to write (as he puts it in the Preface to the Second Edition of *Das Kapital*) 'recipes for the cook-shops of the future'. While this may be part of the reason for Marx's reticence, it cannot alone suffice. For, even if we grant that Marx believed that each stage of society sets a boundary which thought cannot cross (and it is by no means beyond question that he did hold this view in such a strong form) he is also committed to the view that socialism is anticipated within capitalism.

In the Preface to *Zur Kritik der politischen Ökonomie* Marx makes the general claim that new forms of society are always prefigured within the old ones that they replace. 'Mankind', he writes, 'only sets itself such tasks as it is able to solve, since closer examination will always show that the problem itself arises only when the material conditions for its solution are already present or at least in the course of formation'.

Marx describes the process by which capitalism prepares the ground for socialism at the end of the first volume of *Das Kapital*. As the productive forces developed by capitalism grow, he claims, so too does the 'mass of misery, oppression, slavery, degradation, exploitation'. A stage is reached, however, at which the monopoly of capital becomes a 'fetter' on production and 'the centralization of the means of production and the socialization of labour at last reach a point where they become incompatible with their capitalist shell'. At this point, the shell 'bursts asunder', the 'death knell' sounds for capitalism and the 'expropriators are themselves expropriated'.

The first and most obvious difference between capitalism and socialism is that common ownership leads to a quite different pattern of distribution of the products of labour. No longer will the capitalist, in virtue of his ownership of the means of production, be able to exploit the individual producer. In the *Critique of the Gotha Programme* Marx distinguishes two stages of post-capitalist society. In the first, the direct producer receives back from society (after deductions for shared costs and social expenditure) 'what he has given to it as his individual quantum of labour'.

But this, Marx points out, is a principle of distribution that merely rectifies exploitation. It does not remedy the inequalities that arise from contingent differences in natural capacities between individual producers. Later, however, society will move beyond this, Marx claims, and 'the narrow horizon of bourgeois right' will be 'crossed in its entirety'. At this point, the principle upon which society will operate will be: 'From each according to his ability, to each according to his needs!' But socialism is distinguished by more than its principle of distribution. In particular, labour will be organized quite differently from the way that it is organized under capitalism.

One of Marx's few reasonably extensive accounts of the nature of the socialist organization of production is to be found in the section on the fetishism of commodities in *Das Kapital*, as part of a comparison between capitalist and other forms of production. Marx starts with Robinson Crusoe, whose productive activity he describes as 'simple and clear'. For Robinson, Marx says, the organization of production is a purely administrative operation: the end is known, as are the resources available and the techniques by which that end could be attained. Marx then moves from 'Robinson's island, bathed in light', via feudal and patriarchal forms of production, before alighting on: 'a community of free individuals, carrying on their labour with the means of production in common, in which the labour-power of all the different individuals is consciously applied as the combined labour-power of the community'.

Here, Marx says,

> All the characteristics of Robinson's labour are…repeated, but with this difference, that they are social, instead of individual…The social relations of the individual producers to their labour and to the products of their labour remain here transparently simple, in production as well as in distribution.
>
> (1867-: 171-2)

The idea that labour could be 'consciously applied' in a complex modern society - resources and needs coordinated, efficient techniques adopted, innovation managed - with the same 'transparent simplicity' as an individual allocating his time to different tasks on a desert island is astonishingly implausible. And, even if it were not so, the question would still arise how that 'common and rational plan' (as Marx terms it elsewhere) would relate to the individuals whose task it was to carry it out. Would it not, from their point of view, be no less of an 'external' imperative to be followed than the dictates of the market that govern their labour under capitalism?

Arguably, the idea that society under socialism would be spontaneously unified like one great, self-transparent super-individual represents an unacknowledged hangover in Marx's mature thought from Hegel's doctrine of *Geist*. However that may be, the presence of this doctrine goes a long way towards explaining why Marx had so little to say about the problems of socialist economic organization: he simply failed to see the difficulty. Few theoretical omissions, surely, have ever had more disastrous historical consequences.

See also: Chinese Marxism; Civil society; Communism; Economics, philosophy of §3; Explanation in history and social science §§1-2; History, philosophy of; Marxism, Western; Marxist philosophy of science; Marxist philosophy, Russian and Soviet; Marxist thought in Latin America; Needs and interests; Political philosophy, history of; Revolution; Socialism

MICHAEL ROSEN

List of works

Marx, K. and Engels, F. (1975-) *Gesamtausgabe* (*MEGA*), Berlin: Dietz.(This outstanding edition of the collected works is often known as *MEGA* II. However, at the time of writing, *MEGA* II, like *MEGA* I, seems likely to remain uncompleted.)

Marx, K. and Engels, F. (1961-83) *Werke* (*MEW*), Berlin: Dietz.(A very adequate edition in German that contains all the works referred to in the text of the entry.)

Marx, K. and Engels, F. (1975-) *Collected Works*, London: Lawrence & Wishart.(A complete edition in English. Marred, however, by patchy and dogmatic editorial work and poor translations. Other editions are generally to be preferred, if available.)

Marx, K. (1975-) *The Pelican Marx Library*, Harmondsworth: Penguin.(Not a complete edition, but a series that contains particularly good translations of *Das Kapital*, the *Grundrisse*, and the *Early Writings*, among others.)

Marx, K. (1841) *Differenz der demokritischen und epikureischen Naturphilosophie*, trans. *Difference between the Democritean and Epicurean Philosophy of Nature*, in *Collected Works*, London: Lawrence & Wishart, 1975-, vol. 1, 25-107.(Marx's doctoral dissertation.)

Marx, K. (1843a) *Zur Kritik der Hegelschen Rechtsphilosophie*, trans. R. Livingstone and G. Benton, *Critique of Hegel's Philosophy of Right*, in L. Colletti (ed.) *Early Writings*, Harmondsworth: Penguin, 1975, 57-198.(An extensive discussion of Hegel, abandoned by Marx.)

Marx, K. (1843b) 'Zur Judenfrage', trans. R. Livingston and G Benton, 'On the Jewish Question', in L. Coletti (ed.)*Early Writings*, Harmondsworth: Penguin, 1975, 211-42. (An important treatment of the idea of emancipation.)

Marx, K. (1844) *The Paris Manuscripts*, trans. R. Livingston and G. Benton in L. Coletti (ed.) *Early Writings*, Harmondsworth: Penguin, 1975, 279-400.(Contains Marx's ideas on alienation.)

Marx, K. and Engels, F. (1845a) *Die Heilige Familie*, trans. *The Holy Family*, in *Collected Works*, London: Lawrence & Wishart, 1975-, vol. 4, 15-211.(A polemic against Bruno Bauer and his associates.)

Marx, K. (1845b) 'Theses on Feuerbach', trans. R. Livingston and G. Benton in L. Coletti (ed.) *Early Writings*, Harmondsworth: Penguin, 1975, 421-3.(Brilliant aphorisms on philosophy.)

Marx, K. and Engels, F. (1845-6) *Die Deutsche Ideologie*, ed. C.J. Arthur, trans. *The German Ideology*, London: Lawrence & Wishart, 1970.(A well-produced version of this key work.)

Marx, K. (1847) *La Misère de la Philosophie*, trans. *The Poverty of Philosophy*, in *Collected Works*, London: Lawrence & Wishart, 1975-, vol. 6, 105-212.(Polemic against Proudhon.)

Marx, K. and Engels, F. (1848) *Das Kommunistische Manifest*, trans. *The Communist Manifesto*, in D. Fernbach (ed.) *The Revolutions of 1848*, Harmondsworth: Penguin, 1973, 62-98.(The classic statement of Marx's political and social programme.)

Marx, K. (1850) *Die Klassenkämpfe in Frankreich*, trans. *Class Struggles in France*, in D. Fernbach (ed.) *Surveys From Exile*, Harmondsworth: Penguin, 1973, 35-142. (Commentary on contemporary politics.)

Marx, K. (1852) *Das achtzehnte Brumaire des Louis Bonaparte*, trans. *The Eighteenth Brumaire of Louis Bonaparte*, in D. Fernbach (ed.) *Surveys From Exile*, Harmondsworth: Penguin, 1973, 143-249.(Commentary on contemporary politics.)

Marx, K. (1857-8) *Grundrisse der Kritik der politischen Ökonomie*, trans. M. Nicholas, *Grundrisse*, Harmondsworth: Penguin, 1973.(An abandoned draft on political economy - superseded by *Das Kapital*.)

Marx, K. (1859) *Zur Kritik der politischen Ökonomie*, trans. as *A Contribution to the Critique of Political Economy*, in *The Pelican Marx Library*, Harmondsworth: Penguin, 1975-.(Contains the classic statement of 'historical materialism'.)

Marx, K. (1862-3) *Theorien über den Mehrwert*, trans. E. Burns, *Theories of Surplus-Value*, London: Lawrence

& Wishart, 1963, 3 vols.(Marx's studies on economic theory.)
Marx, K. (1867-) *Das Kapital*, trans. B. Foukes, *Capital*, Harmondsworth: Penguin, 1977.(Marx's mature masterpiece.)
Marx, K. (1871) *Der Französische Bürgerkrieg*, trans. *The Civil War in France*, in D. Fernbach (ed.) *The First International and After*, 1973, 187-268.(An important political text.)
Marx, K. (1875) *Kritik des Gothaer Programms*, trans. *The Critique of the Gotha Programme*, in D. Fernbach (ed.) *The First International and After*, 1973, 339-59. (Contains a rare, brief discussion of the nature of socialism.)
Marx, K. (1977) *Selected Writings*, ed. D. McLellan, Oxford: Oxford University Press.(At the time of writing, probably the best of the many good selections from Marx's work available.)

References and further reading

Althusser, L. (1965) *For Marx*, Harmondsworth: Penguin.(The work that initiated the 'structuralist' or 'anti-humanist' interpretation of Marx.)
Callinicos, A. (1985) *Marxism and Philosophy*, Oxford: Oxford University Press.(A sympathetic defence.)
Cohen, G.A. (1978) *Karl Marx's Theory of History: A Defence*, Oxford: Oxford University Press.(A masterpiece of sustained interpretative argument.)
Elster, J. (1985) *Making Sense of Marx*, Cambridge: Cambridge University Press. (Less tightly focused than Cohen, but full of insight and perhaps less one-sided. Contains a particularly good discussion of Marx's economics.)
Kolakowski, L. (1975) *Main Currents of Marxism*, Oxford: Oxford University Press, vol. 1.(A critical treatment, emphasizing the prophetic-metaphysical background to Marxism.)
Lukács, G. (1921) *History and Class Consciousness*, London: Merlin, 1971.(First published in 1921, this is the book that initiated 'Hegelian' or 'humanist' Marxism.)
Popper, K. (1948) *The Open Society and its Enemies*, London: Routledge & Kegan Paul.(An influential critique of Marx's claims to 'science'.)
Roemer, J. (1982) *A General Theory of Exploitation and Class*, Cambridge, MA: Harvard University Press.(An ambitious reconstruction of Marxist economics.)
Rosen, M. (1996) *On Voluntary Servitude: False Consciousness and the Theory of Ideology*, Cambridge: Polity Press.(A critical account of the theory of ideology.)
Torrance, J. (1995) *Karl Marx's Theory of Ideas*, Cambridge: Cambridge University Press.(An extremely detailed account.)
Wood, A. (1981) *Karl Marx*, London: Routledge & Kegan Paul.(Places emphasis on the philosophical aspects of Marx's work. Contains an extended interpretation of Marx's view of morality quite different from the one advanced in this entry.)

Marxism, Chinese

Chinese Marxism is a mixture of elements from Confucianism, German Marxism, Soviet Leninism and China's own guerrilla experience. Because Mao Zedong (1893-1976) was in power longer than any other Chinese communist, the phrase 'Chinese Marxism' is commonly used to refer to Mao's own evolving mixture of ideas from these sources. However, the advocates of Chinese Marxism have come from many different factional backgrounds and have tended to emphasize different aspects in their own thinking. Even Maoism reflects many minds. For example, Mao's two most famous essays, 'Shijianlun' ('On Practice') *and* 'Maodunlun' ('On Contradiction') *(1937) drew heavily from Ai Siqi, the author of the popular philosophical work Dazhong zhexue (Philosophy for the Masses) (1934).*

The goals of the Chinese Marxists included the salvation of China from its foreign enemies and the strengthening of the country through modernization. Accordingly, they selected from other systematic theories those doctrines that appeared to facilitate those goals, and then paired these doctrines with others from theories that were sometimes incompatible. One should not, therefore, look for logical consistency in the relations between the ideas that the Chinese Marxists drew from these various sources.

The foundation of Chinese Marxism was undoubtedly Marx's materialist conception of history, and the concepts of class struggle and control of the forces of production shaped the thinking of many early Marxists. However, faced with the need to accelerate social change through class struggle rather than waiting for the full flowering of capitalism, Marxists such as Li Dazhao began focusing less on materialism or determinism and more on voluntarism. There also arose a doctrine, based on the ideas of Lenin and Trotsky, that right-minded people could 'telescope' the phases of the revolution and hasten the transition through the historical stages. This ultimately led to the doctrine of permanent revolution. First promulgated in China in the late 1920s, it reappeared in the 1950s. After Mao's death, the 'subjectivity' movement within Chinese Marxism sought to move the focus away from classes or groups and onto the individual subject as an active agent.

Throughout the evolution of Chinese Marxism, political struggles played a direct role in the formulation and discussion of philosophical positions. Mao's epistemological essay 'Shijianlun' *clearly reflects the experience of leaders during the guerrilla period, and his theories of knowledge are analogous to the 'democracy' practised by the guerrilla leaders: the people were consulted for their knowledge and opinions, decisions were then made from the centre, and the resulting policies were taken back to the masses through teaching. In the same way, Mao believed, individuals perceive through their senses, form theories in their brains (the centre), and test the resulting theories in a manner analogous to teaching.*

In China, right minds among the people were thought to arise through officials teaching the people. Here pre-modern Confucian legacy becomes important. It helps to explain the endurance of teaching as an official function in the Chinese Marxist discussion of democratic centralism. In Confucianism, the primary function of government was education, although it certainly had other tasks, such as the collection of taxes. All officials, including the emperor, had the task of transforming the character of the people. The education in which the state involved itself, through control of the curriculum and national examinations for the civil service, was moral education. The ultimate aim of state-controlled Confucian education was a one-minded, hierarchical society, meaning that people of all different strata would think the same on important matters. Maoists also sought to create a one-minded people through officially controlled teaching.

If the focus of teaching is on right ideas, which are supposed to motivate people towards socialism, one such idea in later Maoist writing is egalitarianism of social status. This was challenged by others, notably Liu Shaoqi, and following Deng Xiaoping's assumption of power in 1978 it suffered a further blow with the switch in economic policy from central planning to market forces.

An example of the relevance of political struggle to the formulation of ideas was the heightening of the campaign against the philosophy called 'humanism', following a dispute in 1957 between Mao and President Liu Shaoqi. Liu made a speech in April of that year saying that capitalists had changed and so class struggle against them could be minimized; this was followed by a Maoist-inspired attack on humanism as a philosophy. The humanism that the Maoists attacked was a Confucian-inspired belief in a class-transcending humaneness or compassion for

humankind or humaneness. In contrast, in the post-Mao years, the content of humanism has altered, and the term has come to refer to a doctrine inspired by both the early Marx and by the Western psychologist Maslow, namely that the goal of society is the individual's self-realization. This form of humanism is one of several competing positions that claim to carry on the Marxist tradition in new directions, and has been reinforced by one form of the subjectivity movement in the Deng Xiaoping era.

1 Historical materialism

It was the German theories of Marx that furnished the vocabulary from which Chinese Marxists gradually developed their positions. The key expressions are the building blocks of Marx's 'materialist conception of history', found in the preface to his *A Contribution to the Critique of Political Economy* (1858). Here, Marx argues that human societies evolve progressively through historical stages and that there is an agent of change that propels them from one stage to another. This agent is the struggle between social classes ('relations of production') that have differing power ties to property, especially to property used in production. This class struggle begins when the existing class structure prevents further development of the technology used in making tools ('forces of production'). The combination of classes and tool technology constitutes the 'foundation' or base; everything significant that humans believe (their ideologies or 'consciousness') in the realms of legal, political, religious and aesthetic matters, plus the institutions that embody them, are the 'superstructures'. In this formulation, the materialist conception of history suggests that the base determines the superstructure. What we believe is determined by economic factors, the classes and tools.

The class that has the most advanced beliefs leads a society from one stage to the next. However, there is an inevitable sequence in this historical progression. There must first be economic development (in Marx's time, this meant the flowering of capitalism), then the acquisition of consciousness by the emerging social class that capitalism fetters (the proletariat), and only then can that progressive proletarian class carry out its struggle to move society forward.

The founders of the Chinese Communist Party included Li Dazhao (1888-1927) and Chen Duxiu (1879-1942). Li, once chief librarian at Beijing University, was later the first communist to join the Guomindang. Chen had been an editor of the influential journal *New Youth* and initially advocated guild socialism under the influence of John Dewey. Li was at first only critically receptive to Marxism; although he accepted the need for class struggle, he viewed it as a sign of human evil. Marxism itself, he felt, lacked an ethical perspective. Good persons emerge as they learn mutual aid, a principle not stressed by Marx.

The early Chinese Marxists did ultimately retain the idea of a glorious future for China, a future guaranteed by the progressive view of history and embodied in historical materialism and the dynamics of class struggle. However, the sequence of the historical theory, the fact that a full flowering of capitalism was required before the agents of change could begin to guide the struggle to create a new society, caused problems. This process could take a long time in industrially backward China. During the years 1919 and 1920, many of the earliest Chinese Marxists, such as Li Dazhao, focused on economic determinism and were troubled by it. In an attempt to find a solution they turned to other theories, selecting those doctrines that appeared to facilitate their goals.

One such example of selective borrowing took from German Marxism. Li Dazhao and certain other Chinese Marxists played down base-superstructure determinism and the idea of the inevitable sequence of events (right ideas follow upon full economic development, and leadership of revolutionary struggle follows consciousness of those ideas) with which it was associated. They eliminated those parts that comprised the materialist conception of history, as well as the notion that historical stages are of long duration and that classes are to be defined solely in economic terms. Instead, they took the position that human consciousness or the minds of right-thinking, strong-willed persons can determine the future. This is voluntarism, the power of the conscious will (*voluntas*) to change the material world. In taking this position, Li and the others were influenced by non-Marxist intellectual currents in China, including even anarchism.

Although economic determinism in the Chinese Marxist theory of history faded, this did not mean the total disappearance of the economic variables in that determinism from the theories of the Marxist community. For example, enduring variables encompassed some economic determinants used in class analysis. These included differentiating urban factory proletarians from rural peasantry, and defining poor peasants in terms of amount of

farm tools and animals owned. Another example is the explanation of historical stages in terms of the transition from one form of property ownership (bourgeois/capitalist) to another (socialist). Ideological disputes often centred on the nature of some of these economic variables, such as on the duration of stages. For example, Chen Duxiu initially opposed an alliance of the communists and the Guomindang in the 1920s, which was promoted by the USSR. He eventually accepted the alliance, while at the same time trying to preserve the independence of the communists, but he proclaimed that China would need a long capitalist/bourgeois stage before socialism could be achieved. During this period, Chen wanted the communists to focus on education in a multi-party state that protected civil rights. He regarded armed class uprisings at this stage as foolish.

Other factional disputes within the leadership centred on which economically-defined class was to play the dominant role in the revolution. As early as 1926, while the Communist International stressed organizing the urban proletariat, Mao saw the potential of peasant associations, believing that peasants could be successful agents of the communist revolution. In contrast, Li Lisan wanted the urban proletariat to control the rural communist councils. In 1931 came the end of the communist movement in the cities and the shift to rural areas. Mao's doctrine prevailed: the 'peasants' have proletarian ideas, and thus class definition came to be defined partly in economic terms and partly in terms of consciousness.

The voluntarist position that replaced economic determinism gradually became intertwined with another doctrine, that concerning the ability of right-minded people to telescope the process of transition through historical stages. This new doctrine drew elements from a thesis popularly associated with Trotsky, the theory of permanent revolution, from which Mao had nominally distanced himself. Trotsky in turn had borrowed most of the essentials of this theory from Marx and Lenin. This voluntaristic, unacknowledged theory of permanent revolution came to coexist with what was left of historical materialism in China after the unacceptable parts were eliminated. It is completely incompatible with the original German formulation of historical materialism and economic determinism that had so fascinated the first Chinese Marxists.

2 Permanent revolution

Impatient to move China into its safe and prosperous future, first Li Dazhao and then Mao Zedong adopted a stance that had been foreshadowed by Marx (in his 'Address of the Central Committee to the Communist League' in 1850), and then by Lenin (in 'Two Tactics of Social Democracy in the Democratic Revolution'). The voluntarist theories of Li and Mao changed the agent that precipitates class struggle in historical materialism. No longer was this agent the tension between tool technology and class structure; rather, it was the ideas in the minds either of a few elites who teach the masses (in the view of Mao), or of those elites aided by some positive ideas already in the minds of almost all Chinese people (in the view of Li). Armed with these ideas, elites can telescope the course of history, shortening the normal duration of a historical stage. They can do so by taking control of the historical period that really should be led by the capitalists. Although Mao and Liu Shaoqi did not start regularly using the term 'uninterrupted' (permanent) to describe revolution and society until the 1950s, Mao had accepted the principles of this doctrine of permanent revolution in the 1930s.

In this view, then, China does not need to wait for capitalism to flower before the right ideas enter the minds of its people; the elite can acquire the right ideas rapidly and proceed to lead the class struggle, take over leadership of the revolution and introduce socialism. This revolution is permanent in the sense that the attainment of one stage (bourgeois democracy) is not followed by a long period of peace and gradual growth into the next. Rather (and this was Lenin's contribution), the struggle that ushered in the new stage continues, as the few leaders with the right ideas take power away from those that would seem to be in charge (capitalists) and rapidly accelerate people into the new historical (socialist) phase.

Li Dazhao foreshadowed the doctrines of Mao Zedong in important ways. First, he highlighted revolutionary consciousness as being able to arise independently of the existence of either a large capitalist or proletarian class. The correct ideas in the minds of even a small progressive class (the communists) could accelerate China's economic and social development. Second, he defined class partially in terms of the ideas in minds. Thus, as Maurice Meisner (1967) showed, it is permissible to speak of proletarian ('progressive') ideas in the minds of most Chinese people (who are peasants) as a result of their experience with foreign 'imperialists'. Third, in Li's view, China's revolution would be peasant-based rather than based on urban workers. China could change rapidly through class struggle. The agent that would initiate class struggle and modernize China was ideas in the minds of

peasants or of most Chinese, a reversal of the sequence in historical materialism.

Mao differed from Li Dazhao in being China’s quintessential Leninist. In his view, the right ideas are originally the exclusive preserve of a small elite within the Communist Party, rather than being more widely shared among the people. Mao also differed from Trotsky and other advocates of permanent revolution in identifying a democratic or bourgeois stage for China. This is an insignificant difference, however, because for Mao, such a stage (described in his 1940 essay ‘Xin minzhu zhuyilun’ (‘On New Democracy’)) would last only a few years and would be controlled by the communists as much as possible. In fact, according to Mao, socialism was achieved and new democracy ended in 1955.

Not all of the early Chinese Marxists advocated the power of the mind to change objective conditions as strongly as did Li and Mao. This particular aspect of permanent revolution is less pronounced in, for example, the writings of Ai Siqi, the author of the popular philosophical work *Dazhong zhexue (Philosophy for the Masses)* and a resident ideologist in the Yenan guerrilla base.

3 Democracy, centralism and epistemology

The bridge between permanent revolution on the one hand, and democracy and epistemology as topics for study on the other hand, lies in the need for those who already have access to correct ideas to teach and thereby mobilize those who do not. Together, they initiate the struggle that modernizes China. Teaching is required by permanent revolution, a dictum reinforced by Confucian views on the role of rulers. The content of what is taught, or the epistemological element, derives in part from the Chinese Marxist understanding of democracy.

Ideas about democracy and epistemology in Chinese Marxism owe as much to the guerrilla experience of the 1930s and 1940s as they do to Lenin’s earlier insistence on the subordination of party debate to central leadership positions once decisions are made. In north China, the communists were dependent on local peasants for logistical support (avoiding long supply lines that could be cut) and for intelligence about the movements of Japanese and Guomindang troops. Being dependent, they had to maintain good relations with the local people. This is the origin of the communist custom of living with and consulting with the people about non-military and military matters.

Mao believed that the origin of knowledge lies in practice. Although he referred to production and science as forms of practice, uppermost in importance was the practical experience of conducting class struggle. Thus Mao regarded the knowledge of the Chinese people as developing in part during the guerrilla struggles with the Japanese and the Guomindang (in spite of the United Front with the latter, during which the communists and the Guomindang fought together against the Japanese). Democracy and centralism are forms of relationship between party officials and masses that generate knowledge in the course of the class struggle.

Mao understood the term ‘democracy’ to mean letting ordinary people speak out on topics about which they have some direct knowledge. They have useful concrete information that can help officials make good policies. Letting them speak out also helps to unify the party and the people; it promotes cohesiveness. Thus democracy is a useful tool for leaders to employ.

The expression ‘mass line’ arose during the guerrilla period to refer to the custom of consulting with local people or carrying out local investigations prior to expanding operations to a larger area, where that locally-derived information would be applicable. ‘From the masses’ is part of a slogan, and it refers to this grass roots consultation. Thus, for Chinese Marxists this aspect of the mass line practice embodied ‘democracy’. Typical of the Maoist form of Chinese Marxism is the position that ‘masses’ refers primarily to the peasantry, a source of important raw data.

There is another aspect to the mass line, indicated by the rest of the slogan (‘to the masses’), which refers to leadership, guidance or, especially, teaching by officials about policy for the local people. This takes place after prior consultation at headquarters, where the authority of centralism prevails.

In Maoist Marxism, the ideas of democracy and centralism are also intertwined with a theory of knowledge. In the stages of acquiring knowledge, an individual begins with perceptual or empirical knowledge gained through practical experience. Within the brain the individual then generalizes and formulates hypotheses or theories about such concrete data. The resulting rational knowledge is higher or more worthy than perceptual knowledge. Following this, individuals test their theories, observe and, if necessary, revise the theory. This process is

discussed in Mao’s essay ‘Shijianlun’. Chinese Marxists extended the analogy of the individual’s acquisition of knowledge to the mass line or to the leadership style required by democratic centralism. ‘From the masses’ consultation is analogous to sense perception by an individual; policy formulation at headquarters is analogous to theorizing in the brain; and testing/revising of theories is analogous to ‘to the masses’ teaching. Mao discusses this in several places, notably in ‘Guanyu lingdao fangfa di rogan wenti’ (Some Questions Concerning Methods of Leadership), written in 1943.

All of the above is compatible with the Leninist idea of elite officials knowing more than ordinary people. After all, they possess rational knowledge, which is higher than the perceptual knowledge of the masses. That possession justifies their privileges. The fusion of the mass line with a theory of knowledge also meets the demand of the doctrine of permanent revolution for the telescoping of progress through a historical stage. Leaders focus on transforming the minds or consciousness of the people as a way of rapidly modernizing the country socially and economically. Leaders are teachers who help transform the perceptual knowledge of the people into higher, rational knowledge. In consulting and then teaching, they satisfy the requirements of democracy and of centralism. Permanent revolution demands not only changing consciousness, it demands also the class struggle that follows in its wake. Other Marxists in China opposed the focus on struggle, advocating instead what they and their ideological opponents called ‘humanism’ (see §4).

After the communist victory in 1949 the consultative side of the mass line atrophied, but the leaders, members of the Communist Party, retained their instructional role. As during the imperial age, transforming the minds of the people remained a core function of officials. Among the many reasons why the consultative role evaporated was the increased difficulty of the task. Consulting in a small area such as the guerrilla bases was not difficult; consulting in a large country such as China was, especially when that country had no history of representative government. The difficult was ignored. Also, there were no built-in safeguards to the mass line consultation process; in gathering information, officials could easily filter out opinions or data they did not like. Finally, there was no institutionalized time period during which consultation had to be conducted.

As late as 1982, when China’s leaders issued a constitution, there was still no procedure for representative government. The constitution document is not a contract between leaders and people involving the sharing of power; it does more to protect the state from individuals than to grant inalienable rights to participate in government to the people. Nor does the constitution treat the individual as the best judge of the individual’s interests. Such an assumption is part of representative democracy and the obligation of leaders to consult with the people’s representatives.

4 Distributional equality versus material incentives

The content of the right ideas that should motivate people from socialism forward are in part summed up in the terms ‘class analysis’ and ‘class struggle’. However, there is another aspect, the priority of moral over material incentives. Like democracy, to which it is related, the principal Maoist thesis against significant wage differences took shape during the guerrilla period. Like democracy as mass line, its fate was affected by the switch from small societies in the guerrilla base areas during the 1930s and 1940s to its application to China in its entirety after 1949. The Maoist position was attacked intermittently by opponents during the next quarter-century, and finally evaporated after the death of Mao.

Marx himself had no special fondness for spreading wealth around equally, although popular writers often attribute such a concern to him. In *Critique of the Gotha Programme* (1875) he describes inequality of distribution as characterizing the first phase of communist society. Marx states that ‘The right of the producers is proportional to the labour they supply’. The essential trait of communism is not distribution but the elimination of the division of labour. Mao himself criticized those absolute egalitarians who would object to officers riding on horses in periods of guerrilla warfare. Officers have responsibilities and needs that justify some unequal material privileges.

However, Mao idealized the roughly egalitarian society of Yenan and the other base areas during the anti-Japanese war. He wrote that they had a free supply system (take what you need) and minimal contention. After 1949, he stated, the introduction of a differential wage system along with status differences increased quarrelling. In his ‘Reading notes on "The Soviet Union’s Political Economy"’, he wrote: ‘Right up to the early stages of liberation, people on the whole lived an egalitarian life. They worked hard and fought bravely on the battlefront. They

absolutely did not rely on material incentive for encouragement, but on revolutionary spirit'. After 1949, wage differences did continue to exist, but the gap was modest.

There is always the assumption that people can be self-motivating in pursuit of communist goals, especially those whose minds have been transformed. For the Maoist, this obviates the need for material incentives. In Maoist theory, however, social arrangements must coexist with transformed minds in order for this nonreliance on material incentives to prevail. There must be at least the illusion of democratic decision making and equality of status. Here is one place where the topics of democracy and egalitarianism converge.

Critics of this position included economists and some supreme leaders (for example, President Liu Shaoqi in 1962), who questioned its impact on productivity. The argument was that people who have an 'iron rice bowl' (guaranteed wage) have no motive to work harder or to innovate. A free supply system would not work in a large country because it requires an abundance of wealth that can only come later in the development process. Inherent difficulties in the policy of opposing material incentives included the fact that it was incompatible with two values to which its advocates were also committed: social order and self-sufficiency. Social order requires a chain of command and differences between those who give and those who take orders. Efficiency in maintaining order dictates status distinctions between them, often manifest in their access to different styles of uniform, to houses and cars that have different monetary worth, and so forth. With regard to self-sufficiency, if officials ask regions to prize self-sufficiency, they are also telling poor regions to be satisfied with what they have and not ask for handouts from wealthier ones.

The major departure from the Maoist position came after Deng Xiaoping assumed power in 1978. Economic reformers argued that socialist goals could best be attained by rejecting central government planning in favour of market factors such as price and profit. This meant taking profit into account in organizing and running an enterprise, wide differences in wages, legal protection of contracts between economic entities and toleration of individual entrepreneurs. Given China's clan-village social structure, individual entrepreneurs could be either individual persons or villages taking the initiative in starting industries, using village-owned resources. The existing village elites (Communist Party officials) would see their incomes rise most rapidly, along with the incomes of managers working on contract to the village.

5 Humanism and class struggle

Within the ranks of Chinese Marxists, the most serious challenge to Maoism in the 1950s and 1960s came from critics characterized as 'humanists'. After Mao's death in 1976, those Marxists who criticized the Communist Party were also humanists. Some of them decried the alienation of people from leaders caused by the party's dictatorship. Others participated in the growth of a new movement, 'subjectivism', that centred on the autonomy of all individuals as possible and desirable.

'Humanism' stood for opposition to Maoism in the 1950s and 1960s because it was a broad term under which a variety of perspectives could unite. First and foremost, it symbolized repudiation of the Maoist position that class struggle is the principal agent of historical change and the agent that could propel China into the modern world. The doctrine of class struggle assumes that human society is divided into distinctly different classes, that analysis of humans in class terms is the only correct sociological methodology and that class hatred is an inevitable motive in that struggle. In an essay from 1942 entitled 'Zai Yenan wenyi zuotanhui shangdi jianghua' ('Talks at the Yenan Forum on Art and Literature'), Mao had taken two positions that authorize such conclusions. One is that there is no pan-human or universal human nature common to all people. There are only the differing natures of people of different classes; within a given class, natures are somewhat uniform. The other is that love or sympathy that transcends classes is impossible; there is only love for members of one's own class. Class antagonism is therefore inevitable.

Both these positions are totally alien to the dominant Confucian legacy regarding the nature of humans and the ethical position that should flow from that assumption about human nature (see Confucian philosophy, Chinese; Xing). Together, these Confucian positions carry the label of 'humanism', although the term has several meanings in China. In attacking humanism, Maoists were simultaneously supporting class struggle and opposing the most powerful philosophical legacy of the past. This was the teachings of Mencius, who in the fourth century BC first formulated the concept of a universal human nature. The essential features of this universal nature are a mind that

evaluates right and wrong, and a sentiment of compassion. Mencius based an ethical and political position on this psychological analysis, namely that each person has an obligation to implement these features. Acting on the sense of compassion, individuals should extend their affection beyond family members and those who suffer to encompass an increasingly large number of other persons. Class-transcending love and universal traits set Mencius in opposition to Mao. Thus, to attack humanism was to attack the old, or that part of the old that was the greatest ideological threat.

In modern China, three disciplines have inherited the traditional concern with the topic of human nature: literature, philosophy and psychology. During the 1950s and 1960s, each of these was affected by the conflict between class struggle and humanism. In literature, Ba Ren (a pseudonym of Wang Renshu, a literary editor, critic and diplomat) got into trouble in the 1960s for insisting on the right of authors to deal with things that all people universally share, namely the love of the fragrance of flowers and the song of birds, and the desire to eat and to be warm. The popular philosopher Feng Ding of Beijing University was attacked for claiming that there is a survival instinct common to all persons. Beginning in 1958, younger Chinese psychologists shifted away from an interest in neurologically-based character types to focus on the subjective class standpoint or thought that separates people into social groups.

These themes dominated intellectual activity, in so far as it existed during the Cultural Revolution, until Mao's death. They were reinforced by rivalry with the USSR. The Soviet leader Nikita Khrushchev had claimed that the Soviet Union had eliminated class antagonism and was thus a 'state of all the people'. This sounded to Maoists like an arrogant claim that the Soviet Union had leapt ahead of China into communism, and that Soviet leaders were practising class-transcending love where class divisions actually still existed. This Chinese critique appears in an influential essay published in 1963 by Chou Yang (1908-), a leader of the All-China Federation of Literary and Art Circles and a member of the Department of Philosophy and Social Science of the Chinese Academy of Sciences. In this case, international political disputes had spilled over into philosophical debates.

6 Humanism and the end of alienation

With the death of Mao, some prominent Chinese Marxists shifted their concerns away from opposing class struggle towards promoting the free development of the individual. They treated these goals as humanistic goals. However, they shifted the content of humanism away from its Confucian focus on humaneness, being interested instead in the development of other, different human traits. Influenced by the early writings of Marx, long popular in eastern Europe, they spoke of the individual's freedom as the absence of alienation and described it positively as the all-round development of the individual.

This was a monumental leap in China. Mao had manifested no interest in the individual at all, only in classes or groups. His idea of ending the division of labour and promoting all-round development was sending intellectuals to the farm and introducing peasants to schools. This was intended to end the mental-manual division of labour, an ideal with some Marxist justification. However, Marx in his early works had a richer vision, one symbolized in some famous words of *The German Ideology* (1846) that describe a person who can metaphorically 'hunt in the morning, fish in the afternoon, rear cattle in the evening, criticize after dinner', without ever being simply one-sided. This was also the Renaissance humanistic ideal of the fully-developed individual human.

The ideals of the early Marx are also remarkably convergent with the ideals of nineteenth-century British liberals. Indeed, John Stuart Mill was influenced by the same German Romantics that influenced Marx. For those Romantics, the heroes were the artists who put into canvas or into stone their own personalities; they realize themselves in their art (see Goethe, J.W. von; Hölderlin, J.C.F.). These are also the heroes of Marx's *Economic and Philosophical Manuscripts* (1844). These manuscripts became popular with some Chinese Marxists because they directed attention away from class or group to individuals and their self-realization, thus providing the basis for an ethics in which individuals count for something. The *Manuscripts* do so by asserting that all people are born with basic needs (to realize themselves by means of variety in their lives and to be free agents in their productive work). However, most people are unable to satisfy those basic human needs. Marx described communism as humanism because in communism, all individuals are able to satisfy those needs and become fully human. To those Chinese Marxists who prized individual freedom, the contrast with the Maoist vision of communism could not have been more stark.

The most prominent of the advocates of this style of humanism after Mao's death was Wang Ruoshui, a former editor of the official newspaper *People's Daily* and one of the leading reform theorists. Wang accepted the premise (he never tried to prove it) that there are universal human needs, the most basic of which is for self-realization in an all-round, free manner. From this premise, he derived a standard to judge acts or policies. Good acts or policies foster the satisfaction of basic needs; bad ones cause individuals to feel alienated or separated from their natures. Wang used this standard to criticize the Communist Party's dictatorship. In his view, in the 1980s the Party fostered alienation in the Chinese people; individuals had little control over their work and bureaucrats constantly intruded into their lives.

The counterattack on this brand of Chinese Marxist humanism was mounted by Hu Qiaomu, one-time secretary to Mao, author of a history of the Chinese Communist Party and, towards the end of his life, honorary president of the Academy of Social Sciences. With the help of ghostwriters from the Academy's Institute of Philosophy, Hu produced an article entitled 'Guanyu rendaozhuyi ho yihua wenti' (On Humanism and Alienation), which appeared in the journal *Hongai (Red Flag)* in 1984. Hu's thesis was that it is acceptable to incorporate humanism into Chinese Marxism when the term stands simply for routine ethical considerations, such as being kind to an elderly or sick person, but it is not permissible to treat it as a theory of history or as something generating a serious standard. This, he claimed, is what people like Wang Ruoshui had done, by describing humanity's gradual evolution from alienation to need-satisfaction over time and deriving from this process the top normative standard. This is to treat humanism as a theory of history, said Hu, which is wrong. It leads to the impermissible conclusion that the Chinese Communist Party causes alienation, which cannot exist in socialism. The only correct theory of history is historical materialism, with its claims about stages of history, and the standard that one can draw from historical materialism tells whether or not a given policy is right as a function of its appropriateness for a particular stage or moment in history. In Hu's analysis, the dictatorship of the party passes this test as it is appropriate for China's current stage.

7 Subjectivity

The other current away from Maoism that permeated Chinese Marxist philosophy, literary criticism and psychology was subjectivity. In some versions, it too made a place for the individual's worth. Especially in this regard, subjectivity represents a break during the Deng era from the Maoist social class-centred analysis of human beings. It focuses instead on individuals having unique identities. When it began in 1978, the subjectivity movement was a radical departure from the implications of the term 'materialism' in historical materialism. It rejected the idea of humans as puppets whose consciousness is entirely formed by their socio-economic environment; specifically, it was a departure from the Leninist theory of reflection (the inner or mental is simply a reflection of the outer or environmental) that had dominated orthodox Leninist epistemology in China since 1949. Later, the subjectivity movement took two different directions.

Li Zehou, a Marxist philosopher from the Institute of Philosophy, took the lead among those who treated subjectivity as a new direction in epistemology. The objective world, they argued, is shaped by the subjective interests, interpretative paradigms or psychological memories unconsciously sedimented in human minds. The mind is active, not a passive reflector, and objectivity changes as subjects with different interests or (over time) differently sedimented memories come on the scene.

The other view of subjectivity flourished in literary and psychological circles. Here, subjectivity meant the ability and desirability of the individual acting autonomously. The literary critic Liu Zaifu of the Institute of Literature at the Academy of Social Sciences encouraged writers to portray in their works characters who act freely and are conscious of their autonomy. In psychology, this approach was seen as consistent with the worldwide movement away from Freudian psychological reductionism and behaviourism. The autonomous individual's inner needs and choices exist and count. These Marxists wished China to acquire the perspective of the European Enlightenment, wherein the beliefs of individuals do not need to be determined by gods, destiny or rulers. The individual can and should be in control. In working out these various positions on subjectivity, Chinese Marxists drew into their own heritage ideas from Immanuel Kant in the West and Wang Yangming in China.

See also: Confucian philosophy, Chinese; Legalist philosophy, Chinese §2; Theories of history, Chinese §5; Marx, K. §§4, 8-12; Marxism, Western; Political philosophy, history of

DONALD J. MUNRO

References and further reading

Ch'en, J. (1970) *Mao Papers*, London: Oxford University Press.(Writings from 1917 to 69, along with an essay on Mao's literary style.)

Dirlik, A. (1989) *The Origins of Chinese Communism*, New York: Oxford University Press.(Ideology and organization in the early years, including the impact of such diverse strands as anarchism and guild socialism.)

Lenin, V.I. (1888-1924) *Collected Works*, Moscow: Progress Publishers, 1990, 47 vols.(See for the essay 'Two Tactics of Social Democracy in the Democratic Revolution'.)

Mao Zedong (1965) *Selected Works of Mao Tse-tung*, Beijing: Foreign Languages Press, 5 vols.(The official collection of edited works attributed to Mao, including the essays cited in this entry.)

(1968) *Four Essays in Philosophy*, Beijing: Foreign Languages Press. (Includes the two essays 'Shijianlun' ('On Practice') and 'Maodunlun' ('On Contradiction'). These essays owe much to the thought of Ai Siqi.)

(1974) 'Reading Notes on "The Soviet Union's Political Economy"', Joint Publications Research Service no. 61269-2, Washington, DC: US Government Printing Office.(This was not a public document in China; a copy was obtained and published by the US government.)

Marx, K. (1843-83) *Collected Works*, London: Lawrence & Wishart, 1975. (A complete edition of Marx's works in English; see for the works cited in this entry.)

Meisner, M. (1967) *Li Ta-chao and the Origins of Chinese Marxism*, Cambridge, MA: Harvard University Press. (On the reinterpretation of Marxism by one of the founders of the Chinese Communist party.)

Meissner, W. (1990) *Philosophy and Politics in China: The Controversy over Dialectical Materialism in the 1930s*, trans. R. Mann, Stanford, CA: Stanford University Press.(Discusses the meaning of 'philosophy' and 'science' in these years, with special attention to the essays 'On Practice' and 'On Contradiction'.)

Munro, D. (1978) *The Concept of Man in Contemporary China*, Ann Arbor, MI: University of Michigan Press. (The impact of Confucian and Soviet thought on the central theme of traditional Chinese philosophy in Chinese Marxism, and the resultant influence on education.)

Schram, S. (1963) *The Political Thought of Mao Tse-tung*, New York: Praeger. (Topically divided selections from Mao's writings, preceded by a lengthy interpretive introduction.)

Schram, S. (1974), *Chairman Mao Talks to the People: Talks and Letters 1956-1971*, New York: Pantheon. (Directives and statements that came to light in 'Red Guard' tabloids during the Cultural Revolution.)

Schwartz, B. (1958) *Chinese Communism and the Rise of Mao*, Cambridge MA: Harvard University Press.(The interaction of factional conflict and doctrine during the period 1918-33.)

Marxism, Western

Western Marxism is used here as an umbrella term for the various schools of Marxist thought that have flourished in Western Europe since Marx's death in 1883. It is sometimes used more narrowly to refer to those Marxist philosophers whose thinking was influenced by the Hegelian idea of dialectics and who focused their attention on the cultural as opposed to the economic aspects of capitalism. In the broader sense, Western Marxism does not denote any specific doctrine, but indicates a range of concerns that have exercised Marxist philosophers in advanced capitalist societies. These concerns primarily have been of three kinds: (1) epistemological - *what would justify the claim that Marxist social theory and, in particular, the materialist conception of history are true?; (2)*ethical - *does the Marxist critique of capitalism require ethical foundations, and if so, where are these to be discovered; and (3)* practical - *if the economic collapse of capitalism can no longer be regarded as inevitable, who are the agents who can be expected to carry through a socialist transformation?*

In relation to the first issue, the main debate has been between those who, following Engels, adhere to 'scientific socialism' (that is, the view that Marxism is a science in the same sense as the natural sciences), and those who claim that Marxist epistemology relies on a form of dialectics quite distinct from the methods of natural science. The most prominent exponent of this second view was the Hungarian philosopher Georg Lukács, who drew upon the dialectical method of Hegel, with class consciousness replacing Geist *(Spirit) as the vehicle of dialectical reason. Thus, for Lukács the truth of historical materialism and the goodness of communism were both established dialectically, through the class consciousness of the revolutionary proletariat. Lukács' advocacy of dialectics was later taken up and developed by the philosophers of the Frankfurt School.*

In relation to the second issue, early dissenters from the orthodox Marxism of Engels like Eduard Bernstein looked outside Marxism itself, and especially to the philosophy of Kant, for the ethical principles that would justify socialism. The position changed somewhat with the rediscovery of the young Marx's Paris Manuscripts *(1844) from which later Marxists, and in particular those associated with the Frankfurt School, were able to extract a humanistic ethics centred on the notion of alienation.*

In relation to the third issue, most Western Marxists continued to look to the proletariat as the agency of revolutionary change, often distinguishing, as did Lukács, between the true consciousness of that class and the false consciousness that reflected the distorting effects of bourgeois ideology. But in the case of the Frankfurt Marxists, the critical theory that pointed the way to a liberated human future was detached from any specific agency and treated merely as critique. The most original contribution was made by the Italian Marxist Antonio Gramsci, who argued that the working class must use the power of its ideas to establish hegemony over the other classes in bourgeois society, who would then join the proletariat in overthrowing capitalism.

The disintegration of Western Marxism began in the 1960s when the French philosopher Louis Althusser attacked both the use of Hegelian dialectics by Marxists and the various forms of Marxist humanism. Althusser insisted that Marxism was a science which required no ethical foundations. His critique was informed by a conviction that human subjectivity, together with the philosophical problems generated by subject-object dualism, are illusions.

Althusserian Marxism became fashionable in English-speaking universities, but its cavalier and paradoxical style also led, by reaction, to the rise of analytic Marxism in the late 1970s. Analytic Marxists returned to interrogating Marx's texts in more conventional ways, using the methods of analytic philosophy and contemporary social science to reformulate them to withstand academic scrutiny by non-Marxists. A tendency rather than a movement, analytic Marxism perhaps marks the final stage of a process that began with Lukács, that of turning Marxism into a purely academic study remote from politics.

1 Second International Marxism

Marx died in 1883 leaving his philosophical position unclear to his followers. His early philosophical writings were largely unpublished and unknown. He called his approach 'materialist' but it was not certain what this indicated beyond opposition to idealist philosophy and to religion, and a theory of history giving explanatory primacy to the production and reproduction of material life. His place was taken by Engels (§3), whose influence, until his death in 1891, was crucial for the way this uncertainty was resolved by the next generation of Marxist

theorists. Seldom philosophers, they were mostly politically engaged intellectuals involved in the rapidly growing working-class movements.

Engels considered Marx's economics a decisive scientific advance on classical political economy, and claimed that Marx did for history what Darwin had done for biological evolution. He believed these discoveries converged with those of natural science to demonstrate the truth of a dialectical and materialist ontology, and thought that epistemological and ethical questions could be resolved by scientific progress. Hence his description of Marxism as 'scientific socialism'.

Engels also tried to fill in the gaps in Marx's historical theory. He treated historical materialism as a causal theory, with causal interaction between 'base' and 'superstructure'. Development of the productive forces is determining only 'in the last instance'. He also added an 'ideological superstructure' to Marx's 'legal and political superstructure'. To keep scientific socialism from being ideological, he introduced the term 'false consciousness' (defined as a failure to see that the objective cause of supposed wants and ideals is class interest) and the notion that class interests, being the most powerful and uniform real motives, prevail in the long run by cancelling out short-term subjective motives. This was meant to explain why superstructures appear to be independent of, and neutral between, class interests when such is not the case, and why the scientific understanding of history differs from that of participants and chroniclers. But scientific socialism would enable the proletariat to pursue its class interest with the knowledge that its victory would then be inevitable, since it would coincide with the direction of material progress. Thus Engels pushed Marx's ideas rather naïvely towards a closed intellectual system.

Karl Kautsky, the leading theorist of Second International Marxism, interpreted Engels' philosophy in the light of the mixture of materialism, positivism and Darwinism popular among progressive German thinkers in the 1870s. Kautsky's outlook was highly determinist. He relied on the historical inevitability of capitalist crises and proletarian revolution for doctrinaire solutions to theoretical and political problems and failed to see that the inevitability of socialism supplied no moral argument for it. This determinism was allied to a cautious political strategy that welcomed trade unionism and parliamentarism as means of organizing the working class for the ultimate 'decisive hour', but denied that they could improve permanently the situation of the workers or prevent class polarization.

Kautsky's determinism was attacked by his revisionist colleague Eduard Bernstein. Bernstein, whose empiricism and 'evolutionary socialism' were influenced by the English Fabians, aimed to purge Marxism of revolutionary doctrine, which he blamed on the influence of Hegelian dialectics. He advocated a reformist strategy to achieve socialism legally and democratically. Stretching Engels' conception of causal interaction between base and superstructure, he argued that social evolution would give morality an increasing influence over production. He justified his moral position in terms of ethical universalism, in a neo-Kantian manner, denying that a practical doctrine like socialism can be scientific.

Orthodox Marxism was also criticized by sympathetic neo-Kantian philosophers dissatisfied with Engels' outdated materialism. Ethical socialists of the Marburg School, arguing that ethics cannot be based on anthropology, held that the Kantian principle that every human being ought to be treated as an end and not as a means prohibits workers being treated as commodities. Karl Vorländer saw Marx's theory of history and Kant's ethics as mutually complementary: the irreducible value of each individual implies worldwide human solidarity and justifies the connection between proletarian revolution and human emancipation.

The Kantian revival made its greatest impact on the Austro-Marxist school. In the early nineteenth century the leading intellectuals of the Austrian Social Democratic Party, Max Adler, Otto Bauer, Rudolf Hilferding and Karl Renner, shared an empiricist conception of scientific socialism as a system of sociology. Hilferding applied this perspective to economics, Renner to law and Bauer to politics. They shared a neo-Kantian commitment to ethical universalism while Adler supplied the epistemological foundations. Adler claimed, against the prevailing mechanistic interpretation of Engels' causal theory, that since all social phenomena are mental, social causality can be studied empirically only if Marx's concept of 'socialized humanity' is treated as a transcendental category of knowledge which is given a priori. Other Austrian Marxists, influenced by the neo-positivism of Mach and the Vienna Circle, opposed this Kantian Marxism.

Divergences within Marxism reflected new intellectual currents, national diversity and the pull of competing working-class ideologies. Thus Bernstein's revisionism was influenced by liberal reformism and Luxemburg's radicalism by syndicalism (see Luxemburg, R.). Syndicalism, an anarchist brand of trade-union militancy that took shape in France, became a revolutionary alternative to Marxism before the First World War, finding a champion in the unorthodox Marxist Georges Sorel who anticipated some later tendencies in Western Marxism.

Sorel rejected the notion of scientific socialism, together with most other rationalist *motifs* in Marxism, regarding Marx instead as a moralist like himself. He read *Das Kapital* (1867-) as a description of the corrupting economic system responsible for the decadence of bourgeois society and looked to working-class morality for its redemption. Becoming disillusioned with revisionism, he turned to syndicalism and direct action. For Sorel, class war was the central reality, and Marxism a 'myth' whose function was not to predict or guide but to inspire the workers to revolutionary industrial militancy. United and purified by violence, they would acquire the heroic moral qualities needed to destroy capitalism and found a new society of producers.

Meanwhile, Italian Marxism had also received a less orthodox stamp from the work of Antonio Labriola. An academic philosopher and anti-metaphysical Hegelian, Labriola's exposition of the materialist conception of history was opposed to inclusive, determinist and dogmatic interpretations, and his more open-minded approach, at once humanist and empirical, remained influential in Italy.

During this period the concept of ideology lost the negative connotation given to it by Marx and Engels as Marxists were forced to recognize the difference between science and morals. Bernstein argued that if all non scientific ideas are ideological, Marxism must be an ideology since it too is based on a moral ideal. Kautsky admitted the difference between scientific socialism and socialist ideals, but still distinguished the latter from 'bourgeois ideology' on the grounds that only these ideals now correspond with the needs of humanity. For Lenin, however, in exile in Switzerland, class struggle involved a straightforward conflict between bourgeois capitalist ideology and proletarian socialist ideology, whose outcome would be decided by history (see Ideology §1).

2 Western Marxism

Western Marxism in a narrower sense flourished between the Russian Revolution and the death of Stalin, an era of deepening division between Soviet communism and Western socialism. Its keynote was the claim that Marx was a dialectical philosopher in the Hegelian mould who had been misunderstood by Engels and Second International orthodoxy. It originated with Georg Lukács and Karl Korsch, philosophers who wrote after participating in failed revolutions in Hungary and Germany. In *History and Class Consciousness* (1921) Lukács argued that for Marx, as for Hegel, human history was a dynamic totality, an unfolding of reason which must be understood dialectically, and that this understanding is part of the process by which the whole is transformed. Its truth is available only to the historical agent - the proletariat - which becomes identical with the whole through the revolutionary action of consciously and deliberately transforming it. The proletariat is thus doubly privileged by history; by freeing itself it emancipates mankind from exploitation and it resolves the antinomies of philosophy in a transparent unity of theory and practice.

Lukács criticized empiricists and revisionists for isolating social facts from the whole since they can be understood and changed only by a total social transformation. Against Kantian and positivist Marxists he argued that the dualism of 'is' and 'ought' expresses a social separation of contemplation from practice that is overcome in the historical process, and so can be transcended in a dialectical understanding of history. By denying the scientific claims that were important to Marx and Engels, Lukács made the validity of Marxism depend on his own historicist conception of truth. This minimized the difference between social science and ideology; both are expressions of the 'reified consciousness' of the bourgeoisie and reflect the reified social relations of the market to which economics gives an illusory objectivity. There are no 'laws of dialectic' which science proves, as Engels maintained, nor laws of history external to the consciousness of the proletariat. Historical materialism is a critical-revolutionary method, exemplified by Marx's critique of bourgeois political economy or Lukács' own critique of bourgeois philosophy.

But Lukács also warned that the workers, as the 'identical subject-object of history', could only escape reification and false consciousness if enlightened by Marxist leadership. He eventually interpreted this theory of class consciousness in Leninist terms, and reconciled himself to Stalinism as a lesser evil than capitalism. Thus his ideas

had both a critical and an orthodox face, the latter increasingly predominant.

It was their critical aspect that inspired the ’critical theory’ of the Frankfurt School. The Frankfurt Institute of Social Research was established in 1923. Its central figures, active in Germany or the USA for the next forty years, were Max Horkheimer, Theodor Wiesengrund Adorno and Herbert Marcuse. Primarily philosophers, scholars at the Institute none the less developed new Marxist approaches to economics, politics and literature. Erich Fromm, a founder of the culturalist school of psychoanalysis, was also a member (see Frankfurt School).

Frankfurt Marxists were radical in thinking modern civilization diseased beyond hope of reform. They believed in the possibility of a utopian future, but not in a proletarian revolution, and had no political programme. Since they shared Lukács’ belief in the philosophical self-sufficiency of dialectics, their utopianism had to supply the dialectical premises, both epistemological and ethical, for their otherwise antifoundational critical stance. This was directed at both capitalism and Soviet state socialism, as well as at fascism, liberalism and all varieties of ’bourgeois’ philosophy, especially positivism and existentialism (see Critical theory).

Treating Marxism as method rather than theory, they drew on psychoanalysis to explain the sources and mass appeal of authoritarianism and extended Max Weber’s critique of bureaucracy. They emphasized the increasing integration of state and economy (’state capitalism’) and the replacement of traditional ideology with standardized escapist products of the modern ’culture industry’. They carried to implausible extremes the idea that Enlightenment rationalism had culminated in the fetishization of science and technology, which eroded human values and led to a conservative, totalitarian manipulation of a dehumanized society in the name of ’instrumental reason’. This was contrasted, in Hegelian fashion, with dialectical reason, but Adorno’s insistence on critical independence from all absolutes plunged his version of ’negative dialectic’ into obscurity and irrationalism. Although Frankfurt Marxists were mostly pessimistic, Marcuse and Fromm criticized Freud’s pessimism from the standpoint of Marx’s theory of alienation and took a more optimistic view of the possibilities for human enjoyment, creativity and community, appealing to essentialist beliefs about human nature (see Alienation §§3-5).

These German developments were paralleled by the reflections of the Italian Communist Antonio Gramsci during his ten years in a fascist jail. Gramsci had imbibed a neo-idealist critique of positivism from Croce (§4) and emphasized dialectical interaction between base and superstructure to a point where the concepts lost any precise meaning. Rejecting orthodox scientific socialism more radically than Labriola, and relying on Marx’s *Theses on Feuerbach* (1845), he treated Marxism as a historicist philosophy of human praxis, in which thought is a socially conditioned activity and truth is relative to its historical function. Belief in historical laws, he argued, was merely a necessary illusion of the proletariat in its early struggles, an idea close to Sorel’s view of Marxism as myth.

Gramsci’s contribution to the study of ideology was his theory of hegemony. In addition to exercising domination through the coercive apparatus of the state, ruling classes seek hegemony (leadership) through control of the cultural institutions of ’civil society’ - churches, schools, mass media and so on - and are stable only if they rule by consent. Since modern civil society is highly developed, the working class can introduce socialism only by achieving hegemony *before* seizing the state. Each class needs ’organic intellectuals’ to articulate its values, but the weakness of the workers in this respect had left them ’passively consenting’ to bourgeois values, although their actual behaviour showed this to be a type of false consciousness. Gramsci was a major source for the ‘dominant ideology thesis’ and has been criticized for exaggerating the role of ideas.

The strength of socialists and communists in France after 1945 encouraged fellow-travelling by *marxisant* philosophers. This coincided with a vogue for Hegel, the recent availability of Marx’s *Paris Manuscripts*, and a new interest in existentialism and phenomenology. The hybrid Marxism of Jean-Paul Sartre (§5), Maurice Merleau-Ponty and others was the result. For Sartre, Marxism meant historical materialism and class struggle, stripped of naturalistic justifications, and the philosophical task was to ground these existentially in individual praxis. He explained alienation by scarcity; subjective freedom is lost through the historical struggle to produce, resulting in the reified social objectivity of the ’practico-inert’ - a view more Hegelian than Marxian. This condemns individuals to ’serial’ consciousness from which they can be liberated only by the collective consciousness of the ’fused group’. Hence a Sorelian justification of violence for the sake of moral authenticity and, even more questionably, support for Stalinism as a radical denial of bourgeois objectivity (see Violence).

Like Sartre, Merleau-Ponty explicated Marx through Hegel’s dialectic of master and slave. At first, like Lukács, he defended Soviet totalitarianism as a lesser evil, which might in time produce a society without wage slaves. But after Stalin’s death he turned to social democracy and attacked Sartre’s combination of existentialism and Bolshevism. French Marxists of this period affiliated to the Frankfurt School or Lukács rather than existentialism were Henri Lefebvre and Lucien Goldmann.

Leon Trotsky, exiled from Russia in 1919, became an involuntary contributor to Western Marxism. His political analyses were influential, although a book on Marxist ethics made little impact. Together with Western Marxism, Trotskyism contributed to the New Left movement, which arose in English-speaking democracies in reaction to the Cold War. Marcuse’s *One-Dimensional Man* (1964) popularized a sub-Marxist analysis of late capitalism as a technocratic, consumerist totalitarianism. He endorsed militant students, racial minorities and Third World insurgents as the new agents of liberation. A side effect of the New Left was to promote research on subjects relevant to Marxism, and as Western Marxism penetrated east European universities it stimulated new critical studies there also.

3 The disintegration of Western Marxism

The French Communist philosopher Louis Althusser launched his structuralist alternative to dialectical Marxism in 1965 with the publication of his book, *Pour Marx*. It promised a scientific rehabilitation and modernization of Marx, drawing on a new theory of discontinuities in the history of science. Althusser claimed to detect an ’epistemological break’ between the humanist and historicist ideology of Marx’s youth, based on the ’problematic of the subject’, and the science of his maturity, which described a ’history without subjects’. Psychoanalysis suggested a ’symptomatic reading’ of *Das Kapital*, to extract the outlines of the new science, while structural linguistics provided a model for the resulting reinterpretation of historical materialism. This now became an ahistorical theory of ’structures in dominance’, exercising a ’structural causality’ whose effects are borne by human agents. In reality this was a kind of sociological functionalism. Gramsci’s theory of hegemony was stretched to make ideology the universal social cement that reproduces social structures, even under communism, by giving individuals an illusion of free agency. The truth of these theories was treated in a question-begging manner, as an effect of theoretical structures that are scientific and not ideological. Althusser’s approach, implausible as a reading of Marx, reflected a French philosophical interest in deconstruction of the subject which survived the eclipse of his school and the decline of French Marxism (see Structuralism).

Italian Marxism showed a parallel shift away from Hegelian Marxism, but in a different direction. Galvano della Volpe and Lucio Colletti reacted against the historicist aspects of Gramsci’s Marxism in favour of more empiricist readings of Marx. This has led Colletti to reject dialectics and the theory of alienation and, finally, socialism - one example among many of ’post-Marxism’.

Elsewhere, Althusserian Marxism was bypassed by more eclectic theorists who maintained that a critique of modern society drawing on Marx’s values could no longer be merely ’Marxist’. In Germany Jürgen Habermas transformed critical theory by supplementing Marx with theoretical perspectives drawn from sociology, linguistic philosophy and hermeneutics. He criticized Marx for viewing history and human emancipation reductively, in terms of labour, and claimed that interaction, involving communication and social norms, is an independent dimension of social life. A human interest in emancipation is the constitutive ground of the social sciences, whose method is critical because they are implicitly oriented to an ’ideal speech situation’. Their truth, like other human values, is a function of dialogue freed from ideological distortion by the equal participation of all. Habermas’ critics, however, deny that he can demonstrate that free dialogue on equal terms presupposes consensus on values.

Analytic Marxism developed in English-speaking countries in the late 1970s and 1980s, partly as a reaction of academic socialists to Althusserian excesses. It is characterized more by respect for traditional canons of argumentation than by theoretical or political consensus. G.A. Cohen’s cogent demonstration that Marx intended historical materialism to be a functional theory challenges both the causal (Engels) and dialectical (Lukács) interpretation. He defended Marx’s functional explanations in a debate with Jon Elster, for whom functional statements are no more explanatory than dialectics. Elster’s own contributions include a recasting of Marx’s class theory in rational choice terms, and a psychological approach to the explanation of ideology. Other writers have sought to present Marx’s epistemology as a form of scientific realism.

Analytic Marxists were responding in part to the resurgence of economic and political liberalism in the 1980s. This shows in their concern to make Marx's theory of exploitation independent of his discredited economics and, more generally, in a wish to reinstate individuals in Marx's thought - both methodologically, as providing microfoundations for historical explanations, and normatively, as intended beneficiaries of a communist utopia. Hence another debate, this time over whether Marx had, or ought to have had, a theory of justice, and whether it could be consistent with his theory of ideology. Radical critics see analytic Marxists as engaged in a bourgeois academic pastime. But they can also be seen as conservationists, preserving the valid residue of Marxism for a post-Marxian world.

See also: Marxist philosophy, Chinese; Marxist philosophy, Russian and Soviet; Marxist thought in Latin America

JOHN TORRANCE

References and further reading

Althusser, L. (1965) *Pour Marx*, Paris: Maspero; *For Marx*, trans. B. Brewster, London: Allen Lane, 1969. (Perverse, dogmatic, challenging and influential critique of Marxist humanism.)

Anderson, P. (1976) *Considerations on Western Marxism*, London: New Left Books.(Interesting critique by an independent-minded Marxist.)

Bottomore, T. and Goode, P. (1978) *Austro-Marxism*, Oxford: Clarendon Press.(Representative and still interesting selection, including Max Adler on causality and teleology.)

Cohen, G.A. (1978) *Karl Marx's Theory of History: A Defence*, Oxford: Oxford University Press. (Convincing critical reconstruction of Marx's philosophy of history. A demanding text.)

Colletti, L. (1969) *Marxism and Hegel*, trans. L. Garner, London: Verso, 1979.(Wide-ranging and profound critique of the Hegelian heritage within Marxism.)

Engels, F. (1878) *Anti-Dühring*, trans. E. Burns, *Herr Eugen Dühring's Revolution in Science*, Peking: Foreign Languages Press, 1976.(A polemic which provided the most systematic exposition of 'scientific socialism' as understood by Second International Marxists.)

Engels, F. (1888) *Ludwig Feuerbach and the End of Classical German Philosophy*, in K. Marx and F. Engels *Collected Works*, London: Lawrence & Wishart, 1975, ch. xxvi, 353-98.(Source of the orthodox view of how dialectical materialism superseded Hegelian idealism, and of Engels' theory of false consciousness.)

Elster, J. (1985) *Making Sense of Marx*, Cambridge: Cambridge University Press.(Brilliant and dramatic excision of dead wood by a leading analytic Marxist.)

Geras, N. (1985) 'The Controversy about Marx and Justice', *New Left Review* 150: 47-85; and A. Callinicos *Marxist Theory*, Oxford: Oxford University Press, 1989, ch. 7.(Comprehensive critical review of major controversy on Marx's ethics.)

Kolakowski, L. (1978) *Main Currents of Marxism*, vol. 2, *The Golden Age* and vol. 3, *The Breakdown*, trans. P.S. Falla, Oxford: Clarendon Press.(Magisterial and comprehensive critical account up to 1968; thinner on recent French and Italian Marxism.)

Lukács, G. (1921) *History and Class Consciousness: Studies in Marxist Dialectics*, trans. R. Livingstone, London: Merlin, 1971.(Epoch-making essays, seminal for the Western Marxist revival of dialectic philosophy and for critical theory.)

Marcuse, H. (1964) *One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society*, Boston, MA: Beacon Press and London: Routledge. (Popularized Frankfurt Marxism, 1960s style.)

Marx, K. and Engels, F. (1934) *Selected Correspondence 1846-1895*, London: Lawrence & Wishart. (Of particular interest are the letters from Engels to Schmidt (27 October 1890) and Mehring (14 July 1893), which offer influential glosses on the interpetation of Marx's theory of history.)

Merquior, J.G. (1986) *Western Marxism*, London: Paladin.(Readable, sophisticated and engagingly judgmental account from Lukács and Gramsci to 'post-Marxism'; full bibliography.)

Volpe, G. della (1960) *Critique of Taste*, trans. M. Caesar, London and New York: Verso, 1991.(The author at his best, an original approach to Marxist aesthetics.)

Marxist philosophy of science

Marx's approach to science is an intriguing combination of respect for the natural sciences and empirical inquiry, determination to go beyond the description of regularities among observable phenomena, and insistence on the inevitable impact of social circumstances on scientific inquiry. Marx thought that the human sciences and the natural sciences are governed by essentially the same methods, that natural-scientific theories give us enhanced insight into mind-independent reality, and that our most fundamental views are subject to revision through scientific inquiry. Yet Marx rejected the ideal of scientific method according to which rational scientific belief is tied to observational data through a canon of rules as general, timeless and complete as the rules of logical deduction. While traditional empiricists emphasize the economical description of empirical regularities which could, in principle, be used to predict the occurrence of observable phenomena, Marx emphasizes the description of underlying causal structures, employing concepts that are typically irreducible to the vocabulary of mere observation, and causal hypotheses that sometimes do not even sketch means of prediction. Similarly, though Marx shared the optimistic view that science gives rise to long-term improvement in our insight into underlying causes, he disagreed with many epistemic optimists in his insistence that scientific inquiry is inevitably and deeply affected by social interests and relations of social power.

Since Marx's general comments on scientific method are few and scattered, a 'Marxist philosophy of science' consists of the further development of this intriguing mixture.

1 Explanation and theory in Marx

Marx's social theorizing has been attacked as a muddled perversion of science, lacking adequate predictive power and insufficiently connected with empirical data. Consider Marx's materialist conception of history: any stable society's political and cultural institutions have their most important features because their having them serves to maintain the dominant relations of production (the relations of control that dominate material production); an internally-caused change in a society's dominant relations of production is due to processes within the mode of production (the relations of production, work relations and technology) which create desires and powers destroying the old system of relations of production. Marx was well aware that similar networks of relations of production have been stabilized by importantly different political and cultural institutions. He acknowledged that one process (for example, the formation of a proletariat under capitalism) may ultimately doom a social system while a similar process (for example, the formation of the Roman proletariat) has no transformative effect. In a poignant letter to Russian activists who sought his guidance, Marx denied that his general theory provides any 'master key' permitting one to distinguish transformative from non-transformative processes without knowing their outcomes. So all that is rational in his general theory of social change might seem to be trivial: a tautology, that basic change occurs when changes in the mode of production are powerful enough to produce them, together with a banality, that there are important changes within the mode of production in the period leading up to a fundamental change in society as a whole.

A Marxist philosophy of science counters with an alternative view of science. A theory provides a description (sometimes a vague one) of a repertoire of underlying mechanisms, and says that the instantiation of these mechanisms has a certain importance as the cause of certain phenomena. A theory is confirmed by showing that such reliance on its repertoire does a better job than reliance on rival theories of explaining phenomena that stand in need of explanation. Such an argument will partly consist of the explanation of particular phenomena, especially those that have posed long-standing problems for prior explanatory schemes. It will also depend on more general inferences, resting on shared background principles, as to the possible causes of the general kind of phenomenon in question - for example, Marx's arguments that stabilizing tendencies of cultural and political institutions will prevent basic change arising from parts of a social system that are outside the mode of production.

A theory that triumphs in this way need not be useful as a basis for prediction. Marx, of course, *was* committed to certain long-term predictions, but (in this controversial view of Marx which implies a distinctive philosophy of science) these predictions are based on his specific account of capitalism. For Marx, the essence of confirmation is the critical comparison of theories (as Althusser has emphasized), in which the empirical data to be explained are one crucial lever of comparison (which Althusser has been charged with neglecting).

The development of auxiliary hypotheses connecting the general mechanisms in the basic repertoire with relevant concrete phenomena may be an extremely demanding task. In remarks on his economics of capitalism which sometimes liken it to Daltonian chemistry, Marx emphasizes the distance between the processes of class struggle and technological change which he describes using the vocabulary of the labour theory of value, and the economic facts which the underlying processes help to explain, facts which are to be described in the vocabulary of commerce. Although he takes the former processes to be real and fundamental, he thinks they never produce their effects without the mediation of other, market-based processes, the topics of the second and third volumes of *Capital*. Nowak (1980) takes this process of abstraction from, followed by reconstruction of, concrete phenomena to be the essence of Marx's scientific method.

2 Holism

Marx has also frequently been charged with assuming an implausible dominance of social wholes over individual parts in which large-scale social processes have obscure powers to manipulate individual participants (see Holism and individualism in history and social science). Thus, the goals of Cromwell and his followers were mainly religious and political concerns, not concerns to end the feudal restriction of emergent capitalist forces of production. Yet Marx takes the English Revolution to inaugurate the transformation of England into a capitalist society, a transformation he explains as serving to overcome feudal barriers to expansive capitalist production.

Part of Marx's implicit answer to the charge of arcane manipulation is a distinction between explanatory adequacy and causal sufficiency. Marx's historical writings consist largely of the richly detailed description of networks of individualist causes, actions governed by the reasons and resources of individual participants. Yet he thinks such individualistic description lacks the depth needed to explain major social transformations: the transformation would have occurred anyway. Thus, if the political, religious and military events that led to the Commonwealth, the Stuart Restoration and the Glorious Revolution had not transpired, a modernizing, bureaucratic state, advancing the interests of enterprising gentry and manufacturers, would have arisen anyway, on account of prior trends within the mode of production.

Still, the social outcome will result from some network or other of individualist causes. How can large-scale functional characteristics such as the expansive needs of capitalist production regulate these individualist causes when the satisfaction of these functions is not a typical dominant reason for which participants act? Here, Marx is psychologically ingenious in a way that breaks down rigid separations of the structural from the individual. He thinks that the history of ideas (for example, the history of religious differences in seventeenth-century England) establishes the enormous importance of mechanisms moulding belief or desire that do not consist of the subject's reasons for belief or desire. For example, he emphasizes the tendency of beliefs and goals to adjust to material interests, even among those who would accurately deny that these interests are their reasons for believing and desiring as they do. In his favoured explanatory scheme, rational choices within the mode of production have unintended consequences, transforming the pattern of resources and material interests (most importantly, by creating an interest in the destruction of inhibiting social forms on the part of a newly powerful economic class); then, consequent adjustments of goals and beliefs give rise to new psychologies that are the vehicle of social change. Thus Marx's project of explaining large-scale phenomena is also the basis for a novel view of the mechanisms of change in individual participants.

3 Materialism and dialectics

Marx takes both the social and the natural sciences to vindicate a materialist, dialectical view of reality. In its broadest claim, Marx's materialism has it that everything that exists is made of matter and is governed by the causal powers that physics seeks to discover. However, his materialism is not reductive. Within the constraints that the laws of physics impose, other laws regulate phenomena, in ways that cannot be described in the vocabulary of physics (see Dialectical materialism; Reduction, problems of; Unity of science).

Marx's view of cognition is also materialist in that it gives priority to material interests and mind-independent causes. Cognition grows out of the effort to change reality in order to satisfy one's needs. Initially, our concepts mark as relevant the distinctions that are practically important to us. The frustration of our efforts and expectations, together with the social exchange of information which is our main resource for effective change, lead to the continual revision of beliefs and even of the concepts in which they are framed. Like those modern

‘scientific realists’ who emphasize the success of theory-dependent interventions, Marx took the question of whether the whole process of inquiry improves the truth of our beliefs to be settled by its contribution to our capacity to get our way with nature (see Realism and antirealism).

The dialectical side of Marx’s outlook, which reconstructs Hegel’s logic on a materialist basis, was presented much more systematically by his co-worker Engels. In this dynamic view, every part of total reality is at once an active source and a passive object of change and no part is completely distinguishable from the rest. It is of the nature of everything to change from internal causes, because what stabilizes anything is a temporary balance of polar opposites whose conflict dictates transformation and, eventually, self-destruction. (Two alleged examples are the antagonistic relation between the proletariat and the bourgeoisie that characterizes capitalism and the balance of attraction and repulsion in a physical object.) Engels also emphasizes the tendency, in appropriate circumstances, for quantitative changes to become qualitative (for example, when water becomes ice, or a rural English bourgeoisie emerged from the growing prosperity of the better-off Tudor farmers) and the prevalence of spiralling processes of change in which an initial transformation is followed by another that restores aspects of the first stage (the ’negation of the negation’.)

There is much banality and word-juggling in Engels’ illustrations of these principles. But there is also much shrewdness and wide-ranging scientific erudition in his criticism of opposed ‘mechanistic’ approaches, which analyse systems as arrangements of passive, simple, unchanging elements.

4 Objectivity and society

If cultural institutions tend to serve the interests of economically dominant classes, then explanatory frameworks will often be favoured on account of processes with no tendency to produce truth. If cognition grows out of the effort to satisfy human needs in a largely purposeless world, then our ways of characterizing causes will probably always misrepresent nature’s actual causal repertoire. Marx and Engels certainly take scientific inquiry to be subject to these limits. Thus, Marx associates the rise of a politically assertive working class in mid-nineteenth-century Britain with the retreat from hypotheses of conflicting class interests among established economic theorists, while Engels takes reluctance to countenance the evolution of species and celestial bodies to reflect commitment to the fixity of the social order. And both emphasize that any science is bound to misdescribe its subject matter in important ways.

However, mainstream Marxist thinking asserts that scientific inquiry has yielded rational truth claims of broader and broader scope. Outside of a few platitudes, the most that we can rationally claim is that a hypothesis is true within a certain limited sphere, or true enough for certain purposes, or capable of correctly resolving the current controversies with the most important bearing on scientific progress. But (as Lenin put it, in criticizing what would now be called ‘social constructivist’ philosophies of science) it is often rational to claim access to objective truths of this relative kind.

Scientific inquiry is, inevitably, social, and social processes are inevitably affected by social interests in ways of which the participants are not fully aware. Still, in Marx’s view, the impact of social interests does not always distort truth or rationality. For example, he takes the process of increasingly collective resistance which gave rise to class-conscious workers’ movements to be interest-driven*and* to have enhanced truth and rationality in theorizing allied with those movements. Nor is the pressure of social interests intellectually decisive, for him, all the time - as witness his respect for the achievements of the quite bourgeois economists of late eighteenth- and early nineteenth-century Britain.

Even if Marx’s frequent attributions of religious, political and moral trends to truth-distorting social forces are not preliminaries to an anti-objectivist philosophy of science, they might be extended into the social history of philosophies of science and allied epistemologies themselves. For example, Lenin (1908) argues that the currency of Mach’s radical empiricism is partly due to its appeal to those who seek to preserve room for escape into religious wishful thinking: if the most that theoretical science can do is to organize experience in satisfying ways, then religion does as well for the devout.

Lukács’ essay, ‘Reification and the Consciousness of the Proletariat’, is perhaps the most wide-ranging project of this kind. He traces the dilemmas of modern epistemology to the psychopathology of life in a market-dominated society. Routine dependence on labour-markets in which impersonal forces determine the monetary value of the

use of one's labour-power leads people to separate their activities into an immediately experienced subjective sphere and an impersonal objective sphere and to question whether the subjective can reflect objective facts. Market-dominance also gives rise to a formal, calculative conception of rationality, on which people rely to re-establish objective validity. But this reliance only makes the problem more acute, since rationality based on formal principles gives no significance to the content of experience. Lukács subsequently abandoned this approach, on the ground that it treated the real difference between the subjective and the objective as an artefact of ideology, but it has had an extensive influence on subsequent Continental criticisms of positivist philosophy of science, the main twentieth-century legacy of traditional empiricism.

Arguments that presuppose the truth of Marxist social theory have never been important in Anglo-American philosophy of science, and are now virtually extinct. But the larger attitudes toward science of Marx and his successors are very much alive in the search for a replacement for positivism that has dominated philosophy of science since the 1950s. For those who emphasize both the reality of theoretical posits and the need to connect them with observations, the inadequacy of canons of induction and the genuine limits of rational disagreement, the autonomy of large-scale structures and the importance of their micro-constitution, the objective validity of inquiry and its dominance by human interests, the Marxist tradition is as much a home as any and a richer source of inspiration than most.

See also: Neurath, O.

RICHARD W. MILLER

References and further reading

Though none of these works employs extensive technical apparatus, those by Althusser and Lukács are notoriously obscure.

Althusser, L. and Etienne B. (1968) *Reading Capital*, trans. B. Brewster, London: New Left Books, 1970.(The first two parts, by Althusser, derive from Marx a conception of knowledge as internal to 'theoretical practice'.)

Cohen, G.A. (1978) *Karl Marx's Theory of History*, Princeton, NJ: Princeton University Press.(An argument for an interpretation of Marx's theory which would *not* require a distinctive philosophy of science sharply breaking with positivism.)

Engels, F. (1878) *Anti-Duehring*; trans. E. Burns, ed. C.P. Dutt, New York: International Publishers, 1939. (Influential discussions of dialectics and natural science, interspersed with polemics against the long-forgotten Dühring.)

Engels, F. (1934) *Dialectics of Nature*, trans. C. Dutt, Moscow: Progress Publishers.(Notebooks, 1873-86, in which Engels develops a dialectical interpretation of the achievements of natural science.)

Lenin, V.I. (1908) *Materialism and Empirico-Criticism*, Moscow: Progress Publishers.(Criticisms of Russian followers of Mach, Duhem and other pioneers of modern philosophy of science, with plenty of tiresome invective but some shrewd and sensible thrusts.)

Lukács, G. (1971) *History and Class Consciousness*, trans. R. Livingstone, Cambridge MA: MIT Press.(Includes 'Reification and the Consciousness of the Proletariat' 1922, and a 1967 preface in which Lukács criticizes his earlier views as insufficiently materialist.)

Marx, K. (1975) *Texts on Method*, trans. and ed. T. Carver, Oxford: Blackwell.(A useful pairing of Marx's two most extensive discussions of methodology, the Introduction (1857) to the *Grundrisse* and Notes (1879-80) on a textbook by Adolph Wagner. Marx's other, scattered comments on method and natural science are mostly in *Capital*, the *Grundrisse*, *The German Ideology*, the delphic 'Theses on Feuerbach' and his correspondence.)

Miller, R.W. (1984) *Analyzing Marx*, Princeton NJ: Princeton University Press.(Chapters 3-6 develop the interpretation of Marx's theories and methodology and their bearing on modern philosophy of science that was sketched in §§1-2 of this entry.)

Nowak, L. (1980) *The Structure of Idealization*, Dordrecht: Reidel.(Derives from Marx's methods in *Capital* a conception of scientific reasoning as based on idealization.)

Marxist philosophy, Russian and Soviet

The history of Russian Marxism involves a dramatic interplay of philosophy and politics. Though Marx's ideas were taken up selectively by Russian populists in the 1870s, the first thoroughgoing Russian Marxist was G.V. Plekhanov, whose vision of philosophy became the orthodoxy among Russian communists. Inspired by Engels, Plekhanov argued that Marxist philosophy is a form of 'dialectical materialism' (Plekhanov's coinage). Following Hegel, Marxism focuses on phenomena in their interaction and development, which it explains by appeal to dialectical principles (for instance, the law of the transformation of quantity into quality). Unlike Hegel's idealism, however, Marxism explains all phenomena in material terms (for Marxists, the 'material' includes economic forces and relations). Dialectical materialism was argued to be the basis of Marx's vision of history according to which historical development is the outcome of changes in the force of production.

In 1903, Plekhanov's orthodoxy was challenged by a significant revisionist school: Russian empiriocriticism. Inspired by Mach's positivism, A.A. Bogdanov and others argued that reality is socially organized experience, a view they took to suit Marx's insistence that objects be understood in their relation to human activity. *Empiriocriticism was associated with the Bolsheviks until 1909, when Lenin moved to condemn Bogdanov's position as a species of idealism repugnant to both Marxism and common sense. Lenin endorsed dialectical materialism, which thereafter was deemed the philosophical worldview of the Bolsheviks.*

After the Revolution of 1917, Soviet philosophers were soon divided in a bitter controversy between 'mechanists' and 'dialecticians'. The former argued that philosophy must be subordinate to science. In contrast, the Hegelian 'dialecticians', led by A.M. Deborin, insisted that philosophy is needed to explain the very possibility of scientific knowledge. The debate was soon deadlocked, and in 1929 the dialecticians used their institutional might to condemn mechanism as a heresy. The following year, the dialecticians were themselves routed by a group of young activists sponsored by Communist Party. Denouncing Deborin and his followers as 'Menshevizing idealists', they proclaimed that Marxist philosophy had now entered its 'Leninist stage' and invoked Lenin's idea of the partiinost' *('partyness') of philosophy to license the criticism of theories on entirely political grounds. Philosophy became a weapon in the class war.*

In 1938, Marxist-Leninist philosophy was simplistically codified in the fourth chapter of the Istoriia kommunisticheskoi partii sovetskogo soiuza (Bol'sheviki). Kraatkii kurs (History of the Communist Party of the Soviet Union (Bolsheviks). Short Course). The chapter, apparently written by Stalin himself, was declared the height of wisdom, and Soviet philosophers dared not transcend its limited horizons. The 'new philosophical leadership' devoted itself to glorifying the Party and its General Secretary. The ideological climate grew even worse in the post-war years when A.A. Zhdanov's campaign against 'cosmopolitanism' created a wave of Russian chauvinism in which scholars sympathetic to Western thought were persecuted. The Party also meddled in scientific, sponsoring T.D. Lysenko's bogus genetics, while encouraging criticism of quantum mechanics, relativity theory and cybernetics as inconsistent with dialectical materialism.

The Khrushchev 'thaw' brought a renaissance in Soviet Marxism, when a new generation of young philosophers began a critical re-reading of Marx's texts. Marx's so-called 'method of ascent from the abstract to the concrete' was developed, by E.V. Il'enkov and others, into an anti-empiricist epistemology. There were also important studies of consciousness and 'the ideal' by Il'enkov and M.K. Mamardashvili, the former propounding a vision of the social origins of the mind that recalls the cultural-historical psychology developed by L.S. Vygotskii in the 1930s.

However, the thaw was short-lived. The philosophical establishment, still populated by the Stalinist old guard, continued to exercise a stifling influence. Although the late 1960s and 1970s saw heartfelt debates in many areas, particularly about the biological basis of the mind and the nature of value (moral philosophy had been hitherto neglected), the energy of the early 1960s was lacking. Marxism-Leninism still dictated the terms of debate and knowledge of Western philosophers remained relatively limited.

In the mid-1980s, Gorbachev's reforms initiated significant changes. Marxism-Leninism was no longer a required subject in all institutions of higher education; indeed, the term was soon dropped altogether. Discussions of democracy and the rule of law were conducted in the journals, and writings by Western and Russian émigré

philosophers were published. Influential philosophers such as I.T. Frolov, then editor of Pravda, called for a renewal of humanistic Marxism. The reforms, however, came too late. The numerous discussions of the fate of Marxism at this time reveal an intellectual culture in crisis. While many maintained that Marx's theories were not responsible for the failings of the USSR, others declared the bankruptcy of Marxist ideas and called for an end to the Russian Marxist tradition. Following the collapse of the Soviet Union in 1991, it seems their wish has been fulfilled.

1 The beginnings of Russian Marxism

It is not surprising that Marxism captured the imagination of the radical Russian intelligentsia in the late nineteenth century. Questions of social justice, so pressing in Tsarist Russia, dominated philosophical debate. Marxism, a putatively scientific theory of history that promised humanity's self-realization through the liberation of the oppressed masses, appealed to Russian thinkers, preoccupied with human destiny and Russia's role within it, and anxious for a vision that would unite theory and practice. The story of the subsequent domination of Russian philosophy by Marxism is more remarkable, involving a dramatic interplay of philosophy and politics. On the one hand, the Bolshevik project was inspired and guided by a philosophical doctrine; on the other, it resulted in the almost total subordination of philosophical enquiry to political power.

Russian radicals in the 1870s were acutely aware that Marx's theories were written for the industrial nations of Western Europe, not a quasi-feudal autocracy like Russia with a largely peasant economy (*Capital* was published in a Russian translation in 1872). Thus Marxist ideas were at first adopted selectively. The Russian populists, whose leading thinkers were P.L. Lavrov (1823-1900) and N.K. Mikhailovskii (1842-1904), invoked Marx's economic theory in their critique of the horrors of capitalism, but argued that Russia could avoid capitalism altogether by adopting directly a form of communism based on the peasant commune.

The populists differed on how this transformation was to be achieved. While some insisted the masses must engineer their own emancipation, others, like Pëtr Tkachëv (1844-86 NS), favoured a revolution led by a centralized party of conspirators. It was the adventurism of this latter group that pushed some in the populist movement towards a more comprehensive Marxism. Georgii Plekhanov (1856-1918) argued that the socioeconomic forces at work in Russia in fact conformed to Marx's theory of history: capitalism was indeed developing apace and could not be sidestepped. But the admission that capitalism was the unavoidable precondition of a socialist future raised a moral problem. How could revolutionaries endorse the coming of capitalism, when it would bring misery to the masses they sought to liberate? Plekhanov argued that it was futile to resist the inevitable on moral grounds. However, Marxists need not sit idly by; they must awaken the proletariat's self-consciousness. Plekhanov also maintained, percipiently, that the imposition of socialism would result in a 'patriarchal and authoritarian' communism, worse than enduring the evils of capitalism.

Plekhanov sought to justify his economic determinism by placing it in a broader materialist vision, and in *K voprosu o razvitii monisticheskogo vzgliada na istoriiu (The Development of the Monist View of History)* (1895) he articulates a conception of philosophy that came to dominate Russian Marxism. Inspired by Engels, Plekhanov argues that the history of philosophy is a battle between materialism and idealism. The former aspires to explain everything in material terms, the latter in terms of spirit or idea. (Like most Russian thinkers, Plekhanov is profoundly attracted to *monistic* positions; he simply dismisses dualism as unable to explain how its two explanatory principles - matter and spirit - are related (see Dualism).) Endorsing materialism, Plekhanov praises the French thinkers of the eighteenth and nineteenth centuries, especially Helvétius, for holding that ideas are determined by the social environment. These thinkers, however, mistakenly portray human history as the outcome of ideas ('opinions govern the world') or a reflection of an unchanging human nature. Paradoxically, a more plausible materialism was made possible by Hegel's dialectical idealism. In contrast to static, metaphysical conceptions of reality, Hegel stressed that phenomena are interconnected and constantly changing. He saw that gradual, quantitative changes in phenomena eventually result in their abrupt qualitative transformations, that every phenomenon is eventually transformed into its opposite, and that contradiction is the motor of development. Hegel, however, also misunderstood history, representing it as a manifestation of the 'world spirit' (see Hegel, G.W.F.). The genius of Marx and Engels was to invent a form of dialectical materialism (see Dialectical materialism). Their view is materialist because it represents cultural, social, economic, legal, political and ideological relations (the superstructure) as ultimately determined by the level of development of material forces of production (the

economic base). It is dialectical in its dynamic vision of history. Human beings, acting collectively and with the use of tools, are able to change their environment so radically that this in turn transforms their own nature. However, as the level of productive forces increases and human beings master the natural world, they become ever more enslaved to economic forces unleashed by their activity. Marx's theory of history reveals the true character of those forces and how they will be mastered with the emergence of communism. The proletarian revolution represents a transition from the realm of necessity to the realm of freedom. Human beings will no longer labour under alien powers, but will be able knowingly to conform their actions to the dictates of necessity.

2 Orthodoxy and revision before 1917

Though Plekhanov's dialectical materialism quickly became orthodoxy, it struck many as unsatisfactory, for his determinism seemed to preclude human agency and moral criticism. The school of legal Marxists, so called because its representatives published legally under their own names, responded to this failing. These thinkers, led by Pëtr Struve (1870-1944), endorsed Marx's economic theory as proof that Russian absolutism was doomed by the inevitable development of capitalism (which, in contrast to the populists, they saw as a progressive force). They argued, however, that Marxism could not inspire moral renewal unless supplemented with a broader philosophical and ethical perspective. To this end, the legal Marxists turned to Kant (anticipating Eduard Bernstein). The resulting synthesis was fragile, and in 1900 most legal Marxists migrated from social democracy to form the basis of the liberal Constitutional Democratic Party. Struve, N.A. Berdiaev (1874-1948), S.N. Bulgakov (1871-1944) and S.L. Frank (1877-1950) eventually embraced an idealism based on Russian religious thought, which they presented in the symposium *Vekhi (Signposts)* in 1909, provoking a scornful attack by Lenin (see Signposts movement).

Another significant revisionist view was the attempt, by A.A. Bogdanov (1873-1928), A.V. Lunacharskii (1875-1933) and others, to reconcile Marxism with the empiriocriticism of Ernst Mach and Richard Avenarius. The Russian empiriocritics were drawn to Mach's positivist view of science, seeing it not as describing reality beyond appearance, but as providing progressively better ways of organizing experience. Scientific laws do not pick out real necessities in nature, but read order into the deliverances of the senses. The Russians supplemented Mach's neo-Darwinist epistemology with the view that reality is *socially* organized experience, taking this to fit Marx's view in the *Theses on Feuerbach* that objects should be conceived 'as sensuous human activity' (see Russian empiriocriticism).

Bogdanov's *Ėmpiromonizm* (1905-6) maintains that the very dichotomy between materialism and idealism is confused. The mental-physical distinction does not pick out two kinds of entity or substance. It is rather a distinction drawn within experience. Some elements of experience are deemed 'mental' because they are given only to one subject; others are called 'physical' because given to many. Bogdanov argued that under communism, where disagreement and conflict are eradicated, people will share the same modes of organizing experience and the idea of individuals as separate mental worlds will break down. Other consequences of the theory were less extravagant. Bogdanov's interest in organization led him to develop tektology, a precursor of cybernetics and general systems theory.

In 1902, the Russian Social Democratic Party split into two factions. The Bolsheviks embraced the conception of the vanguard party set out in Lenin's *Chto delat' (What Is To Be Done?)* (1902), a centralized and disciplined organization of professional revolutionaries (reminiscent of Tkachëv's view). In contrast, the Mensheviks advocated a broad party embracing many kinds of proletarian support. This conflict reflected different conceptions of revolution. While the Bolsheviks saw revolution as something to be engineered, the Mensheviks awaited a spontaneous uprising precipitated by social forces beyond immediate control.

Although there were empiriocriticists in both factions, the philosophy was associated with Bolshevism, both because of Bogdanov's influence in the faction and because empiriocriticism's emphasis on activity suited the activism and voluntarism of Bolshevism. The association was an embarrassment to Lenin who distrusted empiriocriticism's eschewal of common-sense epistemology and its heady avant-gardism, manifested in Lunacharskii's attempt to supplement socialism with a secular religion celebrating humanity's powers of self-creation ('God-building'). In addition, the Mensheviks stressed that empiriocriticism was a Bolshevik philosophy in order to cast their opponents as revisionists and opportunists. In 1909, when Lenin no longer perceived an alliance with Bogdanov to be politically expedient, he broke a long-standing truce on philosophical

matters and attacked empiriocriticism as a heresy. Thereafter, dialectical materialism was deemed the theoretical foundation of Bolshevism.

In *Materializm i ėmpiriokrititsizm (Materialism and Empiriocriticism)* (1909), Lenin dismisses the idea that empiriocriticism transcends the dichotomy of materialism and idealism. On the contrary, it is a blatant form of idealism that represents matter as a construction of the mind. It therefore falls to the same arguments that defeated Berkeley and Hume. The new empiricism is motivated, Lenin argues, by the latest findings of physics, which suggest to the intellectually naïve that 'matter has disappeared' or that the laws of nature are human inventions. Such conclusions, Lenin maintains, are entirely wrongheaded. He defends a robust form of philosophical realism. The external world exists objectively prior to and independently of thinking subjects. We obtain knowledge through sense perception, on the basis of which we build theories of the world. Lenin wavers between direct and representative realism, suggesting sometimes that the mind is in immediate cognitive contact with reality and sometimes that perceptions and ideas are images or 'photographs' of objects. Either way, Lenin believes our conceptions may accurately represent reality. Objective truth is possible, though our theories are often only relatively, not absolutely, true. The criterion of truth is practice: a theory's truth explains why it is a reliable guide to practice.

Lenin denies that philosophy can dictate to science. Materialism establishes only that matter exists objectively, a view held by anyone who 'has not been an inmate of a lunatic asylum or a pupil of idealist philosophers'. Otherwise, it falls to science to establish what matter is like.

For the most part, Lenin's philosophy differs from Plekhanov's only on minor points (he rejects Plekhanov's view that sensations are 'signs' rather than depictions of objects) and in matters of emphasis (Lenin focuses on philosophy's relation to science rather than the philosophy of history). However, *Materialism and Empiriocriticism* reflects Lenin's Bolshevism in two respects. First, his obsession with objectivity fits his idea of the revolutionary vanguard armed with the truth that will lead the proletariat to victory. Second, his idea of the *partiinost'* ('partyness') of philosophy - the idea that philosophical positions are intrinsically related to class interests - politicizes all philosophical enquiry and suggests that philosophy is a weapon in the class war (see Partiinost').

The influence of *Materialism and Empiriocriticism* was enormous. Under Stalin, it was portrayed as a model of philosophical excellence and the foundation of a new, Leninist stage in Marxist thought (Lenin's debt to Plekhanov was forgotten). The concept of *partiinost'* was invoked to legitimize the criticism of philosophical theories on purely political grounds, and the coarse style of Lenin's writing, with its use of invective and abuse, became the official medium of philosophical discourse.

Materialism and Empiriocriticism was not Lenin's last word on philosophy. Between 1914 and 1916, he made a study of Hegel's philosophy. Lenin's notes, published 1929-30, show him arriving at a greater appreciation of the significance of dialectics. He declares that empiriocriticism was criticized more from the perspective of vulgar Feuerbachian materialism than Marxism, and suggests that the theorists of the 1910s (himself included) failed to understand Marx properly because of their ignorance of Hegel's *Science of Logic*. These and other aphorisms were later much quoted by Soviet philosophers endeavouring to elevate discussion beyond the crude materialism inspired by Lenin's early book.

3 Controversies in the 1920s

Although many areas of Soviet culture flourished after the Revolution, philosophy was beleaguered from the outset. First, the shortage of 'red specialists' meant that the 'philosophical leadership' charged with developing the Marxist ideas that had enabled and empowered the Revolution had to include non-Bolsheviks whom the Party viewed with suspicion. Moreover, when the first Soviet philosophy journal, *Pod znamenem marksizma (Under the Banner of Marxism)*, was launched in 1922, it was attacked on the grounds that the very idea of Marxist philosophy was confused. S.K. Minin argued that since philosophy had only ever served to foster ideological myths in the service of the oppressing classes, it should simply be displaced by science. 'Science to the bridge, philosophy overboard!' was Minin's slogan.

This 'liquidationism' was quickly (and unfairly) dismissed as a species of philistine anti-intellectualism. However,

the issue of philosophy's relation to science was soon the focus of a bitter controversy that divided the Soviet philosophical world: the debate between the 'mechanists' and the 'dialecticians' (or 'Deborinites').

The mechanists included scientists such as A.K. Timiriazev (1880-1955) and philosophers such as I.I. Skvortsov-Stepanov (1870-1928) and L.I. Aksel'rod (1868-1946). Bogdanov was associated with the movement, as was the Bolshevik luminary Nikolai Bukharin (1888-1938), whose *Teoriia istoricheskogo materializma (Historical Materialism)* had been published in 1921. Many mechanists were reductionists, who thought that science would eventually provide an exhaustive account of reality. But others, such as Aksel'rod, were drawn to mechanism by its defence of the autonomy of science. The mechanists argued that philosophy's role was to elucidate the concepts and laws employed by science, but not to interfere with scientific inquiry on a priori grounds.

The dialecticians included I.E. Sten (1899-1937), N.A. Karev (dates unknown), I.K. Luppol (1897-1943), V.F. Asmus (1894-1975) and the single most influential philosopher of the period, A.M. Deborin (1881-1963). Students of the history of philosophy, they saw their principal task as a materialist reworking of Hegelian dialectics. On this view, since the dialectical enterprise seeks to explain the very possibility of cognition, it cannot proceed by empirical generalization from scientific practice. Without philosophy, science cannot understand itself.

Between 1924 and 1929, the two schools fought in the literature and at gatherings in philosophical institutions. Despite the significance of the issues, neither side could move the other by argument and the debate degenerated into political wrangling. Eventually, the Deborinites used their greater institutional power to have their opponents condemned as revisionists.

The dialecticians' triumph was short-lived. In March 1930, a group of young activists at the Institute of Red Professors, led by M.B. Mitin (1901-87), accused their Deborinite teachers of 'formalism' and idealism. Since the mechanists had been (implausibly) portrayed as ideologists of Bukharin's 'right deviationism', Mitin (yet more implausibly) argued that the dialecticians were allied with the Trotskyite 'left deviation'. The young radicals called for a 'battle on two fronts in philosophy', urging that Soviet philosophy had to be 'Bolshevized'. Despite spirited resistance, Deborin and his followers were crushed. In June 1930, *Pravda* published an article by Mitin, Ral'tsevich and Iudin complete with editorial endorsement. Dubbed 'Menshevizing idealists' (Stalin's term), the Deborinites were forced from power.

Fortunately these controversies are not all that occurred in Russian Marxist philosophy in the 1920s and early 1930s. Two thinkers on the fringes of the subject made significant and enduring contributions. The first was the psychologist L.S. Vygotskii (1896-1934), who sought a way between mechanistic, reductionist models of the mind and introspectionist accounts that put the mental beyond systematic analysis. Inspired by Marx, Vygotskii argued that the dominant 'reflexological', stimulus-response models were too unidirectional to account for the human mind. The 'higher mental functions' are *mediated* forms of psychological activity; human beings actively change their environment to create new stimuli that serve to control their behaviour. We fashion artefacts - tools and signs - for manipulating both the world and the behaviour the world calls from us. Vygotskii originally introduced the concept of mediation to supplement the stimulus-response model. However, he became fascinated with the concept of meaning, arguing that semiotic mediation creates not a new class of stimuli, but a completely transformed psychological relation to reality. Vygotskii argued that since the 'mediational means' are cultural creations, preserved in the interpretative practices of the community, mind itself is a cultural artefact. The child's mind develops not through merely natural evolution, but by the 'internalization' of social practices. Vygotskii's writings were blacklisted shortly after his death in 1934, but his legacy was preserved by his followers - including A.R. Luria (1902-77) and A.N. Leont'ev (1904-79) - in the 'cultural-historical' school of psychology.

Similar themes emerge in writings attributed to V.N. Voloshinov (*c*.1895-1936) (which may have been written by M.M. Bakhtin (1895-1975)). In *Marksizm i filosofiia iazyka (Marxism and the Philosophy of Language)* (1929), Voloshinov argues that the problematic relation between base and superstructure can be illuminated by a Marxist theory of language. This is because superstructural phenomena are ideological in kind and ideology is a system of signs. Voloshinov rejects both psychologistic and abstract structuralist views of language, stressing that the linguistic sign exists in verbal utterance: meaning is inherently tied to use. This introduces a social dimension, since language is inherently dialogical. All utterance is explicitly or implicitly directed to another (or to oneself-as-another). This view has psychological consequences, for consciousness exists only in its material

embodiment of signs, and signs exist only in dialogue in a socio-cultural setting. The mind is semiotically constructed in social space, and the structure of thought (inner speech) issues from the internalization of the forms of outer utterance. Voloshinov writes: ‘Individual consciousness is not the architect of the ideological superstructure, but only a tenant lodging in the social edifice of ideological signs’.

Though suppressed in the USSR, Voloshinov’s writings influenced Roman Jakobson and the Prague Linguistic Circle, and, like Vygotskii, his ideas have attracted a significant following in the West (see Bakhtin, M.M. §§1-2).

4 The depths of the Stalin era

In the 1930s, all spheres of Soviet culture were brought to heel and made to glorify the Soviet Communist Party and the deeds of its General Secretary. The philosophers of the ‘new philosophical leadership’ eulogized Lenin’s genius (how fortunate that Stalin was ‘Lenin today’!) and spun a bizarre mythology about class enemies in philosophy, especially the Menshevizing idealists, now ‘discovered’ to be ‘fascists’ and ‘terrorists’.

In 1938, the tenets of Marxism-Leninism were codified in the *Istoriia kommunisticheskoi partii sovetskogo soiuza (Bol’sheviki). Kratkii kurs (History of the Communist Party of the Soviet Union (Bolsheviks). Short Course)*. The infamous fourth chapter, supposedly written by Stalin himself, presents the principles of dialectical materialism in schematic, canonical form, and offers strikingly simplistic examples of its application in historical materialism: that the base determines the superstructure follows from the primacy of the material over the ideal; class struggle reflects the dialectical law of ‘the unity and struggle of opposites’; revolutions are instances of the transformation of the quantitative into the qualitative, and so on. (The law of the ‘negation of the negation’ was omitted, no doubt because it implies the mutability of all things, including Soviet communism.) Both the sense of history, and the thirst for argument, evident respectively in Plekhanov’s and Lenin’s writings, are conspicuously absent. In this primitive form, Marxist philosophy was propagated throughout the Soviet Union. Every literate person was compelled to assimilate the ‘fourth chapter’, and professional philosophers expounded it *ad infinitum*, unable to transcend its horizons for fear of challenging Stalin’s authority.

Just as the Party put science in the service of industry, so the humanities were devoted to ‘the engineering of the human soul’, as Stalin described the role of Soviet literature. Philosophy was placed under the aegis of the Central Committee’s Department of Propaganda and Agitation, directed between 1939 and 1947 by the philosopher G.F. Aleksandrov (1908-61). In this climate, many genuine philosophers turned to the history of philosophy, where they could still study the European tradition. However, before long this area was also the focus of political controversy. In the war years, when ideological constraints were otherwise somewhat relaxed, Soviet works which gave a sympathetic rendition of German classical philosophy were denounced. For this reason, in 1944 the third volume of a collective *History of Philosophy*, which included Aleksandrov himself among its editors, was targeted by the Central Committee, even though its authors had previously received the Stalin Prize.

The post-war years saw a remarkable clampdown throughout Soviet culture. This campaign, led by A.A. Zhdanov, combined fierce allegiance to Marxist-Leninist principles with a virulent form of Russian nationalism. The chauvinistic climate is well illustrated by another controversy surrounding Aleksandrov. In 1947, his textbook *Istoriia zapadnoevropeiskoi filosofii (A History of Western European Philosophy)* (1946) (which is entirely orthodox in spirit) was vilified as ‘abstract’, ‘bookish’, ‘formal’ and ‘lacking *partiinost’*’ largely for its failure to emphasize the supremacy of Russian over Western philosophy. The philosophers who opposed this trend were eventually accused of ‘cosmopolitanism’, ‘sycophancy before bourgeois philosophy’ and ‘kowtowing to the West’.

The Party also used philosophy to interfere in the natural sciences. In the late 1940s, theories in chemistry, cosmology and physics, especially relativity theory and quantum mechanics, were all criticized on ideological grounds when apparently in tension with dialectical materialism. And in the early 1950s, cybernetics was dismissed as a pseudo-science, since its global explanatory pretensions were at odds with Marxism. Chauvinism about Russian scientific achievements reached ridiculous heights, and the myth of a ‘proletarian science’ took hold. The most conspicuous example was the Party’s long-standing support for T.D. Lysenko’s Michurinist genetics, which defended the inheritance of acquired characteristics (suggesting that the qualities of New Soviet Man could be genetically inherited by subsequent generations). These tragicomic episodes did incalculable damage to Soviet science.

5 Renaissance after Stalin

That any kind of philosophical culture survived the Stalin period is a remarkable fact. A significant number of philosophers perished in the purges, but a handful of participants in the debates of the 1920s survived and continued to teach. Although unable to publish much in this period, philosophers like Asmus, O.V. Trakhtenberg (1889-1959), P.S. Popov (1892-1964) and M.A. Dynnik (1896-1971) gave significant courses at Moscow University. B.M. Kedrov (1903-85) was also influential, despite being then estranged from the University for political reasons. In addition, several prominent psychologists, such as Luria, Leont'ev and S.L. Rubinshtein (1889-1960), who were members of the philosophy faculty at Moscow University, made a vital contribution to keeping philosophy alive. These older thinkers found a new generation of philosophers among students resuming their education after the Great Patriotic War. This generation, which includes E.V. Il'enkov (1924-79) and A.A. Zinov'ev (1922-) (who later emigrated, writing the dissident novels, *The Yawning Heights* and *The Radiant Future*), is sometimes called the 'iflitsy', after the unusually vibrant Moscow Institute of History, Philosophy and Literature (IFLI), where many of them had studied before the war. Despite the stifling atmosphere in most academic institutions, and the shortage of materials, the new generation was determined to renew Soviet philosophy by going 'back to Marx'. In the 1950s, their work focused on two principal areas. First, they explored the dialectical method, sketched by Marx in his 'Introduction to a Critique of Political Economy' and deployed in *Capital*. This produced scholarly readings of Marx's texts in which the 'method of ascent from the abstract to the concrete' was developed into a full-blown account of concept acquisition and theory construction, to be contrasted with empiricist views of scientific knowledge. Most significant is Il'enkov's *Dialektika abstraktnogo i konkretnogo v 'Kapitale' Marksa (The Dialectics of the Abstract and the Concrete in Marx's 'Capital')* (1960), but the writings of Zinov'ev, B.A. Grushin (1929-) and G.P. Shchedrovitskii (1929-93) (who formed the Moscow methodological circle) were also influential.

These methodological writings explore the preconditions of scientific understanding, not in abstract Kantian terms, but historically. This project naturally raised the question of the nature and origin of mind itself. How can a material world come to contain beings able to engage in the project of understanding themselves and their world? The question was answered by developing a philosophical anthropology centred on the concept of *activity*. In the 1960s, the Georgian philosopher Merab Mamardashvili (1930-90), who worked in Moscow, produced an influential study of the concept of consciousness in Marx's thought, developing the concept of 'transmuted forms'. And Il'enkov developed his controversial account of 'the ideal', according to which humanity's active engagement with nature results in the objectification of non-material properties and relations in the natural world. We enculturate our material environment, and become conscious beings in so far as we appropriate social activities constitutive of humanity's spiritual culture. In this way, Il'enkov develops the Vygotskian idea that culture is a precondition of the existence of individual consciousness.

The work of this new generation, mostly published in the Khrushchev 'thaw', stimulated lively debate in many philosophical institutions, especially Moscow's Institute of Philosophy. At this time, a translation of Marx's *Paris Manuscripts* (prepared by Il'enkov among others) became available, provoking further interest in a critical, anthropocentric Marxist philosophy. However, many of the Stalinist old guard maintained their positions in the academy, and the Party continued to exercise control over philosophical discussion. Huge editions of one-dimensional Marxist-Leninist textbooks, such as the *Fundamentals of Marxist-Leninist Philosophy*, were used to disseminate the worldview of the Communist Party in all institutions of higher education. Contact with Western philosophers and their works was limited. Moreover, by the mid-1960s, the thaw began to ice over and the mood of creativity and optimism faded.

Although the next two decades lacked the inspiration of the early 1960s, there were many prominent controversies, often conducted on the pages of *Voprosy filosofii (Questions of Philosophy)*, which had been the principal Russian philosophy journal since 1947. Some, such as the discussions of reflection theory and the laws of dialectics, were merely variations on themes of Marxist orthodoxy. Others, like the ongoing debates about the nature of Marxist philosophy and the relation between dialectical and formal logic, were more serious. There was a revival of interest in ethics, which had previously been ruled out by the orthodox view that morality was destined to 'wither away' under communism. Influential figures were O.G. Drobnitskii (1933-73), tragically killed in an air disaster, and Iu.N. Davydov (1929-). Knowledge of contemporary Western philosophy became more widespread, due to the

efforts of such thinkers as A.S. Bogomolov (1927-93) and A.F. Zotov (1931-), though works developing Soviet Marxism in critical dialogue with Western philosophy were few; the most successful was *Sub' 'ekt, ob' 'ekt, poznanie (Subject, Object, Cognition)* (1980) by V.A. Lektorskii (1932-), which ably contrasts an activity-centred epistemology with a variety of Western views.

One significant debate conducted in both academic and popular journals concerned the extent to which human psychological capacities are genetically determined. Many Soviet philosophers and psychologists, such as Il'enkov, Leont'ev, V.V. Davydov (1930-) and F.T. Mikhailov (1930-) argued that Marxism is committed to a conception of individuals as socially constituted beings, whose principal mode of psychological inheritance is cultural rather than biological. In contrast, their nativist opponents, who included scientists as well as such philosophers as D.I. Dubrovskii (1929-), argued that this view contradicted the physical basis of the mental. Though the theoretical value of these polemical exchanges is limited, the political dimensions of the debate are interesting. The nativists resented philosophical interference in what they took to be an empirical matter. In addition, they took their opponents' contructivism to be motivated by dreams of the creation of a 'New Soviet Man', and they turned to genetics to define a sphere where the individual was immune to social influence. The constructivists in turn saw nativism as reactionary, the result of a dehumanizing scientism, inspired by fascination with the so-called 'scientific-technological revolution'. Human beings are not just self-organizing machines to be manipulated by the state, however much they were made so by Stalinism, but beings infinitely capable of self-transcendence and renewal if society can only create the appropriate conditions.

6 *Perestroika* and dissolution

Soon after coming to power in 1985, Gorbachev instigated reforms designed to restructure the Soviet socioeconomic order (*perestroika*) and to encourage more critical debate and public accountability (*glasnost'*). To this end, he disavowed the Party's claim to privileged access to objective knowledge and, calling for greater democracy, urged an end to the 'bureaucratization' of Soviet life. Soviet philosophy underwent significant liberalization. Instruction in Marxism-Leninism became no longer mandatory in Soviet institutions of higher education; indeed, the term 'Marxism-Leninism' was soon abandoned. *Voprosy filosofii* published debates about democracy, the rule of law and civil society, discussions of the history of Soviet philosophy (never previously a possible subject of debate), articles by Western thinkers, and hitherto banned works by Russian philosophers, such as Berdiaev and V.S. Solov'ëv (1853-1900).

The ethos of 'official' Soviet philosophy under *perestroika* is illustrated by the writings of I.T. Frolov (1929-), then advisor to Gorbachev and editor-in-chief of *Pravda*. In his book *Man, Science, Humanism* (1986), Frolov speaks out against Soviet Prometheanism, arguing that although we look to science to improve the conditions of human life, we must recognize that scientific progress has brought dramatic threats to the future of humanity. A critical and humanistic Marxism strives to put science 'in the service of humanity'. Frolov stresses the 'open character' of Marxist theory, urging that its tenets are not absolute truths. He attacks the follies of proletarian science, suggests compromise solutions to entrenched Soviet controversies (for instance over the genetic basis of the mental), and urges dialogue with Western thinkers.

Though more liberal than previous party lines, Frolov's position retains the idea that all issues are best addressed from the perspective of an all-encompassing theory of humanity's place in nature, a theory whose fundamental tenets are beyond dispute. Frolov's writings represent a broadening of Soviet orthodoxy, but the spirit of radical criticism remains absent.

The theme of humanism was central to much philosophical discussion under *perestroika*. The concept of a person (*lichnost'*) as a centre of moral agency came to the centre of discussion, complementing debates about ethical responsibility and legal consciousness. This person-centred approach was reflected in a new official textbook, *Vvedenie v filosofiiu (Introduction to Philosophy)* (1990), focused on problems of human existence instead of the primacy of materialism and the laws of the dialectic. There was a renewed interest in the anthropocentric Marxism that flourished in the Khrushchev thaw, especially Il'enkov's philosophy, which, it was argued, anticipated many themes openly discussed under *glasnost'*.

The new climate precipitated a number of heartfelt public discussions on the question 'Is Marxism dead?' A number of significant philosophers, such as E.Iu. Solov'ëv (1934-), K.M. Kantor (1922-), V.M. Mezhuev (1934-)

and V.I. Tolstykh (1929-) argued that Marx should not be quickly blamed for the failed Soviet experiment; Marxist thought deserves continued study like any other major philosophical system. In contrast, others, like A.S. Tsipko (1941-), poured scorn on the pretensions of Marxist philosophy, stressing the emptiness of Marx's views and their inapplicability to contemporary reality. These discussions reveal an intellectual world in crisis, disoriented by the diminution of its central authorities.

With the collapse of the Soviet Union in 1991, Russia's commitment to Marxism was abandoned entirely. Although Marxist influences are often evident in their work, few Russian philosophers now see themselves as contributing to the Marxist tradition. The thirst for a grand explanatory perspective is no less intense, but Russian thinkers now seek one elsewhere, in the Russian religious philosophy suppressed by the Bolsheviks, or in Western ideas, such as Foucauldian postmodernism (taken up by V.A. Podoroga (1946-) and his associates). Where these developments will lead is an open question. It seems, however, that the history of Russian Marxism is over.

See also: Marxism, Chinese; Marxism, Western; Marxist philosophy of science §3; Marxist thought in Latin America

DAVID BAKHURST

References and further reading

Aleksandrov, G.F.; Bykhovskii, B.E.; Mitin, M.B. and Iudin, P.F. (eds) (1941-3) *Istoriia filosofii (History of Philosophy)*, Moscow: Gospolitizdat, 3 vols.(A collective history of philosophy, written by B.E. Bykhovskii, V.F. Asmus and B.S. Chernyshev among others. Although the book had earlier won the Stalin prize, the third volume was denounced by the Central Committee for its sympathetic treatment of Western philosophy, and plans for a further three were abandoned.)

Aleksandrov, G.F.; Bykhovskii, B.E.; Mitin, M.B. and Iudin, P.F. (1946) *Istoriia zapadnoevropeiskoi filosofii*, Moscow; *A History of Western European Philosophy*, New York, 1949.(A typical textbook history of philosophy by a leading philosopher of the Stalin era. Despite its orthodoxy it was severely criticized in Zhdanov's campaign against 'cosmopolitanism' in 1947.)

Bakhurst, D. (1991) *Consciousness and Revolution in Soviet Philosophy: From the Bolsheviks to Evald Ilyenkov*, Cambridge: Cambridge University Press.(An account of the philosophical culture of the USSR focused on the work of E.V. Il'enkov.)

Bogdanov, A.A. (1905-6) *Ėmpiriomonizm*, Moscow: Dorovatorskii & Charushnikov (Delo), 3 vols.(Bogdanov's principal philosophical treatise; not translated into English.)

Bukharin, N.I. (1921) *Teoriia istoricheskogo materializma: populiarnyi uchebnik marksistkoi sotsiologii*, Moscow; 3rd edn trans. *Historical Materialism: A System of Sociology*, New York: Russel & Russel Inc., 1925.

Davydov, Iu.N. (1982) *Ėtika liubvi i metafizika svoevoliia (The Ethics of Love and the Metaphysics of Self-Will)*, Moscow: Molodaia gvardiia.(A significant and controversial study of moral philosophy in the existentialist tradition.)

Drobnitskii, O.G. (1974) *Poniatie morali (The Concept of Morals)*, Moscow: Nauka.(The principal work of one of the most sophisticated moral philosophers of the Soviet period.)

Frolov, I.T. (1986) *Man, Science, Humanism: A New Synthesis*, Moscow: Progress; repr. New York: Promethius Books, 1990.(A forthright statement of humanistic Marxism by an influential figure in the Soviet establishment under Gorbachev.)

Guseinov, A.A., and Tolstykh, V.I. (eds) (1993) *Studies in East European Thought (Special Issue): Marxism and the Socialist Idea in Russia Today* 45, nos 1-2.(A transcription of the 'Is Marxism Dead?' debate at the Svobodnaia slovo (Free Word) club in Moscow in 1990.)

Il'enkov, E.V. (1960) *Dialektika abstraktnogo i konkretnogo v 'Kapitale' Marksa*, Moscow: Academiia nauk; trans. S. Syrovatkin, *The Dialectics of the Abstract and the Concrete in Marx's 'Capital'*, Moscow: Progress, 1982.(An influential attempt to develop Marx's dialectical method into a comprehensive anti-empiricist theory of knowledge.)

Il'enkov, E.V. (1974) *Dialekticheskaia logika*, Moscow: Politizdat; trans. H. Campbell-Creighton, *Dialectical Logic*, Moscow: Progress, 1977.(Il'enkov's materialist phenomenology of spirit, which includes, as ch. 8, his account of 'the ideal' so influential in the early 1960s.)

Istoriia kommunisticheskoi partii sovetskogo soiuza (Bol'sheviki). Kratkii kurs (1938), Moscow: Gospolitizdat; *History of the Communist Party of the Soviet Union (Bolsheviks). Short Course*, Moscow: Foreign Language Publishing House, 1943.(The main ideological primer of the Stalin era.)

Kelly, A. (1981) 'Empiriocriticism: A Bolshevik Philosophy?', in *Cahiers du monde russe et soviétique* 22:

89-118.(An excellent study of Russian empiriocriticism and its political significance.)
Kolakowski, L. (1978) *Main Currents of Marxism*, trans P.S. Falla, Oxford: Oxford University Press, 3 vols.(A brilliant and comprehensive study of Marxist thought, containing, in vols 2 and 3, detailed discussion of Russian and Soviet Marxism.)
Lektorskii, V.A. (1980) *Sub''ekt, ob''ekt, poznanie*, Moscow: Nauka; trans. S. Syrovatnik, *Subject, Object, Cognition*, Moscow: Progress, 1984.(An insightful discussion of Marxist epistemology in relation to prominent, contemporary Western theories.)
Lenin, V.I. (1895-1916) *Filosofskie tetradi*, in *Polnoe sobranie sochinenii*, Moscow: Gospolitizdat, 1958-69, 5th edn, vol. 29; trans. C. Dutt, *Philosophical Notebooks*, in *Collected Works*, vol. 38, Moscow: Progress and London: Lawrence & Wishart, 1968-9, 4th English edn.(A collection of notes on philosophical texts, the most significant being those taken in the 1910s on various writings of Hegel's.)
Lenin, V.I. (1902) *Chto delat'*, in *Polnoe sobranie sochinenii*, Moscow: Gospolitizdat, 1958-69, 5th edn, vol. 5; trans. J. Fineberg, *What Is To Be Done?*, in *Collected Works*, Moscow: Progress and London: Lawrence & Wishart, 4th English edn, vol. 5.(Lenin's classic statement of his vision of the revolutionary party and its tactics.)
Lenin, V.I. (1909) *Materializm i ėmpiriokrititsizm*, in *Polnoe sobranie sochinenii*, Moscow: Gospolitizdat, 1958-69, 5th edn, vol. 18; *Materialism and Empiriocriticism*, in *Collected Works*, Moscow: Progress and London: Lawrence & Wishart, 4th English edn, vol. 14.(Lenin's defence of dialectical materialism against positivism and other supposed idealist schools; established under Stalin as the central text in Soviet philosophy.)
Mamardashvili, M.K. (1990) *Kak ia ponimaiu filosofiiu (How I Understand Philosophy)*, Moscow: Progress.(An interesting collection of writings and interviews, including Mamardashvili's 'The Analysis of Consciousness in Marx's Works' (the latter is translated in *Studies in Soviet Thought* 32: 101-20 (1986).)
Mikhailov, F.T. (1964) *Zagadka chelovecheskogo 'ia'*, Moscow: Politizdat, 2nd edn 1976; 2nd edn trans. R. Daglish, *The Riddle of the Self*, Moscow: Progress, 1980.(A lively and insightful presentation of a critical Marxist epistemology; conveys the spirit of the works of the Khrushchev thaw.)
O raznoglasii na filosofskom fronte (On a Disagreement on the Philosophical Front) (1930), *Vestnik kommunisticheskoi akademii*, vols 40-1 (12-165), vol. 42 (20-89).(A fascinating transcription of meetings at the Communist Academy where Deborin and his followers were denounced.)
Oznovy marksistko-leninskoi filosofii (1959), Moscow: Politzdat; 5th edn (1980) trans. R. Daglish, *Fundamentals of Marxist-Leninist Philosophy*, Moscow: Progress, 1982.(The standard textbook of official philosophy in the post-Stalin era, written by a team of scholars, including A.S. Bogomolov and T.I. Oizerman, under the direction of F.V. Konstantinov.)
Plekhanov, G.V. (1895) *K voprosu o razvitii monisticheskogo vzgliada na istoriiu*, St Petersburg: Skorokhodova; trans. A. Rothstein, *In Defence of Materialism: The Development of the Monist View of History*, London: Lawrence & Wishart, 1947.(The first comprehensive presentation of Marxist philosophy as dialectical materialism.)
Scanlan, J.P. (1985) *Marxism in the USSR: A Critical Survey of Current Soviet Thought*, Ithaca, NY: Cornell University Press.(A comprehensive survey of Soviet debates in all areas of philosophy in the post-Stalin era and before *perestroika*.)
Swiderski, E. (1993) 'The Crisis of Continuity in Post-Soviet Russian Philosophy', in B. Smith (ed.) *Philosophy and Political Change in Eastern Europe*, La Salle, IL: Monist Library of Philosophy.(A study of the climate in Russian philosophy immediately after the collapse of the USSR.)
Voloshinov, V.N. (1929) *Marksizm i filosofiia iazyka*, Leningrad: Ranion, Nauchno-issledovatel'skii institut sravnitel'noi istorii literatur i iazykov zapada i vostoka; trans. L. Matejka and I.R. Titunik, *Marxism and the Philosophy of Language*, New York: Seminar Press, 1973.(The systematic attempt at a Marxist theory of language; possibly written by M.M.Bakhtin.)
Vvedenie v filosofiiu (Introduction to Philosophy) (1990), Moscow: Politizdat, 2 vols.(A textbook produced in the Gorbachev period by a large team of philosophers, including I.T. Frolov, V.A. Lektorskii and V.Zh. Helle; designed to replace traditional Marxist-Leninist views with a humanistic vision centred on 'the problem of man's place in the world'.)
Vygotskii, L.S. (1934) *Myshlenie i rech'*, Moscow: Gosudarstvennoe sotsial'no-ekonomicheskoe izdatel'stvo; abridged trans. E. Hanfmann and G. Vakar, *Thought and Language*, Cambridge, MA: MIT, 1962; 2nd edn trans. revised and edited by A. Kozulin, Cambridge, MA: MIT University Press, 1986.(Vygotskii's most mature statement of his philosophical psychology; blacklisted during the Stalin era, the text was highly influential in Russia and abroad.)

Marxist thought in Latin America

Marxism is a theory offering a critique of capitalist political economy. Marxism also views itself as an instrument or means of changing the world from a capitalist to a socialist (and/or communist), economic and political order. Given its interest in economic and political change, Marxism involves a philosophy of history which depicts the possibility of and conditions for change from a capitalist to a socialist order. Marxist intellectuals perform the dual task of analysing the failures or limitations of capitalist economic and political structures. The theory also proposes and evaluates socialist alternatives.

Latin American Marxism developed out of its own historical, economic, political and cultural conditions. Influenced by Lenin's analysis of imperialism as the highest stage of capitalism, it directed the critique of capitalist political economy towards the capitalist world market and its disadvantageous effects for the countries, particularly the impoverished classes and social sectors, of the Latin American and Caribbean regions. Latin American Marxism-Leninism argues, on political and economic grounds, that national liberation cannot be achieved without liberation from imperialism.

Marxists believe that although the protagonists of history's political projects are the workers (or if Leninist, the workers together with the peasants), in the end the interests of these groups represent the universal interests of humankind. Marxist political discourse often uses broader categories than those of 'workers' or 'peasants' to designate the agents of political emancipation, employing terms such as 'the people', 'the popular sectors' or 'the revolutionary masses'. In this way Marxism attempts to broaden its political base so as to make its goals more effective. The political discourse of the Cuban Revolution of 1959 and the Nicaraguan Sandinista Revolution of 1979 exemplify this practice.

There are and have been many differences among Marxists because of the different approaches to criticizing capitalism as well as the different conceptions held by those who profess a commitment to the ultimate Marxist goal of creating a nonexploitative socialist society. Representative issues in Latin American Marxism may be illustrated by focusing on three questions: the problem of orthodoxy, the socialist construction of a national identity and socialism's relation to ethics, religion and culture. In addressing these issues, this entry draws significantly from the work of Peruvian Marxist José Carlos Mariátegui, a prominent founder of Latin American Marxism.

1 The problem of orthodoxy

The mechanisms for implementing Marxist orthodoxy were not clear when the ideology was first introduced and then began to take shape between the 1870s and 1920s. The newness of socialist and communist ideas led Marxists to view themselves as future-oriented and revolutionary thinkers. Among the best-known early Latin American Marxists were the Argentine Aníbal Ponce, the Peruvian José Carlos Mariátegui and the Cuban Julio Antonio Mella. Mella was a university student leader strongly influenced by the thought of the Cuban patriot José Martí. Ponce was involved in education and publishing, wrote on socialist humanism and from 1923-5 coedited, with the Argentine positivist philosopher José Ingenieros, the *Revista de Filosofía, cultura, ciencias, educación (Journal of Philosophy, Culture, Science, Education)* (1915-29). Ponce became sole editor after Ingenieros's death in 1925.

Mariátegui, a journalist and social critic, was acknowledged as the most influential of the early Latin American Marxist thinkers. Author of *Seven Interpretive Essays on Peruvian Reality* (1928) and editor of the journal *Amauta: Revista mensuel de doctrina, literatura, arte, polémica (Amauta: Monthly Review of Doctrine, Literature, Art and Polemics)* (1926-30), he developed a highly original position in Latin American Marxism by stressing the importance of ethnicity, or indigenous culture, in the construction of national identity. He did so by maintaining a flexible position with regard to the relationship of Marxism to the latest theoretical currents from Europe and the USA, namely Bergsonian philosophy, pragmatism and psychoanalysis (see Bergson, H.-L.; Pragmatism). He also rejected the notion of an essential antagonism between Marxism and religious thought, thereby breaking with the orthodox Marxist view that a materialist philosophy of history necessarily presupposes a materialist metaphysics. Mariátegui expounded his original ideas as part of his commitment to both a Leninist Marxism and the project of building a socialist society in Peru. As founder of the Socialist Party of Peru, he was criticized during his short life for some of his views by the Communist International. The 'errors' in his ideas were especially criticized during

the Stalinist era, until a resurgence of interest in his work took place in the 1960s and 1970s.

For the first generation of Marxists, orthodoxy was less of a problem than for succeeding generations, since their task was to forge the new political philosophy which was not without codification. Marxism's link between theory and practice inevitably has raised the problem of orthodoxy and heterodoxy. Marxism's role is explanatory and predictive as well as conceptual. It is a theory whose goal is to create a determinate set of political effects. Its difficult aim is to construct a society without class exploitation.

Of fundamental historical importance to Marxism have been the questions of the scientific correctness of the theory designed to change the world and the correct political orientation of the leaders charged with developing and applying the theory in historical conditions, subsequent to those experienced by Marx and Engels. Depending on one's critical approach, Marxist theory can range from a nonpolitical intellectual analysis and critique of various aspects of capitalism as they affect human life and society to a dogmatic exposition and defence of the main theses accepted collectively at Communist Party congresses.

Orthodoxy is a relative concept. A position is orthodox in relation to the dogmas, truths, or methods that are taken to be central to a particular doctrine or school of thought. A Marxist thinker may be a dissident with respect to one school of Marxism while representing an orthodox position in another. The degree to which a Marxist position is free from orthodoxy is determined by the extent of its ability to question the closest Marxist authority with which it is associated. Marxist politics tend to impose a collective discipline over an individual's analysis of social reality. In cases where Marxist philosophy is practised within a socialist state that permits only one political party, the collective constraints over individual thought can be highly exacting. Although not all Marxists defend a one-party political system, those who do assert that the defence of the emerging socialist state against imperialism must take the highest priority.

Traditionally, Marxism relies on the methodologies of dialectical and historical materialism. From Hegel's self-validating account of dialectical thinking, Marxists inherited the view that dialectics provide an insuperable scientific conception of reality (see Hegel, G.W.F.). Using the logic of dialectical materialism, orthodox Marxism holds that reality is material and that it changes by oppositional as well as by qualitative movements governed by dialectical laws known to the human mind. Marxist orthodoxy gave the name of 'scientific socialism' to this perspective. Applying a dialectical and material concept of change to history, historical materialism posits that the development of history is caused by material (primarily economic) factors, that history progresses through stages and that the change from one historical stage to another takes place when internal contradictions lead the old economic structures to be phased out in favour of newer, stronger and more universal elements. Historical materialism postulates that at the point of qualitative change between capitalism and socialism there lies a social revolution that will dismantle the old class divisions within society and lead eventually to a classless society in which the exploitation of human beings by human beings ceases to exist.

As a political philosophy Marxism argued that such exploitation can only be overcome by a social and political revolution in which the working class plays the dominant and leading liberatory role. After Lenin, Marxism-Leninism broadened the class of revolutionary political subjects to include both workers and peasants, students and intellectuals, all of whom had the same revolutionary aspirations. The Italian Marxist Antonio Gramsci (1891-1937) called the revolutionary intellectuals who supported the social revolution 'organic intellectuals'. In Latin America this view influenced many leftist intellectuals to back popular revolutionary movements led by Marxist leaders, as in Cuba and Nicaragua. The Marxist-Leninist-Gramscian influence, augmented in the 1960s by the political impact of leaders such as Ernesto 'Ché' Guevara, distinguished a wide sector of Latin American Marxism from Western European critical Marxism, whose intellectual roots were closer to Hegel, Marx, phenomenology, existentialism and critical theory (see Critical theory; Existentialism; Phenomenology in Latin America). In Western Marxism, the materialist view of the worker and peasant protagonists of history is downplayed while issues such as alienation, the individual and social justice and the combined effects of oppressive class, race, and gender relations in the critique of capitalist economics are foregrounded.

Despite the problem of orthodoxy which delimits the production of Marxist philosophy and social thought, as a critique of capitalism and capitalist social structures Marxism yields some important and original insights which are absent from bourgeois philosophy and ideology. Mariátegui's focus on the construction of an Indohispanic

socialism in Peru is a case in point.

2 The construction of a national identity

In keeping with the theory of historical materialism, Mariátegui (1928) held that societies evolve according to the laws of their socioeconomic development and that historical progress involves the surpassing of feudalism by capitalism and of capitalism by socialism. Apart from this traditional framework of historical materialism, Mariátegui introduced a number of variants into his concept of socialism which, when considered together, lay the foundations for a specifically Latin American approach to socialist theory. Mariátegui questioned the linear concept of history as it applied to the narrative of the superiority of European culture over the indigenous cultures of the Americas. He took the Spanish conquest of the Inca civilization to represent the defeat of a highly competent form of social organization by a less competent one. He claimed the conquest symbolized a rupture within the economic organization which grew out of the experiences of the earliest Peruvian people. These people combined a simple life with a highly sophisticated economic system which was based on hard work, discipline and the satisfaction of the people's material needs. By referring to the pre-Columbian Inca empire as a central, although defeated, player in the history of Peru, Mariátegui demonstrated that Peru's national identity should not be conceived without giving a prominent place to its indigenous people. As a Marxist his goal was to contribute to the formation of an Indohispanic socialism in Peru. This position involved, on the one hand, convincing non-Marxists that socialism was a superior system to capitalism in Peru, while on the other hand, convincing the white and mestizo minorities (including socialists) that the disenfranchised indigenous population of Peru was an essential and irreplaceable part of its national identity and character.

As a member of a generation of Latin Americans critical of positivist philosophy, Mariátegui combined a political interest in Marxism with other artistic and intellectual interests (see Anti-positivist thought in Latin America; Positivist thought in Latin America). He promoted the Peruvian pro-indigenous literary movement of *indigenismo* (indigenism) and the new thinking of the European avant-garde. He supported the merit of thinkers like Friedrich Nietzsche, Sigmund Freud, James Joyce and Miguel de Unamuno and linked aspects of their views and teachings to his socialist interpretation of human and social reality. Mariátegui's appreciation of the value of myth led him to the insight that the relationship between the Indian and the land could not be understood simply in materialist or strictly modern secular terms. It must include a full account of the Indian's deeply-rooted spiritual belief that life springs from the land and returns to the land. He perceived the indigenous peasants as socially oppressed both as a class and in terms of their racial and ethnic status. He argued for a notion of national identity that would allow the Indians to come to prominence. Mariátegui's analysis called for Marxist theory to be reformulated so that it would take into account both class and ethnicity, including the cultural aspects of ethnic beliefs.

3 Ethics, religion and culture

The concept of human nature articulated by Mariátegui included a blend of existential, pragmatist, and Marxist influences (see Existentialism; Pragmatism). It was best characterized by a rejection of Cartesian rationalism and positivism (see Rationalism). He followed the Italian idealist philosopher Benedetto Croce in holding that, without a set of moral principles, specifically a principle condemning human exploitation, Marx's critique of capitalist economy would not make sense (albeit that Marx was not interested in developing a moral philosophy). Influenced by the French labour theorist Georges Sorel, Mariátegui argued that socialism had an ethical function, which was to create a 'morality of producers' in the process of struggling against capitalism. Therefore, Mariátegui was interested in articulating a work ethic based on socialist principles which would raise the moral consciousness of workers as they aimed to transform the exploitative conditions of class society.

Mariátegui departed from a 'scientific' dialectical materialist account of social change by representing the proletariat, or working class, as an affirmative rather than a negative, social and political force. In his early writings Mariátegui argued that the characteristic gesture of the bourgeoisie was to negate, while that of the working class was to affirm. He counterposed bourgeois nihilism and decadence with working class optimism and confidence. In the Hegelian version of the dialectic, negation (rather than affirmation) was the process which moved history forward. This led Marxism to emphasize revolution as the negation of a previously established political order. However, in his account of Peruvian history, Mariátegui argued for socialism on positive grounds. Peruvian capitalism was described as weak and unable to undo the power of the rural conservative sectors, which retarded economic development. Socialism was depicted as the one solution that would incorporate the benefits of

capitalism (creativity, discipline, productiveness) while accommodating the communitarian interests of the indigenous population.

In the light of Mariátegui's portrayal of the workers and indigenous peasants as positive social forces, it could be argued that his concept of ethics, with its connection to socialism, was as close to Nietzsche's idea of a superior morality as to Marx's concept of social revolution. Mariátegui saw the advent of socialism as resulting from its own vigorous and undisputed success, not as a result of a class war. The work ethic he advocated contained a forceful and explicit rejection of what Nietzsche called a 'slave morality', a moral system which was reactive rather than self-initiated in its positing of moral principles and/or practices (see Nietzsche, F. §8). Mariátegui also rejected the notion of a teleological end-state after the achievement of which history would end and all oppression would cease. He declared that no revolution can foretell a subsequent revolution, despite the fact that seeds of political change may have been planted. Like William James, Mariátegui attributed to human beings a basic 'will to believe'. However, he also held, after Sorel, that the object of the will to believe need not be a religion as traditionally understood: it could be a belief in social revolution. Such a belief, like religion, would satisfy the human hope for a better world. Mariátegui therefore subscribed to an open-ended concept of revolution, which favoured an adaptation of Marxist ideas to new historical circumstances, including new trends in theory beyond those with which Marx was acquainted, such as Hegelian philosophy.

Mariátegui's views have been of interest to religious leftists, particularly Christians, motivated by a belief in a just and nonexploitative society, such as those committed to a theology of liberation (see Liberation theology). In the *Seven Interpretive Essays on Peruvian Reality* (1928) he claims that it is a fundamental error for Marxists to attack the clergy and the church as if these were the principal obstacles to socialist change when the real enemy is the socioeconomic structure of society. He argued that the will to believe is a basic factor of human existence and that historical materialism as a theory of historical development should not be confused with philosophical materialism as a comprehensive theory of all reality. This view, conjoined with the view that, according to myth Indians regard the land as their common mother, contributed to Mariátegui's most important theses that the exploitation of the Indians was due to the land tenure system of Peruvian society and that their liberation could only take place through a socialist structure respectful of the Indians' ancient relationship to the land and the inherited collective practices of working it. His position suggested that the acceptance of modern Western standards of development should be tempered by respect for the forms of communal organization and bonding which characterize the lives of Peruvians whose cultural legacy is of pre-Columbian (non-Western) indigenous origin.

Mariátegui united his defence of an Indohispanic Peruvian socialism with a strong Marxist anti-imperialist statement. He argued that as long as imperialism exists, a Latin American society cannot be nationalist unless it is socialist. The economic and political structure of imperialism prevents the full realization of nationality in countries whose economic development is locked into a weak and backward capitalist structure controlled by foreign economic interests. He noted that imperialism implies racism in that the cultural values imposed by an imperialist north-over-south continental order were the values of a white bourgeois class, imported into the south by the white privileged classes of Latin America. The latter, he claimed, failed to question their own involvement in an exploitative and racist system. Mariátegui challenged capitalism in terms that emphasized class, race and a nation's dependent status on the world market without promoting divisiveness or separatism.

Although Latin America has produced some well-known Marxist academic philosophers, including the Heideggerian Marxist Carlos Astrada (1894-1970) and the Marxist aesthetician Adolfo Sánchez Vázquez (1915-), it is José Carlos Mariátegui, a self-educated working class mestizo intellectual, to whom most contemporary scholars turn when the subject of inquiry involves the articulation of a specifically regional, that is Latin American, Marxism.

See also: Chinese Marxism; Liberation philosophy; Marx, K.; Marxism, Western; Marxist philosophy, Russian and Soviet

OFELIA SCHUTTE

References and further reading

Amauta: Revista mensuel de doctrina, literatura, arte, polémica (Amauta: Monthly Review of Doctrine, Literature, Art and Polemics) (1926-30), Lima.(The term *Amauta* is Quechua for 'wise teacher'. Mariátegui's journal took the name to stress the Inca heritage of Peru and to give the term a new vanguardist meaning.)

Astrada, C. (1963) *Existencialismo y crisis de la filosofía (Existentialism and the Crisis of Philosophy)*, Buenos Aires: Devenir.(An account of the status of Western philosophy at mid-century from an existential Marxist standpoint. Advanced reading knowledge of Spanish required.)

Bottomore, T. (ed.) (1983) *A Dictionary of Marxist Thought*, Cambridge, MA: Harvard University Press.(A helpful and substantive philosophical dictionary of Marxist authors and concepts, focusing primarily on Western Marxism.)

Fernández Retamar, R. (1989) *Caliban and Other Essays*, trans. E. Baker, Minneapolis, MN: University of Minnesota Press.(A Cuban theorist's influential interpretation of the figure of Caliban as a symbol of resistance against colonialism.)

Guadarrama González, P. (1994) *América Latina: Marxismo y Postmodernidad (Latin America, Marxism and Postmodernity)*, Santa Clara, Cuba and Bogotá: Universidad de Las Villas and Universidad INCCA.(An orthodox Cuban Marxist's account of the 1990s crisis of Marxism. Reading knowledge of Spanish required.)

Hodges, D.C. (1986) *Intellectual Foundations of the Nicaraguan Revolution*, Austin, TX: University of Texas Press.(A study of the Sandinistas' complex intellectual history and its impact on the 1979 Nicaraguan revolution.)

Liss, S.B. (1984) *Marxist Thought in Latin America*, Berkeley and Los Angeles, CA: University of California Press.(An introductory historical overview of Marxist political thought, classified by countries, from the late nineteenth century up to the 1970s.)

Löwy, M. (ed.) (1982) *El Marxismo en América Latina*, Mexico City: Era, trans. M. Pearlman, *Marxism in Latin America From 1909 to the Present: An Anthology*, Atlantic Highlands, NJ: Humanities Press, 1990.(An informative anthology of Marxist political currents, selected according to political movements, parties, and tendencies, from 1909 to the 1970s. Reading knowledge of Spanish required.)

Mariátegui, J.C. (1928) *Seven Interpretive Essays on Peruvian Reality*, trans. M. Urquidi, Austin, TX: University of Texas Press, 1971.(A clearly written classic in Latin American Marxism, best known for its original treatment of the Indian question. Its contents are the focus of a central portion of this entry.)

Sánchez Vázquez, A. (1965) *Las ideas estéticas de Marx*, Mexico City: Era, trans. M. Riofrancos, *Art and Society: Essays in Marxist Aesthetics*, New York: Monthly Review Press, 1974.(A contemporary account of the opposition between capitalism and artistic creation as developed by a prominent Spanish-born Marxist philosopher. Advanced reading knowledge of Spanish required.)

Sánchez Vázquez, A. (1988-9) 'Marxism in Latin America', *The Philosophical Forum* 20 (1-2): 114-28.(A clear and accessible account of the historical and intellectual roots of Latin American Marxism.)

Santos Moray, M. (ed.) (1985) *Marxistas de América: Julio Antonio Mella, José Carlos Mariátegui, Aníbal Ponce, Juan Marinello*, Havana: Editorial Arte y Literatura.(A selection of Marxist classics focusing on accounts of culture and society. It includes a name index. Advanced reading knowledge of Spanish required.)

Schutte, O. (1993) *Cultural Identity and Social Liberation in Latin American Thought*, Albany, NY: State University of New York Press.(A study of representative intellectual and social movements for liberation from the 1920s to the 1980s, including an analysis of Mariátegui and the origins of Latin American Marxism. Contains an expanded account of some material covered in this entry.)

Masaryk, Tomáš Garrigue (1850-1937)

Masaryk was a philosopher, sociologist, politician and first president of the Czechoslovak Republic (1918-35). Initially he aimed to change the Habsburg monarchy into a democratic federal state, but during the First World War he began to favour the abolition of the monarchy and, with the help of the Allied powers and the Czechoslovakian foreign armed forces, won independence for his nation. Masaryk's philosophy of history posited democracy achieving victory over theocracy as a stage in world evolution. He regarded democracy as both a political system and a humanistic world outlook.

Tomáš Garrigue Masaryk was born into a poor family and gained access to higher education only with difficulty. After grammar school in Brno and Vienna he studied philosophy and classical philology at the University of Vienna where he became a student of Franz Brentano. Having achieved his Ph.D. in 1876, he studied in Leipzig where he met Fechner and Wundt and befriended the young Husserl. Masaryk's 1879 habilitation thesis at the University of Vienna was *Der Selbtstmord als sociale Massenerscheinung der modernen Civilisation (Suicide and the meaning of civilization)*, which contained many ideas later developed and expanded in his philosophy. It concluded that suicide is a result of the prevalence of subjectivism, individualism and one-sided rationalism which had dismantled the Christian worldview. Masaryk sought the solution in an organic cultural synthesis, in education involving the unity of philosophy and science, theory and practice, in a new non-mythical religion, and in the reform of society according to sociological knowledge.

In 1882 Masaryk was appointed professor at the University of Prague. Here his focus went beyond academic activity, bringing supranational criteria to bear on provincial conditions and narrow national interests. Czech cultural life, until then looking exclusively to Germany, was becoming enriched with French and Anglo-American ideas, reflected in Masaryk's first philosophical work, *Základové konkrétné logiky (Fundamentals of concrete logic)* (1885). Although his most theoretical, this was a distinctly practical work, reflecting Masaryk's conviction that theoretical knowledge must serve life and his aim to reform society. Following Comte he created a system from which it is possible to deduce a consistent philosophical and simultaneously scientific worldview, but he placed a greater stress on psychology and ethics and their social role (see Comte, A.). He regarded metaphysics and theism as compatible with scientific knowledge.

Masaryk considered Czech politics of the time to be ineffectual, lacking both a theoretical basis and an ethical dimension. In *Česká otázka (The Czech question)* (1895) and other studies, he sought the *raison d'être* of the Czech nation and formulated a philosophy of Czech history. Its long humanitarian tradition based on religion he dated from the time of the Czech Reformation, continuing during the Czech National Revival at the beginning of the nineteenth century. This tradition, characterized by love of truth and genuine Christian life, was also democratic. Such humanitarian ideas parallel those of other progressive nations: consequently the Czech question represented the world question, and Masaryk believed its politics should be in harmony with this philosophy of history. His immediate target was the transformation of the Austrian multinational monarchy into a democratic federation, securing free evolution for all nations, and he actively pursued this idea until the outbreak of the First World War.

Masaryk's humanism centred around his interest in social problems, and he sympathized with many of the demands of the labour movement. He also concentrated on theory, balancing the philosophical and sociological perspectives of Marxism in *Otázka sociální (The social question)* (1898). He admired Marx's analysis of the importance of human labour, but rejected revolution as a solution to social conflict; rather, he emphasized reform in the spirit of humanism. His outlook is expressed by the aphorism ' Always for the worker, very often with socialism, rarely with Marxism' (Čapek 1938: 214).

Masaryk's interest in Russia culminated in his most extensive work *Rußland und Europa. Studien über die geistigen Strömungen in Rußland (The Spirit of Russia)* (1913). He analysed and critically evaluated the different trends in Russian thinking reflected in theoretical essays and literature. Dostoevskii's works, which he regarded as representing a new direction, challenge Masaryk's own conception of humanitarian philosophy.

The First World War proved a political turning point for Masaryk. He came to the conclusion that the Austrian

monarchy could not be reformed by inner forces, as it directly represented a metaphysical evil which was partially to blame for the outbreak of war. For that reason Masaryk opted for an independent Czech-Slovak state. Elected first president of the Czechoslovak Republic in 1918, he retired from academic life but did not cease to be interested in philosophy. He considered the Czech struggle to be part of a world revolution, a battle against theocracy which democracy can be expected to win, a viewpoint reflected in *Světová revoluce (The Making of a State)* (1924).

See also: Czech Republic, philosophy in; Slovakia, philosophy in; Suicide, ethics of

Translated by G.R.F. Bursa
JOSEF ZUMR

List of works

Masaryk, T.G. (1879) *Der Selbtstmord als sociale Massenerscheinung der modernen Civilisation (Suicide and the meaning of civilization)*, trans. W.B. Weist and R.G. Batson, intro. A. Giddens, Chicago, IL: University of Chicago Press, 1970.(Suicide assessed as a problem of modern mankind.)

Masaryk, T.G. (1885) *Základové konkrétné logiky (Fundamentals of concrete logic)*, Prague: Bursík a Kahout. (Masaryk's theory of science and concept of philosophy, published in German as *Versuch einer concreten Logik* in 1887.)

Masaryk, T.G. (1895) *The meaning of Czech history*, trans. P. Kussi, ed. and intro. R. Wellek, Chapel Hill, NC: University of North Carolina Press, 1974.(Anthology of Masaryk's works including *The Czech Question*, *Palaky's Idea of the Czech Nation*, John Huss and Karel Havlíček.)

Masaryk, T.G.(1898) The Social question: philosophical and sociological foundations of Marxism, in *Masaryk on Marx; an abridged edition of T.G. Masaryk*, ed. and trans. E.V. Kohak, Lewisburg, PA: Bucknell University Press, 1972.(Anthology of Masaryk's work *The Social Question*.)

Masaryk, T.G. (1913) *The Spirit of Russia; studies in history, literature and philosophy*, trans. E. Paul and C. Paul; additional chapters and bibliographies J. Slavik, the former translated and the latter condensed and translated by W.R. Lee and Z. Lee (2nd, 3rd impressions), London: Allen & Unwin; New York: Macmillan, 1961.(This English translation includes the third, unpublished part of Masaryk's manuscript.)

Masaryk, T.G. (1920) *The New Europe (the Slav standpoint)*, ed. W. Preston and W.B. Weist, intro. O. Odlozilik, Lewisburg, PA: Bucknell University Press, 1972.(Project of the new political organization of European states after the First World War.)

Masaryk, T.G. (1924) *The Making of a state; memories and observations, 1914-1918*, ed. H. Wickham Steed, New York: F.A. Stokes Co., 1927.(Translation of *Světová*, reprinted by H Fertig in 1969.)

References and further reading

Beld, A. van den (1976) *Humanity: the political and social philosophy of Thomas G Masaryk*, The Hague: Mouton.(Analysis of Masaryk's humanitarian philosophy).

Branch, M.A. (ed.) (1990) *T.G. Masaryk (1850-1937)*, London: Macmillan, 3 vols, vol. 1: *Thinker and Politician*, ed. S.B. Winters; vol. 2: *Thinker and Critic*, ed. R.B. Pynsent; vol. 3: *Statesman and Cultural Force*, ed. H. Hanak.(Critical assessments of Masaryk's achievements as philosopher, politician and publicist.)

Čapek, K. (1938) *Masaryk on Thought and Life*, trans. M. Weatherall and R. Weatherall, London: Allen & Unwin; New York: Arno Press, 1971.(Interview with Masaryk.)

Kovtun, J.G. (1981) *Tomas G. Masaryk 1850-1937*, Washington, DC: Library of Congress.(A selected bibliography of reading materials in English.)

Novak, J. (ed.) (1988) *On Masaryk*, Amsterdam: Rodopi.(Collection of studies of Masaryk's philosophical and sociological views; texts in English and German.)

Skilling, H.G. (1994) *Against the Current. T.G. Masaryk 1882-1914*, London: Macmillan.(Masaryk's activity and his struggle in the Austrian monarchy until the First World War.)

Zeman, Zbynek A.B. (1976) *The Masaryks: the making of Czechoslovakia*, New York: Barnes & Noble. (Historical monograph about Masaryk's part in the foundation of Czechoslovakia.)

Zumr, J. and Binder, T. (eds) (1992) *T.G. Masaryk und die Brentano-Schule Praha/Graz,* Filosoficky ustav CSAV/Forschungstelle und Dokumentationszentrum für österreichische Philosophie.(Collection of studies directed to philosophy of science and religion and to Masaryk's place in history of philosophy - in English and German.)

Masham, Damaris (1658-1708)

Damaris Cudworth, who became Lady Masham on her marriage to Sir Francis Masham in 1685, was an English moral philosopher who published two short treatises on moral philosophy. These show that she became a disciple of John Locke, although her philosophical background was in Cambridge Platonism. She applied Lockean arguments to defend the education of women; her anti-idealism led her to oppose Malebranche and his English followers, John Norris and Mary Astell; and she also corresponded with Leibniz.

Lady Masham was one of the very first English women to publish philosophical writings. The daughter of Cambridge Platonist Ralph Cudworth, she was also a very close friend of John Locke, who spent his last years at her residence in Essex. Unlike most women of her time, Damaris Cudworth was able to develop her interest in philosophy thanks to the fortuitous circumstances first, of being born into an intellectual family and, second, of becoming the close acquaintance of the leading English philosopher of his generation. Little is known of her education beyond the fact that she learned French but not Latin, which suggests that she had no more than the grooming deemed fit for a lady. In 1685, shortly after Locke's departure for exile, she married Sir Francis Masham, an Essex squire with a family of nine children. She herself had one child.

There are three sources for documenting Lady Masham's activities as a philosopher: her correspondence with Locke, her two short books, and her letters to Leibniz. Her two books were published anonymously and her biography of Locke was printed in abridged form in Moreri's *Grand Dictionnaire Historique* (1728). It is clear from Lady Masham's early correspondence with Locke (dating from 1682) that her introduction to philosophy was through the writings of the Cambridge Platonists, Ralph Cudworth, Henry More and John Smith (see Cambridge Platonism). However, the letters also show that she was receptive to Locke's views, and her two books (published with the encouragement of Locke) show her assimilation of his philosophy. In these she draws on Locke in her opposition to the Malebranchian idealism of Mary Astell and John Norris, and in her arguments for the education of women. The same anti-idealist position informs her critique of Leibniz. But her allegiance to Locke did not entail a repudiation of Cambridge Platonism in all respects: the Leibniz correspondence also includes a defence of her father against Pierre Bayle's imputation of covert atheism (see Bayle, P.).

The connecting thread of Lady Masham's thought is religious and moral. Her published writings are broadly Lockean in epistemology. *A Discourse Concerning the Love of God* (1696) is a critique of the correspondence between John Norris and Mary Astell which had been published as *Letters Concerning the Love of God* (1695). Lady Masham objected that Norris' occasionalist account of human love of God denigrates created things and undermines the bonds of society. By contrast, Lady Masham argued, humans come to know and to love God by observing the created world and comparing ideas 'received from *Sense* and *Reflection*'. She emphasizes practical morality, arguing that it is integral to religion. This position is developed in *Occasional Thoughts in Reference to a Vertuous or Christian Life* (1705) which contains an assessment of the role of reason in religious matters and, in particular, the relationship between religion and morality. Here she insists that civil and religious liberties are necessary for the exercise of virtue and that education is the key means to inculcate virtue. She displays an optimistic view of humankind as rational and social beings for whom the love of happiness is 'the earliest and strongest principle', happiness being the enjoyment of pleasure, and the pursuit of pleasure involving the regulatory exercise of reason which directs us to the greatest happiness. She is critical of Malebranche's pessimistic view of children as natural born sinners as a consequence of each child loving the body of its mother while in the womb. *Occasional Thoughts* concludes by arguing for the education of women on moral grounds: first, instruction of women by precept rather than reason produces the opposite of the desired result (superstition and immorality instead of right belief and virtue); second, since moral principles are best inculcated in childhood, women, as mothers, have an important role as teachers of their children.

In her letters to Leibniz, Lady Masham takes issue with him on a number of points, among them his concept of 'atomes de substance' - whether unextended substances exist and how free will can be maintained within pre-established harmony. Her most important point relates to her critique of Norris: namely that, if matter is denied a causal role in Leibniz's system, the organization of matter is rendered 'superfluous and lost labour'. Although Lady Masham helped to establish a place for women in philosophy, she accepted limitations on their philosophical

activities: in *Occasional Thoughts* she notes that women should not spend their time in unnecessary speculations, but they should endeavour to improve their minds by more than merely doing their duty.

SARAH HUTTON

List of works

Note on correspondence: Lady Masham's correspondence has not been collected, apart from her letters to Leibniz which have been reproduced in *Die Philosophischen Schriften von Gottfried Wilhelm Leibniz* (ed. C.I. Gerhardt, Berlin: Weidmannsche Buchhandlung, 7 vols, 1875-90, vol. 3, 372) and to Locke in *The Correspondence of John Locke*, (ed. E.S. de Beer, 8 vols, Oxford: Clarendon Press, 1976-, vols. 2 and 3). Letters to Shaftesbury are in the London Public Records Office (PRO 30/24/20, fols. 266-7 and 273-4), while Amsterdam University Library has three letters to Philip van Limborch (MS M31c).

Masham, D. (1696) *A Discourse Concerning the Love of God*, London; French trans. *Discours sur l'Amour Divin*, Amsterdam, 1705.(A critique of *Letters Concerning the Love of God* (1695) by John Norris and Mary Astell.)

Masham, D. (1705) *Occasional Thoughts in Reference to a Vertuous or Christian Life*, London.(The second printing in 1747, *Thoughts on a Christian Life*, was misattributed to Locke.)

Masham, D. (1728) 'Locke, (Jean)', in L. Moreri, *Le Grand Dictionnaire Historique*, nouvelle edition, Paris, vol. 3.(This is a shortened version of a memoir of Locke, now in Amsterdam University Library (Remonstrants MSS J. 57a), which also forms the basis of *Eloge de feu Mr. Locke* by Jean Le Clerc in *Bibliotheque choisie*, vol. 6, 1705, 342-411.)

References and further reading

Ballard, G. (1752) *Memoirs of Several Ladies of Great Britain*, Oxford, 379-88; repr. Detroit, MI: Wayne State University Press, 1985.(General and uncritical on her philosophy.)

Hutton, S. (1993) 'Damaris Cudworth, Lady Masham: between Platonism and Enlightenment', *British Journal for the History of Philosophy* 1: 29-54.(A study of Masham's philosophy in relation to Locke and Cambridge Platonism.)

Norris, J. and Astell M. (1695) *Letters Concerning the Love of God*, London. (Correspondence between John Norris and Mary Astell.)

Laslett, P. (1953) 'Masham of Oates', *History Today* 3: 535-43.(Biographical background.)

Simonutti, L. (1987) 'Damaris Cudworth Masham: una Lady della Repubblica delle Lettere', *Studi in Onore di Eugenio Garin (Studies in honour of Eugenio Garin)*, Pisa: Scuola Normale Superiore, 141-65.(A study of Masham's philosophy.)

Mass terms

Mass terms are words and phrases such as 'water', 'wood' and 'white wallpaper'. They are contrasted with count terms such as 'woman', 'word' and 'wild wildebeest'. Intuitively, mass terms refer to 'stuff'; count terms refer to 'objects'. Mass terms allow for measurement ('three kilos of wood', 'much water'); count terms allow for counting, quantifying and individuating ('three women', 'each word', 'that wildebeest over there').

Philosophical problems associated with mass terms include (1) distinguishing mass from count terms, (2) describing the semantics of sentences employing mass terms, and (3) explicating the ontology presupposed by our use of mass versus count terms. Associated with these philosophical issues - especially the third - are the meta-philosophical issues concerning the extent to which any investigation into the linguistic practices of speakers of a language can be used as evidence for how those speakers view 'reality'.

1 Distinguishing mass and count terms

The distinction between mass terms ('water', 'wood' and so on) and count terms ('woman', 'word' and so on) can be seen as syntactic, semantic or pragmatic. If the distinction is seen as syntactic, one might remark that mass terms occur with the quantifiers 'much' and 'little' and with the unstressed article 'some', that they are susceptible to measurement phrases such as 'litres of' and 'amount of', and that they do not exhibit a singular/plural distinction. Conversely, count expressions occur with the quantifiers 'each', 'every', 'many', 'several', 'few' and the stressed 'some', use the indefinite article 'a(n)', are susceptible to counting phrases such as 'five' and 'a score of', and exhibit a singular/plural dichotomy manifested in the count term itself and in agreement with the verb phrase.

If it is seen as semantic - a distinction between the different ways that mass and count terms refer - then one might remark that count expressions refer to discrete, well-delineated objects while mass terms refer without making it explicit how the referent is individuated (some have said that the referents of mass terms are continuous rather than discrete). This feature of mass reference gives rise to the 'cumulative reference test' (any sum of parts which are *M* is also *M*) and to the 'distributive [homogeneous] reference test' (any part of something which is *M* is also *M*).

If the mass/count distinction is seen as a pragmatic one, then one will look to how people use count terms to 'individuate' the world. This gives rise to such tests as whether there is a definite answer to the question 'How many *X*'s are there in such-and-such place?'. In the philosophical literature (following Strawson 1959), terms which pass this counting test are often called 'sortal terms' - although they are equally often called count terms - and ones that fail the test are called mass terms. In this literature, with its emphasis on the pragmatic notions of 'identifying' and 'individuating', it is common to deny that such terms as 'thing', 'object' and so on are sortal (count), there being no definite answer to the question 'How many things are in the room?'. This is so despite the fact that such terms clearly satisfy the syntactic criteria.

All these tests - the syntactic, the semantic and the pragmatic ones - have been challenged. Writers have pointed out that mass terms such as 'wood' can also be used as count: 'a wood' might designate oak or spruce, for example. And 'wildebeest' can be used as a mass term: 'He's not really a vegetarian; he eats wildebeest'. Furthermore, a universal grinder would take an object that an alleged count term referred to - a chair, for example - and grind it up into a powder so that then there would be chair all over the floor. (This last sentence uses 'chair' in a mass manner, thereby showing that the language already has this usage of any count term in the background. And this usage exists despite the fact that we could also have said 'There is wood all over the floor'.)

The criteria just mentioned are usually seen as applying to entire noun phrases as well as to the simple nouns themselves. Furthermore, some writers attempt to apply the mass/count distinction to other syntactic categories, especially adjectives, verbs and adverbs. The application to verbs is especially interesting (see Mourelatos 1978).

2 Semantics and ontology of mass terms

The problem with giving a formal semantic analysis of mass terms arises because first-order predicate logic appears to assume that the entities in the domain of quantification are individuals, so it only makes sense to characterize them with count nouns. When we say, in the quantifier idiom, 'For all *x*, if *x* is *F*…', it is apparently

assumed that the items in the domain have already been individuated. For if *F* were to be interpreted as 'snow', for example, what would be the values of *x*?

Famously, Quine (1960) held that mass terms are ambiguous: when in 'subject position' they are singular terms (names), but when in 'predicate position' they are general terms (predicates) which are 'true of each portion of the stuff in question, excluding only the parts too small to count'. As a name (when in subject position), Quine holds that a mass term 'differs none from such singular terms as "mama"…, unless the scattered stuff that it names be denied the status of a single sprawling object'.

This proposal has not satisfied various authors, who have objected to the nonuniform treatment and to various logical consequences of this approach. For example, on Quine's analysis, 'Water is wet, and this puddle is water' does not imply 'This puddle is wet'; and 'Water is water' does not come out a logical truth. Writers after Quine have proposed many different approaches. Possibly the most popular alternatives involve mereology, according to which the main operator is 'is a part of'. Mass (and other) terms are taken to designate 'mereological wholes'. Some authors have grafted onto pure mereology a notion of 'having certain structural properties', so as to avoid the minimal parts problem alluded to in Quine. (The atoms, *inter alia*, which are part of water are 'too small' to count as water.) But these theories also have not satisfied all those involved in this area, usually because the treatment of certain logical inferences is thought incorrect: the formal semantic analyses do not mirror intuitive beliefs concerning logical consequence.

An alternative is to retain the idea that mass terms name some kind of object - a 'substance' - and to invoke a relational predicate such as 'is constituted of'. This presents a number of tricky issues and there are types of sentences for which such an analysis is not obviously suitable, but still various authors have adopted it. Besides the formal differences entailed by these two approaches (mereological calculus of individuals versus classical logic with a relational constant of constitution), there is an ontological difference, for mereological wholes are generally taken to be physical whereas substances or kinds are often viewed as abstract entities (see Substance).

Another formal semantics of mass terms invokes sets as their denotation. Differences among theorists can then be seen as differences about what the sets contain. One question on which theorists differ is whether the sets contain only 'minimal entities' - the smallest items to which the mass term refers (flakes, maybe, for 'snow'; the items and size vary according to the mass term in question) - or whether it should contain 'ordinary objects' (flakes, drifts, snowmen, snowballs and so on; any object which can be said to be snow). The former proposal has not gained many adherents due to the difficulty of specifying a set of 'minimal entities' for such mass terms as 'garbage', 'speed' and 'information'. The latter proposal runs into difficulties in trying to account for the denotation of definite noun phrases (NPs) such as 'the snow on the table'. It is generally not true that there is exactly one snow-thing on the table. (There is one ball and also many flakes making it up, for example.) So the only reasonable proposal is for the NP to designate *all* the snow-things on the table. But then certain measurement sentences - for example, 'The snow on the table weighs one kilo' - come out wrong, since we will count the same snow-entities many times over.

Theories of mass terms show a fundamental division between those that are committed to abstract substances and those that are physicalistic in nature, invoking mereological wholes. On the physicalistic side are those theories which propose that the ontologically basic objects are the minimal entities, those which claim that the larger entities are 'constructions' out of these minimal entities, and those theories which propose that all these entities are equally basic. On the other side of the gulf are the various styles of substance theories, which usually invoke a lattice structure of kinds. Such ontological issues are discussed in Pelletier and Schubert (1989) and in Burkhardt and Smith (1991).

See also: Logical and mathematical terms, glossary of; Mereology

JEFFRY PELLETIER

References and further reading

Bunt, H. (1985) *Mass Terms and Model-Theoretic Semantics*, Cambridge: Cambridge University Press.(This is the longest and possibly most thorough work on mass terms. It is aimed at a computer implementation of a natural language understanding system which will include mass terms. It invokes 'ensemble theory', which is a type of atomic mereology.)

Burkhardt, H. and Smith, B. (1991) ‘Mass Terms’, in *Handbook of Metaphysics and Ontology*, Munich: Philosophia.(This entry goes into the metaphysical and ontological presuppositions of mass terms in more detail than is possible here. It includes some speculation about how the mass/count distinction might have arisen in natural language.)

Lønning, J.-T. (1987) ‘Mass Terms and Quantification’, *Linguistics and Philosophy* 10: 1-52.(An example of the direction that research into the mereological interpretation of mass terms has taken.)

Mourelatos, A. (1978) ‘Events, Processes, and States’, *Linguistics and Philosophy* 2: 415-34.(Mourelatos draws explicit attention to the relationships that hold between certain types of verbs and verb phrases and the mass/count distinction.)

Pelletier, F.J. (ed.) (1979) *Mass Terms: Some Philosophical Problems*, Dordrecht: Reidel.(This anthology includes many of the classic articles on the topic of mass terms, especially in the philosophical tradition. It also includes a comprehensive bibliography up to 1978.)

Pelletier, F.J. and Schubert, L.K. (1989) ‘Mass Expressions’, in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 4, 327-407, esp. 391-4.(The most thorough summary of the work done on mass terms, both linguistic and philosophical. It also proposes two theories to account for the phenomena, and includes a comprehensive bibliography of work from 1978 to 1988, plus earlier work where relevant.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press, 91-100. (Although this book has only a small section on mass terms, it formed the starting point of most future formal work on the topic. Quine proposed his ‘dual analysis’ here and asserted that mass terms come from an ‘early, primitive stage in linguistic evolution’.)

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of Contemporary English*, London: Longman.(The most thorough descriptive account of the grammar of the English language, with a section about the mass/count distinction, especially as it is used in normal speech. It does not deal particularly with the formal semantics of the distinction.)

Strawson, P.F. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen.(It is here that the distinction between ‘feature-placing universal’ and ‘sortal universal’ was drawn. Many writers wished to put mass terms in the former category, giving rise in the philosophical literature to the opposition ‘mass/sortal’. Other writers thought of mass terms as a type of sortal term, while still others argued for a third category.)

Materialism

Materialism is a set of related theories which hold that all entities and processes are composed of - or are reducible to - matter, material forces or physical processes. All events and facts are explainable, actually or in principle, in terms of body, material objects or dynamic material changes or movements. In general, the metaphysical theory of materialism entails the denial of the reality of spiritual beings, consciousness and mental or psychic states or processes, as ontologically distinct from, or independent of, material changes or processes. Since it denies the existence of spiritual beings or forces, materialism typically is allied with atheism or agnosticism.

The forms of materialism extend from the ancient Greek atomistic materialism through eighteenth- and nineteenth-century scientifically based theories, to recent sophisticated defences of various types of materialism.

1 Materialism

Materialism is the general theory that the ultimate constituents of reality are material or physical bodies, elements or processes. It is a form of monism in that it holds that everything in existence is reducible to what is material or physical in nature. It is opposed to dualistic theories which claim that body and mind are distinct, and directly antithetical to a philosophical idealism that denies the existence of matter. It is hostile to abstract objects, if these are viewed as more than just a manner of speaking (see Abstract objects). An implication of materialism is that the diverse qualitative experiences we have are ultimately reducible to quantitative changes in objects or in our physiological functioning. All the properties of things, including persons, are reducible to properties of matter. Although the terms referring to psychic states such as intention, belief, desire and consciousness itself have a different sense and use than terms referring to material events, a consistent materialist would deny that mentalistic terms have reference to anything other than physical events or physiological changes in our brains. The enormous advances in the sciences have contributed storehouses of empirical data that are often used to support materialism. Many philosophers have been attracted to materialism both because of its reductive simplicity and its association with scientific knowledge.

2 Ancient Greek atomism

Although Leucippus is credited with inventing the atomic theory of matter in the fifth century BC, it was Democritus (fourth century BC) who developed a systematic theory of atomistic materialism. This theory states that matter is composed of separate and minute elements that are ‘uncuttable’ (*atoma*), that these elements move in empty space or the ‘void’. Atoms differ only in shape and volume, and all change occurs by the transfer through direct contact of movement from atoms in motion. These elementary entities are lacking in secondary qualities and are indestructible. Democritus held that things are hot or cold, sweet or bitter, or have different colours ‘by convention’. In reality, ‘there are atoms and the void’.

The essentials of early atomism were retained in Epicurus’ physics, with the exception that Epicurus ascribed freedom to atoms in their movement through space. Epicurean materialism is lucidly expressed in the philosophic poem by Lucretius, *De rerum natura (On the Nature of Things)*, in the first century BC. This popularization of Epicurean thought did much to keep alive both atomistic materialism and what is already recognizable as a naturalistic understanding of humans and world (see Epicureanism).

3 Modern materialism

During the first half of the seventeenth century the atomistic materialism of the Greco-Roman period was revived in a paradoxical way by Pierre Gassendi. He appreciated the scientific interpretation of nature and the methods of science but, at the same time, preserved the Christian idea of the immortality of the soul and conceived of God as the creator of the atoms. The English philosopher Thomas Hobbes presented a systematic theory of nature and human nature that was largely, though not completely, materialistic. Apart from attributing ‘drive’ or *conatus* to human action and sensation, Hobbes virtually banished the concept of ‘incorporeal substance’. In theory and sentiment Hobbes was a materialist thinker, although not a consistent one. The early eighteenth century saw the publication of the first of many works that defended a materialistic and mechanistic interpretation of mankind’s nature on the basis of physiological theory. In *L’Homme machine* (1748) Julien de La Mettrie, a philosophical

physician, described human beings as self-moving mechanisms and sought a neurological basis for mental activity. An advance on previous attempts to develop a systematic materialism is Paul H.D. d'Holbach's 1770 *Système de la nature (The System of Nature)*. Here, a consistent naturalistic materialism is expounded in that cognitive and emotive states are reduced to internal material 'modifications of the brain'. Though not calling it such, d'Holbach presents a form of physiological determinism.

With the rapid growth of the sciences, the astronomical discoveries of Copernicus, the theories of Galileo, and the systematic conception of nature in the physical theory of Isaac Newton, naturalistic interpretations of a variety of phenomena became more and more prevalent. This scientifically founded picture of reality lent greater plausibility to the principles of materialistic theory. The astronomer and mathematician Pierre Laplace (1749-1827) produced a sophisticated astronomical theory which, he thought, illustrated that a supermind, knowing all the states and conditions of every existing entity, could predict the total state of the cosmos in the next moment. When Napoleon I was shown a copy of Laplace's work, he is supposed to have commented on the absence of any mention of God. Laplace replied, 'I have no need of that hypothesis'. Laplace's mechanistic materialism became, in the hands of many thinkers, the definitive explanatory principle of all events.

The formulation of the biological theory of evolution by means of natural selection by Charles Darwin virtually eliminated teleological explanations of biological phenomena and thereby buttressed material and physical interpretations of organic development. With the advances in chemistry achieved by Lavoisier (1743-94) in France and John Dalton (1766-1844) in England, the reductive analysis of natural phenomena to chemical substances, elements and processes bolstered the empirical, naturalistic and materialistic interpretations of phenomena. During the nineteenth century many philosophical thinkers sought to build theories on the foundation of scientific facts, principles or laws. The historical materialism developed by Marx and Engels sought to formulate laws of social, economic and historical development, but did not defend metaphysical materialism (see Dialectical materialism). The general appeal of materialism in the nineteenth century is shown by the popularity of the 1855 work by Ludwig Büchner, *Kraft und Stoff (Force and Matter)*, which passed through sixteen editions. Although philosophically crude, it is an accessible compendium of popular materialism. In 1852, Jacob Moleschott had defended the reduction of force to matter, the doctrine of the conservation of matter, and a species of objective relativism in *Der Kreislauf des Lebens (The Cycle of Life)*. Following the ill-chosen analogy between the brain and thought and the digestive system in Jean Cabanis' *Rapports due physique et du moral de l'homme (Relations of the Physical and the Mental in Man)* (1802), Karl Vogt proclaimed that the brain 'secretes' thought the way the liver secretes bile. Despite such excursions into 'vulgar materialism', the nineteenth century became a period of intense debate for scientists and philosophers alike in regard to the limits of scientific knowledge and the epistemological problems of metaphysical materialism. This was fuelled by a Neo-Kantian movement which, particularly in *Geschichte des Materialismus (History of Materialism)* (1865) by F.A. Lange, held that materialism is a useful methodological principle in science, but questionable as a reductionist metaphysics. The concepts and postulates of science are theoretical entities or conventional notions formed by the mind. Their usefulness does not, according to Lange, warrant their role as bases for materialism.

4 Recent materialism

In the twentieth century, physicalism has emerged out of positivism. Physicalism restricts meaningful statements to physical bodies or processes that are verifiable or in principle verifiable. It is an empirical hypothesis that is subject to revision and, hence, lacks the dogmatic stance of classical materialism. Herbert Feigl defended physicalism in the USA and consistently held that mental states are brain states and that mental terms have the same referent as physical terms (Feigl 1958). The twentieth century has witnessed many materialist theories of the mental, and much debate surrounding them (see Behaviourism, analytic; Functionalism; Mind, identity theory of; Materialism in the philosophy of mind).

In the field of artificial intelligence, the mind is held to be analogous to computers in so far as it functions as an information-processing entity. Daniel Dennett has, in a qualified way, argued that information-processing machines are valid models of the mind. In addition to the scientifically informed arguments for various forms of materialism, including nonreductive materialism, the twentieth-century conception of matter as composed of electrons, protons and other subatomic particles has spawned a rich speculative literature that effectively undermines previous forms of materialism. What the late US philosopher of science, Norwood Hanson, called the

'dematerialization' of matter, raises questions concerning what 'materialism' means in terms of the theories of microphysics. Many of the arguments that sustained earlier forms of materialism (including the assumption of causality as universal in nature) have been put in question. The confluence of contemporary theories about the structure and function of the mind and the nature of matter have introduced a complexity of detail and an array of paradoxical claims that make contemporary materialism a welter of intriguing, but conflicting and perplexing, theoretical elements.

See also: Matter

GEORGE J. STACK

References and further reading

Armstrong, D.M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.(Thorough analysis of mind and mental states in terms of central state physicalism.)

Büchner, L. (1855) *Kraft und Stoff*, Frankfurt: Meidinger; trans. J.F. Collingwood as *Force and Matter*, London: Trübner, 1864.(Popular compendium of the bases of materialism; philosophically unsophisticated, but interesting.)

Cabinis, J. (1802) *Rapports du physique et du morale de l'homme (The relationship between the physical and moral aspects of man)*, Paris.(Cabinis is considered one of the earliest contributors to the science of psychophysics.)

Dennett, D.C. (1978) *Brainstorms: Philosophical Essays in Mind and Psychology*, London: Bradford Books. (Solid treatment of mind conceived of as an information-processing system.)

Feigl, H. (1958) 'The "Mental" and the "Physical"', in *Minnesota Studies in the Philosophy of Science*, vol. 2, Minneapolis, MN: University of Minnesota Press. (Classic statement of physicalism.)

Gassendi, P. (1658) *Syntagma Philosophicum Epicurus*, The Hague.(Exposition and defence of Epicurus' philosophy, criticizing points that conflict with Catholic teachings.)

Hobbes, T. (1989) *Metaphysical Writings of Thomas Hobbes*, ed. M.W. Calkins, La Salle: Open Court.(Includes Hobbes' views on materialism.)

Holbach, P.H.D. de (1770) *Système de la nature*, Amsterdam; trans. H.D. Robinson as *The System of Nature*, Boston, MA, 1868.(Excellent presentation of naturalistic materialism which includes interesting criticisms of freedom of the will.)

La Mettrie, J.O. de (1748) *L'Homme Machine*, Leyden: Elie Luzac Fils; trans as *Man a Machine*, La Salle, IL: Open Court, 1912.(A physician, La Mettrie presented one of the earliest theories of the physiological bases of a variety of psychic states or a strong version of physiological determinism.)

Lange, F.A. (1865) *Geschichte des Materialismus*, Iserlohn: J. Baedeker; trans. E.C. Thomas as *The History of Materialism*, London: Routledge & Kegan Paul, 1925. (Thorough, detailed and insightful tracing of the development of materialism up to the 1870s, vitalized by Lange's penetrating insights and valuable commentary.)

Lucretius (1st century BC) *De rerum natura*; trans. W.H.D. Rouse, Cambridge, MA: Harvard University Press 1975.(Powerful philosophic poem which presents the philosophy of Epicurus in a generally accurate and appealing way.)

Moleschott, J. (1852) *Der Kreislauf des Lebens (The Cycle of Life)*, Mainz.(Study of the circular processes of life in nature with interesting observations on the relativity of perception among nonhuman beings and its application to human knowledge.)

Paulsen, F. (1892) *Einleitung in die Philosophie*, Berlin; trans. F. Thilly as *Introduction to Philosophy*, New York: Henry Holt & Co., 1895.(Lucid presentation of materialism as of the late nineteenth century, with Kantian-like critiques of its central claims.)

Smart, J.J.C. (1963) *Philosophy and Scientific Realism*, London: Routledge.(Clear, well-argued defence of a realist interpretation of the scientific world-picture primarily based on physics.)

Materialism in the philosophy of mind

Materialism - which, for almost all purposes, is the same as physicalism - is the theory that everything that exists is material. Natural science shows that most things are intelligible in material terms, but mind presents problems in at least two ways. The first is consciousness, as found in the 'raw feel' of subjective experience. The second is the intentionality of thought, which is the property of being about *something beyond itself; 'aboutness' seems not to be a physical relation in the ordinary sense.*

There have been three ways of approaching these problems. The hardest is eliminativism, according to which there are no 'raw feels', no intentionality and, in general, no mental states: the mind and all its furniture are part of an outdated science that we now see to be false. Next is reductionism, which seeks to give an account of our experience and of intentionality in terms which are acceptable to a physical science: this means, in practice, analysing the mind in terms of its role in producing behaviour. Finally, the materialist may accept the reality and irreducibility of mind, but claim that it depends on matter in such an intimate way - more intimate than mere causal dependence - that materialism is not threatened by the irreducibility of mind. The first two approaches can be called 'hard materialism', the third 'soft materialism'.

The problem for eliminativism is that we find it difficult to credit that any belief that we think and feel is a theoretical speculation. Reductionism's main difficulty is that there seems to be more to consciousness than its contribution to behaviour: a robotic machine could behave as we do without thinking or feeling. The soft materialist has to explain supervenience in a way that makes the mind not epiphenomenal without falling into the problems of interactionism.

1 From epiphenomenalism to functionalism

Would-be materialists in the latter part of the nineteenth century tended to be epiphenomenalists. They believed that the world was a physical machine, but felt obliged to concede that examination of its machinery, however minute, could never uncover nor explain consciousness. Consciousness was, therefore, an inexplicable left-over (see Epiphenomenalism). Materialism in the twentieth century has largely been concerned to provide a more integrated form of physicalism. The attempt has taken 'hard' and 'soft' forms.

The source of all forms of contemporary hard materialism is behaviourism, which identifies mental states with facts about how people are disposed to respond to external stimuli (see Behaviourism, analytic). The essence of mind is not, therefore, something private to the subject, but something public and observable, a logical product of the relation of stimulus to response. Two fundamental problems plague behaviourism. A phenomenological objection is that we can supposedly tell just from being conscious that experience is more than a mere disposition to behave. A formal objection is that one cannot give necessary and sufficient conditions for being in a given mental state solely in terms of stimulus and response. Behaviourists treated the brain as a 'black box' about which the psychologist must not speculate, but the interdependence of our mental states so complicates the relation of stimulus to response that this relation can only be understood with the aid of a model of our inner workings.

The identity theory of mind, which emerged from Australia in the work of J.J.C. Smart (1959) and D.M. Armstrong (1968), was designed to cope with both problems (see Mind, identity theory of). Armstrong accepted that mental states were dispositions, but identified these, not with abstract relations between stimulus and response, but with the states of the brain that tend to be caused by the appropriate stimulus and to cause the relevant behavioural response. This identifies experience with something occurrent and actual as well as with a disposition. Moreover, by identifying the mind with a complex internal neural structure, it allows mental states to be specified, *not* by any direct relation to stimulus and response but by a complex profile that relates them to stimulus, response *and* all the other mental states that might interact or interfere with their operation.

This theory was influenced by the development of computers and artificial intelligence (see Artificial intelligence). At first, identity theorists had tended to adopt *type identity*, according to which pain, or imagining the Eiffel Tower, would be identical with the same kind of brain state in every creature with these mental states. It was then argued both that such type identity was implausible and that it was not necessary: just as the same programme can be run on different hardware, the same network of causal relations that constitute the mind could be realized in

differently built brains. This is the theory known as functionalism, according to which there is, at most, a token identity between brain and mental states; that is, each mental state is identical with some brain state or other (see Functionalism). Some functionalists prefer to weaken the relation further by talking of the mind being *realized in* the brain rather than being identical to it.

2 Functionalism and consciousness

Functionalism is, therefore, either with or without token identity, that development from - or, perhaps, that developed form of - behaviourism, which is designed to give serious weight to inner processes. However, it faces many problems. It is questionable whether functionalism is phenomenologically any better than normal behaviourism. This doubt is expressed in a group of qualia problems, such as the ‘inverted qualia’ and ‘absent qualia’ problems. But the essence of ‘qualia’ worries seems to have become distilled in the ‘knowledge argument’ (see Qualia). Take someone lacking a certain sensible capacity from birth - they are completely deaf, for example. Suppose, too, that they have learned all that a completed physical science could tell them about the physical processes and the functional organization of the hearing mechanism: call them ‘the Deaf Scientist’. Third, suppose that they then gain their hearing. They would then gain some new knowledge, namely *what it is like to hear*, or *what sound is experientially like*. As they knew all the physical and functional information before, this kind of knowledge must concern something over and above the physical and functional; so the content of experience is something over and above the physical and functional.

Various strategies can be tried against this argument. First, one might argue that if the Deaf Scientist really did know all the physical and computational information about how hearing worked, they would know what sounds sound like. But this seems wrong, because a deaf person could not work this out from current science - more of the same physical and computational information (or the same in greater detail, which could be what a completed science gives), does not seem to be the right kind of thing to tell you what the actual experience of hearing is like. So, second, one might argue that what the formerly Deaf Scientist acquires when they find out what sound sounds like is not new knowledge, but a new way of getting a kind of knowledge they already possess; perception and science give you the same information, one directly through the senses, the other in a propositional form. So the difference is like that between getting the information that Tom is bald by seeing him and getting it by reading about him. But neither does this seem right, for though the Deaf Scientist does indeed acquire a new way of getting knowledge when they gain their hearing, they also gain knowledge of what it is like to hear. If this is not to be new knowledge, it would have to be just the same knowledge as they have as a scientist when they know about the relevant brain process, and this it does not seem to be.

At least it is clear that the knowledge they gain by hearing is not cast in the same neurological terms as the scientific information they already knew, so at least new concepts are involved. Does this imply that new information is acquired? It has been suggested that the necessity for different concepts need not mean that what is being presented are different properties of, or different facts about, the world. The same property or situation is presented in different ways, as when two concepts have the same property as reference while having different senses (for example, ‘blue’ and ‘the colour of the sky’ present the same feature of the world). But unless there is some reductive analysis of how the mental concepts capture the physical properties, such as the functionalist would provide, it is difficult to see why there should be need for new concepts to capture experience, unless it were that they captured different properties from those caught by the physical vocabulary.

Adopting a functionalist account of the mentalistic concepts that actual hearing calls into play is equivalent to admitting that the Deaf Scientist acquires only a kind of ‘knowledge how’ on coming to hear - that is, they come to be able to respond directly to sounds. This toughly neo-behaviourist view is sometimes seemingly alleviated by saying that what the Deaf Scientist comes to learn is how to imagine, or how to remember, the experience. This does not appear too harshly reductive, for what they come to be able to do does not appear to be mere behaviour. But that is only because we still need an account of the mentalistic notions of imaging, and of remembering what an experience was like: taking images (memory or otherwise) as basic is no different from taking ‘raw feels’ as basic. But if these notions are treated in functionalist terms, being finally explained in terms of an ability to say and outwardly do certain things, then the response has become totally reductive and will not convince anyone who thought there was a problem for the functionalist in the first place.

3 Functionalism and matter

It is a largely unexamined assumption of late twentieth-century materialists that the concept of matter, unlike that of mind, is unproblematic. Some philosophers, however, have followed Russell and argued that the conception of the physical world given to us by science is purely functional and formal: science, that is, tells us nothing about the intrinsic nature of anything, only about its behaviour and how to quantify it, and that this alone is an inadequate conception. Qualitative content is given to our conception of the world from the qualia presented in perception, and without that there would be no non-relational and non-formal content to our conception of the world. But functionalism denies that qualia are given in experience and presents the same kind of purely relational conception of experiential states as physical science does of the physical world. If functionalism is correct, we are stuck with a conception of the world which is entirely relational, with no intrinsic content to any of the relata. This may be an incoherent conception.

4 Alternatives to functionalism

One response to the problems faced by these reductive styles of materialism is to abandon the attempt to reduce mental properties and to admit a *dualism* of properties, while prescribing that the mental *supervenes* on the physical in such a tight way that the spirit of materialism is preserved (see Supervenience). This tightness would mean that the dependence was more than merely causal, without being analytic, in the way a reductive account would require.

The first problem this theory faces is to give a proper rationale for supervenience. The provision of a simple definition is easy: the mental supervenes on the physical if there cannot be a mental difference or change without a physical difference or change, and the ‘cannot’ depends on something stronger than a natural law. The problem comes in explaining what this stronger-than-natural necessity is, and why one should believe in it. No generally acceptable account of this seems to have been found. Furthermore, the question remains of whether the matter of the brain is influenced in its behaviour by the supervenient mind. If it is, then the physical system is being influenced by something not itself physical, and this is a species of interactionism. If it is not influenced, then the mind is an epiphenomenon. However, twentieth-century materialism had sought to avoid both interactionism of this kind and epiphenomenalism (see Epiphenomenalism).

There is another and entirely opposite response to the difficulties of functionalism in accommodating the apparent data of thought and experience, and that is to deny that there are any such data. Various traditions came together to bear exotic fruit: the pragmatist doctrine that one should believe whatever is most convenient overall; the post-positivist tendency to deny the distinction between observation and theory; Wittgenstein’s polemic against private data, produce the doctrine that nothing is so plainly revealed in everyday experience that it cannot be overridden by requirements of scientific elegance. Even the seemingly most obvious facts of experience are actually theoretical speculations and can be denied if science is thereby made easier (see Eliminativism; Pragmatism; Private language argument).

Two conceptions of theory seem to be operating in such eliminativism, and they may not always be clearly distinguished. According to one, a theory is any generalization that can be used to provide explanations; according to the other, a theory is something not so blatantly obvious as to be free from revision. This latter, of course, goes naturally with the observation-theory distinction that the eliminativist rejects. But there are many explanatory principles that are totally beyond revision; for example, that wood generally floats, that pigs cannot fly and that cutting off someone’s head kills them. Any higher-order theory that contradicted any of these would be false. Perhaps the use of the term ‘folk psychology’ to characterize our normal psychological generalizations makes them seem more like folklore - more like the belief that comfrey boiled in holy water relieves rheumatism - than they are like the belief that pigs cannot fly. It seems plausible, however, that the law that we shout out because of pain, or eat because we feel hungry, belongs more with the pigs than the comfrey.

5 Cognitive science and intentionality

The development of Artificial Intelligence tended to switch philosophers’ attention from sensation to cognition and, hence, to intentionality. The assumption has been that human behaviour is driven by computational activity in the brain. Two kinds of question have been raised. The first concerns the relation between this computation and the ordinary psychological explanations that we give of behaviour - rather coyly described as ‘folk psychology’ (see

§4 above; Folk psychology; Reductionism in the philosophy of mind). The other is more directly concerned with the problem intentionality poses for materialism. The language of a computer program works entirely as a formal syntax, and brain states, considered simply as biochemical, have no intrinsic meaning; how, then, does neural computation come to have semantic content? Everyone seems to agree that it must be in virtue of the functional relations inner representations have to their typical causes and their behavioural effects, possibly considered in relation to the wider environment and the evolutionary process that produced them (see Functionalism §7). The details of this functional account are disputed, but the main problem is that it is entirely externalist (see Content: wide and narrow). If one thinks of consciousness as something that goes on 'in the head', and that one's own consciousness of one's thoughts is consciousness of one's internal representations, and if those representations have meaning because of the *external* relations they stand in to the world outside the head, then it looks as if consciousness will not reveal to one the contents of any of one's thoughts, nor, indeed, of any kind of mental representation. For all purposes, we will be in the situation of Searle in his 'Chinese room', inspecting symbols we do not understand (see Chinese room argument).

Tyler Burge (1988) has tried to answer this objection to externalism. He says that the holding of the external relations is an *enabling condition* for the representation to have meaning, but that knowing something - in this case the meaning - does not involve explicit knowledge of all the enabling conditions. So we can know the contents of our own thoughts although we do not know the relations that give them their content. The problem with this reply is that, although it is true that we do not, in general, need to know all the things that make something possible in order to know the thing itself, the external relations seem not to be just enabling conditions, but to constitute the content itself. Searle's thought experiment seems to show how ignorance of them actually constitutes ignorance of the meaning of the representations.

6 Materialism and abstract objects

Although materialists are mainly concerned with problems that flow from the philosophy of mind, abstract objects constitute a much less discussed but serious problem. If one believes that it is an irreducible fact that there are any or all of numbers, universals, properties, sets or propositions, then one believes there are things which are not material particulars. It might be that one did not mind adding such light baggage as abstract objects to one's otherwise materialist ontology; but if there are such things as universals and propositions, it looks as if they must enter intimately into our thinking, or, if numbers, into our counting. This would then create a problem for giving a materialist account of these activities, and so accepting abstract objects might have consequences for the philosophy of mind (see Abstract objects).

7 Materialism at the *fin de siècle*

The materialist mood in the twentieth century has been poised between an almost triumphalist self-confidence and a more modest perplexity. The triumphalism is produced by the success of science, which makes materialism seem obviously true. In this mood, materialists are prepared to deny what seem to be the most obvious facts of mental life if their theory requires it. In a more sombre moment, however, some will confess that all attempts to tackle the problems have so far missed the mark. This more sober tendency became stronger in the 1980s and 1990s. Nagel (1974) had already declared that the mind-body problem could only be solved by a conceptual breakthrough we could not, as yet, imagine. McGinn (1991) pronounced the problem insoluble in principle because the mind cannot understand itself. Galen Strawson (1994) has denied that there is any conceptual connection between mind and behaviour. All these philosophers deem themselves to be materialists, of some not-yet-quite-articulable kind. The *Journal of Consciousness Studies* has been set up to 'take consciousness seriously' in a way it is said science has not so far done; but perhaps this underestimates the main reason why consciousness has been sidelined and treated harshly: namely because it seems so clearly impossible to say anything constructive about it within the materialist presuppositions of natural science.

HOWARD ROBINSON

References and further reading

Armstrong, D.M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul. (A classic - full of lucid philosophical argument.)

Burge, T. (1988) 'Individualism and self-knowledge', *Journal of Philosophy* 85 (11): 649-63. (One of a series of

pieces by Burge defending externalism. In this one he makes wide use of the notion of enabling conditions.)
Churchland, P.M. (1984) *Matter and Consciousness*, Cambridge, MA: MIT Press, 2nd edn.(An excellent general account of all the basic ‘isms’.)
Churchland, P.M. (1989) *A Neurocomputational Perspective*, Cambridge MA: MIT Press.(The most thorough-going and consistent eliminativist collection of papers.)
Dennett, D. (1990) *Consciousness Explained*, Harmondsworth: Penguin.(The opposite pole to Armstrong in philosophical style; an entertaining and provocative attempt to use science to bypass the philosophical problems.)
Foster, J. (1991) *The Immaterial Self*, London, Routledge.(A powerful attack on all forms of materialism.)
Hale, B. (1987) *Abstract Objects*, Oxford, Blackwell.(A defence of the irreducibility of abstract objects.)
Jackson, F. (1982) ‘Epiphenomenal Qualia’, *Philosophical Quarterly* 32 (127): 27-36.(The classic source for the ‘knowledge argument’.)
Lockwood, M. (1989) *Mind, Brain and the Quantum*, Oxford, Blackwell.(Is worried by the purely relational scientific conception of matter and makes a serious attempt to rehabilitate Russellian neutral monism as a means to saving materialism. Excellent account of modern science.)
McGinn, C. (1991) *The Problem of Consciousness*, Cambridge, MA and Oxford: Blackwell. (Chapter 1 contains the author’s reason for thinking that the mind-body problem is insoluble. He argues, however, that this is not a disturbing conclusion.)
Nagel, T. (1974) ‘What is it like to be a bat?’, *Philosophical Review* 83 (4): 435-50.(This article achieved great fame by elegantly reminding US philosophers that there was such a thing as subjectivity.)
Papineau, D. (1993) *Philosophical Naturalism*, Oxford: Blackwell.(This contains a well-worked-out version of the teleological-cum-evolutionary theory of representation.)
Robinson, H. (1982) *Matter and Sense*, Cambridge: Cambridge University Press.(A general critique of materialism. Chapter 7 is concerned with problems in the concept of matter, especially the relational versions to which materialists seem forced.)
Searle, J. (1980) ‘Minds, brains and programs’, *Behavioural and Brain Sciences* 3 (3): 417-24. (Contains the ‘Chinese room’ argument, followed by criticisms and a reply.)
Searle, J. (1992) *The Rediscovery of the Mind*, Cambridge, MA and London: MIT Press.(A lucid statement of the view that the emergence of mind is not a problem for materialism, because emergence occurs at all levels in nature.)
Shoemaker, S. (1984) *Identity, Cause and Mind*, Cambridge: Cambridge University Press. (Perhaps the most rigorous development of analytical functionalism. Chapters 10 and 11 defend the view that *all* properties are individuated causally.)
Smart, J.J.C. (1959) ‘Sensations and brain processes’, *Philosophical Review* 68 (2): 141-56. (Important for popularizing the idea that awareness of one’s own experience is grasping an inner state *topic neutrally*, under the aspect of its causal relations to a stimulus.)
Smith, A.D. (1993) ‘Non-reductive physicalism?’, in H. Robinson (ed.) *Objections to Physicalism*, Oxford: Clarendon Press, 225-50.(A clear discussion of the problems of supervenience and of the emergence of the kind favoured by Searle.)
Strawson, G. (1994) *Mental Reality*, Cambridge, MA: MIT Press; London: Bradford Books. (An onslaught against any supposed conceptual connection between minds and behaviour.)

Materialism, Indian school of

'Materialism' stands here for the Sanskrit term Lokāyata, the most common designation for the materialistic school of classical Indian philosophy. However, at the outset 'materialism' and 'Lokāyata' were not equivalent: early materialistic doctrines were not associated with Lokāyata, and early Lokāyata was neither materialistic nor even a philosophical school.

Classical Lokāyata stands apart from all other Indian philosophical traditions due to its denial of ethical and metaphysical doctrines such as karmic retribution, life after death, and liberation. Its ontology, tailored to support this challenge, allows only four material elements and their various combinations. Further support comes from Lokāyata epistemology: the validity of inference and Scriptures is denied and perception is held to be the only means of valid cognition. As offshoots, a fully fledged scepticism and a theory of limited validity of inference developed in response to criticism by other philosophers. Consistent with Lokāyata ontology and epistemology, its ethics centres on the criticism of all religious and moral ideals which presuppose invisible agents and an afterlife. Hostile sources depict its followers as promulgating unrestricted hedonism.

1 Sources

With the single exception of Jayarāśi's *Tattvopaplavasiṃha*, no original Lokāyata works have survived. Our knowledge of Lokāyata is based mainly on fragments and references in often hostile and polemic Brahmanical, Buddhist and Jaina sources. Of the basic text of the materialistic Lokāyata, the *Bṛhaspatisūtra*, only a few fragments remain. Therefore doxographies are of particular importance for our impression of the philosophical system as a whole. Unfortunately, it is not certain to what extent these and other sources represent genuine Lokāyata doctrines. As it was the only philosophical school to deny rebirth and karmic retribution, it was criticized by all other schools. However, many of these refutations are perfunctory and do not add substantially to our knowledge.

2 Lokāyata in early sources

In the earliest references, found in the Buddhist Pāli canon (first committed to writing in the first century BC), Lokāyata designates a Brahmanical branch of learning concerned with the examination of fundamental statements about the nature of the world (equally called *lokāyata*). The *Saṃyuttanikāya* of the Pāli canon mentions four *lokāyatas*: (1) All exists; (2) All does not exist; (3) All is unity; (4) All is severalty. Such pairs of theses and antitheses were topics of public philosophical debate. A later Mahāyāna text, the *Laōkāvatārasūtra*, mentions no fewer than thirty-one *lokāyatas* on topics such as the world, the soul, life after death, liberation (*mokṣa*), momentariness, space, the state between death and birth, and ignorance, desire and karma as causes of the world. Some of them are easily recognized as points of debate among various Hīnayāna Buddhist schools. A connection to materialism is not evident in these sources, nor in another early reference, in the *Arthaśāstra*, which subsumes Lokāyata under those branches of learning which proceed by or are concerned with reasoned investigation (see Kauṭilya). By the second century BC there existed a Lokāyata commentary of unknown content by a certain Bhāguri.

3 Early materialists

The ascetic Ajita Keśakambala (Ajita Kesakambali) was a renowned teacher and contemporary of the Buddha, nicknamed after a blanket made of hair (*keśakambala*) which kept him hot in summer and cold in winter. He maintained that man consists of the four elements earth, water, fire and wind, and that after death his constituents return to the respective elementary mass: what is of earth goes to earth, what is of water to water, and so on; his senses join space. There is no afterlife. Thus sacrificial and moral works, good or bad, do not bear any fruit.

Materialism flourished also at kingly courts. Both the Jaina and Buddhist canons report the views and gruesome experiments of King Paesi (Pāyāsi), who argued that if the soul were different from the body and if a person's fate after death depended on deeds in this life, deceased relatives could be expected to come back from the other world to warn and admonish those left behind. He once had a thief condemned to death put into a hermetically sealed and guarded jar. The prisoner was found dead after a while and the escape of a soul was not observed. On another occasion, Paesi had an executed thief placed in a sealed jar. The jar was later found full of worms, although their

souls could not have entered it. Weighing immediately before and after death and dissection of bodies did not yield evidence for the existence of a soul.

It is noteworthy that both the legendary founder of Lokāyata and the founder of a school of political science (*arthaśāstra*) bear the name Bṛhaspati. It is quite possible, though not yet provable, that Indian materialism developed in kingly and state administration circles as an alternative worldview counterbalancing that of the priestly class.

4 The classical materialistic philosophy

From at least as far back as the sixth century AD onwards, Lokāyata is referred to as a materialistic philosophical school. Its followers, the Lokāyatas or Lokāyatikas/Laukāyatikas, are also called Bārhaspatyas (followers of Bṛhaspati) and Cārvākas. These terms seem to apply only to the followers, not to the school itself. Some modern authors assume a subdivision into 'well-instructed' (*suśikṣita*) and 'cunning' (*dhūrta*) Cārvākas; however, both appellations refer mockingly to the Cārvāka philosopher Udbhaṭa. The designation *nāstika* (nihilist, one who says '…does not exist') is also applied to the materialists. This broader term is used for those who deny life after death, the efficacy of sacrifices and moral deeds, and so on, as well as for those who do not accept the authority of the Veda. The latter usage includes not only materialists, but also Buddhists and Jainas.

Ontology. According to the *Bṛhaspatisūtra*, there are four 'great' elements, earth, water, fire and wind, although some Lokāyatikas accepted space/'ether' as a fifth element (see Matter, Indian conceptions of §1). Certain Lokāyatikas may have admitted the existence of atoms; the majority, however, denied both atoms and space because they are not perceptible. The world in all its diversity is only the result of various combinations of the material elements. There is no determinative principle, such as God or karma, which is responsible for the properties of things. They are due to their own nature; no agent makes fire hot or water cool. Lokāyata causality operates with material causes only, and efficient causes are not recognized (see Causation, Indian theories of §2).

The theory of elements formed the basis for various Lokāyata doctrines of the arising of consciousness. As stated by the *Bṛhaspatisūtra*, consciousness arises from the elements just as the power of intoxication arises from molasses and other substances when a fermenting agent is added. In other words, when certain material substances are mixed something new emerges, be it consciousness or the power of intoxication, that was not there before and could not be produced by these substances severally. The mixture conducive to the production of consciousness is obtained when the elements are transformed into the form of a body. Later Lokāyatikas, such as Kambalāśvatara, refined this theory by assuming further causal factors that support the body in the production of consciousness, notably the vital breaths and the senses. Strong criticism finally led some Lokāyatikas, probably including Udbhaṭa, to abandon the doctrine that consciousness arises from matter and to conceive it as a different entity. Consequently they admitted that consciousness arises only from consciousness and for the sake of, rather than from, the elements; this reverses the relationship between matter and consciousness as conceived in Sāōkhya, according to which primordial matter evolves for the sake of consciousness (see Sāōkhya §§2, 6). However, these Lokāyatikas did not accept any doctrine of rebirth, but claimed that the consciousness of the new-born arises from that of the parents.

Lokāyata ontology seems to be largely subordinated to the school's ethical agenda. The main aim of all theories of elements and consciousness is to deny rebirth and thereby to destroy the cornerstone of Brahmanical, Buddhist and Jaina socio-religious and ethical ideals that presuppose karmic retribution over many lives (see Karma and rebirth, Indian conceptions of).

Epistemology. Indian epistemological doctrines centre on the pivotal concept of means of valid cognition (*pramāṇa*) (see Epistemology, Indian schools of §2; Knowledge, Indian views of §§1-2). Lokāyata claims that there is only one means of valid cognition: sensory perception. Other means, such as inference and verbal testimony, especially Scriptures, are considered invalid. Undoubtedly their rejection has the purpose of refuting the foundation of social and ethical doctrines in any divine or supernatural power.

The inherent difficulties in this initial Lokāyata position were not immediately apparent. However, once epistemological issues became prominent in Indian philosophy (*circa* fifth century AD), the Lokāyatikas found themselves embarrassed over the justification of their single means of valid cognition. As all other means are denied, perception has to establish itself, which would involve a *petitio principii*, or not be established at all.

Further, it became clear that the theory of four elements could not be established by perception alone; also, the denial of an eternal soul, karma and rebirth, or in fact any denial, cannot be arrived at without some form of reasoning.

At least four responses can be discerned to these challenges. Some Lokāyatikas retained the old doctrine, but developed new arguments against inference. By self-destructive logic they tried to prove that all inferences are invalid. For instance, what is the object of the paradigmatic inference from smoke to fire? If it is the universal 'fire-ness', the inference proves what is already proved, since the universal 'fire-ness' was already known before. If, on the other hand, a particular fire is to be proved, the inference fails because the required concomitance does not obtain between the universal 'smoke-ness' and a particular fire. Another response was to grant inference limited validity to avoid the difficulties of the initial position. A philosopher called Purandara claimed that the Cārvākas also admit inferences, but only those that are well known in everyday practice, such as the one from smoke to fire. Inferences meant to establish nonperceptible entities like God or a soul are rejected. To justify only limited use of inference, Purandara emphasized that inference is not an independent means, but depends on perception and therefore cannot transgress the scope of perception. A third response led to the acknowledgement of the indefensibility of perception and a coherent sceptical position. Jayarāśi, whose *Tattvopaplavasiṃha* is the only surviving Lokāyata text, preferred to give up perception rather than to admit the validity of inference. Epistemologically this position is irrefutable, but a high price has to be paid for it. If no means of valid cognition is accepted, the Lokāyata theories of elements and consciousness can no longer be maintained. Jayarāśi accepted this consequence and reinterpreted these doctrines of the *Bṛhaspatisūtra* as the opinions of an opponent that have to be rejected in the final analysis. As an anonymous fragment puts it, the sole purpose of the *sūtras* of Bṛhaspati is to question the opinions of others, and not to establish anything. As a fourth response, Udbhaṭa claimed that the number and characteristics of means of valid cognition cannot be determined. To substantiate his view he adduced examples of knowledge which cannot be derived from any of the recognized means of valid cognition.

Ethics. We are better informed about what Lokāyata condemns than what it prescribes. Many of the reported statements are directed against the efficacy of sacrifices. For instance, if the sacrificed animal goes to heaven, why does the sacrificer not put his own father in its place? The entire Vedic ritual is but a scheme of the Brahmans to make an easy living. Further statements on ethical issues are equally based on the denial of what is imperceptible, notably a transmigrating entity different from the body, an afterlife and the results of one's moral deeds. Thus, among the four 'aims of life' the Lokāyatikas admit only one: pleasure. Sometimes they also consider wealth, as a means to pleasure, to be a legitimate goal; however, the religious and soteriological aims, *dharma* and liberation (*mokṣa*), are always denied (see Duty and virtue, Indian conceptions of §§1-3). No merit or demerit accrues from one's actions; there is neither heaven nor hell and death is the ultimate release. Accordingly, the Lokāyatikas are often described as hedonists. As long as one lives, they say, one should live happily. One should eat meat, drink spirits and indulge in sexual pleasures. That people have an unequal share of pleasure and pain is not due to any unseen force like karma, but to the different capacities of things caused by different combinations of the elements, just as bubbles on the ocean display a diversity of size, hue and duration.

ELI FRANCO
KARIN PREISENDANZ

References and further reading

Chattopadhyaya, D.P. in collaboration with Gangopadhyaya, M.K. (eds) (1990) *Cārvāka/Lokāyata*, New Delhi: Indian Council of Philosophical Research, in association with Ṛddhi-India, Calcutta.(Although most of the materials published here have been published before, this convenient source book contains a good selection of translations concerning classical Indian materialism, as well as studies by modern scholars.)

Frauwallner, E. (1956) *Geschichte der indischen Philosophie (History of Indian Philosophy)*, Reihe Wort und Antwort 6/II, Salzburg: Otto Müller Verlag, 295-309. (A clear and careful presentation of early materialism; contains a summary of the story of King Paesi.)

Jayarāśi (*c*.800) *Tattvopaplavasiṃha*, first half trans. E. Franco, *Perception, Knowledge and Disbelief: A Study of Jayarāśi's Scepticism*, Wiesbaden: Franz Steiner, 1987; Delhi: Motilal Banarsidass, 2nd edn, 1994.(Contains parallel Sanskrit text and English translation of the first half of the *Tattvopaplavasiṃha* with copious notes.)

Jayatilleke, K.N. (1963) *Early Buddhist Theory of Knowledge*, London: George Allen & Unwin; repr. Delhi: Motilal Banarsidass, 1980.(Chapter 1 contains a good discussion of early Lokāyata, especially as referred to in the Pāli canon, but also in later texts such as the *Laōkāvatārasūtra*. Sharp and justified criticism of previous

scholarship, but uncritical acceptance of the interpretation of Lokāyata by late Pāli commentaries as 'the science of casuistry' (*vitaṇḍa-sattha*).)

MacQueen, G. (1988) *A Study of the Śrāmaṇyaphala-sūtra*, Wiesbaden: Otto Harrassowitz.(Remarkable for its methodology, this study contains a convincing attempt to determine which heretical views belong to which heretical teachers, notably Ajita Keśakambala, on the basis of Pāli, Tibetan and Chinese versions.)

Namai, M. (1976) 'A Survey of Bārhaspatya Philosophy', *Indological Review* 2: 29-74; revised and enlarged in Chishō Namai, *Rinne no ronshō (The Establishment of Rebirth)*, Osaka: Tōhō Shuppan, 1996, 561-5, 1-52. (Contains the best compilation to date of *Bṛhaspatisūtra* and other Lokāyata fragments. *Rinne no ronshō* includes further articles, mainly in Japanese, on the controversy between Buddhists and materialists about rebirth; Chishō is Professor Namai's monastic name.)

Purandara (*c.*7th/8th century) *Paurandarasūtra*, trans. E. Franco, 'Paurandarasūtra', in M.A. Dhaky (ed.) *Pt Dalsukhbhai Malvania Felicitation Volume*, Aspects of Jainology 3, Varanasi: P.V. Research Institute, 1991, vol. 1: 154-63.(Only two fragments of this work of uncertain title survive. Franco's essay contains a new interpretation of them on the basis of Jaina sources. For a different interpretation, compare Solomon's study of Udbhaṭa.)

Solomon, E. (1977-8) 'Bhaṭṭa Udbhaṭa', *Annals of the Bhandarkar Oriental Research Institute*, 58-9: 985-92.(A reliable study of one of the most original Lokāyata philosophers.)

Mathematics, foundations of

Conceived of philosophically, the foundations of mathematics concern various metaphysical and epistemological problems raised by mathematical practice, its results and applications. Most of these problems are of ancient vintage; two, in particular, have been of perennial concern. These are its richness of content and its necessity. Important too, though not so prominent in the history of the subject, is the problem of application, or how to account for the fact that mathematics has given rise to such an extensive, important and varied body of applications in other disciplines.

The Greeks struggled with these questions. So, too, did various medieval and modern thinkers. The ideas of many of these continue to influence foundational thinking to the present day.

During the nineteenth and twentieth centuries, however, the most influential ideas have been those of Kant. In one way or another and to a greater or lesser extent, the main currents of foundational thinking during this period - the most active and fertile period in the entire history of the subject - are nearly all attempts to reconcile Kant's foundational ideas with various later developments in mathematics and logic.

These developments include, chiefly, the nineteenth-century discovery of non-Euclidean geometries, the vigorous development of mathematical logic, the development of rigorous axiomatizations of geometry, the arithmetization of analysis and the discovery (by Dedekind and Peano) of an axiomatization of arithmetic. The first is perhaps the most important. It led to widespread acceptance of the idea that space was not merely a Kantian 'form' of intuition, but had an independence from our intellect that made it different in kind from arithmetic. This asymmetry between geometry and arithmetic became a major premise of more than one of the main 'isms' of twentieth-century philosophy of mathematics. The intuitionists retained Kant's conception of arithmetic and took the same view of that part of geometry which could be reduced to arithmetic. The logicists maintained arithmetic to be 'analytic' but differed over their view of geometry. Hilbert's formalist view endorsed a greater part of Kant's conception.

The second development carried logic to a point well beyond where it had been in Kant's day and suggested that his views on the nature of mathematics were in part due to the relatively impoverished state of his logic. The third indicated that geometry could be completely formalized and that intuition was therefore not needed for the sake of conducting inferences within proofs. The fourth and fifth, finally, provided for the codification of a large part of classical mathematics - namely analysis and its neighbours - within a single axiomatic system - namely (second-order) arithmetic. This confirmed the views of those (for example, the intuitionists and the logicists) who believed that arithmetic had a special centrality within human thinking. It also provided a clear reductive target for such later anti-Kantian enterprises as Russell's logicism.

The major movements in the philosophy of mathematics during this period all drew strength from post-Kantian developments in mathematics and logic. Each, however, also encountered serious difficulties soon after gaining initial momentum. Frege's logicism was defeated by Russell's paradox; Russell's logicism, in turn, made use of such questionable (from a logicist standpoint) items as the axioms of infinity and reduction. Both logicism and Hilbert's formalist programme came under heavy attack from Gödel's incompleteness theorems. And finally, intuitionism suffered from its inability to produce a body of mathematics comparable in richness to classical mathematics.

Despite the failure of these non-Kantian programmes, however, movement away from Kant continued in the mid- and late twentieth century. From the 1930s on this has been driven mainly by a revival of empiricist and naturalist ideas in philosophy, prominent in the writings of both the logical empiricists and the later influential work of Quine, Putnam and Benacerraf. This continues as perhaps the major force shaping work in the philosophy of mathematics.

1 Kant's views; reactions

The '*Problematik*' that Kant established for the epistemology of pure mathematics focused on the reconciliation of two apparently incompatible features of pure mathematics: (1) the problem of necessity, or how to explain the apparent fact that mathematical statements (for example, statements such as that $1 + 1 = 2$ or that the sum of the interior angles of a Euclidean triangle is equal to two right angles) should appear to be not only true but

necessarily true and independent of empirical evidence; and (2) the problem of cognitive richness, or how to account for the fact that pure mathematics should yield subjects as rich and deep in content and method, as robust in growth and as replete with surprising discoveries as the history of mathematics demonstrates.

In mathematics, Kant said, we find a ‘great and established branch of knowledge’ - a cognitive domain so ‘wonderfully large’ and with promise of such ‘unlimited future extension’ that it would appear to arise from sources other than those of pure unaided (human) reason (1783: §§6, 7). At the same time, it carries with it a certainty or necessity that is typical of judgments of pure reason. The problem, then, is to explain these apparently conflicting characteristics. Kant’s explanation was that mathematical knowledge arises from certain standing conditions or ‘forms’ which shape our experience of space and time - forms which, though they are part of the innate cognitive apparatus that we bring to experience, none the less shape our experience in a way that goes beyond mere logical processing.

To elaborate this hypothesis, Kant sorted judgments/propositions in two different ways: first, according to whether they required appeal to sensory experience for their justification; and, second, according to whether their predicate concepts were ‘contained in’ their subject concepts. A judgment or proposition was ‘a priori’ if it could be justified without appeal to sensory content. If not, it was ‘a posteriori’. It was ‘analytic’ if the very act of *thinking* the subject concept contained, as a constituent part, the *thinking* of the predicate concept. If not, it was either false or ‘synthetic’. In synthetic a priori judgment - the type of judgment Kant regarded as characteristic of mathematics - the predicate concept was thought not through the mere *thinking* of the subject concept, but through its ‘construction in intuition’. He took a similar view of mathematical inference, believing it to involve an intuition that goes beyond the mere logical connection of premises and conclusions (1781/1787: A713-19/B741-7).

Kant erected his mathematical epistemology upon these distinctions and, famously, maintained that mathematical judgment and inference is synthetic a priori in character. In this way, he intended to account for both the necessity and cognitive richness of mathematics, its necessity reflecting its a priority, its cognitive richness its syntheticity.

Kant applied this basic outlook to both arithmetic and geometry (and also to pure mechanics). He did not regard them as entirely identical, however, since he saw them as resting on different a priori intuitions. Neither did he see them as possessing precisely the same universality (1781/1787: A163-5, 170-1, 717, 734/B204-6, 212, 745, 762). None the less, he regarded their similarities as more important than their differences and therefore took them to be of essentially the same epistemic type - namely, synthetic a priori. In the end, it was this inclusion of geometry and arithmetic within the same basic epistemic type rather than his more central claim concerning the *existence* of synthetic a priori knowledge that gave rise to the sternest challenges to his views.

In the decades following the publication of the first *Critique* (1781/1787), the principal source of concern regarding its views was the growing evidence for and eventual discovery of non-Euclidean geometries. This led many to question whether geometry and arithmetic are of the same basic epistemic character.

The serious possibility of non-Euclidean geometries went back to the work of Lambert and others in the eighteenth century. Building on this work, some - in particular, Gauss (1817, 1829) - stated their opposition to Kant’s views even before the actual discovery of non-Euclidean geometries by Bolyai and Lobachevskii in the 1820s. Gauss’ reasoning was essentially this: number seems to be purely a product of the intellect and, so, something of which we can have purely a priori knowledge. Space, on the other hand, seems to have a reality external to our minds that prohibits a purely a priori knowledge of it. Arithmetic and geometry are therefore not on an epistemological par with one another.

This reasoning became a potent force shaping nineteenth- and twentieth-century foundational thinking. Another such force was the dramatic development of logic and the axiomatic method in the mid- to late nineteenth century and early twentieth century. This included the introduction of algebraic methods by Boole and De Morgan, the improved treatment of relations by Peirce, Schröder and Peano, the replacement of the subject-predicate conception of propositional form with Frege’s more fecund functional conception, and the advances in axiomatization and formalization brought about by the work of Frege, Pasch, Peano, Hilbert and (especially) Whitehead and Russell.

Certain developments in mathematics proper also exerted an influence. Chief among these were the arithmetization

of analysis by Weierstrass, Dedekind and others, and the axiomatization of arithmetic by Peano and Dedekind. Of somewhat lesser importance, though still significant for their effects on Hilbert's thinking, were Einstein's relativistic ideas in physics.

2 Intuitionism

A variety of views concerning the asymmetry of geometry and arithmetic emerged in the late nineteenth and early twentieth centuries. That of the early intuitionists Brouwer and Weyl retained Kant's synthetic a priori conception of arithmetic.

They responded to the discovery of non-Euclidean geometries, however, by denying the a priori status of that part of geometry that could not be reduced to arithmetic by such means as Descartes' calculus of coordinates. They retained, none the less, a type of a priori intuition of time as the basis for arithmetical knowledge (see Brouwer 1913: 127-8). They also emphasized the synthetic character of arithmetical judgment and inference, and sharply distinguished them from logical judgment and inference.

Brouwer described his intuition of time as consciousness of change *per se* - the human subject's primordial inner awareness of the 'falling apart' of a life-moment into a part that is passing away and a part that is becoming. He believed that, via a process of abstraction, one could pass from this basal intuition of time to a concept of 'bare two-oneness', and from this concept to, first, the finite ordinals, then the transfinite ordinals and, finally, the linear continuum. In this way, parts of classical arithmetic, analysis and set theory could be recaptured intuitionistically. (See Brouwer 1907: 61, 97; 1913: 127, 131-2.)

Brouwer thus modified Kant's intuitional basis for mathematics. He also modified his conception of knowledge of existence. Kant believed that humans could obtain knowledge of existence only through sensible intuition. Only this, he believed, had the type of involuntariness and objectivity that assures us that belief in an object is not a mere compulsion or idiosyncrasy of our subjective selves. Like the post-Kantian romantic idealists, however, Brouwer (and Weyl, too) believed as well in knowledge of existence via a kind of 'intellectual intuition' - an intuition carried by a purely internal type of mental construction (1907: 96-7).

The early intuitionists (especially Brouwer and Poincaré) remained Kantian in their conception of mathematical reasoning and took it to be essentially different in character from 'discursive' or logical reasoning. Brouwer believed logical reasoning to mark not patterns in mathematical thinking itself but only patterns in its linguistic representation. It was therefore not indicative of the inferential structure of mathematical thinking itself and had no place within genuine mathematical reasoning *per se*. This was essentially the idea expressed in Brouwer's so-called 'First Act of Intuitionism' (1905: 2, 1981: 4-5).

Thus the early intuitionists (especially Brouwer and Weyl and, to some extent, Poincaré) discarded Kant's view of geometry, revised his conception of arithmetic and existence claims, and preserved his basic stance on the nature of mathematical reasoning and its relationship to logical reasoning. Later intuitionists (for example, Heyting and Dummett) did not keep to this plan. They rejected Brouwer's view of the divide between logical and mathematical reasoning and made a significant place for logic in their accounts of mathematical reasoning. Some of them (Dummett and his 'anti-realist' followers) even went so far as to make the question 'What is the logic of mathematical reasoning?' central to their philosophy of mathematics (see §5 below).

3 Logicism

The view of the logicist Frege (and, to some extent, of Dedekind) accepted Kant's synthetic a priori conception of geometry but maintained arithmetic to be analytic. Russell, another logicist, rejected Kant's views of both geometry and arithmetic (and also of pure mechanics) and maintained the analyticity of both. (See Logicism.)

Frege's logicism differed sharply from intuitionism. First, it differed in the place in mathematical reasoning it assigned to logic. Frege (1884: preface, III-IV) maintained that reasoning is essentially the same everywhere and that even an inference pattern such as that of mathematical induction, which appears to be peculiar to mathematics, is, at bottom, purely logical. Second, it differed in its conception of geometry. Like the early intuitionists, Frege regarded the discovery of non-Euclidean geometries as revealing an important asymmetry between arithmetic and geometry. Unlike them, however, he did not see this as grounds for rejecting Kant's synthetic a priori conception

of geometry (1873: 3; 1884: §89), but rather as indicating a fundamental difference between geometry and arithmetic. Frege believed the fundamental concept of arithmetic - magnitude - to be both too pervasive and too abstract to be the product of Kantian intuition (1874: 50). It figured in *every* kind of thinking and so must, he reasoned, have a basis in thought deeper than that of intuition. It must have its basis in the very core of rational thought itself; the laws of logic.

The problem was to account for the cognitive richness of arithmetic on such a view. How could the ‘great tree of the science of number’ (1884: §16) have its roots in bare logical or analytical ‘identities’? Frege responded by offering new accounts of both the objectivity and the informativeness of arithmetic. The former he attributed to its subject matter - the so-called ‘logical objects’ (§§26, 27, 105). The latter he derived from a new theory of content which allowed concepts to contain (tacit) content that was not needed for their grasp. On this view, analytic judgments could have content that was not required for the mere understanding of the concepts contained in them. Consequently, they could yield more than knowledge of transparent logical identities (§§64-66, 70, 88, 91).

Unlike Kant, then, Frege maintained an important epistemic asymmetry between geometry and arithmetic - an asymmetry based upon his belief that arithmetic is more basic to human rational thought than is geometry. In addition, he departed from Kant in maintaining a realistic conception of arithmetic knowledge despite its analytic character. He saw it as being about a class of objects - so-called ‘logical objects’ - that are external but intimately related to the mind and therefore not the mere expression of standing traits of human cognition. The differences between arithmetical and geometric necessity were to be accounted for by separating the relationship the mind has to the objects of arithmetic from that which it has to the objects of geometry.

Russell’s logicism differed from Frege’s. Perhaps most importantly, Russell did not regard the existence of non-Euclidean geometries as evidence of an epistemological asymmetry between geometry and arithmetic. Rather, he saw the ‘arithmetization’ of geometry and other areas of mathematics as evidence of an epistemological symmetry between arithmetic and the rest of mathematics. Russell thus extended his logicism to the whole of mathematics. The basic components of his logicism were a general methodological ideal of pursuing each science to its greatest level of generality, and a conception of the greatest level of generality in mathematics as lying at that point where all its theorems are of the form ‘*p* implies *q*’, all their constants are logical constants and all their variables of unrestricted range. Theorems of this sort, Russell maintained, would rightly be regarded as logical truths.

Russell’s logicism was thus motivated by a view of mathematics which saw it as the science of the most general formal truths; a science whose indefinables are those constants of rational thought (the so-called logical constants) that have the most ubiquitous application, and whose indemonstrables are those propositions that set out the basic properties of these indefinables (Russell 1903: 8). In his opinion, such a view provided the only precise description of what philosophers have had in mind when they have described mathematics as a necessary or a priori science.

Russell thus accounted for the necessity of mathematics by pointing to its logical character. He accounted for its richness principally by invoking a new definition of syntheticity that allowed all but the most trivial logical truths and inferences to be counted as informative or synthetic. Mathematical truths would thus be logical truths, but they would not, for all that, be analytic truths. Similarly for inferences. An inference would count as synthetic so long as its conclusion was a different proposition from its premises. Cognitive richness was conceived primarily as the production of new propositions from old, and, on Russell’s conception (supposing the criterion of propositional identity to be sufficiently strict), even purely logical inference could produce a bounty of new knowledge from old.

Russell was thus able to account for both the necessity and the cognitive richness of mathematics while making mathematics part of logic. What had kept previous generations of thinkers and, in particular, Kant from recognizing the possibilities of such a view was the relatively impoverished state of logic before the end of the nineteenth century. The new logic, with its robust stock of new forms, its functional conception of the proposition and the ensuing fuller axiomatization of mathematics which it made possible, changed all this forever and provided for the final refutation of Kant. Such, at any rate, was Russell’s position.

4 Hilbert’s formalism

Hilbert accepted the synthetic a priori character of (much of) arithmetic and geometry, but rejected Kant’s account of the supposed intuitions upon which they rest. Overall, Hilbert’s position was more complicated in its

relationship to Kant's epistemology than were those of the intuitionists and logicists. Like Russell, he rejected Kant's specifically mathematical epistemology - in particular, his conception of the nature and origins of its a priori character. Like Russell, too, he rejected the common post-Kantian belief in the epistemological asymmetry of arithmetic and geometry. Hilbert was, however, unique among those mentioned here in endorsing the framework of Kant's general critical epistemology and making it a central feature of his mathematical epistemology. Specifically, he adopted Kant's distinction between the faculty of the understanding and the faculty of reason as the guide for his pivotal distinction between the so-called 'real' and 'ideal' portions of classical mathematics (Hilbert 1926: 376-7, 392).

Hilbert took 'real' mathematics to be ultimately concerned with the shapes or forms (*Gestalten*) of concrete signs or figures, given in intuition and comprising a type of 'immediate experience prior to all thought' (1926: 376-7; [1928] 1967: 464-5). Hilbert proposed this basic intuition of shape as a replacement for Kant's two a priori intuitions of space and time. Like Kant's a priori intuitions, however, Hilbert, too regarded his finitary intuition as an 'irremissible pre-condition' of all mathematical (indeed, all scientific) judgment and the ultimate source of all genuine a priori knowledge (1930: 383, 385).

The genuine judgments of real mathematics were the judgments of which our mathematical knowledge was constituted. The pseudo-judgments of ideal mathematics, on the other hand, functioned like Kant's ideas of reason. They neither described things present in the world nor constituted a foundation for our judgments concerning such things. Rather, they played a purely regulative role of guiding the efficient and orderly development of our real knowledge.

Hilbert did not, therefore, affirm the necessity of either arithmetic or geometry in any simple, straightforward way. Rather, he distinguished two types of necessity operating within both. One, pertaining to the judgments of real mathematics, consisted in the (presumed) fact that the apprehension of certain elementary spatial and combinatorial features of simple concrete objects is a pre-condition of all scientific thought. The other, pertaining to the ideal parts of mathematics, had a kind of psychological necessity, a necessity borne of the manner in which our minds inevitably or best regulate the development of our real knowledge.

This conception of the necessity of mathematics was different from both Kant's and the logicists' and intuitionists'. So, too, was Hilbert's view of the cognitive richness of mathematics, which he attributed both to the objective richness of the shapes and combinatorial features of concrete signs and to the richness of our imaginations in 'creating' complementary ideal objects.

In its overall structure, Hilbert's mathematical epistemology thus resembled Kant's *general* critical epistemology. This included his so-called 'consistency' requirement (that is, the requirement that ideal reasoning not prove anything contrary to that which may be established by real means), which resembled Kant's demand that the faculty of reason not produce any judgment of the understanding that could not in principle be obtained solely from the understanding (1781/1787: A328/B385).

5 Modifications

During the first four decades of the twentieth century, each of the post-Kantian programmes outlined above came under attack. Frege's logicism was challenged by Russell's paradox. Russell's logicism encountered difficulties concerning its use of certain existence axioms (namely his axioms of reducibility and infinity) which did not appear to be laws of *logic*. Both were challenged by Gödel's incompleteness theorems, as was Hilbert's formalist programme. Finally, the intuitionists were criticized both philosophically, where their idealism was called into question, and mathematically, where their ability to support a significant body of mathematics remained in doubt. Various modifications have been proposed.

Modifications of logicism. On the mathematical side, a chastened successor to logicism can perhaps be seen in the model-theoretic work of Abraham Robinson and his followers. They are interested in determining the mathematical content latent in purely 'logical' features of various mathematical structures or the extent to which genuinely mathematical problems concerning these structures can be solved by purely logical (that is, model-theoretic) means. They have been particularly successful in their treatment of various algebraic structures (see Macintyre 1977; Robinson 1979; Hodges 1993).

Philosophically, too, there have been attempts to renew logicism. It re-emerged in the 1930s and 1940s as the favoured philosophy of mathematics of the logical empiricists (see Carnap 1931; Hahn 1933). They did not, however, develop a logicism of their own in the way that Dedekind, Frege and Russell did, but, rather, simply appropriated the technical work of Russell and Whitehead (modulo the usual reservations concerning the axioms of infinity and reducibility) and attempted to embed it in an overall empiricist epistemology.

This empiricist turn was a novel development in the history of logicism and represented a serious departure from both the original logicism of Leibniz (§10) and the more recent logicism of Frege (and Dedekind). It was less at odds with Russell's logicism which saw mathematics and the empirical sciences as both making use of an essentially inductive method (the so-called 'regressive' method - see Russell 1906, 1907).

Like all empiricists, the logical empiricists struggled with the Kantian problem of how to account for the apparent necessity of mathematics while at the same time being able to explain its cognitive richness. Their strategy was to empty mathematics of all non-analytic content while, at the same time, arguing that analytic truth and inference can be substantial and non-self-evident.

Their ideas came under heavy attack by W.V. Quine, who challenged their pivotal distinction between analytic and synthetic truths (1951, 1954). He argued that the basic unit of knowledge - the basic item of our thought that is tested against experience - is science as a whole and that this depends upon empirical evidence for its justification. Mathematics and logic are used to relate empirical evidence to the rest of science and, so, are inextricably interwoven into the whole fabric of science. They are thus part of the total body of science that is tested against experience and there is no clean way of dividing between truths of meaning (analytic truths) and truths of fact (synthetic truths).

Within a relatively brief period of time, Quine's argument became a major influence in the philosophy of mathematics and the logicism of the logical empiricists was largely abandoned. Newer conceptions of logicism have, however, continued to appear from time to time. For example, Putnam (1967) addressed the difficult (for a logicist) question of existence claims, arguing that such statements are to be seen as asserting the possible (as opposed to the actual) existence of structures. They are therefore, at bottom, logical claims, and can be established by logical (or metalogical) means. Hodes (1984) takes a somewhat different approach, arguing that arithmetic claims can be translated into a second-order logic in which the second-order variables range over functions and concepts (as opposed to objects). In this way, commitment to sets and other specifically mathematical objects can be eliminated and, this done, arithmetic can be considered a part of logic.

Field (1980, 1984) also presents a kind of logicist view - namely, that mathematical knowledge is (at least largely) logical knowledge. Mathematical knowledge is defined as that knowledge which separates a person who knows a lot of mathematics from a person who knows only a little, and it is then argued that what separates these two kinds of knowers is mainly logical knowledge; that is, knowledge of what follows from what.

Modifications of Hilbert's programme. Hilbert's programme too has its latter-day adherents. For the most part, these have adopted one of two basic stances: that of extending the methods available for proving the consistency of classical ideal mathematics; or that of diminishing the scope and strength of classical ideal mathematics so that its consistency (or the consistency of important parts of it) can more nearly be proved by the kinds of elementary means that Hilbert originally envisaged.

Some in the first group (for example, Gentzen, Ackermann and Gödel) have argued that there are types of evidence that exceed finitary evidence in strength but which have the same basic epistemic virtues as it. Others (for example, Kreisel 1958; Feferman 1988; Sieg 1988) argue for a change in our conception of what a consistency proof ought to do. They maintain that its essential obligation is to realize an epistemic gain, and that finitary methods are not the only epistemically gainful methods for proving consistency.

Those in the second group - the so-called 'reverse mathematics' school of Friedman, Simpson and others - try to isolate the mathematical 'cores' of the various areas of classical mathematics and prove the consistency of these 'reduced' theories by finitary or related means. So far, significant success has been achieved along these lines. (See Hilbert's programme and formalism §4.)

Modifications of intuitionism. Regarding intuitionism, Heyting's work in the 1930s to formalize intuitionism and

to identify its logic has led to a vigorous programme of logical and mathematical investigation (see Heyting 1956; Troelstra 1973, 1977; Beeson 1985 for descriptions of some of this work). In addition to Heyting and his students, Errett Bishop and his followers have extended a constructivist approach to areas of classical mathematics to which such an approach had previously not been extended (see Bridges 1987 for a survey).

On the more philosophical side, the most important development is the construction by Michael Dummett and his anti-realist followers of a defence of intuitionism based upon - in their view - the best answer to the question ‘What is the logic of mathematics?’. Their answer is based upon what they take to be a proper theory of meaning - a theory which, following certain ideas set out by Wittgenstein in his *Philosophical Investigations*, equates the meaning of an expression with its canonical use in the practice to which it belongs. They then identify the canonical use of an expression in mathematics with the role it plays in the central activity of proof, and from this they infer an intuitionist treatment of the logical operators (Dummett 1973, 1977).

6 Later developments

Along with the modifications of the major post-Kantian viewpoints noted above, two other developments in the second half of the twentieth century are important to note. One of these is the shift towards empiricism that was brought about by Quine’s (following Duhem’s) merging of mathematics and the empirical sciences into a single justificatory unit governed by a basically inductive-empirical method. On this view, mathematics may on the whole be *less* susceptible to falsification by sensory evidence than is natural science, but this is a difference of degree, not kind.

This conception of mathematics dispenses with a ‘datum’ of mathematical epistemology that philosophers of mathematics from Kant on down had struggled to accommodate: namely, the presumed necessity of mathematics. It puts in its place a general empiricist epistemology in which all judgments - those of mathematics and logic as well as those of the natural sciences - are seen as evidentially connected to sensory phenomena and, so, subject to empirical revision.

To accommodate the lingering conviction that mathematics is independent of empirical evidence in a way that natural science is not, Quine introduced a pragmatic distinction between them. Rational belief-revision, he said, is governed by a pragmatic concern to maximize the overall predictive and explanatory power of one’s total system of beliefs. Furthermore, predictive and explanatory power are generally aided by policies of revision which minimize, both in scope and severity, the changes that are made to a previously successful belief-system in response to recalcitrant experience.

Because of this, beliefs of mathematics and logic are typically less subject to empirical revision than beliefs of natural science and common sense since revising them generally (albeit, in Quine’s view, not inevitably) does more damage to a belief-system than does revising its common sense and natural scientific beliefs. The necessity of mathematics is thus accommodated in Quine’s epistemology by moving mathematics closer to the centre of a ‘web’ of human beliefs where beliefs are less susceptible to empirical revision than are the beliefs of natural science and common sense that lie closer to the edge of the web.

In Quine’s view, merging mathematics and science into a single belief-system also induces a realist conception of mathematics. Mathematical sentences must be treated as true in order to play their role in this system, and the world is to be seen as being populated by those entities that are among the values of the variables of true sentences. Mathematical entities are thus real because mathematical sentences play an integral part in our best total theory of experience (see Quine 1948, 1951; Putnam 1971, 1975).

Quine’s views have been challenged on various grounds. For example, Parsons (1980) argues that treating the elementary arithmetical parts of mathematics as being on an epistemological par with the hypotheses of theoretical physics fails to capture an epistemologically important distinction between the different kinds of evidentness displayed by the two. Even highly confirmed physical hypotheses such as ‘The earth moves around the sun’ are more ‘derivative’ (that is, roughly, more theory-laden) than is an elementary arithmetic proposition such as ‘ $7 + 5 = 12$’. It is therefore not plausible to regard the two claims as based on essentially the same type of evidence.

Others have challenged different aspects of Quine’s position. Field (1980) and Maddy (1980), for example, both

question his merging of mathematics and natural science, though in different ways. Field argues that natural science that utilizes mathematics is a conservative extension of it and, so, has no need of its entities. The mathematical part of natural science can thus, in an important sense, be separated from the rest of it. (See Shapiro (1983) for an apt criticism of Field's arguments.). Maddy investigates the possibility that our knowledge of at least certain mathematical objects might not be so diffuse and inextricable from the whole scheme of our natural scientific knowledge as Quine suggests. She argues that perceptual experience can be tied closely and specifically to certain mathematical objects (in particular, to certain sets) in a way that seems out of keeping with Quine's holism.

In addition to Quine, others have suggested different mergings of mathematics and natural sciences. Kitcher (1983), for instance, presents a generally empiricist epistemology for mathematics in which history and community are important epistemological forces. Gödel, on the other hand, argued that mathematics, like the natural sciences, makes use of what is essentially inductive justification ([1947] 1964: 477, 485) when it justifies higher-level mathematical hypotheses on the grounds of their explanatory or simplificatory effects on lower-level mathematical truths and on physics. He allowed, however, that only *some* of our mathematical knowledge arises from empirical sources and regarded as absurd the idea that all of it might do (1951: 311-12).

Another important influence on recent philosophy of mathematics is Benacerraf's 'Mathematical Truth' (1973), in which he argues that the philosophy of mathematics faces a general dilemma. It must give an account of both mathematical truth and mathematical knowledge. The former seems to demand abstract objects as the referents of singular terms in mathematical discourse. The latter, on the other hand, seems to demand that we avoid such referents. There are mathematical epistemologies (for example, various Platonist ones) that allow for a plausible account of the truth of mathematical sentences. Likewise, there are those (for example, various formalist ones) that allow for a plausible account of how we might come to *know* mathematical sentences. However, no known epistemology does both. Towards the end of the twentieth century a great deal of work has been devoted to resolving this dilemma. Field (1980, 1984), Hellman (1989) and Chihara (1990) attempt anti-Platonist resolutions. Maddy (1990), on the other hand, attempts a resolution at once Platonist and naturalistic. To date there is no general consensus on which approaches are the more plausible.

An earlier argument of Benacerraf's (see Benacerraf 1965, but also Dedekind 1888: §73; Hilbert 1900; Weyl 1927; Bernays 1950) was similarly influential in shaping later work. It is the chief inspiration of the position known as 'structuralism' - the view that mathematical objects are essentially positions in structures and have no important additional internal composition or nature (see Resnik 1981, 1982; Shapiro 1983). Apart from the desire for a descriptively more adequate account of mathematics, the chief motivation of structuralism is epistemological. Knowledge of the characteristics of individual abstract objects would seem to require naturalistically inexplicable powers of cognition. Knowledge of at least some structures, on the other hand, would appear to be explicable as the result of applying such classically empiricist means of cognition as abstraction to observable physical complexes. Structures identified via abstraction become part of the general framework of our thinking and can be extended and generalized in a variety of ways as the search for the simplest and most effective overall conceptual scheme is pursued.

Structuralism as a general philosophy of mathematics has been criticized by Parsons (1990) who argues that there are important mathematical objects for which structuralism is not an adequate account. These are the 'quasi-concrete' objects of mathematics - objects that are directly 'instantiated' or 'represented' by concrete objects (for example, geometric figures and symbols such as the so-called 'stroke numerals' of Hilbert's finitary arithmetic, where these are construed as types whose instances are written marks or symbols or uttered sounds). Such objects cannot be treated in a purely structuralist way because their 'representational' function cannot be reduced to the purely intrastructural relationships they bear to other objects within a given system. At the same time, however, they are among the most elementary and important mathematical entities there are.

See also: Antirealism in the philosophy of mathematics; Constructivism in mathematics; Dedekind, J.W.R.; Kronecker, L.; Model theory; Proof theory; Realism in the philosophy of mathematics

MICHAEL DETLEFSEN

References and further reading

Ayer, A. (ed.) (1959) *Logical Positivism*, New York: Free Press.(Includes influential expressions of the positivists' conception of mathematics.)

Barwise, J. (ed.) (1977) *Handbook of Mathematical Logic*, Amsterdam: North Holland.(Useful collection of survey articles on many areas of mathematical logic and foundational studies.)

Beeson, M.J. (1985) *Foundations of Constructive Mathematics*, Berlin: Springer.(Useful survey of recent logical and mathematical developments in constructive mathematics.)

Benacerraf, P. (1965) 'What Numbers Could Not Be', *Philosophical Review* 74: 47-73.(Structuralist argument that arithmetic is not a science concerning particular objects - the numbers - but rather a science elaborating the structure that all arithmetical progressions share in common. Dedekind argued for such a view a century earlier.)

Benacerraf, P. (1973) 'Mathematical Truth', *Journal of Philosophy* 70: 661-80; repr. in Benacerraf and Putnam (1964), 403-20.(Argues that existing philosophies of mathematics are subject to a dilemma - they cannot give a satisfactory account of both mathematical truth and mathematical knowledge.)

Benacerraf, P. and Putnam, H. (eds) (1964) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.(Collection of influential papers in the philosophy of mathematics.)

Bernays, P. (1950) 'Mathematische Existenz und Widerspruchsfreiheit', repr. in *Abhandlungen zur Philosophie der Mathematik*, Darmstadt: Wissenschaftliche Buchgesellschaft, 1976.(A kind of structuralist argument.)

Bishop, E. (1967) *Foundations of Constructive Analysis*, New York: McGraw-Hill.(Influential development of analysis from a constructive vantage somewhat more restrictive than that of intuitionism.)

Bolyai, J. (1832) 'Science of Absolute Space', trans. G.B. Halsted, in R. Bonola (ed.) *Non-Euclidean Geometry: A Critical and Historical Survey of its Development*, New York: Dover, 1955.(Celebrated essay including the first published description of a consistent non-Euclidean geometry.)

Bridges, D. (1987) *Varieties of Constructive Mathematics*, Cambridge: Cambridge University Press. (Mathematical survey of recent developments in various constructivist approaches to mathematics. Requires some background and some knowledge of classical mathematics.)

Brouwer, L.E.J. (1905) *Leven, Kunst, en Mystiek*, Delft; excerpts trans. 'Life, Art and Mysticism', in *Collected Works*, vol. 1, ed. A. Heyting, Amsterdam: North Holland, 1975; unabridged trans. W. van Stigt, *Notre Dame Journal of Formal Logic* 37 (3), 1996.(Early statement of various of Brouwer's philosophical views.)

Brouwer, L.E.J. (1907) *Over de Grondslagen der Wiskunde*, Amsterdam and Leipzig; trans. 'On the Foundations of Mathematics', in *Collected Works*, vol. 1, ed. A. Heyting, Amsterdam: North Holland, 1975. (Brouwer's doctoral dissertation. First sustained presentation of his intuitionist philosophy.)

Brouwer, L.E.J. (1913) 'Intuitionism and Formalism', *Bulletin of the American Mathematical Society* 20: 81-96; repr. in Benacerraf and Putnam (1964), 77-89.(Attempt by Brouwer to distinguish his intuitionist position from a position he labels 'formalism'. Includes an early statement of his antipathy to logical reasoning.)

Brouwer, L.E.J. (1955) 'The Effect of Intuitionism on Classical Algebra of Logic', *Proceedings of the Royal Irish Academy* §A 57: 113-16; repr. in *Collected Works*, vol. 1, ed. A. Heyting, Amsterdam: North Holland.(Perhaps the best and clearest statement of his view separating mathematical from logical reasoning, including intuitionistically valid logical reasoning.)

Brouwer, L.E.J. (1981) *Brouwer's Cambridge Lectures on Intuitionism*, ed. D. van Dalen, Cambridge: Cambridge University Press.(Late statement of Brouwer's intuitionist views, taken from lectures given 1946-51.)

Carnap, R. (1931) 'Die logizistische Grundlegung der Mathematik', *Erkenntnis* 2: 91-105; trans. E. Putnam and G. Massey, 'The Logicist Foundations of Mathematics', in Benacerraf and Putnam (1964), 41-52.(Statement of the type of logicist view that became popular among the logical empiricists of the 1930s and 1940s.)

Chihara, C. (1990) *Constructibility and Mathematical Existence*, Oxford: Oxford University Press. (Argues that classical mathematics does not require commitment either to such linguistic objects as open sentences or to abstract objects (concepts) to which they might be taken to refer.)

Dedekind, R. (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; trans. W.W. Beman (1901), 'The Nature and Meaning of Numbers', in *Essays on the Theory of Numbers*, New York: Dover, 1963.(§73 includes an interesting statement of Dedekind's philosophical conception of the axiomatic method and its relationship to the older genetic method.)

Detlefsen, M. (1986) *Hilbert's Program: An Essay on Mathematical Instrumentalism*, Boston, MA, and Dordrecht: Reidel.(Examination of the arguments of Hilbert's programme using Gödel's incompleteness theorems, especially the second.)

Detlefsen, M. (1990a) 'On an Alleged Refutation of Hilbert's Program using Gödel's First Incompleteness Theorem', *Journal of Philosophical Logic* 19: 343-77.(Critical evaluation of proposed arguments against

Hilbert's programme using Gödel's first incompleteness theorem.)
Detlefsen, M. (1990b) 'Brouwerian Intuitionism', *Mind* 99: 501-34.(Investigation of Brouwer's view of logical reasoning and the distinction that he and Poincaré believed to exist between mathematical and logical reasoning.)
Detlefsen, M. (1996) 'Philosophy of Mathematics in the 20th Century', in *Routledge History of Philosophy*, vol. 9, *Philosophy of Science, Logic and Mathematics in the 20th Century*, London: Routledge.(Survey of the main currents of foundational thinking in the twentieth century.)
Dummett, M.A.E. (1973) 'The Philosophical Basis of Intuitionist Logic', in H.E. Rose and J.C. Shepherdson (eds) *Proceedings of the Logic Colloquium, Bristol, July 1973*, Amsterdam: North Holland, 1975, 5-40; repr. in *Truth and Other Enigmas*, London: Duckworth, 1978, 215-47; and in Benacerraf and Putnam (1964), 97-129. (Argues that an adequate account of the meanings of mathematical propositions reveals the logic of mathematics to be intuitionistic logic. The view of meaning proposed is like that of the later Wittgenstein in that it equates the meaning of a proposition with its canonical usage; takes the canonical usage of a sentence in mathematics to consist in the role it plays in giving proofs.)
Dummett, M.A.E. (1977) *Elements of Intuitionism*, Oxford: Clarendon Press, 1990.(Further development of the argument set out in Dummett (1973). Also gives a basic presentation of various notions of intuitionistic mathematics (for example, choice sequences and spreads) and intuitionistic logic and its metatheory.)
Feferman, S. (1988) 'Hilbert's Program Relativized: Proof-Theoretical and Foundational Reductions', *Journal of Symbolic Logic* 53: 364-84.(Argues that finitary methods are not the only epistemically gainful methods for proving consistency and that significant partial realizations of Hilbert's programme can be realized by exploiting this fact.)
Field, H. (1980) *Science Without Numbers*, Princeton, NJ: Princeton University Press.(Defence of the view that mathematical knowledge is, at least in large part, logical knowledge. Also argues, contra Quine and Putnam, that realism with respect to physics does not entail realism with respect to mathematics.)
Field, H. (1984) 'Is Mathematical Knowledge Just Logical Knowledge?', *Philosophical Review* 93: 509-52. (Argues that mathematical knowledge is predominantly logical in character.)
Frege, G. (1873) 'Über eine geometrische Darstellung der imaginären Gebilde in der Ebene', doctoral dissertation, University of Göttingen; repr. in I. Angelleli (ed.) *Kleine Schriften*, Hildesheim, 1967; partial trans. in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984.(An attempt to vindicate the ultimately Kantian view that, at bottom, geometric laws - even those concerning imaginary points, lines, and so on - are all based on intuition, by showing how to replace imaginary items with real, intuitable items.)
Frege, G. (1874) 'Rechnungsmethoden, die sich auf eine Erweiterung des Größenbegriffes gründen', *Habilitationsschrift*, University of Jena; trans. in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984.(Includes a statement of the basic precept of Frege's logicism; namely, that there can be no intuition of so pervasive and abstract a concept as that of magnitude or number.)
Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Pohle; trans. '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82.(Presentation of Frege's symbolic scheme for logic and mathematics. He hoped that its employment would foster rigour in the conduct of proof and thereby allow for a clearer view of the basic laws upon which arithmetic is founded.)
Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, and Evanston, IL: Northwestern University Press, 2nd edn, 1980.(First full-length presentation of Frege's logicist view of arithmetic; focuses on its philosophical basis.)
Frege, G. (1891) *Funktion und Begriff*, Jena: Pohle; trans. P.T. Geach, 'Function and Concept', in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984.(Basic presentation of Frege's functional conception of propositions.)
Gauss, K. (1817, 1829) letter to Olbers (1817), in *Briefwechsel mit H.W.M. Olbers*, Hildesheim: Olms, 1976; letter to Bessel (1929), in *Werke*, Leipzig: Teubner, 1863-1903, vol. 8, 200.(Includes statements of Gauss' view concerning the 'external' character of geometry versus the 'internal' character of arithmetic.)
Gillies, D. (1982) *Frege, Dedekind and Peano on the Foundations of Arithmetic*, Assen: Van Gorcum. (Useful discussion of some of the similarities and differences between Frege, Dedekind and Peano.)
Gödel, K. (1947) 'What is Cantor's Continuum Problem?', *American Mathematical Monthly* 54: 515-25; revised version repr. in Benacerraf and Putnam (1964).(Basic expression of Gödel's realist and non-empiricist views of

mathematics.)
Gödel, K. (1951) ‘Some Basic Theorems on the Foundations of Mathematics and Their Implications’, in *Collected Works*, vol. 3, ed. S. Feferman, J.W. Dawson Jr, W. Goldfarb, C.D. Parsons and R.M. Solovay, New York and Oxford: Oxford University Press, 1995.(Edited text of Gödel’s 1951 Josiah Gibbs Lecture. Argues for a Platonist conception and against a conventionalist conception of mathematics. Also argues for a non-mechanist view of the human mind. Useful introductory essay by George Boolos.)
Hahn, H. (1933) *Einheitswissenschaft*, vol. 2, *Logik, Mathematik und Naturerkennen*, ed. R. Carnap and H. Hahn; trans. ‘Logic, Mathematics and Knowledge of Nature’, in A. Ayer (ed.) *Logical Positivism*, New York: Free Press, 1959.(Influential statement of the type of logicist view of mathematics popular among the positivists.)
Heijenoort, J. van (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press.(Collection of basic papers in mathematical logic and the foundations of mathematics. Useful forewords and bibliography.)
Hellman, G. (1989) *Mathematics Without Numbers*, Oxford: Oxford University Press.(Detailed development of a non-Platonist account of mathematics.)
Heyting, A. (1930) ‘Die formale Regeln der intuitionistischen Logik’ (The Formal Rules of Intuitionistic Logic), *Sitzungsberichte der preußischen Akademie von Wissenschaften, physikalisch-mathematische Klasse* 42-56. (First symbolic formalization of intuitionist logic.)
Heyting, A. (1956) *Intuitionism: An Introduction*, Amsterdam: North Holland; 3rd revised edn, 1971. (Introduction to philosophical considerations concerning intuitionism and a survey of intuitionist formalizations of logic and various parts of mathematics.)
Hilbert, D. (1900) ‘Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris 1900’, *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse*: 253-97; trans. ’Mathematical Problems’, *Bulletin of the American Mathematical Society* 8: 437-79, 1902.(Interesting remarks concerning the nature of mathematical truth and the axiomatic method.)
Hilbert, D. (1926) ‘Über das Unendliche’, *Mathematische Annalen* 95: 161-90; trans. ‘On the Infinite’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967.(Mature statement of basic philosophical precepts of Hilbert’s programme. Emphasizes the relationship with Kant’s general critical epistemology.)
Hilbert, D. (1928) ‘Die Grundlagen der Mathematik’, *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 6: 65-85; trans. ‘The Foundations of Mathematics’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967.(Repeats some of the material covered in Hilbert (1926); also includes an important clarificatory remark concerning the character of Hilbert’s ‘formalism’ (1967: 475).)
Hilbert, D. (1930) ‘Naturerkennen und Logik’, *Die Naturwissenschaften* 18: 959-63; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer, 1935.(Perhaps the fullest statement of Hilbert’s philosophical ideas; develops his view of the a priori and refines his ‘formalism’.)
Hilbert, D. (1935) *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer; repr. New York: Chelsea, 1965. (Collection of some of the most important papers in Hilbert’s foundational corpus.)
Hodes, H. (1984) ‘Logicism and the Ontological Commitments of Arithmetic’, *Journal of Philosophy* 81: 123-49. (Argues that arithmetical claims can be translated into a second-order logic in which the second-order variables range over functions and concepts.)
Hodges, W. (1993) *Model Theory*, Cambridge: Cambridge University Press.(Thorough and up-to-date survey of model theory. Requires some background in logic.)
Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1929.(Includes the classic statement of Kant’s views in the philosophy of mathematics. The ‘A’ and ‘B’ prefixes to sources refer to pages of the 1781 and 1787 German editions respectively.)
Kant, I. (1783) *Prolegomena to any Future Metaphysics that Shall Come Forth as Scientific*, trans. L. White Beck, Indianapolis, IN, and New York: Bobbs-Merrill, 1950.(Includes an abbreviated statement of the views on the nature of mathematics given in Kant (1781/1787).)
Kitcher, P. (1983) *The Nature of Mathematical Knowledge*, Oxford: Oxford University Press. (Develops an empiricist conception of mathematics which emphasizes its similarities to natural science.)
Kreisel, G. (1958) ‘Hilbert’s Programme’, *Dialectica* 12: 346-72; revised version in Benacerraf and Putnam (1964).(Discussion of the possible significance of a Hilbert-type programme in light of Gödel’s theorems and other more recent work in proof theory.)
Lobachevskii, N.I. (1829) ‘Geometrische Untersuchungen zur Theorie der Parallellinien’, Berlin: Fincke, 1840;

trans. G.B. Halsted, ‘Geometrical Researches in the Theory of Parallels’, in R. Bonola (ed.) *Non-Euclidean Geometry: A Critical and Historical Survey of its Development*, New York: Dover, 1955.(Elaboration of Lobachevskii’s famous work in non-Euclidean geometry.)

Macintyre, A. (1977) ‘Model Completeness’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland.(Survey of basic results and methods linking Robinson’s notion of model completeness to various questions in algebra.)

Maddy, P. (1980) ‘Perception and Mathematical Intuition’, *Philosophical Review* 84: 163-96.(Argues that we can perceive certain kinds of sets.)

Maddy, P. (1990) *Realism in Mathematics*, Oxford: Oxford University Press.(Attempt to show how realism in mathematics can be combined with a naturalistic mathematical epistemology.)

Parsons, C. (1980) ‘Mathematical Intuition’, *Proceedings of the Aristotelian Society* 80: 145-68. (Criticism of Quine in which it is argued that the Quinean view cannot account for an important difference in the kind of evidentness displayed by elementary mathematical truths on the one hand, and hypotheses of theoretical physics on the other.)

Parsons, C. (1983) ‘Quine on the Philosophy of Mathematics’, in *Mathematics in Philosophy*, Ithaca, NY: Cornell University Press.(Critical examination of Quine’s philosophical ideas on mathematics.)

Parsons, C. (1990) ‘The Structuralist View of Mathematical Objects’, *Synthese* 84: 303-46.(Argues that structuralism does not offer an adequate account of certain very elementary mathematical objects, such as geometric figures.)

Peano, G. (1889) *Arithmetices principia, nova methodo exposita*, Turin: Bocca; partial trans. ‘The Principles of Arithmetic’, in J. van Heijenoort (1967), 83-97.(Early attempt by Peano to produce a symbolic formulation of arithmetic.)

Poincaré, H. (1902) *La Science et l’hypothèse*, Paris: Flammarion; trans. in G.B. Halsted (ed.) *The Foundations of Science*, Lancaster, PA: The Science Press, 1946.(Chapter 1 presents Poincaré’s basically Kantian conception of mathematical reasoning.)

Poincaré, H. (1905) *Le valeur de la science*, Paris: Flammarion; trans. *The Value of Science*, New York: Dover, 1958; and trans. in G.B. Halsted (ed.) *The Foundations of Science*, Lancaster, PA: The Science Press, 1946. (Part 1 presents Poincaré’s basically intuitionist view of arithmetic, his so-called conventionalist view of geometry and his Kantian conception of mathematical inference or reasoning.)

Poincaré, H. (1908) *Science et méthode*, Paris: Flammarion; trans. in G.B. Halsted (ed.) *The Foundations of Science*, Lancaster, PA: The Science Press, 1946.(Book 2 gives a statement of Poincaré’s view of logic and its place in mathematics.)

Putnam, H. (1967) ‘Mathematics Without Foundations’, *Journal of Philosophy* 64: 5-22.(Argues that existence claims can be seen as asserting the possible (as opposed to the actual) existence of structures and, therefore, at bottom, can be treated as logical claims.)

Putnam, H. (1971) *Philosophy of Logic*, New York: Harper.(Presentation of a basically empiricist, pragmatist view of logic and mathematics.)

Putnam, H. (1975) ‘What is Mathematical Truth?’, in *Mathematics, Matter and Method: Philosophical Papers*, vol. 1, Cambridge: Cambridge University Press.(Defends a realist, though not immaterialist, view of mathematics. Once again stresses the basic motif that mathematics and the natural sciences are not qualitatively different.)

Putnam, H. (1980) ‘Models and Reality’, *Journal of Symbolic Logic* 45 (3): 464-82; repr. in Benacerraf and Putnam (1964), 2nd edn, 421-44.(Argues against even moderate forms of metaphysical realism in mathematics and defends a verificationist view of meaning in mathematics.)

Quine, W.V. (1948) ‘On What There Is’, *Review of Metaphysics* 2: 21-38.(Argument for mathematical realism.)

Quine, W.V. (1951) ‘Two Dogmas of Empiricism’, repr. in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1953; repr. New York: Harper & Row, 1963.(Influential attack on the analytic/synthetic distinction and empiricist reductionism of the logical empiricists.)

Quine, W.V. (1954) ‘Carnap and Logical Truth’, repr. in Benacerraf and Putnam (1964).(Criticism of Carnap’s conception of logical truth.)

Resnik, M. (1981) ‘Mathematics as a Science of Patterns: Ontology and Reference’, *Noûs* 15: 529-50. (Formulates and defends a structuralist conception of meaning in mathematics.)

Resnik, M. (1982) ‘Mathematics as a Science of Patterns: Epistemology’, *Noûs* 16: 95-105.(Formulates a structuralist view of mathematical knowledge.)

Robinson, A. (1979) *Selected Papers of Abraham Robinson*, vol. 1, *Model Theory and Algebra*, ed. H.J. Keisler *et al.*, Amsterdam: North Holland.(Collection of basic papers of the Robinson school applying model theory to

algebra. See especially ‘On the Application of Symbolic Logic to Algebra’, the paper in which Robinson presented his ideas for applying logic to the resolution of problems in algebra.)

Russell, B.A.W. (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992.(Russell’s first full-length development of his logicist views.)

Russell, B.A.W. (1906) ‘Les paradoxes de la logique’, *Revue de métaphysique et de morale* 14: 627-50; trans. ‘On ‘‘Insolubilia’’ and their Solution by Symbolic Logic’, in *Essays in Analysis*, ed. D. Lackey, London: Allen & Unwin, 1973.(Reply to Poincaré’s views concerning the relationship between logic and mathematics. Also includes an early statement of the hypothetico-deductive conception of the justification of the basic laws of mathematics.)

Russell, B.A.W. (1907) ‘The Regressive Method of Discovering the Premises of Mathematics’, in *Essays in Analysis*, ed. D. Lackey, London: Allen & Unwin, 1973.(Useful statement of the methodological view underlying Russell’s logicism.)

Russell, B.A.W. (1919) *Introduction to Mathematical Philosophy*, London: Allen & Unwin; repr. London: Routledge.(Useful statement of Russell’s post-*Principia* views on logic and mathematics. Discusses some of the weaknesses of his logicist position.)

Russell, B.A.W. (1973) *Essays in Analysis*, ed. D. Lackey, London: Allen & Unwin.(Useful collection of Russell’s writings, some not previously published.)

Shapiro, S. (1983) ‘Mathematics and Reality’, *Philosophy of Science* 50: 523-48.(Statement and defence of a structuralist view of mathematics.)

Sieg, W. (1988) ‘Hilbert’s Program Sixty Years Later’, *Journal of Symbolic Logic* 53: 338-48. (Discussion of the progress and changes in Hilbert’s programme since its formulation in the 1920s.)

Simpson, S. (1988) ‘Partial Realization of Hilbert’s Program’, *Journal of Symbolic Logic* 53: 349-63. (Useful survey and basic exposition of the work of the so-called ‘reverse mathematics’ programme begun by H. Friedman. Argues that though Hilbert’s programme in its original form is refuted by G2, a significant portion of it can still be carried out.)

Troelstra, A. (ed.) (1973) *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, Lecture Notes in Mathematics, vol. 344, Berlin, Heidelberg and New York: Springer. (Metamathematical study of various intuitionist systems.)

Troelstra, A. (1977) ‘Proof Theory and Constructive Mathematics’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland.(Survey of proof-theoretic work on intuitionist systems.)

Weyl, H. (1927) ‘Philosophie der Mathematik und Naturwissenschaft’, in *Handbuch der Philosophie*, Munich: Oldenbourg; revised and expanded trans. *Philosophy of Mathematics and Natural Science*, Princeton, NJ: Princeton University Press, 1949.(Valuable collection of essays in the philosophy of mathematics by one of the twentieth century’s pre-eminent scientific thinkers.)

Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press, 2nd edn, 1925.(The classic symbolic formalization of logic and mathematics; a basis of much of the greatest work in mathematical logic and the foundations of mathematics in the twentieth century.)

Wittgenstein, L.J.J. (1939) *Wittgenstein’s Lectures on the Foundations of Mathematics: Cambridge 1939*, ed. C. Diamond, Chicago, IL: University of Chicago Press, 1976.(A very different view from the logicism of the *Tractatus*.)

Wittgenstein, L.J.J. (1956) *Remarks on the Foundations of Mathematics*, ed. G.H. von Wright, R. Rhees and G.E.M. Anscombe, trans. G.E.M. Anscombe, Oxford: Blackwell, and Cambridge, MA: Harvard University Press, 2nd edn, 1967, part 1, appendix 2, 54-63; 3rd edn, 1978.(Presents statements of the later Wittgenstein’s view that a proof of a mathematical theorem functions to logically exclude doubt through its establishment of a theorem as a norm governing a language game.)

Matter

Viewed as arising within the framework of a more general theory of substance, *philosophical treatments of matter have traditionally revolved around two issues: (1) The nature of matter: what are the distinguishing characteristics of matter or material substance(s) that define it and distinguish it from other substances, if any? (2) The problem of elements: do material things consist of elementary substances, or are there always further constituents? One possible view is that there is no fundamental level - that there are always further constituents, ingredients of ingredients. However, the view most often held by both philosophers and scientists has been that there are indeed fundamental* elements *out of which material things are made. Once this view is adopted, the question arises as to what they are and what properties distinguish them.*

These two issues were introduced, though only gradually, in ancient Greek philosophy. A significant turn came about in the seventeenth century, in which the work of Descartes and Newton led to a picture of matter as passive, inert and dead as opposed to minds and forces, both of which were conceived as being 'active'. Many philosophical problems and doctrines have been formulated in terms of this distinction. However, later developments in science, especially in the twentieth century, have brought about such profound changes that classical concepts of matter are no longer viable. These new developments profoundly alter the statements of philosophical doctrines and problems traditionally associated with matter.

1 Ancient and medieval conceptions of matter

In Aristotle's view, matter is not a substance or an element. His own term for matter was reserved for one aspect of individual substances - particular objects like Socrates - which he saw as having both 'matter' and 'form'. Matter has the potential to receive form; it is never found in the absence of form, being inseparable from it. Thus Aristotle's matter is a bare *capacity* to have properties; it itself does not have any. This is in contrast to the matter of the seventeenth-century philosophers and scientists, which did have essential properties. Thus, 'the problem of matter' as usually conceived in modern philosophy stems from the concept of substance, specifically from the version of that concept which asserts that there are elements. From the perspective of this later view, the problem of matter in Aristotle is the problem of whether individual substances have 'constituents of which they are made' - the problem of elements - and not the problem of the nature of the 'potential to receive form' (see Aristotle §§13-14).

Aristotle attributed to his pre-Socratic predecessors the problem of the ultimate elements of which all things (material substances) consist. He thus saw the Milesian philosophers as maintaining that there is one material substance out of which all things are made - water for Thales, 'the Indefinite' for Anaximander, air for Anaximenes. Aristotle's interpretation of pre-Socratic philosophy is, however, questionable. According to some scholars, the Milesians (at least) were interested in the stuff *from which* things emerge, as things in nature grow from seeds or human beings 'come from' their parents, rather than, as Aristotle saw them, searching for the elementary constituents *of which* things are made. The shift of attention to the problem of elements was likely a gradual one, reaching relative maturity only with Empedocles, the atomists and Aristotle.

In Aristotle's interpretation, the Milesians were *monists*, believing that there is one material substance out of which all things are made. Monism was rejected following Parmenides, who maintained that if there is only one fundamental substance, change is impossible - a consequence his successors could not swallow. Parmenides argued that 'what is', or being, cannot have come into being, cannot change, cannot pass out of existence, is absolutely full (continuous, homogeneous, and as densely packed as possible), and is therefore 'one'. For example, it could only be separated by either nonbeing or being; but since nonbeing is nothing, it cannot separate anything; therefore between any parts of being there can only be more of the same, being. Similarly, if being changed, it could change only into being, that is, itself, since nonbeing is nothing. In order to account for the existence of change, post-Parmenidean philosophers turned to *pluralism*, the doctrine that there is more than one fundamental stuff (see Reduction, problems of). The atomists, Leucippus and Democritus, are clearest on this point. They admitted that ultimate reality must have most of the characteristics attributed to it by Parmenides, changelessness and permanence in particular. But in order to admit the existence of change and motion, they argued that there are many ultimate Parmenidean realities - the atoms - and, further, that nonbeing has some minimal kind of existence.

They identified nonbeing with empty space, through which the intrinsically unchangeable atoms could move and congregate to form the entities of our world (see Atomism, ancient). Other pluralist philosophers responded to the Parmenidean arguments similarly, declaring ultimate reality to be intrinsically unchangeable, the objects of familiar experience not having ultimate existence. Empedocles held that the four elements of earth, air, fire and water, which alone were truly fundamental realities (neglecting for present purposes the agencies love and strife), had the Parmenidean character of being uncreated and indestructible. The entities we know correspond to Parmenidean nonbeing, formed by the intermingling of the four elements through their motions. Anaxagoras said that there are an infinite number of 'seeds', each of which contains a little of everything (flesh, gold, …); larger bodies are formed when one of the infinite components of the seeds becomes dominant.

Plato, in the *Timaeus*, equated his equivalent of matter, the receptacle, with nonbeing, and also referred to it as space. Only the changeless forms truly exist, copies of them being embodied in the receptacle, which, like Aristotle's potentiality, is a mere capacity to receive form. As to the elements, the *Timaeus* presented a geometrical version of atomism in which the four elements, together with the universe, were identified with the five regular solids (see Plato §16).

Aristotle adopted a version of Empedocles' four-element theory, but with three major changes. First, the elements, like other substances, have both matter, in the sense of potentiality, and form, specifically two of the fundamental forms which together constitute the essence of the particular element. Thus the element earth has the fundamental forms 'cold' and 'dry', water 'cold' and 'wet', air 'cold' and 'dry', and fire 'hot' and 'dry'. (Indeed, for Aristotle, each type of entity has a distinct 'essential form' which constitutes its essence, making it the kind of substance it is. Except for the interconvertible elements, different types or species are irreducibly distinct: there simply are different species.) Second, the elements are interconvertible (they were not for Empedocles) by change of form: one fundamental form can 'drive out' its opposite. For example, 'dry' can drive out and replace the 'wet' in water, converting water into air. Third, Aristotle connected the four-element theory with his view of the structure of the universe: the 'natural place' of the element earth is in the central core of the universe, that of water in a spherical layer above earth, of air and fire in layers successively above that.

The Stoics also adopted a version of the four-element theory, but emphasized a distinction, present in Aristotle but not generally emphasized by him, between *passive* and *active* elements. Earth and water are passive elements, unable to move or otherwise act unless acted upon. Air and fire are active elements, capable of self-motion. The analogies between life and the active elements were not missed by the Stoics; indeed, a mixture of the active elements, *pneuma*, was the principle of life and intelligence, and constituted the 'soul of the world' which held the universe together as a unified *cosmos*. Although the Stoics saw the active elements as otherwise on a par with the passive ones in being material, their doctrine presages the Cartesian-Leibnizian distinction of mind and matter (see Stoicism §3).

Aristotle rejected atomism on several grounds, including its dismissal of final causes in nature. Despite a revival of atomism by Lucretius, Aristotle's rejection of the doctrine persisted among his medieval followers. However, atomism was not completely absent in medieval philosophy, being explored particularly among Islamic thinkers who followed the discipline *kalam*. Early *kalam* thinkers developed an atomistic conception of the universe as being freely created by God, but their attempt to affirm God's omnipotence as a causal agent based on his ability to create and sustain matter were refuted by Islamic Aristotelians and later by Al-Ghazali (see Causality and necessity in Islamic thought §3; Islamic theology).

The alchemists were responsible for most of the theoretical and experimental study of matter in the Middle Ages. Although their views were obscure, some were at least consistent with an Aristotelian viewpoint, seeing the purpose of alchemy as bringing the potential of matter to its fulfilment in the perfection of the higher metals and gems. Indeed, some of the alchemists seem to have believed that the formation of metals and gems in the earth was a natural process of growth and maturation, and that, through arcane knowledge, this process could be reproduced in the microcosmic laboratory. Thus they show the persistence of the ancient idea that things 'grow' in and from the ground (see Alchemy).

2 Matter in early modern science and philosophy

According to the conception of matter introduced in the Scientific Revolution of the seventeenth century, it is not

the case that there are many distinct types of material substance, each with its own essential and irreducible nature; rather, there is only one universal type of matter, with universal essential properties or ‘attributes’. All material bodies (and, for some, all entities whatever) are to be understood in terms of the motion of this fundamental matter. According to this view, introduced or at least popularized by Descartes and followed by Newton, matter is a *passive substance*, incapable of inducing its own action. All action of matter must be caused by an external agent. For Descartes, this external agent must be a *contact force*, and specifically an impact (collision); bodies will not move unless impacted (see Descartes, R. §11). In addition to impact forces, Isaac Newton allowed forces acting at a distance to affect matter. For Newton, matter consists of ultimate particles, atoms, which are discrete, localized, inert, ‘solid, massy, hard, impenetrable, movable’ and extended. (Note that the concept of inertness has shifted since Aristotle: whereas for him a body is inert if and only if it is at rest, for Descartes and Newton both rest and uniform rectilinear motion count as ‘inertial’, changeable only by the action of an external agent.) Passivity, inertness, deadness were now among the defining characteristics of material substance in addition to spatial extendedness, mobility and (Descartes excepted) impenetrability. Descartes distinguished such material substance sharply from *mind*, or *mental substance*, the latter being nonspatial and capable of initiating its own activity.

The basic difference between Descartes and Newton regarding matter was that Descartes conceived matter to be continuous and identified it with space, whereas Newton saw atoms as contingent occupants of space and consisting of physically (rather than logically) indivisible atoms occupying only a small proportion of the total space available in the universe. (In many places where he talks about the actual evolution of the universe and about particular physical problems, however, Descartes writes as though matter is a space-filling plenum rather than being identical with space.) Nevertheless, they agreed that it is of the nature of matter to persist in its inertial state of rest or uniform motion unless acted upon by an external force: that it is inert, dead, passive, blindly continuing in its present state, acting only when acted upon. Although Newton sometimes admitted action-at-a-distance (he also sometimes did not), he saw the power of attracting other bodies from a distance as contingently superimposed by God on matter, not part of the essential nature of matter, which was only passive. For most seventeenth-century thinkers, whether atomists or plenists, the basis of action in the universe was the impenetrability or hardness of matter, and action always took place by contact (see Mechanics, classical §4).

Leibniz criticized the Cartesian-Newtonian concept of matter on several grounds. He argued that the merely passive, whether space, time or matter, whether atomistic, plenum, purely space, cannot account for the unity of things, their individuation, change, resistance or impenetrability, or action. For example, the purely passive is incapable of explaining how one thing can affect another. Nor, since impenetrability is itself an activity of resistance, can the notion of a passive matter explain how objects can be affected by other objects. True substances - true fundamental realities - must be capable of action. Space, time and matter, being incapable of action, can only be ‘well-founded phenomena’, epiphenomena of non-material Active Forces or monads, which in their activity are more analogous to minds, capable of initiating their own activity, than to dead matter (see Leibniz, G.W. §11).

The issues raised by the ancient Greeks concerning the nature of matter and change are still present in these disputes between Leibniz and his opponents. The Parmenidean view of ultimate reality as necessarily unchanging is echoed in early modern theories of matter: where matter, whether atomistic or a continuous and space-filling plenum, was conceived of as permanent, indestructible and incapable of action. This view was challenged by the doctrine that the fundamentally real is capable of activity, a view with roots in Anaximenes and Heraclitus, later maintained by the Stoics, and carried by Leibniz to the extreme of asserting that only the active could exist.

3 Modern scientific views of matter and their philosophical relevance

In so far as the problem of matter is concerned with that of elementary constituents, an important development came in the history of chemistry. An empirically applicable concept of ‘element’ was utilized by Lavoisier: a material was to be treated as an element if: (a) with any of the means available to us, we cannot break it down into further constituents, (b) through such experiments, we always find members of the same class of substances as constituents, and (c) we find that other substances can be reconstructed by combinations of one or more of these ultimate breakdown products. Subsequent important steps in the history of chemistry revealed further evidence that there are a small number of such elements and that each element is fundamentally atomic. In the nineteenth century and culminating in the quantum-mechanical theory of the chemical bond, these developments led to a detailed theory of how atoms hold together in molecules and larger configurations (see Chemistry, philosophical

aspects of §4). However, it is now understood that the means which Lavoisier and his nineteenth-century successors used to break down substances (for example, heat and electricity) are low-energy processes, to be distinguished as ‘chemical’ from the forces that hold atomic nuclei themselves together (‘nuclear’ forces). Thus chemical atoms are now understood to be composed of more fundamental particles, electrons, protons and neutrons, the latter two consisting of still more basic particles, quarks. Although there remain dissenters, the view of most present-day particle physicists is that quarks and leptons (a category which includes electrons) are truly fundamental, without further constituents.

Modern scientific studies of matter have introduced further departures from the Cartesian-Newtonian ‘classical’ conception which are of profound relevance to philosophical discussions. While the concept of the indestructibility and conservation of mass is central to classical theories of mechanics and chemistry, relativity theories recognize that mass varies with frame of reference, and with velocity as measured in that reference-frame, and, further, that mass is interconvertible with energy (see Relativity theory, philosophical significance of). While Newton sharply distinguished forces from matter, in quantum mechanics forces are conveyed by transfer of a certain type of particles (the bosons) which has mass, like classical matter-particles. A second class of quantum-mechanical particles (‘fermions’) are intuitively closer analogues of classical matter in exhibiting impenetrability, but the correspondence is only rough. For example, fermions can also be exchanged in interactions, and they also exhibit wavelike (nonlocalized) behaviour. This latter feature is a consequence of deep features of quantum mechanics associated with the Heisenberg indeterminacy principles (see Quantum mechanics, interpretations of). For example, while in classical mechanics a knowledge of the simultaneous precise positions and momenta of the particles in an isolated system allowed prediction of the state of that system at any other time, in quantum mechanics it is impossible to assign simultaneous precise positions and momenta to particles. While the vacuum, or empty space, of classical physics was truly empty ‘nothingness’, particles appear and disappear spontaneously (causelessly, in any classical sense) from the dynamic vacua of modern quantum field theories (see Field theory, quantum §2). Those fields, while explanatory of immense bodies of scientific knowledge, are not classical entities with specific values of each of their assigned properties; the laws describing them are fundamentally probabilistic, treating them as superpositions of all quantum possibilities. Matter as conceived in modern physics is, in short, far different, and far more dynamic, than matter as conceived by Descartes and Newton. Finally, strong evidence indicates that there is far more matter in the universe than is detectable, at least by present instruments. Indeed, some observations and reasoning suggest that as much as 99 per cent of the matter in the universe may be of a kind different from the ordinary (‘baryonic’) matter we know (see Cosmology §3).

Thus the concept of matter has been altered in such drastic ways that certain contemporary ideas correspond at best only roughly to classical scientific and philosophical counterparts. The sharp contrasts between inert, passive permanence and dynamic activity, ultimate constituency and change, matter and force, fail to appear in quantum-theoretical analogues of classical matter. Indeed, the merely approximate analogy of fermions and bosons to classical concepts of matter, the activity of the quantum vacuum, the character of quantum fields as superpositions of possibilities and many other features of the quantum world, together with the existence of dark matter, all conspire to dictate the reformulation of many traditional problems of philosophy, such as the free will problem, the doctrines of materialism and determinism, the distinction between actuality and possibility, and the topic of matter itself.

See also: Matter, Indian conceptions of; Substance; Unity of science

DUDLEY SHAPERE

References and further reading

Descartes, R. (1644) *Principia philosophiae (Principles of Philosophy)*; ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, *The Philosophical Writings of Descartes*, vol. I, Cambridge: Cambridge University Press, 1984- 91.(The most mature statement of Descartes’ views of matter, differing in important respects from those in his earlier works. The cited work contains excerpts.)

Guthrie, W.K.C. (1962-81) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols. (Excellent, clear studies of Greek philosophers, with abundant discussions of problems about matter.)

Jammer, M. (1961) *Concepts of Mass in Classical and Modern Physics*, New York: Dover.(Brief survey of the history of concepts of matter or mass, focusing primarily on scientific aspects. Somewhat technical.)

Koyré, A. (1965) *Newtonian Studies*, Chicago, IL: University of Chicago Press.(Classic treatment of

seventeenth-century science, especially good on relations between thought of Descartes and Newton. Should be required reading for all interested in the historical topics relevant to the concept of matter.)

Krauss, L. (1989) *The Search for the Fifth Essence: The Search for Dark Matter in the Universe*, New York: Basic Books.(Popular survey of the ‘dark matter’ problem mentioned at the end of §3.)

Leibniz, G. (1969) *Philosophical Papers and Letters*, ed. L. Loemker, Dordrecht: Reidel.(Views on matter are scattered throughout; some especially significant discussions are found in ‘On Nature Itself’, ‘Thoughts on the Principles of Descartes’, ‘The Confession of Nature against Atheists’, and the correspondence with Clarke and with De Volder.)

McMullin, E. (ed.) (1963) *The Concept of Matter*, Notre Dame, IN: University of Notre Dame Press.(A collection of articles by historians and philosophers; most papers accessible to non-specialists.)

Newton, I. (1730) *Opticks or a Treatise of the Reflections, Refractions, Inflections and Colours of Light*; repr. New York: Dover, 1952.(The ’Queries’ added to the end of this work give much insight into Newton’s views on matter.)

Plato (*c*.366-360 BC) *Timaeus*; ed. A. Rivaud, *Platon: Timée, Critias*, Budé series, Paris: Les Belles Lettres, 1925; trans. F.M. Cornford, *Plato’s Cosmology*, London: Routledge & Kegan Paul, 1937; New York: Penguin, 1977. (Exposition of Plato’s ’geometrical atomism’.)

Ross, D. (ed.) (1966) *Aristotle, Works*, vol. II, Oxford: Clarendon Press.(In addition to Aristotle’s own views of matter, these writings contain criticisms of the views of his predecessors.)

Sambursky, S. (1956) *The Physical World of the Greeks*, London: Routledge & Kegan Paul.(Written by a physicist-historian, an excellent survey of Greek ideas about matter. Highly accessible.)

Shapere, D. (1991) ‘The Universe of Modern Science and Its Philosophical Exploration’, in E. Agazzi and A. Cordero (eds) *Philosophy and the Origin and Evolution of the Universe*, Dordrecht: Kluwer, 87-202.(Survey of ideas of contemporary physics and its relevance to philosophy.)

Toulmin, S.E. and Goodfield, J. (1962) *The Architecture of Matter*, New York: Harper & Row.(Very readable, clear presentation of the history of matter in philosophy and science.)

Matter, Indian conceptions of

During the long and complex history of Indian philosophy, a number of divergent conceptions of matter have been developed and explored. These conceptions diverge both with respect to the ontological analysis of matter, and with respect to its specific structural characteristics. In terms of ontological conceptions of matter, the rival positions of materialism, idealism and substance-pluralism are all advanced by competing schools of thought. For example, pure materialism is espoused by the Cārvāka school, while absolute idealism is defended by Advaita Vedānta, and varying forms of pluralism are advocated by the Sāōkhya, Yoga, Nyāya and Vaiśeṣika schools.

Regarding the structural characteristics of matter, the most interesting conceptions are advanced by the pluralistic philosophies. In particular, the Vaiśeṣika and Nyāya schools recognize five physical substances, four of which are held to possess atomic structure. According to this conception, matter is composed of imperceptibly small units or paramāṇus*, which constitute the basic substrate in which perceptible qualities inhere. All macroscopic objects are transient composites of atoms, while the* paramāṇus *themselves are indivisible and indestructible. The atoms are held to be naturally at rest, and an external force is required to initiate motion.*

In contrast, the Sāōkhya and Yoga traditions espouse a metaphysical dualism of the two basic categories of matter and consciousness, where the continuity and dynamic transformations of matter are emphasized. All of the diverse phenomena of the physical world result from modifications of a single underlying source known as pradhāna *or primal matter, which is said to be continuous, all pervading, indestructible and imperceptible.* Pradhāna *exists in a balanced and unmanifest state of* pralaya *until it is disturbed by the presence of consciousness. This disturbance leads to an imbalance between the internal constituents of* pradhāna*, and the resulting disequilibrium accounts for the evolutionary transformations of the physical world.*

1 General survey

In the present context it is not possible to provide a comprehensive survey of Indian conceptions of matter, which have developed over several millennia among a number of divergent philosophical schools. Thus the current section sketches some of the most prominent features of the intellectual landscape, and §§2-3 focus in more detail upon two particular theories concerning the structural characteristics of matter, selected on the basis of their conceptual importance and interest.

The metaphysical position of materialism is represented in Indian thought by the heterodox Cārvāka school, which upholds an ontology in which only the physical elements possess ultimate reality, and everything that exists is a composite of these elements formed through the operation of strictly natural processes. In common with ancient Greek theories, the Cārvāka school adopts the four elements of earth, water, fire and air - in distinction to most other schools of Indian philosophy, which accept *ākāśa* or ‘ether’ as a fifth material element. According to the Cārvākas, since only the material elements are real, matter is responsible for thought, and it is the material body which gives rise to consciousness. Therefore, dissolution of the body at death entails the cessation of associated mental activity (see Materialism, Indian school of §4).

At the opposite end of the ontological spectrum, the idealist position that matter is ultimately ‘unreal’ was also advocated, most notably by the Advaita Vedānta school of Śaōkara. According to this view, absolute consciousness (*cit*) is the fundamental reality, and all particular manifestations, including the material world, are at root illusory. *Māyā* is the principle which sustains the appearance of all relative phenomena, and thus matter is ultimately diagnosed as *māyā*, an illusory appearance entertained by consciousness.

However, while the positions of pure materialism and absolute idealism are both represented in the history of Indian thought, some form of metaphysical pluralism is embraced by the majority of schools. Amongst these, matter is conceived as ontologically independent of mind (and vice versa), and this conception naturally motivates a rigorous investigation of the structure of the independent material realm. The atomistic pluralism of the Nyāya and Vaiśeṣika schools will be examined in §2, and the dualism of Sāōkhya-Yoga philosophy will be covered in §3. The heterodox Jaina tradition also embraced a form of atomism in many ways comparable to the Nyāya-Vaiśeṣika theory (see Jaina philosophy §1).

The extended intellectual conflict between the Hindu and Buddhist camps served to stimulate the development and

articulation of their respective positions, and the Hindu views discussed below implicitly assume the presence of Buddhist theory as a competing system. At the heart of Buddhist thought lies a vision of the world as a shifting and transitory array of particles, which lacks any underlying substrate to serve as the common thread holding the moments of existence together. According to the doctrine of momentariness (*kṣaṇikavada*), the world is an evanescent flux which is not materially preserved from one moment to the next, but rather consists in a series of discrete 'instants' (*kṣaṇas*) of existence, followed by complete annihilation before the next instant begins. Material objects, which appear as unitary and enduring, are analysed as mere aggregates, and possess no reality beyond the sum of their elemental parts (*dhātus*). So material objects are conceived as ontologically fragmentary and temporally discontinuous; real existence at each moment is granted only to the basic elements. These elements form momentary compounds, which, when viewed over a series of such moments, give the appearance of integral and continuous objects, much as a series of discrete frames supplies an illusion of continuity in a motion picture (see Buddhism, Ābhidharmika schools of §3; Buddhist philosophy, Indian §5; Momentariness, Buddhist doctrine of).

2 Vaiśeṣika atomism

Kaṇāda, legendary author of the *Vaiśeṣikasūtra*, is regarded as the founder of the Vaiśeṣika school, one of the six *darśanas* of orthodox Hindu philosophy. The name derives from the term *viśeṣa*, meaning 'ultimate particular', and this indicates the distinctively metaphysical orientation of the school. The Nyāya *darśana* of Akṣapāda Gautama is distinguished by its preoccupation with logic and epistemology, and forms the sister school to Vaiśeṣika. The atomistic theory discussed in this section is part of the shared Nyāya-Vaiśeṣika position, although it is primarily the contribution of the latter school (see Nyāya-Vaiśeṣika).

In contrast to Buddhist philosophy, the Vaiśeṣika position is that all qualities must inhere in a substance; properties cannot exist without an underlying substrate (*āśraya*) to ground them. According to this pluralistic account, there are nine such substances: earth, water, fire, air, ether (*ākāśa*), time, space (*dis*), mind (*manas*) and spirit (*ātman*), of which the first five are considered material substances. Each substance forms the ground for a characteristic quality: earth has smell, water has taste, fire has colour, air has touch and ether supports sound. Each substance either has infinite dimension or else is atomic in structure, a structural distinction which cuts across the material/nonmaterial divide. Of the material substances, earth, water, fire and air are atomic, while ether is infinite; and of the nonphysical elements, mind is atomic, while time, space and *ātman* are infinite.

Thus Vaiśeṣika recognizes four kinds of material atom (*paramāṇu*). The various macroscopic objects of the natural world are analysed as composites formed from *paramāṇus* of these four kinds, where the qualities of all objects are due to the qualities of the atoms of which they are composed. *Paramāṇus* have a number of distinguishing characteristics: they are indivisible and without parts (that is, genuinely atomic), they are unalterable and eternal, they are imperceptible, and they in fact lack spatial extension altogether and are held to be infinitesimal. In this manner, all substances are fundamentally either infinite or infinitesimal in magnitude.

The different atoms of a particular substance are qualitatively indistinguishable, but are none the less distinct particulars (*viśeṣas*). And there is only a qualitative difference between the atoms of different substances. As stated above, each substance has its characteristic quality, but this quality is not manifested at the level of the single atom. Rather, two atoms must combine to form a dyad, and three dyads combine to form a triad, which is the smallest level at which qualities are manifest. A triad is said to be of the magnitude of a 'mote of dust in sunlight', and the experienced world of matter is built up from such triads.

In contrast to the doctrine of momentariness, atoms are held to be indestructible and hence must persist through time, as do the composite objects of which they form the parts. But the composite objects, though persistent over finite intervals, are themselves ultimately transient and will eventually decompose. Thus matter at the macroscopic level is unstable, and atoms in motion constantly combine, disperse and recombine in new patterns.

In addition to dissolution at the level of particular objects, there is also a cosmic cycle of dissolution, stasis and reactivation, which forms a recurrent theme in Indian philosophy. According to Vaiśeṣika thought, atoms are naturally passive, so the entire universe eventually runs down. During cosmic stasis (*pralaya*), atoms subsist at rest without producing any effects. Original movement is due to external impact or activation, so the material world is not a closed system. A new cycle of the universe is initiated when an outside and 'unseen' force (*adṛṣta*) supplies

the impetus.

The need to appeal to an external and latently purposive source of activation is an explanatory defect in a theory of matter, and the Vedānta thinker Śaōkara criticizes the Vaiśeṣika account in this regard. The infinitesimal status of atoms is also problematic, since it would seem to entail that no finite collection of atoms could possess spatial magnitude, yet a mere triad is said to have perceptible dimensions. Śaōkara also criticizes the view that the atoms have no parts, since this seems to imply that they cannot structurally combine; if they combine as wholes, then they must collapse into each other and remain infinitesimal, and if they combine in parts then they are not indivisible.

A number of important differences distinguish Kaṇāda's view from the ancient Greek atomistic theories of Democritus (§2) and Epicurus (see Epicureanism §3), indicating that the two conceptions are probably independent. The Greeks hold atoms to be spatially extended, while as mentioned above, Kaṇāda contends that they lack dimension and are hence comparable to points in mathematical space. Additionally, the Greeks hold atoms to differ only quantitatively and not qualitatively, so that qualitative differences reduce to quantitative ones (for example, the figure, size and weight of the different atoms). In contrast, Kaṇāda takes qualitative difference as fundamental, while quantitative differences are analysed as macro-level features. The Greeks also believe atoms to be naturally in motion, as opposed to the Vaiśeṣika account summarized above, in which atoms are naturally at rest. Finally, the Greek view would reduce mind to matter (as would the Cārvākas), but the Vaiśeṣikas take mind (*manas*) as an independent (but none the less atomic) substance (see Atomism, ancient).

3 Sāōkhya dualism

The Sāōkhya *darśana* is one of the oldest philosophical traditions of India, and many of its ideas are traceable to the *Ṛg Veda* and early Upaniṣads. Its historical founder is Kapila, though the original *Sāōkhyasūtra* he is said to have written during the sixth or seventh century BC is now lost, and the most important of the existing texts is the *Sāōkhyakārikā* of Īśvarakṛṣṇa, from around the third century AD. The Sāōkhya tradition has a great many points in common with the classical Yoga *darśana* expounded in Patañjali's *Yogasūtra*, and the conception of matter discussed in the present section is part of their shared philosophical framework (see Sāōkhya).

According to the Sāōkhya-Yoga view, the varied phenomena of the material world are all manifestations of a single underlying source. The metaphysical principle underpinning all physical manifestation is *pradhāna* or primal matter, which can be thought of as a form of pure potentiality, the ultimate base supporting all modifications and transformations. *Prakṛti* is another term used synonymously with *pradhāna* to denote the metaphysical basis of matter. As such, *prakṛti* is the unmanifest (*avyakta*) and uncaused origin of the physical universe. *Prakṛti* is said to be continuous, all-pervading and eternal. Although continuous, it is differentiated into three constituent powers or *guṇas*, and the interaction of the *guṇas* is responsible for the dynamics of manifestation and transformation. The three *guṇas* are *sattva*, *rajas* and *tamas*, which correspond roughly with 'transparency and form', 'energy and motion' and 'inertia and obstruction'. The *guṇas* are opposing and mutually unstable forces intrinsic to the nature of *prakṛti*, and they cannot be separated or merged. All three are present in all material phenomena, and the differences between diverse objects are ultimately traceable to the predominance or proportions of the different *guṇas*.

Prior to the phase of manifestation which results in matter, *prakṛti* exists in the equilibrium state of *pralaya*. This is not a passive state of rest as in Nyāya-Vaiśeṣika, but consists rather in the tension of opposing forces in perfect balance. When this balance is disturbed, the process of becoming is set in motion, and material manifestation is thus seen as a state of disequilibrium. Once out of balance, the 'mechanical' interplay of the *guṇas* results in the evolution of the universe; the destabilized triad of forces counterbalance and readjust in a shifting pattern of transmutation.

Pradhāna gives rise to the material elements, which are said to be atomic in structure. However, unlike the Nyāya-Vaiśeṣika theory, this atomism is a physical rather than a metaphysical thesis, since the atoms themselves are manifestations of a continuous substrate. Thus the atoms are not *metaphysically* discrete, and the five material elements still reduce to the three *guṇas*.

Material actuality emerges from the state of potentiality, and to that source it returns. There is no genuine creation or annihilation in the Sāōkhya view, but rather a sort of metaphysical unfolding and consequent reabsorption into

the realm of latent possibility. Like the Vaiśeṣika atoms, *pradhāna* itself is eternal and indestructible. Eventually the entire universe follows the same course as its individual members, and is reabsorbed into the state of potentiality; the equilibrium of *pralaya* is once again obtained, giving rise to the familiar cycle of cosmic emergence and dissolution.

Prakṛti is inherently unconscious, and is held to be incapable of producing consciousness as an effect; the evolution of matter is a 'blind' mechanical affair. According to the Sāōkhya-Yoga view, subjective awareness belongs to a different ontological category altogether, the realm of *puruṣa* or pure sentience. *Puruṣa* forms the other metaphysical pole in the dualistic system, and is said to consist of undifferentiated consciousness which is immutable and inactive, formless and without parts. However, in sharp contrast to Cartesian dualism, Sāōkhya theory holds that mental faculties and activities are at root material; the mind is a manifestation of *prakṛti* rather than *puruṣa*. Thought is characterized by form and motion, which are distinctive attributes of matter, while consciousness is formless and inactive. Hence the basic metaphysical schism between consciousness and matter then leads to an ancillary dualism between consciousness and mind.

Pure consciousness or *puruṣa* is passive and does not guide the evolution of matter, so it does not guide the progression of thought, which is a strictly material process. However, it is consciousness which triggers the disequilibrium leading from *pralaya* to manifestation. This is not an act of volition or agency on the part of *puruṣa*, but is said to be a mechanical force, like magnetic attraction, which results merely from the proximity of the two substances. So even though consciousness does not influence the specific development of matter, it supplies the initial disturbance that throws *prakṛti* into the disequilibrium of material processes.

Thus *puruṣa* plays a role similar to the Vaiśeṣika *adṛṣṭa* in triggering a cycle of material manifestation, though in the Vaiśeṣika account this is more a purposive or teleological event, while in Sāōkhya thought it is involuntary. Also, in the Sāōkhya picture, the energy which drives the evolution of matter does not derive from an external source, but rather is an expression of *prakṛti*'s inherent powers as they unconsciously unfold. But the Sāōkhya-Yoga picture is still clouded with respect to the precise nature of the interaction between the two substances. The fact that *prakṛti* can be 'mechanically' influenced indicates that it is not a closed system, and the fact that *puruṣa* can affect *prakṛti*, even unintentionally, indicates that consciousness is not purely inactive.

After *pralaya* has been disrupted and the cycle of manifestation has commenced, the two substances must continue their 'interaction' to produce the phenomenon of conscious mental activity. Thought processes and mental events are conscious only to the extent that they receive external 'illumination' from *puruṣa*, and it is the *sattva* constituent of matter which is able to 'absorb' this conscious radiance. The *sattva guṇa* is held to be 'transparent', while consciousness is often compared to a ray of light. The component of the mind known as *buddhi* or pure intellect (which is in some ways comparable to the Greek *nous*) is believed to consist of a preponderance of the transparent *sattva guṇa*, and this enables the material thought process of *buddhi* to become illuminated by the external light of consciousness. However, pure consciousness is held to be entirely independent of the mental structures it illuminates, just as sunlight does not *choose* to pass through particular transparent objects which happen to lie in its path.

The translucent quality of *buddhi* distinguishes thought-stuff from the grosser material objects of thought and perception, which ordinarily contain a preponderance of the 'dark' or opaque *tamas guṇa*. Thus only the thought material of *buddhi* is capable of conscious illumination, rather than the entire material realm, since a preponderance of the *tamas guṇa* renders the external objects of the material realm opaque to the light of consciousness. This is why minds appear to be conscious while stones and tables do not; only the subtle stuff of the *buddhi* is a suitable medium for receiving sentience, and thus it is minds which are the apparent loci of awareness in the material world.

See also: Ontology in Indian philosophy

PAUL SCHWEIZER

References and further reading

Dasgupta, S. (1922) *A History of Indian Philosophy*, Cambridge: Cambridge University Press, 5 vols.(A pioneering compendium of Indian philosophical thought.)

Īśvarakṛṣṇa (3rd century AD) *Sāōkhyakārikā*, in N. Sinha, *The Samkhya Philosophy*, New Delhi: Oriental Books

Reprint Corporation, 1979.(Good general overview of the Saōkhya school.)

Misra, U. (1936) *The Conception of Matter According to Nyāya-Vaiśeṣika*, Allahabad: M.N. Pandey.(Exposition of the shared view of matter.)

Patañjali (4th-2nd century BC) *Yogasūtra*, in H. Aranya, *Yoga Philosophy of Patañjali*, Albany, NY: State University of New York Press, 1983.(Excellent discussion of Yoga philosophy.)

Radhakrishnan, S. (1923) *Indian Philosophy*, London: Allen & Unwin, 2 vols.(A classic survey of Indian philosophy.)

Raju, P.T. (1985) *Structural Depths of Indian Thought*, Albany, NY: State University of New York Press.(Critical exposition of Indian philosophy with a useful glossary of Sanskrit terms.)

Schweizer, P. (1993) ‘Mind/Consciousness Dualism in Saōkhya-Yoga Philosophy’, *Philosophy and Phenomenological Research* 53: 845-59.(Discusses Sāōkhya dualism and compares it with issues in Western conceptions of mind and matter.)

Matthew of Aquasparta (*c.*1238-1302)

Matthew walked in the footsteps of Bonaventure, which were widened by his first followers, Walter of Bruges, John Pecham and William of Mare. For them, the knowledge of God's existence is the first truth implanted in the human mind. God's existence cannot be proved a priori (from something prior to it), since it is the first truth. It is a truth that is immediately known, not in the sense that there is actual knowledge of God implanted in the mind at birth, but because any judgment we make presupposes that the mind has contact with the Truth that is the measure of all truth. Such a first Truth must exist.

Born in the Italian village of Aquasparta, near Todi, Matthew came from the distinguished Bentivenghi family. A Franciscan friar, who often provided a response from the school of Bonaventure to the philosophy of Aquinas, he qualified as a *baccalareus biblicus* (lecturer on the Bible) at Paris in 1268 and *baccalareus Sententiarum* (lecturer on Lombard's *Sentences*) in 1273. He lectured at Bologna from 1273-7 and then became regent master at Paris from 1277-9, before being named lector at the Roman curia from 1279-87. He was elected General Minister of the Franciscans at Montpellier in 1287 and fulfilled this charge until 1289, even after being named a cardinal in 1288. He served the Holy See under Pope Boniface VIII until his death in 1302.

Following Anselm's path, Matthew argues: 'If the first and highest being is the first and highest being, it necessarily follows that the first and highest being exists, for what is first and highest is most actual and most perfect, and for this reason exists *in actu* (actually), since existence is of the very nature of a first and highest being. Indeed, existence is identical with such a being' (*Quaestiones disputatae de productione rerum et de providentia (Disputed Questions on the Production of Things and on Providence)* 11-12). Matthew, however, holds that it is also necessary to approach the question of God's existence from empirical grounds. Such an approach allows us to make more explicit the knowledge of God that is implied in any of our original judgments. He argues, first of all, from the imperfection and mutability of finite beings to the need for their perfect and immutable foundation, and then from the orderly way in which the world runs and the goals things naturally pursue, to a first efficient and final cause (*De productione rerum* 12-17).

As the first and highest being, God knows and loves in the highest degree. Infinitely fulfilled in his own perfection, 'the first being neither makes nor conserves anything unknowingly, nor irrationally, nor by chance, nor by the necessity of his nature, but knowingly and with foresight and in an orderly way' (*De productione rerum* 36). God, as first and highest being, does not need anything. If he created a universe, it could be only to share his supreme goodness. Out of this infinite goodness, Matthew argues, God created rational creatures capable of knowledge and love, and consequently capable of happiness, to manifest and share his boundless perfection (see Creation and conservation, religious doctrine of; God, concepts of).

In treating of creation, Matthew enters into the debate raging in the 1270s between the Averroists contending that the eternity of the world could be rationally demonstrated and the theologians who denied the validity of their proofs (see Averroism). Aquinas, admitting the temporal character of creation as an article of faith, contended that reason could demonstrate neither the temporal nor the eternal nature of creation. Matthew attacked Aquinas' efforts to show that specific arguments against the eternity of the world are not demonstrations. For Matthew, an eternal world would imply the existence of an infinite number of souls, or revolutions of the sun, or generations. These arguments against an eternal creation are necessary reasons, and the attempts of Aquinas to rebut them are sophistical (see Eternity of the world, medieval views of).

Matthew holds that the 'agent' and 'possible' intellects are cognitive faculties belonging to the soul itself. However, they need the assistance of divine illumination to arrive at perfect, evident and necessarily true cognition. In his commentary on Book I of the *Sentences* (d. 35, q. 8), following Guibert of Tournai, he portrays this illumination as a special created light given by God. In his later *Quaestiones disputatae de cognitione (Disputed Questions on Knowledge)*, returning to Bonaventure's position, he argues that this light must be God himself. Special assistance is required because the spiritual nature of the soul prevents it from being affected by sensible objects. The agent intellect, then, assisted by divine illumination, forms intelligible species of sense objects and impresses them on the possible intellect. Objects and the intellect are partial causes of knowledge, but the intellect is the principal cause. In regard to certain and necessary knowledge, however, the intellect is a

secondary efficient cause; the primary cause is the divine light. The soul has a direct knowledge of itself. Concepts of purely immaterial entities, such as God, angels, the good or first principles, are innate or impressed in the soul and are not abstracted from the outside world.

Matthew is very much influenced by Aristotle's philosophy, especially as elaborated by Aquinas. Even when he rejects Aquinas' positions, Matthew's arguments are not simply repetitions of those of Bonaventure and his early followers. They are serious attempts to overcome Aquinas' theses. For instance, when Aquinas holds that the will acts according to the directives it receives from reason or intellect, Matthew objects that, while needing enlightenment from the intellect, the will itself activates the intellect, giving it commands and stimulating it to action. The will is superior to the intellect, first because its perfection, charity, is superior to the intellect's perfection, faith; and secondly, because its object, goodness, is superior to the intellect's object, truth. Will, not reason, is also the centre of freedom. Reason is necessary to enable us to conceive the possible choices open to us, and to deliberate about them; but it is the will which produces free human action, since a person depends on will when deciding to accept or reject what reason proposes (see Free will).

See also: Aristotelianism, medieval; Augustinianism; Bonaventure; God, concepts of; Eternity of the world, medieval views of; Pecham, J.

STEPHEN F. BROWN

List of works

Matthew of Aquasparta (*c.*1238-1302) *Quaestiones disputatae de aeternitate mundi (Disputed Questions on the Eternity of the World)*, q. 9, utrum [mundus] potuit esse ab aeterno vel utrum Deus potuit ipsum ab aeterno producere, ed. E. Longpré, *Archives d'histoire doctrinale et littéraire du moyen-âge* 1, 1926: 293-308.(Edition of disputed questions.)

Matthew of Aquasparta (*c.*1238-1302) *Quaestiones disputatae de anima VI (Six Disputed Questions on the Soul)*, ed. A.-J. Gondras, *Archives d'histoire doctrinale et littéraire du moyen-âge* 32, 1957: 203-352.(Edition of disputed questions on the soul.)

Matthew of Aquasparta (*c.*1238-1302) *Quaestiones disputatae de anima XIII (Thirteen Disputed Questions on the Soul)*, ed. A.-J. Gondras, *Études de philosophie médiévale* 50, 1961.(Edition of disputed questions on the soul.)

Matthew of Aquasparta (*c.*1238-1302) *Quaestiones disputatae de fide et de cognitione (Disputed Questions on Faith and Knowledge)*, ed. Patres Franciscani, Bibliotheca Franciscana Scholastica Medii Aevi I, Quaracchi: Editiones Collegii Sancti Bonaventurae, 1957.(Edition of disputed questions.)

Matthew of Aquasparta (*c.*1238-1302) *Quaestiones disputatae de incarnatione et de lapsu (Disputed Questions on the Incarnation and the Fall)*, Bibliotheca Franciscana Scholastica Medii Aevi II, Quaracchi: Editiones Collegii Sancti Bonaventurae, 1957.(Edition of disputed questions.)

Matthew of Aquasparta (*c.*1238-1302) *Quaestiones disputatae de gratia (Disputed Questions on Grace)*, ed. V. Doucet, Bibliotheca Franciscana Scholastica Medii Aevi XI, Quaracchi: Editiones Collegii Sancti Bonaventurae, 1935. (Contains a detailed biography and examination of Matthew's extensive writings, approximately thirty volumes of 500 pages each.)

Matthew of Aquasparta (*c.*1238-1302) *Quaestiones disputatae de productione rerum et de providentia (Disputed Questions on the Production of Things and on Providence)*, ed. G. Gàl, Bibliotheca Franciscana Scholastica Medii Aevi XVII, Quaracchi: Editiones Collegii Sancti Bonaventurae, 1956.(Questions on metaphysics.)

Matthew of Aquasparta (*c.*1238-1302) *Quaestiones disputatae de anima separata, de anima beata, de ieiunio et de legibus (Disputed Questions on the Separated Soul, the Blessed Soul, Fasting, and Laws)*, Bibliotheca Franciscana Scholastica Medii Aevi XVIII, Quaracchi: Editiones Collegii Sancti Bonaventurae, 1959.(Edition of disputed questions.)

Matthew of Aquasparta (*c.*1238-1302) *Quaestiones disputatae de Christo (Disputed Questions on Christ)*, ed. E. Hocedez, *Quaestio de unico esse in Christo*, Rome: Apud aedes Universitatis Gregorianae, 1933, 44-58. (Questions on the nature of Christ.)

Matthew of Aquasparta (*c.*1238-1302) *Quaestiones disputatae de illuminatione (Disputed Questions on Illumination)*, ed. Patres Franciscani, *De humanae cognitionis ratione*, Quaracchi: Editiones Collegii Sancti Bonaventurae, 87-177.(Matthew's views on illumination.)

Matthew of Aquasparta (*c.*1238-1302) *Commentarium in I-IV Sententiarum (Commentary on the Four Books of Sentences)*, Lib. I, prol., q. 7, ed. L. Amorós, *Archives d'histoire doctrinale et littéraire du Moyen Age* 9, 1934: 284-5; Lib. I, d. 2, qq. 1-2, ed. A. Daniels, *Quellenbeiträge und Untersuchungen zur Geschichte der*

Gottesbeweise im 13. Jahrhunderts, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters 8, 1-2, Münster: Aschendorff, 1905, 51-63; Lib. II, d. 19, q. 1, ed. S. Vanni Rovighi, *L'immortalità dell'anima nei Maestri Francescani del secolo XIII*, Milan: Societa editrice 'Vita e pensiero', 1936, 255-72.(Book I includes Matthew's views on illumination.)

References and further reading

Beha, H.M. (1961) 'Matthew of Aquasparta's Theory of Cognition', *Franciscan Studies* 20: 161-204, 21: 383-465.(Describes Matthew's views on knowledge and its relation to illumination.)

Berubé, C. (1974) 'Henri de Gand et Mathieu d'Aquasparta interprètes de saint Bonaventure' (Henry of Ghent and Matthew of Aquasparta: Interpreters of St Bonaventure), *Naturaleza y gracia* 21: 131-72.(Studies the development of different views of illumination.)

Bettoni, E. (1943) 'Rapporti dottrinali fra Matteo d'Aquasparta e Giovanni Duns Scoto' (Doctrinal Relations between Matthew of Aquasparta and John Duns Scotus), *Studi Francescani XV* (3rd series): 113-30.(Studies the influence on Scotus of Matthew's teaching on the object of the intellect.)

Dowd, J.D. (1974) 'Matthew of Aquasparta's De productione rerum and its Relation to St. Thomas Aquinas and St. Bonaventure', *Franciscan Studies* 34: 34-73.(Shows the influences of Aquinas and Bonaventure on Matthew.)

Emmen, A. (1959) 'Die Glückseligkeitslehre des Matthäus von Acquasparta' (The Teachings on Beatitude of Matthew of Aquasparta), *Wissenschaft und Weisheit* 22: 43-59, 101-18, 174-89.(Explains Matthew's teaching on beatitude.)

Hayes, Z. (1964) *The General Doctrine of Creation in the Thirteenth Century, with Special Emphasis on Matthew of Aquasparta*, Munich: Schöningh.(Matthew's creation doctrine in its context.)

Marrone, S.P. (1983) 'Matthew of Aquasparta, Henry of Ghent and Augustinian Epistemology After Bonaventure', *Franziskanische Studien* 65: 252-90.(Describes the influence of Bonaventure on later thinkers, including Matthew.)

Mazzarella, P. (1969) *La dottrina dell'anima e della conoscenza in Matteo d'Acquasparta (Matthew of Aquasparta's Teaching on the Soul and Knowledge)*, Collana di studi filosofici 17, Padua: Gregoriana. (Matthew's teaching concerning the soul and human knowledge.)

Payne, G.R. (1981) 'Cognitive Intuition of Singulars Revisited (Matthew of Aquasparta versus B.J. Lonergan)', *Franciscan Studies* 41: 346-84. (On Matthew's views on knowledge.)

Prezioso, F. (1950) 'L'attività del soggetto pensente nella gnoseologia di Matteo d'Acquasparta e di Ruggiero Marston' (The Activity of the Present Subject in the Epistemology of Matthew of Aquasparta and Roger Marston), *Antonianum* 25: 259-326.(Indicates the role of the knower as a cause of knowledge.)

Putallaz, F-X. (1991) *La connaissance de soi au XIIIe siécle. De Matthieu d'Aquasparta à Thierry de Freiberg (Self-Knowledge in the Thirteenth Century: From Matthew of Aquasparta to Dietrich of Freiberg)*, Études de philosophie médiévale 67, Paris: Vrin. (Deals with self-knowledge in authors of the late thirteenth century.)

Mauthner, Fritz (1849-1923)

The work of Fritz Mauthner helps document the phenomenon of 'language crisis' or Sprachkrise *in German-Austrian letters at the beginning of the twentieth century. In his Beiträge zu einer Kritik der Sprache (Contributions to a Critique of Language) (1901-2), Mauthner develops a theory of knowledge that draws on empiricism but also redefines certain basic concepts in terms of language. 'Language' refers to more than speech; it is the medium of all cognition and, as such, an instrument of knowledge. Mauthner's reformulation of epistemological questions in linguistic terms does more than replace one topic with another. When language becomes the focus of philosophical debate, the debate cannot help but involve a discussion of the nature and limits of the discussion itself. As epistemology or theory of knowledge gives way to critique of language, it spells the end of philosophy as a foundational discourse for the human sciences.*

1 Language in thought and communication

The philosophical tradition with which Mauthner aligns himself most consistently is that of empiricism (see Empiricism). He acknowledges debts to Kant, Locke and Hume, and sees his work as the continuation or even completion of their inquiries into the possibility of knowledge. Mauthner also stresses his divergence from these three philosophers with his notion of the 'contingent senses' or *Zufallssinne*. He outlines a traditional-sounding sceptical thesis about the limits and accuracy of empirical knowledge, then gives the argument an evolutionistic turn. He holds that the senses are contingent on human needs and have undergone modifications over time corresponding to shifts in those needs. Therefore sense data reflect what is useful, and utility by no means implies accuracy. On the contrary, Mauthner presents the filtering, organizing activity of the senses as a process of distortion.

Out of his scepticism about knowledge through the *Zufallssinne* grows his thesis that language provides only *Zufallsbilder* ('contingent images') of the world. The process of concept formation that begins with sense information also marks the first instance of language. 'Language' is the very medium in which mental pictures of reality are created. According to Mauthner, language is synonymous with reason, cognition and the work of memory. By proclaiming that there is no such thing as 'pure reason' (*reine Vernunft*) but only 'linguistic reason' (*sprachliche Vernunft*), Mauthner echoes a criticism of Kantian epistemology voiced already by Hamann and Herder in the eighteenth century. In this connection he also introduces the term 'metaphor'. By describing the relation between mind and world as metaphorical, Mauthner underlines his point about the linguistic character of all knowledge, since 'metaphor' is a literary trope and an element of rhetoric. The term also serves as a reminder that knowledge is limited: it results from the transformation or translation of sensations into concepts, and something is always lost in translation.

Mauthner's scepticism about an individual's ability to achieve true knowledge also bears directly on his view of language as a means of communication. The notion of a gap between the sensible and intelligible is already introduced with the thesis of the *Zufallssinne*, and the gap widens twice: first, when sense data become schematized in language (here understood in the psychological sense), and again when translated into the language shared by members of a linguistic community. Spoken or written language in Mauthner's view thus stands at a double remove from reality. He contends that because no two people have the same perceptions or thoughts, what they express by means of their communal language is virtually worthless. While admitting that the 'misrepresentations' are adequate for the demands of daily life, Mauthner does not conclude that any other form of knowledge is unnecessary or impossible. He retains instead the premise that knowledge consists in accurate representations of reality, and predictably arrives at sceptical conclusions about the ability of language to provide such knowledge.

2 Philosophy as critique of language

At the beginning of *Beiträge zu einer Kritik der Sprache (Contributions to a Critique of Language)* (1901-2), Mauthner admits that language must play a double role: it is the vehicle of the critique, and also the object of study. But the full extent of the problem becomes clear only when he has developed the thesis about the inadequacy of language as an instrument of knowledge. This argument raises the question of whether it is possible to claim the inadequacy of language without at least implicitly granting the claim itself a special authoritative

status. In other words, does Mauthner's critique itself reflect or bear out the claims about language that it makes? His scepticism would seem to point toward silence, yet Mauthner continues to write after finishing the *Kritik*. Mysticism, too, receives considerable attention in his work, and suggests another possible outcome of language critique. Silence and mysticism, however, remain potential answers rather than actual responses, options that Mauthner discusses but does not choose. The reason lies in the fact that language also has a social dimension which even Mauthner, for all his emphasis on the individual, psychological aspect of language, cannot overlook. When writing about language as a form of communication, he dismisses it as the language of the 'herd', which he feels has nothing to do with the object of his critique. What complicates the situation is that he, too, remains subject to the social force of language. In short, while Mauthner decries the tyranny of language on the first page of his *Kritik*, he still cannot escape it, and he admits as much in several different but related ways.

Naming the individual as the locus of (epistemological) value is itself a gesture that reflects social or communal agreement. The implied community in this case consists of philosophers writing before and at the same time as Mauthner, who likewise concentrate on the individual subject when theorizing about knowledge. In this context he also retains traditional metaphors of vision and representation, and thereby signals once more that he is participating in an ongoing conversation. Perhaps the clearest indication of how the social character of language in effect wins out over the individual-oriented aspect lies in Mauthner's statements about the purpose of his critique. He observes that unless others read his work and adopt his ideas, the critique has absolutely no value, and in his eyes is not even 'real'. He emphasizes his dependence on a community with another metaphor: calling language both a 'social game' (*Gesellschaftsspiel*) and 'the social game of knowing', he characterizes his own efforts as an attempt to change the rules.

On occasion Mauthner seems to grant his own work a voice of authority in traditional Kantian fashion, but other passages reveal that he questions philosophy's ability to provide knowledge or access to truth. He declares for instance that although individual philosophers and philosophies (plural) exist, there is no such entity as 'philosophy' (singular). In his view, philosophy-as-language-critique is indeed the discipline that investigates the working assumptions of other disciplines, but because the inquiry is not written in any sort of metalanguage, philosophy no longer occupies the foundational position it once held. As critique of language it has irrevocably lost its authoritative voice.

3 Reception and later works

Several of Mauthner's later works move away from the emphasis on individual psychology that characterizes his critique. *Die Sprache (Language)* (1906), for example, addresses questions about language as a social phenomenon including attempts to develop an international language or Esperanto, and different forms of translation. Mauthner draws attention to the shift in focus when he notes that the concept of a *sensorium commune* has been borrowed from 'older' psychology and applied to language in a social context, where in his view it becomes more tangible and accessible than when used in reference to the individual mind.

In *Die Sprache* Mauthner mentions an ambitious, now abandoned plan to write an 'international' historical dictionary that would document processes of translation. The somewhat more modest project that takes its place is his *Wörterbuch der Philosophie (Dictionary of Philosophy)* (1910-1), subtitled 'New Contributions to a Critique of Language'. Beyond tracing semantic change in some two hundred philosophical terms, Mauthner hopes to persuade his readers that meanings are anything but stable and that, because there are no fixed referents in language, there is no true knowledge. Because his nominalism skews the perspective decidedly, the dictionary is not always entirely reliable as a reference work. None the less, individual articles contain a wealth of historical information, and also provide condensed versions of arguments advanced in the *Kritik*.

The other major work that Mauthner published during his lifetime was a four-volume history of atheism. Although quite different in style from his philosophical dictionary, this work reflects the same basic argument that linguistic meaning is determined by use, and therefore changes over time. In his autobiographical writings Mauthner admits that at the time he began speculating about the existence of God, he probably did not recognize any link between his religious doubts and his growing interest in language. But as an adult, and more specifically as the author of a work on atheism, he perceives and even insists on the connection. He describes the history as the 'history of a word, the negative word-history of the gradual devaluation of the word "God"' (*Der Atheismus und seine Geschichte im Abendlande*, 1920-3). He intends to prove that belief in 'God' as anything more than a word is a

form of word superstition.

In addition to the dictionary and the work on atheism, Mauthner completed short monographs on Aristotle, Spinoza and Schopenhauer. He also wrote the first (and only) volume of his memoirs, and edited several volumes in the series *Bibliothek der Philosophen* (Library of Philosophers), which was discontinued at the beginning of the First World War. The volumes on which Mauthner worked were on Fritz Jacobi's *Spinozabüchlein*, Agrippa von Nettesheim and the nineteenth-century philosopher of language O.F. Gruppe.

During his lifetime Mauthner enjoyed wide recognition, albeit not as a philosopher. He worked as a professional journalist for over twenty years in Berlin and, prior to writing the *Kritik*, also published novels, novellas and essay volumes in addition to a popular collection of literary parodies. Mauthner's literary and journalistic writings led many academic philosophers to dismiss his language critique as the work of a dilettante, but his work was taken seriously by thinkers including Ernst Mach, Hans Vaihinger and Martin Buber. After his death, Mauthner's work came to attention only gradually, in part because his extreme scepticism was so incompatible with more dominant trends in philosophy such as phenomenology and logical positivism.

Mauthner has the dubious distinction of being mentioned by name in Ludwig Wittgenstein's *Tractatus Logico-Philosophicus* (1922: 4.31), where we read: 'All philosophy is a "critique of language". (Though not in Mauthner's sense.)'. Wittgenstein takes very different stances from Mauthner's on such issues as the possibility of scientific knowledge, and maintains that congruence actually does exist between reality and language (in the form of logical propositions). Despite this important basic difference, the *Tractatus* and Mauthner's *Kritik* both end with a turn toward mysticism. For Mauthner the premise that some unmediated and extralinguistic mode of experience exists leads to a mystical yearning for that kind of experience. Wittgenstein, after working through propositions about what is sayable, declares that some things still elude us, and that these things may ultimately be more important than any knowledge we do have.

How well Wittgenstein knew Mauthner's work is not clear, but there is general agreement that their affinities emerge more clearly if we look at Wittgenstein's later work, particularly *Philosophical Investigations* (1953), where the notions of language game and meaning as use recall Mauthner's positions. Wittgenstein's description of his project as a way of showing the fly the way out of the fly bottle also seems to echo Mauthner, who stresses the salutary effect of critique of language. His attacks on 'word fetishism' and his insistence on the impossibility of attaining knowledge earned him an early reputation as a nihilist, which he himself rejected. Critique of language unmasks and destroys idols in the name of liberation, he argues; it should not be feared as a threat but welcomed as a sorely needed form of philosophical therapy.

See also: Wittgenstein, L.

ELIZABETH BREDECK

List of works

Mauthner, F. (1901-2) *Beiträge zu einer Kritik der Sprache (Contributions to a Critique of Language)*, Leipzig: Felix Meiner, 3rd edn, 3 vols.(Individual volumes on different topics: psychology of language, concepts in linguistics, grammar and logic.)

Mauthner, F. (1906) *Die Sprache (Language)*, in M. Buber (ed.) *Die Gesellschaft*, vol. 9, Frankfurt am Main: Literarische Anstalt Rütten & Loening.(Short essayistic work with emphasis on social rather than psychological dimension of language.)

Mauthner, F. (1910-1) *Wörterbuch der Philosophie. Neue Beiträge zu einer Kritik der Sprache (Dictionary of Philosophy: New Contributions to a Critique of Language)*, Leipzig: Felix Meiner, 2nd edn, 3 vols.(Historical dictionary of philosophical terms, informed by the notion that linguistic meaning is determined by use.)

Mauthner, F. (1920-3) *Der Atheismus und seine Geschichte im Abendlande*, repr. Frankfurt am Main: Eichborn, 1988, 4 vols.(Historical work tracing both the 'struggle for liberation from the belief in God' and the history of the word 'God'.)

Mauthner, F. (1986) *Fritz Mauthner. Sprache und Leben. Ausgewählte Texte aus dem philosophischen Werk*, ed. and intro. G. Weiler, Salzburg: Residenz.(Contains excerpts from *Kritik*, *Sprache*, *Wörterbuch* and *Atheismus*; also from *Totengespräche* (1906), *Der letzte Tod des Gautama Buddha* (1913), *Gespräche im Himmel und andere Ketzereien* (1914), *Erinnerungen* (1918), *Die Philosophie der Gegenwart in Selbstdarstellungen* (1924) and *Die drei Bilder der Welt* (1925).)

References and further reading

Arens, K.M. (1984) *Functionalism and Fin de siècle. Fritz Mauthner's Critique of Language*, Bern: Peter Lang. (Mauthner in relation to nineteenth-century linguistics and various scientific models.)

Bredeck, E. (1992) *Metaphors of Knowledge. Language and Thought in Mauthner's Critique*, Detroit, MI: Wayne State University Press.(Rhetorical analysis, with emphasis on links between Mauthner's thought and contemporary theory.)

Eschenbacher, W. (1977) *Fritz Mauthner und die deutsche Literatur um 1900*, Bern: Peter Lang.(Concentrates on literary-historical context.)

Kühn, J. (1975) *Gescheiterte Sprachkritik. Fritz Mauthners Leben und Werk*, Berlin: de Gruyter.(Contains useful bibliography of all Mauthner's writings including journalism, literary works and philosophical texts.)

Leinfellner, E. and Schleichert, H. (eds) (1995) *Fritz Mauthner. Das Werk eines kritischen Denkers*, Vienna: Böhlau.(Essay collection including photographs, chronology of Mauthner's life and up-to-date bibliography of secondary literature.)

Weiler, G. (1970) *Mauthner's Critique of Language*, Cambridge: Cambridge University Press.(Detailed discussion of philosophical issues, situates Mauthner in the history of philosophy.)

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, trans. C.K. Ogden and F.P. Ramsey, London: Routledge; trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1961.(Quoted in §3. The major work of Wittgenstein's early period and the only book published during his lifetime. The first English translation was revised and approved by Wittgenstein himself, though the later version is now standard.)

Maxwell, James Clerk (1831-79)

For his two achievements of unifying electricity, magnetism and light, and of inventing statistical dynamics, Maxwell stands as the founding mind of modern theoretical physics. More than any other physicist's his also was a mind shaped and informed by a training in philosophy, even though, unlike Heinrich Hertz or Ernst Mach, for example, he never wrote a philosophical treatise. Therein lies the point, however. Mach's and Hertz's best discoveries seem remote from their metaphysics, Maxwell's are bound up with his. Particularly important philosophically are his interconnected uses of relation, analogy and classification. He is also responsible for introducing the word 'relativity' into physics, and for articulating the scientific problematic that led to Einstein's theory.

1 Physics and metaphysics

Born in Edinburgh, Maxwell spent three years at Edinburgh University and learned philosophy there from Sir William Hamilton. Hamilton was a critical rather than systematic thinker, and was influenced by Kant. His greatest gift to Maxwell was a human one - contact with a searching and erudite mind. Most potent among specifics was his doctrine of the 'relativity' of human knowledge - a word he coined in opposition to the post-Kantian idealists Fichte, Schelling and Hegel. Contrasting knowledge with belief, Hamilton held that while we may believe in the absolute we know nothing absolutely. All knowledge involves relations, either our own to the object or of objects to each other.

Maxwell migrated to Cambridge in 1850. There he gained another, different Kantian guide, William Whewell. Whewell's focus was science. In 1830 his friend John Herschel had written a brilliant sketch of early nineteenth-century discovery framed to illustrate Baconian principles - the work from which John Stuart Mill borrowed the five methods of induction now called his (see Science, 19th century philosophy of §4). Whewell demurred, holding that while each science starts with Baconian observation, sooner or later it undergoes transition to a higher Kantian state governed by certain 'appropriate ideas' having an inner authority of their own, independent of the experiments that suggested them. Take Newton's inverse-square law of gravitation. Why is this exact? Kant had guessed a connection with the tridimensionality of space, since the surface area of a sphere in three dimensions is $4\pi r^2$. Whewell gave other examples, arguing that one mark of 'appropriateness' is simplicity, or, as modern physicists say, 'beauty'.

First an artist, then a poet, then a geometer, all before he was fifteen, Maxwell was that paradox, a slowly maturing prodigy. Of his first nine publications other than poems, five treated geometry, two colour vision, one - half-experimental - physics. Not until after his graduation from Cambridge did he join physics to ' metaphysics [in] the rigid high style…about ten times as far *above* Whewell as Mill is *below* him, or Comte or Macaulay *below* Mill, using above and below conventionally like up and down [trains] in Bradshaw['s railway timetable]' ([Letter to R.B. Litchfield 4 July 1856] Campbell and Garnett 1882: 261).

Maxwell's work is a barrage of classifications. Classification of dynamical quantities, of electrical ones, of transport processes, of colours, classification in thermodynamics, various mathematical classifications, classification by dimensional analysis, classifications of instruments, diagrams, experimental techniques, laboratory designs, modes of research, types of scientific paper, even (wittily) of scientific meetings - all occur. With this multiplicity, in but not limited to the classifications, are *dualities*: two ways of classifying dynamical topics, two of vectors, two of colours, two of physical quantities at large, a dual classification of electrical units, dualities in optics, thermodynamics, electromagnetism, two approaches to harmonic analysis, two to potential theory, a dual definition of electromotive force. In electromagnetism, these dualities and multiple classifications connect with Hamiltonian relativity, with Faraday's ideas, and with mathematical analogies by Maxwell's fellow Scot William Thomson (Lord Kelvin), as will now appear.

2 The electromagnetic field

Maxwell's theory united electricity, magnetism and light, it created radio, generated the relativity crisis, and set physics on a new path - field theory. Rich in discovery, new in form, startling in its picture of the electric current as an entity extending beyond the confines of the wire that guides it, this was more than a new theory. It was a

new*kind* of theory.

According to Maxwell, two magnets (or charges or electric currents) do not act on each other directly. Each separately stores up energy in the surrounding space (the 'field') and the forces arise indirectly via the field. A natural question is whether this indirect transmission is instantaneous. To his astonishment Maxwell found that electrical signals propagate through space as waves with a velocity identical with that of light. Moreover, these waves *had to be* transverse. The problem of the longitudinal wave, the nightmare of mechanical ethers, had evaporated (see Electrodynamics; Field theory, classical).

Earlier workers had sought causes; Maxwell sought relations. Instead of *explaining* light or electromagnetism he established a relation between the two, and the central structure of his theory is neither a law nor a set of axioms but a system of relations among electric and magnetic quantities - Maxwell's equations. These he grouped in two ways. One, *electrical*, comprised the four equations we now call his; the other, *dynamical*, included also the 'Lorentz' force and the vector potential A. Compactness plus a belief that A is unphysical made his successors prefer the first. Maxwell saw further. He wrote in 1873: ' Our object at present is not to obtain compactness in the mathematical formulae but to express every relation of which we have any knowledge. To eliminate a quantity that expresses a useful idea would be rather a loss than a gain at this stage of our enquiry' (Maxwell 1873, vol. II: 234).

Electric currents produce not attractions, but *circular* magnetic forces. Why? Earlier opinion had been divided. To Ampère and Wilhelm Weber the effect was secondary, reducible to underlying attractive-repulsive forces, albeit forces entangling strange additional dependencies on angles (Ampère) or relative motions (Weber). To Faraday, the instinctive geometer mapping lines of force, such explanations seemed perverse, a reduction of the simple to the complex, especially when he found that even electrostatic forces act in curved lines. And so to Maxwell's first paper (1856).

Picture - the analogy is Thomson's - a fixed elastic solid in which is embedded a long straight wire. Pull the wire a distance i′. The solid will become cylindrically distorted, with displacements A′ parallel to i′ and consequent twistings B′ about directions circling the wire. Different as they are physically, these quantities i′, B′, A′ correspond geometrically to current i, magnetic force B and Maxwell's A in electromagnetism. Transcending geometry, Maxwell then made a spectacular discovery. Other electromagnetic phenomena also depend neatly on A. Here was a theory, dual to Weber's, advantageous precisely because it did not ' even in appearance, *account for* anything' (Maxwell 1890, vol. I: 207)

A second analogy, between lines of force and fluid motion, enabled Maxwell to mesh Faraday's curved lines of force with classical potential theory, and legitimize Kant's argument about the inverse-square law. Inverse-square forces entail a field satisfying a continuity equation. The law rests on experiment but any departure from it means large changes in theory - a different underlying geometry, say, as with gravity in Einstein's theory, or radically different field concepts as with nuclear forces.

Different equations, Maxwell's and Weber's, cover the same phenomena. Different phenomena, electric force and fluid motion, obey the same equations. Such dualities, with analogy, subvert conventional scientific explanation. Even more peculiar is the union, geometrical not physical, between Coulomb's and Faraday's theories, effected by an analogy unrelated physically to either. It is puzzles like these that destroy the well-known saying by Heinrich Hertz, that ' Maxwell's theory is Maxwell's system of equations', and analogies merely 'gay garments' clothing 'the simple and homely figure' of nature (Hertz 1893: 21, 28). How strangely have Hertz's admirers failed to see his couturierial vision as itself an analogy, and a singularly inept one. For Maxwell's problem is just this - how to link fact, concept, and equation when each holds multiple meanings.

One duality that Maxwell, with magian prescience, illumined was between wave and particle theories of light. ' That theories so fundamentally opposed should have so large a field of truth common to both is a fact the philosophical importance of which we cannot fully appreciate until we have reached a scientific altitude from which the true relation between hypotheses so very different can be seen' (Maxwell 1890, vol. II: 228).

Was all theory analogy? Maxwell's in its dynamical form illustrates the complexities. It was in two senses hybrid. It inserted electrical energies into dynamical equations, and it contained some things literal, others analogical. Field energy was - energy. *Electrokinetic momentum*, Maxwell's new name for his old A, was an analogy.

That A resembles *momentum* for currents and *potential* for forces shows the subtlety of 'scientific metaphor'. Long neglected, the momentum analogy gained new life with quantum mechanics. In 1959, Y. Aharonov and D. Bohm proved, against eighty years of post-Maxwellian wisdom, that through it, A has direct physical meaning.

3 Gases, molecules, statistics

Maxwell did not introduce statistics into physics, or invent gas theory, or first fix molecular sizes. Yet he revolutionized all three, pondered metaphysical consequences and, long before Planck, quanta, and the 'ultraviolet catastrophe', saw here alongside triumph sudden scientific disaster.

It was Clausius who brought in statistics. Recognizing that the fast-moving molecules of gases must constantly collide, he defined in 1858 a characteristic mean *distance* between collisions, taking all molecules to have the same speed. Correcting this, Maxwell treated *speed* also statistically by computing a 'velocity distribution function'. Then came classification. Linking three disjoint phenomena, diffusion, viscosity, thermal conduction, to transport through the gas, respectively, of mass, momentum, and energy, he obtained, among much else, two separate, very good estimates of the free path in air ($\sim 6 \times 10^{-6}\,\mathrm{cm}$), and the astonishing result that viscosity is over a wide range independent of pressure. Investigations of extreme technical sophistication followed, not least the great paper of 1867 on transport processes that electrified Boltzmann.

Are atoms real? If mass and diameter are primary qualities, yes - a reality suddenly unveiled in 1865 by J.J. Loschmidt from data on free paths and condensation-ratios. Here we must repudiate the myth that atomism had no physical basis prior to Einstein's and Perrin's work around 1905. The evidence was overwhelming. Maxwell's 1873 treatment (reprinted in 1890: 343-50) was the best, yielding estimates for molecular masses and dimensions, and Avogadro's number, remarkably near modern values.

Maxwell had found, and Boltzmann elaborated, a most surprising statistical law, the equipartition theorem. It gave for specific heats of gases unequivocally wrong answers. Classical mechanics had failed. Reviewing and rejecting in 1875 all the supposed explanations, Maxwell could only urge that ' thoroughly conscious ignorance which is the prelude to every real advance in knowledge' (1877b).

Since 1925 many people have made quantum uncertainty the door to human free will. Randomness spells choice. Years earlier, Maxwell in an essay on 'Science and Free Will' (1873; Campbell and Garnett 1882: 434-44) explored similar thoughts as issues of analogy. There is science, there is life. Planetary orbits are exactly predictable; gas laws statistically so; other phenomena not at all because tiny changes in initial conditions produce through instability vastly different outcomes. That in essence is chaos theory (see Chaos theory). Not every outcome is chaos, however. Human achievement lies in seizing such singular points, such 'tides in the affairs of men', and our deepest intuitions of free will connect with them. If that only advances us one step (elsewhere Maxwell said, 'states of the will only puzzle me' (1879; Campbell and Garnett 1882: 417)), it is a useful step.

Who but Maxwell would dare the inversion of basing physics on free will? Picking and choosing between fast and slow molecules, Maxwell's demon (in his *Theory of Heat* 1871) uses intelligence rather than work to transfer heat from cold to hot bodies, thereby defeating the second law of thermodynamics. Maxwell thus made entropy incomplete knowledge, refuting attempts by Clausius and Boltzmann to interpret it dynamically. Later (in 1877) Boltzmann derived a famous equation relating entropy to statistical disorder. In 1929 Leo Szilard sought exorcism through information theory, arguing that the demon's ordering of the world is outweighed by a quantum mechanical disordering within his own brain.

4 Relativity, reality, ether

Maxwell's philosophical position is best described as modified scientific realism (see Scientific realism and antirealism). Fundamental to it is Hamilton's distinction between knowledge and belief. Scientific knowledge is real but also limited, partly of the 'rigid high' kind, partly empirical, complicated by analogy, entangled in relations. A way through is classification, or rather multiplicity of classification. It is instructive to compare Maxwell's youthful essay 'Analogy' (1856; reprinted in Campbell and Garnett 1882: 235-44) with two later works, *Matter and Motion* (1877a) and the short 'Note on the Classification of the Physical Sciences' (1873; published 1884). By the 1870s he had defined a programme of dynamical reductionism which, except in its clarity and philosophical restraint, is strikingly modern. *Matter and Motion* is a masterpiece of natural philosophy,

notable especially for introducing into physics the term *relativity* in a passage that combines strenuous scientific insight with a mystical awareness worthy of the fourteenth-century author of *The Cloud of Unknowing*.

Relativity now connotes Albert Einstein. But the issues originated with Maxwell. He suggested the famous Michelson-Morley ether-drift experiment, and his equations necessitate the Lorentz transformation. About the ether Maxwell's problem was again knowledge versus belief. He accepted the ether; he did not finally describe it. His theory is a relational one. For him the ether had also another kind of relativity: it belonged to ' the science of 1876 (which may not agree with that of 1896)'. ([Letter to Bishop C.J. Ellicott, 22 November 1876] Campbell and Garnett 1882: 393) (see Relativity theory, philosophical significance of).

C.W.F. EVERITT

List of works

Maxwell, J.C. (1871) *Theory of Heat*, London: Longmans & Co.(This textbook, which went through 11 editions, contains the first account of Maxwell's demon.)

Maxwell, J.C. (1873) *Treatise on Electricity and Magnetism*, Oxford: Clarendon Press, 2 vols.(The classic work.)

Maxwell, J.C. (1877a) *Matter and Motion*, London: Society for Propagation of Christian Knowledge.(Extremely illuminating both about Maxwell and about physics in the 1870s. Had a strong influence on Poincaré.)

Maxwell, J.C. (1877b) 'Review of "A Treatise in the Kinetic Theory of Gases" by Henry William Watson', *Nature* 18: 242-6.(Reprinted in Garber, Brush, and Everitt, 1995.)

Maxwell, J.C. (1881) *Elementary Treatise on Electricity*, Oxford: Clarendon Press.(Contains an interesting discussion of scientific analogy.)

Maxwell, J.C. (1884) 'Note on the Classification of the Physical Sciences', *Encyclopaedia Britannica*, vol. 19, 1-3.(Written in 1873.)

Maxwell, J.C. (1890) *The Scientific Papers of J. Clerk Maxwell*, Cambridge: Cambridge University Press, 2 vols.(Contains all but fifteen of Maxwell's published papers.)

Maxwell, J.C. (1990, 1995) *Scientific Letters and Papers of James Clerk Maxwell*, vol. 1, *1846-62*, vol. 2, *1862-73*, ed. P. Harman, Cambridge: Cambridge University Press.(Well-edited version of a large selection of Maxwell's previously unpublished writings.)

Garber, E., Brush, S.G. and Everitt, C.W.F. (eds) (1986) *Maxwell on Molecules and Gases*, Cambridge, MA: MIT Press.(See 1995.)

Garber, E., Brush, S.G. and Everitt, C.W.F. (eds) (1995) *Maxwell on Heat and Statistical Mechanics*, Bethlehem, PA: Lehigh University Press.(These two works contain all of Maxwell's published and known unpublished writings in this area, with fairly detailed commentary.)

References and further reading

Brush, S.G. (1972) *The Kind of Motion We Call Heat*, Amsterdam: North Holland. (Covers in technical detail the history of the kinetic theory of gases.)

Buchwald, J. (1985) *From Maxwell to Microphysics*, Chicago, IL: University of Chicago Press.(A thorough treatment of one group of Maxwell's followers. The treatment of the electric current is controversial.)

Campbell, L. and Garnett, W. (1882) *Life of James Clerk Maxwell*, London: Macmillan.(In its account of Maxwell's first twenty-five years one of the most brilliant Victorian biographies. The later parts are less satisfactory.)

Everitt, C.W.F. (1975) *James Clerk Maxwell, Physicist and Natural Philosopher*, New York: Charles Scribner's Sons.(Brief technical account. Contains a detailed bibliography to 1975.)

Hendry, J.M. (1986) *James Clerk Maxwell and the Theory of the Electromagnetic Field*, Bristol: Adam Hilger.(A historical critical overview.)

Hertz, H. (1893) *Electric Waves*, London: Macmillan.(Besides discussing the discovery of radiowaves contains a historically important review of Maxwell's electromagnetic theory.)

Hesse, M.B. (1961) *Forces and Fields*, London: Nelson.(A standard work.)

Hesse, M.B. (1963) *Models and Analogies in Science*, London: Sheed & Ward.(A wide-ranging discussion of scientific analogy with references to Maxwell.)

Hunt, B.J. (1991) *The Maxwellians*, Ithaca, NY: Cornell University Press.(An account of three key Maxwellians: G.F. FitzGerald, Oliver Heaviside, Oliver Lodge.)

Siegel, D.M. (1991) *Innovation in Maxwell's Electromagnetic Theory: Molecular Vortices, Displacement Current and Light*, Cambridge: Cambridge University Press.(Masterly, if controversial, exposition of Maxwell's second

paper on electromagnetism which revealed the connection with light.)
Whittaker, E.T. (1953) *History of the Theories of Aether and Relativity*, London: Nelson, 2 vols.(An indispensable source for nineteenth- and twentieth-century physics at a high intellectual level. Regrettably, the chapter on Maxwell is weak.)

McTaggart, John McTaggart Ellis (1866-1925)

McTaggart was one of the last of the 'British Idealists', the group of British philosophers, such as B. Bosanquet and F.H. Bradley, who took their inspiration from Hegel. In his early writings from the 1890s, McTaggart gave a critical exposition of themes from Hegel's logic before advancing his own distinctive idealist positions concerning time, the mind, and reality in general. But in his writings from 1910 he developed an independent account of the structure of existence from which he then argued for the same idealist positions as before.

The thesis for which McTaggart is now most famous is that of the unreality of time; what is even more difficult to come to terms with is his thesis that the ultimate reality of the world comprises a community of selves wholly constituted by their loving perceptions of each other. This thesis is a manifestation of a mysticism that is an essential element in McTaggart's philosophy; yet this mysticism is combined with a rationalist determination, reminiscent of Spinoza, to vindicate mystical insights by the light of pure reason alone.

1 Hegelian dialectic

John McTaggart Ellis McTaggart's curious name arises from the fact that his father, Francis Ellis, having named his son 'McTaggart' after an uncle of that name, then took the name himself as a surname in order to inherit his uncle's wealth, and thereby passed the name again to his son. McTaggart studied at Cambridge and then taught there throughout his life. In his first work, *Studies in the Hegelian Dialectic* (published in 1896 but largely based on his 1891 Fellowship Dissertation), he defends the use of the Hegelian dialectic as the primary tool of philosophical argument. According to McTaggart, the great strength of Hegel's dialectic is that, by starting from minimal presumptions about the concepts to be applied on the basis of experience, it is possible to unravel significant conclusions concerning the structure of that which experience reveals - reality itself (see Hegel, G.W.F.).

McTaggart's emphasis on the dialectical method is indicative of the fact that he seeks to base his metaphysics on premises that are in a broad sense 'logical'. But he also thinks that there is something essentially mystical about the truths of metaphysics. These two fundamental presumptions of his philosophy are held together by the fact that he holds that the dialectical method itself leads one to recognize the fundamental reality of a domain which transcends ordinary experience. Following Hegel, McTaggart calls that which is thus fundamentally real 'The Absolute', but goes on to make it clear that his detailed conception of the Absolute is not that of Hegel. Thus he insists that the Absolute is not the culmination of a historical process, since the dialectic is not a temporal process whereby the Absolute realizes itself in the world - it is only a method of reasoning whereby we make explicit to ourselves what has *always* been the only fundamental reality. Instead, McTaggart argues in his *Studies in Hegelian Cosmology* (1901), the Absolute comprises a community of selves wholly occupied in loving perceptions of each other; for only within a structure of this kind are the formal requirements of the dialectic fully satisfied. It is not easy to follow McTaggart's reasoning on this issue, and even more difficult to grasp how one can suppose that the loving perceptions of such a community are, in fact, absolutely the only things that there really are. But the challenge of explaining himself on these issues is, in effect, that to which McTaggart dedicated the rest of his life (see Absolute, the §1).

2 Time

The thesis of the unreality of time is one that McTaggart endorses in his early Hegelian writings. In 1908, however, he published a paper on 'The Unreality of Time' in which he argued for this thesis from an independent starting-point.

He starts from the fact that we have two ways of indicating the time of an event such as the Battle of Hastings: we can either just indicate its relation to present experience, by noting that it is in the past; or we can give its date - 1066 AD - and thereby indicate its relationship as earlier or later than the event which is taken as the origin for the system of dates. Where events are ordered in accordance with a system of tenses McTaggart says that they form an 'A-series'; where they are ordered by a system of dates he says that they form a 'B-series'. The first part of his argument is intended to establish that the A-series is more fundamental than the B-series to the conception of such series of events as a series ordered in time. McTaggart argues first that change is essential to time and that while

the A-series conception includes change (in so far as that which was future becomes past), the B-series relations are immutable; second, he argues that time is essentially directed from past to future in a way which involves the A-series but not the B-series. The second part of McTaggart's argument revolves around the claim that the A-series does not after all provide a coherent temporal ordering of events: for the way in which the A-series allows for change leads to the contradiction of supposing that one and the same event is both past and future. Hence, McTaggart concludes, neither A-series nor B-series yields a coherent conception of a temporal ordering of events, and we must conclude that our very conception of events as ordered in time is illusory. In reality, the events are ordered in some other way (as the 'C-series'), and it is this real ordering which we erroneously perceive as a temporal ordering.

McTaggart's critics maintain that the appearance of contradiction inherent in the A-series arises only because McTaggart moves illegitimately from the coherent thought that the Battle of Hastings *was* future and *is now* past to the contradictory thought that it *is* (tenselessly) both future and past. McTaggart was aware of this response, but argued that the use of tenses here merely postpones the difficulty; for not only was the battle future, it was also present and past, so that the same contradiction arises even when tenses are used. Or does it? McTaggart's critics urge that it does not; and, further, that an A-series theorist need never assent to whatever combinations of tenses are incompatible.

Other critics seek to defend the status of the B-series. They argue that although the B-series does not allow for changes in the truth-value of judgments concerning temporal order, it does not follow that no conception of change can be included in such a scheme. Thus Russell (1903) urged that we can say that an object *a* has changed where the judgment that *a* has the property *F* at one date is true and the judgment that *a* has the property *F* at some different date is false. This is perhaps too simple, for it implies that things themselves change when, for example, people alter their attitudes to them; but if some restriction is placed on the relevant properties *F* that occur here (for example, that they be such that losing or gaining such a property has immediate effects for the thing in question), Russell's account looks plausible. Similarly, McTaggart's other claim, that the B-series does not yield an account of the direction of time, is disputed by those who argue that the temporal structure of causation gives time its unique direction.

Whether one wants to be an A-series or a B-series theorist, then, it looks as though there are positions that are preferable to McTaggart's denial of the unreality of time. Perhaps the most commonly held position these days embodies a compromise: it is argued that McTaggart's mistake lay in not recognizing that these two conceptions of time are interdependent - in that the A-series captures the tensed temporal perspective of the subject which is an essential ingredient of practical thought, whereas the B-series captures the tenseless truth-conditions of these A-series judgments, which cannot be coherently captured in A-series terms (seeMellor 1981 for a classic statement of this position) (see Time §1).

3 Idealist metaphysics

McTaggart's 1908 discussion of time shows him already arguing for his idealist conclusions without direct reference to Hegelian dialectic. In 1910 he completed a critical study of Hegel's logic, *A Commentary on Hegel's Logic*, which, although not at all hostile to Hegel's method, revealed substantial defects in the details of his arguments. Hence, McTaggart realized, if he was to secure the idealist conclusions of his earlier work, he needed to find new foundations for them. This was the task that occupied him for the rest of his life, and which he brought to fruition in his final, two-volume work, *The Nature of Existence* (1921, 1927).

In the first volume, McTaggart develops three fundamental ontological principles which, he argues, are a priori conditions of any acceptable metaphysics. McTaggart's account of these principles is obscure, but we can identify them as: (1) the identity of indiscernibles; (2) the infinite divisibility of substances; (3) the principle of 'ontological determinacy' - that is, the principle that the properties possessed by a substance must be fully determinate. In the second volume, McTaggart then argues that these principles can only be satisfied if his previous idealist convictions are correct (see Identity of indiscernibles).

McTaggart's line of thought is best approached by considering why he thinks that material substances fail to satisfy his a priori conditions. His claim is that once one has satisfied conditions (1) and (2) by supposing that a material substance is infinitely divisible into ever smaller parts, each of which differs descriptively from any other

part, one will violate condition (3) since it will no longer be wholly determinate, at any stage in the division, just what properties a substance possesses - for the precise determination of these properties will depend on the properties of its distinct parts *ad infinitum*. There is much here that might be questioned, including each of McTaggart's principles; but it is worth noting in passing that although McTaggart denies the reality of matter ('the result is that matter is in the same position as the Gorgons or the Harpies'), he is quite explicit that he does not dispute the propriety of common-sense talk of hands and so on. So, *pace* G.E. Moore (1939), he cannot be refuted by a simple appeal to the propriety of such talk.

It remains to be seen why McTaggart thinks that a community of at least two selves does satisfy his a priori conditions. The key point is that he takes it that condition (2) is satisfied because perceptions are 'parts' of a self, and the possibility of indefinitely reiterated perceptions of perceptions constitutes, therefore, the possibility of infinite 'division' of a self. Because these ever more complex perceptions are generated by reiteration, condition (1) is satisfied; and he holds that condition (3) is also satisfied because the content of simpler perceptions does not depend on that of the more complex perceptions of perceptions which are their 'parts'. Solipsism is ruled out in favour of a community of selves and their mutual perceptions of each other because a self whose only 'part' was its perception of itself would not have a proper part at all and would therefore violate condition (2); whereas once two or more selves are introduced, each self will have perceptions of the other as well as of itself and thereby introduce a structure of mutually self-reflecting perceptions of perceptions which McTaggart calls 'determining correspondence' and which, he thinks, is the only structure that can satisfy his a priori conditions.

4 Error and love

Because McTaggart's account of this community of selves and their perceptions of each other is supposed to be an account of the only things that there really are, he holds that our ordinary experience is almost entirely erroneous or misunderstood. It is not just that we misconstrue our perceptions of our hands as perceptions of a material world; we also radically mistake the nature of our own experience. We seem to ourselves to have mental states other than perceptions; but in fact we have *only* perceptions - and these perceptions are *only* perceptions of selves and their perceptions. How, then, is this massive error possible?

The conceptual possibility of massive error of this kind is allowed for by the 'perceiving as' idiom: what are in fact just perceptions of selves and their perceptions are perceived as material objects, choices and the like. Furthermore, according to McTaggart, the illusion of time is the ground of these errors. This illusion rests on our misperceiving the ordering of the members of the timeless C-series (which can only be selves and their perceptions) as a temporal ordering. McTaggart expends much ingenuity in setting out the structure of this C-series and suggesting how, once the illusion of time arises, other errors follow; but he acknowledges that he can give no reason why the illusion arises in the first place (in *Studies in the Hegelian Dialectic* he had observed that even the appearance of evil and error poses a deep challenge to the Hegelian dialectic; the deep reason for his abandonment of that method was probably a growing conviction that this challenge is unanswerable).

What, finally, of love? McTaggart's basic thought is that love is the emotion which manifests a sense of intimate union with a self; and since perceptions of ourselves and others provide precisely this sense of intimate union, the community of mutual perceivers is also a city of love. But there is also an element of residual Hegelian teleology here. McTaggart acknowledges that our ordinary perceptions of each other rarely, if ever, give rise to the appropriate sense of intimate union, but he looks forward to a kind of immediate awareness of others which resembles our awareness of ourselves. This possibility requires, however, an extension to our lives, a kind of immortality whereby we exist indefinitely into the future - or rather throughout the C-series, which we mistake for time. At the final stage of this series, McTaggart holds, all perceptions will be clear and distinct; hence here at last the city of love will be realized. We are not now at this stage, for the C-series is real and progress along it appears to us as the passage of time; but in the final ecstatic sentences of *The Nature of Existence* McTaggart looks forward to this ultimate consummation:

> We know that it is a timeless and endless state of love - love so direct, so intimate, and so powerful that even the deepest mystic rapture gives us but the slightest foretaste of its perfection.

(1927: 479)

THOMAS BALDWIN

List of works

McTaggart, J.M.E. (1896) *Studies in the Hegelian Dialectic*, Cambridge: Cambridge University Press.(A critical discussion of the metaphysical significance of Hegel's dialectic.)

McTaggart, J.M.E. (1901) *Studies in Hegelian Cosmology*, Cambridge: Cambridge University Press.(Discusses topics in ethics and political theory and argues that the Hegelian Absolute is the city of love.)

McTaggart, J.M.E. (1906) *Some Dogmas of Religion*, London: Arnold; 2nd edn, 1930, with intro. by C.D. Broad. (McTaggart rejects conventional religion but defends immortality and mysticism.)

McTaggart, J.M.E. (1908) 'The Unreality of Time', *Mind* 17: 457-74.(McTaggart's first statement of his classic argument for the unreality of time; the argument is slightly modified in McTaggart (1927).)

McTaggart, J.M.E. (1910) *A Commentary on Hegel's Logic*, Cambridge: Cambridge University Press.(A critical study of the details of Hegel's dialectic.)

McTaggart, J.M.E. (1921) *The Nature of Existence*, vol. 1, Cambridge: Cambridge University Press. (McTaggart sets out his a priori conditions for the existence of substances.)

McTaggart, J.M.E. (1927) *The Nature of Existence*, vol. 2, ed. C.D. Broad, Cambridge: Cambridge University Press.(Published posthumously; McTaggart argues that only his idealism satisfies the a priori conditions of vol. 1.)

McTaggart, J.M.E. (1934) *Philosophical Studies*, ed. S.V. Keeling, London: Arnold.(A valuable collection of McTaggart's papers, including McTaggart (1908).)

References and further reading

Broad, C.D. (1933, 1938) *Examination of McTaggart's Philosophy*, Cambridge: Cambridge University Press, 2 vols.(A commentary on *The Nature of Existence*; mostly tedious but sometimes helpful; also includes some intrinsically interesting discussions.)

Dickinson, G.L. (1931) *J. McT. E. McTaggart*, Cambridge: Cambridge University Press.(A personal memoir which conveys McTaggart's strange personality.)

Dummett, M.A.E. (1960) 'A Defence of McTaggart's Proof of the Unreality of Time', *Philosophical Review* 69: 497-504; repr. in *Truth and Other Enigmas*, London: Duckworth, 1978.(An idiosyncratic defence of McTaggart.)

Geach, P.T. (1979) *Truth, Love and Immortality*, London: Hutchinson.(A critical exposition of *The Nature of Existence*; much shorter and more helpful than Broad (1933, 1938).)

Mellor, D.H. (1981) *Real Time*, Cambridge: Cambridge University Press.(Defends McTaggart on the unreality of the A-series and expounds the standard A/B compromise.)

Moore, G.E. (1939) 'Proof of an External World', *Proceedings of the British Academy* 25: 273-300; repr. in G.E. Moore: *Selected Writings*, ed. T. Baldwin, London: Routledge, 1993.(A classic attempt at a refutation of McTaggart's idealism.)

Russell, B.A.W. (1903) *The Principles of Mathematics*, London: Allen & Unwin. (See 469-70 for Russell's account of change.)

Stern, R. (1994) 'British Hegelianism: A Non-Metaphysical view?', *European Journal of Philosophy* 2: 293-321. (Sets McTaggart's early work on Hegel in its historical context, and argues that his account resembles recent 'non-metaphysical' approaches to Hegel.)

Mead, George Herbert (1863-1931)

Together with Charles Peirce, William James and John Dewey, George Herbert Mead is considered one of the classic representatives of American pragmatism. He is most famous for his ideas about the specificities of human communication and sociality and about the genesis of the 'self' in infantile development. By developing these ideas, Mead became one of the founders of social psychology and - mostly via his influence on the school of symbolic interactionism - one of the most influential figures in contemporary sociology. Compared to that enormous influence, other parts of his philosophical work are relatively neglected.

Mead was the son of a Protestant clergyman in Massachusetts. He spent the larger part of his childhood and youth at Oberlin College, Ohio, where his father was appointed professor and where he himself studied. After four years of bread-and-butter employment and intense intellectual struggle with the Darwinian revolution and Kant's moral philosophy, Mead entered graduate study at Harvard and continued at the Universities of Leipzig and Berlin, Germany, specializing in questions of psychology and philosophy. After teaching at the University of Michigan (1891-4), Mead was brought by Dewey to the newly founded University of Chicago where he taught until the end of his life in 1931. Publishing very little, but increasingly influential through his teaching and his life as an activist citizen during the Progressive Era, Mead's reputation has grown since his death. All his major works were published posthumously, based partly on student notes, partly on unfinished manuscripts from his remaining papers.

Mead's intellectual development and general outlook are quite similar to Dewey's: from neo-idealism to a pragmatism oriented to the emerging social-scientific disciplines and the debates about the fate of the democratic ideal in fully industrialized societies. Mead went beyond Dewey in his theory of the specificities of human communication and sociality. Human communication is, according to Mead, superior to animal forms of communication primarily because it operates through 'significant symbols': human beings are able to react to their own gestures and utterances in a way that is anticipatory and thus inwardly represents their fellow 'actors'' possible responses. This makes it possible for their own behaviour to be oriented towards the potential reactions of other actors. As these actors, too, are in possession of this ability, common collective action oriented towards a common binding pattern of mutual behaviour expectations becomes possible. This is for Mead the fundamental feature of all human sociality. 'Taking the role of the other' - one of Mead's famous expressions - is the anticipation of the other's behaviour in a specific situation of interaction.

The possibility of communication through the inner representation of the other's behaviour leads to the formation of different instances in the personality structure of the individual. This is due to the fact that individuals observe and estimate their own behaviour similarly to the way they see the behaviour of partners. The individuals are able to look at themselves from the perspective of the other. Alongside the dimension of impulses there is now an opportunity to evaluate those impulses, an instance which arises out of the expectations of the reactions caused by the manifestations of these impulses. In this context Mead speaks about 'I' and 'me'. The concept of the 'I' designates not only the principle of spontaneity and creativity, but also the endowment of the human being with impulses. 'Me' refers to my own mental presentation of the other's image of me, or, on a more primitive level, to my internalization of the other's expectations of me. The 'me' is both an instance of judgment for the structuring of spontaneous impulses and an element of my emerging self-image. As I face several different persons who have significance for me, I acquire several different 'me' images. These have to be synthesized to make a consistent self-image possible. A successful synthesis generates the 'self' as a unitary self-evaluation and as an orientation for action. Simultaneously, a personality structure develops which is stable and certain of its needs.

Mead's ideas about the development of the self can be found *in nuce* in his theory of children's play. Therein, he draws a line between 'play' and 'game'. 'Play' means the child's playful interaction with an imaginary partner in which the child mimes both parts of the interaction. Thereby the capacity for anticipation of behaviour is given practice: the other's behaviour is represented directly, for example, via imitation, and is complemented by the child's own behaviour. The child has reached this stage as soon as it is able to interact with different persons; that means, when the person with a cathectically high significance is no longer the solely important one. This stage of development is followed by the attainment of the ability to participate in 'games', that is, group activities. In

'games' it becomes necessary that the behaviour of all partners - and not of one single partner - can be taken as the guideline for the child's action. The individual actor has to self orientate to a goal that is valid for all the actors concerned, and which Mead calls, trying to express its psychological bases, the 'generalized other'. In the case of games the generalized other's expectations of behaviour are the rules of the game; in general it means the norms and values of a group, differentiated in a specific way to take into account the specific positions in the group and specific situations. The orientation to a particular generalized other, however, reproduces on a new plane the same restriction which the orientation to a particular concrete other has. Thus follows the problem of an orientation to an ever more universal generalized other.

Mead further developed his ideas about human communication in the fields of general cognitive and moral development. The development of communicative abilities becomes a condition for cognitive progress inasmuch as it is the development of role-taking ability that allows the actors to assume a reflexive attitude towards themselves, and substantial cognitive achievements presuppose just such an attitude. Mead elaborates this idea in his late works, particularly with respect to the problem of the constitution of the permanent object. Its ethical implications are expounded at the individual level as universal role-taking ability and performance, and, at the societal level, in a concept of an ideal society with a universal capacity for communication. During the last decade of his life, Mead was particularly engaged in a dialogue with Dewey and Whitehead's new metaphysics and the role of sociality within it (see Dewey, J. §3; Whitehead, A.N. § 3).

See also: Communication and intention; Communicative rationality; Pragmatism; Pragmatism in ethics §4; Self, Indian theories of; Social action; Social norms; Subject, postmodern critique of the; Symbolic interactionism

HANS JOAS

List of works

Mead, G.H. (1932) *The Philosophy of the Present*, ed. A. Murphy, La Salle, IL: Open Court.(Mead's philosophy of temporality.)

Mead, G.H. (1934) *Mind, Self, and Society*, ed. Ch. Morris, Chicago, IL: University of Chicago Press.(Mead's social psychology.)

Mead, G.H. (1936) *Movements of Thought in the Nineteenth Century*, ed. M.H. Moore, Chicago, IL: University of Chicago Press.(Mead's fragmentary intellectual history.)

Mead, G.H. (1938) *The Philosophy of the Act*, ed. Ch. Morris, with J.M. Brewster, A.M. Dunham and D.L. Miller, Chicago, IL: University of Chicago Press.(Collection of manuscripts from Mead's remaining papers.)

Mead, G.H. (1964) *Selected Writings*, ed. A. Reck, Indianapolis, IN: Bobbs-Merrill.(Collection of some articles Mead published during his lifetime.)

References and further reading

Cook, G.A. (1993) *George Herbert Mead. The Making of a Social Pragmatist*, Urbana and Chicago, IL: University of Illinois Press.(Careful study of Mead's biography and intellectual development.)

Joas, H. (1985) *G.H. Mead. A Contemporary Re-examination of his Thought*, Cambridge, MA: MIT Press. (Comprehensive study of Mead's work, comparing it to Continental traditions in philosophy and social theory in general. The second edition of 1997 contains a new preface)

Miller, D.L. (1973) *G.H. Mead. Self, Language and the World*, Austin, TX: University of Texas Press.(Written by a student of Mead, this book gives a good overview of Mead's work and compares it to the tradition of analytic philosophy.)

Meaning and communication

The two fundamental facts about language are that we use it to mean things and we use it to communicate. So the philosophy of language tries to explain what it is for words and sentences to mean things and also what it is for us to communicate by using them. Although it cannot be accidental that meaning and communication go together, it is quite easy to see them as fundamentally distinct. Thus on some accounts the meaning of sentences is conceived in terms of a 'representative' power whereby they stand for either aspects of the world or ideas in the mind and their use in communication is derived from this property: language serves as a vehicle for meaning, itself thought of in independent terms. An alternative approach seeks to link the two more closely, seeing representation as itself only possible through the use of terms of a common language, used in communication.

At its most primitive, communication may be simply akin to infection, as when one animal communicates fear or hostility to another. Simple signalling is also described as communication, as when bees communicate the direction and quantity of pollen to other bees, or when birds communicate territorial or sexual claims. Here the signaller issues a sign and the recipient modifies its behaviour upon perceiving it; the content of the signal is interpreted (by us) to be whatever goal seems to be served by issuing it: maximizing the success of the hive in obtaining honey, or keeping other birds out of a particular space, for example. Humans communicate in a similar sense, through the use of emotional and other signals, unconscious 'body language' or equally unconscious or semiconscious signs, which may include fashion statements or indications of status, for example. When we communicate successfully we share an understanding. Everything is open between us; nothing is concealed nor taken by one person in one way and another in a different way. With linguistic communication this common understanding can be put in terms of meaning: we each know what the other means. This implies that we understand not only the semantics of the utterance, but also the pragmatics: if the utterance states that something is the case then we know what this is, but we also know if the statement was meant ironically, or condescendingly, or with some other intent (see Pragmatics). It is plausible to think that the basic goal of speech is to achieve such communication, although there will be cases where we speak in order to conceal our thoughts, or to mislead the hearer. People fluent in the same language achieve open communication more or less effortlessly, at least where straightforward messages are concerned. But philosophers of language have not found this phenomenon easy to understand. Issues arising include: the place of intention in communication (see Communication and intention); the place of linguistic convention in communication (see Language, conventionality of); the relationship between shared understanding and activities such as translation and interpretation (see Radical translation and radical interpretation); the pervasive possibility of indeterminacy; and the nature of rules and the extent to which language is essentially social - and indeed the extent to which language is necessary to communication (see Meaning and rule-following). In the analytic tradition, major writings on this subject include works by Wittgenstein, Grice, Lewis, Bennett, Kripke, Searle and Tuomela, while a more continental interest in the field is especially illustrated by Habermas (1984, 1987).

The classical tradition, including Aristotle, Hobbes and Locke, thinks of communication in terms of one party having an idea then using language as a medium for transmitting it to others. The meaning of the linguistic vehicle is then derived from the primitive representative powers of the idea. The analytic tradition has sought to bring our ability to represent things closer to our ability to express what we think linguistically, even to the point of reversing the dependency and making thought itself (or at least the kind of thought that could be expressed in language, unlike perhaps the thought of the graphic artist or the musician) dependent upon its linguistic expression. The question then becomes one of saying how language can generate 'intentionality', which is the power of being 'about' external and often absent situations which it represents truly or falsely (see Intentionality). If, for example, language is essentially a shared and social construction, and if it is necessary to thought, then we derive the result that a born Robinson Crusoe would not be capable of thought, at least until he acquired a social identity. The consequence seems uncomfortable, since it is easy to imagine such a Crusoe giving all sorts of signs of intelligent adaptation to his environment, and conceiving and executing complex plans. For some thinkers, such as John Searle (1983), this is because intentionality resides in our biological nature, in advance of and independent of language. For others, such as Jerry Fodor (1987), it is because we come equipped with an innate representative medium conceived on the model of a natural (social) language, but itself explaining the powers of natural languages: a 'language of thought' (see Language of thought).

Arguably too much of the theory of communication treats the process as one of interpretation or translation, in which the task of the hearer is to theorize about the thoughts intended by some act of communication of the speaker. This makes it seem as if each person is secure in their own 'ideolect', with it being relatively problematic whether the same ideolect is shared by others. This clearly distorts what it is to share a language, which surely entails finding ourselves (surprisingly literally) of one mind about meaning. We have a securely shared common understanding, in which I have no privileged access to my own meanings which is not communicable to others. Philosophers have therefore struggled against the vision of the infant, inducted into this shared language, as a kind of outsider or 'little linguist' busy theorizing (in what medium?) about the potential meanings of the utterances made by surrounding people. The proposed alternative derives from the *Verstehen* (literally 'to understand') tradition of humane sciences (see Dilthey, W.), seeing the task of the infant not so much as one of theorizing as one of imitating and simulating the expressions of others. I achieve the openness of full communication with you, on such an account, when I know what it would be to make your words my own.

SIMON BLACKBURN

References and further reading

Bennett, J. (1976) *Linguistic Behaviour*, Cambridge: Cambridge University Press.(A lively development of Grice's work, especially valuable for its discussion of the relation between full intentionality and more primitive signalling systems.)

Blackburn, S. (1984) *Spreading the Word*, Oxford: Oxford University Press.(An attempt at an accessible discussion of most of the central issues, aimed at moderately advanced students.)

Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.(The classic defence of the theory that meaning is the outcome of computational structures in the mind/brain.)

Grice, H.P. (1957) 'Meaning', *Philosophical Review* 66; repr. in *Studies in the Way of Words*, Cambridge, MA, and London: Harvard University Press, 1989.(The seminal article in the field.)

Habermas, J. (1984, 1987) *The Theory of Communicative Action*, trans. T. McCarthy, Boston, MA: Beacon Press, 2 vols.(An attempt to connect the social nature of language to the virtues of free democratic debate.)

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell. (A particularly accessible presentation of the thesis that meaning is essentially social.)

Lewis, D.K. (1969) *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.(The classic modern analysis of the concept.)

Searle, J.R. (1969) *Speech Acts*, Cambridge: Cambridge University Press.(An accessible development of the work of J.L. Austin on the actions that generate meaning.)

Searle, J.R. (1983) *Intentionality: An essay in the philosophy of mind*, Cambridge: Cambridge University Press. (Searle's work is perhaps the most impressive development of the notion of acts performed in speech, but also controversial through its firm belief that the intentional powers of language are derived from a prior biologically engendered capacity.)

Tuomela, R. (1984) *A Theory of Social Action*, Dordrecht: Reidel.(An influential attempt to see collective action as arising from the individual intentions of the different parties.)

Meaning and rule-following

Wittgenstein's discussion of rules and rule-following, and the recent responses to it, have been widely regarded as providing the deepest and most challenging issues surrounding the notions of meaning, understanding and intention - central notions in the philosophy of language and mind. The fundamental issue is what it is for words to have meaning, and for speakers to use words in accordance with their meanings. In Philosophical Investigations and Remarks on the Foundations of Mathematics, Wittgenstein explores the idea that what could give a word its meaning is a rule for its use, and that to be a competent speaker is to use words in accordance with these rules. His discussion of the nature of rules and rule-following has been highly influential, although there is no general agreement about his conclusions and final position. The view that there is no objectivity to an individual's attempt to follow a rule in isolation provides one strand of Wittgenstein's argument against the possibility of a private language.

To some commentators, Wittgenstein's discussion only leads to the sceptical conclusion that there are no rules to be followed and so no facts about what words mean. Others have seen him as showing why certain models of what it takes for an individual to follow a rule are inadequate and must be replaced by an appeal to a communal linguistic practice.

1 Meaning and rules

Words are meaningful only if there is such a thing as using them correctly or incorrectly. To have determinate meanings they must have application in some situations and not others. Words that had application to just anything would be robbed of meaning and could not be distinguished from one another. But what gives a word meaning? And how does this determine what uses do and do not comply with it? Answers to these questions must be compatible with a credible epistemology of meaning: when we understand a word we must be able to use it in accordance with its meaning. If something were to settle the right or wrong use of a term, we must be able to take cognizance of this. The objectivity of meaning and linguistic judgment depends on this. We have to secure both of these simultaneously.

In his later writings, Wittgenstein rejects the idea of meanings as mental or abstract entities to be associated with particular signs. Instead he takes the meaning of a sign, or word, to be its use in a language. However, in equating meaning with use he does not assume that every way of using a sign can contribute to its meaning. To have a meaning there must be some particular range of application that counts as using a word correctly. But what does the correct use of a word consist in? Wittgenstein explores the claim that a word is used correctly when it is used in accordance with a rule. If there are rules governing the use of words, then there are standards speakers have to meet to deploy words competently. Competence requires speakers to know what counts as applying the word correctly or incorrectly on any occasion. The appeal to rules is also thought to guarantee the range of things a word applies to: the rule can be thought of as settling the application of a term not only to cases considered so far, but also to hitherto undiscovered cases.

These features of rules bring out the normative element in meaning. Given that words have meanings there is something that counts as using them in accordance with these meanings: this is how words *should* be used. When meanings are given by rules of use, competent speakers are required to conform to these rules: a rule does not just describe the use we make of an expression, it says how we *ought* to use it.

To satisfy ourselves that there is some substance to the notion of meaning, we need to account for the nature of these rules and the requirements they impose on our correct use of words. We also need an account of how speakers succeed, if they do, in following these rules. The issues are connected. If no account can be given of what it takes to heed the requirements of a rule, it is hard to credit those rules with any normative force in determining the meanings of words in our language. On the other hand, the idea of us conforming our linguistic practice to the requirements of rules cannot be maintained if we can give no substance to the idea of there being rules to be followed in the first place.

2 Rules and rule-followers

Wittgenstein addresses the issue of what it is to follow one rule rather than another by considering the case of a

learner asked to continue the arithmetical series: 2, 4, 6, 8…. There are infinitely many options for them. They could go on by writing 10, 12, 14, or 10, 14, 18, and so on. All they are shown are finitely many cases and from there they must acquire the ability to continue the series the mathematician intended. To go on from any position to the next they need to know the rule for expanding the series. But how do they acquire that from their exposure to the examples? Wittgenstein reviews a number of models for such knowledge of rules, and rejects them as inadequate.

First, we might think of learners giving themselves an instruction (such as $n = x + 2$), which tells them what to do next. But no instruction is sufficient, since it can only be followed if the learners know what it means. If they know this, then we must ask in what their acting in accordance with the meanings or rules governing *that* instruction consists. Either we appeal to further instructions, thus setting up an infinite regress, or else we suppose they just know what the instruction means in which case we have no explanation of their understanding, and we are no further forward.

Perhaps a learner simply reacts to the presented cases in some way, and either does or does not satisfy the mathematician. But if this learner is simply caused to respond, we are unable to say what it is for them to go by the rule as opposed to merely reacting to instances of it. On a future occasion they could be caused to act in a way at odds with the intended series. Also, their reactions may be fortuitous: they may be fitting the rule but not following it, blind to what is required of them. Only if the learner and the mathematician follow the same rule is the possibility of coincidence neutralized.

The justification of each move, if there is one, will depend on the rule the learner is following. But which rule is this? There are infinitely many rules compatible with the learner’s behaviour so far, and every move they subsequently make can be made to accord with some rule or other. So what reason is there to think, before the learner decides on their next move, that there is some particular rule they are following? Perhaps we do not consult anything to decide how we are using a sign (or expanding a series). Nevertheless, the temptation exists to say that there must be something that determines whether we are going right or wrong in what we decide to do.

We might then say the rule itself determines right and wrong moves, because it consists in the full expansion of the series with all the instances fixed. As Wittgenstein puts it, it is as if all the steps had already been taken. This is a picture of rules as rails stretching ahead of us to infinity, with the rules fixed in some abstract or Platonic realm: what is required of us in any given application of them would thus be constituted wholly independently of human propensities to judge. The difficulty with this picture is that it gives us no idea of how we recognize which standards to conform to. If rules were blankly external to the mind, there would be no way to discover what they require of us. However we used words, whether or not we coincided with these objective standards would be a matter settled independently of us. Such a conception makes it unclear how such rules or standards could inform our practices, and it is hard to see how we could attempt to conform to rules that fall outside our cognitive reach.

The Platonist’s picture mislocates the normative dimension of speech. At best our judicious use of words would be a series of stabs in the dark with which we hoped to speak significantly by coinciding with the mind-independent meaning-facts. This picture may afford the idea of an objective standard for our linguistic judgments, but there would be no difference (from the point of view of the speaker) between observing this standard and following one’s subjective inclination to use the words as one felt fit. Without speakers being apprised of the standards of use that regulate their linguistic practice, meaning-scepticism would arise, putting in doubt both the existence of Platonized rules and the possibility of a speaker using words in accordance with their meanings. It seems as if the trouble with Platonism is that it locates the meaning-determining rules wholly outside the mind of the speaker.

Wittgenstein next turns to the temptation to find some self-interpreting item in the mind which serves to stop the regress of interpretations, and by itself points to the correct application of a word. Wittgenstein tells us that when we review the instructions we might give ourselves for deploying a word, it is as if meaning was the last interpretation. The trouble now is that there are no candidates for this role. If we suppose that the meaning itself is an item in mind, or that we have an intuition of what is required of us on any occasion of use, we need to know how such things as items in consciousness could point to the conditions of application for a word. For if intuition leads us to apply the word in one way, it could just as easily mislead us: we are simply reacting in a way that feels right. No items whose properties we can scrutinize consciously seem to carry the right information about how we should use a word, or comply with a rule in unexamined cases. It is utterly mysterious how anything we

consciously consult could do this. It might be said that knowledge of our intention in using the word will serve here, but this also proves unsatisfactory: we are trying to explain how we can have such an intention in the first place, and how we can tell whether or not we are conforming to it. The appeal to an inner item is thus no better than Platonism. Locating the item inside the mind has still not connected our use of a word to a norm we can appreciate and whose requirements we could take into account in our linguistic use.

We appear to have exhausted the options. Does this mean that when we judge how to use an expression there is nothing that counts as getting it right? To concede this would be to surrender the objectivity of linguistic judgments and settle for an absence of genuine constraints on how to use words. However we go on using a word in future, there would be nothing that counted as acting in accordance with the rules, since there would be nothing for the requirements of the rules to consist in. How could thinking about a rule show me what I was meant to do at this point? There may be a subjective impression of constraint, of going on in the right way. But is there anything this answers to? For any way I chose to use the word in a new case, can be made to conform to some rule or other. There are interpretations that can be made consistent with anything I do, as Wittgenstein tells us in *Philosophical Investigations*: 'any interpretation hangs in the air along with what it interprets, and cannot give it any support. Interpretations by themselves do not determine meanings' (1953: §198). If every way of employing a term can be made to conform to a rule, then 'whatever is going to seem right to me is right. And that only means that here we cannot talk about "right"' (1953: §258). Without a distinction between what 'seems right to me' and what *is* right in my use of language, there can be no objectivity to meaning.

This is the conclusion Saul Kripke sees Wittgenstein as reaching at §201 of *Philosophical Investigations*; a conclusion Kripke, in his influential reading of Wittgenstein, calls the sceptical paradox.

3 Kripke's Wittgenstein

As Kripke (1982) formulates it, the sceptical paradox arises for a subject who has already been using the symbol '+' satisfactorily to compute sums of numbers less than 57. Kripke's example supposes that they are then asked to add 68 and 57. They might answer '125', but what fact about them makes this the right answer? The challenge to the subject is to come up with a fact about previous mental, behavioural or physical history that constitutes meaning *plus* rather than the deviant operation *quus*: an operation which shadows addition until we reach numbers greater than 57 at which point the answers are not the sum of those numbers, but always 5. If they meant *plus* by '+', then they are right; if previously they meant *quus*, then to answer correctly they should have answered '5'. They are only justified in answering '125' if there is a fact about what they mean by '+', a fact about which rule they have been following until now and continue to follow.

For Kripke this is not an epistemological problem about there being certain facts to which we cannot have access. He allows the sceptic to suppose that a subject has perfect recall and access to their past behaviour, conscious experiences, memories and so on, but points out that they are still unable to come up with any fact that constitutes their meaning *plus* rather than *quus*. The failure to do so leads Kripke's sceptic to conclude that there is no fact of the matter about a subject's meaning one thing rather than another. There are simply no facts of the matter concerning meaning. (Though Simon Blackburn (1992) argues that this claim cannot be made in the first person without it being self-defeating since the person must be able to distinguish *plus* from *quus* to appreciate the problem.) Although Kripke takes this damaging conclusion to be the upshot of Wittgenstein's rule-following considerations, he suggests that Wittgenstein tries to show that we can live with the consequences. (For Kripke's presentation of what he takes to be Wittgenstein's sceptical solution, see §4.)

Crispin Wright (1989a) and John McDowell (1984) both challenge Kripke's reading of Wittgenstein. They point out that Kripke's sceptic is forcing us to provide a reductionist answer to the question of what my meaning *plus* rather than *quus* consists in. To substantiate the claim to mean one thing rather than another, we are asked to come up with some fact other than that we mean *plus*, or had the intention to use the word as we previously did. McDowell suggests there is no way to avoid an appeal to the fact that we meant *plus* by our words all along, and he insists that we have to take this as a stopping place for explanation. No justification of this fact can be given: we simply have to remind ourselves of the meanings we know our words to have. This approach has come to be known as quietism (see §5). Wright thinks we have to appeal to speakers' previous intentions concerning their use of the word to settle questions of justification, but unlike McDowell he believes there is a constructive explanation of what it is to know one's intentions and conform to them (see §6).

Wittgenstein's remarks at §201 of the *Philosophical Investigations* certainly suggest a sceptical paradox which gives rise to the dilemma McDowell identifies: either we have an item to provide the interpretation and are saddled with Platonism or a regress of interpretations, or in using words as I do I simply act blindly and there is no fact of the matter about what I mean. According to McDowell, the paradox depends on an assumption to be discharged: namely that meaning something by a word must be a matter of attempting to give it an interpretation. In the second half of §201, Wittgenstein goes on to suggest that there must be a way of meaning something that is not a matter of interpretation. This recommends a different way of construing the presence of rules in linguistic behaviour. In following a rule, the individual does not search for an interpretation. He just acts without giving himself supporting reasons. But unlike Kripke's reading, this does not lead to the conclusion that there is no fact of the matter about what a speaker means. We are told at §289, that 'To use an expression without justification is not to use it without right.' What is needed is an alternative account of what it is to act in accordance with a rule and to use our words correctly.

This is where the appeal to the community comes in. Rather than train our sights on facts about an individual, following a rule is a matter of participating in a communal practice. This retains the idea that there must be something that counts as acting in accordance with a rule or going against it on any occasion, but now conceives of it as a matter of obeying a practice or going against it. The idea of rules as consisting in communal practices leaves room for a number of different treatments of the way an appeal to the community can help here. These are reviewed in §4 as attempted solutions to the sceptical paradox, after discussion of another popular response to Kripke: the dispositional response.

4 Attempted solutions

Dispositional response. The dispositionalist attempts to respond to Kripke's sceptic by insisting that while no occurrent facts about a speaker's behaviour or mental history determine what they meant by their words, their disposition to respond in particular ways to cases when confronted by them settles the matter. The dispositionalist is here relying on the fact that when a thinker is asked what '$68 + 57$' is, they will be disposed to reply '125'. However, the disposition of the subject can be tested in only finitely many cases, whereas addition is defined over an infinite range. Also, the subject may not have the disposition to respond in the case of very large sums where the computations are difficult, and they may be disposed to make mistakes. These cases could be dealt with by pointing to a different set of dispositions to work out the sum on paper and to check the results. However, Kripke's main objection to the dispositionalist's solution is that it loses sight of the normative element of meaning. In each case, the dispositionalist tells us how we *would* respond but not how we *should* respond. If I am using a word in accordance with its meaning, there is a way in which I am obliged to use it if I want to apply it correctly. Kripke tells us that the dispositionalist simply equates correctness and performance.

Appeal to the community. One can perhaps resist the conclusion that there is no fact about what a speaker means by their words by bringing in a background of other users against which the use of a word can be measured. We could say that the speaker does attach a certain meaning to a word, if their use of it coincides with the way the majority of people in their linguistic community use it. This response parallels the dispositional response, except that it considers the dispositions of everyone in the linguistic community. The notion of correctness is established in this case not by appeal to a rule that the individual is following, but by appeal to a widespread pattern of use exemplified by fellow language-users. In so far as there is a fairly widespread and consistent pattern of use for a word, the individual will count as using a word correctly just in such cases as their use of the term coincides with this pattern. They need not know that it coincides, or even check that it does. On this view, a speaker's knowing what they mean will not be the effortless and direct matter it is for most people. For this speaker it will be a matter of determining their fit with the uses other people make of these words. We may submit to correction when a number of people point out how our use of a word diverges from theirs, but we can be offered no determinate advice on what is required of us in the future if we are to continue conforming to majority use. For at the community level, there is no fact of the matter about what we mean collectively or what rules we are following (see Wright 1980). The strategy establishes for each individual language-user a sense of their using words correctly or incorrectly, and to that extent determines some objectivity for their linguistic judgments, but the community itself does not go right or wrong, it simply goes. Such a strategy puts in doubt the existence of meaning for shared public languages. And as far as individuals are concerned, it does nothing to guide their use of language.

They will merely coincide or fail to coincide with their fellow language-users.

Kripke's sceptical solution. Kripke's sceptical solution invokes a different appeal to the community. Unlike a straight solution this in effect concedes that there is no answer to the sceptical paradox, but attempts to show us that we can live with the consequences. Kripke maintains that although there are no meaning-facts to settle how my words *should* be used, I will go on using them like other people and be subject to correction by them. Claims about what my words mean will lack truth-conditions but, Kripke insists, they will have assertion-conditions: conditions under which people will be prepared to assert that someone is using a word in conformity with its meaning. There will be nothing about someone's use of a word to which this claim is answerable, but there will be uses we have for such statements that do not require their truth. There is no such thing as individuals taking themselves to mean something in isolation from the community, since there would be no background consensus to establish even assertion-conditions for claims about what they mean. The consensus depends on a community, so the appeal to the community rules out the possibility of someone using a private language.

The sceptical conclusion has been resisted by those who maintain that it is part of Wittgenstein's aim to defuse rather than accept the sceptical paradox. The paradox arises, it is claimed, because of a mistaken understanding of what is required to follow rules competently.

5 McDowell's quietism

As long as there is something that counts as obeying a practice or going against it on any occasion, there is such a thing as correctly following a rule. Rules will no longer be thought of a something I consult inwardly, or by reference to which I interpret what is required of me. Rather, they will simply be enshrined in our practices - the normative practices which inform our linguistic behaviour and impose constraints on the meaningful use of our terms. Individuals follow rules when they participate in communal practices. It has to be by reference to this notion of practice, that our judgments about how to use words come to have objectivity. So what is a practice and how does it provide for the idea of going right or wrong on any occasion?

The simple idea of *de facto* conformity with others cannot serve here. For according to that conception, being right in our use of a word is just a matter of whether we find ourselves in agreement with the majority, and this threatens to deprive the picture of its normative dimension (see §4). How can whether I am in step with my community determine how I should go on using and understanding a term? How can this show me how I ought to use language to be faithful to the meanings of my words?

What we need is to restore the idea of a community of speakers who operate according to the norms of meaning. John McDowell (1984) insists that such a picture is available to us when we observe Wittgenstein's warning not to try to dig below bedrock (the level, Wittgenstein says, at which justifications give out). We can dig down to the ground to find justifications but we will reach a level where no further explanation can be given by reference to what goes on at the level below. This is bedrock. McDowell characterizes bedrock for our linguistic practices as a level of activity at which norms are still in place, and where we must describe one another's linguistic behaviour in the meaningful terms available to us as speakers of the language in question. We will have to describe people's activities as 'saying that *such and such*' - where the words we use to describe what is said (*such and such*) are the very words used by the speaker. This kind of description will not be intelligible to those outside the practice.

This is a resolutely anti-reductionist approach to meaning and rule-following, and we are told that we can expect no explanations of meaningful speech and linguistic practice by reference to something else. We simply have to remind ourselves how familiar we are with the meanings of our words and what they require of us in so far as we participate in these normative practices. No explanation can be given of their significance, and none is called for. This position known is as quietism.

6 Wright's extension-determining approach

Crispin Wright (1989a) opposes both quietism and the sceptical paradox, believing that there is room to show what the rule-governed meaning-facts consist in. Wright accepts the challenge to show how it can be that the requirements of rules governing our words are not simply up to us, while it is also true that they are not constituted wholly independently of us. The aim is to avoid surrendering the objectivity of meaning while at the same time bringing it within epistemic reach.

Wright argues that what constitutes a correct application of a word, including its application to a new case, and how it was used previously, is not something that can be settled independently of human judgment. A contribution from us is required to settle the matter. Wright avoids a slide to subjectivism by claiming that our best opinions about what we previously meant or intended will settle the matter, and since not every opinion we form is best it is not true that whatever we judge to be right is right. Only those judgments we make in the most propitious conditions for judging go into determining the right use of a term. The key point is that our linguistic judgments under these conditions *determine* rather than *reflect* the correct application of our terms. These judgments about how to use words do not involve a tracking epistemology: we are not keeping up with Platonic standards or internalized instructions. According to Wright, if we can specify cognitively ideal circumstances for judging in a substantial and non-circular way, then there is nothing but our opinions in those circumstances to decide what we mean by these words. In this conception, meanings are like secondary qualities. Objections to this account question the plausibility of giving a non-circular specification of the ideal circumstances for judging meanings or conditions of application.

See also: Private states and language; Language, social nature of

BARRY C. SMITH

References and further reading

Blackburn, S. (1992) ’Theory, Observation and Drama’, in *Mind and Language* 7: 187-203.(Aims to show how appeal to the first-person case can defeat Kripke’s sceptic. Referred to in §3.)

Boghossian, P. (1989) ’The Rule-Following Considerations’, *Mind* 98: 507-49.(A good summary of recent work on this topic taking up some issues not dealt with here. Contrasts with these discussions in trying to separate the issue of rule-following from the issue of the speaker’s knowledge of the rules he is following.)

Budd, M. (1984) *Wittgenstein’s Philosophy of Psychology*, London: Routledge. (Chapter 2 provides a clearly written and careful treatment of Wittgenstein’s views on meaning and understanding.)

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell. (Provides a lucid, highly accessible introduction to the problems. Referred to in §§3-4.)

McDowell, J. (1984) ’Wittgenstein on Following a Rule’, *Synthèse* 58: 325-63. (Thorough treatment of the issues discussed, including the position outlined in §§3 and 5. Also provides both an interpretation of Wittgenstein’s thought and a critical discussion of Kripke and Wright.)

McGinn, C. (1984) *Wittgenstein on Meaning*, Oxford: Oxford University Press.(A clear and concise treatment of rule-following and the private language argument.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(The principal source of the issues discussed here; see especially §§185-289, referred to in §§1-3. Notoriously subtle and elusive writings.)

Wittgenstein, L. (1956) *Remarks on the Foundations of Mathematics*, Oxford: Blackwell. (Section 6 includes key passages on rule-following. Intriguing but not straightforward. Referred to in §2.)

Wright, C. (1980) *Wittgenstein On the Foundation of Mathematics*, London: Duckworth.(See chapters 2 and 11 for Wright’s early views on rule-following, mentioned in §4.)

Wright, C. (1989a) ’Wittgenstein’s Rule-following Considerations and the Central Project of Theoretical Linguistics’, in A. George (ed.) *Reflections on Chomsky*, Oxford: Blackwell.(Difficult though rewarding reading, setting out the account described in §§3 and 6.)

Wright, C. (1989b) ‘Critical Notice of *Wittgenstein on Meaning* by Colin McGinn’, *Mind* 98: 289-305.(Section 3 contains a penetrating discussion of the principal themes and arguments from the key passages in Wittgenstein’s *Philosophical Investigations* (1953) and *Remarks on the Foundations of Mathematics* (1956).)

Meaning and truth

Analytic philosophy has seen a resurgent interest in the possibility of explaining linguistic meaning in terms of truth, which many philosophers have seen as considerably more tractable than meaning. The core suggestion is that the meaning of a declarative sentence may be given by specifying certain conditions under which it is true. Thus the declarative sentence 'Venus is red' is true just in case the condition that Venus is red obtains; and this is exactly what the sentence means.

As it stands, however, this suggestion provides us with no explanation of the meanings of the words and phrases that make up sentences, since in general they are not expressions that have truth-conditions. (There are no conditions under which the word 'Venus' is true.) Furthermore, it needs to be supplemented by some method of circumscribing the truth-conditions that embody the meanings of declarative sentences, since there are many conditions under which any given sentence is true: 'Venus is red' is true not merely when Venus is red, but also, for example, when Venus is red and 7 + 5 = 12; but it does not mean *that Venus is red and 7 + 5 = 12.*

Evidently the first problem can be solved only by finding other semantic properties which indicate the meanings of words and phrases. For example, it is sometimes thought that the meaning of a name can be specified by saying what it refers to; and that of a predicate by saying what it is true of. But notice that since the meaning of a declarative sentence can be grasped by first grasping the meanings of its basic components, meaning-indicating ascriptions of semantic properties to those components must entail a meaning-indicating statement of its truth-conditions. Semantic properties such as 'referring to' and 'being true of' satisfy this requirement, at least in the context of what is sometimes called a 'truth theory' for a language.

This still leaves the problem of how to circumscribe the right meaning-indicating statement of truth-conditions for declarative sentences. Indeed we now have a further problem. For if the meanings of the components of sentences are not stated directly, but merely in terms of what they refer to or are true of (say), then we must also find a way of determining which of the many ways of specifying what they refer to, or the conditions under which they are true of something, is meaning-indicating. These problems may arguably be solved by placing an appropriate truth theory for a language in a setting that allows us to appeal to the general psychology of its speakers.

Attempts to elucidate meaning in terms of truth-conditions induce a plethora of further problems. Many are a matter of detail, concerning the kinds of properties we should associate with particular idioms and constructions or, equivalently, how we are to produce truth theories for them. As a result of Tarski's work, we have a good idea how to do this for a wide range of categories of expressions. But there are many which, superficially at least, seem to resist straightforward incorporation into such a framework. More general difficulties concern whether truth should be central at all in the analysis or elucidation of meaning; two objections are especially prominent, one adverting to antirealist considerations, the other to the redundancy theory of truth.

1 The core suggestion and some problems

Three of the most influential approaches to the understanding of linguistic meaning in contemporary analytic philosophy correspond to three different, though overlapping, roles of language: its role in communication; its role in reasoning; and its role in making statements. According to the first of these approaches, linguistic meaning should be analysed or elucidated in terms of the communication intentions of speakers (see Communication and intention; Meaning and communication). According to the second, it should be characterized in terms of the conceptual or inferential role of expressions of the language (see Semantics, conceptual role). And the third, at least in the form in which it has been most fully articulated, requires meaning to be analysed or elucidated in terms of the truth-conditions of sentences whose characteristic role, at least in context, is to make statements. It is the third approach which is the subject of this entry.

The core suggestion (which we owe to Frege (1964)) is that the meaning of a declarative sentence is given by certain conditions under which it is true. Thus 'Venus is red' is true just in case Venus is red; and that Venus is red is precisely what the sentence means. However, this fact leaves us in the dark about the meanings of expressions which are not sentences. For example, none of the words that makes up the sentence 'Venus is red' is the kind of expression that can have truth-conditions. Furthermore, there are countless other conditions under which the

sentence is true that do not reveal its meaning - that Venus is red and $7 + 5 = 12$, for example. So the core suggestion must be supplemented not only by some way of articulating word meaning, but also by a method of picking out those truth-conditions of declarative sentences that embody their meanings - a method which does not make reference to meaning itself. I shall call the problem of articulating word meaning ‘Problem A’, and the problem of picking out the right truth-conditions of sentences ‘Problem B’.

Let us begin with Problem A. It is clear that to solve it we must find further semantic properties which can plausibly indicate the meanings of non-sentential expressions, just as particular truth-conditions are meant to indicate the meanings of declarative sentences. For example, it is plausible to claim that the meaning of a name can be specified by saying what it refers to, and that of a predicate by saying what it is true of: the meaning of the name ‘Venus’ may be specified by stating that its referent is Venus; and that of the predicate ‘is red’ by stating that it is true of something if and only if that thing is red.

However, two difficulties arise with this proposed solution to Problem A. The first - which I shall call D(i) - is an analogue of Problem B. This is that there are many ways of stating the referent of a name, and many ways of stating what predicates are true of, but not all of them are meaning-indicating. Thus, although we may say that the referent of ‘Venus’ is Venus, we may also say that it is Hesperus, or the thing that is both Venus and either green or not green; but only the first statement of its referent specifies its meaning. Equally, although we may say that ‘is red’ is true of something if and only if that thing is red, it is also true of something if and only if the thing is red and $2 + 2 = 4$, and so on; but again only the first biconditional specifies its meaning.

The second difficulty, D(ii), arises from the fact that anyone who understands the components of a given sentence is generally in a position to understand the sentence itself; no further information is needed. For example, if I understand both ‘Venus’ and ‘is red’, I can immediately go on to understand the sentence ‘Venus is red’. Now this is tantamount to saying that the meanings of sentences are determined by the meanings of their components (see Compositionality). Thus the meanings of ‘Venus’ and ‘is red’ determine that of ‘Venus is red’. But if this is right, then any correct specifications of the meanings of a sentence’s components should entail meaning-indicating statements of the truth-conditions of the sentence itself. In particular, the specifications of the meanings of ‘Venus’ and ‘is red’ should entail the meaning-indicating statement of the truth-conditions of ‘Venus is red’, namely that ‘Venus is red’ is true if and only if Venus is red. (If not, the putative meaning-specifications of ‘Venus’ and ‘is red’ must have left something out.) The problem, however, is that even in this simple case it is not immediately clear how to ensure this.

2 Truth theories

It is convenient to tackle difficulty D(i) after we have dealt with D(ii). We can then deal with D(i) along with Problem B (the corresponding difficulty for sentences). Turning to D(ii), then, how are we to ensure that correct meaning-specifications for a sentence’s components entail a correct specification for the sentence itself? Here it is customary to appeal to what is often called a ‘truth theory’ for a language. For it is precisely by employing such a theory that statements of truth-conditions of sentences of a language can be deduced from statements about the semantic properties of their components (see Tarski, A.).

What is a truth theory for a language? Ordinarily, it is taken to consist of three parts: syntactic, semantic and logical. The syntactic part comprises a set of clauses stating how sentences of the language are constructed out of the vocabulary of the language. These clauses allow us to construct a description of any sentence of the language which shows how it is composed out of its vocabulary. (Such descriptions are often called ‘structural descriptions’.) For example, in a very simple subject-predicate language - call it *L* - whose vocabulary consists of the English names ‘Venus’ and ‘Polaris’ and the English predicates ‘is red’ and ‘is bright’, four sentences are constructible. Letting ‘ $x^\wedge y$’ mean the result of concatenating *x* with *y*, we may describe these sentences structurally as follows: ‘Venus’ $^\wedge$ ‘is red’, ‘Polaris’ $^\wedge$ ‘is red’, ‘Venus’ $^\wedge$ ‘is bright’ and ‘Polaris’ $^\wedge$ ‘is bright’. (This notation is rather clumsy, so, for ease of reading, I shall use quotation marks both in the ordinary way and to indicate structural descriptions.)

Next, the semantic part comprises a set of clauses specifying various semantic properties of the vocabulary of the language, together perhaps with a number of semantic properties of complex expressions articulated as a function of the semantic properties of their components. In tandem with the third part, which consists of a number of logical

and set-theoretic principles, these clauses will determine the truth-conditions of sentences of the language. It is this last feature that allows us to see in principle how we might solve D(ii).

Consider the language *L* again. Our aim is to find semantic properties of its names and predicates which will determine the truth-conditions of sentences containing them. And if what we have said so far is right, this should involve a statement of what (according to the theory) the names refer to and the predicates are true of. Thus we might have

(1) The referent of ‘Venus’ = Venus;
(2) The referent of ‘Polaris’ = Polaris;
(3) ‘is red’ is true of something iff that thing is red;
(4) ‘is bright’ is true of something iff that thing is bright.

(Note that strictly speaking the semantic properties here should be relativized to *L*, for in other languages its names and predicates may refer to or be true of very different things. Suppressing the language parameter, however, makes the axioms more readable. The same is true of all semantic properties considered in this entry.)

At this point, of course, we still do not know how these clauses determine specific truth-conditions for sentences. But we may remedy this by adding the following general clause relating the truth of *L*-sentences to the semantic properties of their components:

(5) For all names γ and predicates *F*, ‘ γF’ is true if and only if *F* is true of the referent of γ.

For assuming that the third, logical part of the truth theory includes elementary first-order logic, it is then a simple matter to deduce truth-conditions for the sentences. For instance, we may easily deduce that ‘Venus is red’ is true if and only if Venus is red. For (5) will entail that

‘Venus is red’ is true if and only if ‘is red’ is true of the referent of ‘Venus’;

(1) will entail that

‘is red’ is true of the referent of ‘Venus’ if and only if ‘is red’ is true of Venus;

and (3) will entail that

‘is red’ is true of Venus if and only if Venus is red.

Similar reasoning will enable us to deduce that ‘Polaris is red’ is true if and only if Polaris is red, and so on for the other two.

We are thus able to deduce a statement of what are in fact meaning-indicating truth-conditions for each of the sentences of *L* on the basis of axioms which assign semantic properties to their components, along with a further axiom which characterizes their modes of combination. Of course, this only solves our difficulty for the very limited language *L*; but it does give a hint of how to solve it more generally.

3 Providing correct meaning-specifications

Let us turn now to Problem B - how to circumscribe in a non-circular manner those truth-conditions of declarative sentences which embody their meanings - together with D(i), the corresponding problem for sentence components. And let us begin by introducing a common abbreviation: let us call a biconditional of the form ‘*S* is true iff *p*’ a ‘*T*-sentence’. Since *T*-sentences evidently express the truth-conditions of the declarative sentences mentioned on their left-hand sides, it follows that to deal with Problem B, we need a non-circular way of singling out those *T*-sentences which genuinely indicate the meanings of the declarative sentences from those which do not. One suggestion would be to require that for ‘*S* is true iff *p*’ to be a meaning-indicating *T*-sentence, ‘*p*’ must be a translation of *S*. But although this would certainly ensure that the *T*-sentence correctly indicated the meaning of the relevant sentence, it would not provide a non-circular elucidation of its meaning, since the notion of translation itself directly presupposes the notion of meaning: when *A* and *B* are distinct expressions, *A* is a translation of *B* if and only if *A* and *B* have the same meaning.

A second suggestion is to replace the notion of translation with more general anthropological considerations relating to interpretation. Speech and writing are activities which take place against a backdrop of broadly rational activity, and can be interpreted in the light of such activity with a view to making maximum sense of speakers' behaviour. Consequently, it should be possible in principle to determine whether speakers are saying, for example, that Venus is red when they utter a particular sentence *S* by assessing whether that would be the most sensible thing for them to be saying in the circumstances, and in the light of all their other behaviour. And if this is what they typically or characteristically say when they utter *S*, then (according to the second suggestion) we may identify this with the appropriate truth-condition for *S*.

This prompts the following thought: '*S* is true iff *p*' is correctly meaning-indicating if, other things being equal, the supposition that speakers are saying that *p* in characteristic utterances of *S* makes maximum sense of their behaviour. But even this might be thought to invite a similar criticism to the one which undermined the first suggestion. For does not the use of speech acts such as saying just bring in meaning by the back door? It certainly seems to if saying that *p* is uttering something that means that *p*. But much depends on whether a deeper account of saying - and other speech acts - is possible (see Radical translation and radical interpretation).

A third suggestion would be to forgo such use of speech acts in the characterization of meaning, and focus on speakers' knowledge of truth-conditions. It is clear that those who understand the sentence 'Venus is red' will know not merely that 'Venus is red' is true if and only if Venus is red, but that speakers are expected to know this. Moreover, it also seems clear that they will know that this is the strongest thing speakers are expected to know about the truth-conditions of 'Venus is red'. They will know, for example, that speakers are not expected to know that 'Venus is red' is true if and only if Venus is red and $7 + 5 = 12$. But this suggests the following proposal: that '*S* is true iff *p*' is meaning-indicating if any speaker knows that *S* is true if and only if *p*, that speakers are expected to know this, and that this is the strongest thing speakers are expected to know about the truth-conditions of *S*. Our *T*-sentence which fails to specify the meaning of 'Venus is red' - 'Venus is red' is true if and only if Venus is red and $7 + 5 = 12$ - is then ruled out (see Meaning and understanding).

Turning now to D(i), what of the components of sentences? How can we ensure that their meaning-specifications are correct? A thought that has occurred to several writers is to let whatever is the most satisfactory method of circumscribing the meaning-specifying truth-conditions of sentences do all the work: to say, for example, that the meaning of a non-sentential expression of a language can be specified by its clauses in any true truth theory that meets the above kinds of knowledge-expectation conditions for its sentences. For this to be adequate, however, every true truth theory for a language that meets those conditions would have to assign to those expressions the same or equivalent semantic properties. Unfortunately this not true.

To see this let us return to the truth theory of our language *L*. As we saw in §2, it is possible to use that truth theory to derive a *T*-sentence for each sentence of *L*; and each such *T*-sentence plainly satisfies the relevant knowledge-expectation conditions. What is more, its axioms plausibly specify the meanings of each of *L*'s basic expressions. There are, however, other truth theories for *L* that meet the knowledge-expectation conditions for its sentences, but which do not specify the meanings of its basic expressions. For example, consider a truth theory for *L* which had the following axiom governing 'is red' instead of (3):

(6) 'is red' is true of something if and only if it is red and is either green or not green.

It would certainly be possible to derive all the correct meaning-indicating *T*-sentences for sentences of *L* on the basis of this axiom and the axioms governing the other basic expressions of the sentences. But (6) does not specify the meaning of 'is red' correctly.

It is true that we could rule out such cases by requiring that meaning-indicating *T*-sentences be proved in certain canonical ways - for instance, in the way we derived the *T*-sentence for 'Venus is red' in §2. For a proof that differed from that derivation only in using (6) instead of (3) would result in a *T*-sentence that did not satisfy the knowledge-expectation conditions, namely: 'Venus is red' is true if and only if Venus is red and is either green or not green. But this immediately raises the question why those ways should be thought of as canonical; and since the natural answer is that they ensure that the axioms governing the sentence's basic components specify their meanings correctly, it would seem that we must have some independent grip on what it is for those axioms to be so constrained as to have this property. The question is what provides us with that independent grip.

At this point, we may make a move analogous to the one made earlier in the case of sentences. There we noted that understanding the sentence 'Venus is red' involves knowing that it is true if and only if Venus is red, that speakers are expected to know this, and that this is the strongest thing speakers are expected to know about its truth-conditions. Similarly, we may say that understanding the predicate 'is red' involves knowing it is true of something if and only if that thing is red, that speakers are expected to know this, and that this is the strongest thing speakers are expected to know about the conditions under which it is true of something. (They will know, for example, that speakers will not be expected to know that it is true of something if and only if that thing is red and is either green or not green.) Equally, understanding the name 'Venus' will consist in knowing that it refers to Venus, that speakers are expected to know this, and that this is the strongest thing speakers are expected to know about what it refers to. Notice again that this solves our difficulty only for the very limited language *L*; but we may expect it to be soluble quite generally by means of similar kinds of knowledge-expectation conditions relating to whatever clauses govern other expressions and their modes of combination.

It should be noted that these kinds of knowledge-expectation conditions appear to put a considerable burden on the ordinary speaker: can we really picture them as part of the stock of knowledge possessed by speakers of a natural language such as English? At this point, if we are not to jettison the present approach, there seem to be three strategies. One is to claim that the truth-theoretic clauses are correct, to deny that we have knowledge of them (in the ordinary sense), but to insist that we bear some kind of analogous cognitive relation to them (yet to be fully explicated). The second is again to claim that they are correct, but to try to justify the ascription of tacit knowledge of them in some way. And the third is to see the truth-theoretic clauses as stepping stones to more convincing clauses which we can uncontroversially accept that we know.

4 Recalcitrant idioms and general difficulties

We turn next to a number of problems for the truth-conditional approach to meaning. In the light of Tarski's work, it is now clear how to accommodate within a truth-theoretic framework declarative sentences made up of names, predicates, truth-functors and standard quantifiers. But natural language seems much richer than this. We here consider two idioms which cause problems for the framework, and which in their different ways may require it to be modified. Then we consider two more general difficulties. It should be noted that there are various further issues - such as how the truth-conditional approach melds with doctrines of logical form, transformational grammar or validity - which, for reasons of space, we cannot discuss here.

Until now we have simplified matters by restricting ourselves to languages whose only sentences are declarative. This restriction is convenient, since - ignoring context-sensitivity - it means that all the sentences in the relevant languages have truth-conditions; but it is not warranted by natural languages, which contain sentences of many different moods, including imperatives, interrogatives and optatives. Such sentences do not appear to be truth-evaluable - 'Open the door, Smith!' or 'Where's the butter?' do not look like the kinds of sentence that can be true or false. Yet if they lack truth-conditions, it seems puzzling how they can be accommodated within a truth-theoretic framework.

One response is to add a further component to the semantic theory - what is sometimes called a theory of 'force' (see Pragmatics §8). This states what kind of speech act a given sentence is conventionally or typically geared to express; and it transforms the sentence into a corresponding declarative sentence which can then serve as input to the truth theory. For example, the theory of force for English would tell us that 'Open the door, Smith!' is a command, and that its corresponding declarative sentence is 'Smith opens the door'.

A second response is to introduce a number of further semantic values analogous to truth for each of the new types of sentence: compliance, for instance, for imperatives and optatives. Such values could be integrated into appropriate truth theories by means of such axioms as: *S*! is complied with iff *S* is true, where *S*! is the imperative form of a declarative sentence *S*. And to ensure that we obtain meaning-indicating compliance conditions, we could again use knowledge-expectation conditions, similar to those proposed in the case of declarative sentences, but with compliance replacing truth. Note, however, that both responses have great difficulty accommodating a sentence such as 'Open the door, Smith, and I'll put the kettle on' which contains more than one mood. Associated with this sentence, there seems no single speech act, nor any straightforward value analogous to truth.

Next we come to sentences whose truth-conditions can vary from one context to another. Such sentences, which

typically contain context-sensitive expressions such as personal pronouns, demonstratives, temporal and spatial indexicals, tense indicators and so on, evidently give rise to a problem for the programme of accounting for meaning in terms of truth-conditions. For although the truth-conditions of such sentences can vary from context to context, their meaning need not. The sentence ‘I helped that man yesterday’ means the same thing whoever utters it, whenever it is uttered and whoever the man is; but its truth-conditions will vary with the utterer, the time of utterance and the man indicated. Evidently, this is because in any ordinary way of taking the term ‘meaning’, the meanings of expressions such as ‘I’, ‘yesterday’ and so on are given by such general rules as that ‘I’ refers to the speaker, ‘yesterday’ refers to the day before the day of utterance and so on; but such meanings in themselves do not exhaust the contribution the expressions make to the truth-conditions of sentences containing them.

An amendment to the account is clearly needed. In particular, although we may take the truth-conditions of sentences to be identifiable with their meaning when they have context-free truth-conditions, we must identify those of context-sensitive sentences with a notion of content much stronger than ordinary meaning - roughly, with what the sentences express when the context has been fixed. This means that the semantic theory will in effect have to be divided into two parts: what supplies the context-free information relevant to truth-conditions; and what states the information about specific contexts that determines particular truth-conditions once the relevant context-free information has been supplied. The second part falls into the area of pragmatics (see Pragmatics; Demonstratives and indexicals).

Is this adequate? We here restrict ourselves to noting one general difficulty with the proposal. It is likely that most context-free semantic axioms will involve relativized semantic predicates. The pronoun ‘he’, for instance, might be governed by an axiom such as:

(7) For all *p*, *q*, the referent of ‘he’ as uttered by *p* to indicate $q = q$.

However, it is hard to obtain content-specifying instances of this which satisfy the appropriate knowledge-expectation conditions involving reference (see §3). For suppose I am the speaker and the person I am indicating is Quine. Then to display the content of ‘he’ in my mouth, we need a metalinguistic expression synonymous with it - ‘*a*’, say. This would give us:

The referent of ‘he’ as uttered by me to indicate Quine = *a*.

But even if we had such an expression, we would still face the problem that not all English speakers, even those in my immediate linguistic circle, need know of the referent of ‘he’ in my mouth that it is *a*. Indeed, I may be the only person who knows this. In such extreme circumstances, we could restrict the class of speakers to me. But it would then have to be the case that the strongest thing I expect myself to know about its referent is that it is *a*. And unfortunately there is no reason to think that this is so.

Finally we turn to two general problems for the programme of elucidating meaning in terms of truth-conditions, both of which have considerable contemporary significance. The first, antirealist difficulty focuses on sentences whose truth-conditions we cannot decide. According to one model of understanding, we can acquire and manifest an understanding of a sentence only if we can recognize whatever meaning-indicating condition is associated with it as obtaining when it does obtain and as not obtaining when it does not. But for sentences whose truth-conditions we cannot decide, the supposition that truth-conditions are meaning-indicating conflicts with this requirement. So in the absence of another model, the attempt to elucidate meaning in terms of truth-conditions must be jettisoned (see Intuitionistic logic and antirealism §6; Realism and antirealism §4).

The second difficulty concerns the definition of truth. It is clear that an extant notion of truth is needed to characterize meaning in the way we have been envisaging: we cannot use truth theories like the one described in §2 as analyses of truth, if only because they leave it unsaid how to apply the notion of truth to new languages. However, one of the most plausible independent accounts of truth, the ‘redundancy theory’ (see Truth, deflationary theories of), is arguably incompatible with any attempt to characterize meaning in terms of truth. For on one natural construal, this theory requires that understanding what it is for ‘Venus is red’ (say) to be true is tantamount to understanding what it is to say that Venus is red by using the sentence ‘Venus is red’. But to understand this, we already need to know the meaning of ‘Venus is red’.

Both objections seem to require that we replace truth by some other notion in an elucidation of meaning. One that

appeals to antirealists is that of warranted assertibility, since the conditions under which one would be warranted in asserting a sentence are arguably recognisable as obtaining when they obtain and as not obtaining when they do not. Other options would be to elucidate meaning in terms of communication intentions or conceptual role.

See also: Davidson, D.; Logical and mathematical terms, glossary of

STEPHEN G. WILLIAMS

References and further reading

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press.(These essays present the most influential contemporary defence of a truth-conditional account of meaning; see in particular Essays 2, 9 and 10. Essay 8 discusses the problem of non-declarative sentences; his own account is different from those noted in §4.)

Davies, M.K. (1981) *Meaning, Quantification, Necessity*, London: Routledge. (Includes truth theories for a range of English idioms; chapter 4 has a helpful discussion of tacit knowledge.)

Dummett, M.A.E. (1978) *Truth and Other Enigmas*, London: Duckworth.(Includes a number of seminal essays on antirealism; see in particular Essays 1, 12 and 21. Essay 1 and the Preface argue for the incompatibility of the redundancy theory of truth and the truth-conditions theory of meaning.)

Evans, G. and McDowell, J. (eds) (1976) *Truth and Meaning*, Oxford: Oxford University Press.(Includes an excellent introduction by the editors.)

Frege, G. (1964) *The Basic Laws of Arithmetic*, trans. and ed. and with an intro. by M. Furth, Berkeley, CA: University of California Press.(This is a translation of parts of Frege's *Grundgesetze der Arithmetik*. The suggestion that the meaning of a declarative sentence may be given by the conditions under which it is true is to be found, wrapped up in Frege's own terminology, in §32.)

Higginbotham, J. (1991) 'Truth and Understanding', repr. in C. Peacocke (ed.) *Understanding and Sense*, Aldershot: Dartmouth, 1993, vol. 1.(Presents detailed elucidation of some of the material in §3; in particular suggests circumscribing truth-conditions and so on in terms of knowledge-expectation conditions.)

Larson, R. and Segal, G. (1995) *Knowledge of Meaning: An Introduction to Semantic Theory*, Cambridge, MA: MIT Press.(An excellent survey of truth theories for many English idioms; it also indicates how to integrate this material with current thinking in linguistics. The first three chapters include accounts of some of the basic ideas underlying the truth-conditional approach to meaning; chapter 13 discusses tacit knowledge.)

Lepore, E. (ed.) (1986) *Truth and Interpretation*, Oxford: Blackwell.(This volume includes important essays on and within the framework of truth-conditional theories of meaning. Particularly helpful are those by Higginbotham, Davidson and Dummett.)

McDowell, J. (1977) 'On the Sense and Reference of Proper Name', repr. in M. Platts (ed.) *Reference, Truth and Reality*, London: Routledge, 1980; and in C. Peacocke (ed.) *Understanding and Sense*, Aldershot: Dartmouth, 1993, vol. 1.(Suggests coupling a truth-conditional account with a theory of force in the way indicated briefly in §4. Also fills out material in §3.)

Peacocke, C. (ed.) (1993) *Understanding and Sense*, 2 vols, Aldershot: Dartmouth. (These volumes also include important essays on and within the framework of truth-conditional theories of meaning. In addition to those by Higginbotham and McDowell already recorded, the essays by Davidson on truth, by McDowell on antirealism and by Quine and Davies on tacit knowledge are notable.)

Platts, M. (ed.) (1980) *Reference, Truth and Reality*, London: Routledge.(Includes useful essays on Tarski's account of truth, by Field and McDowell.)

Wright, C. (1987) *Realism, Meaning and Truth*, Oxford: Blackwell, 2nd edn, 1993. (Contains important essays on the antirealist critique of truth-conditional semantics. Introduction is particularly helpful.)

Yourgrau, P. (ed.) (1990) *Demonstratives*, Oxford: Oxford University Press.(A useful selection of essays on context-sensitivity.)

Meaning and understanding

The existence of a close connection between the notions of meaning and understanding can hardly be denied. I may be said to understand you, on a given occasion of utterance, when I know what you then meant - or, at least, when I know the meaning of the words that you then uttered. An important and influential school of thought within the philosophy of language goes much further than these platitudes, however. Its members adhere to the view that questions about meaning are essentially questions about understanding: 'a model of meaning is a model of understanding'. Their approach contrasts with that of those who expect an account of meaning to elucidate the nature of understanding only indirectly - perhaps by explaining meaning in terms of truth, inference, synonymy or self-expression, and only then explaining understanding as the correct recovery of meaning.

1 Dummett's account

The text of this article sketches and constrasts the views of the school's two leading members: Dummett and Davidson. The dictum 'a model of meaning is a model of understanding' was formulated by Michael Dummett. He proceeds to gloss 'a model of understanding' as 'a representation of what it is that is known when an individual knows the meaning' of a possible object of understanding - as it might be, a word, a construction, or a whole language (1973: 217). To provide such a model is to spell out the kind of knowledge that constitutes, or at least sustains, various kinds of linguistic understanding. Dummett's view is that the central philosophical debates about the nature of meaning are properly understood to be debates about the character of such knowledge. Slogans such as 'The meaning of a sentence is the method of its verification' or 'The meaning of a sentence consists in its truth-conditions' mean (if they mean anything at all) that to *understand* a sentence is to know how it may be verified or to know under what conditions it is true. In emphasizing that the central philosophical questions about meaning are questions about understanding, Dummett was self-consciously departing from a prior tradition (of which the most celebrated manifestation is Quine's 'Two Dogmas of Empiricism'); this tradition took the central problem in the area to be that of spelling out the conditions for two expressions to mean the same thing (see esp. Dummett 1981: 74).

Although Dummett canvasses various accounts of knowledge of meaning, it is a presupposition of his discussion that the knowledge in question is propositional. To attempt to produce a model of understanding is to attempt to produce a *theory*, knowledge of which constitutes or sustains possession of a linguistic capacity. In saying so much, Dummett is not committed to denying that an understanding of a word or sentence is in the first instance a matter of knowing how to use it. However, he is committed to denying that this formulation provides a complete account of understanding, he argues that we need a more detailed specification of what such know-how consists in (1973: 222-3). Dummett is also clear that the propositional knowledge in question will in general be implicit. That is to say, it is knowledge that a subject may possess even though he is unable to express the proposition thereby known. From this observation, Dummett derives a consequence that crucially constrains the sort of knowledge which he thinks might constitute or sustain understanding. 'Implicit knowledge', he says, 'cannot meaningfully be ascribed to somebody unless it is possible to say in what the manifestation of that knowledge consists: there must be an observable difference between the behaviour and capacities of someone who is said to have that knowledge and someone who is said to lack it' (1973: 217). In particular, then, the propositional knowledge, implicit possession of which is said to underlie understanding, must be manifestable in behaviour.

The significance of this requirement emerges when we consider the two models of understanding that Dummett centrally considers. Both of these models take to be central the problem of characterizing our understanding of declarative sentences, and according to the first of them - which Dummett labels a 'realist' or 'truth-conditional' model - such understanding 'consists in our grasp of [the sentence's] truth-conditions, which determinately either obtain or fail to obtain, but which cannot be recognized by us in all cases as obtaining whenever they do' (1972: 23). Thus, our understanding of the sentence 'A city will never be built at the North Pole' is taken by the realist to consist in our knowing that it is true just on condition that a city will never be built at the North Pole, a condition that may obtain without our ever knowing so (Dummett 1959: 16-7). This, Dummett argues, is rendered problematical by the requirement of manifestability. 'It is quite obscure', he writes, 'in what the knowledge of the condition under which a sentence is true can consist, when that condition is not one which is always capable of being recognized as obtaining' (1973: 224). The crucial claim, in other words, is that there is

something opaque about the attribution 'Tom knows that sentence *S* is true just on condition that a city will never be built at the North Pole'. Because the occasion may never arise on which somebody is granted the opportunity to give conclusive reasons for accepting the sentence as true or rejecting it as false, 'the knowledge which is being ascribed to one who is said to understand the sentence is knowledge which transcends the capacity to manifest that knowledge by the way the sentence is used'. Accordingly, the truth-conditional model of understanding cannot be one 'in which meaning is fully determined by use' (1973: 225).

Whether this really is a consequence of the truth-conditional model - and whether, if it is, it is sufficient to show that the realist model must be rejected - are questions that remain at the forefront of contemporary debate in the philosophy of language. As he made clear in 'Realism: A Valedictory Lecture' (1992), Dummett is not committed to particular answers. He has, however, articulated an alternative model of understanding which, whatever its other difficulties, is supposed to meet the requirement of manifestability. On this alternative 'antirealist' model, our understanding of a sentence will consist, not in knowing under what conditions it is true, but 'in knowing what recognizable circumstances determine it as true or as false' (1972: 23). Or, to deploy a formulation that Dummett treats as equivalent, it consists in knowing under what circumstances a sentence may be asserted. In addition to meeting the demand of manifestability, this model is supposed to have the merit of leading to a more plausible account of what learning a language involves:

> Thus we learn to assert '*P* and *Q*' when we can assert *P* and can assert *Q*; to assert '*P* or *Q*' when we can assert *P* or can assert *Q*…. We no longer explain the sense of a statement by stipulating its truth value in terms of the truth value of its constituents, but by stipulating when it may be asserted in terms of the conditions under which its constituents may be asserted.
>
> (1959: 17-18)

2 Davidson's account

It may seem as though Donald Davidson is a clear example of a Dummettian realist - someone who takes our understanding of a declarative sentence to consist in knowing the conditions under which it is true. However, while Davidson certainly regards the notion of a truth-condition as central to an account of a subject's linguistic competence, there are nuances in his treatment of these matters that make such a characterization misleading. He begins a celebrated article thus:

> Kurt utters the words 'Es regnet' and under the right conditions we know that he has said that it is raining. Having identified his utterance as intentional and linguistic, we are able to go on to interpret his words: we can say what his words, on that occasion, meant. What could we know that would enable us to do this? How could we come to know it? The first of these questions is not the same as the question what we *do* know that enables us to interpret the words of others. For there may easily be something we could know and don't, knowledge of which would suffice for interpretation, while on the other hand it is not altogether obvious that there is anything we actually know which plays an essential role in interpretation.
>
> (1973: 125)

Davidson is surely right to draw attention to the difference between his central question about knowledge of meaning ('What could we know that would enable us to know what these words mean?') and Dummett's ('What *do* we know when we know what words mean?'). But it is disappointing that he does not give his grounds for doubting that 'there is anything we actually know which plays an essential role in interpretation'. For if there is such actual knowledge, then questions about its character must surely be deemed more urgent and important than the hypothetical question that Davidson chooses to address. We appear, in other words, to need a reason for taking as central the problem upon which he focuses.

Be that as it may, Davidson's own question is a clear one, and his writings contain two rather different answers to it. In his early papers (roughly speaking, those published before 1970, including the influential 'Truth and Meaning' of 1967), we are told that what would suffice for interpreting a speaker *A* at time *t* is knowledge of a truth theory that is interpretative for *A* at *t*. A truth theory, as Davidson conceives it, comprises statements (*T*-theorems) to the effect that a sentence is true (as potentially uttered by such-and-such a speaker at this or that time) just in case a certain condition is met. And such a theory will be interpretative, for a speaker at a time, if (intuitively) the condition that the theory associates with each sentence gives the meaning that it then has in that

speaker's mouth. Plainly, this answer is of little interest without some independent specification of the condition for a theory to be interpretative (independent of the notion of meaning), which Davidson hopes to provide via the figure of the radical interpreter. Such an interpreter is an investigator who constructs a truth theory for a speaker (at a particular time) in accordance with certain general interpretative maxims or constraints, chief among which is the principle of charity (see Davidson 1973; Lewis 1974; Charity, principle of §4; Radical translation and radical interpretation §7-10). The idea, then, is that it is such maxims as these that give content to the requirement that a theory be interpretative, and thereby give content to the notion of meaning.

The model of understanding and meaning that results involves a substantial holism. The condition for a sentence *S* to mean that *p* as a speaker *A* uses it at time *t*, is that a truth theory that is interpretative for *A* at *t* (that is, a truth theory that optimally meets the constraints under which the radical interpreter operates) should contain among its *T*-theorems the result '*S* is true (as potentially uttered by *A* at *t*) just in case *p*'. This makes it clear that the condition for a sentence to have a certain meaning involves reference, not just to the components of *S*, but to the totality of sentences that the relevant speaker is then disposed to utter (see Davidson 1976: 175). This holism does not alarm Davidson, but he has confessed to an important error in his earlier answer to his question. 'My mistake', he wrote in 1976, 'was to overlook the fact that someone might know a sufficiently unique theory [that is a theory that optimally satisfies such constraints as the principle of charity] without knowing that it was sufficiently unique' (1976: 173). He duly amends his account of what suffices for interpretation to correct this mistake: 'what somebody needs to know [in order to interpret speakers of a language *L*] is that some [sufficiently unique] *T*-theory for *L* states that… (and here the dots are to be replaced by a *T*-theory)' (1976: 174).

The occurrence in this formulation of the intentional and linguistic verb 'to state' may lead one to wonder whether the account is invoking a notion to which it is not entitled (compare Forster 1976). After all, if we are allowed to use a verb like 'state' in specifying the knowledge that would suffice for understanding, then why should we not bypass the complexities of truth theories completely, and say simply that what suffices for interpreting a speaker is knowing what he would be stating by affirming declarative sentences, what he would be asking by uttering interrogative ones, and what he would be ordering by uttering imperative sentences? Davidson, however, has a convincing answer to this question. Although the use of the verb 'state' (or, as he eventually prefers, 'entail') 'brings in an appeal to a specifically linguistic notion', it is a notion that has been 'elicit[ed] by placing conditions on a theory of truth' (1976: 178). That is to say, its content must be understood in relation to the conditions for a truth theory to be interpretative.

All the same, one might wonder whether a more direct method of eliciting (or explaining) such a notion was not available. Instead of considering truth theories - and regarding the principle of charity and its ilk as providing conditions for a truth theory to be interpretative - why should we not consider theories entailing such claims as that:

By affirming *S* at *t*, *A* would state that *p*

and regard the relevant principles as conditions for such a theory to be *true*? Davidson's answer appears to be that certain logical difficulties arise in the latter case that are absent from the former. In an early paper, Davidson wrote that 'when we can regard the meaning of each sentence as a function of a finite number of features of the sentence, we have an insight not only into what there is to be learned; we also understand how an infinite aptitude can be encompassed by finite accomplishments' (1965: 8). The context makes it clear that the present occurrence of 'when' means 'when and only when' and the suggestion is that such insight will be gained only when we have an axiomatization - if not finite, then at least recursive - of the body of knowledge in question. Now Davidson also suggests that a theory whose theorems are in the form:

By affirming *S* at *t*, *A* would state that *p*

will resist axiomatization. 'It is reasonable to expect', he writes, 'that in wrestling with the logic of the apparently non-extensional "means that" [and, we might add, the equally non-extensional "would state that"] we will encounter problems as hard as, or perhaps identical with, the problems our theory is out to solve' (1967: 22).

But are such problems insuperable? Surely one can account for somebody's knowing that Kurt uses 'Es regnet und es schneit' to state that it is raining and snowing by reference to his knowing:

(1) that Kurt uses ‘Es regnet’ to state that it is raining
(2) that Kurt uses ‘Es schneit’ to state that it is snowing

and

(3) that (*p*)(*q*)(*S*)(*S′*)(If Kurt uses *S* to state that *p*, and uses *S′* to state that *q*, then he uses ‘*S* und *S′*’ to state that *p* and *q*)?

One source of concern about this last item of knowledge may be that the initial pair of quantifiers must be understood substitutionally (see Quantifiers, substitutional and objectual), but it is noteworthy that on Davidson’s preferred analysis of such constructions as ‘states that’, there is a far more acute problem. On his ‘paratactic’ analysis, what follows the word ‘that’ in an attribution such as ‘Galileo stated that the earth moves’ belongs in a semantically separate sentence from the properly attributive claim ‘Galileo said that’ (see Davidson, D. §6; Propositional attitude statements §3). Applied to (c), then, the substitutional quantifiers ‘(*p*)’ and ‘(*q*)’ would reach into a different sentence. Here, the ‘direct’ approach to theories of meaning seems to take us into uncharted logical waters. Unless Davidson’s analysis of indirect discourse can be shown to be wrong, or unless some way can be discovered of charting these waters, we have a reason for answering his question in the way that he recommends.

See also: Meaning and truth; Meaning and verification

IAN RUMFITT

References and further reading

Davidson, D. (1965) ‘Theories of Meaning and Learnable Languages’, in Y. Bar-Hillel (ed.) *Proceedings of the 1964 International Congress for Logic, Methodology, and Philosophy of Science*, Amsterdam: North Holland; repr. in D. Davidson, *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1984, 3-15. (Gives the rationale for requiring an interpretative truth theory to be finitely axiomatizable.)

Davidson, D. (1967) ‘Truth and Meaning’, *Synthèse* 17: 304-23; repr. in D. Davidson, *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1984, 17-36.(The classic statement of Davidson’s earlier conception of a theory of meaning.)

Davidson, D. (1973) ‘Radical Interpretation’, *Dialectica* 27: 313-28; repr. in D. Davidson, *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1984, 125-39.(Elaborates Davidson’s later conception of theories of meaning.)

Davidson, D. (1976) ‘Reply to Foster’, in G. Evans and J. McDowell (eds) *Truth and Meaning: Essays in Semantics*, Oxford: Oxford University Press, 33-41; repr. in D. Davidson, *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1984, 171-9.(Davidson compares his earlier and later views, explaining the reasons for the change.)

Dummett, M.A.E. (1959) ‘Truth’, *Proceedings of the Aristotelian Society* 59: 141-62; repr. in M.A.E. Dummett, *Truth and Other Enigmas*, London: Duckworth, 1978, 1-19. (An early and influential presentation of the idea that to understand a sentence is to know under what conditions it could be asserted.)

Dummett, M.A.E. (1972) ‘Postscript to Truth’, in M.A.E. Dummett, *Truth and Other Enigmas*, London: Duckworth, 1978, 19-24.(Clarifies the thesis and argument of the previous article.)

Dummett, M.A.E. (1973) ‘The Philosophical Basis of Intuitionistic Logic’, in H.E. Rose and J.C. Shepherdson (eds) *Logic Colloquium ’73*, Amsterdam: North Holland; repr. in M.A.E. Dummett, *Truth and Other Enigmas*, London: Duckworth, 1978, 215-47.(Contains Dummett’s clearest and most influential formulation of the requirement that linguistic knowledge should be manifestable in use.)

Dummett, M.A.E. (1981) *The Interpretation of Frege’s Philosophy*, London: Duckworth.(The chapter entitled ‘Synonymy’ compares Dummett’s way of formulating the questions that a theory of meaning ought to answer with previous conceptions, notably Quine’s.)

Dummett, M.A.E. (1992) ‘Realism: A Valedictory Lecture’, Oxford: Oxford University Press. (Dummett’s discussion of the state of the debate between truth-conditional theories of meaning and their rivals.)

Forster, J.A. (1976) ‘Meaning and Truth Theory’, in G. Evans and J. McDowell (eds) *Truth and Meaning: Essays in Semantics*, Oxford: Oxford University Press, 1-32.(An influential attack on Davidson’s earlier and later theories.)

Lewis, D.K. (1974) ‘Radical Interpretation’, *Synthèse* 27: 331-44.(Spells out explicitly the conditions under which the radical interpreter operates.)

Quine, W.V. (1951) ‘Two Dogmas of Empiricism’, *Philosophical Review* 60: 20-43.(Explains the problems

attending the explication of ‘synonymous’.)
Strawson, P.F. (1970) ’Meaning and Truth: an Inaugural Lecture’, Oxford: Oxford University Press; repr. in P.F. Strawson, *Logico-Linguistic Papers*, London: Methuen, 1971, 170-89.(An influential attack upon the whole school of thought discussed here. Strawson argues that the enterprise of spelling out truth-conditions must be pursued within a theory that explains meaning as relating, in the first instance, to communicative intentions.)

Meaning and verification

The verifiability theory of meaning says that meaning is evidence. It is anticipated in, for example, Hume's empiricist doctrine of impressions and ideas, but it emerges into full notoriety in twentieth-century logical positivism. The positivists used the theory in a critique of metaphysics to show that the problems of philosophy, such as the problem of the external world and the problem of other minds, are not real problems at all but only pseudoproblems. Their publicists used the doctrine to argue that religion, ethics and fiction are meaningless, which is how verificationism became notorious among the general public.

Seminal criticism of verification from around 1950 argues that no division between sense and nonsense coincides tidily with a division between science and metaphysics, as the positivists had claimed. Quine later developed verificationism into a sort of semantic holism in which metaphysics is continuous with science. In contrast, Dummett argues from a reading of Wittgenstein's claim that meaning is use to a rejection of any sort of truth surpassing the possibility of knowledge, and thence to a defence of intuitionistic logic. But the claim that all truths can be known yields in an otherwise innocuous setting the preposterous consequence that all truths actually are known. There are ways to tinker with the setting so as to avoid this consequence, but it is best to conclude by reductio *that some truths cannot be known and that verificationism is false. That in turn seems to show that the prospects for an empiricist theory of meaning are dim, which might well shake a complacent confidence in meaning.*

1 Hume

Perception is the only channel to us from the world whose influx is generally accepted to justify our beliefs about the world. Beliefs figure as units of justification in this empiricist-sounding doctrine. Since justifying something is giving a reason for thinking it true, the units of justification should be the right size for being true or false. Word-sized bits, be they words themselves or their meanings or people's ideas of things, are too small, while paragraph-sized chunks are too big, but sentence-sized stretches are of just the right size and unity for having truth-values. It is natural to take the things believed and justified by experience to be sentence-sized or propositional.

But the unit of Hume's seminal empiricism seems smaller. He begins the *Treatise of Human Nature* (1739-40) with the doctrine that all ideas are derived from impressions. To the extent that Hume made a self-conscious distinction between words and sentences (on the language side), or between concepts and propositions (on the Platonic side), or between ideas and thoughts (on the psychological side), the ideas Hume had in mind seem smaller than sentence-sized. In that way, Hume's empiricism seems narrower than the doctrine with which we began.

Hume claimed ideas are derived from impressions by copying, the copies differing from their originals only in vivacity. Ideas and impressions were thus to be images, the ideas merely fainter reproductions of the impressions. Objections to taking abstract ideas as images predate Hume, but the doctrine at least places ideas in a familiar psychological category. Items in this category are accessible to inner sense or introspection if anything is, and thus are knowable by processes such as those of the five outer senses already empirically certified. Admittedly, images seem too particular to reflect the abstract nature of abstract ideas. But at least acquaintance with images could explain action. If we try to rethink ideas not as images but as capacities or dispositions, we seem left with bare possibilities or potentialities, ungrounded in actualities which could explain action. Faced with a choice between bad theory and none, we often seem to prefer the bad to none; at least bad theory may provide a (risky) basis for getting on, but none provides none. Thinking of ideas on the model of images does not go away.

The doctrine that ideas are images derived from sense impressions by copying can be put to work in two different ways. One might take it for granted that we do indeed have certain ideas we seem to have, and then cast about for impressions from which such ideas might be derived; such a cast might well hook surprisingly unfamiliar epistemic fish. Alternatively, one might start from a catalogue of familiar impressions, and then winnow out ideas that cannot be derived by copying any such impressions; such a purge might well leave a rather sceptical residue.

Hume thought of himself as a sceptic. Causation and the self are two examples of this sceptical pattern in Hume. In

the case of causation, he never seems to have said exactly which impressions are wanting. Kant writes that experience teaches us that a thing is so-and-so, but not that it cannot be otherwise (*Critique of Pure Reason*, 1787: B3). In Hume's terms, we have impressions only of actuality, never of necessity; we see red, but not having-to-be red. But then there is no impression of necessity from which to derive an idea of it, so there is no making sense of causation as necessitation. At most we experience regular succession (with temporal priority and spatial contiguity), so that is all there can be to our idea of causation. Thus Hume's scepticism about necessary connection in nature is linked to his doctrine of ideas. In the case of the self, Hume asserts quite explicitly that there simply is no such thing as an experience of oneself over and above the ideas and impressions revealed to introspection. So, he infers, there can be nothing to the self and our idea of it but the bundle of the contents of consciousness. This bundle theory of the self has always seemed one of the more shockingly sceptical parts of the *Treatise*.

2 Logical positivism

We have now seen how Hume can use his doctrine of impressions and ideas as a sceptical weapon. The Vienna Circle between the two world wars of the twentieth century claimed an intellectual lineage back to Hume, and it is the Circle to whom we credit (or on whom we blame) the verifiability theory of meaning (see Vienna Circle §3). It would be an anachronism to attribute to Hume our preoccupations with language, meaning and reference. But imagine a philosopher already so preoccupied reconsidering Hume's doctrine of impressions and ideas. Many find it difficult to become accustomed to a self-conscious distinction between the meaning of an expression and what a person means by that expression. Without such a distinction at one's fingertips, it is all too easy to slide between the meaning of an expression and the ideas expressed by people using that expression. If our philosopher also admires Hume's empiricism, they may incline towards the doctrine that these ideas must be derived from impressions. The core of empiricism is the view that sense experience is the basic sort of evidence we have for (that is, for verifying) our beliefs about the world. It is then not unnatural for this suspension - of meaning, ideas, impressions and verification - to precipitate out a verifiability theory of meaning, or at least the slogan that meaning is evidence.

The verifiability theory of meaning may not be present in the greatest works of Vienna positivism, such as Carnap's *Aufbau* or his *Logical Syntax of Language* (see Carnap, R. §§1-3). But it does seem to animate his *Scheinprobleme in der Philosophie* (*Pseudoproblems in Philosophy*, 1967) and it is obvious in Ayer's *Language, Truth and Logic* (1936), an enormously noticed popularization of logical positivism (see Ayer, A.J. §§2-3). In so far as the positivists had verificationism, they used it as a weapon of criticism. Hume's scepticism can be taken as criticism of complacent confidence in causes and the self, but it would be a distortion to assimilate verificationist criticism without further ado to Hume's scepticism. On the contrary, Carnap's *Pseudoproblems* shows him using verificationism to criticize certain sceptical lines of thought, the problem of the external world and the problem of other minds.

Descartes (1641) denied the pretensions of sense experience to be utterly basic to knowledge and its justification. To that end, he asks how one knows one is not asleep and dreaming right now. Come to that, perhaps one has been asleep and dreaming all of one's life, and all one's experience, instead of reporting the world of matter it seems to present, is instead merely an interminable hallucination spun by oneself out of whole cloth. This difficulty is the problem of the external world (see Descartes, R. §9). Indeed, perhaps there is no matter at all, but only one's mind and the endless illusions one foists on oneself. Here we have two incompatible hypotheses. One says that there is the matter out there independent of us that there seems to be, subject to the laws of nature. The other says that there is no such matter, but only the hallucination of it one foists on oneself. It is crucial to Descartes' case that the sense experience forecast by these two hypotheses be indistinguishable. For we are to be sure that the matter hypothesis is true and the dreaming hypothesis is false, and since sense experience cannot separate the two, there must be some other way of knowing, perhaps even more basic than experience, since it is what entitles us to take our experience to support the matter hypothesis. Prominent along that way are the *Cogito* and the ontological proof of the existence of God. Many after Descartes are not persuaded that his way leads anywhere. So they are left without an answer to how they know they are not asleep and dreaming.

In the way one feels one's own headache and observes its location, extent and intensity, there is nothing remotely like feeling someone else's pain or apprehending another's experience. Still less can one see or hear or smell or touch or taste another's experience and come to know it in that way as they do. But then for an empiricist there

arises the question of how one knows anyone else even *has* experience and feeling. Why could all one's own experience not be just as it has been and yet all the other human bodies be mindless lumps (acting as if someone with thoughts and feelings and experiences were in there but really) driven only by blind chemical reactions? This is the problem of other minds, and does not seem to go back earlier than John Stuart Mill in the nineteenth century. Mill offered an argument from analogy to solve this problem. One knows first hand from one's own case that it is one's headache that makes one wince just so, and seeing a similar wince on another's face, one infers by analogy a similar cause, namely a headache, in that other. This argument from analogy proceeded from what must be a single case, namely, oneself alone, so can it justify much certainty? Beyond that, the argument seems not even to address the difficulty the problem of other minds may raise for empiricism, namely, how one could know things like another's pain which one simply could not experience.

3 Logical positivism (cont.)

One can make out connections between Hume's doctrine of ideas and the verifiability theory of meaning, but while Hume used his doctrine as a sceptical axe, the positivists used their story of meaning to undercut scepticism. The problems of the external world and of other minds came to figure centrally on the agenda of modern classical philosophy. Philosophical problems rarely if ever receive solutions consensually certified, so it is understandable (even if not quite right) that such problems apparently sit there forever unsolved. In that case, these problems can easily come to look like demonstrations that it is not known whether there is any matter out there and that it is not known whether there is anyone besides oneself. In that way the agenda of modern philosophy can seem to be a compendium of sceptical theses. But the verifiability theory of meaning was used to argue that these so-called problems are not problems deserving solutions, but pseudoproblems deserving only dissolution. If one asks of a bachelor whether his wife is English, both a yes and a no seem bad answers since both seem to confirm a mistake concealed in the question. This trite example is a very simple illustration of how a misconception can present itself as an insoluble problem. Some of what is or was modernist in philosophy was an idea of criticism not as demonstrating the opposite but rather as unmasking, and thus eliminating, fraudulent questions and problems.

How was this to work? Remember that Descartes designed the dreaming hypothesis with the intention that it and the matter hypothesis issue in exactly the same expectations for sense experience; everything would seem the same to one's senses no matter which of the two were true. So by design no sensory experience could settle the question of which is true; both hypotheses are equally well supported by all perceptual evidence. But on the verifiability theory, meaning is evidence. As the old slogan had it, a difference that makes no difference is no difference: claims that cannot be distinguished on the basis of evidence must coincide. So since no sensory experience could discriminate between the matter and the dreaming hypotheses, and since empiricism is right, and sensory experience is the only evidence against which to judge claims about the world, then the two hypotheses have exactly the same evidence, and so, by verificationism, exactly the same meaning. Since they are synonymous, they are not incompatible at all, but as compatible as possible. Hence, there is no question at all of choosing between them, and so, since that question is what the problem of the external world was, there is no such problem. The problem has been dissolved.

In the case of the problem of other minds, it is granted that all we can observe of other people is their physical behaviour (and, for an audience some sixty or seventy years later, electrochemistry in their central nervous systems). The problem is whether they really have minds and thoughts, or are just mindless chemical machines. But it is granted that both these alternatives would issue in exactly the same observations of others, and, by empiricism, such observation exhausts the evidence for either. Hence, the alternatives do not differ in evidence and so do not differ at all. There is thus no choice between them, and so no problem of other minds.

Perhaps the truly devout will hang on to their faith no matter what. In that case it looks as though nothing could count against it. Equally, no experience really counts as evidence for faith. But if faith is free from sensory evidence, then, since by verificationism meaning *is* evidence, religion is meaningless. It was not an accident that believers found such irreligion (and affronts to ethics and poetry as well) in Ayer's potboiler, *Language, Truth and Logic*. (One thing that is distressing here is that some thought they could defend their faith only by showing it empirically verifiable rather than by rebutting verificationism. The devil has no right to pick the battlefield.)

4 Objections to verificationism

One sort of objection to verificationism begins with the observation that evidence for a present claim about ancient events will be contemporary with the claim rather than with the events. Then the claim about the past seems paradoxically to be about the present, and this paradox seems to be an objection to verificationism. But the objection seems to be founded in a confusion between the date of a report and the date, if any, mentioned by the report. It may be mysterious how words, simultaneous with their present meanings, can reach across time to refer to things past (see Proper names §4). But we have yet to be shown how, if present meaning is present evidence, that mystery is any deeper. Yet if present evidence must be present experience, and experience can only be experience of what is (pretty much) contemporary with it, then the objection may be more forceful. (The finiteness of the speed of light means we can see novae long past, but we cannot choose the temporal distance at which we see while we can choose that at which we refer.)

A monument in the criticism of verificationism is Carl G. Hempel's 'Problems and Changes in the Empiricist Criterion of Meaning' (1950). Let us ask in seeming innocence exactly what the verifiability theory of meaning is. It is natural to put the theory by saying that the meaning of a (declarative) sentence is the body of sensory experience that would conclusively establish the sentence; here verification is required to be knock-down. To this form of the theory, Hempel presents three objections. First, natural science, the positivists' paradigm of meaningfulness, is full of universal hypotheses, though, as the problem of induction shows, no genuinely possible body of evidence ever suffices to establish a universal claim conclusively. Second, suppose that *S* typifies what the positivists count as meaningful while *N* typifies what they count as meaningless nonsense. Note that the disjunction, *S* or *N*, follows from *S*; yet if a part is gibberish, one expects the whole to be so too - how can nonsense follow from sense? Third, let *P* be a feature one can observe an object to exhibit or to lack. One observation would suffice to establish conclusively that there is a thing with feature *P*. But then the sentence 'There is something with *P'* is meaningful, while its negation, being universal, is not, though it seems unbelievable that a sentence be meaningful, while its negation is not. These three arguments are not meant to be decisive. Instead they represent a turn in the dialectic both of Hempel's 1950 paper and of verificationism. But through the 1950s and into the 1960s, there was a rather general presumption that verificationism was dead, and that Hempel's paper was its headstone (see Logical positivism §5).

5 Recent developments

But ideas are hardy. After Hume, the gap between experience and theory grows. Perhaps statements which are just about experience can be tested one by one against experience, but the more theoretical a statement, the less it can be so tested individually. Around 1900 Pierre Duhem argued that theoretical statements can be tested against observation only in relatively large, heterogeneous bodies (see Duhem 1906; Duhem, P.M.M. §4). Half a century later, Quine (1961) inferred from verificationism that the smallest unit with meaning is the smallest unit to which empirical evidence attaches. By Duhem's thesis, such units of evidence are large and heterogeneous in the case of theoretical sentences (see Quine, W.V. §3). But that is the more general case, and in it we cannot distinguish different meanings for different sentences; as it were, mostly there are no propositions. In particular, meanings for the mathematics used in theoretical science cannot be distinguished from meanings for the natural laws used there. Thus we lose all basis for saying such mathematics is true by virtue of its meaning (analytic), while those laws are not certified by their meanings alone. Quine thus uses one central tenet of positivism (verificationism) to undercut another (the analyticity of mathematics - see Analyticity §3). This reopens a path to Platonic metaphysics in mathematics, a path positivism thought to block.

Quine develops verificationism towards a semantic holism rather alien from positivism (see Holism: mental and semantic). But from around 1970 Michael Dummett (1975), taking a cue from the later Wittgenstein (§14), revives a verificationism more reminiscent of positivism. The later Wittgenstein sometimes seems to flirt with intuitionism in the philosophy of mathematics. The strictures intuitionism would impose on classical logic hardly seem obvious. But Dummett suggests that those strictures may fall out of a proper theory of meaning (see Intuitionistic logic and antirealism §§4-6). Meaning, Wittgenstein said, is use, and use, Dummett says, is public; the upshot is what Dummett calls the extrusion of meaning (from, perhaps, the mind into, perhaps, public life). Sometimes he presents this extrusion as the proper response to two challenges, one of acquisition, another of manifestation. The point of the first is that no matter what ideas are innate, a child learning its mother tongue has only observations of which words those around it use in which circumstances on which to base its expressions of its ideas. Meaning must be publicly observable to be learned by speakers (see Private states and language). These challenges are

challenges to an account of meaning in terms of truth-conditions, an account of meaning that may have been present in Frege (1884) and that is explicit in Donald Davidson (1984). The idea here is that the meaning of a sentence is given by the conditions for its truth (see Meaning and truth §1). Taken standardly, there is no guarantee that we can tell of what is true that it is true; Dummett describes such truth as ‘verification transcendent’. If the truth-conditions which constitute meaning transcend recognition, then it would seem there is no guarantee that meaning will be manifest enough to be acquired by native speakers. So, since meaning must be knowable for communication to be possible, verification-transcendent truth must be an illusion. But truth so conceived is crucial to the exposition of classical logic; when we say that a statement or its negation is always true, we mean to allow that neither need be known. If we treat from conditions of truth to those of assertibility, as Dummett thinks we must, the logic we get is, he argues, that of intuitionism.

6 Logical problems

It is a consequence of the rejection of verification-transcendent truth that all truths can be known. This may also be a consequence of verificationism itself: any truth is meaningful and so, by verificationism, has attached to it evidence that can be had, thus making the truth known. Dummett is quite clear about the commitment that all truths can be known. Here it seems important to separate possibility from knowledge. There is, presumably, a natural number *n* such that Caesar had exactly *n* hairs on his head at the instant Brutus’s dagger first penetrated Caesar’s body. To say that it can be known that Caesar then had *n* hairs may be true, while it is pretty clearly false that it is in fact known; the possibility of knowledge should allow for the fact of ignorance.

Take a system for the propositional calculus. Add to it a very weak modal logic in which necessary conditionals with necessary antecedents have necessary consequents (Gödel’s axiom) and in which necessitations of theorems are also theorems; the latter is OK if all our axioms are necessary and all our other rules of inference preserve necessity. This weak modal logic is all we need about necessity and possibility on their own. As for knowledge, we make three postulates. A1 says that all truths can be known. This can be written ‘If *p* then it is possible that (‘$\Diamond$’) it is known that *p*’,

$$p \rightarrow \Diamond \mathrm{Kp}.$$

A2 says that what is known is true. This can be written

$$\mathrm{K}p \rightarrow p.$$

A3 says that when a conjunction is known so are its conjuncts. This can be written

$$\mathrm{K}(p \mathbin{\&} q) \rightarrow (\mathrm{Kp} \mathbin{\&} \mathrm{Kq}).$$

A2 is part of all stories of knowledge, and as for A3, it is hard to see how an explicit conjunction could be known without being understood, and so its conjuncts known. Call the whole system *F*. In *F*, we can prove that all truths are known; this can be written

$$p \rightarrow \mathrm{Kp}.$$

Here is how the proof goes. Suppose that *p* but that it is not known that *p*; suppose, in short, that $p \mathbin{\&} {\sim}\mathrm{Kp}$. By A1, $\Diamond \mathrm{K}(p \mathbin{\&} {\sim}\mathrm{Kp})$. A3 says K distributes across conjunction, and weak though our modal logic is, it is strong enough to make the distribution inside the diamond. So $\Diamond(\mathrm{Kp} \mathbin{\&} \mathrm{K}({\sim}\mathrm{Kp}))$. A2 says the outermost K of the second conjunct can be dropped, and weak though our modal logic is, it is strong enough to drop this K inside the diamond. So $\Diamond(\mathrm{Kp} \mathbin{\&} {\sim}\mathrm{Kp})$. This says an explicit contradiction could be true, a possibility refutable even in our weak modal logic. So our original supposition that $p \mathbin{\&} {\sim}\mathrm{Kp}$ has been reduced to absurdity, and thus we have shown that $p \rightarrow \mathrm{Kp}$. (This argument has a curious provenance. It was published by Fitch in 1964, but Fitch attributes it to the anonymous referee of a paper he submitted, but did not publish, in 1945.)

General omniscience is utterly preposterous. If nothing else, it would collapse the distinction between knowledge and truth, leaving nothing to discover. If not all truths are known, then at least one of the postulates of *F* must be mistaken. Faulting the propositional calculus here seems excessive - like using cannon to hunt sparrows. If one is willing to countenance modality at all, then Gödel’s axiom seems acceptable. Whether the rule of necessitation RN is acceptable depends on whether the axioms of *F* are necessary and the other rules preserve necessity. It is

generally held by those who take necessity seriously that the laws of the propositional calculus are necessary and that, say, *modus ponens* preserves necessity. Gödel's axiom too is generally thought necessary by the friends of modality. That leaves only our axioms A1-A3 of knowledge. A2 says that only truths are known; this claim is generally held to be analytic and thus necessarily true. A3 says that the conjuncts of a known conjunction are known. This claim follows, for example, from the more general claim that the logical consequences of what is known are also known. The more general claim is usually rejected since, for example, one seems to be able to know the axioms of number theory but not to know all its theorems. Nevertheless, it is only A3 and not the more general claim that is at issue here. It is at least difficult to imagine how an explicit conjunction might be known, and hence understood, unless its conjuncts also were known. If so, A3 is necessarily true and, thus, true. That leaves only A1, the focus of our original interest. If A1 is not true, then neither is it necessarily true, in which case RN is not to be trusted in *F*. Quite independently of *F*, there is a suspicious hubris in A1. What would give sentient beings the arrogance to suppose themselves so smart as to be able to figure out everything? Perhaps good manners as well as good judgment urge the denial of A1.

Still, one might feel as if a trick were being played on one; can just this dinky little logic chopping really be all it takes to show there absolutely must be ignorance? This seems like too much out of too little to be believed. So one might think that maybe there is a blunder somewhere, and one thing in the argument that bothers some people is the way it nests K inside itself. So one might consider typing (assigning 'levels' or 'types' to) knowledge, rather as Russell (1908) typed classes, or Tarski (1933) levelled languages. It is by no means obvious how to distinguish types, but custom suggests treating them like numbers. Then we would have not a single K, but rather an infinity K_1, K_2 and so on *ad infinitum*. Aping Russell, our syntax might insist that $\mathrm{K_n}$ be applicable to a formula *p* only if, when there are operators $\mathrm{K_m}$ in *p*, *n* is the successor of the largest such *m*. A2 and A3 could be postulated for each such operator uniformly. The novelty would crop up in A1 to block the first step in the proof that $p \rightarrow \mathrm{K_n p}$. Moreover, as Hans Kamp and Charles Parsons showed independently, the result is a consistent system in which $p \rightarrow \mathrm{K_n p}$ fails.

This result depends on sealing off the positions in which types are mentioned from quantification. For otherwise we could re-introduce K so that K*p* means that $(\exists n)\mathrm{K_n p}$, and in this way recover the old proof that $p \rightarrow \mathrm{Kp}$. We are thus landed in an indefinitely extended realm of types (or levels or contexts) that we have to be able to distinguish and mention. Presumably we could introduce a predicate true of all and only the types. This makes it seem like mere artifice or arbitrary convention (like not using one's hands in soccer) not to allow quantification over types. We cannot even say that knowledge is always of some type. In other words, typing without quantification is more like hiding the paradox than solving it. There is nothing special here about knowledge, for types without quantification also seem equally unsatisfactory solutions to the set-theoretic and semantic paradoxes.

Typing is not the only way one might try to diagnose trickery in *F*. A more radical move would be to take our results about *F* as a refutation of A1, as a proof that not all truths can be known and so as a refutation of the verifiability theory of meaning. What might the falsity of verificationism tell us about meaning? Verificationism would fail if there were differences in meaning which no differences in evidence - and thus, granted empiricism, sensory evidence - answered. That would in turn happen were there two sentences that differed in meaning, which would be most convincing if they could or did differ in truth-value, but such that there would be no possible way to tell by looking, listening or otherwise observing, which is true. Here we have reached a description of Descartes' problem of the external world. The matter hypothesis says that there is matter, while the dreaming hypothesis does not. On the very face of it, the two hypotheses are incompatible, and so not synonymous. If a language is to be usable for communication among its speakers, then they must know what its expressions mean. In general, when two of its expressions seem non-synonymous to its speakers, they will be non-synonymous. The dreaming and matter hypotheses seem non-synonymous to us. So if the verifiability theory of meaning says they are synonymous, then so much the worse for verificationism. As a response to the problem of the external world, verificationism was whistling in the dark.

Differences in meaning are not guaranteed to be rooted in sensory differences. Put otherwise, in a more archaic vocabulary, ideas are not images; nor are they faint copies of impressions. This is hardly hot news, and yet it raises in an acute way a worry about whether an empirically acceptable account of what meaning is can be had. What kind of a thing is a meaning? How can that be a difficult question given that since you and I know a natural language, we are both familiar with lots of meanings?

See also: Logical and mathematical terms, glossary of

W.D. HART

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz, 2nd edn, 1946. (Widely read and highly provocative. The second edition includes a famous introduction in which Ayer wrestles with objections to the verification principle.)

Carnap, R. (1928) *Scheinprobleme in der Philosophie*, Berlin; trans. R.A. George, *Pseudoproblems in Philosophy*, with *The Logical Structure of the World*, Berkeley, CA: University of California Press, 1967.(Two seminal works by the leading member of the Vienna Circle.)

Craig, E.J. (1982) 'Meaning, Use and Privacy', *Mind* 91: 541-64.(The first section of this paper is a critical discussion of Moritz Schlick's argument in favour of the verification principle.)

Davidson, D. (1984) 'Truth and Meaning', in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 17-36.(A classic statement of the view that, taken properly, reference fixes sense.)

Descartes, R. (1641) *Meditationes de prima philosophia*, trans., 'Meditations on First Philosophy', in *Descartes: Philosophical Writings*, trans. and ed. G.E.M. Anscombe and P.T. Geach, Indianapolis, IN: Bobbs-Merrill, 1971.(The book that started modern classical philosophy.)

Duhem, P. (1906) *La Théorie physique, son objet et sa structure*, Paris: Chevalier et Rivière, 2nd edn, 1913; trans.P.P. Wiener, *The Aim and Structure of Physical Theory*, Princeton, NJ: Princeton University Press, 1992. (The source of the view that even all possible experience fails to settle all theoretical questions about the world.)

Dummett, M. (1975) 'The Philosophical Basis of Intuitionistic Logic', *Studies in Logic and the Foundations of Mathematics* 80: 5-40; repr. in *Truth and Other Enigmas*, London: Duckworth, 1978.(An authoritative statement of Dummett's contemporary version of 'meaning-empiricism'. No easy read, but the most interesting defence of intuitionism in print.)

Edgington, D. (1985) 'The Paradox of Knowability', *Mind* 94: 557-68.(Discusses issues raised in §6 of this entry.)

Fitch, F.B. (1964) 'A Logical Analysis of Some Value Concepts', *Journal of Symbolic Logic* 28: 135-42.(The first publication of the proof that if all truths can be known, then all truths are in fact known.)

Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, 2nd edn, 1980, esp. Introduction.(Still the best single book in philosophy of mathematics ever written.)

Hempel, C.G. (1950) 'Problems and Changes in the Empiricist Criterion of Meaning', *Revue internationale de philosophie* 4: 41-63; repr. in L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University of Illinois Press, 1952.(Highly influential critique of logical positivism.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. D.G.C. MacNabb, London: Fontana, 1962.(A masterpiece. See book 1 for Hume's sceptical account of meaning.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1963.(Another masterpiece. The second edition of 1787 contains the discussion outlined in §1.)

Quine, W.V. (1961) 'Two Dogmas of Empiricism', in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1980.(Proposes a revamped meaning-empiricism of a holistic rather than molecular or atomistic sort.)

Russell, B.A.W. (1908) 'Mathematical Logic as based on the Theory of Types', *American Journal of Mathematics* 30: 222-62; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992, 59-102.(Perhaps Russell's single best piece of work in logic.)

Tarski, A. (1933) Pojęcie prawdy w językach nauk dedukcyjnych, Warsaw; trans. J.H. Woodger (1956), 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(The start of modern theories of truth.)

Williamson, T. (1982, 1987) 'Intuitionism Disproved?', *Analysis* 42: 203-8, 'On Knowledge and the Unknowable', *Analysis* 47: 154-8, and 'On the Paradox of Knowability', *Mind* 96: 256-61.(Articles which discuss logical issues of the kind described in §6 of this entry.)

Meaning in Islamic philosophy

The discussion of the notion of meaning in Islamic philosophy is heavily influenced by theological and legal debates about the interpretation of Islam, and about who has the right to pronounce on interpretation. The introduction of Greek philosophy into the Islamic world produced a new set of authorities on how to interpret texts, and this led to arguments over the potential benefits of the new approaches as compared with the traditional Islamic sciences. The discussion came to centre on the nature of ambiguity, equivocation and analogy, with different philosophers adopting diverse theories and thus attaining a variety of conclusions about how to interpret meaning. These variations have powerful implications for the understanding of their thought. Not only do the different approaches result in different conclusions, they also represent different approaches to the whole philosophical enterprise. The topic of meaning is not so much an aspect of Islamic philosophy as an interpretation of how to do Islamic philosophy itself. The main issues focus on identifying the people best qualified to interpret texts, valid interpretations of the texts, and the notion of meaning that should be employed in our understanding of the texts.

1 Religious context

Since the Qur'an was transmitted in Arabic, an understanding of the nature of that language is a vital aspect to an understanding of the text itself. Those brought up within the traditions of jurisprudence, grammar and theology were of the opinion that they were the best positioned to pronounce on the meaning of the text (see Islamic theology §1). All scriptures require interpretation, and a wide diversity of views arose within the Islamic community over the correct reading of much of the Qur'an, with the creation of different schools of thought based upon political and religious divisions such as those between the Sunni and the Shiʿi communities and between the Ashʿarites and the Muʿtazilites (see Ashʿariyya and Muʿtazila §1). It was accepted quite early on that while some parts of the Qur'an are clear, others are less easy to grasp and so require more complex interpretations. Some passages are *zahir* (exoteric) while others are *batin* (esoteric), and naturally the commentators disagreed on occasion over which texts fell into which category and how the esoteric texts were to be read. In a religious text, of course, even the apparently plain and commonplace can be given a richer and deeper interpretation, and different interpreters produced different interpretations. None of these was arbitrary. All were based upon argument and a variety of Islamic sources, but this hermeneutic process was capable of arriving at a variety of conclusions.

2 Language versus logic

The arrival of Greek logic in the Islamic world caused great controversy in the fourth century AH/tenth century AD. It brought with it the view that logic is superior to language since the latter deals only with the contingent and conventional, while logic encompasses the necessary and the universal. Language deals only with *alfaz* (utterances), while logic gets to the heart of the matter by analysing *maʿani* (concepts) themselves. This controversy was explored in the celebrated debate between the grammarian Abu Saʿid al-Sirafi and the logician Abu Bishr Matta in Baghdad in AH 320/AD 932. Matta argued that logic was more important than language since the meanings which are embedded in a particular language are analysable without reference to that particular language. Those meanings could exist in a whole variety of languages. Logic is the only rigorous tool for judging when language is used correctly or otherwise, and the logician is the best qualified to adjudicate on such issues.

This was a very important debate, since if Matta were to carry the day it would imply that the traditional approach to Islamic texts rests on an error. Only logic, as understood by the Greeks and a non-Muslim such as Matta, is a sound vehicle for the understanding of texts. Al-Sirafi was clear on the significance of the debate, and he argued that it is impossible to separate language and logic in the way Matta wants. Although logic is useful at one level in dealing with concepts, it is far from comprehensive and is better suited to analysing the Greek language than the Arabic. Problems in understanding Arabic texts can only be answered by a good understanding of the Arabic language and the culture which surrounds it. Logic by itself is not sufficient (see Logic in Islamic philosophy).

This sort of debate continued in the Islamic world for some time, albeit in a more sophisticated form. The debate is highly significant, since it is really about the appropriate notion of meaning to be employed. Must that notion of meaning be taken from the context in which the text to be analysed has itself originated, or can it come from elsewhere? How one answers this question has radical implications for the way in which Islamic philosophy is to

be pursued, and the protagonists of the argument are well aware of this.

3 Ambiguity, equivocation and analogy

There are two main theories of meaning in Islamic philosophy, one broadly Neoplatonic and adopted by Ibn Sina (Avicenna) and al-Ghazali, and one broadly Aristotelian and defended by Ibn Rushd (Averroes). On the former account, the definition of *x* does not include the existence of *x*, so that something else is required to move *x* from the realm of possibility to the realm of actuality if it is to be instantiated. For Ibn Sina the mover is another thing, ultimately God, which causally necessitates the change, while for al-Ghazali it is God who directly established quite arbitrary rules for the behaviour of contingent phenomena. This means that for al-Ghazali it is possible to think of something happening without its customary explanation, that is, without a natural explanation. God could always have brought it about that people on occasion write philosophy books without possessing a head, or that straw does not burn when in contact with fire.

Ibn Rushd operates with a different concept of meaning, in accordance with which the cause of an event is part of its definition or essence. If headless people were to start composing philosophical books we should need to construct an entirely novel conceptual scheme for such a possibility to make sense. Al-Ghazali, by contrast, argues that we can conceive of such changes to our conceptual scheme by using our imagination, and failure to approve the possibility of such changes is merely a result of intellectual laziness, not the flouting of necessity. It is possible, then, for God to do anything he wishes so long as it is logically possible. God can resurrect the dead, he can intervene in the natural world, he can be aware of events in the world, he can create that world at any time he wishes out of nothing. These are all actions which Ibn Rushd argued are not possible even for God, if what is meant when talking about such actions is similar to what we mean when we talk about what we do. Al-Ghazali suggests that any analysis of the properties of God which interprets them as equivocal, ambiguous or metaphorical is a subtle attack upon the notion of God in its religious sense. It involves pretending to make God part of one’s metaphysics but changing the way in which one talks about him to such an extent that he no longer is equivalent to the God of Islam.

For Ibn Rushd, there are serious problems in talking about God using the same sort of language we use about ourselves. God cannot be defined in terms of a genus or species since he cannot consist of a plurality of qualities as we do. Rather, he is the exemplar of all things, and we must work towards a conception of him by thinking analogically about the things in the world of which we have experience. The way in which al-Ghazali talks about God seems to be in line with religion, but really it involves treating the deity as someone much like ourselves but more so, and as a plurality of predicates rather than a complete unity. Ibn Rushd goes along with Aristotle’s argument that there can be no priority or posteriority within the same genus (see Aristotle §§7-8), and so develops a theory of meaning which is based upon the *pros hen* notion of equivocation rather than the genus-species relation. We can use the same language about God as we use about ourselves, but we should realize that the latter type of use is derivative upon the former, since the latter concepts are aspects of the paradigm which is to be found when we talk about God. When we use the same concept to refer to God and to ourselves we are speaking not univocally but equivocally, acknowledging the very real difference which exists between the level of the human and the level of the divine.

4 The main debate

In al-Ghazali’s concept of meaning, the appropriate people to interpret religious texts are those professionally involved in the religion, such as theologians, jurisprudents and so on. They need to understand the religious context of the text and apply the terms univocally to God and to his creatures. The correct way to think analytically is to adopt the methodology of the thought-experiment, holding ideas together in the imagination to see if they can be combined without contradiction. If we can imagine dead people being resurrected and leaving their graves to continue a physical existence somewhere else, in hell or paradise, then there is nothing impossible about that idea and there is no need to suggest that it is a metaphorical or equivocal reference to something else. If the Qur’an refers to physical resurrection, and if we can think about physical resurrection without contradiction, then why not just accept that what is meant by physical resurrection is what we would normally understand by that miraculous event?

It follows from Ibn Rushd’s approach that the best people to interpret difficult theological passages are not the

theologians but the philosophers, since only the latter are skilled in understanding analytical and demonstrative thought. The philosopher can understand how a religious text may embody a whole range of meanings, some intended for a more sophisticated audience and some designed for a more naïve and practically-oriented audience. The latter might understand by physical resurrection that the consequences of what we do outlive us, and the moral status of what we do affects us in this life and others after we are dead. They might find this easier to understand if they can think of themselves, or something like themselves, surviving their death. The more sophisticated audience would understand by the religious text what the philosophers understand by it, what it really means. This sort of audience does not require the rhetorical and poetic language which is capable of moving the largest section of the community to action. Only philosophers are capable of resolving the meaning of a text once and for all. The other theoreticians who work within the Islamic sciences, such as theology and jurisprudence, will constantly disagree about the appropriate meaning, and as a result they will sow discord in the community. The role of the philosopher is to resolve the meaning of the text and then communicate that meaning to different kinds of audiences in different kinds of ways, each suited to the limitations of the audience to understand the real point of the text.

It can thus be seen that the issue of meaning in Islamic philosophy concerns not only the philosophy of language but also politics. It is linked to the question as to who is entitled to derive the meanings of a text and how that meaning may be communicated to others. Although the focus of discussion is generally on the relationship between ordinary language and language about God, it has far broader implications, and leads to a diversity of views on how to do philosophy itself.

See also: Aesthetics in Islamic philosophy; Epistemology in Islamic philosophy; al-Ghazali; Ibn Rushd; Ibn Sina; Logic in Islamic philosophy; Political philosophy in classical Islam

OLIVER LEAMAN

References and further reading

Abed, A. (1991) *Aristotelian Logic and the Arabic Language in Alfarabi*, Albany, NY: State University of New York Press.(A detailed and scholarly investigation of the concept of meaning and the logic versus language debate.)

Abed, S. (1996) ‘Language’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 52, 898-925.(Interesting and detailed account of the role that language plays in Islamic philosophy.)

Ibn Rushd (1179-80) *Fasl al-maqal (Decisive Treatise)*, ed. and trans. G. Hourani, *Averroes on the Harmony of Religion and Philosophy*, London: Luzac, 1976.(An outstandingly clear exposition of the basic themes of Ibn Rushd on different kinds of meaning.)

Ibn Rushd (1180) *Tahafut al-tahafut (The Incoherence of the Incoherence)*, trans. S. Van den Bergh, *Averroes' Tahafut al-tahafut*, London: Luzac, 1978.(A comprehensive comparison of the views of al-Ghazali and Ibn Rushd.)

Leaman, O. (1954-) ‘*Ma'na*’, in *Encyclopedia of Islam*, Leiden: Brill, 347.(An account of the philosophical use of this Arabic term, equivalent in meaning to ‘concept’ or ‘idea’.)

Leaman, O. (1988) *Averroes and His Philosophy*, Oxford: Oxford University Press; 2nd edn, 1997.(A general account of the thought of Averroes (Ibn Rushd), with the emphasis upon his approach to language.)

Leaman, O. (1997) ‘Logic and Language in Islamic Philosophy’, in B. Carr and I. Mahalingam (eds) *Companion Encyclopedia of Asian Philosophy*, London: Routledge, 950-64.(Summary of some of the major approaches to meaning in Islamic philosophy.)

Mahdi, M. (1970) ‘Language and Logic in Classical Islam’, in G.E. von Grunebaum (ed.) *Logic in Classical Islamic Culture*, Wiesbaden: Harrassowitz, 51-83.(The standard discussion of the battle between the logicians and the grammarians.)

Meaning, Indian theories of

The term artha *in Sanskrit is used for the notion of meaning, in the widest sense of the word 'meaning'; it can be the meaning of words, sentences and scriptures, as well as of nonlinguistic gestures and signs. Its meaning ranges from a real object in the external world referred to by a word to a mere concept of an object which may or may not correspond to anything in the external world. The differences regarding what 'meaning' is are argued out by the philosophical schools of Nyāya, Vaiśeṣika, Mīmāṃsā, Buddhism, Sanskrit grammar and Sanskrit poetics. Among these, Nyāya, Vaiśeṣika and Mīmāṃsā have realistic ontologies. Mīmāṃsā focuses mainly on interpreting the Vedic scriptures. Buddhist thinkers generally depict language as giving a false picture of reality. Sanskrit grammar is more interested in language than in ontology, while Sanskrit poetics focuses on the poetic dimensions of meaning.*

Generally, the notion of meaning is stratified into three or four types. First there is the primary meaning. If this is inappropriate in a given context, then one moves to a secondary meaning, an extension of the primary meaning. Beyond this is the suggested meaning, which may or may not be the same as the meaning intended by the speaker. Specific conditions under which these different varieties are understood are discussed by the schools.

The various Indian theories of meaning are closely related to the overall stances taken by the different schools. Among the factors which influence the notion of meaning are the ontological and epistemological views of a school, its views regarding the role of God and scripture, its focus on a certain type of discourse, and its ultimate purpose in theorizing.

1 *Artha* in different Indian traditions

The most common Sanskrit term for meaning is *artha*. In the Western literature on the notion of meaning in the Indian tradition, various terms, such as 'sense', 'reference', 'denotation', 'connotation', 'designatum' and 'intension', have been used to render the Sanskrit. However, these terms carry specific nuances of their own, and no single term adequately conveys the idea of *artha*. *Artha* basically refers to the object signified by a word. In numerous contexts, it stands for an object in the sense of an element of external reality. For instance, Patañjali (second century BC) says that when a word is pronounced, an *artha*, 'object', is understood. So in the case of 'Bring a bull' and 'Eat yoghurt', it is the *artha* that is brought in and the *artha* that is eaten.

The logicians and ontologists belonging to the schools of Nyāya and Vaiśeṣika, and the later combined school of Nyāya-Vaiśeṣika, set up an ontology containing substances, qualities, actions, relations, generic and particular properties, and absences (see Nyāya-Vaiśeṣika §§4-5). With this realistic ontology in mind, they argue that if the relation between a word and its *artha* were a natural ontological relation, there should be real experiences of burning and cutting in one's mouth after hearing words like *agni* ('fire') and *asi* ('sword'). Therefore this relation must be a conventional one (*saṃketa*), the convention being established by God as part of his initial acts of creation. The relationship between a word and the object it refers to is thought to be the desire of God that such-and-such a word should refer to such-and-such an object. It is through this established conventional relationship that a word reminds the listener of its meaning.

The school of Mīmāṃsā represents the tradition of the exegesis of the scriptural Vedic texts (see Mīmāṃsā). However, in the course of discussing and perfecting principles of interpretation, this school developed a full-scale theory of ontology and an important theory of meaning. For the Mīmāṃsakas, the primary tenets are that the Vedic scriptural texts are eternal and uncreated, and that they are meaningful. In this orthodox system, which remarkably defends the scriptures but dispenses with the notion of God, the relationship between a word and its meaning is innate and eternal. Both Nyāya-Vaiśeṣikas and Mīmāṃsakas regard language as referring to external states of the world and not just to conceptual constructions.

While the various schools of Buddhism differ among themselves concerning the nature of the external world, they all seem to agree that language relates only to a level of conceptual constructions which have no direct relationship to the actual state of the world. The tradition represented by Theravāda and the Vaibhāṣikas argued that a word refers to a thing which, in reality, is nothing but a composite entity made up of components which are momentary and in a continual flux (see Momentariness, Buddhist doctrine of). The components, the momentary atomic

elements (*dharma*), are presumed to be more real, but words do not refer to this level of reality. Thus language gives us a less than true picture of what is out there. Other schools of Buddhism, such as Vijñānavāda, reduced everything to fleeting states of consciousness (*vijñāna*). From this point of view, the objects referred to by words are not even composites. They are more like fictions (*vikalpa*) or illusions (*māyā*) created by a magician. The Mādhyamika school of Buddhism focused on the essential emptiness (*śūnyatā*) of all objects which are subject to dependent origination (*pratītyasamutpāda*). This also leaves language far away from the level of reality, which in this case is the level of emptiness (see Buddhist concept of Emptiness). Later Buddhist logicians, such as Dignāga, developed a theory of word-meaning which we can call the 'exclusionary' theory of meaning (*apoha*). Briefly stated, if a word refers only to a conceptual construction, and not to a state of reality, then how are we to construe the conceptual construction? The *apoha* theory proposed by the Buddhist logicians says that external reality ultimately consists of momentary atomic elements which are so individualized and unique (*svalakṣaṇa*) that they are beyond perception and characterization. Our perceptions and conceptions involve generalization (*sāmānyalakṣaṇa*), and hence do not correspond to reality. A concept which corresponds to a given word must be finally construed as being nothing more than the exclusion of all other concepts. This theory ultimately says that all concepts are different from each other, and yet they cannot be defined by making a reference to any level of reality (see Nominalism, Buddhist doctrine of).

The tradition of the grammarians, beginning with Bhartṛhari (fifth century), seems to have followed a middle path between the realistic theories of reference (*bāhyārthavāda*) developed by Nyāya-Vaiśeṣika and Mīmāṃsā on the one hand, and the notional/conceptual theories (*vikalpa*) of the Buddhists on the other. For them, the meaning of a word is closely related to the level of understanding. Whether or not things are real, we do have concepts. These concepts form the content of a person's cognitions derived from language. Without necessarily denying the external reality of objects in the world, the grammarians claimed that the meaning of a word is only a projection of intellect (*bauddhārtha*, *buddhipratibhāsa*). The examples they offer, such as *śaśasṛōga* ('horn of a rabbit') and *vandhyāsuta* ('son of a barren woman'), remain meaningful within this theory. The Sanskrit grammarians are thus not concerned with the truth-functional value of linguistic expressions. For them, the truth of an expression and its meaningfulness are not to be equated.

2 Varieties of meaning

By the middle of the second millennium AD, a certain uniformity came about in the technical terminology used by different schools. The prominent schools in this period were the new school of Nyāya - Navya-Nyāya - initiated by Gaōgeśa, and the schools of Mīmāṃsā, Vedānta, and Sanskrit grammar. While all these schools engaged in pitched battles against each other, they seem to have accepted the terminological lead of the neo-logicians, the Navya-Naiyāyikas. Following the discussion of *artha* by the neo-logician Gadādhara, we can state the general framework of a semantic theory. Other schools accepted this general terminology, with some variations.

It may be said that the term *artha* stands for the object or content of a verbal cognition or a cognition which results from hearing a word (*śābdabodhaviṣaya*). Such a verbal cognition results from the cognition of a word (*śābdajñāna*) on the basis of an awareness of the signification function pertaining to that word (*padaniṣṭhavṛttijñāna*). Depending upon the kind of signification function (*vṛtti*) involved in the emergence of the verbal cognition, the meaning belongs to a distinct type. In general terms:

1 When a verbal cognition results from the primary signification function (*śakti*/*abhidhāvṛtti*/*mukhyavṛtti*) of a word, the object or content of that cognition is called primary meaning (*śakyārtha*/*vācyārtha*/*abhidheya*).
2 When a verbal cognition results from the secondary signification function (*lakṣaṇāvṛtti*/*guṇavṛtti*) of a word, the object or content of that cognition is called secondary meaning (*lakṣyārtha*).
3 When a verbal cognition results from the suggestive signification function (*vyañjanāvṛtti*) of a word, the object or content of that cognition is called suggested meaning (*vyaōgyārtha/dhvanitārtha*).
4 When a verbal cognition results from the intentional signification function (*tātparyavṛtti*) of a word, the object or content of that cognition is called intended meaning (*tātparyārtha*).

Not all schools of Indian philosophy accept all of these different kinds of signification functions for words, and they hold substantially different views on the nature of words, meanings, and the relations between words and meanings. However, the above terminology holds true, in general, for most of the schools. Let us note some of the

important differences. Mīmāṃsā claims that the sole primary meaning of the word ‘bull’ is the generic or class property (*jāti*) - ‘bullness’, say - and the individual object which possesses this generic property, namely a particular bull, is only secondarily and subsequently understood from the word ‘bull’. The school called Kevalavyaktivāda, a segment of the Nyāya school, argues that a particular individual bull is the sole primary meaning of the word ‘bull’, while the generic property bullness is merely a secondary meaning. Nyāya generally argues that the primary meaning of a word is an individual object qualified by a generic property (*jātiviśiṣṭavyakti*), both being perceived simultaneously.

Sanskrit grammarians distinguish between various different kinds of meanings (*artha*). The term *artha* stands for an external object (*vastumātra*), as well as for the object that is intended to be signified by a word (*abhidheya*). The latter - that is, meaning in a linguistic sense - can be meaning in a technical context (*śāstrīya*), such as the meaning of an affix or a stem, or it can be meaning as understood by people in actual communication (*laukika*). Then there is a further distinction. Meaning may be something directly intended to be signified by an expression (*abhidheya*), or it can be something which is inevitably signified (*nāntarīyaka*) when something else is really the intended meaning. Everything that is understood from a word on the basis of some kind of signification function (*vṛtti*) is covered by the term *artha*. Different systems of Indian philosophy differ from each other over whether a given cognition is derived from a word on the basis of a signification function, through inference (*anumāna*), or through presumption (*arthāpatti*). If a particular item of information is deemed to have been derived through inference or presumption, it is not included in the notion of word-meaning.

3 Other dimensions of *artha*

The scope of *artha* is actually not limited in Sanskrit texts to what is usually understood as the domain of semantics in the Western literature. It covers elements such as gender (*liōga*) and number (*saṃkhyā*), as well as such semantic/syntactic roles (*kāraka*) as ‘agentness’ (*kartṛtva*) and ‘objectness’ (*karmatva*). Tenses such as the present, past and future, and moods such as the imperative and optative are also traditionally included in the *arthas* signified by a verb root or an affix.

Another aspect of the concept of *artha* is revealed in the theory of *dyotyārtha*, ‘co-signified’ meaning. According to this theory, to put it in simple terms, particles such as *ca* (‘and’) do not have any lexical or primary meaning. They are said to help other words used in constructions with them to signify special aspects of their meaning. For instance, in the phrase ‘John and Tom’, the meaning of grouping is said to be not directly signified by the word ‘and’. The theory of *dyotyārtha* argues that grouping is a specific meaning of the two words ‘John’ and ‘Tom’, but that these two words are unable to signify it if used by themselves. The word ‘and’ used along with them is said to work as a catalyst that enables them to signify this special meaning.

The problem of use and mention of words is also handled by Sanskrit grammarians by treating the word itself as a part of the meaning it signifies. This is a unique way of handling this problem.

4 Different views regarding sentence-meaning

Most schools of Indian philosophy have an atomistic view of meaning and the meaning-bearing linguistic unit. This means that a sentence is put together by combining words and words are put together by combining morphemic elements such as stems, roots and affixes. The same applies to meaning. The word-meaning may be viewed as a fusion of the meanings of stems, roots and affixes, and the meaning of a sentence may be viewed as a fusion of the meanings of its constituent words. Beyond this generality, different schools have specific proposals. The tradition of Prābhākara Mīmāṃsā proposes that the words of a sentence already convey contextualized/connected meanings (*anvitābhidhāna*) and that the sentence-meaning is not different from a simple addition of these inherently connected word-meanings. On the other hand, the Naiyāyikas and the Bhāṭṭa Mīmāṃsakas propose that the words of a sentence taken by themselves convey uncontextualized/unconnected meanings, and that these uncontextualized word-meanings are subsequently brought into a contextualized association with each other (*abhihitānvaya*). Therefore, sentence-meaning is different from word-meanings, and is communicated through the concatenation (*saṃsarga*) of words rather than by the words themselves. This is also the view of the early grammarians such as Patañjali and Kātyāyana.

For the later grammarian-philosopher Bhartṛhari, however, there are no divisions in speech acts and in

communicated meanings. He says that only a person ignorant of the real nature of language believes the divisions of sentences into words, stems, roots and affixes to be real. Such divisions are useful fictions and have an explanatory value in grammatical theory, but have no reality in communication. In reality, there is no sequence in the cognitions of these different components. The sentence-meaning becomes an object or content of a single flash of cognition (*pratibhā*).

5 Some important conceptions

The terms *śakyatāvacchedaka* and *pravṛttinimitta* signify a property which determines the inclusion of a particular instance within the class of possible entities referred to by a word. It is a property whose possession by an entity is the necessary and sufficient condition for a given word to be used to refer to that entity. Thus the property of potness may be viewed as the *śakyatāvacchedaka* controlling the use of the word 'pot'.

Lakṣaṇāvṛtti (the 'secondary signification function') is invoked in situations where the primary meaning of an utterance does not appear to make sense in view of the intention behind the utterance, causing one to look for a secondary meaning. However, the secondary meaning is always something that is related to the primary meaning in some way. For example, the expression *gaōgāyāṃ ghoṣaḥ* literally refers to a colony of cowherds on the river Ganges. Here it is argued that one obviously cannot have a colony of cowherds sitting on top of the Ganges. This would clearly go against the intention of the speaker. Thus there is both a difficulty in justifying the linkage of word-meanings (*anvayānupapatti*) and a difficulty in justifying the literal or primary meaning in relation to the intention of the speaker (*tātparyānupapatti*). These interpretive difficulties nudge one away from the primary meaning of the expression to a secondary meaning which is related to that primary meaning. Thus we understand the expression as referring to a colony of cowherds on the *bank* of the Ganges.

It is the third level of meaning, or *vyañjanāvṛtti* ('suggestive signification function'), which is analysed and elaborated by authors such as Ānandavardhana (ninth century) in the tradition of Sanskrit poetics. Consider the following instance of poetic suggestion. Her husband away on a long journey, a lovelorn young wife instructs a visiting young man: 'My dear guest, I sleep here and my nightblind mother-in-law sleeps over there. Please make sure you do not stumble at night.' The suggested meaning is an invitation to the young man to come and share her bed. Thus the poetic aspect of language goes well beyond the levels of lexical and metaphorical meanings, and heightens aesthetic pleasure through such suggestions.

6 Why the differences?

The nuances of these different theories are closely related to the markedly different interests of the schools within which they developed. Scholars of Sanskrit poetics were interested in the poetic dimensions of meaning. Grammarians were interested in language and cognition, but had little interest in ontology. For them, words and meanings had to be explained irrespective of one's metaphysical views. Nyāya-Vaiśeṣikas were primarily interested in logic, epistemology and ontology, and argued that a valid sentence was a true picture of a state of reality. The foremost goal of Mīmāṃsā was to interpret and defend the Vedic scriptures. Thus, for Mīmāṃsā, meaning had to be eternal, uncreated and unrelated to a person's intention, because its word *par excellence*, the Vedic scriptures, was eternal, uncreated and beyond any authorship, divine or human. The scriptural word was there to instruct people on how to perform proper ritual and moral duties. The Buddhists, on the other hand, aimed at weaning people away from all attachment to the world, and hence at showing the emptiness of everything, including language. They were more interested in demonstrating how language fails to portray reality, than in explaining how it works. The theories of meaning were thus a significant part of the total agenda of each school and need to be understood in their specific contexts.

See also: Interpretation, Indian theories of; Language, Indian theories of; Language, philosophy of; Meaning in Islamic philosophy

MADHAV M. DESHPANDE

References and further reading

Biardeau, M. (1967) *Théorie de la connaissance et philosophie de la parole dans le brahmanisme classique (Theory of the Understanding and Philosophy of the Word in Classical Brahmanism),* Paris: Mouton.(A comprehensive discussion of theories of meaning and their evolution and interaction in the traditions of

Sanskrit grammar, Mīmāṃsāand Nyāya-Vaiśeṣika; no English translation is available as yet.)
Deshpande, M.M. (1992) *The Meaning of Nouns: Semantic Theory in Classical and Medieval India*, Dordrecht: Kluwer Academic Publishers.(The bulk of the book is an annotated translation of a seventeenth-century Sanskrit text on the meaning of nouns. The introduction covers the history of a number of semantic theories in Sanskrit grammar and philosophy.)
Matilal, B.K. (1971) *Epistemology, Logic and Grammar in Indian Philosophical Analysis*, The Hague/Paris: Mouton.(Very readable general introduction to Indian approaches to meaning, logic and language.)
Matilal, B.K. (1985) *Logic, Language and Reality: An Introduction to Indian Philosophical Studies*, Delhi: Motilal Banarsidass.(More extensive, but still very accessible, discussion of a wide range of issues in Indian philosophy of language.)
Raja, K. (1963) *Indian Theories of Meaning*, Adyar Library Series 91, Madras: Adyar Library and Research Centre.(A somewhat dated, but still very useful, account.)

Measurement, theory of

A conceptual analysis of measurement can properly begin by formulating the two fundamental problems of any measurement procedure. The first problem is that of representation, justifying the assignment of numbers to objects or phenomena. We cannot literally take a number in our hands and 'apply' it to a physical object. What we can show is that the structure of a set of phenomena under certain empirical operations and relations is the same as the structure of some set of numbers under corresponding arithmetical operations and relations. Solution of the representation problem for a theory of measurement does not completely lay bare the structure of the theory, for there is often a formal difference between the kind of assignment of numbers arising from different procedures of measurement. This is the second fundamental problem, determining the scale type of a given procedure.

Counting is an example of an absolute scale. The number of members of a given collection of objects is determined uniquely. In contrast, the measurement of mass or weight is an example of a ratio scale. An empirical procedure for measuring mass does not determine the unit of mass. The measurement of temperature is an example of an interval scale. The empirical procedure of measuring temperature by use of a thermometer determines neither a unit nor an origin. In this sort of measurement the ratio of any two intervals is independent of the unit and zero point of measurement.

Still another type of scale is one which is arbitrary except for order. Moh's hardness scale, according to which minerals are ranked in regard to hardness as determined by a scratch test, and the Beaufort wind scale, whereby the strength of a wind is classified as calm, light air, light breeze, and so on, are examples of ordinal scales.

A distinction is made between those scales of measurement which are fundamental and those which are derived. A derived scale presupposes and uses the numerical results of at least one other scale. In contrast, a fundamental scale does not depend on others.

Another common distinction is that between extensive and intensive quantities or scales. For extensive quantities like mass or distance an empirical operation of combination can be given which has the structural properties of the numerical operation of addition. Intensive quantities do not have such an operation; typical examples are temperature and cardinal utility.

A widespread complaint about this classical foundation of measurement is that it takes too little account of the analysis of variability in the quantity measured. One important source is systematic variability in the empirical properties of the object being measured. Another source lies not in the object but in the procedures of measurement being used. There are also random errors which can arise from variability in the object, the procedures or the conditions surrounding the observations.

1 First fundamental problem: representation

Since the days of the Greek geometers, mathematicians, philosophers and empirical scientists have been discussing the nature of measurement. Eudoxus' development of the theory of proportion, which forms the substance of Book V of Euclid's *Elements,* is probably the first sustained analysis directly relevant to measurement. A systematic approach to the subject of measurement may well begin by formulating the two fundamental problems of any procedure of measurement. The first problem is to justify the assignment of numbers to objects or phenomena.

The early history of mathematics shows how difficult it was to divorce arithmetic from particular empirical structures. Thus the early Egyptians could not think of $2 + 3$, but only of 2 bushels of wheat plus 3 bushels of wheat. From a logical standpoint, there is just one arithmetic of numbers, not an arithmetic for bushels of wheat, and a separate arithmetic for quarts of milk. The first problem for a theory of measurement is to show how this one arithmetic of numbers may be applied in a variety of empirical situations. This is done by showing that the relevant structure of the empirical situation investigated is isomorphic to an arithmetical structure. The purpose of the definitions of isomorphism given for each kind of measurement considered is to make the rough-and-ready intuitive idea of same structure precise. The great significance of finding such an isomorphism of structures is that we may then use all our familiar knowledge of computational methods, as applied to the arithmetical structure, to infer facts about the isomorphic empirical structure.

A procedure of measurement is needed in any area of science when we desire to pass from simple qualitative observations to the quantitative observations necessary for the precise prediction or control of phenomena. To justify this transition we need an algebra of empirically realizable operations and relations which can be shown to be isomorphic to an appropriately chosen numerical algebra. Satisfying this requirement is the fundamental problem of measurement representation.

2 Second fundamental problem: uniqueness

Solution of the representation problem for a theory of measurement is not enough. There is often a formal difference between the kind of assignment of numbers arising from different procedures of measurement. Consider the following three statements:

(1) The number of people now in this room is seven.
(2) Yesterday John weighed 150 lbs.
(3) The maximum temperature tomorrow will be $89°\mathrm{F}$.

Here we may formally distinguish three kinds of measurement. Counting is an example of an ’absolute’ scale. The number of members of a given collection of objects is determined uniquely. There is no arbitrary choice of a unit or zero available. In contrast, the measurement of mass or weight is an example of a ’ratio’ scale. Empirical procedures for measuring mass do not determine the unit of mass. The choice of a unit is an empirically arbitrary decision, but the ratio of any two masses is independent of such choice. The measurement of distance is a second example of a ratio scale. The ratio of the distance between Palo Alto and San Francisco to the distance between Washington and New York is the same whether the measurement is made in miles or metres.

The measurement of temperature is an example of an interval scale. The empirical procedure of measuring temperature by use of a thermometer determines neither a unit nor an origin. In this sort of measurement the ratio of any two intervals is independent of the unit and zero point of measurement. Examples other than measurement of temperature are measurements of temporal dates, geographical longitude and cardinal utility.

In terms of the notion of absolute, ratio and interval scales we may formulate the second fundamental problem for a theory of measurement: determine the scale type of the measurements resulting from the procedure. (A detailed treatment of the theory of fundamental measurement from the standpoint of these two central problems is to be found in Krantz *et al*. the three-volume treatise *Foundations of Measurement*, 1971, 1989, 1990.)

3 Classification of scales of measurement

For the purpose of systematizing some of the discussion of the uniqueness problem in the previous section, we may define a scale as a class of measurement procedures having the same transformation properties. Examples of three different scales have already been mentioned, namely, counting as an absolute scale, measurement of mass as a ratio scale and measurement of temperature as an interval scale.

An ’ordinal’ scale is one which is arbitrary except for order. A number is assigned to give a ranking. The social sciences extensively use such scales. Numbers are also sometimes used for naming or classification. The numerical assignment is completely arbitrary. The numbers on football players’ shirts are examples of this sort of measurement. Such scales are usually called ’nominal’ scales.

In the literature a distinction is often made between those scales of measurement which are fundamental and those which are ’derived’. A derived scale is one which in the procedure of measurement presupposes and uses the numerical results of at least one other scale. In contrast, a ’fundamental’ scale does not depend on other scales. As might be expected, foundational investigations of measurement have been primarily concerned with fundamental scales, and such is the emphasis here (Krantz *et al.* 1971: ch. 10).

Another distinction common in the literature is that between extensive and intensive quantities or scales. For extensive quantities an empirical operation of combination can be given which has the structural properties of the numerical operation of addition. Intensive quantities are characterized by the absence of such an operation.

There is a fairly large literature on the problem of characterizing in what sense fundamental measurement of intensive quantities is possible. An early, significant example is the analysis of Nicole Oresme in the fourteenth

century.

4 Measurement and theory construction

It has become a platitude of scientific method that measurement of some attribute is really only interesting if some theory about that attribute is available. It is common to criticize many psychometric scales in psychology on the ground that application of the scales leads nowhere. Put another way, the criticism is that measurement for measurement's sake is not very fruitful scientifically.

There are two other aspects of the relation between measurement and theory construction that are important and yet are somewhat neglected. The first is that a theory of measurement for some characteristic of empirical phenomena is itself a genuine scientific theory. Moreover, certain theories are practically coextensive with the theory of measurement of the quantities basic to the theory. For example, the heart of the theory of rational behaviour in situations of limited information is the measurement of subjective probability and utility (see Decision and game theory §2).

The second aspect is that the chief stumbling block to applying certain theories is a problem of measurement. A classical example in physics is the general theory of relativity. Of the three 'traditional' pieces of evidence for the general theory, namely, the advancement of the perihelion of Mercury, the deflection of light waves by the gravitational field of the sun, and the red shift of stellar spectra, only the first does not present really difficult measurement problems (see Einstein, A. §3).

5 Ordinal theory of measurement

As a simple example of fundamental measurement, we may consider the ordinal theory of measurement. Models of this theory are customarily called weak orderings, defined as follows. Let A be a nonempty set and R a binary relation on this set. A pair (A, R) is a 'simple relation structure'. A simple relation structure (A, R) is a 'weak ordering' if and only if for every x, y and z in A (i) if xRy and yRx then xRz, and (ii) xRy or yRx.

The definition of isomorphism of simple relation structures illustrates nicely the general concept. A simple relation structure (A, R) is 'isomorphic' to a simple relation structure (A', R') if and only if there is a function f such that (i) the domain of f is A and the range of f is A', (ii) f is a one-one function, and (iii) if x and y are in A then xRy if and only if $f(x)R'f(y)$. To illustrate this definition of isomorphism let us consider the question: 'are any two finite weak orderings with the same number of elements isomorphic?' Intuitively it is clear that the answer is negative, because in one of the weak orderings all the objects can stand in the relation R to each other and not so in the other.

Homomorphism of models. In many cases a representation theorem in terms of isomorphism of models turns out to be less interesting than a representation theorem in terms of the weaker notion of homomorphism. Good examples are provided by theories of measurement. As already noted, when we consider general practices of measurement it is evident that in terms of the structural notion of isomorphism we would, roughly speaking, think of the isomorphism as being established between an empirical model of the theory of measurement and a numerical model. However, a slightly more detailed examination of the question indicates that difficulties about isomorphism quickly arise. In all too many cases of measurement, distinct empirical objects or phenomena are assigned the same number, and thus the one-one relationship required for isomorphism of models is destroyed. Fortunately, this weakening of the one-one requirement for isomorphism is the only respect in which we must change the general notion, in order to obtain an adequate account for theories of measurement of the relation between empirical and numerical models. The general notion of homomorphism is designed to accommodate exactly this situation. To obtain the formal definition of homomorphism for two simple relation structures as previously defined, we need only drop the requirement that the function establishing the isomorphism be one-one.

The concept of homomorphism is what we need to state a representation theorem for weak orders. First, a 'numerical' weak ordering is a weak ordering (A, $\leq$) where A is a set of numbers. The selection of the numerical relation $\leq$ to represent the relation R in a weak ordering is arbitrary, in the sense that the numerical relation $\geq$ could just as well have been chosen. However, choice of one of the two relations $\leq$ or $\geq$ is the only intuitively sound possibility. The following theorem provides a homomorphic representation theorem for finite weak orderings, and thus makes the theory of finite weak orderings a theory of measurement:

(1) Every finite weak ordering is homomorphic to a numerical weak ordering.

(1) was restricted to finite weak orderings for good reason: it is false if this restriction is removed. In order to state the desired theorem for the infinite case one preliminary notion is needed. Let (*A*, *R*) be a simple relation structure, and let *B* be a subset of *A*. Define the strict ordering relation *P* in terms of *R* by the equivalence: xPy if and only if xRy and not yRx. Then we say that *B* is *R-order-dense* in *A* if and only if for every *x* and *y* in *A* and not in *B* such that xPy there is a *z* in *B* such that xRz and zRy. Note that the denumerable set of rational numbers is order-dense in the non-denumerable set of all real numbers with respect to the natural numerical ordering $\leq$. This relationship between the denumerable rational numbers and all real numbers is just the one that is necessary and sufficient for the existence of a homomorphism between an infinite and a numerical weak ordering.

(2) Let (*A*, *R*) be an infinite weak ordering. Then a necessary and sufficient condition that it be homomorphic to a numerical weak ordering is that there is a denumerable subset *B* of *A* such that (i) *B* is *R*-order-dense in *A* and (ii) no two elements of *B* stand in the relation *E*, for example, for any distinct *x* and *y* in *B* either not xRy or not yRx.

The proof is related to the classical ordinal characterization of the continuum by Cantor (1895) (see Continuum hypothesis). The formulation and proof of (2) uses more mathematical apparatus than elementary logic. The restriction of the discussion of theoretical issues in the philosophy of science to the framework of first-order logic has too often meant a restriction of the extent to which the discussion could come into contact with other than highly simplified scientific theories especially constructed for the purpose at hand. An advantage of the set-theoretical formulation of theories such as theories of measurement is that standard mathematical methods are immediately available for both the formulation of the theory and the deductive analysis of the structure of its models (see Theories, scientific §3).

6 Invariance and meaningfulness

As already mentioned, the number assigned to measure mass is unique once a unit has been chosen. A more technical way of putting this is that the measurement of mass is unique up to a similarity transformation. A numerical function is a similarity transformation if there is a positive number αsuch that for every real number *x*, $\phi(x) = \alpha x$. In transforming from pounds to grams, for instance, the multiplicative factor α is 453.6. The measurement of temperature in $^\circ\mathrm{C}$ or F has different characteristics. Here an origin as well as a unit is arbitrarily chosen. Technically speaking, the measurement of temperature is unique up to a linear transformation. A numerical function is a 'linear' transformation if there are numbers α and β with $\alpha > 0$ such that for every number *x*, $\phi(x) = \alpha x + \beta$. In transforming from centigrade to Fahrenheit degrees of temperature, for instance, $\alpha = 9/5$ and $\beta = 32$.

Ordinal measurements are commonly said to be unique up to a monotone increasing transformation. A numerical function is a monotone increasing transformation if, for any two numbers *x* and *y*, if $x < y$, then $\phi(x) < \phi(y)$. Such transformations are also called 'order-preserving'.

An empirical hypothesis, or any statement in fact, which uses numerical quantities is empirically meaningful only if its truth value is invariant under the appropriate transformations of the numerical quantities involved. As an example, suppose a psychologist has an ordinal measure of IQ, and thinks that scores $S(a)$ on a certain new test *T* have ordinal significance in ranking the intellectual ability of people. Suppose further that the psychologist is able to obtain the ages $A(a)$ of the subjects. The question then is: should the following hypothesis be regarded as empirically meaningful? For any subjects *a* and *b*, if $S(a)/A(a) < S(b)/A(b)$, then $\mathrm{IQ}(a) < \mathrm{IQ}(b)$. From the standpoint of the invariance characterization of empirical meaning, the answer is negative. To see this, let $\mathrm{IQ}(a) \geq \mathrm{IQ}(b)$, let $A(a) = 7$, $A(b) = 12$, $S(a) = 3$, $S(b) = 7$. Make no transformations on the IQ data, and make no transformations on the age data. But let be a monotone-increasing transformation of the ordinal measurements $S(a)$ which carries 3 into 6 and 7 into 8. Then we have $3/7 < 7/12$, but $6/7 \geq 8/12$, and the truth value of the hypothesis is not invariant under .

The empirically significant thing about the transformation characteristic of a quantity is that it expresses in precise form how unique is the structural isomorphism between the empirical operations used to obtain a given measurement and the corresponding arithmetical operations or relations. If, for example, the empirical operation is simply that of ordering a set of objects according to some characteristic, then the corresponding arithmetical

relation is that of less than (or greater than), and any two functions which map the objects into numbers in a manner preserving the empirical ordering are adequate. Only those arithmetical operations and relations which are invariant under monotone transformations have empirical significance in this situation.

We can easily state the uniqueness or invariance theorem corresponding to the representation theorem (1) for finite weak orders:

(3) Let (*A*, *R*) be a finite weak order. Then any two numerical weak orderings to which it is homomorphic are related by a monotone-increasing transformation.

Put in other language, the numerical representation of finite weak orders is unique up to an ordinal transformation. Invariance up to ordinal transformations is not a very strong property of a measurement, and it is for this reason that the hypothesis considered above turned out not to be meaningful, because it was not invariant under monotone transformations of the measurement data.

A measurement representation theorem should ordinarily be accompanied by a matching invariance theorem stating the degree to which a representation of a structure is unique. In the mathematically simple and direct cases of measurement it is usually easy to identify the invariant group of transformations. For more complicated structures, for example structures that satisfy the axioms of a scientific theory, it may be necessary to introduce more complicated apparatus, such as the Galilean or Lorentz transformations of physics, but the objective is the same, namely, to characterize meaningful concepts in terms of invariance.

One point needs emphasis. When the concepts are given in terms of the representation, for example a numerical representation in the case of measurement, or representation in terms of Cartesian coordinates in the case of geometry, a test for invariance is needed. When purely qualitative relations are given which are defined in terms of the qualitative primitives of a theory, for example those of measurement or geometry, then it follows at once that the defined relations are invariant and therefore meaningful. On the other hand, the great importance of the representations and the reduction in computations and notation they achieve, as well as understanding of structure, make it imperative that we have a clear understanding of invariance and meaningfulness for representations which may be, in appearance, rather far removed from the qualitative structures that constitute models of the theory (Krantz *et al* 1990: ch. 22).

7 Variability, thresholds and errors

In the standard theory of fundamental measurement as discussed, qualitative axioms are formulated that lead to a numerical assignment unique up to some simple transformation. A widespread complaint about this classical foundation of measurement is that it takes too little account of the analysis of variability in the quantity measured. The sources of such variability can be of several different kinds. One important source is simply variability in the empirical properties of the object being measured. The height of a person for example varies on a diurnal basis and the variability of the weight of a person from day to day is a familiar fact of everyday experience. A quite different source of variability lies not in the object being measured but in the procedures of measurement being used. When the procedures being used lead to variability, these are ordinarily attributed to error. The study of error, especially in physical measurements goes back to the eighteenth century with fundamental work by Simpson, Lagrange, Laplace and, especially, Gauss. In the physical sciences the standard reporting of statistical errors in scientific publications began about the middle of the nineteenth century, especially in the publication of astronomical observations.

At least five kinds of error have been distinguished in astronomical observation, and they illustrate nicely the kind of analysis that can be made of errors in other empirical observations, even though changes in focus must be made when the empirical phenomena are quite different. For astronomers there are instrumental errors arising from imperfections in the production of the instruments of observation used such as telescopes and clocks. There are personal errors due to the response characteristics of the observer, for instance, coordinating a visual observation with the auditory beat of a clock, or with the nearly simultaneous visual observation of a clock, will lead to slightly different results for different observers. One example, Nevil Maskelyne, the fifth astronomer royal of Great Britain, discharged his assistant in 1796 because the assistant had observed the transits of stars and planets about a half-second later than Maskelyne himself. An objective of modern technology is to eliminate as much as possible

such personal errors, but they certainly persist in many areas of empirical observation. Errors of the third kind are 'systematic' and are due to conditions that are themselves subject to observation and measurement. For example, steady winds in a given direction must be taken account of in many kinds of familiar procedures, such as a navigator setting a course of a ship or aeroplane. Fourth, there are 'random' errors which can arise from variability in the conditions surrounding the observations and are not necessarily due either to variations in the object being measured or in the procedures of measurement. Meteorological variations affecting astronomical observations are a good example. Of a different sort, and a fifth kind, are empirical errors of 'computation' once numerical observations have been recorded. It is important to emphasize that there is nothing fixed and a priori about the classification given here. Other classifications can be found in the literature, but there is a natural conceptual basis to support each of the types of error listed (Taylor 1982).

It is not possible here to examine in detail how the foundational investigations of measurement procedures have been able to deal with such problems of error. One point of interest has been the analysis of thresholds in psychological judgments. It is a familiar fact that if a person is asked to judge the weight of two objects that are very close together, psychological discrimination of a difference will not occur when the difference is sufficiently small. There is a large body of work and a large body of theory connected with such psychological thresholds (Krantz *et al*. 1989: ch. 16).

A central objective of foundational analysis in this area is to begin with qualitative axioms, but to expect the representation for a measurement property of an object to be in terms of a random variable with a given probability distribution rather than in terms of a fixed numerical assignment. Such assignment of random variables rather than numbers to objects is already an abstraction from the various kinds of measurement error distinguished above. Part of the reason for this is that the main objective of such analysis is to end up with probability distributions for the errors in the form of some standard probability distributions familiar in mathematical statistics and amenable to detailed statistical treatment.

The theory of measurement has been important in quantum mechanics, the most significant physical theory of the twentieth century, for two distinct but related conceptual reasons. One arises from the Heisenberg uncertainty principle which shows that it is not possible to measure simultaneously with arbitrary precision the position and momentum of particles such as electrons or protons. Closely connected with this is the second aspect of measurement in quantum mechanics, namely the recognition that there is a physical interaction between the measurement instrument and the measurement object, a subject not really developed at all in classical physics (von Neumann 1996; see Quantum measurement problem).

In the social sciences and medicine a familiar kind of error is that which arises from finite sampling. If we wish to know something about the telephone habits, driving habits or reactions to a new medicine of a given population, we ordinarily do not plan to interview or test everyone in the population, but rather use a random sample to infer the distribution of habits or reactions in the entire population. The theory of such sampling is very much a twentieth-century subject (see Statistics and social science §2).

See also: Experiment; Observation; Operationalism; Theory and observation in social sciences

PATRICK SUPPES

References and further reading

Cantor, G. (1895) 'Beiträge zur Begründung der transfiniten Mengenlehre', *Mathematische Annalen* 46: 481-512. (See §5.)

Clagett, M. (ed.) (1968) *Nicole Oresme and the Medieval Geometry of Qualities and Motions*, Madison, WI: University of Wisconsin Press.(Contains Oresme's *Tractatus de configurationibus qualitatum et motuum*, with an introduction, translation and commentary.)

Kisch, B. (1965) *Scales and Weights: A Historical Outline*, New Haven, CT: Yale University Press.(A detailed survey of the more practical aspects, with special emphasis on the history of the instruments used for scales and weights.)

Krantz, D.H., Suppes, P., Luce, R.D. *et al*. (1971, 1989, 1990) *Foundations of Measurement*, vol. I, *Additive and Polynomial Representations*, vol. II, *Geometrical, Threshold, and Probabilistic Representations*, vol. III, *Representation, Axiomatization, and Invariance*, New York: Academic Press.(A comprehensive treatise in which the fundamental problems of measurement are given extensive treatment with many examples of

representation and uniqueness theorems, and extensive references to the literature in many disciplines on the theory of measurement.)

Neumann, J. von (1996) *Mathematical Foundations of Quantum Mechanics*, Princeton, NJ: Princeton University Press.(A classical treatise containing one of the most thorough and important discussions of the interaction between the measurement instrument and the measured object, with a particular reference to quantum mechanics but also applicable to measurement procedures generally.)

Taylor, J.R. (1982) *An Introduction to Error Analysis*, Mill Valley, CA: University Science Books.(An excellent elementary text on uncertainties in physical measurement.)

Mechanics, Aristotelian

The central feature of Aristotle's mechanics is his discussion of local motion, a change of place, which he categorizes as either natural or violent. He further divides natural motion into celestial motion, which is uniform, circular and eternal, and terrestrial motion, which is rectilinear (straight up or down), and finite in both time and distance. All motions which are not natural are classified as violent.

For all motion Aristotle required a force in direct contact with the object being moved. We may represent the Aristotelian law of motion by the modern formula: Velocity = Force (motive power)/Resistance, or $V = kF/R$*. In applying this law of motion to falling bodies, Aristotle associated the weight of the body with the force, and the resistance of the air (or other medium) with the resistance. Thus, Aristotle believed that heavy bodies fall faster than light ones.*

The problem of what force is actually in contact with the body, and causes it to fall, posed a serious difficulty for Aristotle. Aristotle concluded that elements were created with a tendency to move to their natural place, barring any hindrance or interference. Projectile motion posed a similar problem for Aristotle. In the case of a thrown object, the force was provided by the hand of the thrower as long as the object was in contact with the hand. But one needed an explanation of why the object continued to move once it had left the thrower's hand. Aristotle concluded that the medium through which the projectile moved provided the force that kept it moving.

Aristotle also regarded both the existence of a void or any motion in it as impossible. A void contains nothing that could sustain the motion of a projectile once it left the projector. In addition, because a void can provide no resistance, the speed of an object in a void would be infinite.

1 Local motion

This entry concentrates on Aristotle's discussion of motion and its causes as found in his work *Physics*. Aristotle's concept of motion includes not only local motion, a change of place, but any change from potential to actual being. This may include changes in shape, size, or properties of a body such as temperature or colour. The central feature of his mechanics is, however, his discussion of local motion. This includes the Aristotelian law of motion and the questions of falling bodies (and the related problem of the motion of the elements), projectile motion and the existence of a void.

Aristotle categorized local motion as either natural or violent. He further divided natural motion into celestial motion, which is uniform, circular and eternal, and terrestrial motion, which is rectilinear (straight up or down) and finite in both time and distance. The terrestrial motions are governed by the nature of the four elements which make up all substances: earth, air, fire and water, and by the idea of natural place. Earth, which is absolutely heavy, falls toward the centre of the universe (coincidently the centre of the earth), which is its natural place. Fire, which is absolutely light, rises to its natural place, a spherical shell inside the lunar sphere. Thus, the equilibrium distribution of an Aristotelian universe would be a spherical earth surrounded by concentric spherical shells of water, air and fire, surrounded by the heavens. Objects composed of a mixture of elements behave in the manner of their predominant element. All motions which are not natural are classified as violent; for example, projectile motion.

For both natural and violent motions Aristotle required a force in direct contact with the object being moved. This view that a force is necessary for motion to occur at all (everything that is moved is moved by something else) is quite different from the modern view (see Mechanics, classical). For living creatures the motive power required is provided by the soul. Celestial objects and planets are moved by celestial intelligences or spirits. Inanimate objects require a force in contact with the object itself and this posed a difficult problem in dealing with falling bodies, projectiles and with the motion of the elements themselves.

We may represent the Aristotelian law of motion by the modern formula:

$$\text{Velocity} = \text{Force (motive power)}/\text{Resistance} \quad \text{or} \quad V = kF/R.$$

Aristotle never stated his law of motion in this concise form, but rather discussed how either the distance travelled

or the time of travel will change when either the force exerted on the object or the weight of the object changes. He discussed motion in this way because the concept of speed as the ratio of distance to time was not considered a proper ratio. One could have a proper ratio only of similar things such as the ratio of one distance to another distance, or of one time to another time.

If then: the mover *A*, has moved *B*, a distance *C*, in time *D*, then in the same time force *A* will move $\frac{1}{2}B$ twice the distance *C*, and in $\frac{1}{2}D$ it will move $\frac{1}{2}B$ the whole distance *C*: thus the rules of proportion will be observed. Again if a given force move a given weight a certain distance in a certain time and half the distance in half the time, half the motive power will move half the weight the same distance in the same time. Let *E* represent half the motive power *A* and *Z* half the weight *B*: then the ratio between the motive power and the weight in the one case is similar and proportionate to the ratio in the other, so that each force will cause the same distance to be travelled in the same time (*Physics* VII 5).

It should also be emphasized that Aristotelian motion required the action of *both* the force and the resistance, discussed later.

Aristotle recognized that his law of motion did not have universal applicability. He knew that the application of a force might not result in any motion at all, even though his law of motion implies that it would, as would be the case with someone attempting to lift a very large weight. To solve the problem he added a subsidiary condition to the law that required the force to be greater than the resistance.

If, then, (a force) *A* move *B* a distance *C* in time *D*, it does not follow that *E*, being half of *A*, will in time *D*, or in any fraction of it, cause *B* to traverse a part of *C*, the ratio between which and the whole of *C* is proportionate to that between *A* and *E* (whatever fraction of *A* that *E* may be), in fact it might well be that it will produce no motion at all; otherwise one man might move a ship, since both the motive power of the ship-haulers and the distance that they all cause the ship to traverse are divisible into as many parts as there are men (*Physics* VII 5).

In applying this law of motion to falling bodies Aristotle associated the weight of the body with the force, and the resistance of the air (or other medium) with the resistance. Thus, he believed that heavy bodies fall faster than light ones. ’We see that bodies which have a greater impulse either of weight or of lightness, if they are alike in other respects, move faster over an equal space, and in the ratio which their magnitudes bear to each other’ (*Physics* IV 8).

2 The causes of motion

The problem of what force is actually in contact with the body and causes it to fall posed a serious difficulty for Aristotle. This is closely connected to the question of the natural motion of elements. What force causes earth to fall to the centre of the universe and what force causes fire to rise? Aristotle’s discussion of this question is quite complex and involves the change from potentiality to actuality, in this case from potential to actual motion (Lang 1992). Natural place is the actuality that moves the elements. Elements were created with a tendency to move to their natural place, barring any hindrance or interference.

The idea that natural place causes the motion is not an attraction to natural place (that is, up or down) or what we might call the force of gravity, even though Aristotle associated weight with the force that causes a body to fall, because either of these involves action at a distance and not a force in direct contact with the body. Whatever creates the body or element gives it an internal tendency for motion. Aristotle also knew that a falling body did not travel at constant speed, but that it did, in fact, accelerate. He did not offer any explanation of this acceleration, but the problem was considered by several of his later commentators.

Projectile motion posed a similar problem for Aristotle. In the case of a thrown object, the force was provided by the hand of the thrower as long as the object was in contact with the hand. But one needed an explanation of why the object continued to move once it had left the thrower’s hand. Aristotle concluded that the medium through which the projectile moved provided the force that kept it moving. This occurred either by *antiperistasis* (replacement), in which the medium rushes around the body to prevent the formation of a void or vacuum and pushed the body from behind, or by the medium itself having acquired the power to be a mover from the original projector. (The theory of *antiperistasis* is more probably due to Plato and his followers than to Aristotle.)

The medium itself does not have to move, but rather possesses the power to move something else. This power is, however, imperfectly transmitted from one layer of the medium to the next and gradually dies away. Thus, as indicated above, the motion of the projectile seems to be complex, consisting of both its natural motion and the motion imposed on it. It is unclear what Aristotle believed the path of a projectile was. For example, in the case of a projectile thrown horizontally, did it first travel only horizontally, and then fall to its natural place, or did it exhibit both horizontal and vertical motions simultaneously, throughout its motion?

The resistance needed for motion is provided by the resistance, or friction, of the medium, but also seems to include the mass or weight of the projectile. The resistance of the medium is necessary to prevent the motion of the projectile from being infinite in both speed and extent, an impossibility for Aristotle, as discussed below. Aristotle’s use of the medium to both sustain and resist motion seems paradoxical, and was commented on by several medieval critics.

3 The void

It is clear that Aristotle regarded both the existence of a void or any motion in it as impossible. First, a void contains nothing that could sustain the motion of a projectile once it left the projector. Second, because a void can provide no resistance, the speed of an object in a void would be infinite, or as Aristotle states, ’it moves through the void with a speed beyond any ratio’. He further argues that although heavy bodies fall faster than do light ones in a medium, they would move at the same speed in a void, which he regarded as impossible.

In proposing a clear statement of the inertial principle, Aristotle also gave a proof of it by relating the persistence of motion, or inertial motion, to the homogeneity of space. However he turned the argument around and regards it rather as a proof of the impossibility of motion in a void. The inertial consequences of such motion would be (to Aristotle) the unacceptable result of motion that is infinite in extent.

Aristotle’s physics forms the basis of subsequent medieval discussions of these questions. It is fair to say that medieval studies of motion consist of commentaries and criticisms of Aristotle’s work.

See also: Aristotle; Aristotelianism in Islamic philosophy; Aristotelianism, medieval; Aristotelianism, Renaissance

ALLAN FRANKLIN

References and further reading

Aristotle (*c.* mid-4th century BC) *Physics* (Physica); trans. R.P. Hardie and R.K. Gaye, Oxford: Clarendon Press, 1930.(Main source text.)

Clagett, M. (1955) *Greek Science in Antiquity*, New York: Macmillan.(A useful account of Greek science.)

Cohen, M.R. and Drabkin, I.E. (1958) *A Source Book in Greek Science*, Cambridge, MA: Harvard University Press.(A good survey of Greek science, including quotations - in translation - from the original works.)

Lang, H.S. (1992) *Aristotle’s Physics and its Medieval Varieties*, Albany, NY: State University of New York Press.(A detailed discussion of Aristotle’s physics as well as discussion from medieval commentators. Chapter 3 covers the change from potential to actuality mentioned in §2.)

Mechanics, classical

Understood at its most general, 'classical mechanics' covers the approach to physical phenomena that dominated science from roughly the time of Galileo until the early decades of the twentieth century. The approach is usually characterized by the assumption that bodies carry an inherent mass and well-defined positions and velocities. Bodies subsist within a three-dimensional absolute space and influence one another through reciprocal forces. These objects obey the three laws of motion articulated by Isaac Newton in 1686 in a deterministic *manner: once a mechanical system is assembled, its future behaviour is rigidly fixed. Such 'classical' assumptions were eventually rejected by Einstein's theory of relativity, where the assumption of a three-dimensional Euclidean space is abandoned, and by quantum mechanics, where determinism and well-defined positions and velocity are eschewed.*

Classical mechanics is frequently characterized as 'billiard ball mechanics' or 'the theory of mechanism' on the grounds that the science treats its materials in the manner of colliding particles, or clockwork. Such stereotypes should be approached with caution because the basic framework of classical mechanics has long been subject to divergent interpretations that unpack the content of Newton's 'three laws' in remarkably different ways. These differing interpretations provide incompatible catalogues of the basic objects that are supposed to comprise the 'classical world' - are they point masses, rigid bodies or flexible substances? Or, as many writers have suggested, should mechanics not be regarded as 'about' the world at all, but merely as a source of useful but fictitious idealizations of reality?

These foundational disagreements explain why classical mechanics has often found itself entangled in metaphysics. Much of modern philosophy of science is characterized by attitudes that were originally articulated during the nineteenth century's attempts to clarify the grounds of classical mechanics.

1 Newton's laws of motion, first law

Introductory textbooks usually sketch a simple, standardized approach to the content of classical mechanics, which, for reasons discussed in §4, may be called the *point mass interpretation*. Although this interpretation is often presented as 'the correct understanding of classical mechanics', few philosophers or physicists of the classical era would have accepted this reading as adequate. Its reading not only suffers empirical difficulties; it does not describe a universe that would have been pleasing to most practitioners. The point mass reading has gained its textbook centrality largely because alternative foundations for mechanics prove resistant to easy articulation. Indeed, satisfactory alternatives were not fully developed until well into the twentieth century (largely by the engineering community, which requires a trustworthy classical basis for its endeavours). The mismatch between the conceptual simplicity of textbook orthodoxy and the expected characteristics of the subject is apt to cause students of philosophy much difficulty.

The essential content of classical doctrine is commonly held to reside in Isaac Newton's celebrated laws of motion. Unfortunately, these principles, in the form they were articulated by Newton, are subject to a surprising range of interpretations.

Electromagnetism is largely left out of account here. This is partly because 'classical' electromagnetism, as the subject is understood today, fits more neatly into a *relativistic* than a Newtonian framework. In the nineteenth century, light and other electromagnetic phenomena were regarded as vibrations within an 'ether' that was conceived as a material continuum in the manner of §6 (see Field theory, classical).

In Thompson and Tait's (1879) translation of Newton (1726), Newton's laws are: first law: every body continues in its state of rest or of uniform motion in a straight line, except insofar as it may be compelled by force to change that state. Second law: change of motion is proportional to force applied, and takes place in the direction of the straight line in which the force acts. Third law: to every action there is always an equal and contrary reaction: or, the mutual actions of any two bodies are always equal and oppositely directed.

Newton borrowed the first law, also known as 'the principle of inertia', from Galileo and Descartes (although neither would have accepted the doctrine in its fullest generality) (see Galilei, G.). According to it, the *natural motion* of a particle uninfluenced by external forces is always a straight line, whereas earlier scientists had often presumed such natural motions to be circular or directed towards a 'natural home' in the universe. 'Inertia' is the

traditional term for the tendency of a body to retain its present state of motion; it is harder to stop a cannon ball than a baseball even if they are travelling at the same velocity. From a Newtonian point of view, ‘inertia’ is simply another name for ‘mass’, but it is common to speak of ‘inertial *forces*’ as well, a notion discussed in §3.

Henri Poincaré (1902) claimed that the first law (as well as much of the rest of the Newtonian framework) does not report genuine physical content, but instead represents a mere ‘convention’ (see Poincaré, J.H. §3). Newton presumed that a Euclidean space exists independently of matter and understood his first law as delineating how free matter will move upon this geometrical stage. Poincaré objected that we can have no knowledge of spatial structure independent of the behaviour of material objects found within it. The first law is not a straightforward ‘fact’, but merely articulates a *criterion* for erecting a geometry around material objects. No empirical facts can force a rejection of Newton’s first law ‘criterion’ for the presence of ‘straight lines’, although, were the empirical facts otherwise, an alternative ‘convention’ might be favoured where the notions of geometry become coordinated to the movements of matter in a different fashion. This *conventionalism* has proved enormously influential within twentieth-century analytic philosophy and whether or not Poincaré’s implied distinction between a ‘convention’ and a ‘genuine fact of nature’ can be sustained remains a controversial issue (see Conventionalism).

Newton has been widely criticized for assuming the existence of an enduring ‘absolute space’ in the first law, because (i) the notion sounds unduly ‘metaphysical’, (ii) its presence seems inherently unobservable, or (iii) the first law (on its strongest reading) forces a principle of *Galilean relativity*: no law of nature allows a determination of whether a set of objects lies at rest in absolute space or merely moves through it at a constant velocity (reference objects that move in either of these ways define *inertial frames*). One is therefore denied any means of relocating a point in absolute space after an interval of time passes (see Space §§2-3).

During the nineteenth century, followers of Kant saw Galilean relativity as evidence that absolute space does not represent a real condition found in nature, but merely reflects an organizational structure imposed upon reality by the human mind (Poincaré’s conventionalism shares many affinities with such Neo-Kantian thinking) (see Kant, I. §5). But more sophisticated techniques articulated in the twentieth century (chiefly by Cartan) have suggested another resolution to the relativity problem: classical mechanics should follow Einstein’s lead in rejecting the notion of an enduring *space* in favour of a blend of space and time (see Spacetime). In Cartan’s reformulation, the first law is interpreted as demanding a certain geometrical structure within the melded spacetime, avoiding the unobservable aspects of Newton’s persisting space. Although this approach weakens Newton’s geometrical assumptions somewhat, it agrees that spacetime possesses objective features independently of human decision or the placement of material objects within the spacetime.

Cartan also showed that gravitation could be ‘geometrized away’ within certain classical frameworks, in imitation of the techniques of general relativity. It should be mentioned, however, that Galilean relativity was not universally accepted during the heyday of classical mechanics - various electrical effects once seemed as if they might provide criteria for detecting inertial movement within absolute space.

2 Newton’s laws of motion, second law

Newton’s second law is popularly known as ‘ $F = ma$’, where ‘*F*’ stands for ‘force’, ‘*m*’ for ‘mass’ and ‘*a*’ for ‘acceleration’. The claim is that a particle will deviate from its inertial motion only insofar as a force is imposed from without, in which case the particle will accelerate to a degree conditioned by its mass. In a modern treatment (if uncomplicated by ‘constraints’), ‘ $F = ma$’ provides the skeleton upon which the basic *equations of motion* for a collection of particles are constructed. The recipe is this:

(a) Delineate the particles whose interactions one wishes to study.
(b) Determine what kinds of specific force act between these (do they gravitationally attract one another?, do they display electrical attraction or repulsion?, and so on).
(c) For each particle *a*, decompose each of the forces that act upon *a* along a selected Cartesian axis - for example ‘ $f_x^{a,e}$ ’ might express the amount of electrical force that acts upon particle *a* along the *x*-axis.
(d) For each particle *a* and each Cartesian direction *x*, assemble all of the specific forces under (c) into a sum called the ‘total force acting upon the particle acting in direction *x*’.
(e) Set this sum of forces equal to $m\mathrm{d}^2x/\mathrm{d}t^2$ (that is, to what in the calculus is called ‘the second derivative of *a*’s location along the *x*-axis’). The upshot is a system of $3n$ second-order ordinary differential equations for a

collection of *n* particles.

The list under (b) of specific force laws active within the system of particles are called the *constitutive principles* for the system at hand; crudely speaking, they delineate the basic manner in which forces bind the collection together. The equations of motion constructed by this recipe then fix (*modulo* occasional technicalities) how the system will behave at all future times once the positions and velocities of all particles are known at some specific time (such data represent the system's 'phase' or *initial condition*). That is, the equations of motion in classical physics *determine* the future states of the system as it evolves from its initial condition (usually past states are determined as well). The determinism of classical physics is a famous bugbear of ethical philosophy. Quantum mechanics no longer accepts determinism of the classical variety, but it is doubtful whether what it substitutes is preferable from an ethical point of view (see Determinism and indeterminism; Free will).

'Determinism' is not the same as 'predictability' in any reasonable sense. It has been realized for some time that most classical systems are so complicated that no computer could possibly track their behaviour accurately for more than a few seconds. Much of the recent fascination with so-called 'chaos' has evolved from the discovery of new ways to extract useful partial information from these otherwise unpredictable systems (see Chaos theory).

The second law recipe has been described in some detail because Newton's less explicit formulation has proved a ripe source of philosophical confusion. Our 'recipe' approach becomes clearly established only within the investigations of Leonhard Euler in the eighteenth century. Expressing the principle as simply ' $F = ma$' encourages the impression that the second law is merely a definition of 'force'. This objection overlooks the non-arbitrary decomposition into specific forces required in the recipe. Some writers have complained that Newton's laws are 'empty' because they describe systems either subject to one force (the second law) or no forces at all (the first law). But no real life system is like this. In the author's evaluation, this objection, and many others like it, fails to understand the Euler recipe correctly.

But if Newton's second law is understood in this Eulerian vein, physics should be greatly concerned to specify the complete set of specific force laws used in step (b). Newton's law of universal gravitation supplies an admirable example of the kind of law needed, but, beyond that paradigm, classical physicists were somewhat lackadaisical in their efforts to fill out the missing laws (they never agreed, for example, on the nature of the forces that glue matter together). Joseph Sneed (1979) none the less sees the demand for these as yet unarticulated force laws as a vital part of the *content* of classical physics - he hopes that Thomas Kuhn's celebrated notion of a 'scientific paradigm' can be clarified in this manner. In truth, the constraint tradition, described in §5, often herded the actual development of mechanics in directions other than the hypothesized search for specific force laws. As we shall see, the criticism of ' $F = ma$' as an empty definition gains a greater cogency under the point of view of this school.

In general, many of the standard objections to Newton's laws originally grew out of substantial disputes over the fundamental trustworthiness of key mechanical propositions. As these origins were forgotten, the disputes became reinterpreted (particularly by the logical positivist school) as rather trivial squabbles over whether forces are 'observable' or not.

3 Newton's laws of motion, third law

The haziness of Newton's third law - 'action = reaction', as it is popularly known - is palpable in its very formulation. The textbooks often add three additional tenets to the claim: (a) All forces arise between pairs of particles. (b) Such forces are directed along the line between them (the forces are 'central'). (c) The strength of these forces depends only upon the spatial separation between the bodies and not, say, upon their relative velocities. One can then demand that if *a* exerts a specific force *f* upon *b*, then *b* will exert a reciprocating force upon *a* equal in magnitude to *f*, but reversed in direction. Although Newton's own law of gravitation suits these requirements, it is doubtful that he would have accepted the (a)-(c) supplements. Requirement (c) stands in apparent conflict with most frictional forces because their strength depends upon the rate that bodies slip past one another. Requirement (b) seems incompatible with *shearing forces*, which arise when, for example, one layer of water slips over another. Each layer exerts a retarding force upon the other, but these forces point along the layers' interface, not 'the line between them'.

The main motivation for including requirements (a)-(c) is that, on this approach, they are needed to prove

fundamental tenets such as the conservation of energy. For example, (c) justifies the notion of ‘potential energy’. Newton was unconcerned with these requirements because he had no notion that energy conservation holds. Indeed, he often understood the third law in a variant ‘inertial force’ manner.

Consider the sun and its planets. On the (a)-(c) interpretation, ‘action = reaction’ requires that each planet should exert a gravitational force upon the sun equal in strength to the solar gravitational pull that it feels. Since the sun is more massive than the planets, these planetary attractions will not accelerate it greatly. It does not lead to great mathematical error if the sun is treated as if it is the source of an *unreciprocated* force - that is, the solar attractions are described as dependent only upon the planet’s position within an inertial frame. On the (a)-(c) reading, this treatment is only a useful fiction - the reciprocating pulls of the planets must dislodge the sun slightly from its assumed inertial motion. Disregarding the sun’s minor wobblings allows the problem of planetary movements to be *effaced* from the complications of solar motion. When third law requirements can be overlooked in this manner, the force is called *external*; otherwise, they are dubbed *internal*.

However, a long tradition exists of regarding some external forces as nonapproximate in origin. How does a mechanical system balance such a force? Answer: it develops an ‘inertial force of resistance’ - its parts accelerate in such a manner that the sum over all particles of mass *times* acceleration exactly matches the external force. Most modern texts dismiss this approach to the third law as a mistake: no force should be regarded as truly external and inertial reactions do not qualify as forces at all. Indeed, the ‘inertial force’ reading seems merely to repeat the content of the second law over again.

The ‘inertial force’ understanding of the third law becomes more natural within the constraint-based approach of §5, where central forces drop from view. In the mechanics of continua (§6), the ‘principle of material frame indifference’ reveals some further affinities between inertial reactions and forces.

4 Point masses and idealization

Our discussion has involved much hazy talk of ‘particles’; what kind of object is intended by this term? On the interpretation presented so far, a ‘particle’ needs to be a *point mass*: a movable point carrying a fixed mass (and possibly other quantities such as charge). Forces have been required to ‘act along the line between particles’; this makes obvious sense only if ‘particles’ lie at uniquely defined points. Point masses must always remain separate, for the possibility of fusing would permit an infinite source of potential energy. On a point mass interpretation, all forces are of *action-at-a-distance* type, for they originate in one particle, yet act, across an intervening vacuum, somewhere else. *Contact forces*, where one body exerts a pressure immediately upon another through geometrical contact, are not tolerated on this approach to Newton’s laws.

The point mass approach, standard in most modern textbooks, was first proposed by R.J. Boscovich in 1763. Boscovich’s suggestion was rarely embraced during the heyday of classical mechanics. Most practitioners felt that their subject should deal directly with some variety of extended body, and not regard them merely as swarms of points. Newton, although often interested in arguing that objects in apparent contact were actually separated (see his discussion of ‘Newton’s rings’), also believed that atoms possessed hard cores of a certain size.

The point mass presentation of classical physics has gained great favour in modern textbooks partially because of a natural association between its mathematical structures and those of elementary quantum mechanics. ‘Correspondence rules’ exist to convert point mass equations to the partial differential equations of quantum theory. Under this transition, the quantum particles gain a surrogate for ‘size’ due to considerations peculiar to the newer theory, for instance, the exclusion principle. Many authors have been misled by such mathematical relationships into presuming that remarkable successes of classical physics in treating the large-scale phenomena of ordinary life can easily be explained by the way in which point mass theory ‘reduced’ to quantum theory. This impression is misleading, because the point mass interpretation at best provides an awkward basis for setting up the successful models of fluids and elastic bodies that are more adequately established following the programme of §6.

Sometimes it is clearly useful to treat extended bodies *as if* they were point particles. The *locus classicus* of this technique is Newton’s observation that a rigid, perfectly spherical planet behaves gravitationally from its exterior as if it were a single point carrying the same mass. The orbit of a planet can then be calculated without worrying about its interior. Real planets, of course, are neither perfect globes nor rigid and the predictions of point mass

celestial mechanics are slightly inaccurate as a result. None the less, this approximate effacement of the gross behaviour of the planets from their internal complications has inspired a popular ‘as if’ reading of Newton’s laws: they delineate a range of *idealized models* that suit real life situations to a tolerable degree. When greater accuracy is desired, one shifts to modelling the planet as a large swarm of point masses. Each finer scale of idealized modelling provides predictions of increased accuracy, but we should never pretend that our constructions supply anything other than simplified portraits of reality (see Idealizations; Models §3).

On this approach, it becomes difficult to see how Newton’s laws could be shown *false*; at best, they might prove an *inconvenient* series of approximations to utilize. An important distinction should be drawn here: in *any* scientific theory, the successful treatment of everyday phenomena will require artificial simplification because the mathematics pertinent to a fully realistic approach is horrendously complex. But the ‘idealization’ thesis maintains that, quite apart from mathematical intractability, science must study simplified models because of an inherent lack of reality within the very fabric of physics. However, the ‘lack of reality’ in point masses may merely demonstrate that they do not provide the best foundation for a satisfactory classical mechanics.

5 Constraints and rigid bodies

A universe of point masses seems a far cry from standard stereotypes of classical mechanics: is not the subject supposed to treat the universe as a ‘great clock’? Are not clocks composed of gears, rods and springs - extended bodies all? There is a variant tradition, stretching back to antiquity, that exploits the properties of *rigid bodies* in a direct geometrical way (gears and rods, but not springs, are rigid bodies). Typically, such bodies interact with another through contact along their boundaries: two gears intermesh; a bead slides along a wire; a wheel rolls along a plane. These linkages are specified by so-called *constraints*: equations that geometrically link the movements of distinct bodies. The presence of constraints clearly modifies Newtonian behaviour in various ways. For example, the bead will, if no external forces act against it, slide along the wire at a constant velocity forever (if the wire itself is not accelerated). Constraints alter Newton’s law of inertia from ‘travels with constant rectilinear velocity’ to ‘travels at an unchanging speed along the curvilinear constraint of the wire’. Principles of unhampered motion that are adapted to accommodate constraints are called *generalized laws of inertia*.

The existence of constraints, strictly speaking, is incompatible with the point mass reading. Let the bead slide past point *p* upon the wire. What *force* will the wire near *p* exert upon the bead as it slides past? ‘Easy’, the stock reply goes, ‘compute the acceleration *a* needed to move the bead past *p*. Divide *a* by the mass of the bead to obtain the force exerted by the wire.’ But let the bead slide past *p* with a greater velocity. The wire must generate a greater force to hold the bead in its rightful place at *p*. In short, the wire must be able to ‘see’ the bead’s *velocity* before the force needed to maintain the constraint can be exerted. But the point mass understanding of the third law requires that forces be velocity independent; otherwise, conservation of energy might fail.

What should be done? Either constraints are regarded as convenient approximations to point mass reality or our fundamental principles must be modified to tolerate them. In most texts, constraints are introduced in so cavalier a fashion that the policy followed cannot be determined. On the first strategy, ‘rigid bodies’ are simply collections of particles welded together by very strong (but finite) forces. The masses within our wire must shift slightly when a faster bead passes, although these movements may be quite microscopic.

In 1788, however, J.L. Lagrange articulated a version of mechanics that incorporates genuine constraints and rigid bodies within its foundations. He based his approach upon the law of virtual work (sometimes: virtual velocity) and d’Alembert’s principle. The idea behind virtual work is illustrated nicely by the behaviour of a lever. Work can be done if weights are allowed to drop - falling water drives old-fashioned grist mills. String a series of masses along a lever arm; the rigidity of the arm links these weights together by a constraint. How must these weights be arranged if the lever is to stay in *equilibrium* (= at rest)? Answer: any work that might be gained if the weights on one side of the arm are allowed to *fall* must be expended to *raise* the weights strung along the other side. ‘Virtual work’ declares that equilibrium within any constrained set of mechanical parts will obtain just in case their hypothetical abilities to perform work cancel out in the manner described. Laws of this variety (which predate Lagrange) are called *variational principles* because a system’s actual behaviour (equilibrium, in this case) is adduced from a comparison of hypothetical behaviour (that is, imaginary movements given to the weights). For systems not at rest, Lagrange appealed to d’Alembert’s principle. This doctrine claims that if unbalanced external forces are applied to a mechanical system, it will display inertial reactions that exactly balance the external set.

Recall (from §3) that an 'inertial reaction' is simply the acceleration of a body multiplied by its mass. D'Alembert's principle thus maintains that the system 'reacts' to external 'activity' by moving against the activity. We noted that Newton's third law is sometimes interpreted in this manner.

Classical mechanics is often presented using other forms of variational technique, for example, Hamilton's principle and other varieties of 'least action'. Although elementary textbooks often allege that such variants are equivalent to one another and to Newton's original set of laws, this is not true unless a number of background assumptions are made. Hamilton's treatment, especially, renders salient some important features of energy conserving systems, considered with respect to the so-called 'Hamiltonian flow within phase space'. But these matters must be left to standard texts.

Notice that, in dealing with the force upon the bead, we calculated in a direction reverse to what might seem the expected one. In celestial calculations, we compute the acceleration of the earth from a prior knowledge of the *forces* that apply to it. Newton's law of gravitation is the specific force law employed here. But in the bead case, we did not work with any specific force law; instead, the force was calculated merely from the *acceleration* that the wire requires of the bead. Forces induced by the satisfaction of constraints are traditionally called 'forces of reaction'.

Lagrange's approach tolerates both 'forces of reaction' and law specified forces, where the latter are of action-at-a-distance type. It had long been the hope of mechanics, in its 'clockwork' inspired moods, that action-at-a-distance effects might turn out to be the result of contact movements within a hidden medium linking the distant bodies. Descartes and Leibniz believed that the attraction of planets resulted from swimming in a common 'sea' that pushed them together. Appeal to action-at-a-distance forces was considered a retreat to the 'occult' qualities of medieval times. In the nineteenth century, J.C. Maxwell suggested that electromagnetic effects might be communicated between distant bodies through a complicated set of intervening mechanical parts (see Electrodynamics). Such proposals led H. Hertz to reformulate mechanics based upon constraints alone (a wider range of these is needed than appear in standard Lagrangian practice). Hertz (1894) replaced Newton's laws by a simple principle of 'generalized inertia' and identified all forms of energy with actual energy of movement (kinetic energy). Most importantly, the notion of force is essentially discarded in Hertz's system, since only the derivative 'forces of reaction' are tolerated in his framework. Hertz's system provides the best realization of the philosophical dream of a truly machine-inspired mechanics. Interest in Hertz's project faded with the rise of quantum theory, but its distrust of the Newtonian notion of 'force' has proved longer lived. In the author's opinion, 'force' should occasion no unhappiness as long as mechanics remains faithful to a point mass interpretation; it is the frequent appeal to *post hoc* 'forces of reaction' that generates the uneasy sense that 'force' functions as an unnecessary ingredient within mechanics.

In any case, the notion that the universe could actually move as one great *machine* is less plausible than one might expect. Machines are designed (insofar as possible) so that their parts slip over each other without binding. If a collection of rigid parts is randomly assembled, without the guidance of a designer, they are likely to end in *overconstraint*: the expected conditions of sliding, rolling, and so on, among its parts cannot be satisfied all at once. To accommodate the overconstraint, some of the parts will need to distort or fracture, thus losing their postulated rigidity.

In other words, once the basic 'particles' of mechanics are granted any spatial extension at all, they are likely to lose their postulated rigidity at some time or another. How does this eventuality affect Lagrange's programme for the foundations of mechanics? A retreat to an 'idealization' defence is possible: mechanics tries to idealize real physical systems in terms of constrained rigid bodies, despite their lack of permanent reality. Some advantage over point mass idealization is gained, for a better set of modelling tools has been assembled. It would be preferable however, if an approach to mechanics could be found that accepts bodies which are capable of distortion at all levels of analysis. The materials in this reformulated mechanics should be truly flexible entities, such as springs and fluids, which can smoothly distort or flow when the right external pressures are exerted. Such objects are called 'material fields' or *classical continua*.

6 Continua

If one tries to describe continua directly, one quickly discovers that formulating plausible laws for genuinely

flexible bodies introduces many conceptual problems. All of the problems of the 'infinitely divided' that have plagued philosophy and mathematics since the time of Zeno's paradoxes arise (see Zeno of Elea). Mechanics supplements these with some peculiar difficulties of its own.

For example, it seems natural to interpret the term 'particle' in Newton's laws as an 'infinitesimal portion of a continuous body'. If such a 'particle' belongs to a macroscopic body that stretches, must not that 'particle' enlarge infinitesimally under the stretching as well? But how can a point *p* sensibly expand to become a 'bigger point'? And what should ' $F = ma$' require of *p*? If *p* enlarges under an applied force, should its 'mass' (or, more properly, its density) also alter? Traditional arguments concerning 'infinitesimals' were plagued by vagueness of this sort, as critics such as Bishop Berkeley swiftly observed (see Berkeley, G. §§8, 11). Leibniz's celebrated philosophy of 'monads' represented, in part, a sophisticated attempt to surmount some of these problems by moving outside of the arena of space and time altogether (see Leibniz, G.W. §4).

Although infinitesimal readings of Newton's second law are common in textbooks, the considered modern opinion holds that this is a mistake: ' $F = ma$' should apply only to the *extended* parts of a continuous body. This shift in perspective introduces many unexpected conceptual subtleties. Suppose that A is a continuous body with external forces applied at various spots along its outer edge and within its interior. If A is a perfectly *rigid* substance, all of these forces can be added together as if they act at a common location. In a rigid body, a force applied at *p* can be shifted to any other position *q* and still pull A equally to the right. Applying a force at *p* rather than *q* affects only how A will *rotate*, not how it *translates*. Indeed, by defining the so-called 'moment' of A's array of forces (roughly, a measure of their 'off-centredness'), Euler showed how A's rotation could be calculated (the principle employed is called 'balance of angular momentum' or 'moment of momentum'). Euler's two principles - non-infinitesimal ' $F = ma$' and 'moment of momentum' - completely settle how any rigid body will respond to applied forces.

But the success of these principles appears to trade upon A's postulated rigidity. If A is composed of flexible stuff, then how it reacts to a schedule of applied forces will clearly reflect the materials of which it is made - the same set of forces will move rubber and water in completely different fashions. To make further progress, we must consider the *contact forces* that arise upon the surfaces inside A that separate points *p* and *q*. Draw an imaginary boundary shell B around *q* that excludes *p*. An external force applied at *p* can transmit an indirect influence to *q* if it successively alters the conditions of the shells of matter that surround *q*. As the matter outside of B changes its state, a new set of forces will be applied against the *surface* of B *by direct contact*. These 'contact forces' will work their way successively through the shells surrounding *q* and eventually affect *q* itself. The mathematical characteristics of such 'contact forces' are rather different from those of the action-at-a-distance types considered previously.

In the 1820s, Cauchy showed how these two types of force could be harmonized in a fashion that supplies an adequate foundation for continuum mechanics. He first assumed that ' $F = ma$' holds for (almost) every shell B enclosing *q*. '*F*' represents the resultant of summing the contact forces around B, supplemented by whatever external forces operate within B's interior. ' $F = ma$' then assigns an overall 'acceleration' to B. Unfortunately, these observations still allow individual *points* inside B to move in any of the variety of ways that are consistent with the *averaged* acceleration that noninfinitesimal ' $F = ma$' assigns to B. But a framework of laws is not adequate to continua until the principles that govern *point* motion within a flexible body are articulated. At this juncture, Cauchy (1823) had the brilliant insight that, as the volume of B shrinks to 0, Euler's two principles ensure that a rather abstract assemblage of numbers, called the *stress tensor*, needs to be defined at each point *q*. Cauchy's 'stress' turns out to be the novel ingredient that renders the mechanics of continua quite distinct from the other formulations of classical physics we have considered. Popular books often encourage the impression that 'stress' is merely a synonym for 'force', but Cauchy's tensor actually supplies an abstract replacement for the naïve idea that a 'point' like *q* might possess an 'infinitesimal surface' upon which the surrounding materials can tug and shear in different directions. In other words, 'stress' provides a subtle means of avoiding the paradoxical 'points with surfaces' that conceptually bedevilled earlier attempts to discuss continua.

With Cauchy's new notion in hand, the differences between water and rubber can be traced to the distinct manners in which states of stress affect stretching within the materials. The law of stretching for rubber - its so-called 'constitutive equation' - looks quite different from the law suitable for water. Such 'constitutive principles'

together with Euler's two laws form the core of doctrine upon which continuum mechanics rests.

Cauchy's direct approach to continua usually provides a more reliable guide to the behaviour of real life materials than modelling them as arrays of classical points or rigid bodies. Some of the nineteenth-century hostility to 'molecules' stemmed from a recognition of this empirical superiority (the historical dispute over the so-called 'rariconstant' theory of elasticity was a case in point). But Cauchy's programme is complicated, and, without sufficient care, one can quickly return to conceptual absurdities such as 'points with infinitesimal surfaces'. Many modern texts treat continua as large swarms of point masses simply because this approach provides a 'quick and dirty' way of getting some key classical models up and running, without needing to worry about delicate foundational considerations. Handy appeals to the necessity of 'idealization' camouflage some otherwise awkward *non sequiturs*. But if one hopes to appreciate the conceptual issues with which a Leibniz would have struggled, the difficulties of continua need to be addressed more honestly.

In point of historical record, a variant programme for approaching continua based upon 'strain energy' rather than 'stress' was pursued in the late nineteenth century by the so-called 'energetic school'. See Harman (1982) for further details.

7 Retrospect

Which of these approaches to mechanics should the scientists of the nineteenth century have favoured? On the one hand, they had little doubt that the apparent continua of real life - strings, fluids, and so on - are lumpish in composition, possibly suggesting the suitability of rigid bodies as the basic objects of physics. But such molecules also seemed to be capable of vibration, indicating that the molecular 'lumps' might possibly act as flexible bodies at a microscopic level of analysis (the 'ether' that was thought to surround the molecules also acted like a continuum). Such uncertainties spawned lively arguments over the proper basis of classical mechanics, disputes that were never resolved simply because Nature decided to turn quantum mechanical at just the point where experiments might have settled matters.

Unfortunately, many popular accounts of the history of science and philosophy often forget these genuine conceptual worries and presume that 'classical mechanics' is correctly interpreted only from a point mass stance. The many philosophers and scientists who struggled to articulate variant conceptions are frequently dismissed as incompetents blind to the true mechanical light. But the deep philosophical heritage of classical mechanics is better appreciated if its basic internal tensions are more sympathetically understood.

See also: Conservation principles; Mechanics, Aristotelian

MARK WILSON

References and further reading

Boscovich, R.J. (1763) *Theoria Philosophiae Naturalis*, Venice; trans. J.M. Child, *A Theory of Natural Philosophy*, Cambridge, MA: MIT Press, 1966.(First proposition of the point mass approach.)

Cauchy, A.-L. (1823) 'Recherches sur l'équilibre et sur le mouvement intérieur des corps', *Bulletin des Sciences*, Paris.('Contact forces' can be harmonized in a fashion that supplies an adequate foundation for continuum mechanics.)

Dijksterhuis, E.J. (1961) *The Mechanization of the World Picture*, Oxford: Oxford University Press.(Interesting discussion of Newton's understanding of his laws.)

Duhem, P. (1903) *L'évolution de la mécanique*, Paris: Joanin; German trans. P. Frank, *Die Wandlungen der Mechanik und der mechanischen Naturerklärung*, Leipzig: J.A. Barth, 1912; trans. M. Cole, *The Evolution of Mechanics, Monographs, and Textbooks on Mechanics of Solids and Fluids*, Alphen aan den Rijn, The Netherlands and Germantown, MD: Sijthoff & Noordhoff, 1980.(Important evaluation by a prominent scientist.)

Dugas, R. (1955) *A History of Mechanics*, London: Routledge & Kegan Paul.(Survey with many quotations from original sources.)

Friedman, M. (1972) *Foundations of Space-Time Theories: Relativistic Physics and Philosophy of Science*, Princeton, NJ: Princeton University Press.(Account of Cartan's work by a philosopher.)

Harman, P.M. (1982) *Energy, Force, and Matter*, Cambridge: Cambridge University Press.(Readable survey of nineteenth-century developments including the 'energetics' school.)

Hertz, H. (1894) *Die Principien der Mechanik*, Leipzig: J.A. Barth; also *Gesammelte Werke*, vol. 3, Leipzig: J.A.

Barth, 1894; trans. D.E. Jones and J.T. Walley, *The Principles of Mechanics*, London: Macmillan, 1899; repr. New York: Dover, 1956.(The introduction is Hertz's main contribution to philosophy. Discussed in §5, the preface is wonderful.)

Mach, E. (1883) *Die Mechanik in ihrer Entwicklung historisch-critisch dargestellt*, Prague; trans. T. McCormack, 1893; *The Science of Mechanics*, La Salle, IL: Open Court, 1974.(A history of the evolution of science, including a comprehensive attempt to reconstruct classical mechanics on positivist lines. An influential critique.)

Newton, I. (1726) *Isaac Newton's Philosophiae naturalis principia mathematica*; 3rd edn, with variant readings, A. Koyré and I.B. Cohen (eds) with assistance of A. Whitman, Cambridge, MA: Harvard University Press, 1972, 2 vols.(The classic, but hard to read.)

Poincaré, H. (1902) *La Science et L'Hypothése*; repr. *Science and Hypothesis*, New York: Dover, 1952.(Parts II and III outline the 'conventionalism' discussed in §1.)

Sneed, J. (1979) *The Logical Structure of Mathematical Physics*, Dordrecht: Reidel.(Mentioned in §2.)

Thompson, W. and Tait, P.G. (1879) *A Treatise on Natural Philosophy*, Cambridge.(A classic of nineteenth-century thought; quoted in §1.)

Truesdell, C. (1968) *Essays in the History of Mechanics*, New York: Springer.

Truesdell, C. (1977) *A First Course in Rational Continuum Mechanics*, New York: Academic.(The many works by Truesdell supply the best guide to the continuum tradition of §6. The 'First Course' is advanced.)

Medical ethics

Medical ethics was once concerned with the professional obligations of physicians, spelled out in codes of conduct such as the ancient Hippocratic oath and elaborated by contemporary professional societies. Today this subject is a broad, loosely defined collection of issues of morality and justice in health, health care and related fields. The term 'bioethics' is often used interchangeably, though it is also used with its original broad meaning, which included issues in ecology.

The range of concerns grouped under 'medical ethics' begins with the relationship of doctor to patient, including such issues as consent to treatment, truth-telling, paternalism, confidentiality and the duty to treat. Particular moral uncertainty is engendered by contexts which demand divided allegiances of physicians, such as medical experimentation on human subjects, public health emergencies and for-profit medicine. Issues in medical ethics arise in every stage of life, from the fate of defective newborns to the withholding of life-sustaining therapies from the very old. Medical practices with patients who may not be competent to make their own medical decisions, including paediatrics and psychiatry, raise a distinctive set of ethical issues, as does medical genetics, which involves choices affecting family members, future individuals and offspring in utero. *In recent years, medical ethics has broadened its focus beyond the individual physician or nurse to include the organization, operation and financing of the health care system as a whole, including difficult theoretical and practical uncertainties regarding the fair allocation of health care resources.*

Medical ethics is at once a field of scholarship and a reform movement. The latter has campaigned in many countries on behalf of patients' rights, better care of the dying and freedom for women in reproductive decisions. As a field of scholarship, medical ethics addresses these and many other issues, but is not defined by positions taken on any of them. Though ethicists often favour an emphasis on informed consent, oppose paternalism, urge permission to end life-sustaining therapy (or choose suicide) and seek protection of human subjects of experimentation, a diversity of viewpoints finds expression in the medical ethics literature.

1 Development of medical ethics

Medical ethics, in the narrow sense of codes of professional conduct for physicians, has a history dating back to the Hippocratic oath of the fourth century BC (see Hippocratic medicine). But medical ethics became a broad field of scholarship only after the Second World War. The terrible experiments carried out by Nazi scientists on nonconsenting human subjects were publicized and judged at the Nuremberg trials, alerting researchers and the public to the potential for abuse in clinical investigations. This lesson was learned anew in the two decades which followed as reports emerged in the United Kingdom and the United States of wrongs done to patients in the name of science, such as the injection of live cancer cells into unsuspecting patients, and the deliberate infection of institutionalized children with hepatitis. These and other abuses were detailed by the investigators in articles published in leading research journals.

Public attention was drawn to medical ethics, at least in the United States, by the moral dilemmas faced by the world's first kidney dialysis service, which had to choose which of many patients needing treatment would be permitted to live. The dialysis service, at Swedish Hospital in Seattle, created a committee in 1962 to ponder the patients' dossiers and to 'play God' in deciding who was most deserving (the winning candidates were often patients who most resembled the committee members). Even more widely discussed, a decade later, was the fate of Karen Quinlan, a young woman who had fallen into vegetative state. Her family sought permission from the courts to remove her from life-sustaining therapy, exercising a 'right to die'. A series of cases during this period involving the termination of treatment of defective newborns led not only to public debate but to stringent government guidelines for treatment, including, at one point, a national telephone number for reporting suspected doctors. Debates over doctors as agents of death intensified in many countries in the wake of increasingly liberal euthanasia policies in the Netherlands during the 1990s.

The public's fascination with medical ethics, which has increased year by year for over two decades, stems both from advances in medicine and from social changes which affect attitudes towards medical practice (see Technology and ethics §3). Each new drug, device or technique, from the dialysis machine to genetically-engineered human growth hormone, necessitates a decision on its use. Media reports of these advances

have come to include, as a matter of routine, some discussion of attendant ethical issues. A recent example is that of Dolly the sheep, cloned from a cell in another sheep's udder; global reportage of this case prompted ethical questions about the potential use of cloning techniques on humans. Just as important, however, have been issues arising from the reconsideration of traditional medical practices in light of changes in social attitudes. The turbulent 1960s, which gave rise to 'liberation' movements on behalf of many marginalized groups, led to a reconsideration of traditional physician authority, including prerogatives to withhold information and to decide the course of treatment. Thus a variety of traditional medical practices, ranging from choice of which women could be candidates for artificial insemination to the use of charity patients for medical experimentation, became targets of protest and subjects of debate.

The agenda for medical ethics has thus been determined by events outside the discipline itself rather than by its own path of intellectual progress. Anthologies and textbooks require frequent new editions, adding some topics and dropping others, tracking the course of new developments in medicine and the interests of the public. Medical ethics is thus a field of scholarship in public service. It is simultaneously a field of practice. Ethicists, those who are credited with expert knowledge of medical ethics, serve on the staffs of hospitals, in medical research institutes, and as staff members of public agencies (see Moral expertise). A Society for Bioethics Consultation has functioned for many years in the United States, and ethics institutes in the Netherlands and elsewhere in Europe have been likewise composed of philosophers and others who provide ethical advice on a professional basis. National bioethics commissions in over twenty countries offer opinions on issues ranging from the definition of death to the rationing of health care resources (see Death §§1, 3). Medical ethics has become a global concern, and is debated in such bodies as UNESCO's International Bioethics Committee and the International Association of Bioethics, as well as regional groups in Europe, East Asia and South America.

Though the boundaries of the field of medical ethics are vague and shifting, a number of subjects form a recognizable core. Prominent among these are ethical issues arising in reproduction, genetics and death and dying (see Genetics and ethics; Reproduction and ethics; Life and death). Another central issue is the relationship between doctor and patient.

2 The doctor-patient relationship

In traditional medical ethics, the doctor held the reins of authority. The medical mission was to improve or protect the health of the patient. The doctor might accomplish this by giving patients the unvarnished truth, respecting their desire for confidentiality, and pointing out alternative courses of treatment. But if doctors judged these to stand in the way of health, their obligation might be to shade or withhold the truth, share information about patients with others, and manipulate patients into accepting their preferred mode of treatment.

Changes in social attitudes in English-speaking countries have brought demands for a relationship of doctor and patient based instead on respect for the patient's own wishes. This is understood to require the doctor to be entirely candid about the patient's condition, to disclose the full range of alternative therapies, with an account of their benefits and risks, and to share information about the patient with others only when given express permission by the patient.

Much of the early writing in medical ethics in the UK and US was devoted to supporting these new expectations, and self-determination has replaced paternalism as the organizing principle of medicine (see Paternalism). Traditionalist understandings of the doctor-patient relationship persist in other parts of the world, such as Asia and Latin America, where the Anglo-American insistence on self-determination in such matters as disclosure in the diagnosis of fatal illness is regarded as heartless and irresponsible, robbing the patient of hope.

In the older literature of medical ethics, which was chiefly concerned with the duties of the professional as such, rules on such unprofessional practices as fee-splitting, self-referral, and kickbacks played a prominent role in specifying the requirement of fidelity to the patient's interest (see Professional ethics). These issues were replaced at centre stage by the more dramatic questions of life and death which made medical ethics of great interest to the public. However, they have returned to prominence in recent years as innovations in health care financing have deliberately placed physicians, in some countries, in positions of potential conflict of interest in order to motivate them to limit the cost of care. In the United States, 'managed care' organizations often make the conflict explicit, withholding part of the doctor's pay if the latter has failed to be sufficiently frugal in ordering high-cost services

for patients. Critics point to these arrangements as an abandonment of professionalism and a blow to patients' trust of doctors, while proponents point to the doctors' very professionalism as a safeguard against the temptation to exploit the patients' lack of medical knowledge when withholding needed care.

3 Human experimentation

Abuses of human subjects of medical experimentation, most notorious in Nazi medicine but now known to have been practised elsewhere, notably by Japanese physicians acting in occupied China, seem to have abated since the Second World War. Nevertheless, ethical dilemmas persist and have been a continuing concern of scholars and practitioners in medical ethics. The basic norms of humane research have been enforced in the UK and US for some decades and, as with clinical medicine, are founded on the two principles of physician concern for the patient and on patient self-determination, secured in large part through informed consent (see Consent). The former, however, is diluted inevitably by the physician-scientist's concern for future patients, who will receive no benefits unless the experiment is successful in scientific terms.

The potential conflict between the roles of therapist and scientist arises in many randomized clinical trials (RCTs), which is especially troublesome because the RCT has emerged as the gold standard of evidence of clinical efficacy. Patients recruited to RCTs are typically told that the investigators do not know which arm of the trial (for example, a drug or a placebo) is most beneficial. Even if this is true at the beginning of the study, however, RCTs typically enrol patients over a period of months or years, and later recruits may not be informed of trends in accumulating (if inconclusive) data which might lead a clinician to favour one treatment or the other rather than flipping a coin. The conflict in roles between healer and scientist remains.

4 Ethical issues in psychiatry, paediatrics and geriatrics

The currently favoured model of the doctor-patient relationship centres on the doctor's duties of care and advocacy for the patient and the patient's right of self-determination. The latter, however, assumes that patients are competent to make their own decisions. Physicians in psychiatry, paediatrics and geriatrics deal with patients whose decision-making capacities are sometimes lacking, and with others whose mental competence fluctuates over time.

No satisfactory substitute for full decision-making competence has yet been devised, but several strategies are used. For decisions at the end of life, particularly those which need to be made when they are debilitated by illness or are in vegetative state, some patients have executed 'living wills' setting out instructions for their care and have given trusted intimates 'durable power of attorney' to make decisions on their behalf. Unfortunately, these instruments have not proved effective: surveys have shown that physicians tend to overlook them and permit family members to override them. For mental patients, the so-called 'Ulysses Contract' (recalling the island of the Sirens) would permit the patient to waive, in advance, rights to refuse sedation and other treatment in the event of temporary loss of full competence, as in a manic episode. All these strategies require an earlier period of mental competence in which the instruments are created, which makes them inapplicable to paediatric cases. In the latter, and often in psychiatric and geriatric cases as well, the physician must instead rely on others to decide on the patient's behalf or else rely on the decisions of a marginally competent patient. Since self-determination is valuable at least in part because of the protection it affords to the (competent) patient, this last strategy is sometimes self-defeating, though it may be the least unsatisfactory alternative.

5 Allocation of health care resources

The central ethical issues in allocation of health care resources - who should receive what kinds of care, and who should pay for it - are questions of distributive justice (see Justice §2). Though the texts which sparked the revival of philosophical work in distributive justice, particularly Rawls' *A Theory of Justice* (1971), made little mention of health care, attempts to extend the philosophical work to the arena of health were not long in coming. These works addressed a series of questions of both philosophical interest and practical importance. Should health care be distributed in any distinctive way, different from other goods and services? Does it have any particular priority? Is there a floor or a limit on the share of social resources which should be devoted to health care? How equal should provision of health care be, and on what basis should scarce health care goods and services be allocated? What role should markets play in determining health care allocations, and what role democratic political choice?

In the United States, there is considerable debate among philosophers and the public alike on whether a society which fails to guarantee at least a basic level of health care benefits can be called just (in other industrialized countries, this is not questioned seriously). Even where basic health services are available to all, debate exists over the degree to which the remainder of health services ought to be delivered on an egalitarian footing, and over what kinds of allocations qualify as egalitarian.

Under the heading of ‘prioritization’, philosophers have addressed the issues involved in rationing decisions by government authorities, bedside clinicians, and private insurers alike. Ought the goal of rationing be to maximize health, understood as the greatest total number of years of life of all patients affected, discounted by remaining deficits in health? Or should the sickest patients, or those with the most urgent needs, be given priority? Should large gains in health take priority over smaller ones, even if the total gain in health, within a given budget, is achievable in a given instance by the aggregation of the smaller gains? Should the rationing agent consider the self-interest of each patient alone, or should the wishes of third parties (including family members, members of a risk pool, and, most generally, empathic fellow-citizens) also be taken into account?

Health policy has come to rely on the ‘Quality-Adjusted Life Year’ and other measures of the success of health care, but deep divisions over their implications persist in medical ethics. A commitment to maximizing QALYs, for example, tends (*ceteris paribus*) to choose the younger patient over the older, which strikes some theorists as proper but others as arbitrary or discriminatory.

Rationing decisions have been made in the past through a combination of means, including physician discretion, budget allocation and private purchasing in markets. The principles upon which allocation has been based have rarely been explicit. Though little consensus exists in medical ethics over the resolution of these controversies, the rising costs of health care prompt demands for allocation decisions which are principled, public and democratic. A very ambitious attempt to honour these demands, conducted by an agency of the state of Oregon, in the United States, has been scrutinized by health authorities around the globe, and several have initiated smaller-scale initiatives to the same end.

More broadly still, medical ethics has begun to address health-related issues which transcend health care. Enduring differences in health status according to class and race, even in countries such as the UK which provide good health care to all citizens, prompt questions of social justice about other determinants of health. These inquiries are particularly important in consideration of medical ethics in developing countries (see Development ethics).

6 Theory and methodology

With one foot in the academy and the other planted in the ‘real worlds’ of clinical medicine and health policy debate, medical ethics is, of necessity, an untidy discipline. The contexts in which ethicists perform their duties, be they an intensive care ward or a public agency, often preclude elaborate philosophical argumentation. Medical ethicists, moreover, are a varied lot. There is no credential nor any single disciplinary background: philosophers share the title with theologians, lawyers, physicians, social scientists and anyone else who claims it.

It is therefore not surprising that medical ethics lacks theories and a standard methodology. Though contributors to the subject may be informed by the traditions of moral thought, some writing, for example, as utilitarians, others as Kantians, and others as Roman Catholic moral theologians, there is no approach which any appreciable plurality share. The most widely-noted theoretical framework for medical ethics specifies a set of *prima facie* duties for health care professionals (Beauchamp and Childress 1979; Gillon and Lloyd 1994). These principles, however, must be weighted and applied in context by the decision-maker, and in the view of critics fall short of a comprehensive theory. In practice, ethicists approach a clinical case or a public policy issue more or less as anyone else would, except in being conversant with a literature which makes important distinctions, advances and criticizes leading lines of arguments, and dwells in rarefied abstraction. Ethicists begin with no standard set of assumptions, and do not necessarily converge in their conclusions.

What intellectual authority, then, does medical ethics bring to bear on ethical dilemmas in health care? Perhaps none; but this is not where its promise and its value lie in any case. Medical ethics is but one small part of the conversation of humankind, but it is, at its best, of particular interest. It is a formal and considered response to a remarkable challenge from within and without the profession of medicine to debate - publicly, impersonally and prospectively - whether medical practices are right and just. Its theoretical and methodological anarchy

notwithstanding, this is an endeavour which can be commended to other arenas of public life.

See also: Applied ethics; Bioethics; Bioethics, Jewish §2; Medicine, philosophy of; Nursing ethics

DANIEL WIKLER

References and further reading

Beauchamp, T. and Childress, J. (1979) *Principles of Biomedical Ethics*, Oxford: Oxford University Press; 4th edn, 1994.(Influential presentation of one theoretical approach in medical ethics.)

Bell, J.M. and Mendus, S. (1988) *Philosophy and Medical Welfare*, Cambridge: Cambridge University Press. (Papers on health care allocation, including the maximization of 'Quality-Adjusted Life Years'.)

Brock, D. (1993) *Life and Death*, Cambridge: Cambridge University Press. (Influential papers on clinicial issues in medical ethics, and on health care delivery.)

Buchanan, A. and Brock, D. (1990) *Deciding for Others*, Cambridge: Cambridge University Press.(Theory of surrogate decision-making.)

Callahan, D. (1987, 1990) *Setting Limits* and *What Kind of Life?*, New York: Simon & Schuster; repr. with reply to critics, Washington, DC: Georgetown University Press, 1995.(Widely-read discussions of allocation of resources.)

Daniels, N. (1985) *Just Health Care*, Cambridge: Cambridge University Press.(An account of justice in the allocation of health care resources.)

Daniels, N. (1990) *Am I My Parents' Keeper?: An Essay on Justice Between the Young and the Old*, Oxford: Oxford University Press.(Allocation of health care resources across lifetimes.)

Gillon, R. and Lloyd, A. (1994) *Principles of Health Care Ethics*, Chichester and New York: John Wiley & Sons. (A collection of papers on themes in medical ethics which apply Beauchamp and Childress' approach.)

Glover, J. (1977) *Causing Deaths and Saving Lives*, Harmondsworth: Penguin. (Broad, readable discussion of clinical issues in medical ethics.)

Kamm, F.M. (1993, 1996) *Mortality, Morality*, Oxford: Oxford University Press, 2 vols.(Closely argued, extended discussion of allocation of resources, euthanasia and other issues in medical ethics and beyond.)

Potter, V.R. (1971) *Bioethics, Bridge to the Future*, Englewood Cliffs, NJ: Prentice Hall.(Origin of term 'bioethics', broadly defined in ecological terms.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(A work which stimulated recent debate on distributive justice.)

Reich, W. (ed.) (1994) *Encyclopedia of Bioethics*, New York: Macmillan.(Multi-volume, comprehensive set of discussions on major issues in medical ethics.)

Strosberg, M., Wiener, J.M., Baker, R. and Fein, A. (1992) *Rationing America's Medical Care: The Oregon Plan and Beyond*, Washington, DC: The Brookings Institution.(Essays on the Oregon rationing initiative.)

Medicine, philosophy of

The philosophy of medicine can be generally defined as encompassing those issues in epistemology, axiology, logic, methodology and metaphysics generated by or related to medicine. Issues have frequently focused on the nature of the practice of medicine, on concepts of health and disease, and on understanding the kind of knowledge that physicians employ in diagnosing and treating patients.

The history of philosophical reflections concerning medicine reaches back to ancient Greece. Medical knowledge took a further step in the nineteenth century with the introduction of clinical pathological correlations, statistical methods, and systematic experimentation, out of which grew substantive literature exploring the character of medical reasoning and the framing of diagnoses. Debates also developed over contrasting physiological, ontological, nominalist and realist accounts of disease entities.

Contemporary philosophy of medicine has been concerned with the nature of medicine in an increasingly scientific context, a concern that has generated several models of medicine, including George Engel's biopsychosocial model, as well as analyses of the nature of the physician-patient interaction. The longstanding debate over the ontological status of health and disease has been recapitulated and extended by a number of authors, favouring an objective, statistically-based account, while others argue for an irreducible social and valuational element in these concepts. Several approaches to diagnostic logic, including Bayesian and computer-based analyses, have been developed, and sophisticated methods of determining disease causation and therapeutic efficacy, including analyses of the randomized clinical trial, have also been explored. Whether the philosophy of medicine is a distinct discipline or a branch of the philosophy of science has provoked vigorous arguments.

1 Definition and scope

The philosophy of medicine can be generally defined as encompassing those issues in epistemology, axiology, logic, methodology and metaphysics generated by or related to medicine. Though medical ethics is a component of the philosophy of medicine in its broadest conception, this article will not attempt to cover that discipline, except at a few critical points. Medicine is the art or science of restoring or preserving health, and may involve drugs, surgical operations, or other types of intervention. Though medicine has an extensive and continually increasing scientific basis, several physician-philosophers, including Otto Guttentag and Edmund Pellegrino, have seen the essence of medicine in its orientation to one distinct end - the affirmative response of the physician to a plea from the individual patient to help in restoring or preserving that patient's health. Issues in the philosophy of medicine have thus focused on the concept of health (and that of disease), and on understanding the kind of knowledge that physicians employ in diagnosing, treating and caring for patients. These issues have an extensive historical pedigree.

2 History of the philosophy of medicine

There is a rich, complex history of philosophical reflections concerning medicine reaching back to the beginnings of Greek philosophy (see Hellenistic medical epistemology). These can retrospectively be recognized as part of the 'philosophy of medicine', though it was not until the nineteenth century that that expression (and 'medical logic') gained currency. By involving a significant self-critical appreciation of the difficulties in medical knowing, the philosophy of medicine contrasts with various speculative, non-empirical or only quasi-empirical, reflections on the etiology, pathogenesis and cure of disease that have been undertaken in all cultures. An explicit concern with the status of medical theories and medical claims is already present in the Hippocratic corpus (see Hippocratic medicine). The *Precepts*, for example, offers an account of medical theory as a composite memory derived from sense perception; the author underscores the importance of relating medical theorizing to actual experience, thus reflecting a philosophy from the period.

From the time of the Hippocratic corpus into the nineteenth century, mainstream medicine offers an intertwining of speculative and empirical attempts to discover the true nature of disease and proper treatment, set against the background of occasional assessments of medical knowledge. Such intertwinings are found in the work of Galen (129-*c*.210), who underscores the importance of observation and experiment. One finds them as well in René Descartes' attempts to determine the fundamental laws of metaphysics, physics and medicine, especially in the

Treatise on Man (1662) (see Descartes, R. §2). The theoretical disputes of the seventeenth and eighteenth centuries between those using chemistry and those using mechanics, as the basis for theory and practice, were as much speculative as empirical, as were the debates about the nature of life and the status of medical systems. Often physicians who engaged in speculative, largely non-empirical, reflections on medicine, and also attempted to construct medical systems, had significant influence on the practice of medicine. On the other hand, one of the system builders, John Brown (1735-88), attracted the attention of contemporary philosophers such as Kant and Hegel.

In response to the uncritical incorporation of speculative reflections and theorizing into medical theory and practice, Thomas Sydenham (1624-89), the practice partner of John Locke (1632-1704), attempted to nurture a better appreciation of empirical observation in medicine. Inspired by the work of Sir Francis Bacon, Sydenham engendered a cluster of substantial explorations of method in medicine. This literature developed Sydenham's concern not to allow theoretical presuppositions to distort clinical findings. It produced conceptually attentive assessments of medical nosologies and the status of medical knowledge. Among the contributors were Carl von Linnaeus (1707-78) and François Boissier de Sauvages (1701-67). Reflections on the status of medical knowledge took a further step in the nineteenth century with the introduction of clinical pathological correlations, statistical methods, and systematic experimentation. One subsequent line of analysis came from medical scientists, who criticized previous medical knowledge and set out new claims regarding the status of disease entities. To this debate, François-Joseph-Victor Broussais (1772-1838) contributed a special meaning for the term ontological, using it to identify accounts that regarded diseases as in some sense abstract entities. The debates were complex and involved distinguished contributors, such as the cellular pathologist Rudolf Virchow (1821-1902). They encompassed analyses of etiological factors into necessary and sufficient causes. In these debates, distinctions were drawn between physiological and ontological accounts of disease, at times distinguishing physiological and anatomical explanations of disease, and at times nominalist and realist accounts of disease entities. This debate brought together long-standing concerns as to whether 'disease' identified a thing in the world (that is, a disease entity, an *ens morbi*), a recurring constellation of clinical findings, a set of causal factors, or was a term that conventionally identified particular findings.

The changes in medicine during the nineteenth century provoked further assessments of medical knowledge and method. Initially, this produced a mixture of the philosophy of medicine as a specialized speculative philosophy of nature and the philosophy of medicine as an assessment of scientific method in medicine. For example, Blane's *Medical Logick* (1819) incorporated both an enumeration of ten elementary principles of life, as well as a study of fallacies in medical reasoning (including Bartlett's *Philosophy of Medicine* (1844)). By mid-century, this literature compassed a sophisticated appreciation of problems in medical investigation, observation and experiment, which placed concerns about medical knowledge in terms of problems of inductions. Out of this grew a substantive literature exploring the character of medical reasoning and the framing of diagnoses. This tradition of disciplined philosophical reflections concerning medicine and medical knowledge continued as interest in bioethics emerged (see Bioethics). This literature includes the contributions of Ludwik Fleck (1979), who in many respects anticipated the work of Thomas Kuhn in the 1960s. As the second half of this century began, the term 'philosophy of medicine' could identify a complex literature of philosophical reflections.

3 Models of medicine

Medicine has been most generally understood as helping the patient with the care of their health. This perspective makes intelligible why medicine was paradoxically characterized by the nineteenth-century pathologist Virchow as a 'social science', and also suggests why, in spite of the extraordinary advances in the application of reductionistic molecular biology in medicine, many endorse George Engel's 'biopsychosocial' model of medicine as a more appropriate vision of the discipline. This model is not antiscientific, but is viewed as a supplement to a narrower 'biomedical' model. The biopsychosocial model involves a recognition of complex causation, emphasizes that how a disease affects any one individual requires consideration of psychological, social and cultural factors, and stresses individual variability of a disease, which ' reflects as much… [psychological, social and cultural factors] as it does quantitative variations in the specific biochemical defect' (Engel 1977: 132). The centrality of the patient has also led the philosophy of medicine to examine the complexities of the physician-patient relationship, and has explored several different models of this interaction. Consensus, at least in the USA, has since 1970 tended away from a more traditional paternalistic approach to the patient and toward an account that emphasizes the moral

importance of the autonomous and informed patient participating in shared decision making with the physician, tempered by the differing perspectives of each of the parties.

4 Concepts of health and disease

There appears to be a general consensus in philosophy of medicine that the debate over the concepts of health and disease has been the main contender for a 'defining problem' of the field of philosophy of medicine. Reflections concerning concepts of disease, illness, malady and health have since the 1950s recapitulated traditional concerns regarding the status of disease entities. They have assessed the extent to which 'disease', 'illness', 'health', 'disability', and so on, identify states of affairs, reflect cultural values, or direct health care interventions. In 1954 Lester King noted how particular values mark states of health from states of disease. Debates concerning the value-laden and theory-laden character of disease and health concepts, as well as medical notions of normality and abnormality, have provoked attempts to justify culture-independent understandings through appeals to the ends of medicine, the essential functions of a being, the formal conditions of freedom or species-typical (statistically-based) levels of species-typical functions. Such approaches offer the promise of a straightforward critique of the use of medicine as a vehicle of political control, as well as a more unified account of disease and health for both humans and animals. Others have argued for the essentially culture- and value-laden character of concepts of health and disease, as well as the instrumental character of disease and illness, diagnosis, prognosis and treatment. In such accounts, concepts such as disease, illness and malady become warrants for medical interventions (Engelhardt 1996). Distinctions between mental and somatic diseases and illnesses have been developed and criticized, background concepts such as that of malady have been explored, and the role of diagnostic taxons in medical decision-making and artificial intelligence approaches to clinical problem solving and diagnosis have been studied, placing concerns with disease and health in terms of larger themes of medical explanation and reasoning. Recent advances in the molecular genetics of various diseases have occasioned several philosophical analyses of normality and genetic disease (Nordenfelt 1995).

5 Logics of diagnosis, prognosis, and therapy evaluation

Investigations into the logics of diagnosis and prognosis began in the 1950s, stimulated by the thesis of Paul Meehl (1954), that prognoses made by simple statistical models bettered prognoses made by clinicians. More explicit logical analyses of the diagnostic process developed in the 1960s and 1970s, employing both Bayesian inference and branching logics. The former approach was articulated and argued for by Ledley and Lusted in an influential article in 1959, and applied by other clinicians during the 1960s in a variety of medical specialties. The Bayesian approach uses Bayes' theorem to revise initial (or 'prior') probabilities in the light of the outcomes of further clinical tests (see Inductive inference §3; Probability, interpretations of §5). The approach can outperform even experienced clinicians, if confined to narrow areas of inquiry, but requires exponentially growing amounts of data in broader or more complex domains. For many test results, this information is not available.

The branching logic approach is represented by a tree-like structure in which each branch point or node is a decision point for the physician, and can have two or more paths leading from it to additional decision nodes. This approach was defended by Alvan Feinstein in several articles in the 1970s, and applied in neurology by Kleinmuntz, and to the diagnosis of acid-base disorders by Bleich. Relics of this approach can still be seen in various diagnostic flow charts in textbooks of medicine. Branching trees can have 'chance' or probabilistic nodes inserted within them that can be subject to Bayesian inference. This type of structure, supplemented with quantitative measures of benefits and harms (utilities), constitutes the still vigorous subject of 'clinical decision analysis'. In spite of a large number of proponents still committed to the decision-theoretic approach, the need to adhere to a strict order of information presentation, and the lack of empirical information about certain key conditional probabilities, led other investigators of the diagnostic process, such as A. Gorry and J. Kassirer, to look toward the developing area of artificial intelligence for a more tractable approach.

Work in artificial intelligence in medicine (AIM) and philosophical commentary on it flourished during the late 1970s and 1980s, resulting in a number of accomplishments, including the INTERNIST-1 diagnostic program, later known as QMR (Quick Medical Record). The QMR program is now marketed for physicians, and used as a teaching tool. AIM research continues to be actively pursued under the more inclusive rubric of 'medical informatics', which has turned its attention beyond diagnostic reasoning to investigate the possibilities of a unified medical language and integrated medical information systems. In spite of the advances made in diagnostic, prognostic and

therapeutic logics, there is at present no consensus that any one approach has fully captured the physician's reasoning process. Inquiry continues, and includes newer approaches such as connectionist systems, case-based reasoning and fuzzy logics.

Philosophical inquiry has also been directed at methods for determining the causes of disease and evaluating therapies. These issues are also addressed under the rubric of 'clinical epidemiology'. An analysis of how a specific agent can be identified as the 'cause' of a disease has its historical roots in Koch's postulates, which articulated a series of necessary and sufficient conditions for a disease's cause within the context of nineteenth-century bacteriology. Twentieth-century discoveries in microbiology have required the generalization of Koch's postulates, and the development of a more complex approach to the identification and confirmation of disease causation (Evans 1976). A protracted debate about the true cause of AIDS, begun by Duesberg in the late 1980s, has been conducted largely within the context of this inquiry. A related question involving causation, how to determine whether a medical intervention has had the effect of ameliorating or curing a disease, has generated an extensive literature (as well as numerous governmental regulations), most frequently in the area of new therapeutic drugs. Philosophers of medicine have participated, along with statisticians and clinicians, in analyses of various research designs, including the soundness, and the ethics, of what is generally acknowledged to be the 'gold standard' of efficacy determination: the randomized controlled clinical trial (Schaffner 1993) (see Medical ethics §3; Statistics and social science).

6 Relation of philosophy of medicine to philosophy of science

Since the inception of the current era of professional inquiry in the philosophy of medicine - inaugurated by the *Philosophy and Medicine Series* in 1975, *The Journal of Medicine and Philosophy* in 1976 and its sister publication *Theoretical Medicine* (née *Metamed*) in 1979 - an ongoing debate has occurred on whether the philosophy of medicine is a distinctive branch or field of philosophy, or whether it is reducible to or replaceable by some combination of philosophy of science (including the philosophy of biology), philosophy of mind and moral philosophy. Critics of the distinctiveness of the discipline, such as J. Shaffer and A. Caplan, have tended to the view that medicine is essentially the science of biology, and philosophy of medicine is (and should be best pursued as) philosophy of science. What is left out can be assigned to the realm of moral philosophy. Caplan elaborates (1992), arguing that philosophy of medicine is not at present a disciplinary 'field'. It is not a field because it is not integrated into its cognate areas of either philosophy or medicine, it has no set of core readings or 'canon', and it has no core problems (with the exception of the controversy about health and disease as reviewed) about which deep intellectual debates rage. Though Caplan does not deny the theoretical possibility of a philosophy of medicine - in fact, he suggests it should be a necessary foundation for bioethics, could have salutary consequences for philosophy of science, and could be important for medicine in its inquiries into clinical trials, computerized diagnosis and modern molecular genetics - he does not think such a discipline as yet exists.

Defenders of the distinctiveness of a discipline of philosophy of medicine, such as Pellegrino and Wulff, have argued that there are many and varied examples of the application of philosophical approaches and methods to medical topics, some of which would appear to fall into just those areas that Caplan urges be pursued. Sometimes these examples may be viewed as instances of philosophy *and* medicine or as philosophy *in* medicine, but these inquiries all share strong similarities with each other and buttress a more distinctive philosophy *of* medicine. The latter term Pellegrino (1976) prefers to reserve for a philosophical investigation of 'the clinical encounter with a human being experiencing health, illness, neurosis, or psychosis, in a setting which involves intervention into his existence'. Others may favour the broader definition of the term 'philosophy of medicine' used in the present article, which would include philosophical inquiries within medicine generally.

See also: Bioethics, Jewish; Qi

KENNETH F. SCHAFFNER
H. TRISTRAM ENGELHARDT, JR

References and further reading

Bartlett, E. (1844) *An Essay on the Philosophy of Medical Science*, Philadelphia, PA: Lea & Blanchard.(An exploration of medical knowledge medical causality, as well as contemporary theories of medicine, including nosologies; referred to in §2.)

Blane, G. (1819) *Elements of Medical Logick or Philosophical Principles of the Practice of Physik*, London: Thomas & George Underwood.(An overview of the principles of life, as well as principles of reasoning and sources of error in medical reasoning; referred to in §2.)

Broussais, F.-J.-V. (1821) *Examen des doctrines médicales et des systèms de nosologie*, Paris: Méuignon-Marvis, 2 vols.(A contributor to the distinction between so-called physiological and ontological accounts of disease; referred to in §2.)

Caplan, A. (1992) 'Does the Philosophy of Medicine Exist?', *Theoretical Medicine* 13: 67-77.(A critical reflection on the philosophy of medicine as a field; referred to in §6.)

Descartes R. (*c*.1630-3) *Le Monde (The World)*, excerpted in vol. 1 of *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91.(Descartes' first draft of a scientific system, including general physics, cosmology, terrestrial physics and human physiology. The *Treatise on Man* was first published in Latin translation in 1662, then in the French original in 1664.)

Engel, G. (1977) 'The Need for a New Medical Model: A Challenge for Biomedicine', *Science* 196: 129-36.(The classic article developing the concept of the biopsychosocial model of medicine; referred to and quoted in §3.)

Engelhardt, H.T., Jr and Erde, E. (1984) 'Philosophy of Medicine' in P.T. Durbin (ed.) *A Guide to the Culture of Science, Technology, and Medicine*, New York: Free Press, 364-461, 675-7.(The literature of the philosophy of medicine and bioethics is reviewed and an extensive bibliography provided; expansion of the material of §2.)

Engelhardt, H.T., Jr (1996) *The Foundations of Bioethics*, New York: Oxford University Press, 2nd edn.(A systematic exploration of medical explanation and concepts of disease.)

Evans, A.S. (1976) 'Causation and Disease: The Henle-Koch Postulates Revisited', *Yale Journal of Biology and Medicine* 49: 175-95.(The history and development of Koch's postulates; referred to in §5.)

Fleck, L. (1979) *Genesis and Development of a Scientific Fact*, trans. F. Bradley and T.J. Trenn, Chicago, IL: University of Chicago Press.(A classic in the philosophy of medicine that originally appeared in German in 1935; referred to in §2.)

The Journal of Medicine and Philosophy (1976-) Lisse: Swets & Zeitlinger. (Regularly has articles and issues directed to the philosophy of medicine.)

King, L. (1978) *The Philosophy of Medicine: The Early Eighteenth Century*, Cambridge, MA: Harvard University Press.(An overview of the philosophy of medicine in the eighteenth century; material relevant to §§2 and 4.)

Ledley, R.S. and Lusted, L.B. (1959) 'Reasoning Foundations of Medical Diagnosis', *Science* 130: 9-21.(A seminal article arguing for a Bayesian approach to modelling medical diagnosis; refered to in §5.)

Linnaeus, C. von (1763) *Genera morborum, in auditorum usum*, Uppsala: Steinert.(An influential classification of disease offered by the great botanist; referred to in §2.)

Meehl, P.E. (1954) *Clinical Versus Statistical Prediction; A Theoretical Analysis and a Review of the Evidence*, Minneapolis, MN: University of Minnesota Press.(Suggests statistical prediction is better than predictions made by clinicians; referred to in §5.)

Nordenfelt, L. and Lindahl, B.I.B. (eds) (1984) *Health, Disease, and Causal Explanations in Medicine*, Dordrecht: Reidel.(Articles and material relevant to §4 of this entry.)

Nordenfelt, L. (1995) *On the Nature of Health*, Dordrecht: Kluwer.(An exploration of contemporary accounts of health and disease from the perspective of action theory.)

Pellegrino, E. (1976) 'Philosophy of Medicine: Problematic and Potential', *The Journal of Medicine and Philosophy* 1: 5-31.(A seminal article that presents classical and contemporary themes in philosophy of medicine; referred to and quoted in §§1 and 6 of this entry.)

Sauvages de la Croix, F.B. de (1763) *Nosologia methodica sistens morborum classes juxta Sydenhami mentem et botanicorum ordinem*, Amsterdam: Fratrum de Tournes, 5 vols.(A monumental attempt to provide a systematic nosology, compassing some 2,400 diseases; referred to in §2.)

Schaffner, K.F. (1981) 'Modeling Medical Diagnosis: Logical and Computer Approaches', *Synthèse* 47: 163-99. (A summary of contemporary approaches to the logic of diagnosis; contains references to the material of §5.)

Schaffner, K.F. (1992) 'Philosophy of Medicine', in M. Salmon *et al.* (eds) *Introduction to Philosophy of Science,* Englewood Cliffs, NJ: Prentice Hall, 310-45.(A development of the biopsychosocial model and its relation to reduction; expansion of the material of §§1 and 3.)

Schaffner, K.F. (1993) 'Clinical Trials and Causation: Bayesian Perspectives', *Statistics in Medicine* 12: 1,477-94.(Addresses the concepts of causation found in clinical trials within the Bayesian framework; expansion of the material of §5.)

Sydenham, T. (1848) *The Works of Thomas Sydenham, M.D.*, trans. R.G. Latham, London: Sydenham Society, 2

vols.(Explorations of medicine, medical theory and method; referred to in §2.)

Theoretical Medicine (1979) Dordrecht: Kluwer.(Offers articles exploring issues in the philosophy of medicine and bioethics.)

Taylor, F.K. (1979) *The Concepts of Illness, Disease and Morbus*, Cambridge: Cambridge University Press.(A historical and philosophical overview of concepts of disease; articles and material relevant to §4.)

Virchow, R. (1895) *Hundert Jahre allgemeiner Pathologie*, Berlin: Hirschwald.(A synthesis of medical explanation at the end of the nineteenth century by one of its most influential theoreticians; referred to in §2.)

Medieval philosophy

Medieval philosophy is the philosophy of Western Europe from about AD 400-1400, roughly the period between the fall of Rome and the Renaissance. Medieval philosophers are the historical successors of the philosophers of antiquity, but they are in fact only tenuously connected with them. Until about 1125, medieval thinkers had access to only a few texts of ancient Greek philosophy (most importantly a portion of Aristotle’s logic). This limitation accounts for the special attention medieval philosophers give to logic and philosophy of language. They gained some acquaintance with other Greek philosophical forms (particularly those of later Platonism) indirectly through the writings of Latin authors such as Augustine and Boethius. These Christian thinkers left an enduring legacy of Platonistic metaphysical and theological speculation. Beginning about 1125, the influx into Western Europe of the first Latin translations of the remaining works of Aristotle transformed medieval thought dramatically. The philosophical discussions and disputes of the thirteenth and fourteenth centuries record later medieval thinkers’ sustained efforts to understand the new Aristotelian material and assimilate it into a unified philosophical system.

The most significant extra-philosophical influence on medieval philosophy throughout its thousand-year history is Christianity. Christian institutions sustain medieval intellectual life, and Christianity’s texts and ideas provide rich subject matter for philosophical reflection. Although most of the greatest thinkers of the period were highly trained theologians, their work addresses perennial philosophical issues and takes a genuinely philosophical approach to understanding the world. Even their discussion of specifically theological issues is typically philosophical, permeated with philosophical ideas, rigorous argument and sophisticated logical and conceptual analysis. The enterprise of philosophical theology is one of medieval philosophy’s greatest achievements.

The way in which medieval philosophy develops in dialogue with the texts of ancient philosophy and the early Christian tradition (including patristic philosophy) is displayed in its two distinctive pedagogical and literary forms, the textual commentary and the disputation. In explicit commentaries on texts such as the works of Aristotle, Boethius’ theological treatises and Peter Lombard’s classic theological textbook, the *Sentences*, medieval thinkers wrestled anew with the traditions that had come down to them. By contrast, the disputation - the form of discourse characteristic of the university environment of the later Middle Ages - focuses not on particular texts but on specific philosophical or theological issues. It thereby allows medieval philosophers to gather together relevant passages and arguments scattered throughout the authoritative literature and to adjudicate their competing claims in a systematic way. These dialectical forms of thought and interchange encourage the development of powerful tools of interpretation, analysis and argument ideally suited to philosophical inquiry. It is the highly technical nature of these academic (or scholastic) modes of thought, however, that provoked the hostilities of the Renaissance humanists whose attacks brought the period of medieval philosophy to an end.

1 Historical and geographical boundaries

The terms ‘medieval’ and ‘Middle Ages’ derive from the Latin expression *medium aevum* (the middle age), coined by Renaissance humanists to refer to the period separating the golden age of classical Greece and Rome from what they saw as the rebirth of classical ideals in their own day. The humanists were writing from the perspective of the intellectual culture of Western Europe, and insofar as their conception of a middle age corresponds to an identifiable historical period, it corresponds to a period in the history of the Latin West. The historical boundaries of medieval intellectual culture in Western Europe are marked fairly clearly: on the one end by the disintegration of the cultural structures of Roman civilization (Alaric sacked Rome in AD 410), and on the other end by the dramatic cultural revolution perpetrated by the humanists themselves (in the late fourteenth and fifteenth centuries). There is some justification, therefore, for taking ‘medieval philosophy’ as designating primarily the philosophy of the Latin West from about AD 400-1400.

There were, of course, significant non-Latin philosophical developments in Europe and the Mediterranean world in this same period, in the Greek-speaking Byzantine empire, for example, and in Arabic-speaking Islamic and Jewish cultures in the Near East, northern Africa and Spain. None of these philosophical traditions, however, was radically cut off from the philosophical heritage of the ancient world in the way the Latin-speaking West was by the collapse of the Roman Empire. For that reason, those traditions are best treated separately from that of western Europe. Accordingly, they are dealt with in this article only to the extent to which they influence developments in

medieval philosophy in the Latin West.

(See Byzantine philosophy; Hellenistic philosophy; Humanism, Renaissance; Islamic philosophy; Jewish philosophy; Medieval philosophy, Russian; Renaissance philosophy.)

2 Beginnings

The general character of medieval philosophy in the West is determined to a significant extent by historical events associated with the collapse of Roman civilization. The overrunning of Western Europe by invading Goths, Huns and Vandals brought in its wake not only the military and political defeat of the Roman Empire but also the disintegration of the shared institutions and culture that had sustained philosophical activity in late antiquity. Boethius, a Roman patrician by birth and a high-ranking official in the Ostrogothic king's administration, is an eloquent witness to the general decline of intellectual vitality in his own day. He announces his intention to translate into Latin and write Latin commentaries on all the works of Plato and Aristotle, and he gives as his reason the fear that, lacking this sort of remedial aid, his own Latin-speaking and increasingly ill-trained contemporaries will soon lose access altogether to the philosophical legacy of ancient Greece. Boethius' assessment of the situation appears to have been particularly astute, for in fact in the six centuries following his death (until the mid-twelfth century), philosophers in the West depended almost entirely on Boethius himself for what little access they had to the primary texts of Greek philosophy. Moreover, since he had barely begun to carry out his plan when his execution for treason put an end to his work, Boethius' fears were substantially realized. Having translated only Aristotle's treatises on logic together with Porphyry's introduction to Aristotle's *Categories* (see Aristotle; Porphyry) and having completed commentaries on only some of the texts he translated, Boethius left subsequent generations of medieval thinkers without direct knowledge of most of Aristotle's thought, including the natural philosophy, metaphysics and ethics, and with no texts of Plato (though a small portion of the *Timaeus* had been translated and commented on by Calcidius in the fourth century). Medieval philosophy was therefore significantly shaped by what was lost to it. It took root in an environment devoid of the social and educational structures of antiquity, lacking the Greek language and cut off from the rich resources of a large portion of classical thought. Not surprisingly, the gradual reclamation of ancient thought over the course of the Middle Ages had a significant impact on the development of the medieval philosophical tradition.

Medieval philosophy, however, was also shaped by what was left to it and, in particular, by two pieces of the cultural legacy of late antiquity that survived the collapse of Roman civilization. The first of these is the Latin language, which remained the exclusive language of intellectual discourse in Western Europe throughout the Middle Ages and into the Renaissance and Enlightenment. Latin provided medieval thinkers with access to some important ancient resources, including Cicero, Seneca, Macrobius, Calcidius, the Latin Church Fathers (see Patristic philosophy), Augustine and Boethius. These Latin sources gave early medieval thinkers a general, if not deep, acquaintance with classical ideas. Augustine is far and away the most significant of these Latin sources. His thought, and in particular his philosophical approach to Christianity and his Christianized Neoplatonist philosophical outlook, profoundly affect every period and virtually every area of medieval philosophy (see §5).

The second significant piece of late antiquity to survive into the Middle Ages is Christianity. Christianity had grown in importance in the late Roman Empire and, with the demise of the empire's social structures, the Church remained until the twelfth century virtually the only institution capable of supporting intellectual culture. It sustained formal education in schools associated with its monasteries, churches and cathedrals, and provided for the preservation of ancient texts, both sacred and secular, in its libraries and scriptoriums. Medieval philosophers received at least some of their formal training in ecclesiastical institutions and most were themselves officially attached to the Church in some way, as monks, friars, priests or clerks. In the later Middle Ages, the study of theology was open only to men who had acquired an arts degree, and the degree of Master of Theology constituted the highest level of academic achievement. Consequently, most of the great philosophical minds of the period would have thought of themselves primarily as theologians. Moreover, in addition to providing the institutional basis for medieval philosophy, Christianity was an important stimulus to philosophical activity. Its ideas and doctrines constituted a rich source of philosophical subject matter. Medieval philosophy, therefore, took root in an intellectual world sustained by the Church and permeated with Christianity's texts and ideas (see §5).

(See Aristotelianism, medieval; Augustine; Augustinianism; Boethius, A.M.S.; Clement of Alexandria; *Liber de causis*; Marius Victorinus; Nemesius; Origen; Patristic philosophy; Platonism, medieval; Pseudo-Dionysius;

Stoicism; Tertullian, Q.S.F.; Themistius; Translators)

3 Historical development

The full flowering of the philosophical tradition that grows from these beginnings occurs in the period from 1100 to 1400. Two developments are particularly important for understanding the rapid growth and flourishing of intellectual culture in these centuries. The first is the influx into the West of a large and previously unknown body of philosophical material newly translated into Latin from Greek and Arabic sources. The second is the emergence and growth of the great medieval universities.

Recovery of texts. Medieval philosophers before Peter Abelard had access to only a few texts of ancient Greek philosophy: those comprising 'the old logic' (Aristotle's *Categories* and *De interpretatione* and Porphyry's *Isagōgē*) and a small part of Plato's *Timaeus*. Abelard's generation witnessed with great enthusiasm the appearance in the Latin West of the remainder of Aristotle's logical works ('the new logic': the *Prior* and *Posterior Analytics*, the *Topics* and the *Sophistical Refutations*) (see Language, medieval theories of; Logic, medieval). Over the next hundred years, most of Aristotle's natural philosophy (most importantly the *Physics* and *On the Soul*) and the *Metaphysics* and *Ethics* became available for the first time. Not all of these Aristotelian texts were greeted with the same enthusiasm, nor did medieval philosophers find them all equally congenial or accessible (even in Latin translation). However, it is impossible to overstate the impact that the full Aristotelian corpus eventually had on medieval philosophy. The new texts became the subject of increasingly sophisticated and penetrating scholarly commentary; they were incorporated into the heart of the university curriculum, and over time the ideas and doctrines medieval philosophers found in them were woven into the very fabric of medieval thought. Having never before encountered a philosophical system of such breadth and sophistication, philosophers in the thirteenth and fourteenth centuries understandably thought it appropriate to speak of Aristotle simply as 'the Philosopher'.

As medieval thinkers were rediscovering Aristotle they were also acquiring for the first time in Latin translation the works of important Jewish philosophers such as Avencebrol (see Ibn Gabirol) and Maimonides, and Islamic philosophers such as Avicenna (see Ibn Sina) and Averroes (see Ibn Rushd). Some of their works were commentaries on Aristotle (Averroes became known simply as 'the Commentator') whereas some (such as Avicenna's *Metaphysics* and *De anima*) were quasi-independent treatises presenting a Neoplatonized Aristotelianism (see Aristotelianism in Islamic philosophy). Medieval philosophers of this period turned eagerly to these texts for help in understanding the new Aristotle, and they were significantly influenced by them. Averroes's interpretation of Aristotle's *On the Soul*, for example, sparked enormous controversy about the nature of intellect, and Avicenna's metaphysical views helped shape the famous later medieval debates about universals and about the nature of the distinction between essence and existence.

Rise of the universities. As abbot of the monastery at Bec in the 1080s, Anselm of Canterbury addressed his philosophical and theological writings to his monks. By contrast, the great philosophical minds of the next generations, thinkers such as Abelard, Gilbert of Poitiers and Thierry of Chartres, would spend significant parts of their careers in the schools at Paris and Chartres and address a good deal of their work to academic audiences. The growth of these schools and others like them at centres such as Oxford, Bologna and Salerno signals a steady and rapid increase in the vitality of intellectual life in Western Europe. By the middle of the thirteenth century, the universities at Paris and Oxford were the leading centres of European philosophical activity. Virtually all the great philosophers from 1250 to 1350, including Albert the Great, Thomas Aquinas, Bonaventure, John Duns Scotus and William of Ockham, studied and taught in the schools at one or both of these centres. It is partly for this reason that early modern philosophers (who were typically not associated with universities) refer to their medieval predecessors in general as 'the schoolmen'.

The migration of philosophical activity to the universities meant not only the centralization of this activity but also its transformation into an increasingly formal and technical academic enterprise. Philosophical education was gradually expanded and standardized, philosophers themselves became highly trained academic specialists and philosophical literature came to presuppose in its audience both familiarity with the standard texts and issues of the university curriculum and facility with the technical apparatus (particularly the technical logical tools) of the discipline. These features of later medieval philosophy make it genuinely scholastic, that is, a product of the academic environment of the schools.

The philosophical disciplines narrowly construed - logic, natural philosophy, metaphysics and ethics - occupied the centre of the curriculum leading to the basic university degrees, the degrees of Bachelor and Master of Arts. Most of the great philosophers of this period, however, went beyond the arts curriculum to pursue advanced work in theology. The requirements for the degree of Master of Theology included study of the Bible, the Church Fathers and (beginning perhaps in the 1220s) Peter Lombard's *Sentences* (which was complete by 1158). Designed specifically for pedagogical purposes, the *Sentences* is rich in quotation and paraphrase from authoritative theological sources, surveying respected opinion on issues central to the Christian understanding of the world. From about 1250, all candidates for the degree of Master of Theology were required to lecture and produce a commentary on Lombard's text. This requirement offered a formal occasion for scholars nearing their intellectual maturity to develop and present their own positions on a wide variety of philosophical and theological issues guided (often only quite loosely) by the structure of Lombard's presentation.

By virtue of its historical circumstances, medieval philosophical method had from its beginnings consisted largely in commentary on a well defined and fairly small body of authoritative texts and reflection on a canonical set of issues raised by them. Philosophers in the era of the universities took for granted a much larger and more varied intellectual inheritance, but their approach to philosophical issues remained conditioned by an established textual tradition, and they continued to articulate their philosophical views in explicit dialogue with it. Formal commentary on standard texts flourished both as a pedagogical tool and as a literary form. However, other philosophical forms, including the disputation - the most distinctive philosophical form of the thirteenth and fourteenth centuries - were essentially dialectical. In the university environment, the disputation became a technical tool ideally suited to the pressing task of gathering together, organizing and adjudicating the various claims of a complex tradition of texts and positions.

A disputation identifies a specific philosophical or theological issue for discussion and provides the structure for an informed and reasoned judgment about it. In its basic form, a disputation presents, in order: (1) a succinct statement of the issue to be addressed, typically in the form of a question admitting of a 'yes' or 'no' answer; (2) two sets of preliminary arguments, one supporting an affirmative and the other a negative answer to the question; (3) a resolution or determination of the question, in which the master sets out and defends his own position, typically by drawing relevant distinctions, explaining subtle or potentially confusing points, or elaborating the underlying theoretical basis for his answer; and (4) a set of replies specifically addressing the preliminary arguments in disagreement with the master's stated views. A disputation's two sets of preliminary arguments allow for the gathering together of the most important relevant passages and arguments scattered throughout the authoritative literature. With the arguments on both sides of the question in hand, the master is then ideally positioned to deal with both the conceptual issues raised by the question and the hermeneutical problems presented by the historical tradition. Academic philosophers held disputations in their classrooms and at large university convocations, and they used the form for the literary expression of their ideas. Aquinas' *Summa theologiae*, the individual articles of which are pedagogically simplified disputations, is perhaps the most familiar example of its systematic use as a literary device. The prevalence of the disputational form in later medieval philosophy accounts for its being thought of as embodying 'the scholastic method' (see Language, medieval theories of; Logic, medieval).

(On the early Middle Ages (*circa* 600-1100), see Carolingian renaissance; Damian, P.; Encyclopedists, medieval; Eriugena, J.S.; Gerbert of Aurillac; John of Damascus.)

(On the twelfth-century philosophers, see Abelard, P.; Anselm of Canterbury; Bernard of Clairvaux; Bernard of Tours; Chartres, School of; Clarembald of Arras; Gerard of Cremona; Gilbert of Poitiers; Hildegard of Bingen; Hugh of St Victor; Isaac of Stella; John of Salisbury; Lombard, P.; Richard of St Victor; Roscelin of Compiègne; Thierry of Chartres; William of Champeaux; William of Conches.)

(On the thirteenth-century philosophers, see Albert the Great; Alexander of Hales; Aquinas, T.; Averroism; Bacon, R.; Boethius of Dacia; Bonaventure; David of Dinant; Grosseteste, R.; Henry of Ghent; Joachim of Fiore; John of La Rochelle; Kilwardby, R.; Neckham, A.; Olivi, P.J.; Pecham, J.; Peter of Spain; Philip the Chancellor; Pseudo-Grosseteste; Richard Rufus of Cornwall; Siger of Brabant; Thomas of York; Ulrich of Strasbourg; William of Auvergne; William of Auxerre; William of Sherwood.)

(On the fourteenth-century philosophers, see Albert of Saxony; Alighieri, Dante; Aureol, P.; Bradwardine, T.; Brinkley, R.; Brito, R.; Buridan, J.; Burley, W.; Chatton, W.; Crathorn, W.; Dietrich of Freiberg; Duns Scotus, J.; Durandus of St Pourçain; Francis of Meyronnes; Gerard of Odo; Giles of Rome; Godfrey of Fontaines; Gregory of Rimini; Henry of Harclay; Hervaeus Natalis; Heytesbury, W.; Holcot, R.; James of Viterbo; John of Jandun; John of Mirecourt; John of Paris; Kilvington, R.; Llull, R.; Marsilius of Inghen; Marsilius of Padua; Marston, R.; Matthew of Aquasparta; Meister Eckhart; Nicholas of Autrecourt; Oresme, N.; Oxford Calculators; Peter of Auvergne; Richard of Middleton; Suso, H.; Tauler, J.; Vital du Four; William of Ockham; Wodeham, A.; Wyclif, J.)

(On the fifteenth-century philosophers, see Ailly, Pierre d'; Denys the Carthusian; Gerson, J.; Hus, J.; Nicholas of Cusa; Paul of Venice; Thomas à Kempis.)

4 Doctrinal characteristics

At the most basic level, medieval philosophers share a common view of the world that underlies and supports the various specific developments that constitute medieval philosophy's rich detail.

Metaphysics. The common metaphysical ground of medieval philosophy holds that at the most general level reality can be divided into substances and accidents. Substances - Socrates and Browny the donkey are the stock examples - are independent existents and therefore ontologically fundamental. Corporeal substances (and perhaps also certain incorporeal substances) are constituted from matter and form (see Substance). Matter, which in itself is utterly devoid of structure, is the substrate for form (see Matter). Form provides a substance's structure or organization, thereby making a substance the kind of thing it is. Socrates' soul, for example, is the form that gives structure to Socrates' matter, constituting it as the living flesh and blood of a human body and making Socrates a particular human being. Accidents - Socrates' height, for example, or Browny's colour - are also a kind of form, but they take as their substrate not matter as such but a substance: Socrates or Browny. Accidents depend for their existence on substances and account for substances' ontologically derivative characteristics.

Medieval philosophers recognized matter and form, the fundamental constituents of corporeal substances, as fundamental explanatory principles. A thing's matter (or material cause) and its form (or formal cause) provide basic explanations of the thing's nature and behaviour. To these two principles they added two others, the agent (or efficient) cause and the end (or final cause). The agent cause is whatever initiates motion or change; the final cause is the goal or good toward which a particular activity, process, or change is directed.

Medieval philosophers disagreed about extensions and qualifications of this fundamental metaphysical view of the world. They debated, for example, whether incorporeal substances are like corporeal substances in being composed ultimately of matter and form, or whether they are subsistent immaterial forms. They also debated whether substances such as Socrates have just one substantial form (Socrates' rational soul) or many (one form constituting Socrates' body, another making him a living body with certain capacities for motion and cognition (an animal), and another making him a rational animal (a human being)). However, they never doubted the basic correctness of the metaphysical framework of substance and accidents, form and matter, nor are they in any doubt about whether the analytical tools that framework provides are applicable to philosophical problems generally.

(See Aristotelianism, medieval; Augustinianism; Eternity of the world, medieval views of; Platonism, medieval.)

Psychology and epistemology. Medieval philosophers understood the nature of human beings in terms of the metaphysics of form and matter, identifying the human rational soul, the seat of the capacities specific to human beings, with form. All medieval philosophers, therefore, held broadly dualist positions according to which the soul and body are fundamentally distinct. But only some were also substance dualists (or dualists in the Cartesian sense), holding in addition that the soul and body are themselves substances.

Medieval philosophers devote very little attention to what modern philosophers would recognize as the central questions of epistemology (see Epistemology, history of). Until very late in the period, they show little concern for sceptical worries and are not primarily interested in stating the necessary and sufficient conditions for the truth of the claim that some person knows a given proposition. For the most part they assume that we have knowledge of various sorts and focus instead on developing an account of the cognitive mechanisms by which we acquire it. They are especially interested in how we are able to acquire knowledge of universals and necessary truths - objects

or truths that are immaterial, eternal and unchanging - given that the world around us is populated with particular material objects subject to change. The answers medieval philosophers give to this question vary considerably, ranging from Platonistic accounts that appeal to our direct intellectual vision (with the aid of divine illumination) of independently existing immutable entities (such as ideas in the divine mind) to naturalistic accounts that appeal to cognitive capacities wholly contained in the human intellect itself that abstract universals from the data provided by sense perception (see Universals).

(See Aristotelianism, medieval; Augustinianism; Platonism, medieval.)

Ethics. Medieval philosophers share a generically Greek framework of ethical theory, extended and modified to accommodate Christianity. Its main features include an objectivist theory of value, a eudaimonistic account of the human good and a focus on the virtues as central to moral evaluation (see Eudaimonia; Areté; Virtues and vices). According to the metaphysics of goodness inherited by medieval philosophers from Greek thought, there is a necessary connection between goodness and being. Things are good to the extent to which they have being. Evil or badness is not a positive ontological feature of things but a privation or lack of being in some relevant respect. The ultimate human good or goal of human existence is happiness or beatitude, the perfection of which most medieval philosophers identified as supernatural union with God after this life. The ultimate human good is attained both through the cultivation of the moral virtues and through divine grace in the form of supernaturally infused states and dispositions such as faith, hope and charity, the so-called theological virtues (see Theological virtues).

Within this framework, medieval philosophers debated whether human beatitude is essentially an affective state (a kind of love for God) or a cognitive state (a kind of knowledge or vision of God), and whether the virtues are strictly necessary for the attainment of beatitude. They also debated whether the rightness or wrongness of some actions depends solely on God’s will. Contrary to caricatures of medieval ethics, no one unequivocally endorsed a divine command theory according to which the moral rightness (or wrongness) of *all* acts consists solely in their being approved (or disapproved) by God (see Voluntarism).

(See Aristotelianism, medieval; Augustinianism; Platonism, medieval.)

Logic and language. Medieval philosophers devote enormous attention - perhaps more attention than philosophers of any period in the history of philosophy apart from the twentieth century - to logic and philosophy of language. This phenomenon is explained primarily by the uniquely important role played by Aristotle’s logic in the development of medieval thought. Until the early twelfth century, medieval philosophers’ knowledge of Greek philosophy was restricted to a few texts of Aristotelian logic and, by default, those texts largely set the agenda for philosophical discussion. It is a passage from Porphyry’s *Isagōgē*, for example, that enticed first Boethius and, following him, a long line of commentators to take up the philosophical problem of universals (see Universals). The texts of the old logic, which remained a central part of the philosophy curriculum in the later Middle Ages, were eventually supplemented by the remaining treatises of Aristotle’s logic, among which the *Topics* and the *Sophistical Refutations* in particular sparked intense interest in the forms of philosophical argument and the nature of meaning.

(See Aristotelianism, medieval; Language, medieval theories of; Logic, medieval.)

Natural philosophy. Medieval philosophers believed that a complete account of reality must include an account of the fundamental constituents and principles of the natural realm. Their earliest reflections on these matters were inspired primarily by two ancient accounts of the origins and nature of the universe, the biblical story of creation (in Genesis) and Plato’s story of the Demiurge’s fashioning of the world (in the *Timaeus*) (see Plato). The confluence of these ancient sources produced a medieval tradition of speculative cosmological thought paradigmatically expressed in discussions of the six days of creation. This topic in particular gave medieval philosophers opportunity to reflect on the nature of the contents of the universe and the principles governing the created realm.

From the late twelfth century, medieval philosophy is profoundly affected by the new Aristotelian natural philosophy and the new scientific treatises by Islamic philosophers. Aristotle’s *Physics* in particular received enormous attention, and medieval philosophers developed sophisticated tools of logical, conceptual and mathematical analysis to deal with problems raised by Aristotle’s discussions of motion, change, continuity and infinity. Scientific treatises by Islamic thinkers such as Alkindi (see al-Kindi), Alpetragius, Avicenna (see Ibn

Sina) and Alhasen provided the material and impetus for significant developments in astronomy, medicine, mathematics and optics.

(See Aristotelianism, medieval; Eternity of the world, medieval views of; Language, medieval theories of; Natural philosophy, medieval; Oxford Calculators; Platonism, medieval.)

5 Philosophical theology

Christianity is not in itself a philosophical doctrine, but it profoundly influences the medieval philosophical world-view both from within philosophy and from outside it. On the one hand, Christian texts and doctrine provided rich subject matter for philosophical reflection, and the nature and central claims of Christianity forced medieval intellectuals to work out a comprehensive account of reality and to deal explicitly with deep issues about the aims and methods of the philosophical enterprise. In these ways, Christianity was taken up into philosophy, adding to its content and altering its structure and methods. On the other hand, Christianity imposed external constraints on medieval philosophy. At various times these constraints took institutional form in the official proscription of texts, the condemnation of philosophical positions and the censure of individuals.

Augustine laid the foundation for medieval Christian philosophical theology in two respects. First, he provided a theoretical rationale both for Christian intellectuals engaging in philosophical activity generally and for their taking Christian doctrine in particular as a subject of philosophical investigation. According to Augustine, Christian belief is not opposed to philosophy's pursuit of truth but is an invaluable supplement and aid to philosophy. With revealed truth in hand, Christian philosophers are able to salvage what is true and useful in pagan philosophy while repudiating what is false. Moreover, Augustine argued that Christianity can be strengthened and enriched by philosophy. Christian philosophers should begin by believing (on the authority of the Bible and the church) what Christianity professes and seek (by the use of reason) to acquire understanding of what they initially believed on authority. In seeking understanding, philosophers rely on that aspect of themselves - namely, reason - in virtue of which they most resemble God; and in gaining understanding, they strengthen the basis for Christian belief. The Augustinian method of belief seeking understanding is taken for granted by the vast majority of philosophers in the Middle Ages.

Second, Augustine's writings provide a wealth of rich and compelling examples of philosophical reflection on topics ranging from the nature of evil and sin to the nature of the Trinity. Boethius stands with Augustine in this respect as an important model for later thinkers. He composed several short theological treatises that consciously attempt to bring the tools of Aristotelian logic to bear on issues associated with doctrines of the Christian creed. Inspired by the philosophical analysis and argumentation prominent in these writings, medieval philosophers enthusiastically took up, developed and extended the enterprise of philosophical theology.

With the emergence of academic structure in the new European schools and universities of the twelfth and thirteenth centuries, theology became the paramount academic discipline in a formal curriculum of higher education. However, the fact that great thinkers of the later Middle Ages typically studied philosophy as preparatory for the higher calling of theology should not be taken to imply that in becoming theologians they left philosophy behind. As a simple matter of fact, later medieval theologians continued throughout their careers to address fundamental philosophical issues in fundamentally philosophical ways. And it is clear why this should be so: those who took up the study of theology were among the most gifted and highly trained philosophical minds of their day, and they brought to theology acute philosophical sensitivities, interests and skills. Moreover, insofar as they viewed Christianity as offering the basic framework for a comprehensive account of the world, they were naturally attracted to the broadly philosophical task of building on that framework, understanding its ramifications and resolving its difficulties.

Despite the dominance of the Augustinian view of the relation between Christianity and philosophy, religiously motivated resistance to philosophy in general and to the use of philosophical methods for understanding Christianity in particular emerges in different forms throughout the Middle Ages. In the twelfth century, some influential clerics saw the flourishing study of logic at Paris as a dangerous influence on theology and used ecclesiastical means to attack Peter Abelard and Gilbert of Poitiers. In the thirteenth century the new Aristotelian natural philosophy prompted another period of sustained ecclesiastical reaction. In 1210 and 1215 ecclesiastical authorities proscribed the teaching of Aristotle's natural philosophy at Paris, and in 1277 the Bishop of Paris

issued a condemnation of 219 articles covering a wide range of theological and philosophical topics. The condemnation seems largely to have been a reaction to the work of radical Averroistic interpreters of Aristotle. It is unclear how effective these actions were in suppressing the movements and doctrines they targeted.

(See Aristotelianism, medieval; Augustinianism; Averroism; Illumination; Natural theology.)

6 Scholarship in medieval philosophy

Contemporary study of medieval philosophy faces special obstacles. First, a large body of medieval philosophical and theological literature has survived in European libraries, but because many of these collections have not yet been fully catalogued, scholars do not yet have a complete picture of what primary source materials exist. Second, the primary sources themselves - in the form of handwritten texts and early printed editions - can typically be deciphered and read only by those with specialized paleographical skills. Only a very small portion of the known extant material has ever been published in modern editions of a sort that any reader of Latin could easily use. Third, an even smaller portion of the extant material has been translated into English (or any other modern language) or subjected to the sort of scholarly commentary and analysis that might open it up to a wider philosophical audience. For these reasons, scholarship in medieval philosophy is still in its early stages and remains a considerable distance from attaining the sort of authoritative and comprehensive view of its field now possessed by philosophical scholars of other historical periods with respect to their fields. For the foreseeable future, its progress will depend not only on the sort of philosophical and historical analysis constitutive of all scholarship in the history of philosophy but also on the sort of textual archeology necessary for recovering medieval philosophy’s primary texts.

See also: Ancient philosophy; Islamic philosophy; Neoplatonism; Religion, philosophy of; Renaissance philosophy; Islamic philosophy: transmission into Western Europe

SCOTT MacDONALD
NORMAN KRETZMANN

References and further reading

Armstrong, A.H. (ed.) (1967) *The Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press.(A useful survey of medieval philosophy’s historical antecedents, including the philosophical movements of late antiquity, and of the main figures of the medieval period through the beginning of the twelfth century.)

Dales, R. (1992) *The Intellectual Life of Western Europe in the Middle Ages*, Leiden: Brill.(A general historical introduction emphasizing philosophy and theology but giving some attention to broader intellectual culture, including literature and science.)

Dronke, P. (ed.) (1988) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press.(A collection of sixteen articles exploring both the background to the philosophy of the twelfth century and its innovations and major figures; contains a good bibliography and useful bio-bibliographies of twelfth-century philosophers.)

Gilson, E. (1955) *History of Christian Philosophy in the Middle Ages*, London: Sheed & Ward.(A classic one-volume compendium of medieval philosophy, outdated in some respects but still an enormously useful quick reference.)

Kretzmann, N., Kenny, A. and Pinborg, J. (eds) (1982) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(Forty-six articles providing a wide-ranging view of later medieval philosophy and significant recent developments in scholarship on the period; it gives special attention to the philosophy of the thirteenth and fourteenth centuries and to developments in logic and philosophy of language; contains a good bibliography and useful bio-bibliographies for later medieval philosophers.)

Marenbon, J. (1983) *Early Medieval Philosophy (480-1150): An Introduction*, London: Routledge, 2nd edn, 1988.(A readable and informative general introduction to the period.)

Marenbon, J. (1987) *Later Medieval Philosophy (1150-1350): An Introduction*, London: Routledge & Kegan Paul.(Part One provides a clear and succinct account of the pedagogical, institutional, and intellectual developments that shape later medieval philosophy; Part Two of the book is less successful.)

Medieval philosophy, Russian

The term 'philosophy' is itself highly problematic in the context of medieval Russia. Even in its most literal sense of love of learning, it was regarded with ambivalence, its devotees risking persecution. At the same time, Russia at any given point in the Middle Ages possessed what can best be described as a self-consciousness, a sense of its own destiny. Arising from the unusual circumstances of its Conversion in 988, this consciousness continues to draw heavily on Byzantium, with Russia at first in a dependent role but later, following Constantinople's fall, assuming that of the proud successor. The centrality of the Christian element to medieval Russian thought is underlined by the continuing significance both of the monastic movement and of its ancient cradle, Kiev, even as Moscow was being extolled as the Third Rome.

1 The pre-Mongol period

Access to literate culture having been gained relatively late and as part of the Conversion to Byzantine Christianity (988), those at the forefront of intellectual life in pre-Mongol Rus' were concerned primarily with formulaic imitation of Byzantium's Christian culture. This imitative trend, coupled with the emphasis on humility and rejection of worldly glory within Byzantine Christianity itself, resulted in 'philosophy' being used as a virtual synonym for, variously, the foreign, the dangerous and the heretical. While the Greek 'philosopher' who instructs Prince Vladimir, pre-Conversion, on Christian beliefs is presented by the Chronicle as essentially a positive figure, the charges of 'philosophy' levelled at metropolitan Klim Smoliatich in the mid-twelfth century carry clear condemnatory overtones. Similarly, though book learning directed to appropriate ends was regarded as a saintly virtue, parallel cases to that of Klim demonstrate that if taken to excess it was regarded with suspicion. The popularity enjoyed by the figure of the *iurodivyi*, or fool for Christ's sake, who deliberately practised folly in order to free himself from the world and its temptations, is testament to the honouring of humility above learning.

The ideology of rejection of worldly glory proved no barrier to princely propaganda programmes, which were able to base themselves on the new religion both concretely through sumptuous new churches - again copies of the Byzantine - and abstractly by presenting the prince as defender of the Truth, battling with God's enemies.

The spiritual life of the nation, meanwhile, was influenced perhaps above all by the monasteries, headed by the famous Caves Monastery in Kiev: these foundations held a near-monopoly on literacy, learning and contemplation. It was the monasteries with their scribes which acted as a filter for the Byzantine literary heritage, neatly discarding the secular almost in its entirety; and from the monasteries came forth writers, preachers and spiritual leaders. Resonant in their words from the start is a sense of the Rus', as God's chosen, prodigal people, having been called to Christianity at the eleventh hour: prominent examples are Ilarion's *Sermon on Law and Grace* (*c*.1049) and Nestor's *Life of Feodosii of the Caves* (*c*.1088). Feodosii in particular, being one of the earliest Russian saints, and peculiarly accessible in his humility, occupied a place of honour in the developing spiritual tradition.

2 After the Mongol invasions

The sense of Russia's special destiny remains constant through the medieval period. Under the Mongols, the Church as never before became the focus of national identity, and the monastic movement in particular gained unprecedented momentum, developing a number of peculiar tendencies which were to play an important part in the nation's consciousness. Of note were the trend towards withdrawal from the world for contemplative prayer, exemplified by Sergii of Radonezh (14th century), a close advisor to the prince; and the two linked ideologies of Hesychasm and Non-Possession, which developed from it a century later under Nil Sorskii. Opposed to the Non-Possessors was Iosif Volotskii, who defended the extensive estates which the Mongols' tolerant policies had permitted the monasteries. His theory of theocratic absolutism and of the divine origin of power helped him win the self-styled Tsar of all Rus', Ivan III, defeater of the Mongols, to his side.

The growing national consciousness was bolstered by the granting of autocephalous status to the Church (1448) and by the fall of Constantinople five years later. Where Russia had been effectively in the position of a cultural dependency, on the receiving end of a one-way flow of culture and religion, it now found itself the last repository of 'Orthodoxy'. Ivan demonstratively wed Sophia Paleologos, niece of the last Byzantine Emperor, thereby

returning her to Orthodoxy from the Catholicism which, in a final indignity, the Emperor had accepted just before Constantinople fell. The ideology of ‘Moscow the Third Rome’ made its entrance: perceiving itself the sacred home of the Orthodox faith, the city assumed the trappings of the Byzantine court and Ivan took on the role of protector of the faith which had belonged to Byzantium’s emperors. In a parallel move, the Bishop of Moscow was gradually raised to the status of patriarch.

Despite the glorification of Moscow, though, Kiev, both geographically and culturally closer to the West, remained an important centre of thought. In 1631 the Caves Monastery once more demonstrated its significance when Pëtr Mogila, its hegumen, founded what was to become the Academy of Kiev, offering for the first time in Russia schooling in the classics and Western scholastic theology and philosophy. The Academy in addition embarked on translations of philosophical and theological texts, thus making them available for the first time to the Russian readership. The foundation’s significance was to continue far beyond the Middle Ages. Once the Ukraine came under full control of Moscow under Aleksei Mikhailovich, this trend spread into Russia, leading to the reform of the church books which brought about the schism in Russian Orthodoxy, as well as to the founding of a parallel Academy in Moscow and later in the other important Russian cities. Likewise, it was the Kiev Academy which was to exert a major influence over the ideas of the eighteenth-century Hryhorii Skovoroda, who can perhaps be called the first Russian philosopher proper.

See also: Byzantine philosophy; Russian philosophy

CLAIRE FARRIMOND

References and further reading

Edie, J.M. *et al.* (eds) (1965) *Russian Philosophy*, vol. 1, Chicago: Quadrangle Books.(Short section on the beginnings of Russian philosophy.)

Fedotov, G.P. (1946, 1966) *The Russian Religious Mind*, Cambridge, MA: Harvard University Press, 2 vols. (Detailed exploration of the leading figures in medieval Russian Christianity, including Feodosii, Sergii and the fifteenth-century monastic polemicists.)

Fennell, J.L. (1995) *A History of the Russian Church to 1448*, London and New York: Longman.(Detailed account of the Church’s influence on all aspects of life in early Rus’.)

Franklin, S. (1992) ‘O "filosofii" i "filosofakh" v kievskoi Rusi’ (‘Philosophy’ and ‘philosophers’ in Kievan Rus’), in *Byzantinoslavica* 53, 74-196.(Exploration of contemporary use of these terms, of the Byzantine precedent and of their appropriateness in the Kievan Russian context.)

Franklin, S. and J. Shepard (1996) *The Emergence of Rus 750-1200*, London and New York: Longman.(Useful chapters on ideology and imagery, both religious and political, during the early period.)

Ilarion (*c*.1049) *Sermon on Law and Grace*; trans. S. Franklin, in *Sermons and Rhetoric of Kievan Rus’*, Cambridge, MA: Harvard University Press, 1991.(Ilarion’s eloquent analysis of Christianity and the significance of the conversion of Rus’.)

Nestor (*c*.1088) *Life of Feodosii of the Caves*; trans. P. Hollingsworth in *The Hagiography of Kievan Rus’*, Cambridge, MA: Harvard University Press, 1992.(One of the earliest *lives* to be produced in Rus’, and which points to the existence of distinctive tendencies in Russian spirituality from the beginning.)

Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, vol. 1, ch. 1; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, vol. 1, ch. 1.(Outlines the main trends in thought up to Peter the Great’s time.)

Megarian school

The Megarians were a Greek 'Socratic' school of the fourth and early third centuries BC*. After their founder Euclides, whose main doctrine was the unity of the good, the leading Megarian was Stilpo, best known for preaching the self-sufficiency of virtue. They propounded various puzzles and found objections to the possible-actual distinction, the copula and universals.*

Euclides, an intimate associate of Socrates and author of Socratic dialogues, founded the Megarian (more correctly 'Megaric') school in his native Megara, a little west of Athens. His pupils included Eubulides of Miletus, who like his own pupil Alexinus of Elis was a noted propounder of paradoxes. It is not certain that these (or indeed *any* non-natives of Megara) were ever members of the Megarian school. And some other philosophical descendants of Euclides widely assumed to be Megarians, including the celebrated logician Diodorus Cronus from Iasos, in fact classified themselves as 'Dialecticians' (see Dialectical school). After Euclides, the most eminent Megarian was Stilpo of Megara, school head in the late fourth and early third centuries BC and a teacher of the founding Stoic Zeno of Citium. Only scattered evidence of their work survives.

Doctrinally, the chief focus of the Megarians was ethical and metaphysical. Despite their interest in paradoxes, there is no evidence that they shared the Dialecticians' concern with logical theory. They were primarily Socratics, but with some debt to the Eleatic Parmenides too. Euclides held that 'the good is one thing, called by many names: sometimes wisdom, sometimes god, and at other times intellect etc.', and denied the existence of its supposed opposites (Diogenes Laertius II 106). This builds on Socrates' ethical doctrine of the unity of the virtues (see Socrates §5; Stoicism §16), to which the Megarians subscribed, but may also be influenced by Parmenides' metaphysical thesis that all names, even those supposedly opposite to each other, in reality refer to a single being (see Parmenides §7). The upshot seems to be a monism of value (possibly even entailing that the world is entirely good), but not the Eleatic metaphysics that some have attributed to the school.

That they were not considered Eleatic by their contemporaries is clear from Aristotle's treatment of their modal doctrine: only what is actual is possible, for example, only when he is building is a builder capable of building. Aristotle objects (*Metaphysics* IX 3) that this would eliminate all change - an Eleatic-sounding outcome which he implies they would not welcome.

Stilpo resembled the Cynics in his insistence on the autonomy of human good. Asked whether he had lost anything in the sack of Megara, he gave the famous reply that he had lost nothing that was his - meaning that moral and intellectual attainments are inalienable. He raised objections to the copula, dismissing all '*x* is *y*' statements as false identity-statements. An argument of his against Platonic Forms suggests a similar worry: you cannot say 'This is a vegetable', because vegetable (the universal) existed thousands of years ago, whereas this did not.

'Megarian questionings' acquired the reputation of sophistic arguments. Nevertheless, they may have included the important logical puzzles propounded by Eubulides, such as the sorites ('When does a collection of grains become a heap?'), the horned argument ('Have you lost your horns?'), the veiled argument ('You don't know the hooded man? He is your father. So you don't know your father.') and the liar paradox ('Is "I am lying" true or false?').

See also: Socratic schools

DAVID SEDLEY

References and further reading

Aristotle (*c.* mid 4th century BC) *Metaphysics* IX, in J. Barnes (trans. and ed.) *The Complete Works of Aristotle*, vol. 2, Princeton, NJ: Princeton University Press, 1984.(Chapters 3- 4 include a critique of the Megarians on possibility.)

Diogenes Laertius (*c.* early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(Book II 106-20 in volume 1 is a history of the Megarian school.)

Döring, K. (1972) *Die Megariker*, Amsterdam: Gruner.(Collection of testimonies with useful German commentary.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae (Fragments of Socrates and the Socratics)*, Naples:

Bibliopolis.(Volume 1 includes the fullest available collection of testimonies on the Megarians.)

Makin, S. (1996) 'Megarian Possibilities', *Philosophical Studies* 83: 253-76.(Brilliant defence of the Megarians' paradoxical identification of the possible with the actual.)

Zeller, E. (1869-82) *Die Philosophie der Griechen (Greek Philosophy)*, Leipzig: Reisland, 3rd edn; the relevant part is available in English trans. by O.J. Reichel, *Socrates and the Socratic Schools*, Leipzig: Fues' Verlag (Reisland), 3rd edn, 1885. (Pages 250-80 of the translated version are sadly still the best general account of the Megarians available in English.)

Meinecke, Friedrich (1862-1954)

Friedrich Meinecke was a German historian of moral and political ideas who addressed the problems of his age through critical study of the writings of past thinkers. He studied at the universities of Berlin and Bonn, before becoming a state archivist in Berlin for fourteen years. He went on to hold professorships at Strasbourg, Bonn and Berlin. He contributed to political thought through books on cosmopolitanism and nationalism, on the morality of states and on historicism. In these works he re-examined the European tradition of moral and political thought, attempting especially to find a middle way between the universalist and cosmopolitan ethic of the Enlightenment and the nineteenth-century belief in the particularity and historical uniqueness of national cultures. His life spanned the entire period from Bismarck to Hitler, and he sought to bring his deep historical learning to bear on the ethical dilemmas facing Germans during this time.

Friedrich Meinecke was the last great representative of the German Historical School, which began with Niebuhr and Savigny and enjoyed a prolonged flowering in such figures as Ranke, Mommsen, Treitschke, Sybel, Droysen and, in Meinecke’s own day, Otto Hinze, Ernst Troeltsch, Erich Marcks and Hans Delbrueck (and, although more loosely associated with it, Max Weber). His more remote, but none the less omnipresent, intellectual ancestors were Goethe, Herder, and Wilhelm von Humboldt. His awareness of a deep affinity with the great Swiss cultural historian, Jacob Burckhardt, whose profundity of vision and melancholy pessimism contrasted with the more uncomplicated, optimistic outlook of Ranke and the Prussians, only grew with time. He virtually created the history of ideas in Germany as a separate, highly sophisticated and self-conscious discipline bestraddling both history and philosophy; and, although an individualist who respected individuals, displaying no apparent desire to be seen as the founder of a school (his desire for power in this as other matters was practically nonexistent), he nevertheless became the undisputed doyen of several generations of German historical scholars. His many US pupils, as well as the scholar refugees who fled the Third Reich, sowed the seeds of his historical approach in the USA. Never in the political limelight himself, Meinecke did informally advise politicians and public figures, particularly during the Weimar Republic, but his greatest contribution to political thought and practice is to be found in his three major works.

Weltbürgertum und Nationalstaat (Cosmopolitanism and the National State) (1907) was written in that period of liberal bourgeois optimism before the First World War which tended to see a harmony between intellect and power embodied in the state. To uncover the roots of the modern age was a central aim of German historians at this time: as Max Weber had discovered the source of the capitalist world in Calvinism, and Ernst Troeltsch had traced the links between modern culture and Protestantism, so Meinecke went in search of the origins of modern national consciousness. In a series of loosely connected individual studies, he submitted the views of the men and movements that influenced and finally brought about German unification to minute and sensitive analysis. From the cosmopolitan Enlightenment figure of Wilhelm von Humboldt, through early Romantics such as Novalis and Friedrich Schlegel, to the theorists of the state, Fichte and Adam Müller, and the practical reformers, vom Stein and Gneisenau, down to the later Hegel, Ranke and Bismarck, Meinecke sketched forth the gradual transition from German eighteenth-century cosmopolitanism to nineteenth-century nationalist particularism. Elements of the older, universal, humanistic outlook lingered on in the new ideology, but a novel world not wholly favourable to them had been born. Yet, at this stage Meinecke still hoped, rather idealistically, that the old values and the new could co-exist in harmony. Although, in the words of one sympathetic German critic, this book moves ‘in the rarefied air of the highest intellectual regions’, it is considered by many to be Meinecke’s greatest work, and it had a permanent impact on subsequent German historiography.

Meinecke’s second great work was *Die Idee der Staatsräson in der neueren Geschichte (The Doctrine of Raison d’état in the Modern Age)* (1925). Written in the aftermath of the First World War and the collapse of the Hohenzollern monarchy, this is a magisterial survey of the conflict between ethics and politics as exemplified by thinkers and men of action ranging from Machiavelli, Richelieu, and Montesquieu, to Frederick the Great, Hegel, Bismarck and Treitschke. Bringing immense insight and subtlety to the analysis of ideologies, Meinecke presents the history of political thought and action - in Germany at least - as a long chain of triumphs of the naked will to power in the service of the state over common morality, whether Christian or Kantian, and arrives at the bitter insight that, ‘[t]he state, so it seems, must violate the moral law’ (1925 (1957)). The optimistic idealism of

Weltbürgertum und Nationalstaat yields to a sombre, unsentimental recognition of the amorality and cynicism of governments. Nevertheless, Meinecke still insists that wielders of public power should act as guardians and promoters of a common ethical code, and that '[t]he state must become moral and strive for harmony with the universal moral law' (1925 (1957)). A view of history emerges which seeks to avoid the twin extremes, on the one hand, of the positivistic Enlightenment optimism of progress, Hegelian panlogism and Rankean faith in Providence, and, on the other hand, the yawning chasm of Schopenhauerian historical pessimism. The result is a clash between historical experience and the ethical imperative. Yet this tragic view does not exclude the possibility of genuine personal and historical greatness; rather, it requires it and makes it possible. For Meinecke the fact that historical man is constantly called upon to be a hero or an actor in a tragedy is what gives history its universal import and significance.

With the advent of national socialism, Meinecke was dismissed as 'an historical epigone' and divested of his chair and his editorship of *Historische Zeitschrift*. His last major work, *Die Entstehung des Historismus (The Rise of Historicism)* (1936), represents a vast stock-taking. Here Meinecke reaches out beyond the German cultural sphere to embrace Italian, French, and English figures in a grand survey of the emergence of modern historical consciousness itself. Examining such figures as Shaftesbury, Vico, Montesquieu, Herder and Goethe, he traces the profound shift from the static, rationalist, universalist presuppositions of the Renaissance, the Age of Reason, and the Enlightenment, themselves rooted in the much older tradition of Natural Law, to the new historicist, relativist insights into the nature of man and human history that mark our own era. The generalizing, typifying methodology which the older classical tradition shares with the sciences gives way increasingly to the genetic approach with its emphasis on origins, development, organic growth, and on the role of unique creative individuals and groups - nations, cultures, churches, peoples, races, and so on. He thus unearthed the roots of much that came later, including Ranke and the Historical School itself, Romanticism, and twentieth-century currents of relativism and irrationalism.

Die Entstehung des Historismus aimed above all to secure the autonomy of the historical realm and to protect it especially against 'scientistic' incursions, not least by figures such as Spengler with his crude and simplistic biological categories. History begins with creative human freedom 'where the spontaneous factor of human beings acting in the light of values intervenes decisively and thus creates something unique and singular' (1936 (1972)). Not only states and historical epochs are individualities on Meinecke's view, but also ideas and ideals. These are to be studied not only for their causal efficacy, but most of all for their intrinsic value and power to illuminate. They must be infused 'with as much life-blood' as the historian has to give them: 'The essence of history is not the political battle for power but the creation of lasting cultural values' (1936 (1972)).

In *Die Deutsche Katastrophe (The German Catastrophe)* (1946), Meinecke offers a brilliant and profound analysis of Germany's moral and political collapse. Among the ruins of the Bismarck-Hitler era, a path must be found back to the Age of Goethe. Meinecke's last book, *Schaffender Spiegel - Studien zur Deutschen Geschichtsschreibung und Geschichtsauffassung (The Creative Mirror: Studies in German Historiography)* (1948), contains studies on various aspects of 200 years of German historiography and is notable for his agonized reflections on the intractable problems of relativism thrown up by the historicist outlook he did so much to encourage.

Meinecke has often been accused of lacking sharp conceptual clarity and of ignoring material and economic factors in history. Particularly critical are Marxist historians. Yet none of this detracts from his unique achievement in unearthing the genesis and nature of the hitherto unexamined intellectual structures that underlie political thought and action in the modern world, and in describing them with a depth and breadth of minutely detailed learning, and a refined sense of nuance and atmosphere, probably unparalleled in his time or since.

See also: Historicism

ROGER HAUSHEER

List of works

Meinecke, F. (1907) *Weltbürgertum und Nationalstaat (Cosmopolitanism and the National State)*, trans. R.B. Kimber, Princeton, NJ: Princeton University Press, 1970.

Meinecke, F. (1925) *Die Idee der Staatsräson in der neueren Geschichte (Machiavellism: The Doctrine of Raison d'État and its Place in Modern History)*, trans. D. Scott, London: Routledge & Kegan Paul, 1957.

Meinecke, F. (1936) *Die Enstehung des Historismus (Historism: The Rise of a New Historical Outlook)*, trans.

J.E. Anderson, London: Routledge & Kegan Paul, 1972.(Useful foreword by Isaiah Berlin and introduction by Carl Hinrichs.)

Meinecke, F. (1946) *Die Deutsche Katastrophe: Betrachtungen und Erinerungen (The German Catastrophe: Reflections and Recollections)*, London: 1952.

Meinecke, F. (1948) *Schaffender Spiegel - Studien zur deutschen Geschichtsschreibung und Geschichtsauffassung (The Creative Mirror: Studies in German Historiography)*, Stuttgart: Koehler.

References and further reading

Dehio, L. (1953) *Friedrich Meinecke: Der Historiker in der Krise (Friedrich Meinecke: Historian in Crisis)*, Berlin: Dahlem.

Hofer, W. (1950) *Geschichtschreibung und Weltanschauung: Betrachtungen zum Werk Friedrich Meineckes (Historiography and World-View: An Examination of the Work of Friedrich Meinecke)*, Munich: Oldenbourg.

Meinong, Alexius (1853-1920)

Meinong was an Austrian philosopher and psychologist who taught at the University of Graz. He contributed substantially to psychology, epistemology, value theory, ethics and probability theory, but is best known for his theory of objects, in which he advocates the radical view that there are objects which are wholly outside being, including impossible objects. Meinong influencedRussell and the American 'new realists'. Though widely rejected, his views have proved difficult to refute decisively and he has found sympathetic support from a number of logicians and philosophers.

1 Life, works and development

Alexius Meinong studied history at the University of Vienna, where he came under the influence of Franz Brentano, and turned to philosophy and psychology. His earliest works, the *Hume Studien* (Hume Studies) of 1877 and 1882, the first being his *Habilitationsschrift*, examined the nominalism of Locke and Hume and their theories of relations.Meinong thus grew up in the British philosophical tradition, rather than the German. After a short period as *Privatdozent* in Vienna, he moved in 1882 to Graz, where he remained until his death. In 1894 he established Austria's first psychology laboratory. A devoted teacher, he acquired talented students, including Ernst Mally, the psychologist Vittorio Benussi, andStephan Witasek, whose early deathMeinong described as the greatest blow of his life. Meinong's eyesight deteriorated early in life and he preferred to remain in Graz rather than accept a professorship in Vienna. Married and with a son, he was a competent violinist and composed *Lieder*.

Meinong's views developed slowly. Until the turn of the century his main work was in epistemology, psychology and the theory of value. Under the influence of Ehrenfels and Twardowski he turned to ontological concerns. *Über Annahmen (On Assumptions)* (1902, 2nd edn 1910) first introduced objectives, special objects of judgment, while the idea of a science of objects outside being occurs in the programmatic essay 'Über Gegenstandstheorie' ('On the Theory of Objects') (1904). These works attracted the notice of Bertrand Russell, and a controversy developed over the question whether it was logically consistent to assume impossible objects. Meinong responded to Russell's criticisms with refinements, but did not relinquish his position. His ontological views reached their final form with the publication of *Über Möglichkeit und Wahrscheinlichkeit (On Possibility and Probability)* (1915) and *Über emotionale Präsentation (On Emotional Presentation)* (1917). Two volumes of collected essays were published during his lifetime, and an eight-volume *Gesamtausgabe (Complete Edition)* appeared from 1968 to 78.

Meinong's style is clear but ponderous, and few of his many terminological coinages have passed into German philosophical currency. His views have been largely neglected in the German-speaking world, finding more resonance, both positive and critical, in English-language philosophy.

2 Intentionality and objects

FollowingBrentano, Meinong viewed the intentionality of the mental as its defining characteristic, but whereas Brentano originally regarded objects of consciousness as immanent, by 1890 Meinong was distinguishing an immanent from an external sense of 'object' (see Intentionality). FollowingTwardowski, Meinong then reserved the term 'object' (*Gegenstand*) for the external referent, denoting the immanent aspect by 'content' (*Inhalt*). He adoptedTwardowski's view that all mental acts have objects, but that some such objects are nonexistent. This was the basis for the development of the theory of objects outside being. Meinong called the relation between object and subject 'presentation' (*Präsentation*).

Meinong's classification of mental acts also deviated from that of Brentano. He distinguished ideas (*Vorstellungen*), thoughts, feelings and desires. Whereas for Brentano all thoughts are judgments, Meinong added conjectures (*Vermutungen*) and assumptions (*Annahmen*), which may be playful or hypothetical: like judgments, they are positive or negative and may be true or false, but, like ideas, they do not involve intellectual commitment. This distinction between 'serious' and 'phantastic' acts was extended into the emotive and conative spheres. Each kind of act, whether serious or phantastic, came to have its own characteristic kind of object: ideas correspond to things (*Objekte*), for example, Graz, greenness; thoughts correspond to objectives (*Objektive*), for example, that Graz is in Styria; feelings or emotions correspond to dignitatives, such as the beauty of Graz; and desires correspond to desideratives, for instance, that Graz should have less traffic. Things and dignitatives are

designatable by names whereas objectives and desideratives are designatable by clauses.

Meinong's rich ontology is thus the joint product of a unified conception of intentionality as presenting independent objects, and a rich taxonomy of the mental.

3 Objects outside being

Practical life enjoins attention to things around us, but in thought we range beyond what Meinong called the 'actual' or existent. Originally inclined to nominalism, Meinong progressively enlarged his ontology and came to decry the 'prejudice in favour of the actual'. Objectives, which resemble both 'states of affairs' and Russell's 'propositions', divide into those that are 'factual' or obtain, for example, that Graz is in Austria, and those that are 'unfactual' or do not obtain, for example, that Graz is in Slovenia. Meinong called 'true' only those factual objectives which someone grasps as factual. He regarded obtaining (*bestehen*) as a kind of timeless being: unfactual objectives are outside being (*außerseiend*). Similarly there exists no golden mountain, but the golden mountain is just as golden and mountainous as any existent: *what* and *how* a thing is is independent of *whether* it is. The converse does not hold: impossible objects, such as the round square, cannot exist.

Possible but non-actual objects had been accepted before, but Meinong's impossible objects were a largely new departure. Meinong recognized this and proposed a new, existentially uncommitted (*daseinfrei*) science, the theory of objects. It was impossible objects to which Russell took exception: his theory of descriptions aimed to dispense with all non-beings (see Russell, B. §9). Russell claimed thatMeinong's impossible objects offended against the law of contradiction. Meinong agreed but was unperturbed: his understanding of 'law of contradiction' derived from traditional logic, which said that a thing cannot both have and lack a certain property, whereas Russell meant that a proposition and its contradictory could both be inferred. Russell tried again: the *existent* round square is existent, round and square, and so it both exists and does not. Meinong replied that being existent is part of the existent round square's nature, whereas existing is not: being existent does not entail existing. Russell saw no difference and dropped the matter, satisfied he was right, but Meinong's view is in the tradition ofKant. Existence is not a (normal) property, though it has a 'watered-down' counterpart, being-existent. Meinong divided properties generally into two kinds, 'nuclear' ones which are part of the nature of a thing, and 'extra-nuclear' ones, such as existing, being simple or being complete, which are not.

Paradoxical objects such as Russell's set of all non-self-membered sets were harder to accommodate.Meinong called them 'defective objects', but could not decide whether they were just very deviant objects or whether there were none at all.

4 Incomplete objects and modality

'Incomplete' objects lack being for a different reason: they are undetermined one way or the other with respect to at least one property. The golden mountain, for instance, is undetermined with respect to its height. Meinong pressed incomplete objects into service as the senses of predicates and as epistemological intermediaries: in seeing a table I do not see all its characteristics, so the incomplete object corresponding to my perception is an 'auxiliary' object giving me access to the 'target' object, the table itself. An incomplete object with its suite of properties has a parasitic or 'implexive' being in the complete objects which have those properties.

In Meinong's theory of objective possibility and probability, incomplete objects are where possibility is 'at home'. Suppose I am about to throw a (fair) die. The actual throw will be a three, or not a three, but my thought that there is a one in six chance that I will throw a three has an objective foundation in that incomplete object, 'my next throw of the die'. This is subject of the objective 'that my next throw of the die will be a three', which is neither factual, nor unfactual, but 'subfactual'. It has a degree of factuality of 1/6, properly between 1 (factual) and 0 (unfactual). Meinong thus reconciled classical determinism with an objective foundation for probability statements, and showed the possibility of many-valued logic, which was later developed by Jan Łukasiewicz, who attended some of his lectures (see Łukasiewicz, J. §3).

Meinong considers necessity and impossibility to attach to objectives, not because of factuality or unfactuality in all possible worlds, but because their factuality and unfactuality are, he says, 'inhesive' to them. Notably, Meinong never considers God or any other supposedly necessary being.

5 Psychology and epistemology

Meinong carried out rudimentary classroom demonstrations in Vienna and Graz, and from this activity the Graz laboratory arose. His psychological studies were largely in the service of his epistemological interests, and his contributions were more empirical than those of Brentano. He examined among other things the psychology of abstraction, colour and Gestalt perception, phantasy, eye-movement, the feelings accompanying judgments, and sensory fatigue. His analysis of the role of indirect measurement in Weber's psychophysical law was noted by Russell. *On Assumptions* aimed originally at establishing the classificatory proposition that assumptions are different from both judgments and ideas; only in the course of writing did Meinong discover that he needed to consider objectives, the objects of judgment and assumption, for themselves. As object theory came to dominate Meinong's thinking, he left psychological experimentation to others and, like Husserl, rejected undue philosophical reliance on psychology, while allowing it an honourable role. The 1912 essay 'Für die Psychologie und gegen den Psychologismus in der allgemeinen Werttheorie' (For Psychology and Against Psychologism in General Value Theory) shows him striking this balance.

Meinong published several monographs and papers on epistemology, including studies on memory, conjectural conviction, the experiential basis of knowledge and an attempt to prove universal causality by inductive reasoning. The admission of conjectural conviction (*Vermutungsevidenz*), which, by contrast with Brentano's *Evidenz*, is fallible and may be inductively improved, is part of Meinong's general concern to describe those supposedly defective acts of mind often neglected by epistemologists. Our knowledge of the external world is for Meinong conjecturally, not apodictically certain. We have no reason to doubt it, but it is a posteriori and we might conceivably be mistaken. Meinong's work in epistemology has been less influential than in other areas, perhaps because he was never tempted by racy positions such as scepticism or idealism.

6 Value theory and ethics

In Vienna Meinong attended economics lectures by Carl Menger, and his own value theory attempts to extend the theory of value beyond economics, on the basis of our valuing experiences. Rejecting Brentano's assimilation of feeling to desire and will, he considered feeling or emotion to be the predominant source of value. Early lectures in Vienna, systematized in his *Psychologisch-ethische Untersuchungen zur Werth-Theorie (Psychological-ethical Investigations in Value Theory)* (1894) prompted his student Ehrenfels to counter with a subjectivistic desire-based theory, and in their constructive and friendly disagreement they conceded that value might attach *derivatively* to desires (Meinong) or feelings (Ehrenfels).

Meinong's views on value, like those on cognition, became progressively more objectivistic, and in his last works he staunchly upheld what he called 'impersonal' values, which we can recognize but do not create. Meinong's patient dissection of the principles governing values and valuation paved the way for Ernst Mally's first excursion into deontic logic.

7 Philosophy as a discipline

Meinong regarded philosophy as a congeries of disciplines rather than a single science, and his painstaking piecemeal method of philosophizing 'from below' contrasted with bolder systems and programmes. He described himself as an empirical thinker, but not an empiricist, and just this willingness to be led by the data was what recommended him to the young Russell. Nevertheless, his 'empirical thinking' led him to revolutionary conclusions, and towards the end of his life he can be seen connecting the pieces together. Meinong considered *On Emotional Presentation* to be his most important work, spanning psychology, epistemology, object theory and value theory in a single arch.

PETER SIMONS

List of works

Meinong, A. (1968-78) *Alexius Meinong Gesamtausgabe (Complete Edition)*, 7 vols, and *Ergänzungsband (Supplementary Volume)*, Graz: Akademische Druck- und Verlagsanstalt.(Contains all the books and essays that Meinong published, together with his intellectual self-portrait (*Selbstdarstellung*) and major reviews. Volume 7 contains a comprehensive index. The supplementary volume contains a selection of posthumous material including lecture notes, notes on other philosophers' writings, and a lexicon of logic and

epistemology.)
Meinong, A. (1877) ‘Hume-Studien I. Zur Geschichte und Kritik des modernen Nominalismus’, *Sitzungsberichte der philosophisch-historischen Klasse der Kaiserlichen Akademie der Wissenschaften (Wien)* 87: 187-260; published separately, Vienna: Gerold; repr. in *Gesamtausgabe*, 1: 1-72; trans. K.F. Barber, in *Meinong's Hume Studies. Translation and Commentary*, Ann Arbor: University Microfilms, 1966, 98-192.(Distinguishes extreme (linguistic) from moderate (conceptualistic) nominalism and upholds the latter.)
Meinong, A. (1882) ‘Hume-Studien II. Zur Relationstheorie’, *Sitzungsberichte der philosophisch-historischen Klasse der Kaiserlichen Akademie der Wissenschaften (Wien)* 101: 572-752; published separately, Vienna: Gerold; repr. in *Gesamtausgabe*, vol. 1; partially trans. K.F. Barber, in *Meinong's Hume Studies. Translation and Commentary*, Ann Arbor: University Microfilms, 1966, 194-229.(An early investigation of the ontology and epistemology of relations, with reference to Hume.)
Meinong, A. (1894) *Psychologisch-ethische Untersuchungen zur Wert-theorie* (Psychological-ethical Investigations in Value Theory), Graz: Leuschner & Lubensky; repr. in *Gesamtausgabe*, vol. 3, 1-244.(An ethics and value theory based on feelings as the primary bearers of value.)
Meinong, A. (1902) *Über Annahmen*, Leipzig: Barth, 2nd, revised edn, 1910; repr. in *Gesamtausgabe*, vol. 4; trans. and ed. J. Heanue, *On Assumptions*, Berkeley, CA: University of California Press, 1983.(The theory of assumptions and objectives. Probably Meinong's most influential work.)
Meinong, A. (1904) ‘Über Gegenstandstheorie’, in A. Meinong (ed), *Untersuchungen zur Gegenstandstheorie und Psychologie*, Leipzig: Barth, 1-50; trans. and ed. R.M. Chisholm, ‘The Theory of Objects’, in *Realism and the Background of Phenomenology*, Atascadero, CA: Ridgeview, 1982, 76-117.(The programme for the theory of objects. The whole 1904 collection, a showcase for the Graz School, was reviewed by Russell (1905).)
Meinong, A. (1912) ‘Für die Psychologie und gegen den Psychologismus in der allgemeinen Werttheorie’ (For Psychology and Against Psychologism in General Value Theory), *Logos* 3: 1-14; repr. in *Gesamtausgabe*, vol. 3, 267-82.(While values are not all psychological or ‘personal’, psychology is of value to value theory.)
Meinong, A. (1915) *Über Möglichkeit und Wahrscheinlichkeit. Beiträgezur Gegenstandstheorie und Erkenntnistheorie* (On Possibility and Probability), Leipzig: Barth; repr. in *Gesamtausgabe*, vol. 6.(Part 1 deals with ‘objective’ probability, understood in terms of incomplete objects and subfactual objectives. Part 2 concerns our subjective awareness of probability. Meinong's most mature object-theoretic work.)
Meinong, A. (1917) ‘Über emotionale Präsentation’, *Sitzungsberichte der Akademie der Wissenschaften in Wien, philosophisch-historische Klasse* vol. 183 (appeared in 1924); also published separately, Vienna: Holder,1917; repr. in *Gesamtausgabe*, vol. 3, 283-476; trans. and ed. M.-L. Schubert Kalsi, *On Emotional Presentation*, Evanston, IL: Northwestern University Press, 1972.(Emotions present independent value-bearers in much the same way as perception and thought present independent objects. Includes an inconclusive discussion of paradoxical (‘defective’) objects.)
Meinong, A. (1921) ‘A. Meinong [Selbstdarstellung]’, Leipzig: Meiner, vol. 1, 91-150; 2nd edn, 1923, 101-160; repr. in *Gesamtausgabe*, vol. 7, 1-62; biography and section on theory of objects trans. R. Grossmann, in *Meinong*, London: Routledge & Kegan Paul, 224-36.(Short but revealing account by Meinong of his own life, works and views, written shortly before his death.)
Meinong, A. (1965) *Philosophenbriefe. Aus der wissenschaftlichen Korrespondenz von Alexius Meinong (Philosophical Correspondence)*, ed. R. Kindinger, Graz: Akademische Druck- und Verlagsanstalt.(A selection of 43 letters to philosophers written between 1882 and 1920. Includes correspondence withRussell and Husserl.)

References and further reading

Barber, K.F. (1966) *Meinong's Hume Studies. Translation and Commentary*, Ph.D. dissertation, University of Iowa; Ann Arbor, MI: University Microfilms.(A study ofMeinong's earliest works.)
Findlay, J.N. (1963) *Meinong's Theory of Objects and Values*, Oxford: Clarendon Press.(A marvel of sympathetic exposition:Meinong's views in concise form.)
Grossmann, R. (1974) *Meinong*, London: Routledge & Kegan Paul.(Critically examines Meinong's ontology and epistemology, emphasizing his development.)
Lambert, K. (1983) *Meinong and the Principle of Independence*, Cambridge: Cambridge University Press. (Examines, from the standpoint of modern logic, the principle that how an object is is independent of whether it is.)
Routley, R. (1980) *Exploring Meinong's Jungle and Beyond*, Canberra: Australian National University. (Compendious investigations of the theory of objects (items) by an adherent more radical than Meinong himself.)

Russell, B. (1904) ‘Meinong’s Theory of Complexes and Assumptions’, *Mind*, new series 13: 204-19, 336-54, 509-24; repr. in B. Russell *Essays in Analysis*, ed. D. Lackey, London: Allen & Unwin, 1974, 21-76. (Russell’s longest and most appreciative review of Meinong, notable in particular for its praise of his method.)

(1905) ‘A. Meinong, *Untersuchungen zur Gegenstandstheorie und Psychologie*’, *Mind* new series 14: 530-9; repr. in B. Russell *Essays in Analysis*, ed. D. Lackey, London: Allen & Unwin, 1974, 77-88.(This book review containsRussell’s first sally againstMeinong’s impossible objects.)

(1907) ‘A. Meinong, *Über die Stellung der Gegenstandstheorie im System der Wissenschaften*’, *Mind* new series 16: 436-9; repr. in B. Russell, *Essays in Analysis*, ed. D. Lackey, London: Allen & Unwin, 1974, 89-93. (Another book review, andRussell’s last word on impossible objects.)

Meister Eckhart (*c.*1260-1327/8)

More than any other medieval thinker, Eckhart has received widely divergent interpretations. The controversies stem from the fact that his writings fall into two distinct groups, works written in the vernacular and works written in Latin. The German writings, which were intended for a wide audience, established Eckhart's long-standing fame as a mystic. Another, more academic Eckhart emerged when his Latin work was rediscovered in 1886. The study of Eckhart's thought today centres on the unity of the scholastic (Latin) and the popular (German) work.

Eckhart was born about 1260 at Hochheim, Germany, and died in 1327 (or early 1328), probably at Avignon. After entering the Dominican Order, he studied theology at the *studium generale* of the Order in Cologne. In 1302-3 he was Master of Theology at the University of Paris. In the next years he held executive functions within his Order in Germany. A token of the high esteem Eckhart enjoyed is that in 1311 he occupied the Dominican chair of theology at Paris for a second time (an honour that also fell to Thomas Aquinas). Thereafter he was the spiritual master of religious communities and a teacher in Cologne. In 1326 the archbishop of that city started an inquiry into his doctrines on suspicion of heresy. Eckhart appealed directly to the pope in Avignon, but died before the conclusion of the trial. In the bull *In agro dominico* of 27 March 27 1329, Pope John XXII condemned twenty-eight propositions taken from his works and sermons.

Eckhart's main Latin work is the *Opus tripartitum*. As the title indicates, it was intended to consist of three parts: the *Opus propositionum (Work of Propositions)* in which a thousand propositions were to be explained; the *Opus quaestionum (Work of Questions)*, consisting of questions arranged according to the order of Thomas Aquinas' *Summa theologiae*; and the *Opus expositionum (Book of Expositions)*, containing commentaries on scripture and sermons. It is unknown whether Eckhart ever completed this grand project. Only fragments of it have come down to us: the General Prologue and the prologue to the *Opus propositionum*, five questions disputed at Paris, and several commentaries on scripture (Genesis, Exodus, the Book of Wisdom and the Gospel according to John) that were certainly intended to be part of the*Opus propositionum*.

An important clue to Eckhart's self-understanding is to be found in the prologue of his commentary on the Gospel according to John. There he states that in all his works his intention is to expound the doctrines taught by the Christian faith by means of the natural arguments (*rationes naturales*) of the philosophers. This intention is clarified at the beginning of the General Prologue of the *Opus tripartitum* by some preliminary remarks, which Eckhart apparently considers basic to the interpretation of the whole work. According to one of these remarks, the second and third work of the work are so dependent on the *Opus propositionum* that without it they are of little use. He illustrates his method by stating the first proposition ('Being is God'), the first question ('Does God exist?') and the first commentary on a text ('In the beginning God created heaven and earth'), and by showing how the proposition provides the answer to the question and elucidates the text. Eckhart's procedure implicitly criticizes the method of scholastic science. Aquinas' *Summa theologiae*, for instance, consists of thousands of concrete *quaestiones*, whose solution is founded on general principles, which are, however, not explained but presupposed (see Aquinas, T.). In Eckhart's view, the scholastic method requires an axiomatic metaphysics.

A specification of this metaphysics is given in Eckhart's first remark in the General Prologue. The most general terms (*termini generales*), such as 'being', 'one', 'true' and 'good', should by no means be thought of as accidents. 'Being' and the terms convertible with it are not added to things as though posterior to them; on the contrary, they precede everything and are primary in things. The first four treatises of the *Opus propositionum* would deal with these most general terms, which are called *transcendentia* in the Middle Ages. They provide the philosophical foundation of the *Opus tripartitum*. Eckhart identifies the transcendentals with the divine; God alone is properly being, one, true and good. Every other thing is specified through its form to *this* or *that* being.

One of the problems in studying Eckhart is how the identification of God and *esse* in the *Opus tripartitum* can be reconciled with his criticism of ontotheology in the *Quaestiones Parisienses* (the questions disputed at Paris). In the latter, he argues that God is not being (*esse*), but understanding (*intelligere*). He observes ironically that the Evangelist John did not say 'In the beginning was being, and God was being', but 'In the beginning was the Word, and the Word was God'. Being belongs to the domain of the finite, the created.

Eckhart’s German writings comprise numerous sermons and several spiritual treatises: *Reden der Unterweisung (Talks of Instruction)*, *Vom edlen Menschen (The Nobleman)* and *Das Buch der göttlichen Tröstung (The Book of Divine Consolation)*. In one of his sermons (53), Eckhart remarks that he usually preaches about just a few themes. His favourite theme is *Abgeschiedenheit* (detachment), and to it he devoted a separate treatise (*Vom Abgeschiedenheit*), which describes it as the highest virtue, higher even than love. The human person must empty itself and strip off the creaturely conditions of *this* or *that* being. It is by the virtue of detachment that a human being can be most closely united with God. The second theme is the birth of the Son or the Word in the soul. The unity of God the Father and God the Son is for Eckhart the model of the union of the soul with God (see Trinity). The third theme is the ‘nobility’ that God placed in the ‘innermost’ part of the soul. Eckhart calls it the ‘spark of the soul’, which consists in the intellect. ‘If the soul were entirely of this nature, it would be wholly uncreated and uncreatable.’ This proposition was one of those condemned in the trial of Eckhart, but in his apology he denied wiping out the boundaries between God and creature. He wrote that he had never taught that there is something in the soul which is uncreated, ‘because then the soul would be composed of created and uncreated’. A human being is not pure intellect (see Soul, nature and immortality of the).

See also: God, concepts of; Suso, H.; Tauler, J.

JAN A. AERTSEN

List of works

Meister Eckhart (*c*.1260-1327/8) Works, *Die deutschen und lateinischen Werke*, herausgegeben im Auftrage der Deutschen Forschungsgemeinschaft, Stuttgart: Kohlhammer, 1936-.(The modern, almost completed, critical edition of Eckhart’s works.)

Meister Eckhart (*c*.1260-1327/8) *Opus tripartitum*.(Survives only in fragmentary form. The General Prologue, prologue to the *Opus propositionum* (Work of Propositions) and the five questions disputed at Paris are translated by A.A. Maurer in *Parisian Questions and Prologues: Master Eckhart*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies. The prologue and chapter 1 of the commentary on the Gospel of St John, intended for the *Opus expositionum (Work of Expositions)* appear in J.M. Clark and J.V. Skinner, *Meister Eckhart, Selected Treatises and Sermons translated from Latin and German*, London: Faber & Faber, 1958.)

Meister Eckhart (*c*.1260-1327/8) *Reden der Unterweisung (Talks of Instruction)*, trans. in J.M. Clark and J.V. Skinner, *Meister Eckhart, Selected Treatises and Sermons translated from Latin and German*, London: Faber & Faber, 1958.(One of Eckhart’s German spiritual treatises.)

Meister Eckhart (*c*.1260-1327/8) *Vom edlen Menschen (The Nobleman)*, trans. in J.M. Clark and J.V. Skinner, *Meister Eckhart, Selected Treatises and Sermons translated from Latin and German*, London: Faber & Faber, 1958.(One of Eckhart’s German spiritual treatises.)

Meister Eckhart (*c*.1260-1327/8) *Das Buch der göttlichen Tröstung (The Book of Divine Consolation)*, trans. in J.M. Clark and J.V. Skinner, *Meister Eckhart, Selected Treatises and Sermons translated from Latin and German*, London: Faber & Faber, 1958.(One of Eckhart’s German spiritual treatises.)

Meister Eckhart (*c*.1260-1327/8) *Vom Abgeschiedenheit (On Detachment)*, trans. in J.M. Clark and J.V. Skinner, *Meister Eckhart, Selected Treatises and Sermons translated from Latin and German*, London: Faber & Faber, 1958.(Eckhart here describes detachment as the highest virtue.)

References and further reading

Clark, J.M. (1957) *Meister Eckhart: An Introduction to the Study of his Works*, London: Nelson.(A good introduction.)

Goris, W. (1997) *Einheit als Prinzip und Ziel - Versuch über de Einheitsmetaphysik des ‘Opus tripartitum’ Meister Eckharts (Unity as Principle and Objective - An Investigation of the Unity Metaphysics of Meister Eckhart’s Opus tripartitum)*, Studien und Texte zu Geistesgeschichte des Mittelalters 59, Leiden: Brill.(Departing from the systematic perspective of the *Opus tripartitum*, this study gives a new orientation to the question of Meister Eckhart’s philosophy.)

Largier, N. (1989) *Bibliographie zu Meister Eckhart (Bibliography of Master Eckhart)*, Dokimion, vol. 9, Freiburg: Universitätsverlag.(An exhaustive survey of the literature on Eckhart.)

Mojsisch, B. (1983) *Meister Eckhart, Analogie, Univozität und Einheit (Meister Eckhart: Analogy, Univocity and Unity)*, Hamburg: Meiner.(A good example of the philosophical approach to Eckhart. An English translation of this work is expected shortly.)

Melanchthon, Philipp (1497-1560)

Philipp Melanchthon was one of Luther's closest associates, helping to systematize Lutheran theology, and his Loci communes (Commonplaces) (1521) was one of the most influential early works of Protestant theology. He was often a moderating influence in theological debates between Catholics and Protestants. Melanchthon was also involved in controversy over the relationship between human will and God's grace in the achievement of salvation. He was responsible for the reform of Protestant German education in the sixteenth century, through the large number of textbooks which he composed, and through his revisions of the statutes of universities (notably Wittenberg) and schools. As a scholar and reformer of education, he was a staunch follower of the humanism of Agricola and Erasmus, committed to teaching the best Latin authors and the Greek language. Many of his works are textbooks (often produced in different versions), frequently based on lecture notes, summarizing or commenting on classical authors or scripture. Although more important as a summarizer and popularizer than as a source of new ideas, Melanchthon nevertheless made important contributions to the development of logic, rhetoric, ethics and psychology, as well as to aspects of Reformation theology. In logic he contributed to the growth of interest in method. In ethics he established a place for classical moral teaching alongside but subordinate to the teaching of the Bible. His favourite philosopher was Aristotle, and he tended to pour scorn on rival ancient schools of philosophy. In psychology he favoured a simplified Aristotelianism, close to medieval faculty psychology, with strong emphasis on links with biology. He opposed scepticism wherever he encountered it.

1 Reformer of German education

Philipp Melanchthon, born Philipp Schwarzerd at Bretten, received a humanist elementary education, under the influence of his great-uncle, the famous Hebrew scholar and friend of Agricola, Johann Reuchlin. Starting out with Greek and a good knowledge of Latin literature, he was bored by the complex dialectical disputes of the traditional arts course at the University of Heidelberg, where he studied from 1509. His real education came, he said, at the University of Tübingen, where he studied from 1512. In the letter-preface to his *Opera omnia (Complete Works)* (1541) he mentions the importance of Oecolampadius' gift to him, in 1516, of a copy of the first edition of Rudolph Agricola's *De inventione dialectica (On Dialectical Invention)* (1479) (see Agricola, R. §2). In 1518 he was called to the newly founded chair of Greek at Wittenberg, where he came strongly under Luther's influence, enrolling in the theology faculty in 1519. In 1521 he produced the *Loci communes (Commonplaces)*, which attempts a systematic presentation of Protestant theology. Luther described it as 'a book which ought to be esteemed next to the Bible'. Melanchthon surprised his fellow humanists with the strength of his initial endorsement of Luther's views on the corruption of humanity, the enslavement of the will, and justification by faith alone (see Luther, M. §3). In the 1530 version of the *Loci communes* he modified these views somewhat, and in the 1543 version went even further, approaching Erasmus' position on the freedom of the will, and the need for the will to cooperate with the Holy Spirit in order to obtain grace (see Erasmus, D.).

He took part in several of the most famous debates of the early Reformation, including the Colloquy of Marburg (1529) and the Diet of Augsburg (1530). He was the main author of the Augsburg Confession, an authoritative statement of Lutheran doctrine, expressed in the most conciliatory language possible. Strict Lutherans sometimes regarded him as too willing to concede ground to Rome, particularly in the Leipzig Interim (December 1548), an attempt to make peace between Protestants and Catholics. In later life he gave most attention to writing textbooks, to the organization of church and schools, and to controversies with the Calvinists.

Melanchthon's most important contribution to Lutheran theology was his concept of forensic justification, which represents a development of Luther's ideas. According to Melanchthon, the sinful human being is declared to be righteous in God's heavenly court. This is an imputation by God rather than a change in the nature of the individual. This doctrine, which seems to be connected with Melanchthon's humanist interest in Roman legal practice, marks a strong distinction between the Protestant approach to justification and that of Augustine. However (as often in Melanchthon) this firm position is weakened by some of his other comments on the fourth article of the Augsburg Confession which state that a real change does take place.

Melanchthon soon came to dominate Wittenberg University. He was responsible for several reforms of the

Wittenberg statutes, and for giving reformed or new statutes to other Protestant German universities (including Tübingen and Leipzig) and schools (for example in Eisleben, Magdeburg and Nuremberg). As an educational reformer his first commitment was to humanism and to the teaching of ancient languages, but he also placed a strong emphasis on mathematics and physics. Thus in his programme for the Wittenberg arts faculty in 1536, of ten lecturers there were one each for Hebrew, Greek, poetics, Latin grammar and literature, physics and moral philosophy, and two each for rhetoric with dialectic and mathematics. In 1546, the lecturers on poetics and moral philosophy, and one of the rhetoric/dialectic teachers made way for lecturers on physics with botany and two more teachers of Latin literature.

2 Dialectic and rhetoric

Between 1518 and 1547 Melanchthon wrote six books on rhetoric and dialectic. His *De rhetorica libri tres (Three Books on Rhetoric)* (1519) was mainly concerned with invention (the part of rhetoric concerned with generating the material of a composition). He insisted on the close connection between rhetoric and dialectic, as had Rudolph Agricola. He next took the connection further by producing coordinated manuals of both subjects: *Compendiaria dialectices ratio (Compendious Method of Dialectic)* (1520) and *Institutiones rhetoricae (Training in Rhetoric)* (1521). In these books Melanchthon divided the two subjects in a more traditional way than Agricola had done or Ramus was to do. Later in his career he wrote two further coordinated manuals, *Dialecticae libri quatuor (Four Books of Dialectic)* (1528) and *Elementa rhetorices (Elements of Rhetoric)* (1531). These were intended as introductions to Agricola and Aristotle, and presented each subject more fully than his earlier, more elementary textbooks. Finally in *Erotemata dialectices (Questions on Dialectic)* (1547) the dialectic was revised and enlarged to the point where it could be a self-sufficient course, replacing Aristotle altogether. These textbooks included many examples taken from classical literature and the Bible.

In rhetoric Melanchthon's principal innovations were the addition of a fourth genre of oratory (educational, in addition to judicial, deliberative and demonstrative), and the establishment of a simple type of invention for this genre and for 'simple' subjects. This consisted of a small group of questions for the speaker to consider: what is it? what are its parts? what are its causes? what are its effects? which things follow and which oppose? These questions were a selection from the topics, but they can also be seen as a development of the ideas on invention found in the *progymnasmata*, the exercises in short forms of composition which were often used in humanist grammar schools. In later versions he added to the traditional Ciceronian rhetoric syllabus sections on emotional manipulation (summarized from Aristotle), on amplification (based on Erasmus' *De copia*) and on the imitation of Cicero (§2), with particular attention to the theories of prose composition outlined in Cicero's *Orator*.

In dialectic Melanchthon gave particular attention to definition (under which he dealt with the predicables and the categories) and division. As in his rhetoric he introduced an additional but simpler form of invention for easy themes. In the earlier versions this consisted of four topics: definition, causes, parts and effects. By 1547 this became a list of ten questions and had been renamed the method. Method was a topic of considerable interest in the early sixteenth century, on the basis of remarks on the subject by Plato (§§4-5) and Galen (§4), and a reappraisal of Aristotle's *Posterior Analytics* (see Aristotle §§4-6). Melanchthon noticed the connection between method and invention, either through the topics or according to status theory (a technique for determining the key issue at stake in an argument and the best way to argue it) (see Cicero's *On Invention*). His ten questions (what does the word mean? does the thing exist? what is the thing? what are its parts? what is its species? what are its causes? what are its effects? which things are connected to it? what are its cognates? what are its opposites?) represent an amalgamation of the four kinds of question from the second book of Aristotle's *Posterior Analytics*, and a Ciceronian version of the topics. This left a slightly awkward question about the point of the full list of topics, which Melanchthon relegates to the end of his books on dialectic. In 1527 the topics were said to amplify the basic material. In 1547 Melanchthon explained that the method is applied to single words, with topical invention reserved for the proposition. Perhaps the four topics (1527) or the method (1547) were intended to help train beginners, while Melanchthon's basic respect for traditional materials encourages him to include the full list as well. Melanchthon's versions of the topics increase in size in the successive versions of the book. By 1547 they include not only a good deal of material from Agricola's version, but also rules that resemble Boethius' topical maxims (which Agricola rejected). This reflects Melanchthon's eclecticism and the tendency of his dialectic to become more inclusive. By 1547 such strictly scholastic ideas as the theory of supposition have found a place in his dialectic, alongside denunciations of the ancient sceptics and their modes of argument.

Throughout his career Melanchthon insisted on there being connections between rhetoric and dialectic, and on their practical value as guides to reading and composition. He taught manuals of rhetoric and dialectic alongside orations by Cicero or St Paul's letters, so that the precepts could be used to inform reading of the texts, while the texts could illustrate the practical value of the precepts. Melanchthon's published commentaries on Cicero and St Paul make considerable use of rhetoric and dialectic. His major commentary on Romans was entitled *Dispositio orationis in Epistola Pauli ad Romanos (The Disposition of the Oration in St Paul's Epistle to the Romans)* (1530) and devoted itself to the connections between the propositions in that work. In his commentary on Cicero's *Topics*, he analysed a passage from Psalm 1 to explore the similarity between the topics of invention and the figures of rhetoric. This idea goes further than Agricola's analysis of poetic effects in some of his topics entries.

Melanchthon's dialectic textbooks were very influential, going through more than ninety editions in the sixteenth century and becoming required reading in many Protestant German universities, while versions of his rhetoric were printed at least forty times. Many successful textbooks were also composed by his followers, including Thomas Wilson, whose *Arte of Rhetorique* and *Rule of Reason* were the most successful sixteenth-century textbooks of the two subjects in the English language.

3 Ethics

Melanchthon's principal work of moral philosophy was *Philosophiae moralis Epitome (Epitome of Moral Philosophy)* (1538), which he later revised as *Ethicae doctrinae elementorum libri duo (Two Books of the Elements of the Teaching of Ethics)* (1550). Melanchthon creates a space for moral philosophy by making a distinction between God's law and his gospel. Where the gospel is concerned with God's promise to humanity and the forgiveness of sins, the law tells people how they ought to behave in relation to God and other people. He wrote that: 'Moral philosophy is that part of the divine law which teaches about external actions' ([1497-1560] 1963: vol. 16, col. 21). As law, moral philosophy is reliable, but it treats only human external relations, not the relationship between God and the individual. By making clear the primacy of the gospel and allocating a distinct sphere of activity to moral philosophy, Melanchthon both legitimizes teaching the subject and tries to avoid arguments about contradictions between the Bible and philosophy. He produces quotations from scripture to support moral philosophy, and argues that it is useful as part of education, as the basis for human law, that it helps the theologian in interpreting scripture, and that it is a part of the knowledge which gives human beings pleasure.

Melanchthon's approach to moral philosophy is basically Aristotelian (see Aristotelianism, Renaissance §7). Like Aristotle, he sets out by establishing the aim of humankind. In Christian terms this aim is to acknowledge God and take part in his glory. As a result of the Fall, Aristotle was unable to see this, but his idea that the purpose of human life is the exercise of virtue is assimilated by Melanchthon as the aim of external human behaviour. He is even able to accept Aristotle's analysis of virtue as a medium between vices, after a little discussion of the meaning of 'medium'. This contrasts with Luther's hostility to Aristotelian ethics, which he thought had corrupted Christian thought.

As part of his justification of moral philosophy Melanchthon touches on the issue of the freedom of the will. He explains that since philosophy is concerned with behaviour in civil life and not with spiritual issues, it must first be acknowledged that in everyday dealings with other people the will is free, both because scripture says so, and because otherwise there would be no need for laws. He goes on to argue that free will also plays a part in spiritual life, since although no man can feel true remorse or true faith without the help of the Holy Spirit, that help still partly depends on human will.

The list of virtues which Melanchthon discusses seems to have been influenced by Aristotle (for example he gives most space to justice, and he includes friendship), but the actual treatment combines Protestant concerns with the adaptation of Aristotle: there are discussions of the role of the prince in matters of religion, and the religious force of human laws. In contrast to his use of Aristotle, Melanchthon is generally hostile to other ancient schools of philosophy, rejecting the Epicurean position on pleasure and Stoic doctrines on fate and the need to rise above the emotions (see Epicureanism §10; Stoicism §§19-20). He also wrote commentaries on Aristotle's *Ethics* and Cicero's *On Duties (De officiis)*.

4 Psychology and physics

Melanchthon's *Liber de anima (Book on the Soul)* (1540) was intended to be an elementary textbook. As such it was highly successful, going through forty editions and receiving eight commentaries in the sixteenth century. It explicitly avoids ancient and medieval controversies about the interpretation of Aristotle's text, preferring to give a simple exposition of faculty psychology (the theory by which the soul was considered to be divided into three aspects: vegetative, sentient and rational, each having its own range of distinct powers). Scripture takes precedence over philosophy, so for example Melanchthon took for granted the immortality of the individual soul rather than arguing through the implications of Averroes' interpretation of Aristotle and its associated controversies, which were still being argued over in sixteenth-century Italy. At the end of the book he turns away from the classical subject-matter of psychology with a more theological section upholding the freedom of the will (see Aristotelianism, Renaissance §§4-5).

The discussion is conducted in a thoroughly humanist manner, with quotations in Greek, citation of literary texts, and a lengthy argument supporting Cicero's mistaken view (*Tusculan Disputations* I.x.22) that Aristotle's definition of the soul involved reference to *endelecheia* (continuance or eternity) rather than *entelecheia* (actuality) (see Aristotle §17). Melanchthon adds to Aristotle's account of the soul a lengthy discussion of anatomy and physiology (which refers to Vesalius' illustrations of dissections of the human body), a discussion of dreams, and an argument in favour of certainty and against scepticism. His major innovation was to avoid the controversy surrounding the 'possible' and 'agent' intellects by proposing a new, and far simpler, interpretation by which the agent intellect was the faculty of invention (in the rhetorical or dialectical sense) and the possible intellect the faculty of judgment (in the dialectical sense of checking the strength of arguments and putting them in set forms). This idea links his psychology with his dialectic. His work on physics, the *Initia doctrinae physicae (Introduction to Physics)* (1549) is a thoroughly Aristotelian manual, providing an elementary summary which follows the order of the *Physics*, but adding material from other parts of Aristotle's natural philosophy, and a long section on astronomy, according to the Ptolemaic model (see Ptolemy).

See also: Aristotelianism, Renaissance; Humanism, Renaissance §§5-6; Logic, Renaissance §§1, 7; Luther, M.

PETER MACK

List of works

Melanchthon, P. (1497-1560) *Opera omnia (Collected Works)*, in *Corpus Reformatorum*, Brunswick, 1834-60; repr. New York: Johnson, 1963, 28 vols.(The standard edition of Melanchthon's works. Most of the works mentioned in his biography can be found in this edition. For the others, see Mack 1993: 377.)

Melanchthon, P. (1521) *Loci communes (Commonplaces)*, trans. C.L. Hill, Boston, MA: Meador Publishing, 1944.(Useable modern English version of Melanchthon's most influential work.)

Melanchthon, P. (1951-71) *Melanchthons Werke in Auswahl (Selection from Melanchthon's Work)*, ed. R. Stupperich, Gütersloh: Gerd Mohn, 7 vols.(This selected edition includes Melanchthon's *Loci communes*, his commentary on Romans and part of the moral philosophy.)

Melanchthon, P. (1988) *A Melanchthon Reader*, trans. R. Keen, New York: Peter Lang.(English translation of selected works.)

Melanchthon, P. (1991-) *Melanchthons Briefwechsel. Kritische und kommentierte Gesamtausgabe (Melanchthon's Correspondence: a Critical Edition with Commentary)*, ed. H. Scheible, Stuttgart: Frommann. (Excellent new edition of the correspondence.)

References and further reading

Archiv für Reformationsgeschichte, Literaturbericht (1972-), Gütersloh: Gerd Mohn.(Annual bibliography of reformation studies, including Melanchthon, with some annotation.)

Cicero (*c*.80 BC) *De inventione (On Invention)*, trans. H. Hubbell, in Cambridge, MA: Loeb Classical Library, Harvard University Press, 1949.(Latin text and English translation of Cicero's handbook on rhetoric and informal argument, referred to in §2.)

Cicero (late 45 BC) *Tusculanae disputationes (Tusculan Disputations)*, trans. J.E. King, in Cambridge, MA: Loeb Classical Library, Harvard University Press, 1945; trans. A.E. Douglas, Warminster: Aris and Phillips, 1985, 1990.(The King translation gives the whole text in Latin and English, while Douglas provides Books I, II and V with notes. Referred to in §4.)

Green, L.C. and Froehlich, C.D. (eds) (1982) *Melanchthon in English: New Translations into English with a Registry of Previous Translations*, Sixteenth-Century Bibliography 22, St Louis, MO: Center for Reformation

Research.(A bibliography of English translations.)
Hammer, W. (ed.) (1967-81) *Die Melanchthonforschung im Wandel der Jahrhunderte (Melanchthon Research Across the Centuries)*, Gütersloh: Gerd Mohn, 3 vols.(Comprehensive bibliography of materials for the study of Melanchthon.)
Hartfelder, K. (1889) *Philipp Melanchthon als Praeceptor Germaniae (Melanchthon as the Educator of Germany)*, Berlin: A. Hofmann; repr. Nieuwkoop: de Graaf, 1964.(Still the fundamental account of Melanchthon's activities as scholar and teacher.)
Keen, R. (1988) *A Checklist of Melanchthon Imprints through 1560*, Sixteenth-Century Bibliography 27, St Louis, MO: Center for Reformation Research.(Bibliography of early editions of Melanchthon's works.)
Mack, P. (1993) *Renaissance Argument*, Leiden: Brill, 320-33.(Agricola's influence on Melanchthon.)
Manschreck, C.L. (1958) *Melanchthon: The Quiet Reformer*, Nashville, TN: Abingdon.(Full-length study in English.)
Maurer, W. (1967-9) *Der junge Melanchthon zwischen Humanismus und Reformation (The Young Melanchthon between Humanism and Reformation)*, Göttingen: Vandenhoeck & Ruprecht, 2 vols.(A good modern account of Melanchthon as a humanist.)
McGrath, A.E. (1986) *Iustitia Dei: A History of the Christian Doctrine of Justification*, Cambridge: Cambridge University Press, vol. 2, 23-32.(On Melanchthon's doctrine of forensic justification.)
Meerhoff, K. (1994) 'The Significance of Melanchthon's Rhetoric in the Renaissance', in P. Mack (ed.) *Renaissance Rhetoric*, London: Macmillan, 46-62.(Excellent account of the place of rhetoric in Melanchthon's teaching, with references to Meerhoff's other important studies.)
Petersen, P. (1921) *Geschichte der aristotelischen Philosophie im protestantischen Deutschland (A History of Aristotelian Philosophy in Protestant Germany)*, Leipzig; repr. Stuttgart: Frommann, 1964.(Good account of Melanchthon's Aristotelianism.)
Risse, W. (1964) *Die Logik der Neuzeit: I, 1500-1640 (The Logic of the Modern Period)*, Stuttgart and Bad Cannstatt: Frommann.(A survey of sixteenth-century schools of logic.)
Spitz, L.W. (1963) *The Religious Renaissance of the German Humanists*, Cambridge, MA: Harvard University Press.(An old general account of Melanchthon in English.)
Vasoli, C. (1967) *La dialettica e la retorica dell'Umanesimo (Humanist Dialectic and Rhetoric)*, Milan: Feltrinelli.(A good account of humanist rhetoric and dialectic, including a section on Melanchthon.)
Wiedenhofer, S. (1976) *Formalstrukturen humanistischer und reformatischer Theologie bei Philipp Melanchthon (Formal Structures of Humanist and Reformation Theology…)*, Berne.(Links Melanchthon's humanism with his theology.)

Melissus (mid or late 5th century BC)

Melissus was a Greek philosopher from the island of Samos. A second-generation representative of Eleatic metaphysics, he published one work, entitled On Nature or On That-Which-Is, which has been partially reconstructed by editors. It defends a version of Parmenides' monism, but recast with terminology and arguments directly accessible to a readership schooled in the eastern Greek (Ionian) style of physical speculation, as distinct from Parmenides' western Greek background. Although it is uncertain how important Melissus was to his own contemporaries, his prosaic but clear presentation of Eleatic concepts was more widely adopted by later writers than the enigmatic pronouncements of Parmenides.

Melissus argues that that-which-is is: (1) omnitemporal; (2) infinite in extent; (3) one; (4) homogeneous; (5) changeless, that is, without (a) reordering, (b) pain, (c) grief or (d) motion; (6) indivisible; and (7) bodiless. Here (1) - 'it always was what it was, and always will be' - is a departure from Parmenides, who had outlawed past and future in favour of a static present. Likewise (2) contrasts with Parmenides' defence of spatially finite being. The remaining predicates are consonant with Parmenides, although (5)b-c suggest that the being Melissus has in mind is a living one, presumably a deity - an aspect not brought out by Parmenides. Melissus wrote 'If there were many things, they ought to be such as I say the One is' - a remark sometimes thought to have inspired his contemporaries the atomists.

1 Life and work

In his own day, Melissus may have won greater renown as an admiral than as a philosopher. The one firm date in his life is 441 BC, when he led the navy of his native Samos to victory over an Athenian fleet commanded by Pericles. Since his age at the time is unknown, the date of his book might be placed anywhere between 490 and 390 BC. One very suspect source makes his philosophy already known to the Athenian politician Themistocles, who died *c*.462 BC. But the Epicurean Colotes, in a work which after criticizing its principal target Democritus turned the attack onto a further eight philosophers in apparently chronological order, placed Melissus after Socrates, who was active *c*.440-399 BC, a clue which would bring his *floruit* well into the second half of the fifth century - late enough for him to have read Empedocles, Anaxagoras and perhaps even the atomists. The tradition that he studied under Parmenides is not particularly trustworthy, although it cannot be doubted that Parmenides was the dominant influence on his philosophy.

A number of verbatim fragments of his one book have been preserved by Simplicius. They can be pieced together with the help of Simplicius' ancillary paraphrase and another in the Peripatetic treatise *On Melissus, Xenophanes and Gorgias*.

2 Philosophical aims

Melissus' treatise is a methodical defence of Eleatic monism, written in unadorned Ionian prose. The conclusions are by and large Parmenidean, but the arguments are not. There is virtually no sign of Parmenides' most characteristic argument against change - that it would entail not-being, an altogether inapplicable concept (see Parmenides §§2-4). Instead, Melissus relies on the barely controversial premise 'Nothing could ever come to be out of nothing' (fr. 1). Furthermore, whereas Parmenides had in the main inferred each predicate of that-which-is by an independent argument, nearly all Melissus' inferences form a single chain, with each predicate inferred directly from the previous one.

Melissus is not interested in Parmenides' highly refined mode of investigation through the logic of being. He writes as a physical thinker addressing a like-minded audience, and expounds the Eleatic One with arguments appropriate to Ionian cosmology. His title (probably authentic, despite some scholars' hesitation), *On Nature or On That-Which-Is* ['on nature' in Greek is *physis*], in effect announces that this is to be an Eleatic *physics*. His departures from Parmenides, in reverting to ordinary temporal language and in postulating a spatially infinite being, are more symptomatic of this project than of intellectual independence.

3 Temporal and spatial infinity

For the book's beginning we have a probably complete text. It opens as follows: 'It always was what it was, and

always will be. For if it came to be, it is necessary that before it came to be there [or 'it'] was nothing. Well if there [or 'it'] was nothing, nothing could ever come to be out of nothing' (fr. 1). 'Since, then, it did not come to be, it both is and always was and always will be' (fr. 2, beginning).

Unlike Parmenides' highly paradoxical rejection of not-being, Melissus' opening move would be unsurprising to an audience familiar with earlier Ionian cosmologists in the tradition of Anaximander (§2), thinkers whose fundamental tenet had been the permanent existence of an underlying stuff. The commonly assumed causal (rather than ontological) principle 'Nothing could ever come to be out of nothing' was its basis. Equally unsurprising in an Ionian context (see Heraclitus, fr. 30) is Melissus' expression of this permanence in terms of omnitemporality, where Parmenides had chosen to collapse past and future into the present. It is probably misleading to call this a significant philosophical disagreement. Melissus is simply translating Parmenidean thought into the philosophical language which his audience speaks.

As in Parmenides, the rejection of past generation is not followed up with an explicit defence of future imperishability. This is assumed to follow by parity of reasoning. The ground this time would be that 'nothing could perish into nothing'.

In the following part of fragment 2 Melissus moves on from temporal to spatial infinity. (This clear division has been widely missed by scholars who have supposed that the new argument must start at the beginning of fragment 2.): 'And it has no [spatial] beginning or end, but is infinite. For if it had come to be it would have a [spatial] beginning (for it would have begun the process of coming-to-be at some time) and end (for it would have ended the process of coming-to-be at some time). But since it neither began nor ended [the process], and always was and always will be, it has no [spatial] beginning or end'.

The interpretation of this has been much debated. Critics from Aristotle on have detected the fallacious inference 'If *p*, *q*; but not-*p*; therefore not-*q*'. To see why this is unfair, it is crucial to appreciate that, where Parmenides' arguments had evidently addressed an audience accustomed to the concept of a finite universe (see Parmenides §3), Melissus assumes the opposite - that the universe will be infinite unless it can be shown to be otherwise. This is yet another sign of his audience's background in Ionian physics, where the infinity of the universe, prefigured as early as Anaximander, was by Melissus' day enshrined in Anaxagoras' cosmology and on its way to becoming central to atomism (see Anaxagoras §2; Atomism, Ancient.

Melissus' question is: what could have set bounds on that-which-is? If nothing, then it is infinite. Only one thing could have made it finite, and that is a process of generation. Since any generative process would have to be temporally bounded, it could only have produced a spatially finite being. You cannot *create* an infinitely large entity, any more than you can build an infinitely long road, given only that any such process must start at some time (and hence somewhere) and stop at some time (and hence somewhere). Since, therefore, it has already been demonstrated that that-which-is never came to be, there is nothing to limit it spatially, and it becomes infinite by default.

Having demonstrated first its temporal and then its spatial infinity, Melissus adds how the latter is both inferentially dependent on and parallel to the former: 'For what is not *all* would not be able to be *always*' (end of fr. 2). 'But just as it is always, so too it must also always be infinite in magnitude' (fr. 3). 'Nothing is either omnitemporal or infinite if it has a [spatial?] beginning and end' (fr. 4).

The first sentence is particularly obscure. Melissus is perhaps drawing once again on a common assumption of Ionian material monism, that only the underlying stuff of the universe, taken as a whole, is everlasting, while the individual portions of it which constitute animals etc. (*if* there are any, Melissus would add) are temporary. Therefore the entity which Melissus calls everlasting must be the whole of what there is. This, if not a sufficient condition of its being infinite, he clearly regards as a necessary condition.

4 Changeless unity

That-which-is is called 'the One' by Melissus, a new emphasis on unity only weakly prefigured in Parmenides. In fragment 6 Melissus formally infers its unity from its spatial infinity: 'For if there were two they would not be able to be infinite, but would have boundaries in relation to each other'. Oneness here appears to mean uniqueness, and unfortunately it is hard to find any reading of the preceding arguments for its infinity which does not already

assume that there is just one of it. A further objection made by the Peripatetic Eudemus, that the inference depends on its being not merely infinite but infinite in all directions, is less worrying, since the argument for its infinity does indeed imply the latter.

The proof of the next attribute, ‘alike’, that is, homogeneous, survives only in a paraphrase: ‘Being one, it is alike everywhere [or ‘in every way’]. For if it were unlike, there would be a plurality, and no longer one but many’ (*On Melissus, Xenophanes and Gorgias* 974a12-14). Its homogeneity thus follows from its oneness. But whether the preceding arguments have shown it to be ‘one’ in the required sense (= ‘partless’ rather than ‘unique’?) is debatable.

After a brief summary of the results so far (fr. 7.1), there follows a generic argument against change: ‘And it could neither lose anything nor become larger nor be rearranged, nor does it suffer pain or grief. For if any of these happened to it, it would no longer be one. For if it changes, it is necessary that that-which-is is not alike, but that what previously was perishes while that-which-is-not comes to be. So if it were to become changed by a single hair in ten thousand years, it will all perish in the whole of time’.

Formally, this is meant to be an inference from the immediately preceding attribute ‘alike’, in conformity to Melissus’ standard inference pattern (see §2). But this time the likeness has to be invariance over time, not space, and that has not been specifically defended. (The defence might be that variation over time would imply a plurality of temporal parts, bordering on each other and therefore failing to be omnitemporal.) We should perhaps think of the likeness-changelessness inference as dictated by formal considerations, and give primary weight to the more interesting final sentence. This implies a version of the Principle of Plenitude: given infinite time all possibilities are realized. *Any* piecemeal perishing, however nugatory, makes overall perishing a possibility, and therefore a certainty over an infinite timespan - contrary to the earlier demonstration of everlastingness.

Following the generic rejection of change, Melissus adds four arguments against four species of change: reordering, pain, grief and motion (frs 7.3-10). The first three largely reapply the generic objections, but with the added consideration that pain and grief, being weaknesses, are incompatible with everlastingness. Melissus’ concern to deny human forms of suffering to his One has caused surprise. But it must be borne in mind that some Ionian thinkers had already endowed the unitary underlying substance with the characteristics of divinity (see Anaximander §2; Anaximenes §1). If, as suggested, Melissus is working within and correcting that tradition, a later report that he equated his One with God may be well founded.

The refutation of motion can be paraphrased as follows. Since void would be nothing, void clearly does not exist. Therefore that-which-is, lacking any admixture of void, cannot vary in density but is completely full. Therefore it cannot give way at any point. Therefore it cannot (internally) move (frs 7.7-10).

This rejection of void as ‘nothing’ is the nearest Melissus comes to accepting Parmenides’ outlawing of not-being. But he also makes an important advance on Parmenides in rendering explicit the dependence of motion on void.

Two further inferences are recorded. Being motionless, it is indivisible, since division involves motion (fr. 10). And ‘Being one, it must not have [a?] body. If it had bulk, it would have parts and no longer be one’ (fr. 9). The latter is especially puzzling, since the One’s incorporeality would seem to undermine the argument for its immobility on the grounds of ‘fullness’ (that is, solidity). It is likelier that Melissus would be denying that it has *a* body, one with organic parts, by way of objection to anthropomorphic conceptions of divinity. But it is doubtful whether the argument, at least as reported, can support this interpretation.

5 Attack on the senses

In fragment 8, apparently from a separate section of his book, Melissus follows the example set by Parmenides in turning his ontological conclusions against the senses. Since sense perception reports constant change, both between distinct stuffs and between pairs of opposite properties, it cannot be trusted. These many items, if real, would have to obey Melissus’ strictures and be altogether changeless, each of them staying exactly as it seemed when first encountered. In short: ‘If there were many things, they ought to be such as I say the One is’.

6 Influence

This last remark is often conjectured to have inspired early atomism, whose exponents Leucippus and Democritus

(§2) did indeed populate the universe with an infinite plurality of atoms, each internally changeless like a Melissan One. Likewise Melissus' insistence that void is a necessary condition of motion (see §4) may be seen as having helped prompt the atomists' rehabilitation of void.

However, the chronological relation of Leucippus to Melissus remains undetermined. The void-motion connection could as well be one that Melissus learnt from Leucippus as vice versa; and the atomists' introduction of an Eleatic plurality may have needed no impetus from Melissus, just their rehabilitation of Parmenidean not-being in the guise of void, which already implied the separation of Parmenidean being into a plurality of discrete parts. The seductive hypothesis of Melissan influence on the atomists must then remain unproven. Stronger evidence of his public recognition as a philosophical writer in his own day is the fact that his book *On Nature or On That-Which-Is* was evidently a primary target for Gorgias in his *On That-Which-Is-Not or On Nature, a parody of Eleatic Philosophy*, a parody of Eleatic philosophy.

In subsequent generations Melissus never inspired the kind of reverence, or even respect, that Parmenides was accorded. But his generally clear formulations of Eleatic positions are much more widely reflected in later writers, especially Aristotle, than the high-flown obscurities of Parmenides. He can thus be regarded, if not as the driving intellect of Eleatic thought, at least as its leading public voice.

See also: Monism

DAVID SEDLEY

References and further reading

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge & Kegan Paul.(Chapters 10, 11 and 14 include a subtle if sometimes over-generous appraisal of Melissus.)

Barnes, J. (1987) *Early Greek Philosophy*, Harmondsworth: Penguin.(Chapter 10 is a translation of Melissus' fragments complete with the contexts in which they are preserved.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols. (Includes the most detailed and comprehensive English-language history of early Greek thought; the treatment of Melissus in volume 2 pages 101-18 is very lucid.)

Anon. (date unknown) *On Melissus, Xenophones and Gorgias*, trans. W.S. Hett, in *Aristotle, Minor works*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1936.(Contains a summary of Melissus' argument.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(A valuable survey of Presocratic philosophy, including texts and translations; contains the best account of Melissus available in English.)

Melissus (mid or late 5th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 1, 258-76.(The standard collection of the ancient sources; includes Greek text of the fragments with translations in German.)

Reale, G. (1970) *Melisso. Testimonianze e frammenti*, Florence: La Nuova Italia.(The fullest and best guide to the ancient evidence and modern scholarship on Melissus.)

Stokes, M.C. (1971) *One and Many in Presocratic Philosophy*, Washington, DC: Center for Hellenic Studies. (Places Melissus' central theme in its historical context.)

Memory

Memory is central to every way in which we deal with things. One might subsume memory under the category of intellect, since it is our capacity to retain what we sense, enjoy and suffer, and thus to become knowing in our perception and other activities. As intelligent retention, memory cannot be distinguished from our acquisition of skills, habits and customs - our capabilities both for prudence and for deliberate risk. As retention, memory is a vital condition of the formation of language.

Amnesia illustrates dramatically the difference between memory as retention of language and skills, and memory as the power to recollect and to recognize specific things. In amnesia we lose, not our general power of retention, but recall of facts - the prior events of our life, and our power to recognize people and places. Amnesiacs recognize kinds of things. They know it is a wristwatch they are wearing, while unable to recognize it as their own.

This recall of events and facts which enables us to recognize things as our own, is more than just the ability to give correctly an account of them. One might accurately describe some part of one's past inadvertently, or after hypnosis, or by relying on incidental information. Thus, present research on memory both as retention and as recall of specific episodes, attempts to characterize the connection which persists between experience and recall. Neurological or computer models of connectivity owe something to traditional notions of a memory trace, but emphasize also the re-tracing of original memories by later experience and by intervening episodes of recall.

1 History of the interest in memory

It is striking that in two standard compendia of Chinese and of Indian philosophy, there are no entries under memory, remembering, reminiscence or recall. Naturally, there are words used in those cultures which roughly match terms in European languages - any human being will speak of what they did at some time earlier; children begin to speak spontaneously of what they did the day before, and remark with delight or distress at things they have seen earlier. But memory as a philosophical preoccupation is specific to certain cultural periods; memory as our powers of retention is so central to intelligent activity as to be subsumed under intellect.

Within Western philosophy itself, there is no mention of memory in the pre-Socratic fragments. It seems that so long as the poetic history of Homer held sway, memory was not discussed by him or the philosophers. This was not for want of interest in the events of earlier times. Homer aptly and vividly evokes their reality. Then in Plato's Socratic dialogues, while Homeric poetry is attacked as endangering an interest in truth, concern with the past is displaced by a search into memory itself, and Plato's dialogues - his own artful contrivances - are presented as if recollections. Within this new myth of a historical anti-poetic Socrates we learn the philosophical values of truth as against poetry, and of strict recollection against mere fancy about the past.

Plato's Socrates argued initially that all knowledge is recollection. He denied that sensory experience (in the present) produces knowledge, and conjectured that recollection revived knowledge acquired in a life before birth (see Plato §11). Thus not only the Homeric tradition of finding significance in human affairs in a (partly legendary) past is rejected, but also the role of women, those responsible for the child gaining its foundational knowledge, is elided.

In Plato's later *Theaetetus* we find the image of memory as an imprint left in the soul by earlier experience, a trace of what has occurred. In place of Plato's allegory of knowledge as recollection of a state in which knowledge was gained without the impurities of sensory experience we find this more scientific speculation about possible mechanisms of memory, an idea taken up by his student and successor, Aristotle. About eight centuries later these connected concepts of memory and of the past were brought into crisis by the north African Augustine. The present hour consists not only of the present minute, but of past minutes and the minutes to come. In turn, the present minute consists not only of its present second, but of its past seconds and the seconds to come. Thus, successively, the present as an extended period disappears. It remains only as an extensionless cut between the periods of past and future. The picture of present memory as defining the reality of the past is erased (see Augustine §§5, 7).

In the seventeenth century Descartes excepted memory from his radical questioning of the senses, while his notion of instantaneous reasoning was to avoid the reliance of reason on memory in any case. That all understanding of

the past rested on personal memory remained unquestioned by Locke and Hume, and yet all arguments for the validity of memory assumed what they tried to prove. While Locke took common sense as a sufficient reply, Hume drew radically sceptical conclusions. Kant tried to show that the forms of knowledge cast into doubt by Hume were conditions of any coherent experience. Memory, as retaining an awareness of what something is like over some finite period of time, for instance, is such a condition of intelligible experience.

This emphasis on the mind's contribution to the order of things led, however, to an idealism in which the mind knows, not the world itself, but its own forms of representation. Thus the sceptical empiricist tradition set in motion by Descartes, while shaken, was not overturned. The sceptical tradition was empiricist in locating the sources of knowledge within the elements of experience. The emphasis on this experience made it seem that experience itself rather than what we experienced was the primary object of knowledge. Towards the end of the nineteenth century Husserl, inspired by Frege, Brentano and Meinong, created a comprehensive alternative to empiricism - a phenomenology of experience as revealing, not itself but its objects. This phenomenology attempted an unprejudiced study of experience, including memory, as ways in which we grasp the world itself (see Phenomenology, epistemic issues in).

Husserl's enquiries led him finally to think of the human body itself, acting as part of a socially significant world, as the centre of conscious experience. Remembering is one of the activities by which we hold bodily sway in the world. This approach was developed by Heidegger, Sartre and Merleau-Ponty. Alternatives to mind-centred empiricism were also developed during much the same period within analytical philosophy by Wittgenstein, Ryle and J.L. Austin.

2 Contemporary work: retention, imagery and feeling

Retention and recollection. Those who suffer from amnesia, knowing neither who they are, nor any event of their previous lives, retain most of their know-how intact. An amnesiac victim might join in a philosophical discussion, only learning from the impression made on others that they must have practised philosophy during that lost section of their lives. No one is surprised that an amnesiac can speak their mother tongue, nor that they can walk, sing and bargain. Thus, memory as retention of skills is different from the recollection of events, neighbourhoods, smells and sounds.

Imagery. As practised by Wittgenstein, Ryle and J.L. Austin, philosophy broke away from an entrancement with memory as a present image within which one must discern pastness and valid indications of the past. Much of what we call remembering has little to do with the possession of imagery. And, though some events come back to us in flashes of imagery, more is required before one remembers anything coherently. On the other hand, aware of little but the buzz of our conversation we may be recalling things with great accuracy. Similarly, depending on no prior image of the tune, one may hum a melody heard at a concert the previous evening. The recall may leave us with images; images of the past need not have provided the recall. Unable to remember someone's expression, we might mimic it. A flash of an image *may* facilitate that mimesis but, equally, the effort of producing the facial expression may be what lets loose our imagery. Thus images, sometimes central to remembering, may be more product than cause of recall (see Imagery).

Feelings. Though displaced from the centre of our picture of memory, imagery remains strongly associated with our elemental feelings about this power to recall. Perhaps this is because of the importance to us of certain intimate forms of remembering, particularly reminiscing and reliving, which vitally involve either imagery or emotion. True, reliving the past in animated conversation may involve little imagery of it. Yet, in the absence of specific imagery, our feelings - some shadow or reanimation of how we felt at the time, are essential to our reliving the past rather than simply reporting it from memory. These feelings we have in reliving prior experiences themselves have something of the quality of images. So, while memory does not consist in a memory image, the existence of a complex interplay of image and affect in the processes of recall is important to the cluster of ideas brought together under memory.

The dead past. To undergo an experience of recall is to be subject to the return of the past - only a metaphor, yet irreplaceable in a philosophy of memory. (In French, the principal verbs of recall are *se souvenir* - to bring oneself under (the past), and *se rappeller* - to call (the past) back to oneself.) Yet, remembering as bringing back what happened, or as reliving the past, encounters an impasse - accounts of memory are haunted by the notion of the

past as dead, as if requiring memory's miracle of resurrection. The desire to remember thus encounters the limits met by mourning - itself a principal paradigm and emblem for memory in general. As memory would revive the events known to be past, in mourning the impulse is to address, as if to recall, the one known to have died.

3 Contemporary work: causal connections and traces

The initial conceptual analysis of the language of memory concentrated on the everyday contexts in which we learn and use our language of memory, and centred on remembering as successful present performance. But the picture which emerged of remembering as simply getting it right about our own past is unsatisfactory. After all, we may describe events from early childhood, not from memory of them, but from having been told by others. Different aspects of our past, in contrast, may come to mind from our own recall. No one else might have been in a position to tell us of certain details which subsequent evidence corroborates. Hence, to remember something is not simply to represent it correctly later on. Nor can memory be defined by the idea of (Hume's) lively belief about one's past. Sometimes we are recalling something when engaged in what we think of as pure imagination. We might intend to write a fictional story, and others may point out that it matches, incident for incident, some episode of which, hitherto, no one had informed us.

Remembering cannot be defined, either, as the representation of one's past without the need for prompting. After all, quite without assistance, we might merely invent something which, coincidentally, did happen. Conversely, we may not remember something and then, upon being prompted, remember the event for ourselves. A reference to some connection between the past event and the present recall seems to be needed - someone suffering amnesia, but hypnotized to recount a traffic accident which, coincidentally, did happen to them is not thereby remembering it. If they now recount something, not in any way because it happened to them, but simply because of the hypnotist's suggestion, they are remembering only what the hypnotist told them.

This idea of a causal connection between the event and its subsequent recall involves its own difficulties, however. The hypnotist might think of the story only because they recall hearing about the accident from the subject before the subject suffered amnesia. So the subject's original experience of the accident plays a crucial causal role in the subject's now coming to recount the event. The causal connection requires more exact definition. The subject's original experience is causally relevant only for the hypnotist's suggesting the story to them subsequently. Assuming that the hypnosis does not actually revive the memory of the accident, it is causally irrelevant to the subject's accepting the story upon its being suggested to them by the hypnotist.

Alas, complication follows complication. It is easy to imagine that the trauma of the accident might have made the victim more highly suggestible to hypnotists' stories about accidents. So the subject's having been in the accident they subsequently describe might still be relevant to their acceptance of the hypnotist's suggestion, and not only relevant to the suggestion being made to them. It seems one must reinvoke the ancient idea of a memory trace which encodes the information the person gains from an experience. Only where there is an activation of such a trace is there the highly specific causal connection required for true remembering. Such an idea appeals both to common sense and to some contemporary scientific theories of how both general information and specific events might be stored - cortically, and also as conditioned reflexes, making possible muscle memory, as musicians and sports professionals speak of it.

This idea of memory is not universally accepted. Not enough is known about the physiological processes required for the memory of specific events, to say that we can insist upon the presence of such a strict trace. In reply, it will be insisted that while memory may be highly dispersed in alternative neural networks rather than encoded in specific brain cells, it is hard to imagine any alternative to a neural, or neuro-muscular process which would preserve the highly specific information about individual events which makes it possible for us to recall them with all the sensual and emotive feel which they commonly carry.

It has also been objected to such a causal analysis of remembering that memory is a power or disposition to recount things about one's past. Is it not absurd to suggest that some past event should continuously cause the continuation of such a power of recall? Is this not like saying that the capacity of one's living room curtains to keep their colour, day in, day out, is continually held in place by their originally having been dyed? In reply, a causal theorist will point out that it is not the power of recall which is continually upheld by an event in the past. What is required, more simply, is that the trace established by the event remembered should encode all the

information genuinely recalled, and that if prompting is necessary (as normally it is), this prompting should activate that encoding trace.

4 Contemporary work: challenges to fixed ideas of present and past

These approaches, both through everyday concepts and by speculative scientific hypothesis, take for granted memory as reviving an already given past. The phenomenology of Sartre and Merleau-Ponty, too, takes the past as real in itself, not to be reduced in Augustinian fashion to the process of recollection. Nevertheless, their work emphasizes how much what we call past and present owe each other. The past does not so innocently create memories which then we deem to be valid. Perhaps half-consciously, we make nothing of aspects of our past in order to make something of other aspects. We might create a favourable past or, alternatively, one which embodies a harsh vision of our lives. What we remember is not entirely accidental. How we remember and what we recall partially constitutes our present and, conversely, what we countenance as part of our present bears on what we are prepared to recall.

The themes of memory are recalled in a classical enigma. When the present was present, it could not be remembered. Now it is past, its presence is only in recall. This presence in recall cannot be the presence it had when first it occurred. It is in this sense that remembering is, in Derrida's phrase, the raising up of 'a past that was never present'.

See also: Memory, epistemology of

MAX DEUTSCHER

References and further reading

Aristotle (384-322 BC) *On Memory*, in J. Barnes (ed.) *The Complete Works of Aristotle*, Princeton, NJ: Princeton University Press, 1984, 2 vols.(Thorough description of elementary forms of memory, and the development of the idea of a memory trace.)

Augustine (AD 400) *Confessions*, trans. H. Chadwick, Oxford: Oxford University Press, 1991, Book 10.(Vivid portrayal of puzzles about time and recollection, in the confessional mode. Clearly written, easy-to-medium difficulty.)

Derrida, J. (1986) *Memoires for Paul de Man*, New York: Columbia, lectures 1 and 2.(Lectures delivered on the death of Paul de Man. Evocative and moving study of images and concepts of memory, and the relation of mourning to memory.)

Deutscher, M. (1989) 'Remembering "Remembering"', in J. Heil (ed.) *Cause, Mind and Reality*, Dordrecht: Kluwer.(Analyses notions of representation, widens the range of forms of memory, contextualizes the idea of a memory trace.)

Hume, D. (1739) *A Treatise on Human Nature*, any edition, Part III, particularly sections 5 and 6, and Part IV. (Classic statement of the image theory of memory, and of attendant sceptical problems. Very clear, medium difficulty.)

Husserl, E. (1954) *Crisis of European Sciences and Transcendental Phenomenology*, trans. D. Carr, Evanston, IL: Northwestern University Press, 1970, Part II, sections 24-5 and Part III, sections 51-4.(Analyses the pitfalls in Hume's scepticism, the continuity of experience described in terms of detachment (transcendence) and of bodily involvement (sedimentation). Complex, difficult, rewarding.)

Krell, D. (1990) *Of Memory, Reminiscence and Writing*, Bloomington, IN: Indiana University Press.(Close, brilliant study of literary phenomenological, deconstructive and scientific sources of ideas of memory. Valuable bibliography.)

Locke, J. (1690) *Essay on Human Understanding*, Bk II, chs 10, 14, 27.(Ideas of memory as an idea or image, as a storehouse, as a condition for personal identity. Clear and thoughtful. Medium difficulty.)

Martin, C. and Deutscher, M. (1966) '"Remembering"', *Philosophical Review* LXXV 2: 161-96.(Describes wide variety of forms of memory, influential in re-establishing the idea of memory as including an information carrying trace, causally linking past with present.)

Merleau-Ponty, M. (1945) *Phenomenology of Perception*, trans. C. Smith, Routledge & Kegan Paul, 1962, Part I, section 6, Part III, section 2.(The body rather than the mind as centre of experience; analyses metaphors of the flow of time, and the intimate involvement of the present in describing the past. Medium difficulty, evocative, sometimes poetic prose.)

Plato (*c*.428-348 BC) *Phaedo*, trans. D. Gallop, Oxford and New York: Oxford University Press, 1993.(The theory

of knowledge as recollection. Easy to medium difficulty.)
(*c*.428-348 BC) *Theaetetus*, trans. M.J. Levett, revised M. Burnyeat, Indianapolis, IN: Hackett Publishing Company, 1990.(Explanation and criticism of the idea of memory as a trace. Medium to difficult.)
Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson, chs 6 and 8.(Explains the elusive character of the self, and of the present in terms of token reflexive expressions. Describes similarities between imagination and memory, and attacks the image theory of memory. Easy to medium difficulty.)
Squires, R. (1969) 'Memory Unchained', *Philosophical Review* LXXVIII: 178-96. (Witty attack on the idea of memory as involving a causal connection or trace.)

Memory, epistemology of

Memory appears to preserve knowledge, but there are epistemic questions about how this could be. Memory is fallible, and empirical research has identified various ways in which people systematically misremember. Even wholesale error seems possible: Russell (1927) proposed that it is logically possible for the world to have sprung into existence five minutes ago, complete with spurious ostensible memories of earlier times. In light of such possibilities, some sceptics argue that memory cannot yield knowledge.

Assuming that memory provides knowledge, there are serious epistemic issues about how it does this. For instance, does some introspectible quality of remembering provide distinctive evidence for what is remembered, or is it some other feature of memory that secures the epistemic justification needed for knowledge? How readily recollectible must a proposition be in order for it to be known while it is not being recalled? Does a full retention in memory of a previous basis for knowing something assure continuing knowledge of it? Does forgetting an original basis for knowing without replacing it imply a loss of knowledge?

1 Memory and traditional epistemic issues

Epistemic controversies about matters specific to memory are comparatively rare. There are two apparent reasons for this. One is that remembering is a source of reasons for belief with approximately the same problems and prospects as sensory perception (see Perception, epistemic issues in). The two processes share the salient features of being conscious, cognitive episodes that prompt new beliefs without apparent inference. Of the two, perception seems the more promising source of justification. This is principally because conscious sensory characteristics are clearly involved in ordinary sensory perception. They provide something that might confer justification on some resulting perceptual beliefs, though it is problematic how sensory characteristics could justify beliefs. In contrast to the perceptual case, no conscious qualities are clearly distinctive of recollective events. There may be less of a basis on which to explain how memory can furnish justification and knowledge. Thus, the outcome of investigations of the epistemic properties of perception is plausibly thought to carry over to the case of memory, with memory perhaps turning out to have weaker versions of the same epistemic capacities.

A second reason for the relative paucity of epistemic work on memory may be the prevalence in epistemology of disputes about scepticism. In virtually all sceptical controversies, remembering plays at most a derivative epistemic role. Memory on its own serves as a means of knowledge preservation, not knowledge acquisition. The fundamental sceptical issues concern whether justified belief or knowledge can be had in the first place. For instance, if some sceptical argument shows that we can have neither perceptual knowledge nor justified perceptual belief, then no important epistemic question about recalling perceptual beliefs seems to remain. If on the other hand scepticism about some such category of knowledge is refuted, then memory offers only the prospect of retention of knowledge in that category. This is not to say that knowledge retention is unimportant. Retaining nothing, we could know only the facts that we could learn entirely in a moment. But retention seems less epistemically fundamental than do capacities that arguably enable us to acquire new knowledge.

Some sceptical challenges have specifically addressed memory knowledge. The sceptical arguments have generally charged that something about the vulnerability of memory to error shows it not to be a source of reasons sufficiently strong to provide knowledge. But again, memory is not unique in the relevant epistemic respects. Assessments of analogous sceptical arguments concerning sensory perception carry over to this sort of attack on memory knowledge. The familiar mechanisms of deception such as ordinary sorts of error, the evil demon of Descartes and the illusory stimulations of electrodes, are capable of producing as pervasive and deceptive a replication of veridical experiences in the case of sensory experiences as in the case of ostensible memories. Anti-sceptical responses typically deny that such a deception is genuinely possible, or that its possibility shows that we lack knowledge (see Scepticism). These denials have the same merits whether the argument is aimed at knowledge by perception or knowledge by recollection. Although memory knowledge is one of the most important applications of sceptical reasoning, memory does not raise any distinctive basic sceptical issues.

2 An epistemic problem specially involving memory retention

One epistemic problem of considerable interest does essentially involve the capacity of memory to preserve

knowledge. Suppose that, at noon, Smith knows that Jones is smiling. Smith has reasons that are sufficient for him to know this, including clearly seeing Jones smile. At dusk Smith acquires compelling new evidence to the effect that Jones did not smile at noon, receiving sworn testimony from Robertson, a universally acknowledged pillar of integrity. In an honest error, Robertson sincerely and convincingly testifies that Jones has a double who stood in for Jones at noon. It seems plain that as a result of hearing this testimony Smith no longer knows that Jones smiled at noon. This can be puzzling. For one thing, at dusk Smith might perfectly recall his noontime reasons for believing that Jones smiled at noon. If those reasons were adequate for Smith to know this fact at noon, how can exactly the same reasons be inadequate later?

Also puzzling is the fact that at dusk Smith seems to possess a cogent argument for regarding Robinson’s testimony as misleading. Prior to hearing the testimony, Smith unproblematically remembers and continues to know that Jones smiled at noon. This appears to enable Smith to know by inference that any evidence against Jones having smiled at noon is evidence against a truth, and hence misleading. But even if this inference is fresh in Smith’s mind when he hears Robinson’s testimony, at that point Smith would be unreasonable to regard Robertson’s testimony as misleading. Yet why would this attitude be unreasonable, given the apparent cogency of Smith’s argument for it?

Credible answers to these questions will not deny that we ever fully remember our previous reasons for holding some belief. A more promising approach is suggested by a careful look at the structure of epistemic support for a belief. We should observe that conspicuous positive evidence for a proposition cannot by itself justify believing the proposition. Epistemic justification is more holistic than that. In particular, at noon Smith’s justified belief that Jones was then smiling was supported, not merely by Smith’s seeing Jones smile, but also by perceptual indications and background information that argued that the situation was normal. These indications importantly include the lack of any clue of a deception. It is only this sort of comprehensive body of relevant considerations that can epistemically justify a belief. Anything less may fail to take account of pertinent information (see Knowledge, defeasibility theory of).

This structural feature of justification makes it clear that once Smith has new relevant evidence, even a perfect recollection of his previous evidence about Jones at noon does not guarantee that the belief that Jones smiled at noon continues to be justified. At dusk Smith becomes aware of a strong indication that he was deceived at noon. As a result, the new totality of his evidence does not justify the belief that Jones smiled at noon. Such changes in evidence show why retaining a previous basis for knowing does not entail continuing to know.

Similar considerations explain why Smith cannot reasonably regard Robinson’s testimony as misleading once he hears it. Smith possesses a cogent argument that such evidence is misleading only when his evidence on balance supports that conclusion. Otherwise his evidence casts doubt on the conjunction of the argument’s premises. Until Smith hears Robinson’s testimony, what Smith retains from noon supports without challenge an argument for the conclusion that any evidence against Jones having smiled at noon is misleading. Once Smith hears Robinson’s testimony however, any such argument is undercut. Smith’s evidence may still include all that previously supported the conclusion. But now his evidence also includes his hearing of testimony which spoils that support. The premise asserting that Jones did smile at noon is essential to his justifiably inferring that contrary evidence is misleading. When Smith has heard the testimony, this premise ceases to be supported by the entirety of his pertinent considerations. Smith then has no cogent argument for regarding the testimony as misleading.

Epistemic justification for a belief at a time is constituted only by the totality of relevant considerations possessed at the time. Such a totality can mislead by a deceptive expansion. The justification needed for knowledge thus is holistic and transitory. These properties of justification serve to dispel the mystery of how someone might clearly remember his whole basis for knowing a proposition and for accurately classifying all evidence against it as misleading, and yet be straightaway misled by just such evidence and no longer know the proposition to be true.

3 An epistemic problem specially involving memory loss

Someone might come to know the date of the Battle of Waterloo by reading it in a clearly reliable text. Years later the person may continue to believe that the date is 1815, while entirely forgetting the source of this belief and without receiving any subsequent corroboration of that date. Does thus forgetting imply that the knowledge is lost? If so, then it seems that we cease to know most of what we recall, however securely we first came to believe it,

since it is relatively rare for us to retain our original justification for beliefs not recently acquired. Yet if we do continue to know beliefs of forgotten justifying origins, then it seems that we must somehow manage to know them while lacking justification for them. Each of these alternatives is implausible.

An attractive response to this problem occupies a middle ground. It holds that some retained beliefs whose original justification has been forgotten and which are subject to no new doubts are still known, others not. The difference is made by the facts about how currently reasonable to the person is the proposition that continues to be believed. Retention of the original evidence is not required if one has indications that a reliable memory is operative in sustaining the belief. For instance, where the belief is accompanied by an impression of having some reputable source or other for it, and an impression of having often confirmed and seldom disconfirmed the beliefs that thus seem trustworthy, those impressions help to constitute a current justification for the belief. Where someone has nothing more on behalf of a belief than a sense that the proposition was assertively conveyed at some point, perhaps in fiction or even in a dream, the proposition will not be currently believed reasonably enough to be known.

This is an elaboration and application of the appealing view that a reliable memory is what preserves knowledge. The view tells us that we do not continue to know merely by having a recollective impression of a factual proposition that we once knew and to which we now have no substantial objection. Only those of our recollections that are evidently trustworthy can sustain our knowledge of such propositions.

See also: Innate knowledge; Knowledge, tacit; Memory; Reasons for belief; Scepticism

EARL CONEE

References and further reading

Harman, G. (1986) *Change in View*, Cambridge, MA: MIT Press.(Chapter 4 presents a full exposition and suggests a different solution to the problem discussed in §3 above.)

Hume, D. (1739) *A Treatise of Human Nature*, ed. D.G.C. Macnabb, London: Collins, 1970.(Classic philosophical approach to the nature of memory is advocated in Book I.)

Ginet, C. (1980) ‘Knowing less by knowing more’, *Midwest Studies in Philosophy* 5: 51-60.(Presents a full exposition and suggests a different solution to the problem discussed in §2 above.)

Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(Chapter 10 discusses the epistemic significance of some empirical work on the fallibility of memory.)

Russell, B. (1927) *An Outline of Philosophy*, London: Allen & Unwin.(Source of the sceptical hypothesis that the world began five minutes ago, complete with ostensible memories.)

Mencius (4th century BC)

Mencius (Mengzi) was a Chinese Confucian philosopher, best known for his claim that human nature is good. He is probably the single most influential philosopher in the Chinese tradition, in that an interpretation of his thought became the basis of the civil service examinations in China in the fourteenth century and remained so for almost 600 years. The primary source for his thought is the collection of his sayings, debates and discussions known as the Mengzi.

Mencius (known in Chinese as Mengzi (Master Meng) or by his full name, Meng Ke) was born early in the fourth century BC in the state of Zou, now part of Shandong Province. Towards the end of that century he travelled from state to state, seeking a ruler who would put his philosophy into practice, and briefly served as a minister in the state of Qi. The collection of his sayings, debates and discussions is known simply as the *Mengzi*. It was compiled either by Mencius himself or by his disciples soon after his death.

Mencius saw his main intellectual task as defending the doctrines of Confucius against those of the egoist Yang Zhu and the universalistic consequentialist Mozi. In order to do this, he developed a novel and detailed theory of human nature which went beyond anything Confucius had said.

Mencius' claim that human nature is good has been given different interpretations (see Xing). He clearly thought that humans innately have active but incipient tendencies toward virtue, which he describes, using an agricultural metaphor, as sprouts. Each sprout corresponds to one of his four cardinal virtues, and each virtue has an emotion or attitude that is characteristic of it: benevolence is characterized by compassion, righteousness by shame and disdain, wisdom by approval and disapproval, and ritual propriety by either respect or deference. In a famous example, Mencius says that our sprout of benevolence manifests itself in a spontaneous feeling of 'alarm and compassion' when we 'suddenly' see a child about to fall into a well (*Mengzi* 2A6). This same virtue shows itself in one's compassion for a suffering animal, one's service to and love of one's parents and the disinterested concern of virtuous rulers for their subjects.

Righteousness is similarly complicated, manifesting itself in such things as a beggar refusing to accept a handout given with contempt, a wife and concubine being ashamed of their husband's humiliation of himself to obtain luxuries, a chariot driver being ashamed to cheat in a ritual hunt, a person refusing to accept contemptuous forms of address, disdaining to serve base rulers, and obeying and respecting one's elder brother. One passage (*Mengzi* 5A9) suggests that a wise person is a good judge of the character of others, recognizes what is base and avoids it, is a prudent and perceptive judge of policy and a good administrator. Another passage (4A27) describes wisdom as understanding and being committed to benevolence and righteousness. Mencius does not clearly distinguish the virtue of ritual propriety, although it seems to be related to righteousness.

Neo-Confucian philosophers (see Neo-Confucian philosophy) regarded the four sprouts as fully developed virtues whose operation was impeded (in the uncultivated individual) by selfish desires. However, it is more likely that Mencius thought the sprouts were only incipient virtues, which must be cultivated so that they develop into full virtues (see Self-cultivation in Chinese philosophy). That is, we must 'extend' the reactions of the sprouts from the situations in which we already have them to relevantly similar situations in which we should, but do not yet, have them. Mencius seems to advocate several methods for achieving extension, including 'concentration', which probably includes the reflective and joyful exercise of the sprouts. He also taught his students through the study of poetry and 'case studies' of the actions of sages.

Mencius presents what is recognizable as a virtue ethic, but one that is importantly different in terms of its conception of flourishing, its cardinal virtues and its theory of ethical cultivation from classic Western examples such as those of Aristotle and Thomas Aquinas (see Virtue ethics). Like other virtue ethicists, Mencius regards ethics as objective, but eschews decision procedures based upon either rules or the weighing of utility and stresses the context-sensitivity of virtuous responses. As a Confucian, Mencius holds that virtuous individuals will have compassion for all humans, but he also thinks we have special obligations to, and should have greater concern for, those tied to us by particular bonds, including kinship. Mencius is also typically Confucian in his emphasis upon learning from tradition, and in his belief that the virtues first manifest themselves in the family.

In terms of style, Mencius prefers to focus on the specific and the concrete. For example, Mencius explores virtues and their semblances, not through a Socratic search for definitions but through contrasting concrete examples (for examples, *Mengzi* 2A2; 7B37). Furthermore, what he says is geared to the understanding and needs of his interlocutors on particular occasions. So, for instance, he may be oversimplifying his view of self-cultivation for his audience in *Mengzi* 6B2. However, careful study of the *Mengzi* suggests that underlying his concrete comments on specific occasions is a deep and systematic ethical view.

See also: Confucian philosophy, Chinese; Neo-Confucian philosophy; Virtues and vices

BRYAN W. VAN NORDEN

List of works

Mencius (*c.*300 BC) *Mengzi*, trans. D.C. Lau, *Mencius*, New York: Penguin Books, 1970; trans. J. Legge, *The Works of Mencius*, New York: Dover Publications, 1970.(Lau is a good contemporary translation, with helpful appendices. Legge is a reprint of the 1895 translation, including the Chinese text and explanatory notes.)

References and further reading

Graham, A.C. (1990) 'The Background of the Mencian Theory of Human Nature', in A.C. Graham, *Studies in Chinese Philosophy and Philosophical Literature*, Albany, NY: State University of New York Press, 7-66. (Detailed philological study of some key terms in early Chinese thought.)

Ivanhoe, P.J. (1990) *Ethics in the Confucian Tradition: The Thought of Mencius and Wang Yang-ming*, Atlanta, GA: Scholars Press. (A good general introduction to the thought of Mencius, emphasizing his differences from his later interpreters.)

Lau, D.C. (1970) 'On Mencius' Use of the Method of Analogy in Argument', *Mencius*, New York: Penguin Books: 235-63.(Explains some of the more puzzling passages in Book 6 of the *Mengzi*.)

Nivison, D.S. (1996) 'Motivation and Moral Action in Mencius' and 'Two Roots or One?', in B.W. Van Norden (ed.) *The Ways of Confucianism*, La Salle, IL: Open Court, 91-119 and 133-48.(Analyses of *Mengzi* 1A7 and 3A5, respectively.)

Shun, K.L. (1991) 'Mencius and the Mind-lnherence of Morality: Mencius' Rejection of Kao Tzu's Maxim in *Meng Tzu* 2A:2', *Journal of Chinese Philosophy* 18 (4): 371-86.(A careful analytic study.)

Van Norden, B.W. (1996) 'What Should Western Philosophy Learn from Chinese Philosophy?', in P.J. Ivanhoe (ed.) *Chinese Language, Thought and Culture: Nivison and His Critics*, La Salle, IL: Open Court.(Presents an overview of the literature on early Confucianism.)

Yearley, L.H. (1990) *Mencius and Aquinas: Theories of Virtue and Conceptions of Courage*, Albany, NY: State University of New York Press.(An excellent comparative study of Mencian and Thomistic treatments of the virtues.)

Mendelssohn, Moses (1729-86)

A Jewish disciple of Leibniz and Wolff, Mendelssohn strove throughout his life to uphold and strengthen their rationalist metaphysics while sustaining his ancestral religion. His most important philosophic task, as he saw it, was to refine and render more persuasive the philosophical proofs for the existence of God, providence and immortality. His major divergence from Leibniz was in stressing that 'the best of all possible worlds', which God had created, was in fact more hospitable to human beings than Leibniz had supposed. Towards the end of his life, the irrationalism of Jacobi and the critical philosophy of Kant shook Mendelssohn's faith in the demonstrability of the fundamental metaphysical precepts, but not his confidence in their truth. They would have to be sustained by 'common sense', he reasoned, until future philosophers succeeded in restoring metaphysics to its former glory. While accepting Wolff's teleological understanding of human nature and natural law, Mendelssohn placed far greater value on human freedom and outlined a political philosophy that protected liberty of conscience. His philosophic defence of his own religion stressed that Judaism is not a 'revealed religion' demanding acceptance of particular dogmas but a 'revealed legislation' requiring the performance of particular actions. The object of this divine and still valid legislation, he suggested, was often to counteract forces that might otherwise subvert the natural religion entrusted to us by reason. To resolve the tension between his own political liberalism and the Bible's endorsement of religious coercion, Mendelssohn argued that contemporary Judaism, at any rate, no longer acknowledges any person's authority to compel others to perform religious acts.

1 Metaphysics/natural theology

Affectionately called 'the German Socrates' by European admirers, and hailed by some of his Jewish disciples as a modern Maimonides, Moses Mendelssohn (born Dessau, died Berlin), despite his hunchback, cut a towering figure in his generation. Emerging from the ghetto with a solid command of traditional Jewish learning and a wide knowledge of medieval Jewish philosophy, he independently learned ancient and modern European languages, literature and philosophy. In an astonishingly short time he became an important cultural critic and went on to produce notable contributions to German philosophy, especially in metaphysics, ethics and political theory. Unlike Spinoza, the only prior Jewish thinker of modern times to have had a major impact on European thought, Mendelssohn retained his affiliation with Judaism. He actively promoted and fostered modernization within the Jewish community and spoke out forcefully and effectively in governmental and intellectual fora for the bestowal of equal civil rights on Jews. Challenged by adversaries and well-intentioned 'friends' to explain his seemingly unjustifiable loyalty to an outmoded religion, Mendelssohn wrote some of the seminal works of modern Jewish philosophy.

Mendelssohn made almost no claim to originality in metaphysics. He described himself as little more than an exponent of the teachings of the Leibniz/Wolffian school, perhaps lending a more felicitous or up-to-date expression to the affirmations of God's existence and providence and human immortality that had been propounded by Leibniz and Wolff and their disciples. Here and there, however, he admitted modestly that he was providing a new version of an old argument or even saying something that had not been said before.

Mendelssohn made his reputation and set out his basic metaphysical stance in his *Abhandlung über die Evidenz in metaphysischen Wissenschaften (Treatise on Evidence in Metaphysical Knowledge)*, an essay that took first prize in a competition sponsored by the Royal Academy in Berlin in 1763; Immanuel Kant received an honourable mention. The Academy's question was whether 'the truths of metaphysics in general, and the first principles of natural theology and morality in particular' can be shown to be as securely established as those of mathematics. Mendelssohn answered that such principles 'are capable of the same certainty' but are by no means as easily grasped. After discussing the obstacles to such comprehension, he went on to offer cosmological and ontological proofs for the existence of God. He sought to give the ontological argument an 'easier turn' by reversing its usual course and arguing first for the impossibility of God's nonexistence and then against the notion that the most perfect being would enjoy a merely possible existence.

Following Leibniz, Mendelssohn argued in a number of writings that the combination of divine goodness and greatness known as providence brings into being 'the best of all possible worlds'. He was certainly aware of the abundance of evil and injustice in the world. But he found that the good outweighed the evil, and indeed that the

quantity of evil was less than Leibniz himself had supposed. For Mendelssohn joined with other, later representatives of the *Aufklärung* in rejecting the Christian dogma, accepted by Leibniz, that most human beings are destined for eternal damnation. This idea, so much at odds with Leibniz's recognition that the perfection and happiness of rational creatures are critical in God's design, was, he wrote, a notion of which 'neither our reason nor our religion knows anything' (see Providence; Evil, problem of).

What reason teaches about the afterlife is, according to Mendelssohn, altogether consolatory. This view emerges most clearly in what was probably his most famous work, a philosophical dialogue entitled *Phädon, oder über die Unsterblichkeit der Seele (Phaedo, or on the Immortality of the Soul)* (1767). Borrowing from Plato the title and the setting, in Socrates' jail cell, Mendelssohn used the condemned philosopher as the mouthpiece for an argument he had admittedly derived from his own recent predecessors, including such *Aufklärer* as the natural theologian Hermann Samuel Reimarus and the liberal Protestant theologian Johann Joachim Spalding. 'Is it consistent', Mendelssohn's Socrates asks, 'with the Supreme Wisdom to produce a world in order to make the happiness of the creatures inhabiting it arise from the contemplation of its wonders and a moment later deprive them of the capacity for contemplation and happiness?' Pursuing the idea of divine goodness to its logical conclusion, Mendelssohn reasoned that the afterlife must be the arena in which the world's injustices are rectified and *all* rational creatures continue to progress towards ever higher perfection and happiness.

Going beyond his rhetorical question and other arguments of similar provenance, Mendelssohn proudly offered what he saw as an original argument for immortality, a 'proof from the harmony of our duties and rights': An individual who believed that death marked the end of his existence would find that his natural right to preserve his own life at all costs would conflict at times with the state's demand that he be prepared to lay down his life for the common good. But such a conflict can have no objective warrant. For in the mind of God, 'all the duties and rights of a moral entity are in the most perfect harmony'. Since the state's right to call for the supreme sacrifice is irrefutable, it is evident, Mendelssohn concludes, that the human soul does not die with the body.

In time Mendelssohn himself came to see weaknesses at many points in the philosophical structure he had upheld. Confronted, toward the end of his life, by the irrationalism of F.H. Jacobi and by the new critical philosophy of Immanuel Kant, whom he called the 'all-crusher', he felt compelled to acknowledge the insufficiency of rationalistic metaphysics. In his fullest exposition of the philosophy he had set forth, *Morgenstunden, oder Vorlesungen über das Dasein Gottes (Morning Hours, or Lectures on the Existence of God)* (1785), he sorrowfully ceased to reaffirm its irrefutable truth.

Yet, whatever speculative reason might seem to teach, he now argued, common sense still sufficed to orient people and guide them along the path to the most important truths. Just what Mendelssohn meant by common sense has been a subject of much dispute, both among his contemporaries such as Thomas Wizenmann and Kant himself and among modern scholars. But, however he conceived of this faculty, it is clear that he expected our reliance on it to be temporary. For, in the 'cyclical course of things', providence would cause new thinkers to arise who would restore metaphysics to its former glory.

2 Moral and political philosophy

Following Wolff, Mendelssohn affirmed that the fundamental moral imperative is a natural law obliging all rational beings to promote their own perfection and that of others. Unlike Wolff, he did not elaborate all the ramifications of this natural law. But he clearly saw perfection in much the same terms as Wolff, as an unending process of physical, moral and intellectual development, leading naturally to the increase of human happiness.

In sharp contrast to Wolff, Mendelssohn regarded liberty as an indispensable precondition of the pursuit of moral and intellectual perfection. Only a free person, he argued, can achieve moral perfection. For virtue is the result of struggle, self-overcoming and sacrifice, and these must be freely chosen. Intellectual perfection, too, can be attained only by one who is free to err. So, in place of Wolff's tutelary state, Mendelssohn developed a contractarian polity that left individuals largely free to define their own goals. Insisting above all on the inalienable liberty of conscience, he decried any state attempt to impose specific religious behaviour or to discriminate against members of a minority faith. While Mendelssohn's liberalism clearly flowed from his commitment to liberty itself, it was no doubt also informed by his concern for the situation of his fellow Jews, whom he ardently wished to see relieved of a degraded political and social status.

3 Philosophy of Judaism

Mendelssohn faced fewer difficulties in reconciling his Leibniz-Wolffian views with his Judaism than his mentors or their disciples had confronted in reconciling their philosophical views with Christianity. Because he saw little conflict between philosophy and Judaism and because his own position as the first Jew to be accepted by the intellectuals in a fundamentally intolerant society was precarious, he initially refrained from explaining to the general public his continued adherence to his ancestral religion. But he was by no means oblivious to the challenges that Enlightenment scepticism posed to all believers in revealed religions. Nor was he immune to attack. In response to repeated demands from critics that he account for his irrational adherence to Judaism, he wrote his celebrated treatise *Jerusalem* (1783). Subtitled *Über religiöse Macht und Judentum* (On Religious Power and Judaism), *Jerusalem* responds to August Friedrich Cranz's charge that the religious coercion of recalcitrant Jews countenanced in Judaism is blatantly incompatible with Mendelssohn's own political liberalism. In responding to this charge, Mendelssohn proceeded beyond the call of his immediate duty and outlined a general philosophy of Judaism.

Viewed against the backdrop of medieval Jewish philosophy, Mendelssohn's account contains both familiar and unfamiliar elements. Following in the footsteps of Saadiah Gaon, Judah Halevi and Maimonides (and seemingly ignoring the criticisms of Spinoza and various sceptics) he maintained that God's revelation at Sinai was an amply attested historical fact (see Saadiah Gaon; Halevi, Judah; Maimonides, M.). What God disclosed to the people of Israel, Mendelssohn held, was not a 'revealed religion' commanding the acceptance of particular dogmas but a 'revealed legislation' requiring the performance of particular actions. The Israelites at Sinai had already acquired, by unaided reason, a knowledge of the universal religion of humankind. What they learned from revelation was a set of God-given duties. Noting the broad range of these duties and the manifold aspects of life encompassed by them, Mendelssohn reflected at length on the so-called 'ceremonial law', problematized at least since the earliest days of the Christian critique of Judaism. This body of obligations, he reasoned, seems to be intended to combat idolatrous tendencies by ordaining actions that perpetually bring to mind the fundamental truths of natural religion. Observance of these divinely ordained regulations allowed Jews to maintain their grasp of true religion and so to serve as a salutary example to the other peoples of the earth (see Halakhah).

In keeping with the tenets of traditional Judaism, Mendelssohn insisted that the Bible's revealed legislation remained in force perpetually, subject only to such modifications as the changes of time and circumstance rendered advisable. But he parted company with tradition in arguing that the Roman destruction of the Temple in Jerusalem had transformed Jewish law from a corpus of communally enforceable obligations into a code which only the individual Jew had the right to impose on himself or herself. This unprecedented claim parried the charge of illiberalism. Whether it represents a reluctant and merely tactical concession to liberalism, as some have argued, or whether it reflects Mendelssohn's true convictions, as is more commonly believed, it was clearly the only means that Mendelssohn could find to render modern Judaism fully compatible with political liberalism.

See also: Enlightenment, Jewish; Jewish philosophy in the early19th century; Halevi, Judah; Leibniz, G.W.; Saadiah Gaon; Wolff, C.

ALLAN ARKUSH

List of works

Mendelssohn, M. (1929-30) *Moses Mendelssohn Gesammelte Schriften Jubiläumsausgabe*, Berlin: Akademie-Verlag.(The authoritative edition of Mendelssohn's collected works.)

Mendelssohn, M. (1997) *Philosophical Writings*, trans. and ed. D. Dahlstrom, Cambridge: Cambridge University Press.(Collection of Mendelssohn's early philosophical writings, including those on metaphysics, ethics and aesthetics.)

Mendelssohn, M. (1755) *Philosophische Gespräche (Philosophical Speeches)*, Berlin.(In this work, Mendelssohn declares himself a disciple of the school of Leibniz and takes sides with Spinoza.)

Mendelssohn, M. (1755) *Briefe über die Empfindungen (Letters on Feeling)*, Berlin.(Contains a philosophy of the beautiful and forms a chief basis of all philosophical-aesthetic criticism in Germany.)

Mendelssohn, M. (1763) *Abhandlung über die Evidenz in metaphysischen Wissenschaften (Treatise on Evidence in Metaphysical Knowledge)*, Berlin: Mande und Spener, 1786.(Essay comparing the demonstrability of metaphysical propositions with that of mathematical ones.)

Mendelssohn, M. (1767) *Phädon, oder über die Unsterblichkeit der Seele (Phaedo, or on the Immortality of the Soul)*, Vienna: C.F. Schade, 1812.(His chief philosophical work, dealing with the immortality of the soul, following Plato's dialogue of the same name.)

Mendelssohn, M. (1783) *Jerusalem, oder über religiöse Macht und Judentum (Jerusalem, or on Religious Power and Judaism)*, trans. with notes by A. Arkush, *Jerusalem*, Hanover, NH: University Press of New England, 1983.(His epoch-making work calling for the emancipation of the Jews and setting out the conditions of religious freedom in a pluralistic state.)

Mendelssohn, M. (1785) *Morgenstunden, oder Vorlesungen über das Dasein Gottes (Morning Hours, or Lectures on the Existence of God)*, Berlin: C. Voss, 1786.(Sets forth Mendelssohn's fundamental metaphysical beliefs and argues against pantheism.)

References and further reading

Altmann, A. (1969) *Moses Mendelssohns Frühschriften zur Metaphysik*, Tübingen: J.C.B. Mohr.(A detailed study of Mendelssohn's early writings on metaphysics.)

Altmann, A. (1973) *Moses Mendelssohn: A Biographical Study*, Tuscaloosa, AL: University of Alabama Press.(A magisterial account of Mendelssohn's life and thought.)

Arkush, A. (1994) *Moses Mendelssohn and the Enlightenment*, Albany, NY: State University of New York Press.(Studies the relationship between Mendelssohn's general philosophy and his conception of Judaism.)

Mental causation

Both folk and scientific psychology assume that mental events and properties participate in causal relations. However, considerations involving the causal completeness of physics and the apparent non-reducibility of mental phenomena to physical phenomena have challenged these assumptions. In the case of mental events *(such as someone's thinking about Vienna), one proposal has been simply to identify not* 'types' *(or classes) of mental events with* types *of physical events, but merely individual* 'token' *mental events with* token *physical ones, one by one (your and my thinking about Vienna may be 'realized' by different type physical states).*

The role of mental properties *(such as 'being about Vienna') in causation is more problematic. Properties are widely thought to have three features that seem to render them causally irrelevant: (1) they are 'multiply-realizable' (they can be realized in an indefinite variety of substances); (2) many of them seem not to supervene on neurophysiological properties (differences in mental properties do not always depend merely on differences in neurophysiological ones, but upon relations people bear to things outside their skin); and (3) many of them (for example, 'being painful') seem inherently 'subjective' in a way that no objective physical properties seem to be. All of these issues are complicated by the fact that there is no consensus concerning the nature of causal relevance for properties in general.*

1 The problem

Mental causation is central both to our conception of ourselves as thinking and experiencing agents, and to some of the most promising theories of our behaviour. We ordinarily explain, for example, someone's drinking water by citing their thirst and their belief that water would quench it.

Despite the central place of mental causation in both folk and scientific psychology there is a long philosophical tradition that is sceptical about it. This scepticism is based on the apparent scientific intractability of intentionality and consciousness (see Consciousness; Intentionality). Intentionality is the property of mental states whereby they are 'about' things, and can be true or false. For example, the thought that the cat is crying is about the cat and represents it (truly or falsely) as crying. Consciousness involves further subjective and phenomenal properties, such as the taste of beer or the awfulness of pain. It has seemed to some philosophers that it is impossible to fit intentionality and consciousness within the causal order disclosed by physics and biology, and this has led some to the conclusion that there is no genuine mental causation.

2 Cartesian dualism

The *locus classicus* for the problems of mental causation is Cartesian philosophy. Descartes held that mental events and states are distinct from physical events and states: the latter involve changes in the properties (especially spatial properties) of extended bodies, while the former involve changes in the properties (especially properties involving thought) of non-extended mental substance (see Descartes, R. §8; Dualism). He further held that these two kinds of events interact causally. In fact, Descartes thought that mental causes are required to explain linguistic and mathematical behaviour since they exhibit a kind of novelty that he believed could not be accounted for solely in terms of physical causes.

There are many difficulties with Descartes' position, but two are most relevant to our topic. One is that he provided no account of how mental-physical (or mental-mental) causation occurs. He hypothesized that the interaction between mental and bodily events takes place in the pineal gland. While this localizes mental-physical causation, it says nothing about why or how various types of mental states cause, or are caused by, various types of mental or physical states. These connections are posited as basic causal links. But it seems implausible that anything as complicated as mental events should be linked to physical events by basic causal relations.

The second problem is that, contrary to Descartes' view, it is now widely believed that every physical event can be accounted for solely in terms of antecedent physical events (up to whatever objective indeterminism obtains generally) (see Materialism in the philosophy of mind). So, for example, if Alma raises her hand, that movement can be accounted for causally, at least in principle, in terms of prior neurophysiological and other physical events. Alma's conscious decision to raise her hand, if distinct from these physical events, either casually overdetermines the movement or is *epiphenomenal* with respect to it (see Epiphenomenalism). But overdetermination seems *ad*

hoc, motivated only by a desire to provide some causal role for the mental. If overdetermination is rejected then, given the distinctness of mental and physical events, the only causal role available for the former is as the causes of non-physical, that is, other mental events. But even this seems implausible in view of the fact that at least some mental events can be causally accounted for in terms of prior physical events. Indeed, an empirically well-attested claim of modern science is the *causal completeness of physics*: every event, if it is caused at all, occurs by virtue of physical causes (see Determinism and indeterminism). Consequently, the view that mental events are distinct from physical events plausibly leads to 'mental event epiphenomenalism', or the doctrine that mental events have no physical effects.

Mental event epiphenomenalism is radically at odds with our views about our mental lives. But, fortunately, the line of reasoning that leads to it can be resisted by rejecting the Cartesian view that mental events are distinct from physical events. Descartes' arguments for the distinctness are now generally thought to be weak. His view that mathematical and linguistic behaviour could not be accounted for in terms of antecedent physical events has been undermined by, for example, a computational theory of mind. His more philosophical arguments have not fared much better.

3 Identity theories

This leaves the way open for the position that every mental event *is* a physical event (seeMind, identity theory of). The most prominent contemporary version of this view is the 'token identity theory', due to Donald Davidson (1984). Davidson (§3) holds that every particular (or 'token' mental event is identical to some or other particular physical event, but that *classes* (or 'types') of mental events are not identical to natural classes of physical events (see Type/token distinction). For example, every particular event of someone's *wanting water* may be identical to some or other physical event, but the mental *type* event, *wanting water*, may not be identical to any (naturally definable) physical *type* event (see multiple realizability below).

In many people's minds, however, Davidson's view invites the suggestion that, although mental states and events are identical to physical ones, still mental *properties* are distinct from physical properties (see Property theory). Hence, while the token identity theory avoids mental event epiphenomenalism, it raises the problem of whether, and if so how, mental properties participate in casual relations. Consider a much-discussed example from Dretske (1988): when the soprano's singing causes the glass to break, it is the singing's having a certain *frequency*, not its having a certain *meaning* that is relevant to its having that effect. Given that all causation is physical causation, how could a mental property such as meaning become relevant?

This problem has two parts. One is to characterize the difference between a property's being causally relevant or irrelevant in a causal transaction. The other is to determine when, if ever, mental properties are causally relevant. This problem arises because, on the one hand, folk psychology and cognitive theories seem to assume that mental properties are sometimes causally relevant and, on the other hand, certain philosophical accounts of mental properties raise questions of whether they can be. The view that mental properties are not causally relevant is 'mental *property* epiphenomenalism'. If true, it would entail that, even if mental events are causes of behaviour, their mental properties are not causally relevant to it.

If mental properties were *identical* to the physical properties responsible for behaviour, then there would be no special problem of mental property causation, since it is generally thought that neurophysiological properties are the physical properties causally responsible for behaviour. But there are three reasons frequently cited for thinking that mental properties are not identical to neurophysiological ones: multiple realizability, externalism and subjectivity. The first two apply mainly to intentional mental properties and the third to consciousness.

Multiple realizability. To say that a property is 'multiply realizable' is to say that it can be exemplified in virtue of exemplifying different physical properties. Prime examples of such properties are *functional* properties, or properties an object or state has in virtue of its capacity to exemplify a certain pattern of causal relations. For example, a computer has the property of 'executing a print command' by virtue of running a certain program. Since very different physical objects can exemplify the same causal capacities, functional properties are multiply realizable, a fact reflected in the diversity of physical materials (radio tubes, silicon chips) out of which computers can be made.

It has become the received view in philosophy of mind that certain mental properties are functional properties (see

Functionalism). So, for example, to be a belief is to be a state that is generally caused by inferences from other beliefs and has the capacity to participate in further inferences that cause other beliefs, intentions and so forth. If mental properties are functional properties, then they are not *identical* to neurophysiological properties, though they are realized in us by neurophysiological properties. It may well be that properties other than neurophysiological ones can also realize mental properties. In any case, there seems to be a problem concerning how functional properties can be causally relevant to behaviour. The trouble is that neurophysiological properties are sufficient for causally accounting for behaviour. The functional mental properties they realize seem irrelevant to their capacity to produce behaviour.

Externalism. Externalism is the view that intentional mental properties are constituted, in part, by external relations between thinkers and items external to the thinker. Thought experiments due to Putnam (1975) and Burge (1979) have been taken to support externalism. For example, Putnam imagines two individuals, Oscar and twin-Oscar. They are neurophysiologically identical, but Oscar lives in an environment containing H_2O while twin-Oscar lives on another planet much like earth except that instead of H_2O it has a superficially indistinguishable substance XYZ. According to Putnam and most others who have thought about it, Oscar's expression 'water' refers to H_2O and not XYZ while twin-Oscar's expression 'water' refers to XYZ and not H_2O. The point carries over to their thoughts. Oscar's thought that he expresses by 'Here is some water' is true: he is in the presence of H_2O. It would be false should Oscar travel to twin earth and think the thought in the presence of XYZ (see Content: wide and narrow; Methodological individualism).

If externalism is true then intentional mental properties (like thinking 'Here is some water') are externally individuated since what thought a person has depends on features of their environment. Another way of putting this point is that intentional properties fail to *supervene* on neurophysiological properties (see Supervenience of the mental). This means that two individuals can be identical with respect to their neurophysiological properties while differing in their intentional properties. If neurophysiological properties are causally sufficient for behaviour then it is difficult to see how intentional mental properties can be causally relevant.

Subjectivity. Subjectivity involves the idea that experience is from a point of view. It has been argued that consciousness is essentially subjective and that this subjectivity cannot be accounted for in terms of physical (or any essentially objective) properties of any kind. One line of thought in support of this view is due to Nagel (1974) and Jackson (1986). According to them, it is possible to know all the physical and functional facts concerning the operation of human brains without, for example, knowing what it is like subjectively to experience vertigo. Assuming that knowing what it is like to experience vertigo is itself knowing a fact and that knowledge of a fact is not dependent on how the fact is described (both of which assumptions are questionable), it follows that the property of experiencing vertigo is not a neurophysiological or functional property. At best there are basic nomological connections between objective neurophysiological properties and subjective consciousness (see Colour and qualia; Qualia).

4 Causal relevance

If mental properties are not neurophysiological properties, then how can they be causally responsible for behaviour? One possibility is that they possess *emergent* causal powers. Mental property emergentism means that mental and *not* neural properties are responsible for grounding the causation of behaviour. However, this again violates the causal completeness of physics. So mental property emergentism is not a viable option.

Kim (1993) has argued that the causal completeness of physics *excludes* mental properties from causing behaviour. To suppose otherwise, he claims, is to suppose that behaviour is causally overdetermined by both neurophysiological and mental properties. More generally, Kim claims that the causal completeness of physical properties excludes properties not reducible to fundamental physical properties from possessing causal powers. Kim's conclusion contains both bad and good news for mental causation. On the one hand, it asserts that physical properties, and not mental properties, cause behaviour. On the other hand, mental properties are no worse off than other properties not reducible to fundamental physical properties. In particular they are no worse off, as far as Kim's argument is concerned, than neurophysiological properties since it is likely that these are not reducible to fundamental physical properties either.

A natural response to Kim's argument is to grant that non-reducible macro-properties do not possess causal powers in the way that fundamental physical properties do, but then to attempt to characterize a notion of causal relevance that permits macro-properties to be causally relevant. The motivation for this is the fact that we typically distinguish among an event those macroscopic properties which are, and those which are not, *relevant* to its having certain effects. For example, consider the event of Alma's sneezing and two of its properties - discharging a virus into the air and being loud. The first property but not the second is causally relevant to Bill contracting a cold, and so explains why. The second property but not the first is causally relevant to Hilary saying '*Gesundheit!*', and so explains why. The issue of mental property epiphenomenalism then is the question of whether mental properties are causally relevant to behaviour. If they are not, then it follows that mental properties do not causally explain behaviour; and this has significant consequences for our explanatory practices.

Whether or not mental properties are causally relevant to behaviour depends not only the nature of mental properties, but also on the account of causal relevance. Kim (1993) claims that all higher-level causation including mental causation rides piggy-back on more basic causal relations. He calls such higher-level causation 'supervenient causation' (see Supervenience). However, although higher-level causation may be supervenient causation, the relation is too weak to ground a robust notion of mental causation (Lepore and Loewer 1989). A further condition might be supplied by spelling out causal relevance in terms of counterfactuals (see Counterfactual conditionals). An initial, although not completely satisfactory, proposal (Lepore and Loewer 1987; Horgan 1989) is that an instantiation of *F* is causally relevant to an instantiation of *G* if the counterfactual 'If *F* had not been instantiated then *G* would not have been instantiated' is true. Others appeal to the subsumption of events with their properties under general laws, which may be 'strict' (as in Davidson 1984) or looser, *ceteris paribus* ones (as in Fodor 1989). While deciding among these alternatives will make a difference to the question of mental causation, the question clearly depends upon issues independent of the decision.

See also: Determinism and indeterminism; Dualism; Free will; Supervenience of the mental

BARRY LOEWER

References and further reading

Baker, L.R. (1993) 'Metaphysics and Mental Causation', in Heil and Mele (eds) *Mental Causation*, Oxford: Oxford University Press.(Defends mental causal explanations and non-reductive physicalism.)

Block, N. (1990) 'Can the Mind Change the World?', in G. Boolos (ed.) *Meaning and Method*, London: Methuen. (Interesting discussion of the problems posed by functionalism for mental causation.)

Burge, T. (1979) 'Individualism and the Mental', *Midwest Studies in Philosophy* 4: 73-121.(Another source, along with Putnam (1975), of thought experiments which are taken to support externalism.)

Burge, T. (1993) 'Mind-Body Causation and Explanation', in Heil and Mele (eds) *Mental Causation*, Oxford: Oxford University Press.(Defends folk-psychological causal explanations and argues that they are not incompatible with materialism properly understood.)

Davidson, D. (1984) *Essays on Actions and Events*, Oxford: Oxford University Press.(Davidson's important papers on philosophy of mind in which he develops the position he calls 'anomalous monism'.)

Dretske, F. (1988) *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: MIT Press.(Develops his view that intentional mental properties are structuring causes.)

Fodor, J. (1989) 'Making Mind Matter More', *Philosophical Topics* 17: 59-79.(Develops a nomological account of the causal relevance of properties and uses it to defend the causal relevance of intentional properties.)

Horgan, T. (1989) 'Mental Quasation', *Philosophical Perspectives* 3: 47-76.(Develops a counterfactual account of the causal relevance of properties.)

Jackson, F. (1986) 'What Mary Didn't Know', *Journal of Philosophy* 83: 291-5.(A famous argument for the view that subjective mental facts are not necessarily entailed by objective physical facts.)

Kim, J. (1993) *Supervenience and Mind*, Cambridge: Cambridge University Press.(Important papers exploring the extent to which physicalism can be formulated in terms of supervenience and developing the idea of supervenient causation as a solution to the problem of mental causation.)

Lepore, E. and Loewer, B. (1987) 'Mind Matters', *Journal of Philosophy* 84: 630-41.(This paper initiated much of the current discussion. It develops a counterfactual account of causal relevance and argues, contrary to the prevailing view, that Davidson's anomalous monism is compatible with the causal relevance of intentional mental properties.)

Lepore, E. and Loewer, B. (1989) 'More on Making Mind Matter', *Philosophical Topics* 17: 175-91.(Further

development of a counterfactual account of causal relevance.)
Ludwig, K. (1994) 'Causal Relevance and Thought Content', *Philosophical Quarterly* 44: 334-53.(Argues that counterfactual and nomological accounts of causal relevance are inadequate.)
McLaughlin, B. (1989) 'Type Epiphenomenalism, Type Dualism and the Causal Priority of the Physical', *Philosophical Perspectives* 3: 109-35.(Develops a nomological account of causal relevance.)
Nagel, T. (1974) 'What it's Like to be a Bat', in D. Rosenthal (ed.) *The Nature of Mind*, Oxford: Oxford University Press, 1991.(A famous argument for the view that subjective mental facts are not necessarily entailed by objective physical facts.)
Putnam, H. (1975) 'The Meaning of "Meaning"', in *Philosophical Papers*, vol. 2, *Mind, Language, and Reality*, Cambridge: Cambridge University Press.(The classic statement of the 'twin earth' case and the intuitions behind 'externalist' accounts of intentional content.)
Sosa, E. (1984) 'Mind-Body Interaction and Supervenient Causation', *Midwest Studies in Philosophy* 9: 271-81. (Poses difficulties for accounts of mental causation.)
Yablo, S. (1992) 'Mental Causation', *Philosophical Review* 101: 245-80.(Develops a counterfactual analysis of causation and applies it to mental causation.)

Mental illness, concept of

The mad were once thought to be wicked or possessed, whereas now they are generally thought to be sick, or mentally ill. Usually, this is regarded as a benign decision by a more enlightened age, but some see it as a double-edged sword - one that simultaneously relieved and robbed the mad of responsibility for their actions, eventually delivering more compassionate treatment, but also disguising value-laden judgments as objective science. The issue is made more difficult by the diversity of conditions classified as psychiatric disorders, and by the extent to which their causes are still ill understood. But the difficulty is also conceptual: what, after all, is physical illness? People usually agree that it involves abnormal body functioning, but how do we decide what is normal functioning? And even supposing that we know what we mean by a sick body, is there a parallel notion of a sick mind that is more than metaphor?

1 Physical illness and normal functioning

Somatic medicine is both applied biology and a social institution charged with curing and caring for people with complaints that have physiological causes. The ill-health concepts - illness, disease, injury, disability, and so on - reflect this dual nature.

On the one hand, it seems that some degree of biological functional impairment is a necessary condition for ill-health of any kind. That is why childbirth, teething and large noses can cause distress without counting as ill-health, for although they have physiological causes, they do not *per se* involve malfunctioning. On the other hand, the various ill-health concepts introduce additional social, ethical, or clinical considerations. For instance, injuries generally result from a visible collision with the environment, while diseases tend to be heritable or infectious and have less visible causes (for example, micro-organisms, viruses, genes). Furthermore, an illness is something a person has (your little finger can be injured or diseased, but not ill), and illness qualifies you for the sick role, which involves being excused from many normal duties and deserving of special care.

Due to the plethora of ill-health concepts, philosophers often use illness as a term of art - a catch-all that includes all states of ill-health. Neither in this sense, nor in its ordinary sense, is it controversial that it is, in part, a value-laden concept, or that it involves a core of biological dysfunction. The main controversy concerns the notion of dysfunction: is it also value-laden, or can a completely descriptive account of it be given?

Talk about biological functions is clearly normative in some sense. We ask of our bodily parts. Is it functioning properly? What is it supposed to do? Is it malfunctioning? But what is the basis of such norms?

Robert Cummins (1975) argues that functions have their most important explanatory role in biology in functional analysis - a recursive conceptual decomposition of a system and its activities into simpler and simpler subsystems and simpler and simpler tasks. Functions are causal contributions to some activity of a system that is explained by such an analysis, he says. The contribution a part makes is an objective feature of the system so described, but the activity we choose to explain depends on our interests, he adds: in physiology, we are interested in survival and reproduction and contributions to these (see Functional explanation).

Functional analysis is certainly important to biology, but many think that Cummins has not captured its normative dimension. There are numerous problems. For instance, we are most interested in how people die of cancer, but a tumour’s causal contribution to death, although often complexly achieved and a candidate for a Cummins-style analysis, is not an instance of proper functioning.

Normativists claim that normal functioning is further value-laden: while we are deeply interested in both life and death, ease and pain, fertility and infertility, capacity and incapacity, we value the first of each pair, not the second, and that is why contributions to the one, and not the other, are cases of proper functioning, they say. In support, they emphasize the cultural relativity of disease classifications. (Engelhardt’s history of masturbation in Western medicine is a good example, see Engelhardt and McCartney 1981). But this evidence is also compatible with the claim that doctors can make mistakes. If there are descriptive biological norms, it does not follow that doctors will never be biased, confused or even ignorant, when trying to apply them.

Naturalists claim that there is a purely descriptive notion of normal functioning. Christopher Boorse (1976) offered

a theory somewhat like Cummins's, but the goals of survival and reproduction were, he argued, biologically given. As natural selection continually modifies a species-design in favour of an individual's survival and reproduction, these are the goals towards which organisms are goal-directed, he says. To abstract away from idiosyncratic deviations, Boorse suggested that the function of a trait is, roughly, its species-typical contribution to the individual's survival and reproduction, but *prima facie* this seems problematic, for deviations from proper functioning can be typical (for example, in epidemics or pandemics).

Aetiological theories (following Wright 1973) have become popular. According to these, the proper function of a trait is what it was selected for. On this kind of account, it is the function of the kangaroo's pouch to protect her young because that is what the pouches of ancestral kangaroos did that caused the underlying genotype to be selected.

Some are dissatisfied with both of these naturalist theories (see Bigelow and Pargetter 1987, but also Neander 1991) and offer still further suggestions for how biological norms should be understood.

A general feature of naturalist accounts is the claim that systems that have proper functions are the outcome of, or are subject to, selection. For biological functions, the relevant selection process is natural selection: a blind, impersonal and decidedly amoral force. If so, biological norms are not determined by what we value, nor by what God values, but rather by what matters for natural selection. There is no implication, therefore, that what is normal or natural is to that extent desirable. Our values need not be those of natural selection (hence contraception). Indeed, there is no deducing 'ought' from 'is' here, on either type of account. According to the normativist, judgments about normal functioning are value-laden anyway, and so judgments concerning them are as contentious as value-laden judgments generally are.

2 Mental illness

There is no doubt that mental illness can involve great distress and incapacity, or that it can be life-threatening. But does it involve dysfunction in the sense we have been discussing?

A partial answer is easy. Some psychiatric disorders are uncontroversially brain diseases that belong to medicine proper, for they clearly involve straightforward physiological dysfunction - the neurones having atrophied (as in Alzheimer's Disease) or some such thing. The controversial cases are those where, if there is any dysfunction characteristic of the purported disorder, it can only be described in psychological, not physiological, terms (that is, in terms of the conscious or unconscious beliefs and desires of the person). For instance, if a human being desires sex with a five-year-old child, we describe that desire as sick. But is this description merely metaphorical, or is there a genuinely medical judgment to be made here, as well as a moral one? (The causes of mental illness are often complex and multifarious, and there are many mixed cases between the two extremes just mentioned - the organic syndromes and the purely psychological disorders.)

Whether there can be specifically psychological dysfunction will depend partly on the nature of mind and partly on the nature of functions. For instance, if the best theory of functions gives a central place to functional analysis and to what a thing is selected for, it will depend on whether the mind lends itself to a functional analysis of the appropriate kind, and on whether (or to what extent) the mind is subject to selection.

Even if there is specifically psychological dysfunction, (somatic) medicine might still be irrelevant to its treatment, and so psychiatry might still need to be de-medicalized, as some people argue. These days, the mental illness debate is generally based on the assumption that the mind is the brain, and that mental processes are brain processes (see Mind, identity theory of). But there is still much controversy regarding how talk of the one relates to talk of the other. Boorse (1987), for example, argues that psychology and biology are autonomous disciplines. Expertise in somatic medicine is therefore irrelevant to the treatment of psychological dysfunction, he argues, and there is no call for psychiatrists to have (somatic) medical training (although the latter is needed in a supporting role, for example, to first rule out the possibility of physiological causes.)

Others see psychology and biology as more intimately integrated. One popular, but highly controversial, approach to describing the mind is to try to characterize mental states and processes using a biological notion of a proper function (see Semantics, teleological). Many also believe that much of our psychology (learning processes, concepts and so on) is innate and part of our biological endowment (see Nativism), and hence presumably, but not

inevitably, the result of natural selection.

These strands in contemporary philosophy of mind are the ones that promise to provide the most literal interpretation of mental illness, but until we have a more complete understanding of the theoretical details, as well as of the working of the mind, the implications for particular conditions currently classified as psychiatric disorders may well remain obscure.

3 Related ethical issues

Resistance to classifying mental problems as medical problems often centres around concern for the autonomy of, and respect for, persons.

Except in so far as we blame someone for becoming ill, we do not as a rule blame them for the symptoms of their illness (for example, we might blame someone for smoking and so for developing emphysema, but we do not otherwise blame them for coughing). This seems to imply that we are not held responsible for symptoms, and mental illness is often thought to excuse a person from responsibility. The other side of this double-edged sword is that the agency of those labelled mentally ill is assumed to be diminished, and their rights as persons (for example, to choose for themselves what is in their own best interest) may consequently be denied.

There are a number of reasons why the picture just described is too simple, however. For one thing, symptoms are not usually actions. But our behaviour could be an action that is caused by beliefs and desires, and it could also be a symptom, because the beliefs and desires themselves are caused in an abnormal way - they might be caused by depression due to hormonal imbalance, for example. It is unclear why those who act from beliefs and desires that are symptoms of either somatic or mental illness should be excused just because of this fact. If a paedophile’s desire to have sex with children is no more compelling than the embezzler’s desire for money, or the jealous husband’s desire for revenge, paedophiles are equally responsible for their actions, some argue (see Feinberg 1981). On this view, mental illness provides no special ground for excuse. It excuses only if it causes the agent to be blamelessly ignorant or to suffer irresistible compulsion, which are sufficient excuses in their own right.

Similar care has to be taken with the difficult issue of the competence (and civil rights) of the mentally ill. Much mental illness is localized or episodic, and will often be irrelevant to the person’s capacity to make reasonable decisions about, for example, voting in an election, deciding on treatment, or selling the family home. Hard issues arise, however, when a mentally ill person is judged to be generally incompetent. Should we give them similar rights to those we give to children, even if their condition is believed to be irreversible, for example?

See also: Foucault, M. §2; Moral agents; Psychoanalysis, post-Freudian; Responsibility

KAREN NEANDER

References and further reading

Bigelow, J. and Pargetter, R. (1987) ‘Functions’, *Journal of Philosophy*, 84: 181-96.(Clear summary of objections to aetiological theories of functions. Offers alternative naturalist theory.)

Boorse, C. (1976) ‘What a Theory of Mental Health Should Be’, *Journal for the Theory of Social Behaviour* 6: 61-84.(Introductory. Combines functionalist theory of mind with his causal contribution theory of functions.)

Cummins, R. (1975) ‘Functional Analysis’, *Journal of Philosophy* 72: 741-65; repr. in E. Sober (ed.) *Conceptual Issues in Evolutionary Biology*, Cambridge, MA: Bradford MIT, 1994, 2nd ed., 49- 69. (Medium difficulty. Describes functional analysis.)

Engelhardt H.T. and McCartney J. (1981) ‘Concepts of Health and Disease’ in Caplan, A. (ed.) *Concepts of Health and Disease*, Reading, MA: Addison-Wesley.(Useful interdisciplinary collection on the history, analysis and social role of the concepts of health, illness and disease.)

Feinberg, J. (1981) ‘What Is So Special About Mental Illness?’ *Reason and Responsibility: Readings in Some Basic Problems of Philosophy*, Belmont, CA: Wadsworth Publishing Company, 5th edn, 464-73.(Clear and simple introduction to ethical issues concerning mental illness, competence and responsibility.)

Neander, K. (1991) ‘Functions as Selected Effects’, *Philosophy of Science* 58: 168-84.(Medium difficulty. Detailed defence of aetiological theories.)

Szasz, T. (1961) *The Myth of Mental Illness*, New York: Hoeber-Harper.(Best known and rather polemical work of the psychiatrist, Thomas Szasz. Argues that we should de-medicalize psychiatry.)

Wright, L. (1973) ‘Functions’, *Philosophical Review* 82: 139-68; repr. in E. Sober (ed.) *Conceptual Issues in*

Evolutionary Biology, Cambridge, MA: Bradford MIT, 1994, 2nd ed., 27- 47.(Introduction to aetiological theories of functions.)

Mental states, adverbial theory of

According to the adverbial theory, there are no mental objects of experience, no pains, itches, tickles, after-images, appearances. People certainly feel pains and have after-images; external objects certainly present appearances to people viewing them. But pains, after-images, and appearances are not real things. Statements which purport to be about such mental objects have a misleading grammatical form. In reality, such statements are about the ways in which people experience or sense or feel.

1 Motivation

Suppose you are having a tooth drilled. Suddenly you feel intense pain. This is an objective fact about you, not dependent for its existence on anyone else seeing or thinking about your situation. But the pain you are feeling - that particular pain - is private to you. It is yours alone, and necessarily so. No one else could have that particular pain. Of course, conceivably somebody else could have a pain which felt just like your pain, but only you could have that very pain. What is true for this one pain is true for all mental objects of experience, for example after-images, appearances, itches, tickles. None of these items of experience can be shared. I can’t have your images or feel your tickles. Physical objects, however, are not necessarily private in this way. Any physical object that is mine could be yours, my watch, for example, or even my heart (see Privacy).

There is another striking fact about mental objects of experience, or phenomenal objects, as they are often called. Any pain or itch or image is always *somebody’s* pain or itch or image. Likewise, any appearance is always an appearance *to* someone or other. Each mental object of experience necessarily has *an* owner, unlike ordinary physical objects, for example, trees or tables or even legs. Legs can exist amputated, and trees and tables can belong to no one at all.

Any philosopher who wants to view the mind, and its contents, as natural physical phenomena faces a problem with respect to the necessary privacy and necessary ownership of phenomenal objects. The adverbial theorist solves this problem by arguing that while it is certainly true that people feel pains and itches, and have images, this way of talking is misleading. Really there are no such things as pains, itches, and images. The world itself contains no phenomenal objects. So, there is, in reality, nothing to have the physically problematic properties.

2 Elaboration

How can it be true that people feel pains, even though there really are no pains for them to feel? The answer, according to the adverbial theorist, is that our ordinary way of expressing ourselves here is deceptive. Grammatical form is sometimes misleading. Just as it can be true that the average family has 1.2 children even though there really are no families with a fraction of a child, so too you can feel an intense pain even though there really are no pains.

But what is it that we are asserting, then, when we say, in ordinary parlance, that someone *has* a blue after-image or a throbbing pain or a terrible itch? In the case of the claim that the average family has 1.2 children, the answer is reasonably obvious. We are talking about numbers, and what we are really asserting is that the number of children divided by the number of families equals the number ‘1.2’. In the case of phenomenal talk, however, if its grammatical form is indeed misleading, the answer is much less obvious. It is to be found, according to the adverbial theorist in our ordinary talk of other things - of smiles and limps, of voices and stutters.

Suppose we ask ourselves whether there really are, in the world, any such things as limps. If we reply ‘Yes’, then we seem to face some puzzling questions just like the ones raised for phenomenal objects. If Patrick has a pronounced limp, then can Paul have that very same limp (the token limp Patrick has, not one just like it but numerically different)? Can Patrick’s limp exist without anyone at all having it?

One reasonable response to these puzzling questions is to deny that there really are such things as limps, to which people are related by the ’having’ relation. On this view, when we say that Patrick has a pronounced limp, what we are really saying is that Patrick *limps* in a pronounced manner. Likewise for smiles, voices, and stutters. If Patrick also has a charming smile, a loud voice, and a noticeable stutter, what is really the case is that he smiles charmingly, speaks loudly, and stutters noticeably. There are no smiles, voices, and stutters, conceived of as things

people have.

Analogously, according to the adverbial theorist, there are no pains, itches, or visual images, conceived of as objects people have. Instead, to have a terrible itch is just to itch terribly; to have a throbbing pain is to hurt throbbingly or in a throbbing manner. Finally, to have a blue after-image is just to sense in a certain way, a way we may dub ‘bluely’. Now the problem for physicalism supposedly disappears. There are no phenomenal *objects* as such, and so there is no question of trying to accommodate their necessary privacy and ownership in the physical world.

Unfortunately, life is not that easy. To begin with, talk of people sensing bluely or hurting throbbingly is hardly everyday. It cries out for further explanation. Secondly, if there are no phenomenal objects, then just what is it that makes phenomenal talk true? If, when I have a terrible itch, it is really true that I itch terribly, then what exactly is it about me which *makes* this true? Some account is needed here in order to show that the possession by me of a terrible itch has really been repudiated and not just hidden under a verbal smokescreen.

At this point, the adverbial theory of sensation or experience, splits into two. To understand these alternatives and their respective costs and benefits, it is necessary to make some brief general remarks about adverbs.

There are two major theories of adverbs in the philosophical literature: the predicate operator theory and the event predicate theory (see Adverbs). According to the latter view, adverbs, or at least a very wide range of adverbs, are best understood as adjectives that are true of events. Consider, for example,

(1) Patrick is stuttering noticeably.

If we are to bring out what it is that makes (1) true, we should restate it, on the event predicate theory, as

(1a) There is an event of stuttering, which has Patrick as its subject, and that event is noticeable.

Similarly,

(2) Patrick is speaking loudly.

should be recast as

(2a) There is an event of speaking, which has Patrick as its subject, and that event is loud.

So, the adverbs ’noticeably’ and ’loudly’ are really functioning as adjectives or predicates that apply to events, in the one case, to a stuttering, and in the other, to a speaking. What must exist, then, for Patrick to stutter noticeably or speak loudly is a noticeable stuttering or a loud speaking, of which Patrick is the subject.

This view has a number of virtues, not the least of which is that it requires no alteration, or addition, to the materials available in first order quantificational logic, as far as the formal evaluation of arguments in which sentences like (1) and (2) occur (see Predicate calculus). Secondly and relatedly, it provides us with a very simple explanation of why (1), for example, entails

(3) Patrick is stuttering.

(3) is reconstructed as:

(3a) There is an event of stuttering, of which Patrick is the subject.

and (3a) follows from (1) via elementary logical rules.

The alternative view takes adverbs to be operators which turn the predicates they modify into more complicated predicates. In (1), for example, ’noticeably’ is held to operate on the predicate preceding it, namely ’stutters’, and to convert that predicate into a syntactically more complicated predicate, namely ’stutters noticeably’. The idea is that adverbs stand for functions which map the properties expressed by the properties they modify onto other properties. On this view, what makes (1) true is not the fact that Patrick is the subject of an event (a noticeable stuttering) nor the fact that Patrick owns a special object (a noticeable stutter) but simply the fact that Patrick has a certain property, that of stuttering noticeably. So (1) really is a subject-predicate sentence, just as it appears to be.

There is no hidden existential quantification, as there is the event predicate theory proposal.

This approach requires the admission that the resources of standard logic do not suffice to come to grips with arguments containing adverbs, together with an account of operator detachment which will validate inferences like the one from (1) to (3) above. So, it has significant costs. On the other hand, it allows us to take sentences like (1) and (2) entirely at face value. These sentences certainly look to be of the straightforward subject-predicate variety, and, by golly, so they are on the predicate operator proposal.

On the event version of the adverbial theory of sensation, statements putatively about phenomenal objects like pains and itches are really about events like hurtings and itchings. By contrast, on the predicate operator version of the adverbial theory, statements putatively about phenomenal objects are really about people or other sentient creatures and the phenomenal properties they instantiate, properties like itching terribly or hurting throbbingly.

3 Objections

These two developments of the adverbial theory face some serious difficulties. To begin with, the predicate operator theory is hard pressed to account for the truth of ordinary statements putatively about multiple phenomenal objects. For example, suppose that after staring at a bright light and turning away, I report the following:

(4) I have four pink after-images.

It is not a straightforward matter to provide a plausible account of what in the world makes this true, given only phenomenal properties that I possess. The problem becomes especially difficult once it is appreciated that glibly proposing

(4a) I sense four-pink-ly,

or some such contrived statement as the analysis, apparently fails to account for the fact that (4) entails, for example,

(5) I have at least three after-images.

It certainly *looks* here as if, given this entailment, (4) is true if, and only if, there are four pink after-images that I have. Another related difficulty arises in connection with the following two statements

(6) I have both a red, square after-image and a blue, round after-image.

and

(7) I have both a blue, square after-image and a red, round after-image.

These statements evidently are not equivalent: one can be true without the other. But if we analyse the former by

(6a) I sense redly and squarely and bluely and roundly.

and the latter by

(7a) I sense bluely and squarely and redly and roundly.

then the difference between (6) and (7) is lost. For the order of the conjuncts in a conjunction makes no difference to the conditions under which it is true.

These problems, it should be noted, are also challenging for the event version of the adverbial theory. In this case, however, there is a reasonably straightforward reply. Allow there to be as many different sensings as there *appear* to be after-images referred to in the original statements. Then (4) becomes

(4b) There are four pink sensings of which I am the subject.

Now the entailment to (5) is easy to explain. Likewise, there is no difficulty in distinguishing (6) and (7). In both cases, there are two sensings, but they have different properties.

But does it really make sense to talk of my undergoing four simultaneous sensings, each with the same property? Some philosophers have certainly thought not. And it must be admitted that it is not at all easy to find clearcut examples from ordinary life of cases where a single individual is the subject of multiple simultaneous events of exactly the same sort (and not involving any other objects). I can listen to you with my left ear and also simultaneously listen to you with my right, or I can inhale deeply through my left nostril and at the same time inhale deeply through my right, but in both of these cases other objects are also involved (namely ears and nostrils). I can simultaneously be the subject of a singing and a walking and a smiling, say, but apparently I cannot simultaneously be the subject of three singings or three walkings or three smilings.

So, the event version of the adverbial theory faces a serious challenge here, too. And there are other problems. What, for example, can it possibly mean to call an event ’pink’ or ‘square’?

The difficulties I have raised here have been addressed by adverbial theorists. And I think it is fair to say that there is no knockdown objection to the view. But whether it really provides us with the best response to the problems posed by the necessary privacy and ownership of phenomenal objects is a matter of considerable dispute.

Finally, although the adverbial theory is usually associated with sensory experience, it is occasionally generalized so as to apply to all mental states. Under this generalization, there are no beliefs or desires or fears that people *have*, any more than there are after-images or itches or pains. As before, there are again two possible ways of developing this proposal further, one of which countenances believings, desirings, and fearings, and the other of which restricts itself to the attribution of the appropriate psychological properties.

See also: Perception: Propositional attitudes

MICHAEL TYE

References and further reading

Chisholm, R.M. (1957) *Perceiving*, London: Routledge & Kegan Paul.(Classic early presentation of the adverbial theory.)

Ducasse, C.J. (1942) ’Moore’s Refutation of Idealism’, in P.A. Schilpp (ed.) *Philosophy of G.E. Moore*, Chicago, IL: Northwestern University Press, 225-51.(Another good early statement of the view.)

Jackson, F. (1977) *Perception*, Cambridge: Cambridge University Press, 50-87.(Some incisive criticisms of the adverbial theory.)

Sellars, W. (1968) *Science and Metaphysics*, London: Routledge & Kegan Paul, 9-28. (Further development of the adverbial view.)

Tye, M. (1984) ’The Adverbial Approach to Visual Experience’, *Philosophical Review* 93: 195-225.(Detailed formulation of the metaphysics and semantics of the position, and replies to objections.)

(1989) *The Metaphysics of Mind*, Cambridge: Cambridge University Press.(Extension of the adverbial theory to all mental states.)

Mereology

Mereology is the theory of the part-whole relation and of derived operations such as the mereological sum. (The sum of several things is the smallest thing of which they are all parts.) It was introduced by Leśniewski to avoid Russell's paradox.

Unlike the set-membership relation, the part-whole relation is transitive. This makes mereology much weaker than set theory, but gives the advantage of ontological parsimony. For example, mereology does not posit the proliferation of entities found in set theory, such as $\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots$.

Mereology has occasioned controversy: over whether many things really have a mereological sum if they are either scattered or, even worse, of different categories; over the uniqueness of sums; and over Lewis' claim that the non-empty subsets of a set are literally parts of it.

1 Classical mereology

Mereology is the theory of the part-whole relation. It was introduced by Leśniewski (1916) to avoid Russell's paradox (see Paradoxes of set and property). Taking the part-whole relation as primitive, we define a proper part as a part which is not identical to the whole. We define disjointness to mean having no part in common. Then, in addition to the transitivity of the part-whole relation, the following axiom schema, unique composition, characterizes classical mereology.

If there is some *F*, there is precisely one thing (the sum or fusion of the *F*s) of which every *F* is a part and which has no part disjoint from every *F*.

The supplementation principle now follows, namely that if *a* is a proper part of *b*, then *b* has some proper part disjoint from *a*. For otherwise both *a* and *b* are sums of all the parts of *a*, contradicting uniqueness. It also follows that there is a sum of everything, *U*; that any thing *a* which is not *U* has a complement, namely the sum of everything disjoint from *a*; and that any two overlapping things *a* and *b* have a product (or intersection), namely the sum of all things which are parts of both *a* and *b*. Apart from the absence of a null individual, mereology has, with these operations, the structure of a Boolean algebra (see Boolean algebra §1).

Strictly speaking, what we take as the characteristic axiom schema will depend on the logic mereology is an extension of. In the above this is the predicate calculus with identity. Leonard and Goodman's formulation (the calculus of individuals) is an extension of (a fragment of) set theory. In that case, unique composition is the axiom that any non-empty set of individuals has a sum. Or suppose, like Lewis (1991), you are extending a logic which permits plural quantification. Then unique composition could be formulated as:

For any things there is one thing which has those things as parts and which has no part disjoint from those things.

Yet again, in a finitist mereology the formulation would be:

For any *a* and *b* there is a unique *c* of which *a* and *b* are parts and which has no part disjoint from *a* and *b*.

Classical mereology is, therefore, a family of theories, with unique composition adapted to the context.

Both the transitivity of the part-whole relation and unique composition have been rejected as counter-intuitive. Cruse (1979) gives an example, later discussed by Simons (1987: 107), which illustrates the supposedly counter-intuitive character of the transitivity of the part-whole relation: a handle is part of a door, a door is part of a house but, it might seem, the handle is not part of the house. This objection is based on a controversial interpretation of 'part' to mean 'significant part'.

Unique composition may be conveniently divided into the existence of sums and the uniqueness of these sums. Let us first consider uniqueness. The problem, as stated by Simons (1987:115-17), has to do with 'continuants', namely things which endure without having temporal parts. Continuants are themselves controversial, but let us suppose that cats are continuants. Now consider a whole cat, Tibbles, her tail, and the remainder of Tibbles, called Tib. If Tibbles loses her tail then, Simons argues, Tibbles and Tib will be non-identical things with precisely the

same parts. They are non-identical because they are identical, respectively, to Tibbles and Tib prior to tail loss, which are obviously non-identical (in other words, Tibbles is still Tibbles with or without her tail). A similar problem can be stated in modal rather than temporal terms. While there may be several solutions, we might well restrict classical mereology to categories of items which are not continuants. In that case Simons' nonclassical mereology would be appropriate for continuants.

The existence of sums of disparate collections has seemed too strong to many. Here it is important to note that mereology is demonstrably consistent, so there is no analogue of Russell's paradox. The objection, therefore, is not formal. The worst case is the existence of sums of items from different categories, such as: a kettle, the process in the kettle in which water boils, the spatiotemporal region in which the process occurs, the property of being at $100^{\circ}\mathrm{C}$, and the relation of contiguity. Even though all those items might be thought of in connection with the activity of boiling water, it seems queer to say there is something which is the sum of them.

Lewis has defended the existence of disparate sums on the grounds that describing the sum is doing nothing more than describing the parts, for there is no extra ingredient which somehow unifies the parts into a whole (1991: 79-81). He also points out that attempts to avoid the sum of a scattered collection of items of the same category will fail because in such cases queerness is a matter of degree. This leaves open, however, the possibility of restricting unique composition to sums of items all taken from the same category.

2 Mereology, set theory and atoms

An atom may be defined as something which has no *proper* part. We may ask whether there are any atoms and whether every thing is a sum of atoms. If the answer to the first question is negative then a mereology is called 'atomless'; if the answer to the second is positive then it is 'atomistic'. Now mereology does not have to be either atomistic or atomless - it might, for instance, be the case that every object has proper parts even though the space they occupy is a sum of points. It seems rather less plausible that any one category of thing (for example, physical objects) should have both atoms and atomless parts. Mereology restricted to a single category would seem, therefore, to be either atomistic or atomless. For example, mereology restricted to spatiotemporal regions is atomistic if all regions are sums of points, but atomless if points are idealized limiting cases of arbitrarily small regions (see Whitehead 1919; Whitehead §2; see also Mortensen and Nerlich 1978; Clarke 1985).

Mereology is of theoretical interest partly because it has been used to provide either a replacement for, or a formulation of, the theory of sets. Because a sum of atoms has a unique decomposition into atoms, sums behave like sets of atoms. The chief difference between mereology and set theory is that two different sets of non-atoms, say, $\{a+b, c+d\}$ and $\{a+c, b+d\}$, can have the same sum, $a+b+c+d$. It was hoped that the resources of mereology would none the less be enough to replace sets in the foundations of mathematics. Thus Goodman and Quine (Goodman 1972: 173-98) showed how the ancestral (that is, a chain of instances) of a relation could be defined using the part-whole relation, avoiding set theory. They also showed how various statements such as 'There are exactly one-third as many Canadians as Mexicans' can be handled using mereology. Their approach ran into difficulties, however, with quantified numerical formulas, such as 'For all *n*, $n+n=2n$'.

Lewis, in his version of nominalistic set theory, adopted a rather different strategy, namely to imitate set theory using the resources of mereology (see Lewis 1991: 21-8). The chief difficulty seems to have been that without the resources of set theory the materials are lacking with which to construct enough pseudo-sets. Thus a sound Ockhamist reason for preferring mereology to set theory - its inability to generate infinitely many things out of just one thing - restricts its use as a substitute.

Rather different is Lewis' later programme of treating the theory of sets and proper classes as itself an extension of mereology (1991). Excepting the special case of the empty set, we may treat subclasses as parts and thus treat set theory as the extension of mereology obtained by adjoining the singleton operator. This operator forms unit sets, which are always atoms, out of individuals. Lewis is in favour of treating it as primitive, but, relying on results by Burgess and Hazen, he also shows how we can develop an alternative, structuralist account, according to which there are many possible singleton operators, between which set theory does not choose.

See also: Logical and mathematical terms, glossary of; Structuralism

PETER FORREST

References and further reading

Burkhardt, H. (1990) 'Mereology in Leibniz's Logic and Philosophy', *Topoi* 9: 3-13.(Historical background.)

Clarke, B.L. (1985) 'Individuals and Points', *Notre Dame Journal of Formal Logic* 26: 61-75. (Recent discussion of the mereology of space.)

Cruse, D.A. (1979) 'On the Transitivity of the Part-Whole Relation', *Journal of Linguistics* 15: 29-38.(An examination of the linguistic evidence for the failure of transitivity. The author makes a useful distinction between parts and attachments.)

Eberle, R.A. (1970) *Nominalistic Systems*, Dordrecht: Reidel.(Includes simplifications to mereology based on the assumption of atomism.)

Goodman, N. (1956) 'A World of Individuals', in I. Bocheński, A. Church and N. Goodman, *The Problem of Universals*, Notre Dame, IN: University of Notre Dame Press; repr. in *Problems and Projects*, Indianapolis, IN: Bobbs-Merrill, 1972.(Along with the work of Leśniewski and Whitehead, this is one of the classics of mereology.)

Goodman, N. (1972) *Problems and Projects*, Indianapolis, IN: Bobbs-Merrill.(§4 includes two important and accessible papers: 'A World of Individuals' and, with Quine, 'Steps Toward a Constructive Nominalism'.)

Henry, D.P. (1991) *Medieval Mereology*, Amsterdam: Benjamin.(Historical background.)

Le Blanc, A. (1985) 'New Axioms for Mereology', *Notre Dame Journal of Formal Logic* 26: 437-43. (One of a series discussing the axiomatics of mereology.)

Lejewski, C. (1983) 'A Note on Leśniewski's Axiom System for the Mereological Notion of Ingredient or Element', *Topoi* 2: 63-72.(A recent paper by a distinguished mereologist in the school of Leśniewski.)

Leśniewski, S. (1916) 'Podstawy ogólnej teoryi mnogości I', *Prace Polskiego Koa Naukowe w Moskwie, Sekcya matematyczno-przyrodnicza* 2; trans. D.I. Barnett, 'Foundations of the General Theory of Sets I', in *Collected Works*, ed. S.J. Surma, J.T. Srzednicki and D.I. Barnett, Dordrecht: Kluwer, 1992, vol. 1, 129-73.(Classic paper on mereology. The word 'Sets' in the translation of the title is misleading. The paper is not about sets as we now understand them. Consult P. Simons' 'Discovering Leśniewski's Collected Works', *History of Philosophical Logic* 15: 227-35 (1994) for a useful list of typographical errors.)

Lewis, D.K. (1991) *Parts of Classes*, Oxford: Blackwell.(A lucid presentation of set theory as an extension of mereology, using the minimum of technicality; also includes an excellent introduction to mereology (based on plural quantification), a defence of classical mereology, a discussion of nominalistic set theory and some important, though technical, appendices with J.P. Burgess and A.P. Hazen.)

Mortensen, C. and Nerlich, G. (1978) 'Physical Topology', *Journal of Philosophical Logic* 7: 209-23. (Topology based on atomless mereology for space.)

Simons, P. (1987) *Parts: A Study in Ontology*, Oxford: Clarendon Press.(A most comprehensive work discussing both the Leśniewski and the Leonard and Goodman formulations of mereology, providing some criticisms of classical mereology and developing a nonclassical theory. Of necessity, much of the discussion is technical. Includes an extensive bibliography.)

Smith, B. (ed.) (1982) *Parts and Moments: Studies in Logic and Formal Ontology*, Munich: Philosophia.(Includes an extensive bibliography.)

Smith, B. (1985) *Parts and Wholes*, vol. 3, Stockholm: Forskningsrådsnämdnen.(Includes a further bibliography on mereology.)

Sobociński, B. (1934) 'O kolejnych uproszczeniach aksjomatyki ''ontologii'' prof. S. Leśniewskiego', in *Fragmenty Filozoficzne. Księga Pamiątkowa ku Uczczeniu Piętnastolecia Pracy Nauczycielskiej w Uniwersytecie Warszawskim Profesora Tadeusza Kotarbińskiego*, Warsaw: Nakładem Uczniów, 144-60; trans. 'Successive Simplifications of the Axiom-System of Leśniewski's Ontology', in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967.(One of a series, mostly in the *Notre Dame Journal of Formal Logic*, by mereologists in the school of Leśniewski, in which alternative axiomatics for mereology are discussed. For a survey, with references, see Simons (1987: ch. 2).)

Srzednicki, J.T.J., Rickey, V.F. and Czelakowski, J. (eds) (1984) *Leśniewski's Systems: Ontology and Mereology*, The Hague: Nijhoff.(Reprinted articles by mereologists in the tradition of Leśniewski, including Lejewski, Supecki and Sobociński.)

Whitehead, A.N. (1919) *An Enquiry concerning the Principles of Natural Knowledge*, Cambridge: Cambridge University Press.(Includes a formulation of mereology independent of, and almost as early as, Leśniewski's, with application to the theory of atomless spacetime.)

Merleau-Ponty, Maurice (1908-61)

Merleau-Ponty belongs to the group of French philosophers who transformed French philosophy in the early post-war period by introducing the phenomenological methods of the German philosophers Husserl and Heidegger. His central concern was with 'the phenomenology of perception' (the title of his major book), and his originality lay in his account of the role of the bodily sense-organs in perception, which led him to develop a phenomenological treatment of the sub-personal perceptions that play a central role in bodily movements. This account of the sub-personal aspects of life enabled him to launch a famous critique of Sartre's conception of freedom, which he regarded as an illusion engendered by excessive attention to consciousness. None the less, he and Sartre cooperated for many years in French political affairs, until Merleau-Ponty became exasperated by the orthodox Marxism-Leninism of the French Communist Party in a way in which Sartre, who remained a fellow-traveller, did not. As well as several substantial political essays, Merleau-Ponty wrote widely on art, anthropology and, especially, language. He died leaving some important work incomplete.

Although his work is still esteemed within the French academic establishment, his influence in France has waned, because of a tendency there to study his German forebears almost to the exclusion of all else. But elsewhere, and most notably in the USA, Merleau-Ponty's work is widely studied, especially now that questions about the distinction between personal and sub-personal aspects of life have become so prominent.

1 Life

Merleau-Ponty studied philosophy at the École Normale Supérieure, graduating in 1930. As a student he was quickly drawn towards the works of the German phenomenologists, having attended Husserl's 1929 lectures in Paris (the 'Cartesian Meditations') and studied Heidegger's writings (see Husserl, E.; Heidegger, M.). He wrote his first book, *The Structure of Behavior*, before serving briefly in the French Army in 1939-40. During the German Occupation he taught at the Lycée Carnot in Paris and wrote his major work, *The Phenomenology of Perception*, which was published in 1945. In the same year he took a position at the University of Lyon, before returning to Paris in 1950 as Professor of Psychology at the Sorbonne; in 1952 he moved to the prestigious Chair of Philosophy at the Collège de France, a position which he held until his early and unexpected death in 1961.

2 Mind and body

Much of Merleau-Ponty's first work, *The Structure of Behavior* (1942) is devoted to a detailed critical discussion of physiological psychology and the attempt to provide on its basis a reductive explanation of behaviour. In developing his argument Merleau-Ponty draws on Gestalt psychology and especially K. Goldstein's *The Organism* which emphasizes the holistic features of the life of organisms. Merleau-Ponty takes over Goldstein's holism and incorporates it into what he terms a 'dialectical' conception of the structures of behaviour, according to which as organisms evolve and become more sophisticated, higher 'forms' of behaviour develop which transform the life of the organism. So the new capacities characteristic of these higher forms are not simple additions to an otherwise unaltered neurophysiology; instead, through a process of 'dialectical' assimilation, these new capacities bring with them changes in the functioning of the underlying neurophysiology (see Gestalt psychology).

Merleau-Ponty argues that this dialectical approach enables him to reject reductive theories without invoking the kind of vitalism espoused by Bergson (see Vitalism). Equally, he argues, it provides a way of accommodating the phenomena of consciousness without dualist metaphysics: for, he writes, 'man is not a rational animal. The appearance of reason and mind does not leave intact a sphere of self-enclosed instincts in man' ([1942] 1963: 181). Thus there can be no partition of 'mind' from 'body' in human life: the mental is intrinsically bodily, a theme Merleau-Ponty illustrates through the role of the sense-organs in perception. Equally the body is intrinsically mental, in the sense that there can be no adequate understanding of human behaviour that does not conceptualize it as the behaviour of a normally rational agent. Merleau-Ponty's comment on Watson's behaviourism shows his aspirations clearly:

> what is healthy and profound in this intuition of behavior found itself compromised by an impoverished philosophy…. When Watson spoke of behavior he had in mind what others have called *existence*; but the new notion could receive its philosophical status only if causal or mechanical thinking were abandoned for

dialectical thinking.

(Merleau-Ponty [1942] 1963: 226)

3 *The Phenomenology of Perception*

Merleau-Ponty ends *The Structure of Behavior* with some enigmatic remarks to the effect that, just as there is no satisfactory causal account of perception, a realist account of the perceived world is equally unsatisfactory. With these remarks Merleau-Ponty points towards a main theme of his most important work, *The Phenomenology of Perception* (1945). An important difference, however, between the two books is that, as its title and preface (‘What is Phenomenology?’) indicates, the approach adopted in the second book is explicitly ‘phenomenological’. The central theme of Merleau-Ponty’s phenomenology is that a full understanding of the natural and social worlds which we inhabit inevitably leads back to aspects of our experience through which, in some sense, meaning is bestowed upon the objects of experience. Thus he sees the primary task of philosophy as one of quasi-Platonic *anamnesis* (see Plato §§10-11), of re-awakening within us a recognition of these meaning-bestowing aspects of our own experience so that we can grasp how we are integrated into these worlds and how they, in their turn, are dependent upon us: philosophical reflection, he writes, ‘steps back to watch the forms of transcendence fly up like sparks from a fire; it slackens the intentional threads which attach us to the world and thus brings them to our notice’ ([1945] 1962: xiii). In developing this phenomenology Merleau-Ponty often expresses his debts to Husserl’s late works, especially the unpublished papers which he had studied at Louvain. But how far, in fact, Merleau-Ponty is developing themes from Husserl’s writings is disputable. He himself indicates his rejection of the transcendental subjectivism that is a prominent theme of most of Husserl’s writings; similarly when he proclaims the impossibility of completing the phenomenological reduction and affirms the inseparability of essence and existence he is implying that Husserl’s project of essentialist phenomenological analysis can never be accomplished. Thus it is arguable that Merleau-Ponty’s conception of phenomenology is in fact closer to that characteristic of Heidegger’s *Being and Time*, and one can certainly use Merleau-Ponty’s detailed accounts of experience as a way of filling out Heidegger’s all-too-abstract characterization of human life.

Merleau-Ponty’s concentration upon perception rests on a thesis which is fundamental to his phenomenology, namely that perception is ‘transcendental’ in the sense that it cannot be adequately understood from within a fully objective, scientific conception of human life. In part this thesis just draws on the argument of *The Structure of Behavior*; but at a deeper level Merleau-Ponty argues that because perceptual experience is epistemologically fundamental it cannot be the case that perception itself is fully comprehended within the explanatory perspective of natural science, since this employs a conception of the world which draws on features that owe their significance to perceptual experience itself. Merleau-Ponty likes to adapt a phrase from the poet Paul Valéry to express this feature of perception: it is, he says, ‘the flaw in this great diamond’ since ‘we can never fill up, in the picture of the world, that gap which we ourselves are’ ([1945] 1962: 207). Merleau-Ponty’s argument is, therefore, that the realist naturalizing project faces an inescapable blind-spot in connection with perception. Whether Merleau-Ponty is right about this remains disputable - his critics will maintain that, once we separate epistemological from metaphysical priorities, we can see that it is legitimate to regard our own experiences as facts caught up within, and dependent upon, the causal order of the natural world, even if we also acknowledge that our beliefs concerning this causal order rest upon perception.

This is the kind of issue which realists and idealists can debate endlessly, but whatever the outcome of that debate there can also be agreement that Merleau-Ponty’s line of thought leads into one of the most innovative programmes of twentieth-century idealism. Central to this is his discussion of the status of the body. Merleau-Ponty once again uses Goldstein’s work - in this case his studies of a patient (Schneider) who had suffered brain injuries during the First World War. But what is new in *The Phenomenology of Perception* is the use that Merleau-Ponty makes of his phenomenology in the characterization of Schneider’s case and then, by an inversion, his use of Schneider’s case to revise the phenomenological approach itself. Schneider’s injuries, he argues, have brought with them the loss of the sub-personal perceptual fields, bodily spaces and tacit anticipations of possibilities that underpin ordinary human agency. So Schneider’s disability is not merely neurophysiological; nor is it just a disorder of a consciousness detached from behaviour; instead it bears witness to the normal sub-personal union of mind and body through which a form of intentionality is expressed in unreflective but organized bodily movement. Once the further step is taken of recognizing that this is the basic form of

intentionality, it follows that the phenomenological perspective itself needs to be relocated from the personal sphere of explicit thought to the sub-personal domain of bodily movement; thus, as Merleau-Ponty puts it, Schneider's case 'forces us to acknowledge an imposition of meaning which is not the work of a universal constituting consciousness' ([1945] 1962: 147) (see Intentionality).

Merleau-Ponty characterizes this revised phenomenological perspective as one in which the body is conceived as 'a natural self and, as it were, the subject of perception' ([1945] 1962: 206). Because of his transcendentalism concerning perception, the body thus conceived is not the body as characterized within the objective perspective of the natural sciences: 'what prevents it ever being an object…is that it is that by which there are objects' ([1945] 1962: 92). His term for the body thus conceived is 'the phenomenal body' ([1945] 1962: 106) and he argues that by following the intentional threads which link the phenomenal body to the world we will be led to recognize the inadequacy of a purely objective conception of the world as a totality of objects, and to replace it with that of the 'phenomenal field', conceived as 'the horizon latent in all our experience and itself ever-present and anterior to every determining thought' ([1945] 1962: 92).

At this point the similarity with Kantian themes concerning space will be obvious (see Kant, I. §5), and Merleau-Ponty provides an extended account of the a priori status of the practical horizon which, he holds, enters into all our perceptual fields. But what is distinctive about Merleau-Ponty's account is its integration into a phenomenological psychology which preserves the basic metaphysics of transcendental idealism by making the body the transcendental subject. Arguably, however, in *The Phenomenology of Perception* Merleau-Ponty still replicates Husserl's problematic transcendental/empirical division of the self with his distinction between the 'phenomenal' and the 'objective' body. Hence it is notable that when he returns to the topic in his final work, *The Visible and the Invisible*, he attempts to go beyond this distinction by arguing that there is a yet more fundamental conception of the body - as 'flesh', which is a conception of the body that admits both an objective specification as 'a thing among things' and a phenomenal specification as 'what sees and touches' ([1964] 1968: 137). Indeed, despite the title of the book, it is significant that Merleau-Ponty here treats the sense of touch as phenomenologically fundamental and even, implausibly, construes sight as a species of touch (as 'a palpation with the look' ([1964] 1968: 134)); for touch has often been regarded as the sense which by manifesting the independent resistance of things reveals most clearly their objectivity. But because the book is incomplete it is not clear just how far this rebalancing of emphasis, to give equal weight to the 'objective' and the 'phenomenal' sides of life which are 'intertwined' in the human body, implies the need for a reworking of his whole phenomenology in a more objectivist direction. But it is clear, I think, that the position remains recognisably idealist in spirit, partly indeed through his explicit affirmation of the ideality of 'the flesh' ([1964] 1968: 152).

4 Consciousness and freedom

In the final chapters of *The Phenomenology of Perception* Merleau-Ponty uses his account of the sub-personal elements of life to elucidate the characteristic marks of personal life - self-consciousness and freedom. His discussion of these themes is, in part, an argument against their treatment by Sartre in *Being and Nothingness*, which was published two years earlier in 1943. Central to Sartre's account was the thesis that consciousness has a distinctive 'being' as 'being-for-itself' - that is, self-conscious - which is supposed to yield the conclusion that consciousness manifests an inescapable and absolute freedom (see Sartre, J.-P. §3). As against Sartre, therefore, Merleau-Ponty argues that both self-consciousness and freedom are achievements that cannot be understood in isolation from the sub-personal aspects of life from which they are absent. They do not constitute an inescapable fate to which we are condemned; instead they are just the high points in the ebb and flow of our existence.

In the case of consciousness Merleau-Ponty argues that it is because 'all consciousness is, in some measure, perceptual' ([1945] 1962: 395), in the sense that it draws upon our habitual sub-personal experience of the world, that self-consciousness is both capable of genuinely extending our knowledge and also vulnerable to error and illusion. A form of self-consciousness that was not thus vulnerable would be so totally detached from this tacit background that it would lack the concepts with which this background furnishes us and could, therefore, tell us nothing about ourselves. Similarly the exercise of genuine freedom presupposes a background engagement with the world that is not itself an exercise of freedom. If one seeks, as Sartre does, to find freedom in all acts of consciousness, one empties of any content the distinction between freedom and its absence by misconstruing the absence of a choice to act as a choice not to act. Instead, we need to acknowledge the impersonal significance of

things which arises from our sub-personal needs, habits and constraints; only against this background can a ‘strictly individual project’ stand out briefly as an exercise of freedom before its significance too is compromised by its incorporation into the general run of events.

5 Language

For many years Merleau-Ponty aimed to extend his account of our perception of ‘visible’ objects to an account of speech about ‘invisible’ truths (hence the title of his last book). His most extended discussion occurs in *The Prose of the World*, which was written in 1950-2, though published only posthumously. The central claim he makes here is that speech has a transcendental role comparable to that of perception in the constitution of truth. We normally fail to notice this because just as naïve perception encourages the realist illusion of perfect objectivity, our normal uses of language to talk about things encourage a similar realist illusion of language as a transparent vehicle for ready-made meanings. Once, however, we re-awaken an awareness of the genuine phenomenon of ‘speaking language’ within which meanings are forged, it becomes apparent that the expression of truths is a cultural achievement that is only sustained through the use of a common language. Indeed the fact that language is essentially common enables it to play a role comparable to that of the body in relation to perception; it is, Merleau-Ponty says, a form of ‘anonymous corporality’ ([1969] 1974: 140), and as such it sustains our a priori involvement with others in much the way in which our physical corporality sustains our a priori involvement with the physical world in perception. So, in enabling us to share our lives with others, language equally enables us to live in a world of shared, and thus objective, truths.

6 Painting

Merleau-Ponty applied his phenomenology to a great variety of subjects, including issues in history, sociology and politics; but his discussions of painting are especially interesting.

The classical painting of the Renaissance, he argues, with its perspectival delineation of distinct objects, is an art that is informed by the goal of capturing what is seen as the appearance of objects in their own right, related to the eye only in ways legitimated by an objectivist conception of the world. Impressionists such as Cézanne, by contrast, ‘did not want to separate the stable things which we see and the shifting way in which they appear; he wanted to depict matter as it takes on form’ ([1948] 1964: 13). Thus painters such as Cézanne, to whom Merleau-Ponty returns frequently, aim to capture on canvas the true phenomenology of the visual field. In their paintings the elements of the scene depicted emerge only through the interplay of often indistinct colours and shapes, sometimes unfolding without a single perspective; and yet ‘it is Cézanne’s genius that when the overall composition of the picture is seen globally, perspectival distortions are no longer visible in their own right but rather contribute, as they do in natural vision, to the impression of an emerging order’ ([1948] 1964: 14).

Similarly, one may say, it was Merleau-Ponty’s genius to see that when classical phenomenology is combined with psychological insight, a new way of thinking about human life emerges, one that can throw light on modern painting as much as on old questions of metaphysics.

See also: Idealism; Marxism, Western §2; Phenomenological movement

THOMAS BALDWIN

List of works

There is a complete list of Merleau-Ponty’s works, together with what was then a full list of secondary literature in: Lapointe, F. and Lapointe, C. (1976) *Maurice Merleau-Ponty and his Critics: An International Bibliography*, New York: Garland.

Merleau-Ponty, M. (1942) *La Structure de comportement*, Paris: Presses Universitaires de France; trans. A. Fisher, *The Structure of Behavior*, Boston, MA: Beacon Press, 1963.(Merleau-Ponty here develops his ‘dialectical’ approach to psychology.)

Merleau-Ponty, M. (1945) *Phénoménologie de la perception*, Paris: Gallimard; trans. C. Smith, *The Phenomenology of Perception*, London: Routledge, 1962.(Merleau-Ponty’s major work, using psychology to reconstruct classical phenomenology.)

Merleau-Ponty, M. (1946) ‘L’existentialisme chez Hegel’ (‘Existentialism in Hegel’), *Les Temps modernes* 1: 1311-19; repr. in *Sens et non-sens*, Paris: Nagel, 1948; trans. H.L. Dreyfus and P.A. Dreyfus, *Sense and*

Non-Sense, Evanston, IL: Northwestern University Press, 1964, 63-70.(Discusses the existentialist themes in Hegel's thought.)

Merleau-Ponty, M. (1947) *Humanisme et terreur*, Paris: Gallimard; trans. J. O'Neill, *Humanism and Terror*, Boston, MA: Beacon Press, 1969.(Merleau-Ponty's Marxist response to Koestler's *Darkness at Noon*; an unconvincing critique of liberal politics.)

Merleau-Ponty, M. (1948) *Sens et non-sens*, Paris: Gallimard; trans. H. Dreyfus and P. Dreyfus, *Sense and Non-Sense*, Evanston, IL: Northwestern University Press, 1964.(A collection of Merleau-Ponty's early essays, including some on literature and painting.)

Merleau-Ponty, M. (1955) *Les Aventures de la dialectique*, Paris: Gallimard; trans. J. Bien, *Adventures of the Dialectic*, London: Heinemann, 1974.(A collection of essays on Marxism, including Merleau-Ponty's critique of Sartre's 'ultra-bolshevism'.)

Merleau-Ponty, M. (1960) *Signes*, Paris: Gallimard; trans. R. McCleary, *Signs*, Evanston, IL: Northwestern University Press, 1964.(Merleau-Ponty's later essays, including some on language.)

Merleau-Ponty, M. (1964) *Le Visible et l'invisible*, ed. C. Lefort, Paris: Gallimard; trans. A. Lingis, *The Visible and the Invisible*, Evanston, IL: Northwestern University Press, 1968.(Merleau-Ponty's incomplete attempt to rework the central themes of his phenomenology.)

Merleau-Ponty, M. (1969) *La Prose du monde*, ed. C. Lefort, Paris: Gallimard; trans. J. O'Neill, *The Prose of the World*, London: Heinemann, 1974.(The text of a projected book on language, abandoned in 1952.)

References and further reading

Archard, D. (1980) *Marxism and Existentialism*, Belfast: Blackstaff Press.(A useful guide to the quarrel between Merleau-Ponty and Sartre.)

Edie, J. (1987) *Merleau-Ponty's Philosophy of Language: Structuralism and Dialectics*, Washington, DC: University Press of America.(An authoritative discussion by one of America's leading scholars.)

Goldstein, K. (1934) *Der Aufbau der Organismus*, The Hague: Nijhoff; trans. *The Organism*, New York: American Book, 1938.(A classic work on the holistic aspect of the life of organisms, on which Merleau-Ponty relies in *The Structure of Behavior*.)

Hadreas, P. (1986) *In Place of the Flawed Diamond*, New York: Peter Lang.(A short but thoughtful Heideggerian account of Merleau-Ponty's work.)

Husserl, E. (1931) *Méditations cartésiennes*, Paris: Colin; trans. D. Cairns, *Cartesian Meditations*, The Hague: Nijhoff, 1960.(Husserl's Paris lectures, which Merleau-Ponty attended as a student in 1929.)

Madison, G. (1981) *The Phenomenology of Merleau-Ponty*, Athens, OH: Ohio University Press.(A general survey, especially useful in elucidating the significance of themes from Merleau-Ponty's later writings.)

Mallin, S. (1979) *Merleau-Ponty's Phenomenology*, New Haven, CT: Yale University Press.(An account which focuses on Merleau-Ponty as a phenomenologist.)

O'Neill, J. (1970) *Perception, Expression and History*, Evanston IL: Northwestern University Press; repr. as the introduction to J. O'Neil (ed.) *Phenomenology, Language and Sociology*, London: Heinemann, 1974, xi-lxii. (Although this latter volume is primarily a useful selection of Merleau-Ponty's essays, it contains O'Neill's important essay, which focuses on Merleau-Ponty's political writings, and a bibliography of works on Merleau-Ponty up to 1972.)

Olafson, F. (1967) 'A Central theme of Merleau-Ponty's Philosophy', in E. Lee and M. Mandelbaum (eds) *Phenomenology and Existentialism*, Baltimore, MD: Johns Hopkins University Press, 179-205.(An excellent introduction to Merleau-Ponty's philosophy of perception, by a scholar who also contributed an excellent article on Merleau-Ponty to Edwards' *Encyclopedia of Philosophy*.)

Sartre, J.-P. (1943) *L'Être et le néant*, Paris: Gallimard; trans. H. Barnes, *Being and Nothingness*, New York: The Philosophical Library, 1956.(Sartre's classic work of existential phenomenology which Merleau-Ponty criticizes at the end of *The Phenomenology of Perception*.)

Sartre, J.-P. (1961) 'Merleau-Ponty vivant', *Les Temps modernes* 17 (184-5): 304-76; repr. in *Situations IV*, Paris: Gallimard, 1964; trans. B. Eisler, *Situations IV*, New York: Brazilier, 1965.(Sartre's moving and intelligent tribute to his old comrade; the whole issue of *Les Temps modernes* is taken up with tributes to Merleau-Ponty, from Lefort, Lacan and others. Curiously, Sartre's published text is an almost complete replacement of another, even better, text which is published in translation as 'Merleau-Ponty (I)', *Journal of the British Society for Phenomenology* 15 (2): 128-58.)

Waelhens, A. de (1969) *Une Philosophie de l'ambiguité* (revised edition), Louvain: Publications Universitaires de Louvain.(Still the best general account of Merleau-Ponty's work, especially good at situating it in relation to the work of other philosophers.)

Mersenne, Marin (1588-1648)

Marin Mersenne represents a new seventeenth-century perspective on natural knowledge. This perspective elevated the classical mathematical sciences over natural philosophy as the appropriate models of what can be known, of how it can be known and of the cognitive status of that knowledge. His early publications had the apologetic aim not only of combating various forms of heresy, but also of opposing philosophical scepticism, which was widely regarded in Catholic France of the early seventeenth century as undermining the certainty of religious dogma. To that end, Mersenne stressed the certainty of demonstrations in sciences such as optics, astronomy and mechanics, all of which stood as 'mathematical' sciences in the classifications of the sciences stemming from Aristotle. Mersenne's stress on the mathematical sciences contrasted them with natural philosophy in so far as the former concerned only the measurable external properties of things whereas the latter purported to discuss their inner natures, or essences. In accepting the considerable degree of uncertainty attending knowledge of essences, and juxtaposing it to the relative certainty of knowledge of appearances, Mersenne adopted a position (since called 'mitigated scepticism') that combated scepticism by lowering the stakes: in accepting that the essences of things cannot be known, he agreed with the sceptics; but in asserting that knowledge of appearances can, by contrast, be had with certainty, he rejected the apparent intellectual paralysis advocated by the sceptics. In furthering this programme, Mersenne embarked on a publication effort relating to the mathematical sciences, combined with a massive lifelong correspondence on largely philosophical as well as religious topics with a wide network of people throughout Europe.

1 Career and correspondence

Born in France at Oizé, Maine, Mersenne was educated at the Jesuit college of La Flèche soon after its foundation in 1604, where he was an older contemporary of Descartes. He left about 1609 to study theology in Paris at both the Sorbonne and the Collège de France. In 1611 he entered the austere Order of Minims, joining their Parisian convent off the Place Royale (now Place des Vosges) in 1619 where he lived for the rest of his life. His earliest publications date from 1623, and all have apologetic aims, culminating in *La verité des sciences* (1625). From 1626 onwards Mersenne's publications concentrated on the mathematical sciences, including the so-called 'mixed' mathematical sciences that concerned aspects of the physical world rather than simply pure quantity (whether discrete, as in arithmetic, or continuous, as in geometry). He is best known, perhaps, as Descartes' epistolary link to the learned world during the latter's residence in the Netherlands from 1629 until Mersenne's death; he assisted in the publication of Descartes' *Discourse on Method* (1637) and organized solicitation of the 'Objections' to the *Meditations* that appeared in 1641 (see Descartes, R. §1). In addition to Descartes, Mersenne corresponded with many other philosophers throughout Europe. He was the centre of an information network in which he acted as intelligencer and clearing-house, both communicating ideas and sometimes fomenting disputes. This tireless correspondence, amounting to seventeen volumes in its modern edition, has since earned him the label 'one-man scientific journal'.

The correspondence exemplifies aspects of Mersenne's wider religious and social concerns, in so far as it represents his increasing irenicism in the period following his early apologetic writings. Referred to late in his life as 'the Huguenot monk', Mersenne numbered Protestants as well as Catholics among his correspondents, and strove to overcome the doctrinal gaps between those camps. This generally took the form of gentle attempts at converting the Protestants to a moderate, reconciliatory form of Catholicism, and contrasts quite starkly with the ferocity of the assaults on heresy found in his early works. He never changed in his perception of the dangers of scepticism, however, and his so-called 'mitigated' scepticism amounted to a philosophical irenicism to match the religious. Mersenne's wide correspondence could itself proceed only within a broad discursive consensus, whether religious or philosophical.

Mersenne's engagement with the work of some of his illustrious philosophical contemporaries does not appear solely in relation to Descartes. During the 1630s, Mersenne publicized the work of Galileo with two books, as well as in occasional discussions of Galileo's work in other publications and in correspondence. Mersenne travelled several times, to the Netherlands (meeting Isaac Beeckman there), provincial France and Italy. From the latter excursion he brought back news to France, in 1645, of the Torricellian barometric experiment, which led to the famous work of Blaise Pascal on the weight of the air.

2 Apologetics and anti-scepticism

Mersenne's chief apologetic works contain a good deal of philosophical and mathematical content. They date from 1623, and include the recently rediscovered *L'Usage de la raison*, *Quaestiones celeberrimae in Genesim* and *Observationes et emendationes ad Francisci Georgii Venetii problemata* (generally bound together), *L'Impieté des deistes* (1624), and perhaps most significant, *La verité des sciences* (1625), a work that tackles the serious contemporary challenge of philosophical, particularly Pyrrhonian, scepticism with weapons drawn heavily from the mathematical sciences.

Quaestiones, his first major publication, is an uncompleted commentary on Genesis (a continuation exists in manuscript) containing lengthy digressions on such topics as astronomical systems and geometrical optics. Mersenne's ability to use such subject-matter for apologetic ends is typified by his use of optical theory to refute the naturalism of the sixteenth-century Paduan philosopher Pietro Pomponazzi. Naturalism was a doctrine that regarded nature as a self-activating, self-motivating system that had no need of the supernatural to sustain it. It thus seemed to Mersenne, as to many others, to reduce or eliminate the natural world's dependence on God's superintendence. There was an element of circularity in Pomponazzi's position, in so far as what counted as 'natural' was practically defined as whatever occurred in the natural world; thus, by definition, whatever happens is natural rather than miraculous or supernatural. For Pomponazzi and later proponents of this position such as the magician Julius Caesar Vanini (executed for heresy in 1619), even the miracles described in the Bible had been events with natural explanations. Pomponazzi had tried to account in terms of optical illusion for an angelic visitation recorded in the Old Testament. Mersenne responded by using his own knowledge of geometrical optics: he expatiates at length in *Quaestiones* on optics, and especially on the properties of curved mirrors, to show that Pomponazzi's explanation could not possibly work.

Mersenne soon extended the significance for religious apologetics of the so-called 'mixed' mathematical sciences (such as optics, astronomy, mechanics and music theory, which made use of theorems from geometry and arithmetic). This arose from his explicit confrontation of Pyrrhonian scepticism, a form of ancient Greek scepticism that enjoyed a revival starting around the middle of the sixteenth century. The central texts were writings by Sextus Empiricus (*c.*200 AD), including most famously his *Outlines of Pyrrhonism*. The Pyrrhonist maintained that no proposition could be held as absolutely certain - one could always find grounds, however outlandish, for doubt - and that, as a result, the wise man will always suspend judgment on all things, subscribing to no dogmatic beliefs, but simply conforming to prevailing norms. This position contrasts with so-called 'Academic' scepticism, also dating from antiquity, which recommended the exercise of probable judgment (see Pyrrhonism).

Pyrrhonism had become quite popular in some circles in France by the early seventeenth century, where it played a role in religious controversies. By the mid-1620s Mersenne had come to see it as just one example of a whole range of unorthodox philosophical positions that, like Pomponazzi's naturalism, threatened religious, and concomitantly social, stability. The most common tactic of its opponents in France at this time was to counter Pyrrhonian arguments piecemeal, precisely according to the form in which they were offered. Mersenne, by contrast, chose to attack Pyrrhonism by erecting an epistemological structure, with an associated natural philosophy, to act as a comprehensive bulwark; despite the later example of Descartes, reformulations of natural philosophy were not an inevitable tactic of opponents of Pyrrhonism in this period. A specific factor contributing to Mersenne's adoption of such an approach was his interest in natural philosophy and in the mathematical sciences.

For Mersenne, the problem with Pyrrhonism was not so much its use of sceptical arguments as the radical nature of the doubt that those arguments served to produce. Profound and generalized doubt came too close to promoting doubt in religion. Mersenne evidently thought that without the discipline of reason, heresy or unbelief would become too difficult to guard against. If scepticism were applied to sensitive areas like morality, it might not only promote immoral living, but also threaten social stability by questioning the principles on which social hierarchy was established. The difficulties therefore rested on the absoluteness of the challenge to certainty that Pyrrhonian scepticism posed. Mersenne responded by attempting to destroy the threat of total, universal doubt by establishing what might be called bridgeheads of certain knowledge, in the hope that these bridgeheads would prove sufficient to indicate a workable foundation for rational theology and rational, enforceable social order.

The main such bridgehead was that of the mathematical sciences. Mersenne had used the mathematical science of optics in *Quaestiones in Genesim* for apologetic purposes (to show the impossibility of naturalistic explanations of certain miracles), and its advantage there was one shared by all mathematical sciences: mathematics was conventionally treated as the exemplar of certain knowledge in contemporary school teachings, within a general context of a widespread moderate scepticism (greatly to be contrasted with the extremities of Pyrrhonism) promoted in the schools through the medium of humanist dialectic. This stressed the idea of probability rather than certainty in most practical affairs. Mersenne's Jesuit education had similarly emphasized the certainty of mathematics as a contrast to the merely probable knowledge available from qualitative Aristotelian physics. The author of the principal mathematical textbooks used in the Jesuit colleges, Christopher Clavius, tried to promote the status of the mathematical sciences in Jesuit pedagogy by exploiting the apparent contrast between the uncertainty of conclusions in natural philosophy and the certainty of those in mathematics, pointing out that a conclusion in dialectic 'is confirmed only probably', whereas a conclusion in mathematics is a matter of certain demonstration.

Mersenne's stance in *La verité des sciences* flows directly from this impact of humanist probabilism. One of the most characteristic aspects of his response to Pyrrhonian arguments regarding the uncertainty of most questions, and especially on the impossibility of achieving reliable knowledge of the essences of things, is that he generally *accepts* them. These were positions routinely accepted within early seventeenth-century scholastic natural philosophy. Mersenne departed from orthodoxy here only to the extent that his attempt to defeat Pyrrhonism required it. He did not absolutely reject any kind of knowledge that lacked absolute certainty; instead, he tended to discount it only because it did not serve his anti-Pyrrhonian argumentative strategy. He dismissed anything accepted as uncertain, so as to concentrate on the opportunities presented by what *was* certain; in practice, this tended to mean mathematics.

The certainty of mathematics would not in itself have been an adequate basis for an approach to *natural philosophy*. There were, in fact, three commonly-accepted sources of certainty recognized within mainstream humanist-influenced scholasticism by the early seventeenth century: mathematics, ordinary sense-perception and syllogistic logic. Although all three were themselves under fire from Pyrrhonist arguments, together they constituted the main ingredients of Mersenne's answer to Pyrrhonian scepticism. They worked together as interrelated elements of a form of natural philosophy that bypassed the essentialist claims of Aristotelianism. If accepted as reliable, ordinary sense-perception could serve to generate true statements whose consequences, through the application of syllogistic reasoning, might serve in turn to ground mathematical expressions of the appearances of things in the physical world. By a reversal of the process, mathematics then enables the determination of how sensory perceptions of appearances are affected by, or can be deliberately altered by, changes in circumstances or in the conditions of observation. Mersenne's proposal, therefore, is for a mathematical coordination of appearances, which can serve as an entirely satisfactory practical substitute for Aristotelian essentialist physics. Mersenne's prize examples of such a thing were, and remained, geometrical optics and music theory, although he also wrote on mechanics. The last took the form initially of quasi-Archimedean statics, although in the 1630s he took a considerable interest in Galileo's kinematics, an approach to change in the physical world that appealed greatly to Mersenne's project of modelling natural knowledge on the mathematical sciences.

An important aspect of Mersenne's approach is that of operationalism: being able to accomplish some practical task itself acts as a criterion of the possession of knowledge, much as we find in Francis Bacon (see Bacon, F. §3). This in turn indicates a positive evaluation of practical utility as a proper goal of philosophical knowledge. Once again, the theme of the utility of applied mathematics is recurrent in late fifteenth- and sixteenth-century humanist encomiums to mathematics, and forms a significant part of Clavius' defence of mathematical studies.

Richard H. Popkin (1979) has described Mersenne's overall position, as originally laid out in *La verité des sciences*, as 'constructive' or (following Hume) 'mitigated' scepticism. In Mersenne's form, however, it is less a grounded philosophical argument than the outcome of a particular set of philosophical assumptions deployed in a specific polemical context. Mersenne opposes to Pyrrhonism the position of Ciceronian humanist probabilism as widely promulgated in the schools, including the greatly influential Jesuit colleges. He thus emphasizes the kinds of knowledge that this position typically allowed as certain. Since these did not include essentialism - knowledge of the inner natures of things - as sought by (among other systems) Aristotelian natural philosophy, Mersenne

abandoned the latter, even while describing it as the ‘most probable’ such system. Since even a ‘probable’ essentialism was still open to the attacks of the Pyrrhonists precisely for not being certain (still the most important property of genuine knowledge, *scientia*), this left Mersenne upholding his favourite mathematical sciences as the models for a new natural philosophy.

Thus Mersenne should not be seen as substituting, like Descartes, a new philosophy in place of Aristotelianism, one that would be immune to the Pyrrhonist threat. Instead, he simply adapts most of the elements of his position from standard teachings, shifting the emphasis somewhat, in order to secure knowledge of nature (among other things) from Pyrrhonism. It would have been consistent with his arguments in *La verité* to have retained Aristotelian physics on a probabilistic basis. Indeed, his previous publications bear witness to just such a stance. But from *La verité* onwards, Mersenne effectively just ignores it in favour of mathematical sciences because physics implied knowledge of essences, and such knowledge was not certain - hence not an appropriate basis for a true science.

The philosophical inadequacy of Mersenne’s response to Pyrrhonism is made clear by the lack of arguments provided in support of the central planks of his position. Pyrrhonism questioned the reliability of ordinary sense-perceptions; Mersenne merely asserts it, and claims that variations in appearance - such as the standard example of a stick appearing bent when half-immersed in water - could be understood and coordinated through a proper understanding of the contingencies surrounding them (in the case of the stick, a proper understanding of refraction). This response, of course, begs the question of how assertions about refraction itself are to be grounded. Similarly, there were standard Pyrrhonian arguments against the value of syllogistic inference, relying especially on the point that syllogisms are question-begging because the universals on the basis of which they establish their singular conclusions are themselves justified on the basis of just such singulars. Mersenne is content to respond by reasserting the certainty of syllogistic inference and recommending the logical works of Aristotle for further elucidation. In the case of mathematics, for which Pyrrhonists also had a battery of undermining arguments, Mersenne does try a little harder, in particular defending arithmetic and geometry against some specific criticisms found in Sextus Empiricus.

Mersenne’s approach characterizes not just his lifelong opposition to extreme scepticism, but also his opposition to a variety of other deviant philosophies. That opposition had been manifested in previous writings (*Observationes*, directed especially against the recently executed magician Vanini, and the second volume of *L’Impieté des deistes*, which among other targets took on Giordano Bruno). He was also to engage in polemics with Robert Fludd slightly later in the 1620s. Although most of *La verité* takes the form of a dialogue between a ‘Christian Philosopher’ (evidently Mersenne’s mouthpiece) and a ‘Sceptic’ (the latter routinely being deeply impressed by the former’s display of mathematical truths), Book I, which contains most of the work’s philosophical meat, presents a three-way discussion. The additional interlocutor is the ‘Alchemist’. Mersenne’s trick is to allow the Sceptic freely to demolish the foundations of the Alchemist’s position (itself a variety of essentialism), and only when the Alchemist has been vanquished does the Christian Philosopher step in to show why the kind of knowledge that he prizes remains unthreatened by the criticisms directed at the Alchemist. Thus Mersenne’s selective acceptance of certain sceptical arguments flows not simply from his inability to refute them, but also from his wish to use them for his own ends. Orthodoxy was, for Mersenne, threatened more by magical and alchemical philosophies than it would be by the bracketing-off of essentialist Aristotelianism.

3 Mathematics and music

In practice, Mersenne’s work after 1625, both as a correspondent/intelligencer and as writer of books, focused on the promulgation of knowledge in the mathematical sciences and on the consideration of novelties in those fields. He is sometimes called a ‘mechanist’, a label reinforced by its use in the title of Robert Lenoble’s classic study *Mersenne ou la naissance du mécanisme* (1943). The usual understanding of that term for seventeenth-century philosophy is not properly applicable, however. Lenoble’s use of the term ‘mechanism’ equated it with instrumentalist, operational knowledge, whereas the usual meaning now among historians studying this period relates to specific *ontological* commitments, chief among them corpuscularianism or atomism. These stances, associated most closely with Descartes and Gassendi respectively, are of course essentialist, in that they make commitments to views on the underlying natures of things. In Descartes’ case the commitment, at this level of general matter-theory, was meant to be dogmatic rather than conjectural; in Gassendi’s, it was explicitly

hypothetical. But neither position seems to have appealed to Mersenne. His appreciation of sceptical views kept him from wholehearted acceptance of Cartesian mechanism, whereas his desire for certain knowledge made Gassendi's hypotheticalism unattractive. The model of the mixed mathematical sciences provided the perfect *via media*, since they neither committed to a particular ontological structure for the material world nor abandoned the goal of producing rigorous demonstrations of their conclusions.

His first publication specifically on the mathematical sciences was *Synopsis mathematica* (1626), a compendium of material on geometry, optics and also mechanics (a second edition appeared in 1644 under the title *Universae geometriae… synopsis*). In a similar vein he published *Les Mechaniques de Galilée* (1634e), a French paraphrase of an early Galilean work from the turn of the century on statics and simple machines, which had previously circulated only in manuscript; *Les Nouvelles pensées de Galilée* followed in 1638-9. The latter work presented paraphrased material from Galileo's 1638 *Discourses and Demonstrations Concerning Two New Sciences*, and provides further indication of the congeniality to Mersenne of Galileo's approach to issues of natural philosophy developed on the basis of 'mixed mathematics'.

Mersenne's first work devoted to music appeared in 1627, the *Traité de l'harmonie universelle*, on the heels of which he began composition of his great multi-volume *Harmonie universelle*, which finally appeared in 1636-7. These texts betray a further dimension of Mersenne's conception of mathematical knowledge and its relation to natural philosophy. He was much taken with Augustine's Neoplatonic emphasis on harmony as a foundational aspect of the created world's intelligibility. It is important to recall here that 'natural philosophy' in medieval and early-modern Latin culture was a branch of learning that treated the natural world as God's Creation. Music as a branch of mixed mathematics was concerned with ratios and proportions associated with certain privileged qualitative features of the world. It thus served to express the ineffable with mathematical precision. As a consequence, it was fit to act as the ideal basis for Mersenne's rather Augustinian conception of what he called 'universal harmony'. By seeing the constitution of the physical universe as being intelligible in the pre-established terms of musical ratios, Mersenne thought himself to have invested mathematical treatments of nature with genuinely natural-philosophical content. He attempted this extension by more than just drawing analogies or musical metaphors: he analysed the behaviour of balances in statics in terms of musical ratios, so as to represent their action as fulfilling precisely the same formal requirements as those of music, and completed the circle by representing musical consonances themselves as the product of the relative frequencies of series of air pulses. Indeed, in taking these latter as tantamount to sound itself, he went beyond the usual treatment, which usually only *correlated* sound with air disturbances, to *identify* sound with these mathematizable air pulses. He thus retained a kind of Platonic essentialism in his mathematical approach to physical phenomena.

See also: Augustine; Augustinianism; Neoplatonism

PETER DEAR

List of works

Mersenne, M. (1623a) *L'Usage de la raison (Use of reason)*, Paris.(Recently rediscovered work of apologetics.)
Mersenne, M. (1623b) *Quaestiones celeberrimae in Genesim*, Paris.(Uncompleted commentary on Genesis.)
Mersenne, M. (1623c) *Observationes et emendationes ad Francisci Georgii Venetii problemata*, Paris.(Bound together with *Quaestiones*; the two together are a major source on Mersenne's early views on magical and naturalistic systems.)
Mersenne, M. (1624) *L'Impieté des deistes (The impiety of the deists)*, Paris, 2 vols; repr. vol. 1, Stuttgart-Bad Cannstatt: Friedrich Frommann, 1975.(Volume 2 has a lengthy critique of Giordano Bruno.)
Mersenne, M. (1625) *La verité des sciences (The truth of the sciences)*, Paris; repr. Stuttgart-Bad Cannstatt: Friedrich Frommann, 1969. (The classic statement of 'mitigated scepticism'.)
Mersenne, M. (1626) *Synopsis mathematica*, Paris; 2nd edn, 1644.(A compendium based on classical authors.)
Mersenne, M. (1627) *Traité de l'harmonie universelle (Treatise on universal harmony)*, Paris.(On aspects of music theory.)
Mersenne, M. (1634a) *Traité des mouvemens (Treatise on movement)*, Paris; repr. in *Corpus: Revue de philosophie* 2 (1986): 25-58.(On falling bodies.)
Mersenne, M. (1634b) *Questions inouyes*, Paris; repr. ed. A. Pessel, Paris: Fayard, 1985. (Mathematical and physical problems.)
Mersenne, M. (1634c) *Questions harmoniques*, Paris; repr. ed. A. Pessel, Paris: Fayard, 1985. (Musical matters,

including a sceptical treatise by La Mothe Le Vayer.)
Mersenne, M. (1634d) *Questions theologiques*, Paris; repr. ed. A. Pessel, Paris: Fayard, 1985. (Similar to *Questions inouyes*.)
Mersenne, M. (1634e) *Les Mechaniques de Galilée*, Paris; repr. ed. A. Pessel, Paris: Fayard, 1985.(Paraphrase of a Galilean work on statics from the beginning of the century.)
Mersenne, M. (1634f) *Les Preludes de l'harmonie universelle (Preludes of universal harmony)*, Paris; repr. ed. A. Pessel, Paris: Fayard, 1985.(Another discussion of various philosophical issues, with much musical content.)
Mersenne, M. (1636-7) *Harmonie universelle (Universal harmony)*, Paris; repr. Paris: Centre Nationale de la Recherche Scientifique, 1963.(Mersenne's great treatise on music theory, musical instruments, and miscellaneous non-musical matters.)
Mersenne, M. (1638-9) *Les Nouvelles pensées de Galilée (The new thoughts of Galilee)*, Paris.(Paraphrasing some material from Galileo's *Discorsi* of 1638.)
Mersenne, M. (1639) *L'Usage du quadran (The use of the quadrant)*, Paris: Rocolet.(On a mathematical instrument.)
Mersenne, M. (1644a) *Cogitata physico-mathematica*, Paris.(Discusses a wide variety of mathematical subjects, including ballistics and mechanics as well as music.)
Mersenne, M. (1644b) *Universae geometriae, mixtaeque mathematicae synopsis*, Paris.(A new version of the *Synopsis* with optical treatises by Thomas Hobbes and Walter Warner.)
Mersenne, M. (1647) *Novarum observationum…tomus III*, Paris.(A continuation of *Cogitata*.)
Mersenne, M. (1648) *Harmonicorum libri XII*, Paris; repr. Geneva: Minkoff, 1972.(A short Latin condensation of material in *Harmonie universelle*, with some additions.)
Mersenne, M. (1651) *L'Optique et la catoptrique du Reverend Pere Mersenne Minime*, Paris. (A study on optics, published posthumously.)
Mersenne, M. (1932-88) *Correspondance du P. Marin Mersenne, religieux minime*, ed. C. de Waard, R. Pintard, B. Rochot, A. Beaulieu, Paris: Beauchesne, Presses Universitaires de France, Centre Nationale de la Recherche Scientifique, 17 vols.(The great modern edition; a goldmine on seventeenth-century philosophical networks.)

References and further reading

Costabel, P. (1993) 'Marin Mersenne', *Die Philosophie des 17. Jahrhunderts, Band 2: Frankreich und Niederlande*, ed. J.-P. Schobinger, Basel: Schwabe, 637-47, 700-2.(A useful overview in German with good bibliography.)
Coste, H. de (1649) *La vie du R. P. Marin Mersenne, théologien, philosophe et mathématicien, de l'ordre des Pères Minimes*, Paris; repr. in P.T. de Larroque, *Les correspondants de Peiresc* 2, Geneva: Slatkine, 1972, 436-97.(The chief near-contemporary biographical source.)
Crombie, A.C. (1974) 'Mersenne, Marin', *Dictionary of Scientific Biography*, vol. 9, ed. C.C. Gillispie, New York: Scribner's, 316-22.(A brief survey of Mersenne's career and publications.)
Dear, P. (1988) *Mersenne and the Learning of the Schools*, Ithaca, NY: Cornell University Press.(Stresses Mersenne's debt to scholastic, especially Jesuit, sources.)
Duncan, D.A. (1981) *The Tyranny of Opinions Undermined: Science, Pseudo-Science and Scepticism in the Musical Thought of Marin Mersenne*, Vanderbilt University Ph.D. dissertation.(A good treatment of Mersenne's musical work in its broader philosophical context.)
Duncan, D.A. (1986-9) 'An International and Interdisciplinary Bibliography of Marin Mersenne', *Bolletino di Storia della Filosofia dell'Università degli Studi di Lecce* 9: 201-42.(The fullest bibliography currently available.)
Études philosophiques (1994), 'Études sur Marin Mersenne', 1-170.(A miscellaneous collection from a recent French conference on Mersenne.)
Hine, W.L. (1984) 'Marin Mersenne: Renaissance Naturalism and Renaissance Magic', in B. Vickers (ed.) *Occult and Scientific Mentalities in the Renaissance*, Cambridge: Cambridge University Press, 165-76.(Discusses Mersenne's early attitudes towards magical philosophies.)
Lenoble, R. (1943) *Mersenne ou la naissance du mécanisme*, Paris: Vrin, 1971.(The classic study, presenting Mersenne as something of a proto-positivist.)
Popkin, R.H. (1979) *The History of Scepticism from Erasmus to Spinoza*, Berkeley, CA: University of California Press.(Contains the influential account of Mersenne's 'mitigated scepticism'.)
Quatrième centenaire de la naissance de Marin Mersenne, 1588-1988, *Actes du colloqué*, Le Mans: Faculté des Lettres, Univ. du Maine.(A recent miscellany of scholarly articles, largely historical rather than philosophical.)
Thuillier, R. (1709) *Diarium patrum, fratrum et sororum Ordinis Minimorum Provinciae Franciae*, Paris; repr.

Geneva: Slatkine, 1972, vol. 2: 90-113.(Together with De Coste, the main early biographical account, in Latin.)

Messer Leon, Judah (*c*.1425-*c*.1495)

*Messer Leon was a philosopher, physician, jurist, communal leader, poet and orator. Ordained as a rabbi by 1450, Messer Leon was qualified to adjudicate legal cases among Jews and head an academy (*yeshivah*) for advanced studies in Jewish law. He also came close to embodying the Renaissance ideal of* uomo universale. *His learning was formally recognized in 1469, when Emperor Frederick III awarded him a doctorate in medicine and philosophy and granted him the unusual privilege of conferring doctoral degrees in those subjects on Jewish students.*

Messer Leon's contribution to Jewish philosophy was in the field of logic, the art considered by him to be the key to the proper harmonization of religion and philosophy. He regarded scholastic logic to be superior to Arabic logic and wrote supercommentaries on Averroes' logical works as well as an encyclopedia of logic, Mikhlal Yofi (Purest Beauty), in an attempt to shift Jewish philosophical education from the Judaeo-Arabic logical tradition to scholastic logic. Although his encyclopedia became a popular textbook, Messer Leon failed to mould the culture of Italian Jewry as he had intended. In particular, he could not curb the spread of Kabbalah, a tradition which he vehemently opposed because of its underlying Platonic metaphysics.

Most importantly, Messer Leon composed the first manual of Hebrew rhetoric, entitled Nofet Tzufim (The Book of the Honeycomb's Flow). Printed in 1476, this Jewish response to Latin humanism combines the Averroist-Aristotelianism tradition and the Ciceronian-Quintilian one. The appropriation of humanistic rhetoric was given a Jewish meaning when Messer Leon claimed that the Torah, rather than the writings of the pagan, classical orators, exemplified perfect speech because it was a revelation of perfect divine wisdom. By analysing Scripture from the perspective of classical rhetoric, Messer Leon legitimized the study of ancient pagan and Christian orators even as he argued for the supremacy of biblical rhetoric over all merely human eloquence.

1 Life and works

Rabbi Judah ben Yehiel, called Messer Leon, wrote philosophic and scientific works in Hebrew to afford his *yeshivah* students university-level instruction, from which all but a tiny fraction of Jews were excluded. He lived in Ancona, Mantua, Padua, Bologna, Venice and Naples, always relocating his celebrated *yeshivah* wherever he went. Linked by kinship and business ties to Jewish banking families (the Da Pisas of Florence and the Norsas of Mantua) and honoured by titles and privileges in the Jewish and larger Italian community, Messer Leon without question enjoyed the highest social status held by a Jew in Renaissance Italy. He came to think of himself as the leader of Italian Jewry and issued decrees intended to shape Jewish religious practice, belief, and education. He even adopted the ancient title *Ro'sh Golah* (Head of the Diaspora). But his aristocratic pretensions were resented by other rabbis, and his attempts to lead all Italian Jews were hardly universally accepted. The time and place of his death are uncertain, but it is unlikely that he survived the conquest of Naples by the French armies of Charles VIII in 1494/5.

Deeply rooted in scholastic Aristotelianism and staunchly loyal to Maimonides, the paramount figure in medieval Jewish Aristotelianism, Messer Leon saw philosophy as dedicated to the elucidation of authoritative texts. He was dedicated to logic and convinced that reasoning could resolve the apparent conflict between religion and philosophy. Trained by Christian logicians, he found scholastic logic superior to the Arabic logic taught by al-Farabi and Averroes (see Ibn Rushd). Seeking to shift Jewish philosophical training to the Latin model, he wrote supercommentaries on Averroes' middle commentaries on the Organon, painstakingly collating their diverse textual traditions and harmonizing the Hebrew versions with the diverse Latin readings.

His 1454 compendium of Aristotelian logic, entitled *Mikhlal Yofi (Purest Beauty)*, departed from medieval Arabic precedent by omitting the *Rhetoric* and *Poetics*, in effect emancipating these studies from their medieval status as mere arts of persuasion. *Mikhlal Yofi* excerpted several books of scholastic logic, notably the *Logica magna* of Paul of Venice (*circa*1339/72-1429), which was also the primary source of Messer Leon's supercommentary on the *Posterior Analytics*. His supercommentary on Porphyry's *Isagōgē*, similarly, drew heavily on the *Super artem veterem expositio* of Walter Burley (*circa* 1275-1345). The supercommentaries on *Categories*, *De interpretatione* and the *Prior Analytics*, all extant in manuscripts, again rely on scholastic sources. Usefully, Messer Leon wrote summaries (*sefeqot*) of the major philosophical questions debated in Italian universities.

His debt to scholastic authors led to charges by a philosophy instructor in his own *yeshivah*, now known only as David the Spaniard, that his writings were mere translations and that even his glosses bordered on plagiarism. The master publicly rebutted the accusations and dismissed his adversary as a exponent of the views of Jewish Averroists from Provence, chiefly, Levi ben Gershom (see Gersonides), whose apparent denial of God's knowledge of particulars seemed to undermine the idea of providence. Messer Leon tried to suppress Gersonides' recently printed commentary on the Pentateuch, but with little success. It was highly esteemed among native Italian Jews, and Gersonides' *magnum opus*, *Milhamot ha-Shem (The Wars of the Lord)*, became a classic of Jewish philosophy among the educated.

2 Philosophical standpoint

The extant works and the letters of Messer Leon's son David (d. *circa* 1535), cite a number of philosophical and scientific works now lost: supercommentaries on Aristotle's *On the Soul*, *Metaphysics*, *Nicomachaean Ethics*, *Physics* and, possibly, the *Topics*. To prepare his students in medicine, he wrote a commentary on the first and second parts of Avicenna's *Canon*, and a short summary of essential medical information (*Yedi'ot ha-nehuzot le-Rofe'*). Despite his devotion to such studies, Messer Leon, like the medieval Jewish philosophers, did not view philosophy and its allied arts and sciences as ends in themselves, but saw them as preparation for higher study - of the divinely revealed religion. In keeping with his desire to articulate an official Jewish theology, he commented on the two most widely read texts of Jewish philosophy in fifteenth-century Italy: Maimonides' *Guide to the Perplexed* and Yeda'aya Bedersi's *Behinat 'Olam (Examination of the Universe)*. The *Moreh Tzedeq (Teacher of Righteousness)*, his commentary on the *Guide*, is now lost, but his commentary on *Behinat 'Olam* is extant in several manuscripts and awaits systematic study. A reading of it shows Messer Leon's kinship in approach to other Jewish Aristotelians of his time. Like Joseph Ibn Shem Tov (d. *circa* 1460), Abraham Bibago (d. *circa* 1489) and Abraham Shalom (d. *circa* 1492) in Spain, and Elijah Delmedigo (d.1493) in Crete and Italy, Messer Leon insisted on the rationality of a divinely revealed Judaism while recognizing the inherent limitations of natural human reason. Because philosophic understanding rests on sense perception, it is prone to errors, doubts, and uncertainties and cannot bring us to our ultimate goal, individual beatitude in the hereafter. It must be complemented by revealed truths beyond the ken of natural reason. But these suprarational truths cannot contradict the precepts of philosophy. A rational examination of Judaism, which Messer Leon apparently undertook in his now lost commentary on the Pentateuch, would show that the Torah of Moses is in full accord with the teachings of Aristotle, as properly understood.

3 Rhetoric

Messer Leon endorsed the Platonic elitism of the Maimonidean tradition: God's revelation is abstruse; ordinary people lack the intellectual perfection needed to grasp its truths directly. They must receive these through the veils of figurative language. But philosophical wisdom does grasp the inner meaning of the Torah, and this more penetrating understanding of the Torah's message and its tropes endows the philosopher with political authority - an authority that rests on exegesis, and so on the study of grammar and rhetoric.

Messer Leon's *Livenat ha-Sappir (The Sapphire Stone)* systematically applied the categories of European grammar to the Hebrew language. The work relied heavily on the writings of Abraham ibn Ezra (1089-1164). Distinctive in Messer Leon's approach was the influence of Renaissance humanism (see Humanism, Renaissance). He saw grammar not merely as ancillary to logic but as praiseworthy in its own right. Its kinship with rhetoric made it an essential tool of cogent argument and elegant exposition and a *sine qua non* in exegesis.

In 1475, in Mantua, Messer Leon published a manual of Hebrew rhetoric entitled *Nofet Tzufim*, taking his title from the Psalms, where the image is applied to God's speech. Like his works on logic, this inventory of linguistic tropes combined Arabic and Latin traditions: the Averroist-Aristotelian tradition he knew through Todros Todrosi's Hebrew translation of Averroes' middle commentary on Aristotle's *Rhetoric*, and the Ciceronian-Quintilian tradition, which he knew chiefly through the pseudo-Aristotelian *Rhetorica ad Alexandrum*, Cicero's *De inventione*, the pseudo-Ciceronian *Rhetorica ad Herennium*, Quintialian's *Institutio oratoria*, and Marius Victorinus' commentary on Cicero's *De inventione*.

The first articulate Jewish response to the humanist cult of rhetoric and the concomitant revival of classical prose and poetry, *Nofet Tzufim* sparked a new appreciation for eloquence among Italian Jewish scholars. Siding with

Ermolao Barbaro in his dispute with Giovanni Pico della Mirandola, Messer Leon prized eloquence as a complement to moral and intellectual perfection: a good orator must first be a good man and a good philosopher. In place of the 'golden age' Latin held up as a model by Renaissance writers, Messer Leon placed biblical Hebrew, the perfect exemplar of the Ciceronian ideal of a synthesis of wisdom with eloquence (see Cicero, M.T.). By analysing Scripture in this Ciceronian light, Messer Leon legitimized the study of ancient pagan and Christian orators even as he argued for the supremacy of biblical rhetoric over all merely human eloquence.

4 Humanism and Kabbalah

The Renaissance humanism he so admired presented Messer Leon with an acute political and intellectual problem. The humanists of the later fifteenth century, especially in the Platonic Academy of Florence, revived numerous non-Aristotelian cosmological constructs, including Platonic, Neoplatonic and Hermetic frameworks. Urging the unity of all truth, humanists like Marsilio Ficino and Pico della Mirandola sought an ancient syncretic wisdom in these religious and philosophic traditions. A vital part of the *prisca theologia* they sought was the Jewish esoteric and mystical tradition of Kabbalah (see Kabbalah). The humanists' deep respect for Kabbalah contrasted sharply with the disdain of medieval scholars towards rabbinic Judaism, but it was not devoid of missionarizing intent. For the syncretic tradition was seen as culminating in Christianity; and, among the many Jewish intellectuals drawn by the humanist revival of Platonism and fascination with Kabbalah, some became converts to Christianity.

Messer Leon attempted to stem the spread of Kabbalah in Italy, viewing the tradition as neither authentic nor authoritative *vis-à-vis* rabbinic Judaism. It was a mere medieval throwback, a human innovation closely akin to the inferior philosophy of Platonism, which compromised the absoluteness of God's unity and incorporeality. Distressed in particular by the Christian overtones of humanist interest in Kabbalah, Messer Leon forbade its study in his *yeshivah*, but he had little success in curbing its spread in Italy among Jews and non-Jews alike. His son David and two other students of Aristotelian philosophy, Yohanan Alemanno (d. after 1503/4) and Abraham de Balmes (d. 1523), cultivated the Kabbalah and sought to incorporate its doctrines into the framework of Jewish Aristotelianism.

Messer Leon played an essential role in keeping alive the traditions of medieval Aristotelian philosophy among Italian Jews in the later fifteenth century. Modelling the training in his *yeshivah* on the Italian university curriculum, he professionalized the study of philosophy among Italian Jews. His sensitivity to humanist ideals enabled him to create a cadre of Jewish scholars who could offer spiritual leadership and eloquent counsel to the Jewish community. But all his high-handed attempts to mould the culture of Italian Jewry failed. His own students did not accept his categorical preference for Aristotelian logic over Platonic metaphysics, and even his son did not follow him in rejecting Kabbalah. Sixteenth-century Italian Jewish scholars were increasingly influenced by Renaissance Platonism and Kabbalah, rather to the cost of Maimonidean rationalism, as the accepted theology of Judaism.

See also: Alemanno, Y.; Aristotelianism, medieval; Aristotelianism, Renaissance; Humanism, Renaissance; Platonism, Renaissance

HAVA TIROSH-SAMUELSON

List of works

Ben Yehiel, Judah (*c*.1449) *Perush al' Yeda'ya Berdesi's Behinat 'Olam (Commentary on Yeda'ya Berdesi's Examination of the Universe)*, MS Florence: Biblioteca Medicea-Laurenziana Plut. 88.52/2; MS Vienna, 163. 8.(Berdesi's *Behinat 'Olam* was a very popular book among Jewish intellectuals in the first half of the fifteenth century. It articulated the worldview of medieval Jewish Neoplatonized Aristotelianism in a semi-poetic form. By writing a commentary on this text, Messer Leon intended to ground the accepted worldview in a more rigorous interpretation of Aristotelian philosophy.)

Ben Yehiel, Judah (1454) *Livenat ha-Sappir (The Sapphire Stone)*, MS Oxford: Bodleian Library, 1491; MS London: British Museum, Add. 27. 094; MS Paris: Bibliothèque Nationale, Hebrew 1247; MS Parma: Biblioteca Palatina, 2477 (De Rossi 114); MS Cambridge: Trinity College, 88; MS Moscow: Guenzburg, 583 and 1386.(This text, which Abraham Mordecai Farissol copied for the Norsa family in Mantua between 1473 and 1475, was still used as a textbook for study of Hebrew grammar by Jewish scholars in Constantinople during the sixteenth century.)

Ben Yehiel, Judah (1454) *Mikhlal Yofi (Purest Beauty)*, MS Leghorn (Livorno) 41; MS Rome: Biblioteca

Casanatense, 162; MS Florence: Biblioteca Medicea-Laurenziana, 88.52, MS Paris: Bibliothèque Nationale, Hebrew 994 and 998; MS London: British Museum, Or. 5467 and Add. 27.087; MS Vienna, 162; MS Parma: Biblioteca Palatina, 2615 (De Rossi 1355); MS Cambridge: Trinity College, Hebrew Add. 398.(This textbook for the study of logic was very popular among Jewish students in the fifteenth and early sixteenth century, indicating the perpetuation of Aristotelian logic among the Jewish intellectuals.)

Ben Yehiel, Judah (*c*.1470) *Perush le-Mavo-ma'amarot-Melizah (Supercommentary on Isagōgē - Categories-De interpretatione)*, ed. I. Husik, *Judah Messer Leon's Commentary on the Vetus Logica*, Leiden: Brill, 1906. (Husik's is a critical edition of a composite text. Husik used the manuscript which Abraham Mordecai Farissol prepared for the Norsa brothers in Mantua (MS Oxford: Bodleian 1392) and three other manuscripts listed in Bonfil's introduction.)

Ben Yehiel, Judah (1476) *Nofet Zufim (The Book of the Honeycomb's Flow)*, ed. R. Bonfil, *Nofet Zufim on Hebrew Rhetoric*, Jerusalem: The Jewish National and University Library Press and Magnes Press, 1981; ed. and trans. I. Rabinowitz, *The Book of the Honeycomb's Flow: Sepher Nophet Suphim by Judah Messer Leon*, Ithaca, NY: Cornell University Press, 1983.(Written in Hebrew, Bonfil's excellent 'introduction' to his edition lists all the extant manuscripts of Messer Leon and situates Messer Leon in his socio-cultural context. Rabinowitz's superb critical edition and English translation has a detailed introduction, which lists most of the pertinent, albeit controversial, information about Messer Leon's life, as well as giving a partial list of his works extant in manuscripts. Rabinowitz's elegant English translation and notes are especially valuable for situating this work in the context of Western rhetoric.)

References and further reading

Altmann, A. (1981) 'Ars Rhetorica as Reflected in Some Jewish Figures of the Italian Renaissance', in *Essays in Jewish Intellectual History*, Hanover, NH and London: University Press of New England, 97-118; repr. in B.D. Cooperman (ed.) *Jewish Thought in the Sixteenth Century*, Cambridge, MA: Harvard University Press, 1983, 1-22.(This scholarly masterpiece situates Messer Leon's rhetoric in the broader context of ancient, medieval and Renaissance approaches.)

Bonfil, R. (1992) 'The Book of the Honeycomb's Flow by Judah Messer Leon: The Rhetoric Dimension of Jewish Humanism in Fifteenth-Century Italy', in *The Frank Talmage Memorial Volume*, Haifa: Haifa University Press, vol. 2, 21-33.(A superb analysis of Messer Leon's rhetorical and philosophical activities. Indispensable for English readers without access to Bonfil's Hebrew essay.)

Carpi, D. (1972) 'R. Yehudah Messer Leon u-feulato ke-rofe' (R. Judah Messer Leon and His Activity as a Physician), *Michael* 1: 277-301; repr. in *Between Renaissance and Ghetto: Essays on the History of the Jewish in Italy in the 14th and 17th Centuries*, Tel Aviv: University Publishing Project, 1989; trans. D. Carpi, 'Notes on the Life of Rabbi Judah Messer Leon', in E. Toaff (ed.) *Studi sull'ebraismo italiano in memoria di Cecil Roth*, Rome: Barulli, 1975.(This ground-breaking essay established the salient facts of Messer Leon's life and medical work using newly discovered archival material.)

Ivry, A.L. (1983) 'Remnants of Jewish Averroism in the Renaissance', in B.D. Cooperman (ed.) *Jewish Thought in the Sixteenth Century*, Cambridge, MA: Harvard University Press, 243-65.(Surveys some of Messer Leon's younger contemporaries, focusing on Elijah del Medigo.)

Lesley, A.M. (1988) 'Jewish Adaptation of Humanist Concepts in Fifteenth- and Sixteenth-Century Italy', in M.C. Horowitz (ed.) *Renaissance Rereadings: Intertext and Context*, Urbana and Chicago, IL: University of Illinois Press, 51-65.(An important exposition of Jewish polemical adaptation of humanist educational ideals and aesthetic sensibilities.)

Melamed, A. (1978) 'Rhetorica u-Filosofia be-Sefer *Nofet Tzufim*' (Rhetoric and Philosophy in *Nofet Tzufim* by R. Judah Messer Leon), *Italia* 1: 7-38.(Explores Messer Leon's views on the relationship between eloquence and wisdom and highlights the link between rhetoric and political theory in Jewish thought in Italy.)

Ruderman, D. (1981) *The World of a Renaissance Jew: The Life and Thought of Abraham ben Mordecai Farissol*, Cincinnati, OH: Hebrew Union College Press.(A detailed monograph about a student of Messer Leon's who popularized his works on logic among Italian Jews. A rich reconstruction of fifteenth-century Jewish culture in Italy.)

Ruderman, D. (1988) 'The Italian Renaissance and Jewish Thought', in A. Rabil, Jr (ed.) *Renaissance Humanism: Foundations and Forms*, Philadelphia, PA: University of Pennsylvania Press, vol. 1, 382-433. (Excellent overview, relates Messer Leon to the Aristotelian, Renaissance humanist, and Platonic traditions.)

Rosenberg, S. (1973) 'Logica ve-ontologiah ba-filosofia ha-hehudit ba-me'ah ha-arba esreh' (Logic and Ontology in Jewish Philosophy in the Fourteenth Century), unpublished Ph.D. dissertation, Jerusalem: Hebrew University.(Still the only survey of Jewish logic in the late Middle Ages; contains invaluable biographical and

bibliographical information about Jewish logicians and the only analysis of general trends and themes in Jewish logic, including Messer Leon’s attempt to shift Jewish logic from the Judaeo-Arabic to the Latin-scholastic model.)

Tirosh-Rothschild, H. (1991) *Between Worlds: The Life and Thought of Rabbi David ben Judah Messer Leon*, Albany, NY: State University of New York Press.(A monograph on Judah Messer Leon’s son, illustrating the shift from Aristotelian to Kabbalistic trends between 1450 and 1550. Chapter 1 summarizes the educational and political activities and reviews the recent scholarship.)

Metaphor

A standard dictionary definition describes a metaphor as 'a figure of speech in which a word or phrase literally denoting one kind of object is used in place of another to suggest a likeness between them'. Although the theoretical adequacy of this definition may be questioned, it conveys the standard view that there is a difference between literal and nonliteral language; that figurative speech is nonliteral language and that a metaphor is an instance of figurative speech.

The three most influential treatments of metaphor are the comparison, interaction and speech act theories. According to the first, every metaphor involves a comparison; a specific version of this view is that every metaphor is an abbreviated simile. According to the second, every metaphor involves a semantic interaction between some object or concept that is literally denoted by some word, and some concept metaphorically predicated on that word. According to the third, it is not words or sentences that are metaphorical but their use in specific situations; thus, to understand how metaphors function, one must understand how people communicate with language.

1 The comparison theory

The comparison theory asserts that every metaphor compares two things, one of which is designated by a word or phrase used literally and the other of which is designated by a word or phrase used metaphorically. The most interesting version of the comparison view of metaphor is the simile theory. According to its originator, Aristotle (*Rhetoric* III.4), a metaphor is an abbreviated simile: the sentence 'My lover is a red rose' is shorthand for 'My lover is like a red rose'. (The simile theory is also sometimes known as the abbreviation or substitution theory, because a metaphor is represented as being an abbreviation of or a substitution for a simile.) The merit of this view is its double simplicity. First, what looked like two things (metaphors and similes) are said to be one thing; second, the analysis that is given to similes is sometimes quite simple. A simile is true just in case there is some feature that each object has. The feature common to both constitutes the similarity.

One objection to this theory is that what has been represented as its first merit is in fact a misrepresentation. Metaphors and similes are different things because typically if a metaphor were literally asserted, the result would be a falsehood, while (virtually) all similes are true. For example, the metaphor 'My lover is a red rose' is literally false, while the simile 'My lover is like a red rose' is literally true. The pretence of saying something false is a crucial feature of how metaphors work. To hold that metaphorical sentences are in fact similes obscures this feature. This is not to deny that metaphors typically communicate a truth in some way; it is rather to hold that a metaphor communicates its truth in a different way to a simile.

Metaphors (unlike similes) do not state or express truths; they imply them. In order to speak truly, the words of a sentence must be conventionally or semantically appropriate to each other, but the words of a metaphor are not semantically appropriate in this sense. For example, in the metaphor 'Sally is a block of ice', 'Sally' and 'a block of ice' belong to categories so different that to juxtapose them results not merely in a falsehood but in an absurdity. Someone may reply that, since 'Sally is a block of ice' can be used metaphorically, 'Sally is not a block of ice' can also be used metaphorically; and since this latter metaphor results in a true assertion, its words cannot be semantically inappropriate to each other. The proper reply to this objection is to deny that 'Sally is not a block of ice' is a metaphor. (To assert it is to assert something that is obviously true, but that does not mean that the assertion is pointless any more than to assert that 'Business is business' need be pointless. To assert that 'Sally is not a block of ice' may imply that one rejects the metaphorical utterance of 'Sally is a block of ice'.) Sentences that may appear to be negations of metaphorical sentences, such as John Donne's memorable sentiment 'No man is an island', are simply instances of *meiosis* (intentionally saying something logically weaker than the speaker intends to convey, usually by using a negative). Philosophers who consider 'No man is an island' and 'Sally is not a block of ice' to be metaphors, are in effect conflating metaphor and meiosis. They make it difficult or impossible to distinguish something obviously, literally false ('Business is theft') from something obviously, literally true ('Business is business').

To deny an assertion is *ipso facto* to make an assertion; but to deny a metaphor is not to make a metaphor. Assertions are a type of speech act, performed through a literal use of words. But a metaphor is not a kind of

speech act and relies on the literal use of words only in order to indicate that the speaker means to communicate something other than their literal meaning.

Another deficiency of the standard simile theory is that it does not account for the fact that some effort or mental computation is required in order to work out what a metaphor means, in contrast with a simile which usually means what it says. In order to meet this objection, a quasi-simile theory has been proposed. Suppose someone says ‘My lover is a red rose’. To understand this metaphor, one considers the set of red roses plus the speaker’s lover and then tries to work out which property or properties all the members of this set share; in other words, to say that one’s lover is a red rose is to say that they are similar in some respects to red roses. The objection to this theory is the same as one raised to the original theory. As Davidson (1979: 36) points out, a metaphor only intimates what a simile declares.

2 The interaction theory

According to the interaction theory, a metaphor involves an interaction between a literal element in a sentence and a metaphorical element. So there are always two parts to a metaphor. Max Black (1962) calls the literal element the frame and the metaphorical element the focus. Searle (1979) has effectively refuted this version of the theory by pointing out that some metaphorical sentences do not contain any literal element which could plausibly serve the role of frame. For example, the sentence ‘The bad news congealed into a block of ice’ is a mixed metaphor, but it contains no word with a literal meaning that can serve as the frame. Another objection to the interaction theory is that the term ‘interaction’ is itself a metaphor. Since there is no literal interaction between words in a sentence or between concepts, the interaction theory never actually explains or describes how metaphors work. More recent treatments of the theory explain interaction as the placement of the object literally denoted within a conceptual system of commonplaces associated with the metaphorical term. Thus, to say ‘Mary is a wolf’ is to associate Mary and her attributes with the complex of properties associated with wolves: cunning, viciousness, unpredictability and so on. None the less ‘interaction’ seems to be the wrong word to convey this feature. Also, the interaction theory claims that the meaning of the words is affected by this interaction and that seems to be an inaccurate description of the phenomenon.

Donald Davidson has insisted that semantically metaphors mean what the words used to convey them literally mean and nothing more. His thesis constitutes a direct rejection of all views that hold that a distinction must be drawn at the level of words or sentences between two kinds of meaning: literal meaning and metaphorical meaning. While the literal meaning is usually taken as obvious, the metaphorical meaning is often said to be ineffable and ineffability is given as the reason why metaphors are used. Sceptics about ineffability would object to a kind of meaning that is inherently mysterious or unintelligible, even if individual speakers are unable to articulate what they mean. In any case, what is needed is an explanation of how metaphors are understood. Here, Davidson seems to be right when he says that understanding a metaphor is the same kind of activity as understanding any other linguistic utterance and all understanding requires an ‘inventive construal’ of what the literal meaning of the metaphorical utterance is and what the speaker believes about the world. Furthermore, this activity is ‘little guided by rules’. Making a metaphor, like speaking generally, is a ‘creative endeavour’.

3 The speech act theory

It is fair to demand that a theory of metaphor fit into some larger theory of language or communication. While the comparison and interaction theories do not meet this demand, the speech act theory does. H.P. Grice (1989) distinguished between ‘utterance occasion meaning’ (the meaning that an utterance has on some particular occasion of its use) and ‘timeless utterance meaning’ (the meaning an utterance has based upon its usual use and independent of any particular time). This latter form of meaning can be associated with literal or conventional meaning. In a metaphor, the speaker intends to have the audience work out the ‘occasion’ meaning of their utterance in part by understanding its ‘timeless’ meaning (see Grice, H.P.).

How does this happen? Grice points out that normal conversation is directed by the cooperative principle: make your contribution to a conversation appropriate. This principle issues in various conversational maxims. For example, what a person says should be truthful and supported by evidence; it should be as informative as the audience needs it to be and not more so; and the information should be relevant to the topic of conversation. These maxims are related to the content of the message. There are also maxims concerning how the message is to be

expressed: it should be clear, unambiguous, brief and orderly. Of course these maxims are not always fulfilled. Sometimes they go unfulfilled quietly and unostentatiously, such as when a person lies or makes an innocent mistake. Sometimes they are flouted, that is, openly and ostentatiously unfulfilled in order to achieve a certain effect. For example, one way to show disgust for or disappointment with someone is to say ‘You’re a fine friend’. In such instances of sarcasm or irony, the speaker flouts the maxim of truthfulness and thereby within the context of the conversation means exactly the opposite of what appears to be said. Other figures of speech can be explained in similar ways. Hyperbole is intentionally making an exaggerated statement in order to achieve a certain effect. Within the theory being sketched, hyperbole is also achieved by flouting the maxim of truthfulness: the speaker makes a statement that is logically stronger than the proposition they intend to communicate.

Within this theory, meiosis is achieved by flouting the maxim of quantity (which enjoins the speaker to make their statement logically as strong as possible); for example, by saying, ‘John D. Rockefeller was not poor’. The audience, trying to make sense of the speaker’s apparent nonfulfilment of the cooperative principle, infers that the speaker means something stronger than was said.

Metaphor fits into the same pattern of explanation although the calculation of what a metaphor means is more complicated. If someone says ‘My lover is a red rose’ in circumstances that are otherwise unexceptional, it is sensible for the audience to judge that the speaker is flouting the maxim of truthfulness. What then does the speaker mean? The audience’s strategy in the case of metaphor, as in every other case of linguistic understanding, is to achieve a cognitive equilibrium by accepting those propositions that make the most sense of the situation. This involves making new beliefs fit in as smoothly as possible with existing beliefs and sometimes revising old beliefs as part of this process. Briefly, for the example above, it makes sense for the audience to believe that the speaker intends the audience to recognize that the speaker intends the audience to think of certain qualities of a rose - its beauty, value, delicacy, fragrance - and then attribute them to the lover. Such a search for cognitive equilibrium is close to what Davidson meant when he said that understanding language is ‘a creative endeavour’.

See also: Communication and intention; Davidson, D.; Speech acts

A.P. MARTINICH

References and further reading

Aristotle (*c.*mid-4th century BC) *Rhetoric*, trans. G. Kennedy, *Aristotle on Rhetoric: A Theory of Civic Discourse*, New York: Oxford University Press, 1991, III 4.(The first statement of the simile theory.)

Black, M. (1962) *Models and Metaphors*, Ithaca, NY: Cornell University Press. (Contains the most influential presentation of the interaction theory.)

Davidson, D. (1979) ‘What Metaphors Mean’, in S. Sacks (ed.) *On Metaphor*, Chicago, IL: University of Chicago Press, 29-45.(Davidson emphasized that words themselves do not have a metaphorical meaning.)

Goodman, N. (1968) *Languages of Art*, Indianapolis, IN: Bobbs-Merrill.(A sophisticated version of the interaction theory.)

Grice, H.P. (1989) ‘Logic and Conversation’, in *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, 22-40.(Contains Grice’s general theory of conversation, including a treat ment of metaphor.)

Johnson, M. (1980) *Philosophical Perspectives on Metaphor*, Minneapolis, MN: University of Minnesota Press. (A collection of important articles on metaphor.)

Martinich, A.P. (1984) ‘A Theory for Metaphor’, *Journal of Literary Semantics* 13: 35-54; repr. in S. Davis (ed.) *Pragmatics*, Cambridge: Cambridge University Press, 1991, 507-18.(A speech act theory of metaphor with criticisms of Searle’s version.)

Noppen, J.P. van (ed.) (1985, 1990) *Metaphor*, vol. 1, *A Bibliography of Post-1970 Publications*; vol. 2, *A Classified Bibliography of Publications*, Amsterdam: Benjamins.(Continues the bibliographical work begun by Shibles.)

Ortony, A. (ed.) (1979) *Metaphor and Thought*, Cambridge: Cambridge University Press; 2nd edn, 1994. (Contains important articles by philosophers, psychologists, linguists and literary theorists.)

Searle, J. (1979) ‘Metaphor’, in A. Ortony (ed.) *Metaphor and Thought*, Cambridge: Cambridge University Press; 2nd edn, 1994.(A speech act theory of metaphor in an important volume that contains contributions by philosophers, psychologists, linguists and literary theorists.)

Shibles, W. (1971) *Metaphor: An Annotated Bibliography and History*, Whitewater, WI: Language Press.(The most complete bibliography on metaphor up to 1970.)

Metaphysics

Metaphysics is a broad area of philosophy marked out by two types of inquiry. The first aims to be the most general investigation possible into the nature of reality: are there principles applying to everything that is real, to all that is? - if we abstract from the particular nature of existing things that which distinguishes them from each other, what can we know about them merely in virtue of the fact that they exist? The second type of inquiry seeks to uncover what is ultimately real, frequently offering answers in sharp contrast to our everyday experience of the world. Understood in terms of these two questions, metaphysics is very closely related to ontology, which is usually taken to involve both ‘what is existence (being)?’ and ‘what (fundamentally distinct) types of thing exist?’ (see Ontology).

The two questions are not the same, since someone quite unworried by the possibility that the world might really be otherwise than it appears (and therefore regarding the second investigation as a completely trivial one) might still be engaged by the question of whether there were any general truths applicable to all existing things. But although different, the questions are related: one might well expect a philosopher’s answer to the first to provide at least the underpinnings of their answer to the second. Aristotle proposed the first of these investigations. He called it ‘first philosophy’, sometimes also ‘the science of being’ (more-or-less what ‘ontology’ means); but at some point in antiquity his writings on the topic came to be known as the ‘metaphysics’ - from the Greek for ‘after natural things’, that is, what comes after the study of nature. This is as much as we know of the origin of the word (see Aristotle §11 and following). It would, however, be quite wrong to think of metaphysics as a uniquely ‘Western’ phenomenon. Classical Indian philosophy, and especially Buddhism, is also a very rich source (see Buddhist philosophy, Indian; Hindu philosophy; Jaina philosophy).

1 General metaphysics

Any attempt on either question will find itself using, and investigating, the concepts of being and existence (see Being; Existence). It will then be natural to ask whether there are any further, more detailed classifications under which everything real falls, and a positive answer to this question brings us to a doctrine of categories (see Categories). The historical picture here is complex, however. The two main exponents of such a doctrine are Aristotle and Kant. In Aristotle’s case it is unclear whether he saw it as a doctrine about things and their basic properties or about language and its basic predicates; whereas Kant quite explicitly used his categories as features of our way of thinking, and so applied them only to things as they appear to us, not as they really or ultimately are (see Kant, I.). Following on from Kant, Hegel consciously gave his categories both roles, and arranged his answer to the other metaphysical question (about the true underlying nature of reality) so as to make this possible (see Hegel, G.W.F.).

An early, extremely influential view about reality seen in the most general light is that it consists of things and their properties - individual things, often called particulars, and properties, often called universals, that can belong to many such individuals (see Particulars; Universals). Very closely allied to this notion of an individual is the concept of substance, that in which properties ‘inhere’ (see Substance). This line of thought (which incidentally had a biological version in the concepts of individual creatures and their species) gave rise to one of the most famous metaphysical controversies: whether universals are real entities or not (see Species; Natural kinds). In different ways, Plato and Aristotle had each held the affirmative view; nominalism is the general term for the various versions of the negative position (see Nominalism).

The clash between realists and nominalists over universals can serve to illustrate a widespread feature of metaphysical debate. Whatever entities, forces and so on may be proposed, there will be a *prima facie* option between regarding them as real beings, genuine constituents of the world and, as it were, downgrading them to fictions or projections of our own ways of speaking and thinking (see Objectivity; Projectivism). This was, broadly speaking, how nominalists wished to treat universals; comparable debates exist concerning causality (see Causation), moral value (see Emotivism; Moral realism; Moral scepticism; Value, ontological status of) and necessity and possibility (see Necessary truth and convention) - to name a few examples. Some have even proposed that the categories (see above) espoused in the Western tradition are reflections of the grammar of Indo-European languages, and have no further ontological status (see Sapir-Whorf hypothesis).

Wittgenstein famously wrote that the world is the totality of facts, not of things, so bringing to prominence another concept of the greatest generality (see Facts). Presumably he had it in mind that exactly the same things, differently related to each other, could form very different worlds; so that it is not things but the states of affairs or facts they enter into which determine how things are. The apparent obviousness of the formula 'if it is true that *p* then it is a fact that *p*', makes it seem that facts are in one way or another closely related to truth (see Truth, coherence theory of; Truth, correspondence theory of) - although it should be said that not every philosophical view of the nature of truth is a metaphysical one, since some see it as just a linguistic device (see Truth, deflationary theories of) and some as a reflection, not of how the world is, but of human needs and purposes (see Truth, pragmatic theory of; Relativism).

Space and time, as well as being somewhat elusive in their own nature, are further obvious candidates for being features of everything that exists (see Space; Time). But that is controversial, as the debate about the existence of abstract objects testifies (see Abstract objects). We commonly speak, at least, as if we thought that numbers exist, but not as if we thought that they have any spatio-temporal properties (see Realism in the philosophy of mathematics). Kant regarded his things-in-themselves as neither spatial nor temporal; and some have urged us to think of God in the same way (see God, concepts of). There are accounts of the mind which allow mental states to have temporal, but deny them spatial properties (see Dualism).

Be all this as it may, even if not literally everything, then virtually everything of which we have experience is in time. Temporality is therefore one of the phenomena that should be the subject of any investigation which aspires to maximum generality. Hence, so is change (see Change). And when we consider change, and ask the other typically metaphysical question about it ('what is really going on when something changes?') we find ourselves faced with two types of answer. One type would have it that a change is an alteration in the properties of some enduring thing (see Continuants). The other would deny any such entity, holding instead that what we really have is merely a sequence of states, a sequence which shows enough internal coherence to make upon us the impression of one continuing thing (see Momentariness, Buddhist doctrine of). The former will tend to promote 'thing' and 'substance' to the ranks of the most basic metaphysical categories; the latter will incline towards events and processes (see Events; Processes). It is here that questions about identity over time become acute, particularly in the special case of those continuants (or, perhaps, processes), which are persons (see Identity; Persons; Personal identity).

Two major historical tendencies in metaphysics have been idealism and materialism, the former presenting reality as ultimately mental or spiritual, the latter regarding it as wholly material (see Buddhism, Yogācāra school of; Idealism; Materialism; Materialism in the philosophy of mind; Monism, Indian; Phenomenalism). In proposing a single ultimate principle both are monistic (see Monism). They have not had the field entirely to themselves. A minor competitor has been neutral monism, which takes mind and matter to be different manifestations of something in itself neither one nor the other (see Neutral monism). More importantly, many metaphysical systems have been dualist, taking both to be fundamental, and neither to be a form of the other (see Sāōkhya). Both traditions are ancient. In modern times idealism received its most intensive treatment in the nineteenth century (see Absolute, the; German idealism); in the second half of the twentieth century, materialism has been in the ascendant. A doctrine is also found according to which all matter, without actually being mental in nature, has certain mental properties (see Panpsychism).

2 Specific metaphysics

There is also metaphysics that arises in reference to particular subject matters, this being therefore metaphysical primarily with regard to the second question (what are things ultimately like? - or, what kinds of thing ultimately exist?) rather than the first. One of the most obvious cases, and historically the most prominent, is theology; we have already mentioned the philosophy of mind, the philosophy of mathematics and the theory of values. Less obviously, metaphysical issues also intrude on the philosophy of language and logic, as happens when it is suggested that any satisfactory theory of meaning will have to posit the existence of intensional entities, or that any meaningful language will have to mirror the structure of the world (see Intensional entities; Logical atomism). The political theorist or social scientist who holds that successful explanation in the social sphere must proceed from properties of societies not reducible to properties of the individuals who make them up (thereby making a society an entity that is in a sense more basic than its members) raises a metaphysical issue (see Holism and individualism

in history and social science). Metaphysics, as demarcated by the second question, can pop up anywhere.

The relationship with metaphysics is, however, particularly close in the case of science and the philosophy of science. Aristotle seems to have understood his ‘first philosophy’ as continuous with what is now called his physics, and indeed it can be said that the more fundamental branches of natural science are a kind of metaphysics as it is characterized here. For they are typically concerned with the discovery of laws and entities that are completely general, in the sense that everything is composed of entities and obeys laws. The differences are primarily epistemological ones, the balance of a priori considerations and empirical detail used by scientists and philosophers in supporting their respective ontological claims. The *subject matter* of these claims can even sometimes coincide: during the 1980s the reality of possible worlds other than the actual one was maintained by a number of writers for a variety of reasons, some of them recognizably ‘scientific’, some recognizably ‘philosophical’ (see Possible worlds). And whereas we find everywhere in metaphysics a debate over whether claims should be given a realist or an antirealist interpretation, in the philosophy of science we find a parallel controversy over the status of the entities featuring in scientific theories (see Realism and antirealism; Scientific realism and antirealism).

It is true that there has been considerable reluctance to acknowledge any such continuity. A principal source of this reluctance has been logical positivism, with its division of propositions into those which are empirically verifiable and meaningful, and those which are not so verifiable and are either analytic or meaningless, followed up by its equation of science with the former and metaphysics with the latter (see Demarcation problem; Logical positivism; Meaning and verification). When combined with the belief that analytic truths record nothing about the world, but only about linguistic convention, this yields a total rejection of all metaphysics - let alone of any continuity with science. But apart from the fact that this line of thought requires acceptance of the principle about meaninglessness, it also makes a dubious epistemological assumption: that what we call science never uses non-empirical arguments, and that what we regard as metaphysics never draws on empirical premises. Enemies of obscurantism need not commit themselves to any of this; they can recognize the continuity between science and metaphysics without robbing anyone of the vocabulary in which to be rude about the more extravagant, ill-evidenced, even barely meaningful forms which, in the view of some, metaphysics has sometimes taken.

Even the philosopher with a low opinion of the prospects for traditional metaphysics can believe that there is a general framework which we in fact use for thinking about reality, and can undertake to describe and explore it. This project, which can claim an illustrious ancestor in Kant, has in the twentieth century sometimes been called descriptive metaphysics, though what it inquires into are our most general patterns of thought, and the nature of things themselves only indirectly, if at all. Though quite compatible with a low estimate of traditional metaphysics as defined by our two primary questions, it does imply that there is a small but fairly stable core of human thought for it to investigate. Hence it collides with the view of those who deny that there is any such thing (see Postmodernism).

See also: Aristotelianism in Islamic philosophy; Aristotelianism in the 17th century; Aristotelianism, medieval; Aristotelianism, Renaissance; Brahman; Buddhist concept of emptiness; Causality and necessity in Islamic thought; Causation, Indian theories of; Contingency; Hegelianism; Infinity; Materialism, Indian school of; Matter, Indian conceptions of; Negative facts in classical Indian philosophy; Neo-Kantianism; Neoplatonism; Neoplatonism in Islamic philosophy; Nominalism, Buddhist doctrine of; Ontology in Indian philosophy; Platonism, Early and Middle; Platonism in Islamic philosophy; Platonism, medieval; Platonism, Renaissance; Potentiality, Indian theories of; Process philosophy; Religion, philosophy of; Self, Indian theories of; Universals, Indian theories of

EDWARD CRAIG

References and further reading

Kim, J. and Sosa, E. (eds) (1995) *A Companion to Metaphysics*, Oxford: Blackwell. (Encyclopedia-style volume with over 250 entries of varying length devoted to terms, theories, movements and individual philosophers. Some coverage of epistemological issues as well as metaphysics.)

Methodological individualism

Methodological individualism (MI) is the thesis that certain psychological properties are intrinsic *properties, such as 'being made out of iron', rather than externally* relational *properties, such as 'being an aunt'. It has been challenged by influential 'anti-individualist' claims of, for example, Putnam and Burge, according to which the content of an individual's words or thoughts (beliefs, desires) is determined in part by facts about their social or physical environment. Putnam, for example, imagines a planet, 'twin earth', which is identical to the earth in 1750 (prior to modern chemistry) in all respects except that wherever earth had H_2O, twin earth had a different but superficially similar chemical, XYZ. Putnam argues that the English word 'water' in 1750 referred only to H_2O, while the twin word 'water' refers only to XYZ.*

Historically, the term 'methodological individualism' has referred to the thesis that all social explanation must be ultimately expressible in terms of facts about individual human beings; not about economic classes, nations and so on. For a treatment of this subject, see Explanation in history and social science §1.

1 Common arguments for MI

Methodological individualism (MI) is a view about the representational contents of psychological states. There is a distinction between two kinds of such contents: narrow and wide (see Content: wide and narrow). A content is 'wide' if its identity depends upon relations its thinkers bear to things outside their heads. Suppose, for example, that I am looking at a particular banana, *B*, and I think 'That banana is ripe'. My thought is then about that object, *B*. The thought is true if and only if *B* is ripe. In that sense, the content of the thought relates essentially to the banana itself. Had there been a different banana, *A*, my thought would have had a different content: it would have been true if and only if *A* had been ripe.

Content is 'narrow' if it is not relational, but is instead fully determined by intrinsic physical properties of individuals. Thus, any two individuals who were identical in all intrinsic physical respects would be identical in respect of all their narrow contents. Call such individuals 'twins'. Narrow content is content shared by twins (see Identity of indiscernibles §3). The above example of 'twin earth', due to Putnam (1975), challenges the view that the ordinary content we ascribe to people is narrow in this sense: the thought contents expressed by 'Water is wet' by the 1750 earthling and their twin-earthling would seem to be different, the first making a claim about H_2O, the second about XYZ.

MI is the view that, despite such twin cases, every psychological state has a narrow content, and that this content is, or should be, essential for psychological explanations. Thus, consider a pair of twins one of whom is thinking about banana *A* and the other about banana *B*. Each thinks 'That banana is ripe'. The methodological individualist will claim that the two thoughts have a content in common; a narrow content. They concede that the thoughts are about different external objects, but hold that the concepts involved - one twin's concept of *A* and the other's concept of *B* - are exactly the same. And this shared content, the one that includes exactly similar concepts, is crucial to psychological explanation. The same applies to Putnam's twins: their thoughts also have a shared narrow content.

Anti-individualists have presented three kinds of argument to show that contents attributed by both common sense and scientific psychological theory are wide rather than narrow. The first of these involves thought experiments, like Putnam's, in which one conceives of twins who inhabit environments that differ in crucial respects, and it is claimed that there is a strong intuition that the individuals' psychological states would have different contents. The second defends wide content on the basis of independently motivated theories of content, such as causal or teleo-functional theories (see Semantics, teleological; Semantics, informational). The third flow from studies of specific branches of scientific psychology. Since the first two kinds of argument are treated in other entries, the present entry will discuss only the third, considering the question of whether contents attributed by scientific psychology are narrow or wide.

2 MI and the causal role of psychological properties

Jerry Fodor (1987) gives the following argument for MI.

(1) Psychological explanation is causal explanation. For example, we might cite someone's belief that water is good for plants as part of a causal explanation of why they filled a watering-can.
(2) A science that offers causal explanation individuates its subject matter with respect to causal powers. Specifically, it would not classify states as being of different kinds if they have the same causal powers.
(3) The causal powers of the psychological states of twins are the same. So, psychology would classify twins' psychological states as being of the same kind, hence as having the same content.

This argument may be questioned on various grounds (see, for example, Burge (1986); Fodor (1994), who has himself abandoned it). For example, the second premise is not obviously correct. The individuative practice of a science that offers causal explanations might be sensitive to various factors, not all of which are directly relevant to causal powers. For instance, a scientist might cite the action of the heart as part of a causal explanation of the workings of the cardiovascular system. However, the individuation conditions for hearts are not purely causal, but also involve historical, evolutionary considerations. Nevertheless, Fodor's argument has been influential, and methodological individualists may feel that there is something importantly right about it, even if it is not conclusive as it stands (see Mental causation §3).

3 MI and the computational theory of vision

One branch of psychology has been the focus of particular attention by philosophers discussing MI. This is the computational theory of vision, developed by David Marr (1982) and others (see Vision §4). The theory has received much attention because it is among the most developed areas of cognitive science. This makes it possible to look in detail at the actual practice and methodology of a psychological theory and give substance to the debates about the individuation of psychological states.

Tyler Burge (1982; 1986) argues that the theory has an anti-individualist methodology. He argues that the theory assumes that the human visual system has evolved to be successful on earth, and certain of the earthly environmental conditions that account for its success are relevant to the individuation of its representational states. Roughly, Burge's idea is that the content of certain representations is partly determined by what 'normally' cause them, where the 'normal' cause of a type of representation is the type of external condition that has caused it in the past, during the (phylogenetic) evolution of the system, under circumstances where it has been successful.

The argument proceeds by stating that one could envisage alternative possible circumstances in which intrinsically identical visual systems had evolved, but in which the 'normal' causes of some of the representations were different. In such a case, the representational contents, as attributed by the theory, would have been different from what they actually are. Hence MI is violated.

The situation Burge envisages can be schematically represented. On earth (E), the 'normal' cause of a particular physical type, *T*, of representation in the visual system is *N*. Thus *T*s on E represent *N*. In a possible counterfactual circumstance, *C*, the prevailing conditions differ, but the visual system is intrinsically the same. In *C*, the 'normal' cause of *T* is *W*, where *W* is different from *N*. For example, *N* might be small, thin cracks and *W* might be small, thin shadows. When a visual system produces a representation of type *T* on E, it represents the presence of an *N* (its content being something like 'There is a small, thin crack'). When an intrinsically identical visual system in *C* produces a *T*, it represents the presence of a *W* (its content being something like 'There is a small, thin shadow').

Burge's argument is defended by a study of the theory of vision. However, it may be questioned on at least two grounds. First, it can be questioned whether the theory really accords such a crucial role to evolutionary considerations in its attributions of content. While it is true that Marr invokes evolutionary considerations at various points, it is far from clear that these considerations enter into the individuation of the contents of visual states in the decisive way that Burge claims (see Segal 1989). Second, even if the premises of the argument are granted, its validity can be questioned.

One way to question the validity of the argument involves considering in more detail the behaviour of twin subjects when they undergo a *T*. Let us consider two cases.

Case 1. *N*s are circles and *W*s are squares: these are the 'normal' causes of *T*s on E and *C*, respectively. We can suppose that light bends in strange ways in *C*, so that squares there cause the same sorts of retinal images as circles do on E. Suppose we ask subjects who are undergoing a *T* to use their hands to trace in the air the shape they

perceive. Since the subjects in the two environments are twins, they will make the same movements. If they trace a circle, this will indicate that *C* subjects are misperceiving squares as circles. *Mutatis mutandis* if they trace a square. Thus the twins would appear to be undergoing representations with the same content after all.

This first case brings out the point that vision must be successful in both environments, if Burge's argument is to work. The second case brings out a consequence of this point.

Case 2. Suppose that *W*s are thin shadows and *N*s are thin cracks. If the success constraint is to be met, then the subjects must not behave as if they are visually representing the items specifically as cracks or as shadows. If they did, then the behaviour in at least one environment would be inappropriate, and the constraint would be violated. If the constraint is to be met, the behaviour must be indiscriminately appropriate to either shadows or cracks. And this would suggest that the visual representations are not specific to either but represent the items in some more general way. They would represent them as a feature of surfaces that could be either a shadow or a crack: a thin, dark line perhaps.

The second case undermines the validity of Burge's argument, since it allows the premises to remain in place but negates the conclusion. If *T*s represent thin, dark lines in both environments, then it may remain true that the visual systems are successfully adapted and *T* type representations represent their 'normal' causes. For thin cracks and thin shadows are all instances of thin, dark lines. This point can be generalized to the schematic version of Burge's envisaged situation: *T*s do not represent *N*s or *W*s specifically as *N*s or as *W*s, but rather represent them by some more general content, *G*, where *G* includes both *N*s and *W*s. (The objection is given in Segal (1989), and related points are made by McGinn (1989) and Matthews (1988). For extensive discussion, see Egan (1991; 1992), Davies (1991), Wilson (1994) and Segal (1991).)

The debate between methodological individualists and their opponents remains unresolved. This is because the arguments are complete on neither side, both in the particular case of vision, which has been studied in detail, and in the more general, but more complex, case of psychology as a whole. Anti-individualists have tended to rest their positions on arguments the premises and validity of which may easily be doubted, or on general theories of content that are themselves highly controversial. On the other hand, no convincing considerations in favour of a generalized methodological individualism have yet been offered.

See also: Content: wide and narrow; Putnam, H.; Semantics, conceptual role; Semantics, teleological

GABRIEL SEGAL

References and further reading

Burge, T. (1979) 'Individualism and the Mental', in P. French, T. Uehling and H. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 4, Minneapolis, MN: University of Minnesota Press.(The *locus classicus* of the argument that social factors are relevant to the individuation of the content of psychological states. The argument is spelt out in some detail and many objections are anticipated and responded to.)

Burge, T. (1982) 'Other Bodies', in A. Woodfield (ed.) *Thought and Object: Essays on Intentionality*, Oxford, Clarendon Press.(Arguing that the content of non-singular concepts, such as concepts of water or arthritis, should not be thought of as analogous to indexical singular concepts, such as those expressed by 'now', 'here', 'this'.)

Burge, T. (1986) 'Individualism and Psychology', *Philosophical Review* 95 (1): 3-46.(Arguing against the view that scientific psychology must be individualistic and for the view that the computational theory of vision is anti-individualistic.)

Crane, T. (1991) 'All the Difference in the World', *Philosophical Quarterly* 41 (162): 1-26.(A relatively accessible article challenging anti-individualist interpretations of twin earth experiments.)

Cummins, R. (1989) *Meaning and Mental Representation*, Cambridge, MA: MIT Press.(A general work on theories of contents from the perspective of philosophy of science, including a defence of individualism.)

Davies, M. (1991) 'Individualism and Perceptual Content', *Mind* 100 (4): 461- 84.(An article clarifying and defending Burge's argument (1986) that the computational theory of vision is anti-individualistic.)

Egan, F. (1991) 'Must Psychology Be Individualistic?', *Philosophical Review* 100 (2): 179-203.(Criticizing Burge (1986) and arguing that the computational theory of vision is individualistic.)

Egan, F. (1992) 'Individualism, Computationalism and Perceptual Content', *Mind* 101: 443-59.(Developing the individualism of Egan (1991).)

Field, H. (1978) ‘Logic, Meaning and Conceptual Role’, *Journal of Philosophy* 74: 379-409.(A relatively difficult article defending an individualist, functional role theory of psychological states.)

Fodor, J. (1981) ‘Methodological Solipsism Considered as a Research Strategy in Cognitive Science’, in *RePresentations*, Cambridge, MA: MIT Press.(An early defence of individualism based on a computational theory of cognition.)

Fodor, J. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.(A general work on content. Chapter 2 includes an extended defence of individualism.)

Fodor, J. (1991) ‘A Modal Argument for Narrow Content’, *Journal of Philosophy* 78 (1): 5-26.(A relatively difficult article presenting an argument for narrow content based on certain modal considerations.)

Fodor, J. (1994) *The Elm and the Expert*, Cambridge, MA: MIT Press.(A short book in which Fodor withdraws from the individualism of his earlier work, and argues that a certain limited form of anti-individualism is true.)

Grimm, R. and Merrill, D. (eds) (1988) *Contents of Thought*, Tucson, AZ: University of Arizona Press.(Includes several useful articles, including Loar (1988) and Matthews (1988).)

Larson, R. and Segal, G. (1995) *Knowledge of Meaning: An Introduction to Semantic Theory*, Cambridge, MA: MIT Press.(A general introduction to formal semantic theory. Chapter 13 defends individualism with respect to the psychological states involved in understanding language.)

Loar, B. (1988) ‘Social Content and Psychological Content’, in R. Grimm and D. Merrill (eds) *Contents of Thought*, Tucson, AZ: University of Arizona Press.(An influential defence of individualism.)

Marr, D. (1982) *Vision*, New York: Freeman Press.(An introduction to the computational theory of vision. Whether Marr’s work is individualistic or not has been the focus of much debate.)

Matthews, R.J. (1988) ‘Comments on Burge, "Cartesian Error and the Objectivity of Perception"’, in R. Grimm and D. Merrill (eds) *Contents of Thought*, Tucson, AZ: University of Arizona Press.(Discussing aspects of Burge’s anti-individualism).

McGinn, C. (1989) *Mental Content*, Oxford: Blackwell.(A general work on mental content; includes a distinction between different kinds of anti-individualism, and a defence of individualism with respect to the content of perceptual states.)

Millikan, R.G. (1993) *White Queen Psychology and other Essays for Alice*, Cambridge, MA: MIT Press.(A collection of papers, many of which defend and elaborate an anti-individualistic, teleo-functional theory of content.)

Patterson, S. (1991) ‘Individualism and Semantic Development’, *Philosophy of Science* 58 (1): 15-35.(Arguing that a certain influential, psychological theory of concept acquisition is individualistic.)

Putnam, H. (1975) ‘The Meaning of "Meaning"’, in K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science*, vol. 7, *Language, Mind and Knowledge*, Minneapolis, MN: University of Minnesota Press.(The original twin earth experiments.)

Searle, J. (1991) *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.(A general work in the philosophy of mind. Chapter 8 includes an extended defence of individualism.)

Segal, G. (1989) ‘Seeing What is Not There’, *Philosophical Review* 98: 189-214.(Arguing, against Burge (1986), that the computational theory of vision is individualistic.)

Segal, G. (1991) ‘Defence of a Reasonable Individualism’, *Mind* 100 (4): 485-93.(Elaborating and defending the claim that the contents of visual perception are narrow; argues against Davies (1991).)

Wilson, R. (1994) ‘Wide Computationalism’, *Mind* 103: 351-72.(Arguing for anti-individualism with respect to a computational theory of cognition.)

Woodfield, A. (ed.) (1982) *Thought and Object: Essays on Intentionality*, Oxford: Clarendon Press.(A useful collection of relevant articles.)

Mexico, philosophy in

Philosophy has been practised in Mexico for centuries, beginning with Nahuatl thought. Such thinking was rediscovered through laborious translation of surviving fragments of a document of exceptional value known as ‘Coloquio de los Doce’ (Debate of the Twelve) *(1524). Since then, philosophy has come to enjoy a high degree of professionalization and a high quality of academic production. Generally, Mexican philosophical activity has evolved in accordance with world standards of rigour, information and quality of argumentation. In the twentieth century various philosophical groups have been created, namely the Ateneo de la Juventud, the* contemporáneos *and the Hyperion group.*

Leopoldo Zea understood the essence of the Mexican and the Latin American as a historical being with a historically situated consciousness. Zea's history of ideas involved a philosophy of Latin American history which placed the being, destiny and meaning of the history of Mexico and Latin America in the context of the history of the world. The 1940s and 1950s were unusually productive to this end. In the 1970s small groupings of philosophers gathered to focus on problems, traditions, teaching figures, leaders and spheres of influence.

There has been considerable interest in political philosophy, philosophy of history, philosophy of science and ethics. Since the 1980s, works about the history of ideas in Mexico and the history of science and technology have proliferated.

1 Historical overview

Before the arrival of the Europeans, a form of reflection was developed in Nahuatl society which was expressed aphoristically and poetically. Since the 1950s Nahuatl thought has been rediscovered through laborious translation of surviving fragments. A document of exceptional value is one known as ‘Coloquio de los Doce’ (Debate of the Twelve) (1524), a relatively brief fragment located in the Vatican archives. It records the encounter of cultures and the debate sustained by the first twelve Franciscan monk preachers who arrived in New Spain to preach the Gospel. The Franciscans met with native authorities and explained the New Testament to them. The natives believed that the questions were too complex and that they required the intervention of wise, erudite people. Their proposition of calling on their *tlamatinime* (wise persons or philosophers) was accepted by the monks. After listening attentively, the wise Nahuas claimed that the monks appreciated the information and greatly valued the sacrifices made by these foreigners in voyaging to transmit their truths, although the natives themselves had their own truths, which they also set forth. However, the Nahuas conceded that the monks were clearly in charge and could do whatever they wished with the Nahuas anyway. ‘Coloquio de los Doce’ (Debate of the Twelve) is important as testimony of a decisive encounter between the Western world and the so-called New World. Some of the most important topics of debate since the 1980s are addressed in a singularly ceremonial and dramatic manner: the experience of alterity, ethnocentrism, intercultural dialogue and the clash of argumentative styles. These documents reveal a rational, logical, rigorous, erudite and civilized mind on the part of the Nahuas.

After the arrival of the conquistadors from Spain in the sixteenth century a philosophy initially developed in Spain and a system of philosophical production and styles in accordance with the prevailing sociopolitical system was imposed over a period of three centuries (from the early sixteenth century to the mid-nineteenth century). Early on, the Spanish felt it important to establish a university system that would follow the model of the Universities of Salamanca and Alcalá de Henares. The University of Pontificia was founded in New Spain in 1553. The hegemonic philosophical thought in New Spain was fundamentally Scholastic. It was developed in the university, schools and seminaries and was primarily disseminated by the religious orders. Among the most important representatives is Fray Alonso de la Vera Cruz, the first to teach philosophy in the New World at the University of Pontificia in 1553.

Since the early years of the colonial period, two opposing types of reflection have been observed: colonialist and independent thought, the latter of which battles for freedom of thought and action. Sor Juana Inés de la Cruz and the Mexican Jesuits, who collaborated in the formation of a native conscience, were significant proponents of this second position (see Feminist thought in Latin America). At the end of the colonial period the same Scholastics continued to consolidate the thought of the Enlightenment, even when it seemed clear that the inspirational and legitimizing thought of independence was the so-called populist doctrine of Salamanca. The doctrine held that the

people are the source of power and power should devolve to them when the King is not in a position to exercise authority, as was the case with Fernando VII.

The nineteenth century in Mexico has also been re-examined. It has been observed that thought under the rubric of liberalism is not homogeneous and requires multiple approaches. Diverse discourses have been placed in opposition to one another. Towards the end of the nineteenth century, positivism became predominant and remains one of the most debated philosophical movements in Mexican thought (see Positivist thought in Latin America).

In the twentieth century a group of philosophers called the Ateneo de la Juventud quickly launched its anti-positivist spiritualist reaction during the Mexican revolution (1910-17) (see Anti-positivist thought in Latin America). Subsequently, the Contemporáneos constituted another decisive group in the advance of Mexican thought. By the 1950s the Hyperion group had formulated a Mexican philosophy. José Vasconcelos and Antonio Caso of the Ateneo were influential in the world of philosophy, as were Samuel Ramos of the Contemporáneos and Leopoldo Zea, Luis Villoro, Emilio Uranga, Jorge Portilla and Joaquín Sánchez Macgregor of the Hyperion group.

A notable contribution to the development of modern Mexican thought was that by the exiles of the Spanish Republic who arrived in Mexico around 1939. José Gaos, Joaquín Xirau, Juan David García Bacca, José María Gallegos Rocafull, Eugenio Imaz, Eduardo Nicol and Wenceslao Roces remain exemplars of Mexican philosophical life in the late twentieth century. Philosophers of this time are disciples of this aforementioned group who were responsible for defining the philosophical task in Mexico. Each of them has reached international levels of excellence in debate, production, philosophical rigour, argumentation, technique and information.

2 The situation in the 1990s

The situation in the 1990s is not easy to delineate. It cannot be reduced to what has been produced since 1990, nor to what was produced after either the linguistic change, or the final works of the Hyperion group. Focusing on the second half of the twentieth century, it is possible to establish the state of debate and philosophical production in Mexico in the late 1990s.

The careful study of European thought, German in particular, resulted in the development of phenomenology and existentialism in Latin America. Both these schools of thought dominated the principal centre of reflection of the country in the 1950s, which was the department of philosophy and letters at the National Autonomous University of Mexico (UNAM) (see Existentialist thought in Latin America; Phenomenology in Latin America). José Gaos made the very first translation into Spanish of *Being and Time* (1943). In that environment, the efforts began of the Hyperion group, whose name was taken from J.C.F. Hölderlin to signify the joining of heaven and earth. With the tools of phenomenology and existentialism, this group attempted to cast philosophical light on the essence of being Mexican. In an effort lasting around three years, they challenged the task of philosophy in Mexico and internationally, although they did not achieve their intended goals. Emilio Uranga (1952) published his classic analysis of the essence of the Mexican, but it was soon apparent that the achievements of this ontological route were meagre. However, a historicist trend developed and has been productive since the 1950s.

The thought of Leopoldo Zea understands the essence of the Mexican and the Latin American as a historical being with a historically situated consciousness. It is feasible to study the development of that historical consciousness expressed in the history of ideas in the region. From the beginning of that labour of reconstruction derive several modes of philosophizing. In the case of Zea, that history of ideas involved a philosophy of Latin American history which would place the being, destiny and meaning of the history of Mexico and Latin America in the context of the history of the world. The 1940s and 1950s were unusually productive to this end. In the 1970s philosophical positions were well defined, consolidated and manifested in compact groupings. Each one had its own problems, traditions, teaching figures, leaders and spheres of influence.

Leopoldo Zea led the group which focused on Latin America. He emphasized the study of the position of Latin America in the world situation and advanced the history of ideas in what José Martí called ‘Our America’. The antecedents of these teachings, whose points of focus were Latin American identity incorporating a wide cultural focus, were Caso, Vasconcelos, Reyes, Henríquez Ureña, Ramos and Gaos (see Cultural identity).

Marxist reflections were headed by Eli de Gortari and Adolfo Sánchez Vázquez (see Marxist thought in Latin America). De Gortari advanced the study of dialectical logic and the relations between philosophy and the hard

sciences through his seminar on history and the philosophy of science. He also initiated the study of the history of science in Mexico. Sánchez Vázquez formulated a solid argumentation that understood Marxism principally as a philosophy of praxis and a tool for critical aesthetic reflection. Both philosophers employ up-to-date information and argumentative rigour and both have become engaged in the international Marxist debate with a critical, suggestive and stimulating slant. An explicit focus on both the Mexican and Latin American Marxist traditions would highlight the arguments of Lombardo Toledano and José Revueltas.

A third focus has been analytic thought characterized by the linguistic trend (see Analytical philosophy in Latin America). Advanced by Luis Villoro, Fernando Salmerón and Alejandro Rossi, this group incorporated into academic debate the Anglo-Saxon tradition and the classic problems of philosophy of language, mind and science. Strongly professionalized, this group aspires radically to modify philosophical tradition in the region, incorporating modalities and styles of philosophical production in common with Anglo-Saxon countries. Their centre of study is the Institute of Philosophical Research at the National Autonomous University of Mexico. They do not refer to antecedents in the region and maintain a scientific conception of philosophy.

A fourth sector is metaphysical thought, which focuses on ancient philosophy in the tradition of Eduardo Nicol and his principal disciple Juliana González. A preoccupation with the philological and an emphasis on genetic studies in an articulation of aesthetics, ethics and gnoseology mark their thought patterns.

This cartographic outline seemed immutable in those years. Nevertheless, by the end of the 1970s modification of the agenda in Mexico had begun. New generations developed new themes; the trauma of 1968 with its consequences on academic life, along with new social demands and changes in the world began to split this four-part schema, although derivations of their philosophies are still in evidence. The critical reception of L.P. Althusser, political philosophy, epistemology, history, ethics and aesthetics were pushed to the foreground. During the 1980s decisive transformations were produced in contemporary Mexican life and changes at world level affected philosophical reflection which, with few exceptions, was maintained in cloistered areas from an endogenous academic viewpoint. Debates regarding structuralism, Marxism, semiotics and discourse theory; the production of knowledge, liberation, modernism; deconstruction and postmodernism tangentially traversed traditional boundaries which were previously impenetrable.

The philosophical positions developed throughout the 1970s and 1980s have not become outmoded, although they have crossfertilized to a degree and the forum for argumentation and frank debate has expanded. Philosophical activity has expanded outside of the institutions in which it was once confined, namely the department of philosophy and letters and the Institute of Philosophical Research at the UNAM. High-quality work has been carried out at the Autonomous Metropolitan University, especially the Iztapalapa and Xochimilco campuses, at the Jesuit Iberoamerican University, the Intercontinental University and the universities of Chihuahua, Toluca, Guadalajara, Xalapa and Ciudad Juárez.

The topics and areas of study have become richer and more varied. The history of ideas has been pursued with increased interest. There are as many works about Pre-Columbian thought in the Nahuatl field as in the Mayan, as well as in the area of recent indigenous thought, particularly that of the Huicholes. The colonial period is under exploration in an effort to refute its dismissal by positivist historiography. The development of philosophy in the nineteenth and twentieth centuries is under discussion. Monographic studies emphasize access to and retrieval of source material. All this historiographic work has brought advances in the debate over methodological and epistemological questions.

In the 1990s the precarious state of the history of science and technology in Mexico changed markedly. Rescue of sources, new methodological foci and the correlation of local with world history have facilitated work on the relationship between the history of science and ideas. In the area of the philosophy of science, work on the theory of argumentation, scientific language, models of justification and the links between hard sciences and social sciences have been produced. In political philosophy, work has been undertaken related to the idea of democracy, its processes of democratization, its manifestations of power, civil-state relations, mass communications and their prevalence in everyday life, national identity and ethnic and/or regional identities. Feminist thought has grown and developed widely in Mexico. Discussions about gender, ways of knowing and the role of women in the history of philosophy are of particular interest. Also, studies in the philosophy of law, human rights, minority rights and critical criminology have been produced. In the area of ethics, studies have come about stimulated by the

deterioration of living conditions for the vast majority and the need to clarify the rules of social life. Debates regarding historical knowledge, historiographic methodology, the role and meaning of Mexico in the world, the discourses that make up the course of history, have arisen within the sphere of the philosophy of history. There has been an important effort to establish a hermeneutics of Mexican and Latin American culture which surpasses the traditional framework of philosophy of culture and incorporates the study of popular traditions.

Among the established philosophical journals in this area are *Cuadernos Americanos*, *Cr ítica*, *Dialéctica* and *Estudios*. The most active organizations in the promotion of philosophical activities have been the Asociación Filosófica de México (Mexican Philosophical Association), active for twenty-five years and counts among its members the most important and productive philosophers of the country, the Asociación Filosófica Feminista de México (Mexican Feminist Philosophical Association) and the Círculo Mexicano de Profesores de Filosofía (Mexican Circle of Professors of Philosophy).

There have also been numerous academic events in the Mexican philosophical agenda, such as the congresses of the Mexican Philosophical Association, the Mexican Circle of Professors of Philosophy, the Mexican Society for the History of Science and Technology and those dedicated to the study of philosophy in New Spain. The ninth Interamerican Congress of Philosophy took place in Guadalajara in 1985 and the first International Congress of Latin American Philosophy took place in Ciudad Juárez in 1991.

It is worthy of note that philosophy is still taught at the secondary level in Mexico. There is considerable interest in improving the quality of the teaching of philosophy to prepare students for life and future university studies.

3 Prospects

Despite the financial crisis and the budget restrictions that chronically affect Mexican academic institutions, the future of the philosophical agenda in Mexico is promising. Young people with excellent training are slowly becoming involved in philosophical production and other institutional frameworks are surfacing to support philosophy such as the professional associations. The growing social and cultural problems demand inclusion in the philosophical agenda. With a population subject to inhumane deficiencies and with the increase in poverty in real and absolute terms, philosophy is called upon to attempt adequate responses. Also, the process of democratization in the country is demanding philosophical contributions. Cultural creativity continues at its highest level and demands adequate aesthetic reflections which address the phenomenon. Globalization demands new conceptualizations which help address intellectual perplexity.

See also: Argentina, philosophy in; Brazil, philosophy in; Latin America, colonial thought in; Latin America, Pre-Columbian and indigenous thought in

HORACIO CERUTTI-GULDBERG

References and further reading

Bartra, R. (1987) *La jaula de la melancolías: identidad y metamorfosis del mexicano (The Cage of Melancholy: Identity and Metamorphosis of the Mexican)*, Mexico: Grijalbo.(The changing situation of the Mexican is discussed.)

Beuchot, Mauricio, José Rubén Sanabria (compilers) (1994) *Historia de la filosofía cristiana en México*, Mexico, Universidad Iberoamericana.

Bibliografía filosófica mexicana 1991 (1994-), Mexico, National Autonomous University of Mexico Instituto de Investigaciones Filosófica, 2a época, vol. 6,

Cerutti Guldberg, H. (1986) *Hacia una metodología de la historia de las ideas (filosóficas) en América Latina (Towards a Methodology of the History of (Philosophical) Ideas in Latin America)*, Guadalajara: University of Guadalajara.(A discussion of the methodological aspects of the history of philosophy in Mexico and proposals predicting the future of the philosophical task in the country.)

Escobar-Valenzuela, G. (1992) *Introducción al pensamiento filosófico en México (Introduction to Philosophical Thought in Mexico)*, Mexico: Limusa.(A general study of philosophical thought in Mexico.)

Gómez-Martínez, J.L. (ed.) (1993) *Anuario bibliográfico 1990 de historia del pensamiento ibero e iberoamericano*, Athens, Georgia, Series on Hispanic Thought, num. 5.

Léon de Portilla, M. (ed.) (1524) ‘Coloquio de los Doce’ (Debate of the Twelve), in *Coloquios y doctrina Cristiana (Debates and Christian Doctrine)*, Mexico: National Autonomous University of Mexico and Fundación de Investigaciones Sociales, 1986.(An account of the twelve monks from San Francisco sent by

Pope Adriano VI and the Emperor Carlos V to convert the Indians of New Spain. Includes the dialogues of Fray Bernardino de Sahagún and his collaborators Antonio Valeriano de Azcapotzalco, Alonso Vegerano de Cuahutitlán, Mart ín Jacobita and Andrés Leonardo de Tletelolco.)

Martí, O.R. (1989) ‘Mexican Philosophy in the 1980s: Possibilities and Limits’, in *Varios: Philosophy and Literature in Latin America. A Critical Assessment of the Current Situation*, Albany, NY: State University of New York Press, 36-63.(A good overview of Mexican philosophical production.)

Morales, C. (1990) ‘La investigación filosófica en la UNAM’ (Philosophical Investigation at the National Autonomous University of Mexico), in *Varios: Universidad Nacional y Cultura*, Mexico: National Autonomous University of Mexico and Porrúa, 155-78.(An examination of the development of philosophy at the most important university in Mexico after 1968: a key year in the political and intellectual life of the country.)

Osorio, I. (1989) *Conquistar el eco: la paradoja de la conciencia criolla (Conquering the Echoes: the Paradox of the Creole Consciousness)*, Mexico: National Autonomous University of Mexico.(Unites important works about the development of philosophy in New Spain. Its bibliographical references and sources are especially useful.)

Rovira-Gaspar, C. (coordinator) (1993) *Bibliografía mexicana, filosofía y polémica. Primera mitad del siglo XIX*, Mexico, UNAM.

Saldaña, J.J. (1992) *Los Orígenes de la ciencia nacional México*, Sociedad Latinoamericano de Historia de las Ciencias y la Tecnología.

Trabuise, E. (1983) *Historia de la ciencia en México. Estudios y textos*, Mexico, CONACYT, Fondo de Cultura Económica, vol. I, Siglo XVI, 459 pages. Vol. II, *El claroscuro de la ciencia mexicana del siglo barroco*, 298 pages. Vol. III, *La ciencia mexicana en el siglo de las luces,* 513 pages. Vol. IV, *La ciencia mexicana del período nacional*, Vol. V, Apéndices e indices, 591 pages.

Uranga, E. (1952) *Análisis del ser mexicano (Analysis of Being Mexican)*, Mexico: Porrúa y Obregón.(A study of what it is to be a Mexican.)

Vargas-Lozano, G. (1980) ‘Notas sobre la función actual de la filosof ía en México (la década de los setenta)’ (Notes on the Current Trends of Philosophy in Mexico (in the 1970s)), in *Dialéctica* 9 (December): 81-108. (Still relevant, this study decribes the state of philosophy in Mexico in the 1970s.)

Villanueva, E. (1984) ’Philosophical Analysis in Mexico’, in: Varios: *Philosophical Analysis in Latin America Dordrecht*, D. Reidel, pp. 165-171.

Villegas, A. (1993) *El pensamiento mexicano en el siglo XX (Mexican Thought in the Twentieth Century)*, Mexico: Fondo de Cultura Económica.(A comprehensive view of the development of philosophy in Mexico in the twentieth century.)

Villoro, L. (1972) ’Perspectivas de la filosofía en México para 1980’ in: Varios: *El perfil de México* en 1980, Mexico, Siglo XXI.

Meyerson, Émile (1859-1933)

Meyerson rejects the positivism of Comte and Mach, insisting that reason demands more of science than the identification of lawful regularities for the purpose of successful prediction. Reason demands an actual representation of reality based on genuine causal explanation, and causal explanation consists of demonstrating an underlying identity behind the apparent diversity of phenomena. If completely successful, however, such explanation would ultimately lead to the unchanging homogeneous and sterile universe of Parmenides. That it does not is due to what Meyerson calls 'irrationals': inexplicable elements in reality that can never be reduced to identity. Explanation, therefore, is never complete, and scientific laws are 'plausible' rather than universally and necessarily true.

Émile Meyerson was Polish by birth but French by choice. In 1882, after studies in Germany under the celebrated chemist R.W. Bunsen, he began his career in the inorganic chemistry laboratory at the Collège de France. Abandoning chemistry in 1889, he served as a foreign policy editor for the Havas News Agency, then turned to philanthropic work with Jewish emigrants to Palestine, eventually becoming director for Europe and Asia Minor of the Jewish Colonization Association.

Meyerson's erudition was legendary. Although he never held an academic post, he was enthusiastically accepted by the French intellectual community, and his office was a gathering point for eminent philosophers and scientists. When he wrote about the history of science, he had usually read the primary material in the original language. When he wrote about contemporary science, he had often discussed it personally with the scientists who created it.

Although often classified as an epistemologist and/or philosopher of science, Meyerson himself describes his work more broadly as 'philosophy of the intellect' (from 'Philosophie de la nature et philosophie de l'intellect' in 1936), which he differentiates both from logic, because he is concerned with how we actually think rather than how we ought to think, and from psychology, because he is concerned not with direct introspection, but with analysis of the products of thought in order to discover a posteriori the a priori component of our thought processes. On the basis of exhaustive concrete examples, especially from the history of science, but also from the history of philosophy and ordinary common sense reasoning, he identifies an a priori tendency in all reasoning to replace diversity in appearances with identity in reality.

To explain (*expliquer*) an event is to explicate (literally, to unfold) it from its cause to show that it already existed in the cause, and is in this sense identical with it (1921: 9) (see Causation; Explanation). In scientific reasoning, the example *par excellence* of this identity is conservation laws, which maintain that throughout changing appearances the underlying amount of mass, energy, inertia, and so on, remains constant, that is, identical (Zahar 1980) (see Conservation principles). Such laws are both analytic and synthetic: analytic in so far as they exemplify the innate tendency to establish real identity behind apparent diversity, but synthetic in so far as what specifically is posited as identical, for example, momentum, is determined empirically. Meyerson classifies such laws as 'plausible' (1921: 410). This word is to be understood in a strong sense, for, because of their a priori basis, laws have an element of true necessity. But, because of their empirical content, this plausibility falls short of strict necessity.

Other scientific examples abound, such as the persistence of the same chemical elements in the raw material for a chemical reaction and its product, or the permanence of physical atoms throughout physical change, which allows us to explain the diversity of secondary qualities in terms of changes in position in homogeneous space of unchanging material particles. But the a priori tendency toward identity is not confined to formal science. Common sense also seeks to replace the diversity of experience with an underlying permanent reality. For example, we posit the existence of a real table that remains identical in spite of our changing, diverse sensations of the table, and we form the concept of table by ignoring the differences among individual tables and identifying a common (identical) essence.

The identification is never complete, however, whether in scientific or common sense reasoning; at some point we are inevitably confronted with brute diversity which cannot be reasoned away, that is, with one or more

'irrationals' (1908). Perceptual diversity remains even as we reduce the table to its defining geometrical elements, and salt clearly differs from sodium plus chlorine even after we have explained that it is simply sodium chloride. For Meyerson the example *par excellence* of such irrationals is the second law of thermodynamics, and Carnot has the place of honour in Meyerson's scientific pantheon. If there actually were a strict identity between cause and effect, then all physical processes would be reversible. But entropy necessarily increases; there are irreversible processes; identity of cause and effect is not complete (see Thermodynamics). Quantum mechanics, with its indeterminacy principle, presents us with a startling new example of an intractable irrational in physical explanation (see Heisenberg, W.; Quantum mechanics, interpretation of §2), but classical mechanics provides examples as well. Even if the classical mechanist were to succeed in reducing all causal explanation to the physical impact of elementary particles, for example, the causal mechanism of the impact itself - how force or momentum is passed from one particle to another - would continue to defy analysis.

Although a scientific realist, Meyerson insists that his concern is not with metaphysics but only with the philosophy of the intellect. There must be a partial correspondence between our rational theories of reality and reality itself or the success of scientific inquiry could not be explained, but it is the a priori tendencies in our thought processes rather than the nature of physical reality in which he is interested. The correspondence between the structure of our thought and the structure of reality, though necessary, is not complete. Reality is also recalcitrant to our reasoning processes. We are inevitably confronted with irrationals, and there is no way of predicting when they will arise or in what form. We face an epistemological paradox (1921): reason demands identity, but even a Cartesian or Einsteinian world in which matter is reduced to the geometry of space fails to acheive complete identity; nothing less than the sterility of absolute Parmenidean homogeneity will do. Our reasoning is successful only in so far as it does establish identity, but its explanatory power depends also upon its ultimate failure. We succeed in explaining diversity by finding underlying identities while admitting that, nevertheless, there really is change and diversity, which, in the end, we cannot fully explain. Genuine explanation is possible only because the irrational cannot be completely eliminated.

DAVID A. SIPFLE

List of works

Meyerson, É. (1908) *Identité et réalité*, Paris: Alcan; 5th edn, Paris: Vrin, 1951; trans. K. Loewenberg, *Identity and Reality*, New York: Macmillan, 1930.(Meyerson's classic, most concise statement of his thesis - the only book by which Meyerson was known to English readers for over fifty years.)

Meyerson, É. (1921) *De l'explication dans les sciences*, Paris: Payot, 2 vols; 2nd edn in 1 vol., 1927; trans. M.-A. and D.A. Sipfle, *Explanation in the Sciences*, Boston Studies in the Philosophy of Science, vol. 128, Dordrecht: Kluwer, 1991.(A more extensive argument for Meyerson's thesis, with a wealth of examples from current and past science and extensive treatment of Hegel's and Schelling's philosophies of nature.)

Meyerson, É. (1925) *La déduction relativiste*, Paris: Payot; trans. D.A. and M.-A. Sipfle, *The Relativistic Deduction*, Boston Studies in the Philosophy of Science, vol. 83, Dordrecht: Reidel, 1985.(Argues that Einsteinian relativity conforms to Meyerson's theory just as clearly as does classical physics - indeed that, given Meyerson's theory, a theory such as Einstein's is just what one would expect to emerge. A favourable review by Einstein and a fine introduction by MiličČapek are included with the English translation.)

Meyerson, É. (1931) *Du Cheminement de la pensée (The Ways of Thought)*, Paris: Alcan, 3 vols.(Two volumes of text and one of notes. A *tour de force* in which Meyerson extends his thesis to all realms of human thought, including common sense thought and prehistoric thought no less than logical, mathematical, and scientific thought.)

Meyerson, É. (1933) Reél et déterminisme dans la physique quantique (Reality and Determinism in Quantum Physics), *Actualités scientifiques et industrielles, vol. 68*, Paris: Hermann.(Argues for the need for a realistic, but not necessarily deterministic interpretation of quantum mechanics; preface by Louis de Broglie.)

Meyerson, É. (1936) *Essais (Essays)*, Paris: Vrin.(A wide range of journal articles published throughout Meyerson's career and after his death. This volume includes some of Meyerson's most explicit and self-conscious attempts to explain his purpose and his methodology.)

References and further reading

Blumberg, A. (1932) 'Émile Meyerson's Critique of Positivism', *The Monist* 42: 60-79.(Argues that Meyerson's critique of positivism, directed against Comte and Mach, fails when directed against 'the new "logical

positivism" of Wittgenstein, Schlick, and Carnap, developed posterior to and independently of the writings of Meyerson'.)

Boas, G. (1930) *A Critical Analysis of the Philosophy of Émile Meyerson*, Baltimore, MD: Johns Hopkins University Press.(A critical exposition of Meyerson's first three books, written to call attention to a 'thinker, whose importance seemed underestimated in America'.)

Brunschvicg, L. (1926) 'La Philosophie d'Émile Meyerson', *Revue de métaphysique et de morale* 33: 39-63.(A critical study of Meyerson's philosophy, based on his first three books, on the occasion of the publication of *La déduction relativiste*.)

Bryson, P. (1973) 'The Metaphysics of Émile Meyerson: A Key to the Epistemological Paradox', *The Thomist* 37: 119-32.(A critical exposition of Meyerson's epistemological paradox, concluding that Meyerson's account of the explanatory structure of reason is unacceptable because it cannot resolve the paradox.)

Bulletin de la Société française de philosophie (April 1961) 55th year, no. 2, reports the 26 November 1960 meeting devoted to the centenary of Meyerson's birth (and of Gaston Milhaud's); contributions by André Lalande, Louis de Broglie, René Poirier, André Metz and Alexandre Koyré.(Includes biographical details as well as philosophical appraisals.)

Kelly, T. (1937) *Explanation and Reality in the Philosophy of Émile Meyerson*, Princeton, NJ: Princeton University Press.(Argues that the divergence between explanation and reality is so great in Meyerson's earlier works that it should lead to scepticism, but that this problem is alleviated in *Du Cheminement de la pensée*. Includes a comprehensive bibliography.)

Lalande, A. (1922) 'L'Épistémologie de M. Meyerson et sa portée philosophique', *Revue philosophique de la France et de l'étranger* 93: 259-80.(A critical study of Meyerson's philosophy, based on his first two books, on the occasion of the publication of *De l'explication dans les sciences*. §§1-4 are the most valuable.)

Largeault, J. (1992) 'Émile Meyerson, philosophe oublié', *Revue philosophique de la France et de l'étranger* 182: 273-95.(A comprehensive exposition and appreciation of Meyerson's philosophy, looking back long after his influence had waned, nearly sixty years after his death.)

LaLumia, J. (1966) *The Ways of Reason: A Critical Study of the Ideas of Émile Meyerson*, London: Allen & Unwin.(Argues that Meyerson is not doing epistemology or logic, but psychology. Includes a bibliography.)

Loewenberg, J. (1932) 'Meyerson's Critique of Pure Reason', *Philosophical Review* 41: 351-67.(An insightful interpretation of Meyerson's position in Kantian terms. Meyerson expresses his approval in 'Philosophie de la nature et philosophie de l'intellect', *Essais*, 105n.)

Metz, A. (1934) *Meyerson: une nouvelle philosophie de la connaissance*, Paris: Alcan, 2nd edn.(A comprehensive exposition of Meyerson's philosophy by one who knows and admires the man and his work, written for the general reader.)

Zahar, E. (1980) 'Einstein, Meyerson and the Role of Mathematics in Physical Discovery', *British Journal of the Philosophy of Science* 31: 1-43.(§2 briefly explains Meyerson's principle of identity and its application to conservation laws, to the Cartesian reduction of matter to space, and to Einstein's special and general theories of relativity.)

Mi bskyod rdo rje (1507-54)

Mi bskyod rdo rje (Mikyö Dorje) was a Tibetan Buddhist of the Karma bKa'-brgyud (Gagyü) school. Particularly in his earlier thought, he defended the existence of a positive nondual Ultimate Reality (an Absolute) beyond all conceptuality but known in meditation, and apparently understood the purpose of Mādhyamika philosophy as solely the shattering of conceptualization and all philosophical positions as a preparation for this nonconceptual meditation. He was highly critical of Tsong kha pa (Dzongkaba, a notable philosopher of the rival dGe-lugs-pa (Gelukba) school), for being overly concerned with logical analysis of the foundations of everyday experience.

Mi bskyod rdo rje (Mikyö Dorje) was the Eighth Karma pa hierarch of the Karma bKa'-brgyud (Gagyü) school of Tibetan Buddhism. He was born in Damchu (East Tibet) and died of leprosy at Dakpo Shedrup Ling Monastery (East-Central Tibet). The bKa'-brgyud tradition inherited from its Indo-Tibetan yogic origins a dislike of philosophy and logic, which was seen as an over-concern with conceptuality. Emphasis was placed on direct experiential knowing, associated with a meditation system known as *mahāmudrā* ('the Great Seal'). *Mahāmudrā* is said to involve the immediate experience of one's own mind, which is actually the primordial self-luminous empty mirror-like wisdom-Mind of Buddhahood. All things are the spontaneous unobstructed play of this fundamental Mind, which is held to be the primordial Reality beyond all conceptuality and linguistic distortion. The direction of bKa'-brgyud thought is primarily towards moulding a sufficient theoretical framework for this *mahāmudrā* meditation. Nevertheless, Mi bskyod rdo rje devoted impressive intellectual effort to establishing the philosophical position of his school, furnishing the Karma bKa'-brgyud tradition with its own commentaries on some of the principal works used in the Indo-Tibetan monastic curriculum.

At the time of his full ordination as a monk, Mi bskyod rdo rje was exhorted to promulgate 'the view of other-empty' (*gzhan stong*) (see Tibetan philosophy §1). This involves the existence of a true Ultimate Reality, empty of what is other than it in that it is essentially empty of whatever is non-Ultimate, but nevertheless an Absolute, an Ultimate Reality having itself all the intrinsic qualities of Buddhahood. This 'other-empty' position contrasts with a view held within Tibetan Buddhism notably by the dGe-lugs-pa (Geluk) tradition that all things without exception are empty of their own inherent existence (*rang stong*) and exist only relatively. There is no Ultimate Reality at all, as an actual Absolute. There is some suggestion that even at the early age of 23 Mi bskyod rdo rje wished to adopt an other-empty approach in a combative commentary he wrote on the Perfection of Wisdom (*Prajñāpāramitā*) scriptures. At about the same time he also wrote a short independent work on the other-empty perspective. In it he speaks of the other-empty view as that of the Great Madhyamaka. He traces it to Maitreya, Asaōga and Vasubandhu, but he wants to distinguish this Great Madhyamaka view from the 'Mind-only' (*cittamātra*) vision often associated with those figures, which for Mi bskyod rdo rje seeks to establish the flow of relative, tainted ('everyday') consciousness as the Reality (see Buddhism, Yogācāra school of §§1-4, 8). Mi bskyod rdo rje sees the Great Madhyamaka as also the final perspective of Nāgārjuna. He is, however, very critical of certain 'pseudo-Mādhyamikas' (including apparently the renowned scholars Candrakīrti and Haribhadra) for whom there is no Ultimate Reality, and the ultimate truth is itself simply an emptiness of inherent existence (that is, the *rang stong* perspective). Such an emptiness is a mere negation, and could not be the Ultimate referred to in Mahāyāna Buddhist sources, an Ultimate which, as nonconceptual, is realized in direct yogic experience and is not accessible to the negations of logicians like Candrakīrti. The rival view that would make the ultimate *qua* Ultimate Reality nonexistent is a form of nihilism. By way of contrast with the Ultimate Reality, for Mi bskyod rdo rje in reality all conventional, relative, mistaken appearances simply do not exist, and are not known even by the Buddha's omniscience.

However, In Mi bskyod rdo rje's commentary on Candrakīrti's *Madhyamakāvatāra (Supplement to the Middle Way)*, written many years later, he criticizes a number of important Tibetan figures associated with the other-empty tradition, and interprets Candrakīrti's text in a 'self-empty' (*rang stong*) manner. It has been suggested (Ruegg 1988) that this could be seen as Mi bskyod rdo rje's response to dGe-lugs criticisms. On the other hand, since he makes no mention in interpreting the *Madhyamakāvatāra* of any major other-emptiness lineage of interpretation, and given his criticisms of Candrakīrti in his previous work, it seems possible that Mi bskyod rdo rje wished to show that Candrakīrti (or conceptualized Madhyamaka) could only be interpreted in a self-empty and not an other-empty manner. In this commentary Mi bskyod rdo rje shows his impatience with those

who see Candrakīrti as a system-builder. Candrakīrti’s intention is simply absence of delusion, through demolishing all conceptual philosophical systems. We are left with nothing as far as conceptual philosophizing goes, and even terms like ‘emptiness’ are simply conceptual devices to be used pedogogically. Things having thus been cleared away, Mi bskyod rdo rje holds that we should now engage in direct yogic experience through *mahāmudrā*. Although even someone who follows Madhyamaka will continue with the pre-reflective activities of the ‘person in the street’, there can be no satisfactory systematically coherent and rational philosophical foundation for those activities. Mi bskyod rdo rje is particularly critical of Tsong kha pa Blo bzang grags pa (Dzongkaba Losang dragba, 1357-1419) and the dGe-lugs tradition here. Tsong kha pa’s view, as represented by Mi bskyod rdo rje, is that emptiness is the essence of things, *qua* their lack of their own inherent existence. Thus there is a level on which things are coherently founded, in order for them to be empty. Nevertheless *all* things are empty, including emptiness itself. So Tsong kha pa thinks that things can be rationally *shown* to have *an* existential status, although there is no Ultimate Reality at all. For Mi bskyod rdo rje, in the face of any critical analytic investigation nothing can be founded, and it is unseemly to be so concerned with founding the relative and mundane conventional. Also, he says, Tsong kha pa’s emptiness is itself just a ‘limited emptiness’ (*itaretaraśūnyatā*), the relative, mundane emptiness that we see in the everyday context of two different things where ‘*x* is empty of *y*’. Tsong kha pa’s emptiness could not be the Ultimate, the emptiness which could serve as a true basis for the path to liberation.

See also: Buddhism, Mādhyamika: India and Tibet; Buddhist concept of emptiness

PAUL WILLIAMS

List of works

Many of Mi bskyod rdo rje’s writings are at present unavailable. The following unedited texts of philosophical interest have been published by the Tibetan refugee community. None has been translated.

Mi bskyod rdo rje (*c.*1530) *dBu ma gzhan stong smra ba’i srol legs par phye ba’i sgron me (The Lamp Discriminating Properly the System of the ‘Other-Empty’ Mādhyamikas)*, Rumtek: probably Dharmachakra Centre, 1972.(An early work of Mi bskyod rdo rje on the ‘other-empty’ viewpoint.)

Mi bskyod rdo rje (1532-43) *Chos mngon pa’i mdzod kyi ’grel pa rgyas par spros pa grub bde’i dpyid ’jo (A Spring-yield of Accomplishment and Happiness: The Commentary to the Abhidharmakośa Extensively Elaborated)*, New Delhi: T. Tsepal Taikhang, 1975, 2 vols.(Mi bskyod rdo rje’s Commentary on Vasubandhu’s *Abhidharmakośa*.)

Mi bskyod rdo rje (begun 1545) *dBu ma la ’jug pa’i rnam bshad dPal ldan dus gsum mkhyen pa’i zhal lung dvags brgyud grub pa’i shing rta (A Chariot for the Accomplished of the Lineage of Dvags [the bKa’-brgyud], the Oral instructions of Glorious Dus-gsum-mkhyen-pa [the First Karma-pa]: A Commentary to the Madhyamakāvatāra)*, Rumtek: Dharmachakra Centre.(Mi bskyod rdo rje’s Commentary on Candrakīrti’s *Madhyamakāvatāra*.)

References and further reading

Broido, M. (1985) ‘Padma dKar po on the two satyas’, *Journal of the International Association of Buddhist Studies* 8 (2): 7-59.(Mi bskyod rdo rje discussed within the context of the dGe-lugs and bKa’-brgyud treatment of the two truths and Vajrayāna. Fairly technical.)

Douglas, N. and White, M. (1976) *Karmapa: The Black Hat Lama of Tibet*, London: Luzac.(Short biographies based on traditional sources.)

Richardson, H. (1958) ‘The Karma-pa sect: A Historical Note’, *Journal of the Royal Asiatic Society* 3-4: 139-64; documents in J.R.A.S. 1-2: 1-17, 1959.(Historical discussion of the Karma bKa’-brgyud school. A standard work.)

Ruegg, D.S. (1988) ‘A Karma bKa’ brgyud Work on the Lineages and Traditions of the Indo-Tibetan dBu ma (Madhyamaka)’, in G. Gnoli and L. Lanciotti (eds) *Orientalia Iosephi Tucci Memoriae Dictata*, Rome: Istituto Italiano per il Medio ed Estremo Oriente, 1249-80.(Through treating the list of lineages contained at the beginning of the *Madhyamakāvatāra* commentary, a leading authority pictures how Mi bskyod rdo rje saw Madhyamaka doctrine and his place within it. Fairly technical at times.)

Thinley, K. (1980) *The History of the Sixteen Karmapas of Tibet*, ed. D. Stott, Boulder, CO: PrajñāPress.(Short biographies based on traditional sources.)

Williams, P. (1983) ‘A note on some aspects of Mi bskyod rdo rje’s critique of dGe lugs pa Madhyamaka’, *Journal of Indian Philosophy* 11: 125-45.(Mi bskyod rdo rje’s criticisms of the dGe-lugs view of emptiness.

Not particularly technical.)

Midrash

Midrash, a Hebrew word meaning 'investigation' or 'study', denotes both the method used by the Jewish rabbis of the second to sixth centuries AD *to interpret the Bible and the extensive literature that resulted from the application of that method. In rabbinic parlance midrash, or the related term derash, can also designate a homiletic, non-literal way of reading the Bible. Midrash embodies a distinctive hermeneutic which at its most extreme treats the text of Scripture as a set of symbols or signs apparently to be manipulated by the interpreter at will. In recent years midrash has been compared to reader-response literary criticism. It has also been claimed that it represents a 'Judaic' as opposed to a 'Hellenic' mode of thinking which anticipates postmodernist hermeneutics.*

1 Midrash as literature

The extensive midrashic literature falls into two broad types: exegetical midrashim, which comment more or less verse by verse on a biblical book (for example, *Bereshit Rabba* on Genesis), and homiletic midrashim which comment on the Scriptural readings for a series of festivals or special days of the liturgical year (for example, *Pesiqta de-Rav Kahana*). These texts are constructed out of small units of discourse which employ a limited number of highly patterned literary forms. These basic forms are organized into higher forms, larger units of discourse which make a single theological, homiletic, or legal point. The higher forms are then collected into documents or composite forms which usually have some sort of overarching programme. Thus the basic forms of a midrash (in the sense of a single exegetical operation), a *mashal* (parable) and a dictum may be combined into the higher form of a homily, and a number of homilies may in turn be combined into a composite form such the document known as *Pesiqta de-Rav Kahana*. The various documents are closely related to each other and share material. Later stages of the tradition often seem to be developing or reacting to earlier ones. The contents of the documents differ from manuscript to manuscript; and they are not authored texts in any strong sense of that term, but neither are they random collections. They are an evolving literary canon which holds the deposit of the teaching and preaching on the Bible in the rabbinic academies and the synagogues of Palestine in the second to sixth centuries AD.

2 The hermeneutic methods of midrash

In midrash, the Bible is seen as a unique text which originated in the mind of God. The contribution of its human authors is discounted. They wrote under the influence of a divine power ('the holy spirit') which neutralized their human failings and ensured that they transmitted accurately the divine message. The text is fixed, unalterable and authoritative for all time. It is completely consistent and can contain no errors of fact, nor any contradictions. When seeming contradictions appear they are harmonized. The possibility that earlier laws may have been abrogated by later laws is not admitted. All truth that is relevant to human knowledge and action is latent in Scripture and can be discovered therein if the text is investigated in the proper way. The potential of Scripture is maximized. Its text is polyvalent and can have a multiplicity of meanings. It is not seen as a problem if these contradict each other, provided they are properly derived, since it is assumed that the contradiction is resolved in the mind of God. The text of Scripture contains no redundancy. If the same statement is found in two places its meaning will be differentiated: for example, in one place it will be referred to the present age, in the other to 'the age to come' (that is, the Messianic age).

Seen from outside the tradition, the function of midrash is twofold: (a) to apply the fixed text of Scripture to changing historical circumstances; and (b) to validate post-biblical practices, institutions, ideas and laws from the Bible. From within the tradition this second function is rarely openly acknowledged. It is normally assumed that the exegete is drawing out meaning objectively contained within Scripture and not subjectively reading meaning into it. Various techniques are used to explicate or extract meaning from Scripture, some of them manipulative in the extreme. For example, in the device known as *gematria* meaning can be derived by computing the numerical value of the letters. Several lists of the hermeneutical rules (known as Middot) used in midrash are extant. But these listings by no means exhaust the techniques of midrash. Despite the appearances, midrash is a disciplined activity: there must be some objective 'peg' in the biblical text (for example, a lacuna in the narrative, or an unusual linguistic usage) on which to hang the interpretation.

The manipulative approach to the reading of Scripture created an obvious problem of authority, since it would also

allow heretical (for example, Christian) views to be validated from Scripture. This problem was addressed in a number of ways. It was asserted that, whatever midrashic meanings were derived from Scripture, the text could never lose its simple or literal (*peshat*) sense. It was also claimed that the correct interpretation of Scripture could be given only by those who had studied in the rabbinic academies with approved teachers who stood in an unbroken tradition of teaching going back to Moses. The tradition of commentary was elevated to the status of Scripture. It was designated oral Torah in contrast to the written Torah (the Scriptures), and its origins were traced back to Sinai, the supreme moment of revelation when the written Torah was given to Israel.

3 Midrash as a 'Judaic' mode of thinking

The term 'midrash' has passed into contemporary academic discourse, because certain literary critics have suggested that midrash anticipates postmodernist ways of reading texts. They note with approval that midrash often behaves with remarkable freedom towards the text, offering a strong reading, which may at times stand the plain sense on its head. They compare the midrashic postulate that Scripture has a multiplicity of meanings to the postmodern claim that a text does not have a single absolute and objective sense, that its meaning varies from reader to reader and that no one reading can be privileged as more correct than another. Some go further and suggest that midrash embodies a distinctively 'Judaic' mode of thinking which is semiological, auditory and algebraic, in contrast to the ontological, visual and geometrical 'Hellenic' mode of thinking that has dominated Western hermeneutics in modern times.

These claims are historically questionable. The methods of midrash may superficially resemble postmodern literary criticism, but there are important differences. The rabbis regard midrash as an activity that relates to one unique text alone, the Bible. They do not see it as a paradigm of how to read every text. Moreover, they believe that Scripture does have an objective and absolute meaning, although in its entirety this is known only to God. Their mode of interpretation is disciplined, not arbitrary. By the application of the correct procedures the duly accredited and responsible rabbinid exegetes could draw out hidden senses from Scripture, but these senses were all facets of a larger and deeper, *objective* truth. The rabbinic exegetes could no more exhaust the potential of the text than they could exhaustively know the mind of God. But the inexhaustibility of their readings was taken not as a badge of their own creativity but as the emblem of the richness of the text itself. The polarization of Judaic and Hellenic hermeneutics, moreover, ignores the fact that many of the midrashic techniques are found applied to the interpretation of texts in Greek late antiquity.

See also: Bible, Hebrew; Hermeneutics; Hermeneutics, biblical; Meaning in Islamic philosophy; Theology, Rabbinic

PHILIP S. ALEXANDER

References and further reading

Alexander, P.S. (1990) 'Quid Athenis et Hierosolymis? Rabbinic Midrash and Hermeneutics in the Graeco-Roman world', in P.R. Davies and R.T. White (eds) *A Tribute to Geza Vermes*, Sheffield: Sheffield Academic Press, 101-24.(Criticizes Faur and Handelman below.)

Bereshit Rabba (fifth century AD), trans. H. Freedman, in H. Freedman and M. Simon (eds) *Midrash Rabbah: Translated into English*, London: Soncino Press, 3rd edn, 1961, 2 vols.(Exegetical commentary on Genesis. The same edition has the full Pentateuch, translated by various hands.)

Boyarin, D. (1990) *Intertextuality and the Reading of Midrash*, Bloomington, IN: University of Indiana Press. (Considers midrash in relation to intertextuality.)

Faur, J. (1986) *Golden Doves with Silver Dots: Semiotics and Textuality in Rabbinic Tradition*, Bloomington, IN: Indiana University Press.(Examines the semiotics of midrash.)

Goldberg, A. (1985) 'Form-Analysis of Midrashic Literature as a Method of Description', *Journal of Jewish Studies* 36: 159-74.(Outlines a rigorous approach to the description of the literary structure of midrash.)

Handelman, S.A. (1982) *The Slayers of Moses: The Emergence of Rabbinic Interpretation in Modern Literary Theory*, New York: State University of New York Press.(Expands on the argument that midrash has passed into contemporary academic discourse and has anticipated postmodernist ways of reading texts.)

Neusner J. (1987) *What is Midrash?*, Philadelphia, PA: Fortress Press.(A readable non-technical introduction to midrash.)

Pesiqta de-Rav Kahana (sixth century AD), trans. W.G. Braude and I.J. Kapstein, *Pesiqta de-Rab Kahana*, Philadelphia, PA: Jewish Publication Society of America, 1975.(Homiletic commentary.)

Samely, A. (1992) ‘Scripture’s Implicature: the Midrashic Assumptions of Relevance and Consistency’, *Journal of Semitic Studies* 37: 167-205.(An important study of two of the fundamental axioms of midrash.)

Stemberger, G. (1996) *Introduction to Talmud and Midrash*, trans. M. Bockmuehl, Edinburgh: T. & T. Clark. (The standard introduction to the midrashic literature, with full bibliographies.)

Vermes, G. (1973) *Scripture and Tradition in Judaism: Haggadic Studies*, Leiden: Brill.(A pioneering study which adopts a thematic and tradition-historical approach to the analysis of midrash.)

Mikhailovskii, Nikolai Konstantinovich (1842-1904)

Along with Lavrov, N.K. Mikhailovskii, a non-academic social theorist and literary critic, was the most representative and influential thinker of Russian populism. His most distinctive contribution to populist ideology was his attempt to reconcile the 'principle of individuality', so dear to the Russian intelligentsia, with the old, communal 'principles of the people', represented by the non-Westernized Russian peasantry. Unlike Herzen and Lavrov, Mikhailovskii did not see the principle of individuality as a product of Western progress, that is to say, as something which should be introduced from outside to the archaic world of Russian village commune. He challenged the stereotype which associated individualism with the capitalist West; instead, he tried to prove that the principle of individuality was in fact fully compatible with old Russian communalism, and incompatible with Western-type modernization. He did so by a radical redefinition of the very concept of individuality. For him individuality was not the product of a process of 'individualization' in the sense of loosening the communal bonds, making people socially differentiated, functionally specialized and separated from each other. On the contrary, by individuality he meant the 'inner wholeness', that is to say, the non-alienated, many-sided and harmonious development of human beings. Following the romantic critics of modernity, he claimed that individuality, so conceived, was being destroyed by capitalist progress.

Mikhailovskii's notion of individuality was similar to the Slavophile ideal of 'integral personality'. Unlike Slavophilism, however, Mikhailovskii's 'sociological romanticism' was bound up with a secularist worldview and a semi-positivist position in philosophy. For this reason its affinity with Slavophile romantic anticapitalism was relatively inconsequential and remained unnoticed.

1 Anthropological relativism and ethical personalism

Mikhailovskii admired the Russian word for truth - *pravda* - as designating not only theoretical truth but also truth as justice, truth of the moral consciousness. In his view the common root of the two sides of truth was provided by the principle of individuality, treating the concrete human individual as the measure of all things.

The application of this principle to epistemology led to a relativist, or 'subjectivist', theory of knowledge. Absolute truth, or truth in itself, cannot exist; there is only truth for man and, in addition, it is not the same for the entire human species. Concrete truth is always perspectivist; the so-called facts are not something objectively given, their perception and interpretation is preformed by the nature of our social bonds, the type of social cooperation and, last but not least, by the irreducible individual experience of each human being. Hence there is no escape from subjectivism in theoretical knowledge: 'truth is that which satisfies the mental cravings of man'.

However, this epistemological relativism was combined by Mikhailovskii with an unyielding ethicism, treating the human individual as the central and absolute value. For Mikhailovskii this axiological absolutism was simply the other side of his view that the individual was the only basis of truth. If this is so, he reasoned, human knowledge is always relative, but in the realm of values all humans should accept the principle of individuality as the absolute standard for truth as justice. The individual, he argued, is sacred and must never be sacrificed. The individual's need for wholeness was for him a universal phenomenon; therefore the wholeness of individual human beings was in his view the absolute criterion in axiology. Thus, his 'subjective sociology' was based on two tenets: extreme epistemological relativism, often taking the form of a sociological theory of knowledge, and ethical absolutism, elevating the value of the all-round human personality.

2 Theory of progress

The application of Mikhailovskii's axiology to social sciences led to a reversal of the liberal view of progress, as expressed, in particular, in Herbert Spencer's positivistic evolutionism (see Spencer, H.).

In his treatise 'Chto takoe progress' ('What is Progress?') (l869) Mikhailovskii accused Spencer of overlooking the fact that individual progress and social evolution (on the model of organic evolution) were mutually exclusive. Spencer defined progress in the organic world as transition from simplicity (homogeneity) to complexity (heterogeneity); in society the same organic evolution was being achieved through the development of the division of labour. The Russian thinker set against this theory a sociological conception which claimed that the development of the division of labour in society was essentially a retrogression, since it had been achieved at the

expense of individuality, which had suffered fragmentation and disintegration. This was so because only in an undifferentiated, egalitarian society can human individuality be diversified, integral and well-rounded. True progress, Mikhailovskii concluded, ’is the gradual approach to the fullest possible and the most diversified division of labour among man’s organs and the least possible division of labour among men’.

In his other works, especially in the series of articles ’Bor’ba za individualnost’’ (’The Struggle for Individuality’) (1875-6), Mikhailovskii used this formula as an argument for a backward-looking peasant utopia, idealizing the natural economy of the Russian village commune. The Russian peasant, he argued, lived a life that was primitive but full; he was economically self-sufficient, and therefore independent, many-sided, making use of all his capacities, being a tiller and a fisherman, a shepherd and an artist in one person. The absence of complex cooperation made the Russian peasants mutually independent, while simple cooperation (that is, a cooperation in which people were involved as ’whole beings’) united them in a moral solidarity based upon mutual sympathy and understanding. One should distinguish between *levels* and *types* of development. The peasant commune represented a lower level of development than the capitalist factory but was superior to it as a type of development; similarly, Western man belonged to a lower type of development than the Russian peasant, who had not yet lost his primitive ’wholeness’. Hence it was groundless to believe that capitalism had liberated the individual; on the contrary, it turned the individual into a ’mere organ’ of the social organism, mercilessly sacrificing concrete living individuals to the idol of ’maximum production’.

3 The struggle for individuality

Despite his criticism of biological ’organicism’ in sociology, Mikhailovskii yielded to it in his own theory of the ’struggle for individuality’. He attributed to complex societies a tendency to develop into a kind of superorganism whose human members would be reduced to the role of submissive organs. However, he challenged the Darwinian trust in the survival of the fittest with the pessimistic view that ’natural evolution’ was accomplished at the cost of a constant lowering of quality and deserved therefore to be seen as a retrograde process.

Mikhailovskii’s theory was founded on the proposition that there are different stages of individuality that struggle against each other and try to dominate each other. This proposition, derived from Haeckel’s classification of biological organisms (see Haeckel, E.H.), assumed the presence of an inevitable conflict between individual human beings and different supraindividual entities - such as factories (as units of complex cooperation), estates, nations, states and so on. All these *social* individualities could only develop at the expense of man’s freedom and wholeness. Therefore, Mikhailovskii concluded, ’society is man’s chief and worst enemy, an enemy against whom he must always be on guard’.

In his philosophy of history Mikhailovskii combined the theory of the struggle for individuality with Comte’s conception of the phases of intellectual development (see Comte, A.). He divided history into three great periods: the ’objectively anthropocentric’ period, the ’eccentric’ period, and the ’subjectively anthropocentric’ period. The first period knew only simple cooperation and could therefore preserve the original fullness of human beings. The second one, bound up with the social division of labour, was the period of development through alienation, separating the faculties of the species, objectifying them and setting against each other. The third period, which was to be realized as a result of a socialist transformation, would overcome the extreme forms of dependence on the division of labour, liberate individuals from the rule of objectified economic mechanisms and enable them to cooperate with each other in a free, ethical community.

An interesting part of Mikhailovskii’s social philosophy was his view on the question of gender. He thought that sexual differentiation of human beings had been greatly increased as a result of the extreme division of social roles and interpreted love as striving for reintegration through compensatory self-fulfilment in another being. It followed from this that in the homogeneously egalitarian society of the future the differences between sexes would be greatly reduced and the need for love much less important.

4 Mikhailovskii and Marxism

Among the authors who influenced Mikhailovskii’s view on capitalism - the highest phase of the ’eccentric’ period of history - the most important was Marx. As early as 1869, in his article ’Teoriia Darvina i obshchestvennaia nauka’ (’The Theory of Darwin and the Social Sciences’), Mikhailovskii expressed solidarity with Marx’s views

on the negative effects of the division of labour; he referred also to Adam Smith and Ferguson but, as a rule, quoted them after Marx's *Capital*. He was profoundly impressed by Marx's account of capitalist development as necessarily involving the expropriation of the immediate producers. Like other populists, he concluded from this that his country should do its best to avoid the capitalist way. Moreover, the adaptation of Marxism to his own conception persuaded Mikhailovskii that modern socialism and 'medieval forms of production', as represented by the village commune, were but different levels of the same type and, hence, that the workers' cause in Russia was essentially a conservative cause. In his review of the Russian translation of *Capital* (1872) he supported this conclusion by wholeheartedly endorsing Marx's denunciations of the class nature of 'bourgeois democracy'.

A few years later, however, Mikhailovskii became aware that Marxism contained also the notion of 'objective laws of history' and, therefore, could be used as an argument for the necessity of capitalist development in Russia. In his article 'Karl Marks pered sudom g. Zhukovskogo' ('Karl Marx Arraigned Before Mr Zhukovskii'), published in *Otechestvennye zapiski (Notes of the Fatherland)*, 1877, he reacted to this by formulating the dramatic dilemma of the Russian disciples of Marx, who as socialists saw capitalism as morally unacceptable but as historical determinists had to accept 'the capitalist phase', with all its painful consequences for the masses. Marx himself tried to dispute this diagnosis in his 'Pis'mo v redaktsiu *Otechestvennikh zapisok*' ('Letter to the Editor of *Otechestvennye zapiski*') (written 1877, published 1886) but failed to convince his Russian followers. Both Georgii Plekhanov and Pëtr Struve, the leader of the so-called 'legal Marxists' of the l890s, interpreted historical materialism as a rigidly deterministic theory, treating the capitalist development of backward countries as a historical inevitability. This caused Mikhailovskii to launch an energetic campaign against Marxism. He attacked Marxism as an 'inverted Hegelianism', trying to substitute the speculative concept of historical necessity for the empirical explanation of fact and thus justifying all the cruelties of history. Mikhailovskii's arguments exercised an influence on the 'legal Marxists', especially Berdiaev, helping them to evolve in the direction of philosophical idealism.

See also: Herzen, A.I.; Lavrov, P.L.; Slavophilism; Positivism, Russian; Positivism in the social sciences

ANDRZEJ WALICKI

List of works

Mikhailovskii, N.K. (1906-14) *Polnoe sobranie sochinenii (Complete Works)*, St Petersburg: M.M. Stasiulevich, 4th edn, 10 vols.

Edie, J.M., Scanlan, J.P. and Zeldin, M.B. (1965) *Russian Philosophy*, vol. 2, Chicago, IL: Quadrangle Books. (See 170-98 for an English translation of fragments from Mikhailovskii's philosophical writings including his seminal essay 'What is Progress?')

References and further reading

Billington, J.H. (1958) *Mikhailovsky and Russian Populism*, Oxford: Clarendon Press.(The only comprehensive monograph on Mikhailovskii in English.)

Hecker, J.F. (1915) 'The Sociological System of Mikhailovsky', *Russian Sociology*, pt II, ch. 2, 120-55, New York: Columbia University Press.(A systematic presentation of Mikhailovskii's 'subjective sociology'.)

Walicki, A. (1969) *The Controversy Over Capitalism: Studies in the Social Philosophy of the Russian Populists*, Oxford: Clarendon Press.(Contains a comprehensive analysis of Mikhailovskii's social philosophy, with special emphasis on his relationship to Marxism.)

Miki Kiyoshi (1897-1945)

A brilliant young philosopher and critic of his times, Miki Kiyoshi embodied in his life and thought Japan's tortured transition from Westernized modernity to world power. Associated with his teacher, Nishida Kitarō, he is generally regarded as representative of a Marxist turn on the Kyoto School.

As a young student, Miki chanced upon a copy of Nishida Kitarō's *Zen no kenkyū (Study of the Good)*, which turned him to the study of philosophy (see Nishida Kitarō). After studying under Nishida at Kyoto University, he spent two years in Germany, principally with Heidegger, in whose person he came to understand the post-war angst of Europe in the 1920s. He felt drawn to the optimism of thinkers such as Bergson and to the philosophical anthropology of Dilthey and Simmel. In pursuit of his own position he moved to Paris for a year where he threw himself into the works of Pascal, culminating in his maiden work, *Pasukaru ni okeru ningen no kenkyū (Pascal's Anthropology)*, which appeared in 1925.

On returning from abroad in 1924, Miki felt himself out of tune with the 'logic of place' that was occupying Nishida's attention at the time. Eager to bring his humanistic interests in line with the questions of the historical world, he turned to Marxist thought. The following year he took a post as professor at Hōsei University, and began to publish essays on historical materialism. In 1928 he was arraigned in court on charges of making an illegal contribution to the Japan Communist Party and received a suspended two-year sentence. In his own defence he insisted that he had always taken a critical stance to Marxism, which in turn earned him the derision of the Marxist establishment.

Hardened in his resolve both to systematize his own philosophical humanism and to clarify his critique of the rising militant fascism of the time, Miki set to work on a major study, *Kōzōryoku no ronri (The Logic of Structural Power)*. The serialized publication of the work (the final section of which was only published posthumously), coincided with the onset of the Sino-Japanese War. In 1938 Miki joined the controversial Shōwa Study Group where he was the chief architect of its idea of 'commonism' (*kyōdōshugi*) a curious blend of socialist, liberal, and quasi-fascist ideas which was offered as an alternative to the 'co-prosperity sphere' that the military regime was advancing as part of its 'new world order' in Asia. Miki was dismissed as a fool by leading right-wing ideologues and the study group was disbanded.

In 1943 Miki was sent as a war correspondent to the Philippines, an express punishment for his 'red' tendencies. He returned at the end of the year, only to find that almost none of his writings would be accepted for publication. In June 1945 he was arrested on a charge of disrupting public order and the powers behind his detention saw to it that he was denied amnesty at the end of the war. He died in prison.

Miki's overriding concern with philosophy's 'duty to the times' (*jimu*) made for a certain choppiness in his thought. Had he lived to write in time of peace, there is every indication he would have brought the various threads together into a unified body of thought. In what we have of his work, perhaps the most central notion is what he calls *techné* (*gijutsu*), the transformation of reality that serves as the bridge between and final judge of all human logos and pathos. Without it, the angst of the age was apt to lead only to privatized despair and a flight from reality, a tendency he saw in Heidegger. In works like *Tetsugaku nyūmon (Introduction to Philosophy)* and *Gijitsu tetsugaku (The Philosophy of Techné)*, Miki showed how industrial technology and its social problems belong to a wider phenomenon that begins in the natural world and rises all the way to intellectual disciplines and morality in 'the *techné* of the heart' in which existential anxieties are not only accepted but creatively reshaped.

See also: Kyoto School

J.W. HEISIG

List of works

Very little has been written about Miki in English and his works have never been translated.

Miki Kiyoshi (1925) *Pasukaru ni okeru ningen no kenkyū (Pascal's Anthropology)*, in *Miki Kiyoshi Zenshū (The Complete Works of Miki Kiyoshi)*, Tokyo: Iwanami Shoten, vol. 1.(Miki's study of Pascal.)

Miki Kiyoshi (1937-43) *Kōzōryoku no ronri (The Logic of Structural Power)*, in *Miki Kiyoshi Zenshū (The Complete Works of Miki Kiyoshi)*, Tokyo: Iwanami Shoten vol. 8.(Miki's major philosophical work.)

Miki Kiyoshi (1940) *Tetsugaku nyūmon (Introduction to Philosophy)*, in *Miki Kiyoshi Zenshū (The Complete Works of Miki Kiyoshi)*, Tokyo: Iwanami Shoten, vol. 7.(Work which shows much of Miki's later thinking.)

Miki Kiyoshi (1941) *Gijitsu tetsugaku (The Philosophy of Techné)*, in *Miki Kiyoshi Zenshū (The Complete Works of Miki Kiyoshi)*, Tokyo: Iwanami Shoten, vol. 7.(Looks at how industrial technology is part of a wider social phenomenon.)

References and further reading

Akamatsu Tsunihiko (1994) *Miki Kiyoshi: Tetsugakuteki shisaku no kiseki (Miki Kiyoshi: The Locus of Philosophical Speculation)*, Kyoto: Minerva.(Survey of Miki's thought.)

Karaki Junzō (1966) *Miki Kiyoshi*, Tokyo: Chikuma.(Biography of Miki Kiyoshi.)

Kawahara Hiroshi *et al.* (1978) *Kindai nihon no shisō (Modern Japanese Thought)*, Tokyo: Yohikaku, 79-148.(A brief but clear résumé of Miki's major ideas.)

Mill, James (1773-1836)

James Mill, who is today remembered mainly as Bentham's chief disciple and John Stuart Mill's father, was a British philosopher, political theorist, historian, psychologist, economist, educationist and journalist. He was also largely responsible for clarifying and systematizing Bentham's utilitarianism, for introducing a distinction between 'lower', animal pleasures and 'higher', uniquely human ones, and for organizing the small but influential band of Bentham's followers that became known as the 'philosophic radicals'. In politics, he favoured representative democracy as the only practicable system of government capable of maximizing individual and communal happiness.

A prolific author, Mill wrote five books and more than a thousand articles and reviews. The book for which he was best known during his lifetime was his three-volume *History of British India* (1818). His *Essays* (1825) on sundry subjects - including 'Government' and 'Education' - was followed by his *Elements of Political Economy* (1826), his two-volume *Analysis of the Phenomena of the Human Mind* (1829) and *A Fragment on Mackintosh* (1835).

Despite this diversity of topics, Mill's thinking is informed by a unified philosophical vision. In psychology he was an 'associationist' who held that our beliefs and behaviour are the product of pleasurable and painful associations of ideas. In ethics he was a utilitarian who held that the proper aim of human actions and practices was the promotion of human happiness (see Utilitarianism). Unlike Jeremy Bentham (see Bentham, J. §2), however, who held that 'pushpin is as good as poetry' - that is, that a pleasure is a pleasure, whatever its source - Mill believed that happiness is not synonymous with pleasure, particularly of the 'lower' or 'animal' sort. Human happiness comes from developing and using uniquely human powers - reason, logic, the love of truth and beauty - to their fullest extent. Hence his interest in education, which leads people toward the right sorts of associations: to find truth pleasurable and error painful; to delight in doing right and experience pain in doing wrong; and so on.

These themes come together in Mill's political philosophy, which is summarized succinctly in his essay on 'Government'. Government, properly structured, promotes the happiness of the whole community and its individual members. Humans not only naturally desire happiness; they seek to satisfy that desire by investing as little effort as possible in obtaining it. Labour is the means of obtaining happiness. Humans find labour painful; hence they will, if they can, live off the labour of others. But to the extent that some live off the labour of others, the latter's incentive to work is diminished, and the happiness and prosperity of the community is imperilled.

The point and purpose of government is to maximize the happiness of the whole community by minimizing the extent to which some members live off the labour of others. Mill maintains that this aim cannot be achieved in a monarchy, nor in an aristocracy, nor in a direct democracy (the last taking too much time away from labouring). The only system capable of maximizing individual and communal happiness is a representative democracy in which representatives are elected to serve short terms before returning to the ranks of the electorate. So structured, representative government gives representatives every incentive to serve the people instead of themselves.

During his lifetime James Mill was regarded as a radical in politics and philosophy. That his ideas now seem conventional or even conservative shows how the radicalism of one age can become the conservatism of another.

See also: Mill, J.S. §1

TERENCE BALL

List of works

Mill, J. (1818) *History of British India*, ed., abridged and with intro. by W. Thomas, Chicago, IL: University of Chicago Press, 1975, 1 vol.(Written from a rather zealously reforming utilitarian perspective and highly critical of native Indian institutions.)

Mill, J. (1826) *Elements of Political Economy*, London, 3rd edn.(Mill's 'manual' or 'text-book' of Ricardian economics.)

Mill, J. (1829) *Analysis of the Phenomena of Human Mind*, London: Longman, Green, Reader & Dyer, 2 vols.(A classic 'associationist' account of mental phenomena.)

Mill, J. (1835) *A Fragment on Mackintosh*, London.(An acerbic attack on Sir James Mackintosh, a leading Whig apologist and critic of the Philosophical Radicals.)

Mill, J. (1992) *Political Writings*, ed. T. Ball, Cambridge: Cambridge University Press. (Includes Mill's essays on 'Government' and 'Education', T.B. Macaulay's famous attack on the former, and Mill's heretofore unknown reply.)

References and further reading

Bain, A. (1882) *James Mill: A Biography*, London: Longman Green; repr. New York, Augustus M. Kelly Publishers, 1966.(A useful, if somewhat dry, biography.)

Ball, T. (1994) *Reappraising Political Theory: Revisionist Studies in the History of Political Thought*, Oxford: Clarendon Press.(Chapters 6-8 reconsider aspects of James Mill's political philosophy.)

Plamenatz, J. (1958) *The English Utilitarians*, Oxford: Blackwell, 2nd edn.(A concise account and assessment of key themes and thinkers in the utilitarian tradition, including James Mill.)

Thomas, W.E.S. (1979) *The Philosophical Radicals*, Oxford: Clarendon Press.(An excellent history of Bentham and his sometimes-wayward followers, including James and John Stuart Mill.)

Mill, John Stuart (1806-73)

John Stuart Mill, Britain's major philosopher of the nineteenth century, gave formulations of his country's empiricist and liberal traditions of comparable importance to those of John Locke. He united enlightenment reason with the historical and psychological insights of romanticism. He held that all knowledge is based on experience, believed that our desires, purposes and beliefs are products of psychological laws of association, and accepted Bentham's standard of the greatest total happiness of all beings capable of happiness - the principle of 'utility'. This was Mill's enlightenment legacy; he infused it with high Romantic notions of culture and character.

In epistemology Mill's empiricism was very radical. He drew a distinction between 'verbal' and 'real' propositions similar to that which Kant made between analytic and synthetic judgments. However, unlike Kant, Mill held that not only pure mathematics but logic itself contains real propositions and inferences, and unlike Kant, he denied that any synthetic, or real, proposition is a priori. The sciences of logic and mathematics, according to Mill, propound the most general laws of nature and, like all other sciences, are in the last resort grounded inductively on experience.

We take principles of logic and mathematics to be a priori because we find it inconceivable that they should not be true. Mill acknowledged the facts which underlie our conviction, facts about unthinkability or imaginative unrepresentability, and he sought to explain these facts in associationist terms. He thought that we are justified in basing logical and mathematical claims on such facts about what is thinkable - but the justification is itself a posteriori.

What then is the nature and standing of induction? Mill held that the primitive form of induction is enumerative induction, simple generalization from experience. He did not address Hume's sceptical problem about enumerative induction. Generalization from experience is our primitive inferential practice and remains our practice when we become reflectively conscious of it - in Mill's view nothing more needs to be said or can be said. Instead he traced how enumerative induction is internally strengthened by its actual success in establishing regularities, and how it eventually gives rise to more searching methods of inductive inquiry, capable of detecting regularities where enumerative induction alone would not suffice. Thus whereas Hume raised sceptical questions about induction, Mill pushed through an empiricist analysis of deduction. He recognized as primitively legitimate only the disposition to rely on memory and the disposition to generalize from experience. The whole of science, he thought, is built from these.

In particular, he did not accept that the mere fact that a hypothesis accounts for data can ever provide a reason for thinking it true (as opposed to thinking it useful). It is always possible that a body of data may be explained equally well by more than one hypothesis. This view, that enumerative induction is the only authoritative source of general truths, was also important in his metaphysics. Accepting as he did that our knowledge of supposed objects external to consciousness consists only in the conscious states they excite in us, he concluded that external objects amount only to 'permanent possibilities of sensation'. The possibilities are 'permanent' in the sense that they can be relied on to obtain if an antecedent condition is realized. Mill was the founder of modern phenomenalism.

In ethics, Mill's governing conviction was that happiness is the sole ultimate human end. As in the case of induction, he appealed to reflective agreement, in this case of desires rather than reasoning dispositions. If happiness was not 'in theory and in practice, acknowledged to be an end, nothing could ever convince any person that it was so' (1861a: 234). But he acknowledged that we can will to do what we do not desire to do; we can act from duty, not desire. And he distinguished between desiring a thing as 'part' of our happiness and desiring it as a means to our happiness. The virtues can become a part of our happiness, and for Mill they ideally should be so. They have a natural base in our psychology on which moral education can be built. More generally, people can reach a deeper understanding of happiness through education and experience: some forms of happiness are inherently preferred as finer by those able to experience them fully.

Thus Mill enlarged but retained Bentham's view that the happiness of all, considered impartially, is the standard of conduct. His account of how this standard relates to the fabric of everyday norms was charged with the nineteenth century's historical sense, but also maintained links with Bentham. Justice is a class of exceptionally stringent obligations on society - it is the 'claim we have on our fellow-creatures to join in making safe for us the

very groundwork of our existence' (1865b: 251). Because rights of justice protect this groundwork they take priority over the direct pursuit of general utility as well as over the private pursuit of personal ends.

Mill's doctrine of liberty dovetails with this account of justice. Here he appealed to rights founded on 'utility in the largest sense, grounded on the permanent interests of man as a progressive being' (1859: 224). The principle enunciated in his essay On Liberty (1859) safeguards people's freedom to pursue their own goals, so long as they do not infringe on the legitimate interests of others: power should not be exercised over people for their own good. Mill defended the principle on two grounds. It enables individuals to realize their potential in their own distinctive way, and, by liberating talents, creativity and energy, it institutes the social conditions for the moral development of culture and character.

1 Life

Mill was born in London on the 20 May 1806, the eldest son of a Scotsman, James Mill, and an English woman, Harriet Burrow. James educated his son himself - an education made famous by the account John Stuart gave of it in his *Autobiography* (1873). He taught John the classics, logic, political economy, jurisprudence and psychology - starting with Greek at the age of three. John was brought up in a circle of intellectual and political radicals, friends of his father, which included Jeremy Bentham and David Ricardo. In his twenties (not surprisingly, perhaps) he was afflicted by a deep depression from which he recovered partly through reading poetry. In those and subsequent years he also came to know some of the most interesting younger figures in English politics and culture. These included conservative critics of Benthamism, as well as radical adherents of it.

Mill followed his father into the East India Company, where he became an influential official, resigning only in 1858 when, following the Indian Mutiny, the Company was taken over by the Crown and the governance of India became the direct business of the British State. In 1851 he married Harriet Taylor, who in his own account greatly influenced his social philosophy. In the 1860s he was briefly a member of Parliament, and throughout his life was involved in many radical causes. Among them was his lifelong support for women's rights - see *The Subjection of Women* (1869).

Mill made his philosophical reputation with his *System of Logic*, published in 1843. The *Principles of Political Economy* (1848) was a synthesis of classical economics which defined liberal orthodoxy for at least a quarter of a century. His two best-known works of moral philosophy, *On Liberty* and *Utilitarianism*, appeared later - in 1859 and 1861. But he had been thinking about ethics and politics all his life, and it is his moral and political philosophy which is at present most widely read.

2 Language and logic

Nevertheless, Mill's epistemology and metaphysics remain as interesting and relevant as his better-known views in ethics and politics, and it is from these aspects of his philosophy that a general survey must start. In the *System of Logic* Mill distinguishes 'verbal' and 'real' propositions, and correspondingly, 'merely apparent' and 'real' inferences. An inference is merely apparent when no move to a new assertion has been made. For this to be so, the conclusion must literally have been asserted in the premises. In such a case, there can be no epistemological problem about justifying the apparent inference - there *is* nothing to justify. A verbal proposition can now be defined as a conditional proposition corresponding to a merely apparent inference. Propositions and inferences which are not verbal or merely apparent are real.

Mill argues that not only mathematics but logic itself contains real inferences. To demonstrate this he embarks on a semantic analysis of sentences and terms (he calls them 'propositions' and 'names'), of syllogistic logic and of the so-called 'Laws of Thought'. His analysis has imperfections and he does not unify it in a fully general account, but he supplies the foundations of such an account, and in doing so takes the empiricist epistemology of logic and mathematics to a new level.

The starting point is a distinction between the denotation and connotation of names. Names, which may be general or singular, denote things and connote attributes of things. A general name connotes attributes and denotes each object which has those attributes. Most singular names also connote attributes.

There is, however, an important class of singular names - proper names in the ordinary sense, such as

'Dartmouth' - which denote an object without connoting any property (see Proper names §§1, 6). Identity propositions which contain only non-connotative names, such as 'Tully is Cicero', are verbal, in Mill's view. They lack content in the sense that, according to Mill, the only information conveyed is about the names themselves: 'Tully' denotes the same object as 'Cicero' does. Mill's point is that there is no fact in the world to which 'Cicero is Tully' corresponds. But to class these propositions as verbal would require a change in the characterization of verbal propositions given above. Moreover, knowledge that Cicero is Tully is not a priori. We cannot know the proposition to be true just by reflecting on the meaning of the names - whereas Mill's overall intention is that the class of verbal propositions should be identical with the class of propositions which are innocuously a priori *because* they are empty of content. He does not comment on these difficulties.

The meaning of a declarative sentence - 'the import of a proposition' - is determined by the connotation, not the denotation, of its constituent names; the sole exception being connotationless proper names, where meaning is determined by denotation. (Again Mill does not explain how this thesis about the meaning of proper names is to be reconciled with the a posteriority of 'Cicero is Tully'.) Mill proceeds to show how the various syntactic forms identified by syllogistic theory yield conditions of truth for sentences of those forms, when the connotation of their constituent names is given.

Armed with this analysis he argues that logic contains real inferences and propositions. He assumes that to assert a conjunction, 'A and B', is simply to assert A and to assert B. He defines 'A or B' as 'If not A, then B, and if not B, then A'. 'If A then B' means, he thinks, 'The proposition B is a legitimate inference from the proposition A'. From these claims it follows that certain deductive inferences, for example, from a conjunction to one of its conjuncts, are merely apparent. But, Mill holds, the laws of contradiction and excluded middle are real - and therefore a posteriori - propositions. He takes it that 'not P' is equivalent in meaning to 'It is false that P'; if we further assume the equivalence in meaning of P and 'It is true that P', the principle of contradiction becomes, as he puts it, 'the same proposition cannot at the same time be false and true'. 'I cannot look upon this', he says, 'as a merely verbal proposition'. He makes analogous remarks about excluded middle, which turns - on these definitions - into the principle of bivalence: 'Either it is true that P or it is false that P'.

Mill adds an epistemological argument to this semantic analysis. If logic did not contain real inferences, all deductive reasoning would be a *petitio principii*, a begging of the question, and it could produce no new knowledge. Yet clearly it does produce new knowledge. So logic must contain real inferences.

Unfortunately, Mill mixes up this epistemological argument with an interesting but distinct objective. He wants to show that 'all inference is from particulars to particulars', in order to demystify the role that general propositions play in thought. He argues that in principle they add nothing to the force of an argument; particular conclusions could always be derived inductively direct from particular premises. Their value is psychological. They play the role of 'memoranda' or summary records of the inductive potential of all that we have observed, and they facilitate 'trains of reasoning' (for example, as in 'This is A; All As are Bs; No Bs are Cs; so this is not C'). Psychologically they greatly increase our memory and reasoning power, but epistemologically they are dispensable.

This thesis is connected to Mill's rejection of 'intuitive' knowledge of general truths and to his inductivism (see §5 below). But there is also a deeper way in which a radical empiricist must hold that all inference is from particulars to particulars. For consider the inference from 'Everything is F' to 'a is F'. Is it a real or merely apparent inference? It is impossible to hold it real if one also wishes to argue that real inferences are a posteriori. But the only way in which Mill can treat it as verbal is to treat the premise as a conjunction: 'a is F and b is F and…'. If that approach is precluded, then all that remains is to deny that 'Everything is F' is propositional - it must, rather, express an inferential commitment. Both approaches are very close to the surface in Mill's discussion of the syllogism, though neither emerges clearly.

3 Mathematics

The strategy which Mill applies to mathematics is broadly similar to his approach to logic. If it was merely verbal, mathematical reasoning would be a *petitio principii*, but semantic analysis shows that it contains real propositions.

Mill provides brief but insightful empiricist sketches of geometry and arithmetic. The theorems of geometry are deduced from premises which are real propositions inductively established. (Deduction is itself largely a process of real inference.) These premises, where they are not straightforwardly true of physical space, are true in the limit.

Geometrical objects - points, lines, planes - are ideal or ‘fictional’ limits of ideally constructible material entities. Thus the real empirical assertion underlying an axiom such as ‘Two straight lines cannot enclose a space’ is something like ‘The more closely two lines approach absolute breadthlessness and straightness, the smaller the space they enclose’.

Applying his distinction between denotation and connotation, Mill argues that arithmetical identities such as ‘Two plus one equals three’ are real propositions. Number terms denote ‘aggregates’ and connote certain attributes of aggregates. (He does not say that they denote those attributes of the aggregates, though perhaps he should have done.) ‘Aggregates’ are natural, not abstract, entities - ‘collections’ or ‘agglomerations’ individuated by a principle of aggregation. This theory escapes some of the influential criticisms Frege later made of it, but its viability none the less remains extremely doubtful. The respects in which aggregates have to differ from sets if they are to be credibly natural, and not abstract, entities are precisely those in which they seem to fail to produce a fully adequate ontology for arithmetic. (One can, for example, number numbers, but can there be aggregates of aggregates, or of attributes of aggregates, if aggregates are natural entities?)

However this may be, Mill’s philosophical programme is clear. Arithmetic, like logic and geometry, is a natural science, concerning a category of the laws of nature - those concerning aggregation. The fundamental principles of arithmetic and geometry, as well as of logic itself, are real. Mill provides the first thoroughly empiricist analysis of meaning and of deductive reasoning itself.

He distinguishes his view from three others - ‘Conceptualism’, ‘Nominalism’ and ‘Realism’. ‘Conceptualism’ is his name for the view which takes the objects studied by logic to be psychological states or acts. It holds that names stand for ‘ideas’ which make up judgments and that ‘a proposition is the expression of a relation between two ideas’. It confuses logic and psychology by assimilating propositions to judgments and attributes of objects to ideas. Against this doctrine Mill insists that:

> All language recognizes a difference between doctrine or opinion, and the fact of entertaining the opinion; between assent, and what is assented to…. Logic, according to the conception here formed of it, has no concern with the nature of the act of judging or believing; the consideration of that act, as a phenomenon of the mind, belongs to another science.
>
> (1843: 87)

The Nominalists - Mill cites Hobbes - hold that logic and mathematics are entirely verbal. Mill takes this position much more seriously than Conceptualism and seeks to refute it in detail. His main point is that Nominalists are only able to maintain their view because they fail to distinguish between the denotation and the connotation of names, ‘ seeking for their meaning exclusively in what they denote’ (1843: 91) (see Nominalism §3).

Nominalists and Conceptualists hold that logic and mathematics can be known non-empirically, while yet retaining the view that no real proposition about the world can be so known. Realists hold that logical and mathematical knowledge is knowledge of universals which exist in an abstract Platonic domain; the terms that make up sentences being signs that stand for such universals. Versions of this view were destined to stage a major revival in philosophy, and semantic analysis would be their main source, but it is the view Mill takes least seriously.

In the contemporary use of the term, Mill is himself a nominalist - he rejects abstract entities (see Abstract objects §4). However, just as severe difficulties lie in the way of treating the ontology of arithmetic in terms of aggregates rather than sets, so there are difficulties in treating the ontology of general semantics without appealing to universals and sets, as well as to natural properties and objects. We can have no clear view of how Mill would have responded to these difficulties had they been made evident to him. But we can be fairly sure that he would have sought to maintain his nominalism.

However, his main target is the doctrine that there are real a priori propositions (see A priori). What, he asks, goes on in practice when we hold a real proposition to be true a priori? We find its negation inconceivable, or that it is derived, by principles whose unsoundness we find inconceivable, from premises whose negation we find inconceivable. Mill is not offering a definition of what is meant by such terms as ‘a priori’, or ‘self-evident’; his point is that facts about what we find inconceivable are all that lends colour to the use of these terms.

They are facts about the limits, felt by us from the inside, on what we can imagine perceiving. Mill thought he

could explain these facts about unthinkability, or imaginative unrepresentability, in associationist terms, and much of his work claims to do so. This associationist psychology is unlikely nowadays to convince, but that does not affect his essential point: the step from our inability to represent to ourselves the negation of a proposition, to acceptance of its truth, calls for justification. Moreover, the justification itself must be a priori if it is to show that the proposition is known a priori.

4 'Psychologism' and naturalism

Mill is often mistakenly accused of 'psychologism' in his treatment of logic - an accusation which seems to go back to Husserl (and one which Frege does not make). 'Psychologism' is the view that laws of logic are psychological laws concerning our mental processes; or that 'meanings' are mental entities, and that 'judgments' assert relationships among these entities. But Mill's view, as we have seen, is that logic and mathematics are the most general empirical sciences, governing *all* phenomena. He explicitly holds that the distinction between necessary and contingent truths, understood 'metaphysically', is empty. And he dismisses the Conceptualist claim that names refer to ideas and propositions express or assert a psychological relation between them.

What explains, then, the attribution of 'psychologism' to Mill? Husserl quotes a passage from *An Examination of Sir William Hamilton's Philosophy* (1865a), which has been cited many times since:

> Logic is not the theory of Thought as Thought, but of valid Thought; not of thinking, but of correct thinking. *It is not a Science distinct from, and coordinate with Psychology. So far as it is a science at all, it is a part, or branch, of Psychology; differing from it, on the one hand as the part differs from the whole, and on the other, as an Art differs from a Science. Its theoretic grounds are wholly borrowed from Psychology, and include as much of that science as is required to justify the rules of the art.*
>
> (1865a: 359; italics show portion quoted by Husserl)

To give this a psychologistic reading is to take it out of context. Mill means that the logician must formulate rules of reasoning in a manner which will be as helpful as possible to inquirers, and must draw on the psychology of thought to do so. It is in that sense that the art of the logician borrows from the science of the psychologist. How best to promote the art of clear thinking is a psychological question. None the less,

> the laws, in the scientific sense of the term, of Thought as Thought - do not belong to Logic, but to Psychology: and it is only the *validity* of thought which Logic takes cognizance of.
>
> (1865a: 359)

So it is wrong to accuse Mill of psychologism about logic. But there is a sense in which his view of our most basic forms of inductive reasoning is psychologistic, or naturalistic. For how does he respond to the Kantian claim that the very possibility of knowledge requires that there be a priori elements in our knowledge? Even if we accept his inductive account of logic and mathematics, must we not accept the principle of induction itself as a priori?

For Mill, the primitive form of reasoning - in both the epistemological and the aetiological sense - is enumerative induction, the disposition to infer that all As are B from the observation of a number of As which are all B. (Or to the conclusion that a given percentage of all As are B from the observation of that percentage of Bs among a number of As.) We spontaneously agree in reasoning that way, and in holding that way of reasoning to be sound. This method of reasoning, enumerative induction, is not a merely verbal principle. So it cannot on Mill's own account be a priori. Mill says that we learn 'the laws of our rational faculty, like those of every other natural agency', by 'seeing the agent at work'. We bring our most basic reasoning dispositions to self-consciousness by critical reflection on our actual practice. He is right to say that this reflective scrutiny of practice is, in a *certain* sense, an a posteriori process. It examines dispositions which we have before we examine them. Having examined our dispositions, we reach a reflective equilibrium in which we endorse some - and perhaps reject others. We endorse them as sound norms of reasoning. There is nothing more to be said: no further story, platonic or transcendental, to be told.

Unlike Hume, or even Reid, Mill shows no interest at all in scepticism. If one thinks that scepticism is both unanswerable and unserious this may be true philosophic wisdom. But to Mill's epistemological critics, whether they were realists or post-Kantian idealists, it seemed obvious that it was evasion, not wisdom. Naturalism could only seem to differ from scepticism by being uncritical, and in this we find the truth in the allegation that Mill's

system of logic is 'psychologistic'; if it is sound criticism, it is sound criticism of all naturalistic epistemology.

5 Inductive science

Mill does not raise purely sceptical questions about simple generalization from experience; he none the less thinks it a highly fallible method. His aim is to show how reasoning methods can evolve from it which greatly reduce the fallibility of induction, even though they can never wholly eliminate it.

Humankind begins with 'spontaneous' and 'unscientific' inductions about particular unconnected natural phenomena or aspects of experience. As these generalizations accumulate and interweave, they justify the second-order inductive conclusion that *all* phenomena are subject to uniformity, and more specifically, that all have discoverable sufficient conditions. In this less vague form, the principle of general uniformity becomes, given Mill's analysis of causation, the Law of Universal Causation. It in turn provides (Mill believes) the grounding assumption for a new style of reasoning about nature - eliminative induction.

In this type of reasoning, the assumption that a type of phenomenon has uniform causes, together with a (revisable) assumption about what its possible causes are, initiates a comparative inquiry in which the actual cause is identified by elimination. Mill formulates the logic of this eliminative reasoning in his well-known 'Methods of Experimental Inquiry' (Chapter 7, Book 2 of *System of Logic*). (A full account is given in Mackie (1974).) His picture of the interplay between enumerative and eliminative reasoning, and of the way it entrenches, from within, our rational confidence in the inductive process, is elegant and penetrating.

The improved scientific induction which results from this new style of reasoning spills back onto the principle of Universal Causation on which it rests, and raises its certainty to a new level. That in turn raises our confidence in the totality of particular enumerative inductions from which the principle is derived. So the amount of confidence with which one can rely on the 'inductive process' as a whole depends on the point which has been reached in its history - though the confidence to be attached to particular inductions always remains variable.

Mill's inductivism - his view that enumerative induction is the only ultimately authoritative method of inference to new truths - was rejected by William Whewell (see Whewell, W. §2), who argued that the really fundamental method in scientific inquiry was the Hypothetical Method, in which one argues to the truth of a hypothesis from the fact that it would explain observed phenomena (see Inference to the best explanation). Mill had read Whewell's *History of the Inductive Sciences* (1837), and could hardly fail to be aware of the pervasiveness of hypotheses in the actual process of inquiry, or of their indispensability in supplying working assumptions - their 'heuristic' value, as Whewell called it. But what Mill could not accept was that the mere fact that a hypothesis accounted for the data in itself provided a reason for thinking it true.

Yet Whewell's appeal was to the actual practice of scientific reasoning, as observed in the history of science. An appeal of that kind was precisely what Mill, on his own principles, could not ignore. If the disposition to hypothesize is spontaneous, why should it not be recognized as a fundamental method of reasoning to truth, as enumerative induction is?

Mill's refusal to recognize it is not arbitrary. The essential point underlying it is a powerful one: it is the possibility that a body of data may be explained equally well by more than one hypothesis. Mill does not deny the increasingly deductive and mathematical organization of science - he emphasizes it. That is quite compatible with his inductivism, and indeed central to his account of the increasing reliability of the inductive process. He further agrees that a hypothesis can sometimes be shown, by eliminative methods of inductive reasoning which he accepts, to be the only one consistent with the facts. And he allows various other cases of apparently purely hypothetical reasoning which are, in his view, genuinely inductive.

When all such cases have been taken into account, we are left with pure cases of the Hypothetical Method, in which the causes postulated are not directly observable, and not simply because they are assumed to operate - in accordance with known laws, inductively established - in regions of time or space too distant to observe. What are we to say of such hypotheses? For example of the 'undulatory' theory of light? They cannot, Mill says, be accepted as inductively established truths, not even as probable ones.

> An hypothesis of this kind is not to be received as probably true because it accounts for all the known phenomena; since this is a condition sometimes fulfilled tolerably well by two conflicting hypotheses; while

> there are probably many others which are equally possible, but which from want of anything analogous in our experience, our minds are unfitted to conceive.
>
> (1843: 500)

Such a hypothesis can suggest fruitful analogies, Mill thinks, but cannot be regarded as yielding a new truth itself. The data do not determine a unique hypothesis: it is this possibility of underdetermination which stops him from accepting hypothetical reasoning as an independent method of achieving truth.

In seeing the difficulty Mill is certainly on solid ground. What he does not see, however, is how much must be torn from the fabric of our belief if inductivism is applied strictly. So it is an important question whether the difficulty can be resolved - and whether it can be resolved within a naturalistic framework which does not appeal to an underlying idealism, as Whewell did. If naturalism can endorse the hypothetical method, then among other things it can develop a more plausible empiricism about logic and mathematics than Mill's. But the ramifications of his inductivism are even wider, as becomes apparent from an examination of his general metaphysics.

6 Mind and matter

Mill sets out his metaphysical views in *An Examination of Sir William Hamilton's Philosophy*. Hamilton was the last eminent representative of the Scottish Common Sense School, and a ferocious controversialist - in Mill's eyes a pillar of the right-thinking establishment, ripe for demolition. The result is that Mill's discussion of general metaphysical issues is cast in a highly polemical form which leaves important issues shrouded in obscurity. He does however give himself space to develop his view of our knowledge of the external world.

He begins with a doctrine which he rightly takes to be generally accepted (in his time) on all sides: 'that all the attributes which we ascribe to objects, consist in their having the power of exciting one or another variety of sensation in our minds; that an object is to us nothing else than that which affects our senses in a certain manner' (1865a: 6). This is 'the doctrine of the Relativity of Knowledge to the knowing mind'. It makes epistemology, in Mill's words, the 'Interpretation of Consciousness'. He proceeds to analyse what we mean when we say that objects are external to us:

> We mean, that there is concerned in our perceptions something which exists when we are not thinking of it; which existed before we had ever thought of it, and would exist if we were annihilated; and further, that there exist things which we never saw, touched or otherwise perceived, and things which have never been perceived by man. This idea of something which is distinguished from our fleeting impressions by what, in Kantian language, is called Perdurability; something which is fixed and the same, while our impressions vary; something which exists whether we are aware of it or not, constitutes altogether our idea of external substance. Whoever can assign an origin to this complex conception, has accounted for what we mean by the belief in matter.
>
> (1865a: 178-9)

To assign this origin Mill postulates that

> we are capable of forming the conception of Possible sensations; sensations which we are not feeling at the present moment, but which we might feel, and should feel if certain conditions were present.
>
> (1865a: 177)

> These possibilities, which are conditional certainties, need a special name to distinguish them from mere vague possibilities, which experience gives no warrant for reckoning upon. Now, as soon as a distinguishing name is given, though it be only to the same thing regarded in a different aspect, one of the most familiar experiences of our mental nature teaches us, that the different name comes to be considered as the name of a different thing.
>
> (1865a: 179-80)

Physical objects are 'Permanent Possibilities of Sensation' (There is a change in the 'permanent' possibilities of sensation whenever there is change in the world. Mill also uses other terms, such as 'certified' or 'guaranteed'.) We often find that whenever a given cluster of certified possibilities of sensation obtains, then a certain other cluster follows. 'Hence our ideas of causation, power, activity…become connected, not with sensations, but with

groups of possibilities of sensation' (1865a: 181) (see Phenomenalism §1).

However, even if our notion of matter as the external cause of sensations can be explained on psychological principles, it is still possible to hold that good grounds can be given for thinking the notion to have instances. There might be a legitimate inference from the existence of the permanent possibilities and their correlations to the existence of an external cause of our sensations. It is at just this point that Mill's inductivism plays a part. The inference would be a case of hypothetical reasoning, to an explanation of experience which transcended all possible data of experience; and that is precisely what Mill rejects: 'I assume only the tendency, but not the legitimacy of the tendency, to extend all the laws of our own experience to a sphere beyond our experience' (1865a: 187).

If matter is the permanent possibility of sensation what is mind? Can it also be resolved into 'a series of feelings, with a background of possibilities of feeling'? Mill finds in this view a serious difficulty: to remember or expect a state of consciousness is not simply to believe that it has existed or will exist; it is to believe that *I myself* have experienced or will experience that state of consciousness.

> If, therefore, we speak of the Mind as a series of feelings, we are obliged to complete the statement by calling it a series of feelings which is aware of itself as past and future; and we are reduced to the alternative of believing that the Mind, or Ego, is something different from any series of feelings, or possibilities of them, or of accepting the paradox, that something which *ex hypothesi* is but a series of feelings, can be aware of itself as a series.
>
> (1865a: 194)

Thus although Mill is unwilling to accept 'the common theory of Mind, as a so-called substance', the self-consciousness involved in memory and expectation drives him to 'ascribe a reality to the Ego - to my own Mind - different from that real existence as a Permanent Possibility, which is the only reality I acknowledge in Matter' (1865a: 208).

This ontology, Mill thinks, is consistent with common sense realism about the world. Phenomenalism - the conception of matter as possibility of experience - allegedly leaves common sense and science untouched. In particular, mind and experience is still properly seen as a part of the natural order.

Yet if phenomenalism is right, only the experiences are real. Mill thinks we are led to that conclusion by the very standards of reasoning recognized in a naturalistic 'science of science', or 'system of logic'. If he is right, then the naturalistic vision of the world which sees minds as part of a larger causal order is self-undermining. For if we are led to the conclusion that only states of consciousness are real by *an application of naturalism's own standards*, then that conclusion has to be understood on the same level as the naturalistic affirmation that states of consciousness are themselves part of a larger causal order external to them - and therefore as inconsistent with it. Causal relations cannot exist between fictional entities which are mere markers for possibilities of sensation.

This is the fault-line in Mill's epistemology and metaphysics. Either naturalism undermines itself, or there is something wrong with Mill's inductivist analysis of our natural norms of reasoning, or with his endorsement of the doctrine of the Relativity of Knowledge, or with both. It is not obvious that Mill's most fundamental tenet - his naturalistic view of the mind - can be safeguarded by rejecting inductivism and endorsing the hypothetical method. There is still something implausible about hypothesizing the world as an explanation of pure experience. Mill himself explicitly acknowledged that memory, as well as induction, has epistemic authority. Had he analysed the significance of such an acknowledgment more thoroughly, he might have noted a parallel: on the one hand a primitive epistemic norm which warrants assertions about the past based on present memory-experiences; on the other, primitive epistemic norms which warrant assertions about the physical world based on perceptual experience. But perhaps that would have taken him too far in the direction of Reid's principles of common sense.

7 Freedom and the moral sciences

The sixth and last book of the *System of Logic* is a classical statement of methodology in the 'moral' sciences (that is, the human sciences). Its strength derives partly from the fact that Mill was a philosopher who also practised the whole range of these sciences as they then stood. He was mainly known as a political economist, but had strong interests in psychology and in the nascent science of sociology. He thought as an economist as well as a

philosopher about socialism, taxation and systems of property, and he thought in sociological terms about such topics as democracy and the role of moral and intellectual elites. He also took an interest in a variety of psychological topics, including desire, pleasure and will, and the origins of conscience and justice.

The phenomena of mind and society are, in Mill's view, causal processes. If mind and society are part of the causal order, and causation is regular succession, then the general model of explanation he has proposed, according to which explanation subsumes facts under laws linking them to their causal antecedents, will apply, he thinks, to the moral sciences. It may be hard for moral science to live up to it, in view of the complexity of its data, but the model stands as an ideal. Important issues remain about the character of and relationships between the various moral sciences, and Mill treats these issues in detail. But he does not think that the very idea of a moral science raises new metaphysical or epistemological problems (see Explanation in history and social science §3).

Psychological concepts are intentional and, correspondingly, the moral sciences are interpretative. Can laws of individual behaviour be formulated, as Mill assumed, in this interpretative vocabulary? His analysis of the moral sciences takes their fundamental laws to lie in the domain of psychology. He was familiar with a different view, that of Auguste Comte, who held that the fundamental and irreducible moral science was sociology (a term coined by Comte). There was no deeper moral science, no science of psychology; the next level below sociology was the physical science of biology. Mill rejected that view, but enthusiastically shared Comte's vision of a historical sociology. Psychology may be the irreducible theoretical basis of the moral sciences; historical sociology is to be, as far as Mill was concerned, their prime exhibit.

Associationism and a Comtean historical sociology are thus the driving ideas in Mill's logic of the moral sciences. They interlock. Associationism fortifies his belief in the mutability of human nature: different social and historical formations can build radically different patterns of association. The bridge between historical sociology and the invariant laws of associationist psychology can be provided, Mill thinks, by an innovation of his own: a science he calls 'Ethology', which will study the different forms of human character in different social formations. He intended to write a treatise on the subject; significantly, he failed.

How, on this naturalistic view of mind and society, can human beings be free? The question mattered deeply to Mill. The conclusion others drew from the doctrine of determinism, namely, that we have (in Mill's phrase) no 'power of *self*-formation', and hence are not really responsible for our character or our actions, would have destroyed his moral vision. Self-formation is the fulcrum of his ideal of life, and 'moral freedom', the ability to bring one's desires under the control of a steady rational purpose, is a condition of self-formation, of having a character in the full sense (see Free will §§3-4).

Thus Mill had to show how causally conditioned natural objects can also be morally free agents. The sketch of a solution in the *System of Logic* (Book 6, chapter 2), which Mill thought the best chapter in the book, is brief but penetrating. (There is a longer discussion in *An Examination of Sir William Hamilton's Philosophy*, chapter 26.) One of its leading features is a distinction between resistible and irresistible causes; 'in common use', only causes which are 'irresistible', whose operation is 'supposed too powerful to be counteracted at all', are called necessary:

> There are physical sequences which we call necessary, as death for want of food or air; there are others which, though as much cases of causation as the former, are not said to be necessary, as death from poison, which an antidote, or the use of the stomach-pump, will sometimes avert…human actions are in this last predicament: they are never (except in some cases of mania) ruled by any one motive with such absolute sway, that there is no room for the influence of another.

(1843: 839)

An action caused by an irresistible motive (a 'mania') is plainly not free. This is certainly very pertinent. Yet something is added when we move from the idea of motives being resistible by other motives to the idea of moral freedom, the idea that *I* have the power to resist motives. It is the ability to recognize and respond to reasons. I act freely if I could have resisted the motive on which I in fact acted, had there been good reason to do so. A motive impairs my moral freedom if it cannot be defeated by a cogent reason for not acting on it. Mill fails to bring this connection between freedom and reason into clear view, but he relies on it in his ethical writings. He takes it that I am more or less free overall, according to the degree to which I can bring my motives under scrutiny and act on the result of that scrutiny. So I can *make* myself more free, by shaping desires or at least cultivating the strength of

will to overcome them.

> A person feels morally free who feels that his habits or his temptations are not his masters, but he theirs: who even in yielding to them knows that he could resist…we must feel that our wish, if not strong enough to alter our character, is strong enough to conquer our character when the two are brought into conflict in any particular case of conduct. And hence it is said with truth, that none but a person of confirmed virtue is completely free.
>
> (1843: 841)

The identification of moral freedom with confirmed virtue, and (less explicitly) of confirmed virtue with steady responsiveness to reasons, is present in Mill as in Kant (see Kant, I. §11). But Mill does not address crucial questions such as what is it to grasp a reason, or how reason can be efficacious. To vindicate the coherence of his view one would have to show how to answer such questions in a way which is compatible with naturalism. The problem remains central in contemporary philosophy - certainly Mill himself never took full stock of it.

8 Happiness, desire and will

Mill's single ultimate standard of theoretical reason is enumerative induction. His single ultimate standard of practical reason is the principle of utility; its standard is the good of all. But what is the good? According to Mill, it is happiness, understood as 'pleasure, and freedom from pain' (1861a: 210) (see Happiness). His case rests on the following principle of method:

> The sole evidence it is possible to produce that anything is desirable, is that people do actually desire it. If the end which the utilitarian doctrine proposes to itself were not, in theory and in practice, acknowledged to be an end, nothing could ever convince any person that it was so.
>
> (1861a: 234)

Mill is not claiming that the conclusion that happiness is desirable follows deductively from the premise that people in general desire it. He gives some ground for that misinterpretation when he compares the move from 'desired' to 'desirable' to those from 'seen' and 'heard' to 'visible' and 'audible'. Nevertheless, his procedure is simply an appeal to reflective practice, just as in the case of enumerative induction - where again the 'sole evidence' that enumerative induction is an ultimate norm of reasoning is that we acknowledge it as such 'in theory and in practice'.

However, a question which is appropriate by Mill's own principle of method is whether reflective practice shows that happiness is the *only* thing we desire. Do not human beings, in theory and in practice, desire things other than happiness? Mill anticipates this question and responds to it at length. He claims that when we want a particular object for its own sake and with no further end in view (let us say, when we have an underive*d* desire for it), then we desire it because we think of it as enjoyable: 'to desire anything, except in proportion as the idea of it is pleasant, is a physical and metaphysical impossibility' (1861a: 238). But this does not mean that we desire all objects as *means* to our pleasure. The desire for an object is genuinely a desire for that *object*; it is not the desire for pleasure as such. Mill's way of marking this is to say that the object is desired as a 'part' or an 'ingredient' of happiness, not as a means to it. His rejection of psychological egoism was one of the points on which he took himself to be at odds with Bentham (see Bentham, J. §3). When a person does something because they think it will be pleasant - for example a generous person who gives a present - it does not follow that they are acting selfishly (see Egoism and altruism). Generous people take pleasure in the prospect of giving, not in the prospect of getting pleasure; their desire to give is not derived from the desire to get pleasure. Giving is a part of their happiness, not a means to it.

Thus Mill's case for the claim that happiness is the sole human end, put more carefully, is this: 'Whatever is desired otherwise than as a means to some end beyond itself, and ultimately to happiness, is desired as itself a part of happiness, and is not desired for itself until has become so' (1861a: 237). Nothing here assumed Hume's view that every action must ultimately flow from an underived desire. That is a quite separate issue, and Mill's view of it is closer to that of Kant or Reid than to that of Hume. He insists 'positively and emphatically'

> that the will is a different thing from desire; that a person of confirmed virtue, or any other person whose purposes are fixed, carries out his purposes without any thought of the pleasure he has in contemplating them,

> or expects to derive from their fulfilment.
>
> (1861a: 238)

This distinction between purpose and desire is central to Mill's conception of the will. When we develop purposes we can will against mere likings or aversions: 'In the case of an habitual purpose, instead of willing the thing because we desire it, we often desire it only because we will it' (1861a: 238). Every action is caused by a motive, but not every motive is a liking or aversion:

> When the will is said to be determined by motives, a motive does not mean always, or solely, the anticipation of a pleasure or of a pain…. A habit of willing is commonly called a purpose; and among the causes of our volitions, and of the actions which flow from them, must be reckoned not only likings and aversions, but also purposes.
>
> (1843: 842)

The formation of purposes from desires is the evolution of will; it is also the development of character. Mill quotes Novalis: 'a character is a completely fashioned will' (1843: 843). Not that this reflects the whole of his view of character; character for him requires the cultivation of feeling as well as the cultivation of will: 'A person whose desires and impulses are his own - are the expression of his own nature, as it has been developed and modified by his own culture - is said to have a character' (1859: 264). Developed spontaneity of feeling is part of fully-perfected character, but certainly moral freedom is too - 'none but a person of confirmed virtue is completely free'. As noted in §7 above, Mill does not address the crucial question of what it is for a purpose to be informed by reason. Still, the distinction between purpose and desire does allow him to recognize conscientious action, action which flows not from any inclination but solely from a habit of willing; he asserts the possibility and value of a 'confirmed will to do right' (1861a: 238), distinct from motives of anticipated pleasure and pain. That 'virtuous will', however, is not for him an intrinsic good, as it is for Kant. It is

> a means to good, not intrinsically a good; and does not contradict the doctrine that nothing is good to human beings but in so far as it is either itself pleasurable, or a means of attaining pleasure or averting pain.
>
> (1861a: 239)

9 Qualities of pleasure

Happiness - pleasure and the absence of pain - is the sole final end of life. But Mill's idea of it is altogether more romantic and liberal than that of earlier utilitarians. He takes into account the fact that a variety of notions - for example, purity, elevation, depth, refinement and sublimity, and their opposites - enter into our assessments of pleasure. We do not assess pleasures along a single dimension. In his general ethical and political writing, Mill freely draws on that extensive and flexible language. He sees the need to recognize it also in utilitarian theory, but here he does so rather more mechanically by distinguishing 'quality' and 'quantity' of pleasure. From the first publication of *Utilitarianism*, at least three sorts of question have been asked about this famous distinction. The first is whether it is reconcilable with hedonism. The second is epistemological: is there a cogent way of establishing that some pleasures are superior in 'quality'? The third question, perhaps the most challenging, though less often discussed, is how the distinction fits into the framework of utilitarianism.

As to the first question: there is indeed, as Mill says, no reason in logic why more than one characteristic of pleasures should not be relevant to estimating their value - though if we call those characteristics 'quantity' and 'quality', we need to maintain a careful distinction between the quantity and quality of a pleasure on the one hand and its degree of value on the other. All that hedonism requires is that the only things that make a pleasure valuable are its characteristics as a *pleasure* (see Hedonism).

Nevertheless, an impression lingers that Mill's discussion appeals to intuitions which are not hedonistic. For example:

> Few human creatures would consent to be changed into any of the lower animals, for a promise of the fullest allowance of a beast's pleasures; no intelligent human being would consent to be a fool, no instructed person would be an ignoramus, no person of feeling and conscience would be selfish and base, even though they should be persuaded that the fool, the dunce, or the rascal is better satisfied with his lot than they are with

theirs.

(1861a: 211)

He also notes that a 'being of higher faculties requires more to make him happy, is capable probably of more acute suffering, and is certainly accessible to it at more points, than one of an inferior type' (1861a: 212). So a being of higher faculties may be faced with a choice: on the one hand a life of acute suffering, with no access to any of the higher pleasures which its faculties make it capable of appreciating, on the other, a cure (for example, an operation) which relieves its suffering but leaves it only with the pleasures available to a fool or a dunce. Is Mill saying that in *all* such cases the life of suffering should be preferred? He does not say so explicitly and if he does adhere to hedonism he should not. For cases are surely possible in which life after the cure offers a stream of pleasures more valuable overall, taking quality as well as quantity into account, than the life of suffering in which one retains one's higher faculties but is bereft of higher pleasures.

What of the epistemological question? Mill compares assessments of the comparative quality of pleasures to assessments of their comparative quantity: both are determined by 'the feelings and judgments of the experienced' (1861a: 213). But a judgment that the pleasure derived from film A is of a higher kind than that derived from watching film B is clearly, as Mill conceives it, an evaluative judgment. The proper comparison would have been with the evaluative judgment that pleasure as such is desirable. And Mill could have said that with this judgment, as with basic evaluative judgments in general, the only criterion is reflective practice - self-examination and discussion. In such discussion, some people emerge as better judges than others - this is not a circularity but an inherent feature of normative judgment.

Yet now the third question becomes pressing: how are such judgments of the quality of pleasures to be registered in the utilitarian calculus? In requiring utilitarianism to take them into account Mill makes a move of political as well as ethical significance. For what rank do we give to these pleasures in our social ordering - the rank which highly developed human natures attach to them or that which lower human natures attach to them? Mill's answer is unambiguous: it is the verdict of 'competent judges' which stands.

Suppose that beings of highly developed faculties place the pleasures of scientific discovery or artistic creation so much higher than those of material well-being that (above a certain modicum of physical comfort and security) any amount of the former, however small, is ranked by them above any amount of the latter, however large. Suppose, however, that beings of considerably less developed faculties would not share this assessment. And now suppose that the question is put to Mill, how much of the lower pleasure of the less developed being may be sacrificed to maintain the more highly developed being's higher pleasure? Mill's view is that the more highly developed being delivers the correct assessment of the relative value of the higher and lower pleasures. But, by hypothesis, it would be prepared to sacrifice any amount of the lower pleasure, down to a modicum of physical comfort and security, for the smallest amount of the higher. Must the same hold for the interpersonal case? Must it be correct for the utilitarian to sacrifice any amount of the lower pleasures of lower beings, down to a level at which they are provided with the modicum of comfort and security, in order to secure some higher pleasure for a higher being? Mill provides no answer.

10 The utility principle

Though Mill deepened the utilitarian understanding of pleasure, desire, character and will, he never adequately re-examined the principle of utility itself. When he states the utilitarian doctrine before considering what kind of proof can be given of it, he states it thus: 'Happiness is desirable, and the only thing desirable, as an end, all other things being only desirable as means to that end' (1861a: 210). In effect, he takes his task to be that of demonstrating the truth of hedonism. All he has to say about the move from hedonism to the utility principle is that if 'each person's happiness is a good to that person' then 'the general happiness' must be 'a good to the aggregate of all persons'. In a letter in which he explains this unclear remark, he says: 'I merely meant in this particular sentence to argue that since A's happiness is a good, B's a good, C's a good, etc, the sum of all these goods must be a good' (1972: 1414). This contains two inexplicit assumptions. The more obvious point is that an egoist may accept that Mill has shown that 'each person's happiness is a good to that person', but deny that he has shown that happiness is a good tout *court*. The egoist denies that Mill has shown that *everyone* has reason to promote the happiness of anyone. That requires a separate postulate, as Henry Sidgwick pointed out.

The second inexplicit assumption is more subtle. At the end of the last chapter of *Utilitarianism*, 'On the Connexion between Justice and Utility', Mill does explain that he takes 'perfect impartiality between persons' to be part of the very meaning of the Greatest Happiness Principle:

> That principle is a mere form of words without rational signification, unless one person's happiness, supposed equal in degree (with the proper allowance made for kind), is counted for exactly as much as another's. Those conditions being supplied, Bentham's dictum, 'everybody to count for one, nobody for more than one', might be written under the principle of utility as an explanatory commentary.
>
> (1861a: 257)

So here Mill supplies the required postulate of impartiality. However, the concept of impartiality does not, on its own, yield utilitarianism's aggregative principle of distribution. Maximizing the *sum* of individuals' happiness, if it makes sense to talk in this way at all, is one way of being impartial: no individual's happiness is given greater weight than any other's in the procedure which determines the value of a state of affairs as a function of the happiness of individuals in that state of affairs. In this sense the procedure implements the principle, 'Everybody to count for one, nobody for more than one'; but so does maximizing the average of all individuals' unweighted happiness. Here too all individuals count for one and no more than one. In fact a wide variety of non-equivalent distributive principles is impartial in this way. The most one could get from combining a postulate of impartiality with hedonism is that ethical value is a positive impartial function of individual happiness and of nothing else. In a footnote to the paragraph Mill glosses the requirement of perfect impartiality as follows: 'equal amounts of happiness are equally desirable, whether felt by the same or by different persons'. That does yield aggregative or average utilitarianism, but it follows neither from the thesis that happiness is the only thing desirable to human beings, nor from the formal notion of impartiality (see Impartiality §2).

11 Morality and justice

When we turn to Mill's conception of the relationship between the utility principle and the fabric of principles which regulate everyday social life, we find him again at his most impressive. He stresses that a utilitarian standard of value cannot itself tell what practical rules, aims or ideals we should live by. In his autobiography he dates this conviction to the period of his mental crisis. He now 'gave its proper place, among the prime necessities of human well-being, to the internal culture of the individual' (1873: 145-7). The prime task for human beings was to attend to that internal culture - to develop whatever was best in themselves. The indirect role in which he now cast the utility principle became a fundamental structural feature of his moral and political philosophy. For example, he accuses Auguste Comte of committing:

> the error which is often, but falsely, charged against the whole class of utilitarian moralists; he required that the test of conduct should also be the exclusive motive of it…. M. Comte is a morality-intoxicated man. Every question with him is one of morality, and no motive but that of morality is permitted.
>
> (1865b: 335-6)

Mill gives a succinct statement of his own doctrine at the end of the *System of Logic*. As always, he affirms 'that the promotion of happiness is the ultimate principle of Teleology'. But, he continues,

> I do not mean to suggest that the promotion of happiness should be itself the end of all actions, or even of all rules of action. It is the justification, and ought to be the controller, of all ends, but is not itself the sole end…. I fully admit that…the cultivation of an ideal nobleness of will and conduct, should be to individual human beings an end, for which the specific pursuit either of their own happiness or of that of others (except so far as included in that idea) should, in any case of conflict, give way. But I hold that the very question, what constitutes this elevation of character, is itself to be decided by a reference to happiness as the standard.
>
> (1843: 952)

The happiness of all is 'the test of all rules of conduct' - and not only rules of conduct but also of cultivation of feelings. How is the test applied? Here Mill learned more from Coleridge (§2) than from Bentham; that is, from historical criticism directed at the abstract social visionaries of the enlightenment. They did not see that moral sentiments can only grow in a stable tradition and social setting. They did not grasp the conditions necessary for such a tradition and setting - education of personal impulses to a restraining discipline, shared allegiance to some

enduring and unquestioned values, 'a strong and active principle of cohesion' among 'members of the same community or state'. Hence

> They threw away the shell without preserving the kernel; and attempting to new-model society without the binding forces which hold society together, met with such success as might have been anticipated.
>
> (1840: 138)

This feeling for the historicity of social formations and genealogies of morals gives Mill's ethical vision a penetration which is absent from Bentham (and also from the excessively abstract discussions of utilitarianism in twentieth-century philosophy). On the other hand the analysis of morality, rights and justice which Mill fits into this ethical vision owes much to Bentham.

Mill examines the concept of justice in chapter 5 of *Utilitarianism*. Having observed that the idea of something which one may be constrained or compelled to do, on pain of penalty, is central to the idea of an obligation of justice, he notes that it nevertheless 'contains, as yet, nothing to distinguish that obligation from moral obligation in general':

> The idea of penal sanction, which is the essence of law, enters not only into the conception of injustice, but into that of any kind of wrong. We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience.
>
> (1865b: 246)

This is a normative, not a positive, account of morality: the morally wrong is that which *ought* to be punished, by law, social opinion or conscience. It would be a circular account if the 'ought' in question were itself a moral 'ought'. But the utility principle is the ultimate principle of 'Teleology'. Teleology is the 'Doctrine of Ends'; 'borrowing the language of the German metaphysicians', Mill also describes it as 'the principles of Practical Reason' (1843: 949-50). So the 'ought' is the 'ought' of Practical Reason - which, making appropriate use of 'laws of nature', produces the 'Art of Life'. Morality itself is only one department of this art. Moral concepts and judgments issue from the moral sentiments, the sentiments involved in guilt and blame; but are corrigible by a rational doctrine of ends. And that doctrine, in Mill's view, is the utility principle.

From this account of morality Mill moves to an account of rights and justice. A person has a moral right to a thing if there is a moral obligation on society to protect them in their possession of that thing. Obligations of justice are distinguished from moral obligations in general by the existence of corresponding rights:

> Justice implies something which it is not only right to do, and wrong not to do, but which some individual person can claim from us as his moral right…. Whenever there is a right, the case is one of justice.
>
> (1865b: 247)

Upholding rights is one of society's vital tasks. For on it depends our security - which is 'to every one's feelings the most vital of interests':

> This most indispensable of all necessaries, after physical nutriment, cannot be had, unless the machinery for providing it is kept unintermittedly in active play. Our notion, therefore, of the claim we have on our fellow-creatures to join in making safe for us the very groundwork of our existence, gathers feelings around it so much more intense than those concerned in any of the more common cases of utility, that the difference in degree (as is often the case in psychology) becomes a real difference in kind.
>
> (1865b: 251)

In this way the claim of justice comes to be felt as a claim of a higher kind than any claim of utility. Justice, Mill concludes,

> is a name for certain classes of moral rules, which concern the essentials of human well-being more nearly, and are therefore of more absolute obligation, than any other rules for the guidance of life.
>
> (1865b: 255)

Mill spells out in detail what these moral rules should be in his writings on various social questions. In *Utilitarianism*, he is concerned with the more abstract task of showing how justice-rights take priority over the direct pursuit of general utility by individuals or the state, just as they take priority over the private pursuit of personal ends. His position is thus more complex than that of philosophers in a Kantian tradition who assume, in John Rawls' phrase, that the right (or just) is prior to the good. For Mill, good is philosophically prior to right - but politically and socially right constrains the pursuit of good (see Justice §3).

12 Liberty and democracy

The most celebrated part of Mill's social philosophy, his essay *On Liberty*, must be read in terms of this conception of the right and the good. Mill is not a social contract or 'natural rights' liberal. He appeals instead to 'utility in the largest sense, grounded on the permanent interests of man as a progressive being' (1859: 224). He has in mind the higher human nature, capable of development by self-culture, which he believes to be present in every human being. Self-culture opens access to higher forms of human happiness, but it has to be *self*-culture, first because human potentialities are diverse and best known to each human being itself, and second because only when human beings work to their own plans of life do they develop moral freedom, itself indispensable to a higher human nature.

Given the importance free self-culture thus assumes in Mill's idea of human good, and the account of rights which has just been considered, it will follow that individual liberty must be a politically fundamental right. For self-development is one of 'the essentials of human well-being'. Thus Mill is led to the famous principle enunciated in *On Liberty*:

> the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right.
>
> (1859: 223-4)

A society which respects this principle enables individuals to realize their potential in their own way. It liberates a mature diversity of interest and feeling, and it nurtures the moral freedom of reason and will. Throwing open the gates to talent, creativity and dynamism, it produces the social conditions of moral and intellectual progress. This Millian argument remains the strongest defence of any liberalism founded on teleological ethics. It is a resource upon which teleological liberals will always be able to draw, whether or not they accept Mill's hedonistic conception of the human good or his aggregative conception of the good of all.

However, it is also connected with Mill's ambivalence about democracy. Like many other nineteenth-century thinkers, liberal as well as conservative, Mill felt a deep strain of anxiety about democratic institutions and the democratic spirit (see Democracy §2). Certainly he applauded the end of the *ancien régime* and sympathized with the moral ends of the French Revolution - liberty, equality, fraternity - but he learned from it, as had the continental liberals, to fear an enemy on the left, as well as an authoritarian enemy on the right. In its revolutionary form the enemy on the left threatened Jacobin terror, or the disasters which attend any attempt to achieve moral ideals by restarting history at year zero. Its settled form, on the other hand, could be observed in the 'democratic republic' of America: a continuous and unremitting pressure towards conforming mediocrity.

The Romantic-Hellenic ideal of human life both inspired Mill's democratic ideals and fuelled his fears about realized democracy. It was an ideal he shared with left Hegelians like Marx, who experienced less difficulty in combining it with democratic egalitarianism. Mill too had a long-term vision in which the emancipation and education of the working class could bring free self-culture to all human beings. He was able to believe, on the basis of his associationist psychology, that all human beings have an equal potential to develop their higher faculties. This warded off the possibility that utilitarianism might recommend an extremely inegalitarian pursuit of higher forms of well-being as the equilibrium state of a fully-developed human society.

Thus Mill remained more of a democrat than other liberals of the nineteenth century, such as de Tocqueville or Burckhardt, but like them he saw how moral and cultural excellence and freedom of spirit could be endangered by mass democracy. Like them, his attitude to the immediate prospect of democratic politics was decidedly mixed.

What he wanted was a democratic society of freely developed human beings; he did not think it a proximate or certain prospect, and he thought that bad forms of democracy could themselves pose a threat to it by drifting into ‘collective despotism’ - a danger to which America had already succumbed.

His advice for warding off this threat was not less democracy but more liberty:

> If the American form of democracy overtakes us first, the majority will no more relax their despotism than a single despot would. But our only chance is to come forward as Liberals, carrying out the Democratic idea, not as Conservatives, resisting it.
>
> (1972: 672)

This was the importance of the essay on Liberty, and particularly of the defence of liberty of thought and discussion contained therein. Nor were freedom of speech and liberty of the individual the only instruments by which Mill hoped to steer away from bad forms of democracy towards good. Some of his recommendations - plural voting, a public ballot, a franchise restricted by educational qualification - may now seem misguided or even quaint. Others, including proportional representation of minorities and, not least, his life-long advocacy of equal rights for women, make him seem ahead of his time. At any rate, in political philosophy from Plato’s *Republic* to the present day, Mill’s discussion of democracy has few rivals - for its open-mindedness, its historical and psychological awareness, and its underlying ethical power.

See also: Causation §3; Consequentialism; Economics, philosophy of §3; Empiricism; Feminism §3; Freedom and liberty; Good, theories of the; Induction, epistemic issues in; Inductive inference; Liberalism; Nishi Amane; Utilitarianism

JOHN SKORUPSKI

List of works

Mill, J.S. (1991) *Collected Works of John Stuart Mill*, ed. J.M. Robson, London: Routledge and Toronto, Ont.: University of Toronto Press.(The standard edition of Mill’s writings, in thirty-three volumes. The introductions are invariably worth reading. Volume and page numbers given below refer to this edition.)

Mill, J.S. (1838) *Bentham*, in *Collected Works of John Stuart Mill*, London: Routledge, vol. 10, 75-115, 1991. (Mill’s radical assessment of Bentham, usually read in conjunction with the essay on Coleridge.)

Mill, J.S. (1840) *Coleridge*, in *Collected Works of John Stuart Mill*, London: Routledge, vol. 10, 117-63, 1991. (These two essays most accessibly illustrate how Mill wove together Enlightenment and Romanticism.)

Mill, J.S. (1843) *System of Logic: Ratiocinative and Inductive*, in *Collected Works of John Stuart Mill*, London: Routledge, vols 7 and 8, 1991.(The nineteenth century’s most penetrating exposition of a naturalistic philosophy of logic and science, including social science. The pagination of volumes 7 and 8 is consecutive; volume 8 begins at page 639.)

Mill, J.S. (1848) *Principles of Political Economy*, in *Collected Works of John Stuart Mill*, London: Routledge, vols 2 and 3, 1991.(A synthesis of classical economics, this work also contains much interesting social philosophy.)

Mill, J.S. (1859) *On Liberty*, in *Collected Works of John Stuart Mill*, London: Routledge, vol. 18, 213-310, 1991. (One of liberalism’s canonical texts.)

Mill, J.S. (1861a) *Utilitarianism*, in *Collected Works of John Stuart Mill*, London: Routledge, vol. 10, 203-59, 1991.(A central text of moral philosophy - its extraordinary succinctness means that much in it remains to be fully charted and quarried.)

Mill, J.S. (1861b) *Considerations on Representative Government*, in *Collected Works of John Stuart Mill*, London: Routledge, vol. 29, 371-577, 1991.(One-man report on the prospects of democracy.)

Mill, J.S. (1865a) *An Examination of Sir William Hamilton’s Philosophy*, in *Collected Works of John Stuart Mill*, London: Routledge, vol. 9, 1991.(Mill’s main treatment of metaphysical issues, including the nature of mind and matter, free will, logic and thought.)

Mill, J.S. (1865b) *Auguste Comte and Positivism*, in *Collected Works of John Stuart Mill*, London: Routledge, vol. 10, 261-368, 1991.(An assessment, from Mill’s later years, of a philosopher who greatly influenced him in his youth.)

Mill, J.S. (1869) *The Subjection of Women*, in *Collected Works of John Stuart Mill*, London: Routledge, vol. 21, 259-340, 1991.(Manifesto of nineteenth-century liberal feminism.)

Mill, J.S. (1873) *Autobiography*, in *Collected Works of John Stuart Mill*, London: Routledge, vol. 1, 1-290, 1991.

(Famous account of Mill's early education, 'mental crisis', and subsequent intellectual and moral projects.)
Mill, J.S. (1874) *Three Essays on Religion*, in *Collected Works of John Stuart Mill*, London: Routledge, vol. 10, 369-489, 1991.(Posthumously published articles, in which Mill argues that religious hope is legitimate, whereas religious belief is not.)
Mill, J.S. (1879) *Chapters on Socialism*, in *Collected Works of John Stuart Mill* , London: Routledge, vol. 5, 703-53, 1991.(Posthumously published assessment of socialism which complements the *Principles of Political Economy* and *On Liberty*.)
Mill, J.S. (1972) *Later Letters*, in *Collected Works of John Stuart Mill*, London: Routledge, vols 14-17, 1991. (Mill's extensive correspondence after 1848, published in four volumes.)

References and further reading

Berger, F.R. (1984) *Happiness, Justice and Freedom: The Moral and Political Philosophy of John Stuart Mill*, London: University of California Press.(The most comprehensive study of Mill's moral and political philosophy; full of information.)
Crisp, R. (1997) *Mill on Utilitarianism*, London: Routledge.(The most up-to-date study of *Utilitarianism*.)
Gray, J. (1996) *Mill on Liberty: a Defence*, London: Routledge, 2nd edn.(Relates Mill's approach to liberty to his indirect utilitarianism.)
Kahan, A.S. (1992) *Aristocratic Liberalism. The Social and Political Thought of Jacob Burckhardt, John Stuart Mill, and Alexis de Tocqueville*, Oxford: Oxford University Press.(Informative comparative study of the political thought of these three thinkers.)
Lyons, D. (1994) *Rights, Welfare, and Mill's Moral Theory*, Oxford: Oxford University Press.(Important essays on the interpretation of Mill's ethical theory.)
Mackie, J.L. (1974) *The Cement of the Universe*, Oxford: Oxford University Press. (Analyses causation in the spirit of Mill's treatment in the *System of Logic*; the appendix comprehensively surveys and revises his 'Methods of Empirical Inquiry'.)
Ryan, A. (1974) *J.S. Mill*, London: Routledge & Kegan Paul.(Useful survey, not just of Mill's philosophy but of his thought as a whole.)
Scarre, G. (1989) *Logic and Reality in the Philosophy of John Stuart Mill*, Dordrecht: Kluwer.(Study of Mill's metaphysics and philosophy of science.)
Skorupski, J. (1989) *John Stuart Mill*, London: Routledge.(Comprehensive account of Mill's philosophy.)
(ed.) (forthcoming) *The Cambridge Companion to Mill*, Cambridge: Cambridge University Press.(Articles covering various aspects of Mill's philosophy.)
Ten, C.L. (1980) *Mill on Liberty*, Oxford: Oxford University Press.(The most useful student guide to *On Liberty*.)
Whewell, W. (1837) *History of the Inductive Sciences, from the Earliest to the Present Time*, London: J.W. Parker, 3 vols.(Major historical survey, used by Mill.)
Wilson, F. (1990) *Psychological Analysis and the Philosophy of John Stuart Mill*, Toronto, Ont.: University of Toronto Press.(Mill and the history of psychology.)

Millar, John (1735-1801)

John Millar elevated law teaching from mere instruction in technicalities to the level of a genuinely liberal subject, largely by using his teaching to educate students in the science of legislation. A pupil and disciple of Adam Smith, Millar, through his teaching and writing, established his reputation as a key figure in the Scottish Enlightenment movement.

Millar was born on 22 June 1735, a son of the manse of Kirk o'Shotts in Lanarkshire, just as the first signs of 'the great awakening' in Scotland were materializing. He entered Glasgow University in 1746, leaving in 1752 to pursue a legal career. In 1761, at the age of 26, Millar was elected to the Regius Chair of Civil Law in Glasgow and there he taught until his death in 1801.

He was not a prolific writer. There are only two works which can definitively be attributed to him. The first, *The Origin of the Distinction of Ranks* (1771), is the book on which his reputation primarily rests. In *Ranks* Millar broke away from a traditional cyclical view of progress in which the conduct of politics was rooted in a dialectic between virtue and corruption, and presented the development of civil society as a process of unilinear advance. Millar's account constitutes a succinct statement of historical materialism: there are four stages - hunting, pasture, agriculture and commerce - through which societies pass and each form of economic organization gives rise to 'correspondent habits, dispositions and ways of thinking'. Millar argued that this was a natural progression - from ignorance to knowledge, from rude to civilized manners - and, though the general method was not novel, he seems to have been the first to have indicated the causes of growth through each of the sequential phases.

The other book, *An Historical View of the English Government* (1787), is a constitutional history from Saxon times to the Stuarts. Millar here utilizes the method employed in *Ranks*, mainly to challenge the vulgar Whiggism of writers such as William Blackstone who tended to view English constitutional history in terms of a gradual restoration of the ancient Anglo-Saxon constitution which had been undermined by the Normans. Millar, presenting a 'scientific Whig' account, argued that the 'low ranks' secured their liberty not by appeals to a romantic idea of the ancient constitution and not by pieces of paper but by industry and material progress. The origin of liberty in Britain is modern, not ancient; it was a seventeenth-century endeavour culminating in 1688.

Millar's general theoretical approach was historical and sociological rather than conceptual. He rejected any idea that law is determined by higher norms of divine origin or theories of law rooted in social contract. In this sense, he adopted a modern, sociological view that any theory of law must be rooted in a theory of society. Assessments of Millar's originality vary, but it is undoubtedly the case that he made a significant contribution to a remarkable period of, not only Scottish, but European intellectual history.

See also: Enlightenment, Scottish; Jurisprudence, historical §1; Law, philosophy of; Smith, A. §4; Social theory and law

MARTIN LOUGHLIN

List of works

Millar, J. (1771) *The Origin of the Distinction of Ranks: or An Inquiry into the Circumstances which give rise to Influence and Authority, in the Different Members of Society*, Edinburgh: William Blackwood; 4th edn, London: Longman, 1806; repr. London: Thoemmes Press, 1990.(The 1990 edition is prefixed by the 1806 John Craig account of the life and writings of John Millar.)

Millar, J. (1787) *An Historical View of the English Government from the Settlement of the Saxons in Britain to the Revolution in 1688*, Edinburgh: William Blackwood; 3rd edn, 4 vols, London: Longman, 1803.(Constitutional history from Saxon times to the Stuarts.)

Anon. (1796) *Letters of Crito*, Edinburgh: William Blackwood.(Collection of newspaper articles; almost certainly the work of Millar.)

Anon. (1796) *Letters of Sidney*, Edinburgh: William Blackwood.(Occasionally ascribed to Millar but more likely to be the work of John Craig, his nephew and early biographer.)

References and further reading

Forbes, D. (1953-4) ‘Scientific Whiggism: Adam Smith and John Millar’, *Cambridge Journal* 7: 643-70.(Millar as scientific Whig.)

Haakonssen, K. (1985) ‘John Millar and the Science of a Legislator’, *Juridical Review* 30: 41-68.(On Millar’s theory of law, government and property.)

Ignatieff, M. (1983) ‘John Millar and Individualism’, in I. Hont and M. Ignatieff (eds), *Wealth and Virtue: The Shaping of Political Economy in the Scottish Enlightenment*, Cambridge: Cambridge University Press.(On Millar caught between the languages of civic humanism and political economy.)

Lehmann, W.C. (1960) *John Millar of Glasgow*, Cambridge: Cambridge University Press. (The standard biography.)

Mīmāṃsā

The school of Mīmāṃsā or Pūrva Mīmāṃsā was one of the six systems of classical Hindu philosophy. It grew out of the Indian science of exegesis and was primarily concerned with defending the way of life defined by the ancient scripture of Hinduism, the Veda. Its most important exponents, Śabarasvāmin, Prabhākara and Kumārila, lived in the sixth and seventh centuries AD. *It was realist and empiricist in orientation. Its central doctrine was that the Veda is the sole means of knowledge of* dharma *or righteousness, because it is eternal. All cognition, it held, is valid unless its cause is defective. The Veda being without any fallible author, human or divine, the cognitions to which it gives rise must be true. The Veda must be authorless because there is no recollection of an author or any other evidence of its having been composed; we only observe that it has been handed down from generation to generation. Mīmāṃsā thinkers also defended various metaphysical ideas implied by the Veda - in particular, the reality of the physical world and the immortality of the soul. However, they denied the existence of God as creator of the world and author of scripture. The eternality of the Veda implies the eternality of language in general. Words and the letters that constitute them are eternal and ubiquitous; it is only their particular manifestations, caused by articulations of the vocal organs, that are restricted to certain times and places. The meanings of words, being universals, are eternal as well. Finally, the relation between word and meaning is also eternal. Every word has an inherent capacity to indicate its meaning. Words could not be expressive of certain meanings as a result of artificial conventions.*

The basic orientation of Mīmāṃsā was pragmatic and anti-mystical. It believed that happiness and salvation result just from carrying out the prescriptions of the Veda, not from the practice of yoga or insight into the One. It criticized particularly sharply other scriptural traditions (Buddhism and Jainism) that claimed to have originated from omniscient preceptors.

1 History

Mīmāṃsā, one of the six systems (*darśanas*) of Hindu philosophy, was not originally a system of philosophy at all but an auxiliary of Vedic study (see Hindu philosophy §§6-7). Mīmāṃsā literally means 'investigation' or 'analysis'; it refers specifically to the analytic discussion of Vedic ritual. (The Veda is the most ancient scripture of the Hindus.) The *Mīmāṃsāsūtra*, the earliest treatise of the school, which is attributed to Jaimini and dated around the second century BC, is a compendium of debates about how rituals should be carried out; it resolves the debates by carefully analysing the language of the scriptural passages in which the rituals are enjoined. Thus, the school of Mīmāṃsā was first and foremost concerned with exegesis. It developed a system of rules of interpretation that were widely employed elsewhere in Indian philosophical and scientific literature; they were especially used in interpreting and expounding the Dharmaśāstras, works relating to Hindu ethics and law.

At an early stage, Mīmāṃsā took on the broader role of defending Hindu *dharma* as a whole. *Dharma* means what one ought to do and avoid doing, the proper way to conduct oneself in this world and to attain salvation in the next. It includes not just the carrying out of certain rituals but also the observance of the minute details of everyday custom and etiquette. The Brahmanical or priestly conception of *dharma*, which was central to the culture that called itself 'Aryan' in ancient India and which became the core of classical Hinduism, was attacked by various heterodox movements, especially Buddhism and Jainism, from around the fifth century BC. Mīmāṃsā undertook to defend the Brahmanical system by attempting to demonstrate, in particular, the sole authority of the Veda with regard to matters of religion and ethics, and the falsehood of other scriptures. This apologetic enterprise led into the consideration of various philosophical issues, including metaphysical doctrines implied by the Veda and epistemological and linguistic doctrines upon which the claim of the authority of the Veda was based.

As in the case of the other Hindu philosophical systems, Mīmāṃsā was expounded chiefly in commentaries and subcommentaries on a foundational *sūtra* text (a treatise in the form of short statements or aphorisms), namely the *Mīmāṃsāsūtra*. The *Mīmāṃsāsūtra* originally contained material relating both to the interpretation of the ritual portions of the Veda and to the mystical-philosophical texts known as the Upaniṣads. The *sūtras* relating to the latter, however, were at an early date separated off and, known as the Vedānta or *Brahmasūtra*, became the focus of a distinct tradition of philosophical reflection, namely Vedānta. The oldest existing commentary on the Pūrva or 'prior' *Mīmāṃsāsūtra* - the *sūtras* concerned with ritual - is that of Śabarasvāmin, the *Śābarabhāṣya*, dated around

the sixth century AD. Other, older commentaries, such as that of Upavarṣa, have not been preserved. Two principal subschools of Mīmāṃsā thought, the Bhāṭṭa and the Prābhākara, are based on subcommentaries on the *Śābarabhāṣya* by Kumārila (also known as Kumārilabhaṭṭa) and Prabhākara respectively. A trio of commentaries by Kumārila, the *Ślokavārttika*, *Tantravārttika* and *Ṭupṭīkā*, have been preserved; Prabhākara's commentary is titled the *Bṛhatī*. Both authors may be assigned to the seventh century. Another great figure of the early period was Maṇḍanamiśra, who was sympathetic to the views of Kumārila but also composed works on Advaita Vedānta. Other important figures of the Bhāṭṭa school were Umbekabhaṭṭa (eighth century), Pārthasārathimiśra (tenth century), Sucaritamiśra (tenth century), Someśvarabhaṭṭa (twelfth century) and Khaṇḍadeva (seventeenth century). The principal exponent of the Prābhākara school after Prabhākara himself was Śālikanāthamiśra (ninth century). A third school of Mīmāṃsā, which, unlike the other two, accepted the existence of God - the so-called Seśvara Mīmāṃsā- was established by Murārimiśra in about the eleventh century. The following exposition is based primarily on the teachings of the more influential Bhāṭṭa school, although occasional references will be made to distinctive Prābhākara positions.

2 Knowledge and reality

The Mīmāṃsāclaim of the sole authority of the Veda with regard to *dharma* is based on two highly controversial doctrines: the eternality of the Veda and the intrinsic validity of cognitions. According to Mīmāṃsā, any error in cognition derives from defects in the causes of the cognition. A conch shell, which is in fact white, appears yellow to someone whose eye is affected by jaundice; a notion derived from verbal testimony can be false if the witness is dishonest or incompetent. On the other hand, a cognition whose causes are not defective will necessarily be true. Thus, cognitions arising from the Veda must be true, because the Veda is eternal - it was never composed by anyone but has just been handed down from time immemorial, from generation to generation - hence they cannot have a defective cause. They are not in any way contingent on the knowledge and honesty of a fallible author.

Kumārila developed the theory of 'intrinsic validity' (*svataḥ prāmāṇya*) as a general epistemological theory. *All* cognitions, he argued, present themselves initially as true; they arise together with a sense of conviction that matters really are as they indicate. Cognitions are deemed false only when they are directly overturned by other cognitions or undermined by the discovery that their causes are defective. Cognitions of Vedic injunctions, indeed, are never effectively contradicted by other means of knowledge such as perception, inference or testimony, and, as stated, there can be no suspicion that their source is defective, for the Veda is eternal. Prabhākara held not just that all cognitions initially *appear* true but that all cognitions, strictly speaking, *are* true; there is no such thing as error. Every cognition apprehends something real. It is only our judgments, which combine cognitions together, that are mistaken. If I think 'This conch shell is yellow', I am not in fact misperceiving anything. I correctly perceive the conch as 'this', and I correctly perceive the yellow colour, which is due to the jaundice afflicting my eye. It is only the attribution of the yellow to the conch instead of my eye that is wrong. The theory of intrinsic validity evoked a violent reaction from other Hindu, Buddhist and Jaina philosophers, who argued that a cognition can be considered valid only when it is confirmed by another cognition, in particular, a cognition of the 'causal efficacy' (*arthakriyā*) of the object (see Epistemology, Indian schools of §3).

The Mīmāṃsā theory of intrinsic validity was conducive to the adoption of realist positions on a variety of metaphysical issues. Kumārila, along with most other Hindu thinkers, rejected the idealism of the Yogācāra school of Buddhism (see Buddhism, Yogācāra school of §§2-4). His principal argument for realism was simply that the perception of the world as comprising externally existing, physical objects must be considered valid unless and until it is contradicted by other perceptions, in the way dreaming is contradicted by waking consciousness. He also charged that the Buddhist view of experience as comprising self-luminous moments of consciousness - that is, cognitions that reveal, not external objects, but their own inherent forms - violates the principle that the same thing cannot function as both agent and object in the same act (in this case, an act of cognition). More crucially, Kumārila held that we are not aware of cognitions (*jñāna*) at all in experience; we do not, at the time we perceive an object, also perceive a cognition. Our awareness, rather, is completely absorbed in being aware of the object. Hence, there is no basis for suggesting that object and cognition are the same thing, for they never occur together. Kumārila, then, was a direct realist. He believed that we only infer the existence of a cognition (*jñāna*) as the cause of the fact that we have become aware of an external object.

The Prābhākara school, on the other hand, held, like the Buddhists, that consciousness (*saṃvitti*) is self-luminous;

it simultaneously reveals the object, itself, and the knower. Yet it reveals itself always as the conscious principle, never as the object, hence, contrary to the Buddhist teaching, experience itself shows us that object and consciousness are distinct. At the same time, 'cognition' (*jñāna*), in the sense of the act of mind which produces consciousness, is only inferred.

Not only does perception reveal an external world, it reveals its nature. The cognitions we have of objects as endowed with certain properties and being of certain types - as a white man, a walking cow, and so on - are true unless and until overturned by other cognitions. Moreover, such cognitions are perceptual and can be referred to as 'conceptualized perceptions' (*savikalpaka pratyakṣa*); they are caused by the contact of our sense organs with objects, though they are preceded by a moment of nonconceptualized perception (*nirvikalpaka pratyakṣa*) in which the various features of the object are not identified as particular *types* of features or named. Buddhist philosophers, on the other hand, held that conceptualized cognition is not perceptual but mainly involves an act of the imagination by which mentally constructed properties are projected onto things; it is a kind of illusion. Kumārila argued - once again appealing to the principle of intrinsic validity - that the enduring validity of most of our cognitions of objects as qualified by qualities, actions, universals, and so on suggests that in general they are not errors; if they were, they would always be overturned. The specific Buddhist charge that a conceptualized cognition involves the superimposition of a mere *word* upon the object is rejected on the grounds that in fact we first see what a thing is before we apply a word to it. The world is perceived as it is prior to language, and we speak about it in a certain way because of the way we perceive it.

The above teaching implies an acceptance of the reality of universals. Mīmāṃsā philosophers were at the forefront of the debate with the Buddhists over universals. While the Buddhists were nominalists, Hindu and Jaina philosophers held that universals must be posited to account for common notions and common terms, that is, the fact that we cognize various things as being of the same type and refer to them by the same word. Kumārila in particular developed trenchant criticisms of the Buddhist theory that common concepts are just mentally constructed 'exclusion classes', the so-called *apohavāda* (see Universals, Indian theories of §2). The Bhāṭṭa view of universals was somewhat different from that of other realist schools, however, in that it considered universal and particular to be different aspects of one thing and not entirely distinct; the same thing is universal from one angle, particular from another. (It developed a similar view of nonbeing, *abhāva*: a thing is both what it is and what it is not. When we judge that a pot is *not* a cloth, we are apprehending a real, negative aspect of the pot. This tendency of Bhāṭṭa thought is believed by some to reflect the influence of Jainism.) Kumārila criticized the Nyāya-Vaiśeṣika notion that universal and particular are externally related to each other by virtue of 'inherence' (*samavāya*) (see Nyāya-Vaiśeṣika §5). The Prābhākara school, meanwhile, adopted the more conventional stance that universals are distinct from the individuals in which they inhere, but it worked out a more felicitous theory of inherence.

Mīmāṃsā philosophers - again in concert with other Hindu thinkers and against the Buddhists - defended the existence of the soul. An immortal soul is clearly implied by Vedic injunctions of certain sacrifices for the attainment of heaven. Kumārila argued that the existence of an unchanging self is established by the fact that I recognize *myself* as the subject of remembered experiences, hence I must have existed continuously from the past to the present. He also held the self to be essentially self-conscious; the Prābhākara, on the other hand, held that the self becomes manifest only in empirical consciousness and has no consciousness of its own.

Although an apologist for Vedic tradition, the Mīmāṃsaka (Mīmāṃsā philosopher) rejected a supreme being as creator of the universe and author of the Veda. This accords with his staunch empiricism and his belief that the positing of any author of scripture, even a divine one, tends to call it into question. Kumārila extensively refuted the Nyāya arguments for the existence of God as creator. What could his purpose have been in creating the universe? Not the fulfilment of the past karma of living beings, because living beings did not then exist; not 'sport' or amusement, for that would imply some lack in God. Yet if he did it without desire or purpose, he could not be considered intelligent, for no intelligent person acts without a purpose. And so on. Kumārila even rejected the Hindu mythological idea of a periodic creation and destruction of the universe, asserting that we have no evidence that the world has ever been different from how it is now. References to gods in the Veda are not to be taken literally but just as eulogies that serve to recommend certain sacrifices. Heaven is just a region, whose location is indefinite, where one experiences 'pure happiness'. The soteriological goal of *mokṣa*, liberation from the cycle of rebirth, is negatively conceived, not as an experience of transcendental bliss but as the cessation of all

association with a body due to the exhaustion of karma (see Karma and rebirth, Indian conceptions of §§1-4). Thus, Mīmāṃsā is decidedly rationalistic in its interpretation of Vedic religion. The religious life, for Mīmāṃsā, consists not in the practice of yoga or the attainment of insight into the oneness of everything, but in the scrupulous observance of *dharma* as laid down in Vedic prohibitions and commandments.

3 Language and scripture

Mīmāṃsā insisted on the eternality of the Veda on the grounds that its composition has not been observed, nor is there any memory of its having been composed by anyone in the past. Rather, we observe simply that it has been transmitted from generation to generation. What sense, however, can there be in suggesting that there could be a discourse that was never composed by anyone? Language surely always expresses the thoughts of intelligent beings, and depends on the articulation, by living speakers and writers, of symbols that are conventionally linked with certain meanings. Precisely such a view of language is what Mīmāṃsā rejected; in its place, it defended the eternality (*nityatva*) of language in general.

Individual words themselves must be eternal, for we routinely recognize a word we use as ‘the same word’ we used before. We never utter new words but only the same ones over and over. Nor should it be thought that a word is perishable because it consists of momentarily heard sounds. Rather, articulated sounds (*dhvani*) serve only to evoke an awareness in the hearer of the eternal, ubiquitous word; they are not identical with the word itself. In particular, they are not its letters (*varṇa*) - Kumārila insisted, against the grammarian Bhartṛhari, that a word, though eternal, is not an indivisible entity (*sphoṭa*) but comprises a sequence of letters - for letters, too, are omnipresent and eternal. We also recognize a particular letter as ‘the same letter’ as before. It is cognized at a particular time and place due to an alteration in the ear brought about by a disturbance of the air caused by the vocal organs (see Language, Indian theories of §3). The meanings of words are also eternal, according to Mīmāṃsā, for they are universals.

But surely the conventions by which words are assigned certain meanings were fixed by intelligent beings at a certain time, in which case the Veda could not be altogether eternal? The Mīmāṃsaka denied even this; rather, he held that a word does not refer to its meaning by virtue of a convention but rather by an inherent capacity (*śakti*). Kumārila, in particular, argued that the connection of word and meaning cannot be a human construct. A uniform convention devised by humans could only be based on some underlying natural relationship, while diverse conventions observed by different users would render communication impossible. That uniform conventions were established by God at the beginning of time is precluded by doubts about the existence of God. Moreover, it would seem that the devising of conventions by anyone already presupposes the existence of language. Finally, in contrast to the custom with other conventional practices in ancient India, we do not recall the authors of particular conventions when we employ words. (The Mīmāṃsā school holds that in the performance of rituals, for example, we do always recall their authors.) All we observe, in fact, is that we learn what words mean from our elders, who in turn learned the meanings from their elders, and so on; we never observe anyone originally establishing the meanings of common words such as ‘cow’. The Mīmāṃsaka, of course, was cognizant of the fact that other cultures used different words for things, but he insisted that only Aryas speaking proper Sanskrit used the *correct* words! Others use words that do not really mean what they think they mean and whose expressiveness derives only from similarity to Sanskrit words. The fact that there is a natural connection between word and meaning does not entail that everyone must automatically know what a word means, either. Although the connection is natural, it still must be learned.

Nor, finally, is the meaning of a particular utterance dependent on the intention of the utterer. Mīmāṃsā rejected the view, propounded by both Buddhist philosophers and the Hindu Vaiśeṣika school, that verbal testimony is a form of inference, in the sense that we infer facts about the world from what a person says via the belief that what is said expresses the state of mind of the speaker. By contrast, Mīmāṃsā held that language of itself says things. We can, for example, understand what a sentence means without knowing who uttered it or whether that person is reliable or not. A speaker selects a certain pattern of words to express a thought just because those *words* already express a certain state of affairs. Bhāṭṭa and Prābhākara philosophers differed, however, on the matter of how the eternal meanings of individual words combine to form sentence meanings. The Bhāṭṭa view (the so-called *abhihitānvayavāda*) was that individual words first indicate their ‘own meanings’, which in turn combine into the sentence meaning; the Prābhākara view (the *anvitābhidhānavāda*) was that a word originally indicates a meaning

that is modified by the other words with which it stands in syntactic relation in the sentence. The latter theory implies that a word never means exactly the same thing in any two sentences.

Part and parcel of the Mīmāṃsā defence of the authority of the Veda was the effort to debunk other scriptures. The Buddhists and Jainas grounded the authority of their scriptures on the claim that they contained the utterances of omniscient beings - the Buddha and Mahāvīra respectively. Kumārila, in particular, ridiculed the idea of omniscience. We do not know of any omniscient people now, so on what basis should we believe that there were omniscient people in the past? Indeed, we observe that some people are smarter and more talented than others, but that hardly justifies believing in the existence of people who literally knew everything! Moreover, the statements of the Buddha and Mahāvīra, each claimed by his tradition to be omniscient, contradict each other. Given also the observation that humans for the most part say what is false, and that, especially when it comes to religion, charlatans abound, it is altogether safer to rely on a tradition that does not make outrageous claims about the omniscience of its founder, or, for that matter, appeal to any fallible human author at all. Indeed, humans cannot know what *dharma* is through their own faculties. For *dharma* is 'that which is conducive to the good', that is, the happiness of humans, and humans are incapable of knowing through their own faculties which actions yield happiness after death. For that reason, *dharma* is not to be based on rationally conceived, universal ethical principles such as nonviolence or 'benefiting others'. *Dharma* is just what the Veda tells us to do.

See also: Duty and virtue, Indian conceptions of; Interpretation, Indian theories of §4; Meaning, Indian theories of

JOHN A. TABER

References and further reading

Bhatt, G.P. (1962) *Epistemology of the Bhāṭṭa School of Pūrva Mīmāṃsā*, Benares: Chowkhamba Sanskrit Series. (A detailed treatment of Bhāṭṭa views.)

Dwivedi, R.C. (ed.) (1994) *Studies in Mīmāṃsā*, Delhi: Motilal Banarsidass.(A collection of articles by contemporary scholars.)

Jha, G. (1942) *Pūrva Mīmāṃsā in its Sources*, Benares: Benares Hindu University.(A standard, if outdated, survey and comparison of Bhāṭṭa and Prābhākara doctrines.)

Kumārila (7th century) *Ślokavārttika*, trans. G. Jha, Delhi: Sri Satguru Publications, 1983.(Translation of Kumārila's *magnum opus*.)

Prabhākara (7th century) *Bṛhatī*, ed. A. Chinnasvami Sastri, Chowkhamba Sanskrit Series 69, Benares: Chowkhamba Sanskrit Series, 1929-33.(The subcommentary on the *Śābarabhāṣya* that became the foundation of the Prābhākara school of Mīmāmsā.)

Śabarasvāmin (6th century) *Śābarabhāṣya*, trans. G. Jha, Gaekwad's Oriental Series 66, 70, 73, Baroda: Oriental Institute, 1933-6.(Translation of the *Mīmāṃsāsūtra*, together with Śabarasvāmin's commentary.)

Mimēsis

A crucial term in the literary theories of Plato and Aristotle, mimēsis *describes the relation between the words of a literary work and the actions and events they recount. In Plato, the term usually means 'imitation' and suggests that poetry is derived from and inferior to reality; in Aristotle, it loses this pejorative connotation and tends to mean simply 'representation' and to indicate that the world presented in a poem is much like, but not identical with, our own.*

A relation of *mimēsis* to authoritative figures from the past - heroes, fathers, philosophers, authors - was deeply ingrained in ancient Greek and Roman culture, and laid the groundwork for the post-classical notions of the pedagogical value of antiquity as a whole that shaped the classical tradition. *Mimēsis* is usually translated 'imitation', but in fact its central meaning is closer to 'actualization': objects, events, or actions which, because they are divine, past or canonical, belong to a more valuable domain of reality than our quotidian lives but are therefore in some way remote from us, enjoin upon us the obligation to restore their actuality; this is achieved by establishing a privileged sector within our present concerns in which we can (re-)enact them, thereby illuminating our banal world with some of their splendour while rescuing them from the perils of abstraction and irrelevance. Hence the term *mimēsis* is intrinsically ambivalent and, depending upon the circumstances, can emphasize either the actualization's inferiority compared to its model or the relative superiority it acquires by its temporary participation in the model's prestige.

This cultural complex provides the context for Plato's introduction of *mimēsis* as a concept of philosophical aesthetics and literary theory. Three different levels of generalization can be distinguished in Plato's usage of the term in the *Republic* (seePlato §14). First, discussing poetry's content in Book II, he applies the term broadly to the relation of signification between utterances and the entities to which they refer. Then, moving on to poetry's form in Book III, he exploits the fact that in Greek tragedy ordinary human actors played the roles of legendary heroes and gods to invent the novel, highly specific application of *mimēsis* as a technical term to that whole literary genre, drama, in which authors never speak in their own voice, in contrast to the narrative, in which they always so speak, and to mixed forms. Finally, once he has established the tripartition of the soul in Book IV and the way in which lower levels of reality imitate higher ones (the divided line and the analogy of the cave in Books VI and VII) he can return in Book X more fundamentally and drastically to his critique of non-philosophical *mimēsis*; now all poetry, including non-dramatic forms, is deceptively mimetic, in the sense that it is at three removes from true reality. Unless a defender can be found to plead its case in prose, it must be exiled from Plato's ideal city.

In his *Poetics*, Aristotle accepts Plato's challenge (see Aristotle §29). Against *Republic* III he generalizes *mimēsis* beyond the criterion of the dramatic genre and against *Republic* X he empties it of any hint of inferiority with regard to an independent reality: instead, now all imaginative literature, including Plato's dialogues, is mimetic, in the sense that it describes not what really happened but the sort of thing that could happen. Aristotle's *Poetics* is a theory of the poet's making (*poiēsis*) of a representation (*mimēsis*) of characters' actions (*praxis*): in a tragedy, their *praxis* will turn out badly; in a good tragedy, the poet's *poiēsis* will turn out well. If the tragedian directly made actions, people would really commit and suffer outrages and we would be appalled; instead, by making a *mimēsis* of actions, the tragedian can represent them for our considered pleasure. *Mimēsis* is no longer the unique province of the deceived and deceiving artist, but instead an anthropological universal that distinguishes us from most other animals (as in fact imitation does) and provides specifically cognitive pleasures to all humans: we enjoy recognizing what we already knew and saying that 'this man' (whom we see before us) is 'that man' (of whom we had always heard).

Greek aesthetics never comes closer to the modern concept of fictionality than it does in Aristotle's universally generalizing, thoroughly non-derogatory use of the term *mimēsis*. Yet Aristotle, who never analyses the term, remains committed to the view that the world portrayed by poetry is not a fictional world, but a recognizable version of our own real one: poetry does not show us what actually happens or happened, but it does show us the kinds of things that could have happened and might happen. *Mimēsis* prevents the poet's *poiēsis* from making a real *praxis*; but the possibilities it describes are real, not fictional, ones. That is precisely why Aristotle can insist

upon a limited cognitive component in aesthetic pleasure.

See also: Aesthetics; Katharsis; Tragedy

GLENN W. MOST

References and further reading

Aristotle (*c.* mid 4th century BC) *Poetics*, trans. S. Halliwell, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1995.(Greek text with an English prose translation on facing pages.)

Belfiore, E.S. (1992) *Tragic Pleasures: Aristotle on Plot and Emotion*, Princeton, NJ: Princeton University Press, esp. 45-70.(Argues for the cognitive and anthropological importance of *mimēsis* in Aristotle.)

Else, G.F. (1957) *Aristotle's Poetics: The Argument*, Cambridge, MA: Harvard University Press.(A provocative and original critique of widely shared views.)

Else, G.F. (1986) *Plato and Aristotle on Poetry*, ed. P. Burian, Chapel Hill, NC, and London: University of North Carolina Press.(A more generally accessible presentation of some of the arguments in Else (1957).)

Golden, L. (1992) *Aristotle on Tragic and Comic Mimēsis*, Atlanta, GA: Scholars Press.(An eloquent defence of a more cognitive view of *mimēsis* than the one presented here.)

Halliwell, S. (1986) *Aristotle's Poetics*, London: Duckworth.(A general introduction to the treatise with a full treatment of *mimēsis*, pp. 109-37.)

Lucas, D.W. (ed.) (1968) *Aristotle: Poetics*, Oxford: Clarendon Press, 258-72.(The standard English commentary, with a good discussion of *mimēsis*.)

Plato (*c.*370s BC) *Republic*, trans. G.M.A. Grube, Indianapolis, IN: Hackett, 1974.(See Books II-III, X for Plato's treatment of *mimēsis*.)

Rorty, A.O. (ed.) (1992) *Essays on Aristotle's Poetics*, Princeton, NJ: Princeton University Press.(A stimulating collection of recent articles on *mimēsis* and many related issues; good bibliography.)

Mind, bundle theory of

This theory owes its name to Hume, who described the self or person (which he assumed to be the mind) as 'nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement' (A Treatise of Human Nature I, IV, §VI). The theory begins by denying Descartes's Second Meditation view that experiences belong to an immaterial soul; its distinguishing feature is its attempt to account for the unity of a single mind by employing only relations among the experiences themselves rather than their attribution to an independently persisting subject. The usual objection to the bundle theory is that no relations adequate to the task can be found. But empirical work suggests that the task itself may be illusory.

Many bundle theorists follow Hume in taking their topic to be personal identity. But the theory can be disentangled from this additional burden.

1 Origins and nature of the theory

Hume's bundle theory originated in his response to what he took to be a natural illusion about personal identity:

> [S]ome philosophers…imagine we are every moment intimately conscious of what we call our SELF; that we feel its existence and its continuance in existence; and are certain, beyond the evidence of a demonstration, both of its perfect identity and simplicity.
>
> (*A Treatise of Human Nature* I, IV, §VI)

He thought that if that were so, the self would be an item within one's experience. But the self is not such an item: rather, it is that to which our experiences are ascribed. Neither can the existence of the self be inferred from, or found in any other way in, experience. If I introspect, all I find are thoughts, sensations, emotions, not the thing which has them.

This Cartesian self, then, is a myth. But how did the myth arise? Hume's own answer was that relations of resemblance and causation among one's different experiences lead one to suppose, mistakenly, that there is some real underlying unity. (A rough analogy would be getting the impression, in darkness, of a single light's travelling, by seeing adjacent lamps successively illuminated, each going on as its neighbour is extinguished.) This should not be interpreted as his offering some set of conditions for the truth of statements of the form '*x* is the same person as *y*'. Hume regarded it as his job not to give an account of personal identity, but merely to explain how we come to have the notion at all. But subsequent philosophers have conceived the bundle theory as just such an account.

One thing Hume's explanation of the origin of the notion of personal identity lacked was an answer to this question: given that there *are* bundles which constitute particular minds, what is it that ties a particular set of experiences into *that* bundle? Something must do so, because according to Hume, one gets one's false ideas about real identity by being presented with a limited set of experiences *already bundled* (not with all the experiences that there are, for most belong to other people). A Cartesian ego would do the job; but that is already condemned. And it would anyway not do a further job: lying as it does beyond the reach of experience, it would not on its own give one the notion of one's own self. Kant noticed this difficulty (*Critique of Pure Reason*, A103-10), and is sometimes read (for example by Strawson 1966: 163) as suggesting that a necessary element of the linking principle was the possibility of regarding one's experiences as *of* objects independent of those experiences, objects which provide the possibility of experiences which are not one's own (see Kant, I. §6). Be this as it may, bundle theorists have always been exercised by the need to provide suitable unifying links among experiences. Such links have to be capable of accounting for the unity of a single person, both at one time (synchronic unity) and across time (diachronic unity).

Hume's assumption of the identity of self and mind exacerbates his problem. For it multiplies the demands that a bundle theory has to meet. And defenders of the bundle theory have usually been those who have *argued* for this identity, so that they regard themselves as facing the task described of supplying links among experiences adequate to account for both the synchronic and diachronic unity of a person. But it is in principle possible for a bundle theorist of mind to hold, for example, a body-based account of personal identity (see Personal identity). Certain natural objections to the theory would then immediately collapse. One such is this: it seems obviously true

that I might have had a quite different set of experiences from the set I have had. But if I just *am* a set of experiences, this ’obvious truth’ must in fact be false.

Treating the bundle theory, then, as purely a theory of the nature of mind, what sort of theory is it? Initially, little more than a denial of the necessity of an immaterial owner of experiences, the bundle theory is almost defined by its claiming there to be no item in the category of mental substance to which experiences must be attributed. But it is also committed to the denial of the necessity for a material owner. The theory’s most natural form would thus be something like Russell’s post-1919 neutral monism (see Russell, B.A.W. §13; Neutral Monism). But it is compatible with versions of dualism and physicalism that do not have the strong ownership commitments just described. In this sense the bundle theory is only a partial account of the nature of mind.

2 Objections and replies

A common objection is that experiences are identity-dependent on their owners, that the only way in which a particular experience can be identified is as an experience of some person, and that the bundle theory is committed to denying this. But a bundle theorist could without inconsistency admit the claim of identity-dependence, and still ask for proof that the necessary owner must be something other than the rest of the bundle; only this further step would actually refute the theory.

Another, related, objection is that the theory has the supposedly absurd implication that experiences can exist in isolation from each other, for a bundle consists of things that can exist outside the bundle, and can in principle fall apart. If they cannot so exist, then Hume’s atomist version of the theory, where the experiences are mutually independent, would be false. But a non-atomist version of the theory, where the experiences are inextricably interlinked, such as one might find in William James’ neutral monism, could survive this objection (see James, W. §2).

More difficult is the problem of unity - finding suitable candidates for the cross-experiential links. Hume’s suggestions, resemblance and causation, seem more suited to solving the problem of diachronic rather than synchronic unity. But, jointly or severally, these are neither necessary nor sufficient for either. There is more resemblance among different people’s experiences of one thing than among one person’s experiences of different things; and experiences in one mind may have no causal relations between them (for example I hear the doorbell while watching television), whereas experiences in two minds might be cause and effect (for example if telepathy is possible).

Another possibility is that certain experiences embrace others. Synchronic unity would have to be guaranteed by something like a second-order awareness of co-presentation of several first-order experiences. But it is hard to see that such second-order awareness is always present, while its mere possibility seems itself to require explanation in terms of a single mind. (A possible response to this criticism would be that there is no such requirement, that we have reached bedrock.) Further, it would still leave the problem of diachronic unity, which needs another kind of embracing, for example memory’s spanning a sequence of experiences. This, on the face of it, would imply that an experience I do not remember belonged to a different mind. Again, the appeal to possibility seems no help.

Some bundle theorists have postulated diachronic links of a more sophisticated kind: if there is enough psychological connectedness over time, then there is a single mind. Connectedness is exemplified by recall of a previous experience, by the carrying out of a prior intention, or by the persistence of personality traits. It is a matter of degree, and hence can approach zero over the career of a single mind. It thus needs supplementation with psychological continuity, which is the holding of overlapping chains of a high degree of connectedness. But even together these links would allow a single mind to divide, with each successor having an equal claim to be the original, yet their diversity defeating that claim via the logic of identity. (If $b \neq c$, then it cannot be true both that $a = c$ and $a = b$.)

It is not clear what we should infer from this. One possibility is that the connectedness/continuity account is wrong. Another is that we should allow identity to admit degrees, and rewrite its logic. But one might conclude that these constant difficulties in finding links that guarantee both synchronic and diachronic identity of the mind may show merely that the theory’s implication that the unity of mind is an illusion needs to be taken seriously.

From this perspective, better arguments for a bundle theory may lie less in a priori considerations than in empirical

work by psychologists and brain scientists, which reveals striking breakdowns in mental unity (see Split brains). This suggests to some that the mind may best be viewed in terms not of linked experiences, but of hierarchies, systems and modules of monitoring and control of experience and activity, with all the unity there is being provided contingently by its embodiment in a normally functioning single organism. The bundle theory's natural home may thus turn out to be with functionalist accounts of mind.

See also: Consciousness §1; Dualism; Modularity of mind; Persons; Strawson P.F. §§6-7

STEWART CANDLISH

References and further reading

Brennan, A. (1989-90) 'Fragmented Selves and the Problem of Ownership', *Proceedings of the Aristotelian Society* LXXXX: 143-58.(Argues that the theory's failure to provide tight conditions for synchronic and diachronic unity of the self is a reason for accepting rather than rejecting it. Contains many useful references to other discussions.)

Brennan, A. (1994) 'The Disunity of the Self', in J.J. MacIntosh and H. Meynell (eds) *Faith, Scepticism and Personal Identity: Essays in honour of T.M. Penelhum*, Calgary, Alta.: University of Calgary Press.(A defence of the theory against some standard objections; expands some points made very briefly in §2 above.)

Canfield, J. (1990) *The Looking-Glass Self: An Examination of Self-Awareness*, New York: Praeger.(Presents a no-self account of the person, connecting it with the philosophy of Wittgenstein and with Buddhism. Written for the general reader.)

Carruthers, P. (1986) *Introducing Persons*, London: Routledge, ch. 2.(Accessible to beginners without being simplistic; criticizes the bundle theory without treating it as a theory of personal identity.)

Descartes, R. (1641) *Meditations on First Philosophy*, trans. J. Cottingham, R. Stoothoff and D. Murdoch, in *The Philosophical Writings of René Descartes*, Cambridge: Cambridge University Press, 1984.(Classic statement and defence of dualism.)

Giles, J. (1993) 'The No-Self Theory: Hume, Buddhism, and Personal Identity', *Philosophy East and West*, 43 (2): 175-200.(A careful account and defence of Hume. Argues a case like Brennan's, and claims that Buddhist thinkers held a similar view. Contains useful references to Buddhist literature on the topic.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1978, I, IV, VI.(The seminal statement of the theory.)

Kant, I. (1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1933.(An influential but obscure treatment of the unity of the mind's experience.)

Lowe, E.J. (1991) 'Real Selves: Persons as a Substantial Kind', in D. Cockburn (ed.) *Human Beings*, Cambridge: Cambridge University Press.(Attacks the bundle theory on grounds of necessity of ownership. Clear and vigorous, but not for beginners.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press, chaps 10 and 11.(An influential version of the bundle theory as an account of synchronic unity and personal survival, embedded in a wider discussion of personal identity, rationality and ethics. A major work which has provoked much controversy.)

Quinton, A. (1962) 'The Soul', *Journal of Philosophy* 59 (15), repr. in J. Perry (ed.), *Personal Identity*, London: University of California Press, 1975.(Defends a version of the theory in a materialist context. Clearly written.)

Strawson, P.F. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen, ch. 3.(The most definitive statement of the necessity-of-ownership objection to the bundle or 'no-ownership' theory. Difficult but rewarding.)

Strawson, P.F. (1966) *The Bounds of Sense*, London: Methuen.(A widely used but somewhat idiosyncratic essay on Kant's *Critique of Pure Reason*.)

Stroud, B. (1977) *Hume*, London: Routledge & Kegan Paul, ch. 6.(A very good, but intricate, account of Hume's treatment of personal identity. Accessible to undergraduates, but not to complete beginners.)

Mind, child's theory of

Knowledge of other minds poses a variety of unusual problems due to the peculiarly private nature of mental states. Some current views, impressed by the contrast between the apparently direct access we have to our own mental states and the inaccessibility of others' mental states, argue that we understand the mental states of others by imagining that they are our own by 'simulation'. Other current views propose that we infer both our own mental states and the mental states of others by employing a set of conjectures arrived at through general inductive reasoning over experience: a 'folk psychology' or 'theory of mind'.

Experimental studies, by contrast, suggest that we possess an 'instinct' for comprehending the informational mental states of other minds. Children develop mental state concepts uniformly and rapidly in the preschool period when general reasoning powers are limited. For example, children can reason effectively about other people's beliefs before they can reliably calculate that 2 plus 2 equals 4. In the empirical study of the 'theory of mind' instinct there have been three major discoveries so far: first, that normally developing 2-year-olds are able to recognize the informational state of pretending; second, that normally developing children can, by the age of 4 years, solve a variety of false belief problems; and lastly, that this instinct is specifically impaired in children with the neurodevelopmental disorder known as 'autism'.

1 False belief understanding

The study of false belief has come to dominate much of the investigation of the child's theory of mind. Interest in false belief arose initially out of an article published by Premack and Woodruff (1978) on whether or not chimpanzees have a 'theory of mind'. The findings strongly suggested that chimpanzees have some awareness of an agent's goal because they made inferences concerning the use of objects as tools. In so far as understanding the use of tools in relation to an agent's goals is part of 'theory of mind', chimpanzees demonstrated a version of theory of mind. In a commentary on Premack and Woodruff's paper, Dennett (1978) has suggested that a stronger test of an animal's understanding of mental states would involve an understanding of false belief. In understanding false belief, the animal has to understand a belief that is not its own and therefore that does not simply reflect reality as the animal construes it. Such tests with chimpanzees have returned essentially negative results (see Premack 1988; Animal language and thought). However, the idea was taken up by Wimmer and Perner who developed scenarios for testing not chimpanzees, but human children.

In the first of these, Wimmer and Perner (1983) showed that 6-year-old children could predict an agent's behaviour from a false belief attributed by the children to the agent. Wimmer and Perner's original task was somewhat more complex than necessary. When Baron-Cohen, Leslie and Frith (1985) simplified the task, they found that the vast majority of normally developing 4-year-old children could successfully calculate Sally's belief in the following scenario. Sally puts her candy in a cupboard and goes out to play. Anne comes in, finds the candy and moves it to a nearby drawer. Now Sally comes back for her candy. Where does Sally think her candy is? Where will she look for it? Eighty-five per cent of 4-year-olds were able to attribute a false belief to Sally and thus to predict her erroneous behaviour. Children younger than 4, however, tend to predict Sally's behaviour from the candy's current position, as if Sally would know this.

Likewise, 4- but not 3-year-olds succeed in the following scenario from Perner, Leekam and Wimmer (1987). The child is shown a *Smarties* box and asked what it contains. '*Smarties*' [candy] is the invariable reply. The child is then shown that, actually, the box contains only a pencil. Disappointment. The pencil is placed back in the box and the lid closed again. The child is reminded that a friend is outside waiting to come in. Then the test question is asked, 'What will your friend say (/think) is in the box when we first show it to him/her? Again most 4-year-olds correctly predict '*Smarties*', while most 3-year-olds expect the friend to say, 'A pencil'.

The second key finding is that children suffering from the neurodevelopmental disorder known as 'childhood autism' have a focused and specific impairment in the capacity to recognize and reason about mental states. Autism is a life-long disorder resulting from complex biological causes (Frith 1989), characterized by social and communicative impairments beyond that expected on the basis of whatever degree of mental retardation may also be present (Wing and Gould 1979).

Baron-Cohen, Leslie and Frith (1985) presented three groups of children with the Sally/Anne scenario outlined above. One group consisted of normal 4-year-olds, one of Down's syndrome children and one of (older) autistic children. It was deliberately arranged that the autistic group had a mental age considerably higher than that of the Down's group. Furthermore, their IQs were in the borderline to normal range (mean 82), while the Down's children averaged an IQ of 64, in the moderately retarded range. The results of this study showed that, while around 85 per cent of both the normal and Down's groups correctly predicted Sally's false belief, only 20 per cent of the autistic group did so, despite their intellectual and age advantage. Any difficulty the autistic children might have with this task, therefore, cannot be attributed to general intellectual level.

In a follow-up study, Baron-Cohen, Leslie and Frith (1986) compared, across these same three groups, the children's ability to understand different types of events. They used a 'picture sequencing' task. Some picture sequences depicted physical-causal events, and some sequences depicted events which could only be appreciated if one took into account the (false) beliefs of the protagonists. The pictures in each sequence were jumbled and the children were asked to order each sequence so that the story 'made sense'. The results were again quite striking. The autistic children showed a specific difficulty. They performed well on the physical-causal stories but slumped to chance performance on the false belief stories. The younger normal children scored almost 100 per cent on these last stories, while the more retarded Down's group still out-performed the autistic group. Finally, analysis of the children's verbal descriptions of the stories showed that the autistic children produced much more physical-causal language but much less mental state language than the other two groups. Leslie and Frith (1989) subsequently showed that high-ability autistic children had difficulty understanding true-but-incomplete beliefs as well as false beliefs. Perner *et al.* (1989) showed that autistic children failed the *Smarties* task.

The findings of the above initial sets of studies on both normal and autistic children have subsequently been confirmed and extended in a large number of studies (see, for example, Baron-Cohen, Tager-Flusberg and Cohen 1993). Together, they raise the following question. How can the young brain attend to mental states when they cannot be seen, nor heard, nor felt? Leslie (1994) argues that this is the fundamental question to ask, because the child has to be able to attend to mental states before they can learn anything about them. The only plausible answer so far has been to suggest that children are equipped with a specialized cognitive mechanism that appears early in development when encyclopedic knowledge and general reasoning abilities are still severely limited. The pattern of development seen in childhood autism can then be understood as reflecting, at least in part, early biological damage to this specialized mechanism, damage which spares general intelligence (see Modularity of mind).

2 The significance of pretending

There is a certain form of false belief which is understood much earlier even than 4 years. This is 'make-believe' or pretending. Two-year-olds have no trouble recognizing when someone playfully pretends; when someone acts as if they believed an imaginary situation were real. Two-year-olds can readily figure out the specific content of the pretence (Leslie 1994). Likely as not, the 2-year-old will join in the game and proceed to contribute their own appropriate counterfactual inferences. This ability, which first emerges between 18 and 24 months, has been analysed by means of a cognitive model, the so-called 'metarepresentational theory of pretence', which identifies the main properties of the internal representations required by this ability (see Leslie 1987; Aesthetic attitude). These representations turn out to express the key information required to represent propositional attitudes. This led to the suggestion that these representational mechanisms form the specific innate basis of the child's capacity to acquire and elaborate a 'theory of mind'. Autistic children are impaired in their capacity to pretend and, for example, do not spontaneously pretend play. Indeed, this was a key consideration that led to the studies of belief understanding in autism outlined above.

One key feature of the capacity for pretence is that as soon as a child becomes capable of pretending in isolation, by themselves, they also become capable of understanding pretence-in-others (Leslie 1987; 1994). Thus, there is a 'developmental yoking' between being able oneself to pretend that a banana is a telephone and being able to understand that someone else is pretending that a banana is a telephone. This is strikingly different from the case of belief. Most psychologists agree that infants a few months olds can have beliefs, for example, that there is an object behind a screen. But no one thinks that these young infants simultaneously have the capacity to understand that someone else can have that same belief: there is no yoking between the capacity to have beliefs and the capacity to understand beliefs-in-others (see Belief; Propositional attitudes).

The metarepresentational theory of pretence referred to earlier explains the above pattern in the following way. Let us say, for expository purposes, that to have a belief is to stand in a certain kind of functional or processing relation to an internal representation (see Language of thought). Let us call this functional relation the 'belief box'. If a given representation *r* is placed in the belief box, then the organism comes to believe that *r*. In a similar vein, we might talk of a 'desire box', such that if *r* is placed in the desire box, then the organism comes to desire that *r*. In this vein, the metarepresentational theory of pretence says that there is no such thing as a 'pretend box'. Instead, in pretending, a representation 'I pretend that *r*' or a representation 'Mother pretends that *r*' is placed in the belief box and results, respectively, in solitary pretence or in understanding pretence-in-others. What determines the developmental emergence of pretending, then, is not the appearance of a 'pretend box', a new functional relation, but the appearance of new representations of the form '*X* pretends that *r*', that is, the appearance of metarepresentations. The understanding of belief-in-others likewise depends upon the emergence of metarepresentations of the form '*X* believes that *r*' and is delayed relative to the appearance of the 'belief box' at least as long as pretending is.

3 Why do normal 3-year-olds and autistic children fail false belief tasks?

How long after understanding pretence-in-others do children understand belief-in-others? So far what we have said suggests something of the order of two years. However, this may be an over-estimate. The 'standard' tasks used to assess false belief understanding seem to impose general problem-solving demands that 3-year-old children find hard to meet. These performance demands may be masking their underlying competence. Psychologists have been developing versions of the false belief tasks that make reduced demands and 3-year-olds do better on these simplified tasks.

For example, Wellman (1990) found that 3-year-olds could pass a version of the Sally/Anne task in which the position of the target object was not known but simply guessed. Wherever the child guessed, the experimenter said that Sally thought it was some other place. When asked to predict Sally's search behaviour, then, the child has to predict on the basis of Sally's different belief. Under these circumstances most 3-year-olds succeed. Apparently, the child's efforts to represent Sally's belief are not swamped by current reality if reality is only guessed at and not known for sure (see also Zaitchik 1991). In the *Smarties* task, certain memory cues built into the procedure seem to facilitate false belief success in 3-year-olds (Mitchell and Lacohe 1991). Roth and Leslie (1991) showed that normal 3-year-old children, while failing standard false belief tasks, may nevertheless attribute propositional attitudes to participants in a conversation. Roth and Leslie's autistic subjects, however, failed to perform even at the 3-year-olds' level.

If 3-year-olds fail standard false belief tests because of performance limitations, might not the same be true for autistic children? It is nigh impossible to think up, let alone rule out, every potential general processing resource that might be required by these tasks and be impaired in autism. Many such accounts are possible: working memory, language ability, visual imagery ability, abstract reasoning, executive functioning, a cooperative attitude, and so on. Fortunately, such explanations can be ruled out by the following results. Zaitchik (1990) devised a task that closely parallels the standard false belief tasks but instead of an agent forming a belief that subsequently goes out of date, the task involves a camera taking a Polaroid photograph that subsequently goes out of date. So the camera takes a photograph of the candy in the cupboard. The photograph is placed faced down on the table without the child seeing it (after all, the child does not get to see Sally's belief either). Now the candy is moved from the cupboard to the drawer. The usual 'control' questions are asked: 'Where was the candy when I took a photograph of it?' and 'Where is the candy now?'. As usual, the child must get these questions right if they are to be tested further. The child is asked, 'In the photograph, where is the candy?'.

Results with normal children show that, typically, 3-year-olds fail both the out-of-date photograph and belief tasks, while most 4-year-olds pass both. However, if a normal child passes only one of these tasks, there is a reliable tendency for them to pass the false belief rather than the out-of-date photograph task. Leslie and Thaiss (1992) showed that autistic children show near ceiling performance on out-of-date photographs while being severely impaired on the similar out-of-date belief task. (Incidentally, this effect is not restricted to cameras and photographs; it works just as well with maps and drawings.) This is impressive evidence for the domain specificity of autistic impairment. It also contradicts the perhaps natural assumption that understanding mental states and understanding representations should simply be different facets of the same psychological mechanism (see, for

example, Fodor 1992).

4 Alternative theories of 'theory of mind' development

Other current positions in the literature view the development of 'theory of mind' knowledge as the discovery by young children of a succession of 'theories' much as scientists have created and abandoned successive theories of the natural world (for example, Gopnik 1993; Perner 1991; Wellman 1990; see Folk psychology). However, the obvious differences between the historical achievements of talented individuals and the rapid, early and uniform acquisitions of ordinary preschool children imply in the latter case a process more akin to the unfolding of an albeit 'articulated' instinct than to the scientific process. Yet another recent position wants to deny the existence of theory-like knowledge or inference over data structures by substituting the notion that we 'simulate' the mental states of others (Harris 1992; Goldman 1992). Simulation theory proposes that we understand other people's actions by modelling their circumstances and predicting their actions by means of the very mechanisms that generate our own actions. We are said to run our action-planning system 'off-line', supplying it with 'pretend' inputs, to 'simulate' the other person's mental life by, as it were, placing our self 'in the other person's shoes'.

However, there are a number of quite different processes that could be called 'simulation' and their contributions to 'theory of mind' abilities will each have to be determined empirically (see Nichols *et al.* 1996). It may well turn out that 'simulation' is one of the abilities associated with deploying folk psychological knowledge. However, Leslie and German (1995) have pointed out a number of problems for the radical claim that the existence of simulation abilities means there is no need for 'theory of mind' (meta) representations.

Neither of these current alternative views has addressed the fundamental problem of 'theory of mind' development, namely, how is it possible for the young brain to attend to mental states in the first place? The metarepresentational theory was designed for this problem. The theory proposes that the young brain attends to an agent's behaviour and infers an underlying causal state of the agent having the properties specified by the metarepresentation. The metarepresentation is computed by a specialized, probably modular, innate mechanism. This mechanism allows the young brain to attend to and learn about invisible mental states and promotes the early and rapid development of 'theory of mind'. When early occurring biological damage prevents the normal expression of this mechanism in the developing brain, the result is the core symptoms of autism.

See also: Cognition, infant; Modularity of mind; Nativism; Other minds

ALAN M. LESLIE

References and further reading

Baron-Cohen, S., Leslie, A.M. and Frith, U. (1985) 'Does the Autistic Child Have a "Theory of Mind"?', *Cognition* 21: 37-46.(This paper tested the first prediction of the 'metarepresentational' theory and reported the discovery of the 'theory of mind' impairment in autism.)

Baron-Cohen, S., Leslie, A.M. and Frith, U. (1986) 'Mechanical, Behavioural and Intentional Understanding of Picture Stories in Autistic Children', *British Journal of Developmental Psychology* 4: 113-25.(A follow-up to the 1985 paper by the same authors, shows children with autism performing well on a test of understanding stories with mechanical or behavioural themes but poorly on stories with mental state themes.)

Baron-Cohen, S., Tager-Flusberg, H. and Cohen, D. (eds) (1993) *Understanding Other Minds: Perspectives from Autism*, Oxford: Oxford University Press.(A very useful recent collection of articles on the 'theory of mind' impairment in autism.)

Dennett, D.C. (1978) 'Beliefs about Beliefs', *Behavioral and Brain Sciences* 1: 568-70.(This short commentary on Premack and Woodruff (1978) helped shape subsequent psychological research.)

Fodor, J.A. (1992) 'A Theory of the Child's Theory of Mind', *Cognition* 44: 283-96.(Fodor argues that 3-year-olds' failure on the false belief problem is the result of a performance limitation.)

Frith, U. (1989) *Autism: Explaining the Enigma*, Oxford: Blackwell.(An outstandingly interesting and readable account of the nature of autistic disorder.)

Goldman, A. (1992) 'In Defense of the Simulation Theory', *Mind and Language* 7: 104-19.(A leading philosophical proponent argues for 'theory of mind' as simulation.)

Gopnik, A. (1993) 'How We Know our Minds: The Illusion of First-Person Knowledge of Intentionality', *Behavioral and Brain Sciences* 16: 1-14.(Argues for the 'theory-theory' and for the claim that we do not have privileged access to our own mental states.)

Harris, P.L. (1992) 'From Simulation to Folk Psychology: The Case for Development', *Mind and Language* 7: 120-44.(A leading psychological proponent argues for 'theory of mind' as simulation.)

Leslie, A.M. (1987) 'Pretense and Representation: The Origins of "Theory of Mind"', *Psychological Review* 94: 412-26.(First major statement of the metarepresentational theory of normal and abnormal development.)

Leslie, A.M. (1994) 'Pretending and Believing: Issues in the Theory of ToMM', *Cognition* 50: 211-38.(Updated statement of the metarepresentational theory.)

Leslie, A.M. and Frith, U. (1988) 'Autistic Children's Understanding of Seeing, Knowing and Believing', *British Journal of Developmental Psychology* 6: 315-24.(Autistic children understand basic conditions for seeing and not-seeing, but are impaired on understanding knowing and not-knowing, and on understanding falsely believing.)

Leslie, A.M. and German, T.P. (1995) 'Knowledge and Ability in "Theory of Mind": One-Eyed Overview of a Debate', in M. Davies and T. Stone (eds) *Mental Simulation: Philosophical and Psychological Essays*, Oxford: Blackwell, 123-50.(Critique of the radical simulation account of theory of mind.)

Leslie, A.M. and Thaiss, L. (1992) 'Domain Specificity in Conceptual Development: Neuropsychological Evidence from Autism', *Cognition* 43: 225-51.(Reports key findings on 'theory of mind' impairment in autism and refutes the 'representational theory-theory'.)

Mitchell, P. and Lacohe, H. (1991) 'Children's Early Understanding of False Belief', *Cognition* 39: 107-27.(A simple manipulation produces substantially better performance on a standard false belief test in normally developing 3-year-olds.)

Nichols, S., Stich, S., Leslie, A.M. and Klein, D. (1996) 'Varieties of Off-Line Simulation', in P. Carruthers and P.K. Smith (eds) *Theories of Theories of Mind*, Cambridge: Cambridge University Press.(Critical assessment of the simulation theory in the light of available evidence.)

Perner, J. (1991) *Understanding the Representational Mind*, Cambridge, MA: MIT Press.(Major statement of the 'theory-theory' that preschool children have to discover that mental states are really mental representations.)

Perner, J., Frith, U., Leslie, A.M. and Leekam, S.R. (1989) 'Exploration of the Autistic Child's Theory of Mind: Knowledge, Belief and Communication', *Child Development* 60: 689-700.(The *Smarties* task produces excellent performance in specific-language-impaired children but poor performance in verbal-ability-matched autistic children.)

Perner, J., Leekam, S.R. and Wimmer, H. (1987) 'Three-Year-Olds' Difficulty with False Belief: The Case for a Conceptual Deficit', *British Journal of Developmental Psychology* 5: 125-37.(The *Smarties* task.)

Premack, D. (1988) '"Does the Chimpanzee Have a Theory of Mind?" Revisited', in R.W. Byrne and A. Whiten (eds) *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, Oxford: Clarendon Press.(Premack returns to chimpanzee's theory of mind and notes some limitations.)

Premack, D. and Woodruff, G. (1978) 'Does the Chimpanzee Have a Theory of Mind?', *Behavioral and Brain Sciences* 4: 515-26.(This paper coined the term 'theory of mind' and spurred the interest of psychologists.)

Roth, D. and Leslie, A.M. (1991) 'The Recognition of Attitude Conveyed by Utterance: A Study of Pre-school and Autistic Children', *British Journal of Developmental Psychology* 9: 315-30; repr. in G.E. Butterworth, P.L. Harris, A.M. Leslie and H.M. Wellman (eds) *Perspectives on the Child's Theory of Mind*, Oxford: Oxford University Press, 315-30.(Autistic children perform less well even than 3-year-olds, who are more willing, under some circumstances, to attribute false beliefs than are 5-year-olds.)

Wellman, H.M. (1990) *The Child's Theory of Mind*, Cambridge, MA: MIT Press.(Major statement by a leading theory-theorist argues for early emergence of a concept of desire and for an early 'copy theory' of belief.)

Wimmer, H. and Perner, J. (1983) 'Beliefs about Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception', *Cognition* 13: 103-28.(Seminal paper which developed the false belief task.)

Wing, L. and Gould, J. (1979) 'Severe Impairments of Social Interaction and Associated Abnormalities in Children: Epidemiology and Classification', *Journal of Autism and Developmental Disorders* 9: 11-29.(A classic paper establishing the general nature of social impairment in autism.)

Zaitchik, D. (1990) 'When Representations Conflict with Reality: The Preschooler's Problem with False Beliefs and "False" Photographs', *Cognition* 35: 41-68.(Elegant experimental study comparing young children's performance on the false belief task with a highly similar 'false photograph' task, with some unexpected results.)

Zaitchik, D. (1991) 'Is Only Seeing Really Believing?: Sources of True Belief in the False Belief Task', *Cognitive Development* 6: 91-103.(Experimental study showing that 3-year-old children find the attribution of false belief to someone else easier if they are not completely confident in their own belief.)

Mind, computational theories of

The computational theory of mind (CTM) is the theory that the mind can be understood as a computer or, roughly, as the 'software program' of the brain. It is the most influential form of 'functionalism', according to which what distinguishes a mind is not what it is made of, nor a person's behavioural dispositions, but the way in which the brain is organized. CTM underlies some of the most important research in current cognitive science, for example, theories of artificial intelligence, perception, decision making and linguistics.

CTM involves a number of important ideas. (1) Computations can be defined over syntactically specifiable symbols (that is, symbols specified by rules governing their combination) possessing semantic properties (or 'meaning'). For example, addition can be captured by rules defined over decimal numerals (symbols) that name *the numbers. (2) Computations can be analysed into 'algorithms', or simple step-by-step procedures, each of which could be carried out by a machine. (3) Computation can be generalized to include not only arithmetic, but deductive logic and other forms of reasoning, including induction, abduction and decision making. (4) Computations capture relatively autonomous levels of ordinary psychological explanation different from neurophysiology and descriptions of behaviour.*

1 Syntax, semantics and algorithms

Crucial to understanding computational theories of mind (CTM) is the distinction between the syntax and the semantics of a symbol system. *How* symbols may be combined comprises the syntax of a system and what the symbols *mean* or *name* comprises its semantics. For example, consider three different notational systems for numbers: decimal numerals ('1', '2', '3',…), Roman numerals ('I', 'II', 'III',…) and binary numerals ('1', '10', '11',…). These different types of *numerals* name the same *numbers* ('5' names the same number as 'V' and '101') but the rules for combining the numerals in calculations are very different. In other words, these systems each have a different syntax, but (approximately) the same semantics.

When children learn arithmetic, they learn reliable rules, or 'algorithms', for computing *syntactically* with numerals (symbols) in ways that mirror their *semantics* (in particular, the arithmetical relations among the corresponding numbers). Children doing calculations with decimal numerals should get the same answers as a computer working in binary, but the algorithms each uses look very different.

An algorithm is a 'mechanical' step-by-step procedure operating on syntactically well-defined symbols in a way that captures relations among the things the symbols represent. Alan Turing developed a general account of algorithms, conceiving what are called 'Turing machines', perfectly mechanical computing devices that arrive at a determinate answer or 'output' to a question, given certain data or 'input' (see Turing machines). According to the influential 'Church-Turing thesis' Turing machines can systematically compute anything that is intuitively 'computable' (see Church's thesis). This idea is the inspiration for much of the modern computer industry. CTM extends it to psychology: the suggestion is that all intelligent processes can be systematically decomposed into algorithms consisting of steps that can be executed by primitive processors of a machine.

2 Primitive processors

The specific architecture of a Turing machine involves a 'tape' consisting of an infinite number of individual cells and a 'scanner', whose primitive processes consist in registering whether it is scanning a '1' or a '0', and then moving left and right from cell to cell (see Turing machines). This is only one among many possible computational architectures. What is essential to a computer is merely that the primitive processes be realizable by some physical means or other. To make computational theories of mind (CTM) remotely plausible as an account of the *human* mind, we will think here in terms of the kinds of electrical devices that seem at least to be available in the human nervous system (but are also exploited in commercial computers).

A good example of a primitive processor is a 'gate'. An 'and'-gate is a device that accepts two inputs and emits a single output. If both inputs are '1's (that is, sum to 2), the output is a '1'; otherwise, the output is a '0'. An 'exclusive-or' (either but not both) gate is a 'difference detector': it emits a '0' if its inputs are the same and it emits a '1' if its inputs are different (or sum exactly to 1).

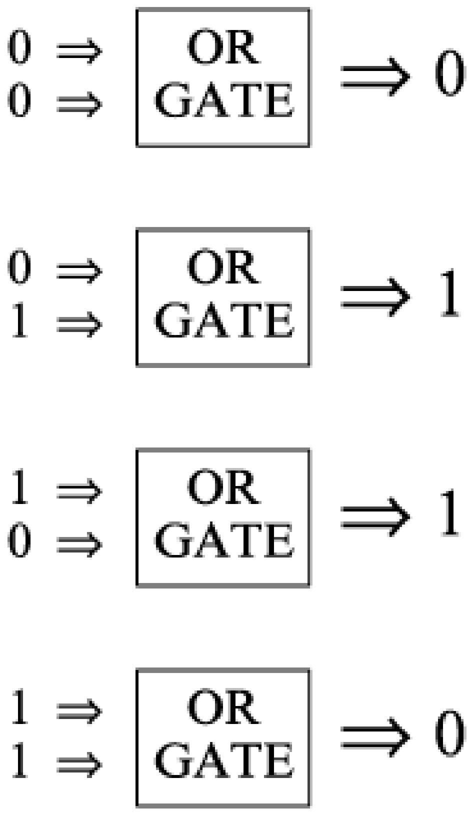

Figure 1a. Firing pattern of an “exclusive-or”-gate

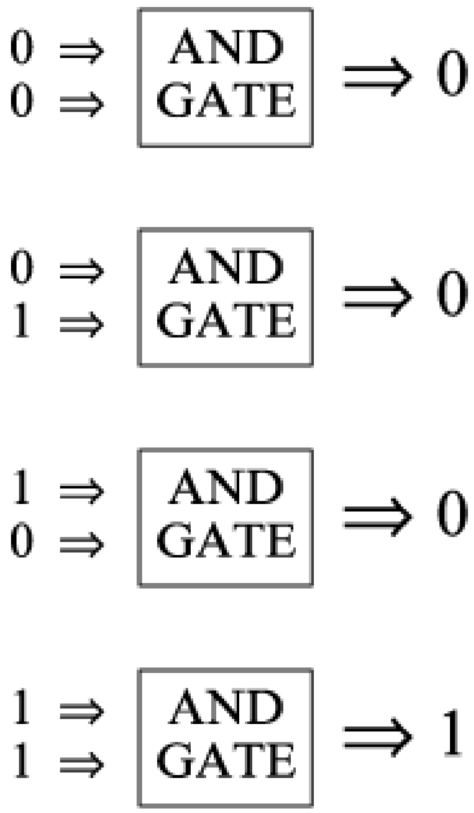

Figure 1b. Firing pattern of an “and”-gate

This talk of ‘1’ and ‘0’ is a way of thinking about the ‘bi-stable’ states of computer representers: they are almost always in one or the other of two states; the in-between states are not exploited computationally.

Using these gates, one could construct a binary ‘adder’ that operated according to the following rules of binary addition:

$0 + 0 = 0$

$1 + 0 = 1$

$0 + 1 = 1$

$1 + 1 = 10$

The two digits to be added are connected both to an ‘and’-gate and to an ‘exclusive-or’-gate as illustrated.

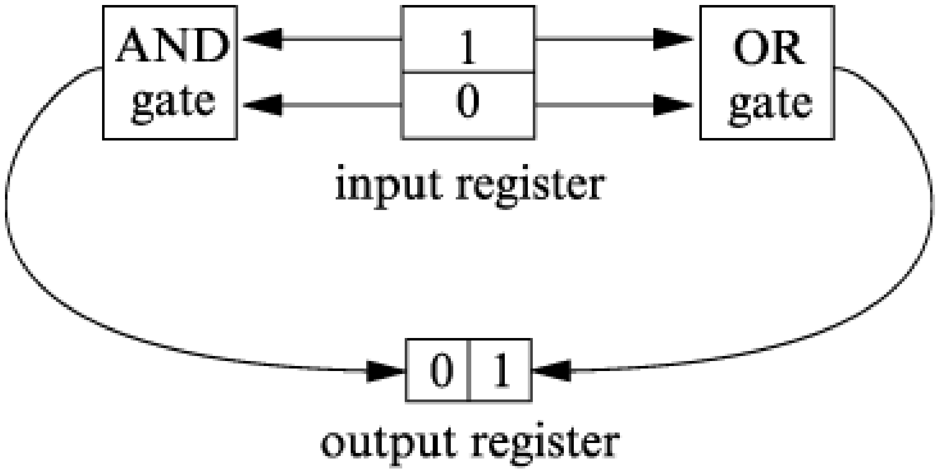

Figure 2a. Adder computing 1 + 0 = 1

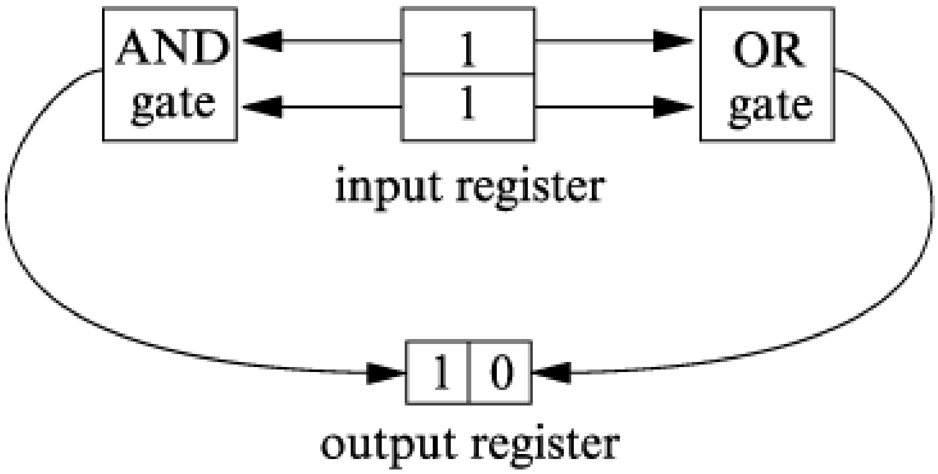

Figure 2b. Adder computing 1 + 1 = 10

Consider figure 2a first. The digits to be added are ‘1’ and ‘0’, and they are placed in the input register (the top pair of boxes). The ‘exclusive-or’-gate sees different things, and so outputs a ‘1’ to the right-most box of the answer register (the bottom pair of boxes). The ‘and’-gate outputs a ‘0’ except when it sees two ‘1’s, so it now outputs a ‘0’. In this way, the circuit computes ‘ $1 + 0 = 1$’. For this problem, as for ‘ $0 + 1 = 1$’ and ‘ $0 + 0 = 0$’, the ‘exclusive-or’-gate does all the real work. The role of the ‘and’-gate in this circuit is ‘carrying’, which is illustrated in figure 2b. The digits to be added, ‘1’ and ‘1’, are again placed in the top register. Now, both inputs to the ‘and’-gate are ‘1’s, so the ‘and’-gate outputs a ‘1’ to the left-most box of the answer (bottom) register. The ‘exclusive-or’-gate puts a ‘0’ in the right-most box, and so we have the correct answer, ‘10’.

3 Syntactic engines

The key idea behind the adder is that of a mirroring, or an isomorphism, between syntactic relations among numerals and their semantics, namely the relations among the numbers the numerals represent (see §1 above). This idea need not be confined to arithmetic. In their development of modern ‘formal’ logic, Frege and Russell showed how deductive logic can be formalized: that is to say, it can be characterized in terms of syntax, and this syntax

can be characterized in terms of the physical *form* of sentences that receive a systematic semantic interpretation. In particular, it can be shown that a certain set of syntactically and formally specifiable rules is adequate to capture a significant class of deductively valid arguments (see Formal languages and systems).

Ordinarily, sentences in logic are manipulated by people consciously following explicit rules, such as *modus ponens*, that are defined over their syntactic form, and it is often feared that CTM is committed to 'homunculi' in the brain executing the rules. However, the rules could also be obeyed by a Turing machine or, alternatively, a machine constructed along the lines of our simple adder: for example, whenever the machine detected sentences of the form 'P' and '$P \Rightarrow Q$' as input, it could print 'Q' as output, thereby obeying *modus ponens*. In general, Turing showed that there could be a sequence of physical processes that realized any finite piece of deductive reasoning simply as a consequence of its physical organization. Consequently, one's reasoning could be in accordance with the rules of logic, not by virtue of someone representing and following them, but as a result of the causal organization of the brain. This is what is meant by the expression a 'syntactic engine driving a semantic one' (see Fodor 1975, 1980; Newall 1980). As Dennett (1975) put it, intelligent processes are broken down into separate processes so stupid that a mere machine could execute them.

Note that, although the machine's operations can be defined without mentioning the meaning of symbols, this does not entail that the symbols are meaningless, any more than the fact that bachelors are defined without reference to hair entails that bachelors are bald. Indeed, computations are standardly defined over symbols that are presumed to have some meaning or other (a point often missed by critics of CTM such as Searle (1984: 33); see Chinese room argument).

How do symbols in computers acquire meaning? In the case of most artificial computers, their semantics are pretty much whatever their designers say they are. However, this might not always be so: philosophers have proposed a wide variety of conditions, involving both the internal states of a system, and their (for example, causal and co-variant) relations to external phenomena, that arguably would provide natural conditions for the states of a machine - or of an evolved animal like ourselves - having specific meanings (see Semantics, informational; Semantics, teleological; Semantics, conceptual role). These theories could provide semantic interpretations for the syntactic states postulated by CTM.

4 Other forms of reason

Deductive arguments were the inspiration, and remain the clearest success, for CTM. Similar hopes are entertained by many philosophers and cognitive scientists for a logic of induction, abduction, practical reason and decision making (see Inductive inference; Inference to the best explanation; Rationality, practical). The following affords an extremely simple example of what they have in mind.

Imagine that a computer is provided with programs that perform operations on sentences in a formal language like that of modern symbolic logic, and that it is supplied with certain hypotheses, say, about circles and squares, to which it assigns certain initial probabilities. A video camera presents it with some inputs that directly cause it to access those hypotheses, deducing from each of them, one by one, the consequences regarding the character of the input that would be expected if the hypothesis were true. These consequences are compared in each case with the actual input that is received, and that hypothesis is 'accepted' which satisfies some threshold value of a function of the initial probabilities and the highest degree of fit between its consequences and that input.

Imagine also that the machine is capable of performing certain basic actions: for example, whenever certain sentences (for example, 'Move right!') are in certain 'command' registers, then its 'arms' (say) are caused to move in a corresponding way. And suppose that the sentences that end up in these command addresses are the result of a program that spells out familiar decision-theoretic reasoning ('If more squares are wanted, and the probability of getting some by moving right is greater than by moving left, then move right') that are applied to an arbitrary set of basic preferences that are wired into the machine (say, that it prefers to replicate squares and obliterate circles).

If the machine now operated according to these programs, its actions would seem to be explainable along familiar mentalistic lines. Although reasoning in most animals is massively more complex than this, these programs would seem to capture what is essential to such processes as perception, hypothesis testing, belief, desire and decision making. That, at any rate, is the claim of CTM.

5 Gödel's incompleteness theorem

It is often thought that Gödel's famous 'incompleteness' theorem, which proved that no formal system can prove all the truths of arithmetic (see Gödel's theorems §§3-4), showed that the human mind cannot be a computer, and that CTM must therefore be false (see Lucas 1961; Penrose 1989). After all, Gödel appreciated the truth of the very sentences he is supposed to have proved to be unprovable!

There are a great many problems in this reasoning (for a start, CTM is not committed to all reasoning being deductive), but perhaps the most basic one arises from not appreciating the essentially disjunctive character of Gödel's theorem: what he proved was that *either* a formal system of arithmetic was inconsistent *or* it was incomplete. If people managed to do something that counted as proving the relevant undecidable sentence for their own system of arithmetic, they would do this at the expense of no longer being consistent: a contradiction could now be derived from their axioms. But this would not be a particularly remarkable fact: they would not explode or die. They would simply be disposed to believe a contradiction (a disposition many of us possess already). It is true that there are things no machine can do, namely, prove all the truths of arithmetic without contradicting itself; but it has yet to be established that any human being can do them either (see Putnam 1960; Lewis 1969, 1979).

6 Different modes of computation

There are many different ways in which computations can be carried out. Along lines closest to the original computers described here, there are the 'classical', 'symbolic' approaches that typically involve a 'language of thought' that is encoded and manipulated in the hardware of the computer (see Fodor 1975; Newall 1980; Language of thought). But more recently there have emerged 'non-symbolic', 'connectionist' approaches that eschew sentence-like structures, and exploit networks of interconnected cells subject to varying degrees of excitation (see, for example, Churchland and Sejnowski 1992; Connectionism). However, the two approaches need not be exclusive: through ingenious encoding, connectionist programs can be (and standardly are) run on computers constructed along classical lines, and vice versa. This is why one cannot tell simply by looking at the physical architecture of the brain whether the mind 'really' has a classical or connectionist '*cognitive* architecture'.

7 Different explanatory levels

This last point highlights an issue central to computational approaches to the mind: that there may be many different explanatory levels for describing the brain (see Explanation). Note, for example, that what is essential to the success of the computations of even our little adder (see §§2-3 above) does not depend upon the gates being realized electronically. An indefinite variety of devices (composed of pipes of water, or of mice in traps) could obey the same rules. Two primitive processors (such as gates) count as computationally equivalent if they have the same input-output function, that is, the same actual and potential behaviour, even if one works hydraulically and the other electrically. So CTM describes the organization of the brain at a level that abstracts from most of its biology.

This is not to say that the computer model is incompatible with a biological approach. Sometimes important information about how a computer works could be gained by examining its circuits. As Lycan (1981) emphasizes, there may well be indefinitely many different levels of description and explanation of a system's processes, within both CTM and biology itself.

CTM is thus a paradigm of a functionalist approach to the mind, whereby mental states are individuated by their role in an organization (see Functionalism). Unlike dualism and physicalistic reductionism, it is not committed to any particular claims about the substance out of which minds might be composed. But, unlike behaviourism, CTM is not entirely indifferent to what goes on inside the head.

In particular, although computational equivalence of primitive processors is defined in terms of their input and output, the equivalence of *non*-primitive devices is not. Consider two multipliers that work via different programs. Both accept inputs and emit outputs only in decimal notation. One of them converts inputs to binary, does the computation in binary, and then converts back to decimal. The other does the computation directly in decimal. These are not computationally equivalent multipliers despite their identical input-output functions. Thus, not only might creatures made of different stuff have the same mental lives, but behaviourally indistinguishable creatures

might have very different ones.

See also: Artificial intelligence; Chinese room argument; Cognitive architecture; Vision

NED BLOCK
GEORGES REY

References and further reading

Bechtel, W. and Abrahamsen, A. (1991) *Connectionism and the Mind*, Oxford: Blackwell.(The most comprehensive philosophical discussion of the connectionist approach to cognition.)

Block, N. (1990) 'The Computer Model of the Mind', in D. Osherson and E. Smith (eds) *An Invitation to Cognitive Science*, vol. 3, *Thinking*, Cambridge, MA: MIT Press.(A fuller exposition of CTM.)

Churchland, P.S. and Sejnowski, T.J. (1992) *The Computational Brain*, Cambridge, MA: MIT Press/Bradford Books.(An excellent introduction to the study of neural computation using artificial neural networks.)

Dennett, D. (1975) 'Why the Law of Effect Will Not Go Away', in *Brainstorms*, Cambridge, MA: MIT Press. (Discussion of how computational processes do not presuppose intelligence to execute them.)

Fodor, J. (1975) *The Language of Thought*, New York: Crowell.(Influential exposition and defence by a philosopher of the hypothesis that thought consists in computations over syntactically specified, semantically valuable representations encoded in the brain.)

Fodor, J. (1980) 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology', *Behavioral and Brain Sciences* 3: 417-24.(Argues that psychology need only concern itself with internal computations over formally specified representations, independent of their external interpretation.)

Lewis, D. (1969) 'Lucas against Mechanism', *Philosophy* 44: 231-3.(A reply to Lucas' argument against CTM.)

Lewis, D. (1979) 'Lucas Against Mechanism II', *Canadian Journal of Philosophy* 9 (3): 373-5.(A further reply to Lucas.)

Lucas, J. (1961) 'Minds, Machines and Gödel', *Philosophy* 36: 112-27.(One of the original efforts to argue against CTM on the basis of Gödel's theorem.)

Lycan, W. (1981) 'Form, Function and Feel', *Journal of Philosophy* 78: 24-50.(Useful discussion of how 'multi-realizability' recurs at innumerable levels of scientific explanation.)

Newall, A. (1980) 'Physical Symbol Systems', *Cognitive Science* 4 (2): 135-83.(Discussion by a leading computer scientist of the hypothesis that thought consists of computations defined over syntactic objects.)

Penrose, R. (1989) *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, New York: Oxford University Press.(A recent attack on CTM, based in part on Gödel's theorem.)

Putnam, H. (1960) 'Minds and Machines', in *Philosophical Papers*, vol. 2, *Mind, Language and Reality*, Cambridge: Cambridge University Press, 1975.(An early, influential statement of CTM.)

Rey, G. (1997) *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Oxford: Blackwell.(An exposition and defence of CTM as an approach to philosophy of mind quite generally.)

Searle, J. (1984) *Minds, Brains and Science*, Cambridge, MA: Harvard University Press.(Discussion of the 'Chinese room' objection to CTM.)

Mind, identity theory of

We know that the brain is intimately connected with mental activity. Indeed, doctors now define death in terms of the cessation of the relevant brain activity. The identity theory of mind holds that the intimate connection is identity: the mind is the brain, or, more precisely, mental states are states of the brain. The theory goes directly against a long tradition according to which mental and material belong to quite distinct ontological categories - the mental being essentially conscious, the material essentially unconscious. This tradition has been bedevilled by the problem of how essentially immaterial states could be caused by the material world, as would happen when we see a tree, and how they could cause material states, as would happen when we decide to make an omelette.

A great merit of the identity theory is that it avoids this problem: interaction between mental and material becomes simply interaction between one subset of material states, namely certain states of a sophisticated central nervous system, and other material states. The theory also brings the mind within the scope of modern science. More and more phenomena are turning out to be explicable in the physical terms of modern science: phenomena once explained in terms of spells, possession by devils, Thor's *thunderbolts, and so on, are now explained in more mundane, physical terms. If the identity theory is right, the same goes for the mind. Neuroscience will in time reveal the secrets of the mind in the same general way that the theory of electricity reveals the secrets of lightning. This possibility has received enormous support from advances in computing. We now have at least the glimmerings of an idea of how a purely material or physical system could do some of the things minds can do.*

Nevertheless, there are many questions to be asked of the identity theory. How could states that seem *so different turn out to be one and the same? Would neurophysiologists actually see my thoughts and feelings if they looked at my brain? When we report on our mental states what are we reporting on - our brains?*

1 Origin of the identity theory

The identity theory of mind holds that each and every mental state is identical with some state in the brain. My desire for coffee, my feeling happy, and my believing that the dog is about to bite are all states of my brain. The view is not that mental states and brain states are correlated but that they are literally one and the same. Despite its name, the identity theory of mind is strictly speaking not a view about the mind as such, but about mental phenomena. However, most protagonists of the identity theory, and most contemporary philosophers of mind if it comes to that, hold some version of the view that the mind is a construction out of its states in somewhat the way that an army is a construction out of its soldiers.

The identity theory of mind arose out of dissatisfaction with dualism, and with behaviourism as an attempt to avoid dualism. According to dualism, mental states are quite distinct from any material states, including brain states (see Dualism). The most famous challenge to dualism is to give a satisfactory account of the causal interactions between mental states and material states, and most especially to give a satisfactory account of causation from mental states to bodily occurrences. We believe that sometimes my desire for ice cream causes my arm to move in such a way that an ice cream is in my mouth, that my pangs of hunger cause me to tighten my belt, and so on. But how do states allegedly 'outside' the material world cause material goings on like arm movements? How do they do this in such a way as to avoid violating the various conservation laws in physics? And how do they do this in such a way as not to conflict with what the physical sciences, and especially neuroscience, tell us about how bodily movements are caused?

The last question is particularly pressing. The success of the physical sciences in explaining phenomena in their own, physical, terms has been striking. We now know that lightning is not caused by Thor's actions, that epilepsy is not caused by demonic possession, and that plants do not grow because they contain a vital essence but because their cells divide (see Vitalism). It is hard to believe that bodily behaviour is unique in resisting in-principle explanation in purely physical terms. The dualist can respond to this challenge by denying the common-sense view that mental states sometimes cause bodily behaviour, a position known as epiphenomenalism. This position holds that although physical states on occasion cause mental states, mental states themselves never cause anything, being mere epiphenomena of the brain states that, along with the appropriate material surroundings, are the true causes of the behaviour we associate with mental states. Apart from flying in the face of common-sense, this position makes it hard to see why the mind evolved (see Epiphenomenalism).

Behaviourism treats mental states in terms of behaviour and dispositions to behaviour. Its inspiration comes from facts such as these: that those creatures we credit with mental states are precisely those manifesting sophisticated behaviour and possessing sophisticated behavioural capacities; that psychology became a serious science when psychologists started to investigate the mind via the investigation of behavioural capacities; and that there are conceptual links between mental states and behaviour - it is, for instance, part of the concept of an intention that having an intention goes along with behaving in a way that tends to fulfil it, and it is part of the concept of intelligence that the intelligent are better problem solvers than the unintelligent (see Behaviourism, analytic).

Behaviourists delight in pointing out that it is hard to see how dualists could explain the last two points. Why should investigation of an immaterial realm be especially aided by looking at behaviour? How could the way things are in some immaterial realm be *conceptually* linked to brute behavioural facts? However, for our purposes here the crucial point is that behaviourists, like dualists, have trouble over the causation of behaviour.

The common-sense position is that mental states are causally responsible for behaviour - the itch causes the scratching - and responsible for behavioural dispositions and capacities - people's intelligence is responsible for their capacity to solve hard problems. But then mental states are not the same as behaviour and behavioural dispositions and capacities, being rather their underlying causes. And so they are, says the identity theorist. Mental states are those brain states that all the scientific evidence points to as being causally responsible for the behaviour, and behavioural dispositions and capacities, distinctive of those creatures we credit with a mental life.

2 Early objections

Here are some of the many objections that greeted the identity theory when it became well known through, especially, Smart (1959).

It was objected that the ancients knew about mental states while knowing next to nothing about the brain. How could this be if mental states are identical with brain states? However, as identity theorists observed, science has established many identities that were unknown to the ancients. They did not know that lightning is identical with an electrical discharge, temperature in gases is mean molecular kinetic energy, and water is H_2O. The identity of mental states with brain states is, identity theorists urged, of a piece with scientific identities in general. We learned, for example, that temperature is mean molecular kinetic energy by discovering that it is mean molecular kinetic energy that is responsible for the phenomena we associate with temperature. In the same general way, identity theorists claim, we have discovered that mental states are brain states by discovering that brain states are responsible for the behavioural phenomena that we associate with mental states, and we will discover which particular brain states are which particular mental states by discovering which particular brain states are responsible for the behavioural phenomena associated with those mental states.

Identity theorists also noted that water, lightning, and temperature do not present themselves to us *as* H_2O, electrical discharge, and mean molecular kinetic energy, respectively; so it cannot be an objection to their theory that mental states do not present themselves *as* brain states.

Many objected that the identity theory violates Leibniz's Law that if x = y, then x and y share all properties (see Identity), on the ground that mental states and brain states differ in their properties. An after-image is, say, yellow and in front of my face, but my brain states are not yellow, and are inside my head; again, an itch is, say, in the middle of my back, but my brain is not in the middle of my back. They also pointed out that brain states are at a certain temperature but that it is surely absurd to hold that my belief that the Earth is not flat is at a certain temperature.

Identity theorists replied by arguing that our talk about mental states is, from a logical point of view, misleading. We talk as if mental states involved relations to mental objects. We say, for example, 'I have a headache', and this apparently has the same logical structure as 'I have a car': in both cases we seem to be asserting a relation between a person, me, and an object - a headache in one case, a car in the other. However, as identity theorists noted, cars are very different from headaches: cars can exist without their owners (if I die suddenly, my car will go on existing), but headaches cannot exist independently of being experienced (if I die, any headache of mine necessarily dies with me). Headaches, and mental states in general, are necessarily someone's. In consequence, they argued, we should think of headaches in the way we think of limps: limps are not things we have when we

limp. When I say that I have a limp, I am simply saying that I limp; similarly, to say that I have a headache is to say that my head aches. Attributing properties to aches and limps is really attributing properties to ach*ings* and limp*ings*. To have a bad limp is to limp badly, and to have a winning smile is to smile winningly; likewise, to have a bad headache is to have a head that aches badly (see Mental states, adverbial theory of). Strictly speaking, there are no mental objects, and so, in particular, no after-images to be yellow and in front of faces, and no itches to be in backs; there are experiences of having after-images and of having one's back itching. These experiences are not yellow, not in front of us or in our backs, and so the fact that brain states are not yellow, not in front of us and not in our backs is no objection to the identity theorist's claim that these experiences are identical with brain states.

The discussion of belief followed a different course. There is an important distinction between my believing something, my state of belief, and what I believe. My believing that there is a tiger before me is, most likely, caused by seeing a tiger and is a state of mine, whereas what I believe - that there is a tiger before me, the *proposition* believed, as it is often put (see Propositional attitudes) - is not caused by my seeing the tiger and is not a state of mine. Again, there is my belief that the Earth is not flat, thought of as a state of mine that causes me, say, to reassure a traveller that they do not have to worry about falling off the edge, and there is the proposition that the Earth is not flat, which is *what* I accept. Now, the identity theory is not a view about the objects of belief, the propositions; it restricts itself to claiming that the state of believing is a state of the brain. And, identity theorists argued, although it would be absurd to hold that the objects of belief are at a certain temperature, it is not absurd to hold that the believings are.

Finally, some philosophers objected that the behaviour associated with having a mind, and in particular, that associated with intelligence, rationality and free action, displays a flexibility and sophistication incompatible with a purely material etiology. There are complex issues here, but we can note two serious problems for this objection. First, computers have enlarged our conception of the behavioural flexibility and sophistication compatible with a purely material etiology; and, second, it is hard to see how having an *immaterial* etiology would make any difference to the conceptual issues at stake (and, of course, quantum mechanics has broken the tie between having a material etiology and being determined).

3 Qualia

There was, however, one early objection to the identity theory that proved harder to dismiss. It concerns a perennial problem in the philosophy of mind: the nature of conscious experience and the sensuous side of psychology. But first, some stage setting.

The identity theory of mind is typically seen as part of the programme of giving a purely naturalistic or physicalistic view of the mind (see Materialism in the philosophy of mind). It is not a species of dual attribute theory of mind, according to which, although mental states are brain states they are brain states with special, non-physical properties, properties quite distinct from the kinds of properties neuro-science in particular, and the physical sciences in general, might attribute to them. The problem with dual attribute theories is that the question of how, in the light of what science is teaching us about the physical nature of the causation of behaviour, these non-physical properties could be causally relevant to behaviour seems just as pressing as the question, raised earlier, of how non-physical states could be causally relevant to behaviour. In consequence, most identity theorists see their theory as a purely naturalistic account of the mind: the denial of 'spooky' properties or attributes is as much part of the theory as the denial of 'spooky' entities or states.

The perennial objection is that the identity theory, when seen, as it should be, as part of a purely naturalistic view of the mind, leaves out the nature of conscious mental experience, the phenomenal side of psychology. We distinguish those mental states that are not associated with a characteristic 'feel' from those that are. Paradigmatic examples are belief, on the non-sensuous side, and bodily sensations and sensory perceptions, on the sensuous side. My belief that the world is round does not have a characteristic conscious feel available to introspective awareness; my itch and my sensing of a red sunset do. But, runs the objection, no amount of neurophysiological information about our brains tells us what it is like to itch, see a sunset, or smell a rose. Protagonists of this objection often use the term 'qualia' ('quale' is the singular) for the special properties that they insist that the identity theory leaves out of account. Typically, they hold that these qualia are epiphenomenal features. They avoid the implausibility of denying that pains and sensings of red *per se* are causally inefficacious but acknowledge that they have to allow that the *distinctive feel* of pains and sensings of red is causally irrelevant (see

Qualia).

4 Functionalism and the identity theory

When identity theorists first discussed the qualia objection they urged that when we itch, smell a rose, and quite generally when we are in a mental state, we are not introspectively aware of the intrinsic nature of the mental state. For they granted, as we noted earlier, that mental states do not present themselves to us as states of the brain. The identity theorists held that what happens when we introspect is that we are aware of highly relational properties of our mental states, properties like being like what goes on in me when a pin is stuck in me (in the case of pain), being like what goes on in me when I see blood and geranium petals (in the case of sensing red), and being of a kind apt for the production of scratching behaviour (in the case of itches). In bringing relational properties into the picture, the early identity theorists can be seen as precursors of functionalist theories of mind. For functionalism is a theory according to which what makes something the mental state it is is a highly relational feature of it, that feature known as its functional role.

According to functionalism, we can think of mental states as causal intermediaries between inputs from the environment, outputs in the form of behavioural responses, and other mental states. Pain, for instance, is an internal state that is typically caused by bodily damage, and typically causes the desire that it itself cease along with behavioural responses that tend to minimize the damage. The perception that there is coffee in front of me is an internal state typically caused by coffee in front of me, which in turn causes belief that there is coffee in front of me, and this belief, when combined with desire for coffee, typically causes movement towards the coffee. Mental states are, that is, specified in terms of their place in a huge network of interlocking states (of which we have just described a tiny fragment) (see Functionalism). Our concern here is with the implications of this general approach for the identity theory.

Functionalism can be, and often is, regarded as entirely consonant with the identity theory, being indeed a good way of arguing for it; or functionalism can be, and often is, regarded as a major objection to the identity theory. It all depends on the kind of identity theory.

If mental states are defined by their place in a network, then the question of what some given mental state *is* comes down to the question of what state occupies the relevant place in the network. An analogy: money can be defined in terms of a characteristic functional role, a role we are all only too familiar with through our knowledge of what you can do if you have money, and cannot do if you do not have money. In consequence, the question of what money is is the question of what plays, or occupies, the money-functional role, be it paper notes, coins, cowrie shells, or whatever. But, identity theorists observe, by far the most plausible candidates for what play the functional roles associated with mental states are various states of the brain. Thus, functionalism gives us a simple argument for the identity theory of the following structure: pain = what plays the pain-role; what plays the pain-role = a state of the brain (say, C-fibres firing); therefore, pain = C-fibres firing. And likewise for all the mental states. Identity theorists who see the identity theory as a natural offshoot of functionalism, often refer to the identity theory as ‘central state materialism’.

Approaching the identity theory through functionalism commits identity theorists to an anti-essentialist theory of mind. For the brain state that occupies the pain-role and so is, according to them, pain, might not have done so, and so might not have been pain. And the same goes for all mental states. Early expositions of the identity theory made this point by insisting that the identity of pain with, say, C-fibres firing was a *contingent* identity, and drew an analogy in this regard with the scientific identities we mentioned before. However, at least some of these identities are arguably necessarily true. It is arguable that water is necessarily H_2O. Certainly, we had to discover the identity of water and H_2O by empirical investigation, which makes it an a posteriori matter, but what we discovered was an essential feature of water. However, the identity theory must hold that the identities it posits are contingent. They are like the identity of the President of the United States in 1997 with Bill Clinton, which is contingent because that which occupies the role definitive of being President in 1997 is a contingent matter. (Dole was swimming against an economic tide, not a logical one.)

Although functionalism is a good way of arguing for the identity theory, it is a major objection to one kind of identity theory. We noted that identity theorists appeal to scientific identities in explaining and introducing their theory. These identities concern kinds or *types*. When scientists tell us that lightning is an electrical discharge, they

are not merely telling us that the instance or *token* of lightning we saw last night is an instance or token of an electrical discharge; they are telling us, in addition, about what kind of happening lightning in general is. Again, the claim that temperature is mean molecular kinetic energy is a claim about kinds or types; it tells us what the property of temperature (in gases, anyway) is - namely, mean molecular kinetic energy. This suggests that we should think of the (mind-brain) identity theory as a type-type identity theory. Moreover, the theorists' favourite illustration - 'Pain = C-fibres firing' - is a type-type identity statement.

The problem is that different types of state might occupy the pain-role in different creatures. Perhaps it is C-fibres firing in humans but D-fibres firing in dolphins. But dolphins with their D-fibres firing would then be just as much in pain as we are when our C-fibres are firing. It is the role occupied, not the occupier, that matters for being in pain according to functionalism. And the point is independently plausible. We feel sorry for dolphins that exhibit all the signs of pain despite not knowing in any detail how intrinsically alike our and their brains are. But the identity theorist cannot allow both that pain = C-fibres firing, and that pain = D-fibres firing. That would, by the transitivity of identity, lead to the false contention that C-fibres firing = D-fibres firing.

Two responses are possible. Identity theorists can retreat to a token-token identity theory. Each and every token or instance of mental state M is some token brain state, but mental types are not brain types, being instead functional types. Alternatively, they can allow that the identities between mental types and brain types may need to be restricted. Think again of the example of money. Although different types of things are money in different societies, we can make true identity claims about the types of things that are money in the different societies. For instance, money in our society = notes and coins produced by the mint, whereas money in early Polynesian society = cowrie shells (or whatever). Similarly, although temperature in gases = mean molecular kinetic energy, in substances of which the molecules do not move freely it is something else. In the same way, if indeed it is C-fibres in us but D-fibres in dolphins that play the pain-role, then identity theorists must restrict themselves to 'Pain in humans = C-fibres firing' and 'Pain in dolphins = D-fibres firing'. The question of what humans in pain and dolphins in pain have in common would remain, of course, for they would not *ex hypothesi* share the same kind of brain state. And the identity theorists' answer must be that what they would have in common would be that each has a state inside them playing the pain-role, although not the same state.

See also: Reductionism in the philosophy of mind

FRANK JACKSON

References and further reading

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.(Classic presentation of the 'central state' version of the identity theory, combined with detailed discussions of alternative theories and analyses of various key mental concepts.)

Borst, C. V. (ed.) (1970) *The Mind/Brain Identity Theory*, London: Macmillan.(Useful collection of articles for and against the identity theory.)

Jackson, F., Pargetter, R. and Prior, E. (1982) 'Functionalism and Type-Type Identity Theories', *Philosophical Studies* 42: 209-25.(Detailed discussion of the choice between type-type and token-token identity theories from a functionalist perspective.)

Lewis, D. (1972) 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50: 249-58.(Presents the identity theory in the context of a general and powerful treatment of theoretical identification in general.)

Rosenthal, D.M. (ed.) (1991) *The Nature of Mind*, London: Oxford University Press. (Large collection of important articles and extracts from books covering the philosophy of mind in general. Contains a comprehensive bibliography.)

Smart, J.J.C. (1959) 'Sensations and Brain Processes', *Philosophical Review* LXVIII: 141-56.(Classic early statement of the identity theory.)

Mind, Indian philosophy of

Despite the enormous complexity of the Indian philosophical tradition, all the different schools developed within a common worldview mapped out by the three ideas of saṃsāra, karma *and* mokṣa *(or* nirvāṇa *in the case of Buddhism). This soteriological context, which informs much of Indian philosophy, is of particular importance for the philosophy of mind, giving it a distinctive character unparalleled in the Western tradition. Speculations about the nature of mind originated in Upaniṣadic teachings that salvific knowledge comes from looking inwards. We see in the earlier Upaniṣads (from about the eighth to the fifth centuries* BC*) a careful classification of normal states of consciousness, but eventually liberation from the cycle of rebirth and suffering was framed in terms of the individual's ability to manipulate and ultimately transcend such states through the pursuit of a set of ascetic practices known as* yoga. *These ascetic practices led to a liberating state of consciousness which the Upaniṣads equated with the realization of a transcendental Self known as the* ātman. *With the development of Buddhist thought in India (from the fifth century* BC*), the philosophical tradition became divided. Generally, Buddhist schools of thought were united in their opposition to the existence of the* ātman, *whereas the so-called orthodox Hindu schools continued to favour the Upaniṣadic position. The practical quest for liberation from suffering remained central, however, to the entire philosophical tradition as the Upaniṣads gave way to a more systematic philosophizing. Subsequently, Buddhists and non-Buddhists alike continued to accept the results of meditative practice as being a legitimate concern for philosophical speculation. A dialectical relationship between theory and practice meant that philosophical disagreements created not just differences in the interpretation of meditative experiences, but also shaped such practices themselves in different ways. The apparent empirical vein of the Upaniṣads was also continued in all schools of thought, leading to richer and more detailed phenomenological classifications of experience. This rich ontological landscape is what gives Indian philosophy of mind its distinctive character.*

1 Philosophy of mind in the Upaniṣads

The Upaniṣads viewed the human individual as a composite of psychological and physical elements. Upaniṣadic thinkers set themselves the task of sifting through this psychophysical complexity in the belief that the solution to human suffering lay in some aspect of human existence itself. The urgency and practical nature of this project yielded a rich variety of classificatory schemes as one solution was discarded in favour of another. The Bṛhadāraṇyaka Upaniṣad, for example, distinguishes three generically different types of consciousness in the experiential life of all individuals. The first of these, the waking state, is characterized by pain and pleasure and its content is determined by the external world. It provides the raw materials for the second state, that of dreaming. Dreaming, however, differs from wakefulness, for in dreams we assume a greater power to construct our own worlds. Yet dreaming still subjects us to pain and pleasure and is viewed as a transition to the third state, that of deep sleep. This third state is characterized by serenity and lack of desire, and, being without content, is also a state in which the distinction between subject and object disappears. This Upaniṣadic classification of experience is thus both descriptive and hierarchical, ranking the three states of consciousness against the requirements of what a liberatory state should be. This tripartite division of everyday experience remained common currency in subsequent philosophy of mind but it was discarded as a solution to the problem of endless rebirth into a world of suffering. The Māṇḍūkya Upaniṣad thus extends the threefold classification to a fourfold classification of experience with the addition of a fourth, liberatory, state of consciousness (*turīya*). This fourth state is a specialized state of consciousness which lies beyond the everyday experience of the individual and is attainable only on the basis of yogic techniques developed for that purpose.

Parallel to the phenomenological analysis, the Upaniṣads also analysed the individual in terms of constituent elements. Many such schemes are to be found in the Upaniṣads, but perhaps the most important for later thought is that presented in the Kaṭha Upaniṣad. This uses an image of a horse-drawn chariot to present the human individual as a composite of different elements in a hierarchical relationship to each other. The horses are likened to the senses (*indriya*), the reins to the mind (*manas*), the charioteer to the intellect (*buddhi*), the chariot to the body (*śarīra*), and the passenger riding in the chariot is the Self (*ātman*). Just as a charioteer needs to gain control of his horses using the reins, so too an individual aspiring to liberation must gain control of the senses using the mental disciplines of *yoga*.

Whether it be in positing a fourth state of consciousness, or in viewing the *ātman* as a passenger in the chariot rather than as the charioteer, it is clear that Upaniṣadic thinkers sought to separate the agent of normal phenomenal experience from the transcendental *ātman*. Indeed, the very failure to distinguish between these two selves is at the root of human suffering, and liberation from rebirth demands that the individual ultimately negate and transcend empirical individuality.

Upaniṣadic thought is thus of necessity dualist in nature. If liberation is to enable the individual to transcend the spatiotemporal world, then there must be some aspect of the individual's existence which lies beyond such a world. On the other hand, if liberation is to be attainable, then it must be amenable to individual experience. There is a tension here between the need to keep these two domains apart and the need to bring them together at some point. It was left to later schools of thought to wrestle in a more systematic way with the consequences of this Upaniṣadic worldview. For example, because of its immutability, the *ātman* is not subject to rebirth. This demands an explanation of the precise mechanism of rebirth. What is it that gets reborn? It also makes the very notion of liberation itself fraught with difficulty. Liberation cannot simply be a liberation of the *ātman* because of course the *ātman* must already lie beyond the spatiotemporal world. Who then is liberated and how? The Sāōkhya-Yoga and Advaita Vedānta schools gave particular attention to solving the problems of their Upaniṣadic legacy and, in doing so, needed to address the question of the relationship between the *ātman* and the empirical individual (see Sāōkhya §§2-3, 6; Vedānta §§1-3).

The dualism of Upaniṣadic thought is distinct from the dualism between the mind and the body in Western philosophy. In Cartesian dualism, the two domains that are sharply distinguished are the mental and the physical. While Indian philosophers certainly recognized such a distinction, they were more impressed by what the mental and the physical had in common, namely that they were both thought to be causally conditioned, with cause preceding effect in a regular, predictable way. Indeed, it was the regular, predictable nature of the mind that enabled spiritual practitioners to manipulate it to their desired goal of quiescence, once the causal mechanisms were understood. Later schools of thought, therefore, developed a rich understanding of the empirical workings of the mind. This is seen in the more complex classificatory schemes that Indian philosophers developed. In the later Yoga school, for example, the fourth state of consciousness described in the Māṇḍūkya Upaniṣad becomes a complex of eight different types of meditative experience. The sharpest distinction arising from Upaniṣadic thought occurs in the opposition of the *ātman* to both the mental and the physical, inasmuch as the latter are firmly located in the empirical world.

A second important difference between the Cartesian model of the mind and that developed by orthodox Indian philosophers is the distinction the latter make between the *ātman* and the 'I' given in ordinary experience. As one of the psychophysical elements of the individual, the 'I' is just another temporary causally conditioned aspect of the individual with no ultimate value. Again, it was left to later systems of thought to work out the exact nature of the relationship between the *ātman* and the empirical sense of individuality. In Upaniṣadic and subsequent philosophy of mind, there is no transcendental ego in the Cartesian sense, only a transcendental impersonal Self.

2 Philosophy of mind in early Buddhism

Buddhism inherited and propagated the Upaniṣadic view of the human individual as a composite of psychological and physical elements, and the scholasticism of early Buddhist thought presents us with a huge variety of such classifications. Buddhism, however, abandoned the Upaniṣadic idea of an enduring Self. Indeed, the very purpose of Buddhist classifications was meant to reveal the absence of any such entity. In an early post-canonical Buddhist text, the *Milindapañha (Questions of Milinda)*, the monk Nāgasena prompts King Milinda towards the correct view. Milinda is led, through a series of questions, to realize that the term 'chariot' designates nothing more than a pole, axle, set of wheels, chariot body, banner-staff, goad and pair of reins. Similarly, the term 'Nāgasena' is but a way of designating a collection of ever-changing physical and psychological elements. The passenger who rides in the chariot of the Kaṭha Upaniṣad is here conspicuous by his absence. There is no permanent substratum underlying our existence.

A personal proper name thus refers to a stream of causally connected elements. As Buddhist thought developed, philosophers of all schools presented increasingly detailed and distinctive descriptions and classifications of these elements. Prevalent in these schemas is a broad distinction between those elements that are mental and those that are physical. This is seen in the distinction made in the early Buddhist texts between name (*nāma*) and form

(*rūpa*), and also in the widely influential classification of the elements into five categories (*skandha*). This latter classification, which purported to be an exhaustive description of the elements, divided them into five different types: physical form (*rūpa*), sensation (*vedanā*), conceptualization (*saṃjñā*), dispositions (*saṃskāra*) and consciousness (*vijñāna*). Four of the five categories relate to different types of mental events, indicating the importance of the latter, even though Buddhist philosophers from the different schools understood these categories somewhat differently.

It is important to understand exactly what Buddhists wished to deny with their 'no-self' theory (*anātma-vāda*) and this cannot be done without developing an adequate understanding of the complexity of Buddhist practice and the functions of different types of Buddhist texts. The no-self theory did not prohibit our usual ways of talking about the self in literature of a narrative and ethical nature. Where the theory is important is in the more technical, philosophical literature, where any talk of a permanent Self is prohibited by Buddhist ontology (see Buddhist philosophy, Indian §5; Buddhist concept of emptiness §1). In the great systematic treatises of Buddhism, such as the fifth-century Vaibhāṣika text, the *Abhidharmakośa (Treasury of Metaphysics)*, we see the attempt to construct a comprehensive Buddhist ontology in which the objects of conventional discourse will disappear. Vasubandhu, its author, construes ultimate reality in terms of seventy-five different types of unique, self-sufficient, momentary particulars known as *dharmas* (see Buddhism, Ābhidharmika schools of).

The no-self theory was also important in terms of its influence on meditational practice, since theory and practice were not seen as two separate activities. The Buddhist doctrine itself was an important focus for one type of meditational practice, in which the aim was to augment a mere intellectual understanding of the doctrine with an actual change in the cognitive and perceptual apparatus of the practitioner by internalizing the doctrinal categories of thought. For example, the *dharma* analysis of the Vaibhāṣika system, as well as being an intellectual enterprise, was also linked to meditational practices aimed at deconstructing the categories of everyday experience. The Buddhist rejection of the *ātman* also meant that Buddhism did not equate meditational states of consciousness with the realization of a liberatory Self. Although valuable, meditational states were still causally conditioned, and hence temporary, states of existence. As such, they could not be the final solution to the problem of suffering.

Despite their differences, it should be remembered that Buddhism and Upaniṣadic thought were both responses to a common problem, the suffering of the individual caught up in an endless round of rebirth. Buddhism rejected the idea of an enduring Self largely on the grounds that, as an idea, it was unhelpful in freeing the individual from suffering. Both traditions viewed empirical individuality as problematic. Buddhist philosophers did not frame the no-self theory as a challenge to the idea of rebirth. Rather, they themselves were challenged to address the intellectual problems that arose once these two ideas were brought together. If the theory could not be reconciled with the belief in rebirth, then so much the worse for the no-self theory. As with Upaniṣadic thought, it was left for later systematic philosophy to address some of these questions.

3 Personal identity in Indian philosophy of mind

In the Western analytic tradition, personal identity and the mind-body problem have been central and related questions in the philosophy of mind. In the context of Indian philosophy of mind, parallel questions do arise, but in such a different context that the cross-cultural philosopher cannot make any simple comparisons. Personal identity is largely problematic in Indian philosophy because of the mistaken value individuals attach to it. For both Buddhist and non-Buddhist alike, any account of personal identity must at the same time reveal its lack of absolute value for that individual. This point is especially important for the correct interpretation of karma and rebirth doctrines, which should not necessarily be interpreted as doctrines requiring the personal survival of bodily death. Buddhist philosophers, for example, did have a problem reconciling their no-self view with the belief that actions performed in one life can have repercussions in another, but it was largely a problem of explaining how the past can be causally efficacious in the present. This was not so much a problem of making sense of the notion of moral responsibility in the absence of either bodily or memory continuity, but, rather, was a mechanistic problem of bringing a cause and its temporally distant effect together. This became a particularly intractable problem given the Buddhist metaphysics of constant change and the Buddhist view of consciousness as being intentional in nature. Together these two ideas could lead to the view that an individual is but a flux of events conditioned only by the present. The 'warehouse consciousness' (*ālaya-vijñāna*) posited by the Yogācāra school of Buddhism was one attempt to solve this problem. Although momentary in nature, the warehouse consciousness was nonintentional

and thus provided a locus for ‘karmic seeds’ until they were ripe for maturation (see Buddhism, Yogācāra school of §§6-8).

Some Western philosophers have believed that a sense of self is given directly as a separate content of our experience. The word for ‘I’ in Indian philosophy is *ahaṃkāra*, meaning literally ‘I-maker’. This indicates a view common to both Buddhist and non-Buddhist philosophers, that the ‘I’ is made on the basis of something else. Indian philosophers of mind were interested in discovering how experience gives rise to this sense of ‘I’. The word *ahaṃkāra* also indicates the view that it is through our constant utterances of ‘I’ or ‘this is mine’ that we solidify our sense of empirical individuality. Thus even where non-Buddhist philosophers - for example, in the Advaita Vedānta school - used arguments strikingly similar to the *cogito ergo sum* argument of Descartes, indicating that the self really is a genuine datum of experience, they would want to add that the self given by such an argument is the self of waking consciousness, and hence a necessarily limited self in relation to the *ātman*. Philosophers of the Advaita Vedānta school were concerned to develop an understanding of the relationship between the *ātman* and the self as given in the schema of three generically different kinds of consciousness inherited from the Upaniṣads (waking, dreaming, deep sleep). Some philosophers of the Advaita Vedānta school viewed the empirical self as a distorted reflection of the *ātman*, while others viewed it as the result of limitations being placed upon the *ātman*. Ignorance (*avidyā*) is in each case responsible for the misperception (see Monism, Indian §§1-2).

Buddhist philosophers were resolute in their resistance to converting the sense of ‘I’ given in empirical experience into any kind of theoretical or absolute reality. Consciousness is a changing, conditioned phenomenon and there is no self beyond any given moment of conscious experience. Thus, for example, the *dharma* theory of the Vaibhāṣika tradition was also a meditational technique enabling practitioners to view their experience in impersonal rather than subjective terms.

4 The mental and the physical in Indian philosophy of mind

While Indian philosophers certainly recognized the distinction between the mental and the physical, both phenomenologically and often ontologically, this difference did not become of such overriding importance as it did in Western philosophy of mind. Rather, the different schools of thought focused on the complex task of mapping out the network of causes underlying the occurrence of any individual mental event, which could be the product of a complex which might include both mental and physical causes. Perceptual events, for example, were generally viewed as the result of a causal interaction involving both mental and physical causes. Six sense organs were acknowledged by Indian philosophers, those of vision, hearing, smell, taste, touch and the mind. The mind (*manas*) was classified along with other sense organs and was responsible for inner cognition, as well as being able to coordinate the other sense organs (see Sense perception, Indian views of §§1-2). Where the schools differed in their understanding of perception, it was largely on account of differences in their wider metaphysical beliefs concerning the status of physical objects and the existence of an external world. There would be no uniform response from Indian philosophers to the Lockean proposition that it is impossible to perceive without perceiving that one perceives. This was a point of considerable contention between the different schools. It was agreed, however, that waking consciousness was intentional in nature: being awake meant being awake to something. Despite the many differences between the schools and the complexity of the tradition, Indian philosophers were united in their characterization of consciousness as a complex continuum of causally conditioned mental events fuelled by the inherent restlessness of the human condition. Their need to understand the principles governing this continuum lay at the heart of Indian philosophy of mind.

See also: Awareness in Indian thought; Karma and rebirth, Indian conceptions of; Matter, Indian conceptions of §3; Self, Indian theories of; Xin (heart-and-mind)

JOY LAINE

References and further reading

Biardeau, M. (1965) ‘Ahaṃkāra in the Upaniṣads’, *Contributions to Indian Sociology* 8. (A discussion of the important distinction between the Self (*ātman*) and the ego principle (*ahaṃkāra*) in the Upaniṣads.)

Collins, S. (1982) *Selfless Persons*, Cambridge: Cambridge University Press. (Describes in rich detail the Buddhist notion of a person in the light of its no-self theory.)

Deutsch, E. (1969) *Advaita Vedānta: A Philosophical Reconstruction*, Honolulu, HI: University of Hawaii Press.

(A good philosophical account of this important orthodox school of Indian philosophy, with a chapter on the place of the Self in the overall system.)

Griffiths, P. (1986) *On Being Mindless: Buddhist Meditation And The Mind-Body Problem*, La Salle, IL: Open Court.(In discussing the problems raised by a particular type of Buddhist meditative experience, the attainment of cessation, Griffiths also succeeds in giving a comprehensive account of Buddhist theories of the mind.)

Matilal, B.K. (1986) *Perception: An Essay on Classical Indian Theories of Knowledge*, Oxford: Clarendon Press. (A comprehensive account of theories of perception and consciousness in relation to the epistemological theories of the different schools of Indian philosophy.)

Milindapañha (*c*.150 BC), trans. I.B. Horner, *Milinda's Questions*, Sacred Books of the Buddhists 22-3, London: Pali Text Society.(A good translation of an important early Theravāda text.)

Vasubandhu (5th century) *Abhidharmakośa*, ed. D. Śāstri, *Abhidharmakośa and Bhāṣya of Acārya Vasubandhu with Sphuṭārthā Commentary of Acārya Yaśomitra*, Bauddha Bharati Series 5-8, Varanasi: Bauddha Bharati, 1970-3; repr. 1981, 2 vols.(A small portion of the *Abhidharmakośabhāṣya* is translated as Appendix B in Griffiths 1986.)

Mind, philosophy of

'Philosophy of mind', and 'philosophy of psychology' are two terms for the same general area of philosophical inquiry: the nature of mental phenomena and their connection with behaviour and, in more recent discussions, the brain.

Much work in this area reflects a revolution in psychology that began mid-century. Before then, largely in reaction to traditional claims about the mind being non-physical (see Dualism; Descartes), many thought that a scientific psychology should avoid talk of 'private' mental states. Investigation of such states had seemed to be based on unreliable introspection (see Introspection, psychology of), not subject to independent checking (see Private language argument), and to invite dubious ideas of telepathy (see Parapsychology). Consequently, psychologists like B.F. Skinner and J.B. Watson, and philosophers like W.V. Quine and Gilbert Ryle argued that scientific psychology should confine itself to studying publicly observable relations between stimuli and responses (see Behaviourism, methodological and scientific; Behaviourism, analytic).

However, in the late 1950s, several developments began to change all this: (i) The experiments behaviourists themselves ran on animals tended to refute behaviouristic hypotheses, suggesting that the behaviour of even rats had to be understood in terms of mental states (see Learning; Animal language and thought). (ii) The linguist Noam Chomsky drew attention to the surprising complexity of the natural languages that children effortlessly learn, and proposed ways of explaining this complexity in terms of largely unconscious mental phenomena. (iii) The revolutionary work of Alan Turing (see Turing machines) led to the development of the modern digital computer. This seemed to offer the prospect of creating Artificial intelligence, and also of providing empirically testable models of intelligent processes in both humans and animals. (iv) Philosophers came to appreciate the virtues of realism, as opposed to instrumentalism, about theoretical entities in general.

1 Functionalism and the computational theory of mind

These developments led to the emergence in the 1970s of the loose federation of disciplines called 'cognitive science', which brought together research from, for example, psychology, linguistics, computer science, neuroscience and a number of sub-areas of philosophy, such as logic, the philosophy of language, and action theory. In philosophy of mind, these developments led to Functionalism, according to which mental states are to be characterized in terms of relations they bear among themselves and to inputs and outputs, for example, mediating perception and action in the way that belief and desire characteristically seem to do. The traditional problem of Other minds then became an exercise in inferring from behaviour to the nature of internal causal intermediaries.

This focus on functional organization brought with it the possibility of multiple realizations: if all that is essential to mental states are the roles they play in a system, then, in principle, mental states, and so minds, could be composed of (or 'realized' by) different substances: some minds might be carbon-based like ours, some might be computer 'brains' in robots of the future, and some might be silicon-based, as in some science fiction stories about 'Martians'. These differences might also cause the minds to be organized in different ways at different levels, an idea that has encouraged the co-existence of the many different disciplines of cognitive science, each studying the mind at often different levels of explanation.

Functionalism has played an important role in debates over the metaphysics of mind. Some see it as a way of avoiding Dualism and arguing for a version of materialism known as the identity theory of mind (see Mind, identity theory of). They argue that if mental states play distinctive functional roles, to identify mental states we simply need to find the states that play those roles, which are, almost certainly, various states of the brain. Here we must distinguish identifying mental state tokens with brain state tokens, from identifying mental types with brain types (see Type/token distinction). Many argue that multiple realizability shows it would be a mistake to identify any particular kind or type of mental phenomenon with a specific type of physical phenomenon (for example, depression with the depletion of norepinepherine in a certain area of the brain). For if depression is a multiply realized functional state, then it will not be identical with any particular type of physical phenomenon: different instances, or tokens, of depression might be identical with tokens of ever different types of physical phenomena (norepinephrine deletion in humans, too little silicon activation in a Martian). Indeed, a functionalist could allow (although few take this seriously) that there might be ghosts who realize the right functional organization in some

special dualistic substance. However, some identity theorists insist that at least some mental state types - they often focus on states like pain and the taste of pineapple, states with Qualia (see also the discussion below) - ought to be identified with particular brain state types, in somewhat the way that lightning is identified with electrical discharge, or water with H_2O. They typically think of these identifications as necessary a posteriori.

An important example of a functionalist theory, one that has come to dominate much research in cognitive science, is the computational theory of mind (see Mind, computational theories of), according to which mental states are either identified with, or closely linked to, the computational states of a computer. There have been three main versions of this theory, corresponding to three main proposals about the mind's Cognitive architecture. According to the 'classical' theory, particularly associated with Jerry Fodor, the computations take place over representations that possess the kind of logical, syntactic structure captured in standard logical form: representations in a so-called Language of thought, encoded in our brains. A second proposal, sometimes inspired by F.P. Ramsey's view that beliefs are maps by which we steer (see Belief), emphasizes the possible role in reasoning of maps and mental Imagery. A third, recently much-discussed proposal is Connectionism, which denies that there are any structured representations at all: the mind/brain consists rather of a vast network of nodes whose different and variable excitation levels explain intelligent Learning. This approach has aroused interest especially among those wary of positing much 'hidden' mental structure not evident in ordinary behaviour (see Ludwig Wittgenstein §3 and Daniel Dennett).

The areas that lend themselves most naturally to a computational theory are those associated with logic, common sense and practical reasoning, and natural language syntax (see Common-sense reasoning, theories of; Rationality, practical; Syntax); and research on these topics in psychology and Artificial intelligence has become deeply intertwined with philosophy (see Rationality of belief; Semantics; Language, philosophy of).

A particularly fruitful application of computational theories has been to Vision. Early work in Gestalt psychology uncovered a number of striking perceptual illusions that demonstrated ways in which the mind structures perceptual experience, and the pioneering work of the psychologist, David Marr, suggested that we might capture these structuring effects computationally. The idea that perception was highly cognitive, along with the functionalist picture that specifies a mental state by its place in a network, led many to holistic conceptions of mind and meaning, according to which parts of a person's thought and experience cannot be understood apart from the person's entire cognitive system (see Holism: mental and semantic; Semantics, conceptual role).

However, this view has been challenged recently by work of Jerry Fodor. He has argued that perceptual systems are 'modules', whose processing is 'informationally encapsulated' and hence isolatable from the effects of the states of the central cognitive system (see Modularity of mind). He has also proposed accounts of meaning that treat it as a local (or 'atomistic') property to be understood in terms of certain kinds of causal dependence between states of the brain and the world (see Semantics, informational). Others have argued further that Perception, although contentful, is also importantly non-conceptual, as when one sees a square shape as a diamond but is unable to say wherein the essential difference between a square and a diamond shape consists (see Content, non-conceptual).

2 Mind and meaning

As these last issues indicate, any theory of the mind must face the hard topic of meaning (see Semantics). In the philosophies of mind and psychology, the issue is not primarily the meanings of expressions in natural language, but of how a state of the mind or brain can have meaning or content: what is it to believe, for example, that snow is white or hope that you will win. These latter states are examples of Propositional attitudes: attitudes towards propositions such as that snow is white, or that you will win, that form the 'content' of the state of belief or hope. They raise the general issue of Intentionality, or how a mental state can be about things (for example, snow) and properties (for example, white), and, particularly, 'about' things that do not exist or will not happen, as when someone believes in Santa Claus or hopes in vain for victory.

There have been three main proposals about mental content. A state might possess a specific content: (i) by virtue of the role it plays in reasoning (see Semantics: conceptual role); (ii) by virtue of certain causal and lawful relations the state bears to phenomena in the world (see Semantics: informational; Functionalism); or (iii) by virtue of the function it plays in the evolution and biology of the organism (see Semantics: teleological; Functional

explanation). Related to these proposals are traditional philosophical interests in Concepts, although this latter topic raises complicating metaphysical concerns with Universals, and epistemological concerns with A priori knowledge.

Special problems are raised by indexical content, or the content of thoughts involving concepts expressed by, for example, 'I myself', 'here', 'now', 'this', and 'that' (see Content, indexical; Demonstratives and indexicals; Propositional attitudes §3). Does the thought that it is hot here, had in Maryland, have the same content as the thought that it is hot here, had in Canberra? The conditions under which such thoughts are true obviously depends upon the external context - for example, the time and place - of the thinking.

This dependence on external context is thought by many to be a pervasive feature of content. Drawing on recent work on reference (see Reference; Proper names), Hilary Putnam and Tyler Burge have argued that what people think, believe and so on depends not only on how they are, but also upon features of their physical and social environment. This raises the important question of whether an organism's psychology can be understood in isolation from the external world it inhabits. Defenders of methodological individualism insist that it can be (see Methodological individualism); Putnam, Burge and their supporters that it can't. Some theorists respond to the debate by distinguishing between wide and narrow content: narrow content is what 'from the skin in' identical individuals would share across different environments, whereas wide content might vary from one environment to the next (see Content: wide and narrow). These theorists then give distinctive roles to the two notions in theoretical psychology, although this is a matter of great controversy.

3 Alternatives to functionalism

Not everyone endorses functional and computational theories of mind. Some, influenced by Ryle and the later Wittgenstein, think that such concern with literally inner processes of the brain betrays a fundamental misunderstanding of mental talk, which, they argue, rests largely on outward Criteria. Others think that computational processes lack the means of capturing the basic properties of Consciousness and Intentionality that are essential to most mental phenomena. John Searle, in particular, regards his Chinese room argument as a devastating objection to computational approaches. He thinks that mental phenomena should be understood not functionally, but directly in biological or physical terms.

The hardest challenge for functionalism is posed by Qualia - the properties that distinguish pain, the look of red, the taste of pineapple, and so on, on the one hand, from mental states like belief and understanding on the other. (See also Bodily sensations; Sense-data; Perception). Some argue that unnecessary problems are produced in this area by an excessive reification of inner experience, and recommend instead an adverbial theory of mental states (see Mental states, adverbial theory of). However, some problems persist, and can be made vivid by considering the possibility of 'inverted qualia'. It seems that two people might have colour experiences that are the complements of one another (red for green, yellow for blue, etc.), even though their behaviour and functional organization are identical. This issue is explored in Colour and qualia and leads inevitably to the hard problems of Consciousness: What is it? What things have it? How do we tell? What causal role, if any, does it play the world?

There is also an issue for functionalists over Mental causation. A principal reason why Dualism has few adherents today is the problem of explaining how non-physical or non-natural phenomena can causally affect a physical world. And although some dualists retreat to Epiphenomenalism, the view that mental phenomena are caused by, but do not themselves cause any physical phenomena, this is widely seen as implausible. However, functionalists also have a problem. Even though they can and do insist that functional states are realized physically, arguably the functional states *per se* do no causing; what does the causing would seem to be the underlying physical properties of the physical realization. So, although functionalists avoid giving causal roles to the 'non-natural', it seems they must allow that mental properties *per se* do no causing.

Although the view that the mind is a natural phenomenon is now widely accepted (principally because of the causal problem for dualism), what this implies is highly contentious. Some hold that it simply means that mental phenomena supervene on physical nature in the sense that there can be no mental difference without a physical difference (see Supervenience of the mental). Donald Davidson thinks this can be true without there being any strict laws connecting the physical and the mental (see Anomalous monism). Others insist that a naturalist about the mind must reduce the mental to the physical in somewhat the way thermodynamics has been reduced to

statistical mechanics, so delivering neat lawful biconditionals linking the mental and the physical (see Reductionism in the philosophy of mind).

Much in this discussion turns on the status of Folk psychology, the theory of mind allegedly implicit in ordinary (folk) thought and talk about the mind. On one view, mental states are simply the states that fill the roles of this implicit theory, and the reduction consists in finding which internal physical states fill the roles and are, thereby, to be identified with the relevant mental states. However, defenders of Eliminativism, noting that any theory - especially a folk one - can turn out false, argue that we should take seriously the possibility that the mental states postulated by folk psychology do not exist, much as it turned out that there are no witches or phlogiston.

4 Issues in empirical psychology

Empirical psychology has figured in philosophy not only because its foundations have been discussed in the above ways, but also because some of its specific findings have been relevant to traditional philosophical claims. Thus, experiments on Split brains have undermined traditional conceptions of Personal identity (see also Mind, bundle theory of), and research on the reliability of people's self-attribution of psychological states has cast doubts on introspection as a source of specially privileged knowledge about the mind. The work of Freud on psychopathology (see Mental illness, concept of; Psychoanalysis, post-Freudian; Psychoanalysis, methodological issues in) and of Chomsky in linguistics, suggests that the states of most explanatory interest are not introspectively accessible (see Knowledge, tacit; Unconscious mental states). Chomsky's ideas also seem to revive Rationalism's postulation of innate knowledge that was long thought to have been discredited by Empiricism (see also Nativism; Innate knowledge; Language, innateness of). And they have stimulated research beyond knowledge of grammar, into infant cognition (see Cognition, infant) generally (some of which treats the Molyneux problem of whether newly sighted people would be able to recognize shapes that they had previously only touched). Other questions about the basic categories in which people understand the world have benefited from work on how these categories are understood and evolve in childhood (see Piaget, J.; Cognitive development; Moral development). A particularly important issue for the philosophy of mind concerns the origin of our mental concepts, a topic of lively current research (see Mind, child's theory of) that affects our understanding of Folk psychology.

5 Philosophy of action

Whether or not it is ultimately vindicated by empirical research, folk psychology is a rich fund of distinctions that are important in human life. The examination of them has tended to focus on issues in the explanation of Action, and, in a related vein, on psychological issues relevant to ethics (see Moral psychology).

The traditional view of action, most famously advocated by David Hume, is that an action needs both a desire and a belief. The desire provides the goal, and the belief the means of putatively achieving it (see also Reasons and causes; Desire; Belief). But what then is the role, if any, of Intention? Are intentions nothing more than some complex of belief and desire? And how, if at all, do we find a place in the Humean picture for the will? Is it something that can somehow act independently of beliefs and desires, or is it some kind of manifestation of them, some kind of 'all things considered' judgment that takes a person from dithering to action? (See Will, the.) Notoriously difficult questions in this regard concern whether there actually is anything as Free will, and how it is possible for a person to act against their better judgement, as they seem to do in cases of Akrasia, or 'weakness of will'.

Beliefs and desires seem intimately connected with many other mental states. Belief about the past is of the essence of Memory. Perception delivers belief about how things are around one, and Dreaming seems to be the having of experiences during sleep akin to (rather fragmented) perceptions in the way they tend to make you believe that certain things are happening. Even emotions and bodily sensations seem to have belief and desire components (see Emotions, nature of; Bodily sensations): anger involves both a belief that one has been wronged and a desire to do something about it, and pain involves the belief that something is amiss and the desire that it stop. Much contemporary philosophy of mind and action is concerned with teasing out the relationship between beliefs and desires and various other mental states, although approaches in cognitive science often focus upon more computationally active states, such as: noticing, deciding, and 'on line' processes of reasoning.

See also: Akan philosophical psychology; Awareness in Indian thought; Emotions, nature of; Imagination; Jung, C.G.; Materialism in the philosophy of mind; Mind, Indian philosophy of; Neutral monism; Nous; Pleasure;

Psychology, theories of; Secondary qualities; Self-deception; Vygotskii, L.S.

FRANK JACKSON
GEORGES REY

References and further reading

Braddon-Mitchell, D. and Jackson, F. (1997) *Philosophy of Mind and Cognition*, Oxford: Blackwell.(Discusses most of the live positions in the philosophy of mind; sympathetic to analytical functionalism.)

Guttenplan, S. (ed.) (1994) *A Companion to the Philosophy of Mind*, Oxford: Blackwell. (Collection of substantial essays by some the best-known figures in the philosophy of mind.)

McGinn, C. (1997) *The Character of Mind*, Oxford: Oxford University Press, 2nd edn.(Very readable introduction to philosophy of mind for the beginner, with particular emphasis on issues of consciousness and the self.)

Rey, G. (1997) *Contemporary Philosophy of Mind*, Oxford: Blackwell.(Discusses most of the live positions in the philosophy of mind; sympathetic to the representational theory of mind.)

Smith, P. and Jones, O.R. (1986) *The Philosophy of Mind*, Cambridge: Cambridge University Press.(An excellent, elementary introduction to philosophy of mind and action, from the standpoint more of traditional philosophy than of cognitive science.)

Mir Damad, Muhammad Baqir (d. 1631)

Mir Damad is primarily a gnostic philosopher, arguing that the activity of the mind makes possible the experience of spiritual visions, while visionary experience gives rise to rational thought. He brings together a variety of different traditions in Islamic philosophy, incorporating both the sort of philosophy advocated by Aristotle and its later development by the Neoplatonists, and combining them with the mystical views of Islamic thinkers. The principles of his thought are the backbone of the celebrated 'School of Isfahan', which developed this rich mixture of philosophical traditions even further. His approach to the analysis of being was a considerable extension to previous views on this subject, and enabled him to make important contributions to the notion of time. Mir Damad's philosophical style is characterized by a treatment of abstract concepts behind which lies the living experience of the mystic.

1 Life and philosophy of being

Mir Burhan al-Din Muhammad Baqir Damad, whose poetic *nom de plume* was 'Ishraq' and who was also referred to as 'the Third Master' (after Aristotle and al-Farabi) was born into a distinguished religious family. Another honorific title by which Mir Damad has been known is Sayyid al-Afadil, 'Prince of the Most Learned'.

Mir Damad was born in Astarabad but grew up in Mashhad, the religious capital of Shi'i Persia. He received his early education there, and studied Ibn Sina's texts closely (see Ibn Sina). Prior to coming to Isfahan during the reign of Shah 'Abbas, he also spent some time in Qazvin and Kashan. In Isfahan, Mir Damad continued his education, paying equal attention to rational and transmitted sciences. He died in AH 1041/AD 1631 when he fell ill on his way to Karbala, in the entourage of Shah Safi. He was buried in Najaf.

As is evident from his contemporary sources, Mir Damad was recognized simultaneously as a jurist, a mystic and a philosopher - a rare but not altogether impossible combination in Muslim intellectual history. His writings reflect his comprehensive and encyclopedic interests. He wrote on philosophy and theology, prophetic and Imami traditions, Shi'i law, Qur'anic commentary, ethics and mysticism as well as on logic. Mir Damad's ascetic exercises have been noted particularly by some of his biographers (see the introduction to the 1977 edition of *Kitab al-qabasat (Book of Embers)*, page xxviii). These exercises are combined, if his biographers' sometimes hyperbolic tone is to be believed, with a precocious attention to philosophy.

Despite his prominent status as both a mystic and a jurist, an uneasy combination made possible by certain specific features of the School of Isfahan, it was principally as a philosopher that Mir Damad recognized, praised and distinguished himself, as in his many self-praising poems, for example:

> I conquered the lands of knowledge, I lent old wisdom to my youth. So that I made the earth with my *al-Qabasat* The envy of the heavenly abodes
>
> (Hadi 1984: 134)

Al-Qabasat is Mir Damad's most significant philosophical work, containing the essence of his philosophy. *Al-Qabasat* consists of ten *qabas* ('a spark of fire') and three conclusions. Its central question is the creation of the world and the possibility of its extension from God. The first *qabas* discusses the variety of created beings and the divisions of existence. In the second he argues for a trilateral typology of essential primary notions and the primacy of essence. The duality of perspectives through which essence is subdivided and an argument to that effect through pre-eternal principles constitutes the third *qabas*. In the fourth, Mir Damad provides Qur'anic evidence, as well as references from the Prophetic and Imami traditions, to support his preceding arguments. The fifth *qabas* is devoted to a discussion of the primary dispositions through an understanding of natural existence. The connection (*ittisal'*) between 'time' and 'motion' is the subject of the sixth, where Mir Damad also argues for a 'natural order' in time. Here too he argues for the finality of numeral order and against the infinity of numbers in time-bound events (*al-hawadith al-zamaniyya*). He then devotes the seventh *qabas* to a refutation of opposing views. In the eighth, he verifies the divine authority in the establishment of such orders and the role of reason in ascertaining this truth. The ninth proves the archetypal substance of intellect (*al-jawahar al-'aqliyya*); in this chapter also, Mir Damad argues for the existence of an order in existence, a cycle of beginning and return. Finally, in the tenth *qabas*, he discusses the matter of divine ordination (*al-qada' wa al-qadar*), the necessity of

supplication, the promise of God’s reward, and the final return of all things to his judgment.

In *al-Qabasat*, Mir Damad engages in the age-old debate over the priority of ‘essence’ (*mahiyya*) versus that of ‘existence’ (*wujud*). He ultimately decides on the priority of essence, a position that would later be fundamentally disputed by his distinguished pupil Mulla Sadra. *Al-Qabasat* has remained a central text of Islamic philosophy since its first appearance. A number of philosophers of later generations have written commentaries about it, including Mulla Shamsa Gilani and Aqa Jani Mazandarani (Ashtiyani 1972). Mir Damad wrote *al-Qabasat* in response to one of his students, who had asked him to write a treatise proving that the Creator of creation and being is unique in his pre-eternality, pre-eternal in his continuity, continuous in his everlastingness, and everlasting in his post-eternality. In this text, therefore, he set himself to prove that all existent beings, from archetypal models to material manifestations, are ‘contingent upon nothingness’ (*masbuqun bi’l-‘adam*), ‘inclined towards creation’ (*tarifan bi’l-huduth*), ‘pending on annihilation’ (*marhunun bi’l-halak*), and ‘subject to cancellation’ (*mamnuwwun bi’l-butlan*) (*Kitab al-qabasat*: 1). The question of the pre-eternity (*qidam*) or createdness (*huduth*) of the world is one of the oldest and most enduring questions of Islamic philosophy, deeply rooted in the early Mu‘tazilite codification of Islamic theology (Watt 1962: 58-71; Fakhry 1983: 67-8; Leaman 1985: 11-12, 132-4). Mir Damad reminds his readers that even Ibn Sina considered the debate on this question to be ‘dialectical’ (*jadali*) rather than based on ‘proof’ (*burhan*).

For Mir Damad, being is circulated through a cycle of emanation from the divine presence to the physical world and then a return to it. In a progression of distancing emanations, the material world gradually emanates from the divine presence. From the Light of Lights (*nur al-‘anwar*) first emanate the archetypal lights (*‘anwar qahira*), of which the Universal Intellect (*‘aql-i kull*) is the first component. From this stage emanate the ‘heavenly souls’ (*nufus-i falakhyya*), the ‘ruling lights’ (*‘anwar-i mudabbira*), of which the ‘universal souls’ (*nafs-i kull*) is the primary member. The ‘natural souls’ (*nufus-i muntabi‘a*) were subsequently created by the ‘universal soul’. The archetypes of the heavens, planets, elements, compounds and the four natures were thus created. The final stage of this ontological emanation of being is the creation of matter from these archetypal origins. There is then a reverse order through which matter is sublimated back to light. Through this order, absolute or irreducible body (*jism-i mutlaq*) is advanced to the mineral stage of compound compositions. The minerals are then sublimated to the vegetative stage and then upward to the animal. Man is the highest stage of this upward mobility before the absolute matter rejoins the Light of Lights.

‘Creation’ (*ibda‘*) is the ‘bringing into being’ of something from absolute-nothing. That which is ‘evident’ (*ma’lum*), if left to its own ‘essence’ (*dhat*), would not be. It is only by virtue of something outside it (in other words, its cause) that it is or, more accurately, is brought into being. Things in their own essence have an essential, not a temporal, primacy over things that are located outside of them, such as their cause for becoming evident and manifest. Thus the secondariness of the caused over the primacy of its cause is an essential, not a temporal, secondariness. From this it follows that unless the relation between the cause and the caused is a temporal one, not every caused is created in time, that is, not every *ma‘lul* (caused) is a *muhaddath* (created-in-time). Only that caused is created-in-time which is contingent upon time (*zaman*), motion (*haraka*), and change (*taghayyur*) (*Kitab al-qabasat*: 3). That created-being which is not subsequent to time is either subsequent to absolute nothingness, whose creation is called *ibda‘* (or ‘brought into beginning’), or subsequent to not-absolute-nothingness, in which case its creation is called *ihdath* (or ‘brought into being in time’). If the created being is subsequent to time, it can have only one possibility, which is its being-in-time subsequent to its being-in-nothingness (*Kitab al-qabasat*: 3-4) (see Causality and necessity in Islamic thought).

2 Philosophy of time

There is also a hierarchical conception of time that Mir Damad begins to develop, mostly from arguments used by Ibn Sina, Nasir-i Khusrow and Khwajah Nasir Al-Tusi. First there is time (*zaman*), superior and more expansive than which is the atemporal (*dahr*) and ultimately the everlasting (*sarmad*). This hierarchy of timespan is also to be understood in terms of relationship. *Sarmad* postulates the relation of the permanent to the permanent; *dahr*, the relation of the permanent to the changing; and *zaman*, a relation of the changing to the changing. From this trilateral conception of time, Mir Damad reaches for his unique understanding of creation. Both *huduth* (creation) and *qidam* (pre-eternity) are of three kinds; *dhati* (essential), *dahri* (atemporal), and *zamani* (temporal). Essential pre-eternality (the counterpart of the essential createdness) is that whose being and actuality is not subsequential to

its not-being (*laysiyya*) and/or nothingness (*'adam*). Atemporal pre-eternality (the counterpart of the atemporal createdness) is that whose being and actuality are not subsequential to its absolute nothingness in the span of the atemporal. On the contrary, from pre-eternity it is in-being. Finally, temporal pre-eternity (the counterpart of temporal createdness) is that temporal thing whose being is not specific to a time and whose already-being (*husul*) is constantly present in the course of all time, and of whose being has no temporal beginning.

As a believing Muslim, Mir Damad must accept the createdness of being. Neither essential createdness (*al-huduth al-dhati*) nor temporal createdness (*al-huduth al-zamani*) is subject to disagreement among philosophers because they are self-evident. It is only over the question of atemporal createdness (*al-huduth al-dahri*) that disagreement arises. God's creation of the universe, Mir Damad concludes, is of the *ibda'* (brought into beginning) and *sun'* (brought into createdness) kind as it pertains to atemporal createdness and of the *ihdath* (brought into being in time) and *takwin* (brought into existence) kind as it pertains to temporal createdness.

Mir Damad proceeds to distinguish between three kinds of 'worlds'. First is the Everlasting World (*al-'alam al-sarmadi*), which is the space for Divine Presence, his essence and attributes; second is the Atemporal World (*al-'alam al-dahri*), which is the space for the pure archetypes (*al-mujarradat*); and third is the Temporal World (*al-'alam al-zamani*), which is the space for daily events, created beings and generation and corruption. There is a hierarchical relationship among these three worlds: the Everlasting World encompasses the Atemporal and the Temporal. The Temporal World is the weakest and least enduring of the three.

As temporal events are contingent upon time - that is, there are times when they are not and then they are 'produced', or brought into being, in time - the same contingency governs the hierarchical order of *sarmad* (everlasting), *dahr* (atemporality) and *zaman* (temporality). Every inferior stage, such as *zaman*, is in actual state of non-being to its superior state, in this case *dahr*. The real existence of the superior stage is identical to the actual non-being of the inferior stage. Reversing the order, the accidental defectiveness of the inferior stage - *zaman* to *dahr*, or *dahr* to *sarmad* - is not present in the superior stage. In other words, the in-itself existence of the superior stage is the *ipso facto* non-existence of the inferior stage in-itself. Mir Damad then concludes that the contingent non-being of the world of the archetypals of the *dahri* stage in the stage of *sarmadi* existence is a real and self-evident non-being. Thus all created beings and their archetypals are consequent to real and self-evident non-being. Their creation is an atemporal (*dahri*) creation and not, as theologians maintain, a temporal (*zamani*) creation. From this it follows that beyond their essential creation (*al-huduth al-dhati*), all temporal events are contingent upon and consequent to three real modes of non-existence: temporal, atemporal and everlasting. All the archetypal beings in the stage of temporal being are also contingent upon and consequent to one kind of non-being, namely the everlasting; and, of course, the everlasting world is not contingent upon or consequential to anything.

What Mir Damad achieves through this systematic separation of a trilateral stipulation of existence is the effective isolation of God at the top of the hierarchy, where he can initiate and sustain the world and yet not be subsequent to temporal corruption, to which all visible creations must yield. Moreover, the necessary contingency of an agent of creation, which is evidently active in the *zamani* and *dahri* stages of existence, is not necessary in the superior stage of *sarmadi*. As one commentator rightly observes, 'by devising the concept of *huduth-i dahri* (atemporal creation), he has succeeded in establishing a compromise between the theologian and the philosopher, in other words, between the religious law and reason' (Musawi Bihbahani in introduction to 1977 edition of *Kitab al-qabasat*: lxix).

3 Philosophical style

Mir Damad's philosophical discourse is indexical and suggestive, symbolic and referential. He relies heavily on a thorough knowledge of previous Islamic philosophy, and has a particular penchant for obscure Arabic words. The legendary difficulty of his philosophical prose should be understood as a response to the anti-philosophical climate of the period promoted by the politically powerful nomocentric jurists. Perhaps the greatest philosopher of this period, Mulla Sadra, was forced to leave the capital city of Isfahan at the instigation of the high clerical establishment precisely because of the articulate clarity of his writing. A story in *Qisas al-'ulama* is illustrative: Mulla Sadra saw Mir Damad in a dream and asked why people condemned him as a blasphemer when he had just repeated what Mir Damad had already said. 'The reason is', Mir Damad answered, 'that I wrote about philosophical matters in such a way that the religious authorities (*'ulama*) could not understand them, and that nobody other than philosophers would comprehend them. But you have vulgarized the philosophical issues and

expounded them in such a way that a teacher at an elementary religious school can understand them. That is why they have called you a blasphemer and not me' (Tunikabuni 1985: 334-5).

See also: Being; Cosmology; God, concepts of; Illuminationist philosophy; Mulla Sadra; Mystical philosophy in Islam; Soul in Islamic philosophy; Time

HAMID DABASHI

List of works

Mir Damad (before 1631) *Kitab al-qabasat (Book of Embers)*, ed. M. Mohaghegh, T. Izutsu, M. Bihbahani and I. Dibaji, Tehran: McGill University, Institute of Islamic Studies, Tehran Branch, 1977. (Important edition of Mir Damad's work, together with extensive commentaries by some of the main authorities and a valuable introduction by the translators.)

References and further reading

Ashtiyani, S. (1972) *Muntakhabati az athar-i hukama-yi ilahi-yi Iran: az 'asr-i Mir Damad va Mir Findiriski ta zaman-i hadir (Selection from Iranian Metaphysical Thinkers: Mir Damad, Mir Findiriski and Their Contemporaries)*, Tehran and Paris: Département d'Iranologie de l'Institut Franco-Iranien de Recherche. (Accounts of some of the commentaries on *al-Qabasat*.)

Corbin, H. (1972) *En Islam iranien: Aspects Spirituels et Philosophique (Iranian Islam: Spiritual and Philosophical Aspects)*, Paris: Gallimard, 4 vols.(Detailed acocunt of some of the main Persian thinkers, including Mir Damad.)

Dabashi, H. (1996) 'Mir Damad and the Founding of the School of Isfahan', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 34, 597-634.(An authoritative and detailed discussion of this crucial period of Persian thought.)

Fakhry, M. (1983) *A History of Islamic Philosophy*, New York: Columbia University Press. (A standard work on this subject.)

Hadi, A. (1984) *Sharh-i hal-i Mir Damad va Mir Findiriski (Commentary on the Thought of Mir Damad and Mir Findiriski)*, Isfahan: Maytham Tamar Publications.(Many interesting details of Mir Damad's life and contemporaries.)

Leaman, O. (1985) *An Introduction to Medieval Islamic Philosophy*, Cambridge: Cambridge University Press. (Account of some of the main philosophical and theological debates.)

Mulla Sadra, S. (*c.*1628) *al-Shawahid al-rububiyya (Divine Testimonies)*, ed. S. Ashtiyani, Mashhad: Mashhad University Press, 1967.(Discussions by Mir Damad's most famous pupil which show the differences in their views, with annotations and introduction by the editor.)

Nasr, S.H. (1966) 'The School of Isfahan', in M.M. Sharif (ed.) *A History of Muslim Philosophy*, Wiesbaden: Otto Harrassowitz, vol. 2: 904-32.(General account of the intellectual period by the leading expert.)

Nasr, S.H. (1978) *Sadr al-Din Shirazi and His Transcendent Theosophy*, Tehran: Imperial Iranian Academy of Philosophy.(Clear account of the thought of this period.)

Savory, R. (1980) *Iran: Under the Safavids*, Cambridge: Cambridge University Press.(The historical context of Mir Damad's life and works.)

Tunikabuni, M.M. (1985) *Qisas al-'ulama' (Stories of the Scholars)*, Tehran: Intisharat-'illmiyyah Islamiyyah. (Amusing and intriguing accounts of stories about thinkers including Mir Damad.)

Watt, W.M. (1962) *Islamic Philosophy and Theology*, Edinburgh: Edinburgh University Press.(Concise guide to the general set of questions which formed Mir Damad's intellectual background.)

Miracles

Does God at times miraculously intervene in earthly affairs? That is, do some events occur because God has entered our space-time continuum and directly modified or circumvented the relevant natural laws? Few philosophers today deny that this is possible. But many question whether we could ever justifiably maintain that such intervention has taken place.

According to some philosophers, it is not even necessary to grant that the types of events believers label miracles - for instance, healings or resurrections - actually occur as reported. Since the evidence supporting the occurrence of such events is the personal testimony of a few, possibly biased, individuals, while the basis for doubt is the massive amount of objective research upon which the relevant laws are based, it is always justifiable, according to this view, to conclude that such reports are erroneous. Others contend, however, that the presence of some forms of evidence - for instance, independent confirmation from reputable sources - could make it most reasonable in some cases to acknowledge that even the most unexpected of events had actually occurred.

Some philosophers also deny that we could ever justifiably conclude that an event could not have been produced by natural causes alone. Since we will never be in a position to identify all that nature can produce, they declare, it will always be most reasonable for the scientist facing a currently unexplainable counterinstance to a natural law to continue to look for a natural explanation. Many believers, however, are quite willing to grant that nature could in principle produce any event, since what they wish to maintain is only that nature does not do so in the case of miraculous interventions.

Finally, while many philosophers acknowledge that belief in direct divine intervention may at times be justifiable for those who already believe that God exists, some also argue that no single event or series of events could ever compel all thoughtful individuals to acknowledge the existence of a perfectly good supernatural causal agent, given all we experience - for instance, the tremendous amount of horrific evil in our world. Many believers, though, are also willing to grant this point.

1 Definition

The term 'miracle' is sometimes used in ordinary discussions to refer to the occurrence of any unexpected event - from the sudden discovery of a lost possession to the unanticipated passing of an exam. Within philosophical circles, however, 'miracle' is almost always discussed in its more restricted sense: as a designation for an unusual event that is the result of direct divine circumvention or modification of the natural order.

Philosophers, as well as religious believers, differ on the exact nature of the conceptual relationship between miraculous divine interventions and the natural order. For those who understand miracles to be violations of natural laws, a miracle is not simply an event that nature *did* not alone produce. It is an event that nature *could* not have produced on its own - an event that will always be incompatible with the relevant natural laws (see Laws, natural). For example, as proponents of the violation model understand it, to maintain that someone has miraculously been healed, it is not sufficient to maintain simply that God was directly involved. It is also necessary to maintain that the state of affairs in question could not have occurred naturally (that no totally natural explanation could be forthcoming).

Other philosophers, and many believers, however, deny that a miraculous divine intervention must be defined as an event for which no plausible natural explanation is, or could be, available. It is sufficient, they believe, to maintain that God was directly involved. For example, to maintain that someone's cancer has miraculously entered remission, it is not necessary to hold that nature alone could not have brought it about (to maintain that it could not have happened naturally). It is sufficient to maintain that nature alone did not do so in this case.

2 The possibility of miracles

Some philosophers (for example, McKinnon 1967) have claimed that the concept of a miracle, if defined as a violation of a natural law, is incoherent. Natural laws, they point out, are really only generalized descriptions of what does in fact happen. That is, these laws summarize for us the actual course of events. Accordingly, to claim that an occurrence is a violation of a natural law is to claim that the event in question is a suspension of the actual

course of events and this is, of course, impossible. Events may well occur, they acknowledge, that seem at present to be incompatible with how we believe things normally happen. But a true counterinstance to what we now believe to be a natural law only shows the law to be inadequate. Since natural laws, by definition, only summarize what actually occurs, we must always be willing in principle to expand our laws to accommodate any occurrence, no matter how unusual. We can never have both the exception and the rule.

Others, however, take this line of reasoning to be based on a confusion. To maintain that a natural law accurately describes the natural order, they point out, is to say only that it correctly identifies that which will occur under a specific set of natural conditions. But to maintain that an event is a miraculous counterinstance to a natural law is not to maintain that some event has occurred under the exact set of natural conditions covered by this law and nothing more. To say that water has miraculously turned into wine, for example, is not to say that water has turned into wine only under the exact set of natural conditions under which the relevant laws tell us this will not occur. It is to maintain that an additional non-natural causal factor, namely direct divine activity, was also present in this case. Accordingly, these philosophers contend, unless it is assumed that supernatural activity is impossible, it cannot be assumed that a miraculous counterinstance to a natural law - a counterinstance produced in part by divine circumvention or modification of the natural order - is conceptually impossible. That is, unless it is assumed that supernatural intervention is impossible, we can have both the exception and the rule.

Of course, many individuals do in fact deny the existence of any type of supernatural being. And even some who affirm the existence of such a being - for example, process theists (see Process theism) - deny that this being can unilaterally intervene in earthly affairs in the sense necessary to produce miraculous events. However, few philosophers today maintain that the existence of a supernatural being, or the ability of such a being (if it exists) to intervene, can be demonstrated to be impossible. That is, while most philosophers agree that the existence of a supernatural being who intervenes in earthly affairs can justifiably be denied, most also agree that it is possible to maintain justifiably that such a being does exist. Consequently, few deny that miracles, even if defined as violations of natural laws, could occur. Since the time of David Hume (§2), however, philosophers have continued to debate vigorously a number of questions related to our ability to *identify* miraculous events.

3 The credibility of personal testimony

One such question is whether we need even acknowledge that alleged counterinstances to well-confirmed natural laws actually occur. Most philosophers agree that reports of repeatable counterinstances - counterinstances that can in principle be produced by anyone under a specified set of natural conditions - cannot justifiably be dismissed. But there are a number of philosophers (most notably Flew 1961) who believe that if the events in question are nonrepeatable - if they cannot be reproduced under specifiable natural conditions - the situation is quite different. It is clearly possible, they acknowledge, that nonrepeatable counterinstances to well-confirmed natural laws have occurred (or will occur). They acknowledge, for instance, that nonrepeatable counterinstances to our current laws describing the properties of water or human tissue may have occurred (or might occur). However, the evidence supporting the adequacy of laws of this type, they point out, is very strong. These laws not only can be, but are, tested and reconfirmed daily by people with no vested interest in the outcome.

On the other hand, they are quick to add, reports of presently nonrepeatable counterinstances to such laws - a claim, for instance, that water has turned into wine or that someone has been raised from the dead - will be supported at best only by the personal testimony of a few people who may well have a vested interest in the outcome. Consequently, as long as alleged counterinstances remain nonrepeatable, we can never possess better reasons for believing that the events in question have actually occurred as reported than for believing that they have not. And therefore, following the Humean maxim that the wise person proportions belief to the evidence, these philosophers conclude that it is always justifiable to deny the accuracy of such reports.

However, there are those (for instance, Swinburne 1967) who believe that this conclusion is much too strong. They acknowledge that reports of seemingly nonrepeatable counterinstances to well-established laws must be approached with appropriate scepticism, since deception or misperception is always possible. But from their perspective it is unreasonable to assume that the evidence supporting even the most highly confirmed laws would *always* furnish a sufficient basis for dismissing reports of counterinstances to them.

First and foremost, they argue that to make this assumption fails to take into account the *prima facie* reliability of

our visual belief-forming faculties. We all rely on these faculties daily and, in general, they serve us quite well. In fact, the general reliability of such faculties must be presupposed by those formulating our natural laws. Thus, in cases where we had no reason to doubt the reliability of these belief-forming faculties - for instance, if we were to observe a seeming counterinstance ourselves or if it were directly observed by a friend whose character and objectivity were beyond question - it is not clear, they maintain, that it would always be justifiable to decide in favour of the natural laws in question, even if they were very well established.

Moreover, these philosophers add, we might in some cases have compelling physical traces to consider. In the case of an alleged healing that runs counter to well-established laws, for instance, we might have more than personal testimony. We might have objective data - photographs or videotapes or X-rays or medical records - that would stand as strong evidence for the occurrence of the event in question, evidence so convincing that it would be unreasonable to reject it. Thus they conclude that decisions concerning the accuracy of reports of alleged counterinstances - even if the events in question are nonrepeatable - must be made on a case-by-case basis.

4 Miracles as events unexplainable by natural causes

Even if some occurrences can justifiably be labelled counterinstances to our current laws, could we ever be in a position to maintain justifiably that any such event is permanently unexplainable scientifically? That is, could we ever be in a position to maintain that an acknowledged counterinstance is a state of affairs that nature could never produce on its own?

In addressing this question, it is important to clarify a potential ambiguity that has been glossed over so far in this entry. By definition, no specific state of affairs produced even in part by direct supernatural activity (by direct circumvention or modification of the natural cause/effect patterns) could ever be given a totally natural explanation. Accordingly, if we were ever in a position to maintain justifiably that some event was actually a direct act of God, we would automatically be in a position to maintain justifiably that this specific occurrence was, itself, permanently unexplainable scientifically.

As currently understood by most philosophers, however, the primary purpose of natural science is not to determine what nature has in fact produced. The main objective of science, rather, is to determine what nature is capable of producing - what can occur under solely natural conditions. For instance, the primary purpose of natural science is not to determine whether natural factors alone actually did cause any specific person's cancer to enter remission. The primary purpose of science is to determine whether natural factors alone could have done so.

Hence, when philosophers ask whether we could ever be in a position to maintain justifiably that an event is permanently unexplainable scientifically, they are not asking whether we could ever be in a position to maintain justifiably that a specific state of affairs *was* not produced by nature alone. They are asking, rather, whether we could ever be in a position to maintain justifiably that a specific event *could* not have been produced by nature alone.

In considering this question, it should first be noted that no philosopher believes that we as human beings are in a position to state with absolute certainty what nature could or could not produce on its own. All acknowledge that the scientific enterprise is continually discovering new, often startling and unexpected, information about the causal relationships that obtain in our universe. And all freely admit that the annals of science record numerous instances in which supposed counterinstances to natural laws were later demonstrated to be consistent with such laws or revisions of them.

However, as some philosophers (such as Swinburne 1967 and Holland 1965) see it, some of our natural laws are so highly confirmed that any modification we might suggest to accommodate counterinstances would be clumsy and so *ad hoc* that it would upset the whole structure of science. For example, from their perspective, to attempt to modify our current laws relating to the properties of water to allow for the possibility that water could turn into wine naturally, or to attempt to modify our current laws relating to the properties of nonliving human tissue to allow for the possibility that a dead body could be resuscitated naturally, would make these laws of little practical value. Consequently, if we were in a position to maintain justifiably that a counterinstance to a law of this type had actually occurred, we would be required, for the sake of the scientific enterprise, to maintain that this event was permanently unexplainable by natural causes - that this event could never have been produced by nature on its own.

Critics (for instance, Basinger and Basinger 1986) consider this line of reasoning to contain a false dilemma. If faced with an acknowledged counterinstance to a natural law, even one that was very highly confirmed, we would not, they contend, be required at that moment either to modify the law to accommodate the occurrence or to affirm the adequacy of the law and declare the event permanently unexplainable by natural causes. Rather, since only naturally repeatable counterinstances falsify natural laws, the appropriate initial response to the occurrence of any seeming counterinstance to any law, no matter how highly confirmed, would be to acknowledge both the law and the counterinstance while further research was undertaken.

Moreover, these critics argue that such research could never make it most reasonable to conclude that something beyond the ability of nature to produce had actually occurred. If it were discovered that the seeming counterinstance was naturally repeatable - if it were found that the event in question could be produced with regularity under some set of purely natural conditions - a revision of the relevant laws would indeed be necessary. But then this event would no longer be naturally unexplainable. On the other hand, if natural repeatability could not be achieved, the appropriate response, they contend, would still not be to maintain that this occurrence was permanently unexplainable. Since nonrepeatable counterinstances do not present us with competing hypotheses to the relevant law(s), the appropriate response, rather, would be to label the counterinstance an anomaly while continuing to accept the functional adequacy of the law(s) in question.

Even if this line of reasoning is correct, however, nothing of significance follows for those who maintain only that a miracle is an event that would not have occurred at the exact time and in the exact manner it did if God had not somehow directly circumvented or modified the natural order in the specific case in question. Only those who believe that a miracle must be a violation of a natural law - who believe that a miracle must be an event that nature could not have produced - are affected.

5 Miracles as acts of God

Regardless of the perceived relationship between miracles and nature, however, questions concerning our ability (or inability) to identify events as direct acts of God remain important. For many philosophers, the most significant question of this sort continues to be whether there are imaginable conditions under which all rational individuals would be forced to acknowledge that God has directly intervened. And although most philosophers believe the answer to be no, some (for example, Larmer 1988) believe an affirmative response is required. They acknowledge that with respect to many states of affairs which believers do in fact maintain have been brought about by God - for example, many alleged cases of divine healings - it is possible for a rational person to grant that the event has occurred as reported and yet justifiably deny that it was the result of direct divine intervention. But let us assume that someone who has been dead for twenty-four hours is raised from the dead when divine intervention is requested. Or let us assume that the missing fingers of a leper instantaneously reappear following a prayer for healing. In such cases, they argue, there would be very strong evidence supporting supernatural causation and no evidence supporting purely natural causation. In fact, the evidence would be so strong that to continue to hold out indefinitely for a totally natural explanation in such contexts would be unjustified in that this would simply demonstrate an unreasonable a priori naturalistic bias.

In response, critics (for example, Basinger and Basinger 1986) do not deny that there might be conceivable cases which, if considered in isolation, would appear to make divine intervention a very plausible causal hypothesis. However, to acknowledge that God exists and has beneficially intervened in some specific case(s), they point out, is also to acknowledge that God’s existence is compatible with all we experience - for example, that it is compatible with the tremendous amount of horrific human pain and suffering that appears to fall disproportionately on the innocent and disadvantaged. And even if it is possible to claim justifiably that God’s existence is compatible with all we experience, it cannot be argued successfully that everyone must agree. Disbelief in God also remains a justifiable response (see Evil, problem of §6). Consequently, these critics conclude, the belief that there exists a solely natural cause for any specific occurrence always remains a justifiable option, regardless of the extent to which it may appear that divine intervention was involved.

For many philosophers, though, the crucial question is not whether there are imaginable conditions under which all rational individuals would be compelled to acknowledge divine intervention but rather whether there are conditions under which those who already believe in God would be justified in doing so. Even if it is true that the

occurrence of no single event (or set of events) can justifiably compel belief in divine intervention, it is also true (so philosophers such as Wainwright 1988 and Abraham 1985 contend) that the occurrence of no event (or set of events) - for instance, no amount of evil - can rule out justified belief in God's existence as a supernatural causal agent in our world. And given this fact, it is argued, as long as believers themselves possess good theistic reasons for assuming that God has directly intervened in a given case - for instance, because the occurrence appears clearly to fit an accepted pattern of divine action - they are justified in making this assumption.

It must be added, however, that even if this is correct, an important inverse relationship between miracles and evil remains. For instance, to respond to evil by claiming that God cannot both grant humans significant freedom and yet beneficially intervene on a consistent basis is, at the same time, to cite a reason why miracles should not be expected with frequency. And to respond to evil by claiming that 'God's ways are above our ways' places the believer in a less secure position to say when and where miraculous intervention has occurred.

See also: Deism §§1-2; Nahmanides, M. §2; Occasionalism; Religion and science; Revelation §§1-2

DAVID BASINGER

References and further reading

Abraham, W.J. (1985) *An Introduction to the Philosophy of Religion*, Englewood Cliffs, NJ: Prentice Hall, ch. 13.(A very accessible response to the Humean arguments against miracles.)

Basinger, D. and Basinger, R. (1986) *Philosophy and Miracle: The Contemporary Debate*, Lewiston, NY: Edwin Mellen Press.(An accessible summary and assessment of contemporary thought on all aspects of the issue.)

Flew, A. (1961) *Hume's Philosophy of Belief*, New York: Humanities Press.(A strengthened contemporary restatement of Hume's critique.)

Holland, R.F. (1965) 'The Miraculous', *American Philosophical Quarterly* 2 (1): 43-51.(An influential defence of our ability, in principle, to identify the miraculous.)

Houston, J. (1994) *Reported Miracles: A Critique of Hume*, Cambridge: Cambridge University Press.(A thoughtful assessment of Hume's position.)

Hume, D. (1777) 'On Miracles', in *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 2nd edn, 1972.(The classic critique of miracles, used by almost all as a starting point in the discussion.)

Larmer, R.A.H. (1988) *Water into Wine*, Kingston, Ont.: McGill-Queen's University Press.(A thoughtful defence of belief in the miraculous.)

Larmer, R.A.H. (ed.) (1996) *Questions of Miracle*, Kingston, Ont.: McGill-Queen's University Press.(An accessible collection of articles on many key issues.)

McKinnon, A. (1967) '"Miracle" and "Paradox"', *American Philosophical Quarterly* 4 (3): 308-14.(An influential challenge to the coherence of the concept of miracle.)

Swinburne, R. (1967) *The Concept of Miracle*, London: Macmillan.(A short, accessible response to Humean criticism.)

Swinburne, R. (ed.) (1989) *Miracles*, New York: Macmillan.(A balanced collection of essays on all aspects of the issue.)

Wainwright, W.J. (1988) *Philosophy of Religion*, Belmont, CA: Wadsworth, ch. 2. (A brief, objective discussion of the basic issues.)

mKhas grub dge legs dpal bzang po (1385-1438)

mKhas grub dge legs dpal bzang po (Kaydrup gelek belsangbo) was one of the early masters of the dGa'-ldan-pa (Gandenba) or dGe-lugs-pa (Gelukba) school of Tibetan Buddhism. His importance derives primarily from his close association with the founder of that school, Tsong kha pa (Dzongkaba, 1357-1419), whose religious and philosophical tradition he was instrumental in preserving and transmitting. A prolific writer, whose interests ranged across the entire spectrum of Buddhist doctrine, from the exoteric (Sūtra) to the esoteric (Tantra), his work is highly regarded for its ability to encapsulate lucidly entire fields of knowledge (for example, Tantra, the Mādhyamika doctrine of emptiness, and the pramāṇa *tradition of logic and epistemology).*

mKhas grub dge legs dpal bzang po (Kaydrup gelek belsangbo; often called mKhas grub rje) was born in Western Tibet in 1385. Like his younger brother, Ba so Chos kyi rgyal mtshan (Baso chogyi gyeltsen, 1402-73), another important figure in the early dGe-lugs-pa school, he became a novice monk at an early age and studied under some of the most distinguished masters of the Sa-skya-pa school, including the great Red mda' ba gZhon nu blo gros (Rendawa Shönnu lodrö, 1349-1412), one of Tsong kha pa's own teachers, from whom mKhas grub rje received full ordination at the age of twenty (see Tsong kha pa Blo bzang grags pa). As was the custom, he made the rounds of the great Tibetan monastic universities of his day, pitting his knowledge of the Buddhist doctrinal tradition against that of his peers and establishing a reputation as a great debater. Especially important in this regard was his encounter (at a mere sixteen years of age) with the renowned Bo dong Phyogs las rnam rgyal (Bodong choklay namgyel, 1375-1450), in debate with whom he is said to have defended, with great flair, the views of Sa skya Paṇḍita (Sagya paṇḍita, 1182-1251).

In 1407, on Red mda' ba's advice, mKhas grub rje went to central Tibet in search of Tsong kha pa (Dzongkaba). An instant rapport was established between the two when they met. In the dGe-lugs-pa tradition to this day, mKhas grub rje is known as the closest of Tsong kha pa's spiritual sons. From the time of his meeting with Tsong kha pa, mKhas grub rje began to apply himself earnestly to the study of the latter's teachings, receiving instruction as well from Tsong kha pa's more senior student and direct heir, rGyal tshab dar ma rin chen (Gyeltsap darma rinchen, 1364-1432). He succeeded rGyal tshab rje as the heir to the lineage of Tsong kha pa, becoming the third holder of the throne of dGa'-ldan monastery, a position that he held until his death in 1438. mKhas grub rje's philosophical and religious views are preserved in the eleven volumes of his collected works (*gSungs 'bum*).

Most of mKhas grub rje's philosophical and religious writings have yet to be analysed by Western scholars. From the work done to date, however, we know that, following Tsong kha pa, the following themes were central to his philosophical systematization of Buddhist doctrine: (1) that the public study of the great textual tradition of India, and not the private oral instruction lineages prevalent in Tibet, should form the basis for Buddhist study and practice; (2) the importance of ethics, as epitomized in the monastic way of life, as the basis for Buddhist practice; (3) the importance of self-generation, that is, the visualization of oneself in the form of a deity, as a defining characteristic of Buddhist Tantra; (4) the fact that the emptiness that is the object of meditation in both the Sūtra (that is, the Madhyamaka) and Tantra (for example, in the Kālacakra) traditions is identical; (5) that even though the analytical understanding of emptiness must be transformed, through the practice of meditation, into direct yogic experience, the object of both the analytical and yogic insights - the emptiness itself - is the same, so that emptiness as the object of analytical/conceptual thought is not a lower or corrupt form of emptiness (as maintained by some of Tsong kha pa's opponents) but is instead, like the object of the more sublime yogic consciousness, the full-blown reality of things, the ultimate truth that is the true mode of existence of persons and phenomena; (6) that the methods of logic and epistemology as systematized in the Prāmāṇika tradition of India are indispensable to the spiritual path, and, especially, to the initial understanding of emptiness; (7) that the highest expression of the Madhyamaka theory of emptiness, that of Candrakīrti's Čast;Prāsaōgika, is compatible with the overall thrust of the Indian logical/epistemological (*pramāṇa*) tradition, and, especially, that the principle of contradiction is valid even in the Madhyamaka system, so that the understanding of emptiness does not require a deviant logic which repudiates that principle; (8) that since conventional reality is the basis for the acquisition of the merit essential to spiritual progress, emptiness is not a repudiation of the existence of things, but only the negation of a specific false quality attributed to things by the ignorance innate in all sentient beings; and (9) that systems of meditation that repudiate analysis and conceptualization represent quietist/nihilist corruptions of the Buddha's teachings.

It is important to note that much of mKhas grub rje’s philosophical work is at least partially polemical in tone. This derives on the one hand from a personal predilection for a combative style of argumentation acquired and refined on the debating grounds of Tibet’s great monasteries. On the other hand, his polemical style is a direct outcome of the task he set himself, namely, that of clearly demarcating, systematizing and defending the doctrines of his master, Tsong kha pa.

See also: Buddhist concept of emptiness; Mādhyamika: India and Tibet; Tibetan philosophy

JOSÉ IGNACIO CABEZÓN

List of works

mKhas grub dge legs dpal bzang po (1385-1438) *gSungs ’bum (Collected Works)*, ed. Lhasa Zhol, Dharamsala: Shes rig par khang, 1981; ed. sKra shis lhun po, New Delhi: Ngawang Gelek Demo, 1983.(Two different editions of the eleven volumes of mKhas grub rje’s collected works in the original Tibetan.)

mKhas grub dge legs dpal bzang po (1385-1438) *sTong thun chen mo (The Great Digest on Emptiness)*, trans. J.I. Cabezón, *A Dose of Emptiness: An Annotated Translation of the sTong thun chen mo of mKhas grub dge legs dpal bzang*, Albany, NY: State University of New York Press, 1992.(Translation of mKhas grub rje’s great Madhyamaka classic, containing an introduction to his philosophy, a brief biography based on five Tibetan sources and oral traditions of the dGe-lugs-pa school, and a bibliographical survey of literature on him.)

mKhas grub dge legs dpal bzang po (1385-1438) *rGyud sde spyi rnam (General Exposition of the Division of the Tantra)*, trans. F.D. Lessing and A. Wayman, *Mkhas grub rje’s Fundamentals of the Buddhist Tantras*, The Hague: Mouton, 1968.(A critical edition and translation, with an introduction.)

References and further reading

Hopkins, J. (1983) *Meditation on Emptiness*, London: Wisdom Publications.(A general introduction to the dGe-lugs-pa interpretation of Madhyamaka.)

Kuijp, L.W.J. van de (1985, 1986) ’Studies in the Life and Thought of Mkhas grub rje’, part 1, ’mKhas grub rje’s Epistemological Oeuvre and his Philological Remarks on Dignāga’s Pramāṇasamuccaya I’ in *Berliner Indologische Studien*, (1985) 75-105; part 4, ’mKhas-grub-rje on Regionalisms and Dialects’ in *Berliner Indologische Studien*, (1986) 23-49.(Very little secondary literature is devoted to mKhas grub rje. Of what is available, these two articles are the most interesting and accurate.)

Ruegg, D.S. (1981) *The Literature of the Madhyamaka School of Philosophy in India*, Wiesbaden: Otto Harrassowitz.(Contains some useful material on Tibet.)

Snellgrove, D. (1987) *Indo-Tibetan Buddhism*, 2 vols, Boston: Shambhala.(A useful scholarly overview, especially with regard to Tantric Buddhism.)

Williams, P. (1989) *Mahāyā Buddhism: The Doctrinal Foundation*, London and New York: Routledge.(A general overview, containing some useful material on Tibetan Buddhism.)

Modal logic

Modal logic, narrowly conceived, is the study of principles of reasoning involving necessity and possibility. More broadly, it encompasses a number of structurally similar inferential systems. In this sense, deontic logic (which concerns obligation, permission and related notions) and epistemic logic (which concerns knowledge and related notions) are branches of modal logic. Still more broadly, modal logic is the study of the class of all possible formal systems of this nature.

It is customary to take the language of modal logic to be that obtained by adding one-place operators ' $\Box$*' for necessity and '* $\Diamond$ *' for possibility to the language of classical propositional or predicate logic. Necessity and possibility are interdefinable in the presence of negation:*

$\Box A \leftrightarrow \neg\Diamond\neg A$ and

$\Diamond A \leftrightarrow \neg\Box\neg A$

hold. A modal logic is a set of formulas of this language that contains these biconditionals and meets three additional conditions: it contains all instances of theorems of classical logic; it is closed under modus ponens *(that is, if it contains A and* $A \rightarrow B$ *it also contains B); and it is closed under substitution (that is, if it contains A then it contains any substitution instance of A; any result of uniformly substituting formulas for sentence letters in A). To obtain a logic that adequately characterizes metaphysical necessity and possibility requires certain additional axiom and rule schemas:*

$$\begin{array}{rl} K & \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B) \\ T & \Box A \rightarrow A \\ 5 & \Diamond A \rightarrow \Box\Diamond A \\ Necessitation & A/\Box A. \end{array}$$

By adding these and one of the $\Box$*-* $\Diamond$ *biconditionals to a standard axiomatization of classical propositional logic one obtains an axiomatization of the most important modal logic, S5, so named because it is the logic generated by the fifth of the systems in Lewis and Langford's* Symbolic Logic *(1932). S5 can be characterized more directly by possible-worlds models. Each such model specifies a set of possible worlds and assigns truth-values to atomic sentences relative to these worlds. Truth-values of classical compounds at a world w depend in the usual way on truth-values of their components.* $\Box A$ *is true at w if A is true at all worlds of the model;* $\Diamond A$*, if A is true at some world of the model. S5 comprises the formulas true at all worlds in all such models. Many modal logics weaker than S5 can be characterized by models which specify, besides a set of possible worlds, a relation of 'accessibility' or relative possibility on this set.* $\Box A$ *is true at a world w if A is true at all worlds accessible from w, that is, at all worlds that would be possible if w were actual. Of the schemas listed above, only K is true in all these models, but each of the others is true when accessibility meets an appropriate constraint.*

The addition of modal operators to predicate logic poses additional conceptual and mathematical difficulties. On one conception a model for quantified modal logic specifies, besides a set of worlds, the set D_w *of individuals that exist in w, for each world w. For example,* $\exists x\Box A$ *is true at w if there is some element of* D_w *that satisfies A in every possible world. If A is satisfied only by existent individuals in any given world* $\exists x\Box A$ *thus implies that there are necessary individuals; individuals that exist in every accessible possible world. If A is satisfied by non-existents there can be models and assignments that satisfy A, but not* $\exists x A$*. Consequently, on this conception modal predicate logic is not an extension of its classical counterpart.*

The modern development of modal logic has been criticized on several grounds, and some philosophers have expressed scepticism about the intelligibility of the notion of necessity that it is supposed to describe.

1 History

Modal logic in the narrow sense was a topic of considerable interest to ancient and medieval philosophers. It occupied two chapters of Aristotle's *De Interpretatione*, and a substantial part of the *Prior Analytics*. Discussion of argument forms involving necessity and possibility that included, and sometimes transcended, commentary on

Aristotle's was standard fare in Hellenistic and medieval treatises on logic (see Logic, medieval). From our vantage point the ancient and medieval discussion can be interpreted as including distinctions among various kinds of possibility and necessity and investigations of the logical relations among them, as well as logical investigations of the interactions between modalities and negation, modalities and conditionals or consequence, and modalities and quantifier expressions. Aristotle determines in *De Interpretatione*, for example, that 'It may be' and 'It cannot be' are contradictories, as are 'It may not be' and 'It cannot not be' (see Logic, ancient). Furthermore, 'from the proposition ''It may be'' it follows that it is not impossible' and in one sense 'the proposition ''It may be'' follows from the proposition ''It is necessary that it should be'''. In another sense (which we might gloss as 'It is merely possible that'), 'It may be' is logically incompatible with 'It is necessary that it should be'.

Besides these purely modal principles, Aristotle and his commentators were concerned with arguments that we might think of as mixing time and modality. A notorious example is the fallacious 'sea battle' argument for determinism that he tries to debunk in *De Interpretatione* (see Aristotle §20). In addition to the admixture of temporal considerations, one should observe that the notion of necessity involved in these discussions is not likely to be the same as the one whose logical behaviour was summarized above. Aristotle himself catalogues four senses of the word 'necessary' in *Metaphysics* (V 5), and makes other distinctions elsewhere.

Although necessity and possibility have never ceased to play an important role in philosophical discourse, their logical properties were largely neglected in modern philosophy until the beginning of the twentieth century. The contemporary revival was sparked by C.I. Lewis' critique of Whitehead and Russell's *Principia Mathematica*. The logical system elaborated in *Principia* contained as theorems the formulas $p \supset (q \supset p)$ and $\sim p \supset (p \supset q)$, which Whitehead and Russell understood as asserting the apparently paradoxical propositions that if a sentence is true it is implied by any sentence and if a sentence is false it implies any sentence. Lewis maintained that these propositions, while unavoidable and unobjectionable with respect to Russell and Whitehead's understanding of implication, were false with respect to a more natural 'strict' sense of implication (see Relevance logic and entailment §§1-2).

Lewis embarked on a project of determining the appropriate axioms of strict implication with which to supplement this system. In *Principia* the 'material' implication $p \supset q$ is considered true unless *p* is true and *q* false. In Lewis' systems, the strict implication $p \Rightarrow q$ is considered true only if it is *impossible* that *p* is true and *q* false. Thus Lewis' strict implication can be defined from Russell and Whitehead's conjunction and negation signs and a new connective, '$\Diamond$', of possibility.

$$p \Rightarrow q = \sim\Diamond(p \mathbin{\&} \sim q)$$

Conversely, possibility can be defined from strict implication.

$$\Diamond p = \sim(p \Rightarrow \sim p)$$

Hence Lewis' project of finding the correct logical principles for his notion of strict implication is tantamount to that of finding the correct logical principles for possibility or, equivalently, those for necessity.

Lewis and Langford's *Symbolic Logic* (1932) describes five different axiom systems as candidates for the logic of strict implication. Much effort was expended in the first half of the century investigating these systems and variations of them. Even showing that all five are distinct (in the sense that they produce different classes of theorems) required considerable ingenuity. Fifteen years after the publication of *Symbolic Logic*, Carnap (1947) gave a non-axiomatic characterization of 'logical' necessity. $\ulcorner \text{Necessarily } A \urcorner$ is true, according to Carnap, if *A* is 'L-true', that is, if *A* is true in all state descriptions. (A 'state description' is a kind of canonical inventory of the primitive relations that hold and fail to hold of each sequence of individuals.) Thus Carnap can be seen as making precise the old idea that necessity is truth in all possible worlds. This idea is usually associated with Leibniz, but it is traced to Descartes by Curley (1984) and to Duns Scotus by Knuuttila (1982).

The logic determined by Carnap's interpretation turned out to be S5, the fifth of the Lewis and Langford systems. In the late 1950s and early 1960s, several authors proposed interpretations that refined and generalized Carnap's idea by introducing something like the accessibility relation described above (see Kanger 1957; Montague 1960; Hintikka 1963; Kripke 1963). 'Kripke models', which are essentially the models described above, are the neatest formulation of this idea. Kripke and a continuing line of successors have shown that a great variety of modal

systems can be characterized by models of this kind. This enormously simplified the kinds of investigations of axiomatic systems mentioned above and opened new lines of research.

Kripke models were particularly fruitful in the study of modal logic in the broader sense. It had long been noted that the pairs ‘it will always/sometimes be the case that’, ‘it is obligatory/permitted that’ and ‘it is known/consistent with knowledge that’ exhibit logical behaviour similar to that of ‘necessarily/possibly’. The success of Kripke’s treatment of necessity encouraged analogous treatments of these other notions.

GA (⌜It will always be the case that A⌝) is true at time *t* if *A* is true at all times after *t*.
FA (⌜It will sometimes be the case that A⌝) is true at *t* if *A* is true at some times after *t*.
OA (⌜It is obligatory that A⌝) is true at world *w* if *A* is true at all worlds at which the obligations of *w* are discharged.
PA (⌜It is permitted that A⌝) is true at *w* if *A* is true at some such worlds.
KA (⌜It is known that A⌝) is true at world *w* if *A* is true in all worlds consistent with what is known at *w*.

The resulting systems are labelled ‘tense logic’, ‘deontic logic’ and ‘epistemic logic’ to distinguish them from the original ‘alethic’ modal systems for necessity. In the last two areas this account probably takes the analogy with necessity too far, but it still serves as a useful point of departure (see Deontic logic; Epistemic logic).

Among other broadly modal systems that have received attention recently are the ‘dynamic logics’ or ‘logics of computation’ for reasoning about computer programs. Here worlds become computation states, which can be thought of as specifications of all the program variables at a particular time, and relative possibility becomes program accessibility, which holds between two states if a program can start in the first and terminate in the second. Such logics may be useful in verifying, without interminable testing, that a large program is ‘correct’, that is, that it does what it is supposed to do (see Dynamic logics).

The development of accessibility semantics led naturally to general questions about the classes of systems that can be characterized by various versions of it. Some of these are discussed below. Before the development of the accessibility interpretation, questions about the scope of axiomatic systems were often answered by devising suitable algebraic interpretations, and algebraic methods remain important tools for studying more general questions. Possible worlds semantics, however, seems less *ad hoc* than the algebraic methods. It is not clear whether they provide an analysis of necessity and possibility, or whether the notions that they incorporate, possible world and accessibility, are themselves to be analysed in terms of necessity and possibility. Either way, there is a close fit between the meanings of modal terms and their possible-worlds interpretations. The notion of possible world has proved useful in philosophical discussions on topics other than necessity and possibility: supervenience, causality, and the nature of propositions, properties and relations, to mention a few examples. The general utility of possible worlds has, in turn, inspired modal languages more expressive than the standard ‘box and diamond’ variety.

2 Propositional S5

As suggested above, S5 is the set of formulas provable from the axioms of classical propositional logic (PL) together with the axiom schemas *K*, *T*, *5* and $\Diamond A \leftrightarrow \neg\Box\neg A$ (henceforth $\mathrm{Df}\Diamond$) and the rules *modus ponens* (MP) and necessitation (Nec). It is important to understand that Nec states that $\Box A$ is a theorem if *A* is, and not that $\Box A$ is true if *A* is. Given *T* and the replacement of equivalents, the latter condition would make $\Box A$ and *A* freely interchangeable, rendering modal logic pointless.

To identify a modal system in this way with its theorems and to refer to such systems as modal logics is to follow standard, though misleading, practices. One would expect a ‘logic’ to indicate which conclusions follow logically from which premises, and perhaps the deductions by which they so follow. Furthermore, although S5 and other modal systems are intended to represent formal truths, that is, sentences true in virtue of form, it is not clear that they are intended to represent *logical* truths, that is, sentences true in virtue of logical form, for it is not clear that necessity is a logical constant (see Logical form; Logical constants). The first of these concerns can be somewhat reduced by stipulating that *A* follows from a set of formulas with respect to a logic *L* if, and only if, *L* contains a conditional whose antecedent is a conjunction of those formulas and whose consequent is *A*.

A model (for S5) is a pair (W, V) where *W* is a non-empty set (the possible worlds) and *V* is a function (the

valuation function) that assigns a truth-value (T or F) to each sentence letter and each possible world $w \in W$. If $M = (W, V)$ is a model and $w \in W$, the notion '*A* is true at *w* in *M*' (written $(M, w) \models A$) is defined inductively, the key clauses being:

$(M, w) \models \Box B$ if, for all $v \in W$, $(M, v) \models B$; and
$(M, w) \models \Diamond B$ if, for some $v \in W$ $(M, v) \models B$.

A formula is true in *M* if it is true at all possible worlds in *M*, and it is valid if it is true in all S5-models.

The 'soundness theorem' for S5 says that the semantics respects the logic in the sense that every formula in the logic is valid. The 'completeness theorem' states that every valid formula is in the logic. To say that S5 is both sound and complete with respect to the interpretation given is to say that validity and theoremhood coincide. Soundness is proved by an easy inductive argument appealing to the axiomatization: each axiom is observed to be valid and the two rules are shown to preserve validity. Completeness requires more ingenuity. A common approach is to adapt Henkin's proof of the completeness of classical logic. (For details see Lemmon (1966) or Chellas (1980).) One can get an idea of the value of an interpretation and completeness theorem by trying, with and without them, to demonstrate that $\Box p \rightarrow \Box\Box p$ is a theorem of S5.

In S5, all 'nesting' or 'iteration' of modality can be eliminated. For example, if a formula has the form $M_1 \ldots M_n A$ where each M_i is $\Box$, $\Diamond$ or negation, an equivalent formula can be obtained by deleting all the $\Box$'s and $\Diamond$'s except the innermost. Taking a modality to be a string of $\Box$'s, $\Diamond$'s and $\neg$'s, it follows that there are only six non-equivalent modalities in S5: the empty string, $\neg$, $\Box$, $\Diamond$, $\neg\Box$ and $\neg\Diamond$, corresponding to simple truth, falsity, necessity, possibility, non-necessity and impossibility.

One important schema of classical propositional logic that is not provable in S5 is extensionality:

(Ext) $(A \leftrightarrow A') \rightarrow (B \leftrightarrow B')$, where B' is the result of replacing an occurrence of the subformula *A* in *B* by A'.

For example, if *p* and *q* are both true at world *w* but only *q* is true at world *v*, then

$$(p \leftrightarrow q) \rightarrow (\Box p \leftrightarrow \Box q)$$

is false at *w*. By completeness it follows that

$$(p \leftrightarrow q) \rightarrow (\Box p \leftrightarrow \Box q)$$

is not a theorem of S5. Ext says that replacement of one subformula by another of the same truth-value will not affect the truth-value of the whole. Its failure is often viewed as characteristic of modal systems in general. Note, however, that provably equivalent formulas can be substituted for each other.

S5 is a relatively strong modal system. Its only extensions are the trivial logic, containing all instances of $\Box A \leftrightarrow A$ and, for every natural number *n*, the '*n*-possibility logic', containing all instances of

$$(\Diamond A_1 \wedge \ldots \wedge \Diamond A_{n+1}) \rightarrow \text{Dis},$$

where *Dis* is the disjunction of all formulas $\Diamond(A_i \wedge A_j)$ such that $i < j \leq n + 1$. (*n*-possibility logic is complete with respect to the class of all models with at most *n* possible worlds.)

The formulas of (propositional) S5 correspond to formulas of classical monadic predicate logic in one variable. For example, $\Diamond(p_1 \wedge \Box p_2)$ corresponds to $\exists x(P_1 x \wedge \forall x P_2 x)$. Since decision procedures for monadic predicate logic are known, this correspondence allows one to determine effectively whether a formula is in S5 by testing the corresponding monadic formula for validity. A more direct proof of decidability rests on the result that S5 has the 'finite model' property: every non-theorem is false in some model with finitely many worlds. To test whether *A* is in S5, one checks first whether *A* is provable in one step or falsifiable in a one-world model, then whether it is provable in two steps or falsifiable in a two-world model, and so on. Each step can be completed in finite time. By the finite model property, there is some step *n* at which the process yields the desired answer.

3 Philosophical questions about S5

The most important philosophical question about S5 is whether it captures the inferences and truths it was intended

to. This may depend, of course, on the kind of necessity that '$\Box$' is supposed to represent. On the usual view this is broadly logical or metaphysical necessity. Truths necessary in this sense include those true in virtue of logical form (the logical truths), those true in virtue of meaning (the analytic truths) and a more problematic category of those true in virtue of the basic nature of things. The last category has been said to contain truths of mathematics, the proposition that water is H_2O, and the proposition that Queen Elizabeth came from the egg and sperm that she did (see Essentialism). All these examples are controversial, but those who argue whether particular propositions are metaphysically necessary may nevertheless share a common conception of what it is to be metaphysically necessary. Furthermore, examples of propositions that lack metaphysical necessity seem uncontroversial: 'Napoleon invaded Russia', 'Asbestos is carcinogenic', 'Paris is the capital of France'.

The question of whether S5 (or any other system) is the right 'logic' for metaphysical necessity can be divided into two parts, corresponding to the two parts of the completeness theorem. Say that S5 is 'correct' if every theorem represents a formal truth about metaphysical necessity. Say that it is 'adequate' if every formal truth (with 'and', 'not', 'or' and 'if' as well as 'necessarily' and 'possibly') is represented by a theorem. Correctness and adequacy, then, are philosophical counterparts of soundness and completeness.

Correctness can be established by an argument similar to that for soundness: first show that the axioms represent formal truths and then that the rules transform formulas representing formal truths into formulas representing formal truths. Instances of axiom schema *T*, for example, clearly represent formal truths. ('If necessarily $87+25=112$ then $87+25=112$' is true in virtue of its form.) The difficult cases are axiom schema 5 and the rule Nec. For the former case we need to establish that sentences of the form ⌜If possibly *S* then necessarily possibly S⌝ are true in virtue of their form. One argument is as follows. To say possibly *S* is to say that ⌜$\mathrm{Not}\ S$⌝ does not follow logically or analytically from a description of the natures of things or from logical truths or analytic truths, that is, that *S* is logically and analytically consistent with the basic natures of things. But a proposition that something is consistent with logical laws, meanings and basic natures in this way is true or false in virtue of those logical laws, meanings and basic natures. So ⌜$\mathrm{Possibly}\ S$⌝, if true, is necessarily true. And since this argument does not appeal to *S*, the conditional is true in virtue of its form. An argument for the rule Nec might go as follows. Once it is established that all the axioms represent formal truths and that MP preserves formal truth, we know that everything proved without Nec represents a formal truth. Since formal truths are either logical or analytic, they are necessarily true. Since the argument that any of these sentences *S* is necessary relies only on the form of *S*, necessarily *S* is true in virtue of its form. This establishes that the first application of Nec results in formulas that represent formal truths. But if a sentence *S* proved with one application of Nec is shown to be a formal, and hence necessary, truth by an argument that appeals only to *S*'s form, then ⌜$\mathrm{Necessarily}\ S$⌝ must be true in virtue of its form. In this way, the argument can be extended to any subsequent application of Nec.

Adequacy might be established indirectly. Suppose that S5 were not adequate for metaphysical necessity. Then there would be some formal truth *S*, with 'necessarily', 'possibly', 'and' and so on, represented by a formula *A* that is not a theorem of S5. We should, then, be able to 'improve' the adequacy of S5 by adding *A* as an axiom. Furthermore, all the substitution instances of *A* will represent sentences with the form of *S*, which, if *S* is a formal truth, will also be formal truths. So we should be able to add all the substitution instances of *A* as well, obtaining an extension of S5. But, as was noted above, the extensions of S5 must contain either $\Box A \leftrightarrow A$ or, for some *n*, $(\Diamond A_1 \wedge \ldots \wedge \Diamond A_{n+1}) \rightarrow \mathrm{Dis}$. All these are incorrect for metaphysical necessity.

4 Quantified S5

Consider a language obtained by adding '$\Box$' and '$\Diamond$' and a special predicate *E* of existence to a version of predicate logic with predicates $P_1, P_2, \ldots,$ individual constants, $t_1, t_2, \ldots,$ and the identity sign, '$=$'. A model is a triple (W, D, V) where *W* is a non-empty set (the possible worlds), *D* is a function that assigns to each $w \in W$ a set D_w (the domain of *w*) and *V* is a function that assigns to each constant *t* a member of the union $\cup D$ of the sets D_w (the possible object 'denoted' by t_i) and to each world and *n*-place predicate an *n*-ary relation on $\cup D$. A definition of truth at a world is obtained from the previous one by replacing the base clause and adding clauses for '$\forall$', '$\exists$', '$=$' and *E*. The quantifier clauses state that, for example, $\forall x Px$ is true at *w* if, for every *d* in D_w, *Pt* is true at *w* when *t* denotes *d*, and that $\exists x Px$ is true at *w* if, for some *d* in D_w, *Pt* is true at *w* when *t* denotes *d*. $s=t$ is true at *w* if *s* and *t* denote the same possible objects. *Es* is true at *w* if $V(s) \in D_w$.

This interpretation reflects several choices. First, constants are ‘objectual’, that is, *V* assigns a possible object directly to each constant. $\Box Pcd$ is true at *w* if the possible objects $V(c)$ and $V(d)$ are related, at every world *w*, by $V(P, w)$. Thus $\Box Pcd$ expresses a *de re* necessity: it asserts that particular objects, independently of their descriptions, are necessarily related. This treatment of constants makes the schemas

$$(c = d) \rightarrow \Box(c = d) \qquad (\text{‘necessary identity’}) \text{ and}$$
$$\neg(c = d) \rightarrow \Box\neg(c = d) \qquad (\text{‘necessary difference’})$$

valid. Kripke has influentially argued that this is appropriate if *c* and *d* represent proper names of natural language, but generally inappropriate if they represent definite descriptions (see Proper names). The alternative to treating constants objectually is to allow their denotations to vary from world to world. An (individual) concept is a function from worlds to individuals. For example, the concept ‘first person to reach the South Pole’ might assign Amundsen to this world, Scott to the possible world in which Scott wins his race with Amundsen, and nothing to possible worlds in which the Pole is never (or always) occupied by people. We can regard a constant with a non-objectual interpretation as denoting a concept rather than an individual. On this ‘conceptual’ interpretation, necessary identity and necessary difference are not valid. Quantification may be treated conceptually as well: $\forall xA$ is declared true at *w* if every individual concept assigns to *w* an object to which *A* applies at *w*. In that case, quantified necessary identity and difference formulas,

$$\forall x \forall y(x = y \rightarrow \Box x = y) \text{ and}$$
$$\forall x \forall y(\neg(x = y) \rightarrow \Box\neg(x = y)),$$

lose their validity as well. The quantification described above, by contrast, is objectual.

Second, quantification is ‘actualist’. The truth of $\forall xPx$ at *w* requires only that the objects in D_w have the appropriate property. It follows that the Barcan formula

$$\forall x\Box Px \rightarrow \Box\forall xPx$$

and its converse

$$\Box\forall xPx \rightarrow \forall x\Box Px$$

are both invalid. Suppose, for example, that for some world *w*, *P* is a predicate that holds everywhere of just the objects that exist in *w*, and that D_u contains something not in D_w. Then at *w*, ‘Everything is necessarily *P*’ is true while ‘Necessarily, everything is *P*’ is false. Conversely, suppose that, at every world *w*, *P* holds of just the objects that exist at *w*, and that D_v contains something not in D_u. Then at *v* ‘Necessarily everything is *P*’ is true, while ‘Everything is necessarily *P*’ is false. More generally, $\forall x\Box\forall y\Box A$, $\Box\forall x\forall y\Box A$, $\Box\forall x\Box\forall yA$, and $\Box\forall x\Box\forall y\Box A$ are logically distinct.

If quantification is ‘possibilist’, that is, if $\forall xA$ means that *A* is true of all objects in $\cup D$, then these formulas are all equivalent and the Barcan formula and its converse are valid. The distinction between actualist and possibilist quantification is significant only because the models defined above have domains that vary from world to world. If one stipulates ‘constant domains’, that is, that $D_u = D_v$ for all worlds *u* and *v*, then the possible objects are just the actual ones, and the distinction collapses. Barcan and its converse are again valid.

Third, predicates can be truthfully applied to constants denoting non-actual objects. Since quantification is actualist, this implies that the classical theorems $Pc \rightarrow \exists xPx$ and $\forall xPx \rightarrow Pc$ are not valid. These principles can be retained by insisting that the application of predicates to constants denoting non-actuals is always false (defining models so that $V(P_i, w)$ is a relation on D_w). But to do so would be to adopt a kind of atomism according to which the properties and relations expressed by atomic formulas with free variables had a special status. Furthermore, although it would save the particular classical theorems above, some of their substitution instances, such as $\neg Ex \rightarrow \exists x\neg Ex$, would still fail. Another alternative is to stipulate that formulas in which predicates are applied to constants denoting non-existents lack truth-value. If validity is taken as ‘false in no model’, both the formulas and their instances are saved. On the other hand, that approach would seem to make $\Box Ec$ valid (if necessity is ‘nowhere false’) or to make $\Box(Fc \vee \neg Fc)$ invalid (if necessity is ‘everywhere true’).

Finally, domains are permitted to be empty. This implies that

$\Diamond\exists x(Fx \vee \neg Fx), \Box\exists x(Fx \vee \neg Fx)$ and $\exists x(Fx \vee \neg Fx)$

are all invalid. If one regards the first as expressing a formal truth, one can require that some world have non-empty domain; if the second, that all worlds do. If one regards the third as expressing a formal truth one can require that each model specify - in addition to the possible worlds, domains, and valuation - a particular possible world (the actual) that has a non-empty domain. Truth in a model is then redefined as truth at the actual world of the model.

5 Weaker systems

A Kripke model is a triple (W, R, V), where *W* and *V* are as in §2, and *R* (accessibility) is a binary relation on *W*. Truth at a world is defined as before except for the '$\Box$' and '$\Diamond$' clauses:

$$(M, w) \models \Box A \text{ if, for all } v \in W \text{ such that } wRv, \quad (M, v) \models A$$
$$(M, w) \models \Diamond A \text{ if, for some } v \in W \text{ such that } wRv, \quad (M, v) \models A$$

Truth in a model and validity are defined as before. The formulas valid in this sense comprise the logic K. K can be axiomatized by the schemas PL, *K* and $\mathrm{Df}\Diamond$ and the rules MP and Nec. Since it lacks the schema $\Box A \rightarrow A$, K is not adequate for necessity under any construal. It occupies an important position in modal logic in the broader sense, however, because many well-known modal systems are simple extensions of it. The systems in the leftmost column of the table, for example, are obtained by adding the schemas in the middle column to PL, *K* and $\mathrm{Df}\Diamond$ and the rules MP and Nec.

System	Characteristic axioms	Conditions on *R*
D	D: $\Box A \rightarrow \Diamond A$	seriality: $\forall x \exists y Rxy$
T	*T*	reflexivity: $\forall x Rxx$
S4	*T 4*: $\Box A \rightarrow \Box\Box A$	reflexivity transitivity: $\forall x \forall y \forall z(Rxy \mathbin{\&} Ryz \rightarrow Rxz)$
S4.3	*T, 4* H: $\Diamond A \mathbin{\&} \Diamond B \rightarrow \Diamond(A \wedge \Diamond B) \vee \Diamond(\Diamond A \wedge B) \vee \Diamond(A \wedge B)$	reflexivity, transitivity connectedness: $\forall x \forall y(Rxy \vee Ryx)$
GL	W: $\Box(\Box A \rightarrow A) \rightarrow \Box A$	transitivity no infinite chains: $Rx_1x_2 \mathbin{\&} Rx_2x_3 \mathbin{\&} \ldots \rightarrow \exists i(x_i = x_{i+1})$

The schema *D* (see the table) is formally true when '$\Box$' and '$\Diamond$' are read 'it is obligatory that' and 'it is permitted that', and the system D is known as the standard deontic logic. T was one of the earliest modal logics to be characterized precisely (see Feys 1937, 1938) and it seems to be the weakest logic in which '$\Box$' can plausibly be regarded as representing a reading of 'it is necessary that'. If, as some have suggested, the remaining S5 axioms are not correct for physical necessity, T would be a plausible candidate for the logic of that notion. S4 was the fourth of the Lewis systems. Gödel (1933) gave it a characterization like the one above and showed it to be intertranslatable with intuitionistic propositional logic. S4.3 (so named, in part, because it is intermediate in strength between S4 and S5) is correct and adequate for a reading of '$\Box$' as 'it is and always will be the case that' if time is assumed to be linear (see Tense and temporal logic). If sentence letters represent statements about numbers and '$\Box$' is interpreted as 'provable in arithmetic' then the system GL contains exactly the formulas that are themselves provable in arithmetic (where a statement about provability is 'provable' if its Gödel number is the number of a theorem of arithmetic - see Gödel's theorems). By adding all instances of $\Box A \rightarrow A$ to the theorems of GL (without allowing any new applications of the rule Nec) one obtains the system GLS, which may be regarded as the logic of arithmetic provability. Whereas GL comprises the arithmetically provable formulas about provability, GLS comprises the truths about provability (see Provability logic).

Each of these logics can also be characterized semantically, just as we have done with K. A ‘D-model’ is a Kripke model whose accessibility relation is ‘serial’, that is, every world is related to some world or other, and models appropriate for each of the other systems can be similarly defined by the conditions in the table. In each case, soundness and completeness results like those for S5 sketched above can be given.

Not all broadly ‘modal’ systems can plausibly be interpreted by Kripke models. Suppose $\Box A$ is read ‘usually *A*’. There is no relation on times such that ‘usually *A*’ is true now if *A* is true at all related times. Rather, the truth of ‘usually *A*’ depends on the number, and perhaps the distribution, of all the times at which *A* is true. This suggests a more general kind of modal semantics, one formulation of which is the ‘neighbourhood semantics’. A ‘neighbourhood model’ is a triple (W, R, V), where *W* and *V* are as before, and *R* is a relation between worlds and sets of worlds (the ‘neighbourhoods’ of those worlds, although there is no requirement that the neighbourhood of a world contain the world itself, or even that it be non-empty). The definition of truth at a world is as before except that $\Box A$ is true at *w* just in case *w* is related to the ‘truth-set’ of *A*, that is, the set of worlds at which *A* is true; and $\Diamond A$ is true at *w* if *w* is unrelated to the ‘falsity-set’ of *A*.

The modal system determined by the set of all such models is the system E, axiomatized by the schemas PL and $\mathrm{Df}\Diamond$ and rules MP and ‘equivalents’ (RE): ‘If $A \leftrightarrow B$ is provable, so is $\Box A \leftrightarrow \Box B$’. Like K, E provides a convenient base from which a variety of systems of interest can be constructed, rather than a characterization of the formal truths for some particular reading of ‘ $\Box$’. This idea can be carried even further. Operator logic (OL) is the system in the language with ‘ $\Box$’ that is axiomatized just by PL (see Kuhn 1981). The modal operators of S5, K, E and OL can be regarded as being successively more schematic and having successively less content, the theorems of OL being the modal sentences that are true in virtue of their logical form, and the stronger modal systems being theories of operators based on that logic.

6 General results

Much of the contemporary study of modal logic is directed, not towards investigating any particular one of the systems described above in more detail, but towards a general understanding of classes of such systems. The modal logics, as defined above, form a lattice structure. A Kripke model $M = (W, R, V)$ can be viewed as the addition of a valuation *V* to the ‘Kripke frame’ (W, R). *M* is then said to be ‘based on’ (W, R). Similarly, the neighbourhood model (W, N, V) is based on the neighbourhood frame (W, N). A formula is valid in a frame if it is true in all models based on the frame, and it is valid in a class of frames if it is valid in all frames of the class. A ‘frame for *L*’ is a frame in which all the theorems of *L* are valid. The set of formulas valid in a class of frames is a logic, the logic determined by that class. A logic is sound for a class of frames if every formula in the logic is valid in the class; it is sufficient for the class if every formula valid in the class is a member of the logic; it is complete (for the class) if it is sound and sufficient.

The logic determined by a class of frames is, by definition, complete for that class. It might also be complete for other classes. K, for example, is determined by (and therefore complete for) the irreflexive Kripke frames as well as all Kripke frames. Fine (1974) and Thomason (1974) independently showed that there are finitely axiomatizable extensions of K that are not determined by any class of Kripke frames. Such logics are incomplete: formulas true in every frame that verifies the axioms are unprovable. It is now known that the incompleteness phenomenon is widespread. For every extension *L* of the logic T there are uncountably many incomplete logics whose frames are exactly the frames for *L*. (This result holds for both Kripke frames and neighbourhood frames; see Benton 1985.) The incomplete logics that have been exhibited in the literature are generally complex and *ad hoc*, but one simple example is obtained from GL by replacing the first conditional of the schema *W* by a biconditional. In the other direction, results of Bull (1966) and Fine (1971) imply that every extension of S4.3 that admits the necessitation rule is complete and decidable.

Much work has also been done on the correspondences between modal and classical formulas. The schema *T* corresponds to the classical formula $\forall x Rxx$ in the sense that the frames for the former are just the first-order models for the latter. Most modal schemas that have arisen naturally in philosophical discussion correspond similarly to first-order formulas. The schemas in the second column of the table, for example, correspond to formulas in the third column. The McKinsey schema, $\Box\Diamond A \rightarrow \Diamond\Box A$, on the other hand, corresponds to no first-order formula (see van Benthem 1975; Goldblatt 1975), and no modal schema corresponds to irreflexivity:

$\forall x \neg Rxx$. (The latter fact follows from the remark above that K itself is complete for the irreflexive frames.)

The general study of modal logic encompasses a variety of other topics, including: model theory (transformations of frames and models may preserve truth-values of classes of formulas); boundary investigations (some logics have properties that all their extensions lack); expressive power (some classes of modal connectives can be defined from a few representatives); and connections with non-modal logics (such as that between S4 and intuitionistic logic). The completeness and correspondence investigations discussed above, however, have formed the core of contemporary investigations.

See also: Logical and mathematical terms, glossary of; Modal logic, philosophical issues in

STEVEN T. KUHN

References and further reading

Aristotle (*c*.mid 4th century BC) *De Interpretatione*, *Prior Analytics* and *Metaphysics*, trans. and ed. J. Barnes, in *The Revised Oxford Translation*, vols 1 and 2, Princeton, NJ: Princeton University Press, 1984.(Referred to in §1.)

Benthem, J.F. van (1975) ‘Some Connections Between Modal and Elementary Logic’, in S. Kanger (ed.) *Proceedings of the Third Scandinavian Logic Symposium*, Amsterdam: North Holland.(Referred to in §6.)

Benton, R. (1985) ’General Modal Incompleteness and Finite Axiomatizability’, Ph.D. dissertation, University of Michigan. (Referred to in §6.)

Boolos, G. (1993) *The Logic of Provability*, Cambridge: Cambridge University Press.(Thorough and readable exploration of the arithmetic provability interpretation of □ discussed in §5.)

Bull, R.A. (1966) ‘That all Normal Extensions of S4.3 Have the Finite Model Property’, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 12: 341-44.(Referred to in §6.)

Carnap, R. (1947) *Meaning and Necessity: A Study in Semantics and Modal Logic*, Chicago, IL: University of Chicago Press; 2nd, enlarged edn, 1956.(Referred to in §1.)

Chellas, B.F. (1980) *Modal Logic: An Introduction*, Cambridge: Cambridge University Press. (Referred to in §2. Careful and clear introductory text on propositional modal logic, emphasizing completeness and decidability.)

Curley, E.M. (1984) ‘Descartes on the Creation of the Eternal Truths’, *Philosophical Review* 93: 569-97.(Referred to in §1.)

Feys, R. (1937, 1938) ‘Les logiques nouvelles des modalités’, *Revue Néoscholastique de Philosophie* 40: 517-53, 41: 217-52.(Referred to in §5.)

Fine, K. (1971) ‘The Logics Containing S4.3’, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 17: 371-6.(Referred to in §6.)

Fine, K. (1974) ‘An Incomplete Logic Containing S4’, *Theoria* 40: 23-9.(Referred to in §6.)

Fine, K. and Kuhn, S. (forthcoming) *Modal Logic*.(Text and survey that is attentive to philosophical concerns, including the issues of correctness and adequacy discussed in §3.)

Gabbay, D. and Guenthner, F. (eds) (1984) *Handbook of Philosophical Logic*, vol. 2, *Extensions of Classical Logic*, Dordrecht: Kluwer.(The articles by Bull and Segerberg, van Benthem, Garson and Cocchiarella constitute an advanced survey of central topics. The volume also contains excellent articles on modality and time by R. Thomason, modality and self-reference by Smorynski, and others on topics somewhat more peripheral to the present one.)

Gödel, K. (1933) ‘Eine Interpretation des intuitionistischen Aussagenkalküls’, *Ergebnisse eines Mathematischen Kolloquiums* 4: 39-40; trans. ‘An Interpretation of the Intuitionistic Propositional Calculus’, in *Collected Works*, ed. S. Feferman, J.W. Dawson, W. Goldfarb, C.D. Parsons and R.M. Solovay, Oxford and New York: Oxford University Press, 1995, vol. 3, 296-302.(Referred to in §5.)

Goldblatt, R. (1975) ‘First-order Definability in Modal Logic’, *Journal of Symbolic Logic* 40: 35-40. (Referred to in §6.)

Hintikka, J. (1963) ‘The Modes of Modality’, *Acta Philosophica Fennica* 16: 65-81; repr. in *Models for Modalities*, Dordrecht: Kluwer, 1969.(Referred to in §1.)

Hughes, G.E. and Cresswell, M.J. (1968) *An Introduction to Modal Logic*, London: Routledge; revised edn, *A New Introduction to Modal Logic*, London: Routledge, 1996.(The classic modern textbook; still useful.)

Hughes, G.E. and Cresswell, M.J. (1984) *Companion to Modal Logic*, London: Routledge.(A sequel to the previous book that incorporates subsequent developments in the field, including an incompleteness result.)

Kanger, S. (1957) *Provability in Logic*, Stockholm: Almqvist & Wicksell.(Referred to in §1.)

Knuuttila, S. (1982) ‘Modal Logic’, in N. Kretzmann, A. Kenny and J. Pinborg (eds) *Cambridge History of Later*

Medieval Philosophy, Cambridge: Cambridge University Press.(Referred to in §1. Brief survey of some medieval thought on modality, including that of Boethius, Scotus and Buridan.)

Knuuttila, S. (1993) *Modalities in Medieval Philosophy*, London and New York: Routledge.(More comprehensive survey than Knuuttila (1982), with an extensive bibliography.)

Kripke, S.A. (1963) 'Semantical Analysis of Modal Logic I: Normal Modal Propositional Calculi', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9: 67-96.(Referred to in §1.)

Kuhn, S. (1981) 'Constants, Logical Symbols and Operator Logic', *Journal of Philosophy* 78: 487-99. (Referred to in §5.)

Lemmon, E.J. (1966) *An Introduction to Modal Logic*, ed. K. Segerberg, Oxford: Blackwell, 1977. (Referred to in §1. Draft of initial chapters of a planned collaboration with D. Scott. In fewer than 100 pages, it includes a historical survey and many applications of the Henkin completeness method.)

Lewis, C.I. and Langford, C.H. (1932) *Symbolic Logic*, New York: Century.(Appendix II, which was written by Lewis, lays out the axiom systems S1-S5 for strict implication that are mentioned in §1.)

Mints, G. (1992) *A Short Introduction to Modal Logic*, Stanford, CA: Center for the Study of Language and Information, Stanford University.(A concise treatment of the systems S5, S4 and T that emphasizes refutation procedures inspired by Gentzen's sequent calculi, rather than the axiomatic characterizations presented here.)

Montague, R. (1960) 'Logical Necessity, Physical Necessity, Ethics and Quantifiers', *Inquiry* 4: 259- 69; repr. in *Formal Philosophy*, New Haven, CT: Yale University Press, 1974.(Referred to in §1.)

Thomason, S.K. (1974) 'An Incompleteness Theorem in Modal Logic', *Theoria* 40: 30-4.(Referred to in §6.)

Modal logic, philosophical issues in

In reasoning we often use words such as 'necessarily', 'possibly', 'can', 'could', 'must' and so on. For example, if we know that an argument is valid, then we know that it is necessarily *true that if the premises are true, then the conclusion is true. Modal logic starts with such modal words and the inferences involving them. The exploration of these inferences has led to a variety of formal systems, and their interpretation is now most often built on the concept of a possible world.*

Standard non-modal logic shows us how to understand logical words such as 'not', 'and' and 'or', which are truth-functional. The modal concepts are not truth-functional: knowing that p is true (and what 'necessarily' means) does not automatically enable one to determine whether 'Necessarily p' is true. ('It is necessary that all people have been people' is true, but 'It is necessary that no English monarch was born in Montana' is false, even though the simpler constituents - 'All people have been people' and 'No English monarch was born in Montana'- are both true.)

The study of modal logic has helped in the understanding of many other contexts for sentences that are not truth-functional, such as 'ought' ('It ought to be the case that p') and 'believes' ('Alice believes that p'); and also in the consideration of the interaction between quantifiers and non-truth-functional contexts. In fact, much work in modern semantics has benefited from the extension of modal semantics introduced by Richard Montague in beginning the development of a systematic semantics for natural language.

The framework of possible worlds developed for modal logic has been fruitful in the analysis of many concepts. For example, by introducing the concept of relative possibility, Kripke showed how to model a variety of modal systems: a proposition is necessarily true at a possible world w if and only if it is true at every world that is possible relative to w. To achieve a better analysis of statements of ability, Mark Brown adapted the framework by modelling actions with sets of possible outcomes. John has the ability to hit the bull's-eye reliably if there is some action of John's such that every possible outcome of that action includes John's hitting the bull's-eye.

Modal logic and its semantics also raise many puzzles. What makes a modal claim true? How do we tell what is possible and what is necessary? Are there any possible things that do not exist (and what could that mean anyway)? Does the use of modal logic involve a commitment to essentialism? How can an individual exist in many different possible worlds?

1 What makes a modal claim true?

Alice ate vegetable soup today. What makes that true? Alice's activities and their relationship to vegetable soup. Alice *could have* eaten cream of tomato soup. What makes that true? This is less clear. No activity of Alice's seems to make this statement about possibilities true.

If we ask what makes something possible, the context of the question sometimes sets a standard: it was possible for Alice to have eaten cream of tomato soup because it was on the menu, there was plenty of it available, there was nothing barring her from ordering it and nothing standing in the way of the waiter's bringing it to her. Yet it seems that there is a more absolute sense of possibility, according to which it was at least logically possible for Alice to have eaten cream of rutabaga soup (even if no one had ever made such a thing), because no logical or conceptual barrier prevents its existence or her eating it.

In contrast, Alice could not eat cream of tomato soup without eating it, and she could not have vegetable soup that involved no vegetables at any stage of its preparation. For logical or conceptual reasons, these are not activities open to Alice. Not every description picks out something that is possible.

Once we get a view of this broader notion of possibility, it seems that possibility can get a negative characterization. A description of an event or state of affairs describes something possible if no logical or conceptual consideration rules out what it describes. It is possible if it is not impossible; and the impossible is whatever is ruled out by logical or conceptual considerations. If a proposition is necessary, then its negation is impossible. The proposition

If Alice eats tomato soup, then Alice eats tomato soup

is necessary; and so a negation of that, a proposition equivalent to

Alice eats tomato soup without eating tomato soup,

is impossible. It is necessary that

All vegetable soup involves vegetables in its preparation;

so it is impossible that

Something is vegetable soup and does not involve vegetables in its preparation.

Understanding what is impossible will follow upon an understanding of what is necessary.

All of this suggests that we can understand what makes something possible if we can get a general understanding of what makes something a necessary truth. An event or state of affairs *p* is not possible if and only if *p* is impossible; *p* is impossible if and only if the negation of *p*, $\sim p$, is necessary. Ordinarily the box symbol is used before a sentence to stand for necessity (‘ $\Box p$’ stands for ‘It is necessary that *p*’) and the diamond is used to stand for possibility (‘ $\Diamond p$’ stands for ‘It is possible that *p*’), and we have just taken note of these related equivalences:

$\Diamond p$ iff $\sim\Box\sim p$.
$\sim\Diamond p$ iff $\Box\sim p$.

Because of these logical relationships between possibility and necessity, we can understand what makes something possible by developing an understanding of what makes something necessary. A state of affairs is possible if and only if its negation is not necessary. So what makes something a necessary truth? Our examples so far have consisted of logical and conceptual truths, but it seems that the category of necessary truths is broader. My pet cat could not exist without being an animal; it is necessary that if she exists, then she is an animal. I could not have existed if a certain sperm and ovum had never existed; it is necessary that if I exist, then those cells existed. No right hand can precisely fill the space taken up by any left hand; it is necessary that if a space *S* is filled by a left hand, then *S* cannot be precisely filled by any right hand. Logic and definitions do not seem to suffice as the source of the necessity of these truths.

In exploring logical relationships, logicians have introduced the concept of a possible world. This has been of enormous benefit in understanding logical relationships among modal claims. A statement is a necessary truth if and only if it is true at every possible world; a statement is possibly true if and only if it is true at at least one possible world. Thus our symbols can be connected with a systematic semantics for modal language in which possible worlds are reference points:

Any sentence of the form ‘ $\Box p$’ is true iff *p* is true at every possible world.
Any sentence of the form ‘ $\Diamond p$’ is true iff *p* is true at at least one possible world.

Kripke’s elaboration of this, employing the notion of relative possibility, has been very important in exploring a range of modal systems that give a range of different results concerning the logical relationships among modal sentences (see Kripke, S.A.; Modal logic). With further elaboration, introducing a ‘closeness’ relation, this framework has been extended (see Lewis 1973, Stalnaker 1968) to provide a better semantics for counterfactuals: roughly, a counterfactual ‘If it were the case that *p*, then it would be the case that *q*’ is true at *w* if and only if *q* is true at all of the closest worlds to *w* in which *p* is true.

Such systematizations seem to give us an account of the necessity and possibility of propositions that can help us to see more clearly what is and what is not possible. This account, however, is not an ultimate answer to the question of what makes something possible, because there is no independent account of what a possible world is. We are left with the question of what makes it the case that Alice’s eating rutabaga soup is included in some possible worlds, whereas my cat’s being a plant is not. At a possible world, every necessary truth must be true, every state of affairs which obtains must be possible, and the possible world must meet some maximality condition (for example, that every proposition or its negation must be true at that world) - but then judging what is included in a possible world takes us back to the question of judging what is possible. So we have a circle of concepts. The

exploration of such a circle of concepts can be very enlightening concerning their logical relationships, and it can help us to make more consistent judgments about the application of the concepts, but it cannot give an independent answer to the question of what makes a modal statement true (see Possible worlds; Semantics, possible worlds).

Thus we are led to a root philosophical problem about modality, the problem of giving a general account of what makes modal claims and related claims true - or, which amounts to the same problem, the problem of saying in general what makes a set of propositions true of a possible world (or what makes a set of states of affairs constitute a possible world). Any answer that is non-modal seems inadequate, and any answer that is modal merely enlarges the circle. Perhaps the kind of understanding that we get by enlarging the circle and exploring the relationships among modal concepts is the best we can hope for (see Blackburn 1993).

2 The merely possible

Even when we work within the circle of modal concepts, attempting to get a coherent, systematic semantics for modal claims, we come upon many puzzles. Understanding the relationship between the actual and the possible is at the root of much of the difficulty.

It seems clear that there could have been more carrier pigeons. If people had not shot so many of them, they would be flying now, perhaps in flocks so large that they would darken the skies. In saying this - in using the word ‘they’ to refer to some carrier pigeons that *would be* flying now - I seem to be referring to carrier pigeons that do not exist. Are there carrier pigeons that do not exist, that is, some nonexistent, merely possible carrier pigeons? We should hope not. Yet in a systematic treatment of modality, it is difficult to avoid the problem of requiring the existence of merely possible objects, and various strategies have been employed to deal with this problem. (Not everyone regards a commitment to nonexistent objects as a problem; see Meinong, A.)

One strategy is to note a distinction between existence and actuality. Some things - possibilia - exist without being actual, and the merely possible carrier pigeons are among them. In fact, it might be said that if possible worlds exist without being actual, why should we regard it as a new problem to accept the existence of the things that make up those worlds, such as possible carrier pigeons? If the possible worlds exist, it seems that the things in them should exist as well. The use of modal language commits us to both possible worlds and possibilia, according to this view.

The view that there are some merely possible objects, just as there are merely possible worlds, seems grounded in a puzzle, however, and we can separate that view from the thesis that other possible worlds exist. A distinction between existence and actuality makes sense for possible worlds, on most conceptions, because a possible world is really an abstract representation; a way things might be, rather than a way in which they are. (The conception for which David Lewis argues is an important exception to this; see Lewis 1986.) Thus we can say that there are worlds very much like this one except that Quine becomes a full-time geographer and never becomes a philosopher. There are such ways for things to be, but they are not actualized: things are not that way. This distinction between existing and being actualized holds among worlds as it does for stories or dreams, because a world is a way things might be; a maximal state of affairs or a maximally consistent set of propositions. This distinction does not carry over to individuals such as Quine and carrier pigeons. If non-actual carrier pigeons are carrier pigeons, then they are not the sort of thing that can exist without being actualized. This is the puzzle of possibilia, because we cannot see how any carrier pigeon can exist without being actual (in the way that a story can exist without being actual), and the idea that possible worlds and possible objects are on an existential par seems unacceptable. Possibilia cannot simply ride in on the coat tails of possible worlds.

One way out of this is to interpret the semantics as being about something more abstract than individuals - concepts or individual essences (see Plantinga 1974, for example). Instead of using the ordinary concept of satisfaction in the semantics, we can use a concept ‘satisfactionČast;’, saying that an individual essence *e* satisfies* ‘*x* is a geographer’ at a world *w* if and only if *e* is an essence of something that would be a geographer if that world, *w*, were actual. These individual essences can then be things that exist even when they are not actualized and that are actualized by the same individual at every world at which they are actualized. The individual essences of many carrier pigeons exist without being actualized. That is why it is true that other carrier pigeons could have existed.

On this view, individual essences rather than individuals are components of propositions. This allows the

propositions to exist even when the individuals do not exist. This conflicts with David Kaplan's view that a typical singular proposition is contingently existent, existing only in those possible worlds in which the individual that it is about exists (see Kaplan 1989; Demonstratives and indexicals §2). But it need not conflict with his more central views: that singular propositions do not contain concepts that can be used to pick out the individual the proposition is about; and that the propositional component is determined by the individual rather than the other way around. This would only mean that possible worlds in which an individual does not exist provide no resources for singling out that individual's essence. Nothing in this approach conflicts with that.

The approach in terms of essences leaves puzzles, too. What are these things? How many of them are there? These questions are not easy to answer, and so we are left with some discomfort in understanding modal claims no matter how we try to deal with the problem of possibilia.

3 Quantification, modality and essentialism

Many of those who have considered the concept of necessity have regarded it as a logical concept, applying to sentences in virtue of their form. An idea like one of the following would be typical.

A sentence of the form ‘ $\Box p$’ is true iff *p* is a logical truth.
A sentence of the form ‘ $\Box p$’ is true iff *p* is analytic.

W.V. Quine has expressed grave misgivings about this at two levels. First, he questions the intelligibility of the notion of analyticity, and at times seems even to question the concept of logical truth. Thus he will reject any notion of necessity construed in this way. Quine's misgivings about logical truth and analyticity are explored elsewhere (see Analyticity).

Quine (1961, 1966) has a further objection to the intermingling of quantifiers and modalities. Even if one grants a concept of analyticity or logical truth to serve as the basis for the concept of necessity, there are two problems for quantified modal logic: the problem of quantifying into modal contexts; and the related problem of intersubstituting coreferential names.

We can see the problem of quantifying in more easily if we consider first the problem of quantifying into truth-functional contexts. This is a real problem, because we ordinarily give an account of truth-functional connectives in terms of the truth-values of the constituents. Thus ‘ $Fa \,\&\, Ga$’ is true if and only if ‘*Fa*’ is true and ‘*Ga*’ is true. However, if we quantify, writing ‘ $(\exists x)(Fx \,\&\, Gx)$’, then ‘Čamp;’ no longer connects two sentences that have a definite truth-value. Some new interpretation for ‘Čamp;’ must be found, so that we can understand how the semantic value of the unit ‘ $Fx \,\&\, Gx$’ is determined.

Tarski (1933) found a systematic way to do this. Even though ‘*Fx*’ has no truth-value, it has a truth-value *relative to an assignment of an individual i to* ‘*x*’. ‘ $Fx \,\&\, Gx$’ is true relative to an assignment if and only if ‘*Fx*’ is true relative to that assignment and ‘*Gx*’ is true relative to that assignment. We can also say that an assignment or an individual satisfies the open sentence ‘ $Fx \,\&\, Gx$’ if and only if it satisfies ‘*Fx*’ and satisfies ‘*Gx*’. This solves the problem of quantifying into truth-functional contexts (see Tarski's definition of truth).

To interpret ‘ $\exists x \Box Fx$’ we need to know when an individual *i* satisfies ‘ $\Box Fx$’. It is not so immediately evident how to account for this, especially if you think that the necessity of ‘*Fa*’ comes from features of ‘*Fa*’ (for example, an analytic association of ‘*a*’ with ‘*Fx*’) rather than from features of the individual *i* that ‘*a*’ denotes. An individual can be singled out by many different conditions, and perhaps some of those are analytically associated with ‘*Fx*’ though others are not. For example, the condition ‘ $x = 3^3$’ has ‘ $x > 7$’ as a necessary consequence (that is, it is a necessary truth that whatever satisfies one satisfies the other); but ‘*x* numbers the planets’ does not have this necessary consequence. The condition ‘ $x > 7$’ is necessarily connected with some ways of specifying nine but not with others. We can ask whether nine is necessarily greater than seven, but Quine finds no grounds for answering this question one way rather than another, because he believes that all necessity must be grounded in analytic consequence relationships and he finds no basis for favouring one specification of nine over the other. By this standard, either answer to the question of whether nine is necessarily greater than seven seems equally justified. Beginning with an account of necessity in terms of analyticity or logical truth should lead one to Quine's suspicions about quantification into this context.

The attempt to explain satisfaction of a modal condition in terms of analytic consequence seems doomed. But how else? Saying directly that some properties are necessary and others are not repels Quine. Quine sees such distinctions as an objectionable essentialism (see Essentialism), and accordingly he dismisses quantified modal logic.

Kripke provides an alternative account of what it is for an individual to satisfy ‘ $\Box Fx$’. The account, in terms of possible worlds, has puzzled Quine; and puzzlement about existence of an individual in different possible worlds is closely related to Quine’s puzzlement about essentialism. For existence in multiple possible worlds to have significance, it seems that there must be something that makes it possible for an individual to exist with some sets of properties and not others - some are essential and some are not. We must distinguish between properties that an individual has: for example, my cat satisfies ‘*x* is an animal’ and also ‘*x* has eaten chicken livers’, but she has the first property (being a cat) essentially and the second (having eaten chicken livers) only contingently. Kripke provides at least the formal framework for an account (1963), and in later work (1972) does much to defend the substantive metaphysics that goes with it.

Terence Parsons (1969) has pointed out that we can make at least one kind of distinction between essential and non-essential properties in a way that even Quine should not object to. Properties expressed by sentences of the form ‘If *Fx*, then *Fx*’ will be essential to everything. Such essential properties seem unproblematic. We encounter problematic essentialism only if we find two individuals that differ in their essential properties. This point is limited in its scope, however. Within any fixed model verifying the assumption that all individuals have the same essential properties, any sentence with quantifying in is true in all of the same worlds as some sentence that lacks quantifying in. The resources that troubled Quine - quantification into modal contexts - seem to serve no expressive purpose (see McKay 1975 and Fine 1978b).

4 Counterparts

Although Barney is eating a sandwich for lunch, he could have had a salad instead. Most modal theorists take that claim to involve the existence of a certain non-actual but possible situation: the possibility that Barney eats a salad for lunch.

David Lewis (1968) thinks that no such possible situation exists. Barney could have eaten a salad, but that really means that there is a possible world very much like this one in which the person most like Barney eats a salad. A non-actual world will never include Barney; it may, however, include a counterpart of Barney, an individual that is more like him than is any other in that world. The fact that one of his counterparts eats a salad makes it true that Barney could eat a salad, according to Lewis.

This approach has seemed unintuitive and ill-motivated to many (Plantinga (1974) and (Kripke 1972), for example). When considering what is possible for Barney, it seems natural to suppose that we are considering possibilities that involve Barney rather than someone else. In addition, it seems that in another possible world Barney could have been very different from the way he is while someone else was more like our actual Barney. As arguments against counterpart theory, however, such intuitive considerations are inconclusive. They might be answered by denials (citing a conflict of intuitions) or, for the last objection, by clarifications of the theory that elaborate the relevant similarity notion.

Formal considerations are also significant, however. The identity relation is transitive, but there is no reason to think that the counterpart relation is transitive, and this will produce different valid formulas. In the modal systems S4 and S5 (see Modal logic), for example, the following is ordinarily valid:

$$\forall x(\Diamond\Diamond Fx \rightarrow \Diamond Fx).$$

But as long as the counterpart relation is not transitive, we could not expect that to be valid, even in S5. (An individual *i* might have a counterpart that has a counterpart that satisfies ‘*Fx*’ even though no counterpart of *i* satisfies ‘*Fx*’.) Whether this difference would be an advantage or a drawback would need further consideration.

One technical feature is directly related to some of the intuitive considerations against counterpart theory. Lewis allows that an individual in one world might have two counterparts within a single world. If that is the case, then some laws involving identity will fail. Ordinarily, the principle of necessary identity is valid:

$\Box\forall x\forall y(x = y \rightarrow \Box x = y)$.

And there seems to be a good argument for this, based on two other fundamental principles, the necessity of self-identity and Leibniz's Law (see Identity):

$$\Box\forall x\Box x = x$$
$$\forall x\forall y(x = y \rightarrow (\phi x \leftrightarrow \phi y))$$

One substitution instance of Leibniz's Law is this:

$$\forall x\forall y(x = y \rightarrow (\Box x = x \leftrightarrow \Box x = y)).$$

Given the assumption of the necessity of self-identity, the principle of necessary identity then follows. To defend counterpart theory, one must locate a fault with the necessity of self-identity or with this application of Leibniz's Law, or one must add features to the counterpart theory that will validate the principle of necessary identity (see Forbes 1985).

5 The logic of ability

Having surveyed some of the difficulties in interpreting modal logic, we now return to an illustration of the use of the possible worlds framework to shed light on inferences that are important for philosophy.

Although 'might' and 'can' are often thought of as possibility operators, they are clearly different. 'John might hit the target' and 'John can [is able to] hit the target' are very different, and Mark Brown's semantics for a logic of ability extends the possible worlds resources to make it clear what the difference is (1988).

Suppose that John is throwing darts, but does not reliably hit the target. Is John *able* to hit the target? He might hit it, but he does not have an ability that enables him to hit it reliably. In ordinary discussion, 'John is able to hit the target' is perhaps ambiguous. It could represent 'He might hit it', but is more likely to mean 'He has the ability to hit it reliably'. This suggests that there are two possibility operators, and these are distinguishable in Brown's system.

We can also give a more formal motivation for such a system. When we consider the context 'He has the ability to…reliably', we find that it does not obey the following law $\mathrm{K}\Diamond$ of the most fundamental normal modal system K:

$$\Diamond(p \vee q) \rightarrow (\Diamond p \vee \Diamond q).$$

If John has the ability to hit the target reliably, then he is able to bring it about that he hits either the right half or the left half. It need not be true that he has the ability to hit the right half reliably, and it need not be true that he has the ability to hit the left half reliably. So $\mathrm{K}\Diamond$ does not seem true, if '$\Diamond$' is to represent an ability operator.

Also, in almost all standard modal systems (all that include T), we have this principle: $p \rightarrow \Diamond p$. But suppose that John will, by luck, hit the target. It does not follow that he has the ability to do that reliably. So $\mathrm{T}\Diamond$ is not true, if '$\Diamond$' is to represent an ability operator.

Brown introduces a family of new modal operators that we can add to the standard ones. We will use these symbols: ⟐, ◈, ⊡ and ⊟ The semantics is based on a binary relation *S*, relating worlds in *W* to sets of worlds (that is, $S \subseteq (W \times \wp(W))$). We can define the new modal operators as follows (where we use '*K*' as a variable for sets of possible worlds):

⟐*p* is true at *w* iff $\exists K$ such that wSK and $\forall w'$ such that $w' \in K$, *p* is true at w'. (Informally: John is able to bring it about that *p* reliably. John can perform some action such that *p* is sure to be true if that action is performed.)
◈*p* is true at *w* iff $\exists K$ such that wSK and $\exists w'$ such that $w' \in K$, *p* is true at w'. (John might bring it about that *p*. Some action that John can perform allows for the possibility of *p*.)
⊡*p* is true at *w* iff $\forall K$ such that wSK, $\exists w'$ such that $w' \in K$, *p* is true at w'. (No matter what John does, *p* might be true.)
⊟*p* is true at *w* iff $\forall K$ such that wSK, $\forall w'$ such that $w' \in K$, *p* is true at w'. (No matter what John does, *p* will

be true.)

Given this understanding of these operators, it is natural to make at least the following assumptions about the relation *S*, validating the formulas $⟐p \rightarrow ◈p$ and $p \rightarrow ◈p$:

$\forall w \forall K$ such that wSK, $K \neq \emptyset$. (Validates D ⟐: $⟐p \rightarrow ◈p$)

(Nothing John can do annihilates all possibilities. Everything John can do has some possible outcomes.)

$\forall w \exists K$ such that wSK and $w \in K$. (Validates T ◈: $p \rightarrow ◈p$)

(The actual world is a possible outcome of at least one of John's actions.)

Focus first on ⟐, the modal operator representing ability. If we think of an action as being represented by a set of possible outcomes, then we can consider each set *K* of possible worlds such that wSK to correspond to the outcomes of some action that the agent under consideration might take. Each such set *K* would be a set of possible worlds that might be actual if the agent performed some particular action. So if the agent can hit the dartboard, but has no control over where, then a set of possible worlds that vary in where the agent hits the dartboard would be a set *K* such that wSK, and it could correspond to the action of throwing a dart. To say that $⟐A$ is to say that there is some action (that is, some set *K* of outcomes; that is, some set *K* such that wSK) that guarantees *A* (that is, *A* is true throughout *K*). To say that the agent can hit the dartboard is to say that there is some such action (that is, some set *K* such that wSK, and throughout *K* the agent hits the dartboard).

The word 'could' is associated with possibility and ability. Three of our new modal operators involve possibility, and we can give an example that is three ways ambiguous:

He could win the lottery.

This could mean that he bought a ticket, and so nothing he can do would guarantee to make him a non-winner, that is, $⊠A$. This could mean that he can buy a ticket, and then he might win, that is, $◈A$. Or this could mean that he has special powers such that if he exercised them he would win, that is, $⟐A$. (The King could win the lottery (if he wanted to), but he hates to interfere in the peasants' fun.)

Thus we have a resource for disambiguation and for the exploration of logical relationships that are central to some philosophical work in ethics and metaphysics. Even though we have seen how many puzzles arise in the acceptance of the possible worlds framework, we should not lose sight of the many reasons we have to value it.

See also: Logical and mathematical terms, glossary of; Modal operators; Montague, R.M.

THOMAS J. McKAY

References and further reading

Blackburn, S. (1993) 'Morals and Modals', in *Essays in Quasi-Realism*, Oxford: Oxford University Press. (Discusses the problem of §1: what makes modal claims true.)

Brown, M. (1988) 'On the Logic of Ability', *Journal of Philosophical Logic* 17: 1-26.(An extension of standard modal logic to accommodate statements of ability, such as 'John can [is reliably able to] hit the target', and related concepts. Discussed in §5.)

Fine, K. (1978a) 'Model Theory for Modal Logic. Part I: The *De Re*/*De Dicto* Distinction', *Journal of Philosophical Logic* 7: 125-56.(Relates technical results to some philosophical considerations.)

Fine, K. (1978b) 'Model Theory for Modal Logic. Part II: The Elimination of the *De Re*', *Journal of Philosophical Logic* 7: 277-306.(Technical results and philosophical issues related to those discussed in §3.)

Fine, K. (1981) 'Model Theory for Modal Logic. Part III: Existence and Predication', *Journal of Philosophical Logic* 10: 293-307.(Technical results and philosophical issues related to a version of actualism, a view that ascribes a special status to actual or existent objects. Related to issues discussed in §2.)

Forbes, G. (1985) *The Metaphysics of Modality*, Oxford: Oxford University Press. (A discussion of metaphysical issues associated with modal logic.)

Kaplan, D. (1989) 'Demonstratives', in *Themes from Kaplan*, ed. J. Almog, J. Perry and H.K. Wettstein, Oxford: Oxford University Press.(A classic discussion of demonstrative reference, with a semantics that builds on possible worlds models. Makes clear the distinction between necessity and analyticity mentioned in connection

with §3.)
Kripke, S.A. (1963) 'Semantical Considerations on Modal Logic', *Acta Philosophica Fennica* 16: 83-94; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971.(This presents the semantic framework for modal logic.)
Kripke, S.A. (1972) 'Naming and Necessity', in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel, 252-355.(This important article illuminates the concept of metaphysical necessity and identifies important principles of identity. There is material relevant to every section of this entry.)
Lewis, D.K. (1968) 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* 65: 113-26. (Presents and defends counterpart theory. See §4.)
Lewis, D.K. (1973) *Counterfactuals*, Cambridge, MA: Harvard University Press.(In discussion of the main topic, conditionals, there is a valuable development of the possible worlds framework. There is also some general discussion of the nature of possible worlds and of counterpart relations.)
Lewis, D.K. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.(Further development of Lewis' view of possible worlds, with some important challenges to other views of modality.)
Linsky, L. (ed.) (1971) *Reference and Modality*, Oxford: Oxford University Press.(A collection of important articles on the topics of §§2, 3.)
McKay, T. (1975) 'Essentialism in Quantified Modal Logic', *Journal of Philosophical Logic* 4: 423-38.(Results mentioned in §3, concerning Quine's linking of modal logic and essentialism.)
Montague, R. (1974) *Formal Philosophy*, New Haven, CT: Yale University Press.(In these papers, Montague developed the semantic framework of possible worlds in beginning to give a more general semantics for natural language.)
Parsons, T. (1969) 'Essentialism and Quantified Modal Logic', *Philosophical Review* 78: 35-52; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971, 73-87.(Anti-essentialist models introduced as part of a response to Quine. See §3.)
Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Oxford University Press.(An excellent introduction to topics in the metaphysics of modality. Relevant to §§1-4 of this entry.)
Quine, W.V. (1961) 'Reference and Modality', in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1980; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971.(Misgivings about the interaction of modality with reference and quantification. Discussed in §3.)
Quine, W.V. (1966) 'Three Grades of Modal Involvement', in *The Ways of Paradox*, New York: Random House. (Misgivings about the interaction of modal sentence operators with reference and quantification. Discussed in §3.)
Stalnaker, R. (1968) 'A Theory of Conditionals', in N. Rescher (ed.) *Studies in Logical Theory*, Oxford: Blackwell.(A very readable presentation of a theory of conditionals that builds on the semantic framework of possible worlds.)
Tarski, A. (1933) *Pojęcie prawdy w jezykach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956) 'On the Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(The semantics for quantification for non-modal logic. See §3.)

Modal operators

Modal logic is principally concerned with the alethic modalities of necessity and possibility, although this branch of logic is applied to a wide range of linguistic and conceptual phenomena, including natural language semantics, proof theory, theoretical computer science and the formal characterization of knowledge and belief. This wide range of application stems from the basic form of modal assertions, such as 'it is necessarily the case that ϕ*', where an entire statement* ϕ *is embedded within a context possessing rich logical structure.*

When constructing a formal representation of these embedding contexts, there are several choices concerning their specific symbolic form. The most standard approach symbolizes modal contexts as operators, which combine directly with formulas of the object language to yield new formulas. The primary alternative to this approach is to treat modal contexts as predicates, which attach not to formulas directly, but to names of formulas, and thereby attribute a metalinguistic property to a syntactic object. A variation on the operator approach, which assumes the interpretive framework of possible worlds semantics, is to treat modal contexts as quantifications over possible worlds. Finally, a variation on the predicate approach is to analyse modal contexts as predicates of propositions rather than as predicates of syntactic objects.

1 Monadic connectives

On the most standard approach, the modal operators '$\Box$' for necessity and '$\Diamond$' for possibility are introduced as new logical constants (although only one need be taken as primitive). These modal operators are syncategorematic expressions, since they are not assigned an explicit reference, rather their semantic import is carried in a recursive clause of the formal definition of truth. In terms of the syntactic formation rules, the modal operators behave as monadic connectives, structurally comparable to negation in classical logic. The modal assertion 'it is necessarily the case that ϕ' will have the symbolic form '$\Box\phi$', where the necessity connective 'operates' on the formula ϕ to yield the new formula '$\Box\phi$'. However, unlike the negation connective, the modal operators are not truth functional: it is not possible to determine the truth-value of a modal statement simply on the basis of the truth-value of the formula it embeds. In this sense, introduction of the modal operators constitutes a significant departure from the pure extensionality of classical logic, and this has led to criticism of their legitimacy, most notably from W.V. Quine.

Quine has argued that non-extensional contexts are logically permissible only if construed as 'referentially opaque', that is, the constituent symbols within these contexts are treated as occurring non-referentially. Quotation is a paradigmatic case of such a context since, for example, in the true statement '''John walks'' is a short sentence', the occurrence of the string 'John' does not refer to anyone (see Use/mention distinction and quotation). Quine (1953) argues that since modal contexts are non-extensional, they are properly viewed as quotational, wherein the embedded sentence ϕ is mentioned rather than used (this theme will be continued in §3; see Propositional attitudes §2). However, treatment of modal contexts as operators suggests that the logical structure of the embedded formula is semantically relevant, even though it occurs within a non-extensional context. Indeed, since '$\Box$' operates at the same linguistic level as the formula ϕ, it follows that the semantic value of ϕ should directly contribute to the compositional value of '$\Box\phi$'. Thus the '$\Box$' operator must utilize information not contained in the extension of the embedded expression.

2 Sorted quantifiers

According to possible worlds semantics for modal operators, introduced by Saul Kripke (1959), the non-extensional value of modal statements is determined by appeal to the extension of the embedded expression in alternative possible worlds (see Semantics, possible worlds). The sentence '$\Box\phi$' is evaluated as true at a given world just in case ϕ is true in all worlds possible relative to the given world, while the sentence '$\Diamond\phi$' is true just in case there is at least one possible world in which ϕ is true. So in this standard semantic framework, the truth definition for modal statements is specified in the metalanguage by quantifying over the set of possible worlds; the truth of an assertion of necessity is expressed in terms of a universal quantification, while an assertion of possibility is evaluated as an existential quantification. Somewhat more formally, a Kripke model M comes equipped with a set W of worlds (or points or indices) and a binary relation R on W such that a world $w' \in W$ is possible relative to a world $w \in W$ just in case wRw'. In the simplest case, where R is universal (that is, every

world in *W* is possible relative to itself and every other), reference to *R* can be dropped. Then '$\Box\phi$' is true in *w* just in case ϕ is true in all worlds $w' \in W$, and '$\Diamond\phi$' is true in *w* just in case there is at least one $w' \in W$ in which ϕ is true.

It is therefore possible to recast the logical form of modal statements as quantifiers rather than operators, where the structure of the truth definition is explicitly manifested in the structure of the object-language formula. In the general case, this will require the use of many-sorted quantificational logic, where the domain of discourse is divided into distinct subdomains, and typographically different variables are used to quantify over these respective subsets. In the modal case, the set W of possible worlds will form one such subdomain, and a distinct type of intensional variable will range over this set. Thus where the extension of an expression ψ varies according to the index of evaluation, this same 'intensional' effect can be mirrored by adding a variable to yield the expression $\psi(i)$, where i ranges over possible worlds. The interpretation function will assign to ψ a value which gives the standard type of extension when applied to worlds as arguments. To translate operator modality into the quantifier formulation, a straightforward mapping is defined where $(\psi)^* = \psi(i)$ for nonlogical constants ψ, where logical form is preserved for the extensional connectives and where $(\Box\phi)^* = \forall i(\phi)^*$ and $(\Diamond\phi)^* = \exists i(\phi)^*$. In this manner, the necessity operator is explicitly recast as a universal quantifier, and possibility is represented as an existential quantifier. The quantifier approach is perhaps most attractive as an alternative to intensional type theory (see Montague, R.M.), where the different sorts of variables correspond to distinct logical types.

3 Predicates of expressions

While the quantifier interpretation is a natural variation on the use of operators (given the framework of possible worlds semantics), the predicate approach to modality constitutes a more divergent analysis. Under this development, modal contexts are symbolized by a monadic predicate, say *N*, which is categorematic and receives a normal first-order extension (that is, a set of objects). But admissible extensions of *N* must satisfy the constraints imposed by its intended meaning as a modal device, and in this sense *N* is comparable to the binary predicate symbol 'Čequals;' normally used to designate the identity relation.

On the predicate approach to modality, syntactic expressions are themselves named by singular terms, and the necessity predicate *N* attaches to such terms to yield new atomic formulas. Quotation is often adopted as a convenient method for constructing terms that name expressions. Thus if $\ulcorner\phi\urcorner$ is the official quotation term denoting the object-language formula ϕ, then the assertion that ϕ is necessary will have the form $N(\ulcorner\phi\urcorner)$. In this manner, modal formulas are viewed as attributing a metalinguistic property to a linguistic expression treated as an object of discourse. The two most natural properties to attribute in this context are that the object-language formula is valid with respect to some class of models, or that it is provable in some formal system. On this latter reading, the predicate approach is quite suitable for applications to proof theory (see Gödel's theorems §6) and for applications to the symbol-processing accounts of epistemic contexts common in artificial intelligence.

On the predicate analysis, the named formula is mentioned rather than used, and this is amenable to the Carnap-Quine position that the logic of modalities should be expressed in an extensional metalanguage (see Carnap 1937), and that the attempt to express modality in an intensional object language rests on a confusion between use and mention. It is also compatible with Quine's view (see §1) that modal contexts are referentially opaque. The fact that the modal predicate attaches to a term denoting a formula indicates that there are no free variables inside a modal context, and this would prohibit the articulation of quantified modal logic. Therefore, Quine has argued that the only legitimate development of 'intensional' logic is comprised by that fragment which can be rendered in terms of the extensional clarity of a metalinguistic predicate.

However, Montague (1963) has subsequently demonstrated that for any language endowed with the 'self-referential' capacities of elementary number theory, the addition of even very weak modal assumptions leads to inconsistency under the predicate approach. In systems that extend the theory *Q* of Robinson arithmetic, Gödel's diagonal lemma establishes that for any open formula $\beta(x)$ with one free variable, there exists a closed formula θ such that $\vdash \theta \leftrightarrow \beta([\theta])$, where $[\theta]$ is the Gödel numeral code for θ. If necessity is formalized as a predicate $N(x)$ within such a system, then the diagonal lemma establishes that there will be a sentence μ such that $\vdash \mu \leftrightarrow \neg N([\mu])$, where μ 'says of itself' that it is not necessary. μ is the modal analogue of the celebrated Gödel sentence that asserts its own unprovability, and of the liar sentence that underpins Tarski's theorem on the undefinability of truth (see Semantic paradoxes and theories of truth). And in a manner analogous to the

Gödel-Tarski results, μ can be used to deduce a contradiction, given fairly minimal principles governing the logic of *N*.

In light of these negative results, the predicate view of modality must be approached with some care. A crucial point is that not all the formulas in which *N* occurs will correspond to statements that can be articulated using the resources of the operator. In particular, there is no operator counterpart to the 'self-referential' sentence μ, since the '$\Box$' connective must apply to formulas rather than to their names. And if radically new formulas such as μ are allowed to instantiate the relevant modal schemata, then new axioms are added to the modal system, and this in turn yields an unwanted increase in deductive power. Thus it is possible to avoid inconsistency by inverting the Quinean perspective and defining the predicate system so that it precisely mirrors the expressive-deductive power of the operator.

4 Predicates of propositions

In addition to treating them as predicates of syntactic objects, it is also possible to formalize modal contexts as predicates of semantic objects. In systems of higher-order intensional logic such as Montague's system IL, semantic value is divided along Fregean lines into both an extensional and an intensional component (see Sense and reference). Montague's scheme utilizes Carnap's insight that the intensional component of semantic value can be modelled as a function from possible worlds to extensions. The intension of a sentence is a proposition, while its extension is a truth-value, and hence a proposition is modelled as a (characteristic function of a) set of possible worlds, namely that set of worlds in which the sentence is true. The object language is equipped with an intensional operator 'ˆ' which, when attached to an expression, yields a term that denotes the intension of the expression. So the term $\hat{\phi}$ denotes the proposition expressed by ϕ, which is the set of possible worlds in which ϕ is true. Necessity can then be rendered as a higher-order predicate *Nec*, which applies to sets of worlds and is meant to assert that the set in question is identical to the set of all worlds in the model. In this framework, the statement that ϕ is necessary will then have the form $Nec(\hat{\phi})$, which asserts that the set of worlds constituting the intension of ϕ is identical to the set of all worlds in the model, and this is equivalent to the standard operator interpretation that ϕ is true in all possible worlds.

See also: Logical and mathematical terms, glossary of; Modal logic; Modal logic, philosophical issues in; Possible worlds §2

PAUL SCHWEIZER

References and further reading

Carnap, R. (1937) *The Logical Syntax of Language*, London: Routledge & Kegan Paul. (Expounds the view that modality should be expressed in an extensional metalanguage.)

Dowty, P., Wall, R. and Peters, S. (1981) *Introduction to Montague Semantics*, Dordrecht: Reidel. (Provides an introduction to Montague's system of intensional type theory.)

Gallin, D. (1975) *Intensional and Higher-Order Modal Logic*, Amsterdam: North Holland. (Gives a treatment of modal operators as quantifiers within the framework of type theory.)

Kripke, S.A. (1959) 'A Completeness Theorem in Modal Logic', *Journal of Symbolic Logic* 24: 1-15. (Introduces possible worlds semantics for modal logic.)

Montague, R. (1963) 'Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability', in R. Thomason (ed.) *Formal Philosophy*, New Haven, CT: Yale University Press, 1974. (Presents inconsistency proofs for predicate version of modal logic.)

Quine, W.V. (1953) 'Three Grades of Modal Involvement', in *The Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press, 1966. (Advocates the view that modal contexts are quotational.)

Schweizer, P. (1992) 'A Syntactical Approach to Modality', *Journal of Philosophical Logic* 21: 1-31. (Addresses Montague's argument that the predicate approach to modality is inconsistent.)

Skyrms, B. (1978) 'An Immaculate Conception of Modality', *Journal of Philosophy* 75: 77-96. (Addresses Montague's argument that the predicate approach to modality is inconsistent.)

Model theory

Model theory studies the relations between sentences of a formal language and the interpretations (or 'structures') which make these sentences true or false. It offers precise definitions of truth, logical truth and consequence, meanings and modalities. These definitions and their consequences have revolutionized the teaching of elementary logic.

Model theory also forms a branch of mathematics concerned with the ways in which mathematical structures can be classified. This technical work has led to philosophically interesting results in at least two areas: it has thrown light on the nature of the set-theoretic universe, and in nonstandard analysis it has suggested new forms of argument (where we prove something different from what we intended, but then use a general model-theoretic argument to change the result into what we wanted).

The word 'model' has many other uses. For example, model theory is not about scientific theories as models of the world. It is also a controversial question - not considered here - how model theory is connected with the 'mental models' which appear in the psychology of reasoning.

1 Basic ideas

In 1954 Alfred Tarski announced that a new branch of metamathematics had come into being. He called it 'model theory'; it studied 'mutual relations between sentences of formalized theories and mathematical [structures] in which these sentences hold'. Model theory quickly became an active area of research on the boundary between logic and other branches of mathematics (see §5 below). The name 'model theory' is sometimes applied more broadly, particularly by philosophers, to include any inquiry which uses formal languages, structures and a truth definition. For example, a model-theoretic approach to English semantics is one which treats English as a formal language and attaches meanings to English phrases in terms of possible worlds, regarding these worlds as mathematical structures.

When Tarski spoke of a 'sentence of a formalized theory', he meant what logicians have also called a 'sentence schema', that is, a string of meaningful words and meaningless symbols (sometimes called schematic variables), which becomes a meaningful sentence when the symbols are given meanings. For example, the string

(1) Everything which is a P is a Q

is a sentence schema. If we interpret 'P' as meaning 'cow' and 'Q' as meaning 'quadruped', then the schema comes to mean the same as the sentence 'Every cow is a quadruped'. In model theory we always assume that the word 'everything' needs interpretation too; the interpretation will say what set of things counts as everything, and any set may be allowed. For example, if we take 'everything' to range over the set of objects in China, then the original string comes to mean 'Every object in China which is a cow is a quadruped'.

A 'mathematical structure' is a way of interpreting the symbols in a schema. It consists of a set of objects which is the range of quantifier words such as 'everything' and 'there exists'; and a family of sets, relations, functions and so on which express the meanings of the schematic variables. The set of objects is called the 'domain' or 'universe' of the structure, and its members are called the 'elements' of the structure. For example, the structure which gives our interpretation of (1) has the set of objects in China for its domain; it attaches the set of cows in China to the symbol 'P' and the set of quadrupeds in China to the symbol 'Q'. (More formally, we can regard a structure M as an ordered pair $\langle \mathrm{dom}(M), Y \rangle$ where dom(*M*) is the domain of *M* and *Y* is a function which takes each interpreted symbol *s* to its interpretation $Y(s)$, also written s^M. Many mathematical objects can also be thought of as structures in this sense; thus a group *G* is a pair $\langle X, Y \rangle$ where *X* is the set of elements of *G* and *Y* is a function taking the group product symbol '$\cdot$' to the group multiplication '$\cdot^G$'.)

In our schema (1), the meaningful expressions are those written in English. Model theorists often use logical symbols for the meaningful words. For example, 'and' is often written 'Čamp;' or '$\wedge$'; for 'it is not true that' one writes '$\sim$' or '$\neg$'; one writes '(x)' or '$\forall x$' to mean 'for every element *x*'. When this is done, a sentence schema is the same thing as a sentence of formal logic with uninterpreted non-logical symbols (see Predicate calculus).

2 Truth and satisfaction

The central device of model theory is the truth relation, written ‘ $\models$’. (Pronunciations vary, but ‘double turnstile’ is safe, using ‘turnstile’ for ‘ $\vdash$’.) Suppose ϕ is a sentence schema and *M* is a structure which interprets ϕ. Interpreted, ϕ says something which is either true or false. If it is true, we say that ϕ is true in *M*, or that *M* is a model of ϕ; in symbols, $M \models \phi$. Tarski showed how one can give a precise mathematical definition of the relation ‘ $\models$’. His definition is correct in the sense that it agrees with our intuitive notions about when an interpretation *M* makes a formula ϕ into a true sentence. (See Tarski’s definition of truth for the context of Tarski’s definition. See also Semantics, game-theoretic for an interesting variant.)

In defining ‘ $\models$’, Tarski assumed that each structure *M* is a set-theoretic object and that the schemas ϕ are sentences of a conventional formal logic. Also *M* must interpret every non-logical symbol in ϕ. It is usual to ensure this by stating Tarski’s definition relative to a set *S* of non-logical symbols; we assume that *M* interprets just these symbols, and that ϕ uses no non-logical symbols besides those in *S*. This set *S* is then called the ‘signature’ (or similarity type) of *M* and ϕ.

In fact Tarski defined a slightly more complicated relation, namely satisfaction. Suppose *M* is a structure, $a_1, \ldots, a_n$ are elements of *M* and $\phi(x_1, \ldots, x_n)$ is a formula with free variables $x_1, \ldots, x_n$. We write $M \models \phi[a_1, \ldots, a_n]$ (pronounced ‘ a_1 to a_n satisfy ϕ in *M*’) to mean that when *M* is used to interpret the symbols of ϕ apart from the free variables, and x_i is read as the name of the element a_i for $1 \leq i \leq n$, then ϕ becomes a true sentence. The truth relation is the special case where $n = 0$. (The set of all *n*-tuples $(a_1, \ldots, a_n)$ such that $M \models \phi[a_1, \ldots, a_n]$ holds is called the relation (or set) defined in *M* by ϕ.)

Tarski’s definition of satisfaction has several clauses, depending on the form of the formula ϕ. For example, if ϕ is of the form $\psi \wedge \chi$ then $M \models \phi[a_1, \ldots, a_n]$ holds if and only if $M \models \psi[a_1, \ldots a_n]$ and $M \models \chi[a_1, \ldots, a_n]$ both hold. If ϕ is of the form $\forall x_2 \psi(x_1, x_2)$, then $M \models \phi[a_1]$ holds if and only if for every element *b* of *M*, $M \models \psi[a_1, b]$ holds. These and similar clauses allow us to break ϕ down into its smallest component formulas, known as atomic formulas. For atomic formulas, Tarski’s definition refers directly to the interpretations of symbols given by the structure *M*.

By applying Tarski’s definition, we can reduce the statement that $M \models \phi[a_1, \ldots, a_n]$ to a purely set-theoretic statement. To illustrate this, let $\phi(x)$ be the first-order formula ‘ $\forall y(y = c \wedge Py \rightarrow y = x)$’, and let *M* be a structure with a set P^M to interpret the symbol ‘*P*’ and an element c^M to interpret the constant *c*. Suppose *b* is an element of *M*. Then according to Tarski’s definition of satisfaction (with a very small simplification),

(2) $M \models \phi[b]$

holds if and only if

(3) $\{c^M\} \cap P^M \subseteq \{b\}$.

(Intuitively, every member of the set P^M which is equal to c^M is equal to *b*. The schema ϕ defines in *M* the set $\{c^M\}$ if c^M is in P^M, and the set of all elements of *M* otherwise.)

3 Classes defined by axioms

Model theory grew out of a development in mathematics. Around the turn of the twentieth century it was becoming common to define classes of structures in terms of the axioms which are true in them; such classes are said to be axiomatically defined. The classes of groups, rings and Boolean algebras are three examples (see Boolean algebra). Certain questions arise naturally in this context. Let us note three. First, what can we tell about an axiomatically defined class by looking at the syntactic form of its axioms? Second, given an axiomatically defined class, when is one of the symbols redundant in the sense that its meaning in any structure in the class is determined by the meanings of the other symbols? Third, given a set of axioms *T*, one of which is ϕ, when is ϕ redundant in the sense that the class axiomatically defined by *T* would not be altered if we left out ϕ? These are all typical questions of model theory.

By a theory we mean a set of sentences of a formal language. We say that a structure *M* is a model of the theory *T* if *M* is a model of every sentence in *T*. So the class of all models of *T* is the axiomatically defined class whose set

of axioms is *T*. First-order model theory, which is by far the largest part of the subject, studies the case where *T* is a set of first-order sentences. Many interesting mathematical classes of structures - including groups, rings and Boolean algebras - are defined by first-order axioms.

Here are two sample theorems of first-order model theory, which illustrate the first and second questions above. We shall turn to the third question in the next section.

Two structures are said to be ‘elementarily equivalent’ if they are models of exactly the same first-order sentences. Our first theorem states that if *M* is a structure with infinitely many elements and λ is an infinite cardinal, then we can cut down *M* (if it has more than λ elements) or add new elements (if it has fewer) so as to form a structure *N* which is elementarily equivalent to *M* and has exactly λ elements. This theorem is known as the downward Löwenheim-Skolem theorem (when elements are removed) or the upward Löwenheim-Skolem theorem (when elements are added). The upward and downward theorems together are equivalent to the axiom of choice, which is an indication that in general it is impossible to give an explicit description of *N*. The upward theorem follows from the compactness theorem of §4.

If *T* is a first-order theory, and a symbol *R* of the signature of *T* is redundant in the sense described at the start of this section, then we say that *R* is implicitly definable relative to *T*. Our second theorem states that if *R* is implicitly definable relative to *T*, then it is explicitly definable in the sense that there is some first-order sentence which is true in all models of *T* and which says, ‘*R* means such-and-such’. (See Beth’s theorem and Craig’s theorem for further details.) In practice this theorem is generally used backwards: we are given that *R* is not explicitly definable relative to *T*, and we use the theorem to find two models of *T* which agree in the interpretations of all symbols except *R*.

4 Semantic and syntactic consequence in first-order logic

The question of redundant axioms goes back several hundred years. Already in the thirteenth century Persian geometers were trying to show Euclid’s parallel postulate to be redundant in his axioms for geometry. In fact it is not, and proofs of this fact by Beltrami and others in the late nineteenth century gave a boost to the development of model theory. Beltrami introduced the word ‘interpretation’ in 1868.

Some notation will help. If *T* is a theory and ϕ a sentence, we write

(4) $T \models \phi$

to mean that every model of *T* is also a model of ϕ. (4) is sometimes read ‘ ϕ is a semantic consequence of *T*’. (This double use of ‘ $\models$’ is unfortunate, but the usage has become firmly entrenched. In (4) there is a theory to the left of ‘ $\models$’, not a structure as in (2).) So ϕ is redundant in the theory consisting of *T* and ϕ, if and only if $T \models \phi$.

The central theorem here is Gödel’s completeness theorem, proved in his doctoral dissertation of 1929 and published in 1930. It states that if *T* and ϕ are first-order, then (4) holds if and only if ϕ is deducible from *T* in a standard proof calculus for first-order logic. We write $T \vdash \phi$ when this condition holds; so Gödel’s theorem says that the structural relation ‘ $\models$’ (in the sense of (4)) is equivalent to the purely syntactic relation ‘ $\vdash$’. Like all later logicians, Gödel proved his theorem by assuming that ϕ is not deducible from *T* and constructing a model of *T* in which ϕ is false. Later logicians have usually built the model out of the expressions of the first-order language; Gödel, perhaps influenced by Hilbert’s constructions of counterexamples in geometry, used numbers as the elements of the model and showed that appropriate relations could be found to interpret the symbols of the language. (See Gödel’s theorems. Gödel proved his theorem for one particular proof calculus and a countable signature, but both these restrictions are unnecessary.)

Of the many applications of Gödel’s completeness theorem, two are worth noting here. First, when *T* is the empty theory in (4), we leave it out and write simply ‘ $\models \phi$’; and likewise with ‘ $\vdash$’. Compare the forms of the following two statements.

(5) $\vdash \phi$, that is, there is a proof of ϕ
(6) $\models \phi$, that is, there is not a model of $\neg\phi$.

By Gödel’s theorem, (5) and (6) are equivalent, both saying that ϕ is a theorem (or valid sentence) of first-order

logic. The difference between the forms of (5) and (6) is very convenient. If ϕ is a theorem, according to (5) we can show it by giving a proof of ϕ. If ϕ is not a theorem, then by (6) we can show it by giving a structure in which ϕ is not true. Either way, we know what we have to produce in order to establish the fact. (A strengthening of the completeness theorem, proved independently by Hasenjäger and Kleene in 1952, allows us to do better. If ϕ is a first-order sentence but not a theorem, then there is a model of $\neg\phi$ which consists of the natural numbers together with some arithmetically definable sets, relations and functions. So in the second case we can specify the model by giving some arithmetical formulas.)

Now that we know Gödel's theorem, we can try to prove a first-order sentence ϕ by showing that its negation $\neg\phi$ has no model. This idea leads quickly to tableau proofs (see Natural deduction, tableau and sequent systems §4). Of course the tableau method itself is a purely syntactic calculus. But it can be motivated very directly in terms of the model-theoretic relation (4); with a good class of students one can explain the principle and leave it as a homework exercise to work out the rules.

Second, suppose *T* is a first-order theory and every finite subset of *T* has a model. Then there is no finite subset T' of *T* such that every model of T' is a model of the contradictory statement '$\neg\forall x\ x = x$' (since no structure is a model of this statement). So by Gödel's theorem, *T* has no finite subset from which we can formally prove '$\neg\forall x\ x = x$'. But then we cannot prove '$\neg\forall x\ x = x$' from *T* at all, since any formal proof would use only a finite part of *T*. Hence by Gödel's theorem again, it is not true that every model of *T* is a model of the contradiction; in other words, *T* has a model. By this roundabout route we have deduced the 'compactness theorem': if every finite subset of a first-order theory has a model, then the whole theory has a model. This is the single most important theorem of model theory. All the applications in the next section use it.

5 First-order model theory: four applications

This section relates some of the mathematical successes of first-order model theory. It can be skipped without breaking continuity.

First, suppose *T* is a first-order theory, and for every natural number *n* there is a model M_n of *T* in which there are at least *n* elements which satisfy the formula *Px*. For each *n* we can write a first-order sentence which says, 'There are at least *n* elements *x* such that *Px*'; let *U* be the set of these sentences for all *n*. Our assumption about *T* implies that if we take just finitely many sentences from *U*, then there is some model of *T* in which all these sentences are true. Then the compactness theorem tells us that there is a model of *T* in which all the sentences are true, so that infinitely many elements satisfy *Px*. This argument has many applications. It shows, for example, that if *T* has arbitrarily large finite models then *T* has an infinite model. It also shows that there is a structure *N* which is elementarily equivalent to the field *R* of real numbers (and in fact contains it), but which has infinitesimal elements; these are elements greater than 0 but less than 1, $\frac{1}{2}$, $\frac{1}{3}$ and so on. Since *N* contains infinitesimals, we can do calculus in *N* in the way that Leibniz intended, calculating $\mathrm{d}y/\mathrm{d}x$ as a ratio of two infinitesimal numbers d*y* and d*x* (this is a slight oversimplification). Suppose that we can use infinitesimals to prove that a certain first-order sentence ϕ is true in *N*. Then ϕ must be true in *R* too, since *R* and *N* are elementarily equivalent. This ingenious way of doing calculus has sprouted into a new branch of mathematics (see Analysis, nonstandard).

In 1982 Shelah gave a very different application of model theory. Confining himself to structures of countable signature, he looked at elementary equivalence classes, that is, classes that consist of a structure and every other structure elementarily equivalent to it. Two structures are said to be isomorphic if one is an exact copy of the other. (Formally, *M* is isomorphic to *N* if there is a bijection from the domain of *M* to that of *N*, which takes the interpretation s^M of each symbol *s* of the signature to the interpretation s^N.) Shelah proved that a dichotomy holds: each elementary equivalence class is either very good or very bad. 'Very good' means that we can classify all the structures in the class, up to isomorphism, by means of a few numerical invariants (such as dimension in the case of vector spaces). 'Very bad' means that we can construct two structures in the class which are not isomorphic but are extremely hard to distinguish from each other. Shelah sees this work as a step towards bringing the more chaotic areas of mathematics under control.

In 1992 Hrushovski and Zil'ber proved a theorem which says that algebraically closed fields can be recognized through certain model-theoretic properties of their Zariski topologies. Hrushovski then used this result to prove an open case of the geometric Mordell-Lang conjecture (an important conjecture in diophantine geometry). Zil'ber

had long been convinced that there must be some kind of characterization of algebraically closed fields in terms of their model-theoretic properties, partly because these fields are such fundamental objects in mathematics. It was also partly because Macintyre had already shown that if *M* is an infinite field, then *M* is algebraically closed if and only if the set *T* of all first-order sentences true in *M* is λ-categorical for every uncountable cardinal λ - that is, that any two models of *T* with exactly λ elements must be isomorphic. (A theory is said to be categorical if all its models are isomorphic. By the upward Löwenheim-Skolem theorem, if a first-order theory has a model with infinitely many elements, then it cannot be categorical, though it may be λ-categorical for some cardinal λ.)

Finally, one achievement of the twentieth century has been to base all mathematics on axiomatic set theory. A theorem is reckoned to be proved if we know in principle how to write it as a formal deduction from the axioms of ZF (Zermelo-Fraenkel set theory). Unfortunately ZF seems to leave unanswered many natural questions about sets and infinite cardinals. Since the 1960s, set theorists have studied these questions by examining models of the ZF axioms. For example, in 1963 Paul Cohen showed that the axiom of choice is not deducible from the other axioms of ZF, by constructing models of these other axioms in which the axiom of choice is false. Model theorists have devised several techniques for constructing interesting models of a given theory, and two of these techniques have led to other advances in set theory. Ultraproducts, which are good for constructing 'fat' models in which there are as many types of element as possible, are an essential tool for studying the colossal (but possibly non-existent) cardinals that set theorists call large cardinals. The technique of indiscernibles, which constructs 'thin' models with as few types of element as possible, is used to show the enormous differences between Gödel's constructible universe and a universe of sets which contains very large cardinals. (See Set theory.)

6 Limitations of first-order model theory in mathematics

First-order logic has a delicate balance of expressiveness: it can say things about a fixed number of elements at a time, but it is very bad at expressing statements which involve infinitely many, or arbitrarily large finite numbers of, elements. For example, it is easy but boring to write down a first-order sentence which says, 'The number of elements *x* such that *Px* is exactly 10,556,001'. By contrast there is no first-order sentence that says, 'There are infinitesimals'. (The definition of infinitesimals involves the infinitely many elements 1, $\frac{1}{2}$, $\frac{1}{3}$ and so on.) Also we saw that no first-order sentence expresses 'There are just finitely many *x* such that *Px*'. Even if we restrict attention to finite structures, there is no first-order sentence which is true in just those structures for which the set of elements satisfying the formula '*Px*' has an even number of elements.

We have other formal languages in which these things can be expressed. For example, there are languages, known as 'infinitary' languages, in which we can form infinite conjunctions ' S_1 and S_2 and S_3 and… '. The statement that an even number of things satisfy '*Px*' can be written as an infinite conjunction 'There is not exactly one thing *x* such that *Px*, and there are not exactly three things *x* such that *Px*, and there are not exactly five things… '. Infinitary languages can say more than first-order languages; but by the same token it is harder to construct models of theories in infinitary languages, and this makes them less useful for applying model-theoretic arguments in mathematics (see Infinitary logics). The study of different kinds of formal languages and their expressive power is called 'generalized model theory' (see the encyclopedic volume of Barwise and Feferman 1985).

In 1969 Lindström proved some theorems which compare first-order languages with other languages that have greater expressive power. For each signature *S* there is a first-order language $L(S)$; it consists of just those formulas which can be built up from the symbols in *S* using the first-order logical expressions. Suppose that we add some new expressions, for example, quantifiers which express that an even number of things have this or that property. Then each signature *S* will give rise to a richer language $L'(S)$. One of Lindström's results is that if the upward and downward Löwenheim-Skolem theorems (quoted in §3 above) hold for 'elementary equivalence' in the sense of the languages $L'(S)$ instead of $L(S)$, then for every sentence of $L'(S)$ there is already a sentence of $L(S)$ which has exactly the same models. Roughly speaking, any logic which is at least as strong as first-order logic, but not strong enough to distinguish one infinite cardinal from another, must be first-order logic. Lindström's theorems show that the choice of first-order languages as the main languages of model theory was not just a happy accident. But they also raise the question of whether there might be purposes for which first-order languages are not the best choice. We may find such purposes outside mathematics.

7 Truth and possible worlds

The interest of model theory for philosophers rests mainly on two items: Tarski's truth definition and Tarski's characterization of logical deduction. Let us begin with the truth definition. (One philosophical text published in 1992 even defines model theory as the programme of giving a truth definition for quantified predicate logic!)

In model-theoretic versions of Tarski's truth definition we define a relation '$\models$' between structures and sentences; '$M \models \phi$' says that the sentence ϕ is true in the structure *M*. But a typical sentence of English, for example,

(7) Pollution is not the sole cause of the rise in asthma

can be assessed as true or false; no structure is involved.

Some writers contrast absolute truth with model-theoretic truth; the latter is relative to a structure, the former is not. This is a misleading contrast. Truth means the same in model theory as it means anywhere else - the difference lies in the sentences whose truth is being assessed. A typical sentence of a formal language of logic is a schema with schematic variables in it; by itself it does not make a statement. A structure is needed to give meanings to the variables. By contrast, the sentence (7) has no variables, so it needs no structure.

There are some analogues of schematic variables in natural languages. The closest analogues are indexical expressions such as 'I', 'he', these' and 'now' (see Demonstratives and indexicals). To assess the truth of the sentence

(8) He took those

we need to know who he is and what those are. An answer of the form 'If he is Christopher Naples and those are my football boots, then (8) is true' has almost exactly the same content as the statement that the schema '*X* took *Y*' is true in a certain structure.

Nor does (8) determine the time and place referred to. Maybe the sentence was false yesterday but will be true tomorrow. When we assess its truth, we normally take the default time and place - namely here and now; but any account of the truth or otherwise of (8) needs to make some assumption at least about the time. (Strictly there may also be a question about domains of quantification: in (7) what range of possible causes are we considering?)

Some philosophers go further: we need to know whether (8) is taken to be about this world or some other possible world. Even further: there are sentences such as

(9) It is possible that it will rain tomorrow

whose truth now in this world may depend on what is true in other possible worlds. (Is there a possible continuation of this world in which it will rain tomorrow?)

None of these relativities depends on model theory. But model theory offers a format for handling them. Sentence (8) is true in the structure where the possible world is this one, the time and place are as follows, he is so-and-so and those are such-and-such. Oversimplifying a little, sentence (9) is true in possible world *i* if and only if there is a world *j* which is possible relative to *i*, such that 'It will rain tomorrow' is true in *j*. Montague (a student of Tarski) offered a truth definition for substantial fragments of English, using devices of exactly these kinds (see Montague, R.M.). Similar ideas appear in the semantics of many languages used in philosophy and artificial intelligence; see, for example, Modal logic; Tense and temporal logic; Intensional logics. (See also Semantics, possible worlds.)

There may be ontological problems about what a possible world is, but model theory neither adds to nor subtracts from these problems. Most of the structures that one meets in model theory are set-theoretic objects which have nothing to do with possible worlds.

8 Logical truth and consequence

As we saw in the previous section, model theory provides a setting for studying how the truth of a statement is affected if we alter some features of the statement. One traditional way of defining and classifying necessary truths is by noting that they stay true under various kinds of alteration. For example, a necessary truth stays true if we change the possible world; an eternal truth stays true if we change the time; a logical truth stays true if we

systematically reinterpret all the words in it except the logical ones. Classifications along these lines are sometimes described as model-theoretic.

In 1936 Tarski offered just such a definition of logical consequence. On his account, a sentence *S* is a logical consequence of sentences *T* if one cannot reinterpret the non-logical words in *S* and *T* so as to make *S* false and all the sentences of *T* true. He was deliberately vague about which words are logical; he suggested that different choices might lead to different notions of consequence (such as analytic consequence). (See Consequence, conceptions of §5.)

Tarski's definition makes no mention of structures, since he applied it directly to meaningful sentences. To adapt it to a model-theoretic style, we first have to consider formal languages *L* whose symbols are all either logical expressions or schematic variables. Then, as in §4, if *U* is a theory and ϕ a sentence in such a language *L*, we write '$U \models \phi$' to mean that every model of *U* is a model of ϕ. Finally, we say that *S* is a logical consequence of *T* if and only if there are a theory *U* and a sentence ϕ in some such language *L* such that $U \models \phi$ and T, S is a substitution instance of U, ϕ. (We say that a sentence *S* is a 'substitution instance' of a schema ϕ if *S* comes from ϕ by replacing each schematic variable by a meaningful expression - the same expression at all occurrences of the variable; and likewise for a theory.)

9 Model theory in computer science

Computers are good at handling formulas; they are not so good at set theory. So when model theory appears in computer science, it usually has a supervisory role, for example, to describe what computers or robots are supposed to be doing, or how a computer language is intended to work, or what kind of information a database is capable of retrieving.

There is virtually no place in computer science where the class of all models of a theory has any importance at all; in fact it is hard to see how large uncountable structures could possibly be relevant to computing. But sometimes (a) one is interested in the finite models of some theory. For example, these can represent the possible states of a database, or the possible states of a machine during a computation. Sometimes (b) the relevant models are those in which every element is named by a term; these appear as datatypes, or as representations of the behaviour of a machine through time. Another important case is (c) where one is only interested in a single structure, and the question is which sets and relations can be defined in this structure. The model-theoretic semantics of PROLOG is a case in point: the structure in question is the term algebra, and the relations are defined by PROLOG programs. Computer science is open-ended and fast developing, so that other uses of model theory are very likely to appear.

In the 1980s some theoretical computer scientists started to voice a damaging criticism of first-order model theory. The model-theoretic questions which arise naturally in computer science are to do with defining classes of finite structures, or single structures, or relations within a given structure. The classes are defined in terms of the behaviour of a machine or program. For model theory one needs logical languages in which these same classes can be defined. In practice, first-order languages usually turn out to be grossly unsuitable.

For example, if *M* is a finite structure representing the information in a database, and the relations of *M* include a set *P* and a binary relation *R*, then it is easy to compute whether the number of elements in *P* is even, or whether it is possible to get from one given element *a* to another element *b* by a chain: $aRc_1R\dots Rc_nRb$. But neither of these questions can be phrased as a first-order sentence. Or, again, the relations defined by a PROLOG program are rarely definable by first-order formulas; this is because the program can perform recursion, so that the relations are built up inductively.

On the other hand, the properties of first-order logic that Lindström pointed to (see §6), such as the Löwenheim-Skolem theorems, are largely irrelevant for computer science.

No definitive replacement for first-order logic has appeared, but computer scientists have begun to develop some very interesting new languages, for example, languages in which one can express the transitive closure of relations. There is also a thriving branch of model theory devoted to finite structures. Here the interesting classes of structure are, for example, the class of structures which give the answer 'yes' to a certain database query, or the class of those structures which can be picked out by an algorithm in a certain complexity class. (See Complexity, computational; Quantifiers, generalized.)

See also: Logical and mathematical terms, glossary of

WILFRID HODGES

References and further reading

Barwise, J. and Feferman, S. (eds) (1985) *Model-Theoretic Logics*, New York: Springer.(A mathematical survey of various logics, unfortunately before there was much model-theoretic work on logics for computer science.)

Doets, K. (1996) *Basic Model Theory*, Stanford, CA: CSLI Publications.(A good introduction for readers who know some logic but are not mathematicians.)

Ebbinghaus, H.-D., Flum, J. and Thomas, W. (1984) *Mathematical Logic*, Berlin and New York: Springer, 2nd edn, 1994.(A readable general introduction to mathematical logic which emphasizes model theory; includes proofs of Lindström's theorems.)

Hintikka, J. (1988) 'On the Development of the Model-Theoretical Tradition in Logical Theory', *Synthese* 77: 1-16.(A philosophical discussion of model theory by a model theorist.)

Hodges, W. (1986) 'Truth in a Structure', *Proceedings of the Aristotelian Society* 86: 135-51.(Another philosophical discussion of model theory by a model theorist.)

Hodges, W. (1993) *Model Theory*, Cambridge: Cambridge University Press.(A general mathematical text of model theory.)

Tarski, A. (1954) 'Contributions to the Theory of Models I', *Indagationes Mathematicae* 16: 572-81. (The paper which first named model theory.)

Tarski, A. (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. J.H. Woodger, Oxford: Clarendon Press; ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(Includes important papers on the conceptual background of model theory, in particular, 'The Concept of Truth in Formalized Languages' (1933), 'On the Concept of Logical Consequence' (1936) and 'On Definable Sets of Real Numbers' (1931).)

Models

Of the many kinds of things that serve as 'models', all function fundamentally as representations of what we wish to understand or to be or to do. Model aeroplanes and other scale models share selected structural properties with their originals, while differing in other properties, such as construction materials and size. Analogue models, which resemble their originals in some aspect of structure or internal relations, are important in the sciences, because they can facilitate inferences about complicated or obscure natural systems. A collection of billiard balls in random motion is an analogue model of an ideal gas; the interactions and motions of the billiard balls are taken to represent - to be analogous to - the interactions and motions of molecules in the gas.

In mathematical logic, a model is a structure - an arrangement of objects - which represents a theory expressed as a set of sentences. The various terms of the sentences of the theory are mapped onto objects and their relations in the structure; a model is a structure that makes all of the sentences in the theory true. This specialized notion of model has been adopted by philosophers of science; on a 'structuralist' or 'semantic' conception, scientific theories are understood as structures which are used to represent real systems in nature. Philosophical debates have arisen regarding the precise extent of the resemblances between scientific models and the natural systems they represent.

1 Types of model

Generally, we use models as representations of various things, including desires, hopes and actions, as well as objects and systems in the real world. Not only do many distinct types of thing serve as models, but they also relate in a variety of ways to the world and to us. Fundamentally, to be a model is to serve in a triadic relation: *a person* takes *something* as a model of *something else* (Wartofsky 1979).

One very general meaning of 'model' is as an ideal or paradigm case or exemplar, for example a 'model student', curious and energetic. Here, models serve as ideals towards which people are encouraged to adapt themselves (see Examples in ethics).

Scale models, for example, model aeroplanes and other miniaturized representations, are objects whose properties and relative proportions resemble their originals. Only some properties, however, are the same in the model and the original; usually the construction materials differ, and the functions of the model and the original differ. Still, scale models exhibit the basic relations which enable one thing to stand in for another; in Charles S. Peirce's terminology, the model is an *icon* which represents the original. Icons are different from symbols in their relationship to the things represented: symbols are tied to their originals purely by convention, and any thing can serve as a symbol for any other thing; icons, in contrast, bear some resemblance or similarity to the things that they represent to us. That is, there must be something about the icon itself which is similar or analogous to the object or system of which it is a model. Thus, scale models are a simple kind of iconic model.

Analogue models are usually more abstract than scale models, and resemble their originals primarily in some aspect of structure or internal relations, rather than in materials or appearance. A collection of billiard balls in random motion is an analogue model of an ideal gas in virtue of certain internal relations and interactions among the billiard balls, and not because any thing in the substance or appearance of gas molecules is *like* billiard balls.

Mathematical models are even more abstract than analogue models, in that the objects and their interrelations in the model are all mathematical entities. In the sciences, mathematical models are used to represent a wide variety of real objects and situations.

Finally, there are metamathematical models, which are technical entities used to understand formal systems, such as mathematical systems themselves.

2 Models in metamathematics

In metamathematics (or 'metalogic', the study of the features of formal systems), a 'model' is a structure that makes all of the sentences in a theory true, where a 'theory' is a set of sentences in a language, the various terms of which are mapped onto objects in the system and their relations. Even in metamathematics, though, 'model' is used in several ways: most generally, a model is a special kind of interpretation, where an interpretation of a

theory consists in both the assignment or mapping of terms in the theory to objects, and to a structure, which consists of objects and their relations to one another. More specifically, an interpretation of a language specifies: (1) a domain (universe of discourse), for example, the range of any variables that occur in any sentence in the language; (2) a designation (denotation, bearer, reference) for each name in the language; (3) a function f, which assigns a value in the domain for any sequence of arguments in the domain, for each function symbol in the language; (4) a truth-value for each sentence letter in the language; and (5) a characteristic function for each predicate letter (Boolos and Jeffrey [1974] 1980). In these cases, models are defined as those interpretations under which all the sentences in the theory are true.

Alternatively, models are simply the structures themselves under a specified mapping assignment (where the mapping assignment is taken to be external to the model; Robinson 1965; Tarski 1941). Models can be any kind of structure; in metalogic, models are usually expressed in terms of set theory, but technically any group of objects and their relations could serve as the model for a theory, provided that it displayed the right structure.

For example, take as a theory the sentences: ‘object A is touching object B’; ‘object C is touching object B’; and ‘object C is not touching object A’ (where A, B and C are terms in the theory, and ‘is touching’ is a relation between two or more of the terms). We can easily construct or imagine a structure which ‘satisfies’ or makes true all of these sentences in the little theory: it could consist of three objects in a row, 1, 2, 3, each one touching only the next. Notice that they could be any kind of object, including cats, jars of jam, and so on, or some mixture of these. The *mapping assignment* for the theory might map A onto 1, B onto 2 and C onto 3. (Equally, it could map A onto 3, B onto 2, and C onto 1.) On the usual logical definition, then, the objects 1, 2, 3 constitute one model of the theory, because a ‘model’ is a structure which can be interpreted so as to make all the sentences in a specific theory true (see Model theory).

3 Models in science

Several uses of ‘model’ are standard in the sciences. In each case, models are substitute systems used to investigate and understand the real systems they model. In biology, certain organisms are selected as ‘models’ on the basis of ease of investigation and manipulation; they are investigated intensively, in the hope that the results generalize to other organisms. In biomedical research, for instance, mice are often used as models or stand-ins for studies of drug effects in human beings.

One of the most basic uses of models in physics involves mechanical models of natural processes. Thus, billiard balls in random motion are taken as a model for a gas. This modelling relation does not imply that billiard balls are like gas particles in all respects, simply that gas molecules are analogous to billiard balls. Under the model, some properties of billiard balls ought to be ascribed to gas molecules, that is, motion and impact (the ‘positive analogy’, in Mary Hesse’s terms), while other properties of billiard balls ought not be ascribed to molecules, such as colour and hardness (the ‘negative analogy’). There is also the ‘neutral analogy’, used when we do not know whether the properties are shared; according to Hesse, these properties are what allow us to make new predictions (Hesse 1966: 9). Max Black argued that ’use of a particular model may…help us to notice what otherwise would be overlooked, to shift the relative emphasis attached to details - in short, to *see new connections*’ (Black 1962: 237). N.R. Campbell (1920) claimed that during the development of the kinetic theory of gases, this mechanical billiard ball model of the theory played an essential part in its extension, and thus, that the availability of a working model of a theory is essential to successful theorizing in the sciences. Pierre Duhem, in contrast, argued that such use of models in science was preliminary, optional, and potentially misleading, and that proper scientific theories were expressed abstractly and systematically (Duhem 1914).

Many of the philosophical and scientific discussions about the roles of models in science seem to have mixed together several distinct issues. All participants in these debates agree that there must be some analogy between a model used in the sciences (whether mechanical or not) and the phenomena it is being used to explain. One primary concern was how realistically and completely a model - especially a mechanical model - is supposed to represent the aspect of reality under investigation. Hence, Peter Achinstein (1968) suggests a hierarchy of models based on their ontological commitments: at one end, we have models which are simply supposed to provide possible mechanisms for how natural systems might be operating, while at the other end, we have concrete claims that the real world is thoroughly like the entities and dynamics in the model. In the latter case, the choice of model amounts to a metaphysical commitment regarding the contents of the universe - which things and relations really

exist. A great deal of discussion has thus centred around which of these attitudes - from purely instrumental to strongly ontological - scientists and philosophers should take to the models they use.

The degree to which a model is taken as exactly and adequately representing reality is one of the most significant ways that claims about models in science can differ. What is frequently called an ‘instrumental’ use of models amounts to treating the models as calculating devices; slightly more committed, ontologically, are idealized models, which can be seen as either false-but-handy simplifications of processes in the natural system being discussed, or as approximately true representations of some of the forces operating in the natural system (see Idealizations).

One link between the metamathematical and scientific uses of models lies in the notion of interpretation. In the sciences, models are sometimes used as tools for making sense of theories which are not otherwise immediately comprehensible; a model can thus help the community of scientists articulate and pursue scientific theories. The wave and particle interpretations of light and of quantum mechanics are well-known examples; both the wave and particle models offer ways to understand the theoretical (and well-confirmed) equations, but they present different and incompatible features (see Quantum mechanics, interpretations of).

Finally, models are generally acknowledged to be psychologically valuable to scientists; they serve the heuristic functions of helping scientists envisage very complex systems, and they help simplify inferences about those systems. Even models that are known to be misrepresentations of the real world may sometimes possess these virtues (Wimsatt 1987).

4 Models in philosophy of science

While discussion of the above uses of models in the sciences occupied philosophers of science for many decades, one of their basic assumptions was that models are essentially different from scientific theories. This distinction has come under scrutiny since the 1950s, and much work in recent philosophy of science going under the name of ‘structuralist’ or ‘semantic’ approaches has centred on analysing scientific theories in terms of metamathematical models.

For the middle third of the twentieth century, the reigning logical positivist approaches to science understood scientific theories as ’sets of deductively connected sentences in a formal language’ combined with rules for interpreting some of the terms of that language. In contrast to this linguistic view of theories, some philosophers advocated viewing theories as structures, ‘which are propounded as standing in some representational relationship to actual and physically possible phenomena’ (Suppe 1979: 320). More specifically, scientific theories are understood as presenting models in the metamathematical sense, that is, arrangements of objects and their relations. These structures function as iconic models for the scientists using them: they characterize, in idealized circumstances, the systems that they represent.

Take evolutionary population genetics as an example. Population geneticists tend to present their theories in the form of mathematical models. This means that, given the mathematical models, it is possible to examine the structures that instantiate that theory. Under this ‘semantic approach’ to theories, the focus is on the structures themselves, rather than on an attempt to reconstruct, in some theoretical language, the sentences of the theory, as demanded by the axiomatic and positivist approaches to theory structure. (It remains a matter of debate whether there is an important difference between viewing a theory as a set of sentences and viewing it as a set of models. Some philosophers have suggested that the difference is merely pragmatic or heuristic: scientists sometimes find it more natural to regard a theory one way rather than another, but the two views are in principle intertranslatable. See Schaffner 1993: 99-125.)

This use of metamathematical models is an extension from their use in interpreting formal systems. When models serve as icons, some aspects of the model’s entities and their interrelations are taken to be similar to the phenomena in the natural world that are being investigated. In metamathematics, the more precise notion of isomorphism is central to evaluating the similarity of models. Two models are isomorphic if it is possible to make a one-to-one mapping from each element in one model to an element in the other model, and to make a one-to-one mapping from each relation among elements in one model to a relation among corresponding elements in the other model. While the notion of similarity between model and nature may be intuitive, precision demands that the natural system be represented, abstracted or measured in some fashion, if evaluations of isomorphism between the

theoretical model and nature itself are to be attempted. Typically, measurements or data from the natural system are arranged into a ‘data model’, which provides the basic entities and relations to which the theoretical models are compared (Suppes 1962).

Many questions can be raised about the relations between scientific theories - when understood as models - and the real world. For instance, it is possible that there are relations or entities that appear in the models that have no correspondence in the real world. Alternatively, there may be mechanisms represented in the model which are hypothesized to represent the way the natural world really works; the model might then be interpreted as an approximation or idealization of the real world systems, including mechanisms that we cannot observe directly. Debates among philosophers about ‘scientific realism’ and ‘antirealism’ have centred on the issues of what sorts of inference are necessary, or are justified, by the successful application of a scientific model to the real world (Churchland and Hooker 1985; van Fraassen 1980; Giere 1988; Suppe 1989).

See also: Scientific method; Scientific realism and antirealism; Theories, scientific

ELISABETH A. LLOYD

References and further reading

Achinstein, P. (1968) *Concepts of Science*, Baltimore, MD: Johns Hopkins University Press.(Referred to in §3. Careful analysis of various conceptions of models and their uses in the sciences.)

Black, M. (1962) *Models and Metaphors: Studies in Language and Philosophy*, Ithaca, NY: Cornell University Press.(Referred to in §3. Insightful discussion of philosophical understanding of models and the uses of models in scientific thought; not very technical.)

Boolos, G. and Jeffrey, R.C. (1972) *Computability and Logic*, New York: Cambridge University Press; 2nd edn 1980.(Second edition referred to in §2. Standard text for metalogic. Requires familiarity with predicate logic.)

Burks, A.W. (1949) ‘Icon, Index, and Symbol’, *Philosophical and Phenomenological Research* 9: 673-89. (Review of Peirce’s semiotics, good explication of the concept of ‘icon’ proposed by Peirce.)

Campbell, N.R. (1920) *Physics, the Elements*, Cambridge, Cambridge University Press; repr. as *Foundations of Science*, New York: Dover, 1957.(Referred to in §3. Controversial early advocacy of the centrality of models in scientific research.)

Carnap, R. (1942) *Introduction to Semantics*, Cambridge, MA: Harvard University Press.(Provides more background and discussion than many other basic texts. Useful introductory chapters and examples of metamathematical models.)

Churchland, P.M. and Hooker, C.W. (1985) *The Image of Science*, Chicago, IL: University of Chicago Press. (Referred to in §4. Useful collection of essays addressing the interpretation of models in science.)

Duhem, P. (1914) *The Aim and Structure of Physical Theory*; trans. P.P. Wiener, Princeton, NJ: Princeton University Press, 1954, 2nd edn.(Referred to in §3. Influential criticism of the role of mechanical models in physics.)

Fraassen, B.C. van (1980) *The Scientific Image*, Oxford: Clarendon Press.(Referred to in §4. Influential discussion of models in science. Defence of empiricist interpretation of scientific theories.)

Giere, R. (1988) *Explaining Science: A Cognitive Approach*, Chicago, IL: University of Chicago Press.(Referred to in §4. Accessible discussion and applications of the semantic conception of scientific theories. Excellent introductory chapters.)

Hesse, M.B. (1966) *Models and Analogies in Science*, Notre Dame, IN: University of Notre Dame Press.(Referred to in §3. Original and influential discussion of the scientific and logical uses of models.)

Peirce, C.S. (1934) *Collected Papers of Charles Sanders Peirce*, vol. II, Cambridge, MA: Harvard University Press.(Original, foundational work on semiotics, icons, symbols; fairly challenging.)

Robinson, A. (1965) *Introduction to Model Theory and to the Metamathematics of Algebra*, Amsterdam: North Holland.(Referred to in §2. Authoritative text, primarily technical.)

Schaffner, K.F. (1993) *Discovery and Explanation in Biology and Medicine*, Chicago, IL: University of Chicago Press.(Referred to in §4. Discusses the semantic view and alludes to the debate about differences between the semantic and syntactic views.)

Suppe, F. (1977) *The Structure of Scientific Theories*, Urbana, IL: University of Illinois Press, 2nd edn.(Excellent review of scientific and philosophical discussions concerning models. Comparison of logical positivist and semantic conceptions of theory structure. Outstanding bibliography.)

Suppe, F. (1979) ‘Theory Structure’, *Current Research in Philosophy of Science*, East Lansing, MI: Philosophy of Science Association.(Quoted in §4; useful brief review of some central issues in philosophy of science.)

Suppe, F. (1989) *The Semantic Conception of Theories and Scientific Realism*, Urbana, IL: University of Illinois Press.(Referred to in §4. Best source for history and context of treatments of models in philosophy of science. Careful discussion of issues regarding how extensively scientific theories represent the natural world.)

Suppes, P. (1962) 'Models of Data', in E. Nagel, P. Suppes and A. Tarski (eds) *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, Stanford, CA: Stanford University Press.(Referred to in §4. Sophisticated but accessible analysis of the extraction and organization of quantitative evidence from nature, and of its role in theory testing.)

Suppes, P. (1967) 'What is a Scientific Theory?', in S. Morgenbesser (ed.) *Philosophy of Science Today*, New York: Basic Books.(Influential defence of the use of metamathematical models in philosophy of science. Clear and accessible.)

Tarski, A. (1941) *Introduction to Logic and to the Methodology of the Deductive Sciences*, New York: Oxford University Press; 2nd edn, 1946.(Referred to in §2. Presentations of formal logic and models by one of the developers of metamathematics.)

Wartofsky, M. (1979) *Models: Representation and the Scientific Understanding*, Dordrecht: Reidel.(Referred to in §1. Original essays discussing the metaphysical, epistemological and social significance of various types of model. Excellent and readable.)

Wimsatt, W. (1987) 'False Models as Means to Truer Theories', in N. Nitecki and A. Hoffman (eds) *Neutral Models in Biology*, London: Oxford University Press. (Referred to in §3. Insightful discussion of a variety of attitudes towards models in the sciences.)

Modernism

As a period in cultural history, modernism usually denotes advanced or avant-garde European and American art and thought, though it has also been used to describe more general social conditions and attitudes. Most historians of literature and the plastic arts - the fields in which the term has most play - date it from the late 1880s to the Second World War. Modernism is thus distinguished from the 'modern' of 'modern history' (understood as anything since medieval history), 'modern life' (popular contemporary attitudes and difficulties), and other broad uses of the term 'modern'. In fact, recognition of the ism *in modernism is a key to understanding it - intense self-awareness being an essential characteristic or value, allied to modernism's complex engagement with avant-garde status. Other values that consistently underpin modernism include a propensity to create 'culture shock' by abandoning traditional conventions of social behaviour, aesthetic representation, and scientific verification; the celebration of elitist or revolutionary aesthetic and ethical departures; and in general the derogation of the premise of a coherent, empirically accessible external reality (such as Nature or Providence) and the substitution of humanly devised structures or systems which are self-consciously arbitrary and transitory.*

1 Epistemic trauma

In religious history, the term modernism still refers narrowly to the late nineteenth- and early twentieth-century movement within Roman Catholicism, and also to analogous developments in Protestantism and reformed Judaism. It constituted an attempt to bring nineteenth-century critical methods, especially those of pragmatism, to bear on subjects formerly regarded as beyond their influence, including the interpretation and application of sacred texts and other doctrines (see Pragmatism). It was condemned as a heresy in 1907 by Pope Pius X. This religious movement shares some of the values of the period in cultural history now commonly referred to as modernism, but remains quite distinct from it.

Given the historical intensity and complexity of the period, it is helpful to abstract several general values that characterize most manifestations of modernism. Such values are not specifically substantive or thematic, but dynamic and structural. They are the dominant qualities that sustain and distinguish the most advanced intellectual activity from the 1880s to the Second World War.

The most readily apparent of these values is that of *epistemic trauma*. This formulation signifies a kind of primary or initial difficulty, strangeness or opacity in modernist works; a violation of common sense, of laboriously achieved intuitions of reality; and an immediate, counter-intuitive refusal to provide the reassuring conclusiveness of the positivist realism that preceded modernism. This traumatic otherness stems in part from a conscious refusal by modernist artists and other thinkers to give their audiences the kind of spatial and temporal orientation that art and literature had been providing since the Renaissance and that had reached a high finish in the mid-nineteenth century, when novelists took pains to provide their 'dear Reader' with temporal and spatial coordinates and when the subject matter of most paintings was generally accessible. In a surprising and historically sudden contrast to this traditional solicitude, the cutting-edge artistic culture of modernism - and much contemporary social and scientific thinking - offered this quality of trauma everywhere. In the painting of Picasso and Braque, in the music of Stravinsky and Schoenberg, in the fiction of Kafka and Faulkner, in the poetry of Yeats and Eliot, the immediate difficulty, the epistemic trauma, is a given of the modernist aesthetic.

The *kind* of difficulty we find in modernist artistic culture resembles the kind of difficulty contemporaneous advances in mathematics and the natural sciences presented to the scientific establishment. In both spheres, the difficulty arises not so much from developments and complications of traditional techniques (what might be called *baroque* difficulty), but more often from what is left out. Relativity Theory (viewed as a vehicle for cultural values, like works of art and literature) provides a notorious example. For many physicists, the initial difficulty of the Special Theory lay in the fact that Einstein found the nineteenth-century hypothesis of an ether 'superfluous', thus radically pruning physics of a laborious but comfortably familiar hypothesis. In all manifestations of the modernist breakthrough, thinking people missed those qualities or techniques on which they had customarily relied for meaning, such as single-point perspective in painting, tonality in music, neutral and uniform time in narrative, and unvarying temporal and spatial reference frames in physics.

Out of this fundamental value of epistemic trauma a number of cognate characteristics emerged. Modernism

quickly developed an affinity to what seemed (when viewed from the tradition of realism) an addiction to gratuitous *difficulty* and *distortion*, either for their own sakes or for the sake of being avant-garde. In fact, discussions of modernism regularly equate it with the avant-garde; and while this equation ignores functional distinctions between the two terms, modernism's commitment to being at the cutting edge, to being shocking and difficult, quickly became a major value. In early modernism, when Cézanne began to depart from single-point perspective in the interest of greater truth to our visual experience of Nature, when Henry James began to introduce ambiguities of motive intolerable to the realist tradition, or when Max Planck introduced energy quanta, the apparent distortions and difficulties, the elitist value of being avant-garde, was not a principal motive or effect. But in a very short period, by the time Picasso had re-represented the human figure in *Les Demoiselles d'Avignon* and Kafka had transformed his protagonist into an insect in *The Metamorphosis* and Einstein had employed obscure mathematics to model a finite but unbounded universe in his General Theory, the antagonistic relation of modernism to popular culture was irreversible: it gradually led to the almost complete bifurcation of serious and popular culture. ' Modern [that is, modernist] art', wrote Ortega y Gasset in 1925 (1968: 5), ' will always have the masses against it. It is essentially unpopular; moreover, it is anti-popular' (see Ortega y Gasset, J.).

2 Reality and observation

One of the most prominent and lasting achievements of modernism in all its manifestations is the devaluation of the premise that we occupy an 'objective' reality, accessible to but independent of human perception. In traditional realism, artists and critics, social thinkers and scientists were thought to make direct statements about this reality, whether it be natural, social or psychological. Modernism essentially turns away from this realist enterprise and towards discussion and analysis of human measurement or observation.

This rich and profound departure is difficult to describe in brief, but an example from modernist physics helps to illustrate it. In the natural sciences, the great model of an objective, constant, external universe was based on Newtonian mechanics. Einstein undermined this coherent and rational structure by redirecting attention from the nature of reality to the nature of measurement, from what was taken to be our direct contact with nature to our observation of it. In his 1905 paper on Special Relativity, Einstein did not ask what time *is*: he asked how we *measure* it. He asked what we *mean* by the time of an event. In Relativity Theory, a measurement is neither a subjective impression (a unique event in a single mind that cannot be fully communicated) nor a constant, necessary description of an independent external object or event. Instead, it may be seen as a kind of middle ground - literally a mediation - between the observer and the observed phenomenon. And this middle ground is the characteristic epistemological location of modernism: its focus is on neither subject nor object but on the act of human observation of a reality presumed but not proved to be external to the observer.

In modernist painting we can see other manifestations of this shift in value. Cubism - with its subject-matter of bottles, tables, shreds of newspaper, musical instruments - clearly deflects our interest from the subject to its representation. In this process it derogates the specific importance of the historically or religiously or sentimentally significant subject (Christ on the Cross, the *Mona Lisa*, *The Rape of the Sabines*) to the problematics of representation, the aesthetics of composition, the formal language. What is significant about Picasso's *Portrait of Ambroise Vollard*, for example, is not the subject (Vollard's distinctive physical features, his support of modernist art, his aesthetic intuition) but how Picasso *presents* Ambroise Vollard. Just as Relativity Theory focused attention away from the nature of reality towards the nature of measurement and observation, so Cubism focused attention away from what was being represented towards how it was being represented.

Similar illustrations of this shift can be found in most movements of modernist art and literature, as well as in modernist developments in philosophy (see Phenomenological movement). But what is of greatest significance here is the liberating nature of this change in values. The practitioners of modernism felt themselves no longer locked into the limiting dichotomies of object and representation, world and observer. They began to move from one side to the other, to explore without interruption the unceasing interaction between the object and its space, or the event and its temporality. The old categories lost their integrity and the artist and subject, or observer and object, came to inhabit a middle ground of observation itself. In this sense of liberation, the modernist model for reality becomes the *field* (as in scientific field models), where observers are also participants (as in cubist painting or quantum theory), where readers help to create the text (as in Kafka's *The Castle* or Faulkner's *Absalom, Absalom!*), or in the many other modernist constructions in which all constituents are interdependent and in which

all participate and interrelate without privilege.

3 Major manifestations

The major literary and artistic movements most often cited as major manifestations of modernism include Post-Impressionism, Fauvism, Cubism, Expressionism, Futurism, Symbolism, Imagism, Vorticism, Dadaism and Surrealism. But such a list will be regarded as necessarily incomplete, and by definition it omits many modernist developments in architecture, philosophy, social science, psychology and the natural sciences - all areas sustained by and manifesting the values discussed above. But a catalogue of movements has the advantage of demonstrating one more value of modernism, that of *reflexivity*: the characteristic of self-awareness and of conscious programmatic direction. We have been regarding modernism as a historical period in high culture and therefore as what Astradur Eysteinsson (1990) calls a ‘cultural force’, but to the extent that each of the movements contained in modernism was aware of its newness and its cohesiveness (that is, of being avant-garde) it can be viewed as an ‘aesthetic project’, the self-conscious endeavour of a group of innovators to further certain values and achieve recognition.

4 Towards postmodernism

Cultural historians and theorists of modernism tend to view what is now called postmodernism in one of two general ways: as part of the original revolutionary gesture in which postmodernism constitutes a late and distinguishable development; or as a new ethos and a new aesthetic (new collection of ‘values’). Preference for one view or the other depends largely on the size of one’s historical canvas, but in either case such discussion proves helpful in defining modernism itself. This is because we can see changes in or departures from the values that constituted modernism, and these changes help to establish their original identities.

In the place of epistemic trauma, postmodernism affirms a denial of prescriptive norms and derogates the *value* of the normative. Where modernism wrestled with the difficulties caused by the absence of universal temporal, spatial, and ethical coordinates, postmodernism adopts without struggle the surreal, bizarre, and meta-natural. Where modernism negotiated the difficult transition from examination of reality to examination of observation, postmodernism accepts the premise that all values are ‘constructions’ on the model of language - local, contained, self-referential. Where modernism sought depth and abstraction, postmodernism turns toward surface and particularity. For what was in modernism a dominant value of ‘order’, postmodernism substitutes ‘design’ and ‘pattern’. Finally, the reflexivity of modernism, its self-awareness, becomes in postmodernism a more radical and complete self-referentiality moving toward visions of total (and so completely free) self-containment (see Postmodernism).

See also: Art, abstract; Relativity theory, philosophical significance of

THOMAS VARGISH

References and further reading

Bradbury, M. and McFarlane J. (eds) (1976) *Modernism: 1890-1930*, London: Penguin. (Excellent historical essays.)

Ermarth, E.D. (1992) *Sequel to History: Postmodernism and the Crisis of Representational Time*, Princeton, NJ: Princeton University Press.(So far the most accurate and theoretically complete study of postmodernism.)

Eysteinsson, A. (1990) *The Concept of Modernism*, Ithaca, NY and London: Cornell University Press.(A full account of critical usage.)

Kern, S.J. (1983) *The Culture of Time and Space: 1880-1918*, Cambridge, MA: Harvard University Press.(Unruly but indispensable and comprehensive cultural history of modernist technology and culture.)

Ortega y Gasset, J. (1923) *The Modern Theme*, New York: W.W. Norton, 1933. (Ortega’s assessment of his contemporary culture which remains valuable, despite being partial and opinionated.)

Ortega y Gasset, J. (1925) *The Dehumanization of Art and Other Writings on Art and Culture*, Princeton, NJ: Princeton University Press, 1968.(A major statement on the subject of modernism that has itself become an important modernist document.)

Vargish, T. and Mook, D. (1998, forthcoming) *Inside Modernism: Relativity Theory, Cubism, Narrative*.(An analytic description of modernism through the identification of values common to physics, painting, and narrative.)

Modularity of mind

A common view in recent philosophy of science is that there is no principled distinction between theoretical and observational claims, since perception itself is thoroughly contaminated by the beliefs and expectations of the observer. However, recent psychological and neurological evidence casts doubt on this latter claim and suggests, instead, that perceptual processing is to a significant extent 'cognitively impenetrable': it takes place in informationally encapsulated 'modules' that cannot be rationally influenced by beliefs or other 'central' cognitive states, or even other portions of the perceptual system.

1 Claims and distinctions

It might appear that what we see depends heavily on what we expect and what we know about the situation at hand. That is how magicians manage to deceive us, and how we sometimes deceive ourselves into hearing and seeing what we have reason to believe is there, missing, for example, typographical errors that we recognize only when they are explicitly brought to our attention. This idea has also been supported by hundreds of experiments carried out in the 1950s, and has been the basis for a widespread view in psychology associated with Jerome Bruner (1957) and referred to as the 'New Look in Perception'. It has also been very influential among philosophers of science such as Hanson (1958) and Kuhn (1962), who have taken for granted that how we perceive the world is thoroughly conditioned by our prior expectations and beliefs (see Observation §§3-4; Perception).

In recent years, however, doubt has been cast on the premise that such processes as visual perception, face recognition, language comprehension are equally 'cognitively penetrable'. The issue is whether the flow of information among cognitive functions is governed by fixed architectural principles, or whether failures to access relevant information in the course of perceptual processing, though widespread, are adventitious (see Cognitive architecture). Despite the widespread acceptance of the idea that perception is contaminated by cognition, a careful examination of the psychological evidence actually favours the view that perception operates autonomously from the system of beliefs and inferences. Jerry Fodor (1983) has argued that certain cognitive capacities seem to be 'modular' or 'informationally encapsulated'. He focuses particularly on what he calls 'input systems', which include cognitive systems for perception and language (see Fodor, J.A.).

To make Fodor's views clear, however, we need to make some critical distinctions.

(1) The distinction between 'what we see' and what we come to believe about what we are looking at.

(2) The distinction between perceptual and post-perceptual effects such as those that depend on the exigencies of memory and decision processes, including the decision to make a particular response. (Sometimes post-perceptual effects can be mathematically factored out using a signal-detection method that allows one to separate 'detection' from 'response selection' effects - the latter of which clearly depend on beliefs and utilities; see Samuel (1981).)

(3) The distinction between *extra*-modular *cognitive* penetration of perception, and other top-down but *intra*-modular effects. It is clear that how certain local features are perceived may depend on more global aspects of percepts - for example, how a phoneme is perceived may depend on word-level acoustical perception - but these generally are not effects of cognition but of broader aspects of perception. There are plenty of examples in vision where how a local feature is perceived depends on how the larger scene is perceived, but where the latter itself is insensitive to more general beliefs and inferences.

2 Evidence

Perhaps the most striking evidence for modularity of perception is the imperviousness of various illusions to knowledge: even when you have every reason to believe that something you are seeing is an illusion, you cannot make the illusory percept disappear. Moreover, perception appears to have a logic all its own: if you see some part of a pattern in a certain way you will automatically tend to see other parts of the pattern in a compatible way. For example, if you see the reversing Necker cube (see Gestalt psychology §2) in one orientation you will see each of its edges in the appropriate relative location, and will even see the apparently-further face as bigger.

There is a great deal of evidence for the existence of pre-recognition stages in perception that are clearly immune

to cognitive influences. (Stich (1978) calls the output of such stages, characterized by their lack of 'inferential promiscuity', 'sub-doxastic states'.) In addition there appear to be even finer submodules within the visual system. For example, there is considerable independence of the colour perception system, the motion detection system, and the form recognition system. Stereo depth and motion-induced depth perception (so-called kinetic depth effect) do not require prior recognition of form. Examples of this sort are commonplace, and all point to there being architectural constraints on communication among parts of the visual system and between the visual system and the rest of cognition. In the face of this kind of evidence it appears that at least the principle of modularity is well established, even when the exact locus of the modules and the type and extent of encapsulation may be open to debate.

Related evidence is also provided by various pathologies of perception. In these cases of dysfunction (due to brain damage or illness) patients frequently show signs of dissociation of functions, providing clear evidence that a certain function, which is absent in particular patients, is not required in order to accomplish some other function, which remains intact in those patients. For example, colour-blindness can occur without form-blindness.

Although functional decomposition is commonplace, it is the special case of the dissociation of much of perception from the rest of the cognitive system that is of particular importance to philosophy. What such cognitive impenetrability suggests is that there is reason to distinguish between *data*-driven and *theory*-driven aspects of perception. And if that is the case, then the observation/theory distinction may have some empirical support after all, and one of the arguments for epistemological holism may also be undermined. (See the exchange between Fodor (1984; 1988) and Churchland (1988) for discussion of this issue.)

See also: Cognitive architecture; Perception, epistemic issues in; Perception; Vision

ZENON W. PYLYSHYN

References and further reading

Bruner, J. (1957) 'On Perceptual Readiness', *Psychological Review* 64: 123-52.(Early defence of the claim that observation was thoroughly penetrated by theory.)

Churchland, P. (1988) 'Perceptual Plasticity and Theoretical Neutrality: A Reply to Jerry Fodor', *Philosophy of Science* 55: 167-87.(Reply to Fodor 1984.)

Fodor, J. (1983) *The Modularity of Mind: An Essay on Faculty Psychology*, Cambridge, MA: MIT Press.(The main statement of the modularity thesis.)

Fodor, J. (1984) 'Observation Reconsidered', *Philosophy of Science* 51: 23-43.(Fodor's effort to use his claims about modularity to resuscitate the observation/theoretic distinction in the philosophy of science.)

Fodor, J. (1985) 'Précis of *Modularity of Mind*', *Behavioral and Brain Sciences* 8: 1-42.(Short version of Fodor (1983), with commentaries by many philosophers and psychologists and replies by the author.)

Fodor, J. (1988) 'A Reply to Churchland's "Perceptual Plasticity and Theoretical Neutrality"', *Philosophy of Science* 55: 188-98.(Reply to Churchland's 1988 reply to Fodor 1984.)

Garfield. J. (ed.) (1987) *Modularity in Knowledge Representation and Natural Language Understanding*, Cambridge, MA.: MIT Press.(A diverse collection of essays by philosophers and psychologists on the issue of modularity.)

Hanson, N. (1958) *Patterns of Discovery*, Chicago, IL: University of Chicago Press.(Early philosophical effort to show how observation was inseparable from theory.)

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(Highly influential attack on the suggestion that there was any theory-neutral observations that could provide a neutral basis for comparing theories.)

Samuel, A. (1981) 'Phoneme Restoration: Insights from a New Methodology', *Journal of Experimental Psychology: General* 110 (4): 474-94.(Subtle discussion of how to distinguish experimentally between 'signal detection' and 'response selection'.)

Stich, S. (1978) 'Beliefs and Sub-Doxastic States', *Philosophy of Science* 45: 499-518.(Introduces the idea of isolated belief-like states that are relatively independent of the rest of one's belief system, as in the case of the kind of 'knowledge' of grammar that Chomsky attributes to normal language speakers.)

Mohist philosophy

Mohist philosophy describes the broad-ranging philosophical tradition initiated by Mo Ti or Mozi (Master Mo) in the fifth century BC. Mozi was probably of quite humble origins, perhaps a member of the craft or artisan class. Early in life, he may have studied with followers of Confucius. However, he went on to become the first serious critic of Confucianism.

Mozi's philosophy was part of an organized utopian movement whose members engaged in direct social action. He was a charismatic leader who inspired his followers to dedicate themselves to his unique view of social justice. This required them to lead austere and demanding lives, as he called upon them to participate in such activities as the military defence of states unjustly attacked.

Mozi is arguably the first true philosopher of China. He was the first to develop systematic analyses and criticisms of his opponents and present carefully argued positions of his own. This led him and his later followers to develop an interest in and study of the forms and methods of philosophical argumentation, which contributed significantly to the development of early Chinese philosophy.

Mozi saw ideological differences and the factionalism they spawned as the primary source of human suffering, and he hotly criticized the familially-based ethical and political system of Confucius for its inherent partiality. In its place he advocated three basic goods: the wealth, order and the population of the state. Against the Confucians, he argued for jian'ai *(impartial care).* Jian'ai *is often translated as 'universal love', but this is misleading. Mozi saw the central ethical problem as an excess of partiality, not a lack of compassion; he was interested not in cultivating emotions or attitudes, but in shaping behaviour. He showed remarkably little interest in moral psychology and embraced an extremely thin picture of human nature, which led him away from the widely observed Chinese concern with self-cultivation. His general lack of appreciation for psychological goods and the need to control desires and shape dispositions and attitudes also led him to reject the characteristic Confucian concern with culture and ritual.*

Mozi believed human beings possess an extremely plastic and malleable nature, and he advocated a strong form of voluntarism. For several different reasons, he believed that people could be induced to take up almost any form of behaviour. First, he shared a common early Chinese belief in a psychological tendency to respond in kind to the treatment one receives. He further believed that, in order to win the favour of their rulers, many people are inclined to act as their rulers desire. Those who do not respond to either of these influences can be motivated and controlled by a system of strict rewards and punishments, enforced by the state and guaranteed by the support of Heaven, ghosts and spirits. Most important of all, Mozi believed that rational arguments provide extremely strong if not compelling motivation to act: presented with a superior argument, thinking people act accordingly.

The social and political movements of the later Mohists lasted until the beginning of the Han Dynasty (206 BC). They continued Mozi's early interests and developed sophisticated systems of logical analysis, mathematics, optics, physics, defensive warfare technology and strategy and a formal ethic based upon calculations of benefit and harm. All the philosophical concerns of the later Mohists can be found in the early strata of the Mozi, and seem to reflect the teachings of the tradition's founder.

1 Nature of the work

The present text of the *Mozi* consists of fifteen books unevenly divided into seventy-one chapters, eighteen of which are now lost (for a complete list, see the table of contents of Y.P. Mei's translation (1929)) (see Mozi). Contemporary scholars tend to divide the book into five parts: the Epitomes (chapters 1-7), Essays (8-39, seven of which are lost), Logical Chapters (40-5), Dialogues (46-50) and Military Chapters (51-71, nine of which are lost). The Epitomes are clearly of later origin and, in several cases, of questionable provenance. The Essays are the core of the text and offer a complete and consistent account of Mohist philosophy. Depending on one's view, they present either ten or eleven different themes: 'elevating the worthy', 'identifying with one's superior', 'impartial care', 'against aggressive warfare', 'economy in expenditures', 'economy in funerals', 'heaven's intention', 'explaining ghosts', 'against music', 'against fate' and 'against Confucians'. For both textual and philosophical reasons, the last does not seem to fit with this group and will not be considered here as part of the Essays' corpus.

Originally there were three separate versions of each of the remaining ten themes, which are thought to represent three different sects of later Mohism. The existence of three later Mohist sects is attested to in collateral sources, though we know very little about their distinctive natures or their relationship to these chapters. Graham (1985) argued for a rearrangement of the chapters, which he believed revealed distinct doctrinal differences, but his analysis remains controversial and the philosophical differences he identified are not significant. The repetitive and unadorned style of these triadic chapters has led several scholars to speculate that they may be different versions of recorded sermons or lectures, or part of an original oral tradition. In any event, they present the central doctrines of early Mohist philosophy and may even be an accurate record of the teachings of Mozi himself. Later Mohists tend to write in a similarly unembellished fashion but ostensibly for a definite purpose: they sought to eliminate all stylistic flourish, believing it could only obscure the weakness or interfere with the strength of an argument.

The Dialogues seem to be of slightly later date and, like the *Analects* of Confucius (see Confucius), record conversations between Mozi and various disciples and opponents and stories about his exploits and travels. The Logical Chapters and Military Chapters both appear to be of considerably later date, and develop in systematic fashion Mozi's early concern with the forms of argumentation and the practical aspects of defensive warfare.

2 Central philosophical doctrines

Like Xunzi (and also like Thomas Hobbes), Mozi described a state of nature which is an unfulfilling war of all against all. However, he did not see this state as the result of our having shared and unruly desires which lead us into competition and conflict over scarce goods. For Mozi, human beings are led to fight with one another because they hold individual and irreconcilable notions of what is *yi* (right). The fundamental problem is a fragmentation of values. Hence the most pressing task is to get people to agree on a single notion of what is right. Mozi believed that the right consists in maximizing the collective well-being of the group: those who support and work for this unified goal put the common good ahead of their personal well-being and hence are, by definition, the most 'worthy' people for public office. Thus they deserve to be promoted to positions of authority in the state and deferred to by those they supervise. Mozi argues for these ideas in the first two groups of synoptic chapters: 'Elevating the Worthy' and 'Identifying with One's Superior'.

Mozi believed that the control and redirection of human desires was not a problematic task, nor did he believe one needed to develop virtuous attitudes and dispositions in order to behave ethically. Since he saw no need to develop or shape the self through ritual practice, introspection or insight, he was not, like most early Chinese philosophers, a self-cultivationist (see Self-cultivation in Chinese philosophy). Nevertheless, he was not a Hobbesian-style contractarian either. He believed that people could be influenced to adopt a non-competitive mode of behaviour through a variety of means. Hence, in theory at least, he avoided the difficulties associated with problems like the prisoner's dilemma (see Rational choice theory §3).

Mozi relied upon four basic techniques to influence people's behaviour. First, he appealed to a rather mechanistic version of the generally accepted early Chinese belief that people are inclined to respond in kind to the treatment they receive. Second, he believed that many would act as their ruler desired, simply to curry favour. Third, he advocated a system of well-defined rewards and punishments designed to move people away from certain types of behaviour and toward others. Consistent with his moral psychology, his list of rewards and punishments contained only material goods, with the single exception of 'honour' or 'prestige'. In this respect he anticipated later Legalist philosophers such as Han Feizi, who like him were not self-cultivationists and who relied upon the power of the state to shape behaviour through a publicly-proclaimed system of rewards and punishments (see Legalist philosophy, Chinese). However, Mozi differed from these later thinkers in believing that his system of rewards and punishments had Heaven's active and direct support (see Tian). In the seventh group of synoptic chapters, 'Heaven's Intention', he claimed that Heaven directly intervenes in human affairs to reward and punish, and in the eighth group, 'Explaining Ghosts', he argued that ghosts and spirits act as Heaven's agents to ensure that justice is upheld.

Mozi's fourth basic approach to influencing people's behaviour entailed an extreme form of voluntarism. He believed that most, or at least many, people could simply take up a form of behaviour and would do so if they were given good reasons for adopting it. Mozi argued that anyone who truly understood that a given form of behaviour does indeed maximize the common good (as he understood it) would immediately act accordingly. This belief in

the inexorable power of argument is of a piece with his sparse moral psychology and belief in the plasticity of human nature, and helps to explain his active and persistent interest in the forms and method of philosophical debate.

3 Central philosophical doctrines (cont.)

The most important philosophical position Mozi sought to establish is *jian'ai* (impartial care). He argues for this view in the third group of synoptic chapters, which take this theme as their title. Impartial care is the key to Mozi's philosophical system and the basis for many of his other teachings. He believed it is the only way to correct the destructive tendency to form factions and show preference to oneself, one's family and one's group. The unity that results when everyone adopts a policy of impartial care optimizes the wealth, population and order of the state.

Mozi often asserts or assumes that maximizing these collective goods will result in maximizing individual good. At times he does argue for this connection: for example, he argues that a truly filial man will embrace impartial care as the best way to serve his parents, wife and children. Like Confucius, Mozi develops his ethical philosophy from the perspective of groups rather than individuals. Here, their major difference is that for Confucius the family is primary while for Mozi the collective or state is most important. Impartial care does not entail egalitarianism. Mozi saw *jian'ai* as perfectly compatible with a strict social hierarchy and an absolute monarchy. Like almost every early Chinese philosopher, he appealed to a past golden age when his own proposed policy had been actually practiced. Such an appeal was thought to show first, that his theory was feasible, and second, that the practice of this theory was the reason for the success of earlier society (see History, Chinese theories of).

Mozi further believed that Heaven itself practised impartial care, and desired human beings to do the same. He argued for this view in the seventh group of synoptic chapters, 'Heaven's Intention'. In early China, it was widely held that if one could discern and follow Heaven's plan, one would be in accord with a pattern inherent in the universe itself, which would normally lead to ease in action and to success. Mozi accepted this general idea and insisted that following Heaven's plan would guarantee reward. As he argued in the tenth group of synoptic chapters, 'Against Fate', to believe anything less was to succumb to a pernicious form of fatalism. Since Heaven practises impartial care and is omniscient and all-powerful, good people will always be rewarded for their good deeds and bad people will always suffer punishment. Since the ideal state is modelled on the pattern of Heaven, rewards and punishments will be the strict policy of any just ruler.

Another important and related Mohist assumption is that the good is always useful, in the sense of directly producing material benefit for the world. One can see this throughout Mozi's philosophical teachings. Impartial care is good and directly benefits the world. A belief in 'fate' is bad and entails the negative consequences of complacency and defeatism. This linkage of the good with the useful occasionally led Mozi to conflate the notions of the 'good' and the 'true'. For example, as part of the justification for his claim that ghosts and spirits exist, Mozi argues that believing in their existence is useful: it influences people to be ethical. It is tempting to claim that Mozi is merely relying on a 'useful fiction' in order to achieve a greater good, but such an interpretation must ignore or explain away considerable textual evidence supporting his belief in ghosts and spirits, in particular the entire eighth group of synoptic chapters. Such a view must also ignore or explain away Mozi's belief that Heaven guarantees that there is absolute justice in the world. Finally, such an interpretation must ignore the fact that Mozi did not support this claim with an appeal to its utility alone. In fact, he employed three basic tests for any doctrine.

In the last group of synoptic chapters, 'Against Fate', Mozi presents the three tests that any theory should pass. It is significant that the word translated here as 'test' is literally a gnomen, an astronomical instrument used to determine the direction of the rising and setting sun. Thus Mozi's three tests allow one to orient oneself to the way things are in the world: 'We must set up a standard of judgement, for to try to speak without a standard of judgement is like trying to establish the direction of sunrise and sunset with a revolving potter's wheel…' (*Mozi* 35, in Watson 1963: 117-8). These three tests are precedent, evidence and utility.

The first test is an appeal to the policies and beliefs of idealized past sages. While the Mohists recognized the need for technological innovation, they share the common Chinese belief that the important theoretical discoveries regarding self and society were all known to a select group of past sages (see Chinese Classics). The Mohists tended to appeal to the earlier sage-king Yu rather than to the favourite Confucian exemplar, the Duke of Zhou, but

they often appealed to the same individuals as the Confucians did, ascribing to them different beliefs and practices. Any proffered theory ought to find precedent in the practices and beliefs of these former sages. Such precedent establishes both its feasibility and efficacy. The second test is an appeal to commonly held beliefs and practices, something like an appeal to 'common sense'. Any theory that is not supported by evidence derived from the 'eyes and ears' of the people is not to be accepted. Finally, the third test concerns measuring the consequences the theory has for the common good: the aggregate wealth, order and population of the state. As Graham has pointed out (1989: 37-8), the theories presented in the ten synoptic chapters themselves are all supported by these three tests. However, for most, only the first and third tests are applied. The second test is only employed to establish Mozi's claims regarding the existence of ghosts and spirits and the non-existence of destiny.

Mozi did not rank the three tests in any way. While it is true that he often pointed to the purported good consequences of his proposals, this may have been motivated more by rhetorical than logical considerations. It seems most reasonable to believe that for Mozi and his followers the truth of their theories was strongly over-determined: it could never be the case that the precedents of the sages or the common-sense views of the people could conflict with optimizing the collective good.

The three tests can be seen as related to another principle which Mozi developed and employed both to criticize his opponents and to establish his own case, namely, consistency. For example, in condemning aggressive wars Mozi argued that someone who steals one peach or kills one person is recognized as guilty, but those who steal and slaughter whole states are not (this argument was later parodied in the *Zhuangzi*). Here the call to consistency is combined with an appeal to common sense. In another chapter from the same group, Mozi argues that while aggressive warfare has worked for a few states, it comes only at the expense of many others. To claim that, as a principle, aggressive war is right '…is rather like the case of a doctor who administers medicine to over ten thousand patients but succeeds in curing only four' (*Mozi* 19, in Watson 1963: 59). Here the principle of consistency is combined with an appeal to the common good.

4 Central philosophical doctrines (cont.)

The remaining four groups of synoptic chapters, 'Against Aggressive War', 'Economy in Expenditures', 'Economy in Funerals' and 'Against Music', are all reactive policies. The first is directed against the general practice of rulers in his time, and the last three are specifically aimed at his Confucian competitors. While motivated as criticisms, these chapters are perfectly consistent with and representative of Mohist philosophy as a whole. All are justified by appeals to the three tests of a doctrine as well as Mozi's principle of consistency. Moreover, they all reveal his characteristic lack of appreciation for the value of psychological goods.

For example, in 'Against Aggressive War', Mozi deploys his full array of arguments: aggressive war fails all three tests and the asymmetry between the condemnation of killing one person and the celebration of killing whole populations violates his principle of consistency. As Lowe (1992: 18) has noted, Mohists were not really pacifists although they are often described as such. Mozi and his followers had a just-war theory and participated in what they regarded as ethically warranted warfare, but, as these chapters reveal, they were adamantly opposed to wars of expansion. What is striking in Mozi's searing and well-argued condemnation of aggressive war is the complete absence of descriptions of the horror of war. All Mozi seems to see is the unprofitability of war. His overriding concern with *li* ('benefit' or 'profit') led him to ignore a broad range of psychological goods and harms (see War and peace, philosophy of §1).

This absence of any consideration of psychological consequences is manifested throughout the remaining three groups of synoptic chapters, in which Mozi argues against the elaborate cultural embellishments, funeral practices and court music advocated by Confucians. While a reasonable case could be made against extravagance in any of these pursuits, Mozi does not acknowledge the possibility that such activities, properly moderated, might possess some value. He not only rejects the idea of their inherent value, he does not even consider that they might be instrumentally valuable in the task of shaping human behaviour.

Mozi's complete rejection of psychological goods shows that it is inaccurate to describe him as a utilitarian, at least in the classical sense (see Utilitarianism). He is not working to optimize aggregate pleasure and pain. In the absence of this classical utilitarian concern, it is at first glance difficult to see what justification or motivation Mozi can provide to support his ethical theory: in particular his call for impartial care. However, there is a clear

foundation for his ethical vision: his belief that it accords with or, more accurately, mirrors what Heaven intends (Lowe 1992: 132).

Mozi often appeals to a kind of divine sanction as justification for his ethical views, but it is important to understand that his notion of Heaven is not strongly theistic. While Heaven has intentions, these are a set of natural patterns or forces in the world, not acts of will originating from an independent agent standing outside the world and sustaining it. Heaven did not create the world and, while Heaven provides a model for human activity by showing impartial care for all, it is always reactive in showing its support or displeasure. Heaven is not a personal God who loves individuals, but a force which ensures that certain actions are met with appropriate reactions.

The relationship between human beings and Heaven is quite revealing and representative of Mozi's general orientation. The first point to note is that a person's attitude toward Heaven is not devotional: there is little or no emotion - never mind passion - involved. The intercourse between people and Heaven is a simple form of behaviour, a kind of transaction. There is nothing mysterious or awe-inspiring about Mozi's Heaven. As long as people act in accordance with the Heavenly pattern of impartial care, they can be assured that they will receive Heaven's support. If they deviate from this plan, they act against Heaven and will not escape its punishment.

Confucius believed that Heaven was on the side of good people and that acting in accordance with Heaven's will was the only way to realize certain personal and social goods. However, he also believed that no one was guaranteed Heaven's support. Heaven sometimes acts in ways that were inscrutable to human beings and this should not affect the way one behaves. He also believed that good people tend to have a strong influence on others that elicits from them reciprocal acts of goodness, and that this is the only legitimate way to influence others to be good; but again, one could never be guaranteed that such reciprocity is forthcoming. In either case, there was something fundamentally wrong with attempting to manipulate Heaven or humans by consciously working these natural tendencies to personal advantage. Even self-consciously pursuing virtue ran the risk of tainting one's character with thoughts of self-aggrandizement and pride (see Confucian Philosophy, Chinese).

Mozi's philosophy opposed almost all of these fundamental tenets of Confucius' thought. He believed that if there were not a direct and immediate guarantee of reward or punishment, the idea that Heaven supported ethical conduct was empty. If Heaven proved to be unknowable or unreliable, people would lose all faith in affecting their future improvement and would resign themselves to whatever they perceived as their fate. In a similar way, Mozi believed that people could be motivated to 'care' for others only by embracing a wholly impersonal system of the equitable distribution of material goods which guaranteed them treatment in kind. Mozi and his followers could never see virtue as, in any way, its own reward (see Virtue ethics).

5 Later Mohist philosophy

The later Mohists developed and extended the philosophical interests of their tradition's founder. Some of their most important advances came in what might be described as the science of argumentation (see Logic in China). They laid out definitions of necessary and sufficient conditions, inference and a priori, developed a version of the law of excluded middle, and analyzed and discussed such notions as universals and particulars. However, the way they approached these issues was distinctive and in some cases significantly different from the way early Greek thinkers worked (see Zeno of Elea). The Mohists were more empirical than abstract in their approach. They were trying to sort out certain problems; their work bore a greater resemblance to careful accounting, taxonomy or cartography than theoretical mathematics. It is significant, as Graham (1989: 60) has pointed out, that although the later Mohists took geometry as paradigmatic of clear and exact thinking, they never developed a discipline of geometric proof. They were content simply to illustrate that certain relationships and regularities obtained by appeal to geometric paradigms.

Owing at least in part to the nature of the classical Chinese language, where many logically significant distinctions within sentences such as number and tense are unmarked, the later Mohists did not develop formal logic. Like most Chinese philosophers, they began with *ming* (names) and sought to map these onto the things and events in the world in a clear and consistent fashion. Working in this way, they built up terms and strings of names to form sentences which would then either be 'so' (*ran*) or 'not so' (*buran*) of a given case. They noticed and sought to make clear that semantic differences - particularly those arising from the idiomatic meanings of compound terms - generate considerable problems when one attempts to draw parallels and analogies between sentences and arrange

them into classes.

One well-known example is the later Mohist treatment of the parallel sentences: ‘Robbers are people’ and ‘Killing robbers is not killing people’. While the second sentence simply appears false in English, this is because it has been rendered literally, thereby obscuring the semantic ambiguity of key terms. In Chinese, *shadao* (killing robbers) has the sense of ‘executing robbers’, while *sharen* (killing people) has the sense of ‘murder’. Thus the immediate sense of the sentence is something like, ‘Executing robbers is not murdering people.’ Many related cases were also studied, such as the sentences, ‘Increasing the number of robbers is not increasing the number of people’ and ‘Caring for robbers is not caring for people.’

The later Mohist approach to these problems is largely an attempt to define criteria for class membership. On the primary level of things, they sought to establish definite and distinct *lei* (‘categories’ or ‘classes’) based on observable similarities and differences. They were not seeking Platonic essences nor even the generic definitions of Aristotle; they were thoroughgoing and somewhat naïve nominalists (see Nominalism). Nevertheless, since they insisted on a strict accounting of what does and does not merit inclusion in a given category, they avoided many problems that have bedevilled Western philosophy concerning the notion of what is ‘true’, while still having a workable and functionally equivalent concept. They used a similar approach on the level of sentences, which they viewed as strings of names.

In order to argue for the similarity of separate cases, the later Mohists also developed the notion of ‘inference’ (*tui*). Again, Graham (1989: 150) is highly instructive in noting that the later Mohist notion of inference differs from induction, whereby one moves from the known to the unknown. Rather, it is used to argue for the *actual* similarity between cases. At their most abstract, the later Mohists employed an established category as a kind of template to assess the status of some actual thing or situation, to see whether or not it warrants inclusion within a given set. As is clear from the examples involving ‘people’ and ‘robbers’ presented above, class membership can be critically important (see Categories).

Mohist ideas regarding ‘categories’ and ‘inference’ had a profound effect on early Chinese philosophy. For example, they were adopted and played a major role in the thought of Mencius. As well as using Mohist logical terms of art, Mencius adapted them, in some cases enlarging their sense, to make specific points critical to his philosophical position. For example, he argued that the sage is the same in kind (*lei*) as other human beings (*Mengzi* 2A2, 6A7). He further argued that in order to ‘fill out the category’ of human being, one must ‘infer’ (*tui*) or ‘draw the analogy’ between paradigmatic cases of moral action and relevantly similar ones (*Mengzi* 1A7, 2A9). In a justly famous passage (*Mengzi* 1B8) strongly reminiscent of the Mohist examples presented above, Mencius argues that past sages who rebelled against and killed their evil rulers did not perpetrate the crime of regicide, since these rulers were not ‘kings’ but rather ‘mutilators’ and ‘outcasts’.

In the later Mohist material, there is a discussion of the relationship between benefit and harm and liking and disliking that appears to present a genuinely utilitarian ethic. However, closer inspection reveals that what is of value are not raw states of pleasure or pain nor even more lofty feelings of altruism or compassion. What is ‘liked’ or ‘disliked’ is determined a priori by ‘the sage’, who enjoys what produces the greatest collective good for the state. What is new is the explicit use of a functionally defined ideal observer, the sage. Earlier Mohists appealed to the authority of sages, but their appeal was always to actual past sages.

This was not the only change that later Mohists made in their ethical theory. In *Mengzi* 3A5, we find the Mohist Yi Zhi proposing a modified form of voluntarism (Nivison 1980). The debate begins with Mencius criticizing Yi Zhi for advocating frugal burials while at the same time providing lavishly for the burial of his own parents (essentially accusing Yi Zhi of failing to meet his own Mohist criterion of consistency). Yi Zhi replies first by quoting a classical passage (thus invoking the third ‘test’ and employing an appeal the Confucians themselves should accept), which he presents as supporting the Mohist teaching of ‘impartial care’. He then argues that while such care is the ideal and ultimate end, it naturally begins in the home. We start by caring for our parents and then, by applying the teachings of Mozi, we are to universalize this feeling until we care equally for all. Thus Yi Zhi softens the strong voluntarism of early Mohism and attempts to ground it in a Mencian-style appeal to innate moral inclinations.

Although they shared many important assumptions and concerns with other early Chinese thinkers, the later

Mohists were remarkably different in their approach. They did not seek for wisdom guided by the contemplation and synthetic grasp of historical paradigms, nor did they turn to intuitions about or reflections upon human nature. They trusted in careful observations and precise analyses with the aim of attaining a comprehensive and systematic understanding. The later Mohists were highly successful for several centuries but then, for reasons that remain obscure, their movement died out rather suddenly around the beginning of the Han Dynasty (206 BC). Nevertheless, they live on in at least one sense, in the profound and pervasive effect they have had on the Chinese philosophical tradition.

See also: Chinese Classics; Chinese philosophy; Confucian philosophy, Chinese; Daoist philosophy; Law and ritual in Chinese philosophy; Legalist philosophy, Chinese; Logic in China; Mozi

PHILIP J. IVANHOE

Reference and further reading

Graham, A.C. (1978) *Later Mohist Logic, Ethics and Science*, Hong Kong: The Chinese University Press, and London: School of Oriental and Asian Studies.(A remarkable reconstruction and analysis of the later Mohist sections of the *Mozi* which concern the systematic study of argumentation, ethics and science.)

Graham, A.C. (1985) *Divisions in Early Mohism Reflected in the Core Chapters of Mo-tzu*, Singapore: National University of Singapore, Institute of East Asian Philosophies. (An intriguing though not altogether persuasive study which argues that the synoptic chapters can be rearranged to reveal three schools of later Mohist thought, each with a distinct political agenda and philosophical position.)

Graham, A.C. (1989) *Disputers of the Tao: Philosophical Argument in Ancient China*, La Salle, IL: Open Court Press.(Contains an insightful study of the Mohist movement within the context of early Chinese thought.)

Lowe, S. (1992) *Mo Tzu's Religious Blueprint for a Chinese Utopia*, Lewiston, NY: Mellen Press.(A general study of early Mohist thought. Offers brief though incisive criticisms of most of the contemporary scholarship available in English.)

Mozi (*c*.470-*c*.391 BC) *Mozi*, trans. Mei Yi-pao, *The Ethical and Political Works of Mo Tzu*, London: Arthur Probsthain, 1929; trans. B. Watson, *Mo Tzu: Basic Writings*, New York: Columbia University Press, 1963.(Mei is the most complete translation of Mozi's works available in English. Watson is the most readable, selective English translation of the *Mozi*; it contains all of the synoptic chapters plus the two 'Against Confucians' chapters.)

Needham, J. (ed.) (1956) *Science and Civilisation in China*, vol. 2, repr. Taipei: Caves Books Ltd, 1985: 165-203. (A study of Mohist thought from the perspective of its contributions to science and logic.)

Nivison, D.S. (1980) 'Two Roots or One?', *Proceedings and Addresses of the American Philosophical Association* 53 (6): 739-61.(A philosophically sophisticated discussion of the ethical thought and theory of moral action of the later Mohist, Yi Zhi.)

Schwartz, B.I. (1985) *The World of Thought in Ancient China*, Cambridge, MA: The Belknap Press, 1985.(An incisive account of the Mohists from a sociological-historical perspective.)

Molecular biology

Molecular biology is the study of the structure, function and kinetics of biologically important molecules. Historically, molecular biology has often been identified with molecular genetics. Similarly, the chief philosophical concern with molecular biology has been the possibility of the reduction of classical genetics to molecular genetics. The nature and boundaries of molecular biology, however, are themselves disputed. To some, molecular biology seems to be a morass of molecular details without any overarching theory. To others molecular biology is an integrated interlevel theory. How philosophical issues, such as reduction, are addressed can depend importantly on how molecular biology is initially characterized.

1 Molecularizing biology

Although molecular biology can be broadly defined as the study of the structure, function and kinetics of biologically important molecules, the origins of molecular biology are usually traced back to two specific research traditions: the informational school and the structural school. This historical approach has been strongly disputed. Nevertheless, it provides a starting place for historical analysis.

The informational school is identified with the study of bacterial viruses (bacteriophage or phage) begun in the late 1930s by Max Delbruck, Salvador Luria and Alfred Hershey. The phage group, as they would come to be known, took as its goal the elucidation of the physical basis of heredity through careful experimental study of virus self-replication within bacterial hosts. The problem of replication was itself a means of investigating how molecules could store and transmit genetic information. Chromosomes were the accepted material basis of heredity and were known to be mixtures of nucleic acids and proteins. Proteins, by virtue of their known linear arrangement of specific units, were thought to be better candidates for information storage and transfer than were nucleic acids. This emphasis on proteins was overthrown by Alfred Hershey and Martha Chase’s experiments published in 1952 demonstrating that deoxyribonucleic acid (DNA) alone was injected into host bacteria. With the use of their Waring blender, Hershey and Chase had ingeniously shown that DNA, not protein, was responsible for virus replication.

At roughly the same time that the informational school was starting out in the 1930s, the structural school began extending traditions in structural chemistry to the study of biological molecules. In the USA, Linus Pauling used his expertise as a structural chemist to explain the helical structure of polypeptides essential to important biological proteins such as hemoglobin and myoglobin. In England, W.T. Astbury and J.D. Bernal began using X-ray crystallography to study the internal structure of proteins. This programme was expanded as the Cavendish Laboratory at Cambridge and King’s College in London began X-ray diffraction studies of biological molecules.

These two schools of thought came together in the 1950s with the collaboration of James Watson and Francis Crick. Watson’s knowledge of phage genetics and Crick’s knowledge of X-ray crystallography both contributed to their now famous discovery of the double helical structure of DNA in 1953. The two-stranded model of DNA with its complementary base pairs explained the structural features of the DNA molecule and suggested how DNA could both store and transfer information.

The neat division of the origins of molecular biology into these two schools has been disputed on historical as well as historiographical grounds. The structural and informational schools were introduced as historical categories in the 1960s as biologists struggled to define molecular biology. Gunther Stent, a member of the phage group, strongly identified the phage group with the origins of molecular biology. John Kendrew, who won the Nobel Prize in 1962 for determining the structure of myoglobin, had a completely different experience of the emergence of molecular biology. Kendrew introduced the idea that there were two schools. Gunter Stent agreed and extended Kendrew’s analysis in his reply. At the time Kendrew and Stent were writing, the emphasis in molecular biology was on integration of structure and function in biological molecules à la Watson and Crick. In effect Kendrew and Stent projected the reality of 1960s molecular biology into the past and in doing so created a history that legitimated present practice (Abir-Am 1985).

The structural and informational schools do not encompass the origins or the present state of molecular biology. It is not even clear to what extent the loose association of individuals doing structural work on biological molecules

can be considered a school. The origins of molecular biology as well as its current practice represent a much more diverse array of traditions.

Molecular biology, as it was first defined by the National Science Foundation in 1954, covered a wide variety of biological phenomena, such as the:

> identification and structure of particulate matter such as mitochondria, chloroplasts, chromosomes, viruses, enzyme structure and kinetics - chemistry of coenzymes, electrochemical phenomena, membranes and fibers, solid and liquid state phenomena, reactions of proteins - long range forces, mathematical approaches to biological problems
>
> (quoted in Zallen 1993: 81).

Clearly, molecular biology could be seen as drawing on long-standing traditions in biochemistry, cell biology and bioenergetics, as well as those in phage research and structural chemistry. The history of molecular biology may be dominated by the history of molecular genetics, but the practice of molecular biology is and has been much more diverse.

2 Interlevel theories

As philosophers tried to come to grips with molecular biology and particularly the question of the reduction of classical genetics to molecular genetics, they were faced with the problem of how to characterize the structure of molecular biology (see Genetics §4). While some thought of molecular biology as a scientific theory, characterized as a set of general laws or a set of axioms, others began to conceive of molecular biology in terms of interlevel theories and practices. These different approaches to theory structure can have important consequences for other philosophical issues, such as reduction.

Kenneth Schaffner has argued that most biomedical theories, including molecular biological theories, should be thought of as a series of overlapping interlevel temporal models (Schaffner 1993a, 1993b). This proposed analysis treats theories as families of models, as polytypic aggregates with some specified core characteristics. The entities represented by these models are usually undergoing some process and so are temporal. The theory is an interlevel theory because the entities represented by the theory can be grouped according to level of aggregation. Entities at one level of aggregation may share parts with entities at lower levels, but the defining properties of the entity at the higher level require organizing principles not found at the lower level.

It is often assumed that the successful reduction of a higher level theory to a lower level theory will result in the replacement of the higher level theory by the lower level theory. This strict reduction-replacement approach requires that the lower level theory use only lower level terms and entities. Schaffner’s interlevel theories require that this strict model of reduction-replacement be relaxed to allow partial or patchy reductions as a result of more complex connections between parts of different interlevel theories. With the complex interlevel processes considered in molecular biological theories, one should not expect complete unilevel reduction of a higher level theory to a lower level biochemical theory, which itself is often an interlevel theory. Molecular biological explanations, according to this view, are facilitated by partial reductions and causal generalizations regarding the temporal sequence of events represented by the relevant model (Schaffner 1993b).

3 Practices

Schaffner’s appreciation of the complexity of molecular biology is shared by Sylvia Culp and Philip Kitcher’s analysis of molecular biology in terms of practices. A practice consists of a language used by the scientific community of interest, the set of statements that community accepts, the set of questions they take to be important, the patterns of reasoning they use to answer those questions, the methodological directives or standards they use to evaluate solutions and experiments, and a set of experimental techniques (Culp and Kitcher 1989). Using this account of practices, Culp and Kitcher argue that the contemporary practice of molecular biology can be understood in terms of a hierarchy of questions concerning why some biologically important process occurs, and how accepted statements and experimental techniques contribute to the solution of that problem. The interlevel complexity described by Schaffner is captured in Culp and Kitcher’s practices by the hierarchy of questions. In the case of cell biology, for instance, Culp and Kitcher start out with the fundamental question of ‘how do organisms move?’. This question is then followed by the supposition that ‘motion requires contraction and extension of

muscles' and the related question of 'how do muscles contract?'. This question is then followed by the supposition that 'muscle cells contain actin and myosin' and the related question of 'how do actin and myosin contribute to the contraction of a cell?' (Culp and Kitcher 1989). This nested set of questions leads to a nested set of explanations. Cellular theories of motion do not reduce to molecular biology within this hierarchy; instead the explanations of the action of myosin and actin extend the cellular explanatory scheme. An explanatory scheme is a schematic argument with specific filling instructions.

What makes explanatory extension different from strict reduction-replacement is that it allows different parts of an explanatory scheme to be extended in different directions by different theories. Molecular biology does not provide the only route for explanatory extension and it is not the case that all explanatory extensions will ultimately end with molecular biology (Kitcher 1984). Schaffner's complex reduction-replacement model with its partially overlapping interlevel theories also allows theories and explanations to be elaborated in a number of ways, and agrees with much of Kitcher's explanatory extension approach, although Schaffner and Kitcher and Culp continue to disagree on how to characterize theories in molecular biology (Schaffner 1993a).

One further aspect of Culp and Kitcher's characterization of molecular biology deserves comment: namely, their emphasis on experiment. One of the most striking features of molecular biology is the prominence of experimental techniques (see Experiment). A vital aspect of the structure of molecular biology is the presence and exportation of sets of techniques to address a vast array of different biological problems. While philosophers of science have been traditionally concerned with theories, the ongoing study of sciences like molecular biology is fuelling the growing concern with experimentation, so much so that it now seems difficult to characterize molecular biology without discussing the central role of experimental techniques, such as X-ray diffraction, electrophoresis, or polymerase chain reactions (Zallen 1993).

See also: Reduction, problems of; Theories, scientific

MICHAEL R. DIETRICH

References and further reading

Abir-Am, P. (1985) 'Themes, Genres, and Orders of Legitimation in the Consolidation of New Scientific Disciplines: Deconstructing the Historiography of Molecular Biology', *History of Science* 23: 73-117. (Historiographical analysis of four major traditions in the history of molecular biology, including the history of the informational and structural schools.)

Cairns, J., Stent, G. and Watson, J. (eds) (1992) *Phage and the Origins of Molecular Biology*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.(Volume of papers in honour of Max Delbruck. The expanded edition includes Kendrew and Stent's papers contesting the informational and structural schools.)

Culp, S. and Kitcher, P. (1989) 'Theory Structure and Theory Change in Contemporary Molecular Biology', *British Journal of the Philosophy of Science* 40: 459-83.(An analysis of structure and change in molecular biology based on a case study of the discovery on enzymatic RNA.)

Judson, H.F. (1979) *The Eighth Day of Creation: Makers of the Revolution in Biology*, New York: Simon & Schuster.(A detailed history of molecular biology notable for its attention to the individual contributions and accomplishments of a diverse array of scientists.)

Kitcher, P. (1984) '1953 and All That: A Tale of Two Sciences', *Philosophical Review* 93: 335-73.(An antireductionist argument based on intertheoretic relationships and the concept of explanatory extension.)

Olby, R. (1974) *The Path to the Double Helix*, Seattle, WA: University of Washington Press.(A detailed reconstruction of the historical precursors to Watson and Crick's discovery of the structure of DNA. It further develops the idea of the structural and informational schools.)

Schaffner, K. (1993a) *Discovery and Explanation in Biology and Medicine*, Chicago, IL: University of Chicago Press.(A broad survey and analysis of a number of important issues such as reduction, theory testing and explanation.)

Schaffner, K. (1993b) 'Theory Structure, Reduction, and Disciplinary Integration in Biology', *Biology and Philosophy* 8: 319-47.(An analysis of theory structure in terms of integrated interlevel theories that allow for partial reduction and integration.)

Zallen, D. (1993) 'Redrawing the Boundaries of Molecular Biology: The Case of Photosynthesis', *Journal of the History of Biology* 26: 65-87.(A convincing argument for the reconception of molecular biology as a more diverse and complex field based on a case study of the history of bioenergetic research on the process of photosynthesis.)

Molina, Luis de (1535-1600)

A leading figure in sixteenth-century Iberian scholasticism, Molina was one of the most controversial thinkers in the history of Catholic thought. In keeping with the strongly libertarian account of human free choice that marked the early Jesuit theologians, Molina held that God's causal influence on free human acts does not by its intrinsic nature uniquely determine what those acts will be or whether they will be good or evil. Because of this, Molina asserted against his Dominican rivals that God's comprehensive providential plan for the created world and infallible foreknowledge of future contingents do not derive just from the combination of his antecedent 'natural' knowledge of metaphysically necessary truths and his 'free' knowledge of the causal influence - both natural (general concurrence) and supernatural (grace) - by which he wills to cooperate with free human acts. Rather, in addition to God's natural knowledge, Molina posited a distinct kind of antecedent divine knowledge, dubbed 'middle knowledge', by which God knows pre-volitionally, that is, prior to any free decree of his own will regarding contingent beings, how any possible rational creature would in fact freely choose to act in any possible circumstances in which it had the power to act freely. And on this basis Molina proceeded to forge his controversial reconciliation of free choice with the Catholic doctrines of grace, divine foreknowledge, providence and predestination.

In addition to his work in dogmatic theology, Molina was also an accomplished moral and political philosopher who wrote extensive and empirically well-informed tracts on political authority, slavery, war and economics.

1 Life and writings

Luis de Molina was a leading figure in the remarkable sixteenth-century revival of scholasticism on the Iberian peninsula that also produced the likes of Pedro da Fonseca, Domingo de Soto, Domingo Báñez and Francisco Suárez. After entering the Jesuit noviciate in 1553, Molina studied philosophy and theology for ten years at Coimbra and Évora in Portugal, and then taught at these same colleges until retiring in 1583 to devote himself to writing. He spent the next sixteen years in Évora, Lisbon and Cuenca, before being summoned just before his death to teach moral theology in Madrid.

The most famous and controversial of Molina's three published works was the *Liberi arbitrii cum gratiae donis, divina praescientia, providentia, praedestinatione et reprobatione concordia (A Reconciliation of Free Choice with the Gifts of Grace, Divine Foreknowledge, Providence, Predestination and Reprobation)* (first published in 1588). Popularly known simply as the *Concordia*, this work was in large part extracted from the *Commentaria in primam divi Thomae partem (Commentaries on the First Part of Aquinas' Summa theologiae)*, subsequently published at Cuenca in 1592. Molina also wrote a five-volume work on political philosophy, *De Justitia et Jure (On Justice and Law)*, the first complete edition of which appeared only posthumously. Although there are also modern editions of a few unpublished pieces, most of Molina's shorter tracts and commentaries survive only in manuscript.

The publication of the first edition of the *Concordia* ignited a fierce controversy about grace and human freedom that had already been smouldering for two decades between the youthful Society of Jesus (founded in 1540) and its theological opponents. At Louvain, the Jesuit Leonard Lessius had been assailed by the followers of Michael Baius for harbouring views on grace and freedom allegedly contrary to those of St Augustine. In Spain and Portugal, the Jesuits were accused of doctrinal novelty by theologians of the more established religious orders, especially the Dominicans, led by the redoubtable Báñez.

When the dispute began to jeopardize civil as well as ecclesiastical harmony, political and religious leaders in Iberia implored the Vatican to intervene. In 1597 Pope Clement VIII established the *Congregatio de Auxiliis* (Commission on Grace) in Rome, thus initiating a ten-year period of intense investigation - including eighty-five hearings and forty-seven debates - that rendered the *Concordia* one of the most carefully scrutinized books in Western intellectual history. At first, things did not go well for the Jesuits; Molina died in Madrid amid rumours that he was being burned in effigy in Rome. However, due to the efforts of Cardinals Robert Bellarmine and Jacques du Perron, Molina's views emerged unscathed in the end. In 1607 Pope Paul V issued a decree allowing both parties to defend their own positions but enjoining them not to call one another's views heretical.

2 Grace and freedom

Both sides of the dispute over grace and freedom agree that in addition to creating and conserving contingent beings, God acts as an immediate efficient cause of all the effects produced by created or secondary causes, including acts of free choice (see Grace; Free will).

First, God acts as an immediate 'general' or 'universal' cause of all the natural effects produced through the powers rooted in the essences of natural substances. This immediate causal contribution is called God's general concurrence (*concursus generalis*) because even though in any given case the effect proceeds as a whole from both God and the relevant secondary causes, the fact that the effect is of one species rather than another is primarily traceable not to God's concurrence but to the natures and causal contributions of the created agents, which for this reason are called 'particular causes'. So, for instance, when a gas flame makes a pot of water boil, the fact that the effect is the boiling of water rather than, say, the blossoming of a flower is traceable primarily not to God's causal contribution, but rather to the specific natures and causal contributions of the secondary causes (gas, water, and so on). Likewise, any defectiveness in the effect is traced back causally to a defect or impediment within the order of secondary causes rather than to God's causal contribution, which is always, within its own order, causally sufficient (in the sense of 'enough') for its intended effect, even when that effect is not produced. When the intended effect is produced, God's concurrence is said to be 'efficacious' with respect to it; when it is not produced, God's concurrence is said to be 'merely sufficient' with respect to it. If a defective effect is instead produced, its defectiveness is something that God merely permits rather than intends.

This account applies straightforwardly to morally good and evil acts emanating from the power of free choice that rational creatures are endowed with by nature. No such act can occur without God's general concurrence, and in causally contributing to it God intends that the act be morally upright rather than sinful. None the less, because of defects in the free agents with whom God cooperates, his general concurrence is often merely sufficient - and thus inefficacious - with respect to the morally good act he intends. So even though God must concur causally in order for even a sinful act to be elicited, none the less, the act's defectiveness is traced back to the free created agent rather than to God, who permits the defect without intending it (see Sin). By contrast, when a morally good act is freely elicited, God's concurrence is efficacious with respect to it.

Second, God also cooperates with free human acts by means of the particular causal influence of supernatural grace, merited for the human race by the salvific death and resurrection of Jesus Christ. By this grace, God empowers and prompts human beings to elicit free acts of will that are supernaturally salvific, and he cooperates as a simultaneous cause in the very effecting of these acts. By eliciting such free acts of faith, hope, charity and the other infused virtues, human agents are able to attain and foster that intimate friendship with God which in its fullness constitutes their highest fulfilment as rational beings. Still, in so far as grace operates prior to the consent of human free choice, it can be freely resisted; and when it is resisted, it is said to be inefficacious, or merely sufficient, with respect to the salvific act God intends (see Salvation §4).

All this is accepted by both Molina and his Báñezian opponents (see Báñez, D.). Their dispute has to do with the 'intrinsic' character of God's simultaneous causal cooperation with free human acts. Regarding the natural order, Báñezians insist that there is an intrinsic ontological difference between efficacious and merely sufficient concurrence, so that of itself efficacious concurrence necessarily attains its intended effect, whereas of itself merely sufficient concurrence necessarily fails to attain its intended effect. So God grants intrinsically efficacious concurrence when and only when the human agent freely elicits the morally good act that God intends; and God grants intrinsically merely sufficient concurrence when and only when the human agent freely fails to elicit the morally good act that God intends. An immediate consequence is that God infallibly foreknows whether or not a human agent will freely elicit a morally good act at a given time, simply by virtue of knowing whether or not he himself will grant intrinsically efficacious concurrence with respect to that act.

The Báñezians hold a parallel position concerning the grace by which God operates to elicit supernaturally salvific acts. God grants intrinsically efficacious grace when and only when the human agent elicits the supernaturally salvific act God intends; and God grants intrinsically merely sufficient grace when and only when the human agent resists and thus fails to elicit the supernaturally salvific act that God intends.

Molina argues that this doctrine is incompatible with human freedom and falls into the strict determinism

advocated by the Lutherans and Calvinists. For even though the Báñezians, like Molina, insist that a free act of will cannot result by natural necessity from antecedently acting causes, they none the less assert that an act of will can be free even if God has predetermined to cooperate with it contemporaneously by a concurrence or grace that is intrinsically efficacious (or inefficacious). This Molina denies: ‘That agent is called free who, with all the prerequisites for acting having been posited, is able to act and able not to act, or is able to do one thing in such a way that he is also able to do some contrary thing’ (*Concordia*: 14). Numbered among these prerequisites is God’s fixed intention to confer his general concurrence and grace. So if God has decided to confer only intrinsically efficacious (or intrinsically inefficacious) grace or concurrence in a given situation, the created agent’s freedom is destroyed.

Molina’s alternative thesis is that God’s grace and general concurrence are intrinsically neither efficacious nor inefficacious. Rather, they are intrinsically ‘neutral’ and are rendered efficacious or inefficacious ‘extrinsically’ by the human agent’s free consent or lack thereof. Báñezians retort that this position savours of Pelagianism and violates the Catholic doctrine that God is the primary source of morally good and supernaturally salvific acts (see Pelagianism).

3 Providence and middle knowledge

Molina’s views about grace and freedom seem at first glance to jeopardize the Catholic tenets that God is perfectly provident and has infallible and comprehensive knowledge of future contingents.

According to the traditional doctrine of divine providence (see Providence §1), God freely and knowingly plans, orders and provides for all the effects that constitute the created universe with its entire history, and he executes his chosen plan by playing an active causal role that ensures its exact realization (see Providence). Since God is the perfect craftsman, not even trivial details escape his providential decrees. Whatever occurs is specifically decreed by God; more precisely, each effect produced in the created universe is either specifically and knowingly intended by him or, in concession to the defectiveness of creatures, specifically and knowingly permitted by him. Divine providence thus has both a cognitive and a volitional aspect. By his pre-volitional knowledge God infallibly knows which effects would result, directly or indirectly, from any causal contribution he might choose to make to the created sphere. By his free will God chooses one from among the infinity of total sequences of created effects that are within his power to bring about and, concomitantly, wills to make a causal contribution that he knows with certainty will result in his chosen plan’s being effected down to the last detail (see Omnipotence; Omniscience).

This much is accepted by both Molina and the Báñezians. They further agree that it is because he is perfectly provident that God has comprehensive foreknowledge of what will occur in the created world. That is, God’s speculative post-volitional knowledge of the created world - his so-called ‘free knowledge’ or ‘knowledge of vision’ - derives wholly from his pre-volitional knowledge and his knowledge of what he himself has willed to do. Unlike human knowers, God need not be acted upon by outside causes in order for his cognitive potentialities to be fully actualized; he does not have to, as it were, look outside himself in order to find out what his creative act has wrought. Rather, he knows ‘in himself’ what will happen precisely because he knows just what causal role he has freely chosen to play within the created order and just what will result given this causal contribution. In short, no contingent truth grasped by the knowledge of vision can be true prior to God’s specifically intending or permitting it to be true or to his specifically willing to make the appropriate causal contribution towards its truth.

Molina’s problem can now be stated succinctly. As noted above, he affirms against the Báñezians that God’s cooperating grace and general concurrence are intrinsically neither efficacious nor inefficacious. So if God has only his antecedent ‘natural knowledge’ of metaphysically necessary truths plus knowledge of his own total causal contribution to the created world, he cannot specifically provide for or know any actual or ‘absolute’ future contingents. His natural knowledge tells him only what each free secondary cause would be able to do, not what it would in fact do, in any possible situation in which it could act; and if his grace and general concurrence are intrinsically neutral, then his own causal contribution to the contingent effects of secondary causes cannot uniquely determine what those effects will be. How, then, can God be perfectly provident and have comprehensive knowledge of future contingents?

Molina replies that since there is genuine freedom and indeterminism in the created world, God can be perfectly

provident in the way demanded by orthodoxy only if his pre-volitional knowledge includes comprehensive knowledge of which effects would in fact result from causal chains involving free or indeterministic created causes. That is, God must have infallible and comprehensive pre-volitional knowledge of ‘conditional future contingents’ or *futuribilia* - metaphysically contingent propositions specifying how any possible indeterministic created cause would in fact act in any set of circumstances it might find itself in. Since this knowledge is of metaphysically contingent truths, it is not part of God’s natural knowledge; since it is pre-volitional, it is not part of God’s free knowledge. It stands ‘midway’ between natural knowledge and free knowledge - hence its title, ‘middle knowledge’ (*scientia media*).

According to Molina, then, the basis for God’s providence and his foreknowledge of absolute future contingents is threefold: (1) his pre-volitional natural knowledge of metaphysically necessary truths, (2) his pre-volitional middle knowledge of *futuribilia*, and (3) his post-volitional knowledge of the total causal contribution he himself wills to make to the created world. By (1) he knows which spatio-temporal arrangements of secondary causes are possible and which contingent effects might possibly emanate from any such arrangement. By (2) he knows which contingent effects would in fact emanate from any such arrangement. By (3) he knows which secondary causes he wills to create and precisely how he wills to cooperate with them via his intrinsically neutral cooperating grace and general concurrence. So given God’s pre-volitional natural knowledge and middle knowledge, he is able to choose a comprehensive providential plan; and given further his post-volitional knowledge of what his own causal contribution to the created world will be, he has free knowledge of all absolute future contingents.

The Bañezians counter by denying that any metaphysically contingent propositions, including any *futuribilia*, can be true prior to God’s freely making them true by his predetermining decree. This dissatisfaction with middle knowledge provides another reason for Bañezians to reject Molina’s strongly libertarian account of freedom, which engendered the need for middle knowledge in the first place. (Interestingly, by way of contrast, the main opponents of middle knowledge in twentieth-century analytic philosophy of religion retain Molina’s account of freedom and reject instead the traditional doctrines of divine providence and foreknowledge.)

4 Political philosophy

Molina holds a ‘translation’ account of political authority (see Political philosophy, history of). Unlike the ecclesiastical authority of the pope, which is bestowed immediately by God on the person designated by the ecclesiastical community, the governing authority of the political ruler is bestowed immediately by a community which has established itself as a commonwealth and in which political authority resides by natural law. Hence, unlike the ecclesiastical community, the political commonwealth not only designates its rulers but also legitimately limits the power it transfers to them and restricts their use of that power. Because of this, the individual members of the commonwealth can, under certain extreme circumstances, legitimately resist and defend themselves against tyrannical rulers.

Molina also addresses at length a wide variety of moral questions regarding war, slavery and economic matters such as taxation, free markets, monetary policy and price regulation. Given his historical circumstances, Molina’s ruminations about slavery are especially interesting. Like his Aristotelian predecessors, he believes that slavery is morally justifiable under certain limited circumstances. For example, those lawfully condemned to death may legitimately have their death sentences commuted to perpetual servitude; enemy populations conquered in a just war may legitimately be enslaved in restitution for damages to the victors; and free and willing adults may legitimately sell themselves into slavery. None the less, Molina’s many conversations in Lisbon with the captains of slave ships led him to conclude that the African slave trade as it was actually being carried out by the Portuguese was ‘unjust and wicked’ and that those who engaged in it, both sellers and buyers, were almost surely ‘in a state of eternal damnation’ (see Slavery).

See also: Báñez, D.; Freedom, divine; God, concepts of; Grace; Molinism; Omniscience; Predestination; Suárez, F. §1

ALFRED J. FREDDOSO

List of works

Molina, L. de (1592) *Commentaria in primam divi Thomae partem (Commentaries on the First Part of Aquinas’ Summa theologiae)*, Cuenca.(Provided the basis for Molina’s more famous work, the *Concordia*.)

Molina, L. de (1588) *Liberi arbitrii cum gratiae donis, divina praescientia, providentia, praedestinatione et reprobatione concordia (A Reconciliation of Free Choice with the Gifts of Grace, Divine Foreknowledge, Providence, Predestination and Reprobation)*, Lisbon, 1st edn; 2nd edn, Antwerp; ed. J. Rabeneck, Oña and Madrid: Collegium Maximum, 1953; trans. A.J. Freddoso, *On Divine Foreknowledge: Part IV of the 'Concordia'*, Ithaca, NY: Cornell University Press, 1988.(Translation of and introduction to Molina's theory of middle knowledge, including references to twentieth-century discussions.)

Molina, L. de (1614) *De Justitia et Jure (On Justice and Law)*, Venice.(The first complete edition of this work appeared posthumously in 1614.)

Molina, L. de (1935) *'Neue Molinaschriften' (New Writings by Molina)*, ed. F. Stegmüller, *Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters*, vol. 32, Münster: Aschendorff.(A collection of unpublished treatises and letters.)

References and further reading

Adams, R.M. (1977) 'Middle Knowledge and the Problem of Evil', *American Philosophical Quarterly* 14: 109-17.(Introduced the debate over middle knowledge into twentieth-century analytic philosophy of religion.)

Brodrick, J. (1928) *The Life and Work of Blessed Robert Francis Cardinal Bellarmine, S.J.*, vol. 2, New York: Kenedy.(Chapter 19 contains an extensive historical account of the sixteenth-century dispute over grace and freedom.)

Costello, F. (1974) *The Political Philosophy of Luis de Molina, S.J.*, Spokane, WA: Gonzaga.(Helpful and clear introduction to Molina's political thought.)

Flint, T. (1988) 'Two Accounts of Providence', in T. Morris (ed.) *Divine and Human Action*, Ithaca, NY: Cornell University Press, 147-81.(Clear exposition by a Molinist of some key differences between Molinists and Thomists.)

Garrigou-Lagrange, R. (1943) *The One God*, St Louis, MO: Herder.(Contains a trenchant neo-Thomist critique of Molinism.)

Hasker, W. (1989) *God, Time, and Knowledge*, Ithaca, NY: Cornell University Press. (Contains the best critique of middle knowledge in analytic philosophy of religion.)

Pegis, A. (1939) 'Molina and Human Liberty', in G. Smith (ed.) *Jesuit Thinkers of the Renaissance*, Milwaukee, WI: Marquette, 75-131.(Comparison of Molina with Aquinas.)

Molinism

Molinism, named after Luis de Molina, is a theological system for reconciling human freedom with God's grace and providence. Presupposing a strongly libertarian account of freedom, Molinists assert against their rivals that the grace whereby God cooperates with supernaturally salvific acts is not intrinsically efficacious. To preserve divine providence and foreknowledge, they then posit 'middle knowledge', through which God knows, prior to his own free decrees, how any possible rational agent would freely act in any possible situation. Beyond this, they differ among themselves regarding the ground for middle knowledge and the doctrines of efficacious grace and predestination.

Molinism is an influential system within Catholic theology for reconciling human free choice with God's grace, providence, foreknowledge and predestination. Originating within the Society of Jesus (the Jesuits) in the late sixteenth and early seventeenth centuries, it encountered stiff opposition from Báñezian Thomists (see Báñez, D.) and from the self-styled Augustinian disciples of Michael Baius and Cornelius Jansen.

Molinism's three distinguishing marks are a strongly libertarian account of human freedom (see Freedom and liberty §1); the consequent conviction that the grace whereby God cooperates with supernaturally salvific free acts is not intrinsically efficacious (see Grace); and the postulation of divine middle knowledge (*scientia media*), by which God knows, before any of his free decrees regarding creatures, how any possible rational being would freely act in any possible situation (see Molina, L. de §§2-3; Omniscience).

Beyond this, Molinists disagree about three important issues. The first is the question of how God knows the 'conditional future contingents' or *futuribilia* that constitute the objects of middle knowledge. Molinists cannot accept the Báñezian claim that God knows *futuribilia* by virtue of his freely decreeing their truth, since according to Molinism *futuribilia* have their truth prior to any free divine decree. Nor can Molinists claim that God knows *futuribilia* simply by virtue of comprehending all possible creatures, if 'comprehending' a creature means just understanding all the metaphysical possibilities involving it. For such comprehension is insufficient for knowing how a possible creature would freely act - as opposed to how it could act - in any possible situation.

Molina himself claims that because God's cognitive power infinitely surpasses the natures of creatures, God is able to know those natures 'in a more eminent way than that in which they are knowable in themselves' (*Concordia*: 343). So God not only comprehends possible creatures but also 'super-comprehends' them, as later Molinists put it, and in this way knows *futuribilia* involving them. One corollary, explicitly defended by Molina, is that God does not know *futuribilia* concerning his own free decrees, since his cognitive power does not infinitely surpass his own nature.

Other Molinists retort that no amount of insight into the natures of possible creatures can yield infallible knowledge of *futuribilia*, since such natures are exhausted by their metaphysical possibilities and do not include *futuribilia*. Instead, God has direct knowledge of *futuribilia*, unmediated by his knowledge of natures, and this is simply because the *futuribilia* are true and hence intelligible to an infinite intellect. From this viewpoint there is no reason why God should not know *futuribilia* concerning his own free decrees - a result Molina takes to be incompatible with God's freedom.

The second dispute concerns the reason for the efficaciousness of the grace whereby God cooperates with supernaturally salvific acts of free choice. Suppose that in circumstances *C*, influenced by grace *G*, Peter freely elicits salvific act *A*. All Molinists agree that God places Peter in *C* with *G* knowing full well that Peter will freely elicit *A*; and they also agree that *G* is not intrinsically efficacious and hence does not causally predetermine *A*. However, there is strong disagreement about whether or not it is Peter's free consent alone that 'extrinsically' renders *G* efficacious in *C* with respect to *A*.

One possibility is that God first resolves absolutely that Peter should freely elicit *A* in *C* and then, as it were, consults his middle knowledge to see just which particular graces would, if bestowed on Peter in *C*, obtain his free consent and thus issue in *A*. It follows that, given his antecedent resolution, God would have conferred some grace other than *G* if he had known by his middle knowledge that *G* would turn out to be 'merely sufficient' with respect to *A*, that is, that Peter would not freely consent to *G* in *C*. So *G* is rendered efficacious not only by Peter's free

consent but also, and indeed more importantly, by God's antecedent predetermination to confer a 'congruous' grace that will guarantee Peter's acting well in *C*. This model, which brings Molinism more into line with Báñezianism, is known as Congruism and was worked out in detail by Robert Bellarmine and Francisco Suárez (§1). In 1613 Congruism was mandated for all Jesuit theologians by the Father General Claude Aquiviva.

Another possible model, which seems to be Molina's own, is that God simply wills to put Peter into *C* with *G*, knowing that Peter will freely elicit *A* but not having absolutely resolved beforehand that Peter should freely elicit *A*. This model does not entail that if God had known that Peter would act badly in *C* with *G*, he would have conferred some grace other than *G* in order to guarantee *A*. Accordingly, it is Peter's free consent alone that renders *G* efficacious.

An analogous dispute arises over predestination, where the question concerns not one or another of Peter's acts, but his eternal salvation. Some Molinists, including Bellarmine and Suárez, agree with the Bañezians that God antecedently elects certain people to eternal glory and only then consults his middle knowledge to discover which graces will guarantee their salvation. Thus, in Peter's case, God would have chosen different graces if those he actually chose had been foreknown to be merely sufficient and not efficacious for Peter's salvation. Other Molinists, including Molina himself, vigorously reject any such antecedent absolute election of Peter to salvation. They insist instead that God simply chooses to create a world in which he infallibly foresees Peter's good use of the supernatural graces afforded him, and only then does he accept Peter among the elect in light of his free consent to those graces.

See also: Báñez, D.; Freedom, divine; Molina, L. de; Predestination; Providence; Suárez, F. §1

ALFRED J. FREDDOSO

References and further reading

Bellarmine, R. (1873) *Opera Omnia (Complete Works)*, vol. 5, ed. J. Fèvre, Paris: Louis Vivès. (Contains the tract *De Gratia et Libero Arbitrio*, one of the main sources for Congruism.)

Garrigou-Lagrange, R. (1952) *Grace*, St Louis, MO: Herder.(Chapters 7 and 8 contain a Báñezian assessment of the Molinist and Congruist accounts of efficacious grace.)

Molina, L. (1953) *Liberi arbitrii cum gratiae donis, divina praescientia, providentia, praedestinatione et reprobatione concordia (A Reconciliation of Free Choice with the Gifts of Grace, Divine Foreknowledge, Providence, Predestination and Reprobation)*, ed.J. Rabeneck, Oña and Madrid: Collegium Maximum; trans. A.J. Freddoso, *On Divine Foreknowledge: Part IV of the 'Concordia'*, Ithaca, NY: Cornell University Press, 1988. (Translation of, and introduction to, Molina's theory of middle knowledge.)

Pohle, J. (1947) *Grace: Actual and Habitual*, ed. A. Preuss, St Louis, MO: Herder. (Gives a weaker characterization of Congruism than that laid out above in order to classify Molina himself as a Congruist.)

Suárez, F. (1963) *Opera Omnia (Complete Works)*, vols 7-11, ed. C. Berton, Brussels: Culture et Civilisation, 1963.(Suárez's voluminous treatise *De Gratia*, containing the most sophisticated explication and defence of Congruism.)

Molyneux problem

The origin of what is known as the Molyneux problem lies in the following question posed by William Molyneux to John Locke: if a man born blind, and able to distinguish by touch between a cube and a globe, were made to see, could he now tell by sight which was the cube and which the globe, before he touched them? The problem raises fundamental issues in epistemology and the philosophy of mind, and was widely discussed after Locke included it in the second edition of his Essay concerning Human Understanding.

1 The origin of the problem

William Molyneux (1656-98) was a leading Irish scientist whose wife became blind shortly after marriage. His consequent interest in vision gave rise to *Dioptrica Nova (New optics)* (1692), a treatise on optics which also touches on the psychology of vision. The main problems for seventeenth-century theories of vision arose from the apparent conflict between physical theories about light and the phenomenology of perception. For example, optics demonstrates the inversion of the retinal image, but we perceive the object the right way up. Moreover, the retinal image is flat, but we perceive distance and depth. In response to these problems, Descartes postulated an innate judgment by which the soul corrects the inverted retinal image, and he ascribed the perception of distance partly to 'natural geometry'. We unconsciously calculate the distance of objects from the angles of the incoming light rays and the distance between our eyes.

Molyneux's treatment of the perception of distance or depth falls into two parts. Perception of the distance of far-off objects, he agrees, involves an act of judgment: 'For Distance of itself, is not to be perceived, for 'tis a Line (or a Length) presented to the Eye with its End towards us, which must therefore be only a Point, and that is Invisible' (Molyneux 1692: 113). But perception of near-distance is assigned, somewhat vaguely, to unconscious natural capacities. Near objects are perceived 'by the turn of the eyes, or by the angle of the optic axes' explanations going back, via Descartes and Kepler, to medieval theories of perception.

While writing *Dioptrica Nova* Molyneux read the French abstract, published in 1688, of Locke's *Essay concerning Human Understanding*, and sent the author the following problem:

> A Man, being born blind, and having a Globe and a Cube, nigh of the same bigness, committed into his Hands, and being taught or Told, which is Called the Globe, and which the Cube, so as easily to distinguish them by his Touch or Feeling; Then both being taken from Him, and Laid on a Table, let us suppose his Sight Restored to Him; Whether he Could, by his sight, and before he touch them, know which is the Globe and which the Cube? Or Whether he could know by his sight, before he stretched out his Hand, whether he Could not Reach them, tho they were Removed 20 or 1000 feet from him.
>
> (Letter to Locke, 7 July 1688)

Despite receiving no reply, Molyneux dedicated his book to the 'incomparable Mr. Locke' and sent him a copy. In the ensuing correspondence Molyneux again posed the problem. This time Locke judged it an 'ingenious problem', including it in the second edition of his *Essay* (1694).

2 Philosophical responses

Philosophers' answers to Molyneux's problem and their conception of its significance reflect their general epistemology. Both Locke's empiricism and his requirement that any idea is such as the subject perceives it to be rule out the possibility of the sort of innately guided, unconscious reasoning postulated by Descartes. The perception of distance is therefore an acquired ability and the answer to Molyneux's question is negative. This reply may seem inconsistent with Locke's doctrine that ideas of primary qualities, such as shape or motion, resemble their causes: visual and tactual ideas of such qualities have common causes, which implies that they resemble each other. However, Locke saw Molyneux's problem simply as showing that 'the Ideas we receive by sensation are often in grown People alter'd by the Judgement, without our taking notice of it' (1689: ch. II, §8). The ideas we receive by sight are allegedly only two-dimensional patterns of light and colour. In order to be able to recognize a three-dimensional shape or body in these patterns, we need to know how to correlate them with information provided by touch. The visual idea is like a word which induces in a hearer the associated idea, but is

itself hardly noticed (see Locke, J. §§2, 4).

Leibniz (1765) accepted that the connection between the visual and the tactual sensations is not innate. However, if the blind person has been given the information that a cube and a globe lie in front of him, they can work out which is which from the geometric features the visual and the tactual ideas share. In the cube there are eight points distinguished from all the others, whereas in the globe there are none. This shows, Leibniz concludes, how important it is to distinguish sensory images from ‘exact ideas constituted by definitions’ (see Leibniz, G.W. §8). Another of Locke’s contemporaries, Edward Synge (1695), insisted on a somewhat similar distinction between image and sensation on the one hand, and idea and perception on the other. He argued that, although the images acquired by sight and touch are different, the ideas are identical. In recognizing a cube by sight and by touch only one conceptual ability is involved.

Berkeley took a contrary view, rejecting Locke’s account of primary qualities and arguing that ideas of sight and touch are radically heterogeneous, connected only by contingent correlations known through sense experience. His reasons for endorsing the negative answer to the problem, and for his enthusiastic development of the analogy with language, are thus far more radical than Locke’s (see Berkeley, G. §4).

The Molyneux problem was widely discussed in the eighteenth and nineteenth centuries (indeed Ernst Cassirer has claimed that it was the fundamental question for both continental and Anglo-saxon epistemology and psychology in the eighteenth century). Among the contributors were Condillac and Diderot whose *Lettre sur les aveugles (Letter on the blind)* contains a fascinating description of the difference between the ‘life-world’ of the blind and that of the sighted. Various attempts have been made to answer the question empirically, the first being ‘Cheselden’s case’, named after the surgeon who, in 1728, reported an operation performed on a boy with congenital cataract. Since it took some time for the boy to perceive and recognize objects by sight, Cheselden’s case was supposed to confirm the negative reply. Yet even Leibniz had foreseen that the subject might be ‘dazzled and confused by the strangeness’. In any case, since blindness is a symptom resulting from different diseases of the eyes and along the optical tract, it is difficult to assess what bearing empirical data have on the Molyneux problem.

On any interpretation, the Molyneux problem raises the issue of the relation between the perceptual representation of space attributable to the blind and that made available by sight. Berkeley’s radical assertion of heterogeneity assigned primary spatial experience to touch, treating visual ideas as mere signs of impending or available tactual ideas. Conversely, it might be questioned whether the blind possess genuine spatial concepts, on the grounds that they only have successions of tactual experiences. Possession of spatial concepts, however, involves thought of distinct objects existing not in succession but simultaneously. Yet both these views ignore the fact that there is only one behavioural space for a subject to move around in. Simply to act, a subject - whether blind or sighted - must have experience of egocentric space involving an ability to locate what is perceived, however perceived, in relation to their body. If that is correct, the chief question remaining is how readily a blind subject, made to see, is able to generalize spatial concepts to visual experiences.

MENNO LIEVERS

References and further reading

Ayers, M.R. (1991) *Locke*, London: Routledge, 2 vols; repr. as *Locke. Epistemology and Ontology*, 1993.(Locke’s response to the Molyneux problem is discussed in the context of an extensive examination of Locke’s notion of an idea.)

Berkeley, G. (1709) ‘An Essay towards a New Theory of Vision’, in *Philosophical Works*, intro. and notes M.R. Ayers, London, J.M. Dent & Sons Ltd, Everyman’s Library, 1989: 3-59.(Berkeley’s first work in which he expounds his theory of vision, which is the foundation for the idealist philosophy he developed in his later works. In this work he discusses and refers to Molyneux’s *Dioptrica Nova*.)

Brandt Bolton, M. (1994) ‘The Real Molyneux Question and the Basis of Locke’s Answer’, in G.A.J. Rogers (ed.) *Locke’s Philosophy: Content and Context*, Oxford: Oxford University Press, 75-99.(Establishes the importance of situating the Molyneux problem within the context of seventeenth-century theories of vision.)

Degenaar, M. (1996) *The Problem of Molyneux*, Dordrecht: Reidel.(A discussion of eighteenth- and nineteenth-century - and more recent - responses to the Molyneux problem, with emphasis on empirical attempts to settle the question; extensive bibliography.)

Diderot, D. (1749) *Lettre sur les aveugles*; trans. as *The Letter on the Blind for the Benefit of Those Who See*, Albuquerque, NM: American Classical Collection Press, 1983.(A fascinating work arguing that being blind affects not only a person's sight but also their metaphysical, religious and moral convictions.)

Evans, G. (1985) 'Molyneux's Question', in *Collected Papers*, Oxford: Oxford University Press, 364-99.(Argues convincingly that the heart of the Molyneux problem is whether the spatial concepts possessed by the blind are identical to those possessed by the sighted.)

Leibniz, G. W. (1765) *New Essays on Human Understanding*, trans. and ed. P. Remnant and J. Bennett, Cambridge: Cambridge University Press, 1981.(A commentary on Locke's *Essay*, unpublished in Leibniz's lifetime, which contains an excellent treatment of the Molyneux problem on pages 136-8).

Locke, J. (1688) 'Essai philosophique concernant l'Entendement où l'on montre quelle est l'étendue des connaissances certaines, et la manière dont nous y parvenons', *Bibliothèque Universelle et Historique*, January-March: 49-142, Amsterdam, ed. J. Le Clerc.(Locke's first major publication in the journal of his Swiss friend Jean Le Clerc, whom he met in The Netherlands. It is a substantial abridgement of his as yet unpublished *Essay concerning Human Understanding* and often referred to with the term Abrégé, because it was later republished separately and translated into English in 1692 as an extract of a book entitled *A Philosophical Essay upon Human Understanding.*)

Locke, J. (1689) *Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1985. (Locke's magnum opus, breathtaking in its scope and depth, still worth consulting on every topic it discusses. The second edition appeared in 1694.)

Molyneux, W. (1688) Letter to Locke of 7 July 1688, in E.S. de Beer (ed.) *The Correspondence of John Locke*, Oxford: Clarendon Press, 1978, vol. 3, 482, letter no. 1064.

Molyneux, W. (1692) *Dioptrica Nova*, London: Benjamin Tooke.(A treatise in the tradition of similar works by Kepler and Descartes, which is mainly devoted to the physical aspects of optics, containing also several remarks on the psychology of vision.)

Synge, E. (1695) *The Correspondence of John Locke*, vol. 5, ed. E.S. de Beer, Oxford: Oxford University Press, 1974.(Letter 1984 (pages 494-6) from Molyneux to Locke, quotes in full the interesting letter by Synge - later Archbishop of Tuam - which anticipates several later objections to Locke's reply.)

Momentariness, Buddhist doctrine of

*The object of the Buddhist doctrine of momentariness is not the nature of time, but existence within time. Rather than atomizing time into moments, it atomizes phenomena temporally by dissecting them into a succession of discrete momentary entities. Its fundamental proposition is that everything passes out of existence as soon as it has originated and in this sense is momentary. As an entity vanishes, it gives rise to a new entity of almost the same nature which originates immediately afterwards. Thus, there is an uninterrupted flow of causally connected momentary entities of nearly the same nature, the so-called continuum (*santāna*). These entities succeed each other so fast that the process cannot be discerned by ordinary perception. Because earlier and later entities within one continuum are almost exactly alike, we come to conceive of something as a temporally extended entity even though the fact that it is in truth nothing but a series of causally connected momentary entities. According to this doctrine, the world (including the sentient beings inhabiting it) is at every moment distinct from the world in the previous or next moment. It is, however, linked to the past and future by the law of causality in so far as a phenomenon usually engenders a phenomenon of its kind when it perishes, so that the world originating in the next moment reflects the world in the preceding moment.*

At the root of Buddhism lies the (never questioned) conviction that everything that has originated is bound to perish and is therefore, with the exception of factors conducive to enlightenment, ultimately a source of frustration. There is no surviving textual material that documents how this law of impermanence came to be radicalized in terms of momentariness. It seems that by the fourth century the doctrine of momentariness had already assumed its final form. Characteristically, the debate became more and more dominated by epistemological questions, while the metaphysical aspect faded into the background.

1 Exposition

The doctrine of momentariness entails that entities are too shortlived to undergo change. Thus, if an entity has always engendered a new entity of exactly the same kind and with exactly the same properties, the worlds arising at every moment anew would be identical, so that there could be no evolution. This, however, is not the case because the process of reproduction of a given entity may be manipulated by outside factors in such a way that the newly created entity differs qualitatively from the preceding entity. If exposed to fire, for instance, a wood entity does not give rise to an identical wood entity when it perishes, but to a wood entity which bears the mark of impairment by fire and so is slightly charred. (According to later parlance, the wood as the main cause forms, together with the fire as a subsidiary cause, a causal complex which produces the slightly charred wood entity.) Thus, change is not constituted by the transmutation of persisting entities, but by the qualitative difference between earlier and later entities within a series.

Not only the transformation of series but also their cessation (that is, what is ordinarily conceived of as the utter annihilation of temporally extended objects), is caused by an external agent, which affects the process of reproduction of the object exposed to it in such a way that this process comes to a complete standstill. Hence, in the case of murder, the victim dies because the murderer affects the final moment of the breath of life (*prāṇa*), that is, the vital principle accounting for the body's animation, in such a way that it fails to reproduce itself. Since the final moment (like all preceding ones) passes out of existence automatically, murder is, microscopically speaking, not destruction but the interception of the process of reproduction. In this way the teaching that all entities pass out of existence spontaneously without depending for this upon any external cause is reconciled with the observation on a macroscopic level that wood is burnt by fire, or that one dies when knifed by a murderer.

Independently of the doctrine of momentariness, the Buddhists, like many other Indian schools, also dissected everything spatially into atoms (see Matter, Indian conceptions of). Thus, in the final analysis, the world is made up of momentary atoms, which by their spatial arrangement and by their concatenation with earlier and later atoms of the same kind, give rise to the illusion of persisting compact things. This analysis of existence can be illustrated by referring, anachronistically, to cinematography. Just as the rapid projection of distinct pictures evokes the illusion of continuous action on the screen, so the fast succession of distinct momentary entities gives rise to the erroneous impression that the world around us (and we ourselves) exist continuously without undergoing destruction and being recreated every moment. Similarly, as the change of events on the screen is caused by the

qualitative difference between earlier and later pictures on the film reel, so the change in the world is brought about by the qualitative difference between earlier and later entities. Moreover, as people vanish from the screen because they are not featured in the subsequent frame, so things cease to exist because they stop reproducing themselves. Finally, just as each projected picture only consists of differently shaded points, which by their specific arrangement give rise to the perception of composite shapes, so the world around us consists of nothing but distinct atoms which are arranged in such a way that they convey the impression of compact bodies.

2 Relevance

The Buddhist doctrine of momentariness does not challenge our experiences of macroscopic events as such, but only our interpretation of these events on a microscopic level. The claim that macroscopic objects are constituted by a succession of distinct momentary units only affects the intuitive conception of these objects as self-identical units (think of the notion of an uninterrupted line in contrast to one made up of distinct but contiguous points), but it does not affect the question of how these macroscopic objects behave, whatever their analysis on a microscopic level.

The doctrine of momentariness was not viewed as a purely metaphysical theory without practical relevance. On the contrary, the contemplation of the constant rise and fall of phenomena was employed to induce a particularly poignant experience of their impermanence, thus revealing the unsatisfactory nature of all existence. Moreover, in a Mahāyāna context this contemplation served as a tool for undermining (but not negating) the substantial existence of phenomena. However, since only advanced yogins seem to have been able to perceive momentariness directly, the soteriological significance of this doctrine remains very limited. This explains why it only played a marginal role in the wider context of Buddhist spirituality.

3 Development

The doctrine of momentariness is postcanonic and may have originated in the first century. It is for the first time presupposed in the *Vibhāṣa* (both in the Chinese translation by Xuanzang (Hsüan-tsang) and by Buddhavarman) of the Sarvāstivādins, one of the major Hīnayāna schools of Buddhism. In this scholastic compendium, the better part of which was probably compiled in the second century, the momentariness of all phenomena is not treated as a topic in its own right, but is frequently taken for granted when dealing with other issues.

Doxographical reports and other evidence confirm the impression that it was in the milieu of the Sarvāstivādins that all phenomena, more precisely all conditioned entities (*saṃskṛta*, *saṃskāra*), came to be looked upon as momentary. (The Sarvāstivādins treated space and two forms of suppression of certain factors as unconditioned entities (*asaṃskṛta*) which have never been created and, hence, are not subject to the law of impermanence, hence the specification at this point that momentariness only applies to conditioned entities and not to all phenomena.) Although the Sarvāstivādins reduced the duration of all phenomena to a moment, they still conceived of their existence much in the same way as they had done before the introduction of the doctrine, insisting that even within one moment they first originate, then persist and decay and finally perish. This treatment violated the common conception of the moment (*kṣaṇa*) as the shortest conceivable unit of time and consequently was rejected by the Dārṣṭamacr;(ntikas and Sautrāntikas, who are closely related to the Sarvāstivādins and may have evolved from them. These two schools argued that contradictory events cannot take place within one moment. From this they concluded that all things perish as soon as they have originated. Since destruction was conceived of as the spontaneous cessation of existence and not as a time-consuming process, the existence of entities was reduced to mere acts of origination, flashes into existence.

With this radicalization of the instantaneous nature of existence, the doctrine of momentariness assumed its final form. Such a form was adopted by the Yogācāras, one of the two main Indian schools of Mahāyāna Buddhism, and came to be known by other Buddhists and non-Buddhists. The Yogācāras with a Mahāyāna orientation, however, only accepted the doctrine of momentariness as valid on the level of relative truth. In so far as the doctrine affirms the existence of discrete entities (although they are reduced to mere point instants), it is characteristic of the realism of Hīnayāna Buddhism and at odds with their Mahāyāna stance that all phenomenal entities are ultimately, on the level of highest truth, unreal (see Buddhism, Yogācāra school of).

4 Doctrinal background

The surviving sources do not record how the doctrine of momentariness originated. Hence the reconstruction of this process has to be hypothetical. The anti-substantialist tendency characteristic of Buddhism denies that entities have a substantial core beyond the sum of their properties and thus equates the properties with the entities themselves. Hence, phenomena in Buddhism are called *dharmas*, a term with a wide range of meanings which is used in this context because it may stand for ‘property’ and ‘quality’. Since change was viewed as the replacement of one quality for another, the identification of property and entity led to the position that any qualitative change implies numeric difference, that is, the substitution of one entity for another. When Buddhists applied this understanding of change to their analysis of ageing they were bold enough to conclude that the ageing body must at every moment vanish to be replaced by a new, slightly modified body. As all things were conceived of as constantly changing, momentariness had to be attributed in this way not only to bodily matter but also to all other things. The conviction that everything is always changing (in as much as it is always subject to ageing) had resulted from the contemplation of the law of impermanence.

The discovery of the doctrine of momentariness in this way was possible because at a much earlier stage the momentariness of all mental entities had already been established in an apparently analogous way. This way was the denial of a permanent Self, a cardinal tenet of scholastic Buddhism which led to the conception of the mind as a flow of mental events conceived of as entities in their own right (see Buddhist concept of emptiness). Their momentariness was probably deduced from the speed with which mental events normally follow each other. The establishment of the doctrine of momentariness may have benefited from the testimony of yogins who are reported to have access to the direct experience of the incessant rise and fall of phenomena at every moment.

5 Proofs

Such a doctrine, fundamentally at odds with the appearance of the world, met great opposition. Initially, it was rejected by large sections of the Buddhist community, notably the Vātsīputrīyas and related schools. Later, when it had gained ground among Buddhists, it was fervently opposed by the Brahmanical schools as it contradicted their postulation of eternal entities of one sort or another (souls, atoms, primary matter, a supreme deity). This rejection made it necessary to defend the doctrine by argumentation.

The oldest transmitted proofs of momentariness are recorded in early Yogācāra sources. They are still primarily directed against other Buddhists and derive the momentariness of all phenomena in three different ways. First, it is presupposed that the mind is momentary - this stance is also shared by Buddhist opponents who do not accept the momentariness of matter - and on this basis it is concluded that matter, too, has to be momentary. This conclusion is based on the demonstration that mind and matter can only depend upon each other and interact as they do because they have the same duration. Second, by referring to ageing and similar processes it is proved that everything changes all the time and thus undergoes origination and destruction at every moment. This argument rests on the presupposition that any form of transformation implies the substitution of one entity for another. This proof from change reflects the presumable doctrinal background underlying the formation of the doctrine of momentariness. Third, it is argued that everything has to perish as soon as it has originated because, otherwise it would persist eternally. This would be at odds with the law of impermanence. The argument rests on the presupposition that destruction cannot be brought about from without and that it is impossible for an entity to perish on its own account after it has persisted as this would require a change of nature. The latter presupposition reflects the view that self-identical entities cannot change.

Vasubandhu (fourth-fifth century) marks the gradual transition between the earlier phase when the debate was still confined to Buddhism and the later phase when it was carried out between Buddhists and non-Buddhists. Vasubandhu only adopted the third type of proof, deducing momentariness from the spontaneity of destruction. He developed this idea further with the argument that destruction cannot be caused since, as mere nonexistence, it does not qualify as an effect. Up to the time of Dharmakīrti (*c*.600-60) and to a lesser extent thereafter, this proof of momentariness, the so-called inference from perishability (*vināśitvānumāna*), dominated the controversy.

With Dharmakīrti, the doctrine entered a new phase. He developed a new type of proof, the so-called inference from existence (*sattvānumāna*), that derives the momentariness of all entities (without presupposing their impermanence) directly from the fact that they exist. On the basis of the premise that existence entails causal efficiency, Dharmakīrti demonstrates that all existing things have to be momentary as it is impossible for nonmomentary entities to function as efficient causes. This impossibility derives from the idea that, if the entities

already produce their effect in the first moment, they also have to produce it again and again at all subsequent moments of their existence. This is an absurd position as their nature then does not differ from their nature in the first moment; nor can they discharge their causal efficiency gradually. If they were not able to produce their effect completely from the beginning, neither should they be able to do so later as this would entail a change of nature. This argument is also based on the premise that one and the same entity cannot change its properties.

The inference from existence became more prominent than the inference from perishability, although it never superseded it completely. Its prominence can be explained partly by the logical peculiarity of this proof which gave rise to an epistemological debate about the correct form of a valid syllogism. Since momentariness is to be proved for everything, all entities are the subject of inference (*pakṣa*). Thus, the inference from existence fails to fulfil two of the three classical conditions (*trairūpya*) for a valid syllogism, namely a positive and negative exemplification of the logical nexus (*vyāpti*) between the reason (to be existent) and the argued property (to be momentary) outside the subject (see Inference, Indian theories of).

Among other responses, this problem led to the modification of the conditions of a syllogism in such a way that those *vyāptis* also became accepted as valid where the logical relation between reason and argued property is not induced from other cases. This solution was already developed by Dharmakīrti himself, however, it was neglected until the time of Ratnākaraśānti (eleventh century). He argued that in those syllogisms where the proving property is intrinsic to the subject (*svabhāvahetu*), the logical nexus is to be established by demonstrating that the proving property cannot inhere in a locus that is lacking the argued property.

Frequently, as a corollary of these proofs of momentariness, the Brahmanical arguments against this doctrine are refuted. The most prominent argument - that the recognition of phenomena disproves their contended momentariness - is invalidated by the contention that recognition is a mixture of perception and memory and does not therefore qualify as a valid means of knowledge (*pramāṇa*). The related argument, that the mind cannot be a mere stream of momentary mental entities because memory and the discernment of causal relationships presuppose an enduring subject, is rejected. It is so on the grounds that the knowledge of the past is, by the principle of causal concatenation, passed on from one mental entity to the next. Thus, it is transmitted down to the present moment in a way which we may compare to the transmission of historic data from generation to generation.

Over the centuries the debate on the doctrine of momentariness developed to such an extent that Ratnakīrti (eleventh century) felt the need to deal with the inference from perishability, the inference from existence and the refutation of the proof of duration each in a separate treatise.

See also: Buddhism, Ābhidharmika schools of; Buddhist philosophy, Indian; Mujō; Potentiality, Indian theories of

ALEXANDER VON ROSPATT

References and further reading

The pertinent textual material has to be accessed in the following studies where it is presented in reliable translations (in English, French, or German) which can also be consulted by readers without Indological training.

Mimaki, K. (1976) *La Refutation bouddhique de la Permanence des Choses (sthirasiddhidūṣaṇa) et la Preuve de la Momentanéité des choses (kṣaṇaōgasiddhi) (The Buddhist Refutation of the Permanence of Things and the Proof of the Momentariness of Things)*, Paris: Publications de l'institut de civilisation indienne, fasc. 41.(Deals with later proofs of the doctrine of momentariness.)

Oetke, C. (1993) 'Bemerkungen zur buddhistischen Doktrin der Momentanheit des Seienden. Dharmakīrtis Sattvānumāna' (Remarks on the Buddhist Doctrine of Momentariness. Dharmakīrti's Sattvānumāna (Inference from Existence)), *Wiener Studien zur Tibetologie und Buddhismuskunde* 29, Vienna: Arbeitskreis für Tibetsich und Buddhistische Studien Universität Wien.(Employs contemporary analytical logic for the analysis of the *Sattvānumāna*.)

Rospatt, A. von (1995) 'The Buddhist Doctrine of Momentariness: A Survey of the Origins and Early Phase of this Doctrine up to Vasubandhu', *Alt- und Neu-indische Studien* 47, Stuttgart: Steiner.(Deals with the early phase of the doctrine of momentariness and analyses it doctrinal background on the basis of earlier proofs of momentariness; presents the most complete collection of primary sources (all translated into English) available on the subject matter.)

Steinkellner, E. (1968/9) 'Die Entwicklung des Kṣaṇikatvānumāna bei Dharmakīrti' (The Development of the Kṣaṇikatvānumāna (Proof of Momentariness) by Dharmakīrti), *Wiener Zeitschrift für die Kunde Südasiens*

12-13: 361-77.(Traces the development of the *Sattvānumāna* by Dharmakīrti.)

Monboddo, Lord (James Burnett) (1714-99)

In speculating that orang-utans' vocal organs must have been designed for speech, Monboddo was convinced that these creatures were primitive humans who had not yet entered society. His chief contribution to the history of linguistics and anthropology turns upon two propositions: that language is not natural to man, and that close physical resemblance between species is evidence of biological relation.

By training as a jurist and through his writings as a linguist and anthropologist, Monboddo was one of the most learned figures in eighteenth-century Scotland. Appointed a law lord or judge on the Scottish Court of Session in 1767 (from which his title derives), he drew lifelong inspiration from the classicism of Thomas Blackwell, whose writings on Homer and Augustus had helped to convince him of the decline of modern man and the decadence of modern forms of speech by contrast with the heroism of the ancient Greeks and Romans and the poetic resonance of their languages. More hostile to the empiricist tendencies of contemporary British philosophy than any other predominantly secular writer of the Scottish Enlightenment, he sought to rescue the glorious achievements of ancient science, ethics and rhetoric in reformulating an essentially Aristotelian interpretation of the human faculties, published in six volumes from 1779 to 1799, entitled*Antient Methaphysics* (see Enlightenment, Scottish).

A similar enthusiasm for classicism over modernity is manifest in Monboddo's more influential work of roughly the same period, *Of the Origin and Progress of Language*, also published in six volumes between 1773 and 1792. In addition to commenting on the splendours of ancient Latin and Greek, this text discusses the nomenclature of a variety of exotic languages, including Huron, Carib, Eskimo and Tahitian, which Monboddo had learned through dictionaries and travellers' reports. His attempt to trace the natural history of languages as an expression of both the universal capacities of the human mind and the specific genealogies of diverse cultures drew Monboddo in the direction of the nascent sciences of etymology and historical linguistics along lines developed by Sir William Jones (1746-89), with whom he corresponded. But he was even more drawn to the anthropological linguistics sketched in the *Discours sur l'inégalité (Discourse on the Origin of Inequality)* (1755) of Jean-Jacques Rousseau, from which Monboddo adopted and developed two main propositions: first, that language must be cultivated and mastered in society and hence is not natural to man; and, second, that the 'orang-utan' (in the Enlightenment a generic term for all the great apes) is human, since the inarticulacy of this creature so similar to a human being is attributable to its not yet having had the opportunity to enter society and therein to exercise its larynx, pharynx and other organs of speech. Monboddo imagined that analogous physical traits characteristically signify homologous functions, so that unless Nature had been so uneconomical as to engage in redundant design, the apparently human vocal organs with which orang-utans are endowed must have been intended for speech. It is a wonder that some of our forebears invented language at all, he claimed, since language does not spring spontaneously from our lips as does sight from our eyes or hearing from our ears. In time, and with proper tuition, he concluded, as some primatological researchers still do today, orang-utans would bridge the putative gulf created by language between human nature in society and the conduct of apes outside it (see Animal language and thought).

See also: Human nature, science of in the 18th century

ROBERT WOKLER

List of works

Monboddo (1773-92) *Of the Origin and Progress of Language*, Edinburgh: Balfour.
(1779-99) *Antient Metaphysics*, Edinburgh: Balfour.

References and further reading

Cloyd. E.L. (1972) *James Burnett: Lord Monboddo*, Oxford: Clarendon.(A concise intellectual biography.)
Sherwin, O. (1958) 'A man with a tail - Lord Monboddo', *Journal of the History of Medicine*, 23: 435-67.(A bibliographically detailed study of Monboddo's primatological anthropology.)
Wokler, R. (1988) 'Apes and races in the Scottish Enlightenment: Monboddo and Kames on the nature of man', in P. Jones (ed.), *Philosophy and science in the Scottish Enlightenment*, Edinburgh: John Donald.(A comparison of two conjectural histories of anthropology.)

Monism

'Monism' is a very broad term, applicable to any doctrine which maintains either that there is ultimately only one thing, or only one kind of thing; it has also been used of the view that there is only one set of true beliefs. In these senses it is opposed to the equally broad term 'pluralism'. But it is also often contrasted with 'dualism', since so much philosophical debate has focused on the question whether there are two different kinds of thing, mind and matter, or only one.

The ending '-ism' suggests a particular theory or school; but in the case of 'monism' this is somewhat misleading. Less misleading is to think of certain philosophies as monistic in respect of some central doctrine, realizing that which doctrine this is may differ widely from case to case. This approach gives the flexibility to see that a philosophy which is from one perspective a variety of monism may well be, in respect of some other central doctrine, not monistic at all. For instance, Spinoza took the view that there was and could be only one substance (see Substance). But he also held that this substance had infinitely many attributes, of which all but two (the mental and the physical) are unknown to us. There is nothing wrong with calling Spinoza a monist - provided it is said with an eye to a particular feature of his system.

With that warning in mind, we can say that the most common use of 'monism' is to describe philosophies which maintain that there is, ultimately, only one thing, and that 'the Many' are aspects of it or, to a more radical way of thinking, simply an illusion resulting from our mis-perception of the One. What they are not, for an ontological monist, is a collection of independent existences. Bertrand Russell once wrote that Hegel thought reality was like a tin of treacle, whereas he (Russell) thought it was like a heap of shot - the metaphor is even better if we keep the treacle in mind and forget the tin.

A spectacular early example of monism is found in the thought of Parmenides, who obscurely argued that reality could consist only of one thing, changeless and undifferentiated, and that the appearance of plurality was illusory (see Parmenides §7). In later antiquity, Plotinus influentially declared that everything emanated from another such changeless and undifferentiated entity, 'the One'. In the nineteenth century the view (associated in particular with F.H. Bradley) became widespread that relations between distinct entities cannot ultimately be real: for if two such entities are related by some relation *R*, there would be need of further relations to relate the relation *R* to each of them, and so on into infinite regress. Therefore there can only really be one thing, for if there were more they would have to be in some way related to each other. A similar argument occurs in Indian thought (see Monism, Indian §3). Those who are sceptical will suspect that it rests on the tacit and dubious assumption that a real relation would have to be a kind of thing, like the things it relates.

A very different group of doctrines is also often called 'monism': those doctrines which assert, not that there is only one thing, but that there is ultimately only *one kind* of thing, one basic stuff of which the many kinds we observe are variant forms. Tradition unreliably has it that Thales, the earliest Greek philosopher of whom we have record, taught that this basic substance was water; a little later Anaximenes held a similar view, but chose air as his elemental stuff. They were therefore monists in this second sense, as is anyone who believes that the variety of the world is really nothing more than different arrangements of atoms which are themselves identical (see Atomism, ancient).

One version of the 'one stuff or more?' debate has been especially prominent in modern philosophy: are there two kinds of substance - matter and mind - or only one, and, if so, which? Here the question is not 'are there one or more?' but 'are there one *or two*?', and when philosophers speak of monism as the alternative to dualism (without saying one or two *what?*) it is nearly always this issue that they are referring to. (A quite different use of the term 'dualism' is used in religious thought to refer to the view that the world is the work of two powers, one good and one bad - see Manicheism §§1, 4.) In modern times, Descartes is the classic exponent of mind-body dualism, though hardly its inventor - he was only giving precision and proof to the very old idea that as well as matter there are spirits or souls, and that the latter are very different in kind from the former. Ranged on the other side are various styles of monist. There is idealist monism, typified by Berkeley, which holds that there are only minds or spirits, and that material bodies are nothing but a way of speaking about mental states (see Phenomenalism); there is material monism, steadily more popular with the rise of the natural sciences, which views everything as

material, and reduces the supposedly mental to facts about matter (see Materialism in the philosophy of mind). And there is a doctrine associated in particular with the names of Mach, James and Russell, which sees the mental and the physical as different aspects of a single basic stuff that in itself is neither mental or physical (see Neutral monism).

Because of its very general meaning, ‘monism’ lends itself to a wide variety of uses easily understood from their context. Thus one may see ‘value-monism’ - meaning an ethical system which recognizes only one thing as having value in itself, for example happiness (see Utilitarianism). And one may speak of monism about truth, meaning that there is one unique set of truths - in opposition to pluralism, relativism and perspectivalism, according to which conflicting or incommensurable views may be equally true.

See also: Pluralism

EDWARD CRAIG

References and further reading

Bradley, F.H. (1897) *Appearance and Reality*, London: Allen & Unwin, 2nd edn, Appendix II.(A *locus classicus* for the argument to monism from the unreality of relations.)

Gallop, D. (1984) *Parmenides of Elea*, Toronto, Ont.: Toronto University Press.(Text and translation of the *c.*150 extant lines by Parmenides.)

James, W. (1909) *A Pluralistic Universe*, London: Longman, Green & Co.(Chapter 2, ‘Monistic idealism’, is a lively polemic against nineteenth century monists. Accessible.)

Moore, G.E. (1922) ‘External and Internal Relations’, in *Philosophical Studies*, London: Routledge & Kegan Paul, 276-309.(Moore resists the view - dear to the ‘tin of treacle’ type of monist - that nothing can be the thing it is without standing in all the relations to other things it actually does stand in.)

Plotinus (*c.*250-70) *Enneads*, trans. S. MacKenna, London: Penguin, 1991.(See in particular Enneads IV-VI; for readers in a hurry, VI.9: 535-49. Difficult reading.)

Spinoza, B. (1677) *Ethics*, in *The Collected Works of Spinoza*, trans. and ed. E. Curley, Princeton, NJ: Princeton University Press, 1985.(Part I, ‘Of God’, contains Spinoza’s proof that there is only one substance.)

Monism, Indian

The prominent classical and modern Indian philosophy known as Advaita Vedānta, which insists on the single reality of Brahman (the Absolute), is often identified as Indian monism. But the monism of Advaita is only a portion, albeit central, of the Advaita view. Furthermore, a monism in theology (Brahman as God) is important to almost all expressions, classical and modern, of Indian theism.

The monism of Advaita is principally psychological. Nondual awareness is considered the true self; that is to say, in the self's native state, the object of awareness and awareness itself are identical. This kind of awareness is claimed to be presupposed by all dualistic consciousness. Moreover, it is said that only self-aware self-awareness itself cannot be revealed by experience to be illusory. And according to Advaita, a supreme mystical experience, popularly called liberation, does in fact, when it occurs, reveal self-awareness to be the sole reality. A dialectical Advaita adds the further contention that it is impossible to define and explain coherently diverse appearances. This contention is cashed out by long and intricate attacks on the pluralistic ontologies of rival schools, particularly Nyāya-Vaiśeṣika.

The monism of Indian theism centres on the reality of God, who is constrained by metaphysical law to create out of the single spiritual substance that God is. The world is commonly said to be God's body. Various ramifications of God's being in some way everything can be discerned in Indian theology.

1 The Advaita view of self

Advaita Vedānta is a scripturally derived philosophy centred on the proposition, first found in early Upaniṣads (800-300 BC), that Brahman - the Absolute, the supreme reality - and the self (*ātman*) are identical. The identity is understood as an objectless consciousness, as awareness nondualistically self-aware. Arguments in support of the view that nondual awareness is the sole reality are developed by classical and modern Advaitins, from Gauḍapāda (*c*.600 AD) and Śaōkara (*c*.700 AD), in hundreds of texts. Some of these are suggested in Upaniṣads.

Scripture - that is, Upaniṣadic passages - is the principal reason cited why, in the first place, awareness aware of nothing but itself should be considered a reality. A supreme mystical experience, deemed the goal of life, is said to be just such a state, and scripture is interpreted as teaching this. But that so-and-so reports such mystical experience is not cited, at least not commonly, in support of the view that such self-awareness occurs. Scripture is the testimony of the occurrence.

That nondual awareness is the only possible *self*-awareness is defended by a *reductio* argument. If a further awareness C_2, having C_1 as content, is required for self-awareness, then since there would be no awareness of C_2 without awareness C_3, *ad infinitum*, there could be no self-awareness, that is, unless the self is to be understood as limited to past awareness only. For self-awareness to be an immediate awareness, self-awareness has to be nondual. It is a further question how self-aware awareness accounts for our everyday sense of self and our commitment to the reality of self, a question of psychology that in particular a few late classical Advaitins struggle with. It is problematic, for example, how our native self-awareness, aware only of itself, accounts for our everyday sense of ourselves as a body. Moreover, self-awareness, as understood by Advaita, is said to be presupposed by all consciousness, but the precise connection - other than as an ongoing possibility of meditative self-absorption - is not clearly worked out. Nevertheless, the argument that self-awareness would be impossible without nondual awareness is forceful, it is said, for all who admit self-awareness, and forceful, so it is presumed by many Advaitin advocates, independently of the question of the precise nature of the tie to the everyday sense of self.

In support of awareness as self-illumining, the great Advaitin Śaōkara (often taken to be the very founder of Advaita) argues that no one says 'I am not.' This Advaita argument for self-illumination is often compared to Descartes' *Cogito*. As self-illumining, awareness blocks doubt: the doubt never occurs, 'Am I aware or not?' In this way an appeal is made to what is sometimes called pragmatic contradiction: there would be contradiction between essence and content were someone to have an awareness verbalizable as 'I am not aware.' (Compare a speaker saying 'I am not speaking.')

From self-illumination it is argued that an epistemic principle known as self-certification (*svaprāmāṇya*) is right

concerning all questions about awareness, since only an awareness itself has, so to say, access to itself. Awareness itself is the only consideration relevant to any question about awareness itself, its existence or its nature. Self-certification is thus grounded in self-illumination. And self-certification is key to further Advaita arguments.

2 Sublatability

The most commonly voiced argument in defence of the Advaita view of self-awareness as the single reality centres on a presumed sublatability of all indications of awareness other than self-awareness itself. Perceptual illusion shows that there is no worldly content of an experience that could not prove unreal. In contrast, (again) no one says, or could say (in good faith), 'I am not.' The self, unlike objects presented perceptually, all of which might not be, cannot, if presented, possibly not be. Self cannot be sublated. Perceptual illusion shows that anything else can be. The appearances of a rope as a snake, of mother-of-pearl as silver, and so on, show that the worldly content of experiences is not to be trusted. Only self-experience is sure (see Error and illusion, Indian conceptions of §1).

Thus awareness self-certifyingly illumines self-awareness but not anything else. External objects are not self-certifyingly illumined because any presentation of an external object is sublatable, whereas the self-illumination of awareness itself is nonsublatable. All awareness may be *prima facie* self-certifying, but only self-awareness is *ultima facie* self-certifying, for all content other than self-illumination is sublatable, negatable experientially.

Advaitins move from these Cartesian-like considerations to embrace a very strong criterion for ontic commitment. An absolute nonsublatability is the mark, they say, of the real, or existent, *sat*. That which is not sublatable by any manner or means is to be accepted as real. Only self-awareness meets the criterion. Self-awareness is the single reality.

There is another argument centring on a sublatability of a different sort, not the experiential sublatability of presentations other than self, but a propositional defeasibility. Brahman, the Absolute, with which in the Upaniṣads the self is declared to be identical, is presented in the Upaniṣads as also including everything (*sarvam idaṃ brahma*, 'All this is Brahman'). The scripturally conveyed cognition of Brahman is not the same as self-awareness aware only of itself, for the scripturally conveyed cognition of Brahman is verbal, conveyed linguistically. It is only self-awareness that cannot possibly be sublated experientially and thus it is only that awareness that indicates reality in the truest sense of the word. But (unfortunately) we do not live our lives in such a state of self-absorption. Within the sphere of everyday life, we adopt a modified self-certificationalism: A cognition is to be accepted as true unless defeated by countervailing considerations. That is to say, innocence, or an assumed warrantedness, is the epistemic default. Now this holds concerning the scripturally conveyed monistic cognition 'Brahman' (Brahman is everything). But also, Advaitins argue, Brahman cognition is in principle indefeasible, for two reasons. First, it is the proper province of scripture, and of scripture only, to tell us things of this order, as the sun (says Śaōkara, who is echoed in dozens of texts) is the proper illuminer of colour. Second, a proper challenger to Brahman cognition would similarly have to be about everything. But if a cognition were truly about everything (including itself) it would be identical with Brahman cognition. Thus there can be no challenger of Brahman cognition other than Brahman cognition itself, which is no challenger. Brahman cognition is both *prima facie* warranted (innocence is the justificational default) and *ultima facie* warranted, in that it cannot be challenged.

The monism of an all-inclusive Brahman as declared in the Upaniṣads is not the heart of the Advaita position, which is instead the psychological monism of self-awareness. But the two are related, as we can see by looking at another argument, namely that the *summum bonum*, popularly termed 'liberation' and by philosophers 'knowledge (or mystical experience) of Brahman' (*brahma-vidyā*), does in fact, when it occurs, reveal self-awareness to be the sole reality. The fundamental reason why scripture teaches that Brahman is everything is to provide a meditational support helpful in achieving the supreme good. There may seem to be some truth in the idealist view that Brahman is the ground and support of everything - indeed its material reality, as the material reality of a gold bracelet is just the gold - in that every appearance appears just to the self. However, there is no more truth in this view than in the theistic position that God is the creator and sustainer of things, a teaching also helpful for some. Scripture's purport is thoroughly soteriological; that is to say, it aims at bringing us to the realization of self-illumining self-awareness. Scripturally conveyed cognitions are sublated, like every cognition other than self-awareness itself, in the supreme experience. But the monism of an all-inclusive Brahman declared

by scripture is, within the sphere of spiritual ignorance (*avidyā*), a most helpful teaching for transcending spiritual ignorance - as well as one that is indefeasible by everyday experience and knowledge. (Advaita's exegetical strategy is similar, as suggested, with regard to Upaniṣadic teachings about God.)

There are more or less obvious problems with these arguments, problems that classical and modern opponents of Advaita have identified. We shall mention just a few. First, it is not clear that self-illumining self-awareness is nonsublatable experientially, particularly as Advaitins understand (experiential) sublation - that is, on analogy with an experiential correction of a perceptual illusion. A nonveridical perception of a snake is sublated and corrected by a veridical perception of a rope. For Advaitins, it seems to be the posteriority of the correcting perception that is key. Why then could not a fuller experience of self as embodied, or being in the world, follow a meditational experience of self-absorption? Second, Brahman is declared in the Upaniṣads to be bliss, among other attributions, not only self-awareness. Could there not be, then, a distinct challenger to the scriptural teaching about Brahman which was similarly about everything: for example, 'All-inclusive Brahman is not bliss?' And finally, there are Buddhists and others who say that the supreme mystical experience does not reveal a self, but, for example, a no-self. Whom should we believe?

3 Dialectical Advaita

Although Śaōkara and other early Advaitin philosophers concern themselves principally with the logic of illusion along with exegetical questions, they also launch dialectical attacks against rival schools, and specifically against the category of external relationality crucial to all realist and pluralist ontology. In the twelfth century, an Advaitin named Śrīharṣa ingeniously expands the earlier polemics, aiming in particular at the realist system known as Nyāya-Vaiśeṣika and purporting to dismantle it plank by plank. Śrīharṣa takes more seriously the all-inclusiveness and simplicity of Brahman presumed to be taught by scripture, and believes that the reality of Brahman means that it is impossible to define and explain coherently apparent diversities. He cashes out this commitment with an onslaught upon Nyāya's pluralist ontology, an onslaught in which attacks on relations figure prominently (see Nyāya-Vaiśeṣika §§4-5). Later Advaitins (including modern philosophers) rework some of his arguments to explode other views.

If there are two things *a* and *b* related by the relation *R* - thus *aRb* - in virtue of what is *R* related to *a*, and likewise to *b*? A second relation and a third, *ad infinitum*, would seem to be required. Śrīharṣa employs this logic (in the West sometimes called the Bradley problem (see Bradley, F.H. §5)) in a variety of contexts, for example, regarding the relation between a property-bearer and a property:

> If the property is unrelated to the property-bearer, there is an obvious problem; [if, on the other hand, it is related] there will be an endless number of relations and thus infinite regress. Or if at the beginning or the end the relation is admitted to be of the very nature of one of the terms [property or property-bearer], then since even the other term of the relation would enter into the very nature of that [the combined relatee-relator], nothing but nondistinctness would result.
>
> (*Khaṇḍanakhaṇḍakhādya*, 107)

Three options are sketched: (1) a property, such as blue, is unrelated to its bearer, such as a pot; (2) if there is a relation that relates them, such as inherence, then there have to be further relations to relate the inherence to each of the terms, the blue and the pot, *ad infinitum* (*aRb, aR′R, aR″R′R, ad infinitum*, likewise with the second term); unless (3) it is the very nature of one of the terms to link with the other: such linkage would amount to nondistinctness. The third seems the only viable option. Nondistinctness, however, is at odds with Nyāya pluralism, and thus the argument (along with others), Śrīharṣa concludes, shows that there is no coherent Nyāya challenger to the monism of Brahman taught by the Upaniṣads.

In sum, distinct things have to be in relation, but relations are impossible. Thus by presupposing fundamentally distinct things, pluralists run into a host of difficulties, as indicated by this demonstration of incoherence (and others). Gaōgeśa (*fl. c.*1325), a later logician of the Nyāya school, responds that it is an error to treat a relator as the same type of thing as its relatees, and it is this mistake that engenders the appearance of vicious regress. Of course, Advaitin followers of Śrīharṣa find further problems with the response, and the controversy, one of the finest preserved in Sanskrit, continues for several more centuries.

4 Theological monism

The cornerstone of the monism of Indian theism is found in an Upaniṣadic story of a father teaching his son about Brahman:

> Just being, my son, was this [universe] in the beginning, one alone, without a second. There are, to be sure, some that say, "Just non-being was this [universe] in the beginning, one alone, without a second. From that non-being, being arose." But how, my son, could it be so?… How from non-being could being arise? On the contrary, just being, my son, was this [universe] in the beginning, one alone, without a second.
>
> (Chāndogya Upaniṣad 6.2.1-2)

Here we find an early expression (700 BC?) of the metaphysical law *ex nihilo nihil fit*, 'nothing from nothing'. The Upaniṣads weave several fanciful stories about how the One becomes many, stories that are interpreted by later theologians as underscoring emanationism. God looses forth (*sṛjate*) gods, Vedas and Upaniṣads, humans, animals, plants, rocks and dust out of God's own substance. The universe is God's body.

Indian theists commit themselves to a thoroughgoing reality of a plurality of things; spiritual monism none the less provides an absolute boundary, or a resource, for their theorizing (except with the Dvaita, 'Dualist', Vedānta school). For example, evil, a challenge to any view that would uphold the goodness of a Creator (see Evil, problem of), has a different look within a theological monism as opposed to a theism that has God separate from the world, creating *ex nihilo*. God, being everything, shares in everyone's suffering; evil looks more like masochism than a matter of God's harming others. Moreover, God's being everything is thought to guarantee a universal harmony. It is impossible for society, for example, to disintegrate into chaos. If there were such a threat, God would appear (would be forced to appear) in a special manifestation, *avatāra*, to assure a return to harmony. This is a principal message of the *Bhagavad Gītā (Song of God)* and other theistic texts.

Finally, the monistic dimension of Indian theism spawns, in contrast with the illusionism of Advaita Vedānta and much Buddhism as well, a distinct attitude towards the body and physical things. Nature is enchanted, the body a mystic instrument. To be sure, asceticism, much admired in classical Indian society in accord with Advaita and Buddhist teachings, colours the religious practices of Indian theism. But worship and other practices reflecting belief in a divine immanence tend to prevail.

See also: Awareness in Indian thought §3; Brahman; God, Indian conceptions of; Monism; Self, Indian theories of; Vedānta

STEPHEN H. PHILLIPS

References and further reading

Aurobindo, Sri (1914-20) *The Life Divine*, Pondicherry: Sri Aurobindo Ashram Trust, revised edn, 1943-4, 2 vols.(Aurobindo is a modern Indian theist who expresses the logic of theological monism particularly well.)

Bhagavad Gītā (200 BC - AD 200, disputed), trans. F. Edgerton, Harvard Oriental Series 38-9, Cambridge, MA: Harvard University Press, 1944, paperback repr. 1972. (Though Edgerton's interpretive essay - orginally volume 39 in the HOS - too much reflects outworn assumptions of nineteenth-century indologists, his translation is excellent, faithful and elegant. There are, however, dozens of acceptable translations into English and other modern languages.)

Dasgupta, S. (1922-55) *A History of Indian Philosophy*, Cambridge: Cambridge University Press, 5 vols.(Volume 4 has a section on the late classical debate between dialectical Advaita and the Dualist school of Vedānta, whose pluralist ontology has much in common with Nyāya.)

Phillips, S.H. (1995) *Classical Indian Metaphysics: Refutations of Realism and the Emergence of 'New Logic'*, La Salle, IL: Open Court.(Focuses on dialectical Advaita and responses provoked.)

Potter, K.H. (ed.) (1981) *Encyclopedia of Indian Philosophies*, vol. 3, *Advaita Vedānta*, Princeton, NJ: Princeton University Press.(Contains an excellent introduction to Advaita philosophy as well as summaries of works in the early Advaita school.)

Śrīharṣa (12th century) *Khaṇḍanakhaṇḍakhādya*, ed. N. Jha, Kashi Sanskrit Series 197, Varanasi: Chowkhamba, 1970.(The title translates as 'Delectable Refutations'; this is *the* classic of dialectical Advaita Vedānta.)

Thibaut, G. (trans.) (1890) *The Vedānta Sūtras of Bādarāyaṇa, with the Commentary by Śaṃkara*, New York: Dover.(An excellent, highly readable translation of Śaōkara's principal philosophical work.)

Upaniṣads (800-300 BC), trans. P. Olivelle, *Upaniṣads*, Oxford: Oxford University Press, 1996.(A new translation

that may well be the best; readable and accurate. The translations in the present entry are by Stephen Phillips.)

Monotheism

Judaism, Christianity and Islam are usually cited as the major monotheistic religions. These are religions which acknowledge only a single god, and which construe that god as transcendent - that is, as a being who is distinct from the ordinary world and superior to it. They also construe this god as a person or as very much like a person. The polytheistic religions agree with monotheism, for the most part, in construing the gods in personalistic terms, but they acknowledge a plurality of gods. Pantheists, on the other hand, usually accept the singularity of the deity, but reject the transcendence, identifying the deity more closely with the ordinary universe, perhaps as a certain aspect of the universe or as the totality of the universe considered in a certain way. They also are likely to reject the personalistic idea of the deity.

This entry will discuss some philosophical aspects of the contrast between monotheism and polytheism. There are different ways of understanding 'acknowledge' in the characterization of monotheism given above. One may believe that there exists only one god. Or one may believe that more than one god exists, but worship only one, or hold that it is wrong to worship more than one. A closely related issue concerns the concept of deity that is employed in such beliefs and claims. The 'high' Anselmian conception of Christianity's god - a being than whom no greater can be conceived - is only one account. Perhaps of most philosophical interest is the problem as to whether logical argument can demonstrate polytheism to be false and establish the truth of the Anselmian account.

1 Types of monotheism

Monotheism contrasts on one side with atheism and on the other with polytheism. Philosophers, initially at least, may take monotheism to be a doctrine to the effect that there exists exactly one god, no more and no less. Atheism, then, would be the doctrine that there exist no gods at all, and polytheism would be the claim that there are several gods (see Atheism §1). A monotheist would be a person who accepts the monotheistic doctrine, and similarly for atheists and polytheists.

All three of these 'isms', however, are ambiguous in an important way. Consider another contrasting pair of notions, which mark an important distinction among various social systems - monogamy and polygamy. Hardly anyone would take seriously the suggestion that we should construe a monogamist as a man who believes that there is only one wife, or that we should take monogamy to be the doctrine that one and only one wife exists. It would be much more plausible to construe monogamy simply as a practice - that of having only one wife - and a monogamist would then be a man who had only one wife. Polygamy, of course, would then be the contrasting practice of having more than one wife. And some people might want to add a somewhat more theoretical and doctrinal element to these conceptions, perhaps saying that monogamy includes a normative claim, the claim that it is not proper for a man to have more than one wife; polygamy would then involve the contrasting normative claim.

In an analogous way, there is a religiously important sense of monotheism in which it consists simply in having only one god - that is, in committing oneself to, worshipping, obeying, and so on, only one god. There is a second sense in which monotheism is the doctrine that there is something improper and illegitimate in worshipping more than one god (and perhaps also that there is something improper about not worshipping any god). And then there is a third sense, that in which monotheism is the doctrine that there exists one and only one god. These will be referred to in this entry as, respectively, *cultic monotheism*, *normative monotheism* and *descriptive monotheism*.

It is useful to keep these distinctions in mind when considering classical texts dealing with this general idea. Pre-Christian Judaism, for example, is often cited as an early example of monotheism, and indeed as the precursor of monotheistic Christianity. But consider a classic monotheistic Hebraic text, the first of the Ten Commandments: 'Then God spoke all these words: I am the LORD your God, who brought you out of the land of Egypt, out of the house of slavery; you shall have no other gods before me' (Exodus 20: 1-3). Pretty clearly, cultic monotheism is here being enjoined on Israel. They are to 'have' no other gods than the LORD (Yahweh). And perhaps the fact that this commandment is attributed to God himself may suggest that there is also a normative element in this monotheism. (Even so, this text would not support the view that it is improper for anyone to have more than one god, or one of those other gods. This command is explicitly addressed to Israel.) But the text seems to have no suggestion in it at all of descriptive monotheism. In fact, it may well be construed as positively envisaging the possibility of the existence of other gods.

Cultic and normative monotheism are logically compatible with descriptive polytheism, and apparently they are religiously compatible with it also. There have been, and probably still are, religions whose adherents clearly recognize the existence of gods to whom they themselves give no allegiance. An interesting and explicit biblical example comes from the monarchic period of Israel's history. In that story, some Aramean army officers undertake to explain to their king why they were defeated by the Israelites. 'Their gods are gods of the hills,' they say, 'and so they were stronger than we.' And they propose a different strategy for the next encounter: 'Let us fight against them in the plain, and surely we shall be stronger than they' (I Kings 20: 17-25). The Arameans of this story express no doubt about the existence of the Israelite gods - in fact, they attribute their own defeat to the power of those foreign gods in the hill country of northern Israel. But they themselves have no allegiance to those gods. They do not profess to worship them or to serve them. In fact, they are plotting to defeat them at the next opportunity, by fighting on the plains where their own gods are at home. These Arameans were descriptive polytheists with respect to the Israelite gods, and probably they were both cultic and descriptive polytheists with regard to another set of gods, their own.

2 Concepts of deity

If the story of the Arameans seems strange and awkward, this reflects the fact that the word 'god' (and 'God') involves an ambiguity similar to the one we have been considering. There is a sense of this word, fairly common now, in which it has a cultic element built into it. In this sense, to use the word 'god' to describe or identify some being is *ipso facto* to profess a certain attitude - worship, obedience, reverence, and the like - towards that being. Sometimes this cultic element is made explicit in something like a definition. Perhaps the best-known modern definition of this sort is that attributed to Paul Tillich (§3), that God is the object of ultimate concern. A person who accepted that account would speak paradoxically, to say the least, if they were to say 'Of course God exists, but he is of no concern to me.' Some students of early Buddhism, however, think that just this combination may well have been the view of Gautama the Buddha (see Buddha). They suspect that Gautama held that (perhaps only probably) there are gods, but that they have no concern with human life and so we need not concern ourselves with them. A similar view is sometimes attributed to Epicurus (see Epicureanism §9).

This raises a question about the concept of a god. Monotheism, polytheism and atheism differ with respect to the number of gods, but all are equally in need of some elucidation of the concept of a god. Anyone who professes one of these 'isms' seems to need some idea of the nature of a god. What sort of thing is it of which only one, or more than one, or none at all is supposed to be worshipped? What sort of thing is it of which there is just one, or more than one, or none at all? Throughout the Christian period, Western philosophical theology has gone very strongly to extremely 'high' conceptions of the deity. Perhaps the best-known formulation of a high conception is that of Anselm of Canterbury (§4), who understood God to be a being than whom no greater can be conceived (*Proslogion*, ch. 2). Whether this conception guarantees descriptive monotheism will be considered in §3 (see God, concepts of §§3-6).

For the moment, however, it is useful to recognize that there have surely been religions, and perhaps there still are some, that do not construe their gods in this high way. The Arameans referred to earlier certainly cannot have had anything remotely like Anselm's idea in mind when they spoke about gods, either their own or those of the Israelites. For they construed those gods as beings whose power would probably run out forty miles from home, in the transition from the hill country to the Syrian plains. The gods of pre-Christian Greek and Roman popular religion also do not seem to make a plausible fit with anything like the Anselmian notion. And so on. Is there any other notion of a god, something radically non-Anselmian, which would be applicable to a wide range of actual religions? Some philosophers think so. Richard Swinburne, for example, says 'I understand by a god a non-embodied rational agent of great power' (1970: 6). Probably the gods as the Arameans construed them, along with the gods of ancient Greece and Rome, and many others, would fit this definition fairly well. And so, for that matter, would God as he is construed in Christian theology. The Anselmian concept might be a more adequate account of the Christian deity, and perhaps a more interesting and provocative account. But God, as he is described by orthodox Christian theologians, would also satisfy Swinburne's definition. He too is a non-embodied rational agent of great power.

With reference to Swinburne's definition, then, the Arameans of the biblical story were probably doubly polytheistic in the descriptivist sense, believing in the existence of several gods to whom they owed allegiance and

also in the existence of several to whom they did not owe allegiance. The ancient Greeks and Romans were apparently also descriptive (and cultic) polytheists. Perhaps more interestingly, most Christians are descriptive polytheists in Swinburne's sense of 'god' because angels (and devils also) fit his definition perfectly well. But of course Christians are not, on that account at least, cultic polytheists. For though they often profess to believe in the existence of angels and devils, they do not characteristically profess to worship these beings.

Many Christians, of course, will feel uncomfortable with Swinburne's definition, and would not like to be identified as polytheists. This perhaps reflects the strongly cultic connotations which the word 'god' now has, especially when it is capitalized. These Christians, quite properly, do not want to be taken for worshippers of angels, or even for people who think that it might be appropriate to worship an angel. But they are, for the most part, content to be known as people who believe in the existence of angels. And being a descriptive polytheist, in Swinburne's sense of 'god', need not involve anything more than that.

3 Rationales for monotheism

The most common sort of rationale given for cultic monotheism is strongly religious in character. So in the Hebraic account, for example, God is said to have rescued the people of Israel from their bondage in Egypt, exercising his power on their behalf. He then makes a covenant with them, and one element of this covenant is that they will have no other god. The idea of this covenant is deeply embedded in the self-identification of this people. People who accept that self-identification as their own thereby accept a commitment to that deity. That god, and no other, will be their god.

Sometimes the commitment is made on a more personal level. The biblical story of Ruth, for example, is the story of a young widow who decides to leave her home in Moab and to move to Israel with her mother-in-law. Declaring her decision, Ruth says to Naomi, 'your people shall be my people, and your God my God' (Ruth 1: 16). She has made her decision to take up a new life in a new place and a new community, and part of that decision involves making a commitment to a new god and abandoning whatever commitment she may have had to the gods of Moab. She may have been a cultic polytheist in Moab, but now, she resolves, she will commit herself to the one god whom Naomi worships.

Rationales for descriptive monotheism often have a more philosophical flavour, appealing to more general principles, such as logic. They may thus be of more general philosophical interest. Especially for attempts to establish descriptive monotheism, it is crucial to fix somehow on the relevant concept of deity. And it is important also to remember that descriptive monotheism has a contrast on each side of it - atheism and polytheism. Establishing monotheism must involve ruling out both of these alternatives.

Perhaps it seems easiest to attack descriptive polytheism first. The most attractive strategy for this project is analytic or definitional. One looks for a concept of deity which is such that it would be logically impossible for more than one actual existent to satisfy it. Of course, Swinburne's concept will not be satisfactory for this purpose. There is nothing in it which rules out the possibility that several beings could satisfy it. But maybe there is some 'higher' concept which guarantees singularity.

Christian theologians have often included the attribute of omnipotence in their idea of God, and this may seem to provide a convenient connection with singularity (see Omnipotence §1). Initially at least, it is plausible to think that there could not possibly be two omnipotent beings. Which one would prevail, after all, if they came to blows? A little more formally, it might be thought that an omnipotent being ought to be able to overcome any external obstacle and also to thwart any external initiative. So, if there were two distinct omnipotent beings, *A* and *B*, then *A* should be able to overcome *B*, and *B* should also be able to resist *A* and thwart him. But that combination seems logically incoherent; if so, then it may constitute a *reductio ad absurdum* disproof of the suggestion that there could be two distinct omnipotent beings.

The concept of omnipotence, however, is itself surprisingly difficult to clarify, and that in turn generates a difficulty for this line of argument. At least since the time of Thomas Aquinas, most Christian theologians have construed omnipotence in such a way that it does not involve the power to do logically impossible tasks. Given this understanding of omnipotence, it is hard to see how to carry out the *reductio* successfully. Assume (for the *reductio*) that both *A* and *B* exist and are omnipotent. Perhaps it can then be assumed that the omnipotence of *A* should enable him to overcome any resistance offered by *B*. Well and good. But, on this understanding of

omnipotence, the failure of *B* to resist *A* successfully does not count against *B*'s omnipotence. For, on this reading, thwarting *A* is a logically impossible task. After all, *A* is omnipotent; he cannot be resisted. And the impossibility of the task removes it from the sphere of omnipotence. Therefore, the fact that *B* cannot succeed in the logically impossible project of resisting *A* does not represent a failure of omnipotence in *B*. There is no contradiction in the admission that *B*, though omnipotent, cannot thwart *A*. And so the *reductio* fails.

Suppose we reject the 'majority' view of omnipotence associated with Aquinas and opt instead for the minority view. On this view, omnipotence enables its possessor to do even those things which are logically impossible. Given this reading of omnipotence, *B* is not excused from the task of thwarting *A* merely because that task is impossible. But now the *reductio* runs into difficulty at a different point. For now we can say that *A*, being omnipotent, will certainly succeed in his project, and *B*, also omnipotent, will surely be able to thwart *A*. Of course, that seems thoroughly incoherent, the description of a logically impossible state of affairs. It entails that *A* both succeeds and fails to succeed, and also that *B* both succeeds and fails. But given the minority understanding of omnipotence, what would be the significance of discovering this logical incoherence or impossibility? After all, on the minority view, impossible things fall within the power of omnipotence. If *A* and *B* are omnipotent in the minority sense, then they may well be able to bring about this logically impossible state of affairs. Here too the attempt to construct the desired *reductio* seems to fail. As far as these arguments go, then, we are left with the possibility that there might indeed be two omnipotent beings.

On neither of these views of omnipotence, therefore, is it clear that we can rule out descriptive polytheism simply by including omnipotence in our concept of a god. We might, of course, try some other attribute which theologians generally include in a high conception of the deity. But in fact there would seem to be an easier way to succeed along these lines. We can simply attach an explicit singularity requirement, perhaps by adding the phrase 'the one and only', to some already formulated specification of a concept of deity. We can, for example, start with Anselm's account, and generate from it the revised formula 'the one and only being than whom no greater can be conceived'. Let us call this the 'singularity' concept of deity. If we adopt the singularity concept, then at least one half of the descriptive monotheist position is guaranteed. There is not, and there cannot be, more than one actually existent being who satisfies this description. Furthermore, we can say that if there is exactly one existing being who satisfies Anselm's original concept, then that same being also satisfies the singularity version. And if nothing satisfies Anselm's concept, then nothing satisfies the singularity version either.

This strategy, however, also costs something. It provides a comparatively easy way to rule out descriptive polytheism (for the chosen concept of deity, of course), but it makes it harder to rule out atheism. In the example just given, for example, we can notice that if two (or more) existent beings satisfy Anselm's concept, then nothing at all satisfies the singularity concept. For in that case there would be no being who was the one and only being than whom no greater can be conceived. There would instead be several beings who were tied for the position of maximal conceivable greatness. But, of course, none of these beings would satisfy the singularity requirement.

Anselm, of course, argued that there is an existent being who satisfies his concept - that is, that there actually is a being than whom no greater can be conceived. And there have been some impressive and provocative recent reformulations of the Anselmian line of argument (for example, Plantinga 1974). But neither Anselm nor his recent followers clearly address the question of whether there may be more than one being than whom no greater can be conceived. We might say, therefore, that if their arguments are successful they may refute 'Anselmian atheism' and establish Anselmian theism, where Anselmian atheism and theism are defined by reference to Anselm's concept of God. But there is also 'singularity atheism', which consists of the claim that nothing satisfies the singularity concept. The arguments of Anselm and his followers do not seem to refute singularity atheism. Consequently, they do not establish descriptive monotheism according to either the Anselmian or the singularity concept of deity.

There are, of course, other lines of argument that might be explored. We might specify, for example, that we think of God as the creator of the world, and then go on to argue that the universe exhibits so much unity that it is plausible to attribute it to a single creator. That strikes some philosophers as rather attractive. But David Hume (§6) suggested that it is just as plausible to think that the universe was created by a committee (1779: part V). So this line of argument is not convincing to everyone. In fact, it is likely that no sort of philosophical argument for descriptive monotheism will be universally attractive. If we are concerned about descriptive monotheism at all, it

might be just as well to rely here too on distinctively religious considerations, perhaps indeed on some appeal to divine revelation (see Revelation §§1-2).

See also: God, arguments for the existence of; Pantheism; Religious pluralism; Trinity §§1-2

GEORGE I. MAVRODES

References and further reading

Anselm (1077-8) *Proslogion*, trans. J. Hopkins, *A New Interpretive Translation of St Anselm's Monologion and Proslogion*, Minneapolis, MN: Arthur J. Banning Press, 1986.(This short book contains the most widely known formulation of the ontological argument for the existence of God, as well as a very influential formulation of the concept of God.)

Aquinas, T. (1266-73) *Summa theologiae*, trans. Fathers of the English Dominican Province (1911), Westminster, MD: Christian Classics, 1981, 5 vols.(A concise discussion and defence of descriptive monotheism by one of the greatest philosopher-theologians of the medieval period; see particularly Ia, q.11, a.3.)

Hume, D. (1779) *Dialogues Concerning Natural Religion*, ed. N.K. Smith, Indianapolis, IN: Bobbs-Merrill, 1947.(This is an extensive discussion and critique of arguments which purport to prove the existence of God on the basis of the order and design in the world.)

Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Clarendon Press.(Much of this book consists of rather technical discussions of modal logic, but the chapter on the ontological argument (196-221) contains one of the most influential modern versions of that argument. It requires careful reading, but it is understandable without much background in modal logic.)

Swinburne, R. (1970) *The Concept of Miracle*, London: Macmillan.(This is mainly a discussion of the possibility of miracles, but it includes an interesting and provocative specification of the concept of a god.)

Montague, Richard Merett (1930-71)

Richard Montague was a logician, philosopher and mathematician. His mathematical contributions include work in Boolean algebra, model theory, proof theory, recursion theory, axiomatic set theory and higher-order logic. He developed a modal logic in which necessity appears as a predicate of sentences, showing how analogues of the semantic paradoxes relate to this notion. Analogously, he (with David Kaplan) argued that a special case of the surprise examination paradox can also be seen as an epistemic version of semantic paradox. He made important contributions to the problem of formulating the notion of a 'deterministic' theory in science.

Much of Montague's work in philosophical logic consisted in bringing the resources of possible worlds semantics to bear on metaphysical and logical problems. He pioneered the idea of generalizing possible worlds semantics to include tense logic and the logic of demonstratives, in which an expression is evaluated for truth not just at a possible world, but at an index consisting of a world, a time, a location, an addressee, and so on. He called this general framework 'pragmatics'. Constraining the indices in various ways yields traditional logics: limiting them to times alone yields traditional tense logic, limiting them to possible worlds alone yields traditional modal logic, and taking them without limitation yields a general kind of logic in which one may capture the interrelations, for example, of tense, modality and demonstratives. As part of this enterprise he studied different ways of formulating nonextensional notions, examining the connections between languages in which modal notions are captured by modal operators and languages in which these same notions are captured by explicit characterization of propositions. In the former, the sentence 'Necessarily, snow is white' has the form ' $\Box(\text{Snow is white})$'; in the latter it has the form ' $\exists p(p \ = \ \text{that} - \text{snow} - \text{is} - \text{white} \ \& \ p \ \text{isnecessary})$'.

Montague espoused the metaphysical reduction of problematic philosophical entities to constructions out of the metaphysical ingredients needed for possible worlds semantics. He assumes that properties are functions that map worlds to sets of entities, mapping each world to the set of entities that have the property in that world. He suggests that pains may be reduced to relations between persons and moments of time, that events may be reduced to properties of moments (or intervals) of time, and so forth. These reductions are to be justified by their logical fruit; for example, we can analyse 'an event of Caesar's running happened at *t*' as 'the property of being a time at which Caesar runs is possessed by *t*'. He extends this idea to mass terms - holding, for example, that 'water' denotes the property of being a body of water.

Montague's most influential work lies in the semantics of natural language. In 'Universal Grammar' (1970), he formulated a quite abstract, general framework for semantics, and in 'English as a Formal Language' (1970) and 'The Proper Treatment of Quantification in Ordinary English' (1973), he gave particular semantical systems that represent fragments of English.

'Proper Treatment of Quantification' contains the suggestion, contra Russell, that denoting phrases do have meanings of their own, but that they stand not for individuals but for intensions of sets of properties of individuals. An intension of this sort is a function that picks out from each possible world a set of properties, not necessarily the same set in every world. For example, 'no giraffe' stands for the function that maps each world to the set that contains each property of individuals that is not possessed by any giraffe in that world; then, 'No giraffe runs' is true in a world *w* just if the property of running is in the set to which that function maps *w*.

He used this idea to address the 'opaque' readings of transitive verbs, such as 'seek', which may take as objects terms which do not have an actual referent. Simplifying somewhat, the denotation of 'seek' (on this reading) is a relation between individuals (the seekers) and intensions of sets of properties, the sorts of intensions that will be denoted by a direct object of the verb 'seek'. With no additional constraints, this permits 'seek a unicorn' to be true of individuals who 'seek unicorns' without there actually being any unicorns; the seekers of unicorns in a world *w* are those individuals who are appropriately related (in *w*) to the intension that in every possible world picks out the set of properties possessed by some unicorn in that world. (The appropriate relation is not further defined; it is to be given by our understanding of the meaning of 'seek'.) An ordinary nonopaque verb has the same kind of logical form, but special principles (called 'meaning postulates') reinstate traditional inferences for them. For example, let *find** be the denotation of 'find', so that *find** is a relation between individuals (the finders) and intensions of sets of properties. Then a meaning postulate for the verb 'find' would be:

There is an ordinary relation, *find*, between individuals such that necessarily *find** relates (in world *w*) an individual *i* to an intension *P* of a set of properties just if the property of being a thing such that *i finds* it is a member of the set that *P* picks out in world *w*.

It is a consequence of this postulate that if you find a unicorn, then there is a unicorn that you find, but this consequence will not follow without the meaning postulate. Since there is no similar meaning postulate for 'seek', you may seek a unicorn without there being a unicorn that you seek.

These essays, particularly 'Proper Treatment of Quantification', led to a tradition dubbed 'Montague Grammar', in which one studies natural language by giving a rigorous semantics for a formal system that resembles a fragment of a natural language; much fruitful work has been pursued within this tradition, often based on some version of the 'pragmatic' model theory discussed above. This has been an influential element in the development of semantical work by theoretical linguists, as well as the popularity among philosophers and linguists of appealing to possible worlds in informal semantical accounts.

See also: Intensional logics; Modal logic; Modal logic, philosophical issues in; Modal operators; Possible worlds; Semantics, possible worlds

TERENCE PARSONS

List of works

Montague, R.M. (1974) *Formal Philosophy: Selected Papers of Richard Montague*, ed. R.H. Thomason, New Haven, CT, and London: Yale University Press.(Contains most of Montague's influential writings in philosophy and semantics. Formidable reading for the layman.)

Montague, R.M. (1970) 'Universal Grammar', *Theoria* 36: 373-98.(Formulates a quite abstract, general framework for semantics.)

Montague, R.M. (1970) 'English as a Formal Language', in R. Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, CT, and London: Yale University Press, 188-221.(Develops a formal semantics for significant fragments of a natural language.)

Montague, R.M. (1973) 'The Proper Treatment of Quantification in Ordinary English', in R. Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, CT, and London: Yale University Press, 1974, 247-70.(Also known as 'PTQ'. Develops a formal semantics for significant fragments of a natural language.)

Hintikka, J., Moravcsik, J. and Suppes, P. (eds) (1973) *Approaches to Natural Language*, Dordrecht: Reidel. (Contains Montague's only work on mass terms as well as comments by others on the proper treatment of quantification.)

References and further reading

Partee, B. (1989) *Possible Worlds in Model-Theoretic Semantics: A Linguistic Perspective*, in S. Allen (ed.) *Possible Worlds in Humanities, Arts, and Sciences: Nobel Symposium 65*, Berlin and New York: de Gruyter. (Sketches the fruit of Montague's ideas in linguistic semantics and contains useful references.)

Montaigne, Michel Eyquem de (1533-92)

Montaigne was a sixteenth-century French philosopher and essayist, who became known as the French Socrates. During the religious wars between the Catholics and the Protestants in France, he was a friend and adviser to leaders of both sides, including the Protestant leader Henri de Navarre, who converted to Catholicism and became King Henri IV. Montaigne counselled general toleration for all believers, a view promulgated by the new king in the Edict of Nantes (1598). His main literary work was in the form of essais *(a word originally meaning 'attempts'), or discussions of various subjects. In these he developed various themes from the sceptical and Stoic literature of antiquity, and in his unique digressive way presented the first full statement in modern times of Pyrrhonian scepticism and cultural relativism. In particular, he presented and modernized the ancient sceptical arguments about the unreliability of information gained by the senses or by reason, about the inability of human beings to find a satisfactory criterion of knowledge, and about the relativity of moral opinions. His advocacy of complete scepticism and relativism was coupled with an appeal to accept religion on the basis of faith alone. His writings became extremely popular, and the English translation by John Florio, first published in 1603, was probably known to Shakespeare and Francis Bacon. Montaigne, whose essays provided the basic vocabulary for modern philosophy written in vernacular languages, was one of the most influential thinkers of the Renaissance, and his works are regarded as classics of literature and philosophy.*

1 Life

Michel Eyquem de Montaigne was born into a very well-to-do family. His father, Pierre Eyquem, was a Catholic merchant and his mother, Antoinette de Louppes, came from an important Jewish family which had fled to southern France from Spain towards the end of the fifteenth century. Montaigne studied at the Collège de Guyenne in Bordeaux, an elite institution that had been recently set up by Portuguese New Christian refugees (that is, Jewish converts to Christianity) to educate the new middle-class students in the city. (Montaigne's distant cousin, Francisco Sanches, the physician and sceptical philosopher, attended the same school, but not at the same time.) The teaching staff included the Scottish poet and theological leader, George Buchanan, and some of the leading Portuguese humanistic scholars of the period who had studied at the University of Paris. After being educated at the Collège, Montaigne may also have studied briefly at the University of Toulouse (where Sanches became a famous professor). Montaigne studied law, and became a magistrate and a counsellor to the *parlement* of Bordeaux. After a lengthy retreat from public life, followed by a period of travel to Switzerland, Germany and Italy, described in his *Journal de Voyage*, he was Mayor of Bordeaux from 1581-5. During the religious wars which divided France for a good part of the sixteenth century, Montaigne was a friend of many of the leading personalities of the Reformation and the Counter-Reformation in France, including Henri de Navarre, the Protestant leader who became the Catholic King Henri IV.

Montaigne's first work was a letter portraying the death in 1563 of his very close friend, Etienne la Boétie, who had written against slavery and for toleration, views Montaigne continued advocating after his friend's death. In 1569 he published a translation of the *Theologia Naturalis sive Liber Creaturarum (Natural Theology or Book of Creatures)*, a treatise by the Spanish theologian Raymond Sebond (Ramon Sibiuda, d. 1436), who had taught at the University of Toulouse, and who claimed that almost all of the Christian dogmas could be proved by rational means. Montaigne's father, who thought highly of the work, had requested that his son undertake this task. It was during his period of retreat from public life (1571-80) that Montaigne began writing on a variety of subjects. He presented his ideas in the form of *Essais*, and he was one of the first persons to introduce this literary style of expression. He was also one of the very first thinkers to write on philosophical subjects in French rather than in Latin, the standard language for philosophy and theology for centuries. In so doing, Montaigne had to create a philosophical vocabulary in French. His essays are usually very digressive, starting with a quotation, or a thought, and rambling on in different directions to related subjects. The first version of his *Essais* was published in 1580, and he added to and revised them in later editions. The final and most complete edition appeared posthumously in 1595.

2 Scepticism and fideism

Montaigne's philosophy has to be gleaned from various of the essays, for he was not a systematic thinker who

presented his views in didactic form. His longest essay by far, the 'Apology for Raymond Sebond' of which large parts were written 1575-6, is the most extended treatment of his philosophy. It was composed shortly after Montaigne had read the works of the ancient Greek sceptic, Sextus Empiricus, whose writings had been published in Latin in 1562 and 1569. At the same time he was undergoing a personal sceptical crisis, in which he found that everything he had previously believed was in doubt. Montaigne carved mottoes from Sextus' text in the rafter beams of his study, so that if he was attacked by any dogmatic thoughts, he could lean back and regain his sceptical outlook by looking at the ceiling.

The 'Apology' begins by saying that many readers, especially women, had asked Montaigne to explain and defend Sebond's views in the face of criticisms levelled against them. Montaigne then offered as the defence of Sebond that one cannot expect his reasons to be convincing when nobody else's reasons are defensible, and when nobody is able to achieve certainty through reasoning. In order to show this, Montaigne then proceeded to develop in a gradual manner the many kinds of problem that make people doubt the reliability of human reason. He pointed out that human beings believe that they are able to understand the cosmos without the aid of Divine Light, and that they alone of the creatures on the planet can comprehend the world which they think was made for their own benefit. But if human beings are compared with the animals, it is obvious that they have no special faculties that are lacking among the animals. The vaunted rationality of the human being is just a form of animal behaviour. To show this, Montaigne cited materials from Sextus Empiricus, including the story (attributed to Chrysippus) of the dog who is supposed to have worked out a disjunctive syllogism. Montaigne further contended that even religion appears in animals as well as people, citing evidence that elephants seem to pray.

The comparison of animal abilities to those of humans presumably creates doubt about human intellectual pretensions. Montaigne contrasted the glories of the animal kingdom with the vain, stupid, immoral activities of human beings. Our so-called knowledge has not helped us to create a better world or to solve our problems. In spite of all our learning we are ruled by bodily demands and passions. What we consider our wisdom is just presumption which achieves nothing for us. 'The plague of man is the opinion of knowledge. That is why ignorance is so recommended by our religion as a quality suitable to belief and obedience' ([1563-92] 1957: 360). Wisdom has never actually helped anyone. On the other hand, the recently discovered natives of Brazil, the noble savages, manage to live in admirable simplicity, uneducated, with neither laws nor kings nor religion. Each society's laws are just the product of customs, with no genuine basis in reason. People are Christians in France and Muslims in Turkey by the accident of where they were born. So, according to Montaigne, one should take seriously the Christian message that we should cultivate complete ignorance and believe by faith alone. 'The participation that we have in the knowledge of truth, whatever it may be, has not been acquired by our own powers. God has taught us… his admirable secrets. Our faith is not of our acquiring, it is a pure present of another's liberality' ([1563-92] 1957: 369). Montaigne then quoted St Paul (I Corinthians 1) to the effect that God 'will destroy the wisdom of the wise, and bring to nothing the understanding of the prudent', and will save those who believe 'by the foolishness of preaching'.

After this introduction to scepticism and advocacy of complete fideism, Montaigne turned to the arguments from Sextus Empiricus about Pyrrhonian scepticism, and explained their value for religion. The Pyrrhonists, he insisted, are not negative dogmatists like the Academic sceptics. The Pyrrhonists suspend judgment about all propositions or assertions, even the one that says all is subject to doubt. They oppose all knowledge claims, and if their opposition has merit, it shows the ignorance of the opponent. Alternatively, if it has no merit, then it shows the ignorance of the Pyrrhonists. In either case, human ignorance must be acknowledged. Once the state of complete doubt is reached, these sceptics then live according to nature and custom. This state is the finest human achievement, and is the one that is most compatible with religion, for the human being becomes 'a blank tablet prepared to take from the finger of God such forms as he shall be pleased to engrave on it' ([1563-92] 1957: 375) (see Pyrrhonism).

These ancient sceptics had not only reached the summit of human wisdom, they also unknowingly provided a defence for Catholicism against the Reformation. If a person has no positive views, they cannot have wrong views. The sceptics, accepting the laws and customs of their society, would naturally accept Catholicism (as the customary religion of sixteenth-century France), and would have no reason for changing their faith. Moreover, the complete sceptic is in the perfect state to receive whatever revelation God cares to give.

3 Philosophical scepticism

Montaigne contrasted his monumental picture of Pyrrhonism with the deplorable history of dogmatic philosophers whose endless disputes and heterodox views exhibit nothing but human stupidity and credulity. In the best tradition of Renaissance humanists, Montaigne cited the vast range of opinions of ancient thinkers from Greece and Rome that had recently been discovered (see Humanism, Renaissance §1). In view of the enormous diversity of points of view, can one really determine what to believe or to accept as true? In all fields of inquiry the dogmatists have finally had to confess their ignorance, and their inability to come to definitive and unquestionable conclusions. Indeed, even the Pyrrhonists have fallen into the trap of asserting that they doubt, when what they need is a negative non-assertive language. It is better to question, and ask 'What do I know?' ('Que sçay-je?'), a phrase that Montaigne adopted as his motto (see Scepticism, Renaissance §5).

After surveying all kinds of philosophy, from ancient times to the present, Montaigne concluded that 'philosophy is but sophisticated poetry' ([1563-92] 1957: 401). He also used the information presented by explorers ancient and modern concerning the great variety of customs and behaviour in different parts of the planet in order to raise the question of whether there was any way of determining right or true standards, or whether one had to recognize that customs and behaviour just are relative to the cultures in which they occur.

Science fared no better than philosophy and morality. What thinkers have been offering through their theories are just human inventions, and philosophers never find out what actually happens in nature. What occurs is that some traditional opinions have been taken as authoritative, indubitable principles. Questioners are told that one cannot dispute with people who deny first principles, but, Montaigne observed, 'there cannot be first principles for men, unless the Divinity has revealed them; all the rest - beginning, middle, and end - is nothing but dreams and smoke' ([1563-92] 1957: 404). Even the new scientists of the Renaissance such as Copernicus or Paracelsus are just offering personal opinions that will probably be replaced by other people's opinions at some future time.

Having built up a general sense of why one should be a sceptic, Montaigne finally turned to the fundamental philosophical reasons for accepting the Pyrrhonian view that all is in doubt. These concern the unreliability of information gained by the senses or by reason, and the inability of human beings to find a satisfactory criterion of knowledge. Montaigne argued that if the dogmatists claim that human reason can know and understand things, they should show how in fact this occurs. If the claim is that sense experience provides this knowledge and understanding, the dogmatists should make clear what it is that we actually experience and whether we do in fact experience the very things that we think we experience. In fact, the senses are clearly subject to illusion, and we ourselves are constantly changing as our physical and emotional conditions alter. Once it becomes clear that the senses are full of uncertainty, the dogmatists will appeal to reason, but 'no reason can be established without another reason'. The dogmatists will have to provide a criterion of right reasoning, and a criterion of the criterion, and so on *ad infinitum*.

Montaigne suggested that people should suspend judgment on all matters, and then wait until God reveals principles to them. Until then, he said, one should just follow customs, traditions and social rules undogmatically, and one should be tolerant of other people's views. Religious beliefs should be based solely on faith rather than on dubious evidence (see Faith §§4-5).

4 Influence

Montaigne's rambling presentation of Pyrrhonian scepticism quickly became the best known and most influential statement of this view. Francis Bacon, Descartes, and Pascal, among other readers, were greatly influenced by him. His pleasant literary presentation of the sceptical attitude as well as the arguments behind it effectively provided many with a basis for rejecting the entire intellectual world of the time. He also provided reasons for accepting the status quo: his advocacy of accepting customary views because there was no adequate reason to change them became a defence of Catholicism against the Reformation during the seventeenth century in France. Leading churchmen, such as Bishop Jean-Pierre Camus, the secretary of St François de Sales, and Father François Veron, the official defender of the faith under Louis XIV, employed sceptical arguments as 'a machine of war' to defeat the intellectual claims of the Calvinists. On the other hand, Montaigne's leading disciples, Pierre Charron, Marie de Gournay, and François de La Mothe Le Vayer, are usually treated as being *libertins érudits*, possible secret non-believers in any form of Christianity, whose scepticism prefigured that of the Enlightenment, and who were

really trying to undermine religion.

The question of what Montaigne himself actually believed has been debated for the last four centuries with some interpreters insisting that he was trying to undermine all belief, including any religious belief, and other readers insisting that he was a genuine Christian believer who was offering a sceptical defence of Catholicism in the age of the Counter-Reformation.

It is extremely difficult to determine the actual views of someone centuries after he lived, and we have conflicting testimonies from his friends and from early readers. We know that Montaigne came from a family whose views encompassed the main religious beliefs of the time. His mother was half Jewish, and it has been suggested by Donald Frame that his tolerant, cosmopolitan, humanistic attitude may in part have been derived from his Jewish heritage (Frame 1965: 28). His father was a Catholic, and he had Protestant siblings. He was related to leading Calvinists in Belgium as well as to some important Catholics in France. He reported in his *Journal de Voyage* that when he visited Rome the Catholic censors seemed quite pleased when they read the manuscript of the first draft of the *Essais*. At the time they actually invited the author to stay in Rome and to 'assist the Church with [his] eloquence'. (The *Essais* were put on the Index late in the seventeenth century, principally because of Montaigne's suggestion that elephants engage in prayer.) Montaigne turned down the invitation, preferring retirement at his chateau near Bordeaux where he could gently muse about the human comedy, and discuss it in his many essays. He said at one point that in the essays 'I paint myself'.

What kind of a self was thereby indicated? Commentators have offered all sorts of readings. Without going into details about these many kinds of interpretations, it can be said that the sceptical fideism that is offered in the 'Apology for Raymond Sebond' is compatible with both a religious and an irreligious evaluation of the author's intentions, as either defending or undermining Christian belief. Essentially Montaigne's avowed fideism rests upon a *non sequitur*, namely that since all is in doubt, we should therefore accept Christianity on faith alone. This has been said by extremely religious persons like Pascal and Kierkegaard. And it has been said, presumably ironically, by agnostic or irreligious thinkers like Hume. In evaluating people who make this fideistic assertion we usually appeal to other information about their character and their activities in order to ascertain whether or not they are being sincere. Montaigne does not appear to have been particularly devout or fervent in his religious life, and his own acceptance of Catholicism seems tepid. He was generally interested in the variety of religious experience, ancient and modern, but he appears to have been indifferent to the major spiritual tendencies of his day, and was much more concerned with creating a peaceful social world in which all people could believe and practice what they wished. He was active in trying to bring about toleration of different religious practices and views, and he opposed and ridiculed different kinds of religious fanaticism and superstition. Pascal, who was very much influenced by reading Montaigne's *Essais*, considered him a sceptical non-believer who exhibited the misery of man without God. However, Montaigne himself said that to philosophize is to learn to die, and his sceptical outlook may have prepared him to live undogmatically in a troubled world, and to accept whatever might come afterwards.

See also: Humanism, Renaissance; Political philosophy, history of §7; Scepticism, Renaissance

RICHARD H. POPKIN

List of works

Montaigne, M. de (1563-92) *Oeuvres complètes (Complete Works)*, ed. A. Thibaudet and M. Rat, Paris: Bibliothèque de la Pléiade, 1962.(Standard edition of Montaigne in French.)

Montaigne, M. de (1563-92) *The Complete Works of Montaigne*, trans. D.M. Frame, Stanford: Stanford University Press, 1957.(Includes Montaigne's letters, as well as the *Essais* and *Journal de Voyage*; contains an annotated bibliography and fine introduction.)

References and further reading

Boase, A.M. (1935) *The Fortunes of Montaigne: A History of the Essays in France, 1580-1669*, London: Methuen.(A study of the impact and influence of Montaigne's writings.)

Brush, C.B. (1966) *Montaigne and Bayle. Variations on the Theme of Skepticism*, The Hague: Martinus Nijhoff. (A study and comparison of two important sceptics.)

Busson, H. (1957) *Le rationalisme dans la littérature française de la Renaissance (1533-1601) (Rationalism in*

French Literature of the Renaissance), Paris: Vrin.(An interpretation of Montaigne as a free-thinker, and a participant in rationalist irreligious currents.)

Dréano, M. (1936) *La pensée religieuse de Montaigne (Montaigne's Religious Thought)*, Paris: Beauchesne; revised edn, *La religion de Montaigne*, Paris: Nizet, 1969.(Portrays Montaigne as a sincerely religious thinker.)

Frame, D.M. (1955) *Montaigne's Discovery of Man. The Humanization of a Humanist*, New York: Columbia University Press.(A study of the development of Montaigne's thought.)

Frame, D.M. (1965) *Montaigne: A Biography*, New York: Harcourt Brace.(Biography by leading Montaigne scholar.)

Hallie, P.C. (1966) *The Scar of Montaigne, an Essay in Personal Philosophy*, Middleton, CT: Wesleyan University Press.(Important interpretation of Montaigne's philosophy.)

Laursen, J.C. (1992) *The Politics of Skepticism in the Ancients, Montaigne, Hume and Kant*, Leiden, New York and Cologne: Brill.(Tries to connect Montaigne's scepticism with his political actions.)

Malvezin, T. (1875) *Michel de Montaigne, son origine, sa famille (His Origins and his Family)*, Bordeaux: Lefebvre.(Deals with Montaigne's family background and his milieu.)

Popkin, R.H. (1979) *The History of Scepticism from Erasmus to Spinoza*, Berkeley, CA: University of California Press.(Places Montaigne in the course of Renaissance scepticism.)

Schiffman, Z.S. (1984) 'Montaigne and the Rise of Skepticism in Early Modern Europe: A Reappraisal', *Journal of the History of Ideas* 45: 499-516.(Attempts to trace Montaigne's scepticism to his early education before he came in contact with Pyrrhonian scepticism.)

Shaefer, D.L. (1990) *The Political Philosophy of Montaigne*, Ithaca, NY, and London: Cornell University Press. (Argues that Montaigne is best understood from the perspective of political philosophy and that he anticipates classical political liberalism.)

Villey, P. (1908) *Les sources et l'évolution des Essais de Montaigne (The Sources and Development of Montaigne's Essais)*, Paris: Hachette.(The basic study of the sources used by Montaigne, showing how they were employed in the development of the *Essais*.)

Montesquieu, Charles Louis de Secondat (1689-1755)

Montesquieu, one of the greatest figures of the Enlightenment, was famous in his own century both in France and in foreign lands, from Russia to the American colonies. Later generations of French philosophes *took for granted his concern to reform the criminal laws, to replace the Inquisition with a reign of tolerance, and to repudiate the vicious conquests of the Spaniards in the Americas. They also accepted his finding that Protestant, commercial, and constitutionalist England and Holland represented all the best possibilities of Europe; whereas Catholic, economically backward, and politically absolutist Portugal and Spain represented the worst of the Western world and constituted a warning to the French.*

Although the findings and specific reforms proposed by Montesquieu were repeated by many another figure of the French Enlightenment, his work in certain respects remained unique in the circles of the most advanced thinkers. In his efforts to think systematically about politics and to do so by employing the comparative method, he stands virtually alone in his age. Other thinkers sharing his commitments resorted to the universalizing language of natural rights when they ventured into the realm of political philosophy. Or, like Voltaire, they tied their thoughts about politics to a succession of specific issues, each essay bearing so indelibly the imprint of specific time and place that there was no room for theory in their writings. Finally, as is true of Diderot or D'Alembert, many of the philosophes *were slow to recognize what Montesquieu knew from the outset, that if Enlightenment does not extend to politics it is futile.*

Steeped in Montaigne's scepticism, Montesquieu found that in the absence of absolutes there were good reasons to appreciate the 'more than/less than' and 'better than/worse than' judgments of comparative analysis. In his notebooks he commented that the flaw of most philosophers had been to ignore that the terms beautiful, good, noble, grand, and perfect are 'relative to the beings who use them'. Only one absolute existed for Montesquieu and that was the evil of despotism, which must be avoided at all costs.

Montesquieu wrote three great works, each teaching lessons about despotism and freedom,The Persian Letters (1721), the Considerations of the Grandeur of the Romans and the Cause of Their Decline (1734), and The Spirit of the Laws (1748).

1 *The Persian Letters*

A year before his death Montesquieu wrote a new preface to *The Persian Letters* (first published in 1721). Looking back, he recommended that the letters be read not individually but in their interconnections, since they are held together by a 'chain'. The letters form 'a kind of novel'; they tell a story and that story is the meaning of the book. Hence he repeatedly refused requests to add new letters, because more letters would signify less meaning.

As the story begins we learn that Usbek, a Persian, is travelling in the West with his countrymen Rica and Rhedi. The letters we read are exchanges among these tourists or between them and the persons they have left behind; many of the letters are the correspondence of Usbek with his wives or with the eunuchs who govern the household in his name during his absence. Between the first letter and the last, some nine years pass, during which the Persians learn much about Europe. The more they understand a foreign culture, the more they gain insight into their own as well. One cannot imagine a more entertaining or convincing argument for the comparative method than that which Montesquieu sets forth by example in *The Persian Letters*.

As Montesquieu indicated in 1754, his strategy was to portray the 'genesis and development' of the ideas of the Persians. Initially the Oriental visitors comprehend little of what they observe, so their earliest commentaries afford Montesquieu ample opportunity to satirize Western ways, as when Rica naïvely writes that the Pope is a magician who has the king believing that three are one and that bread is not bread, wine not wine. Eventually the Persians become so familiar with Western mores, customs, science and philosophy that the 'otherness' of Europe fades into their pasts. At the same time, however, their native Persia seems ever more foreign and even hideous from the standpoint of their transformed beliefs. When he first stepped foot on European soil in Italy Usbek sensed, 'even in the most insignificant of details, something that I feel and cannot express'. Later he and his comrades come to realize that it is freedom they have encountered in their new environment, and that for them to utter the word 'freedom' in reference to the West is inevitably to speak of 'despotism' when designating their own

country.

Virtually alone in his circle, Montesquieu recognized that to hold enlightened beliefs is not necessarily to be a tolerant person. Who could be more of a *philosophe* than Usbek, author of the most intellectually sophisticated of the Persian letters? And who could have better intentions than this same Usbek who dared to carry truth to the court of the Persian king, in consequence of which his days would have been numbered had he not found an excuse to visit foreign countries? Yet Usbek, after writing a series of letters diagnosing the malady of Oriental despotism, turns around and demands that his eunuchs initiate a reign of terror against his wives.

Jealousy is Usbek's undoing, a jealousy that must not be mistaken for the reverse side of love, because, as Usbek admits in an unguarded moment, he is incapable of loving his wives. In Persia women are the property of men; and although a man may fear losing his woman, exactly as he is disturbed by the loss of any of his prized possessions, it is impossible for him to love a creature who by the very nature of the mores and laws is not fully a human being. Near the end of the novel Usbek, provoked to hear of growing disorder in his seraglio, decides to return to Persia in order to reclaim his possessions, and in the meantime he deputizes the chief eunuch to rule his wives by fear and terror.

Far from an island of peace and contentment, the Persian family is a replication in miniature of the most insidious characteristics of the surrounding society and polity. Where Oriental despotism reigns, the family, too, is despotic. It is the curse of arbitrary power that it infects all human relationships, affairs of the heart no less than affairs of state.

Self-deception is a recurring theme in the novel. Usbek has come to know Oriental despotism for what it is, but he lacks the strength to apply the lessons he has learned to his personal relations. The wives are even more determined to avoid the truth. Slaves by the laws of their country, they prefer to regard themselves as enslaved by their love of Usbek. Only the eunuchs stand outside the web of self-deception; they, in order to feel less diminished by castration, dominate the women by self-consciously manipulating oppressive Persian notions of virtue.

The story ends with the suicide of Roxane, the one woman who mattered to Usbek. By holding herself aloof and refusing to be treated as property, she aroused Usbek's ardour. During his absence Roxane avenges herself by taking a lover and dies by her own hand rather than submit to Usbek's despotic rule. Her act proves a truth Usbek himself had enunciated, that under despotism everyone loses, the despot as well as his long-suffering subjects. It was Usbek who proclaimed that systematic injustice is self-destructive, little realizing that he would one day demonstrate his point at his own expense.

2 The *Romans*

Montesquieu's monograph on the Romans, *Considerations of the Grandeur of the Romans and the Cause of Their Decline* (1734), usually figures in modern scholarship as a warming up exercise for the causal reasoning found in his magnum opus, *The Spirit of the Laws* (1748). Such an interpretation is not incorrect but it is wanting in so far as it permits his third book to define the meaning of his second, as if his *Considerations on the Romans* had no significance in its own right. One might do better to see his second major publication as a variation on the theme of his first book. Previously he had argued that because of its internal injustice Oriental despotism is inherently self-destructive. Now he adds that the ancient Roman republic, though sometimes reasonably just internally, destroyed itself through the unrelenting external injustice of its foreign policy.

Montesquieu's Romans knew 'not even the justice of brigands'. After inflicting military defeat upon a prince, they exacted excessive war reparations that left him with no choice but to collect exorbitant taxes - 'a new kind of tyranny that forced the prince to oppress his subjects and lose their love'. Similarly, the Romans destroyed Carthage, 'saying they had promised to save the people but not the city itself'. Whenever possible the Romans fostered factional strife so as to weaken and eventually conquer foreign cities. 'If princes of the same blood were disputing the crown, the Romans sometimes declared them both kings. If one of them was under age, they decided in his favour'. A city that lost its battle with Rome inevitably found itself compelled to fight with Rome to subdue another people, after which the Romans stripped their erstwhile allies of their remaining freedoms. These practices of the Romans were 'in no way just particular actions occurring by chance; these were ever-constant principles'.

If nothing could sound more Machiavellian than Montesquieu's Romans, that is because his account of ancient history is taken directly from the great Florentine's *Discourses on Livy* (see Machiavelli, N.). Why not follow the lead of Machiavelli in turning the Roman senators into the most Machiavellian of ruling classes if to do so is the best way to discredit power-politics? Each Roman conquest, Montesquieu insisted, brought Rome that much closer to the day when the republic would perish, and Rome would begin to terrorize itself in much the same manner as it had long terrorized other city-states.

The more Rome expanded, the less it could sustain the integrity of its republican way of life. Civic mindedness yielded to civic neglect; private pursuits replaced public concerns. Frugality gave way to luxury and ostentation when returning soldiers carried plunder and foreign mores back to Rome. Removed from Rome for ever longer campaigns, soldiers no longer identified with their city but rather with their commander who offered them riches in return for their assistance in elevating him to political power.

Faced with evidence that the ruthlessness of the senators had added the name of Rome itself to the list of republics undone by Roman imperialism, Machiavelli hastened to place the blame on *fortuna*, which eventually undoes even the best laid plans. Montesquieu countered by denying that chance had anything to do with the demise of the republic. What was at work, he insisted, was nothing less than the operation of causal necessity. Any republic which grows beyond a certain point necessarily undermines its civic culture. Since the republic had to perish, it was only a question of how, and by whom, it was to be overthrown.

> It is not chance that rules the world…There are general causes, moral and physical, which act in every [regime], elevating it, maintaining it, or hurling it to the ground.
>
> (*Considerations on the Romans*, chaps 11, 18)

Montesquieu's objective in the *Considerations* was to set forth a causal argument that eliminated not just *fortuna* from historical explanation but Providence as well. Bishop Bossuet had preceded Montesquieu both in accepting Machiavelli's reading of Roman history and also in drawing the conclusion that the Roman republic destroyed itself by the systematic injustice of its foreign policy. Dedicated to vindicating the Christian view of history, Bossuet in the *Discourse on Universal History* (1681) deliberately set out to expunge pagan *fortuna* from the historical record. To that end he announced that 'the true science of history' was one in which the investigator identified the underlying causes which account for the revolutions of empires. Those causes turn out to be much the same as the ones cited later by Montesquieu, especially the decline of civic life and the demise of the large, frugal, agrarian middle class which, as Aristotle noted, was the backbone of a popular republic.

When Montesquieu approached Roman history he found that Bossuet had prepared the way for his *Considerations*. Specifically, the good bishop had driven *fortuna* off the historical stage and effectively pleaded the case for causal reasoning. All Montesquieu had to do to arrive at a historical science fit for the Enlightenment was to bracket off Providence as an unnecessary hypothesis, and with that act he had moved to a stage of thought which was both post-pagan and post-Christian.

3 The *Spirit of the Laws*

The centrepiece of Montesquieu's most comprehensive and systematic treatise is its typology of sociopolitical forms. Monarchy in politics is matched in society by an ethos revolving around the 'principle' of aristocratic honour; Oriental despotism is characterized by the principle of fear; and the ancient republic was enlivened by civic virtue.

By 'monarchy' Montesquieu understands feudal government in its final incarnation during the Old Regime. Absolute in theory, the power of the king is thwarted in fact by 'intermediary bodies' - the first and second estates (clerics and nobles) and the *parlements*, which were judicial institutions, not legislative bodies. Corporate and hierarchical, the social order is so thoroughly under the sway of the privileged estates that the members of the middle class, far from challenging the aristocrats, want nothing more than to amass enough wealth to buy a noble title and to pay someone - as did Molière's bourgeois gentleman - to teach them how to put on aristocratic airs.

Voltaire believed that *The Spirit of the Laws* was in reality nothing more than an apology for the privileged classes to which Montesquieu, a noble and former member of the Bordeaux *parlement*, belonged. The ancient republic was something lost and gone forever, and Oriental despotism was not only the worst of all possible worlds but a

constant reminder that to remove the intermediary bodies from a monarchy is to invite disaster. Hence, Voltaire concluded, Montesquieu was a formidable apologist for all the abuses of the Old Regime. If Montesquieu championed the English example, Voltaire assumed the hidden agenda was to suggest that power be taken away from the French monarch and turned over to nobles seated in a legislature.

Modern historians, Marxists and all those who wish for a ready-made link between social and intellectual history, have repeated Voltaire's charges. It is therefore crucial to note that other *philosophes* refused to follow Voltaire's lead, and to ask why. The best place to begin our reexamination of *The Spirit of the Laws* is with the recognition that, in its pages, England figures not only as the most free country in the world but also as the only monarchy that no longer possesses intermediary bodies, whereas Spain is portrayed both as the nation in which the intermediary bodies could not be more powerful and as a country plagued by a peculiarly Western variety of despotism.

'Abolish the privileges of the lords, the clergy, and cities in a monarchy', wrote Montesquieu, 'and you will soon have a popular state, or else a despotic government'. Although in despotism Montesquieu found nothing but evil, he did have considerable sympathy for the popular state that had emerged from the ashes of a feudal past. 'The English, to favour their liberty, have abolished all the intermediate powers of which their monarchy was composed.' During the upheavals of the seventeenth century, the remnants of feudalism perished in England, with the consequence that in Montesquieu's day the English had a monarch but no Old Regime. England was the first new nation.

Call England a monarchy and Montesquieu will not object, for such it once was and such it still is on the surface. But on a closer look England is much better described as a 'republic hiding under the form of a monarchy'. Unlike the republics of antiquity, but like Holland across the Channel, England is a commercial republic. Politically, however, there is a significant difference between Holland and England: the Dutch live in a confederate republic, the English in a centralized republic.

As a result of purging the feudal 'intermediary bodies' England is, at one and the same time, the most free country and the one in which freedom is in greatest danger of declining into its opposite. No sooner has Montesquieu applauded English liberty than he issues a stern warning: the English 'have a great deal of reason to be jealous of this liberty; were they ever to be so unhappy as to lose it, they would be one of the most servile nations upon earth'. Under the conditions of a post-feudal world, where power is centralized as never before, only a 'system' of liberty will suffice, a constitutional structure wherein there is a sharp separation of legislative, executive and judicial functions. In the famous Book XI, chapter 6, Montesquieu outlines his proposal for a system of freedom and deems it especially appropriate for the British; in Book XIX, chapter 27, he makes it clear that the spoils system (of Robert Walpole) is what actually exists in England.

Whatever his misgivings about England and Holland, Montesquieu was convinced that they were not only the most free but also the most powerful nations. Each age, he remarked in his notebooks, has its spirit, and the spirit of the modern age is that of commerce. Conquests weaken the conqueror, trade strengthens the most economically progressive nations. Both Holland and England are small; neither has much of a standing army; and yet these two commercial nations may well be the most formidable powers of Europe.

Altogether different is the situation of once proud Spain. Fields lie uncultivated, trade is nonexistent, and the Spaniards sit by idly while 'the rest of the European nations carry on in their very sight all the commerce of their monarchy'. It is the very power of the 'intermediary bodies', the nobility and the clergy, that accounts for the increasing powerlessness of the nation which in early modern history had been the leading country in Europe. The Church does nothing with its ever growing holdings of land, nor do the nobles produce anything since the aristocratic code of honour forbids titled persons to engage in commercial activity.

Enrichment through plunder, conquest, and murder is 'honourable'; indeed, it is holy as well when combined with forcible baptism. Thus the Church blessed the bloody conquests of greedy and cruel Spanish nobles in the Americas, as did the state which, following a mercantilist policy, was only too happy with the mounds of gold the conquistadores shipped back to the old world. Which is to say, the Spaniards mistook a symbol of wealth for wealth itself. As the quantity of gold in the king's coffers piled up, the value of gold declined: 'Spain behaved like the foolish king who desired that everything he touched might be converted into gold, and who was obliged to beg

of the gods to put an end to his misery' (*The Spirit of the Laws*, XXI, ch. 22). Well in advance of Adam Smith, Montesquieu declared that the wealth of nations depended on the production of goods and services.

A regressive economy was one consequence of the influence of Spain's powerful intermediary bodies; the Inquisition was another. That the French government might well follow Spain in permitting the Catholic Church to dictate oppressive public policies, even at the cost of undermining the power of the state, was indicated, Montesquieu believed, by the Revocation of the Edict of Nantes (1685). The Huguenots sent into exile by Louis XIV were the most able economic producers in France; in eliminating them Louis inflicted serious damage on the economic resources without which there can be little political power. Let the heirs of Cardinal Richelieu talk as much as they wish about reason of state; policy will never be reasonable when the king's confessor advocates the destructive measures dictated by the fanatical politics of divine right.

From his earliest writings to his last, Montesquieu was constantly attentive to Spain, the country which was a living example of what form despotism was likely to assume in the Western setting. In the Orient it is the absence of intermediary bodies that makes despotism so devastating. In the Western world it is the very triumph of those same bodies that fosters a less overt and less total despotism but a despotism nonetheless.

When Montesquieu looked to Spain and England he saw countries which, like France, had evolved from the embryo of a feudal past. Both those countries were further down the road of historical development than his native land. Would France duplicate the Spanish example, as Montesquieu feared? Or would it break with its feudal past, as had England? The best possibility, to Montesquieu's mind, was for France to chart its own course, one that would be constitutionalist, assuredly, but more politically decentralized than the British government. How to accomplish this task of reconstruction, he was the first to admit, was far from obvious. Surely the Catholic Church would not readily relinquish its stranglehold in France, nor would the nobility readily learn how to succeed economically or how to share political power. Historical and comparative analysis tell us what our problem is; they provide no easy solutions.

4 Legacy and reputation

The *philosophes* regarded Montesquieu, quite rightly, as an eloquent and penetrating spokesman for the outlook of the Enlightenment. For Diderot and D'Alembert, editors of the *Encyclopédie* (1751-72), it was a great victory for the party of humanity that the illustrious Montesquieu agreed to submit an article on taste to their undertaking in collective and committed publication. Yet most of the *philosophes*, despite their admiration of *The Spirit of the Laws*, denied that climate, a physical rather than a 'moral' cause, played the dominant role in the non-Western world that Montesquieu attributed to it. Frequently, too, the *philosophes* displayed some impatience that Montesquieu spent so much time discussing what is, when they were more interested in what ought to be.

In the nineteenth century Hegel, who acknowledged his intellectual debts, praised Montesquieu as superior to other writers of his age in taking an historical approach to his subject matter and in 'always treating the part in its relation to the whole'. Among noteworthy French authors Benjamin Constant's writings make incessant use of insights and formulations gleaned from Montesquieu; this includes Constant's frequently cited *Liberty of the Ancients Compared with that of the Moderns* (1819). Not only did Constant agree with Montesquieu in championing modern notions of freedom as privacy, but the two writers were alike again in their estimation that the freedom to pursue our individual desires will never be safe if we completely abandon the public and civic conceptions of liberty which prevailed in antiquity.

Finally, Tocqueville must be mentioned. His entire manner of conducting comparative analysis is reminiscent of *The Spirit of the Laws* (1835), and the specific objective of *Democracy in America* (1840), to fathom the political, economic, social, and cultural possibilities of a post-feudal republic, is already present a century earlier in Montesquieu's examination of England.

There are signs that the scholarship of our age is finally removing the onus with which Voltaire burdened Montesquieu's reputation. No longer is it common to see Montesquieu diminished to the lowly stature of an apologist for the privileged orders of his day. Unquestionably, he now figures as a far more profound thinker than Voltaire, and has begun to compete with Diderot and Rousseau for the title of the greatest French thinker of his time. Despite the historical specificity of his method, Montesquieu continues to speak to all persons who adhere to the ideals and aspirations of the Enlightenment.

See also: History, philosophy of

MARK HULLIUNG

List of works

Montesquieu, C. Baron de (1951) *Oeuvres complètes*, ed. Caillois, R. Paris: Pléiade, 2 vols.(Contains the major works.)

(1721) *Lettres persanes*, trans. J.R. Loy, *The Persian Letters*, New York: Meridian Books, 1961.(Includes an introductory essay by Loy.)

(1734) *Considérations sur les causes de la grandeur des Romains et de leur décadence*, trans. D. Lowenthal, *Considerations on the Grandeur of the Romans and the Cause of Their Decline*, Ithaca, NY: Cornell University Press, 1965.(The only English translation.)

(1748) *L'Esprit des lois*, trans. A. Cohler *et al.*, *The Spirit of the Laws*, Cambridge: Cambridge University Press, 1989.

References and further reading

Althusser, L. (1959) *Montesquieu, la politique et l'histoire (Montesquieu, Politician and Historian)*, Paris: Presses Universitaires de France.(A Marxist account making Montesquieu the spokesman for a social class.)

Hulliung, M. (1976) *Montesquieu and the Old Regime*, Berkeley, CA: University of California Press.(A refutation of the efforts of Voltaire and his descendants to reduce Montesquieu to his social class.)

Shackleton, R. (1961) *Montesquieu: a Critical Biography*, Oxford: Oxford University Press.(The standard biography.)

Shklar, J. (1987) *Montesquieu*, Oxford: Oxford University Press.(The best brief overview.)

Starobinski, J. (1953) *Montesquieu par lui-même (Montesquieu by Himself)*, Paris: Editions du Seuil.(Autobiographical jottings and commentary.)

Moore, George Edward (1873-1958)

G.E. Moore was one of the most influential British philosophers of the twentieth century. His early writings are renowned for his rejection of idealist metaphysics and his insistence upon the irreducibility of ethical values, and his later work is equally famous for his defence of common sense and his conception of philosophical analysis. He spent most of his career in Cambridge, where he was a friend and colleague of Russell, Ramsey and Wittgenstein.

The best-known thesis of Moore's early treatise on ethics, Principia Ethica (1903), is that there is a fallacy - the 'naturalistic fallacy' - in almost all previous ethical theories. The fallacy is supposed to arise from any attempt to provide a definition of ethical values. The validity of Moore's arguments is much disputed, but many philosophers still hold that Moore was right to reject the possibility of a reductive definition of ethical values. The book is also renowned for Moore's affirmation of the pre-eminence of the values of Art and Love.

Moore's later writings concern the nature of the external world and the extent of our knowledge of it. In opposition to idealist doubts about its reality and sceptical doubts concerning our knowledge of it, Moore defends 'common sense' by emphasizing the depth of our commitment to our familiar beliefs and criticizing the arguments of those who question them. But although he insists upon the truth of our familiar beliefs, he is remarkably open-minded concerning their 'analysis', which is intended to clarify the facts in which their truth consists.

1 Ethics

Moore studied philosophy at Cambridge University in the 1890s. At this time the idealist philosophy of F.H. Bradley was the dominant influence within British universities, and, under the influence of J.M.E. McTaggart, Moore became a disciple of Bradley. After graduating in 1896, however, Moore began to turn against the idealist tradition in philosophy. His first writings concern the foundations of ethics, and in the course of a critical study of Kant's ethical theory Moore argued that the idealist programme was deeply flawed: it was, he thought, a mistake to suppose that the fundamental principles of ethics require justification by reference to the will, reason or any similar criteria (see Kant, I. §9; Kantian ethics §1). For they concern a domain of values which is as objective as any physical fact, though Moore held that ethical values are not themselves physical or 'natural' facts; in his view the fundamental truths of ethics are abstract necessary truths concerning the intrinsic value of different types of state of affairs. He famously propounded this position in *Principia Ethica* (1903), where he argued that there is a fallacy, the 'naturalistic fallacy', in all theories which offer naturalistic accounts of ethical values and that much the same fallacy afflicts idealist theories of value. Moore assumed that such accounts always involve a definition, or analysis, of ethical value, so the thesis he sought to establish was that goodness, which he believed to be the fundamental ethical value, is indefinable. His main argument was that any supposed definition of goodness (for instance, as that which we desire to desire) can be seen to be incorrect when we recognize that it incorporates a substantial ethical thesis (for example, that whatever one desires to desire is *ipso facto* good). For, Moore maintained, if the original definition had been correct, then there should seem to us to be no more of an 'open question' concerning its truth than there is concerning the result of applying the definition to itself, that is, concerning the thesis that whatever we desire to desire we desire to desire.

As Moore's many critics have observed, this argument is problematic. It is not clear that definitions need to be obviously self-applicable in the way that Moore seems to assume (a point he later acknowledged in his discussions of the 'paradox of analysis'). His conclusion is also vulnerable to revisionists who offer an alternative, perhaps naturalistic, set of ethical concepts. These critics will argue that even if Moore is correct concerning the indefinablity of our concept of goodness, then so much the worse for our concept; we would do better to abandon it in favour of something less metaphysically demanding. Hence Moore's thesis can really only be made secure when it is embedded in a broader metaphysics which elucidates the nature of ethical concepts and explains thereby why they are indefinable. Moore himself offers a Platonist metaphysics of goodness as a simple, non-natural property whose a priori relationships to the natural properties of things we are able to discern through reflective ethical intuitions. Not surprisingly, few of his successors were able to accept this. As a result many of those who accepted his thesis that ethical concepts are indefinable held that the distinctive feature of ethical statements arises, not from any special ethical facts they purport to describe, but from the special role of these statements as expressions of emotion, or as recommendations and prescriptions (see Emotivism; Hare, R.M. §1; Prescriptivism).

This linguistic version of Moore's thesis is no longer popular because it does not appear to do justice to the content of ethical judgments. The reaction against it has not, however, led to a return to Moore's abstract conception of ethical values; instead it is Moore's anti-naturalist argument that is questioned, in particular his assumption that ethical naturalists have to offer a reductive definition of ethical value (see Naturalism in ethics §1). In this case, however, Moore might respond that his thesis (implicit in *Principia Ethica*, explicit in later writings) that the ethical value of a state of affairs supervenes upon its natural properties shows that this naturalist position cannot be correct (see Supervenience §1). For if ethical values are just simple natural properties, it is quite unclear why they should supervene on other natural properties. The ethical naturalist might well reply *ad hominem* that supervenience looks just as much a problem for Moore's conception of ethical value as a simple non-natural property. More deeply, however, the naturalist may want to claim that ethical values have an explanatory role within human psychology through their involvement in reasons for action which explains both their supervenience and their irreducibility. So at this point the debate about ethical naturalism connects with other debates about reductionism and naturalism in the philosophy of mind (see Reductionism in the philosophy of mind).

Although Moore's ethical theory is now primarily studied because of his critique of naturalism, at least two other aspects of it merit attention - his ideal utilitarianism and his emphasis on the value of art and love (see Utilitarianism). Moore's ideal utilitarianism owes much to another of his teachers, Henry Sidgwick, who had argued that utilitarianism, which he endorsed, involves two distinct doctrines: (1) the right action is that which makes the world as good as possible; (2) pleasure is the ultimate good. Moore agreed with Sidgwick on the first point (which is constitutive of ideal utilitarianism) but rejected the second, arguing that once we recognize that pleasure occurs within 'organic wholes' which include an affective response to perceived or anticipated states of affairs we can understand how there are evil as well as good pleasures (see Hedonism §1). None the less, Moore's own supreme goods, art and love, are themselves pleasures, so it is basically a refined hedonism that Moore himself commends; and it was certainly as such that his position was understood by his friends within the Bloomsbury Group (such as J.M. Keynes, Lytton Strachey and Leonard Woolf). Admittedly, in *Principia Ethica* itself Moore maintained that certain states of affairs not involving human consciousness possess intrinsic value; but in subsequent writings he affirmed the thesis, which became characteristic of the Bloomsbury Group, that it is only certain pleasant states of consciousness which have positive intrinsic value.

2 The rejection of idealism

Moore's criticisms of idealist ethical theory are accompanied by similar criticisms of idealist metaphysics (see Idealism). Moore directs his criticisms at three different idealist positions: he argues (1) that Berkeley's thesis that *esse* is *percipi* confuses the 'object' of perception with its 'content' (see Berkeley, G. §3); (2) that Kant's 'Copernican Revolution' wrongly treats a priori truths as if they were dependent upon the nature of human reason; and (3) that Bradley's 'Absolute Idealism' rests upon the assumption that all relations are internal, which itself only reflects confusions concerning identity and difference. For Moore these last two points connect closely with the distinction in whose terms the first is couched - that between the 'content' and 'object' of thought. His rejection of Kantian and Bradleian idealism is founded upon the thesis that the objects of thought are 'propositions', possible states of affairs which, in the case of a true proposition, actually obtain. The existence and truth of such propositions is altogether independent of our thought or knowledge of them; so there is no internal relation between the truth of a proposition and anyone's grasp of it, even in the case of a priori propositions.

Moore's initial reaction against idealist metaphysics led him to an extreme realist position concerning the existence of propositions and sensory objects, on the basis of which he held that traditional sceptical arguments could be easily rejected. But in his 1910-11 lectures *Some Main Problems of Philosophy* (1953) he recognized that matters cannot be quite so simple. He here abandons belief in the existence of propositions because he is no longer able to accept the existence of false ones (which, on his view, would be non-actual states of affairs). He also abandons his previous naïve realism concerning sensory objects, and takes the view that the objects of sense experience, which he here calls 'sense-data' (thereby introducing the term into the philosophy of perception), are non-physical representations of the physical world. And, having accepted this last point, he acknowledges that it is a good deal more difficult to find the fault in traditional sceptical arguments than he had previously supposed.

It is in discussion of these issues that Moore begins to develop two new methods of argument which combine to give his later writings their distinctive style - the appeal to common sense and the use of philosophical analysis.

3 A defence of common sense

Moore's appeal to common sense is not a dogmatic affirmation of the truth of the 'common-sense view of the world' in the face of critical or sceptical arguments to the contrary. Instead it belongs within a higher-order scepticism concerning the epistemic status of philosophical arguments themselves. Moore's claim is that the general principles upon which critical and sceptical philosophies rely are answerable to our everyday employment of the concepts they involve in our 'common-sense' judgments concerning particular matters of fact. Hence, Moore argues, critical and sceptical philosophers undermine the reasons for accepting the general principles on which they base their arguments. We are more certain that we know such things as that 'this is a pencil' than we can be of the premises of any sceptical argument which disputes such knowledge.

Moore's critics urge that this is too simple; after all, most sceptical arguments begin by exploiting entirely familiar grounds for doubt. And the status of 'common sense' is not itself unchallengeable; Moore, an agnostic, would not have welcomed an attempt to defend religious knowledge by an appeal to 'common-sense' religious convictions; but then he would have had to explain why common-sense beliefs about the physical world have an epistemic authority that simple religious convictions lack. In fact towards the end of his life Moore himself came to feel that more needed to be said in order to deal properly with sceptical arguments. Although in his famous lecture 'A Proof of an External World' (1939) he argued that, contrary to the claims of idealists, one could prove the existence of an external world by demonstrating the existence of one's hands, in his comments on this lecture Moore denied that one could prove in a similar way that one possessed *knowledge* of the existence of an external world. To do this one had to be able to refute sceptical arguments, and this could not be done by simply waving one's hands at them. Moore himself attempted this task of refutation in two of his very last papers, but in fact by the end of the second one he seems almost ready to concede defeat, while equally maintaining that scepticism about the external world is literally incredible.

4 Philosophical analysis

As Moore's 'A Defence of Common Sense' (1925) shows well, his appeal to common sense is typically combined with an emphasis on philosophical analysis. The background to Moore's conception of analysis lies in his early conception of a proposition as both the object of thought and a possible state of affairs. Thus conceived propositions are complex structures of objects and properties, and an analysis of their structure as objects of thought is equally a metaphysical account of the structure of reality. But Moore's conception of the role of analysis was greatly enhanced by his appreciation of Russell's work in logical theory, especially the theory of logical fictions (see Russell, B.A.W. §9). One issue to which Moore applied this theory was that of propositions themselves. For although he regularly represented himself as concerned with the analysis of propositions, he did not think that his talk of propositions had to be taken as carrying with it a commitment to the existence of propositions, as genuine entities to which reference is made in the analysis of propositional attitude idioms. But whether his treatment of this issue was altogether consistent can reasonably be doubted (see Propositional attitudes).

The context beyond all others to which Moore applied his conception of analysis was that of perception: he maintained that although there is no doubt concerning our *knowledge* of such propositions as 'This is a hand', their *analysis* is deeply puzzling. For Moore their analysis led directly into the philosophy of perception since he took it that the demonstrative 'This' refers here to a sense-datum, whose relationship to one's hand requires clarification through further analysis. Moore took it that there were three alternative positions here: (1) a direct realist position, according to which the sense-datum is part of the hand (for instance, part of its surface); (2) an indirect realist position, according to which the sense-datum is a non-physical representation of one's hand; (3) a phenomenalist position, according to which all talk of one's hand turns out, when fully analysed, to involve no more than descriptions of actual and possible sense-data (see Perception §2). Moore moved through different versions of these positions at different stages of his career without ever finding one that commanded his wholehearted assent, though in later writings he seems to have favoured the phenomenalist alternative. Moore's indecision is perhaps best regarded as a *reductio ad absurdum* of his initial assumption that demonstratives always refer to a sense-datum. For even though this assumption is intended to be compatible with direct realism, the fact that Moore assumes that there cannot be a failure of reference, even when there is no appropriate physical object to be demonstrated, shows that his conception of a sense-datum is that of an object of experience whose existence is

independent of the physical world.

Moore was always emphatic that he did not think that all the problems of philosophy could be resolved by means of philosophical analysis. He resisted the conception of logical analysis propounded by Wittgenstein in his *Tractatus Logico-Philosophicus* (1921) (even though he regarded the book, to which he gave the name by which it is now known, as a work of genius); and he was equally suspicious of Wittgenstein's later attempts to show that philosophical problems arise from a misunderstanding of our language. Yet there is no doubt that analysis did play a central role in Moore's philosophical thought: again and again in his writings, he approaches a question by offering an analysis of the propositions which enter into his initial formulation of it. One can best appreciate the role of such analyses by reflecting upon the fact that Moore was inclined to favour a phenomenalist analysis of propositions such as 'This is a hand' at the very time at which he was propounding his anti-idealist proof of an external world by demonstrating the existence of his hands to his audience. For this shows that Moore's appeal to common sense was more limited than is sometimes appreciated. According to Moore, we can employ this appeal to assure ourselves of the truth of such propositions as 'This is a hand'; but such an appeal tells us little about what their truth consists in. For this, we require an analysis of these propositions, and this can be as radically unfamiliar as one likes, as long as the truth values of our common-sense judgments are respected. Thus the apparent conservatism of Moore's defence of common sense is somewhat deceptive. To put the matter in Fregean terms, common sense constrains us to respect the reference of our ordinary judgments, but it does not thereby reveal their sense, whose elucidation rests instead upon the outcome of philosophical analysis.

For this reason, Moore is rightly regarded as a paradigm 'analytic philosopher'. Indeed his early writings played a crucial part in the development of analytic philosophy, since they implied that the analysis of the logical structure of propositions could be of direct metaphysical significance. But because Moore never subscribed to the broad metaphysical doctrines of logical atomism and logical positivism, his conception of philosophical analysis remained somewhat pragmatic, piecemeal and idiosyncratic. None the less, particularly because of his emphasis upon common sense, his work constitutes one of the essential elements within that broad conception of philosophy which we think of as 'analytic philosophy'.

See also: Analytical philosophy §§1-2

THOMAS BALDWIN

List of works

Moore, G.E. (1899) 'The Nature of Judgment', *Mind* new series 8: 176-93.(The paper which signals Moore's youthful rejection of idealism.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press; revised edn, ed. T. Baldwin, 1993. (Moore's classic work of ethical theory, largely concentrating on the issue of the 'naturalistic fallacy'.)

Moore, G.E. (1912) *Ethics*, London: Williams & Norgate.(Moore's attempt at an elementary exposition of ethical theory, interesting in parts but rather pedestrian in style.)

Moore, G.E. (1922) *Philosophical Studies*, London: Routledge.(A collection of Moore's important papers from the period 1903-19.)

Moore, G.E. (1925) 'A Defence of Common Sense', in *Contemporary British Philosophy*, 2nd series, ed. J.H. Muirhead, London: Allen & Unwin; repr. in Moore (1959) and Moore (1993).(Moore's classic formulation of his complex attitude to 'common sense'.)

Moore, G.E. (1939) 'Proof of an External World', *Proceedings of the British Academy*, 25: 275-300; repr. in Moore (1959) and Moore (1993).(Moore's famous 'proof', which consisted in holding up his hands and declaring 'Here is one hand and here is another'.)

Moore, G.E. (1953) *Some Main Problems of Philosophy*, London: Allen & Unwin.(The text of lectures given in 1910-11 which sets the agenda for Moore's later thoughts.)

Moore, G.E. (1959) *Philosophical Papers*, London: Allen & Unwin.(A collection of Moore's important papers from the period 1923-55)

Moore, G.E. (1966) *Lectures on Philosophy*, ed. C. Lewy, London: Allen & Unwin.(A selection of Moore's lecture notes for the period 1929-34.)

Moore, G.E. (1986) *G.E. Moore: The Early Essays*, ed. T. Regan, Philadelphia, PA: Temple University Press. (Contains most of Moore's papers from the period 1897-1904.)

Moore, G.E. (1993) *Selected Writings*, ed. T. Baldwin, London: Routledge.(A collection of Moore's important papers and some further selections.)

References and further reading

Ambrose, A. and Lazerowitz, M. (eds) (1970) *G.E. Moore: Essays in Retrospect*, London: Allen & Unwin. (Contains useful critical essays on Moore's work.)

Baldwin, T.R. (1990) *G.E. Moore*, London: Routledge.(The only critical study of all of Moore's work.)

Keynes, J.M. (1949) 'My Early Beliefs', in *Two Memoirs*, London: Hart-Davis; also in J.M. Keynes, *Collected Writings X*, London: Macmillan, 1972.(Provides a classic account of the influence of*Principia Ethica* on the Bloomsbury Group.)

Levy, P. (1979) *Moore: G.E. Moore and the Cambridge Apostles*, London: Weidenfeld & Nicolson.(Describes Moore's friendships with the Bloomsbury Group.)

Regan, T. (1986) *Bloomsbury's Prophet*, Philadelphia, PA: Temple University Press. (Discusses Moore's early writings.)

Schilpp, P.A. (ed.) (1942) *The Philosophy of G.E. Moore*, Evanston, IL: Northwestern University Press; 3rd edn, La Salle, IL: Open Court, 1968.(Rather old-fashioned, but contains some classic papers by Bowsma, Ducasse and Frankena, together with Moore's autobiography and his reply to his critics.)

Stroud, B.A. (1984) *The Significance of Philosophical Scepticism*, Oxford: Clarendon Press.(Chapters 3 and 4 provide a sympathetic discussion of Moore's epistemology; but see chapter 9 of Baldwin (1990) for criticism of Stroud.)

Sylvester, R.P. (1990) *The Moral Philosophy of G.E. Moore*, ed. R. Perkins and R. Sleeper, Philadelphia, PA: Temple University Press.(A scholarly discussion of Moore's treatment of ethical values.)

Moral agents

Moral agents are those agents expected to meet the demands of morality. Not all agents are moral agents. Young children and animals, being capable of performing actions, may be agents in the way that stones, plants and cars are not. But though they are agents they are not automatically considered moral agents. For a moral agent must also be capable of conforming to at least some of the demands of morality.

This requirement can be interpreted in different ways. On the weakest interpretation it will suffice if the agent has the capacity to conform to some of the external requirements of morality. So if certain agents can obey moral laws such as 'Murder is wrong' or 'Stealing is wrong', then they are moral agents, even if they respond only to prudential reasons such as fear of punishment and even if they are incapable of acting for the sake of moral considerations. According to the strong version, the Kantian version, it is also essential that the agents should have the capacity to rise above their feelings and passions and act for the sake of the moral law. There is also a position in between which claims that it will suffice if the agent can perform the relevant act out of altruistic impulses. Other suggested conditions of moral agency are that agents should have: an enduring self with free will and an inner life; understanding of the relevant facts as well as moral understanding; and moral sentiments, such as capacity for remorse and concern for others.

Philosophers often disagree about which of these and other conditions are vital; the term moral agency is used with different degrees of stringency depending upon what one regards as its qualifying conditions. The Kantian sense is the most stringent. Since there are different senses of moral agency, answers to questions like 'Are collectives moral agents?' depend upon which sense is being used. From the Kantian standpoint, agents such as psychopaths, rational egoists, collectives and robots are at best only quasi-moral, for they do not fulfil some of the essential conditions of moral agency.

1 Agents versus recipients

Moral agents should be distinguished from moral recipients (see Moral standing). Moral agents are those who are morally accountable for at least some of their conduct. They are subject to moral duties and obligations, and, therefore, to moral praise and blame. Moral recipients are those who are owed moral consideration for their own sakes (see Respect for persons). On certain views moral agents and moral recipients are coextensive. Thus according to Kantians persons are the only moral agents and the only moral recipients (see Kant 1785). We should not be cruel to animals, not because animals are owed anything for their own sakes, but because cruelty to animals may indirectly harm persons who are the only ends in themselves (see Animals and ethics). According to utilitarians all sentient beings are owed consideration to the extent that they have feelings; so on this view there are moral recipients who are not moral agents - for instance, animals, or at least their feelings. Can there be moral agents who are not moral recipients? It would seem not, unless perhaps one uses moral agency in a weak sense which includes among moral agents nonsentient entities such as robots and corporations.

2 Understanding

The view that moral agents must have the capacity to conform to some of the external demands of morality is consistent with the view that there are parts of morality that they cannot conform to. Thus kleptomaniacs do not have the capacity to conform to certain moral requirements about not stealing, but it does not follow that they are not moral agents or that they should not be held morally responsible for murders that they might commit (see Responsibility). Moral agents must be morally responsible for some of their conduct, not necessarily for all. What, then, are the conditions of moral agency?

One essential condition is that the agent must have the relevant understanding (or capacity for understanding) of what the external requirements of morality are (see Moral knowledge §1). Thus in the case of murder one must understand that murder is wrong and that a particular act is an instance of murder. Exactly how much understanding is required is not easy to specify. There are plenty of borderline cases, but there are clear cases on both sides of the line. A baby does not have the relevant understanding of any of the requirements of morality, while an average adult citizen does at least sometimes. Of course even average citizens are very ignorant in many matters, but that at most is relevant to assessing their responsibility in these matters; it does not prevent them being

moral agents.

Some existentialist philosophers insist that moral agency requires the ability to create and choose one's own values, unconstrained by objective or rational considerations (see Existentialist ethics). It is objected that such creation involves a capricious freedom, since the agents have no guide as to how they should choose their values. Charles Taylor (1982) attempts to overcome this problem by suggesting that moral agency requires that one should have the capacity to choose one's values, after reflection, in accordance with one's deepest and most authentic nature (see Taylor, C. §§4-5). Does this requirement provide the necessary guidance for the agent? Critics would point out that this just shifts the problem. In what sense are we responsible for our deepest nature?

It seems that moral accountability requires that the agents should have an objective basis for choosing their moral values. They could then be held morally accountable to the extent that they have the capacity to find out what the relevant moral requirements are and, to the extent that they have the capacity, to conform to such requirements in the relevant ways. People who cannot reason properly (such as the severely mentally ill) or those who lack certain volitional abilities lack the capacity to conform to the relevant moral requirements.

In order to be morally accountable, an agent does not always have to know or even have the correct opinion about what the moral requirements are. The capacity for finding out such things can be enough. For instance, some Nazis who persecuted Jews may have thought sincerely that they were doing the right thing; but if they could and should have known better then they can be censured for moral negligence. Had they thought things through, which they could and should have done, they would have realized how wrong such acts were. Or so it is believed by those of us who want to hold them morally responsible.

To have the capacity to find out that something is morally wrong does not necessarily involve having the capacity to know why. It is plausible to distinguish having right opinions on moral matters from knowledge of them. The person who has knowledge in moral matters not only has the right opinions but also has them for the right reasons. Even after a study of moral philosophy many people do not know why things like stealing and murder are morally wrong; they do not understand the grounds for such judgments. True, theories have been advanced to answer such problems, but there is no general agreement on correct answers. An ideal moral agent, who exists only in the imagination of philosophers, might have knowledge of all moral matters, but ordinary moral agents have only opinions (for example, that stealing is wrong) and the capacity to find out such opinions in areas where they are held morally responsible.

According to some philosophers it is not enough to have the intellectual ability to tell right from wrong. There are psychopaths who are quite intelligent in general, and can even talk intelligently about morality. They might be able to tell us what things are wrong and even why, but they lack moral sentiments, such as remorse and consideration for others, and are unable to act for the sake of moral considerations. Many would say that they are not moral agents, and therefore not subject to moral condemnation, nor to punishment in so far as punishment presupposes moral condemnation. Bradley (1894) thought that psychopathic killers are not moral agents and so we cannot morally condemn them or punish them in the way we punish moral agents, but we have the right to use social surgery, even to kill them if that is necessary for social welfare, somewhat as we have a right to kill dangerous animals; considerations of justice do not apply to those who are not moral agents. Bradley forgot that considerations of humanity may still apply to them.

Bradley's view has been endorsed by Jeffrie Murphy (1972) but with two important qualifications. First, he points out that there is the danger of abuse in such a system. If we were permitted to go in for social surgery against nonmoral agents, moral agents might sometimes be wrongly diagnosed as nonmoral agents. Second, many people become psychopaths because of bad social conditions, and not of their own free will. Murphy rejects Aristotle's view that psychopathic wickedness is like a disease that people are responsible for acquiring by their voluntary conduct. He argues that psychopaths are those whose potential moral agency has been destroyed by society, and that they should therefore not be treated too harshly.

John Rawls (1971) maintains that only those who can give justice are owed justice (see Rawls, J.). And he stresses the importance for moral personality of acting from the sense of justice. On this view those who cannot act from a sense of justice would not be moral agents and so would not be owed duties of justice. But there is a weaker thesis according to which it will suffice to be owed justice if people can give justice even if they are not motivated by the

sense of justice. In the case of potential criminals, even if they are incapable of acting out of a sense of justice, or for the sake of moral considerations, we can apply considerations of justice to them and respect their rights if they respect the rights of others, even if their reasons for doing so are egoistic.

Indeed some Hobbesian philosophers contend that rational egoists can set up a just society, without the aid of a moral sense or a sense of justice (see Contractarianism §§2-3). Similarly, one could operate with a sense of moral agency, according to which agents have a capacity to conform to some of the external requirements of morality, but may lack the capacity to act for the sake of the moral law. They could sustain something like a morality. From a pragmatic point of view it would not matter too much why people conformed to moral requirements as long as they continued to do so. Jeremy Bentham (1817) thought that human beings were primarily egoistic, and that basic human nature was unalterable. He suggested that the setting up of the right institutions, such as representative democracy, and sanctions, such as punishment, would lead all members of the public, including the rulers, to see that contributing to the common good would be in their own best interests; thus rulers who acted against the public interest would be unlikely to be re-elected (see Bentham, J.). But would such rational egoists be moral agents? This is partly a verbal matter. They would satisfy the requirement that people should have the capacity to conform to some of the external requirements of morality, but they would not meet the Kantian requirement about being able to act for the sake of moral considerations.

3 The inner life and the Kantian view

The substantial question is whether the Kantian requirement is legitimate. Different moral theories give different answers to this question. Critics of the Kantian view might ask: if members of your family rescue you from a burning house, would you not think better of them if they did this out of affection for you rather than out of a sense of duty? Kantians would reply that if they did it out of affection, this would not at that time involve an exercise of their moral agency. This is consistent with the view that there may have been some exercise of moral agency in the past if the agent cultivated the right feelings and dispositions out of a sense of duty. Some people through good luck have better feelings and dispositions than others. Their qualities may even be wonderful, but do they deserve *moral* credit? Wittgenstein thought that G.E. Moore’s lack of vanity and his innocence generally might be loveable but Moore did not deserve any moral credit, for he was not ‘talking of the innocence a man has fought for, but of an innocence which comes from a natural absence of a temptation’ (see Malcolm 1966: 80). But what of the family members who fight against a selfish temptation to run away from the burning house where you are, and overcome it not out of a sense of duty but because of their love for you?

Persons who, through no fault of their own, have temptations to commit serious crimes are at a serious disadvantage compared to those of us who are lucky and do not have these temptations (see Moral luck). The former may deserve moral credit for conquering their temptations and for attempting to cultivate the right dispositions. Moral agency in one important sense of the term is to be contrasted with what happens as a result of luck. The conduct of individual agents is produced by a combination of factors: heredity, environment and free will. On the Kantian view, since personal choice is the only contribution to conduct made by agents themselves, it is only for this last factor that they are to be held morally accountable (see Free will).

The Kantian view of moral agency presupposes an enduring self that has the power of acting freely in a strong libertarian (or nondeterministic) sense. The enduring self on this view is different from an enduring character. The self is that which has the character and is autonomous in the sense that it has the power after reflection to change or not to change the character to some degree (see Autonomy, ethical). It is only to the extent that it has this power that it is held morally accountable for actions that issue from it or its character. If there is no enduring self, if the later self is a different self from the earlier one, then the later self cannot be morally accountable for the acts of its predecessor any more than a child can be for the deeds of its parents. Critics point out that the free will and the enduring self that are presupposed are incoherent. Either the character is determined by various factors that are ultimately outside the agent’s control or, if there is a break in the causal chain, then the act is a chance or random event and so again there can be no freedom.

Kantians reply that this objection loses force once we acknowledge the existence of the inner standpoint of the moral agent. C.A. Campbell (1957) has pointed out that from the internal standpoint we can make sense of a free act that is neither determined by factors outside our control nor a random event. Campbell appeals to our phenomenological experience of moral effort in the face of moral temptation, in which we can make sense of the

creative agency of the self, when the self has the power to go beyond its formed character. On this view what is really admirable about moral agents is that they can obey the moral law, rising above their feelings and passions by efforts of will made for the sake of the moral law. So the existence of the inner life is considered, at any rate on the Kantian view, to be another essential requirement of moral agency. It is this requirement that is not met by robots, corporations, states and other groups. Even if they instantiate rational systems or functional systems such that it makes sense to attribute actions (in a functionalist sense) to them, they do not have an irreducible inner phenomenology (see Functionalism). Thus a corporation or a state is not joyous and does not suffer (in the phenomenological sense) except in the sense that is reducible to the suffering and joys of its members. To say this is consistent with the view that such entities are extremely important in the influence they have over their members and that their behaviour and the laws governing them are not reducible to the behaviour and the laws governing the behaviour of their members.

William James (1907) pointed out that what gives significance to human life is that we can set ourselves ideals or goals and then pursue them with zest, overcoming obstacles in the way (see James, W. §3). If there were no struggle in human endeavour there would be nothing heroic about us. When we admire individuals for their struggle against temptation, or against disease, or their heroic attempt at conquering mountains or solving mathematical problems, we are appealing to an inner life. If a computer solves a mathematical problem, however ingenious its solution, there is nothing heroic about it.

It might be objected that one can understand the duty versus temptations battle without appealing to an inner life; one may give a dispositional analysis of temptation as well as of overcoming it. But Kantians would reply that we can only make sense of moral effort and free will if we assume an internal point of view. If we look at things from a purely objective point of view neither the presence of determinism nor its absence can make sense of our free will and of our moral agency.

4 Collectives and moral agency

There is sometimes disagreement about whether such categories as the insane, children, robots and collectives are moral agents. This disagreement can partly reflect different standards of moral agency being used. It can also be due to different views about the facts. Thus some of us may deny that robots are moral agents on the ground that they lack an inner life. Others may disagree with us on the grounds that an inner life is not essential to moral agency; and some may argue that even though an inner life is essential, robots of the sophisticated variety might be constructed in the future who have inner lives in the sense that human beings do.

People often wonder whether collectives such as nations, states and corporations are moral agents. On the assumption that they lack an irreducible inner life or that they lack a permanent or enduring self, they will not fulfil an essential condition for moral agency in the Kantian sense. But if we use moral agency in a weaker sense, then they too could be moral agents. Kantians would say that they are at best quasi-moral agents, but critics of the Kantian approach would complain that the Kantian view is too stringent; it has presuppositions that are unverifiable and controversial. Some of these critics prefer a practical solution where terms like free will and moral agency are used in a weaker sense.

On a Humean view people can be given moral praise if they meet the external requirement of morality by promoting social welfare, provided their conduct reflects their character (see Hume, D. §4). Similarly, they can be apportioned moral blame and punished if they produce social misery, provided their conduct stems from their character (rather than by accident). No permanent or enduring self in any deep sense is required. What is required is that the individual being blamed has the same character traits (in the relevant respects) as the one who committed the act. We are free to the extent that our conduct reflects our character. Deep problems like whether we are responsible for acquiring our characters are bypassed on this view. It implies that collectives too could be moral agents in the way that individual persons are.

Indeed Hume explicitly compared the person’s self to a republic or a commonwealth. So the argument that groups are less eligible for moral agency than individual human beings collapses on this view. Individual human beings can have character traits that persist; but so can groups. And though groups do not have an irreducible inner life, neither do individual human beings on the Humean view, according to which the ultimate constituents are the items of experience. The person does not persist over time in any deeper sense than groups do. The person’s

relation to the individual items of experience is like the relation of the republic to individual members who compose it (see Mind, bundle theory of). So on this view an individual can be morally accountable in much the same sense as collectives; this is one of the corollaries of his view that Hume did not notice.

Much communal violence presupposes ideas of collective moral agency. For instance, during the partition of India, when some Muslims in one part of the country persecuted Hindus, another group of Hindus would take retributive actions against some other Muslims in another part of India. The Hindus would often justify such conduct on the grounds that the people they attacked shared common characteristics with their coreligionists who committed the evil deeds elsewhere. The Muslims would use a similar justification against the Hindus and the vicious circle would continue. Jonathan Edwards (1758) thought that human beings are now morally accountable for the sins of Adam, for they belong to the same collective, humanity, and share the same fallen nature as Adam (see Edwards, J.). On the Kantian view (as well as on some other individualist views) this is not enough. Persons now are separate moral agents and have a separate centre of consciousness from Adam, even if they have similar character traits; the self is different from and transcends the character. Such requirements are often violated by those who mete out retribution at a collective level.

But we do seem to make moral demands upon collectives. Thus one might say that the IMF (International Monetary Fund) ought to provide more facilities for the poorer countries. And it has been claimed that some of these demands are not reducible to demands upon individuals. If this claim is correct, it would follow that at least a part of our moral language is addressed to those who are moral agents in the weak sense but not in the Kantian sense. This can also be seen in the case of young children, with whom we sometimes use moral language before they have become moral agents in the full Kantian sense. Indeed it is partly by participating in such use of language that children gradually acquire moral sentiments.

The Kantian sense of moral agency is presupposed only by parts of our moral system, especially that part which is concerned with apportioning moral desert from the point of view of cosmic fairness. If certain wrongdoers are not Kantian moral agents, the view that they deserve to suffer for their conduct is undermined. We do say to a child of two that it ought not to kill. We may even punish the child for its wrongs if that does some good; but we do not think that if it did not suffer for its wrongs in this life, it would be fair if there were another world, a hell, where it would be made to suffer, in the way that it would be fair if Stalin (assuming that he was a Kantian moral agent) were made to suffer in hell. As for the IMF, even God would not be able to send it to hell or make sense of its deserving to suffer (in the irreducible phenomenological sense) as a collective.

See also: Action; Desert and merit; Kantian ethics; Moral justification; Moral motivation; Morality and identity; Praise and blame

VINIT HAKSAR

References and further reading

Although they involve intricate argument, none of these items is particularly technical. All except the Kant are fairly easy to follow.

Bentham, J. (1817) *Plan for Parliamentary Reform*, in J. Bowring (ed.) *The Works of Jeremy Bentham*, vol. 3, Edinburgh, 1843.(Proposes that the setting up of the right institutions, such as representative democracy, and sanctions, such as punishment, would lead all members of the public, including the rulers, to see that contributing to the common good would be in their own best interests, and therefore that rulers who acted against the public interest would be unlikely to be re-elected.)

Bradley, F.H. (1894) ‘Some Remarks on Punishment’, *International Journal of Ethics*; repr. in *Collected Essays*, Oxford: Oxford University Press, 1935, vol. 1, essay 7.(Advocates Darwinism and social surgery against nonmoral agents.)

Bradley, F.H. (1927) *Ethical Studies*, Oxford: Oxford University Press, 1927, ch. 1.(Outlines some of the conditions of moral responsibility.)

Campbell, C.A. (1957) *Selfhood and Godhead*, London: Allen & Unwin, 376-85. (A Kantian defence of free will and moral agency.)

Edwards. J. (1758) ‘Original Sin’, in C.A. Holbrook (ed.) *Works of Jonathan Edwards*, New Haven, CT: Yale University Press, 1970, part IV, ch. 3.(Argues in favour of collective moral accountability.)

Haksar, V. (1964) ‘Aristotle and the Punishment of Psychopaths’, *Philosophy* 39: 323-40; repr. in J. Walsh, and H. Shapiro (eds) *Aristotle’s Ethics*, Belmont, CA: Wadsworth, 1967, 80-101.(Discusses whether psychopaths

are morally accountable for their actions and examines Aristotle's views on moral agency.)

Haksar, V. (1991) *Indivisible Selves and Moral Practice*, Edinburgh: Edinburgh University Press, chaps 2 (section 2), 9 and 10.(Examines the analogy between individuals and collectives.)

James, W. (1907) *Talks to Teachers on Psychology and to Students on Some of Life's Ideals*, London: Longman's Green and Co., 265-301.(Discussion of what makes life significant.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(Advocates 'strong' interpretation of conditions required for moral agency.)

Malcolm, N. (1966) *L. Wittgenstein, A Memoir*, Oxford: Oxford University Press.(Includes Wittgenstein's views on G.E. Moore's 'innocence'.)

Murphy, J. (1972) 'Moral Death: A Kantian Essay on Psychopathy', *Ethics* 91: 284-98.(Discussion of psychopathy from a Kantian standpoint.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, ch. 8.(Expounds the view that a moral personality must have the capacity to act from a sense of justice.)

Taylor, C. (1976) 'Responsibility for Self', in A. Rorty (ed.) *The Identities of Persons*, Berkeley, CA: University of California Press, 281-99; repr. in G. Watson (ed.) *Free Will*, Oxford: Oxford University Press, 1982, 111-26. (Suggests moral agency consists in the capacity to choose one's values in accordance with one's authentic nature.)

Moral development

The concept of moral development has its roots in Plato's metaphor of ascent from the dark recesses of the cave to the initially blinding sight of the form of the good. Influenced by the developmental theories of Jean Piaget, Lawrence Kohlberg proposed a sequence of stages beginning with two stages of egoism, followed by stages of conventionalism, contractarianism, consequentialism, and finally a Kantianism emphasizing the role of universalizable laws. Recent empirical work has not, however, corroborated moral stage theory. There seems not to be a unified mode of thought that applies to all and only moral problems.

In the 1920s, Jean Piaget articulated the idea that cognitive development in a number of domains (regarding, for example, substance, space, time) involved the development of progressively more adequate ways of thinking about each domain. Children's thinking could be classified according to Piaget in terms of unified modes of thought at each of four stages, in terms of what he called a 'genetic epistemology'. Like the Platonic picture of ascent (see Plato §14), stage progression involved progress towards seeing the nature of things more and more clearly.

In 1932 Piaget published *The Moral Judgment of the Child*, a book he called a 'preliminary piece of work', in which he presented two discoveries of note. First, young children, but not older children, judge the goodness or badness of actions more on outcomes than on motives: accidentally breaking a lamp is worse than intentionally breaking a less expensive object. Second, children initially think of prescriptive rules as unchangeable, and as originating outside human convention and social life; only later do they come to see rules as conventions designed for profitable human interaction and as adjustable accordingly. But, by the time a child is 10 or 11, rules are seen as emanating from humans engaged in complex social interactions, trying to harmonize interaction and avoid conflict. The rules are not so much forced upon us from the outside as are chosen by reasonable persons trying to engage in productive interaction. Adopting Kantian terminology, Piaget referred to this shift in the moral understanding of the child as a progression from a stage of 'heteronomy' to the stage of 'autonomy'.

Beginning in the 1950s, Kohlberg (1981, 1984) pursued Piaget's Kantian assumptions, claiming that individuals at the highest stage of moral development respond to moral problems 'in moral words such as duty or morally right and use them in a way implying universality, ideals, and impersonality'. In Kohlberg's hands, Piaget's two stages yielded to a six-stage scheme:

(1) egoism: right is what is rewarded and what is wrong is what is punished;
(2) instrumentalism: right is what serves one's needs and satisfies fair agreements;
(3) conventionalism: right is what conforms to age, gender, occupational and social role conventions;
(4) social contract: right is conceived in terms of the conventions of the society as a whole, especially the legal conventions;
(5) consequentialism: right is what promotes the general welfare even if this might involve breaking the law, for example, laws that discriminate on the basis of race or gender;
(6) Kant's 'categorical imperative': right is acting in accordance with rules that you would be willing to recognize as universal laws.

The descriptive sequence of moral stages follows a universal and irreversible sequence, and although not everyone reaches the highest stage of moral development, those who do will see the highest stage they reach as more adequate than the previous stages they occupied. Thus, a stage (5) consequentialist who is exposed to stage (6) Kantianism will judge it an improvement.

The empirical sequence of the stages, and the normative preferences of those occupying the stages, were thought by Kohlberg to increase the plausibility of the claim that the stages of moral development were in fact increasingly more adequate from a moral point of view. Although Kohlberg did not test the relationship between moral judgment and moral action, he explicitly assumed that there was a significant correlation between the two.

At the present time, moral stage theory is in a state of considerable disarray. Starting in the late 1970s, Carol Gilligan (1982) argued that Kohlberg's longitudinal study, based as it was on an all-male sample, was not listening to the ways women speak, and thus think, about moral problems. She claimed that males deliberate from a perspective that weights heavily respect for impersonal rules and rights, whereas females are more likely to find

the particular needs of others salient and to respond to moral problems out of direct care, concern and feeling for the other. Although Kohlberg claimed in only one co-authored paper to have found small gender differences between males and females, and generally concurred with finding of colleagues that there were none using his way of testing, Gilligan's critique opened the possibility of gender bias. Kohlberg was eventually forced to acknowledge that his was a theory of 'justice reasoning', but not of moral reasoning more generally (see Feminist ethics §1).

Gilligan's theory did not emerge unscathed, however, since subsequent analyses of male and female moral reasoning using a wide variety of testing procedures and types of problems show that the only significant gender difference has to do with the moral problems males and females think they face, but not with how they reason about different kinds of moral problems (Walker 1984).

Another problem for stage theory is the evidence of moral regression: people who at one time exhibit stage (6) Kantian reasoning may later endorse more conventionalist views. During the 1980s Kohlberg and his colleagues revised their scoring manual to eliminate regressors. However, in doing this they also unfortunately eliminated any confirmed cases of stage (6) justice reasoning.

Although the idea of moral development is still alive, the prospects of a unified moral stage theory seems dim. Moral philosophers and moral psychologists increasingly agree that the domain of morality contains heterogeneous types of moral problems and responsiveness to different types of problems may well require different special-purpose competencies, or 'virtues', not some general 'moral faculty'.

See also: Cognitive development; Feminist ethics; Justice; Karma and rebirth, Indian conceptions of; Moral education

OWEN FLANAGAN

References and further reading

Flanagan, O. (1991) *Varieties of Moral Personality*, Cambridge, MA: Harvard University Press.(An argument for a psychologically realistic ethics containing a critical assessment of moral stage theories, the debate about morality and gender, as well as discussions of several alternative approaches to the study of moral psychology.)

Gilligan, C. (1982) *In a Different Voice*, Cambridge, MA: Harvard University Press.(A challenge to Kohlberg's theory on the grounds that it privileges a conception of morality that is gender-biased.)

Kagan, J. and Lamb, S. (eds) (1987) *The Emergence of Morality in Young Children*, Chicago, IL: University of Chicago Press.(A collection of mostly empirical papers offering an alternative to moral stage theory in which temperament, moral emotions and the like are highlighted.)

Kittay, E. and Meyers, D. (eds) (1987) *Women and Moral Theory*, Totowa, NJ: Rowman & Littlefield.(An interesting collection of papers, mostly by philosophers, on the topic of gender differences in morality.)

Kohlberg, L. (1981, 1984) *Essays on Moral Development*, New York: Harper & Row, 2 vols.(All of Kohlberg's most important papers covering the empirical evidence for moral stages and his various defences of the philosophical conclusion that the highest stage is the most adequate from the moral point of view.)

Piaget, J. (1932) *The Moral Judgment of the Child*, New York: Free Press.(A study of how children conceive of prescriptive rules, truth-telling and the relative importance of motives and intentions versus consequences.)

Walker, L. (1984) 'Sex Differences in the Development of Moral Reasoning: A Critical Review', *Child Development* 57: 677-91.(A meta-analysis of all the research on gender differences in moral reasoning with the finding that there are none, except in the types of problems males and females think they characteristically face.)

Moral education

This entry looks at three contemporary approaches to moral learning and education, all of which have roots in the history of philosophy. The first holds that just as children grow, or develop, in a physical sense, so they also develop in their moral dispositions or judgments. A central issue here is whether the concept of development is applicable outside its biological home.

The second sees moral learning not as a natural process, but as a deliberate induction into socially approved norms or values. On one version of this view, it is not enough to bring children to follow the rules enshrined in conventional moral codes as they need to learn to sift these in the light of higher-order rational principles. Problems arise here both about moral motivation and about whether morality is wholly to do with rules and principles. For other theorists moral education is more a matter of shaping children's nature-given desires and emotions into settled dispositions or virtues on Aristotelian lines. While the 'rational principle' view focuses on the morally autonomous individual, this view has its roots in communal moral traditions.

Despite Plato's belief that only knowledge is teachable, and therefore that it is doubtful whether moral goodness can be taught at all, the third view of moral learning maintains that it must include the acquisition of relevant knowledge and understanding, and cover the formation of dispositions. All this bears on how moral education should feature in schools - on the role of school ethos, learning by example, and the contribution of the whole curriculum.

1 Moral development

Let us start with a biological approach to moral growth in children. On this view morality develops in us from within, given appropriate external circumstances, from innate seed through to maturity. Rousseau's *Émile* (1762) has been a key text here for modern developmentalists, although few have followed his specific suggestion that the moral sentiments appear only with adolescence (see Rousseau, J.-J. §4). Jean Piaget charted the development of moral judgment (as distinct from moral behaviour) in children, identifying three main stages through which they pass: from a pre-moral lack of awareness of moral rules, via a perception of them as sacrosanct, to a more reflective understanding of their point in fostering cooperation and mutual flourishing. Subsequent psychological investigations by Lawrence Kohlberg (1981) of children's moral judgments, as revealed in their responses to stories involving moral dilemmas, told a broadly similar story (see Moral development).

Developmentalist accounts have appealed to that brand of 'child-centred' educator wary of attempts to mould or indoctrinate children into socially approved codes and attitudes and eager to leave as much as possible to children's own resources. They have not been without their critics. Carol Gilligan (1982) has drawn attention to the fact that Kohlberg's subjects were all boys; when girls' judgments were studied, their responses showed less evidence of detached appeal to principles and more of a caring response to particular individuals' needs and desires. While Gilligan's work has been important in psychology in pointing to gender differences in moral learning, it also raises the philosophical questions of what is meant by 'moral maturity' and how this is to be identified (see Feminist ethics §1). The claim that the morally highest state is consciously to be guided by abstract principles might be accepted by philosophers of a Kantian persuasion, but others might disagree (see Universalism in ethics §3).

On one view, the notion of moral development, if taken strictly, is incoherent. This view sees moral learning as a matter of being shaped in certain desired directions by one's parents, teachers and perhaps social influences more generally. It rejects the atomic individualism implicit in the developmental model, where social phenomena are at best environmental influences nurturing moral growth rather than intrinsic features of human personhood (see Human nature §2).

2 Moral education: the place of rules and principles

An uncomplicated notion of moral education favoured by some governments, religious bodies and no-nonsense members of the public identifies it as bringing children's behaviour into line with moral rules to do with such things as telling the truth, refraining from physical harm, and keeping one's promises. In especially uncomplicated versions, these rules are 'moral absolutes', allowing of no exceptions.

Many philosophers have argued that moral education must go further than mere rule-following. For one thing, the view of what is morally wrong - that masturbation is sinful, for instance - might itself be baseless. Moral education must help young people critically to assess the acceptability of these and other rules, otherwise it may degenerate into indoctrination. Moral autonomy is the central aim: indeed, as long as pupils are 'blind' rule-followers, they cannot strictly be said to be acting morally at all, but at best sub-morally (see Autonomy, ethical; Moral agents). This position is clearly akin to Piaget's and Kohlberg's, except that acting according to moral principles is not the end point of a developmental process, but the product of conscious social direction. As to how the moral principles should be characterized, philosophers of education have held different views. While some have been attracted by utilitarianism, others have been more Kantian (see Utilitarianism; Kantian ethics). The latter group include the early R.M. Hare (1952), despite his un-Kantian insistence on individual 'decisions of principle', and Peters (1981), who would replace Kant's monistic emphasis on the categorical imperative with a broader array of 'transcendentally deduced' ultimate principles, such as impartiality, benevolence, truth-telling and liberty.

Critics of this general approach have pointed to weaknesses in its central demand that moral beliefs be based on higher-order principles. Given the disagreements just mentioned about their content and capability of rational justification, what happens if the learners, encouraged to find an incontrovertible, rational basis for their moral beliefs, fail to find one? Moral autonomists are only a hair's-breadth away from moral sceptics: if there is, after all, no good reason for being moral, why strive to be so (see Moral scepticism; Moral justification §3)?

The question is as old as Plato, whose rationalist approach to moral education has contemporary echoes in the view under discussion. Plato's *Republic* is partly an answer to Thrasymachus' question at the outset asking why it is in his interests to be just. The account of the education of the guardians later in the work shows them learning to base their conduct not on taken-for-granted orthodoxy, but on the sure foundations of knowledge of the Good. Whether Plato's arguments here and elsewhere in the *Republic* succeed in answering Thrasymachus is disputable, as is also his assumption that if the guardians attain a proper knowledge of the Good, this is enough to guarantee their acting well (see Plato §14). Contemporary rationalists tend to rate moral knowledge a necessary, rather than a sufficient, condition of moral learning, for in addition children have to acquire appropriate dispositions to bring their behaviour into line with what they know to be right (see Moral knowledge §§1-2; Moral motivation §§1-2). Like Plato, however, their whole account is vulnerable if its foundations prove shaky.

This approach to moral education has also come under fire for the narrowness of its conception of morality. In concentrating so heavily on the rules and principles that one morally ought to follow, it does less than justice to other features of the ethical life, especially the virtues (see Virtues and vices §§1-3). True, it puts weight on intellectual virtues like clarity and independence of thought, as these will be necessary when one makes the abstraction to moral principles; true, too, that the names of proposed principles, like benevolence or impartiality, are also names of moral virtues; and that the great weight put on getting learners to see what they morally ought to do and act accordingly may be said to imply a master virtue of moral conscientiousness. But what can the theory say about children's learning to be courageous, kind, self-controlled, moderate in their bodily appetites, friendly, generous, cooperative, confident, or equipped with a proper measure of self-esteem? Even where benevolence comes into the rationalist account, as in Peters, it presupposes learners who finally come to act benevolently because this is what they morally ought to do, and not out of some more spontaneous fellow-feeling for the other. We come back to Gilligan's difficulty with Kohlberg, as well as, more broadly, to the question of the relative roles of reason and emotion in moral motivation (see Morality and emotions).

3 Moral education: the place of virtues and community

Children need to learn how to regulate their feelings of anger when frustrated, as well as their bodily appetites for food, drink and in due course sex. Contemporary approaches to moral education influenced by Aristotle (for example, Carr 1991) start, as his does in *Nicomachean Ethics*, with biological features of human nature such as emotions and physical desires like these and see the educator's task as helping to shape them in desirable directions (see Aristotle §22). Acquiring the virtue of self-control, for instance, is a matter of learning such things as when to feel anger or express it in appropriate behaviour, to what degree, on what occasions, and towards which people (see Self-control).

Similar points could be made about the bodily appetites in relation to the virtue of moderation, fear in relation to

courage, our inbuilt reactions to others as social animals in relation to virtues like friendliness or generosity. In each case children need habituation in responses fitting the particular occasion. These cannot be subsumed under rules as they require flexible and intelligent adjustment to circumstances (see Virtue ethics).

On such an account, moral education starts by getting children to behave in the way which a virtue would require in different particular situations. At first, they can have little or no understanding of why they have to do this, but they gradually learn to make the flexible responses just mentioned and thereby come to see what they feel and do under the aspect of, say, the courageous or the considerate, rather than, more primitively, as eliciting pleasure or preventing pain. Rationality is already at work in the shape of the practical *nous* which generates the flexible responses. It comes in, too, at a deeper level as children begin to grasp how self-control, generosity and the other virtues fit together in a wider picture of their wellbeing, Aristotle's *eudaimonia* (see Eudaimonia). For Aristotle, indeed, practical wisdom is the master virtue, embedded in the other virtues, but also helping us, against a larger canvas, to lead a flourishing life overall. It presupposes a theoretical understanding of whatever material is necessary for our appropriate responses. One contemporary illustration: in coming to terms with their sexual desires and knowing what to do when, young people these days have to have some knowledge of biological functions, contraception, AIDS, and their own and others' psychology.

On this view rationality has to do not with subsumption of actions under rules and rules under principles, but with relating particular responses to more global pictures of one's wellbeing (see Virtue ethics §6). The latter, in Aristotle, is inextricable from the flourishing of wider groups. While the rationalist account of moral education focuses on individuals and what they morally ought to do, the Aristotelian starting point is the ethical life of a political community (see Community and communitarianism). Learners learn the subtleties of what to do and feel within not fully articulated traditions of response embedded in their family and broader communal life (see Family, ethics and the §6). In addition, the virtues they acquire - courage, temperance, justice and others - are necessary not only for their own flourishing, but also for that of those around them.

There are differences of opinion about how far this Aristotelian approach is applicable to moral education not only in the Greek city state, but also in societies of scores or hundreds of millions of people like the political units in which we live today. Has the tradition of virtues-education been eroded irreparably except within favoured religious and other communities not destroyed by the ravages of modernism, as Alasdair MacIntyre has suggested?

Two other issues remain. First, could virtues-education be the whole of moral education? Should it be combined with elements of a rules-and-principles approach? And are there actions which do not feature in children's moral upbringing as such, but belong beyond the pale, in the realm of the unthinkable - perhaps things like murder, physical harm, or theft? Second, should children be brought up, as in the Aristotelian tradition, to see their own flourishing as inextricable from that of others? Or is 'morality' to be kept apart from 'prudence' (see Prudence)? If the former, is there still any place for moral education as a distinguishable part of education?

What seems clear is that there is a great diversity of ethical values jostling for a place in children's upbringing - Aristotelian virtues, utilitarian considerations, Kantian values of impartiality, autonomy and respect for persons, role-obligations, beyond-the-pale prohibitions, commitments to one's personal projects, and more besides. If these, or most of these, are to be included, are children to be taught to arrange them, in the rationalist mode, in some kind of systematic hierarchy of importance? Or are they to be encouraged, more after the Aristotelian pattern, to resolve conflicts intelligently within and between irreducible values in the light of specific circumstances (see Examples in ethics §2)?

4 Procedures of moral education

Plato first raised the question in the *Protagoras* whether morality - including the virtues of good citizenship - can be taught. Socrates doubted this on the grounds that there are no specialist teachers in this area as there are in, say, architecture or naval design (see Socrates §2). Socrates was assuming that teaching is a matter of transmitting knowledge. On this assumption, there is indeed a problem - despite Plato's later position in the *Republic*, described above - since the acquisition of knowledge, either propositional knowledge or skills, is no proof against wickedness. Children need to become good, not merely to know about the good. Hence the vital importance of cultivating desirable dispositions in them - which may count as teaching morality on a broader definition of 'teaching' than Socrates'. As a part of this enterprise the narrower kind of teaching of empirical knowledge is

important, too, as the example of sex education shows, but only as an adjutant to the dispositions. Some moral educators would also put weight on giving children - perhaps even younger children - a more philosophical understanding of the nature of moral issues; but not all would agree with this.

As to how the dispositions and their attendant knowledge - of whatever sort - are best acquired, answers will differ according to the emphasis one puts on a communal starting point for morality and to one's views on the separability of morality from personal wellbeing. Most writers on the topic stress learning by example. As for school learning in particular, some would start with the educative ethos of the whole institution, others with classroom activities. For some, like Dewey (1909) for instance, moral education is not a separate subject, but should pervade every aspect of the life of the school community, while for others it is closely allied to religious education. For an Aristotelian the teaching of literature could be especially important in revealing conflicts of ethical value and subtleties of judgment (see Examples in ethics §3).

See also: Education, philosophy of; Education, History of Philosophy of; Learning

JOHN WHITE

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985.(Book II discusses the moral virtues in general; later books look at particular virtues.)

Carr, D. (1991) *Educating the Virtues*, London: Routledge.(A contemporary Aristotelian account of moral education, including discussions of opposing viewpoints.)

Dewey, J. (1909) *Moral Principles in Education*, Carbondale and Edwardsville, IL: Southern Illinois University Press, 1975.(A spirited short account, in the Aristotelian tradition, of the school's contribution to moral education.)

Gilligan, C. (1982) *In a Different Voice: Psychological Theory and Women's Development*, Cambridge, MA: Harvard University Press, chaps 1 and 2.(These chapters of this psychological study of the distinctive nature of women's moral development discuss Kohlberg's theory.)

Hare, R.M. (1952) *The Language of Morals*, Oxford: Oxford University Press, ch. 4. (Raises questions about moral learning and teaching.)

Kohlberg, L. (1981) *Essays on Moral Development*, New York: Harper & Row, 2 vols.(Volume I is on the philosophy of moral development, volume II on its psychology.)

Peters, R.S. (1981) *Moral Development and Moral Education*, London: Allen & Unwin.(A collection of Peters' essays, including discussions of Piaget and Kohlberg.)

Plato (*c*.386-380 BC) *Protagoras*, trans. W.K.C. Guthrie, in E. Hamilton and H. Cairns (eds) *Plato, The Collected Dialogues including Letters*, Oxford: Oxford University Press, 1975.(A spirited Socratic dialogue on the theme of whether virtue can be taught.)

Plato (*c*.380-367 BC) *Republic*, trans. P. Shorey, in E. Hamilton and H. Cairns (eds) *Plato, The Collected Dialogues including Letters*, Oxford: Oxford University Press, 1975.(The first treatise on education firmly based on a philosophical system.)

Rousseau, J.-J. (1762) *Émile: ou, de l'éducation*, trans. A. Bloom, *Emile: or, On Education*, Harmondsworth: Penguin, 1991.(An account of an ideal education in accordance with nature.)

White, J. (1990) *Education and the Good Life*, London: Kogan Page.(Explores issues raised in §3 and §4 in the context of a discussion of the aims of education as a whole.)

Moral expertise

Moral experts are best defined as those who have studied moral questions carefully, know the main theories developed in response to such questions, and (where possible) know and are able to offer arguments that would convince reasonable people.

In scientific and technical areas, one important feature of a successful answer is that it works, in the sense that it makes accurate predictions. We can say that successful answers to moral questions take the form of arguments which, if examined carefully, would persuade reasonable people and lead to convergence in their moral views.

The moral responsibility of individuals for themselves does not preclude the role of moral advisor. Many self-pronounced moral experts might be interfering, condescending and hypocritical, but such characteristics need not accompany moral expertise. Probably no one could claim a high degree of expertise in all areas of ethics.

1 Attempts at a definition

Moral expertise certainly involves more than merely knowing what moral views are prevalent within a profession or a society (see Morality and ethics §2). There are people who study different professions' 'codes of ethical practice' (see Professional ethics). There are social scientists who study and know a lot about the moral views prevalent in one or more societies. Are such people moral experts? Only the most extreme form of moral relativist will think that these people must be moral experts. The crucial question here is whether to know what a given group's moral views are is necessarily to know what actions or practices would really be morally right. Extreme moral relativism seems implausible. The moral views prevalent in some groups certainly seem morally worse than those prevalent in other groups (see Moral relativism). Moral expertise presumably involves some ability to sort the better from the worse moral views.

We might go further and say that, if there is a distinction between better and worse moral views, then the people with the best moral views could qualify as moral experts. However, if what makes certain people moral experts is that their moral views are best, then in order to determine who the moral experts are we would first have to determine which moral views are best. Yet once we find out for ourselves which moral views are best, why do we need the notion of moral experts? If we make having the best moral views the criterion for moral expertise, it is unclear what role moral expertise has left to play.

We could try the view that experts are those who have studied the relevant questions and problems and who know the competing theories about them and whatever evidence there is. We do not have to know which controversial view in physics is right to know who the experts in physics are. Likewise, we might think of experts in some area of morality (such as business ethics or medical ethics) as those who have studied the area's moral problems, considered the opposing moral arguments, and even tested their own arguments and principles on informed audiences. The experts are sometimes proved wrong about what the correct view is, but they are the ones to whom we would reasonably go for help.

2 Moral expertise and scientific expertise

The comparison between morality and science immediately raises the question of whether a difference between them makes the idea of expertise appropriate in science but not in morality. There is broad agreement that one important feature of successful answers to scientific questions is that they make predictions that turn out to be correct. In applied sciences like medicine and mechanics, a successful answer is the one that cures the patient or makes the machine do what we want. We can then say that an expert in such fields is someone who regularly returns answers that are successful in that they achieve some agreed upon aim. So the question will be whether in ethics there is an agreed upon aim that would specify what would count as successful answers to moral questions.

Some would say that the agreed upon aim is to find ways of living and behaving that reasonable people would accept as justified. Just as an expert in medicine might be expected to give advice which, if followed, would maximize our chances of a long life, a moral expert might be expected to provide arguments which, if examined carefully, would persuade reasonable people and produce convergence in their moral views.

The criterion of success in moral arguments cannot be mere success in convincing people (after all, sophistry can

be good at convincing people). Rather, the arguments must be ones whose careful examination would resolve moral disagreements. What counts as careful examination? Certainly, the arguments must be ones that would stand up to logical scrutiny. They must also be arguments whose premises and conclusions would still seem compelling after people with powerful imaginations teased out all the implications.

We should not go so far as to claim that no one can have moral expertise unless they can present moral arguments that will persuade everyone who is reasonable and considers the arguments carefully. In other disciplines, there can be expertise without convergence of opinion, so why not in ethics? But where there are arguments whose careful consideration would persuade every reasonable person, then those who can mount these arguments have greater moral expertise in this respect than those who cannot.

The stress on good arguments in ethics suggests another difference from science, especially from applied natural science. With scientific questions, we sometimes might merely want the answer that works and not really be interested in the background theories and evidence that support the answer. In ethics, however, asking for the answer to a moral question without wanting to know the reasoning behind it looks like an attempt to duck responsibility.

3 Some concerns about moral expertise

What is the point of looking for moral experts if each of us must take responsibility for the morality of our own actions and way of life? That we must take responsibility for the morality of our own actions and way of life does not preclude a role for moral advisors. Indeed, when we face difficult moral decisions, consulting judicious advisors might often be the responsible thing to do.

Nevertheless, in mainstream Western culture to call someone a moral expert would seem ironic. It would typically be thought to suggest that the person is judgmental, interfering, condescending, self-important, hypocritical and perhaps close-minded about morality. Self-righteousness is almost never attractive. But moral expertise does not inevitably lead to an unpleasant personality. On the contrary, honesty will require those with moral expertise to acknowledge their own moral mistakes and fallibility.

It will also require them to acknowledge their limitations. Few could plausibly claim a high degree of moral expertise in all of many different areas. Think how much time and energy it would take to study in depth (even just most of) the problems and theories in the ethics of medicine, business, law and government, not to mention military ethics, ethics in international affairs, academic ethics, and the ethics of personal relationships. Although there may be some features common to problems in all these areas, there are also morally relevant differences between them. So expertise in one such area does not carry over into expertise in another. Part of having an expertise can be knowing its limits.

See also: Applied ethics; Moral knowledge; Theory and practice

BRAD HOOKER

References and further reading

Maclean, A. (1993) *The Elimination of Morality: Reflections on Utilitarianism and Bioethics*, London: Routledge. (Contends that modern, mainly utilitarian, philosophers have been wrong to claim that they are uniquely qualified to teach moral expertise. Maclean holds that philosophy does not settle moral questions, and that philosophers do not possess special moral expertise and so cannot teach such expertise to others.)

Plato (*c*.395-387 BC) *Gorgias*, trans. W.D. Woodhead, in E. Hamilton and H. Cairns (eds) *The Collected Dialogues*, Princeton, NJ: Princeton University Press, 1961.(Classic discussion of the relationship between possessing virtue and teaching it.)

Plato (*c*.386-380 BC) *Meno*, trans. W.K.C. Guthrie, in E. Hamilton and H. Cairns (eds) *The Collected Dialogues*, Princeton, NJ: Princeton University Press, 1961.(Investigation of the teachability of virtue.)

Plato (*c*.386-380 BC) *Protagoras*, trans. W.K.C. Guthrie, in E. Hamilton and H. Cairns (eds) *The Collected Dialogues*, Princeton, NJ: Princeton University Press, 1961.(Another dialogue focusing on the relationship between possessing virtue and teaching it.)

Williams, B. (1995) 'Truth in Ethics', *Ratio* 8 (2): 227-42; repr. in B. Hooker (ed.) *Truth in Ethics*, Oxford: Blackwell, 1996.(Rejects the idea that having a certain kind of degree (for example, in medical ethics) qualifies someone for deciding certain moral matters. But Williams goes on to explain how there is a place for the idea

of a helpful advisor on whether a particular ‘thick ethical concept’ (such as ‘treacherous’, ‘kind’, ‘honest’) applies in a given case.)

Moral judgment

The term 'moral judgment' can refer to four distinguishable things. First, the activity of thinking about whether a given object of moral assessment (be it an action, person, institution or state of affairs) has a particular moral attribute, either general (such as rightness or badness) or specific (insensitivity, integrity). Second, the state that can result from this activity: the state of judging that the object has the attribute. Third, the content of that state: what is judged by us, rather than our judging it. And fourth, the term can be read as commendatory, referring to a moral virtue that we might also call 'moral discernment' or 'moral wisdom'. There are three principal questions regarding moral judgment. The first asks what kind of state the state of moral judgment is, and in particular whether this state is to be characterized, either wholly or in part, as a state of belief. The second is concerned with the activity of moral judgment, investigating especially the role within this activity that is played by the application of rules. The third examines the conditions under which a person is justified in making a moral judgment with a given content.

1 Cognitivism versus noncognitivism

When I think that something has a certain moral attribute (that an action is wrong, say), the state I am in seems to have features both of cognitive (belief-based) and noncognitive states (see Analytic ethics §1). First, my moral attributions aspire to objective truth - to being true independently of my attitude towards them. For my moral attitudes can take the form 'This action is wrong, and would have been wrong even if I had approved of it', committing me to moral facts of the matter that I can be right or wrong about (see Moral realism §1). And a state that aims to be true to independently obtaining facts must be thought of as a belief, it seems (see Belief). However, attributing a moral or any other evaluative feature to something seems to go beyond being in a merely cognitive state: it seems to involve orienting myself towards or away from the object of the attribution. In particular, it is often maintained that there is a certain sort of 'internal connection' between the state of moral judgment and the judge's motivation: it is part of the concept of judging something to be wrong that if I judge that it is wrong to perform this action in these circumstances, then (provided I am not weak-willed, and other things equal) I am motivated to avoid it in these circumstances (see Moral motivation §§1-2).

Thus cognitivism about the state of moral judgment - the view that it consists essentially in a belief - seems to be supported by its objectivity-presupposing character, while noncognitivism (the denial of cognitivism) seems to be supported by the internal connection to motivation. The problem is not simply resolved by characterizing moral judgment as a compound state, comprising both a belief and a noncognitive companion state. For this would seem to require as an isolable component a pure belief that the action is wrong, and the possibility of such a belief is precisely what is at issue.

Beyond this, four main options present themselves. Each side of the dispute can either deny the claim being adduced in support of the other, or accept the claim while denying the relationship of support. Thus, first, a cognitivist might deny the internal connection, insisting that to the (unsurprisingly widespread) extent that moral judges tend to care about morality, this is only contingently true. A second, less plausible view is a noncognitivism that simply denies that moral judgments presuppose objectivity. Perhaps, as emotivism suggests, objectivity-presupposing moral judgments ought to be rejected and replaced with something else (see Emotivism); it seems hard to deny that moral judgment as it actually exists does have this objectivity-presupposing character. This leaves two possibilities, each of which attempts to allow for both the objectivity-presupposing nature of moral judgment and its internal connection to motivation. According to the cognitivist internalism of John McDowell (1979), moral judgments are indeed beliefs, but beliefs of a special kind, in being internally connected to the believer's motivational states. And according to the noncognitivist objectivism of Simon Blackburn (1984) and Allan Gibbard (1990), the objectivity-presupposing character of moral judgments can be explained without construing them as beliefs. The central thought here is that the attitude expressed by 'This action is wrong, and would have been wrong even if I had approved of it' can be explained as a noncognitive attitude that ranges across counterfactual as well as actual possibilities - as an attitude of approval of the action not only in the world as it is, but in the world as it would have been if I had happened to approve of the action but the world were otherwise unchanged.

It remains unclear, however, whether noncognitivism can supply a satisfactory treatment of the full range of semantic features of moral judgment-contents, and in particular, of unendorsed occurrences of moral terms, such as the first occurrence of ‘wrong’ in ‘If blasphemy is very seriously wrong, then it is wrong to associate with blasphemers’ (see Analytic ethics §2). If not, we should be cognitivists about our actual states of moral judgment. Whether sense can be made of the objective moral facts seemingly presupposed by moral judgment is a further question - the question of moral realism (see Moral realism).

2 Moral rules and deliberation

What is the role of rule-application in good deliberation about the moral attributes of things? A widespread view is this: good moral thinking identifies correct moral principles under which the moral judge then subsumes particular instances to produce a moral verdict about them (see Logic of ethical discourse §6; Universalism in ethics §3).

On this view, though, not much actual moral deliberation can qualify as good. Few moral judges are equipped with an exhaustive set of exceptionless moral principles by reference to which all their moral judgments are made. Most of the time, when we deliberate about whether a case instantiates a moral concept, we are engaged in an activity of moral discernment or judgment that is not simply a matter of applying a fully determinative, independently articulable rule to the case. Moreover, there seem to be good reasons not to conceive of moral deliberation as consisting ideally in the application of an exhaustive set of exceptionless moral rules. This view confronts a dilemma. Either our set of principles would be small and readily comprehensible - perhaps consisting of a single member, as in direct utilitarianism (see Utilitarianism) - but would need to be framed in such general terms as to make their application to even the simplest cases a difficult matter to determine; or they would be framed in terms specific enough to make the assessment of their application to particular circumstances straightforward, but would thereby need to be so numerous and highly qualified as to be unusable.

This does not yet show that good deliberation cannot consist solely in the application of rules, however. On R.M. Hare’s consequentialist view (1981), the response to this problem should be to equip ourselves with a set of familiar moral rules of thumb (for instance, ‘Promise-breaking is wrong, other things being equal’), together with a meta-rule that in cases where these rules conflict we should apply a fundamental principle of direct consequentialism (see Consequentialism; Hare, R.M.). Alternatively, Barbara Herman’s Kantian theory (1985) tells us to equip ourselves with an equally familiar set of rules of moral salience - rules identifying certain features of the maxims on which actions are performed as those showing that the actions bear the burden of moral justification - together with a meta-rule that, when an action has a morally salient feature, its permissibility is to be judged in terms of the fundamental categorical imperative principle (see Kantian ethics).

3 Moral rules and justification

The views just described allow that it is proper for much of our deliberation to proceed by appealing to merely *prima facie* rules - rules specifying defeasible reasons (reasons capable of being outweighed) for making a moral attribution - but they supplement them with more fundamental, fully determinative rules governing judgment in cases where the *prima facie* rules conflict. But why should we seek rules of the fully determinative kind? The natural answer is this: good moral deliberation issues in judgments that are justified; if an all-things-considered judgment I make, in a case where my *prima facie* rules conflict, is to be justified, it cannot be arbitrary; and a judgment of mine can only be non-arbitrary if it instantiates a fully determinative rule.

When it is said that judgments are only justified by the existence of rules, two things might be intended:

(a) What makes it right to apply this concept to this object is a rule for doing so.
(b) What warrants me in judging that this concept applies to this object is a rule for doing so.

Claim (a), we might say, concerns the ‘constitutive justification’ of a judgment-content, whereas (b) concerns the epistemic justification of a state of judging (see Justification, epistemic).

Claim (a) pictures constitutive justification as consisting in the subsuming of instances under rules. This picture has been attacked, on the ground that it betrays a fallacy about rule-following (see Wittgenstein, L.J.J. §10; Meaning and rule-following). Our ability to discern whether a given case falls under a rule cannot consist in our grasp of an independent meta-rule, on pain of regress. This applies to the rule ‘Identify instances falling under

concept *c*’ as much as to any other. Proceeding correctly in the application of a concept cannot require sensitivity to an independently articulable rule which one’s practice of applying the concept follows. If so, then this leaves open the possibility that there are no such further rules that justify our practice of moral concept-application. The onus is on a proponent of such rules to produce compelling examples that do square with our practice; the criticism is that they have failed to do so.

According to opponents of (a), then, there is no obstacle to the claim of ‘moral pluralists’ that the only defensible moral rules are *prima facie* ones (see Moral pluralism). A further claim is that there are not even rules of this limited kind. This is held by moral ‘particularists’, who claim that there are no properties that always count as reasons for the same moral attribution. Often, the fact that an action of mine will harm someone is a reason for the wrongness of the action; but sometimes (for instance, if I am rightly administering a just punishment), it will count morally as a reason for its rightness (see Logic of ethical discourse §6). If so, we need to exercise a form of moral judgment that goes beyond the application of rules not only in order to determine how moral reasons relate to each other, but to determine when a consideration counts as a moral reason (see Virtue ethics §6)

4 The virtue of moral judgment and moral epistemology

If claim (a), with its subsumptive picture of constitutive moral justification, ought to be rejected, then so should (b), concerning epistemic justification. If there need be no independently articulable rule for a correct moral attribution, then there need be no such rule for my being warranted in judging that the attribution is correct. This still leaves it open to claim a relation of mutual epistemic support between my judgments concerning *prima facie* moral principles and my judgments concerning the existence of moral reasons on particular occasions, given a coherence conception of epistemic justification (see Knowledge and justification, coherence theory of). However, concerning the justification of my judging that, all things considered, a particular object has a given moral attribute, it seems that we must say simply that this is justified provided it is the attribution that would be made by a person with the virtue of moral judgment or discernment.

How can one ever be warranted in believing this? If the correctness of a moral judgment depends on its endorsement by an agent with a virtue the conditions for the possession of which are not independently specifiable, then it seems that any claims to epistemic justification will amount to dubious claims of self-evidence (see Intuitionism in ethics). To progress beyond this, we should need to be able to cite the convergence in judgment of those whom we have independent reason to identify as good moral judges (see Moral expertise).

See also: Epistemology and ethics; Feminist ethics; Moral justification; Moral knowledge; Objectivity

GARRETT CULLITY

References and further reading

Blackburn, S. (1984) ‘Evaluations, Projections, and Quasi-Realism’, in *Spreading the Word*, Oxford: Clarendon Press, 181-223.(A prominent defence of noncognitivism against the objection that it cannot accommodate the objectivity-presupposing character of moral judgment.)

Dancy, J. (1993) *Moral Reasons*, Oxford: Blackwell, chaps 4-6.(A defence of particularism about moral reasons.)

Gaut, B. (1993) ‘Moral Pluralism’, *Philosophical Papers* 22: 17-40.(A defence of moral pluralism.)

Gibbard, A. (1990) *Wise Choices, Apt Feelings*, Oxford: Clarendon Press.(Another prominent defence of noncognitivism.)

Hare, R.M. (1981) *Moral Thinking*, Oxford: Clarendon Press.(A utilitarian account of moral judgment as exhausted by rule-application.)

Herman, B. (1985) ‘The Practice of Moral Judgment’, *Journal of Philosophy* 82: 414-36.(A Kantian account of moral judgment as exhausted by rule-application.)

Larmore, C. (1981) ‘Moral Judgment’, *Review of Metaphysics* 35: 275-96.(Argues that moral deliberation is not rule-governed.)

McDowell, J. (1979) ‘Virtue and Reason’, *The Monist* 63: 331-50.(A defence of cognitivism and the need for the virtue of moral judgment for the possession of moral knowledge.)

McDowell, J. (1981) ‘Non-Cognitivism and Rule-Following’, in S.H. Holtzman and C.M. Leich (eds) *Wittgenstein: To Follow a Rule*, London: Routledge & Kegan Paul, 141-62.(An attack on the subsumptive conception of justification which seeks to support moral cognitivism.)

Moral justification

Questions of justification arise in moral philosophy in at least three ways. The first concerns the way in which particular moral claims, such as claims about right and wrong, can be shown to be correct. Virtually every moral theory offers its own account of moral justification in this sense, and these accounts naturally differ from each other. A second question is about the justification of morality as a whole - about how to answer the question, 'Why be moral?' Philosophers have disagreed about this, and about whether an answer is even possible. Finally, some philosophers have claimed that justification of our actions to others is a central aim of moral thinking. They maintain that this aim provides answers to the other two questions of justification by explaining the reasons we have to be moral and the particular form that justification takes within moral argument.

1 Three questions of justification

Questions of justification arise in moral philosophy in at least three different ways. First, it is generally taken as a central task of moral philosophy to clarify the structure of first-order moral thinking. A theory should, for example, explain 'what makes acts right'(see Morality and ethics §1). Is the status of an action as morally right or wrong determined by its actual or expected consequences, by its conformity to certain rules or standards, or in some other way (see Consequentialism; Deontological ethics)? Any theory that answers these questions provides us with an account of moral justification, that is to say, of justification *within* morality.

A second task of moral philosophy is to explain why we should care about morality and give its requirements priority over other considerations. This can be seen as a matter of providing a justification of morality.

Third, some philosophers have argued that it is a central aim of morality (or of some part of it, such as principles of justice) to provide a basis for justifying our actions or our institutions to one another, and that this aim provides a basis for understanding the importance of morality and for determining its content. According to these writers, justification is an aim of morality.

2 Justification within morality

Every moral theory offers its own account of moral justification. There is, however, a more general methodological question that transcends particular accounts of 'what makes acts right', namely the question of how the claim of any such account to be the correct account of moral justification can itself be justified.

One answer to this general methodological question has been formulated by Rawls (1971) as the method of 'reflective equilibrium' (see Rawls, J.). According to this method, one begins by identifying one's 'considered moral judgments', those judgments in which one has the highest degree of initial confidence. These may be judgments of any degree of abstraction: judgments about the rightness or wrongness of particular actions, about the correctness of general moral principles, or about the moral relevance of various considerations. One then tries to find the set of principles (the account of the particular area of morality in question) that best fits with these judgments. This fit will almost certainly be imperfect, so one then proceeds to consider whether and how to modify the principles one has formulated to yield a better fit with one's considered judgments, and whether and how to modify these judgments in the light of the principles at which one has arrived. This process of revision leads to a new set of principles, and hence to a new stage of adjustments of principles and judgments. The process continues until 'equilibrium' between principles and judgments is reached; the principles one has formulated at that stage are to be taken as justified as an account of the area of morality in question.

This account of how moral principles are to be justified has been attacked, chiefly on the ground that it gives too much authority to the considered judgments with which the process begins. This is sometimes put in the form of a charge that it is circular to base an argument for a particular form of moral justification on some of the very beliefs that are to be justified. It is also objected that two people, using this same method, might arrive at quite different principles, because they began with a high degree of confidence in different sets of judgments. In reply, it can be said that the charge of circularity is misplaced, since the aim of the process is not to supply a needed justification for every moral belief we have, but rather to come up with the best general account of the subject matter. How else could we arrive at justified conclusions about a subject except by relying, at the outset, on those beliefs about that subject that seem to us most likely to be true? Moreover, in the search for reflective equilibrium no particular

judgment is assumed, irrevocably, to be correct. Any reasons we might have for mistrusting particular beliefs are taken account of in the process as described, either as possible reasons for denying them the status of considered judgments to begin with or as reasons for abandoning them later when their incompatibility with the best account of our other beliefs comes to light.

3 Justification of morality

Turning now to the question of justifying morality itself, circularity looms when the bases of a proposed justification seem themselves to be moral in character. One cannot answer the question ‘Why care about right and wrong?’ by saying that it would be morally deficient not to. There is a dilemma here, however, since a justification for morality that appeals ultimately to values that are clearly not moral (such as those of self-interest) does not seem to give the right kind of explanation of morality’s importance. It does not seem that virtuous people care about what is morally right because doing so is conducive to their self-interest.

Prichard (1917) concluded from this dilemma that moral philosophy, or at least that part of it that seeks to provide a justification of morality, ‘rests on a mistake’ (see Prichard, H.A. §2). A less pessimistic conclusion would be that an adequate justification, or explanation, of the authority of moral requirements must appeal ultimately to considerations which are evidently relevant to morality yet have a significance which is not wholly dependent on it. The appeal of utilitarianism for some people, and the conviction of others that morality must have a basis in religion, can be explained in part by the fact that the ideas of the greatest happiness of the greatest number and of divine will appear, to members of these groups, to be the only plausible solutions to this problem. That is to say, these reasons have seemed to represent values both relevant to morality and able to supply it with the proper authority, but not (so it is claimed) dependent on it. One question is whether there are other kinds of reasons that have this status.

The task of finding such reasons is made more difficult by morality’s claim to unconditional authority, and to priority over all other values. A justification for such authority seems to require starting points that are peculiarly inescapable, and it may seem that reasons grounded in what is merely one value among others could not meet this test. Accordingly, many have attempted to justify morality by showing that it is entailed, or presupposed, by something to which everyone is committed. So, for example, Kant (1785) argued that we must take ourselves to be bound by morality in so far as we regard ourselves as rational agents, and Habermas (1990) has maintained that moral requirements are entailed by the idea of communication with others (see Habermas, J. §3; Kant, I. §9).

4 Justification as an aim of morality

Moral and political theories that may be broadly classed as ‘contractualist’ emphasize justification as an aim of morality (see Contractarianism §9). They argue, for example, that our thinking about right and wrong is guided by the idea of what could be justified to others on grounds they could not reasonably reject (Scanlon 1982), or that principles of justice are to be defended on the ground of their suitability to serve as public standards for the assessment of claims against the basic institutions of a society (Rawls 1993).

These ideas are related to questions of justification of the two kinds already discussed. First, it can be claimed that the aim of justifiability to others determines the content of justification within morality. Rawls, for example, concludes from the thesis just mentioned about the aim of justice that principles of justice and the basic institutions of a society must be justifiable on a basis that is independent of controversial world views and conceptions of the good life. Second, it can be argued that the value of being able to justify one’s actions to others has the right combination of connection with morality and independence of it to serve as an explanation of the significance that morality has for us. These claims are, however, controversial. Against them it can be maintained that the idea of justifiability to others lacks sufficient importance to account for the authority of morality and that some deeper standard of right and wrong must be presupposed as a standard of justifiability.

See also: Epistemic Relativism §2; Fallibilism; Moral Judgment; Moral motivation §§4-7; Moral scepticism

T.M. SCANLON

References and further reading

Brandt, R. (1979) *A Theory of the Good and the Right*, New York: Oxford University Press, esp. ch. 1.(Criticizes the method of reflective equilibrium and defends an alternative view of justification, appealing to ideas of

rationality and to empirical psychology.)

Brink, D. (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press, esp. ch. 5.(Critical discussion of issues concerning moral justification, defending a method like that of reflective equilibrium.)

Daniels, N. (1979) 'Wide Reflective Equilibrium and Theory Acceptance in Ethics', *Journal of Philosophy* 76: 256-82.(Reformulation and defence of Rawls' notion of reflective equilibrium.)

Habermas, J. (1990) 'Discourse Ethics: Notes on a Program on Philosophical Justification', in *Moral Consciousness and Communicative Action*, trans. C. Lenhardt and S. Nicholson, Cambridge, MA: MIT Press, 1990, 43-115; repr. in S. Benhabib and F. Dallmayr (eds) *The Communicative Ethics Controversy*, Cambridge, MA: MIT Press, 1990, 60-110.(Argues that moral requirements are entailed by the idea of communication with others.)

Hare, R.M. (1973) 'Critical Study: Rawls' Theory of Justice', in two parts, *Philosophical Quarterly* 23: 144-55, 241-52; repr. in N. Daniels (ed.) *Reading Rawls*, Oxford: Blackwell, 1975.(Attacks the methods of justification employed in Rawls' *A Theory of Justice*.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(Classic attempt to base morality on rationality.)

Prichard, H.A. (1917) 'Does Moral Philosophy Rest on a Mistake?', *Mind* 21 (81): 21-37; repr. in *Moral Obligation, Essays And Lectures*, ed. and with a note by W.D. Ross, Oxford: Clarendon Press, 1949.(Argues that it is a mistake to attempt to justify morality by offering an answer to the question, 'Why be moral?')

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, section 9.(Describes the method of reflective equilibrium, and defends principles of justice by arguing that they would be agreed to in a suitably defined 'original position'.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press, esp. lectures IV, VI.(Defends a view of how basic institutions and claims of justice should be justified in a pluralistic society.)

Scanlon, T.M. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press, 103-28.(Argues that justifiability to others is the guiding aim of our ideas of right and wrong.)

Moral knowledge

One possesses moral knowledge when, but only when, one's moral opinions are true and held justifiably. Whether anyone actually has moral knowledge is open to serious doubt, both because moral opinions are so hard to justify and because there is reason to think moral opinions are expressions not of belief (which might be evaluated as true or false) but of taste or preference. A successful defence of the view that people do have moral knowledge requires assuaging these doubts. Attempts in this direction standardly emphasize the respects in which our moral opinions, and the evidence we have for them, are analogous to the opinions and evidence we have concerning nonmoral matters, such as logic, mathematics, science, psychology and history. In the process they attempt to show that we do have good reason to think some of our moral opinions are true.

1 Confidence and evidence

People frequently, and quite casually, assume they have moral knowledge, in so far as they stand ready to act on the moral views they embrace. Yet people are circumspect when it comes to claiming moral knowledge, largely because so much of morality is controversial, but also because it is disturbingly difficult to defend in detail any particular moral view. Indeed, the confidence people have in their own moral convictions, although not easily shaken, often looks poorly backed by anything reasonably thought of as evidence or argument.

This apparent mismatch between confidence and evidence is largely explainable by noting that morality requires knowing how - how to act, react and feel - and not knowing that - that injustice is wrong, courage is laudable, kindness a virtue. One might well have the requisite abilities and skills without being able to articulate, let alone offer evidence for, one's moral views. Appreciating the practical dimension of morality is crucial to understanding its intimate connection with action. At the same time, it suggests that a proper account of morality must avoid an excessively intellectualized picture of the capacities it requires.

Still, on most accounts, people count as having moral knowledge, as opposed to moral virtue, only if (a) their moral beliefs are true and (b) those beliefs are justified. Thought of in this way, moral knowledge is an intellectual accomplishment. Against this background, arguments for thinking no one has moral knowledge commonly take one of two forms. Some hold that moral convictions are not properly seen as beliefs at all, which would mean we cannot satisfy (a). Others grant that people do have moral beliefs, but hold that such beliefs are all unjustified, which would mean we do not satisfy (b). A defence of moral knowledge must meet both threats (see Moral scepticism §3).

2 Noncognitivism

The first threat, posed by noncognitivists, usually travels with a positive account of moral convictions, according to which moral judgments reflect the preferences, or desires, of those who make them, but do not express beliefs. Noncognitivism finds support from the fact that moral opinions seem not to be empirically verifiable. As Hume put the point (1739/40), there appears to be no way to derive an 'ought' from an 'is', no way to establish a moral conclusion by appeal only to empirical premises (see Hume, D. §4; Logic of ethical discourse). G.E. Moore (1903) pressed this observation against all attempts to define moral claims in empirical terms, using what has come to be called the 'open question' argument (see Moore, G.E. §1). He maintained that all definitions of a moral term (say, 'good') using nonmoral terms (say, 'pleasure') are undermined by the fact that it always makes sense, in a way that it would not if the terms had the same meaning, to ask whether something describable by the empirical terms was describable using the moral term. It makes sense, for instance, to ask whether something pleasant is good in a way that it does not make sense to ask whether something pleasant is pleasant. Moore himself concluded that moral claims describe non-empirical, *sui generis* facts (see Intuitionism in ethics). Noncognitivists, in contrast, have held that if a claim is not empirically verifiable, it does not describe anything whatsoever. They conclude that the open question argument shows that moral claims, precisely because they cannot be defined in empirical terms, have no (descriptive) meaning and so cannot express beliefs (see Analytic ethics).

Noncognitivists frequently come to the same conclusion via another route by appealing to the apparently intimate connection between thinking that some action is morally required or good and being motivated (at least to some degree) to perform it. This connection is easily explained, the noncognitivists note, once we see moral 'beliefs'

not as genuine beliefs but instead as motivational states akin to desires or preferences (see Moral motivation §§1-2).

Whichever way one comes to the conclusion that moral claims do not express beliefs, the implications for moral knowledge are the same. For if a person's moral view fails to express a belief then that view, however salutary, cannot rightly be dignified as knowledge.

Against this threat, many cognitivists argue that, on a suitably sophisticated understanding of empirical verification, moral claims actually are verifiable. And they maintain that the open question argument establishes at most that moral claims are not synonymous with empirical claims, not that they are empirically unverifiable. Others, though, grant that moral claims are not empirically verifiable, but argue against empirical verifiability as a proper criterion for meaningfulness. As for the connection between moral conviction and motivation, cognitivists either deny that the connection is as intimate as the noncognitivists suppose, or argue that the intimate connection can be just as well explained without abandoning the view that moral convictions are expressions of belief. However these particular responses play out, cognitivists regularly stress the extent to which moral convictions resemble beliefs. We see them as answerable to evidence and argument, talk of them as true or false, and distinguish between those who know right from wrong and those who do not. All of this suggests that adopting a moral conviction is a matter of forming a belief about how things are morally.

3 Scepticism

The second threat to moral knowledge, posed by sceptics, leaves unchallenged the contention that we have moral beliefs. Instead, it attacks our right to see these beliefs as justified. We may believe that deliberate cruelty is wrong, and that slavery is unjust, but (the sceptic maintains) we have, at bottom, no good reason for thinking such beliefs are true.

The sceptical challenge, to the extent it raises a special problem for *moral* knowledge, depends on drawing a contrast between moral and nonmoral beliefs that shows the former to be especially suspect. Three considerations are standardly marshalled in support of such an unflattering contrast.

The first is that even a cursory glance across cultures, or back through history, reveals an astonishing diversity of moral views (see Moral relativism). That people so frequently come to such different, and incompatible, opinions makes the thought that they form their opinions in response to evidence or argument radically implausible. At the same time, the diversity exhibits a telling pattern. It looks as if the opinions people hold by and large reflect either their own interests or the interests of the powerful. This recommends seeing the moral views people embrace as mere rationalizations or as reflections of systems of social control and regulation (see Ideology). Either way, moral beliefs emerge as views we have no reason to think true - even when the views happen to be advantageous.

The second consideration picks up on Hume's point that there is an unbridgeable gap between 'is' and 'ought' and highlights the way in which, in defending one moral belief against others, we seem inevitably to appeal to other moral commitments that are equally contentious. In arguing with particular people we sometimes find common ground and can proceed, using our shared assumptions, to resolve a debate. The resolution is available, however, only because we share some moral view with them that others could reasonably challenge. To meet such a challenge, rather than ignore it or oppose it with force, we will (again inevitably) need to appeal to some other moral view in order to defend the one in question. Yet we will then just be pushing the issue back, finding ourselves appealing to another assumption which is itself either unsupported or supported by an undefended moral assumption. Thus the is/ought gap appears to entail that the only defence one might offer of one's moral beliefs is one that illegitimately presupposes others. So again our moral beliefs emerge as views we have no reason to think true - even when we can appeal to some in defending others.

The third consideration focuses on what the world would have to be like in order for our moral beliefs to be true and justified. Moral facts putatively stand as authoritative guides to action. Yet it is hard to make sense of this authority at all, and harder still to reconcile such authority with a plausible view of the natural world. Moreover, if there are moral facts, they are presumably connected in some way to various nonmoral facts. After all, we often appeal to nonmoral facts to explain moral ones, saying for instance that some action is good because it causes pleasure. But the connection between these nonmoral facts and the moral facts looks to be utterly mysterious. Perhaps most troubling of all, there seems to be no good way to explain how we discover these peculiar facts and

their connection to the more mundane facts upon which they supposedly depend. These difficulties all provide reason to think the world is not the way it would have to be in order for our beliefs to be true and justified (see Moral realism §6; Naturalism in ethics). This time our moral beliefs emerge not only as views we have no reason to think true, but as views we have positive grounds for rejecting as false.

Against the sceptics, those who hold that moral beliefs are sometimes justified standardly note, to start, that moral disagreement is not as widespread as the sceptics suggest nor noticeably more common than disagreements concerning empirical matters. At least when it comes to fundamental moral principles (concerning, say, the value of life, the importance of honesty, the nobility of courage) there is actually a striking consensus. Disagreements at a less fundamental level are, of course, common. Yet these disagreements frequently turn on differences of opinion concerning plain matters of fact. When they do not, the arguments we offer on behalf of a particular moral view will, as the sceptic says, inevitably implicate further moral assumptions. However, in this respect, moral arguments resemble nonmoral arguments in that the latter too must always be advanced against a background provided by additional (in these cases, nonmoral) assumptions. Indeed, whether the disagreements are moral or not, almost any argument one side offers will appeal to considerations the other side might intelligibly call into question. Thus, it seems, neither of the first two considerations advanced by sceptics stands as an especially strong argument for thinking our moral beliefs suffer by contrast with other sorts of beliefs we readily acknowledge as (sometimes) justified.

In response to the third consideration, many argue that neither moral facts nor our epistemic access to them are as mysterious as sceptics maintain. To a large extent, the suggestion that moral facts, and so our access to them, would have to be mysterious rests on rejecting the view that such facts are reducible to, or constituted by, natural facts. Yet the standard arguments against naturalism, which appeal to the open question argument and the is/ought gap, lose much of their force in light of recent developments in semantics and metaphysics. Such developments suggest, first, that the open question argument mistakenly assumes that moral and nonmoral terms can refer to the same property only if they are synonymous and, second, that our inability to infer moral conclusions from nonmoral premises is no more significant than our similar inability to infer biological or psychological conclusions from nonbiological or nonpsychological premises. If naturalism concerning moral facts is defensible, then our epistemic access to moral facts is no more puzzling than is our access to nonmoral facts, even if it is sometimes more tenuous.

Whether or not naturalism is, in the end, defensible, many argue that we have strong evidence for some of our moral views anyway. They make their case by emphasizing the significant respects in which the evidence we have for our moral views is analogous to the evidence we can marshal for certain nonmoral views. Moral sense theorists, for instance, see some moral beliefs as analogous to perceptual judgments and suggest these justify our other moral views in much the way perception provides support for our scientific theories (see Moral sense theories). Alternatively, intuitionists and rationalists maintain that we come to know certain moral truths - truths that can then provide evidence for others - in just the way we come to know certain fundamental truths of mathematics or logic, by relying on reason and reflection. These particular proposals usually play out against the assumption that justification is possible at all only if at least some fundamental beliefs can serve to justify others without themselves needing support. But even those who reject this foundationalist view, and embrace instead a coherentist conception of justification, argue that moral beliefs, no less than various nonmoral beliefs, are justified to the extent that they stand in relations of mutual support (see Knowledge and justification, coherence theory of; Foundationalism). The common thread in all these suggestions is the persistent pressure to recognize that our grounds for accepting particular moral views are not significantly different in kind from the grounds we have for other views that are standardly thought to be unproblematic. And the suggestion is that experience, reason and reflection, which are the main tools we have for learning about the world, may help us as well in coming to a justified view of morality.

See also: Common-sense ethics; Emotivism; Justification, epistemic; Knowledge, concept of; Moral expertise; Moral Judgment; Prescriptivism

GEOFFREY SAYRE-McCORD

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz; 2nd edn, 1946. (Classic appeal to verifiability

considerations in defence of noncognitivism.)
Brink, D. (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.(Offers a defence of cognitivism combined with a coherentist epistemology.)
Daniels, N. (1979) 'Wide Reflective Equilibrium and Theory Acceptance in Ethics', *Journal of Philosophy* 76: 256-82.(Articulates method of 'reflective equilibrium' with special attention to its epistemological implications.)
Ewing, A.C (1947) *Ethics*, London: Macmillan.(Defends a fallibilist version of intuitionism.)
Harman, G. (1977) *The Nature of Morality*, New York: Oxford University Press. (Mobilizes forcefully the worry that moral beliefs might be unjustified because they do not figure in the best explanations of our experiences.)
Hudson, W.D. (1969) (ed.) *The Is/Ought Question*, London: Macmillan.(Collection of papers exploring the plausibility and significance of the gap between 'is' and 'ought'.)
Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978.(Seminal expression of arguments often taken to support either moral scepticism or noncognitivism, but also (on some interpretations) a defence of moral sense theory.)
Kant, I.(1785) *Grundlegung zur Metaphysik der Sitten*, trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(Classic articulation of the view that morality can be grounded in rationality.)
Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin. (Defends moral scepticism by appeal to the ontological and epistemological difficulties that attend a belief in moral facts.)
Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press.(Classic defence of intuitionism that has shaped most of twentieth-century meta-ethics.)
Ross, W.D. (1930) *The Right and the Good*, Oxford: Oxford University Press. (Important extension and elaboration of intuitionism.)
Sayre-McCord, G. (ed.) (1988) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press.(Collection of papers devoted to issues related to noncognitivism and scepticism.)
Sinnott-Armstrong, W. and Timmons, M. (eds) (1996) *Moral Knowledge?*, New York: Oxford University Press. (Collection of contemporary papers on moral knowledge.)
Stevenson, C.L. (1937) 'The Emotive Meaning of Ethical Terms', *Mind* 46: 14-31. (Influential defence of noncognitivism that highlights the pragmatic role of moral language.)

Moral luck

The term 'moral luck' was introduced by Bernard Williams in 1976 to convey the idea that moral status is, to a large extent, a matter of luck. For example, that Bob grows up to be vicious and Tom to be virtuous depends very much on their different family conditions and educational background. Following Williams, Thomas Nagel widened the scope of moral luck. The position taken by both stands in stark contrast to the widely-held view, influenced by Kant, that one is morally accountable only for what is under one's control, so that moral accountability is not a matter of luck. This idea is so deeply entrenched in our modern concept of morality that rejecting it would call for a rethinking and reformulation of the most basic notions of morality. Some have argued that the paradox of moral luck provides a strong reason to abandon traditional moral theories, and lends support to virtue ethics.

1 Domains of moral luck

Bernard Williams introduced the term 'moral luck' in 1976, to convey the idea that moral status is largely a matter of luck. Williams' main interest lies in how luck may influence agents' reflective assessment of their actions, that is, in their ability to justify rationally their decisions and actions to themselves. He uses the example of Gauguin, who abandoned his family to pursue his artistic aspirations in Tahiti. In Williams' view, whether Gauguin's choice can be justified depends on the success of his venture, a success not entirely under his control and, thus, in part, a matter of luck. Two obstacles block the agent's capacity to guarantee retroactive justification in such decisions. First, at the time of the decision-making the agent can never know how things will turn out. Second, in projects like that of Gauguin's, the project often changes the very character and attitudes of the agent. The Gauguin at the time of the decision has no way of knowing whether Gauguin after the decision will approve retroactively of the project (see Williams, B.A.O. §2)

While Williams is concerned mainly with luck from the 'subjective' point of view, namely, with how agents relate to the justification of their own actions, Thomas Nagel seeks to show the influence of luck from a more 'objective' point of view. According to Nagel (1976), luck affects morality in four ways. The first is 'constitutive luck', luck in the kind of person one is, which depends, at least partially, on factors beyond one's control (such as heredity and environment). The second is 'circumstantial luck', luck in the kind of problems and situations one faces. Nagel offers the example of some young Germans who were morally lucky to have left Germany in the early 1930s. Had they stayed, they would have been faced with a terrible moral test which many of them would probably have failed. Thus, according to Nagel, whether or not a person became a Nazi murderer was, to a significant extent, a matter of circumstantial luck.

The other two ways luck is believed to influence morality have to do with the causes and effects of actions. One is concerned with the way antecedent circumstances determine the acts of the will itself (see Will, the). If determinism is true, then our very choices are beyond our control, and thus a matter of (good or bad) luck (see Free will). The other and last kind of luck, sometimes referred to as 'resultant luck', is luck in the outcome of one's actions and projects. This includes cases of decisions made with uncertainty, as when a political leader decides to go to war knowing full well at the time of the decision that if the venture fails, the decision can never be justified retroactively; and cases of negligence, such as that of two reckless drivers, one of whom is lucky, not causing any injury, the other being unlucky, hitting a child crossing the road. The independence of these different kinds of luck from one another makes life difficult for critics of moral luck. To sustain the alleged immunity of morality to luck, they have to show that each and all of them have no effect of morality.

Attempts to apply the contemporary debate on moral luck to ancient philosophy focus on another aspect of luck, namely, its relation to wellbeing. Some ancient schools, especially those of the Cynics and Stoics, argued that immunity to luck is a necessary condition for wellbeing, while others, notably Aristotle, emphasized the essential vulnerability of human happiness (eudaimonia) (see Cynics; Stoicism; Aristotle §21; Eudaimonia). According to Martha Nussbaum (1986), their different attitudes towards the role of luck mark an important difference between Aristotle and Plato (see Plato).

2 Arguments for and against moral luck

To resolve the paradox of moral luck, some have criticized the vivid examples presented by Williams and Nagel. With regard to the lucky and unlucky drivers, for example, it has been argued that our different ways of relating to them are not to be explained by reference to the drivers' ultimately different moral status - they were equally reckless - but to our epistemic shortcomings. In most cases, we know about other people's personalities and intentions only through their actual behaviour. Hence, contra Williams and Nagel, luck does not affect one's real desert, but only our knowledge of it.

Proponents of moral luck have argued that any attempt to detach ourselves from the unintentional and unforeseen aspects of our actions would make it impossible for us to retain our identity and character as agents. Since every aspect of our moral life results from factors which are beyond our control, there seems to be nothing left for us to be responsible for and the area of genuine agency, the traditional locus of moral judgments, shrinks to an extensionless point. Such a detachment also carries with it unwanted normative implications. If responsibility were to be limited strictly to what is under our control, we would be able to walk away from the sick, the old, and the otherwise helpless, provided their helplessness was not an intended result of any agent's actions. A world without moral luck, that is, a world where the boundaries of our responsibilities are determined by what is under our control, would thus be - on this argument - a place in which most of us would not want to live (see Responsibility §§2-3; Desert and merit §§2-3).

3 Theoretical implications

Acceptance of the existence of moral luck has significant implications for ethical theory. Williams and others have argued that the paradox of moral luck provides us with a powerful reason to abandon the ethics of duty and to adopt some kind of virtue ethics (see Virtue ethics). Virtue ethics is expected to escape the problem of moral luck by focusing on good and bad traits of character, instead of on notions of moral responsibility and blame. While we refrain from attributing responsibility and blame in cases where luck is involved in the action under question, we do not change our judgment of character if having a certain character is a matter of luck.

Allowing luck to encroach into the realm of morality means assigning less importance to the category of the voluntary, as the boundaries of our responsibility and obligations would on this view extend beyond the limits of our voluntary actions. If moral luck exists, we might be accountable for results we could not control, and we might be responsible for other people even if we never voluntarily committed ourselves to them. Since reference to the voluntary versus the nonvoluntary seems to be embodied in the very notion of the 'moral', acceptance of moral luck would imply abandoning, or assigning much less significance to, the common distinction between the moral and the nonmoral. Such a move is indeed made by most proponents of moral luck.

See also: Praise and blame

DANIEL STATMAN

References and further reading

Kenny, A. (1993) 'Aristotle on Moral Luck', in R.W. Sharples (ed.) *Modern Thinkers and Ancient Thinkers*, London: University College London.(An attempt to relate the modern debate on moral luck to its ancient and medieval antecedents.)

Nagel, T. (1976) 'Moral Luck', *Proceedings of the Aristotelian Society*, supplementary vol. 50: 137-51.(Responds to Williams' essay of the same title (published in the same issue of this journal), disagreeing on some points but accepting and further expanding the notion of the vulnerability of morality to luck.)

Nussbaum, M. (1986) *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*, New York: Cambridge University Press.(Discusses attitudes to luck in ancient literature and philosophy, especially in the work of Plato and Aristotle.)

Statman, D. (1991) 'Moral and Epistemic Luck', *Ratio* 4: 146-56.(Explores the relation between the problem of moral luck and that of epistemic luck.)

Statman, D. (ed.) (1993) *Moral Luck*, Albany, NY: State University of New York Press. (Includes the articles of Williams and Nagel, as well as most of the important discussions of the topic, a postscript of Williams to his 1976 paper, and a comprehensive introduction by the editor.)

Williams, B. (1976) 'Moral Luck', *Proceedings of the Aristotelian Society*, supplementary vol. 50: 115-36; repr. in *Moral Luck: Philosophical Papers 1973-80*, Cambridge: Cambridge University Press, 1981, 20-39. (Introduces the term 'moral luck' and explains the central role played by luck in morality, especially in

retroactive justification.)

Moral motivation

Questions about the possibility and nature of moral motivation occupy a central place in the history of ethics. Philosophers disagree, however, about the role that motivational investigations should play within the larger subject of ethical theory. These disagreements surface in the dispute about whether moral thought is necessarily motivating - 'internalists' affirming that it is, 'externalists' denying this.

The disagreement between externalists and internalists reflects a basic difference in how the subject matter of ethics is conceived: externalism goes with the view that ethics is primarily about the truth of theories, construed as sets of propositions, while internalists see morality as a set of principles meant to guide the practical deliberations of individual agents. Internalists interpret questions of objectivity in ethics as questions of practical reason, about the authority of moral principles to regulate our activities. Here controversy has centred on whether the authority of practical principles for a given agent must be grounded in that agent's antecedent desires, or whether, instead, practical reason can give rise to new motivations.

There are also important questions about the content of moral motivations. A moral theory should help us to make sense of the fact that people are often moved to do the right thing, by identifying a basic motive to moral behaviour that is both widespread and intelligible, as a serious source of reasons. Philosophers have accounted for moral motivation in terms of self-interest, sympathy, and a higher-order concern to act in accordance with moral principles. But each of these approaches faces difficult challenges. Can egoistic accounts capture the distinctive character of moral motivation? Can impartial sympathy be integrated within a realistic system of human ends? Can we make sense of responsiveness to moral principle, as a natural human incentive?

1 Internalism and externalism

Internalism states that moral considerations are necessarily motivating for those who grasp them. The most common version of this position is the claim that sincere acceptance of a moral judgment implies that the person who accepts it has some motive - not necessarily overriding - for compliance with the judgment. Externalists deny this, insisting that one might sincerely accept a moral judgment without any corresponding motivation. On this view, psychological questions about human motivation are completely distinct from issues about the truth of moral judgments or the nature and constitution of moral facts (see Moral judgment §1; Moral realism §1).

The categories 'internalism' and 'externalism' were introduced only in the middle of the twentieth century. It is often unclear how exactly these categories are to be understood, inviting confusion about their precise extension. (Which moral judgments are supposed to be motivationally effective? How are we to understand the notion of a motive that is present, but not necessarily overriding?) So it is often hard to classify moral philosophers as internalists or externalists as such.

Nevertheless, David Hume is presumably giving voice to an internalist point of view when he makes central to his theory the idea that 'morals excite passions, and produce or prevent actions' (1739/40: 457). In characteristic internalist fashion, Hume interprets this idea as placing the important constraint on an account of moral distinctions that it must explain how grasping moral distinctions could immediately give rise to passions and actions, in the way it seems to do. By contrast, the hermetic distinction drawn by John Stuart Mill (1861), between the proof of the principle of utility and the sanctions for compliance with it, suggests an externalist position. On this position, it would be possible to accept a moral principle without having any tendency to act in accordance with it; the question of the truth of a moral principle is thus taken to be completely distinct from the psychological issue of whether and how the principle can be rendered effective in people's motivations (see Hume, D. §4; Mill, J.S. §§7-8).

2 Implications

The issue of moral motivation is a crux in discussions about the nature and status of moral judgments and moral values. Noncognitivists, who hold that moral judgments express emotions or preferences rather than report genuine truths, often start out by affirming that moral judgment is directly motivating, a version of internalism (see Moral Judgment §1; Emotivism). Underlying the position that only noncognitivism can explain this fact is the assumption that the cognitive states through which we grasp judgments capable of genuine truth are motivationally inert.

In a more metaphysical vein, irrealists or error theorists, who deny the reality of moral values in the world, similarly argue from internalist assumptions. Real moral values would have to have the distinctive property of affecting the will directly. But, it is argued, items with this property would be mysterious-utterly unlike the other kinds of things with which we are familiar in the world-and so we should be reluctant to include them in the inventory of reality (see Moral realism §7).

Philosophers who wish to resist these conclusions may either accept internalism, and try to show that cognitivist or realist views can account for the motivational effects of ethical reflection, or reject internalism (see Moral realism §§4-5). Those who pursue the second option need not deny that people are often motivated to do what they judge to be morally right or good. Like Mill, however, they insist that motivations of this kind are not necessarily built into moral reflection; rather it is the task of moral education to supply them.

The interest of this dispute lies in its connection with fundamentally divergent ways of thinking about moral objectivity and the subject matter of ethics. Externalism goes naturally with the view that moral philosophy is essentially about moral theories, construed as sets of propositions setting out what it is morally right or good to do. On this conception, questions of objectivity in ethics turn primarily on the truth of the propositions that make up competing moral theories. But whether a given proposition is true or false seems independent of the psychological question of whether and how people may be moved to act in accordance with it. Taking this approach seriously, we may be led to the conclusion that the true moral theory should be esoteric, so that the theory itself provides good reasons for discouraging people from accepting it, a conclusion countenanced by Sidgwick and other utilitarians (see Sidgwick, H. §2).

Internalist approaches, by contrast, conceive of moral philosophy as dealing with common principles for public moral discourse and practical deliberation. On this view, the issue of objectivity in ethics is conceived primarily as an issue of practical reason, about the title of moral principles to regulate our activities. Moral reflection is accordingly thought of as a form of practical deliberation, yielding verdicts about what is to be done, and questions about the truth of moral propositions assume a subsidiary role. If we conceive of the subject in this way, then some form of internalism will seem inevitable, since it is the distinguishing mark of practical deliberation that it affects the will directly. Furthermore, it will hardly make sense to suppose that the criteria of right defined by the correct moral theory should be esoteric, since the kind of objectivity aimed at primarily in ethical theory is precisely an objectivity within practical reason (see Practical reason and ethics).

3 Reason and desire

Within the internalist camp, there are striking disagreements about the degree to which the motivational consequences of practical reason effectively constrain our philosophical options. On one side are the Humeans, who hold that practical reasoning must always be grounded in the noncognitive psychological states of individual agents - what Bernard Williams calls their ‘subjective motivational sets’ (1980) (see Williams, B.A.O. §3). Combined with internalism, this view has the consequence that moral principles are hostage to antecedent, empirical facts about the motivational sets of the agents to whom the principles apply. This conclusion in turn exerts some pressure in the direction of relativism, since motives to comply with moral principles may, as a matter of empirical fact, fail to show up in the subjective sets of all agents (see Moral relativism).

On the other side, Kantians deny that empirical facts about people’s subjective motivational sets place substantive restrictions on the content or normative force of moral principles. They point out that any theory of practical reason must allow for the phenomenon of motivational irrationality (see Akrasia). Thus I may fail to take steps that I know to be necessary if I am to attain some end that I hold dear. The principle that I ought to pursue those means that are necessary for realizing my ends does not entail that I will necessarily be motivated accordingly, but only that I will be motivated accordingly in so far as I am rational. This has been pointed out by Christine Korsgaard (1986). If she is right, then it cannot be inferred from the fact that a given agent lacks an empirical motive for compliance with a candidate principle that the principle is not binding on the agent, since it is possible that the agent is irrational. To decide whether that is the case, however, we will need to determine - in a way that is not circumscribed by antecedent facts about people’s motives - the content of principles of practical reason.

A central question in this dispute concerns the role of desires in the explanation of motivation. Humean resistance to the idea of practical principles that are not grounded in people’s subjective sets stems from the idea that

motivation is essentially a noncognitive orientation to the world. Kantians - most notably including Thomas Nagel (1970) - have granted that motivation always involves a state of desire, but suggested that reasoning in accordance with practical principles may give rise to new desires, so that the validity of such principles for a given agent need not be constrained by the items already contained in that agent's subjective motivational set (see Nagel, T. §5).

4 Explanation and justification

However this dispute is resolved, it cannot be denied that the disposition to respond to moral principles is reliably awakened and strengthened in the normal course of human psychological development. Furthermore, moral reasons often weigh heavily with people in their practical deliberations, leading them to take remarkable steps to avoid doing what would be wrong. It seems a reasonable constraint on moral theories that they should help us to make sense of these facts. A theory on which it appeared utterly mysterious that people naturally respond to moral considerations and take them seriously in their deliberations would be to that extent implausible.

What is needed is an account of the content of our moral incentives. Moral theories typically identify some common pattern of motivation, such as sympathy or self-interest, and try to show how motivations of that type might lead people to comply with principles of right, as the theory characterizes them. In practice, human motives are heterogeneous, complex and fluid; why should we suppose that there is any general account to be had of our responsiveness to moral considerations? This supposition is linked to two controversial normative assumptions. First, there is the assumption that moral considerations hang together in such a way that a uniform account can be provided of what makes actions morally wrong (in terms, for instance, of the effects of the actions on general utility, or the failure of universalizability). Philosophers who accept this assumption naturally expect to find a complementary general pattern of moral motivation, while those who deny it - some virtue theorists, perhaps - will be more sceptical.

Second, it is commonly assumed that modern morality aspires to a kind of objectivity within practical reason, providing public principles that make claims on virtually all agents. Here controversy centres not so much on whether morality represents itself as authoritative in this way, as on its title to such authority. How this controversy is resolved will depend in part on whether we can identify incentives to moral behaviour that are both widespread in human nature and intelligible as serious sources of reasons, for the availability of such common motivational structures would seem a necessary condition for the objective authority of morality.

It would not, however, be sufficient. If Nietzsche and Freud are right, for instance, then there are common motivational structures at the root of moral responsiveness, involving internal mechanisms for enforcing compliance with moral norms that draw on the redirection of aggressive impulses back against the self (see Nietzsche, F. §§8-9; Freud, S. §8). These accounts suggest a ready explanation of the urgency with which moral demands are invested in the deliberations of many agents; indeed, they have the further advantage of being able to explain the characteristic pathologies of the moral life, such as moralistic forms of extreme self-denial. But the explanations provided by these accounts seem to undermine rather than to vindicate the authority of morality, suggesting that the common patterns of moral motivation are essentially harmful for the individuals subject to them.

This shows that the authority of morality will depend not just on the availability of a common incentive to moral behaviour, but on the content of the common incentive that is identified. An account of moral motivation congenial to morality's normative ambitions would, at a minimum, identify a motive to moral behaviour that can be cultivated without necessarily causing dire psychological harm to the individual who has the motive. Beyond this, an effective moral motive should be capable of being integrated within an agent's overall system of ends, so as to produce a stable and self-reinforcing personality.

5 Egoism and self-interest

Very few philosophers accept egoism, construed as the psychological thesis that our motives are uniformly self-interested. This bald thesis flies in the face of the many occasions on which people act without any apparent concern for their own interests, however broadly conceived. Even when our motives are self-interested, the specification of our interests would seem to presuppose some core of nonegoistic concerns, a point well made by Joseph Butler (1726) (see Butler, J. §3; Egoism and altruism §3).

Still, self-interest is an undeniably widespread and powerful pattern of motivation, and this makes it an obvious candidate in terms of which to reconstruct moral concern. Thus, suppose that compliance with moral principles could be shown to conduce to the long-term interest of each individual. We would then have a ready explanation to hand of the ease with which moral incentives emerge in the course of psychological development, and of the urgency with which moral considerations present themselves in practical deliberation: these phenomena would reflect the natural concern of humans for their own wellbeing, and the equally natural use of practical reason in the service of this concern. This explanation would in turn support rather than undermine the authority of morality. Those who make morality regulative of their activities would thereby achieve integration and stability of personality in an exemplary degree, since morality, far from being inimical to our other interests, would represent the condition for their effective pursuit.

This broadly Hobbesian strategy for understanding morality rests on the idea that each of us benefits enormously from the availability of certain public goods - such as peace, security, and trustworthiness - that can be secured only through general compliance with moral norms (see Prudence §2; Hobbes, T. §5). Discussions of the strategy have become highly sophisticated, drawing on advances in the theory of rational choice (see Rational choice theory). But the increasing sophistication of its formulation cannot compensate for the inadequacy of the strategy as an account of moral motivation. The strategy might show that something like moral norms make claims on all of us. But to the extent that our incentive to comply with these norms is self-interest, the norms themselves seem to lose their distinctively moral character, for it is part of our understanding of moral requirements that they are to be followed even when it is inexpedient for us to do so. The problem is made vivid by the figure of the prospective ‘free rider’, who asks why they should comply with social norms when sufficiently many other people follow them to secure the public goods of cooperation. However this question may be answered, the very fact that it needs to be asked shows that the prospective free rider is not an ordinary moral agent at all.

6 Altruism and sympathy

A genuinely moral agent - as opposed to an egoist - is altruistic, in the sense of having some immediate concern for the interests of others. Furthermore, the capacity for sympathetic identification seems natural to human beings; it is the rare person who is not directly moved by exposure to human suffering, and this tendency to sympathetic response can be nurtured through moral education. For these reasons, many have followed Hume and Schopenhauer (1840) in taking sympathy to be the paradigmatic moral motivation (see Hume, D. §4; Schopenhauer, A. §6; Moral sentiments §1).

If sympathy is to play this role, however, then it will need to be corrected and refined. Actual sympathy is notoriously partial and erratic, directed (for instance) at those individuals with whom one has some personal connection, whereas morality seems to demand impartiality of response. Thus, whether morality requires that I help a given person should not depend just on whether I happen to know them, but also (for instance) on the comparative urgency of the claims of others. The standard way to deal with this problem is to subject the mechanism of sympathetic identification to the discipline of impartial reflection. One is to adopt the perspective of an informed spectator, abstracting from one’s knowledge of one’s own position in the world, and treating each person’s welfare as equally important; the morally right action is what sympathy, refined by this procedure of reflection, would then lead one to approve of. Though liable to many different interpretations, this general approach to moral motivation has been associated traditionally with utilitarianism (see Impartiality §§1-2; Utilitarianism).

The utilitarian approach yields an impersonal conception of ethical impartiality, on which moral requirements take an agent-neutral rather than agent-relative form. Consider the familiar moral prohibition on murder. This is ordinarily understood as an agent-relative requirement, proscribing my killing another person even if by doing so I could prevent a number of other murders. But a prohibition of this kind would not survive utilitarian impartial reflection. Once we abstract from our knowledge of our local position in the world, it is no longer possible to sustain a proprietary concern for the character of our own actions. When sympathy is subjected to a filter of impersonal reflection, it can only endorse requirements that are agent-neutral in form, and hence capable of being accepted and acted on by any other agent as well (an injunction to prevent murder, say, instead of the conventional agent-relative prohibition on killing people oneself). In this way, the impersonal interpretation of impartiality defines a standpoint that is available to be occupied by any moral agent.

Questions arise, however, about the authority and motivational effectiveness of the claims made from this point of view. To take the latter first, it is supposed to be an advantage of the utilitarian approach that it accounts for moral motivation in terms of the familiar operations of sympathy. But there is a vast difference between the immediate, impulsive sympathy that develops naturally in human beings and the impersonal concern to maximize the good. The utilitarian claims that sympathy is transformed into a responsiveness to agent-neutral reasons as a result of moral reflection, but one may wonder whether sympathy, thus transformed, remains a natural pattern of motivation.

As for authority, the question is whether impersonal sympathy can be integrated into a person's system of ends. It gives us a standing first-order aim - the maximization of the good - which is set over against our other first-order aims, and likely to conflict fundamentally with them (given the demands that the maximization of the good notoriously imposes on individuals). The result is that we must either radically curtail the pursuit of our ordinary interests for the sake of morality, or resign ourselves to living in a way that is not morally justifiable. Utilitarians reply that morality is a hard thing, and that we should not expect to be able to accommodate its demands within the contours of ordinary bourgeois life. But the difficulty of integrating impersonal sympathy within a normal system of human ends is likely to raise a doubt about the authority of a morality construed in these terms: by what right does it demand this degree of sacrifice of us?

7 Principle-dependent motives

Kantian approaches typically trace moral concern to a higher-order motive, such as the concern to act only in ways that are permitted by moral principles (see Kantian ethics). Motives of this type will sometimes give an agent first-order ends, namely when moral principles require a certain course of action (keeping a promise, say). But when moral principles do not impose requirements in this way it will be open to agents to pursue their first-order ends while at the same time completely satisfying their higher-order moral concern. Morality regulates our first-order pursuits, but because the moral motive does not automatically supply a standing first-order end it is not inevitably in conflict with those pursuits, and so the possibility of integrating moral concern within a realistic system of human ends is left open.

Whether this possibility is realized will depend on the content of moral principles. If those principles instructed us always to maximize the good, then they would in effect determine a standing first-order end. In that case the higher-order concern to act in ways that are permitted by moral principles might be as difficult to integrate with our other ends as impersonal sympathy would be. But Kantians reject such maximizing interpretations of morality, taking moral requirements to derive from reflection on the universalizability of our aims, and contending that these universality procedures yield specific agent-relative obligations that function as limiting conditions on our activities (see Universalism, ethical §5).

The idea that morality supplies a higher-order motive invites the charge that it is self-indulgent, reflecting a fastidious concern for one's own virtue rather than a direct responsiveness to the needs of others. Further, the involvement of moral principles in Kantian moral motivation seems to deny moral worth to spontaneous and heartfelt actions, and threatens to erect an alienating screen of reflection between moral agents and their projects. A more fundamental question is whether the motive of duty can be rendered intelligible, as a natural human incentive. Why should the concern to act in ways that are universalizable emerge so readily in moral development, and present itself to us as a serious source of reasons?

To this, Kant (1785) replies that the motive of duty is constitutively rational, and that acting on it enables us to realize the supreme good of autonomy (see Kant, I. §9). Somewhat less ambitiously, T.M. Scanlon (1982) observes that people commonly want to be able to justify their conduct to others on grounds that could not reasonably be rejected, arguing that moral principles tell us what we have to do to satisfy this familiar desire. Either of these strategies might derive support from the readiness with which people respond to 'golden rule' arguments, in which we invite them to consider what it would be like to change positions with those who would be affected by their actions (see Universalism, ethical §4). This phenomenon suggests that the concern to act in ways that are universalizable may be a powerful and *sui generis* pattern of human motivation.

See also: Moral justification §3; Moral motivation

R. JAY WALLACE

References and further reading

Baier, A.C. (1987) 'Hume, the Women's Moral Theorist?', in E. Kittay and D. Meyers (eds) *Women and Moral Theory*, Lanham, MD: Rowman and Littlefield, 37-55; repr. in *Moral Prejudices: Essays on Ethics*, Cambridge, MA: Harvard University Press, 1994, 51-75.(Offers an account of reflective sympathy interestingly different from the standard utilitarian interpretation discussed in the text.)

Butler, J. (1726) *Fifteen Sermons Preached at the Rolls Chapel*, Sermons I, II, III, XI, XII; repr. in S. Darwall (ed.) *Five Sermons Preached at the Rolls Chapel and A Dissertation Upon the Nature of Virtue*, Indianapolis, IN: Hackett Publishing Company, 1983, Sermon XI.

Foot, P. (1985) 'Utilitarianism and the Virtues', *Mind* 94: 196-209; repr. in S. Scheffler (ed.) *Consequentialism and its Critics*, Oxford: Oxford University Press, 1988, 224-42.(Suggests that benevolence and sympathy have a place within the moral life, as one of the virtues, but that they are distorted when made the global basis of moral motivation.)

Freeman, S. (1991) 'Contractualism, Moral Motivation, and Practical Reason', *Journal of Philosophy* 88: 281-303.(Sophisticated contractualist account of moral motives as highest-order desires that are principle-dependent, and based in practical reason.)

Herman, B. (1981) 'On the Value of Acting from the Motive of Duty', *Philosophical Review* 90: 359-82; revised version repr. in *The Practice of Moral Judgment*, Cambridge, MA: Harvard University Press, 1993, 1-22. (Argues that the Kantian motive of duty functions as a higher-order limiting condition, leaving scope for action on other first-order ends.)

Herman, B. (1983) 'Integrity and Impartiality', *Monist* 66: 233-50; revised version published as 'Rules, Motives, and Helping Actions', *Philosophical Studies* 45 (1984): 369-77; revised version repr. in *The Practice of Moral Judgment*, Cambridge, MA: Harvard University Press, 1993, 23-44.(Defends Kantian morality against the charge that the commitment to moral principle would alienate us from our projects and emotional attachments.)

Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978.(Book III emphasizes the motivational consequences of moral thought, and develops an account of moral motivation in terms of sympathy; book II, part 3 argues that practical reason has its roots in our desires.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. L.W. Beck, *Foundations of the Metaphysics of Morals*, Indianapolis, IN: Bobbs-Merrill, 2nd edn, 1995, section 2.(Argues that the motive of duty is constitutively rational, and that acting on it enables us to realize the supreme good of autonomy.)

Korsgaard, C. (1986) 'Skepticism about Practical Reason', *Journal of Philosophy* 83 (1): 5-25.(Important discussion of internalism; denies that the motivational dimension of practical reason places constraints on the content of practical principles.)

Mill, J.S. (1861) *Utilitarianism*, ed. G. Sher, Indianapolis, IN: Hackett Publishing Company, 1979.(Distinguishes between the question of the truth of a moral principle and the question of the sanction or motive for compliance with it.)

Nagel, T. (1970) *The Possibility of Altruism*, repr. Princeton, NJ: Princeton University Press, 1978.(Challenging and suggestive argument that practical reason can give rise to new motivations, and that altruism is based in reflection from an impersonal point of view.)

Nagel, T. (1986) *The View from Nowhere*, New York: Oxford University Press.(Chaps 8 and 9 discuss the prospects for objectivity in practical reason, and the differences between agent-centred and agent-neutral requirements, while ch. 10 deals with the prospects for reconciling moral concern with one's personal ends; helpful bibliography.)

Scanlon, T.M. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press, 103-28.(Interprets moral motivation in terms of the desire to be able to justify one's conduct to others on grounds they could not reject; includes a good general discussion of the demands made by moral motivation on moral theory.)

Scheffler, S. (1992) *Human Morality*, New York: Oxford University Press.(Subtle and wide-ranging discussion of naturalistic approaches to moral motivation and their implications for the content and authority of moral demands; contains extensive bibliographic references.)

Schopenhauer, A. (1840) *Über die Grundlage der Moral*, trans. E.F.J. Payne, *On the Basis of Morality*, Providence, RI and Oxford: Berhahn Books, 1995, sections 16, 22.(Suggests that sympathy is the paradigmatic moral motivation.)

Smith, M. (1994) *The Moral Problem*, Oxford: Blackwell.(Sophisticated discussion of internalism and how it functions in theoretical and metaphysical debates in ethics; helpful bibliography.)

Williams, B. (1976) 'Persons, Character and Morality', in A.O. Rorty (ed.) *The Identities of Persons*, Berkeley,

CA: University of California Press, 197-216; repr. in *Moral Luck: Philosophical Papers 1973-80*, Cambridge: Cambridge University Press, 1981, 1-19.(Influential argument that Kantian impartial motivations cannot be made regulative of one's projects without alienating one unacceptably from them.)

Williams, B. (1980) 'Internal and External Reasons', in R. Harrison (ed.) *Rational Action*, Cambridge: Cambridge University Press; repr. in *Moral Luck: Philosophical Papers 1973-80*, Cambridge: Cambridge University Press, 1981, 101-13.(Contends that practical reasons must be grounded in an agent's 'subjective motivational set'.)

Moral particularism

Moral particularism is a broad set of views which play down the role of general moral principles in moral philosophy and practice. Particularists stress the role of examples in moral education and of moral sensitivity or judgment in moral decision-making, as well as criticizing moral theories which advocate or rest upon general principles. It has not yet been demonstrated that particularism constitutes an importantly controversial position in moral philosophy.

Moral particularism is the view that general moral principles play less of a role in moral thought than has often been claimed. In its most extreme form, particularism states that there are no genuine moral principles, and that therefore moral agents who attempt to guide their action by reference to moral principles, and philosophers who attempt to construct moral theories based on principles, are seriously mistaken (see Moral realism §5; Situation ethics).

Many particularists are influenced by the view of Wittgenstein that one acquires a concept not by being taught some universal rule for its application, but through introduction into a human practice and a way of seeing things (see Wittgensteinian ethics §3). The particularist view of moral education will stress the importance of examples and actual experience of individual moral cases rather than the learning of universal moral rules under which particular cases can be subsumed (see Examples in ethics §2; Moral education §3).

A link is often drawn between moral particularism and so-called 'antitheory' in ethics. Antitheorists suggest that ethical theorists are in error in postulating principles according to which actions are right to the extent that they are in accord with these principles. Utilitarianism, for example, claims that acts are right to the extent that they maximize utility (see Utilitarianism). A further link is often alleged between particularism, anti-theory and virtue ethics (see Virtue ethics). But this link may rest on a confusion between particularism about moral theory and particularism about moral agency. Virtue ethics has its own principle: right actions are those that the virtuous person would do. The virtuous person will indeed not in practice proceed by attempting to apply this principle directly. Here, however, virtue ethics and utilitarianism are in agreement, since most utilitarians have claimed that moral agents should not attempt to apply utilitarianism in practice.

Moral particularism can also emerge out of the theory of reasons for action. On the view of Dancy (1993), reasons are not universalizable across cases, so that what counts as a reason in one case need not be assumed to function as a reason in the same way in other cases (see Logic of ethical discourse §7). My enjoying giving you a present counts in favour of my action; but my enjoying torturing you counts against. Against this, it can be suggested that reasons are universalizable at a higher level (*innocent* pleasure, perhaps, always counts as a reason), and that to deny this is to embrace a form of irrationalism.

According to less extreme particularists, principles can play some role in theory and in practice (see Casuistry). On one view, they serve as useful generalizations, but there is always a need for judgment in particular cases (see Moral judgment §2; Theory and practice §2). Aristotle is best seen as such a particularist. Once again, it is not clear that, for example, utilitarians or Kantians would wish to deny such a role for judgment (see Kantian ethics; Universalism in ethics §3).

See also: Aesthetics and ethics

ROGER CRISP

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985, book 6.(Contains an account of moral judgment or 'practical wisdom'.)

Dancy, J. (1993) *Moral Reasons*, Oxford: Blackwell.(Central outline and defence of modern particularism. Difficult.)

McDowell, J. (1979) 'Virtue and reason', *Monist* 62: 331-50. (Defence of a form of particularist virtue ethics involving both Aristotle and Wittgenstein. Difficult.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Clarendon Press, chaps 1-2. (Has been influential on the development of modern particularism.)

Moral pluralism

Moral pluralism is the view that moral values, norms, ideals, duties and virtues are irreducibly diverse: morality serves many purposes relating to a wide range of human interests, and it is therefore unlikely that a theory unified around a single moral consideration will account for all the resulting values. Unlike relativism, however, moral pluralism holds that there are rational constraints on what can count as a moral value. One possible, though not necessary, implication of moral pluralism is the existence of real moral dilemmas. Some philosophers have deemed these to be inconceivable; in fact, however, they do not constitute a serious threat to practical reason. Another possible implication of moral pluralism is the existence within a society of radically different but equally permissible moralities. This poses a challenge for political philosophy, and might justify a liberal view that particular conceptions of the good life ought not to be invoked in the formulation of public policy.

1 Moral pluralism and moral theory

Moral pluralism is the view that moral values, norms, ideals, duties and virtues cannot be reduced to any one foundational consideration, but that they are rather irreducibly diverse. As such, moral pluralism is a metaphysical thesis, in that it tells us what moral considerations there are. Pluralist moral philosophers disagree as to exactly what the plural sources of moral value are. For example, Sir David Ross (1930) distinguished six species of duty, including duties of fidelity and reparation, of gratitude, of justice, of beneficence, of self-improvement and of non-maleficence; and Thomas Nagel (1979) has claimed that the conflicts among diverse moral principles are due to there being five distinct sources of value - special allegiances, universal rights, utility, perfectionist ends of self-development, and individual projects (see Nagel, T. §5). Despite the differences between these accounts of the sources of moral value, the resolution of which constitutes a challenge for substantive moral theory, these thinkers can be seen as united in the view that morality has developed to protect and promote basic interests related to human wellbeing and flourishing, but that since there is no unique form that human wellbeing must take, there can consequently not be a theory of morality unified around one supreme value (see Happiness §3; Welfare).

This is not to say that the truth of moral pluralism disqualifies any attempt at formulating a moral theory (see Morality and ethics §§1-2). Among the many moral values which human beings pursue, there are undoubtedly some that can be grouped together and accounted for in terms of some more general value relevant to the particular set of human interests with which they are all in one way or another concerned. Moral pluralism implies simply that none of these values could plausibly claim hegemony over the entire set of moral considerations.

Historically, moral pluralism has been linked with controversial positions in moral epistemology and the ontology of value, according to which moral facts are real and non-natural, and are given as self-evident to a distinct human faculty of moral intuition (see Moral realism; Intuitionism in ethics). It is in fact compatible with a wide range of philosophical positions on these issues, including anti-realism and naturalism.

Some philosophers have argued that the diversity of our moral concepts is a distinctive feature of modernity. Alasdair MacIntyre, for example, has claimed that the plurality of conflicting considerations which make up the moral lexicon of the modern agent is a sign of cultural decay. Modern morality is for MacIntyre a congeries of concepts which have been inherited by modern agents from past forms of life, but which have been torn from the coherent concrete human practices within which they originated, and in the context of which alone they have any real meaning. Moral pluralism is therefore in his view a symptom of our moral discomfiture.

Moral pluralism has also been challenged by defenders of classical single-principle moral theories. Yet there have also been signs of theoretical rapprochements. Indeed, many modern consequentialists are abandoning the simple view of human wellbeing embodied in classical utilitarianism in favour of more multifaceted accounts (see Consequentialism). And many deontologists can be read as formulating rational priority rules ranking deontological constraints over other types of moral considerations, rather than as banishing the latter completely from the realm of the moral (see Deontological ethics).

2 Relativism and moral dilemmas

It is important to distinguish moral pluralism from a thesis with which it has too often been confused, namely that of moral relativism. A relativist claims that the truth of moral judgments is relative to the conventions of the social

group (or even to the individual whim) of the person issuing the judgment, and that these conventions or whims are not themselves subject to any further criterion of adequacy. There are therefore according to relativists no rational constraints on what can count as a moral value, and it is therefore senseless in their view to speak of the truth, falsity or justification of moral judgments (see Moral relativism). Moral pluralism in contrast holds that while the variety of moral principles applying to human beings is irreducible, it is not infinite. Rather, there are constraints on what can count as a moral value (and there is therefore sense in speaking of moral truth and falsity). These constraints might, for example, have to do with the (inherently diverse but not infinite) forms which human wellbeing and flourishing can take.

Moral pluralism is therefore compatible with the existence of rational constraints upon moral thought.But the fact that a number of statements to do with moral value might all be true while apparently recommending incompatible actions, and that real, as opposed to merely apparent, moral dilemmas emerge as a real possibility, has been thought by some philosophers to disqualify it as a theory of value by making it sin against basic axioms of deontic logic. These include the principle that ‘ought implies can’ and the principle of agglomeration, which states that if I ought to do A and I ought to do B, then I ought to do A and B. This troubling apparent consequence of moral pluralism must, however, be qualified by a number of observations. First, as Michael Stocker has observed (1990), the premise that statements about moral value are always act evaluations is an unvindicated assumption of much modern moral theory, yet moral pluralism only leads to moral dilemmas if this assumption is granted. There might be a number of true moral descriptions of a situation, emphasizing different moral considerations present in it. Any one of these might well on its own give rise to an ‘ought’ statement, but given the presence of other moral considerations it may give rise only to what Ross has called a *prima facie* obligation. As the latter are not directly action-guiding, they need not conform to the strictures of deontic logic. Second, certain axioms of deontic logic might actually embody controversial first-order moral propositions. The fact that they conflict with the hypothesis of moral dilemmas does not therefore automatically place the burden of proof upon defenders of the latter. Third, the reality of moral dilemmas need not, as philosophers such as Bernard Williams (1965) have suggested, put paid to all attempts to rationally order our, at times conflicting, moral values. There are reasons supporting both sides of a moral dilemma, and the presence of a moral dilemma, rather than signalling the necessary end of moral inquiry, can point to the need to undertake inquiry into these reasons in a more fine-grained manner. Moral pluralism involves the denial of the existence of a supreme value from which all others might be derived; it does not entail incommensurability, the view that moral considerations cannot be compared and ranked. Thus, for example, there may be rational priority rules allowing us to order the claims of different moral values.

3 Moral pluralism and political philosophy

The plurality of moral values can manifest itself in a number of different ways. Most relevantly from the point of view of political philosophy, it can involve the existence within a society of a number of equally acceptable moral forms of life. This form of social pluralism poses a set of challenges for political philosophers, suggesting that there may be no simple way of adjudicating conflicts between adherents of equally admirable moral forms of life, or of engaging in the interpersonal welfare comparisons often seen as necessary for the formulation of theories of distributive justice. Moral pluralism has been seen by many philosophers, including John Rawls (1971), Thomas Nagel and Charles Larmore (1987), as calling for the liberal doctrine of state neutrality, the view that particular conceptions of the good ought not to be invoked in the formulation of public policy.

See also: Axiology; Duty; Ideals; Pluralism; Religious Pluralism; Values; Virtues and vices

DANIEL M. WEINSTOCK

References and further reading

Berlin, I. (1969) *Four Essays on Liberty*, Oxford: Oxford University Press.(The classic twentieth-century statement of moral pluralism.)

Larmore, C. (1987) *Patterns of Moral Compexity*, Cambridge: Cambridge University Press.(Argues that a purely political conception of liberalism flows from the plurality of moral forms of life in a society.)

MacIntyre, A. (1984) *After Virtue*, Notre Dame, IN: Notre Dame University Press, 2nd edn.(Argues that moral pluralism and conflict result from moral concepts no longer being embedded in concrete social forms.)

Nagel, T. (1977) ‘The Fragmentation of Value’, in H.T. Englehardt, Jr. and D. Callahan (eds) *Knowledge, Value and Belief*, Hastings-on-Hudson, NY: Institute of Society, Ethics and the Life Sciences; repr. in *Moral*

Questions, Cambridge: Cambridge University Press, 1979.(A clear statement of the different sources of moral value.)
Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, 34-40.(A standard modern argument against a form of moral pluralism identified as ‘intuitionism’.)
Ross, W.D. (1930) *The Right and the Good*, Oxford: Oxford University Press, 21. (The classic statement of an intuitionist moral pluralism.)
Stocker, M. (1990) *Plural and Conflicting Values*, Oxford: Oxford University Press. (Argues that value pluralism does not threaten the possibility of sound practical reason.)
Williams, B. (1965) ‘Ethical Consistency’, *Proceedings of the Aristotelian Society*, supplementary vol. 39; repr. in *Problems of the Self: Philosophical Papers 1956-72*, Cambridge: Cambridge University Press, 1973.(Raises problems for moral reasoning caused by the plurality of values.)

Moral psychology

Moral psychology as a discipline is centrally concerned with psychological issues that arise in connection with the moral evaluation of actions. It deals with the psychological presuppositions of valid morality, that is, with assumptions it seems necessary for us to make in order for there to be such a thing as objective or binding moral requirements: for example, if we lack free will or are all incapable of unselfishness, then it is not clear how morality can really apply to human beings. Moral psychology also deals with what one might call the psychological accompaniments of actual right, or wrong, action, for example, with questions about the nature and possibility of moral weakness or self-deception, and with questions about the kinds of motives that ought to motivate moral agents. Moreover, in the approach to ethics known as 'virtue ethics' questions about right and wrong action merge with questions about the motives, dispositions, and abilities of moral agents, and moral psychology plays a more central role than it does in other forms of ethical theory.

1 Psychology and the possibility of morality

We can divide the main traditional concerns of moral psychology as a distinctive philosophical discipline into questions about the psychological assumptions necessary to the validity of morality, or moral rules, in general, and questions about the psychological concomitants or underpinnings of particular actions evaluated either as good/right or as bad/wrong. The question of whether human beings have free will or freedom of choice (see Free will) naturally falls within the first of these areas. If human beings lack free will, then, it has over the millennia typically (though not universally) been assumed, they cannot be held responsible for their actions and cannot be bound by moral obligations any more than are animals or small children. Thus those who have systematically elaborated one or another view of moral right and wrong have usually thought it necessary to defend, or at least explicitly assume, the existence of human free will, or freedom of choice.

In the first instance this standardly involves saying something about free agency in relation to universal causal/nomic determinism. If the world is universally governed by causal or physical laws, then it is unclear how anyone could possibly have behaved otherwise than they in fact did, and so have been responsible for what they did. So defenders of (objective) morality typically feel called upon either to argue that human beings are in important ways not subject to iron-clad causal determination, or else to show that causal determinism does not in fact deprive us of free will (see, for example, Watson 1982; Determinism and indeterminism).

Another metaphysical or quasi-metaphysical issue that looms large in moral psychology concerns the human capacity for morality. Most moral codes and moral philosophies require, for example, that people occasionally put aside self-interest in the name of honour, fairness, decency, compassion, loyalty or the general good. But if one believes in psychological egoism, one will hold that people lack the capacity for these forms of self-sacrifice, and it then becomes problematic whether human beings really have the obligations that various non-egoistic moral views/theories claim that they do (see Egoism and altruism).

But even if one rejects both psychological and ethical egoism (for example, Butler 1950), there are moral-psychological issues about how much morality can fairly be demanded of people. These issues arise especially in connection with utilitarianism (see Utilitarianism) and Kantianism (see Kantian ethics). Utilitarianism is usually stated in a 'maximizing' form that treats it as a necessary and sufficient condition of right action that one do the best/most one circumstantially can to advance the happiness of humankind (or sentient beings). But such a doctrine seems to entail that if one is in a position to relieve the suffering, hunger or disease of others, one is morally obliged to do so, even if that means giving up one's own life plans and most of what one really cares about in life. The traditional utilitarian moral standard is thus very demanding, and some philosophers have questioned whether morality can properly, or, one might say, fairly, require so much of people (see, for example, Stocker 1976; Swanton 1993). In particular, it may be wondered whether most people have the capacity to live up to such a stringent morality as is presented by maximizing utilitarianism.

Certain doctrines of Kantian ethics can likewise be seen as going against the grain of human nature or capacities, not by demanding too much of a sacrifice of self-interest, but by laying down rather stringent or narrow psychological conditions for the moral admirability of actions. According to Kant, if one helps someone in need or in trouble out of fellow-feeling or friendship, one's act lacks all moral worth, because it was not performed out of a

sense of duty and respect for the moral law. And many ethical thinkers have either implicitly or explicitly held such a view of moral virtue to be too narrow and out of keeping with realistic human psychology, and have argued that certain primary, immediate or natural motives such as compassion or friendly feeling not only have moral worth, but can often be even more praiseworthy than a cool appeal to duty or to something as abstract as the moral law (see, for example, Martineau 1891; Stocker 1976). But Kantians question the durability and reliability of mere feeling or emotion, and there remains considerable debate in this area (Blum 1980).

2 Psychology and moral judgment

Another topic that has recently occupied moral psychologists and ethicists is the relevance of so-called ‘moral emotions’ to moral judgments. Some philosophers have considered the validity of one or another moral theory to depend, in part, on whether people tend to feel guilty for violating its dictates (see, for example, Williams 1981). But others have held that there are situations where guilt is inevitable, but that that fact constitutes no sort of evidence of wrongdoing. Indeed, it is sometimes argued that the question of whether guilt for some action is *appropriate* or *justified* is separate from the issue of whether that action was actually wrong (Greenspan 1988). (Is it, for example, appropriate to feel guilty about injuries one caused in a traffic accident that was someone else’s fault entirely?) Other philosophers, however, seek to connect the moral emotions with morality by arguing that the seeming *ineradicability* of our emotions of anger and resentment provides some sort of underpinning for human judgments about justice and injustice, praiseworthiness and blameworthiness (see Strawson in Watson 1982; Morality and emotions; Moral sentiments).

In addition, moral psychologists concern themselves with various forms of situational moral failure, such as self-deception and weakness of will (see Self-deception; Akrasia). The very existence of these phenomena is often questioned on grounds of their paradoxical character. If one is motivated to deceive oneself about some difficult matter (for example, the needs of one’s children), how can one possibly succeed unless one knows what it is one has to deceive oneself about - in which case, in what way is one actually deceived? Similarly, if moral weakness, or weakness of will more generally, means allowing anger or passion to lead one to act against one’s better judgment, is it not natural to reconfigure our understanding of what is happening so as to suppose that the reasons for anger or passion warp one’s sense of what really is best (for one) - in which case, in what way has one actually acted against one’s (however momentary) better judgment?

Both weakness of will and self-deception are, therefore, inherently problematic, and moral psychology has sought ways of either making sense of these moral phenomena or showing them definitively to be incoherent and consequently impossible (see, for example, Pears 1984). Either ‘solution’ has bearing on the moral assessment of actions (and desires). If self-deception is possible, what sort of blame attaches to it when it leads to bad consequences (for example, neglect of children)? Should it be counted as a deliberate, intentional act; or it is more like negligence, or heedlessness? But if self-deception is not possible, then what masquerades as such may be more like intentional wrongdoing, and perhaps even more blameworthy, than it initially appears (see Self-deception, ethics of). Then too, if weakness of will is not possible, many cases we are inclined (at most) partially to excuse on grounds of weakness may be cases in which the agent is actually compelled, psychologically, to act wrongly - or else perhaps momentarily ignorant of the difference between right and wrong. Such conclusions are bound to affect the character and/or severity of the blame or punishment directed at those who initially seemed to be acting weakly.

Another set of issues in moral psychology arises within the tradition of moral theory known as virtue ethics, which treats issues of moral psychology as essential to our understanding of right and wrong action (see Virtue ethics). Thus, rather than basing morality on moral rules or on the production of ‘good consequences’, Aristotelian virtue ethics denies the possibility of universal moral rules and thinks of the virtuous individual as someone who intuitively perceives what is right or noble in various situations and, fairly effortlessly, acts accordingly. To that extent, the Aristotelian tradition in virtue ethics approaches the idea of moral rightness in some measure indirectly, by focusing on the character, habits and abilities of the virtuous individual who tracks rightness in thought and actions (see Feminist ethics).

In addition, there is a more radical tradition of virtue ethics in which the moral evaluation of actions and moral rules is directly derived from characterizations of good and bad motivation and how these may be expressed or realized in someone’s actions. Perhaps the best-known historical example of such an approach can be found in

Plato's *Republic* (Book IV), where it is said that good actions are those that enhance or support the health and harmony of the soul (see Plato §14).

A more recent and 'purer' example (because it does not mention anything like the Form of the Good) can be found in James Martineau's *Types of Ethical Theory* (1885, 1891). Martineau ranks all human motives on an absolute scale - for example, compassion is placed above ambition, the latter above sexual desire, and the latter, in turn, above vindictiveness - and claims that right action is action that comes from the higher or highest motive operating in any given situation of moral choice. This then allows moral rules of thumb to be derived from generalizations about which motives are likely to operate in various morally familiar situations.

These varying virtue-ethical approaches put moral psychology at the very centre of ethics. However, for the foreseeable future, the chief role of moral psychology will probably be as a discipline ancillary to, rather than as the main focus of, moral theory.

MICHAEL SLOTE

References and further reading

Baron, M. (1984) 'The Alleged Moral Repugnance of Acting from Duty', *Journal of Philosophy* 81: 197-219.(A critical examination of views about the moral worth of various kinds of motivation.)

Blum, L. (1980) *Friendship, Altruism and Morality*, London: Routledge.(An extended critique of Kantian ideas about the moral significance of the emotions.)

Broadie, S. (1991) *Ethics with Aristotle*, New York: Oxford University Press.(A thoroughgoing examination of Aristotle's ethics.)

Butler, J. (1950) *Fifteen Sermons Preached at the Rolls Chapel*, New York: Bobbs-Merrill.(A classic refutation of the arguments for psychological egoism.)

Garcia, J. (1993) 'The New Critique of Anti-Consequentialist Moral Theory', *Philosophical Studies* 71: 1-32.(An account of virtue ethics that bases morality in motives, rather than consequences.)

Greenspan, P. (1988) *Emotions and Reasons: An Inquiry into Emotional Justification*, New York: Routledge.(A full-scale examination of the rational status of emotions.)

Hursthouse, R. (1991) 'Virtue Theory and Abortion', *Philosophy and Public Affairs* 20: 223-46.(A defence of neo-Aristotelian virtue ethics.)

Kagan, S. (1989) *The Limits of Morality*, Oxford: Oxford University Press.(A defence of utilitarianism generally and against the charge that it is 'too demanding'.)

Martineau, J. (1885, 1891) *Types of Ethical Theory*, Oxford: Clarendon Press, 2 vols.(The classic defence of motive-based morality.)

Pears, D. (1984) *Motivated Irrationality*, Oxford: Oxford University Press.(An extended analysis of weakness of will and other forms of irrationality.)

Slote, M. (1993) 'Virtue Ethics and Democratic Values', *Journal of Social Philosophy* 24: 5-37.(A defence of current-day political ideals in virtue-ethical terms.)

Stocker, M. (1976) 'The Schizophrenia of Modern Ethical Theories', *Journal of Philosophy* 73: 453-66.(A critique of moral theories that make happiness difficult or that treat conscientiousness as the morally best of motives.)

Swanton, C. (1993) 'Satisficing and Virtue', *Journal of Philosophy* 90: 33-48.(A critique of the 'overdemandingness' of Aristotelian virtue ethics and various theories of practical reason.)

Watson, G. (ed.) (1982) *Free Will*, Oxford: Oxford University Press.(An anthology of recent articles on free will and moral responsibility.)

Williams, B. (1981) *Moral Luck*, Cambridge: Cambridge University Press.(A collection of essays that, among other things, questions the realism of much recent ethical theory.)

Moral realism

Moral realism is the view that there are facts of the matter about which actions are right and which wrong, and about which things are good and which bad. But behind this bald statement lies a wealth of complexity. If one is a full-blown moral realist, one probably accepts the following three claims.

First, moral facts are somehow special and different from other sorts of fact. Realists differ, however, about whether the sort of specialness required is compatible with taking some natural facts to be moral facts. Take, for instance, the natural fact that if we do this action, we will have given someone the help they need. Could this be a moral fact - the same fact as the fact that we ought to do the action? Or must we think of such a natural fact as the natural 'ground' for the (quite different) moral fact that we should do it, that is, as the fact in the world that makes it true that we should act this way?

Second, realists hold that moral facts are independent of any beliefs or thoughts we might have about them. What is right is not determined by what I or anybody else thinks is right. It is not even determined by what we all think is right, even if we could be got to agree. We cannot make actions right by agreeing that they are, any more than we can make bombs safe by agreeing that they are.

Third, it is possible for us to make mistakes about what is right and what is wrong. No matter how carefully and honestly we think about what to do, there is still no guarantee that we will come up with the right answer. So what people conscientiously decide they should do may not be the same as what they should do.

1 Realism, objectivism, cognitivism

These three terms are hard to keep separate, but it is worth the effort. Cognitivism is the claim that moral attitudes are cognitive states rather than noncognitive ones. The distinction between cognitive and noncognitive states is not clear; the best way of drawing it is by appeal to the distinction between two 'directions of fit'. Beliefs, which are the paradigm examples of a cognitive state, have one direction of fit; desires, which are the paradigm examples of a noncognitive state, have the other. A belief, that is, has to fit the world; the world is given, as it were, and it is the belief's job to fit that world, to get it right. A desire is not like that; the desire's job, if anything, is to get the world to fit it, to make things be the way it wants them to be. Crucially, a desire is not at fault if things are not as it wants them to be; a belief is at fault if things are not as it takes them to be. The question whether moral attitudes are cognitive states or noncognitive ones is the question whether they have the direction of fit of a belief, or that of a desire. They could, of course, be complex states with a mixture of both; but noncognitivism is the view that moral attitudes have either wholly or partly the direction of fit of a desire. This is normally expressed more briefly as the view that moral attitudes either are or at least contain desires; to think an action right is a sort of 'pro-attitude', and pro-attitudes are wantings (see Belief; Desire; Moral Judgment §1).

Realists, believing that there are distinct moral facts, are likely to be cognitivists, since the appropriate attitude to a fact is belief rather than desire. It is for this reason that the opposition to realism is normally called 'noncognitivism'. Realists, holding there to be moral facts, maintain for that reason that moral attitudes are beliefs; noncognitivists, holding that moral attitudes either are or include desires, claim for that reason that there are no facts to be the objects of those attitudes.

Objectivism is harder to distinguish from realism, since the two are very closely linked. Objectivity is something to do with independence from us. Realism, as characterized above, combines three theses: a distinctness thesis, a metaphysical thesis and an epistemological thesis. The metaphysical thesis is the claim that moral facts are objective. If moral facts are independent in some way, what exactly is it that they are independent of? To say 'independent of us' is little help at best, and straightforwardly wrong at worst. Moral facts concern agents and actions, which are human matters, and so they are not completely independent of us; if we were different, different actions would be wrong. The phrasing used above, 'independent of any beliefs and thoughts we might have about them', attempts to be more precise, at the cost of excluding too much. Is the limitation involved in 'about them' justified? If not, what other limitation would be better? This matter is very difficult to resolve (see Values; Objectivity).

Realism, on this showing, is a complex of claims. The distinctness claim carves out a distinctive subject matter for

ethics. The independence claim tells us something about the sort of fact that ethics is concerned with. The epistemological claim tells us that we have a less than perfectly secure grip on those facts. Moral realists have, however, generally been willing to say that we are capable of moral knowledge, even if we do not achieve it very often. They would all agree that are we capable of justified moral beliefs (if they are cognitivists).

Readers should be aware that the characterization of moral realism is a matter of hot debate. In particular, as well as the sort of account offered in the present entry, there is a form of realism that, taking its start from the claim that there are facts of the matter in ethics, and so moral truths, holds that there must therefore be ‘truth-makers’ - things that make the truths true. In the moral case, what makes moral truths true must be the possession of moral properties by suitable agents and actions. Realism, so understood, is the commitment to moral properties and relations as no less ‘real’ than other properties. It is possible, however, to combine these two strands of realist thinking without strain.

2 Arguments for realism, and an outline of its history

Corresponding to the three elements of realism, there are three things commonly urged by realists in favour of their position. In different ways, they all suggest that we should take seriously the way things initially appear to us. Realists try to hold that things are as they appear, despite noncognitivist arguments that they cannot possibly be. This is sometimes called ‘the appeal to phenomenology’.

First, realists claim that moral thought appears to have its own subject matter, distinct from science and all natural inquiry. Second, they argue that moral judgment appears to be an attempt to determine a matter of fact that is independent of any beliefs we might have about it; the fact is one thing, and what we think about it another. Third, realists hold that moral judgment presents itself to the judger as risky and fallible. When facing a difficult choice, especially, we have a sense of thin ice; we know that, with the best will in the world, the view we come to may be wrong. Only the second two of these three claims may properly be termed ‘phenomenological’.

Most moral theorists have been realists, from Plato and Aristotle through Price and Hutcheson to Sidgwick and Mill, and then, in the twentieth century, to Moore and the intuitionist tradition (see Intuitionism in ethics). It is the opposition to realism that needs to be documented. Here the patron saint of noncognitivism is Hume, whose work flowered, though altered in many respects, in the noncognitivist tradition of Stevenson, Ayer and Hare (see Emotivism). Leading contemporary noncognitivists are Blackburn and Gibbard.

Advances in the philosophy of language, philosophy of mind and philosophy of science in the 1960s and 1970s persuaded many that the noncognitivist arguments against realism, which had dominated the intellectual scene since 1930, were less powerful than they had appeared (see Analytic ethics §2). Two quite distinct forms of moral realism emerged, American and British. (This means only that most proponents of the first form are American, and most proponents of the second are British.) In the next two sections, the differences between these two positions will be charted in three areas: metaphysics, epistemology and theory of motivation.

3 American moral realism

Metaphysics. American moral realists are naturalists: they suppose that moral facts are either natural facts or configurations of natural facts (see Naturalism in ethics §1). As suggested in the initial summary, it is possible that the fact that if I do this I will have helped someone and harmed nobody else just is the fact that I ought to do this action. Perhaps, however, the moral fact is more complex than this natural fact (and than any other single natural fact), without this meaning that it is not some combination of natural facts. If so, those natural facts will have to be combined in the right sort of way - in the sort of way that they are here - if they are together to make the moral fact that I ought to do this action. Then the moral fact will be identical with this configuration of natural facts.

This form of naturalism in ethics is often, though not always, accompanied by some form of consequentialism. A certain sort of natural fact is a moral fact because there is a relation between that fact and certain consequences (see Consequentialism). Naturalism in ethics is now a live option again because of a growing sense that G.E. Moore’s ‘open question’ argument is flawed (see Moore, G.E. §1).

Epistemology. American moral realists, seeing moral facts as natural facts, suppose that they are knowable in whatever ways natural facts can be known, including science. To identify those facts as moral facts, we will need

to combine the best scientific theory with the best moral theory.

Theory of motivation. How is it possible to reconcile the claim that moral facts are natural facts with the widespread sense that moral facts have an intrinsic authority - that they make demands on us to which we should respond, whatever our personal choices and preferences? This ‘intrinsic authority’ is hard to understand in any detail. Perhaps the best attempt is Kant’s distinction between categorical and hypothetical imperatives (see Kant §1; Kantian ethics). Moral imperatives such as, ‘Help those less fortunate than you’, are categorical, in the sense that one cannot escape their relevance to oneself by saying, ‘I just don’t care very much about that sort of thing’. By contrast, a hypothetical imperative such as, ‘Use fresh eggs to make an omelette’, has no grip on those who just do not care about how their omelettes taste.

However we understand it, the idea of the authority of moral facts does not sit easily with ethical naturalism. Natural facts, however configured, do not seem able to have any such authority over the will (see Moral motivation §§1-2). American realists generally respond to this by doubting the claim that any fact could have that sort of authority. The world, whether in its moral or its more obviously natural clothes, is one whose grip on us depends on our bringing to it a sort of moral concern (which will have been the product either of evolution or of education). If we lack that concern, we will be unmotivated by moral distinctions, even though we will be still perfectly capable of discovering which actions are right and which are wrong. To know the right is one thing, and to bend one’s will to it another. Moral judgment is cognitive; it is the discovery of facts. But facts are motivationally inert; whether one is motivated by them depends not so much on them as on what one cares about. So moral imperatives are hypothetical, despite appearances.

In these thoughts, American realists adopt what is sometimes called a ‘Humean’ theory of motivation. Neither belief nor desire alone can lead to action; only a combination of the two can do that. (For an action we need two mental states, one with each direction of fit.) Moral facts are the objects of belief. No such belief can motivate alone; for there to be an action, the agent must have some desire or preference as well as the belief. Moral facts cannot motivate in their own right, therefore; their ability to make a difference to how we act depends upon the independent contribution of a desire. This being so, they cannot have such a thing as an intrinsic authority over us; for whether they can make a difference to how we act depends on something over which they have no control (see Practical reason and ethics§2).

4 British moral realism

The British variety of moral realism denies everything that the American variety claims.

Metaphysics. British moral realism is non-naturalist: moral facts are not natural facts, nor are they are natural configurations of natural facts. They may be non-natural configurations of natural facts, but that is another matter. Natural facts are relevant to moral ones, of course, since they are the reasons why actions are right or wrong. It follows from this that any two situations that are naturally indistinguishable must be morally indistinguishable (see Supervenience). But this sort of supervenience is a far cry from any identity between the natural and the moral.

In terms of the characterization of moral realism given earlier, the British thus attribute far greater distinctness to moral facts, considering them to be metaphysically distinct from natural ones. The Americans have a harder task in showing what is distinctive about the moral, though not an impossible one. They can say, for instance, that moral facts are distinguished by their subject matter, or by the sort of configuration of natural facts that they are.

Epistemology. If moral facts are not natural facts, the normal methods of finding out how things are will not suffice for the discovery of the moral. Admittedly, British realists have been prone to talk of seeing that an action is right. But it appears that they mean by this neither that rightness is visible, nor that there is a moral sense in addition to the normal five senses. Talk of seeing that the action is right is intended to echo Aristotle’s remark that right and wrong are not matters of rules so much as of the nature of the case before us, and that to discern what is right we have to concentrate on the details of the present situation (see Virtue ethics §6). They deny, therefore, that moral judgment is a matter of subsumption, of bringing the present case under some moral rule. Moral judgment is the application of concepts, but those concepts are not rules. Indeed, it is characteristic of British realists to be sceptical about the very possibility of moral rules or principles. For them, moral judgment is a matter of recognizing the reasons for action as they present themselves in the present case, and responding to them as such (see Universalism in ethics §3; Logic of ethical discourse). This sort of recognition is not perceptual, but it is

cognitive and practical at once - for what one is recognizing is a reason for action, that is, a normative state of the world.

Theory of motivation. American realists are ‘externalists’ about moral judgment. They hold that the ability of a moral judgment to motivate, that is, to make a difference to how one acts, is dependent on the presence of a quite different mental state, namely some sort of desire. British moral realists are generally ‘internalists’, holding that a moral judgment motivates in its own right, and does not get its ability to affect action from a desire that is present at the same time (see Moral motivation §1). To make this out, they have to reject the standard Humean picture of motivation, though they do not agree among themselves about quite how to do this. Wiggins (1987) and McDowell (1978) hold that action does require a combination of belief and desire, but that in the moral case it is belief that leads and desire that follows. Dancy (1993) suggests that it is belief alone that motivates, since mere recognition of relevant reasons should be enough for action; he sees desire as a state of being motivated (by the reasons), not as what motivates. Either way, the suggestion that there are genuinely normative and non-natural facts is combined with the claim that recognition of these facts motivates in its own right, in a way that is not dependent on the presence of an independent desire.

This raises problems. The first is that Humeanism is more or less received wisdom; so the British are fighting an uphill battle in rejecting it. The second is that it is very hard to make sense of the idea of a state of the world, or fact, that stands in some intrinsic relation to the will. It is normal to think that facts are motivationally inert; to recognize them is one thing, and to bend the will to them another. The British hold that the response to the fact, which we call recognition, is itself motivational. The preferred way of doing this is by appeal to the ‘dispositional conception of value’ (see McDowell 1985). This conception is inspired by a supposedly similar dispositional conception of colour: a red object is one that is disposed to cause in us a characteristic sort of experience (see Colour, theories of §2). Similarly, a valuable object is conceived as one which is disposed to elicit a certain response from us, an inclination of the will. As such, it is not totally independent of us, since it consists (at least partly) in a certain relation to us; and this means that it is not fully objective, if objectivity is to be understood in terms of independence from us. But it is still objective in a weak sense, since value can still be conceived as there for us to recognize and there whether we recognize it or not (see Value, ontological status of).

There are two difficulties with this appeal to dispositions. The first is that the analogy with colours is hotly disputed. The second is that, in order to keep values in the world, their objectivity has had to be diminished. For some, this weaker conception of objectivity is hardly to be distinguished from subjectivity.

5 Realism and minimalism

In this entry, realism has been seen as a combination of three distinct theses. But there is an alternative account of what realism is that sees it as nothing more than a claim about truth. The realist, on this account, holds that moral statements are capable of truth, and indeed that some are true. If we say this, we can still distinguish between realism and objectivism in ethics. Realism is the claim that moral judgments are sometimes true; objectivism is the claim that the sort of truth they have is objective truth.

We can distinguish two sorts of opposition to this form of realism. The first accepts that moral statements are capable of truth, but holds that all are false (see Moral scepticism §3); the second holds that truth is not the appropriate form of success for a moral judgment, and that we would do better to think of them as sincere or insincere, or as more or less well connected with other judgments, or as ones that we ourselves would agree with. One should not try to combine these two sorts of opposition.

Crispin Wright (1992) has suggested that, if this is what is at issue between realism and noncognitivism, the matter will be quickly resolved in favour of realism. In his view, the mere fact that moral discourse is assertive, and that moral utterances are governed by norms of warranted assertibility, is enough to establish that we make no mistake in calling some true and others false. The question should not be, then, whether moral judgments are capable of truth, since everyone really admits that they are. Instead, the debate about realism should focus on other questions. According to Wright, among these questions are:

(1) Is it a priori that differences of moral opinion can only be explained in terms of divergent input, unsuitable conditions or malfunction (such as prejudice and dogma)?
(2) Do the supposed moral facts serve to explain anything at all? Suppose that they explain (some of) our moral

beliefs. Do they explain anything else in a way that is not mediated by our beliefs? That is to say, do moral facts directly explain anything about how the world goes?

Wright suggests that we only get a ‘full-blooded’ moral realism if our answer to these questions is ‘yes’. There will, therefore, be degrees of realism, and in a way the question is not whether we should be realists, but what sort of realists we should be. How far should our realism go?

6 Arguments against realism

Since realism comes in different forms, arguments against it are more likely to attack some particular form than to attack realism as such. The main difficulties for the American and the British realist schools have already been mentioned. In both cases they were metaphysical. The Americans have difficulty in keeping moral facts both natural and moral. The British have difficulty in explaining how the world can be other than ‘motivationally inert’.

The two general challenges to realism that are most often mentioned are those made by John Mackie and Gilbert Harman. Harman (1977) asks what, if anything, is explained by moral facts that cannot be equally well explained by moral beliefs. If moral belief alone is enough for all such explanations, why suppose that the facts exist in addition to the beliefs? The facts appear to be explanatorily redundant. We might suggest that at least the facts explain the beliefs, but Harman replies that the beliefs can be equally well explained in other ways, for example, by appeal to upbringing and education. This leaves the facts explaining nothing; they are mere metaphysical danglers, hanging in the air and not related to anything else at all. We are better off without such things. (This is different from Wright’s view above, because Wright allows that moral facts could explain moral beliefs; he only asks whether they could explain anything else ‘directly’.)

Mackie (1977) suggests that values, if they existed, would be very peculiar things, unlike anything else in the universe; so queer are they that, if they existed, we would need a special faculty of moral perception or intuition to perceive them. Their queerness lies in the idea that an objective value would necessarily be pursued by anyone who recognized it, because such values have ‘to-be-pursuedness’ built into them. Even if such things are possible, which nobody influenced by Hume would allow, something of that sort is of a different order from anything else with which we are acquainted.

Mackie also asks about the supposed relation between moral facts and natural facts. We ordinarily say, for instance, that an action was wrong because it was cruel. But ‘just what *in the world* is signified by this "because"?’ (1977: 41; original emphasis). Not only is there the wrongness and the cruelty, but also a totally mysterious ‘consequential link’ between the two.

These arguments of Mackie’s are answered in different ways by the different varieties of realism. The Americans deny the possibility of ‘to-be-pursuedness’; the British admit it, but try to explain it by appeal to the dispositional theory of value. As for the mysterious ‘consequential link’, both sides would, in their different ways, try to say that the wrongness is somehow ‘constituted’ by the cruelty.

See also: Logic of ethical discourse; Moral justification; Moral knowledge; Moral relativism; Projectivism; Realism and antirealism

JONATHAN DANCY

References and further reading

Brink, D. (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.(The only book-length presentation of American realism.)

Dancy, J. (1993) *Moral Reasons*, Oxford: Blackwell.(A recent full-scale expression of British realism.)

Darwall, S., Gibbard, A. and Railton, P. (1992) ‘Towards *Fin de Siècle* Ethics’, *Philosophical Review* 101 (1): 115-89.(An advanced survey of late twentieth-century ethics.)

Harman, G. (1977) *The Nature of Morality*, New York: Oxford University Press, ch. 1.(An influential introductory text.)

Little, M. (1995) ‘Recent Work in Moral Realism’, *Philosophical Books* 35 (3): 145-53; 35 (4): 225-33.(A two-part comparison of British and American realism.)

Mackie, J. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin, ch.1.(A very influential introductory text, which starts by attempting to undermine realism.)

McDowell, J. (1978) ‘Are Moral Requirements Hypothetical Imperatives?’, *Proceedings of the Aristotelian Society* supplementary vol. 52: 13-29.(Gives reasons for rejecting a Humean conception of moral motivation.)

McDowell, J. (1979) ‘Virtue and Reason’, *The Monist* 62: 331-50.(Attacks ‘subsumptive’ accounts of moral rationality, in favour of ‘perceptual’ ones.)

McDowell, J. (1985) ‘Values and Secondary Qualities’, in T. Honderich (ed.) *Morality and Objectivity*, London: Routledge, 110-29.(Gives a dispositional account of normativity in the world, appealing to the analogy with secondary qualities.)

Pettit, P. (1992) ‘Realism’, in J. Dancy and E. Sosa (eds) *A Companion to Epistemology*, Oxford: Blackwell, 420-4.(An accessible account of what it is to be a realist in any area, which sees realism as a complex combination of claims; the present entry is much influenced by Pettit’s account.)

Wiggins, D. (1987) ‘Truth, Invention and the Meaning of Life’, in *Needs, Values, Truth: Essays in the Philosophy of Value*, Oxford: Blackwell.(This and the other papers in the work represent the ethical thought of another leading British realist.)

Wright, C. (1992) *Truth and Objectivity*, Cambridge, MA: Harvard University Press.(Sets out Wright’s sophisticated conception of how the debate between realists and their opponents should be conducted, in a way that focuses on, but is not at all restricted to, *moral* realism.)

Moral relativism

Often the subject of heated debate, moral relativism is a cluster of doctrines concerning diversity of moral judgment across time, societies and individuals. Descriptive relativism is the doctrine that extensive diversity exists and that it concerns values and principles central to moralities. Meta-ethical relativism is the doctrine that there is no single true or most justified morality. Normative relativism is the doctrine that it is morally wrong to pass judgment on or to interfere with the moral practices of others who have adopted moralities different from one's own. Much debate about relativism revolves around the questions of whether descriptive relativism accurately portrays moral diversity and whether actual diversity supports meta-ethical and normative relativism. Some critics also fear that relativism can slide into nihilism.

1 Descriptive relativism

From the beginnings of the Western tradition philosophers have debated the nature and implications of moral diversity. Differences in customs and values the Greeks encountered through trade, travel and war motivated the argument attributed to the sophist Protagoras in Plato's *Theaetetus*: that human custom determines what is fine and ugly, just and unjust (see Protagoras). Anthropologists in the twentieth century, such as Ruth Benedict (1934), have emphasized the fundamental differences between the moralities of small-scale traditional societies and the modern West. For example, many traditional societies are focused on community-centred values that require the promotion and sustenance of a common life of relationships, in contrast to both the deontological morality of individual rights and the morality of utilitarianism that are the most prominent within modern Western moral philosophy. Within this philosophy itself moral diversity is represented by the debates between utilitarians and deontologists, and more recently criticism of both camps by defenders of virtue theory and communitarianism (see Deontological ethics; Utilitarianism; Virtue ethics; Community and communitarianism). Such differences have motivated the doctrine of *descriptive relativism*: that there exists extensive diversity of moral judgment across time, societies and individuals, and that it concerns central moral values and principles.

Critics of descriptive relativism argue that it fails to account for important moral similarities across cultures such as prohibitions against killing innocents and provisions for educating and socializing the young. A relativist response given by Michael Walzer (1987) is to argue that shared norms must be described in an extremely general way and that once one examines the concrete forms they take in different societies, one sees significant variety, for example, in which persons count as 'innocent'. The descriptive relativist might go so far as to assert that no significant similarities exist, but an alternative position is that broad similarities exist that are compatible with significant differences among the moralities human beings have held.

Critics of descriptive relativism also argue that many moral beliefs presuppose religious and metaphysical beliefs, and that these beliefs, rather than any difference in fundamental values, give rise to much moral diversity (see Religion and morality §3). Also, differences in moral belief across different societies may not arise from differences in fundamental values but from the need to implement the same values in different ways given the varying conditions obtaining in these societies. One relativist reply is that while such explanations apply to some moral disagreements, they cannot apply to many others, such as disagreements over the rightness of eating animals or the moral status of the foetus or the rightness of sacrificing an innocent person for the sake of a hundred more.

2 Meta-ethical relativism

The most heated debate about relativism revolves around the question of whether descriptive relativism supports *meta-ethical relativism*: that there is no single true or most justified morality. There is no direct path from descriptive to meta-ethical relativism; the most plausible argument for meta-ethical relativism is that it is part of a larger theory of morality that best explains actual moral diversity.

Critics of meta-ethical relativism point out that moral disagreement is consistent with the possibility that some moral judgments are truer or more justified than others, just as disagreement among scientists does not imply that truth is relative in science. Some relativists are unimpressed by the analogy with science, holding that disagreements about the structure of the world can be sufficiently radical to undermine the assumption that there is an absolute truth to be found. This defence of meta-ethical relativism amounts to founding it upon a

comprehensive *epistemological relativism* that expresses scepticism about the meaningfulness of talking about truth defined independently of the theories and justificatory practices of particular communities of discourse (see Epistemic relativism).

An alternative relativist response is to take a nonrelativist stance towards science and to drive a wedge between scientific and moral discourse. Defenders of such a *morality-specific meta-ethical relativism* argue that scientific disagreements can be explained in ways that are consistent with there being a nonrelative truth about the structure of the physical world while moral disagreements cannot be treated analogously. For example, much scientific disagreement may be traced to insufficient or ambiguous evidence or distortions of judgment stemming from personal interests. Relativists have argued that such explanations will not work for moral disagreements such as the ones mentioned above concerning the eating of animals, abortion, and the sacrifice of an innocent to save more lives.

In offering alternative explanations of moral disagreement, morality-specific relativists tend to adopt a 'naturalistic' approach to morality in the sense that they privilege a scientific view of the world and fit their conceptions of morality and moral disagreement within that view. They deny that moral values and principles constitute an irreducible part of the fabric of the world and argue that morality is best explained on the theory that it arises at least in part from custom and convention. On Wong's view (1984), for example, a good part of morality arises out of the need to structure and regulate social cooperation and to resolve conflicts of interest. Meta-ethical relativism is true because there is no single valid way to structure social cooperation.

Morality-specific relativism divides into *cognitive* and *non-cognitive* versions (see Moral judgment §1). On C.L. Stevenson's emotivist view (1944), for example, moral discourse merely expresses emotion and influences the attitudes and conduct of others (see Emotivism). Cognitive relativists, such as Mackie, Harman, Foot and Wong, interpret moral judgments as expressing belief, on the grounds that moral judgments are often argued or judged true or false on the basis of reasons. Within cognitive relativism, there are those who believe that there is no single true morality because more than one morality is true, and those who believe that there is no single true morality because all are false. J.L. Mackie (1977) represents the latter camp, on the ground that while morality actually arises out of custom and convention, the meanings of moral terms presuppose a mistaken reference to *sui generis* properties that provide everyone with a reason for acting according to morality (see Value, ontological status of). Other cognitive relativists see no need to construe moral terms as containing a reference to nonexistent properties and instead tie their cognitive content to certain standards and rules.

According to such a standards relativism, moral language is used to judge and to prescribe in accordance with a set of standards and rules. Different sets of standards and rules get encoded into the meaning of ethical terms such as 'good', 'right' and 'ought' over time, and into individuals, groups, or societies in such a way that two apparently conflicting moral beliefs can both be true. Though under a relativist analysis the beliefs express no conflicting claims about what is true, they do conflict as prescriptions as to what is to be done or as to what kinds of things are to be pursued. The disagreement is purely pragmatic in nature, though parties to the disagreement may not be aware of this if they erroneously assume they share the relevant standards.

Another crucial question for the standards relativist concerns *whose* standards and rules apply when someone makes a moral judgment. Suppose that Jones makes a moral judgment about what Smith ought to do, but that the standards Jones applies to guide his own conduct are not the same as the standards Smith uses to guide hers. One possibility is that Jones uses Smith's standards to judge what she ought to do. Another possibility offered by Harman in some of his writing about relativism is that one must judge others by standards one shares with them. His theory is that morality consists of implicit agreements for the structuring of social cooperation. Moral judgments implying that the subjects have a reason to do what is prescribed make sense only as prescriptions based on what the speakers and subjects (and the intended audience of the judgments) have agreed to do. Other standards relativists observe that people use their own standards in judging the conduct of others, whether or not they believe these others to share their standards.

There are *radical* and *moderate* versions of meta-ethical relativism. Radical relativists hold that any morality is as true or as justified as any other. Moderate relativists, such as Foot (1978), Walzer and Wong (1984), deny that there is any single true morality but also hold that some moralities are truer or more justified than others. On Wong's view, for instance, certain determinate features of human nature and similarities in the circumstances and

requirements of social cooperation combine to produce universal constraints on what an adequate morality must be like. It may be argued, for example, that a common feature of adequate moralities is the specification of duties to care and educate the young, a necessity given the prolonged state of dependency of human offspring and the fact that they require a good deal of teaching to play their roles in social cooperation. It may also be a common feature of adequate moralities to require of the young reciprocal duties to honour and respect those who bring them up, and this may arise partly from role that such reciprocity plays in ensuring that those who are charged with caring for the young have sufficient motivation to do so. Such common features are compatible with the recognition that adequate moralities could vary significantly in their conceptions of what values that cooperation should most realize. Some moralities could place the most emphasis on community-centred values that require the promotion and sustenance of a common life of relationships, others could emphasize individual rights, and still others could emphasize the promotion of utility.

3 Normative relativism

Does meta-ethical relativism have substantive implications for action? *Normative relativism* - the doctrine that it is morally wrong to pass judgment on or to interfere with the moral practices of others who have adopted moralities different from one’s own - is often defended by anthropologists, perhaps in reaction to those Western conceptions of the inferiority of other cultures that played a role in colonialism. It also has application to disagreements within a society such as that concerning the morality of abortion, where the positions of the disputing parties seem ultimately to be based on fundamentally different conceptions of personhood.

As in the case of descriptive and meta-ethical relativism, however, there is no direct path from metaphysical to normative relativism. One could hold consistently that there is no single true morality while judging and interfering with others on the basis of one’s own morality. Wong has proposed a version of normative relativism consistent with the point that nothing normative follows straightforwardly from meta-ethical relativism. Meta-ethical relativism needs to be supplemented with a liberal contractualist ethic to imply an ethic of nonintervention. A liberal contractualist ethic requires that moral principles be justifiable to the individuals governed by these principles. If no single morality is most justified for everyone, *liberal normative relativism* may require one not to interfere with those who have a different morality, though the requirement of noninterference may not be absolute when it comes into conflict with other moral requirements such as prohibitions against torture or the killing of innocents (see Liberalism).

4 Relativism and moral confidence

A reason why relativism has been feared is the thought that it could easily slide into moral nihilism. Could one continue living according to one’s moral values, which sometimes require significant personal sacrifice, if one can no longer believe that they are truer or more justified than other values that require incompatible actions? One relativist response is that one may reasonably question the importance of certain features of one’s morality upon adopting a view of their conventional origin. Consider that duties to give aid to others are commonly regarded as less stringent than duties not to harm them. Gilbert Harman (1975) has proposed that this difference results from the superior bargaining position of those with greater material means in the implicit agreement giving rise to morality. Those with lesser material means may reasonably question this feature of morality, if they are persuaded of Harman’s explanation. Notice, however, that it is not merely the supposition that this feature arose from convention that may undermine one’s confidence in it. With regard to other features of one’s morality, one may adopt a relativist view of them and continue to prize them simply because they are as good as any other and because they help to constitute a way of life that is one’s own.

Admittedly, people who condemn torture and unremitting cruelty as an offence against the moral fabric of the world may possess a certitude not available to relativists and may find it easier to make the personal sacrifices morality requires. Moral certitude has its own liabilities, however, and has itself contributed to the unremitting cruelty that human beings have inflicted upon each other.

See also: Morality and ethics; Relativism; Social relativism

DAVID B. WONG

References and further reading

Benedict, R. (1934) *Patterns of Culture*, New York: Penguin.(Argues that different cultures are organized around different and incommensurable values.)

Foot, P. (1978) *Moral Relativism (The Lindley Lectures)*, Lawrence, KS: University of Kansas Press.(Defends a form of moderate relativism.)

Harman, G. (1975) 'Moral Relativism Defended', *Philosophical Review* 84: 3-22.(Argues that morality is founded on implicit agreement and that moral 'ought to do' judgments presuppose that speaker, subject and intended audience share the relevant moral standards.)

Harman, G. (1984) 'Is There a Single True Morality?', in D. Copp and D. Zimmerman (eds) *Morality, Reason and Truth: New Essays on the Foundations of Ethics*, Totowa, NJ: Rowman and Allanheld.(Discusses the relation between a naturalistic approach to morality and relativism.)

Harman, G. and Thomson, J. (1996) *Moral Relativism and Moral Objectivity*, Cambridge, MA: Blackwell.(Most comprehensive statement of Harman's relativism. Modifies some earlier positions taken.)

Herskovits, M. (1972) *Cultural Relativism: Perspectives in Cultural Pluralism*, New York: Vintage Books. (Anthropologist argues for meta-ethical and normative relativism.)

Krausz, M. (1989) *Relativism: Interpretation and Confrontation*, Notre Dame, IN: University of Notre Dame Press.(Besides the articles from this volume specifically identified here, this is a good survey of different perspectives on descriptive and meta-ethical relativism.)

Ladd, J. (1973) *Ethical Relativism*, Belmont, MA: Wadsworth.(A collection of philosophical and anthropological essays on descriptive and meta-ethical relativism.)

MacIntyre, A. (1988) *Whose Justice? Which Rationality?*, Notre Dame, IN: University of Notre Dame Press. (Accepts a strong version of descriptive relativism in which different moral traditions contain incommensurable values and standards of rational justification, but argues against meta-ethical relativism on the grounds that traditions may be compared with respect to their ability to resolve internal problems and to explain why other traditions have failed to solve their own problems.)

Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin.(Defends a sceptical form of relativism under which moral judgments lack the objectivity they purport to have. Hence no standard moral judgments are true.)

Nagel, T. (1986) *The View from Nowhere*, New York: Oxford University Press.(Criticism of arguments for meta-ethical relativism from moral diversity.)

Plato (*c*.380-367 BC) *Theaetetus*, in *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1961.(Statement of a conventionalist and relativist view of morality attributed to Protagoras.)

Scanlon, T.M. (1995) 'Fear of Relativism', in R. Hursthouse, G. Lawrence and W. Quinn (eds) *Virtue and Reasons: Philippa Foot and Moral Theory*, Oxford: Clarendon Press.(Discussion of why relativism appears to be a threat to the importance of morality.)

Stevenson, C.L. (1944) *Ethics and Language*, New Haven, CT: Yale University Press.(Defends a noncognitivist theory of moral judgment.)

Walzer, M. (1987) *Interpretation and Social Criticism*, Cambridge, MA: Harvard University Press.(Defence of moderate meta-ethical relativism based on the theory that the meaning of general values is given through specific practices.)

Williams, B. (1972) *Morality: An Introduction to Ethics*, New York: Harper & Row.(Criticism of some versions of meta-ethical and normative relativism.)

Wong, D. (1984) *Moral Relativity*, Berkeley, CA: University of California Press.(A defence of moderate relativism based on a naturalistic approach. Some chapters presuppose contemporary philosophy of language that some may regard as technical.)

Wong, D. (1991) 'Three Kinds of Incommensurability', in M. Krausz (ed.) *Relativism: Interpretation and Confrontation*, Notre Dame, IN: University of Notre Dame Press.(Discusses ways in which value differences between cultures may result in different criteria for the rationality of belief about the world.)

Wong, D. (1996) 'Pluralistic Relativism', *Midwest Studies in Philosophy* 20: 378-400.(More discussion about the constraints that all adequate moralities would have to meet.)

Moral scepticism

Scepticism in general is the view that we can have little or no knowledge; thus moral scepticism is the view that we can have little or no moral knowledge. Some moral sceptics argue that we cannot have moral knowledge because we cannot get the evidence necessary to justify any moral judgments. More radical moral sceptics argue that we cannot have moral knowledge because in morality there are no truths to be known. These radical sceptics argue either that moral judgments are all false because they erroneously presuppose the real existence of 'objective values', or that moral judgments aim to express feelings or influence behaviour instead of stating truths. Critics of moral scepticism, in turn, argue that in at least some cases moral judgments aim to state truths, some of these judgments are in fact true, and we have enough evidence to say that we know these moral truths.

1 Types of moral scepticism

Moral scepticism is the view that we can have little or no knowledge about ethics, about right and wrong, good and bad, virtue and vice (see Moral knowledge §3). It is one of a cluster of views that would challenge or limit traditional philosophical accounts of ethics. Moral relativism claims that moral truth is relative to cultural convention or personal commitment (see Moral relativism); moral subjectivism claims that moral truth is determined by the preferences, feelings, and other subjective states of persons; moral egoism is the claim that it is always obligatory to advance one's own interests (see Egoism and altruism). These positions are motivated by some of the same problems about morality that motivate moral scepticism, but they are not properly speaking sceptical positions, since they are all consistent with the claim that we have extensive knowledge of right and wrong, good and bad, virtue and vice.

Theoretically, at least, sceptics about morality can be divided into global sceptics and specifically moral sceptics. That is, some philosophers are sceptical about morality because they are sceptical about everything, claiming that no one has knowledge of morals because no one has knowledge about anything (see Scepticism). Other philosophers claim that we can and do have knowledge of many things, but not of morality in particular. Global scepticism is not of interest here, and in any event has been comparatively rare in the history of philosophy. Specifically moral scepticism has been more common, mainly because of the intractability of some moral disagreements, the difficulty in providing any plausible account of how we could be causally related to the objects of moral knowledge, and the perception that morality suffers badly when compared with natural science, widely accepted as the paradigm of knowledge. Moral sceptics can be divided into total moral sceptics, who claim that we can have no moral knowledge, and partial moral sceptics who claim that we can have knowledge of some aspects of ethics, but not of others (for example, of good and bad, but not of right and wrong).

Moral sceptics may differ in their accounts of why we fail to have moral knowledge, and these differences can be explained in terms of the traditional analysis of knowledge (see Knowledge, concept of). Traditionally, a judgment has been counted knowledge only when it is both true and justified. Thus, moral sceptics generally argue that in ethics (or some aspect of ethics) one or the other of these conditions is unsatisfiable. Some find fault with the justification of moral judgments; others find fault with their truth.

2 The problem of justification

The first sort of moral sceptic holds that we lack moral knowledge because we lack adequate justification for the moral judgments in question. This inadequacy may be characterized in terms of one of the following: (1) A slim evidential basis for moral judgments, comprising, for instance, only empirical and logico-mathematical truths. (2) Limited kinds of inference admitted in moral reasoning, such as deduction and induction. (3) A strong characterization of moral judgments such that if there were any true moral judgments, they would have to be universal, necessary, and action-guiding. (4) Doctrines such as the 'is/ought gap', according to which no good inferences are possible, from the slim evidential basis to strongly characterized moral conclusions, via the limited kinds of inference admitted (see Logic of ethical discourse).

In antiquity, doubts about the adequacy of justification led some to total moral scepticism. Pyrrhonists such as Sextus and Philo argued that one and the same thing appears good to one set of people and bad to another, due to differences in their customs, laws, mythical beliefs, and so on (see Philo of Larissa; Pyrrhonism; Sextus

Empiricus). These appearances are the only justification we have for the moral value of the things in question, but since a thing cannot be both good and bad simultaneously, and since we have no reason to trust one set of appearances over another, no moral judgment has more rational support than its contradictory, and the appropriate attitude toward all moral matters is suspension of belief.

In modern philosophy, doubts about justification have led some to partial moral scepticism. G.E. Moore, for example, held in *Principia Ethica* (1903) that we could have knowledge of the good, and that right actions were by definition those that produced the most good, but that we could not know which actions were right, because we could not know enough about their causal consequences in the distant future (see Moore, G.E. §1). Similarly, it has been argued by W.D. Ross (1930) that we could know the general principles of *prima facie* duty, but we could rarely, if ever, know in particular circumstances which action was our actual duty, because of the moral complexity of every particular set of circumstances (see Ross, W.D.). In both cases, these philosophers hold that we lack moral knowledge simply because we lack the factual or moral information needed to justify the moral judgments in question.

3 The problem of truth

The second, and more radical, sort of total moral sceptic holds that we lack moral knowledge because in matters of morality there are no truths to be known. Some, often called 'noncognitivists', argue that moral judgments have no truth-value, because they do not express cognitively meaningful propositions (see Analytic ethics; Moral judgment §1). Others, often called 'error theorists', hold that moral judgments have a truth-value, but are in fact all false. Noncognitivism and error theory are often classified as 'anti-realist' theories of ethics, since they reject the realist's claim that some moral judgments are literally true, independently of anyone's beliefs or attitudes about them. It should be noted, however, that they are not the only possible anti-realist positions in ethics, and that anti-realism in ethics need not entail scepticism in ethics.

Noncognitivism was perhaps first suggested by David Hume, who argued that morality was the object of feeling not of reason, and that 'when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it' (1739/40: 469) (see Hume, D. §4). Following Hume, early emotivists such as A.J. Ayer (1936) argued that moral judgments such as 'Stealing is wrong' cannot be literally true or false, since they express emotions, as in 'Stealing! Boo!', instead of cognitively meaningful propositions (see Ayer, A.J. §5; Emotivism). Later emotivists, such as C.L. Stevenson (1944), argued that moral judgments serve to express the speaker's attitudes and evoke similar attitudes in the hearer, and prescriptivists such as the early R.M. Hare (1952) argued that they are prescriptions, rather than statements of fact (see Stevenson, C.L.; Hare, R.M. §1; Prescriptivism).

Error theory is the claim that all moral judgments are false, because they aim to state facts about good and bad or right and wrong, but that these facts do not obtain because the objective value entities they presuppose do not exist. J.L. Mackie (1977) argues against objective moral values on the grounds that, if they were to exist, they would be utterly unlike any other sort of entity with which we are familiar. Similarly, Gilbert Harman (1977) argues against objective moral facts and properties on the grounds that they play no role in the best explanation of any observations (see Naturalism in ethics §3; Moral realism).

4 Responses to moral scepticism

Thinking moral scepticism an encouragement to immorality, and an outrageous denial of obvious human experience, numerous philosophers have defended moral knowledge from sceptical criticisms of both types. Defences of the justification of moral judgments can be classified as responses to the four factors outlined in §2.

(1) Some have argued for an expansion of the evidential basis so as to include, for instance, moral intuitions or theological information. (Proponents of the method of 'reflective equilibrium' in ethics claim that moral intuitions should be admitted as the data for moral theorizing just as perceptual judgments are admitted as the data for scientific theorizing - see Moral justification §2.) (2) Some have argued for a richer view of the kinds of inference admissible in moral reasoning. (Stephen Toulmin (1970) has argued that the standards of reasoning appropriate for ethics and other areas of everyday life are different from those of the natural sciences.) (3) Others have tried to lessen the distance between premises and conclusion by reducing the content of the moral judgments in need of

justification. (A utilitarian might try to meet Moore's scepticism about right action by redefining 'right' in terms of 'maximal expected utility' instead of 'maximal utility' (see Rationality, practical).) (4) Still others, including many in the Aristotelian tradition, have argued against doctrines such as the 'is/ought gap', by providing examples of valid 'is-to-ought' inferences, or denying the sharp distinction between empirical and moral judgments which such doctrines appear to presuppose (see Logic of ethical discourse §3).

Similarly, defences of the truth of moral judgments can be classified as responses either to noncognitivism or to error theory. Critics of noncognitivism emphasize the similarity between moral judgments and other types of fact-stating judgment, minimize the tension between the fact-stating and action-guiding functions of moral judgments, or highlight the inadequacies of particular emotivist accounts. (Moral philosophers such as Peter Geach (1972) have criticized noncognitivism as unable to account for the meaning of judgments in unasserted contexts such as the antecedents of material conditionals.) Certain philosophers who are widely thought of as noncognitivist, such as R.M. Hare, have worked to reconcile prescriptivism or emotivism with the possibility of true, hence knowable, moral judgments. In his essay 'Objective Prescriptions' (1993), Hare argues that any moral judgment that fully satisfies the requirements implicit in the logical meaning of moral words, and takes into account all the relevant empirical information, may be called 'true', even though prescriptions are not ordinarily thought of as 'stating facts'. Realist critics of error theories, such as David Brink, Geoffrey Sayre-McCord, and Nicholas Sturgeon, play down the differences between moral facts and properties on the one hand, and physical facts and properties on the other, arguing that the former have a legitimate role in the best explanation of moral and other phenomena. Other critics, such as Ronald Dworkin, Thomas Nagel and John McDowell, reject explanatory potency as an appropriate criterion for existence of values. Still others, such as Crispin Wright, argue for minimalist or coherentist accounts of truth in ethics, which allow moral judgments to be true (hence in principle knowable) despite their failure according to more robust correspondence theories of truth (see Moral realism §6).

See also: Moral expertise; Moral justification; Value, ontological status of

MARK T. NELSON

References and further reading

Annas, J. and Barnes, J. (1985) *The Modes of Scepticism*, Cambridge: Cambridge University Press, esp. 151-71. (Contains excerpts of the Pyrrhonist sceptical arguments and useful commentary.)

Ayer, A.J. (1936) 'Critique of Ethics and Theology', in *Language, Truth and Logic*, London: Gollancz; 2nd edn, 1946; 2nd edn repr. in G. Sayre-McCord (ed.) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press, 1988, ch. 1.(Probably the most accessible introduction to noncognitivism in ethics.)

Brink, D.O. (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press, esp. chaps 2, 5, 6.(A difficult but comprehensive treatment of moral realism, containing a detailed response to error theory.)

Butchvarov, P. (1989) *Skepticism in Ethics*, Bloomington, IN: Indiana University Press.(A systematic defence of partial moral scepticism.)

DePaul, M.R. (1993) *Balance and Refinement: Beyond Coherentism in Moral Inquiry*, London: Routledge.(A sophisticated discussion of the method of reflective equilibrium, and a defence of a nonsceptical, related method for moral inquiry.)

Dworkin, R. (1996) 'Objectivity and Truth: You'd Better Believe It', *Philosophy and Public Affairs* 25 (2): 87-139.(A sustained and moderately accessible attack on various versions of moral scepticism.)

Geach, P. (1972) 'Assertion', in *Logic Matters*, Oxford: Blackwell.(A difficult but influential criticism of noncognitivism.)

Hare, R.M. (1952) *The Language of Morals*, Oxford: Oxford University Press.(The original, and most influential, presentation of universal prescriptivism.)

Hare, R.M. (1993) 'Objective Prescriptions', in A.P. Griffiths (ed.) *Ethics*, Cambridge: Cambridge University Press, 1-17.(A moderately accessible attempt to reconcile the prescriptivity and objectivity of moral judgments.)

Harman, G. (1977) *The Nature of Morality*, New York: Oxford University Press.(An accessible introduction to ethics, containing an influential 'explanatory' criticism of moral knowledge.)

Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978, esp. 455-76.(A classic source of sceptical arguments, both in general and with regard to ethics.)

Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin, esp. 1-49.(The original, and most influential, source of error theory in ethics.)

McDowell, J. (1985) 'Values and Secondary Qualities', in T. Honderich (ed.) *Morality and Objectivity*, London: Routledge & Kegan Paul, 110-29; repr. in G. Sayre-McCord (ed.) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press, 1988, ch. 8.(A difficult but important criticism of Mackie's error theory.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press, esp. ch. 5.(An influential presentation of intuitionism about value, including partial moral scepticism about obligation.)

Nagel, T. (1980) 'The Limits of Objectivity', in S.M. McMurrin (ed.) *The Tanner Lectures on Human Values*, Salt Lake City, UT: University of Utah Press, 77-139.(A difficult but influential piece, critical of Mackie.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Oxford University Press, esp. ch. 2.(A classic exposition of pluralist intuitionism about obligation, combined with partial moral scepticism.)

Sayre-McCord, G. (1988) 'Moral Theory and Explanatory Impotence', in *Midwest Studies in Philosophy* 12: 433-58; repr. in G. Sayre-McCord (ed.) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press, 1988, ch. 11.(An accessible argument for the indispensability of evaluative facts.)

Sinnott-Armstrong, W. and Timmons, M. (eds) (1996) *Moral Knowledge? New Essays in Moral Epistemology*, Oxford: Oxford University Press.(A collection of moderately difficult but important essays on moral knowledge and moral scepticism, with a useful annotated bibliography.)

Stevenson, C.L. (1944) *Ethics and Language*, New Haven, CT: Yale University Press.(An early but sophisticated version of noncognitivism.)

Sturgeon, N. (1984) 'Moral Explanations', in D. Copp and D. Zimmerman (eds) *Morality, Reason and Truth: New Essays on the Foundations of Ethics*, Totowa, NJ: Rowman and Allanheld, 49-78; repr. in G. Sayre-McCord (ed.) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press, 1988, ch. 10.(A moderately accessible criticism of Harman's explanatory critique of moral realism.)

Timmons, M. (ed.) (1990) 'Spindel Conference 1990: Moral Epistemology', *Southern Journal of Philosophy*, suplementary vol. 29.(A wide-ranging collections of essays on moral knowledge and moral scepticism, with an extensive bibliography, post-1971.)

Toulmin, S.E. (1970) *Reason in Ethics*, Cambridge: Cambridge University Press, esp. part III.(The original, and most accessible, presentation of the 'good reasons' approach in ethics.)

Wright, C. (1996) 'Truth in Ethics', in B. Hooker (ed.) *Truth in Ethics*, Oxford: Blackwell, 1-18.(A moderately accessible introduction to a minimalist conception of truth in ethics.)

Moral sense theories

In Leviathan (1651), Thomas Hobbes argued that since good and evil are naturally relative to each individual's private appetites, and man's nature is predominantly selfish, then morality must be grounded in human conventions. His views provoked strong reactions among British moral philosophers in the seventeenth and eighteenth centuries. Moral sense theories comprise one set of responses. A moral sense theory gives a central role to the affections and sentiments in moral perception, in the appraisal of conduct and character, and in deliberation and motivation. Shaftesbury and Francis Hutcheson argued that we have a unique faculty of moral perception, the moral sense. David Hume and Adam Smith held that we cultivate a moral sensibility when we appropriately regulate our sympathy by an experience-informed reason and reflection.

1 The moral sense

Moral sense theory originated in the work of Shaftesbury (1711), as he drew on aspects of the seventeenth-century English rationalist tradition that dealt with the emphasis Hobbes had placed on self-interest (see Shaftesbury; Rationalism). Hobbes (1651) had argued that natural circumstances lead each man to favour his own interests at the expense of more socially-directed affections, and when men enter into the social compact and obey the dictates of morality they do so primarily from self-interested motives (see Contractarianism §2). Following the Cambridge Platonist Henry More, Shaftesbury conceived of the moral sense as a part of intellect. According to More (1667), our 'boniform faculty' distinguishes goodness, discovering in it a 'sweetness and flavour' that in turn induces us to act virtuously. In his *Inquiry Concerning Virtue and Merit* (1711), Shaftesbury combines the notion of an intellectually grounded taste with Locke's notion of reflection, and argues that our moral sense directs reflexive sentiments of praise or censure to the form of virtue or vice in our affections and resolutions, the psychological springs of conduct (see Locke, J. §3; Virtues and vices §2-5). Certain 'pre-conceptions' in the mind guarantee a correct correspondence between the pleasure of moral admiration and virtuous affections. The virtuous agent is one who achieves a balance of the self-interested and social affections. Shaftesbury's argument concerning how we achieve this balance links obligation with self-interest, and is intended to show both that Hobbes was wrong about the weakness of our other-regarding motives, and that self-love obliges us to virtue. The moral sentiments are disinterested evaluations of virtuous and vicious affections. Our moral admiration of our social affections nevertheless enriches the latter, rendering them 'master-pleasures' that govern all our other motives. Moreover, since such pleasures in turn yield the highest 'self-enjoyment', virtuous action undertaken from other-regarding motives provides agents with the chief means for promoting their own interest.

In his *Inquiry into the Original of Our Ideas of Beauty and Virtue* (1725), Francis Hutcheson defended Shaftesbury's views against the criticisms of Mandeville. Hutcheson followed his predecessor in stressing the instinctively reflexive character of our moral sentiments. In employing an empiricist framework and analogizing the moral sense to our physical senses, however, he severed the intimate connection forged by Shaftesbury between the moral sense and intellect. With Locke, he asserted that reason could not be the source of any new ideas, including our moral perceptions (see Locke, J. §2; Moral knowledge §3). Approbation and blame consist in perceptions of pleasure and pain, and these simple ideas are the immediate and necessary responses of our moral sense to an agent's benevolent or unkind affections. We rely on reason to draw inferences about an agent's character and to inform us of the tendencies of the actions and affections that we evaluate morally, but it cannot by itself determine our moral response.

To counter the claim advanced by rationalists such as Clarke and Wollaston that our rational apprehension of goodness obliges us to virtuous conduct, Hutcheson constructed new arguments concerning reason's limits in *Illustrations on the Moral Sense* (1728). He drew a distinction between reasons for approving or blaming and reasons for acting, while showing that both presuppose our affective nature. In his account of justifying reasons, he acknowledged that we need to justify our particular moral evaluations of actions and characters. Justifying reasons establish the grounds for our approval or disapproval by stating truths about the tendencies of particular actions and motives. Since our approval of benevolence (which Hutcheson regarded as the sole virtue) is uniquely determined by the moral sense, not by the truth of a justifying reason, the justificatory function of such reasons presupposes the moral sense. Exciting reasons are truths concerning the motives of agents, and therefore presuppose affections. Since we may present a justifying reason for approving some action in the distant past that

cannot possibly motivate us, or have exciting reasons for acting in ways that we neither approve nor blame, our reasons for evaluating and our reasons for acting are distinct from one another.

Hutcheson rejected the view of Hobbes and some of the natural lawyers that moral obligation needed the backing of sanctions (see Natural law). He also minimized Shaftesbury's emphasis on the connection between obligation and interest, since it made self-love, not itself a virtuous motive, the sole obligation to virtue. He conceded that there is an interested sense of obligation that concerns the happiness of the agent and presupposes self-love. But he redefined moral obligation to mean the determination of the agent disinterestedly to approve and perform virtuous actions. The unease agents feel if they omit a required benevolent action makes them conscious of their failure with respect to their obligation, without invoking either sanctions or advantage to themselves.

2 Regulated sympathy

In *A Treatise of Human Nature* (1739/40), Hume borrowed Shaftesbury's and Hutcheson's arguments concerning our natural sociality, added to Hutcheson's arguments about the limits of reason, and revised his doctrine of moral obligation (see Hume, D. §§3-4). Both Hume and Adam Smith agreed that our moral sentiments distinguish between virtue and vice, and take as their objects the characters of agents. Aware of rationalist criticisms of the notion of an original moral sense, they appealed instead to the principle of sympathy, regarding it as the foundation for a moral sensibility that is essentially social in character, and requires cultivation through education and experience.

In an important concession to Hobbes, Hume argued that prior to establishing conventions of justice, value responses are relative to each individual in part because of the natural limitations of our social affections and sympathy. In order to resolve the conflicts stemming from each individual's partiality, a more reflective form of interest naturally obliges us to establish rules of justice. Our extended sympathy with all who participate in the convention makes us feel uneasy when others disapprove of our breaking the rules, and this gives rise to our sense of moral obligation. As in Hutcheson's account, moral obligation is connected with our moral sentiments, and does not depend on sanctions. The extensive sympathy underpinning our moral sensibility enables us to adopt general perspectives from which we correct our individual biases and form shared sentiments of praise and blame. Although this more cultivated moral sensibility depends on human conventions, Hume argued that it results from the natural and necessary 'progress of the sentiments'.

Hume distinguished between artificial virtues whose motives arise only when certain conventions are in place, and natural virtues that presuppose natural affections. He also emphasized the importance to us of self-regarding virtues, and argued that we cannot draw a fine line between our approval of virtue and our approval of agents' talents or natural beauty. He agreed with Shaftesbury and Hutcheson that our interested contemplation of the pleasures yielded by moral approval gives us a reason to endorse having a virtuous character. But he also urged that his version of our moral sensibility gains an advantage for virtue because it accommodates our disinterested approval of the experience-informed reason and sympathy that lead us to correct and improve our moral sentiments.

In *The Theory of Moral Sentiments* (1759), Smith refined Hume's account of sympathy, distinguishing between judgments of propriety and judgments of merit. The former arise when impartial spectators sympathize with the agent's motive and approve of its propriety, the latter when they sympathize with both the agent's proper motive and the gratitude elicited by the recipient of the virtuous action. Smith used the distinction to argue for a greater variety of moral sentiments. Hume and Smith argued that our shared moral sensibility enables us to see our conduct and character mirrored in one another's sentiments. But Smith was less sceptical than Hume about the influence of moral judgment on deliberation. When we become the impartial spectator of our own character, we develop a conscience. We can rely on the internal promptings of conscience to guide our own conduct (see Conscience).

Towards the end of the eighteenth century, Reid's connection between moral choice and free will, and the emphasis on the consequences of action for social reform in the utilitarianism of James Mill and Bentham, displaced the moral sense theorists' concentration on our internal sentiments (see Reid, T.; Utilitarianism).

See also: Cambridge Platonism §4; Common-sense ethics; Moral motivation

JACQUELINE TAYLOR

References and further reading

Cragg, G.R. (ed.) (1968) *The Cambridge Platonists*, Oxford: Oxford University Press.(Contains More's *Enchiridion ethicum*, and other rationalist responses to Hobbes.)

Darwall, S. (1995) *The British Moralists and the Internal 'Ought': 1640-1740*, Cambridge: Cambridge University Press.(Discusses developments of the notion of obligation, and relates the British moralist tradition to contemporary ethics.)

Hobbes, T. (1651) *Leviathan*, in W. Molesworth (ed.) *The Collected Works of Thomas Hobbes*, vol. 3, London: Routledge/Thoemmes Press, 1994.(Ground-breaking work sparking debate about the reality of morality and the source of obligation.)

Hope, V.M. (1989) *Virtue by Consensus*, Oxford: Clarendon Press.(Discusses the moral sense and sympathy in Hutcheson, Hume and Smith.)

Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978, bk III.(Presents Hume's account of our sense of morality.)

(1748, 1751) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 3rd edn, 1975.(A more concise and philosophically sophisticated account of morality than that given in the *Treatise*.)

Hutcheson, F. (1725) *An Inquiry Into the Original of Our Ideas of Beauty and Virtue*, in *Collected Works of Francis Hutcheson*, vol. 7, Hildesheim: Olms, 1969. (An account of various internal senses, including the aesthetic and the moral sense.)

Hutcheson, F. (1728) *Illustrations Upon the Moral Sense*, ed. B. Peach, Cambridge, MA: Harvard University Press, 1971.(Especially good for Hutcheson's criticism of the rationalist tradition.)

More, H. (1667) *Enchiridion ethicum (An Account of Virtue)*, trans. as *An account of virtue: or, Dr. Henry More's abridgement of morals, put into English*, London, 1690; repr. in G.R. Cragg (ed.) *The Cambridge Platonists*, Oxford: Oxford University Press, 1968.

Raphael, D.D. (1947) *The Moral Sense*, Oxford: Oxford University Press.(Good introductory work. Contains chapters on Hutcheson, Hume and Reid.)

Raphael, D.D. (ed.) (1969) *British Moralists: 1650-1800*, Oxford: Clarendon Press, 2 vols, 1969; repr., Indianapolis, IN: Hackett Publishing Company, 2 vols, 1991.(Excellent selections from the major works.)

Schneewind, J.B. (ed.) (1990) *Moral Philosophy from Montaigne to Kant*, Cambridge: Cambridge University Press, 2 vols.(Useful compilation that places the work of British moralists in a broader European context. Excellent bibliography.)

Shaftesbury, A.A.C. (1711) *Inquiry Concerning Virtue and Merit*, in *Characteristics of Men, Manners, Opinions, and Times*, ed. D. Walford, Manchester: Manchester University Press, 1977.(Contains Shaftesbury's major ethical writings.)

Smith, A. (1759) *The Theory of Moral Sentiments*, ed. D.D. Raphael and A.L. Mackie, Oxford: Clarendon Press, 1976.(An edited reprint of the 6th edn of the work, published in 1790, this is Smith's major contribution to ethics. In part VII, Smith contrasts contemporary views of virtue with those of the ancients, and criticizes the rationalist, selfish and moral sense schools of moral thought.)

Moral sentiments

Moral sentiments are those feelings or emotions central to moral agency. Aristotle treated sentiments as nonrational conditions, capable of being moulded into virtues through habituation. The moral sense theorists of the Enlightenment took sentiments to provide the psychological basis for our common moral life. Kantian approaches deny the primacy of sentiments in moral personality, and treat moral sentiments as conditioned by our rational grasp of moral principles.

A central issue is whether moral sentiments incorporate moral beliefs. Accounts which affirm a connection with moral beliefs point to the complex intentionality (object-directedness) of such states as resentment or indignation. Against this, some observe that moral emotions may be felt inappropriately.

Of special interest are the sentiments of guilt and shame. These seem to reflect different orientations towards moral norms, and questions arise about the degree to which these different orientations are culturally local, and whether either orientation is superior to the other.

1 Historical perspectives

Sentiments have played various roles in ethical theory. On Aristotle's view, set out in *Nicomachean Ethics*, moral virtue is essentially a condition of the sentiments. It is a state of character, involving a disposition to feel the emotions connected with pleasure and pain in ways appropriate to one's circumstances. Moral virtue is distinguished from other states of character by the fact that the virtuous person's sentiments are responsive to practical reason, which determines what it is right to do or to feel in any given situation. One acquires this state of character through habituation, which refines one's susceptibility to pleasure through repeated performance of virtuous actions, and indirectly affects one's ability to deliberate effectively about the true goods of human life (see Aristotle §23). This complex view of moral sentiments treats them as parts of the nonrational soul, which nevertheless both influence and are influenced by one's capacities for practical reflection; it was opposed in the ancient period by the Stoics, who interpreted sentiments essentially as rational judgments (see Stoicism §19).

A different approach was taken by the sentimentalist theorists of the eighteenth century. Setting themselves against the apparent scepticism and egoism of Thomas Hobbes, these philosophers grounded our autonomous moral faculties not in reason but in the moral sense. The moral sense was interpreted as an innate susceptibility to feel sentiments of approval when confronted with morally virtuous character traits, under conditions of suitably impartial reflection. Virtuous traits were understood primarily in terms of benevolence (or other traits, such as justice, which are useful on the whole); and later proponents of the approach took the moral sense that approves of those traits to rest on a mechanism of sympathetic identification with others (see Moral sense theories).

This conception is like Aristotle's in treating moral sentiments as nonrational phenomena - moral approval cannot incorporate moral judgments, because such judgments would presuppose a rational grasp of moral distinctions of the sort the sentimentalists denied. In this respect, the sentimentalists were very much the precursors of modern noncognitivists, who hold, for instance, that moral judgments serve to express emotions of approval or disapproval (see Emotivism; Analytic ethics §1). In contrast with Aristotle, however, the sentimentalists treated the susceptibility to moral sentiments as an innate part of our common human nature, the emergence of which does not require elaborate habituation; this befits the role of the emotions in moral sense theory, as providing a common psychological basis for moral life (see Morality and emotions §1).

Kant and his followers ground our common moral capacities in reason rather than the emotions. This approach is popularly thought to be hostile to the idea that sentiments contribute positively to the moral life - Kant himself, for instance, notoriously claimed that actions motivated by sentiment have no moral worth (see Kantian ethics). But the denial that moral sentiments play a foundational role in moral thought and motivation does not mean that they have no contribution to make.

Outlined in his *Critik der practischen Vernunft (Critique of Practical Reason)* (1788), Kant's account of the sentiment of respect for the moral law suggests that distinctively moral emotions are structured by our independent rational grasp of moral principles: this approach can be extended to such emotions as indignation, resentment and guilt, which are usually responses to the violation of moral principles. The susceptibility to such moral sentiments

shows that one has internalized moral principles, and it is an important task of moral theory to consider whether the principles it proposes can be so internalized that a society organized around them will be a stable and self-sustaining one.

2 Sentiment and moral belief

A leading issue is whether the moral sentiments incorporate moral beliefs. Philosophers who deny this (such as the sentimentalists, and modern noncognitivists) tend to emphasize the passivity of the moral sentiments, and interpret questions of emotional justification primarily in strategic terms ('Is it good for the agent to be subject to sentiments of this type?'). Kantians and others who take moral sentiments to rest on moral beliefs see greater scope for active control of the moral sentiments, through exercise of our rational capacities for moral judgment; and they treat moral sentiments as open to nonstrategic forms of assessment ('Is a particular moral sentiment really appropriate to its object?').

This difference in approach is difficult to resolve in the abstract: moral sentiments are of many different kinds, ranging from pleasure or approval to such highly structured emotions as pride or regret, and there is no reason to expect any single account to apply to all of them. It is more fruitful to explore the difference with reference to a restricted class of moral sentiments.

Resentment, for instance, is characteristically a reaction to a moral offence, and this aspect of the emotion is naturally accounted for by supposing the emotion to rest on the moral belief that one has been treated wrongly. This moral belief gives the emotion a propositional object, and opens it to assessment in nonstrategic terms ('Was the action I resent really wrong?'). Those who favour this approach can readily explain the difference between resentment and closely related sentiments, such as indignation or disdain, in terms of the moral beliefs that the different emotions incorporate. Opponents point out that resentment is often felt irrationally, on occasions when one does not really take oneself to have been morally wronged. This phenomenon seems hard to reconcile with the claim that episodes of resentment necessarily incorporate a moral belief.

An alternative approach treats resentment and other moral sentiments as refinements of underlying biological syndromes, to be identified not by the moral beliefs they incorporate but by their characteristic causes and behavioural effects. By depriving moral sentiments of their propositional objects, however, this approach seems to deny something essential to states of resentment, namely that they involve a distinctive way of thinking of the person one resents, different from the way one conceives the targets of indignation or anger (which may have similar causes and behavioural effects).

3 Guilt and shame

A different set of questions concerns the degree to which moral sentiments are culturally conditioned (see Moral relativism). Many theories hold that a susceptibility to distinctively moral sentiments such as sympathy or even resentment is built into human nature, providing a common basis for moral discourse and response. But other sentiments have seemed more parochial. Guilt, for instance, is sometimes conceived of as a sentiment available only within cultures whose ethical systems are organized around the notions of law, obligation and individual freedom. In other cultures, where ethical precepts are taken to describe a public ideal of honourable behaviour, shame is the characteristic emotional sanction.

Early statements of the contrast between guilt and shame cultures took the latter to be the more primitive, lacking the resources for fully internalizing moral demands. This 'progressivism' has been reversed in more recent treatments of the distinction, which have seen in shame a moral sentiment that lacks the objectionable aspects of guilt - the metaphysical assumption of absolute freedom before the law, and the pointless cruelty that seems to characterize the bad conscience.

This contrast may be sustained even by those who remain agnostic on the question of whether there are any pure shame cultures. Shame and guilt seem to reflect two different orientations towards moral norms: the personality in which guilt predominates sees them as harsh and punitive requirements, while those in whom shame is the dominant moral sentiment view moral norms as articulations of an ideal way of life, a set of excellences which they aspire to exemplify. (This opposition is sometimes expressed in psychoanalytic literature in terms of a contrast between the super-ego and the ego-ideal; see Freud, S. §§8-9.) Important questions here are whether we

can really do without the guilt-orientation to moral norms, and whether it is as metaphysically and morally suspect as its critics have alleged.

See also: Conscience; Moral judgment; Moral knowledge; Moral motivation; Rectification and remainders

R. JAY WALLACE

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. W.D. Ross, revised by J.L. Ackrill and J.O. Urmson, Oxford: Oxford University Press, 1980, books II-V. (Discusses the role of sentiments in moral virtue, or excellence of character.)

Baier, A.C. (1993) 'Moralism and Cruelty: Reflections on Hume and Kant', *Ethics* 103: 436-57; repr. in *Moral Prejudices: Essays on Ethics*, Cambridge, MA: Harvard University Press, 1994, 268-93.(Attacks guilt-based moralities as cruel, by comparison with moralities oriented around shame, and associates the two with the theories of Kant and Hume respectively.)

Gibbard, A. (1990) *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Cambridge, MA: Harvard University Press.(Argues that moral sentiments do not rest on moral judgments, and that moral norms are norms for the appropriateness of moral sentiments.)

Hume, D. (1751) *An Enquiry Concerning the Principles of Morals*, ed. J.B. Schneewind, Indianapolis, IN: Hackett Publishing Company, 1983.(Characteristic and accessible statement of the moral sense theory.)

Kant, I. (1788) *Critik der practischen Vernunft*, ed. and trans. L.W. Beck, *Critique of Practical Reason*, New York: Macmillan, 3rd edn, 1993, part 1, book I, ch. 3.(Discusses respect for the law as a sentiment conditioned distinctively by the rational grasp of moral principles.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, part III.(Contains an important discussion of the nature and genesis of moral sentiments, and considers the prospects for internalizing competing principles of justice in the sentiments of individual agents.)

Strawson, P.F. (1962) 'Freedom and Resentment', *Proceedings of the British Academy* 48: 1-25; repr. in G. Watson (ed.) *Free Will*, Oxford: Oxford University Press, 1982.(Influential defence of the view that our moral relations are thoroughly structured by our susceptibility to moral sentiments, drawing implications for the issue of freedom of will.)

Taylor, G. (1985) *Pride, Shame, and Guilt: Emotions of Self-Assessment*, Oxford: Clarendon Press.(Subtle discussion of the sentiments mentioned in the title, emphasizing their cognitive dimensions.)

Wallace, R.J. (1994) *Responsibility and the Moral Sentiments*, Cambridge, MA: Harvard University Press.(Offers an interpretation of guilt, resentment and indignation, and traces the role of these moral sentiments in our practice of holding people morally responsible.)

Williams, B. (1993) *Shame and Necessity*, Berkeley, CA: University of California Press.(Argues that the shame morality depicted in ancient Greek literature is far more sophisticated than its critics have grasped.)

Moral standing

Towards whom is it appropriate to direct fundamental moral consideration? This is the question of moral standing. Many different answers have been offered: all and only those creatures that are themselves capable of extending moral concern and consideration; all humans, whether capable of functioning as moral agents or not; humans plus certain other 'higher' animals (such as gorillas, chimpanzees and porpoises) that can think, reason and be self-aware; creatures capable of feeling sensations such as pain, no matter how otherwise rudimentary their psychological existence; living beings, whether sentient or not; 'holistic' entities such as political states, cultural traditions, biological species, natural ecosystems.

The moral standing issue has great significance both practically and theoretically. Earnestly reconsidering who, or what, counts morally could change what we eat, how we clothe ourselves, the extent to which we spread out over the land. Even those who believe that only humans count morally must still address the theoretically challenging question of what it is about humans that warrants such exclusive concern.

To try to resolve this difficult issue, philosophers have pursued several different strategies. Some work from plausible convictions about particular cases (such as '"Normal" healthy adult humans have moral standing, if anyone does') and then, subject to the demand for principled consistency, try to extrapolate to a more general account of what confers moral standing. Others try to articulate a broad view of the general nature of moral consciousness (as involving, for example, the disposition to empathize with others or promote their good) and work back from that to an account of the most basic conditions of the possibility of being conscientiously considered (for instance, having feelings with which others can empathize or a good which they can take into account).

1 The nature of moral standing

Philosophers often grapple with two major issues. Which considerations would guide and shape a morally decent life, and of whom is morally decent living required? But there is a third very large philosophical concern about the nature of morality: towards whom is it appropriate to direct fundamental moral consideration? This is the issue of basic moral standing: which entities matter morally, not merely for the sake of others, but in their own right?

Essentially the same issue has been formulated in a variety of different vocabularies. What does it take to be a 'moral patient' (as opposed to a 'moral agent')? Who or what is 'morally considerable'? Who can be regarded properly as an 'object' of moral concern (as distinct from the active 'subject' who entertains the concern)? But there are also less appropriate ways to pose the question. Thus it is sometimes rendered as 'Who or what deserves moral consideration?', or 'Who is entitled to such consideration?'. These formulations unduly narrow the scope of the question: being deserving and having a right are only two of the grounds on which beings who matter morally can be taken into consideration for their own sake (see Desert and merit; Rights).

The use of the expression 'moral standing' is modelled on, but in some measure also diverges from, the notion of standing as employed in the context of the law. To have legal standing is to be legally authorized to raise, or have raised on one's behalf, a complaint in court. Thus it is possible to have standing (a) with respect to some legal matters, but not others; and (b) within certain judicial venues only. The notion of moral standing parallels such possibilities: (a) a creature capable of suffering but incapable of thinking and communicating might have a morally significant claim on others not to be treated cruelly but no meaningful claim to 'freedom of thought and speech'; (b) the prospect of being pressed into service as a beast of burden could give an animal moral standing in the 'court' of humanity's conscience; attacks by nonhuman predators in the wild might not. Despite these parallels, moral and legal standing are not equivalent. There can be moral standing without the corresponding legal standing (consider, for example, the emotional support that close friends can rightly expect but not legally compel); and legal standing without fundamental moral standing (consider the fiction of business corporations as legal 'persons').

2 Positions on moral standing

Who, then, has basic moral standing? Many different answers have been offered. At one end of the spectrum is 'all and only those creatures that are themselves capable of extending moral concern and consideration'. But

common sense rejects this answer as too restrictive in scope. Young children are thought to fall within the domain of direct moral concern even before they are able to be morally considerate themselves; perhaps their standing derives from the fact that they are in the process of developing the ability to take others into moral consideration. But the same answer will not do with respect to humans who are profoundly, and permanently, retarded or demented; nor for animals who have desires and feelings, but are not capable of moral understanding and motivation.

At the other extreme, the answer ‘everything has moral standing’ raises at least two difficulties. (1) The greater the number and variety of entities that are to be regarded as having standing, the greater is the potential for serious conflicts of putatively legitimate need and interest. The task of resolving such conflicts justly becomes immensely, perhaps hopelessly, complicated. (2) If even insensate and inanimate entities are to be counted as falling within the domain of basic moral consideration, it remains to be explained what the nature and basis of that consideration could be. Just what ought to be done for the sake of a rock or a mud-hole?

Between these extremes several other answers have been proposed: for example, any creature capable of experiencing sensations such as pain; beings who, though not rational, are more than merely sentient, that is, who have enough psychological structure and complexity to count as ongoing ‘subjects of a life’ or as havers (however nonreflectively) of reasons; all living beings, whether sentient or not; biological species and/or natural ecosystems; societies, political states, cultures and traditions.

Thus, in trying to clarify who or what has moral standing, we will be prompted to investigate other difficult issues: What is it to be a moral agent - is it to be the conscious and self-conscious subject of a life, or to have reasons? What is involved in being sentient and how is sentience different from mere reactivity? What is it to be alive? How is biological life relevantly different from homeostatic mechanism? What is a species, an ecosystem? What differentiates one species from another, one ecosystem from another?

For each putatively relevant criterion (such as rationality, self-awareness, conscience, sentience and life), there are difficult questions as to which beings actually satisfy it: Who or what has the capacity for self-awareness and rational reflection - do porpoises and chimpanzees? Who or what is sentient - are clams and oysters? Could plants, slow-responding though they may be, be said to experience anything? Who or what is alive? Do viruses count? Are ecosystems living (super-) organisms too? Shall we say that the earth, or at least the outer crust and atmosphere known as the biosphere, is a living being (‘Gaia’) in its own right?

Perhaps there is some wisdom in more than one such view. As there are several different dimensions of moral concern, so it might be the case that different kinds of entities warrant different kinds of concern. This possibility is illustrated by Hume’s claim (1751) that while humans must treat animals kindly, they are not required to treat animals in accordance with rules of justice (which for Hume have to do primarily with property and contract) (see Hume, D. §5; Justice). Whatever we may think of the specifics of Hume’s position, the general insight behind it seems plausible. How a morally virtuous human agent ought to treat beings of another kind depends on the nature and capacities of the being in question: towards a sentient creature lacking the capacity to form expectations and harbour hopes, there is a duty to refrain from cruel treatment but no duty to keep ‘promises’; towards a living being incapable of experiencing pain and suffering, there is no duty to refrain from cruel and inhumane treatment, but there may yet be a duty to preserve its life, or the possibility of reproducing its kind.

Thus moral standing need not be all or nothing. Unfortunately, deciding on which entities are to be included within the domain of basic moral consideration does not by itself decide the question of how to resolve conflicts of legitimate need and interest. To help meet this problem, some philosophers argue that in the event of conflict moral priority ought to be given to beings with self-consciousness or with the capacity to reason or to creatures of the same species (or culture or community) as oneself. Other thinkers urge a more egalitarian approach.

3 Theoretical and practical import

The question of moral standing should have profound theoretical significance. Even ethicists who prefer to limit their focus to the domain of humankind must come to terms with the question: what is it about fellow humans that warrants such (exclusive) concern or consideration? Is it because they are capable of having experiences, feelings, hopes, desires? Then it would seem to follow that at least some nonhuman beings ought to be taken into consideration as well. Or is it a matter of ‘species-solidarity’ - the mere fact that we humans are all members of the

same biological species? From this it follows that intelligent beings visiting from another planet would be entitled to conquer and enslave us, or at least to be completely indifferent to how their actions affected us, because we and they are not all members of the same biological species. It seems more plausible to suggest that as thinking, feeling beings, humans do matter morally and hence that it would be unreasonable of other-worldly but morally responsible agents, even those of a different biological species, not to extend to us at least a measure of respect and consideration.

At the same time, the moral standing issue also holds more than purely theoretical interest. Agriculture, urban development, biomedical research, energy technology - these are some of the respects in which human practice could be altered profoundly by earnest reconsideration of who or what has basic moral standing (see Agricultural ethics; Bioethics; Technology and ethics). Even within the human sphere, there are difficult cases to address: future generations (see Future generations, obligations to); the deceased; people suffering from earthquake, famine or persecution, who live in distant places (see Help and beneficence); pre-natal human organisms (see Reproduction and ethics); human beings who are profoundly retarded or demented; and human organisms in a persistent vegetative state (see Life and death §3).

Of course, even if plants, animals, endangered species and ecosystems turned out not to have basic moral standing, many of the same restrictions on human conduct could conceivably be generated by purely ‘anthropocentric’ considerations, such as the concern not to undermine environmental conditions crucial to human survival; the desirability of keeping open the range and variety of humanly valued (emotional, aesthetic and even scientific) experiences; and the importance of reinforcing inhibitions and compunctions (against such factors as cruelty and wastefulness) that humans sorely need to cultivate in their dealings with one another.

Arguments such as these make the moral status of nonhuman entities contingent upon putative facts (psychological and ecological) of human life. Animal liberationists, reverence-for-life theorists, and ‘deep ecology’ advocates typically seek to ground the moral significance of nonhuman entities in properties wholly intrinsic to the entities themselves (see Animals and ethics; Environmental ethics). They dismiss human-regarding defences of respect for nonhuman nature as too limited, uncertain and shallow. Despite the large theoretical gulf that separates the two opposing schools of thought, the extent to which they might nevertheless converge in their specific policy prescriptions remains an open, and largely empirical, question.

4 How to decide moral standing: possibilities and pitfalls

How, then, can we make progress towards resolving this profoundly difficult issue? Philosophers often start with intuitively plausible convictions about particular cases (such as ‘"Normal" healthy adult humans have moral standing, if anyone does’) and then try to extrapolate, subject to the demand for principled consistency, to a more general account of what confers moral standing. Working from the other direction, philosophers sometimes try to trace out the implications of a well-established normative theory. Taking a utilitarian perspective, for example, one might reason that what confers moral standing is the capacity to experience pleasure and pain, or to form preferences whose satisfaction or frustration could be said to matter to oneself (see Utilitarianism). From a Kantian perspective, it might be argued that the only beings who ought to be taken into respectful consideration for their own sake are those who are (actually or potentially) capable of reasoning about what to do and of regulating their conduct accordingly (see Kantian ethics). Somewhere in between, there is the possibility of working back from such ‘mid-level’ common-sense considerations as kindness, fidelity and fairness, to the conditions of the possibility of being thus considered. For example, if there is a duty not to be cruel, then it applies to all and only those creatures who can suffer; if there is a duty to be honest, then creatures who have the capacity to form beliefs and expectations ought not to be deceived and defrauded. A fourth possibility is to articulate a broader, more procedural account of what is involved in being morally thoughtful. For example, moral consciousness seems to involve a willingness (1) to empathize with others, to try to imagine what it is like to be in their position; (2) to take their good or wellbeing into account; and (3) to deal with them in ways that can be justified to them. From these observations it seems to follow that nothing can be treated with morally thoughtful consideration unless there is something it feels like to be that being, a perspective of its own with which to empathize; or unless it has a good or wellbeing that can be taken into benevolent consideration; or unless it has a mind or will that can be respectfully consulted and addressed.

Such approaches are not without their pitfalls. Relying on (1) the paradigm of humanness, we might well be

accused of being anthropocentric, extending moral concern only to those beings that resemble ourselves to a degree sufficient to inspire our parochial interest and concern. Relying on (2) a familiar theoretical approach or on (3) everyday mid-level moral considerations, we run the risk of working from a perspective that is already tailored to an overly narrow conception of the domain of proper moral consideration. But charges of self-bias, anthropocentrism, speciesism and the like cannot be pressed indefinitely without losing some of their force. For suppose it were suggested that a grain of sand has moral standing and that to deny this is only to betray an unduly biased perspective. To respond to this challenge one could reasonably raise any or all of the following questions. What is the perspective of the grain of sand and what would it be like to empathize with it? In what does its good or wellbeing consist? What could count as respectfully consulting or addressing its will, or as justifying oneself to it?

See also: Respect for persons

ARTHUR KUFLIK

References and further reading

Aquinas, T. (*c*.1259-65) *Summa contra gentiles (Synopsis [of Christian Doctrine] Directed Against Unbelievers)*, trans. V.J. Bourke, Notre Dame, IN: University of Notre Dame Press, 1975, III.2, c.12.(Argues that we must refrain from cruelty to nonrational animals lest we become cruel to one another.)

Aristotle (*c*. mid 4th century BC) *Politics*, trans. T.J. Saunders, Oxford: Clarendon Press, 1995, I 5, 8.(Contends that plants 'exist for the sake of' animals and 'animals for the sake of man'.)

Baier, A.C. (1985) 'Knowing our Place in the Animal World', in *Postures of the Mind*, Minneapolis, MN: University of Minnesota Press.(Shows why, on Humean grounds, the virtues of honesty and loyalty as well as of kindness ought to shape our conduct towards some nonhuman animals.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, revised F. Rosen, Oxford: Clarendon Press, 1996, ch. 17, section 1.(Classic utilitarian account which insists that moral concern not be limited to rational beings but extended to any creature capable of suffering.)

Callicot, J.B. (1989) *In Defense of the Land Ethic: Essays in Environmental Philosophy*, Albany, NY: State University of New York Press.(Appeals to human sentiments (but not to human self-interest) to explain why species and ecosystems should matter to us in their own right.)

Coady, C.A.J. (1992) 'Defending Human Chauvinism', in C. Mills (ed.) *Values in Public Policy*, Fort Worth, TX: Harcourt Brace.(An entertainingly pointed reply to those who deny that human beings should have moral priority.)

Feinberg, J. (1974) 'The Rights of Animals and Unborn Generations', in W.T. Blackstone (ed.) *Philosophy and Environmental Crisis*, Athens, GA: University of Georgia Press, 43-68; repr. in *Rights, Justice, and the Bounds of Liberty*, Princeton, NJ: Princeton University Press, 1980.(Modern classic arguing that to have moral rights a being must have, or be capable of having, interests and that to have interests it must be capable of having conscious experiences, desires and beliefs.)

Feinberg, J. (1978) 'Human Duties and Animal Rights', in R.K. Morris and M.W. Fox (eds) *The Fifth Day: Animal Rights and Human Ethics*, Washington, DC: Acropolis Books, 45-69; repr. in *Rights, Justice, and the Bounds of Liberty*, Princeton, NJ: Princeton University Press, 1980.(Like the earlier work, this piece focuses on what qualifies a being as having moral rights.)

Goodpaster, K. (1978) 'On Being Morally Considerable', *Journal of Philosophy* 75 (6): 308-25.(Careful defence of the claim that living beings, whether sentient or not, have genuine interests that ought to be taken into basic moral consideration.)

Guyer, P. (1993) 'Duties Regarding Nature', in *Kant and the Experience of Freedom: Essays on Aesthetics and Morality*, Cambridge: Cambridge University Press, ch. 9.(Fine scholarship tracing the development of Kant's views on the moral status of animals and inanimate nature.)

Hume, D. (1751) *Enquiry Concerning the Principles of Morals*, ed. L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press, 1978, section 3, part I.(Maintains that while humans must be kind to animals, animals are not within the protective scope of justice.)

Johnson, E. (1983) 'Life, Death and Animals', in H.B. Miller and W.H. Williams (eds) *Ethics and Animals*, Clifton, NJ: Humana Press, 23-33.(Critique of key arguments for attributing moral priority to humans or any other beings capable of forming second-order preferences.)

Johnson, L.E. (1992) *A Morally Deep World: An Essay in Moral Considerabiity and Environmental Ethics*, New York: Cambridge University Press.(Argues that species and ecosystems are living beings in their own right and hence are to be respected in much the same way as the individual organisms that comprise them.

Kant, I. (1797) *Metaphysische Anfangsgründe der Tugendlehre*, trans. M.J. Gregor, *The Doctrine of Virtue*, New York: Harper & Row, 1964.(Argues that duties regarding nonhuman animals and inanimate nature are grounded in a duty to oneself, namely, to cultivate and improve one's own moral character.)

Korsgaard, C.M. (1996) *Sources of Normativity*, Cambridge: Cambridge University Press, lecture IV.(Makes a Neo-Kantian case for our having duties to, not merely regarding, nonhuman animals.)

Leopold, A. (1949) 'The Land Ethic' and 'On a Monument to the Pigeon', in *A Sand County Almanac with Other Essays on Conservation*, Oxford: Oxford University Press; repr. New York: Sierra Club/Ballantine, 1970. (Suggests that we must respect the way in which various forms of life are interconnected ecologically, and argues that if humans have a claim to superiority it is because of their seemingly unique capacity to care about the extinction of other species.)

Lovelock, J.E. (1979) *Gaia*, New York: Oxford University Press.(Makes a case for regarding the biosphere itself as a living organism.)

Naess, A. (1974) *Økologi, samfunn og livsstil. Utkast til en økosofi*, Oslo: Universitetsforlaget; trans. D. Rothenberg, *Ecology, Community and Lifestyle: Outline of an Ecosophy*, Cambridge: Cambridge University Press, 1989.(The classic statement of 'deep ecology' - a nonanthropocentrically based respect for ecosystems.)

O'Neill, O. (1996) *Towards Justice and Virtue: A Constructive Account of Practical Reasoning*, Cambridge: Cambridge University Press.(Suggests how, instead of trying to discover comprehensive and definitive criteria, we might 'construct' a solution to the problem of moral standing that is adequate for practical purposes.)

Partridge, E. (ed.) (1981) *Responsibilities to Future Generations*, Buffalo, NY: Preometheus Books.(A fine anthology of essays concerning the moral standing of future generations of humans.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, esp. 512.(A contractarian account of political justice; suggests that to be owed justice one must have the capacity for a sense of justice, but insists that another account of 'right conduct in regard to animals and the rest of nature' must still be provided.)

Regan, T. (1983) *The Case for Animal Rights*, Berkeley, CA: University of California Press.(A defence of why we must treat adult mammals as ends-in-themselves.)

Rolston, H. (1988) *Environmental Ethics: Duties to and Values in the Natural World*, Philadelphia, PA: Temple University Press.(Argues on nonanthropocenctric grounds for the significance of whole species and ecosystems.)

Routley, R. and Routley, V. (1980) 'Human Chauvinism and Environmental Ethics', in D. Mannison, M.A. McRobbie and R. Routley (eds) *Environmental Philosophy*, Canberra: Australian National University.(Offers a 'last person on earth' thought-experiment in defence of the claim that nonhuman (and nonsentient) forms of life matter in their own right.)

Scanlon, T.M. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds) *Beyond Utilitarianism*, Cambridge: Cambridge University Press.(Includes an explanation of how a contractarian approach can still account for the moral standing of nonhuman animals and other beings who would not be able to make or keep agreements.)

Singer, P. (1993) *Practical Ethics*, Cambridge: Cambridge University Press, 2nd edn, chaps 3, 5, 6, 10.(Clear discussions of the pros and cons of attributing moral standing to animals, plants, species, ecoystems and, within the human sphere, foetuses, future generations and impoverished persons in other societies.)

Stone, C. (1972) *Should Trees Have Standing?*, Los Altos, CA: William Kaufmann.(A defence of the claim that living beings have a good of their own that we can discern and ought to promote.)

Taylor, P.W. (1986) *Respect for Nature: A Theory of Environmental Ethics*, Princeton, NJ: Princeton University Press.(Holds that each living being ought to be taken into respectful consideration; offers some guidelines for resolving conflicts of need and interest that are likely to arise.)

VanDeVeer, D. (1979) 'Interspecific Justice', *Inquiry* 22: 55-79.(Proposes that when otherwise comparable interests conflict, 'the interests of beings with more complex psychological capacities' should be given greater weight.)

Moralistes

The moralistes constitute a tradition of secular French writing about human nature and political and social behaviour principally in the context of the court and the salon. Their non-systematic observations about mankind are couched in literary forms, such as the maxim and the pen-portrait, appropriate to the social context from which they emerged. The four principal moralistes of the ancien régime *were La Rochefoucauld, La Bruyère, Vauvenargues and Chamfort. La Rochefoucauld's Maximes (1665) constitute a sharp attack on the neo-Stoic moral optimism of the first half of the seventeenth century, and determine self-love to be the mainspring of all human behaviour. La Bruyère's Caractères (1688) is a more diverse work in both form and content: it contains a satire of the follies and vices of his age, as well as vivid pen-portraits. There are implicit contradictions in the moral norms governing this often indignant denunciation of men and society. Vauvenargues, writing some fifty years later, expresses more confidence in human nature, rehabilitating the passions and arguing for the moral value of self-love of a certain kind. This optimism is not shared by Chamfort, whose Maximes et pensées (1795) reverts to the cynical tone of his seventeenth-century predecessors in the genre. These writers do not attempt to systematize their thoughts, and they choose to express themselves in urbane and witty ways rather than in sober prose, but they carry out the Cartesian programme of employing 'common sense' and native intellectual powers to the end of uncovering aspects of human nature and behaviour accessible to observant people free from moral or religious preconceptions.*

1 'Moralistes'

The term 'moraliste', which first became current in its modern sense in the mid-eighteenth century, designates thinkers (usually not professional philosophers) who study humanity in a non-systematic, secular way, examining personal, social and political conduct with clear-sightedness and objectivity. In some accounts, this predominantly French tradition is said to stretch from Montaigne in the late sixteenth century to Paul Valéry in the twentieth; but it is more usual to see the heyday of the moralistes as the *ancien régime*, and especially the latter half of the seventeenth century. The practice of setting down dispassionate observations about human motives and behaviour in non-specialized terminology has very diverse roots, among which are: the educational exercise of keeping commonplace-books, in which students recorded pithy sentences from their reading; the practice of dialectical reasoning, which included the collection of points or *loci* to support a given conclusion; the existence of popular collections of proverbs, adages, and apophthegms; the habit of extracting realistic political doctrine in the form of precepts from historical writing and debunking the claims of Stoic virtue, whose most notorious practitioner was Machiavelli (see Stoicism); the development of manuals of civility and books of advice to courtiers based not on the morality of behaviour but its practical benefit to the individual and to society as a whole; the cult of witty conversation at court and in the salons; and finally, the invitation of Descartes to all people to examine their own thoughts with the free use of their innate powers of reasoning.

The mode of writing most associated with the moralistes is the maxim, although they also express themselves in short moral reflections and essays. The maxim can be distinguished from the proverb and apophthegm by its greater abstraction and elegance of form although, like the proverb, its author is not identified and it is given no specific context in time or place. It very often embodies a paradox, either of the kind which refutes a commonly received opinion about human nature and institutions (women are said to be better than men, for example, and tyranny than monarchy), or it is a paradox of a linguistic or terminological kind which breaches the rule of non-contradiction (such as the claim that it is foolish to be wise). In either form it is designed to surprise readers, by revealing to them some hitherto unnoted aspect of human society or by unmasking one of the ways in which human beings are dupes of their own nature. The more discursive literary forms adopted by moralistes do not necessarily have all these features, but there is the same combination of neutral, dispassionate, anonymous, general presentation with an implied social context of witty, secular, non-pedantic, urbane conversation. The four principal moralistes are usually said to be La Rochefoucauld, La Bruyère, Vauvenargues and Chamfort: all were either members of the highest level of French salon and court society, or its clients.

2 La Rochefoucauld

François, duc de La Rochefoucauld (1613-80) was born into the highest aristocracy. After a career in the army,

followed by hectic involvement in domestic political intriguing, he suffered a short period of banishment from Paris and the court before returning to both in the 1650s. His *Réflexions ou sentences et maximes morales (Reflections or moral aphorisms and maxims)* had its first authorized edition in 1665, having been produced in consultation, if not collaboration, with Jacques Esprit and Mme de Sablé: its fifth, much emended, edition appeared in 1678. The preface to the second edition of the work makes explicit its secular nature, but the Augustinian (or Jansenist) tone and vocabulary of much of the work are manifest. The preface also invites the reader to engage in a search for coherence which the form of the maxim frustrates, through its use of qualifying adverbs (often, usually, rarely, sometimes, mostly), its ambiguous syntax and inconsistent or vague use of terms, and its changes in tone, which varies from the cynical to the urbane; but a certain number of conclusions can be drawn from this deliberately disordered collection of discrete aphorisms. It is uniformly anti-Stoic: the heroic ethic which had dominated French literary and moral writing in the first half of the seventeenth century is systematically debunked. Human claims to know themselves through the exercise of reason, human contempt for death and for fortune in all its guises, human mastery over the will and passions, human love of virtue for its own sake, are all shown to be illusory or false; the subjection to their *amour-propre*, to their physiological nature and to the demands made upon them by the social environment in which they live are amply demonstrated. Much of this material is specific to the cultural context from which it emanates, being the product of a clear-sighted member of a disillusioned generation of aristocratic adventurers living in a society which placed great value on wit and urbanity. Its presentation in the form of aphorisms lends it however that air of universality which is traditionally associated with the classical age in France.

3 La Bruyère

Jean de la Bruyère (1645-96) was a member of the bourgeoisie and a client of the high nobility and of the theologician Jacques-Bénigne Bossuet, whose religious and political views he shared. He published his *Caractères ou les moeurs de ce siècle (The characters or mores of our age)* in 1688, on the coat-tails of a translation of *Characters* by Theophrastus; a ninth, much augmented edition appeared in 1696, by which time an unofficial key to the pseudonymous society portraits it contained had been published. The fifteen, rather disparate sections of the work consist of a mixture of short moral essays, maxims and pen-portraits; they were said by La Bruyère to culminate in the last section which purports to be a refutation of free-thinking, although it is difficult to substantiate, from internal evidence, the claim that such a refutation is the main point of the work.

La Bruyère’s work differs from La Rochefoucauld’s in its broader consideration of society and its preoccupation with writers and literary criticism. There are many apparent inconsistencies in the work: it claims not only to inculcate moral lessons, but also characterizes men as incorrigible; it favours the recognition of personal merit and social mobility in one section, only to denounce violently the social ascension of bourgeois financiers and the abandonment of the old feudal order in another; it strongly defends the use of the classics as aesthetic yardsticks at one point, only to engage in ‘modern’ techniques of literary composition at another. The work may best be viewed as a satire, in which La Bruyère deploys a wide range of rhetorical voices and devices, sometimes presenting himself as a detached, impartial observer, sometimes as a professional writer, sometimes as a representative of the bourgeoisie, sometimes as La Bruyère the oppressed client of the aristocracy; in this way he sets out to elicit a wide range of responses in his readers, ranging from indignation at the manifest injustices of society to detached amusement at its less harmful follies. Except in the rather unsatisfactory last section in which he sets out to prove the existence of God, he chooses to depict vice and folly rather than to enunciate the (often contradictory) moral norms which their denunciation implies. In his pen-portraits, which show him to be a very talented observer and painter of his age, he succeeds in using concrete details to flesh out the sometimes schematic moral character being described. He was one of the great stylists of his day, and a writer committed to promoting the cause of art and artists, but never at the expense of his satirical purposes.

4 Vauvenargues

Luc de Clapiers, marquis de Vauvenargues (1715-47) was, like La Rochefoucauld, a member of the aristocracy who did not receive a formal education but, unlike La Rochefoucauld, he played no part in the court or the salons of Paris; he served for a time as an army officer, before premature retirement and death from ill-health. His *Introduction à la connaissance de l’esprit humain, suivie de réflexions et maximes (Introduction to the knowledge of the human mind, followed by reflections and maxims)* (1746) did not receive the same attention as the works of

La Rochefoucauld and La Bruyère when they were published. In three books he deals with the mind, the passions, and the vices and virtues; his maxims were written very much with La Rochefoucauld and Pascal in mind, but where his predecessors had stressed human's enslavement to passions and the impotence of reason, Vauvenargues claims for humankind a degree of moral autonomy and a capacity for virtue. Passions are not to be seen in exclusively negative terms, but rather as the mainspring of moral thought and action; they are the very sources of intellectual activity, not forces which pervert it. Before Rousseau, he made the distinction between *amour-propre* (self-love) and *amour de nous-mêmes* (love of self), seeing the latter sort of love as compatible with virtue. This anti-Jansenist stance, which Vauvenargues shares with Voltaire and which he derives to some degree from Spinoza, is accompanied by a tolerant Deism, making Vauvenargues representative of his age, and suggesting parallels with the aphoristic productions of Voltaire himself (in his *Dictionnaire philosophique (Philosophical dictionary)* of 1764) and of Diderot (in the more contentious *Pensées philosophiques (Philosophical thoughts)* of 1746).

5 Chamfort

Nicolas-Sébastien Roch (1741-94), alias de Chamfort, has more in common with La Bruyère, being like him a client of aristocratic society and a reputed wit. A fervent supporter of the French Revolution, he none the less became suspect under the Terror, attempted to take his own life, and died shortly thereafter. His *Maximes et pensées, caractères et anecdotes (Maxims and thoughts, characters and anecdotes)* - also published with the ironic title *Produits de la civilisation perfectionnée (The products of perfected civilisation)* - first appeared posthumously in 1795. They claim to steer a middle course between the pessimistic depiction of human nature found in such thinkers as La Rochefoucauld and the optimism of idealist philosophers such as Shaftesbury, but the often cynical or pungent tone suggests more affinity with the former. Chamfort expresses great admiration for English philosophy, and especially Bacon; he recommends a radical inductive approach to social and political philosophy as well as to science, but most of his observations are directed at the corrupt aristocratic society of the last years of the *ancien régime*, and suggest an attitude of resignation rather than revolution, taking particular delight in representing social and political life through the metaphor of the theatre.

6 Conclusion

These moralistes embody a philosophical approach in so far as they set out to liberate themselves from preconceptions and observe humankind in its social and political context impartially; but what they have to say and the forms through which they express themselves are closely linked to their envisaged readership. Even the maxim form, which appears to state abstract, general truths outside any social context, is impregnated with the intention to shock, unsettle and surprise. It is therefore reasonable to claim that moralists reflect the thinking of their day in moral matters, but much less plausible to characterize them as independent thinkers making significant contributions to the history of moral thought.

See also: Cynics

IAN MacLEAN

References and further reading

Arnaud, C. (1988) *Chamfort: une biographie*, Paris: Laffont; trans. D. Dusinberre as *Chamfort: a biography*, Chicago, IL and London: Chicago University Press, 1992.(A recent biobibliography.)

La Bruyère, J. de (1688) *Caractères ou les moeurs de ce siècle (The characters or mores of our age)*, ed. R. Garapon as *Caractères*, Paris: Garnier, 1962; trans. J. Stewart as *Characters*, Harmondsworth: Penguin, 1970. (A good scholarly edition.)

Chamfort, de (Nicolas-Sébastien Roch) (1795) *Maximes et pensées, caractères et anecdotes (Maxims and thoughts, characters and anecdotes)*, Paris: Gallimard, 1982; trans. E.P. Mathers as *Maxims and Considerations*, Waltham St Lawrence: Golden Cockerel Press, 1926, 2 vols.(Plain-text edition.)

Clapiers, L. de, marquis de Vauvenargues (1746) *Introduction à la connaissance de l'esprit humain, suivie de réflexions et maximes (Introduction to the knowledge of the human mind, followed by reflections and maxims)*; repr. as *Réflexions et maximes*, Paris: Gallimard, 1971; trans. F.G. Stevens as *The reflections and maxims*, London: Humphrey Milford, 1940.(Plain-text edition.)

Delft, L. van (1971) *La Bruyère moraliste: quatre études sur les Caractères (La Bruyère as moralist: four studies on the Characters)*, Geneva: Droz.(A mainly literary analysis of the Caractères.)

Fine, P.M. (1974) *Vauvenargues and La Rochefoucauld*, Manchester: Manchester University Press.(A study in influence.)
Mourgues, O. de (1984) *Two French moralists*, Cambridge: Cambridge University Press.(The best available introduction to La Rochefoucauld and La Bruyère, aimed at an undergraduate readership.)
La Rochefoucauld, F., duc de (1665) *Réflexions ou sentences et maximes morales (Reflections or moral aphorisms and maxims)*, ed. J. Truchet as *Maximes,* Paris: Garnier, 1967; trans. L.W. Tancock as *Maxims*, Harmondsworth: Penguin, 1959.(The best available scholarly edition.)

Morality and emotions

Emotions such as anger, fear, grief, envy, compassion, love and jealousy have a close connection to morality. Philosophers have generally agreed that they can pose problems for morality in a variety of ways: by impeding judgment, by making attention uneven and partial, by making the person unstable and excessively needy, by suggesting immoral projects and goals.

The place of emotions in moral theories depends on whether they are conceived of merely as impulses without thought or intentional content, or as having some sort of cognitive content. Plato argued that emotions form a part of the soul separate from thought and evaluation, and moved, in the course of his writings, from a sceptical view of their contribution to morality to a more positive appraisal. Aristotle connected emotions closely with judgment and belief, and held that they can be cultivated through moral education to be important components of a virtuous character. The Stoics identified emotions with judgments ascribing a very high value to uncontrolled external things and persons, arguing that all such judgments are false and should be removed. Their cognitive analysis of emotion stands independent of this radical normative thesis, and has been adopted by many philosophers who do not accept it.

Modern theories of emotion can be seen as a series of responses and counter-responses to the Stoic challenge. Descartes, Spinoza, Kant and Nietzsche all accepted many of the Stoics' normative arguments in favour of diminishing the role played by emotions in morality; they differed, however, in the accounts of emotion they proposed. Focusing on compassion or sympathy, Hutcheson, Hume, Rousseau, Adam Smith and Schopenhauer all defended the role of some emotions in morality, returning to a normative position closer to Aristotle's (though not always with a similarly cognitive analysis).

Contemporary views of emotion have been preoccupied with the criticism of reductive accounts that derive from behaviourist psychology. By now, it is once again generally acknowledged that emotions are intelligent parts of the personality that can inform and illuminate as well as motivate. Philosophers' views have been enriched by advances in cognitive psychology, psychoanalysis and anthropology. Feminist accounts of emotion differ sharply, some insisting that we should validate emotions as important parts of moral character, others that emotions shaped by unjust conditions are unreliable guides.

1 Morality and conceptions of emotion

Emotions are usually considered to be distinct both from bodily appetites, such as hunger and thirst, and from objectless moods, such as irritation and endogenous depression (see Emotions, nature of). Major members of the class have traditionally been love, anger, grief, fear, envy, jealousy, guilt and pity (compassion) (see Love; Moral sentiments; Rectification and remainders). Some philosophers have adopted an account of these experiences that makes them rather close to appetites after all: emotions are surges of affect or energy in the personality, unreasoning movements that push people into acting without being very much connected to their thoughts about the world. This view is frequently connected to the idea that emotions derive from an animal part of our nature, often by thinkers who do not have a high regard for animal intelligence.

Seen this way, emotions figure in morality only as forces that either advance or impede the purposes of moral judgment or will. They can be trained and to some extent conditioned, but in a relatively mechanical manner. Thus they will never be what makes a virtuous character virtuous, though a virtuous person can take pride in having exercised an appropriate degree of discipline over them (see Self-control; Virtues and vices §3).

Other philosophical theories about emotions hold that they involve interpretation and belief. An emotion such as grief is not simply a mindless surge of painful affect: it involves a way of seeing an object, an appraisal of that object as important, and the belief that the object is lost. Fear involves the belief that bad events are impending, and that one is not fully in control of warding them off. Such theories hold that changes in the relevant beliefs entail changes in emotion: one who learns that danger is not really at hand will cease to fear. When we scrutinize the beliefs involved in the emotions, it emerges that they are of two sorts: beliefs about what is happening in the world (Is the person really dead? Is the enemy really at hand?) and evaluative beliefs (Is the person really worth getting upset about? Is the danger serious?). Many of the evaluative beliefs derive from social teaching, some of it

moral in nature. We learn what insults are worth getting upset about, what losses are serious, what damages are to be avoided.

Seen this way, emotions become intelligent parts of the moral personality, which can be cultivated through a process of moral education (see Moral education §3). Such a process will aim at producing adults who not only control their anger and fear, but experience anger and fear appropriately, towards the appropriate objects at the appropriate time in the appropriate degree. Merely self-controlled persons look to these theorists like those whose moral development is incomplete or imperfect. If we find them hating foreigners, but controlling their behaviour towards them, we will judge that there is some further moral work they should be doing before they can claim to be fully virtuous.

2 Emotions and character-based theories

All moral theories have some occasion to talk about emotions, but some focus on their cultivation more than others. Theories that focus on the calculation of utility will see emotions as forces that either promote or eclipse such calculations and the appropriate acts (see Utilitarianism). Theories that focus on duty can have a substantial place for an account of emotion, since emotions will be seen as motives that either promote or impede action in accordance with duty (see Deontological ethics; Duty). But if the account of duty is not supplemented with an account of virtue, it will not be clear how these motives should be integrated into the personality. Kant's theory of virtue (1797) shows us that a duty-based view need not neglect emotions as elements in virtue (see Kant, I. §10; Emotions, philosophy of §3). Since this theory has a noncognitive conception of emotion, and also a rather negative view of the contributions made by emotion, it gives emotions a rather restricted role; but there is nothing to prevent a virtue-theory such as Kant's from giving more positive attention to the cultivation of emotion.

Theories of character based on Aristotle and other Greek thinkers, however, have been Western philosophy's most substantial sources for thought about the proper cultivation of emotion within morality (see Emotions, philosophy of §2). Aristotle's norm of a reasonable person is one whose character is infused completely by the correct reasons for action, which have shaped all their motives and attitudes. Because he aims to describe the cultivation of a whole person and way of life, rather than simply to prescribe a list of duties, he has ample scope for discussing emotional self-shaping.

3 Ancient Graeco-Roman theories of emotion: Plato, Aristotle, Stoics

Plato argues in the *Republic* that the soul contains three distinct parts: the calculative part, the appetitive part, and the 'spirited' part (see Plato §14). This last part seems to be where Plato locates emotions such as anger, grief and fear. Although his account is cryptic, he suggests that the emotions differ from appetites in their responsiveness to changes in belief, and also in the fact that they are less brutish and more discriminating. But he refuses to construe them as belonging to the 'part' that performs evaluative reasoning. In normative terms, the *Republic* argues that all pity, fear and grief should be prevented from developing by a process of moral education that teaches young people that there is nothing for a good person to fear or grieve over; anger is apparently allowed to remain, but is channelled for military purposes. In the *Phaedrus*, Plato's account of the emotional 'part' seems more positive: focusing on emotions of reverence and awe, he shows how this element in the personality makes a crucial contribution to virtue and understanding.

For Aristotle, emotions are combinations of a feeling of pleasure or pain with a belief (or perhaps a more rudimentary cognitive attitude, a seeing *x* as *y*). Fear, for example, combines painful feeling with the thought that there are bad events impending. The combination is not casual: the pain is pain *at the thought of* that impending danger. Therefore, changes in belief will change emotions. (Aristotle makes these arguments in *Rhetoric*, showing how an orator can manipulate the passions of the audience.) In normative terms, Aristotle argues that the virtuous person is one who has attained balance and appropriateness in emotion as well as action. A person who completely lacked anger at an insult to loved ones, for example, would be culpably deficient; but excessive anger is strongly criticized, and the condition to aim for is 'mildness of temper' (see Aristotle §§22-4).

Epicurus and his school (including the Roman poet Lucretius) presented impressive accounts of several emotions, in particular the fear of death (see Epicureanism §13). They argued that this fear poisons individuals' lives and causes social distress. Since they agreed with Aristotle about its cognitive basis, they argued that it could be

completely removed by teaching people that death is not a bad thing for the person who has died.

The Greek and Roman Stoics were the great passion theorists of antiquity. They produced impressive analyses and taxonomies of the passions, arguing powerfully in favour of the view that they are essentially evaluative judgments that ascribe to things or persons outside our control great importance for our flourishing (see Stoicism §19). This analysis is continuous with Aristotle's, but it goes a step further, denying that there is any bodily feeling over and above the cognition that is essential to the identity of a particular emotion. This apparently counter-intuitive view is rendered plausible by extensive consideration of the power of thought to transform the personality.

One might accept this analysis of emotion and still hold, with Aristotle, that many of the judgments involved in emotions are true and appropriate: when one's child dies, for example, it is right to think that something of enormous importance has been lost. The Stoics, however, held that all the judgments involved in all the passions are false, because all ascribe too much importance to external things and persons, making people dependent on the world for their happiness. To this argument they added several others: emotions weaken the personality, robbing it of force and integrity; they make one prone to excessive and violent acts; and, finally, they are not necessary to motivate good acts, which can be chosen out of duty alone.

Their conclusion is thus that the emotions should be extirpated from human life. Moved especially by the damage done by anger in social life, and convinced that there is no getting rid of anger without getting rid of the attachments to externals that are also involved in love and grief and fear, they came to the radical conclusion that we can stop cruelty and violence only by cultivating utter detachment from everything that used to matter to us. They then strove valiantly to show that we can motivate an active concern for humanity without relying on emotion. Removing emotions is not expected to be like removing false beliefs about more trivial matters: because the evaluations involved are transmitted early through social and parental teaching, they have become deeply habitual, and can be changed, if at all, only through a lifetime of patient effort. Their hope was that as this effort is exercised with greater success, the personality as a whole would become enlightened.

Because their analyses are so compelling and humanly rich, and because they focus with convincing examples on the depredations of emotion in politics, writers such as Seneca and Marcus Aurelius continue to make this radical thesis compelling.

4 Modern supporters of the Stoic normative thesis

The Stoics set the agenda for early modern moral theories of emotion. Most thinkers, though not all, accept some version of the cognitive Stoic analysis of passion; they then differ sharply over the normative thesis. Descartes (1649) modified the Stoic analysis to fit into his divided account of mind and body, but his definitions of the passions are related closely to those of the Stoics; his normative judgments, while not quite as severe, are similar (see Descartes, R. §10).

Spinoza accepted the Stoic analysis and the radical normative thesis, holding, with Seneca, that philosophical therapy can free us from bondage to our emotions (see Spinoza, B. de §9). He gave the normative programme new urgency by insisting that all strong attachments are essentially ambivalent: we love an external person or thing because we find that it assists our efforts to flourish. But anything that can assist us, so long as it is separate from us, can also frustrate us; so all love is mixed with hate. Like the Stoics, Spinoza held that there is a type of joy that is not an emotion in the sense that it is not based on a overvaluation of the significance of externals; it is the contemplative joy with which one regards the deterministic system of the universe as a whole. Like the Stoics, Spinoza identified that order with god; thus the good state is designated the 'intellectual love of god'.

Kant did not accept the Stoic analysis of the passions. Oddly enough, he presented no arguments against it, and no substantial analysis of his own; but he plainly conceived of passions as pre-rational, impulsive and undiscriminating. Thus, the only role he saw passions playing in virtue was a rather mechanical one, as forces that either aid or impede duty. In general, he thought they impede it, and he therefore conceived of virtue as a kind of strength of will, in which the will maintains its control over the potentially disrupting passions.

Kant was well aware of the Stoic normative view. To a great extent he shared it, praising the Stoics for cultivating detachment from passion, and for promoting active beneficence rather than relying on compassionate emotion. But he showed considerable ambivalence, and was unwilling to dismiss compassionate emotion utterly. He understood

that it may be difficult to motivate beneficence without this emotion, that motives of duty by themselves may not suffice. He therefore urged people to seek out experiences in which they will naturally be moved to compassion, so that their beneficence will have the strength of passion behind it.

A surprising ally of the Kantian position was Nietzsche (1881), who objected to pity as an emotion that insults the dignity of the suffering person by implying that this is a person who really needs the things of this world. Equipped with a rather romantic picture of strength and self-sufficiency, Nietzsche believed that the truly strong can rise above life's ills through will, and therefore do not need compassion, grief or fear. Citing the Stoics, Spinoza and Kant as his predecessors, Nietzsche denounced the Christian/democratic tradition of praising compassion as a basic social motive.

5 Modern anti-Stoics: the defence of sympathy

Opposition to the Stoic normative thesis began in the ancient world, with Augustine, who insisted that the goal of self-sufficiency was an inappropriate one for a Christian to pursue. Characterizing the ideal Stoic wise man as obtuse in his detachment from longing and pain, he concluded: 'The fact that something is tough does not make it right; and the fact that something is inert does not make it healthy.' Subsequent Christian views have agreed in characterizing moral error as due to wrongly placed love, and in making love itself an essential ingredient of the cure. Dante's *Divina Commedia (The Divine Comedy)* (1313-21) famously ends with the visionary experience of a harmony of emotion and will. Because emotional error is understood (in an anti-Stoic fashion) to be rooted in the body itself and its sexuality, it is expected that this perfected state will arrive only in paradise, after the purifications of purgatory. Most important Christian thinkers make pity or compassion central parts of Christian life, arguing that these emotions are appropriate expressions of and responses to our vulnerable and imperfect earthly condition.

The eighteenth century saw the flowering of a number of distinct approaches to the sympathetic emotions. Although Christian views did not cease to exert an influence, the notion of original evil was contested increasingly in favour of a view of natural goodness in which basic emotional equipment played a salient role. Hutcheson and Hume made the capacity for sympathy a basic part of human nature (see Hutcheson, F. §2; Hume, D. §3). Hume (1739/40) made many valuable observations about the role of sympathy in motivating moral conduct (see Moral motivation §6). But his conviction that desires and passions are basically noncognitive, and that reasoning is capable only of devising means to ends set by desire, led him to short-change many aspects of the passions that even his own concrete analyses at times acknowledge. Although he connected passions with a characteristic object, he seems not to have treated the connection as essential to the identity of the passion: a passion is simply a particular type of impression caused by the object. Hume's enormous influence has led to the sharp split between passion and cognition that characterizes much modern Anglo-American thought (see Emotions, philosophy of §§3-4).

Rousseau gave the emotion of *pitié*, compassion, a central role in social morality (see Rousseau, J.-J. §4). Describing the education of young Émile (1762), he shows that the experience of being astonished and pained at the pain of another is an essential foundation for all society. Compassion leads us to recognize the vulnerability that all humans share, and to appreciate the pain that disasters of various types cause others. Thus, the emotion leads us to be sceptical of distinctions that situate some people as high above others, their fortunes as vastly more secure. Rousseau evidently thought that emotions, involving complex imagining and thought, convey moral information that it would be difficult to obtain in any other way. Accepting a version of the Stoic idea of original human innocence, he seems to have held that excesses in self-love derive from experience of unequal social conditions.

Adam Smith's *The Theory of Moral Sentiments* (1759) provides one of the modern tradition's most detailed analyses of the moral contributions of emotion. Heavily indebted to the Stoics, Smith develops a richly detailed account of the cognitive content of passions such as anger and sympathy. He differs with the Stoics on the normative question, taking up a more Aristotelian view about attachments to family and loved ones. His device of the 'judicious spectator', modelled on Stoic conceptions of self-scrutiny and conscience, gives the moral agent a way of estimating the point of propriety in passion: we are to ask what a concerned spectator, not personally involved, would feel at the events, and this will help us identify bias and irrationality deriving from our own personal immersion. Despite his positive view of sympathy and other emotions, Smith thought them inconstant and

unreliable as guides to social choice, since he observed that we are most easily moved by events closest to ourselves: an earthquake in China means less to a person than an injury to his own finger (see Smith, A. §§2-3).

Another philosopher who defended the fundamental role of compassion in morality was Schopenhauer (1840). Criticizing Kant, he argued that all genuinely moral action must be grounded in other-directed emotion (see Schopenhauer, A. §6). Imagining compassion as involving a mysterious union of the self with the other, he leaves the reader unclear as to precisely why it is not, therefore, a form of self-concern. Schopenhauer's account had, nevertheless, tremendous influence in a culture dissatisfied with duty-based views.

6 Contemporary views: influences from psychology and anthropology

The middle of the twentieth century saw the rise of noncognitive views of emotion in both behaviourist psychology and in Freudian psychoanalysis (see Behaviourism, methodological and scientific; Freud, S.; Psychoanalysis, post-Freudian). Philosophers such as Anthony Kenny, Robert Solomon, Robert Gordon, William Lyons, Ronald de Sousa and Michael Stocker have by now demonstrated the poverty of such accounts and our need for some type of cognitive view (see Emotions,nature of §2). (Kenny also in the process offered cogent criticisms of Hume's elusive and influential account.)

During this period, cognitive psychology was itself evolving. By now, psychologists no longer expect to be able to give purely behavioural accounts of emotion in terms of stimulus and response, and most of the dominant accounts hold that emotions are a form of intelligent interpretation in which an animal takes in news of how things are in the world with respect to its most important goals and projects. Richard Lazarus (1991) has argued that emotion's evolutionary contribution is best explained this way: emotions give animals information that is essential to survival. Psychologists generally agree that we have reason to credit many animals at least with the capacity for complex appraisals suited to this account of emotion. Their work has enriched the philosophical debate.

At the same time, anthropologists have produced fascinating accounts of the role played by social norms in shaping the emotion-categories of different societies. These accounts raise questions about the extent to which an emotional repertoire is malleable; at least with respect to specific sub-types and internal demarcations within categories, there appears to be considerable cross-cultural variation in emotion-types, and also, presumably, at least some variation in emotional experience (see Moral sentiments §3; Morality and ethics §4).

This returns us to the moral problem with which the Stoics grapple: for if we see that emotions are learned with the learning of social norms, and we are convinced that our society is not perfect, then it might be unwise to trust the emotions too much as guides to conduct. One society described by anthropologist Jean Briggs (1971), the Utku Eskimos, actually embodies a relatively successful Stoic programme for the elimination of anger, thus prompting us to ask how we should think about our own goals and moral projects.

7 Feminist accounts of morality and emotion

Women have frequently been devalued on the grounds of being emotional and making decisions by consulting emotion. This has led many feminist thinkers to defend the contribution that emotions might play in moral reasoning, arguing that it is the (male) denigration of emotion that is mistaken. Philosophers such as Annette Baier, Lawrence Blum, Sara Ruddick, Virginia Held and Nel Noddings have defended a normative ethical view in which care for others plays a central role, and have seen this as a way of promoting feminist goals (see Feminist ethics).

On the other hand, feminists who consider the social origins of the appraisals involved in emotion have reasons for doubt. If women have a great propensity to care for others, is this always a good thing? Is it not the case that men have standardly urged women to see themselves as care-givers, rather than as sources of worth and agency in their own right? J.S. Mill (1869) argued that even women's desire to please men has a social origin and is a legacy of women's subordination. This line of argument has been continued recently in the work of Catharine MacKinnon (1989), who argues that women's emotions and desires are in significant ways created by inequality and injustice, and that we should therefore be highly sceptical of the claim that women's instincts of care are always trustworthy and morally valuable.

It is not necessary to choose between these extremes. We can hold that many emotions are valuable parts of the moral life without giving implicit trust to any that have been shaped in imperfect and unjust social conditions -

judging, with Aristotle and Adam Smith, that emotions can be good guides as elements in a life organized and examined by critical reflection.

See also: Emotivism; Family, ethics and the

MARTHA C. NUSSBAUM

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985, books II-IV.(On the role of passions in virtue.)

Aristotle (*c*. mid 4th century BC) *Rhetoric*, trans. G. Kennedy, *Aristotle on Rhetoric: A Theory of Civic Discourse*, New York: Oxford University Press.(Illustrates Aristotle's view that, since emotions are combinations of a feeling of pleasure or pain with a belief, changes in belief will change emotions.)

Baier, A.C. (1994) *Moral Prejudices: Essays on Ethics*, Cambridge, MA: Harvard University Press.(Discusses the role of care in moral theory.)

Blum, L. (1980) *Friendship, Altruism, and Morality*, London: Routledge.(Defends the role of compassion in moral judgment.)

Briggs, J. (1971) *Never in Anger*, Cambridge, MA: Harvard University Press.(Anthropological study of a society that removes anger.)

Dante (1313-21) *Divina Commedia*, trans. J. Ciardi, *The Divine Comedy*, New York: E.P. Dutton, 3 vols, 1989. (Highly influential medieval Christian account of love, which ends with the visionary experience of a harmony of emotion and will.)

Descartes, R. (1649) *Les passions de l'âme*, trans. *The Passions of the Soul*, in *The Philosophical Writings of Descartes*, ed. and trans. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, vol. 1, 1984.(Analysis of passions in accordance with mind-body dualism.)

De Sousa, R. (1987) *The Rationality of Emotion*, Cambridge, MA: MIT Press.(Excellent account of the cognitive content of emotions and its evolutionary role.)

Gordon, R. (1987) *The Structure of Emotions*, Cambridge: Cambridge University Press.(Valuable philosophical analysis of emotion/cognition relationship.)

Held, V. (1993) *Feminist Morality: Transforming Culture, Society, and Politics*, Chicago, IL: University of Chicago Press.(Care ethics and social justice.)

Held, V. (ed.) (1995) *Justice and Care: Essential Readings in Feminist Ethics*, Boulder, CO: Westview Press. (Collection of different feminist viewpoints on care ethics.)

Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978.(Influential account of the relationship between emotion and action.)

Kant, I. (1797) *Metaphysische Anfangsgründe der Tugendlehre*, trans. J.W. Ellington, *Metaphysical Principles of Virtue*, Indianapolis, IN: Hackett Publishing Company, 1964; repr. in *Immanuel Kant: Ethical Philosophy*, Indianapolis, IN: Hackett Publishing Company, 1983.(Discussion of role of emotions in virtue; critique of pity as moral emotion.)

Kenny, A. (1963) *Action, Emotion and Will*, London: Routledge.(Influential and devastating critique of behaviourist and Humean conceptions of desire.)

Klein, M. (1921-45) *Love, Guilt, and Reparation and Other Works, 1921-45*, London: Tavistock, 1985.(Central psychoanalytic account of emotion and morality.)

Lazarus, R. (1991) *Emotion and Adaptation*, Oxford: Clarendon Press.(Critique of behaviourism and account of the field's return to cognitive conceptions of emotion by leading cognitive psychologist.)

Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press, 2 vols.(Basic sources for Hellenistic views of morality and emotion, with fine commentaries.)

Lutz, C. (1988) *Unnatural Emotions: Everyday Sentiments on a Micronesian Atoll*, Chicago, IL: University of Chicago Press.(Valuable anthropological study of role of social norms in emotion.)

Lyons, W. (1980) *Emotion*, Cambridge: Cambridge University Press.(Well-argued defence of a cognitive-Aristotelian account.)

MacKinnon, C. (1989) *Feminism Unmodified*, Cambridge, MA: Harvard University Press.(Argues that women's subordination to men has shaped their emotions and desires.)

Mill, J.S. (1869) *The Subjection of Women*, in J. Gray (ed.) *On Liberty and Other Essays*, Oxford: Oxford University Press, 1991.(Argues that women's emotions are shaped by unjust social conditions.)

Nietzsche, F. (1881) *Morgenröte*, trans. R. Hollingdale as *Daybreak*, ed. M. Clark and B. Leiter, Cambridge: Cambridge University Press, 1997.(The central locus for Nietzsche's attacks on pity.)

Noddings, N. (1984) *Caring: A Feminine Approach to Ethics and Moral Education*, Stanford, CA: Stanford University Press.(Argues that women's experience of maternal care should be the basis for ethics; opposes abstract principles.)

Nussbaum, M. (1994) *The Therapy of Desire: Theory and Practice in Hellenistic Ethics*, Princeton, NJ: Princeton University Press.(Discussion of ancient Graeco-Roman views of passion and morality.)

Oatley, K. (1992) *Best-Laid Schemes: The Psychology of Emotions*, Cambridge: Cambridge University Press. (Valuable account by a cognitive psychologist.)

Plato (*c*.386-380 BC) *Symposium*, trans. A. Nehamas and P. Woodruff, Indianapolis, IN: Hackett Publishing Company, 1989.(Describes the ascent of love to contemplation of the immortal form of beauty.)

Plato (*c*.380-367 BC) *Republic*, trans. G.M. Grube, Indianapolis, IN: Hackett Publishing Company, 1992, book IV. (Account of the structure of the soul and the role of the passions in virtue.)

Plato (*c*.366-360 BC) *Phaedrus*, trans. A. Nehamas and P. Woodruff, Indianapolis, IN: Hackett Publishing Company, 1995.(Illustrates how emotions play a positive role in connecting the person to beauty and truth.)

Rousseau, J.-J. (1762) *Émile: ou, de l'éducation*, trans. A. Bloom, *Emile: or, On Education*, Harmondsworth: Penguin, 1991, book IV.(Contains a fundamental account of compassion's role in social morality.)

Ruddick, S. (1989) *Maternal Thinking: Toward a Politics of Peace*, Boston, MA: Beacon Press.(Balanced account of the moral importance of maternal care.)

Schopenhauer, A. (1840) *Über die Grundlage der Moral*, trans. E.F.J. Payne, *On the Basis of Morality*, Providence, RI and Oxford: Berhahn Books, 1995.(Argues that all proper moral action must be grounded in other-directed emotion.)

Seneca, L.A. (before AD52) *On Anger*, in J.M. Cooper and J. Procopé (eds) *Seneca: Moral and Political Essays*, Cambridge: Cambridge University Press, 1995.(Influential account of the removal of anger as social goal.)

Sherman, N. (1989) *The Fabric of Character: Aristotle's Theory of Virtue*, Oxford: Clarendon Press.(Excellent account of role of emotion in virtue.)

Smith, A. (1759) *The Theory of Moral Sentiments*, ed. D.D. Raphael and A.L. Macfie, Oxford: Clarendon Press, 1976.(Rich account of balance in passion, and of role of sympathy in morality.)

Solomon, R. (1976) *The Passions: Emotions and the Meaning of Life*, Indianapolis, IN: Hackett Publishing Company; repr. 1993.(Influential philosophical account of emotion.)

Spinoza, B. (1677) *Ethica Ordine Geometrico Demonstrata (Ethics Demonstrated in a Geometrical Manner)*, trans. E. Curley, *Ethics*, Harmondsworth: Penguin, 1996.(Analyses our 'bondage' to passions as cause of strife, urges a therapeutic programme to remove passions.)

Stocker, M. (1996) *Valuing Emotions*, Cambridge: Cambridge University Press.(Valuable analysis of evaluative dimensions of emotion.)

Morality and ethics

Morality is a distinct sphere within the domain of normative thinking about action and feeling; the whole domain, however, is the subject of ethics.

How should the moral sphere be characterized? The three most influential suggestions are that morality should be characterized by its function, *by the* supremacy *of the moral, or by the distinctive moral* sentiments. *It is plausible that moral codes have a social function, such as that of maintaining beneficial cooperation; but it does not seem an a priori truth. In contrast, it may be true a priori that moral obligations are supreme - accepting an obligation as moral is accepting that it should be carried out whatever else may be said against doing so. But even if this is a priori, it does not provide a criterion for demarcating the moral. A better characterization takes an obligation to be moral if and only if certain sentiments, those invoked in blame, are justified towards an agent who fails to comply with it.*

This provides a criterion for demarcating the moral, but only if the sentiments can be identified. The sentiment at the core of blame is sometimes held to be a species of anger - indignation, for example. However it seems that one may feel the sentiment involved in guilt or blame without feeling indignation. A view deriving from Hegel's conception of wrongdoing may be more accurate. Whereas indignation disposes to aggressive restorative action, the sentiment of blame itself disposes to withdrawal of recognition, expulsion from the community. Punishment can then be seen, with Hegel, as a route whereby recognition is restored.

Criticisms of morality are broadly of two kinds, though they often overlap: that moral valuation rests on incoherent presuppositions, and that morality is a dysfunctional system. The leading source of the first kind of criticism (and one source of the second) is Nietzsche; in contemporary philosophy related ideas are developed by Bernard Williams. One of Williams' criticisms centres on something which does indeed seem to be presupposed by moral valuation, at any rate in modern moral thought: that moral obligations exist independently of one's desires and projects yet of themselves give one a reason to act. Other doubts about the coherence of the moral focus on a conception which, again, may be distinctively modern - being associated particularly with some forms of Protestant Christianity and with Kant; the conception takes it that all are equally autonomous and that the only true worth is moral worth. Criticisms of this conception occur (in different ways) in Nietzsche's treatment of modern morality and in Hegel's treatment of what he calls Moralität.

The idea that morality is dysfunctional, that blame and guilt deny life or impose pain without securing compensating gains, has considerable influence in contemporary culture (as does the idea that they are compromised by the interests of those who can shape them). Such criticism must come from a conception of ethical value, and assume that there is an alternative to morality. Unless one believes in the possibility of a communal life unmediated by any disciplinary forces at all, the assumption being made must be that there could be a discipline which was better, ethically speaking, than the discipline of guilt and blame.

1 Morality, practical deliberation and ethics

Morality is but a part of the whole domain of normative thinking about action and feeling. Many questions about what one should do or what kind of activity or character one should admire are not ordinarily held to be moral questions. Furthermore, it may be asked whether one should do what one has a moral obligation to do, or why one should do it. The question is not empty: those who ask it are not asking whether or why they should do what they should do, or have a moral obligation to do what they have a moral obligation to do. They are deploying two distinct concepts.

It is possible to go further and be sceptical about the very tenability of the notion of moral obligation, without denying that there are reasons for acting (see Moral scepticism). In a well-known passage, Nietzsche denies that any moral judgment is true. He denies morality, he says, as he denies alchemy. But he adds:

> It goes without saying that I do not deny - unless I am a fool - that many actions called immoral ought to be avoided and resisted, or that many called moral ought to be done and encouraged - but I think that the one should be encouraged and the other avoided *for other reasons than hitherto*. We have *to learn to think differently* - in order at last, perhaps very late on, to attain even more: *to feel differently*.

(1881: §103; original emphasis) (see Nietzsche, F. §§8-9)

We may use the term ‘ethics’ broadly to cover the normative theory of conduct, of what reasons there are for doing this rather than that. Through its concern with reasons for action, ethics will also be concerned with ideas about ends and character-ideals in so far as these bear on action (see Ideals). And by that route it also becomes concerned with questions about what there is reason to feel, and how reasons to feel connect with reasons to act. Inevitably, then, it shades into aesthetic questions; but not all questions of value dealt with in aesthetics fall within its scope. The dominant concern of ethics is conduct, right action - right feeling comes in in so far as it bears on right action rather than as a fully general subject of study in its own right. In this broad sense Nietzsche’s repudiation of morality and insistence that we have to learn ‘to feel differently’ remains an ethical teaching.

2 Characterizing the ‘moral’

How then is morality to be characterized? A number of suggestions have been put forward.

(1) By its *function*. Morality is that set of convictions whose function is to promote human flourishing, to enable us to live together on terms of mutually beneficial cooperation - or whatever one’s doctrine as to its function may be. (2) By its *supremacy* in relation to other deliberative conclusions. Morality is a subsystem of deliberation about what one ought to do, whose conclusions in principle override all others. To take an obligation as a moral obligation is to take it that one should fulfil it whatever else can be said against doing so. (3) By the distinctive *sentiment* which is involved in moral valuations (see Moral sentiments). On this view, a precept is moral if particular sentiments are appropriate towards those who do not comply with it - the sentiments involved in guilt and blame (see Praise and blame; Rectification and remainders §4).

According to (1), then, the morality of a society will consist in those of its generally acknowledged rules or convictions which can be characterized as functioning in it to produce a certain outcome. If this is taken as a definition, ‘morality’ is a functional term, like ‘tuner’. A tuner in a sound system is by definition the device in the system whose function is to enable a user to select and receive radio signals. It is not obvious, however, that there is any function which morality by definition has. Certainly it is plausible that the existence of commonly acknowledged moral convictions has the effect of facilitating, for example, beneficial cooperation. But to say that that is their function, as against their consequence, is to say something stronger: roughly, that that is what they are there for. This is a substantial thesis. It may be true, for instance, in virtue of natural selection, or God’s design (Evolution and ethics; Religion and morality §3). Even if moral convictions do have a particular function, though, it is still not obvious that that is what it is, by definition, for them to be *moral* convictions.

(2) holds that the criterion of morality is the supremacy of moral obligations. In a positive sense the morality of a person or society is that which they take to be supreme; in a normative sense morality is that which is supreme. In the normative sense, then, accepting that an obligation to do X is a moral obligation is accepting that the obligation is of such a kind that one should do X whatever else can be said against doing X.

But what kind of obligation is that? In one sense any deliberative conclusion about what I should do, all things considered, is supreme: if my conclusion is correct then I should do that thing, whatever it is, despite everything which can in fact be said against it. However the idea of supremacy is that the total set of considerations, moral and nonmoral, cannot outweigh the moral ones alone. This may be true, and indeed true a priori. But it cannot serve as a definition because it already presupposes a distinction between moral reasons and reasons other than moral ones.

Consider (3). Hume said:

> When you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it.
>
> (1739-40: 469)

He appears to define moral censure in terms of the feeling or sentiment of blame - but too simply. Suppose I pronounce something to be vicious without having any feeling or sentiment of blame towards it. It does not follow that my pronouncement is false. The question of its viciousness turns rather on whether it is appropriate or reasonable to blame it. So a better analysis would make the normative element explicit: when you pronounce any

action to be vicious you are claiming that the feeling or sentiment of blame is an appropriate response to its agent. This explicitly normative line was taken by J.S. Mill (1861); he implied that believing oneself the object of justified blame is already a form of punishment. This 'idea of penal sanction' enters into the notion of moral wrongness: an action is morally wrong if it is appropriate to respond to it with the penal emotions involved in blame (see Mill, J.S. §11).

3 Blame

Blame is an act which stands to a specific feeling, the blame-feeling or sentiment of disapproval, as the act of apology stands to feeling sorry, the sentiment of regret. In each case the action is a judgment that a certain feeling is appropriate towards an action. This is a narrow notion of blame, blame as censure. It is not the wider notion in which we can blame the cheap pen for the ink blot or the weather for the traffic delays. For we are not then saying that a particular penal sentiment is appropriate towards them, but only that they were the relevant cause of a bad outcome (see Praise and blame §1). Note also that since blame is an action, not a feeling, there may be situations in which it is unreasonable to blame explicitly, or blame at all, even though it is reasonable to feel the sentiment which is invoked in blame. If someone is in an emotionally distressed state it may be wrong to burden them with blame even if they are indeed blameworthy. And it might be expedient to put aside guilt feelings, refrain from self-blame, when time is short and there are things to be done, even though the feeling is reasonable. So defining the morally wrong as the blameworthy - understood as that towards which it is reasonable to have the blame-feeling - leaves open the possibility that one should not blame an action even though it is morally wrong.

If the morally wrong is characterized as the blameworthy it becomes possible to characterize the other moral concepts by reference to the sentiment of blame. The morally right is that which it is morally wrong not to do. 'X is morally obligatory' will hold just if non-performance of X is blameworthy. We admire people for going beyond the call of duty even though we do not blame them for not doing so: admiration of such actions is moral admiration if we admire them for the reasons that impelled them and those reasons are moral reasons (see Supererogation). And we can say that they are moral reasons - as against, say, prudential or aesthetic reasons - when their absence from a person's mind beyond a certain point becomes blameworthy. The virtues involve sensitivity to various types of reason for acting which we recognize as moral reasons; they are also traits of character which we could be blamed for not attempting to attain or lose, when it is possible for us to do something to attain or lose them (see Virtues and vices §§2-3). This sets them apart from admirable qualities in general.

The last point is controversial. Hume (1748, 1751) treated the distinction between virtues and talents, vices and defects, as 'merely verbal', and understood blame in a correspondingly wide sense, as referring to any disvaluing emotion. He took himself in this respect to be following the ancients. Aristotle, for example, argued that we disapprove of those who become ugly through 'want of exercise and care' (*Nicomachean Ethics* 1114a23-25). (He is translated as saying that we 'blame' or 'censure' them, but these words may be too narrow for the kind of disapproval, thinking less of, that he has in mind.) But he added that we do not disapprove of those who are naturally ugly - maintaining a link between disapproval and the voluntary even if he does not distinguish moral defects from disapprovable defects in general (see Aristotle §20). Now, however, we do make the distinction and it plays a role in our feelings. We would blame a person who is wantonly ugly only in contexts which generated an obligation to look after one's looks (they accepted a job as a model and bankrupted the company). Otherwise their ugliness, whether it be natural or wanton, may provoke derision or disesteem - and disapproval if it is wanton, though it still remains their business. Thus we make a three-way distinction between moral blame, more general forms of disapproval which link with the voluntary, and disvaluing emotion as such. If Aristotle seems to conflate the first two elements in the distinction, Hume seems to conflate all three.

On the other hand Mill's notion that moral disapproval is penal picks out the relevant difference. But even if we accept that there is a distinctively penal sentiment of blame it does not follow that the morally wrong can be defined in terms of it as that towards which it is a justified response. This could not, in the first place, be a definition if it proposed that the morally wrong is that which there is a moral obligation to blame. We have distinguished, however, between the act of blame and the sentiment of blame itself; to judge that the sentiment is justified is to make a judgment about what one ought to feel in the sense of what feeling is reasonable. That need not be, arguably cannot be, a moral 'ought'.

A more serious objection arises from the intentional content of the blame-feeling. Is not the blame-feeling directed

properly only at that which is believed to be morally wrong? This circularity is harder to set aside; it may force the conclusion that 'morally' in 'morally wrong' is a semantically primitive modifier (see Skorupski 1993). But even so the thesis that an action is morally wrong if and only if the sentiment of blame towards its agent is justified (or some refinement of that thesis) may be a conceptual truth which supplies a criterion for demarcating the moral system of evaluations in a society from others, by reference to the distinctive sentiment involved in guilt and blame - so long as we can get some independent grip on that sentiment.

Characterizing morality in this way is consistent with holding that it has a function. Our spontaneous responses as to what is reasonably blamed serve to maintain beneficial social cooperation; it is not implausible that they persist in their character and content because they do that. What of the supremacy of morality? Is the claim that moral obligations are overriding consistent with characterizing the moral in terms of sentiments of blame? Suppose it would be blameworthy not do X. Does it follow that one should do X, whatever can be said against doing it? At least this much seems true: if there is no reason why a person should not do X - where 'reason' is being used in the general deliberative way - then there is no reason to blame that person for doing X. This is a point about the intentional content of the blame sentiment, like the point that if something is in no way dangerous there is no reason to fear it. If it is correct, and if an action is morally wrong just if it is reasonable to blame a person who does it, then it follows from the fact that it is morally wrong that there is reason not to do it, without need of any extra premise. We might call this the thesis that morality is constraining. To say that morality is supreme is stronger. It requires something like this: if doing X is, all things considered, optimal (there is nothing there is more reason to do) then there is no reason to blame a person for doing X. If that is intrinsic to the intentional content of the blame sentiment, then so is the supremacy of morality.

4 Do all societies have morality?

All sustainable societies have spontaneous disciplinary systems maintaining solidarity and prohibition - in the sense that they flow from spontaneously shared favourable and hostile attitudes to actions and are not codified systems of law enforced by instituted penalties. Artificial disciplines psychologically presuppose them and could not sustainably replace them, not just because (outside utopia) the enforcement costs would be too great.

But need these spontaneous disciplines be based on the emotions involved in guilt and blame? Here questions on the borders of philosophy and anthropology arise. Among them are: How do guilt and blame relate to or depend on religious emotions of transgression and expiation or redemption, or to emotions mobilized by the ideas of purity and pollution? And how are they related to shame and disdain? A distinction has been proposed between shame cultures and guilt cultures; spontaneous discipline is realized in the former through the experience of shame rather than of guilt (Benedict 1947). E.R. Dodds (1951) suggested that the development from Homeric to classical Greece was a development from shame to guilt culture; the Christian West is often thought of as *par excellence* a guilt culture (see Moral sentiments §3).

In attempting to pin down a sentiment one looks to its intentional content and the action to which it disposes. Some have suggested that the core of blame is a species of anger - indignation or resentment. Resentment, however, is specifically occasioned by what is taken as injury to oneself. Indignation, it is true, is more general: it is occasioned by what is taken to be wrongdoing whether or not it involves injury to oneself. Yet it seems mistaken to make it the emotional core of blame, even though wrongdoing is its object. Rather, indignation is what people of spirit feel on witnessing a religious or moral violation. It impels them to right the wrong done, aggressing if necessary against the violator. The blame-feeling as such does not dispose to aggressive restorative action: it resembles, in extreme cases, a kind of horror akin to the reaction to pollution and disposes to withdrawal of recognition, casting out of the community (see Recognition). It further disposes to punishment; but punishment is not the expression of anger. When it is not a simple deterrent discipline it is an atonement - a regaining of recognition. Or rather, punishment as such is never just deterrence. The concept is not applicable to the entirely wild, though it can apply in some degree to animals which are incorporated into the social order. The blameworthy but unpunished agent is in a liminal state between community and wilderness. Hegel held punishment to be the right of the wrongdoer; his view is understandable when punishment is seen as the mechanism for regaining recognition. Expulsion is the first moment of blame, redemption and return the second and third. The punishment must be accepted: guilt - self-blame - is a withdrawal of recognition from oneself and a movement to redeeming self-punishment.

This human pattern which lies at the core of blame can plausibly be held to exist in all cultures. But that does not mean that institutional and ideological structures which mobilize and channel it do so in the same way or degree. For example many societies connect transgression to impurity or pollution, and associate with them rituals of cleansing such as the scapegoat carrying its burden of evil into the wilderness (see Atonement). (The impure can have unexpected connections with the sacred: the impure transgresses law, the sacred is above it.) Christianity's heaven and hell take integration into the community or expulsion from it to their furthest limit; this life becomes a liminal phase preceding final integration or expulsion. But we should not infer too easily that the very notions of wrongdoing and punishment depend on such religious or magical extensions. They remain the root notion of criminal as against civil law, for example, even when law is shorn of magic and religion.

How then do guilt and blame contrast with shame and disdain? Central to those latter feelings is the notion of personal standing, status or rank. The behaviour to which disdain disposes is not exclusion from the community but demotion within it, not withdrawal of recognition as such but loss of status. Shame, disdain directed towards oneself, is the experience of loss of standing in one's own eyes. With loss of standing there is no analogue to the recognized antidote, so to speak, provided for exclusion by atonement. Presumably that is because standing requires qualities that may be non-voluntary whereas membership only requires compliance (see Honour). Blame relates to the quality of the will, something redeemable of whose redemption one can give a sign; whereas disdain need not. A shameful defect is not necessarily, or even typically, a defect of will and it may not be possible to remedy it by effort.

The responses of exclusion and demotion are elemental, but develop and shade into each other in many ways. A contrast between 'blame cultures' and 'shame cultures' may be too simple to capture the diversity and continuity among cultures that is actually found. Yet it is plausible that modern Western culture stands out by the marked reliance it places on the discipline of guilt and blame and the way it isolates or 'purifies' that discipline, shearing it of any magico-religious setting, distinguishing it from customary law, and disentangling it from systems of shame and disdain, which it radically downgrades (the most recent step in this process has been the removal of shame from the sexual domain). Theologically, this tendency was led by Protestant Christianity; philosophically, it is epitomized in the abstract, though not the practical, ethics of Kant. One can seek to explain it in various ways: by the internal dynamics of the West's Judaeo-Christian legacy, or as part of the transition from aristocratic values to modern bourgeois notions of individual responsibility, or simply by the general modern tendency to differentiate social functions (see Responsibility). These questions belong to the humanities as a whole. But there is also a question of ethical assessment which must be of special interest to moral philosophy. Granting that the modern West is singular in the degree to which, and the way in which, it isolates and relies on guilt and blame, is this reliance wise? Or does it throw up illusions and downgrade valuable ideals? Does morality as we have come to understand it - 'modern morality' - play an indefensible role in our ethical thought? It was Nietzsche's achievement to put these questions firmly on the agenda. In recent moral philosophy they have been pursued by Bernard Williams (1985; 1993).

5 Criticisms of morality

Criticisms of morality - or of 'modern morality' - fall into two kinds, claiming respectively that it is incoherent and that it is dysfunctional.

The object of the blame-feeling seems to be, as one may call it, transgression. Transgression is avoidable wrongdoing. It is recognizable as such by a responsible agent; responsibility is the capacity to respond to it as 'constraining' in the sense explained in §2 - to acknowledge that there is reason to avoid it as such. Moreover transgressive wrongness is a wrongness independent of the agent's desires and projects. Bernard Williams criticizes this idea of wrongdoing on the ground that there are no 'external reasons' - nothing which one has reason to avoid as such, irrespective of one's desires or projects. If transgression is indeed the intentional object of the blame-feeling, his criticism undermines the coherence of blame: it would follow that nothing is blameworthy, in the way that if nothing is dangerous it follows there is nothing to fear. Williams (1985; 1995), however, holds this notion of wrongdoing, and the corresponding notion of responsibility, to be a fiction peculiar to what he calls 'the morality system' - in present terms, modern morality (see Williams, B.A.O. §§2-3).

Whether there are 'external reasons' is a controversial issue in epistemology and the philosophy of mind (see Moral motivation). Williams' rejection of external reasons stems from a neo-Humean view of those issues which

has been widely influential but is also widely rejected. However there is a distinct criticism, of modern morality at least, also sketched by Williams. Modern morality seems to hold that those who are bound by the moral law at all are bound by it equally. Is this inherent in the blame-feeling as such? Certainly the blame-feeling presupposes no doctrine of degrees of rank, or of ideals of life underpinning such degrees - disdain in contrast works with a notion of what can be expected from the superior and the inferior. But modern morality seems markedly to strengthen this feature of the blame-feeling. Kant (1785) stressed that the moral law must be accessible to any autonomous being and it is part of his metaphysics of the person that everyone has an equal potential for rational autonomy (see Autonomy, ethical). More generally, the problem of 'free will' is typically presented in modern moral philosophy as an all or nothing affair: either all lack free will and thus responsibility (because of the truth of determinism, for example) or all are equally free; either way there are no degrees of moral agency (see Free will). A connected and very powerful modern thought is that no system of valuation has the existential finality of the moral. Its extreme formulation is Kant's thesis that the only real human worth is moral worth; it fits onto the thesis that all are equally capable of acknowledging and acting on the moral law. Taken together the two theses imply that differences of human worth stem exclusively from what is equally within all human agents' control. One may call the idea that this must be so the egalitarianism of pure desert, or of pure autonomy.

The right response to it may be that moral agency does come in degrees and that there are admirable or contemptible forms of agency other than the moral. This may be incompatible with the egalitarianism of pure autonomy but is it incompatible with morality as such? Ordinary moral thought is quite willing to allow that there can be diminished responsibility and that it implies diminished blame. It also allows that natural virtues may be differentially distributed and it does not extirpate systems of valuation other than the moral. Perhaps the egalitarianism of pure autonomy or desert arises from the profoundly Christian, or more narrowly, Protestant character of modern ethical thought and not from moral valuation, grounded in the sentiment of blame, as such.

These lines of response take account of Nietzsche's criticism of Christianity and modern morality (see §1) but seek to disentangle it from his rejection of moral valuation itself. Another influential critique, which focuses on somewhat related aspects of modern morality, is found in Hegel's *Philosophy of Right* (1821). Here too a contrast between Hellenic and modern ethical conceptions plays a vital role. Hegel divides the ethical sphere, which he thinks of as the sphere of freedom, into three aspects or phases - 'abstract right', *Moralität* and *Sittlichkeit* (see Hegel, G.W.F. §8). Abstract right treats of the external, juridical regulation of persons conceived of as property owners and the rights implied thereby. The latter two terms are usually translated as 'morality' and 'ethics' or 'ethical life', but in the present context those translations are misleading, for both terms denote aspects or kinds of morality. *Moralität* develops the subjective side of moral agency. Hegel takes as its characteristic form the Kantian notion of a self-legislating subject evolving, entirely from its own reflection, the content of the moral law. He sees this, in particular its idea that true moral agency arises from rational insight into universal law, as a moment in the development of modern freedom from primitive *Sittlichkeit*, or communal moral order. But he also holds that it is doomed to emptiness by its own purism. Taken as rejection of the very possibility of abstract ethical theory, as it sometimes is, Hegel's criticism would be vitiated by its excessive focus on the specifically Kantian treatment of the issue. But if it is seen as a criticism of the ahistorical and asocial distortions which modern morality's egalitarianism imposes it is on target. It is a critique of the one-sided subjectivity or individualism of certain modern conceptions of moral valuation - but Hegel also thinks that this one-sidedness is bound to be overcome. In modern *Sittlichkeit*, the objective ethical life to which Hegel thinks modernity eventually tends, moral valuation is relocated in institutions whose rationality enables one to achieve freedom by taking up a social role.

Rejecting some of the distinctive components in the ideology of modern morality, then, is not necessarily rejecting morality as such. But it may still be asked whether it is a good thing to entrench the emotions of guilt and blame into a social discipline. The question must come from some conception of ethical value, and it must assume that there is an alternative. To think morality is dysfunctional one must have an ethical end in mind which is not itself moral. For example, a utilitarian might claim that the blame-sentiment should be downgraded as dysfunctional, perhaps after some level of social development has been reached (see Utilitarianism). Blame and guilt are after all hostile feelings which cause pain. Yet is there any conception of ethical value which does not make the idea that we would be better off without the social authority those feelings presently carry utopian? Can human beings reach a level of rationality or goodness at which the need for any discipline, spontaneous or artificial, drops away? If not, the choice lies between spontaneous systems and costly and repressive artificial ones, and then within the

spontaneous, between the system of guilt and blame and other systems such as that of shame and disdain. A society which shifted emphasis towards the latter would also have to shift - how far is hard to tell - from modern individualism and egalitarian respect.

See also: Duty; Moral agents; Moral Judgment; Moral realism; Universalism in ethics

JOHN SKORUPSKI

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. W.D. Ross, Oxford: Oxford University Press, 1954, book III.(Discusses the conditions of responsibility.)

Benedict, R. (1947) *The Chrysanthemum and the Sword: Patterns of Japanese Culture*, London: Secker & Warburg.(Perceptive comments on guilt and shame in Japan and the West.)

Dodds, E.R. (1951) *The Greeks and the Irrational*, London, Berkeley and Los Angeles, CA: University of California Press.(Influential discussion of ethical concepts in Homeric and classical Greece.)

Gibbard, A. (1990) *Wise Choices, Apt Feelings*, Oxford: Clarendon Press.(Includes extended discussion of the moral sentiments.)

Hare, R.M. (1963) *Freedom and Reason*, Oxford: Clarendon Press.(Principles are moral if they are universal, prescriptive and supreme ('overriding').)

Hegel, G.W.F. (1821) *Naturrecht und Staatswissenschaft im Grundrisse. Grundlinien der Philosophie des Rechts* (Natural Law and Politics in Outline. The Principles of the Philosophy of Right), trans. H.B. Nisbet and ed. A. Wood, *Elements of the Philosophy of Right*, Cambridge: Cambridge University Press, 1991, §§82-104. (Discussion of punishment, which deals with Wrong (*Das Unrecht*), a concept whose development, according to Hegel, marks the transition from 'abstract right' to *Moralität*.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978.(Book 3, part I, sections 1 and 2 give Hume's view of the connection between moral valuation and the sentiments. For his view of reasons for action, see book 2, part II, section 3.)

Hume, D. (1748, 1751) 'Of some verbal Disputes', in *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 3rd edn, 1975, appendix IV.(Dismisses the distinction between moral virtues and other good qualities as merely verbal.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. L.W. Beck, *Foundations of the Metaphysics of Morals*, Upper Saddle River, NJ: Prentice Hall, 2nd edn, 1995, ch. 1.(Emphasizes that the moral is available to the common understanding of all.)

Mackie, J.L. (1977) 'The object of morality', in *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin, ch. 5.(A clear statement of the functional view of morality.)

Mill, J.S. (1861) *Utilitarianism*, in J. Robson (ed.) *The Collected Works of John Stuart Mill*, vol. 10, London and Toronto: Routledge and University of Toronto Press, 1969, 203-59.(Mill's characterization of morality comes as part of his account of justice.)

Nietzsche, F. (1881) *Morgenröte*, trans. R. Hollingdale with an introduction by Michael Tanner, *Daybreak, Thoughts on the Prejudices of Morality*, Cambridge: Cambridge University Press, 1982.(This work, with the following two, gives a comprehensive picture of Nietzsche's critique of morality.)

Nietzsche, F. (1886) *Jenseits von Gut und Böse*, trans. R.J. Hollingdale with an introduction by Michael Tanner, *Beyond Good and Evil: Prelude to a Philosophy of the Future*, Harmondsworth: Penguin, 1990.(Contrasts modern morality, conceived as incipiently nihilistic, with Nietzsche's conception of what is noble.)

Nietzsche, F. (1887) *Zur Genealogie der Moral*, trans. with an introduction and notes by D. Smith, *On the Genealogy of Morals: A Polemic by Way of Clarification and Supplement to my Last Book, Beyond Good and Evil*, Oxford: Oxford University Press, 1996.(Three essays dealing with the origins of moral concepts, the nature of guilt, and the meaning of asceticism.)

Skorupski, J. (1993) 'The definition of morality', in A. Phillips (ed.) *Ethics*, Cambridge: Cambridge University Press, 121-44.(Argues that morality can be characterized - though not defined - in terms of blame, and discusses the nature of blame.)

Taylor, G. (1985) *Pride, Shame and Guilt: Emotions of Self-Assessment*, Oxford: Clarendon Press.(Detailed discussion of these related emotions.)

Wallace, G. and Walker, A.D.M. (eds) (1970) *The Definition of Morality*, London: Methuen.(Articles covering a variety of approaches to the definitional problem.)

Williams, B. (1985) 'Morality the Peculiar Institution', in *Ethics and the Limits of Philosophy*, Cambridge, MA:

Harvard University Press and London: Fontana, ch. 10.(The fullest statement of his critique of morality.)
Williams, B. (1989) 'Internal Reasons and the Obscurity of Blame', *Logos* 10: 1-12; repr. in *'Making Sense of Humanity' and Other Philosophical Papers 1982-1993*, Cambridge: Cambridge University Press, 1995. (Describes a 'proleptic' mechanism of blame by which he explains the categoricity and apparent unavoidability of moral requirements.)
Williams, B. (1993) *Shame and Necessity*, Berkeley, CA: University of California Press.(A discussion of Greek and modern ethical thought. Endnote 1 concerns 'Mechanisms of shame and guilt'.)

Morality and identity

Philosophers have drawn connections between morality and identity in two ways. First, some have argued that metaphysical theories about personal identity - theories about what makes one the same person over time - have important consequences for what ought to matter to a rational agent. Second, others have argued that understanding the concrete identities of persons - the social contexts and personal commitments that give life substance and meaning - is essential if moral philosophy is to address real human concerns.

How are metaphysical questions about personal identity supposed to bear on morality? The thought is that what unifies a series of experiences into a single life illuminates what we are, and what we are helps determine how we ought to live. More broadly, it is natural to seek coherence in our metaphysical and our moral views about persons. This pursuit of a comprehensive account has its dangers; perhaps we will tailor a metaphysical view to fit our moral prejudices, or distort moral philosophy and judgment to fit a false metaphysics. But the pursuit has its attractions too; perhaps we will come to understand what we are, and how we ought to live, in a single package.

Philosophers who attend to concrete rather than metaphysical identity characterize persons as committed by social and historical circumstances to a particular range and ordering of values, and as committed by proximity and affection to a particular circle of other persons. These concrete and individual characteristics at least constrain what morality can reasonably demand. But this interpretation suggests that morality stands back from the rich texture of each life, and moderates its demands to accommodate that life. Some philosophers think of morality instead as part of the texture, as intimately connected to, rather than constrained by, concrete identity.

1 Some historical background about personal identity

Does anything about a person survive bodily death? This is a natural way for metaphysical questions about identity to acquire practical and moral urgency (see Personal identity). For example, Plato argued for the existence of an immortal soul, and used this metaphysical view to buttress his moral claims (see Plato §13). Thus, it is always wrong to harm others because doing so also harms your soul, the most enduring and valuable part of yourself. Theologians too have an abiding interest in such issues. For whether persons survive in resurrected bodies or as permanently disembodied souls, they must survive in *some* form to receive God's justice, or God's mercy. The thought of such justice and mercy plays an important moral role for many religious believers (see Soul, nature and immortality of the).

In his paper, 'Of Identity and Diversity' (1694), John Locke focuses not on what might survive bodily death, but on what accounts for identity over time during our earthly lives (see Locke, J. §9). Locke invites us to imagine these possibilities: two sets of character traits and memories (or, for brevity's sake, two psychologies) alternate in possession of a single body; a single psychology is transferred from one body to another; a single soul is associated with two different psychologies; a single psychology has 'underlying' it not a single soul, but rather a rapid succession of souls. Locke argues that our best responses to these imagined cases indicate that neither sameness of body nor sameness of soul is necessary or sufficient for identity. Instead, the right kind of psychological continuity determines personal identity over time.

Locke's positive theory of personal identity has been dubbed the memory theory: at time t_2, I am identical with a person at some earlier time t_1 if and only if I can remember the t_1 experiences *as my own*. This theory gives rise to numerous problems; Locke himself broaches some of them. For instance, Locke admits that, were he to lose irretrievably the memories of certain experiences, he would no longer be the same person that had those experiences. But he tries to soften this odd-sounding claim. He says that he would be the same biological being ('man') in spite of his forgetting; but he would not be the same *person*, because a person is an amalgam of the actions for which that person can take responsibility, and one cannot take responsibility for what one cannot remember. So Locke's sense of the moral role of the concept of the person shapes his metaphysical account of personal identity.

Such manoeuvres did little, however, to forestall criticism of the memory theory. Some criticisms focused simply on identity: Thomas Reid, for instance, elaborated variations on the theme that forgetfulness does not destroy identity (see Reid, T. §7). But some criticisms were moral: Joseph Butler argued that the memory theory provides

such a loose sort of identity that it renders our ordinary self-concern mysterious (see Butler, J. §6). The deeper aim of such critics was to restore the soul to its rightful identity-producing role. But they failed to counter Locke's arguments against the soul's having such a role.

David Hume, in his *Treatise of Human Nature* (1739/40), extended Locke's criticisms of the enduring-soul theory of personal identity by arguing that the concept of the soul has no empirical content. He also revised Locke's positive account of identity by supplementing memory with causation. It is appropriate to speak of a previous experience as my experience, Hume says, even if I cannot remember it, if it played some suitable role in causing other experiences of mine. Hume further claims that all experiences or perceptions are, strictly speaking, 'loose and separate', so that the idea of personal identity over time is a fiction, even if an indispensable fiction (see Mind, budle theory of). But Hume draws no moral conclusions from this extreme or sceptical view. Instead he draws moral conclusions from his attack on the soul, supporting his opposition to religious values ('monkish virtues') by denying the favourite religious account of personal identity.

2 Parfit's metaphysical and moral revisionism

In *Reasons and Persons* (1984), Derek Parfit revives and extends Lockean and Humean themes. Parfit argues that ordinary beliefs about personal identity are mistaken, and that correct beliefs about identity might engender altruism and support consequentialism (which says that an action is right if it promotes the best outcome overall, impersonally considered) (see Consequentialism; Egoism and altruism).

Parfit first describes ordinary belief. We do not believe that the changing surface facts about us account for personal identity over time. Instead, some unchanging 'further fact' constitutes personal identity. When this further fact has received any explication, it has been understood as the Christian soul or the Cartesian ego.

However, Parfit says, there is no good reason to believe in any such further fact. To bear out this and other claims, Parfit produces thought experiments along Lockean lines, but fleshed out with science-fiction details; for instance, he imagines cases of teletransportation and of brain fission. The former procedure involves destroying a body at one place and simultaneously creating a precisely similar new body, realizing a precisely similar psychology, at another place; the latter procedure involves splitting the brain in two and transplanting each half into a new body. Examining our best responses to such cases shows us that we ought to be *reductionists* about personal identity, and that the right kind of reductionism is *psychological* (see Personal identity §3). Personal identity over time is constituted by the holding of certain psychological relations (which Parfit labels 'Relation R'), whether the physical cause of these relations is the stuff of science fiction, or is familiar and normal. (The normal cause of Relation R is the spatiotemporal continuity of the body, including the brain, that realizes the related psychologies). There is, though, this proviso: however extensive the psychological relations that hold between A at t_1 and B at t_2, A and B are not the same person if there exists at t_2 another person, C, whose psychological relations to A are equally extensive. In Parfit's terminology, identity requires that there be *no branching*.

Parfit claims that, from his general account of the metaphysics of identity, it follows that identity (1) is not all-or-nothing and (2) is not 'what matters' (what is most valuable to a rational agent). These conclusions challenge ordinary practical reasoning, which gives pride of place to the determinate and meaningful self.

(1) Identity (hence survival) is a matter of degree, rather than all-or-nothing, for psychological relations that are themselves a matter of degree constitute identity. We can imagine cases where many identity-producing psychological relations hold, but many do not, so that we are *unsure* whether these are cases of identity and survival; our uncertainty shows that in such borderline cases there is no fact of the matter about whether the later person is identical to the earlier person. But even in mundane cases, the unity of a person's past, present, and future 'selves' is a matter of degree, constituted by Relation R. And Relation R holds not just *within* persons, but also *across* persons; for two people can share many values and projects and character traits. So the same relation that imperfectly unifies the self also connects us with other people.

(2) In branching cases, the kind of psychological relation that obtains in unproblematic instances of identity obtains twice. But this very surplus rules out identity-claims; for we should not say that two numerically distinct persons are *both* identical to some earlier person, and we have no grounds for preferring one identity-candidate to the other. Now Parfit believes that few people would say that a branching case is as bad as ordinary death; and

most people would say that such a case is about as good as ordinary survival. (The imagined future contains two agents who share your values, and will pursue your projects as you would have.) Thus branching cases contain what matters in ordinary survival, without containing identity. Parfit concludes that, even in ordinary cases, what matters in survival is not identity, but the obtaining of the right psychological relation with some future person.

What, then, is Parfit’s desired practical and moral upshot? He hopes that the reductionist will have more reason to be *impartial* and *impersonal* than the non-reductionist does (see Impartiality). For reductionism lessens the gap between persons. It shows that my relation to other people is more like my relation to my own past and future than non-reductionist common sense allows. So reductionism undermines the belief that it is most rational for me to pursue my own long-term self-interest; and believing reductionism might cause me to promote my values and projects (part of the content of Relation R) without special regard for the particular people involved. Thus reductionism might move us towards greater *altruism*; and, because this altruism focuses not on other enduring individuals but on impersonally promoting certain values and projects, reductionism might move us closer to *consequentialism*.

However, Parfit qualifies these conclusions, in two ways. First, he admits that a rational person could believe that only a further fact could be what matters, so that reductionism would show that nothing matters. Parfit believes this ‘extreme view’ is rationally indefeasible, but also not rationally compelling. That is, Parfit believes his own optimistic view about what matters in the absence of a ‘further fact’ is as rational as the extreme view is. Second, Parfit hopes not that people will become selfless, only that they will become somewhat less self-interested, when they learn the reductionist truth about personal identity. He suggests that there are evolutionary reasons for thinking in terms of identity, and for acting in self-interested ways (see Evolution and ethics §2). But, where hope for psychological revolution might be naïve, hope for psychological reform might be well-grounded. Parfit hopes for reform, in the direction of consequentialism, once people become reductionists about personal identity.

3 Criticisms of Parfit

We now consider four main criticisms of Parfit.

(1) The most straightforward and ambitious criticism says that personal identity *does* consist in the holding of some ‘further fact’ (the existence of an unchanging soul). Making out this criticism involves showing that the scientific picture of human beings is irremediably incomplete, or at least that Parfit offers no good reason for thinking otherwise. Assessing this ambitious criticism is beyond this entry’s scope; but it is worth noting that the criticism attacks Parfit at his strong point, and where he can find many allies.

(2) A subtler criticism of Parfit’s metaphysics agrees that some form of reductionism is correct, but disagrees with his permissive attitude towards the cause of Relation R and with his exclusively psychological reductionism. The critics (including Peter Unger (1990)) argue that Parfit underestimates the importance of the normal cause of Relation R, namely the physical continuity of a living body. The most plausible reductionism refers to both psychological relations and physical continuity. This partly-physical reductionism captures our deepest notions about what makes for the same person over time in our usual circumstances; the appeal of disembodied Cartesian egos or of Parfitian teletransported psychologies is shallow, compared to the appeal of an embodied this-worldly self. Partly-physical reductionism also does *not* commit us to substantially revising our deepest values. Instead, it helps explain why we value whole persons and maintain clear distinctions between persons, rather than valuing isolated psychological characteristics and blurring the distinctions between persons.

(3) Parfit believes that it is possible and desirable to become less committed to particular persons (oneself included) and more committed to general features of people (character traits, patterns of concern). But perhaps it is a practical necessity that we conceive and order our commitments in terms of whole and enduring persons - enduring, that is, at least over the course of a lifetime. Or perhaps, though we *might* come to think of the human world as a collection of psychological characteristics variously related, this new way of thinking would impoverish our lives. For instance, it might might make it hard for us to conceive of love or friendship.

(4) Parfit argues that ordinary normative practices are founded on false metaphysics, and that coming to believe the true metaphysics gives us opportunity and incentive to revise our practices. But perhaps our practices are founded not on any metaphysical view about persons, but on our circumstances and needs. Perhaps ordinary non-reductionist metaphysics is just window-dressing.

To each of these last three criticisms, Parfit's best response might be a partial accommodation. He has been willing to entertain the possibility that the best reductionist account of personal identity refers to the normal cause of Relation R. He admits that there are limits to our ability to excise identity from practical reasoning, and that even a partial excision must be judged by its consequences. Finally, it would indeed be unreasonable to claim that *all* our ordinary practices are *entirely* shaped by false metaphysical views; but it is possible, perhaps likely, that *some* practices are *partly* shaped in this way. (But which practices depend on false metaphysics? To what extent? Can we revise the practices to fit the true metaphysics? Is this revision morally desirable?)

4 Concrete identity

Advocates of 'concrete identity', the second broad approach to connecting issues of morality and identity, side with Parfit's critics. Parfit's view is, according to concrete identity theorists, one manifestation of a widespread mistaken approach to ethics - an approach that abstracts away from the richness of actual psychology and social and historical context, thus losing sight of real human lives. Kantians appeal to a general and formal feature of persons, their rational agency, in constructing moral theories and testing moral injunctions (see Kantian ethics). Consequentialists see persons simply as causes of good and bad consequences, or the locations where those consequences are embodied. Such artificially restricted views of the person distort one's view of morality, the concrete-identity theorists argue; to see morality clearly, we must see people whole.

Starting from this complaint about abstraction, philosophers have pursued two paths: the first psychological, the second social and historical. Bernard Williams has put the psychological case forcefully, arguing that abstract moral theories miss or dismiss important moral phenomena, and fundamental facts about human motivation (see Williams, B.A.O. §2). Effective reasons for action are deeply personal; how can the stripped-down agents of Kantianism and consequentialism have reasons to act at all? It is absurd to require that agents abandon the commitments that give their lives shape and meaning, in order, for example, to produce the best consequences overall. Close personal attachments have independent moral value; it is impertinent, perhaps monstrous, to demand that we justify our special attachments in light of abstract moral principles (see Friendship; Love §1). In short, moral insight into a person, and effective moral appeals to that person, had better be based on actual features of *that* person's rich psychology, rather than on abstract general features of the person or of moral agency.

Charles Taylor (1989) makes the case for a social and historical approach to concrete identity (see Taylor, C. §§4-5). Taylor does furnish a general theory of sorts about identity: each person shapes an identity through self-interpretative activity, activity that employs the interconnected resources of community, language, and a conception of the good life. This theory connects morality and identity by connecting the questions 'What do I value?' and 'Who am I?'; it also directs our attention towards social and historical particulars. So Taylor's method for investigating our moral situation is to see how the modern identity, that is the modern range of conceptions of the good life, has been shaped and enriched over a long history. The method is not to pare down the concept of identity to its essentials, but to build it up in all its complexity.

It would be folly to deny that concrete identity theorists point at important phenomena. Moral philosophers of all bents should strive at least to show that their theories are consistent with the concrete aspects of identity and of moral life; even better, moral philosophers might strive to connect the abstract and the concrete. But concrete identity theorists often aim at more than providing a salutary reminder to other moral philosophers. They suggest that abstract moral theorizing distorts our lives beyond recognition, that there is no use for the sort of abstraction that whittles people down to a few essential characteristics. In response to these ambitious criticisms, traditional moral philosophers can claim that to abstract away from psychology and culture and history, for special purposes and for the sake of clarification, is not to deny the existence or importance of the things we abstract away from. Understanding our moral lives might require that our attention move back and forth between general features of persons, and persons in their particularity.

This survey has examined two ways in which questions of identity matter to moral philosophers, and two approaches to moral philosophy. Ranged on one side are philosophers who focus on general and trans-historical aspects of being a person or a moral agent. On the other side are philosophers who focus on human beings and societies in their particularity, and who eschew abstraction (or practise a different kind of abstraction, rooted in psychological and historical rather than metaphysical facts). In spite of the apparent conflict between the abstract

and the concrete, it is unclear whether these two approaches to thinking about morality and identity are mutually exclusive, or complementary.

See also: Ideals; Moral agents; Self-realization; Self-respect

IRA SINGER

References and further reading

Flanagan, O. and Rorty, A.O. (eds) (1990) *Identity, Character, and Morality: Essays in Moral Psychology*, Cambridge, MA: MIT Press.(The essays in this volume discuss many of the issues grouped under the label ‘concrete identity’; the volume includes an extensive bibliography.)

Hume, D. (1739/40) ‘Of the Immateriality of the Soul’ and ‘Of Personal Identity’, in *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978, book 1, part IV, sections 5-6.(In three essays, Hume attacks the belief in a substantial or enduring self.)

Johnston, M. (1992) ‘Reasons and Reductionism’, *Philosophical Review* 101 (3): 589-618.(A thorough and forceful criticism of Parfit’s attempt to draw moral conclusions from his metaphysical account.)

Kymlicka, W. (1991) *Liberalism, Community, and Culture*, Oxford: Oxford University Press.(Discusses and responds to criticisms of liberal political philosophy for neglecting concrete identity and community.)

Locke, J. (1694) ‘Of Identity and Diversity’, in *An Essay Concerning Human Understanding*, extended 2nd edn, ed. P.H. Nidditch, Oxford: Oxford University Press, 1975, book II, ch. 27.(In this paper Locke proposes and defends the memory theory of personal identity.)

Parfit, D. (1984) *Reasons and Persons*, New York: Oxford University Press, part III. (Elaborates a reductionist view of personal identity, and argues that adopting this view might bring about morally desirable consequences. Parfit’s work has been the most influential recent discussion of the metaphysics of personal identity, and of the connections between metaphysical identity and morality.)

Perry, J. (ed.) (1975) *Personal Identity*, Berkeley, CA: University of California Press. (Includes the passages from Locke, Butler, Hume, and Reid, referred to in §1; modern revisions and criticisms of Locke’s memory theory; an article by Parfit anticipating the themes of *Reasons and Persons*; and an article by Bernard Williams, ‘The Self and the Future’, suggesting that a bodily criterion of identity retains a strong hold on our imaginations even when it conflicts with a psychological criterion. The collection is an excellent introduction to the issues discussed in §§1-3.)

Rorty, A.O. (ed.) (1976) *The Identities of Persons*, Berkeley, CA: University of California Press.(See especially Perry, ‘The Importance of Being Identical’, for reflections on Parfitian topics, and Williams, ‘Persons, Character, and Morality’, for a seminal discussion of concrete identity.)

Taylor, C. (1989) *Sources of the Self*, Cambridge, MA: Harvard University Press.(A wide-ranging, detailed, yet accessible book, narrating the historical growth of the modern concrete identity.)

Unger, P. (1990) *Identity, Consciousness, and Value*, New York: Oxford University Press.(A sustained and detailed criticism of Parfit, which argues that our deepest intuitions favour a physical standard of personal identity. This volume contains a useful bibliography about metaphysical identity.)

Wolf, S. (1986) ‘Self-Interest and Interest in Selves’, *Ethics* 96 (4): 704-20.(Argues that Parfit’s desired moral reforms may not be possible, probably would not be desirable, and in any event have to be evaluated by normative rather than metaphysical standards.)

More, Henry (1614-87)

The English philosopher Henry More was one of the leaders of the movement known as Cambridge Platonism. Like his Cambridge colleague Ralph Cudworth, More elaborated a constructive metaphysics which, although deeply informed by the new philosophy and science of the seventeenth century, recovered what More saw as an ancient truth or 'cabbala'. The articulation of this truth was an exercise of reason, guided by innate notions or inherent, God-given cognitive propensities. More's ultimate aim as a philosopher was religious or ethical. His 'one main Design*', he explained, was 'The knowledge of God, and therein of true Happiness,*so far as *Reason* can cut her way through those darknesses and difficulties she is encumbred with in this life*' (1662: iv). Among the central themes of the ancient truth More rediscovered and defended were the existence of a God whose leading attributes are wisdom and goodness; the immateriality and immortality of the human soul (the hope of immortality being, as More explained in the Preface to his poem Psychathanasia (1642), 'the very nerves and sinews' of religion); a dualism of active spirit and passive matter that differed significantly from the dualism of Descartes, despite More's early enthusiasm for (and continuing engagement with) Cartesianism; the animation of matter by an immaterial but unthinking spirit of nature; and the existence of an infinite, substantial space, really distinct from matter, in which God is everywhere present and everywhere potentially active.*

More's appeals to experiment in defence of the spirit of nature provoked criticism from Robert Boyle. His doctrine of infinite, substantial space was (in the opinion of some historians) an important influence on Isaac Newton. Space seems, on More's portrayal, to be something divine; this troubled George Berkeley, who thought that by assigning space the 'incommunicable' or unshareable attributes of God, More in the end encouraged the atheism he worked so hard to defeat.

*More is usually represented as a rationalist in religion: '*I conceive*', he once wrote, '*Christian Religion rational throughout*' (1662: iv). It is important to distinguish, however, between More's appeal to reason as a writer defending Christianity, and his appraisal of reason's role in an ordinary Christian life. More was conscious of living in 'a Searching, Inquisitive, Rational and Philosophical Age', and he saw it as his duty to serve God by 'gaining or retaining the more Rational and Philosophical' of his contemporaries in the Christian faith (1664: 482). Rational and philosophical genius was not, however, required of every Christian: although More's accounts of faith vary somewhat from work to work, they typically call not for a rational assessment of argument and evidence, but for moral purity, and for a belief in (and devotion to) a relatively short list of 'essentials'. More sought a statement of these essentials that would reach across Protestant sectarian divides. He also defended a liberty of conscience or religion that was, he said, the natural right of every nation and every person. It was, however, a liberty that could be forfeited, and More thought it* had *been forfeited by some (atheists, for example, and at least some Catholics and Muslims) who might be found claiming its protection.*

1 Life and works

More gives an absorbing account of his early life in the general preface to his *Opera omnia* (1675-9). He was born in Grantham, England, to parents who were 'earnest Followers of *Calvin*'. He himself could never 'swallow down that hard Doctrine concerning *Fate*' because it seemed so at odds with God's justice. As a student at Eton and at Christ's College, Cambridge (where he arrived in 1631), More was possessed by an 'insatiable Desire and Thirst… after the *Knowledge* of things'. At first he took this to be a love of knowledge simply for its own sake, but when he discovered that even the most eager study of philosophy did not quench it, he determined that it was, at bottom, really a form of self-love. He resolved to extinguish or annihilate his own 'proper *will*' and seek 'full union' with the will of God, and was rewarded with the intellectual satisfaction he had earlier failed to find - a 'greater *Assurance*', he said, 'than ever I could have expected'. More became a fellow of Christ's College in 1641, and was ordained in the Church of England in the same year. His career as a writer can be divided into three broad periods, marked not by shifts in doctrine but by changes in the shape and scope of his literary projects.

In the 1640s More published a series of philosophical poems defending the Neoplatonic Christianity he embraced when he came to Cambridge - influenced not only by his study of Plato, but by Renaissance Neoplatonists such as Ficino and Pompanazzi (see Neoplatonism). These poems argue for the soul's immortality, for its pre-existence and for its consciousness after death. Memories ('the bundle of the souls duration') preserve the soul's identity

after it parts from its earthly vehicle. Memories also account, at least in part, for one soul's distinctness from others. It was in the middle of this period, around 1645, that More first read Descartes. In *Democritus platonissans* (1646: [iii]), a poem defending the infinity of worlds on new-scientific and Platonic principles, More describes Descartes as a '*sublime and subtill Mechanick*'. '*Though he seem to mince it*,' More continues, Descartes '*must hold infinitude of worlds, or…one infinite one*', because his 'indefinite' extension is simply infinite extension under another name. This is one of several criticisms More developed in letters he exchanged with Descartes in 1648-9. More's side of the correspondence is deeply respectful - almost reverential - but his disagreements with Descartes were fundamental, and as his awareness of their implications grew, More's attitude towards Descartes and Cartesianism became increasingly hostile.

In the 1650s More translated the philosophy of his poems into prose. Works published in this second period include *An antidote against atheisme* (1653a), *Enthusiasmus triumphatus (Enthusiasm Defeated)* (1656), and *The immortality of the soul* (1659). The period culminated in 1662 with the appearance of *A collection of several philosophical writings of Dr. Henry More*. This incorporated the three earlier prose works (the books for which More is now best known), the *Conjectura cabbalistica* (1653b), and More's correspondence with Descartes.

From 1662 until his death, More completed three main projects in philosophy: Latin summaries of his views on ethics (*Enchiridion ethicum (Handbook of Ethics)* 1667) and metaphysics (*Enchiridion metaphysicum (Handbook of Metaphysics)* 1671); the first volume of *Divine dialogues* (1668), the most accessible presentation of his proofs of God's existence, his account of the natural world, and his theodicy; and the translation of his English philosophical works into Latin (in volume 2 of his *Opera omnia*). These were labours of codification more than of inquiry; as Ward suggests in an unpublished manuscript, in this third period More turned towards theology, proving the 'Truth and Nature of *reveald Religion* as before of *natural*' (Gabbey 1982: 222). More's theological works of philosophical interest include *An explanation of the grand mystery of godliness* (1660), *A modest enquiry into the mystery of iniquity* (1664), and the second volume of *Divine dialogues* (1668).

More remained a fellow of Christ's until his death. He consistently refused to consider larger and more prestigious responsibilities such as the mastership of his college (a position filled by Cudworth) and the provostship of Trinity College Dublin. More's second home, where he often rushed as soon as term time ended, was Ragley, north of Oxford, the estate of his friend Anne Conway. More's correspondence with Conway, published in *The Conway Letters*, is an intimate record of a philosophical friendship. When More, consoling Conway on the death of her two-year-old son from smallpox, tells her that 'there is nothing tragicall to those that are truly good' (1930: 169), his theodicy is tested more severely than it ever could be tested in a treatise.

2 Theory of knowledge

More recognized three sources of knowledge: common notions (which he defined as 'whatever is *Noematically* true, that is to say, true at first sight to all men in their wits, upon a clear perception of the Terms'); external sense (including memory, its 'faithfull Register'); and the 'undeniable deductions' of reason ([1659] 1925: 61). More portrays these sources as clear and compelling and he often claims to demonstrate his main conclusions. But he does not expect his arguments to convince a complete sceptic. 'Perfect *Scepticisme*', he wrote, 'is a disease incurable, and a thing rather to be pitied or laughed at, [than] seriously opposed' ([1659] 1925: 59). When he promises to demonstrate that there is a God, he explains, he does not mean that the reader 'shall be forced to confesse that it is utterly unpossible that it should be otherwise'. It is possible, after all, that '*Mathematical evidence* it self may be but a constant undiscoverable Delusion'. But More's arguments 'shall be such as shall deserve *full assent*, and win *full assent* from any unprejudic'd mind' ([1653a] 1925: 4-5).

More sometimes presents arguments in geometric form, drawing attention to their dependence on an ordered list of 'axioms'. Many of these axioms are laid down as noematic truths but, as More makes clear in several places, his axioms are sometimes merely probable. Some of his axioms are rules for the guidance of inquiry: a crucial axiom in the *Immortality*, for example, declares that whatever is clear from common notions, external sense or rational deduction must '*be held undoubtedly true*' when none of the three sources speaks against it ([1659] 1925: 61). In an *Apology* attached to the *Mystery of iniquity* (1664), More lays down several rules for the conduct of a public intellectual concerned to preserve the peace and authority of the established church. A theorem, he cautions, should be embraced 'no farther then as Rational'; we should be 'wary' in asserting its 'absolute reality and truth' (1664: 488).

Though any truth must be clear of contradiction, we can know that something is the case without being able to explain it. ‘*Our Senses*’, for example, ‘do assure us of such things that no faculty can conceive how they are such as *our Senses* warrant them to be’ (1664: 454). This principle plays an important role in More’s defence of the possibility of spirit and the authority of revelation.

Although he admits that thinking is first occasioned by external objects, More thinks it is vital to distinguish between the ‘extrinsecall Occasions’ of our ideas and their ‘adequate or principal Causes’. There are ‘*Innate Notions and Ideas*’ in the mind; the soul is no *tabula rasa* or blank slate. Innate ideas do not ‘[flare] to the *Animadversive Faculty*, like so many *torches* or *Starres*’. There is, instead, ‘an active sagacity in the Soul, or quick recollection, as it were, whereby some small businesses being hinted unto her, she runs out presently into a more clear and large conception’ ([1653a] 1925: 14). Damaris Cudworth (later Masham) quoted these words of More in a letter to Locke. She followed More in suggesting that even though there are no actual representations in the mind innately, the mind has a native power to form conceptions that reach beyond the experience that prompts them. More himself illustrated this dispositional interpretation of innate ideas by comparing the soul to a sleeping musician who can finish a tune after a friend hums the opening notes.

Geometrical ideas are among those that cannot be explained by experience alone; no object of experience approaches the accuracy of our idea of a circle or triangle. Relative notions provide further instances. Cut off half of a pound of lead, More points out, and another pound, at first its equal, becomes its double, without any change in its corporeal makeup. Equal and unequal, he concludes, are not physical affections of matter, from which it follows that they cannot enter the mind through the corporeal organs of sense.

Besides single ideas, there are innate complex ideas (including whole truths or ‘common notions’). The idea of God is More’s leading example, and in the *Antidote*, its innateness plays what seems to be a crucial role in his primary argument for God’s existence. Yet in the *Appendix* added to the second edition of the *Antidote* in 1655, More claims that it is ‘all one’ to his argument whether the idea of God is innate or not. He does go on to insist, however, that the idea is not ‘a crafty Figment of *Politicians*’. It does seem important to More, then, as it had been to Descartes, that the idea of God should not be arbitrary or factitious. But More offers no clear test of factitiousness. To the objection that existence can be derived from the idea of a being which is absolutely evil, he is content to reply that we ‘can easily discover the difference betwixt such arbitrarious and forced Figments and fancies’ as this, and the ‘naturall and consistent *Ideas* of our own Mind’ ([1653a] 1662: 156).

3 The possibility and existence of spirit

The greatest obstacle to the belief in immortality, More believed, was ‘that confident opinion in some’ - he was thinking above all of Hobbes - that ‘the very notion of *a Spirit*’ is ‘a piece of Non-sense and perfect Incongruity in the conception thereof’ ([1659] 1925: 65). In the *Immortality*, More set out to free both spirit in general and the several kinds of spirit from ‘the imputation of either obscurity or inconsistency’ ([1659] 1925: 96).

More defined body as a substance impenetrable and *discerpible* (that is, separable into parts), and spirit as a substance penetrable and *indiscerpible* (not separable into parts). His main argument on behalf of spirit is that every term in its definition is as ‘intelligible and congruous’ as its counterpart in the definition of body. The ‘precise Notion’ of substance (elsewhere defined as any being subsisting by itself) appears in both. ‘And it is as easy to understand what *Penetrable* is as *Impenetrable*’, he writes, ‘and what *Indiscerpible* as *Discerpible*’ ([1659] 1925: 66). If the positive denominations were all on the side of body, it could be said that the intelligibility of spirit, being derived by negation, was second-rate. But each of More’s definitions, as he points out, is partly negative and partly positive.

More’s definition of body is in fact misleading because, according to the *Immortality*, the smallest bodies - atoms or ‘*minima corporalia*’ - are indiscerpible. As the young Isaac Newton wrote in his notebooks, ‘that matter may be so small as to be indiscerpible the excellent Dr More in his book of the soul’s immortality has proved beyond all controversy’ (Newton 1644-5: 341). More’s indiscerpible atoms resist separation into parts, as Newton realized, only because they are so tiny. Despite their small size they remain, in More’s view, ‘purely’ or ‘mathematically’ divisible; it is simply that no force in nature has the power to break them apart. ‘It is of the very Essence of *Matter* to be *divisible*’, More wrote, ‘but it is not at all included in the essence thereof to be *discerpible*’ ([1659] 1925: 77).

Perhaps the most perplexing feature of spirit as More understood it was *extension*: not the extension distinctive of matter, which involves a juxtaposition of parts, but 'a certain *Amplitude of Presence*' that enables the human soul, for example, to be 'at every part [of the body] at once' ([1653a] 1662: 172). Spirit, More reasoned, has the power to move matter, and it must be present wherever that power is exerted. But if the extension of spirit reaches as far as its power, why is spirit not discerpible? In reply More offers an analogy: a 'luminous orb' diffused in all directions from a single and inexhaustible point of light. The spreading light can of course be blocked or contracted by an opaque body, but according, for example, to the Aristotelians (whose hypothesis can be accepted for the sake of unfolding the analogy), it is impossible for the rays to be 'clipt off, or cut from this *lucid point*, and be kept apart by themselves' ([1659] 1925: 71). In a similar way, More proposes, a spirit's 'Centre of Life' gives rise to a 'Secondary Substance' spread around it. The inmost point is an 'emanative cause' of the surrounding extension, one that 'merely by Being, no other activity or causality interposed', produces an effect ([1659] 1925: 74). The effect must therefore be instantaneous. Because the centre of life is an *absolutely* indivisible point, it is, in the end, less perplexing than an atom of matter, whose indiscerpibility coexists uneasily with its (absolute) divisibility.

If spirit is extended, extension cannot be the mark of material substance, as Descartes (for example) proposed. And if God's power is unlimited, God must be extended throughout space. In a letter to More in April 1649, Descartes distinguished between the extension of a substance and the extension of its power. God, he granted, is everywhere, but only in virtue of his power. More replied that the power of God is a mode of God, and because a mode cannot occur apart from the substance it modifies, it is necessary that God is everywhere, a point he repeated years later in the *Enchiridion metaphysicum*.

More defended the intelligibility of both God and finite spirits. He defines a finite spirit, for example, as 'a substance Indiscerpible, that can move it self, that can penetrate, contract, and dilate it self, and can also penetrate, move, and alter the Matter'. More supported the various characteristics in this definition by comparing them to the equally perplexing characteristics of matter. Those who deny spirit, for example, are in no position to quarrel with spirit's ability to move itself, because they themselves take matter to be self-moving (since it obviously moves, and there is, on their view, nothing but matter to get things going in the first place). In the end, we may not understand how spirit moves matter and unites with it. But it is no easier to understand how matter moves matter and how it holds itself together.

More defines God as a being fully and absolutely perfect ([1653a, 1659] 1925: 10, 68). Because necessary existence is among the attributes entailed by full perfection, it follows that God exists. More supports this version of the ontological argument (derived from Descartes) with a second: any being we can contemplate, God included, is either contingent, impossible or necessary. But God is neither contingent (contingency being 'incompetible to an *Idea of a Being absolutely Perfect*') nor impossible (the idea of God 'being compiled of no Notions but such as are *possible* according to the Light of Nature') ([1653a] 1925: 22). Hence God exists, a conclusion confirmed by the natural remorse of conscience, our sense of hope and our feelings of religious veneration. More also employed versions of the cosmological argument and the argument from design: 'No *Matter* whatsoever of its own Nature has any active Principle of *Motion*' ([1659] 1925: 102). Motion is 'superadded' to matter by God, who then conserves it. A 'blind *impetus*' could never produce the admirable order we observe in nature and in ourselves.

There are four main categories of finite or created spirits: seminal forms, the souls of brutes, human souls, and the spirits actuating angels. A seminal form is 'a created Spirit organizing duly-prepared Matter into life and vegetation proper to this or the other kind of Plant' ([1659] 1925: 83). An animal soul is the source of both vegetation and sensation. More's argument for the existence of an incorporeal human soul turns on our awareness of faculties and operations 'utterly incompetible to *Matter* considered *at large* without any *particular* organization' and beyond the capacity of matter (or the 'tumbling of atoms') however organized. These faculties include sense, reason, and liberty, which More defines as the power, notwithstanding external assaults, to cleave to virtue. (Our liberty is known by '*Internal Sense*'. In both the *Immortality* and the *Enchiridion ethicum*, More argues at length against Hobbesian determinism.) The soul is also responsible for the frame of the human body and its vital powers. The human soul as More understands it is therefore not just a mind (as it was for Descartes), but an agent of bodily development and maintenance. More must therefore repudiate Descartes' main argument for dualism. He faults Descartes for inferring a 'real precision' (or real distinction) from a precision in thought. Our vital powers can be conceived apart from our power to reflect and reason, but it does not follow that we have two

souls.

More viewed apparitions (‘speakings, knockings, opening of doors when they were fast shut, sudden lights’ - [1659] 1925: 105) and the efficacy of witchcraft as further proofs of spirit. He contributed reports of apparitions to the 1681 edition of Joseph Glanvill’s *Saducismus Triumphatus*. ‘Those that lay out… well-attested Stories of Witches and Apparitions’, More wrote in a letter included in the volume, ‘*do real service to true Religion and sound Philosophy*’ (Glanvill 1681: 26). The phenomena reported, though not miraculous, cannot be accounted for mechanically (see Glanvill, J.).

4 More and Descartes

More’s initial enthusiasm for Descartes had moderated by the late 1650s, but even as late as 1662 he was able to describe the books in his *Collection* as combining Platonism and Cartesianism. The *Divine dialogues* marks a clear break. The book was published without More’s name, but his earlier books are cited in the dialogues themselves and in a prefatory letter from ‘the Publisher’ - a letter actually written by More. The letter takes note of the author’s ‘*over-plain and open opposing that so-much-admired*…Descartes’ (1668: [x]). Some earlier writers admit that there may be ‘*some few effects*’ in nature that are ‘*purely Mechanicall*’. (Here the publisher refers to the author of the *Epistola ad V.C.* (1662) - the More of the late 1650s.) But the present book, the publisher vows, shows abundantly that ‘*there is no purely-Mechanicall* Phenomenon *in the whole Universe*’ (1668: [xii]). In the ensuing dialogues, Descartes’ views are represented by the character Cupuphon, a ‘zealous, but Airie-minded, *Platonist* and *Cartesian*, or *Mechanist*’. Cartesian views are shown to be incompatible with scripture and serviceable to religion only in so far as they display (against their intention) the limits of mechanism. The publisher does credit Descartes with discovering the ‘most credible Material Causes’ of phenomena; More never abandoned the Cartesian view that matter is of one uniform or ‘homogeneal’ nature, and that change in body is, in itself, nothing but local motion. But the explanatory *causes* of bodily change - the causes that prompt and guide it - were another matter. More had many other quarrels with Descartes; in the fiercely critical *Enchiridion metaphysicum*, for example, he scornfully crowns Descartes prince of the ‘nullibilists’, a party of philosophers who endanger religion by asserting that spirit is nowhere (see Descartes, R. §8).

5 The spirit of nature and the nature of space

Changes in body that cannot be ‘resolved into mere Mechanical powers’ are, according to More, the work of the spirit of nature, the ‘Vicarious power of God upon… matter’ ([1659] 1925: 169; 1662: 13). It is an immaterial substance without sense or thought, pervading the ‘whole matter’ of the universe. It accounts for departures from the laws of mechanism but also, as a ‘natural Transcript’ of those laws (1662: xvi), for matter’s submission to them. Among the phenomena that argue powerfully for the spirit of nature are sympathetic pains, assuagements and cures. But even the falling of a stone or bullet is a sign of its influence because, given only the laws of mechanism, a stone or bullet should recede from the earth instead of falling towards it.

In the *Enchiridion metaphysicum* More presented some of Robert Boyle’s hydrostatic experiments as evidence of the spirit of nature. Boyle, in his *Hydrostatical discourse* (1672), defended a mechanistic interpretation of his results. There Boyle stopped with the claim that More’s ‘hylarchical’ principle is unnecessary, but in his *Inquiry into the vulgarly received notion of nature* (1686) he made broader criticisms (without naming More). Some of Boyle’s points were elaborated by Robert Hooke (1677), who asked, in effect, by what laws More’s spirit operates, since it is surely true, even on More’s view, that it does not simply do what it pleases (see Boyle, R.).

More’s primary argument in defence of immaterial substance in the *Enchiridion metaphysicum* was simply that there is an infinite, substantial space or ‘immobile extension’ really distinct from matter. He proceeded to assign to it many of the names or attributes traditionally reserved for God: infinite, incomprehensible, uncreated, absolutely independent. Infinite space is, he concluded, a ‘rough representation of the divine essence’. Other remarks strongly suggest that it *is* God, or at least an aspect of God. More did caution that the attributes on his list pertain not to God’s ‘life and operation’ but only to God’s presence. He promised to take up the connection between the divine life and the divine presence in a sequel to the *Enchiridion* (which was published as ‘Part One’), but a second part never appeared. The influence of More’s divinized space on the absolute space of Newton is still a matter of debate.

6 Religion

The main theme of More's writing on religion, aimed against atheists on one side and 'enthusiasts' on the other, was the 'Reasonableness' of Christianity. *Enthusiasmus triumphatus* (1656) was a satire on the vice of enthusiasm, which More defined as the 'misconceit of being inspired', or 'moved in an extraordinary manner by the power or spirit of God to act, speak, or think what is holy, just, and true' (1656: 2). Enthusiasm is caused, More argued, by an imagination overpowered by melancholy; its remedy consists in temperance, humility and reason.

The *Mystery of godliness* (1660) was More's major work in defence of the Christian religion. 'A *Mystery*', he explains in the Preface, 'is a piece of Divine knowledge *measurably Abstruse*, whereby it becomes *more Venerable*, but yet *Intelligible* that it may be *Communicable*, and *True* and *Certain* that it may win*firm Assent*, and lastly *very Useful* and *Effectual* for the *perfecting* of the Souls of men, and restoring them to that *Happiness* which they anciently had faln from' (1660: 2). His aim is to represent the fundamentals of Christianity - the faith, plainly propounded in scripture, which is all that is required for salvation. He therefore says little about forms of worship and nothing about church government, but he ends the book with a plea for liberty of conscience. Toleration should be extended to all who believe in God and in punishment and reward in a life to come, provided their beliefs do not call for violations of moral law or disloyalty to their rightful rulers.

See also: Cambridge Platonism

KENNETH P. WINKLER

List of Works

More, H. (1925) *Philosophical writings of Henry More*, ed. F. Mackinnon, The Wellesley Semi-Centennial Series, New York: Oxford University Press; repr. New York: AMS Press, 1969.(Remains the best anthology of More's philosophical writings. Selections from the *Antidote* and the *Immortality*, together with the final chapters of the *Enchiridion Metaphysicum* as they appeared in the *Saducismus Triumphatus*. Mackinnon's 'Outline Summary of More's Philosophical Theory' on pages 159-71 will be especially useful to anyone reading More for the first time.)

More, H. (1642) *Psychodia platonica; or, a platonicall song of the soul, consisting of foure severall poems; viz., Psychozoia, Psychathanasia, Antipsychopannychia, Antimonopsychia*, Cambridge: R. Daniel.(Verse-arguments for the soul's immortality and pre-existence, its consciousness after death, and its distinctness from other souls.)

More, H. (1646) *Democritus platonissans; or, an essay upon the infinity of worlds out of Platonick principles*, Cambridge; repr. Los Angeles, CA: William Andrews Clark Memory Library, University of California, 1968.(Philosophical poem defending the infinity of worlds.)

More, H. (1653a) *An antidote against atheisme, or an appeal to the natural faculties of the mind of man, whether there be not a God*, London; 2nd edn, with appendix, London, 1655, repr. Bristol: Thoemmes Press, 1997.(Arguments, a priori and empirical, for '*a Being absolutely Perfect*'.)

More, H. (1653b) *Conjectura cabbalistica*, London; repr. Bristol: Thoemmes Press, 1997.(An attempt to lay bare the hidden meanings - scientific, metaphysical, and moral - in the first three chapters of *Genesis*.)

More, H. (1656) *Enthusiasmus triumphatus: or, a discourse of the nature, causes, kinds, and cure, of enthusiasme; written by Philophilus Parresiastes*, London and Cambridge; repr. Los Angeles, CA: Clark Memorial Library, University of California, 1996; repr. Bristol: Thoemmes Press, 1997.(The cause and cure of enthusiasm, or 'misconceit of being *inspired*', in religion and philosophy.)

More, H. (1659) *The immortality of the soul, so farre forth as it is demonstrable from the knowledge of nature and the light of reason*, London and Cambridge; repr. Bristol: Thoemmes Press, 1997; modern edn, ed. A. Jacob, *Henry More. The immortality of the soul*, Dordrecht: Martinus Nijhoff, 1987.(Defends the possibility of spirit and the existence and immortality of an immaterial human soul.)

More, H. (1660) *An explanation of the grand mystery of godliness; or, a true and faithful representation of the everlasting gospel of our Lord and Saviour Jesus Christ*, London and Cambridge; repr. Bristol: Thoemmes Press, 1997.(More's main defence of the Christian religion.)

More, H. (1662) *A collection of several philosophical writings of Dr. Henry More*, London and Cambridge; repr. New York: Garland, 1978.(Contains the *Antidote* and its *Appendix*, *Enthusiasmus triumphatus*, the *Immortality* (with some additions), More's correspondence with Descartes (introduced by an exchange of letters between More and Claude Clerselier), and the *Epistola ad V.C.*, an appraisal of Cartesianism written around 1658 and first published here.)

More, H. (1664) *A modest enquiry into the mystery of iniquity, the first part, containing a careful and impartial delineation of the true idea of antichristianism*, London and Cambridge.(An attack on the idolatry of Catholics.)

More, H. (1667) *Enchiridion ethicum (Handbook of ethics)*, London and Cambridge; trans. as *An account of virtue: or, Dr. Henry More's abridgement of morals, put into English*, London, 1690: repr. New York: The Facsimile Text Society, 1930; repr. Bristol: Thoemmes Press, 1997.(More's only book devoted to ethics. He defines virtue as an intellectual power by which the soul overrules '*Right Reason*', an innate copy of an eternal and immutable morality. The soul is also equipped with a '*Boniform Faculty*', enabling it to 'relish' what is best.)

More, H. (1668) *Divine dialogues…. Collected and compiled by the care and industry of Franciscus Palaeopolitanus*, London, 2 vols.(Volume 1 contains More's most accessible presentation of his proofs of God's existence. Volume 2, which is more theological, contains *A brief discourse of the true grounds of the certainty of faith in points of religion*, where the certainty of faith is said to presuppose certainty of sense and certainty of reason.)

More, H. (1671) *Enchiridion metaphysicum;…. Pars prima: de existentia & natura rerum incorporearum in genere (Handbook of metaphysics;…. Part one: concerning the existence and nature of incorporeal things in general)*, London and Cambridge; repr. Bristol: Thoemmes Press, 1997; modern translation, ed. A. Jacob, *Henry More's Manual of Metaphysics. A Translation of the Enchiridium Metaphysicum with an introduction and notes*, Hildesheim: Georg Olms, 1995, 2 vols.(Latin summary of More's metaphysics.)

More, H. (1675-9) *Opera omnia (Complete works)*, London, 3 vols; repr. vols 2 and 3, Hildesheim: Georg Olms, 1966.(Volume 1 is theological, volumes 2 and 3 philosophical. More's autobiography appears in the *Praefatio generalissima*, volume 2.)

More, H. (1930) *The Conway Letters: The Correspondence of Anne, Viscountess Conway, Henry More, and their Friends, 1642-1684*, ed. M.H. Nicolson, revised edn, ed. S. Hutton, Oxford: Clarendon Press, 1992.(Contains More's correspondence with Conway, with connecting commentary by Nicolson. Hutton adds several letters of philosophical importance, as well as an introduction that is, in part, a brief survey of recent work on More.)

References and further reading

Boyle, R. (1772) *The Works of the Honourable Robert Boyle*, London: J. & F. Rivington, 6 vols.(*An hydrostatical discourse, occasioned by the objections of the learned Dr. Henry More* appears in volume 3, pages 596-628. *A free inquiry into the vulgarly received notion of nature* appears in volume 5, pages 158-254, where the spirit of nature seems to be under attack on, for example, page 191.)

Burtt, E.A. (1954) *The Metaphysical Foundations of Modern Physical Science*, Garden City, NY: Doubleday. (Chapter 5 contains a highly readable account of More's views on the extension of spirit, the spirit of nature, and absolute space. More's influence on Newton is discussed in chapter 7.)

Colie, R. (1957) *Light and Enlightenment*, Cambridge: Cambridge University Press. (More's relation to Spinoza and Spinozism is discussed in chapters 5 and 6.)

Copenhaver, B.P. (1980) 'Jewish Theologies of Space in the Scientific Revolution: Henry More, Joseph Raphson, Isaac Newton and their Predecessors', *Annals of Science* 37: 489-548.(A richly documented account of More's sources and his influence.)

Coudert, A. (1975) 'A Cambridge Platonist's Cabalist Nightmare', *Journal of the History of Ideas* 36: 633-52.(On More and Conway, their friend Francis Mercury van Helmont, and the cabala. More's eventual disenchantment with cabalism is discussed on pages 645-52.)

Descartes, R. (1991) *The Philosophical Writings of Descartes*, vol. 3, ed. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press.(Contains English translations from Descartes' letters to More.)

Glanvill, J. (1681) *Saducismus triumphatus: or, full and plain evidence concerning witches and apparitions*, 2nd edn.(Contains several contributions by More, including the two final chapters of *Enchiridion metaphysicum*, translated into English.)

Gabbey, A. (1982) *'Philosophia Cartesiana Triumphata: Henry More (1646-71)'*, in T.M. Lennon, J.M. Nicholas and J.W. Davis (eds) *Problems of Cartesianism*, Kingston and Montreal, Que.: McGill-Queen's University Press, 171-250.(The leading account of More's engagement with Descartes.)

Greene, R.A. (1962) 'Henry More and Robert Boyle on the Spirit of Nature', *Journal of the History of Ideas* 23: 451-74.(Traces the rise and fall of the experimental defence of the spirit of nature.)

Hall, A.R. (1990) *Henry More: Magic, Religion and Experiment*, Oxford: Blackwell. (The most thorough study of More's relation to the science of his day.)

Hooke, R. (1677) *Lampas: or, descriptions of the some mechanical improvements of lamps and waterpoises*, in R.T. Gunther (ed.) *Early Science in Oxford*, vol. 8, Oxford: R.T. Gunther, 1931.(Hooke criticizes More's *Enchiridion Metaphysicum* on pages 182-96.)

Hutton, S. (ed.) (1990) *Henry More: Tercentenary Studies*, Dordrecht: Kluwer.(The single most useful secondary source on More. Among the most valuable contributions are Crocker's 'Henry More: A Biographical Essay' and 'A Bibliography of Henry More', Gabbey's 'Henry More and the Limits of Mechanism', Henry's 'Henry More versus Robert Boyle: The Spirit of Nature and the Nature of Providence', and Coudert's 'Henry More and Witchcraft'.)

Koyré, A. (1957) *From the Closed World to the Infinite Universe*, Baltimore, MD: Johns Hopkins University Press.(An especially influential study. More is discussed in chapters 5 and 6; his relation to Newton is touched on in chapters 7 and 8.)

Lichtenstein, A. (1962) *Henry More: The Rational Theology of a Cambridge Platonist*, Cambridge: Harvard University Press.(A study of tensions in More's theology.)

Newton, I. (1664-5) *Certain Philosophical Questions: Newton's Trinity Notebook*, ed. J.E. McGuire and M. Tamny, Cambridge: Cambridge University Press, 1983. (Newton refers to More on pages 341 and 393.)

Norris, J. (1688) *The theory and regulation of love*, Oxford.(Includes *Letters philosophical and moral between the author and Dr. Henry More*.)

Ward, R. (1710) *The life of the learned and pious Dr. Henry More*, London, 1710; modern edn ed. M.F. Howard, London: Theosophical Society, 1911; repr. Bristol: Thoemmes Press, 1997.(Includes long passages, translated from the original Latin into English, from More's account of his early life in the general preface to his *Opera Omnia*.)

Moscow-Tartu School

*The Moscow-Tartu School of semiotics (theory of signs) was formed when a diverse group of scholars joined informally from the 1950s to 1980s to provide alternatives to the regnant Soviet approaches to language, literature and culture. Their work develops the linguistics of Saussure, elaborated by Trubetzkoi and Hjelmslev, with its central notions of sign as union of signifier and signified, its distinction between language as system (*langue*) and language as utterance (*parole*), and its analysis in terms of the significant differences between paired equivalent elements in a system (that is, meaning is a matter not of individual elements, but of the relationship between comparable elements). In its early stages members of the Moscow-Tartu School did intricate analyses of lyric poetry and of highly conventional prose works (such as detective stories) using statistical and linguistic methods. They subsequently came to treat art works and other cultural artefacts as the products of 'secondary modelling systems', that is, as elements arranged according to rules that could be seen as language-like and hence accessible to analysis by the procedures of structuralist linguistics. The group shared an interest in Western and pre-Stalinist Russian literary theory - especially in the Russian formalists - and in contemporary linguistics, semiotics and cybernetics. In a time of pervasive intellectual stagnation this loose confederation sought to formulate objective and exact methods for literary scholarship, to republish works of Russian theory that had been repressed from the 1930s to 1950s, and to bring scholarship in the humanities into line with developments in other scholarly fields. During the 1970s prominent members of the group, such as Iu.M. Lotman and B.A. Uspenskii, turned from more theoretical and formalized work to historical studies of culture as a system of semiotic systems.*

1 History: the Soviet context and the rise of semiotics

It is difficult to comprehend the extent to which the policies of the Stalin era hindered the unfolding of academic life. A vigilant censorship, the prescriptions of party congresses, and academic hierarchies established not by intellectual distinction but by survival skills, all combined to discourage critical thinking. Literary scholarship, which had rejected the formalists' attempts to examine literature as a process with its own dynamics and evolution (see Russian literary Formalism), became particularly vulnerable to the demands of the political apparatus. By an irony of cultural history Soviet literary criticism had followed the pattern of innovation and mechanization outlined by the formalist theory of literary evolution: the innovatory and socially committed literary criticism of the nineteenth-century Russian intelligentsia, with its utilitarian view of literature and its understanding of art as a reflection of reality, had become institutionalized and stultifying, a barrier to re-examining the considerable role of literature in Russian society. Yet some disciplines, because of their remoteness, abstractness (linguistics) or technological promise (cybernetics, machine translation) were relatively exempt from the blatant ideologization of academic life. It was to these disciplines that a number of young linguists and 'philologists' (scholars trained in literature, historical linguistics and the history of ideas) turned for support in renewing the humanities. Their fascination with the 'exact methods' and 'objective scholarship' that these disciplines offered ultimately must be understood in the context of the impasse that scholarship had reached.

One of the group's most erudite members, the linguist V.V. Ivanov, would in 1976 trace the rise of Soviet semiotics to the research and theories of Russian and foreign linguists and anthropologists. He specifically identified the psychologist Vygotskii, the cinematographer Eisenstein, the information theorists A.N. Kolmogorov and Claude Shannon (see Information theory), the phenomenologist G.G. Shpet, the philologist M.M. Bakhtin and his colleague V.N. Voloshinov, for their work in reconstructing ancient cultures, in developing theories of the sign and in studying the structures and levels of art. The group's acquisition of these interests was achieved through conferences, informal gatherings and harsh debates between 'physicists' and 'lyricists' in scholarly and popular periodicals.

During the 1950s structuralist analysis slowly began to develop in the Soviet Union, enabled in part by Stalin's 1950 articles on linguistics, which in effect freed the discipline from Marxism's base-superstructure model, licensing the study of grammar and internal laws of language development. Two conferences brought the movement into focus and lent it its particular Soviet concern with cybernetics and information theory: a conference in Gorky (1961) 'On the Application of Mathematical Methods to the Study of the Languages of Artistic Literature' and a symposium in Moscow (1962) 'On the Structural Study of Sign Systems'. The first conference centred around the interest of Academician Kolmogorov in poetry's special potential for conveying information,

but already the papers showed the group’s wide range of interests and approaches: I.I. Revzin used Noam Chomsky’s *Syntactic Structures* (1957) to speculate on the generative structures of literary texts, Ivanov surveyed the achievements of Western structuralism, and A.K. Zholkovskii constructed a genealogy from the works of the Russian Formalists and Sergei Eisenstein. He and Iu.K. Shcheglov would subsequently develop their own ‘generative poetics’, a pragmatic one which focused on the effects generated by the poetic text. The Moscow Symposium, organized by Ivanov, included many of the same participants, but had a more semiotic and less cybernetic focus. Danish semiotic theory (‘glossematics’) (see Structuralism in linguistics §3) offered the promise of interaction between mathematics and semiotics and the possibility of joining the human sciences in a common and exact methodology. A group of papers on art as a semiotic system served as a beginning of the movement beyond verbal texts; papers on etiquette, games, and fortune-telling began the group’s investigations of behaviour and popular culture.

The Tartu presence in the Moscow-Tartu School began with Iu.M. Lotman’s 1963 essay ‘On the Delimitation of the Concept of Structure in Linguistics and Literary Scholarship’ and his *Lektsii po struktural’noi poėtike (Lectures on Structural Poetics)*, which appeared in the first issue of an important new series, *Trudy po znakovym sistemam (Works on Sign Systems)* (1964). As Uspenskii notes in his account of the School’s beginnings, Tartu University brought a different scholarly culture to the movement: more oriented towards canonical literature and the history of ideas, less concerned with linguistics and with popular genres. Tartu University also brought the legacy (however constrained by Soviet reality) of having been the Russian Empire’s most European university, with a tradition of academic autonomy that the other Russian universities had not always enjoyed.

The high point of the Moscow-Tartu School’s activity was its five summer schools (1964-74), held at Kääriku in rural Estonia and in Tartu itself. Described by their participants as ‘carnivalesque’, ‘utopian’ and ‘hermetically sealed’, they centred around lively discussions rather than formal presentations, and their proceedings give brief lists of theses rather than lengthy formal papers. The summer schools brought together the Moscow semioticians and the philology faculty from Tartu, occasional guests (such as Jakobson in 1966), and a number of younger scholars who would continue to develop the school’s themes and concepts after it ceased to exist.

The publications of the Moscow-Tartu School, *Trudy po znakovym sistemam* and other periodicals from Tartu, appeared in shoddy editions whose print runs could not come close to meeting the domestic and international demand for them. Their opponents in Soviet academic life published in much more widely circulated publications and levelled a consistent series of charges against the semioticians, as Peter Seyffert has noted: that semiotics was mere fashion, formalism couched in incomprehensible shibboleths; that its claim to universal applicability could not be sustained; and that ‘mathematical methods’ and ‘exact science’ were promises often unrealized and unrealizable. Certainly the School’s focus on the sign and on language opposed the entrenched Russian view of art as ‘thinking in images’, and certainly its view of culture as a semiotic mechanism rather than a reflection of the means of production, challenged the dogma of Soviet Marxism.

For a variety of internal and external reasons, the Moscow-Tartu School had ceased to function by the mid-1970s. Growing government opposition to intellectual unorthodoxy met with opposition from the Tartu faculty, especially over the invasion of Czechoslovakia in 1968. A number of members of the School emigrated from the Soviet Union and made successful careers in the West. The growing interest of members of the School in questions of culture, which began with studies of typology and increasingly branched out into highly specific research with little attention to theory, led to methodological diffuseness. Nevertheless, the publications of the Moscow-Tartu School continued to appear in periodicals and, increasingly, in book form. By the 1990s Lotman was starring in a highly regarded television series on Russian culture. By another irony of cultural history the Moscow-Tartu School, condemned for neglecting traditional Russian views on literature and society, had become the principal interpreter of both, and of their interrelations for the new Russian Republic. After Lotman died in 1993, this canonization was completed by generous obituaries in the Russian press and by the republication of his semiotic studies.

2 Doctrine: general topics and the work of Uspenskii and Lotman

Because so many of the School’s works were experimental and because their work was so far-ranging, it can be difficult to generalize about their activity. During the two decades of their most intense activity the School tried out many variants of semiotic analysis, moving from an early interest in mathematical modelling, cybernetics and

information theory to work more akin to cultural history and cultural anthropology, two disciplines which were virtually non-existent in the Soviet Union. Nevertheless, a common feature was their use of linguistic models. This was also the case with American, Danish, French and Italian structuralism, but the use of linguistic models stands out even more sharply in the Soviet context because the regnant approach to literature focused on ‘images’, not ‘signs’, and on social determinism rather than on self-regulating sign systems. Against this background the Moscow-Tartu School could be seen as a movement, even though its members were drawing upon a number of different linguistic and semiotic theories: C.S. Peirce’s theory of the sign, Saussure’s understanding of semiotic systems, Danish glossematics, Trubetzkoi’s phonemics, Jakobson’s poetry of grammar and Chomsky’s transformational grammar, to name the most important (see Chomsky, N. §1; Peirce, C.S. §8; Saussure, F. de §2; Structuralism in linguistics §§2-3; Syntax §3). Yet members of the School tended not to pursue linguistic models as exhaustive sets of discovery procedures, as did Jakobson in his studies in the ‘poetry of grammar’: they moved from the mechanisms of meaning-production towards interpretation and beyond the study of closed systems towards contextual studies. As the group moved beyond language to other semiotic systems - ‘secondary modelling systems’, as Uspenskii termed them - the ties to linguistic theory could become tenuous. Nevertheless, even in these later essays one detects an enduring fascination with language, language use (pragmatics), and analytic operations learned from formal linguistics (such as analysis in terms of binary opposition). The problems that the group addressed tended to follow logically from this fundamental appeal to linguistic models: how does poetry convey information with its formal resources, how does poetry use the resources of language to model reality and the artist’s worldview, how may other cultural spheres (including social behaviour) can be analysed using the model of linguistic analysis, what is the text that is available for semiotic analysis, how do different cultures in different historical periods use signs differently, what is the impact of such secondary modelling systems (such as literature, art, or the theatre) upon human cognition and behaviour? The logic of their methodology did not, however, lead them into exploring deeply such subjects as ‘the death of the author’ (who would become a ‘guest of the text’ for such French semioticians as Roland Barthes) or the theory of tropes. Unlike Western structuralists and post-structuralists, the Moscow-Tartu group tended to view the text as a unity, however multilayered and however rich its resources for generating interpretations. Language retained for them its ability to represent reality, complexities of coding notwithstanding; and the human subject was inevitably seen as the constitutor, or at least the user, of codes, more than constituted by them. In this the Moscow-Tartu semioticians, especially those who worked on such canonical figures as Pushkin and Gogol’, contributed a semiotic version to the traditional Russian heroic portrayal of authors and to the traditional Russian veneration of high cultural texts.

Members of the School did innovatory work on many subjects: on various aspects of semiotic theory and on text theory, on mythology and the reconstruction of ancient symbol systems (Ivanov, V.N. Toporov, E.M. Meletinskii, D.M. Segal), on musical semiotics (B.M. Gasparov) and on lyric poetry (Zholkovskii, Revzin, Ivanov). Borrowing a favourite technique from the School’s analytic procedures, description by analysis of pairs of equivalent elements in a system, one may treat the group’s concerns by focusing on the work of two leading members, one from each centre: Iu.M. Lotman, by training as a literary scholar and intellectual historian (he had studied in Leningrad with surviving formalists) and B.A. Uspenskii, by training as a linguist and Slavic philologist (he had studied briefly in Denmark). Lotman’s primary research area was Russian literature and culture of the 1780s to 1830s; Uspenskii’s was the Russian language of the late medieval period and of the eighteenth century, although each ranged far afield in search of telling instances and challenging sign systems. By the late 1960s, however, each had moved towards the interests of the other, towards the investigations of ideology and point of view that marked Lotman’s early work, towards studies of multiple linguistic codes that had interested Uspenskii, and some of their best work was collaborative.

Lotman’s changing interests may be seen in four studies of Pushkin’s novel in verse, *Evgenii Onegin*. The first (1960) analysed the construction of character in the novel; this piece is richly contextual, relating character development to contemporary trends in Russian thought, but treating the novel as an ‘organic whole’. In 1966 he took a radically different approach in a paper on the artistic structure of *Evgenii Onegin*: here the operative concepts are modelling, a hierarchy of relationships, multiple points of view, system and anti-system, relationships, binary oppositions, and the inevitability of multiple interpretations given the text’s multiplication of segments through the organization of its complex stanzas. Interpretation and the reader’s role in the literary process become prominent, although they are governed by the internal relationships of the text. Lotman published a separate monograph on the novel in 1975. Here he achieves a synthesis of his work in intellectual history, his

semiotic analysis of paired opposition, and his Bakhtin-inspired study of multiple points of view as embodied in multiple discourses. He continues to study relations between the literary text and what lies outside it and the text's potential impact on a reader. In 1980 Lotman's commentaries to the novel treated culture in an anthropological sense, as the rituals, customs and conventions of Pushkin's Russia. But analysis by binary opposition - gentry/folk, Russian/European - remains an operative principle.

Uspenskii's work follows a similar trajectory, moving from the analysis of point of view towards studies of Russian culture in a broad, anthropological sense. His *O semiotike ikony (Semiotics of the Russian Icon)* (1971) subjects the 'language' of icons to linguistic analysis (syntactic, semantic, pragmatic), paying particular attention to the 'internal perspective' of medieval icons, according to which painter and viewer establish their point of view within the painting, not outside it. *Poetika kompozitsii (A Poetics of Composition)* (1970) had treated point of view more generally, and, although terms are not always defined with desirable precision, the book suggests an innovatory alternative to plot-based narratology, namely, the analysis with respect to four levels of point of view (ideological, phraseological, spatial and temporal, psychological).

For Lotman, as for Uspenskii, the text is both a sign (in the secondary modelling system) and a sequence of signs in verbal language. Both continued to focus their attention on problems of communication and point of view. Lotman's *Struktura khudozhestvennogo teksta (Structure of the Artistic Text)* (1970) and *Analyz poeticheskogo teksta (Analysis of the Poetic Text)* (1972) share some of the same problems of definition and focus that appeared in Uspenskii's *A Poetics of Composition*. But Lotman, the literary scholar, pays more attention to linguistic analysis and to information theory (the latter generally used metaphorically) than the linguist Uspenskii. While problems of integration in moving from level to level and in moving from analysis to interpretation are not solved theoretically, these books offer many suggestive ideas and fragments of literary analysis. Lotman's treatment of 'event' in the first book as 'the shifting of a character across the borders of a semantic field', as a transgression, or as the violation of an expectation, has been widely influential in studies of narrative.

From these large synthetic works on artistic texts, Lotman and Uspenskii moved to speculative discussions on an even higher level of synthesis, the typology of culture. In their 1971 essay, 'On the Semiotic Mechanism of Culture', culture becomes a system of sign production marked off from non-culture, a system of constraints and prescriptions. Here they differentiate cultures by modes of sign production and, predictably, Lotman and Uspenskii identify two opposed types of culture: cultures oriented towards expression (which view themselves in terms of an aggregate of texts) and cultures oriented towards content (which see themselves as a system of rules). In a more specific and controversial essay, 'Binary Models in the Dynamics of Russian Culture' (1977), they treat Russian culture as marked by a binary opposition between the 'sacred' and 'profane', an opposition which dictates diametrically opposed modes of behaviour. The same terms, for instance new/old, can in different periods occupy different positions in this hierarchy. Lotman and Uspenskii oppose this binary structure to a European ternary one, in which a neutral middle sphere allows for new systems to develop gradually, not catastrophically.

Subsequent research by Lotman and Uspenskii in the concrete phenomena in Russian cultural history address topics of moral and political significance: lying, theatricality, obscenity, religious dissent, pretenders to the throne and the use of multiple languages. Lotman's penultimate book, *Kul'tura i vzryv (Culture and Explosion)* (1992), published just after the fall of the Soviet Union, returns to speculative semiotic-based historiography to re-examine the relationship of culture to what is outside it and to explore Russia's new possibility of entering the 'European ternary system', with its concomitant development by gradual change as opposed to change by violent overthrow. Lotman's last book, his *Besedy o russkoi kul'ture (Conversations About Russian Culture)* (1994), returns to the turn of the nineteenth century to give a thick description of the daily life and traditions of the Russian gentry - its rituals, fashions and social patterns and the meaning that these had for those who practised them. The principled theoretical argumentation of these books, and their rigorous wide-ranging research and openness to experiment, encompass the main achievement of the Moscow-Tartu School.

See also: Semiotics; Structuralism in linguistics

WILLIAM MILLS TODD III

List of works

Baran, H. (ed.) (1976) *Semiotics and Structuralism: Readings from the Soviet Union*, White Plains: IASP.

(Well-introduced selection of papers on culture, mythology and literature.)
Lotman, Iu.M. (1960) ‘K èvoliutsii postroeniia kharakterov v romane ‘‘Evgenii Onegin’’’ (On the Evolution of the Construction of Characters in the Novel *Eugene Onegin*), in *Pushkin: Issledovaniia i materialy*, vol. 3, Moscow-Leningrad: Nauka, 131-73.(An excellent example of Lotman’s early work in intellectual history.)
Lotman, Iu.M. (1964) *Lektsii po struktural’noi poėtike (Lectures on Structural Poetics)*, in *Trudy po znakovym sistemam (Works on Sign Systems)*, vol. 1, Tartu: Uchënye zapiski Tartuskogo Gos. Universiteta, No. 160. (Introduction to the structure of verse.)
Lotman, Iu.M. (1966) ‘Khudozhestvennaia struktura ‘‘Evgenii Onegin’’’ (The Artistic Structure of ‘Eugene Onegin’) in *Uchënye zapiski Tartuskogo Gos. Universiteta* 184: 5-32.(An important early example of Lotman’s structuralist analysis.)
Lotman, Iu.M. (1970) *Struktura khudozhestvennogo teksta*, Moscow; trans. G. Lenhoff and R. Vroon, *The Structure of the Artistic Text*, Ann Arbor, MI: Michigan Slavic Contributions, 1977.(The most complete discussion of structuralist poetics, together with topics relevant to prose, such as plot and character; difficult to read.)
Lotman, Iu.M. (1972) *Analyz poeticheskogo teksta*, Leningrad; trans. D.B. Johnson, *Analysis of the Poetic Text*, Ann Arbor, MI: Ardis, 1976.(Further development of the theory of verse together with close analyses of a dozen lyric poems.)
Lotman, Iu.M. (1973) *Semiotika kino i problemy kinoėstetiki*, trans. M.E. Suino, *Semiotics of Cinema*, Ann Arbor, MI: Michigan Slavic Contributions, 1976.(A brief, accessible outline of selected topics in a grammar of cinematography.)
Lotman, Iu.M. (1975) *Roman v stikhakh Pushkina ‘Evgenii Onegin’ (Pushkin’s Novel in Verse ‘Eugene Onegin’)*, Tartu.(An example of Lotman’s structuralist analysis, more extensive and less technical than the 1966 article on *Eugene Onegin*.)
Lotman, Iu.M. (1980) *Roman A.S. Pushkina ‘Evgenii Onegin’: Kommentarii (A.S. Pushkin’s Novel ‘Eugene Onegin’: Commentaries)*, Leningrad.(Semiotic analysis, tending towards cultural anthropology; very accessible.)
Lotman, Iu.M. (1990) *Universe of the Mind: A Semiotic Theory of Culture*, trans. A. Shukman, Bloomington, IN: Indiana University Press.(A well-integrated collection of papers on text, rhetoric, space and culture.)
Lotman, Iu.M. (1992) *Kul’tura i vzryv (Culture and Explosion)*, Moscow: Gnosis.(A semiotic approach to culture and historical transformation.)
Lotman, Iu.M. (1992-3) *Izbrannye stat’i (Selected Articles)*, Tallinn: Aleksandra, 3 vols.(Prepared by Lotman himself, includes essays on semiotics, cultural history and literature.)
Lotman, Iu.M. (1994) *Besedy o russkoi kul’ture: Byt i traditsii russkogo dvorianstva (XVIII-nachalo XIX veka) (Conversations About Russian Culture: Everday life and traditions of the Russian gentry (18th to early 19th century))*, St Petersburg: Iskusstvo.(Popular, easily accessible discussions of such topics as duels, cardplaying, balls and gender relations.)
Lucid, D.P. (1977) *Soviet Semiotics: An Anthology*, Baltimore, MD: Johns Hopkins University Press.(Papers on general concepts, modelling systems, communication, art and culture.)
Matejka, L. *et al.*, (eds) (1977) *Readings in Soviet Semiotics (Russian Texts)*, Ann Arbor, MI: Michigan Slavic Publications.(Excellent selection, especially of the School’s early papers.)
Nakhimovsky, A.D. and A.S. (eds) (1985) *The Semiotics of Russian Cultural History: Essays by Iurii M. Lotman, Lidiia Ia. Ginzburg, Boris A. Uspenskii*, Ithaca, NY: Cornell University Press.(Well-chosen papers on cultural typology and history.)
Paperno, I. (1988) *Chernyshevsky and the Age of Realism: A Study in the Semiotics of Behavior*, Stanford, CA: Stanford University Press.(Highly accessible and intelligent development of the School’s ideas on text, behaviour, modelling and culture.)
Shukman, A. (ed.) (1978) *Soviet Semiotics and Criticism: An Anthology*, special issue of *New Literary History* 9 (3).(Includes ‘On the Semiotic Mechanism of Culture’ and well-chosen studies in the semiotics of literature.)
Shukman, A. (ed.) (1984) *The Semiotics of Russian Culture*, Ann Arbor, MI: Michigan Slavic Contributions. (Comprehensive set of papers on topics in cultural history.)
Tsiv’ian, T.V. (1987) *Issledovaniia po strukture teksta (Studies in the Structure of the Text)*, Moscow: Nauka. (Important papers by V.V. Ivanov and others on the theory, linguistics and analysis of texts.)
Uspenskii, B.A. (1970) *Poetika kompozitsii*, Moscow; trans. V. Zavarin and S. Wittig, *A Poetics of Composition*, Berkeley, CA: University of California, 1973.(An original approach to artworks in terms of point of view.)
Uspenskii, B.A. (1971) *O semiotike ikony (The Semiotics of the Russian Icon)*, Tartu; ed. S. Rudy, Lisse: Peter de Ridder, 1976.(An important application of semiotics to a non-verbal art form.)

Uspenskii, B.A. (1994a) *Izbrannye trudy (Selected Works)*, Moscow: Gnosis, 2 vols.(Selected papers on semiotics and cultural history.)

Uspenskii, B.A. (1994b) *Kratkii ocherk istorii russkogo literaturnogo iazyka (XI-XIX vv.) (A Brief Outline of the History of the Russian Literary Language (eleventh-nineteenth centuries))*, Moscow: Gnosis.(Important, controversial papers on diglossia and bilingualism.)

References and further reading

Gasparov, B. (1985) ‘Introduction’, in A.D. and A.S. Nakhimovsky (eds) *The Semiotics of Russian Cultural History: Essays by Iurii M. Lotman, Lidiia Ia. Ginzburg, Boris A. Uspenskii*, Ithaca, NY: Cornell University Press.(An elegant, insightful account of the School’s history and work on culture.)

Ivanov, V.V. (1976) *Ocherki po istorii semiotiki v SSSR (Outlines of the History of Semiotics in the USSR)*, Moscow: Nauka.(An important attempt to draw a genealogy for the School.)

Koshelev, A.D. (1994) *Iu.M. Lotman i Tartusko-moskovskaia semioticheskaia shkola (Iu.M. Lotman and the Tartu-Moscow School of Semiotics)*, Moscow: Gnosis.(Includes memoirs by members of the School, and a bibliography of Tartu works on semiotics, 1964-92.)

Seyffert, P. (1985) *Soviet Literary Structuralism. Background. Debate. Issues*, Columbus, OH: Slavica.(An invaluable, exhaustive discussion of the Soviet debates about semiotics up to 1971.)

Shukman, A. (1977) *Literature and Semiotics: A Study of the Writings of Yu.M. Lotman*, Amsterdam: North Holland. (The most comprehensive and intelligent survey of Lotman’s work, together with useful lists of papers from the Gorky conference, the 1962 symposium, and the first four summer schools.)

Motoori Norinaga (1730-1801)

*Motoori Norinaga was a pivotal figure in Japan's 'Native Studies' or 'National Learning' (*kokugaku*) movement. An accomplished philologist, he helped decipher the idiosyncratic eighth-century orthography of the Japanese chronicle of history and myth, the Kojiki (Records of Ancient Matters). This was part of his broader scholarly project of defining the nature of the ancient Japanese sensitivity or 'heart-and-mind' (*kokoro*). In so doing, he articulated an influential religious philosophy of Shintōand an axiology of traditional Japanese values, which he considered as primarily emotivist and aesthetic.*

The eldest son in a merchant family, Motoori relinquished his business interests to concentrate on scholarly pursuits. In order to support himself he also practised traditional, Chinese-based medicine. Motoori's driving passion, however, was the study of ancient Japanese culture. Although his times were dominated by a fascination with Chinese thought, Motoori gravitated towards the 'Native Studies' movement and its appreciation of, and romanticized return to, the values of pre-Sinified Japanese culture (see Japanese philosophy §7). In this regard he felt himself to be the disciple of Kamo no Mabuchi (1697-1769), even though he had only once met the latter in person.

Motoori was also a devout follower of Shintō, believing his own birth to have been the answer to his parents' prayer to the gods. The early Native Studies scholars such as Kamo no Mabuchi had often specialized in the historical philology necessary for studying Japan's most ancient texts, primarily literary or poetic. Because of his religious interests, however, Motoori applied his philological expertise to interpreting the *Kojiki (Records of Ancient Matters)*, a text written in an experimental, short-lived orthography that had been virtually unintelligible for a millennium. Motoori believed the *Kojiki* contained the actual accounts of creation. For four decades, therefore, he analysed the text line by line, translating it into language intelligible to the ordinary reader of his own day.

The connection between Motoori's religious and aesthetic theories lay in his understanding of *kokoro* and its relation to both language and reality (see Kokoro). Motoori argued that the ancient Japanese believed in the self-expressive character of things. In human beings, the *kokoro* is the responsive centre of both emotions and intellect. According to his interpretation of the classical Japanese world view, however, things and words also have *kokoro*. Having a 'genuine *kokoro*', the person is touched by the '*kokoro* of things', thus activating the spiritual and poetic 'power of words' (*kotodama*). If the audience lets themselves be touched by those words with a pure heart, they encounter the *kokoro* of the original expressive event. In short, language is ideally not simply a human product, but the outgrowth of a special resonance among the involved persons, things and words.

This view of ideal expression was intimately related to Motoori's reverence for the *Kojiki*. He believed the tradition that the *Kojiki* was a literal transcription of the ancient Japanese oral tradition, which itself supposedly went back to the original words spoken at the time of creation and arising out of the *kokoro* of that moment. Therefore, if his philological studies could recapture those words, Motoori believed he could, in effect, ritually participate in the re-enactment of creation.

Of course, such a cosmogony necessarily privileged ancient Japan, its language, its gods and the role of its emperor as the intermediary between the gods and humanity. If those claims about origins seemed incredible, Motoori argued, that very incredible appearance only made their truth more likely. The more rationalistic accounts of China (or the West) were suspect precisely because they better met human standards of reasoning; generations of narrators could have corrupted the miraculous facts to make them more acceptable to human reason. Motoori argued that the sacred should inspire awe rather than rational appreciation. Because the *Kojiki* had been written in an orthography that had quickly become unintelligible to the ancients, its account preserved the irrational character of the miracle of creation (see Creation and conservation, religious doctrine of).

Although Motoori himself might not have been an ultranationalist, his account of Japan in world history lent itself to such a world view. Shortly after his death, the Native Studies movement took a more ethnocentric and nationalistic turn. In that way, Motoori's Shintōphilosophy became intertwined with both national and international political agenda.

In his theory of poetics, Motoori focused on the role of aesthetic community, the culture that both articulates and serves as audience for literary expression (see Aesthetics, Japanese). In this respect, he regarded the culture of the Heian court (794-1185) to be the zenith of aesthetic culture. According to Motoori, the genuine *kokoro* of the ancient period was best expressed when there were elaborate rules and a common aesthetic regimen governing the interaction of writer and audience. Such a context, he believed, could generate a heightened sensitivity, subtlety and elegance.

In conclusion, through his philological studies, his formulation of Shintōthought and his detailed theory of poetics, Motoori articulated many principles that continue to be influential in Japanese attempts at cultural or ethnic self-understanding.

See also: Aesthetics, Japanese; Japanese philosophy; Kokoro; Shintō

THOMAS P. KASULIS

List of works

Motoori Norinaga (1747-1801) *Motoori Norinaga zenshū (Complete Works of Motoori Norinaga)*, ed. Ōkubo Tadashi and Susumu Ōno, Tokyo: Chikuma Shobō, 1968-93, 23 vols.(The standard edition of the original works.)

Motoori Norinaga (1771) *Naobi no mitama (The Way of the Gods)*, trans. Nishimura Sey, 'The Way of the Gods: Motoori Norinaga's *Naobi no mitama*', *Monumenta Nipponica* 42, 1991: 21-41.(An essay section of Motoori's commentary on the *Kojiki*, outlining his understanding of the Shintō worldview.)

Motoori Norinaga (1781) *Tamakushige (The Jeweled Comb-Box)*, trans. J.S. Brownless, 'The Jeweled Comb-Box: Motoori Norinaga's *Tamakushige*', *Monumenta Nipponica* 43, 1988: 35-61.(Translation of Motoori's response to a neo-Confucian critique of Native Studies.)

Motoori Norinaga (1798) *Uiyamabumi (First Steps into the Mountains)*, trans. Nishimura Sey, 'First Steps into the Mountains: Motoori Norinaga's *Uiyamabumi*', *Monumenta Nipponica* 42, 1987: 449-93.(Translation of Motoori's classic essay about the methodology of Native Studies.)

References and further reading

Harootunian, H.D. (1988) *Things Seen and Unseen: Discourse and Ideology in Tokugawa Nativism*, Chicago, IL: University of Chicago Press.(Provocative Marxian critique of the Native Studies movement from the standpoint of political ideology.)

Maruyama Masao (1952) *Nihon seiji shisōshi kenkyū*, trans. Mikiso Hane, *Studies in the Intellectual History of Tokugawa Japan*, Tokyo: University of Tokyo Press, 1974.(A classic study of the relation between Tokugawa thought and politics with sections I: 4 and II: 5 especially addressing Motoori's thought.)

Matsumoto Shigeru (1970) *Motoori Norinaga 1730-1801*, Cambridge, MA: Harvard University Press.(Helpful intellectual biography examining the connections between Motoori's life and thought.)

Nosco, P. (1990) *Remembering Paradise: Nativism and Nostalgia in Eighteenth-Century Japan*, Harvard-Yenching Institute Monograph Series 31, Cambridge, MA: Harvard University Press.(Good study of the romantic qualities in the Native Studies scholars' treatment of literature and history.)

Yoshikawa Kōjirō (1975) *Jinsai Sorai Norinaga*, trans. Y. Kikuchi, *Jinsai Sorai Norinaga: Three Classical Philologists of Mid-Tokugawa Japan, Tokyo: The TōhōGakkai (The Institute of Eastern Culture)*, 1983.(The essay on Motoori is an excellent overview of his thought with a special emphasis on his theory of language.)

Mozi (5th century BC)

Mozi was the first philosopher to question the ideas of Confucius. Scholarly debate centres around the issue of whether Mozi was a 'weak' or a 'strong' utilitarian, an 'act' or 'rule utilitarian', and whether he was a 'language utilitarian' or rather placed the religious authority of a personalized Heaven at the centre of his system. He is noteworthy for being the first thinker to develop a tripartite methodology for verifying claims to knowledge and for attacking the Confucian emphasis on ritual and the centrality of the family as the basis for social and political action.

Little is known about the life of Mozi (otherwise known as Modi or Zi Mozi (the Master Mozi)) except what is recorded in later, apocryphal stories, and scholars are even unsure what his name 'Mo' means or designates. Although he is credited in the 'Treatise on Bibliography' in the *Hanshu (History of the Former Han Dynasty)* with being a member of the aristocratic elite of *daifu* rank in the state of Song, some argue that 'Mo', literally 'black ink', refers to the fact that he was a carpenter of low social status; others claim that he was at some time a convict and branded on the face.

Mozi elaborated ten theses in opposition to Confucian doctrine, now found in the book *Mozi*, which contains all that is left of writings from the Mohist tradition. These theses are 'elevating the worthy', 'conforming with superiors', 'universal love' (also interpreted as 'concern for everyone' and 'impartial love'), 'rejecting aggression', 'thrift in use', 'thrift in funerals', 'heaven's will' (or intent), 'elucidating ghosts', 'rejecting music' and 'rejecting destiny'.

Mozi gathered numerous disciples, founding a sect organized on military lines with a single designated leader who had the powers of life and death over his followers. The Mohists specialized in defensive warfare and hired themselves out as experts to rulers of cities under attack. Towards the end of the Warring States period the Mohists may have divided into three separate groups, each of which criticized the others as being heretical; yet all were deeply involved in defending their master's original vision from attack by contemporary philosophers, producing a second major stage of philosophical theorizing.

Severely damaged, Mohist writings were almost totally ignored throughout Chinese history until restored by the Qing scholar and official Bi Yuan (1730-97). Since then, the greatest effort has gone into rediscovering and interpreting later Mohist thought, with its forays into disciplines now identifiable as logic, ethics and science, especially geometry, optics, mechanics and economics. This later Mohist corpus, written in three stages - 'Expounding the Canons', the 'Canons' and 'Explanations', and 'Names and Objects' - developed a rigorously logical, rational and systematic discourse for determining the three sources and four kinds of knowledge, eliminating the grounding of the philosophy in the divine.

See also: Chinese philosophy; Confucian philosophy, Chinese; Logic in China; Mohist philosophy

ROBIN D.S. YATES

List of works

Mozi (5th century BC) *Mozi*, ed. and trans. Mei Yi-Pao, *The Ethical and Political Works of Motse*, London: Arthur Probsthain, 1929; trans. A. Forke, *Me Ti: des Sozialethikers und seiner Schuler philosophische Werke (Mo Di: Philosophical Works of Social and Ethical Thinkers and of the Mohist School)*, Berlin: Kommissionsverlag der Vereinigung Wissenschaftlicher Verleger, 1922; ed. and trans. B. Watson, *Mo Tzu: Basic Writings*, New York: Columbia University Press, 1963.(Forke is the first translation of the *Mozi* into a Western language, with a detailed introduction and explanatory and textual comments, that attempts, prematurely, to indentify an 'Ur-text'. Watson is a more up-to-date English translation.)

References and further reading

Graham, A.C. (1978) *Later Mohist Logic, Ethics and Science*, Hong Kong: The Chinese University Press, and London: School of Oriental and African Studies.(A highly technical reconstruction, translation and analysis of the later Mohist corpus, including a valuable analysis of the linguistic features of Mohist writings.)

Graham, A.C. (1985) *Divisions in Early Mohism Reflected in the Core Chapters of Mo-tzu*, Singapore: National

University of Singapore, Institute of East Asian Philosophies.(A preliminary attempt to identify three different schools of Mohism on the basis of a technical linguistic analysis of the triads of chapters expounding the ten theses of the *Mozi*.)

Lowe, S. (1992) *Mo Tzu's Religious Blueprint for a Chinese Utopia: The Will and the Way*, Lampeter: The Edwin Mellen Press.(Analysis of the essay and dialogue sections of *Mozi* from a religious perspective that argues that Mohist religion is a coherent, practical system, based on the Will of Heaven, for transforming society and establishing an ideal state.)

Needham, J. and Yates, R.D.S. (1994) *Science and Civilisation in China*, vol. 5, *Chemistry and Chemical Technology*, Part VI, 'Military Technology: Missiles and Sieges', Cambridge: Cambridge University Press. (Extensive analysis and discussion of Mohist defensive tactics in historical and cross-cultural perspective.)

Sun Yirang (ed.) (1936) *Mozi xiangu (Casual commentaries on Mozi)*, Shanghai: Shangwu yinshuguan.(The most detailed and valuable edition and commentaries on the *Mozi* by a Chinese scholar.)

Yates, R.D.S. (1980) 'The City under Siege: Technology and Organization as Seen in the Reconstructed Text of the Military Chapters of Mo Tzu', unpublished Ph.D. dissertation, Harvard University.(A technical reconstruction, with introduction, and translation of chapters 14-15 of the *Mozi* on defensive warfare.)

Yan Lingfeng (1975) *Wuqiubeijai Mozi jicheng (Collected Editions of Mozi from the Studio that does not Seek for Completion)*, Taibei: Chengwen chubanshe, 46 vols.(Collection of the most important editions of and commentaries on the *Mozi* from pre-twentieth and twentieth century sources.)

Mujō

A Japanese word originating in Buddhism, mujō *means impermanence, transience or mutability. It characterizes all phenomena of experience, but is especially significant for human endeavours to achieve happiness. In the Buddhist analysis of existence, all things arise and perish through dependent origination; they are impermanent, without substance and continually subject to change. This presents human beings with an imperative existential problem, for they ignore or fail to realize the pervasiveness of impermanence. Endeavouring to secure lasting satisfaction, they cling to what is transient and mutable. The result is suffering, the unsatisfactory nature of ordinary - unenlightened - human existence.*

The problem of impermanence is fundamental to Buddhism from its inception, but interpretations of the concept varied with the evolution of other doctrines, most notably the Mahāyāna *notion of emptiness. The indigenous Japanese sensitivity to the transience of life and nature interacted with Buddhism to articulate, often in aesthetic terms, not only the threats but also the contributions of impermanence to meaningful human existence. The most thorough examination of* mujō *is offered by the Zen Buddhist Dōgen, who in his exposition of the radical temporality of existence as being-time identifies the full realization of impermanence with Buddha-nature or enlightenment.*

1 The problem of impermanence

Mujō is the Japanese word for the Buddhist concept of impermanence (the corresponding Sanskrit term is *anitya*). The traditional Buddhist statement of impermanence is ‘all conditioned things are impermanent’, meaning that they arise, pass away and are subject to change (see Momentariness, Buddhist doctrine of). *Mujō* may thus be translated as ‘transience’ or ‘mutability’. The two Chinese characters making up the word literally mean ‘without constancy’, which further indicates the possible irregularity of change and implies a spatial sense of instability. *Mujō* can also mean evanescence; things may be so fleeting as to have only a tenuous, dreamlike existence.

Originally, the concept of impermanence was descriptive of all experienced phenomena. In the early Buddhist formulation of the fundamental doctrine of dependent origination (*pratītyasamupāda*), all things arise and cease dependent upon causes or conditions. All existence is thus impermanent and nonsubstantial. Importantly, a person has no permanent self but is only a stream of changing interdependent factors. These claims are made on the basis of a close inspection of experience, including meditative states.

The primary implication of impermanence - the problem it presents - is that ordinary ways of seeking satisfaction or meaning in human life are ultimately unsatisfactory and induce suffering (*duḥkha*). That is, human beings mistakenly formulate the concept of a permanent self and seek lasting and secure satisfaction through it and its attachment to things, all of which - self, satisfactions, attachments and things - are actually transient and mutable, hence undependable. The continual pursuit of these dependent and undependable satisfactions is *saṃsāra*; in contrast, *nirvāna* is nondependent or unconditioned (see Nirvāṇa).

The cause of suffering, however, is not impermanence *per se* but ignorance of ‘things as they are’, for example in their conditional arising and impermanence, and the disposition to seek satisfaction through attachments or dependencies. Since the causes of suffering arise conditionally, they too are impermanent and removable. Their elimination, however, requires a thorough transformation of one’s whole person. A key factor in this is the existential awareness of impermanence, for as it deepens so too does the realization of the futility of attachments and the motivation for release from them (see Suffering, Buddhist views of origination of).

2 Background developments in Buddhism

To enhance awareness of impermanence, the scholastic Buddhism that evolved in the centuries following the Buddha’s death (*circa* 483 BC) postulated a theory of moments as part of an analysis of things into their ultimate, atomistic constituents (*dharma*). The Sautrāntika, for example, dissolved the apparent temporal durations of ordinary experience into a succession of discrete, virtually durationless moments, resulting in such a discontinuous conception of change that it raised serious problems in accounting for dependent origination. The Sarvāstivāda theory, on the other hand, attempted to account for continuity by arguing - surprisingly for Buddhists - that all *dharmas* are unchanging real existents at all times and change is mere appearance. Other disputes arose over

theories that tried to mitigate the problem of impermanence by introducing a permanent self, principle, or transcendent form of the Buddha. Since impermanence was associated with *saṃsāra*, *nirvāṇa* was often viewed as permanent.

Mahāyāna Buddhism sought to avoid such controversies by reformulating the nonsubstantiality originally implied by impermanence and dependent arising through the more spatially derived notion of emptiness (śūnyatā) (see Buddhist concept of emptiness). This occurred especially in the Perfection of Wisdom literature (*circa* 100 BC-AD 500) and the works of Nāgārjuna (*circa* AD 150-250). Emptiness means lack of any independent self-nature or defining essence. Everything exists only in relative interdependence, and linguistic distinctions cannot be assumed to designate independently real or substantially existing referents. This empties all ontological categories of any claims to ultimacy, and thus of anything to which one could form attachments.

The emptiness of all conceptual designations means that impermanence and permanence are relatively dependent and neither can be taken as describing the true nature of things. The *Heart Sutra* (*circa* AD 350), for example, states this in the language of nonduality as all *dharmas* marked with emptiness are equally not produced and not destroyed. The distinctions between past, present and future are likewise relatively dependent and not absolute. Dependent origination itself is consequently not conceived in temporal terms of arising and passing away but rather as relative interdependence, or emptiness. Nāgārjuna argues that arising, abiding and perishing are like an illusion or dream. While the language of impermanence may have practical usage at the conventional level, it is not descriptively true, as was posited in early Buddhism, of things as they are.

The lack of defining essence or self-nature in temporal concepts led to relational theories of time in China such as the Tiantai notion of ‘three thousand worlds in one instant of thought’ and the Huayan conception of ‘ten ages’. The ten ages are the present, past and future to each of the three time periods of past, present and future, together with their inclusion in one instant. That is, each moment is interdependently related to a vast field of time, and all time originates interdependently with each moment. The interplay of relations within this temporal field, which seem modelled on spatial relationships, are elaborately explained to show how each time is what it is in relation to all other times, and yet is just itself right now. In Huayan terms, the times interpenetrate each other but do not obstruct each other (see Buddhist philosophy, Chinese).

The actual aim of such metaphysical theories is to return one to spontaneous and harmonious participation in the immediate dynamics of experience, free from delusory mediation, by substantializing concepts and dispositional striving. Impermanence is still conjoined with dissatisfaction and suffering if one holds to a linear conception of time as sequential arising and passing away, but this is only a partial understanding which is to be supplanted by a holistic, relational view of time in light of emptiness.

3 Japanese affirmation of impermanence

The Japanese have often viewed impermanence not only as a problem but also as a fundamental condition contributing to the realization of meaningful human existence. There is a long-standing tradition that emphasizes a deep emotional response to transience and often sees evanescence as heightening the appreciation of beauty.

Affective responses of sorrow or longing can affirm the value and meaningfulness of that which is lost or longed for and, through revealing the shared nature and fate of all things in their transience, allow identification with them. This affective affirmation is enhanced by focusing on the most poignantly beautiful aspects of natural change, typically on such things as autumn leaves, spring blossoms or morning dew, which are beautiful in and through their transience. Evanescence further accentuates this beauty through such qualities as temporal rarity, purity or dreamlike delicacy (see Aesthetics, Japanese).

The aesthetic appreciation of transience stands in complex relation to the primary thrust of Buddhism, for while both see the affective response to impermanence as sorrow, the former treats emotion as revelatory of dimensions of value, whereas the latter views emotion as leading to attachment and enmeshment in delusion. This tension was creatively explored in classical and medieval literature where impermanence was a major theme, and where the modes of temporality could be investigated and expressed through literary form as well as content.

As Japanese responses to transience became more fully informed by Buddhist conceptions in the twelfth to sixteenth centuries, the view of change widened to include such grim aspects as social upheaval and natural

catastrophe. An austere aesthetic of impermanence evolved that admired first the muted and then the cold, desolate beauty associated with the elemental reduction of things through attrition and loss. This aesthetic contributed to theoretical developments in Nō drama, the tea ceremony and linked poetry. The reduction of life to its most extreme conditions reveals the latent, regenerative power of nature and balances perishing and dissolution against arising and creativity. Neither side can exist without the other. In this delicate balance the inconstancy, incompleteness and nonsubstantiality associated with transience become conditions and means for creatively engaging and affirming an impermanent world. As affective responses shade more fully into the Buddhist view, they move from attachment to compassion.

4 Dōgen

The Japanese Zen master Dōgen gives impermanence special prominence by identifying *mujō* (impermanence) with Buddha-nature and offering a descriptive exposition of the radical temporality of existence in terms of being-time. In Dōgen's view, 'The expounding, practising and realizing (enlightening) of impermanence by the impermanent themselves all must be impermanent'. That is, the problem of impermanence is resolved through deepening one's realization of impermanence, which occurs most fundamentally through the practice of Zen meditation.

The identification of impermanence with Buddha-nature means that there are not separate realms where one achieves enlightenment (the realization of Buddha-nature) by escaping or transcending impermanence. Furthermore, Buddha-nature is not to be construed as an essence or attribute which one has or acquires. Rather, it is the occurrence of things as they are, free from the imposition of delusory notions of fixed essences and attachments. As identified with impermanence, it is ongoing, continually occurring and being relinquished in each situation. Experientially, the event of Buddha-nature has an original wholeness and fullness which we reflectively objectify and divide through such distinctions as self and world, being and time, now and then.

Dōgen clarifies the temporality of *mujō* in terms of being-time, stating that 'time, just as it is, is being, and being is all time'. Time and being are inseparable; there is only temporal existence. This reaffirms the nonsubstantiality, or emptiness, of all things through their temporality. Things are better understood as events taking place as constituents of the interrelational temporal-spatial field of an occasion.

Dōgen offers two primary characterizations of the temporality of being-time: right now (*nikon*) and seriatim passage (*keireki*). The former is the immediate presence of events as they are within the temporal occasion, seen in discontinuity with the before and after of other times. It has a dimension of constancy as it is 'always' right now, but it is not like a point on a line or a container within which things occur. Such views, according to Dōgen, are inadequate to lived experience and contribute to the separation of time from being, and from oneself, which results in the problem of impermanence.

Each present occasion can also be viewed in its timefulness, that is, in its continuity with other times, their relations contributing to what it is as in the Huayan view. Seriatim passage indicates this continuity or interdependency with the extended field of other times. Dōgen depicts these interrelations as multidirectional, but unlike Huayan, he presents his views through descriptions of the temporality of experience rather than through metaphysical speculation.

Principally through renewed interest in Dōgen by twentieth-century Japanese philosophers, most notably of the Kyoto school (see Kyoto School), the concept of *mujō* has begun to enter comparative discussions with Western views of temporality and time.

See also: Buddhist concept of emptiness; Buddhist philosophy, Chinese; Buddhist philosophy, Japanese; Dōgen; Momentariness, Buddhist doctrine of

MONTE S. HULL

References and further reading

Abe Masao (1971) 'Dōgen on Buddha Nature', repr. in Masao Abe, *Zen and Western Thought*, ed. W.R. LaFleur, Honolulu, HI: University of Hawaii Press, 1985, 25-68. (General introduction to Dōgen, with discussion of identification of impermanence and Buddha-nature.)

Conze, E. (1958) *Buddhist Wisdom Books*, repr. New York: Harper, 1962.(Includes translation and commentary

on the *Heart Sutra*.)
Dōgen (1240) *Shōbōgenzō*, trans. N. Waddell, 'Being Time, Dōgen's *Shōbōgenzō Uji*', *The Eastern Buddhist* new series 12 (1) 1979: 114-29.(Reliable translation of a demanding text.)
Dōgen (1241) 'Dōgen's *Shōbōgenzō Buddha-nature*', part 2 trans. N. Waddell and Masao Abe, *The Eastern Buddhist* new series 9 (1) 1976: 87-105.(Identifies impermanence and Buddha-nature. Japanese text of both fasciles available in *Dōgen* I, ed. T. Terada and Y. Mizuno, *Nihon shisōtaikei* vol. 12, Tokyo: Iwananami Shoten, 1970.)
Fung Yu-lan (1953) *A History of Chinese Philosophy*, trans. D. Bodde, Princeton, NJ: Princeton University Press, vol. 2, 354-5, 370-4.(A standard history, giving the Huayan view of time.)
Inada, K. (1970) *Nāgārjuna: A Translation of his Mulamadhyamakakarika*, Tokyo: Hokuseido Press.(See especially chapters 7 and 19.)
Kalupahana, D.J. (1992) *A History of Buddhist Philosophy: Continuities and Discontinuities*, Honolulu, HI: University of Hawaii Press.(An introductory history; useful on early and scholastic Buddhist interpretations of impermanence.)
Karaki, J. (1965) *Mujō (Impermanence)*, Tokyo: Chikuma Shobo.(A history of impermanence in Japanese thought and literature, with special attention to Dōgen. Part 1, section 6 on the aesthetics of impermanence is translated and adapted in 'Wafting Petals and Windblown Leaves: Impermanence in the Aesthetics of Shinkei, Sōgi and Bashō', trans. A. Birnbaum, *Chanoyu Quarterly* 37 (1980): 7-27.)
LaFleur, W.R. (1983) *The Karma of Words: Buddhism and the Literary Arts in Medieval Japan*, Berkeley, CA: University of California Press.(Includes treatment of impermanence by selected literary figures.)
Nishitani Kenji (1982) *Religion and Nothingness*, trans. J. Van Bragt, Berkeley, CA: University of California Press.(A Kyoto school philosopher who discusses impermanence together with time and emptiness.)
Stambaugh, J. (1990) *Impermanence is Buddha-nature: Dōgen's Understanding of Temporality*, Honolulu, HI: University of Hawaii Press.(Interprets and compares Dōgen with Western philosophers.)

Mulla Sadra (Sadr al-Din Muhammad al-Shirazi) (1571/2-1640)

Sadr al-Din al-Shirazi (Mulla Sadra) is perhaps the single most important and influential philosopher in the Muslim world in the last four hundred years. The author of over forty works, he was the culminating figure of the major revival of philosophy in Iran in the sixteenth and seventeenth centuries. Devoting himself almost exclusively to metaphysics, he constructed a critical philosophy which brought together Peripatetic, Illuminationist and gnostic philosophy along with Shi'ite theology within the compass of what he termed a 'metaphilosophy', the source of which lay in the Islamic revelation and the mystical experience of reality as existence.

Mulla Sadra's metaphilosophy was based on existence as the sole constituent of reality, and rejected any role for quiddities or essences in the external world. Existence was for him at once a single unity and an internally articulated dynamic process, the unique source of both unity and diversity. From this fundamental starting point, Mulla Sadra was able to find original solutions to many of the logical, metaphysical and theological difficulties which he had inherited from his predecessors. His major philosophical work is the Asfar (The Four Journeys), which runs to nine volumes in the present printed edition and is a complete presentation of his philosophical ideas.

1 The primacy of existence

Sadr al-Din Muhammad ibn Ibrahim ibn Yahya al-Qawami al-Shirazi, known variously as Mulla Sadra, Sadr al-Muta'allihin, or simply Akhund, was born in Shiraz in central Iran in AH 979-80/AD 1571-2. He studied in Isfahan with, among others, Mir Damad and Shaykh Baha' al-Din al-'Amili, Shaykh-e Baha'i, before retiring for a number of years of spiritual solitude and discipline in the village of Kahak, near Qum. Here he completed the first part of his major work, the *Asfar (The Four Journeys)*. He was then invited by Allah-wirdi Khan, the governor of Fars province, to return to Shiraz, where he taught for the remainder of his life. He died in Basra in AH 1050/AD 1640 while on his seventh pilgrimage on foot to Mecca.

Safavid Iran witnessed a noteworthy revival of philosophical learning, and Mulla Sadra was this revival's most important figure. The Peripatetic (*mashsha'i*) philosophy of Ibn Sina had been elaborated and invigorated at the beginning of the Mongol period by Nasir al-Din al-Tusi, and there existed a number of important contributors to this school in the century before Mulla Sadra. Illuminationist (*ishraqi*) philosophy, originated by Shihab al-Din al-Suhrawardi, had also been a major current (see Illuminationist philosophy). The speculative mysticism of the Sufism of Ibn al-'Arabi had also taken firm root in the period leading up to the tenth century AH (sixteenth century AD), while theology (*kalam*), particularly Shi'ite theology, had increasingly come to be expressed in philosophical terminology, a process which was initiated in large part by al-Tusi (see Mystical philosophy in Islam; Islamic theology). Several philosophers had combined various strands from this philosophical heritage in their writings, but it was Mulla Sadra who achieved a true fusion of all four, forming what he called 'metaphilosophy' (*al-hikma al-muta'aliya*), a term he incorporated into the title of his magnum opus, *al-Hikma al-muta'aliya fi'l-asfar al-'aqliyya al-arba'a (The Transcendent Wisdom Concerning the Four Intellectual Journeys)*, known simply as the *Asfar*.

Mulla Sadra made the primacy of existence (*asalat al-wujud*) the cornerstone of his philosophy. Aristotle (§§11-12) had pointed out that existence was the most universal of predicates and therefore could not be included as one of the categories, and al-Farabi added to this that it was possible to know an essence without first knowing whether it existed or not, existence thus being neither a constitutive element of an essence nor a necessary attribute, and that therefore it must be an accident. But it was Ibn Sina who later became the source for the controversy as to how the accidentality of existence was to be conceived. He had held that in the existence-quiddity (*wujud-mahiyya*) or existence-essence relationship, existence was an accident of quiddity. Ibn Rushd had criticized this view as entailing a regress, for if the existence of a thing depended on the addition of an accident to it, then the same principle would have to apply to existence itself. This was merely an argument against the existence-quiddity dichotomy, but al-Suhrawardi had added to this another argument, asserting that if existence were an attribute of quiddity, quiddity itself would have to exist before attracting this attribute in order to be thus qualified. From this, al-Suhrawardi deduced the more radical conclusion that existence is merely a mental concept with no corresponding reality, and that it is quiddity which constitutes reality.

It was this view, that of the primacy of quiddity (*asalat al-mahiyya*), which held sway in philosophical writing in Iran up to Mulla Sadra's time. Indeed, Mir Damad, Mulla Sadra's teacher, held this view. However, Mulla Sadra himself took the opposite view, that it is existence that constitutes reality and that it is quiddities which are the mental constructs. By taking the position of the primacy of existence, Mulla Sadra was able to answer the objections of Ibn Rushd and the Illuminationists by pointing out that existence is accidental to quiddity in the mind in so far as it is not a part of its essence. When it is a case of attributing existentiality to existence, however, what is being discussed is an essential attribute; and so at this point the regress stopped, for the source of an essential attribute is the essence itself.

2 The systematic ambiguity of existence

A concomitant of Mulla Sadra's theory that reality and existence are identical is that existence is one but graded in intensity; to this he gave the name *tashkik al-wujud*, which has been usefully translated as the 'systematic ambiguity' of existence. Al-Suhrawardi, in contrast to the peripatetics, had asserted that quiddities were capable of a range of intensities; for example, when a colour, such as blue, intensifies it is not a new species of 'blueness' which replaces the old one, but is rather the same 'blue' intensified. Mulla Sadra adopted this theory but replaced quiddity with existence, which was for him the only reality. This enabled him to say that it is the same existence which occurs in all things, but that existential instances differ in terms of 'priority and posteriority, perfection and imperfection, strength and weakness' (making reality similar to al-Suhrawardi's Light). He was thus able to explain that it was existence and existence alone which had the property of combining 'unity in multiplicity, and multiplicity in unity'.

Reality is therefore pure existence, but an existence which manifests itself in different modes, and it is these modes which present themselves in the mind as quiddities. Even the term 'in the mind', however, is merely an expression denoting a particular mode of being, that of mental existence (*al-wujud al-dhihni*), albeit an extremely attenuated mode. Everything is thus comprehended by existence, even 'nothingness', which must on being conceived assume the most meagre portion of existence in order to become a mental existent. When reality (or rather a mode of existence) presents itself to the mind, the mind abstracts a quiddity from it - being unable, except in exceptional circumstances, to grasp existence intuitively - and in the mind the quiddity becomes, as it were, the reality and existence the accident. However, this 'existence' which the mind predicates of the quiddity is itself merely a notion or concept, one of the secondary intelligibles. It is this which is the most universal and most self-evident concept to which the Aristotelians referred, and which al-Suhrawardi regarded as univocal. But in reality there are not two 'things', existence and quiddity, only existence - not the concept, but the reality - and so 'existence' cannot be regarded as a real attribute of quiddity; for if this were possible quiddity would have to be regarded as already existent, as al-Suhrawardi had objected.

3 Substantial motion

Another of the key properties of existence for Mulla Sadra is its transubstantiality, effected through what he termed motion in substance (*al-haraka fi'l-jawhar*) or substantial motion (*al-haraka al-jawhariyya*). The peripatetics had held that substance only changes suddenly, from one substance to another or from one instant to another, in generation and corruption (and therefore only in the sublunar world), and that gradual motion is confined to the accidents (quantity, quality, place). They also held that the continuity of movement is something only in the mind, which strings together a potentially infinite series of infinitesimal changes - rather in the fashion of a film - to produce the illusion of movement, although time as an extension is a true part of our experience. What gives rise to movement is an unchanging substrate, part of the essence of which is that it is at an indefinite point in space at some instant in time; in other words, movement is potential in it and is that through which it becomes actual. Mulla Sadra completely rejected this, on the grounds that the reality of this substance, its being, must itself be in motion, for the net result of the peripatetic view is merely a static conglomeration of spatio-temporal events. The movement from potentiality to actuality of a thing is in fact the abstract notion in the mind, while material being itself is in a constant state of flux perpetually undergoing substantial change. Moreover, this substantial change is a property not only of sublunary elemental beings (those composed of earth, water, air and fire) but of celestial beings as well. Mulla Sadra likened the difference between these two understandings of movement to the difference between the abstracted, derivative notion of existence and the existence which is reality itself.

Existence in Mulla Sadra’s philosophical system, as has been seen, is characterized by systematic ambiguity (*tashkik*), being given its systematic character by substantial motion, which is always in one direction towards perfection. In other words, existence can be conceived of as a continual unfolding of existence, which is thus a single whole with a constantly evolving internal dynamic. What gives things their identities are the imagined essences which we abstract from the modes of existence, while the reality is ever-changing; it is only when crucial points are reached that we perceive this change and new essences are formed in our minds, although change has been continually going on. Time is the measure of this process of renewal, and is not an independent entity such that events take place within it, but rather is a dimension exactly like the three spatial dimensions: the physical world is a spatio-temporal continuum.

All of this permits Mulla Sadra to give an original solution to the problem which has continually pitted philosophers against theologians in Islam, that of the eternity of the world. In his system, the world is eternal as a continual process of the unfolding of existence, but since existence is in a constant state of flux due to its continuous substantial change, every new manifestation of existence in the world emerges in time. The world - that is, every spatio-temporal event from the highest heaven downwards - is thus temporally originated, although as a whole the world is also eternal in the sense that it has no beginning or end, since time is not something existing independently within which the world in turn exists (see Eternity).

4 Epistemology

Mulla Sadra’s radical ontology also enabled him to offer original contributions to epistemology, combining aspects of Ibn Sina’s theory of knowledge (in which the Active Intellect, while remaining utterly transcendent, actualizes the human mind by instilling it with intellectual forms in accordance with its state of preparation to receive these forms) with the theory of self-knowledge through knowledge by presence developed by al-Suhrawardi. Mulla Sadra’s epistemology is based on the identity of the intellect and the intelligible, and on the identity of knowledge and existence. His theory of substantial motion, in which existence is a dynamic process constantly moving towards greater intensity and perfection, had allowed him to explain that new forms, or modes, of existence do not replace prior forms but on the contrary subsume them. Knowledge, being identical with existence, replicates this process, and by acquiring successive intelligible forms - which are in reality modes of being and not essential forms, and are thus successive intensifications of existence - gradually moves the human intellect towards identity with the Active Intellect. The intellect thus becomes identified with the intelligibles which inform it.

Furthermore, for Mulla Sadra actual intelligibles are self-intelligent and self-intellected, since an actual intelligible cannot be deemed to have ceased to be intelligible once it is considered outside its relation to intellect. As the human intellect acquires more intelligibles, it gradually moves upwards in terms of the intensification and perfection of existence, losing its dependence on quiddities, until it becomes one with the Active Intellect and enters the realm of pure existence. Humans can, of course, normally only attain at best a partial identification with the Active Intellect as long as they remain with their physical bodies; only in the case of prophets can there be complete identification, allowing them to have direct access to knowledge for themselves without the need for instruction. Indeed, only very few human minds attain identification with the Active Intellect even after death.

5 Methodology

Even this brief account of Mulla Sadra’s main doctrines will have given some idea of the role that is played in his philosophy by the experience of the reality which it describes. Indeed he conceived of *hikma* (wisdom) as ‘coming to know the essence of beings as they really are’ or as ‘a man’s becoming an intellectual world corresponding to the objective world’. Philosophy and mysticism, *hikma* and Sufism, are for him two aspects of the same thing. To engage in philosophy without experiencing the truth of its content confines the philosopher to a world of essences and concepts, while mystical experience without the intellectual discipline of philosophy can lead only to an ineffable state of ecstasy. When the two go hand in hand, the mystical experience of reality becomes the intellectual content of philosophy.

The four journeys, the major sections into which the *Asfar* is divided, parallel a fourfold division of the Sufi journey. The first, the journey of creation or the creature (*khalq*) to the Truth (*al-haqq*), is the most philosophical; here Mulla Sadra lays out the basis of his ontology, and mirrors the stage in the Sufi’s path where he seeks to control his lower *nafs* under the supervision of his *shaykh*. In the second journey, in the Truth with the Truth, the

stage at which the Sufi begins to attract the divine manifestations, Mulla Sadra deals with the simple substances, the intelligences, the souls and their bodies, including therefore his discussion of the natural sciences. In the third journey, from the Truth to creation with the Truth, the Sufi experiences annihilation in the Godhead, and Mulla Sadra deals with theodicy; the fourth stage, the journey with the Truth in creation, where he gives a full and systematic account of the development of the human soul, its origin, becoming and end, is where the Sufi experiences persistence in annihilation, absorbed in the beauty of oneness and the manifestations of multiplicity.

Mulla Sadra had described his blinding spiritual realization of the primacy of existence as a kind of ‘conversion’:

> In the earlier days I used to be a passionate defender of the thesis that the quiddities are the primary constituents of reality and existence is conceptual, until my Lord gave me spiritual guidance and let me see His demonstration. All of a sudden my spiritual eyes were opened and I saw with utmost clarity that the truth was just the contrary of what the philosophers in general had held.… As a result [I now hold that] the existences (*wujudat*) are primary realities, while the quiddities are the ‘permanent archetypes’ (*a‘yan thabita*) that have never smelt the fragrance of existence.

(*Asfar*, vol. 1, introduction)

Therefore it is not surprising that Mulla Sadra is greatly indebted to Ibn al-‘Arabi in many aspects of his philosophy. Ibn Sina provides the ground on which his metaphilosophy is constructed and is, as it were, the lens through which he views Peripatetic philosophy. However, his work is also full of citations from the Presocratics (particularly Pythagoras), Plato, Aristotle, the Neoplatonists (see Neoplatonism in Islamic philosophy) and the Stoics (taken naturally from Arabic sources), and he also refers to the works of al-Farabi, and Abu’l Hasan al-‘Amiri, who had prefigured Mulla Sadra’s theory of the unity of intellect and intelligible. This philosophical heritage is then given shape through the illuminationism of al-Suhrawardi, whose universe of static grades of light he transformed into a dynamic unity by substituting the primacy of existence for the latter’s primacy of quiddity. It is in this shaping that the influence of Ibn al-‘Arabi, whom Mulla Sadra quotes and comments on in hundreds of instances, can be most keenly felt. Not only is that apparent in Mulla Sadra’s total dismissal of any role for quiddity in the nature of reality, but in the importance which both he and Ibn al-‘Arabi gave to the imaginal world (*‘alam al-mithal*, *‘alam al-khayal*).

In Ibn Sina’s psychology, the imaginal faculty (*al-quwwa al-khayaliyya*) is the site for the manipulation of images abstracted from material objects and retained in the *sensus communis*. The imaginal world had first been formally proposed by al-Suhrawardi as an intermediate realm between that of material bodies and that of intellectual entities, which is independent of matter and thus survives the body after death. Ibn al-‘Arabi had emphasized the creative aspects of this power to originate by mere volition imaginal forms which are every bit as real as, if not more real than, perceptibles but which subsist in no place. For Mulla Sadra, this world is a level of immaterial existence with which it is possible for the human soul (and indeed certain higher forms of the animal soul) to be in contact, although not all the images formed by the human soul are necessarily veridical and therefore part of the imaginal world. For Mulla Sadra, as also for Ibn al-‘Arabi, the imaginal world is the key to understanding the nature of bodily resurrection and the afterlife, which exists as an immaterial world which is nevertheless real (perhaps one might say more real than the physical world), in which the body survives as an imaginal form after death.

Philosophy has always had a tense relationship with theology in Islam, especially with the latter’s discourse of faith (*iman*) and orthodoxy. In consequence, philosophy has often been seen, usually by non-philosophers, as a school with its own doctrines. This is despite the assertions of philosophers themselves that what they were engaged in was a practice without end (for, as Ibn Sina had declared that what is known to humankind is limited and could only possibly be fulfilled when the association of the soul with the body is severed through death), part of the discipline of which consisted in avoiding *taqlid*, an uncritical adherence to sects (see Islam, concept of philosophy in). It is the notable feature of Mulla Sadra’s methodology that he constantly sought to transcend the particularities of any system - Platonic, Aristotelian, Neoplatonic, mystical or theological - by striving to create through his metaphilosophy an instrument with which the soundness of all philosophical arguments might be tested. It is a measure of his success that he has remained to the present day the most influential of the ‘modern’ philosophers in the Islamic world.

See also: Existence; Ibn al-‘Arabi; Ibn Sina; Illuminationist philosophy; Islamic philosophy, modern;

Metaphysics; Mir Damad; al-Sabzawari

JOHN COOPER

List of works

Mulla Sadra [Sadr al-Din al-Shirazi] (*c.*1628) *al-Hikma al-muta'aliya fi-'l-asfar al-'aqliyya al-arba'a (The Transcendent Wisdom Concerning the Four Intellectual Journeys)*, ed. R. Lutfi *et al.*, Tehran and Qum: Shirkat Dar al-Ma'arif al-Islamiyyah, 1958-69?, 9 vols; vol. 1, 2nd printing, with introduction by M.R. al-Muzaffar, Qum: Shirkat Dar al-Ma'arif al-Islamiyyah, 1967.(This is Mulla Sadra's major work, often known simply as *Asfar (The Four Journeys)*. The full edition includes partial glosses by 'Ali al-Nuri, Hadi al-Sabzawari, 'Ali al-Mudarras al-Zanuzi, Isma'il al-Khwaju'i al-Isfahani, Muhammad al-Zanjani and Muhammad Husayn al-Tabataba'i.)

Mulla Sadra [Sadr al-Din al-Shirazi] (*c.*1628) *Kitab al-masha'ir (The Book of Metaphysical Penetrations)*, ed., trans. and intro. by H. Corbin,*Le livre des pénétrations métaphysiques*, Paris: Départment d'Iranologie de l'Institut Franco-Iranien de Recherche, and Tehran: Librairie d'Amerique et d'Orient Adrien-Maisonneuve, Bibliothèque Iranienne vol. 10, 1964; French portion re-edited Lagrasse: Verdier, 1988; ed. and trans. P. Morewedge, *The Metaphysics of Mulla Sadra*, New York: Society for the Study of Islamic Philosophy and Science, 1992. (Corbin is a synopsis of Mulla Sadra's ontology, with a useful bibliography of Mulla Sadra's writings and introduction by Corbin. Morewedge provides a parallel Arabic-English edition; the translation is based on Corbin's edition of the text.)

Mulla Sadra [Sadr al-Din al-Shirazi] (*c.*1628) *al-Hikma al-'arshiyya (The Wisdom of the Throne)*, ed. with Persian paraphrase by G.R. Ahani, Isfahan, 1962; trans. and intro. J.W. Morris, *The Wisdom of the Throne: An Introduction to the Philosophy of Mulla Sadra*, Princeton, NJ: Princeton University Press, 1982.(A useful summary of Mulla Sadra's views on theology and eschatology; the introduction to the English translation provides an informative general introduction to Mulla Sadra work.)

References and further reading

Izutsu Toshihiko (1971) *The Concept and Reality of Existence*, Studies in the Humanities and Social Relations 13, Tokyo: Keio Institute of Cultural and Linguistic Studies.(Although concerned primarily with the philosophical ideas of Mulla Sadra's principal nineteenth century follower, Mulla Hadi al-Sabzawari, this work contains an extremely valuable exposition of the history of the existence-essence controversy in metaphysics, and deals with Mulla Sadra's views in many places.)

Nasr, S.H. (1978) *Sadr al-Din Shirazi and His Transcendent Theosophy: Background, Life and Works*, Tehran: Imperial Academy of Philosophy.(The first part of a planned, but so far uncompleted, two-volume work, the second volume of which is intended to deal with Mulla Sadra's philosophical ideas; contains the best bibliography of Mulla Sadra's works.)

Nasr, S.H. (1996) 'Mulla Sadra: His Teachings', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 643-52.(Short summary of Mulla Sadra's thought.)

Rahman, F. (1975) *The Philosophy of Mulla Sadr (Sadr al-Din al-Shirazi)*, Albany, NY: State University of New York Press. (To date, the only full-scale study of Mulla Sadra's philosophy in English.)

Ziai, H. (1996) 'Mulla Sadra: His Life and Works', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 635-42.(Biographical essay discussing Mulla Sadra's influence and works.)

Multiculturalism

Multicultural political philosophy explores ways of accommodating cultural diversity fairly. Public policies often have different consequences for members of different cultural groups. For example, given the importance of language to culture, and the role of the modern state in so many aspects of life, the choice of official languages will affect different people very differently. Similar issues arise concerning the cultural content of education and the criminal law, and the choice of public holidays. To avoid policies that create unfair burdens, multicultural theory turns to abstract inquiries about such things as the relation between culture and individual wellbeing, or the relation between a person's culture and the appropriate standards for judging them. Multiculturalism raises related questions for democratic theory also. Culture may be important to deciding on appropriate units of democratic rule and to the design of special mechanisms for representing minorities within such units. Each of these questions is made more difficult in the context of cultures that reject the demands of liberty or equality. The challenge for philosophers is to develop a principled way of thinking about these issues.

1 Multiculturalism and political thought

Political philosophers from Plato to Mill largely managed to ignore the question of culture. The societies in which they wrote enabled them in large part to take it for granted that the appropriate unit about which to ask questions concerning justice or democracy was a society that shared a culture. As a result, problems of justice or democracy were seen as superimposed on a homogeneous community. Although Mill saw nation states as inevitable, he believed that members of smaller cultural groups would readily give up their inherited culture to join other, stronger nations.

Demographic and political changes throughout the twentieth century have made traditional assumptions about the relations between culture and politics largely irrelevant to public life. Almost all of the world's countries now have substantial minorities from more than one culture. As a result, the ideals of political philosophers, whether concerning democracy, justice or membership, must be redeemed in settings very different to those for which they were first proposed.

Different states are multicultural in different ways. Belgium, Switzerland and Canada are federations of different language groups. Australia, Canada and the USA have substantial aboriginal populations. And most developed countries have large but geographically dispersed immigrant populations. Geographical concentration makes federalism and secession viable options for dealing with cultural difference. In its absence, what is needed is some way of building a political culture that does not exclude those who are different.

2 Liberal theories of cultural membership

Theories of cultural rights divide on a number of questions. While not all cultures are individualistic, most prominent views about multiculturalism share a root commitment to the view that culture matters because of its role in individual people's lives. They differ, however, in their detailed accounts of how and why it matters. Three accounts are prominent. First, some have suggested that culture is largely incidental to political life. While this cosmopolitan view does not deny that people have strong cultural attachments, it supposes that the state has no business taking up any sort of stance in relation to them. Instead the state should treat culture in the way many states treat religion, as a private matter on which the state is officially neutral (see Neutrality, political). Ideally people will be able to help themselves to the resources and possibilities of all of the world's cultures. Second, some have assigned each person's own culture an important role, either as an important aspect of identity or as a precondition of that person's ability to develop autonomously. On this view, the demands of justice are the same across cultures, but among those demands is the protection of the cultural conditions essential to individuality. Third, some have claimed that politics is an essential form of cultural expression, and that institutions should be designed to allow cultures to express themselves politically.

Cosmopolitanism is in many ways an attractive ideal, but there is some reason to doubt that it provides a realistic model for organizing political life. Most of the world's people have neither the resources nor the opportunities to move readily between cultures. Perhaps as important, few have any real inclination to do so. In the light of these circumstances and the central role of the modern state in coordinating economic and other activities, states have no

real alternative to adopting policies with significant cultural consequences. Official languages must be chosen and the legal framework within which people go about their affairs defined. Any response to these issues will make it easier for some cultural groups and more difficult for others. While advocates of cosmopolitanism might doubt the long-term significance of such effects, important questions of fairness arise concerning the effects of any such choices.

Multicultural liberals make fairness the starting point for their accounts of cultural policies. They insist that accommodating cultural difference is fundamentally different from discredited policies of segregation that sought to exclude minorities. For example, Will Kymlicka (1995) argues that aboriginal peoples should be given extra resources and political protection in order to enable them to maintain their cultural context. Because that context provides the background against which they are able to become autonomous, the state has a special responsibility to make sure that it is secure. Otherwise they would be unfairly deprived of something others receive free. The only way to protect that interest is to allow them to exclude outsiders from their society. In the same way, boundaries between states should be drawn in such a way as to enable as many people as possible to find themselves with a secure culture (see Nation and nationalism) Within self-governing areas, immigrants may be chosen with an eye to the society’s cultural stability. Others have made parallel arguments emphasizing the extent to which cultural identification provides a secure anchor in a changing world.

Still, any way of drawing boundaries will leave some people as members of dispersed minorities. Kymlicka advocates integrative policies for such groups so that maintaining some of their traditional practices does not pose a barrier to successful participation in the larger society in which they find themselves. Examples of such policies include allowing religious groups special exemptions from common pause days (when most businesses tend to be closed) or providing heritage language classes for the children of immigrants. In the UK, some have advocated extending anti-blasphemy laws to non-Christian religions. On similar grounds, many liberals advocate an inclusive educational curriculum. Such policies are sometimes said to offer advantages for members of the dominant culture as well as for minorities. In the longer term, they may well lead to the disappearance of many aspects of the cultures they aim to protect. As a result, the ultimate consequence of multicultural liberalism may well be the development of a cosmopolitan culture, in which cultural difference has a status little different from that of religion. While such a loss of diversity may be regarded as unfortunate, multicultural liberals defend their policies in terms of the interests of the current generation, and so have no commitment to diversity as such.

For multicultural liberals, cultural rights are always understood as rights that individuals have to certain cultural conditions. They are not rights that cultures have over their members. Those who wish to reject their inherited culture are always entitled to do so, however difficult they may find it. As a result, multicultural liberals insist that illiberal cultural practices are outside the bounds of toleration. While there are sometimes pragmatic and humanitarian grounds for limiting the use of force to eradicate such practices, communities that coerce dissident members are not entitled to support in so doing. Thus, everything from arranged marriages to coerced participation in religious rituals is outside the range of special protection or even toleration, however important they may be to the survival of the cultural context. Rights of exit are important for the same reasons (see Liberalism §3; Toleration §2).

Differing conceptions of property rights also complicate liberal attempts to protect minority cultures. Many aboriginal cultures (and some religious groups) have traditionally held land in common. If members wish to leave to join the surrounding society, they cannot take their share of the land with them without endangering the material basis of the culture. In such cases, there may be no way of protecting culture without sacrificing individual liberty.

3 The politics of identity

Advocates of identity politics insist that various forms of membership are far more important than cosmopolitans or multicultural liberals suppose them to be. It is sometimes suggested that cultures, as such, are worthy of respect, and that, as such, they require some form of political expression. So put, such a position is attributed more often to opponents than advocated seriously. But something like it marks a fundamental divide in thought about multicultural politics. For liberals, politics should aim to be fair, and so must be fair in its treatment of cultural difference. For others, politics aims to be expressive, and the suggestion that political institutions could ever avoid being carriers of a particular culture is at best naïve.

Expressivist views of politics and culture lead to a variety of policy suggestions. Many of these aim at ensuring long-term cultural survival, while others are supposed to ensure special political representation for groups whose views have been neglected in the past. The differences between these and the policies advocated by multicultural liberals are sometimes unclear. However, the bases for the claims are very different. In particular, where multicultural liberals insist on treating cultural membership as voluntary and so protect rights of exit, advocates of identity politics sometimes suggest the state has a special responsibility to protect found identities, or at least to tolerate groups that seek to do so. In this they come close to the philosophical position of communitarians, and adapt their political views to accommodate the fact of diversity (see Community and communitarianism).

Identity politics derives some of its appeal from the fact that many of the groups in whose name it is put forward have been treated badly by liberal societies. Perhaps its most important shortcoming is its inability to articulate a philosophical basis that does not apply equally well to cultural practices that have in the past been oppressive, especially to women. This problem is particularly pronounced when it is allied to a commitment to cultural survival as a fundamental political goal. Many of the most important aspects of human self-definition, from religious dissent through choice of sexual orientation, are called into question by the idea that cultural groups have some sort of claim over their members.

Recent developments in world politics have led some to scepticism about the prospects for multicultural societies and to arguments that liberal societies can only survive in the context of a shared culture. Only with a shared culture, it is suggested, will people be able to live together in peace and harmony, and only with a shared culture will there be enough social solidarity to call forth the sacrifices necessary to sustain fair terms of social cooperation. At present, such pessimism seems premature. The psychology of sacrifice is at best very poorly understood. While goodwill and the absence of grudges are surely needed for peaceful coexistence, and familiarity doubtless facilitates cooperation, the relation of these to culture in any more robust sense is unclear. A shared culture bears no clear relation to the more everyday demands of justice. (Were the relationship direct, it is probable that culturally homogeneous countries would have achieved a higher level of justice.) Justice needs a model of citizenship and a civic culture. One possible form such a culture can take is a multicultural world in which people learn to appreciate difference and to respect each other. Such a civic culture is perhaps the best hope for a world made up of multicultural societies.

See also: Citizenship (§3); Culture

ARTHUR RIPSTEIN

References and further reading

All of these works are written in nontechnical language, accessible to those with little or no background in political philosophy. Because of the explosion of interest in the topic, many new books and special issues of journals continue to appear.

Gutmann, A. (1993) ‘The Challenge of Multiculturalism in Political Ethics’, *Philosophy and Public Affairs* 22 (3): 171-206.(An examination of different proposals for dealing with culturally plural societies.)

Kukathas, C. (1992) ‘Are There Any Cultural Rights?’, *Political Theory* 20 (1): 105-39.(A critique of Kymlicka, advocating toleration of illiberal communities.)

Kymlicka, W. (1995) *Multicultural Citizenship*, Oxford: Clarendon Press.(A defence of liberal approaches to cultural difference.)

Multiculturalism (1992), special issue of *University of Michigan Journal of Law Reform* 25 (3).(A wide variety of legal and philosophical perspectives on multiculturalism.)

Parekh, B. (1986) ‘The Concept of Multicultural Education’, in S. Mogdil, G. Verma, K. Mallick and C. Mogdil (eds) *Multicultural Education: The Interminable Debate*, London: Falmer Press, 19-32.(A defence of multicultural education for both minority and non-minority students.)

Parekh, B. (1990) ‘The Rushdie Affair’, *Political Studies* 38 (4): 695-709.(A discussion of the political debates raised by the attempt to ban Salman Rushdie’s *The Satanic Verses* as blasphemous.)

Rorty, A.O. (1992) ‘The Hidden Politics of Cultural Identification’, *Political Theory* 22 (1): 152-66.(An examination of the ambiguities in the notion of cultural membership.)

Taylor, C. (1994) *Multiculturalism: Examining the Politics of Recognition*, ed. A. Gutmann, Princeton, NJ: Princeton University Press.(A leading essay by Taylor offers cultural survival as a rationale for state policies towards culture. Responses to Taylor’s essay, by Kwame Anthony Appiah, Jürgen Habermas, Stephen

Rockefeller, Michael Walzer and Susan Wolf cover most of the theories and issues described in this article.)
Waldron, J. (1992) ‘Minority Cultures and the Cosmopolitan Alternative’, *University of Michigan Journal of Law Reform* 25 (3): 751-93.(A defence of the cosmopolitan approach described in §2.)
Young, I.M. (1990) *Justice and the Politics of Difference*, Princeton, NJ: Princeton University Press. (A defence of differentiated citizenship.)

Multiple-conclusion logic

Ordinary arguments can have any number of premises but only one conclusion. Multiple-conclusion logic also allows for any number of conclusions in an argument, regarding them as setting out the field of possibilities among which the truth must lie if the premises are true. Such an argument counts as valid if it is impossible for all the premises to be true and all the conclusions false. Anything that can be said about premises can now be said, mutatis mutandis, *about conclusions, and much of the interest of the subject comes from exploiting this duality. Putting conclusions on a par with premises reflects the idea that truth and falsity, and likewise acceptance and rejection, are polar notions standing on a par with one another.*

1 Multiple-conclusion arguments

Multiple conclusions can best be understood by analogy with multiple premises. Premises function collectively: a number of premises taken together have a strength which far exceeds the sum of their separate strengths. The very existence of deductive logic turns on this fact, which is not undermined by the fact that a finite number of premises are always equivalent to a single one, namely their conjunction. For this equivalence itself relies on the working of the rule 'From *A* and *B* infer $A \,\&\, B$', understood as involving two separate premises (not one conjunctive one), so it does nothing to establish the redundancy of arguments with more than one premise.

Multiple conclusions also function collectively: a set of conclusions will typically be deducible although no single conclusion is deducible. They are best thought of disjunctively, as setting out the field within which the truth must lie on the assumption that the premises are all true. Just as enlarging a set of premises strengthens them (makes more things follow from them), so enlarging a set of conclusions weakens them (makes them follow from more things): as the field containing the truth becomes larger it becomes less informative. Of course, a finite number of conclusions are equivalent to a single one, namely their disjunction, but this no more makes multiple conclusions redundant than the analogous fact about conjunction makes multiple premises redundant.

The definition of logical consequence or implication is now that a set of premises implies a set of conclusions if it is impossible for all the premises to be true and all the conclusions false. The transitivity of implication is a key property of it, which in conventional single-conclusion logic takes the form of the 'cut' principle; that if a set of premises Γ implies a conclusion *A*, and if Γ, *A* imply *B*, then Γ implies *B*. The corresponding multiple-conclusion principle takes this novel form: if $\Gamma,\ \Theta_1$ implies $\Theta_2,\ \Delta$ for every partition of the set Θ into two disjoint parts Θ_1 and Θ_2, then Γ implies Δ.

Logicians often depict arguments by diagrams in which a horizontal line indicates that the sentence immediately below is inferred from those immediately above, so that the diagram branches upwards from one conclusion. Multiple-conclusion arguments can be represented in a similar fashion, except that now the graph can branch downwards as well as upwards. So (1) shows one variety of 'proof by cases', where one starts with a premise of the form '*A* or *B*' and presents the alternatives as distinct cases as a preliminary to establishing the conclusion by separate, parallel arguments for case *A* and case *B*.

(1) $$\begin{array}{cc} \multicolumn{2}{c}{A \vee B} \\ \hline A & B \\ \vdots & \vdots \\ C & C \end{array}$$

(2) $$\begin{array}{cc} \multicolumn{2}{c}{A \vee B} \\ \hline A & B \\ \hline \multicolumn{2}{c}{A \,\&\, B} \end{array}$$

(1) is untypically simple, for in general the combination of upward and downward branching means that branches which have split may subsequently rejoin, creating circuits. Moreover, it turns out that it is essential to allow for circuits (or some equivalent sophistication) if one is to obtain an adequate variety of arguments. But some circuits are vicious, (2), for example. Much of the mathematical development of the subject has therefore been concerned with finding criteria for a diagram to have a valid pattern, in the sense that any argument which has the same bare layout (that is, abstracting altogether from the sentences involved and paying attention only to the way the steps are arranged relative to one another) is bound to be valid provided its component steps are individually valid. This is a good example of a new concept which, because patterns of single-conclusion argument are so rudimentary by comparison, is scarcely appreciable within the conventional framework.

2 Motivation and applications

Multiple-conclusion logic commends itself to the attention of the philosophical logician because putting conclusions on a par with premises is a natural corollary of acknowledging that truth and falsity are polar notions which stand on a par with one another, and likewise acceptance and rejection. Deduction is no longer tilted towards acceptance alone: an argument regarded as going from premises Γ to conclusions Δ with an eye to acceptance can equally be regarded as going from Δ to Γ with an eye to rejection. The connection can be stated in suitably symmetrical terms, namely that if the argument is valid it is irrational to accept all of Γ and reject all of Δ.

For the mathematical logician the removal of the asymmetry embodied in single-conclusion logic means that some anomalies disappear, and distinctions and results which needed qualification become unconditional. Recall, for example, the conventional way of developing a system of logic. First the semantics is set out, then implication is defined in semantic terms and finally, where possible, it is axiomatized, that is, shown to be equivalent to formal deducibility according to some system of rules of inference. This exposes a deficiency. Evidently the notion of implication can be used to say things about the assignments of truth-values determined by the semantics - for example, that none of them makes certain premises true and a certain conclusion false. But the information that can be conveyed in this way is only a fraction of the whole. For example, there is no way of saying in terms of implication alone whether the set of all sentences is satisfiable or not. And even if all the notions with which logicians work - logical truth, implication, satisfiability and so on - are taken together, the situation is similar. What we would surely like, but do not have, is a logical notion or notions capable of saying all there is to say about the semantics at the level appropriate to sentences.

In a similar vein, Carnap (1943) pointed out that, although one may manage to axiomatize a logic initially given semantically, one can never work back and retrieve the semantics from the axiomatization. For example, any axiomatization which is sound and complete for classical propositional logic will equally be sound and complete with respect to a nonstandard semantics in which, in addition to the classically possible assignments of truth-values, every sentence (and its negation) is simultaneously satisfiable. In short, there is no way, in purely formal deductive terms, of expressing the fact that '$\sim$' signifies negation. The situation is akin to the non-categoricity of theories, whereby a first-order theory of real numbers will also have a nonstandard interpretation in a merely countable domain (see Löwenheim-Skolem theorems and nonstandard models). Those who are rightly concerned about the implications of that notorious result might do well to reflect that things are no better for negation.

Multiple-conclusion logic remedies both deficiencies. Implication is now an adequate vehicle for saying all there is to say about any semantics at the level of sentences, for every fact of the form 'It is (is not) possible for these sentences Γ to be true and those ones Δ false' is expressible by the statement that Γ does not (does) imply Δ. And by using multiple-conclusion rules one can obtain axiomatizations from which the intended semantics can be uniquely reconstructed. For example, one can work back from the rules 'From *A*, $\sim A$ infer' and 'Infer *A*, $\sim A$' to the classical truth table for '$\sim$'.

Multiple-conclusion rules of inference can also provide an elegant alternative to the often complicated rules involving assumptions and their discharge, which characterize natural deduction systems. (1) above illustrates the idea by using a multiple-conclusion rule in place of the natural-deduction rule of $\vee$-elimination. A profound application of this idea can be found in Gentzen's work on 'sequents' if one interprets his sequents metalogically as statements about multiple-conclusion implication, though this was not how Gentzen himself regarded them (see Natural deduction, tableau and sequent systems).

See also: Logical and mathematical terms, glossary of

TIMOTHY SMILEY

References and further reading

Carnap, R. (1943) *Formalization of Logic*, Cambridge, MA: Harvard University Press.(The first explicit appearance of the subject.)

Kneale, W.C. (1956) 'The Province of Logic', in H.D. Lewis (ed.) *Contemporary British Philosophy*, 3rd series, London: Allen & Unwin; abridged in W. Kneale and M. Kneale, *The Development of Logic*, Oxford: Clarendon

Press, 1962, ch. 9, §3.(Another pioneering piece, with a very readable introduction.)
Shoesmith, D.J. and Smiley, T.J. (1978) *Multiple-Conclusion Logic*, Cambridge: Cambridge University Press.(A systematic treatise, including a discussion of the history of the subject.)

Music, aesthetics of

The aesthetics of music comprises philosophical reflection on the origin, nature, power, purpose, creation, performance, reception, meaning and value of music. Some of its problems are general problems of aesthetics posed in a musical context; for example, what is the ontological status of the work of art in music, or what are the grounds of value judgments in music? Other problems are more or less peculiar to music, lacking a clear parallel in other arts; for example, what is the nature of the motion perceived in music, or how can the marriage of music and words best be understood?

Attempts to define the concept of music generally begin with the fact that music involves sound, but also posit such things as cultural tradition, the fulfilment of a composer's aims or the expression of emotions as essential features of music. Perhaps any plausible concept, though, has to involve the making of sounds by people for aesthetic appreciation, broadly conceived. In deciding what is meant by a musical work, further considerations come into play, such as might lead to the identification of it with a sound structure as defined by a given composer in a particular musico-historical context.

In what sense can a piece of music be said to have meaning? Some hold it has meaning only internally - in its structure as an arrangement of melodies, harmonies, rhythms and timbres, for instance - while others have claimed that its meaning lies in the communication of things not essentially musical - such as emotions, attitudes or the deeper nature of the world. The most popular of these beliefs is that music expresses emotion. This is not to say, however, that the emotion expressed in a work is necessarily experienced by those involved in its composition or performance: composers can create peaceful or furious music without themselves being in those states, and the same goes for the performance of such music by performers. Also, the emotions evoked in listeners seem of a different nature from those directly experienced: negative emotions expressed in music do not preclude the audience's appreciation, and in fact commonly facilitate it. Ultimately, a work's expressiveness should be seen as something directly related to the experience of listening to that work. Music is often said to have value primarily in so far as it is beautiful, its beauty being whatever affords pleasure to the listener. But the quality of a work's expressiveness, its depth, richness and subtlety, for example, also seems to form an important part of any value judgment we make about the work.

1 The concept of music

A fundamental question in the aesthetics of music is 'What is music?', understood as a request for a definition or delineation of the general concept of music. Theorists have adopted a number of different approaches to the question, often depending on their further purposes in asking it. Perhaps the only thing that all theorists agree on is that music is necessarily sound.

The most conservative approach seeks to define music in terms of the standard features of most music, to wit, melody, harmony, rhythm, metre, instruments, voices and tone production; but in addition to being inadequate to a number of modes of contemporary composition - serialism, minimalism, *musique concrète*, computer music, aleatory music - it ignores many practices remote from us in space or time that we acknowledge as musical. A more liberal approach to defining music - a structuralist one - issues in the formula of music as 'organized sound'; though music so conceived need not exhibit the standard features of music noted above, for on this conception something is music in virtue of its intrinsic properties alone. On another approach, which may be termed 'experiential' (or 'phenomenological'), music is any sounds that are heard as music, it then being incumbent on the theorist to say what makes such hearing musical; salient implications of this approach are, first, that natural and accidental sounds may be music, and second, that the status of an item as music is relative to the listener and the occasion of listening.

Attempts have been made to define music as a type of sound-involving activity distinguished by certain cultural or sociological traits, for example, a particular cultural function, such as the accompaniment of ritual or the enhancement of group memory, or particular social relationships, such as apprenticeship. Alternatively, music may be defined in an essentially historical manner as those sound-involving items, activities and practices that have evolved, historically and reflexively, from certain earlier such items, activities and practices, and so on, it being incumbent on the theorist to indicate some non-question-begging way of picking out the musical strand of human

history from all others (for instance, the linguistic one) that are sound-involving.

Finally, one may attempt to characterize music intentionally, from the producer's point of view, by appeal to distinctive aims or purposes on the part of makers of sound. One approach of long standing conceives music as sounds made in order to express, evoke or elicit emotions or feelings. Another conceives it as sound used as a vehicle of communicable but non-linguistic thought. But the most common approach of this kind proposes simply that music is sounds made or arranged for aesthetic appreciation.

If one is concerned to define music as an art, to preserve a measure of objectivity for the status of music, and yet to avoid making music's import necessarily either emotional or intellectual, one could do worse than to accept this last suggestion. However, in order comfortably to cover a wide range of cross-cultural phenomena easily recognized by us as music, but in which we would be hard put to discern a norm of aesthetic appreciation in operation, a more inclusive notion of the aim with which sounds are made must be invoked. The following suggests itself: music is sounds humanly made or arranged for the purpose of enriching experience via active engagement (such as through performing, listening, dancing), with the sounds regarded primarily as sounds. Such a definition looks to be adequate to cover virtually everything intelligibly accounted music.

The proposal accommodates even John Cage's notorious *4′33″*, an ostensibly soundless musical composition. Designating a period of silence is (though a limiting case) an organization of sounds in a given span of time, presumably done here with a view to heightening consciousness. In addition, as the composer clearly envisaged, even a span of time specified as one to be left silent by the performer is inevitably filled with sounds of various sorts originating from the environment in which the piece is performed.

2 The ontology of music

The central question of the ontology of music, in musical domains where the notion of a work - a repeatable and non-occasion-bound musical entity - has purchase, is that of what manner of thing a musical work precisely is.

It is clear that the standard musical work of the Western tradition is not any physical object or event whatsoever. In particular, it is not identical with any performance of it. It is equally not to be identified with any score of it, whether original manuscript or mass-produced copy, for such things are evidently seen and not heard. Furthermore, a musical work generally predates any of its performances, and can survive the destruction of all of its scores. Yet scores and performances remain of great importance: musical works in this tradition are largely defined by the former and experienced through the latter.

If a work of music is not a physical entity, what is it? There are four views on this question with some currency, three of which hold that a musical work is some variety of abstract entity. The first is that a musical work is a set or class of performances. The second is that it is a universal or pure type, such as a sound structure or pattern. The third is that it is a mental rather than an abstract entity, something existing properly in the minds of composers as well as perhaps of their interpreters and audiences. And the fourth is that a musical work is a qualified or contextualized type, akin to other products of culture in being creatable and bound to specific persons, times and places of origin. Finally, it is also possible to take an eliminativist view of musical works, denying that there really are any such things, and recognizing only scores, performances, intentions and associated practices.

Eliminativism aside, of the four views indicated above there are arguably conclusive objections to all but the last. Focusing on the second view, the idea that a musical work is simply a sound structure (or tonal-instrumental structure) confronts grave problems. One is that such a thing does not admit of being created; the total musical structure ultimately settled on by the composer, the upshot of various acts of choosing and arranging tones and instrumentations, already exists as an abstract object within the musical system in which composition takes place, and so cannot be brought into being by the composer. But a more important problem is that, conceived as simply a sound structure, a musical work cannot support the complex of aesthetic and artistic predications justifiably made of it. Two musical works, composed within the same musical system, may be identical in musical structure and yet differ in the aesthetic or artistic features that are ascribed to them, for example, brashness, wittiness or originality. This is due to the different contexts in which they were composed, and the different contexts of correct performance, audition, and understanding that this entrains. Only a view such as the fourth noted above, that individuates musical works more finely - for example, as composer-initiated types whose identity is bound up with person, time and place - can be adequate to musical works as they figure in our experience and description of

them. The act of composing a standard musical work is thus one of the composer indicating, in a specific musico-historical context, a musical structure, so creating the work which is precisely the structure-as-indicated-by-the-composer-in-that-context (see Art works, ontology of §§2-3).

3 Musical performance

A number of questions about the performance of music have received attention. Foremost is ‘What is it to perform?’ or ‘What criteria must be satisfied if a performance can be said to have taken place?’ One issue is whether performing requires a pre-existing work, or whether improvising is not also a kind of performing. Another issue is what sort of intentions a performer must have in order for their actions to count as a performance of such-and-such a work, what sorts of means must be employed, and what success conditions *vis-à-vis* a target audience might also need to be satisfied. The question of what it is to perform a work at all shades readily into the question of what it is to perform a work correctly and of what it is to perform it well. Issues such as these about the nature of musical performance and issues concerning the ontology of music are closely interrelated.

Much discussed is the matter of historical authenticity in performance - both what is required in order to achieve it, and why, or even whether, achieving it is desirable. Writers have variously emphasized the specific sound of original performance, the means of sound production employed, the manner of playing then current, or the physical venue and social circumstances of such performance. Alternatively, authority may be claimed to rest with the composer’s intentions, though this also raises problems: there are officially declared intentions, represented by scores, and privately held intentions; low-order intentions, merely implementational in nature, and high-order intentions, definitive of aesthetic goals; and intentions regarding what is essential for proper performance, in contrast to intentions in regard to achieving optimal performance.

The nature of the interpretation involved in performing has been queried, and compared and contrasted with critical interpretation. The status of musical performance as an art of its own has been explored, as have analogies between performing and other actions, for example, those of quoting, displaying, translating or enacting. The phenomenon of virtuosity has attracted attention, as attaching not only to performances but to works themselves. Finally, issues regarding the recording of music and its effect on the performance and the reception of music are beginning to be widely addressed (see Art, performing).

4 Musical form and musical perception

Questions about the basic form of music and the nature of the perception involved in grasping music at ground level are closely intertwined. For the basic form of music is arguably that in virtue of which it is heard as music, or that which is necessarily tracked in the course of perceiving the music.

Though all agree that music is sound which is organized and apprehended in time and thus that the form of music must be both audible and temporal, there is disagreement on many further points. Some hold the fundamental form of music to be local, and to reside in moment-to-moment connections between small-scale parts, while others hold global form, governing large-scale and temporally distant sections of a piece, to be equally basic. Some regard musical form as *sui generis*, involving irreducible and specifically musical qualities, while others take notions, particularly spatial ones, rooted in other domains - such as balance, proportion, symmetry and overall shape - to apply directly to musical form.

Debate also flourishes concerning the nature of the fundamental musical perception by which musical form is grasped. Some philosophers think registration of even the most basic musical features, such as tones, rhythms, motifs and chords, or at least the experience of musical connectedness and movement requires either a special mode of perception, or the metaphorical projection of concepts not literally applicable to sequences of sound, or the exercise of a species of imagination; others hold such posits to be unnecessary, claiming that ordinary perception is adequate to the phenomena at hand. A compromise position about musical movement may be this: even if such movement fails to result from either metaphorical projection or aural imagining, music is *heard as* moving, most notably in its melodic rise and fall, harmonic progression and rhythmic propulsion, despite failing to contain anything that literally moves in the way it is heard as moving.

Currently there is much empirical work on the cognitive psychology of music worthy of the attention of philosophers, concerning principles of grouping, the grasp of melodic contour, the mechanisms of memory and

attention, and the limits of sensitivity to key relationships. Also of interest is the role of unconscious processing in the perception of music, including assignment by a musical processor of syntactic or semantic structure to music as it is heard.

5 Musical understanding

What does the listener's understanding of music consist in? It is arguably a species of knowing-how, or experiential knowing, as opposed to knowing-that, or propositional knowing. Reflection suggests that understanding music is basically a matter of hearing the music in a certain way, of registering or responding to certain aspects of the music. The understanding of music, in other words, is continuous with perception of music; it is not a matter of technical analysis, causal explanation or discursive construal, but of a certain sort of listening experience (see Art, understanding of §5).

This is not to deny that there are kinds and levels of musical understanding, some of which may essentially involve propositional knowledge, that understanding at any level may be shallow or deep, and that different sorts of music may call for different kinds of understanding. But the ability to describe music in articulate terms is not a *sine qua non* of core musical understanding, even if ability to describe music is requisite for the kind or level of musical understanding appropriate to a teacher or critic.

Still, there is a question of the extent to which the listener's understanding of music is informed by concepts of various sorts, in what manner and with what degree of recoverability by the subject. There is also the issue of how similar the understanding of music is to the understanding of language, and of how the understanding of music relates to the appreciation and evaluation of it.

6 Musical meaning: general

It is common to divide views on the meaning or significance of music into two sorts, autonomist and heteronomist. The autonomist position is that music has no meaning, or else that it means only itself (thus yielding what is sometimes called 'intra-musical' meaning). The heteronomist position is that music has some sort of meaning that is other than the music itself (sometimes denominated 'extra-musical' meaning).

It is difficult to find thinkers whose views wholly exemplify either position. Perhaps Hanslick, who regarded music as essentially just a glorious succession of tones, and who held music incapable of conveying anything more than the dynamic qualities exhibited indifferently by phenomena of various sorts, is closest to a complete autonomist; Schopenhauer, who regarded music as an image of the inner nature of the world and held it to signify the infinite varieties of willing or striving, is perhaps closest to a complete heteronomist (see Hanslick, E.; Schopenhauer, A. §5). Gurney's view, contrary to common belief, is as much heteronomist as autonomist, though he regards music's expression of mental states as detachable from, and less important than, its purely musical beauty or impressiveness (see Gurney, E. §2). Langer's view is substantially heteronomist, postulating that music is symbolic of emotional life, though Langer denies that such symbolism carries as far as individual emotions. The view of Leonard Meyer is substantially autonomist, in that it accords pride of place to embodied meaning, consisting roughly in the implications that musical events have for other musical events in a musical fabric; yet room is also made for designative meaning of an emotional sort, explained as a by-product of the play of expectations involved in sensing the implications for continuation that a musical composition at every point presents (Meyer 1956; 1967).

It seems reasonable to take musical meaning and understanding to be correlative concepts, so that the meaning of a stretch of music would comprise whatever is understood in understanding it. Viewed from that perspective, the question of whether music has any meaning beyond itself becomes that of whether in understanding music we need register or respond to anything more than purely musical events and relationships. The answer to this would seem clearly to be yes; though order and connectedness in purely musical dimensions is the basis of musical discourse, it does not exhaust it, and a comprehending experience of that discourse accordingly goes beyond a grasp of musical relationships *per se*.

7 Musical meaning: expression

The sort of extra-musical meaning in music that has seemed to most observers to be of greatest importance is

expressive meaning, that is, the expression by music of psychological states. The states music has been held to be capable of expressing include emotions, feelings, moods, attitudes and traits of personality.

Philosophical explication of the concept of musical expressiveness must not be conflated with investigation of the grounds or causes of such expressiveness. The identification of factors involved in making music expressive - tempo, timbre, major or minor mode, or similarities between music and vocal utterance, for example - is one thing, and the logical analysis of what musical expressiveness consists in quite another.

The expressiveness of music, though evidently related to the literal expression of psychological states by persons through behaviour, countenance and demeanour, must yet be clearly distinguished from it. Taking emotions as the paradigm states involved, the expression of emotion by a person is a dated occurrence, involving outward manifestations that warrant a reasonable inference to the person's being in the given emotional state, and requires that the person actually feel the emotion being expressed. When, however, a musical passage is expressive of an emotion, the emotion is not an occurrent one, and any inference to emotion felt on the part of the composer would be unwarranted, and probably mistaken; nor is a musical passage literally a behavioural manifestation of any sort. Musical expressiveness, in short, is a property of music, not of individuals who happen to be connected to the music.

The logical distinctness of musical expressiveness and personal expression still allows for a remarkably persistent hypothesis, sometimes called the ‘Expression Theory’ of music, to the effect that the emotion a piece of music is expressive of is always as a matter of fact one that was experienced by the composer, and that the expressiveness of music is always in effect the expression of the composer's own emotion, imparted to the music in the act of composing. But apart from its romantic appeal, the hypothesis has little to recommend it. The composing of music is typically too indirect, intellectually mediated, temporally extended and discontinuous for such a generalization to hold, even roughly. A composer in a sanguine frame of mind can very well write sanguinary music, or the reverse, with this being as comprehensible as the only somewhat more frequent match of music and mood mistakenly posited as universal by the Expression Theory. And talented composers can craft music expressive of a number of emotional states without themselves being in any of them at the time.

The emotional expressiveness of music must also be distinguished from its power to arouse or evoke corresponding feelings in listeners. There are numerous reasons for this, but two of the more important seem to be these. First, the expressiveness of music presents itself as a manifest property of it, or at any rate, something about how it is readily heard, rather than as an inferred power to raise affect in us. And second, the affects that music does produce in us while listening will often differ, in degree, kind and polarity, from those which it is expressive of; if they did not, it would be puzzling both that music expressive of negative emotion was as popular as it was, and that the aesthetic appreciation of expressive music of any sort, which calls for a certain amount of clear-headed attention, was possible at all.

As for what emotional expressiveness in music might consist in, a number of ideas are currently under consideration. They include music's having a sound resembling the behavioural expression of an emotion; music's metaphorically exemplifying some emotion; music's sounding the way some emotion feels; music's presenting the appearances of an emotion; music's corresponding to an emotional state or being seen as suitable to its expression; music's being imagined to be the gestures of an individual experiencing emotion; and music's being hearable as, or as if it were, the personal expression of an emotion. In any event, it seems clear that the expressiveness of a musical passage should be conceived as something not detachable from, or experienceable apart from, the passage that possesses it.

Some theorists who fully acknowledge the expressive dimension of music balk at describing this as a species of musical meaning. Music's expressiveness would only be a form of meaning, they maintain, if in addition to possessing and exhibiting expressive qualities music also referred to, denoted or was otherwise about the emotions or other states corresponding to such qualities. But the issue can be mooted by simply speaking of music's expressive content, rather than meaning (see Artistic expression; Emotion in response to art §§4, 6).

8 Musical meaning: representation

Musical representation means the depiction by music of concrete persons, things or events. This is least problematic for sounds, or objects with characteristic sounds, that music can obviously imitate or approximate. But

the scope of musical representation plausibly extends beyond this to concrete phenomena that can be heard in music, or that a passage of music might be heard as, which hearing may obtain in the absence of notable sonic resemblance between music and the phenomenon in question, but in virtue instead of certain structural parallels or isomorphisms between them.

There is an important distinction between musical depictions that attuned listeners recognize as such unaided, and those that are only so recognized upon provision of a verbal cue or label, but there seems little reason to identify that with the distinction between representational and nonrepresentational music *per se*. The issue has also been raised of whether musical representation requires the composer’s intention to represent the putative subject, and whether this intention must be signalled by a composer-given title, or otherwise publically recorded.

Some have questioned whether even the most explicitly programmatic music is ever truly representational, on the grounds that representation requires more than securing perceptual reference to some object or event, representation being held to require also a characterization of what is referred to or the attribution of predicates to it, which music is claimed powerless to accomplish. But it is not clear either that such a demand is warranted, or, if it is, that music cannot meet it.

Philosophers have also been concerned to assess the importance of such musical representation as there is. It has been asked whether, when a piece of music is representational, its representational aspect must be grasped if the piece is to be understood, whether such music is not sometimes equally enjoyable if its representational aspect is ignored, and whether the representational and expressive properties of music are not often interdependent - in which event, assuming expressive properties are always of artistic moment, the appreciative relevance of the representational aspect of music in such cases would be assured.

9 Musical meaning: other aspects

The expression of mental states and the representation of concrete objects and events by no means exhausts the modes of meaning music has been held to possess. Additional meanings that have been ascribed to music fall mostly into one of two broad categories, metaphysical and sociocultural.

Metaphysical meanings include suggestions of a fairly moderate sort, such as that music symbolizes, in addition to or in lieu of individual psychological states, the sphere of feeling generally (Langer 1942), or that music exemplifies general patterns of continuation, growth or development in the natural world (Beardsley 1981), or that music is capable of modelling states of mind in which truths of human existence stand revealed (Sullivan 1927). Metaphysical meanings of a more radical sort have also been posited: Schopenhauer, for example, thought of music as a direct image of the cosmic will, and Nietzsche viewed music as divulging the pain and suffering at the root of earthly existence.

Cultural meanings are implicated in suggestions to the effect that music possesses ideological content, reflecting or endorsing existing political arrangements or social structures, or moral content, personifying and displaying virtues of character, or that it embodies a society’s attitudes towards sex, reason or religion. Unlike the expressive and representational meanings considered above, meanings of this sort, where justifiably ascribed, may be carried by whole genres of music, or by acts or contexts of presentation, rather than by specific works.

Supporters of broader meanings for music must indicate criteria for the possession of such meanings by music, mechanisms by which music may conceivably effect them, and reasons for holding them to be intersubjectively valid. The more extrinsic the kind of meaning proposed as attaching to music, the more difficult it will be to hold both that the understanding of music requires grasp of such meanings, and that such understanding will be primarily experiential in nature.

10 The value of music

The value of individual pieces of music must be distinguished from the value to a culture of music as a whole - though the latter will likely depend, in various ways, on the former. The artistic value of an individual piece of music may plausibly be identified, in the main, with the intrinsic value of correct experience of it.

Chief among the artistic values of music is presumably beauty - what Gurney labelled simply ‘musical impressiveness’, and what Hanslick regarded as inherent in ‘tonally moving forms’. Musical beauty, it might be

said, is that in virtue of which music is inherently pleasurable to follow, apprehend or contemplate. It is a point of contention whether such beauty is susceptible to explanation in terms of principles, if any, governing the emergence of beauty in other spheres, or is rather uniquely musical (see Beauty).

The other obvious artistic value of music derives from its expressiveness or extra-musical dimension generally. Music seems more valuable the more expressive it is, or the more richly, finely or profoundly it embodies whatever content it possesses beyond the purely musical.

Broadly speaking, views on the value of music divide into views that locate such value primarily in music's presentation of an autonomous world of sound, one whose forms and qualities can be appreciated without reference to those of life, and whose virtue consists in removing one satisfyingly from ordinary human affairs, and views that locate such value primarily in music's reflection of the real world, whose virtue is the effecting of a deeper or more intense immersion in the world, revealing or clarifying aspects of life that ordinarily remain obscure or only dimly grasped.

See also: Opera, aesthetics of

JERROLD LEVINSON

References and further reading

Adorno, T. (1938) 'On the Fetish Character in Music and the Regression in Listening', in A. Arato and E. Gebhardt (eds) *The Essential Frankfurt School Reader*, New York: Continuum, 1982.(Condemns 'easy listening' music and habits, partly for sociopolitical reasons.)

(1948) *Philosophie der neuen Musik*, trans. A. Mitchell and W. Bloomster, *The Philosophy of Modern Music*, London: Sheed & Ward, 1973.(A defence of advanced modes of composition, and of the superiority of art music to popular music generally.)

Alperson, P. (1984) 'On Musical Improvisation', *Journal of Aesthetics and Art Criticism* 43: 17-30.(Identifies the distinguishing features of improvised music.)

Alperson, P. (ed.) (1987) *What is Music?*, New York: Haven; reprinted University Park, PA, Penn State Press, 1994.(A collection of essays, mostly by philosophers, with substantial introductions and bibliographies.)

Beardsley, M. (1981) 'Understanding Music', inK. Price (ed.) *On Criticizing Music*, Baltimore, MD: Johns Hopkins University Press.(Distinguishes different types of understanding and explores the possibilities for semantic understanding of music in accord with Goodman's notion of artistic exemplification.)

Blacking, J. (1973) *How Musical Is Man?*, Seattle, WA: University of Washington Press.(An ethnomusicological perspective on some fundamental questions about music, focusing on the role of music in society and of society in music.)

Budd, M. (1985) *Music and the Emotions*, London: Routledge.(A masterly critical survey of influential theories of musical expression and evocation.)

Budd, M. (1985) 'Understanding Music', *Proceedings of the Aristotelian Society Supplement* 59: 233-48. (Explores dimensions of musical understanding, and criticizes Scruton (1983).)

Budd, M. (1989) 'Music and the Communication of Emotion', *Journal of Aesthetics and Art Criticism* 47: 129-38. (Examines the idea that musical expressiveness might consist in the musically mandated imagining of emotion.)

Budd, M. (1995) *Values of Art*, London: Allen Lane, The Penguin Press.(Advances in its last chapter an important positive account of the relationship of music and emotion.)

Busoni, F. (1911) *Sketch of a New Aesthetic of Music*, trans. T. Baker, New York: G. Schirmer.(Explores, in loose-jointed prose, the ideal of absolute music.)

Callen, D. (1982a) 'The Sentiment in Musical Sensibility', *Journal of Aesthetics and Art Criticism* 40: 381-93. (Analyses musical expressiveness as representation of fictive acts of emotional expression.)

Callen, D. (1982b) 'Making Music Live', *Theoria* 48: 139-68.(Insightful discussion of what is involved in performative interpretation of music beyond merely playing 'the notes'.)

Cavell, S. (1969) 'Music Discomposed', in *Must We Mean What We Say?*, New York: Charles Scribner's Sons. (Charges that certain modern modes of music-making are artistically fraudulent.)

Clifton, T. (1983) *Music as Heard*, New Haven, CT: Yale University Press.(An approach to musical comprehension from a phenomenological perspective, drawing mainly on the ideas of Husserl.)

Collingwood, R. (1938) *The Principles of Art*, Oxford: Oxford University Press.(Contains important reflections on the creative process in art and the operation of imagination in expression and appreciation.)

Cone, E. (1974) *The Composer's Voice*, Berkeley, CA: University of California Press. (Develops a thesis about musical communication in terms of the composers' musical personae, in both song and instrumental music.)
Cone, E. (1989) *Music: A View from Delft*, Chicago, IL: University of Chicago Press.(A collection of essays by the prominent music theorist, half of which have a philosophical dimension.)
Cook, N. (1990) *Music, Imagination and Culture*, Oxford: Oxford University Press.(A stimulating and provocative look at music from the point of view of both composer and listener, and a defence of the primacy of the latter.)
Cooke, D. (1959) *The Language of Music*, Oxford: Oxford University Press.(Proposes that music is indeed the language of the emotions and seeks through a wide range of examples to locate the individual units of emotive meaning in music; responds to Hindemith (1952).)
Copland, A. (1952) *Music and Imagination*, Cambridge, MA: Harvard University Press.(A composer's sensible view of the creation and appreciation of music, particularly good in its portrait of the gifted or ideal listener.)
Cox, R. (1985) 'Are Musical Works Discovered?', *Journal of Aesthetics and Art Criticism* 43: 367-74.(Gives interesting reasons for a negative answer to the question posed.)
Dahlhaus, C. (1981) *Esthetics of Music*, trans. W. Austin, Cambridge: Cambridge University Press.(A historically informed survey of the subject by the leading German musicologist-historian.)
Dahlhaus, C. (1990) *The Idea of Absolute Music*, trans. R. Lustig, Chicago, IL: University of Chicago Press.(A historical and philosophical examination of the idea of 'absolute' instrumental music.)
Davies, S. (1987) 'Authenticity in Musical Performance', *British Journal of Aesthetics* 27: 39-50.(Locates authenticity of performance in fidelity to the composer's publically expressed and correctly understood intentions.)
Davies, S. (1991) 'The Ontology of Musical Works and the Authenticity of Their Performance', *Nous* 25: 21-41. (A sensitive discussion of the interrelationship of these issues, with significant attention to non-Western musical traditions.)
Davies, S. (1994) *Musical Meaning and Expression*, Ithaca, NY: Cornell University Press.(A comprehensive and painstaking survey of its topics, covering some of the same ground as Budd (1985), but with a focus on work of the past twenty years.)
DeBellis, M. (1995) *Music and Conceptualization*, Cambridge: Cambridge University Press.(Explores the nature of expert and nonexpert hearing of music, and the role and degree of conceptualization in each; draws on work in cognitive psychology.)
Dewey, J. (1934) *Art as Experience*, New York: Putnam.(Contains important discussions of the creative process applicable to musical composition.)
Dipert, R. (1980) 'The Composer's Intentions: An Examination of Their Relevance for Performance', *Musical Quarterly* 66: 205-18.(Argues for the importance of distinguishing low-, middle- and high-level intentions on a composer's part.)
Elliott, R.K. (1966) 'Aesthetic Theory and the Experience of Art', *Proceedings of the Aristotelian Society* 67: 111-26.(Illuminatingly contrasts the experience of music 'from within' with the experience of music 'from without'.)
Evans, M. (1990) *Listening to Music*, London: Macmillan.(A recent discussion of some problems of musical meaning and comprehension, informed by the philosophy of Wittgenstein.)
Ferguson, D. (1960) *Music as Metaphor*, Minneapolis, MN: University of Minnesota Press.(Claims that music, which is capable of resembling emotions in two different respects, provides metaphors for emotions and so represents them.)
Fisher, J. (1991) 'Discovery, Creation, and Musical Works', *Journal of Aesthetics and Art Criticism* 49: 129-36. (Argues cogently for a creationist view of musical composition.)
Godlovitch, S. (1988) 'Authentic Performance', *The Monist* 71: 258-77.(Canvasses a large number of defences of authenticity in musical performance.)
Goehr, L. (1992) *The Imaginary Museum of Musical Works*, Oxford: Oxford University Press.(Surveys with dissatisfaction current accounts of the nature of a musical work, proposes instead adoption of a historical-cultural method, and suggests that the full-blown notion of a musical work may not predate 1800.)
Goldman, A. (1992) 'The Value of Music', *Journal of Aesthetics and Art Criticism* 50: 35-44. (Suggests that the central value of music is its presenting us with another world, removed from that of everyday life.)
Goldman, A. (1995) 'Emotions in Music: A Postscript', *Journal of Aesthetics and Art Criticism* 53: 59-69. (Assesses the debate between Kivy and Radford on musical arousal of emotion.)
Goodman, N. (1968) *Languages of Art*, New York: Bobbs-Merrill.(A seminal treatise in contemporary aesthetics,

containing influential proposals regarding the relationship of works and performances and the nature of musical expression.)

Gracyk, T. (1992) ‘Adorno, Jazz, and the Aesthetics of Popular Music’, *The Musical Quarterly* 76: 526-42.(A critical response to Adorno’s elitist strictures on music.)

Gurney, E. (1880) *The Power of Sound*, New York: Basic Books, 1966.(His major work, an important contribution to aesthetics of music in the nineteenth century.)

Hanslick, E. (1891) *Vom Musikalisch-Schönen*, trans. G. Payzant, *On the Musically Beautiful*, Indianapolis, IN: Hackett Publishing Company, 1986.(The classic statement of formalism about music.)

Higgins, K. (1991) *The Music of Our Lives*, Philadelphia, PA: Temple University Press. (Charges that contemporary philosophy of music largely ignores certain dimensions of musical experience, and argues in particular for the ethical role and significance of music.)

Hindemith, P. (1952) *A Composer’s World*, Cambridge, MA: Harvard University Press.(A searching, if one-sidedly formalistic, examination of some fundamental issues of musical aesthetics.)

Ingarden, R. (1957) *The Work of Music and the Problem of Its Identity*, trans. A. Czerniawski, Berkeley, CA: University of California Press, 1986.(Advances a view of the musical work as an intentional object, one culturally shaped and sustained, created in time and subject to dissolution.)

Katz, R. and Dahlhaus, C. (eds.) (1987) *Contemplating Music: Source Readings in the Aesthetics of Music*, New York: Pendragon Press, 4 vols.(An immense compendium of source readings, though somewhat idiosyncratically arranged.)

Kivy, P. (1984) *Sound and Semblance*, Princeton, NJ: Princeton University Press.(A careful study of the basis of musical representation, and the different species thereof.)

Kivy, P. (1989) *Sound Sentiment*, Philadelphia, PA: Temple University Press.(A partly historical and widely influential account of the nature and causes of musical expressiveness, with supplementary essays defending the author’s thoroughgoing cognitivism about music. Includes *The Corded Shell*, Princeton University Press, 1980.)

Kivy, P. (1990) *Music Alone*, Ithaca, NY: Cornell University Press.(An account of the experience of ‘music alone’, focusing on the role of explicit, articulable cognitions in the understanding of music.)

Kivy, P. (1993) *The Fine Art of Repetition*, Cambridge: Cambridge University Press.(A collection of essays, the most important of which concern the ontology of music and the curious status of music as an art.)

Kivy, P. (1995) *Authenticities*, Ithaca, NY: Cornell University Press.(A philosophical analysis of the whys and wherefores of the currently ascendant norm of ‘authentic performance’.)

Krausz, M. (ed.) (1993) *The Interpretation of Music*, Oxford: Oxford University Press.(A collection of new essays by nineteen philosophers.)

Langer, S. (1942) *Philosophy in a New Key*, Cambridge, MA: Harvard University Press. (Advances the view of music as a presentational, rather than discursive symbol, and claims that what it symbolizes, though non-specifically, is the human life of feeling.)

Langer, S. (1953) *Feeling and Form*, New York: Charles Scribner’s Sons.(A further development of ideas in Langer (1942) ranging across all the arts and featuring prominently the notions of ‘virtual space’ and ‘virtual time’.)

Levinson, J. (1990) *Music, Art, and Metaphysics*, Ithaca, NY: Cornell University Press. (Contains six essays devoted to philosophy of music, covering the definition, ontology, meaning, performance and appreciation of music.)

Levinson, J. (1992) ‘Composition, musical’, in *A Companion to Aesthetics*, ed. D. Cooper, Oxford: Blackwell. (Treats the questions of what it is to compose a musical work, and of what, if anything, characterizes the process of musical composition.)

Levinson, J. (1996) *The Pleasures of Aesthetics*, Ithaca, NY: Cornell University Press.(Contains four essays devoted to philosophy of music, including ones on song, musical literacy, musical expressiveness and musical interpretation.)

Levinson, J. (1997) *Apprehending Music*, Ithaca, NY: Cornell University Press.(An account, inspired by Edmund Gurney, of what the listener’s core understanding of music consists in.)

Levinson, J. (1998) *Music in the Moment*, Ithaca, NY: Cornell University Press.(Contains four essays devoted to philosophy of music, including ones on song, musical literacy, musical expressiveness and musical interpretation.)

Levinson, J. and Alperson, P. (1991) ‘What Is a Temporal Art?’, *Midwest Studies in Philosophy* 16: 439-50. (Details fourteen different notions of temporality in art, many applicable to music.)

Lippman, E. (1977) *A Humanistic Philosophy of Music*, New York: New York University Press.(An investigation

of music from a cultural and historical point of view.)
Lippman, E. (1994) *A History of Western Musical Aesthetics*, Lincoln, NB: University of Nebraska Press.(The latest historical survey of the subject, beginning with ancient thought, but covering writers as recent as Leonard Meyer and Roger Scruton.)
McClary, S. (1991) *Feminine Endings: Music, Gender, and Sexuality*, Minneapolis, MN: University of Minnesota Press.(A feminist reinterpretation of the tradition of Western tonal music, raising many philosophical questions about the basis and grounds of musical meaning.)
Maconie, R. (1990) *The Concept of Music*, Oxford: Oxford University Press.(An engaging and wide-ranging examination of the nature and importance of music.)
Mark, T. (1980) 'On Works of Virtuosity', *Journal of Philosophy* 77: 28-45.(Proposes that the distinctive mark of works of musical virtuosity is a kind of musical reflexiveness or self-reference.)
Mark, T. (1981) 'Philosophy of Piano Playing: Reflections on the Concept of Performance', *Philosophy and Phenomenological Research* 41: 199-324.(Likens a performance of a work to a quotation of it.)
Maus, F. (1988) 'Music as Drama', *Music Theory Spectrum* 10: 54-73.(Proposes that musical discourse be conceived as essentially dramatic, as an arena in which musical characters perform acts of various sorts.)
Mew, P. (1985) 'The Expression of Emotion in Music', *British Journal of Aesthetics* 25: 33-42.(Defends an evocation view of musical expressiveness.)
Meyer, L. (1956) *Emotion and Meaning in Music*, Chicago, IL: University of Chicago Press.(The first of Meyer's books, setting out his view in psychologistic terms, in which the notion of expectation figures prominently.)
Meyer, L. (1967) *Music, the Arts, and Ideas*, Chicago, IL: University of Chicago Press.(Ranges beyond music as such to other arts, and reframes some of Meyer's key ideas in terms of information theory.)
Meyer, L. (1974) *Explaining Music*, Berkeley, CA: University of California Press.(Highlights, in the course of numerous analyses, the notion of hierarchic levels of musical implication.)
Newcomb, A. (1984) 'Sound and Feeling', *Critical Inquiry* 10: 614-43.(A critical discussion of Kivy (1984).)
Nietzsche, F. (1872) *The Birth of Tragedy*, trans. W. Kauffmann, New York: Vintage Books, 1967.(Discusses the conflict of the Apollonian and the Dionysian at the heart of Greek tragedy, and the relation of this to 'the spirit of music'.)
Pearce, D. (1988) 'Intensionality and the Nature of a Musical Work', *British Journal of Aesthetics* 28: 105-18. (Suggests that musical works have the kind of existence accorded to mathematical objects by intuitionism in the philosophy of mathematics.)
Pratt, C. (1931) *The Meaning of Music*, New York: McGraw-Hill.(The work of an important psychologist of music, who advanced the view that music sounds the way emotions feel.)
Putman, D. (1987) 'Why Instrumental Music Has No Shame', *British Journal of Aesthetics* 27: 55-61.(Argues for and diagnoses the incapacity of music to express cognitively complex emotions.)
Radford, C. (1989) 'Emotions and Music: A Reply to the Cognitivists', *Journal of Aesthetics and Art Criticism* 47: 69-76.(Replies to Kivy (1989), and argues for the power of music to induce simple emotions and moods in listeners.)
Radford, C. (1991) 'How Can Music Be Moral?', *Midwest Studies in Philosophy* 16: 421-38.(Suggests that musical works reflect morally relevant states of mind and traits of personality.)
Raffman, D. (1993) *Language, Music and Mind*, Cambridge, MA: MIT Press.(Argues for three sorts of ineffability in connection with music, the most important being carried by uncategorizable nuances of sound as produced in concrete performance.)
Rantala, V., Rowell, L. and Tarasti, E. (eds) (1988) *Essays on the Philosophy of Music*, Helsinki: Acta Philosophica Fennica.(A collection of twenty essays, representing a variety of approaches: analytic, semiotic, historical and phenomenological.)
Ridley, A. (1995) *Music, Value and the Passions*, Ithaca, NY: Cornell University Press.(An impressive recent contribution, defending an emotionalism about music that mediates between cognitivist and arousalist approaches.)
Robinson, J. (1994) 'The Expression and Arousal of Emotion in Music', *Journal of Aesthetics and Art Criticism* 52: 13-22.(Emphasizes the importance of music's capacity to affect the feelings directly, in cognitively unmediated ways.)
Robinson, J. (ed.) (1997) *Music and Meaning*, Ithaca, NY: Cornell University Press.(A collection of ten recent essays by philosophers and theorists of music.)
Rowell, L. (1983) *Thinking About Music*, Amherst, MA: University of Massachusetts Press. (An introduction to the philosophy of music with an ethnographic flavour.)
Said, E. (1991) *Musical Elaborations*, New York: Columbia University Press.(Three interrelated essays by a

musically erudite cultural critic, focusing on issues of performance.)
Schoenberg, A. (1950) *Style and Idea*, New York: Philosophical Library.(Definitive statements by the founder of 12-tone music.)
Schopenhauer, A. (1819, 1844) *Die Welt als Wille und Vorstellung*, trans. E.F.J. Payne, *The World as Will and Representation*, New York: Dover Publications, 1966, 2 vols.(Contains a powerful brief for the metaphysical significance of musical process.)
Scruton, R. (1976) ‘Representation in Music’, *Philosophy* 51: 273-87.(Argues that music is incapable of representation in a robust sense.)
Scruton, R. (1976) ‘Representation in Music’, *Philosophy* 51: 273-87.(Argues that music is incapable of representation in a robust sense.)
Scruton, R. (1983) ‘Understanding Music’, in *The Aesthetic Understanding*, London: Methuen. (Argues that musical understanding is a species of intentional understanding, and that musical features are strictly distinct from sonic ones.)
Scruton, R. (1987) ‘Analytical Philosophy and the Meaning of Music’, *Journal of Aesthetics and Art Criticism* 46: 169-76.(Argues for a conception of musical meaning as inseparable from the experience of music.)
Serafine, M. (1987) *Music as Cognition*, New York: Columbia University Press.(A cognitive psychologist’s view of music.)
Sessions, R. (1950) *The Musical Experience of Composer, Performer, Listener* , Princeton, NJ: Princeton University Press.(Offers exactly what its title suggests, and is illuminating in regard to each of its topics.)
Shepherd, J. (1991) *Music as Social Text*, Oxford: Polity Press.(Explores music as a repository of social and cultural meanings.)
Sloboda, J. (1985) *The Musical Mind: The Cognitive Psychology of Music*, Oxford: Oxford University Press.(A valuable recent survey.)
Sparshott, F. (1994) ‘Music and Feeling’, *Journal of Aesthetics and Art Criticism* 52: 23-35. (Sensitive to the many senses of ‘feeling’ invoked in claims about musical expression and evocation.)
Sparshott, F. (1994) ‘Music and Feeling’, *Journal of Aesthetics and Art Criticism* 52: 23-35.(Sensitive to the many senses of ‘feeling’ invoked in claims about musical expression and evocation.)
Storr, A. (1992) *Music and the Mind*, New York: Free Press.(Genial meditations by a psychotherapist on the origin, power and beauty of music.)
Stravinsky, I. (1956) *Poetics of Music*, New York: Vintage Books.(Reflects an extreme formalism about music, arguably inconsistent with the composer’s own practice.)
Sullivan, J.W.N. (1927) *Beethoven: His Spiritual Development*, London: Jonathan Cape. (Suggests that some music, such as the last compositions of Beethoven, embodies and communicates to a listener spiritually and cognitively valuable states of mind.)
Tanner, M. (1985) ‘Understanding Music’, *Proceedings of the Aristotelian Society Supplement* 59: 215-32. (Emphasizes a hierarchy of levels of understanding music, with higher levels presupposing lower ones if they are to have any point.)
Tarasti, E. (1994) *A Theory of Musical Semiotics*, Bloomington, IN: Indiana University Press.(A comprehensive work by a leading advocate of semiotics as applied to music.)
Thom, P. (1993) *For an Audience: A Philosophy of the Performing Arts*, Philadelphia, PA: Temple University Press.(Advances a naturalistic understanding of works in the performing arts.)
Toch, E. (1948) *The Shaping Forces in Music*, New York: Criterion Music Corp.(A searching, lesser-known composer’s examination of the bases of musical form, and the special role of movement and connectedness therein.)
Tormey, A. (1971) *The Concept of Expression*, Princeton, NJ: Princeton University Press.(A seminal philosophical analysis, containing a devastating attack on the Expression Theory of music.)
Tormey, A. (1974) ‘Indeterminacy and Identity in Art’, *The Monist* 58: 203-15.(Explores the sort of ontology of music suggested by, or appropriate to, certain modes of avant-garde composition.)
Treitler, L. (1993) ‘History and the Ontology of the Musical Work’, *Journal of Aesthetics and Art Criticism* 51: 483-97.(Explores the relevance of historical scolarship in musicology to questions of musical ontology.)
Walton, K. (1998) *Looking, Listening, Imagining*, Oxford: Oxford University Press.(A collection of essays by a leading aesthetician, some of which bring the author’s doctrine of make-believe to bear on issues about musical understanding and content.)
Wolterstorff, N. (1980) *Works and Worlds of Art*, Oxford: Oxford University Press.(Contains important proposals regarding the ontology and composition of music.)
Zuckerkandl, V. (1956) *Sound and Symbol: Music and the External World*, Princeton, NJ: Princeton University

Press.(A wide-ranging examination of the phenomenon of music, focusing on the paradox of tonal motion and the experience of musical time.)

Zuckerkandl, V. (1973) *Man the Musician*, Princeton, NJ: Princeton University Press.(A sequel to *Sound and Symbol*, focusing on the notion of musicality as exhibited in listening, performing and composing.)

Musonius Rufus (1st century AD)

Gaius Musonius Rufus was a Stoic philosopher who taught in Rome. Active on the margins of political life, he was twice exiled and recalled. His surviving work focuses on practical ethics. Besides his distinctive views on marriage, sexual morality and women's education, Musonius is important for his influence on Epictetus.

Born before AD 30 in Etruria, Musonius was a member of the wealthy equestrian class, not the senatorial aristocracy. He taught in Greek, and his influence in Neronian and Flavian Rome was considerable. After being exiled by Nero for his links to Stoic senators, his fame grew and followers joined him on the island of Gyara. Recalled to Rome at Nero's death, he resumed lecturing and successfully prosecuted the persecutor of a Stoic senator. He was later exiled and recalled again under the Flavian emperors. By the time of his death, probably before AD 100, he had become a symbol of the philosophical life for the Roman aristocracy.

Although Musonius wrote nothing, accounts of many of his lectures were taken down and published after his death. Extracts from twenty-one survive in a late anthology; these and some short quotations and anecdotes are all that remain today. His thought is rooted in the Socratic tradition. Happiness (see Eudaimonia) is the ultimate goal; the human species is defined by its natural rationality and desire for the good. The virtues (see Aretē) include justice, courage, wisdom and self-control (*sōphrosynē*), though emphasis is given to self-control, especially in matters of sex, food and personal adornment. Virtues are understood partly in terms of the traditional craft analogy. Training (*askēsis*) and endurance are emphasized; practical application overshadows theory.

As a Stoic, Musonius held that human nature is essentially rational and that our function is to act in accordance with reason and the providential plan for the world laid down by the gods (see Stoicism). Human reasoning is to be used to support and complement our nature, never to transcend it or transform it. Since women have the same nature as men, they need and deserve the same education. Although their social roles differ, men and women have equal need of philosophy to guide their lives; similarly kings should practise philosophy, though the manual work of farming is held to be the ideal activity for a philosopher.

Musonius advanced the distinctive view that the natural function of sexual activity is only to produce offspring; hence all other sexual relations, heterosexual or homosexual, are improper and reveal a lack of self-control. Even in marriage, sex for pleasure alone is wrong. The goal of marriage is the rearing of many children (abortion and exposure of infants being contrary to nature) and the cultivation of companionship between spouses, which is both a symbol of and the foundation for social relations generally. Though we may suspect the influence of Roman social values on this teaching, Musonius supports his position with arguments based on natural teleology. His central theme is the importance of self-control in the service of a rationally articulated understanding of human nature.

BRAD INWOOD

References and further reading

Arnold, E.V. (1911) *Roman Stoicism*, Cambridge: Cambridge University Press. (Standard work, somewhat dated but still useful.)

Laurenti, R. (1989) 'Musonio, maestro di Epitteto' (Musonius, Master of Epictetus), in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36 3: 2,105-46.(Good survey of Musonius' views with emphasis on his relationship to Epictetus.)

Musonius Rufus (1st century AD) Surviving Writings in O. Hense, *C. Musonius Rufus Reliquiae*, Leipzig: Teubner, 1905; C.E. Lutz, 'Musonius Rufus "The Roman Socrates"', *Yale Classical Studies* 10: 3-147, 1947; A. Jagu, *Musonius Rufus. Entretiens et fragments. Introduction, traduction et commentaire*, Hildesheim and New York: Olms, 1979; I. Andorlini and R. Laurenti, *Corpus dei papiri filosofici Greci e Latini*, Florence: Olschki, 1992, vol. 1, 480-92.(Hense is the standard critical text; Lutz includes the text and a complete English translation of the discourses preserved by Stobaeus, with useful introduction; Jagu offers French translation and commentary; Andorlini and Laurenti edit further fragments found on papyrus.)

Van Geytenbeek, A.C. (1963) *Musonius Rufus and the Greek Diatribe*, Assen: Van Gorcum.(A useful treatment of his life, intellectual affiliations and distinctive views.)

Mystical philosophy in Islam

Mystical philosophy has an intimate connection with the mainstream of Islamic philosophy. It consists of several main strands, ranging from Isma'ili *thought to the metaphysics of al-Ghazali and Ibn al-'Arabi, and with a continuing powerful presence in the contemporary Islamic world. Although mystical thinkers were aware that they were advocating an approach to thinking and knowledge which differed from much of the Peripatetic tradition, they constructed a systematic approach which was often continuous with that tradition. On the whole they emphasized the role of intellectual intuition in our approach to understanding reality, and sought to show how such an understanding might be put on a solid conceptual basis. The ideas that they created were designed to throw light on the nature of the inner sense of Islam.*

1 Mystical philosophy as Islamic philosophy

It is important at the outset to ask what is meant by mystical philosophy in the context of the Islamic philosophical tradition. The term in Arabic closest to the phrase 'mystical philosophy' would perhaps be *al-hikmat al-dhawqiyya*, literally 'tasted philosophy or wisdom', which etymologically corresponds exactly to sapience from the Latin root *sapere*, meaning to taste. As understood in English, however, the term 'mystical philosophy' would include other types of thought in the Islamic context, although *al-hikmat al-dhawqiyya* was at its heart. *Al-hikmat al-dhawqiyya* is usually contrasted with discursive philosophy, or *al-hikmat al-bahthiyya*. Mystical philosophy in Islam would have to include all intellectual perspectives, which consider not only reason but also the heart-intellect, in fact primarily the latter as the main instrument for the gaining of knowledge. If this definition is accepted, then most schools of Islamic philosophy had a mystical element, for there was rarely a rationalistic philosophy developed in Islam which remained impervious to the distinction between reason and the intellect (as *nous* or *intellectus*) and the primacy of the latter while rejecting altogether the role of the heart-intellect in gaining knowledge.

This entry concentrates on those schools which not only include but emphasize *noesis* and the role of the heart-intellect or illumination in the attainment of knowledge. We shall therefore leave aside the Peripatetic school, despite the mystical elements in certain works of al-Farabi, the 'oriental philosophy' of Ibn Sina (Nasr 1996b) and the doctrine of the intellect adopted by the Muslim Peripatetics (*mashsha'un*) in general. Instead, the discussion will concentrate primarily upon the Isma'ili philosophy so closely connected with Hermetic, Pythagorean and Neoplatonic teachings, the school of Illumination (*ishraq*) of al-Suhrawardi and his followers, certain strands of Islamic philosophy in Spain and later Islamic philosophy in Persia and India. However, it would also have to include the doctrinal formulations of Sufism and its metaphysics from al-Ghazali and Ibn al-'Arabi to the present.

2 Isma'ili and Hermetic philosophy

Isma'ili philosophy was among the earliest to be formulated in Islam going back to the *Umm al-kitab (The Mother of Books)* composed in the second century AH (eighth century AD). It expanded in the fourth century AH (tenth century AD) with Abu Hatim al-Razi and Hamid al-Din Kirmani and culminated with Nasir-i Khusraw (Corbin 1993, 1994). By nature this whole philosophical tradition was esoteric in character and identified philosophy itself with the inner, esoteric and therefore mystical dimension of religion. It was concerned with the hermeneutic interpretation (*ta'wil*) of sacred scripture and saw authentic philosophy as a wisdom which issues from the instructions of the Imam (who is identified on a certain level with the heart-intellect), the figure who is able to actualize the potentialities of the human intellect and enable it to gain divine knowledge. The cosmology, psychology and eschatology of Isma'ilism are inextricably connected with its Imamology and the role of the Imam in initiation into the divine mysteries. All the different schools of Isma'ili philosophy, therefore, must be considered as mystical philosophy despite notable distinctions between them, especially, following the downfall of the Fatimids, between the interpretations of those who followed the Yemeni school of Isma'ilism and those who accepted Hasan al-Sabbah and 'The Resurrection of Alamut' in the seventh century AH (thirteenth century AD).

Two of the notable philosophical elements associated with Shi'ism in general and Isma'ilism in particular during the early centuries of Islamic history are Hermetism and Pythagoreanism, the presence of which is already evident in that vast corpus of writings associated with Jabir ibn Hayyan, who was at once alchemist and philosopher. The philosophical dimension of the Jabirian corpus is certainly of a mystical nature, having incorporated much of

Hermeticism into itself, as are later works of Islamic alchemy which in fact acted as channels for the transmission of Hermetic philosophy to the medieval West. When one thinks of the central role of Hermeticism in Western mystical philosophy, one must not forget the immediate Islamic origin of such fundamental texts as the *Emerald Tablet* and the *Turba Philosophorum*, and therefore the significance of such works as texts of Islamic mystical philosophy. Obviously, therefore, one could not speak of Islamic mystical philosophy without mentioning at least the Hermetical texts integrated into Islamic thought by alchemists as well as philosophers and Sufis, and also Hermetic texts written by Muslim authors themselves. It should be recalled in this context in fact that the philosopher Ibn Sina had knowledge of certain Hermetic texts such as *Poimandres* and the Sufi Ibn al-ʻArabi displays vast knowledge of Hermeticism in his *al-Futuhat al-makkiyya (The Meccan Illuminations)* and many other works (Sezgin 1971).

As for Pythagoreanism, although elements of it are seen in the Jabirian corpus, it was primarily in the *Rasa'il* (Epistles) of the Ikhwan al-Safa' in the fourth century AH (tenth century AD), who came from a Shiʻite background and whose work was wholly adopted by later Ismaʻilism, that one sees the full development of an Islamic Pythagoreanism based upon the symbolic and mystical understanding of numbers and geometric forms (Netton 1982) (see Ikhwan al-Safa'). What is called Pythagorean number mysticism in the West had a full development in the Islamic world, and was in fact more easily integrated into the general Islamic intellectual framework than into that of Western Christianity (see Pythagoreanism).

3 Illuminationist philosophy

Perhaps the most enduring and influential school of mystical philosophy in Islam came into being in the sixth century AH (twelfth century AD) with Shihab al-Din al-Suhrawardi, who founded the school of *ishraq* or Illumination. Al-Suhrawardi's basic premise was that knowledge is available to man not through ratiocination alone but also, and above all, through illumination resulting from the purification of one's inner being. He founded a school of philosophy which some have called theosophy in its original sense, that is, mystical philosophy through and through but without being against logic or the use of reason. In fact, al-Suhrawardi criticized Aristotle and the Muslim Peripatetics on logical grounds before setting about expounding the doctrine of *ishraq*. This doctrine is based not on the refutation of logic, but of transcending its categories through an illuminationist knowledge based on immediacy and presence, or what al-Suhrawardi himself called 'knowledge by presence' (*al-ʻilm al-huduri*), in contrast to conceptual knowledge (*al-ʻilm al-husuli*) which is our ordinary method of knowing based on concepts (Ha'iri Yazdi 1992).

In his masterpiece *Hikmat al-ishraq (The Philosophy of Illumination)*, translated by the foremost Western student of al-Suhrawardi, Henry Corbin, as *Le Livre de la Sagesse Orientale (The Book of Oriental Wisdom)*, the Master of Illumination presents an exposition of a form of mystical philosophy which has had a following up to the present day. Based upon the primacy of illumination by the angelic lights as the primary means of attaining authentic knowledge, the school of *ishraq* in fact was instrumental in bestowing a mystical character upon nearly all later Islamic philosophy, which drew even closer to Islamic esotericism or Sufism than in the earlier centuries of Islamic history without ever ceasing to be philosophy. Although the wedding between philosophy and mysticism in Islam is due most of all to the gnostic and sapiential nature of Islamic spirituality itself, on the formal level it is most of all the school of Illumination or *ishraq* which was instrumental in actualizing this wedding, as eight centuries of later Islamic philosophy bears witness (see Illuminationist philosophy).

4 Philosophy in the Maghrib and Spain

The rise of intellectual activity in the Maghrib and, especially, Andalusia was associated from the beginning with an intellectual form of Sufism in which Ibn Masarra was to play a central role. Most of the later Islamic philosophers of this region possessed a mystical dimension, including even the Peripatetics Ibn Bajja and Ibn Tufayl. The former's *Tadbir al-mutawahhid (Regimen of the Solitary)*, far from being a political treatise, deals in reality with man's inner being. Ibn Tufayl's *Hayy ibn Yaqzan (Living Son of the Awake)*, interpreted by many in the West in naturalistic and rationalistic terms, is a symbolic account of the wedding between the partial and universal intellect within the human being, a wedding which results consequently in the confirmation of revelation that is also received through the archangel of revelation, who is none other than the objective embodiment of the universal intellect. Moreover, this mystical tendency is to be seen in its fullness in less well-known figures such as Ibn al-Sid of Badajoz who, like the Ikhwan al-Safa', was devoted to mathematical mysticism, and especially the

Sufi Ibn Sab'in, the last of the Andalusian philosophers of the seventh century AH (thirteenth century AD), who developed one of the most extreme forms of mystical philosophy in Islam based upon the doctrine of the transcendent unity of being (*wahdat al-wujud*) (Taftazani and Leaman 1996). Andalusia was also the home of the greatest expositor of Sufi metaphysics, Ibn al-'Arabi (see §6).

5 Illuminationist thought in the East

In eastern lands of the Islamic world and especially Persia, which was the main theatre for the flourishing of Islamic philosophy from the seventh century AH (thirteenth century AD) onward, primarily mystical philosophy was dominant during later centuries despite the revival of the discursive philosophy of the *mashsha'i*s, such as Ibn Sina, by Khwajah Nasir al-Din al-Tusi and others. It was in the East in the seventh and eighth centuries AH (thirteenth and fourteenth centuries AD) that the doctrines of *ishraq* with its emphasis on inner vision and illumination were revived by al-Suhrawardi's major commentators, Shams al-Din al-Shahrazuri and Qutb al-Din al-Shirazi, who was also a master of Ibn Sinan philosophy. The next three centuries saw mystical ideas and doctrines become ever more combined with the philosophical theses of the earlier schools, and figures such as Ibn Turkah Isfahani sought consciously to combine the teachings of Ibn Sina, al-Suhrawardi and Ibn al-'Arabi.

This tendency culminated in the tenth century AH (sixteenth century AD) with the establishment of the School of Isfahan by Mir Damad and the foremost metaphysician of later Islamic thought, Mulla Sadra, in whom the blending of ratiocination, inner illumination and revelation became complete (Corbin 1972). In this school the most rigorous logical discourse is combined with illumination and direct experience of ultimate reality, as seen so amply in Mulla Sadra's masterpiece *al-Asfar al-arba'ah (The Four Journeys)*. This later Islamic philosophy is certainly mystical philosophy, relying as it does on 'experiential' knowledge and direct vision of ultimate reality and the angelic worlds, a vision that is associated with the eye of the heart (*'ayn al-qalb* or*chism-i dil*). However, it is also a philosophy in which the categories of logic are themselves seen as ladders for ascent to the world of numinous reality in accordance with the Islamic perspective, in which what would be called Islamic mysticism from a Christian perspective is of a gnostic (*'irfani*) and sapiental nature, Islamic mysticism being essentially a path of knowledge of which love is the consort, rather than a way of love exclusive of knowledge.

In any case it was this type of philosophy, associated especially with the name of Mulla Sadra, that has dominated the philosophical scene in Persia during the past few centuries and produced major figures such as Hajji Mulla Hadi al-Sabzawari and Mulla 'Ali Zunuzi in the thirteenth century AH (nineteenth century AD), both of whom were philosophers as well as mystics. It is also this type of philosophy that continues to this day and has in fact been revived during the past few decades. Nearly all philosophers in Persia associated with the school of Mulla Sadra, which is also known as *al-hikmat al-muta'aliya* (literally the 'transcendent theosophy'), have been and remain at once philosophers and mystics.

In India likewise, Islamic philosophy began to spread only after al-Suhrawardi and during the past seven centuries most Islamic philosophers in that land have been also what in the West would be called mystics. It is not accidental that the school of Mulla Sadra spread rapidly after him in India and has had expositors there to this day. Perhaps the most famous of Muslim intellectual figures in India, Shah Waliullah of Delhi, exemplifies this reality (see Shah Wali Allah). He was a philosopher and Sufi as well as a theologian, and his many writings attest to the blending of philosophy and mysticism. It can in fact be said that Islamic philosophy in India is essentially mystical philosophy, despite the attention paid by the Islamic philosophers there to logic and in some cases to natural philosophy and medicine.

6 Sufism and the Akbarian tradition

No treatment of mystical philosophy in Islam would be complete without a discussion of doctrinal Sufism and Sufi metaphysics, although technically speaking in Islamic civilization a clear distinction has always been made between philosophy (*al-falsafa* or *al-hikma*) and Sufi metaphysics and gnosis (*al-ma'rifah*, *'irfan*). However, as the term 'mystical philosophy' is understood in English, it would certainly include Sufi metaphysical and cosmological doctrines which were not explicitly formulated until the sixth and seventh centuries AH (twelfth and thirteenth centuries AD) although their roots are to be found in the Qur'an and hadith and the sayings and writings of the early Sufis. The first Sufi authors who turned to an explicit formulation of Sufi metaphysical doctrines were Abu Hamid Muhammad al-Ghazali in his later esoteric treatise such as *Mishkat al-anwar (The Niche of Lights)*

and *al-Risalat al-laduniyya (Treatise on Divine Knowledge)*, and ‘Ayn al-Qudat Hamadani who followed a generation after him.

The writings of these great masters were, however, a prelude for the vast expositions of the master of Islamic gnosis Muhyi al-Din Ibn al-‘Arabi, perhaps the most influential Islamic intellectual figure of the past seven hundred years. Not only did he profoundly influence many currents of Sufism and establish an ‘Akbarian tradition’ identified with such later masters as Sadr al-Din Qunawi, ‘Abd al-Rahman Jami and, in the last century, Amir ‘Abd al-Qadir and Shaykh Ahmad al-‘Alawi. He and his school also influenced formal philosophy to such an extent that a figure such as Mulla Sadra would not be conceivable without him. The Ibn al-‘Arabian doctrines of the transcendent unity of being, the universal man, the imaginal world and eschatological realities are not only esoteric and mystical doctrines of the greatest significance in themselves for the understanding of the inner teachings of Islam, but are also sources of philosophical meditation for generations of Islamic philosophers to the present day, who have cultivated diverse and rich schools of mystical philosophy during the past eight centuries and brought into being currents of philosophical thought that are still alive in the Islamic world. One need only think of such fourteenth century AH (twentieth century AD) figures as ‘Alalamah Tabataba’i in Persia and ‘Abd al-Halim Mahmud in Egypt to realize the significance of the wedding between philosophy and mysticism in the Islamic intellectual tradition, not only over the ages, but as part of the contemporary Islamic intellectual scene (see Islamic philosophy, modern).

See also: Gnosticism; Ibn al-‘Arabi; Illuminationist philosophy; Mysticism, history of; Mysticism, nature and assessment of; al-Suhrawardi

SEYYED HOSSEIN NASR

References and further reading

Chittick, W. (1989) *The Sufi Path of Knowledge*, Albany, NY: State University of New York Press.(The standard account of the nature of mystical knowledge.)

Chittick, W. (1994) *Imaginal Worlds: Ibn al-‘Arabi and the Problem of Religious Diversity*, Albany, NY: State University of New York Press.(An analysis of the concept of the *mundus imaginalis*.)

Chodkiewicz, M. (1993) *Seal of the Saints - Prophethood and Sainthood in the Doctrine of Ibn ‘Arabi*, trans. L. Sherrard, Cambridge: Islamic Texts Society.(Close account of the key concepts of prophecy and sainthood.)

Corbin, H. (1972) *En Islam iranien (On Persian Islam)* Paris: Gallimard.(The most important collection of sources of Persian philosophy.)

Corbin, H. (1980) *Avicenna and the Visionary Recital*, trans. W. Trask, Houston, TX: Spring Publications.(Ibn Sina’s account of mystical perception.)

Corbin, H. (1993) *The History of Islamic Philosophy*, in collaboration with S.H. Nasr and O. Yahya, trans. P. Sherrard, London: Kegan Paul International.(The first history to lay proper emphasis on Persian philosophy.)

Corbin, H. (1994) *Trilogie ismaélienne (Isma‘ili Trilogy)*, Paris: Verdier.(Discussion of some of the most important Isma‘ili texts.)

Cruz Hernández, M. (1981) *Historia del pensamiento en el mundo islámico (History of Thought in the Islamic World)*, Madrid: Alianza Editorial.(Excellent general account of Islamic philosophy.)

Ha’iri Yazdi, M. (1992) *The Principles of Epistemology in Islamic Philosophy - Knowledge by Presence*, Albany, NY: State University of New York Press.(The best account of *‘ilm al-huduri*, knowledge by presence.)

Knysh, A. (1993) ‘The Diffusion of Ibn ‘Arabi’s Doctrine’, in S. Hirtenstein and M. Tiernan (eds) *Muhyiddin ibn ‘Arabi - A Commemorative Volume*, Shaftesbury: Element, 307-27.(Discussion of the influence of Ibn al-‘Arabi.)

Nanji, A. (1996) ‘Isma‘ili Philosophy’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 9, 144-54.(Examination of Isma‘ili philosophy including the influence of Neoplatonism.)

Nasr, S.H. (1975) *Three Muslim Sages*, New York: Delmar.(Excellent introductions to Ibn Sina, al-Suhrawardi and Ibn al-‘Arabi.)

Nasr, S.H. (1978) *Islamic Life and Thought*, Albany, NY: State University of New York Press. (General introduction to the role of mysticism in Islamic culture.)

Nasr, S.H. (1996a) ‘Ibn Sina’s Oriental Philosophy’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 247-51.(Argument for the existence and importance of the ‘oriental philosophy’.)

Nasr, S.H. (1996b) *The Islamic Intellectual Tradition in Persia*, Richmond: Curzon Press.(Deals with the Persian contribution to philosophy and mysticism.)

Netton, I. (1982) *Muslim Neoplatonists: An Introduction to the Thought of the Brethren of Purity*, London: Allen & Unwin.(The standard account of the Ikhwan al-Safa'.)
Sezgin, F. (1971) *Geschichte des arabischen Schrifttums (History of Arabic Literature)*, vol. 4, Leiden: Brill. (Sources on Hermetism in Islamic literature.)
al-Suhrawardi (1154-91) *Hikmat al-ishraq (The Philosophy of Illumination)*, trans H. Corbin, *Le livre de la sagesse orientale*, Paris: Verdier, 1986.(Very important illuminationist text.)
Taftazani, A. and Leaman, O. (1996) 'Ibn Sab'in', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 346-9.(Discussion of the significance of the thought of Ibn Sab'in.)
Ziai, H. (1990) *Knowledge and Illumination*, Atlanta, GA: Scholars Press.(Very clear account of the links between illuminationist philosophy and epistemology.)

Mysticism, history of

Contemporary authors generally associate mysticism with a form of consciousness involving an apparent encounter or union with an ultimate order of reality, however this is understood. Mysticism in this sense, it is argued, can be found in virtually all cultures and religious traditions, and is perhaps as old as humanity itself. None the less, there is no agreement on the identifying characteristics of mystical states; the term 'mysticism' and its cognates have undergone long evolution and been used in a bewildering variety of ways.

Such ongoing disputes about the nature and significance of mysticism only underscore both the challenge and the importance of studying its history. On the one hand, without consensus on a definition, scholars disagree on which texts and figures merit inclusion in a historical survey of mysticism. On the other hand, arguments about whether mystical experiences are 'everywhere the same' can hardly be settled apart from attention to the historical evidence.

1 Methodological issues

Increasingly, scholars debate not only the data but also the proper methodology for studying mysticism's history. Lack of consensus on a definition of mysticism gives rise to corresponding disagreements over who and what to classify as truly mystical, a problem compounded by the modern tendency to identify mysticism exclusively with certain subjective experiences involving 'pure undifferentiated consciousness' or 'union'. We have no direct access, after all, to the states of consciousness of past mystics, but only to the texts and other artefacts they have left behind, from which it is generally difficult to extract clear 'phenomenological' descriptions of the sort modern commentators seek.

Another problem is that if, according to some definitions, mysticism is a fundamental mode of human consciousness present in every era, or even a tacit dimension of all human experience, a comprehensive history of mysticism would have to include all human history and experience, something obviously impossible. Still, we may suppose that for every mystic recognized today, hundreds more have left behind no traces (for example, because they lived in pre-literate communities, or their works were lost, or they felt no urge to record their experiences). Thus any account of the history of mysticism has an essentially modest scope, since it illuminates only a small portion of the totality of human mystical experience, most of which will remain forever inaccessible. At the same time, however, modern scholarship is opening up a wealth of new material that has been forgotten or under-utilized in the tradition. Feminists, for example, are retrieving the testimony of women mystics, noting what is distinctive about their experience. Such research continues to enlarge and modify our understanding of mysticism's history.

The present entry briefly traces several strands of mysticism according to the religious traditions out of which they emerged. No claim is made that this is the best or most complete approach; mystical experiences may occur to people of no formal religious affiliation. Nevertheless, even the most unorthodox of the great mystics generally have their roots in a specific religious tradition, and cannot be adequately understood apart from it.

2 Mysticism of ancient and indigenous communities

So-called 'extrovertive' or 'nature' mysticism, involving an experience of a sacred unity in and with all things, is a relatively common human phenomenon, even where evidence for other forms of mystical experience is wanting (an example here is the alleged 'mystical' quality of many prehistoric cave paintings or ancient carvings). The spiritualities of many archaic cultures and indigenous peoples involve the celebration (and sometimes restoration) of a communal and individual sense of 'oneness' with the whole created order.

Eliade (1964) and others discern a mystical element in shamanism, and in the 'spirit possession' of African mediums. Native American peoples of the Central Plains region retain the tradition of the 'vision quest', in which seekers undertake an ascetic discipline to encounter their 'spirit guide'. Though the results of these practices may not be an experience of 'undifferentiated unity', the processes involved and the ecstatic states attained bear certain affinities to the mystical disciplines of more familiar traditions.

3 Indian mysticism

Among the earliest recorded manifestations of mysticism are those from the Indian subcontinent. Even before the second millennium BC, the pre-Aryan civilization apparently practised a form of yoga. The polytheistic religion of the invading Aryans centred on ritual sacrifice, but gradually evolved in a mystical direction as its practitioners reflected upon the inner significance of the sacrifices, and Brahman (originally identified with the sacred power contained in the rites) came increasingly to be understood as a universal principle or power underlying all reality.

Out of this context, and over many centuries, emerged the Vedas, a fluid canon of sacred writings accepted as revelation (*śruti*) by orthodox Hindus; among these the Upaniṣads propose a path of mystical insight allowing liberation (*mokṣa* or *mukti*) from the vicissitudes of human existence. The Upaniṣads are concerned with the relationship between Brahman (the Supreme Soul or Absolute) and *ātman* (the eternal inner soul or self), a relationship summarized in the classic formulation *tat tvam asi*, 'thou art that'; in other words, the eternal self, realized through meditation and asceticism, is ultimately one with the Absolute (see Brahman).

The Upaniṣads themselves, however, incorporate a variety of tendencies from the preceding traditions. The Śvetāśvatara Upaniṣad, for example, contains an unusually strong theistic strain, and speaks of the path to union with the Supreme Self through devotion (*bhakti*); other Upaniṣads seem to support a strict identity between *ātman* and Brahman. The Māṇḍūkya Upaniṣad, often cited in Western philosophical discussions of mystical experience, speaks of a state of consciousness beyond dreamless sleep: 'unperceived, unrelated, incomprehensible, uninferable, unthinkable, and indescribable', 'all peace, all bliss, and nondual', 'this is *ātman*, and this has to be realized' (Māṇḍūkya Upaniṣad, 7).

Other movements emerging between 800 and 500 BC developed approaches to mysticism that were atheistic or agnostic in their implications. Like the Sāōkhya philosophy associated with classical yoga, Jainism maintains the ultimate status not of a single divine monad but of an infinity of individual souls; the goal of the mystical journey is to isolate this self-monad from its immersion in materiality. Early Buddhism, by contrast, developed the doctrine of 'no self' (*anatta*) and impermanence (*anicca*), analysing persons as a series of transitory states and denying the existence of any underlying eternal self. According to Buddhist tradition, Prince Gautama Siddhartha (born perhaps in the latter half of the sixth century BC) left home and family to seek the cause of human suffering and the way to liberation from it. After years of harsh asceticism and searching, he discovered the Middle Path (between sensuality and excessive austerity) and, while meditating beneath a *bo* (or *bodhi*) tree, attained *nirvāṇa*, that is, supreme peace and enlightenment (*bodhi*) - hence the title 'Buddha' ('Enlightened One'). The Buddha taught his disciples that by following the Noble Eightfold Path and attaining *nirvāṇa*, the cycle of rebirth, the illusion of selfhood, and the suffering caused by ignorance and craving can be overcome. In succeeding centuries, as it flourished and spread to other parts of Asia, Buddhism underwent extensive scholastic systematization (for example, in the dialectical method developed by Nāgārjuna in the second century) and evolved into several schools, most notably Theravāda (probably closer to the Buddha's original doctrine and found today in Sri Lanka, Myanmar and southeast Asia) and Mahāyāna (in China, Japan and Korea).

Crucial to the development of later Indian mysticism is the *Bhagavad Gītā*, a famous mystical poem (in the form of a dialogue between Lord Arjuna and Krishna) which presents three paths to salvation - the way of knowledge, the way of works, and the way of devotion (*bhakti*) or loving adoration - and which (on some interpretations) seems to give a higher place to *bhakti* than to contemplative yoga. Since Krishna is a manifestation of Vishnu, the poem clearly contains a theistic element, but there are monistic strains as well.

Such diverse tendencies eventually gave rise in the medieval period to three main schools of Vedānta (the 'end' or systematic explanation of the Veda) (see Vedānta). Śaōkara gave the Upanishadic formula 'thou art that' its most radical nondualistic (*advaita*) interpretation, insisting on a strict numerical identity between the soul and Brahman; though allowing for worship and devotion on the level of appearances, he believed these to be transcended in union. Against this view, the eleventh-century thinker Rāmānuja developed a position of 'qualified nondualism'; he recognized an impersonal mysticism of undifferentiated unity, but ranked it inferior to loving communion with a personal God. Two centuries later, Madhva propounded a dualistic (*dvaita*) system of thought. He argued for a radical pluralism among God (*īśvara*), individual souls and non-intelligent substances, and maintained that it is God who decides the destinies of all finite selves.

More recently, Aurobindo Ghose has attempted to combine elements of traditional Hindu theology with an evolutionary mysticism. But perhaps the best-known modern representative of Vedāntic spirituality is

Ramakrishna (1834-86), whose life and message of universalism were popularized in the West by Swami Vivekananda (1862-1902), organizer of the Ramakrishna Mission, which combined Advaita philosophy with a deep concern for social issues (see Ramakrishna Movement).

4 Mysticism in China and Japan

Despite its influence on later Hinduism, Buddhism largely died out within India itself after the revival of Vedantism. Elsewhere in Asia, however, Buddhism continued to spread and evolve as it encountered other traditions.

With its focus on harmonious relations between the 'superior individual' and a well-ordered society, early Confucianism appears as a form of ethical humanism, though a mystical element emerges more clearly in the *Book of Mencius* and in later neo-Confucianism, under Daoist influence (see Chinese philosophy §§4-5). By contrast, the *Daodejing* (ascribed to Laozi, considered the founder of Daoism) has a more mystical tone. Here the Dao or Way is regarded as both the natural order of the universe, and a life lived in harmony with that order. Laozi advises the sage to act spontaneously and effortlessly, through non-action (*wuwei*). The *Daodejing* is relatively silent on mystical disciplines or states of consciousness, which are treated in more detail in the later *Book of Zhuangzi* (see Daodejing).

Daoism exerted a powerful influence on *Chan* (Japanese *Zen*) Buddhism, a school of Mahāyāna Buddhism developed in China and Japan (see Japanese philosophy §5). Here there is a strong emphasis on sudden illumination (*satori*), often after prolonged sitting meditation (*zazen*), concentrating under the direction of a master (*roshi*) on an assigned *koan* (a paradoxical question or answer designed to stop ordinary conceptual thinking). The two principle surviving Zen schools are Rinzai and Sōtō, brought to Japan by Eisai (1141-1215) and Dōgen respectively. Dōgen and the Sōtōschool emphasize the centrality of the Zen meditation practice itself, not as a means to an end, but as the realization of one's own already present Buddha-nature. By contrast, *roshis* of the Rinzai school have traditionally used *koans* and even startling and eccentric behaviour (slaps, sudden gestures, and so on) to evoke in their pupils a condition of great doubt and mental tension leading to the sudden breakthrough of *satori*. Revitalized in the eighteenth century by Hakuin (1685-1768), who systematized the use of *koans* and left detailed descriptions of his own mystical experiences, Rinzai has become the best-known branch of Zen in the West, largely through the efforts of D.T. Suzuki (1870-1966).

5 Jewish mysticism

Though the Hebrew Scriptures describe many encounters with the God of Israel (for example, Moses before the burning bush, Elijah at Horeb, Isaiah's Temple vision), encounters often treated as paradigmatic in later Jewish and Christian mystical texts, many modern scholars prefer to classify these as numinous states, since they seem to involve the overwhelming sense of a sacred presence external to the subject, rather than an absorption into the divine. In the post-exilic period, Jewish spirituality became increasingly eschatological, even apocalyptic, giving rise in the first century BC to *merkavah* (chariot) mysticism, involving Gnostic-influenced mystical speculations on Ezekiel's vision of the throne-chariot (Ezekiel 1), and the possibility of ascending through various spheres to the divine throne. Though *merkavah* mysticism declined in the Middle Ages, it left its imprint on the more popular movement of medieval Hasidism, which through prayer and religious practices cultivated intense awareness of the omnipresent creator (see Hasidism).

The first-century philosopher Philo of Alexandria (§§1, 4), so deeply influenced by Greek thought, is usually ranked among the Hellenistic philosophers rather than within the main currents of Jewish spirituality. Nevertheless, he encourages pursuit of the direct vision of God, and develops a highly allegorical method for discerning mystical meanings in scriptural texts.

The Jewish spiritual movement known as Kabbalah found classic expression in the *Zohar* (Splendour), a massive and esoteric work claiming great antiquity but more probably dating from the thirteenth century. According to its complex speculations, echoing Gnostic and Neoplatonic themes, God is conceived as the Eyn Sof (Infinite), without qualities, but emanating ten ideal qualities, or Sefirot, the lowest of which includes the created order. Creation occurs within God, through a divinely initiated self-differentiation, and each soul contains some of the Sefirot. But since Adam's Fall, the *Shekinah*, or divine presence, has been exiled from the Eyn Sof, and so the goal

of the devout is to become aware of the divine presence within and thus help reunite the *Shekinah* with the Eyn Sof, restoring the cosmic order. Isaac Luria (1534-72), a major figure of this mystical school, conceived Adam as a cosmic figure (Adam Kadmon) whose Fall brought about the shattering of the ideal universe into the present material world, with the divine light broken up into the many sparks of individual souls; through prayer and concentrated devotion (*kevannah*) human beings could take an active role in the ultimate reintegration (*tikkun*) of all things (see Kabbalah).

More popular than speculative, modern Hasidism originated in the eighteenth century with Israel Baal Shem Tov (1700-60) and his successor Baer of Meseritz (1710-72). This mystical and devotional movement emphasizes especially the role of the Tzaddik, or perfectly righteous person, who alone can guide others effectively in their spiritual journey. Hasidism remains a strong force within contemporary Judaism.

6 Christian mysticism

Christianity rests upon Jesus of Nazareth, and his consciousness of a deep continual communion with the God of Israel, whom he boldly addresses as ‘Abba’ (‘dear Father’). This so-called ‘Abba-experience’ has been compared with the experiences of mystics, though Jesus himself (according to the evidence of the Gospels) seems to have claimed an utterly unique and unprecedented intimacy with the divine (Christian dogma identifies him as the sole incarnation of the Second Person of the Trinity) (see Incarnation and Christology §1).

Strikingly, though states arguably ‘mystical’ in the contemporary sense are reported among Jesus’ first followers, the word *mystikos* itself appears nowhere in the New Testament; on the other hand, ‘mystery’ (*mysterion*) figures prominently in the Pauline letters. Past historians of religion detected here the influence of ancient mystery cults, and claimed that Christian mysticism was largely imported. Louis Bouyer (1990) and others have argued, however, that while the secrecy of the mystery religions pertained only to their rituals, St Paul uses ‘mystery’ in a different sense, Semitic in origin, to refer to God’s hidden plan for the salvation of the world, now revealed in Christ.

None the less, Hellenistic influences on the subsequent evolution of Christian spirituality and mystical theology are undeniable, beginning with ancient Alexandria, where Plotinus (founder of Neoplatonism) and Origen (Christianity’s first great theologian) shared the same teacher, Ammonius Saccas. The views of Plotinus (§§3, 6) on the immanent presence within the intelligible and material order of the divine One from which they have emerged, and his goal of a mystical return to the One through asceticism and contemplation, struck a responsive chord with Augustine and many others. Origen (§2) was especially concerned with discovering the ‘mystical sense’ of Scripture, by which he understood the hidden presence of Christ within the biblical text. Later Christian authors of the first four centuries extended the scope of *mystikos* to Christ’s hidden presence in the sacraments, particularly the Eucharist.

With the conversion of Constantine, the early Christian ideal of martyrdom was recast in terms of austere self-sacrifice; the Desert Fathers and Mothers, as well as the major figures of the monastic movement, such as Anthony the Great (*c*.251-356), Pachomius (*c*.290-346), Basil the Great (*c*.330-79), John Cassian (*c*.360-435) and Benedict (480-543), provided guidance for a life dedicated to prayer, asceticism and the search for God.

In terms showing thorough familiarity with Neoplatonic thought, the mystical implications of God’s incomprehensibility were explored by Gregory of Nyssa (*c*.335-95), and even more thoroughly by the (apparently) late-fifth-century Syrian monk now known as Pseudo-Dionysius, so called because his writings purported to be from the Dionysius converted by Paul in Athens (Acts 17: 34). The presumed apostolic pedigree gave the Dionysian texts enormous authority throughout the Middle Ages and into modern times. In *The Mystical Theology*, this author proposes a *via negativa* or ‘apophatic’ ascent into the ‘dazzling darkness’ of the ‘superessential’ divinity by successively rejecting the applicability of all predication to it; elsewhere he outlines a corresponding *via affirmativa* or ‘kataphatic’ approach, attempting to explain in what sense God may be called perfect, good, omnipotent, and so on.

Later mystics of the Eastern Church, such as John of Damascus (*c*.674-749) and especially Simeon the New Theologian (949-1022), stressed the role of the divine Light in the divinization of the human person, restoring the divine image obscured by Adam’s Fall. Simeon in turn influenced the development of hesychasm (from *hesychia*, ‘quietness’), involving the constant repetition of the mantra-like Jesus Prayer (‘Lord Jesus Christ, Son of God,

have mercy on me, a sinner') in pursuit of the vision of divine Light. Hesychasm was successfully defended by Gregory Palamas (*c.*1296-1359), who identified the divinizing Light with the uncreated 'energies' of God.

In the vastly influential works of Augustine (§11) comes a shift towards a more psychological approach to mystical themes. Augustine traces the image of the Trinity in the faculties of the soul (intellect, memory and will), and shows interest in the differences between various types of visionary and ecstatic phenomena. The Christocentric love-mysticism of the Cistercian Bernard of Clairvaux (1090-1153) characterizes loving union with the Incarnate Word in affective rather than metaphysical terms, not as a 'union of essences' but a 'concurrence of wills', leading none the less to an experience of apparent absorption in God. Meanwhile, the mysticism of the renowned Benedictine abbess Hildegard of Bingen (1098-1179), as well as the Cistercian nuns Mechthild of Magdeburg (*c.*1210-80), Gertrude the Great (*c.*1256-1302) and Mechthild of Hackeborn (1240-98), involved numerous visionary experiences and revelatory messages. Francis of Assisi (d. 1226) and his Franciscan followers (for example, Clare of Assisi, Jacopone da Todi, Angela of Foligno, and Bonaventure) ushered in a new era of intense devotion to the humanity of Christ, with a correspondingly higher appreciation of the created order in its own right as intrinsically worthy of holy human love because the object of God's love. Mystical themes were sometimes clothed in narrative form in the popular (if often fanciful) medieval *vitae* of Christian saints. Jacobus de Voragine's *Legenda aurea (Golden Legend)*, one of the most famous books of the Middle Ages, recounts the life of the noble and learned Catherine of Alexandria, regarded as patroness of scholars and philosophers because she was allegedly martyred in the early fourth century after converting the family of Maxentius and defeating by her wise argumentation fifty philosophical opponents handpicked by the Emperor. Despite the lack of any early evidence that she ever existed, medieval accounts of her mystical marriage to Christ helped popularize her cult.

The most significant figure among Rheno-Flemish mystics of the thirteenth and fourteenth centuries is the brilliant Dominican theologian and preacher Meister Eckhart, whose more startling and paradoxical expressions in vernacular sermons brought accusations of heresy. His mystical writings are still subject to the most divergent interpretations, and parallels with Hindu and Buddhist mysticism are often noted. For Eckhart, only God truly is, since being in the strictest sense is God alone. Human creatures possess existence only 'in and through God', who is immanent in the soul's uncreated apex or ground, which is eternally one with the divine. The mystical goal is therefore to divest oneself of everything, so that God may bring the Word to birth within the soul. Eckhart likewise speaks of a Godhead beyond God, a completely unknowable and indescribable 'Ground' beyond the distinction into a Trinity of Divine Persons. Despite the official condemnations, Eckhart's influence continued in John Tauler (*c.*1300-61), Henry Suso (*c.*1295-1366), and the Friends of God movement, whose teachings were crystallized in the anonymous mystical classic, the *Theologica Germanica*.

For Jan van Ruysbroeck (John Ruusbroec, 1293-1381), the empty desert of the Godhead itself has a Trinitarian orientation, and mystics who attain union with this nameless unity are not static, but move out with the Father into the world created in the divine image. Catherine of Siena (1347-80) provides an impressive example of the apostolic force of Christian love mysticism, since her 'mystical marriage with Christ' bore fruit in a remarkable involvement with contemporary social needs, as did the mystical experiences of Catherine of Genoa (1447-1510). The fourteenth-century English mystics - among them Richard Rolle (*c.*1300-49), Walter Hilton (d. 1395), Julian of Norwich (*c.*1342-after 1413) and the author of *The Cloud of Unknowing* - offer an appealing and approachable spirituality which has found a wide audience in our own times.

The famed *Spiritual Exercises* of St Ignatius of Loyola (*c.*1491-1556) appear at first glance to have little mystical about them, but are designed to lead to a direct encounter with God. The Spanish Carmelites Teresa of Avila (1515-81) and John of the Cross (1545-91) were declared Doctors of the Church for their spiritual teaching. John is best known for his classic analysis, in *The Dark Night of the Soul* and other works, of the passive purifications endured in the search for God, though his descriptions of the joys of mystical union are even more eloquent. In various writings Teresa sought to identify, through their spiritual and psychological effects, the various stages and pitfalls in the spiritual quest; her masterpiece, *The Interior Castle*, describes the process as a journey inwards through seven progressively more interior dwelling places, leading finally to a permanent union with the Triune God already present in the soul's centre.

Francis de Sales (1567-1622), a key figure in the seventeenth-century French school of spirituality, described how a 'devout life' could be lived in any state, without requiring withdrawal from the world. Pierre de Bérulle

(1575-1629) offered a theocentric and Christocentric spirituality, directed towards conformity with the 'states' or inner dispositions of the Incarnate Word. Madame Jeanne Guyon (1648-1717) advocated quiet prayer to the exclusion of all other methods, while her disciple and director, François Fénelon (1651-1715), defended a mysticism of 'pure love' of God, free of all self-interest, even in one's own salvation; critics charged that such teachings opened the door to complete moral and religious indifferentism. Within Catholic circles, the condemnation of quietism created a distrust of mystical spirituality that persisted into the twentieth century, while evoking an interest in the suspect doctrines and authors among many Protestants.

Ironically, despite its emphasis on the experience of personal salvation, the Protestant tradition has often regarded mysticism (especially in its Roman Catholic expressions) with mistrust, as a form of 'works righteousness', or a misguided attempt to bypass the mediatorial role of Jesus. Nevertheless, Protestantism has had its own mystics, such as Jakob Boehme (1575-1624), George Fox (1624-91), William Law (1686-1761) and John Woolman (1720-72), and certain Protestant movements, such as Quakerism and pietism, have a decidedly mystical flavour.

7 Islamic mysticism

Scholars who distinguish sharply between mystical and prophetic experiences do not usually classify the prophet Muhammad as a mystic, although his experiences and certain texts in the Qur'an have lent themselves to later mystical interpretations. But Islamic mysticism seems to have first emerged on a broader scale as part of a reaction against the worldliness of the expanding Muslim empire, stressing deep personal devotion and an ascetic lifestyle in the quest for inner illumination and loving communion with God. The term 'Sufism', by which Islamic mysticism is generally known, may itself have been inspired by this lifestyle, since *suf* refers to the undyed wool worn as a sign of simplicity.

Rabi'ah (d. 801) 'is generally regarded as the person who introduced the element of selfless love into the austere teachings of the early ascetics and gave Sufism the hue of true mysticism' (Schimmel 1975: 38). Subsequent Sufi mysticism is love mysticism, and uses much of the same love imagery found in the Christian mystical tradition. The search for *marifa* (direct knowledge of God), however, sometimes led to seemingly unorthodox results. Abu Yazid (d. 875) introduced the notion of *fana*, the passing away of the empirical self, a crucial and sometimes controversial theme in later Sufi mysticism; his famous words 'Glory to me, how great is my majesty!' seemed to make a blasphemous claim to divinity. Al-Hallaj (854-922) was in fact put to death for blasphemy, after seeming to assert identity with God. Al-Ghazali (§4) was able to reconcile Sufism with orthodoxy by explaining the mystics' sense of identity with God as a kind of passing illusion brought on by the intensity of the experience.

The pantheistic Sufism of Ibn al-'Arabi (§5), though unorthodox, exerted a strong influence on the mystical poet Rumi (1207-73), who founded the Mevlevi *darwish* order, whose members engage in ritual swirling movements to induce states of intense religious devotion (the so-called 'dervish dances'). After a long decline, Sufism has enjoyed a resurgence in the West in recent years (see Mystical philosophy in Islam).

See also: Bonaventure; Hasidism; Mysticism, nature of; Pantheism; Religious experience; Religious pluralism

STEVEN PAYNE

References and further reading

Most of the Western mystical authors and works mentioned in this entry (including the *Mystical Theology* of Pseudo-Dionysus, the *Zohar*, the Ignatian *Spiritual Exercises*, and St Teresa's *Interior Castle*) can be found in 'Classics of Western Spirituality', an ongoing series from Paulist Press. Readers are also recommended to refer to Crossroad's *World Spirituality: An Encyclopedic History of the Religious Quest*, a 25-volume series on all the major traditions of spirituality, including their mystical elements.

Bhagavad Gītā (200 BC -AD200, disputed), trans. D. White, *The Bhagavad Gītā*, American University Studies, Series 7, Theology and Religion, vol. 39, New York: Peter Lang, 1989, 1993. (Also includes a commentary on the text.)

Bouyer, L. (1980) 'Mysticism: An Essay on the History of the Word', in R. Woods (ed.) *Understanding Mysticism*, Garden City, NY: Image Books.(Outlines evolution of the term 'mystical' from Greek origins through patristic era. Translation of 1952 article from *La Vie Spirituelle*.)

Bouyer, L. (1990) *The Christian Mystery: From Pagan Myth to Christian Mysticism*, Edinburgh: T. & T. Clark. (Reviews the development of Christian mysticism, arguing against its origin in mystery religions.)

Bouyer, L. *et al.* (1969) *A History of Christian Spirituality*, New York: Seabury, 3 vols.(A classic survey of Christian spirituality.)

Certeau, M. de (1992) *The Mystic Fable*, vol. 1: *The Sixteenth and Seventeenth Centuries*, Chicago, IL: University of Chicago Press.(Difficult but rewarding discussion of the term 'la mystique' and the fundamental change it represents in Western approaches to the sacred. Incorporates semiotics, linguistics and psychoanalytic theory.)

Daodejing (*c*.350-250 BC), trans. D.C. Lau, *Tao Te Ching*, Baltimore, MD: Penguin, 1963.(This popular work is traditionally - although now controversially - ascribed to the sage Laozi.)

Dasgupta, S. (1927) *Hindu Mysticism*, New York: Unger.(Dated but useful overview of the subject, first presented as six lectures at Northwestern University.)

Dumoulin, H. (1988-90) *Zen Buddhism: A History*, New York: Macmillan, 2 vols. (Widely regarded as the standard work on the history of Zen, now greatly amplified and updated.)

Dupré, L. (1987) 'Mysticism', in M. Eliade (ed.) *The Encyclopedia of Religion*, New York: Macmillan.(Useful survey of the nature and history of mysticism, organized according to its own distinctive typology.)

Eliade, M. (1964) *Shamanism*, London: Routledge.(Shows links between elements of shamanism and mysticism.)

Freemantle, A. (1964) *The Protestant Mystics*, New York: New American Library.(A study of an important and neglected topic, but includes figures not ordinarily considered 'mystical'.)

Inge, W.R. (1899) *Christian Mysticism*, London: Methuen.(Classic early survey, with famous appendix listing twenty-six definitions of 'mysticism' and 'mystical theology'. Marred by author's biases.)

Katz, S.T. (ed.) (1978) *Mysticism and Philosophical Analysis*, New York: Oxford University Press.(Important articles by Katz, Peter Moore and others criticizing the view that mysticism is 'everywhere the same', and stressing the role of mystics' beliefs and expectations in shaping the very quality of their experience.)

Katz, S.T. (ed.) (1983) *Mysticism and Religious Traditions*, New York: Oxford University Press.(Contributors further explore themes from the preceding anthology, arguing that mysticism can only be understood in the context of the particular religious and cultural traditions within which it occurs.)

Lewis, I.M. (1971) *Ecstatic Religion*, Baltimore, MD: Penguin.(A noted social anthropologist describes 'mystical' aspects of African spirit possession.)

Lossky, V. (1957) *The Mystical Theology of the Eastern Church*, Cambridge and London: James Clarke & Co. (Classic study of Orthodox mysticism.)

McGinn, B. (1991) *The Presence of God: A History of Western Christian Mysticism*, vol. 1, *The Foundations of Mysticism: Origins to the Fifth Century*, New York: Crossroad.(This book, the first in a projected four-volume history, contains a long and valuable appendix on theological, philosophical, comparativist and psychological approaches to mysticism.)

Merlan, P. (1969) *Monopsychism, Mysticism, Metaconsciousness: Problems of the Soul in the Neoaristotelian and Neoplatonic Tradition*, The Hague: Martinus Nijhoff. (Technical but rewarding discussion of the often overlooked or misinterpreted tradition of 'rationalistic mysticism' in the West.)

Otto, R. (1970) *Mysticism East and West*, New York: Macmillan.(Problematic but important comparison of Śaōkara and Eckhart.)

Parrinder, G. (1976) *Mysticism in the World's Religions*, New York: Oxford University Press.(Problematic in some of its positions, but contains useful material.)

Schimmel, A. (1975) *Mystical Dimensions of Islam*, Chapel Hill, NC: University of North Carolina Press. (Standard work on Sufi mysticism.)

Scholem, G.G. (1946) *Major Trends in Jewish Mysticism*, New York: Schocken. (Classic study of Jewish mystical traditions, but interprets 'mystical' more broadly than many other authors.)

Smart, N. (1967) 'Mysticism, History of', in P. Edwards (ed.) *Encyclopedia of Philosophy*, New York: Macmillan.(Excellent brief survey, somewhat coloured by Smart's own views on the 'common core' of mystical experience.)

Smith, M. (1976) *Way of the Mystics: The Early Christian Mystics and the Rise of the Sufis*, London: Sheldon Press.(Popular work on the connections between Christian and Islamic mysticism.)

Spencer, S. (1963) *Mysticism in World Religions*, Baltimore, MD: Penguin.(A good introduction to the major mystical traditions.)

Stace, W.T. (1960) *The Teachings of the Mystics*, New York: New American Library. (Anthology of mystical writings from various religious traditions, with introductions presenting the author's much discussed and disputed views on the nature and significance of mysticism.)

Upaniṣads (800-300 BC), trans. P. Olivelle, *Upaniṣads*, Oxford: Oxford University Press, 1996.(A readable and accurate translation.)

Various authors (1933-95) *Dictionnaire de spiritualité*, Paris: Beauchesne, 21 vols.(Massive dictionary of major figures and themes in Christian spirituality.)

Voragine, Jacobus de (13th century) *Legenda aurea*, trans. W.G. Ryan, *The Golden Legend: Readings on the Saints*, Princeton, NJ: Princeton University Press, 1993.(One of the most popular and influential books of the Middle Ages.)

Wiseman, J.A. (1993) 'Mysticism', in M. Downey (ed.) *The New Dictionary of Catholic Spirituality*, Collegeville, MN: Liturgical Press.(Brief survey with theological reflections upon the history of mysticism.)

Wiseman, J.A. (1970) *Concordant Discord*, Oxford: Clarendon.(The author's Gifford Lectures contain further discussion of issues raised in the previous work.)

Woods, R. (ed.) (1980) *Understanding Mysticism*, Garden City, NY: Doubleday. (Perhaps the best and most comprehensive anthology of notable articles on the nature of mysticism, its various forms in world religions, and its scientific, philosophical and theological appraisal.)

Zaehner, R.C. (1961) *Mysticism: Sacred and Profane: An Inquiry into Some Varieties of Preternatural Experience*, New York: Oxford University Press.(Influential study sharply differentiating three fundamental types of mysticism: panenhenic, monistic and theistic.)

Mysticism, nature of

Mysticism continues to elude easy definition, and its nature and significance remain the subject of intense debate. The terms 'mystic', 'mystical' and 'mysticism' have been used in an astonishing variety of ways by different authors in different eras.

Nevertheless, modern philosophical discussions have tended to focus on so-called 'mystical experiences', understood as certain states or modes of awareness, allegedly found within (and even outside) virtually all faith-traditions, and variously characterized as 'consciousness without content', 'the experience of absolute oneness', 'union with the transcendent', 'immediate consciousness of the presence of God', and so on. Philosophers are particularly interested in whether such experiences constitute a 'way of knowing', and whether they provide any support for either traditional religious beliefs or unusual metaphysical claims made by certain mystics (for example, that time is illusory). Some authors argue affirmatively, on the basis of an alleged 'universal consensus among mystics', for example, or the parallels between mystical consciousness and other modes of experience accepted as cognitive. Others, however, challenge these views, noting that mystics often appear to disagree precisely along the lines of their prior religious convictions, that mystical awareness seems capable of explanation in terms of natural causes, that mystical claims (like claims about one's private feelings) do not admit of ordinary testing, or that the alleged 'ineffability' of mystical states frustrates any attempt at rational analysis.

These concerns, then, tend to shape the kinds of questions typically addressed in contemporary philosophical discussions of mysticism, such as: What is mysticism? What are the identifying characteristics of mystical experience? Is mysticism 'everywhere the same', and if so, in what sense? Are there different types of mystical experience? What is the relationship between mystical awareness and its interpretation? Are mystical experiences a 'way of knowing'? Do they involve some form of union or contact with God? Are mystical experiences 'ineffable' or 'nonlogical', and in what sense? Can drugs or other natural stimuli induce mystical experiences, and would that affect their cognitive value?

Finally, in light of the increasingly technical nature of much of the philosophical debate, in which the primary mystical sources themselves often play a relatively minor role (except as mined for brief 'proof texts'), there have been calls for renewed attention to the larger historical, cultural and religious contexts from which mysticism and mystical literature emerge, and within which they must be interpreted.

1 Background

Like 'mystery' (Greek *mysterion*), the adjective 'mystical' (Greek *mystikos*) is derived from the Greek verb *myein* ('to close', especially the eyes and lips), suggesting something secret or concealed. The term was used by Christian authors of the patristic era to refer to the 'mystical sense' of Scripture, the hidden presence of Christ in the biblical text and, by extension, in the sacraments, especially the Eucharist. Only in the late medieval and modern periods did the use of the term begin to shift to reflect a growing interest in the 'private' experiences of those alleged to have encountered the divine. The noun 'mysticism', as Michel de Certeau (1992) has shown, is itself a seventeenth-century French creation (*la mystique*), marking a turn 'away from the liturgical and scriptural context of patristic and medieval Christianity to a situation in which private illumination and unusual psychosomatic experiences became the criteria' (McGinn 1991: 312).

The academic study of mysticism developed in the late nineteenth century, with William James' classic *The Varieties of Religious Experience* (1902) offering one of the most influential early contributions to the field (which also included important works by William Ralph Inge, Friedrich von Hügel, Joseph Maréchal, Rudolf Otto, Evelyn Underhill and others). James devotes considerable attention to mysticism, offers numerous examples of mystical reports, and identifies 'four marks which, when an experience has them, may justify us in calling it mystical': ineffability, noetic quality, transiency and passivity. Of these, the 'noetic quality', or the recipient's impression that these are 'states of knowledge', is particularly important, and James argues that 'mystical states, when well developed, usually are, and have the right to be, absolutely authoritative over the individuals to whom they come', though not to those who stand outside the experience (James [1902] 1936: 414) (see James, W. §4).

Comparativist R.C. Zaehner’s famous work *Mysticism: Sacred and Profane* (1961) draws upon cross-cultural differences in the descriptions of various mystical states to support a typology of three fundamentally distinct kinds of mysticism: panenhenic (associated with ‘nature mysticism’), monistic (or ‘soul’ mysticism) and theistic (allegedly the highest form). Despite his greater familiarity with the primary texts, Zaehner’s typology has often been criticized as a form of special pleading on theism’s behalf, and for improbably lumping together, under the heading of ‘soul mysticism’, traditions with such radically divergent views on the soul as Sāōkhya-Yoga, Advaita Vedānta and Buddhist mysticism.

More influential in shaping contemporary Anglo-American philosophical discussions of mysticism has been Walter T. Stace’s *Mysticism and Philosophy* (1960a). Stace first identifies two main types of mystical consciousness: ‘extrovertive’, involving an apprehension of the oneness of all things, and ‘introvertive’, involving ‘Unitary Consciousness, from which all the multiplicity of sensuous or conceptual or other empirical content has been excluded, so that there remains only a void and empty unity’. He goes on to argue, however, that these states are related as ‘two species of one genus’, and share a ‘universal core’ of common characteristics, including ‘the apprehension of an ultimate nonsensuous unity in all things’ (though realized differently in extrovertive and introvertive states), blessedness, paradoxicality, alleged ineffability, and a sense of objectivity (Stace 1960a: 110, 131-2; 1960b: 14). Most importantly and controversially, Stace explains apparent conflicts among mystical reports from different religious traditions as the result of later interpretations superimposed on what are essentially the same experiences. He is particularly critical of Zaehner for failing to grasp the implications of the experience/interpretation distinction, and for simply concluding that Christian and Indian mysticism are different ‘from the mere fact that the *beliefs* which Christian mystics based upon their experiences are different from the *beliefs* which the Indians based on theirs’ (Stace 1960a: 35-6).

Stace’s own views on the relation between experience and interpretation have since been strongly criticized, especially by some of the authors represented in Steven Katz’s anthologies *Mysticism and Philosophical Analysis* (1978) and *Mysticism and Religious Traditions* (1983), who insist that the mystics’ religious backgrounds and prior beliefs play a much greater role in actually shaping the specific character of the experience. More recently, the contributors to Robert Forman’s collection *The Problem of Pure Consciousness* (1990) have criticized in turn this ‘constructivist’ position, defending the possibility of ‘pure consciousness events’ experientially indistinguishable across different religious traditions. Meanwhile, Peter Moore (Katz 1978), Nelson Pike (1992) and others have pointed out that many mystics themselves describe not a single mystical state, but successive degrees of mystical union, and that the ordered pattern of these distinct stages may itself have epistemic significance. At the same time, Bernard McGinn (1991) has argued that ‘presence’ is a far more useful category for the understanding of certain mystical traditions than that of ‘union’, despite the latter’s prominence in philosophical analyses (though it is not yet clear whether McGinn is simply replacing what is arguably too narrow a category with one too broad).

2 Defining mysticism

What, then, is mysticism, and what are the identifying characteristics of mystical states? Following the lead of certain mystics themselves, modern commentators often sharply distinguish mystical awareness from sensory experience, discursive thought, the ordinary reflective exercise of mental imagery and concepts, and such extraordinary phenomena as visions, voices and levitations. More controversially, some authors (for example, Ninian Smart 1965) also exclude numinous and prophetic experiences from the ‘mystical’ category, though in fact all of these experiences are hard to disentangle in the lives and testimony of actual mystics. Many would agree with James that mystical experiences seem to their recipients to be states of knowledge or insight, to possess what is variously called a ‘noetic quality’, a ‘sense of objectivity or reality’, or a ‘perception-like’ character. Most importantly, it is often claimed (though seldom in terms equally acceptable to all religious traditions) that those undergoing mystical experiences feel themselves to have attained some supreme goal, and to be in the presence of, or in union with, the Absolute, the Transcendent, the One.

However, as the brief overview in §1 suggests, a deeper philosophical consensus on mysticism’s scope and nature seems unlikely at this point, at least in part because the terminology and focus of the discussion have changed so dramatically over time, and attempts at more precise definitions now appear almost unavoidably tendentious; philosophers are naturally disinclined to count as truly ‘mystical’ anything which does not seem to fit their

particular theories. Indeed, some (for example, in Katz 1983) have argued that the whole modern preoccupation with identifying the internal features of these relatively rare and unusual states of consciousness has tended to impede a more comprehensive philosophical understanding of mysticism in all its dimensions and manifestations.

3 Mystical experience and its interpretation

Is mysticism 'everywhere the same'? This apparently simple question provokes endless philosophical debate, not only because of disagreements about the nature of mysticism, but also in part because of a lack of clarity about the kind of sameness or difference at issue. Presumably any two experiences are the same at least to the extent that they are both experiences, and different at least to the extent that they are distinguishable as two. Presumably even otherwise indistinguishable states of 'pure consciousness' or 'undifferentiated unity' may differ in intensity, duration, clarity, time of occurrence, and so on.

But what primarily interests philosophers, it seems, is whether mystical states are sufficiently alike across different periods and cultures to offer some *prima facie* support for belief in their cognitive value (or, correspondingly, to undermine any one tradition's claim to be the sole possessor of authentic mysticism). The underlying assumption here is that, other things being equal, the greater the consensus, the stronger the case in mysticism's favour (just as we tend to give greater credence to reports of unusual sightings when there is multiple attestation from independent sources). Of course, this assumption can be challenged; there are certainly cases of mass delusion. None the less, even many otherwise sceptical authors, as well as those inclined to make exclusivist claims for particular mystical traditions, have found the apparent cross-cultural similarities striking, and something that requires explanation.

Yet first impressions can deceive. As James admits, when we examine the broader field of mystical reports more closely, 'we find that the supposed unanimity largely disappears…[Mysticism] is dualistic in Sankhya, and monistic in Vedanta philosophy' ([1902] 1936: 416). Christian mystical literature speaks of union with God; Theravāda Buddhism does not. In short, mystics of different backgrounds describe their experiences in radically different ways.

To maintain the 'common core' hypothesis in the face of conflicting descriptions, authors typically distinguish the mystical experience as such from its interpretation. Ninian Smart, for example, argues that 'phenomenologically, mysticism is everywhere the same', but that 'different flavours accrue to the experiences of mystics because of their ways of life and modes of auto-interpretation' (Smart 1965: 87). Stace, as noted above, maintains that all authentic introvertive mystical states are in fact experiences of 'pure undifferentiated unity', but that 'the same mystical experience may be interpreted by a Christian in terms of Christian beliefs and by a Buddhist in terms of Buddhistic beliefs' (1960b: 10). Thus 'union with God', according to Stace, 'is not an uninterpreted description of any human being's experience', but 'a theistic interpretation of the undifferentiated unity' (Stace 1960a: 103-4).

Interestingly, Stace's approach is sometimes turned on its head in popular writings on mysticism, by authors who seem to imply that all mystics in fact experience union with God, but that some interpret this union in non-theistic terms. Most proponents of the 'common core' hypothesis, however, seem to favour 'undifferentiated unity' or 'pure consciousness without content' as the essential feature of all introvertive mystical states.

Stace in turn has been roundly criticized for his naïve approach to the experience/interpretation distinction. Peter Moore notes that, besides the 'raw experience' (that is, 'features of experience unaffected by the mystic's prior beliefs, expectations, or intentions'), the interpretive element found in mystical reports may be not only retrospective (as in Stace's view) but also reflexive ('spontaneously formulated either during the experience itself or immediately afterwards') or incorporated ('caused or conditioned by a mystic's prior beliefs, expectations and intentions') (Katz 1978: 108-9). Steven Katz goes further, insisting that 'there are NO pure (i.e., unmediated) experiences'. Rather, 'the experience itself as well as the form in which it is reported is shaped by concepts which the mystic brings' to the experience, so that Hindus have Hindu mystical experiences, Jews have quite distinct Jewish mystical experiences, and so on (Katz 1978: 26). As critics have pointed out, Katz seems to assume rather than prove the 'constructed' character of all experience, and makes his case more by way of persuasive examples than conclusive demonstration. Pressed to its limits, his radically pluralist, contextual and constructivist approach would seem to have the counterintuitive consequence of making all mystical states (and indeed all experiences) unique and incommensurable.

In response, Forman and others have argued that so-called ‘pure consciousness events’, involving ‘a wakeful though contentless (non-intentional) consciousness’, can and do occur, and that such mystical states cannot be adequately accounted for on the ‘constructivist’ model, since they contain no inner content to be ‘constructed’ in the requisite way. Moreover, in so far as such experiences occur across different traditions and cultures, they would seem to provide a clear instance of what Katz claims is impossible: a type of mystical consciousness virtually the same everywhere. Whether Forman has identified a convincing counterexample to the ‘constructivist’ position, or whether ‘pure consciousness events’ are what most classic mystical texts are actually describing, is still under debate.

Other authors, without adopting Katz’s radical pluralism, nevertheless sort out mystical states into a smaller number of basic types. Here, too, the question arises of which differences in the way such states are described correspond to important differences among the experiences themselves. Many surveys of mysticism divide their material according to the major religious traditions, although ordinarily for expository reasons alone. Zaehner, as we have seen, sharply distinguishes theistic ‘love’ mysticism from the monistic or ‘soul’ mysticism found in certain Asian religions. Neo-Thomist Jacques Maritain (1944) likewise sharply differentiates ‘supernatural mystical experience, by means of affective connaturality with the deity’, from the purely natural mysticism of Hindu yogis, involving ‘an intellectual (negative) experience of the substantial esse of the soul’. Critics charge that such distinctions are dictated as much by Maritain’s and Zaehner’s theological presuppositions as by careful attention to the mystical reports themselves. Yet while acknowledging such criticisms, Wainwright and others have defended the legitimacy in principle of a typological approach, and pointed the way towards a more sophisticated typology of mystical states.

Such ongoing disputes have spawned a vast and often highly technical literature on the relationship between experience and its interpretation, with important implications not simply for the understanding of mysticism but for epistemology in general. These discussions underscore the complexity of the hermeneutical task in interpreting mystical reports, and the necessity of focusing not just on isolated psychological states, but on the larger historical and cultural context in which they occur.

4 Mysticism and God

Although at times the ‘consensus of the mystics’ has been invoked on behalf of monism, idealism and other less well-known metaphysical doctrines, Anglo-American philosophers have tended to approach mysticism primarily in terms of what it might contribute to questions about the nature and existence of God. Authors debate whether mystical states in particular, and religious experiences in general, should be considered cognitive, and whether they provide any support for theistic belief. Most often, therefore, mysticism is discussed in conjunction with the so-called ‘argument from religious experience’, as a possible source of premises upon which one could base an inference (causal, explanatory or otherwise) to the existence of God (see Religious experience §2). Alston and others take a different approach, arguing that putative mystical perceptions of God are part of a justified belief-forming practice, in which ‘people sometimes do perceive God and thereby acquire justified beliefs about God’ (Alston 1991: 3), in much the same way that we acquire justified beliefs about external objects by directly perceiving them, and not just by an inference from internal sensory impressions.

Such an approach presumes, of course, that at least some mystical experiences have an intentional structure, and that it makes sense to talk of ‘theistic mystical experiences’. This assumption may be challenged. Forman’s ‘pure consciousness events’, for example, are characterized as ‘contentless’ and ‘non-intentional’. Stace’s states of ‘pure undifferentiated unity’ are similarly described. If one holds, with Stace, that all introvertive mystical consciousness is of this nature, then it is difficult to see how such experiences could be phenomenally ‘of’ anything. Pike has argued that even an experience of undifferentiated unity might count as ‘phenomenologically theistic’ if this ‘climax moment of the paradigm union experience is preceded by specifically theistic experience having dualistic structure’ (Pike 1992: 164). That is, just as we might correctly describe our experience of ‘stun-stars and fading consciousness’ as ‘the experience of being hit by a baseball’ if in fact we had observed the course of ball before impact, so too an undifferentiated unity might be correctly described as (apparent) union with God if it comes as the culmination of a sequence of states in which the subject-object distinction seems gradually to disappear as God and the soul draw closer.

None the less, some philosophers apparently want to claim something stronger: that the climactic experiences

described by mystics such as Ruusbroec and St Teresa are not merely states of undifferentiated unity having a ‘theistic’ phenomenological ancestry, but are themselves ‘perception-like’, involving a nonsensuous (putative) presentation of God to the mystic. The cognitive value of such experiences can then be defended by invoking some variant of what Richard Swinburne calls the ‘principle of credulity’, that ‘in the absence of special considerations what one seems to perceive is probably so’, and ‘how things seem to be is good grounds for a belief about how things are’ (Swinburne 1979: 254).

But are there in fact ‘special considerations’ that might limit the application of this principle? Antony Flew (1966), Ronald Hepburn (1967a), C.B. Martin (1959) (most famously) and others have argued that religious experiences (and thus, by implication, mystical states) do not qualify as a ‘way of knowing’ at least in part because they do not meet certain standards of testability that experiences must satisfy if they are to count as cognitive. If someone reports seeing a skunk in the basement, for example, I know how to confirm or disconfirm this (by looking, smelling, setting a trap, and so on), but ‘when someone claims to have direct awareness of God, to encounter, see, or intuit the divine, we are not able to suggest a test performance of an even remotely analogous kind’ (Hepburn 1967a: 166).

Alston (1991), Wainwright (1981) and others have responded that, in fact, religious communities use many checks and testing procedures for evaluating the authenticity of alleged mystical perceptions of God, such as: the consequences of the experience for the mystic and for others, the orthodoxy of the claims based on the experience, the similarity to paradigmatic mystical states, the judgment of qualified religious authorities, and so on. Critics reply that such tests are inadequate, especially since a sceptic can admit that a mystical state satisfies them all while denying that it constitutes an experience of God. To this, however, the usual counter-response is that standards of testability must be appropriate to the nature of what is allegedly perceived; it is not clear that different or more stringent checking procedures can reasonably be demanded for alleged perceptions of an immaterial, all-good, all-powerful personal God free to decide the occasion and recipient of divine self-manifestations.

Again, one crucial test for evaluating most perceptual claims is whether they agree with those of other qualified observers. Ordinarily, the testimony of virtually all normal observers asked to verify the same putative sensory perception (for example, of a skunk in the basement) would be more or less unanimous. By contrast, critics charge, mystics disagree even among themselves, and theistic mystical experiences seem to occur only to those who already believe in God, monistic experiences to those from monistic traditions, and so on. This suggests that mystical ‘perceptions’ of God are shaped more by the mystic’s prior beliefs and expectations than by anything independent. One possible response is to challenge the extent of the disagreements among mystics of different backgrounds, or the conformity to the recipient’s prior beliefs. Mystics seem to recognize each other across denominational boundaries, and most speak of the novelty and wonder of mystical union; in many cases, recipients have no particular religious affiliation or prior faith. The underlying issue here may be whether the actual distribution and frequency of theistic experiences across cultures is what we might antecedently anticipate if in fact God exists and is perceived in these states, or whether we would otherwise expect such experiences to be more common; philosophers argue somewhat inconclusively over God’s possible reasons for more or less self-disclosure.

Another ‘special consideration’ often mentioned is that the occurrence of mystical states seems capable of other explanations (physiological, psychological, psychoanalytical, and so on). Some authors have claimed that mystical experiences can be experimentally evoked through sensory deprivation, ‘deautomatization’ (a term coined by A.J. Diekman to describe the undoing of habitual patterns of perception by techniques such as staring steadily at an object or image; see Diekman in Woods 1980), taking hallucinogens, and so on; there is, in fact, an extensive literature on the question of drugs and mysticism. One may argue that ‘instant chemical mysticism’ is somehow too cheaply bought, and dispute whether the states thus produced actually correspond with what the mystics describe. But the mere fact that physiological and psychological processes may be involved in a mystical experience hardly discredits it, since such processes are involved in ordinary perception as well, and one can always argue that drugs simply open the doors of mystical perception rather than entirely creating the mystical experience. Of course, in most perceptual theories, putative perceptions must have the right kind of causal connection with what is allegedly perceived, but given the ‘primary causality’ traditionally attributed to God, any experience will be grounded in God’s causal activity, whatever the ‘secondary causes’ involved. To be sure, a putative mystical experience after taking drugs may be discredited, but on other grounds (for example, its bad

effects).

In short, one important line of defence of the cognitive value of mysticism, and of the practice of forming beliefs based on theistic mystical experiences, appeals to the analogy between mystical experience and other modes of experience accepted as cognitive (especially sensory experience). How plausible one finds this appeal will depend on how much significance one attaches to the various points of similarity and difference. Critics remain unconvinced, arguing that the dissimilarities outweigh the similarities.

5 Other questions

Mystics' comments about the strangeness of their experiences or their difficulties in describing them have sometimes provoked elaborate philosophical analyses of the 'ineffability' and 'paradoxicality' of mystical states. Yet few mystics actually claim that their experiences are radically 'ineffable' or 'paradoxical' in the self-defeating sense so easily refuted in philosophical discussions. Perhaps these discussions are most useful as reminders of the complex and creative use of language in mystical literature: sometimes descriptive, sometimes prescriptive, often evocative (the frequent recourse to poetry and symbolism is notable).

Again, while some authors assert that mystics experience union with the Supreme Good and the source of all morality, others claim that mystics pass beyond all oppositions, including the distinction between good and evil. Not surprisingly, therefore, philosophers have shown some interest in the relationship between mysticism and ethics.

Certain neo-scholastics (Jacques Maritain, Réginald Garrigou-Lagrange) and 'transcendental Thomists' (Karl Rahner, Bernard Lonergan), as well as authors from other traditions (process thought, for example), have developed their own distinctive approaches to the understanding of mysticism, relatively neglected by contemporary analytic philosophers perhaps because couched in unfamiliar categories and terminology. More generally, the 'turn to the subject' in modern theology and efforts (since Schleiermacher) to ground theological claims in religious experience raise important philosophical issues as well. A particularly delicate question is to determine how mystical and religious experience might provide more than mere reinforcement of what is already believed, without adding to or subtracting from the original revelation.

Finally, despite analytic philosophy's tendency to focus on mystical states of awareness, it should not be thought that this is all mysticism has to offer philosophy. Historically, philosophers have often been influenced by their reading of mystics (for example, Schelling by Boehme, Schopenhauer by the Upaniṣads, Heidegger by Eckhart) in much more complex ways. Mysticism, understood broadly, may have contributions to make in many areas, such as the philosophical understanding of the self and personal identity, for example, or aesthetic theory. Most importantly, contemporary philosophers need to recall that, for all their interest in the internal characteristics of mystical consciousness, they generally have no direct access to mystical states, but only to mystical texts, composed in a particular genre, for a particular audience, in a particular cultural and historical setting, and so on. Without greater attention to the special hermeneutics of mystical texts, contemporary philosophers will continue to run the risk of analysing with ever greater sophistication a 'mysticism' that exists nowhere in reality, and seems far removed from the actual experience and testimony of actual mystics.

See also: God, arguments for the existence of; Mysticism, history of; Otto, R.; Pantheism; Religion and epistemology; Religious experience; Religious pluralism

STEVEN PAYNE

References and further reading

Alston, W.P. (1991) *Perceiving God: The Epistemology of Religious Experience*, Ithaca, NY: Cornell University Press.(Sustained defence of the contribution of religious experience to the grounds for belief in God. Contains valuable discussion of many epistemological issues and the possibility of 'mystical perceptions of God'.)

Bouyer, L. (1980) 'Mysticism/An Essay on the History of the Word', in R. Woods (ed.) *Understanding Mysticism*, Garden City, NY: Image Books.(Outlines evolution of the term 'mystical' from Greek origins through patristic era. Translation of 1952 article from *La Vie Spirituelle*.)

Certeau, M. de (1992) *The Mystic Fable*, vol. 1: *The Sixteenth and Seventeenth Centuries*, Chicago, IL: University of Chicago Press.(Difficult but rewarding discussion of the term 'la mystique' and the fundamental change it represents in the Western approach to the sacred. Incorporates semiotics, linguistics and

psychoanalytic theory.)
Coward, H. and Penelhum, T. (eds) (1977) *Mystics and Scholars: The Calgary Conference on Mysticism 1976*, Waterloo, Ont.: Wilfrid Laurier University Press. (Contains useful articles by John Hick, Walter Principe, Terence Penelhum and others.)
Davis, C.F. (1989) *The Evidential Force of Religious Experience*, Oxford: Clarendon. (Treats mysticism in the context of religious experience, as part of a cumulative case for theism. Discusses many issues raised in this article.)
Dupré, L. (1976) *Transcendent Selfhood: The Rediscovery of the Inner Life*, New York: Seabury.(Valuable discussion in chapter 8 of mysticism's contributions regarding the nature of the self. Revision of longer article reprinted in Woods (1980).)
Flew, A. (1966) *God and Philosophy*, New York: Delta Books.(Chapter 6 is a sustained attack on arguments from religious experience, raising the issues of testability and agreement.)
Forman, R.K.C. (ed.) (1990) *The Problem of Pure Consciousness: Mysticism and Philosophy*, New York: Oxford University Press.(The contributors criticize Katz's 'constructivism' and argue for new models of mystical experience. Authors defend the possibility and actual occurrence of 'pure consciousness events'.)
Gale, R.M. (1991) *On the Nature and Existence of God*, New York: Cambridge University Press.(Chapter 8 challenges religious experience arguments, with a long critique of Alston and Wainwright on the analogy between religious and sensory experience.)
Hepburn, R.W. (1967a) 'Religious Experience, Argument for the Existence of God', in P. Edwards (ed.) *Encyclopedia of Philosophy*, New York: Macmillan, vol. 7.(Deals with issues of testability, agreement, natural explanations, and so on.)
Hepburn, R.W. (1967b) 'Mysticism, Nature and Assessment of', in P. Edwards (ed.) *Encyclopedia of Philosophy*, New York: Macmillan, vol. 5.(Discusses conflicts of interpretation, mystical paradoxes, and so on. Sympathetic towards mysticism as a valuable human phenomenon, but agnostic concerning its religious implications.)
Hügel, F. von (1908) *The Mystical Element of Religion as Studied in Saint Catherine of Genoa and her Friends*, London: James Clarke & J.M. Dent, 2 vols, 1961.(Unusual mixture of biography, theology and philosophy. Difficult and not entirely successful, but rewarding.)
Inge, W.R. (1899) *Christian Mysticism*, London: Methuen.(Classic early survey with famous appendix listing twenty-six definitions of 'mysticism' and 'mystical theology'. Marred by the author's biases.)
James, W. (1902) *The Varieties of Religious Experience: A Study in Human Nature*, New York: Modern Library, 1936.(Lectures 16 and 17 treat the characteristics and significance of mystical states, with a wealth of examples. Enormously influential.)
Katz, S.T. (ed.) (1978) *Mysticism and Philosophical Analysis*, New York: Oxford University Press.(Important articles by Katz, Peter Moore and others criticizing the view that mysticism is 'everywhere the same', and stressing the role of mystics' beliefs and expectations in shaping the very quality of their experience.)
Katz, S.T. (ed.) (1983) *Mysticism and Religious Traditions*, New York: Oxford University Press.(Contributors further explore themes from the preceding anthology, arguing that mysticism can only be understood in the context of the particular religious and cultural traditions within which it occurs.)
Maréchal, J. (1926, 1937) *Studies in the Psychology of the Mystics*, trans. A. Thorold, Albany, NY: Magi Books, 1964.(Difficult but rewarding analysis of epistemological preconditions for mystical intuition of God, by a precursor of transcendental Thomism. This is a partial translation of the two-volume French edition.)
Maritain, J. (1944) 'The Natural Mystical Experience and the Void', in *Redeeming the Time*, London: Geoffrey Bles Centenary Press.(Goes beyond the following work in allowing the possibility of natural mystical experiences, associated primarily with certain Asian religions.)
Maritain, J. (1959) *Distinguish to Unite, or The Degrees of Knowledge*, trans. G.B. Phelan, New York, Scribner, 4th edn.(Magisterial study by a pre-eminent Neo-Thomist which presents mystical knowledge as the summit and goal of all knowing.)
Martin, C.B. (1959) *Religious Belief*, Ithaca, NY: Cornell University Press.(Chapter 5, 'Seeing God', is an influential and often reprinted critique of religious experience as a 'way of knowing', on the grounds that appropriate tests are lacking. An earlier version appeared in *Mind* in 1952.)
McGinn, B. (1991) *The Presence of God, A History of Western Christian Mysticism*, vol. 1: *The Foundations of Mysticism: Origins to the Fifth Century*, New York: Crossroad.(The first in a projected four-volume history, this contains a long and valuable appendix on theological, philosophical, comparativist and psychological approaches to mysticism.)
Murphy, M. and Donovan, S. (1988) *The Physical and Psychological Effects of Meditation*, San Rafael, CA:

Esalen Institute.(As the subtitle indicates, this is ‘a review of contemporary meditation research with a comprehensive bibliography, 1931-1988’. Useful overview of recent scientific studies, focusing mainly on Transcendental Meditation, but with some attention to mystical experience.)

Otto, R. (1958) *The Idea of the Holy*, New York: Oxford University Press.(Classic attempt to root religion in a prerational feeling of the ‘numinous’. Includes a discussion of mysticism.)

Otto, R. (1970) *Mysticism East and West*, New York: Macmillan.(Problematic but important comparison of Śaōkara and Eckhart.)

Pike, N. (1992) *Mystic Union: An Essay in the Phenomenology of Mysticism*, Ithaca, NY: Cornell University Press.(‘Phenomenography’ of successive states of infused contemplation as described by Christian mystics, and analysis of their philosophical implications.)

Proudfoot, W. (1985) *Religious Experience*, Berkeley, CA: University of California Press.(Chapter 4 argues controversially that ‘ineffability’ and ‘noetic quality’ are not simply qualities of the mystic experience, but ‘conceptual constraints on what experiences are identified as mystical’.)

Rahner, K. (1983) *The Practice of Faith: A Handbook of Contemporary Spirituality*, New York: Crossroad. (Sections 13 and 14 briefly outline some of Rahner’s views on mystical theology and ‘everyday mysticism’.)

Smart, N. (1965) ‘Interpretation and Mystical Experience’, *Religious Studies* 1: 75-87.(An attempt to define the scope of mysticism.)

Stace, W.T. (1960a) *Mysticism and Philosophy*, London: Macmillan.(This classic defence of the ‘common core’ hypothesis and other controversial claims has set the agenda for most subsequent philosophical discussions.)

Stace, W.T. (1960b) *The Teachings of the Mystics*, New York: New American Library. (Companion anthology to the preceding work.)

Swinburne, R. (1979) *The Existence of God*, Oxford: Clarendon Press.(Chapter 13 invokes the ‘principle of credulity’ in defending the argument from religious experience.)

Thompson, W.M. (1987) *Fire and Light: The Saints and Theology*, New York: Paulist Press.(An essay ‘on consulting the saints, mystics, and martyrs in theology’, with a useful review of how contemporary theologians are incorporating the contributions of Christian mystics.)

Underhill, E. (1965) *Mysticism: A Study in the Nature and Development of Man’s Spiritual Consciousness*, Cleveland, OH, and New York: World, 12th edn.(Classic popular introduction to mysticism and the stages of the mystical journey, focusing primarily on Christian mystics.)

Wainwright, W. J. (1978) *Philosophy of Religion: An Annotated Bibliography of Twentieth Century Writings in English*, New York: Garland.(Pages 367-438 offer concise and perceptive evaluations of over 200 works on mysticism and religious experience.)

Wainwright, W. J. (1981) *Mysticism: A Study of its Nature, Cognitive Value and Moral Implications*, Madison, WI: University of Wisconsin Press.(Dense, masterly discussion of many issues raised in the present entry.)

White, J. (ed.) (1972) *The Highest State of Consciousness*, Garden City, NY: Doubleday & Co.(Uneven but useful anthology of psychological essays on transcendent states of awareness, including mystical consciousness.)

Woods, R. (ed.) (1980) *Understanding Mysticism*, Garden City, NY: Doubleday. (Perhaps the best and most comprehensive anthology of notable articles on the nature of mysticism, its various forms in world religions, and its scientific, philosophical and theological appraisal.)

Yandell, K.E. (1993) *The Epistemology of Religious Experience*, New York: Cambridge University Press. (General defence of evidential force of religious experience, with discussion of ineffability and other topics.)

Zaehner, R.C. (1961) *Mysticism: Sacred and Profane: An Inquiry into Some Varieties of Preternatural Experience*, New York: Oxford University Press. (Influential study, sharply distinguishing three fundamental types of mysticism: panenhenic, monistic and theistic.)

Zaehner, R.C. (1970) *Concordant Discord*, Oxford: Clarendon.(Author’s Gifford Lectures, containing further discussion of issues raised in Zaehner 1961.)

Næss, Arne (1912-)

As professor of philosophy in Oslo between 1939 and 1970, Arne Næss contributed to a strengthening of the position of philosophy in Norwegian academic life. During the German occupation (1940-5) he played an active part in the resistance movement. In the 1940s and 1950s he was the inspiration for and centre of a group of students of philosophy and social science, the 'Oslo School', whose members became influential in the later development of these fields. His philosophical thinking passed through an early 'scientistic' period of radical empiricism to 'possibilist' and pluralist views, and an undogmatic scepticism. After resigning his professorship in 1970, he became the protagonist of a version of ecological philosophy, 'deep ecology'. He has always been an admirer of Spinoza and has also sought inspiration in Spinozism for his ecological philosophy.

Although the Norwegian philosopher Arne Næss had contact with the Vienna Circle of logical positivists in 1934-6, his early views were more influenced by pragmatism than by positivism (see Logical positivism; Pragmatism). He did not accept that empirical methods have no place in philosophy, and he had reservations about the logical positivists' use of a priori formal-logical methods. But he shared their high esteem of science and the idea of making philosophy more scientific, and he supported the international movement for unified science.

In his first major work, *Erkenntnis und wissenschaftliches Verhalten* (Cognition and Scientific Behaviour, written in Vienna 1934-5), Næss sketched a programme for an empirical science of science, attacking 'subjective models' in traditional accounts of cognition and research. During his first years as professor of philosophy in Oslo he developed a programme of empirically based foundation research involving criticism of unclear use of language. He provided the groundwork for a scientific semantics by introducing a conceptual framework which made semantic hypotheses testable by empirical investigations of actual usage. Næss' empirical semantics became a common methodological basis for a circle of students from different fields who attached themselves to his programme of foundation research - the 'Oslo School'. In this milieu, key formulations in the history of philosophical ideas were regarded not as expressions of definite doctrines but as highly ambiguous formulations which might be interpreted in an indefinite number of directions, and which could be made more precise.

Næss was inclined to consider the difference between normative and descriptive statements as one of degree rather than a clear dichotomy. He rejected the view that moral philosophy must be meta-ethics, a science *about* norms and their function and not a doctrine *in* morals. He envisaged a value-theoretical foundation research that sought justifications for norms, compared norm systems and sharpened normative formulations. Such a view was the basis for his later investigations of norm systems such as Gandhi's ethics of non-violence (see Gandhi, M.K.).

Næss employed concepts and methods of empirical semantics on politically controversial subjects, in several investigations of uses of the term 'democracy', for example, in studies supported by UNESCO (see Democracy); he also undertook empirical studies of philosophically relevant terms.

In the early 1950s Næss moved away from the semantically and empirically oriented platform of the Oslo School. He continued his Spinoza studies, but also started work on a comprehensive history of philosophy, making extensive use of his semantic methods on philosophical texts. He also began to reflect critically on empiricism as a fundamental position. In a series of lectures in 1953 (see Næss 1959) he discussed the presuppositions of empiricism, and considered standpoints resulting from negating or abolishing one or more such premises (see Empiricism). He sketched a series of different views, from 'narrow scientism' via 'empiricism without dogmas' and 'probabilism' to extremely wide 'possibilism'. His work in the history of philosophy changed his conception of philosophy, and inspired him to a pluralistic doctrine of philosophical systems and 'total views', which was connected with his possibilism. He espoused a thesis of 'incommensurability' of total systems (see Incommensurability).

In *Scepticism* (1968) Næss discussed Pyrrhonism, which he considered the most radical and important form of scepticism, taking his point of departure from Sextus Empiricus (see Pyrrhonism). Næss' own position was a kind of spectator standpoint, he was a 'sympathetic metasceptic' trying to present the sceptic in as favourable and coherent a light as possible. But the sceptical attitude obviously appealed to Næss and was in harmony with deep tendencies in himself: the need for distanced, theoretical non-attachment, except in questions in which he was

emotionally strongly engaged.

A tension between engagement and distance seems to mark his development. After theoretical work in philosophy of science in the 1930s, Næss played an active part in the resistance movement during the war, at considerable personal risk. After technical work in empirical semantics, logic and the philosophy of language, studies of Spinoza and the history of philosophy, and investigations of norm systems in the 1950s and 1960s, he became strongly interested in ecological philosophy from about 1970. He took part in nonviolent action for the preservation of nature, and he became known among environmentalists all over the world as a protagonist of 'deep ecology'.

In his famous 1973 paper, 'The Shallow and the Deep, Long-Range Ecology Movement', Næss distinguished between a shallow and a more 'revolutionary' deep ecology movement. The purpose of the former is to fight against pollution and resource depletion; its central objective is to promote the health and affluence of people in the developed countries. In contrast, the deep ecology movement rejects the distinction between humans and environment, emphasizing their intrinsic relationship and total-field image. Further, the deep ecology movement stresses biospherical egalitarianism in principle, promoting a deep-seated respect for ways and forms of life; it emphasizes principles of diversity and symbiosis, repudiates class differences, favours complexity over complication, local autonomy and decentralization against centralization - and fights against pollution and resource depletion, but without losing sight of the other points. In the theoretical foundation for deep ecology, Næss integrates elements of his earlier philosophical achievements and adds new points. In 'The World of Concrete Contents' (1985) he suggests that the difference between environment conservers and developers is one of ontology rather than ethics. Clarification of differences in ontology may contribute significantly to the clarification of different policies and their ethical basis. As an ontological basis for deep ecology, Næss explores a kind of neutral monism identifying the world (reality) with the set of 'concrete contents' in the form of Gestalts, as against abstract structures such as those developed in scientific theories. This 'Gestalt ontology' is contrasted with an ontology distinguishing things in themselves with 'objective' primary qualities and relations from 'subjective' secondary and tertiary qualities (see Primary-secondary distinction). *In* the concrete contents there are no subject-object, primary-secondary or fact-value distinctions (see Fact-value distinction).

See also: Ecological philosophy; Scandinavia, philosophy in §5

INGEMUND GULLVÅG

List of works

Næss, A. (1936) *Erkenntnis und wissenschaftliches Verhalten (Cognition and Scientific Behaviour),* Oslo: Det Norske Videnskapsakademi i Oslo. Skrifter. II. Hist.-filos.kl. (1936: 1).(Outlines and discusses a radical programme for an 'objectivist' science of science to replace traditional 'subjectivist' philosophical disciplines, for example, epistemology and theory of meaning. Næss' project in this work is a parallel to, and anticipates, later eliminativist projects of replacing 'folk psychology' by non-mentalist terminology and theory. The programme is reconsidered and moderated in (1965a) below.)

Næss, A. (1938) *'Truth' as conceived by those who are not Professional Philosophers*, Oslo: Det Norske Videnskapsakademi i Oslo. Skrifter. II. Hist.-filos.kl. (1938: 4).(Report on empirical investigations of what non-philosophers understand by 'truth'. The study revealed that all the main philosophical 'theories of truth' were represented in non-technical versions among 'ordinary people'. In Næss' opinion, his study showed that philosophers' claims about what 'the ordinary man' means or does not mean by truth were unfounded and irresponsible.)

Næss, A. (1947) *Endel elementære logiske emner (Some elementary logical topics),* Oslo; trans. A. Hannay as *Communication and Argument. Elements of Applied Semantics*, Oslo: Scandinavian University Books, 1966. (Elementary introduction to applied semantics - ambiguity, interpretation, precisation (disambiguation), definition, misunderstandings in communication - argumentation and debating.)

Næss, A. (1953a) *Interpretation and Preciseness. A Contribution to the Theory of Communication*, Oslo: Det Norske Videnskapsakademi i Oslo. Skrifter II. Hist.filos.kl. (1953: 4).(A technical treatment of the same topics as in his 1947 work. Here, Næss endeavours to provide the conceptual and methodological foundation for an empirical semantics.)

Næss, A. (1953b) *Filosofiens historie. En innføring i filosofiske problemer (The History of Philosophy. An Introduction to Philosophical Problems)*, Filosofiske Problemer 16, Oslo: Akademisk forlag.(A not-entirely-elementary history of philosophy.)

Næss, A. (1956) *Democracy, Ideology and Objectivity. Studies in the Semantic and Cognitive Analysis of Ideological Controversy*, with J.A. Christophersen and K. Kvalø, Oslo: Oslo University Press.(Report on empirical studies of east-west ideological conflicts.)

Næss, A. (1959) ‘Refleksjoner om helhetssyn, sannhet og kunnskap’ (Reflections on Total Views, Truth and Knowledge), *Festskrift til A.H. Winsnes på syttifemårsdagen*, Oslo, 285-308.(A first publication of ideas originally discussed in lectures in Oslo in 1953 - ideas further developed in various later publications on possibilism and pluralism.)

Næss, A. (1960) *Gandhi og atomalderen*, Oslo: Universitetsforlaget; trans. A. Hannay as *Gandhi and the Nuclear Age*, Totowa, NJ: Bedminster Press, 1965.(A discussion of the relevance of Gandhi’s principles of non-violence in the nuclear age.)

Næss, A. (1963) *Logikk og metodelære. En innføring*; trans. A. Hannay as *Introduction to Logic and Scientific Method*, Paterson, NJ, 1968.(An elementary introduction to logic and scientific method.)

Næss, A. (1964) ‘Pluralistic Theorizing in Physics and Philosophy’, *Danish Yearbook in Philosophy*, 1, Copenhagen, 101-11.(Development on Næss (1959) above.)

Næss, A. (1964-5) ‘Reflections about Total Views’, *Philosophy and Phenomenological Research* 25: 16-29. (Development on Næss (1959) above.)

Næss, A. (1965a) ‘Science as Behaviour: Prospects and Limitations of a Behavioural Metascience’, in B.B. Wolman (ed.) *Scientific Psychology. Principles and Approaches*, New York, 50-67.(Development on Næss (1936) above.)

Næss, A. (1965b) *Moderne filosofer*, Stockholm: Almqvist & Wiksell; trans. A. Hannay as *Four Modern Philosophers: Carnap, Wittgenstein, Heidegger, Sartre*, Chicago, IL: The University of Chicago Press, 1969. (Elementary introductions to the philosophies of Carnap, Wittgenstein, Heidegger and Sartre.)

Næss, A. (1966) ‘Psychological and Social Aspects of Pyrrhonian Scepticism’, *Inquiry* 9: 301-21.(Outline of a non-dogmatic form of scepticism ascribed to Pyrrho; discussion of possible beneficial effects of this kind of non-dogmatic attitude.)

Næss, A. (1968) *Scepticism*, London: International Library of Philosophy and Scientific Method.(A thorough discussion of non-dogmatic, Pyrrhonian scepticism, the kind claimed by Næss to be generally overlooked in modern critiques of scepticism.)

Næss, A. (1969) *Hvilken verden er den virkelige? (Which World is the Real One?)*, Oslo: Universitetsforlaget. (Discussion of total views and problems involved in describing, comparing and evaluating near-total views.)

Næss, A. (1972a) *Freedom, Emotion and Self-Subsistence. The Structure of a Central Part of Spinoza’s ‘Ethics’*, Oslo: Oslo University Press, 1975.(A somewhat technical discussion of a part of Spinoza’s *Ethics*.)

Næss, A. (1972b) *The Possibilist and Pluralist Aspects of the Scientific Enterprise*, Oslo: Oslo University Press. (Development on Næss (1959) above.)

Næss, A. (1973) ‘The Shallow and the Deep, Long-Range Ecology Movement’ *Inquiry* 16: 95-100.(A famous early expression of Næss’ ideas of ‘deep ecology’.)

Næss, A. (1974) *Økologi, samfunn og livsstil. Utkast til en økosofi*, Oslo: Universitetsforlaget; trans. D. Rothenberg as *Ecology, Community and Lifestyle: Outline of an Ecosophy*, Cambridge: Cambridge University Press, 1989.(A more comprehensive discussion of the philosophy of ‘deep ecology’.)

Næss, A. (1974) *Gandhi and Group Conflict. An Exploration of Satyagraha - Theoretical Background*, Oslo: Oslo University Press.(A more thorough discussion of the same themes as developed in Næss (1960) above.)

Næss, A. (1975) *Freedom, Emotion and Self-Subsistence. The Structure of a Central Part of Spinoza’s ‘Ethics’*, Oslo: Oslo University Press.(A further discussion of Spinoza’s *Ethics*.)

Næss, A. (1979) ‘Towards a Philosophy of Wide Cognitivism’ in *Theory of Knowledge and Science Policy*, Ghent.

Næss, A. (1985) ‘The World of Concrete Contents’, *Inquiry* 28: 417-28.(An outline of an ontology of ‘neutral monism’ as a foundation for ‘deep ecology’.)

Næss, A. (1986) ‘The Deep Ecology Movement: Some Philosophical Aspects’, *Philosophical Inquiry* 8; repr. in G. Sessions (ed.) *Deep Ecology for the Twenty-First Century*, Boston, MA: Shambhala, 1995.(Further development of themes discussed in (1973) and (1974a) above.)

Næss, A. (1995a) ‘Deepness of Questions and the Deep Ecology Movement’, in G. Sessions (ed.) *Deep Ecology for the Twenty-First Century*, Boston, MA: Shambhala.(Further development of themes discussed in (1973) and (1974a) above.)

Næss, A. (1995b) ‘Self-Realization: An Ecological Approach to Being in the World’, in G. Sessions (ed.) *Deep Ecology for the Twenty-First Century*, Boston, MA: Shambhala. (Further development of themes discussed in (1973) and (1974a) above.)

References and further reading

Glasser, H. (1995) ‘Deep Ecology Clarified: A Few Fallacies and Misconceptions’, *The Trumpeter Journal of Ecosophy* 12: 138-42.

Gullvåg, I. and Wetlesen, J. (eds) (1982) *In Sceptical Wonder. Inquiries into the Philosophy of Arne Naess on the Occasion of his 70th Birthday*, Oslo: Universitetsforlaget and Columbia University Press.

Light, A. and Rothenberg, D. (eds) (1996) ‘Arne Naess’s Environmental Thought’, special issue of *Inquiry* 39. (Essays on Næss’ thought, published in honour of his eighty-fifth birthday.)

Reed, P. and Rothenberg, D. (eds) (1993) *Wisdom in the Open Air; the Norwegian Roots of Deep Ecology*, Minneapolis, MN: University of Minnesota Press.(Traces the Norwegian roots of ‘deep ecology’, with contributions from many prominent Norwegian thinkers, including Næss.)

Nāgārjuna (*c.* AD 150-200)

Nāgārjuna was the first Buddhist philosopher to articulate and seek to defend the claim that all things are empty, that is, devoid of their own essential nature. A native of South India, as the founder of the Madhyamaka school of Mahāyāna Buddhism he exerted a profound influence on the further development of Buddhist thought in South and East Asia. When he claimed that all things are empty, he denied that anything exists solely in virtue of its own inherent nature. If, as all Buddhists hold, existents only arise in dependence on other existents, then nothing may be said to have a determinate nature apart from its relations to other things. Yet prior developments in Buddhist philosophy had presumably shown that anything lacking an independent nature is a conceptual fiction and not ultimately real. Thus if all things are empty, nothing is ultimately real. Still Nāgārjuna claimed not to be a nihilist. Emptiness is rather the defeat of all metaphysical theories, all attempts at grasping the ultimate nature of reality - including nihilism. Insight into emptiness is said to free us from our tendency conceptually to construct an ultimate truth, a tendency that bolsters our sense of self. Thus realization of emptiness is, Nāgārjuna held, required in order to attain full liberation from the suffering caused by clinging.

1 Historical context

Nāgārjuna was the first philosopher systematically to develop the assertion made in the *Prajñāpāramitā* (Perfection of Wisdom) literature of early Mahāyāna Buddhism that all things are empty (*śūnya*). His principal work, *Mūlamadhyamakakārikā(Fundamental Treatise on the Middle Path)* serves as the foundational text of Madhyamaka, one of the two chief philosophical schools of Mahāyāna. Another surviving text, *Vigrahavyāvartanī(Refuting the Opponents)*, replies to various objections to the doctrine of emptiness. Modern scholars are divided over the authorship of several other works traditionally attributed to Nāgārjuna.

Nāgārjuna's claim that all things are empty, that is, devoid of 'own essential nature' (*svabhāva*), is in part a response to earlier developments in Buddhist philosophy. Early Buddhists held a reductionist view of persons, according to which the continued existence of a person just consists in the occurrence of a causal continuum of impersonal, impermanent, physical and psychological elements. In the Abhidharma or scholastic phase of Buddhism, this reductionist tendency was developed into a thoroughgoing denial of all partite entities as mere conceptual fictions, things that are commonly thought to exist only because of our use of such convenient designations as 'forest', 'chariot', and 'person'. Yet our use of such convenient designations does have utility, and this fact requires explanation. Thus it was thought that while the chariot is ultimately unreal, there must be some set of ultimately real impartite entities, the behaviour of which explains our use of 'chariot'. Such an ultimately real entity is a *dharma*, defined as that which bears its own essential nature. This definition was arrived at by way of the fact that the essential properties of a complex entity are all borrowed from its parts: that a chariot serves as a mode of transportation is to be explained in terms of facts about its parts and their relations. Since such explanation must terminate in entities that are ultimately real, these must be conceived of as bearing their own essential nature. Thus when Nāgārjuna states that all things are empty, he is denying that there are any ultimately real entities.

2 Arguments for emptiness

Nāgārjuna's basic strategy in arguing for the emptiness of all existents is to seek to show that the assumption that real entities bear their own essential nature leads to absurdities: implications that either contradict the official position of some Abhidharma theory, or else fly in the face of common sense. The argument of *Mūlamadhyamakakārikā* 3 is an instance of the first sort of *reductio*. It was a received view among Ābhidharmikas that there are six sense faculties (the sixth being a kind of inner sense), each with its own distinct class of sense-object (see Buddhism, Ābhidharmika schools of §§3-4). Since consciousness was claimed to arise in dependence on contact between sense-faculty and sense-object, this doctrine could be used to demonstrate the radical impermanence of consciousness (and thus to support the view that there is no self). But Nāgārjuna points out that if, for example, the faculty of vision is thought to be a *dharma*, then it must bear its own essential nature of seeing. Yet vision cannot see itself, as no (impartite) entity may operate on itself. Vision manifests its essential nature of seeing only in the presence of a visible sense-object and light. Thus vision may not be said to bear its own essential nature of seeing; rather, it borrows its essential nature from the conditions on which seeing depends.

Vision is not, then, an ultimately real entity, but merely a conceptual fiction. And if consciousness does arise, it can only do so in dependence on causes and conditions that are themselves ultimately real. Thus the Ābhidharmika may not claim that consciousness arises in dependence on vision, or any of the other sense faculties either.

The view that ultimately real existents bear their own essential nature also contradicts common sense, for example, the fact that existing things undergo alteration. For if all existents are indeed impermanent, then alteration must ultimately require that things lose their essential nature, which is incoherent. If being young is the essential nature of the youth, and being old the essential nature of the aged, who is it that undergoes ageing (*Mūlamadhyamakakārikā* 13.5)? Indeed, it would appear that if what makes an entity ultimately real is that it bears its own essential nature, then the ultimately real could never be created or destroyed. For creation involves the acquisition of essential nature from the causes and conditions responsible for creation, which would make essential nature ‘borrowed’ and not ‘its own’. And since destruction involves loss of essential nature, there can be no ultimately real subject of destruction.

An additional though related problem with the notion of a *dharma* concerns the relation between essential nature and the bearer or substrate of essential nature (*Mūlamadhyamakakārikā* 5). If it is the function of essential nature to characterize the bearer, then presumably the bearer must itself be an existent. But if the bearer is itself of some determinate nature, then the function of essential nature becomes redundant. Yet it seems incoherent to suppose that the bearer is in itself utterly devoid of any nature; we only seem able to imagine such a thing because we tacitly supply the bearer with the nature of being devoid of determinate nature. Finally, it would be wrong to suppose that a *dharma* consists only of essential nature without underlying bearer, since then we should be unable to distinguish between distinct instances of a given *dharma*.

Through arguments such as these, Nāgārjuna seeks to show that all things are empty, that is, that it is not the case that there are ultimately real entities. This is not, however, nihilism, for neither is it the case that ultimately there is nonexistence. Nāgārjuna claims that in order to grasp the concept of the nonexistent, we must understand the notion of the existent. What we have presumably discovered through the investigation of the concept of a *dharma* is that the notion of the ultimately existent is incoherent. It follows, then, that we cannot give coherent content to the nihilist’s claim that reality is ultimately devoid of existents. To say that all things are empty is not to propound a metaphysical thesis, is not to seek to characterize the ultimate nature of reality, but rather to reject the very idea. Nāgārjuna calls those for whom emptiness is itself a metaphysical thesis ‘incurable’ (*Mūlamadhyamakakārikā* 13.8). Emptiness is not itself a property of ultimately real things; emptiness is just a conceptual fiction, it too is empty.

To this it was objected (*Vigrahavyāvartanī* 5-6) that Nāgārjuna cannot consistently seek to demonstrate that all things are empty, since such demonstration would require that there ultimately exist valid means of knowledge, as well as objects on which these means or instruments operate. Any ‘proof’ of emptiness would then be fallacious, in that it would presuppose the existence of ultimately real entities. To this objection Nāgārjuna replies (*Vigrahavyāvartanī* 30-51) by first denying that he seeks to prove a thesis. He then argues that there can ultimately be no such things as valid means of knowledge: any attempt to demonstrate that a given procedure is a means of knowledge will either be a mere dogmatic assertion, or else be subject to either circularity or infinite regress. Thus all supposed means of knowledge are likewise empty: none has the nature of a valid means of knowledge independently of other existents.

3 Interpretations of Nāgārjuna

Nāgārjuna’s response to the above objection, to the effect that his position is self-stultifying, is ambiguous. He might be taken as claiming that it is futile to assert any thesis concerning the ultimate nature of reality because its true nature is only obscured through such conceptual constructions as the so-called means of knowledge. In this case he does recognize such a thing as the ultimate nature of reality, though perhaps as apprehended only through some non-discursive means such as mystical intuition. Alternatively, he might be understood as a kind of global anti-realist, claiming that since a cognitive procedure could count as a valid means of knowledge only in the context of certain concrete human institutions and practices, the very idea of ‘the ultimate nature of reality’ is incoherent, since it requires that we be able to give content to the notion of how things are independently of any human institutions and practices.

Nāgārjuna's claim that emptiness is itself empty results in a paradox: the ultimate truth is that there is no ultimate truth. This and related paradoxes came to play a central role in Mahāyāna Buddhist practice. The two different ways of understanding Nāgārjuna are reflected in two different ways of understanding such paradoxes. One way is to see them as pointing beyond our conventional discursive practices to a reality that cannot be grasped conceptually. But on the global anti-realist reading, such paradoxes are instead meant to force us to abandon the quest for ultimate truth by getting us to see the very idea of ultimate truth as the final obstacle to attaining enlightenment. On this interpretation, the ultimate truth is that there is only conventional truth, a set of discursive practices that originate in dependence on contingent human needs and interests.

Nāgārjuna suggests the first interpretation when he speaks of ultimate reality as 'at peace, not populated with conceptual fictions, devoid of construction, not having many meanings' (*Mūlamadhyamakakārikā* 18.9). But it is the second interpretation that is supported when he goes on to equate such soteriologically privileged terms as *nirvāṇa*, the Buddha and liberation with the ostensibly opposed mundane concepts of, respectively, cyclic rebirth, the unenlightened and bondage. These equations - made on the grounds that both members of each pair are equally empty - suggest that enlightenment through realization of emptiness should not be construed as insight into a reality somehow purified of all conceptual contamination. To realize emptiness is just to see that our practices do not require grounding in things that bear their own essential nature.

One final piece of evidence derives from the frequency with which Nāgārjuna appears to violate either non-contradiction or excluded middle, as in his claim that an existing substance such as earth or space is not existent, not nonexistent, not both, and not neither (*Mūlamadhyamakakārikā* V.6-7). This sort of statement has sometimes been taken to show that Nāgārjuna rejects logic altogether in his account of reality. The Buddha, however, used a similar locution to emphasize the point that the subject of some predication fails to refer: just as we would not ask concerning a fire that has gone out in what direction it has gone, so we cannot say of, for example, the enlightened person that they are reborn, not reborn, both, or neither - since ultimately there are no persons. Thus those who favour reading Nāgārjuna as a global anti-realist would claim that when he says reals neither arise nor fail to arise nor both nor neither, he is best seen as simply denying that coherent sense can be made of the notion of how things ultimately are.

See also: Buddhism, Mādhyamika: India and Tibet §1; Buddhist concept of emptiness; Buddhist philosophy, Indian; Knowledge, Indian views of §4; Seng zhao

MARK SIDERITS

List of works

Nāgārjuna (*c.* AD 150-200) *Mūlamadhyamakakārikā(Fundamental Treatise on the Middle Path)*, ed. P.L.Vaidya, under the title *Madhyamakaśāstra (Mūlamadhyamakakārikā)*, India: Mithila Institute, 1960; trans. J.L. Garfield, *The Fundamental Wisdom of the Middle Way: Nāgārjuna's Mūlamadhyamakakārikā*, New York: Oxford University Press, 1995.(Contains the chief arguments for the doctrine of emptiness; *Madhyamakaśāstra* is an alternative title. Garfield supplies a commentary in addition to his translation, which is based largely on the Tibetan text.)

Nāgārjuna (*c.* AD 150-200) *Vigrahavyāvartanī (Refuting the Opponents)*, ed. and trans. K. Bhattacharya, *The Dialectical Method of Nāgārjuna*, Delhi: Motilal Banarsidass, 1978. (Contains replies to a variety of objections, including the charge that Nāgārjuna's position is self-stultifying.)

References and further reading

Lindtner, C. (1982) *Nagarjuniana: Studies in the Writings and Philosophy of Nāgārjuna*, Copenhagen: Akademisk Forlag.(Contains translations of a number of works traditionally attributed to Nāgārjuna but concerning which there is no modern scholarly consensus.)

Murti, T.R.V. (1955) *The Central Philosophy of Buddhism*, London: Allen & Unwin. (A classic statement of the thesis that Nāgārjuna holds ultimate reality to be ineffable.)

Ruegg, D.S. (1977) 'The Uses of the Four Positions of the Catuṣkoṭi and the Problem of the Description of Reality in Mahāyāna Buddhism', *Journal of Indian Philosophy* 5: 1-71.(Contains a careful discussion of Nāgārjuna's treatment of non-contradiction and excluded middle.)

Siderits, M. (1980) 'The Madhyamaka Critique of Epistemology I', *Journal of Indian Philosophy* 8: 307-35. (Discusses the debate between Nāgārjuna and classical Indian realists over the possibility of a theory of

knowledge.)
Streng, F.J. (1967) *Emptiness: A Study in Religious Meaning*, Nashville, TN: Abingdon. (Clear presentation of the anti-realist interpretation of Nāgārjuna.)
Tuck, A.P. (1990) *Comparative Philosophy and the Philosophy of Scholarship: On the Western Interpretation of Nāgārjuna*, New York: Oxford University Press. (Discusses the variety of interpretive approaches to Nāgārjuna taken by modern scholars.)
Wood, T. (1994) *Nāgārjunian Disputations: A Philosophical Journey Through an Indian Looking Glass*, Monographs of the Society for Asian and Comparative Philosophy 11, Honolulu, HI: University of Hawaii Press.(Defends a nihilist interpretation of Nāgārjuna’s doctrine of emptiness.)

Nagel, Ernest (1901-85)

Ernest Nagel was arguably the pre-eminent American philosopher of science from the mid 1930s to the 1960s. He taught at Columbia University for virtually his entire career. Although he shared with Bertrand Russell and with members of the Vienna Circle a respect for and sensitivity to developments in mathematics and the natural sciences, he endorsed a strand in the thought of Charles S. Peirce and John Dewey that Nagel himself called 'contextual naturalism'. Among the main features of contextual naturalism is its distrust of reductionist claims that are not the outcomes of scientific inquiries.

Nagel's contextual naturalism infused his influential, detailed and informed essays on probability, explanation in the natural and social sciences, measurement, history of mathematics, and the philosophy of law. It is reflected, for example, in his trenchant critiques of Russell's reconstruction of the external world and Russell's epistemology as well as cognate views endorsed at one time or another by members of the Vienna Circle.

1 Life

Ernest Nagel was born on 16 November 1901 in Novomesto, Slovakia. He came to New York City in 1911. Educated in the New York City school system, he received a BS degree from The College of the City of New York in 1923. While teaching mathematics in the New York City public schools, he received an MA from Columbia University in 1925 and a Ph.D. in 1931 for a dissertation on the logic of measurement. After spending one year as an instructor at The College of the City of New York in 1930-1, he moved to Columbia University in 1931. He became a full professor in 1945, was appointed to the John Dewey professorship in 1955 and to a University professorship from 1967 until he retired in 1970. He served as an editor of *The Journal of Philosophy*, *Philosophy of Science* and *The Journal of Symbolic Logic*. He died on 20 September 1985.

2 Philosophy of science

According to Nagel, science, understood as an 'institutionalized art of inquiry' has contributed to the articulation and realization of 'aspirations generally associated with the idea of a liberal civilization' (1961: vii). As a consequence, the relatively recent identification of philosophy of science as a special branch of study should not let us forget that topics covered under this rubric 'are continuous with those that have been pursued … for centuries under such traditional divisions of philosophy as "logic", "theory of knowledge", "metaphysics" and "moral and social philosophy"'. Nagel wrote interestingly on all of these subjects; but his most extensive and best-known contributions are focused on topics in what he called 'the logic of science'. As with the later Peirce and with Dewey, Nagel understood logic to cover methodological issues as well as formal syntactic and semantic questions. He thought the logic of science could be divided conveniently into three parts: (a) the nature of scientific explanations, including not only their formal characterization but also the relations of the different types of explanation to one another, their functions in inquiry and how they contribute to systematizing knowledge; (b) the structure of scientific concepts, including their linkages to observational data, conditions of meaningfulness and explication via methods of definition and measurement; (c) evaluation of claims to knowledge, including the structure of probable inference, criteria for weighing evidence and the validation of inductive arguments. Nagel's best-known and most extensive publications have focused on topics related to explanation, probability and probable reasoning, but his many articles cover aspects of all three parts of this classification (see Scientific method).

3 Observation and theory

In 'Verifiability, Truth and Verification' of 1934 (Nagel 1956: ch. 7) and his critique of Russell's phenomenalism in the 1940s, Nagel insisted that the reports articulating the observed results of experiment and controlled observation used in scientific inquiry assume background information. In this sense, they are 'theory laden'. However, in*The Structure of Science* (1961) and in his comments on M. Hesse and P. Feyerabend in the late 1960s and early 1970s, Nagel denied that understanding reports of outcomes of tests of a hypothesis need or should ever presuppose the hypothesis under test. Laws that are explained by a given theory are experimental in that context when their terms may be interpreted observationally and can be confirmed or disconfirmed without presupposing the theory. Universal generalizations within a theory that cannot meet these context-relativized requirements for

being experimental utilize concepts whose 'structural' features are characterized by the basic principles of the theory (that Nagel thought could be formally presented as a 'calculus' at least in principle). How such structural features are characterized is first elaborated in two papers Nagel wrote in the 1930s, on the emergence of the notion of pure geometry as distinct from a theory of spatial relations and pure mathematics as distinct from a science of quantity. Nagel sought to bypass the conflict between his two teachers, the realist M.R. Cohen and the instrumentalist J. Dewey, by arguing that realist and instrumentalist understandings of scientific theories and theoretical entities do not reflect a substantive disagreement (see Scientific realism and antirealism; Observation; Laws, natural).

4 Explanation

Nagel claimed that the 'distinctive aim' of scientific inquiry is 'to provide systematic and responsibly supported explanations'. Because explanation is a distinctive aim of inquiry, Nagel recognized that a clarification of the common features of scientific explanations is also a clarification of common features of the aims of scientific inquiries. Consequently, an account of scientific explanation is inevitably going to require an examination of pragmatic aspects of scientific explanation. For example, Nagel insisted that deductive explanations of laws require that the law to be explained be less general than the explaining laws in a manner that required reference to the intended domains of application of the laws whose levels of generality are being compared.

In addition to deductive explanation, Nagel wrote on probabilistic, functional and statistical explanation. According to his classic discussion of goal-directed systems and functional explanation in biology and the social sciences, the claim that a system *S* is goal directed with respect to goal *G* is explained by showing it to be a logical consequence of a set of explanatory premises including causal laws that show how goal state *G* is realized relative to a variety of antecedent conditions. Ascribing a function *F* to the system *S* presupposes that *S* is goal directed to some goal and that *F* contributes to the realization of that goal. Nagel applied his account of functional explanation both to biology and to anthropology and sociology where he undertook a critical examination of the notions of functional explanation promoted by Malinowski and Radcliffe-Browne (see Functionalism in social science).

Nagel's relativization of the deterministic or nondeterministic character of theories to how states of systems characterized by such theories and their laws are specified allowed him to claim that theories such as classical statistical mechanics may be viewed as deterministic when the states of aggregates are represented statistically and as indeterministic when the states are represented as mechanical states (see Determinism and indeterminism). The contextual naturalism infusing his informed discussion of reduction of one theory to another is manifested by the remark: 'if and when the detailed physical, chemical and physiological conditions for the occurrence of headaches are ascertained, headaches will not thereby be shown to be illusory' (1961: 366) (see Explanation; Reduction, problems of).

5 Probability, confirmation and induction

Nagel's classic monograph *Principles of the Theory of Probability* (1939) is a philosophical survey of issues pertaining to applications of the calculus of probability as these issues were seen in the 1940s and 1950s. Nagel tended to favour the limit of relative frequency interpretation of Venn and Von Mises as a way of understanding applications of the calculus of probability within scientific theories. Although he was sympathetic to efforts to clarify conceptions of weight of evidence and evidential support, he was a trenchant critic of the efforts of Keynes, Carnap and Reichenbach to use various interpretations of the calculus of probability to develop notions of degree of confirmation and evidential support. In this connection, the methodological parts of Nagel's collaboration with M.R. Cohen, *An Introduction to Logic and Scientific Method*, should also be mentioned (see Probability, interpretations of).

6 Other topics

Nagel wrote extensively on the methodology of the social sciences as well as the natural sciences. He was concerned to defend the autonomy of the aims of scientific inquiry from those who insisted that science was laden with moral, political or social agenda. He was a sceptical critic of psychoanalytic theory, and had interesting things to say about philosophy of law.

Nagel shared with many members of the Vienna Circle the notion that extending the methods of inquiry used in

the natural and social sciences to social and political problems would promote an approach to such problems free of the totalitarian tendencies so prevalent in Europe during much of his career. However, his association with the logical empiricists did not reflect a sympathy with the their positivist tendencies. He was, as he put it, a contextualist naturalist throughout.

See also: Logical positivism; Naturalized philosophy of science

ISAAC LEVI

List of works

Cohen, M.R. and Nagel, E. (1934) *An Introduction to Logic and Scientific Method*, London: Routledge & Kegan Paul.(Widely used textbook for two decades, which pioneered in presenting elementary introductions to topics in propositional logic. The discussion of the distinction between deductive and inductive reasoning, Mill's methods and elementary topics in statistics continue to have interest.)

Nagel, E. (1939) *Principles of the Theory of Probability*, International Encyclopedia of Unified Science, vol. 1, number 6, Chicago, IL: University of Chicago Press.(A survey of a variety of interpretations of the calculus of probability focused primarily on controversies concerning the domain of application of frequency interpretations.)

Nagel, E. (1954) *Sovereign Reason*, Glencoe, IL: Free Press.(A selection of essays and reviews including critical discussions of Peirce, Dewey, Whitehead, Russell, Eddington, and a sustained critique of Reichenbach's views on probability that retains its relevance.)

Nagel, E. (1956), *Logic without Metaphysics*, Glencoe, IL: Free Press.(A selection of essays and reviews including the well-known 'Logic without Ontology' and 'A Formalization of Functionalism' together with essays - especially chapters 7, 8 and 9 - that should afford a more accurate picture of Nagel's relations to analytic philosophy and the logical empiricists than has been offered in recent commentaries on his work.)

Nagel, E. and Newman, J.R. (1960) *Gödel's Proof*, New York: New York University Press.(An effort at providing a popular presentation of Gödel's incompleteness theorem.)

Nagel, E. (1961) *The Structure of Science: Problems in the Logic of Scientific Explanation*, New York: Harcourt Brace.(Presents Nagel's classification of types of explanation in science and discussions of them, his conception of theories and their cognitive status, determinism, reduction, teleological, functional and statistical explanation in biology and social science, subjectivity and value in social inquiry, methodological individualism and a chapter on the 'logic' of historical inquiry.)

Nagel, E. (1979) *Teleology Revisited and Other Essays in the Philosophy and History of Science*, New York: Columbia University Press.(Contains two classic essays on the history of mathematics, Nagel's John Dewey Lectures of 1977, a critique of Carnap's theory of inductive logic and several other essays including a discussion of Popper and Feyerabend.)

References and further reading

Morgenbesser, S., Suppes, P and White, M. (eds) (1969) *Science and Method: Essays in Honor of Ernest Nagel*, New York: St Martin's Press.(Contains a full bibliography.)

Reck, A.J. (1968) *The New American Philosophers: An Exploration of Thought Since World War II*, Baton Rouge, LA: Louisiana State University Press; repr. New York: Dell, 1970.

Nagel, Thomas (1937-)

The comprehensiveness of Thomas Nagel's approach to philosophy sets him apart among late-twentieth-century analytic philosophers. Nagel develops a compelling analysis of the fundamental philosophical problems, showing how they result from our capacity to take up increasingly objective viewpoints that detach us from our individual subjective viewpoints as well as from the viewpoints of our community, nation and species. Our essentially dual nature, which allows us to occupy objective as well as subjective viewpoints, poses unsolvable problems for us because subjective and objective viewpoints reveal conflicting facts and values. Our ability to undertake increasingly detached viewpoints from which objective facts come into view indicates that we are contained in a world that transcends our minds; similarly, our ability to examine our values and reasons from a detached or impartial objective viewpoint implies that moral values are real in the sense that they transcend our personal motives and inclinations. Yet Nagel also holds that our capacity for objective thought is limited by the fact that we cannot detach ourselves completely from our own natures in our attempts either to know our world or to act morally. Subjective facts are equally a part of reality and our moral outlook is essentially the outlook of individual agents with personal and communal ties. Consequently, Nagel argues against any form of reductionism which holds that only objective facts and values are real or which attempts to explain subjective facts and values in terms of objective ones.

1 Life

Thomas Nagel was born in 1937 in Belgrade, Yugoslavia, becoming a naturalized US citizen in 1944. He was educated at Cornell University, Oxford University and Harvard University. His career began at the University of California, Berkeley, where he spent three years before moving to Princeton University in 1966 and then to New York University in 1980.

Nagel stands out among his contemporaries for the breadth of his work, which encompasses metaphysics, epistemology and moral philosophy, and for his development of a comprehensive realistic approach. He is also distinguished by the nontechnical nature of his argumentation, which is accessible to non-professionals.

2 Metaphysics

Nagel's realism is complex in that it advances one intuitive idea at the core of the realist outlook, the idea of a world that contains and radically transcends us, at the cost of the intuitive and equally core idea of a single objective reality. For Nagel, the 'pure idea of realism' is the idea 'that there is a world in which we are contained' (1986: 70). He understands the idea of containment in terms of mind-independence. Thus, the world is independent of us not only causally, but also epistemically: the way things are might completely transcend our ability to know them, even to think about them. His development of this idea in terms of the distinction between subjective and objective standpoints entails that there is no single way that things are in themselves. If, as Nagel holds, reality includes all the myriad subjective viewpoints and the subjective facts they make available, then there is no one way that things are. It is not possible to give a single account from the completely detached or objective view, because such an account would fail to include subjective facts (that are only available from irreducibly distinctive subjective viewpoints).

Hence Nagel's approach is not only anti-verificationist and anti-idealist in its commitment to the intuitive idea of a mind-independent world. His realism is also anti-reductionist, because of its denial of the idea of a single objective reality. As both reductionism and idealism are prevailing tendencies in contemporary philosophy, Nagel's views on many issues challenge the prevailing positions. On the one hand, Nagel objects that many contemporary approaches, such as that of the later Wittgenstein, are essentially idealist in that they involve 'a broadly epistemological test of reality'. On the other hand, as we will see, Nagel opposes a variety of reductionist accounts by arguing that the phenomena at issue are subjective facts or values that cannot be described or explained from a more objective standpoint in terms of objective facts or values. Nagel's anti-reductionist arguments thus depend on whether he is right that the recalcitrant phenomena are essentially subjective (see Realism and antirealism).

3 Theory of knowledge

Nagel's realist construal of the distinction between subjective and objective viewpoints affirms a traditional

understanding of the problem posed for theories of knowledge that does not allow for revisionist resolution. The problem, according to Nagel, is that while knowledge needs to be objective and so requires that we undertake increasingly objective standpoints, we can only do so to a limited extent, since any outlook that we can occupy will be to some extent subjectively bounded.

Since we cannot attain the completely detached viewpoint from which we could examine our own attempts at understanding the world, Nagel affirms the possibility of sceptical doubt, opposing the anti-sceptical trend which holds that sceptical worries are unwarranted or even incoherent because they impose a theoretical demand for justification of our system of beliefs as a whole from an unobtainable, completely objective standpoint. Moreover, since Nagel urges that knowledge should be of what is correct, even if it is inaccessible to us, he opposes reductive theories which attempt to meet scepticism by reducing the aim or scope of knowledge to concern the way the world appears from the limited objective viewpoints that we can undertake rather than the way it really is. Though he believes that the possibility of limited objectivity is associated with the possibility of scepticism, Nagel endorses a heroic approach to knowledge, like that of Descartes, which acknowledges that we cannot detach from our perceptual and cognitive faculties completely, but attempts to use a priori argumentation to form a conception of our situation as knowers none the less (see Descartes, R. §3; Scepticism).

4 Theory of mind

Commitment to the irreducibility of subjective facts leads Nagel to argue that the mind-body problem could only be solved, if at all, by a complete and as yet unimaginable revision of our conception of the basic constituents of the universe. Nagel explains that the relation of the mind and the body seems problematic if, as a result of the success of physical science, one holds a physical conception of objective reality, and espouses a reductionist programme according to which all phenomena must be explainable reductively in terms of physical objective facts. In opposition, he argues that the qualitative nature of conscious experience is a subjective feature of reality available only from a first-person subjective viewpoint. Hence, though the qualitative aspects of conscious experiences are part of reality, they cannot be explained in terms of objective facts in general or physical objective facts in particular.

The precise force of Nagel's denial of the possibility of psychophysical reduction is important. He is not denying that we are physiological organisms or that the central nervous system (in vertebrates) and its physical states is necessary for subjective mental states. He is not denying that conscious organisms are made up of the same ultimate constituents that constitute all the physical phenomena in the universe. Rather, Nagel is proposing that our conception of the ultimate constituents as physical is too narrow for explaining that some combinations of those constituents have conscious experience. He is arguing for a dual-aspect approach, according to which the ultimate constituents are dual in nature - both physical and mental- in the sense that they can be combined to form physical things which have physical properties and conscious organisms which have the subjective properties of conscious experience (see Consciousness §2; Qualia).

5 Moral philosophy

Nagel uses the distinction between subjective and objective viewpoints to articulate a non-metaphysical understanding of moral realism, to explain the bifurcation in moral philosophy between consequentialist and deontological approaches and to argue against consequentialism, even while showing that the tension between more and less impartial moral outlooks often cannot be reconciled (see Consequentialism; Deontological ethics). Nagel's discussion of these issues illuminates the differences between the role of the objective standpoint in moral as opposed to factual or empirical thought.

According to Nagel, 'the view that values are real is not the view that they are real occult entities or properties, but that they are real values: that our claims about value and about what people have reason to do may be true or false independently of our beliefs and inclinations' (1986: 144). To be ethical is to examine one's reasons for acting from more than one's own particular perspective. Hence, unlike factual understanding, moral understanding 'is not a question of bringing the mind into correspondence with an external reality which acts causally on it, but of reordering the mind itself in accordance with the demands of its own external view of itself' (1986: 148). Consequently, while the truth about the empirical world may completely transcend our theoretical reason, the truth about how we should act could not completely transcend our practical reason. Hence, unlike metaphysical versions

of moral realism, Nagel's normative realism is not associated with the possibility of moral scepticism according to which the objective viewpoint shows that there are no values, only facts about people's inclinations and motives.

Since consequentialism has been understood as advocating an impartial point of view, that of an impartial spectator, Nagel's discussion of moral theory in terms of the possibility of undertaking an impartial or objective viewpoint is not entirely new. However, Nagel gives a more complete and nuanced treatment of both deontological and consequentialist approaches by distinguishing between agent-relative and agent-neutral reasons in terms of two correlative distinctions that are also generated by our dual viewpoints: the distinction between what we do and what happens and between choosing actions as opposed to choosing states of the world. Reasons that are relative to the agent are 'specified by universal principles which nevertheless refer ineliminably to features or circumstances of the agent for whom they are reasons'. Reasons that are neutral with respect to the agent 'depend on what everyone ought to value, independently of its relation to himself' (1991: 40). As agents, we act on agent-relative reasons because even though actions affect what happens in the world, in the first instance one's choice is necessarily between one's own actions. However, each of us also has an objective self which views the world in detachment from one's particular perspective. Consequently, the objective self chooses between different possible states of the world and its choice is based on agent-neutral reasons. So, moral conflict is due to the fact that 'every choice is two choices', every choice is at once a choice between actions and between states of the world. Moral conflict arises when the agent's choice concerning what to *do* conflicts with the objective self's choice concerning what should *happen*.

Nagel thus explains that consequentialism gives primacy to the agent-neutral values on the basis of which the objective self chooses between world-states. In contrast, deontological theories give primacy to certain agent-relative reasons which restrict agents from acting in certain ways. His own position is both complex and modest. It does not allow for a general championing of one kind of reason over another. Against consequentialism, Nagel argues that not all values are agent-neutral. But he also voices some uncertainty about whether there really are agent-relative reasons. However, while he does not argue for their existence directly, he explains them, especially the most problematic deontological ones, in lucid and compelling terms. The modesty of Nagel's position lies in his conviction that we are far from a developed moral outlook. In contrast to his view that subjective and objective facts are irreconcilable, Nagel believes that it is possible, in principle, to develop our agent-relative values to be more consonant with agent-neutral values. However, as we will see in Nagel's discussion of political theory, he finds little hope that the actual conditions of human life will allow us to school our moral outlook (see Moral realism).

6 Political philosophy

While political theory is typically understood as dealing with the relationship of the individual and society, Nagel holds that it deals with the relation of each individual to themselves, since each one of us occupies both a particular individual viewpoint and the detached standpoint of the collectivity. So the problem of designing morally justified social institutions that would reconcile the conflicts between the individual and society is the problem of designing social institutions that would allow each of us to reconcile those two standpoints within ourselves.

Nagel argues that a justified political system would be a unanimously supported, strongly egalitarian one, because in taking on a more objective standpoint, we see that everyone is equal not only in the sense that no person's life matters any more than any other's, but in the sense that it is more important to improve the lives of the worse off than to add to the advantages of the better off. The conflict between this impersonal, strongly egalitarian dictate and our personal motivation to lead our own lives could be resolved by general principles that can be universalized in a Kantian manner: principles that each of us can will to be universal laws. That is, we need to find our way to wanting 'to live by principles that anyone can accept' (1991: 48).

Nagel holds that we need political institutions to help integrate and develop our dual motivations. He suggests that the 'general form' of a solution is 'a moral division of labour between individuals and institutions'. According to such a division, the social institutions in which we participate would allow us to act on our impartial, egalitarian motives. This would allow us to act on our personal motives outside of our social roles. However, Nagel argues pessimistically against the possibility of developing such institutions given the initial duality of motives. In sum, since he is drawn to a strongly egalitarian social ideal, to whose recognition the duality of standpoints is necessary,

but to whose realization the duality of standpoints seems to present great obstacles, Nagel does not see how to embody that ideal in a morally and psychologically viable system (see Equality).

See also: Reduction, problems of §2

SONIA SEDIVY

List of works

Nagel, T. (1970) *The Possibility of Altruism*, Oxford: Clarendon Press.(Incorporates material from Nagel's B.Phil. and Ph.D. theses. Focuses on the motivational basis for morality, arguing that altruism is a basic rational requirement on moral action. This first work gives a stronger role to the objective standpoint and its deliverances than does Nagel's subsequent work in moral philosophy.)

Nagel, T.(1979) *Mortal Questions,* Cambridge: Cambridge University Press.(Essays on a wide variety of topics that were published between 1969 and 1979, including the highly influential 'Moral Luck' and 'What is it like to be a bat?'. Several essays deal with issues of public policy and are motivated by events and policies associated with the Vietnam War.)

Nagel, T., Cohen, M. and Scanlon T. (eds) (1980) *Marx, Justice and History*, Princeton, NJ: Princeton University Press.(Essays by a variety of authors in *Philosophy and Public Affairs* from 1972 to 1979 that bring together contemporary analytic moral philosophy and Marxist social theory.)

Nagel, T. (1986) *The View From Nowhere*, Oxford: Oxford University Press.(Presents Nagel's systematic and comprehensive treatment of the fundamental philosophical issues in terms of the essential duality between objective and subjective points of view and the distinctive facts and values that they make available.)

Nagel, T. (1987) *What Does It All Mean?*, Oxford: Oxford University Press.(A concise and 'direct' or nonhistorical introduction to nine basic philosophical problems.)

Nagel, T.(1991) *Equality and Partiality*, Oxford: Oxford University Press.(Incorporates Nagel's 1990 John Locke Lectures. A detailed development of Nagel's proposal that the central problem of political theory- that of reconciling the conflicts between the individual and society - is posed by the division between these standpoints within each individual and needs to be addressed accordingly.)

Nagel, T. (1995) *Other Minds: Critical Essays, 1969-1994*, Oxford: Oxford University Press.(These essays focus on the views of particular philosophers in the philosophy of mind and in ethics and political theory. In the philosophy of mind, Nagel addresses the views of Freud, Wittgenstein, Chomsky, Fodor and Dennett among others. In ethics and political theory, Nagel treats the views of philosophers such as Rawls, Nozick, MacIntyre and Williams as well as Schelling and Kolakowski. The book opens with a fascinating intellectual autobiography entitled 'Introduction: The Philosophical Culture'.)

References and further reading

Darwall, S. (1987) 'The View From Nowhere', *Ethics* 98: 137-57.(Focuses on Nagel's normative realism, §5 of this entry. Criticizes the account of agent-relative reasons while arguing that Nagel's views on autonomy provide the resources for a successful account.)

McGinn, C. (1987) 'The View From Nowhere', *Mind* 96: 263-72.(Focuses primarily on Nagel's discussion of the mind-body problem, §4 of this entry.)

Peacocke, C. (1989) 'No Resting Place: A Critical Notice of The View From Nowhere by Thomas Nagel', *Philosophical Review* 98: 65-82.(Focuses on Nagel's account of conscious states and his normative realism, §§4 and 5 of this entry.)

Raz, J. (1990) 'Facing Diversity: The Case of Epistemic Abstinence', *Philosophy and Public Affairs* 19: 3-46. (Criticizes John Rawls' as well as Nagel's attempt to accommodate the diversity of views in pluralistic societies with the notion that justified political arrangements and actions require individual consent, §6 of this entry.)

Nahmanides, Moses (1194-1270)

One of the most influential medieval Jewish thinkers to engage with the philosophical tradition, Nahmanides was also a leading Talmudist, biblical exegete, and a founding figure of the Jewish mystical tradition (Kabbalah) that emerged in thirteenth-century Spain. Generally critical of Aristotle, he was deeply influenced by predecessors such as Moses Maimonides. As the leading rabbi in Catalonia, Nahmanides played a central role in the 'Maimonidean controversy' of 1232-3, a dispute that raged over the permissibility of philosophical study. He was also at the centre of the 'Barcelona disputation' of 1263, conducted with the apostate Pablo Christiani over the issue of philosophically motivated allegories.

Unlike his Provençal contemporaries, Nahmanides wrote neither free-standing philosophical treatises nor commentaries on Graeco-Arabic philosophical texts. Instead, he developed his original metaphysical views in sermons and his highlt influential biblical commentaries treating such topics as miracles, providence and idolatry. The commentaries proved especially influential through their thematic treatments of philosophical questions raised by the biblical text and for their suggestive expositions of mystical and theosophical ideas.

1 Relation to the philosophical tradition

Moses Nahmanides, Rabbi Moses ben Nahman, known in Hebrew texts by the acronym 'Ramban' and in Christian literature as Bonastrug de Porta was born in Gerona, Catalonia, and died at Acre in Palestine. He became the leading Catalonian thinker in the intellectually tumultuous period following the publication of Maimonides'*Guide to the Perplexed*, whose Hebrew translation in 1204 by Samuel ibn Tibbon brought about the first major encounter of Jewish thinkers in Christian Spain and Provence with Graeco-Arabic and specifically Aristotelian philosophy. Although standardly portrayed as a conservative defender of rabbinic traditions, Nahmanides showed a nuanced ambivalence towards rationalism in general (which had a long indigenous Andalusian tradition) and the thought of Aristotle and Maimonides in particular. In the 'Maimonidean controversy' of 1232-3 Nahmanides joined ranks with the anti-rationalist camp to forbid unrestricted public study of philosophy, but he also vigorously defended Maimonides and attempted to prevent the banning of the *Guide* and Maimonides' philosophical oopening sections of his fourteen-volume code of Jewish law, the *Mishneh Torah*. Comparable dualities run throughout his thought. He repeatedly criticized Aristotle for his empiricism and doctrine of the world's eternity and reproved Maimonides for his challenges to rabbinic theological traditions. But his own doctrines are constructively influenced by the philosophy of the *Guide*. His decision to present his ideas primarily in commentaries on Scripture is exemplary of his outlook: philosophical reasoning may be essential to the strengthening of belief but it must remain subservient to divine revelation and the revealed tradition. Nahmanides repeatedly emphasizes the importance of rational understanding. In *Sha'ar ha-Gemul (The Gate of the Reward)*, he recalls the words of Saadiah Gaon, in *Kitab al-mukhtar fi'l-amanat wa'l-'i'tiqadat (The Book of Critically Chosen Beliefs and Opinions)*, attacking those who spurn philosophical reflection as 'fools who despise wisdom' (see *Kitvei ha-Ramban* II: 21).

Nahmanides' cautiously critical stance towards philosophy stands out in his influential views on midrash *aggadah*, the immense literature of non-legal (non-*halakhic*) rabbinic exegesis, homilies and imaginative writing (see Midrash), whose interpretation and authority were centrally at issue in both the Maimonidean controversy and the 'Barcelona disputation' of 1263. In the latter event, Nahmanides was summoned to engage in a public argument with the apostate Pablo Christiani. Maimonists viewed the rabbinic midrash (and the non-legal sections of the Hebrew Bible) as philosophical allegories, contemplation of whose truths was the highest form of worship. Anti-Maimonists insisted on the primacy of practice, the life of the divine commandments or *mitzvot*. Rejecting philosophically motivated allegories as a potential source of antinomianism, grounded in a suspected substitution of contemplation for religious observance, they urged that midrash *aggadah* be accepted virtually at face value, even when its mythic and supernatural claims violated reason. In the Barcelona disputation, Christiani put *aggadic* and scriptural materials to allegorical use in behalf of Christianity, rendering the allegorical method all the more suspect. Nahmanides' position in both confrontations was novel. His rationalist sensibilities forbade him to interpret all *aggadah* literally. Following his Andalusian predecessor Abraham ibn Ezra, he subjected *aggadah* to the same interpretive standards of philology, grammar and reason that he applied to biblical interpretation itself. But he undercut the antinomian potential of philosophical allegorization and the evangelical impact of Christian misappropriations by denying that midrash *aggadah* is authoritative. Relying on some imaginative philology, he

argued that *aggadah* is merely homiletic - or, even less, ‘things one man tells another’. It is never authoritative in belief or practice, as are *halakhic* texts. This conservative and anti-Maimonidean approach deeply influenced later Jewish dogmatics and polemics by restoring action rather than belief as the essence of Judaism. The immediate consequences for Nahmanides were not, however, favourable. While the disputation was a theological victory for him, his Dominican opponents responded by charging him with abusing Christianity and he was forced to flee to Palestine.

Nahmanides’ turn to mystical theosophy reflects his opposition to the Maimonists’ assumption that the deepest truths of the Law are somehow coextensive with the received tradition of Aristotelian physics and metaphysics. Nahmanides presents Kabbalah as an indigenous Jewish alternative to Graeco-Arabic philosophy, a *prisca philosophia* (archaic or antidiluvean philosophy) based on a revealed (and hitherto oral) tradition rather than reason, and a fixed canon of theosophic exegeses of biblical verses and commandments (see Kabbalah). Thus, in place of Maimonides’ account of evil in terms of matter, he proposes a theodicy based on transmigration; instead of Maimonides’ anthropological-historical account of the Temple sacrifices, he turns to the theory of divine emanations (Sefirot).

2 Metaphysics: miracles and divine providence

Perhaps the most original and philosophically interesting of Nahmanides’ doctrines is his theory of revealed and hidden miracles (see Miracles). Revealed miracles such as the splitting of the sea in Exodus 14 violate the observed regularities of nature and cannot be explained by natural causes but must be attributed to divine intervention. Nahmanides believed that the Bible provides reliable evidence of such manifest miracles. Typically witnessed by the entire community, they provide empirical verification for the three fundamental principles of Judaism: *ex nihilo* creation, divine knowledge, and providence. Such empirical appeals, rather than rational demonstrations, are Nahmanides’ primary epistemic grounds for faith. In the case of creation, for example, he argues against Aristotelian eternalism, suggesting that if the world were eternal, nature as it is would be necessary and God would be unable to shorten the wing of a fly if he wished. But this consequence is baldly disconfirmed by the experience of revealed miracles in which God has violated presumed necessities of nature. Eternalism, then, is refuted and, Nahmanides concludes, the world must have been created *ex nihilo*. To be sure, this conclusion does not follow. As Maimonides (from whom Nahmanides drew this argument) observed, the existence of miracles does not exclude a third alternative, the Platonic theory of *formatio mundi*, creation from eternal matter. Indeed, from other statements, in which Nahmanides claims that what was created *ex nihilo* was matter, or potentiality, without form, it is not clear how well he understood Aristotelian metaphysics or its differences from Platonic or Platonizing doctrines. However, his argument for creation *ex nihilo*, which appeals to miracles as empirical data, is still of interest. After repeatedly describing Aristotle as an empiricist who believed only in what our senses perceive, Nahmanides tries to turn empiricism against Aristotle, arguing, in effect, that one cannot be both an empiricist and an eternalist.

Nahmanides’ doctrine of hidden miracles underlies his theory of providence - specifically, of divine rewards and punishments (see Providence). Such miracles violate no observed regularities and may even have obvious explanations in terms of intermediate natural causes. The same events, however, can be explained by attributing them to God, who brings them about as rewards or punishments of the absolutely righteous and wicked. Such direct attributions, Nahmanides argues, are simpler and more sensible than the naturalistic alternatives: while caused by God under conditions we may never know in full, hidden miracles follow in law-like patterns as consequences of human actions; and because they occur only in response to the acts of absolutely righteous and wicked persons, not the ordinary individuals who fall between these two extremes, Nahmanides can uphold a nature whose laws govern these ordinary people. So despite some strong rhetoric in which he appears to deny the existence of nature *tout court*, Nahmanides does not seem to be an occasionalist. Rather, his doctrine seems to expand upon Maimonides’ theory of ‘natural’ providence, whereby all species (and ordinary individuals as their members) are endowed with natures to provide for their wellbeing, and certain perfected individuals can transcend the vicissitudes of nature, achieving an additional kind of individual providence. The main differences from Maimonides here turn on the nature of that individual perfection and on the role of miracles. For Maimonides, individual providence is consequent on intellectual development: only those who become ‘acquired intellects’ rise (in that measure) above matter and the conditions of natural evil. Miracles have no place in this process. Nahmanides, by contrast, makes individual reward and punishment consequent on moral character and religious

practice. He offers no metaphysical explanation of its operations but makes miracles its unseen vehicles. He takes over the structure of Maimonides' account but substitutes more traditional rabbinic categories for Aristotelian ones.

Paralleling the individual providence that attends the actions of the absolutely righteous and wicked, Nahmanides affirms a special providence over the nation and land of Israel. This special providence is described by Nahmanides in relation to his views on astrology and astral determinism, theories shared by many of his Andalusian predecessors and contemporaries, with the notable exception of Maimonides. On Nahmanides' theory, there exist real forces associated with the constellations created by God to govern the nations. There is also a hierarchy of angels and demons. The people and land of Israel, however, remain directly under the dominion of God. It is forbidden as idolatry for Jews to worship any other power, not because such powers do not exist but because no other power is the rightful Lord of Israel. Similarly with the complex hierarchy of powers, Kabbalistic Sefirot, *within* the deity. It is a kind of idolatry to worship any of these emanations apart from the deity as a whole. Such worship is tantamount to denying God's unity.

3 Hermeneutics and explanation of the divine commandments

Nahmanides pursues three distinct levels of biblical exegesis with exceptionally systematic care. He typically offers (1) a literal interpretation (*peshat*) according to the philologically determined senses of words, syntax, historical context and narrative structure; (2) an interpretation containing moral or political wisdom (*hokhmah*); and (3) a riddle-like secret theosophic interpretation (*sod*), which he calls 'the way of truth'. At this third level, he was the first to incorporate Kabbalistic mysteries, even if only allusively, in an exoteric commentary. Some of his most original and enduring exegetical contributions, however, work at the first two levels. He was the first, and perhaps the only, medieval Jewish exegete to offer typological interpretations like those found in Christian exegesis. In reaction to Maimonides' radical devaluation of literal meanings, he also did much to restore the status of *peshat* and to enrich its traditional scope, making it more thematic and conceptual, and at times even subsuming the philosophical and mystical exegeses of the text.

Like Maimonides, Nahmanides was a vigorous advocate of inquiry into the reasons for the *mitzvot*, although he often disagreed with the explanations of the *Guide to the Perplexed*. Applying his exegetical method, he argued that, besides their moral, political and commemorative purposes, the commandments have an inner theurgic function in serving divine needs. Here, as in his exegesis, his multi-levelled explanations, ranging from the popular to the esoteric, and his brilliance as a stylist made Nahmanides an exceptionally influential writer.

See also: Bible, Hebrew; Kabbalah; Maimonides, M.; Midrash; Providence

JOSEF STERN

List of works

Nahmanides, Moses (1194-1270) Works, ed. C.B. Chavel, *Kitvei Ha-Ramban (Writings and Discourses of Nahmanides)*, Jerusalem: Mosad Ha-Rav Kook, 1963, 2 vols; trans. C.B. Chavel, New York: Shilo, 1978.(A collection of Nahmanides' speculative writings, including his sermons, a record of the 'Barcelona disputation', and a commentary on the Book of Job, as well as some misattributed works. The translation is not always sensitive to philosophical usage.)

Nahmanides, Moses (1194-1270) *Perush Ha-Ramban 'al ha-Torah (Commentary on the Torah)*, ed. C.B. Chavel, Jerusalem: Mosad ha-Rav Kook, 1959-63, 2 vols; trans. C.B. Chavel, *Commentary on the Torah*, New York: Shilo, 1976.(Nahmanides' highly influential commentary on the Torah.)

References and further reading

Berger, D. (1983) 'Miracles and the Natural Order in Nahmanides', in I. Twersky (ed.) *Rabbi Moses Nahmanides (Ramban): Explorations in His Religious and Literary Virtuosity*, Cambridge, MA: Harvard University Press. (An excellent analysis of Hahmanides' theory of revealed miracles, his theory of providence, and conception of nature.)

Funkenstein, A. (1982) 'Nahmanides' Symbolical Reading of History', in J. Dan and F. Talmage (eds) *Studies in Jewish Mysticism*, Cambridge, MA: Association for Jewish Studies.(An insightful study of Nahmanides' use of typological interpretation.)

Idel, M. (1983) 'We Have No Kabbalistic Tradition on This', in I. Twersky (ed.) *Rabbi Moses Nahmanides*

(Ramban): Explorations in His Religious and Literary Virtuosity, Cambridge, MA: Harvard University Press. (A careful discussion of Nahmanides on the relations of mysticism and philosophy.)

Maimonides, Moses (*c*.1185) *Mishneh Torah (The Law in Review)*, Books I and II trans. M. Hyamson, *The Book of Knowledge and the Book of Adoration*, Jerusalem: Feldheim, 1974; Books III-XIV trans. as *The Code of Maimonides*, Yale Judaica Series, New Haven, CT: Yale University Press, 1949-72.(Maimonides' highly influential fourteen-volume codification of rabbinic law, which ranges from beliefs through ritual and civil law to the Temple cult and laws of war and political governance.)

Maimonides, Moses (*c*.1190) *Dalalat al-Ha'irin (Moreh Nevukhim, Guide to the Perplexed)*, ed. S. Munk, *Le Guide des Égarés*, Arabic text, critically edited, with annotated French translation, Paris, 1856-66, 3 vols; repr. Osnabrück: Zeller, 1964; trans. S. Pines, with an introductory essay by L. Strauss, *The Guide of the Perplexed*, Chicago, IL: University of Chicago Press, 1969.(Maimonides' major philosophical work, which served as the starting-point for all subsequent medieval Jewish philosophical thought.)

Novak, D. (1992) *The Theology of Nahmanides Systematically Presented*, Atlanta, GA: Scholars Press.(Exposition of Nahmanides' thought on the soul, commandments, miracles, eschatology, and other topics, with key texts culled from his works and organized thematically.)

Saadiah Gaon (early 10th century) *Kitab al-mukhtar fi 'l-amanat wa 'l-'i'tiqadat (The Book of Critically Chosen Beliefs and Opinions)*, trans. J. Kafih, *Sefer ha-Nivar ba-Emunot uva-De'ot*, Jerusalem: Sura, 1970; trans. S. Rosenblatt, *The Book of Beliefs and Opinions*, New Haven, CT: Yale University Press, 1948.(The passage to which Nahmanides refers can be found on p. 13 of the Rosenblatt translation. This is the first systematic work of Jewish rational theology; it shows the influence of Islamic *kalam*.)

Septimus, B. (1983) '"Open Rebuke and Concealed Love": Nahmanides and the Andalusian Tradition', in I. Twersky (ed.) *Rabbi Moses Nahmanides (Ramban): Explorations in His Religious and Literary Virtuosity*, Cambridge, MA: Harvard University Press.(A fine historical essay on Nahmanides' place in Spanish Jewish thought, with special attention to his rationalism, literary style and role in the 'Maimonidean controversy'.)

Stern, J. (1995) 'Nahmanides' Conception of Ta'amei Mitzvot and its Maimonidean Background', in D. Frank (ed.) *Community and Covenant: New Essays in Jewish Political and Legal Philosophy*, Albany, NY: State University of New York Press.(On the relation of Nahmanides to Maimonides, with special attention to their accounts of the reasons for the commandments.)

Wolfson, E.R. (1989) 'By Way of Truth: Aspects of Nahmanides' Kabbalistic Hermeneutic', *AJS Review* 14: 103-78.(An important analysis of Nahmanides' exegetical method.)

Nancy, Jean-Luc (1940-)

Jean-Luc Nancy has disclosed significant political and social dimensions to the general project of deconstructing Western philosophy. Existence does not precede essence, according to Nancy; existence is without essence, and it is therefore impossible both to represent existence and to exist alone. Being is always 'being in common'. The task of philosophy consists in rethinking the commonality of being without relying on any prior conception of identity, unity or wholeness. 'Being in common' means that nothing - no substance, no identifiable trait- is held in common. The absence of a common substance or spiritual identity does not then generate a command to make up for this lack by means of socially useful work. As the exposure of each singularity to its ungrounded commonality, 'being in common' is the always surprising 'fact' upon which all of Nancy's investigations into philosophy, literature, psychoanalysis and political phenomena are oriented.

As a professor of philosophy at the University of Human Sciences in Strasbourg, Jean-Luc Nancy remains on the periphery of contemporary academic philosophy in France. The proximity of Strasbourg to Germany is reflected in many aspects of his work. Not only has Nancy translated seminal texts of German philosophy and literary theory, his writings constantly engage those post-Kantian German thinkers - Hegel, Marx, Nietzsche, Freud, Heidegger - who declare, each in his own way, that philosophy has come to an end. Both his critical commentaries on philosophical texts and his attempts to rethink the terms in which political and social philosophy are cast take their point of departure from what Heidegger called 'the end of philosophy'. This phrase means, for Nancy, that the resources of representation have been exhausted: philosophy can no longer claim to bring the essence of a phenomenon into view on the basis of a unified subject of any sort. Since philosophy has come to an end, and since it cannot be simply overcome, there is no choice but to repeat the history of philosophy. There are two ways to undertake this repetition: either ignore the end of philosophy and proceed to rework previous philosophical positions, or expose philosophy to its end and thus develop a 'finite' thought of community, sharing, meaning, freedom - to name a few of the topics Nancy has explored in detail. By re-examining these topics from the perspective of what Jacques Derrida has called the 'closure of metaphysics', Nancy has made a significant contribution to the general project of deconstructing the history of Western metaphysics.

Nancy's readings of philosophical texts closely match those of Derrida. What distinguishes his writings from other versions of 'deconstruction' (see Deconstruction) is his decision to rethink certain topics often associated with existentialism (see Existentialism). By reconsidering such topics as existence, abandonment, embodiment, freedom, community and communism on the basis of a more rigorous reading of Hegel and Heidegger than any of the existentialists could claim, Nancy has been able to disclose significant political and social dimensions to the project of deconstruction. His attempts to rethink these terms are all oriented on one 'fact': existence is without essence. Nancy does not proclaim, as does Sartre, that 'existence precedes essence' but that existence precedes, exceeds and succeeds itself, for existence is sheer non-coincidence with itself. From its inception, philosophy has interpreted the 'fact' that existence precedes itself by representing it in terms of a subject that underlies and is thus prior to its various attributes and accidents. Even when the subject is viewed as an ideal or as an incomplete project, the basic structure of 'infinite' and thus 'onto-theological' thought remains intact: the subject grounds itself in its prior or future essence. The 'finite thought' Nancy proposes never proceeds beyond the 'fact' that existence is without essence and without ground. Forever receding from representation, existence consists in an exposure to the groundlessness of being.

One of the most innovative aspects of Nancy's work is the way in which it brings the existential analysis of *Being and Time* into conjunction with Heidegger's late reflections on *Ereignis* ('event of appropriation') (see Heidegger, M. §5). Nancy presents being (*être*) not as the essence of things or the ground of phenomena but as a 'free' and 'generous' event: being takes place as existence. Since existence cannot support itself, it can never be alone or on its own. 'Being in common' thus defines 'finite' existence. No one participates in any*thing* when existence takes part in, or 'shares' being. Existence can even be understood as co-exposure to the absence of anything held in common as long as this absence is not then represented as something communal or 'socially useful' work could seek to repair. If 'being in common', which Nancy associates with 'the political' as opposed to 'politics', is represented in terms of a programme for securing what society lacks, it - and therefore the 'political' - are destroyed. Drawing on Kant's thesis of radical evil, Nancy presents the destruction of 'being in common' in

L'Expérience de la liberté (1988b) (*The Experience of Freedom*, 1993) as 'wickedness'.

Many of Nancy's later writings explore the bodily character of 'being in common'. Thought itself is presented as bodily: to think (*penser*) is not simply to weigh (*peser*) various options but to be weighed down. The kind of thinking that takes its point of departure from the 'fact' that existence cannot be represented is indistinguishable from touching, for, according to Nancy, to touch and to be touched are two modes of being at the limit of representation. Nancy thus counters the Kantian conception of thought as spontaneity with a 'passive' thinking whose 'passion' consists in an exposure to an always antecedent sense. *Sens* ('sense', 'meaning'), according to Nancy, precedes and exceeds all signification; it cannot be grasped, granted, discovered or produced. As long as the body is the locus of sense, it cannot be securely located, only absolved, shared, and parcelled out: 'There is no whole, no totality of the body - but its absolute separation and sharing out. There is no such thing as *the* body. There is no body' (1993b: 207). All of Nancy's explorations of the political and social dimensions of deconstruction assign themselves the same task as his writings on 'the body': they gesture toward what allows for and what destroys 'being in common'.

See also: Deconstruction; Heidegger, M.;Lacoue-labarthe, P.

PETER FENVES

List of works

Nancy, J.-L. and Lacoue-Labarthe, P. (1972) *Le Titre de la lettre*, Paris: Éditions Galilée; trans. D. Pettigrew and F. Raffoul, *The Title of the Letter: A Reading of Lacan*, Albany, NY: State University of New York Press, 1992.(One of the best and most difficult analyses of Lacan's thought, that shows its relation to Heidegger's conception of truth as the interplay of concealment and unconcealment.)

Nancy, J.-L. (1973) *La Remarque spéculative: un bon mot de Hegel (The Speculative Remark: A Hegelian Witticism)*, Paris: Éditions Galilée.(An analysis of the relation between certain 'speculative' terms in Hegel's work - especially *Aufhebung* - and his explorations of certain excessive explosions of 'wit' and 'frivolity'.)

Nancy, J.-L. (1976) *Le Discours de la syncope (The Discourse of Syncopation)*, vol. 1, *Logodaedalus*, Paris: Aubier-Flammarion.(A reading of Kant oriented towards the 'syncopation', or the punctuating rhythm, which makes it possible for consciousness to identify itself as itself but which cannot be understood either as a form or a content of consciousness.)

Nancy, J.-L. and Lacoue-Labarthe, P. (1978) *L'Absolu littéraire: théorie de la littérature du romanticisme allemand*, Paris: Éditions du Seuil; trans. P. Barnard and C. Lester, *The Literary Absolute: The Theory of Literature in German Romanticism*, Albany, NY: State University of New York Press, 1987.(A wide-ranging presentation of early German romanticism's conception of the relation between literature and philosophy. The French edition includes extensive translations of German texts.)

Nancy, J.-L. (1979) *Ego Sum*, Paris: Flammarion; 'Dum scribo' trans. I. Macleod, *Oxford Literary Review* 3 (2), 1978; 'Larvatus pro Deo' trans. D. Brewer, *Glyph* 2, 1977; 'Mundus est fabula', trans D. Brewer, *Modern Language Notes* 93, 1978.(A reading of Descartes oriented towards those 'autobiographical' moments in Cartesian texts where the indexical 'ego' marks a dislocation and interruption of 'the ego', that is, the thinking subject.)

Nancy, J.-L. (1982) *Le Partage des voix*, Paris: Éditions Galilée; 'Sharing Voices', trans. G. Ormiston in *Transforming the Hermeneutic Context: From Nietzsche to Nancy*, ed. G. Ormiston and A. Schrift. Albany, NY: State University of New York Press, 1990.(An incisive reading of Plato's *Ion* that develops out of a reading of Heidegger's brief remarks on the use of the word *hermeneus* in this dialogue.)

Nancy, J.-L. (1983) *L'Impératif catégorique (The Categorical Imperative)*, Paris: Flammarion; '''Our Probity!'' On Truth in the Moral Sense in Nietzsche', trans. P. Conner, in *Looking after Nietzsche*, ed. L. Rickets, Albany, NY: State University of New York Press.(A collection of essays, the first of which is an extensive examination of Kant's 'fact of reason'. Other essays analyse moments in the repercussion of Kant's moral theory in Nietzsche, Heidegger and Derrida.)

Nancy, J.-L. (1986a) *L'Oubli de la philosophie (The Forgetting of Philosophy)*, Paris: Éditions Galilée.(A sustained critique of contemporary philosophical trends that call for a 'return to the subject'. These calls are shown to arise from a desire to evade an exposure to 'sense' by securing an order of 'signification'.)

Nancy, J.-L. (1986b) *La Communauté désoeuvrée*, Paris: Christian Bourgois; trans. P. Connor, L. Garbus, M. Holland and S. Sawhney, *The Inoperative Community*, ed. C. Fynsk. Minneapolis, MN: University of Minnesota Press, 1991.(An analysis of what happens to the thought and practice of community once it is no longer understood in terms of a common identity or common substance. A community thus understood does not

belong to the order of representation and it cannot be a community for which anyone works. It consists of those who are co-exposed to their own lack of identity and substantiality.)

Nancy, J.-L. (1988a) 'L'Offrande sublime', in *Du Sublime*, ed. J.-L. Nancy. Paris: Belin; trans. J. Librett, 'The Sublime Offering', in *Of the Sublime: Presentation in Question*, Albany, NY: State University of New York Press, 1993.(An innovative reading of Kant's Analytic of the Sublime in the *Critique of Judgment*.)

Nancy, J.-L. (1988b) *L'Éxpérience de la liberté*, Paris: Éditions Galilée; trans. B. McDonald, intro. P. Fenves, *The Experience of Freedom*, Stanford, CA: Stanford University Press, 1993.(One of the landmarks of contemporary continental philosophy. Drawing on Heidegger's shift from a discourse of freedom in his early work to a thought of an open - or free - spatiality in his later thought, the book explores in an interconnected series of essays the modern project of grounding philosophical thought and political practice on the idea of freedom.)

Nancy, J.-L. (1990) *Une Pensée finie (A Finite Thought)*, Paris: Éditions Galilée.(A collection of essays on philosophy and literature, many of which are translated in *The Birth to Presence*, 1993.)

Nancy, J.-L. (1991a) *Le Poids d'une pensée (The Weight of a Thought)*, Québec-Grenoble: Presses Universitaires de Grenoble; 'Les Iris', trans. M. Syrotinski, *Yale French Studies* 81, 1992.(A collection of essays, most of which are concerned with contemporary artists and artistic practices.)

Nancy, J.-L. (1991b) 'Of Being in Common', trans. J. Creech in Miami Theory Collective (eds) *Community at Loose Ends*, Minneapolis, MN: University of Minnesota Press. (A succinct exposition of the thought and practice of community once it is freed from the categories of identity, substantiality and work.)

Nancy, J.-L. and Baily, J.-C. (1991) *La Compuration (politique à venir)*, Paris: Christian Bourgois; 'La Compuration/The Compearance: From the Existence of ''Communism" to the Community of "Existence"', trans. T. Strong, *Political Theory* 20 (3), 1992.(An analysis of what happens to 'communism' now that the project of maintaining a 'real, existing socialism' has collapsed. The book is an exploration of the political dimensions of 'being-in-common'.)

Nancy, J.-L. and Lacoue-Labarthe, P. (1991) *Le Mythe nazi*, La Tour d'Aigue: L'Aube; trans. B. Holmes, 'The Nazi Myth', *Critical Inquiry* 16 (2), 1990.(An analysis of Nazi ideology from the perspective of modern German philosophy and intellectual history.)

Nancy, J.-L. (1992) *Corpus*, Paris: Métailié.(An extensive analysis of embodiment, corporeality, incarnation, and the dislocated body as the site of a 'sense' that forever evades all philosophical and scientific attempts to make it into a place of 'signification'.)

Nancy, J.-L. (1993a) *Le Sens du monde*, Paris: Éditions Galilée; trans. J. Librett, *The Sense of the World*, Minneapolis, MN: University of Minnesota Press, 1996.(A series of short chapters in which the thought of a single 'fact' is pursued in a variety of ways: there is no sense of the world as long as 'sense' is understood as 'meaning', but the world itself is 'sense' if this word is understood as the limit of 'signification'.)

Nancy, J.-L. (1993b) *Birth to Presence*, trans. B. Holmes, T. Harrison, C. Laennec, M. Syrotinski, M. Caws, P. Caws *et al*., Stanford, CA: Stanford University Press.(A wide range of essays, including an exemplary analysis of 'decision' in Heidegger's *Being and Time*, the place of the monarch in Hegel's *Philosophy of Right*, and an explication of what Nancy calls 'exscription': the inscription of an ineluctable exteriority in literary texts.)

Nancy, J.-L. (1994) *Les Muses (The Muses)*, Paris: Éditions Galilée; trans. P. Kamuf, *The Muses*, Stanford, CA: Stanford University Press, 1966.(A collection of essays on art and art theory that begins with the question: 'Where are there several arts and not simply one?')

References and further reading

Blanchot, M. (1988) *The Unavowable Community*, trans. P. Joris, Barrytown, NY: Station Hill Press.(A response to Nancy's earliest essays on the 'inoperative community' by one of France's most influential writers and critics.)

Kamuf, P. (1993) *Paragraph: On the Works of Jean-Luc Nancy*, Edinburgh: Edinburgh University Press.(A volume of essays devoted to analysing the work of Nancy, with difficult but rewarding contributions from Derrida, Hamacher, Garciá-Düttmann, and others; the only extensive bibliography of Nancy's writings.)

Miami Theory Collective (eds) (1991) *Community at Loose Ends*, Minneapolis, MN: University of Minnesota Press.(Several essays discuss Nancy's 'Of Being in Common' and consider the consequences of his thought for contemporary political and social philosophy.)

Sparks, S., Sheppard, S and Thomas, C. (1997) *The Philosophy of Jean-Luc Nancy*, London: Routledge.(A collection of essays that analyses Nancy's work, assesses and develops certain lines of his thought.)

Narrative

Narrative, in its broadest sense, is the means by which a story is told, whether fictional or not, and regardless of medium. Novels, plays, films, historical texts, diaries and newspaper articles focus, in their different ways, on particular events and their temporal and causal relations; they are all narratives in the above sense. Accounts of mathematical, physical, economic or legal principles are not. A narrower sense of narrative requires the presence of a narrator mediating between audience and action, and contrasts with imitative discourse wherein the action is presented directly, as in drama. The boundary between narration narrowly construed and imitation is disputed, some writers arguing that apparently imitative forms are covertly narrated.

Attempts have been made to characterize fictional narratives in linguistic terms; another view is that fictions are things which have certain (intended or actual) effects on the audience. Theorists of narrative have mostly concentrated on narratives of the fictional kind and have developed a complex taxonomy of the various narrative devices, some of which are discussed in more detail below. Recently, pressure has been placed on the distinction between historical and fictional narrative by those who believe that history is nothing distinct from the various and conflicting narrative versions we have. It has also been argued that value accrues to an agent's life and acts when those acts conform to a conception of that life as exemplifying narrative.

1 Narrative and fiction

On one view, narratives are identifiable as fictional by their use of distinctively fictional language. But while devices like free indirect discourse (for example, 'She was damned if she would go') may be peculiar to fiction, there are fictions which do not employ them and which are linguistically indistinguishable from nonfiction. Nor can fictionality be explained in terms of lack of reference or truth, since fictions often concern real people and events, while much nonfiction is false. An alternative view is that fictional narratives function in a certain way. On one version of this theory, the reader is intended to *imagine* the occurrence of the events described in the narrative, rather than to treat those descriptions as assertions judgeable by standards of evidence and truth. This suggestion will be employed in the next section.

2 Narrative theory

Work on the theory of narrative has largely concentrated on fictional narratives; the results so far constitute a suggestive taxonomy rather than a systematic theory. An important distinction should be made between the story presented and the discourse that presents it. For example, *anachrony* occurs when the temporal order of events as they happen in the story is not reflected in the discourse; we hear first of A and then of B, though B occurred before A. Again, it may be a feature of the discourse that the story is told by a certain narrator, or from the point of view of a certain character; these are not features of the story itself unless it is part of the story (or, as I shall say, it is fictional) that someone with knowledge of these events performs an act of telling.

Narration and point of view are themselves distinct concepts; the narrator may tell the story from the point of view of a character, with shifts in point of view without change of narrator. The narrative is said to be 'focalized' through a character when the narrator presents the story as it is experienced by that character. It is sometimes said that the story is focalized through the narrator when there is no character through whom it is focalized. This is confusing, because the narrator's act of telling is not in general an act of experiencing the events of the story. Other unhelpful extensions of focalization have been made; it has been said that when character A hears character B speak of character C, C is focalized at the third degree. But C's relation to B and to A in this case is not at all like the relation of a character to the narrator who narrates from that character's perspective (Chatman 1986). Focalization needs a treatment of greater care and parsimony than it has yet received.

The narrator is often said to be distinct from the author. (Until now, I have been using 'narrator' in a sense which does not distinguish them.) The narrator is a construct of the work itself, and either within the world of the story tells us what they know or believe to be true (or, as in an unreliable narrative, tries to mislead us about what has happened), or, alternatively, from outside that world, tells us a fictional story. Watson is the internal or 'intradiegetic' narrator of the Sherlock Holmes stories, while the narrator of *Tom Jones* is external or 'extradiegetic'. The author is also distinguished from the implied author, who is an imagined creature constructed

by the reader as the personality who seems to have written this work; the implied author of *Anna Karenina* might be more tolerant, generous and understanding than the real Tolstoy apparently was.

Could the implied author take on the role otherwise assigned to the extradiegetic narrator? Advocates of the extradiegetic narrator point to works in which the narrator's outlook is one the reader seems to be expected to reject; it is natural to say that there is in such cases a tension between narrator and implied author and consequently that we must conceptualize the work by invoking two distinct personalities. But cases of this kind can be handled without appeal to the extradiegetic narrator, by assuming that the implied author is speaking ironically; it is the implied author who speaks, but what is said is intended by that same speaker to be ridiculed or rejected. We can, therefore, dispense with the idea of an extradiegetic narrator in favour of the implied author. Intradiagetic narrators, on the other hand, are certainly present in some fictions; Watson, in the Holmes stories, is one. In other cases no narrator within the story is evident. Theorists have sometimes postulated hidden narrators, intra- or extradiegetic, for such works on the grounds that the events of the story must be told to us by someone (Chatman 1990: 133). Since the argument just given abolishes the extradiegetic narrator, let us concentrate on the claim that every fictional narrative requires an intradiegetic narrator, even where no such narrator is apparent in the text. In assessing this postulate, it is important to distinguish between what is true and what is fictional. It is true of all stories that they are told by someone, namely the author. The author's act of telling is an invitation to us to imagine various things, and these things jointly constitute what is fictional in the story. Among the things we may be asked to imagine is that someone with knowledge of the events of the story is telling them to us as fact; in that case we have a fictional story with an intradiegetic narrator. On the other hand, we may be invited to imagine merely that the events of the story occurred and that no one in particular is telling them to us. Once we distinguish really telling a fictional story (something the author does) from fictionally telling a true story (something an intradiegetic narrator does) we see that while every fictional story is told as fiction, not every such story is fictionally told as fact. If the reasoning of this paragraph is correct, no works have extradiegetic narrators, only some have intradiegetic ones, and consequently some narratives have no narrator.

In dispensing with a narrator we need not reject the idea that a narrative needs to be understood as the product of an intending agent. But some writers in the structuralist tradition have claimed to identify a kind of writing that lacks the devices which signal the presence of a writer or speaker, such as indexicals or the past tense. They say that narrative in this sense is objective (or purports to be), and the events of the story 'speak themselves' (Benvenista 1966). This is a mistake. Even writing that contains no verbal indicators of agency presupposes the existence of one who speaks or writes, for something counts as language - and not as meaningless marks or sounds, - only if it was produced by an agent intending to conform their sound/mark production to the conventions of a language.

3 Narrative and antirealism

Antirealists (see Realism and antirealism §2) reject the idea that there is a determinate reality, largely independent of human activity and conception, to which our cognitive powers give us only limited access. They claim that the world is indeterminate in those regions inaccessible to our investigations or - in a more extreme version - multiple and conflicting in a way that mirrors our own multiple and conflicting conceptions of the world. Some antirealists have supposed that science, biography and history are species of fiction, giving rise to narratives unconstrained by correspondence with reality.

One version of this argument holds that historical narrative imposes on a series of events features that the series does not intrinsically possess, by means of selectivity, teleology and closure. Narrative selects among events, highlighting some and placing others in the background; but the events themselves do not stand in relations of salience and no event is intrinsically more important than any other. Narrative is teleological in so far as the reason for including events in a narrative is their contribution to outcomes distant from those events and identifiable, once again, only in terms of interests that are variable across cultures, times and individuals. Narrative imposes closure on events by claiming to identify a natural stopping place for the narrative; in fact there are no natural stopping places intrinsic to the events themselves, but only relative to the events as they are ordered and selected by interest. Let us call such features 'plot-like', so as to remind us of the antirealist's presumption that they are features which push historical writing into the realm of the fictional.

One realist response would be to argue that plot-like features, while not intrinsic to the historical events

themselves, are none the less real, relational features of those events, much as colours are said to be real, relational properties of things, since the colour of a thing depends on the kind of visual experience it causes us to have. And just as an account of the colours of things can be thought of as reporting matters of fact, so the attribution of plot-like features to events and their relations in an historical narrative counts as (potentially) fact-stating. On this view, narratives which seem to disagree in their attributions of plot-like features may genuinely contradict one another, and only one narrative can be right. But another possibility for the realist is that the appearance of disagreement is the result of the failure of the plot-like features to be properly relativized. Such a conflict would dissipate if what is said to be significant according to one narrative and insignificant according to another is intended to be understood as significant from one point of view according to the first narrative, and insignificant from another point of view according to the second.

A more moderate realist response admits that plot-like features are not genuinely features of the historical events themselves. It insists, however, that there is another set of features (hereafter 'truth-making' features), which are genuine features of the events themselves, and that historical narratives can be thought of as correct if and only if they attribute these features correctly. Correctness of a narrative in this sense is independent of its plot-like aspects, in that narratives will only count as rivals when they differ over attributions of truth-making features. When they differ over the attribution of plot-like features, they merely count as alternatives. Thus two correct but different narratives would be like two photographs of the same scene taken from different angles. Each photograph/narrative is correct as far as it goes, though one might be more comprehensively correct than the other. Perspective-related differences between the pictures that affect the apparent spatial relations of objects will not count as disagreements between them. Likewise, differences between the narratives in respect of plot-like features will not count as genuine contradictions.

4 Narrative and value

It has been argued that our lives constitute narratives, and that this gives them value. It is difficult to decide what literal truth there is in this idea. How could a life *be* a narrative, rather than simply contain the material for a narrative? If narrative is what someone, perhaps the agent, tells of that life, in what sense does the value of the narrative constitute the value of the life? We know from literature that there can be valuable narratives of bad or otherwise valueless lives. Perhaps a life can be a narrative in this sense: the agent's actions are determined by the agent's conception of past and present actions and future goals as part of a coherent narrative; the agent's life would then be lived 'as narrative'. We can only guess how many lives are in this sense narrative, and in what proportions value is distributed between lives lived as narrative and lives not: of the former many would be reduced in value to the extent that the subject's narrativization of their actions and goals was the result of misunderstanding or self-delusion. Another suggested association between narrative and value is that our present and future actions can affect the value of past actions, for instance in cases where a bad act is to some extent redeemed by its role in the actor's subsequent change of heart. To make a thesis about narrative of this it is not, however, sufficient to say that later acts can affect the value of earlier ones, for that could be true even when the change of heart is related to any story-telling on the subject's part or on anyone else's part. It must be claimed in addition either that (1) value is transmitted from later acts to earlier ones *only* when the later act is the outcome of the agent's conceiving the past act within a narrative framework, or, more modestly, that (2) while value can accrue to the past act without any such narrativizing taking place, additional value accrues when the later act *is* the outcome of such narrativizing.

See also: Barthes, R.; History, philosophy of; Nietzsche, F.; Structuralism in literary theory

GREGORY CURRIE

References and further reading

Barthes, R. (1966) 'Introduction à l'analyse structurale des récits' *Communications* 8; trans. S. Heath, 'Introduction to the Structural Analysis of Narrative', in *Image, Music, Text*, New York: Wang & Hill, 1977; London, Fontana, 1977.(An influential version of the structuralist approach to narrative.)

Benvenista, E. (1966, 1974) *Problèmes de linguistique générale*, Paris: Gallimard, 2 vols; vol. 1 trans. M.E. Meek, *Problems in General Linguistics*, Coral Gables, FL: University of Miami Press, 1971.(A source for the idea of narrative without utterance.)

Booth, W.C. (1983) *The Rhetoric of Fiction*, Harmondsworth: Penguin, 2nd edn.(A classic study of the forms of

narrative, with emphasis on the idea of the implied author. Excellent bibliography.)

Carroll, N. (1990) ‘Interpretation, History and Narrative’, *The Monist* 73: 134-66.(A careful exposition and criticism of some antirealist views about historical narrative.)

Chatman, S. (1986) ‘Characters and Narrators’, *Poetics Today* 7: 189-204.(Criticizes some extensions of the notion of focalization.)

Currie, G. (1993) ‘Interpretation and Objectivity’, *Mind* 102: 413-28.(Argues that there is a degree of relativism inherent in the attempt to interpret narratives.)

(1994) ‘Unreliability Refigured: Narrative in Literature and Film’, *Journal of Aesthetics and Art Criticism* 53: 19-29.(Critical of standard assumptions of narrative theory concerning the role of the implied author and of the narrator.)

Genette, G. (1972) ‘Discours du récit’, in *Figures III*, Paris: Editions du Seuil; trans. J.E. Lewin, *Narrative Discourse*, Ithaca, NY: Cornell University Press, 1980.(A classic development of narrative theory through close readings of Proust.)

MacIntyre, A. (1985) *After Virtue*, London: Duckworth, 2nd edn.(Argues for ‘a concept of a self whose unity resides in the unity of a narrative’.)

Nehamas, A. (1985) *Nietzsche: Life as Literature*, Cambridge, MA: Harvard University Press.(An interpretation and defence of Nietzsche assimilation of ‘the ideal person to the ideal literary character and the ideal life to the ideal story’.)

Savile, A. (1996) ‘Instrumentalism and the Interpretation of Narrative’, *Mind* 105: 553-76.(Argues for the objectivity of narrative interpretation.)

White, H. (1987) *The Content of Form*, Baltimore, MD, and London: Johns Hopkins University Press.(The work of an influential historical antirealist.)

Nation and nationalism

No one observing political events in the world today could deny the continuing potency of nationalism. Many of the most intractable conflicts arise when one national community tries to break away from another, or when two such communities lay claim to the same piece of territory. Yet to outsiders the basis for such conflicts often seems mysterious. People are prepared to fight and die for their nation, yet what exactly is this 'nation' that commands such loyalty? Why should it matter so much that a person is governed by leaders drawn from one community rather than from another?

Philosophers are often inclined to dismiss nationalism as having no rational basis, but as resting merely on tribal instincts and brute emotions. Such a response overlooks the different forms that nationalism has taken: in particular, the contrast between authoritarian nationalism, which allows national cultures to be imposed by force and which may justify acts of aggression against neighbouring peoples, and liberal nationalism, which upholds the rights of individuals to form political communities with those with whom they feel identified and to protect their common culture. We need to examine carefully the arguments that have been advanced by nationalist thinkers in order to decide which form of nationalism, if any, is rationally defensible.

1 What is a nation?

Defining 'nation' has proved a difficult task for political philosophers. In common speech 'nation' is often used loosely to mean 'state', as it does in phrases such as 'United Nations'. But this loose usage is not helpful in political philosophy, one of the central questions of which is whether there are indeed such things as nations, and if there are, whether nations have a valid claim to be self-governing - or to put it more crudely, whether each nation is entitled to have its own independent state. When questions such as these are asked, 'nation' must refer to something that exists apart from and prior to the political institutions of the state.

A nation, we may say, is a community of people who recognize that they are distinct from other communities and wish to control their own affairs. But a community of this sort is clearly very different from the small face-to-face communities that may first spring to mind when the idea of community is invoked (see Community and communitarianism §1). Since I shall never know more than a tiny fraction of my fellow countrymen personally, how can I say that we form a community? What is the source of unity that holds nations together?

Broadly speaking, two answers have been given to this question. On one side stand those who take a subjective view, claiming that nations are essentially voluntary associations of people held together simply by the continuing will of their members. We form a nation because we want to be politically associated. The reasons behind this desire are irrelevant: all that finally matters is that each of us wants to associate with this group of people rather than that. On the other side we find those who maintain that nations are marked out by certain objective characteristics that their members share - by racial descent, by language, by religion, by common traits of character, and so forth. On this view it makes sense to speak of dormant nations whose members have in common whatever is taken to be the essential defining characteristic of nationality, but who as yet display no consciousness of nationhood or desire to form a political community.

Neither of these answers seems adequate as it stands. Objective accounts of nationhood fall down when we take the proposed characteristics one by one and see that none is adequate to distinguish all those communities that we recognize to be separate nations (Renan 1939). If we take the objective view and focus on a feature such as language, we find on the one hand that there can be distinct nations with a shared language - the English-speaking countries for instance - while on the other hand there can be bi- or tri-lingual nations such as Belgium or Switzerland. The subjective account avoids this pitfall, but this does not clear up the mystery of why people should care so much who they associate with politically. A more adequate answer must combine elements of both viewpoints. A nation exists when its members recognize one another as belonging to the same community and as bearing special obligations to one another, but this is by virtue of characteristics that they believe they share: typically, a common history, attachment to a geographical place and a public culture that differentiates them from their neighbours.

Because nations are communities on a large scale, they are inevitably 'imagined communities', in Benedict

Anderson's evocative phrase, meaning that they depend for their existence on collective acts of imagining which take place through media of communication such as books, newspapers and television (Anderson 1990). It is from these media that we acquire our understanding of what it means to belong to this particular nation. Does this imply that national communities are after all spurious? This by no means follows, although it may justify a degree of scepticism about the received content of national identities. First, such identities usually incorporate a certain proportion of historical myth, in the sense that past events are interpreted in such a way as to highlight their continuity with events in the present. Although we actually know very little about our thirteenth-century ancestors, for example, we are invited to see their heroic struggles against their neighbours as the precursors of our own. Second, nations tend to exaggerate their own homogeneity, supposing that every member conforms more or less precisely to the cultural stereotype that forms part of the national identity. Third, national identities are to some degree able to be manipulated by powerful groups, especially political leaders, whose command of the media enables them to refashion the content of the imagining to suit their purposes. No political leader can invent a nation where one does not exist; but it may be possible to persuade a nation that its legitimate territorial boundaries stretch more widely than it had hitherto imagined, or that a particular subcommunity - an ethnic group for instance - that has long lived happily as part of the nation does not after all belong there. The practical consequences of such refashioning may be very serious, as experiences of the twentieth century readily illustrate.

For reasons such as these, most philosophers have been ambivalent both about the validity and about the practical relevance of national identities.

2 The evolution of nationalism

Nationalism is primarily a phenomenon of nineteenth- and twentieth-century political thought. Its constituent ideas, however, stretch back much further in time. Among the Jews of the Old Testament and among the ancient Greeks we find the sense of belonging to a distinct people marked off from the rest of the world by special characteristics - an embryonic sense of national superiority (Kohn 1944, ch. 1). In the classical period, moreover, patriotism was regarded as a leading virtue. Patriotism is not the same as nationalism: it means a love of one's native country and a willingness to sacrifice oneself to defend it, but it need not invoke the idea of a self-conscious nation discussed in §1, nor need it imply that the country in question should enjoy political self-determination. Like nationalism, however, patriotism involves the idea that one owes special obligations to one's compatriots, that one should be willing to protect and defend them at the expense of outsiders - an ethical stance that many philosophers have found difficult to accept.

Nationalism proper began to emerge in the second half of the eighteenth century, when to the ideas of distinct nationhood and patriotic loyalties were added the idea that nations are the source of legitimate political authority. There is a close link here between nationalism and the doctrine of popular sovereignty, which replaced the older belief that political authority flowed downwards from the king or the emperor (Kamenka 1976). Rousseau (§3) provides a striking illustration of this link, holding on the one hand that the general will of the people was the only legitimate source of legislative authority, and on the other that every citizen must be taught to love their native country before all else: 'That love makes up his entire existence: he has eyes only for the fatherland, lives only for his fatherland;… the moment he has no fatherland, he is no more' (Rousseau 1770-1: 19).

In these words, Rousseau anticipates the more extreme forms of nineteenth-century nationalism. The contribution of the early nineteenth century was the Romantic belief that each nation formed an organic unity, with its own soul and its own special destiny. Of particular importance was Herder (§§2, 6), whose expressive theory of language underpinned his idea that each nation was the bearer of a unique culture, immersion in which was the only path to human self-fulfilment. Herder's nationalism was primarily cultural rather than political; he wanted to see each national culture thriving in peaceful coexistence with its neighbours, and he had no sympathy with the aggressive militarism of the Prussian state. None the less he continued to link nationality and popular self-government: 'The roses for the wreath of each nation's liberty must be picked with its own hands, and must grow happily out of its own wants, joys, and love' (Kohn 1944: 431). This message found a ready response among those cultural minorities of Europe struggling for political independence from long-established empires.

Cultural nationalism and political self-assertion coalesced in the later thought of Fichte (§7). Fichte combined three doctrines: the idea that nations were organic wholes, each with its own language and culture; that each person owed supreme loyalty to their own nation - that indeed freedom consisted of identifying oneself with the

higher cause represented by the nation; and that each nation had its own peculiar mission. Fichte also attenuated the link between nationalism and popular self-government: national leaders might be justified in using coercive methods of education in order to create strong and united peoples. Here we see the foreshadowing of an authoritarian form of nationalism which is willing to override liberties domestically in the name of national unity, and which may justify acts of aggression against neighbouring states in the name of national destiny.

The same period also witnessed the growth of a liberal form of nationalism, according to which the cause of national independence and the cause of political liberty go hand in hand. This was the view of Mazzini, the father of Italian nationalism, and it was echoed by J.S. Mill, who in *Considerations on Representative Government* (1861) emphasized the role played by national sympathies in checking the power of government and making free institutions workable. Liberal nationalism treats all nations as having equally good claims to self-determination, resists ideas of ’national destiny’ and argues that national cultures will flourish best when freedom of expression and other liberties are secure.

The subsequent history of nationalism can be seen as a struggle between authoritarian and liberal forms of nationalism. The latter can be seen at work in the doctrine of national self-determination promulgated by Woodrow Wilson at the end of the Second World War and embodied in the League of Nations; the former found its final expression in the fascist movements of the inter-war years (see Fascism). Thus nationalism continues to present us with two very different faces. On the one hand it stands for a people’s right to protect and develop their inherited cultures and to be politically independent in association with those they regard as their compatriots. On the other hand it stands for forcible indoctrination in the national culture and the promotion of national interests abroad at the expense of other peoples. Once again we see why philosophers have very often reacted ambivalently to nationalist ideas.

3 Problems of nationality

Nationalist ideas raise a number of important questions in ethics and political philosophy. This section addresses three. Do we owe special obligations to our compatriots? How can the idea of national self-determination be justified? How should we resolve disputes about borders, and in particular secessionist demands by national minorities?

The first question arises because the claim, common to all forms of nationalism, that we owe special loyalties to our compatriots, seems to collide directly with the idea that morality requires us to show an equal regard for every human being, without making distinctions. Most moral philosophers, especially utilitarians and Kantians, embrace the idea of ‘equal regard’, which may be labelled ethical universalism. We should weigh the interests of all human beings equally, or show each person equal respect regardless of their particular relationship to us. This points to a form of cosmopolitanism (see International relations, philosophy of §4). But many of these philosophers, recognizing how demanding and how far removed from ordinary moral ideas strict universalism is, have also tried to make room for particular loyalties, including national loyalties. They argue, for instance, that duties that are universal in scope may be discharged most effectively if each us takes responsibility for the welfare of a small number of others. Or they argue that members of particular communities may be regarded as contractually obliged to give one another special assistance, the justification for such contracts (or quasi-contracts) again being universal in form.

The weakness of these arguments is that they have difficulty showing why nations should be singled out as the groups to which special obligations are owed. If I am a utilitarian, for instance, why should I acknowledge a special responsibility to contribute to the welfare of the poorer members of my own national community when there are people much worse off in other countries? Even if my sympathies are presently biased towards those closest to me, do I not have a duty to correct this bias for the sake of maximizing overall utility? It seems that if we are to recognize national obligations, we have to abandon universalism as a complete picture of ethics and allow particular obligations to enter at ground level, so to speak. We need a picture which allows us to connect questions about personal identity - who we are, where we belong - with questions about our responsibilities to others (MacIntyre 1984).

Turning next to national self-determination, critics of nationalism such as Kedourie (1966) have argued that nationalist appeals to a national or popular will are fundamentally misguided, and that our concern should be with

the quality of government - whether it observes the rule of law and protects our rights, for instance - rather than with its source. If we are governed well by outsiders, we have no justified grounds for complaint. This challenge forces us to look more closely at the reasons for favouring self-determination, of which there are several.

One such reason has to do with our interest in protecting and fostering the common culture that provides us with the resources for a meaningful life. This is especially the case when the culture in question is the culture of what Margalit and Raz (1990) have called an ‘encompassing group’. The best guarantee that such cultures will not be destroyed or forcibly altered is to be governed by representatives who also participate in the national culture. Given the pressures currently exerted on national cultures by worldwide commercial forces, this argument has obvious force.

A second reason concerns the part played by nationality in fostering trust between the members of a political community. Where a shared national identity creates mutual trust, it becomes easier to agree on particular disputed issues - each side having more incentive to reach a compromise - and there is also a better chance of winning support for policies that aid one section of the community at the expense of others. Of course this argument can be turned on its head by those who favour a state with minimal responsibilities: if your aim is to disable the state from doing very much, you should prefer it to be multinational in character, so that each community will try to block policies that favour the others (this argument was advanced by Lord Acton (1907)). But those for whom social justice is an important value, and those who place a premium on achieving political agreement by democratic means, will favour political communities that are held together by national solidarity.

Finally we come to the problems posed for nationalist doctrines by disputed boundaries and demands for secession. If the peoples of the world were neatly divided into separate nations living on well-defined territories, the principle of national self-determination would be easy to apply. But in reality populations are very often interspersed in such a way that any territorial division will leave many people living under national governments with which they do not identify; and national identities, too, may be somewhat ambivalent. What are we to say about groups such as the Basques, the Québecois and the Scots, who appear to identify in part with larger nations but also have distinct identities of their own that they want to express politically?

Problems such as these have led many liberals to conclude that appeals to nationality cannot settle the question of where the borders of states should be drawn. We must be guided by other values: we should decide borders on the basis of individual consent, allowing any subcommunity to secede from the state by majority vote as long as it extends the same right to its own minority groups (Beran 1984); or we should permit secession when the seceding group can show that it is the victim of serious injustice or that it is threatened with cultural extinction (Buchanan 1981).

These responses may underestimate the versatility of nationalist principles, however. The solution prescribed by these principles will vary depending on the facts of each particular case. Where a state currently houses two or more distinct and well-defined national communities, the case for a peaceful divorce will be very strong. Where a small minority lacks a territorial base and is dispersed throughout a larger national community, the best solution may be to open up the existing national identity to that minority by removing or playing down divisive elements - for instance, by diminishing the role played by religion - so that in the longer term it will be able to integrate without annihilating its own culture. Where we find subcommunities nesting inside larger communities, as in the examples given above, national self-determination suggests devolving power to these subcommunities so that they are able to protect and express their own culture, without breaking up the state as such, since this continues to give expression to their wider identities and loyalties.

There will none the less be tragic cases in which no redrawing of borders or restructuring of political authority is able to give divided communities adequate protection and self-determination. Conflicts such as that created by the partition of Bosnia cannot be resolved by appeals to these principles. Such cases remind us that nationalism cannot by itself be a complete political philosophy. At the very least we need to add to it principles such as human rights, which set limits to what may be done to communities and individuals in the name of the nation.

See also: Justice, international; State, the; Zionism

DAVID MILLER

References and further reading

Acton, Lord (1907) 'Nationality', in J.N. Figgis (ed.) *The History of Freedom and Other Essays*, London: Macmillan.(A defence of multinational states from a liberal perspective.)
Anderson, B. (1990) *Imagined Communities*, London: Verso, revised edn.(Wide-ranging study of national identities, emphasizing the part played by mass media in their creation and dissemination.)
Beitz, C. (1983) 'Cosmopolitan Ideals and National Sentiment', *Journal of Philosophy*, 80 (10): 591-600.(Careful exploration of the question whether universalist reasons can be given for acknowledging special obligations to compatriots.)
Beran, H. (1984) 'A Liberal Theory of Secession', *Political Studies* 32 (1): 21-31.(Appeals to the principle of consent to defend a universal right of secession.)
Berlin, I. (1981) 'Nationalism: Past Neglect and Present Power', in H. Hardy (ed.) *Against the Current*, Oxford: Oxford University Press.(An interpretation of nineteenth-century nationalism, and an attempt to explain the continuing appeal of nationalist ideas.)
Buchanan, A. (1981) *Secession*, Boulder, CO: Westview Press.(Comprehensive critical assessment of liberal arguments for and against secession.)
Fichte, J.G. (1922) *Addresses to the German Nation*, Chicago, IL: Open Court.(The main statement of Fichte's German nationalism and appeal for a national scheme of education.)
Gellner, E. (1983) *Nations and Nationalism*, Oxford: Blackwell.(An interpretation of the rise of nationalism, hinging upon the need for a common high culture in modern industrial societies.)
Greenfeld, L. (1992) *Nationalism: Five Roads to Modernity*, Cambridge, MA: Harvard University Press.(A comparative study of the evolution of nationalist ideas in Britain, France, Germany, Russia and the USA.)
Herder, J.G. (1969) *Herder on Social and Political Culture*, ed. F.M. Barnard, Cambridge: Cambridge University Press.(An anthology of texts covering Herder's explanation of the cultural diversity of humankind and his exploration of the links between national cultures and political communities.)
Kamenka, E. (1976) 'Political Nationalism - The Evolution of the Idea', in *Nationalism: The Nature and Evolution of an Idea*, London: Edward Arnold.(Explores the origins of nationalism, stressing its connection with the idea of popular sovereignty.)
Kedourie, E. (1966) *Nationalism*, London: Hutchinson.(Critical study of nationalism by a classical liberal.)
Kohn, H. (1944) *The Idea of Nationalism*, New York: Macmillan.(A classic study of the development of nationalist ideas, beginning with ancient Israel.)
MacIntyre, A. (1984) 'Is Patriotism a Virtue?', in R. Beiner (ed.) *Theorising Citizenship*, New York: State University of New York Press, 1995.(Critique of contemporary ethical theories in terms of their incapacity to understand patriotism as a virtue.)
Margalit, A. and Raz, J. (1990) 'National Self-determination', *Journal of Philosophy* 87 (9): 439-61.(Argues for a right to self-determination to protect the interests that people have as members of 'encompassing groups'.)
Mill, J.S. (1861) *Considerations on Representative Government*, London: Dent, 1971.(Chapter 16 contains Mill's argument that for representative government to work successfully, state boundaries should coincide, if possible, with national boundaries.)
Miller, D. (1995) *On Nationality*, Oxford: Clarendon Press.(Sympathetic discussion of nationalist ideas, exploring in greater depth the issues addressed in this entry.)
Nathanson, S. (1993) *Patriotism, Morality and Peace*, Lanham, MD: Rowman & Littlefield.(An attempt to separate defensible 'moderate' patriotism from indefensible 'extreme' patriotism.)
Renan, E. (1939) 'What is a Nation?', in A. Zimmern (ed.) *Modern Political Doctrines*, London: Oxford University Press.(Much-cited essay, critical of objective definitions of nationality.)
Rousseau, J.-J. (1770-1) *The Government of Poland*, trans. W. Kendall, Indianapolis, IN: Bobbs-Merrill, 1972.(Rousseau's most explicit account of how and why citizens must be imbued with national loyalty.)
Scruton, R. (1990) 'In Defence of the Nation', in *The Philosopher on Dover Beach*, Manchester: Carcanet.(A conservative defence of the nation as an object of loyalty.)
Tamir, Y. (1993) *Liberal Nationalism*, Princeton, NJ: Princeton University Press.(An attempt to show that liberalism and nationalism can be reconciled with one another.)
Walzer, M. (1990) 'Nation and Universe', in G.B. Petersen (ed.) *The Tanner Lectures on Human Values*, Salt Lake City, UT: University of Utah Press, vol. 11.(Separates forms of nationalism which proclaim the superiority of a particular culture from those that extend the same recognition to all.)

Native American philosophy

Native American philosophies are multiple and multiply different, though with some palpable commonalities. Cosmologically, Native American thought posits phased differentiations of form from a primordial void. Form is fundamentally animate, and widespread metaphysical premises indicate a pervasive animism. The relationship between language and reality is grounded in the semiotic manipulation of animate forces. Encompassing force - like Siouan wakan *or Iroquoian* orenda *- is the originating source of human ontology and subjectivity, which are strongly characterized by ideas of consciousness and will. Mind is critically informed by transcendental experience (dreams, visions and so on) as well as by reason. Ethically, the harnessing of individual energies to specified social goals is discursively and dramatically prominent especially in ritual performance. Ethical principles are extended to non-human 'persons' (particularly animals and deific forms) who operate within the human moral compass. Native American reason emphasizes analogy (especially in Lévi-Strauss' interpretations) and aetiology, and there is a central concern with formal composition and decomposition. The latter is strongly evident in the ubiquitous philosopher-figure 'Trickster' - a subversive transgressor of constituted order.*

1 Background

Prior to 1492 Native North America included ten to fifteen million people, speaking 250 distinct languages (none written), belonging to more than twenty largely unrelated language families. There were upwards of 300 separate 'cultures' - that is, forms of life constitutively differentiated according to language, livelihood, social form and belief. Larger groupings into 'culture areas', united by geographic proximity, subsistence practices and customs, minimally include (with examples of named peoples in parentheses): Northeast Woodlands (Mohawk, Seneca, Pequot, Delaware), Southeast (Cherokee, Creek, Choctaw, Chickasaw), Plains (Lakota, Cheyenne, Comanche, Kiowa), Southwest (Hopi, Zuni, Navajo, Pima-Papago), Great Basin (Paiute, Shoshone, Washo, Ute), Plateau (Nez Perce, Klamath, Yakima, Kutenai), California (Yurok, Pomo, Yokuts, Maidu) Northwest Coast (Tlingit, Kwakiutl, Tsimshian, Nootka), Sub-Arctic (Koyukon, Beaver, Cree, Montagnais-Naskapi), and Arctic (numerous Eskimo/Inuit groups and Aleut).

While philosophical forms cannot be read from economic adaptation, certain limitations are imposed by population size and density, and ensuing social contexts. State-level societies, for example, seem prerequisite for higher mathematics and, correlatively, symbolic logic. Some North American number and calendrical systems were quite complex: the Pomo, who participated in a wide-ranging shell-disc currency, had a counting capacity up to 40,000; likewise, prehistoric Anasazi astronomical observations were notably fine-grained. But both mathematical and astronomical differentiation were significantly greater in the Classic Maya state, with its interlinked day- and year-counts, observatories and writing-system.

Each culture-area is characterized by ecosystems allowing particular modes of adaptation. Maize-beans-squash agriculture, the foundation of stratified states in Mesoamerica, occurred in the Northeast, Southeast and small areas of the Plains and Southwest; in these regions status differences distinguished ritual priest-leaders as the intellectual specialists. Elsewhere, dependence on hunting (Plains, Plateau, Sub-Arctic and Arctic), and wild plant collection (Great Basin, California and parts of the Southwest) are associated with smaller populations and more egalitarian societies; here, individual shamans ('medicine-men') are the philosophical experts. Interestingly, the fishing societies of the Northwest Coast produced fairly dense hierarchical communities that marked the metaphysical qualities of high-ranking individuals with titles, crests, masks and other insignia, all subject to contestation in elaborate ceremony (the 'potlatch').

Sociocultural diversity is accompanied by extensive historical differences. Archaeological remains attest to states in the Southeast - the 'Mississippian' - and the Southwest - 'Anasazi' and 'Hohokam'. These civilizations collapsed before 1492, though elements of the former persisted into the seventeenth century (especially with the Natchez), and elements of the latter are still found in Pueblo and Pima-Papago societies. European contact was devastating. Population declined massively (up to ninety per cent) from introduced diseases and genocidal wars. Today many languages are dead and numerous systems of thought are irretrievable. Yet in some areas indigenous culture remains strong; many Hopis and Navajos, for example, still speak their own languages, and some cultural forms are largely unaltered from pre-European times.

The corollary of a traumatic history is the erasure of records. There was little genuine interest in Native thought, peremptorily dismissed as ‘savage’ or ‘primitive’. The meditations of Montaigne, Locke or Rousseau on Native social and natural philosophy were largely oppositive projections of European conceptions onto people who typified the ‘state of nature’ - ornamented with frequently misconceived observations by explorers or missionaries (Lafitau’s *Moeurs des sauvages amériquains* (1724) is a partial early exception). Since the languages were oral, philosophical thought often did not survive unless translated and transcribed by chroniclers literate in a European language. Most early intellectually-oriented inquisitors were missionaries, whose epistemological intents were not neutral. Recorded occasions of genuine philosophical dialogue are sparse (those of turn-of-the-century physician James R. Walker with George Sword and other Lakota shamans are noteworthy). Even with sincere interest, interlocutors have typically had an insufficient grasp of notoriously complex linguistic forms for effective understanding. One of the few academic philosophers to undertake field research likens his dialogue with an elderly monolingual Navajo, translated by younger bilinguals, to ‘trying to talk to Plato with an Attican peasant as interpreter!’ (Ladd 1957: 201). In sum, the difficulties of knowing much genuine Native American philosophy are manifold, owing to its multiplicity and diversity, as well as to problems of translatability, both linguistic and social, compounded by historical circumstances.

2 Cosmology, metaphysics and ontology

For many Native American philosophies, a pervasive animism - denoted by Siouan *wakan*, Iroquoian *orenda*, Algonquian *manitou*, and Kwakiutl *nawalak* - is fundamental. This widespread idea posits, in all things and processes, an inherent vital force which has compulsive effects on events and conditions, may become attached to or the possession of individual ‘persons’ (not only human), and was the originating source of all phenomena. Iroquoian *orenda*, for example, is a:

> hypothetic principle…conceived to be immaterial, occult, impersonal, mysterious in mode of action, limited in function and efficiency, and not at all omnipotent, local and not omnipresent, and ever embodied or immanent in some object, although it was believed that it could be transferred, attracted, acquired, increased, suppressed, or enthralled by the orenda of occult, ritualistic formulas endowed with more potency.
>
> (Hewitt 1906: 147)

Likewise, Siouan *wakan* is ‘the animating force of the universe, the common denominator of its oneness’ (DeMallie 1987: 28); the totality of life-forces are *wakan tanka* (‘the great *wakan*’, popularly translated ‘Great Spirit’). Conceptually, *wakan* is typically characterized by its impenetrability, and in use has radiating synecdochic and metaphoric referents. But its multiplex significations are formally divided by shamans into sixteen hierarchical (four times four) aspects (see Figure 1). Each of these has particular powers of creation and transformation in the founding of the universe and its continual unfolding in human experience and practice.

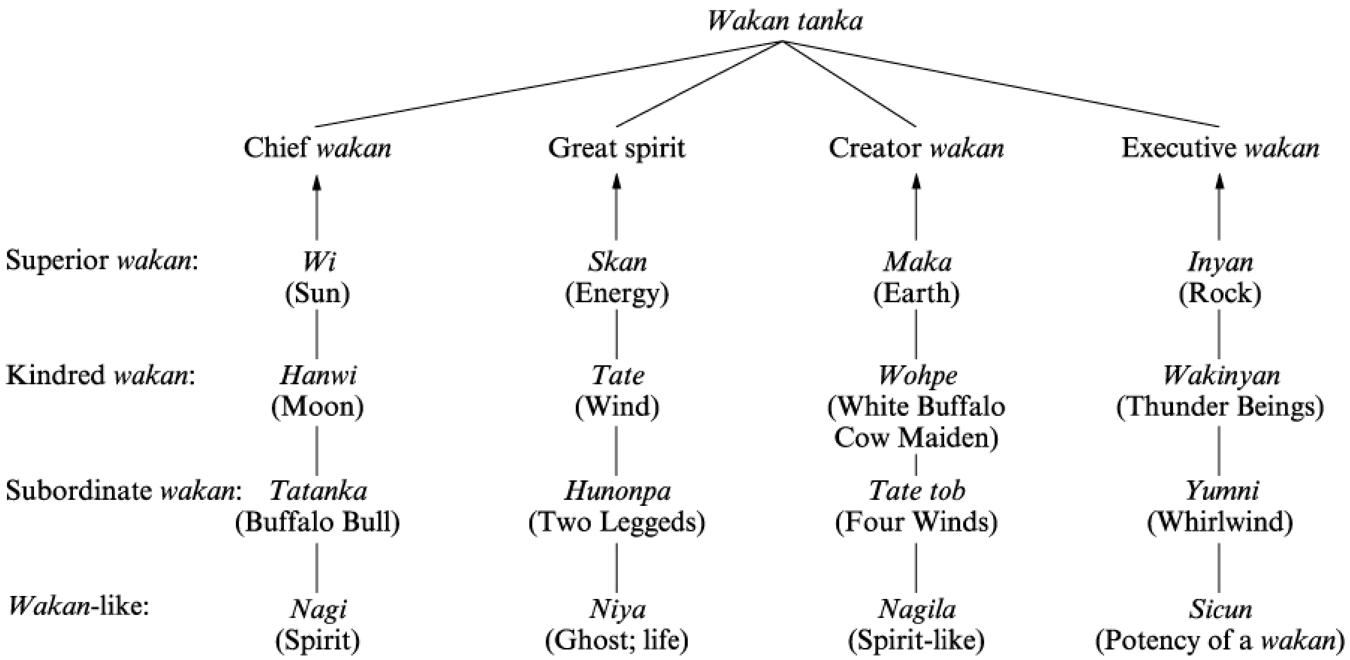

Figure 1. Tobtob kin, the four times four. (DeMallie and Lavenda 1977)

Here the abstraction of universal power converges with the constitution of human subjectivity. The Sioux person is animated at birth by three *wakan* elements given by *skan* ('Energy'): 'consciousness or will' (*sicun*, a guardian against evil); a ghost 'which comes from the stars'; and 'a spirit, evidently an immaterial but immortal reflection of the body'. 'After the death, the guardian escorts the spirit to the spirit world beyond the Milky Way; the guardian and the ghost return to the places from which *skan* originally got them' (DeMallie and Lavenda 1977: 157). These elements of ontogeny and ontology depict a human essence and matter particulate to universal totality. Simultaneously, a sense of phylogeny is central.

Narratives of origin vary, but many identify consecutive phases (for Zuni there are four, for Hopi three, for Tewa two). The natural and cultural world is progressively differentiated out of formlessness through active intervention by mystical forces (for Navajo these include coloured mists, swallows, and insects). At intermediate junctures, transformer-heroes appear in various guises, like Changing Woman and the monster-slayers, her twin sons by the Sun, for the Navajo and Apache. Gradually transformers produce sufficient order for humanoid forms, which evolve from primitivity to modernity (for example, Brightman 1990).

The Tewa (Pueblo) world was formed and ordered after the emergence of proto-humans from a primordial lake (Ortiz 1969). Various culture-heroes (including Mountain Lion the Hunt Chief, Summer Chief, Winter Chief, and six pairs of varicoloured sibling deities of the directions) were sent up by the originating Mothers, 'Blue-Corn Woman near to summer', and 'White-Corn Woman near to ice'. They harden the ground and order time, space, topography, seasonality, biology and social mores. The constructed world is bounded at its corners by four sacred mountains (with 'earth-navels' on the peaks); towards the centre these are successively linked to four flat-topped hills, four shrines at the village-margins, and finally four dance plazas in each Tewa village. The world ordering affects all aspects of Tewa existence and is dramatically reiterated in ritual; the culture heroes are impersonated, and all ritual trajectories - in speech, song, dance-movements and so on - insistently re-mark the tetradic structures, both temporally and spatially. In short, origin-narratives both explain the world and set the terms for human action within it.

Hopi metaphysics and ontology, evidenced in linguistic forms, emphasizes 'events', or 'eventings', that are objectively 'expressed mainly as outlines, colours, movements, and other perceptive reports', in contrast with the static substances, masses and entities foremost in European languages (Whorf 1956: 147). According to Whorf:

> Every language contains terms that have come to crystallize in themselves the basic postulates of an unformulated philosophy…. Such are our words 'reality, substance, matter, cause,' and… 'space, time, past, present, future.' Such a term in Hopi is the word most often translated 'hope' - *tunatya* - 'it is in the action of

> hoping, it hopes, it is hoped for, it thinks or is thought of with hope,'…. The verb *tunatya* contains in its idea of hope something of our words 'thought,' 'desire,' and 'cause,' which sometimes must be used to translate it. The word is really a term which crystallizes the Hopi philosophy of the universe in respect to its grand dualism of objective and subjective: it is the Hopi term for SUBJECTIVE. It refers to the state of the subjective, unmanifest, vital and causal aspect of the Cosmos, and the fermenting activity toward fruition and manifestation with which it seethes - an action of HOPING; i.e. mental-causal activity, which is forever pressing upon and into the manifested realm.
>
> (Whorf 1956: 61-2)

The difficulties of translating Native philosophies are partly descriptive. With some adjustments, Whorf's rendering of Hopi metaphysics calls to mind Heidegger's theory of being, particularly of existence as an emergence-into-presence (see Heidegger, M.).

Despite valuable criticism (for example, Malotki 1982), much of Whorf's analysis holds true. In Whiteley's (1988) research, one Hopi philosopher, Yoywayma, emphasized a triadic principle of idea, intention, and action - *tunatya*, *pasiwni*, *okiw antani*. In combination with *natwani*, an exteriorization of self, this triad suggests that - as for Husserl - intentionality and its worldly engagement is the defining feature of a conscious self (see Intentionality). *Tunatya*, explained with a characteristic horticultural metaphor, refers to a 'seed of thought' - the germ of an idea, we might say - that occurs in one's consciousness. *Pasiwni*, planning, means bringing together and setting into motion all the necessary technical and symbolic resources to germinate the seed and bring it to fructifying life. *Okiw antani*, 'let it be this way,' refers to proper mental attitude and is also a prayer that sanctifies the intention to control deific forces.

Pueblo thought discloses a common metaphysics (the 'Pueblo' Indians include the Hopi, Zuni, Tewa, Tiwa, Towa and Keres, all with entirely different languages but other cultural similarities):

> The general Pueblo conception of causality is that everything - animate and inanimate - counts and everything has its place in the cosmos. All things are thought to have two aspects, essence and matter. Thus everything in the cosmos is believed to be knowable and, being knowable, controllable. Effective control comes only from letter-perfect attention to detail and correct performance, thus the Pueblo emphasis on formulas, ritual, and repetition revealed in ritual drama. Among human beings, the primary causal factors are mental and psychological states; if these are harmonious, the supernaturals will dispense what is asked and expected of them. If they are not, untoward consequences will follow just as quickly, because within this relentlessly interconnected universal whole the part can affect the whole, just as like can come from like. Men, animals, plants and spirits are intertransposable in a seemingly unbroken chain of being.
>
> (Ortiz 1972: 143)

In Navajo metaphysics, causality and consciousness are linked through speech. Speech, especially in ritual, has causal power, simultaneously *doing* what it says: 'Ritual language does not describe how things are, it determines how they will be. Ritual language…disperses evil, reverses disorder, neutralizes pain, over-comes fear, eliminates illness, relieves anxiety, and restores order, health, and well-being'. Ritual speech derives from foundational knowledge of which reality itself is composed. In Navajo creation narratives:

> language is the means by which form is projected onto substance…. The symbol was not created as a means of representing reality; on the contrary, reality was created or transformed as a manifestation of symbolic form. *In the Navajo view of the world, language is not a mirror of reality; reality is a mirror of language*.
>
> (Witherspoon 1977: 34, emphasis in original)

3 Being and mind

The (Iroquoian) Huron believed in a soul both 'divisible and physical' with terms for the divisible parts. One part, *,aata*, is equated with the body, and with the body's life-force, but also with aspects of 'the self', particularly in moral interrelationship with other selves, in emotions and in personality. It also operates conjunctively, in an overall conception that is fundamentally dualistic:

> One major Huron conception of the soul, that of the perceptive or sensitive-animating one, is linked and is

> comprised of the emotive soul, *eiachia* or heart…, the general body-soul, *,aata*, and the general life-soul, *onnhek8i*. The second Huron conception, the rational or intellectual free soul, combined the intellectual soul, *,andi,onra* (mind), and the free soul, oki.
>
> (Pomedli 1991: 60)

The intellectual soul is the means of thought. In European philosophy, thinking is generally depicted as a process in which mind absorbs sensory impressions from external reality and then deliberates internally with the abstracted residue:

> Hurons, however, emphasize a more extrinsic method of thinking. For them, the mind, in the process of thinking, transports itself or is transported into the object or event in question…. Instead of the mind using its power to assimilate, it rather has the power to move into an object, enrich it, name it, make it meaningful, and empower it. The object or event remains out there, nevertheless, but is now changed and made similar to the mind because of the free-moving thinking process…. While for Europeans thinking involves the conscious and deliberative, that is, discursive activities, for Hurons it focuses on the imaginative and intuitive, that is, excursive activities.
>
> (Pomedli 1991: 65-8)

In conscious thought and in dreams there is an internal dialogue - an ‘immanent intersubjectivity’ - among Huron soul forms. A conception of multiple, particularly dual, souls (where one - the free soul or psyche - is detachable from the animate body) is widespread throughout North America.

Li’maqsti, ‘mind’, among the Nuu-cha-nulth (or Nootka) of the Northwest Coast, is conceptualized as a column of ten connected spiritual homunculi resident in the spinal cord. *Li’maqsti* incorporates ‘reason’ and ‘will’ and serves as a sort of transducer between sensation, mentation and action, between ‘internal and external states’. The realization of action from intention involves a colloquy:

> Each of the [ten connected] persons is an individual…and each is capable of autonomous thought. A man acts, in effect, by majority rule. If the ten persons are in complete agreement he knows just what to do; if they are in balanced disagreement, then he is incapable of acting at all.
>
> (Golla 1988: 4)

Concepts of mind emerge also in the widely remarked Native attention to dreams. Iroquois dream theory is ‘basically psychoanalytic’:

> They recognized conscious and unconscious parts of the mind…. [They made] the distinction between the manifest and the latent content of dreams and employed what sounds like the technique of free association to uncover the latent meaning. And they considered that the best method for the relief of psychic and psychosomatic distresses was to give the repressed desire satisfaction, either directly or symbolically.
>
> (Wallace 1969: 63)

Dream theories vary, but there are numerous echoes of Iroquois views (for example, Lincoln 1935), especially with regard to dreams as guides for conscious action. For the sub-Arctic Cree:

> Dream images may be understood as prophetic, as literal or symbolically disguised representations of future states of affairs. Memories of the dream are then retained by the waking self, providing foreknowledge of future occurrences….
>
> [But] Crees say that dream events may *determine* the occurrence of like events - their worldly simulacra - that have not yet happened but that will transpire in waking life only as a result of their having initially been dreamed. The dreamer…[further] conceives him or herself as capable of influencing worldly outcomes by choosing certain courses of action in the anticipatory dream.
>
> (Brightman 1993: 97-9)

In other sub-Arctic hunting societies, dreams may link a hunter’s psyche with that of an individual prey animal, or with a master-soul of the species, who guides the hunter to a waking bodily form as part of a reciprocal life-death compact between human beings and game animals from originary times. But in all cases, interpretive criticism and

failure are centrally acknowledged, belying an inference of naive mysticism.

In some cultures, a *totem* (an Ojibwa word from the Great Lakes - far from the misnamed ‘totem poles’ of the Northwest Coast), or guardian-spirit, watches over a person’s life. It may be appointed at birth or acquired in a ‘vision-quest’, especially at puberty. The widespread vision-quest, for both sexes, is a socially sanctioned will-to-power, involving fasts on mountain tops or other secluded spots deemed powerful, where through privation the individual seeks a transcendent vision, which, though more intense, is like ordinary dreaming in its informing of conscious action. If successful, a spirit-animal or deity representative appears, and conveys a personal power that may subsequently be invoked at will.

4 Ethics

Major visions were used to guide community actions and were ritually enacted in collective performance. The mutual dependence of self and society is also strikingly illustrated in the Plains Sun Dance. Individual men may elect to undergo extreme physical pain by piercing muscles with rawhide thongs attached to a pole representing the *axis mundi*, and pulling away until the flesh is broken or they lose consciousness - an act of self-sacrifice for the spiritual benefit of the whole community.

The individuated yet relational self is often associated with concrete ethical manifestations in objective reality. For the Hopi, ethical values prescribe modesty, self-effacement, interpersonal and environmental responsibility, industry, courtesy, circumspection, and humility. *Natwani*, the exterior reflection of one’s performance, may refer to crops, children or other fruits of personal effort; if these turn out well, they accrue to the individual’s virtue as ‘*hopi*’ (an ethical term in itself, encapsulating the values mentioned). Ethically composed and judged selves in intersubjective relationship engender and reproduce moral community. After detailed field study, philosopher Richard Brandt (1954: 72) concluded that, ‘it is doubtful whether there is any evidence for ethical attitudes or remorse in Western society which cannot be matched among the Hopi’.

Ladd (1957: 212-13) interprets Navajo ethics as ‘rational’ (especially in terms of moral justification for action), ‘prudential’ (particularly regarding the individual’s welfare), ‘egoistic’ (promoting individual satisfaction), and ‘utilitarian’ (aimed at happiness for all) (see Egoism and Altruism; Utilitarianism). The central goal is *Są’ah Naagháii Bik’eh Hózho* which refers, *inter alia*, to the continuous reproduction of beauty, health, long-life and happiness for self and community (Witherspoon 1977: 13-46).

In many Native societies, high value is placed on both individual autonomy and harmonious cooperation. Coercion is rare and social adherence depends on individual assent. Conformism is perhaps most strongly emphasized in Pueblo values, although the standard Hopi phrase ‘*pi um ‘i’*’ (‘it is up to you’), stresses individual choice. Individual autonomy is especially strong in northern hunting societies, but here too is accompanied by a powerful sense of mutual reliance. While there are exceptions (as in the institution of partial slavery in Northwest Coast societies), this conjoint ethos of individual autonomy and interpersonal responsibility is so widespread it represents a deep-seated moral choice in the Native American *conscience collective*, and an implicit rejection of the Enlightenment social contract, conceived as the surrender of ‘natural rights’ to a sovereign, from which much contemporary Western social and political philosophy stems.

Ethical principles are also extended to non-human ‘persons’, particularly game-animals. According to Penobscot (Eastern Algonquians):

> The animals know beforehand when they are to be slain, when their spirits have been overcome by the hunter’s personal power or magic. They exact also a certain respect, which is shown to them in different ways, through generally by the proper treatment of their remains after they have been killed for their flesh or skins. If this respect is not accorded to them, the animals refuse to be killed or their souls may not be reborn.
>
> (Speck 1935: 32-3)

How, though, do these beliefs differ from mere anthropomorphism? Rock Crees ‘ascribe animateness, self-awareness, intelligence, and sometimes covert humanoid characteristics to many non-human beings and objects’. But they simultaneously depend on a detailed pragmatic knowledge that emphasizes human-animal differences that are ‘constitutive of and endlessly objectified in the human practice of killing and eating animals’ (Brightman 1993: 77, 179). The dualism of ethical consubstantiation and pragmatic differentiation reflects the

central metaphysical dichotomy between ‘matter and essence’ noted above for Pueblo world-view. It is the essential forms of nature that are included within the ethical sphere of ‘persons’; the material forms are objects of pragmatic interaction.

5 Epistemology and reason

Native knowledge systems are grounded in the metaphysical and ontological principles so far described. In Hopi, knowledge as an abstraction is denoted by *navoti*, the closest equivalent to ‘philosophy’, but also implying such European paradigms as *scientia*, theology, epistemology, psychology, hermeneutics and with a determinist cast in which events are fore-ordained (with a causal nexus in mystical forces). What is known through *navoti* may, by the same token, be acted upon. *Navoti*, especially in ritual use, is, like Navajo speech, causative: a *pouvoir/savoir* addressed to supernatural causes attributed with conscious animate force.

Many Native American societies denominate social units by natural species and elements: the Hopi have more than thirty matrilineal clans, including Bear, Spider, Greasewood, Badger, Butterfly, Sun, Parrot and Eagle. Lévi-Strauss has produced a penetrating analysis of this ‘totemic logic’, which, he argues, is fundamentally relational:

> The homology they [totemic institutions] evoke is not between social groups and natural species but between the differences which manifest themselves on the level of groups on the one hand and on that of species on the other. They are thus based on the postulate of a homology between *two systems of differences*, one of which occurs in nature and the other in culture.
>
> (Lévi-Strauss 1966: 115)

Natural classes are thus ‘good to think’, that is, as metaphorical vehicles for conceptual differentiation - behaviourally, epistemologically and socially. The inferred mode of thought is fundamentally analogical, operating with a ‘system of concepts embedded in images’ that serve to build ‘mental structures that facilitate an understanding of the world in as much as they resemble it’ (Lévi-Strauss 1966: 264, 263).

Lévi-Strauss extrapolates Native American systems of thought from mythological narratives (especially), social forms, arts and architecture. In myth, embedded codes of meaning operate both episodically (‘syntagmatically’) and thematically (‘paradigmatically’). The narratives unfold via a series of binary oppositions ranged along multiple intersecting planes, including geography, economy, sociology and cosmology - ‘each one of these levels together with the symbolism proper to it, being seen as a transformation of an underlying logical structure common to all of them’ (Lévi-Strauss 1968: 1). The oppositions derive from contradictory principles in social experience. Myths, then, seek to ‘resolve’ paradoxes and contradictions, which are in fact irresolvable, by correlating oppositions from the different planes and postulating syntheses. In a Tsimshian (Northwest Coast) story, for example, Asdiwal, a supernatural transformer-hero, is the figure of attempted synthesis between (*inter alia*): endogamy versus exogamy, famine versus plenty, salmon versus candlefish, mountain-hunting versus sea-hunting, female versus male, movement versus immobility, patrilocal residence versus matrilineal descent, downstream versus upstream, mother versus daughter, and heaven versus earth. The mythic interplay of these oppositions configures the attempt epistemologically to overcome their individual contradictions. This ‘fundamental notion of dualism in a perpetual state of disequilibrium’ Lévi-Strauss sees as pervading Native American systems of thought:

> In Amerindian thought a sort of philosophical bias seems to make it necessary for things in any sector of the cosmos or society to not remain in their initial state and for an unstable dualism to always yield another unstable dualism, regardless of the level on which it might be apprehended.
>
> (Lévi-Strauss 1995: 231)

Lévi-Strauss’ position is persuasive. Dual forms of social and conceptual organization are common, from the Red-White social binaries of the Creek confederacy to paired masks of the Kwakiutl and Salish, the Winter-Summer Tewa moieties, diametric town-plans of the Winnebago and oppositive species in Keresan emergence myths. Ortiz (1969) confirms a pervasive dualism in Tewa culture, but challenges its supposed insolubility. Tewa social oppositions are resolved cyclically through temporal alternations where one or the other of the Summer-Winter pair has successive executive authority. Herein lies an indication of the limits of

Lévi-Strauss' interpretation of Native thought, removed as it is from first-hand empirical inquiry.

The Hegelian and Kantian foundations of Lévi-Strauss' approach are clear: the former in his dialectical reason and the latter with his tacit reliance on a transcendental subject. Both principles are, then, inferred as key components of Native American thought *sui generis*. But if a major logical process in Native thought systems is autogenously dialectical, this operates largely without the conscious intervention of any Native thinker: Lévi-Strauss (1970) holds that mythological thematic operators (oppositions, or figures like Asdiwal himself) *think themselves* out through individual minds and tellings rather than vice versa.

While of great heuristic value, Lévi-Strauss' semiotic reductions result in an abstract logical and aesthetic formalism, where the endless interplay of binary oppositions becomes a hall of mirrors (see Lévi-Strauss, C.). In use in Native discourse, species names and denotations are both literal and often multiply symbolic. A re-analysis of Hidatsa names with Frege's notions of sense and reference opens out the understanding of Native signification considerably further (Barnes 1980). Similarly, for the Hopi, who continue to gather fledgling eagles each May, species meanings are more than mere logical abstractions. Eagles are 'persons' from a Hopi point of view and reincarnate clan ancestors. Eagles are, then, good to think, but they are also sentient spirits who are good to be with and who mediate prayers to deity, in a cosmos saturated with sentient subjects rather than rote mechanistic forms.

Particularly for shamans, narratives of personal relationship with animals and elements are frequent. Eskimo shamans, for example, were thought capable of detaching part of their selves from their bodies and visiting the under-sea homes of deities. Some shamanic initiations among Tikigaq Eskimos involve metamorphosis into animal forms, like whales. Yet indigenous scepticism of both particular claims and general principles is clearly apparent (Lowenstein 1992: 91, 78-84). A well-known example of scepticism involves a Kwakiutl man, Quesalid, who doubts shamanic claims. He proceeds, as an empirical test, to apprentice himself, and goes about exposing the fraud of others. But he also becomes a highly successful curer himself. Quesalid's apprenticeship does not entail a transition to uncritical belief, but rather a comprehension of the social and psychological contexts in which shamanic techniques productively operate.

If in origin narratives a central concern is ordering by categorial differentiation, and if ritual's key intent is to exert control over a deterministic world, these are always accompanied by a simultaneous indeterminacy principle. In myths, this is instantiated by Trickster, and in rituals, by formal clowning and games. Trickster resists structure, is instinctual and without ethics or conscience. He/she/it is (willy-nilly) a major creative force in some instances, in others ancillary to more prominent creators. But everywhere, Trickster - embodied as Coyote, Hare or Raven, or a bodiless figure of metamorphosis - represents the chance element, the subversion and transgression of order, and, through selfish desire, vanity, lust, greed and gluttony, is both inverse ethical exemplar and an avenue of comic release and communal therapy. Trickster is a protohuman cheerleader for flawed humanity, who mocks the claims to righteousness of gods or pious humans. But Trickster is, reflexively, a satire of creatural vanity, displaying an ironic, ambivalent sense of the human condition marked by a continual failure to live up to ideals, and he/she/it ostensively defines the provisionality of all established forms:

> The symbol which Trickster embodies is not a static one. It contains within itself the promise of differentiation, the promise of god and man…. He represents not only the undifferentiated and distant past, but likewise the undifferentiated present within every individual. This constitutes his universal and persistent attraction. And so he became and remained everything to every man - god, animal, human being, hero, buffoon, he who was before good and evil, denier, affirmer, destroyer and creator. If we laugh at him, he grins at us. What happens to him happens to us.
>
> (Radin 1956: 168-9)

In Winnebago narratives, *wadjunkaga*, 'the tricky one', passes through a series of episodes in a developmental cycle; along the way, his left and right hands fight and cause serious injury to each other, he wounds his own (personified) anus for failing to guard a brace of ducks, eats his own intestines, punishes a chipmunk with his penis that gets chewed off in return, changes sex to marry a chief's son, claims to be a war chief but systematically violates all the taboos, and so on. The cycle depicts a journey from a formless, childlike id ruled by immediate sensory and intellectual gratification, in which everything is misconceived - biology, cosmology, social roles, sexuality, mortality and morality - towards biological, social, and psychological definition. In a conscious

mythological world where other creatures comment adversely on his exploits, Trickster the protohuman blunders about, asocial and isolated, largely unconscious of the effects of his actions. Logically and cosmologically, trickster is antithesis, subversion, imbalance.

Mixed-blood Anishinabe philosopher Gerald Vizenor extends Trickster into his own work, showing further how it operates as a sign. He writes both *of* and *as* Trickster against the reification of social-scientific analysis:

> Tricksters, the wild ironies of survivance, transformation, natural reason, and liberation in stories, are marooned as obscure moral simulations in translations…. *The trickster is a language game*, a counter causal liberation of the mind.
>
> Trickster is androgynous, a comic healer and liberator in literature: the whole figuration that ties the unconscious to social experience. The trickster sign is communal, an erotic shimmer in oral traditions; the narrative voices are *holotropes* [total figures] in a discourse.
>
> (Vizenor 1994: 76-7, 89, first emphasis added)

Trickster frees from structure, a semiotic sign forever darting in and out of established meaning and reference (including its own), and a Native counterpart to postmodernism's free play of signification and subversion of master-narratives (see Postmodernism).

Hopi clowns play out the same philosophical drama in a two-day ceremony. To begin with, clowns, *tsutskut* ('the pointed ones', after a feature of their costume, but also 'those who are making a point' - the double meaning also works in Hopi) specifically symbolize undifferentiated consciousness, humanity in a childlike state innocent of understanding, virtue or evil. During a line-dance performance by kachinas (masked representations of spirits, which here embody pure essences of the differentiated cosmos) unmasked clowns appear in counterpoint, lampooning everything sacred or normative in Hopi values, and following Trickster's trajectory of semi-successful category learning by the end of the ceremony. At a critical moment, under threat of death by moralizing warrior kachinas, they agree to try order, and to demonstrate this they imitate the kachinas in the line-dance. After a few tries, however, they fall apart, hopelessly incapable.

Several key themes of Native American philosophy are conjoined in this performance. When they first appear in the plaza, clowns are the children of the Sun arriving in a formless void (initially they do not notice the kachinas) where they start iconically building the world and their place within it, though with frequent reversals of contemporary significations. Their transgressions are comically played out for the Hopi audience, which then reflexively inspects the world's categories and codes, re-notes their arbitrariness and provisionality, delights in the cathartic subversion of the intending human self, its acts and foibles, and is reminded of the virtue and vitality of created natural and social order. This play of order and disorder dialectically deconstructs the human condition - cosmologically, metaphysically, ontologically, biologically, ethically and epistemologically. Trickster, in a word, is the philosopher.

See also: Latin America, philosophy in

PETER M. WHITELEY

References and further reading

Bahr, D., Smith, J., Smith Allison, W., and Hayden, J. (1994) *The Short, Swift Time of the Gods on Earth: The Hohokam Chronicles*, Berkeley, CA: University of California Press.(An oral record of a prehistoric Hohokam origin myth, which attests to the depth of phased emergence cosmology in Southwest thought.)

Barnes, R.H. (1980) 'Hidatsa Personal Names: an Interpretation', Plains *Anthropologist* 25 (90): 311-31. (Philosophically sophisticated, cogent analysis of Hidatsa naming and subjectivity, employing Frege's conception of sense and reference.)

Basso, K.H. (1976) ''Wise Words' of the Western Apache: Metaphor and Semantic theory', in K.H. Basso and H.A. Selby (eds) *Meaning in Anthropology*, Albuquerque, NM: University of New Mexico Press.(A meticulous account of Apache metaphor that instances unsuspected modes of analogical reason in Apache thought.)

Basso, K.H. (1988) ''Speaking with Names'; Language and Landscape among the Western Apache', *Cultural Anthropology* 3 (2): 99-130.(Analyzes Apache discursive use of place-names as metonyms of moral narratives; shows how the intersubjectively cognized landscape may serve as a philosophical text in oral societies, suggesting that argument is not limited to the subjective authority of interior memory.)

Basso, K.H. (1996) *Wisdom Sits in Places: Landscape and Language among the Western Apache*, Albuquerque, NM: University of New Mexico Press.(Deeply insightful analysis of Apache discursive use of place-names as metonyms of moral narratives; shows how the intersubjectively cognized landscape may serve as a philosophical text in oral societies, suggesting that argument is not limited to the subjective authority of interior memory; assesses some Apache concepts of mind versus brain.)

Benedict, R. (1934) *Patterns of Culture*, Boston, MA: Houghton Mifflin.(Argues for modal personality types in the autonomy versus conformity interplay, famously contrasting Pueblo ('Appollonian' - sober, moderate, ritually compulsive, and individually repressed) and Kwakiutl ('Dionysian' - rivalrous, ecstatic, and marked by delusions of grandeur) in a manner oversimplified at best.)

Boas, F. (1930) 'The Religion of the Kwakiutl', *Columbia University Contributions to Anthropology* 10 (II): 1-41.

Brandt, R.B. (1954) *Hopi Ethics: a Theoretical Analysis*, Chicago, IL: University of Chicago Press.(A thorough field study in the 1940s by an academic philosopher interviewing numerous Hopis.)

Brightman, R.A. (1990) 'Primitivism in Missinippi Cree Historical Consciousness', *Man* (n.s.) 25: 399-418. (Insightful, approachable reflection on Cree evolutionary epistemology.)

Brightman, R.A. (1993) *Grateful Prey: Rock Cree Human-Animal Relationships*, Berkeley, CA: University of California Press.(An exemplary ethnographic inquiry into Cree natural philosophy, ethics and metaphysics.)

Brody, H. (1981) *Maps and Dreams*, New York: Pantheon.

Closs, M.P. (1986) 'Native American Numbering Systems', in M.P. Closs (ed.) *Native American Mathematics*, Austin, TX: University of Texas Press.

DeMallie, R.J. (1987) 'Lakota Belief and Ritual in the Nineteenth Century', in R.J. DeMallie and D.R. Parks (eds) *Sioux Indian Religion: Tradition and Innovation*, Norman, OK: University of Oklahoma Press.(Fine synthetic analysis of Lakota metaphysics and ontology within a religious frame of reference.)

DeMallie, R.J. and Lavenda, R.H. (1977) '*Wakan*: Plains Siouan Concepts of Power', in R.D. Fogelson and R.N. Adams (eds) *The Anthropology of Power: Ethnographic Studies from Asia, Oceania and the New World*, New York: Academic.(Lucid discussion of the concept of pervasive animate force suffusing the cosmos in Sioux thought.)

Derrida, J. (1977) *Of Grammatology*, trans. G.C. Spivak, Baltimore, MD: Johns Hopkins University Press. (Includes a trenchant critique of the spectral transcendental subject in Lévi-Strauss' method, but which fails to displace his explication of binarism and semiosis in Native thought.)

Devereux, G. (1951) *Reality and Dream: Psychotherapy of a Plains Indian*, New York: International Universities Press.

Farella, J.R. (1984) *The Main Stalk: a Synthesis of Navajo Philosophy*.(A useful summary of Navajo religious and mythological thought.)

Goldman, I. (1981) *The Mouth of Heaven: an Introduction to Kwakiutl Religious Thought*, Huntington, NY: Krieger, 2nd edn.(A somewhat controversial but profoundly insightful analysis of Kwakiutl metaphysics and symbolism.)

Golla, S. (1988) 'Persons, Names and Selves among Traditional Nuu-chah-nulth', paper presented to the symposium 'Persons and Selves in Pueblo and Northwest Coast Societies: Marcel Mauss Reconsidered', American Anthropological Association Annual Meetings, Phoenix, AZ.(Penetrating, user-friendly analysis of Nuu-chah-nulth concepts of mind, subjectivity and action.)

Hallowell, A.I. (1966) 'The Role of Dreams in Ojibwa Culture', in *Contributions to Anthropology: Selected Papers of A. Irving Hallowell*, Chicago, IL: University of Chicago Press, 1976.

Heidegger, M. (1927) *Being and Time*, trans. J. Macquarrie and E. Robinson, New York: Harper & Row, 1962.

Hewitt, J.N.B. (1906) 'Orenda', in F.W. Hodge (ed.) 'Handbook of American Indians', *Bureau of American Ethnology Bulletin* 30 (II): 147-8.(A still-valuable summary of Iroquois metaphysical thought; the author, an anthropologist, was Tuscarora (Iroquois).)

Hultkrantz, A. (1952) *Conceptions of the Soul among North American Indians: a Study in Religious Ethnology*, Stockholm: The Museum of Sweden.(A wide- ranging summary of Native American ideas of the soul, but rather difficult reading for its minute detail.)

Joseph, A. (ed.) (1992) *America in 1492: the World of the Indian Peoples before the Arrival of Columbus*, New York: Alfred A. Knopf.(Provides an overview of Native cultures and languages.)

Kan, S. (1989) *Symbolic Immortality: the Tlingit Potlatch of the Nineteenth Century*, Washington, DC: Smithsonian Institution Press.(Analytically sophisticated and ethnographically rich - particularly interesting on Tlingit concepts of reincarnation and their implications for concepts of the self.)

Keeling, R. (1992) *Cry for Luck: Sacred Song and Speech among the Yurok, Hupa, and Karok Indians of Northwestern California*, Berkeley, CA: University of California Press.(A nice synthesis of 'World Renewal'

ceremonies in Northwest California, illustrating the attempt in ritual to recapitulate primordiality and evolutionary transformation in the (re)production of natural and cultural order.)

Ladd, J. (1957) *The Structure of a Moral Code: A Philosophical Analysis of Ethical Discourse applied to the Ethics of the Navaho Indians*, Cambridge, MA: Harvard University Press.(A philosophically detailed and informed attempt at intercultural interpretation of ethics; Ladd is criticized by Navajo specialists for his reliance on a single informant, yet the inferences are often sophisticated and ring true, although perhaps excessively dependent on a Western analytical model.)

Lafitau, J.F. (1724) *Moeurs des sauvages amériquains comparées aux moeurs des premiers temps*, Paris, 4 vols; trans. and ed. W.N. Fenton and E. Moore as *Customs of the American Indians compared with the customs of primitive times*, Toronto, Ont.: Champlain Society, 1974-7.(One of the earliest serious attempts at ethnographic description of ethics and beliefs, based on the long-term work of the Jesuit missions among Huron and Iroquois - see Thwaites (below).) *

Lame Deer, J.P. and Erdoes, R. (1972) *Lame Deer, Seeker of Visions*, New York: Simon & Schuster.(An engaging, genial biography of a twentieth-century Lakota shaman.)

Lévi-Strauss, C. (1963) 'Do Dual Organizations Exist?', *Structural Anthropology*, trans. C. Jacobson and B.G. Schoepf, New York: Basic Books.

Lévi-Strauss, C. (1966) *The Savage Mind*, Chicago, IL: University of Chicago Press. (Arguably the twentieth-century's most brilliant attempt to engage intellectually with the epistemology and ontology of non-Western cultures, with some particular emphasis on thought in the indigenous Americas. Occasionally intractable, Lévi-Strauss is consistently provocative, philosophically speaking, and simultaneously keeps faith with some basic premises of Amerindian conception.)

Lévi-Strauss, C. (1968) 'The Story of Asdiwal', trans. N. Mann, in E.R. Leach (ed.) *The Structural study of Myth and Totemism*, London: Tavistock.(Perhaps the most cogent single demonstration by Lévi-Strauss of the multiple planes on which Native American mythological dialectics operate.)

Lévi-Strauss, C. (1970) *The Raw and the Cooked: Introduction to a Science of Mythology*, London: Jonathan Cape.(First of the four-volume magnum opus *Mythologiques* (the others are *From Honey to Ashes*, *The Origin of Table Manners*, and *The Naked Man*), analysing the mythology of the Americas.)

Lévi-Strauss, C. (1982) *The Way of the Masks*, trans. S. Modelski, Seattle, WA: University of Washington Press. (A structuralist analysis of ritual mask symbolism in Northwest Coast societies, especially Kwakiutl and Salish.)

Lévi-Strauss, C. (1995) *The Story of Lynx*, trans. C. Tihanyi, Chicago, IL: University of Chicago Press.(An analysis of Native American thought via myths from the Plateau and Northwest Coast.)

Lincoln, J.S. (1935) *The Dream in Primitive Cultures*, London: Cresset.(Though limited interpretatively by its time, contains substantive discussion of psychology and philosophy of dreams, with detailed cases from Native North America.)

Lowenstein, T. (1992) *The Things That Were Said of Them: Shaman Stories and Oral Histories of the Tikigaq People told by Asatchaq*, trans. Tukummiq and T. Lowenstein, Berkeley, CA: University of California Press.(A sensitive approach to Eskimo thought through oral literature.)

Malotki, E. (1982) *Hopi Time: a Linguistic Analysis of the Temporal Concepts in the Hopi Language*, The Hague: De Gruyter & Mouton.(A detailed linguistic analysis that effectively dislodges Whorf's view of timelessness in Hopi linguistic forms.)

McNeley, J.K. (1981) *Holy Wind in Navajo Philosophy*, Tucson, AZ: University of Arizona Press.(Focuses on *nil'chi*, 'wind', 'air', or 'atmosphere', the Navajo version of pervasive animate force.)

Neihardt J.G. (1932) *Black Elk Speaks: Being the Life Story of a Holy Man of the Oglala Sioux*, New York: William Morrow.(A now controversial, but widely popular, account of a Lakota shaman's life and thought, which remains valuable for its insights into Sioux metaphysics; the controversy derives from the extent of editorial intervention that, *inter alia*, neglects Black Elk's active Catholicism by the time his autobiography was dictated.)

Ortiz, A. (1969) *The Tewa World: Space, Time, Being and Becoming in a Pueblo Society*, Chicago, IL: University of Chicago Press.(A superb, detailed analysis of Tewa cosmology and metaphysics by an anthropologist from San Juan (a Tewa) Pueblo.)

Ortiz, A. (1972) 'Ritual Drama and the Pueblo World View' in A. Ortiz (ed.) *New Perspectives on the Pueblos*, Albuquerque, NM: University of New Mexico Press.(A development of the above, focused on Pueblo ritual.)

Pomedli, M.M. (1991) *Ethnophilosophical and Ethnolinguistic Perspectives on the Huron Indian Soul*, Lewiston, NY: Edwin Mellen.(One of the best analyses of seventeenth and eighteenth century Jesuit records from the Huron missions - see Thwaites below.)

Radin, P. (1927) *Primitive Man as Philosopher*, 2nd revised edn, New York: Dover, 1957.(Originally published in , this is now dated, but incudes insightful discussion of Radin's philosophical inquiries with Native Americans, particularly Winnebago, and useful interpretations of the philosophies of other Native societies.)

Radin, P. (1945) *The Road of Life and Death: a Ritual Drama of the American Indians*, Bollingen Series V, New York: Pantheon.(A thorough, albeit authorially idiosyncratic, account of a key Winnebago ritual, The Medicine Rite, with useful inferences particularly regarding metaphysics and eschatology.)

Radin, P. (1956) *The Trickster: A Study in American Indian Mythology*, New York: Philosophical Library.(The classic account of the Winnebago Trickster narrative cycle, with comparisons to other Native American Trickster myths, and both philosophically and psychologically inflected interpretations; C.G. Jung offers his usual archetypalist view.)

Rasmussen, K. (1929) 'Intellectual Culture of the Iglulik Eskimos', *Report of the Fifth Thule Expedition* VII (1): 1921-24, Copenhagen: Gyldendalske Boghandel, Nordisk Verlag.

Ridington, R. (1987) 'Fox and Chickadee', *Little Bit Know Something: Stories in a Language of Anthropology*, Iowa City, IA: University of Iowa Press, 1990.(The whole volume brings together some of Ridington's most useful writings on northern Athapaskan religious and philosophical thought, written in a clear, jargon-free style.)

Ridington, R. (1988) *Trail to Heaven: Knowledge and Narrative in a Northern Native Community*, Iowa City, IA: University of Iowa Press.(An insightful examination of Beaver mystical and pragmatic thought.)

Ridington, R. and Ridington, T. (1975) 'The Inner Eye of Shamanism and Totemism', in D. Tedlock and B. Tedlock (eds) *Teachings from the American Earth: Indian Religion and Philosophy*, New York: Liveright.

Rushforth, S. (1992) 'The Legitimation of Beliefs in a Hunter-gatherer Society: Bearlake Athapaskan Knowledge and Authority', *American Ethnologist* 19: 483-500.(A sharply focused examination of the social and epistemological contexts of Bearlake truth-evaluation and judgment of knowledge-claims.)

Sartre, J.-P. (1958-9) *Critique of Dialectical Reason*, London: New Left Books, 1976.(Includes an attack on Lévi-Strauss' dialecticalism, but as with Derrida (1977) fails to address the persuasive analogical disclosures in Lévi-Strauss' analyses of native representations.)

Sebag, L. (1971) *L'invention du mond chez les indiens pueblos*, Paris: Maspero.(A penetrating structuralist analysis of Keresan Pueblo emergence myths.)

Speck, F. (1935) 'Penobscot Tales and Religious Beliefs', *Journal of American Folklore* 48: 1-107.(Nice illustration of Native American moral extensions to animate environmental forms.)

Sturtevant, W. (ed.) (1979-) *Handbook of North American Indians*, Washington, DC: Smithsonian Institution Press, 20 vols.(This emerging twenty-volume work is the principal source to begin all inquiries into Native American culture, history, and thought.)

Sullivan, L.E. (ed.) (1989) *Native American Religions: North America*.(Includes very useful articles excerpted from the 1987 Macmillan *Encyclopedia of Religion*.)

Tanner, A. (1979) *Bringing Home Animals: Religious Ideology and Mode of Production of the Mistassini Cree Hunters*, New York: St Martin's Press.(A particularly fine-grained and thoughtful explication of Cree thought.)

Tedlock, B. (1992) *The Beautiful and the Dangerous: Encounters with the Zuni Indians*, New York: Viking.(A personalized account of the author's long-term ethnographic work at Zuni Pueblo: contains nuanced interpretations of Zuni aesthetics, and natural and social philosophy.)

Tedlock, B. and Tedlock, D.(eds) (1975) *Teachings from the American Earth: Indian Religion and Philosophy*, New York: Liveright.(A useful collection of interpretations, with discussions of several philosophical perspectives.)

Tedlock, D. (1979) 'The analogical Tradition and the Emergence of a Dialogical Anthropology', *Journal of Anthropological Research* 35: 387-400.

Tedlock, D. (1983) *The Spoken Word and the Work of Interpretation*, Philadelphia, PA: University of Pennsylvania Press.(Contains several searching analyses (including the preceding item) of complexities of translation, particularly from oral languages to written, with a major focus on Zuni.)

Thwaites, R.G. (ed.) (1896-1901) *The Jesuit Relations and Allied Documents, Travels and Explorations of the Jesuit Missionaries in New France, 1610-1791*, Cleveland, OH: Burrows Brothers, 73 vols.(*The* source for early European interpretation of Iroquoian philosophy, from accounts of the missions.)

Vizenor, G. (1989) '*Trickster Discourse*: Comic Holotropes and Language Games', in G. Vizenor (ed.) *Narrative Chance: Postmodern Discourse on Native American Indian Literatures*, Albuquerque, NM: University of New Mexico Press.(An inquiry into the meaning of Trickster as a sign in Native discourse.)

Vizenor, G. (1994) *Manifest Manners: Postindian Warriors of Survivance*, Hanover, NH: Wesleyan University Press.(A literary and philosophical critique of representations of Native identity and experience, including

further reflections on Trickster. Vizenor, a mixed-blood Anishinabe, is one of the most philosophically engaged contemporary Native scholars, and his writings, often Trickster-esque, offer subtle insights into Native thought, informed by poststructuralist theory.)

Walens, S. (1981) *Feasting with Cannibals: an Essay on Kwakiutl Cosmology*, Princeton, NJ: Princeton University Press.(A searching examination of Kwakiutl metaphor and metaphysics, based on the ethnography of Franz Boas.)

Walker, J.R. (1896-1914a) *Lakota Belief and Ritual*, ed. R.J. DeMallie and E.A. Jahner, Lincoln, NB: University of Nebraska Press, 1980.(Walker's field notes of interviews with Lakota intellectuals, collected while he was Agency physician at the Pine Ridge Reservation.)

Walker, J.R. (1896-1914b) *Lakota Myth*, ed. E.A. Jahner, Lincoln, NB: University of Nebraska Press, 1983.(More of Walker's field notes of interviews with Lakota intellectuals)

Walker, J.R. (1917) 'The Sun Dance and Other Ceremonies of the Oglala Division of the Teton Dakota', *Anthropological Papers of the American Museum of Natural History* 16 (2): 51-221.(A classic account of Oglala religious belief and practice, with some direct commentary by Oglala shamans.)

Wallace, A.F.C. (1969) *The Death and Rebirth of the Seneca*, New York: Random House.(A historically focused interpretation of Seneca (Iroquois) religious revitalization, with a psychologically informed analysis of Seneca thought.)

Whiteley, P.M. (1988) *Deliberate Acts: Changing Hopi Culture through the Oraibi Split*, Tucson, AZ: University of Arizona Press.(Extends some of the author's observations on Hopi intentionality and philosophy of action in social and historical context.)

Whiteley, P.M. (1992) 'Burning Culture: Auto-da-fe at Oraibl', *History and Anthropology* 6: 46-85.(Examines intentionality and action in a Hopi historical event, drawing especially on the formulations of Donald Davidson and Richard Wollheim.)

Whorf, B.L. (1956) *Language, Thought and Reality*, ed. J.B. Carroll, Cambridge, MA: MIT Press.(Selected writings of Benjamin Lee Whorf who, with Edward Sapir, is the key figure in linguistic relativism, which suggests that differences in thought are systemically conditioned by language differences, resulting in incommensurable phenomenologies: this anthology contains Whorf's central arguments, with specific foci on Hopi language and thought.)

Witherspoon, G. (1977) *Language and Art in the Navajo Universe*, Ann Arbor, MI: University of Michigan Press. (A linguistically and culturally well-informed philosophical analysis of Navajo metaphysics and epistemology.)

Nativism

Traditional empiricism claims that the mind is initially equipped only with the capacity for experience and the mechanisms that make it possible for us to learn from experience. Nativists have argued that this is not enough, and that our innate endowment must be far richer, including information, ideas, beliefs, perhaps even knowledge.

Empiricism held the advantage until recently, partly because of a misidentification of nativism with rationalism. Rationalists such as Descartes and Leibniz thought nativism would explain how a priori knowledge of necessary truths is possible. However, the fact that something is innate does not establish that it is true, let alone that it is necessary or a priori.

More recently, nativism has been reanimated by Chomsky's claims that children must have innate language-specific information that mediates acquisition of their native tongue. He argues that, given standard empiricist learning procedures, the linguistic data available to a child underdetermines the grammar on which they converge at a very young age, with relatively little effort or instruction.

The successes in linguistics have led to fruitful research on nativism in other domains of human knowledge: for example, arithmetic, the nature of physical objects, features of persons, and possession of concepts generally.

1 Nativism and empiricism

The fact that we human beings understand the world to the extent that we do depends on two factors: our inherent nature as thinkers, and the specific course of our experience. One of the great oppositions in the history of philosophy has centred on the relative contributions of each of these factors. Empiricists hold that experience provides all the raw materials out of which knowledge and understanding is fashioned: the mind is often compared to a *tabula rasa* (blank slate) initially equipped only with the capacity for experience and the ability to build upon and learn from past experience. Nativists have argued that this is not enough, and that our innate endowment is richer than the empiricist allows (see Empiricism).

Although nativists agree that empiricism is mistaken, they differ widely over the sorts of things that are innate - for example, ideas, principles, knowledge, dispositions, biases, representations, concepts, strategies, theories - as well as about the domains in which innateness must be invoked: Leibniz argued for it primarily for logic and mathematics; Plato for all *genuine* knowledge (see Innate knowledge). The matter is further complicated by the fact that empiricists themselves must inevitably incorporate some nativist elements in their own theorizing. At the very least, the ability to experience the world and the basic machinery that makes building on such experience possible must be innate and not learned. But empiricists typically argue that such commitments are innocuous, not specific to any domain, and (especially in recent empiricism) do not involve the elaborate mental machinery that nativists standardly posit.

Perhaps what unifies all nativist positions is a reliance on what Noam Chomsky has called the 'poverty of the stimulus' argument scheme. Nativists point to some discrepancy between our experience (the stimulus) and the knowledge (or ideas, concepts, principles, and so on) that we eventually come to have, and argue that this discrepancy cannot be made up by invoking the sorts of simple learning mechanisms posited by empiricists - association and conditioning, for example. It follows that we must be contributing something substantial 'from our side' in the construction of knowledge, and this is what is innate.

The earliest nativist argument is in Plato's *Meno*, where Socrates' questioning enables Meno's uneducated slave to discover 'for himself' a theorem of geometry (*Meno* 80a-86c). Socrates argues that his questions have 'reminded' the slave of what he knew innately. Indeed, for Socrates, all knowledge (properly construed) is acquired by recollecting what is innate in the soul from an earlier existence (the doctrine of anamnesis - see Plato §11). In the *Phaedo* (73c-78b) is a related argument: Socrates claims that the notion of equality involved in perceiving a pair of sticks as equal could not have its source in experience and therefore must be innate.

A similar pattern of argument pervades seventeenth-century rationalist defences of innateness. Descartes and Leibniz argue that our knowledge of the necessary truths of mathematics seems to be justifiable a priori, or 'independently of experience': mathematicians do not need to perform experiments to test whether $2+2=4$.

They thought that nativism would explain how such knowledge was possible (see Rationalism).

However, this is a dubious explanatory strategy. Nativism is a claim about the *origins* of knowledge; *rationalism* is a claim about its *justification*. The fact that something is innate does not establish that it is true, let alone that it is necessary or a priori: someone might be born believing that space is Euclidean or that smiling faces are to be trusted, neither of which is even true.

Empiricists attacked nativism in two ways. One approach challenged the legitimacy of nativism as a plank in a theory of knowledge. Locke, for instance, argued that nativism reduces either to the nonsensical view that we are, in an everyday sense of the term, literally born knowing things, or to the trivially true and empty view that we have an innate capacity to think and to discover truths. Of course, he then needed an account of how knowledge was possible. He and later empiricists offered what may be termed 'adequacy of the stimulus' counter-arguments (see Behaviourism, methodological and scientific; Hume, D.; Logical positivism; Quine, W.V.). They constructed theories of the mechanisms of inductive inference, of concept formation and so on, and applied them to specific cases (for example, our idea of infinity, our knowledge of general causal laws) to argue that such mechanisms can indeed generate our knowledge given the raw materials of experience. This philosophical effort was in the twentieth century reinforced by the concurrent growth of an independent science of associationist-behaviourist psychology, as well as, more recently, connectionist models of cognitive processing, all based on empiricist principles (see Connectionism).

2 Linguistic nativism

The dominant philosophical view for most of the twentieth century was that the empiricists had won. But all this changed in the 1960s, due in large measure to Noam Chomsky's arguments that empirical research in linguistics vindicated the rationalist innateness doctrine (as well as its attendant mentalism) over its empiricist counterpart (see Chomsky, N.; Language, innateness of).

Chomsky's ground-breaking work has been largely responsible for the development of linguistics as a deeply theoretical empirical science. He argued that language users have an internal representation of a set of rules that relate utterances to meaning and thereby enable them to understand and produce sentences of their language. These rules constitute the grammar of the particular language in question, and determine which patterns constitute well-formed sentences of the language and which do not. The central finding of modern linguistics is that the grammatical rules that govern natural languages are surprisingly complex and non-obvious: they cannot be known by mere casual reflection on one's own linguistic practice, and bear only a remote connection to the traditional schoolbook grammars. Indeed, their specification seems to require a very sophisticated technical apparatus, and what they allow and disallow cannot be accounted for in terms of any generic requirements on communication (see Syntax).

Furthermore, if we compare the linguistic data available to the child to the grammar of the specific language the child comes to master, we see that the data severely underdetermines the grammatical rules. There are many alternative rules that could claim as strong inductive support from the child's data as the rules children actually master. Nevertheless, all normal children who learn a language arrive at the same set of rules at a very young age, with relatively little effort, and with little (if any) relevant explicit instruction. Despite disagreement within linguistics about how to specify the particulars of this picture, the preponderant view of the evidence as it now stands is that such a nativist approach is unavoidable.

Philosophers committed to empiricism have responded to Chomsky's linguistic nativism in a variety of ways. Some claim that knowing a language is not a matter of knowing *that*, but rather knowing *how*: one can know how to walk without knowing the specific physical principles that govern walking. Empiricists can then freely grant that there is innate know-how without having to claim there are any innate ideas. Some, echoing Locke, have challenged innate 'Universal Grammar' because of the remoteness of such rules from consciousness (see Consciousness §7; Knowledge, tacit; Unconscious mental states).

Still others have claimed that the most that Chomsky-style nativist arguments show is that speakers of a language have various innate *beliefs*, but no *knowledge*; for, after all, the neonate's beliefs are not *justified* (see Knowledge, concept of). Finally, Hilary Putnam has argued that Chomsky's conclusions are premature because we still have a poor understanding of the general inductive principles that make learning from experience possible (see Inductive

inference). Until and unless we understand how induction works, we will be in no position to argue that it cannot account for the acquisition of a grammar. Notice, however, that at the very least, this argument concedes the force of the nativist evidence and admits that the long-dominant empiricist paradigm remains underdeveloped.

A more forceful but speculative nativist line of response would challenge the whole notion of general inductive principles. The idea would be that, for any domain of knowledge, we either have special-purpose innate structures that mediate its acquisition, or else we adapt one of the special-purpose structures already in place and refit it for a new domain (for example, applying innate combat strategies to chess). Such a hypothesis does posit some sort of general purpose 'refitting' function, but it clearly has a central nativist core.

3 Nativism in philosophy and empirical science

The success of the Chomskian paradigm in linguistics has strongly influenced research in cognitive development, and there are a number of empirical research programmes aimed at showing that, contrary to the empiricist and Piagetian paradigms, different domains of human knowledge rest on domain-specific innate bases (see Cognitive development; Cognition, infant; Mind, child's theory of; Piaget, J.).

For example, recent findings suggest that 5-month-olds already have some expectations about the results of simple arithmetic operations, manifesting surprise (measured in looking time) at scenes displaying one object where two were to be expected, as a result of either addition or subtraction of objects by hands disappearing behind screens (Wynn 1992). Similar work suggests that the infant's notion of what constitutes a single physical object and certain elementary physical principles may be innately understood (Kelman and Spelke 1983; Baillargeon 1987). There seem to be innate ideas about specific kinds of objects: newborn babies (even less than 10 minutes old!) will look longer at a schematic of a normal human face than at a face with scrambled features (see Morton and Johnson 1991).

Of course, showing that a concept is already present in a 6- or even a 4-month-old does not show that it is innate (although it is hard to extend such scepticism to an infant only 10 minutes old). But as the evidence mounts for earlier and earlier concepts and knowledge, the burden of argument shifts to the anti-nativist camps (which generally predict later onset).

An important philosophical development associated with the Chomskian revolution is Jerry Fodor's radical concept-nativism, which also has roots in Plato and Descartes. Fodor (1981) argues that no empiricist theory of concept learning is viable; all our concepts (to a first approximation) are therefore innate. He points out that empiricist theories implicitly assume that that there is a set of unlearned primitive concepts, typically sensory ones such as hard or red, and propose 'reducing' all other concepts to these (see Concepts §6). However, he argues, such 'reductionist' efforts have been notorious failures, as leading empiricists including Quine (1953) and Putnam (1965) have conceded. Moreover, recent experimental studies indicate that what the empiricist would take to be complex concepts are not represented in the mind as complex mental representations. So, Fodor concludes, these too must be primitive, and therefore innate. Needless to say, these arguments have stirred substantial controversy, both about the resources of reduction, and the interpretation of these latter experimental studies.

In place of the empiricist picture of concepts learned from experience, Fodor offers a triggering model that is familiar in cognitive ethology. Ethologists have discovered that for some bird species, the young bird has only to hear some song or other for it to begin singing its species-characteristic song (see Learning §2). This is plainly not a matter of learning from experience; experience is more of a release or trigger. Fodor's suggestion is that concepts - like some birds' song - are innate but dormant, waiting patiently for the right experience to trigger them and bring them into cognitive play. This would doubtless be music to Socrates' ears.

See also: Cognitive development; Innate knowledge; Mind, child's theory of

JERRY SAMET

References and further reading

Baillargeon, R. (1987) 'Object Permanence in 3- and 4-Month-Old Infants', *Developmental Psychology* 23: 655-64.(Study of infants' expectations about enduring physical objects.)

Carruthers, P. (1992) *Human Knowledge and Human Nature*, Oxford: Oxford University Press.(An interesting attempt to divorce nativism from classical rationalism and join it to classical empiricism.)

Chomsky, N. (1965) *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press.(Chapter 1 lays out a blueprint for linguistics that puts the question of language acquisition - and therefore of nativism - at the heart of the enterprise.)

Chomsky, N. (1988) *Language and Problems of Knowledge*, Cambridge, MA: MIT Press.(One of many recent expositions and defences of nativism that are accessible to the non-specialist - see especially chapters 1, 2, 5.)

Descartes, R. (1641) *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes*, trans. and ed. J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press, 1984-91, 3 vols.(Descartes' nativistic proof of the existence of God is in Meditation III.)

Edgley, R. (1970) 'Innate Ideas', in *Knowledge and Necessity*, Royal Institute of Philosophy Lectures, vol. 3, London: Macmillan.(Includes an important discussion of the interplay of nativism, empiricism and realism touched on at the end of §2.)

Fodor, J. (1981) 'The Present Status of the Innateness Controversy', in *Representations*, Cambridge, MA: MIT Press.(A tour-de-force defence of a very radical form of concept nativism.)

Hook, S. (ed.) (1969) *Language and Philosophy: A Symposium*, New York: New York University Press.(Includes a number of philosophical responses - mostly negative - to Chomsky's nativism and mentalism, along with Chomsky's replies. See especially papers by Chomsky, Quine, Schwartz, Nagel, Harman and Goodman.)

Kelman, P.J. and Spelke, E.S. (1983) 'Perception of Partially Occluded Objects in Infancy', *Cognitive Psychology* 15: 483-524.(Study of the conditions under which infants take what is given in a visual array to constitute a single object.)

Leibniz, G.W. (1704) *New Essays on Human Understanding*, trans. and ed. P. Remnant and J. Bennett, Cambridge: Cambridge University Press, 1981.(A paragraph-by-paragraph rationalist response to Locke's empiricist *Essay* (1689). The preface and first chapters are especially important.)

Locke, J. (1689) *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Oxford University Press, 1975.(Chapter 1 includes a blistering anti-nativist polemic. It has never been clear precisely whom Locke was attacking, but its effect was to make nativism a virtual dead letter for 250 years.)

Morton, J. and Johnson, M.H. (1991) 'CONSPEC and CONLEARN: A Two-Process Theory of Infant Face Recognition', *Psychological Review* 98: 164-81.(Evidence that newborn infants have a specialized system for face recognition.)

Piattelli-Palmarini, M. (ed.) (1980) *Language and Learning*, Cambridge, MA: Harvard University Press.(Papers from a conference bringing together Piaget and Chomsky and hoping to find common ground. The conference was not successful in its stated purpose, but Chomsky and Fodor set out the nativist position(s) and the book closes with a spirited debate between these two and Hilary Putnam.)

Pinker, S. (1994) *The Language Instinct*, New York: William Morrow.(An extremely readable account of the state of the art in our understanding of language; defends the nativist view.)

Putnam, H. (1965) 'The Analytic and the Synthetic', in *Collected Papers*, Cambridge: Cambridge University Press, vol. 2.(With Quine (1953), a classic discussion of the empiricist effort to reduce all concepts to sensory ones.)

Quine, W.V. (1953) 'Two Dogmas of Empiricism', in *From a Logical Point of View*, New York: Harper & Row.(The seminal paper attacking the empiricist conception of meaning. Although Quine does not react to the difficulty by embracing nativism - he defends empiricism in a different way - Fodor draws his conclusions in part from this attack.)

Samet, J. (forthcoming) *Nativism*, Cambridge, MA: MIT Press.(An extended study covering all of the material in this entry (and more) in depth.)

Scott, D. (1996) *Recollection and Experience: Plato's Theory of Learning and its Successors*, Cambridge: Cambridge University Press.(Discussion of the historical nativist-empiricist debate, with special attention to the ancient world.)

Starkey, P., Spelke, E.S. and Gelman, R. (1990) 'Numerical Abstraction by Human Infants', *Cognition* 36: 97-127.(Presents evidence that infants respond differentially to visual arrays with different numerosities.)

Stich, S. (ed.) (1975) *Innate Ideas*, Berkeley, CA: California University Press.(A very useful but out-of-print collection of historical texts and contemporary discussions of innateness, with an excellent introduction by Stich. A number of papers address the relevance of linguistics to the debate.)

Wynn, K. (1992) 'Addition and Subtraction by Human Infants', *Nature* 358: 749-50.(Presents evidence that infants have some sort of grasp of elementary arithmetical operations.)

Natural deduction, tableau and sequent systems

Different presentations of the principles of logic reflect different approaches to the subject itself. The three kinds of system discussed here treat as fundamental not logical truth, but consequence, the relation holding between the premises and conclusion of a valid argument. They are, however, inspired by different conceptions of this relation. Natural deduction rules are intended to formalize the way in which mathematicians actually reason in their proofs. Tableau systems reflect the semantic conception of consequence; their rules may be interpreted as the systematic search for a counterexample to an argument. Finally, sequent calculi were developed for the sake of their metamathematical properties.

All three systems employ rules rather than axioms. Each logical constant is governed by a pair of rules which do not involve the other constants and are, in some sense, inverse. Take the implication operator '$\rightarrow$', for example. In natural deduction, there is an introduction rule for '$\rightarrow$' which gives a sufficient condition for inferring an implication, and an elimination rule which gives the strongest conclusion that can be inferred from a premise having the form of an implication. Tableau systems contain a rule which gives a sufficient condition for an implication to be true, and another which gives a sufficient condition for it to be false. A sequent is an array $\Gamma \vdash \Delta$, where Γ and Δ are lists (or sets) of formulas. Sequent calculi have rules for introducing implication on the left of the '$\vdash$' symbol and on the right.

The construction of derivations or tableaus in these systems is often more concise and intuitive than in an axiomatic one, and versions of all three have found their way into introductory logic texts. Furthermore, every natural deduction or sequent derivation can be made more direct by transforming it into a 'normal form'. In the case of the sequent calculus, this result is known as the cut-elimination theorem. It has been applied extensively in metamathematics, most famously to obtain consistency proofs. The semantic inspiration for the rules of tableau construction suggests a very perspicuous proof of classical completeness, one which can also be adapted to the sequent calculus. The introduction and elimination rules of natural deduction are intuitionistically valid and have suggested an alternative semantics based on a conception of meaning as use. The idea is that the meaning of each logical constant is exhausted by its inferential behaviour and can therefore be characterized by its introduction and elimination rules.

Although the discussion below focuses on intuitionistic and classical first-order logic, various other logics have also been formulated as sequent, natural deduction and even tableau systems: modal logics, for example, relevance logic, infinitary and higher-order logics. There is a gain in understanding the role of the logical constants which comes from formulating introduction and elimination (or left and right) rules for them. Some authors have even suggested that one must be able to do so for an operator to count as logical.

1 Natural deduction

Systems of natural deduction were first devised in the late 1920s by Gerhard Gentzen (1935), a student of Hilbert's, and, independently, by Stanislaw Jaśkowski. Their derivations formalize proofs from assumptions and are most conveniently represented as tree-like structures with the conclusion at the root and an assumption at the top of each branch. Assumptions may be discharged by the application of a rule in the course of the derivation, and the conclusion will depend only on the undischarged assumptions. Here is one formulation of the introduction and elimination rules for the usual logical operators:

(1) $\dfrac{A \quad B}{A \wedge B}$ (2) (a) $\dfrac{A \wedge B}{A}$

(b) $\dfrac{A \wedge B}{B}$

(3) (a) $\dfrac{A}{A \vee B}$ (4) $\dfrac{A \vee B \quad \begin{matrix}[A] \\ C\end{matrix} \quad \begin{matrix}[B] \\ C\end{matrix}}{C}$

(b) $\dfrac{B}{A \vee B}$

(5) $\dfrac{\begin{matrix}[A] \\ B\end{matrix}}{A \rightarrow B}$ (6) $\dfrac{A \quad A \rightarrow B}{B}$

(7) $\dfrac{\begin{matrix}[A] \\ \bot\end{matrix}}{\neg A}$ (8) $\dfrac{A \quad \neg A}{\bot}$

(9) $\dfrac{A(a)}{\forall x A(x)}$ (10) $\dfrac{\forall x A(x)}{A(a)}$

(11) $\dfrac{A(a)}{\exists x A(x)}$ (12) $\dfrac{\exists x A(x) \quad \begin{matrix}[A(a)] \\ C\end{matrix}}{C}$

Rule (3a), for example, adds a new root to the tree, transforming a derivation of *A* from unspecified assumptions into a derivation of $A \vee B$ from those same assumptions. Introductions are written on the left, eliminations on the right. ‘ $\bot$’ is a constant for falsity. A formula enclosed in square brackets indicates that assumptions of this form in the derivation of the premise displayed below it may be discharged by the inference in question.

In rules (9) and (12), *a* cannot occur in $A(x)$ and is called the ‘proper parameter’ of the inference; it must satisfy the following restrictions. In (9), *a* cannot occur in any assumption on which the premise depends; in (12), *a* cannot appear in *C* nor in any assumption upon which the premise *C* depends except for those of the form $A(a)$. Because they allow the discharge of assumptions, it is better to view applications of the rules $\rightarrow$-introduction and $\vee$- and $\exists$-elimination as operating on whole derivations rather than simply on their conclusions. A similar remark can be made about $\forall$-introduction because it requires that the derivation of $A(a)$ in its premise satisfy an extra condition. This system of introductions and eliminations constitutes minimal logic. A system equivalent to Heyting’s axiomatization of intuitionistic logic is obtained by adding a rule which allows any formula to be inferred from a falsehood.

$\dfrac{\bot}{A}$

Classical logic can be obtained by adding to the intuitionistic system as axioms all instances $A \vee \neg A$ of the law of the excluded middle, or a rule such as the following.

$\dfrac{\neg\neg A}{A}$

The major premise of an elimination is the formula whose principal operator is displayed in the statement of the rule. A ‘maximum formula occurrence’ in a derivation is both the conclusion of an introduction and the major premise of an elimination. It constitutes a needless detour and can be removed. This can be illustrated in the case of implication by the following pair of derivations (where Π_1 and Π_2 are derivations of *A* and *B*, respectively):

$$\dfrac{\dfrac{\Pi_1}{A} \quad \dfrac{\begin{matrix}[A] \\ \Pi_2 \\ B\end{matrix}}{A \rightarrow B}}{B}$$

The one on the right is obtained from Π_2 by replacing each occurrence of the assumption *A* discharged by the

application of $\rightarrow$-introduction by a copy of the derivation Π_1 of *A*. If no assumption is discharged, the derivation on the left is just replaced by Π_2. In either case, the result is a more direct derivation of the same conclusion from the same or fewer assumptions. A transformation of this kind is called a 'reduction' step, and similar reductions can be defined for the other operators. The special form of the elimination rules for '$\vee$' and '$\exists$' give rise to what are called 'maximum segments'. These consist of consecutive occurrences of the same formula such that the first is the conclusion of an introduction, the last is the major premise of an elimination and all except the first are conclusions of an application of $\vee$- or $\exists$-elimination. To remove such segments it is necessary to permute applications of $\vee$- and $\exists$-elimination with other inferences.

A derivation is said to be 'normal' if it contains no maximum formula occurrence or segment. Gentzen's normal form theorem states that for every derivation in the intuitionistic system, there is a normal one of the same conclusion from the same or fewer assumptions. It is proved by prescribing a method for systematically removing all detours from a derivation, and thus a stronger result is established - sometimes called the 'normalization theorem' - namely, that every derivation Π can be *reduced* to one in normal form. The reduction steps can be so chosen that every sufficiently long sequence of them beginning with Π terminates in a unique normal derivation. This is called the strong normalization theorem. Roughly speaking, a normal derivation is constructed by first breaking up its assumptions using eliminations and then building up its conclusion using introductions. That is why every formula appearing in the derivation will be a subformula of an assumption or the conclusion. All three forms of the theorem have been exploited to yield various proof-theoretic results for formalized mathematical theories. In addition, the significance of the normalization procedure itself has been studied in a more general context. It has been argued that the steps required to reduce a derivation to normal form do not change its meaning. In consequence, derivations which reduce to the same normal form have the same meaning or express the same proof. More controversially, the converse has also been claimed to hold.

2 The sequent calculus

The theorems discussed at the end of the previous section do not extend straightforwardly from the intuitionistic to the classical version of natural deduction. By exploiting the definability of disjunction and existential quantification in classical logic, however, they can be shown to hold for a system of natural deduction adequate for classical logic. A rather more sophisticated argument is required to establish them for a classical natural deduction calculus in which all the usual logical constants are taken as primitive. Gentzen, however, did not extend his theorem to the classical system described above. He chose rather to reformulate it as a calculus of sequents. It is a trivial matter to rewrite the introductions and eliminations as sequent rules by placing the assumptions on which a formula depends on the left-hand side of the sequent. The rules for '$\rightarrow$', for example, will look as follows.

$$\frac{A,\Gamma\vdash B}{\Gamma\vdash A\rightarrow B} \qquad\qquad \frac{\Gamma\vdash A\rightarrow B \quad \Delta\vdash A}{\Gamma,\Delta\vdash B}$$

Although assumptions have been eliminated in favour of axioms of the form $A \vdash A$, nothing else of significance has been changed. In fact, the resulting calculus is just a notational variant of natural deduction. Gentzen's idea was to rewrite only the introduction rules in this way, and to replace the eliminations by new rules for introducing a formula on the left side of a sequent. These left rules are displayed below. (*a* cannot occur anywhere in the conclusion of the last rule.)

$$\frac{A, \Gamma \vdash C}{A \wedge B, \Gamma \vdash C} \qquad \frac{B, \Gamma \vdash C}{A \wedge B, \Gamma \vdash C}$$

$$\frac{A, \Gamma \vdash C \quad B, \Gamma \vdash C}{A \vee B, \Gamma \vdash C}$$

$$\frac{\Gamma \vdash A \quad B, \Delta \vdash C}{A \rightarrow B, \Gamma, \Delta \vdash C}$$

$$\frac{\Gamma \vdash A}{\neg A, \Gamma \vdash}$$

$$\frac{A(a), \Gamma \vdash C}{\forall x A(x), \Gamma \vdash C}$$

$$\frac{A(a), \Gamma \vdash C}{\exists x A(x), \Gamma \vdash C}$$

The rules may be read in different ways. Their connection with natural deduction is best brought out by interpreting the symbol ‘ $\vdash$’ as derivability. Then the last of them, for example, asserts that a derivation of *C* from $A(a), \Gamma$ can be transformed into one of *C* from $\exists x A(x), \Gamma$ (subject to the restriction on *a*).

In addition, Gentzen introduced some structural rules: ‘interchange’ for permuting formulas on a list, ‘contraction’ for amalgamating two occurrences of the same formula, and ‘thinning’ or ‘weakening’ for adding formulas to a list. Because an empty list on the right of a sequent can be interpreted as falsity, there is no longer any need for the constant ‘ $\bot$’. Furthermore, the intuitionistic negation rule can now be derived because thinning on the right justifies the inference from $\Gamma \vdash$ to $\Gamma \vdash A$. In the presence of the structural rules, it does not matter whether the left rule for ‘ $\vee$’ is formulated as above or as follows.

$$\frac{A, \Gamma \vdash C \qquad B, \Delta \vdash C}{A \vee B, \Gamma, \Delta \vdash C}$$

And likewise for the other two premise rules, both left and right. Recently J.-Y. Girard, in his work on linear logic (see Girard 1995), has embarked upon a searching investigation of the proof-theoretic significance of the structural rules, and of the consequences of omitting or weakening them in various ways. In their absence, he distinguishes two versions of disjunction, for example: an additive and a multiplicative one according to which formulation of its left rule is adopted.

3 Comparing sequents and natural deduction

It is easy to translate a derivation of the sequent $\Gamma \vdash A$ into a derivation in natural deduction of *A* from assumptions included in the list Γ - into a normal one, in fact. Call this derivation its $\mathcal{N}$ translation. Applications of right rules translate directly into introductions. The situation with left rules is only slightly more complicated. Consider conjunction, for example, and suppose that Π is the $\mathcal{N}$ translation of a derivation of $A, \Gamma \vdash C$, then the required derivation of *C* from assumptions in $A \wedge B, \Gamma$ is obtained by substituting copies of the two-line derivation

$$\frac{A \wedge B}{A (\dagger)}$$

for the appropriate occurrences of the assumption *A* in Π.

Converting natural deduction derivations into sequent ones is less straightforward. Introductions correspond as before to applications of a right rule. To translate eliminations, however, we need not only the corresponding left rule, but also a rule called ‘cut’.

$$\frac{\Gamma \vdash A \qquad A, \Delta \vdash C}{\Gamma, \Delta \vdash C}$$

With the addition of this rule every sequent derivation will still have an $\mathcal{N}$ translation, although we can no longer be sure that it will be normal. The $\mathcal{N}$ translation of a derivation which ends with the cut displayed above is obtained by substituting copies of the $\mathcal{N}$ translation of the derivation of its left premise for the appropriate occurrences of the assumption *A* in the $\mathcal{N}$ translation of the derivation of its right one.

To see why cut is needed for translating the elimination rules, consider again the case of conjunction and suppose that we have already succeeded in translating a derivation Π of $A \wedge B$ from assumptions Γ into a derivation δ of the sequent $\Gamma \vdash A \wedge B$. Then what needs to be shown is that δ can be converted into a derivation of $\Gamma \vdash A$. This is accomplished by a cut applied to the conclusions of δ and a derivation whose $\mathcal{N}$ translation is $(\dagger)$ above, as shown in the following figure.

$$\dfrac{\begin{matrix}\delta\\ \Gamma \vdash A \wedge B\end{matrix} \qquad \dfrac{A \vdash A}{A \wedge B \vdash A}}{\Gamma \vdash A}$$

This translation exploits the fact that the result of applying $\wedge$-elimination to Π is the same as substituting Π for the assumption $A \wedge B$ in the derivation $(\dagger)$ above. The other elimination rules are handled similarly.

Clearly, every formula appearing in the premise(s) of a left or right rule is a subformula of a formula appearing in its conclusion, and the same holds for the structural rules. Consequently, any derivation constructed only by the use of these rules will be such that it contains only subformulas of formulas occurring in its final sequent. Gentzen's normal form theorem in this context takes the form of a cut-elimination theorem, and he was able to prove that every derivation can be transformed into a cut-free derivation of the same sequent. His proof, although similar in kind to the one for natural deduction, does involve additional complications. Furthermore, cut-free forms are not in general unique. It would appear therefore that little advantage has been gained by passing from natural deduction to a sequent calculus, especially since sequent derivations are for the most part less intuitive.

In favour of sequent derivations is the fact that the absence of assumptions makes them easier to work with from a metamathematical standpoint. To verify the correctness of a sequent derivation it is enough to check it locally, that is, to look at each sequent and its immediate successors individually. There are no global conditions on the derivation that need to be satisfied (as, for example, that a certain parameter cannot appear in any assumption). This also makes coding derivations by means of Gödel numbers much easier. A more significant advantage, however, concerns the relationship between the calculus described above and classical logic. Gentzen observed that, by the simple expedient of allowing a list containing more than one formula to appear on the right-hand side of a sequent, he could obtain a system adequate for classical logic. The classical versions of the right and left rules for ' $\rightarrow$', for example, read as follows.

$$\frac{A, \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \rightarrow B} \qquad\qquad \frac{\Gamma \vdash \Delta, A \qquad B, \Psi \vdash \Phi}{A \rightarrow B, \Gamma, \Psi \vdash \Delta, \Phi}$$

All the remaining rules, including cut, are similarly generalized; no new rules need to be added. To see that the resulting calculus is adequate for classical logic, note that the sequent $\vdash A \vee \neg A$ is now easily derivable. Applying $\neg$-right to $A \vdash A$ yields $\vdash A, \neg A$, from which the desired sequent follows by two applications of $\vee$-right and a couple of structural inferences. It is possible therefore to give a single proof of the cut-elimination theorem for both classical and intuitionistic systems. The proof for the classical system involves no new ideas, and it specializes immediately to a proof for the intuitionistic one. One need simply observe that each reduction prescribed in the proof yields an intuitionistic derivation when applied to one.

Gentzen's theorem has been fruitfully applied to yield a variety of results in proof theory. The Craig interpolation theorem, for example, follows from it easily, as do various decidability results. Gentzen himself utilized the method of cut elimination to prove the consistency of formal number theory, that is, of a system obtained by augmenting the sequent calculus with appropriate arithmetical principles. (In fact, his original proof was for a theory based on natural deduction, but he subsequently rewrote it for the sequent calculus.) His methods have since been adapted and extended, in many different ways and to a variety of different situations. A version of the cut-elimination theorem for second-order logic, for example, has yielded a consistency proof for analysis, formalized as a theory of second-order arithmetic (see Proof theory §4). The information provided by results such as these depends upon the methods required to establish them, and their significance is understood differently in different foundational traditions. Nevertheless, there is no doubt that the methods pioneered by Gentzen have been one of the mainstays of proof theory in the tradition of Hilbert, as it has developed in the wake of Gödel's theorems.

Although cut-free derivations are in some sense simpler and more direct than their counterparts containing cuts, the elimination procedure described by Gentzen increases the length of a derivation each time a cut is removed

from it. Furthermore, this turns out to be an essential feature of any cut-elimination procedure. The same holds true for the elimination of maximum formula occurrences from natural deduction derivations and, as one might expect, the efficiency of the tableau method described below is correspondingly increased when an analogue to the cut rule is added to the system. This simple fact has given rise to a great deal of research on the complexity of derivations, and very refined bounds on the length of cut-free ones have been established for a variety of systems and theories. The subject of complexity has attracted particular attention from researchers interested in implementing automated reasoning systems (see Complexity, computational).

4 Tableau systems

The discussion of the previous sections was based on an interpretation of the sequent calculus as characterizing the relation of derivability under assumptions in natural deduction. With the passage from intuitionistic to classical sequents, this interpretation is no longer possible. The classical rules are easily seen to be correct under a semantic interpretation, however, according to which $\Gamma \vdash \Delta$ is taken to mean that for every model of Γ there is at least one formula in Δ that is true in it. If the rules are read backwards, from conclusion to premise(s), they represent steps in the search for a counterexample to $\Gamma \vdash \Delta$, a model in which all the formulas in Γ are true, while all those in Δ are false. To find a counterexample to $\Gamma \vdash \Delta, A \wedge B$, for example, we may search in one of two directions, either for a counterexample to $\Gamma \vdash \Delta, A$ or for one to $\Gamma \vdash \Delta, B$. The first person to exploit this feature of the rules seems to have been E.W. Beth. It underlies the tableau system for classical logic which he formulated in the mid-fifties (see his later work of 1964 for details). However, the general idea of a no-counterexample treatment of logic is also to be found in the works of other logicians of the period, including J. Hintikka, S. Kanger and K. Schütte.

A semantic tableau for the sequent $\Gamma \vdash \Delta$ begins as a figure consisting of two columns, with the formulas in Γ listed at the head of the left one and the formulas in Δ at the head of the right. The rules extend the tableau by operating on individual formulas in these lists. A left rule operates on *A* in Γ and allows us to enter into the tableau subformulas of *A* in such a way that the truth of the left entries together with the falsity of the right ones together constitute a sufficient condition for *A* to be true. Similarly, a right rule operates on *A* in Δ and allows us to enter subformulas whose truth and falsity are sufficient for *A* to be false. As before, there are pairs of rules for each of the logical constants, and some of these rules require the tableau to split into a pair of subtableaus, each with its own left and right column. Since all the classical connectives and quantifiers are definable in terms of ‘ $\neg$’, ‘ $\rightarrow$’ and ‘ $\exists$’, it will be sufficient to list tableau rules for just these three operators:

If $\neg A$ appears in the left column, write A in the right one.
If $\neg A$ appears in the right column, write A in the left one.
If $A \rightarrow B$ appears in the left column, split the tableau into two subtableaus, write B in the left column of one subtableau and A in the right column of the other.
If $A \rightarrow B$ appears in the right column, write B below it and A in the left column.
If $\exists x A(x)$ appears in the left column, write $A(a)$ in the left column, where a is a new parameter.
If $\exists x A(x)$ appears in the right column, write $A(a)$ in the right column, where a is any parameter that has already been introduced into the tableau (or a new one if none has been introduced).

The left rule for ‘ $\rightarrow$’, for example, reflects the fact that in classical logic there are two ways to verify an implication: its antecedent may be false or its consequent true. The right rule reflects the fact that there is only one way to refute one: its antecedent must be true and its consequent false. In all cases except the last, the formula to which the rule has been applied is cancelled to indicate that the rule cannot be applied to it again. In the case of right ‘ $\exists$’, the possibility of applying the rule with respect to a different parameter is left open. A subtableau is closed if there is at least one formula which occurs in both its left and right column. A tableau is closed if all its subtableaus are closed. $\Gamma \vdash \Delta$ is derivable if it has a closed tableau. Finally, a tableau has been terminated if it is not closed, but no further rules are applicable to any of its subtableaus.

By way of example, the following closed tableau establishes the validity of the sequent

$\neg \exists x \neg (A(x) \rightarrow B) \vdash \exists x A(x) \rightarrow B.$

$\neg\exists x\neg(A(x) \rightarrow B)$		$\exists x A(x) \rightarrow B$	
$\exists x A(x)$		B	
$A(a)$			
		$\exists x\neg(A(x) \rightarrow B)$	
		$\neg(A(a) \rightarrow B)$	
$A(a) \rightarrow B$			
1	2 B	1 $A(a)$	2

Each line is obtained from a previous one by applying one of the rules for tableau construction. Notice that the initial tableau splits into two subtableaus after line six. (The first and third cells of line seven are the left and right columns of one, the second and fourth of the other.) Both subtableaus are closed, the one because $A(a)$ appears on the left in line three and on the right in line seven, and the other because *B* appears on the right in line two and on the left in line seven.

The rules for tableau construction are really nothing other than upside-down versions of the classical sequent rules. Indeed, one need do little more than invert a closed tableau to convert it into a cut-free derivation in the sequent calculus. There is no need to augment the tableau rules by a rule analogous to cut because their deductive adequacy can be demonstrated directly without reference to the derivations of any other system. The rules may be applied to a sequent in any order (although it is convenient to postpone applying right '$\exists$' for as long as possible) and three possibilities can arise:

(1) The tableau closes after a finite number of steps, in which case the disjunction of the formulas in Δ is a classical consequence of the formulas in Γ.
(2) The tableau is terminated after a finite number of steps. This will be because every uncancelled formula (other than existential ones to which right '$\exists$' is no longer applicable) is atomic. It is then easy to construct a model in which each atomic formula appearing on the left of an open subtableau is true and each one appearing on the right is false. Such a model turns out to be a counterexample to the original sequent in the sense that all the formulas in Γ are true in it while those in Δ are false.
(3) The tableau is never terminated. This possibility arises because right '$\exists$' may be applied repeatedly to a formula as needed. Nevertheless, as Beth showed, we can prove generally that a counterexample can also be constructed from any such infinite tableau.

The upshot is that the rules of tableau construction provide a complete proof procedure for classical logic. Despite their semantic motivation, it should be emphasized that these are purely syntactic rules - and hence that there is a cut-free derivation of every derivable sequent in the sequent calculus as well. This semantic approach to the cut-elimination theorem is less elementary than Gentzen's original proof. More significantly, however, the connection between a derivation and its cut-free form is lost.

Because tableaus may split into subtableaus, the method is cumbersome to use. A considerable simplification was effected by Raymond Smullyan, who introduced the idea of 'signed' formulas. Given $\Gamma \vdash \Delta$, the formulas in Γ receive the sign T and those in Δ the sign F. A tableau now begins with a single list of signed formulas, and left and right rules are reformulated as rules for formulas with the signs T and F respectively. The result is a single tree, rather than a series of tableaus, which branches instead of splitting. A closed branch will contain an occurrence of some formula *A* with the sign T and another occurrence with the sign F, and a tree is closed when every branch is closed. (Alternatively, one can dispense with signs and form a single list from the formulas in Γ together with the negations of those in Δ.) It is in this form, rather than the original one, that the tableau method has found its way into the literature. An analogous modification of the sequent calculus was introduced by Kurt Schütte. Here the effect is to make derivations less intuitive, although the calculus is more concise and better suited for some applications.

Just as natural deduction seems to be a more appropriate system for intuitionistic logic, so tableaus seem better adapted to express classical consequence. Nevertheless, as Beth showed, they can be modified to yield a proof procedure for intuitionistic logic. Subsequently, tableau techniques were extended to modal systems, by Kripke among others, and what are now called Kripke models have their origin in constructions deriving from these.

See also: Logical and mathematical terms, glossary of

A.M. UNGAR

References and further reading

Beth, E.W. (1964) *The Foundations of Mathematics*, Amsterdam: North Holland; New York: Harper & Row, 2nd edn, 1966.(§§67-70, 144-5 include accounts of tableaus for classical and intuitionistic logic, respectively; one of the few places where tableaus are presented in their original two-column form.)

Boolos, G. (1984) 'Don't Eliminate Cut', *Journal of Philosophical Logic* 13: 373-8.(Includes a striking illustration of just how much shorter than its cut-free counterpart a derivation with cut can be.)

Došen, K. (1993) 'A Historical Introduction to Substructural Logics', in K. Došen and P. Schroeder-Heister (eds) *Substructural Logics*, Oxford: Clarendon Press, 1-30.(A brief and readable survey of the motivations for and effects of restricting structural rules in sequent calculi, with a useful bibliography.)

Fitting, M. (1983) *Proof Methods for Modal and Intuitionistic Logics*, Dordrecht: Reidel.(A thorough and self-contained presentation of natural deduction, sequent and tableau methods for a variety of logics, including classical.)

Gentzen, G. (1935) 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* 39: 176-210, 405-565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969; trans. in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 68-131. (Gentzen's original, and very readable, account of sequent and natural deduction systems for classical and intuitionistic logic.)

Girard, J.-Y. (1995) 'Linear Logic: Its Syntax and Semantics', in J.-Y. Girard, Y. Lafont and L. Regnier (eds) *Advances in Linear Logic*, Cambridge: Cambridge University Press, 1-42.(An introductory, albeit challenging, account of linear logic by its originator.)

Kleene, S.C. (1967) *Mathematical Logic*, New York: Wiley.(Chapter 4 describes a very direct proof of the completeness of the classical cut-free sequent rules.)

Prawitz, D. (1965) *Natural Deduction: A Proof-Theoretical Study*, Uppsala: Almqvist & Wiksell.(A thorough but elementary presentation of natural deduction and the normalization theorem.)

Prawitz, D. (1971) 'Ideas and Results in Proof Theory', in J.E. Fenstad (ed.) *Proceedings of the Second Scandinavian Logic Symposium*, Amsterdam: North Holland, 235-307.(A survey of some proof-theoretic applications of natural deduction.)

Schwichtenberg, H. (1977) 'Proof Theory: Some Applications of Cut-Elimination', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 867-95.(A survey of some proof-theoretic applications of the sequent calculus; requires more background than the previous item.)

Stålmarck, G. (1991) 'Normalization Theorems for Full First Order Classical Natural Deduction', *Journal of Symbolic Logic* 56 (1): 129-49.(This article is rather technical.)

Sundholm, G. (1983) 'Systems of Deduction', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 1, 133-88.(A more detailed survey of the systems discussed in this article, with a useful bibliography; recommended as the next thing to read.)

Sundholm, G. (1986) 'Proof Theory and Meaning', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 3, 471-506.(Another well-written survey article; explains how natural deduction and related systems may contribute to a theory of meaning which does not depend on the notion of truth.)

Natural kinds

Objects belonging to a natural kind form a group of objects which have some theoretically important property in common. For example, rabbits form a natural kind, all samples of gold form another, and so on. Natural kinds are contrasted with arbitrary groups of objects such as the contents of dustbins, or collections of jewels. The latter have no theoretically important property in common: they have no unifying feature. Natural kinds provide a system for classifying objects. Scientists can then use this system to predict and explain the behaviour of those objects. For these reasons, the topic of natural kinds is of special interest to metaphysics and to the philosophy of science.

1 The notion of natural kinds

Objects belonging to a natural kind form a group of objects which have some theoretically important property, or properties, in common. Standard examples of natural kinds include biological species such as rabbits, oaks and whales, chemical elements and compounds such as oxygen, carbon and aluminium, and stuffs such as salt, wool and heat. An object's membership of a given natural kind bears upon how that object will behave, the properties it can or cannot acquire, and the types of interaction in which it can figure. Natural kinds thereby provide a scheme of classification for science to chart (see Taxonomy). Scientists can use such a system to predict, explain and (perhaps) control the behaviour of those objects. Knowledge of natural kinds can also provide a basis for inductive and counterfactual scientific reasoning: knowledge about how past members of a natural kind behaved may warrant beliefs about how present or future members of that kind will behave, as well as how hypothetical members of that kind would behave.

Closely related is the question regarding which predicates are 'projectible': the question of which predicates are admissible in inductive reasoning. Predicates such as 'is a rabbit' and 'is an oak' are admissible, but not every predicate is. The predicate 'is a roak', which is hereby defined as being true of an object if and only if it is a rabbit or an oak, seems inadmissible. The notion of a natural kind may help in explicating this distinction between projectible and non-projectible predicates (see Goodman, N. §3).

Again, a theory's terms are apparently open to deviant interpretations. A theory containing (say) the predicate 'is a rabbit' is standardly interpreted as meaning rabbit, but this predicate can be deviantly interpreted as meaning (say) rabbit or oak. To meet this problem, we might invoke the principle that among meaning-assignments we should prefer those in which the assigned meaning concerns natural kinds, rather than non-natural ones.

2 Two major questions

The topic of natural kinds raises two major related questions. There is a metaphysical question: what is a natural kind? Natural kinds are standardly distinguished from arbitrary groups of objects, such as what you had for breakfast. They are also standardly distinguished from so-called artefact kinds such as books, tables or cars. A full account of the metaphysics of natural kinds should clarify these distinctions and explain the rationale behind them. Furthermore, it should explain the sense in which different members of a natural kind have a property in common. On one possible account, different members of a natural kind share a universal: they have a property which is literally identical between them (see Universals). But on a rival account, different members of a kind have a property in common only in the sense that each member bears a certain similarity relation to every other member of the kind.

There is also an epistemological question: how do we tell what natural kinds there are? Every object has many properties, yet we think that only some properties determine natural kinds. But what reasons guide our thinking that certain properties rather than others determine natural kinds? Furthermore, our beliefs about what natural kinds there are, and which objects belong to them, may be inaccurate to some degree. These beliefs may only approximately describe what natural kinds there are, and what their members are. But then how are we to establish how accurate these beliefs are?

Such metaphysical and epistemological questions are also of interest outside philosophy. For instance, it is a live issue among biologists whether species are natural kinds, and there is a further issue of whether it is morphological, reproductive or phylogenetic factors which determine a creature's membership of a given species

(see Species).

3 Essentialism

Locke (1689) and, following him, Kripke (1972) and Putnam (1975), each offer an account of how our understanding of a given natural kind can develop over time. At an initial stage, speakers classify certain objects as members of the same natural kind because they perceive those objects to have certain properties in common. Such common properties form the ‘nominal essence’ (Locke) or ‘stereotype’ (Putnam) associated with that kind. The speakers, however, further believe that these objects are members of the same natural kind if and only if they have some underlying, though perhaps currently unknown, common property. This is the ‘real essence’ of that kind. In the case of chemical elements, it will perhaps be some microstructural property. In the case of organisms, it will perhaps be some genetic structure. Moreover, given how laws of nature relate a kind’s real essence to its nominal essence, an object’s real essence explains why that object has the nominal essence that it does. It then falls to empirical science to discover the real essences of objects.

In sum, an object’s nominal essence serves as a heuristic, a guide for thinking whether that object is a member of a certain natural kind. But whether in fact that object is a member of the kind in question is determined by whether it has the appropriate real essence. For instance, various objects may once have been classified as members of the natural kind lemon because of their yellow appearance, tart taste, thick skin and oval shape. These properties form the kind’s nominal essence. However, scientific inquiry may subsequently reveal the real essence of lemons - a common genetic structure, let us suppose. Since something is a lemon if and only if it has this genetic structure, we may revise to some degree our original classification of which objects are lemons. On this view, fruit which are superficially similar to lemons but which have a different genetic structure, are not lemons. In contrast, unripe, mouldy, or squashed fruit with the same genetic structure as lemons are lemons, even though superficially they do not appear to be so.

Kripke and Putnam further argue that it is (metaphysically) necessary that something is a member of a given kind if and only if it has a certain real essence. Hence if the real essence of, for example, gold is a certain atomic number, then, necessarily, something is gold if and only if it has that atomic number. Kripke and Putnam conclude that there are certain *de re* necessities which can be discovered only empirically (see De re/de dicto; Essentialism). Other philosophers have since argued that essentialism about natural kinds entails that the laws of nature which govern kinds are also *de re* necessary.

4 Conventionalism

Kripke and Putnam assume that what natural kinds there are, and which objects are members of those kinds, are matters which obtain independently of our attempts at classification. Conventionalism about natural kinds denies this assumption. It says that our classification of objects into natural kinds does not reflect any pre-existing divisions in nature. Instead, it is based on human convenience, and its goal is the prediction and control of observable events, not the discovery of underlying essences.

Locke ultimately took this view. Frequently, he is interpreted as taking it for an epistemological reason, namely that real essences, if there are any, would be unobservable and so unknowable. More recently, Bas van Fraassen (1980) has taken a similar view, but for a different reason, believing that science need be concerned only with observable entities, and so not with real essences. Hence the latter are redundant: they are ‘metaphysical baggage’. The dispute between realists and conventionalists about natural kinds is then located within the more general dispute between scientific realism (which claims that science describes an unobservable world) and constructive empiricism (which is agnostic about whether science does this). A further reason for scepticism about essentialism about natural kinds, and one which Locke and van Fraassen share, is a suspicion of what *de re* or metaphysical necessity is supposed to be. The suspicion is that the thesis that there are necessities in nature is obscure, and that instead the only source of necessity lies in human conventions. Accordingly, the thesis that objects have real essences in virtue of which they are members of certain natural kinds is held to be infected with this obscurity.

See also: Mass terms; Reference §§3-4

CHRIS DALY

References and further reading

Ayers, M. (1981) ‘Locke versus Aristotle on Natural Kinds’, *Journal of Philosophy* 78 (5): 247-72.(Careful and thorough discussion of Locke’s views on the topic.)

Fraassen, B. van (1980) *The Scientific Image*, Oxford: Clarendon Press.(Chapters 1-3 present van Fraassen’s brand of empiricism.)

Goodman, N. (1954) *Fact, Fiction, and Forecast*, Cambridge, MA: Harvard University Press.(See chapters 2 and 3 for the problem of projectibility and one possible solution.)

Kornblith, H. (1992) *Inductive Inference and its Natural Ground*, Cambridge, MA: MIT Press.(Part I presents a response to Locke’s views. Part II is an extended discussion of the epistemological issues concerning natural kinds.)

Kripke, S. (1972) ‘Naming and Necessity’, in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel, 253-355; repr. Cambridge, MA: Harvard University Press and Oxford University Press, 1980.(Kripke’s views on natural kinds and natural-kind terms are developed from his more general views about how names refer. Lecture III presents his views on natural kinds, but begins with a helpful summary of the conclusions reached in the preceding lectures.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, Oxford: Oxford University Press, 1975.(Locke’s views on natural kinds are found particularly in book 2 chapter 27, and in book 3 chapters 3 and 6.)

Putnam, H. (1975) *Mind, Language, and Reality: Philosophical Papers volume II*, Cambridge: Cambridge University Press. (Chapters 8, and 10-12 are relevant and best read in that order.)

Riggs, P. (ed.) (1996) *Natural Kinds, Laws of Nature, and Scientific Methodology*, Dordrecht: Kluwer Academic Press.(Collection of papers including sophisticated defences of essentialism about both natural kinds and laws of nature.)

Schwartz, S.P. (ed.) (1977) *Naming, Necessity, and Natural Kinds*, London and Ithaca, NY: Cornell University Press.(Collection which reprints important papers by Kripke, Putnam, and others. It also contains a helpful introduction and a comprehensive bibliography.)

Wilkerson, T.E. (1995) *Natural Kinds*, Aldershot: Avebury Press.(Lucid defence of essentialism about both natural kinds and laws of nature.)

Wolfram, S. (1989) *Philosophical Logic: An Introduction*, London and New York: Routledge.(Chapter 7 §1 is extremely clear and a good place to start. The end of the chapter lists further reading.)

Natural law

When made within the discourse of ethics, political theory, or legal theory or philosophy of law, the claim that there is a natural law is an offer to explain and defend certain claims often made, in different terms, in the discourse of moral argument, politics or law. In pre-theoretical moral discourse, certain choices, actions or dispositions may be asserted to be 'inhuman', 'unnaturally cruel', 'perverse' or 'morally unreasonable'. In pre-theoretical political discourse, certain proposals, policies or conduct may be described as violations of 'human rights'. In international law and jurisprudence, certain actions may be described as 'crimes against humanity' and citizens may claim immunity from legal liability or obligations by appealing to a 'higher law'. A natural law theory offers to explain why claims of this sort can be rationally warranted and true. It offers to do so by locating such claims in the context of a general theory of good and evil in human life so far as human life is shaped by deliberation and choice. Such a general theory can also be called a general theory of right and wrong in human choices and actions. It will contain both (1) normative propositions identifying types of choice, action or disposition as right or wrong, permissible, obligatory and so on, and (2) non-normative propositions about the objectivity and epistemological warrant of the normative propositions.

1 Why called natural? Why called law?

Theorists who describe their theory of good and evil, right and wrong, as a 'natural law theory' are not committed to asserting that the normative propositions which they defend are 'derived from nature' or 'read off' or 'inspected in the nature of things'. Still less are natural law theorists committed to claiming that the normative propositions they defend stand in some definite relationship to, or are warranted by, the 'laws of nature' in the sense of the regularities observed, and explanatory factors adduced, by the 'natural sciences'. Nor is the typical natural law theory (classical, medieval or contemporary) concerned with any alleged 'state of nature', in the sense of some golden age or state of affairs prior to human wrongdoing or to the formation of human societies or of states or political communities (see Laws, natural).

As for the term 'law', as understood in the phrase 'natural law', it does not connote that the relevant principles and norms have their directive force precisely as the imperatives or dictates of a superior will. Even when a natural law theorist argues (as most do) that the ultimate explanation of those principles and norms (as of all other realities) is a transcendent, creative, divine source of existence, meaning and value, the theorist usually will maintain also that moral precepts divinely commanded are commanded because fitting and obligatory, not fitting or obligatory because commanded (or that the source of their obligation is rather divine wisdom than divine will). Rather, the term 'law' in the phrase 'natural law' refers to standards of right choosing, standards which are normative, that is rationally directive and 'obligatory', because they are true and choice otherwise than in accordance with them is unreasonable.

The term 'natural' in this context signifies any one or more of the following: that the relevant standards are not 'positive', that is, are directive prior to any positing by individual decision or group choice or convention; that the relevant standards are 'higher' than positive laws, conventions and practices, that is, provide the premises for critical evaluation and endorsement or justified rejection of or disobedience to such laws, conventions or practices; that the relevant standards conform to the most demanding requirements of critical reason and are objective, in the sense that a person who fails to accept them as standards for judgment is in error; that adherence to the relevant standards tends systematically to promote human flourishing, the fulfilment of human individuals and communities.

2 Critique of scepticism and dogmatism

Historically, natural law theories have been articulated as part or product of a philosophical critique of ethical scepticisms (see Scepticism), and have included in their critique a differentiation of the rationally grounded norms of natural law from moral dogmatism or conventionalism. In contemporary thought, scepticism about natural law (and about all moral theories claiming to be objective or true) is very often based upon a logically illicit and rationally unwarranted inference from certain propositions about what 'is' the case to certain propositions about what is good or obligatory. This particular form of invalid reasoning was assiduously employed in the ancient manuals of scepticism (notably by Sextus Empiricus, *c*. AD 200), and reintroduced into European discourse in the

sixteenth century. Sextus Empiricus uses the term ‘dogmatism’ to insinuate that all non-sceptical ethical theories, such as natural law theory, are uncritical; but this charge is not well grounded, and the term is used here to refer to moral positions held without openness to critical questions.

Examples of the invalid reasoning commonly encountered include:

X is not universally regarded as good/obligatory; therefore X is not good/obligatory.
In modern thought X is widely regarded as not good/obligatory; therefore X is not good/obligatory.
In contemporary society X is widely regarded as good/obligatory; therefore X is good/obligatory.
I have a sentiment of approval of X; therefore X is good (or worthwhile…or obligatory…), at least for me.
I have opted for or decided upon or am committed to the practical principle that X ought to be done; therefore X ought to be done, at least by me.

As this list of *non sequitur*s suggests, there is a link between ethical scepticism (at least in its popular forms) and ethical conventionalism. There are many natural law theories, on the other hand, which are not guilty of these or other fallacies, fallacies that consist in concluding to a normative judgment from premises which include no normative proposition.

David Hume (1740) suggests that ‘every system of morality’ prior to his critique illogically purports to infer ought or ought not from is or is not (see Hume, D.). The suggestion is ungrounded, though the illicit inference may perhaps be detected in certain eighteenth-century rationalists (especially Samuel Clarke) whom Hume seems to have had prominently in mind when writing this part of his *Treatise*. In so far as Hume’s predominant view concerning the nature and basis of moral judgments was that they are judgments about what characteristics and actions arouse approval or disapproval, Hume was himself plainly guilty of this kind of illegitimate inference. The same cannot be said of Plato, Aristotle and Aquinas, for example.

Max Weber is often, and with some plausibility, said to have introduced the modern form of the ‘fact/value distinction’ and denial of the ‘objectivity of value judgments’. Weber’s primary argument for this denial seems to have been simply that people, or educated people, in fact disagree with each other about values. This argument has no more validity than the popular arguments listed above; the fact of disagreement no more disproves a proposition than the fact of agreement proves it. Weber had three other lines of argument. The first was a Neo-Kantian argument that all judgments must rest upon presuppositions, and that selection among competing presuppositions must be non-rational. This argument is self-refuting in its general form, and there is no way to limit it (as Weber sought to do) to judgments of value. Weber’s next argument pointed, like Sartre after him, to certain supposed political and/or ethical dilemmas before which reason is supposed to fall silent. In fact, however, such ‘dilemmas’ are no more than a stimulus to a more nuanced and resourceful practical reasoning, which may identify more than one option as morally acceptable in the situation. Weber’s final argument alleged that there are distinct and incommensurable ‘spheres’ of practical judgment, such as the political, the erotic and the ethical, each with its own ultimate values between which reason cannot adjudicate because reason operates only within spheres. This final claim is never defended by Weber, and is contradicted by his own acknowledgement that the spheres interpenetrate one another.

Perhaps the most fruitful critique of ethical scepticism is one which takes its cue from Aristotle’s critique of general scepticism (*Metaphysics* IV). If sceptics are willing to affirm their position as rationally warranted they can be shown to be denying (1) something given, instantiated, in the activity of rationally considering and proposing their or any other position, and/or (2) some proposition to which their assertion of their position rationally commits them. In the case in hand, the relevant givens include the rationality norms which guide the rational inquiries of those who choose to follow them and to resist temptations to reach conclusions by sub-rational processes; and the propositions to which an argued assertion commits the person asserting it include the proposition that knowledge of truth (at least the truth of such propositions as the one asserted) is a good worth pursuing and instantiating in that argument and assertion, a good worthwhile for its own sake as well as for any instrumental advantage it might yield.

3 Cognitivism and natural law theories

Not every non-sceptical ethics is appropriately called a natural law theory. Natural law theories are distinguished

from the broader set of cognitivist or objectivist ethical theories in four main ways.

First, they are differentiated from any ethics in a Kantian mould by their willingness to identify certain basic human goods, such as knowledge, life and health, and friendship, as the core of substantive first principles of practical reasoning (see Kantian ethics). Taken together, these basic human goods give shape and content to a conception of human flourishing and thus, too, to a conception of human nature. For an axiom of Aristotle's method, deployed more generally by Aquinas, shows that while nature is metaphysically (ontologically) fundamental, knowledge of a thing's nature is epistemically derivative: an animate thing's nature is understood by understanding its capacities, its capacities by understanding its activities, and its activities by understanding the objects of those activities. In the case of the human being the 'objects' which must be understood before one can understand and know human nature are the basic goods which are the objects of one's will, that is, one's basic reasons for acting, that give reason for everything which one can intelligently take an interest in choosing.

Second, natural law theories are distinguished from any theory which asserts that moral truths are known essentially by discrete 'intuitions' (see Intuitionism in ethics §3). Rather, these theories argue that specific moral judgments concerning obligation or right are applications or specifications of higher principles. The first principles of the 'system' are known by insight without deduction via any middle term (they are *per se nota*: Aquinas, *Summa theologiae* IaIIae.94. 2). But the insights whose content is the self-evident principles of practical knowledge are not intuitions - insights without data. Rather they are insights whose data are, in the first place, natural and sensory appetites and emotional responses. These data are subsequently enriched by theoretical knowledge or true opinion about possibilities (for example, about what threatens and enhances health, or about what knowledge is available), and by experience of disharmony (frustrated intentions). It should be added that the classical natural law theorists would reject Jacques Maritain's theory that the knowledge of first practical principles is 'connatural' or 'by affinity or congeniality' or 'through inclination' rather than conceptual. There is compelling reason not to accept Maritain's claim that Aquinas must have been referring to such 'knowledge' when saying (in the passage cited above) that 'reason naturally understands as good all the objects of man's inclinations'. For Aquinas all knowledge is conceptual (*De veritate* 4.2.5), and the understanding of basic human goods is quite ordinary, unmysterious, *conceptual* understanding (Grisez 1965), albeit practical, that is, directive or prescriptive ('is-to-be-done-and-pursued'), rather than theoretical or descriptive ('is'). The first principles of natural law are not inclinations, but fundamental human goods understood as reasons for action.

Third, natural law theories are distinguished from any fundamentally aggregative conception of the right and the just. For viable natural law theories postulate no one end to which all human actions might be effective means, no one value in terms of which one might commensurate alternative options as simply better or worse, and no one principle which, without further specification in other principles and norms, should guide deliberation and choice. Rather they claim to identify a number of basic human goods, none of which is simply a means to or simply a part of any other such basic good. A number of principles are also identified to guide ('morally') the choices necessitated by the variety of basic goods and reasons for action and the multiple ways of instantiating these goods and acting on these reasons for action by intelligent and creative choice (as well as by misguided choices whose primary motivation is not reasons but emotion).

Fourth, natural law theories typically differ from other ethical theories by offering to clarify not only the normative disciplines and bodies of discourse, but also the methods of the descriptive and explanatory social theories (political theory or political science, economics, legal theory and so on). How best can human societies and their formative concepts be understood, without illusions, but in a general way (as in a project such as Aristotle's *Politics* or Max Weber's *Wirtschaft und Gesellschaft (Economy and Society)*)? Could such projects be 'value-free'? Or must even descriptive-explanatory theorists, in selecting concepts, rely upon some definite conceptions of what is important in human existence? Must they not use such conceptions as criteria for selecting topics for study, and concepts for describing those topics? Must they not also employ such criteria in judging some types and instantiations of human institutions or practices to be the 'central cases' of such institutions or practices, and also in judging some uses of terms such as 'law' or 'constitution' or 'authority' to be, for critical descriptive theory, the 'focal' uses and senses of those terms? And must not such conceptions and criteria of importance be the subject, not of selection by 'demonic' personal preference (Weber) or silent conformism to academic fashion or political *parti pris*, but rather of an open, public, critical justification? Natural law theories of the classical type, such as those of Aristotle and Aquinas, claim to offer such a justification (see Finnis 1980).

4 Derivation of positive law

The history of moral and political philosophy, and of natural law theory, is much affected by the lacunae in Aquinas' work. He was the thinker who most clearly articulated and developed the elements of practical philosophy in Aristotle (and thus, in a sense, in Plato), so as to represent them as propositions of or about natural law. Yet he failed to give a clear, full, careful and consistent statement of the first principles of practical reason, and a satisfying set of illustrations of the way in which first practical principles (such as 'human life is a good to be advanced and respected') or first moral principles (such as the 'golden rule') are given specificity in less fundamental principles and norms. This failure contributed significantly to the spread of voluntarist and fideistic currents which virtually overwhelmed the tradition of natural law theory as a living school of thought.

Aquinas was the first philosophical thinker to exploit the concept of 'positive law' which emerges in mid twelfth-century theological and juristic speculation. He gave an original and helpful sketch, scarcely surpassed even today, of the differing types of relationship between the high-level principles of natural law and (1) the more specified principles which can be attained by practical reasoning about the implications of those principles in recurrent types of human predicament and opportunity; and (2) norms, rules, precepts and juridical institutions which cannot be said to be required as the conclusion of any course of practical reasoning, yet which are in some other way rationally connected to, or derived from, principles of the first two sorts, and are authoritative for upright judges and good citizens if and when chosen and promulgated ('posited') by an appropriate authority. Laws, rights and institutions derived in the second way are 'positive', yet have much the same moral force in conscience as the higher-level principles of natural law, provided that their positing was by a person or body constitutionally authorized to make such decisions and making them without serious disregard either for other relevant moral principles and rights or for the 'common' (as opposed to partisan) good of all the members of the community subject to their authority.

This second mode of derivation was called by Aquinas *determinatio*, which might be translated 'concretization'. It is best understood by a comparison such as he offered by way of explanation. From his commission ('build a maternity hospital for our town') an architect can deduce various specifications (the building must be more than one metre in height, and must include doors, means of warming the spaces in winter, and so on). But no amount of attention to the commission (the 'principle' of his work) and to the circumstances (including the 'nature of things') will yield a single, rationally required answer to the unavoidable questions. Should the doors be 2.1 or 2.2 metres high? Of this metal or that? Questions of this kind must be answered by decisions (*determinationes*) which are rationally under-determined, yet intelligibly related to (and in this weak sense 'derived from') the master principle (the commission taken with its more or less necessary implications).

In this account the sheer positivity of much of a country's legal system is fully acknowledged. But at the same time the moral significance of the law's directives and institutions is affirmed and explained. The affirmation is conditional. If the relationship of 'derivation', albeit nondeductive and 'free', is broken by *ultra vires* (lack of authority) or unjust discrimination or violation of inviolable rights, the positive law's proper moral authority is eliminated. From the point of view of the conscientious judge, citizen or indeed legislator, this lack of the proper ('normal') moral authority was marked by the sayings found in Plato, Cicero, Aquinas and most of the philosophical, theological and juridical tradition down to the nineteenth century: an unjust law, albeit valid by the legal system's own criteria of validity (and in this sense properly described as a law), is not really a law, or is not a law *simpliciter*, that is, without qualification. It can have some collateral obligatoriness, in that it may be unfair to others to disobey it publicly (for example, because doing so would encourage others to disobey other laws without good reason). This classical thesis, *lex iniusta non est lex* (an unjust law is not a law), properly interpreted, is fully compatible with the jurist's or historian's wish to employ an amoral criterion of 'legal validity' or historical cognizability (see Aquinas, T. §13).

5 Inviolable human rights

The collapse of Christianity and other religious cultures, as the matrix for contemporary legal and political orders, has posed a challenge to those who wish to affirm that there is a natural law. In general, the challenge is to show that, even without the support of a religiously warranted ethic (revealed divine law) having an identical or overlapping content, there are philosophically sound reasons to affirm the truth of non-posited principles or norms

which, although not claiming to be authoritative because posited, are sufficiently definite to exclude gross ‘crimes against humanity’ (Nuremberg, 1945; chattel slavery; nonvoluntary euthanasia; and so forth) and to underpin the main institutions of civilized societies (family, property, religious liberty and so forth).

In particular, the challenge is to show that aggregative conceptions of moral and political reasoning, such as classical or contemporary utilitarianism or consequentialism or ‘economic analysis of law’, are unsound (see Utilitarianism §5; Consequentialism §2; Law, economic approach to §3).

Contemporary work in natural law theory therefore includes extensive attempts to criticize aggregative ethics, by showing that their conceptions of value overlook some or all of the basic human goods, and/or that their master principle of maximizing value overlooks the incommensurability not only of one basic human good relative to other such goods, but also of particular instantiations of even one and the same good in alternative options for morally significant choice. The supposition of those undertaking these critiques is that, if they are successful, the way will be open to identifying a richer set of moral principles guiding choice. A working postulate has been that any principle which has been the organizing or dominant principle of an entire philosophical ethic (for example, Kantian universalizability, Epictetan detachment, the principle of commitment expounded by Royce and Marcel) has some place in a developed natural law theory.

But theoretical reflection has yielded a more systematic and unifying ‘master principle of morality’. This principle is reached by way of the consideration that, so far as it is in one’s power, one should allow nothing but the principles corresponding to the basic human goods to shape one’s practical thinking. Aquinas’ first practical principle, ‘Good is to be done and pursued and evil avoided’, taken as it stands, requires only that one not act pointlessly, that is without reason; it requires only that one take at least one of the principles corresponding to a basic human good and follow through to the point at which one somehow instantiates that good through action. The first moral principle makes the stronger demand, not merely that one be reasonable enough to avoid pointlessness, but that one be entirely reasonable in one’s practical thinking, choice and action. It can be formulated: in voluntarily acting for human goods and avoiding what is opposed to them, one ought to choose and otherwise will those and only those possibilities whose willing is compatible with a will towards integral human fulfilment (that is, the fulfilment of all human persons and communities).

Integral human fulfilment, so defined, is not a goal towards which actions could be organized as means. But it is not empty of critical force, in identifying choices motivated by emotion fettering and deflecting reason. Unfairness, acting contrary to the golden rule (a principle superior in determinacy to Kant’s substitute, universalizability (see Kantian ethics §4)), is one sort of willing that is incompatible with openness to integral human fulfilment. A choice to destroy some instantiation of basic human good out of hatred is another. A choice precisely to destroy or damage some instantiation of a basic human good as a means to some other end is yet another, which comes into view as a demand of reason once the critique of aggregative theories has demonstrated the rational incommensurabilities overlooked by those who think that the goods promised by the end ‘outweigh’ the destruction or damage to human personal goods which is intended in a choice of this type.

This third moral principle corresponds not only to the traditional ‘the end does not justify these means’ but also to the Kantian ‘treat humanity always as an end and never as a means only’ - though ‘humanity’ is to be understood without Kant’s dualistic restriction of it to the rational faculties as such. It is a principle specified (made more specific) in the form of the traditional moral norms against murder and fraud. As such it is the backbone, so to speak, of traditional and modern legal systems. Articulated from the viewpoint of the beneficiaries of its moral protection, it is specified in those human rights which are not only fundamental but also, properly speaking, inalienable. It is the resultant of the classical principle that reason neither need nor should be the slave of the passions, together with both a precise understanding of free choice (enabling a firm and definite distinction between what is chosen and what is only accepted as a side-effect), and an understanding, developed in dialectic with Enlightenment and post-Enlightenment aggregative ethics, of the incommensurabilities of the goods, persons and probabilities at stake in alternative options for choice. In these and related ways one can think that natural law theory of the classic type is capable of philosophically warranted development (see Rights §3).

See also: Grotius, H.; Law, philosophy of; Legal idealism; Legal positivism

JOHN FINNIS

References and further reading

(1256-9) *De veritate*, *Opera omnia*, vol. 22, Rome: Commissio Leonina, 1970-4; trans. R.W. Mulligan *et al.*, Chicago: Regnery, 1952-4.(In which Aquinas argues that all knowledge is conceptual.)

(1266-73) *Summa theologiae (Synopsis of Theology)*, Milan: Edizioni Paoline, 1988; trans. T. Gilby, London: Eyre and Spottiswoode and New York: McGraw Hill, 60 vols, 1963-73.(Aquinas' greatest and most characteristic work.)

Aristotle (*c*.384-322BC) *Metaphysics* (*Metaphysica*), W.D. Ross (ed.), 2 vols, Oxford: Oxford University Press, 1924; Books IV-VI, trans. C.A. Kirwan, Oxford: Oxford University Press, 1971.(Include's Aristotle's critique of general scepticism.)

Finnis, J. (1980) *Natural Law and Natural Rights*, Oxford: Clarendon Press.(A wide-ranging and fundamental treatment with some, but not exclusive, emphasis on legal philosophical issues; suitable, though difficult, for a first approach to the subject.)

(ed.) (1991) *Natural Law*, vols 1 and 2, Aldershot: Dartmouth Press; New York: New York University Press.(An extensive collection of previously published essays by various Anglo-American authors stating the case for and against natural law theory, and including several debates; gives a broad coverage of recent discussions.)

Grisez, G. (1965) 'The First Principle of Practical Reason: A Commentary on the *Summa theologiae*, 1-2, Question 94, Article 2', *Natural Law Forum* 10: 168-201. (Also in Finnis 1991, vol. 1. A classic article reinterpreting Aquinas and launching what might be called 'new classical natural law theory'; a seminal essay.)

(1978) 'Against Consequentialism', *American Journal of Jurisprudence* 23: 21-72.(Also in Finnis 1991, vol. 2. A wide-ranging critique of utilitarianism, consequentialist, proportionalist, and other aggregative ethical methods.)

Grisez, G., Boyle, J. and Finnis, J. (1987) 'Practical Principles, Moral Truth, and Ultimate Ends', *American Journal of Jurisprudence* 32: 99-151.(Also in Finnis 1991, vol. 1. A restatement and defence of new classical natural law theory against mainly conservative philosophical critics.)

Hume, D. (1740) *A Treatise of Human Nature*, 2nd edn, ed. L.A. Selby-Bigge and P.H. Nidditch, Oxford: Oxford University Press, 1978.(Book III contains the classic dismissal of natural law theory for alleged derivation of 'ought' from 'is', and for treating reason as a fundamental ground of action.)

Rommen, H. (1947) *The Natural Law: A Study in Legal and Social History and Philosophy*, St Louis, MO, and London: Herder Book Company.(A full and careful statement of implications and applications of natural law theory in politics and economics as seen from a mid twentieth-century Catholic perspective.)

Natural philosophy, medieval

Medieval Latin natural philosophy falls into two main periods, before the rise of the universities (mainly in the twelfth century, when works were produced in connection with aristocratic patrons, monastic institutions or cathedral schools) and after their rise. In the earlier period, the dominant Greek influence is that part of Plato's Timaeus which had been translated into Latin and commented on by Calcidius. In the university period, the central works are those of Aristotle, often together with commentaries by Averroes.

Before the twelfth century, there was very little that could be described as natural philosophy. Such work as existed fell mainly into the genres of natural history (encyclopedic works using Pliny and the like as sources), didactic works (perhaps following a question and answer format on the model of Seneca's Natural Questions) or biblical commentary (especially commentaries on the Hexaemeron, or six days of creation). In the twelfth century, however, there are a number of original texts that may be considered as natural philosophy; examples include William of Conches' Philosophia mundi (Philosophy of the World), Bernard Sylvester's Cosmographia or Hildegard of Bingen's Scivias (Know the Ways). Greek natural philosophy also reached the Latin West through its influence on medical works and on art, for example on drawings of the cosmos, heaven, angels and hell.

The high and late Middle Ages (thirteenth and fourteenth centuries) was perhaps the preeminent period in all of history for natural philosophy. Natural philosophy was an official area of study in the arts faculties of medieval universities, alongside and distinct from the seven liberal arts (the trivium - grammar, rhetoric and logic - and the quadrivium - arithmetic, geometry, astronomy and music), moral philosophy or ethics, and first philosophy or metaphysics. As a subject of the arts faculty, natural philosophy was also defined as distinct from the subjects studied in the graduate faculties of theology, medicine and law.

The most common approach to natural philosophy in the thirteenth and fourteenth centuries was to comment on, or to dispute questions arising from, the natural works of Aristotle, especially his Physics, On the Heavens, On Generation and Corruption, Meteorology and On the Soul, as well as his various works in biological areas and the so-called Parva Naturalia, a group of short works on psychological topics. Medieval investigations of the cosmos that were largely mathematical - for example, most of astronomy - were considered in the Middle Ages to belong not to natural philosophy but to the quadrivium or perhaps to the so-called 'middle sciences' (such as optics, statics or the newly developed 'science of motion'). What little medieval experimental science there may have been (for instance that appearing in Peter Peregrinus' De magnete (On the Magnet), in Frederick II's De arte venandi cum avibus (On the Art of Hunting with Birds) and perhaps in some works on alchemy) seems not to have been done within the university setting. In the fourteenth century the new methods of medieval logic (supposition theory, propositional analysis or exposition, rules for solving sophismata and so on) are prominently used in natural philosophy.

Thirteenth- and fourteenth-century natural philosophy began with the general assumption of the Aristotelian world view, but later medieval natural philosophers did not hold to the Aristotelian view rigidly or dogmatically. In some cases, Christian faith seemed to contradict or to add to Aristotle's ideas, and natural philosophers tried to resolve these contradictions or to make the appropriate additions, as in the case of heaven and hell and angels. A number of difficulties, inconsistencies and sticking points in Aristotle were special subjects for discussion and received new resolutions as time went on.

Within the medieval university, natural philosophy was considered to be a part of general education, but it was also thought to be useful as a tool for theology and medicine. In northern universities such as Paris and Oxford, some of the most fundamental original work in natural philosophy was done in connection with the investigation of theological problems, for which natural philosophy, together with the other disciplines of the arts faculty, served as important aids. In Italian universities, where faculties of theology were less prominent or non-existent, natural philosophy was similarly tied to the resolution of medical questions.

European libraries contain many manuscript commentaries on Aristotelian works that still await modern analysis. The medieval university system did not as a rule identify, encourage or reward originality or uniqueness. Many natural philosophers claimed to be explaining Aristotle's meaning, even when they were introducing a novel interpretation of or variation on his ideas. When they made use of the ideas of earlier commentators, they rarely

mentioned them by name. What we now know about medieval natural philosophy is not a mirror reflection of what happened in the thirteenth and fourteenth centuries, because modern scholars have chosen to study those subjects and individuals relevant to their own present situations: Dominicans have emphasized the history of Dominican natural philosophy in such thinkers as Albert the Great and Thomas Aquinas, Franciscans have studied Franciscans such as John Duns Scotus and William of Ockham, historians of science have studied those individuals who had something to say about the subjects of modern science such as bodies, forces, velocities and resistances, logicians have studied logic, and so on. Because natural philosophy as such is not the focus of attention of many modern philosophers or other scholars, much medieval natural philosophy remains unread, sometimes in large-scale and handsomely produced commentaries on Aristotle's works, sometimes in hastily scribbled student notebooks.

1 Natural philosophy before and outside universities

From the beginning of the Middle Ages until the rise of the universities, relatively few people had time or opportunity for natural philosophy. At the beginning of the Middle Ages, Boethius set out to translate the works of Plato and Aristotle from Greek to Latin, and also proposed to write textbooks in Latin on the seven liberal arts. Before his imprisonment for treason in mid-career, he had succeeded in translating some of Aristotle's logical works and in editing textbooks of arithmetic, geometry and music. While in prison awaiting execution, he completed the classic *De consolatione philosophiae (The Consolation of Philosophy)*, in which he reflects on the pleasures of philosophy. Unfortunately, Boethius' early death prevented him from translating all of Aristotle's works or any of Plato's, and he did not translate any of Aristotle's natural philosophy.

Earlier, the Neoplatonist Calcidius had translated part of Plato's *Timaeus* into Latin. In this dialogue, Plato told a myth about the origins and nature of the cosmos and transmitted a number of Greek ideas (for example, that there are four elements, earth, water, air and fire, which can be transformed into each other). Because Aristotle's physical works were not available in Latin until the twelfth century, Calcidius' version of *Timaeus* was heavily used by early medieval thinkers as their source of knowledge about natural philosophy.

From 500-1000 AD, most of those who had the training and opportunity to study natural philosophy were connected with the church, as monks, priests, bishops or nuns. A century after Boethius, the bishop Isidore of Seville wrote his encyclopedic work *De rerum natura (On the Nature of Things)*, and a work of the same name was produced a further century later by the Venerable Bede in Northumbria. Both men were attempting to improve the low state of knowledge by condensing and transmitting the natural knowledge found, for example, in Pliny's *Natural History*, an encyclopedic compilation dating from the first century AD (see Encyclopedists). By the year 800, Charlemagne had established his court school under the direction of Alcuin of York, where the latter supervised the copying of manuscripts and oversaw the correction of existing copies of the Latin Bible. Among the Carolingians, Hrabanus Maurus again wrote the encyclopedic work *De rerum natura (On the Nature of Things)* as part of his programme for reforming the education of the clergy. Although Isidore of Seville, the Venerable Bede and Alcuin were among the best-educated people of their times, the intellectual communities in which they worked were small and isolated, consisting of monasteries and a few court-based schools (see Carolingian Renaissance).

By the turn of the tenth century, however, perhaps partly as a result of the efforts of Charlemagne and Alcuin, the situation was changing. John Scottus Eriugena, recruited from the British Isles into the Frankish court of Charlemagne's grandson, was able to make use of sources written in Greek. Through Eriugena, many of the ideas of the Neoplatonists, particularly those of the mysterious Pseudo-Dionysius the Areopagite, were imported into the West in Latin translation. Through Eriugena's translation of Pseudo-Dionysius' *De coelesti hierarchia (On Celestial Hierarchy)*, the West inherited its elaborate notion of nine orders of angels (angels, archangels and principalities, dominions, powers and authorities, and cherubim, seraphim and thrones). Though some of Eriugena's work was condemned as heretical in the thirteenth century, and although Pseudo-Dionysius was later shown not to have been Paul's Athenian disciple as had been thought (but rather someone living after the time of Proclus), nevertheless an elaborate taxonomy of angels was accepted as an intrinsic part of the cosmos throughout the later Middle Ages. Angels were considered no less real by medieval scholars than neutrinos or quarks are to educated people of the late twentieth century, despite their invisibility to the ordinary eye.

By the year 1000, Gerbert of Aurillac, tutor to the Holy Roman Emperor Otto III, had already travelled to Spain to bring Islamic knowledge to the rest of Europe. Gerbert was perhaps the first person to be able to operate on an

intellectual level comparable to that of Boethius, at least as regards the seven liberal arts (especially the trivium). It was at about this time that something that might be called natural philosophy began to appear in the schools established in cathedrals in urban centers such as Chartres, Paris and Laon. Such natural philosophy might take the form of a creation myth, such as Bernard Sylvester's *Cosmographia*, a work that combined ideas of Genesis with ideas of Plato's *Timaeus* to describe the origin and organization of first the macrocosm (natural world) and then of the microcosm (the human being) (see Bernard of Tours). Some twelfth-century authors, for example William of Conches, attempted to give an explicitly natural explanation of the cosmos, albeit within an assumed Christian context. The general world picture that these Platonistic Christian authors passed on to the natural philosophers of later centuries was one in which God first establishes and then maintains order in the cosmos, so that there is a natural design which natural philosophers may study using purely human, rational and empirical means.

Before the rise of the universities in the twelfth century, then, there was a well-established intellectual culture in the Latin West, largely dependent on clergymen but extending also to a few nuns and a few of the aristocratic laity of both sexes. Apart from the subjects of elementary education, manuscripts might concern religious topics, medicine (including compilations about medical plants), animal lore (bestiaries), basic astronomy for calendrical purposes (*computus*), or astrology and meteorology. Clerical readers might make use of animal lore in preparing sermons, taking examples of courage from lions or of compassion from pelicans and the like.

In the twelfth century, the available writings on nature were suddenly vastly increased by a spate of translations from Arabic to Latin, either of Greek works that had previously been translated into Arabic, or of original Arabic works (see Translators; Islamic philosophy: transmission into Western Europe). Thus in the late twelfth and early thirteenth centuries, as the universities were beginning to develop, philosophers suddenly had access to the natural philosophical works of Aristotle, which had previously been unavailable in Latin. Although the works of Aristotle were over fifteen hundred years old by the year 1200, they represented a level of natural philosophical sophistication well above what had been available in the earlier Middle Ages. They were therefore highly attractive to the clerical scholars who read them despite the fact that they contained ideas at least superficially contradicting Christianity, such as Aristotle's purported proof that the universe must always have existed. Initially, popes and bishops tried to forbid the reading or teaching of Aristotle's natural works, but by the mid-thirteenth century these works became the basic teaching texts at the University of Paris and other universities. In the later thirteenth century, Albert the Great, Thomas Aquinas, Robert Grosseteste and Roger Bacon were among the first Latin commentators to lecture to university audiences Aristotle's *Physics*, *On the Heavens* and other natural works. In doing this, they often made use of translations of the Arabic commentaries of Averroes (see Ibn Rushd), composed in Muslim Spain in the previous century. They also used Latin translations of the works of other Islamic and Jewish authors such as al-Kindi, al-Farabi, Avicenna (see Ibn Sina) and Maimonides.

Finally, there was the study of subjects such as astrology and alchemy, which were at least peripherally related to natural philosophy but were not part of official university curricula. Both these disciplines made considerable use of Aristotelian ideas of the cosmos, celestial influences on earth, the elements, chemical change and so forth. The situation is confused by the fact that there are works on astrology, alchemy and even magic attributed to prominent university scholars such as Albert the Great and Thomas Aquinas. These works, if authentic, cast considerable light on the natural philosophical views of their authors; even if they are not by the individuals to whom they are ascribed, these works can tell us much about the general ambience of medieval natural philosophy, and particularly about the perceived connection between the heavens and the earth.

2 Natural philosophy in medieval universities: sources

In the thirteenth and fourteenth centuries, natural philosophy was done most directly in the form of disputing questions arising from Aristotle's natural works. Although, as has been stated, a distinction was maintained in the medieval university between the discipline of natural philosophy on the one hand and the disciplines of the liberal arts and higher faculties of theology and medicine on the other, nevertheless some natural philosophy, including some of the most advanced and innovative work, was also done under the umbrellas of theology (especially in northern universities) or medicine (especially in southern universities).

The primary sources upon which thirteenth- and fourteenth-century natural philosophers drew were the natural works of Aristotle and later commentaries on these works. Aristotle himself often discussed the views of his predecessors, and his medieval commentators followed him in this practice. Thus a medieval commentator might

contrast Aristotle's views with those of the Presocratics - including Thales, Anaximander, Anaximenes, Heraclitus, Pythagoras, Parmenides, Zeno of Elea, Anaxagoras, Empedocles, Leucippus and Democritus - or he might compare Aristotle's ideas to those of Plato. Additional sources of ideas included medical works, such as those of Hippocrates (see Hippocratic medicine) and Galen among the Greeks as well as various medieval medical authors. Medieval natural philosophers also knew something about the ideas of the Stoics (see Stoicism), the Epicureans (see Epicureanism), the Neoplatonists, including Simplicius and Philoponus (see Neoplatonism), and Roman and medieval Latin authors from Pliny and Seneca to Calcidius, Macrobius, Martianus Capella, Boethius, Isidore, Alcuin, John Scottus Eriugena, Gerbert of Aurillac and various others (see Encyclopedists). Among Islamic authors the most influential were al-Kindi, Avicenna (see Ibn Sina) and Alhazen as well as Averroes (see Ibn Rushd) and his many Islamic predecessors whom Averroes discussed in his commentaries on Aristotle, such as Avempace (see Ibn Bajja). A few Jewish authors, such as Maimonides, were also used.

Although they were not likely to cite them in purely philosophical work, medieval commentators on Aristotle's natural philosophy also knew earlier, specifically Christian writings such as the various works of Augustine and other church fathers, including commentaries on Genesis and discussions of predestination, the nature of the soul, the sacraments and so on (see Patristic philosophy). As medieval theology developed, it not infrequently led to partly philosophical speculations which, in turn, had an influence on natural philosophy. Theological discussions of God's power, eternity, infinity, the Eucharist, the soul and the angels produced results that subsequently fed into the further development of natural philosophy. Thus for example, the idea that in the Eucharist the qualities of the bread and wine are present without inhering in any substance led to the conclusion that it was not self-contradictory that there might be accidental forms that existed independently.

3 Natural philosophy in medieval universities: goals

The purpose of natural philosophy as taught in medieval universities was to understand nature: that is, the created world, or cosmos, and everything contained in it. As indicated above, the materials used by university scholars for their study of the cosmos were the writings of their predecessors, pre-eminently the works of Greek philosophers (especially Aristotle and Plato and their followers), but also Muslim and Jewish philosophers insofar as their works had been translated into Latin, and earlier writers in Latin. It was quite unusual for medieval natural philosophers to draw new empirical information into their work except by reference to common knowledge and daily experience.

Theology, however, did provide what were considered new facts to be taken into account by natural philosophers insofar as they took theology to be a science comprised of truths about the world. For example, theology taught that angels were real created beings; consequently, medieval natural philosophers took angels into account in forming general metaphysical or natural philosophical theories. If angels were thought to be both indivisible and moveable, then natural philosophers might question Aristotle's argument that it is impossible for an indivisible to move.

Natural philosophy, along with the liberal arts, was thus a 'handmaiden' or service discipline to theology, but it would be a mistake to confuse the situation in medieval universities with that found among some fundamentalists today. By and large, intellectual integrity and exactness were hallmarks of medieval university scholars including the members of the theology faculties, who were nothing if not rational. Going as far back as Augustine (and to Philo Judaeus before him), Christian scholars had come to the conclusion that if Christians were not well-educated they could not win the respect of non-Christians in their societies, who were often, as in the case of the Greeks, highly educated themselves. Moreover, Christian scholars wanted to learn the truth about the world in its natural as well as its divine aspects, and they did not expect that true philosophy would contradict true religion. They had not yet imagined a 'warfare' of Christianity and science, such as later occurred between the Copernican view of the universe and the Catholic church in the case of Galileo, or between Darwinian evolution and creationism. Medieval scholars understood that God gave human beings reliable minds and sense organs so that humans could learn about the cosmos, as well as about God's plan for it.

The goal of medieval natural philosophy was therefore to discover the truth about the world for its own sake. Medieval natural philosophy differed from modern science not in being tied to religious dogma and authority, but rather in the fact that it was *not* tied to technological application or economic spin-off. Modern scientific critics sometimes discredit medieval natural philosophy on the grounds that it was not sufficiently experimental or

mathematical, and seems to have spun fantastical theories solely on the basis of cloistered speculation. Medieval natural philosophers might defend themselves against this criticism on the ground that their job was to teach future priests, ecclesiastical or secular administrators, lawyers and physicians for whom scholastic methods were well suited. Had they been asked to teach future engineers, their approach might certainly have taken more mathematical and empirical directions. To ask whether medieval natural philosophers should have had a technological goal is beyond the scope of this article: the relevant point here is that the goal of medieval natural philosophy was knowledge of the world for its own sake and as a sturdy and flexible tool for use in the higher faculties, not for use in technology (the modern Western assumption that rapid technological progress is possible and good is the exception rather than the rule within world history).

4 Natural philosophy in medieval universities: practitioners

Because natural philosophy was mainly a university subject, and because medieval students and faculty were considered to be clergymen of a sort, the practitioners of medieval natural philosophy were almost always celibate men. Rarely and marginally, a work with natural philosophical content was written by a medieval woman, usually a nun - Hildegard of Bingen, for instance, who was the author of books of visions including visions of the cosmos - or by a non-clerical man - for example, the Emperor Frederick II, who was the author of a work on falconry including natural philosophical content.

Modern scholars, writing from Marxist, feminist or third world points of view, have faulted medieval natural philosophy (and medieval scholasticism in general) for its narrowness, arguing that because its study was limited to celibate white Christian men, it represented the interests and biases of only a very small segment of society. Certainly this narrowness may explain why the study of natural philosophy was devoted to a certain kind of knowledge, with no concern for the material well-being of society as a whole. Then as now, however, some authors were motivated by narrow self-interest while others had a more generous point of view. Because the medieval church was a powerful social institution that placed a high value on intellectual inquiry and supported scholars, many highly intelligent men were attracted into the universities and were given the institutional support to devote their entire attention to intellectual matters.

5 Natural philosophy in medieval universities: literary genres

The forms in which medieval natural philosophy was written were mainly determined by the teaching practices of medieval universities. This teaching fell into two main types, the lecture and the disputation. Lectures in natural philosophy most frequently involved the reading (*lectio*) of one of the books of Aristotle, either ‘cursorily,’ simply reading the book rather quickly, or in the more formal ‘ordinary’ fashion, with pauses in which the lecturer explained the structure of the text, emphasized Aristotle’s main points or conclusions, cleared up special difficulties and so on. In a disputation, on the other hand, not only the lecturer but several disputants, often including advanced students, took part. Some disputations were tied to the texts covered in the lectures and concerned questions that arose from the text. One or more individuals (the ‘opponent’) would raise problems about the question, and one or more individuals (the ‘respondent’) would reply. After this had gone on for some time, the master overseeing the disputation would give what he considered to be the correct answer to the question (the ‘determination’), organizing logically the main points that had been made in the disputation, reinforcing the arguments that supported his view and responding to arguments for alternative positions. More rarely there were disputations on more general themes, not closely tied to any one text. One type of disputation detached from texts was the disputation on sophismata (*de sophismatibus*), in which students tried to untangle the possible truth-values of seemingly paradoxical statements, some of which concerned natural subjects such as heating and cooling or the traversal of distances. Contrary to stereotypes deriving from the trial of Galileo in the seventeenth century, medieval university disputants pursued their inquiries with relentless logic and were not constantly looking over their shoulders lest the expression of heretical views might lead to their condemnation. In fact, medieval university disputants were often themselves the authorities whose job it was to determine what was true and what was not.

These oral lectures and disputations sometimes led to written accounts, and masters of arts also occasionally wrote expositions of texts or determinations of questions in advance of lecturing or instead of taking part in oral disputations. There are also marginal glosses and brief expositions on the manuscripts of Aristotle’s natural philosophical works, which formed a large percentage of the medieval natural philosophical textbooks; an example is Robert Grosseteste’s notes or commentary on Aristotle’s *Physics*. Some scholars, such as Albert the Great,

wrote paraphrases of the Aristotelian texts, perhaps to make the meaning clearer. (The Latin translations of Aristotle, either directly from the Greek or indirectly through an intermediate Arabic translation, were often difficult to understand simply because of the wording.) More important in terms of the specifically medieval contribution to natural philosophy were compilations of ‘questions’ on the Aristotelian texts. Written in the standard ‘scholastic’ form, these ‘questions’ began with a question related to the text and then gave arguments for and against a number of possible answers to the question, before going on to the author’s preferred answer. Almost always one finds the question and then one or more ‘principal arguments’ that are opposed to the position that the author will ultimately adopt. Then there are brief arguments for the other side (*in oppositum* or *sed contra*), followed by the ‘determination.’ This determination may begin with clarifications or definitions, then give the preferred answer, put forth objections to this answer and answer the objections. Finally, there are responses to the principal arguments put forth at the start.

In sum, medieval natural philosophy was predominantly studied in medieval universities, based on texts of Aristotle together with a series of later commentaries on the latter’s works. Medieval philosophy began by assuming the truth of the general view of the cosmos advocated by Aristotle, but over time natural philosophers modified this view in many respects, although without overturning it. When the natural works of Aristotle first became available in Latin in the early thirteenth century, they were rejected, most notably at Paris, because of their conflicts with Christianity. Before long, however, Aristotle’s natural works became widely-used textbooks, and their incompatibilities with Christianity provided the grist for repeated disputations and efforts to discover where the truth lay. By the end of the Middle Ages, a few scholastics composed their own treatises on natural philosophical topics (for example John Dumbleton’s *Summa logicae et philosophiae naturalis (Summa of Logic and Natural Philosophy)*, but these almost always dealt with Aristotelian themes (see Logic, medieval).

6 The Aristotelian cosmos

Medieval natural philosophy was therefore predominantly a series of more or less far-reaching variations on the theme of Aristotelian natural philosophy. Basing themselves on Aristotle, with a lesser reliance on Plato, medieval natural philosophers assumed that the natural world or cosmos is finite and spherical, with a spherical earth at the centre surrounded by spherical shells of water, air and fire (with some mixing, especially at the boundaries, such as the surface of the earth). The moon and all the rest of the heavenly bodies are contained in spherical shells of a fifth element called aether, turning in uniform circular motion about the centre of the cosmos, at which the earth is located. Technical astronomy, in order to account exactly for apparent motions of the planets, sun and moon, modified this idea by allowing additional aether spheres that rotated about centres other than the center of the cosmos. Such additional spheres were called eccentrics and epicycles.

For Aristotle, the cosmos ended with the outermost aether sphere (*primum mobile*), which contained the stars and rotated at a rate of approximately once per day, carrying all the other heavenly bodies with it. Outside the *primum mobile* was nothing, not even empty space. Medieval theologians and authors such as Dante in his *Divine Comedy* reconciled this Aristotelian picture with Biblical accounts of heaven and hell by placing hell in the center of the earth and by adding a last, unmoved sphere, the empyrean heaven containing the souls of the blessed, outside the *primum mobile*, and fourteenth-century theologians also supposed that God was outside the cosmos. Medieval technical astronomers also sometimes added an extra empty sphere to account for the precession of the equinoxes, the apparent motion relative to the stars of the points at which the sun crosses the celestial equator. The *primum mobile* could then be thought to be an empty sphere rotating approximately once a day, inside of which would be the sphere of the fixed stars, carried around by the *primum mobile* outside it but with its own proper rotation in the opposite direction about a different axis, with a period of about 26,000 years.

Following Plato and Aristotle, medieval natural philosophers assumed that everything in the cosmos (or everything except the movers of the heavenly spheres) is composed fundamentally of prime matter, which has no characteristics of its own, and forms, which give characteristics to matter. The technical term for an individual thing existing in the cosmos was ‘substance’; for example, a human being is a substance, a tree is a substance, a stone is a substance and so on. Each substance has a substantial form that makes it what it is or gives it its so-called ‘essential’ characteristics. For example, the substantial form of human beings makes them rational animals: anything that is not rational or not an animal is not human. Each substance also has many so-called ‘accidental’ forms such as colours or temperatures. These accidental forms happen to characterize a substance, but

are not essential to it, as it is not essential for a human being to have blue eyes or even five fingers on each hand.

Following the ideas of Aristotle's *Physics*, medieval natural philosophers also took it for granted that all or most things in the cosmos move, understanding by the term 'motion' not only local motion but also alteration (or change of quality) and augmentation and diminution (or change of size). The task of the natural philosopher was to explain or give the causes why the bodies in the cosmos move or change as observed. Thus medieval natural philosophy tried to answer such questions as, why do the sun, moon and stars move through the sky as they appear to move? Do animals move themselves, or is there an external cause of their motion; for instance, the food they see in front of them? How are metals formed? How do animals reproduce? How does nutrition occur?

The sorts of causes that natural philosophers looked for were usually forms, either substantial forms or accidental forms. In the terrestrial realm, from the centre of the earth up to the sphere of the moon, there were four elements - earth, water, air and fire - and many compound bodies formed by the interaction of these elements, including minerals, plants, animals and humans. Each of the terrestrial elements was thought to have two basic qualities: earth is cold and dry, water is cold and wet, air is hot and wet, fire is hot and dry. If one asked, 'Why does water evaporate and become steam or air?' the answer might be that fire, which is hot, heats the water so that its coldness is changed to warmth, with the result that the cold and wet water becomes hot and wet air. When they considered that an element such as water has a substantial form that makes it water as well as basic qualities that make it cold and wet, then natural philosophers might debate whether it is the heat of the fire or the substantial form of the fire that heats the water. Assuming further that the first result of heating water is hot water, natural philosophers might try to understand more exactly when and how the hot water changes its substantial form to become air. The usual answer was that the substantial form does not change until the very end of the process (when, in our terms, evaporation occurs): a less common answer was that, in reality, water and air have the same substantial form, differing only in their accidental qualities.

When the elements interacted to form minerals, plants or animals, questions about changes of substantial form became more serious. Gold, for instance, was assumed to have a substantial form that made it gold as well as being generated by a combination of earth, water, air and fire or by interactions of hot, cold, wet and dry. One not uncommon Platonic or Neoplatonic model was that the terrestrial elements interacted up to a certain point and then influences from the celestial realm, the planets and stars, led to the introduction of the substantial form into the prepared matter. For those thinking in a specifically Christian context, the idea was that God sometimes chooses to act using the heavenly bodies as instruments. In the case of humans, the rational soul might be supposed to be introduced into the embryo in this way, leading to 'quickening' and noticeable motion within the womb. This would be the point in time at which the embryo became a human being. This whole scheme provided a rationale for astrology, the study of the configurations of the heavenly bodies to predict a person's nature and likely fortune, since the exact configuration of the heavens at the time of the introduction of the rational soul would likely affect its nature at least accidentally.

In considering the whole realm of compound bodies on the surface of the earth, medieval natural philosophers followed Aristotle and Greek medical theorists in assuming that there are levels of souls or faculties active in living things, where a 'soul' is understood as a source of motion or activity in the substance or thing (see Hippocratic medicine). Plants have a 'nutritive soul' that explains their growth and reproduction; animals have a 'sensitive soul' that explains their motion and sensation as well as their growth and reproduction; and humans then have a 'rational soul' that explains their ability to think as well as all of the lower functions. It was a matter of disagreement whether each living thing has only one substantial form, which is the source of all its motions or actions, or whether the higher living things have forms that make them bodies and also souls. In more physiological or medical contexts, philosophers also made use of the concept of 'spirit' (*pneuma*), with the idea that the soul, conceived as pure form, uses spirit, conceived as the thinnest possible gas-like stuff, to move the body, whether in nutrition, sensation or local motion.

In the inanimate realm (things without souls of any sort), forms were also used to explain the natural motions of the elements and compound bodies. The elements earth and water were understood to be heavy, that is, to have a form called 'gravity' which leads them to move downward whenever this is not prevented or whenever they are not as close as possible to the centre of the cosmos. Likewise air and fire were supposed to be characterized by lightness or 'levity' as a form, leading them to move upward whenever they are below their natural places and not

hindered from moving higher. Thus it might be an internal characteristic (form) of a heavy or light body that explained its downward or upward natural motion, toward or away from the centre of the cosmos. The cause of gravity was not ‘attraction’ between bodies across empty space, as it would later be understood following Newtonian physics. For an Aristotelian, there is no empty space, and, even if there were, there could be no action at a distance without any intermediary. From the Aristotelian point of view, the only cases of apparent action at a distance between inanimate bodies were the attraction of magnets for iron and electrostatic attraction or repulsion, as between amber and chaff. Even these phenomena were explained without supposing action at a distance between inanimate bodies; instead, bodies were supposed to multiply forms of a sort (so-called ‘species’) into the medium surrounding them and up to the other bodies on which they might act.

Occasions on which elements, or compounds in which given elements were predominant, did not move in their natural directions (heavy bodies downward and light bodies upward) were also explained by medieval natural philosophers along Aristotelian lines, arguing that there must be an external force causing these elements or compounds to move in an unnatural direction and that this external force must be in contact with them. This idea followed obviously from everyday experience: loaded carts, for instance, do not move along the road unless some person or animal is in contact with them and pulls or pushes them. In the later books of the *Physics*, Aristotle had broadened this common conception to argue that everything that is moved is moved by something else (*omne quod movetur ab alio movetur*). In an apparent inconsistency with what he said elsewhere about souls being sources of motion in substances that have them, Aristotle argued in the *Physics* that living things and even elements in their natural motions have outside movers. Medieval Christian Aristotelians sometimes resolved this apparent inconsistency by concluding that only human beings, because of their free will, can act independently of outside forces.

In the celestial realm, it was supposed to be natural for aether to move in a circle around the centre of the universe. Apparently there was nothing to resist this motion since, according to Aristotle, it had been going on forever with no change of velocity whatsoever. An ox pulling a cart becomes fatigued because of the resistance to its pulling and, after a time, slows down. The celestial movers, however, must experience nothing similar since their motion continues unretarded for time without end. Because what happens on earth and what happens in the heavens are quite different, the explanation for the motion of the heavenly spheres is different from the explanation of terrestrial motion. To make a rough comparison, the movers of the heavens were supposed to move the celestial spheres in the way that hay causes a horse to move toward it, that is, by arousing in the moved body the desire to move. Unlike hay, however, the movers of the heavenly spheres were supposed to be immaterial, as Aristotle had argued they must be if they are to remain unmoved themselves; they must be so-called ‘intelligences’ or ‘separated substances’ (separated, that is, from matter). In more thoroughly Christianized versions of these theories, the celestial intelligences were sometimes identified as angels. In a view that was condemned by the Bishop of Paris in 1277, the celestial spheres were supposed to have their own souls, united to their matter, by which they perceived the separated substances and were therefore motivated to rotate. While this view of souls of the spheres helped to complete the analogy between the cause of their rotation and the cause of the motion of animate beings such as animals and humans, it was opposed to the Christian conception of the cosmos; if it were true, it would seem to imply that the celestial spheres were the most perfect beings in the cosmos, more perfect than humans because their matter was more perfect, although they were alike in having rational souls.

However, if in the Aristotelian cosmos the heavens were thought to be very different from the terrestrial realm below the sphere of the moon, the two realms were not disconnected. Influences from the heavens upon the earth were supposed to occur constantly, through the agency of light or other similar forms. In the Aristotelian explanation of light, a light source (*lux*) does not send out particles or anything else material; rather, the light source affects the medium immediately surrounding it, producing in it an illumination (*lumen*), which in turn causes the next part of the medium to light up, and so on. Thus no matter is moved in the spreading of light, but rather a form reproduces itself in decreasing degrees in successive contiguous parts of a transparent medium. In this model of light, other forms (for example, the form that causes magnetism) might be thought to reproduce themselves in contiguous bodies. This process, called by the generic name ‘the multiplication of species’, occurred constantly throughout the cosmos. In this usage, the word ‘species’ referred to forms that copy the form that causes them while not being entirely identical or not having identical power, like images in a mirror (*speculum*). From the time of the Greeks, optics, or the geometrical study of the ways in which light propagates itself, had been highly

developed, including studies of reflection, refraction and the decrease in intensity with distance. In the later Middle Ages, the model of optics was widely presupposed in the understanding of how other forms, such as heat and cold, might act on neighbouring bodies. Some alchemists believed that the substantial form of gold might under proper circumstances multiply itself in a similar way, transforming other metals into gold. Pharmacists likewise assumed that drugs act on the body not through the material of the drug incorporating itself into the human body (as happens with food), but through the substantial and accidental forms of drugs heating, cooling, moistening or drying the tissues with which they came into contact, or by the drug multiplying some other 'species' or effect of a specific form into parts of the body.

The medieval Aristotelian cosmos, then, was almost the opposite of the universe of the ancient atomists (see Atomism, ancient). Whereas for the atomists the universe is infinite, for Aristotelians it is finite. Whereas for the atomists the universe is made up of particles alike in their matter separated by empty space, for Aristotelians bodies differ from each other intrinsically by their substantial and accidental forms, not only by their size, position and motion; moreover, empty space is impossible. The atomists further theorized that fundamental particles have a size but cannot be cut into pieces ('atom' means uncuttable in Greek), while for the Aristotelians there is no limit to divisibility; no matter how finely something has been divided, it can still be divided further. This was expressed by saying that matter is potentially but not actually infinitely divisible; in other words there is no limit to the number of times one can divide a body into smaller and smaller pieces (this is the potential infinity), but one could not actually finish dividing the body infinitely many times (to finish an infinite process was considered to be a self-contradiction).

7 Self-consciousness of philosophers and the philosophic life

This, then, was the basic view of the cosmos accepted by medieval natural philosophers. From about the middle of the thirteenth century, when Aristotle's books were adopted as the textbooks for natural philosophy, philosophers began to clarify and suggest modifications in the received Aristotelian view. Sometimes obscure even in Greek, these texts had become far more difficult in the process of translation, especially when there had been an Arabic intermediary between the Greek and Latin; the differing vocabularies and grammatical structures of Arabic and Greek required many modifications in order to translate as literally as possible from Greek to Arabic, and with similar problems again occurring during the later translation from Arabic to Latin.

The difficulty of the Aristotelian texts was further compounded when one asked whether or how they fit with Christian faith. What should a person teaching natural philosophy do with the texts in which Aristotle ostensibly proves that the world has always existed, when Genesis says it was created? There were also the commentaries of Averroes who, although he was a monotheist, could not be expected to adhere to specifically Christian beliefs (see Ibn Rushd). Commentators on Aristotle often held degrees in arts but not in theology (though they might simultaneously be students in a faculty of theology). Lecturers on philosophy were not authorized to discuss specifically theological points in the classes of the arts faculty, but they were taken as authoritative on purely natural issues.

In these circumstances, Boethius of Dacia at the University of Paris argued that a philosopher should always resolve every philosophical difficulty solely from the point of view of philosophy. In his view, this was a matter of disciplinary integrity. Aristotle had said in the first book of the *Physics* that a geometer need not argue with someone who denies his principles or axioms; if a person denies one of the axioms of geometry, it will be impossible to prove something to him using those axioms or anything better known than the axioms. So too, Aristotle said, a physicist need not argue with someone such as Parmenides, who denies that motion occurs. That all or some things move, Aristotle argued, is the most obvious fact of physics, so that if a person denies that things move, it will be impossible to find a more obviously true principle on which to build a physical proof that motion occurs. In a similar way, each discipline has fundamental principles that are not proved but are taken as obvious for the purposes of that discipline, and on the basis of these principles, the practitioners of that discipline reason to conclusions. When people are practising a given discipline, they can and should assume that its principles are true.

Arguing on this basis, Boethius of Dacia claimed the right to conclude, when he was working in natural philosophy, that the world had no beginning because, rationally speaking and accepting the principles of Aristotelian natural philosophy, Aristotle proves that the world, if it exists at all, must be eternal or perpetual: something does not come from nothing; something is not caused by nothing and so on (see Eternity of the world,

medieval views of). On the other hand, as a Christian, Boethius of Dacia believed that the world was created by God as told in the book of Genesis; but, in his opinion, this did not affect the cogency of Aristotle’s philosophical argument for the eternity of the world. When Boethius was speaking as a natural philosopher, he was speaking not as a Christian but as someone who accepted the principles of natural philosophy.

In the furore that these views caused among Boethius of Dacia’s contemporaries, some thought that he could not and should not assert, even when doing natural philosophy, something that contradicted Christian faith and that he should be disciplined for teaching such a heresy. In reply, Boethius tried to explain why he should be able to assert without blame, when acting as a philosopher, conclusions contradicting Christian beliefs that he himself shared. In the excitement of defending his position, Boethius added to his armament many claims about the joys of the philosophical life, such as that no one lives a higher or happier human life than the philosopher.

In opposition to Boethius of Dacia, Thomas Aquinas developed a position allowing for a synthesis or merging of philosophical and Christian views. Arguing that Boethius’ position amounted to claiming that there can be two truths that contradict each other, the truth of philosophy and the truth of Christianity, Aquinas emphasized that there is only one truth and that true philosophy cannot contradict true Christianity. In the view of knowledge proposed by Aquinas, the truths of philosophy and the truths of faith are partially overlapping and partially separate. There are some things we know only by reason, some things we know only by faith and some things we know both by reason and by faith. That there are four terrestrial elements, for instance, we know only by natural philosophy and experience of the physical world, and that God is a trinity of Father, Son and Holy Ghost, we know only by faith; but that God exists, as the creator, mover and highest good of the universe, we know both by reason and by faith. Thus Aquinas proposed several cosmological and teleological arguments, generally founded on natural philosophy, which he claimed prove rationally the existence of God (see God, arguments for the existence of).

With regard to Aristotle’s supposed proof of the eternity of the world, Aquinas adopted the strategy of Maimonides’ *Guide to the Perplexed*, arguing that Aristotle’s proof is not demonstrative. Though Aristotle may have thought he had proved that the world is eternal, the argument could be refuted on rational grounds by pointing out that creation is after all supposed to be a supernatural rather than natural act. But, neither can it be proved philosophically that the world was created a finite time in the past. Thus for Aquinas the question of the eternity or newness of the world is philosophically neutral or undecided. Since philosophy leaves the question open, Aquinas could determine on the basis of faith that the world is created, with no inconsistency with natural philosophy.

Thus Aquinas, like Boethius of Dacia, argued that philosophy has an important role to play for the Christian as well as for the non-Christian. By dint of long study, Aquinas became one of the most knowledgeable and skillful Aristotelian philosophers of his day. On the basis of his synthesis of Aristotelianism and Christianity, other thinkers, prominently Dante in his *Divine Comedy*, developed a view of the cosmos that included both Aristotelian and Christian elements.

However, Aquinas’ proposal of a Christian-Aristotelian synthesis did not convince everyone, nor did it survive for long unmodified after his death. In the 1270s, the Bishop of Paris issued two condemnations which included theses put forth by Aquinas as well as by Boethius of Dacia (see Aristotelianism, medieval). In the later of these condemnations, a primary target was those natural philosophers who claimed that even God could not do what was impossible according to Aristotelian physics. If Aristotle proved that a vacuum cannot exist, so said these natural philosophers, then even God cannot make a vacuum. To the contrary, implied the condemnation of 1277, God is not limited by what is possible or impossible according to Aristotelian natural philosophy.

In arguing that the world cannot have had a beginning in time, Aristotle had adopted what has been called the statistical interpretation of modality or of possibility and necessity. On this interpretation, what is impossible never happens, what is necessary is always the case, and what is contingent or possible is sometimes the case and sometimes not. This implies that what is possible must sooner or later occur. As Maimonides reported in his *Guide to the Perplexed*, the *mutakallimun*, in defence of monotheistic religion, had adopted a different conception of possibility, making the possible not what sometimes happens, but what can be imagined without contradiction (see Maimonides).

In the Latin West, John Duns Scotus was the first to make use of the concept of ‘logical possibility’ (*possibilitas logica*), saying that what is contingent is not what is neither necessary nor impossible, but rather such that its opposite could have happened at the very same time that it did. Something is possible if it is not self-contradictory, whether or not it ever exists in reality. Following along these lines, fourteenth-century natural philosophers and theologians often discussed counterfactual cases according to imagination (*secundum imaginationem*). Even if there are no vacuums in the cosmos, they asked, what would happen if God made a vacuum and if a projectile were placed within it: would it move infinitely fast, with a finite velocity, or not at all? Or what would happen if God rotated the cosmos as a whole? Since there would then be no fixed point of reference, how could the rotation of the cosmos be distinguished from rest? In such discussions *secundum imaginationem*, natural philosophers often made a distinction between God’s absolute power (*potentia Dei absoluta*) and God’s ordained power (*potentia Dei ordinata*). God’s ordained power referred to the ways in which God has chosen to create and sustain the universe, as pictured in twelfth-century cosmogonies making use of Plato’s *Timaeus*, with its image of the Demiurge or Craftsman who forms an orderly cosmos out of a pre-existing chaos. God’s absolute power, on the other hand, covered everything that God might have chosen or done, even though God has not so chosen in fact. God chose to make a single cosmos with the earth in the centre, but God could have made more than one cosmos. Although there are no vacua in the universe, if God had wanted to create vacua, God could have done so.

In this distinction between God’s absolute and ordained powers, natural philosophy was understood to focus on the world as it has been formed by God’s ordained power. On the other hand, arguments *secundum imaginationem* or *de potentia Dei absoluta* might be used to help clarify what things are and are not distinct, or independent of each other. So, William of Ockham might argue, if motion or quantity are really distinct things, then God can create motion without any moving body, or quantity without anything that is extended or quantified. If God cannot cause motion to exist without a mobile (the very concept being self-contradictory), so Ockham might argue, it follows that the word ‘motion’ does not refer to anything other than substances or qualities. In this way, arguments *secundum imaginationem* involving God’s absolute power were used to support nominalism or a minimalist ontology without necessarily implying that God actually puts his absolute power into effect as imagined.

Thus in the late medieval understanding, God establishes but is not limited by the laws of nature. The only things that God could not do are those whose existence would be a self-contradiction. God could not make a square circle, because a square circle is logically impossible. God can and does, however, create accidents that do not inhere in any substance: this is physically impossible, but not logically or supernaturally impossible. Although naturally there are never accidental forms where there is no substance, such as whiteness where there is nothing to be white, in the Eucharist God miraculously causes the accidental forms of the bread and wine to persist without inhering in any substance (the substances of bread and wine having become the body and blood of Christ, without Christ’s body and blood having taken on the accidental forms of bread and wine).

The significance of all this for the status of natural philosophy was that nearly everyone in the later medieval university accepted a view that had elements of the views of both Boethius of Dacia and Thomas Aquinas. Aristotelian natural philosophy was understood to describe the cosmos as established by God’s ordained power. One could understand that within the cosmos as it currently exists, creation is impossible, and so Aristotle’s argument for the eternity of the world is valid given the laws of nature as they are currently established. On the other hand, the creation of the world as recounted in Genesis could be understood as an event that occurred not within the laws of nature as currently constituted, but rather supernaturally as the effect of God’s absolute power, the occasion upon which God’s ordering of the world was established. Likewise, miracles could be understood as departures from the laws of nature; which departures, as such, did not show that these laws of nature do not exist. God is the first cause of everything that happens, but, except for creation and miracles, events also have secondary causes that act instrumentally according to the patterns described by Aristotelian natural philosophy.

Thus later medieval natural philosophers found a *modus vivendi* in which they were free to follow the dictates of reason and observation in further developing their natural philosophy. They developed a number of methods and theories that were impressively successful, so that fourteenth-century theologians often adopted the methods of the faculty of arts for their own purposes. At the same time, later medieval theology not infrequently posed new concerns which natural philosophers subsequently took into account. If, according to the theologians, God could create a vacuum *de potentia absoluta*, then natural philosophers considered what implications the existence of a vacuum might have for the basic laws and concepts of Aristotelian natural philosophy. If in the Eucharist there are

qualities not inhering in any substance, then natural philosophers took this possibility into account. If, according to faith, God sends messages to earth using angels, then natural philosophers assumed that angels or separate substances were included among the types of entities that exist and whose properties may be discussed by natural philosophy (in fact, they supposed, the Aristotelian prime movers of the heavenly spheres are angels, even if Aristotle does not call them by this name).

Although in this way there was general agreement among the scholastics about how natural philosophy should operate, the results were by no means monolithic. If all agreed that, on the whole, the cosmos consists of substances made up of matter and form, they did not all agree what should be said, for instance, about the matter of the heavens as opposed to that of the sublunar elements. On each of the many problems that could be raised concerning Aristotelian physics, alternative solutions were proposed. University officials did not require natural philosophers to adhere to a single party line. The Dominicans might favour a position different from that of the Franciscans, and students and masters not belonging to any of the mendicant orders might hold still other positions.

8 Medieval variations on Aristotelian natural philosophy: new methods

Students in the faculty of arts in medieval universities perhaps spent more time studying logic than any other discipline, and it was from logic that later medieval natural philosophy adopted many of its most widely successful and transforming methods (see Language, medieval theories of; Logic, medieval). Because modern science uses mathematics rather than logic as its predominant tool (leaving aside experiment and data collection), historians of science have rarely recognized or understood how logic could play such an important and transformative role in medieval natural philosophy.

Put simply, medieval grammar, logic, natural philosophy and metaphysics together developed scholastic Latin to the point where it could map or model the world as exactly as mathematics does in modern science, albeit without the flexibility to describe very great complexity. This was especially the case for those logicians who took a nominalist or Ockhamist point of view, asserting that of the Aristotelian categories, only substance and quality refer to things existing outside the mind (see Nominalism). The basic idea was that words in propositions stand for objects in the physical world, while the syntax of the proposition indicates exactly the relationships between those objects represented in the proposition's assertion. Thus in one view, in the proposition, 'All humans are featherless,' the word 'humans' stands for all and only existing humans, the word 'featherless' stands for all and only existing featherless things, and the syntax of the sentence asserts the claim that every one of the things for which the word 'humans' stands falls also within the group of things for which the word 'featherless' stands.

This general strategy, known as the theory of the 'supposition' of terms in propositions, was then elaborated to apply to the exact interpretation of the propositions of natural philosophy. Many words, the so-called 'categorematic terms,' were supposed to stand for entities falling within the ten Aristotelian categories, especially for substances and qualities in the outside world or for concepts in the mind. Other words important to natural philosophy were explained to be 'syncategorematic terms' or logical operators, words that do not themselves stand for objects in the external world but rather affect the way that other terms appearing after them in the same proposition stand for things in the world. A given syncategorematic term was then associated with a paired 'exposition', in which a single proposition was shown to be expoundable into two or more propositions, the terms of each of which stood for things in the external world in a more standard way than did the terms of the original proposition.

Thus in the proposition, 'Socrates begins to be white', the term 'begins' was explained to be the sort of term that requires exposition, and the proposition was said to mean either (1) Socrates is not now white and immediately after this he will be white, or (2) Socrates is now white and immediately before this he was not white. Then, depending on the characteristics of the predicate term 'white', the natural philosopher/logician could determine which of the two expositions was the correct one. So-called 'permanent' characteristics, like whiteness, were said to have a first instant of being, but no last instant of non-being at their start, while so-called 'successive' characteristics, like running, were said to have a last instant of non-being at their start, but no first instant of being.

Natural philosophy, then, was understood to be exact talk about the things in the external world and their motions

and changes, and the propositions of natural philosophy were formulated and interpreted to make this exact description possible. In the case of the nominalists (William of Ockham, John Buridan, Albert of Saxony and others), this was associated with a great clarity about and minimization of the number of types of things that were supposed to exist in the cosmos. Following Aristotle's *Categories*, the world was supposed to be filled with individual substances, each of which could be qualified by various accidental forms. All the categories except substance and quality (and possibly also quantity) were supposed to involve only manners of speaking or modes of existence of the primary substances and qualities, not additional entities in the universe; thus time, place, relation, position and the rest were not separate things in the universe that natural philosophy had to deal with. All propositions that referred to time, for instance, could be expounded into propositions involving only substances and qualities and perhaps quantities. Aristotle had defined time as 'the number of motion'. Scholastic natural philosophers following Aristotle could say that, since this is what the word 'time' means, time exists only if there is a body in motion and a mind observing that motion and measuring it or assigning a number to it. (In this view, numbers also are not independently existing things in the universe, but require a mind counting substances.) Understanding that this is what time is could serve in part to solve the problem of whether the cosmos had a beginning in time. The cosmos has existed for all time, it might be said, because when there was no cosmos but only God in existence, there was no motion and hence no time. Time begins when God creates the moving cosmos, and it exists fully only if there is a mind numbering that motion (see Cosmology; Time).

A similar precision concerning the supposition of terms in propositions might be used to help solve problems of the possibility that there is a vacuum or empty space outside the cosmos, because 'space' in Aristotelian physics is an apparently significant term that in fact does not and cannot stand for anything to which its definition applies. 'Space', as used by the atomists, is supposed to be an extension in which there is nothing, but according to the medieval understanding of Aristotle, extensions can only be extensions of things; of substances or, in the special case of the Eucharist, of qualities. If it was supposed that God, according to God's absolute power, could create an empty space, that it was not a contradiction in terms, then it was typically concluded that, even so, there would be no extension there, or only an imaginary extension, because only bodies are extended. Even if God was supposed to be in infinite empty space outside the cosmos (as Thomas Bradwardine and Nicole Oresme both suggested), this did not necessarily imply that there would be extension there because God was understood to be wholly present at every point, not extended with one part in one place and another part in another. Just as God's existence 'before' the creation of the cosmos did not mean that there was time before creation, so God's existence 'outside' the cosmos did not mean that there was real space or extension there.

If the most widespread methodological innovation of medieval natural philosophy was the application of the precision of supposition theory, natural philosophers did also make progress in the application of mathematics. In an influential application, Thomas Bradwardine, writing at Oxford in 1328, found a new way to apply the Euclidean theory of ratios as it had been applied to the understanding of musical harmony, to describe the ratios of forces, resistances and velocities in motions. This innovation was so pleasingly apt, appearing very simple yet overcoming important drawbacks of the more authentically Aristotelian view, that it was adopted by nearly everyone. Paired with this innovation in the understanding of the relation of motion to its causes came progress in the exact mathematical description of motion in terms of its effects in the traversal of space, heating, change of size and so on. Corresponding to distances traversed in space, the concepts of 'latitude' and 'degree' were developed to measure intensive changes or the so-called intension and remission of forms. In later decades, the disputations on sophismata that had previously been used at Oxford to demonstrate students' abilities to apply the rules of logic were expanded to involve the new uses of mathematics as well as logic or, more rarely, grammatical theory. From this came the work of authors such as Richard Kilvington, William Heytesbury, Richard Swineshead and the other Oxford Calculators (see Oxford Calculators). Absolutely typical of later medieval natural philosophy and theology was the widespread use of what have been called 'analytical languages' or 'conceptual algorithms', that is, standard patterns of analysis repeatedly used to solve almost any problem. These 'analytical languages' included the Bradwardinian theory of the ratios of velocities in motions, the analysis of first and last instants, and analysis in terms of maxima and minima.

In their commentaries on Aristotle's *Posterior Analytics*, medieval natural philosophers were clear that the starting point for natural philosophy is experience. One starts from observation and works back to find the causes of the effects observed, demonstrating that (*quia*) the causes exist. Natural philosophy as a demonstrative science was

supposed to be complete when the philosopher could start from the causes and demonstrate why (*propter quid*) the effects as originally observed occur. However, if medieval natural philosophers believed in empiricism, they did not believe it was their role to actually conduct observations, let alone experiments. In practice, what medieval natural philosophers did was to try to perfect Aristotelian natural philosophy, making the meaning of its propositions more exact through logic and arguing the merits of various interpretations and modifications of Aristotle through formal public disputations.

9 Medieval variations on Aristotelian natural philosophy: traditions and research programmes

Our current knowledge of medieval natural philosophy has been built up in the twentieth century, for the most part through the combined efforts of historians of science, theology and philosophy. In the last twenty years, important work relevant to the history of medieval natural philosophy has also been done by historians of logic. From the perspective of the history of science, the original motivation, following the groundbreaking work of Pierre Duhem (1985), was to show how the roots of many of the ideas of early modern science could be found in the Middle Ages. For example, the medieval idea of impetus was seen as leading to the early modern idea of inertia, fourteenth-century analyses of uniformly accelerated motion were seen to foreshadow Galileo's analysis of balls rolling down inclined planes (see Galilei, Galileo), and medieval ideas of God's omnipresence were seen to have contributed to Newton's conception of absolute space (see Newton, I.).

In contrast to historians of science, some historians of theology have taken an interest in the history of medieval natural philosophy insofar as it interacted with theology, contributing, for example, to the medieval Christian-Aristotelian synthesis and to the subsequent threat to this synthesis posed by the Copernican thesis that the earth rotates and orbits the sun (see Aristotelianism, medieval; Copernicus, N.). In telling the story of medieval natural philosophy, some historians have taken Aquinas' Christian-Aristotelian synthesis as the high-water mark of medieval Christian philosophy, portraying the nominalist path as a falling away from or dissolution of this synthesis. On the other hand, some historians of the Reformation have taken William of Ockham as the hero of medieval natural philosophy, and have portrayed his nominalism or preference for a minimalist ontology as a fortunate paring away of the overdevelopment of previous scholastic thought. The theological position known as 'voluntarism' has been seen as equally salutary. While Aquinas thought he could reason how the cosmos must be on the grounds that God, as the highest good and all powerful, must have done things in the best possible way, the voluntarists argued that God had a choice, that there were many more things God could have done by his absolute power than he chose to do by his ordained power. If, according to the voluntarists, humans cannot reason what God must have done, they must instead observe the world as it actually is to determine how God in fact chose to create it. Thus, it is suggested, voluntarism in theology led to empiricism in science, and Ockham led to Copernicus and Galileo.

By the sixteenth century, natural philosophers often described their predecessors as followers of the realist way (*via realium*) or followers of the nominalist way (*via nominalium*), but thirteenth- and fourteenth-century natural philosophers less commonly labelled themselves according to philosophical schools. On the other hand, the Dominican and Franciscan orders ran their own houses of study, and a Dominican or a Franciscan friar would likely have studied a somewhat different curriculum before appearing on the scene at Oxford, Cambridge or Paris. Early on, the Dominicans adopted Thomas Aquinas as their intellectual model and so upheld the theory of the unity of the substantial form, that each substance is composed of a substantial form inhering directly in prime matter and that there are no intermediary forms, such as Averroes' 'indeterminate dimensions' (*dimensiones interminatae*) preparing the matter to receive a substantial form.

Some recent historians question the validity or usefulness of categorizing medieval thinkers as nominalists, terminists or Ockhamists. They have also claimed for earlier and less well-known thinkers many of the ideas previously associated particularly with Ockham, and have questioned whether nominalism was as important in the fourteenth century as earlier historians believed. Similarly, some recent historians have questioned the usefulness of the term 'Averroist,' previously applied to medieval natural philosophers to signify either philosophers who followed Averroes' opinion on certain questions - for instance, the question of the nature of the human soul - or philosophers such as Boethius of Dacia who chose to emphasize a purely natural or secular approach without obvious Christian modifications (see Averroism). The usefulness of labels such as 'nominalist' or 'Averroist' varies depending on where one's attention is directed, whether to issues of logic, psychology or theology.

The organization of medieval natural philosophy into schools and trends is also made more complicated by the fact that it was not common for medieval philosophers to claim originality. Someone such as Roger Bacon, who was quick to criticize his contemporaries for what he considered their ineptness, was the exception rather than the rule. Not only did scholastics generally not claim originality, but there were a number of scholastic practices that tended to further obscure their differences. To begin with, the same tools of logic and mathematics - the same analytical languages - could be used by individuals of very different theological persuasions. In commenting on Aristotle, many authors claimed that they were only describing and not advocating the opinions of Aristotle and Averroes. They said that they had recited the arguments for a particular point of view for the sake of (their own or the students') exercise, and not because they necessarily held them to be persuasive, and they often argued that more than one position on a given issue might be 'probable'; that is, one could develop a respectable argument on either side of a disgreement, and no one could demonstrate that one and only one solution was correct. Some texts existed as handbooks to prepare students for obligatory disputations, so they include the source arguments for alternative positions that a student might try to defend.

Sometimes church leaders advocated a sort of decorum so that clergymen would not be seen by the laity to disagree, but this did not apply in general to discussion of natural philosophy within the university because it could be assumed that these would not spill over into public consciousness. Within the university, then, scholars did disagree over natural philosophy issues, but this might not attract much attention unless the given issue had repercussions for theology and individual thinkers were much more likely to fall into camps over differences of opinion in theology than in natural philosophy.

In sum, because natural philosophy was rarely a life or death matter in the medieval university and because the answers to many questions really did remain open, many individual natural philosophers could present views that were somewhat tentative or that evolved over time. On the other hand, if a modern historian follows the discussion of a particular point over time, there is often a definite trend and progress in the solutions given, with some previous difficulties resolved to general agreement. Some of the most prominent commentaries on Aristotle have been studied, but many anonymous or poorly labelled commentaries on Aristotle's natural works still lie unstudied in manuscript form. Historical conceptions of the trends in medieval natural philosophy are likely to continue to evolve. John Duns Scotus seems to have been a pivotal thinker for many natural philosophical topics, even though his influence was exerted through commentaries on Peter Lombard's *Sentences* rather than through commentaries on Aristotle. Walter Burley has been labelled an Averroist and a realist in contrast to Ockham's nominalism, but Burley's position remains unclear, perhaps because there is no modern group such as the Dominicans or Franciscans that takes him as a special ancestor. Depending upon the perspectives of modern historians, new images of medieval natural philosophy can be expected to emerge in the future.

See also: Albert the Great; Alchemy; Aquinas, T.; Aristotelianism, medieval; Averroism; Bacon, R.; Bradwardine, T.; Copernicus, N.; Cosmology; Galilei, Galileo; Eternity of the world, medieval view of; Language, medieval theories of; Logic, medieval; Mechanics, Aristotelian; Medieval philosophy; Neckham, A.; Oresme, N.; Oxford Calculators; Religion and science; Science in Islamic philosophy; Scientific method

EDITH DUDLEY SYLLA

References and further reading

1 Representative primary sources in translation

Aquinas, Thomas (1257-9) *Expositio super librum Boethii De trinitate (Commentary on Boethius' De trinitate)*, trans A. Maurer, *Faith, Reason and Theology: Questions I-IV of his Commentary on the De Trinitate of Boethius*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1987.(A central medieval expression of the organization of the sciences.)

Aquinas, Thomas (1266-73) *Summa theologiae*, vol. 10, *Cosmogony*, ed. and trans. W.A. Wallace, New York: McGraw-Hill, 1967.(An example of how medieval natural philosophy appears in a theological context.)

Aquinas, Thomas (1271) *De aeternitate mundi (On the Eternity of the World)*, trans. C. Vollert, L. H. Kendzierski and P.M. Byrne, Milwaukee, WI: Marquette University Press, 1964.(The eternity of the world was perhaps the most debated question of thirteenth century natural philosophy in conjunction with theology; this was Aquinas' contribution.)

Bacon, R. (before 1292) *De multiplicatione specierum (On the Multiplication of Species)*, ed. and trans. D.

Lindberg in *Roger Bacon's Philosophy of Nature*, Oxford: Clarendon Press, 1983.(The multiplication of species, or action of a body on its surroundings, was a key concept in medieval natural philosophy.)

Bernard of Tours (*c*.1174) *Cosmographia*, trans W. Wetherbee, *The Cosmographia of Bernardus Sylvestris*, New York: Columbia University Press, 1973.(A good example of twelfth-century natural philosophy in the form of mythography, as associated with the cathedral of Chartres.)

Boethius of Dacia (*c*.1275) *De summo bono (On the Supreme Good)*, trans. J.F. Wippel, *On the Supreme Good, On the Eternity of the World, On Dreams*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies,1987.(Expresses Boethius of Dacia's high regard for the philosophic life.)

Boethius of Dacia (*c*.1275) *De aeternitate mundi (On the Eternity of the World)*, trans. J.F. Wippel, *On the Supreme Good, On the Eternity of the World, On Dreams*, Toronto, Ont.: Pontifical Institute of Medieval Studies, 1987.(Boethius' views on creation.)

Bonaventure (1250-73) *De aeternitate mundi (On the Eternity of the World)*, trans. C. Vollert, L. H. Kendzierski and P.M. Byrne, Milwaukee, WI: Marquette University Press, 1964.(Bonaventure's contribution to the debate on the eternity of the world.)

Maimonides, Moses (*c*.1190) *Dalalat al-Ha'irin (Moreh Nevukhim, Guide to the Perplexed)*, trans. S. Pines, Chicago, IL: University of Chicago Press, 1963.(Good example of the sorts of ideas from Islam and Judaism available in Latin from the mid-thirteenth century, *c*.1230-5.)

Oresme, Nicole (1377) *Le livre du ciel et du monde (The Book of the Heaven and Earth)*, ed. A.D. Menut and A.J. Denomy, trans. A.D. Menut, Madison, WI: University of Wisconsin Press, 1968.(Oresme's translation into French and commentary on Aristotle's *On the Heavens*; shows both what fourteenth-century Aristotelian commentaries were like and how a commentary on Aristotle might be modified for a lay audience.)

Oresme, Nicole (*c*.1351-55?) *Tractatus de configurationibus qualitatum et motuum (Treatise on Configurations of Qualities and Motions)*, ed. and trans. M. Clagett, *Nicole Oresme and the Medieval Geometry of Qualities and Motions*, Madison, WI: University of Wisconsin Press, 1968.(Represents the peak of fourteenth-century development of a science combining natural philosophy and quantification.)

Siger of Brabant (*c*.1272) *De aeternitate mundi (On the Eternity of the World)*, trans. C. Vollert, L. H. Kendzierski and P.M. Byrne, Milwaukee, WI: Marquette University Press, 1964.(Siger's contribution to the debate on the eternity of the world.)

2 Secondary sources

Cadden, J. (1995) 'Science and Rhetoric in the Middle Ages: The Natural Philosophy of William of Conches', *Journal of the History of Ideas* 56: 1-24.(How twelfth-century natural philosophy reflects the audience addressed.)

Caroti, S. and Souffrin, P. (1997) *La nouvelle physique du xiv^{e} siecle (The New Physics of the Fourteenth Century)*, Biblioteca di Nuncius, Studi e Testi XXIV, Florence: Olschki.(Eleven articles related to such topics as natural philosophy and sophismata in the thirteenth century, the debate over Ockham's physical theories at Paris, God's absolute power, reaction and so on.)

Chenu, M.-D. (1968) *Nature, Man and Society in the Twelfth Century*, ed. and trans. J. Taylor and L.K. Little, Chicago, IL: University of Chicago Press.(Classic expression of twelfth-century world view.)

Clagett, M. (1979) *Studies in Medieval Physics and Mathematics*, London: Variorum. (Key studies of medieval natural philosophy from the point of view of the history of science.)

Courtenay, W. (1984) *Covenant and Causality in Medieval Thought*, London: Variorum.(Theological trends with an impact on natural philosophy.)

Crombie, A.C. (1953) *Robert Grosseteste and the Origins of Experimental Thought*, Oxford: Clarendon Press.(Classic statement of the argument that medieval thinkers developed an empirical epistemology.)

Davidson, H. (1987) *Proofs for Eternity, Creation and the Existence of God in Medieval Islamic and Jewish Philosophy*, Oxford: Oxford University Press.(A superb survey of Islamic and Jewish sources for the subjects listed.)

Dronke, P. (ed.) (1988) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press.(Essential compendium for the twelfth century.)

Duhem, P. (1913-59) *Medieval Cosmology: Theories of Infinity, Place, Time, Void and the Plurality of Worlds*, ed. and trans. R. Ariew, Chicago: University of Chicago Press, 1985.(Translation of selections from Duhem's *Système du monde*, Paris, 10 vols. A groundbreaking work for twentieth century interest in medieval cosmology.)

Flint, V.I.J. (1991) *The Rise of Magic in Early Medieval Europe*, Princeton, NJ: Princeton University Press.(An essential reminder that the context of medieval natural philosophy was shaped in part by belief in magic.)

Funkenstein, A. (1986) *Theology and the Scientific Imagination from the Middle Ages to the Seventeenth Century*, Princeton, NJ: Princeton University Press.(A stimulating series of insights into the ways in which ideas of medieval and early modern theology and science may have been connected with one another.)

Goddu, A. (1984) *The Physics of William of Ockham*, Leiden: Brill.(An assessment of the role of Ockham in medieval natural philosophy.)

Grant, E. (1981a) *Much Ado About Nothing: Theories of Space and Vacuum from the Middle Ages to the Scientific Revolution*, Cambridge: Cambridge University Press.(A survey in depth of medieval ideas of place and space.)

Grant, E. (1981b) *Studies in Medieval Science and Natural Philosophy*, London: Variorum.(Collected articles especially relevant to cosmology, the impact of the condemnations of 1277, the roles of John Buridan and Nicole Oresme in fourteenth century natural philosophy, and the longevity of the Aristotelian world view.)

Grant, E. (1994) *Planets, Stars and Orbs: The Medieval Cosmos 1200-1687*, Cambridge: Cambridge University Press.(See especially the catalogue of scholastic questions asked about the cosmos from 1200 to 1687 and its analysis.)

Grant, E. (1996) *The Foundations of Modern Science in the Middle Ages, their Religious, Institutional and Intellectual Contexts*, Cambridge: Cambridge University Press.(Argues that the translations from Greek and Arabic, the development of the universities and the activities of theologian-natural philosophers prepared the way for modern science.)

Kretzmann, N., Kenny, A. and Pinborg, J. (eds) (1982) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(In addition to chapters by Weisheipl, Grant, Sylla and Murdoch on natural philosophy, provides essential background in logic, metaphysics, epistemology and other philosophical subjects essential to natural philosophy.)

Maier, A. (1982) *On the Threshold of Exact Science: Selected Writings of Anneliese Maier on Late Medieval Natural Philosophy*, trans. S.D. Sargent, Philadelphia, PA: University of Pennsylvania Press.(A small selection from Maier’s fundamental studies of late medieval natural philosophy published in German, 1952-77. Maier’s work is still the most important single authority on medieval natural philosophy.)

Molland, A.G. (1995) *Mathematics and the Medieval Ancestry of Physics*, Aldershot and Brookfield, VT: Variorum.(Collected articles especially on the relation of mathematics to natural philosophy.)

Moody, E. (1975) *Studies in Medieval Philosophy, Science and Logic: Collected Papers 1933-1969*, Berkeley, CA: University of California Press.(Represents pioneering work on medieval natural philosophy.)

Murdoch, J.E. (1969) ‘Mathesis in Philosophiam Scholasticam Introducta. The Rise and Development of the Application of Mathematics in Fourteenth Century Philosophy and Theology’, in *Actes du Quatrième Congrès International de Philosophie Médiévale*, Montreal, Que.: Institut d’Études Médiévales, 215-54.(How medieval natural philosophers introduced quantification into their work.)

Murdoch, J.E. (1975) ‘From Social into Intellectual Factors: An Aspect of the Unitary Character of Medieval Learning’, in J.E. Murdoch and E.D. Sylla (eds) *The Cultural Context of Medieval Learning*, Dordrecht: Reidel, 271-339.(Describes in some detail the conceptual algorithms or analytical languages used throughout much of fourteenth-century philosophy and theology and connecting philosophy to theology through common methods.)

North, J. (1984) ‘Natural Philosophy in Late Medieval Oxford’, in J.I. Catto and R. Evans (eds) *The History of the University of Oxford*, Oxford: Oxford University Press, vol. 2, 65-174.(Provides North’s angle on late medieval natural philosophy and the mixed sciences in general, given the difficulty of distinguishing what belonged to Oxford and what to other centres of intellectual activity. See also North’s article ‘Astronomy and Mathematics’ and other articles on theology and logic in the same volume.)

Pederson, O. (1953) ‘The Development of Natural Philosophy, 1250-1350’, *Classica et Medievalia* 14: 86-155. (An early attempt at synthesis.)

Schmitt, C. (1984) *The Aristotelian Tradition and Renaissance Universities*, London: Variorum.(Aristotelian natural philosophy in its Renaissance manifestations.)

Sorabji, R. (1983) *Time, Creation and the Continuum: Theories in Antiquity and the Early Middle Ages*, Ithaca, NY: Cornell University Press.(Wide-ranging discussion of the essential late Hellenistic background to medieval natural philosophy.)

Sylla, E.D. (1993) ‘Aristotelian Commentaries and Scientific Change: The Parisian Nominalists on the Cause of the Natural Motion of Inanimate Bodies’, *Vivarium* 31: 37-83.(A case study of how commenting on Aristotle worked as a form of scientific - or natural philosophical - practice.)

Tachau, K. (1988) *Vision and Certitude in the Age of Ockham: Optics, Epistemology and the Foundations of Semantics, 1250-1345*, Leiden: Brill.(Argues for the centrality of the optical metaphor in medieval theories of

knowledge.)

Weisheipl, J.A. (ed.) (1980) *Albertus Magnus and the Sciences: Commemorative Essays, 1980*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(Albert the Great is essential to understanding the nature of Aristotelian natural philosophy in the thirteenth century.)

Wilson, C. (1956) *William Heytesbury: Medieval Logic and the Rise of Mathematical Physics*, Madison, WI: The University of Wisconsin Press.(Demonstrates the interaction of logic and sophismata with medieval natural philosophy.)

Wolfson, H.A. (1929) *Crescas' Critique of Aristotle: Problems of Aristotle's 'Physics' in Jewish and Arabic Philosophy*, Cambridge, MA: Harvard University Press.(Classic study by a renowned scholar of a fifteenth-century Hebrew commentary on Aristotelian physics, useful as a comparison to Latin sources.)

Natural theology

Natural theology aims at establishing truths or acquiring knowledge about God (or divine matters generally) using only our natural cognitive resources. The phrase 'our natural cognitive resources' identifies both the methods and data for natural theology: it relies on standard techniques of reasoning and facts or truths in principle available to all human beings just in virtue of their possessing reason and sense perception. As traditionally conceived, natural theology begins by establishing the existence of God, and then proceeds by establishing truths about God's nature (for example, that God is eternal, immutable and omniscient) and about God's relation to the world.

A precise characterization of natural theology depends on further specification of its methods and data. One strict conception of natural theology - the traditional conception sometimes associated with Thomas Aquinas - allows only certain kinds of deductive argument, the starting points of which are propositions that are either self-evident or evident to sense perception. A broader conception might allow not just deductive but also inductive inference and admit as starting points propositions that fall short of being wholly evident.

Natural theology contrasts with investigations into divine matters that rely at least in part on data not naturally available to us as human beings. This sort of enterprise might be characterized as revelation-based *theology, in so far as the supernatural element on which it relies is something supernaturally revealed to us by God. Revelation-based theology can make use of what is ascertainable by us only because of special divine aid. Dogmatic and biblical theology would be enterprises of this sort.*

Critics of natural theology fall generally into three groups. The first group, the majority, argue that some or all of the particular arguments of natural theology are, as a matter of fact, unsuccessful. Critics in the second group argue that, in principle, natural theology cannot succeed, either because of essential limitations on human knowledge that make it impossible for us to attain knowledge of God or because religious language is such as to make an investigation into its truth inappropriate. The third group of critics holds that natural theology is in some way irrelevant or inimical to true religion. They argue in various ways that the objectifying, abstract and impersonal methods of natural theology cannot capture what is fundamentally important about the divine and our relation to it.

1 A strict conception of natural theology

An influential traditional understanding of natural theology is based on a model of theoretical inquiry derived from Aristotle's account of demonstrative science (see Aristotle §6). The model of demonstrative science describes the logical structure of the truths comprising a given area of theoretical inquiry and the corresponding epistemic structure of the beliefs one has when one has attained theoretical knowledge of that area. On this model, theoretical knowledge is constituted by two sorts of truths: those that are basic or fundamental and those that are dependent on them. Truths of the first sort are facts whose nature is such that they are open to our immediate cognitive gaze. Our knowledge of them consists in our having direct access to or awareness of them, and so they can be thought of as *evident*. They are self-evident if they are accessible to us in virtue of being immediately present to our strictly intellective capacities, or evident to sense perception if they are accessible in virtue of being immediately present to our sensory capacities. When we are directly aware of evident truths in either of these ways, we are epistemically justified in holding them just in virtue of our immediate grasp of their truth. Truths of this sort are epistemically privileged in the sense that one's epistemic justification for holding them is independent of one's justification for any other truths one might happen to hold. They can suitably ground other sorts of knowledge because they require no epistemic ground beyond themselves.

The second sort of truths constitutive of theoretical knowledge in demonstrative science are those derivable from the basic truths by means of demonstrative proofs or deductively valid arguments. These truths are accessible to us by virtue of our capacity for discursive reasoning. In so far as they are known as conclusions of demonstrative proofs they are epistemically dependent, mediated by our awareness of the truth of the relevant premises and the validity of the relevant inferences. We are inferentially justified in holding these truths by virtue of our having derived them in appropriate ways from truths we are independently and non-inferentially justified in holding.

Conceiving of natural theology as a kind of demonstrative science allows us to characterize it precisely: it consists

of truths about God which are either (1) self-evident or evident to sense perception, or (2) derived by deductively valid proofs the (ultimate) premises of which are self-evident or evident to sense perception. Thomas Aquinas (§14) is perhaps the most famous Christian theologian to develop Aristotle's account of demonstrative science and adapt it to theological inquiry. His famous proofs for the existence of God (the so-called five ways) and his arguments establishing such truths as that God is one, immutable and eternal - his proofs for what he sometimes calls the preambles of faith - are paradigms of natural theology understood as an Aristotelian demonstrative science (he himself prefers to call this enterprise divine science). Aquinas' influential articulation and development of this model of natural theology in the opening sections of his two great theological works, *Summa theologiae* and *Summa contra gentiles*, have made this a dominant conception of that enterprise.

Because its method - deductive argument - is clear and subject to precise and well-understood standards, intellectual inquiry that conforms to the model of demonstrative science is open to the sort of rigorous assessment and evaluation crucial to theoretical objectivity. Moreover, when successful, demonstrative science offers for its results the strongest sort of certification possible. By deriving its conclusions from evident truths by means of necessarily truth-preserving forms of reasoning, a demonstrative science guarantees their truth; we can be certain of them and can claim to have conclusively established them. But these virtues are purchased at a price: a demonstrative science's criteria for admissible data and methods are very strict. Indeed, few philosophers today are sanguine about the prospects of satisfying them outside the realms of logic and mathematics.

2 A broader conception of natural theology

Philosophers have always recognized the importance of arguments that fall short of the paradigm of demonstrative proof - that is, arguments that carry epistemic weight but in some way fall short of being utterly truth-guaranteeing. Following a long tradition, we might call arguments of this sort dialectical arguments. The criteria for good or acceptable dialectical arguments will be more generous than those definitive of demonstrative science. On the one hand, dialectic allows as data propositions that we have some appropriate degree of epistemic justification in believing, though they need not be evident either to reason or sense perception. On the other hand, dialectic admits the use of methods of reasoning other than deduction. It allows inductive reasoning broadly conceived - enumerative induction, probabilistic reasoning, argument from analogy, inference to the best explanation and so on. These requirements on the data and methods of dialectic mean that dialectical reasoning might fall short of demonstrative proof in any of three ways: by having evident premises but lacking a deductive argument form; by having a deductive argument form but lacking evident premises; or by lacking both evident premises and a deductive argument form. Of course these specifications of the data and methods of dialectic are imprecise, and philosophers might well disagree about what propositions are admissible as data (that is, have an appropriate degree of justification) and whether certain particular inductive arguments are strong or weak.

Our engaging in inquiry using only methods and data satisfying the criteria of dialectical reasoning requires us to use nothing beyond our natural cognitive resources. Hence, we ought to acknowledge a corresponding natural theology - dialectical natural theology - that consists in philosophical investigation, of the sort allowed by dialectic, into divine matters. We might call the genus of which strict, demonstrative natural theology and dialectical natural theology are species *broad* natural theology. The practitioner of broad natural theology will attempt to establish truths or acquire knowledge about God using deductive or non-deductive arguments the premises of which are evident or justified to some appropriate degree.

Two important contemporary projects in the philosophy of religion demonstrate some of the different ways natural theology might extend beyond the limits of the strict conception. In *The Existence of God* (1979), Richard Swinburne develops inductive versions of the traditional proofs for God's existence and treats theism as a theoretical hypothesis that claims to offer the best explanation for certain phenomena in the world. Swinburne shares with the strict natural theologian the aim of providing good arguments for the truth of certain theological propositions, in particular that the theistic God exists. But he rejects strict natural theology's requirement that limits the natural theologian to arguments satisfying the strict criteria of demonstrative proof, relying instead on techniques and standards derived from recent developments in the philosophy of science and the logic of probability.

By contrast, philosophers such as Alvin Plantinga and Nicholas Wolterstorff have objected to what they see as the epistemological assumptions underlying both strict natural theology and the work of many practitioners of broad

natural theology. Reformed epistemology, as they call it, challenges the view that establishing or defending the rationality of belief in God requires the development of arguments, whether strictly or broadly conceived, for certain theological propositions. They claim that it can be rational to believe in God even if one has *no* evidence for God's existence (see Plantinga and Wolterstorff 1983). Despite this seemingly bold denial of the need for evidence of the sort natural theology might provide, Reformed epistemology does not reject traditional natural theology's general aim of securing appropriate epistemic grounding for certain theological propositions; it rejects only its understanding of what can constitute appropriate grounding. Reformed epistemology allows that belief in God can be properly basic, that is, that the belief that God exists (and other religious beliefs) might appropriately be included among the starting points or data of natural theology. It is therefore a particular kind of defence of the epistemic rationality of our belief in God, and so a kind of broad natural theology (see Religion and epistemology §§1-3.)

Broad natural theology is not a new phenomenon. It has a long history and includes among its practitioners some of the best-known representatives of strict natural theology. Aquinas, for example, recognizes that some theological truths can be established by means of what he calls merely probable, persuasive or dialectical arguments. John Locke (§7), too, allows that reasoning about theological matters that depends solely on our natural faculties can depend on probabilities as well as on necessary reasoning.

3 Reason, revelation and faith

A strict conception of natural theology provides grounds for a clear distinction between natural and revelation-based theology. This is because revelation-based theology, in so far as it is based on divine revelation, is ultimately based on someone's testimony (God's, a divinely authorized agent's or that of historical witnesses to revelatory events). But no proposition whose epistemic grounding essentially involves testimony is either evident or strictly demonstrable. Hence, parts of revelation that are accessible only because they have been revealed are inaccessible to the strict natural theologian and belong exclusively to the domain of revelation-based theology. (The Christian doctrines of the Trinity and the Incarnation have been taken to be paradigm examples.)

Augustine (§11) and the long line of thinkers who have followed him in defining faith as belief based on authority have held that the contents of revelation are appropriate matters for faith and that the truths accessible only by revelation and not also by strict natural theology are strictly matters of faith and cannot be objects of theoretical knowledge or understanding. They have held, however, that natural theology can provide us with theoretical knowledge or understanding of some matters that can be held by faith. When we acquire a rational proof for a proposition previously held on testimony or come to see the theoretical explanation of it, our faith (with respect to that proposition) is thereby replaced with or converted to knowledge or understanding. Hence, to the extent that natural theology is successful we will have knowledge or understanding, but in so far as we must rely on revelation and testimony for truths about God we will have only faith (see Faith §4).

If we adopt a broader conception of natural theology, however, we lose the basis for a clear distinction between faith and knowledge or understanding. This is because, unlike demonstrative science, some forms of dialectical inquiry can make use of arguments that appeal to legitimate authority. Augustine points out that virtually everything we believe about the past, other parts of the world, and other people's thoughts and attitudes is based on someone's testimony. Since we typically suppose that we have good reasons for believing things of this sort (because we typically suppose we have good reason to accept the testimony on which those beliefs are based), it follows that propositions whose epistemic grounding essentially involves testimony may not be in principle inaccessible to the natural theologian who uses dialectic. Broad natural theologians might find themselves in a position to establish, solely on the basis of our natural cognitive resources (broadly construed), the legitimate authority of some body of revelation (for example, the historical reliability of the biblical accounts of extraordinary manifestations of divine activity). In that case, the truth of any proposition contained in that body of revelation, including those that must remain for the strict natural theologian strictly matters of faith, might be accessible to the broad natural theologian.

It might turn out, then, that with respect to their contents, there is no distinction between broad natural theology and revelation-based theology. That is to say, it might be that there is no territory open to the practitioner of revelation-based theology that is not in principle also open to the practitioner of broad natural theology. Even so, we might still distinguish between the two enterprises on epistemological grounds. Practitioners of broad natural

theology will accept a proposition only in so far as they take it to be an acceptable datum for natural theology or grounded by a suitable argument in it. By contrast, practitioners of revelation-based theology can accept a proposition simply because they believe it to be revealed, regardless of whether that belief or the proposition itself is adequately grounded for them.

4 Critiques of natural theology

Most philosophical attacks on natural theology are directed at particular arguments in natural theology - at one or another of the proofs for God's existence, for example. Objections to Anselm's ontological argument date back to Anselm's contemporary, the eleventh-century monk Gaunilo (see Anselm of Canterbury §4). René Descartes' proofs for God's existence found a contemporary critic in Thomas Hobbes (§3) (see Descartes, R. §6). David Hume (§6) offered a sustained critique of the argument from design popular among Enlightenment thinkers and developed his empiricist principles into an attack on the possibility of miracles and the evidential value of appeals to their alleged occurrence. The Humean project has been elaborated and extended in the twentieth century by the influential work of J.L. Mackie (1982).

Other critics have objected not only to the particular arguments of natural theology but also to the enterprise itself. Immanuel Kant (§8) argued that human knowledge is attainable only within the realm of possible human experience, and he concluded that since God is an object outside that realm, human knowledge cannot extend to divine matters. Hence, in so far as natural theology aims at attaining knowledge of the divine, it must fail. This rejection of natural theology, however, might be taken as relevant only to a certain kind of strict natural theology. In fact if Kant's own 'practical' arguments for the existence of God and the immortality of the soul can be understood as providing some epistemic justification for their conclusions (though not justification sufficient to give us knowledge strictly speaking), then they might be understood as arguments in dialectical (or broad) natural theology.

Some twentieth-century philosophers, following contemporary accounts of the nature of language, have developed analyses of religious discourse that undermine natural theology. Developing one strain of empiricism, A.J. Ayer (§2) and R.B. Braithwaite, for example, have defended non-cognitivist views of religious language. According to Ayer (1946), to say that God exists is to make a metaphysical utterance that is empirically unverifiable and so neither true nor false. Religious discourse, in so far as it purports to describe a transcendent reality, is therefore literally meaningless. Working in the same tradition, Braithwaite (1955) has argued that although religious beliefs are neither true nor false, they are expressive of a practical commitment to a certain religious way of life. On views of this sort, there are no propositions about God for natural theology to establish (see Religious language §3).

Following Wittgensteinian insights into the nature of language, philosophers such as Norman Malcolm (1977) have argued that religious discourse constitutes a particular language game. As with any language game, there can be no question about the truth, justifiability or legitimacy of the game itself. Questions about the truth or falsity of the beliefs that constitute a language game can be asked only from within the game itself. On this view, a natural theology that aims at providing ultimate grounding for religious beliefs is fundamentally misconceived.

Some critics have had religious grounds for scepticism about the very enterprise of natural theology. Søren Kierkegaard (§§4-5), for example, claimed that objective reflection on the truth of divine matters of the sort that is essential to natural theology necessarily fails to attain religious truth. According to Kierkegaard, the highest truth attainable by an existing individual is an infinite subjective passion that is itself paradoxical from the objective point of view. Natural theology will be not just useless but inimical to the attainment of religious truth. Developing both Kantian and Kierkegaardian objections, theologians such as Karl Barth (§§2-4) have claimed that the Christian God is wholly other and, as such, knowable only in virtue of God's own self-revelation. Barth (1960) concluded that any enterprise that relies solely on natural human resources cannot arrive at knowledge of God.

5 Natural atheology and apologetics

Natural theology is typically thought of as aiming at a positive result - that is, at proving that God exists, establishing that theism is true, or justifying religious belief in general or particular religious beliefs. Arguments in natural theology that are unsuccessful fail to achieve these positive results. But the truth about God (or divine matters generally) might be negative, that is, it might be that there is no God, that theism is false and so forth. So

we might think of arguments that have these negative results as also being part of natural theology. Alternatively, we might think of them as part of a distinct enterprise, natural *atheology*, that aims at establishing the falsity of theism or of certain sorts of religious belief.

Natural atheology has had two prominent strands. One strand argues that the concept of God is in some way incoherent or internally inconsistent. Some philosophers have supposed that the concepts of certain attributes commonly thought to be essential to God are incoherent, and that therefore no being possessing such attributes can exist. The well-known paradoxes of omnipotence - for example, that an omnipotent being must be able to create a stone too big for an omnipotent being to lift - purport to show that no being could be omnipotent (see Omnipotence §2). Others have maintained that certain of the divine attributes are incompatible with one another. It has been argued, for example, that no being could be both omniscient and immutable, since an omniscient being's knowledge must be constantly changing as facts about the world change (see Immutability §3).

The other main strand of natural atheology argues that the existence of God or the truth of some set of religious beliefs is incompatible with things we know about the world. By far the most famous atheological argument, the argument from evil, attempts to show that the existence of an omnipotent, omniscient, perfectly good God is incompatible with the existence of evil, or that the kinds and amount of evil we find in our world in some way make improbable the existence of the theistic God or render the theistic hypothesis less probable than some rival hypothesis (see Evil, problem of).

Defending theism against atheological attack has been the traditional role of apologetics. Defence might consist merely in showing that atheological arguments do not succeed in establishing their negative conclusions, or it might involve the more ambitious project of showing how the difficulties alleged against theism can be accommodated and explained by theism. The latter, more ambitious sort of defence against the argument from evil is theodicy - the attempt to justify God's goodness by identifying the morally sufficient reason that justifies God in causing or permitting the evil that in fact exists. Apologetics generally and theodicy in particular can be thought of as parts of natural theology.

6 Clarificatory and philosophical theology

Natural theology, whether strictly or broadly conceived, is justificatory in nature; that is, it is concerned with establishing the truth of certain theological propositions using standard techniques of reasoning starting from propositions that have some appropriate degree of epistemic justification. Many sorts of philosophical reflection, however, are primarily clarificatory rather than justificatory in nature. Philosophical reflection on and explication of the nature and content of a particular theory typically involves the analysis of the concepts central to that theory, the examination of the theory's coherence and internal consistency, and the assessment of the theory's relations to other theories and beliefs. These kinds of analytic and systematic tasks can be and often are carried out apart from questions about the theory's truth or our justification in accepting it; indeed the successful completion of some of these clarificatory tasks is in some ways a prerequisite for assessing the truth of a theory. We can therefore identify an enterprise whose aim is the philosophical clarification and explication of theories about divine matters or the particular beliefs that comprise such theories. We might call this enterprise *clarificatory* theology.

Clarificatory theology is like natural theology with respect to its methods: it uses standard techniques of reasoning (where those include tasks such as definition and conceptual analysis as well as techniques for constructing and evaluating arguments). But it is unlike natural theology in placing no epistemic restrictions on the sorts of propositions or theories that it can take as its starting points. Clarificatory theology might, for example, take the Christian notion of atonement as a datum and engage in philosophical clarification and explication of that notion without requiring that we justifiably believe or know that God exists, that an atonement has in fact occurred or anything of the sort.

If we think of the justification and clarification of religious beliefs and theories as two ways in which our natural cognitive resources can be brought to bear on divine matters, then we can take natural theology and clarificatory theology to be species of a single genus that we might call *philosophical* theology. Philosophical theology is the enterprise that attempts to understand divine matters using the techniques and methods of human reason, particularly as those have been developed within philosophy.

See also: Agnosticism; Atheism; God, arguments for the existence of; God, concepts of §§2-6; Mendelssohn, M.;

Religion, history of philosophy of; Revelation §§1-2

SCOTT MacDONALD

References and further reading

Anselm of Canterbury (1077-8) *Proslogion*, in *Anselm of Canterbury*, ed. and trans. J. Hopkins and H. Richardson, London: SCM Press, 1974, vol. 1.(Contains Anselm's famous presentation of the ontological argument; Gaunilo's reply *On Behalf of the Fool* is included in this volume, as is Anselm's *Monologion*, which contains his development of the concept of God.)

Aquinas, T. (1257-9) *Faith, Reason and Theology* (qq. I-IV of the commentary on Boethius' *De trinitate*), trans. A. Maurer, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(Aquinas' systematic reflections on the nature and methods of intellectual investigation of divine matters.)

Aquinas, T. (1259-65) *Summa contra gentiles* I, trans. A.C. Pegis, Notre Dame, IN: University of Notre Dame Press, 1975.(A classic reflection on and defence of natural theology.)

Aquinas, T. (1266-73) *Summa theologica* I, trans. Fathers of the English Dominican Province, Westminster, MD: Christian Classics, 1981.(Aquinas' most mature attempt at natural theology; contains the famous five ways.)

Augustine (392) *On the Usefulness of Believing*, trans. J.H.S. Burleigh, *Augustine: Earlier Writings*, Philadelphia, PA: Westminster Press, 1953, 291-323.(An argument for the epistemological and practical justification for believing certain propositions when one lacks conclusive proof.)

Ayer, A.J. (1946) *Language, Truth and Logic*, London: Victor Gollancz. (Ayer's famous statement of the verificationist principle of meaning and his application of it to metaphysical, ethical and religious language.)

Barth, K. (1960) *Anselm: Fides quaerens intellectum*, trans. I.W. Robertson, Richmond, VA: John Knox Press.(A manifesto for Barth's theological position, but controversial as an interpretation of Anselm.)

Braithwaite, R.B. (1955) *An Empiricist's View of the Nature of Religious Belief*, Cambridge: Cambridge University Press.(A fully developed account of a non-cognitivist theory of religious language.)

Descartes, R. (1641) *Meditations on First Philosophy*, ed. G.E.M. Anscombe and P. Geach, *Philosophical Writings*, Indianapolis, IN: Bobbs-Merrill, 1971.(Descartes' proofs can be found in the third and fifth of his *Meditations*. Hobbes' objections are published together with the *Meditations*.)

Hume, D. (1748) *Enquiry concerning Human Understanding*, ed. C.W. Hendel, Indianapolis, IN: Bobbs-Merrill, 1955, sect. X.(Hume's famous discussion of miracles.)

Hume, D. (1779) *Dialogues concerning Natural Religion*, ed. R. Popkin, Indianapolis, IN: Hackett Publishing Company, 1980.(An engaging dialogue focusing on the argument from design.)

Kant, I. (1781/7) *Critique of Pure Reason*, trans. N.K. Smith, London and Basingstoke: Macmillan, 1978, esp. A567/B595-A642/B670.(Kant's general critique of natural theology and his objections to the proofs for God's existence.)

Kant, I. (1788) *Critique of Practical Reason*, trans. L.W. Beck, Indianapolis, IN: Bobbs-Merrill, 1978.(Contains Kant's 'practical' arguments for the existence of God and the immortality of the soul.)

Kierkegaard, S. (1846) *Concluding Unscientific Postscript*, trans. D. Swenson and W. Lowrie, Princeton, NJ: Princeton University Press, 1941, esp. 177-88.(Kierkegaard's most detailed defence of the subjectivity of religious truth.)

Kretzmann, N. (1997) *The Metaphysics of Theism: Aquinas' Natural Theology in Summa contra gentiles I*, Oxford: Oxford University Press.(An excellent philosophical assessment of Aquinas' natural theology as presented in *Summa contra gentiles*, book 1.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P. Nidditch, Oxford: Clarendon Press, bk IV, ch. 18.(His position as outlined in §2.)

Mackie, J.L. (1982) *The Miracle of Theism: Arguments for and against the Existence of God*, Oxford: Clarendon Press.(A lucid, critical discussion of the main arguments for the existence of God.)

Malcolm, N. (1977) 'The Groundlessness of Belief', in S. Brown (ed.) *Reason and Religion*, Ithaca, NY: Cornell University Press, 143-57.(An analysis of the nature of religious belief relying on the Wittgensteinian notion of a language game.)

Morris, T. (1987) *Anselmian Explorations: Essays in Philosophical Theology*, Notre Dame, IN: University of Notre Dame Press.(A collection of essays developing perfect-being natural theology from an Anselmian perspective.)

Plantinga, A. (1991) 'The Prospects for Natural Theology', in J. Tomberlin (ed.) *Philosophical Perspectives 5: Philosophy of Religion*, Atascadero, CA: Ridgeview Publishing.(Plantinga's assessment of the project of natural theology in light of his religious epistemology.)

Plantinga, A. and Wolterstorff, N. (eds) (1983) *Faith and Rationality: Reason and Belief in God*, Notre Dame,

IN: University of Notre Dame Press.(A seminal collection of essays presenting the foundations of the anti-evidentialist movement in religious epistemology.)

Swinburne, R. (1979) *The Existence of God*, Oxford: Clarendon Press.(Defence of inductive versions of many of the major arguments for theism within the context of a cumulative-case argument that, given our total evidence, theism is more probable than not.)

Webb, C.C.J. (1915) *Studies in the History of Natural Theology*, Oxford: Clarendon Press.(An older but illuminating general study of different historical approaches to natural theology.)

Naturalism in ethics

Ethical naturalism is the project of fitting an account of ethics into a naturalistic worldview. It includes nihilistic theories, which see no place for real values and no successful role for ethical thought in a purely natural world. The term 'naturalism' is often used more narrowly, however, to refer to cognitivist naturalism, which holds that ethical facts are simply natural facts and that ethical thought succeeds in discovering them.

G.E. Moore (1903), attacked cognitivist naturalism as mistaken in principle, for committing what he called the 'naturalistic fallacy'. He thought a simple test showed that ethical facts could not be natural facts (the 'fallacy' lay in believing they could be), and he took it to follow that ethical knowledge would have to rest on nonsensory intuition. Later writers have added other arguments for the same conclusions. Moore himself was in no sense a naturalist, since he thought that ethics could be given a 'non-natural' basis. Many who elaborated his criticisms of cognitivist naturalism, however, have done so on behalf of generic ethical naturalism, and so have defended either ethical nihilism or else some more modest constructive position, usually a version of noncognitivism. Noncognitivists concede to nihilists that nature contains no real values, but deny that it was ever the function of ethical thought to discover such things. They thus leave ethical thought room for success at some other task, such as providing the agent with direction for action.

Defenders of cognitivist naturalism deny that there is a 'naturalistic fallacy' or that ethical knowledge need rest on intuition; and they have accused Moore and his successors of relying on dubious assumptions in metaphysics, epistemology, the philosophy of language and the philosophy of mind. Thus many difficult philosophical issues have been implicated in the debate.

1 Naturalism versus other issues

The focus in contemporary debates about ethical naturalism is on the relation of ethics to a naturalistic worldview, one that takes the emerging scientific picture of the world as approximately accurate and that rejects belief in the supernatural. Ethical nihilists deny that real values can fit into such a picture. Cognitivist naturalists reply that they can, and that we can learn about them by methods similar to those by which we learn of other natural facts. Noncognitivists disagree with cognitivists on both counts; they distinguish themselves from nihilists, however, by insisting that the primary function of ethical thought is not to discover ethical truths but rather to direct action, a task at which it may succeed (see Analytic ethics; Moral realism).

These debates need to be distinguished from others with which they are sometimes confused. The ancient Sophists initiated a discussion, for example, about whether morality exists by nature or by convention (see Sophists; Nature and convention). Although answers to this somewhat ill-defined question may bear on contemporary issues about ethical naturalism, the relation is not simple. That morality requires 'following nature' is sometimes a version of cognitivist naturalism, but sometimes not (as when 'following nature' is explained in theological terms). Many versions of cognitivist naturalism take no stand on this thesis, moreover, and others, such as versions of social relativism, appear to come down on the 'convention' side of the older debate (see Social relativism).

Questions about naturalism should also be distinguished from more general issues about 'facts and values' (see Fact/value distinction). One question is whether there are ethical facts. Cognitivist naturalists answer yes, but so did Moore and so do theological moralists. Noncognitivists such as R.M. Hare (1952), to be sure, reject cognitivist naturalism because they think values could not be facts of any sort, even if there were (or are) supernatural facts; and in this they can claim inspiration from Moore (1903), who introduced his peculiar category of the 'non-natural' to accommodate ethical facts precisely because he thought such facts could be neither natural nor 'metaphysical' (his term for the supernatural) (see Moore, G.E. §1). But others who applaud Moore's arguments against cognitivist naturalism, such as J.L. Mackie (1977), think there would be no bar to taking certain supernatural facts (for example, about Plato's forms), if there were such facts, to be ethical facts.

So there seems no thesis about whether there are, or could be, ethical facts, which uniquely distinguishes either generic ethical naturalists, cognitivist naturalists or their critics. Another question is whether, from premises entirely about nonethical facts, there are reasonable inferences to any ethical conclusions (see Logic of ethical discourse §§2-4). Here all critics of cognitivist naturalism answer no. (The doctrine they thereby affirm is called

the autonomy of ethics.) Many writers, again beginning with Moore, and including some defenders of cognitivist naturalism, have thought the latter doctrine committed to answering yes. Indeed, in the third quarter of the twentieth century, this question was often understood to be the central issue between cognitivist naturalists and their critics. For reasons which shall be noted, however, many cognitivist naturalists after that time have seen no need to deny the autonomy of ethics. Thus, on this issue, there is agreement among critics of cognitivist naturalism but none among its defenders.

2 Vindicating and debunking

We have distinguished ethical naturalism in the generic sense from significant subcategories: nihilist, cognitivist, noncognitivist. Within these latter two categories we can note one other important distinction. All versions of cognitivist naturalism will allow for discovering that some ethical statements are true, but they may nevertheless differ greatly on how important a ‘success’ they thereby make possible for ethical thought. For some of these views - for example, the suggestion that judgments of right and wrong merely report local social conventions - seem intended to represent ethical considerations as far less important and less worthy of respect than they have seemed; whereas others, such as ideal contractarian or Aristotelian views, seem designed not to debunk but instead to vindicate the practical importance conventionally ascribed to ethics. Positions at the debunking end of this scale thus join with nihilism in deflating the importance of ethical thought, though on the basis of a thesis about what makes ethical beliefs true rather than on the grounds that they are never true.

A similar distinction can be drawn within noncognitivism. Any noncognitivist theory will specify some role, such as guiding conduct, at which ethical discourse may succeed. But early versions, such as A.J. Ayer’s (1936), did not suggest that this would be success at anything rational or likely to command respect (see Emotivism). Later versions, however, such as those of Simon Blackburn (1993) or Allan Gibbard (1990), disown carefully any intention of representing ethical thought as defective or ‘second-rate’. Crucial to their case is the claim that, even on a noncognitivist understanding, ethical thought can systematically mimic most of the features - such as answerability to evidence and to requirements of consistency - that make it appear to aim at discovering ethical truths. This defence suggests a way, however, in which even the most vindicatory version of noncognitivism is likely to seem at a disadvantage when compared with a comparably motivated cognitivist naturalism, if the latter is truly available. For why settle for mimicry if one could have the real thing? The answer, according to these noncognitivists, is that the real thing is not available; and their reasons are mostly familiar from the tradition noted above, of thinking cognitivist naturalism vulnerable to objections of principle. So we need to turn to those objections.

3 Metaphysics

These arguments may for convenience be divided into the metaphysical and the epistemological, although the best-known of them, the ‘open question’ argument, through which Moore tried to expose the fallacy in all ‘naturalistic ethics’, draws a metaphysical conclusion from partly epistemic premises. Moore assumed, in what is arguably already an oversimplification, that any version of cognitivist naturalism would rest on a reductive property-identification such as ‘goodness = pleasure’. Further, he thought (1) that any such identification could be true only if the expressions flanking the identity-sign are synonyms for a speaker who understands them both, and (2) that the terms could be shown not to be synonyms by the mere fact that such a speaker could conceivably doubt the identity-statement: by the fact, in other words, that it remains in this minimal sense an ‘open question’ whether the identification is correct. Thus, from the mere fact that doubt about this sample doctrine is conceivable, Moore took it to follow that it is false. Even on these assumptions, it is unclear why he was confident that every version of cognitivist naturalism would fail his test; but his assumptions have in any case met severe criticism. Assumption (2) has been attacked as too stringent a test for synonymy, assumption (1) questioned on the deeper ground that synonymy seems in any case not required for the truth of identifications of this form, as illustrated by such significant discoveries as that water is H_2O or that heat is molecular motion.

Critics of cognitivist naturalism have asked in reply how good a model these scientific identifications could provide for any proposed in ethics. Certainly, there could be little resemblance if Gilbert Harman (1977) were right that ethical notions never seem useful in explaining any natural facts or events. For these scientific identifications are accepted partly because of an approximate match between, for example, the effects common sense explains by

reference to heat and the ones attributed to molecular motion. But cognitivist naturalists have replied to Harman that the explanatory use of ethical notions is endemic to ordinary thought. Social justice has long been conceived as a condition that, without undue deception or coercion, will stabilize a society and permit it to flourish; moral decency is taken to be a trait that keeps some people from doing things that come easily to others.

Another cluster of objections concerns the reason-giving and motivating roles of moral facts. Some critics deny that natural facts could possess the rational authority for action expected of ethical facts. Others, especially noncognitivists, deny that natural facts could motivate as ethical facts would have to do when recognized (see Moral motivation). To both challenges cognitivist naturalists typically offer a two-part response. First, any vindicatory version will attempt to show about at least some ethical values why, on a full understanding, they would command a strong allegiance from most reflective agents. Anyone who knows or assumes that such an account is available would tend to be motivated by ethical considerations, moreover.

Second, even naturalists who hope to vindicate much of ethics in this way will oppose attempts to dismiss debunking accounts prematurely and for the wrong reason. Conventional thought may accord ethical considerations rational authority, but as with many conventional beliefs about ethics it remains to be seen how well this assumption will withstand critical challenge. Plato has Thrasymachus in the *Republic* say that justice is nothing but the interest of the ruler; on a plausible reading, Karl Marx holds (to oversimplify considerably) that it is nothing but the interest of a ruling class (see Wood 1981). These are both on their face naturalistic but deliberately debunking accounts. It strikes cognitivist naturalists as too easy a victory over such disturbing suggestions to hold that they can be rejected a priori simply because they threaten the conventional assumption that justice merits respect. Critics sometimes maintain that at least the most general ethical categories (the good, the right) must be immune from such debunking challenges, but cognitivist naturalists argue that the history of ethical debate suggests otherwise. If they are right, then they can point to agents, those advancing the debunking positions, who are recognized as joining issue with more conventional views, but for whom these ethical considerations are not motivating.

4 Epistemology

Moore thought that ethical knowledge would have to rest on ethical intuition, a source unlikely to be welcomed by an ethical naturalist (see Intuitionism in ethics). But there is a simple argument for Moore's view from two premises that have been widely accepted by philosophers. One is that ethics is autonomous, that there is no reasonable inference to any ethical conclusion from entirely nonethical premises (see Autonomy, ethical §5). The other is a traditional and very influential doctrine about knowledge called foundationalism (see Foundationalism). According to foundationalism, everything we know to be true must either be based on a reasonable inference from other things we know, or else must be known directly, without inference; and all knowledge of the first sort must ultimately rest entirely on a foundation provided by the other, noninferential kind. Together these doctrines entail that if we have any ethical knowledge at all, some of it must be noninferential (or, in equivalent traditional terminology, intuitive). For if we have any ethical knowledge by inference, that knowledge must (by foundationalism) be obtainable by a reasonable inference from things we know without inference; and (by the autonomy of ethics) the premises of that inference must be partly ethical.

For much of the twentieth century, most cognitivist naturalists, when faced with this argument, agreed with their critics that they would have to deny the autonomy of ethics. But that doctrine has continued to seem more plausible than the arguments often proposed in its defence. Prominent defences of it have appealed to the controversial theses already mentioned, deriving it, for example, from the principle (itself not obvious) that action-guiding conclusions cannot be inferred from premises that severally lack this property. The resilience of the doctrine is arguably due less to such defences than to its being an instance of what has come to be recognized as a more general phenomenon: that when evidence from one area is brought to bear on a second, it typically has to be applied in light of assumptions one already has about the second area. Thus, to take an important example, it was a fond hope of empiricist philosophy of science to display standard scientific reasoning as proceeding from a foundation of empirical observation, guided only by methodological principles that are entirely neutral on scientific issues; but there has come to be near-consensus that this cannot be done. Scientific reasoning instead adjusts information about empirical evidence along with a diverse body of theory already tentatively in place, in the direction of a better overall fit. The process is quite similar to what John Rawls (1971) has called the search for

a ‘reflective equilibrium’ among our views at different levels of generality, although Rawls described the procedure only for ethics (see Moral justification §2).

One moral drawn from these considerations by many philosophers is that foundationalism is mistaken. Although there is controversy over what account of knowledge to substitute, moreover, the reasonable alternatives have in common that, even when conjoined with the autonomy of ethics, they do not require that ethical knowledge rest on intuition. One suggestion, for example, is that scientific reasoning can lead to knowledge because the process of progressive readjustment described above is generally reliable, in that it tends over time to eliminate error and bring the body of scientific beliefs closer to the truth; and that scientific reasoning has had this feature partly because enough of the theoretical beliefs from which it begins have been at least approximately true. Some cognitivist naturalists have defended a similar thesis about ethical reasoning; and their position no more requires noninferential knowledge of goods or duties than the comparable thesis about science requires intuitions of quarks or electrons.

Each of these objections to cognitivist naturalism can be further elaborated, as can the replies. If a fallacy is a mistake easily agreed upon once it is pointed out, there is no naturalistic fallacy. There remains, however, a complex debate about whether any form of cognitivist naturalism can prove satisfactory, a debate that quickly takes one into all the major areas of philosophy.

See also: Moral judgment; Moral knowledge; Value, ontological status of

NICHOLAS L. STURGEON

References and further reading

Many of these readings are demanding but, except as noted, are not especially technical.

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz; 2nd edn, 1946, ch. 6.(Presents one of the earliest explicit versions of noncognitivism.)

Blackburn, S. (1993) *Essays in Quasi-Realism*, New York: Oxford University Press, essays 6-11.(Defends a version of noncognitivism, here called ‘projectivism’, and of ‘quasi-realism’, the thesis that this understanding will largely vindicate rather than debunk ethical thought.)

Boyd, R. (1988) ‘How to be a Moral Realist’, in G. Sayre-McCord (ed.) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press, 181-228.(A defence of cognitivist ethical naturalism, especially important on the comparison of ethical to scientific epistemology; the volume includes an extensive bibliography, much of it relevant to debates about ethical naturalism.)

Foot, P. (1978) *Virtues and Vices*, Berkeley, CA: University of California Press.(Essays by a prominent defender of cognitivist naturalism. Chaps 7-9 question the autonomy of ethics, 10-13 its rational authority.)

Gibbard, A. (1990) *Wise Choices, Apt Feelings*, Cambridge, MA: Harvard University Press.(Perhaps the most carefully worked out version of noncognitivism, which extends that doctrine beyond ethics to cover judgments of rationality; a sympathy for ‘Moore-like’ dismissals of cognitivist naturalism is conjoined with the reflection that the author’s disagreement with that position may in the end be empirical. Pages 150-250 explore the extent to which, on the account presented, normative judgment ‘mimics the search for truth’ (218).)

Hare, R.M. (1952) *The Language of Morals*, Oxford: Clarendon Press.(A highly influential presentation of noncognitivism, which attempts to derive the autonomy of ethics from the distinctively action-guiding character of ethical language. Like most writers influenced by Moore, Hare uses the term ‘naturalism’ for what this entry calls cognitivist naturalism.)

Harman, G. (1977) *The Nature of Morality*, New York: Oxford University Press, chaps 1-2.(Presents the argument that supposed moral facts seem irrelevant to naturalistic explanation, without committing himself to this conclusion.)

Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin, ch. 1.(Makes a vigorous naturalistic argument for ethical nihilism.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press, sections 1-17.(Dense, technical and, by Moore’s later admission, sometimes confused, this passage presents the ‘open question’ argument and claims to refute cognitivist naturalism, called by Moore ‘naturalistic ethics’.)

Plato (*c*.380-367 BC) *Republic*, trans. G. Grube, revised by C. Reeve, Indianapolis, IN: Hackett Publishing Company, 1992, book I.(Presents a classic debate about the nature and value of justice.)

Railton, P. (1992) ‘Some Questions about the Justification of Morality’, in J.E. Tomberlin (ed.) *Ethics, Philosophical Perspectives* 6, Atascadero, CA: Ridgeview Publishing Co., 27-53.(A subtle discussion of how

vindicatory a naturalistic theory of ethics can be, and should be required to be.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Belknap Press, ch. 1, section 9.(Describes the process of seeking a ‘reflective equilibrium’ among one’s ethical and other beliefs. This is accepted as an approximately accurate account of thoughtful reasoning in ethics by many who then disagree about the philosophical implications.)

Sturgeon, N.L. (1985) ‘Moral Explanations’, in D. Copp and D. Zimmerman (eds) *Reason, Truth and Morality*, Totowa, NJ: Rowman and Allenheld, 49-78.(A defence of ethical naturalism, with special reference to the explanatory relevance of ethical facts.)

Wood, A.W. (1981) *Karl Marx*, London: Routledge & Kegan Paul, chaps 9, 10.(Defends a careful version of an interpretation according to which Marx holds (to put it much less carefully) that justice is nothing but the interest of a ruling class.)

Naturalism in social science

Naturalism is a term used in several ways. The more specific meanings of 'naturalism' in the philosophy of social sciences rest on the great popular authority acquired by modern scientific methods and forms of explanation in the wake of the seventeenth-century scientific revolution. For many of the thinkers of the European Enlightenment and their nineteenth-century followers the success of science in uncovering the laws governing the natural world was used as an argument for the extension of its methods into the study of morality, society, government and human mental life. Not only would this bring the benefit of consensus in these contested areas, but also it would provide a sound basis for ameliorative social reform. Among the most influential advocates of naturalism, in this sense, was the early nineteenth-century French philosopher Auguste Comte.

The authority of the new mechanical science, even as an account of non-human nature, continued to be resisted by romantic philosophers. However, the more limited task of resisting the scientific 'invasion' of human self-understanding was taken up by the Neo-Kantian philosophers of the latter part of the nineteenth century, in Germany. Followers and associates of this tradition (such as Windelband, Rickert, Dilthey and others) insist that there is a radical gulf between scientific knowledge of nature, and the forms of understanding which are possible in the sphere of humanly created meanings and cultures. This view is argued for in several different ways. Sometimes a contrast is made between the regularities captured in laws of nature, on the one hand, and social rules, on the other. Sometimes human consciousness and self-understanding is opposed to the non-conscious 'behaviour' of non-human beings and objects, so that studying society is more like reading a book or having a conversation than it is like studying a chemical reaction.

1 Varieties of naturalism

A 'naturalistic' approach to the social sciences may express a view about the nature and limits of our knowledge of the different subject matters, or about the appropriateness of certain methods of inquiry, or about the essential sameness of the subject matters of natural and social sciences. So, we may usefully distinguish epistemological naturalism, methodological naturalism and ontological naturalism. Also, in the case of ontological naturalism, it is important to distinguish between, on the one hand, 'reductionist' versions, and, on the other, 'non-reductionist' versions, such as 'emergent powers' or 'dialectical' naturalisms.

In most of the literature of the philosophy of social science, naturalism is taken to be primarily an epistemological doctrine, and, frequently, to be more or less identical with 'positivism' (see Positivism in the social sciences §1). Epistemological naturalism is the thesis that the social life of humans is knowable in just the same sense, and will take the same form, as knowledge of the natural world. So, the natural sciences - physics, chemistry and biology - are taken as paradigmatic for social scientific knowledge. Historically, the most influential accounts of the natural sciences have been empiricist ones, according to which scientific knowledge is characterized by its openness to empirical testing (generally by the use of experimental method), the law-like character of its general statements, the symmetry of explanation and prediction, and the clear distinction between value-judgements and factual statements. The nineteenth-century French sociologist, Auguste Comte is generally credited as the founding figure of a positivist movement, devoted to extending scientific knowledge, on this model, to the spheres of psychology and human social life. The positivists and their latter-day followers have also been committed to a view of the social sciences as technically applicable knowledge, in much the same way as knowledge in the physical sciences has enabled a degree of control and regulation of natural phenomena.

Critics of positivism have most commonly been adherents of epistemological or ontological anti-naturalism. However, the proliferation of non-empiricist accounts of the natural sciences, especially, in the Anglophone world, since the work of Thomas S. Kuhn (see Kuhn, T.S.) has opened up the possibility of versions of epistemological naturalism that reject positivism. 'Conventionalist' accounts of empirical testing serve to undermine the fact/value distinction as a distinguishing mark of science, while, various realist approaches to the nature of the natural sciences put the emphasis on the role of analogy and metaphor in the discovery of mind-independent causal mechanisms and structures in nature.

These non-empiricist accounts of natural scientific knowledge open up a complex range of possible forms that might be taken by the social sciences within the general terms of non-empiricist epistemological naturalism. The

work of Rom Harré and his associates in the field of social psychology, and of Roy Bhaskar in developing a 'critical realist' reconstruction of historical materialism are two notable developments along these lines (see Critical realism).

2 Methodological naturalism

Methodological naturalism may be defined as the commitment to applying natural science methodology for purposes of developing the social sciences. Again, methodological naturalism is most commonly associated with the positivist tradition. The central priority for such an approach is to render social processes and relations measurable and so mathematically analysable. It is widely accepted that the direct use of experimental methods is rarely possible in the social sciences - either for ethical or legal reasons, or because of the impossibility of controlling extraneous variables. So, methodological naturalists tend to opt for two main substitutes. One of these, typical of historical sociology, is the comparative method. Societies that differ in only a small number of significant features are compared with respect to some historical outcome (for example, the emergence of dictatorships or democracy) whose causes are sought. Alternatively, analysis may be carried out on large-scale data-sets to discover statistical associations (for example, between social class and morbidity or mortality) which may allow of causal inferences. Many of these techniques are common to both social and (some) natural sciences (such as epidemiology and population biology). Émile Durkheim's *Suicide* (1897) is a classic text for methodological naturalism in sociology (see Durkheim, É. §3).

Opponents of methodological naturalism commonly argue that the relation, in the social sciences, between researcher and subject matter is quite different from that which obtains in the natural sciences. Participants in human social life do so on the basis of their own understandings of what they are doing. The forms of understanding developed by social scientists are, therefore, secondary, 'understandings of understandings'. Social scientific understanding, on this view, is more like textual interpretation or, perhaps, dialogue, than it is like the external relation between the natural scientists and the physical structures and processes they study. The hermeneutic tradition that flourished in Germany in the late nineteenth century, and which returned to the philosophy of Immanuel Kant for ways to resist positivist 'invasion' of the humanities, has been the most powerful source of opposition to methodological naturalism. The British philosopher, Peter Winch, in his influential *The Idea of a Social Science* (1958), deployed his reading of the later work of Wittgenstein to reach very similar conclusions. The subsequent 'linguistic turn' in the social sciences, and the associated introduction of literary modes of social analysis under the banner of 'postmodernism' have served to put the proponents of methodological naturalism very much on the defensive.

Again, however, it may be argued that an identification of 'methodological naturalism' with comparative method and statistical analysis as experiment substitutes, relies on a narrow and distorted view of the range of methods of investigation used by natural scientists themselves. There are, for example, many parallels between the fieldwork methods of animal ethologists and the 'qualitative' research of social scientific ethnographers.

3 Ontological naturalism

However, the more fundamental objections to both epistemological and methodological naturalism tend to turn on matters of ontology. For anti-naturalists, humans, their symbolic creations, their social relations and institutional forms, are radically different in kind from the order of nature. In their different forms, anti-naturalisms may emphasize free choice and spontaneity (ruling out deterministic laws and scientific prediction), self-definition (cognitive authority of participants, not observers), rule-following (as against causal determination), or the creation of meaning (requiring interpretation, as distinct from explanation) as the key ontological feature that defeats the naturalist project. Sometimes the case for human-social ontological uniqueness is made by way of a dualistic opposition between physical mechanisms or processes, and those of the human social world, or, alternatively, some version of human/animal dualism may be relied upon.

The case for a dualistic opposition between humans and animals (what has come to be called, by its opponents, 'human exemptionalism') may rest on a range of supposedly unique features of human individuals, or of the species as a whole. The attributes of self-consciousness, language-use, and moral agency are most commonly emphasized, but especially in traditions of thought influenced by Marx and Hegel, such distinctive properties of human groups as social production (rather than mere collection), tool-use, and historicity may be emphasized.

Against this, ontological naturalists have available two broad strategies. One is to focus on those aspects of their lives which humans have in common with (other) animal species. The other is to rely on modern research in the life sciences that seems increasingly to have ‘closed the gap’ in just those areas of human ‘uniqueness’ which the anti-naturalists prize. The Darwinian revolution established the phylogenetic kinship between humans and other animal species, and set the scene for later approaches to the study of animal behaviour in the field, and of the ecology of the human species itself. Tool-use and the learning of elements of human sign-language among chimps, as well as widespread evidence of high levels of social cooperation, emotional complexity and a rich psychic life in other animal species are now difficult to deny.

Proponents of naturalism sometimes accuse anti-naturalists of an arrogant ‘species-chauvinism’, and point to the inability of ‘human-exemptionalist’ explanatory strategies to deal with the material dimensions of human ecological destruction. Advocates of ontological naturalism also point out how the reduction of human social life to ‘text’, which is implicit in the ‘linguistic turn’, obscures or renders marginal the material dimensions of human deprivation, oppression, and unnecessary suffering that has been at the heart of much critical social science.

But the anti-naturalists can, with some force, point out that, in its various forms, ontological naturalism in the social sciences has itself served oppressive interests. The Malthusian ‘law of population’, social Darwinism, Nazi racial science, and, in our own day, sociobiology have all been used as means of oppression. Naturalisms of various kinds have also been widely used to represent as ‘natural’ and therefore unalterable all manner of unjust and exploitative institutions.

4 Non-reductionist naturalisms

It may be that no single solution may be found to the dilemmas and tensions posed by this confrontation between naturalisms and their opponents. It is certain that there are no easy or simple ways forward. However, from the side of ontological naturalism a number of intellectual strategies are available which show some promise of answering the main objections made by anti-naturalists. It is quite possible, for example, within a broadly naturalistic framework, to acknowledge the specificity of the human species without denying our kinship with other species. Indeed, human powers of symbolic communication, autonomous agency and capacity for history can themselves be seen not as *sui generis*, but as consequences of our biological peculiarities - ‘premature’ birth, prolonged infantile dependency, bipedalism, highly developed central nervous system, and so on.

The social world humans create as their ‘second nature’ may also be understood as irreducibly heterogeneous in its make-up - some of it is, literally, text, while much more of it is text-like. Humans live all their social relations through some form of symbolic understanding, so that a hermeneutic aspect, or moment, is necessary to all social science. However, to suppose that human social life is composed wholly of understandings, as if there were nothing beyond discourse which these ‘understandings’ are understandings (or misunderstandings) of, would be to commit a linguistic reduction as difficult to defend as its polar opposite naturalistic reductionism.

For an advocate of naturalism to go this far in the direction of the anti-naturalist case would require a complex and open-textured ontology, incompatible with the forms of reductionist naturalism widely criticized by the linguistic and hermeneutic traditions. In other words, to be defensible against the main anti-naturalist critique, naturalism would need to avoid the ‘reductionist’ project of representing the subject matter of the social sciences as the direct or unmediated effect of the operation of laws and mechanisms proper to the natural sciences.

A non-reductionist naturalism, making use of the ideas of a hierarchy of more or less autonomous levels of organization of matter, each with its own, qualitatively new, ‘emergent’ powers or properties has been one fruitful way of sustaining the main insights of a naturalistic approach, without falling foul of what is valid in the anti-naturalistic critique.

Such hierarchical, ‘emergent powers’ ontologies enable their advocates to recognize in the various subject matters of the different natural and social sciences, more or less discrete and autonomous object-domains, while at the same time making no concessions to spiritualistic, vitalist, or supernatural beliefs. ‘Emergent powers’, or ‘dialectical’ naturalisms remain committed to the doctrines that the beings making up the object-domains of the various sciences are all ultimately composed of the same basic ‘stuff’, and that at whatever level of organization they exist, they behave consistently with the laws governing the lower levels of organization.

On this view, social (economic, political, and so on) structures and practices are held to exist independently of our (social scientific) knowledge of them, and as objects of study in their own right, with their own distinctive sets of properties, powers, tendencies, and so on. Nevertheless, and this is the hallmark of a naturalistic approach, these structures and practices, with their distinctive properties, powers and tendencies exist only in virtue of their status as more or less enduring forms of organization of the lower-level entities (people, buildings, tools, animals and plants), and so on down the hierarchy of organization.

Clearly, ideas such as 'emergent powers' enable people who want to work within a naturalistic framework to take account of the very powerful arguments used by anti-naturalists against reducive forms of naturalism (such as sociobiology and Malthusianism). To some extent this may lead to a convergence of views between non-reductive naturalisms and the less extreme forms of anti-naturalism. However, in research-practice in the social sciences, this difference of view tends to take the shape of a division of labour between different topics and sub-disciplines. So, for example, cultural anthropology and sociology of culture have tended to be dominated by anti-naturalistic approaches, while studies of social 'systems', power-structures, social class and stratification and so on, have more commonly been governed by epistemological and methodological, if not ontological naturalism.

See also: Dilthey, W.

TED BENTON

References and further reading

Andreski, S. (ed.) (1974) *The Essential Comte*, London: Croom Helm.(Translation of part of Comte's classic *Course in Positive Philosophy*, which outlines the case for a naturalistic social science.)

Benton, E. (1991) 'Biology and Social Science', *Sociology* 25 (1): 1-29.(Argues that sociologists should take account of biological insights, but can do so without becoming committed to naturalistic reductionism.)

Bhaskar, R. (1979) *The Possibility of Naturalism*, Hemel Hempstead: Harvester, 1989.(Argues for an epistemological naturalism that recognizes both methodological and ontological peculiarities of the human sciences.)

Catton, W.R. and Dunlap, R.E. (1978) 'Environmental Sociology: a New Paradigm', *American Sociologist* 13: 41-9.(Introduces the term 'human exemptionalism' to characterize anti-naturalistic sociology, and proposes an environmentally oriented sociology as an alternative to it.)

Dilthey, W. (1883) *Introduction to the Human Sciences*, trans. R.J. Betanzos, Hemel Hempstead: Harvester, 1988. (Another classic statement of an anti-naturalistic foundation for the social sciences.)

Durkheim, É. (1897) *Suicide: A Study in Sociology*, trans. J.A. Spaulding and G. Simpson, Glencoe, IL: Free Press, 1953.(A methodological classic, which uses official statistics to establish the role of social facts in the explanation of suicide rates.)

Kuhn, T.S. (1970) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(Major work, that did more than any other English-language work to undermine the authority of empiricist and 'positivist' views of the nature of scientific knowledge.)

Rose, S. (ed.) (1982) *Towards a Liberatory Biology*, London: Allison and Busby.(An influential collection of articles devoted to developing non-reductionist or 'dialectical' versions of biology.)

Rose, S., Kamin, L.J. and Lewontin, R.C. (1982) *Not in Our Genes*, Harmondsworth: Penguin.(Wide-ranging critique of naturalistic reductionism and attempt to develop an alternative naturalistic approach.)

Wilson, B. (ed.) (1970) *Rationality*, Oxford: Blackwell.(A classic collection of papers devoted to issues posed by anti-naturalist arguments, centring on the possibility of cross-cultural criteria of rationality, and the intelligibility of 'alien' cultures.)

Winch, P. (1958) *The Idea of a Social Science*, London: Routledge & Kegan Paul. (Another classic anti-naturalist work, which makes use of the later philosophy of Wittengenstein.)

Naturalized epistemology

The term 'naturalized epistemology' was coined by W.V. Quine to refer to an approach to epistemology which he introduced in his 1969 essay 'Epistemology Naturalized'. *Many of the moves that are distinctive of naturalized epistemology were made by David Hume, but Quine's essay fixes the sense of the term as it is used today.*

Naturalized epistemology has critical as well as constructive thrusts. In a critical spirit, 'naturalists' (theorists who identify with the label 'naturalized epistemology') abandon several assumptions that are part of the tradition. They reject Descartes' vision of epistemology as the attempt to convert our beliefs into an edifice resting on a foundation about which we have complete certainty. Descartes is wrong to equate knowledge with certainty, and wrong to think that knowledge is available through a priori theorizing, through reasoning which makes no use of experience. Nor should epistemology continue as David Hume's attempt to rest knowledge on an introspective study of the mind's contents. Moreover, the global sceptic's claim that there is no way to justify all our views at once, should either be conceded or ignored.

On the constructive side, naturalists suggest that in investigating knowledge we rely on the apparatus, techniques and assumptions of natural science. Accordingly, naturalized epistemology will be a scientific (and hence neither indefeasible nor a priori) explanation of how it is that some beliefs come to be knowledge. Issues of scepticism will be addressed only when they come up in the course of a scientific investigation.

Quine's seminal essay lays out the core of naturalized epistemology, but subsequent naturalists disagree on the appropriate responses to several issues, among them the following: First, may theories be tested on the basis of (independently plausible) theory-neutral observation, or are observations simply more theory? Second, after being naturalized, does epistemology survive as an autonomous discipline? Quine argues that epistemology should become a subfield of natural science, presumably a part of psychology, so that there is no separate field left specifically to philosophers. But can all our questions about knowledge be answered by natural scientists? Third, the claim that epistemology explains how knowledge comes to be suggests that epistemology will merely describe the origins of beliefs we take to be known; but what is the relationship between such descriptive issues and normative issues such as that of how we ought to arrive at our views? Fourth, to what extent is the new approach to epistemology susceptible to sceptical concerns such as those that so plagued traditional epistemologists, and how effective a response can be made to those concerns?

1 Hume's anticipation of naturalized epistemology

The naturalizing movement in epistemology is the continuation of Hume's rebellion against Descartes' view about knowledge. Like Descartes, Hume wanted to conduct an investigation of the mind and its operations, including 'the operations we perform in our reasonings'. But five features of Hume's approach place him far closer to contemporary naturalism than to Cartesianism.

First, while Descartes wished to leave no room for doubt, Hume explicitly took for granted the trustworthiness of the very faculties whose operations he wanted to investigate. Later naturalists (theorists who identify with the label 'naturalized epistemology') make a parallel move: they trust the techniques and assumptions of science even while investigating how (scientific) knowledge is possible. Hume assumed that our mental faculties are trustworthy because it would be pointless to attempt to test their accuracy; after all, any test required their use. In fact, it was *because* Hume supposed that the mental faculties generated knowledge (or at least rational belief) that he thought the clarification of their workings would shed light on the normative question of what an epistemic agent ought to believe. His project still had a critical edge, for when Hume found any belief that could not be satisfactorily accounted for in terms of standard human faculties, he recommended throwing it out as baseless.

Second, like later naturalists, Hume modelled his epistemology after the emerging natural sciences, where empirical confirmation served as the basis for claims. He thought that knowledge encompasses everything we can discover using all our mental faculties, including experience and what we can discover by applying our mental faculties to themselves.

Third, like contemporary naturalists, Hume was prepared to say that some knowledge is the product of purely causal mechanisms rather than reason (or reasoning), since through introspection Hume thought he could detect a

causal mechanism at work. The mechanism he detected itself produces knowledge of causal relations, on the basis of which we believe in matters of fact, which are facts that hold contingently and whose negations indicate real possibilities (see Hume, D. §2; Causation §1). Custom, prompted by experience, is the mechanism through which we form our suppositions concerning causal relations. Hume does not recommend doubt about the products of custom. After all, the mind is functioning in a healthy, normal way when it is under the influence of custom, and reasoning can begin only after custom does its work. We should begin to question our beliefs only when we find that they are arrived at while the mind is not functioning in the normal way science describes.

Fourth, like many contemporary naturalists, Hume explains some of the mechanisms responsible for knowledge (such as custom) in terms of survival value. The linking of causes to effects is so important to human survival that it would have been a mistake for ‘nature’ to entrust it to our reason ‘which is…extremely liable to error’. Better to entrust it to ‘some instinct or mechanical tendency, which may be infallible in its operations, may discover itself at the first appearance of life and thought, and may be independent of all the laboured deductions of the understanding’ (Hume [1748/51] 1975: 55). Years later, W.V. Quine and other naturalists will speak in the same vein. Quine (1974: 20) explains induction in terms of natural selection, and disavows any claim to have justified induction. ‘In the matter of justifying induction we are back with Hume, where we doubtless belong’.

Fifth, Hume is with contemporary naturalists in their reactions to scepticism. As he had to, since he put his trust in his faculties, Hume rejected Descartes’ idea that to know that our beliefs are true is to be in a position to place all our beliefs beyond doubt at once. Hume saw that it is not even possible to *justify* all our views at once. No more than Descartes could Hume use his faculties to assess his epistemic prospects without first assuming that they were reliable. In the last section of the *Enquiry Concerning Human Understanding* (1748/51), Hume mentions scepticism and simply acknowledges that truly global doubts would be ‘entirely incurable’.

2 Internalism versus externalism

Like Hume, contemporary naturalists are empiricists. For example, in ‘Epistemology Naturalized’ Quine explicitly affirms a version of Humean empiricism:

> It was sad for epistemologists, Hume and others, to have to acquiesce in the impossibility of strictly deriving the science of the external world from sensory evidence. Two cardinal tenets of empiricism remained unassailable, however, and so remain to this day. One is that whatever evidence there is for science is sensory evidence. The other…is that all inculcation of meanings of words must rest ultimately on sensory evidence.
>
> (Quine 1969: 75)

But contemporary naturalists do not have in mind the introspective sort of empiricism advocated by Hume but, rather, intersubjective empiricism. Instead of simply taking our mental faculties and the data of introspection for granted and accounting for knowledge from that perspective, Quine and other naturalists suggest that we take a more full-bodied version of scientific practice for granted, and account for knowledge from *that* perspective. The empiricist science on which naturalists rely is conducted in the public domain. Scientists use microscopes and other instruments to extend their senses, and the observations upon which they rely are not data of introspection but rather observations that are at least in principle publicly confirmable (see Empiricism; Introspection).

Naturalists think that the stimulation of sensory receptors helps to determine whether or not people know the truth of beliefs that are causally linked to those stimulations. Yet the stimulations themselves are usually not noticed by the people in whom they occur. Naturalists are *externalists*, defined by Laurence BonJour (1985), following D.M. Armstrong (1973), as theorists according to whom facts that are external to an agent’s conception of the situation can serve to justify that agent’s beliefs in a way that is sufficient for knowledge (see Internalism and externalism in epistemology). *Internalists*, by contrast, would insist that all knowledge is based on justifications that are in some sense in the cognitive possession of the knower. A related view is that epistemology, or knowledge about knowledge, is based on such justifications. Consider the following arguments for these internalist positions.

First, as BonJour emphasizes, justifications that are in no way possessed by an agent are completely arbitrary (unsupported), at least so far as the agent can tell from the agent’s own point of view, and accepting arbitrary beliefs is a bad idea from the standpoint of getting to the truth, which is the goal of the epistemic agent. Yet externalists claim that it is possible for agents to acquire knowledge through sources about which those agents

believe little or nothing. Their sources might be a causal chain, a reliable belief-formation process, an information channel, or all three (see Knowledge, causal theory of; Information theory and epistemology; Reliabilism). Internalists disagree, since it is epistemically irresponsible to believe something through some avenue without checking out the truth-conductivity credentials of that avenue.

The main problem with this first line of thought is that the internalists' assumptions seem to lead immediately to sceptical results. For internalists want an avenue to the truth but they want to accept nothing except what is justifiable, and nothing as justification except what is available to them 'from the inside'. That means they need an avenue to the truth that can be defended as such on the basis of what is available 'from the inside', a task that appears to be impossible. No empirical premise will serve in the defence since empirical premises will either be accepted without justification (arbitrarily) or justified on the basis of some other empirical premise, thus initiating a regress.

A second line of argument gives up on the attempt to show that all knowledge conforms to internalist assumptions, but purports to show that epistemology fits the internalist view: since logic is a priori accessible if any body of knowledge is, the thought that there is an inductive as well as a deductive logic may lead theorists to suggest that knowledge is the product of reasoning that conforms to valid deductive and inductive forms of argument. Some of the assumptions that are fed into the arguments might not conform to internalist assumptions, but the reasoning does. So we can understand epistemology to be an a priori study of the argument forms to which reasoning ought ideally to conform.

One problem here is that, unlike deductive logic, inductive logic is not the study of argument forms. Ultimately inductive logicians invoke appeals to the way the world is, appeals which internalists want to avoid. Except in trivial cases, that one statement makes another probable is a claim about the world, not about the forms of those statements. More seriously, in the phrase 'reasoning that conforms to valid argument forms' there is a conflation of argument with reasoning. As Gilbert Harman (1986) and others have emphasized, these are quite distinct things. The study of arguments and argument forms is logic; the study of reasoned belief revision is the study of the belief-management practices governing those revisions. Principles of belief management would tell us when epistemic agents should retain their beliefs and when and how they should revise them. The inference rules of logic do not tell us when revision is appropriate. Inference rules can tell us many things; for example, suppose that we believe the premises of an argument - inference rules can tell us that the premises of that argument entail its conclusion. But that is not the same thing as a recommendation that (say) we affirm the conclusion. In fact, belief-management principles might tell us that we ought to drop one of the premises rather than accept the conclusion.

Once we distinguish between logic and belief management, like Harman we may become quite sceptical about the existence of anything like an inductive logic. An inductive logic would be a logic that resembles deductive logic except that the conclusions of 'valid' inductive arguments are merely made probable by their premises, not entailed by them. That such a logic exists is not entailed by the existence of inductive reasoning, and the fact that so little progress has been made towards developing an inductive logic might lead us to suspect that there is no such subject to be investigated (see Inductive inference). Moreover, if inductive logic does not exist, good inductive reasoning cannot be simply reasoning that conforms to inductive logic. (Some have held that probability theory provides the basis of inductive logic - see Probability theory and epistemology.)

So the epistemic activity of inductive reasoning might not be illuminated by advances in 'inductive logic'. A related point is that much epistemic activity probably does not involve beliefs or any other proposition-related states at all. As Paul Churchland (1979: 128) has argued, a great deal of information is processed by infants, and their prelinguistic behaviour does not invite ascriptions of propositional attitudes to them. 'Rational… intellectual development in an infant cannot be…usefully represented by a sequence of sets of sentences suitably related'. Information processing in infants, as well as the processing by virtue of which sensory stimulations are assimilated by adult brains prior to the formation of beliefs, would presumably be articulated by principles of which epistemic agents themselves are entirely unaware and to which they could not purposely conform even if they wanted to. Churchland takes such facts to suggest that any account of knowledge in terms of 'sentences suitably related' is bound to be superficial; a proper theory would deal in more primitive parameters that apply to pre-linguistic processing, and would account for the way epistemic agents deal with beliefs as a derivative case.

In sum, critics of naturalized epistemology who say that knowledge or epistemology is based on justifications in the cognitive possession of knowers face some severe difficulties.

3 The role of observation

Naturalists disagree on many points, among them the role of observation in knowledge, the degree to which natural science is separable from epistemology, the role of normativity in epistemology, and the significance of scepticism (these latter three points are discussed in subsequent sections). One view about the role of observation in knowledge, that of Quine, is that ultimately the 'evidence for' or 'test of' our theory of the world is the prediction of the stimulations of exteroceptors. The components of our view that are most directly tested by these stimulations are observation sentences such as 'It is raining' or 'There is a rabbit'. These are directly tested by stimulations because what each *means* is a set of stimulations that prompts assent to and a set that prompts dissent from the sentence. The test of our theory of the world is therefore its ability to predict the observation sentences that verbalize our stimulations (see Quine, W.V. §2).

But many naturalists reject Quine's version of the empiricist claim that our theory rests on observation. One concern is that it is unclear how sensory stimulations can serve as *evidence* for a theory. It is relatively clear how a sentence can serve as evidence for a theory, so that we can see how observation sentences test theories, but stimulations are not sentences nor, for all we can see, any sort of propositional attitude. Moreover, the relationship between our stimulations and our assenting to observation sentences is causal, not evidential.

Another problem concerns the distinction between observation sentences (which are purported to have a special plausibility due to their intimacy with experience, so that they may serve to test theories) and the sentences in which theory is couched. Several philosophers (for example, N.R. Hanson 1968 and Paul Feyerabend 1986) have argued that there is no theory-observation distinction. There is no such thing as observation sentences whose meanings are determined by stimulations or the like. 'Observation' statements are just theoretical claims themselves, though ones we are accustomed to. So it will not do to say that our theory of the world rests on or is tested by non-theoretical observation statements. It is theory all the way down (see Observation §§3-4).

4 The autonomy of epistemology

All naturalists reject the Cartesian view that epistemology is entirely separable from and prior to other disciplines such as the sciences. But just how much autonomy epistemology retains is controversial. The most extreme view is that of Quine, who suggests that epistemology has no autonomy at all. Once we drop the idea that 'the epistemologist's goal is validation of the grounds of empirical science' we can and should, he says, 'surrender the epistemological burden to psychology'. We are to construe epistemology as the attempt to 'understand the link between observation and science', and consider ourselves 'well advised to use any available information, including that provided by the very science whose link with observation we are seeking to understand' (Quine 1969: 75-6). Hence epistemology collapses without residue into sciences such as biology and psychology. Its work is turned over to evolutionary epistemologists such as Campbell (1974) and genetic epistemologists such as Jean Piaget who attempt to explain the development of knowledge, the former explicitly in terms of the biological theory of evolution. But exactly how the sciences will distribute the work of studying knowledge is unclear. The scientific fields that study knowledge are somewhat fledgling and are not clearly differentiated. Nor is it clear how these fields are related to the sociology or sociobiology of knowledge, fields which also can be construed as part of naturalized epistemology (see Sociology of knowledge).

Some philosophers sympathetic to the naturalizing movement reject the claim that all legitimate epistemic issues can be investigated scientifically: (1) Alvin Goldman (1986) says that while epistemology and natural science are relevant to and continuous with each other, the former is not contained entirely in the latter. For example, the philosophy of logic is a branch of epistemology, but Goldman doubts that logic can be investigated scientifically. (2) Another argument is that epistemology progresses in part by studying the basic epistemic concepts, such as 'justification', 'rationality', 'warrant', 'certainty', 'knowledge' and so on. Such study is the basis for claims about what constitutes knowledge, yet philosophers, not scientists, undertake it. (3) Moreover, Goldman and many others would say, one task of epistemology is to settle normative questions such as that of how we ought to arrive at our conclusions, but science is unequipped to handle normative issues. It seems limited to telling us how we *do* arrive at our conclusions. Accordingly, epistemology might survive as an autonomous discipline in so far as it continues

to deal with deductive logic and the normative issues involved in knowledge.

Quine would reject each of these attempts to preserve an autonomous form of epistemology. To deny the first two, he would use arguments developed in 'Two Dogmas of Empiricism' (1953). (For a discussion of his approach to the third, see §5 below.) (1) That even claims of logic must face the tribunal of experience is suggested by the Quine-Duhem thesis: all statements face the 'tribunal of sense experience not individually but only as a corporate body' (Quine: 1953: 41). Hence 'no statement is immune to revision. Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics' (Quine: 1953: 43) (see Quine, W.V. §3). (2) Progress in epistemology based on the analysis of terms is made very difficult, but is not ruled out, by Quine's holism - his claim that the meanings of particular statements and *a fortiori* terms overlap with the meanings of other statements, so that describing the meaning of a particular term requires describing its role in a complex of theory (see Atomism, ancient; Holism: mental and semantic). Moreover, analysis merely clarifies the epistemic concepts we have been employing. It does not rule out the possibility that we ought to employ different concepts, such as ones suggested by an empirical study of the brain's workings.

5 The role of normative issues

Epistemology's *descriptive* task is to identify how people actually arrive at beliefs. But what people actually do is not necessarily what they ought to do. Epistemology's *normative* task is to identify how people ought (rationally) to arrive at their beliefs, and this seems to go well beyond the descriptive task (see Normative epistemology). Not all naturalists hope that they can do away with traditional epistemology entirely and replace it with natural science. But those with this hope have trouble finding a place in their project for the normative task of epistemology, since science seems incapable of prescription. Radical naturalists could argue that the normative issue is not worth pursuing, so that reducing epistemology to its descriptive task leaves out nothing worth doing. But this is an implausible option on its face. However, their only other choice is daunting: they must argue that science can tell us how we ought to arrive at our beliefs.

Naturalists who want to argue that science can answer the normative question have two options. In the spirit of Hume they can assume that the way we arrive at our beliefs *is* (more or less) the way we ought to. (But is this assumption a scientific or an epistemological claim?) Epistemology's descriptive task is clearly within the province of science; if we completed the normative by completing the descriptive task, then science could handle both sides of traditional epistemology. Unfortunately, the claim that we ought to maintain our existing belief-management practices more or less as they are faces difficulties. The main problem lies in the data that psychologists have already gathered. These data show that normal human cognitive processes are shot through with faulty logic, bad probabilistic inferences and wishful thinking (Nisbett and Ross 1980; Taylor 1989). Beliefs with such origins do not count as knowledge.

Perhaps radical naturalists need not assume that people ought to maintain their actual belief-management practices, however. Instead, they could try arguing that discoverable facts about things other than our actual belief-management practices will allow us to accomplish epistemology's normative task (but will they be able to make these discoveries without relying on the belief-management practices they question?). For example, scientists might be able to determine that all (rational?) human beings have a common epistemic goal, whether it is reaching the truth or predicting the future course of sensory stimulation or whatever. If so, then, as Quine (1992) has pointed out, epistemology's normative task could be performed by engineers. Engineers could work out the best ways available to people (given our limited faculties and resources) for reaching their epistemic goal. We could say that these efficient methods are the ones people ought to adopt, even at the expense of fairly radical changes in their actual practices. Normative epistemology becomes part of engineering science, not a branch of epistemology that is outside of science.

But while engineers investigate efficient ways to do such things as to transport or kill people, they do not investigate the issue 'Ought we to transport or kill people?' Only after it is established that it is important to achieve some goal does engineering come into play. The issue of what epistemic agents ought to aim at is not an engineering issue. It remains a philosophical issue which cannot be absorbed into science.

The problem would be rather trivial if scientists discovered that there is a single, unvarying goal (or prioritized set of goals) which everyone wants to reach by managing their beliefs as they do, and which does not look silly on its

face. Even if the goal is ‘wired in’, we would worry about committing a ‘naturalistic fallacy’ and say ‘The fact that something is everyone’s goal does not entail that it ought to be’. Even the fact that a goal is forced on everyone does not show that we have grounds for pursuing it. Still, the temptation to take this goal as the one we ought to pursue would probably be overwhelming. However, it is by no means obvious that such a single, unvarying goal exists. In different cultures at the same time, and in the same culture at different times, people might be aiming to accomplish a variety of things by believing what they do, and the (conscious and unconscious) belief-management principles they employ might also differ widely. Given the possibility of significant diversity, epistemologists need some way to decide which competing goals and sets of belief-management principles are right for cognitive agents. Perhaps the diversity would be reduced if we focused on the goals of scientists during those times when they are being scientific. But even if it were so reduced, we would need some way of arguing that the goals of the scientists are the ones which epistemic agents ought to pursue. Several ways to handle such normative issues are open to radical naturalists.

First, they could say that belief-management principles should be evaluated from the point of view of natural selection, so that the greater the survival value of these principles, the better the principles. Once this (epistemological?) claim is made, then science can take over, by describing the belief-generating mechanisms of human beings (and other animals?) and explaining how they have the survival value they do. (But how will naturalists absorb the growing data that suggest that wishful thinking, a paradigm case of irrationality, is adaptive?)

Second, naturalists could exploit the elastic boundaries of science in order to call conceptual analysis science. Then they could use conceptual analysis to defend some view about what epistemic agents’ goals or evaluative concepts are, and make the (epistemological?) claim that belief-management principles should be evaluated from the standpoint of those goals or concepts. Supposing, as several epistemologists have, that conceptual analysis reveals truth or predictive power to be the goal of epistemic agents, then (theoretical) scientists can go on to clarify the extent to which people naturally achieve the epistemic aim. And engineering will be useful in that it might help people to find better ways to achieve the epistemic goal.

Third, Stephen Stich (1988) suggests a pragmatic approach to normative issues. First he criticizes the strategy of using conceptual analysis to discover the goals and concepts of epistemic agents. The goals and evaluative epistemic concepts that are part of ordinary language are as likely as belief-management principles themselves to vary from culture to culture (see Cognitive pluralism). So it is arbitrary to rely on them when we select management principles. ‘In the absence of any reason to think that the locally prevailing notions of epistemic evaluation are superior to the alternatives, why should we care one whit whether the cognitive processes we use are sanctioned by those evaluative concepts?’ (Stich 1988: 406). Then Stich points out that there are many common values, such as happiness or reproductive success, that are not epistemic values but can be considered relevant to our cognitive lives. Stich suggests a pragmatic approach: he makes the (epistemological? ethical?) claim that we should evaluate belief-management principles from the standpoint of these non-epistemic values. Here again theoretical science will help us to evaluate our actual belief-forming mechanisms and engineering will help us to improve upon them.

6 The significance of scepticism

Like Hume, contemporary naturalists view epistemology as the attempt to clarify how the apparatus people use in investigating the world works when used in applications for which, we assume, it is reliable, and to identify what ought and ought not count as knowledge by identifying what sorts of beliefs are endorsed by the proper use of that apparatus. Contemporaries depart from him chiefly in thinking that more must be taken for granted than the reliability of mental faculties. Recent naturalists help themselves to the whole of natural science, which can be thought of as the combination of our mental faculties with techniques and devices that extend them.

Accordingly, there is a compelling case for saying that naturalists cannot hope to put global scepticism to rest. Global scepticism says that our belief schemes (including science) are irrational because: (1) ultimately our beliefs are based on arbitrary assumptions, claims that, even upon some reflection, we cannot link to a consideration that suggests it is true; and (2) it is irrational to make arbitrary assumptions (see Scepticism §5). To argue that scientific apparatus is reliable after having simply assumed that it is would be circular, so naturalists seem committed to granting (1). Perhaps this is why naturalists rarely confront global scepticism.

Attempts, none the less, have been made. One approach involves *coherentism*, which is the claim that beliefs may derive justification by cohering one with another (see Knowledge and justification, coherence theory of). If coherentism were correct, then since it sanctions some circular justifications, naturalists could use it against (1), and argue that all our beliefs can be justified at once. Another common approach is to turn the tables on the global sceptic and point out that, like everyone else who investigates knowledge, sceptics, too, must take for granted the reliability of their investigative apparatus. As an attack on (1) this table-turning would not work. Sceptics can retort: 'Yes, all of us are in the same boat: there are assumptions we simply take for granted'. But it might well prove useful as part of an attack on (2).

It is important to notice that the global sceptic needs both (1) and (2). Once we do, we can see that even if naturalists cannot defeat (1), they can still respond to scepticism if they defeat (2). Naturalists could accept the sceptic's discovery that ultimately our views are, perforce, arbitrary, and insist that it is sometimes all right, it is sometimes rational, to believe things we simply take for granted (Luper-Foy 1990).

See also: Knowledge, concept of; Naturalized philosophy of science; Social epistemology

STEVEN LUPER

References and further reading

Armstrong, D.M. (1973) *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.(Referred to in §2 above. A difficult, technically sophisticated but groundbreaking defence of an externalist theory of knowledge.)

BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.(Discussed above in §2. Important and highly accessible defence of an internalist approach to the theory of knowledge.)

Campbell, D.T. (1974) 'Evolutionary Epistemology', in P.A. Schilpp (ed.) *The Philosophy of Karl Popper*, bk 1, La Salle, IL: Open Court.(Referred to in §4 above. Helpful account of the field of evolutionary epistemology.)

Churchland, P. (1979) *Scientific Realism and the Plasticity of Mind*, Cambridge: Cambridge University Press. (Discussed in §2 above. Reasonably accessible and influential attack on the idea that knowledge is 'sentences suitably related'.)

Feyerabend, P.K. (1986) *Realism, Rationalism and Scientific Method: Philosophical Papers Volume 1*, Cambridge: Cambridge University Press.(Discussed in §3 above. Demanding essays in the philosophy of science.)

Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(Discussed above in §4. Reasonably accessible and influential defence of a moderately naturalized epistemology.)

Hanson, N.R. (1968) *Patterns of Discovery: An Enquiry into the Conceptual Foundations of Science*, New York: Cambridge University Press.(Discussed in §3. Accessible introduction to the philosophy of science.)

Harman, G. (1986) *Change in View: Principles of Reasoning*, Cambridge, MA: MIT Press.(Discussed in §2 above. Highly accessible defence of a moderately accessible naturalized epistemology.)

Hume, D. (1748/51) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, revised P.H. Niddich, Oxford: Clarendon Press, 3rd edn, 1975.(Referred to in §2 above. Accessible and extremely influential.)

Kitchener, R.F. (1986) *Piaget's Theory of Knowledge: Genetic Epistemology and Scientific Reason*, New Haven, CT: Yale University Press.(Exhaustive discussion of genetic epistemology.)

Kornblith, H. (ed.) (1985) *Naturalizing Epistemology*, Cambridge, MA: MIT Press. (Contains an introductory essay which defends the idea that the distinctive feature of naturalized epistemology is its position on the relationship between the normative question of how we ought we to arrive at our beliefs, and the descriptive issue of how we do we arrive at our beliefs. Also contains leading naturalist essays and an exhaustive bibliography up to 1984. Most of the material is accessible to non-specialists.)

Kornblith, H. (1993) *Inductive Inference and its Natural Ground: An Essay in Naturalistic Epistemology*, Cambridge, MA: MIT Press.(Accessible and lucid explanation of inductive inference from the perspective of naturalistic epistemology.)

Luper-Foy, S. (1990) 'Arbitrary Reasons', in *Doubting*, Dordrecht: Kluwer.(Referred to above in §6. Argues that it is rational to accept arbitrary beliefs.)

Nisbett, R. and Ross, L. (1980) *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, NJ: Prentice Hall.(Referred to in §5 above. Accessible survey of literature on irrationality.)

Quine, W.V. (1953) 'Two Dogmas of Empiricism', in *From a Logical Point of View*, New York: Harper Torchbooks.(Referred to in §4 above. Spells out much of the theory that suggested naturalized epistemology.

Most of it is accessible.)
Quine, W.V. (1969) 'Epistemology Naturalized', in *Ontological Relativity & Other Essays*, New York: Columbia University Press.(Referred to in §§2 and 4. The definitive essay of naturalized epistemology. Accessible.)
Quine, W.V. (1974) *The Roots of Reference*, La Salle, IL: Open Court.(Referred to above in §1. Discusses induction and hints at an externalist refutation of scepticism. Accessible.)
Quine, W.V. (1992) *Pursuit of Truth*, Cambridge, MA: Harvard University Press.(Referred to in §5 above. Section 8 of this book elaborates (in the minimalist fashion typical of Quine) on the role of normativity in naturalized epistemology.)
Stich, S. (1988) 'Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity', *Synthese* 74 (3): 391-415.(Discussed above in §5. Defends a pragmatic approach to belief management. Accessible.)
Taylor, S. (1989) *Positive Illusions: Creative Self-Deception and the Healthy Mind*, New York: Basic Books. (Referred to in §5 above. Very accessible argument for the claim that self-deception is adaptive.)

Naturalized philosophy of science

Naturalized philosophy of science is part of a general programme of naturalism in philosophy. Naturalists reject all forms of supernaturalism, holding that reality, including human life and culture, is exhausted by what exists in the causal order of nature. Naturalists also reject any claims to a priori knowledge, including that of principles of inference, holding instead that all knowledge derives from human interactions with the natural world. Philosophically, naturalists identify most closely with empiricism or pragmatism. David Hume was a naturalist. So was John Dewey. The logical empiricists were naturalists regarding fundamental ontological categories such as space, time and causality, but non-naturalists about scientific inference, which they came to regard as a branch of logic. Most naturalists now dismiss searches for 'philosophical foundations' of the special sciences, treating the basic principles of any science as part of scientific theory itself.

The main objection to naturalism has been at the level of general methodological principles, particularly those regarding scientific inference. Here non-naturalists object that, being limited to 'describing' how science is in fact practised, naturalists cannot provide norms for legitimate scientific inferences. And providing such norms is held to be one of the main goals of the philosophy of science. Naturalists reply that the only norms required for science are those connecting specific means with the assumed goals of research. These connections can be established only through further scientific research. And the choice of goals is, for naturalists, not a scientific question but a matter of practical choice guided by an empirical understanding of what can in fact be achieved.

Among naturalists, the main differences concern the relative importance of various aspects of the practice of science. These aspects exist at different levels: neurological, biological, psychological, personal, computational, methodological, social and cultural. Each of these levels has its champions among contemporary naturalists, some insisting that everything ultimately be reduced to their favourite level. Perhaps the wisest approach is to allow that influences at all levels are operative in any scientific context, while admitting that some influences may be more important than others in particular cases, depending on what one seeks to understand. There may be no simple, one-level, naturalistic theory of science.

1 Naturalism and the philosophy of science

Philosophical naturalism is the view that an adequate philosophical account of the natural world, including humans, can be given solely in terms of objects and processes occurring in the natural causal order. No appeal to anything supernatural or transcendent is required. Materialism and other forms of reductionism imply naturalism, but naturalists need not be reductive materialists. Thus a naturalist might deny the reducibility of psychology to biology so long as psychology itself functions as a natural science in its methods and forms of explanation (see Reduction, problems of; Unity of science).

Naturalism is not new. David Hume was clearly a naturalist in holding that the operations of the mind could be understood solely in terms of natural processes such as perception, memory and the association of ideas. His goal has been rightly characterized as that of becoming the Newton of the mind. J.S. Mill continued the naturalist tradition in Britain. Descartes, on the other hand, counts as a non-naturalist because his theory of knowledge required an immaterial soul with special powers for ascertaining truths involving clear and distinct ideas. Kant was a more sophisticated non-naturalist. For Kant the operations of individual minds take place in the causal order of appearances. But all perception and judgment is constrained by categories that are discoverable only by the special philosophical method of transcendental deduction. In the mid-nineteenth century, other German scientist-philosophers, such as Hermann Helmholtz attempted to naturalize Kant's categories, particularly those pertaining to spatial perception, through physiological and psychological investigations.

The late nineteenth-century debate in Germany between naturalists and non-naturalists was echoed in the USA where philosophers also attempted to assimilate the findings of the sciences, particularly psychology. John Dewey exemplified this debate within his own philosophical development. Beginning as a Hegelian who saw psychology as applying only to individuals while philosophy uncovered the universal laws of rational thought, Dewey ended up a pragmatic naturalist who looked to science for all knowledge of the world and to philosophy primarily for investigation into the social values that should direct the employment of scientific knowledge for human betterment.

The non-naturalist cause found a major champion in Gottlob Frege who, beginning in the 1880s, insisted on a sharp distinction between patterns of inference actually employed by individuals and objectively valid arguments. The former may be studied as part of empirical psychology; the later are the province of logic, an a priori discipline which is part of philosophy (see Frege, G. §4). Frege's views were incorporated into the programme of scientific philosophy promoted by the Vienna Circle led by Moritz Schlick and Rudolf Carnap, and by the Berlin Society for Scientific Philosophy led by Hans Reichenbach (see Vienna Circle §3). The major thrust of scientific philosophy was the naturalization of metaphysics, that is, the replacement of Neo-Kantian philosophical metaphysics with a scientific metaphysics. In particular, the fundamental properties of space, time and causality, they argued, cannot be learned through any sort of purely philosophical inquiry, but only through the study of the appropriate scientific theories, particularly relativity theory and the quantum theory (see General relativity, philosophical responses to §3). The task left to philosophy was the logical analysis of scientific theories, a purely a priori undertaking. Eventually included in this analysis was the *epistemological* relationship between statements of evidence and statements of theory. Following Frege, this came to be regarded as the logical study of inductively valid or rationally justified inference, independent of any inferences actual scientists might in fact make (see Inductive inference).

In the face of German National Socialism, most of the scientific philosophers left the German-speaking world and the vast majority, including Carnap and Reichenbach, ended up in the USA. Their incorporation into the American philosophical world was fostered primarily by pragmatist philosophers. By the mid-1950s, what in Germany had been a radically anti-establishment scientific philosophy had become in North America the established philosophy of science. Ironically, pragmatism and its associated naturalism were eclipsed by a rival empiricist programme incorporating a militantly non-naturalistic view of scientific method (see Pragmatism; Logical positivism §§3-4).

The revival of naturalism owes much to the works of Kuhn and Quine during the 1960s. In *The Structure of Scientific Revolutions* Kuhn attempted to explain the historical development of science in terms borrowed from fields such as linguistics, psychology, and sociology - a thoroughly naturalistic project. (see Kuhn, T.S. §2). Quine, by contrast, argued for a naturalized epistemology by demonstrating the impossibility of carrying out any programme of reducing all knowledge to sense impressions, a programme that had come to be synonymous with a Fregean approach to epistemology (see Quine, W.V. §2; Naturalized epistemology). Unfortunately, the empirical resources available in the 1960s, gestalt psychology, behaviourism and structural-functional sociology of science, were not sufficient to develop the naturalistic programmes of either Kuhn or Quine. Since the mid 1970s, however, developments in the cognitive sciences and constructivist sociology of science have provided far richer resources (see Constructivism). By the mid-1980s, a naturalized philosophy of science had become once again a viable option.

2 The problem of epistemic norms

The most persistent objections to the re-emergence of naturalism in the philosophy of science involve the status of epistemic norms, or principles of rational inference. The problem may be stated as follows: a naturalistic philosophy of science, it is said, can at most provide *descriptions* of actual scientific practice. But science requires *normative principles* which distinguish those claims legitimately warranted by available evidence from those not so warranted. So no naturalized philosophy of science can be a complete philosophy of science. A more pointed way of framing the same sort of objection is this: a naturalized philosophy of science appeals to established theories of various sciences. But one of the conclusions of any philosophy of science should be an account of what makes some theories warranted and others not. So a naturalized philosophy of science either begs a central question or remains viciously circular.

Naturalists respond to the more pointed objection by questioning the presupposition that it is possible to justify any principles of scientific inference a priori. Here naturalists can point to nearly 300 years of attempts by empiricist philosophers to find an adequate justification of induction that does not beg the question (see Induction, epistemic issues). Such considerations, of course, do not constitute an a priori proof that no such justification will ever be forthcoming. But where, the naturalist asks, might such a justification be found? In logic, in intuition,or in an a priori theory of human judgment? These have all been tried without acknowledged success. Where could we turn next? The naturalist concludes that there most likely is no place beyond the province of science that could provide such a justification. The supposition that there must be some such place looks to be the product of an unwarranted

yearning for some sort of absolute foundation for scientific claims.

Positively, naturalists recommend a non-foundational, pragmatic approach to epistemology. Individually and collectively we can do no better than begin where we are, with the theories and methods we have inherited, historically contingent though these may be. Any theory or method may be questioned, though not without a specific reason, and not all at once. The general goal is not to ground what we think we know, but to improve upon it. As for all-encompassing sceptical worries, naturalists take comfort in evolutionary biology. We as a species would not be here if our evolved sensory and motor skills were not good enough to secure the necessities of life and procreation. This reflection begs the question against a radical sceptic of course, and does little to explain how we achieved our current advanced state of scientific knowledge, including that of evolutionary biology. But it provides a naturalistic starting point.

As for the general problem of epistemic norms, naturalists insist that categorical principles of rational inference are not to be had. But that still leaves conditionally normative principles of the form: if you wish to accomplish goal *G*, try method *M*. The warrant for such principles consists in finding empirical evidence that doing *M* in fact promotes achievement of *G*. In this sense, all knowledge has potential normative implications, for example, as nuclear physics had implications for how to go about building a powerful bomb.

Contemporary naturalists disagree about the kind of principle to be sought, and the kind of evidence relevant to normative claims. Some, for example Larry Laudan, look for general methodological rules such as: successful prediction of novel facts provides stronger evidence for a theory than does explanation of previously known facts. And the relevant evidence is said to be found in historical instances in which theories which generated successful predictions turned out to be better than others which did not. Other naturalists doubt the usefulness of rules at this level of generality either for the practice of science or for explaining how science works. And they doubt that the historical record can in fact furnish the required evidence. Rather, it is argued, we need a deeper theory of how scientists do science and of the methods they employ. Such a theory could help us explain why some scientists were successful and others not. It could even provide some general normative guidance for scientists themselves, or for those who determine priorities for funding scientific research. And it could explain the circumstances in which more general rules, like that regarding the value of successful predictions, may be operative. The evidence for such a theory of science would have to go beyond mere correlation between supposed applications of a rule and the end results. It would be more like evidence which supports any general theory in such areas as biology, statistics, the cognitive sciences or sociology. Here it is important that statistics and experimental design be seen as themselves scientific theories relating experimental designs, experimental results, and correct decisions among alternative hypotheses.

Finally there is the question of the general epistemic goals of science. Naturalists reject attempts to determine a priori what the epistemic goals of science should be. They doubt there are any justifiable principles on which to base such attempts. Rather, we can investigate only what epistemic goals have in fact been pursued by various scientists and, perhaps, the empirical consequences of pursuing various goals. One need not presume that all scientists everywhere have pursued the same epistemic goals, or even that their professed goals are the same as their actual goals. Naturalists thus reject general philosophical arguments for or against ’realism’ or for or against ‘empiricism’ (see Scientific realism and antirealism §3). The historical record shows that scientists have pursued both sorts of goal, sometimes successfully and sometimes not.

3 The bases for naturalization

Naturalists differ among themselves mainly on the sorts of scientific knowledge they deem most relevant to a naturalistic understanding of science. Individual naturalists typically stress one type of knowledge over all others, and the types of knowledge tend to represent different levels of organization. Paul Churchland, for example, attempts to understand scientific explanation and theorizing at the neuro-computational level. Paul Thagard argues that conceptual revolutions can be explained by relationships of explanatory coherence among statements that can be modelled in a quasi-connectionist computer program. Many naturalists, following in the footsteps of Darwin himself, appeal to evolutionary biology, though in different ways (see Evolution, theory of). Michael Ruse thinks that evolution yielded general methodological rules we follow intuitively and discover by reflecting on our practice. David Hull (1988) argues for strict isomorphisms among organic evolution, the evolution of scientific communities, and the evolution of scientific concepts. Others, including Stephen Toulmin, Donald Campbell and

Abner Shimony make less strict use of evolutionary ideas in their philosophies of science. Still other naturalists, such as Nancy Nersessian, focus on concepts and results developed in cognitive psychology. Some would add here a theory of human judgment and experimental design. Finally, others, such as Steve Fuller, find the most explanatory power at the level of social interaction and communication. What is an aspiring naturalist to conclude?

The wise course is to recognize that there can be no such thing as a homogeneous theory of science. Science is far too complex a phenomenon for that. Rather, we must settle for a diverse set of models in terms of which we can understand how science works in specific instances. These models will be inspired by models first developed in a wide variety of sciences. The best to be hoped for is that the various models employed are complementary. Where applications overlap, the results should at least be compatible. But we must also realize that there is a limit to theoretical naturalistic explanations of science. Some features of various sciences will have no deeper explanation than the existence of various historical contingencies in their origin and development. As a human activity, even as an activity expressly designed to produce faithful representations of nature, science will always contain elements of historical contingency.

RONALD N. GIERE

References and further reading

Giere, R.N. (1988) *Explaining Science: A Cognitive Approach*, Chicago, IL: University of Chicago Press.(A comprehensive naturalistic theory of science emphasizing contributions from the cognitive sciences.)

Giere, R.N. (ed.) (1992) *Cognitive Models of Science*, Minnesota Studies in the Philosophy of Science, vol. 15, Minneapolis, MN: University of Minnesota Press.(Includes original essays by leading cognitive naturalists mentioned in §3, including Churchland, Fuller, Nersessian and Thagard.)

Hatfield, G. (1989) *The Natural and the Normative: Theories of Perception from Kant to Helmholtz*, Cambridge, MA: MIT Press.(A masterly though demanding history of naturalism in German philosophy.)

Hull, D. (1988) *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*, Chicago, IL: University of Chicago Press.(A detailed elaboration of an evolutionary model of science. See §3.)

Kitcher, P. (1992) 'The Naturalists Return', *Philosophical Review* 101: 53-114.(A survey of contemporary philosophical naturalism emphasizing the influence of Frege.)

Kornblith, H. (ed.) (1985) *Naturalizing Epistemology*, Cambridge, MA: MIT Press.(Includes classic papers by Campbell (§3) and Quine (§1).)

Laudan, L. (1987) 'Progress of Rationality? The Prospects for Normative Naturalism', *American Philosophical Quarterly* 24: 19-31.(Laudan's answer to the problem of epistemic norms in a naturalistic philosophy of science, §2.)

Laudan, L. *et al* (1986) 'Scientific Change: Philosophical Models and Historical Research', *Synthèse* 69: 141-223. (Sets out Laudan's programme for the empirical study of scientific change §2.)

Ruse, M. (1986) *Taking Darwin Seriously*, Dordrecht: Reidel.(Contains an evolutionary account of science grounded in sociobiology, §3.)

Wilson, D.J. (1990) *Science, Community, and the Transformation of American Philosophy, 1860-1930*, Chicago, IL: University of Chicago Press.(A historian's view of naturalism in American philosophy before the logical empiricists.)

Nature and convention

The nature-convention distinction opposes instinctual or 'spontaneous' modes of comportment (those which follow from 'human nature') to those which are socially instituted or culturally prescribed. Its philosophic interest resides in its use to justify or contest specific forms of human behaviour and social organization. Since the 'conventional' is opposed to the 'natural' as that which is in principle transformable, the adherents of a particular order in human affairs have standardly sought to prove its 'naturality', while its critics have sought to expose its merely 'conventional' status. Relatedly, 'conventions' may be associated with what is distinctive to 'human', as opposed to 'bestial' nature, or denounced for their role in repressing our more 'natural' impulses.

The nature-convention antithesis is one of a number of cognate dualisms (nature-culture; natural-artificial; animal-human) through which Western thought conceptualizes what is distinctive to the 'being' of humanity, and demarcates between that which is independent of and pre-given to human activity and that which is humanly achieved or contrived. Though many would dispute whether any clear-cut division can be drawn between the 'human' and the 'animal', the very posing of the question of our affinity with or distinction from other creatures is reliant on a prior appreciation of those features which are specific to humans. One of the most notable of these is that human beings establish 'conventions': customs, rules, or norms of behaviour which are socially instituted as opposed to naturally dictated. The establishment of conventions is thus intimately bound up with those capacities for moral thought and action which mark us off from other animals. To follow (or contest) a convention presupposes an ability to understand prescription, and to act otherwise than in obedience to the dictate of instinct.

The nature-convention opposition has its origins in Greek antiquity, in the fifth century BC, where the distinction between *physis* and *nomos* becomes the site of a dispute about the cultural relativity of political and moral values (see Plato; Sophists) which has remained a central concern of philosophy ever since. Thus it is by reference to the multiplicity of human conventions, that relativists have often challenged claims about the existence of a common 'human nature' or universally applicable morality (see Relativism §1).

Conventions may be very culturally specific, transient, and readily transformed codes of conduct, or more perduring customs or institutional forms which are none the less particular to given societies. In some instances the conventional prescriptions or prohibitions may be trans-cultural and deeply entrenched (as in the case of the law against incest, which Claude Lévi-Strauss presents as marking the arrival of the 'cultural'), while remaining in principle revisable as norms of conduct precisely in virtue of their being humanly ordained.

It is the double association of the conventional both with what is 'right' for human beings and with what is transformable, and of the 'natural' both with the 'bestial' and with what is deemed immutable, which accounts for the complex ideological disputes that attach to the distinction. The rules that human beings set themselves (against murder, or incest, for example) have very frequently been viewed as preserving them from the degeneracy or 'wickedness' of nature's ways (a position argued very forcibly by John Stuart Mill in his essay 'Nature' (1874)). But in the Rousseauan or Romantic conception, nature is viewed as a source of goodness or authenticity, and hence invoked as a model from which humanity should take its cue. These differing perspectives, however, may both seek confirmation in terms that tend to blur or confound the nature-convention distinction: those forms of behaviour which distinguish us from the beast being endorsed as qualities of human nature (and hence denied the status of 'mere' convention); and those which nature inspires being pitted against those supposedly natural rules of conduct which it is claimed are more properly to be viewed as repressive convention.

The philosophic interest of the nature-convention distinction lies therefore in the normative disputes that are conducted in its terms: that which is socially instituted is at once both defended by its adherents as part of the order of nature, and thus untransformable, and exposed by its critics as conventional and hence mutable. Relations of class, and discriminations on grounds of gender, race and sexuality, for example, have all been said to be natural by those seeking to sustain their social hierarchies, and denounced as illegitimate and oppressive norms by those seeking to overthrow them. Within this challenge, however, to the naturalization of what is socially instituted, we may distinguish between those who would contest what is said to be natural in the name of the more authentic naturality of what they seek to institute; and those (like Michel Foucault) who would insist upon the always normative and constantly revisable quality of what is said to be natural to human societies at any point in time. For

example, one may contest existing social relations, institutions and sexual 'norms' in terms of their 'alienation' from or 'repression' of nature, or dispute the very existence of natural, as opposed to socially 'constructed' or inculcated needs and desires. It is arguable, however, that without recourse to an idea of nature there can be no coherent or self-consistent critique of custom or convention.

See also: Conventionalism; Culture; Naturalism in social science

KATE SOPER

References and further reading

Horigan, S. (1988) *Nature and Culture in Western Discourses*, London: Routledge.(A critical survey of the ways in which the nature-culture divide has been conceptualized in anthropology.)

Lovejoy, A.O. (1948) 'Nature as Aesthetic Norm', in *Essays in the History of Ideas*, Baltimore, MD: Johns Hopkins University Press.(Briefly lists and comments on the inflections acquired by the term 'nature' when used in reference to human conduct.)

Mill, J.S. (1874) 'Nature', in *Three Essays on Religion*, London: Longman.(Usefully distinguishes between non-normative and normative uses of the term, and illustrates the tensions to which the latter are subject.)

Nussbaum, M.C. (1986) *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*, Cambridge: Cambridge University Press.(For many insights on classical Greek thought concerning the distinction. See especially pp. 397-421.)

Soper, K. (1995) *What is Nature? Culture, Politics and the Non-human*, Oxford: Blackwell. (Discusses the contested nature of 'nature' in contemporary theory and social movement politics.)

Taylor, C. (1989) *Sources of the Self*, Cambridge: Cambridge University Press, part IV, pp. 305-68.(Illuminating discussion of Enlightenment, Rousseauan and Romantic conceptions of 'nature' and its role as guide to human conduct.)

Nature, aesthetic appreciation of

In the Western world, aesthetic appreciation of nature and its philosophical investigation came to fruition in the eighteenth century. During that time, aestheticians made nature the ideal object of aesthetic experience and analysed that experience in terms of disinterestedness, thereby laying the groundwork for understanding the appreciation of nature in terms of the sublime and the picturesque. This philosophical tradition reached its zenith with Kant, while popular aesthetic appreciation of nature continued primarily in terms of the picturesque.

In the late twentieth century, renewed interest in the aesthetics of nature has produced various positions designed to avoid assimilating appreciation of nature with traditional models for aesthetic appreciation of art. Three are especially noteworthy. The first holds that the appreciation of nature is not in fact aesthetic; the second rejects the traditional analysis of aesthetic experience as disinterested, arguing instead that the aesthetic appreciation of nature involves engagement with nature; the third attempts to maintain the traditional analysis, while distinguishing aesthetic appreciation of nature by dependence on scientific knowledge.

These positions have a number of ramifications. In freeing aesthetic appreciation of nature from artistic models, they pave the way for a general environmental aesthetics comparable to other areas of philosophy, such as environmental ethics. Moreover, the significance given to scientific knowledge in the third position both explains the aesthetic appreciation associated with environmentalism and provides aesthetic appreciation of nature with a degree of objectivity that may make aesthetic considerations more effectual in environmental assessment.

1 History

Although a tradition of viewing art as the mirror of nature has existed since antiquity, the idea of aesthetically appreciating nature itself is sometimes traced toPetrarch's novel passion for climbing mountains, simply in order to enjoy the prospect. Yet from the dawn of the Renaissance to the present, the development of aesthetic appreciation of nature has been uneven and episodic. Initially such appreciation, together with its philosophical investigation, was hamstrung by a religious tradition that saw mountains as despised heaps of wreckage left behind by the flood, wilderness regions as fearful places for punishment and repentance, and all of nature's workings as a poor substitute for the perfect harmony lost in humanity's fall. It took the rise of a secular science and an equally secular art to free nature from such associations and open it up for aesthetic appreciation. Thus, in the Western world, the evolution of aesthetic appreciation of nature has been intertwined with the objectification of nature achieved by science and the subjectification of it completed by art.

Early in the eighteenth century, British aestheticians initiated a tradition that gave theoretical expression to the connection between aesthetic appreciation of nature and scientific objectivity. Empiricist thinkers, such asJoseph Addison and Francis Hutcheson, took nature rather than art as the ideal object of aesthetic experience and developed the notion of disinterestedness as the mark of such experience. In the course of the century, this notion was elaborated so as to exclude from aesthetic experience an ever-increasing number of associations and conceptualizations. Thus, the objects of appreciation favoured by the tradition - British landscapes - were eventually severed not only from religious associations, but from an appreciator's personal, moral and economic interests. The upshot was a mode of aesthetic appreciation that looked upon the natural world with an eye not unlike the distancing, objectifying eye of science. In this way, the tradition laid the groundwork for the idea of the sublime, by which even the most threatening of nature's manifestations, such as mountains and wilderness, could be distanced and appreciated rather than simply feared and despised (see Sublime, the §§2-3).

However, if disinterestedness laid the groundwork for the sublime, it also cleared the ground for another, quite different idea: the picturesque. This idea secured the connection between aesthetic appreciation of nature and the subjective renderings of nature in art. The term literally means 'picture-like' and indicates a mode of appreciation in which the natural world is divided into scenes, each aiming in subject matter or in composition at an ideal dictated by art, and especially by poetry and landscape painting. Thus, while disinterestedness and the sublime stripped and objectified nature, the picturesque dressed it in a new set of subjective and romantic images: a rugged cliff with a ruined castle, a deep valley with an arched bridge. Like disinterestedness, the picturesque had its roots in the theories of early eighteenth-century aestheticians, such as Addison, who thought 'works of nature' more appealing the more they resembled art. The trend of the picturesque culminated later in the century, when,

popularized primarily by William Gilpin and Uvedale Price, it became the aesthetic ideal of English tourists as they pursued picturesque scenery in the Lake District and the Scottish Highlands. Indeed, the picturesque remains the preferred mode of aesthetic appreciation for the form of tourism that sees the natural world in the light of travel brochures, calendar photos and picture postcards.

While the picturesque lingered on as a popular appreciative mode, theoretical study of the aesthetics of nature, after flowering in the eighteenth century, went into decline. Many of the main ideas, such as disinterestedness and the centrality of nature rather than art, reached their climax withKant (Sublime, the §3), and in his third critique these ideas received such exhaustive treatment that a kind of closure was achieved. In the new world order initiated by Hegel, art was a means to the Absolute, and it, rather than nature, was destined to became the favoured subject of philosophical aesthetics.

2 Background to the present

However, as theoretical study of the aesthetics of nature declined, a new view of nature was initiated that eventually gave rise to a different kind of aesthetic appreciation. This mode of appreciation was rooted in the nature writings of Henry David Thoreau and reinforced by George Perkins Marsh's recognition that humanity is the major cause of the destruction of nature's beauty. It achieved its classic realization at the end of the nineteenth century in the work of the American naturalist John Muir.Muir saw all nature and especially wild nature as aesthetically beautiful and found ugliness only where nature was subject to human intrusion. These ideas strongly influenced the North American wilderness preservation movement and continue to shape the aesthetic appreciation of nature associated with contemporary environmentalism. This kind of appreciation may be called positive aesthetics and, in so far as it eschews humanity's marks on the landscape, contrasts sharply with picturesque appreciation, with its delight in the signs of human presence. It became and remains the rival of the picturesque as the popular mode of aesthetic appreciation of nature, although popular appreciation is typically an uneasy balance of the two.

In spite of these developments in popular appreciation of nature, however, philosophical aesthetics, with few exceptions, ignored nature during the nineteenth and much of the twentieth centuries. Friedrich Schelling and a few thinkers of the Romantic movement dealt with it to some extent, and George Santayana and John Dewey both discussed it. But, by and large, aesthetics was dominated by an interest in art. By the mid-twentieth century, philosophical aesthetics in the analytic tradition was virtually equated with philosophy of art. Moreover, when aesthetic appreciation of nature was mentioned, it was treated, by comparison with that of art, as a messy, subjective business of little philosophical significance. In this context Ronald Hepburn's work in the 1960s had special importance. Hepburn argued that aesthetic appreciation of nature was as significant and as rich as that of art, and that it could be, though not a carbon copy of appreciation of art, similar to that of art - shallow or deep, appropriate or not. It required serious philosophical study.

3 Positions and problems

In the latter part of the twentieth century, renewed interest in the aesthetics of nature has generated a number of different positions on aesthetic appreciation of nature. Many of them share the view that such appreciation cannot simply be assimilated with standard accounts of aesthetic appreciation of art. In this they affirmHepburn's insight that the appreciation of nature, although comparable to that of art, does not mimic it. The rejection of some of the traditional modes of aesthetic appreciation is implied. For example, it is argued that appropriate appreciation of nature cannot involve viewing natural objects in the same way that we frequently view traditional art objects such as works of sculpture - that is, as passive objects of contemplation divorced from both their and our spatial and temporal contexts. By contrast, natural objects belong in an environment that not only surrounds and engages them, but surrounds and engages the appreciator. Thus, the physical or contemplative removal of natural objects or ourselves from their environments results in inappropriate appreciation. Likewise, it is held that viewing nature as we might view a landscape painting, that is, as a static scene that we behold from a predetermined distance, is equally inappropriate in that it misleadingly reduces nature to a series of two-dimensional scenic views.

Given the rejection of traditional modes of aesthetic appreciation of nature, an alternative approach becomes necessary. The most radical of such alternatives elevates the neglect of nature by analytic aesthetics to a philosophical position, denying either implicitly or explicitly the possibility of aesthetic appreciation of nature.

When explicitly developed, this position leaves traditional modes of appreciation intact for art, but concerning nature argues as follows: aesthetic appreciation necessarily involves aesthetic judgment, which entails judging the object of appreciation as the achievement of a designing intellect. However, since nature is not the product of a designing intellect, appreciation of it is not aesthetic. In the past, the appreciation of nature was deemed aesthetic only because of the assumption that nature is the handiwork of a designing creator; this assumption is either false, or at least inadequate for grounding an aesthetic of nature. This position is problematic, however, since it runs against both the orthodox view that everything is open to aesthetic appreciation and the common-sense idea that at least some natural objects, such as certain landscapes and sunsets, constitute paradigm objects of aesthetic appreciation.

A second alternative rejects traditional modes of aesthetic appreciation not only for nature but for art as well. This position argues that the disinterestedness tradition involves a mistaken analysis of the aesthetic and that this is most evident in the field of aesthetic appreciation of nature. In such appreciation the quality of disinterestedness, with its isolating, distancing and objectifying gaze, is out of place, for it is the means by which both natural objects and appreciators are wrongly abstracted from the environments in which they properly belong and in which appropriate appreciation is achieved. Thus, this position replaces disinterestedness with engagement, distance with immersion, and objectivity with subjectivity, and calls for a new aesthetic of engagement. Such an aesthetic, it is argued, yields a mode of appreciation not only more appropriate for nature, but also for most works of art, at least for those of the twentieth century. However, this second position is also problematic. First, since disinterestedness constitutes the favoured analysis of the aesthetic, its rejection may constitute a rejection of the aesthetic itself, reducing this second alternative to a version of the first. Second, this position seemingly embraces an unacceptable degree of subjectivity in aesthetic appreciation of both nature and art.

A third position attempts to maintain much of the traditional analysis of the aesthetic, while yet accommodating features unique to nature. The motivation for this alternative is not only to find a mode of aesthetic appreciation especially appropriate for nature, but also to avoid excessive subjectivity. This position notes that in aesthetic appreciation of art both appropriateness of appreciative mode and objectivity are achieved by tying appreciation to understanding in terms of art-historical and art-critical knowledge. It thus adopts a parallel approach for aesthetic appreciation of nature. However, although the appropriate knowledge for art is derived from art history and art criticism, nature is not art and such knowledge is not relevant to nature's appreciation. For nature, it is argued, the relevant knowledge is that which provides our understanding of the natural world, that is, the knowledge given by the natural sciences. For example, just as art-historical knowledge of mid-twentieth century schools of painting endows appreciation of an abstract expressionist work with appropriateness and objectivity, so knowledge of geology, biology and ecology similarly endows appreciation of an alpine meadow. In this way, aesthetic appreciation of nature is once again linked to the objectifying eye of science.

Like the other two alternatives, however, this third alternative faces some philosophical difficulties. One is that a scientific approach may not yield real aesthetic appreciation. However, unlike the second alternative, the third position does not reject traditional resources, such as disinterestedness, as constitutuents of aesthetic appreciation. Thus, a science-based mode of appreciation can be sufficiently disinterested to be truly aesthetic. Indeed, the problem may be the opposite one, that such a mode of appreciation, given its essential dependence on science, is too distancing, too objectifying. It thus runs the risk of wrongly abstracting natural objects from their environment and in this way making appropriate appreciation impossible. In reply it is argued that since aesthetic appreciation necessarily involves distancing and thus some degree of abstraction, the issue is not whether to abstract. Rather it is how to supplement the required abstraction so as to achieve rich and appropriate appreciation. The problem with traditional approaches is not that they involve abstraction, but that, having abstracted nature, they reconstitute it by means of artistic models, such as the picturesque, and not in terms of relevant knowledge, such as science provides. Moreover, science is not, or at least is no longer, the severely abstracting endeavour it is popularly taken to be. Modern natural science, as exemplified by ecology, increasingly views nature in holistic terms.

4 Ramifications

These approaches to the aesthetics of nature have a number of ramifications. One is that, concerning popular appreciation of nature, they seemingly support positive aesthetics as opposed to the picturesque. Appreciating nature in terms of artistic models, such as the picturesque, is rejected by all these approaches, and the scientific

approach in particular accords well with the positive aesthetic model. When nature is aesthetically appreciated in virtue of natural science, positive aesthetic appreciation seems singularly appropriate, for, on the one hand, humanity's marks on the landscape appear especially out of place, while, on the other, wild, pristine nature becomes an aesthetic ideal. Moreover, as the natural sciences increasingly find, or at least appear to find, unity, order and harmony in nature, nature itself, when appreciated in the light of such knowledge, increasingly appears more profoundly beautiful. Science replaces the once despised wreckage of God's deluge with inspiring manifestations of orderly geological forces; and the harmony lost in humanity's fall is superseded by a perfect harmony of science's own making.

Two other ramifications concern aesthetics and philosophy more generally. First, all three positions (and particularly the scientific approach), in rejecting artistic models for appreciation of nature and in replacing them with a dependence on knowledge of the object of appreciation, provide a blueprint for aesthetic appreciation in other non-art cases. It is made evident that in aesthetic appreciation of, for example, toys, furniture, people, cornfields, pets, neighbourhoods, shoes and shopping malls, what is appropriate is not an imposition of various artistic models, but rather a dependence on knowledge that is pertinent to the nature of the object of appreciation. This suggests a general and universal aesthetics, which may be termed environmental aesthetics, distinct from traditional aesthetics, which is too easily equated with philosophy of art. Second, there is an alignment of this universal aesthetics with other areas of philosophy in which there is a rejection of archaic, inappropriate models and a new-found dependence on knowledge relevant to the particular objects in question. For example, environmental aesthetics parallels environmental ethics in its rejection of anthropocentric models for the moral assessment of the natural world, replacing them with models based on science, most notably ecology (see Environmental ethics).

Lastly, if approaches such as the scientific one described above endow aesthetic appreciation of nature with a degree of objectivity, they might have some practical relevance in a world increasingly engaged in landscape and environmental assessment. Individuals making such assessments are often reluctant to acknowledge the importance of aesthetic considerations, regarding them at worst as totally subjective or based at best on culturally relative artistic modes of appreciation. A scientific approach suggests the need to re-evaluate such opinions.

See also: Aesthetic attitude §§1-3; Hegel, G.W.F. §8

ALLEN CARLSON

References and further reading

Berleant, A. (1992) *The Aesthetics of Environment*, Philadelphia, PA: Temple University Press.(An elaboration of the second position mentioned in §3 of this entry.)

Berleant, A. and Carlson, A. (eds) (1998) Special Issue on Environmental Aesthetics, *Journal of Aesthetics and Art Criticism* 56.(A special issue of the journal, with a number of essays on the aesthetics of nature.)

Carlson, A. (1979) 'Appreciation and the Natural Environment', *Journal of Aesthetics and Art Criticism* 37: 267-76.(A sketch of the third position mentioned in §3 of this entry.)

(1981) 'Nature, Aesthetic Judgment and Objectivity', *Journal of Aesthetics and Art Criticism* 40: 15-27.(A discussion of the possibility of objectivity in aesthetic appreciation of nature.)

(1984) 'Nature and Positive Aesthetics', *Environmental Ethics* 6: 5-34.(A discussion of the positive aesthetics mode of nature appreciation and its relationship to science.)

Elliot, R. (1982) 'Faking Nature', *Inquiry* 25: 81-93.(An example of the first position mentioned in §3 of this entry.)

Hepburn, R.W. (1966) 'Contemporary Aesthetics and the Neglect of Natural Beauty', *British Analytical Philosophy*,ed. B. Williams and A. Montefiore, London: Routledge & Kegan Paul, 285-310.(Hepburn's classic discussion mentioned in §2 of this entry.)

Kemal, S. and Gaskell, I. (eds) (1993) *Landscape, Natural Beauty and the Arts*, Cambridge: Cambridge University Press.(A useful collection of essays by a number of the main contributors to this area of research.)

Sepänmaa, Y. (1993) *The Beauty of Environment*, Denton: Environmental Ethics Books, 2nd edn.(A discussion of a number of key issues in the field, together with a useful bibliography.)

Naturphilosophie

Naturphilosophie *refers to the philosophy of nature prevalent especially in German philosophy, science and literary movements from around 1790 to about 1830. It pleaded for an organic and dynamic worldview as an alternative to the atomist and mechanist outlook of modern science. Against the Cartesian dualism of matter and mind which had given way to the mechanist materialism of the French Encyclopedists, Spinoza's dual aspect theory of mind and matter as two modes of a single substance was favoured. The sources of this heterogeneous movement lie in the philosophy of German idealism as well as in late classicism and Romanticism. The leading figure, Schelling, assimilated and stimulated the major trends and ideas through his work.*

After the death of Hegel (1831) and of Goethe (1832), Naturphilosophie *quickly disappeared from the mainstream. Yet it survived in various different forms, especially as an undercurrent of German culture and science, until the twentieth century.*

1 Kant and Fichte

The leading role among the various trends in *Naturphilosophie* was played by Schelling's philosophy of nature. To understand it, it is necessary to examine also the philosophy of Kant and subsequently that of Fichte.

Kant had a threefold influence on the movement of *Naturphilosophie*: through his *Metaphysische Anfangsgründe der Naturwissenschaft (Metaphysical Foundations of Natural Science)* (1786) where he took forces as the defining properties of matter; through his third critique, the *Kritik der Urteilskraft (Critique of Judgment)* (1790), where he treated the concept of purpose in nature, and finally in providing a frame for these ideas through the transcendental idealism of his first critique, the *Kritik der reinen Vernunft (Critique of Pure Reason)* (1781/87).

In his *Metaphysical Foundations* Kant tried to clarify the concept of matter as it has to be presupposed by all possible scientific investigation. In contesting Newton's views, he saw matter as constituted by the opposing forces of attraction and repulsion. Matter fills space by its repulsive force. The resulting dispersion is counteracted and limited by an attractive force. Thus Kant arrives at a dynamic view of matter which takes active opposing forces as fundamental and which avoids the conception of the atom and of empty space. His theory advanced a dynamic view of nature ('dynamism') as a research programme in physics, and made similar conceptions of the sciences appear in a new light (for example, positive and negative forces in magnetism and electricity).

In the *Critique of Judgment*, Kant tried to come to terms with the teleological structure of nature. Organisms as products of nature cannot be completely explained in the causal way in which one would explain a mechanical system. For in organisms not only do the parts determine the whole, but the purpose guiding the organism as a whole also determines its parts. In understanding organism and nature as a whole we have thus to grasp it as fulfilling a purpose, to see it as a self-organizing system which has the cause and effect of its action in itself (Kant 1790: §64). However, since we can talk of purposes only in relation to free action and since nature is not an intelligent being we can ascribe purposes to nature only in a metaphorical, regulative sense, *as if* it were a free acting creature with its own purposes. Our grasp of organic nature is therefore inherently subjective and can never reach the objective validity of explanations by mechanical causality. Later *Naturphilosophen* attempted to overcome this Kantian limitation and to furnish nature with an objective, purposive structure which is constitutive of it, and thus to see nature as an autonomous and even indeterministic system.

In the *Critique of Pure Reason*, Kant had claimed that experience is not passively received by our mind but is determined and even produced by categories intrinsic to our own understanding. The unity of our self-consciousness, the synthetic apperception, as the 'highest point' of philosophy is responsible for our sense impressions being synthesized into genuine knowledge. Thus every philosophy has to start with an inquiry into the transcendental conditions of the mind itself which make knowledge possible.

Before he became acquainted with Kant's writings, Johann Gottlieb Fichte believed determinism to be unavoidable. His view changed dramatically, however, when he realized that Kant's transcendental idealism can be radicalized so as to make room for absolute human freedom. Nature is a system of necessity only because it is conceived by us that way. In Kant, the unity of self-consciousness was only the *formal* condition of our knowledge; some material must be given to our senses which ultimately stems from the thing-in-itself. For Kant,

the process of being affected by objects does not depend on the self. Fichte, however, came to reject the idea of the thing-in-itself and claimed that the unity of self-consciousness is also the *material* condition of our knowledge. The self with its capacity for synthesizing knowledge is not reacting to something given, it is autonomously producing its reality, including itself. Knowledge of objects is derived from an initial act of self-positing; objects are the result of a law-governed production of the ego. The ‘I’, the self, is in this case not to be thought of as the consciousness of an individual person, but as a general structure independent of individual realization.

2 Schelling and the *Naturphilosophen*

Schelling originally started in complete harmony with Fichte’s viewpoint, but later gradually shifted away from it. Whereas Fichte stressed the certainty of which the I disposes in positing itself, Schelling tried to uncover the origin of the I, the kind of being from which it stems. ‘*I am*! My I comprises a being which precedes all thinking and conceiving’ (Schelling [1795] 1856-61, vol. 1: 167). This being is now seen as more comprehensive than the ego: namely nature. The I is not its own product, as Fichte would have it, but the I and the material objects are produced by nature as a ground deeper than the I. ‘Nature’s reality stems from herself - she is her own product - a whole which is organized out of itself and organized by itself’ (Schelling [1799] 1856-61, vol. 3: 17). With this move, Schelling could ascribe to nature the structure which Fichte had claimed for the I. Schelling also thought that this would make it possible to overcome the limitations which Kant thought were necessary in our grasp of nature. Nature can now be seen as a reality which is a self-determining system for which purpose is a constitutive idea and not a regulative one as Kant had it.

From 1797 to 1806 Schelling made various attempts to formulate his philosophy of nature. This necessarily brought him into conflict with Fichte’s soulless mechanical conception of nature. For Fichte nature was just the ‘system of necessity’, the inanimate realm of the ‘not-I’ or ‘non-ego’, posited by the I itself. Schelling, however, envisaged nature as a vital process from whose activity the consciousness of the ego and the mind emerges. Whereas Fichte asked how the self gets at nature, Schelling had the problem of how nature finally arrives at the I.

At the beginning of his development away from Fichte, Schelling envisaged *Naturphilosophie* as being of equal status with transcendental philosophy; from 1801 onwards, however, *Naturphilosophie* became primary for Schelling and independent from it. He claimed that we cannot grasp nature by merely investigating her products as they present themselves to our experience. This would be to see nature merely as an inanimate object, as *natura naturata*. Nature as ‘productivity’, or ‘absolute activity’, in its ‘infinite becoming’ requires studying her as *natura naturans*, as a subject, a productive process which eventually gives rise to mind and crystallizes in its products (Schelling [1799] 1856-61, vol. 3: 284). This view eventually flowed into a critique of Newtonian mechanical physics taken as the paradigm of modern science. Mechanism starts from material bodies which are supposed to have the property of attracting and/or repulsing each other. Yet, in order to conceive nature as an entity that is able to produce the self and the objects in like manner, it is necessary to give attraction and repulsion a more basic status and to take them as defining properties from which material objects and also consciousness are derived. Schelling did not attack the experimental and empirical side of science but its premature objectivism which leaves out the subjective element in nature and its developmental aspect.

In carrying out his programme Schelling was influenced by the experimental physicist and founder of electrochemistry, Johann Wilhelm Ritter (1776-1810), and influenced him in turn. Ritter pleaded for a strong connection of galvanism and chemistry and saw the living organism as a system of nested galvanic chains. When he found that the galvanic process is not necessarily bound to organic material, he considered galvanic action to be the key to the activity of the whole universe.

Hegel was the other post-Kantian idealist who occupied himself with *Naturphilosophie*. He was decisively stimulated by Schelling before he came to work out a fundamentally different system. For Hegel, nature as the realm of the idea in its otherness is not to be grasped as an evolutionary process but as a conceptual development of the idea on the way to self-knowledge.

3 Major doctrines of *Naturphilosophie*

Schelling’s *Naturphilosophie* set the tone for many similar undertakings. Although these attempts sometimes differ very much in detail from each other, we can distinguish several recurrent tenets. First, and foremost, there is the

view of the 'unity of mind and matter'. Schelling and many *Naturphilosophen* argued that nature is animated, that there is an original unity or 'identity' of nature and mind which allows us to infer nature's laws from the laws of the mind and vice versa: 'The system of nature is at the same time the system of our mind' and 'Nature must be visible spirit, and mind invisible nature' (Schelling [1797] 1856-61, vol. 2: 39, 56). This view was primarily directed against both Cartesian dualism and eighteenth-century materialism. It had a great impact on the Romantic movement, both in literature and science. Hans Christian Ørsted, for example, who discovered electromagnetism in 1820, claimed that all objects are materialized ideas. And according to the biologist Lorenz Oken, it was the task of *Naturphilosophie* 'to show that the laws of the mind are not different from the laws of nature but that both represent each other. Philosophy of nature and philosophy of mind range, therefore, parallel to each other' (Oken 1843: §13). Later, one of the sharpest critics of *Naturphilosophie*, Hermann von Helmholtz saw the principal error of this movement precisely in the mingling of the necessary laws of nature with the spontaneous activity of mind. A last remnant of the identity view of nature and mind can be found in the dual aspect view of mind and body as it was embraced by a large number of physiologists, psychologists and many philosophers up to and including Rudolf Carnap's *Der Logische Aufbau der Welt (The Logical Construction of the World)* (1928).

A second doctrine is the prevalence of the organism as an explanatory model. For Schelling (as for Kant), mechanical explanations of nature have to be supplemented by organic ones. It is not possible, Schelling claimed, for the universe to be simply a causal mechanism as its organic structure is logically prior to its mechanical one. 'One and the same principle' is at work in both organic and inorganic nature which binds all of it 'into a general organism' (Schelling [1798] 1856-61, vol. 2: 350). Organisms are causally self-contained, they are their own cause and effect and they are self-movable, whereas causal mechanisms can only be moved from the outside and are therefore essentially passive. Carl Friedrich Kielmeyer, who decisively influenced Schelling in his organic thought, explained the organism in terms of three organic forces: irritability, reproduction and sensibility (Kielmeyer 1793). Irritability is organic activity or movement, sensibility is organic receptivity or sensation, and reproduction refers to the metabolism, to growth and procreation.

A third related doctrine of *Naturphilosophie* is the unity of nature and its forces, and thereby also of science. There was a strong tendency to exhibit nature in all its different aspects as a development from a small number of fundamental forces. Ørsted, whose dissertation had examined Kant's *Metaphysical Foundations of Natural Science*, had claimed that there is a 'unity of all the forces' of heat, chemistry, electricity and magnetism and that 'the former physical results [*Kenntnisse*] thus combine into a physics of one piece' (Ørsted 1803: 209). Similar claims have been made by scientists such as Faraday, Mohr, Grove, J. R. Mayer and others, which eventually led to the establishment of the energy conservation principle. The inorganic as well as the organic realms were seen as directed by the same principles, so that there is no essential difference between them. This does not mean that nature reduces to mechanism, but that the inorganic world has to be conceived as the product of an inherent organic natural process. *Naturphilosophie* developed the tendency to impose the organic categories found in the study of nature even on culture, society and the state. In this sense, it also helped to prepare the ground both for the later materialist and naturalist movement as well as for organic conceptions of the state and society.

In viewing nature as a unity, the movement of *Naturphilosophie* was also influenced by Goethe's search for an underlying idea of the organic and inorganic realms of nature. In morphology, he argued that all plants are modifications of one primordial plant, the *Urpflanze*, and that all plant organs were variations of a single fundamental organ, the leaf (see Goethe, J.W. von §3).

Another doctrine of *Naturphilosophie* concerns the developmental aspect of nature. Instead of viewing nature and its parts in static terms, *Naturphilosophie* tried to conceive of it as the outcome of an evolutionary graduated succession. Very often this process was seen as being initiated and maintained by a polar antagonism of forces. Models for conceiving this process were either taken by way of analogical reasoning from some philosophical source (Kant's *Metaphysical Foundations*, Fichte's idea of the self determining itself by positing the I and the not-I at the same time) or some scientific theory of the time (magnetism, galvanism, electricity, chemistry, physiology). Other sources included Goethe and Herder who also held influential theories of development. In all this, the concept of polarity played a prominent role and was seen as the continuous principle for every development and activity. Oken, for example, viewed living organisms as governed by two opposing processes: an individualizing one which increases the organism's vitality, and a universalizing one that leads to death and is a drive to the absolute (see Absolute, the).

See also: Romanticism, German §6; German idealism

MICHAEL HEIDELBERGER

References and further reading

Amrine, F. (ed.) (1986) *Goethe and the Sciences: A Reappraisal*, Dordrecht: Kluwer. (Goethe in the history of science and his relevance to philosophical treatment of nature today. Annotated bibliography.)

Bonsiepen, W. (1997) *Die Begründung einer Naturphilosophie bei Kant, Schelling, Fries und Hegel. Mathematische versus spekulative Naturphilosophie (Foundations of the philosophy of nature in Kant, Schelling, Fries and Hegel: mathematical versus speculative philosophy of nature)*, Frankfurt: Klostermann. (Detailed and thorough account of German idealism's different philosophies of nature and of the opposing mathematical philosophy of nature of J.F. Fries and his school. Much philosophical and historical context.)

Böhme, G. (ed.) (1989) *Klassiker der Naturphilosophie. Von den Vorsokratikern bis zur Kopenhagener Schule (Classics of the philosophy of nature: from the Presocratics to the Copenhagen School)*, Munich: Beck.(Essays on the leading philosophers of nature and philosophical cosmologists, from the Presocratics to Einstein and the Copenhagen interpretation of quantum mechanics. Includes entries on Kant, Goethe, Schelling, and Hegel.)

Caneva, K.L. (1997) 'Physics and *Naturphilosophie*: A Reconnaissance', *History of Science* 35: 35-107.(Critical and thorough review of claims for the influence of *Naturphilosophie* on early nineteenth-century physics. Valuable and extensive bibliography.)

Carnap, R. (1928) *Der logische Aufbau der Welt*, Berlin-Schlachtensee: Weltkreis-Verlag; trans. R. George as *The Logical Structure of the World*, Berkeley, CA: University of California Press, 1967.(Attempt to carry out the programme of early Logical Empiricism to reduce all concepts to elementary experiences. Gives a solution to the mind-body problem in terms of a double-aspect theory that is reminiscent of the Spinozism of *Naturphilosophie*.)

Cohen, R.S. and Wartofsky, M.W. (eds) (1984) *Hegel and the Sciences*, Dordrecht: Reidel.(Essays on Hegel's philosophy of nature, its relation to logic and on his treatment of various sciences.)

Cunningham, A. and Jardine, N. (eds) (1990) *Romanticism and the Sciences*, Cambridge: Cambridge University Press.(Essays on Romanticism in general and on its relation to the sciences of the organic and inorganic, as well as on Romantic literature and the sciences.)

Engelhardt, D. von (1978) 'Bibliographie der Sekundärliteratur zur romantischen Naturforschung und Medizin 1950-1975' (Bibliography of secondary literature on Romantic science and medicine, 1950-1975), in R. Brinkmann (ed.) *Romantik in Deutschland. Ein interdisziplinäres Symposion (Romanticism in Germany: An interdisciplinary symposium)*, Stuttgart: Metzler, 307-30.(Comprehensive bibliography for the given period. The volume contains other relevant material.)

Engelhardt, D. von (1988) 'Romanticism in Germany', in R. Porter and M. Teich (eds) *Romanticism in National Context*, Cambridge: Cambridge University Press, 109-33.(Places *Naturphilosophie* in the wider context of the German Romantic movement.)

Engelhardt, D. von and Musto, R. (eds) (1993) *Annali. Sezione Germanica* NS 3: 1-3. (A collection of nineteen essays, in Italian, on a wide range of topics concerning science and philosophy in the age of Romanticism; the collection has no special title.)

Esposito, J.L. (1977) *Schelling's Idealism and Philosophy of Nature*, Lewisburg, PA and London: Associated University Presses.(Short account of Schelling's philosophy of nature. Useful chapter on Schelling's contemporary critics and on his influence on nineteenth-century America)

Fichte, J.G. (1794) *Grundlage der gesammten Wissenschaftslehre*, Jena/ Leipzig: Gabler; repr. in *Sämmtliche Werke*, ed. I.H. Fichte, Berlin: Veit, 1845-6, vol. 1, 83- 328; trans. P. Heath and J. Lachs as *The Science of Knowledge*, Cambridge: Cambridge University Press, 1982.(Theoretical foundation of Fichte's philosophical system.)

Fischer, K. (1902) *Schelling's Leben, Werke und Lehre (Schelling's life, work, and doctrine)*, Heidelberg: Winter, 3rd edn.(Elaborate and vivid account of Schelling's life and work. Sometimes dated but still valuable, with a wealth of information.)

Frank, M. (1985) *Eine Einführung in Schellings Philosophie (An introduction to Schelling's philosophy)*, Frankfurt: Suhrkamp.(Short but useful introduction to Schelling's work which puts stress on his *Naturphilosophie*.)

Giovanni, G. di (1979) 'Kant's Metaphysics of Nature and Schelling's *Ideas for a Philosophy of Nature*', *Journal of the History of Philosophy* 17: 197-215.(Argues for the closeness of Schelling's *Ideen* to Kant.)

Gloy, K. and Burger, P. (eds) (1993) *Die Naturphilosophie im Deutschen Idealismus (The philosophy of nature in German idealism)*, Stuttgart: Frommann-Holzboog.(Collection of essays on German idealism's concept of

nature: on Kant, Fichte, Schelling, Hegel and the Romantics.)

Gode-von Aesch, A. (1941) *Natural Science in German Romanticism* New York: Columbia University Press; repr. New York: AMS Press, 1966.(Deals with science and literature as two intertwined aspects of the German Romantic movement.)

Gower, B. (1973) 'Speculation in Physics: The History and Practice of *Naturphilosophie*', *Studies in History and Philosophy of Science* 3 (4): 301-56. (Authoritative and influential survey which includes treatment of Ritter and Ørsted.)

Hasler, L. (ed.) (1981) *Schelling. Seine Bedeutung für eine Philosophie der Natur und der Geschichte (Schelling: his significance for a philosophy of nature and history)*, Stuttgart: Frommann-Holzboog.(Includes essays on Schelling's relation to the sciences of his day and on his philosophy of history.)

Heckmann, R., Krings, H. and Meyer, R.W. (eds) (1985) *Natur und Subjektivität. Zur Auseinandersetzung mit der Naturphilosophie des jungen Schelling (Nature and subjectivity: a discussion of the young Schelling's Naturphilosophie)*, Stuttgart: Frommann-Holzboog.(Essays on the place of Schelling's *Naturphilosophie* in German idealism, its reception by its contemporaries and an assessment from a present-day point of view.)

Hegel, G.W.F. (1817) 'Naturphilosophie', part 2 of *Enzyklopädie der philosophischen Wissenschaften im Grundrisse*, Heidelberg: Oßwald; 2nd edn repr. in *Gesammelte Werke*, Hamburg: Meiner, 1989, vol. 19, 183-283; trans. M.J. Petry as *Hegel's Philosophy of Nature*, London: Allen & Unwin, 1970, 3 vols.(Hegel's major work on his own specific version of *Naturphilosophie*.)

Helmholtz, H. (1855) *Ueber das Sehen des Menschen (On human vision)*, Leipzig: Voss; repr. in *Vorträge und Reden*, 5th edn, 2 vols, Braunschweig: Vieweg, 1903, vol. 1: 85-117.(One of the first documents of rising Neo-Kantianism. Kant's epistemology, interpreted in terms of sensory physiology, is proposed as a remedy for the extravagances of *Naturphilosophie*.)

Heuser-Keßler, M.-L. (1986) *Die Produktivität der Natur. Schellings Naturphilosophie und das neue Paradigma der Selbstorganisation in den Naturwissenschaften (Nature's productivity: Schelling's philosophy of nature and the paradigm of self-organization in the natural sciences)*, Berlin: Duncker & Humblot. (Account of Schelling's conception of (especially organic) nature and its autogeneric ordering processes. Schelling's work is related to conceptions of self-organization in present-day science, in the work of I. Prigogine and H. Haken.)

Heuser-Keßler, M.-L. and Jacobs, W.G. (eds) (1994) *Schelling und die Selbstorganisation. Neue Forschungsperspektiven (Schelling and self-organization: new research perspectives)*, *Selbstorganisation*, vol. 5, Berlin: Duncker & Humblot. (Deals with Schelling's concept of self-organization, its contemporary and present-day reception.)

Horstmann, R.-P. and Petry, M.J. (eds) (1986) *Hegels Philosophie der Natur. Beziehungen zwischen empirischer und spekulativer Naturerkenntnis (Hegel's philosophy of nature: relations of empirical and speculative natural science)*, Stuttgart: Klett-Cotta.(Essays on Hegel's *Naturphilosophie*.)

Kant, I. (1781/1787) *Kritik der reinen Vernunft*, Riga: Hartknoch; trans. N.K. Smith as *The Critique of Pure Reason*, New York: St Martin's Press, London: Macmillan, 1964.(Explanation and justification of the a priori forms of reason that constitute the basis of any theoretical knowledge. Kant's doctrine of the spontaneity of the subject and the autonomy of reason was especially relevant for *Naturphilosophie*.)

Kant, I. (1786) *Metaphysische Anfangsgründe der Naturwissenschaft*, Riga: Hartknoch; trans. J. Ellington as *Metaphysical Foundations of Natural Science*, Indianapolis, IN: Bobbs-Merrill, 1970.(In his metaphysics of nature, Kant formulates the principles that specify the transition from the critique of pure reason to actual empirical science. His dynamic definition of matter from attractive and repulsive forces was important for *Naturphilosophie*.)

Kant, I. (1790) *Kritik der Urteilskraft*, Berlin: Lagarde; trans. J.C. Meredith as *Kant's Critique of Judgement*, Oxford: Clarendon Press, 1952.(In a way, the cornerstone of Kant's philosophy. Tries to show that the realms of nature and freedom, of causality and teleology are compatible. The second part, 'Critique of teleological reason', became especially relevant for *Naturphilosophie*.)

Kanz, K.T. (ed.) (1994) *Philosophie des Organischen in der Goethezeit. Studien zu Werk und Wirkung des Naturforschers Carl Friedrich Kielmeyer (1765-1844) (Philosophy of the organic during the time of Goethe: studies in the work and influence of the naturalist Carl Friedrich Kielmeyer)*, Stuttgart: Steiner.(Essays on Kielmeyer's vision of the organism and biology in general, its context and influence.)

Kielmeyer, C.F. (1793) *Ueber die Verhältnisse der organischen Kräfte unter einander in der Reihe der verschiedenen Organisationen, die Geseze und Folgen dieser Verhältnisse (On the relation of the organic forces to each other in the series of the different organisms, the laws and the successions of this relation)*, Stuttgart and Tübingen: Heerbrand; repr., intro. K.T. Kanz, Marburg: Basilisken-Presse, 1993. (Living systems fulfil three main functions: sensation, motion and self-preservation. Organic forms differ from each other by the

different amount in which these functions are fulfilled. A concise statement of the problem situation at the end of the eighteenth century.)

Knittermeyer, H. (1928) *Schelling und die romantische Schule (Schelling and the Romantic movement)*, Munich: Ernst Reinhardt.(Lengthy account of Schelling's *Naturphilosophie* and of the ideas of many of his followers, especially from philosophy.)

Kuhn, Th. S. (1959) 'Energy Conservation as an Example of Simultaneous Discovery', in M. Clagett (ed.) *Critical Problems in the History of Science*, Madison, WI: University of Wisconsin Press, 321-56; repr. in *The Essential Tension. Selected Studies in Scientific Tradition and Change*, Chicago, IL: University of Chicago Press, 1977, 66-104.(In this influential piece, Kuhn shows how *Naturphilosophie* served as a philosophical background for the simultaneous discovery of energy conservation by many different scientists. Compare the comments by Caneva 1997.)

Mutschler, H.-D. (1990) *Spekulative und empirische Physik. Aktualität und Grenzen der Naturphilosophie Schellings (Speculative and empirical physics: actuality and limits of Schelling's Naturphilosophie)*, Stuttgart: Kohlhammer.(A critique of Schelling's *Naturphilosophie* from a present-day point of view.)

Oken, L. (1843) *Lehrbuch der Naturphilosophie*, Zurich: Schultheß, 3rd. edn; repr. Hildesheim: Olms, 1991; trans. A. Tulk as *Elements of Physiophilosophy*, London: Ray Society, 1847.(Exposition of a *naturphilosophisches* system of the life sciences; a mixture of an unscrupulous usage of polarity principles and of an effusive naturalism.)

Ørsted, H.Ch. (1803) *Materialien zu einer Chemie des Neunzehenten Jahrhunderts (Materials for a chemistry of the nineteenth century)*, Erstes Stück, Regensburg: Montag & Weiß; repr. in *Naturvidenskabelige Skrifter*, ed. K. Meyer, Copenhagen: Høst 1920, vol. 1, 133-210.(Enthusiastic exposition of the chemical system of J.J. Winterl, which claimed the principles of electricity as fundamental for acids, bases, heat, light and magnetism.)

Ørsted, H.Ch. (1849) *Aanden i naturen*; trans. K.L. Kannegiesser as *Der Geist in der Natur*, Leipzig: Lorck, 1850; trans. L. Horner and J.B. Horner as *The Soul in Nature*, London: Bohn, 1852, repr. London: Dawsons of Pall Mall, 1966.(Series of articles exhibiting physical nature as permeated and dominated by the mind.)

Petry, M.J. (ed.) (1987) *Hegel und die Naturwissenschaften (Hegel and the natural sciences)*, Stuttgart: Klett-Cotta.(Collection of essays expounding Hegel's relation to the sciences.)

Poggi, S. (1996) 'La "Naturphilosophie" dell'idealismo e del romanticismo: lo stato della ricerca', (The Naturphilosophie of idealism and of Romanticism: The state of research), *Rivista di Filosofia* 87 (1): 111-28. (Concise overview of contemporary research on *Naturphilosophie*.)

Poggi, S. and Bossi, M. (eds) (1994) *Romanticism in Science: Science in Europe, 1790-1840*, Dordrecht: Kluwer. (Essays on Romanticism and *Naturphilosophie* in biology, chemistry, psychology, philosophy and mathematics.)

Sandkühler, H.J. (ed.) (1984) *Natur und geschichtlicher Prozeß. Studien zur Naturphilosophie F.W.J. Schellings (Nature and historical development: studies in Schelling's Naturphilosophie)*, Frankfurt: Suhrkamp.(Collection of essays on Schelling's *Naturphilosophie* and its relation to history.)

Schelling, F.W.J. (1795) *Vom Ich als Princip der Philosophie oder über das Unbedingte im menschlichen Wißen*, Tübingen: Herrbrandt; repr. in *Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart, 1856-61, vol. 1, 149-244; trans. F. Marti as *The Unconditional in Human Knowledge: Four Early Essays 1794-6*, Lewisburg, PA: Bucknell University Press, 1980.(On the absolute ego as the foundation of philosophy. A confrontation of Fichte's philosophy with that of Spinoza.)

Schelling, F.W.J. (1797) *Ideen zu einer Philosophie der Natur als Einleitung in das System dieser Wissenschaft*, Leipzig: Breitkopf & Härtel; repr. in *Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart 1856-61, vol. 2, 1-343; trans. E.E. Harris and P. Heath, intro. R. Stern, *Ideas for a Philosophy of Nature as Introduction to the Study of that Science*, Cambridge: Cambridge University Press, 1988.(Schelling's major work on *Naturphilosophie* aiming to overcome the split of mind and nature, of subject and object. Follows Kant in constructing matter from expansive and contractive forces.)

Schelling, F.W.J. (1798) *Von der Weltseele, eine Hypothese der höheren Physik zur Erklärung des allgemeinen Organismus (On the world-soul: a hypothesis of higher physics in order to explain the general organism)*, Hamburg: Perthes; repr. in *Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart 1856-61, vol. 2, 345-583. (Explanation of organic nature in terms of a general dualism of opposing forces.)

Schelling, F.W.J. (1799) *Erster Entwurf eines Systems der Naturphilosophie (Outline of a system of the philosophy of nature)*, Jena/Leipzig: Gabler; repr. in *Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart 1856-61, vol. 3, 1-268.(Elaboration of a developmental theory of life as proposed in the *Weltseele*. Influenced by Kielmeyer and Herder.)

Schelling, F.W.J. (1994) *Ergänzungsband zu Werke Band 5 bis 9. Wissenschaftshistorischer Bericht zu Schellings*

Naturphilosophischen Schriften 1797-1800 (Supplementary volume of vols 5 to 9 of the works: The scientific background of Schelling's works on Naturphilosophie 1797-1800), Stuttgart: Frommann-Holzboog. (Supplement to the critical edition of the Bavarian academy of the sciences (1976-) ed. H.M. Baumgartner *et al*. Essays on chemistry, magnetism, electricity, galvanism and physiology in Schelling's day aiming to recover the contemporary scientific context of Schelling's *Naturphilosophie*.)

Schmied-Kowarzik, W. (1993) 'Selbst und Existenz. Grundanliegen und Herausforderung der Naturphilosophie Schellings' (Self and existence: the challenge of Schelling's Naturphilosophie), in H.M. Baumgartner and W.G. Jacobs (eds) *Philosophie der Subjektivität? Zur Bestimmung des neuzeitlichen Philosophierens (Philosophy of subjectivity? on the characteristics of modern philosophy)*, Stuttgart: Frommann-Holzboog, vol. 1, 111-30.(Thoughtful defence of Schelling's approach and an appreciation of its relation to modern science.)

Snelders, H.A.M. (1970) 'Romanticism and Naturphilosophie and the Inorganic Natural Sciences 1797-1840: An Introductory Survey', *Studies in Romanticism* 9 (3): 193-215.(A clear presentation of the influence of Romantic thinking and *Naturphilosophie* on the inorganic sciences in Germany. Claims that the influence was positive and fruitful.)

Snelders, H.A.M. (1994) *Wetenschap en intuitie: Het Duitse romantisch-speculatief natuuronderzoek rond 1800 (Science and intuition: German Romantic and speculative natural science around 1800)*, Baarn: Ambo. (Balanced and authoritative account of the impact of *Naturphilosophie* on science and vice versa from about 1790 to 1830. Includes vivid and expressive accounts of central and marginal figures of the time. Helpful annotated bibliography.)

Strack, F. (1994) *Evolution des Geistes: Jena um 1800. Natur und Kunst, Philosophie und Wissenschaft im Spannungsfeld der Geschichte (Evolution of the spirit: Jena around 1800. Nature and art, philosophy and science under the tension of history)*, Stuttgart: Klett-Cotta.(Essays on the most brilliant period of the University of Jena as the origin of the German Romantic movement in philosophy, art, literature and the sciences. Includes essays on philosophical ideas about nature in Goethe, Hölderlin, Novalis, Schelling and Ritter.)

Wilson, A.D. (1997) 'Die romantischen Naturphilosophen' (The romantic Naturphilosophen), in K. von Meyenn, *Die grossen Physiker: Erster Band: Von Aristoteles bis Kelvin (Great Physicists. Vol. 1: From Aristotle to Kelvin)*, Munich: Beck, 319-35.(The role of *Naturphilosophie* in the physics of Ritter, Ørsted and Faraday.)

Necessary being

Many think that God is perfect, or free from defect, and that being able not to exist is a defect. These infer that God is not able not to exist - that is, that God exists necessarily. Some add that what makes God perfect also makes him exist necessarily, and so trace his necessity to his immateriality (Aristotle), eternity (Plotinus) or simplicity (Aquinas). Others trace God's necessity to his relation to creatures (Ibn Sina, Anselm). Spinoza and Leibniz held that what makes God necessary explains his very existence.

Many have thought that if God exists necessarily, there is a sound ontological argument for God's existence, or that if there is a sound ontological argument for God's existence, God exists necessarily. But both claims are false. Some have used philosophical views of the nature of necessity - for example, that all necessity is conventional, a matter of how we choose to use words - to challenge God's necessary existence. But the theories which best support these challenges have fallen from favour, and in fact, even if one accepts the theories, the challenges fail.

1 Contingent and necessary existence

Our lives depend on many things. Had our parents not had us, we would not exist. Were there not enough food, or had we the bad luck to be there when a bomb goes off, we would not continue to exist. Our lives are thus contingent on parents, food and other things. To be contingent in this way is to be vulnerable. Such vulnerability seems a defect. So if we think God perfect, we will deny that God exists contingently.

If God exists and does not exist contingently, then:

(GN) God exists necessarily.

Thus (GN) emerges when we think about God's perfection. (GN) asserts that God exists, could not have failed to exist and cannot fail to exist. Equivalently, (GN) asserts that God exists and would always exist, no matter what. One can read (GN) as:

(GN1) the proposition 'God exists' is necessarily true,

or

(GN2) there necessarily is some individual or other who is God,

or

(GN3) whoever is God exists necessarily,

or

(GN4) the individual who is God exists necessarily.

If

(GN5) whoever is God is necessarily God

is false, (GN1) and (GN2) are equivalent, but logically independent of (GN3) and (GN4), and (GN3) entails (GN4), but (GN4) does not entail (GN3). If (GN5) is true, (GN1)-(GN4) are equivalent.

(GN) uses the concept of necessity. So as philosophers' grasp of necessity changed, what they meant by (GN) changed. But the impulse to ascribe maximal perfection to God always pushed towards claiming for God's existence the strongest sort of necessity a philosopher's modal theory would allow. Sometimes stronger doctrines of divine necessity subsume weaker ones; for instance, a God necessary in Leibniz's sense is also necessary in Aristotle's.

2 The evolution of divine necessity

Aristotle (§16) held that God exists throughout infinite time (*Metaphysics* XII, 7), and that:

(1) if something exists throughout infinite time, it exists necessarily (*Generation and Corruption* II, 11).

Aristotle rests (1) on a claim that any *x* is possibly *F* only if something is able to make it *F*, and a principle of plenitude, that any genuine ability (at least of certain sorts) sooner or later is used (*On the Heavens* I, 11). Thus for Aristotle, if a thing can cease to exist, it sooner or later does so. So if something exists throughout infinite time, it must be unable to cease to exist. A thing which exists and cannot not exist exists necessarily. Aristotle thinks that God's immateriality explains his necessity (*Metaphysics* XII, 6): things can cease to exist only if made of matter, for a thing *A* ceases to exist only if the matter making up *A* loses the constituting ('substantial') form whose presence in that matter makes *A* exist.

Plotinus (§3) asserts (GN) of *Nous*, a being like Aristotle's God (*Enneads* VI). *Nous* is atemporal due to its fullness and perfection of being (*Enneads* III). So *Nous* has no potentialities, since potentialities for Plotinus as for Aristotle are abilities to *become* different and so require time to be actualized (*Enneads* II). If *Nous* has no potentialities, it has no potentiality not to exist, and so exists necessarily.

Ibn Sina (§5), by contrast, held that (GN) 'means that God is a being without a cause, and…is the cause of other than himself' (*al-Risalat al-'arshiya*, 1951: 32). Other Muslim and Jewish philosophers added that God needs no cause or cannot be caused to exist. So read, (GN) does not entail that God or God's nature explains God's existence, or that anything explains God's existence, or therefore that there is an explanation for the fact that there is something rather than nothing at all.

This remains true even for Aquinas (§11), who held that 'God exists' is 'self-evident in itself' (*Summa theologiae* Ia, 2, 1), a status which entails broadly logical necessity. For that 'God exists' is self-evident in itself entails only that its truth would be self-evident to one who perceived God's nature, not that the perceiver would see why it is true or that God's nature would explain that truth. For Aquinas, 'God exists' is 'self-evident in itself' due to the doctrine of divine simplicity (see Simplicity, divine §§1-2). Given God's simplicity, God, his nature and his existence are identical (*Summa theologiae* Ia, 3, 3 and 4). If:

(2) God = God's existence,

Aquinas reasons, the subject and predicate of 'God exists' have the same reference. Thus 'God exists' is in effect an identity-claim, 'self-evident in itself' as all identities are: 'the reason for the necessity of God's existence is that His essence is His existence' (*Quaestiones disputatae de veritate*, 3, 1, *ad*. 7).

Thus Aquinas in effect traces God's necessity of existence to the necessity of identity. The necessity of identity is logical: statements of the form '$A = A$' instance the logical law of identity $(\forall x)(x = x)$, and so their contradictories are inconsistent with theorems of logic. Therefore even if Aquinas did not have this concept of necessity or mean to ascribe it to God, there is a sense in which, for Aquinas, God's existence is logically necessary. It may be that this holds for divine simplicity's many earlier medieval friends as well. Note that in Thomas' hands, (GN) does not *explain* God's existence. Divine simplicity makes (GN) true by entailing (2). But (2) *presupposes* that God exists. (If there is no Sherlock Holmes, 'Holmes = Holmes' is not true.) So divine simplicity and (GN) cannot explain God's existence.

Anselm's treatment of (GN) begins from the claim that God's will is the source of all necessity (*Cur Deus homo* II, 18). If it is, God's will makes God's existence necessary. How can it do so? To Anselm, 'necessarily, God is *F*' really says that nothing distinct from God can make it the case that God is not *F* (*Cur Deus homo* II, 18). If so, to Anselm, (GN) means 'nothing distinct from God can make it the case that God does not exist'. God makes this claim true if he creates only creatures who lack the requisite power, or if he does not create. For Anselm, what is so by necessity, properly speaking, is so because of compulsion, external constraint or some genuine inability (*Cur Deus homo* II, 5). So for Anselm, (GN) is an indirect way to say that God's power to exist cannot be overcome. For Aquinas, (GN) rests on something about God's own makeup. For Anselm, (GN) is true due to God's relations to creatures. For neither does (GN) explain God's existing (see Anselm of Canterbury §§2-4, 8).

Duns Scotus (§§7-11) popularized the view that a state of affairs is possible just in case it does not entail a contradiction, and is necessary just in case its negation entails a contradiction (*Commentaria oxoniensa* I, 36). This

view evolved into our current concept of broadly logical necessity. So like many writers after Scotus, Descartes took (GN) to assert God’s broadly logical necessity. Descartes held both the doctrine of divine simplicity and a theory of God’s responsibility for modal facts which may have been more radical than Anselm’s (see Plantinga 1980). So for Descartes, (GN) is overdetermined. As Descartes held that God causes the truth of logically necessary (‘eternal’) truths, being logically necessary did not for him entail being uncreated or wholly independent. Thus the meaning Descartes gave (GN) was wholly discrete from the meaning it had for such medievals as Ibn Sina.

In other Continental rationalists, (GN) also took on a new import. Spinoza (§§2-3) held that each existing thing’s existence must be explained (*Ethics* I, 8, n.2). In the case of God, who is necessarily uncaused, only his intrinsic nature can provide this explanation. So Spinoza infers that God’s nature entails and explains his existence (*Ethics* I, 11). Spinoza adds that ‘existence…insofar as it is conceived to follow necessarily from the definition alone of the [existing thing]…is conceived as an eternal truth’ (*Ethics* I, df.8). Spinoza’s eternal truths, like Descartes’, are broadly logically necessary. So for Spinoza, something exists necessarily only if and because its nature entails and explains its existence, and (GN) asserts or entails that such an explanation exists. Leibniz (§§3, 7) holds a principle of sufficient reason even stronger than Spinoza’s (1717: 16), and reasons much as Spinoza does (1697: 85-6). Neither Leibniz nor Spinoza held to divine simplicity as Aquinas and other medievals understood it. Leibniz held the eternal truths to depend on God’s intellect (*Theodicy*, 184), and so also divorced necessity and independence.

3 Divine necessity and the ontological argument

Some writers (for example, Penelhum 1971: 183-4) nearly equate (GN) with the claim that there is a sound ontological argument for God’s existence (see God, arguments for the existence of §§2-3). In this they follow Spinoza: ‘God is substance, which necessarily exists, i.e. to whose nature it pertains to exist, or (what is the same) from whose definition it follows that He exists’ (*Ethics* I, 19).

But there could be a sound ontological argument even if (GN) were false. Some such arguments (including Anselm’s in *Proslogion* 2) end in contingent propositions asserting God’s existence. An ontological argument is an a priori existence-proof. Descartes’ *Cogito* may try to prove a priori that a contingent being exists. If the *Cogito* does, it is not therefore absurd. In Kaplan-style indexical logics, ‘I exist’ is both contingent and an analytic or logical truth; there are also such truths in systems of modal logic employing actuality-operators (see Zalta 1988). One can hold both that it is provable a priori or even broadly logically necessary that a certain property is exemplified (as when mathematicians prove that there exist numbers with certain properties) and that the item having that property exists contingently (as on the theory that numbers are sets of sets of concrete things).

Conversely, even if (GN) is true, there might be no sound ontological argument: if (as Gödel reminds us) not every necessary truth is provable, perhaps ‘God exists’ is not. Some who take it that (GN) is true only if there is a sound ontological argument for God’s existence assume that if ‘God exists’ is necessary, ‘God does not exist’ is or entails a contradiction. But this is debatable. Arguably, both

(IO) there can be an immovable object, and
(IF) there can be an irresistible force

are necessarily true if true at all. At most one is true: if nothing can resist a certain force, the supposed immovable object cannot, and so is not after all immovable. But neither (IO) nor (IF) seems to entail a contradiction.

4 Objections to divine necessity

Some philosophers (for example, Findlay 1955) argue that:

(3) All necessary truths are true by convention.
(4) If (GN) is true, ‘God exists’ is necessarily true.
(5) So if (GN) is true, ‘God exists’ is true by convention.
(6) Merely conventional truths do not assert real facts.
(7) So if (GN) is true, ‘God exists’ asserts no real fact.

Findlay thinks (GN) a conceptual truth. So having argued for (7), he infers that ‘God exists’ asserts no real fact -

that is, that God does not really exist. Theists might instead take (3)-(7) as a *reductio* of (GN). But (3) no longer seems patently true. Further, even if conventionalism is true, (5) is not. On conventionalism, if (GN) is true, our ways of using words determine the modality of ‘God exists’ and so determine whether (GN) is true. It doesn’t follow that our usage determines whether God *exists*. Perhaps we have the power Anselm thought God has, to make God’s existence necessary but not to make God exist.

If ‘God exists’ is necessary but conventionally so, (GN) entails ‘God exists’ without explaining or grounding it. If this is the case, one has to ask why our conventions have the good luck to affirm as necessary what really does exist. But such questions are not unanswerable. (For instance, if there is a God, he might want us to believe (GN), and might design us to do so.) If (GN) is conventionally true, (GN) is not a good reason to affirm that God exists; until we find on other grounds that God exists, we have no guarantee that we do not have the wrong convention. But this does not even entail giving up all forms of ontological argument, so it has little significance.

Some philosophers (for example, Edwards 1976) argue that:

(8) if (GN) is true, ‘God exists’ is analytic, or true due to its meaning.
(9) If ‘God exists’ is analytic, it is true because its subject-concept (God) contains its predicate-concept (existence).
(10) So if ‘God exists’ is analytic, existence is a predicate.
(11) Existence is not a predicate (as Kant argued).
(12) So (GN) is false.

(8)-(11) may even imply that (GN) is conceptually incoherent, since it assigns ‘exists’ to the wrong semantic category. Edwards rests (8) on the theory that broadly logical necessity is analyticity. But this theory is now widely rejected. Rather, a proposition is broadly logically necessary if and only if true in all broadly logically possible worlds. Some such propositions seem not to be analytic - for example, that everything green has some spatial property (Adams 1987: 45-6). Some might argue, following Anselm, that ‘God exists’ is true in all possible worlds because whatever ‘possible worlds’ are, they depend on God for their existence (see Morris and Menzel 1986).

A different sort of objection to (GN) arises because (GN) implies that ‘something exists’ is necessarily true. We seem able to conceive (in some sense) there being nothing at all. Being able to conceive (in some sense) that this is true is defeasible evidence that it is possible. If possibly there is nothing, ‘something exists’ is not necessarily true. Still, there are many sorts of conceivability. They do not all have equal weight as evidence for possibility. Adams (1971: 285-6) suggests that conceiving that *P* is greater evidence that *P* is possible the closer it is to imagining the actual experiences one would have were *P* true. One cannot do this for ‘nothing exists’, for if nothing existed, there would be no experiences, or minds to have them.

See also: God, concepts of; Natural theology

BRIAN LEFTOW

References and further reading

Adams, R.M. (1971) ‘Has It Been Proven That All Real Existence is Contingent?’, *American Philosophical Quarterly* 8 (3): 284-91.(Clear reply to challenges from Findlay, Edwards and others.)

Adams, R.M. (1987) ‘Divine Necessity’, in T. Morris (ed.) *The Concept of God*, New York: Oxford University Press, 41-53.(A rich, important defence, rigorous but not technical.)

Anselm (1077-8) *Proslogion*, trans. J. Hopkins and H. Richardson, in J. Hopkins and H. Richardson (eds and trans), *Anselm of Canterbury*, Toronto, Ont.: Edwin Mellen Press, 1974, vol. 1.(*Locus classicus* of perfect-being theology.)

Anselm (completed 1098) *Cur Deus homo*, in F.S. Schmitt (ed.) *S. Anselmi opera omnia*, Edinburgh: Nelson, 1946, vol. 2.(Important text for Anselm’s theory of modality.)

Aquinas, T. (1256-9) *Quaestiones disputatae de veritate*, in *Quaestiones disputatae*, Rome: Taurini, 1927, vol. 3. (Questions 1-3 reveal much of Aquinas’ metaphysics of modality.)

Aquinas, T. (1266-73) *Summa theologiae*, New York: Benziger Brothers, 1948. (Question 2 treats the modal and epistemic status of the proposition ‘God exists’.)

Aristotle (mid 4th century BC) *On the Heavens (De caelo)*, trans. J. Stocks, in R. McKeon (ed.) *The Basic Works*

of Aristotle, New York: Random House, 1941.(His view of the physical world; ties everlasting and necessary existence.)

Aristotle (mid 4th century BC) *Generation and Corruption (De generatione et corruptione)*, trans. H.H. Joachim, in R. McKeon (ed.) *The Basic Works of Aristotle*, New York: Random House, 1941. (Analysis of coming and ceasing to exist; ties everlasting and necessary existence.)

Aristotle (mid 4th century BC) *Metaphysics*, trans. W.D. Ross, in R. McKeon (ed.) *The Basic Works of Aristotle*, New York: Random House, 1941.(Book 12 treats God's existence and nature, and ties immateriality and necessity.)

Curley, E.M. (1984) 'Descartes on the Creation of the Eternal Truths', *Philosophical Review* 93 (4): 569-97.(Fine, careful reading of Descartes.)

Duns Scotus, J. (*c*.1300-7) *Commentaria oxoniensa*, Florence: Quaracchi, 1912, vol. 1.(Distinctions 2, 36, 39 and 43 are good sources for Scotus' views of modality and God's necessity.)

Edwards, P. (1976) 'The Cosmological Argument' in P. Angeles (ed.) *Critiques of God*, Buffalo, NY: Prometheus, 43-58.(Interesting critique of attempted proofs of God's existence.)

Findlay, J.N. (1955) 'Can God's Existence Be Disproved?' in A. Flew and A. MacIntyre (eds) *New Essays in Philosophical Theology*, New York: Macmillan, 47-56.(Classic argument that God must be, yet cannot be, 'logically' necessary.)

Ibn Sina (Avicenna) (early 11th century) *al-Risalat al-'arshiya*, in A. Arberry (trans.) *Avicenna on Theology*, Westport, CT: Hyperion, 1951.(Treats many issues in philosophical theology.)

Knuuttila, S. (1993) *Modalities in Medieval Philosophy*, New York: Routledge. (Goldmine of information on evolving concepts of necessity. At times technical.)

Leibniz, G.W. (1697) 'On the Ultimate Origination of the Universe', in P. Schrecker and A. Schrecker (trans.) *Leibniz: Monadology and Other Philosophical Essays*, Indianapolis, IN: Bobbs-Merrill, 1965.(His cosmological argument for God's existence.)

Leibniz, G.W. (1710) *Theodicy*, trans. D. Allen, Indianapolis, IN: Bobbs-Merrill, 1966. (Important discussion of God's nature and goodness.)

Leibniz, G.W. and Clarke, S. (1717) *The Leibniz-Clarke Correspondence*, ed. H.G. Alexander, Manchester: University of Manchester Press, 1956.(Rich exchange on space, time and God, dense but not technical.)

Morris, T. and Menzel, C. (1986) 'Absolute Creation', *American Philosophical Quarterly* 23 (4): 353-62. (Proposes that God is the source of all necessity.)

Penelhum, T. (1971) 'Divine Necessity' in B. Mitchell (ed.) *The Philosophy of Religion*, New York: Oxford University Press, 179-90.(Highly critical.)

Plantinga, A. (1974) *The Nature of Necessity*, New York: Oxford University Press. (Classic account of broadly logical necessity, with application to God and the ontological argument. Exact and exacting.)

Plantinga, A. (1980) *Does God Have a Nature?*, Milwaukee, WI: Marquette University Press.(Full, careful treatment of God's relation to necessity. Includes interesting exegesis of Descartes, with references to important literature.)

Plotinus (*c*.250-66) *Enneads II, III, VI*, trans. A.H. Armstrong, Cambridge, MA: Loeb Classical Library, Harvard University Press, 1979, 1980, 1988.(Neoplatonist metaphysics, discusses modal matters intermittently. Not technical but sometimes obscure.)

Spinoza, B. (1676) *Ethics*, in E. Curley (trans.) *The Collected Works of Spinoza*, Princeton, NJ: Princeton University Press, 1985.(Book I is the main source for his view of God. Sometimes obscure.)

Zalta, E. (1988) 'Logical And Analytic Truths That Are Not Necessary', *Journal of Philosophy* 85 (2): 57-74. (Somewhat technical argument for contingent analytic truths.)

Necessary truth and convention

Necessary truths have always seemed problematic, particularly to empiricists and other naturalistically-minded philosophers. Our knowledge here is a priori - grounded in appeals to what we can imagine or conceive (or can prove on that basis) - which seems hard to reconcile with such truths being factual, short of appealing to some peculiar faculty of a priori intuition. And what mysterious extra feature do necessary truths possess which makes their falsity impossible? Conventionalism about necessity claims that necessary truths obtain by virtue of rules of language, such as that 'vixen' means the same as 'female fox'. Because such rules govern our descriptions of all cases - including counterfactual or imagined ones - they generate necessary truths ('All vixens are foxes'), and our a priori knowledge is just knowledge of word meaning. Opponents of conventionalism argue that conventions cannot ground necessary truths, particularly in logic, and have also challenged the notion of analyticity (truth by virtue of meaning). More recent claims that some necessary truths are a posteriori have also fuelled opposition to conventionalism.

1 The linguistic theory of necessity

Many of the clearest necessary truths - such as 'Squares are rectangles' - seem simply to reflect the meaning of the words involved: we need not postulate some metaphysical essence of squareness which the necessity of 'Squares are rectangles' reflects. Thus, the position common to conventionalists is that necessary truths are explained by conventional rules of language, rather than a mind-independent modal structure of the world. For this reason, the general conventionalist position has often been called the linguistic theory of necessity (or, of the a priori). This was the dominant view about necessity from the 1930s-50s, particularly under the influence of logical positivism (see Logical positivism). Drawing largely upon claims they found in Wittgenstein's *Tractatus* (see below), the positivists rejected Kant's claim that there are synthetic a priori truths, trumpeting the slogan 'The a priori is analytic' (see Analyticity). What exactly analyticity was - beyond, in some way, truth by virtue of the meanings of words - was not entirely clear, which gave rise to different versions of the linguistic theory. But for most, conventionalism itself was beyond doubt and, apart from the few explicitly engaged in trying to spell out just how conventions explained necessity, most writers used different ways of articulating the view ('All necessary truths are tautologies', 'Necessity simply reflects linguistic rules') interchangeably simply to invoke the linguistic theory.

An early expression of conventionalism is found in Locke's proposal that all essences are merely nominal - they simply reflect the meaning of the subject term, as in 'Gold is a metal'. In this, Locke meant to contradict the then-received view that essences are determined by a mind-independent sorting of things. Similarly, he says essences 'are made by the mind' (1689: III.vi.26), and since 'kinds' - or in Locke's terms, 'sorts' - are the kinds they are because of their essences, Locke says the sorts themselves are 'the Workmanship of the Understanding' (1689: III.vi.12) (see Locke, J. §4). This well exemplifies the metaphysically deflationary thrust of conventionalism.

While there are various conventionalist views, they agree negatively both in their suspicion of appeals to imagination and conceivability as sources of factual knowledge, and in their finding the notion of necessary facts - and a distinction between essential and accidental properties - rather mysterious (see Essentialism). On the positive side, they agree in seeing appeals to imagination and conceivability as appeals to what we would *say* in various imagined situations, which they see as a way to find out what rules of language we are implicitly employing.

The version of conventionalism presented in Wittgenstein's *Tractatus* focuses on the notion of 'tautology', that is, truth-functional compound statements, such as 'If it is raining, then it is raining', which are true for all assignments of truth-values to the propositions from which they are built (1922: 4.46). Wittgenstein argues that tautologies say nothing about the world; rather, the symbols, given their meanings, themselves determine the truth of such statements (1922: 4.461-4661, 5.142). He claims that all truths of logic are tautologies (1922: 6.1-112) and seems to hold this for all a priori truths (for example, 1922: 5.134, 5.634). There are necessary truths which are not truths of logic, such as 'All bachelors are unmarried'. However, since 'bachelor' means 'unmarried man', the statement, when analysed, says, 'All unmarried men are unmarried', the form of which is 'Whatever is *F* and *G* is *F*', which is a tautology. One problem this version of conventionalism faced was accounting for necessary truths such as 'Nothing can be simultaneously both red and green all over' - for it is hard to see how to assign meanings to 'red'

and 'green' whereby this can be reduced to a tautology.

Other conventionalists have concentrated more on the general idea that truths such as 'All squares are rectangles' express rules of language. Their 'truth by convention' could perhaps serve as a model for all necessary truths. One version of this idea was that 'Necessarily, all squares are rectangles' simply means 'It is a rule of language that "square" applies to something only if "rectangle" also applies'; A.J. Ayer wrote that necessary propositions 'simply record our determination to use words in a certain fashion' (1936: 84). This view was regularly criticized for identifying necessary a priori truths with contingent empirical statements about a language or its speakers. A subtler variant was that one should not really speak of necessary truths at all - what look to be assertions are really, covertly, prescriptions ('Do not apply "square" to anything you would not apply "rectangle" to') or perhaps decisions of some sort. One problem here is that necessary statements can stand in logical relations, can figure essentially in truth-functional compounds, and so on, which seems to require that they have truth-values, unlike prescriptions. The most common version of this view is simply that the rule determines the truth of the sentence; given the rule, nothing can count as a square but not a rectangle. This view does not require that all rules have the form of definitions which allow us, via substitution, to reduce all necessary truths to syntactical tautologies; a rule prohibiting the simultaneous use of 'red all over' and 'green all over' to an object need not be rooted in a full, necessary-and-sufficient-conditions style definition of some constituent term (see Carnap 1952).

While many attacks on conventionalism only properly address particular versions, certain problems challenge any sweeping conventionalism. By the late 1950s, not only was conventionalism no longer dogma but, to many, seemed simply mistaken. One central problem concerns the fact that whether some convention obtains is always a contingent matter - but how can something which is contingent make anything necessary? For instance, even conventionalists (generally) acknowledge that different conventions would not affect the necessity of: if if *p* then *q*, and *p*, then *q*. But if different conventions would not change what is necessary, how can the conventions be responsible? Different conventions can change which proposition is expressed - but not the necessity of the proposition actually expressed. Another general problem concerns the idea that logical truths obtain by convention. The conventionalist wants to say that, given certain rules, it follows that certain statements are true (and necessary). But 'follows' is an expression of necessitation. Quine (1966) and others argue that in order to derive the truths of logic from conventions, one must apply logical rules - one must presuppose logic in order to derive logic. Additionally, attacks on the analytic/synthetic distinction, particularly vivid in arguments put forward by Putnam (1975) that apparently analytic truths are actually revisable empirical claims ('Cats are animals'), cast further doubt on the analytic explanation of necessity. If there are no analytic truths, they cannot explain necessity (see Analyticity).

2 Necessity a posteriori

In *Naming and Necessity* (1980), Saul Kripke argued that the supposition that all necessary truths are a priori is mistaken. His examples included empirical identity statements, such as 'Hesperus is Phosphorus', statements of material origin, such as 'This lectern was originally made of (a particular piece of) wood' and statements of kind essence and identity, such as 'Gold has atomic number 79'. Since these are not a priori, they cannot simply be true by convention, and if conventionalism was already on the wane, Kripke's examples, and similar ones from Putnam (1975), seemed to kill it decisively. Because these truths are not logically necessary, but are still necessary absolutely, they are sometimes called 'metaphysically necessary'.

However, it has been noted that even if these necessary truths are not true by convention, they may still owe their necessity to convention (see Coppock 1984, Sidelle 1989). The arguments for proposed necessary a posteriori truths each involve commitment to some general principle, for instance, that if a material object originates in bit of matter *M*, then it essentially originates in *M*. These principles - in so far as the arguments for the necessities are plausible - are plausibly a priori, for the arguments rest on the familiar appeals to what we can imagine, or would be willing to say. The conventionalist proposal, then, is that these general principles are analytic, and that while these conventions do not determine which non-modal statements are true (what bit of matter something did originate in), they are responsible for these truths being necessary (the principle determines that nothing with a different origin can count as this object). So conventionalism does not require that all necessary truths are *true* by convention, but only that this is why they are necessary.

See also: Conventionalism; EssentialismLanguage, conventionality of; Logical and mathematical terms, glossary of

ALAN SIDELLE

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz; repr., New York: Dover, 2nd edn, 1946, 16-18 and ch. 4.(Influential and often cited presentation of conventionalism about necessity (and a priori truth); seen by many as claiming that necessary truths say that there is a convention (but see Broad 1936).)

Broad, C.D. (1936) ‘Are There Synthetic A Priori Truths?’, *Proceedings of the Aristotelian Society*, supplement: 102-16.(Expresses scepticism about Ayer’s conventionalism.)

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press, esp. chaps 1 and 5.(A rigorous attempt to formulate a system in which it is clear how necessary truth arises from the language.)

Carnap, R. (1952) ‘Meaning Postulates’, *Philosophical Studies* 3: 65-73. (Introduces a device for trying to account for non-tautological necessary truths. Also appears in Carnap 1947.)

Coppock, P. (1984) ‘Review of N. Salmon, *Reference and Essence*’, *Journal of Philosophy* 81: 261-70.(Sketches a conventionalist account of necessary a posteriori truths.)

Ewing, A.C. (1940) ‘The Linguistic Theory of A Priori Propositions’, *Proceedings of the Aristotelian Society*: 207-44.(Classic criticism of conventionalism.)

Dummett, M. (1959) ‘Wittgenstein’s Philosophy of Mathematics’, *Philosophical Review* 68: 324-48.(Critical discussion of Wittgenstein’s conventionalism (1956).)

Forbes, G. (1985) *The Metaphysics of Modality*, Oxford: Clarendon Press, ch. 9.(Defends modal concepts against scepticism and psychological accounts, and offers a broadly conventionalist view.)

Fraassen, B.C. van (1977) ‘The Only Necessity is Verbal Necessity’, *Journal of Philosophy* 74: 71-85.(Argues against any irreducibly natural necessity.)

Kneale, W. (1946) ‘Truths of Logic’, *Proceedings of the Aristotelian Society*: 207-34.(Argues against a conventionalist view of logic.)

Kneale, W. (1947) ‘Are Necessary Truths True By Convention?’, *Proceedings of the Aristotelian Society*, supplement: 118-33.(Forceful negative argument, emphasizing the conflict between ‘necessary’ and ‘conventional’, and critically discussing the conventionalist’s appeal to alternative logics and algebras.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.(Argues for the conceptual separation of necessity and a prioricity, and for necessary a posteriori truths.)

Lewis, D.K. (1969) *Convention*, Cambridge, MA: Harvard University Press.(Argues that sense can be made of the idea of non-explicit conventions, contra Quine 1953.)

Locke, J. (1689) *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1979. (Referred to in §1. Locke’s discussion of substance in III.vii classically expounds his conventionalism about essence and kinds.)

Malcolm, N. (1940) ‘Are Necessary Propositions Really Verbal?’, *Mind* 49: 189-203.(Representative expression of conventionalism, particularly concerned to avoid the claim that necessary statements express the proposition that some convention holds, or state any fact.)

Pap, A. (1958) *Semantics and Necessary Truth*, New Haven, CT: Yale University Press, esp. ch. 7.(An extended argument against conventionalist views of necessity.)

Putnam, H. (1975) *Philosophical Papers II: Mind, Language and Reality*, Cambridge: Cambridge University Press, 1979.(‘Is Semantics Possible?’ (referred to in §1) argues against the traditional view of meaning upon which analyticity rests; development of Quine 1953. ‘The Meaning of "Meaning"’ (referred to in §2) argues for necessary a posteriori status for certain claims about natural kinds, such as ‘Tigers are mammals’ and ‘Water is H_2O’.)

Quine, W.V. (1953) ‘Two Dogmas of Empiricism’, in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1980.(Most often cited argument against the analytic-synthetic distinction, and so against truth by convention.)

Quine, W.V. (1966) ‘Truth By Convention’, in *The Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press.(Referred to in §1. A classic argument against accounting for truth by convention alone.)

Schlick, F.A.M. (1931) ‘Gibt es ein materiales Apriori?’, trans. ‘Is There a Factual A Priori?’, in H. Feigl and W.S. Sellars (eds) *Readings in Philosophical Analysis*, New York: Ridgeview, 1982.(Highly representative example of assertion that necessity and a prioricity can only derive from linguistic convention.)

Sidelle, A. (1989) *Necessity, Essence and Individuation: A Defense of Conventionalism*, Ithaca, NY: Cornell University Press.(Referred to in §2. Defends a conventionalist account of necessary a posteriori truths.)

Wittgenstein, L.J.J. (1922) *Tractatus Logico-Philosophicus*, trans. C.K. Ogden and F.P. Ramsey, London: Routledge; trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1961.(Highly influential source of the view that all necessary truths are tautologies and that tautologies say nothing about the world.)

Wittgenstein, L.J.J. (1956) *Remarks on the Foundations of Mathematics*, Oxford: Blackwell.(Critical remarks about the necessity involved in mathematical truths and proofs. Helpful interpretation in Wright (1980) and Dummett (1959).)

Wright, C. (1980) *Wittgenstein on the Foundations of Mathematics*, London: Duckworth, and Cambridge, MA: Harvard University Press, esp. Part 3.(Sympathetic interpretation of Wittgenstein's conventionalism about necessity.)

Yablo, S. (1993) 'Is Conceivability a Guide to Possibility?', *Philosophy and Phenomenological Research* 53: 1-42.(Tries to defend the objectivity of our a priori methods of investigating modality, thus responding to the epistemological arguments for conventionalism.)

Neckham, Alexander (1157-1217)

Alexander Neckham is one of the leading thinkers in the English appropriation of the new science made available during the twelfth century. His best known writings, especially De naturis rerum (On the Natures of Things), show a prodigious acquaintance with natural history. Neckham was most concerned, however, that the study of the natural world be made to serve the purposes of theology. He thus strove not only to draw moral lessons from nature, but to apply to theological method the doctrines of the new logic, especially Aristotle's Topics. While Neckham cannot be said to have mastered the texts that were flooding into the Latin West, he certainly did realize the challenges and the possibilities that they offered to inherited theologies.

Alexander Neckham (also Neckam or Nequam) displays in his writings that zeal for new learning that characterizes the 'renaissance' of the twelfth century in Latin-speaking schools of western Europe. Neckham lived out his life in these schools. He studied and then taught in Paris from about 1175 to 1182, where his subjects were the liberal arts and theology, with some law and more medicine. Neckham remarks that he was a student of the school on the Petit-Pont, and his writings reflect the doctrines of the school's most famous master, Adam of Petit-Pont. From Paris, Neckham returned to England to teach in the grammar schools of Dunstable and St Albans. Not long after 1190 he was in the schools at Oxford, which were just then beginning to reorganize themselves in ways that would make them into a university. There are many references in Neckham's writing to Scholastic techniques, such as the exercise of the *disputatio* or disputed question, and there survives a set of theological questions that he determined. The appropriation of newer techniques was also perfectly compatible with the cultivation of older ones. Neckham established himself at Oxford as an important preacher and as the author of a commentary on the Athanasian creed. About 1200, he withdrew to the monastery of Augustinian canons in Cirencester, where he was elected abbot in 1213. Neckham's most extended and important works were written at the monastery. He died there, as abbot, in 1217.

Neckham is best known to historians of philosophy as a student of natural philosophy. Part of this reputation is due to his being the recipient of lavish praise in a contemporary treatise, Alfred of Shareshill's *De motu cordis (On the Motion of the Heart)*. Alfred wrote the treatise at Neckham's prompting while the latter was still at Oxford. However, Neckham's reputation as a student of natural philosophy is due more substantially to several of his own works, especially *De naturis rerum (On the Natures of Things)* and *De laudibus divinae sapientiae (Praise of Divine Wisdom)*.

De naturis rerum takes the form of an immense prologue to a commentary on Ecclesiastes. The relation of 'prologue' and body is not entirely clear, especially since *De naturis rerum* itself begins as a commentary on the account of creation in Genesis. Whatever Alexander's exact motive, the work quickly enough becomes something like an encyclopedia of moralized natural philosophy. It is loosely organized around the four elements of ancient physics - fire, air, water, earth - but the range of erudition is enormous. In 272 chapters over two books, Neckham rehearses most of the natural history known to him, from problems in astronomy, through the hydrodynamics of fountains and the properties of precious stones, to the behaviour of animals, wild and domestic. Although much of the material is taken from ancient sources, such as Pliny, Neckham frequently adds information or anecdote of his own. More importantly, he regularly reads the facts he recites so as to yield a moral lesson. Indeed, the whole of *De naturis rerum* ends with a sustained attack on human vices, thus making a bridge to the reading of Ecclesiastes.

Neckham returned to these topics in a metrical compendium, his *De laudibus divinae sapientiae*. It too is ordered around the four elements, but ranges widely. In it the reader finds Neckham reciting catalogues of the stars and of Europe's main rivers, as well as descriptions of jewels, herbs and trees and the liberal arts. Still not satisfied with his comprehensiveness, Neckham returned to natural history again in a *Suppletio defectuum (Remedy of Defects)* appended to *De laudibus*, where he offers new descriptions of animals and their habits.

The ordering of natural knowledge to theology, always a concern, seems to have occupied Neckham increasingly. He tries to show the right ordering of them in his last great work, the *Speculum speculationum (Mirror of Speculations)*. The work is announced and begun as a refutation of the heresies of the Cathars. It is in fact, and from the beginning, organized roughly along the lines of Peter Lombard's *Sentences* (see Lombard, P.). It expands on the topics of the *Sentences*, increasing the number of arguments and the length of their analysis. Neckham does

much, for example, to bring in the still relatively little known works of Anselm of Canterbury. Of course, some of his erudition is drawn from theological *florilegia*, anthologies of excerpts drawn from the Christian Bible, the Church Fathers or medieval theologians of the Latin tradition. Neckham also knows and uses the Platonic and Neoplatonic texts popular among philosophical writers earlier in the century: Plato's *Timaeus* with Calcidius's commentary, Boethius with his medieval commentators, and at least some of Thierry of Chartres, William of Conches and Bernard Silvestris (see Bernard of Tours; Platonism, medieval).

What is more striking, because it is much newer, is Neckham's appropriation for theology of the 'new logic' and newly translated works by Aristotle (see Aristotelianism, medieval). Certainly Neckham knew the logical works well. In constructing his dialectical theology, he makes much use of Aristotle's *Topics*, so becoming one of the first theologians to capitalize on this recently rediscovered text. His relation to other Aristotelian works is more difficult to discern. Neckham refers by name to *On the Soul*, *Metaphysics* and *Nicomachean Ethics*, and he quotes without reference from *On the Heaven and the Earth*. Such early access to these works suggests that Neckham may have brought them back with him from the continent. Equally striking is an implicit citation of Avicenna's *On the Soul*, one of the new wave of Arabic works just finding their way into Latin and into the northern schools. To say that Neckham mentions or sometimes uses these texts is not to say that he has mastered them. On the contrary, he stands at the beginning of what will be a very long effort to read and then to understand the texts that were pouring into Latin so rapidly from both Greek and Arabic (see Translators).

See also: Chartres, School of; Natural philosophy, medieval

MARK D. JORDAN

List of works

Neckham, Alexander (before 1182?) *De nominibus utensilium (On the Names of Utensils)*, ed. A. Scheler, 'Trois traités de lexicographie latine du XIIe et due XIIIe siècle', *Jahrbuch für romanische und englische Literatur 7*, 1866: 58-74, 155-73. (An elementary survey of technical terms from arts and crafts as encountered in ordinary use; modelled after a work by Adam Parvipontanus.)

Neckham, Alexander (before 1182?) *Corrogationes Promethei (Collections of Prometheus, that is, of one who teaches rudimentary arts)*, excerpts ed. P. Meyer, 'Notice sur les Corrogationes Promethei d'Alexandre Neckham', *Notices et extraits des Manuscrits de la Bibliothèque Nationale* 35 (2), 1896: 641-2.(An introduction to grammatical theory, based on teaching in the Paris schools; incomplete.)

Neckham, Alexander (1194-1204) *De naturis rerum (On the Natures of Things)*, ed. T. Wright, Rolls Series, London: Longman, Green, Longman, Roberts and Green, 1863; Books 3-4 never printed.(A comprehensive survey of knowledge about the natural world; the work is the first two parts of a five-part commentary on Ecclesiastes, the remainder of which has never been printed.)

Neckham, Alexander (*c*.1213) *De laudibus divinae sapientiae (Praise of Divine Wisdom)*, ed. T. Wright, Rolls Series, London: Longman, Green, Longman, Roberts and Green, 1863. (A metrical rehearsal of the order of nature.)

Neckham, Alexander (1216-7) *Suppletio defectuum (Remedy of Defects)*, excerpts ed. H. Walther, 'Zu den kleineren Gedichten des Alexander Neckham', *Mittellateinisches Jahrbuch* 2, 1965: 111-21.(A correction and supplement to the *De naturis rerum*; incomplete.)

Neckham, Alexander (1201?-17) *Speculum speculationum (Mirror of Speculations)*, ed. R.M. Thomson, Oxford: Oxford University Press, 1988.(Treatment of all the major topics in Christian theology, with dialectical analysis and ample use of the pertinent authoritative texts; incomplete.)

References and further reading

Alfred of Shareshill [Sareshel] (*c*.1195) *De motu cordis (On the Motion of the Heart)*, ed. C. Baeumker, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters 23.1-2, Munich: Aschendorff, 1923. (This edition is neither complete nor critical.)

Hunt, R.W. (1984) *The Schools and the Cloister: The Life and Writings of Alexander Nequam*, ed. and revised M.T. Gibson, Oxford: Clarendon Press.(Revised from a doctoral dissertation completed in 1936.)

Needs and interests

To have an interest in something is to have a stake in how that thing goes. Needs can be thought of as interests instrumental to a specified purpose, as an artist will need paint, or as general essential interests, like the needs associated with physical survival. Both needs and interests can be contrasted with wants, in being more objectively assessable. A concrete specification of needs and interests is better done by means of a list than by a criterion stipulated in terms of goods that would be chosen in a hypothetical situation. The concept of need can be used to justify a claim to economic resources, even though it only states a minimal demand. However, when total resources will not meet all needs, it is necessary to work with the principle of the equal consideration of interests rather than that of needs, as in the case where health care needs may outstrip available resources. The concept of interests can be misused in defence of authoritarian political systems, but does not have to be used in this way. Although citizens may have an interest in performing their duty, it is likely that there will always be a continuing conflict between duty and interests in politics.

1 The concepts of needs and interests

As concepts, needs and interests are closely related, although not identical. If I say that I need to go to the dentist, I will normally be taken to imply that dental treatment would be in my interest. However, not everything in a person's interest is a need. For example, securing tax allowances may be in the interests of the rich, but it is not something that they need.

It would be difficult to imagine political, legal, moral and practical discussion without the concept of interests. Political parties advance certain interests, legal institutions arbitrate between interests, moral systems restrain the scope of self-interest and practical reason considers how actions will promote interests. Indeed, on one theory of action, agents are moved only by their own interests.

It is possible to distinguish between something being in someone's interest and someone taking an interest in something. I may take an interest in something purely for the sake of curiosity, for example the performance of my local sports team, without my interests being affected by that performance, as they would be if I had a financial stake in their results. However, there are many aspects of life where the two notions coalesce. For example, parents will have an interest in their child's career success if that affects their own household income, but typically they will also take an interest in it, in the sense of being personally concerned.

What then are interests? In a well-known definition John Locke (1689: 15) defined interests as 'life, liberty, health, and indolency of body; and the possession of outward things, such as money, lands, houses, furniture and the like'. This definition takes the form of a list of things that are thought to be the generalized means, in the form of personal conditions and external possessions, to some supposed human good. Feinberg (1984) has expressed the matter by saying that to have an interest in something is to have a stake in it, and to have a stake in *X* is to stand to lose or gain depending on what happens to *X*.

Needs may be regarded as a subset of interests, namely those interests that are in some sense essential or necessary to one's affairs. As necessary conditions, needs can be stipulated in a weak or strong sense. In the weak sense, they are simply the conditions required to achieve any goal or general aim. In this sense, an artist will need paints and brushes in order to work. In such cases the claims of need depend upon an evaluation of the end state to which the need is instrumental.

In the strong sense, needs are the conditions required to achieve some basic goal like survival. In this case needs designate those items necessary to having any purposes at all, or to avoid serious harm. These may thus be called vital interests, their claims being correspondingly higher than instrumental needs.

Both needs and interests may be contrasted with wants. I may not want to go to the dentist, but I may need to, because my health will suffer if I do not. One way of putting this point is to note a logical difference in the inferences that can be drawn from sentences concerned with wants and those concerned with needs. Wants are always characterized from the point of view of those with the wants, and in consequences their full implications may remain obscure to their possessor. For example, I may want to take the fastest train to my destination, but not know that I want to take the 1300, because I do not know that it is the fastest. Thus, one cannot validly infer from

my wanting *X* that I want *Y*, which is identical with *X*, unless I am also known to identify *X* and *Y*. By contrast, if I need *X* it can be validly inferred that I need *Y* identical to *X*, independently of my own state of mind.

A similar point holds in respect of interests. There is therefore an objectivity about needs and interests that there is not about wants. The contrast is not absolute, and the concept of interest can sometimes come close to a more subjectivist notion of welfare or want-satisfaction, since how people conceptualize their interest or balance conflicting interests can vary (see Welfare §1). So, we cannot say that the religious who pursue their spiritual development at the cost of poverty are acting against their own interests. But that is partly because people can have an interest in the conditions of their own personal development, and this can clash with material interests.

Since interests need to be balanced in this way, can we state a general criterion for identifying how this would be done? One suggestion is to identify interests thus balanced with the informed hypothetical choices that agents might make. On this account, the balance of interests represented by policy *X* is more in my overall interests than that produced by policy *Y*, were I to prefer the results of *X* to those of *Y* when I experience them both. However, unless the hypothetical choice occurs, it would seem that the only way we could know what choices someone would make is to have a sense of what their interests already involve, so that the prediction of choice is grounded in our understanding of needs and interests and not vice versa. In this sense, interests are logically prior to choice.

2 Needs, interests and economic distribution

The concept of need is frequently linked to questions of economic justice. For example, modern health and welfare services are often said to be based on the principle of meeting need irrespective of the ability to pay. But can the principle of need be used in this way to underscore a claim for resources?

Of the two meanings of need that we have so far distinguished, it is clear that the sense relevant is the strong one stated in terms of means to basic goals like survival. Distribution according to need thus means no more than that people should be assured of the means of subsistence and survival.

It may be urged that as a principle of distribution the appeal to needs is too undemanding, since the means of survival are widely available in modern societies, and therefore economic justice should be concerned with broader claims of justice, for example that there should be an equal consideration of interests above the minimum needed for mere survival (see Justice).

However, the proposal to dispense with a principle of need in economic matters has to be treated with some caution. The notion of survival is not itself clear-cut, and it can include the means to secure self-respect in society. Adam Smith (1776) pointed out that, although Romans did without a linen shirt, even the poorest labourer in Smith's day would be ashamed to appear in public without one. Thus, what is necessary to survival has a cultural, as well as a material, component which varies from society to society, so that needs will expand over time as an economy grows.

Moreover, in the context of claims for the international redistribution of resources, the appeal of the principle of need is strong, since on a global scale not all are assured of the means of survival even in the physical sense. In this context, a basic-needs strategy of economic development would entail some significant redistribution, and the principle would be neither empty nor without force (see Justice, international).

Yet the principle of the equal consideration of interests may come to supplant the needs principle in some circumstances. For example, distribution of care according to need has been a traditional principle in medicine. Yet modern technology makes it possible to meet an increasing range of medical needs, but only at impossibly high cost. Where resources are inadequate to meet the sum of needs, certain needs will have to remain unmet, and an alternative principle found for rationing resources, for example that all are entitled to an equal consideration of their health interests.

3 Interests in politics

The concept of interests in politics has historically been associated with the democratic idea that government should act in the interests of the governed, rather than acting, say, for an abstract ideal like national glory or spiritual purity. However, the appeal to interests can also be ambiguous in democratic terms, since a government

may well appeal to the idea of interests as a way of ignoring popular sentiment and opinion. Exploiting the logical space opened up by the distinction between wants on the one side and needs or interests on the other, a government might claim to know what is in the interest of people better than the people itself. A strong notion of interests might therefore seem to lend itself to a paternalist or authoritarian notion of government.

The danger of paternalist or authoritarian government arising from this substitution of elite judgement for the opinion of the people has been widely feared. However, it is difficult to dispense with the idea that there may be some occasions when governments should act according to the interests of citizens rather than being responsive merely to expressed wants. For example, when people underinsure themselves against age dependency or ill health, the government might properly compel individuals to act in their own long-term interests by instituting compulsory pension and health insurance schemes. Indeed, even the most liberal government will not allow individuals to sell themselves into slavery (see Paternalism).

We can distinguish between people wanting a policy and people wanting the results of a policy. One way in which it is possible to guard against the paternalist or authoritarian tendencies implicit in the idea that governments should act in the interests of the people is to say that these interests should be related to the policy results that people were seeking. Thus, if citizens value economic security, requiring compulsory savings to achieve this may well be justified as it would not be justified to impose puritan policies restricting individual liberty of action when spiritual purity was desired only by a few.

4 Interests and moral powers

So far we have considered interests as personal conditions and external possessions determining how individuals fare. But Rawls (1993) has proposed the concept of ‘higher-order interests’ (see Rawls, J. §1). Citizens with a conception of their own good and a sense of justice also have these higher-order interests in the development and exercise of these two powers. On this conception, then, instead of thinking of interests as the basis of claims on resources, we are to think of them as motivations to cooperative social action, making it in the interests of citizens to observe principles of justice.

To extend the notion of interest in this way has certain advantages, not least in understanding how citizens might have an interest in acting justly. Conflicts between the duties of justice and acting on one’s own self-interest would, in effect, become conflicts within one’s own sense of self-interest. To adopt this construction would have the disadvantage, however, of running contrary to the common observation that acting justly may often involve acting against one’s self-interest. Moreover, as a statement of philosophical psychology, it is hard to see how it is empirically supported.

Yet, even if interests are often opposed to duty, individuals can think of their interests as being wider than simply their own personal concerns and can identify with the interests of their community. In this sense, the idea of moral powers may point to a socializing of interests that is essential to any workable conception of justice.

ALBERT WEALE

References and further reading

Barry, B. (1965) *Political Argument*, London: Routledge & Kegan Paul; repr. Berkeley, CA: University of California Press.(Has an excellent chapter on the concept of interest.)

Braybrooke, D. (1987) *Meeting Needs*, Princeton, NJ: Princeton University Press.(A sophisticated discussion of the concept.)

Doyal, L. and Gough, I. (1991) *A Theory of Human Need*, Basingstoke: Macmillan.(Usefully brings together philosophical and social science analyses of needs.)

Feinberg, J. (1984) *Harm to Others*, vol. 1 of *The Moral Limits of the Criminal Law*, New York and Oxford: Oxford University Press.(Chapter 1 discusses interests in the context of discussing harm as a set-back to interests.)

Feinberg, J. (1986) *Harm to Self*, vol. 3 of *The Moral Limits of the Criminal Law*, New York and Oxford: Oxford University Press.(Discusses the conditions under which paternalist legislation in people’s own interests may be permissible.)

Locke, J. (1689) *A Letter Concerning Toleration*, ed. M. Montuori, The Hague: Martinus Nijhoff, 1963.(Contains a definition of civil interests.)

Miller, D. (1976) *Social Justice*, Oxford: Clarendon Press.(Discusses need in the context of theories of social justice.)

Plant, R. (1991) *Modern Political Thought*, Oxford: Blackwell.(Has a good discussion of various approaches to the concept of needs.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(Lecture II uses the concept of higher-order interests and Lecture V sets out the view of primary goods as citizen’s needs.)

Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, Oxford: Clarendon Press, 1979. (Book V, chapter 2, has the definition of necessaries.)

Thomson, G. (1987) *Needs*, London and New York: Routledge & Kegan Paul.(A good philosophical discussion of the concept of needs, defending a strong account of needs as the basis of justified claims.)

Negative facts in classical Indian philosophy

Like their European counterparts, the philosophers of classical India were interested in the problem of negative facts. A negative fact may be thought of, at the outset at least, as a state of affairs that corresponds to a negative statement, such as 'Mr Smith is not in this room.' The question that perplexed the philosophers of India was: How does someone, say Ms Jones, know that Mr Smith is not in the room? There are essentially four possible metaphysical positions to account for what it is that Ms Jones knows when, after entering a room, she comes to know that her friend is not present there. Each of the positions has been adopted and defended by certain classical Indian philosophers. On the one hand, some take the absence of the friend from the room as a brute, negative fact. Of these, some hold knowledge of this fact to be perceptual, while others hold it to be inferential. On the other hand, some hold that the absence of the friend from the room has no real ontic status at all, and believe that what there really is in the situation is just the sum of all the things present in the office. These latter philosophers hold that knowledge of one's friend's absence is just knowledge of what is present, though some believe the knowledge results from perception, while others believe it to result from inference. These four positions were maintained by, respectively, the Nyāya *philosopher Jayanta, the* Mīmāṃsā *philosophers Kumārila Bhaṭṭa and Prabhākara, and the Buddhist Dharmakīrti.*

1 Introduction to the problem

Negative facts have perplexed Western philosophers ever since the time of Plato. But the philosophers of Europe and America have not been the only philosophers to have been perplexed by them; classical Indian philosophers too have pondered their nature. The problem, as raised by classical Indian philosophers, is easy enough to state. Suppose that someone, say Ms Jones, enters a room and her friend, Mr Smith, is not there. How does Ms Jones know that Mr Smith is not there? One obvious answer, tendered by the classical Indian philosopher Kumārila Bhaṭṭa, is that Ms Jones fails to perceive Mr Smith (*Ślokavārttika* 9.11). While this answer has the ring of common sense, there are reasons to baulk at accepting it. To begin with, perception and inference are, in one sense, faculties or cognitive capabilities, whereas nonapprehension (*anupalabdhi*), the act of failing to perceive something, is not. There is, however, a more important reason to abandon the view that nonapprehension of something is a means of knowledge of its absence - something may be present yet not be perceived. Suppose that when Ms Jones enters a room, it is night, the lights are out, and the room is completely dark. Under these conditions, she would fail to see Mr Smith, even though he might very well be present.

Four main responses to the problem of negative facts can be discerned in classical Indian thought: that they are correctly analysed as straightforward positive facts known through inference; that there are real negative facts but that they can only be known through inference; that negative knowledge comes directly from perception of positive presences; and that we directly perceive real negative facts. Below, arguments for and against these positions are stated, although no attempt is made to track down their metaphysical origins and the related arguments. Rather, the exposition is confined to the authors whose treatments of the issue are the earliest ones which are both complete and sustained. They are those of Dharmakīrti (*c*.600-60), a Buddhist; of the followers of Kumārila Bhaṭṭa (*c*.620-80), an adherent of Mīmāṃsā; of Prabhākara (*c*.650-720), another adherent of Mīmāṃsā; and of Jayanta (*fl.* 890), an adherent of Nyāya.

2 The view of Dharmakīrti

The first classical Indian philosopher known to have understood and addressed the problem which confronts the view of Kumārila Bhaṭṭa was his older contemporary, Dharmakīrti. According to Dharmakīrti, both in his *Nyāyabindu (Drop of Logic*; 2.12-) and his *Pramāṇavārttika (Commentary on Valid Cognition*; 3.3-6 and commentary), one knows something when one correctly reports, having entered a room under the right conditions, that one's friend is absent from it; but what one knows is not a negative fact and is not known by perception. Rather, what one knows is a simple (positive) fact and it is known by inference.

A form of inference which comes to mind here is one advocated by Raphael Demos (1917) in an article in which he criticized Bertrand Russell's espousal of the existence of negative facts in his *The Philosophy of Logical Atomism*. Demos suggested that knowledge of absences is based on the relation of inconsistency between, on the one hand, the simple (positive) fact one observes upon, say, entering a room, and, on the other hand, the state of

affairs of one’s friend’s presence.

To be sure, Dharmakīrti holds that the relation of inconsistency plays a role in many inferences which warrant one’s claim to know some allegedly negative facts. For example, he holds that from the presence of heat one can infer an absence of cold, since heat and cold are inconsistent (with respect to the same locus). But it is not his view that inference based on the relation of inconsistency plays a role in the case where Ms Jones enters a room and comes to know that her friend is not there. In such a case, Dharmakīrti maintains, Ms Jones reasons as follows: the causal conditions of perception known are such that, if her friend were present in the room, she would see him; she does not see him; therefore, he is not present. The inference is counterfactual (see Gillon 1986 for further discussion).

Does Dharmakīrti’s counterfactual inference obviate the need to invoke the existence of absences? It seems not. Suppose, for example, one knows that one’s friend is absent from a room. This can be symbolized as $\neg p$. According to Dharmakīrti, one knows this because one has inferred it from two premises which one already knew: first, the counterfactual conditional that if one’s friend were present in the room, then one would know that the friend were present there; and second, the fact that one does not know that the friend is present in the room. These facts can be symbolized as $p \rightarrow Kp$ and $\neg Kp$ (where ‘*K*’ denotes the epistemic operator corresponding to the English ‘*a* knows that’ and ‘ $\rightarrow$’ denotes the connective for counterfactual conditionals). But this gives rise to a dilemma: how does one know that one does not know that one’s friend is present in the room? In other words, how does one know that $\neg Kp$? On the one hand, if Dharmakīrti holds that one perceives that one does not know that one’s friend is present in the room, that is to say, that one perceives that $\neg Kp$, then Dharmakīrti accepts thereby not only the existence of negative facts, in this case, negative mental facts, but also their perceptibility, both of which he wants to deny. On the other hand, if Dharmakīrti holds that one infers that one does not know that one’s friend is present in the room, that is to say, that one infers that $\neg Kp$, then there must be an inference to ground that claim, just as there is an inference to ground the initial claim that one’s friend is not present in the room, that is, $\neg p$. An infinite regress of inferences becomes inescapable, despite Dharmakīrti’s protestations to the contrary (see *Pramāṇavārttika* 3.3 and commentary).

3 The view of Kumārila Bhaṭṭa’s followers

A different position regarding negative facts and knowledge of them is the one adopted by the followers of Kumārila Bhaṭṭa. Though they maintained, unlike Dharmakīrti, that there are negative facts, they none the less seem to have adopted his view that knowledge in such cases is derived only from counterfactual inferences. These philosophers, unlike Dharmakīrti, but like most other classical Indian philosophers, considered counterfactual reasoning (*tarka*) neither as a form of inference (*anumāna*) nor, in general, as a standard means of knowledge (*pramāṇa*). The case of counterfactual reasoning yielding knowledge of negative facts was set aside by them as a means of knowledge unto itself, called ‘nonapprehension’ (*anupalabdhi*). Their view, of course, suffers from the very same dilemma just raised with respect to Dharmakīrti’s view, namely the dilemma of having either to concede the perception of some negative facts or to accept, for any negative fact said to be known, an infinite regress of inferences of negative facts.

4 The view of Prabhākara

A younger contemporary of Kumārila Bhaṭṭa, Prabhākara, who shared many of his philosophical assumptions, adopted the view that what one knows when, upon entering a room, one comes to know that one’s friend is absent from the room is some visually perceived, positive fact, that is, some presence. But what would be the plausible candidate for such a presence? Well, suppose Ms Jones’ friend, Mr Smith, usually sits in some chair in the room. According to Prabhākara, the pertinent presence would be the presence of the chair Mr Smith usually sits in. What else could there be to the absence of Mr Smith from his chair than the presence of the chair he usually sits in? After all, as Gerd Buchdahl (1961: 176-7) has observed in considering this problem in the Western philosophical tradition, one needs to draw only one picture to draw a picture either of the chair Mr Smith usually sits in or of the same chair without him in it. Yet Prabhākara’s metaphysical opponents were sceptical of the validity of this retort: they pointed out that if Mr Smith’s absence were identical with the presence of the chair he usually sits in, then any time Ms Jones perceives the chair, including the times when Mr Smith is in it, she should know that Mr Smith is absent. This, according to the opponents, is patently absurd.

Prabhākara apparently sought to extricate himself from this objection by denying that the presence of Mr Smith in the chair he usually sits in is the same as the presence of his chair without him in it: in the former case there are two presences, that of Mr Smith and that of the chair he usually sits in; but in the latter there is only one, namely that of the chair Mr Smith usually sits in. Opponents suspected that this distinction cannot be made without directly or indirectly relying on an assumption of the existence of absences; for, as they pointed out, to distinguish the presence of one's friend in his chair from the presence of only his chair requires the use of such expressions as 'only' (*eva*) or 'merely' (*mātra*), which are implicitly negative expressions, since to say that there is only one presence is to say that there is at least one presence and there is no more than one presence.

Prabhākara faced another objection. How could he account for the fact that, when one sees simply one's friend's chair, what one comes to know is the friend's absence and not one of the infinity of other absences which are also then present? One reply is to maintain that there is a difference between what one sees and what one remarks upon, and that this difference obtains equally in cases of perception of what are unquestionably presences as well as in cases of perception of what are allegedly absences. For the very same reason that when one looks at a desk and remarks that there is a pad of paper on it, but fails to remark that there is a pencil on it, one remarks on Mr Smith's absence from a chair he usually sits in, but one does not remark on the absence of, say, an elephant from it.

5 The view of Jayanta

An alternative to Prabhākara's position is one advocated by Jayanta. He agrees with Prabhākara that perception is the relevant means for Ms Jones to know that Mr Smith is absent from a room, but disagrees with him that all that is relevant to Ms Jones' knowledge of Mr Smith's absence is her perception of some presence. Rather, Jayanta holds that absences are brute facts unto themselves and that Mr Smith's absence in the case described is known through perception. Like others of the same school (Nyāya), he thought of perception as a causal process, the linchpin of which is contact between the object of perception and the appropriate sense organ. Thus, for example, in the tactile perception of the pencil in one's hand, the key causal link is the contact of the organ, the skin, with the pencil. This view of perception, however, raised an immediate problem: if contact between an object of perception and the appropriate sense organ is the key link in the causal process of perception, how can there be perception of absences, for surely it is absurd to think of any sense organ as coming into contact with an absence?

Jayanta's response to this objection was to revise the definition of sense perception: he restricted the above definition to the perception of positive facts and provided a special clause to account for perception of negative facts. According to Jayanta, this move was not *ad hoc*; and at least within his overall metaphysical framework there are independent grounds for its adoption, for he believed that, though universals do not come into contact with the sense organs, they none the less can be perceived. In general, something is perceived just in case it is a physical thing suited to a sense organ and in contact with it, or it is a universal inhering in a physical thing and both the universal and the physical thing are suited to the sense organ, or it is an absence related to a place and both the absence and the place are suited to the sense organ.

See also: Knowledge, Indian views of; MīmāṃsāČsect;Čsect;1-2; Nyāya-Vaiśeṣika §6

BRENDAN S. GILLON

References and further reading

Buchdahl, G. (1961) 'The Problem of Negation', *Philosophy and Phenomenological Research* 22: 167-78.(A clear and simple presentation of the problem of knowledge of negative facts within the Western philosophical tradition.)

Datta, D.M. (1932) *Six Ways of Knowing: A Critical Study of the Vedānta Theory of Knowledge*, Calcutta: Calcutta University Press, 2nd rev. edn, 1960.(Book 3, chapter 4 provides a comprehensive modern survey of absences and their knowability.)

Demos, R. (1917) 'A Discussion of a Certain Type of Negative Proposition', *Mind* 26: 188-196.(A historically important article on negative facts, not recommended to the philosophical neophyte.)

Dharmakīrti (*c*.600-60) *Pramāṇavārttika*, ed. R. Gnoli, *The Pramāṇavārttikam of Dharmakīrti: The First Chapter with the Autocommentary*, Serie Orientale Roma 23, Rome: Istituto Italiano per il Medio e Estremo Oriente, 1960; partly trans. R.P. Hayes and B.S. Gillon, 'Introduction to Dharmakīrti's Theory of Inference as Presented in *Pramāṇavārttika Svopajñavṛtti* 1-10', *Journal of Indian Philosophy* 19: 1-73.(A very difficult and important work on logic and metaphysics in the Indian philosophical tradition. The Gnoli edition provides

critical notes as well as the Sanskrit text.)

Dharmakīrti (*c*.600-60) *Nyāyabindu*, ed. D. Malvania, *Paṇḍita Durveka Miśra's Dharmottarapradīpa*, *Being a Sub-commentary on Dharmottara's Nyāyabinduṭīkā*, *a Commentary on Dharmakīrti's Nyāyabindu*, Tibetan Sanskrit Works Series 2, Patna: Kashiprasad Jayaswal Research Institute, 1955; 2nd edn, 1971; trans. T. Stcherbatsky, in *Buddhist Logic*, vol. 2, New York: Dover Publications, 1962.(A clear but all too concise compendium of its author's views on logic and metaphysics.)

Gillon, B.S. (1986) 'Dharmakīrti and his Theory of Inference', in B.K. Matilal and R.D. Evans (eds) *Buddhist Logic and Epistemology*, Dordrecht: Reidel, 77-87.(A self-contained explanation of how Dharmakīrti treats knowledge of absences as a form of inference.)

Gillon, B.S. (1997) 'Negative Facts and Knowledge of Negative Facts', in P. Bilmoria and J.N. Mohanty (eds) *Relativism, Suffering, and Beyond: Essays in Memory of Bimal K. Matilal*, Delhi: Oxford University Press, 1969, 128-49.(A contemporary philosophical treatment of the issue, which takes into account both Western and Indian views.)

Jayanta (*c*.890) *Nyāyamañjarī*, ed. K.S. Varadacharya, *Nyāyamañjarī of Jayanta Bhaṭṭa with Ṭippaṇi - Nyāyasaurabha by the Editor*, Mysore: The Oriental Research Institute, University of Mysore, 1969, 2 vols; trans. J.V. Bhattacharya, *Nyāya-Mañjarī: The Compendium of Indian Speculative Logic*, New Delhi: Motilal Banarsidass, 1978, 102-32.(This work provides a clear and comprehensive overview of the issues of classical Indian philosophy. The last fifth of the first āhnika (pages 102-33 of the English translation, 130-67 of the Sanskrit edition) provides a useful summary of views on the knowledge of absences.)

Kumārila Bhaṭṭa (*c*.620-80) *Ślokavārttika*, ed. G.Ś. Musalgaonkar, *Mīmāṃsā Darśana Ślokavārttika of Kumārilabhaṭṭa with Śābarabhāṣya, the Commentary Nyāyaratnakara and Notes*, New Delhi: Bharatiya Vidya Prakashan, 1979, 2 vols; trans. G. Jhā, *Śloka-vārttika: Translated from the Original Sanskrit with Extracts from the Commentaries of Sucarita Miśra (The Kāśikā) and Pārthasārathi Miśra (Nyāyaratnākara)*, Bibliotheca Indica 146, Calcutta: Asiatic Society of Bengal, 1907; repr. New Delhi: Sri Satguru Publications, 1983.(This work is the fundamental text of the school of Mīmāṃsā (founded by Kumārila Bhaṭṭa.)

Russell, B. (1918) *The Philosophy of Logical Atomism*, ed. D. Pears, La Salle, IL: Open Court, 1985.(This important philosophical work contains a passage which is a rejoinder to Demos' proposed treatment of negative facts.)

Stcherbatsky, T. (1930) *Buddhist Logic*, Bibiotheca Buddhica 26, Leningrad: The Academy of Sciences of the USSR, 2 vols; repr. New York: Dover Publications, 1962. (This important yet tendentious and often unclear introduction to Buddhist logic contains a usable translation of Dharmakīrti's *Nyāyabindu*.)

Negative theology

*The term 'negative theology' refers to theologies which regard negative statements as primary in expressing our knowledge of God, contrasted with 'positive theologies' giving primary emphasis to positive statements. The distinction was developed within Muslim, Jewish and Christian theism. If the negative way (*via negativa*) is taken to its limits, two questions arise: first, whether one may speak of God equally well in impersonal as in personal terms (blurring the distinction between theism and, say, the philosophical Hinduism of Śaōkara); and second, whether it leads ultimately to rejecting any ultimate being or subject at all (blurring the distinction between theism and, say, the atheism of Mahāyāna* Buddhism*). However, within their original theistic context, positive and negative statements about God are interdependent, the second indispensably qualifying the first, the negative statements taken alone useless.*

Negative qualifications on positive statements attributing so-called 'perfections' to God - for example, existence, life, goodness, knowledge, love or active power ('strength') - are obviously necessary if God is unimaginable. If his presence is always of his whole being and life all at once, in each place in space and time, he must be non-spatial and non-temporal in being and nature, and clearly he must be unimaginable. However, his supposed 'simplicity' and 'infinity' imply that he is much more radically outside the reach of understanding or 'comprehension', imposing the negative way at a deeper level than mere unimaginability. This unimaginability and incomprehensibility are key to theistic accounts of prayer and the mystical life.

1 The negative way in older Christian tradition

Within Christian theology, the negative way (called 'apophatic' from Greek) has two very different roles. First, if God is unimaginable and incomprehensible, this negative way is integral to any right understanding within theology as a 'theoretical' ('speculative') study considering models and using discourse and argument to assist understanding of God and his plan for creation (so far as this understanding is linguistically expressible). Second, the negative way is essential to the use of the intellect as directed practically towards union with God, that is, to the theology of spirituality and what Lossky (1957) calls 'mystical' theology. However, such theology seems to depend on positive statements about God, first for its context - some particular teaching and liturgy - enabling the mind to be directed in attention and love towards God and no other, and second for its verbal expression. This being directed towards God is typified when the mind is moved towards a loving, non-conceptual knowledge of God, setting aside images and concepts; this is illustrated in the mystical writings of Gregory of Nyssa and the poetry and commentaries of John of the Cross, their teaching being 'practical' in the sense explained by Jacques Maritain (1959). This love, whether realized in mystical prayer or as the unseen root of more active forms of life and prayer, is essentially directed to God, one who is holy or 'wholly other' (Otto 1917), not to any finite object of which we might form an image, or grasp or comprehend in nature.

The pre-Christian idea of a name as a means of control over what is named, bringing it within reach of human science and manipulation, is echoed in the negative emphasis both in the Old Testament and the Greek Fathers, who said that God 'has no name', or is 'ineffable'. This negative perspective was reinforced by Exodus 33: 18-23, which represents Moses as unable to see God's face, but only his back, and is echoed in the Gospels: 'no man has seen God at any time' (John 1: 18). Basil speaks of God as descending to us in his operations and John of Damascus suggests three possible etymologies deriving the word *theos* (God) from an operation. From these suggestions and the fifth-century writings of Pseudo-Dionysius, later Eastern tradition developed the idea that we never say anything positive of God's 'essence' (God as he is in himself), but only about his 'uncreated energies' - a development especially marked in Gregory Palamas and Eastern councils in 1351 and 1357. Latin tradition never accepted any distinction between God's essence and uncreated energies, but insisted that when we refer to God's existence, life, goodness, greatness, love, strength and suchlike, we refer to God's essence, that is, to God as he is in himself ('*substantialiter*' in Aquinas); however, it still maintained a parallel negative bent, which is highlighted in the Fourth Lateran Council's statement in 1215 that 'between Creator and creature no similarity can be expressed without including a greater dissimilarity' (Tanner 1990: 232).

2 Negative theology in Thomas Aquinas

The classic Western solution of the problem was presented by Thomas Aquinas (*Summa theologiae* Ia, qq.1-13).

He explains that we speak of things according to the way in which we know them (q.13, a.1), and speak of God as 'we know him from created things as their cause (*principium*), by way of excellence and by negation (*remotio*)'. In this traditional phraseology, he was not referring to three distinct kinds of statement about God, but to how we are to understand positive statements attributing 'perfections' to God, for example, existence, life, goodness, knowledge, love and active power. In each case, God as cause or principle of creatures possessed the perfection beforehand in a more eminent way, and not in any of the finite ways in which the perfection concerned is exemplified in creatures (see God, concepts of §§1-6).

Any such approach, allowing certain terms to be applied both to God and to creatures, but in different ways, absolutely involves some flexibility in our use of these terms - Aquinas saw this flexibility in the possibility of middle ways between terms being used 'equivocally' and being used 'univocally' (in identical senses in different cases). 'Analogy' was his blanket term covering all such middle ways.

For Aquinas, the words we use to signify things belonging to the essence of a particular sort of creature express things stateable in its definition, whereas what they signify in God's case cannot be contained in a definition. A definition makes the thing defined the object of human comprehension in some particular science. In Aquinas' conception, no kind of theology can yield comprehension (meaning master grasp, rather than loose understanding), and there is no science of God in the sense applicable to other sciences. Theology proper can only arise because by revelation God gives us some participation in his own knowledge of himself and of the way to him, participation in a style appropriate to human nature, and therefore using human language and concepts, including metaphors (*Summa theologiae* Ia, q.1). Aquinas has no conception of any self-contained natural theology, but only of a general metaphysics reaching some theological conclusions which, despite their limited character, give metaphysics some fuller integration (see Aquinas, T. §§9, 14).

In Aquinas' view, we can know the relevant statements about God to be true, some through philosophical reflection, and these and others through revelation, but we cannot comprehend God or the way he exists, lives, knows, loves and is good, or his mode of relation or action towards creatures, and so we have no comprehension or imagination of what makes these statements true. Terms for such 'perfections' apply truly to God in their proper or literal sense - it would, for example, be absurd to suppose that the cause of other things existing existed in a less real or proper sense than its effects. None of these most general positive terms has one definition even with regard to all creatures, each being used of different kinds of creature according to different definitions, or - in modern parlance - according to different criteria. Each such term is used of different types of thing 'analogically'. Although we learn to use them of creatures first, as applying in this or that way or respect, they apply unrestrictedly only to God. With regard to God, they have no 'scientific' definition. (The special term 'God' applies only to God and presents different but related problems: 'God' is the name of a nature, grammatically parallel to terms such as 'angel', 'animal', 'human being' and 'horse', and is not a formal or category term such as 'substance' or 'quality'; but still neither 'God' nor any other term applying to God alone has any real definition.)

For Aquinas as for Leibniz, without some positive predicates there would be no subject, and he rejects the more extreme negative way of Maimonides, for whom affirmative statements had only a negative content. Aquinas argues that this would not explain why we prefer some affirmative statements to others, would imply that God existed, was good, and so on, only in a secondary sense, and would be contrary to the intention of those speaking about God. Maimonides may have thought that our statements lock onto God as their object only because they are given direction by an adherence to God's law inspired from the heart. Aquinas would have replied that the love of heart or will, as distinct from sensory love, presupposes some knowledge by intellect, a reply also telling against more recent noncognitive approaches (see Mitchell 1971).

The completeness of Aquinas' embracing of the negative way appears in his exposition of God's simplicity (*Summa theologiae* Ia, q.3), where he successively announces different facets of the unlikeness between God and creatures immediately after proving God's existence (see Simplicity, divine §§1-2). This unlikeness is above all plain in God's different way of possessing existence; God alone possesses it intrinsically, is able to communicate it (*Summa theologiae* q.45, a.5; q.104, a.1), contains only active rather than passive potentiality, and is the ultimate cause of the being and continuance of anything with any real kind of contingency. Aquinas views God's act of existing (identical with his acts of living, knowing and loving) as identical with what he is (his *essentia*, his stuff, being or nature).

This last embodies Aquinas' rejection of self-causation, but its key importance is in marking the misleadingness of the ordinary structures of human language when applied to God. The subject/predicate structure of our statements does not make them false with regard to God, but makes them misleading. True, God exists, lives and loves, but our linguistic way of speaking ('by composition and division') suggests distinctions between the subject and the attribute predicated which are not real in the case of God. A term's 'way of signifying' (*modus significandi*) differs according to whether it signifies concretely ('God' or *Deus*, 'wise' or 'is wise') or abstractly ('deity', 'wisdom'), although the thing signified (*res significata*) is the same. Although we can make true statements about God, our language-formed way of thinking is in this way radically inappropriate, suggesting (for example) that God is one among many others who live, are good, know, act and exist. On the contrary, he cannot be grouped with them in the same genus - even the widest - as if the terms concerned might apply in the same sense or way. Rather, God not only exemplifies these attributes more perfectly, but also is his own life, goodness, and so on, and is the archetype (cause and *principium*) of life, goodness, and so on, in others - and this eminence implies the denial of every finite mode of possession of these attributes. Such is Aquinas' conception. At the same time, he always assumes the concreteness of God, the three divine persons being three primary substances in Aristotle's sense, rejecting the idea that God might be Existence, Life and Love in the Platonic sense of abstracted common properties shared by things which exist, live or love.

3 Later developments

The importance of the negative way appears in its contrast with rival, more positive theologies. Duns Scotus insisted that the terms for the divine perfections are used in the same sense of God and of creatures (univocally, not merely analogously), a view extended in Descartes' conception that we have a clear and distinct idea of infinity as attributed to God, as well as of these perfections. More recently, an empiricist natural theology has tended to model accounts of the relation between God and the world on a dualist conception of the relation between mind and body, and to rely on arguments for God from design (thinking of God's motivations in a parallel way to human ones) more than on cosmological arguments. This seemingly more common-sense approach rejects traditional ideas of God's simplicity, non-temporal nature and foreknowledge, and conceives his eternity only as an everlasting existence in parallel with creation, and therefore as compatible with a changing or developing God, as in process theology (see Process theism §§1-3). Such empiricism tends to conceive infinity in terms of magnitude, rather than in terms of the negation of all finite properties and finite modes of realization of 'perfections' (see Mitchell 1971, papers 1-3).

This 'common-sense' theology thinks of our capacity to speak of God as based on a supposed likeness of God to creatures, supposing the terms we use to apply literally only to creatures and to God in a derived or secondary sense. Yet this might seem absurd or even blasphemous in the case of the so-called 'perfections' - as if God could exist in a less real sense than his creatures, live, know, understand, love, be good or have active or causal power in a less full sense than creatures. In the older view the terms applied unrestrictedly to God, but only in some limited ('finite') respect to creatures.

These tendencies encourage an idolatry of creature-likeness in place of the Platonic idolatry that makes God an abstract entity; both these were rejected by Aquinas, and both invited Pascal's reaction, that the God of Abraham, Isaac and Jacob is not the god of the philosophers. Since 1700, the antipathy of theologians and spiritual men and women towards philosophical theology has steadily intensified. This reaction, rejecting Hegel's idea that God's nature and purposes can be read from history, reached its extremes in Kierkegaard and Karl Barth.

4 The negative way and arguments about meaningfulness

Does the negative way destroy the possibility of religious language's having any cognitive meaning? In the classic picture, since we know various positive statements about God to be true, and thence also know them to be meaningful (in the strong sense of being coherent enough to be capable of truth), some content remains, and not mere nonsense, even after all our negative qualifications. What we say, whether from revelation or tradition or from these assisted by philosophy, constitutes a pointer, and, in a structure of related statements, embodies some beginnings of thought or understanding, albeit not comprehension. It may contain paradox, defeating the imagination, but not intrinsic self-contradiction. Atheism and agnosticism, however, transform this situation: if we do not know that what is said is true, we have no way of knowing that any cognitive content at all remains behind after so many negative qualifications together with the adverbs 'infinitely' or 'unsurpassably'. The idea of the

infinite as a degree of intensive magnitude is unhelpful, since the atheist or agnostic has been given no reason why there should be any maximum in modes of existence, life, goodness, knowledge, love or causal power. There is then no ground for supposing these statements meaningful in any fuller sense than science-fiction stories or fairy tales - indeed they lack even the benefit of the imaginability that gives these stories their power.

This problem, of whether religious language has any cognitive meaning, is not just insoluble for the atheist and agnostic, but for any philosophy of religion which makes it a point of method always to consider questions about the meaning of types of statement (in this case, about God) in advance of considering their truth - as if to agree a neutral ground specific to the philosophy of religion from which to debate with agnostics or atheists. The supposed problem of meaningfulness is only solvable if knowledge of the truth of at least some of a class of statements is prior to knowledge of meaningfulness - for example, because given by revelation (the only solution admitted in Karl Barth's theology), by philosophy (questions of the meaningfulness of the terms we use being resolved in general philosophy, before we turn to using the same terms with regard to God) or somehow by 'experience'. Always some have assumed truth on the basis of faith, independently of and prior to considering how meaningfulness arises, and this has allowed a negative theology grounded in faith to be endemic, sometimes in distinguished figures scarcely noticed by later tradition, such as Nicholas of Cusa.

Older Jewish, Christian and Muslim traditions are strongly set against appeals to 'experience' as the key to knowledge of truth and meaningfulness. In this life, we know God by faith, not by sight; inner or outer experience can always be supposed in retrospect or by another person to have a different explanation from that which the experiencer thought certain at the time, and therefore cannot be the basis of the certainty of faith. If we follow the phenomenologists and Wittgenstein, no two 'religious experiences' described in quite different terms with different kinds of object or relations to different systems of motivation could be the same. The writings of R.C. Zaehner are typical of many suggesting just such differences (see Zaehner 1958). But with the increase of emphasis on 'experience' in religion over the nineteenth and twentieth centuries (see Phenomenology of religion §§1-2), and the decrease in the value put on teaching, ritual and reasoning in this sphere, the negative way has eased the way towards supposing that all the main world religions are set on the same objects and values, whether the 'ultimate' is described in impersonal or personal terms.

See also: Mysticism, history of §6; Patristic philosophy §2; Religion and epistemology; Religious language; Religious pluralism

DAVID BRAINE

References and further reading

Aquinas, T. (1266-73) *Summa theologiae*, London: Blackfriars, 1961-70, Ia, qq.1-13, 44-5, 104.(Q.3 presents the basis of the negative way in Aquinas, which is illustrated in qq.7-12, and is the key to his understanding of analogy in q.13; the utter difference of God's way of possessing being is best explained in q.45, a.5 and q.104, a.1. Q.4 and q.13 explain how a statement's positive and non-metaphorical character can remain despite negative qualification, and q.1 explains the role of metaphor.)

Dupré, L. (ed.) (1990) *Nicholas of Cusa*, *American Catholic Philosophical Quarterly* 64 (1), Washington, DC: Catholic University of America.(Useful introductory scatter of papers.)

John of the Cross (1966) *Collected Works*, ed. K. Cavanaugh and O. Rodriguez, London: Nelson; trans. A. Peers, *Complete Works*, London: Burns & Oates, 1934.(The poems themselves, reminiscent of the Scriptural 'Song of Songs', show how deeply the negative way is set within the context of a personal love-relationship.)

Lossky, V. (1957) *Mystical Theology of the Eastern Church*, London: James Clarke. (See especially pages 8-11 and chapters 2 and 4 for a common Eastern Orthodox perspective.)

Maritain, J. (1959) *The Degrees of Knowledge*, London: Geoffrey Bles.(Part II, especially chapters 8-9, focuses on the epistemological situation of John of the Cross' mystical writings.)

Mitchell, B. (ed.) (1971) *The Philosophy of Religion*, Oxford: Oxford University Press.(On the meaning of religious assertions: papers 1 (parts A and C), 2 and 3 present or discuss some modern cognitive approaches; papers 1 (part B), 4 and 7 present some modern noncognitive ones.)

Otto, R. (1917) *The Idea of the Holy*, trans. J.W. Harvey, London: Oxford University Press, 1923.(Philosophically expressive of the idea that emotions have to have an object.)

Tanner, N.P. (ed.) (1990) *Decrees of the Ecumenical Councils*, vol. 1, London: Sheed & Ward, and Washington, DC: Georgetown University Press.(This standard, widely available work provides both original text and translation.)

Zaehner, R.C. (1958) *At Sundry Times*, London: Faber & Faber; Westport, CT: Greenwood, 1971.(Detailed contrasting of the differences in object of different types of mysticism; see also his *Mysticism: Sacred and Profane*, London: Oxford University Press, 1957.)

Nemesius (*fl. c.*390-400 AD)

Nemesius' treatise De natura hominis (On the Nature of Man) is the first work by a Christian thinker dedicated to articulating a comprehensive philosophical anthropology. Like many of his pagan and Christian contemporaries, Nemesius employs Platonic, Aristotelian and Stoic ethical psychologies in developing his views, but he also provides remarkably detailed accounts of the physiological structure and function of the sense-faculties based on extensive medical knowledge.

Information on the life of Nemesius is limited. He probably became bishop of Emesa in Syria during the last decade of the fourth century AD, and his philosophical treatise *De natura hominis (On the Nature of Man)*, written in Greek, dates from this same period. As late as the Middle Ages this work was attributed to Gregory of Nyssa, his better known contemporary, though seventh-century writers such as Maximus Confessor recognized Nemesius as the author. Such confusion was common in a literary tradition which stipulated an impersonal style of writing, often accompanied by an injunction to avoid originality.

Nemesius received an excellent liberal Greek education, which was broadened to include extensive knowledge of the medical tradition stemming from Hippocrates and Galen and deep familiarity with the spiritual metaphysics of late antique Neoplatonism (see Hippocratic medicine; Neoplatonism). Scientific knowledge about the human body and its physiological functions is harmonized with the Platonic doctrine of the soul; Aristotle's theory of free choice (see Aristotle) is reconciled with Christian providence. Nemesius relies heavily on the voluminous writings of Galen on anatomy and physiology. His treatise is a rich source of information on a wide variety of topics and on lost works by Galen and by other Christian authors, such as the *Commentary on Genesis* of Origen. The localizations of the sense and cognitive faculties in various parts of the brain were transmitted to Albert the Great and Thomas Aquinas in the thirteenth century by Latin translations of *De natura hominis*.

Despite the rich philosophical and scientific content of the work, its focus is religious and spiritual. It is clearly a piece of Christian apologetic addressed to Hellenic intellectuals not yet committed to Christianity. Nemesius' informative, multi-layered doxography, with its extensive medical knowledge of the body, made for an anthropology sophisticated enough for educated Hellenes yet also consonant with Christian doctrines. This interest in the concrete is reflected in his theological adherence to the school of Antioch, which stressed the historical character of the Gospel and the humanity of Christ.

The long and important first three chapters of the treatise aim at harmonizing Platonic and Aristotelian psychology and Platonic and Stoic metaphysics. On the latter topic, Nemesius endorses the Platonic/Neoplatonic hierarchy and continuity of being, but stresses the continuity from one level to the next, from inanimate entities through animals to human beings and beyond to the intelligible world. He objects to the radically transcendent character of Neoplatonism because it leaves discontinuous gaps in the structure of reality. This more vitalistic and less hierarchical metaphysic has been traced by some to the influence of the first-century BC Platonizing Stoic Posidonius of Apamea, parts of whose largely lost works may have been preserved in Origen's lost *Commentary on Genesis*. At the centre of the picture, Nemesius situates human nature as the link between the spiritual realm and the physical universe.

In the important second chapter on the soul, Nemesius begins with an extensive doxographical survey of psychological theories from the Presocratics through Aristotle, the Stoics, Epicureans, Neoplatonists, various Christian authors and the Manicheans (see Presocratic philosophy; Stoicism; Epicureanism; Manicheism). Adopting Plato's view that the soul is an independent substance, he rejects Aristotle's claim that the soul is a part of the composite human being. For Nemesius, then, the soul is a self-moving, incorporeal and immortal principle which employs the body as its instrument; it is not the body's perfection or entelechy, as Aristotle held. Chapter 3 of *De natura hominis* tackles a basic problem: if soul and body comprise some sort of unity, how can the soul be immortal? Nemesius relies on Neoplatonists who argued that in its essence the soul is an immutable substance which, when united with a body, suffers no change or alteration. He illustrates the relation between soul and body by comparison with Christological doctrine: the being of the Logos is not diminished when united with human nature (see Soul, nature and immortality of the).

Subsequent chapters take up the location of various faculties in bodily structures and their physiological description. Chapters 6-11 respectively cover imagination, sight, touch (which is defined as the ‘consciousness of feeling due to nerves proceeding from the brain and spreading into every part of the body’), taste, hearing and smell, which are located in the frontal lobes of the brain; intellect is located in the middle of the brain (Chapter 12), and memory in the back lobes (Chapter 13). A distinctive feature of the localization of the sense-faculties in parts of the brain is the citation of empirical evidence (for example, brain lesions) derived from Galen. Chapter 13 also reveals Nemesius’ commitment to the Neoplatonic doctrine of the pre-existence of the soul in which souls recollect things they learned in the intelligible world before taking birth in bodies.

Various models of parts of the soul are reviewed in chapter 15. His analysis of the passions and pleasures in Chapters 16-18 relies on Epicurus’ distinction among natural, necessary and non-necessary desires which is grafted on to the Platonic tripartite model of the soul, consisting of rational, emotional and appetitive parts. Discussion of the nutritive, generative and life-sustaining faculties or capacities in Chapters 23-8 closely follows Galen. Chapters 29-34 develop a largely Aristotelian theory of action which adroitly employs Aristotle’s accounts of the voluntary, involuntary, deliberation and choice. In the ensuing Chapters 35-40, Nemesius relies on the power of human deliberative rationality to ground his attack on fatalism and his argument in favour of the freedom of the will (see Free will). Some have seen in this account of freedom and rationality evidence of Pelagian optimism about human nature (see Pelagianism). Nemesius insists, however, that free choice is a divine gift and that it is exercised exclusively within the orbit of God’s providence (Chapters 42-4).

See also: Neoplatonism; Patristic philosophy; Platonism, medieval; Soul, nature and immortality of the

JOHN BUSSANICH

List of works

Nemesius (*c*.390-400) *De natura hominis (On the Nature of Man)*, ed. M. Morani, Leipzig: Teubner, 1987; trans. W. Telfer in *Cyril of Jerusalem and Nemesius of Emesa*, Library of Christian Classics vol. 4, London: SCM Press, 1955.(This is the standard critical edition of the Greek text of *De natura hominis*.)

References and further reading

Jaeger, W. (1914) *Nemesios von Emesa*, Berlin: Weidmann.(Comprehensive treatment of the Greek philosophical influences on Nemesius.)

Telfer, W. (1955) *Cyril of Jerusalem and Nemesius of Emesa*, Library of Christian Classics vol. 4, London: SCM Press, 203-455.(Useful introduction and valuable chapter-by-chapter commentary accompanying the translation.)

Neo-Confucian philosophy

Chinese neo-Confucian philosophy, or 'neo-Confucianism', is a term which refers to a wide variety of substantially different Chinese thinkers from the Song dynasty (960-1279) through the Qing dynasty (1644-1911). In at least one respect the term is misleading, for unlike Neoplatonists, most neo-Confucians saw themselves as reviving, not revising, the earlier Confucian tradition. What united all these thinkers was a common allegiance to Confucius and his thought. Many of the central debates within the tradition concern the issue of who could claim to be Confucius' legitimate heir.

Despite their shared dedication to Confucius' legacy, a number of the central beliefs of neo-Confucians were unknown to Confucius and his early followers and appear to be at odds with early Confucian views. Many of these beliefs were part of a novel, elaborate and comprehensive metaphysical scheme linking human beings (as microcosm) to the universe (as macrocosm). Such cosmological theories provided a new ground for Confucian ethical claims and strengthened a tendency towards the mystical identification of the self with the universe. These changes also helped to transform the earlier Confucian concern with self-cultivation and steady moral improvement to a more dramatic quest for spiritual enlightenment, replete with a distinctly Confucian style of meditation.

One widely accepted account of the rise of neo-Confucianism sees it as a reaction on the part of Confucian scholars to the perceived dominance of Buddhist thought. On this view, Confucians had become complacent in the period between the end of the Han dynasty, in the early third century AD, to the beginnings of the neo-Confucian movement late in the ninth century. Fearing that their way of life was in peril, these later Confucians resolved to overcome their Buddhist competitors and revive their tradition. Purportedly, these later Confucians realized that in order to accomplish these goals they would need to develop and deploy an account of the Confucian tradition that could compete with and overcome the complex metaphysical schemes the Buddhists had used to argue against them. For these strategic reasons, early neo-Confucians focused their attention on texts like the Mengzi (Mencius), Daxue (Great Learning) and Zhongyong (Doctrine of the Mean) and began to evolve a version of Confucianism that was supported by the kind of comprehensive and complex metaphysical system described above.

A different account of the rise of neo-Confucianism can be given, however, in which the selection of certain texts and the development of new styles and practices of reasoning are not seen as self-conscious strategic borrowings from Buddhism and Daoism. Rather, these characteristic features of neo-Confucianism are consequences of the profound and pervasive effect that Buddhist and Daoist thought had exerted for centuries upon Chinese intellectuals. There was no organized or distinct group of thinkers who identified and thought of themselves as Confucians in the centuries immediately prior to the neo-Confucian revival. Rather, Chinese intellectuals had come to accept a wide range of philosophical ideas and spiritual practices as part and parcel of literati culture. When certain of these broadly read and eclectically trained individuals began looking back to the writings of early Confucian figures, they did so through the categories and with the concerns and approaches of their age. Their angle of vision had changed, and as a result they saw these earlier sources differently. This new orientation led them to elevate certain texts to canonical status; in some cases, these were texts that were unknown to the sages they claimed to follow. Moreover, these later Confucians developed particular aspects of the classical writings in ways that would have been unrecognizable to the founding figures of the tradition they so adamantly defended.

The great impetus for the revival of Confucianism came with a series of social and political crises that came to a head roughly in the middle of the eighth century. Under the dual pressures of internal rebellion and economic distress and external attack on a number of different fronts, Chinese intellectuals began to question their fundamental beliefs and practices. Rather than calling for reform or progress, such reflection led them back to their roots. Increasingly, they came to believe that the primary source of their troubles was that they had lost their Way. Like the foreign enemies who plagued them from without, they were being undermined by a foreign force from within: the non-Chinese religion of Buddhism and its spiritual cousin Daoism. These had led them to abandon the true source of their strength and former glory: the culture of classical China. Their course was clear. They must retrieve and revive the classical culture that was their very essence. In so doing they would replay the roles their most revered sages, Confucius and Mencius, had played in their own degenerate ages. This call to defend and promote 'this culture' served as the rallying point for the neo-Confucian movement.

Towards the end of the neo-Confucian period, Chinese Confucians themselves began to recognize the degree to which earlier neo-Confucians had incorporated beliefs and styles of reasoning that were alien to Confucius' way of thinking, specifically ideas drawn from Buddhism and Daoism, into the tradition. A number of these Qing dynasty thinkers saw their task primarily in terms of purging the tradition of these foreign elements in an effort to reconstitute a purer form of Confucius' original vision.

1 Historical background

During the 'Northern and Southern Dynasties' (AD 317-588), China experienced a prolonged period of instability and upheaval. Buddhism, which had arrived in China as early as the first century AD, became the dominant religious and intellectual force in the country (see Buddhist philosophy, Chinese). Daoism also flourished and became a powerful and elaborate institutional religion (see Daoist philosophy). Under the short-lived Sui dynasty (589-618), Buddhism was adopted and promoted as the state ideology, and it continued to flourish and develop in the succeeding Tang dynasty (618-905). This period witnessed prolific growth in Buddhist philosophical ideas and systems of thought, several of which were to have a profound effect on neo-Confucian philosophy. Daoism also continued to exert a strong and pervasive influence throughout Chinese society. The Tang imperial family claimed one of the mythical founders of Daoism, Laozi, as an ancestor and actively supported and practiced Daoism.

These three traditions, Confucianism, Daoism and Buddhism, did not always coexist peacefully. There were a number of Buddhist persecutions instigated by Confucian and Daoist elements within the imperial court. The worst of these occurred during the Tang dynasty, from 841-5, when a Daoist emperor launched a severe and wide-ranging persecution of Buddhism. To a large extent, this reaction against Buddhism was brought on by the decline of Chinese political and economic fortunes in the closing years of the Tang. This anti-Buddhist sentiment was to play a major role in the resurgence of Confucianism.

The founding of the Sui dynasty had marked the revival of a centralized Chinese state. With it came bureaucratic institutions and practices, among them the civil service examinations, modelled on paradigms drawn from the earlier Han dynasty (206 BC-AD 220). The Sui rulers turned to these neglected Confucian institutions and practices as the most effective available means of organizing and governing their state. However, there was little concomitant interest in the philosophical ideas behind these institutions and practices. Buddhism and Daoism continued to be the primary interests of intellectuals during the Sui dynasty and throughout most of the Tang dynasty.

In the middle of the Tang period, the Chinese empire began to experience intense economic and political strain. Because of rapid population growth, increased fluidity between socio-economic classes, the monetary drain of large, tax-exempt monastic estates, rising military expenditures and a lack of leadership by the imperial court, the tax system began to break down. In addition to these economic woes, in 751 the Chinese state suffered defeat in two important military campaigns, against a Thai army in the south and an Arab army in the west. These military debacles precipitated a disastrous internal rebellion by a young general named An Lushan in 755. After this, the Tang court was never again fully in control of China.

Not long after these events, Han Yu began what was to become the Confucian revival. His blistering criticisms of Buddhism as a socially parasitic and quintessentially anti-Chinese religion are among the earliest and most powerful statements of sentiments that were coming to the surface during the later Tang dynasty, feelings fuelled by the indignation and frustration arising from the economic and political difficulties described above. Han Yu's condemnation of Buddhism as a foreign system of thought which had invaded China and weakened it - laying the ground for the foreign invaders who were now plaguing Chinese territory - provided a focus and point of departure for what was to come.

Han Yu argued that the Chinese needed to rid themselves of this foreign influence and return to their true roots. These ran back to Confucius and the idealized past culture which he preserved and advocated. Daoism and Buddhism were both accused of being 'other-worldly' and incapable of providing guidance in the all too real world of domestic and foreign challenges. Han Yu himself was not philosophically inclined. His great talent was literature, and his belief in the need to revive ancient literary style expresses both his own predilection and the degree to which he saw ideas in terms of the specific cultural practices of an idealized past. Nevertheless, he provided the founding vision and voice for the Confucian revival.

2 Central concepts and terms

Traditional accounts of the rise of neo-Confucianism agree in describing it as a reaction against Buddhism and Daoism. However, most overlook the economic and political events which precipitated this reaction. Moreover, they contend that the central texts and ideas which neo-Confucians embraced were consciously chosen for their efficacy in the fight against Buddhist and Daoist thought. According to such accounts, texts such as the *Mengzi (Mencius)*, *Daxue (Great Learning)* and *Zhongyong (Doctrine of the Mean)* were chosen because these were the best sources from which to draw in defending Confucianism (see Daxue; Mencius; Zhongyong). The present account differs from this view in seeing the choice of texts and ideas as a consequence of the pervasive and profound influence exerted by Buddhism and Daoism throughout the Chinese intellectual world of this period. When Confucians of the Song dynasty looked back to the early sources of their tradition, they saw them through many layers of Buddhist and Daoist influence. This altered their perception of what was most central to the tradition as well as their understanding of traditional philosophical concepts. In order to understand neo-Confucianism, it is important to realize how some of these ideas evolved.

Through its interaction with indigenous Chinese thought, particularly Daoism, Buddhism changed dramatically and developed distinctively Chinese forms. The early Buddhist belief that all imperfect aspects of reality are ultimately unreal became transformed into the idea that our less savoury aspects are not really part of our nature. According to this view, our fundamental nature is equally the nature of all things, part of a transpersonal *foxing* (Buddha-nature) which is reflected in each and every thing in the world. Once one realizes that one is part of all reality, one no longer will feel the inevitable anxiety and fear of the eventual loss of one's self at death; instead one comes to see the common belief in an enduring and separate self as the source of all suffering. A person with this 'right' view of things manifests universal compassion towards all sentient beings, as manifestations of Buddha-nature (see Buddhist philosophy, Chinese).

Since the beginning of their tradition, Confucians had offered competing theories about the true character of human nature (see Xing). It was therefore natural that the notion of Buddha-nature had a profound influence on neo-Confucian views. Neo-Confucians began to talk about *benxing* (original nature) and *qizhi zhixing* (material nature), the former perfect and complete, the latter flawed and needing refinement. These terms are not found in pre-Buddhist Chinese philosophy. Within this conceptual scheme, Mencius' claim about the original goodness of human nature came to be understood as referring to the original, perfect and pure state of human nature rather than to certain of its nascent tendencies (see Mencius).

In order to appreciate fully this new view of human nature, we need to understand two additional neo-Confucian terms of art: *li* and *qi*. *Li* ('pattern' or 'principle') appears in early Daoist texts and means the underlying pattern running throughout the world. Different things, including human beings, possess their own individual structures or 'patterns' and contribute to a larger, grand 'pattern' or scheme (see Li). Under the influence of Buddhist philosophy, particularly the notion of Buddha nature as described above, this concept changed dramatically. The underlying pattern of the universe came to be seen as completely present in every mote of dust, with each aspect of reality reflecting all others.

However, neo-Confucians believed that any given thing only manifests certain particular *li* and this is what makes a thing the thing it is. Only certain *li* are manifested because of the effect of *qi* (ether). *Qi* is what makes up the world. It is a kind of lively - not inert - matter which exists in various grades of purity. As a function of its purity, the *qi* of different things obscures, to varying degrees, the *li* that are within them, only allowing some to shine through. So while all things equally possess all the *li*, their different endowments of *qi* make them different. Human beings are unique in that they alone have the ability to become aware of the *li* within them by refining their *qi* to the highly tenuous state which allows all the *li* within to shine forth. Other creatures are stuck at different levels of 'clarity'; the *qi* of inanimate things is so 'dense' that they lack consciousness (see Qi). Neo-Confucians tended to equate *li* with the 'original nature' of human beings and contrasted this with the idea of 'material nature', which is *li* embedded in *qi*. The task of moral self-cultivation was to refine one's *qi* and move from relative ignorance to complete and comprehensive knowledge (see Self-cultivation in Chinese philosophy).

3 The Song dynasty (960-1279)

These terms of art and the basic concepts were developed and made systematic during the Song. In particular, the

speculative metaphysics that came to serve as the foundation for neo-Confucian philosophy was forged during the early part of the dynasty. These early Song thinkers did not always share the social, political and ethical concerns of later neo-Confucians.

Shao Yong was one of the first early Song figures associated with the rise of neo-Confucianism in the eleventh century. His major interest was explaining the cosmology of the *Yijing (Book of Changes)* (see Yijing). Shao is distinctive for linking his cosmological speculations to an elaborate system of numerology. In this regard, his thought is not unlike that of the Pythagoreans (see Pythagoreanism) or Leibniz, who was familiar with some of his ideas.

Shao assumed that there was a discernable correlation between the hexagrams of the *Yijing* and the structure and function of the universe. Hence, anyone with a proper understanding of the text could not only understand but foretell events in the world. He wove his philosophical system around a passage from the *Xicizhuan (Commentary on the Appended Phrases, also known as the Dazhuan (Great Commentary))* of the *Yijing*, which he believed described this relationship: ‘And so within the Changes is the *taiji* [Supreme Ultimate]. It produces the two Principles [symbolized by *yin* and *yang*]. The two Principles produce the four Forms. The four Forms produce the eight Trigrams’ (*Xicizhuan* A11). Shao read this passage in terms of numerology, relying upon the notion of *shen* (spirituality). *Shen* is an active and impersonal conscious power which both animates things in the world and allows one to understand them. It is able to act immediately at a distance, and as understanding can instantly ‘penetrate’ throughout the universe. Shao believed that *shen* produces numbers, corresponding to the initial distinction of *yin* and *yang* (see Yin-Yang), and that from number the ‘images’ of things (the four Forms) arise. These then generate the actual things and events in the world. Human beings have *shen* as part of their nature; they can therefore tap into its activity and understand its movement by grasping the underlying numbers which give rise to all things.

Following this scheme, Shao saw the world in terms of sets of four (corresponding to the four Forms): there were four earthly substances (water, fire, earth, stone), four livings things (animals, birds, grasses and trees) and so on. In addition to these fundamental classes of things, Shao used this scheme to describe four phases of history, which recur in unending cycles of fixed numerical length. He showed little interest in many of the basic issues that later came to characterize neo-Confucianism, and was not regarded by Zhu Xi (see §6) as part of the orthodox Confucian revival. Shao was neither anti-Buddhist nor anti-Daoist; he was not a moralist, a political or social reformer or a self-cultivationist. His most important work, the *Huangji jingshi (Supreme Principles Governing the World),* concerns as its title suggests the underlying patterns and principles of the universe. It is primarily a work of cosmology.

4 The Song dynasty (960-1279) (cont.)

Zhou Dunyi shared Shao Yong’s interest in the cosmology of the *Yijing*. However, he did not share Shao’s numerological theories, and his cosmological system placed human beings more clearly at the centre of the universe. He served briefly (1046-7) as a teacher of Cheng Hao and Cheng Yi (see §5). His two most important works, the *Taiji tushuo (Diagram Explaining the Supreme Ultimate)* and the *Tongshu (Comprehending the Book of Changes)* were instrumental in the development of neo-Confucian metaphysics. Zhu Xi regarded him as the founding figure of the neo-Confucian revival.

In the *Taiji tushuo*, Zhou presents a chart with accompanying commentary explaining how the Supreme Ultimate (*taiji*) contains within itself the two modes of stillness and activity. These give rise to *yin* and *yang* respectively. These produce the *wuxing* (five elements), which in turn generate the two fundamental principles: the heavenly principle, *qian*, and the earthly principle, *kun*. These then produce all the things in the world. Human beings are unique among creatures in that they receive the most pure forms of the five elements and thus are able to play a critical role in both understanding and guiding the course of the universe.

An important feature of Zhou Dunyi’s thought concerns the relationship between the unity of the *taiji* and the variety of things in the world. According to Zhou, the *taiji* produces the two fundamental *qi* ‘ethers’, *yin* and *yang*. As things take shape, they move through a progression of increasing distinctiveness; and yet, while they become separate individual things, all partake of the original unity of *taiji*. Zhou makes this idea explicit in his *Tongshu*, where he talks about how the myriad things are really one and the one is present in each of the myriad things. We

can see the ethical implications of this idea in a well-known anecdote about Zhou. He refused to cut the grass in front of his window, saying that he thought of the grass as he thought of himself. The idea that the universe is coextensive with oneself and that its comprehensive unity (*taiji*) is present in each and every aspect of it, becomes a cornerstone of neo-Confucian thought.

A source of considerable controversy among the neo-Confucians who followed Zhou was his equation of the Supreme Ultimate with *wuji*, the 'infinite' or 'ultimate of non-being'. Both these readings of *wuji* present problems. The first seems to be a tautology; the second seems perilously close to the Daoist claim that existence comes from non-existence, a claim most neo-Confucians would emphatically deny. However, *wuji* can be understood as describing the tenuous state of the universe prior to there being any discrete things present. Zhou would then be saying that this original undifferentiated state is the origin of all things.

Another point of controversy concerns Zhou's claim that one should make *jing* (stillness) one's guiding principle. The idea is that one should cultivate one's mind in order to attain the lucid and tenuous state of the *taiji* itself. The mind then can sympathetically detect the subtle, emerging form and direction of things, and one can act to maintain a state of universal harmony and balance. Cheng Yi, and later Zhu Xi, took exception to this aspect of Zhou's thought and sought to replace his emphasis on 'stillness' with an emphasis on another notion, represented by a different word (also pronounced *jing*) meaning 'reverential attention' (see §§5, 6).

Traditionally, Zhang Zai is the second great figure of neo-Confucianism. He was the uncle and teacher of Cheng Hao and Cheng Yi. His most important works include the *Zhengmeng (Correcting Delusions)* and the *Ximing (Western Inscription)*. The latter was originally part of the former work, but was inscribed on the west wall of Zhang's study and came to be regarded as a free standing text.

Zhang's thought was based on his interpretation of the *Yijing* and the *Zhongyong*. While he owed much to Zhou Dunyi and Shao Yong, he did not employ the charts that were central to their method of explication, nor did he show any interest in numerology. Zhang's interest in metaphysics was primarily as a guide to morals. He simplified earlier metaphysical theories, arguing that the *taiji* was simply undifferentiated *qi* (ether) which arises from an inchoate, primordial state called *taixu* (the supremely tenuous). The nature of human beings and all things contains all that is in the *taiji* and the human mind can attain knowledge of these things; the self and the universe are fundamentally one. However, in order for people to realize this unity they must attain a state of lucidity, purity and tenuousness, like that of the *taixu* itself, through a process of self-cultivation.

The driving thrust of Zhang's philosophy is this imperative to realize the fundamental unity of human beings with the rest of the universe. Zhang Zai described this goal as the task of forming one body with all things. Perhaps it would be more accurate to say the goal is realizing (both in the sense of grasping it intellectually and manifesting it personally) the underlying unity between the self and the universe. This idea is expressed beautifully in the opening section of his famous *Ximing*:

> Heaven is my father, earth my mother; even an insignificant creature such as I finds an intimate place between them. And so, all that fills the universe I take as my body and what gives it direction I take as my nature. All people are my brothers and sisters, all creatures my companions.

(*Ximing*)

5 The Song dynasty (960-1279) (cont.)

Cheng Hao and Cheng Yi were brothers responsible for the mature statement of neo-Confucianism. They elevated the notion of *li* ('pattern' or 'principle') to a pre-eminent position and identified the 'pattern' within things and events with *tianli* (heavenly principle), the pattern of the entire universe. Here we see an explicit statement of the idea (discussed in §2 above) of the universe being present in each of its parts. Cheng Yi is credited with the famous dictum: '*Li* is one but its manifestations are many.' The Chengs also explicitly equate this universal pattern or principle with *xing* (nature) and *xin* (mind) (see Xing; Xin (heart-and-mind)).

The Chengs do not talk about the notion of *taiji* and show little concern with how the universe evolved. Their metaphysical scheme consists of two fundamental notions: *li* (pattern) and *qi* (ether) (see Li; Qi). They are dualists but of a special kind. While they believe the universe is composed of these two constituents, they insist that *li* cannot be found apart from *qi*, and *qi* without *li* would lack any shape or meaning. In a sense, the distinction

between them is logical and not actual. For both brothers, the mind is *li*, and we come to understand the things of the world when we match up the *li* in our minds with the *li* of individual things and events (thus explaining the notorious problem of how we can recognize something as right). Most people are born with impure endowments of *qi* which prevents the *li* of their minds from properly matching up with things. Moral self-cultivation is the process of refining one's *qi* to remove the impediments to the *li* within. The ultimate result is complete and perfect knowledge of both self and the world.

The brothers differ most in their views concerning how to carry out this process of self-cultivation. Cheng Hao was the more mystical of the two and emphasized the power of moral intuition (in the sense of a feeling rather than an insight). He believed in a universal, creative spirit of life, *ren* (benevolence), which permeates all things just as *qi* permeates one's body. Playing on the multiple senses of the word *ren*, he likened an 'unfeeling' (that is, non-benevolent) person to one with an 'unfeeling' (paralyzed) limb. Neither realizes the unifying 'oneness' of themselves (in the case of the former, this involves failing to feel that he is 'one body' with all things). For Cheng Hao, the task is to locate and pay attention to this inner feeling and allow it to guide one throughout one's life.

Cheng Yi presented a more developed and detailed philosophical system. His method of self-cultivation urges one to awaken the *li* within the mind by perceiving the *li* within the world. One could do this by carefully attending to one's daily affairs, but since the classic texts of Confucianism present these 'patterns' in their clearest and most accessible form, the primary source for such understanding was study of the classics. In either case, as one came to understand a given *li* one was to *tui* ('extend' or 'infer') its interconnection with other *li* until one achieved a complete and comprehensive understanding of them all. This process must be carried out with an attitude of *jing* (reverential attention) in order to insure that one's knowledge is affectively appropriate as well as cognitively accurate. Such knowledge is *zhenzhi* (real knowledge) as opposed to *changzhi* (ordinary knowledge) (see Knowledge, concept of). Cheng Yi illustrates this distinction with an allegory about people who 'know' (that is, they have heard) that tigers are dangerous versus one who 'knows' because he has been mauled.

Both brothers employ a scheme of learning sketched in another text that was to become central to neo-Confucianism, the *Daxue (Great Learning)*. One was to *zhizhi* (extend knowledge) by *gewu* (investigating things). For Cheng Hao, this involved a regimen of introspection and internal self scrutiny with the aim of correcting any errant thoughts. For Cheng Yi, the process was more externally directed: one came to understand the world by a careful inspection of paradigmatic cases in the classics and practical problem solving in one's own life (see Moral education; Moral development; Self-cultivation in Chinese philosophy).

6 The Song dynasty (960-1279) (cont.)

Zhu Xi is rightfully regarded as the greatest neo-Confucian synthesizer. While he drew a great deal of his thought from Cheng Yi (see §5), he gave it novel form and introduced significant innovations of his own design. The system of thought that emerged quickly became one of the two primary 'schools' of neo-Confucianism: Cheng-Zhu Learning or *lixue*, the 'Learning of Principle'.

Zhu reintroduced Zhou Dunyi's term *taiji* (see §4) and explained that it is the sum of all the *li* in the universe. It exists before the universe comes into being and is reflected in every aspect of reality. He illustrated this latter point with the metaphor of the moon being reflected in countless bodies of water. Zhu was a stronger dualist than Cheng Yi, insisting that *li* existed before the universe did and would remain even if it were destroyed. However, at times he presents this as merely a notional, not real, possibility. For all practical purposes, one can never find *li* and *qi* apart from one another. This question generated a long and acrimonious debate among later neo-Confucians in China and Korea.

Zhu also emphasized a distinction between *daoxin* (the mind of the Way) and *renxin* (the human mind). The former was pure *li* and hence perfect; the latter was *li* embedded in *qi* and hence necessarily obscured and prone to error. The concept of *daoxin* was in a certain sense a limiting notion. Zhu did believe that people could attain this level of refinement (and become a sage), but even the minds of sages are a mixture of *li* and *qi* (albeit an extremely refined and limpid *qi*), and so they are still susceptible to error should they grow complacent. Given this view, even the sage must engage in a life of constant moral scrutiny.

Among Zhu Xi's many innovations was his grouping together of four texts, the *Analects*, *Mencius*, *Zhongyong* and *Daxue*, as a set called the 'Four Books'. This set soon became the core of neo-Confucian learning. Zhu's text and

commentaries to the Four Books served as the basis for public education and the civil service examinations from 1313 until 1905. Zhu also helped to shape the Chinese education system by working to establish *shuyuan* (academies), Confucian institutions modelled on Buddhist monastic schools. These served a critical role in the neo-Confucian tradition until they were absorbed into a system of state schools during the Qing dynasty.

Zhu Xi also established the notion of *daotong*, the 'transmission of the Way', as a central neo-Confucian concern. It was he who argued that the orthodox understanding of Confucius' teaching had been lost after Mencius and only recently retrieved by Zhou Dunyi, then handed down through Zhang Zai, the Cheng brothers and (by implication) to Zhu himself. Zhu's influence on later Confucian thinking and Chinese society cannot be overstated. His view of the later tradition and its place in the larger context of Chinese history still greatly influences contemporary scholarship on Chinese philosophy throughout the world.

Lu Xiangshan represented what became the main alternative 'school' within neo-Confucianism known as *xinxue*, the 'learning of the mind'. It was later associated with the thought of Wang Yangming (see §9) and so also came to be called 'Lu-Wang Learning'. Lu's metaphysical disagreements with Zhu Xi are slight but significant. They mainly concern the issue of the nature of the mind: Zhu held that the mind was *li* embedded in *qi*, whereas Lu insisted that the mind is *li*. In a related disagreement, Lu rejected Zhu's distinction between the 'mind of the Way' and the 'human mind', arguing that there is only one mind and that it is principle. These doctrinal differences reflect deep and important disagreements about the character of human nature and the proper method of moral self-cultivation. Lu believed that our true mind is an innate moral mind which we can access and bring to bear in all our thoughts and action. The deluded mind exists only because of a kind of self-deception regarding our true nature.

Lu saw Zhu Xi's strong dichotomy between *li* and *qi* as establishing a corresponding dichotomy between moral knowledge and human desire. Lu thought that if one embraced this dichotomy, one would come to see self-cultivation in terms of the suppression of one's emotions and the accumulation of disconnected empirical facts about the world instead of the discovery and cultivation of one's innate moral feelings. One would then come to confuse broad learning with moral wisdom. The effort to acquire such learning would lead one into competition with others, which would only further obscure one's innate moral feeling of benevolence for all. The knowledge being talked about here was primarily knowledge of classical texts, the kind of knowledge that was the primary prerequisite for success in the civil service exams; and so, another danger associated with Zhu's method was the confusion of the quest for moral wisdom with the search for worldly success.

In contrast to Zhu, Lu argued that the human mind (in particular, our moral sense) was both necessary and sufficient for self-cultivation. Without it, one would fail to develop the affective dimension of 'real knowledge'. If one could only keep one's innate moral sense before one, it would guide one to identify and remove and selfish desires obscuring the *li* of the mind. Lu's difference with Zhu can be seen in his attitude towards the classics. Lu was not opposed to study of the classics *per se*, but he thought that a deep and personal understanding of some small part of them was preferable to the comprehensive knowledge Zhu Xi recommended. In particular, Lu advocated coming to a personal understanding, a kind of *verstehen*, of the mind of the sages as revealed in the classics. His position follows easily from his belief in a transpersonal 'mind' shared by all people and is neatly summed up in his well-known maxim regarding study of the classics: 'If only one understands what is fundamental, the six classics are all one's footnotes' (see Chinese Classics; Xin (heart-and-mind); Self-cultivation in Chinese philosophy).

The differences between Lu and Zhu were made clear in two debates between them, held at Goose Lake Monastery in 1175. In these debates, the intimate relationship between their metaphysical differences and their views on moral self-cultivation are evident. Lu accused Zhu Xi of ignoring the task of *zundexing* (honoring the virtuous nature) at the expense of *dao wenxue* (pursuing inquiry and study). In effect, he was accusing Zhu of forsaking what was commonly regarded as the most fundamental tenant of Confucianism: the innate goodness of human nature. In reply, Zhu accused Lu of mistaking his own subjective view of things as universal truth. In essence, this was to accuse Lu of the ultimate neo-Confucian error: selfishness.

7 The Yuan dynasty (1280-1368)

The Yuan dynasty marked the beginning of nearly one hundred years of Mongol rule over China. During this

period, elite Chinese culture gave way to the nomadic, tribal customs of the conquering Mongols. In addition, the Chinese were exposed to a variety of other, equally alien cultural influences that came in the wake of far-ranging Mongol conquests. At the same time, Chinese influences were carried along in the flow of Mongol victories and spread to other lands.

While Mongol influence was strong at the court and higher echelons of Chinese society, it often coexisted peacefully with indigenous Chinese beliefs and practices and probably had little influence on the lives of the vast majority of Chinese. On the one hand, this was the result of the Mongol practice of general non-intervention in local customs and practices, due in part to their inability to manage their immense realm on a local level and in part to their tendency to tolerate diversity as long as it posed no threat. On the other hand, this policy was simply a function of the substantial numerical superiority of the Chinese. Like earlier conquerors of China, the Mongols realized that existing Confucian institutions were indispensable for ruling their newly-gained empire. These institutions helped preserve much of neo-Confucian thought and even contributed to its acceptance as the orthodox state ideology and its spread to foreign lands.

Cheng-Zhu Learning did particularly well during this period and in 1313 was officially adopted as the state orthodoxy. The reason for the triumph of this ‘school’ of neo-Confucianism and the decline of the teachings of Lu Xiangshan had a good deal to do with the relative superiority of Zhu Xi’s system as a basis for the civil service examinations. It was much more detailed, systematic and accessible. Zhu had composed extensive, meticulous and persuasive commentaries on all of the Chinese classics. He had provided a handy summary of Confucian learning in the form of the text of and accompanying commentaries on the Four Books, and he had also produced numerous anthologies and primers of Confucian learning. One of these, the *Xiaoxue (Elementary Learning)*, intended for the moral education of the young, found wide appeal among the Mongols themselves.

Zhu Xi’s thought provided a well-defined system that was basically quite conservative in nature, well-suited to the role of a bureaucratic ideology under occupation. This is not intended to slight the genuinely moral aspects of almost all of Zhu Xi’s writings; had his philosophy not provided concerned Chinese intellectuals with a powerful moral calling, it would undoubtedly not have succeeded as well as it did throughout the Yuan dynasty and on into our own time. Its great power lay precisely in its ability to channel moral and patriotic feelings in constructive directions under a variety of circumstances. Given the situation at the time, Zhu’s philosophy provided a way for the Chinese to continue their own culture and to a significant degree Sinicize the conquering Mongols. However, Zhu Xi’s philosophy is not easy to see as a revolutionary creed, and during the Yuan dynasty we see little if any of the philosophical creativity and debate that marked the preceding dynasty and the one to follow.

8 The Ming dynasty (1368-1644)

The Ming period marked a revival and development of Lu Xiangshan’s challenge to the Cheng-Zhu orthodoxy. One of the earliest figures in this movement was Chen Xianzhang, also known as Chen Baisha (1428-1500). Chen argued that the strong dualism of the Cheng-Zhu school did not reflect the central neo-Confucian belief in the inherent unity between human beings and the universe. In particular, the strong distinction between *li* and *qi* tended to value speculative theory and the intellect at the expense of practical concerns and the emotions. This resulted in a general distrust of feeling and intuition and a gross overemphasis on the intellect and study of the classics as the proper methods of moral self-cultivation.

Chen argued for the inherent unity of human beings and the universe. He also promoted the idea that human beings spontaneously move in harmony with the natural realm and fail to do so only when they impose their own selfish ideas upon the world. In this latter view in particular, as well as in several other aspects of Chen’s thought, one can see the influence of the early Daoist philosopher Zhuangzi.

Chen’s criticism of the Cheng-Zhu school’s overreliance on intellect and study and his belief in the reliability of spontaneous intuitions to guide human action shifted the focus of moral self-cultivation away from study of the classics and speculative philosophy and towardss reflection and self-scrutiny. One could describe this as a shift towards ‘subjectivity’, but only with the qualification that he believed the intuitions of properly cultivated individuals would agree in every significant respect. He was not advocating individual expression in the modern Western sense of expressing one’s unique individuality; he was encouraging the discovery and personal expression of a shared, transpersonal mind. Any strong sense that one’s actions were uniquely one’s own was a sign of

selfishness - not of the moral mind. Chen's views greatly diminished the prestige of the classics as the ultimate source of moral knowledge. In his view, individuals possessed a moral sense that was at least equal in value as a moral guide.

A related feature of Chen's views is his belief that moral judgments are highly context-sensitive. The difficult situations one encounters in life are so complex, nuanced and specific that no appeal to rules or precedents could ever provide adequate guidance. Fortunately, each person possesses an inherent moral guide which one can make contact with and engage through a process of reflection and inner self-scrutiny. This aspect of Chen's thought further eroded the status of the classics and with it the plausibility of the Cheng-Zhu school's entire approach to moral self-cultivation.

9 The Ming dynasty (1368-1644) (cont.)

Many of these ideas were taken up and developed into a powerful challenge to the Cheng-Zhu school by the most important Ming thinker, Wang Yangming. Like Chen, Wang emphasized that the goal of moral self-cultivation was to realize that one was 'one body' with all things. Also like Chen, he believed in an innate faculty of moral intuition and the critical need to rely upon this faculty in making moral judgments. And like both Lu Xiangshan and Chen, Wang believed that the mind itself is *li* (principle) (see Li). However, Wang developed these ideas in novel ways and combined them with others of his own design to produce a much more sophisticated and powerful philosophical system than either of these earlier thinkers.

Wang borrowed a term of art from Mencius in order to describe our innate and infallible moral sense: *liangzhi* (pure knowing). *Liangzhi* is an ever-present faculty that will spontaneously guide us in making proper moral judgments if only we can succeed in eliminating the selfish desires which normally obstruct it. We eliminate these by bringing them to complete awareness, where they are consumed by the light of *liangzhi*. The task is to maintain a constant state of internal scrutiny, monitoring our nascent thoughts to ferret out and eliminate any existing or emerging selfish thoughts.

Given this picture, Wang took issue with Zhu Xi's teachings on moral self-cultivation (see §6). Zhu believed that the major focus of one's effort should lie in coming to understand *li* by engaging in *gewu*, the 'investigation of things' (in other words, investigating the principles in the classics and in the world). Wang first objected to Zhu Xi's emendation and rearrangement of the text of the *Daxue* in support of this view, and advocated the original form of this classic. According to Wang, the original text shows that *gewu* is not a preliminary step in a process of acquiring knowledge of *li*; rather, it refers to the act of preserving the inherent integrity of the mind by 'correcting one's thoughts'. The task of moral self-cultivation thus does not rely upon the acquisition of knowledge through empirical inquiry, but rather the restoration and preservation of the mind through the elimination of incorrect (that is, selfish) thoughts (see Self-cultivation in Chinese philosophy).

According to Zhu Xi's interpretation of the text, one is to first cultivate one's virtue in order to care for all people. Wang insisted that these are simply different aspects of the same event. One cannot cultivate oneself *without* caring for others, and properly caring for others *is* cultivating oneself. Zhu Xi's view created a false division between the moral mind and the universe; Wang insisted that they were one. Zhu Xi's view also created a division between moral knowledge and action which Wang insisted was both wrong and pernicious. Wang's response to Zhu on this issue resulted in his most well-known teaching: the unity of knowledge and action.

Wang relied upon the distinction, first developed by Cheng Yi, between 'real knowledge' and 'ordinary knowledge' to argue that anyone who really knew any moral truth must necessarily act upon it. Such action was a necessary constituent of real knowledge. One actually had to engage in filial activity in order to know what filial piety really is, and all those who possessed such knowledge would spontaneously act in accordance with their knowledge when faced with an appropriate situation. People who claim to know filial piety but do not act filially simply do not really know.

Wang saw his approach as addressing the most severe consequence of the Cheng-Zhu view. Wang believed that Zhu Xi's approach led people to regard moral self-cultivation as one thing and the affairs of their lives as another. Wang saw Zhu as counselling people to study the classics and calm themselves through quiet sitting in order to improve themselves morally, and then go out and face the world. Wang insisted that this could never work. First, like Lu and Chen before him, he insisted that the moral problems we face are extremely context-sensitive; no

appeal to precedents or rules of conduct can provide us with the guidance we need. As Wang was fond of pointing out, the sages had no such precedents when they acted.

Wang's deeper objection concerned the efficacy of Zhu Xi's method of self-cultivation. Wang did not believe that studying moral theory or contemplating moral paradigms contributed to one's moral development. Such pursuits often had the exact opposite result, for they easily became sources of additional selfish thoughts. In order to improve oneself morally, one needed to work on the actual moral problems of one's own life. Only these engaged one both cognitively and affectively, and only solving such 'real' problems would result in moral improvement.

This antinomian aspect of Wang's thought opened up moral self-cultivation to a much wider range of people, for one no longer needed to be a highly-educated intellectual to grasp the true meaning of the classics (in fact, such an approach would actually lead one farther from self-cultivation). It also gave to his thought an existential flavour; one was made profoundly aware of the weight of one's individual responsibility for one's own moral well-being. The anxiety of such an awareness may even have been critical to gaining true moral understanding. However, Wang's 'existentialism' differs in significant ways from most of its Western advocates. Unlike Kierkegaard, Wang held 'God' to be within each individual and reflected in every feature of the universe; unlike Sartre, existence for Wang does not precede essence for they are one and the same: the mind *is* principle. Properly cultivated individuals faced with similar situations would render similar judgments.

The idea that the mind itself is *li* led Wang to one of his most controversial teachings, known as the Four Sentence Teaching. In his *Chuanxilu (Instructions for Practical Living)*, he said, 'There is neither good nor bad in the mind itself. Good and bad are the activity of thoughts. To know good and bad is pure knowing. To do good and eliminate bad is *gewu* [correcting one's thoughts]'. The first sentence of this teaching seems to contradict not only the central neo-Confucian belief in the innate goodness of the mind, but also Wang's own claims about *liangzhi*. Perhaps Wang is saying that in itself the mind, being the sum of all the principles in the world, is simply how things are in their natural state. In such a state the predicates good and bad are simply inappropriate; they apply only in cases where there is an agent striving either to do good or bad actions. But the mind in itself has no intentions; it just is as it is. Similarly, the people who attain the highest state of moral self-cultivation are not aware that they are doing good. Those who have such thoughts are still striving for goodness and run the risk of becoming 'obscured' by selfish attachment to their own moral progress. The actions of sages, on the other hand, simply happen *through* them. Thus while the actions that result from the fully cultivated mind (that is, the mind itself) will all be good when seen from the perspective of those who observe them, the mind itself is neither good nor bad. Even to say that what the mind does is good is to view its actions from the perspective of the unenlightened. While this view of things profoundly diminishes a person's sense of individual agency, it also lends to their actions a remarkable feeling of necessity.

In the closing years of the Ming dynasty, Wang's ideas - in particular the Four Sentence Teaching - were interpreted by several of his later followers as warranting idiosyncratic judgments of right and wrong. Several of these followers became quite controversial, and a few even died in prison. Later neo-Confucians were to blame these individuals and, by implication, Wang himself for the eventual downfall of the dynasty. While such claims are clearly absurd, they do reveal the perception among the educated elite that the more radical of Wang's later followers showed a degree of independence that many found uncomfortable if not dangerous. These thinkers threatened the order and hierarchy characteristic of Confucian society and advocated such 'radical' ideas as the intellectual and moral equality of women. With the fall of the Ming, these ideas and the intellectual tendencies that gave rise to them came to a halt in a general and dramatic reaction against what was viewed as the Song-Ming drift into radical subjectivism. In its place emerged a movement that again called for a return to 'true' Confucianism based upon a solid 'objective' approach to the classics.

10 The Qing dynasty (1644-1911)

When the Manchus conquered China and established the Qing Dynasty, Chinese intellectuals once more found themselves facing the problem of how to explain the failure of the Confucian Way. Again, their answer was that the Way had been misunderstood and poorly practised. So began a prolonged and wide-ranging attack on the whole of the neo-Confucian tradition, from the Song through the Ming. One of the first and most important figures in this new movement was Wang Fuzhi.

Wang attacked both the Cheng-Zhu and Lu-Wang schools. Most of his objections turn on his rejection of their basic metaphysical schemes. He insisted that there was no *li* ('pattern' or 'principle') apart from *qi* (ether). *Li* is simply the 'pattern' of 'actual things and events'. Wang rejected the notion that there exists something called the *taiji* or anything like it which is the sum of all *li* and the related claim that all things somehow reflect all the *li* in the world. This of course meant that he also rejected the Lu-Wang school's equation of principle and mind. He believed that both schools tended to reify *li*, turning people's attention away from the actual things and events in the world and towards the search for some fanciful, ultimate *li*. He further believed that this 'other-worldly' speculation was a direct consequence of Daoist and Buddhist corruption of Confucianism and the primary cause of the defeat of the Chinese at the hands of the Manchus.

Wang's views had further quite profound implications. For example, while he admitted that the actual things and events in the world tended to fall into natural types and patterns and are related through an ongoing process of evolution, his strong denial of any ahistorial universal 'principles' and his belief in the steady historical improvement of social institutions and practices led him to advocate quite radical notions of social reform.

The next important figure in the Qing dynasty is Yan Yuan. Like Wang Fuzhi, he rejected both of the Song-Ming Schools. One of his main arguments against the Song Confucians rests upon his claim that they mistook the meaning of the word *wen* (culture), interpreting it instead as 'writing'. The Chinese graph he is referring to appears throughout the classics and can have either of these meanings, but Yan argued that Confucius' primary interest was 'culture'. Song neo-Confucians mistakenly believed that they were following Confucius when they engaged in the extensive writing of commentaries and explanatory essays on the classics. Yan responded by saying that Confucius only engaged in these activities when it became evident that no ruler in his time would employ him and put the Way into practice. He turned to the task of preserving the *dao* only when faced with this intractable fate; and he did so after having mastered the *dao* through years of diligent practice.

Yan advocated a return to the institutions and practices of the sages, and on this issue he differed dramatically from Wang Fuzhi and shows important similarities to Xunzi. One should practise the six arts which Confucius had taught: ceremony, music, archery, charioteering, reading and mathematics. Such concrete practice would lead one to sagehood. For Yan, *gewu* was neither the discovery of some underlying 'principle' (Zhu Xi) nor the rectification of one's 'thoughts' (Wang Yangming). It was the task of mastering the skills and abilities of the sages. For Yan, sages were individuals of sound mind, spirit and body: men of action. Yan harshly criticized the Confucians of his day for being effeminate and ineffective intellectuals, and he saw their personal moral failure manifested in the defeat of the Chinese state. Like many Confucians, Yan believed that those who were really following the Way could not but create a flourishing, peaceful and strong society. At times he almost seems to be arguing for a kind of pragmatism. However, what appears as a 'pragmatic' appeal is the widely held Confucian belief in the efficacy of the Way. In the face of defeat at the hands of the Manchus, Yan's natural response was that his predecessors had lost the true Way.

11 The Qing dynasty (1644-1911) (cont.)

The greatest philosopher of the Qing dynasty was Dai Zhen. Dai had a lifelong dedication to philosophy, but until recently was not appreciated for his philosophical contributions. He was however a highly respected mathematician, geographer, astronomer and philologist. This is largely a result of the particular *Zeitgeist* of the Qing, which had a profound appreciation of 'hard sciences' such as textual study and little tolerance for speculative philosophy. Ahead of his time, Dai denied this distinction.

One can gain a significant insight into Dai's views by studying the form of his two major philosophical works: the *Yuanshan (On the Good)* and the *Mengzi ziyi shuzheng (The Meaning of Terms in the Mencius Explained and Verified)*. Both of these are works of philosophical philology, commentaries on the meaning of key philosophical terms taken from the Confucian classics, which argue against common interpretations and provide both argument and philological evidence for a new view. Since Dai believed that the classics were the key to understanding the *dao*, his philosophical method necessarily contained a strong philological component. For one cannot understand the classics without understanding the words of these texts, and only sound philological method allows one to do this. On the other hand, Dai insisted that philological inquiry only had value when employed in the service of philosophy, for all inquiry should be directed towards an understanding of the *dao*.

His philological method reflects or perhaps more accurately prefigures his philosophical views. Dai believed one clearly cannot rely upon raw intuition in deciding the meaning of a term; one must resort to careful study of its uses throughout the classics. As one refines one's initial theory in the light of different cases, one moves closer to the correct meaning, the one that accounts for all the cases. This approach is clearly manifested in the two works cited above, where he provides extensive evidence to argue against both the Cheng-Zhu and Lu-Wang interpretations of the key terms of Confucian philosophy. Dai convincingly demonstrates how the interpretations of both schools are not attested in the classical sources and further argues for their Daoist and Buddhist origins. Both the Cheng-Zhu and Lu-Wang schools were overly subjective in the sense that they read into the classical texts their own personal views. This is one example of a general type of error that was the focus of much of Dai's attention.

For Dai, the most common error people make is to mistake their private 'opinions' (*yijian*) as 'unchanging standards' (*buyi zhize*). This applies to determining the meaning of a philosophical term as well as to making an ethical judgment; for Dai, these were but two sides of a very thin coin. In the case of ethical standards, one moved from 'opinion' to 'unchanging standard' by applying the Confucian Golden Rule: not doing to others what one would not want done to oneself (see Confucius). This mirrors the logic of moving from one's initial impression of what a term means to the correct interpretation: both involve a progression from personal opinion to a comprehensive or universal view. In the specific case of ethical judgments, Dai believed that this movement led one from what is 'spontaneous' (*ziran*) to what is 'necessary' (*biran*). That is, one begins with one's spontaneous reaction to a situation and, through an appeal to the Golden Rule, tests this reaction to see if it passes muster as a universal standard. If it does then the reaction is certified as 'correct', something that one 'necessarily' must do. If it fails this test, it is revealed as merely one's own selfish 'opinion' (see Confucian philosophy, Chinese).

Dai offered a remarkably novel and powerful interpretation of the Confucian vision. He is rare if not unique among Confucians for his strongly rational and intellectualist approach towards moral self-cultivation. Yet, while he provided compelling criticisms of the Daoist and Buddhist elements within earlier neo-Confucian thought, he himself was not without such influences. Unlike early Confucians who believed that one needed to acquire or at least extend and expand one's moral sensibility, Dai seems to have held that one already possesses the ability to make proper moral judgments. The problem for him is not that we lack the proper sense, but rather that we also have many improper reactions that must be identified and eliminated.

It is tempting to see Dai as one step away from a kind of Kantian view of ethics with his appeal to a type of universalizability criterion (see Kant, I.), but this would be to misrepresent Dai's thought. For one thing, he believed in an inextricable link between philosophy and philology: one could not find ethical truth apart from understanding of the classics. Dai did not see the Confucian Golden Rule (itself an idea derived from the classics) as free-standing; it did not generate all and only correct ethical judgments. It was a winnowing process to determine which among our spontaneous reactions were in fact in accord with the *dao*. These reactions were feelings about the rightness or wrongness of a given act, and if they were found to be genuine 'necessary' moral feelings, then this recognition produced a profound feeling of joy within the individual. Dai was very much a Confucian defending his tradition.

12 Conclusion

This survey of neo-Confucian thinkers is necessarily selective in terms of the figures covered. It also is confined to what might be called pre-modern philosophers. In the closing period of the Qing dynasty several extremely interesting and innovative thinkers such as Kang Youwei, Tan Sitong and, in the twentieth century, Feng Yulan produced new versions of Confucian philosophy which incorporated and synthesized their understanding of Western philosophy, religion and science. These new forms of Confucian thought continue to be developed in contemporary times, but as they represent a new and distinct stage in the Confucian tradition, they warrant a separate and detailed study in their own right.

See also: Cheng Hao; Cheng Yi; Chinese Classics; Buddhist philosophy, Chinese; Confucian philosophy, Chinese; Chinese Philosophy; History, Chinese theories of; Dai Zhen; Daoist philosophy; Knowledge, concept of; Li; Lu Xiangshan; Mencius; Moral education; Moral development; Qi; Self-cultivation in Chinese philosophy; Shao Yong; Wang Fuzhi; Wang Yangming; Xunzi; Zhang Zai; Zhou Dunyi; Zhu Xi

PHILIP J. IVANHOE

References and further reading

Birdwhistell, A.D. (1989) *Transition to Neo-Confucianism*, Stanford, CA: Stanford University Press.(An intellectual historical account of Shao Yong and his thought with philosophical sensibility.)

Black, A.H. (1989) *Man and Nature in the Philosophical Thought of Wang Fu-chih*, Seattle, WA: University of Washington Press.(An intellectual historical account of Wang Fuzhi and his thought.)

Chan Wing-tsit (1963) *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press. (Selective translations and brief introductions of most of the major neo-Confucian thinkers.)

Chan Wing-tsit (trans.) (1963) *Instructions for Practical Living and Other Neo-Confucian Writings by Wang Yang-ming*, New York: Columbia University Press.(A translation of Wang's most important works.)

Chan Wing-tsit (ed.) (1986) *Chu Hsi and Neo-Confucianism*, Honolulu, HI: University of Hawaii Press.(An anthology of essays on Zhu Xi and his thought. Note in particular Graham's contribution.)

Elman, B. (1984) *From Philosophy to Philology*, Cambridge, MA: Harvard University Press.(An insightful historical account of the cultural and institutional context of Qing Dynasty Confucians.)

Ewell, J.W., Jr (1990) *Reinventing the Way: Dai Zhen's Evidential Commentary on the Meanings of Terms in Mencius (1777)*, Ph.D. dissertation, University of California at Berkeley. (The best translation of Dai Zhen's most important work; includes a long introduction.)

Gardner, D.K. (1990) *Learning to be a Sage: Selections from the Conversations of Master Chu, Arranged Topically*, Berkeley, CA: University of California Press. (Translation with commentary of selections from Zhu Xi's works.)

Graham, A.C. (1992) *Two Chinese Philosophers*, La Salle, IL: Open Court.(The best study available on the Cheng brothers, containing many translated passages. Philosophically astute though it does not demonstrate a full appreciation of the role played by Buddhist and Daoist philosophy).

Huang Siu-chi (1977) *Lu Hsiang-shan: A Twelfth Century Chinese Idealist Philosopher*, Westport, CT: Hyperion Press.(A brief and quite dated study of Lu Xiangshan's thought. Still useful.)

Ivanhoe, P.J. (1990) *Ethics in the Confucian Tradition: The Thought of Mencius and Wang Yang-ming*, Atlanta, GA: Scholars Press.(A comparative study that seeks to provide a philosophical introduction to the thought of both Mencius and Wang which illustrates the contrasts as well as the continuities between them.)

Ivanhoe, P.J. (1993) *Confucian Moral Self-cultivation*, Frankfurt: Peter Lang.(A study of the origin and development of this characteristically Chinese philosophical concern in the Confucian tradition. Chapters 4-6 are devoted to Zhu Xi, Wang Yangming and Dai Zhen.)

Kasoff, I.E. (1984) *The Thought of Chang Tsai (1020-1077)*, Cambridge: Cambridge University Press.(A sinological study of Zhang Zai.)

Metzger, T.A. (1977) *Escape From Predicament*, New York: Columbia University Press.(An insightful study of neo-Confucianism. One of the few works that shows the proper appreciation of their metaphysical commitments and the relationship of these to their ethical views.)

Nivison, D.S. (1973) 'Moral Decision in Wang Yang-ming: The Problem of Chinese "existentialism"' *Philosophy East and West* 23: 121-38.(A very helpful study demonstrating that there are at best thin similarities between Wang and Western existentialists.)

Tu Weiming (1976) *Neo-Confucian Thought in Action: Wang Yang-Ming's Youth (1472-1509)*, Berkeley, CA: University of California Press.(A detailed historical study of Wang Yangming's early life and his struggle to attain sagehood.)

Neo-Kantianism

In contrast to earlier research, which chose to distinguish up to seven schools of thought within the field of Neo-Kantianism, more recent scholarship takes two basic movements as its starting point: the Marburg School and the Southwest German School, which are based respectively on systematically oriented works on Kant published during the 1870s and 1880s by Hermann Cohen and Wilhelm Windelband.

Cohen held that Kant's concern in all three Critiques was to reveal those a priori moments which above all give rise to the domains of scientific experience, morality and aesthetics. Windelband on the other hand held that Kant's achievement lay in the attempt to create a critical science of norms which, instead of giving a genetic explanation of the norms of logic, morality and aesthetics, aimed instead to elucidate their validity. In both approaches, an initial phase during which Kant's doctrines were appropriated subsequently developed into the production of systems. Thus Cohen published a 'System of Philosophy' during the early years of the twentieth century, which consisted of the Logik der reinen Erkenntnis (Logic of Pure Knowledge) (1902), the Ethik des reinen Willens (Ethics of Pure Will) (1904) and the Ästhetik des reinen Gefühls (Aesthetics of Pure Feeling) (1912) and which radicalized the operative approach of his work on Kant. Later, Cohen conceived a Religion der Vernunft aus den Quellen des Judentums (Religion of Reason from the Sources of Judaism) (1919). Windelband, on the other hand, who made a name for himself primarily in the sphere of the history of philosophy, understood philosophy to be essentially concerned with value, anchored in transcendental consciousness. He emphatically linked the classical division of philosophy into logic, ethics and aesthetics to the values of Truth, Goodness and Beauty and also tried to situate the philosophy of religion in this context.

Apart from Cohen, the Marburg School is represented by Paul Natorp and Ernst Cassirer, whose early works followed Cohen's philosophical views (compare Natorp's interpretation of the Platonic doctrine of ideas and Cassirer's history of the problem of knowledge), but whose later works modified his approach. Nevertheless, their extensions and developments can also be explained within the framework of the original Marburg doctrines. The ontological turn which Natorp undertook in his later years can be seen as a radicalization of Cohen's principle of origin, which Natorp believed could not be expressed in terms of pure intellectual positing, and the operative moment introduced by Cohen lives on as a theory of creative formation in Cassirer's theory of symbolic forms.

In addition to Windelband, the Southwest German school of Neo-Kantianism is represented by Heinrich Rickert, Emil Lask, Jonas Cohn and Bruno Bauch. Windelband instigated the systematic approach of the Southwest School, but it was left to Rickert to develop it fully. Unlike Windelband, who traced the difference between history and science back to the difference between the idiographic and the nomothetic methods, Rickert distinguished between the individualizing concepts of history and the generalizing concepts of science. During his middle period he turned his attention to the problem of articulating a system of values. In his later works, Rickert also turned towards ontology, a development which should not necessarily be interpreted as a break with the constitutional theories of his early years. In concrete terms, building on his earlier theories concerning the constitutive role of concepts in experience, Rickert henceforth distinguishes not only the realm of scientific and cultural objects and the sphere of values, but also the further ontological domains of the world of the free subject and the metaphysical world, which is the object of faith and which can only be comprehended by thinking in symbols.

Lask's theoretical philosophy was characterized by a turn to objectivism. In contrast to the classical Neo-Kantian conception of knowledge, according to which everything given is determined by the forms of cognition, Lask sees matter as that element which determines meaning. Accordingly, at the centre of his theory of knowledge is not the subject's activity in constituting the object, but the subject's openness to the object. In the final stage of his philosophy, however, he once more attributed to the subject an autonomous role in the actualization of knowledge. Cohn contributed to Southwest German Neo-Kantianism not only his Allgemeine Ästhetik (General Theory of Aesthetics) (1901), but also works on the philosophy of culture and education as well as on the systematic articulation of values and the problem of reality. During the 1920s Cohn moved towards dialectics. In contrast to Hegel, however, he understood this to mean critical dialectics inasmuch as it does not aim to sublate or overcome opposition, but merely sets itself the unending task of attempting to resolve irreconcilable contradictions.

Finally, Bauch can be regarded as the most essentially synthetic thinker of the Southwest German Neo-Kantian school. He tried to demonstrate the inseparable connectedness of individual problems which had generally been

treated separately. Apart from his great Kantian monographs, these ideas are also put forward in his systematic works on the questions of theoretical and practical philosophy, such as his study Wahrheit, Wert und Wirklichkeit (Truth, Value and Reality) (1923) and his Grundzüge der Ethik (Fundamentals of Ethics) (1935).

Despite the one-sidedness of its reception of Kant's doctrines, Neo-Kantianism was important for the momentum it gave to research into Kantian philosophy during the twentieth century. Its systematic achievement lies in its development of the normative concept of validity and its programmatic outline for a philosophy of culture.

1 Delimitations and affiliations

The term 'Neo-Kantianism' arose in the 1870s and owes its origins to the need to characterize that overall reassessment of the theories of Immanuel Kant which was then taking place in various different forms. The terminological problem concerns how broadly or narrowly the expression should be applied. Earlier research tended to work with definitions which encompassed ever-increasing circles. Thus T.K. Oesterreich claimed to distinguish seven different approaches to Neo-Kantianism: a physiological approach (of Hermann von Helmholtz and Friedrich Albert Lange), a metaphysical approach (of Otto Liebmann and Johannes Volkelt), a realistic approach (of Alois Riehl), a logicist approach (of Hermann Cohen, Paul Natorp and Ernst Cassirer), an approach based on a theory of values (of Wilhelm Windelband and Heinrich Rickert), a relativist approach (of Georg Simmel) and a psychological approach (of Leonhard Nelson). This classification is problematic for a number of reasons. Not only is it doubtful whether Simmel and Nelson should be included among the Neo-Kantians in the first place; it is also questionable whether the specific attitudes of Lange's particular understanding of Kant justifies his inclusion among the supporters of the physiological approach. Furthermore, it is unfortunate that, in the chronological list of the individual approaches, the metaphysical approach of Liebmann is listed before the logicist Neo-Kantians, for Liebmann's turn towards a metaphysical approach came after the beginning of the Marburg School, which based its theories on a logicist approach. Finally, this list gives no mention of the representatives of the philological approach to the doctrines of Kant. More recent research has moved away from comprehensive definitions of this nature and now distinguishes between the Marburg School and the Southwest German School of Neo-Kantianism, and a Kantian movement which extends beyond these limitations. The latter movement did not remain confined to Germany but could also be found in other European countries such as France (in the philosophy of Charles Renouvier) or Italy (with Carlo Cantoni), where it demonstrated its debt to Kant in a variety of ways.

Even if we accept that Neo-Kantianism is restricted to these two main schools, we must ask which writers should be ascribed to which school. To all intents and purposes, the Marburg School is limited to Cohen and Natorp and the Southwest German School to Windelband, Rickert and Emil Lask. However, should one include among the Neo-Kantians Cassirer in his middle and later periods along with Jonas Cohn and Bruno Bauch, though they are both writers whose important works were mostly published after the First World War? There is no consensus among Neo-Kantian scholars on this point. Some writers operate with a very narrow definition of Neo-Kantianism, while others apply the term in a much wider sense. The representatives of the first position argue that the First World War marked the end of the Neo-Kantian movement. Therefore they take all publications which appeared after the end of the First World War and illustrated a Neo-Kantian approach as in one way or another influenced by Neo-Kantian thought. The scholars who adopt a wider definition of Neo-Kantianism distinguish on the other hand between first- and second-generation Neo-Kantianism and assume in this context that Neo-Kantianism did not simply come to an end after the First World War. The first approach is supported by the fact that the First World War did also represent the end of an era in the philosophical world. After 1918 other philosophical trends gained ground and forced Neo-Kantianism out of the dominant position which it had enjoyed since the late nineteenth and early twentieth century, at least in Germany. Significant factors in this respect are the phenomenology of Edmund Husserl, who demonstrated the validity of transcendental thought in a different way from the Neo-Kantians; a logical positivism which challenged Neo-Kantianism on its own ground, in other words, epistemology and theory of science; attempts at a metaphysical interpretation of Kant which then developed; and cultural-critical philosophies of life which dominated public discussion. At the same time, a number of movements set themselves apart from the main trend. Nikolai Hartmann distanced himself from his Neo-Kantian origins in his attempt to re-establish an ontology, as did Martin Heidegger in his attempt to find a solution to the question of being. Richard Kroner, a pupil of Rickert, consciously made the transition to Neo-Hegelianism (see Hegelianism §6). However, if one considers not only the changed background conditions but also the internal theoretical

development within both schools, it will become clear that the second position also has good arguments to support it. For it cannot be maintained that the followers of the Neo-Kantian school failed to react to the changes in philosophical climate as a whole. This is shown not only in the later works of Natorp and Rickert: the philosophical thought of Cassirer, Bauch and Cohn must also be interpreted in the light of this approach. It is clear that these developments in the growth of the theory are just as much a part of the history of Neo-Kantianism as the phase of theoretical systems produced before the First World War. For this reason, this extended definition of Neo-Kantianism will be taken as the starting point for the following discussion.

Cohen, Natorp and Cassirer, on the one hand, and Windelband, Rickert, Lask, Bauch and Cohn, on the other, are merely the most important representatives of Neo-Kantianism. Apart from them, a large number of other philosophers can be included among the followers of both main trends, or at least associated with them. Earlier research assumed a close-knit school containing the various proponents of these doctrines; more recent approaches, however, have been more cautious in this respect. As far as Marburg Neo-Kantianism is concerned, it is possible to distinguish between a group of scholars closely associated with Cohen, including - apart from Natorp and Cassirer - Karl Vorländer and Albert Görland, and a wider circle of disciples among whom were Hartmann and Heinz Heimsoeth. None the less, Natorp’s attitude to a term like ‘The Marburg School’ was decidedly ambivalent. Although he accepted this definition for the purposes of academic debate, he wanted it to be understood merely as a term of collaborative association.

The relationships within the Southwest German School of Neo-Kantianism are similar. The labelling of Windelband and Rickert as the two leaders of the school is problematic inasmuch as Windelband had only sketched his systematic position, whereas Rickert must be credited with the true development of the Southwest German system. A circle of disciples also grew up around the latter in Heidelberg, among whom were Eugen Herrigel and Hermann Glockner. Furthermore, as in the case of Marburg Neo-Kantianism, a whole series of writers, such as Georg Lukács and Max Weber, found important inspiration in the philosophy of Southwest German Neo-Kantianism. There are also links with the work of Gottlob Frege. Finally, of considerable significance for the spread of Southwest German Neo-Kantianism was the fact that the periodical *Logos* provided a publication outlet from 1910.

For an understanding of Neo-Kantianism it is important to recall that the discussion of specifically Neo-Kantian ideas also continued outside the framework of the Marburg and Southwest German approaches. Riehl’s pupil Richard Hönigswald adopted, on the one hand, a critical approach to the Neo-Kantian logic of scientific method, because in his eyes the latter was based upon an undifferentiated concept of experience and pursued a different path from the Neo-Kantian approach to subjectivity by projecting a theory of the concrete subject within the context of his philosophical psychology. On the other hand, he continued to endorse the Neo-Kantian concept of normative validity, even though he attempted to develop this concept further in terms of a novel theory concerning the differentiation of spheres of validity.

Hönigswald, moreover, like his Jewish colleagues Cohn and Cassirer, was deprived of his professorial chair and had to emigrate after Hitler’s rise to power. The expulsion of the leading representatives of Neo-Kantianism from German universities was mainly responsible for bringing to a halt philosophical research into the problems posed within the Neo-Kantian doctrines, and the Second World War exacerbated this situation. In the post-war development of German philosophy only a few philosophers, such as Wolfgang Cramer, Hans Wagner and Rudolf Zocher, chose to continue in this theoretical tradition.

2 The origins of Neo-Kantianism

In earlier research the overriding tendency was to trace Neo-Kantianism back to certain founding figures. Thus, for example, the role of inaugurator of the Neo-Kantian movement was attributed to Liebmann. This view was based mainly on his early work *Kant und die Epigonen (Kant and the Epigones)* (1865), in which he had accused all branches of post-Kantian philosophy of working with a variously interpreted notion of the thing-in-itself. In the true Kantian view this was a false approach, and in each case he concluded his criticism of the schools of idealism (Fichte, Schelling, Hegel), realism (Herbart), empiricism (Fries) and transcendentalism (Schopenhauer) with a demand that scholars should return to Kant. In the same way, Helmholtz and Lange were credited with having founded the Neo-Kantian movement, since they both based their doctrines on the Kantian a priori, which they admittedly (mis)interpreted in terms of innate species-specific capacities (*Gattungsorganisation*), turning their

backs on the simplistic objectivism popular during the 1850s in the context of the debate concerning materialism. Finally, Eduard Zeller was also frequently associated with the origins of Neo-Kantianism. In his lecture *Über die Bedeutung und Aufgabe der Erkenntnistheorie (On the Meaning and Purpose of the Theory of Knowledge)* (1877) he was the first philosopher to win academic acceptance for epistemology or the theory of knowledge, which played a central role within the framework of the Neo-Kantian system.

Modern research sees the birth of Neo-Kantianism as being more clearly determined by certain developments which were characteristic of the philosophy of the post-Hegelian period. In the first instance a central role was played by a departure from pure thought and from the systematic philosophy (*Systemphilosophie*) of the idealist tradition, as exemplified in the works of Friedrich Schleiermacher or Adolf Trendelenburg. The latter are important for the development of Neo-Kantianism, first because the concept of theory of science is articulated as early as the second edition of Trendelenburg's *Logische Untersuchungen (Logical Investigations)* (1840), and second because with his *Dialektik (Lectures on Dialectics)* (1839) Schleiermacher can be regarded as the originator of the theory of knowledge. Furthermore, the thesis that German philosophy from Fichte to Hegel pursued a mistaken course which could only be corrected by establishing a direct link with the Kantian tradition was not first formulated by Liebmann, but can be found in the earlier doctrines of Christian Weisse and Friedrich Beneke. A third problem area in post-Hegelian philosophy is the relationship between philosophy and the individual sciences. The attempts by Helmholtz and Lange to provide Kant's theory of knowledge with a sensory-physiological basis against the background of reinstating cooperation between philosophy and science was just one way of overcoming the problem. There was also the opposite tendency, which aimed at rejecting the inclusion of individual and special scientific knowledge in an attempt to maintain the autonomy of philosophy. This attitude can be found in the works of Kuno Fischer, Liebmann's teacher, who propagated a neo-idealistic interpretation of Kant in the manner of Fichte and played an important role in the foundation of Southwest German Neo-Kantianism. Fourth, in addition to the critical approach to the idealist point of view there was also the tendency to insist on the necessity for a critique of knowledge in the hope that it would finally provide a means of settling disputes about fundamental questions of worldview or '*Weltanschauung*'. Such a tendency is to be found not only in the works of Zeller and Lange, but also in the those of J.B. Meyer, a hitherto largely unknown philosopher in the field of Neo-Kantian research. In his essay *Über Sinn und Wert des Kritizismus (On the Meaning of Criticism)* (1857), Meyer emphasizes that the relationship between body and soul belongs, like all metaphysical questions, to those problems which transcend the horizon of our possible knowledge, and that the significance of '*kritizistischen Kant*' ('Kant the Criticist') lay in the way in which he had demonstrated this fact. Similarly, in his programmatic deliberations concerning the theory of knowledge, Zeller repudiated the exploitation of specific, scientific knowledge for the purpose of establishing a worldview, just as Lange was eventually concerned to refuse universal claims based on such a worldview, of whatever kind they were.

Therefore, even if the extensive pre-history of the Kantian movement is taken into account, there can be no doubt that its real breakthrough occurred during the 1870s and 1880s. In 1871 Cohen published his work *Kants Theorie der Erfahrung (Kant's Theory of Experience)*, which set out to establish Kant's doctrine of the a priori. Cohen's central thesis is that only those things can be seen to be objective which are produced a priori by subjectivity. Thus he understands Kant's *Critique of Pure Reason* to be a theory of experience in which Kant's aim was to reveal those a priori moments which make scientific experience possible. In the same vein he maintains that in ethics and aesthetics Kant was also concerned with the revelation of those a priori elements which in the first instance generate the sphere of morality and art. This concept of generation thus provided the basis for a new form of systematic appropriation of Kant which was significantly different from the subjective-idealistic Kant interpretation of Kuno Fischer or Lange's naturalistic interpretation. Cohen considers both interpretations inadequate because neither manages to provide a stable basis for the sphere of the ideal. In the case of Lange in particular, Cohen criticized the fact that, although he attempted in his *Geschichte des Materialismus* (*History of Materialism*) to save ethics in opposition to the mechanistic-deterministic worldview, he nevertheless failed to justify the idealist point of view in a scientific manner. Cohen believed that he had found the means of achieving this in his concept of generation. In his view, modern natural science is the best evidence for the fact that reality as objective has been created by us, since we are only capable of knowing that which we have put into things.

This realist attitude was given a different emphasis in the Kantian research of Friedrich Paulsen and Alois Riehl. Paulsen's *Versuch einer Entwicklungsgeschichte der Kantischen Erkenntnistheorie (Attempt at a History of the*

Development of Kant's Theory of Knowledge) (1875) aims, as the title indicates, at a historical and systematic reconstruction and not, like Cohen's work, at a new and improved presentation of Kant's doctrine of the a priori. Thus Paulsen's goal was not a productive assimilation of Kantian thought against a background of the specifically nineteenth-century conflict between idealism and materialism, but an attempt to understand the philosophy of Kant from its historical context, against a background of the contrast between rationalism and empiricism during the eighteenth century. Riehl's three-volume work *Der philosophische Kritizismus und seine Bedeutung für die positive Wissenschaft (Philosophical Criticism and its Significance for Positive Science)* (1876-7) aimed to place Kant within a tradition of critical thought which began with John Locke. In his view, Kant's critical position tends towards a 'Phenomenalism of Intuitions' which is at the same time a 'Realism of Things'. Riehl regarded the epistemological scepticism represented by David Hume as the precursor of such a position, since it can be seen as a means of securing a positive content for thought. Thus, in his presentation of Kant, Riehl undertakes an attempt to bridge the gap between Kantianism and positivism. This approach can also be seen in his history of philosophy, which assumes as its basic concept a parallelism between scientific and philosophical history. Benno Erdmann completes the spectrum of Kant interpretations during this period. His works use a purely historical approach to reconstruction and rejected all attempts to update Kantian thought.

Finally, during this period Windelband founded the Kantian approach of the Southwest German School, and this represents a second type of systematic Kantian assimilation beside that of Cohen. Mindful of the overwhelming significance of Kant for all subsequent philosophical endeavours, Windelband investigates the nature of the critical method put forward by Kant. According to him, what is essential in this respect is the contrast with every type of genetic method. Unlike the genetic method, it is not a question of explaining how certain norms in the realm of logic, ethics and aesthetics were established. Instead, Windelband's critical method questions the correctness of the norms that have come into existence in these areas. In other words, its aim is the clarification of the question of justification. Windelband is convinced that Kant and his three Critiques made a significant contribution to the critical science of norms. Thus philosophy, if it wishes to counter the relativism which inevitably results from a consistent application of the genetic method, is obliged to return to Kant. At the same time, however, Windelband emphasizes that such a 'return' to Kant cannot involve simply a revival of the historically conditioned form of his philosophy, but remains a task for all contemporary engagement with the philosopher from Königsberg: to understand Kant means to go beyond him.

3 The Marburg School: Cohen

Cohen's three-part philosophical system serves as a paradigm for the development of the Marburg School system. It consists of the *Logik der reinen Erkenntnis (Logic of Pure Knowledge)* (1902), the *Ethik des reinen Willens (Ethics of Pure Will)* (1904) and the *Ästhetik des reinen Gefühls (Aesthetics of Pure Feeling)* (1912) and attempts to demonstrate in a consistent manner the point of view of pure generation, that is, creation which does not make appeal to any particular preconceptions.

This is particularly clear in the *Logik*, which forms the first part of his system. Cohen differs markedly from Kant in refusing to preface his logic with a doctrine of sensibility. Indeed, his attempt at logic can be characterized by the claim, 'We take thought as our starting point. Thought cannot have any origin outside itself.' According to the 'Logic of Origin' as conceived by Cohen, therefore, thought is grounded solely in itself and is thus a hypothesis in the original Greek sense. The denial of the Kantian theorem of the dual origin of knowledge (intuition and thought) results in the abandonment not only of the transcendental aesthetic as a doctrine of the principles of sensibility, but also of the transcendental deduction (see Kant, I. §§5-6). What remains of Kant's approach is no more than the programme of a transcendental method for the foundation of science, which is obliged to take the existence of science as its starting point. In his doctrine of judgment Cohen differentiates between judgments involving the laws of thought (Origin, Identity, Contradiction), of mathematics (Reality, Plurality, Totality), of mathematical natural science (Substance, Law and Concept) and of method (Possibility, Reality, Necessity). In his doctrine of categories, he starts from the necessary incompleteness and essential extendibility of the system of categories, since he is convinced that new scientific problems will also make new categories necessary.

Cohen sees ethics, which he seeks to base on the fact that a science of right exists, as a doctrine of humanity. The ideal collective subject which ethics deals with is humankind as a whole. It is important that this term is not taken to mean the human species in a biological sense, but rather an ethical expression encompassing humans as moral

individuals. Such an interpretation of the concepts of universality also permits a new definition of the concept for the individual as the idea of the human being in every person. However, the medium which allows us to establish a correlation between these two concepts is the state. For within the state the common will can be realized through the harmony achieved between the actions of the individual and the actions of everyone else. The model of the state which Cohen takes as the basis for his discussion has a cooperative constitution. The essential characteristic of cooperatives is that the common will is legally capable of action. This distinguishes it from communities which only recognize the ruling will of the individual. Against this background Cohen developed the concept of an ethical socialism which aims at overcoming the existing 'state dominated by the estates and the ruling classes' (Cohen 1904: 582), without deteriorating, like the materialist historical perspective, into economic determinism.

In *Ästhetik*, Cohen modifies his system, no longer linking aesthetics back to the fact of an existing science of art, but rather attempting to deduce the a priori law-like character of pure feeling directly from works of art. In doing so he specifically defines pure feeling itself as the feeling of love towards human nature. This is the presupposition not only of the creation of art but also of its reception. With regard to the objectivity of aesthetic experience, Cohen emphasizes that although there are no further objects as such apart from nature and morality, it is possible for the aesthetic orientation of consciousness to shape and transform these spheres of objectivity and thus 'bring forth beauty as a new object' (Cohen 1899: 100).

It is a feature of Cohen's development that during his later years he devoted much of his attention to questions concerning the philosophy of religion. Although during his Marburg years he had argued for the incorporation of religion into ethics, he later recognized that religion also makes a special contribution to the problem of culture. In his essay *Der Begriff der Religion im System der Philosophie (The Concept of Religion in the Philosophical System)* (1915) he attempts to determine this contribution on the basis of the three systematic disciplines of logic, ethics and aesthetics. If logic is concerned with the being of nature, then religion is concerned with the unique being of God, and if the God of ethics stands for the endless continuation of the moral tasks of the human race, then the God of religion must ensure that the moral striving of the individual does not ultimately come to nothing. If aesthetic love is related to humankind as a type, then religious love embraces the human being as individual. Irrespective of this special contribution to culture in its entirety which is accomplished by religion, the latter is not regarded as being independent of culture, but rather as possessing a special character within the framework of culture as a whole. The quest for a non-reductionist understanding of religion is also one of the dominant themes in Cohen's posthumous work *Religion der Vernunft aus den Quellen des Judentums* (*Religion of Reason out of the Sources of Judaism*) (1919). It cannot be said, however, that Cohen achieved a position of ultimate clarity with regard to religion. Evidence for this can be found, for example, in the interpretative dispute concerning an idealistic or dialogical interpretation of the term 'correlation', which plays a central role in Cohen's later philosophy of religion. Is Cohen envisaging simply a relationship between two concepts which are mutually required (in much the same way as the concept of the sinful individual necessarily calls for the concept of God, who forgives sins), or does he have in mind a real contrast of two partners as is suggested by the definition of prayer as 'dialogue with God'? Furthermore, there is a tension in Cohen's posthumous work between the ethically orientated interpretation of the religious statements which he attempts to provide and the simultaneous reversion to statements reflecting a metaphysical doctrine of divinity. On the one hand he states that the doctrine of divine attributes should not be interpreted in terms of ontological statements about God, but rather in terms of statements which make clear how human behaviour should be governed. On the other hand, Cohen characterizes God as the one and only unique Being as opposed to all Becoming, as the Infinite as opposed to all spatial limitations, as the source of all movement, and as the Eternal in contrast to all Change. Cohen was unable to carry out his plans for a fourth work to complete his 'System of Philosophy', a psychology conceived on a science of the unity of cultural consciousness.

4 The Marburg School: Natorp

Natorp was able to present his thoughts in a clearer and more comprehensible form than Cohen, and hence he was largely responsible for the spread of the doctrines of the Marburg School of Neo-Kantianism. Unlike Cohen, he is not famous for a comprehensive range of studies on Kant. None the less, he also made a name for himself by virtue of his works on the history of philosophy. Of greatest relevance in this context is his controversial and subsequently modified interpretation of Plato, in which he understood the Platonic 'idea' as a hypothesis from the perspective of the modern concept of the laws of nature.

In his systematic thought, Natorp at first operated systematically within the terms laid down by Cohen. None the less, even in *Kant und die Marburger Schule (Kant and the Marburg School)* (1912a) he emphasized that philosophy did not merely refer back to scientific facts (natural sciences and humanities) but was also related fundamentally to morality, art and religion. Its subject is therefore 'the entire creative work of culture', which is expressed in the 'theoretical-scientific description of phenomenal appearance' as well as in the 'practical formation of social orders', 'in artistic creativity' and 'in the inmost forms… of religious life'.

In his *Allgemeine Psychologie nach kritischer Methode (General Psychology According to the Critical Method)* (1912b) Natorp goes further than Cohen in setting a subjective foundation for knowledge alongside the objective one. He is guided by the following consideration: in every phenomenal appearance we can differentiate between a relationship to the object and a relationship to our consciousness. The former is the subject matter of the theory of knowledge which establishes the laws of cognitive objectification; the latter is the subject matter of psychology, which proceeds back to the objective phenomenon of knowledge from the already accomplished objectifications, since the 'immediacy of consciousness' does not permit any immediate psychological comprehension.

In his doctrine of knowledge, Natorp attempts to interpret Cohen's concept of origin in the context of Kant's synthetic unity. According to this theory, that which characterizes the thinking process is the basic structure of the relation between the one and the many. In this fundamental relation there are two tendencies: a differentiating tendency, which aims at quantity, and a unifying tendency, which aims at quality. Quantity results when the manifold as such is emphasized in the fundamental relation; quality results, on the other hand, when thought is directed towards the unity of the manifold. The relation arises with the establishment of interdependency between the unities, which are determined in a quantitative-qualitative manner, whereby the modality determines the contribution of each level of the 'thought-functions' constituting the object to the knowledge of that object. In contrast to Cohen, Natorp does not base this system of fundamental logical functions on a system of sciences, although he asserts that the systematic logical structure it contains will prove its validity in the fact that the basic scientific facts can be connected with the logical ones. In fact, Natorp starts out from a reciprocal interrelationship between science and the logic of knowing. While the latter tends towards the centre, in other words towards the basic law of synthetic unity, the former tends towards the periphery, that is, the manifold diversity of the particular instances of knowledge. The difference between this and Cohen's position also becomes clear since, for Natorp, the reason belonging to science proves itself in the foundation of the objective sciences from the perspective of synthetic unity, while for Cohen reason proves itself in demonstrating the origin of all knowledge out of pure thought.

Natorp's social educational theories, which he developed after becoming professor of philosophy and education in Marburg, were very influential. The subtitle of *Sozialpädagogik* (1899), his major work on the subject, is *Theorie der Willenserziehung auf der Grundlage der Gemeinschaft (Theory of the Formation of Will on the Basis of the Community)*: it makes clear that Natorp sees the formation of will as the most important aspect of education, particularly inasmuch as the former is determined by life within the community and has an effect on the community. In the development of will Natorp distinguishes three stages of activity: instinct, will and the rational will. Whereas instinct is wholly governed by the sensory object, will on the other hand provides for the possibility of choice in this respect. True moral orientation, however, occurs only in the case of the rational will. Although a restructuring of society in accordance with the principles of ethical socialism implies changes within its economic and political institutions, Natorp none the less sees the real motor for change in the sphere of social education, since only education can bring about a change of consciousness.

At the heart of Natorp's later philosophy stands his attempt at a new definition of the doctrine of categories, whose goal is a unified foundation of every kind of positing of an object ('*Gegenstandssetzung*'). Natorp starts concretely with a closed system of basic categories - modality, relation and individuation - which are to serve as the basis for the construction of an open system of categories. By adopting this distinction between categories and basic categories he attempts to do justice both to Kant's assumption of a comprehensible number of basic concepts of understanding and to the fact that the process of investigation means that a number of categories cannot be determined in advance. Natorp's system of basic categories is also built up in a different way from Kant's, inasmuch as he begins here with the categories of modality (Possibility, Necessity and Reality) and places the categories of Quantity, Quality and the intuitive forms of Space and Time in the third group of categories. The second group of categories (Substance, Causation and Interaction) corresponds on the other hand with the original

Kantian architectonic which Natorp had mainly used as a starting point in his early attempt at a system of categories, *Die logischen Grundlagen der exakten Wissenschaften (The Logical Foundations of the Concrete Sciences)* (1910). The order of categories within the individual groups occurs in each case according to the scheme: universality, particularization, individuality. This structure applies not only to the theoretical sphere but also to the practical and what Natorp calls the 'poetic' (that is, productive) spheres. Natorp's later philosophy is characterized by the insight that the process of categorization assumes the acceptance of the as yet totally undetermined original fact of 'it is' and ends with the determination of the individual as the concrete objective existent. This ontological modification of Natorp's late philosophy need not necessarily be interpreted - as it frequently is - as a break with the Marburg theoretical outlook; it can also be seen in terms of a radicalization of the original Marburg theory.

5 The Marburg School: Cassirer

Like Cohen and Natorp, Cassirer, who was Cohen's principal disciple, made a name for himself initially in the history of philosophy. After his doctoral thesis under Cohen on *Leibniz' System in seinen wissenschaftlichen Grundlagen (Leibniz's System in its Scientific Foundations)* (1902), he enjoyed immediate success with the first two volumes of his four-volume *Geschichte des Erkenntnisproblems (History of the Problem of Knowledge)* (1906-20). In it he demonstrated in particular the emergence of a new definition of knowledge during the Renaissance which reached its climax in the philosophy of Kant. Thus Cassirer attempted to make plausible Cohen's claim concerning the harmony of his basic ideas with the historical development of scientific and philosophical thought.

Cassirer also turned his attention to the theory of the exact sciences, publishing his *Substanzbegriff und Funktionsbegriff (The Concept of Substance and the Concept of Function)* (1910) at the same time as Natorp's investigation *Die Logischen Grundlagen der exakten Wissenschaften*. Here, Cassirer makes it plain that any attempt to represent the whole of knowledge ends in certain ultimate form-concepts which express the various possible ways in which contents can be related to one another. In this respect, what we call the object of knowledge is dissolved in a network of relations. What applies to the concept of the object in general is also relevant for the concept of the objects with which the individual sciences operate. They too do not represent an absolute beyond the logical forms of knowledge, but that which is to be expressed through them is nothing but a functional relation within these forms. In concrete terms, Cassirer sees in the serial ordering which manifests itself in the relations of space and time, of size and number and of dynamic interaction and interrelationship of events, the moment in the process of knowledge which determines its real empirical content. The logical determination of the object of knowledge thus leads ultimately to an original relation which can be understood in various ways: as a relation of form and matter, of the universal and particular, and of being or normative validity. What is important is that this relation between reciprocally determining elements remains irreducible. Taken as a whole, what distinguishes Cassirer's concept of knowledge is the insight that concepts are intellectual constructs which cannot be understood as an imitation of the objective given but as 'symbols for relations and functional connections within the real'.

A third major theme in Cassirer's philosophical work is the philosophy of culture. His *Philosophie der symbolischen Formen* (*The Philosophy of Symbolic Forms*) (1923-9) is relevant in this respect. Cassirer understands by the term 'knowledge' not only that mathematical-scientific form of knowledge which remained paradigmatic for the Marburg School, but rather 'any intellectual activity in which we construct a world for ourselves with a characteristic form, order and particular character' (1956: 208). The concept of symbolic form implies on the one hand a formal function, for it is essential that 'an intellectual meaning is attached to a concrete sensory sign and internally allocated to this sign' (1956: 175). On the other hand, Cassirer also marks out in this manner as characteristic areas of such formation, such concrete areas of culture as language, myth, art and science. Although the symbolic activity linked to signs is that activity which runs through all productive achievements of consciousness as a unifying theme, further differentiations with regard to the basic functions of the symbolic form will result. Here Cassirer distinguishes between the function of expression, the function of presentation and the function of meaning. The almost physical function of expression is particularly relevant in the case of myth, the distancing presentation function is the preserve of language, and the function of meaning is found in particular in the world of science. The systematic implications of Cassirer's theory of symbolic forms are further developed in his final works, *An Essay on Man* (1944) and *The Myth of the State* (1946), produced after his emigration to

America.

Like the later works of Natorp, the works of Cassirer's middle and later periods raise the question of whether they should really belong to the Marburg School of Neo-Kantianism. This is a subject of lively debate even today. It is certainly the case that Cassirer's middle and late works transformed Cohen's approach, which revolved around the problem of knowledge and science. Similarly, it cannot be maintained that Cassirer in his later development actually abandoned his Neo-Kantian beginnings, for even in his conception of a philosophy of symbolic forms he retains the typical emphasis upon the generative-theoretical approach characteristic of the Marburg School.

6 The Southwest German School: Windelband

Although the real development of the Southwest German system was the work of Rickert, and Windelband merely sketched his systematic approach in a collection of essays with the descriptive title *Präludien (Preludes)* (1884) as well as in his *Einleitung in die Philosophie (Introduction to Philosophy)* (1914), the importance of Windelband as an initiator is not to be underestimated. Windelband understands philosophy as a whole to be concerned with values. He rejects a purely objectivist conception of value and attempts instead to solve the problem of the objectivity of values in a transcendental manner by having recourse to an archetypal consciousness (*Normalbewußtsein*) which is able to make value judgments, and for which the Kantian 'Pure Consciousness' (*Bewußtsein überhaupt*) serves as a foil. He traces the traditional division of philosophy into the disciplines of logic, ethics and aesthetics back to the way in which these disciplines are orientated to the values of Truth, Goodness and Beauty. Like Cohen, Windelband also attributes a special role to religion. According to him, the sphere of human value judgments is exhausted by these logical, ethical and aesthetic values. Religion, whose proper concern is with the sacred, does not represent a separate cultural domain. Correspondingly, the sacred does not represent a specific class of values with universal validity alongside truth, goodness and beauty; it embodies all these values inasmuch as they are all related to a supersensible reality.

Like the philosophers of the Marburg School, Windelband also sees the significance of Kant in his re-establishment of the relationship between science and philosophy. At the same time he criticizes Kant - and this distinguishes him from the Marburg School - for his bias in favour of the mathematically orientated natural sciences. In the face of the triumphant emergence of the historical sciences throughout the nineteenth century he sees this as an anachronism. Therefore his efforts are directed towards bringing out their special structure in contrast to the natural sciences. In his speech as Rector of Strasbourg University he defined his approach: the natural sciences operate with universal, apodeictic judgments. They aim at universal statements, which explains their interest in a reality which always remains constant. Their approach is nomothetic, because their aim is to reveal relationships between laws. The price paid for this method is a tendency to abstraction. The historical sciences mitigate this by means of singular, assertive judgments. They are directed towards the particular, which explains their interest in what reveals itself as a unique, intrinsically determined content of reality. Their approach is idiographic, because what they aim to make visible are concrete configurations. For this reason their procedure depends upon intuitive self-presentation.

The main emphasis of Windelband's publications lay in the sphere of the history of philosophy. He was the author of, among other works, a two-volume *Geschichte der neueren Philosophie (History of Modern Philosophy)* (1878-80) and a *Lehrbuch der Geschichte der Philosophie (Textbook of the History of Philosophy)* (1892), which was continued by Heimsoeth and is still reprinted regularly today. In Windelband's view, the history of philosophy has a threefold task. It should:

1 Establish precisely what can be ascertained concerning the living conditions, the intellectual development and the doctrines of the individual philosophers from the existing sources;
2 Reconstruct from this information the genetic process in such a way that in the case of each philosopher it is possible to understand the relationship of his teachings to those of his predecessor, to the general ideas current at the time, to his individual character, culture and education;
3 To judge from an observation of the whole what value can be ascribed to the teachings which have been discovered and whose origins have been explained in the light of the history of philosophy in its totality.
(Windelband 1902: 13)

7 The Southwest German School: Rickert and Lask

As far as Rickert's philosophical development is concerned, his main interest originally lay in questions related to the theory of knowledge and methodology. In his habilitation essay, *Der Gegenstand der Erkenntnis (The Object of Knowledge)* (1921), he criticizes the customary definition of knowledge, which is based on the opposition between real being and a consciousness which groups such being as it is reflected in our representations. He puts forward instead a new concept of knowledge. His starting point for this is a reflection on the primary aim of knowledge, namely the striving after truth. Truth, for Rickert as for Aristotle, lies not in representation but in judgment. However, even if we have to relinquish the concept of a world that is independent of our representations, if judgment as the locus of truth is to be ascribed a central role in the process of knowledge, this approach will allow us to demonstrate a world that is independent of the judging subject, because every judgment is based on the recognition of an 'ought'. Although this 'ought' represents a transcendental order as far as knowledge is concerned, it is not a matter of the order of a transcendent reality, for what we are able to discover is no more than the correct order belonging to the content of awareness, that is to say 'the relationships of representations to each other, which ought to exist and which must therefore be affirmed'.

In his scientific methodology Rickert does not distinguish as Windelband does between nomothetic and idiographic methods, but rather between generalizing and individualizing concept formation. Even if all terms are general inasmuch as they bind general elements together to form conceptual unities, these unities are of a universal nature in the sphere of the sciences, but individual in the sphere of history. For Rickert, the formation of such individual conceptual unities can be explained by their relationship to universal values. This permits a selection to be made from the immeasurable variety of individual given facts within the world of experience. However, the historian's value-orientated procedure should not be understood as representing a subjective valuation of historical events. It is not the historian's job to pronounce on the worth or lack of worth of a particular event such as the French Revolution, but rather to enquire as to the significance of such an event for the history of Europe as a whole. Admittedly, that such historical concept-formation is not arbitrary naturally presupposes the existence of certain generally recognized cultural values. This conclusion, which Rickert recognized was hardly self-evident in his own period, was none the less essential for his argument. His methodological reflections thus culminate in an objective doctrine of value which claims that values enjoy the ontological mode of 'validity' as distinct from the ontological mode of 'existence'.

In his middle period, Rickert drafted a *System der Philosophie (Philosophical System)*, which, however, remained a fragment. In it he conceives of philosophy as articulating a worldview intended to illuminate the meaning of life for us. This, however, involves actually bringing to consciousness the values which make life meaningful. Rickert therefore projects a system of values which he conceives of as an open system. His goal is not 'a complete grasp and enumeration of all valid values', but merely 'a complete formal classification of various types of value' (Oakes 1990: 134). Specifically, this means that he distinguishes between a sphere of contemplative, non-social and substantive values on the one side and a sphere of active, social and personal values on the other.

The attempt to engage with ontological questions is characteristic of Rickert's later phase of thought, yet this need not be interpreted as a break with his original theoretical approach to the problem of transcendental constitution. Indeed, it could be seen as an attempt to pursue this line of thought more deeply. Relevant to Rickert's later philosophy are his *Thesen zum System der Philosophie (Theses on the Philosophical System)* (1932). Here he made it clear that the task of the individual sciences is to comprehend the world of the senses 'either in a universalizing manner (as nature, in so far as it is governed by a system of laws) or in an individualizing manner (as the history of "culture")' (Rickert 1932: 99). On the other hand, he sees a system of values as necessary in order to articulate the intelligible world, because 'meaning and significance can only be properly and rigorously explained by recourse to concepts of value' (1932: 99). Admittedly this 'dualist idea of the world' (*Weltdualismus*) immediately prompts the question of whether there might be an encompassing unity which could bind the two worlds of being and value together. Rickert recognizes this unity of reality and value primarily in the sphere of the non-objectifiable being of human subjects and distinguishes therefore, in addition to the world of experience of sensible and intelligible subjects, a third 'pro-physical' type of worldly being. Of course, the question of the unity of the world is not adequately resolved by this concept. It is impossible to avoid the question concerning a 'transcendent' unity of the world in which value and reality are even more closely connected than in the pro-physical sphere. While in the latter value and reality are only connected through the free acts of the subjects, and otherwise remain separate from one another, here the problem concerns the possibility of an original unity of reality and value in the

supersensuous ground of the world. Rickert leaves us in no doubt that this metaphysical transcendence can only be grasped with the aid of symbolic thought, which interprets material taken from this world as a simile or an image for the beyond.

Lask wrote his doctoral thesis under Rickert on *Fichtes Idealismus und die Geschichte (Fichte's Idealism and History)* (1923-4) and then completed his postdoctoral teaching qualification under Windelband with a study of the philosophy of law. This is an important work for the philosophical system of the Southwest German School because in it Lask understood law as an independent sphere of validity and developed at the same time a methodology of the science of law. However, questions of theoretical philosophy stand at the centre of Lask's philosophical works. For him, everything that can be conceived is divided between the categories of existence and validity. Correspondingly it is necessary to distinguish between knowledge of being and knowledge of validity. Form bears validity, while material being, on the other hand, is utterly foreign to form and therefore does not enjoy the status of validity. The sphere of the sensuously given is then revealed as a realm which lacks this status. For what is sensuously given cannot be deduced by means of logical processes, but can only be acknowledged in its factuality.

It is a characteristic of Lask's understanding of the doctrine of form and material being that a particular form of validity no longer bears a direct relationship to the material, but rather corresponds to a specific material to which it clings (as it were). For Lask, therefore, material being functions as an element which determines meaning. By virtue of this theory of the differentiation of meaning, an independence is attributed to the material of knowledge which stands in direct opposition to the classical Neo-Kantian concept of knowledge, according to which the entirety of the given is determined by the forms of cognition. Correspondingly, at the centre of Lask's conception of the subject we find, not the constitutive activity of the subject with respect to objects, but rather the subject given over to the object. For the relationship of form to the material, which the subject is required to comprehend, precedes all subjectivity. Lask's objectivistic tendency can also be clearly seen in the fact that for him validity is in the first instance free of a relationship to the subject. This relationship only arises when validity is experienced by the subject as an 'ought'. Inasmuch as the subject experiences validity in the shape of a demand, then the conflict of values also comes into play. This occurs because the activity demanded of the subject as regards the recognition of the 'ought' can either succeed or fail. If it succeeds, then the truth attained through this activity appears as a positive value; if it fails, as a negative value. The actual relationship to the validity-content occurs as the judging act which connects subject and predicate and simultaneously contains the affirmation and denial of this connection. It is now important for Lask that the truth lies outside the judgment in the configuration of matter and form. The judgment can correspond to this structure or contradict it. Therefore we can only speak of the judgment's appropriateness to truth or untruth, but not of its truth or falsity as such.

During the final phase of his philosophical development Lask abandons such pure objectivism. Knowledge as a subjective act is once again seen as an independent problem-area of philosophy, although Lask was not prepared to take back any part of his claim concerning the significance of the objectively true with regard to the process of knowing.

8 The Southwest German School: Cohn and Bauch

Cohn's turn towards dialectics is important for his philosophical development. Cohn defines as 'dialectical' all knowledge which 'uses the appearance, the resolution and the reappearance of contradiction as a means of acquiring knowledge'. In contrast to Hegel, Cohn supports a 'critical' dialectic in that he does not presuppose the dogma of the complete transparency of the world to reason. His dialectic is only a 'thinking towards the absolute', but not the kind of thought which could expound the absolute speculatively.

This conception of thought is also evident in Cohn's treatment of the problem of reality. First he distinguishes between three 'forms of reality': (1) the immediate experience of reality, which in his view is the beginning and the point of departure for all other conceptions of reality and thus also the basis for the process of objectivization, (2) the so-called existing factual reality, which represents a first stage in the process of objectivization, because here 'the non-sensational aspects of immediate experience' are excluded, and finally (3) physical reality, which as the final stage of the process of objectivization embodies a 'reality of maximum objectivity' (Flach and Holzhey 1980: 59). But what about the unity of reality in the face of these diverging conceptions of reality? According to Cohn, although this unity is the ultimate aim of human knowledge, it remains unrealizable. It therefore remains an

idea in the Kantian sense: it retains the character of a task which is as essential to knowledge as it is unachievable. Thus Cohn's posthumously published work, which is focused upon the implications of the theory of knowledge for the concrete sciences, is entitled *Wirklichkeit als Aufgabe (Reality as Task)* (1955).

Cohn is not merely interested in questions of theoretical philosophy, as is shown by his major investigation, *Voraussetzungen und Ziele des Erkennens (Presuppositions and Goals of Knowledge)* (1908). Because in general he understands philosophy as a discipline concerned with questions of value, in addition to the realm of logical value he also investigates the realm of aesthetic values; his *Allgemeine Ästhetik (General Aesthetics)* (1901) presents the 'only major work on aesthetics in the Southwest German School' (Holzhey 1992: 45). Questions concerning the systematization of values are the subject of a major work entitled *Wertwissenschaft (Science of Values)* (1932), in which he rejects the thought of a closed value system and interprets the relationship of values to each other in terms of a '*teleological completion*' (*teleologische Ergänzung*).

In connection with the problem of values, Cohn also turned his attention to the philosophy of culture, and on the eve of the First World War he published a paper entitled *Der Sinn der gegenwärtigen Kultur (The Meaning of Contemporary Culture)* (1914). This takes as a starting point the theory that while culture does indeed arise from aspects of human behaviour based on values, it is not created by man's conscious purposive behaviour, because the individual always already finds himself in a specific configuration of cultural life, 'from which he draws substantial content and value' (Cohn 1923: 14), and which he can, of course, seek to transform. With regard to the development of culture in general this produces the following schema: all culture is in the first instance 'bound', that is to say the individual lives totally within the forms which he takes over as being both necessary and self-evident. His consciousness is directed towards interpreting, maintaining and actualizing these forms. In this process his intelligence is strengthened and comes to regard the status quo critically. This means that human beings are presented with new possibilities for shaping their environment, albeit at the price of spiritual homelessness. Against this background, Cohn interprets the present as a time in which the liberated spirit strives to find a new form of fulfilment in life after recognizing this homelessness.

For Cohn, the practical discipline corresponding to the philosophy of culture is pedagogy or theory of education. For this reason he also attempted to establish a philosophical basis for this discipline and saw his treatise *Der Geist der Erziehung (The Spirit of Education)* (1919) as supplementing his study on the meaning of contemporary culture. In his view, the central task of education is the conscious 'development and perpetuation of culture' (*Fortbildung der Kultur*). Education must both encourage the autonomy of the pupil and at the same time ensure their membership of a historical cultural community. This dialectic of liberation and integration is essential to the process of education.

Cohn's importance lies in his development of the initial theories of the Southwest German Neo-Kantianism in the direction of a critical dialectics. Bauch, on the other hand, can be considered as the synthesizer of Southwest German Neo-Kantianism. His major monograph on Kant set out to understand the individual problems of Kant interpretation from the standpoint of the 'Kantian spirit in its entirety'. His systematic works all have a similar aim, suggested by the title of his main philosophical work *Wahrheit, Wert und Wirklichkeit (Truth, Value and Reality)* (1923): he tries to combine various approaches to problems which are generally tackled separately.

On the problem of knowledge, Bauch claims that we must assume a correlation between knowledge of the object and the object of knowledge. Knowledge and object do not simply coincide, but neither can they be completely separated from each other. However, this presupposes that they are both subject to the same conditions. Bauch finds these shared conditions in truth. He analyses individually the structural forms of truth in judgment, category and concept. In the case of judgment he distinguishes between the actual judgment as a 'connection of representations' (*Beziehung der Vorstellungen*), and the logical judgment as an 'objective relationship of validity' (*objektive Geltungsrelation*). The factual judgment can succeed or fail, in other words it is either valid or invalid. It is only valid when it connects up with the validity relation given in the logical judgment. For this, however, it is necessary that it subordinate itself to the latter's validity or, in Bauch's words, that it 'recognizes it'. On the category, Bauch agrees with Kant that 'the category determines the judgment… because the judgment is inseparable in its logical significance from the function of unity which the category represents'. Since the category only makes possible a one-sided determination of the object, if we are to arrive at a complete determination then we must appeal to the concept which, in contrast to the category, refers to the object as concretely given in

intuitive isolation. However, the concept may not be considered in isolation, but must be seen as part of a network of concepts. Bauch names this network ‘idea’.

From the concept of ‘idea’ Bauch makes the transition to the concept of the ‘world’, which for him is nothing more than the appearance of the idea. In this context the concept of system also becomes comprehensible together with the concept of the world. If the idea can be thought of as a system of mutually conditioning concepts, then the world can be seen as a ‘system or relation of conditional things’. In the final analysis, the world cannot be conceived of without an ‘I’, a subject which conceives of the world as having arisen from the idea.

In his practical philosophy Bauch assumes that we encounter the ethical principle *qua* value principle in three forms: (1) as a demand on the will which addresses itself to the will of every rational creature, (2) as a demand for action in the light of the fact that there are certain duties which are only meaningful if specific possibilities of acting exist (consider the Hippocratic oath taken by doctors) and (3) as an essential demand, by which he means the obligation of the individual person to realize their specific essential being. Depending on whether the values as tasks and demands apply to the community as a whole or to individual members of the community, Bauch distinguishes between communal and personal values, which can no more be thought of isolated from one another than can the terms ‘personality’ or ‘community’. On the totality of values, Bauch leaves us in no doubt that values form an ‘interconnected sphere’ and do not exist alone in isolation from each other. Of course this interconnection does not entail enclosure and rigidity. Just as in the development process of knowledge new categories can be created, so the discovery of new values cannot be excluded.

9 The significance of Neo-Kantianism

Neo-Kantianism has often been dismissed as sterile academic philosophy, but this fails to do justice to its real significance. Despite the one-sidedness of its approach, the Neo-Kantian attitude to Kant profoundly influenced the twentieth-century study of Kant’s works. From the perspective of systematic philosophy, the significance of Neo-Kantianism lies first in its development of the concept of ‘validity’ and second in its projected philosophy of culture.

On the first point, the essence of Neo-Kantian thought lies in the strict division it makes between the realm of objects to be grounded and the principles of validity which essentially perform this grounding. According to Neo-Kantianism these principles must show themselves to possess an entirely different character.

If we understand what is principally grounded as the ‘world’, then the transcendental subject which grounds it principally is ‘extramundane’. The consequence of this conception is that all empirical features of the subject must be ascribed to the principal grounded domain, and further that all ontology which regards being as an immediate object of access in itself must be declared obsolete. It was essential to question the radical feasability of the development and transformation of Neo-Kantianism in terms of both the subject and ontology. In the long term, therefore, the issues of the subject and of an ontological presupposition of any kind were equally impossible to ignore.

The significance of the second contribution of Neo-Kantianism, the projected philosophy of culture, lies first in its attempted philosophical reflection on the existence and nature of science. The Marburg School placed its emphasis here on the so-called exact sciences, while the Southwest German School concentrated more on the so-called human sciences. It would be a mistake, however, to reduce the Neo-Kantian philosophy of culture to a specific kind of philosophical theory of value, for Neo-Kantianism devoted itself also to questions of practical philosophy, philosophy of art and philosophy of religion. It is disputed just how far Neo-Kantianism succeeded in its philosophy of culture in gaining a perspective on the political and social reality of its time. It is unquestionable, however, that Neo-Kantianism attempted to advance in this direction, as can clearly be seen, for example, in Cohen’s concept of an ethical socialism, Natorp’s pedagogy, Cassirer’s engagement with the myths of the twentieth century and Cohn’s philosophy of culture. Aside from the question of whether these attempts succeeded, however, it cannot be ignored that the school of Neo-Kantianism also included representatives of a regressive and nostalgic conservatism, such as Bruno Bauch, who in their later works openly defended a populist-naturalist position.

Despite the undeniable limitations of the Neo-Kantian systematic approach and the historical distance between Neo-Kantian and contemporary thinkers, the concern and themes of the Neo-Kantian philosophers have lost none

of their relevance. For now, as then, the issue at stake is the defence of the autonomy of the subject. Since this was seen to be in danger, the Neo-Kantians of both the Marburg and the Southwest German schools insisted on the generative moment inherent in the scientific and cultural form of objectivization, or on an evaluative element which necessarily comes into play with such objectivization. If philosophy today does not wish to capitulate to the current trend towards the repudiation of the subject, then it must follow the strategy already preached by the Neo-Kantians, namely to return to Kantian insights and develop Kant's thought further in a systematic manner.

See also: Cohen, H.; Nishida Kitarō; Neo-Kantianism, Russian

Translated from the German by Jane Michael and Nicholas Walker
HANS-LUDWIG OLLIG

References and further reading

1 Primary literature

Bauch, B. (1917) *Immanuel Kant*, Leipzig: Goschen.(General interpretation of Kantian thought from the perspective of the Southwest German School of Neo-Kantianism.)

Bauch, B. (1923) *Wahrheit, Wert und Wirklichkeit (Truth, Value and Reality)*, Leipzig: Meiner.(Deals with the problems of systematically linking theories of truth, reality and ethics.)

Bauch, B. (1926) *Die Idee (The Idea)*, Leipzig: Reinecke.(Presentation of the Idea as a unity of logical functioning.)

Bauch, B. (1931) *Selbstdarstellung (Self-Portrayal)*, in H. Schwarz (ed.) *Deutsche systematische Philosophie nach ihren Gestaltern (German Systematic Philosophy According to its Creators)*, Berlin: Junker & Dünnhaupt, vol. 1, 237-9.(A philosophical autobiography leading up to the 1930s.)

Bauch, B. (1935) *Grundzüge der Ethik (Fundamentals of Ethics)*, Stuttgart: Kohlhammer.(An outline of the ethics of value-formation.)

Cassirer, E. (1902) *Leibniz System in seinen wissenschaftlichen Grundlagen (Leibniz' System in its Scientific Foundations)*, Marburg: N.G. Elwertsche Verlagsbuchhandlung.(An epistemological interpretation of Leibniz's thought.)

Cassirer, E. (1906-20) *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neuen Zeit (The Problem of Knowledge in the Philosophy and Science of the Modern Age)*, Berlin: Bruno Cassirer, 4 vols.(A history of epistemology from a scientific-historical perspective.)

Cassirer, E. (1910) *Substanzbegriff und Funktionsbegriff (The Concept of Substance and the Concept of Function)*, Berlin: Bruno Cassirer.(Describes the transition from the concept of substance to that of function characteristic of modern thought.)

Cassirer, E. (1923-9) *Philosophie der symbolischen Formen*, Darmstadt, 3 vols, 2nd edn, 1977-85; trans. R. Manheim, *The Philosophy of Symbolic Forms*, New Haven, CT: Yale University Press, 1955.(Outline of a comprehensive cultural philosophy based on his theory of symbolism.)

Cassirer, E. (1944) *An Essay on Man*, New Haven, CT, and London: Yale University Press.(Outline of a philosophical anthropology.)

Cassirer, E. (1946) *The Myth of the State*, New Haven, CT: Yale University Press.(Outline of a social philosophy.)

Cassirer, E. (1956) *Wesen und Wirkung des Symbolbegriffs (The Nature and Influence of the Concept of Symbol)*, Darmstadt: Wissenschaftsliche Buchgesellschaft.(A collection of important essays on Cassirer's understanding of 'symbol'.)

Cohen, H. (1871) *Kants Theorie der Erfahrung (Kant's Theory of Experience)*, Berlin: Dümmler, 2nd edn, 1885. (Reconstruction of Kant's first critique.)

Cohen, H. (1877) *Kants Begründung der Ethik (Kant's Foundation of Ethics)*, Berlin: Dümmler.(Reconstruction of Kant's second critique.)

Cohen, H. (1899) *Kants Begründung der Ästhetik (Kant's Foundation of Aesthetics)*, Berlin: Dümmler. (Reconstruction of Kant's third critique.)

Cohen, H. (1902) *Logik der reinen Erkenntnis (Logic of Pure Knowledge)*, Berlin: Bruno Cassirer.(Reconstruction of the foundation of a theory of knowledge.)

Cohen, H. (1904) *Ethik des reinen Willens (Ethics of Pure Will)*, Berlin: Bruno Cassirer. (Theory of the foundations of ethics.)

Cohen, H. (1912) *Ästhetik des reinen Gefühls (Aesthetics of Pure Feeling)*, Berlin: Bruno Cassirer, 2 vols.(Theory

of the foundations of ethics.)

Cohen, H. (1915) *Der Begriff der Religion im System der Philosophie (The Concept of Religion in the Philosophical System)*, Gießen: Töpelmann.(Interpretation of the relationship between religious thinking and Neo-Kantian system concepts.)

Cohen, H. (1919) *Religion der Vernunft aus den Quellen des Judentums*, Frankfurt am Main: J. Kaufmann; trans. S. Kaplan, *Religion of Reason out of the Sources of Judaism*, New York: Ungar, 1972.(An attempt to synthesize Neo-Kantian systems and Jewish religious thought.)

Cohn J. (1901) *Allgemeine Ästhetik (Universal Aesthetics)*, Leipzig: Engelmann. (Outline of a theory of aesthetics based on moral philosophy.)

Cohn J. (1908) *Voraussetzungen und Ziele des Erkennens (The Presuppositions and Ends of Knowledge)*, Leipzig: Engelmann.(Outline of a morally oriented theory of knowledge.)

Cohn J. (1914) *Der Sinn der gegenwärtigen Kultur (The Meaning of Contemporary Culture)*, Leipzig: Meiner.(An interpretation of the cultural situation on the eve of the First World War.)

Cohn J. (1919) *Geist der Erziehung. Pädagogik auf philosophischer Grundlage (The Spirit of Education: Pedagogy on a Philosophical Basis)*, Leipzig: Teubner.(Outlines an educational philosophy based on moral philosophy.)

Cohn J. (1923) *Selbstdarstellung (Self-Presentation)*, in R. Schmidt (ed.) *Die deutsche Philosophie der Gegenwart in Selbstdarstellungen (Self-Presentation of Contemporary German Philosophy)*, Leipzig: Meiner, vol. 2, 61-81. (Autobiographical account of Cohn’s philosophy until the 1920s.)

Cohn J. (1926a) ‘Erlebnis, Wirklichkeit, Unwirkliches’ (Experience, Reality and the Unreal), *Logos* 15: 194-215. (An essay on the problems of understanding the concept of reality.)

Cohn J. (1926b) *Befreien und Binden, Zeitfragen der Erziehung überzeitlich betrachtet (Liberation and Commitment, Contemporary Questions of Education from a Super-temporal Perspective)*, Leipzig: Quelle & Meyer.(Collection of essays on educational philosophy.)

Cohn J. (1932) *Wertwissenschaft (Science of Value)*, Stuttgart: Frommann.(Outline of a systematic moral philosophy.)

Cohn J. (1955) *Wirklichkeit als Aufgabe (Reality as Task)*, Stuttgart: Kohlhammer. (Outline of a science of reality.)

Lange, F.A. (1866) *Geschichte des Materialismus und Kritik seiner Bedeutung in der Gegenwart*, Iserlohn: J. Baedeker; 2nd edn, Iserlohn and Leipzig, J. Baedeker, 1873-5, 2 vols; trans. E.C. Thomas, *History of Materialism*, Boston, MA, Osgood, 1877, 3 vols.(A detailed study of the history of materialism from the ancient Greeks to the 1870s including material on contemporary developments in physics, biology and social theory, and their impact on a Neo-Kantian thought.)

Lask, E. (1923-4) *Gesammelte Schriften (Collected Writings)*, Tübingen: J.C.B. Mohr, 3 vols.(Includes posthumous writings. *Fichtes Idealismus und die Geschichte (Fichte’s Idealism and History)*, Lask’s doctoral thesis referred to in §4, can be found in volume 1, pages 1-273.)

Liebmann, O. (1865) *Kant und die Epigonen (Kant and the Epigones)*, Stuttgart: Carl Schober.(A critique of post-Kantian philosophy and its un-Kantian orientation towards the thing-in-itself.)

Meyer, J.B. (1857) *Über Sinn und Wert des Kritizismus (On the Meaning and Value of Critical Philosophy)*. (Pleads for a self-limitation of philosophy regarding ideological questions.)

Natorp, P. (1884) *Forschungen zur Geschichte des Erkenntnisproblems im Altertum (Investigations into the History of the Problem of Knowledge in Antiquity)*, Berlin: Hertz.(An investigation into classical epistemology.)

Natorp, P. (1899) *Sozialpädagogik (Social Pedagogy)*, Stuttgart: Frommann.(Outline of a concept of social pedagogy.)

(1902) *Platons Ideenlehre (Plato’s Doctrine of Ideas)*, Leipzig: Meiner; 2nd edn, 1922.(Epistemological interpretation of Platonic idealism.)

Natorp, P. (1910) *Die logischen Grundlagen der exakten Wissenschaften (The Logical Foundations of the Exact Sciences)*, Leipzig; Berlin: Teubner.(A study of the foundational problems in the exact sciences.)

Natorp, P. (1911) *Die Philosophie, ihr Problem und ihre Probleme (Philosophy, Its Problem and Its Problems)*, Göttingen: Vandenhoek & Ruprecht.(General overview of philosophy from a Marburg Neo-Kantian perspective.)

Natorp, P. (1912a) *Kant und die Marburger Schule*, Berlin: Reuther & Reichard.(Presents the Marburg School’s interpretation of Kant.)

Natorp, P. (1912b) *Allgemeine Psychologie nach kritischer Methode (General Psychology According to the Critical Method)*, Tübingen: J.C.B. Mohr.(Outlines a philosophical psychology.)

Natorp, P. (1925) *Vorlesungen über praktische Philosophie (Lectures in Practical Philosophy)*, Erlangen: Verlag der philosophischen Akademie.(Foundations for a practical philosophy.)

Natorp, P. (1958) *Philosophische Systematik (The Systematic Structure of Philosophy)*, Hamburg: Meiner. (Outlines a theory of categories.)

Paulsen, F. (1875) *Versuch einer Entwicklungsgeschichte der Kantischen Erkenntnistheorie (Contribution Towards a History of the Development of Kant's Theory of Knowledge)*, Leipzig: Fues's Verlag (R. Riesland). (Reconstructs the genesis of Kantian epistemology.)

Rickert, H. (1921) *Der Gegenstand der Erkenntnis (The Object of Knowledge)*, Tübingen: J.C.B. Mohr, 5th edn, improved.(Outline of the general foundation of the theory of knowledge.)

Rickert, H. (1896-1902) *Die Grenzen der naturwissenschaftlichen Begriffsbildung (The Limits of Scientific Concept Formation)*, Tübingen: J.C. Mohr, 2 vols.(The classic work of the Southwest German School of Neo-Kantianism on the logic of a historical science.)

Rickert, H. (1915) *Kulturwissenschaft und Naturwissenschaft*, Tübingen: J.C.B. Mohr, 3rd edn; trans. G. Reisman, *Science and History*, Princeton, NJ: Van Nostrand.(Presents the dualistic epistemology characteristic of the Southwest German School of Neo-Kantianism.)

Rickert, H. (1920) *Die Philosophie des Lebens (The Philosophy of Life)*, Tübingen: J.C.B. Mohr.(Discussion of the philosophy of life.)

Rickert, H. (1921) *System der Philosophie (System of Philosophy)*, Tübingen: J.C.B. Mohr, vol. 1.(Develops a philosophical system, which remained unfinished.)

Rickert, H. (1930) *Die Logik des Prädikats und das Problem der Ontologie (Predicate Logic and the Problem of Ontology)*, Heidelberg: Winters Universitätsbuchhandlung.(Develops a liguistic-philosophical approach to ontology.)

Rickert, H. (1932) *Thesen zum System der Philosophie (Theses on the Systematic Philosophy)*, *Logos* 4: 97-102. (Sketches a general outline of philosophy.)

Rickert, H. (1934) *Grundprobleme der Philosophie, Methodologie, Ontologie, Anthropologie (Fundamental Problems of Philosophy, Methodology, Ontology, Anthropology)*, Tübingen: J.C.B. Mohr.(Essential philosophical problems from a methodological, ontological and anthropological point of view.)

Riehl, A. (1876-7) *Der philosophische Kritizismus und seine Bedeutung für die positive Wissenschaft (Philosophy of Criticism and its Significance for Positive Science)*, Leipzig: Engelmann Verlag, 3 vols.(Develops of a realistic interpretation of Kant.)

Schleiermacher, F.D.E. (1839) *Dialektik (Dialectic)*, ed. Jonas, Berlin: G. Reimer. (The first post-Kantian attempt to construct a theory of knowledge based on pure thought as well as on sensory perception.)

Trendelenburg, A. (1840) *Logische Untersuchungen (Logical Investigations)*, Leipzig: Hirzel Verlag, 1870, 2 vols.(Develops a non-speculative logic oriented towards the particular sciences.)

Windelband, W. (1878-80) *Geschichte der neueren Philosophie in ihrem Zusammenhange mit der allgemeinen Kultur und den Wissenschaften (History of Modern Philosophy in its Relation to General Culture and the Sciences)*, Leipzig: Breitkopf & Haertel, 2 vols.(A history of philosophy from the Italian Renaissance to the psychologism of Fries and Beneke.)

Windelband, W (1884) *Präludien (Preludes)*, Tübingen: J.C.B. Mohr, 2 vols.(A collection of essays on philosophical-historical and systematic questions.)

Windelband, W (1902) *Lehrbuch der Geschichte der Philosophie (Textbook of the History of Philosophy)*, Tübingen: J.C.B. Mohr, 17th edn, 1994.(Reviews problems in the history of philosophy.)

Windelband, W (1914) *Einleitung in die Philosophie (Introduction to Philosophy)*, Tübingen: J.C.B. Mohr. (Systematic introduction to moral philosophy.)

Zeller, E. (1877) *Über Bedeutung und Aufgabe der Erkenntnistheorie (On the Significance and Task of the Theory of Knowledge)*, in *Vorträge und Abhandlungen (Addresses and Essays)*, Leipzig: Fues's Verlag (R. Reisland), 479-96.(Presents a theory of knowledge as a formal basis for the whole of philosophy.)

2 Secondary literature

Brelage, M. (1965) *Studien zur Transzendentalphilosophie (Studies on Transcendental Philosophy)*, Berlin: de Gruyter.(Places Neo-Kantianism in the context of the problem of 'Transcendental Philosophy and concrete Subjectivity'.)

Flach, W. and Holzhey, H. (1980) *Erkenntnistheorie und Logik im Neukantianismus (Theory of Knowledge and Logic in Neo-Kantianism)*, Hildesheim: Olms.(Contains, apart from a collection of important source texts, an introduction to the theoretical philosophy of Neo-Kantianism.)

Holzhey, H. (1992) *Der Neukantianismus*, in A. Hügli and P. Lübke (eds) *Philosophie im 20. Jahrhundert*

(Philosophy in the Twentieth Century), Hamburg: Rowohlt-Verlag, vol. 1, 19-51.

Holzhey, H. and Orth, E.-W. (1994) *Neukantianismus. Perspektiven und Probleme. Eine Zwischenbilanz der jüngsten Neukantianismusforschung (Neo-Kantianism: Perspectives and Problems; An Assessment of Contemporary Research into Neo-Kantianism)*, Hildesheim: Königshausen & Neumann.(Attempt at a summary of recent international research into Neo-Kantianism.)

Köhnke, K.C. (1986) *Entstehung und Aufstieg des Neukantianismus (The Development and Rise of Neo-Kantianism)*, Frankfurt am Main: Suhrkamp. (Comprehensive investigation into the beginnings of Neo-Kantianism.)

Oakes, G. (1990) *Die Grenzen der kulturwissenschaftlichen Begriffsbildung (The Limits of the Formation of Expression in the History of Culture)*, Frankfurt am Main: Suhrkamp.(Discusses the relationship of Rickert's philosophy to Weber's methodology.)

Oelkers, J., Schulz, W.K. and Tenorth, H.-E. (1989) *Neukantianismus. Kulturtheorie, Pädagogik und Philosophie (Neo-Kantianism: Theory of Culture, Pedagogy and Philosophy)*, Weinheim: Deutscher Studien Verlag.(Contains contributions to an appreciation of Neo-Kantian educational theory.)

Ollig, H.-L. (1987) *Materialien zur Neukantianismus-Diskussion (Material on the Discussion of Neo-Kantianism)*, Darmstadt: Wissenschaftliche Buchgesellschaft. (Collection of important essays on research into Neo-Kantianism.)

Willey, T. (1987) *Back to Kant. The Revival of Kantianism in German Social and Historical Thought 1860-1914*, Detroit, MI: Wayne State University Press.(Casts light on the return to Kant in its contemporary context.)

Neo-Kantianism, Russian

A rather amorphous movement, Russian Neo-Kantianism, in the first decades of the twentieth century, found its most visible and enduring representatives in A. Vvedenskii and his student/disciple I. Lapshin, both of St Petersburg University, who together took a distinct stance within the movement as a whole. Both were chiefly concerned with epistemological issues although their respective publications revealed a much wider field of interests. Both maintained an allegiance to the spirit of Kantian philosophy while devoting little attention to the intricacies of the three Kantian Critiques. In addition, at Moscow University G. Chelpanov adhered to the ideality of the Kantian categories although he upheld an ultimate realism. In the social sciences P. Novgorodtsev and B. Kistiakovskii defended positions revealing a significant debt to the Baden School of German Neo-Kantianism, and as such they were deeply involved in methodological investigations. Several others (N. Berdiaev, P. Struve, G. Shpet) within Russia also briefly espoused a Neo-Kantianism after abandoning Marxism and before moving on to a usually more idealistic stance. Still another, younger group a decade later, centred chiefly around the journal Logos, maintained an idiosyncratic, though not fully articulated, amalgam of Kantian and Hegelian doctrines. The Bolshevik Revolution led to a decimation of what remained of the movement within Russia and a dispersal of its members, who became increasingly isolated from current philosophical trends and issues.

1 Overview of the movement

Russian Neo-Kantianism was neither a unified nor a fully self-conscious movement. Largely as the result of being sent to Germany for further graduate study many young Russian-born scholars returned to their homeland imbued with the dominant trend of their host institution. Like their German counterparts (see Neo-Kantianism) the Russians themselves often shared little more than a conviction that genuine philosophy must begin, though not necessarily end, with Kant. Also similarly to their German counterparts, the Russian Neo-Kantians were not above quarrelling. Yet with their attention fixed on the scene abroad, particularly in Germany, and with their individual, narrow concerns they usually ignored one another, never developing anything similar to the trenchant mutual critiques we find in the competing German Neo-Kantian schools. Although several Russians showed an impulse to systematization, none, with the possible exception of their last representative (Veideman), ever produced a unified work systematically addressing all major philosophical issues. Nor did the Russian Neo-Kantians pen methodic commentaries on any of the three Kantian ’Critiques’ despite one notable excursus into Kant-philology by their leading representative. Partly because they were few in number, but probably more importantly because of the sheer political difficulty of establishing a periodical during the Tsarist era, the Neo-Kantians in Russia had no journal of their own, let alone one for each separate group as in Germany. Even the Russian edition of *Logos*, which was securely in the hands of a young generation of Neo-Kantians, proclaimed itself open to all philosophical directions. The events of 1917 did much to hasten the demise of Neo-Kantianism, although by then its younger members were rapidly moving towards a neo-Hegelian stance.

2 Epistemological Neo-Kantianism

Some scholars trace the beginnings of Russian Neo-Kantianism as far back as the early writings of Pётr Lavrov in the 1860s and Vladimir Lesevich in the 1870s. Nevertheless the first explicit stirrings date from the following decade. Although, unlike the Germans, the Russian Neo-Kantians in general paid little attention to the natural sciences, the initiator, leader and philosophically most influential member of the movement Aleksandr Vvedenskii (1856-1925) started his career with a lengthy master’s thesis - he never attained a doctorate - on the concept of matter. In it he upheld the basic thrust of Kant’s dynamic model of elementary physical processes against the backdrop of contemporary theories. In his mature years he came to call his own stance ’logicism’, which he conceived as a logical proof of the impossibility of metaphysical knowledge and, not immodestly, as the distinctive Russian contribution to philosophy on a par with British empiricism, French positivism and German Kantianism in their opposition to such knowledge. Logicism, in Vvedenskii’s eyes, presupposed no theory of a priori judgment or conclusions about the nature of space and time. Yet he believed it possible to demonstrate, without the elaborate reasoning and considerations found in Kant’s *Critique of Pure Reason* (see Kant, I. §§4-8), that the objects studied by the natural sciences are mere representations of ’true being’, the thing-in-itself, which as such remains totally transcendent.

As all metaphysical hypotheses are irrefutable some can be incorporated into a worldview, though purely on the basis of faith. There is, however, another group of ’views’ that consists neither of metaphysical hypotheses, since the views make no reference to transcendent objects, nor of knowledge claims, since they concern activity in the immanent world. This group is that of our moral tenets. We must integrate into our worldview only those metaphysical hypotheses that are demanded by our moral tenets, only those that we must hold to be true if we are to believe in the ’absolute obligatory nature of moral duty’. Among these hypotheses, or postulates of practical reason, are those discussed by Kant (see Kant, I. §§9-11), but in addition there is one he largely failed to elaborate, namely the existence of mental activity similar to my own in the other.

The Cartesian thesis of the privacy of the mental, that there are no objective criteria by which I know that the other has thoughts and feelings like mine, is one Vvedenskii continued to defend throughout his long professional career at St Petersburg University. The argument from analogy can be of use only in explaining how I extend my conviction in the existence of the other’s mental activity to a number of others but not how I come to my initial recognition (see Other minds). Just as we speak of a moral sense in referring to the recognition of moral obligation, so too can we speak of a metaphysical sense in referring to my acknowledgement of the other’s mental processes. Furthermore, since moral obligation is contingent on the existence of at least one other animate being, I am as certain of the existence of mental activity in the other as I am of the obligatory nature of my duties. In fact, the metaphysical and the moral sense are the same; the former is merely what Vvedenskii called ’the cognitive aspect’ of the latter. In his later years Vvedenskii spoke less of a metaphysical sense and more of accepting the other’s mental activity as a ’working hypothesis’ that proved ’convenient’ in the construction of a scientific psychology.

Not surprisingly, then, Vvedenskii held that psychology as a science yielding knowledge relies necessarily and fundamentally on personal self-observation. Nevertheless he allotted a role, albeit on a secondary level, in a properly conceived psychology to experiment and what he termed ’collective or comparative self-observation’, the accumulation and comparison of introspective observational data by a number of individuals.

Despite Vvedenskii’s wish to found a distinct school of thought his only disciple was Ivan Lapshin (1870-1952), who also taught at St Petersburg University prior to his expulsion from the Soviet Union in 1922. Extending Kant’s original transcendentalism, Lapshin held that even formal logic is grounded in an a priori categorical synthesis. Consequently, anyone who rejects Kantian (or, rather, Neo-Kantian) epistemology has no basis for using the laws of logic. As those laws cannot justifiably be applied in contexts beyond possible experience metaphysics lies beyond the pale of science. Such is the purest expression of ’critical idealism’. Disagreeing with Kant, Lapshin maintained that space and time were essentially not pure intuitions but categories of cognition along with causality, substantiality and so on. Intuited parts of space are merely conceptual ’illustrations’ of space. As a category space is intimately involved in all judgments including analytical ones such as the law of non-contradiction.

Like Vvedenskii, Lapshin devoted considerable attention to the problem of ’the other’. He too claimed that my experience of an immediate givenness of another’s intimate self is illusory. For the most part we analogically infer the other’s mental states from observed bodily movements or positions. Even whether all other rational beings think in accordance with the same laws of logic must be answered agnostically. We are forced to turn to faith. Yet faith can be either immanent or transcendent. A transcendent faith in the other’s self, in other words, that the other has mental states exactly as I do and that they are the ’essence’ of the empirical being before me, is both unnecessary and philosophically unjustified, since it treats the other’s self as a thing-in-itself. On the other hand, immanent faith means we have supreme confidence in the empirical reality, the ’objective significance’, of positing mental states within the other’s empirically observed body.

3 Epistemological Neo-Kantianism (cont.)

While not a disciple of Vvedenskii, Georgii Chelpanov (1863-1936), a professor first at Kiev and then at Moscow University, is almost invariably classed as a Neo-Kantian, although even this is not incontestable. As active in psychology as philosophy he upheld Wundt’s programme in experimental psychology (see Wundt, W.) and defended psychophysical parallelism throughout his career. Vvedenskii himself sharply rebuked Chelpanov for his reliance on J.S. Mill’s ideas in an elementary logic textbook. Yet Chelpanov, unlike many others, sought to retain a role for the thing-in-itself, viewing it as a transcendent something that ’evokes’, along with the forms of

consciousness, a particular representation of an object. Likewise, something analogous to our representation of space ’evokes’ in us that representation. Explicitly in opposition to those who evaded the issue of how to account for a particular representation and called such a position ’transcendental idealism’, Chelpanov termed his stance ‘critical realism’ or ‘ideal-realism’.

While uninterested in the problem of other minds Chelpanov sought to defend the Kantian apriority of space in the light of non-Euclidean geometry by arguing that our particular spatial relations and therefore a particular set of geometrical axioms are formed under the influence of experience. Nevertheless as idealizations the axioms of Euclidean geometry are produced in and by consciousness. Another set of spatial relations will produce another set of axioms. The concept of space simpliciter, however, is a priori in the sense that it is a condition of perception in general. Every space is characterized by its exteriority, and the particular dimensionality of a space is derived from the establishing of a relation between points.

We should also mention Chaim Flekser (pseudonym Akim Volynskii) (1863-1926), a prominent literary critic and editor for a time of the important journal *Severnyi vestnik*. Despite his support for an idiosyncratic, but undeveloped, mysticism Volynskii in an early and lengthy article in that journal held that Kant had conclusively resolved the problems concerning cognition. Kant’s practical philosophy, on the other hand, being based on a ’practical faith’, he held to be imbued with the purest dogmatism.

4 Neo-Kantianism in the social sciences

Pavel Novgorodtsev (1866-1924) took an active role in politics in addition to his duties as professor of jurisprudence at Moscow University. He came under the influence of the Baden and Marburg schools (see Neo-Kantianism §§3-8) during a lengthy stay in Western Europe, particularly in Freiburg. Although endorsing Kant’s epistemology Novgorodtsev’s interests lay mainly in practical philosophy. Whereas Kant’s morality is centred around the notion of duty (see Kantian ethics), Novgorodtsev sought to reconstruct it around that of the unique, harmonious person understood as the sole norm. He seeks to retain the rigour of Kant’s ethical doctrine while recognizing the force of Hegel’s criticism of its formalism by allowing, especially in his early writings, for natural law with a changing content. All practical actions are necessarily limited and incomplete and must be evaluated with their finitude in mind. Nevertheless, laws themselves must be framed keeping in mind the moral foundation of law, the indisputable authority of moral consciousness. In his later writings as the First World War approached Novgorodtsev retreated from his Neo-Kantian programme to a somewhat more traditional conception of natural law.

Bogdan Kistiakovskii (1868-1920) studied for a time under Simmel in Berlin and completed his doctoral studies under Windelband in Strasbourg. After returning to Russia he taught at several institutions including for a short time Moscow University. Kistiakovskii termed his stance, basically a defence of Windelband’s, ’scientifico-philosophical idealism’. He held that just as the basic category in the natural sciences is natural necessity, so in the various disciplines of ’scientific philosophy’ it is found in a consciousness of what should be. Broadly speaking, logic determines what is valuable in a cognitive relation, whereas ethics is concerned with the valuable in practical activity. The normative character of logic is Kant’s greatest discovery in the *Critique of Pure Reason*. Epistemology is concerned not with the very existence of scientific knowledge as a given fact but with the means by which humanity obtains such truths, and with the justification of science.

Nevertheless, Kistiakovskii’s chief interest was not logic but philosophy of the social sciences, where he stood close to the position of Rickert and Weber in Germany. For the social sciences to be legitimate, separate branches of inquiry we must understand how concepts are formed in them. Kistiakovskii denied that categories applicable in the natural sciences can be immediately applied in the social sciences unaltered. Absolute necessity, for example, in the natural sciences is something sought, whereas in the social sciences the search for necessity is but one viewpoint.

Also among those influenced by the Baden school was the historian Aleksandr Lappo-Danilevskii (1863-1919), who taught at St Petersburg University. Although chiefly noted for his concrete historical studies he accorded a singular role to epistemology in the elaboration of principles and methods without which scientific advancement is inconceivable. While sharing the Baden view that the natural sciences are nomothetic and the social sciences primarily but not exclusively idiographic, he also saw a difference in their goal. The natural scientist sought to

explain nature whereas the social scientist strives for understanding of a phenomenon.

The most partisan Russian disciple of the Marburg School, particularly in the philosophy of the social sciences, was Vasilii Savalskii (1873-1915). In his chief work from 1908 he engaged in polemics against Kistiakovskii and Novgorodtsev while expounding the position of Hermann Cohen in ethics and jurisprudence.

5 Transitional Neo-Kantians

Several notable Russian intellectuals early in their respective careers abandoned Marxism and briefly adopted a Neo-Kantian position before moving on to another, usually more idealistic philosophy. Perhaps best known among these was Nikolai Berdiaev (1874-1948). Although acknowledging the primacy of practical reason in one of his earliest publications from 1900, he nevertheless saw this as a psychological truth. Berdiaev held Kant's atemporal postulates of practical reason to be 'temporary postulates of people of a specific socio-psychological formation'. At a different time other interests could predominate leading to a different set of postulates. However, Berdiaev came to the view that there were unalterable moral values which, while atemporal, were in the course of history expounded by different classes. Along these lines he attempted to combine Kantian idealism with a lingering sympathy for Marxian sociology.

Pëtr Struve (1870-1944), a dynamic personality in Russian politics as well as intellectual history, came to Neo-Kantianism largely out of dissatisfaction with the account of freedom and necessity in Marxian philosophy of history. During his Neo-Kantian phase he retained a belief in the validity of a thorough-going determinism for the phenomenal realm while asserting that freedom, conceived as indeterminism, was a 'reality' in the activity of the human individual.

In his first published article from 1907 on Hume and Kant, Gustav Shpet (1879-1937) sided with the Baden-school outlook that while the validity of the law of causality was undemonstrable, we must believe in it for the sake of cognition. Within a few years, however, he abandoned Neo-Kantianism altogether, drawing close to Husserlian phenomenology.

6 The journal *Logos*

A Russian edition of the international journal *Logos* began in 1910 under the responsible editorship of Sergei Hessen (1887-1950) and Fëdor Stepun (1884-1965). In 1911 Boris Iakovenko (1884-1948) was added. Hessen studied in Heidelberg and Freiburg with among others Windelband and Rickert, under whom he wrote a notable dissertation. After returning to Russia he taught in St Petersburg and Tomsk. Subsequently after the Bolshevik Revolution he taught in Prague and Warsaw. Although later in emigration he wrote on pedagogy and philosophy of law his early writings reveal a strong debt to the Baden School. He argued that in the natural sciences to subsume a particular natural event under a general law, thereby illuminating it in a causal sequence, is to understand it. The individual or historical event too is marked with necessity even though the degree of generality appears markedly different in historical and scientific explanations. To understand an event in history is to understand it in its concreteness, its individuality but also its necessity no less than in the natural sciences.

Stepun studied in Heidelberg under Windelband under whom he wrote a slim dissertation on Solov'ëv. Shortly after the Revolution he briefly served as cultural director at a state theatre in Moscow. Later he taught sociology at Dresden University and from 1946 in Munich. His early affiliation with Neo-Kantianism consists chiefly in his view that philosophy must examine all cultural fields, distinguishing in them a 'transcendental-formal element' from that which is merely contingent. This formal element is a manifestation of the absolute and as such is what gives cultural 'creations' their universality and necessity. Unfortunately Stepun failed to develop these thoughts more precisely or fully.

After 1917 Iakovenko first lived in Italy and then Berlin and Prague. Despite his many writings he devoted little attention to enunciating his own position either in detail or rigorously. In short he held to an absolute monism but proclaimed transcendental philosophy to be the intellectual path to the Absolute. In it philosophy has attained a self-consciousness. The next and last step for philosophy is to return to the original unity revealed in mythology. When philosophy has come full circle it will then be critical and rational.

7 The last Russian Neo-Kantians

Vasilii Sezeman (1884-1963) was born in Finland and studied in St Petersburg. After graduation from the university there he went abroad to Marburg and Berlin for further study. Returning to Russia he taught during the First World War in Petrograd and later in Saratov. In the early 1920s he lived briefly in Berlin and then in Lithuania, where he taught philosophy at the Universities of Kaunas and Vilnius. During this later period his major writings were composed in German, although a few years prior to his death he published a Lithuanian translation of Aristotle's *De Anima*. Particularly interested in bestowing a transcendental significance on the concept of evolution without falling into psychologism, Sezeman held that knowledge at any particular time is incapable of seizing the transcendent object. Absolute knowledge remains the ideal, but it itself is inseparable from the path leading to the latter. Each stage in the advance of knowledge functions as a constitutive element in this ideal. As truth is a regulative idea there are no absolute, self-evident truths; all propositions are subject to evolution including the categories or seemingly ultimate elements of knowledge. Sezeman thus rejects interpreting the thing-in-itself as some causative transcendent reality and views it instead as an ideal, self-contained system.

Aleksandr Veideman (1879-?) published in Latvia in 1927 the only attempt at a systematization in Russian Neo-Kantianism which nevertheless stands as the 'swan song' of the movement. Situating himself between Kant and Hegel, Veideman faulted the former for limiting himself to the possibility of mathematical and scientific knowledge, not knowledge in general. Had Kant done so he would have affirmed that all being is created by thought dialectically. As thought here is conceived to be a logical not individual process, solipsism is averted. The impetus for creative cognition receives its answer in ethical striving. Aesthetics represents the field wherein the finite is raised to an ideal.

As in the parallel German movement concern with the historical Kant was greatest during the early years of the respective protagonists and gradually diminished. In some cases this was simply a matter of an evolving thought-process, whereas in others external events surely played a decisive role. Vvedenskii distanced himself from the intricacies of Kant's 'First Critique', seeking an independent and 'easier proof of philosophical Criticism'. Chelpanov, witnessing the consolidation of Marxism in Russia in the 1920s, sought above all to defend moderation and the toleration of competing viewpoints in philosophy. To this end he emphasized his own hostility and that of Marx to transcendent metaphysical explanation. Lapshin, increasingly isolated in Prague, turned ever more towards studies of Russian culture. As the years turned into decades Sezeman, perhaps alone among the Neo-Kantians in the Russian diaspora, continued working without thematic abatement on problems traditionally conceived as philosophical, keeping abreast of the latest developments.

THOMAS NEMETH

List of works

Berdiaev, N. (1900) 'F.A. Lange i kriticheskaia filosofiia' ('F.A. Lange and Critical Philosophy'), in *Mir bozhii* (1900) July: 224-54.(Reveals a side to his thought far different from his later 'existentialist' turn.)

Chelpanov, G. (1896-1904) *Problema vospriiatiia prostranstva v sviazi s ucheniem ob apriornosti i vrozhdënnosti (The Problem of the Perception of Space in Connection with Theories of Apriority and Innateness)*, Kiev: Kushnerov, 2 vols.(Of his numerous works his most scholarly.)

Chelpanov, G. (1905) *Vvedenie v filosofiiu (Introduction to Philosophy)*, 6th edn, Moscow: V.V. Dumnov, 1916. (Admirably clear and simple prose.)

Hessen, S. (1909) *Individuelle Kausalität. Studien zum transzendentalen Empirismus (Individual Causality. Studies in Transcendental Empiricism)*, Berlin: Reuther & Reichard.(A treatment of an issue central to the Baden School of Neo-Kantianism.)

Kistiakovskii, B. (1899) *Gesellschaft und Einzelwesen. Eine methodologische Untersuchung (Society and the Individual. A Methodological Investigation)*, Berlin: Verlag Otto Liebmann.(A Neo-Kantian examination of the foundations of sociology that paralleled much of Weber's methodological work.)

Kistiakovskii, B. (1916) *Sotsial'nye nauki i pravo (Social Sciences and the Law)*, Moscow.(A general theory of law from the standpoint of sociology.)

Lappo-Danilevskii, A. (1910-13) *Metodologiia istorii (The Methodology of History)*, St Petersburg, 2 vols.(A monumental study by a distinguished historian.)

Lapshin, I. (1906) *Zakony myshleniia i formy poznaniia (Laws of Thinking and Forms of Cognition)*, St Petersburg.(Despite its frequent departure from academic style this is his basic work.)

Lapshin, I. (1924) 'Oproverzhenie solipsizma' ('A Refutation of Solipsism'), repr. in *Filosofskie nauki* (1992) 3: 18-45.(A valiant though vain attempt to overcome the limitations of his own essentially methodological

solipsistic stance.)

Novgorodtsev, P. (1901) *Kant i Gegel' v ikh ucheniiakh o prave i gosudarstve (Kant and Hegel on Law and the State)*, Moscow: Moscow University.(Far more than simply an examination of Kant and Hegel.)

Novgorodtsev, P. (1909) *Krizis sovremennogo pravosoznaniia (The Crisis of Contemporary Legal Consciousness)*, Moscow: Moscow University.(The major statement of his fully developed position.)

Savalskii, V. (1908) *Osnovy filosofii prava v nauchnom idealizme (Foundations of the Philosophy of Law in Scientific Idealism)*, Moscow: Moscow University.(Highly polemical.)

Sezeman, V. (1911) 'Ratsional'noe i irratsional'noe v sisteme filosofii' ('The Rational and the Irrational in a Philosophical System'), in *Logos* 2: 93-122.(An early generally mediocre essay.)

Shpet, G. (1907) *Problema prichinnosti u Iuma i Kanta, (The Problem of Causality in Hume and Kant)*, Kiev: St Vladimir University. (Unfortunately Shpet states his own position only at the very end.)

Veideman, A. (1927) *Myshlenie i bytie (Thinking and Being)*, Riga.(An almost unknown, albeit sketchy, amalgam of Neo-Kantianism with elements of Hegelianism.)

Volynskii, A. (1889) 'Kriticheskie i dogmaticheskie ėlementy v filosofii Kanta' ('Critical and Dogmatic Elements in Kant's Philosophy'), in *Severnyi vestnik*, (1889) 7: 67-87; 9: 61-83; 10: 89-109; 11: 51-72; 12: 55-78. (Surprising support for Kant from an unexpected corner.)

Vvedenskii, A. (1901) *Filosofskie ocherki (Philosophical Essays)*, 2nd edn, Prague, 1924.(Contains a number of valuable pieces illuminating his general philosophical outlook.)

Vvedenskii, A. (1909) *Logika, kak chast' teorii poznaniia (Logic as a Part of a Theory of Cognition)*, 3rd edn, St Petersburg: Stasjulevich, 1917. (His basic text on epistemology.)

References and further reading

Putnam, G. (1977) *Russian Alternatives to Marxism*, Knoxville, TN: University of Tennessee Press.(Useful and very readable discussion of Novgorodtsev.)

Vucinich, A. (1976) *Social Thought in Tsarist Russia*, Chicago, IL: University of Chicago Press, 106-52.(Incisive and highly informative particularly on Kistiakovskii.)

Walicki, A. (1987) *Legal Philosophies of Russian Liberalism*, Oxford: Clarendon Press, 291-465.(Exciting and learned discussion of Novgorodtsev, Kistiakovskii and Hessen.)

Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, vol. 2, Paris: YMCA-Press; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, vol. 667-705.(Despite its age this remains the best secondary source.)

Neoplatonism

Neoplatonism was the final flowering of ancient Greek thought, from the third to the sixth or seventh century AD*. Building on eight centuries of unbroken philosophical debate, it addressed questions such as: What is the true self? What is consciousness and how does it relate to reality? Can intuition be reconciled with reason? What are the first causes of reality? How did the universe come into being? How can an efficient cause retain its identity and yet be distributed among its effects? Why does the soul become embodied? What is the good life?*

There were several flavours of Neoplatonism, reflecting the concerns and backgrounds of its practitioners, who ranged from Plotinus and his circle of freelance thinkers to the heads of the university schools of the Roman Empire, Proclus, Ammonius and Damascius. In the later, more analysed form, we see a rich scheme of multi-layered metaphysics, epistemology and ethics, but also literary theory, mathematics, physics and other subjects, all integrated in one curriculum. Neoplatonism was not just a philosophy but the higher education system of its age.

The Neoplatonism that came to dominate the ancient world from the fourth century was an inseparable mixture of inspired thought and scholastic order. To this may be traced some of its internal conceptual conflicts: for example, the free individual soul versus the ranks of being, personal experience versus demonstrative knowledge. To this may also be traced its appeal to polar audiences: mystics and mathematizing scientists, romantics and rationalists.

To the Neoplatonist, knowledge consists of degrees of completion. Take the example of tutor and student. Both study the same things, but the tutor has a wider and more intimate knowledge. The tutor opens the student's mind to the breadth and intricacies of the study-matter, and corrects the student's deliberations. So it is with the Neoplatonic levels of knowledge. Every level has access to the entire spectrum of what there is to know, but each with its appropriate adverbial modifier. At the 'lower' level an individual comprehends things 'particularly' and is concerned with the 'images' or presentations of mind and sense-impressions of the qualities of physical things. At the 'higher' level, an individual apprehends things 'wholly', as universal statements (often called 'laws' and 'canons'). The concern is with propositions about what is true or false, self-grounded and logically necessary. Thus the higher level corrects and supplies the 'criterion' for the lower level.

Knowledge, however, is not an end in itself, but a means to salvation. Increasing awareness puts us in touch with the levels of reality of which we ourselves are part. The ultimate reality is none other than the fundamental unity out of which all came into being: God. In this union we recover our true good.

As the summation of ancient Greek philosophy, Neoplatonism was transmitted to Byzantium, Islam and western Europe. It was the prime intellectual force behind the protagonists of the Italian Renaissance, and its influence was felt until the nineteenth century.

1 Who, what, when?

The modern label 'Neoplatonism' conveniently describes the philosophy of late antiquity, and at once misrepresents it. It was neither 'new', a departure from Plato, nor simply a 'Platonism' (see Plato; Platonism, Early and Middle). The more we learn of the philosophers from the first century BC to the second century AD, such as Cicero, the Middle Platonists and Neo-Pythagoreans, the more we discern a continuous line of thought (see Cicero; Neo-Pythagoreanism). How far back it stretches is still intriguing and controversial, since several leading scholars have indicated independently a proximity in thought between Neoplatonism and the late Plato and his immediate successors in the Old Academy (see Speusippus; Xenocrates). The Neoplatonists were known in their time as 'Platonists'; those at Athens regarded themselves the true successors to Plato, and called their school the 'Academy'. But these ancient labels mislead as much as the modern, and to understand the sources of Neoplatonism we have to look also to the Stoics and the commentators on Aristotle (see Stoicism; Aristotle Commentators). Plato's and Aristotle's schools disappeared as institutions in the first century BC. Freelance philosophers took up the different causes. In the second century AD the Roman emperor-philosopher Marcus Aurelius (§1) established chairs of philosophy at Athens (apparently bearing the title 'successor') in Platonism, Aristotelianism, Stoicism and Epicureanism. Alexander of Aphrodisias, in his endeavour to unify the Aristotelian corpus, identified the active intellect (*nous*) with god and with its intelligible object, unintentionally easing the

absorption of Aristotle in mild Platonism. By the beginning of the third century imperial funding for the chairs had vanished, and by the end of that century only Platonists survived. Some, who were more inclined to scholasticism, flourished at the prestigious teaching centres of the late Roman Empire, Athens and Alexandria. Others, particularly those of the Neo-Pythagorean tendency, such as Numenius, flourished at other centres, notably Apamea (Syria). Both sorts took up relevant ideas from past Greek philosophies, including Aristotelian theories of logic and mind, and rejected those they considered fruitless (Epicureanism) or wrong (for example, the dualism of the Platonist Plutarch of Chaeronea §3).

This is the philosophy that Plotinus, Porphyry and Iamblichus turned into a fresh, highly integrated school of thought, now called, for lack of a better term, Neoplatonism. Each of the three founders contributed something different, often in reaction to their teacher. Plotinus abandoned scholastics in favour of a Pythagorizing Platonist, Ammonius Saccas, at Alexandria. Later, he set up his own circle of followers in Rome, which was to prove highly influential. Porphyry first studied under the leading scholastic at Athens (Longinus) but then studied under Plotinus. Only five years later he abandoned Plotinus and pursued his explicit 'harmonization' of Aristotle with Plato. Iamblichus taught philosophy at Apamea. He had read Porphyry and may have been his student, but disagreed with him over the approach to the divine. One of Iamblichus' disciples, Aedesius, founded a celebrated school at Pergamum, and proselytized the emperor Julian. By the fifth century the Iamblichean version had spread to the main teaching establishments at Athens (see Damascius; Proclus; Simplicius) and Alexandria (see Ammonius, son of Hermeas; Philoponus), as well as Gaza (which was Christian).

Neoplatonists saw their metaphysics in all forms of expression, provided they could find the appropriate level of interpretation. Language is the mediator of concepts and realities, and so even literary works (for example, those by Homer and Hesiod) allude to metaphysical truths. Similarly, revelatory proverbs, the Orphics and the Chaldaean Oracles, point to truths that transcend ordinary description (see Orphism; Chaldaean Oracles). The Neoplatonists did not admit irrationalism, however, remaining within the Greek rationalist manner of explanation.

Their writings are mostly commentaries on classical greats, Plato and Aristotle - although there are notable exceptions, including the *Enneads* of Plotinus, Porphyry's *On Abstinence from Animal Food*, Iamblichus' *On the Mysteries*, Proclus' treatises *On Providence, On Fate*, *On Evil* and *Platonic Theology* and Damascius' treatise *Problems and Solutions on First Principles*. As a textual form, the commentary is no evidence of lack of originality: it was the conventional form of scholarly writing of the middle and late imperial period, and was useful for teaching. Good interpretative works contained critical appraisals and innovative theories using the set text as a frame for the commentator's own reflections. That two Athenian heads of school (Proclus and Damascius) wrote most of the commentaries on Plato, while the Alexandrians concentrated on Aristotle, was not due to an ideological split, but to avoidance of duplication. The Alexandrian professors were envious of the rich endowments given to the Athenians and the Athenians looked with disdain on the money-making compromises of the Alexandrians, but the two schools enjoyed close bonds, and students from the one regularly travelled to take positions at the other (for example, Hierocles, Hermeias, Ammonius, Damascius and Simplicius).

On the Latin side, Plotinus' and Porphyry's works were translated by the pagan Macrobius and the Christian convert Marius Victorinus. The latter transformed Neoplatonism to a form suited to Roman Christianity and influenced Augustine and Boethius (see Marius Victorinus; Augustine; Boethius, A.M.S.).

On the Greek side, Neoplatonists supported Hellenism as a civilizing culture and as a universal, pagan religion, at a time of Christian emperors. The school at Athens ended with the emperor Justinian's imperial ban on pagan public teaching in 529. There was an unsuccessful sojourn to Persia, but it is still a mystery what finally happened to the last Athenian scholarchs, Damascius, and especially Simplicius, who flourished after the ban. At Alexandria, Neoplatonic teaching continued with the pagan Olympiodorus, and then with Christians (Elias and David). In the seventh century, Alexandrian teaching transferred to the Byzantine capital, when Stephanus was invited to teach at the university of Constantinople (see Byzantine philosophy §§2,3).

2 The curriculum of knowledge and virtue

The Neoplatonic curriculum was established by Iamblichus, and maintained with minor changes by Athenians and Alexandrians over three centuries until the end of independent Neoplatonist schools (whence it was transmitted to Byzantium). It consisted of thematic topics tuned to educational 'scopes'. Beginning with Aristotle's logic (first

through Porphyry’s *Introduction*) and practical ethics, it progressed to physics, the mathematical sciences, Aristotle’s metaphysics, and twelve of Plato’s dialogues culminating with the *Timaeus* and *Parmenides*. After approximately six years of university education, the student graduated as a philosopher and a virtuous person.

From Socrates and Aristotle, Neoplatonists accepted that *aretē* (virtue) can be taught (see Aretē): however, this is not to say that teachers impart goodness to students, but that they make the students fully aware of their true nature. Each grade of the curriculum was meant to stimulate the appropriate virtue. According to the Platonic *Alcibiades I*, *psychē* (soul/life/mind) is the human essence (see Psychē). Educating the person means educating the soul as to its true origins and place in the scheme of things. Thus, education and ethics are grounded in epistemology and metaphysics. Concentrating on the essence of things involves a certain detachment from incidentals, the temporal, physical events. The soul must be ’unaffected’ (*apathēs*) by externals. Eventually a sense of completion and full awareness bring about the state of wellbeing (*eudaimonia*) that is the goal of human life (Eudaimonia). Ultimately this is nothing else than ’becoming similar to God’.

As there are many degrees of awareness, so there are many levels of virtue. The physical (*physikai*) virtues relate to a well-maintained body: for example, genetic characteristics, beauty, state of health, stamina. Ethical (*ēthikai*) virtues relate to the sense of justice, courage, temperance and truth (the four Platonic cardinal virtues) and, generally to the temperament of the individual, including, it seems, the quality of memory. Social (*politikai*) virtues measure regard for one’s fellow human beings, participation in politics and affairs at large: for example, standing up for human welfare and dignity against state officialdom (compare Aristotle’s definition of the human being as social animal). The three types outlined above belong to the ’lower’ level because they refer to passing matters, not to the atemporal essence of life. The first of the ’higher’ virtues are the purifying (*kathartikai*) virtues. The detachment of the soul can be assisted through a simple lifestyle consisting of abstinence and vegetarianism (Porphyry was the first to write on its philosophy), fasting and praying. Erotic love was not shunned, and most of the Neoplatonist scholarchs enjoyed fertile marriages. The aim was to redirect desire towards the eternal. Next are the intellectual (*theōrētikai*) virtues: learning about names (*onomata*) as signifying concepts (*noēmata*) of real objects (*pragmata*), both natural and divine. Iamblichus added also a sixth type, the theurgic virtue. What exactly he meant is still controversial, but it included initiation in magic, prediction of natural disasters, out-of-body visions and, in general, acting like god while being a mortal. The leap to the transcendent absolute was theurgy’s highest aim. Finally, there was the ‘paradigmatic’ virtue, which in earlier Neoplatonism denoted the exemplary state of divine union but in later Neoplatonism was reserved solely for divine beings.

3 Epistemology, metaphysics and the daring soul

Pure sense-perception ranks low in the scale of certainty, because its information is incidental and transient. Mind arrives at an opinion or judgment (*doxa*) by assimilating sense-impressions into images (*phantasiai*) and comparing them with preconceived patterns of thought, which are intelligible and can be articulated, thus forming the mind’s internal objects. Certain, scientific knowledge (*epistēmē*) is the result of reasoning. This is the mental discourse (*dianoia*) in which mind weighs concepts, some abstracted from physical properties, others found inherent in formal relations (for example, mathematics) or in the method of pure reasoning. Above discursive reasoning Neoplatonists placed ’intellection’ (*noēsis*), which is contemplation. They saw it as pure thinking (*nous*) in the sense of grasping unmediated truths as a ’simultaneous whole’. For Plotinus this meant that the knower identifies with what is to be known. Others asserted that the two coincide (Proclus, *Platonic Theology* I, ch. 21, 97.22-98.5) so that we can still articulate propositions: perfect identity comes at a higher level. Grasping the ultimate truth, which has no differentiation, entails a complete unity, an ’illumination’. Plotinus described it as mental rapture, but Iamblichean Neoplatonists deduced that it is fully supra-intellectual, brought about by ’faith’ (*pistis*).

We assimilate knowledge with our *psychē* (soul, or the personal mind). It is what makes us alive and thinking. Not localized in a part of body or anywhere in the cosmos, it is as individual as the body it animates. *Psychē* exercises diverse activities, non-rational, vital, intellectual and spiritual. It can combine divine, intellectual inspiration and sensory perceptions into one mental object. Plotinus made *psychē* very flexible. The rest kept it between intellect and body, and sometimes distinguished its non-rational ’nature’ (*physis*).

Education, knowledge, science, ethics and psychology are thus grounded in metaphysics: the kind of existence that substantiates them. Degrees of learning, awareness or ’consciousness’ and virtue, reflect degrees of being.

Humans, plants, gods and every existent thing are made up of a ’bundle’ of the appropriate layers of being. In the basic scheme these are: physicality, soul, intellect, the One. Intellect (*nous*) is the objective, self-absorbed intelligence that creates by imposing ’form’. On a universal scale it is the Creator. The unqualified One is the supreme principle of all that exists in some sense: that is, the ultimate God. In itself it is beyond the system, transcending determination. Later Neoplatonists distinguished aspects within soul and intelligence. In this way, they arrived at a richer scheme: pure body, physicality, rational soul, intellect, life-principle (process), substantive essence (object-pattern) and the One. They also distinguished class properties as ’wholes’ and ’parts’. For example, the complete soul is an unanalysed whole, but its ’parts’ are the souls resident in individual living beings. This distinction helped specify ’unity’ (*henas*) as the ’root’ of the characteristic identity of a being, whether universal or particular. In later Neoplatonism (notably in Proclus) this is the spark of divinity.

The grades of being can be imagined as spheres of increasing comprehensiveness. Bodily existence is perfected by soul, this by the objective intellect, and so on. The One embraces everything. Conversely, there is a cascade of causation, from the One to the ’many’. Neoplatonists were still faced with problems: How can diversity emerge from undifferentiated unity? Why should a perfect state give rise to others? In response they proposed that such causation is not voluntary but a necessary consequence of perfection, which has no external limits (so divine action is not a mystery, unlike in the revealed religions). Existence ’unfolds’. This dispersion is like a light ’emanating’ from its source and diffusing into space, or as a ’procession’ of numbers from one to two, three, four…. Things (and levels) lose their foremost identity by gaining separate existence. They preserve the character that determines them by an act of ’returning’. However, Neoplatonists were not satisfied with the One being both the original cause of intelligible reality and transcending it, and to the end argued over solutions (as indeed cosmologists and theologians do to this day).

The problem also appeared at the level of soul. Why does *psychē* leave its own proper level of being and ’descend’ to a mortal body? Ontology requires an organizing principle to preserve the ever-disintegrating body. But something complete in itself cannot want something like the material body. Soul’s self-consciousness gives a sense of ’dare’ (*tolma*) that enables its departure and blind fall. Later Neoplatonists pointed out that perfect souls (for example, the moving and organizing principle of the cosmos) do not suffer such a fate. The ’particular’ souls are not completely aware what they are, and so leave their own order: either to ’fall’ to body, or to ’ascend’ towards the One. In this way the daring soul acquires a wide but personal experience.

4 Philosophy of science

The Neoplatonists taught the mathematical quadrivium of arithmetic, geometry, astronomy and harmonics, and the physics of Aristotle and Plato: Plato’s *Timaeus* belonged to the highest section of the curriculum. In the sixth century we also see Galen’s medicine appearing at the Alexandrian school, and we have a certain Stephanus of Athens writing as a philosopher-physician.

The programme of mathematization can be traced to Iamblichus’ reform (with the influence of Neo-Pythagoreans), although Pythagorizing mathematics is found in Plato himself. Part of the enterprise was the mathematization of physics and matter. Proclus and Simplicius declared that ’quantity’ is more fundamental than ’quality’. Iamblichus and later Neoplatonists were excited about mathematics because mathematical knowledge alone proceeds by the ’solidity of scientific demonstration’ (Elias, *On Porphyry’s Introduction* 28.24-9). Religious speculation and the physicist’s empiricism do not, and therefore need mathematics to support them. Mathematics thus became a pillar of Neoplatonic metaphysics and epistemology. Proclus pointed out that mathematical proofs rely on indemonstrable axioms, and so philosophy has to turn to a higher science - Platonic dialectic.

In ancient philosophy the concept of time related to change, movement and the psychological experience of events. Place was associated with the container where change happens and so with theories of matter. Neoplatonists stayed true to the Platonic outline, that time is the ’moving image of eternity’, but in fleshing it out they arrived at startling conclusions. For Plotinus, time is the product of psychological movement. It is the life of the soul as it moves from one activity to another ‘restlessly’. Since soul is not confined to individuals but is what gives life and movement to the universe, there is a cosmic time. By this account, eternity is the timelessness ’of that quiet life…which belongs to that which exists and is in being, all together and full, completely without extension or interval’ (*Enneads* III.7.3). The later Neoplatonists rejected Plotinus and followed Iamblichus in giving time its own reality. In this way they finally broke with Greek tradition. The time of ordinary experience is not true time

but the ’flowing’ appearance of it. Even the cosmic clocks, the stars and planets and the revolution of the universe as a whole, do not reveal what time is. Like the ’centre of a rotating wheel’ (Proclus, *Commentary on Plato’s Timaeus* 3.27), primary time is singular and motionless, and abides before creation. Damascius took Neoplatonic innovation further by proposing that ordinary time flows not by atomic instants but as a series of ’strides’ or ’leaps’ (each of which is a temporal part).

Neoplatonists rejected Aristotle’s view of place as a boundary. In developing instead Plato’s ‘receptacle’ of creation, they arrived at the full identification of place with three-dimensional space (see Space). Proclus equated space with massless body, and contemplated universal space as the absolute frame of cosmic motion. Others saw space as incorporeal. For the Christian Neoplatonist Philoponus, pure space is more fundamental than matter, and is a void. Iamblichus elevated place to the realm of Forms, and made space the instrument by which the world-soul binds things in their proper positions. For Damascius place really was the power that keeps the members of the universe in their arrangement. Simplicius returned to space as extension, and distinguished relative from absolute space, the latter being indistinguishable from essence. In examining the relationship of ’when’ and ’where’, he concluded that the two are ’siblings’, thereby anticipating the modern equal ranking of time and space.

5 Influence

Neoplatonists were influential through their philosophy, through their preservation of ancient Greek thought (for example, Simplicius is one of our two major sources for the Presocratics), and in their capacity as the educators of late antiquity. Their curriculum passed to the Byzantines, and a Latin fifth-century Neoplatonist, Martianus Capella, devised an influential allegory on the seven liberal arts. Indeed, it is hard to find thinkers of the early medieval period who were not influenced by Neoplatonism. In eleventh-century Byzantium, still find scholars arguing for (Psellus) or against ancient Neoplatonists, and so to the end in the fifteenth century (Plethon).

Theologians found it appealing that rational argument can sustain religion and a life of selfless goodness. However, several philosophical conclusions - for example, that God the transcendent is distinct from the Creator, involuntary divine action, the perpetuity of the physical world, the unimportance of accidental events (incarnation) - conflicted with religions based on the Bible. Moreover, the Neoplatonists tended to defend the Hellenic religion: Porphyry wrote an extensive scholarly refutation of Christianity. As a result Neoplatonist influence often went uncredited. A striking example of this was, the late fifth century theologian who wrote under the name ’Dionysius the Areopagite’ (see Pseudo-Dionysius). He produced an entire Christian theology, ’hierarchy’, by adapting the Athenian multi-layered system. Between the seventh and ninth centuries this was taken up by leading theologians of the Greek East and the Latin West. It is even claimed that a prototypical Gothic cathedral, St Denis, was built on Pseudo-Dionysian/Neoplatonic principles. Another example was a collection of works attributed to Aristotle but compiled from Plotinus and Proclus. They first occur in Arabic. Later they were translated into Latin. Both the Pseudo-Dionysian and pseudo-Aristotelian corpora were correctly identified by Aquinas.

Muslims came into contact with Neoplatonism with their first conquests of Syria and Egypt (seventh century). In the next three centuries they translated and paraphrased most of the Greek texts they found. With the Muslim expansion into Spain, the Arabic books became a source of Greek ideas for western Europeans. Some Neoplatonic texts still survive only in Arabic. Neoplatonism influenced Islamic philosophy and religion (Ismailism) (see Ibn Sina; Neoplatonism in Islamic philosophy).

Jewish thinkers had already participated in Platonism with Philo of Alexandria. Much later, living in the Muslim dominion, the eleventh century thinkers Gabirol (see Ibn Gabirol) and Gersonides developed Plotinian themes on intelligible matter and the active intellect. In the twelfth century, Neoplatonism was influencing the emerging Kabbalah in the south of France (see Kabbalah). Next we find it in Renaissance Kabbalism (see Bruno, G.; Pico della Mirandola, G.), and finally in the famous Dutch Jew, Spinoza.

Western Europe first inherited Plotinian-Porphyrian Neoplatonism through Augustine and Boethius, and later received Byzantine Neoplatonism. They influenced both mystics and logicians, for example, Duns Scotus, Anselm of Canterbury, Abelard, Eriugena, Albert the Great, Grosseteste and Thomas Aquinas. The Renaissance marked the recovery of Greek texts in the original, and a new wave of influence began (see Platonism, Renaissance). Proclus had been recognized since Aquinas’ time; now Plotinus (and Plato) became finally known. The Byzantine Neoplatonist Plethon was invited by the Medici to set up the Platonic Academy of Florence, where Ficino and Pico

della Mirandola taught. Their impact extended to European art, literature and landscape theory. The Hellenic rationalism of Neoplatonism appealed to the Humanists, particularly the idea that individuals possess their own means for salvation. This was also the message of Christian mystics such as Eckehart: even theologians such as Nicholas of Cusa upheld it. The late Neoplatonists' love for mathematics appealed to scientists who disliked medieval explanations by qualities. In art, the Neoplatonic light metaphysics and theories on mathematics and representation proved fruitful in the invention of perspective.

When Plato was separated from Platonists, (Neo)platonic influence declined. Notable exceptions were the Cambridge Platonists (see Cambridge Platonism) and the Idealists - both the British (see Berkeley, G.; Bradley, F.H.) and, especially, the Continental (see Bergson, H.-L.; Schelling, F.W.J. von; Hegel, G.W.F.). In science, Schelling and Hegel contributed to *Naturphilosophie*, which saw the forces of nature (magnetism, electricity, gravity) as the unfolding of one unified force (see Naturphilosophie). In literature, the English Romantics (Blake, Shelley) were influenced through Thomas Taylor. Coleridge was interested both in *Naturphilosophie* and, preoccupied with the origin of inspiration, in Schelling's investigations into the 'unconscious'.

See also: Hypatia; Kabbalah; Numenius; Renaissance philosophy

LUCAS SIORVANES

References and further reading

Anon., (6th-7th century AD) *Prolegomena to Platonic Philosophy*, ed. L.G. Westerink, Amsterdam: North-Holland, 1962; *Prolégomènes à la philosophie de Platon*, eds L.G. Westerink, J. Trouillard and A. Segonds, Paris: Les Belles Lettres, 1990.(1962 is parallel Greek text and English translation of a late Alexandrian work on the Neoplatonic curriculum. 1990 is a new edition of the Greek text with a parallel French translation and extensive annotation.)

Armstrong, A.H. (ed.) (1970) *Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press.(Excellent introduction, with chapters on links with Plato, on Alexander of Aphrodisias and on Christian thinkers.)

Blumenthal, H.J. (1996) *Aristotle and Neoplatonism in Late Antiquity: Interpretations of the 'De Anima'*, Ithaca, NY: Cornell University Press.(Key study on the assimilation of Aristotle in Neoplatonic psychology and metaphysics, and its influence.)

Bowersock, G.W. (1990) *Hellenism in Late Antiquity*, Ann Arbor, MI: University of Michigan Press.(Concise examination of the historical, religious and social context.)

Gersh, S. (1978) *From Iamblichus to Eriugena*, Leiden: Brill.(Detailed account of Neoplatonic metaphysics and the transition to medieval theology, including pseudo-Dionysius; major themes are illustrated with diagrams.)

Gersh, S. (1986) *Middle Platonism and Neoplatonism: The Latin Tradition*, Notre Dame, IN: University of Notre Dame Press.(Two source volumes for Latin Platonism from Cicero to medieval period; volume two covers Neoplatonism.)

Gersh, S. (ed.) (1992) *Platonism in Late Antiquity*. Notre Dame, IN: University of Notre Dame Press.(Articles mainly on metaphysics and religion.)

Haas, F.A. (1997) *John Philoponus' New Definition of Prime Matter: Aspects of its Background in Neoplatonism and the Ancient Commentary Tradition*, Leiden and New York: Brill.(New account of the theories of matter and space in the context of Neoplatonic thought.)

Lamberton, R. (1986) *Homer the Theologian: Neoplatonist Allegorical Reading and the Growth of the Epic Tradition*, Berkeley and Los Angeles, CA: University of California Press. (On the Neoplatonic theory of literary interpretation.)

Lindberg, D.C. (1986) 'The Genesis of Kepler's Theory of Light: Light Metaphysics from Plotinus to Kepler', *Osiris* 2: 5-42.(Definitive study on the influence of Neoplatonic light metaphysics in medieval and early modern science.)

Lloyd, A.C. (1990) *The Anatomy of Neoplatonism*, Oxford: Clarendon Press.(The best analytical account of Neoplatonic logic and metaphysics; for the advanced philosophy reader.)

O'Meara, D.J. (1982) *Pythagoras Revived: Mathematics and Philosophy in Late Antiquity*, Oxford: Clarendon Press.(Key study on later Neoplatonic metaphysics and theory of knowledge.)

Saffrey, H.D. (1990) *Recherches sur le néoplatonisme aprés Plotin (Studies of Neoplatonism After Plotinus)*, Paris: Vrin.(Collected articles about Porphyry, Iamblichus and Proclus, on theological themes, by a leading scholar in the field.)

Siorvanes, L. (1996) *Proclus: Neoplatonic Philosophy and Science*, Edinburgh and New Haven, CT: Edinburgh

and Yale University Presses.(Covers Neoplatonism to Philoponus and Simplicius, suitable for newcomers.)
Steel, C. (1978) *The Changing Self: A Study on the Soul in Later Neoplatonism: Iamblichus, Damascius, Simplicius*. Brussels: Royal Academy.(Seminal study of late Neoplatonic psychology.)
Wallis, R.T. (1972) *Neoplatonism*, London: Duckworth.(Compact but thorough account suitable for a newcomer to the subject.)
Westerink, L.G. (1964) 'Philosophy and Medicine in Late Antiquity', *Janus* 51: 169-77.(On medical interests, primarily at the late Alexandrian school.)

Neoplatonism in Islamic philosophy

Islamic Neoplatonism developed in a milieu already saturated with the thought of Plotinus and Aristotle. The former studied in Alexandria, and the Alexandrine philosophical syllabus included such figures as Porphyry of Tyre and Proclus. Associated with these scholars were two major channels of Islamic Neoplatonism, the so-called Theology of Aristotle and the Liber de Causis (Book of Causes). Other cities beloved of the philosophers at the time of the rise of Islam in the first century AH *(seventh century* AD*) included Gondeshapur and Harran.*

Islamic Neoplatonism stressed one aspect of the Qur'anic God, the transcendent, and ignored another, the creative. For the Neoplatonists, all things emanated from the deity. Islamic philosophers were imbued to a greater or lesser degree with either Aristotelianism or Neoplatonism or, as was often the case, with both. Al-Kindi, the father of Islamic philosophy, has a Neoplatonic aspect, but the doctrine reaches its intellectual fruition in the complex emanationist hierarchies developed by al-Farabi and Ibn Sina. Their views are later developed (or metamorphosed) by later thinkers into an emanative hierarchy of lights, as with Shihab al-Din al-Suhrawardi, or the doctrine of the Unity of Being espoused by Ibn al-'Arabi. While al-Ghazali and Ibn Rushd both vigorously opposed Neoplatonic views, the latter attacked the former for his general opposition to the philosophers.

Neoplatonism itself had a major impact on that sectarian grouping of Muslims known as the Isma'ilis, and became the substratum for its theology. Historically, Neoplatonism in Islam achieved its climax with the Fatimid Isma'ili conquest of Egypt towards the end of the fourth century AH *(tenth century* AD*). While Neoplatonism later declined in philosophical importance in the face of rampant Aristotelianism and Hanbalism, it may be said to have bequeathed an important religious, historical and cultural legacy to the Islamic world, which in the Isma'ili movement endures to this day.*

1 Milieu and sources

Islamic Neoplatonism developed in a milieu which was familiar with the doctrines and teachings of Plotinus. The city of Alexandria, into which the Arab armies of Islam marched in AD 642, had down the centuries been home to many philosophies and philosophers: Plotinus himself, the founding father of Neoplatonism, studied in Alexandria for eleven years under the scholar Ammonius Hierocles. The Alexandrian philosophical syllabus was imbued with Neoplatonism and coated with Aristotelianism. The works of important Neoplatonists such as Porphyry and Proclus were studied there. Two works, whose exact authorship is unclear but which became associated with Porphyry and Proclus respectively, were the famous *Theology of Aristotle* and the work which became Latinized as the *Liber de causis* (see Liber de causis). Both these works, regardless of their actual authorship, were major channels of Islamic Neoplatonism. The *Theology of Aristotle*, despite its name, had nothing to do with Aristotle but summarized, with some additions, Books IV-VI of Plotinus' *Enneads*. The *Liber de causis (Book of Causes)* had its basis in the *Elements of Theology* by Proclus. The Neoplatonic themes in both the *Theologia* and the *Liber* are not difficult to identify, ranging from the key doctrine of emanation through references to the hypostases such as the Universal Intellect and Universal Soul (*Theology of Aristotle*) or the Procline hypostases of the One, Existence, Intellect and Soul (*Liber de causis*), to the sublime attributes of the One (see Neoplatonism).

However, Alexandria was not the only major city in the Middle East to foster the rise of Neoplatonism before the rise of Islam. Another was Gondeshapur, a great centre of Greek Byzantine learning, especially in the fields of philosophy and medicine, where Aramaic rather than Persian appears to have been the dominant language. This city, built by Shapur I in the mid-third century AD, acted as a magnet to many Middle Eastern intellectuals in both the pre-Islamic and Islamic periods. The great father of the Arabic translation movement, Hunayn ibn Ishaq, studied there; earlier, Nestorian scholars had fled to that city after the Council of Ephesus in AD 431. These scholars knew the work of Aristotle, but they had also studied Porphyry and so were familiar with the teachings of Neoplatonism. The closeness of Gondeshapur to what became Baghdad meant that the former city was able to infiltrate the latter, when it became an Islamic seat, with a variety of Greek elements.

Then there was Harran in northern Syria, a city which was home to the star-loving Sabaeans, a pagan sect whose transcendent theology was imbued with Neoplatonic elements. In the third century AH (ninth century AD) Harran was visited by refugee scholars from the schools of Alexandria; in the following century these scholars moved from Harran to Baghdad, bringing to that last city elements, both Aristotelian and Neoplatonic, from the rich

philosophical heritage of both Alexandria and Harran. Of course, cities such as Alexandria, Gondeshapur and Harran were not the only sources of Neoplatonic thought, but their examples serve to illustrate the ease with which the expanding Arab-Islamic Empire came into contact with Greek thought, especially in its Aristotelian and Neoplatonic incarnations. And it was between the latter that the pendulum of Islamic philosophy frequently swung in the writings of the individual Islamic philosophers, when they were not actually mixing the two in a glorious intellectual syncretism as happened with the thought of the Ikhwan al-Safa' (see Ikhwan al-Safa').

2 The God of Islamic Neoplatonism

The description of God in the Qur'an is by and large fairly clear, though it did give rise to complexities of interpretation in Islamic theology centring on such matters as anthropomorphism, God's omnipotence and man's free will, and the attributes of God. The Qur'anic God, however, is both transcendent and immanent. There is none like him but he is also closer to man than man's jugular vein. He intervenes in human history to reveal himself to man, for example in the revelations of the Qur'an, and sends angels to fight for his prophet Muhammad, as at the Battle of Badr in AD 624. Here he is often like the God beloved of today's process theologians (see Process theism). Above all, however, the Qur'anic God is one who creates *ex nihilo*. There is no concept of Neoplatonic emanation in the Qur'an. In contrast, for Aristotle, God is the Unmoved First Mover. The emphasis in Aristotelian theology is much more on God's movement rather than on his creation, which is limited in any case to his production of form in prime matter which has existed eternally. With the Neoplatonists, the emphasis moves from the concept of creation to that of eternal emanation: God or the One or the Good - however he is to be characterized - does not create *ex nihilo* but 'engages' in eternal emanation of all that is below him.

Thus in the Middle East at the time of the rise and spread of Islam there were at least three different 'theologies' vying for space, emphasising different qualities of their deity. There is the Qur'anic God as creator *ex nihilo*; there is the Aristotelian God as first Mover; and there is the Neoplatonic God as eternal emanator. The debate which was engendered about the relationship between God and the rest of observable and intangible reality and phenomena became a fundamental characteristic of the writings of the Islamic philosophers. The Qur'anic God was linked to his creation by the sheer power of creativity, the Aristotelian God was linked - much less feelingly - with that which moved, while the Neoplatonic God bridged, or attempted to bridge, the huge gulf between transcendence and corporeal reality by the device of emanation. A brief survey of the thought of some individual Islamic philosophers will serve to illustrate how the debate featured in their writings, and thus in the general development of Islamic philosophy.

Abu Yusuf ibn Ishaq al-Kindi is universally acknowledged by scholars of Islamic philosophy as the 'Father of Islamic Philosophy'. Al-Kindi's God has four faces or aspects. Doctrinally, he is classically rooted in and derived from the Qur'an, and bears such epithets as 'creator' and 'active'. God has an essential unity which does not derive from anything else. He also has Aristotelian aspects - he is, for example, unmoved - but of course al-Kindi's deity is much more than a mere Mover. God's attributes are also discussed by al-Kindi in Mu'tazilite terms and al-Kindi espouses a Mu'tazilite antipathy towards anthropomorphism (see Ash'ariyya and Mu'tazila §4). Finally, we can detect a Neoplatonic influence on al-Kindi's thought. He was the first major Islamic philosopher to reflect significant aspects of the Neoplatonic tradition, and is a bridge to the thought of philosophers such as al-Farabi and Ibn Sina.

It is with these latter two philosophers that Islamic Neoplatonism reaches its apotheosis, where such fundamental Neoplatonic concepts as hierarchy and emanation are fully developed and integrated into a metaphysics of being. Al-Farabi is rightly regarded as the father and founder of Islamic Neoplatonism, while Ibn Sina, though less original, is often considered to be Islam's greatest Neoplatonic philosopher. While the deity that he portrays certainly has other aspects, it is the Neoplatonic aspects which draw our attention. Like al-Farabi, Ibn Sina has a complex scheme of emanation with ten intellects emanating from the Necessary Being. Again as with al-Farabi, emanation constitutes a bridge between the unknowable God of Neoplatonism and earthbound humanity. However, the theological terminology deployed in Ibn Sina's thought is perhaps less negative than that of al-Farabi; this is particularly true of the mystical dimension of Ibn Sina's thinking.

Neoplatonism in Islam may be said to have reached its furthest limits of development in the thought of Isma'ili theologians such as Hamid al-Din al-Kirmani on the one hand, and that of Shihab al-Din Yahya al-Suhrawardi and Muhyi al-Din Ibn al-'Arabi on the other. Al-Kirmani espouses a Farabian elaboration of God and ten intellects in

his Neoplatonic emanationist hierarchy. Al-Suhrawardi, 'the Master of Illumination' (*shaykh al-ishraq*), as he became known, established an extraordinary complex Neoplatonic hierarchy of lights in which the divine and quasi-divine are seen all in terms of light. God is the Light of Lights (*nur al-anwar*), and from him emanates the First Light from which emanates the Second Light and so on; but bound into the whole system is a complex three-tier system of Angelic Lights. Because of the doctrine of emanation, the lights (or intellects) have an ontological or noetic precedence, the one over the other, but not a temporal precedence. By contrast, Ibn al-'Arabi employs Neoplatonic terminology to bolster his doctrine of the Unity of Being (*wahdat al-wujud*). The circularity of his thought, however, precludes the elaboration of a classical system of emanation following Plotinian, or even Farabian, lines. It may be argued that the terms 'theophanies' or 'manifestations' (*tajalliyat*) of the divinity, rather than 'emanations', are a more accurate rendering of his thought.

3 Reaction and counter-reaction

The reaction and counter-reaction to the infiltration of Neoplatonism into Islamic thought and philosophy may usefully be studied in the writings of the great Abu Hamid al-Ghazali, Sunni theologian and mystic, on the one hand and those of Islam's most notable Aristotelian, Abu'l Walid Muhammad ibn Rushd, also known as Averroes, on the other. In his *Tahafut al-falasifa (Incoherence of the Philosophers)*, al-Ghazali (§3) attacked both the Neoplatonists and Aristotle. He rebutted, for example, the idea that the world was eternal, and tried to show mathematically that the thesis of the Neoplatonists was illogical. He believed that the Neoplatonists had failed to prove that God was One, and attacked their beliefs about a variety of other fundamental and crucial points such as divine knowledge and the question of the immutability of God. The two principal philosophers whose views al-Ghazali attacked were al-Farabi and Ibn Sina. As al-Ghazali himself put it:

> However, the most faithful - as Aristotle's translators - and the most original - as his commentators - among the philosophizing Muslims are al-Farabi Abu Nasr, and Ibn Sina. Therefore, we will confine our attention to what these two have taken to be the authentic expression of the views of their mis-leaders…. Therefore, let it be known that we propose to concentrate on the refutation of philosophical thought as it emerges from the writings of these two persons.
>
> (Tahafut al-falasifa: 5)

In all, al-Ghazali itemized twenty particular problems 'in whose discussion in this book we will expose the contradiction involved in the philosopher's theories'.

If al-Ghazali represents Islamic theology's most biting attack on philosophy and severest reaction to Neoplatonism in Islam, Ibn Rushd represents the counter-reaction. This is not to say that the latter wholeheartedly espoused the views of the Neoplatonists: indeed, very far from it. In his *Tahafut al-tahafut (Incoherence of the Incoherence)*, referring to the incoherence of al-Ghazali's *Tahafut al-falasifa*, Ibn Rushd wrote a work which Fakhry has described as 'the product of Ibn Rushd's maturest thought [which] constitutes a systematic rebuttal of al-Ghazali's critique of Greco-Arab philosophy' (Fakhry 1983: 276). Al-Ghazali is accused of misunderstanding, and it is clear that Ibn Rushd is concerned to defend the merits of philosophy as a mode of non-heretical thought while at the same time not accepting the theses of the Neoplatonist philosophers. Despite his intention in the *Tahafut al-tahafut* of defending the philosophical targets of al-Ghazali's wrath, Ibn Rushd, as Bello points out, 'more often than not…does not, in fact, defend al-Farabi and Ibn Sina…. Instead, he shows to what extent they have departed from the authentic Aristotelian philosophical doctrines, and sometimes joins his voice with that of Ghazali in convicting them of heresy' (Bello 1989: 15). Thus Ibn Rushd, despite his defence of philosophy and philosophers, is more than happy to declare open war on Neoplatonism.

4 The influence and legacy of Islamic Neoplatonism

While it is certainly untrue to say that Islamic philosophy came to a sudden end with the death of Ibn Rushd, we can say that his death in AH 595/AD 1198 marks the approaching end of the great debates about Neoplatonism in Islamic thought. By then the kind of peripateticism espoused by Ibn Rushd may be said to have at least revived, if not definitely triumphed over, other forms of philosophy. The death of Ibn al-'Arabi in AH 638/AD 1240 marks that triumph, for the latter's doctrine of *wahdat al-wujud* was perpetuated by only a few faithful disciples. Furthermore, other movements had arisen in competition with Neoplatonism and Aristotelianism, such as the literalism of the Spanish Muslim Ibn Hazm, and the Hanbalism of Ibn Taymiyya. Neoplatonism as a radical system of

philosophical thought with a controversial theological agenda was enshrined in the writings of such thinkers as the Ikhwan al-Safa’, but generally speaking its greatest surviving influence was, and is, on the theology of the Isma‘ili sect in Islam, one of the three great divisions of Shi‘ism. This sect achieved its political apotheosis with the coming to power of the Fatimid Isma‘ili dynasty in North Africa and Egypt in the fourth and fifth centuries AH (tenth and eleventh centuries AD). The Mosque-University of al-Azhar was a beacon of Isma‘ili (and thus Neoplatonic) thought before the Ayyubids took possession of Egypt and returned it to the Sunni fold. Today, Neoplatonism in Islam survives principally as the philosophical substratum which underpins the theology of the Isma‘ilis, a group which, though itself split over issues of leadership, nonetheless holds many theological and philosophical points in common.

If we examine the impact of Neoplatonism on Islamic thought generally, it is clear that this philosophy served to emphasize that transcendent aspect of God which is to be found clearly in the Qur’an, sometimes at the expense of the immanent. The impact of Neoplatonism on the course of Islamic history itself has been considerable in some regions. Among many examples we may note that the Fatimid dynasty came to power in Egypt and ruled there from AH 297-567 (AD 969-1171); the Isma‘ili Assassins flourished at the Castle of Alamut from AH 483-654 (AD 1090-1256); and a Nizari Isma‘ili imamate later moved from Persia to India. Theologically then, it is clear that a body of doctrines, so many of which seemed at odds with mainstream Islamic teaching, served at times to highlight the Qur’anic emphasis on transcendence and was actually absorbed by one, albeit heterodox, sect to become the foundation for that sect; while historically, Neoplatonism from its mainly theoretical Middle Eastern origins in Alexandria, Harran, Gondeshapur and elsewhere became sufficiently powerful to ‘hijack’ an entire dynasty, the Fatimid.

We may extrapolate from all this, then, the paradigm of an ‘alien’ cult which becomes ‘Middle-Easternized’ and ‘Islamicized’ and which acts on occasion as theological stimulus, irritant, gadfly or foundation, and in so doing ultimately inserts itself from a variety of perspectives into the broad and multivalent fabric of Islam. Alternatively, we may choose to examine the phenomenon of Neoplatonism rather more closely, assess its emphasis on order, structure, emanation, hierarchy, transcendence, intellect and soul, and extrapolate a rather different paradigm, perhaps more akin to that preferred by the Isma‘ilis. According to this view, Neoplatonism would not be regarded as a foreign or invasive growth within the body politic of Islam but rather as something which, despite its emphasis on emanation rather than *creatio ex nihilo* and other real differences from mainstream Islamic theology, addressed an aspect or aspects of Islam which had been neglected or overlaid by other matters in the development of that faith. It is useful perhaps to ponder Lenn and Madeleine Goodman’s observation: ‘Emanation was perfected by the neo-Platonists, quite consciously as an alternative to creation because the learned neo-Platonic philosophers did not choose to redescend into the anthropomorphic cosmogenies from which Aristotle had rescued them with great difficulty only a few centuries earlier’ (Goodman and Goodman 1983: 31).

See also: Aristotelianism in Islamic philosophy; al-Farabi; al-Ghazali; Greek philosophy: impact on Islamic philosophy; Ibn Rushd; Ibn Sina; Illuminationist philosophy; Neoplatonism

IAN RICHARD NETTON

References and further reading

Bello, I.A. (1989) *The Medieval Islamic Controversy Between Philosophy and Orthodoxy*, Islamic Philosophy and Theology Texts and Studies vol. III, Leiden: Brill. (Deals with the conflict between al-Ghazali and Ibn Rushd with special reference to *ijma‘* and *ta’wil*.)

Fakhry, M. (1983) *A History of Islamic Philosophy*, 2nd edn, London: Longmans; New York: Columbia University Press.(A superb introduction to the whole field.)

al-Ghazali (1058-1111) *Tahafut al-falasifa (Incoherence of the Philosophers)*, English trans. S.A. Kamali, Lahore: Pakistan Philosophical Congress, 1963.(A translation of al-Ghazali’s attack on the philosophers.)

Goodman, L.E. and Goodman, M.J. (1983) ‘Creation and Evolution: Another Round in an Ancient Struggle’, *Zygon* 18 (1): 3-43.(A fascinating and thought-provoking article.)

Henry, P. and Schwyzer, H.R. (eds) (1959) *Plotini Opera*, vol. 2, Paris: Desclée de Brouwer; Brussels: L’Édition Universelle.(Contains G. Lewis’ English translation of the *Theology of Aristotle*.)

Ibn Rushd (*c.*1180) *Tahafut al-tahafut (Incoherence of the Incoherence)*, trans. and intro. S. Van Den Bergh, *Averroes’ Tahafut al-Tahafut (The Incoherence of the Incoherence)*, ‘E.J.W. Gibb Memorial’ new series XIX, London: Luzac, 1978.(Ibn Rushd’s famous response to al-Ghazali.)

Leaman, O. (1988) *Averroes and his Philosophy*, Oxford: Clarendon Press; 2nd edn, Richmond: Curzon, 1997. (An introduction to the philosophy of Ibn Rushd arranged according to metaphysics, practical philosophy, and reason, religion and language.)

Nanji, A. (1996) ‘Isma‘ili Philosophy’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 9, 144-54.(Examination of Isma‘ili philosophy including the influence of Neoplatonism.)

Netton, I.R. (1989) *Allah Transcendent: Studies in the Structure and Semiotics of Islamic Philosophy, Theology and Cosmology*, London and New York: Routledge. (Contains major chapters on al-Farabi and Ibn Sina.)

Shayegan, Y. (1996) ‘The Transmission of Greek Philosophy into the Islamic World’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 6, 98-104.(Detailed account of how the transmission took place, paying particular attention to the Persian background.)

Neo-Pythagoreanism

Neo-Pythagoreanism is a term used by modern scholars to refer to the revival of Pythagorean philosophy and way of life in the first century BC*. It coincides with the redevelopment of Platonic thought known as Middle Platonism. Neo-Pythagoreans elaborated a mathematical metaphysics in which the highest level of being was occupied by a transcendent principle, equated with 'the One' or 'the Monad' and regarded as the source of all reality. Neo-Pythagorean anthropology reaffirmed the ancient Pythagorean belief in the immortality of the soul. Although Neo-Pythagoreanism is often indistinguishable from Middle Platonism, it is characterized by a tendency to see Pythagoras as the father of all true philosophers, including Plato. In the third century* AD *Neo-Pythagoreanism was absorbed into Neoplatonism.*

As the name implies, Neo-Pythagoreanism is a renewal of Pythagoreanism (see Pythagoreanism). Since the reanimators were philosophers of a Platonic stamp, it is particularly the kind of Pythagoreanism found in Plato and the early Academy that came to be renewed and developed further. Plato incorporated Pythagorean number theory in the *Timaeus* (see Plato §16) to provide a mathematical structure for the universe. In the so-called 'Unwritten Doctrines' he posited two first principles, the One (or Monad) and the Indefinite Dyad, essentially developments of the Pythagorean primal opposites, Limited and Unlimited (see Philolaus §2; Plato §5; Platonism, Early and Middle §2-4; Pythagoreanism §2). The One, identified with the Good, is the determining principle, which acts upon the Dyad (the Two), the passive principle. Their interaction generates numbers that are equated with the Platonic Forms (the Form-numbers), of which the physical world is but a reflection. The One and the Dyad were taken over, with certain modifications, by Plato's immediate successors, Speusippus (§2) and Xenocrates (§2). In the subsequent history of Platonism, until the first century BC, Pythagorean metaphysical speculation went into abeyance, notably during the sceptical phase of the Academy when all forms of dogmatism, which would include the positing of mathematicals as the eternal and invariable archetypes of reality, were rejected (see Academy; Arcesilaus §2).

Although the Hellenistic age marks a fallow time for the serious pursuit of Pythagoreanism, it was a period remarkably productive of Pythagorean texts, known as pseudepigrapha. Their authors either preferred to remain anonymous or purported to be older Pythagoreans. Some of these compositions are religious-didactic, some philosophical but with little original inspiration. None the less, their concern to show that the teachings of Plato and Aristotle were of ancient Pythagorean provenance found great resonance with Neo-Pythagoreans, and several pseudepigrapha were considered authoritative texts.

Pythagorean philosophy resurfaced in earnest in the early first century BC after Antiochus issued his call 'to follow the ancients'. In the course of a renewed interest in the original works of Plato, the *Timaeus* gained particular prominence, a dialogue that now came to be seen virtually as a Pythagorean tract. A notable commentator on the *Timaeus* was Eudorus of Alexandria. His first principles are Monad and Dyad, but above these he posits a supreme One, also called the Supreme God. 'Becoming like God' was the goal of life. While Eudorus is a Platonist, he presents his doctrines as 'Pythagorean' and thus counts as an important witness to the 'Pythagoreanizing' current in Platonism (see Platonism, Early and Middle §§4,8-9).

The first *bona fide* Neo-Pythagorean was Moderatus of Gades. In his *Pythagorean Lectures* he establishes a system of three Ones: a One beyond being, a One in the realm of the Ideas that acts as a divine craftsman (Demiurge) upon the Dyad, bringing matter into existence, and a One that participates in the first two and permeates the world (the World-Soul). Moderatus' triad foreshadows the three *hypostases* - One, Intellect, Soul - of Neoplatonism (§3).

Nicomachus of Gerasa has a similar triad. His first God, the Monad, appears, however, to be identified with the Demiurge, who in combination with the Dyad produces the Logos - or the Forms, equated in accentuated Pythagorean-Platonic fashion with mathematicals (see Logos §1). Much of Nicomachus' philosophy must be extracted from his *Theological Arithmetic*, a work in which he identifies the traditional Hellenic gods with numbers. This number-mysticism excited the interest of later Neoplatonists, especially Iamblichus.

With Numenius of Apamea (Syria) we reach the zenith of Neo-Pythagorean thought. While a Platonist from our

viewpoint, Numenius put Pythagoras, Socrates and Plato on an equal footing and was generally regarded as a Pythagorean in antiquity. He also proclaimed three gods, but in his attempts to separate the Monad, the absolute Good, from all association with creative activity and matter (the Dyad for Numenius is co-eternal with the Monad and embodies an evil principle), his philosophy assumes a starkly dualistic character, reminiscent of ancient Pythagoreanism.

In addition to metaphysics, Neo-Pythagorean philosophers reaffirmed the immortality of the soul and transmigration (see Pythagoreanism §3). There are also indications that, along with the renewal of Pythagorean philosophy, came a revival of the Pythagorean lifestyle with its emphasis on spiritual and bodily purity. This is evidenced especially in Roman Pythagoreanism. Cicero credits one Nigidius Figulus with renewing an extinct Pythagoreanism. Although we know nothing of the speculative thought of Figulus, Cicero's testimony accords well with the general picture of a Pythagorean reawakening in the first century BC, which in the Roman world undoubtedly took a very practical form. In the next century, under the emperor Augustus, a circle of Pythagorean philosophers formed around Quintus Sextius, the 'Sextii', whose ethical precepts had considerable impact on Seneca (§1). The emperor Tiberius employed a Pythagorean court-philosopher named Thrasyllus.

After flourishing in the hands of Numenius, Neo-Pythagoreanism came to an end with the waning of the second century AD. But with Middle Platonism, with which it had always gone hand in hand, it lived on as major force in Neoplatonism. Plotinus, Porphyry, Iamblichus, Proclus, to name just the most famous Neoplatonists, not only wrote on Pythagorean themes (we owe two of our extant biographies of Pythagoras to Porphyry and Iamblichus) but also were indelibly marked by the philosophy that bore the name of Pythagoras.

See also: Orphism; Platonism, Early and Middle §1

HERMANN S. SCHIBLI

References and further reading

Armstrong, A.H. (ed.) (1970) *The Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press.(In chapter 5 Merlan briefly discusses the Pythagorean pseudepigrapha and several Neo-Pythagoreans, among them Moderatus, Nicomachus and Numenius; for a more comprehensive treatment of these authors see Dillon (1977).)

Dillon, J.M. (1977) *The Middle Platonists: A Study of Platonism*, 80 BC to AD 220, London: Duckworth.(Chapter 1 includes valuable pointers to Pythagorean currents in Plato, Speusippus and Xenocrates; chapter 7 covers the individual Neo-Pythagoreans against the background of Middle Platonic thought.)

Guthrie, K.S. (1987) *The Pythagorean Sourcebook and Library*, Grand Rapids, MI: Phanes Press.(Contains English translations of many of the pseudo-Pythagorean texts, including those that were highly influential on Neo-Pythagoreanism, such as the fragments of pseudo-Philolaus and pseudo-Archytas, *On the Nature of the Universe* by Ocellus Lucanus, *On the World and the Soul* by Timaeus of Locri, and the *Golden Verses of Pythagoras*.)

Kingsley, P. (1995) *Ancient Philosophy, Mystery, and Magic: Empedocles and Pythagorean Tradition*, Oxford: Oxford University Press.(In chapter 20 Kingsley examines the later history of Pythagoreanism, arguing that Neo-Pythagoreanism was not so much a revival as a continuation of early Pythagorean traditions that were kept alive in small circles of practising Pythagoreans.)

O'Meara, D.J. (1989) *Pythagoras Revived: Mathematics and Philosophy in Late Antiquity*, Oxford: Oxford University Press.(A short introduction to Neo-Pythagoreanism is followed by an account of the Neoplatonic adaptation of Pythagorean philosophy, particularly the mathematization of reality, by Iamblichus, Hierocles, Syrianus and Proclus.)

Reale, G. (1990) *A History of Ancient Philosophy IV: The Schools of the Imperial Age*, trans. J.R. Catan, Albany, NY: State University of New York Press.(Pages 237-72 discuss the revival of Pythagoreanism, stressing the incorporeal and transcendental aspects of Neo-Pythagoreanism.)

Thesleff, H. (1961) *An Introduction to the Pythagorean Writings of the Hellenistic Period*, Åbo Finland: Åbo Akademi.(A scholarly account of the Pythagorean pseudepigrapha and their sources. Although the dating of these texts is a subject of controversy, Thesleff's discussion of the genre remains useful.)

Neumann, John von (1903-57)

Von Neumann was one of the great mathematical minds of the twentieth century. His work has affected philosophy on several fronts, including logic and the philosophy of science. He also had great influence upon developments in the philosophy of mind: the computer model of mind employed during the middle-to-late twentieth century was explicitly based upon the von Neumann computer architecture. Although late twentieth-century philosophy of mind has largely rejected the von Neumann machine as a model of brain activity, his pioneering work in cellular automata has provided a basis for subsequent development in 'distributed' or 'connectionist' computer architectures.

John von Neumann was described by mathematician Jacob Bronowski as 'the cleverest man I ever knew, without exception. And he was a genius, in the sense that a genius is a man who has *two* great ideas'. But von Neumann was even cleverer than that, doing foundational work in logic and set theory, quantum mechanics, computer design and the theory of computation, and game theory, to mention only those contributions of immediate interest to philosophers. In pure mathematics, von Neumann also produced important work in measure theory and the spectral theory of operators in Hilbert space, as well as introducing the theory of 'rings of operators' (now known as von Neumann algebras). In applied mathematics, he did groundbreaking work in numerical analysis in connection with fluid dynamics, the theory of turbulence and the physics of shock waves.

Born in Budapest, Hungary, von Neumann was educated at the Lutheran *Gymnasium* and was then tutored privately in mathematics when his extraordinary abilities became apparent. Though enrolled at the University of Budapest, von Neumann appeared there only to take exams, instead spending his time in Germany and Switzerland studying physics and chemistry and visiting David Hilbert in Göttingen to discuss mathematics. He took a chemical engineering degree at the Federal Institute of Technology in Zürich in 1925 and a doctorate in mathematics at Budapest in 1926 (his dissertation concerned the axiomatization of set theory). He was named *Privatdozent* at the University of Berlin (1927-9), and then at Hamburg (1929-30). In 1930-1 he became a visiting lecturer at Princeton University in New Jersey, and subsequently professor of mathematics there until he joined the Institute for Advanced Study (also at Princeton) in 1933, becoming its youngest permanent member. During the Second World War, von Neumann participated in various war-related scientific enterprises for the USA, in particular the atomic bomb project at Los Alamos. In 1954 von Neumann was appointed to the USA's Atomic Energy Commission. He died of bone cancer in Washington, DC in 1957.

Among von Neumann's accomplishments in set theory (see Set theory) are his elegant treatment of ordinal numbers (in which each ordinal number is the set of all smaller ordinal numbers), the first rigorous treatment of definition by transfinite induction, and an axiomatization of set theory more technically intricate than that of Ernst Zermelo and Abraham Fraenkel. This axiomatization and its successors have turned out to be useful in mathematical research. With respect to mathematical logic, von Neumann participated in Hilbert's programme in metamathematics, publishing proofs of consistency for parts of arithmetic until Kurt Gödel's 1931 results made such activity moot (see Gödel's theorems).

In quantum theory, von Neumann provided the first rigorous axiomatization of quantum mechanics, contributed to measure theory and proved an ergodic theorem for quantum systems. His *Mathematische Grundlagen der Quantenmechanik (Mathematical Foundations of Quantum Mechanics*, 1932) gives his argument that no introduction of 'hidden parameters' can eliminate the statistical character of statements of quantum theory unless the theory itself is fundamentally changed (see Quantum mechanics, interpretation of).

Von Neumann introduced the theory of games in a 1928 paper, providing a mathematical description of games and strategy and stating and proving the minimax theorem (see Decision and game theory). (Game theory is also known as 'decision theory' because a game in von Neumann's sense includes bluffing and, in selecting a move, taking into account what you think your opponent thinks you are going to do.) His later collaboration with Oskar Morgenstern (1944) extended game theory to mathematical economics.

Computer design and theory of computation are the areas of von Neumann's work which have had the most widespread influence among philosophers. Von Neumann's computer design was influential for two main reasons:

the machine architecture appears ‘brain-like’ at an abstract level (or so many philosophers thought); and the design could be implemented using high-speed components available at the time (resulting in an actual computer which could perform impressive computations).

A von Neumann machine consists of three primary components (a ‘central control unit’, a ‘central arithmetic unit’ and a finite uniform ‘memory’) and input and output ‘organs’. (The analogy with the human body is discussed in von Neumann’s ‘First Draft of a Report on the EDVAC’ (1945).) The central control unit interprets instructions while the central arithmetic unit executes operations on data contained in the memory; a crucial feature of a von Neumann machine is that the instructions are also stored in the uniform memory. Thus the functional status of the contents of any particular memory location is indeterminate: if it is executed, it is an instruction; if it is operated upon, it is data. A von Neumann machine exploits this indeterminacy: the machine can modify its own program during execution. This feature both allows the high-speed implementation of basic programming constructs and explains why the von Neumann architecture was regarded as ‘brain-like’.

Subsequent developments in artificial intelligence and neuroscience, however, have led to the rejection of the von Neumann machine as a model of brain activity. Basic cognitive processes (for example, pattern recognition) are computation-intensive, and neurons are too slow (and too unreliable) to complete them accurately in real time given a von Neumann computer organization. If the brain is a digital computer, it must have a non-von Neumann architecture.

Interestingly, in the 1940s von Neumann was already working on developing a general theory of natural and artificial automata; his particular areas of concern were self-reproducing automata and the synthesis of reliable systems from unreliable components. His early work on McCulloch-Pitts ‘neural networks’ (see McCulloch and Pitts 1943) made von Neumann himself a pioneer in non-von Neumann computer architectures.

See also: Logical and mathematical terms, glossary of

BRIAN ROSMAITA

List of works

Neumann, J. von (1961-3) *The Collected Works of John von Neumann*, ed. A.H. Taub, Oxford and New York: Pergamon Press, 6 vols.(Organized chronologically by topic, with an exhaustive bibliography. Of particular interest to philosophers: papers on logic, set theory and quantum mechanics in volume 1; computer design and automata theory in volume 5; and game theory in volume 6.)

Neumann, J. von (1986) *Papers of John von Neumann on Computing and Computer Theory*, ed. W. Aspray and A. Burks, The Charles Babbage Institute Reprint Series for the History of Computing, vol. 12, Cambridge, MA: MIT Press.(The best introductory source for von Neumann’s writings on computing; includes relevant papers from the *Collected Works*, including the 1945 and 1946 papers on von Neumann architecture, some papers not in that collection, and the first part of von Neumann (1966), with useful introductions by A. Burks. Includes a bibliography of von Neumann’s publications in all subjects, and a convenient bibliography of computer research contemporary with von Neumann and building on his work.)

Neumann, J. von (1928) ‘Zur Theorie der Gesellschaftsspiele’, *Mathematische Annalen* 100: 295-320; trans. S. Bargmann, ‘On the Theory of Games and Strategy’, in *Contributions to the Theory of Games*, vol. 4, Princeton, NJ: Princeton University Press, 1959, 13-42.(The paper in which von Neumann invented game theory; includes the first statement and proof of the minimax theorem.)

Neumann, J. von (1932) *Mathematische Grundlagen der Quantenmechanik*, Berlin: Springer; trans. R. Beyer, *Mathematical Foundations of Quantum Mechanics*, Princeton, NJ: Princeton University Press, 1955; repr. in the Princeton Landmarks in Mathematics and Physics series, Princeton, NJ: Princeton University Press, 1983. (Von Neumann’s axiomatization of quantum mechanics and his development of a mathematical theory of measurement in quantum mechanics; includes his argument that if the basic structure of quantum theory is to remain unchanged, the statistical nature of statements of the theory cannot be eliminated by the introduction of ‘hidden parameters’.)

Neumann, J. von and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press.(The foundational work in mathematical economics. Chapter 1, which explains the basis for the extension of game theory to economics, is accessible to a general audience, but the remainder very quickly becomes technically intricate.)

Neumann, J. von (1945) ‘First Draft of a Report on the EDVAC’ (for the US Army Ordnance

Department/University of Pennsylvania), repr. in *Papers of John von Neumann on Computing and Computer Theory*, ed. W. Aspray and A. Burks, Cambridge, MA: MIT Press, 1986; corrected version, *IEEE Annals of the History of Computing* 15 (4): 27-75, 1993.(This paper specifies the first logical design of an electronic computer whose program could be both stored and modified electronically. It incorporated, without credit, work by J.P. Eckert, Jr and J. Mauchly, which led to a dispute over who really originated the 'von Neumann' architecture. See Goldstine (1972) and especially Williams (1985) for accounts of the controversy. For justification of the corrections in the 1993 version, see Godfrey and Hendry (1993).)

Neumann, J. von, Burks, A. and Goldstine, H. (1946) 'Preliminary Discussion of the Logical Design of an Electronic Computing Instrument' (US Army Ordnance Department/Institute for Advanced Study, Princeton, NJ), repr. in *The Collected Works of John von Neumann*, ed. A.H. Taub, Oxford and New York: Pergamon Press, 1963, vol. 5, 34-79; and in *Papers of John von Neumann on Computing and Computer Theory*, ed. W. Aspray and A. Burks, Cambridge, MA: MIT Press, 1986, 97-142.(This paper completes the description of what became known as the 'von Neumann architecture' by incorporating a random access memory along with the stored program concept articulated in von Neumann (1945).)

Neumann, J. von (1950) *Functional Operators*, vol. 1, *Measures and Integrals* (Annals of Mathematics Studies 21), vol. 2, *The Geometry of Orthogonal Spaces* (Annals of Mathematics Studies 22), Princeton, NJ: Princeton University Press.(Notes compiled from lectures on measure theory given at the Institute for Advanced Study at Princeton in 1934.)

Neumann, J. von (1951) 'The General and Logical Theory of Automata', in L. Jeffress (ed.) *Cerebral Mechanisms in Behavior: The Hixon Symposium*, New York: Wiley, 1-31; repr. in *The Collected Works of John von Neumann*, ed. A.H. Taub, Oxford and New York: Pergamon Press, 1963, vol. 5, 288-328; and in *Papers of John von Neumann on Computing and Computer Theory*, ed. W. Aspray and A. Burks, Cambridge, MA: MIT Press, 1986, 391-431.(A slightly revised version of a paper delivered at the Hixon Symposium (1948) reflecting on the influence of McCulloch and Pitts (1943) on the design of digital computers, and speculating on the direction von Neumann would later take in his work on self-reproducing automata; accessible to a general audience and a nice introduction to his work in this area.)

Neumann, J. von (1958) *The Computer and the Brain*, New Haven, CT: Yale University Press.(The text of the 1956 Mrs Hepsa Ely Silliman memorial lectures, which von Neumann was too ill to deliver at Yale University; accessible to a general audience.)

Neumann, J. von (1960) *Continuous Geometry*, Princeton, NJ: Princeton University Press.(Notes from lectures given at Princeton 1935-7 on a topic von Neumann invented in 1935 by considering geometric interpretations of von Neumann algebras.)

Neumann, J. von (1966) *Theory of Self-Reproducing Automata*, ed. A. Burks, Urbana, IL: University of Illinois Press.(Based on lectures given at the University of Illinois in 1949.)

References and further reading

Arbib, M. (1964) *Brains, Machines, and Mathematics*, New York: McGraw-Hill; 2nd revised edn, New York: Springer, 1987.(An introduction to the theory of automata and computation, including discussion of von Neumann machines, self-reproducing automata and neural nets. The second edition includes material on connectionism.)

Aspray, W. (1990) *John von Neumann and the Origins of Modern Computing*, Cambridge, MA: MIT Press.(A detailed intellectual biography of von Neumann, concentrating on his contributions to the theory of computation, computer design and scientific computer applications. Provides accurate glosses of the technical work for a general audience.)

Brink, J. and Haden, C. (eds) (1989) *The Computer and the Brain: An International Symposium in Commemoration of John von Neumann (1903-1957)*, special issue of *Annals of the History of Computing* 11 (3).(Selected and revised essays from a 1987 symposium of biologists, computer scientists, electrical engineers, historians, linguists, philosophers, physicists and psychologists. Topics include aspects of von Neumann's life, his correspondence with Hungarian physicist Rudolph Ortvay (in which von Neumann worked out some of his ideas in automata theory), his contribution to computing and the effect of his influence on subsequent developments in the field.)

Glimm, J., Impagliazzo, J. and Singer, I. (eds) (1990) *The Legacy of John von Neumann*, Proceedings of Symposia in Pure Mathematics, vol. 50, Providence, RI: American Mathematical Society.(Twenty-two articles examining the mathematical ideas of von Neumann and tracing their subsequent development and evolution; includes a short note on von Neumann's philosophical development by his brother. A useful successor to Oxtoby, Pettis and Price (1958).)

Godfrey, M. and Hendry, D. (1993) 'The Computer as Von Neumann Planned It', *IEEE Annals of the History of Computing* 15 (1): 11-21.(Discusses the motivation for the logical architecture and design in the 'First Draft of a Report on the EDVAC' (1945) and contrasts the machine described with the EDVAC actually constructed. Argues for some corrections to von Neumann's original first draft and points out errors which have crept into subsequent reprints.)

Goldstine, H. (1972) *The Computer from Pascal to von Neumann*, Princeton, NJ: Princeton University Press.(A thorough history of digital computing with particular attention paid to the circumstances surrounding von Neumann's 1945 and 1946 papers.)

Heims, S. (1980) *John von Neumann and Norbert Weiner: From Mathematics to the Technologies of Life and Death*, Cambridge, MA: MIT Press.(Part biography and part exploration of the connections between science, technology and politics; provides a nontechnical discussion of von Neumann's work in historical context. Of particular biographical interest is chapter 14, which presents a non-hagiographic portrait of von Neumann.)

McCulloch, W. and Pitts, W. (1943) 'A Logical Calculus of the Ideas Immanent in Nervous Activity', *Bulletin of Mathematical Biophysics* 5: 115-33; repr. in M. Boden (ed.) *The Philosophy of Artificial Intelligence*, New York: Oxford University Press, 1990, 22-39.(The original paper describing how computations could be performed by networks of threshold logic units (artificial neurons). Accessible to those familiar with logical formalisms; a more gentle (though still rigorous) introduction is Arbib (1964).)

Oxtoby, J., Pettis, B. and Price, G. (eds) (1958) *John von Neumann, 1903-1957*, special issue of the *Bulletin of the American Mathematical Society* 64 (3, part 2).(Includes Ulam (1958) and seven articles which provide good introductions to particular aspects of von Neumann's mathematical work, including quantum theory, game theory and automata theory; most trace developments arising from von Neumann's work and each has a selected bibliography.)

Poundstone, W. (1992) *Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb*, New York: Doubleday.(A partly biographical essay which presents the theory of games for a general audience; includes a bibliography.)

Ulam, S. (1958) 'John von Neumann, 1903-1957', *John von Neumann, 1903-1957*, special issue of the *Bulletin of the American Mathematical Society* 64 (3, part 2): 1-49.(A good introductory survey of von Neumann's mathematical work, particularly useful as Ulam explains and assesses von Neumann's contributions while giving specific pointers to his published work; also includes a brief biography and a bibliography of von Neumann's publications.)

Williams, M. (1985) *A History of Computing Technology*, Englewood Cliffs, NJ: Prentice-Hall.(A reliable and very accessible introduction to computing devices, with particular attention paid to mechanical details; includes a balanced account of the attribution controversy over the von Neumann architecture.)

Neurath, Otto (1882-1945)

An Austrian socialist philosopher, economist, sociologist and historian, Neurath was a charismatic orator and an energetic cultural activist. Deeply concerned with education as a tool for social progress and cooperation, he founded museums and created the Isotype language for visual education. During the Bavarian Revolution he headed a plan for the centralized socialization of the Bavarian economy. A determined supporter of the scientific attitude and opponent of metaphysics (which he believed to have pernicious political and social consequences), he was a founder member of the Vienna Circle and a heterodox proponent of logical empiricism. He spearheaded the Unity of Science Movement that launched the project of an encyclopedia of unified science.

1 Life

Otto Neurath was born on 10 December 1882, in Vienna, son of Gertrud Kaempfert and Wilhelm Neurath, social reformer and political economist. He received a doctoral degree in Berlin in the history of economics in 1906. A study of the Balkans led to his theory of war economy as a natural (non-monetary) economy upon which he based his programme for full socialization. In 1919 the short-lived Bavarian socialist government appointed him head of the Central Planning Office. From 1920 to 1934 Neurath participated actively in the direction of Viennese socialist politics, especially in housing and adult education. He founded the Social and Economic Museum of Vienna, which created the 'Vienna method' of picture statistics and the Isotype language ('International System of Typographic Picture Education'); the Verein Ernst Mach, which, with the publication of an intellectual manifesto in 1929, became the public face of the Vienna Circle; the International Foundation for Visual Education at the Hague; and the International Unity of Science Movement. The latter, inspired by the ideas of the French Encyclopedists, launched the project of an Encyclopedia for Unified Science. Together with the pictorial languages, this scientific encyclopedia would promote social solidarity and progress at an international level. Neurath fled from the Nazis, first to The Hague, then in 1940 to England, where after nine months of internment he resumed activities related to the Isotype method and the unity of science. He died in Oxford on 22 December 1945.

2 Science and society: empiricism in the social sciences (1910-31)

Neurath championed 'the scientific attitude'. He denied any value to philosophy over and above the pursuit of work on science, within science and for science. His views on the language, method and unification of science were led throughout by his interest in the social life of individuals and their well-being. This concern is reflected in his engagement both in political life and in social science. For Neurath, to theorize about society is inseparable from theorizing for and within society (see Marxist philosophy of science). Science is in every sense a social and historical enterprise. It is about social objectives as well as objects, about social realizations as well as reality.

Neurath drew upon the spirit - if not the letter - of two major turn-of-the-century thinkers: Ernst Mach and Karl Marx. Mach introduced a radical antimetaphysical approach into the analysis of science which Neurath embraced, in part because he believed that metaphysical obscurantism underwrote social institutions that attacked enlightenment values such as equality, freedom and progress. Neurath's aim was to apply the empirical attitude throughout the social sciences. He attacked the distinction drawn by Dilthey, Rickert and Weber between the natural and the social sciences, which rested on non-empirical practice (like the 'empathic' method). A purely empiricist language would represent one big step towards unified science. Methodologically, the job of the social sciences, like the natural sciences, would be to establish correlations - statistical when possible - and to deduce predictions about the future.

Neurath saw in Marx a model of social science without metaphysics and also of unification - of sociology, economics and politics. Marx inspired in him the belief in the social and historical contingency of language and concepts, including those of science ('there is no *tabula rasa*'), and the enlightenment idea that the scientific attitude possesses a practical and socially redemptive (revolutionary) value: the social scientist is also a social engineer. This influence is reflected in Neurath's work in natural economy and his activities as social planner. During the 1910s he articulated, following ideas of Popper-Lynkeus (see Wachtel 1955), an 'ecological' economic programme based on central planning and non-monetary structure. The goal was not the increase of wealth, but the allocation of exhaustible resources and goods in a way that increased the standards of living of individuals; the method, an empirical calculus of needs and satisfaction that introduced 'qualitative exactness' in place of the

quantitative monetary calculations and a central organization in place of a market and monetary unit. The centralized programme required, Neurath realized, a practical, cooperative unification of the sciences: unity at the point of action.

Neurath's ideas also reflect the influence of Henri Poincaré and P.M.M. Duhem (see Conventionalism). A multiplicity of theories can accord equally well with the same data. Science is methodologically open-ended. For Neurath the objective was not to determine which theory is true but which theory, or combination, should be used for a given purpose. This decision is pragmatic: the social scientist is to examine all possible theories and predictions that fit the data, in order, like the engineer, to design (social) machines that have never yet been built.

Neurath rejected all pictures of 'ideal' science as gross metaphysics. His attention to the inextricable link between science and society had several consequences: the amount of knowledge available can never be exhaustive; the uncertainty involved in justifying decision-making 'scientifically' can only be honestly and rationally eliminated through extra-empirical 'auxiliary' motives (to deny this limitation in the power of scientific justification amounts to 'pseudorationalism'); the empirical language of science cannot be a precise and atomic Machian one, on pain of losing descriptive, predictive and explanatory power; the introduction of vague, ambiguous terms into the 'physicalist' language is inevitable; abstract analysis prompts the adoption of a multiplicity of conventional idealized concepts - social factors or indicators - which are value-laden and historically contingent.

3 Scientific language and scientific method: the Vienna Circle and Neurath's logical empiricism (1931-5)

Neurath's later views on the language and method of science represented his simultaneous response to problems in the social sciences and to problems addressed by the Vienna Circle between 1928 and 1934, and by Karl Popper (see Vienna Circle §§2-3). A primary aim of the Vienna Circle was to account for the objectivity of scientific knowledge claims without resort to metaphysics. The method was to investigate the logical and linguistic frameworks of scientific knowledge (the linguistic turn). The 'orthodox' position was that revisable theoretical scientific statements stand in appropriate logical relations to unrevisable elementary observation statements (the data), called 'protocol statements'.

Neurath took the objectivity of scientific knowledge to be provided by the public and social nature of its representations and rules for acceptance. By 1931 his proposed scientific language was 'radical physicalism': scientific statements must speak of material things (not necessarily from the domain of physics) in space and time. Objectivity and the avoidance of metaphysics require that sentences be compared only with sentences, not with 'reality' ('it is impossible to step outside language'). Protocol statements preserve strict 'linguistic empiricism' by avoiding both reference to 'reality' (contra Schlick and Popper) and to subjective (phenomenalistic) experience (contra Carnap and Schlick). They are not items of personal experience but public reports of scientific observations, like a laboratory notebook. They are thus syntactically complex and 'rich in theory', not atomic and purely observational, unavoidably incorporating into science complex and vague cluster-concepts from natural language that lack the 'ideal' precision of theoretical terms. This makes scientific language a historically provided 'jargon'.

Neurath rejected both Carnap's methods of verification and Popper's method of falsification (see Scientific method). With Duhem he believed first that not one but only groups of hypotheses can be confronted with empirical data, hence leaving undetermined how to assign praise or blame within the group (Duhem-Quine problem), and second, that any number of such groups can always fit the same data (see Quine, W.V.; Underdetermination §3). Moreover, protocol statements are revisable (see Fallibilism). In case of conflict between the prediction of theory and a protocol statement, either the theory or the protocol can be rejected (the Neurath principle). Finally, in many cases the presence of imprecise terms will leave the protocol statements in an indeterminate logical relation to the theoretical ones, thus rendering both verification and falsification logically inconclusive. Neurath concluded that the belief in a general unambiguous scientific method is a 'pseudorational' ideal and that all theory choice is ultimately pragmatic. His views are illustrated with his image of the ship: 'We are like sailors who have to rebuild their ship on the open sea, without ever being able to dismantle it in dry-dock and reconstruct it from its best components' (1983: 92).

4 Unity of science and the encyclopedia-model (1934-45)

The ship metaphor also illustrates Neurath's idea of the unification of the sciences as a historical and communal enterprise. In 1934 he articulated the ideas on the unity of science that had been developing in his works since 1910 (see Unity of science). He opposed the myth of the 'system-model': a unique, axiomatic, deductively closed and complete hierarchy of sciences. Instead he proposed the more realistic 'encyclopaedia-model': the coherent totality of the accepted scientific statements at a given time, in flux, incomplete, not void of imprecise statements and logical gaps, unified linguistically by the physicalist 'jargon', and methodologically by the general use of certain techniques (statistics, probabilities and so on) and the empiricist spirit; on the whole, in Neurath's terms, like a 'mosaic', an 'aggregation' of the sciences, or an interdisciplinary 'orchestration' of the scientists' efforts.

See also: Holism and individualism in history and social science; Naturalism in social science; Vienna Circle §4

JORDI CAT
NANCY CARTWRIGHT

List of works

Neurath, O. (1981) *Gesammelte philosophische und methodologische Schriften (Collected Philosophical and Methodological Writings)*, ed.R. Haller and H. Rutte, Vienna: Holder-Pichler-Tempsky.(This work does not include Neurath's economic writings, which are currently being edited by R. Haller, his numerous political writings nor those on the Isotype language.)

Neurath, O. (1939) *Modern Man in the Making*, New York: Knopf.(Popular socio-historical-cum-philosophical account of the rise of modernity, with ISOTYPE illustrations.)

Neurath, O. (1944) *Foundations of the Social Sciences*, International Encyclopaedia of Unified Science vol. 2, no. 1, Chicago, IL: University of Chicago Press.(An account of the place of the social sciences in his controversial 'encyclopedic' conception of unified science.)

Neurath, O. (1973) *Empiricism and Sociology*, ed. R.S. Cohen and M. Neurath, Dordrecht: Reidel.(A collection of memoirs of Neurath and a selection from his philosophical, scientific, pedagogic and economic writings; includes Vienna Circle manifesto and Anti-Spengler.)

Neurath, O. (1983) *Philosophical Papers 1913-1946*, ed. R.S. Cohen and M. Neurath, Dordrecht: Reidel. (Selection of philosophical writings, predominantly from Neurath's Vienna Circle period.)

Neurath, O. (1991) *Gesammelte bildpädagogische Schriften*, ed. R. Haller and R. Kinross, Vienna: Hölder-Pichler-Tempsky.(Collected writings on picture statistics and visual education, 1925-45.)

Neurath, O. (forthcoming) *Gesammelte ökonomische Schriften*, ed. R. Haller, Vienna: Hölder-Pichler-Tempsky. (Collected economic writings, 1904-45.)

Neurath, O. (forthcoming) *Selected Economic Writings*, ed. R.S. Cohen and T.E. Uebel, Dordrecht: Kluwer. (Selections from Neurath's economic writings, 1909-45.)

References and further reading

Cartwright, N., Cat J., Fleck, K. and Uebel, T. (1995) *Between Science and Politics: The Philosophy of Otto Neurath*, Cambridge: Cambridge University Press.(A study of the joint development of Neurath's social science, his political activities and his philosophical point of view.)

Nemeth, E. and Stadler, F. (eds) (1996) *Otto Neurath: Encyclopedia and Utopia*, Dordrecht: Kluwer.(This volume contains recent works reassessing Neurath's views on philosophy, the social sciences and education.)

Uebel, T.E. (ed.) (1991) *Rediscovering the Forgotten Vienna Circle*, Dordrecht: Kluwer.(A collection and translation of recent German and Austrian scholarship.)

Uebel, T.E. (1992) *Overcoming Logical Positivism from Within*, Amsterdam: Rodopoi.(A critical philosophical history and investigation of Vienna Circle views.)

Wachtel, H.I. (1955) *Security for All and Free Enterprise: A Summary of the Social Philosophy of Josef Popper-Lynkeus*, New York: Philosophical Library.(An account of Popper-Lynkeus's life and work, based on his *Die allgemeine Nährpflicht als Lösung der socialen Frage* (1912).)

Zolo, D. (1989) *Reflexive Epistemology: The Philosophical Legacy of Otto Neurath*, Dordrecht: Kluwer.(A philosophical and historical account focusing on Neurath's views about knowledge and method.)

Neutral monism

Neutral monism is a theory of the relation of mind and matter. It holds that both are complex constructions out of more primitive elements that are 'neutral' in the sense that they are neither mental nor material. Mind and matter, therefore, do not differ in the intrinsic nature of their constituents but in the manner in which the constituents are organized. The theory is monist only in claiming that all the basic elements of the world are of the same fundamental type (in contrast to mind-body dualism); it is, however, pluralist in that it admits a plurality of such elements (in contrast to metaphysical monism).

Historically, neutral monism had its origins in Ernst Mach's sensationalism (see Mach, E.). According to Mach (1885), the relative permanence of both bodies and the ego are due to their being composed of large numbers of elements not many of which are replaced at any one time. He called these elements 'sensations', though the phenomenalist suggestions carried by this term are misleading, for Mach's sensations are neither mental nor physical.

Mach directly influenced both William James and Bertrand Russell, the philosophers most widely regarded as neutral monists. James' early psychological work, notably his *Principles of Psychology* (1890), was explicitly dualist, although his neutral monism began to appear soon afterwards - for example, in his account of colour perception in *Psychology. A Briefer Course* (1892: 332) - and some commentators have suggested the dualism of the *Principles* was merely methodological. James' monism emerges clearly in 'The Knowing of Things Together' (1895), where he claimed (rather loosely) that the object known and the mind that knows it were ' only two names that are given to the one experience, when, taken in a larger world of which it forms a part, its connections are traced in different directions' (1895: 379). For the next couple of years, in classes and notebooks, James developed these ideas into the view, very similar to Mach's, that there was no 'pure ego' and no material substance, only experience. This, however, remained unpublished until, in 1904-5, he wrote most of the essays which were posthumously collected in *Essays in Radical Empiricism* (1912), and which contains his fullest statement of the theory (see James, W. §6).

It seems to have been Russell who coined the term 'neutral monism' in the course of criticizing the theory in 1913. (James always referred to it as 'radical empiricism', though he had applied the term 'neutral' to experiences in notes written in the 1880s.) Russell did not embrace neutral monism until 1919. In doing so he was influenced to lay aside his earlier objections partly by Hume's arguments for the unintrospectibility of the self (which also impressed James), but most importantly by Wittgenstein's (1921) treatment of belief, which 'shows…that there is no such thing as the soul' (see Wittgenstein, L. §§3-5). James' treatment of belief had been a particular target of Russell's earlier criticism and Russell's first work as a neutral monist was an account of belief (Russell 1919; 1925).

Russell treated the theory more broadly in *The Analysis of Mind* (1921), where he attempted to show how both the mind and material objects could be constructed out of elements he called 'events', the differences between mind and matter arising from the irreducibly different types of causal law each obeyed (see Russell §13). As parts of material objects, events were related by physical causal relations expressed ultimately by differential equations; as parts of the mind they were related by 'mnemic' causal laws, in which a past state of the mind is among the proximate causes of its current state. Although Russell changed details in subsequent works, he never abandoned this general picture (Russell 1956).

In developing his theory, Russell was influenced by recent advances in physics (especially as interpreted by the subjectivist philosophy of A.S. Eddington) and by behaviourism. He said that modern physics had shown matter to be less material than formerly supposed, while modern psychology had shown mind to be less mental. He thought of neutral monism as the natural outcome of these two developments. Despite the evident influence of behaviourism, he rejected the behaviourist attempt to reduce mental phenomena to physical ones. Indeed, Russell insists that mental states, despite their status as logical constructions, are causally efficacious. This has led Baldwin (1995) to suggest that Russell's neutral monism is best seen as an early form of functionalism.

Historically, however, the most obvious heir of neutral monism was the brain-mind identity theory. Here,

neurophysiology takes the place of behaviourist psychology and the account is (typically) given a more materialist slant than the neutral monists would have accepted. None the less, many neutral monist ideas reappear in the identity theory, as identity theorists have acknowledged (Feigl 1967).

See also: Behaviourism, analytic; Functionalism; Mind, Identity theory of; Monism

NICHOLAS GRIFFIN

References and further reading

Ayer, A.J. (1968) *The Origins of Pragmatism: Studies in the Philosophy of Charles Sanders Peirce and William James*, San Francisco, CA: Freeman, Cooper.(Pages 224-336 provide a readable and comprehensive survey of James' neutral monism.)

Ayer, A.J. (1971) *Russell and Moore. The Analytical Heritage*, London: Macmillan.(Chapter 5 contains a good, brief account of Russell's neutral monism.)

Baldwin, T. (1995) 'Introduction', in B. Russell *The Analysis of Mind*, London: Routledge. (A very useful, brief survey of *The Analysis of Mind*, linking Russell's ideas to current work.)

Feigl, H. (1967) *The 'Mental' and the 'Physical'. The Essay and a Postscript*, Minneapolis, MN: University of Minnesota Press.(A nearly comprehensive survey of the literature on the brain-mind identity theory to 1967, which recognizes Russell as a precursor - but, surprisingly, not James.)

James, W. (1890) *The Principles of Psychology*, New York: Dover, 1950, 2 vols.(James' pre-neutral monist psychology.)

James, W. (1892) *Psychology. A Briefer Course*, ed. F.H. Burkhardt and I.K. Skrupskelis, Cambridge, MA: Harvard University Press, 1984.(Contains James' first hints of neutral monism.)

James, W. (1895) 'The Knowing of Things Together', in *Essays in Philosophy*, ed. F.H. Burkhardt and I.K. Skrupskelis, Cambridge, MA: Harvard University Press, 1978, 71-89. (James' address to the American Psychological Association; it contains the first full statement of his neutral monism.)

James, W. (1912) *Essays in Radical Empiricism*, ed. F.H. Burkhardt and I.K. Skrupskelis, Cambridge, MA: Harvard University Press, 1976.(The *locus classicus* for James' neutral monism. Two essays, 'Does "Consciousness" Exist?' and 'A World of Pure Experience' are particularly important.)

Mach, E. (1885) *The Analysis of Sensations and the Relation of the Physical to the Psychical*, 5th edn, 1906, trans. C.M. Williams and S. Waterlow, New York: Dover, 1959. (The first appearance of neutral monism.)

Perry, R.B. (1938) *In The Spirit of William James*, New Haven, CT: Yale University Press. (Contains a sympathetic account of James' position by a friend, follower and fellow neutral monist.)

Russell, B. (1913) *Theory of Knowledge. The 1913 Manuscript*, ed. E.R. Eames and K. Blackwell in *The Collected Papers of Bertrand Russell*, vol. 7, London: Allen & Unwin, 1983.(Chapter 2 contains Russell's criticisms of neutral monism.)

Russell, B. (1919) 'On Propositions: What They Are and How They Mean', *Aristotelian Society* supplementary vol. 2: 1-43; reprinted in *The Collected Papers of Bertrand Russell*, vol. 8, ed. J.G. Slater, London: Allen & Unwin, 1986, 278-306.(Russell's first published neutral monist work, concerned especially with the analysis of belief and other 'propositional' attitudes.)

Russell, B. (1921) *The Analysis of Mind*, London: Routledge, 1995.(Russell's first full statement of his neutral monism.)

Russell, B. (1925) 'Truth Functions and Others', Appendix C in B. Russell and A.N. Whitehead *Principia Mathematica*, 2nd edn, 1925-7, 3 vols, Cambridge: Cambridge University Press, 659-66.(Russell's most detailed attempt at a neutral monist analysis of belief.)

Russell, B. (1956) 'Mind and Matter' in *Portraits from Memory and Other Essays*, London: Allen & Unwin, 1956, 140-60.(Russell's last statement of his neutral monism.)

Sainsbury, R.M. (1979) *Russell*, London: Routledge.(Pages 261-8 give a good critical survey of Russell's early objections to neutral monism and his subsequent response to them.)

Wild, John (1969) *The Radical Empiricism of William James*, New York: Doubleday.(A good introductory account that emphasizes the affinities of James' radical empiricism with phenomenology.)

Wittgenstein, L. (1921) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London: Routledge & Kegan Paul, 1966.(Propositions 5.541-5.5421 give the main reasons that led Russell to embrace neutral monism.)

Neutrality, political

The principle of political neutrality, which requires the state to remain neutral on disputed questions about the good, is an extension of traditional liberal principles of toleration and religious disestablishment. However, since neutrality is itself a contested concept, the principle remains indeterminate: is it, for example, a requirement of neutral reasons for legislation (or neutral legislative intentions) or is it a more exacting requirement of equal impact in so far as legislative consequences are concerned? The answer must surely reflect the deeper values that are used to justify the neutrality principle. This raises further problems, however. If the principle is based upon certain value commitments - such as the importance of equality or individual autonomy - then it cannot require us to be neutral about all values. It requires some sort of distinction between principles of right (of which neutrality is one) and conceptions of the good (among which neutrality is required). Critics believe that liberal principles of right are symptomatic of a deeper liberal bias in favour of individuality as a way of life. Perhaps liberals should embrace this point, and accept that the neutrality they advocate is quite superficial compared to the depth of their own value commitments.

1 The neutrality principle

In recent political philosophy, the idea of neutrality has been used by Bruce Ackerman, Ronald Dworkin, John Rawls and others to express a liberal position concerning the possibility of bias or commitment on the part of the state in ethical and religious controversies. The citizens of modern pluralist societies disagree on almost everything, but particularly on issues of faith, philosophy and value. The principle of political neutrality requires the state to stay out of all such controversies, and to refrain from throwing the weight of its authority - or worse still, the coercive power at its command - behind any particular view about what makes life worth living.

2 Defending political neutrality

Neutrality may be understood as a reformulation and extension of the traditional principles of religious toleration and disestablishment. Liberal philosophy has always involved an affirmation of political secularism and a rejection of the idea of a confessional state. Political neutrality extends that to encompass the rejection not just of religious establishment, but of ethical perfectionism also (see Perfectionism; Toleration).

The extension is motivated in part by a sense that people’s individual ethical values play much the same role in their lives as their religious convictions. To impose a religious orthodoxy is to strike at the heart of a person’s freedom, for it threatens to displace the criteria which they regard as most important in organizing and disciplining their life. For many people, however, this central role is played not by religious beliefs but by ethical convictions. The broader doctrine of neutrality therefore requires the state to stand back and let people live their individual lives on whatever terms they choose, whether or not those terms have the aura of sanctity associated with religion.

Neutrality can also be defended by arguing that the means which the state characteristically uses to achieve its goals - namely, force and the threat of legal sanctions - are inappropriate for promoting religious or ethical objectives. Proponents of religious toleration such as John Locke believed that a faith imposed coercively from the outside was as good as no faith at all - ’I cannot be saved by a religion that I distrust, and by a worship I abhor’ (1689: 32) - and the development in the nineteenth and twentieth centuries of the idea of ethical individuality led to a similar conclusion about values generally. As Ronald Dworkin put it: ’No life is a good one lived against the grain of conviction. It does not help someone else’s life but spoils it to force values upon him he cannot accept but can only bow before out of fear or prudence’ (1993: 168).

Alternatively, neutrality may be defended on pragmatic or consequentialist grounds. It has not been our experience that those who wield governmental power are endowed with profound ethical insight. In the real world, non-neutral government may well lead to the inculcation and enforcement of superficial or inappropriate ethical standards. More affirmatively, defenders of neutrality may appeal to the case made by J.S. Mill in the second chapter of *On Liberty* (1859), where he argued that the existence of a diverse array of ethics and lifestyles, competing with one another in the marketplace of ideas, is the best way of arriving at the truth, if there is any truth, about the good for human beings. Non-neutral action on the part of the government amounts to loading the dice in favour of one view or set of views in a way that is likely to distort the market-like processes whereby truths and

half-truths are winnowed out from error.

3 Intentions, reasons or consequences?

The justification of the neutrality principle is crucial for our understanding of what, exactly, it requires. Does neutrality simply mean *not taking sides*, or does it mean *going out of one's way to ensure that one does not have an unequal impact*, in the struggle or controversy in which one is supposed to be neutral? Is the liberal state supposed to be neutral in virtue of its intentions or neutral in virtue of the consequences of its laws and policies?

A number of writers - for example, Alan Montefiore (1975) and Joseph Raz (1986) - have interpreted liberal neutrality as concerning primarily the consequences of legislative action. For example, the legislator must take care that the laws do not increase the chances of a hedonistic lifestyle flourishing at the expense of adherence to traditional Christian values. It must enhance or retard the prospects of these lifestyles to the same degree. But this conception gives rise to a number of difficulties. It involves the postulation of some baseline, relative to which differential effects of state action may be measured. And in practical terms it is a very difficult requirement to live up to, because it is so hard to predict what the effect of a law is going to be on lifestyles and mores. If that is how neutrality were to be understood, we should have grave doubts about whether it was ever reasonable to require legislators to be neutral.

Instead of neutrality of impact, the liberal may be talking about neutrality of intention - that is, neutrality in relation to the motives and reasons that the legislator uses to justify the laws. On this account, the fact that a law against Sunday trading would accord with the requirements of a Sabbatarian faith is not a good reason for having such a law; but the fact that it is necessary to prevent shop employees from being overworked may be. The latter reason can be a good reason, and the legislation neutral on that account, even though the law undoubtedly benefits Sabbatarianism over other sects.

This interpretation also faces difficulties however. It is often hard to determine the real intentions behind what lawyers would call a 'facially neutral' statute; a law banning animal sacrifice might be promulgated in the name of public health, but what if some of the lawmakers voted in favour of it because of their antipathy to an animal sacrifice cult? It may be wiser to reformulate neutrality of intention in terms of a principle governing what counts as a good reason for a law or policy. Such a principle would not condemn a piece of legislation if there were in fact neutral reasons in favour of it, whether or not those reasons actually figured in the conscious intentions of the lawmakers.

Which of these conceptions - neutral impact or neutral reasons - should the modern liberal adopt? This must depend on the grounds upon which neutrality is ultimately valued. For example, the Lockean argument about the impossibility of coercing religious faith may yield or justify one conception of neutrality - neutrality of reasons - since it is only the use of coercion to *achieve* religious objectives that is said to be self-defeating. The 'marketplace-of-ideas' argument, on the other hand, may be more congenial to consequentialist concerns. If ethical diversity is valued as a medium in which truth can emerge and progress be made, then we ought to be careful that our political action (whatever its intentions) does not accidentally diminish the diversity of lifestyles.

Something similar applies to the question of non-coercive state action. Is the state required to be neutral in everything it does, or only in its use of force? In England, there is an established state Church, but nowadays no one is obliged to belong to it. Does this violate the neutrality constraint? What about the official institution of (heterosexual) marriage? What about government subsidies to the arts? These policies and institutions do not compel any citizen to do anything in particular, but arguably they still represent the state taking sides in ethical and religious disputes. Can the liberal principle be violated then in this purely symbolic way? It is impossible to find answers to these questions by considering the word 'neutrality' (or, for that matter, the word 'coercion'). The answers must be dictated by the deeper concerns that lie behind our choice of this particular terminology in which to phrase liberal commitments.

4 The distinction between the right and the good

The move from religious toleration to a broader neutrality about values generally poses one other difficulty for liberal political theory. Liberals believe that governments are set up to do justice, to respect rights, to promote the general welfare and to protect freedom. Thus they cannot hold that the state should be neutral in relation to *all*

moral values. After all, neutrality is itself a value: it is a normative position, a doctrine about what legislators and state officials ought to do. Liberals favour neutrality, in Ronald Dworkin's words, 'not because there is no right and wrong in political morality, but because that is what is right' (1978: 142).

It follows that a principle of political neutrality has to be able to justify certain discriminations among values or principles. We have to be able to distinguish value controversies in which the liberal legislator is required to get involved (such as controversies about justice), from those value controversies in which they are required by the principle not to get involved (such as controversies about lifestyle, sin or salvation). The distinction is sometimes characterized in terms of the difference between 'the right' and 'the good' (see Right and good). Neutrality itself and the principles associated with it, such as justice, freedom and individual rights, are principles of political right: their point is to define a framework within which individuals ought to be able to pursue any conception of the good they please. The conception they choose must of course fit within the framework of right, for the framework exists in order to ensure that each person's pursuit of the good (as they see it) is compatible with everyone else's pursuit of the good (as they see it). In this sense, the right is said to be prior to the good, even though the right exists in order to accommodate the good.

But the distinction between the right and the good has itself been subject to heavy criticism, most notably by Michael Sandel (1982), Charles Taylor (1985) and other communitarian writers (see Community and communitarianism §2). They doubt whether any sensible system of right can be defined without reference to a determinate conception of human good: how, for example, can we define a set of rights to cope with the fact of our mutual vulnerability, without some detailed account of what counts as harm and thus some correlative account of interests, needs, health and a person's good? They also note that many people's conceptions of the good life already involve a view of the good society and they may therefore already implicate principles of right: think of an Englishman yearning for a life of traditional service in the established Church, or an Algerian who believes that the practice of his religion is impossible except under a system of Islamic law. Since the liberal doctrine directly challenges these aspirations, it is evidently *not* neutral between all conceptions of human fulfilment, but only (or at most) between those that fit a certain individualistic pattern.

This is not a surprising result. Even the old principle of toleration never claimed to be neutral between all religions: it condemned human sacrifice, the torture of heretics and compulsory tithing, even though these were - strictly speaking - religious practices. It was neutral only between non-coercive faiths. The liberal proponent of political neutrality rejects (and refuses to accommodate) communitarian conceptions of the good for similar reasons.

Still, the critics are right to insist that the liberal framework of right itself embodies certain assumptions about the good. Liberals believe that the shaping of human lives by the individuals who are living them is a good thing, and so they adopt principles of right (including the principle of neutrality) which seek to minimize the coercive impact of other people's views on that process. It is true that some individuals choose to shape their lives (or would like to shape their lives) in accordance with a communitarian conception. But that choice amounts in part to a proposal to govern or interfere with others' shaping of their lives, and this the liberal cannot tolerate. In other words, far from the liberal being required to be neutral between communitarian and non-communitarian conceptions of the good, the liberal's adoption of the neutrality principle is an expression of opposition to the very idea of a communitarian conception of the good.

See also: Liberalism

JEREMY WALDRON

References and further reading

Ackerman, B. (1980) *Social Justice in the Liberal State*, New Haven, CT: Yale University Press.(An argument that liberal justification consists in dialogue constrained by a principle of neutrality concerning conceptions of the good.)

Barry, B. (1965) *Political Argument*, London: Routledge & Kegan Paul, 75-9.(The argument, alluded to in §3 of this entry, that political neutrality is impossible, on the assumption that it means neutrality of impact.)

Dworkin, R. (1978) 'Liberalism', in S. Hampshire (ed.) *Public and Private Morality*, Cambridge: Cambridge University Press.(An argument that the basic liberal commitment to equality is best understood in terms of a principle of political neutrality.)

Dworkin, R. (1993) *Life’s Dominion: An Argument about Abortion, Euthanasia and Individual Freedom*, New York: Alfred A. Knopf.(Arguing for an extension of the principles of religious toleration and disestablishment to cover secular views about the value of life.)

Goodin, R. and Reeve, A. (eds) (1989) *Liberal Neutrality*, London: Routledge.(A collection of original essays on the subject of neutrality, applied to various institutions of liberal society, such as the law, the market and the university.)

Larmore, C. (1987) *Patterns of Moral Complexity*, Cambridge: Cambridge University Press, esp. 40-68.(A strong defence of the neutrality principle, purportedly on grounds that are themselves also neutral.)

Locke, J. (1689) *A Letter Concerning Toleration*, ed. J. Horton and S. Mendus, London: Routledge, 1991(The classic defence of religious toleration on the grounds of the inefficacy of coercion to produce sincere belief, referred to in §§2 and 3 of this entry.)

Mill, J.S. (1859) *On Liberty*, ed. C.V. Shields, Indianapolis, IN: Bobbs-Merrill, 1956.(A defence of individuality on the basis of the likelihood that truth will emerge from the free interplay of rival lifestyles and conceptions of the good.)

Montefiore, A. (ed.) (1975) *Neutrality and Impartiality*, Cambridge: Cambridge University Press.(A discussion of the neutrality of the modern university.)

Nagel, T. (1973) ’Rawls on Justice’, *Philosophical Review* 82: 220-34.(A version of the argument, alluded to in §4 of this entry, that liberal theories of justice are decidedly non-neutral between individualistic and communitarian conceptions of the good life.)

Rawls, J. (1971) *A Theory of Justice*, Oxford: Oxford University Press.(A liberal discussion of social justice, emphasizing the distinction between the right and the good, discussed in §4 of this entry.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(An argument to the effect that liberal principles must not rest on controversial philosophical principles, but must command overlapping support among a wide variety of ethical and philosophical conceptions in a modern pluralist society.)

Raz, J. (1986) *The Morality of Freedom*, Oxford: Clarendon Press, esp. 107-162 and 369-430.(A critique of political neutrality and an argument that the value of individual autonomy commits liberals to ethical perfectionism.)

Sandel, M. (1982) *Liberalism and the Limits of Justice*, Cambridge: Cambridge University Press.(A communitarian critique of the liberal distinction between the right and the good.)

Taylor, C. (1985) *Philosophy and the Human Sciences: Philosophical Papers 2*, Cambridge: Cambridge University Press, esp. 187-247.(The argument, mentioned in §4 of this entry, to the effect that liberal principles are incoherent unless underpinned by a distinctive conception of human flourishing.)

Waldron, J. (1989) ’Legislation and Moral Neutrality’, in R. Goodin and A. Reeves (eds) *Liberal Neutrality*, London: Routledge.(A defence of the neutrality principle; along the lines of this entry.)

Newman, John Henry (1801-90)

John Henry Newman was the principal architect of the Catholic revival (the Oxford or Tractarian movement) within the Church of England in the 1830s, and went on to become probably the most seminal of modern Roman Catholic thinkers. Although primarily a theologian, Newman regarded his defence of religious belief in terms of a philosophical justification of non-demonstrable certainty as his most important life work.

Newman was born in London, the elder son of a banker. Elected a fellow of Oriel College, Oxford, in 1822, he became vicar of the university church of St Mary the Virgin in 1828. The five sermons he preached there between 1839 and 1841 on the relation of faith to reason, published in*Sermons, chiefly on the Theory of Religious Belief, preached before the University of Oxford* (1843), are a brilliantly original challenge to the received Lockean understanding of rationality. His defence of the reasonableness of religious belief, which he completed in *An Essay in Aid of a Grammar of Assent* (1870), involves an epistemology which gives both works a wider philosophical significance.

Newman's reception into the Roman Catholic Church in 1845 was followed by the publication of the *Essay on The Development of Christian Doctrine*, his most famous theological work which itself presupposes the anti-Enlightenment theory of knowledge he had sketched in the *Sermons*. Founding President of the new Catholic University of Ireland from 1851 to 1858, Newman's *The Idea of a University* (1873) propounds a philosophy of education concerned with the development of the whole mind rather than only those faculties of utilitarian value. Created a cardinal in 1879, he died in Birmingham in 1890 at the Oratory he had founded there.

The target of the *Sermons* is essentially the restriction of reason to Cartesian rationalism and Lockean empiricism, that is, the assumption that knowledge must either be deduced from logical, a priori truths or derived a posteriori from sense experience by induction. Newman argues that religious faith is only a particular instance of the many perfectly reasonable convictions which we reach in a variety of areas. This kind of reasoning depends not so much on specific arguments and evidence as on pre-existing assumptions and expectations. Someone may legitimately reason in this informal sense without necessarily being able to provide any analysis of their reasoning or even produce their reasons. Thus people may argue badly and yet reason perfectly well. In some kinds of reasoning, and especially in matters of religious belief, the moral factor is especially important because it largely determines the relevant underlying assumptions that in turn form the actual reasons for believing something to be true. Newman thought that the philosopher and the factory worker who are believers share the same 'antecedent probabilities', which in turn rest on the same 'first principles'. The fact that people may not be able to state clearly what they believe and why is because of the distinction between what Newman calls 'explicit' and 'implicit' reasoning. Nor is there one single model of informal reasoning, since there are as many different types of reasoning as there are subjects on which to reason.

If religious belief involves the same kind of informal reasoning that we unhesitatingly employ in other intellectual matters, still the question remains whether certainty is possible and how it differs from the objective certainties of logically necessary propositions and empirically verifiable factual statements. The *Essay in Aid of a Grammar of Assent* is Newman's final attempt to answer the problem. He begins by distinguishing between the assent one makes to a proposition and the conclusion one draws from an inference: assenting and inferring are thus two distinct kinds of activity, however close they may sometimes seem to be. Assent may be 'notional' or 'real', depending on whether the apprehension of a proposition is notional or real. Newman's usage is confusing because, though 'notional' seems to mean what is abstract or general and 'real' what is concrete and individual, nevertheless the distinction is not in fact between sense perceptions and mental abstractions but between experiential and non-experiential knowledge. For one can give a real assent to an apparently abstract idea if that idea becomes a real 'thing' in the imagination (an 'image'), because one then grasps the force of the idea in an experiential and personal way. Indeed, a mental act may bring before the mind a more vivid image than a sensible object. Thus Newman's argument from conscience for the existence of God is that the intimations of conscience are the echoes of a magisterial voice suggestive of a God of whom we gain a real image from and in these dictates of conscience.

Given that assent is not simply the conclusion to an inference (from which it is distinct), then it does not follow, as

Locke claimed, that the only certainty is logical certainty. Propositions that are true by definition normally command assent, but even here assent may be withheld until the conclusion is accepted. Where formal logical inference is impossible, Newman argues that it is the cumulation of probabilities which leads to certainty (see Butler, J.). But assent to the truth of nonlogical propositions involves personal judgment as the mind evaluates both the strength of the individual probabilities and their combined strength. This judgment, which is akin to Aristotle's *phronesis*, Newman calls the 'illative sense', and it operates more or less implicitly and instinctively, without formal verbal analysis.

In his sensitivity to the actuality and complexity of human rationality and its linguistic expression, Newman clearly anticipates the central concerns of the later Wittgenstein. The other factor which has given his epistemology a new significance is his refusal to take the path of Schleiermacher in conceding to science all factual knowledge and claiming for religious utterances merely emotional, imaginative and existential significance; he has found unexpected support from modern philosophers of science who, while accepting the validity of scientific method, nevertheless recognize that the truth claims of science are not as self-evident and unproblematic as used to be assumed.

See also: Religion and epistemology

IAN KER

List of works

Newman, J.H. (1843) *Sermons, Chiefly on the Theory of Religious Belief Preached before the University of Oxford*, London and Oxford: Rivington & Parker; repr. (1872) with a new preface and notes and an additional sermon as *Fifteen Sermons preached before the University of Oxford, between A.D. 1826 and 1843*, London: Rivingtons, in the collected uniform edition of Newman's works in 36 vols, 1868-81, all the volumes of which from 1886 were published by Longmans, Green & Co. of London.(All these editions are out of print, but there have been modern paperback reprints.)

Newman, J.H. (1870) *An Essay in Aid of a Grammar of Assent*, London: Burns, Oates.(In the uniform edition cited above. There is a modern critical edition by I.T. Ker, Oxford: Clarendon Press, 1985, as well as various paperback reprints.)

Newman, J.H. (1969-70) *The Philosophical Notebook of John Henry Newman*, ed. E.J. Sillem, Louvain: Nauwelaerts, 2 vols.(The first volume, a somewhat wooden and wordy introduction to Newman's philosophy, includes a survey of the background influences. Although the notebook is fragmentary and was never intended for publication, it contains some penetrating and suggestive observations that are sometimes uncannily similar to the gnomic utterances of Wittgenstein, who appreciated Newman's originality.)

Newman, J.H. (1976) *The Theological Papers of John Henry Newman on Faith and Certainty*,ed. H.M. de Achaval, S.J. and J.D. Holmes, Oxford: Clarendon Press.(Exploratory papers on faith and certainty in preparation for the *Grammar of Assent*.)

References and further reading

Ferreira, M.J. (1980) *Doubt and Religious Commitment: The Role of the Will in Newman's Thought*, Oxford: Clarendon Press.(The best full-length book on certain aspects of Newman's philosophy. Readable but philosophically demanding.)

Hick, J. (1967) *Faith and Knowledge*, Ithaca, NY: Cornell University Press, 69-91. (Straightforward introduction to Newman's main ideas.)

Kenny, A. (1990) 'Newman as a Philosopher of Religion', in *Newman: A Man for our Time*, ed. D. Brown, London: SPCK, 98-122.(Useful exposition of Newman's ideas in their context.)

Ker, I.T. (1985) Introduction to *An Essay in Aid of a Grammar of Assent*, Oxford: Clarendon Press.(Includes survey and discussion of secondary literature.)

Ker, I.T. (1990) *The Achievement of John Henry Newman*, Notre Dame, IN: University of Notre Dame Press; London: Collins, ch. 2.(Succinct introduction to Newman's philosophy.)

Mitchell, B. (1990) 'Newman as a Philosopher', in *Newman after a Hundred Years*, ed. I. Ker and A.G. Hill, Oxford: Clarendon Press, 223-46.(An extremely readable and important account of Newman's philosophical significance.)

Owen, H.P. (1969) *The Christian Knowledge of God*, London: Athlone Press. (Contains stimulating critical discussions of Newman's ideas of the illative sense, probability and assent. Highly readable.)

Price, H.H. (1969) *Belief*, London: Allen & Unwin; New York: Humanities Press, 130-56, 315-48.(Two lucid,

vigorous chapters devoted to Newman’s theory of assent.)

Newton, Isaac (1642-1727)

Newton is best known for having invented the calculus and formulated the theory of universal gravity - the latter in his Principia, the single most important work in the transformation of natural philosophy into modern physical science. Yet he also made major discoveries in optics, and put no less effort into alchemy and theology than into mathematics and physics.

Throughout his career, Newton maintained a sharp distinction between conjectural hypotheses and experimentally established results. This distinction was central to his claim that the method by which conclusions about forces were inferred from phenomena in the Principia made it 'possible to argue more securely concerning the physical species, physical causes, and physical proportions of these forces'. The law of universal gravity that he argued for in this way nevertheless provoked strong opposition, especially from such leading figures on the Continent as Huygens and Leibniz: they protested that Newton was invoking an occult power of action-at-a-distance insofar as he was offering no contact mechanism by means of which forces of gravity could act. This opposition led him to a tighter, more emphatic presentation of his methodology in the second edition of the Principia, published twenty-six years after the first. The opposition to the theory of gravity faded during the fifty to seventy-five years after his death as it fulfilled its promise on such issues as the non-spherical shape of the earth, the precession of the equinoxes, comet trajectories (including the return of 'Halley's Comet' in 1758), the vagaries of lunar motion and other deviations from Keplerian motion. During this period the point mass mechanics of the Principia was extended to rigid bodies and fluids by such figures as Euler, forming what we know as 'Newtonian' mechanics.

1 Life

Isaac Newton entered Trinity College Cambridge in 1661, and began investigations of mathematics in 1664. These investigations culminated two years later in the binomial theorem and the fundamentals of the calculus. During the so-called *annus mirabilis* of 1666, while the university was closed because of the plague, and in the years immediately following, he extended his mathematical work; he also conducted optical experiments and worked on several basic problems in mechanics, including impact and circular motion. He became Lucasian Professor of Mathematics at Cambridge in 1669.

Although some of Newton's mathematical manuscripts were in circulation, yielding him some renown, his only notable publications before the *Principia* were a series of communications in the *Philosophical Transactions* of the Royal Society from 1672 to 1676 on his experiments on light and colours and the reflecting telescope. The debate which this work provoked led Newton to begin articulating what he called his 'experimental philosophy', which focused on *establishing* propositions by means of experiment.

In an exchange of letters in late 1679, Robert Hooke, himself an eminent scientist, asked Newton to use his mathematical methods to determine the trajectory of a body under a combination of inertial motion and an inverse-square force directed towards a central point - that is, the force Newton later named 'centripetal'. But the intense effort that culminated in the publication of the *Principia* (1687) did not begin until 1684, after a visit from young Edmond Halley, who later became Astronomer Royal.

Newton spent most of the years after 1689 in London. He was elected to represent Cambridge University in Parliament in 1689 and again in 1701, the year in which he resigned his professorship. He became Warden of the Mint in 1696, and Master of it in 1699. In 1703 he became President of the Royal Society, a post he held until he died. He was knighted in 1705.

During his London years Newton engaged in an acrimonious dispute with Leibniz over who had priority for inventing the calculus. One element fuelling this dispute was Newton's failure to publish his work, save for a three-page summary of a handful of results in Book II of the *Principia*. His first formal publications on the calculus appeared in 1704, when two earlier manuscripts were included as supplements in the first edition of the *Opticks*. (A Latin edition of the *Opticks* appeared two years later.)

Newton gave some thought to a restructured edition of the *Principia* in the early 1690s. But the second edition was not published until 1713, after four years of effort under the constructively critical eye of its editor, Roger Cotes. A third edition followed in 1726. These editions sharpened the contrast between his approach and that of Leibniz and

the Cartesians. The second English edition of the *Opticks* (1717/18) included Queries that summarized his conjectures on atomism. These Queries end with a concise statement of his method for establishing scientific knowledge on the basis of experiment and induction; so too does his final riposte in the priority dispute with Leibniz, his anonymous 'An Account of the Book Entitled *Commercium epistolicum*'.

2 Experimental philosophy in the light and colours debate

Newton's 1672 paper on light and colours reported only a small fraction of the optical experiments he had conducted. The debate it initiated concerned what the reported experiments had established. According to Newton, these experiments had conclusively shown that the oblong shape of the image cast by sunlight that has passed through a round hole and has then been refracted by a prism is caused by sunlight's consisting of rays that are refracted in different degrees by the prism. (The correspondence between these different refrangibilities and different colours led Newton to invent the first reflecting telescope, which eliminates the problems of chromatic aberration that had marred the refractive telescopes of the era.)

Hooke, interpreting Newton as claiming that the experiments established a corpuscular theory of light, insisted that a wave theory could account for the results just as well. Newton responded that the hypothesis that light is a body was put forward only as a conjecture suggested by the experiments, and not as part of what he claimed to have been established by them. He granted that Hooke's wave hypothesis could explain the conclusion the experiments had established; but this conclusion spoke of light only abstractly as 'rays' propagating in straight lines from luminous bodies, with no commitment to any specific 'mechanical' hypothesis.

His Dutch contemporary, Christiaan Huygens argued that Newton had failed to show the nature and difference of colours because he had offered no 'hypothesis by motion' to explain them. Newton responded that he ' never intended to shew, wherein consists the nature and difference of colours, but only to shew that *de facto* they are original and immutable qualities of the rays which exhibit them' (1958: 144).

Newton's contemporaries had trouble understanding his attempt to construe light rays abstractly in a way that would allow experiments to decide claims about them - this, independently of any mechanical account of light. In his replies, Newton outlined how, according to his experimental philosophy, diligently establishing properties of things by experiment takes precedence over framing hypotheses to explain them. Yet he also made clear that the propositions he regarded as conclusively established by experiment were nevertheless subject to correction based on detailed criticism of the experimental reasoning that had established them or on further experimental results challenging them (see Optics §§1-2).

3 Space, time and the laws of motion

Two aspects of the *Principia* provoked philosophical controversy in the decades following its publication: first, the appeal to absolute space and motion, and second, the insistence on establishing properties of gravity, especially its universality, without appeal to any mechanical hypothesis that could begin to explain how gravity is produced.

The *Principia* opens with two short sections, 'Definitions' and 'Axioms, or Laws of Motion', that have drawn philosophical fire ever since. The distinctions between absolute and relative time, space and motion are drawn in the first of these, following his introduction of the concept of mass and definitions pertaining to quantity of motion and force. For Newton, the distinction between absolute (or true) and relative (or apparent) motion is primary, and the parallel distinctions concerning space and time serve mostly to support this one (see Mechanics, classical; Space §§2-3). Newton was acutely aware of the empirical difficulties raised by such distinctions:

> It is certainly very difficult to find out the true motions of individual bodies and actually to differentiate them from the apparent motions, because the parts of that immovable space in which the bodies truly move make no impression on the senses. Nevertheless, the case is not utterly hopeless. For it is possible to draw evidence partly from apparent motions, which are differences between the true motions, and partly from the forces that are the causes and effects of the true motions …. But in what follows, a fuller explanation will be given of how to determine true motions from their causes, effects, and apparent differences, and, conversely, how to determine from motions, whether true or apparent, their causes and effects. For this was the purpose for which I composed the following treatise.

(Newton 1687: Scholium to Definitions)

The reference here is to the laws of motion and their corollaries, which immediately follow this last remark, as well as to the ninety-eight demonstrated Propositions of Book I and the fifty-three of Book II.

Like Descartes, Newton appealed to forces to distinguish true from apparent motions. And, again like Descartes, the true motion of greatest importance to him in the sequel is curvilinear motion, most notably the true motion of the planets that would distinguish between their equivalent relative motions in the Copernican and Tychonic systems. Unlike Descartes, however, Newton refused to offer hypotheses concerning the forces in question (see Descartes, R. §11). He required that the forces be inferred from phenomena with the help of the mathematical principles of Books I and II, many of which licence inferences from observed motions to measures of force. Inconsistencies among the inferred quantities of force or the motions subsequently inferred from them would indicate a failure to be dealing with true motions. But this is an empirical question, to be decided by carrying out the investigation of motions under forces to its fullest extent, insisting on no less than complete agreement between observed and calculated motions. Thus, the successes, and also the limitations, of the appeal to absolute space, time and motion were, for Newton, empirical issues that the long-term development of an exact science of motion would decide, and not something he thought was open to a priori resolution.

4 Inferences from phenomena and rules of natural philosophy

The Propositions of Books I and II are powerful resources for establishing conclusions about forces from phenomena of motion. For example, according to Propositions 1 and 2, Kepler's area rule holds if and only if the force acting on the moving body is centripetal. A corollary adds that the areal velocity is increasing when the force is off-centre in the direction of motion and decreasing when it is in the opposite direction. *The variation of the areal velocity is thus a measure of the direction of the force*. Similar systematic dependencies are involved in the inferences from Kepler's $3:2$ power rule and the absence of discernible orbital precession to the inverse-square variation of celestial centripetal forces (see Kepler, J.).

Rules of reasoning, which in the second and third editions are singled out at the beginning of Book III under the title *Regulae philosophandi*, strengthen the inferences that can be drawn from phenomena by licensing inductive generalizations (see Scientific method §2). The first two rules, for example, underlie the inference that the force holding the moon in orbit is terrestrial gravity - this, on the basis of the inverse-square relation between the centripetal acceleration of the moon and the acceleration of gravity at the earth's surface. The third rule, appearing for the first time in the second edition, supports the inference that all bodies gravitate towards each planet with weights proportional to their masses - this, on the basis of pendulum experiments and the common acceleration of Jupiter and its satellites toward the sun.

The fourth rule authorizes the practice of treating propositions that are supported properly by reasoning from phenomena as 'either exactly or very nearly true notwithstanding any contrary hypotheses, until yet other phenomena make such propositions either more exact or liable to exceptions'. It was added in the third edition to justify treating universal gravity as an established scientific fact in the face of complaints that it was unintelligible without an explanation of how it results from mechanical action by contact. This rule, and the related discussion of hypotheses at the end of the General Scholium added in the second edition, distinguish Newton's experimental philosophy most sharply from the mechanical philosophy of his critics.

5 Gravity as a universal force of interaction

The systematic dependencies by means of which Keplerian phenomena become measures of celestial forces are one-body idealizations. Universal gravity entails interactions among bodies, producing perturbations that require corrections to the Keplerian phenomena. The *Principia* includes a successful treatment of two-body interactions and some promising, though limited, results on three-body effects in lunar motion. But the full significance of the inferences that could be drawn from universal gravity did not become clear until Clairaut's analysis of the precession of the lunar orbit in 1749, and Laplace's determination in 1785 of the roughly 880-year 'Great Inequality' in the motions of Jupiter and Saturn.

Such successful treatments of perturbations do more than provide corrections to Keplerian phenomena. They also show that Newtonian measurements of inverse-square centripetal forces continue to hold to high approximation in the presence of perturbations. Interactions with other bodies account for the precessions of all the planets except

Mercury. Even in the case of Mercury, the famous 43 seconds-of-arc-per-century residual in its precession yields -2.000000157 as the measure of the exponent, instead of the exact -2 measured for the other planets. That such a small discrepancy came to be a problem at all testifies to the extraordinary level to which Newton's theory of gravity demonstrates an ideal of empirical success.

6 Mathematics

Newton engaged in extensive mathematical research throughout much of the period from 1664 until he left for London in 1696, and even after that he produced new results on some problems. Besides the many lines along which he developed and applied the calculus, he made substantial discoveries in algebra and in pure as well as analytic geometry. The mathematics of the *Principia* is itself a new form of synthetic geometry, incorporating limits. (Contrary to the still-persisting myth, there is no evidence that Newton first derived his results on celestial orbits within the symbolic calculus and then recast them in geometric form.)

Newton's invention of the calculus grew out of his attempts to solve several distinct problems, often employing novel extension of the ideas and methods of others. For example, his initial algorithms for derivatives of algebraic curves combined Cartesian techniques with the idea of an indefinitely small, vanishing increment. He exploited a method of indivisibles that John Wallis had used in obtaining integrals of algebraic curves; but he reconceptualized the method to represent an integral that grows as the curve extends incrementally, and then joined this with the binomial series to yield integrals via infinite series of a much wider range of curves. Geometrical representations of these results revealed the inverse relation between integration and differentiation. Then, adapting his Lucasian predecessor Isaac Barrow's idea to treat curves as arising from the motion of a point, Newton recast the results on derivatives, replacing indefinitely small increments with his 'fluxions'. The first full tract on fluxions, 'To Resolve Problems by Motion', is dated 1666, but the first manuscript to circulate was *De analysi per aequationes infinitas* of 1669.

Mathematicians in England used Newtonian methods and notation into the nineteenth century. But the Leibnizian tradition had gained so much momentum by the time Newton's works appeared that the calculus, as we know it, stems far more from that tradition. Ironically, it was such figures as the Bernoullis and Euler (belonging to the Leibnizian tradition) who recast the *Principia* into the language of the calculus.

7 Studies in alchemy and theology

Newton's unpublished manuscripts contain voluminous studies on alchemy, theology, prophecy and Biblical chronology. His alchemical work led to a number of elaborate chemical experiments carried out from the mid-1670s until 1693. His notes from these efforts display his great skill as an experimenter, but they appear to include nothing that would have altered the course of chemistry had they become public at the time (see Alchemy).

He first became preoccupied with theology in the early 1670s, probably in response to the requirement that he accept ordination to retain his Trinity fellowship. (He was granted a dispensation in 1675.) By 1673 he had rejected the doctrine of the Trinity and concluded that Christianity had become a false religion through a corruption of the Scriptures in the fourth and fifth centuries. He returned to these studies in subsequent decades, especially in the last years of his life. During his lifetime, however, he conveyed his radical views to only a few. They became widely known only when *Observations upon the Prophecies* was published in 1733.

Recent investigations of the alchemical and theological writings suggest that Newton's hope in natural philosophy was to look through nature to see God, that it was to be part of a larger investigation that would give meaning to life. His engagement with these larger issues may have helped him to free himself from the narrower restraints of the mechanical philosophy.

See also: Cosmology §2; Field theory, classical; Relativity theory, philosophical significance of; Space §2; Theories, scientific

WILLIAM L. HARPER
GEORGE E. SMITH

List of works

Newton, I. (1687) *Philosophiae naturalis principia mathematica*, London: Joseph Streater (for the Royal Society);

2nd edn, Cambridge, 1713; 3rd edn, London: Guil. & Joh. Innys (for the Royal Society), 1726.(Newton's monumental contribution, forever transforming science.)

Newton, I. (1715) 'An Account of the Book Entitled *Commercium epistolicum*', *Philosophical Transactions* 29 (342): 173-224.(Newton's anonymous review of the Royal Society's findings in the dispute over whether he or Leibniz had priority for inventing the calculus. Reprinted in Hall 1980.)

Newton, I. (1730) *Opticks or a Treatise of the Reflections, Refractions, Inflections and Colours of Light*, 4th edn; New York: Dover, 1952.(Based on the fourth edition of 1730. Newton's other major contribution to science.)

Newton, I. (1733) *Observations upon the Prophecies of Daniel, and the Apocalypse of St John*, London; repr. W. Whitla, *Sir Isaac Newton's Daniel and the Apocalypse with an Introductory Study… of Unbelief, of Miracles and Prophecy*, London: John Murray, 1922.(Published posthumously, Newton's most influential work in theology.)

Newton, I. (1958) *Isaac Newton's Papers and Letters on Natural Philosophy and Related Documents*, ed. I.B. Cohen, assisted by R.E. Schofield, Cambridge, MA: Harvard University Press.(Contains the publications in the dispute on light and colours as they appeared in the *Philosophical Transactions* of the Royal Society. Second revised edition 1978.)

Newton, I. (1959-77) *The Correspondence of Isaac Newton*, eds H.W. Turnbull, A. Scott, A.R. Hall and L. Tilling, Cambridge: Cambridge University Press, 7 vols.(Newton's correspondence from 1661 to 1727, including correspondence by others and related papers.)

Newton, I. (1962) *Unpublished Scientific Papers of Isaac Newton*, eds A.R. Hall and M.B. Hall, Cambridge: Cambridge University Press; repr. 2nd edn, 1978.(Contains several early papers in mechanics as well as the first full tract on the calculus, 'To Resolve Problems by Motion'.)

Newton, I. (1965) *The Background to Newton's Principia: A Study of Newton's Dynamical Researches in the Years 1664-84*, ed. J.W. Herivel, Oxford: Clarendon Press.(Contains most early papers and fragments leading to the *Principia*, with commentary and analysis.)

Newton, I. (1967-81) *The Mathematical Papers of Isaac Newton*, ed. D.T. Whiteside, Cambridge: Cambridge University Press, 8 vols.(Newton's mathematical manuscripts from 1664 to 1722, authoritatively annotated in invaluable detail.)

Newton, I. (1972) *Isaac Newton's Philosophiae naturalis principia mathematica* (from the 3rd edn of 1726) with variant readings, eds A. Koyré and I.B. Cohen, with assistance of A. Whitman, Cambridge, MA: Harvard University Press, 2 vols.(Includes not only all variants from the first two editions, but also those from Newton's manuscripts and annotated copies.)

Newton, I. (1983) *Certain Philosophical Questions: Newton's Trinity Notebook*, eds J.E. McGuire and M. Tamny, Cambridge: Cambridge University Press.(Newton's notebook from his early years at Cambridge, displaying his scientific style before it had matured.)

Newton, I. (1984) *The Optical Papers of Isaac Newton*, vol. 1, ed. A.E. Shapiro, Cambridge: Cambridge University Press.(The first volume of Newton's extensive unpublished writings on optics, detailing his numerous experiments.)

Newton, I. (1995) *Newton - A Norton Critical Edition*, eds I.B. Cohen and R.S. Westfall, New York: W.W. Norton.(A selection of Newton's writings from all areas of his work, with commentaries that extend across four centuries.)

Newton, I. (1997) *Isaac Newton's Mathematical Principles of Natural Philosophy*, trans. I.B. Cohen and A. Whitman, Los Angeles: University of California Press.(The first entirely new English translation of Newton's *Principia* since Andrew Motte's of 1729.)

References and further reading

Bricker P. and Hughes R.I.G. (eds) (1990) *Philosophical Perspectives on Newtonian Science*, Cambridge, MA: MIT Press.(Among the many excellent collections inspired by the 300th anniversary of the first edition of the *Principia*, this one is especially focused on philosophy.)

Cohen, I.B. (1974) 'Isaac Newton', in Ch. Gillispie (ed.) *Dictionary of Scientific Biography*, New York: Charles Scribner's Sons, vol. 10, 42-103.(A comparatively brief, but authoratitive survey of Newton's life and work.)

Cohen, I.B. (1980) *The Newtonian Revolution*, Cambridge: Cambridge University Press.(An analysis of Newton's style in science, as exemplified in the *Principia*.)

Dobbs, B.J.T. (1990) 'Newton as Alchemist and Theologian', in N.J.W. Thrower (ed.) *Standing on the Shoulders of Giants: A Longer View of Newton and Halley*, Berkeley, CA: University of California Press.(An extensive discussion of the topics of §7.)

Earman, J. (1989) *World Enough and Space-Time: Absolute versus Relational Theories of Space and Time*,

Cambridge, MA: MIT Press.(This together with Stein’s paper in Palter 1971 is good background for §3.)

Gjertsen, D. (1986) *The Newton Handbook*, London: Routledge & Kegan Paul.(Contains a complete listing of Newton’s works and an extensive bibliography of secondary sources.)

Hall, A.R. (1980) *Philosophers at War: The Quarrel Between Newton and Leibniz*, Cambridge: Cambridge University Press.(The history of the priority dispute over the calculus; includes Newton 1715.)

Hall, A.R. (1992) *Isaac Newton: Adventurer in Thought*, Oxford: Oxford University Press.(The best concise biography of Newton.)

Harper, W.L. (1997) ’Isaac Newton on Empirical Success and Scientific Method’, in J. Earman and J. Norton (eds) *The Cosmos of Science*, Pittsburgh, PA: University of Pittsburgh Press.(This paper gives further development of material in §§3-5.)

Harper, W.L. and Smith, G.E. (1995) ’Newton’s New Way of Inquiry’, in J. Leplin (ed.) *The Creation of Ideas in Physics: Studies for a Methodology of Theory Construction*, Dordrecht: Kluwer.(This paper expands on material in §§2, 4-5.)

Harrison, J. (1978) *The Library of Isaac Newton*, Cambridge: Cambridge University Press.(A listing of the books in Newton’s personal library.)

Koyré, A. (1968) *Newtonian Studies*, Chicago, IL: University of Chicago Press.(Classic material especially for §§ 3, 4.)

Palter, R. (ed.) (1971) *The Annus Mirabilis of Sir Isaac Newton 1666-1966*, Cambridge, MA: MIT Press.(This classic collection contains papers still relevant to most sections.)

Stein, H. (1991) ’From the Phenomena of Motions to the Forces of Nature: Hypothesis or Deduction?’, in A. Fine, M. Forbes and L. Wessels (eds) *PSA 1990*, East Lansing, MI: Philosophy of Science Association, vol. 2, 209-22.(Relevant to material in §§4-5.)

Theerman, P. and Seeff A.F. (eds) (1993) *Action and Reaction*, Newark, DE: University of Delaware Press. (Includes papers relevent to §7 as well as to most other sections.)

Westfall, R.S. (1980) *Never at Rest: A Biography of Isaac Newton*, Cambridge: Cambridge University Press.(The most thorough intellectual biography of Newton, also available abridged as *The Life of Isaac Newton*, 1993.)

Whiteside, D.T. (1991) ’The prehistory of the *Principia* from 1664-1686’, in *Notes and Records, The Royal Society of London* 45: 11-61.(A review of the steps from Newton’s earliest investigations of motion to the first edition of the *Principia*.)

Wilson, C. (1989) ’The Newtonian Achievement in Astronomy’, in R. Taton and C. Wilson (eds) *Planetary Astronomy from the Renaissance to the Rise of Astrophysics*, Cambridge: Cambridge University Press.(A critical summary of the contributions Newton’s *Principia* made to planetary astronomy.)

Nichiren (1222-82)

Fiery prophet, religious reformer, founder of a major religious movement, brilliant preacher and erudite writer, Nichiren is one of Japan's most controversial religious figures. His thought derived from the Tendai tradition, from which he inherited both his veneration of the Lotus Sutra and the idea of an eternal Buddha woven into the stuff of all reality.

Born on the southeastern seacoast of Japan, the son of a fisherman, Nichiren entered a local Buddhist monastery at eleven and was ordained four years later with the monastic name Rencho. He subsequently studied Pure Land and Zen in Kamakura, and Shingon on Mount Kōya. In 1253 he returned to his home temple, changed his name to Nichiren (Sun Lotus), and announced that he would devote his life to ridding Buddhism of the heresy and corruption that had overtaken it. Like many in the Kamakura period, he believed that Buddhism was entering the age of *mappō* (latter days of the Dharma, a period of ten thousand years beginning with the third millennium after the death of the Buddha), which only a faith of the utmost purity could hope to survive.

For Nichiren, the loadstone of that purity was the *Lotus Sutra*, the chanting of whose name (*Namu-myōhō-rengekyō*) and living out of whose teachings alone would insure salvation. Proclaiming himself the embodiment of the original *bodhisattva*, Jōgyō, who had first received that scripture from the Buddha himself, Nichiren took on himself the mission of saving the soul of Japan.

Traditionally Buddhism recognized two means of spreading the Dharma: accommodation (*shōju*) and confrontation (*shakabuku*). In the age of *mappō*, only the latter would do, and Nichiren set out to demonstrate this. He rejected Zen meditation, Pure Land *nenbutsu*, Shingon rituals, and Ritsu (*vinaya*) discipline as diabolical. So sweeping were his condemnations that some contemporary scholars have even questioned whether the sects that venerate him as founder deserve to be called Buddhist at all. Loyalty to the *Lotus Sutra* took precedence over everything else in his life and provoked one conflict after another with civil and religious authorities.

As Nichiren saw it, belief in 'provisional' Buddhas, central to the first half of the *Lotus Sutra*, ended up in superstition for the ordinary faithful and abstract debate for the scholars. Only single-minded faith in the reality of the one, absolute Buddha, central in the second half of the *sūtra*, could awaken the world from its dark ignorance, and only the recitation of the sacred name of the *sūtra* could bridge the gap between the two. Further, since he was persuaded that Japan was destined to replace India as the centre from which Buddhism would spread to the rest of the world, he insisted that the scripture be read not only with the lips but with the whole body: that is, that it lead to the transformation of society. Perhaps the most systematic statement of Nichiren's thought appears in a work entitled *Kanjin honzon shōomacr;((A Clear Mind on Supreme Reality)* and in the graphic portrayal of his *Great Mandala*, both completed in exile in the year 1273.

See also: Buddhist philosophy, Japanese

J.W. HEISIG

List of works

Nichiren (1273) *Kanjin honzon shōomacr;((A Clear Mind on Supreme Reality)*, ed. P.B. Yampolsky and B. Watson, in *Selected Writings of Nichiren*, New York: Columbia University Press, 1990.(This edition is part of a reliable and easily accessible selection of Nichiren's writings.)

References and further reading

Anesaki Masaharu (1966) *Nichiren: The Buddhist Prophet*, Gloucester, MA: Peter Smith.(The classic source in English for the life and thought of Nichiren, written in easy prose for the non-specialist.)

Matsunaga, A. and Matsunaga, D. (1988) *Foundation of Japanese Buddhism*, Tokyo: Buddhist Books International, vol. 2, 137-81.(Includes a reliable and objective account of Nichiren's major teachings.)

Nicholas of Autrecourt (*c.*1300-69)

Unlike most of his late medieval contemporaries, Nicholas of Autrecourt did not subscribe to Aristotelianism. Instead, he radically challenged the foundations of Aristotle's metaphysics and epistemology by asking two questions: first, how are we supposed to explain the basic constituents of the world, given that the Aristotelian categories are mere theoretical constructions and not immediately perceivable? Second, what can we know with absolute certitude, given that sense perception - the starting point in Aristotelian epistemology - is fallible and unreliable? Focusing on these two questions, Nicholas elaborated an atomistic metaphysics and defended an epistemology that emphasizes knowledge of a logical principle (the principle of non-contradiction), dismissing all knowledge based on inductive reasoning as uncertain.

1 Life and works

Nicholas of Autrecourt (also Autricuria or Ultricuria), a scholastic theologian and philosopher, was born around 1300 in the diocese of Verdun. He studied at the University of Paris around 1330 and lectured there on Peter Lombard's *Sentences*. In 1340 he was summoned to Avignon to answer charges of heresy. His trial was interrupted by the death of Pope Benedict XII but was resumed under Pope Clement VI, by Cardinal William Curti. In 1346, the papal jury concluded that Nicholas was unworthy to continue his academic career. He was forced to recant many of his philosophical and theological theses and was expelled from the university. In 1347 his works were publicly burned. His appointment as deacon at the cathedral of Metz in 1350 may be seen as an indication that he was no longer considered a heretic. He probably died in Metz in 1369.

Only a small portion of Nicholas' works has been transmitted. Extant works include two complete letters addressed to the Franciscan Bernard of Arezzo, mostly dealing with epistemological problems, along with fragments of seven other letters to Bernard; a letter written to a certain Giles, also concerned with epistemological issues; the treatise *Exigit ordo executionis (The Universal Treatise)*, surviving in a single incomplete manuscript, which is inspired by a decisive anti-Aristotelianism and focuses on metaphysics and natural philosophy; and a short theological question. In addition, the condemned articles contain important information about some of Nicholas' semantic theses.

2 Epistemology

Nicholas' main goal was to defend the possibility of evident, certain knowledge against Bernard of Arezzo's sceptical challenges. According to Bernard, one can never infer '*x* exists' from '*x* is perceived and thought of', because one may have a mere thought object that is not immediately caused by or founded upon a real, actually existing object. For instance, one may think about Caesar although Caesar is not actually existing and may never have existed, or one may have a perception of whiteness although there is no white thing in reality.

In order to reject Bernard's conclusion that no certain knowledge is possible, Nicholas maintained that there is one kind of absolutely certain knowledge (apart from the certitude of faith): knowledge of the principle of non-contradiction (de Rijk 1994: 58). This is the first principle in a negative sense because nothing is prior to it, and the first principle in a positive sense because all other certain knowledge is reducible to it. One can have certain knowledge of an implication (*consequentia*), Nicholas claimed, only if the consequent is identical with or part of the antecedent. Otherwise it would not be evident that the antecedent is inconsistent with the denial of the consequent. For instance, one can have certain knowledge of 'If the house exists, then the wall exists' only if the wall is a part of the house: otherwise it would not be evident that 'If the house exists, then one of its parts does not exist' violates the principle of non-contradiction. Nicholas held that we have to confine ourselves to this basic knowledge if we aim at certitude. From the fact that one thing is known to exist, the existence or non-existence of another thing can never be known with absolute certitude. Nor can one have any certain knowledge based on inductive reasoning: 'In the past, my hand was warmed when I put it toward the fire; therefore it will be warmed now if I put it toward the fire' cannot be known with certitude. My present action is not identical with or part of my past action, and no violation of the principle of non-contradiction results if I say 'Although in the past my hand was warmed when I put it toward the fire, it will not be warmed now when I put it toward the fire'.

It is clear that Nicholas' strict limitation to knowledge of the first principle involves a radical critique of theories of

causality and of induction. Strictly speaking, one can never know that a certain effect *e* is brought about by a cause *c*, and one can never predict that *e* will be brought about if *c*, not hindered by any impediment, will be active. All one can do is have a sense perception of *c* and *e*, and then evaluate whether or not the two contradict each other. Because of this radical claim, Nicholas has been labelled ‘a medieval Hume’ by some modern commentators (Rashdall 1906-7). Yet it is important to see that Nicholas had no sceptical intentions. In pointing out that absolutely evident, though narrowly limited knowledge is possible, he in fact aimed to refute scepticism.

3 Metaphysics

Humans would live better lives, Nicholas polemically claimed, if only they would abandon their studies of Aristotle (de Rijk 1994: 154-5). Aristotle and his successors posited many entities that are mere theoretical constructions and cannot be known with certitude. All we can immediately experience are (1) the objects of our five exterior senses, and (2) our own acts. It is not evident on the basis of (1) that there are entities belonging to the Aristotelian categories (substances, qualities and so on); all we can say is that we have sense experience of constantly changing objects. Nor is it evident on the basis of (2) that there is a soul with three powers. We are only entitled to say that we are aware of inner acts having certain functions (thinking, willing and so on). We have no justification for introducing any concrete or abstract entities beyond the objects of our external or internal experience.

Nicholas also harshly attacked the Aristotelian thesis that there is generation and corruption of things. We do not know with certitude that substances or qualities come into and go out of existence; all we know is that things we perceive appear to change. This appearance can be explained in terms of an atomistic theory, wherein the world consists of eternally existing particles which constantly change their constellation; when things appear to change, or to come into or go out of existence, there is nothing but aggregation and separation of the atomic particles (see Atomism, ancient).

In this manner, Nicholas broke radically from the then prevailing Aristotelian metaphysics. He used atomism as a decisive tool in his natural philosophy, in particular in his theory of the continuum. Contrary to many Aristotelians, he held that a continuum is not composed of infinitely divisible parts, but that it has atomic particles as its ultimate constituents. Although radical, however, this thesis was not completely new; indivisibilism can be traced back to several early fourteenth-century authors including Henry of Harclay, Walter Chatton and Gerard of Odo.

4 Semantics

No work documenting Nicholas’ semantics has been transmitted, but some of his most controversial theses can be reconstructed on the basis of the condemned articles. Article 1.1 states: ‘I have said and written that this sentence "Man is an animal" is not necessary according to faith, not paying attention for the moment to the necessary linkage between the foresaid terms’ (de Rijk 1994: 146). This article is clearly connected with a famous late medieval sophism that raises the problem of whether the truth of a sentence requires the actual existence of the things referred to by the terms of the sentence. Some authors (for example, Roger Bacon and Boethius of Dacia) gave an affirmative answer: if no man exists, they claimed, the term ‘man’ has no reference, and if there is no referent, the sentence is false. Therefore the sentence is not necessarily true. Others (such as Siger of Brabant and Robert Grosseteste) held that the terms in this sentence refer not to individual men and animals but to the eternally existing species and genus; therefore ‘man’ always has a referent, even if no man actually exists, and the sentence is necessarily true. Nicholas obviously chose the former explanation, but this was taken to be heretical (and indeed had already been condemned at Oxford in 1277). Contrary to Christian doctrine, the sentence ‘Christ is a man’ is not necessarily true if one claims that terms exclusively refer to actually existing things. During the three days in the tomb Christ was not an actually existing man; therefore the term ‘man’ did not have a reference during this period and the sentence was temporarily false.

Another important semantic issue is raised in Article 16.4: ‘[I have said] that what is signifiable in a complex way (*complexe significabile*) through the sentence "God and creature are distinct" is nothing (*nihil*)’ (de Rijk 1994: 160). This article raises the question of whether the entire sentence signifies an entity that is not reducible to any other entity. Does ‘God and creature are distinct’ signify the state of affairs that God and creature are distinct, something reducible neither to the extramental things (that is, God, creature and their relation) nor to the concepts

of these things? Nicholas clearly gave a negative answer. A state of affairs is nothing: there is no third entity in addition to extramental and mental entities. Thus, Nicholas did not deny the existence of God and creature, as his judges may have thought; rather, he rejected the claim (supported by Gregory of Rimini and other Parisian authors) that a state of affairs is a peculiar kind of thing.

The semantic theses show clearly that Nicholas tackled some of the most prominent problems of his time. Although a radical critic of the dominant Aristotelianism and excluded from the intellectual centre, Nicholas was not simply an academic outsider. Most of his criticisms are developed within the framework of prevailing philosophical doctrines, and attack either the framework itself (especially in epistemology and metaphysics) or some widely accepted theses within it.

See also: Aristotelianism, medieval; Epistemology, history of; Language, medieval theories of; Natural philosophy, medieval

DOMINIK PERLER

List of works

Nicholas of Autrecourt (*c.*1300-69) *Exigit ordo executionis (The Universal Treatise)*, ed. J.-R. O'Donnell in 'Nicholas of Autrecourt', *Mediaeval Studies* 1, 1939: 179-280; trans. L. Kennedy, R. Arnold and A. Millward, *The Universal Treatise of Nicholas of Autrecourt*, Milwaukee, WI: Marquette University Press, 1971.(This edition is taken from the single incomplete manuscript which has survived.)

Nicholas of Autrecourt (*c.*1300-69) Letters, ed. L.M. de Rijk, *Nicholas of Autrecourt: His Correspondence with Master Giles and Bernard of Arezzo*, Leiden: Brill, 1994.(Critical edition with English translation and extensive commentary.)

References and further reading

de Rijk, L.M. (ed.) (1994) *Nicholas of Autrecourt: His Correspondence with Master Giles and Bernard of Arezzo*, Leiden: Brill.(Critical edition of Nicholas' letters with English translation, accompanied by an extensive commentary.)

Imbach, R. and Perler, D. (1988) *Nicolaus von Autrecourt: Briefe (Nicholas of Autrecourt: Letters)*, Hamburg: Meiner.(Contains a reprint of the condemned articles as well as an edition of Nicholas' letters.)

Lappe, J. (1908) *Nicolaus von Autrecourt, sein Leben, seine Philosophie, seine Schriften (Nicholas of Autrecourt, His Life, Philosophy and Writings)*, Beiträge zur Geschichte der Philosophie des Mittelalters 6.2, Münster: Aschendorff.(First edition of all the letters and condemned articles, and also a pioneering study of Nicholas' life and intellectual context.)

Rashdall, H. (1906-7) 'Nicholas de Ultricuria, a Medieval Hume', *Proceedings of the Aristotelian Society* 7: 1-27.(Pioneer study of Nicholas' alleged scepticism.)

Scott, T.K. (1971) 'Nicholas of Autrecourt, Buridan, and Ockhamism', *Journal of the History of Philosophy* 9: 15-41.(Discussion of Nicholas' epistemology in the light of the Ockhamist and Anti-Ockhamist movements in Paris.)

Tachau, K.H. (1988) *Vision and Certitude in the Age of Ockham: Optics, Epistemology and the Foundations of Semantics 1250-1345*, Leiden: Brill.(Chapters 11-12 focus on Nicholas' condemnation in the context of the 1340s and shed new light on the anti-sceptical strategy.)

Thijssen, J.M.M.H. (1990) 'The "Semantic Articles" of Autrecourt's Condemnation', *Archives d'histoire doctrinale et littéraire du moyen-âge* 65: 155-75.(Detailed analysis of most semantic articles; interesting comparisons with fourteeth-century debates.)

Weinberg, J. (1948) *Nicholas of Autrecourt*, Princeton, NJ: Princeton University Press.(Most comprehensive, still important study that takes into account both Nicholas' epistemology and his metaphysics.)

Nicholas of Cusa (1401-64)

Also called Nicolaus Cusanus, this German cardinal takes his distinguishing name from the city of his birth, Kues (or Cusa, in Latin), on the Moselle river between Koblenz and Trier. Nicholas was influenced by Albert the Great, Thomas Aquinas, Bonaventure, Ramon Llull, Ricoldo of Montecroce, Master Eckhart, Jean Gerson and Heimericus de Campo, as well as by more distant figures such as Plato, Aristotle, Proclus, Pseudo-Dionysius and John Scottus Eriugena. His eclectic system of thought pointed in the direction of a transition between the Middle Ages and the Renaissance. In his own day as in ours, Nicholas was most widely known for his early work De docta ignorantia (On Learned Ignorance). In it, he gives expression to his view that the human mind needs to discover its necessary ignorance of what the Divine Being is like, an ignorance that results from the infinite ontological and cognitive disproportion between Infinity itself (that is, God) and the finite human or angelic knower. Correlated with the doctrine of docta ignorantia *is that of* coincidentia oppositorum in deo*, the coincidence of opposites in God. All things coincide in God in the sense that God, as undifferentiated being, is beyond all opposition, beyond all determination as* this *rather than* that*.*

Nicholas is also known for his rudimentary cosmological speculation, his prefiguring of certain metaphysical and epistemological themes found later in Leibniz, Kant and Hegel, his ecclesiological teachings regarding the controversy over papal versus conciliar authority, his advocacy of a religious ecumenism of sorts, his interest in purely mathematical topics and his influence on the theologian Paul Tillich in the twentieth century. A striking tribute to Nicholas' memory still stands today: the hospice for elderly, indigent men that he caused to be erected at Kues between 1452 and 1458 and that he both endowed financially and invested with his personal library. This small but splendid library, unravaged by the intervening wars and consisting of some three hundred volumes, includes manuscripts written in Nicholas' own hand.

1 Biography

Of Nicholas' early life, little is known. He was one of four children born to Johan Krebs and Katharina Roemer Krebs. Still unconfirmed is the speculation that his father, a successful ferryman and hauler of merchandise up and down the Moselle, sent him to Deventer in Holland for part of his primary education. As for higher learning, Nicholas studied the liberal arts at the University of Heidelberg during 1416-7. Subsequently, he transferred to the University of Padua, where he came under the influence of the Italian humanists and acquired familiarity with the latest physical and mathematical theories. At Padua he studied principally canon law, receiving his *doctor decretorum* in 1423. In the spring of 1425 he continued his study of canon law at the University of Cologne where, apparently, he also gave lectures and where Heimericus de Campo introduced him to the ideas of Ramon Llull and Pseudo-Dionysius. In 1425 and 1426 he held benefices from the archbishopric of Trier, and by 1427 he was secretary to the archbishop himself, Otto von Ziegenhain. In 1428 and again in 1435 he was offered a chair in canon law at the University of Louvain, but on both occasions he declined, preferring to continue with his canonical and administrative work on behalf of the archdiocese and - at the Council of Basel - with his work as a member of the Council's committee on matters of faith.

During his 'Basel period' (1432-8, not all of which time he was in Basel) he wrote, among other things, *De communione sub utraque specie (On Communion Using Both the Bread and the Wine)*, against the Hussites, in 1433; *De concordantia catholica (The Catholic Concordance)* in 1433 and *De auctoritate praesidendi in concilio generali (On Presidential Authority in a General Council)* in 1434, both of which concern the demarcation of papal authority; and *De reparatione kalendarii (On Amending the Calendar)* in 1436, a proposal for revising the Julian calendar. At Basel he met John of Torquemada, whose treatise against the Muslims, *Tractatus contra principales errores perfidi Machometi* (Tractate Against the Principal Errors of the Infidel Muhammad) (1459), was to influence his own *Cribratio Alkorani (A Scrutiny of the Qur'an)* in 1461. Also affecting his response to the incursion of Islam into the West were his first-hand experiences of Islam in 1437, while present in Constantinople for two months (24 September-27 November). The delegation to which he belonged had been sent by the minority party of the Council of Basel to encourage both the Byzantine emperor and the Greek patriarch to reunite with the Roman patriarchate. It was on his return voyage from Constantinople, while still at sea, that Nicholas came to embrace the notion of learned ignorance.

Sometime between 1436 and 1440, Nicholas was ordained a priest. On 20 December 1448 he was named cardinal, and to him was assigned, on 3 January 1449, the titular church of St Peter ad Vincula in Rome. Some fifteen months thereafter (23 March 1450) he was designated bishop of Brixen in Tirol, where later his attempts to introduce monastic and diocesan reforms engendered enmity with Archduke Sigismund, whose threats forced him twice to flee to Rome. His death came on 11 August 1464 in Todi, Italy, as he was travelling from Rome to Ancona.

2 Two basic themes

The themes of *docta ignorantia* and *coincidentia oppositorum* were misunderstood by Nicholas' contemporary and adversary John Wenck of Herrenberg, who in *De ignota litteratura* (On Unknown Learning) inveighed against Nicholas' treatise *De docta ignorantia (On Learned Ignorance)*. In a response, *Apologia doctae ignorantiae (A Defence of Learned Ignorance)*, Nicholas made clear that the doctrine of learned ignorance is not intended to deny knowledge of the existence of God but only to deny all knowledge of God's nature (except on the part of God himself). All our discourse about God is metaphorical; moreover, when we are conceiving positively of God, we are obliged to conceive of him in accordance with those metaphors that signify the highest perfections. The doctrine of *docta ignorantia* goes on to teach that the quiddities of created objects are unknowable by us precisely; that is, they are unknowable exactly as they are in themselves. Yet, although the objects are unknowable exactly *in themselves*, the *objects themselves* are not unknowable. Rather, they are apprehensible by us *in coniectura*, Nicholas' expression for what is knowable only imperfectly and perspectivally. The doctrine of learned ignorance constitutes a radical break with the scholastic affirmation of a real, but remote, resemblance between God and his creation. The late scholastic William of Ockham had repudiated the validity of all alleged analogical knowledge of God's nature, though he nonetheless maintained that the predicate 'being' is used univocally in discourse about divine being and created being. By contrast, Nicholas regards even 'being' as an equivocal and metaphorical predicate when used of God.

The doctrine of *coincidentia* does not mean, as in John Wenck's construal of it, that the universe coincides with God, nor does it serve to contravene the Aristotelian principle of non-contradiction within the legitimate domain of that principle's application - that is, its application to the realm of finite affairs. All things are present in God ontologically prior to their creation, as illustrated by an effect's being present in the priority of its cause. But just as, in its cause, an effect is the cause, so in God all things are God rather than being their differentiated, finite selves. Nicholas nowhere teaches, unqualifiedly, that all things are God; indeed, he expressly rejects pantheism and its variants. Though he speaks of God as not *other* than (*non aliud a*) the universe, he does not mean that God is identical with (*non aliud quam*) the universe, but rather that God transcends the domain of comparison with all finite objects.

3 Ontology

Nicholas' ontology has been interpreted along two very different lines. According to one line of interpretation, Nicholas regards finite objects as having no respective positive essence of their own, for God himself is the self-identical essence of each finite thing. Accordingly, the relative identity of each thing is said to consist only in its difference from all other finite things, all such differences being accidental and all relations between these things being internal relations. Thus in this view, propounded in Europe by Heinrich Rombach (1965) and in North America by Thomas P. McTighe (1964), created things are not substances and do not differ from one another substantially. According to the second, more generally accepted line of interpretation - one taking its lead from Nicholas' *Apologia* - some created things are substances, and these do have positive essences of their own. In accordance with their respective essences, finite things differ essentially from one another. God is the Essence of these things in that he is the Essence of the respective essences of these things, even as he is *being* itself (*entitas*), which sustains in existence whatever is *a* being (*ens*). This distinction between *being* and *beings* Nicholas draws from Meister Eckhart; and it is one of the distinctions that influence both Paul Tillich and Martin Heidegger.

In *De filiatione Dei (On Being a Son of God)* 3, Nicholas' likening of all created beings to mirrors that mutually reflect one another prefigures Leibniz's systematic use of a similar theme (see Leibniz, G.W.). Also parallel with tenets in Leibniz are the assertions (1) that each kind of thing - though not each individual within that kind - is as perfect as that kind of thing can be and (2) that physical objects (which, for Leibniz, are 'well-founded appearances') are, in principle, infinitely divisible. Nicholas' various utterances regarding *nihil* (not-being) in

relation to *esse* (being) have struck some interpreters as inchoately Hegelian, in a dialectical sense (see Hegelianism). Others have seen Nicholas not as working with logico-ontological categories but as being a metaphysical nominalist. In the last analysis, his theory of universals (if not his theory of transcendentals) must be reckoned to be a moderate realism that is cognate with that of Aristotle and Aquinas.

4 Cosmology

At times Nicholas denies that the universe is finite and also denies that it is infinite; at other times he asserts that it is both finite and infinite. It is not finite, he tells us, in that it is not physically limited by anything outside its dimensions; on the other hand, it is finite in that it has a definite measure, known to God alone. In other words, the universe is finite but unbounded; yet, Nicholas terms it *privatively* infinite, in contradistinction to God, who is *negatively* infinite, since he is not even conceivably limited.

Because there is no perfect precision in the world, explains Nicholas, the universe's circumference is not perfectly spherical, nor is the earth a perfect sphere. Consequently, neither the universe nor the earth has an exact physical centre and, thus, the earth is not the fixed centre of the universe. Indeed, the earth is not stationary but, like all stars and planets, has a circular movement of its own, though this movement is undetectable by us, its inhabitants. Likewise, the earth appears to us to be at the centre of the universe, and we infer that the 'fixed' stars are at the outer circumference. However, these stars also have inhabitants, who will judge themselves to be at the centre and the earth to be on the periphery. Furthermore, both the earth and the moon, as also the other heavenly bodies, emit a light and a heat of their own, even though both the earth and its moon also reflect the sun's light and heat. Nicholas' cosmological claims do not anticipate those of Copernicus, but they do deviate in important respects from those of Aristotle and Ptolemy.

5 Epistemology

In *De visione Dei (The Vision of God)* 24, Nicholas states that 'there cannot be in the intellect anything which is such that it was not first in the senses', thus rejecting the notion of innate concepts. Yet, in *Idiota de mente (The Layman on the Mind)* and elsewhere, he allows that the mind, through construction, can make concepts of things (such as a spoon) whose forms are not found in nature. Moreover, the mind has an innate power of judgment (*vis iudiciaria*) by which to weigh the strength of rational considerations and through which it is familiar, a priori, with basic moral principles (*Compendium* 10). Although human beings cannot know the exact quiddity of any given thing, they do nevertheless perceive the things themselves, through the medium of sensory images and conceptual forms. In places, Nicholas emphasizes (1) sensory and conceptual perspectivity, (2) the role of reason (*ratio*) as synthesizer and (3) the conditioning operations of the intellect (*intellectus*). (One must be careful, however, not to exaggerate his dim prefiguring of Kant.) Mathematical knowledge he regards as a priori and certain, because mathematical entities are conceptual entities elicited from, and known in a precise way by, our reason (*De possest (On Actualized-possibility)* 43-4). In *Idiota de mente* 6 he seems to endorse the doctrine that these conceptual numbers are 'images', as it were, of God's thought. The mathematical puzzle with which he is principally preoccupied is that of how to square the circle: in other words, how in a finite number of steps to construct, using only an unmarked straight-edge and a drawing compass, a square whose area is equivalent to that of a given circle.

6 Christology

Nicholas' Christology is orthodox, inasmuch as he considers Jesus to be the incarnate second member of the Trinity, uniting in his divine person his divine nature with a human nature. In accordance with the divine nature, Jesus is the Absolute Maximum; in accordance with his human nature, he is the contracted maximum individual. The Absolute Maximum is Infinity itself and is such that nothing greater than it can be conceived; the contracted maximum is an individual creature that is of maximum perfection within its kind. (By 'contracted', Nicholas means restricted, differentiated or delimited, so as to be *this* thing and not *that* thing.) Thus Jesus' human nature is the perfection of the possibility of human nature. Furthermore, since human nature, which consists of a body and a rational soul, is a middle nature - a nature in between the higher intelligible natures (namely, angels) and the lower corporeal natures (namely, brute animals and non-animate things) - Jesus' perfect human nature is 'the fullness of all the perfections of each and every thing'. Thus, reminiscent of Eriugena's *De divisione naturae*, the perfect world that went forth from God in creation is reunited to God through the subsumption of Christ's human nature in

his divine nature.

7 Ecclesiology

In *De concordantia catholica*, Nicholas seeks to demarcate the respective authority of a universal council and of the pope. A universal council (in contrast to a patriarchal council) is one to which bishops and select ecclesiastics from all five patriarchates have been called - a call issued usually by a patriarch or a patriarchal council, or an agent of either. A universal council may depose a heretical pope. A decision (on matters of faith) reached with a high degree of agreement by a universal council is more trustworthy than would be a decision on such matters by the pope alone. Yet, on matters of faith, the pope's consent as well as the council's is required for declaring a decision infallible. Nicholas speaks of degrees of infallibility, and he relates these to degrees of consensus: the greater the consensus within a universal council and beween this council and the pope, the greater the infallibility of the decision reached. Nicholas later was to abandon his claim that the authority of a universal council in which there is substantial agreement is superior to the pope's authority.

8 Ecumenism

The dialogue *De pace fidei (On Peaceful Unity of Faith)* serves to challenge the various leaders of the different religions to adopt *religio una in rituum varietate*, a common religion that admits of diverse rites. The spirit of this 'ecumenism' is reinforced in *Cribratio Alkorani*, where Nicholas seeks to find common ground between Christianity and Islam. Yet Nicholas' is not a genuine ecumenism, for he aims to show how other religions can be led to discern the truth of the doctrines essential to Christianity while retaining many of their own diverse rituals and liturgies. Through his programme of learned ignorance, he envisions the possibility of worshipping a God who, as infinite, is neither trine nor one in any sense in which trineness and oneness are positively conceivable by us. By thus adapting Eckhart's distinction between God and Godhead, Nicholas can regard the doctrine of the Trinity as no longer a stumbling block to Jews and Muslims.

9 Conclusion

Nicholas' system of thought is often misapprehended, sometimes in a way that detracts from his real genius, sometimes in a way that is all too flattering of his intellectual accomplishments. A proper grasp of his ideas will locate him, no doubt, somewhere between the invective of John Wenck, who condemns him as a corrupter of Aristotelianism, and the praise of Heinrich Rombach, who sees him as a new Aristotle for the modern period, as someone whose anti-substantive, functionalist ontology is as revolutionary as was Aristotle's introduction of hylomorphism in the first place.

See also: Cosmology; Christology and Incarnation; Ontology; Tillich, P.

JASPER HOPKINS

List of works

1 Major Latin editions

Nicholas of Cusa (1433-64) *Collected Works*, Strasbourg: Martin Flach, 1488. (Based upon codices Latini Cusani 218 and 219. Newly reprinted as *Werke*, ed. Paul Wilpert, Berlin: Walter de Gruyter, 1967, 2 vols.)

Nicholas of Cusa (1433-64) *Collected Works*, Paris: Jacques Le Fèvre d'Etaples, 1514; repr. Frankfurt: Minerva, 1962.(Adds to the Strasbourg edition *De concordantia catholica*, *De deo abscondito*, *Coniectura de ultimis diebus*, certain mathematical *opuscula* and excerpts from the *Sermones*. Contains many editorial emendations. The later Basel edition of 1565 follows the wording of the Paris edition but adds several mathematical works; like previous editions, it lacks *De li non aliud*.)

Nicholas of Cusa (1433-64) *Collected Works*, Heidelberg Academy of Letters edition, Leipzig/Hamburg: Meiner, 1932-.(Series remains in progress. The most reliable of the editions.)

Nicholas of Cusa (1445-59) *Mathematical Writings*, trans. J. Hofmann, *Die mathematischen Schriften*, Hamburg: Meiner, 2nd edn, 1979.(A German translation of the mathematical works, together with an introduction, extensive notes and important manuscript information.)

2 Individual works

Nicholas of Cusa (1433) *Opusculum contra Bohemorum errorem: De usu communionis (A Short Work Against the Error of the Hussites: On the Celebration of Communion)*, in the Paris edition of *Collected Works*, vol. II, 2, ff. 5^{r}-13^{v}.(This is a single work, though the Paris edition divides it into two works.)

Nicholas of Cusa (1433) *De maioritate auctoritatis sacrorum conciliorum supra auctoritatem papae (On the Superiority of the Authority of the Sacred Councils to the Authority of the Pope)*, text ed. E. Meuthen, printed in Abhandlungen der Heidelberger Akademie der Wissenschaften, Philosophisch-historische Klasse, 1977, Abhandlung 3.(One of Nicholas' early works, on papal authority.)

Nicholas of Cusa (1433) *De concordantia catholica (The Catholic Concordance)*, trans. P. Sigmund, Cambridge: Cambridge University Press, 1991.(On the demarcation of papal authority.)

Nicholas of Cusa (1433) *Libellus inquisitionis veri et boni (A Short Book Investigating the True and the Good).* (No longer extant.)

Nicholas of Cusa (1433 or 1434) *De modo vero habilitandi ingenium ad discursum in dubiis (On the Correct Manner of Training the Mind for Reasoning about Doubtful Matters)*.(No longer extant.)

Nicholas of Cusa (1434) *De auctoritate praesidendi in concilio generali (On Presidential Authority in a General Council)*, trans. H.L. Bond *et al.*, *Church History* 59, 1990: 19-34.(On the demarcation of papal authority.)

Nicholas of Cusa (1436) *De reparatione kalendarii (On Amending the Calendar)*, text ed. V. Stegemann, in *Die Kalenderverbesserung. De correctione kalendarii*, Schriften des Nikolaus von Cues, Heidelberg: Kerle, 1955. (Nicholas' proposals for reforming the calendar.)

Nicholas of Cusa (1440) *De docta ignorantia (On Learned Ignorance)*, trans. J. Hopkins, *Nicholas of Cusa on Learned Ignorance*, Minneapolis, MN: Banning, 1985.(Nicholas' best-known work, setting out his doctrine of 'learned ignorance'.)

Nicholas of Cusa (1441-2) *De coniecturis (On Surmises)*, trans. W.F. Wertz, *Toward a New Council of Florence*, Washington, DC: Schiller Institute, 1993.(Recent translation.)

Nicholas of Cusa (1444 or earlier) *De Deo abscondito (On the Hidden God)*, trans. J. Hopkins, *A Miscellany on Nicholas of Cusa*, Minneapolis, MN: Banning, 1994.(Recent translation.)

Nicholas of Cusa (1445) *De quaerendo Deum (On Seeking God)*, trans. J. Hopkins, *A Miscellany on Nicholas of Cusa*, Minneapolis, MN: Banning, 1994.(Recent translation.)

Nicholas of Cusa (1445) *De filiatione Dei (On Being a Son of God)*, trans. J. Hopkins, *A Miscellany on Nicholas of Cusa*, Minneapolis, MN: Banning, 1994. (In Book 3, Nicholas likens all created beings to mirrors.)

Nicholas of Cusa (1445) *De geometricis transmutationibus (On Geometrical Transformations)*, in the Paris edition of *Collected Works*, vol. II, 2, ff. 33^{r}-53^{v}.(Edition of mathematical work.)

Nicholas of Cusa (1445) *De arithmeticis complementis (On Arithmetical Complements)*, in the Paris edition of *Collected Works*, vol. II, 2, ff. 54^{r}-58^{v}.(Edition of mathematical work.)

Nicholas of Cusa (1445-6) *De dato patris luminum (The Gift of the Father of Lights)*, trans. J. Hopkins, *Nicholas of Cusa's Metaphysic of Contraction*, Minneapolis, MN: Banning, 1983. (Recent translation.)

Nicholas of Cusa (1446) *Coniectura de ultimis diebus (A Surmise Regarding the Last Days)*, in the Heidelberg Academy edition of *Collected Works*, vol. IV, *Opuscula*. (Reliable edition.)

Nicholas of Cusa (1447) *Dialogus de genesi (On the Genesis [of All Things])*, trans. J. Hopkins, *A Miscellany on Nicholas of Cusa*, Minneapolis, MN: Banning, 1994.(Nicholas' writings on creation.)

Nicholas of Cusa (1449) *Apologia doctae ignorantiae (A Defence of Learned Ignorance)*, trans. J. Hopkins, *Nicholas of Cusa's Debate with John Wenck*, Minneapolis, MN: Banning, 1988. (Includes text and translation of Wenck's *De ignota litteratura (On Unknown Learning)*.)

Nicholas of Cusa (1450) *Idiota de sapientia (The Layman on Wisdom)*, trans. J. Hopkins, *Nicholas of Cusa on Wisdom and Knowledge*, Minneapolis, MN: Banning, 1996.(Further work on epistemology.)

Nicholas of Cusa (1450) *Idiota de mente (The Layman on Mind)*, trans. J. Hopkins, *Nicholas of Cusa on Wisdom and Knowledge*, Minneapolis, MN: Banning, 1996.(Nicholas allows that the mind can make concepts of things whose forms are not found in nature.)

Nicholas of Cusa (1450) *Idiota de staticis experimentis (The Layman on Experiments Done with Weight-Scales)*, trans. J. Hopkins, *Nicholas of Cusa on Wisdom and Knowledge*, Minneapolis, MN: Banning, 1996.(Edition of mathematical work.)

Nicholas of Cusa (1450) *De circuli quadratura (On Squaring the Circle)*, Munich, Staatsbibliothek, Codex Latinus Monacensis 14213, ff. 101^{r}-104^{v}. (Edition of mathematical work.)

Nicholas of Cusa (1450) *Quadratura circuli (Squaring the Circle)*, in the Basel edition of Collected Works, 1091-95.(Includes what the Basel edition entitles *De sinibus et chordis*.)

Nicholas of Cusa (1452) *Tres epistolae contra Bohemos (Three Letters against the Hussites)*, in the Strasbourg edition of Collected Works, vol. II, 674-97.(The Strasbourg, Paris and Basel editions have the wrong ordering.

Corrected order: Letter I is in the Strasbourg edition, vol. II, 675-9, and Paris edition, vol. II, 2, ff. 14ʳ-15ᵛ. Letter II is the Strasbourg edition, vol. II, 679-82, and Paris edition, vol. II, 2, ff. 15ᵛ-16ᵛ. Letter III is in the Strasbourg edition, vol. II, 674-97, and Paris edition, vol. II, 2, ff. 13ᵛ-22ʳ. Letter III contains, as inserted material, the first two letters. See H. Hallauer, 'Das Glaubengespräch mit den Hussiten', *Mitteilungen und Forschungsbeiträge der Cusanus-Gesellschaft* 9, 1971: 72-3.)

Nicholas of Cusa (1453) *De pace fidei (On Peaceful Unity of Faith)*, trans. J. Hopkins, *Nicholas of Cusa's De Pace Fidei and Cribratio Alkorani*, Minneapolis, MN: Banning, 1994.(Recent translation.)

Nicholas of Cusa (1453) *De visione Dei (The Vision of God)*, trans. J. Hopkins, *Nicholas of Cusa's Dialectical Mysticism*, Minneapolis, MN: Banning, 1988.(Recent translation.)

Nicholas of Cusa (1453) *De mathematicis complementis (On Mathematical Complements)*, in the Strasbourg edition of Collected Works, vol. II, 388-430.(A second, fuller version was produced in 1454.)

Nicholas of Cusa (1453) *De theologicis complementis (Complementary Theological Considerations)*, trans. J. Hopkins, *Nicholas of Cusa: Metaphysical Speculations*, Minneapolis, MN: Banning, 1997. (Recent translation.)

Nicholas of Cusa (1454?) *Declaratio rectilineationis curvae (An Explication of the Straightening of a Curved Line)*, in the Basel edition of the Collected Works, 1100-1. (Edition of mathematical work.)

Nicholas of Cusa (1454?) *De una recti curvique mensura (On a Single Measurement of the Straight and the Curved)*, in the Basel edition of the Collected Works, 1101-6.(Edition of mathematical work.)

Nicholas of Cusa (1457) *Dialogus de circuli quadratura (Dialogue on Squaring a Circle)*, in the Basel edition of the Collected Works, 1095-8.(Pages 1099-1100 contain a letter from Master Paul to Nicholas.)

Nicholas of Cusa (1457) *De caesarea circuli quadratura (The Imperial Squaring of a Circle)*, Milan, Biblioteca Ambrosiana, Codex Latinus Mediolanensis G 74 inf., ff. 3ʳ-4ʳ. (Edition of mathematical work.)

Nicholas of Cusa (1458) *De beryllo (On [Intellectual] Eye-Glasses)*, trans. J. Hopkins, *Nicholas of Cusa: Metaphysical Speculations*, Minneapolis, MN: Banning, 1997.(Translation of mathematical work.)

Nicholas of Cusa (1458) *De mathematica perfectione (On Mathematical Perfection)*, in the Strasbourg edition of Collected Works, vol. II, 698-709.(Edition of mathematical work.)

Nicholas of Cusa (1459) *De aequalitate (On Equality)*, trans. J. Hopkins, *Nicholas of Cusa: Metaphysical Speculations*, Minneapolis, MN: Banning, 1997.(Recent translation.)

Nicholas of Cusa (1459) *De principio (On the Beginning)*, trans. J. Hopkins, *Nicholas of Cusa: Metaphysical Speculations*, Minneapolis, MN: Banning, 1997.(Recent translation.)

Nicholas of Cusa (1459) *Aurea propositio in mathematicis (The Golden Proposition in Mathematics)*, Milan, Biblioteca Ambrosiana, Codex Latinus Mediolanensis G 74 inf., ff. 2ᵛ-3ʳ. (Edition of mathematical work.)

Nicholas of Cusa (1460) *De possest (On Actualized-possibility)*, trans. J. Hopkins, *A Concise Introduction to the Philosophy of Nicholas of Cusa*, Minneapolis, MN: University of Minnesota Press, 1978; 3rd edn, Minneapolis, MN: Banning, 1986.(Reliable translation.)

Nicholas of Cusa (1461) *Cribratio Alkorani (A Scrutiny of the Qur'an)*, trans. J. Hopkins, *Nicholas of Cusa's De Pace Fidei and Cribratio Alkorani*, Minneapolis, MN: Banning, 1994.(Recent translation of Nicholas' writings against Islam.)

Nicholas of Cusa (1461 or 1462) *De li lon lliud (On Not-other)*, trans. J. Hopkins, *Nicholas of Cusa on God as Not-other*, Minneapolis, MN: University of Minnesota Press, 1979; 3rd edn, Minneapolis, MN: Banning, 1987.

Nicholas of Cusa (1462) *De figura mundi (On the Shape of the World)*.(No longer extant.)

Nicholas of Cusa (1462-3) *De ludo globi (The Game of Spheres)*, trans. P. Watts, *Nicholas de Cusa. De Ludo Globi. The Game of Spheres*, New York: Abaris Books, 1986. (Translation of mathematical work.)

Nicholas of Cusa (1462 or 1463) *De venatione sapientia (On the Pursuit of Wisdom)*, trans. J. Hopkins, *Nicholas of Cusa: Metaphysical Speculations*, Minneapolis, MN: Banning, 1997. (Recent translation.)

Nicholas of Cusa (1464) *Compendium*, trans. J. Hopkins, *Nicholas of Cusa on Wisdom and Knowledge*, Minneapolis, MN: Banning, 1996. (Recent translation.)

Nicholas of Cusa (1464) *De apice theoriae (Concerning the Loftiest Level of Contemplative Reflection)*, trans. J. Hopkins, *Nicholas of Cusa: Metaphysical Speculations*, Minneapolis, MN: Banning, 1997.(Recent translation.)

References and further reading

Euler, W. (1990) *Unitas et Pax. Religionsvergleich bei Raimundus Lullus und Nikolaus von Kues (Unity and Peace: Ramon Llull's and Nicholas of Cusa's Comparison of Religions)*, Würzburg: Echter Verlag.(A careful study of Llull's influence on Cusa.)

Hagemann, L. (1976) *Der Kur'ān in Verständnis und Kritik bei Nikolaus von Kues. Ein Beitrag zur Erhellung islamisch-christlicher Geschichte (The Qur'an as Understood and Criticized by Nicholas of Cusa: A*

Contribution to the Illumination of Islamic-Christian History), Frankfurt: Knecht. (Major study of Nicholas' *Cribratio Alkorani* and of his approach to Islam.)

Harries, K. (1975) 'The Infinite Sphere: Comments on the History of a Metaphor', *Journal of the History of Philosophy* 13: 5-15.(Explores Nicholas' emphasis on perspective and his use of the metaphor of God as a sphere whose centre is everywhere and whose circumference is nowhere.)

Haubst, R. (1991) *Streifzüge in die cusanische Theologie (Forays into Nicholas of Cusa's Theology),* Münster: Aschendorff.(A collection of Haubst's major essays analyzing Nicholas' theological views.)

Haubst, R. *et al.* (eds) (1961-) *Mitteilungen und Forschungsbeiträge der Cusanus-Gesellschaft (Reports and Research-Contributions of the Cusanus Society),* Mainz: Mattias-Grünewald.(A continuing series; bibliographies in vols. I, III, VI, X, XV.)

Jacobi, K. (1969) *Die Methode der cusanischen Philosophie (Cusanus' Philosophical Method),* Munich: Alber. (An extensive survey of previous interpretations of Cusa; Jacobi himself is influenced by H. Rombach.)

McTighe, T. (1964) 'Nicholas of Cusa and Leibniz's Principle of Indiscernibility', *Modern Schoolman* 42: 33-46. (Gives the gist of one competing interpretation of Nicholas' ontology.)

Meuthen, E. (1979) *Nikolaus von Kues, 1401-1464. Skizze einer Biographie (Nicholas of Cusa 1401-1464: A Biographical Sketch),* 4th edn, Münster: Aschendorff.(Provides, concisely, the essential facts about Nicholas' life.)

Meuthen, E. and Hallauer, H. (eds) (1976, vol. I, fascicle 1, and 1983, fascicle 2) *Acta Cusana. Quellen zur Lebensgeschichte des Nikolaus von Kues*, Hamburg: Meiner. (Other volumes in progress. A wealth of crucial details concerning Nicholas' life and works.)

Rombach, H. (1965) *Substanz, System, Struktur. Die Ontologie des Funktionalismus und der philosophische Hintergrund der modernen Wissenschaft (Substance, System, Structure: Functionalist Ontology and the Philosophical Background of Modern Science)*, vol. I, Freiburg: Alber.(Contains an extreme, but extremely intriguing, construal of Cusa's metaphysics as a metaphysics of functionalism.)

Sigmund, P. (1963) *Nicholas of Cusa and Medieval Political Thought*, Cambridge, MA: Harvard University Press.(A balanced examination of Nicholas' theory of conciliar and papal authority.)

Vansteenberghe, E. (1920) *Le Cardinal Nicolas de Cues (Cardinal Nicholas of Cusa),* Paris; repr. Frankfurt: Minerva, 1963.(An early monumental, pathbreaking work.)

Niebuhr, Helmut Richard (1894-1962)

After early writings interpreting the social dimensions of US religious life from a Protestant point of view, Helmut Richard Niebuhr came increasingly to focus on theology, the interpretation of the Word and experience of God, and ethics, the attempt to live a responsible life grounded in the deepest values. More successfully than any contemporary, he brought into creative tension two major strands of modern Christian thought. From the liberal tradition, he came to stress the relativity of all statements about God and accented the situation of the believer who makes them. Niebuhr matched or countered this perspectival approach with an inheritance from 'neo-orthodoxy' which stressed the otherness of God, the distance between God and all human experiences, statements and perspectives. While preoccupied with witness to 'God beyond the gods', Niebuhr was also devoted to understanding the human in the light of the experience of God. He developed comprehensive views of the responsible self as the focus of ethics and emphasized the communal dimensions of human life.

1 The sovereignty of God

Helmut Richard Niebuhr was born in Wright City, Missouri, on 3 September 1894. He acquired theological interests early as the son of a well-read small-town pastor; his brother, Reinhold Niebuhr, also became a prominent theologian. The sermons he heard as a child reflected German Protestant traditions, which were nurtured in the college and seminary life of the Evangelical and Reformed tradition. After receiving a PhD at Yale in 1924 and studying at Berlin and Marburg in 1930, he taught at Yale Divinity School (adjacent to a secular university), where seminarians from many denominations studied. His thought was never devoted to sectarian interests, but reflected and kept in mind the ecumenical context, and addressed the world of those who cared little for the Christian faith. He died on 5 July 1962.

Despite the broadness of Niebuhr's thought, many of its accents are developments of the Reformed background. In the tradition associated with John Calvin (§3), he witnessed to a God who is not the product of human imagination, not one among many deities, not merely immanent and able to be taken captive within human affairs, and not capable of compromising or being compromised. This is the 'God beyond the gods', not to be confused with the idols and deities manufactured by humans to justify their nations, Churches or other institutions. The ethical human being is to draw their values from reference to this sovereign God.

Niebuhr balanced this view of God, which in philosophical terms he called 'radical monotheism', with a very realistic view of the way human beings grasped the activity of God. God is indeed active in the world in a distinctive way through Jesus Christ; he is testified to in the Bible, engaged in human affairs through the life of the Church and manifestly holds humans responsible for actions that could be valued as good.

2 Social history and the kingdom of God

Not a biblical scholar, Niebuhr was frankly a philosophical theologian at a time when American Protestant elites often denigrated philosophy. He was not so much a precise or seminal philosophical thinker as someone who carried on a lifelong dialogue with philosophers, at first with social and political thinkers. This led him to write a sociological and then a historical interpretation of modern and especially American religious life. In *The Social Sources of Denominationalism* (1929), he displayed an almost deterministic view of what social class did to shape Churchly response. At that time he was strongly influenced by the German social philosopher Ernst Troeltsch, who led Niebuhr towards relativism, or at least towards the need to see the relativity of all human witness to God. He later wrote his own critique of this sociological determinism, *The Kingdom of God in America* (1937). In it, he accented the way the experience of God and thinking about God shaped the life of the American nation.

A third historical work, *Christ and Culture* (1951), saw him move towards a classifying system for relating the believing community to the culture around it. He tended to favour two views: 'Christ Transforming Culture', in which responsible humans in community set out to realize something of the coming kingdom of God; and 'Christ and Culture in Paradox', a dialectical view which showed Christians at times affirming and at times negating the main elements of culture. These reflect Niebuhr's own view of the limits of human achievement, and his belief in the need to be responsible within culture.

3 Theology and Jesus Christ

While writing these three more or less historical works, Niebuhr came to favour a more abstract and philosophical approach to theology. Through it all, he kept two poles of thought in the front of his mind. One was witnessed to in *Radical Monotheism and Western Culture* (1960), a judgment upon all efforts to reduce God to a cultural artefact or to claim God for favoured human causes. His last book, *The Responsible Self* (published posthumously in 1963), demonstrated that he had never left behind his preoccupation with conscience, community and responsibility. He died before being able to complete a synthetic work. Yet his life and teachings were sufficiently consistent for one to be able to deduce the elements of a complete systematic theology. Two generations of Niebuhr students went from Yale to strategic positions in theological schools, there to develop these poles of his thought.

As a Christian philosophical theologian, Niebuhr naturally focused on the figure of Jesus Christ. He connected Jesus chiefly with the redeeming activity of God. Niebuhr spent little time revisiting ancient creedal categories, largely shaped under Greek influence, in order to deal with the ways the human Jesus is to be regarded as a distinctive expression of divine action. He chose historical more than philosophical language for this purpose. Jesus was first of all a pious Jew, a person of his times. He was a proclaimer of the sovereignty of God, of radical monotheism focused on the one he called 'Father'. He was also a mediator of faith, who proclaimed and brought about for human beings a new relation to God.

4 Theory of value and the responsible self

Rather than make original contributions to what Christians call Christology, Niebuhr related Jesus to Niebuhrian theories of values and valuation. Jesus embodied the revelation of God and used his situation to offer a new covenant with God, a revolutionary impulse that called others who saw him as an exemplar of faith to change the world around them. One of Niebuhr's students, Hans Frei, summarized well the way Niebuhr could speak of 'the teaching and acts of Jesus Christ, his moral virtues' as offering 'the direct clue to his being'. In these, one may find 'in faith, hope and love, the unique moral Sonship to God of one who is completely at one with men' (Frei 1957: 115).

Niebuhr is usually described as a confessional theologian, that is, a thinker about God who interprets the 'confessions' or creeds of the believing community. In fact, he presented two perspectives. One is indeed 'internal', the 'we' language of the community. But there is also the perspective of the sociologist, historian or analyst, who looks at the community from the 'external' point of view and sees a very different sort of reality. It is this understanding that has led Niebuhr to be described, second, as a 'perspectival' thinker - not a mere relativist, but someone acutely conscious of the relativity of all human experience and discourse, the perspectives through which people interpret God's action.

On the basis of these interpretations, Niebuhr can be seen, third, as a philosophical theologian of 'value'. For him, the experience of God and the witness to God are not ends in themselves, having no connection with action in the public world. On the contrary, the witness to God should impel responsible selves to give expression to that which they value. God is the source and centre of value, and knowledge of God leads believers to patterns of loyalty that help them to develop appropriate scales of value with respect to the things of the world. For the believer, knowing is a matter that involves valuing. The act of valuing is not capricious, not a matter of mere preference, but is related to the deepest commitments of life. Being and value are inseparable in Niebuhr's thinking. While this could have meant 'Whatever is, is right', for Niebuhr it had to be 'Whatever is, is good', since 'Being' for him meant the Being of the creator God who is good and creates good. The person of faith becomes the one who knows, in a limited way at least, to trust Being, God, the source and centre of values. This view of values and valuation relates again to Niebuhr's lifelong interest in the God who holds human beings responsible, and in human beings who, for all their limits, should be responsive to God; this means that they should be bound in loyalty to the signals of the 'Creative Action of God', the 'Governing Action of God', and the 'Redemptive Action of God'.

Philosophers interested in thinkers who would prove the existence of God and describe the divine attributes would be frustrated upon encountering Niebuhr, who never tried to do either. In *The Meaning of Revelation* (1941), another major work, he describes the experience and confession of a revealing God who, in the ambiguities of history, summons people to trust and to responsible action. Niebuhr elaborated this perspective of the believing community and its 'internal' history as well as anyone in American Protestantism in his time, and remained a

consistent witness to radical monotheism.

MARTIN MARTY

List of works

Niebuhr, H.R. (1929) *The Social Sources of Denominationalism*, New York: Holt.(More sociological than theological. A categorization of social contexts that shape religious responses.)

Niebuhr, H.R. (1937) *The Kingdom of God in America*, New York: Harper & Row.(A theological interpretation of religion in the USA from a Reformed Protestant point of view.)

Niebuhr, H.R. (1941) *The Meaning of Revelation*, New York: Macmillan.(Reflects on the way believers respond to what they regard as a transcendent message.)

Niebuhr, H.R. (1951) *Christ and Culture*, New York: Harper & Bros.(A classic typology of the ways in which Christianity relates to its surrounding culture.)

Niebuhr, H.R. (1960) *Radical Monotheism and Western Culture*, New York: Harper & Bros.(A collection of essays on faith and intellectual life, centring on a rigorous call to monotheistic expression.)

Niebuhr, H.R. (1963) *The Responsible Self*, New York: Harper & Row.(An 'ethics of responsibility' for the individual in a complex society.)

References and further reading

Frei, H. (1957) 'The Theology of H. Richard Niebuhr', in P. Ramsey (ed.) *Faith and Ethics*, New York: Harper Torchbooks, 115.(A compressed early critique, focusing on Niebuhr's methods and classic Christian doctrines.)

Fowler, J.W. (1985) *To See the Kingdom: The Theological Vision of H. Richard Niebuhr*, New York: University Press of America.(A clear elaboration of Niebuhr's main theological themes.)

Irish, J.A. (1983) *The Religious Thought of H. Richard Niebuhr*, Atlanta, GA: John Knox Press.(A reliable popular interpretation, and the best introduction to the main themes of Niebuhr's thought.)

Ramsey, P. (ed.) (1957) *Faith and Ethics*, New York: Harper & Bros.(A collection of critical essays by colleagues.)

Niebuhr, Reinhold (1892-1971)

Reinhold Niebuhr is widely regarded as the foremost public theologian in twentieth-century America. A 'public' theologian is one who is responsive to the biblical tradition and responsible to the Christian Church, but who also responds to the social and political concerns of the world and seeks to effect change in that world. In the case of Niebuhr, this meant being attuned to life in the USA as it passed through times of prosperity, depression, war, cold war and cultural complacency. It also meant that he had to be as alert to the secular philosophy and expression of the times as to the biblical and creedal traditions of the Church, and he managed to correlate and connect these in ever-changing ways.

Niebuhr is known as a developer of a school of thought often called 'Christian realism'. Though shaped first by the more optimistic liberal thought of his teachers' generation, which stressed the immanence of God and the potential for goodness in human beings, Niebuhr came to witness to the otherness of God and the drastic limits of human potential. He even helped resurrect the term 'original sin' to describe the human condition, well aware that the term was scorned by most philosophers of his time. Yet over the decades, his realism came to be seen as so appropriate to descriptions of human actions, especially in situations of power, that he attracted a following far beyond Church communities. Niebuhr thus influenced both domestic and international policies. He was seen both as a self-critical theologian who uttered judgments on the Christian Church and the American nation, and as a theologian of the Cold War who issued devastating critiques of Soviet Communism. In the time before the Second World War, when most of the more notable Protestant clerics leaned towards pacifism, Niebuhr let his realistic vision and his opposition to totalitarian powers lead him to argue for 'preparedness' for war and to scold those who refused to see with him a need for military response to the threat of dictators.

1 A theology informed by public life

Reinhold Niebuhr, brother of the equally notable theologian H. Richard Niebuhr, was born in Wright City, Missouri, on 21 June 1892 and died in Stockbridge, Massachusetts, on 1 June 1971. He first learned Reformed Protestant theology from his German-born father, a small-town pastor. Schooled in the rather provincial sphere of his denomination's college and seminary, he moved on to Yale Divinity School and University and from there to a thirteen-year pastorate in industrial Detroit. His next move was to the interdenominational Union Theological Seminary in New York, where he taught from 1930 to 1960. During all these years, he expressed himself through thousands of editorials and articles in religious and secular publications alike, plunged actively into politics, and lectured and preached widely.

At the same time, Niebuhr kept schooling himself in theology and philosophy, though he made no secret of his limits as a philosopher or his lack of interest in its more abstract formulations. However, he was sufficiently disciplined and prestigious as a philosophical theologian to be invited to give the Gifford Lectures at the University of Edinburgh in 1939; these were published as his most systematic work, *The Nature and Destiny of Man* (1941, 1943). Yet for all the philosophical power and passion displayed in that work, Niebuhr came to be more noticed for his comments on the USA and world affairs. He made his mark with a contentious work published at the beginning of the Depression, *Moral Man and Immoral Society* (1932), a book that seemed to be intended to insult liberal theologians. For several years afterwards he was avowedly informed by Marxian critical ideas, though he never advocated radical socialist or communist programmes and soon passed critical judgment on those ideas as well. He developed his reasoned perspective on domestic and international affairs at mid-century in *The Irony of American History* (1952). Although hampered by the effects of a stroke, he continued to comment on the Cold War scene and on American culture for more than a decade after he could no longer teach.

2 The witness to God

The term 'theology' etymologically combines witness (*logos*) with God (*theos*). Theologians thus try to interpret the life of believing communities and the world around them not only in the light of experience, but also with reference to a transcendent God. Niebuhr conformed to this model, since his whole career was devoted to seeing public life, including that of the Churches, in the light of divine activity. His choice to be a theologian of the engaged sort showed an understanding of his own personality, talents and interests, but was also a sign of his limitations. He spent little or no energy on some classic philosophical themes such as the proofs for the existence

of God.

Instead, Niebuhr's God was the God witnessed to in the Hebrew Scriptures and the New Testament, the Bible of the Christian world. Uncommonly informed by Jewish thought at a time when many Christians went their separate way, Niebuhr had identified with rabbis in Detroit who shared his concern for social justice. With them, he explored the tradition of the Hebrew prophets, who, also using the transcendent reference to God to judge human affairs, usually put more energy into confronting their rulers and the religious powers around them than their enemies. Niebuhr's God was therefore the God of the prophets, the one who measures human affairs and finds them wanting.

In *The Irony of American History*, Niebuhr defined the relation of God to human power and potential as succinctly as he did anywhere. It was a self-consciously ironic - as opposed to tragic, comic or pathetic - understanding. He liked to quote Psalm 2: 4, in which the transcendent God is pictured as laughing at and even holding in derision earthly rulers and powers when they are over-proud and unmindful of their limits. At the same time, according to Niebuhr, this God was not witnessed to as being above human affairs or eager to demean and annihilate people. Instead, this God honoured human aspirations and held people responsible.

Niebuhr was often criticized for refusing to do basic theological work. He shunned concepts associated with ontology, metaphysics and epistemology, took much for granted and spent his energies witnessing to the God revealed in the Bible. He was also criticized for not having developed a sustained and sophisticated hermeneutic. There was an often-expressed American pragmatic bent to his use of the Bible. Thus Niebuhr argued that one is drawn to the Bible in human affairs because it seems to represent the most accurate and appropriate description of human nature and the mortal plight. It was a profoundly relevant address to the world he and the realists around him perceived.

Niebuhr's God was active in creating the world and sustaining it, and in judging human pride and pretension. But the same God was also present in history through revelation in Jesus Christ. Again, Niebuhr avoided agendas common among theologians in that he spent little time expounding the natures of the divine and the human in Jesus. He focused on Jesus' ethical proclamation, which he saw to be severe, 'an impossible possibility'. And he saw in the figure of Jesus crucified not only a love that redeemed humans, but also a sign of the evil of a world that could not be just and did not welcome love. So he could say to the 'atheists for Niebuhr', as some of the people who did not share his theology but identified with his social and political outlook (such as Arthur Schlesinger and Hans Morgenthau) waggishly called themselves, that for a full understanding of him and his thought they had to know that he preached Christ crucified, Jesus on the cross.

3 The nature of the human

While God can only be witnessed to, humans are accessible to social scientists, ordinary people making observations, and theologians who use divine perspectives to study them. It was on these grounds that Niebuhr suggested that 'original sin' was the only empirically observable Christian doctrine. However, after disagreements with philosophers such as John Dewey, who wanted to dismiss Niebuhr as a grouchy preacher of doom, he expressed regret for having retrieved the notion, or at least for having chosen the term 'original sin'. He did not want it to suggest how morally low humans could sink, or to shift responsibility from them to someone else (such as Adam in the Genesis myth). Instead, it was intended to signal that all humans have limits that cause them to express their pride and to rebel against God (see Feminist theology §1). These limits were never fully explained; the question as to where evil comes from was of the sort that Niebuhr preferred to hand over to the philosophers, though expecting little yield from their answers. Instead, Niebuhr insisted, the limits are simply 'there' and have to be reckoned with. And though all that humans do is tainted, their evil inclinations and pride do not leave them free to be irresponsible.

If human nature is marked by the sin of pride, the evil effects are most noted in collective life; hence the title of Niebuhr's *Moral Man and Immoral Society*. Less concerned with petty vice or personal errancy than with public life - he believed redemptive love to be at its most effective in more intimate spheres - Niebuhr asked what the particular evils are to which Churches, corporations, political parties and especially nations are prone. He saw little possibility of large-scale self-purification by societies and this led him to stress justice more than love in collective life. Societies have a way of turning even their good intentions into proud and self-interested purposes. For this

reason, Niebuhr came to believe in balances of power, of checking the set of interests of one group by countering them with those of another. It was this pragmatic realism, with its blunt readiness to use power, that so offended Christian and other pacifists who opposed Niebuhr.

4 Christian realism

To those philosophical and cultural critics who thought Niebuhr simply wrong in his assessment of human nature and strategically misguided because his pessimism could lead to passivity, the adjective 'Christian' could be interpreted in either of two ways. To some it was the key to explaining why Niebuhr had gone wrong. The Bible, from which he drew so many ideas, saw human nature against a cosmic and mythic backdrop in which God warred against Satan, heaven was opposed to hell, and humans were utterly evil unless rescued by Christ; anyone bringing such a view to ordinary human affairs would inevitably distort them. On the other hand, fellow Christians often complained that Niebuhr did not do justice to the themes of redemption, love and peace. The central figure of the Christian faith, Jesus of Nazareth, represented and taught the way of peace and was an exemplar of love. In traditional creedal terms, Jesus Christ, as the divine Son of God, effected redemption and helped people begin to realize a new age, a new ordering of being. Why, these critics asked, did Niebuhr so often act and teach as if none of this had happened? Was not his Christianity mere realism, indeed, pessimism?

In the face of such criticisms Niebuhr spent many hundreds of pages pointing to what he thought were the specific Christian dimensions of his philosophy and action. Yes, he would be a realist, not hiding from view the excessively proud antisocial aspects of humans, especially in their collective actions. No, no efforts to educate or reform would bring people to a situation where they could live in an era of progress or move towards Utopia. But the word 'Christian' did signal a different interpretation of realism. Niebuhr liked to think in polar terms, and just as he set redemption against sin, and the kingdom of God against human striving, so he introduced a kind of idealistic note to counter ordinary realism. This ideal was not some abstract or remote notion. It had been present in the utterances of the prophets and embodied and proclaimed by Jesus, who presented humans with something of God's love and therefore of God's very nature. So Niebuhr could praise *agapē*, the Greek New Testament's name for the spontaneous, unmotivated, generous and even reckless love of God, who showed concern for humans even in their waywardness and who made it possible for them to express such love.

The presence of *agapē*, even if one cannot expect to see much of it in the spheres of human self-interest, is realized just often enough in personal affairs to inspire idealistic action in public ones. The ethical teachings of Jesus in the Gospels, for instance, expressed such idealism in the form of love. Niebuhr succinctly addressed the polarity between ideal and reality by observing that if realists are not idealists they are 'inclined to obscure the residual moral and social sense even in the most self-regarding men and nations'; on the other hand, idealists who are not realists 'are inclined to obscure the residual individual and collective self-regard either in the "saved" or in the rational individuals and groups' (1965: 31-2).

Niebuhr was criticized by many orthodox Christians for not revisiting classic ways of discussing the divine nature of Jesus Christ or the metaphysical aspects of the divine-human transaction in the death of Jesus on the cross. Other Christians accused him of drawing too severe a line between the potential for sacrificial love in personal life (a love transcending mere self-interest) and the self-interest that realists saw as marking human affairs so much that the love of Christ could hardly enter political and international transactions. But it is always the hazard of a dialectical thinker who tries to do justice to two contrasting, if not overtly contradictory, poles to be vulnerable from both sides. Niebuhr took this chance, for his worldview was so decisively shaped by such poles as 'idealism' and 'realism', and 'self-interest' and 'sacrificial love' that he could not have avoided dealing with them. In a lifelong devotion to this vision, applied to countless changing situations, Niebuhr left a stamp on American political and Church life deeper than that of any other religious thinker of his times.

MARTIN MARTY

List of works

Niebuhr, R. (1932) *Moral Man and Immoral Society: A Study in Ethics and Politics*, New York: Charles Scribner's Sons.(A strong indictment of liberal optimism in politics and religion.)

Niebuhr, R. (1935) *An Interpretation of Christian Ethics*, New York: Harper & Bros.(Discusses the 'impossible possibility' of ethics based on the teachings of Jesus.)

Niebuhr, R. (1941, 1943) *The Nature and Destiny of Man: A Christian Interpretation*, New York: Charles Scribner's Sons, 2 vols.(The author's most formal statement in theology, stressing human limitations and responsibility to the divine.)

Niebuhr, R. (1949) *Faith and History: A Comparison of Christian and Modern Views of History*, New York: Charles Scribner's Sons.(Philosophy of history on a grand scale, stressing distinctively Christian aspects.)

Niebuhr, R. (1952) *The Irony of American History*, New York: Charles Scribner's Sons.(A Cold War-era interpretation of US history, criticizing the pride of nations.)

Niebuhr, R. (1955) *The Self and the Dramas of History*, New York: Charles Scribner's Sons.(Niebuhr confronts individualism, a neglected topic for him, and locates it in social processes.)

Niebuhr, R. (1965) *Man's Nature and His Communities*, New York: Charles Scribner's Sons.(A late statement discussing pluralism and international affairs.)

References and further reading

Fox, R. (1985) *Reinhold Niebuhr: A Biography*, New York: Pantheon.(The most comprehensive biography, essential for understanding the career, though criticized by some for slighting the theology.)

Harland, G. (1960) *The Thought of Reinhold Niebuhr*, New York: Oxford University Press.(A fair-minded synthesis of the main themes.)

Kegley, C. and Bretall, R.W. (eds) (1984) *Reinhold Niebuhr: His Religious, Social and Political Thought*, New York: Pilgrim Press.(A revised and expanded second edition of an extensive set of critiques first published in 1956.)

Rasmussen, L. (ed.) (1989) *Reinhold Niebuhr: Theologian of Public Life*, London: Collins.(A well-introduced collection of Niebuhr's writings.)

Richards, P. (1984) *Annotated Bibliography of Reinhold Niebuhr's Works*, Madison, NJ: American Theological Library Association.(Includes succinct and helpful comments on hundreds of works.)

Stone, R.H. (1972) *Reinhold Niebuhr: Prophet to Politicians*, Nashville, TN: Abingdon.(Stresses the political effects of Niebuhr's theological and philosophical career.)

Nietzsche: impact on Russian thought

Nietzsche's thought had a massive influence on Russian literature and the arts, religious philosophy and political culture. His popularizers were writers, artists and political radicals who read his works through the prism of their own culture, highlighting the moral, psychological and mythopoetic aspects of his thought and their sociopolitical implications, and appropriating them for their own agendas. Literature addressed to a mass readership disseminated crude notions of a master morality and an amoral Superman.

Russians discovered Nietzsche in the early 1890s. His admirers regarded him as a proponent of self-fulfilment and an enemy of the ''slave morality' of Christianity. Two of them, Dmitri Merezhkovskii (1865-1941) and Maksim Gor'kii (real name Aleksei Peshkov, 1868-1936), were the progenitors of the two main streams of Nietzsche appropriation - the religious and the secular. Merezhkovskii was the initiator of Russian Symbolism. In 1896 he began trying to reconcile Nietzsche and Christianity; this attempt led him to propound an apocalyptic Christianity in 1900 and to found the Religious-Philosophical Society of St Petersburg (1901-3, 1906-17). Its members, the so-called God-seekers, included artists and intellectuals who were also attracted to Nietzsche. As for Gor'kii, his early short stories featured vagrant protagonists who personified crude versions of the slave and the master morality. In 1895 Gor'kii began to dream of a Russian Superman who would lead the masses in a struggle for liberation and imbue them with respect for Man, which he always wrote with a capital letter. During the Revolution of 1905, he and Anatolii Lunacharskii (1875-1933), a Bolshevik admirer of Nietzsche, constructed a Marxist surrogate religion to inspire heroism and self-sacrifice. They believed, as did most Symbolists and some philosophers, that art could transform human consciousness.

New literary schools emerged after 1909. The Futurists exaggerated Nietzsche's anti-rationalism, anti-historicism and cultural iconoclasm. The Acmeists propounded a non-tragic Apollonian Christianity and idealized classical antiquity and 'world culture'. After the Bolshevik Revolution, Nietzsche was considered an ideologue of reaction and his books were removed from the People's Libraries, but his ideas, not identified as such, continued to circulate and pervaded Soviet literature, the arts and political culture.

1 Symbolists

In Russia, literature and literary criticism were the venues for discussing social and political ideas in coded Aesopian language. This applied to Nietzsche as well. Successive generations of writers and critics developed their own interpretations of Nietzsche, polemicizing with their contemporaries and with preceding generations, and basing their arguments on different Nietzsche texts or on variant readings of the same text. Strictly speaking, Nietzsche had no Russian disciples. Rather, his ideas fructified various literary and artistic schools, religious philosophy and Bolshevik political culture. The early twentieth-century religious renaissance was in large part a response to Nietzsche, as was the Soviet obsession with creating a new culture and a new man (*chelovek*); the word is gender neutral but discussions focused on men.

For the Symbolists, the key Nietzsche text was *The Birth of Tragedy*. They were dazzled by Nietzsche's aesthetic justification of the world and human existence, his celebration of the Dionysian, and his belief that myth is essential to the health of a culture. Much of their imagery and their perception of Nietzsche as a mystic and a prophet, derived from *Thus Spoke Zarathustra*.

The first generation of Russian Symbolists - Merezhkovskii, his wife Zinaida Hippius (1869-1945), Konstantin Balmont (1867-1942), Valerii Briusov (1873-1924), N.M. Minskii (N.M. Vilenkin, 1855-1937) and their allies, the modernists of *Severnyi vestnik* (Northern Herald, 1885-98) and the aesthetes of *Mir iskusstva* (The World of Art, 1898-1904) - discovered Nietzsche at a time when the appeal of populism, the surrogate religion of the intelligentsia, was fading. Nietzsche's aestheticism helped fill the spiritual gap. Art became their religion. They saw themselves as warriors for culture and against arid rationalism and emotional repression, as creators who smashed the old 'tables of values' and inscribed their own values on new tables. Balmont regarded the artist as Superman. Later on, Merezhkovskii referred to the Symbolists of the 1890s, including himself, as Decadents.

In a famous 1892 essay on the causes of the decline of Russian literature, Merezhkovskii attacked positivism and populism for debasing and despiritualizing art and argued that Symbolism would lead to new truths (note the

plural) that would unite the atheistic intelligentsia and the Christian people (*narod*). Unlike Nietzsche, most Russian Symbolists believed that higher truths did exist somewhere beyond the Dionysian flux. Only two years later, Merezhkovskii propounded an aesthetic individualism in which self-expression and the creation of beauty were the highest values. In essays and poems, and in his historical novel *Otverzhennyi (The Outcast)* (1895) later published as *Smert' bogov: Iulian Otstupnik (Death of the Gods: Julian the Apostate)*, about the Roman Emperor who tried to restore paganism, he celebrated the liberation of the passions and instincts repressed by Christianity, sexuality in particular, and the Nietzschean theme of laughter. But Merezhkovskii's inordinate fear of death, together with other factors such as his discovery of the cruelty of Roman civilization, led him to reconsider Christianity. In 'Pushkin' (1896), he declared that Christianity articulated the 'truth of heaven' (love and personal immortality) and paganism (really Nietzscheanism), the 'truth of the earth' (enjoyment of worldly pleasures) - two truths that must be reconciled by an all-encompassing higher truth. Asserting that Pushkin had reconciled them unconsciously, Merezhkovskii set about trying to find Pushkin's 'secret'. In the same essay, he also stated that Pushkin was the perfect combination of Apollo and Dionysus, in other words, a Superman in the popular conception of the term. Nietzsche himself expected the Superman to emerge only in the remote Future. Merezhkovskii's essay stimulated new readings of Pushkin well into the Soviet period.

Merezhkovskii's equally seminal book *Tolstoi i Dostoevskii* (1900-2) compared the Russian writers with one another and with Nietzsche. This book was Merezhkovskii's response to Nietzsche's critique of Christianity, which Merezhkovskii now claimed applied only to 'historical Christianity' (Christianity as preached in the churches). A Third Testament or Third Revelation was forthcoming. Jesus Christ himself would show humankind how to reconcile Christianity and paganism. The Revolution of 1905 politicized Merezhkovskii. He regarded it as the beginning of the apocalypse and labelled the Autocracy, the Orthodox Church and middle-class philistinism (*meshchanstvo*) guises of the Beast. Nietzschean elitism helped this Russian aristocrat justify his contempt for the middle class. Like his fellow Symbolists, Merezhkovskii considered Nietzsche an anarchist.

A second generation of Symbolists - Viacheslav Ivanov (1866-1949), Andrei Belyi (1880-1934) and Aleksandr Blok (1880-1921) - emerged after 1900. Ivanov maintained, in 'Ėllinskaia religiia stradaiiushchego Boga' (The Hellenic Religion of the Suffering God) (1904), that Dionysianism was an aesthetic-psychological-religious phenomenon, and that Dionysus was a precursor of Christ. In subsequent essays, Ivanov renounced 'Nietzschean' individualism and the Superman ideal and urged the Symbolists become myth-creators - to articulate the myths around which Russians could unite and thereby end social conflict. During the Revolution of 1905, Ivanov supported a doctrine called Mystical Anarchism, the brainchild of Georgii Chulkov, an anarchist and a former editor of Merezhkovskii's revue *Novyi put'* (New Path, 1902-4). Mystical Anarchism demanded the abolition of all authorities external to the individual - government, law, morality and social custom. Post-apocalyptic society would be characterized by freedom, beauty and love (*eros*, not *agapē*) and would be cemented by passion, myth and sacrifice. These invisible bonds would be forged in a theatre-temple modelled on the Theatre of Dionysus and devoted to myth-creation (*mifotvorchestvo*). The Symbolists would serve as high priests and the spectators would become a chorus and participate actively in the myth-creating process. In the orgiastic ecstasies of the rites, they would shed their separateness and become one, thereby bringing about the communitarian consciousness necessary for a society without coercion. Ivanov's ideas on a Dionysian theatre impressed his contemporaries, and were a major source of the mass festivals and political theatre of the early Soviet period. Ivanov's Christ/Dionysus archetype became a Symbolist trope.

One product of the post-1905 easing of the censorship laws was a vulgar Nietzscheanism (*Nitsheanstvo*) of pornography, nihilism and decadence in literature and in life. Appalled by this phenomenon, Ivanov and Merezhkovskii distanced themselves from Nietzsche. Nevertheless, his ideas continued to shape their thought, informing, for example, the Symbolist concept of life-creation (*zhiznetvorchestvo*) which exalted art as a theurgical activity that would transfigure man and the world. Different Symbolists had their own versions.

Nietzsche's ideas were an important component of a mystical revolutionism that associated the people with Dionysus and the intelligentsia with Apollo and bourgeois civilization. Blok predicted, even welcomed, the destruction of the intelligentsia by the enraged people. He, Belyi and a group called the Scythians perceived the Bolshevik Revolution as the beginning of a 'revolution of the spirit' that would engender a new culture and a new man - an artist (see Russian religious-philosophical renaissance §5).

2 Philosophers

Philosophic debate on Nietzsche was initiated by the respected journal *Voprosy filosofii i psikhologii (Questions of Philosophy and Psychology)* and centred on his critique of Christian morality. In October 1892, the journal published 'Friedrich Nietzsche: The Critique of Altruism' by V.P. Preobrazhenskii (1864-1900) who praised Nietzsche as a moralist, a defender of freedom and an advocate of the elevation and strengthening of the human person. Three articles in the next issue (January 1893) attacked Nietzsche as an amoralist and a decadent. The discussion spilled over into other journals. Nikolai Mikhailovskii (1842-1904), editor of the Populist journal *Russkoe bogatstvo (Russian Wealth)* and a long-time advocate of the efficacy of individual action, regarded Nietzsche as an ally in his struggle against Marxist collectivism and historical determinism, as an advocate of a higher morality. As a socialist, Mikhailovskii recognized and objected to the anti-democratic implications of the Superman ideal. Vladimir Solov'ëv (1853-1900), the most influential philosopher of this period, described Nietzsche's idea of the Superman as 'mangodhood', the obverse of his own idea of *Bogochelovechestvo* (*Godmanhood*). Decrying Nietzsche's separation of beauty and power from religion, Solov'ëv maintained that only religion can free beauty from death. He also argued that even though the Superman ideal was demonic, it was religiously significant because it expressed a desire for human perfection, a yearning to be more than human. The Symbolists combined Solov'ëv and Nietzsche in various ways, in the above-mentioned concept of life-creation for example.

Nietzsche was of central importance to Nikolai Berdiaev (1874-1948), Semën Frank (1877-1950) and Lev Shestov (1866-1938). Nietzsche confirmed and reinforced their objections to positivism and rationalism, and challenged them to think about goodness, truth and beauty in new ways.

In their contribution to the symposium *Problemy idealizma (Problems of Idealism)* (1902), Berdiaev and Frank, two former Marxists turned Neo-Kantians, tried to accommodate Nietzsche to Kant. Berdiaev wanted to replace Kant's 'morality of duty' with a higher Nietzschean morality of free desire, but in the same essay he declared that man had a duty to become a Superman. He regarded the Superman as a religious-metaphysical ideal and envisaged a spiritual aristocracy roughly comparable to Nietzsche's 'new nobility'. Frank interpreted Zarathustra's 'love for the most distant' as love for distant goals such as truth, justice and beauty, as opposed to love for the nearest, for one's neighbour. Frank also talked about a 'new nobility', a spiritual or cultural elite, that struggled to realize its ideals. Years later, Frank credited Nietzsche with revealing to him the reality of the psyche and of spiritual life.

Berdiaev came to consider Nietzsche the forerunner of a new religious anthropology and the unwitting prophet of a new Christianity of freedom ungrounded in material reality, of limitless creativity, unprecedented beauty, and self-overcoming. In *Smysl tvorchestva: opyt opravdanie cheloveka (The Meaning of the Creative Act: An Attempt at the Justification of Man)* (1916), he maintained that the Third Revelation will be about man and will be accomplished in a free creative act. Nietzsche's goal, he said, was 'not to justify creativity by life but to justify life by creativity' ([1916] 1954: 110). At a time when most of his contemporaries had distanced themselves from Nietzsche, Berdiaev lauded *Thus Spoke Zarathustra* as 'the most powerful human book without Grace; whatever is superior to Zarathustra is so by Grace from on high' ([1916] 1954: 90).

Shestov maintained that reason can neither explain nor assuage misfortune, suffering or ugliness. In *Dobro v uchenii gr. Tolstogo i Fr. Nitshe (The Good in the Teaching of Tolstoi and Nietzsche)* (1900) and *Dostoevskii i Nitshe - filosofiia tragedii (Dostoevsky and Nietzsche: The Philosophy of Tragedy)* (1903), he used Nietzsche to negate the 'rational morality' of self-renunciation preached by Kant and Tolstoi, and to argue against rational systems in general. Unlike the philosophers mentioned above, Shestov did not consider Nietzsche an advocate of a higher morality; rather Shestov insisted that Nietzsche was not a moralist at all, for he showed that people must seek God rather than 'the good'. In subsequent works, Shestov attacked philosophical systems that impose a non-existent unity on the world and gloss over the horrors of life. His religious existentialism reminds one of Kierkegaard's, whom he discovered in the late 1920s.

3 Bolsheviks

Nietzsche's most important Bolshevik popularizers - Gor'kii (even though he did not join the Party), Anatolii Lunacharskii and Aleksandr Bogdanov (Aleksandr Malinovskii, 1873-1928) - wanted to restore heroism and voluntarism to Marxism and to supplement Marx's teachings with new findings in the social and the physical

sciences. All of them maintained that under socialism, Kant’s ’morality of duty’ would be superseded by a morality of free desire, a ’master morality’, so to speak, but for everyone. They invoked the power of creative will, emphasized the collective aspects of Dionysianism (self-forgetting) and quoted Zarathustra’s dictum ’man is a bridge and not a goal’. Nietzsche’s powerful future-orientation and his related espousal of instrumental cruelty (as distinct from sadism), were taken up without acknowledgement by the Leninist-Marxists, especially Trotsky.

Nietzsche helped Gor’kii substitute an ethos of self-development for the Populist dictum that the writer must serve the people. Later on, Gor’kii tried to develop a Nietzschean aesthetic for a socially committed literature that promoted reason and science as guides to life. After he allied with the Bolsheviks in 1905, he denounced what he took to be Nietzschean individualism as philistinism. Gor’kii’s favourite Nietzsche text was ’On War and Warriors’ in *Thus Spoke Zarathustra*, which includes the line ‘man is something that must be overcome’. Also important to Gor’kii was Zarathustra’s injunction ’be hard - all creators are hard’. For most of his life, Gor’kii regarded the masses as sources of Dionysian energy, to be channelled by an Apollonian elite.

Lunacharskii discovered Nietzsche in 1901. Nietzsche helped him reconcile his love of the arts with his commitment to revolution. He claimed that art is a powerful stimulus to revolutionary ardour, described the ’will to power’ as the ’will to creativity’ (including social creativity), and called the Marxist ideal of a just and harmonious society an aesthetic ideal.

During the Revolution of 1905, he and Gor’kii developed the Marxist surrogate religion of *Bogostroitel’stvo* (God-building). It extolled the heroic proletariat as saviour of humanity, preached worship of collective humanity, and promised collective immortality to encourage people to risk death fighting for socialism. The basic texts of God-building are Lunacharskii’s *Religiia i sotsializm (Religion and Socialism)* (2 vols, 1908, 1911), Gor’kii’s novel *Ispoved’ (Confession)* (1908) and his essay ’The Destruction of the Person’ (1909), written against individualism. *Confession* includes a scene in which the energies of an assembled crowd raise a paralyzed girl; it also includes a Superwoman, aptly named Christina.

Strictly speaking, Bogdanov was not a God-builder, but he did try to create a new myth. He wrote the utopian novels *Krasnaia zvezda (Red Star)* (1908) and *Inzhener Menni (Engineer Menni)* (1913), about a perfected communist society on the planet Mars, to inspire the workers to continue their struggle by showing them what they were fighting for. Bogdanov’s vision was based on an ultra-collective, ultra-egalitarian interpretation of Marxism which denigrated the individual as a ‘bourgeois fetish’. Paradoxically, Bogdanov claimed that communism would facilitate self-development. He hated authoritarianism and militarism. His ideal world had no experts, no hierarchies and no would-be Supermen. It was a world of proletarian renaissance persons (he was a partisan of gender equality) who had conquered nature or at least kept it at bay. They were guided by ‘expediency norms’ that changed as conditions changed, as distinct from ‘coercive norms’ (absolute commandments) and eternal truths. Bogdanov’s first ‘expediency norm’ was ‘there shall be no herd instinct’.

Bogdanov maintained that the Revolution of 1905 failed because the workers were culturally and psychologically unprepared to assume power. Therefore, he argued, cultural revolution must precede and accompany political revolution. Lenin disagreed. In 1909, he had Bogdanov expelled from the Party. The expulsion gave Bogdanov time to develop his concept of cultural revolution, by which he meant: the psychological and spiritual emancipation of the proletariat from bourgeois dominance, systematic revaluation of the cultural legacy from a proletarian perspective, and the construction of a distinctively proletarian art and science. Bogdanov’s views inspired Proletkult (proletarian culture), an extra-Party movement that he, Lunacharskii and others founded in October 1917. Gor’kii was not among them; he wanted culture to be above class. At its peak in 1921, Proletkult numbered half a million members and had an extensive network of schools, art studios and theatres. The movement collapsed after Lenin attacked it, but crude versions of Bogdanov’s ideas were perpetuated by self-styled proletarian groups during the culture wars of the 1920s, and helped inspire Stalin’s cultural revolution of 1928-31. Gor’kii championed Socialist Realism, the Stalinist form of myth-creation and the official aesthetic of the Soviet Union from 1934, when it was officially adopted by the first Congress of Soviet Writers, until well after Stalin’s death. Lunacharskii was involved in preparatory activities but he died before the Congress met (see Russian empiriocriticism).

4 Futurists and Acmeists

To the Futurists, the key Nietzsche text was the madman’s announcement of the death of God in *The Gay Science* (§125) and the chaotic, directionless, universe and infinite horizons that ensue. Their foundational myth was *Pobeda nad solntsem* (Victory Over the Sun) (1913), the title of an operetta about the capture of the sun (Apollo), which occurs offstage, by Futurist strongmen. These new people are male or androgynous, but definitely not female. They speak a new language - *zaum* - literally ‘beyond the mind’, a concept inspired, in part, by Nietzsche’s attack on ‘reason in language’ and his association of language and power. The Futurists advocated a complete break with history, exaggerating the anti-historicism and celebration of youth in Nietzsche’s *The Advantage and Disadvantage of History for Life*, and they wished to jettison the cultural legacy. The bellicosity of their language and their aggressive masculinism echoed some of Nietzsche’s most ferocious passages. Their unabashed primitivism and orientation to Asiatic Russia stemmed from a literal reading of Nietzsche’s ‘new barbarians’ and indicated their desire to create *de novo*. Paradoxically, they were enthralled by technology, large cities and noisy crowds. *Pobeda nad solntsem* was performed together with *Vladimir Maiakovskii: Tragediia*, by Vladimir Maiakovskii (1893-1930), which was structured like a Dionysian dithyramb and included Maiakovskii’s version of the Christ/Dionysus archetype. Maiakovskii knew *Thus Spoke Zarathustra* well; his pre-Revolutionary poetry is permeated with images and themes drawn from it. Many Futurists were also painters; their alogism was a pictorial counterpart to *zaum*.

The Futurists brought the issues of language and form to the centre of literary debates, where they remained for over two decades. Their cultural iconoclasm and anti-bourgeois stance were easily politicized. After the Bolshevik Revolution they achieved quasi-official status. They maintained that a ’revolution of the spirit’ was needed to complete the political and social revolution and tried to establish their own dictatorship of taste.

The Acmeists, a short-lived group of poets that included Nikolai Gumilëv (1886-1921) and Osip Mandel’shtam (1891-1938), developed the concept of the Apollonian in a way that enabled them to reconcile Hellenism with Orthodox Christianity in a poetic exaltation of earthly reality. Rejecting theurgy and mysticism, their aesthetics emphasized clarity in language, concreteness and a visual orientation. They defended Christian ethical values (Apollo is the ethical deity), moral courage and ‘manly will’. Gumilëv’s warrior pose and conquistador persona stemmed from *Thus Spoke Zarathustra*; also important to him was Nietzsche’s concept of *amor fati*. Mandel’shtam’s statement that ’to remember also means to invent’ encapsulated his own interpretation of Nietzsche’s views on history and myth and affirmed past values and ’world culture’. The Acmeists were persecuted during the Soviet period. Gumilëv was executed for treason; Mandel’shtam died in the Gulag.

5 The Soviet period

Unacknowledged Nietzschean ideas (ideas mediated by Nietzsche’s Russian popularizers) helped shape Soviet literature and culture, including political culture, in ways too numerous and complex to be detailed here. They inspired literary, artistic and architectural experimentation, a ruthless revolutionary morality and cultural construction. In addition, they underlay the Lenin Cult, the Stalin Cult, socialist realism and the grandiose projects of the first and second five-year plans. Nietzschean themes were taken up by now-forgotten poets and playwrights and by such Soviet luminaries as the writer Isaac Babel, the theatrical director Vsevolod Meierhol’d and the film director Sergei Eisenstein. The religious stream of Nietzsche appropriation was perpetuated by Boris Pasternak, Mikhail Bakhtin and other disaffected intellectuals.

Nietzsche’s name was linked with ’isms that the government denounced: bourgeois individualism and nihilism in the 1920s, fascism in the 1930s, and bourgeois hedonism in Brezhnev’s time. The few books and articles about Nietzsche published after 1927 quoted *Beyond Good and Evil* and *The Will to Power* extensively, but his works remained inaccessible to ordinary readers. Under Gorbachev’s policy of *glasnost’* Nietzsche’s writings were reprinted in press runs of 100,000, which sold out immediately.

See also: Nietzsche, F.

BERNICE GLATZER ROSENTHAL

References and further reading

For additional references see the entries Russian religious-philosophical renaissance and Shestov, L.

Berdiaev, N. (1916) *Smysl tvorchestva: opyt opravdanie cheloveka*, Moscow; trans. D. Lowrie, *The Meaning of the Creative Act: An Attempt at the Justification of Man*, New York: Harper, 1954.

Clowes, E. (1988) *The Revolution of Moral Consciousness: Nietzsche and Russian Literature, 1890-1914*, De Kalb, IL: Northern Illinois University Press.(Treats popularizers and vulgarizers of Nietzsche, mystical Symbolists and revolutionary Romantics.)

Gor'kii [Gorky], M. (1908) *Ispoved'*, St Petersburg: Znanie; trans. R. Strunsky, *Confession*, New York: Stokes, 1916.

Gunther, H. (1993) *Der sozialist Übermensch: Maksim Gor'kij und der sovjetische Heldenmythos (The Socialist Superman: Maksim Gorkii and the Soviet Hero-Myth)*, Stuttgart: Metzler.(Treats Gor'kii's changing perceptions and utilizations of Nietzsche.)

Ivanov, V. (1904) 'Ėllinskaia religiia stradaiiushchego Boga' (The Hellenic Religion of the Suffering God), serialized in *Novyi put'*.

Kline, G.L. (1968) '"The God-builders": Gorki and Lunacharsky', in *Religious and Anti-Religious Thought in Russia*, 103-27.(Traces the roots of God-building to nineteenth-century Russian radicalism, Feuerbach and Nietzsche.)

Kline, G.L. (1969) '"Nietzschean Marxism" in Russia', in *Demythologizing Marxism*, ed. F. Adelmann, Boston, MA: Boston College Studies in Philosophy, vol. 2, 166-83.(Discusses Lunacharskii, Bogdanov, V.A. Bazarov and S. Volskii.)

Lunacharskii, A. (1908, 1911) *Religiia i sotsializm (Religion and Socialism)*, St Petersburg: Shipovnik, 2 vols.

Merezhkovskii, D. (1914) *Pol'noe sobranie sochinenii (Collected Works)*, Moscow: D. Sytin, 24 vols.(Contains all Merezhkovskii's works listed in this entry.)

Merezhkovskii, D. (1895) *Otverzhennyi (The Outcast)*, in *Severnyi vestnik*; subsequently retitled *Smert' bogov: Iulian Otstupnik*; trans. C. Johnson, *Death of the Gods: Julian the Apostate*, London: Constable.(This historical novel is the first of a trilogy of historical novels titled *Khristos i Antikhrist (Christ and Antichrist)*.)

Merezhkovskii, D. (1896) 'Pushkin', in 'Filosofskie techeniia russkoi poėzii', ed. P.P. Pertsov, St Petersburg: Merkushev.

Merezhkovskii, D. (1900-2) 'Tolstoi i Dostoevskii', in *Mir iskusstva*; *Tolstoy as Man and Artist, with an Essay on Dostoievsky*, Westminster: Constable, 1902.

Rosenthal, B.G. (ed.) (1986) *Nietzsche in Russia*, Princeton, NJ: Princeton University Press.(Multi-author work. Chapters on Symbolism, religious philosophy, Bolshevism, popular literature, Bakhtin; list of publications by and about Nietzsche 1892-1919; historical introduction by the editor.)

Rosenthal, B.G. (1991) 'A New Word for a New Myth: Nietzsche and Russian Futurism', in P.I. Barta (ed.) *The European Foundations of Russian Modernism*, Lewiston, NY: Mellen, 219-50.(The interconnection of word and myth.)

Rosenthal, B.G. (ed.) (1994) *Nietzsche and Soviet Culture: Ally and Adversary*, Cambridge: Cambridge University Press.(Multi-author work. Chapters on Futurism, Acmeism, occult doctrines, Soviet literature, art and architecture; adaptations of Nietzsche in Soviet ideology; Nietzsche among disaffected writers and thinkers; his influence on Hebrew writers before the Revolution. Editor's introduction includes discussion of methodological problems of assessing a buried influence.)

Nietzsche, Friedrich (1844-1900)

Appointed professor of classical philology at the University of Basel when he was just 24 years old, Nietzsche was expected to secure his reputation as a brilliant young scholar with his first book, Die Geburt der Tragödie (The Birth of Tragedy) (1872). But that book did not look much like a work of classical scholarship. Bereft of footnotes and highly critical of Socrates and modern scholarship, it spoke in rhapsodic tones of ancient orgiastic Dionysian festivals and the rebirth of Dionysian tragedy in the modern world. Classical scholars, whose craft and temperament it had scorned, greeted the book with scathing criticism and hostility; even Nietzsche eventually recognized it as badly written and confused. Yet it remains one of the three most important philosophical treatments of tragedy (along with those of Aristotle and Hegel) and is the soil out of which Nietzsche's later philosophy grew. By 1889, when he suffered a mental and physical collapse that brought his productive life to an end, Nietzsche had produced a series of thirteen books which have left a deep imprint on most areas of Western intellectual and cultural life, establishing him as one of Germany's greatest prose stylists and one of its most important, if controversial, philosophers.

Nietzsche appears to attack almost everything that has been considered sacred: not only Socrates and scholarship, but also God, truth, morality, equality, democracy and most other modern values. He gives a large role to the will to power and he proposes to replace the values he attacks with new values and a new ideal of the human person (the Übermensch *meaning 'overhuman' or 'superhuman'). Although Nazi theoreticians attempted to associate these ideas with their own cause, responsible interpreters agree that Nietzsche despised and unambiguously rejected both German nationalism and anti-Semitism. Little else in his thought is so unambiguous, at least in part because he rarely writes in a straightforward, argumentative style, and because his thought changed radically over the course of his productive life. The latter is especially true of his early criticism of Socrates, science and truth.*

Nietzsche's philosophizing began from a deep sense of dissatisfaction with modern Western culture, which he found superficial and empty in comparison with that of the ancient Greeks. Locating the source of the problem in the fact that modern culture gives priority to science (understood broadly, including all forms of scholarship and theory), whereas Presocratic Greece had given priority to art and myth, he rested his hopes for modern culture on a return to the Greek valuation of art, calling for a recognition of art as 'the highest task and the truly metaphysical activity of this life'.

He soon turned his back on this early critique of science. In the works of his middle period he rejects metaphysical truth but celebrates the valuing of science and empirical truth over myth as a sign of high culture. Although he had earlier considered it destructive of culture, he now committed his own philosophy to a thoroughgoing naturalistic understanding of human beings. He continued to believe that naturalism undermines commitment to values because it destroys myths and illusions, but he now hoped that knowledge would purify human desire and allow human beings to live without preferring or evaluating. In the works of his final period, Nietzsche rejects this aspiration as nihilistic.

In his final period, he combined a commitment to science with a commitment to values by recognizing that naturalism does not undermine all values, but only those endorsed by the major ideal of value we have had so far, the ascetic ideal. This ideal takes the highest human life to be one of self-denial, denial of the natural self, thereby treating natural or earthly existence as devoid of intrinsic value. Nietzsche saw this life-devaluing ideal at work in most Western (and Eastern) religion and philosophy. Values always come into existence in support of some form of life, but they gain the support of ascetic religions and philosophies only if they are given a life-devaluing interpretation. Ascetic priests interpret acts as wrong or 'sinful' because the acts are selfish or 'animal' - because they affirm natural instincts - and ascetic philosophers interpret whatever they value - truth, knowledge, philosophy, virtue - in non-natural terms because they share the assumption that anything truly valuable must have a source outside the world of nature, the world accessible to empirical investigation. Only because Nietzsche still accepted this assumption of the ascetic ideal did naturalism seem to undermine all values.

According to his later thought, the ascetic ideal itself undermines values. First it deprives nature of value by placing the source of value outside nature. Then, by promoting the value of truth above all else, it leads to a denial that there is anything besides nature. Among the casualties of this process are morality and belief in God, as

Nietzsche indicated by proclaiming that 'God is dead' and that morality will gradually perish. Morality is not the only possible form of ethical life, however, but a particular form that has been brought about by the ascetic ideal. That ideal has little life left in it, according to Nietzsche, as does the form of ethical life it brought about. Morality now has little power to inspire human beings to virtue or anything else. There is no longer anything to play the essential role played by the ascetic ideal: to inspire human beings to take on the task of becoming more than they are, thereby inducing them to internalize their will to power against themselves. Modern culture therefore has insufficient defences against eruptions of barbarism, which Nietzsche predicted as a large part of the history of the twentieth and twenty-first centuries.

But Nietzsche now saw that there was no way to go back to earlier values. His hope rested instead with 'new philosophers' who have lived and thought the values of the ascetic ideal through to their end and thereby recognized the need for new values. His own writings are meant to exhibit a new ideal, often by exemplifying old virtues that are given a new, life-affirming interpretation.

1 Life

Nietzsche was born in Rocken, a small village in the Prussian province of Saxony, on 15 October 1844. His father, a Lutheran minister, became seriously ill in 1848 and died in July 1849 of what was diagnosed as 'softening of the brain'. His brother died the following year, and Nietzsche's mother moved with her son and daughter to Naumberg, a town of 15,000 people, where they lived with his father's mother and her two sisters. In 1858, Nietzsche was offered free admission to Pforta, the most famous school in Germany. After graduating in 1864 with a thesis in Latin on the Greek poet Theogonis, he registered at the University of Bonn as a theology student. The following year he transferred to Leipzig where he registered as a philology student and worked under the classical philologist Friedrich Ritschl. The events of his Leipzig years with the most profound and lasting influence on his later work were his discoveries of Schopenhauer's *Die Welt als Wille und Vorstellung (The World as Will and Representation)* and F.A. Lange's *Geschichte des Materialismus (History of Materialism)*, and the beginning of a personal relationship with Richard Wagner (see Schopenhauer, A. §§3-6; Lange, F.A. §2). Nietzsche became Ritschl's star pupil, and on Ritschl's recommendation he was appointed to the Chair of Classical Philology at Basel in 1869 at the age of twenty-four. Leipzig proceeded to confer the doctorate without requiring a dissertation.

Basel's proximity to the Wagner residence at Tribschen allowed Nietzsche to develop a close relationship with Richard and Cosima Wagner. Sharing with the composer a deep love of Schopenhauer and a hope for the revitalization of European culture, he initially idealized Wagner and his music. His first book, *Die Geburt der Tragödie (The Birth of Tragedy)* (1872), used Schopenhauer's philosophy to interpret Greek tragedy and to suggest that Wagner's opera constituted its rebirth and thereby the salvation of modern culture. Torn between philology and philosophy since shortly after his discovery of Schopenhauer, Nietzsche devoted much of his teaching to the texts of ancient Greek and Roman philosophy and hoped that his first book would establish his credentials as a philosopher. Instead, its unorthodox mixture of philosophy and philology merely served to damage his reputation as a philologist.

In 1879, he resigned his chair at Basel because of health problems that had plagued him for years. In the meantime, he had become progressively estranged from Wagner, a process that culminated in the 1878 publication of the first volume of *Menschliches, Allzumenschliches (Human, All Too Human)*, a positivist manifesto that praised science rather than art as indicative of high culture. The ten productive years left to him after his retirement were marked by terrible health problems and a near absence of human companionship. Living alone in Italian and Swiss boarding houses, he wrote ten books, each of which has at least some claim to being a masterpiece. His last seven books mark a high point of German prose style.

In January 1889, Nietzsche collapsed in Turin. He wrote a few lucid and beautiful (although insane) letters during the next few days, and after that nothing of which we can make any sense. Following a brief institutionalization, he lived with his mother and then his sister until his death in Weimar on 25 August 1900.

2 Writings and development

During the sixteen years between his first book and his last productive year, Nietzsche's thinking underwent remarkable development, usually with little notification to his readers. The traditional grouping of his writings into

three major periods is followed here, although there is significant development within each period. In addition to *The Birth of Tragedy* his early work consists of four essays of cultural criticism - on David Strauss, history, Schopenhauer, and Wagner - published separately but linked together as *Unzeitgemäße Betrachtungen (Unfashionable Observations)* (1873-6), plus a number of largely finished essays and fragments that belong to the *Nachlaß* of the period. The most important of the essays are 'Über Wahrheit und Lüge im außermoralischen Sinne' ('On Truth and Lies in a Nonmoral Sense'), 'Homer's Wettkampf' ('Homer's Contest'), and 'Die Philosophie im tragischen Zeitalter der Griechen' ('Philosophy in the Tragic Age of the Greeks').

These early writings sound a note of great dissatisfaction with European (Enlightenment) culture, of which Socrates is taken as the earliest representative and continuing inspiration. At the base of Socratic culture Nietzsche finds the belief that life's highest goal is the theoretical grasp of truth at which science and philosophy aim. Theory's claim to provide truth has been undermined, he thinks, by the doctrine of Kant and Schopenhauer that discursive thought gives access not to things-in-themselves but only to 'appearance' (see Kant, I. §5; Schopenhauer, A. §§2, 4). Nietzsche's suggestion for saving European culture is that art should replace theory as the most valued, the 'truly metaphysical', human activity. At first, his main argument for elevating art is that it is more truthful than theory. But he also suggests a very different argument: that theory is destructive of culture unless it is guided and limited by the needs of life which art serves. In his essay on history, the second argument has largely replaced the first. He argues that when practised as autonomous theory, devoted solely to truth, history destroys the limited and mythical horizons required by life and action. And if we emphasize for another generation the naturalistic understanding of human beings at which Socratic culture has now arrived (for example, the denial of a cardinal distinction between humans and other animals), we will only further our culture's disintegration into chaotic systems of individual and group egoism. Nietzsche suggests that history can be harnessed to serve the needs of life, for instance when it is written to emphasize great lives and other aspects that encourage individuals to set lofty and noble goals for themselves. Such history is as much art as it is theory or science.

Nietzsche turns decisively away from such criticism of pure theory in the writings of his middle period, *Human, All Too Human* (1878-80) and *Morgenröte (Daybreak)* (1881). He here celebrates as a sign of high culture an appreciation of the little truths won by rigorous method, and presents his own philosophy as a form of natural science that serves only truth. He also commits himself to the truth of the naturalism he earlier considered so dangerous: there is no cardinal distinction between humans and other animals; everything about human beings, including their values, can be explained as a development from characteristics found among other animals. At the beginning of this period Nietzsche struggles with how naturalism can be compatible with a commitment to values, for he sees it as exposing and thereby undermining the illusions that are needed in order to find value in life. In *Human*, his hope is that knowledge will gradually purify 'the old motives of violent desire' until one can live 'as in *nature*', without preferring or evaluating, but 'gazing contentedly, as though at a spectacle' (*Human* §34 [*Werke* IV.2: 50]). Nietzsche later has Zarathustra mock this spectator conception of knowledge and life as 'immaculate perception' (*Zarathustra* II: §15 [*Werke* V.1: 152-5]).

Nietzsche's final period begins with *Die Fröhliche Wissenschaft (The Gay Science)* (1882), which replaces the spectator conception with one in which the 'knower' belongs to the dance of existence and is one of its 'masters of ceremony' (*The Gay Science* §54 [*Werke* V.2: 90]). This formulation expresses his new confidence that naturalism, which he often calls 'knowledge', is compatible with commitment to values. In this period, Nietzsche once again celebrates art, criticizes Socrates and denies the autonomy of theory, suggesting to some that he has reverted to the viewpoint of his early period. Evidence is provided throughout this entry for an alternative interpretation: the later Nietzsche does not deny that theory can provide truth, and he remains as committed to the pursuit of truth as he was in his middle period. The difference is that he now recognizes in even the apparently autonomous theory of his middle period a commitment to an ideal that is external to and served by theory, namely, the ascetic ideal. Nietzsche returns to the suggestion of his first book, that theory is not autonomous; however, he now objects not to theory, but only to the ideal that theory has served (see §§6 and 7 of this entry). The works of Nietzsche's final period are largely devoted to uncovering, criticizing and offering an alternative to that ideal.

Gay Science was followed by *Also Sprach Zarathustra (Thus Spoke Zarathustra)* (1883-5), a fictional tale used as a vehicle for Nietzsche's most puzzling and infamous doctrines, including the overhuman (*Übermensch*), will to power and eternal recurrence. He considered this to be the deepest work in the German language and suggested that chairs of philosophy might one day be devoted to its interpretation. Our surest guides to it at present are the

other books of his final period, especially the two that followed it: *Jenseits von Gut und Böse (Beyond Good and Evil)* (1886) and *Zur Genealogie der Moral (On the Genealogy of Morality)* (1887). These masterpieces show Nietzsche at the height of his powers as a thinker, an organizer and an artist of ideas. Yet some prefer his last five books. At the beginning of 1888, Nietzsche published *Der Fall Wagner (The Case of Wagner)* and then composed four short books before the year was out: *Die Götzen-Dämmerung (The Twilight of the Idols)*, an obvious play on Wagner's *Götterdämmerung (Twilight of the Gods)*, puts the finishing touches to his accounts of knowledge and philosophy and offers his final critique of Socrates; *Der Antichrist (The Antichrist)*, a critique of Pauline Christianity, offers a relatively sympathetic portrait of Jesus; and *Ecce Homo*, Nietzsche's own portrait of his life and work under such chapter headings as 'Why I write such good books'. It is easy to hear signs of his impending insanity in the shrill tone and self-promotion that sometimes takes over in these books (although not in the very funny and brilliantly anti-Socratic chapter headings of *Ecce Homo*), and perhaps also in the fall-off in organizational and artistic power from the masterpieces of the previous two years. *Nietzsche Contra Wagner*, which he dated Christmas 1888, leaves a different impression. Nietzsche's shortest and perhaps most beautiful book, it is a compilation of passages from earlier works, with a few small improvements, as if aiming at perfection. He collapsed nine days later.

3 The *Nachlaß*

Nietzsche left behind a large body of unpublished material, his *Nachlaß*, which technically should include *The Antichrist* and *Ecce Homo*, published by his sister in 1895 and 1908 respectively. However, Nietzsche had made arrangements for their publication and prepared a printer's copy of each. For purposes of understanding his philosophy, these are therefore accorded the same status as his earlier works and are not usually considered to be part of his *Nachlaß*. This entry gives a very secondary role to the remainder of Nietzsche's *Nachlaß*, which includes the relatively polished essays written in the early 1870s (mentioned in §2 of this entry). These essays are informative about Nietzsche's early views, and are sometimes also thought to provide a clearer statement of his later views of truth and language than do the works he published. The interpretation of Nietzsche's development offered here supports a very different view: that Nietzsche chose not to publish these essays because he soon progressed beyond them to quite opposed views.

Another issue that divides interpreters concerns the weight to give to the notes of Nietzsche's later years. Many treat them as material he would have published if he had remained productive for longer. But since he might instead have rejected and disposed of much of this material, others advise great caution in its use. Further, we often cannot determine the use Nietzsche had in mind for particular notes even when he wrote them. Nietzsche *composed* his books to lead prepared readers to certain views. The rich context and clues for reading supplied by his books, when they are attended to, provide a check on interpretive licence and a basis for getting at Nietzsche's own thinking that has no parallel in the case of the *Nachlaß* material. This applies to the entire contents of *Der Wille zur Macht (The Will to Power)*, which some have regarded as Nietzsche's magnum opus. Although he did announce it as 'in preparation', there is evidence that he dropped his plans to publish a work of this title; the book we have is actually a compilation of notes from the years 1883-8 selected from his notebooks and arranged in their present form by his sister and editors appointed by her. Such notes may sometimes help in understanding what Nietzsche actually did publish. But it is difficult to justify giving them priority when they suggest views that differ from and are even contrary to those suggested by a careful reading of Nietzsche's books (see §§11-12 of this entry).

4 Truth and metaphysics

In the writings of his early and middle periods, Nietzsche often appears to deny that any of our theories and beliefs are really true. By the end of his final period, he denies only metaphysical truth. The rejection of metaphysics forms the cornerstone of his later philosophy.

What Nietzsche rejects as metaphysics is first and foremost a belief in a second world, a metaphysical or true world. *Human, All Too Human* offers a genealogy of this belief. Receiving their first idea of a second world from dreams, human beings originally share with 'everything organic' a belief in the existence of permanent things (substance) and free will. When reflection dawns and they fail to find evidence of these in the world accessible to empirical methods, they conclude that these methods are faulty, and that the real world is accessible only to non-empirical methods. They thus take the empirical world to be a mere appearance or distortion of a second

world, which is thereby constituted as the true one. Metaphysics is purported knowledge of this non-empirical world. *The Birth of Tragedy* affirms metaphysics in this sense - 'an artists' metaphysics' he later called it - in the suggestion that perception and science confine us to mere appearance, whereas truth is accessible in the special kind of preconceptual experience characteristic of Dionysian art.

Human, All Too Human sets out to undermine metaphysics by showing that knowledge of a non-empirical world is cognitively superfluous. Nietzsche's Enlightenment predecessors had already established the adequacy of empirical methods to explain what goes on in the nonhuman world. However, belief in a metaphysical world persisted because that world is assumed to be necessary to account for the things of the highest value in the human world. Nietzsche sought to explain the origin of this assumption and to undermine it. The assumption was made, he claims, because thinkers were unable to see how things could originate from their opposites: disinterested contemplation from lust, living for others from egoism, rationality from irrationality. They could deny this origination only by positing for 'the more highly valued thing a miraculous source in the very kernel and being of the "thing-in-itself"'. Nietzsche offers a naturalistic account of higher things, which presents them as sublimations of despised things and therefore as 'human, all too human'. Once it is clear that we can explain their origin without positing a metaphysical world, he expects the interest in such a world to die out. We cannot deny the bare possibility of its existence, however, because 'we view all things through the human head and cannot cut this head off; yet the question remains what of the world would still be there if we had cut it off' (*Human* §§1, 9 [*Werke* IV.3: 19, 25]).

Nietzsche later goes a step further and denies the very existence of a metaphysical world. His history of the 'true' world in *Twilight of the Idols* offers a six-stage sketch of how the metaphysical world came to be recognized as a 'fable'. Stage Four corresponds to the position of *Human*: the 'true' world is cognitively superfluous. In Stage Five, its existence is denied. Stage Six adds that without a true world, there is no merely apparent world either: the empirical world originally picked out as 'merely apparent' is the only world there is. Nietzsche thus makes clear that he has moved beyond the assumption that there *might* be a metaphysical world to a positing of the empirical world as the only one. He dismisses the whole idea of a second world as unintelligible. The books after *Beyond Good and Evil* proceed on this assumption: they no longer claim that the empirical world is a mere appearance or, what amounts to the same thing, that empirical truths are illusions or falsifications.

5 Knowledge

The position on knowledge to which Nietzsche is led by his rejection of metaphysics is a combination of empiricism, antipositivism and perspectivism. Claiming in his later works that 'all evidence of truth comes only from the senses', and that we have science 'only to the extent that we have decided to *accept* the testimony of the senses - to the extent to which we sharpen them further, arm them, and think them through', he considers the rest of purported knowledge 'miscarriage and not yet science', or formal science, like pure logic and mathematics (*Beyond Good and Evil* §134 [*Werke* VI.2: 96]; *Twilight* III §3 [*Werke* VI.3: 69-70]). The latter, he now insists, departing from his earlier stand that they falsify reality, make no claim about reality at all. Nietzsche's empiricism amounts to a rejection of any wholesale disparaging of sense experience, an insistence that the only bases for criticizing or correcting particular deliverances of the senses are other sense experiences or theories based on them.

Nietzsche's antipositivism involves a rejection of two aspects associated with some other versions of empiricism. First, he rejects foundationalism. Anticipating many later critics of positivism, he denies that there is any experience that is unmediated by concepts, interpretation or theory. Sense experience, our only evidence of truth, is always already interpreted, and knowledge is therefore interpretation, as opposed to the apprehension of unmediated facts. Nietzsche also avoids the problem of needing an a priori theory to establish his empiricism, which he bases instead on his genealogy of the belief in a metaphysical world (a genealogy that is itself empirical in that it accepts the testimony of the senses) and a diagnosis and working-through of the intellectual confusions that have locked previous philosophers into that belief. Clearing away these confusions (especially pictures of knowledge that set the world's true nature over against its appearances) removes all intellectual basis for considering sense experience in principle problematic, and all intellectual motivation for pursuing a priori knowledge. Philosophers may, however, still have non-intellectual motives for this pursuit (see §6 of this entry). The upshot of Nietzsche's antipositivism is that what counts as knowledge is always revisable in the light of new or improved experience. This reinforces his empiricism, and in no way devalues empirical theories or denies that

they can give us truth.

Nietzsche’s perspectivism is often thought to imply that empirical knowledge offers us ‘only a perspective’ and not truth. But is perspectivism itself only a perspective? If not, it is false; if so, it is not clear why we should accept perspectivism rather than some other perspective. And Nietzsche himself puts forward as truths not only perspectivism, but also many other claims.

We can avoid saddling Nietzsche with these problems by recognizing that, at least in its mature and most important formulation (at *Geneology* III §12 [*Werke* VI.2: 381-3]), perspectivism is a claim about knowledge; it is not a claim about truth, and it does not entail that truth is relative to perspective. Further, ‘perspectives’ are constituted by affects, not beliefs. The point is *not* that knowledge is always from the viewpoint of a particular set of beliefs and that there are always alternative sets that would ground equally good views of an object (see Relativism). Such a view inevitably saddles perspectivism with relativism and problems of self-reference. Nietzsche’s explicit point in describing knowledge as perspectival is to guard against conceiving of knowledge as ‘disinterested contemplation’.

His early essay ‘Truth and Lie’ did use the impossibility of disinterested knowledge to devalue empirical knowledge, arguing that the latter was only a perspective and an illusion. But the point of the *Genealogy*’s claim that there is ‘only a perspective knowing’ is quite the reverse: to guard against using the idea of ‘pure’ knowing to devalue the kind of knowledge we have. The metaphor of perspective sets up disinterested knowing as the equivalent of the recognizably absurd notion of seeing something from nowhere. If the conception of knowledge ruled out by perspectivism really is absurd, however - and Nietzsche insists that it is - then it excludes only a kind of knowledge of which we can make no sense and which we could not really want. This explains why so many find perspectivism obvious and even self-evident; but so interpreted, it does nothing to devalue empirical knowledge.

Why does Nietzsche deny the possibility of disinterested knowledge? That surely does not follow from the impossibility of seeing something from nowhere. His early basis for this denial was Schopenhauer’s doctrine that the intellect originates as servant to the will, but he accepted the same doctrine in later works on the basis of a thoroughgoing Darwinian naturalism. Human cognitive capacities exist because of the evolutionary advantage they confer on the species, and no such advantage is to be found in attending to any and all features of reality. The intellect must be directed to certain features - initially at least, those most relevant to human survival and reproduction. Affect - emotion, feeling, passion, value orientations - turns the mind in a particular direction, focusing its attention on certain features of reality and pushing it to register them as important; knowledge is only acquired when the intellect is so pushed and focused. Nietzsche’s perspectivism is a metaphorical formulation of this naturalistic understanding of knowledge.

Because knowledge is always acquired from the viewpoint of particular interests and values, there are therefore always other affective sets that would focus attention on different aspects of reality. Nietzsche’s use of the metaphor of perspective thus implies that knowledge is limited in the sense that there are always other things to know, but not that perspectives block our access to truth. Affects are our access, the basis of all access to truth. If its perspectival character raises any problems for knowledge, it is only because being locked into a particular perspective can make one unable to appreciate features of reality that are apparent from other perspectives. Nietzsche’s solution is simple: the more affects we know how to bring to bear on a matter, the more complete our knowledge of it will be.

This does not mean that true knowledge requires assuming as many perspectives as possible. Knowledge does not require complete knowledge, and complete knowledge is not Nietzsche’s epistemological ideal. In fact, he suggests that the greatest scholars tend to serve knowledge by immersing themselves deeply and thoroughly in some particular perspective, so much so that they damage themselves as human beings. The situation is different for philosophers because their ultimate responsibility is not knowledge, but values. To undertake the task to which Nietzsche assigns them, they need practice in shifting perspectives. This explains much that is distinctive about his way of writing philosophy: why it involves so much affect and seems so given to extremes of expression. He uses different affective stances - assuming them for a while - in order to show us features of reality that are visible from them. More importantly, by moving from one perspective to another, he attempts to show philosophers the kind of ‘objectivity’ that is required for their task: objectivity understood not as disinterested contemplation, but as a

matter of not being *locked into* any particular valuational perspective, as an ability to move from one affective set to another.

6 Philosophy and the ascetic ideal

According to Nietzsche, philosophy has been understood as an a priori discipline, a deliverance of pure reason. Given his empiricism, what role can he allow philosophy? In *Human, All Too Human* he claims to practise 'historical philosophy' and denies that it can be separated from natural science, suggesting that he counts empirical theories as philosophy if they illuminate topics of traditional philosophical concern. Attention to the more conceptual aspects of such theories might especially count as continuing the philosopher's traditional role (§§8-9 of this entry provide an example).

Further, Nietzsche's thinking on topics of traditional philosophical concern (§§4-5 of this entry) is philosophical in a more traditional sense, to the extent that it deals with conceptual as opposed to empirical matters. Such 'pure' philosophy is a matter of battling the images and pictures that beguile the mind and lead philosophers into thinking that there are purely philosophical questions to be answered concerning knowledge, truth and reality. Philosophy in this sense functions as therapy, and to the extent that Nietzsche practises it, he counts as a forerunner of Wittgenstein (see Wittgenstein, L. §§9-12).

Like Wittgenstein, Nietzsche gives language a major role in generating the problems and confusions of previous philosophy. He sometimes seems to criticize language itself for falsifying reality, holding the subject-predicate structure of Indo-European languages responsible for philosophers' propensity to think that reality itself must consist of ultimate subjects that could never be part of the experienced world: God and the ego, or indivisible atoms of matter. But he would probably say of language what he ultimately says of the senses: only what we make of their testimony introduces error. Language misleads us into traditional philosophy only if we erroneously assume that linguistic structure offers us a blueprint of reality that can be used to challenge the adequacy of empirical theories. This is similar to Wittgenstein's diagnosis that philosophical problems arise when language is taken away from the everyday tasks for which it is suited and expected to play a different game. Nietzsche's philosopher forces language to play the 'game' of affording insight into a non-empirical world.

Unlike Wittgenstein, however, Nietzsche fights the confusions of traditional philosophy to free us not *from* the need to do philosophy, but *for* what he considers the true task of genuine philosophy. And this task is not a matter of offering empirical theories. In his later work, Nietzsche insists that philosophers should not be confused with scholars or scientists, that scholarship and science are only means in the hands of the philosopher. Gradually it has become clear to him, he says, that philosophers' values are the 'real germ' from which their systems grow. While pretending to be concerned only to discover truth, philosophers have actually been wily advocates for prejudices (values) they call 'truths'. They interpret the world in terms of their own values, and then claim that their interpretation, which they present as objective knowledge, gives everyone reason to accept these values - as the Stoics justified their ideal of self-governance on the grounds that nature itself obeys laws, an interpretation they arrived at by projecting their ideal of self-governance onto nature (*Beyond Good and Evil* §§1-9 [*Werke* VI.2: 207-16]).

Because Nietzsche believes that interpretations of the world in terms of values provide something that is as important as truth, he wants the philosophy of the future to preserve this function of traditional philosophy. He does not, however, wish to preserve two aspects of the way in which previous philosophers have gone about this task: the lack of courage evident in their failure to recognize that they were reading values into the world rather than discovering truth, and the particular values they read into the world.

These values, he claims, have been expressions of the ascetic ideal, the ideal that takes the highest human life to be one of self-denial, denial of the natural self. Behind this ideal, which he finds in most major religions, Nietzsche locates the assumption that natural or earthly existence (the only kind he thinks we have) is devoid of intrinsic value, that it has value only as a means to something else that is actually its negation (such as heaven or nirvana). He claims that this life-devaluing ideal infects all the values supported by most religions (although *The Antichrist* retracts this in the case of Buddhism). Having come into existence in support of some form of life, values gain the support of the ascetic priest only if they are given a life-devaluing interpretation. Acts are interpreted as wrong or 'sinful', for instance, on the grounds that they are selfish or animal, that they affirm natural instincts. Traditional

('metaphysical') philosophers are successors to the ascetic priest because they interpret what they value - truth, knowledge, philosophy, virtue - in non-natural terms. In the background of their interpretations Nietzsche spies the assumption of the ascetic ideal: that whatever is truly valuable must have a source outside the world of nature, the world accessible to empirical investigation. What ultimately explains the assumption that philosophy must be a priori, and therefore concerned with a metaphysical world, is philosophers' assumption that nothing as valuable as philosophy or truth could be intimately connected to the senses or to the merely natural existence of human beings.

The philosophy of Nietzsche's early and middle periods can itself be diagnosed as an expression of the ascetic ideal. We can understand his devaluation of human knowledge in 'Truth and Lie' (his claim that human truths are 'illusions') as a response to the recognition that knowledge is rooted in the world of nature and thereby lacks the 'purity' demanded by the ascetic ideal. And we can surmise that he considered Darwinian naturalism dangerous because he saw that it deprives human life of value - if one accepts the ascetic ideal (see Darwin, C.R.). Indeed, when he embraced the truth of naturalism in *Human* (to the extent of accepting philosophy itself as an empirical discipline), he drew the conclusion that follows from the combination of naturalism and the ascetic ideal: that human life is without value. From the viewpoint of Nietzsche's later philosophy, it is hardly surprising that his early philosophy turns out to be another expression of the ascetic ideal. According to his *Genealogy*, the ascetic ideal is the only ideal of any widespread cultural importance human beings have had so far; it has dominated the interpretation and valuation of human life for millennia. To have escaped the ascetic ideal without having to work through its influence on him would have been impossible.

On the other hand, Nietzsche was also fighting the ascetic ideal. Naturalism works against the supernatural interpretation of human life that has been promoted by the ascetic ideal, and as modern science increasingly shows how much of the world can be understood in naturalistic terms, the influence of the ascetic ideal wanes. Or, rather, it goes underground. Nietzsche denies that science and the naturalism it promotes are themselves opposed to the ascetic ideal. The commitment to science is actually the latest and most noble form of the ascetic ideal, based as it is on the Platonic/Christian belief that God is truth, that truth is divine. This amounts to the assumption that truth is more important than anything else (for instance, life, happiness, love, power) an assumption Nietzsche traces to the ascetic ideal's devaluation of our natural impulses. Thus the development of science and naturalism has been promoted by inculcating the discipline of the scientific spirit - the willingness to give up what one would like to believe for the sake of what there is reason to believe - as the heir to the Christian conscience cultivated through confession. This has thereby worked against the exterior of the ascetic ideal, against the satisfaction it has provided - in particular, the sense that this life, and especially its suffering, has a meaning, that it shall be redeemed by another life. But to work against such satisfaction is not to oppose the ascetic ideal; it is simply to require more self-denial.

Nietzsche believes that we need a new ideal, a real alternative to the ascetic ideal. If philosophers are to remain true to the calling of philosophy and not squander their inheritance, they must create new values and not continue merely to codify and structure the value legislations of ascetic priests. To create new values, however, it will be necessary for philosophers to overcome the ascetic faith that truth is more important than anything else, for truth is not sufficient support for any ideal. Although they must therefore overcome the ascetic ideal to create a new one, undertaking this task responsibly requires the training in truthfulness promoted by the ascetic ideal. The overcoming of the ascetic ideal that Nietzsche promotes is thus a self-overcoming.

7 The 'death of God' and nihilism

Nietzsche is perhaps best known for having proclaimed the death of God. He does in fact mention that God is dead, but his fullest and most forceful statement to this effect actually belongs to one of his fictional characters, the madman of *Gay Science* 125 (*Werke* V.2: 158-60). Nietzsche's madman declares not only that God is dead and that churches are now 'tombs and sepulchres of God', but also that we are all God's 'murderers'. Although the madman may accept these statements as literally true, they clearly function as metaphors for Nietzsche. The 'death of God' is a metaphor for a cultural event that he believes has already taken place but which, like the death of a distant star, is not yet visible to normal sight: belief in God has become unbelievable, the Christian idea of God is no longer a living force in Western culture.

Nietzsche views all gods as human creations, reflections of what human beings value. However, pagan gods were constructed from the qualities human beings saw and valued in themselves, whereas the Christian God was given

qualities that were the opposite of what humans perceived in themselves, the opposite of our inescapable animal instincts. Our natural being could then be reinterpreted as ‘guilt before God’ and taken to indicate our unworthiness. Constructed to devalue our natural being, the Christian God is a projection of value from the viewpoint of the ascetic ideal (see §6 of this entry). That this God is dead amounts to a prediction that Christian theism, along with the ascetic ideal that forms its basis, is nearing its end as a major cultural force and that its demise will be brought about by forces that are already and irreversibly at work.

One such force, to which Nietzsche himself contributed, is the development of atheism in the West, a development that stems from Christian morality itself and the will to truth it promotes. The will to truth, a commitment to truth ‘at any price’, is the latest expression of the ascetic ideal, but it also undermines the whole Christian worldview (heaven, hell, free will, immortality) of which ‘God’ is the symbol. Inspired by the will to truth, philosophy since Descartes has progressively undermined the arguments that supported Christian doctrines, and science has given us reason to believe that we can explain all the explicable features of empirical reality without appealing to God or any other transcendent reality. Theism has thus become cognitively superfluous. In this situation we can justify atheism without demonstrating the falsity of theism, Nietzsche claims, if we also have a convincing account of how theism could have arisen and acquired its importance without being true. Even if there is no cognitive basis for belief in God, however, might not one still accept something on the order of William James’ will to believe? (see James, W. §4). Nietzsche nowhere treats this option as irrational, but he does deny that it is now a serious option for those who have taken most strictly and seriously Christianity’s ascetic morality. It may not be irrational, but it is psychologically impossible, Nietzsche thinks, to accept theism if the commitment to truthfulness has become fully ingrained, if hardness against oneself in matters of belief has become a matter of conscience. Atheism is ‘the awe-inspiring *catastrophe* of a two-thousand year discipline in truth that finally forbids itself the *lie involved in belief in God*’ (*Genealogy* III §27 [*Werke* VI.2: 427]).

Although atheism, especially among the most spiritual and intellectual human beings, undoubtedly weakens Christianity, depriving it of both creative energy and prestige, it does not bring about the death of God by itself. The modern world, as Kierkegaard had seen already, contains many other factors that weaken the influence of Christianity and its ideal; among these Nietzsche includes the development of money-making and industriousness as ends in themselves, democracy, and the greater availability to more people of the fruits of materialistic pursuits. Zarathustra’s statement that ‘when gods die, they always die several kinds of death’ suggests that just as the ascetic ideal has been accepted by different kinds of people for different reasons, the death of God and the ascetic ideal is also brought about by a multiplicity of causes that operate differently on different kinds of people. What matters, says Zarathustra, is that ‘he is gone’ (*Zarathustra* IV §6 [*Werke* VI.1: 320]).

According to Nietzsche, the loss of belief in God will initiate a ‘monstrous logic of terror’ as we experience the collapse of all that was ‘built upon this faith, propped up by it, grown into it; for example, the whole of our European morality’ (*Gay Science* §343 [*Werke* V.2: 255]). In notes made late in his career (published in *The Will to Power*), Nietzsche calls this collapse of values ‘nihilism’, the ‘radical repudiation of value, meaning, and desirability’. He predicts ‘*the advent of nihilism*’ as ‘the history of the next two centuries’, and calls himself ‘the first perfect nihilist of Europe’. However, he adds that he has ‘lived through the whole of nihilism, to the end, leaving it behind’ (*Will to Power* Preface [*Werke* VIII.2: 431-2]). Nihilism is therefore not his own doctrine, but one he diagnoses in others (including his own earlier self). He *does not* believe that nothing is of value (or that ‘everything is permitted’) if God does not exist, but that this form of judgment is the necessary outcome of the ascetic ideal. Having come to believe that the things of the highest value - knowledge, truth, virtue, philosophy, art - must have a source in a reality that transcends the natural world, we necessarily experience these things as devoid of value once the ascetic ideal itself leads to the death of God, to the denial that any transcendent reality exists.

8 Morality

Nietzsche’s criticism of morality is perhaps the most important and difficult aspect of his later philosophy. Calling himself an ‘immoralist’ - one who opposes all morality - he repeatedly insists that morality ‘negates life’. He turned against it, he claims, inspired by an ‘instinct that aligned itself with life’ (*Birth of Tragedy Preface 5* [*Werke* III.1: 13]). Whatever Nietzsche might mean by suggesting that morality is ‘against life,’ his point is *not* that morality is ‘unnatural’ because it restricts the satisfaction of natural impulses. He finds what is natural and ‘inestimable’ in any morality in the hatred it teaches of simply following one’s impulses, of any ‘all-too-great

freedom': it teaches 'obedience over a long period of time and in a single direction' (*Beyond Good and Evil* §188 [*Werke* VI.2: 110-11]). Nietzsche analyses the directive to 'follow nature' as commanding something that is either impossible (if it means 'be like the nonhuman part of nature') or inevitable (if it means 'be as you are and must be').

His objection to morality sometimes seems to be not that it is 'against life', but that it promotes and celebrates a kind of person in which he finds nothing to esteem: a 'herd animal' who has little idea of greatness and seeks above all else security, absence of fear, absence of suffering. To complicate matters still further, he sometimes uses 'morality' to refer to what he approves of, for instance, 'noble morality' and 'higher moralities'.

The last of these interpretive problems can be resolved by recognizing that Nietzsche uses 'morality' in both a wider and a narrower sense. Every ethical code or system for evaluating conduct is 'a morality' in the wider, but not in the narrower sense. A system that determines the value of conduct solely in terms of 'the retroactive force of success or failure', for instance, is an instance of 'morality' in the wider sense, but Nietzsche counts it as 'pre-moral' in the narrower sense (*Beyond Good and Evil* §32 [*Werke* VI.2: 46-7]). And it is the narrower sense Nietzsche is using when he commits himself to 'the overcoming of morality' and claims that it 'negates life'. His immoralism does not oppose all forms of ethical life. Although he opposes morality in the narrower sense, Nietzsche accepts another ethical system in terms of which he considers himself 'bound' or 'pledged'. Indeed, he claims that, contrary to appearances, 'we immoralists' are human beings 'of duty', having 'been spun into a severe yarn and shirt of duties [which we] *cannot* get out of' (*Beyond Good and Evil* §226 [*Werke* VI.2: 168]).

Why didn't Nietzsche just say that he opposed *some* moralities and call his own ethical system his 'morality'? He undoubtedly thought that would be more misleading than his use of the term in a dual sense because it would trivialize the radical nature of his position. He called himself an 'immoralist' as a 'provocation' that would indicate what distinguishes him from 'the whole rest of humanity' (*Ecce Homo* IV: §7 [*Werke* VI.3: 369]). And it could so function, he thought, even though he actually opposes morality only in the narrower sense, precisely because this is the sense 'morality' has had until now. That word has been monopolized, he thinks, for a particular kind of ethical system on which all our currently available choices for an ethics are mere variations.

Genealogy provides a genealogy of morality in the narrower sense (the sense 'morality' will have hereafter in this entry) and a complex and sophisticated analysis of that concept of morality. Although there is no agreed-upon definition, we all have a feeling for what 'morality' in this sense means. But both the feeling and the 'meaning' are actually products of a complicated historical development that synthesized meanings of diverse origins into a unity, one that is difficult to dissolve or analyse and impossible to define. If conceptual analysis were a matter of formulating necessary and sufficient conditions for the use of a term, we might analyse the concept of morality by specifying the characteristics that are both necessary and sufficient to qualify a code of conduct as 'a morality'. But this approach has never delivered great clarification, and Nietzsche's understanding of concepts explains why: our concepts need clarification precisely because they are products of a complicated historical development. Different strands have been tied together into such a tight unity that they seem inseparable and are no longer visible as strands. To analyse or clarify such a concept is to disentangle these strands so that we can see what is actually involved in the concept. History can play a role in analysing a concept because at earlier stages the 'meanings' that constitute it are not as tightly woven together and we can still perceive their shifts and rearrangements. Looking at the history of the corresponding phenomenon can therefore make it easier for us to pick out the various strands that make up the concept and better able to recognize other possible ways of tying them together. Nietzsche's genealogy of morality aims to show that there are distinct aspects of morality, each with a separate pre-moral source, which makes the synthesis we call 'morality' something that can be undone, so that its strands might be rewoven into a different form of ethical life.

The three essays of *Genealogy* separate out for examination three main strands of morality: the good (in the sense of virtue), the right (or duty), and a general understanding of value. Each essay focuses on the development of one strand without paying much attention to the other two, even though any developed form of ethical life will actually involve all three aspects in some form and interconnection. The overall account of morality constitutes a 'genealogy' precisely because it traces the *moral* version of each strand back to pre-moral sources - thus to *ancestors of morality*. Its upshot is that what we call 'morality' emerged from these pre-moral ancestors when the right and the good become tied together under the interpretation of value provided by the ascetic ideal. This

explains why Nietzsche claims that morality 'negates life': morality is an ascetic interpretation of ethical life.

The first essay in *Genealogy* finds the central pre-moral ancestor of morality's idea of goodness among politically superior classes in the ancient world whose members called themselves 'the good' and used 'good' and 'virtue' as their marks of distinction, the qualities that distinguished them from commoners or slaves. 'Good' and 'virtuous' were the same as 'noble' in the sense of 'belonging to the ruling class'; their contrasting term was 'low-born' or 'bad' (the German, *schlecht*, originally meant 'simple' or 'common').

As Nietzsche uses 'bad' (he does not claim to reflect contemporary usage), it involves no connotation of blame, whether applied to the poor person or the liar. 'Bad' certainly expresses a value judgment: that the person so described is inferior. The nobles regard themselves as superior and look down on the bad (sometimes with contempt, sometimes with pity). But they do not blame them for being inferior, or think that the inferior ought to be good (much less that inferiority deserves punishment or goodness a reward). Such judgments make sense only if one is judging inferiority in moral terms - that is, if 'bad' has become 'morally bad' or 'evil'.

To explain the origin of the good/evil (the specifically moral) mode of valuation, Nietzsche postulates a 'slave revolt in morality', a revaluation inspired by *ressentiment* (grudge-laden resentment) against the nobles. Nietzsche does not claim that the nobles' actions were considered wrong because they were resented. He is dealing only with ideas of goodness or virtue in this essay; he seeks to explain how goodness became connected to praise and blame, reward and punishment. His postulated 'slave revolt' was led not by slaves but by priests, the 'great haters' in human history precisely because their spirituality is incompatible with the direct discharge of resentment and revenge. They hated the nobles not because they were oppressed by them but because the nobles considered themselves superior and had been victorious over them for the respect and admiration of the people. Because this hatred could not be expressed directly, it grew to monstrous proportions until it finally found an outlet in revaluing the nobles and their qualities as inferior. As a result, certain qualities - useful to those in a slavish or dependent position - were called 'good', not because anyone found them particularly admirable, but from a desire to 'bring down' people with the opposite qualities. Simply 'looking down' at the nobles and their qualities would not have done the trick, especially since the majority envied and admired them. Only through the transformation of bad into evil, of inferiority into something for which one could be blamed, could the revaluation succeed. Pent-up *ressentiment* could then be vented in acts of blaming and moral condemnation, which Nietzsche sees as acts of 'imaginary revenge' that 'bring down' hated opponents 'in effigy' and elevate those who do the blaming, at least in their own imagination. Blaming, for Nietzsche, evidently involves the judgment that the person blamed is deserving of punishment, in this case for their inferiority. Therefore, once 'bad' is transformed into 'evil', God and his judgment along with heaven and hell can be used to support the revaluation by winning over to it those who would not feel sufficiently elevated by mere moral condemnation of the nobles. Nietzsche suggests that this is how the issue of free will became connected to morality. Blaming or holding people responsible for their actions does not raise the issue; it is raised by holding them responsible for what they are. And that is precisely what was required for the revaluation of noble values.

9 Morality (cont.)

Priests did not invent the idea of 'evil' on the spot, however. The notion of blame required for the revaluation emerged in a quite different sphere, that of right conduct or duty, the development of which Nietzsche sketches in the second essay of *Genealogy*. The pre-moral ancestor to which this essay traces moral versions of right and wrong, duty or obligation, is the ethics of custom (*Sittlichkeit der Sitte*), an early system of community practices that gained the status of rules through the threat of punishment. These rules were perceived as imperatives, but not as moral imperatives: violation was punished, but not considered to be a matter of conscience or thought to incur guilt.

Nietzsche finds an ancestor of guilt in the realm of trade, in the creditor-debtor relation. Guilt arises, Nietzsche claims, when the idea of debt is put to the uses of the 'bad conscience', the sense of oneself as unworthy, which develops when the external expression of aggressive impulses becomes restricted to such an extent that they can be expressed only by being turned back against the self. This internalization does not take place automatically, however; human beings must learn techniques that promote it, and Nietzsche views priests as the great teachers in this field. One such technique exploits the idea that a debt is owed to ancestors (who eventually come to be perceived as gods) for the benefits they continue to bestow and for violations of community laws which represent

their will. Priests use this idea to teach the people that they must make difficult sacrifices to the gods - for example, to sacrifice one's first-born - and that certain instances of apparent bad luck and suffering constitute the extraction of payment for violations of divine law, hence are deserved punishments.

So conceived, the debts are still mere debts, material rather than moral 'owings'. The moralization of debt (and thereby of duty) removes the idea that it can simply be paid off and connects it to one's worth or goodness. This moralization takes place by means of the third strand of morality analysed in *Genealogy*, the understanding of value, which in the case of morality is guided by the ascetic ideal. We enter what Nietzsche calls the 'moral epoch' only when the divine being to whom the debt is owed is considered the highest being and is conceived in non-naturalistic or ascetic terms, as a purely spiritual being and thus as a repudiation of the value of natural human existence (see §§6-7 of this entry). What must now be sacrificed to the divine is 'one's own strongest instincts, one's "nature"' (*Beyond Good and Evil* §55 [*Werke* VI.2: 72]). The affirmation of these instincts is conceived as rebellion against God, and the normal sufferings of human life as punishments for this rebellion. The debt is now owed precisely for what one is and continues to be, for being part of the natural world. This debt can no longer be considered material, a mere debt, for while it is owed and payment must be made, it can never be paid off. And the punishment one deserves is now completely bound up with one's (lack of) goodness or virtue, which is interpreted in ascetic terms as self-denial, the denial of one's natural impulses, or at least as selflessness.

The priest now has the notion of 'evil' required for the revaluation of the noble values: the moralized notion of virtue as self-denial provides the standard against which the nobles could be judged inferior, whereas the moralized notion of debt provides the basis for blaming the nobles for that inferiority. Both notions (of virtue and duty) were moralized by being tied together under the understanding of value provided by the ascetic ideal. Morality connects duty and virtue in such a way that blameable violations of duty are taken to show lack of virtue and lack of virtue is blameable (luck has nothing to do with it). Because he sees this connection as having been brought about by means of the ascetic ideal, Nietzsche regards that ideal as a major element of morality.

His own ideal is a very different one. Named after the Greek god Dionysus, Nietzsche's ideal celebrates the affirmation of life even in the face of its greatest difficulties, and thus gives rise to a doctrine and valuation of life that is fundamentally opposed to the one he finds behind morality. Committed to finding the sources of value in life, he rejects all non-naturalistic interpretations of ethical life, those that make reference to a transcendent or metaphysical world. It therefore seems likely that what he opposes in morality is not the idea of virtue, or standards of right and wrong, but the moralization of virtue and duty brought about by the ascetic ideal. Morality 'negates life' because it is an ascetic interpretation of ethical life. By interpreting virtue and duty in non-natural terms, it reveals the assumption of the ascetic ideal: that things of the highest value must have their source 'elsewhere' than in the natural world. This is why Nietzsche says that what 'horrifies' him in morality is 'the lack of nature, the utterly gruesome fact that *antinature* itself received the highest honours as morality and was fixed over humanity as law and categorical imperative' (*Ecce Homo* IV: §7 [*Werke* VI.3: 370]).

But how is this connected to Nietzsche's complaints against 'herd morality'? 'Herd' is his deliberately insulting term for those who congregate together in questions of value and perceive as dangerous anyone with a will to stand alone in such matters. He calls the morality of contemporary Europe 'herd animal morality' because of the almost complete agreement 'in all major moral judgments'. Danger, suffering, and distress are to be minimized, the 'modest, submissive, conforming mentality' is honoured, and one is disturbed by 'every severity, even in justice'. Good-naturedness and benevolence are valued, whereas the 'highest and strongest drives, if they break out passionately and drive the individual far above the average and the flats of the herd conscience,' are slandered and considered evil (*Beyond Good and Evil* §§201-2 [*Werke* VI.2: 123-7]).

This morality does not seem to involve the ascetic ideal. In fact, it is more likely to be packaged as utilitarianism, which offers a naturalistic, and therefore presumably unascetic, interpretation of duty and virtue, in terms of happiness (see Utilitarianism). We might, in fact, formulate Nietzsche's main objection to herd morality as a complaint that there is nothing in it to play the role of the ascetic ideal: to hold out an ideal of the human person that encourages individuals to take up the task of self-transformation, self-creation, and to funnel into it the aggressive impulses, will to power and resentment that would otherwise be expressed externally. Although it horrifies him, Nietzsche recognizes the greatness of the ascetic ideal. It is the only ideal of widespread cultural importance human beings have had so far, and it achieved its tremendous power, even though it is the 'harmful

ideal *par excellence*', because it was necessary, because there was nothing else to play its role. 'Above all, a counterideal was lacking - until Zarathustra' (*Ecce Homo* III GM [*Werke* VI.3: 352]).

The problem is that the ascetic ideal is now largely dead (as part of the 'death of God'). Nietzsche thinks we need something to replace it: a great ideal that will inspire the striving, internalization, virtue, self-creation that the ascetic ideal inspired. 'Herd animal morality' is what we are left with in the absence of any such ideal. It is what morality degenerates into once the ascetic ideal largely withdraws from the synthesis it brought about. The virtuous human being no longer is anything that can stir our imagination or move us. For Nietzsche, this is the 'great danger' to which morality has led: the sight of human beings makes us weary.

10 The overhuman (*Übermensch*)

Nietzsche's apparent alternative to 'herd-animal morality' is his most notorious idea, the Übermensch. (There is no really suitable English translation for this term: 'overhuman' has been chosen instead of 'superman' or 'overman' because it seems best able to bring out the idea of a being who overcomes in itself what has defined us as human.) The idea actually belongs to the protagonist of *Thus Spoke Zarathustra*, Nietzsche's work of philosophical fiction, and it can never be assumed that Zarathustra's ideas are the same as those of Nietzsche. As the story opens, Zarathustra is returning from ten years of solitude in the wilderness, bringing human beings a gift: his teaching that humanity is not an end or goal, but only a stage and bridge to a higher type of being, the overhuman. He teaches that now that God is dead, it is time for humanity to establish this higher type as the goal and meaning of human life, a goal that can be reached only if human beings overcome what they now are, overcome the merely human.

The idea of becoming a higher kind of being by overcoming one's humanity can seem frightening. For some, it calls up images of Nazi stormtroopers seeking out 'inferior' human beings to annihilate. However, *Zarathustra* suggests that Nietzsche has something very different in mind. 'Zarathustra' is another name for Zoroaster, the founder of Zoroastrianism (see Zoroastrianism). Nietzsche claims that the historical Zarathustra 'created the most calamitous error, morality', because his doctrine first projected ethical distinctions into the metaphysical realm (as a cosmic fight between good and evil forces). Nietzsche bases his character on Zarathustra because the creator of the error 'must also be the first to recognize it' (*Ecce Homo* IV §3 [*Werke* VI.3: 365]).

Zarathustra is thus the story of a religious leader, the inventor of one of the world's oldest religions, who comes to recognize the 'error' of traditional (moralized) religions. Far from turning against every aspect of traditional religion, however, Zarathustra commits himself to its central task: urging human beings to raise their sights above their usual immersion in materialistic pursuits to recognize the outlines of a higher form of being that calls them to go beyond themselves, to become something more than they are. Zarathustra's overhuman can thus be seen as a successor to the images of 'higher humanity' offered by traditional religions. His teaching is not intended to encourage human beings to throw off the constraints and shackles of morality (something Nietzsche sees as well underway without his help). Its point is, rather, to combat the forces of barbarism by encouraging us to take on a more demanding ethical task than modern morality requires: that of becoming what Nietzsche had earlier called a 'true human being'. When he used that phrase, however, Nietzsche believed it applied only to 'those no-longer animals, the philosophers, artists, and saints'. Animal (purely natural) existence was a senseless cycle of becoming and desire, and only those who escape it by extinguishing egoistic desire counted as truly human. The saint in particular counted as 'that ultimate and supreme becoming human', in which 'life no longer appears senseless but appears, rather, in its metaphysical meaningfulness' (*Unfashionable Observations* III §5 [*Werke* III.1: 371-9]). From the viewpoint of his later philosophy, early Nietzsche's conception of true humanity is an obvious expression of the ascetic ideal; it devalues natural existence relative to something that is its opposite. Once one recognizes this opposite as unattainable (as Nietzsche did in *Human*), the conception can be seen for what it really is: a devaluation and condemnation of human life.

Nietzsche never abandoned his early belief that the modern world is threatened by forces of both conformism and barbarism and that our great need is therefore for educators who will inspire human self-overcoming by the force of a lofty ideal. But since he rejects the ascetic ideal, he must abandon his earlier image of a true human being. At the end of *Genealogy*'s second essay, Nietzsche suggests that what the overhuman overcomes is not the 'natural' inclinations against which the ascetic ideal has been directed, those that make apparent our connection to other animals, but rather the 'unnatural' inclinations, the aspirations to a form of existence that transcends nature and animality. In other words, the overhuman must overcome all the impulses that led human beings to accept the

ascetic ideal, an ideal that has so far defined what counts as 'human'. As we will see in §2, however, Zarathustra's call for the overcoming of the human is still too bound up with the old ideal.

11 The will to power

Zarathustra teaches that life itself is will to power, and this is often thought to be Nietzsche's central teaching as well. However, will to power first appears in Nietzsche's work in *Daybreak* (1881), and there it is one human drive among others, the striving for competence or mastery. It is usefully thought of as a second-order drive or will: a need or desire for the effectiveness of one's first-order will. In *Daybreak*, Nietzsche finds this drive at work in large areas of human life: in asceticism, revenge, the lust for money, the striving for distinction, cruelty, blaming others, blaming oneself. He explains the drive's apparent omnipresence in human life by saying not that life is will to power (or that power is the only thing humans want), but that power has a special relation to human happiness. He calls love of power a 'demon' because human beings remain unhappy and low-spirited if it is not satisfied even if all their material needs are satisfied, whereas power can make them as happy as human beings can be, even if everything else is taken away (*Daybreak* §262 [*Werke* V.1: 211]). In *Genealogy* (1887) he expresses a similar idea in more positive terms when he calls the will to power 'the most life-affirming drive', that is, the one whose satisfaction contributes most to finding life worth living (*Genealogy* III §18 [*Werke* VI.2: 383]).

Zarathustra claims that this 'will to be master' is found in all that lives, and that this explains why life is 'struggle and becoming', always overcoming itself, always opposing what it has created and loved: 'Verily, where there is perishing and a falling of leaves, behold, there life sacrifices itself - for power' (*Zarathustra* II §12 [*Werke* VI.1: 144-5]). But this seems a clearly anthropomorphic conception of life, the projection of the human will to power onto nonhuman nature. Nietzsche rejects anthropomorphic conceptions of nature, insists that will is to be found only in beings with intellects, and complains that Schopenhauer's idea of will 'has been turned into a metaphor when it is asserted that all things in nature possess will' (*Human* II §5 [*Werke* IV.3: 18]).

Yet Nietzsche does say that life, and even reality itself, is will to power. The idea seems to be that reality consists of fields of force or dynamic quanta, each of which is essentially a drive to expand and thus to increase its power relative to all other such quanta. However, almost all the passages to this effect are found in Nietzsche's notebooks. He actually argues that reality is will to power in only one passage he chose to publish, and this passage gives us good reason to doubt that Nietzsche actually accepted the argument. He neither says nor implies that he accepts its conclusion, and he argues against its premises in earlier passages of the same book (*Beyond Good and Evil* §36 [*Werke* VI.2: 50-1]).

Why would Nietzsche construct a rather elaborate argument from premises he clearly rejects? Perhaps it was to illustrate the view of philosophy presented earlier in the same book. Philosophers' ultimate aim, he claims, is not to obtain knowledge or truth, but to interpret the world in terms of their own values (see §6 of this entry) - to 'create [in thought] a world before which [they] can kneel' (*Zarathustra* II §12 [*Werke* VI.1: 143]). Yet they present their interpretations as true, and argue for them on the basis of amazingly 'little': 'any old popular superstition from time immemorial', a play on words, a seduction by grammar, or 'an audacious generalization of very narrow, very personal, very human, all too human facts' (*Beyond Good and Evil* Preface [*Werke* VI.2: 3-4]). This seems an apt diagnosis of Nietzsche's own argument, since he elsewhere identifies its first premise as 'Schopenhauer's superstition' and the exaggeration of a popular prejudice, and its second and third premises as part of the 'primeval mythology' Schopenhauer 'enthroned' (*Beyond Good and Evil* §§16-19 [*Werke* VI.2: 22-6]; *Gay Science* §127 [*Werke* V.2: 160]). Furthermore, the effect of the argument is an 'audacious generalization' to the whole universe of the will to power, which Nietzsche originally understands as one human drive among others. In generalizing this drive, Nietzsche can be seen as generalizing and glorifying what he values, just as he claims philosophers have always done and must do. For Nietzsche's own answer to 'what is good?' is 'everything that heightens the feeling of power in human beings, the will to power, power itself' (*Antichrist* §2 [*Werke* VI.3: 168]).

Why does Nietzsche value the will to power? He certainly came to recognize it as responsible for the violence and cruelty of human life and as the prime ingredient in what he had earlier called the 'cauldron full of witches' brew' that threatens the modern world with 'horrible apparitions' (*Unfashionable Observations* III §4 [*Werke* III.1: 363]). But he also saw it as 'the most life-affirming drive' and as responsible for the great human accomplishments - political institutions, religion, art, morality and philosophy. His basic psychological claim is that human beings are subjected to intense experiences of powerlessness and that such experience leads to depression unless some

means is found for restoring a feeling of power. What we call ‘barbarism’ is largely a set of direct and crude strategies for restoring the feeling of power by demonstrating the power to hurt others. What we call ‘culture’ is a set of institutions and strategies for achieving the same feeling in a sublimated or less direct fashion. The most important strategies have all involved directing the will to power back against the self. Such internalization is responsible for all the ethical achievements of human life, all the ways in which human beings have changed and perfected their original nature by taking on a new and improved nature. But the internalization of the will to power has been promoted by the ascetic ideal’s condemnation of our original nature, especially of the will to power. This is what Zarathustra attempts to overcome with his overhuman teaching, which directs the will to power back against the self to overcome the inclinations that led to the old ideal. He therefore does not condemn the will to power, but celebrates it.

12 Eternal recurrence

Nietzsche identifies himself above all else as the teacher of eternal recurrence, which is often interpreted as a cosmological theory to the effect that the exact history of the cosmos endlessly repeats itself. Although he did sketch arguments for such a theory in his notebooks, he actually does not argue for or commit himself to a recurrence cosmology in any work he published. And, although he presents eternal recurrence as the ‘basic conception’ of *Zarathustra*, he does not commit its protagonist to a cosmology. He identifies this ‘basic conception’ not as a cosmology, but as ‘the highest formula for affirmation that is at all attainable’ (*Ecce Homo* III §1 [*Werke* VI.3: 335]).

As first articulated in *Gay Science*, eternal recurrence is a heuristic device used to formulate Nietzsche’s Dionysian ideal (see §9 of this entry). How well disposed we are to life is to be measured by how we would react upon being told by a demon (in a manner designed to induce uncritical acceptance) that we will have to live again and again the exact course of life we are now living. Would we experience despair or joy, curse the demon or greet him as a god? Nietzsche’s ideal is the *affirmation of eternal recurrence*, to be a person who would respond to the demon with joy. This is *not* equivalent to having no regrets, since it has no implication concerning how to respond if given the choice of variations on history. Nietzsche’s ideal is to love life enough to be joyfully willing to have the whole process repeated eternally, including all the parts that one did not love and even fought against. Eternal recurrence gives him a formula for what it is to value the process of life as an end and not merely as a means.

Nietzsche’s special self-identification with eternal recurrence can be explained in terms of his view regarding the importance of the ascetic ideal and his explanation of its power: ‘a *counterideal* was lacking - until Zarathustra’. There are only two plausible candidates for the counterideal Zarathustra offers, the overhuman and the affirmation of eternal recurrence. The overhuman is one who overcomes the ascetic ideal. But, as Zarathustra first preaches it, the overhuman ideal can be seen as another variation on the ascetic ideal. Like the ascetic priest, Zarathustra treats our lives as valuable only as a means to a form of life that is actually their negation. Like the ascetic priest, he turns his will to power against human life and takes revenge against it (for the powerlessness it induces) by excluding it from what he recognizes as intrinsically valuable. The ideal of affirming eternal recurrence, in contrast, values the whole process of living, and thereby overcomes the ascetic ideal’s devaluation of human life, even while pushing us to go beyond its present form. It provides us with the image of a higher form of human life, but does not take revenge against the latter by refusing to call its higher form ‘human’. It therefore appears to be Zarathustra’s true alternative to the ascetic ideal.

It may seem, however, that happy pre-moral barbarians should be able to affirm eternal recurrence. How then can it provide an image of a higher form of human life towards which to strive, one that could inspire internalization, virtue and self-creation comparable to that inspired by the ascetic ideal? One relevant factor is Nietzsche’s hope for new philosophers who will create new values. Perhaps he did not expect his counterideal to provide the full content of new values, just as the ascetic ideal did not provide the full content of the old values. The ascetic priests did not create their values from scratch. They took over virtues, duties, forms of life that were already there and gave them a new interpretation, one that denied the value of natural human existence. Nietzsche seems to hope that new philosophers will do something comparable - that they will provide a new life-affirming interpretation of virtues, duties and forms of life that are already there. Eternal recurrence would function as the form of new values, a test that they must pass to count as non-ascetic or life-affirming. The test for teachers of new values would be: can you endorse and teach these values while affirming eternal recurrence?

If this suggestion is correct, Nietzsche's relation to the modern world is not quite as revolutionary as it sometimes appears. The role of his new philosophers is not to overturn everything, but to take what is in pieces due to the dissolution of the old interpretation of value and to provide a new interpretation. This kind of philosophizing is not just for the future, but is found in Nietzsche's own writings. He praises old virtues - justice and generosity, for instance - but gives us a new interpretation of them, a different way of seeing them as valuable. Generosity is valuable not because it is selfless, but because it exhibits the soul's richness and power. And justice is perhaps the greatest virtue not because it is disinterested or obeys a higher law, but because it is the rarest and highest mastery that is possible on earth. And Nietzsche does not merely talk about these matters. His writings show us a new kind of person and a new kind of philosopher in the virtues he exhibits in them, not least of all in the interpretations he gives of his virtues. Truthfulness or honesty, justice, generosity are all exhibited in his writings, but are given life-affirming interpretations that bring to our attention the role of the will to power in them.

This is not to say that Nietzsche's new values are simply repackaged old ones. Nietzsche's ideal leads him to value qualities that he claims have never before been considered part of greatness, such as malice, exuberance and laughter. But even in their new interpretations of old values, the aim of Nietzsche's new philosophers is to push culture in new directions, for instance, towards giving explicit expression at the higher levels of culture to what the old ideal excluded from the highest forms of life. This is what Nietzsche exhibits, for instance, in the positive and negative emotion, the exuberance and malice, the aggression and eros, that permeates his writings. At this level, his philosophy is art, but it is an art that completes and is no longer used to devalue knowledge, which can now be recognized as its sometimes contentious partner in Nietzsche's soul and writings.

See also: Genealogy §2; Nietzsche: impact on Russian thought

MAUDEMARIE CLARK

List of works

Nietzsche, F. (1967-84) *Werke: Kritische Gesamtausgabe (Complete critical works)*, ed. G. Colli and M. Montinari, Berlin: de Gruyter.(Now the standard German edition of Nietzsche's works. Volumes of the critical apparatus are still appearing, but the books and *Nachlaß* were completed by 1984. A paperback edition of the essentials appeared in 1980 as *Sämtliche Werke*.)

Nietzsche, F. (1872) *Die Geburt der Tragödie*; trans. W. Kaufmann and R. Hollindale as *The Birth of Tragedy*, in *Basic Writings of Nietzsche*, New York: Modern Library, 1968. (Nietzsche's first book, written under the influence of Wagner and Schopenhauer, with 'Attempt at a Self-Criticism', a preface written for the second edition of 1886. Important source for Nietzsche's early views of art and of the Dionysian, but too bound up with the thesis that art is more truthful than science to represent Nietzsche's later view.)

Nietzsche, F. (1873-6) *Unzeitgemäße Betrachtungen*; trans. R. Gray as *Unfashionable Observations*, Stanford, CA: Stanford University Press, 1995.(Four essays of cultural criticism from Nietzsche's early period, in which he is still trying to elevate art over science. These essays exhibit Nietzsche's early philosophy as a critique of modern culture, and the essays on history and Schopenhauer point forward to his later ideas of eternal recurrence and the overhuman. This work has also been translated as 'Untimely Meditations' and 'Unmodern Observations'.)

Nietzsche, F. (1878-80) *Menschliches, Allzumenschliches*; trans. R. Hollingdale as *Human, All Too Human*, intro. R. Schacht, Cambridge: Cambridge University Press, 1996. (Represents the beginning of Nietzsche's middle period, his turn away from Wagner, art and metaphysics, and his embrace of science and a thoroughgoing naturalism. Contains his first, relatively crude, attempts at a naturalistic understanding of moral values.)

Nietzsche, F. (1881) *Morgenröte*; trans. R. Hollingdale as *Daybreak*, ed. M. Clark and B. Leiter, Cambridge: Cambridge University Press, 1997.(The second work of Nietzsche's middle period, it represents the true beginning of Nietzsche's own way in philosophy. He called it the beginning of his 'campaign against morality,' and it contains much of his psychology of the will to power. This work has also been translated as 'Dawn'.)

Nietzsche, F. (1882) *Die Fröhliche Wissenschaft*; trans. W. Kaufmann as *The Gay Science*, New York: Vintage, 1974.(The work that inaugurates Nietzsche's final period, it announces the 'death of God' and contains Nietzsche's first formulation of his idea of eternal recurrence. Book Five, added in the second edition of 1887, contains some of Nietzsche's most important reflections on science and its connection to democracy. This work has also been translated as 'Joyful Wisdom'.)

Nietzsche, F. (1883-5) *Also Sprach Zarathustra*; trans. W. Kaufmann as *Thus spoke Zarathustra*, in *The Portable Nietzsche*, New York: Viking, 1954.(Nietzsche's work of philosophical fiction and the major source for the cosmological version of the will to power and the ideas of the overhuman and eternal recurrence. Not easily

understood without knowledge of Nietzsche's other books.)

Nietzsche, F. (1886) *Jenseits von Gut und Böse*; trans. W. Kaufmann and R. Hollingdale as *Beyond Good and Evil*, in *Basic Writings of Nietzsche*, New York: Modern Library, 1968.(Nietzsche at the height of his powers on philosophy and science, religion, morality, politics, nationality and nobility. A true masterpiece, but difficult.)

Nietzsche, F. (1887) *Zur Genealogie der Moral*; trans. M. Clark and A. Swensen as *On the Genealogy of Morality*, Indianapolis, IN: Hackett Publishing Company, 1997.(Nietzsche's most detailed analysis of and most important treatment of morality, considered by him as probably the most accessible introduction to his work and the 'touchstone' of what 'belongs' to him. Contains his most sustained discussion of the ascetic ideal, and was intended to clarify the title and contents of *Beyond Good and Evil*. This work has also been translated as 'On the Genealogy of Morals'.)

Nietzsche, F. (1888a) *Der Fall Wagner*; trans. W. Kaufmann and R. Hollingdale as *The Case of Wagner*, in *Basic Writings of Nietzsche*, New York: Modern Library, 1968.(Finished at the beginning of his final year, this is an often very funny book, and a major source for Nietzsche's views of Wagner and art.)

Nietzsche, F. (1888b) *Die Götzen-Dämmerung*, 1889; trans. W. Kaufmann as *The Twilight of the Idols*, in *The Portable Nietzsche*, New York: Viking, 1954.(Contains Nietzsche's final accounts of truth and knowledge, as well as important sections on morality and art. Fairly straightforward and more accessible than many of his other works.)

Nietzsche, F. (1888c) *Der Antichrist*, 1895; trans. W. Kaufmann as *The Antichrist*, in *The Portable Nietzsche*, New York: Viking, 1954.(Nietzsche's version of 'Why I am not a Christian'. Distinguishes original Christianity from the Pauline version and gives a relatively sympathetic portrait of the former and of the one who lived it. Fairly straightforward and accessible.)

Nietzsche, F. (1888d) *Ecce Homo*, 1908; trans. W. Kaufmann and R. Hollingdale in *Basic Writings of Nietzsche*, New York: Modern Library, 1968.(Nietzsche's autobiographical work, written in the middle of his final year, includes his own invaluable accounts of all his earlier works under the title 'Why I write such good books'. An important source for his view of art since the autobiography is clearly intended as a work of art.)

Nietzsche, F. (1888e) *Nietzsche contra Wagner*, 1895; trans. W. Kaufmann in *The Portable Nietzsche*, New York: Viking, 1954.(Nietzsche's last and shortest book, this is a compilation of passages from earlier works designed to show that Nietzsche and Wagner are 'antipodes'. A source for Nietzsche's final view of art and of the ways in which the art he promotes differs from Wagner's.)

Nietzsche, F. (1906) *Der Wille zur Macht*,; trans. W. Kaufmann as *The Will to Power*, New York: Viking, 1967. (Although treated by many, including Nietzsche's sister, as his magnum opus, this is not a book by Nietzsche. Rather, it is a collection of outlines, notes and jottings from his notebooks of 1883-8, selected and arranged by his sister and editors appointed by her.)

Nietzsche, F. (1975-84) *Briefwechsel: Kritische Gesamtausgabe*, ed. C. Colli and M. Montinari, Berlin: de Gruyter.(An edition of the complete letters to and from Nietzsche (the Nietzsche side of the correspondence was published in 1986 as *Sämtliche Briefe*). A 1969 English edition of selected letters, translated by Christopher Middleton, is reprinted as *Selected Letters of Friedrich Nietzsche*, Indianapolis, IN: Hackett Publishing Company, 1996.)

Nietzsche, F. (1979) *Philosophy and Truth: Selections from Nietzsche's Notebooks of the early 1870's*, trans. D. Breazeale, Atlantic Highlands, NJ: Humanities Press.(A valuable collection of material from the *Nachlaß* of the period surrounding the publication of *The Birth of Tragedy*. Includes 'Über Wahrheit und Lüge im aussermoralischen Sinne', trans. as 'On Truth and Lies in a non-moral sense', a widely known essay in which Nietzsche argues that, except for tautologies, all truths are 'illusions'.)

References and further readings

Ansell-Pearson, K. (1994) *An Introduction to Nietzsche as Political Thinker*, Cambridge: Cambridge University Press.(A lively introduction to Nietzsche's political thought.)

Clark, M. (1990) *Nietzsche on Truth and Philosophy*, Cambridge: Cambridge University Press.(An account of the development of Nietzsche's thinking about truth which forms the basis of the present entry. Includes chapters on truth and the ascetic ideal, the will to power and eternal recurrence.)

Deleuze, G. (1962) *Nietzsche et la philosophie*; trans. H. Tomlinson as *Nietzsche and Philosophy*, New York: Columbia University Press, 1983.(A reading of Nietzsche as anti-Hegelian; known especially for its influential distinction between active and reactive forces. An often difficult read.)

Derrida, J. (1978) *Éperons: Les Styles de Nietzsche*; trans. B. Harlow as *Spurs: Nietzsche's Styles*, Chicago, IL and London: University of Chicago Press, 1979.(Responds to Heidegger's interpretation of Nietzsche as a

metaphysician by playing with Nietzsche's texts rather than attempting to establish a truth about them. An example of philosophy as deconstruction. The translation is a French-English edition.)

Heidegger, M. (1961) *Nietzsche*, 2 vols; trans. D. Krell, San Francisco, CA: Harper & Row, 1979-82, 4 vols.(This interpretation of Nietzsche as a critic of metaphysics who was himself unable to avoid metaphysics has greatly influenced later Continental readings of Nietzsche; for advanced students who are as interested in Heidegger as in Nietzsche.)

Janz, C.P. (1978) *Friedrich Nietzsche Biographie*, Munich: Carl Hanser Verlag, 3 vols.(The standard biography; an English translation is in progress.)

Kaufmann, W. (1950) *Nietzsche: Philosopher, Psychologist, Antichrist*, Princeton, NJ: Princeton University Press, 4th edn, 1974.(A landmark study that went a long way towards destroying the picture of Nietzsche as a Nazi sympathizer and an anti-Socratic irrationalist that once dominated the English-speaking world. Still useful for this purpose.)

Lange, F.A. (1866) *Geschichte des Materialismus und Kritik seiner Bedeutung in der Gegenwart*; trans. E.C. Thomas as *History of Materialism*, Boston, MA: Osgood, 1977, 3 vols.(A history of materialism from ancient Greece to 19th century Europe which Nietzsche read during his student days at Leipzig. A major influence on his views of science and knowledge.)

Morgen, G. (1941) *What Nietzsche Means*, New York: Harper & Row.(Still a valuable and very accessible introduction to all the main themes of Nietzsche's philosophy.)

Mueller-Lauter, W. (1971) *Nietzsche: Seine Philosophie der Gegensätze und die Gegensätze seiner Philosophie (Nietzsche's philosophy of contradictions and the contradictions of his philosophy)*, Berlin: de Gruyter. (Perhaps the most important German work on Nietzsche written after the Second World War; distinguishes the apparent from the real contradictions of Nietzsche's philosophy of the will to power, and argues that Nietzsche's thinking leads him to two very different versions of the overhuman. English translation in progress.)

Nehamas, A. (1985) *Nietzsche: Life as Literature*, Cambridge, MA: Harvard University Press.(Important attempt to relate Nietzsche's philosophical views to his literary styles, and the most important examination of how Nietzsche's books *exhibit* what this entry calls Nietzsche's 'ideal'. Its accounts of the will to power and eternal recurrence are alternatives to those given above.)

Richardson, J. (1996) *Nietzsche's System*, Oxford: Oxford University Press.(A systematic, challenging account of Nietzsche's thought that provides a more metaphysical alternative to the naturalistic Nietzsche of this entry, especially in its accounts of the will to power and perspectivism. Influenced by Continental readings of Nietzsche, especially those of Heidegger and Deleuze.)

Schacht, R. (1983) *Nietzsche*, London: Routledge & Kegan Paul.(Careful and detailed survey of all the main themes of Nietzsche's philosophy. Its Nietzsche has a basically naturalistic orientation, but nevertheless accepts the will to power as the basic principle of life.)

Schopenhauer, A. (1818) *Die Welt als Wille und Vorstellung*; trans. E.F.J. Payne as *The World as Will and Representation*, New York: Dover, 2 vols, 1969.(Perhaps the major philosophical influence on Nietzsche's thought and development.)

Young, J. (1992) *Nietzsche's Philosophy of Art*, Cambridge: Cambridge University Press.(Interesting account of Nietzsche's view of art; highly critical of Nietzsche *vis-à-vis* Schopenhauer.)

Nifo, Agostino (*c.*1470-1538)

Agostino Nifo was a university teacher, medical doctor and extremely prolific writer. His books included many commentaries on Aristotle's logic, natural philosophy and metaphysics, as well as original works on topics ranging from elementary logic to beauty and love. However, his most important works had to do with the human intellect, and with Averroes' view that there is just one intellect shared by all human beings. Although he never accepted Averroes' position as true, he did initially believe that Averroes correctly interpreted Aristotle on this point. He also entered into public controversy with Pomponazzi on the question whether human immortality could be proved. Nifo's Aristotelianism reflects his interest in many different traditions of commentary on Aristotle, including medieval Latin commentators, especially Thomas Aquinas, medieval Arab commentators and their Latin followers, especially John of Jandun, but most of all the Greek commentators. Here he shows the strong influence of Renaissance humanism, which made the Greek texts available. It was when Nifo himself learned Greek that he came to abandon the notion that Averroes was an accurate interpreter of Aristotle. Nifo was also very interested in Plato and Platonism, particularly as presented by Marsilio Ficino. His careful presentations of other people's doctrines were popular in university circles for much of the sixteenth century.

1 Life and works

Agostino Nifo (Eutychus Augustinus Niphus Suessanus) was born at Sessa Aurunca, near Caserta in southern Italy. After studying the liberal arts at Naples, he went on to Padua, where he received his degree around 1492. During his stay at Padua he gained the enmity of the Franciscan theologian Antonio Trombetta, a Scotist. He also engaged in a dispute with his teacher, Nicoletto Vernia, regarding the correct interpretation of Alexander of Aphrodisias' doctrine on the human soul (see Alexander of Aphrodisias). Presumably he was also on poor terms with Pietro Pomponazzi, who was his competitor by 1495. Pomponazzi scornfully criticizes Nifo in his own early questions on the soul.

Nifo left his post at Padua in 1499 and returned to his native town. There he witnessed the difficulties and trials resulting from the Spanish occupation. Nifo, who was also a practising medical doctor, was the personal physician of the Spanish Grand Captain in 1504. He seems to have taught at Naples in 1505 (where he moved in humanist circles), at Salerno in 1506-7, and again at Naples in 1507-8, when he unsuccessfully attempted to regain his post at Padua. In 1514 he taught at Rome. Subsequently he entered into controversy with his old rival Pomponazzi. Encouraged by Ambrogio Flandino to write a reply to Pomponazzi's *De immortalitate animae (On the Immortality of the Soul)*, which had been published in 1516, Nifo wrote a work with the same title that appeared in 1518. Pomponazzi replied in turn with his *Defensorium* (1519).

Nifo taught at Pisa for the Florentines from 1519 to 1522. He seems to have then taught at Salerno from 1522 to 1526 and again from 1533 to 1535; there is evidence that he taught at Naples in 1531-2. In 1535 he declined Pope Paul III's invitation to teach natural philosophy at Rome. He was elected mayor of his native town and extended a formal welcome to the Emperor Charles V when he visited Sessa Aurunca on 24 March 1536.

Nifo was an extraordinarily prolific writer, though very little of his work is available in a modern edition. His first work was an edition of the works of Aristotle and Averroes that appeared at Venice in 1495 and 1496. In 1497 this was followed by his own commentary on the *Destructio destructionum* (the medieval Latin translation of Averroes' *Tahafut al-tahafut (The Incoherence of Incoherence)*) and a treatise on the agent sense. While at Padua, he apparently finished a commentary on a Latin translation of Aristotle's *On the Soul* that was published against his wishes in 1503, the same year that he published his own important treatise on the intellect (*De intellectu*), which differs significantly from the commentary. He wrote a long sequence of commentaries on Aristotle's logical works, on Aristotle's works in natural philosophy, and on Aristotle's *Metaphysics* (see Aristotle). In 1520 he published a textbook in elementary logic, the *Dialectica ludicra (Playful Dialectics)*. He paid some attention to political philosophy, and published *De regnandi peritia (On the Practical Knowledge of Ruling)* (1522), a free adaptation of Machiavelli's *Il principe (The Prince)* (see Machiavelli, N. §2). He also wrote on humanist topics, including a work on beauty and love which he finished at his villa Niphanus in 1529. In them he attempted to give an Aristotelian theory of beauty and love in competition with that of Plato and his followers. He finished his commentary on Aristotle's books *On Animals* in 1534. Many of Nifo's Aristotle commentaries continued to be

published for at least three decades after his death.

2 Natural philosophy

Nifo showed an interest in natural philosophy from the beginning. In his early commentary on Averroes' *Destructio destructionum* he emphasizes that inasmuch as natural philosophy takes the senses as its starting point, it cannot reach a knowledge of God's act of creation, which is a form of making that is totally beyond sense knowledge. It was thus a grievous error on the part of the medieval Christian philosopher-theologians, the Latin Expositors (*expositores latini*), to have attributed a doctrine of creation to Aristotle. He was also concerned about the relationship between natural philosophy or natural science (*naturalis scientia*) and metaphysics, arguing that natural philosophy was subordinate to metaphysics. It is true that the existence of God and the separate substances (or purely immaterial beings) is demonstrated by natural philosophy on the basis of eternal motion. However, the existence of these beings is studied in a more complete fashion in metaphysics, which thus takes us beyond natural philosophy.

Nifo's commentary (1508) on Aristotle's *Physics* is remarkable for what it shows of his changing use of sources in response to a newly acquired mastery of Greek. Nifo openly admits that in his youth he asserted that the position of Averroes on the intellect is the same as that of Aristotle (see Averroism), but he explains that after having read and examined Aristotle's words in the Greek he came to see that Averroes' position is nonsense (*deliramentum*). Because Averroes, who did not know Greek, was working with erroneous translations which he could not correct, he misrepresents Themistius and generally acts like a drunkard. Nifo even stresses that he prefers himself to err with the Greeks in explicating a Greek author, rather than to think correctly with barbarians who lack command of the language. In the commentary, Nifo makes frequent reference to the Greek commentators on Aristotle, namely Alexander of Aphrodisias, Simplicius, Themistius and John Philoponus, who are all mentioned in his preface. In the same place however, he states his strong commitment to Aquinas, his 'trustworthy leader', to whom he gives, following John of Jandun and others, the honorific title of 'the Expositor'.

In both the commentary on Averroes and the commentary on the *Physics*, Nifo also mentions the Oxford Calculators (see Oxford Calculators). In the earlier commentary, he dismisses their work on the grounds that they follow mathematical principles and not natural principles. In the later work, he remarks that the Peripatetics (see Peripatetics) never dreamed of such a doctrine as that of the intension and remission of forms; it is a new question invented by idle men who want to be noticed.

3 Psychology

Both of Nifo's earliest works, the treatise on the agent sense and the commentary on Averroes' *Destructio destructionum*, treat psychological topics. The first treatise takes up the question of whether an agent sense is required in the process of sensation. Aristotle had argued that an agent intellect is required to initiate the process of understanding, and John of Jandun had argued that, in a parallel manner, an agent sense is required to initiate the process of sense perception. The activity of external objects and their impact on our passive sense organs via the reception of sensible species is not enough to explain how perception can occur. Nifo argues that Jandun was wrong. In the second treatise, he argues that Averroes' doctrine of the unity of the intellect - which is that all human beings share one intellect - is the correct interpretation of Aristotle. He calls Averroes 'Aristotle transposed' and 'Aristotle's priest'.

In his early commentary (published in 1503) on Aristotle's *On the Soul*, Nifo continues his praise of Averroes, whom he calls 'the Arab Aristotle' (see Ibn Rushd), and his attacks on John of Jandun. This time he asserts that John of Jandun was wrong in believing that Averroes acknowledged the existence of intelligible species as part of the process of understanding. Other striking positions that Nifo adopts are his belief that according to philosophy one can achieve an intuitive knowledge of God during this life, and his insistence that according to Aristotle and Averroes there is no real distinction between psychological faculties, notably the intellect and the will. He uses the phrase 'potestative whole' (*totum potestativum*) of the human soul, thus emphasizing his view that it is a single power, and also reflecting the influence of Albert the Great. In the commentary, Nifo shows a close knowledge both of Themistius' paraphrases on Aristotle's *On the Soul* in the translation of Ermolao Barbaro and of a now lost Latin translation of the commentary on *On the Soul* traditionally attributed to Simplicius. He compares both these sources to Plotinus, whom he knew from Marsilio Ficino's translation of the Enneads and commentary (see Ficino,

M.).

In 1503 Nifo published *De intellectu (On the Intellect)*, his major psychological work. Book I is divided into five treatises (*tractatus*) that are devoted to fundamental topics: the origin and immortality of the soul, the nature of the soul's separability from the body, the unity of the soul, the agent and possible intellects, and the speculative and practical intellects. At the end of each treatise Nifo sets forth what he calls the 'true position'. In fact, these passages are borrowed word for word from Albert the Great's *De natura et origine animae (On the Nature and Origin of the Soul)*. The same practice is continued in Book II, which concerns the felicity or 'state' (*status*) of the soul, a term found earlier in Albert and Jandun. While Nifo bases the 'true position' primarily on Albert the Great, it should be noted that Aquinas and Ficino play secondary roles. Nifo's remarks on individuation are basically Thomistic, and his emphasis on moral arguments for immortality seem to be inspired by his study of Ficino.

Some themes from Nifo's earlier writing reappear in *De intellectu*. He consistently attacks Jandun as misrepresenting Averroes' doctrines, especially by postulating that Averroes accepted intelligible species in his theory of intellectual cognition. He also rejects Aquinas' doctrine of a real distinction between the intellect and the will. But the most outstanding feature of *De intellectu* is Nifo's complete reversal of his position on Averroes. Although he spends much effort on determining the correct interpretation of Averroes' views on the soul and intellect, he now rejects Averroes as a true interpreter of Aristotle. He offers arguments against Averroes culled from Albert the Great, Thomas Aquinas and Marsilio Ficino, drawing especially on Aquinas' *De unitate intellectus contra Averroistas (On the Unity of the Intellect Against the Averroists)*, and on Ficino's *Theologia platonica (Platonic Theology)*. He states decisively that the immortality of the soul can be demonstrated, and he offers arguments drawn from Albert, Aquinas and Ficino. He attacks Alexander of Aphrodisias for denying immortality, and criticizes Vernia for proposing Alexander as a proponent of immortality. In like fashion, without mentioning him by name, he rejects Duns Scotus' view that immortality was a neutral topic because equally good arguments can be given for and against it (see Duns Scotus, J.).

A remarkable feature of *De intellectu* is Nifo's command of the thought of Plotinus, Alexander, Themistius and Simplicius, and his attempt to delineate their positions on the different topics that he treats. He also summarizes the contents of Siger of Brabant's *De intellectu*, thus serving as the primary source for knowledge of that now lost work (see Siger of Brabant).

4 Metaphysics

As early as his commentary on the *Destructio destructionum*, Nifo adopted a conceptual scheme of the hierarchy of being and a participation metaphysics that reflected his dependence on the thought of Aquinas (§9). God exists by virtue of his essence, and all other beings exist by virtue of their participation in God's existence. Moreover, this participation comes in degrees: higher creatures participate more fully in existence than do lower creatures, and so we get the hierarchically ordered system that has been called the Great Chain of Being. What is noteworthy about Nifo's position here is not so much the basic doctrines as his attribution of them to Averroes. In this attribution he was followed by others, such as Marcantonio Zimara.

Another of Nifo's early treatments of metaphysics is a commentary on Book XII of Aristotle's *Metaphysics*, published in 1505. Among the doctrines discussed by Nifo in his commentary are the movement of the heavenly bodies, the nature of divine cognition, the union of the human intellect with God, the infinity of God's power, and the relation of God and the Intelligences. Two constant themes are Aquinas' ability to interpret even corrupt texts correctly, and Averroes' misreadings, owing to his ignorance of Greek and consequent reliance on faulty translations. Nifo added that he was none the less forced to expound Averroes, since no one is considered an Aristotelian unless he is an Averroist. This early commentary was later incorporated into a commentary on the *Metaphysics* as a whole, published at Venice in 1547 after Nifo's death.

Nifo's own major work on metaphysics is his *Metaphysicarum disputationum dilucidarium (Clarification of Metaphysical Disputations)*, completed at Naples in 1510. One of the themes that he takes up is that of the transcendentals, being (*ens*), good (*bonum*), one (*unum*) and true (*verum*), so-called because they transcend Aristotle's ten categories. Nifo argues that no order can be established among the transcendentals, since they are all equally convertible with being. Nifo also notes that he has never read Aristotle calling being and one transcendentals (*transcendentia*) at all. Another topic of discussion was the distinction between existence (*esse*)

and essence (*essentia*) in created things. Nifo takes Aquinas to hold that in every caused thing existence and essence are really and simply distinct, principally on the grounds that that which participates is distinct from that in which it participates. Nifo argues that the Peripatetic position is that there can be no composition of essence and existence in the separate substances or in the heavenly bodies, but that Aquinas, as a Catholic though not as a natural philosopher, was forced to posit that essence and existence were distinct even in the Intelligences, since he considered them able not to exist. Nifo is quick to add that as a good Peripatetic, Aquinas rarely if ever departs from the thinking of Aristotle, and he adds that he himself rarely if ever departs from Aquinas.

See also: Aristotelianism, Renaissance; Aristotle Commentators; Averroism; John of Jandun; Pomponazzi, P.

EDWARD P. MAHONEY

List of works

For details of the publication of Nifo's Aristotle commentaries, see Lohr (1988).

Nifo, A. (1495, 1497) *Destructiones destructionum Averroys cum Augustini Niphi de Suessa expositione. Eiusdem Augustini quaestio de sensu agente (Agostino Nifo's Exposition of Averroes' Destructio destructionum. The Same Agostino's Question on the Agent Sense)*, Venice: Octavianus Scotus, 1497.(Nifo began his commentary on the medieval Latin version of Averroes' *Tahafut al-tahafut (Incoherence of the Incoherence)* in 1494 and finished it at Padua in 1497. He accepts Averroes' interpretation of Aristotle on the unity of the intellect and ascribes to him a participation metaphysics. His question on the agent sense, completed at Padua in 1495, was written to discredit Jandun's postulating such a faculty. God provides the spiritual factor needed for sensation to occur.)

Nifo, A. (1503) *Super tres libros De anima (On the Three Books On the Soul)*, Venice: Petrus de Quarengis. (Completed at Sessa in 1498 and published against Nifo's wishes. Although he knows Greek and cites Themistius and Simplicius, he considers Averroes to have correctly interpreted Aristotle on the unity of the intellect.)

Nifo, A. (1503) *Liber de intellectu (Treatise on the Intellect)*, Venice: Petrus de Quarengis.(Examination of such topics as immortality, unity of the intellect and cognitive union of humans to separate intellects and God. Rejects Averroes as reliable interpreter of Aristotle and gives arguments against unity of intellect and for personal immortality. True position on all major topics derived from Albert the Great.)

Nifo, A. (1505) *In duodecim metà tà physiká seu metaphysices Aristotelis et Averrois volumen (Volume on Book XII of the Metaphysics of Aristotle and Averroes)*, Venice: Alexander Calcedonius.(This is a commentary only on Book XII of Aristotle's *Metaphysics*. Nifo questions Averroes' interpretations.)

Nifo, A. (1508) *Aristotelis physicarum acroasum liber interprete atque expositore Eutyco Augustino Nypho (Aristotle's Book of Physical Discourses, Agostino Nifo the Translator and Expositor)*, Venice: Octavianus Scotus, 1508.(Completed in 1506 after only one year of work, the commentary reveals particular interest in Themistius and Aquinas.)

Nifo, A. (1510) *Metaphysicarum disputationum dilucidarium (Clarification of Metaphysical Disputations)*, Naples: Sigismundus Mayr Alemanus, 1511.(Begun at Salerno in 1507 and finished at Naples in 1510. Studies such topics as metaphysical participation, distinction of essence and existence, transcendentals, individuation, divine infinity and intuitive knowledge of God. The Venice 1559 edition of the *Dilucidarium* was reprinted by Minerva in 1967.)

Nifo, A. (1518) *De immortalitate animae libellus (Small Book on the Immortality of the Soul)*, Venice: Heirs of Octavianus Scotus.(Written against Pietro Pomponazzi's work *De immortalitate animae* and dedicated to Pope Leo X. Insists that Aristotle held to personal immortality. Offers arguments for immortality from different traditions: from Plato and Plotinus but also from Albert the Great and Aquinas. Human free choice shows the soul to be immaterial and immortal.)

Nifo, A. (1520) *Collectanea ac commentaria in libros de anima (Collected Remarks and Commentaries on the Books on the Soul)*, Venice: Heirs of Octavianus Scotus, 1522.(The *Collectanea* are Nifo's early commentary published in 1503; the *Commentaria* are a later commentary completed at Pisa in May 1520. Nifo at this point considers Averroes' explications of Aristotle worthless since he used bad translations. He therefore follows Themistius and Simplicius.)

Nifo, A. (1520) *Dialectica ludicra tyrunculis atque veteranis utillima (Playful Dialectics Useful for Beginners and Veterans)*, Florence: Heirs of Philippus Juncta.(An introduction to logic expressly opposed to recent logicians and claiming to be in harmony with real Aristotelians, the Greek commentators. Refers the reader to his commentary on the *Prior Analytics* (1524).)

Nifo, A. (1522) *De regnandi peritia (On the Practical Knowledge of Ruling)*, Naples: Aedes Dominae Catherine

de Silvestro, 1523; trans. S. Pernet-Beau, *Une réécriture du Prince de Machiavel: le De Regnandi Peritia de Agostino Nifo*, Paris: Université Paris X-Nanterre, 1987.(Completed at Sessa in 1522. Derivative from Machiavelli's *Il principe (The Prince)*.)

Nifo, A. (1524) *Prioristica commentaria (Commentary on the Prior Analytics)*, Naples: Evangelista Papiensis, 1528.(Completed at Salerno in 1524 and dedicated to Cajetan, that is, Thomas de Vio. Heavily dependent on John Philoponus and overall critical of Latin commentators with the exception of Aquinas.)

Nifo, A. (1529) *De pulchro et amore liber (On Beauty and Love)*, Rome: Antonius Bladus, 1531; trans. F. Sotas, *Sobre la belleza y el amor*, Seville: Secretariado de Publicaciones de la Universidad de Sevilla, 1990. (Completed at Niphanus, near Sessa, November 1529, this work attempts to construct a theory of beauty and love based on Aristotelian concepts and competing with the theory of Marsilio Ficino. Notable erotic element.)

Nifo, A. (1534) *Expositiones in omnes Aristotelis libros: De Historia animalium lib. IX; De Partibus animalium et earum causis lib. IV; ac De Generatione animalium lib. V (Expositions on All Aristotle's Books On Animals)*, Venice: Hieronymus Scotus, 1546.(Completed at Salerno in 1534 but published only after Nifo's death. Follows John Philoponus. Says Albert the Great 'babbled' and 'stuttered'.)

Nifo, A. (1535) *Aristotelis Stagiritae Topica inventio a magno Augustino Nipho Medice interpretata atque exposita (Aristotle's Invention, the Topics, Translated and Explained by Agostino Nifo Medici)*, Venice: Octavianus Scotus.(Completed at Niphanus in 1535. Considers it evil to attempt to explain Aristotle's works without knowing Greek and using the Greek commentators. Both Averroes and the Latins followed erroneous translations.)

Nifo, A. (1547) *Expositiones in Aristotelis libros metaphysices (Expositions on Aristotle's Metaphysics)*, Venice: Hieronymous Scotus.(A close textual study published posthumously. He cites his own Greek manuscripts and his *Dilucidarium* (1510), showing great esteem for Aquinas. Covers participation, analogy, infinity of God, divine Ideas and divine knowledge. The Venice 1559 edition of the *Expositiones* was reprinted by Minerva in 1967.)

References and further reading

Ashworth, E.J. (1976) 'Agostino Nifo's Reinterpretation of Medieval Logic', *Rivista critica di storia della filosofia* 31: 354-74.(A study of Nifo's logic textbook, the *Dialectica ludicra* (1520), and how it modified medieval doctrines.)

Lohr, C.H. (1988) *Latin Aristotle Commentaries. II. Renaissance Authors*, Florence: Olschki, 282-7.(Listing of all Nifo's Aristotle commentaries in various editions and where they are to be found. Publishers are not uniformly indicated.)

Mahoney, E.P. (1971) 'Agostino Nifo's *De Sensu Agente*', *Archiv für Geschichte der Philosophie* 53: 119-42. (Discusses Nifo on Averroes, Albert the Great, Giles of Rome and John of Jandun regarding the notion of an agent sense.)

Mahoney, E.P. (1983) 'Philosophy and Science in Nicoletto Vernia and Agostino Nifo', in A. Poppi (ed.) *Scienza e filosofia all'Università di Padova nel quattrocento*, Contributi alla storia dell'Università di Padova 15, Padua: Lint, 135-202.(Examines Nifo's early discussions on subalternation of the sciences, nature of matter, intension and remission of forms, composition of the heavenly bodies, and the velocity of their movement; pages 173-200 focus on Nifo rather than Vernia.)

Mahoney, E.P. (1986) 'Marsilio Ficino's Influence on Nicoletto Vernia, Agostino Nifo and Marcantonio Zimara', in G.C. Garfagnini (ed.) *Marsilio Ficino e il ritorno di Platone: Studi e documenti*, vol. 2, Florence: Olschki, 509-31.(Emphasizes Nifo's heavy use of Ficino's work, especially on the soul and the intellect.)

Mahoney, E.P. (1989) 'Plato and Aristotle in the Thought of Agostino Nifo (ca. 1470-1538)', in G. Roccaro (ed.) *Platonismo e aristotelismo nel mezzogiorno d'Italia (secc. XIV-XVI)*, Palermo: Officina di Studi Medievali, 81-102.(Complete survey of the role of Plato and Aristotle in Nifo's writings with a useful bibliography.)

Mahoney, E.P. (1993) 'Agostino Nifo and Neoplatonism', in P. Prini (ed.) *Il Neoplatonismo nel Rinascimento*, Rome: Istituto della Enciclopedia Italiana, 205-31.(Detailed examination of Nifo's works delineating his acquaintance with the writings of Plotinus and other Neoplatonists.)

Nardi, B. (1958) *Saggi sull'aristotelismo padovano dal secolo XIV al XVI (Essays on Paduan Aristotelianism from the 14th to the 16th Century)*, Università degli Studi di Padova: Studi sulla tradizione aristotelica nel Veneto 1, Florence: Sansoni.(A series of studies on Aristotelianism at the University of Padua which help to place Nifo in context.)

Poppi, A. (1966) *Causalità e infinità nella scuola padovana dal 1480 al 1513 (Causality and Infinity in the Paduan School from 1480 to 1513)*, Padua: Antenore, 222-36.(Critical discussion of Nifo on God's infinity.)

Nihilism

As its name implies (from Latin nihil*, 'nothing'), philosophical nihilism is a philosophy of negation, rejection, or denial of some or all aspects of thought or life. Moral nihilism, for example, rejects any possibility of justifying or criticizing moral judgments, on grounds such as that morality is a cloak for egoistic self-seeking, and therefore a sham; that only descriptive claims can be rationally adjudicated and that moral (prescriptive) claims cannot be logically derived from descriptive ones; or that moral principles are nothing more than expressions of subjective choices, preferences or feelings of people who endorse them.*

Similarly, epistemological nihilism denies the possibility of justifying or criticizing claims to knowledge, because it assumes that a foundation of infallible, universal truths would be required for such assessments, and no such thing is available; because it views all claims to knowledge as entirely relative to historical epochs, cultural contexts or the vagaries of individual thought and experience, and therefore as ultimately arbitrary and incommensurable; because it sees all attempts at justification or criticism as useless, given centuries of unresolved disagreement about disputed basic beliefs even among the most intelligent thinkers; or because it notes that numerous widely accepted, unquestioned beliefs of the past are dismissed out of hand today and expects a similar fate in the future for many, if not all, of the most confident present beliefs.

Political nihilism calls for the complete destruction of existing political institutions, along with their supporting outlooks and social structures, but has no positive message of what should be put in their place. Cosmic nihilism regards nature as either wholly unintelligible and starkly indifferent to basic human concerns, or as knowable only in the sense of being amenable to scientific description and explanation. In either case, the cosmos is seen as giving no support to distinctively human aims or values, and it may even be regarded as actively hostile to human beings. Existential nihilism negates the meaning of human life, judging it to be irremediably pointless, futile and absurd. Cosmic and existential nihilism are the focus of this entry.

1 Historical background

Scattered uses of the term 'nihilism' can be found in philosophical, theological, political and literary writings of the late eighteenth and early nineteenth centuries in Europe. In that period, the term was sometimes used to refer to atheism and its alleged inability to provide support for knowledge and morality, or to impart purpose to human life; sometimes to any sort of uncommitted, sceptical or despairing outlook on existence; sometimes to philosophical idealists (especially proponents of the 'critical' philosophy of Immanuel Kant), on the basis that they negate any possibility of knowing or relating to objective, in-itself facts of the world by insisting that all experienced objects are constructions of the mind; and sometimes to socialists or anarchists, because their revolutionary philosophies were judged to be destructive of all political and social order. But the term first came into wide use in the period extending from the 1870s into the early years of the twentieth century, due largely to the influence of three writers: the Russian novelists Ivan Turgenev and Fëdor Dostoevskii, and the German philosopher Friedrich Nietzsche.

A character named Bazarov, in Turgenev's novel*Fathers and Sons* (1862), proclaims himself a nihilist and explains that nihilists 'act by virtue of what we recognize as beneficial'. What is most beneficial for the present, he continues, is 'negation, and we deny…everything'. Bazarov's nihilism quickly became famous in Russia and was warmly endorsed by certain revolutionary groups there in the 1860s, as well as being denounced by defenders of traditional beliefs and existing religious and political institutions. Dostoevskii was one such adamant defender, and his novel *The Possessed* (1871-2) gives a lurid portrayal of the lives of three members of a nihilist society, two of whom die by their own hand, the third at the hands of his comrades. Their existence is wretched and empty, the author implies, because they have lost faith in God and think that they must now arrogate to themselves God's absolute freedom. But theirs is an absurd freedom, without guidance or norms. It ravages their lives and brings them to pointless deaths. Dostoevskii propounded the theme of ruinous nihilism following inevitably from atheism through much of his literary career.

In notebooks composed between 1883 and 1888, parts of which were later published under the title *The Will to Power* (1901), Nietzsche announced 'the advent of nihilism' in European culture and ruminated at length on its character and causes (see Nietzsche, F. §7). There was one sense, he thought, in which nihilism could be attributed

to the declining influence of Christianity and loss of faith in God, for without God human life seemed deprived of purpose and value. But in another sense, Western civilization was now coming to realize that Christianity had stripped the world of immanent meaning - thus containing in itself the seeds of the most thoroughgoing nihilism - by insisting on the necessity for a transcendent ground of truths and values, and focusing most of its attention on attainment of a paradise beyond the grave where the trials and sorrows of earthly existence would be redeemed.

The task to which Nietzsche devoted much of his writing, therefore, was twofold. The first part was to do relentless battle with Christianity and lay bare what he regarded as its implicit nihilism, thus freeing European culture from its spell. The second was to find some way to move beyond the nihilistic desolation that Nietzsche's character Zarathustra foresees in *Thus Spoke Zarathustra* (1883-85), following hard on the heels of his startling discovery of 'the death of God'. Nietzsche condemned as 'passive nihilism' the despairing resignation with which Arthur Schopenhauer and other radical pessimists responded to the looming crisis of the West. He sought to develop a vision of an 'active nihilism' that would fully acknowledge important findings of epistemological, moral and cosmic nihilism, but convert them into stepping stones to a new affirmation of life.

2 Cosmic nihilism

The absolute form of cosmic nihilism denies to the universe any sort of intelligibility or meaning. It is blank and featureless, giving no response to the age-old human search for understanding and no support to distinctively human aims, aspirations or purposes. All attempts to comprehend the world, including those of the various natural sciences, are doomed to failure. Whatever meanings or values we may think we have found in it stem from unconscious projections of our own suppositions and wishes, because it contains no knowable traits, principles or patterns that would enable it to be understood, and it is utterly lacking in discernible value. This total alienation of human understanding and need from the surrounding universe is asserted by the German thinker Max Stirner, when he states that the world 'cannot be regarded as a comprehensive structure of objective meanings' but must rather be seen 'as a metaphysical chaos'. This conclusion follows from Stirner's radical nominalism, which views the world as a loose aggregation of unique, self-contained particulars which exhibit no intelligible relations to one another. As for values, Stirner's nominalism led him to the view that we must fend for ourselves; individuals are left isolated in a chaotic world to find as much private enjoyment and achieve as much power over others as they can.

A similar bleak view of the cosmos is taken by Antoine Roquentin, a character in Jean-Paul Sartre's novel *Nausea* (1938). The fundamental discovery he makes is that 'the world of explanations and reasons is not the world of existence'. An uncrossable chasm yawns between the concepts, descriptions, explanations and valuations of human beings on the one hand, and the world on the other. From the perspective of his recurring experiences of 'nausea,' Roquentin perceives the world to be like a thick paste or oozing slime, without character or distinction, and impenetrable to the mind (see Sartre, J.-P. §4). In much the same fashion, Albert Camus speaks in *The Myth of Sisyphus* (1942) of the world as 'dense' and 'strange', a 'vast irrational' wholly unresponsive to human attempts to find a home within it or to comprehend its meaning. Schopenhauer, in the early nineteenth century, went even further than these thinkers in denuding the world of value. He argues in *The World as Will and Idea* (1818) that the cosmos is steered by a blind impulse or energy that is not merely oblivious to matters of good and evil but actively hostile to human beings, generating a horrible preponderance of sufferings over pleasures and disappointments over satisfactions.

The relative form of cosmic nihilism sees the world as fully open to scientific understanding but denies to it any normative significance. The real universe, in this outlook, is the purely descriptive one being brought steadily into view by the natural sciences, a universe of factual principles and laws that gives no place or support to the moral, aesthetic and religious concerns that preoccupy human life. The French biochemist Jacques Monod develops this idea in *Chance and Necessity* (1970), where he identifies 'the essential message of science' as an objective portrayal of the cosmos that 'outrages values' by making them 'seem to melt into the world's uncaring emptiness'. Humans must somehow reconcile themselves to living 'on the boundary of an alien world' just as indifferent to their hopes as it is to their sufferings or crimes.

3 Existential nihilism

Existential nihilists contend that human existence has no purpose, value, or justification. There is no reason to live,

and yet we persist in living. The human situation is therefore absurd. The philosophical position of existential nihilism is not just that this or that person may fail to find meaning in life; it is rather that a genuinely meaningful life is impossible. In the face of this conclusion, Schopenhauer counsels us to quench the flame of affirmation and desire, to resign ourselves to the span of our days without hope of respite and calmly await an annihilating death, the last sure sign of the utter futility of our existence. Sartre too hammers at the absurdity of existence but counsels us in *Being and Nothingness* (1943) to invent meanings for our lives through sheer acts of freedom, creating those meanings - in the manner of the God of traditional religion - out of nothing. Camus, in *The Myth of Sisyphus*, finds the absurdity of the human situation to lie in our restless, futile search for comprehensive meaning in a universe that has no discoverable significance or value. He advises us not to commit suicide when we realize that our lives have no point - that would be the coward's way out. Instead, we should heroically rebel against the abyss of meaninglessness and in that very act of defiance find some semblance of a reason for being (see Existentialism).

A relentless, uncompromising exponent of existential nihilism is the Romanian philosopher E.M. Cioran. In *A Short History of Decay* (1949), he notes that most of us strive throughout our lives to ' keep deep down inside a certitude superior to all the others: life has no meaning, it *cannot* have any such thing'. Civilizations, philosophies, religions - all are ways of masking this inescapable truth, of seeking to divert our attention from its shattering impact. How, then, can we cope with it? As against Schopenhauer, Cioran insists that we cannot escape torturing ourselves with awareness of the 'disease', 'shame' and 'curse' of our existence. Our probing, questioning minds give us no peace; only those living in shallowness and illusion can avoid their constant torture. As against Camus, Cioran argues that we all *should* commit suicide: this is the only consistent way of dealing with the absurdity of our lives. But we foolishly compound the absurdity by refusing, in our cowardice, to do away with ourselves. As against Sartre, Cioran contends that ' the intoxication of freedom is only a shudder within a fatality, the form of [our]…fate being no less regulated than that of a sonnet or a star'. For beings born only to experience the crushing inevitabilities of disappointment, suffering and death, a freedom thrown defiantly against the void can give no respite. From 'spermatozoon to sepulchre' we are pawns of a taunting fate that arbitrarily selects some for good fortune and others for bad. Each life is a useless 'hyphen' between birth and death. Human history, as evolutionary biology and scientific cosmology show, is that of one more organic species doomed, like all the rest, to extinction, and of a mere fleeting breath in the universe's onrush to entropic death. Analogous to Sartre's descriptions of 'nausea' is Cioran's acknowledgement of a sickening 'disgust' welling up from such recognitions: ' that *negative* superfluity which spares nothing…[and] shows us the inanity of life'.

4 Critical comment

That the threat of nihilism is still very much with us is indicated by Loyal Rue's book, *By the Grace of Guile: The Role of Deception in Natural History and Human Affairs* (1994). Rue states that he is convinced of two things: that nihilism is true and that human beings cannot live with its truth. Rue does not so much systematically argue for nihilism as confess himself to be firmly in its grasp. The solution he proposes is that of a 'Noble Lie', that is, a programme of mythical re-enchantment of the universe which will deceive people into believing that it and they both have significance, even though neither has. This desperate and implausible strategy shows the continuing power of nihilism, here regarded as a kind of foregone conclusion.

Martin Heidegger, a prominent critic of nihilism in its various forms, attributes what he regards as the nihilistic malaise of the West - astutely predicted by Nietzsche - to an inherited tangle of epistemological and metaphysical misconceptions that have led to a disastrous 'forgetfulness of being'. Only if we go back to the philosophical roots of the Western world, in the profound reflections of the Presocratic philosophers that Heidegger considers to have been distorted by Plato and Aristotle (these distortions having plagued us ever since), can we come to understand how the objectifying, separating, controlling and ego-centred modes of present thought and experience have led us to nihilism. Heidegger's solution lies in our learning radically to 're-vision' reality, to become receptive to the healing presence of being that shines through all beings and binds us to one another and to the world (Heidegger 1953). A powerful explication and defence of Heidegger's general approach to nihilism is contained in Levin (1988).

Arguments for cosmic and existential nihilism, like those for the other forms of nihilism, are complex and varied. Detailed analysis and criticism of some of them can be found in Crosby (1988), which probes beneath the arguments themselves to the crucial assumptions that underlie them, many of which stem from the legacy of

seventeenth- and eighteenth-century Western thought (some examples of these assumptions are given below). The book concludes that, while philosophical nihilism has important lessons to teach us, it is ultimately a distorted and one-sided view of ourselves and the experienced world, one that turns on highly questionable assumptions.

Many such assumptions lie behind some of the arguments put forward in support of cosmic nihilism. Proponents believe that if claims about the universe cannot be convincingly shown to correspond to what it is in and of itself, apart from human experience and conceptualization, then we cannot be said to have any significant understanding of the universe. Additionally, they sometimes think that the presence of purposive activity in the universe (for example, in human beings and other animals on earth) is not enough to give it meaning - for this there must be a discernible purpose of the universe *as a whole*. Such thinkers often demand that the universe must have a ‘human face’, have its ground in a personal divine Being and be centred on humans and their distinctive concerns. Others rely on natural science as the final arbiter of the nature of nature and assume that, since science discovers no values in nature but only descriptive facts and relations, nature is devoid of value. Some suppose that if, as natural science claims, the universe had an accidental beginning (the Big Bang) and will eventually come to an absolute end (its inevitable entropic death, assuming that the universe is a closed system) then it follows that the universe can have no meaning.

Similarly, proponents of existential nihilism use arguments which are informed by a number of disputable assumptions. They may believe that a life which ends in annihilating death cannot be said to have meaning, or that there is too much suffering, or constant threat of suffering, in human life for it to be meaningful. They sometimes conceive of time as a series of disconnected moments, and hence believe that our lives in time can have no cumulative pattern or significance. Often they assume that meaningful values, including the existential ones that relate to the course and direction of our lives, must be conferred on us from on high and thus have a transcendent source and sanction. Others believe that values, including existential ones, can have no significance or weight unless they are absolute, unconditioned, timeless and universal, or that our lives are meaningless if they are the outcome of a process of biological evolution, guided by the principle of natural selection that is non-purposive and non-designed. That human beings are locked in their individual subjectivities and doomed to invent arbitrary meanings for themselves, with no objective rules or standards, is yet another assumption that often underlies the existential nihilist’s argument.

All these suppositions are open to serious criticism. When they are brought under careful scrutiny, arguments that depend on them lose much of their force. Still, there is considerable truth about the world and the human condition that nihilistic arguments in general can bring home to us. Reflection on such arguments can sharpen our philosophical vision in fundamental and positive ways.

See also: Anarchism; Life, meaning of; Moral scepticism; Nishitani Keiji; Scepticism

DONALD A. CROSBY

References and further reading

Camus, A. (1942) *The Myth of Sisyphus and Other Essays*, trans. J. O’Brien, New York: Alfred A. Knopf, 1955. (Sisyphus is a symbol of human beings learning to live defiantly despite their inability to find meaning in the universe; see §2 above; relatively easy.)

Carr, K. (1992) *The Banalization of Nihilism: Twentieth-Century Responses to Meaninglessness*, Albany, NY: State University of New York Press.(Examines Nietzsche’s diagnosis of nihilism as an impending crisis of the West and compares responses to it given by Karl Barth and Richard Rorty; warns that nihilism has become so ‘banal’ in its postmodern form as to sanction a dangerous complacency about existing structures of thought, practice and value; moderately difficult.)

Cioran, E.M. (1949) *A Short History of Decay*, trans. R. Howard, Oxford: Blackwell, 1975. (Plumbs the depths of the mood and outlook of existential nihilism; mentioned in §3 above; relatively easy.)

Crosby, D.A. (1988) *The Specter of the Absurd: Sources and Criticisms of Modern Nihilism*, Albany, NY: State University of New York Press.(Explicates and critically examines arguments for moral, epistemological, cosmic and existential nihilism, with an emphasis on existential nihilism and ways the other types relate to it; see §4 above; moderately difficult.)

Dostoevskii, F. (1871-2) *The Possessed*, trans. C. Garnett, New York: Heritage Press, 1959. (A novel about the destructive lives of members of a revolutionary nihilist society in nineteenth century Russia; see §1 above; relatively easy.)

Evans, F.J. (1993) *Psychology and Nihilism: A Genealogical Critique of the Computational Model of Mind*, Albany, NY: State University of New York Press.(Argues that cognitive psychology not only reflects the currents of nihilism rampant in our time, but also reinforces them with truncated, technocratic, inadequate conceptions of human mind, human language, human creativity and human beings themselves; moderately difficult.)

Heidegger, M. (1953) *An Introduction to Metaphysics*, trans. R. Manheim, Garden City, NY: Doubleday & Company, Anchor Books, 1959.(Traces the rise of nihilism in the West to epistemological and metaphysical ideas going back to the Hellenic Age in Greece; see §4 above; difficult.)

Levin, D.M. (1988) *The Opening of Vision: Nihilism and the Postmodern Situation*, New York and London: Routledge.(Argues that the advent of nihilism is both cause and consequence of a corrupted egocentric and patriarchical vision; pleads for a new metaphysical and social vision that draws on the thought of Heidegger, Merleau-Ponty, Foucault and others; see §4 above; difficult.)

Monod, J. (1970) *Chance and Necessity*, trans. A. Wainhouse, New York: Vintage Books, 1972.(Natural science's 'postulate of objectivity' shows the world and human life to be devoid of value, so values must be arbitrarily created; see §2 above; moderately difficult.)

Nietzsche, F. (1883-5) *Thus Spoke Zarathustra*, trans. W. Kaufmann, New York: Viking, 1966.(Nietzsche's prophet Zarathustra proclaims the death of God and its consequences; see §1 above; moderately difficult.)

Nietzsche, F. (1901) *The Will to Power*, trans. W. Kaufmann and R.J. Hollingdale, ed. W. Kaufmann, New York: Vintage Books, 1968.(Excerpts from Nietzsche's notebooks in the 1880s that announce the advent of nihilism in the West and defend 'active' nihilism against 'passive' nihilism; see above §1; moderately difficult.)

Ritter, J. and Gründer, K. (eds) (1971-) 'Nihilismus', *Historisches Wörterbuch der Philosophie (Historical dictionary of philosophy)*, Basel/Stuttgart: Schwabe & Company, vol. 6, 846-54.(Extensive survey of uses of the term 'nihilism' in literature, philosophy, theology and politics, going back to the twelfth century.)

Rosen, S. (1969) *Nihilism: A Philosophical Essay*, New Haven, CT: Yale University Press. (Attributes the presence of nihilism in today's world to a series of past philosophical decisions; especially attacks the modern detachment of 'reason' from its traditional close association with 'good'; moderately difficult.)

Rue, L.D. (1994) *By the Grace of Guile: The Role of Deception in Natural History and Human Affairs*, New York: Oxford University Press.(Accepts the truth of what he regards as epistemological, moral and cosmic nihilism, and argues that only deception or illusion can enable human beings to live in the face of this truth; discussed in §4 above; moderately difficult.)

Sartre, J.-P. (1938) *Nausea*, trans. L. Alexander, New York: New Directions, 1964.(A novel about a character who lives nihilism through in its various forms and experiences a sickening loathing for himself, other humans, and the world; see §2 above; relatively easy.)

Sartre, J.-P. (1943) *Being and Nothingness*, trans. H. Barnes, New York: Washington Square Press, 1966.(An approach to existential meaning, epistemic truth and moral values that turns on a radical doctrine of human freedom; see §3 above; difficult.)

Schopenhauer, A. (1818) *The World as Will and Idea*, trans. R.B. Haldane and J. Kemp, 2nd edn, London: Routledge & Kegan Paul, 3 vols, 1957.(Portrays the world and humans as expressions of the inexorable workings of a cosmic will that is without purpose or meaning and causes an overbalance of suffering in all forms of sentient life, including our own; see §2 above; difficult.)

Stirner, M. (1845) *The Ego and His Own*, trans. J. Carroll, New York: Harper & Row, 1971. (Presents an ego-centred nominalism implying moral, epistemological and cosmic nihilism; see §2 above; relatively easy.)

Tillich, P. (1953) *The Courage to Be*, New Haven, CT: Yale University Press.(Draws on the ideas of Heidegger, among others, to discover resources for coping with the threats of 'non-being' in the forms of fate and death, guilt and moral self-condemnation, emptiness and meaninglessness. Finds 'the courage to be' in the 'God above God,' that is, the power of 'being-itself'; moderately difficult.)

Turgenev, I. (1862) *Fathers and Sons*, trans. M.R. Katz, New York: W.W. Norton, 1994.(A novel about the seductions of nihilism for youths in nineteenth-century Russia; see §1 above; relatively easy.)

Nihilism, Russian

The term 'Nihilist', although it was first used in Russian as early as 1829, only acquired its present significance in Turgenev's novel Ottsy i deti (Fathers and Sons) (1862), where it is applied to the central character, Bazarov. Thereafter Nihilism quickly became the subject of polemical debate in the journal press and in works of literature. The Nihilists were the generation of young, radical, non-gentry intellectuals who espoused a thoroughgoing materialism, positivism and scientism. The major theorists of Russian Nihilism were Nikolai Chernyshevskii and Dmitrii Pisarev, although their authority and influence extended well beyond the realm of theory. Nihilism was a broad social and cultural movement as well as a doctrine.

Russian Nihilism negated not the normative significance of the world or the general meaning of human existence, but rather a particular social, political and aesthetic order. Despite their name, the Russian Nihilists did hold beliefs - most notably in themselves and in the power of their doctrine to effect social change. It is, however, the vagueness of their positive programmes that distinguishes the Nihilists from the revolutionary socialists who followed them. Russian Nihilism is perhaps best regarded as the intellectual pool of the period 1855-66 out of which later radical movements emerged; it held the potential for both Jacobinism and anarchism.

1 The Nihilist movement: materialism as way of life

Russian Nihilism is a phenomenon of cultural history just as much as a current in Russian thought. Educated Russians in the nineteenth century had a marked propensity not only to reflect on but also to live by ideas, and nowhere is this more evident than in the radical movement of the 1860s. In the decade following the death of Tsar Nicholas I the new class of educated commoners (the *raznochintsy*) felt the delayed intellectual impact of materialism, which they converted into an all-consuming worldview. Characteristic of the Nihilists was a rejection of all authorities other than materialism itself. In *Ottsy i deti (Fathers and Sons)* (1862) a Nihilist is defined as someone who 'does not bow down before any authorities, who does not take a single principle on faith'. Turgenev's Bazarov trusts only the exercise of his own rational faculties in the spirit of scientific inquiry.

The leading lights of Russian Nihilism were Nikolai Chernyshevskii and Dmitrii Pisarev, whose articles appeared in *Sovremennik (The Contemporary)* and *Russkoe slovo (The Russian Word)* (both of these radical journals were closed for good after the attempted assassination of the Tsar in 1866). Chernyshevskii wrote prolifically, but by far his most influential work was *Chto delat' (What Is To Be Done?)* (1863), which became the catechism of the Nihilist 'new men'.

This novel offered them less a utopian vision than a practical programme of social action and a code of moral conduct. The main characters embodied the Nihilist ideals of equality, personal liberation and 'rational egoism'. The novel's artlessness, moreover, provided an excellent illustration of the new 'realist' aesthetics propounded by the Nihilists. Among Pisarev's most famous essays is 'Bazarov' (1862), a review of *Fathers and Sons* where he accepted Turgenev's depiction of the younger generation and thereby sparked an acrimonious debate within the Nihilist camp. Pisarev argued that Bazarov had enormous significance as a representative of the new worldview, and drew particular attention to his rejection of abstraction and his commitment to immediate and concrete action. Bazarov, unlike the earlier socially alienated 'superfluous men' of Russian literature, had both the will to act and the knowledge to act effectively. The Nihilist movement belonged to the young. It contained an element of protest not only against the existing order but also against the older generation, the intellectual 'fathers' who had been brought up on German Idealism rather than Feuerbach. The Nihilists gave short shrift to anything that smacked of sentimentality or romanticism. Rather like protesting youth in the twentieth century, the Nihilists cultivated their own style of dress and behaviour, which is particularly evident in contemporary descriptions of women radicals (*nigilistki*). One conservative newspaper noted that they were:

> usually very plain, exceedingly ungracious, so that they have no need to cultivate curt, awkward manners; they dress with no taste and in impossibly filthy fashion, rarely wash their hands, never clean their nails, often wear glasses, always cut their hair, and sometimes even shave it off.
>
> (Moser 1964: 44)

The 'woman question' did in fact occupy a prominent place in Nihilist social thought. In *What Is To Be Done?*

Chernyshevskii affirmed the importance of women’s personal liberation and sexual freedom. The Nihilist striving for women’s emancipation was reflected in the phenomenon of fictitious marriages: in the 1860s a young woman wanting to gain independence from her family might search out a young man willing to go through the nuptial rites with her, and would then be free to join a sewing cooperative or disappear abroad to study.

For the young Nihilists materialism was less a philosophical doctrine than an article of faith and a guide to action: it carried with it a new ethic of scientific inquiry, personal liberation and social activism. The Nihilists were convinced that the methods of the natural sciences would bring benefits if applied to the study of man, art and society. They asserted the prime social value of work and advocated the introduction of rational forms of social cooperation, most notably in the communal movement. But perhaps most importantly of all, they were blessed with boundless faith in themselves and their ideas. As Pisarev wrote in his essay ‘Mysliashchii proletariat’ (‘The Thinking Proletariat’) (1865), the new man ‘believes in his reason and in his reason alone’. For the Nihilist the intellectual emancipation of the individual came before social emancipation. Enlightenment was to be spread throughout society by the actions of an elite group of ‘new men’.

The contemporary observer Nikolai Strakhov suggested that ‘Nihilism itself hardly exists, although there is no denying the fact that Nihilists do’. The Nihilists undoubtedly did form a significant social movement, but their activities were underpinned by a materialist doctrine. The philosophical content of Nihilism will be examined in greater detail in the next section.

2 Nihilist doctrine

The 1860s were once described by Trotsky as ‘a brief eighteenth century’ in Russian thought. The Nihilist thinkers sought to assimilate and resynthesize the main trends in Western materialism and positivism. As usual in Russia, imported ideas were treated selectively and deployed in quite distinctive intellectual formations.

The best-known and most scandalous statements of the Nihilist position came in the field of aesthetics. In Russia from the 1840s onwards literature could be both the instrument and the subject of social, philosophical and political debate: for the Nihilists too it was both a target and a weapon. Chernyshevskii’s master’s thesis, ‘Esteticheskie otnosheniia iskusstva k deistvitel’nosti’ (‘The Aesthetic Relations of Art to Reality’) (1855), identifies the main criteria of artistic value in the new ‘realist’ aesthetic. Chernyshevskii counters idealism in aesthetics, and argues that art is always aesthetically inferior to reality. Art’s worth is thus defined by its faithfulness to reality. The artist, however, does more than provide naturalistic snapshots of life: the work of art contains not only accurate depiction, but also explanation and assessment of reality. Chernyshevskii’s views helped to spawn a new school of civic-minded criticism which assessed works of literature for the faithfulness and ‘typicality’ of their depiction of contemporary Russian society. The most brilliant representative of this school was Chernyshevskii’s friend and disciple Nikolai Dobroliubov, whose essays included an impassioned sociological reading of Goncharov’s *Oblomov* (this novel contained perhaps the most celebrated Russian depiction of social superfluousness). Pisarev summed up well the demands the Nihilists made of the artist: ‘The true, "useful" poet must know and understand everything that at a given moment interests the best, the most intelligent, and the most enlightened representatives of his age and nation’. Realist aesthetics, in its concept of the ‘typical’, did not in fact manage to part entirely with the Hegelian notion of the ideal: art was expected not simply to depict reality, but to single out the ‘most real’ part of reality.

For the Russian Nihilists, reality was the world of nineteenth-century science. Taking its lead from the contemporary German materialists Büchner, Vogt and Moleschott, their worldview left no room for anything other than force and matter. In his essay, ‘Antropologicheskii printsip v filosofii’ (‘The Anthropological Principle in Philosophy’) (1860), Chernyshevskii mounts an assault on philosophical dualism and abstraction, arguing that materialism proves the ‘unity of the human organism’. He concedes the distinction between ‘moral’ and ‘material’ phenomena, but maintains that it is one of degree rather than of kind. Chernyshevskii’s ‘anthropological principle’ was intended to prove that the methods of the natural sciences were applicable to the study of man. As these methods became more sophisticated, the moral sciences would inevitably attain complete understanding of the causes of human behaviour.

To his scientism and determinism Chernyshevskii added an ethics based on ‘rational egoism’. ‘Good’ and ‘bad’, he argues, are not inherent in human nature. Man always acts in his own interests and does what is ‘useful’ for

him. Good is thus defined as 'the superlative of utility'. Chernyshevskii's theory of rational self-interest is underpinned by the optimistic assumption that the interests of society and those of the individual naturally coincide. If individuals, acting as always in their own interests, commit 'bad' deeds, then it is the social order that must be at fault. Chernyshevskii's determinism was thus profoundly polemical. In his essay 'The Thinking Proletariat' (1865) Pisarev makes the similar argument that selfishness is useful if it is allowed to find a productive outlet, and that work is the best way of harmonizing the interests of all members of a society. Idleness and exploitation, on the other hand, lead man's natural selfishness to take destructive forms. In the Nihilist 'new men' individual qualities and the general interest are inseparable: 'goodness and truth, honesty and knowledge, character and reason are identical'.

Largely because of this faith in the natural state of harmony between the individual and society, the Nihilists tended to leave their social programmes vague. They were certainly committed to a form of secular socialist society with guaranteed rights for the individual, but their main contribution to social debate was their rejection of existing authority. The Nihilists were, however, unswerving in their commitment to social change through the spread of enlightenment and in particular the natural sciences. They believed that if true 'understanding' was brought to people through the dissemination of knowledge, social progress would inevitably follow. Pisarev laid particular emphasis on the role of education in social change. In his essay 'The Realists' (1864) he argues that science must find its way into the everyday life of the common people if they are to break out of the vicious circle of poverty and ignorance. With a characteristically Russian faith in the power of ideas, he asserts that the wise investment of 'intellectual capital' can 'diminish the quantity of human pains'.

See also: Nihilism; Positivism, Russian; Russian Materialism: 'the 1860s'

STEPHEN LOVELL

List of works

Chernyshevskii, N.G. (1939-53) *Polnoe sobranie sochinenii (Collected Works)*, Moscow: Khudozhestvennaia literatura, 16 vols.(The standard Soviet edition of Chernyshevskii's works.)

Chernyshevskii, N.G. (1863) *Chto delat'*, St Petersburg; trans. B.R. Tucker, *What Is To Be Done?*, New York: Vintage Books, 1961.(An English translation of Chernyshevskii's famous and highly influential novel.)

Chernyshevskii, N.G. (1953) *Selected Philosophical Essays*, Moscow: Foreign Language Publishing House.(A useful selection of Chernyshevskii's main writings translated into English.)

Dobroliubov, N.A. (1956) *Selected Philosophical Essays*, Moscow: Foreign Language Publishing House. (A selection of essays by Chernyshevskii's friend and disciple.)

Pisarev, D.I. (1958) *Selected Philosophical, Social and Political Essays*, Moscow: Foreign Language Publishing House.(A useful selection of Pisarev's writings which includes all the essays mentioned in this entry.)

Pisarev, D.I. (1955) *Sochineniia v chetyrëkh tomakh (Works in Four Volumes)*, Moscow: Khudozhestvennaia literatura.(The standard Soviet edition of Pisarev's works.)

References and further reading

Coquart, A. (1946) *Dmitri Pisarev (1840-68) et l'idéologie du nihilisme russe (Dmitri Pisarev (1840-68) and the Ideology of Russian Nihilism)*, Paris.(A good, detailed account of Pisarev's thought and its place in the social and philosophical debate of the 1860s; has a useful bibliography.)

Edie, J.M. *et al.* (eds) (1965) *Russian Philosophy*, Chicago, IL: Quadrangle Books, 3 vols.(A very useful edited selection of Russian philosophy; vol. 2 contains extracts from important works by Chernyshevskii and Pisarev, along with short introductions.)

Masaryk, T.G. (1919) *The Spirit of Russia*, London: Allen & Unwin, 2 vols.(An old but excellent account of Russian philosophy, with a substantial chapter on Nihilism in vol. 2.)

Moser, C. (1964) *Antinihilism in the Russian Novel of the 1860's*, The Hague: Mouton.(Although the main focus of this book is anti-Nihilism, it has two excellent introductory chapters on the history and background of Nihilism itself.)

Paperno, I. (1988) *Chernyshevsky and the Age of Realism: A Study in the Semiotics of Behavior*, Stanford, CA: Stanford University Press.(An important study of the cultural significance of Chernyshevskii's *What Is To Be Done?*)

Stites, R. (1991) *The Women's Liberation Movement in Russia: Feminism, Nihilism, and Bolshevism 1860-1930*, Princeton, NJ: Princeton University Press, 2nd edn.(This fine social and intellectual history of the women's

movement has an interesting chapter on the Nihilist contribution to the cause.)
Turgenev, I. (1862) *Ottsy i deti*; trans. M.R. Katz, *Fathers and Sons*, New York: W.W. Norton, 1994.(Referred to in §1.)

Nirvāṇa

The aim of the spiritual life was already described as nirvāṇa *before the rise of Buddhism around the fifth century BC, but it is in the Buddhist context that it is most well known. In earlier Buddhist works and in popular usage to the present day it refers to the goal of Buddhist discipline, reached by systematic training in morality, meditation and intellect. That goal consists of the final removal of the disturbing mental elements which obstruct a peaceful and clear state of mind, together with a state of awakening from the mental sleep which they induce. Such an awakening (often referred to in English as 'enlightenment') enables a clear perception of fundamental truths, the understanding of which is essential to mental freedom. Later, the term was often applied more narrowly to a specific aspect of the awakened condition - that aspect of this experience which was considered to be unchanging, that is, an element which is not the product of either mental construction in particular, or of causes and conditions in general.*

1 Origins and etymology of the word *nirvāṇa*

Nirvāṇa (or *nibbāna* in the Pāli language) means literally 'blowing out' or 'quenching'. However, since the term is probably pre-Buddhist, its etymology is not necessarily conclusive for determining its exact meaning as the highest goal of early Buddhism. In fact, many Buddhists have traditionally preferred to explain it as absence of the weaving (*vāna*) activity. Here weaving is a metaphorical description of the mind's manner of operation when distorted by ignorance of the true nature of things and by craving for possessions and states of being. It is probably not accidental that such a definition does not tell us what exactly *nirvāṇa* is. Buddhists have generally avoided any verbal description of their goal and have tended to doubt whether it is possible to describe it.

The earliest Buddhist texts seem to use the word *nirvāṇa* to designate the final end (*accantaniṭṭhā*) of the path taught by the Buddha. As such it is by definition supreme happiness and ultimate peace. Probably they shared this with other religious-philosophical schools of the period. The same texts generally go to great lengths to avoid discussing what happens to the attainer of this goal after death. The Buddha is presented as 'not having declared' that such a one exists after death or does not exist then. Nor did he declare that both of these were the case. Even a denial of both possibilities was ruled out as an option. Indeed, concern with the nature of that goal of happiness and ultimate peace is presented as unhelpful to the task of achieving it. The simile of the man injured by a poisoned arrow is commonly cited in this context: in the immediate situation he should focus upon obtaining medical treatment rather than be concerned with specific details of the arrow, poison or archer. Similarly the urgency of attaining *nirvāṇa* is such that it is better not to be overconcerned with details of its nature and consequences.

From the origins of the word itself it is possible to assess the earlier position which is being reacted to here. *Nirvāṇa* derives from the root 'to blow' and means the blowing out or quenching of a lamp flame or a fire. In one use it is the fires of passions, such as greed and hate, which are put out by the enlightened sage. In another usage, applicable to the time of death, the flame of the lamp covers all types of ordinary human activity. These cease at the death of the enlightened saint, like the flame of a lamp. In pre-Buddhist usage this would probably have been understood as meaning that the fire had returned to the latent or potential state from which it arose in the first place. Contrary to its modern ascription, the metaphor would not have had connotations of annihilation. The Buddha himself seems to have resisted this kind of understanding, perhaps because it was too closely associated with the pantheistic self of the Upaniṣads, or the life-monad of the Jains and others (see Jaina philosophy). Rather, the Buddhist texts tend to suggest that language is inapplicable to the case since words are in the last analysis derived from sensory experience, which is taken to include memory and imagination.

2 Earlier accounts

Most mentions of *nirvāṇa* refer to it as the goal, or as supreme happiness with many synonyms which make clear that it is conceived as the ultimate security, purity and peace. A few of the early texts, mostly preserved in Pāli or Chinese translations, provide a little more information. Someone who experiences *nirvāṇa* is explicitly declared to be conscious, but not conscious of 'anything which the mind has seen, heard, sensed, felt, obtained, sought or explored' (*Anguttara Nikāya (The Book of Gradual Sayings)* IV 320; 353-8). This experience is differentiated from various kinds of meditative attainment.

Later a clear distinction between *nirvāṇa* attained in life and *nirvāṇa* entered into at the point of death begins to develop. The term *parinirvāṇa* is often used for the latter although it does not acquire this meaning until a later date - in the earliest sources it is simply an alternative to *nirvāṇa*. (Strictly, *nirvāṇa* is the state of release; *parinirvāṇa* is the act of attaining release.) However, relatively early sources (although not the very oldest) do contain the important simile of the ocean which shows no sign of filling or diminishing no matter how much water is poured into it by streams or rainfall. Similarly, even if many enter *parinirvāṇa* by means of the element of *nirvāṇa* without any remaining 'clinging', the element of *nirvāṇa* shows no sign of filling or diminishing. Here 'clinging' is a technical term referring to the appropriation or identification which occurs as a result of desire. Since ordinary human existence and mental process are understood as the result of such appropriation, there can be no continuation in the familiar conditions of existence once it has ceased. The point of the simile is no doubt to emphasize that *nirvāṇa* transcends normal rules.

3 European understanding of the term

The *Oxford English Dictionary* (first edition) defines *nirvāṇa* thus: 'In Buddhist theology, the extinction of individual existence and absorption into the supreme spirit, or the extinction of all desires and passions and attainment of perfect beatitude'. Although versions of this definition have been widely used, it confuses the Buddhist with Jain and Hindu understandings of the term (see Karma and rebirth, Indian conceptions of). Such descriptions of *nirvāṇa* are rather old in European thought. By the seventeenth century the French ambassador to the Siamese court, Simon de La Loubère referred to Portuguese authors as having translated a closely related Siamese term as 'it is annihilated' and 'it has become a God'. He rightly indicated that in the opinion of the Siamese: 'this is not a real Annihilation, nor an Acquisition of any divine Nature'. In fact, La Loubère translated *nirvāṇa* as 'disappearance', meaning disappearance from the world of birth and death. This is a very accurate understanding of the term and cannot be bettered today.

However, little attention was paid to La Loubère's more reliable account of the term, and the views he corrected remained current. Part of the reason for this was the tendency, especially in the seventeenth and eighteenth centuries, but still current in some circles, to understand Buddhism as a form of paganism and then to apply to it the critique which the Church Fathers directed against pagan philosophy. While there are some similarities between the two areas, such as the belief in reincarnation characteristic of some forms of classical philosophy, overall this is inappropriate.

4 Schools of mainstream Buddhism

By the time of the development of the first forms of organized Buddhist exegesis in the third or fourth centuries BC slightly more systematic ideas were found (see Buddhist philosophy, Indian). *Nirvāṇa* as the element 'without any remaining clinging' was contrasted with the element 'with remaining clinging' (*saupādisesa*). This latter perception of *nirvāṇa* connoted the stages of partial enlightenment known as 'entering the stream'. As earlier texts most often referred to *nirvāṇa* 'without any remaining clinging' in the context of the Buddha's death, it began to be assumed that if what was attained at the death of the saint was the element of *nirvāṇa* "without any remaining clinging', then there must be a corresponding element which was achieved at enlightenment. Henceforth, the element of *nirvāṇa* 'with remaining clinging' regularly referred to the achievement of enlightenment, normally envisaged as followed by a life of compassionate teaching activity. As a result of a later mistranslation or misunderstanding, the word translated as 'clinging' is often replaced by a word which may refer to 'possessions' (*saupadhiśeṣa*) and taken as referring to the mind and body as possessions appropriated during the process of rebirth. Alternatively, both *saupādisesa* and *saupadhiśeṣa* could derive from **upādhisesa* and **upātisesa* respectively, meaning 'small remainder'. Thus, originally the meaning of the terms connoting 'clinging' could have been a reference to the small number of defilements remaining after 'entering the stream'.

Older sources are reticent on the subject of *nirvāṇa*. However, the same is not quite true of the canonical *abhidhamma* literature which took form between the second and fourth centuries BC. A characteristic feature of this literature is the analysis of mind and body in terms of a changing process of interacting and multiple events. These were referred to as *dhammas*, that is, particular truths as opposed to truth in general. (Earlier the word was used for all objects of mental awareness other than sense objects themselves.) The *abhidhamma* texts compiled open-ended lists of these events and tried to analyse which events occurred in which mental states. In this analysis they

included *nirvāṇa* which they explained as differing from all other events in that it was not categorizable in spatial or temporal terms. At this stage there does not seem to have been any controversy as to its fundamental nature, nor was it conceived as mere absence: '*Nibbāna* exists; it does not change its nature. *Nibbāna* is permanent, constant, eternal, not subject to change' (*Kathāvatthu (Points of Controversy)* 122).

It is probable that the main concern of the *abhidhamma* texts was with the experience of *nirvāṇa* in life in terms of awakening or enlightenment. In that context *nirvāṇa* would be understood as the object or support for the enlightened mind, itself a composite characterized by the repeated presence of such events as faith and wisdom. Like the texts which preceded them, the works of this period did not attempt to comment upon the fate of the enlightened saint after death. No doubt the tradition of the undeclared questions was still too firmly established to allow this.

5 In the Mahāyāna systems

New tendencies in Indian religion, developing over some time, led to the formation of a new kind of Buddhism in the period AD 100-300: the Mahāyāna or 'great vehicle'. Attracting only minor interest in the beginning, the new movement gradually became increasingly influential. Devotional elements were strong in this school which arose essentially from an elaboration of the idealized biography of the Buddha and an extension of that into an ideal for all. An important feature was the prolongation of the path to enlightenment in order to enable the development of the superior qualities seen as necessary to achieve Buddhahood and work more effectively for the benefit of others. A side effect of this shift in focus was that *nirvāṇa* receded somewhat into the background, although in some subsequent schools of Mahāyāna Buddhism, such as Chan/Zen or Tantra, it was ultimately to return to the foreground as something attainable in this life.

New schools of thought developed within the overall Mahāyāna with particular perspectives. Probably the oldest of these was the Madhyamaka (see Buddhism, Mādhyamika: India and Tibet). This laid stress on the idea that there was no difference between *nirvāṇa* and existence in the round of rebirths (*saṃsāra*). They seem to have meant by this that *nirvāṇa* is simply a name for ordinary existence with ignorance removed. In other words it is not to be viewed as an entity existing somewhere else. It is rather a transformation of normal experience.

Some later Mahāyāna writings placed an emphasis on '*nirvāṇa* without foundation' as part of the path to the highest goal. This means that true enlightenment involves neither acceptance nor rejection of birth and death. More remarkably, it implies neither accepting nor rejecting *nirvāṇa* itself, but instead encouraging minds to be free of dependency upon or need for any objects at all, whether those objects are internal or external.

6 Synonyms for *nirvāṇa*

A common error in examining a concept such as *nirvāṇa* is to focus too much on the exact denotation of the term at the expense of its wider associations and context, not taking into account the number of synonyms frequently used to describe it. In a practical, experiential context it is referred to as 'freedom' or 'awakening'; 'peace' or 'bliss'. In each case the implication is that it is this in the ultimate possible degree. Words which imply safety or complete protection are commonly used in relation to it, as are words which imply permanence, unchangingness or unendingness. These latter denotations are close to the general terms used in Indian religions to describe the highest goal of the spiritual life. Many such words are regularly used to describe *nirvāṇa*. So it is liberation (*mokkha*/*mokṣa*), wholeness (*kevala*/*kaivalya*), truth and the like; it is the deathless and the imperishable. Many of these terms are given a specific Buddhist sense, but their general connotation in the world at large should not be neglected. A specific example might be that *nirvāṇa* is the *amṛta*, or the deathless, but it is important that this refers to the nectar which confers immortality upon the gods. In the Buddhist context it refers to a condition in which there is no death, although it is clearly intended to have the positive associations of Indian myth.

See also: Buddhism, Ābhidharmika schools of; Buddhist concept of emptiness; Cosmology and cosmogony, Indian theories of; God, Indian conceptions of; Heaven, Indian conceptions of; Momentariness, Buddhist doctrine of; Tibetan philosophy

L.S. COUSINS

References and further reading

Hare, E.M. (trans.) (1955) *Anguttara Nikāya, or More Numbered Suttas (The Book of Gradual Sayings)*, vol. 4, *The Books of the Sevens, Eights and Nines*, London: Pāli Text Society, 1955.(One of the major Pāli canonical works.)

Harvey, P. (1995) *The Selfless Mind: Personality, Consciousness and Nirvāna in Early Buddhism*, Richmond: Curzon Press.(Analyses most of the early sources.)

Norman, K.R. (1991) ‘Death and the Tathāgata’, *Studies in Buddhism and Culture* (in honour of Professor Egaku Mayeda), Tokyo: 1-11.(An essay that appears in a volume in honour of Mayeda.)

Norman, K.R. (1994) ‘Mistaken Ideas about *Nibbāna*’, in T. Skorupski and U. Pagel (eds) *The Buddhist Forum*, London: School of Oriental and African Studies, vol. 3: 211-25.(Two technical discussions by a leading specialist.)

Rahula, W. (1967) *What the Buddha Taught*, Bedford: Gordon Fraser, 2nd edn, 1959.(An influential account of the basic issues by a Sinhalese Buddhist monk.)

Saddhātissa, H. (1970/1) *Buddhist Ethics*, London, New York: Allen & Unwin/Braziller.(A Sinhalese Buddhist monk taking a slightly more traditionalist approach. Chapter 8 is especially interesting.)

Slater, R.H.L. (1951) *Paradox and Nirvāṇa*, Chicago, IL: University of Chicago Press.(An older work which deserves to be better known.)

Tissa Moggalīputta (*c.*246 BC) *Kathāvatthu (Points of Controversy)*, trans. Shwe Zan Aung and C.A.F. Rhys Davids, London: The Pāli Text Society, 1979.(The standard English translation of the earliest collection of points disputed by Buddhists during the first two centuries after the Buddha’s career, part of the *abhidhamma* section of the Pāli canon.)

Vajirañāṇa, P.M. (1962) *Buddhist Meditation in Theory and Practice: A General Exposition According to the Pāli Canon of the Theravāda School*, Colombo: M.D. Gunasena & Co.(Includes a convenient list of synonyms, some of which are discussed in §6. Pages 469-80 are especially interesting.)

Welbon, G.R. (1968) *The Buddhist Nirvāṇa and Its Western Interpreters*, Chicago, IL: University of Chicago Press.(Expansion of the material in §3.)

Williams, P. (1989) *Mahāyāna Buddhism: The Doctrinal Foundations*, London, New York: Routledge, esp. 52-4; 67-9; 181-4.(Presents the fundamental ideas of the Mahāyāna, as discussed in §4.)

Nishi Amane (1829-97)

Among the campaigners for Japanese enlightenment in the early Meiji era, Nishi Amane was prominent for his philosophical achievements. He introduced European philosophy into Japan, especially the positivism of Auguste Comte and the utilitarianism of John Stuart Mill. The academic philosophy of modern Japan owes its origins to him.

The son of a court physician of the Tsuwano fief, Nishi was ordered by his lord to learn neo-Confucianism. Quite unsatisfied by this old-fashioned doctrine, he became aware of the necessity for studying Western science and fled the fief. In 1857 he was employed at the Foreign Books Research Institute by the Tokugawa government, and in 1862 he was sent by the government to Leiden in Holland on a mission to study the European social sciences. He studied law, statistics and economics formally under S. Vissering, but he also attended lectures by C.W. Opzoomer, then the leading historian of philosophy in Holland. Opzoomer introduced him to the works of Auguste Comte and J.S. Mill.

Having returned home in 1865, Nishi continued to work for the Tokugawa government and, after the Restoration, was appointed to the Military Department of the Meiji government in 1870. He contributed much to the establishment of the modern Japanese military during his bureaucratic career, while as a private individual he introduced European philosophy and social sciences through his writing and teaching. From 1870 onwards he ran a private school called the Ikueisha. In 1873 he took part in the inauguration of the Meirokusha (the Meiji Six Society), an active scholarly society, and promoted the campaign for the enlightenment of the Japanese people.

Nishi's philosophical writings were motivated by criticism of Confucianism and praise for Western scientific thinking. In *Hyakuichi-shinron (New Theory of the Hundred and One)*, published in 1874, he stressed the distinctions between morality and law and between human and natural principles. Embracing a dualistic attitude, he declared himself against the monism of the Confucians, especially the neo-Confucian world view. In *Jinsei-sanposetsu (Theory of the Three Human Treasures)*, he enumerated health, knowledge and wealth as the *sine qua non* for happiness. He opposed these to the feudalistic values of docility, naivety, humility, deference, unselfishness and freedom from avarice. Thus he replaced the group-oriented morality with a utilitarian-individualistic one. In *Hyakugaku-renkan (Links of All Sciences)*, which consists of drafts of his lectures at Ikueisha, he showed the classification and harmony of all disciplines, assigning the pivotal position to philosophy. He also explained Comte's theory of three stages and Mill's method of induction. Moreover, he published the first Japanese textbook on formal logic, *Chichi-keimo (Logic and Enlightenment)* in 1874, and a translation of Mill's *Utilitarianism* in 1877. As Thomas R.H. Havens states:

> Because Nishi's contributions… were the transmission of European ideologies and the invention of new academic terminology, rather than original thought systems of his own, Nishi did not found any new school of Japanese philosophy. However, his influence is to be found in the works of…various important philosophers in Meiji era.
>
> (Havens 1970: 218-9)

See also: Japanese philosophy; Logic in Japan

HIMI KIYOSHI

List of works

Nishi Amane (*c.*1866-89) *Nishi Amane zenshū(The Complete Works of Nishi Amane)*, ed. T. Okubo, Tokyo: Munetaka Shobo, 1960-6, 3 vols.(Standard edition of Nishi's collected works.)

Nishi Amane (1874) *Hyakuichi-shinron (New Theory of the Hundred and One)*, in *Nishi Amane zenshū*, ed. T. Okubo, Tokyo: Munetaka Shobo, 1960-6.(Stresses the distinctions between morality and law and between human and natural principles.)

Nishi Amane (1876) *Jinsei-sanposetsu (Theory of the Three Human Treasures)*, in *Nishi Amane zenshū*, ed. T. Okubo, Tokyo: Munetaka Shobo, 1960-6.(Discusses the prerequisites for human happiness.)

Nishi Amane (1877) *Hyakugaku-renkan (Links of All Sciences)*, in *Nishi Amane zenshū*, ed. T. Okubo, Tokyo: Munetaka Shobo, 1960-6.(Drafts of lectures in which Nishi assigns positions to all the disciplines.)

Nishi Amane (1874) *Chichi-keimo (Logic and Enlightenment)*, in *Nishi Amane zenshū*, ed. T. Okubo, Tokyo: Munetaka Shobo, 1960-6.(The first Japanese textbook on formal logic.)

References and further reading

Havens, T.R.H. (1970) *Nishi Amane and Modern Japanese Thought*, Princeton, NJ: Princeton University Press.(A detailed argument about Nishi's contribution to the Japanese revolution of ideas in the latter half of the nineteenth century.)

Miyagawa, T. (1962) *Kindai Nihon no Tetsugaku (Japanese Modern Philosophy)*, Tokyo: Keisoshobo.(An explanation of the role played by Nishi at the beginning of modern Japanese philosophy; see especially pages 17-47.)

Nishida Kitarō (1870-1945)

Considered Japan's first original modern philosopher, Nishida not only transmitted Western philosophical problems to his contemporaries but also used Buddhist philosophy and his own methods to subvert the basis of traditional dichotomies and propose novel integrations. His developmental philosophy began with the notion of unitary or pure experience before the split between subject and object. It developed to challenge other traditional opposites such as intuition and reflection, fact and value, art and morality, individual and universal, and relative and absolute. In its organic development, Nishida's philosophy reacted to critiques that it neglected the social dimension with political essays that sometimes aligned it with Japanese imperialism. It culminated in the 'logic of place', a form of thinking that would do justice to the contradictory world of human actions.

While a teacher in his twenties and thirties, Nishida fervently practiced Zen Buddhism under masters of the Rinzai lineage. After the publication in 1911 of his groundbreaking work *Zen no kenkyū (An Inquiry Into the Good)*, he held professorships at Kyoto Imperial University until his retirement in 1929. Nishida produced numerous essays, developing a logic of consciousness and, from the 1930s, a philosophy of history and culture. In the 1940s, government leaders called upon the prestigious philosopher to write political tracts in support of Japanese expansionism. These writings were too abstract for politicians and military leaders, but were attacked as pro-Western by right-wing philosophers of the day, and later as imperialist by leftist critics. Because of his interest in Zen, his adaptation of Mahāyāna Buddhist logic and his influence on younger Kyoto colleagues such as Tanabe Hajime and students such as Nishitani Keiji, Nishida became known as the father of the 'Kyoto School' that dominated Japanese philosophy of religion up to the 1990s and informed much Buddhist-Christian dialogue (see Kyoto School). It is possible, however, to understand 'Nishida philosophy' in a different light, as a systematic deconstruction of logical relations.

Four methods are evident in the development of Nishida's philosophy and help to clarify it. First is his way of laying a foundation; then, undermining or undergirding that foundation with something more basic, and ultimately, undermining any positive determination of a foundation. This method has similarities with that of Nāgārjuna, except that Nishida works on broad foundations in several major treatises instead of particular propositions in a single analysis. Second, there is his way of subverting anthropocentric assumptions about the nature of knowledge and reality. Third, there is his method of exhibiting the self-reflective or self-mirroring structure of many logical and metaphysical categories. Fourth, there is his reversal of the order of things in many traditional hierarchies or explanatory schemes.

The preface to *Zen no kenkyū* expresses Nishida's initial project to explain everything on the basis of pure experience as the single reality. Pure or immediate experience is the moment or act of experiencing before there is an awareness of any difference between an object that is experienced and a subject who does the experiencing, before there is any thought, reflection or meaning. This singular and unitary foundation develops to include derivative and differentiated moments such as thought and reflection that seek a higher unity in reality. Reality objectified and unified as a whole is what is meant by nature, but as the activity of unifying it is spirit. Nishida undermines subjectivist views of experience by claiming that it precedes and founds the individual human subject; he subverts appeals to an ultimate difference between reality and consciousness, and he distances himself from anthropocentric views of consciousness by using the word 'God' for the infinite activity of unification. Unlike many commentators who see the notion of pure experience as the basis of all of Nishida's philosophy, however, he himself soon abandoned this notion, perhaps because of its psychological connotations, or perhaps because of its inherent logical difficulty in accounting for a meaningful world that arises from a meaning-less foundation.

Jikaku ni okeru chokkan to hansei (Intuition and Reflection in Self-Consciousness), written serially from 1913 to 1917, at first continues the project to explain immediate experience as a dynamic system that develops into reflective thought and judgment. Nishida follows neo-Kantians in undercutting the traditional fact-value distinction; factual judgments intend truth that serves as a value or norm (see Neo-Kantianism). He appeals to the notion of intellectual intuition as a way to reconcile immediate experience with reflection, which coincide in the case of intuited self-awareness. By the end of this work he has shifted the foundation to what he considers the primary instance of self-awareness, namely absolute free will, because its activities of creating and valuing are

more basic than intellectual knowing (see Free will). In *Ishiki no mondai (Problems of Consciousness)*, Nishida reverses the usual primacy not only of knowing over valuing, but also of feeling over willing and sensing over feeling, by proposing that the latter term of each pair is the more concrete reality. *Geijutsu to dōtoku (Art and Morality)* names the absolute will as the source of moral decision making and creative artistic activity. In later works, however, Nishida eventually abandons the notion of absolute will as a basis, because the activity of the will eludes reflection and thus would be an unknowable foundation.

Essays in *Hataraku mono kara miru mono e (From That Which Acts to That Which Sees)* and *Ippansha no jikakuteki taikei (The Self-Conscious System of Universals)* begin to develop Nishida's 'logic of place' (*basho*). His notion of a universal (see Universals) as a place may be interpreted as a founding level of discourse that resolves contradictions on less comprehensive levels. Ultimately, the place that holds differences or contradictions together is absolute nothingness, the place beyond discursive determination. This notion undermines previous attempts to locate a foundation with a positive description such as pure experience, self-awareness, or absolute will. Yet the self-reflective structure of self-awareness, in which a whole is mirrored or reflected in a part of itself, is still evident in universals of all levels, including the ultimate nothingness. We may read 'self-reflective' for the term 'self-conscious' (*jikakuteki*) in the above works, or in *Mu no jikakuteki gentei (The Self-Conscious Determination of Nothingness)*, published in 1932.

In *Tetsugaku no kompon mondai (The Fundamental Problems of Philosophy)* and seven volumes of *Tetsugaku ronbunshū (Philosophical Essays)*, Nishida shifts from the scheme of place within place (in other words, less and more inclusive contexts) to the notions of 'absolutely contradictory self-identity' and 'the world as a dialectical universal'. The former notion lets opposites form unities through continual self-negation, for example, that of the world or the self, each a contradictory unity of many and one, and a 'continuity of absolute discontinuities'. The notion of the dialectical universal implies that the world is not another entity but rather a place of mediation between individuals that co-arises and is codetermined with them. The exemplary case of individuality is not a material (and abstract) object like an atom, but rather the concrete human self, the focal point where the world reflects itself in itself. The self is not a subject that can be defined by adding specific differences to a genus (see Aristotle §§7-8), but rather something that forms itself through continual self-negation of its previous identities, in other words, through 'acting intuition', acting arising from direct seeing. Self and world both find their immediate 'ground' in absolute nothingness, for nothing outside them can serve as their basis and yet they do not serve as a foundation for themselves.

Partially in tension with this individualism is the theme that the world in its most concrete manifestation is a historical world determined by the interaction of different selves. Nishida's turn towards a philosophy of history from 1931 onwards culminated in a political and cultural philosophy, pronounced in the early 1940s, that envisaged a new, truly global world of nation-states (see Nation and nationalism). Each has its own principle that forms it in difference from and in interaction with all others. The state is the source of moral values. Nishida's own warning that Japan should not become imperialist is belied in his view that Japan is the nation-state that can awaken other nations, particularly Asian ones, to their truly global identities. The historical position of the Emperor is the empty centre of the body of the nation (*kokutai*) and the place from which emerge the creative possibilities of Japan as a nation and of being Japanese. In contrast to a view of history that sees present circumstances determined by past events and decisions oriented toward the future, Nishida sees history forming out of the 'absolute present' in which past and future are simultaneously co-present, each negating the other to allow infinite possibilities.

Nishida's last writings continue this nationalism but also make more explicit the religious themes latent throughout. In particular, *Bashoteki ronri to shūkyōteki sekaikan (The Logic of Place and the Religious Worldview)*, published in 1945, addresses personal concerns of religious awareness and the death of the individual self. There the basic Other to oneself is not another self, nor the world as dialectical universal, nor even absolute nothingness, but rather the absolute or God (see Absolute, the; God, concepts of). The relation between the individual self and God is a kind of contradictory self-identity, more properly called an 'inverse correspondence' (*gyakutaiō*). The self properly relates to God only by dying to itself, to ego (a death that is not the same thing as the nihilation of the person), and so becomes its true self. For its part, the absolute arises through its self-negation in the relative self; it does not exclude the relative, nor even evil. Although God 'is absolute being only because he is absolute nothingness', this essay's implicit accommodations to Christian theism and Pure Land

Buddhist faith seem at odds with Nishida's earlier anti-foundation of absolute nothingness.

See also: Japanese philosophy; Logic in Japan; Perception, epistemic issues in; Political philosophy, history of

JOHN C. MARALDO

List of works

Nishida Kitarō (1911-45) *Nishida Kitarō zenshū (Complete Works of Nishida Kitarō)*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn.(Complete works of Nishida Kitarō in nineteen volumes.)

Nishida Kitarō (1911) *Zen no kenkyū (An Inquiry into the Good)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 1; trans. Masao Abe and C. Ives, *An Inquiry Into the Good*, New Haven, CN: Yale University Press, 1990.(Nishida's first major work on consciousness.)

Nishida Kitarō (1915) *Shisaku to taiken (Contemplation and Experience)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 1.(On consciousness and perception.)

Nishida Kitarō (1917) *Jikaku ni okeru chokkan to hansei (Intuition and Reflection in Self-Consciousness)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 2; trans. V. Viglielmo with Y. Takeuchi and J. O'Leary, Albany, NY: State University of New York Press, 1987.(Appeals to the notion of intellectual intuition in a further attempt to explain the relations of thought and experience.)

Nishida Kitarō (1920) *Ishiki no mondai (The Problem of Consciousness)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 3.(Proposes that valuing, willing and feeling are more concrete than knowing and sensing.)

Nishida Kitarō (1923) *Geijutsu to dōtoku (Art and Morality)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 3; trans. D. Dilworth and V. Viglielmo, Honolulu, HI: University Press of Hawaii, 1973. (Names the absolute will as the source of moral decision making and creative activity.)

Nishida Kitarō (1927) *Hataraku mono kara miru mono e (From That Which Acts to That Which Sees)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 4.(Early development of Nishida's logic of place.)

Nishida Kitarō (1930) *Ippansha no jikakuteki taikei (The Self-Conscious System of Universals)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 5; 'Sōsetsu' (General Summary) trans. in R. Wargo, 'The Logic of Basho and the Concept of Nothingness in the Philosophy of Nishida Kitarō', Ann Arbor, MI: University Microfilms, 1972; 'Eichiteki sekai' (The Intelligible World) trans. in R. Schinzinger, *Intelligibility and the Philosophy of Nothingness*, Westport, CN: Greenwood Press, 1973.(The translations are of two central essays of the volume.)

Nishida Kitarō (1932) *Mu no jikakuteki gentei (The Self-Conscious Determination of Nothingness)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 6.(On self-reflection and self-awareness.)

Nishida Kitarō (1933-4) *Tetsugaku no kompon mondai (Fundamental Problems of Philosophy)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 7; trans. D. Dilworth, *Fundamental Problems of Philosophy*, Tokyo: Sophia University, 1970.(Later work on universals and self-identity.)

Nishida Kitarō (1935-46) *Tetsugaku ronbunshū I-VII (Philosophical Essays I-VII)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vols 8-11; 'Sekai no jikodōitsu to renzoku' and 'Rekishiteki keiseisayōtoshite no geijutsuteki sōsaku', trans. E. Weinmayr as 'Selbstidentität und Kontinuität der Welt' and 'Das künstlerische Schaffen als Gestaltungsakt der Geschichte', in *Die Philosophie der Kyōto-Schule*, ed. R. Ohashi, Munich: Karl Alber, 1990; 'Kokka riyū no mondai', trans. A. Jacinto Zavala as 'El problema de la razón de estado', in Nishida Kitarō, *Estado y Filosophía*, Zamora: El Colegio de Michoacan, 1985; 'Bashoteki ronri to shūkyōteki sekaikan', trans. Michiko Yusa as 'The Logic of *Topos* and the Religious Worldview', *The Eastern Buddhist* 19 (2):1-29 and 20 (1): 81-119, 1986-7.(Further work on universals and the self.)

Nishida Kitarō (1937) *Zoku shisaku to taiken (Contemplation and Experience, Continued)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 12.(Revisiting earlier work on perception.)

Nishida Kitarō (1938) *Zoku shisaku to taiken igo (Sequel to Contemplation and Experience, Continued)*, in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 12.(Follows on from *Zoku shisaku to taiken*.)

Nishida Kitarō (1938) 'Nihon bunka no mondai' (The Problem of Japanese Culture) in *Nishida Kitarō zenshū*, Tokyo: Iwanami Shoten, 1987-89, 3rd edn, vol. 12; trans. A. Jacinto Zavala, 'La crisis de la cultura Japonesa', in Nishida Kitarō, *Estado y Filosophía*, Zamora: El Colegio de Michoacan, 1985; trans. P. Lavelle, *La Culture Japonaise en Question*, Paris: Publications Orientalistes de France, 1991.(Essay on culture.)

References and further reading

Abe Masao (1988) ‘Nishida’s Philosophy of Place’, *International Philosophical Quarterly* 28 (4): 355-71.(A clear exposition from a Zen philosopher’s point of view. Includes a comprehensive bibliography of translations and works on Nishida.)

Arisaka, Y. (1996) ‘The Nishida Enigma: "The Principle of the New World Order"’, *Monumenta Nipponica* 51 (1): 81-105.(A clarification of the background of Nishida’s political views and translation of his 1943 essay on the new world order.)

Dilworth, D. (1973) ‘Nishida Kitarō: Nothingness as the Negative Space of Experiential Immediacy’, *International Philosophical Quarterly* 13 (4): 463-84.(An advanced exposition of Nishida’s philosophy of nothingness.)

Dilworth, D. (ed.) (1987) *Last Writings: Nothingness and the Religious Worldview*, Honolulu, HI: University of Hawaii Press.(Comparative essays on Nishida and a free translation of ‘Bashoteki ronri to shūkyōteki sekaikan’ and a fragment, ‘Concerning My Logic’.)

Feenberg, A. and Arisaka, Y. (1990) ‘Experiential Ontology: The Origins of the Nishida Philosophy in the Doctrine of Pure Experience’, *International Philosophical Quarterly* 30 (2): 173-205.(A clear account of the main ideas of Nishida’s philosophy in light of his first work.)

Heisig, J. and Maraldo, J. (ed.) (1994) *Rude Awakenings: Zen, the Kyoto School and the Question of Nationalism*, Honolulu, HI: University of Hawaii Press.(The articles by Ueda, Yusa, Jacinto Zavala and Feenberg in particular examine the role of politics in Nishida’s philosophy.)

Mafli, P. (1996) *Nishida Kitarō‘s Denkweg (Nishida Kitarō‘s Way of Thought)*, Munich: Icidium.(A systematic survey in German of Nishida’s entire ‘way of thought’.)

Maraldo, J. (1989) ‘Translating Nishida’, *Philosophy East and West* 39 (4): 465-96. (Summaries of some of Nishida’s major works and terms, as well as a review of English translations.)

Nishitani Keiji (1991) *Nishida Kitarō*, trans. Yamamoto Seisaku and J. Heisig, Berkeley, CA: University of California Press.(Clearly translated essays on Nishida’s person and thought, by a renowned student and successor.)

Nishitani Keiji (1900-90)

Nishitani Keiji is generally regarded as the leading light of the 'second generation' Kyoto School of modern Japanese philosophy. Influenced by Zen thinkers from Chinese and Japanese Buddhism as well as by figures from the Western mystical and existential traditions, he is a pre-eminent voice in East-West comparative philosophy and late twentieth-century Buddhist-Christian dialogue. Primarily a philosopher of religion, Nishitani strove throughout his career to formulate existential responses to the problem of nihilism.

A student of the foremost figure in modern Japanese philosophy, Nishida Kitarō, Nishitani Keiji spent most of his life writing and teaching in Kyoto and became one of the most prominent members of the Kyoto School of philosophy (see Kyoto School). His earliest thinking developed under two different sets of influences. A classical education acquainted him with the Chinese Confucian and Buddhist traditions as well as with the philosophies of the Japanese schools of Zen Buddhism (see Buddhist philosophy, Japanese), where he found the thinking of Dōgen, Hakuin and Takuan most congenial. At the same time he read widely among Western authors (he mentions in particular Nietzsche and Dostoevskii, Emerson and Carlyle, as well as the Bible and St. Francis), and soon began to tackle such thinkers as Plotinus, Meister Eckhart, Schelling and Heidegger. By his own account, his engagement with philosophy was not the result of an intellectual decision but was precipitated by a severe existential crisis in his youth.

In 1936, Nishitani went to study philosophy in Germany and ended up in Freiburg, where he worked closely with Heidegger, returning to Japan on the eve of the Second World War. The next several years saw the publication of a number of texts on political themes (some of which were clearly written under external pressure). Although some internationalist elements in these writings provoked the ire of the right wing at the time, Nishitani was subsequently relieved of his teaching duties at Kyoto University for several years on the grounds of the inflammatory 'nationalist' tone of his wartime writings. While he has been reviled by Marxists in Japan and the USA for his 'rightwing' politics, Nishitani's published works of the period are mild compared to those of some other members of the Kyoto School; and among the critics, political innuendo has generally substituted for contextualized readings of the relevant texts.

Nishitani's long-standing engagement with the problem of nihilism was furthered in a series of talks subsequently published in 1949, the Japanese title of which was simply *Nihirizumu (Nihilism)*. In the framework of a synoptic account of nihilism as a historical phenomenon arising in nineteenth-century Europe, the author goes on to take it as an immediately existential problem. With the main focus on Nietzsche and a secondary concern with such thinkers as Max Stirner and Heidegger, he shows their common concern with the issue of what he calls 'the fundamental integration of creative nihilism and finitude'. By this is meant that when the self is plumbed to sufficient depth, and if one can endure the experience of the *nihilum*, or void, at its base, that 'nihility' may eventually be realized as a creative nothingness. Turning at the end of the book to nihilism in modern Japan, Nishitani argues that the Nietzschean project of 'the self-overcoming of nihilism' has close parallels to 'the standpoint of Buddhism, and in particular to the standpoint of emptiness in the Mahāyāna tradition' (see Buddhist concept of Emptiness; Buddhist philosophy, Japanese). On this view, the nothingness of the abyss in Nietzsche would be similar to the nothingness (*mu*) or fertile emptiness (*kū*) that is central to Japanese Buddhist thinking.

The text that is widely regarded as Nishitani's most important work, *Shūkyōto wa nani ka? (What Is Religion?)*, came out of a series of essays written in the late 1950s and was first published in book form in 1961. The emphasis is now very much on the Zen Buddhist tradition, with a secondary focus on Christianity and certain figures in Western philosophy (Meister Eckhart, Nietzsche, Heidegger). Like his teacher Nishida and several other senior members of the Kyoto School, Nishitani sees the Zen notion of 'absolute nothingness' as central to both philosophy and religion. True to the Mahāyāna Buddhist tradition, he stresses that absolute nothingness cannot be properly grasped by intellect alone but must rather be thought and experienced as the field or place from which all our thoughts and actions can flow. He thereby elaborates a radically non-anthropocentric worldview that allows us to realize ourselves as fully human beings-in-the-world, and from which the nothingness discussed by Nietzsche and Heidegger can be seen to 'fall short' of the Mahāyāna notion. However, to avoid entering a debate in competitive (rather than comparative) philosophy, let it simply be said in conclusion that the 'double genealogy' of

Nishitani's thinking - incorporating strains of Confucian, Daoist and Buddhist thinking on the one hand, and Neoplatonic, Christian mystical, and German idealist and existential philosophy on the other - gives rise to a uniquely formidable philosophy.

See also: Buddhist philosophy, Japanese; Dōgen; Japanese philosophy; Kyoto School; Nihilism

GRAHAM PARKES

List of works

Nishitani Keiji (1949) *Nihirizumu (Nihilism)*, trans. G. Parkes and Setsuko Aihara, *The Self-Overcoming of Nihilism*, Albany, NY: State University of New York Press, 1990. (The main focus is Nietzsche, with chapters on Stirner, Heidegger and Russian authors.)

Nishitani Keiji (1961) *Shūkyōto wa nani ka? (What is Religion?)*, trans. J. Van Bragt, *Religion and Nothingness*, Berkeley, CA: University of California Press, 1985.(The excellent translator's introduction includes a brief biographical account.)

Nishitani Keiji (1985) *Nishida Kitarō*, trans. Yamamoto Seisaku and J.W. Heisig, Berkeley, CA: University of California Press, 1991.(A collection of essays from several decades, beginning in the mid-1930s.)

References and further reading

In Memoriam Nishitani Keiji (1992), special edition of *The Eastern Buddhist* 25 (1).(A collection of essays and personal reminiscences conveying a lively sense of Nishitani's personality.)

Unno Taitetsu (ed.) (1989) *The Religious Philosophy of Nishitani Keiji: Encounter with Emptiness*, Berkeley, CA: Asian Humanities Press.(A collection of conference papers; of broad range but uneven quality.)

Waldenfels, H. (1980) *Absolute Nothingness: Foundations for a Buddhist-Christian Dialogue*, trans. J.W. Heisig, New York: Paulist Press.(A pioneering work, now out of print but still a good introduction.)

Nominalism

'Nominalism' refers to a reductionist approach to problems about the existence and nature of abstract entities; it thus stands opposed to Platonism and realism. Whereas the Platonist defends an ontological framework in which things like properties, kinds, relations, propositions, sets and states of affairs are taken to be primitive and irreducible, the nominalist denies the existence of abstract entities and typically seeks to show that discourse about abstract entities is analysable in terms of discourse about familiar concrete particulars.

In different periods, different issues have provided the focus for the debate between nominalists and Platonists. In the Middle Ages, the problem of universals was pivotal. Nominalists like Abelard and Ockham insisted that everything that exists is a particular. They argued that talk of universals is talk about certain linguistic expressions - those with generality of application - and they attempted to provide an account of the semantics of general terms rich enough to accommodate the view that universals are to be identified with them.

The classical empiricists followed medieval nominalists in being particularists, and they sought to identify the kinds of mental representations associated with general terms. Locke argued that these representations have a special content. He called them abstract ideas and claimed that they are formed by removing from ideas of particulars those features peculiar to the particulars in question. Berkeley and Hume, however, attacked Locke's doctrine of abstraction and insisted that the ideas corresponding to general terms are ideas whose content is fully determinate and particular, but which the mind uses as proxies for other particular ideas of the same sort.

A wider range of issues has dominated recent ontological discussion, and concern over the existence and status of things like sets, propositions, events and states of affairs has come to be every bit as significant as concern over universals. Furthermore, the nature of the debate has changed. While there are philosophers who endorse a nominalist approach to all abstract entities, a more typical brand of nominalism is that which recognizes the existence of sets and attempts to reduce talk about other kinds of abstract entities to talk about set-theoretical structures whose ultimate constituents are concrete particulars.

1 Introduction

In one use, 'nominalism' refers to a cluster of loosely-related philosophical and theological themes articulated by certain late fourteenth-century thinkers who were influenced by William of Ockham. These thinkers expressed doubts about the Aristotelian metaphysics, in particular its use in proving God's existence. They gave priority to faith over reason and emphasized the omnipotence of God in ways that often led to a Divine Command theory in ethics and a general scepticism about our knowledge of both causal relations and the substance-accident distinction.

Used this way, the term has its roots in the more common use which refers to a general theoretical orientation to questions about the existence and nature of abstract entities, an orientation exemplified in the work of Ockham himself. Those who are nominalists in this sense reject a Platonistic or realistic interpretation of discourse about things as diverse as properties, kinds, relations, propositions, sets, states of affairs and modality. Sometimes the nominalist is said to endorse the view that discourse of the sort in question is metalinguistic and that talk about the so-called abstract entities is really just talk about *nomina* or linguistic expressions. Characterized in this way, nominalism is sometimes said to stand opposed to conceptualism, another reductionist approach to ontological questions about abstract entities. The claim is that while the conceptualist insists that it is necessary to make reference to the activity of conceptual representation to accommodate the recalcitrant discourse, the nominalist denies this. But, first, not all of those who are called nominalists endorse a metalinguistic interpretation of the disputed discourse; and, second, since few philosophers have thought it possible to characterize language without referring to conceptual activity, to take nominalism as a view opposed to conceptualism yields the result that few of those thinkers normally taken to be nominalists turn out to deserve that label. Accordingly, it has become customary to construe as nominalistic any reductive approach to ontological questions about abstract entities that stands opposed to the Platonistic view.

2 The medieval period

The orientation we call nominalism is typically traced back to medieval debates over universals. A major source of

these debates was Boethius' commentary on Porphyry's *Isagoge*, where we find a detailed discussion of the ontological status of universals (see Boethius, A.M.S.; Porphyry §1). Boethius' commentary became a pivotal text for medieval philosophers, and by the twelfth century the debate over universals had become a dominant topic of philosophical concern. Two opposing views emerged. One, championed by William of Champeaux, was an extreme form of realism. On this view, a genus or a species is literally the same in all its members; the individuals falling under the kind are rendered distinct by the addition of forms to the common essence, and those forms are predicates of the essence. The opposing view, defended by Roscelin of Compiègne, represented an extreme version of nominalism. Insisting that all entities are particulars, Roscelin argued that talk of universals is merely talk about linguistic expressions that can be applied to a number of different particulars, and he held to an austere interpretation of linguistic expressions where universals are mere *flatus vocis* or vocalizations.

Abelard attacked both views (see Abelard, P. §5). He argued that since the forms which allegedly diversify any common essence are contraries, William's realism commits us to the view that a single entity can simultaneously exhibit incompatible properties; and he appealed to Aristotle's definition of the universal as that which can be predicated of many to call into question the claim that nonlinguistic entities can be universals. That definition, he contended, applies exclusively to things that can function as predicates in subject-predicate sentences, to what we nowadays call general terms. But while agreeing with Roscelin that we can 'ascribe universality to words alone' Abelard rejected Rocelin's claim that universals are mere *flatus vocis*, insisting that they are expressions that are significative or meaningful; and he argued that any adequate account of universals must show how, in the absence of a common essence, general terms can be meaningful. In doing so, it must answer two questions: (1) what is the cause of the imposition of a common name; and (2) what is it that we grasp when we grasp the signification of a common name?

Abelard's answer to the first question is disarming. The things to which the term 'man' applies, for example, all agree in being men. Their being men is the ground of the imposition of the common name 'man'. Abelard denies, however, that this agreement involves any common entity. He takes it to be an irreducibly primitive fact that all the things called men agree in that they are all men. In response to the second question, Abelard argues that what we grasp when we understand the common name 'man' is not any of the particular men named by that term; nor is it the collection consisting of all those particulars. To explain the kind of cognition associated with general terms, Abelard appeals to the distinction between perception and intellection. In perception, we grasp the particular named by a proper name; the cognition associated with general terms is, however, intellective. Here, the mind is directed toward an object of its own making, a *res ficta*. The *res ficta* is a kind of image, one that is *communis et confusa* (common and indifferent). It is common to all the items named by the associated general term and proper to none. Accordingly, it represents them all indifferently. Since it is something distinct from any of the particulars that fall under the Aristotelian categories, it is neither a substance nor an accident. It is the product of the intellect's activity of abstraction, and it is what is signified by the associated general term - it is what we grasp when we understand such a term.

The century after Abelard brought a number of different developments. The appearance of the complete Aristotelian corpus gave a clearer picture of Aristotle's views on universals, and the rich framework of semantical concepts associated with the developing terminist logic made possible the articulation of a more powerful form of nominalism than those defended by Roscelin and Abelard. That articulation came from William of Ockham (see William of Ockham §6). Following terminist logicians, Ockham distinguished between categorematic and syncategorematic terms. Categorematic terms are expressions whose significance derives from their having a 'definite and determinate signification'. Syncategorematic terms, by contrast, do not serve as signs of objects; their significance derives from the roles they play when used in conjunction with categorematic terms. Categorematic terms are further divided into discrete and common terms, where this is the contrast between expressions that signify just one object and expressions that signify many and so are predicable of many.

Like Abelard, Ockham identifies universals with common terms. Insisting that every existing thing is a particular, he construes the distinction between universals and particulars as a distinction between categorematic terms that signify just one thing and those that signify many. But where Abelard takes conceptual representations to be nonlinguistic items that function as the significata of the various common terms, Ockham wants to claim that a common term like 'man' signifies the various particulars of which it is truly predicable, and that the conceptual representation corresponding to the term 'man' is itself a linguistic entity. His idea is that thinking is inner

dialogue, best understood by way of the familiar concepts appropriate to spoken and written language. Thus, concepts are mental terms, judgments are mental propositions, and inferences are mental syllogisms. Conceptual linguistic items differ, however, from spoken or written words in that while the latter are only conventionally significant, the former are natural signs. The phoneme ‘man’ has the meaning it does only by virtue of a complicated system of conventions, but the concept *man* is something whose intrinsic nature it is to play the linguistic role it does; and although we describe conceptual representations by the use of concepts derived from our characterization of spoken and written language, mental language is prior to both. Just as written language is an outgrowth of spoken language, spoken language is an extension of mental language, a sort of ‘thinking out loud’.

The distinction between discrete and common terms, therefore, applies to conceptual representations, so that there are conceptual universals - conceptual representations that are predicable of many - and these are the genuine universals. Since they are naturally significant, the common terms of mental language are in their intrinsic nature items predicated of many; and their universality is the root of the merely conventional universality of the common terms of spoken and written language. Corresponding to the spoken/written term ‘universal’ there is a mental common term *universal*. To bring out the contrast between his own form of nominalism and its realist alternatives, Ockham tells us that this mental term is a term of second rather than first intention; it is a term that signifies not extramental entities but, rather, intentions of the soul, those that are in their intrinsic nature signs of many.

In characterizing the conceptual items that are in their intrinsic nature universals, Ockham mentions three possible views. One harkens back to Abelard and construes the mental term as a *res ficta* not found in any Aristotelian category; a second construes mental terms as qualities of the soul that serve as objects of its acts of understanding; the third identifies the mental term with the act of understanding itself. Over the course of his career, Ockham wavers between these views, but he ultimately comes to endorse the third view on grounds of theoretical simplicity.

Ockham’s nominalism extends beyond a concern with universals and the distinction between common and discrete terms. He is also interested in the distinction between concrete and abstract terms, between terms like ‘man’ and ‘humanity’, ‘courageous’ and ‘courage’, and he is concerned to undermine what initially appears to be a plausible account of this distinction. A natural response to the distinction is to say that whereas concrete terms signify familiar concrete particulars (the particulars which are men and courageous), their abstract counterparts signify the abstract entities (humanity and courage) those particulars exhibit. Since the distinction between concrete and abstract terms is found in all ten Aristotelian categories, the view unfolds into the claim that for each category there is a categorically different kind of abstract entity such that, in virtue of exhibiting an entity of that kind, a particular comes to be characterized by the appropriate concrete term.

In combating this view, Ockham argues that the categories do not represent a classification of nonlinguistic objects; they are, rather, a classification of linguistic expressions according to their mode of signification. As he sees it, there are nonlinguistic objects corresponding only to the categories of substance and quality, and the entities in question are all particulars. So there are particular substances (like this man) and particular qualities (like the whiteness of this piece of paper). Abstract terms from the category of substance do not signify anything distinct from the particular substances signified by their concrete counterparts. Abstract terms from the category of quality do tend to signify entities distinct from the familiar substances that we say are white and courageous; but a term like ‘courage’ does not signify some one quality that all courageous individuals share. ‘Courage’ is better construed as a general term signifying individual qualities, the various courages in virtue of which individual human beings are called courageous. In none of the other Aristotelian categories do abstract terms signify any entities distinct from those signified by their concrete counterparts. Indeed, Ockham wanted to claim that abstract terms from categories other than that of quality are eliminable from discourse, that sentences incorporating terms like ‘paternity’ and ‘burglary’ can be replaced, without loss of content, by sentences in which those terms do not appear, but their concrete counterparts (‘father’ and ‘burglar’) do; and a significant portion of his ontological/logical works is dedicated to showing how these translations are to go (see Aristotelianism, medieval).

3 Classical British empiricism

The classical empiricists followed Abelard and Ockham in denying that general terms signify universals. Thus, Hobbes sounds a familiar theme when he tells us that the only things that exist are particulars and that the terms

'general' and 'universal' are just 'names of names'. Like their medieval forbears, the empiricists recognized that the plausibility of this view hinges on our ability to provide a satisfactory account of the relation between general terms and the inner representations or ideas corresponding to them. Locke, who agrees that extramental entities are one and all particulars, argues that words signify ideas and that the ideas corresponding to general terms are abstract ideas - ideas formed from our ideas of particulars by separating out the features peculiar to this or that particular, retaining 'only what is common' to all the things to which a given general term applies. Berkeley goes further in his nominalism, denying that we have abstract ideas of the sort Locke describes. On Locke's view, the process of forming an abstract idea of a triangle, for example, consists in separating out all those features with respect to which triangles differ; and the result of this process is an idea of a triangle that is 'neither oblique nor rectangle, neither equilateral nor equicrural nor scalenon, but all and none of these at once'. Berkeley challenges us to identify an idea that corresponds to this characterization. As he sees it, our ideas are determinate in all their features and, accordingly, particular in their content (see Hobbes, T. §3; Locke, J. §6; Berkeley, G. §2).

While Berkeley attacks the view that ideas are general in virtue of being abstract, he concedes that there are general ideas; but he insists that the generality of an idea is a function of its role in thinking rather than any special kind of content. Ideas are general not because they result from abstraction in Locke's sense, but because the idea is made 'to represent or stand for all other particular ideas of the same sort'. So the mind takes an idea that is fully determinate and particular in its content and makes it stand for other ideas of the same kind. Hume wholeheartedly endorses Berkeley's attack on abstraction and his account of generality, telling us that general ideas are ' in themselves individual, however they may become general in their representation. The image in the mind is only that of a particular object, though the application of it in our reasoning be the same as if it were universal' (Hume [1740] 1978: 20).

4 The twentieth century

Like ontological debates in earlier periods, ontological discussions in early analytic philosophy typically focused on the problem of universals. Thus, Frege, Moore and Russell were all anxious to undermine nominalistic theories that seek to analyse subject-predicate discourse without reference to nonlinguistic universals; and when the later Wittgenstein attacks the view that the use of a general term like 'game' is grounded in the antecedent recognition of a property or set of properties common to all the items to which the term applies, he is, among other things, challenging their Platonistic accounts of subject-predicate discourse (see Wittgenstein, L. §9). Although concern with universals has continued throughout the twentieth century, the investigations of recent nominalists bear on a wider range of issues than those of their medieval and classical modern forbears. In addition to concern with universals, contemporary nominalists attempt to provide reductive accounts of things as diverse as the mathematician's sets, propositions, states of affairs, events and possible worlds, and philosophers of a nominalistic spirit take different attitudes towards different items on this list. Some, for example, are nominalists with regard to the traditional universals while insisting on a Platonistic account of sets; others insist on the irreducibility of events while providing reductive accounts of discourse apparently about propositions, states of affairs and universals. Indeed, few philosophers have been willing to defend a nominalistic approach to all of the so-called abstract entities. One exception is Wilfred Sellars (see Sellars, W. §2).

The account given by Sellars is an elaboration of Ockham's suggestion that talk about abstract entities is metalinguistic discourse. This suggestion had previously been elaborated in a proposal by Rudolf Carnap *The Logical Syntax of Language* that we construe talk about abstract entities as pseudo-material mode discourse, discourse apparently, but not really about nonlinguistic objects (see Carnap, R. §3). Carnap's concern is with sentences of the following sort:

(1) Courage is a property.
(2) Mankind is a kind.
(3) Paternity is a relation.
(4) That two plus two equals four is a proposition

His proposal is that we treat these sentences as disguised ways of making claims about the syntax of certain linguistic expressions. Thus, (1)-(4) become:

(1′) 'Courage' is an adjective.

(2′) ‘Man’ is a common noun.
(3′) ‘Father of’ is a many-place predicate.
(4′) ‘Two plus two equals four’ is a declarative sentence

The difficulty with this proposal is that (1)-(4) turn out to be claims about *English* expressions. The proposal forces us to take the Spanish counterparts of (1)-(4), for example, to be claims about *Spanish* words, so that (1)-(4) and what are supposed to be their Spanish translations do not even agree in reference. Sellars responds to this problem by introducing a kind of quotation that cuts across languages, called dot quotation. Whereas standard quotation of the sort we meet in (1′)-(4′) creates metalinguistic expressions that apply exclusively to words in the quoting language, the application of Sellars’ dot quotation to an expression creates a metalinguistic common noun that is true of all those expressions, regardless of language, which play the same linguistic role that the quoted expression plays in the base language. Thus, ‘ ·man·’ is a common noun true of ‘hombre’, ‘homme’, and ‘Mensch’. In their respective languages, these terms play the same role that ‘man’ plays in English; they are all ‘ ·man·’s. Now, Sellars wants to claim that using the machinery of dot quotation, we can provide a satisfactory reconstruction of (1)-(4) as:

(1″) ·Red·s are adjectives.
(2″) ·Man·s are common nouns.
(3″) ·Father of·s are many-place predicates.
(4″) ·Two plus two equals four·s are declarative sentences.

As Sellars understands them, (1″)-(4″) represent claims about linguistic expressions construed as tokens rather than types; and talk of linguistic tokens can be recast as talk about speakers and inscribers. Even the apparent Platonism involved in talk about linguistic roles is illusory since talk about linguistic roles can be eliminated by reference to talk about the linguistic rules that govern the use of terms. Accordingly, talk apparently about abstract entities is consistent with the most austere nominalism; it is merely metalinguistic discourse that cuts across languages.

Sellars believes that the sort of account he proposes for (1)-(4) can be extended to handle all discourse involving the so-called abstract entities. A slightly less radical form of nominalism is found in the writings of W.V. Quine (see Quine, W.V. §6). Early in his career, Quine espoused a nominalism as austere as that developed by Sellars, but by the time he wrote *Word & Object* (1960), he had concluded that there is one kind of abstract entity whose existence we have to acknowledge, the mathematician’s set or class. Quine remains unwilling to recognize things like properties, relations, kinds, and propositions, however. Unlike sets, these alleged entities lack clear-cut identity conditions and should play no role in our ontology.

Most contemporary philosophers agree with Quine that we must endorse an ontology of sets. This view provides the backdrop for the reductive approach to universals defended by G.F. Stout and D.C. Williams. They hold that there are particular as opposed to general qualities or properties, things like the whiteness of this piece of paper. So there are abstract entities besides sets; but they are one and all particulars. Williams calls these abstract particulars *tropes* and he tells us that they constitute ‘the alphabet of being’. Tropes are ontologically primitive, and items from other categories are constructions out of them. Thus, the universal of the Platonist is a set of resembling tropes; and familiar concrete objects are bundles of tropes that contingently enter into a relation of ‘collocation’.

Although Williams’ trope-theoretic nominalism continues to enjoy some popularity, the most prominent form of nominalism in the contemporary arena is that influenced by developments in the semantics of modal logic, where we meet the idea that the actual world is just one of infinitely many possible worlds and that the totality of possible worlds constitutes the subject matter for talk about necessity and possibility (see Possible worlds). Contemporary nominalists claim that the framework of possible worlds provides the resources for a genuinely reductive account of things like properties and propositions. These philosophers propose that we take possible worlds as primitive. Each such world, they claim, can be characterized in nominalist terms as a totality of concrete particulars, and they argue that we can provide a nominalist treatment of things like properties and propositions by identifying them with set-theoretical entities of a transworld sort. We can identify properties with functions from worlds to sets of objects, relations with functions from worlds to sets of ordered n-tuples and propositions with sets of worlds or functions from worlds to the truth values. The most prominent proponent of this sort of view is David Lewis (see

Lewis, D. §3). He has invoked the framework of possible worlds not simply to provide an account of properties and propositions, but to clarify the concept of meaning, to state truth conditions for counterfactuals, and to provide an analysis of causation (1986).

See also: Abstract objects; Intensional entities; Nominalism, Buddhist doctrine of; Universals

MICHAEL J. LOUX

References and further reading

Abelard, P. (before 1120) 'Logica "Ingredientibus"', in R. McKeon (ed.) *Selections from Medieval Philosophers*, New York: Scribners, 1959, 202-58.(Provides a survey of Abelard's approach as outlined in §2 above.)

Armstrong, D. (1978) *Universals and Scientific Realism*, Cambridge: Cambridge University Press, 2 vols. (Contains useful discussions of nominalism.)

Berkeley, G. (1710) *A Treatise Concerning the Principles of Human Knowledge* in C. Turbayne (ed.) *Principles, Dialogues, and Correspondence*, New York: Bobbs-Merrill, 1965. (The introduction to this work is especially relevant.)

Carnap, R. (1934) *Logische Syntax der Sprache (The Logical Syntax of Language)*, trans. A. Smeaton, Patterson, NJ: Littlefield & Adams, 1959.(Carnap's attempt to clarify the structure of language.)

Goodman, N. (1956) *The Problem of Universals*, Notre Dame, IN: University of Notre Dame Press.(The chapter entitled 'A World of Universals' is especially relevant.)

Hobbes, T. (1651) *Leviathan*, ed. M. Oakeshott and R. Peters, New York: Collier, 1962. (Part 1, chapter 4 contains Hobbes' exposition of the topic.)

Hume, D. (1740) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, 2nd edn (with revisions by P.H. Nidditch), Oxford: Clarendon, 1978.(Book I, part 1, section 7 contains Hume's response to Berkeley on abstract ideas.)

Lewis, D. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.(Uses possible worlds as the framework for a new perspective, as outlined at the end of §4 above.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, collated and annotated by A.C. Fraser, Oxford: Oxford University Press, 1894.(See Book II, chapter 9, sects 9-10 and Book III, chapters 3 and 4.)

Loux, M.J. (1978) *Substance and Attribute*, Dordrecht: Reidel.(Contains discussions of a variety of different forms of nominalism.)

Ockham, William of (*c*.1329) 'Summa Logicae', part I, in M.J. Loux, *Ockham's Theory of Terms*, Notre Dame, IN: Notre Dame University Press, 1974.(A translation of the clearest formulation of Ockham's approach to ontology.)

Quine, W.V. (1953) *From a Logical Point of View*, Cambridge, MA: Harvard University Press.(The essays in this volume express the earliest and most austere form of Quine's nominalism.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Gives the 'sets only' doctrine.)

Sellars, W. (1967) 'Abstract Entities', in *Philosophical Perspectives*, Springfield, IL: Charles C. Thomas. (Exposition of the line of thought outlined in §4 above.)

Stout, G.F. (1921) *The Nature of Universals and Propositions*, London Oxford University Press.(British Academy Lecture which discusses 'Trope Theory', outlined in §4 above.)

Williams, D.C. (1953) 'Elements of Being', parts I and II, *Review of Metaphysics*, 6: 3-18, 171-93.('Trope theory' outlined in §4 above.)

Wittgenstein, L. (1953) *Philosophical Investigations*, trans. G.E.M. Anscombe, London: Macmillan.(The famous passage on 'family resemblances' begins at §65.)

Nominalism, Buddhist doctrine of

Buddhist nominalism refers to the nominalist ontology and semantics developed especially by the Indian Buddhist philosophers Dignāga and Dharmakīrti. Elaborating on the arguments of their Buddhist predecessor Vasubandhu, they critically examine the notions of spatial and temporal extension. For Dignāga and Dharmakīrti, spatially and temporally composite entities are constructed through concepts and language and as such those entities exist only nominally or conventionally. Their semantics rejects the realist position that expressions refer to real, extra-mental universals that are instantiated in each particular of the class formed by the respective universal. Instead, these philosophers developed the unique theory of 'exclusion' whereby expressions convey meaning by the exclusion of some particulars from those which do not have the expected causal capacities. Dharmakīrti's nominalism is credited with a greater impact on Indian philosophy than Dignāga's.

1 Ontology

One can reduce the concerns of Buddhist philosophers to a central preoccupation: the cessation of suffering. According to Buddhist philosophers, the elimination of suffering (*duḥkha*) requires the elimination of its cause, which is thought to be the belief in entities that in fact do not exist (see Suffering, Buddhist views of origination of). Of particular concern is the belief in an eternal, unchanging self or soul (*ātman*), for this belief motivates the behaviour that produces suffering. Buddhists before Dharmakīrti often employed part or whole arguments to refute the notion of self and eventually to reject all composite entities. These arguments, which inform much of Dharmakīrti's nominalism, can be summarized by briefly examining a composite entity, such as a table.

Although a table is apparently a single entity composed of various parts, Buddhists question whether any such entity actually exists. They ask whether the whole that is the table is the same as or different from its parts. If the whole differs from its parts, what evidence do we have for its existence? Immediate sensory perception yields information only about the parts. If the whole is the same as its parts, then it must be either fully or partially instantiated in each part. If it is fully instantiated in each part, then any single part is an entire table. If the table is only partially instantiated in its parts, then we should speak of many partial tables rather than one single table. If we were to claim that all the parts together are the table, then a heap of table-parts should be a table. Finally if the parts in a particular configuration are the table, one could again ask whether that configuration is the same as or different from the configured parts: an infinite regress ensues.

Expanding on this critique of spatial extension, Dharmakīrti also attacks temporal extension. In doing so, he affirms the doctrine of momentariness, whereby an ultimately real entity can only exist for an infinitesimal period of time (see Momentariness, Buddhist doctrine of). This doctrine is supported in part by the argument that an entity which does not change over time cannot produce any effects, for a cause must change from a pre-production state to a post-production state; otherwise, it would either produce its effects at all times eternally, or it would never produce any effect. Furthermore, Dharmakīrti maintained a causal model of sense perception whereby a sensed entity causes the content of one's sensory cognition. Since a nonmomentary entity could never act as a cause, it could never be perceived. As a result, one could never indubitably establish its existence, for in Dharmakīrti's words, 'to exist is to be perceived' (*sattvam upalabdhir eva*) (*Pramāṇavārttikasvopajñavṛtti on Pramāṇavārttika* I.3ab).

In short, Dharmakīrti denies that any composite entity can be real in the strictest sense. However, following his predecessors, he admits that composite entities can be said to exist from the perspective of conceptual and linguistic conventions. These modes of existence yield two important categories: things that exist in the strictest sense are 'ultimately' real (*paramārthasat*), while those contingent upon linguistic and conceptual conventions are only 'nominally' (*prajñaptisat*) or 'conventionally' real (*saṃvṛtisat*).

2 Critique of universals

Dharmakīrti's most thorough analysis of semantic issues appears in the *Svopajñavṛtti (Interpretative Commentary)*, his lengthy remarks on the first chapter of his *Pramāṇavārttika (Comments on Instrumental Knowledge)*. Much of his work concerns the problem of repeatability or universality (*anvaya*). In other words, when I use the expression 'cow', for example, what is it about all cows that allows me to apply this one word to all

of them? On the pre-reflective intuitions suggested by language, it seems that there is something the same about all cows. Elaborating on this intuition, most Indian philosophers maintained that there is in fact some real entity - in this case, 'cowness' - that is instantiated in each cow. Entities such as 'cowness' are called 'universals' (*sāmānya, jāti*) and they are instantiated in 'particulars' (*vyakti, svalakṣaṇa*), the things that impinge on our senses and which we seem to identify as 'cows'. Here, the technical term 'referent' might be used in a way that differs from some of its more common uses in Western philosophy. That is, to describe the view of Indian realism, the actual referent of the expression 'cow' might best be considered the universal 'cowness' itself (see Universals, Indian theories of).

When Dharmakīrti attacks this view, he maintains that an expression is successful in its semantic function only when it has induced an action (*pravṛtti*) towards the intended particular in the person who has apprehended that expression. Such action might be some physical manipulation of the particular or a mere cognitive act such as recognition. That being the case, ultimately real universals make meaning impossible because expressions which refer to ultimately real universals would be incapable of inducing action towards particulars.

To demonstrate this conclusion, Dharmakīrti raises the problem of identity and difference as a species of the part or whole analyses previously mentioned. He asks whether the universal is the same as or different from the particulars in which it is instantiated. If the universal is different, then the expression 'cow' would not induce acts towards any particular because it refers only to the universal 'cowness', which is entirely distinct from any particular. One might still insist that the expression can induce action towards the intended particular, but then one must admit that an expression can induce action towards any entity, whether it be the same as or different from its referent.

On the other hand, if the universal is identical with the particulars, then the universal would lose its universality or repeatability. In other words, a cow-particular must be distinct from all other cow-particulars, otherwise, one would be unable to distinguish one cow from any other. Hence, a cow-particular is not repeatable - one cow-particular does not occur in or as any other cow-particular. But if the universal 'cowness' is identical with any cow-particular, then it would also be unrepeatable. One would need a new universal for every instance and the expression 'cow' could only be applied to that one particular, which is identical to the universal 'cowness'.

In response, some Indian realists admit that the particular and universal must be distinct, but they claim that an expression which refers to a universal can induce action towards a particular by virtue of the particular's relation to the universal. Among the more compelling criticisms of this position is Dharmakīrti's contention that any form of relation necessarily leads to either a contradiction or an infinite regress. That is, a relation's relata must be either the same or different. If they are the same, then there is no relation, for relations presuppose difference. If, on the other hand, the relata are different, then they cannot be in relation, for if entirely distinct entities can stand in relation, then one can haphazardly relate any entity to any other entity. The notion of relation would thus be meaningless. Only a contradiction remains: the relata are both the same and different. If in response one posits a subsistent relation that 'ties' the relata together, one must then explain how the relata are related to the relation. If one then speaks of some second-order relation that connects the relata to the relation, one falls into an infinite regress.

3 Theory of meaning

Dharmakīrti's ontology can be reduced to the claim that the ultimately real is necessarily unique and unrepeatable. Hence, he rejects realist semantics because a universal must be in some sense repeatable in each of its instances. Yet without real universals, how do we explain our ability to use language? The Buddhist answer is that expressions convey meaning by the exclusion of some particulars from those which do not have the expected causal capacities. In this respect, meaning has three components: the cognitive image (*pratibhāsa, ākāra*), the particular and the exclusion (*vyāṛtti*).

For Dharmakīrti, language operates in the same fashion as conceptual thought. We can most easily examine the role of cognitive images by turning to the cognition called 'recognition' (*pratyabhijñāna*) - the conceptual act whereby a sensed object is identified or labelled. For recognition to occur, one must have learned the conventions that govern the appropriate expression or concept. If, for example, one knows the conventions that govern 'cow', one can have the recognition, 'that is a cow'. For Indian philosophers the important question is, what is it that we label with concepts or expressions such as 'cow'?

In discussing recognition, Dharmakīrti assumes a causal model of sense perception: when some particular impinges upon the senses, that particular creates an image in the mind. This image or sensum is what one knows in sense perception. This means that what we recognize as a 'cow' is not some particular that impinged upon the senses; rather, it is the image produced in our minds by that particular. Hence, if expressions and concepts give us any knowledge of particulars, they do so only by the mediation of the cognitive images caused by those particulars.

Objectors to this position point out that if concepts and language yield knowledge just of cognitive images, then actions based on such knowledge would be focused upon images and not upon particulars. Dharmakīrti responds that ordinary persons have a deeply inculcated cognitive habit of mistaking the images that arise from particulars for those particulars themselves. This claim leads to his assertion that all linguistic and conceptual knowledge is flawed in that it rests upon the conflation of the image with the particular that caused it. Dharmakīrti remarks that this psychological apparatus explains how images can induce action towards particulars. He further notes that an image nevertheless can yield useful information because it directs one only towards the particular which acted as its cause.

In appealing to the causal link between images and particulars, Dharmakīrti makes a crucial statement about images. That is, Dharmakīrti argues that any entity which acts as a cause or effect is a particular and since images are effects, they must be mental particulars. As particulars, images are never repeatable, so images cannot be the same in all cases when we use a given concept or expression. Instead, Dharmakīrti maintains that what is actually the same in all cases are 'exclusions'.

His theory begins with the claim in the *Pramāṇavārttikasvopajñavṛtti on Pramāṇavārttika* (I, V 107-9; 119-43) that all images of tables, for example, produce an effect that no non-table produces: a second-order cognition called the 'determination of sameness' (*ekapratyavamarśajñāna*) in which any table-image is determined to be the same as every other table-image in that they are all imagined to be tables. Hence, all table-images are the same in that they all produce a cognition of their sameness. Since those images were produced by certain particulars, we can likewise say that those particulars are the same because they all produce images that all produce a second-order cognition of their sameness.

Some Indian realists maintain that this claim leads to an infinite regress. That is, in this respect all images of tables are the same because they produce the same effect, namely, second-order cognitions of their sameness. But since those second-order cognitions are all effects, they must be unique mental particulars. Thus, we must also prove that those second-order cognitions are the same. To do so, we must show that they all produce the same third-order cognition and an infinite regress ensues. In response, Dharmakīrti admits that the sameness of a certain set of images is constituted by their production of the same second-order cognitions. Nevertheless, the sameness of the second-order cognitions is not constituted by the sameness of their effects. Rather, they are all said to be the same in that they determine the first-order cognitions to be the same. This is an appeal to experience: the fact that the second-order cognitions include this determination of sameness is established by the apperceptive aspect of those second-order cognitions themselves.

It might appear that Dharmakīrti has opted for a realist theory of resemblance: certain particulars are the same in that they have the same causal capacity to produce images that in turn produce a determination of their sameness. But he rejects the notion of some shared causal capacity precisely for the reasons mentioned above: such an entity would have to be either the same as or different from the particulars in which it is instantiated and neither possibility makes sense.

Given that there can be no single entity that is the same in all cases, how then can we use this information that all tables, for example, produce the same effect? If we cannot even conclude that there is some repeatable causal capacity, then what will provide the commonality in all table-particulars such that we can use the expression 'table' for each one of them? Dharmakīrti answers that we need not find any repeatable entity at all. Instead, we need only admit that if all tables produce the same effect, this necessarily distinguishes them from all other particulars. In short, what is the same about all tables is that they are excluded from non-tables.

This move seems tantamount to calling a table a non-non-table and one might wonder why we should resort to this double negative: why not simply say that a table is a table? Although modern interpreters have taken this objection

seriously, Dharmakīrti noted that this qualm is misplaced for it fails to see how anti-realist ontological claims inform his semantics. Dharmakīrti's theory rests on the argument that repeatable entities are impossible. Since we undeniably experience particulars in such a way that we are able to classify them, our cognitions must group particulars into classes without having to pinpoint any repeatable entity. But since it is impossible to group particulars positively, we must do so negatively. The point here is that while we cannot posit any single, repeated entity that accounts for the fact that the particulars of a given class produce cognitive images that in turn produce the determination of their sameness, we are still able to distinguish between particulars that lead to that determination and particulars that do not.

The Mīmāṃsaka philosopher Kumārila (*fl*. *c*.625) responds that the reasoning employed is circular. That is, if Buddhists define 'tree' as the negation of 'non-tree', then one must be able to specify non-trees. Either this requires that one has already specified what trees are, or that one is somehow able to specify directly what non-trees are. In the former case, the theory of exclusion is superfluous since one can specify trees before excluding them from non-trees. In the latter case, one has specified what are non-trees without the negation of their complement. If this is possible in the case of non-trees, why should it not also be possible in the case of trees? The answer is the one stated above: we know that images are different by the apperceptive second-order cognitions that they induce. Nevertheless, we cannot specify any repeatable entity possessed by a certain set of images and not possessed by the set of all other images. We can only appeal to the difference and we use that difference to define both sets. Dharmakīrti suggests that if this theory is circular, then no theory of semantics will be able to avoid circularity.

Clearly, the theory of exclusion assumes some cognitive means of comparing images. Dharmakīrti notes that this process involves imprints (*vāsanā*) placed in the mind by previous experiences and that in the act of comparison those imprints become active. Another crucial factor enters into comparison: the intentions and expectations present in one's mind when one uses concepts and expressions. That is, since any cognitive image is necessarily unique it can be excluded from an infinite range of other images. For example, any table-image is unique and one can focus on its uniqueness by excluding it from all other images, even other table-images. If we wish to say that the particular which produced that image is a 'table', we ignore that image's difference from other table-images. Also, if we wish to speak of it as 'furniture' we widen our scope by ignoring its difference from chairs and such while maintaining its difference from all non-furniture, such as ping-pong balls. This process is a function of the goals, expectations and other disposition that we bring to any use of language and concepts. While accounting for the importance of dispositions in the determination of meaning, the theory of exclusion also allows Dharmakīrti to account for repeatability and reference without making any ontological commitments to universals. As far as the Indian realist account is concerned, it would seem that exclusions are universals instantiated in images. Hence, Dharmakīrti would face the same identity/difference objection that he levelled against the realists. To avoid this problem, Dharmakīrti resorts to the aforementioned distinction between ultimate and conventional reality. As a species of negation, exclusions cannot exist ultimately. Hence, no expression has any real referent. Nevertheless, in conventional terms exclusions can be said to qualify cognitive images and since any image is the unique effect of some particular, that causal link enables Dharmakīrti to claim that the particulars which cause images are the indirect referents of expressions. His semantic theory enables him to avoid the realist doctrine of universals when he claims that the information derived from language can lead to successful activity in the world.

See also: Buddhist philosophy, Indian; Dignāga; Meaning, Indian theories of; Nominalism; Vasubandhu

JOHN DUNNE

References and further reading

Dharmakīrti (*c*.625) *Pramāṇavārttikasvopajñavṛtti on Pramāṇavārttika (Svopajñavṛtti) (Pramāṇavārttikam: The First Chapter with Autocommentary)*, Serie Orientale Roma 23, ed. R. Gnoli, Rome: Istituto Italiano per il Medio ed Estremo Oriente, 1960.(Sanskrit edition of the most important source for views discussed in §2 and §3. Relevant portions translated in Dunne (1997).)

Dharmakīrti (*c*.625) *Pramāṇavārttika of Ācārya Dharmakīrti with Manorathanandin's Commentary*, ed. S. Dvārikādāsa Śāstrī, Bauddha Bharata Series 3, Varanasi: Bauddha Bharati, 2nd edn, 1984.(Sanskrit edition; portions of the chapter on direct awareness (*pratyakṣa*) are particularly relevant to §1.)

Dreyfus, G. (1997) *Recognizing Reality: Dharmakīrti's Philosophy and its Tibetan Interpretations*, Albany, NY: State University of New York Press.(An important work that offers some competing interpretations on these

issues.)
Dunne, J.D. (1997) 'Rending the Web of Concepts: Dharmakīrti's Nonessentialist Approach to Language and Concepts', unpublished Ph.D. dissertation, Harvard University, MA.(Expands on the presented here; includes a translation of Dharmakīrti's *Pramāṇavārttikasvopajñavṛtti* up to verse 231.)
Hayes, R.P. (1988) *Dignāga on the Interpretation of Signs*, Dordrecht: Kluwer.(An excellent study of Dharmakīrti's predecessor. Chapters 2, 3 and 5 are particularly relevant.)
Kumārila Bhaṭṭa (7th century) *Ślokavārttika: Translated from the Original Sanskrit with Extracts from the Commentaries 'Kāśikā' of Sucarita Miśra and 'Nyāyaratnākara' of Pārthasārathi Miśra*, trans. G. Jha, New Delhi: Śri Satguru Publications, 1983.(Reprint of 1908 edition; although at points problematic, this translation provides access to the views of an influential opponent of Buddhist nominalism.)
Shah, N.J. (1967) *Akalaōka's Criticism of Dharmakīrti's Philosophy: A Study*, Lalbhai Dalpatbhai Series 11, D. Malvania and A.P. Shah (eds), Ahmedabad: L.D. Institute of Indology.(Includes a useful summary of Dharmakīrti's views relevant to §§2-3.)
Siderits, M. (1991) *Indian Philosophy of Language: Studies in Selected Issues*, Studies in Linguistics and Philosophy 46, Dordrecht: Kluwer.(Discusses the issues raised in §3, while presenting a different approach to the apparent problem of double negation).
Steinkellner, E. (1971) 'Wirklichkeit und Begriff bei Dharmakīrti' (Reality and Concept in Dharmakīrti), *Wiener Zeitschrift für die Kunde Südasiens und Archiv für Indische Philosophie* 15: 179-211.(This seminal article, although technical at points, expands on §1 and §3.)

Non-constructive rules of inference

For some theoretical purposes, generalized deductive systems (or, 'semi-formal' systems) are considered, having rules with an infinite number of premises. The best-known of these rules is the ' ω*-rule', or rule of infinite induction. This rule allows the inference of* $\forall n\Phi$*(n) from the infinitely many premises* Φ*(0),* Φ*(1),... that result from replacing the numerical variable n in* Φ*(n) with the numeral for each natural number. About 1930, in part as a response to Gödel's demonstration that no formal deductive system had as theorems all and only the true formulas of arithmetic, several writers (most notably, Carnap) suggested considering the semi-formal systems obtained, from some formulation of arithmetic, by adding this rule. Since no finite notation can provide terms for all sets of natural numbers, no comparable rule can be formulated for higher-order arithmetic. In effect, the* ω*-rule is valid just in case the relevant quantifier can be interpreted substitutionally; looked at from the other side, the validity of some analogue of the* ω*-rule is the essential mathematical characteristic of substitutional quantification.*

1 The ω-rule

In a standard formalized deductive system, the axioms form a decidable set of formulas (or, under Gödel-numbering, a recursive set of numbers), and each rule is a decidable relation between (conclusion) formulas and finite sets of (premise) formulas. (A set, or relation, is 'decidable' if there is an algorithm for determining whether a given item belongs to it, or given items stand in it.) A proof in the system can therefore be thought of as a tree of formulas, with the theorem to be proved at the root, axioms at the leaves, and at each node a formula which follows by a rule from its predecessors.

For some theoretical purposes, however, 'semi-formal' systems (generalized deductive systems) are considered, having rules with an infinite number of premises. The best-known of these rules is that called the rule of infinite induction, or the 'ω-rule'. This rule allows the inference of $\forall n\Phi(n)$ from the infinitely many premises $\Phi(0), \Phi(1), \ldots$ that result from replacing the numerical variable *n* in $\Phi(n)$ with the numeral for each natural number. About 1930, in part as a response to Gödel's demonstration that no formal deductive system had as theorems all and only the true formulas of arithmetic, several writers (most notably, Carnap) suggested considering the semi-formal systems obtained, from some formulation of arithmetic, by adding this rule.

An axiomatic system enriched with the ω-rule can, obviously, prove all the truths of first-order arithmetic: without loss of generality we may think of the atomic sentences of arithmetic as equations between terms made up out of numerals and the signs for addition and multiplication, so propositional logic and elementary algorithms (which can be simulated in standard axiomatic systems) suffice to prove all quantifier-free sentences of arithmetic, after which any true existential quantification can be inferred by the standard rule of existential quantifier introduction from one of its true instances, and any true universal quantification by the ω-rule from the set of all its true instances. (Proofs of true 'prenex' sentences (that is, those made up of a string of quantifiers followed by an atomic sentence) have this form exactly; other truths can be deduced from their prenex equivalents by standard means.)

If the language of the system goes beyond first-order arithmetic the situation is more complicated. Within a few years of its proposal, Rosser (1937) showed that the addition of the ω-rule to a standard system of higher-order arithmetic did not suffice to render the system complete. There are, however, some partial completeness results. Schütte (1960) showed that standard axioms and rules enriched with the ω-rule remain complete when the language of first-order arithmetic is supplemented with free variables for sets of numbers. Earlier, Orey (1956) had shown that any set of sentences of second-order arithmetic consistent with the ω-rule has an 'ω-model', that is, a model in which the second-order quantification (quantification over sets of numbers) can have a nonstandard interpretation in the manner of Henkin (1950), but in which the first-order quantifiers range over the genuine natural numbers. (Note that Schütte's result is that every sentence true in the standard model is a theorem of a certain theory, whereas Orey's is more analogous to completeness theorems for logics in specifying only a kind of model.)

The finite trees of formulas counting as proofs in a given formal system themselves form a decidable set of arrays, and can be Gödel-numbered; the set of (Gödel-numbers of) theorems of the system is thus defined by a Σ^0_1 formula, that is, a formula obtained from one expressing a recursive relation by existentially quantifying over

numbers (see Recursion-theoretic hierarchies). This is the general form of a definition of a recursively enumerable set, and, as shown most elegantly by Smullyan (1961), any recursively enumerable set can be represented by the set of theorems of a formal system. A proof in a semi-formal system can be represented by a set of numbers, so it is tempting to define the set of (Gödel numbers of) theorems of a semi-formal system by a formula starting with an existential quantifier over sets, by analogy with the definition of a recursively enumerable set by existential quantification over numbers.

But an ω-proof is a *well-founded* tree of formulas: it may have infinitely many branches, though each branch is only finitely long. So specifying just which sets of numbers correspond to infinite proofs would require a second quantifier over sets, and this would yield a Σ^1_2 formula as definiens. It is better to define the set of theorems directly, without reference to the infinite proofs, as the set of formulas belonging to every set which contains the axioms, is closed under the ordinary rules of inference, and contains a universal (numerical) quantification if it contains all its instances. This yields a Π^1_1 formula (a formula obtained by prefixing existential quantifiers over sets to a formula containing at most first-order quantifiers) as a definiens. Since Frege (1879) it has been standard to use a formula of this form in giving an explicit definition of an inductively defined notion, and sets definable by Π^1_1 formulas are sometimes called 'inductive' sets (see Aczel (1977) or Moschovakis (1974) for details). Any Π^1_1 set can be represented by the set of theorems of a semi-formal system.

Formal systems of arithmetic can be given an elegant formulation analogous to natural deduction or sequent-calculus formulations of pure logic, with the principle of mathematical induction taking the form of a rule of inference similar to the rules for the quantifiers. When this is done, it turns out that Gentzen's cut-elimination theorem for pure logic (1935) does not extend in its general form to arithmetic, but his later proof of the consistency of arithmetic can be seen as proceeding via a special case of cut-elimination that does hold. Kurt Schütte noted that it was possible to modify Gentzen's methods to prove a general cut-elimination theorem for various systems of arithmetic, provided that the cut-free proofs are allowed to employ the ω-rule. He and his colleagues have developed this into a useful tool for analysing formal systems of arithmetic: the complexity of the cut-free ω-proofs corresponding to proofs in a particular formal system turns out to be a measure of the strength of that system, correlating in particular with the class of number-theoretic functions whose totality can be proved in the system.

2 Related rules

Stronger infinitary rules, producing semi-formal systems of second-order arithmetic which are still incomplete but are stronger than those formed by adding the ω-rule to standard formal systems, have been studied by Mostowski (1961) and Enderton (1967). No satisfactory analogue of the ω-rule for introducing the universal quantifications over sets in classical higher-order arithmetic or set theory is possible because a formal language cannot contain terms for all the uncountably many sets in the standard model, but higher-order analogues of the ω-rule have been studied in the context of non-classical set theories. Schütte (1960) describes a type-free theory inspired by Ackermann; and Fitch, in publications starting in the late 1940s, obtained results closely related to those of later workers on the inductive semantics of type-free theories by studying semi-formal systems with rules similar to the ω-rule.

In the other direction, various weaker forms of the ω-rule have been studied. It is sometimes possible to prove in a finitary metatheory (or in a theory like Peano arithmetic used as its own syntactic metatheory) that all the instances of a certain quantification are provable in a formal system. A rule allowing the inference of a universal (numerical) quantification from the statement that all its instances are provable is sometimes called the 'formalized' ω-rule. Such a rule, unlike those discussed above, is of finitary character; added to a formalized system of arithmetic it amounts to a kind of reflection principle. Feferman (1962) showed that every true sentence of first-order arithmetic is provable in some formal system in a transfinite hierarchy of systems obtained from a standard axiomatic system by iterated addition of reflection principles. (Much earlier, Hilbert (1931) had proposed a restricted version of the formalized ω-rule in what seems to have been an attempt to get around Gödel's incompleteness theorems.)

A somewhat related idea is the 'constructive' ω-rule: take $\forall n\Phi(n)$ to be a theorem if there is a recursive function assigning to each natural number (the Gödel number of) a formal proof of the corresponding instance of $\Phi(n)$. Despite the name, this is a non-constructive rule, since it is not in general a decidable matter whether or not there is a recursive function of the sort required: Shoenfield (1959) shows that all truths of first-order arithmetic are

provable with it.

A.P. HAZEN

References and further reading

Aczel, P. (1977) 'An Introduction to Inductive Definitions', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 739-82.(A survey of results on inductive definability; shorter than Moschovakis (1974).)

Carnap, R. (1935) 'Ein Gültigskeitskriterium für die Sätze der klassischen Mathematik', *Monatshefte für Mathematik und Physik* 42: 163-90; incorporated in *The Logical Syntax of Language*, trans. A. Smeaton, London: Kegan Paul, 1937.(Uses the ω-rule to give a definition of mathematical 'validity'.)

Dunn, J.M. and Belnap, N.D. (1968) 'The Substitution Interpretation of the Quantifiers', *Noûs* 2: 177- 85.(An influential study of the notion of substitutional quantification, pointing out that analogues of the ω-rule are valid for it.)

Enderton, H. (1967) 'An Infinitistic Rule of Proof', *Journal of Symbolic Logic*, 32: 447-51.(Treats a non-constructive rule more nearly analogous to the ω-rule than, but weaker than, that studied in Mostowski (1961).)

Feferman, S. (1962) 'Transfinite Recursive Progressions of Axiomatic Theories', *Journal of Symbolic Logic* 27: 259-316.(Study of iterated addition of reflection principles.)

Feferman, S. (1984) 'Toward Useful Type-Free Systems I', *Journal of Symbolic Logic* 49: 75-111.(Includes a historical survey describing the Ackermann-Schütte and Fitch type-free systems.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82.(First formalization of second-order logic; replacement of recursive definitions with explicit definitions quantifying over sets.)

Gentzen, G. (1935) 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* 39: 176- 210, 405-565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969; trans. 'Investigations into Logical Deduction', *American Philosophical Quarterly* 1: 288-306, 2: 204-18, 1964; repr. in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 68-131.(Cut elimination.)

Gentzen, G. (1936) 'Die Widerspruchsfreiheit der reinen Zahlentheorie', *Mathematische Annalen* 112: 493- 565; trans. 'The Consistency of Elementary Number Theory', in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 132-213.(Proof of consistency of arithmetic.)

Henkin, L. (1950) 'Completeness in the Theory of Types', *Journal of Symbolic Logic* 15: 81-91. (Introduces nonstandard models for higher-order logic and arithmetic.)

Hilbert, D. (1931) 'Die Grundlegung der elementaren Zahlenlehre', *Mathematische Annalen* 104: 485-94.(An early proposal of a restricted version of the formalized ω-rule.)

Moschovakis, Y.N. (1974) *Elementary Induction on Abstract Structures*, Amsterdam: North Holland. (Inductive definition.)

Mostowski, A. (1961) 'Formal Systems of Analysis Based on an Infinitistic Rule of Proof', in *Infinitistic Methods: Proceedings of Symposium on Foundations of Mathematics*, Warsaw: PWN, and New York, Oxford, London, Paris: Pergamon Press, 141-66; repr. in *Foundational Studies* vol. 1, Amsterdam: North Holland, 1979. (Investigates systems of second-order arithmetic validated by a certain subclass of ω-models.)

Orey, S. (1956) 'On ω-Consistency and Related Properties', *Journal of Symbolic Logic* 21: 246-56. (Completeness theorem.)

Pohlers, W. (1987) 'Contributions of the Schütte School', appendix in G. Takeuti, *Proof Theory*, Amsterdam: North Holland, 2nd edn.(Survey of work by Schütte and colleagues.)

Pohlers, W. (1989) *Proof Theory: An Introduction*, Berlin: Springer.(Introductory textbook on the work of the Schütte school.)

Rosser, J. (1937) 'Gödel Theorems for Non-Constructive Logics', *Journal of Symbolic Logic* 2: 129- 37. (Incompleteness theorem.)

Schmerl, U. (1982) 'Iterated Reflection Principles and the ω-Rule', *Journal of Symbolic Logic* 47: 721-33. (Relates ideas of Feferman (1962) to more explicitly stated rules.)

Schütte, K. (1960) *Beweistheorie*, Berlin: Springer; trans. J.N. Crossley, *Proof Theory*, Berlin and New York: Springer, 1997; also trans. with modifications, S. Toledo, *Tableau Systems for First Order Number Theory and Certain Higher Order Theories*, Berlin: Springer, 1975(Treatise on proof theory. Toledo's book is a 'pseudo-translation' of the first edition, helpfully substituting Beth- Smullyan tableaux for Schütte's

idiosyncratic formulation of basic logic.)
Shoenfield, J. (1959) 'On a Restricted ω-Rule', *Bulletin de l'Académie Polonaise des Sciences* 7: 405- 7.(Result on effective ω-rule.)
Smullyan, R. (1961) *Theory of Formal Systems*, Princeton, NJ: Princeton University Press.(Elegant presentation of basic recursion theory in terms of simple formal systems.)
Tarski, A. (1933) 'Einige Betrachtungen über die Begriffe ω-Widerspruchsfreiheit und der ω-Vollständigkeit', *Monatshefte für Mathematik und Physik* 40: 97-112; trans. 'Some Observations on the Concepts of ω-Consistency and ω-Completeness', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J.H. Woodger, Oxford: Oxford University Press, 1956.(Introduces the ω-rule.)

Non-monotonic logic

A relation of inference is 'monotonic' if the addition of premises does not undermine previously reached conclusions; otherwise the relation is non-monotonic. Deductive inference, at least according to the canons of classical logic, is monotonic: if a conclusion is reached on the basis of a certain set of premises, then that conclusion still holds if more premises are added.

By contrast, everyday reasoning is mostly non-monotonic because it involves risk: we jump to conclusions from deductively insufficient premises. We know when it is worthwhile or even necessary (for example, in medical diagnosis) to take the risk. Yet we are also aware that such inference is 'defeasible' - that new information may undermine old conclusions. Various kinds of defeasible but remarkably successful inference have traditionally captured the attention of philosophers (theories of induction, Peirce's theory of abduction, inference to the best explanation, and so on). More recently logicians have begun to approach the phenomenon from a formal point of view. The result is a large body of theories at the interface of philosophy, logic and artificial intelligence.

1 Sources of non-monotonicity

A relation of inference is 'monotonic' if the addition of premises does not undermine previously reached conclusions; otherwise the relation is non-monotonic. Theories of non-monotonic inference locate the source of non-monotonicity in various forms of calculated risk-taking in the course of seeking information. Many kinds of reasoning involving 'degrees of belief' provide familiar examples. Suppose that a set *A* of assumptions raises the credibility of an event described by a sentence *b* above a certain threshold. (Lower case letters will stand for sentences in a given language, upper case letters denote sets of sentences.) Then it is quite possible that the credibility of *b* falls below that threshold if further assumptions are added to *A*. Thus, while one may be prepared to accept *b* on the assumptions *A*, one may prefer to remain agnostic about *b* given additional information. Credibility here may be measured by ordinary probability or by more general measures, such as Shackle's (1979) non-additive degrees of belief (where the credence of mutually exclusive and jointly exhaustive alternatives need not add up to unity). Indeed, most advocates of non-monotonic formalisms insist that they model a wider class of phenomena than can be captured by ordinary probability functions.

A particularly simple source of non-monotonicity is the 'closed world assumption' (CWA). With respect to a given collection of data about some subject matter it states that the collection is complete; that no relevant information is left out. Hence, if *a* is not part of the collection, it is safe to assume $\neg a$. Since *a* may be derivable from a larger set of data, derivability under the CWA is non-monotonic. The CWA is behind the treatment of 'negation as failure' in Prolog ('Programming in logic', a programming language): Prolog derives $\neg a$ from a set of data *D* just in case a query for *a* given *D* fails. (Caution: *a* must be of a suitably simple syntactic kind. For complex sentences, negation as failure is a delicate and potentially hazardous matter.)

Another source of non-monotonicity is the licence to draw from a pool of presumptions ('defaults') as long as these are 'safe' to use (that is, consistent). For example, if one is told that tea is served, one will naturally act on the presumption that it is served plain or with milk and/or sugar. That presumption may become unsafe as additional information is supplied (such as that tea is served with brake fluid or salt).

Default logic (Reiter 1980) is one of the first and most influential theories responding to this source of non-monotonicity. It serves well as a principal example of a non-monotonic logic because it exhibits very clearly three characteristic features of almost all theories of non-monotonic reasoning: first, closure under rules with premises that require tests for non-derivability ('non-Horn' rules); second, a non-unique process of extending premises using defaults; and third, some method for reducing the multitude of extensions and for specifying a univocal set of conclusions.

A 'default theory' (D, A) (giving rise to a default logic) consists of a set *A* of premises (or 'facts') and a set *D* of default rules ('defaults' for short), each of the form

(1) $(B : C) \Rightarrow a.$

These rules are to be read as follows: infer *a* from *B* unless *C* is inconsistent with what can be known by default on

the basis of *A*.

A similar and, under certain conditions, equivalent way of expressing defaults is by modal formulas such as

(1′) $\Box b \wedge \neg\Box\neg c \rightarrow a,$

as in Moore's (1985) auto-epistemic logic, where $\Box a$ may be read '*a* is known'.

The rule (1) is non-Horn, since the 'unless' clause must be verified by showing that no contradiction can be derived from *C* together with the closure of *A* under the default rules.

Closing the premises *A* under the default rules in *D* issues in an extension of the default theory (D, A). More precisely, let $\Sigma = (S_0, S_1, \ldots)$ be a sequence of sets of formulas with $S_0 = A$, let *S* be the union of all its members, and let each S_i in Σ be such that

$$\begin{aligned} S_{i+1} = \{a\text{: there is a default } (B : C) \Rightarrow a \text{ in } D,\\ B \subseteq Cn(S_i) \text{ and}\\ C \text{ is consistent with } S\} \end{aligned}$$

(where *Cn* denotes the operation of generating the deductive consequence of a given set). For each such sequence Σ, $Cn(S)$ is called an 'extension' of the default theory (D, A).

To see how a single default theory (D, A) gives rise to a multitude of extensions, let $A = \{a\}$ and consider a set *D* of two defaults,

(2) $(a : \neg b) \Rightarrow c$
(3) $(a : \neg c) \Rightarrow b$

If we first close *A* under (2), we obtain a set A_1 containing *c*, so *b* cannot be inferred using (3). If we first close *A* under (3) we obtain a set A_2 containing *b*, so *c* cannot be inferred using (2). Thus, the result of extending a set of premises ('facts') by using default rules depends on the order in which the rules are applied. In general there are many extensions of a set of facts under default rules; sometimes there are none (Poole 1994: §7).

The multiplicity can be resolved by intersecting a set of selected extensions. There are three basic options for selecting: according to the 'sceptical' (full meet) solution, all extensions should be selected; the 'brave' (maxi-choice) approach recommends selecting a single extension; and the 'cautious' (partial meet) approach is less committal in requiring only that some extensions be selected. In each case a set of premises *A* is said to entail a conclusion *b* by default *D* just in case *b* follows deductively from the intersection of all selected extensions of *A* under *D*. How should one select among possible extensions? Some approaches recommend that the basis should be syntactic information about the applied defaults: for example, that extensions generated by more specific defaults should take precedence over those generated by more high-handed ones. Others assume that defaults are to be ordered, either by a qualitative preference ranking or by a numerical assignment of degrees of belief.

2 Patterns and models of non-monotonic inference

Inference is 'defeasible' if new information may undermine old conclusions. Formal theories of non-monotonic reasoning differ greatly both as to the kind of defeasibility they focus on and as to their chosen level of generality. They also differ, at first sight, from what are typically referred to as 'logics'. Still, the term logic is appropriate because all these theories can be used to specify a relation in some formal language which exhibits many of the patterns usually associated with relations of inference.

It is not enough to characterize defeasible inference negatively by the possible failure of monotonicity. We also need to ask: which conditions are compatible with defeasibility and which conditions should be satisfied by any relation of inference? These questions point towards continuing the abstract study of consequence relations (or operations) initiated by Tarski (1930). By abstracting from the particular ways in which inference relations are generated we turn to properties they may share.

Let $Cn(A)$ stand for the set of classical and $C(A)$ for the set of defeasible consequences of some set *A*. Among the more stable properties for *C* that have emerged are the following:

to A$\subseteq C(A)$(inclusion) $C(A) = C(C(A))$(idempotence) If $A \subseteq B \subseteq C(A)$ then $C(A) = C(B)$(cumulativity) $Cn(A) \subseteq C(A)$(supraclassicality) $C(A) \cap C(B) \subseteq C(Cn(A) \cap Cn(B))$(distribution)

In the presence of the other conditions, cumulativity is equivalent to the slightly more perspicuous: from $A \subseteq C(B)$ infer $C(A \cup B) = C(A)$. The properties of *C* may be recast in terms of a relation $A \mid\!\sim b$, defined to mean $b \in C(A)$. For example, cumulativity combines the classical principles of cut,

$$\text{If } (\forall b \in B) A \mid\!\sim b \text{ and } A \cup B \mid\!\sim c, \text{ then } A \mid\!\sim c (\text{cut})$$

with a cautious form of monotonicity,

$$\text{If } (\forall b \in B) A \mid\!\sim b \text{ and } A \mid\!\sim c, \text{ then } A \cup B \mid\!\sim c. (\text{cautious monotonicity})$$

Inference operations (and their associated relations) satisfying the first three conditions are called ‘cumulative’; if they also satisfy the last two conditions, they are called ‘classically cumulative’. Note that although the above list describes the common core of various theories of non-monotonic inference, theories based on high probability rules for acceptance are not among them. If $A \mid\!\sim b$ is defined to hold just in case the probability of *b* given all members of *A* reaches a certain threshold, then $\mid\!\sim$ satisfies neither cut nor cautious monotonicity, nor can $\mid\!\sim$ be both idempotent and supraclassical; for that would entail the objectionable left absorption principle $C = CnC$. (Objectionable because of ‘lottery paradoxes’ where both *a* and *b* are accepted, while their logical consequence $a \wedge b$ is not (see Kyburg 1961).) Still, high probability accounts of $\mid\!\sim$ can be made to score better on the above properties, if the probability required for acceptance is pushed high enough, that is, infinitely close to unity. Such accounts have been developed by Adams (1975) and Pearl (1988).

Cumulative inference can be characterized in terms of *preferential models* (see Shoham 1987; Kraus, Lehmann and Magidor 1990; Makinson 1994). These consist of a set of ‘states’ ordered by some relation $<$ and a mapping $|\ |$ assigning to each sentence a set of states. Intuitively, the mapping takes a sentence to a state just in case the sentence truly describes the state, and the relation $<$ orders states with respect to their ‘extravagance’ (conversely, ‘naturalness’). For example, a state in which tea is served with a dash of brake fluid is less natural (more extravagant) than a state in which tea is served with salt, which in turn is still less natural than serving tea with sugar or with milk or with both. The most natural states are the states we can expect to obtain unless there is evidence to the contrary.

If $|a|$ is the set of all states at which *a* holds, then the minimal elements of $|a|$ under $<$ represent the most natural (‘preferred’) ways for *a* to be true. The condition of ‘stopperedness’ (sometimes called ‘smoothness’; see also the ‘limit’ assumption in Lewis 1973) requires that there always are minimal elements.

For each set *A* of sentences, let $|A|_<$ denote the set of minimal states at which all sentences in *A* hold. Then *A* *preferentially* entails *b* just in case $|A|_< \subseteq |b|$. In contrast to ordinary entailment, preferential entailment restricts the domain of quantification: not all possible but only the most natural ways in which all elements of *A* can be true are considered. Since a most natural way for *A* to be true may not be a most natural way for $A \cup B$ to be true, preferential entailment is not monotonic. To illustrate: we may find that all most natural ways of making the sentence ‘tea is served’ true will please someone in the room. But that finding may well be undermined, if we move to considering all ‘natural’ ways of serving tea with brake fluid.

It can be shown that (stoppered) preferential entailment is a cumulative inference relation and that every cumulative inference relation can be presented as a relation of preferential entailment. (For the representation results mentioned here see Makinson 1994: §3.4.) The condition of stopperedness guarantees that preferential entailment will be cautiously monotonic. (To see how cautious monotonicity may fail in non-stoppered models let the states be indexed by the natural numbers and let s_m be more natural than s_n just in case *m* is greater than *n*. Consider three sentences a, b, c such that *a* holds everywhere, *c* fails everywhere and *b* holds at the first state only. Then $|a|_< = \emptyset$ because there is no most natural state for *a*. So, on the one hand we have both $|a|_< \subseteq |b|$ and $|a|_< \subseteq |c|$ vacuously. On the other hand, there is exactly one state - namely the first - that satisfies both *a* and *b*. Hence, $|\{a, b\}|_<$ - having the first state as its (only) member - cannot be contained in $|c|$ since the latter was assumed to be empty.)

If the mapping $|\ |$ is classically well-behaved (that is, satisfies the equations $|a \wedge b| = |a| \cap |b|$ and $|\neg a| = \overline{|a|}$), then each state will be closed under classical consequence whence preferential entailment will be supraclassical. Moreover, it will also be distributive and thus satisfy the conditions for being classically cumulative. Conversely,

each classically cumulative inference can be represented as preferential entailment in preferential models that respect the truth-functional connectives.

Supraclassicality is a rather common property of non-monotonic reasoning. It gives formal expression to the idea that such reasoning goes beyond what can be inferred 'safely' (deductively) from a given body of knowledge. However, some theories of non-monotonic inference (like those based on defeasible inheritance nets or on reason maintenance systems) fail to fully extend classical logic because they operate in impoverished fragments of classical languages. Such restrictions are usually introduced for computational reasons; they are not based on principled objections to supraclassicality.

Note also that in so far as non-monotonic logics are supraclassical, they are not rivals but extensions of classical logic. This sets them apart from certain other logics which fail to be monotonic, such as relevance, linear and many paraconsistent logics. These logics are *sub*classical and their source of non-monotonicity is different, having nothing to do with 'risk-taking'.

3 Related work

A sentence *b* may be (defeasibly) inferred from a set *A* in the light of background assumptions *B* just in case *b* follows logically from *A* together with as much of *B* as is compatible with *A*. The right-hand side of this equivalence can be seen as involving a step of belief change: from the background beliefs *B* to a *revision* of *B* so as to accommodate *A* (with a minimal amount of change). Belief revision and defeasible inference thus appear to be two sides of the same coin.

The correspondence between belief revision and non-monotonic reasoning has been investigated in depth by Gärdenfors and Makinson (1994). In contrast, Levi (1995) argues that the correspondence can only be partial because the sole purpose of belief revision is to accommodate belief-contravening evidence. Such accommodation obeys a maxim of minimal change. But in so far as non-monotonic reasoning also covers ampliative inference, it must go beyond the maxim of minimal change governing belief revision. Hence, according to Levi, properties of non-monotonic inference as generated by belief revision on the one hand and as generated by ampliative inference on the other hand cannot be expected to coincide.

There is a close relation between counterfactual conditionals and belief revision via the 'Ramsey test': accept a conditional 'If *a* were the case, then *b* would be the case' in a state of belief *A* just in case you would accept *b* if *A* had to be revised so as to contain *a*. Given the correspondence, noted in the last paragraph, between belief revision and non-monotonic reasoning, one expects - and will not be disappointed - to find similarities between counterfactuals and non-monotonic reasoning.

A more direct route to such similarities is opened by comparing preferential models with the Stalnaker-Lewis semantics for counterfactuals (see Counterfactual conditionals §§3-4). In this semantics a counterfactual is pronounced true in a 'world' *w* just in case the consequent is true in all the closest possible worlds (that is, those worlds which depart as little as possible from *w*) that make the antecedent true. There is an obvious similarity between the condition for evaluating preferential entailments and that for evaluating counterfactuals: not all antecedent-worlds are examined but only those that are minimal under a certain relation. But there are also differences. Because each model for counterfactuals is equipped with as many relations of closeness as there are worlds to be close to, counterfactuals can be evaluated (differently) at single worlds. By contrast, preferential models have only one relation; accordingly, preferential entailments are assessed not with respect to single worlds but only in a model as a whole. It is a controversial matter how to interpret such subtle differences (see Makinson 1993): whether preferential entailments are essentially valid counterfactuals frozen at the first degree (that is, with the conditional connective joining truth-functional formulas only) or whether the differences are more substantial.

See also: Logical and mathematical terms, glossary of

ANDRÉ FUHRMANN

References and further reading

Adams, E. (1975) *The Logic of Conditionals*, Dordrecht: Reidel.(Derives a relation of defeasible inference (*p*-entailment) from limiting probability valuations.)

Brewka, G. (1991) *Non-Monotonic Reasoning: Logical Foundations of Commonsense*, Cambridge: Cambridge

University Press.(A concise introduction to the topic.)
Gabbay, D.M., Hogger, C.J. and Robinson, J.A. (eds) (1994) *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, *Non-Monotonic Reasoning*, Oxford: Oxford University Press.(A collection of authoritative surveys.)
Gärdenfors, P. and Makinson, D. (1994) 'Non-Monotonic Inference Based on Expectations', *Artificial Intelligence* 65: 197-245.(Investigates correspondences between belief change and non-monotonic reasoning.)
Kraus, S., Lehmann, D. and Magidor, M. (1990) 'Non-Monotonic Reasoning, Preferential Models and Cumulative Logics', *Artificial Intelligence* 44: 167-207.(A study of models for non-monotonic inference relations extending the approach of Shoham (1987).)
Kyburg, H. (1961) *Probability and the Logic of Rational Belief*, Middleton, CT: Wesleyan University Press.(An early treatise of defeasible inference as in high probability accounts of acceptance. See also Kyburg's contribution to Gabbay *et al*. (1994).)
Levi, I. (1995) *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference and Non-monotonic Reasoning*, Cambridge: Cambridge University Press.(Critically examines non-monotonic logic in the light of philosophical studies of inductive inference and rational belief change.)
Lewis, D.K. (1973) *Counterfactuals*, Oxford: Blackwell.(Seminal study of models for counterfactual conditionals.)
Makinson, D. (1993) 'Five Faces of Minimality', *Studia Logica* 52: 339-79.(Compares various ways of using the notion of 'minimality' in semantics for non-monotonic logic, semantics for conditionals and for conditional obligation, belief change and updating.)
Makinson, D. (1994) 'General Patterns in Non-Monotonic Reasoning', in D.M. Gabbay, C.J. Hogger and J.A. Robinson (eds) *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, *Non-Monotonic Reasoning*, Oxford: Oxford University Press.(Focuses on the material of §2.)
Moore, R.C. (1985) 'Semantical Considerations on Non-Monotonic Logic', *Artificial Intelligence* 25: 75-94. (Develops the modal version of default logic.)
Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems*, San Mateo, CA: Morgan Kaufmann.(Derives non-monotonic inference from limiting probability valuations. Rediscovers and further develops Adams' (1975) theory of *p*-entailment (under the name of ε-entailment).)
Poole, D. (1994) 'Default Logic', in D.M. Gabbay, C.J. Hogger and J.A. Robinson (eds) *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, *Non-Monotonic Reasoning*, Oxford: Oxford University Press.(A survey of default logics.)
Reiter, R. (1980) 'A Logic for Default Reasoning', *Artificial Intelligence* 13: 81-132.(Classic presentation of default logic.)
Shackle, G.L.S. (1979) *Imagination and the Nature of Choice*, Edinburgh: Edinburgh University Press. (Shackle's writings have informally anticipated many of the models of non-monotonic reasoning.)
Shoham, Y. (1987) 'A Semantical Approach to Non-Monotonic Logics', in *Readings on Non-Monotonic Reasoning*, ed. M. Ginsberg, Los Altos, CA: Morgan Kaufmann.(First presentation of preferential models.)
Tarski, A. (1930) 'Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften I', *Monatshefte für Mathematik und Physik* 37: 361-404; trans. J.H. Woodger (1956), 'Fundamental Concepts of the Methodology of Deductive Sciences', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 60-109.(Classic study of abstract consequence operations.)
Ullman-Margalit, E. (1973) 'On Presumptions', *Journal of Philosophy* 80: 143-63.(Develops informally many of the ideas underlying formal theories of non-probabilistic non-monotonic reasoning.)

Normative epistemology

There are three kinds of normative work in epistemology. The first is the provision of epistemic advice, which offers guidance towards improving the cognitive condition of an individual or community. This advice often concerns science. Philosophers in the tradition of Francis Bacon have sought to identify and advocate proper forms of scientific research and explanation. More generally, according to some philosophers, a principal epistemological task is that of finding and recommending ways to improve the whole range of our individual and collective cognitive activities.

A second kind of epistemology is classified as normative because evaluative concepts figure in explanations. For example, A.J. Ayer explains knowledge partly in terms of having a right to be sure. Other evaluative notions enter into work in this category, such as intellectual duties, responsibilities and virtues. Some of these are specifically ethical notions; some are non-ethical evaluative notions such as proper cognitive functioning and intellectual excellence.

Epistemic concepts such as justification and rationality appear to be normative, or at least evaluative, in a way that contrasts with purely desciptive concepts. One tendency in naturalistic epistemology is to seek either to explain away this appearance or to reconcile it with a scientific worldview. Non-naturalistic efforts in epistemology commonly find no reason to undertake this project, and are consequently often counted as normative. Most historical epistemology is normative by this standard.

1 Advisory epistemology

Francis Bacon and René Descartes are among the prominent historical philosophers who advocated methods of cognitive improvement. Bacon (1620) defended methodological departures from the scholastic intellectual traditions that prevailed in Europe in his time. The investigative practices that Bacon recommended constituted early modern scientific methodology (see Bacon, F. §§4-8).

Descartes (1701) elaborated a system of rules to guide inquiry. His most characteristic enterprise was a search for certainty in his beliefs. He advocated a method of doubt for achieving certainty. In Descartes' view, beliefs that withstand the strongest reasons for doubt are certain. Such beliefs are distinguished by being clearly and distinctly perceived. His rules are designed to facilitate achieving this (see Descartes, R. §4).

In the twentieth century, Alvin Goldman among others has supported advisory projects for epistemologists. Goldman (1986) has championed 'epistemics': an alliance between epistemology and cognitive psychology. Epistemologists contribute to the joint effort by identifying epistemically desirable magnitudes that apply to belief formation process types, such as the truth-to-falsehood ratio of the doxastic output of a process type. Psychologists determine the degree to which actual psychological processes exemplify these magnitudes. They also contribute to epistemics by studying how, and to what extent, our cognitive capacities allow the magnitudes to be exemplified to other degrees. In light of such psychological findings epistemologists engaged in epistemics propose strategies for effecting cognitive improvements.

Advisory epistemology has problematic characteristics. For one thing, philosophical thought does not seem to constitute a suitable preparation for recommending successful ways to conduct empirical science. Effective ways to learn the contingent facts that are the subject matter of science seem themselves to be matters of contingent empirical fact. Bacon and other philosophers of science may well have happened to acquire the empirical information useful in making helpful recommendations in this area. But the actual efficacy of any given empirical research strategy seems to be an empirical question rather than a philosophical one.

More fundamental difficulties affect epistemological advice on all topics. These problems concern whether there is some distinctive epistemic standard, and whether there is any general justification for recommending cognitive conditions or processes that have some positive epistemic status. Epistemic advice seems to include identifying and asserting effective techniques for gaining various epistemically favourable outcomes, such as reasonable beliefs, justified beliefs, true beliefs, and knowledge. This multiplicity of positive potential outcomes does not always coincide, and that leads to the first problem. Note that it might be feasible to get someone to believe some difficult truth for a bad reason and feasible to give them good reason to believe the negation of that truth, but not

feasible to get them to believe the truth for a good reason. When there is some such competition between apparently valuable epistemic states, what is ‘epistemically advisable’? A single overriding ranking of epistemic outcomes would answer this sort of question. But does such a ranking exist, and if so, why does it override?

The other fundamental problem concerns justification for issuing any epistemic advice. Is some such advice somehow justified even when the advised outcome is not sought and is not an effective means for attaining anything that is sought? For instance, suppose that a friend is interested in forming a belief about what fraction of the inhabitants of Tibet are Buddhists. It might seem to be sensible epistemic advice to recommend forming this belief on the basis of a large and randomly selected sample. Believing on this sort of basis may well give such beliefs a relatively high truth-to-falsehood ratio. Does this make that basis ‘epistemically advisable’? Suppose that a friend wants to form a fanciful belief on the topic, with no concern for its truth. If so, what justification would there be for offering the hypothesized advice? Or suppose that the friend seeks certainty and nothing less. The advised procedure will not yield that. Would the same recommendation be apposite, and if so, why?

Recommendations can be justified in various ways. Advice can be justified for the sake of meeting several sorts of standards, such as moral ones, legal ones and goal-oriented ones. So various beliefs and inferences might be morally, legally or instrumentally advisable. But there seems not to be any unified notion of ‘cognitive betterment’ that could provide some distinctively epistemic standard justifying the issuing of doxastic advice (see Scientific method).

2 Evaluative epistemology

In a tradition which is traceable to John Locke, some philosophers have theorized about epistemic notions in plainly evaluative terms. One line of thinking, directly initiated by Locke and recently represented by Roderick Chisholm, appeals to intellectual duty. Another approach, stemming from Cartesian and Lockean origins and present in recent work by Laurence BonJour (1985) and Hilary Kornblith (1980), explains a kind of epistemic justification for belief in terms of responsible doxastic conduct.

Criticism has been offered of requiring any such duty fulfilment or responsible conduct in order for believing to have some positive epistemic status. Some objections are based on the assumption that epistemic duties or responsibilities are limited by our ability to comply with them. A common sort of objection adds the premise that we seldom if ever have control over what we believe. These assumptions together imply that we seldom if ever have such duties or responsibilities to believe otherwise than we actually do. Yet believing otherwise frequently does have some positive epistemic status. Thus dutiful or responsible doxastic conduct is not required for having this status.

There is a reasonable reply to this criticism. It is doubtful that all duties are limited by the ability to comply with them. Duties from various sources seem to exceed this limit. Political entities appear to impose legal duties beyond the abilities of their citizens always to comply. The duties taken on by agreeing to perform a job with certain assigned responsibilities may turn out not to be feasible to accomplish. So the general principle that duties can always be met seems unacceptable. Even within the sphere of purely ethical duties, it has been credibly argued that obligations can exceed abilities. Consequently, the principle that epistemic duties in particular are limited by ability stands in need of some special defence.

Alvin Plantinga (1993) has offered a different criticism of a reliance on intellectual duties in a theory of knowledge. Plantinga’s topic is a theory of warrant. ‘Warrant’ is his term for a magnitude the bearing of enough of which by a true belief constitutes a case of knowledge. Plantinga objects to the thought, derived from work by Chisholm, that believing a proposition as a result of trying one’s best to fulfil one’s epistemic duties is sufficient for being warranted in the belief to some significant extent. Plantinga describes the following case. Paul is given by a malicious demon a nearly overwhelming tendency to believe that whenever it sounds to him as though a church bell is ringing what he is hearing is orange in colour. All who are around Paul have been given this same tendency. Plantinga contends that, by acceding to this strong inclination when hearing a church bell sound and believing that what he hears is orange, Paul would fulfil his epistemic duties. Yet this belief would have little or no warrant for Paul. Plantinga concludes that complying with intellectual duty is not sufficient for being significantly warranted in the resulting belief.

There is a promising reply to this objection. It can be maintained that having even a nearly overwhelming tendency

to believe something does not generate an intellectual duty to believe it. What intellectual duties require is conformity to epistemic reason. A sound and a colour are conspicuously diverse. In Plantinga's example Paul is ascribed no *reason* for thinking that the two go together. So, intellectual duty calls for Paul to refrain from forming the colour belief. He would thereby respond rationally to the manifest lack of any evidential connection with the sound. He has no epistemic ground, and hence no duty, to give in here to a powerful but unreasonable impulse. Thus, this sort of case does not seem to pose an insuperable problem for the claim in question that complying with intellectual duty is sufficient for having significant warrant.

3 Non-naturalistic epistemology

Naturalism in epistemology has no widely accepted definition. Naturalistic epistemologists variously ally themselves with science in ontology or method. They tend to seek explanations of epistemic facts that employ properties with actually or prospectively good scientific credentials, such as causal relations and statistical probabilities, and they tend to classify the epistemological claims that they make as empirical rather than a priori (see Naturalized epistemology). Sometimes it is counted as adequately naturalistic to postulate entities that have some suitable sort of ontological dependence on physical reality, whether or not the postulated entities themselves play a role in physical science (see Supervenience).

It is notoriously difficult to see how ethical features such as intrinsic goodness and moral obligation can be part of the sort of reality that science describes. Like such ethical evaluations, some epistemological categories seem to resist assimilation to a scientific worldview. Several of the main epistemological characterizations, such as 'justified', 'rational' and 'reasonable', seem to make comparably evaluative appraisals of the mental states or processes to which they are applied. It is comparably unclear how such epistemic attributes of belief and reasoning fit into the world of natural science. Naturalistic epistemologists characteristically attempt to clarify this fit or eliminate the recalcitrant epistemic features.

Other epistemologists do not make this attempt. At one non-naturalistic extreme there is work like that of Alvin Plantinga, who argues that God plays a role in the nature of warrant. Plantinga conceives of warrant in terms of properly functioning cognitive mechanisms. He argues that only the intentional design of these mechanisms by God is adequate to explain their proper function. This sort of view relies on a defiantly non-naturalistic ontology.

Many non-naturalist epistemologists simply theorize in terms of whatever concepts they find adequate to their theoretical tasks, without paying heed to naturalistic constraints. For instance, a coherence relation among propositions figures in some theories of epistemic justification, and often no effort is made to square the existence of this coherence relation with a scientific ontology (see Knowledge and justification, coherence theory of).

Non-naturalistic methodology is common in epistemology. Explicit appeals are made to a priori considerations in defence of epistemological theses, typically by claiming support by appeal to intuitive judgments about hypothetical cases. This methodological departure from natural science marks the work as non-naturalistic.

Some epistemologists thus do not abide by the ontological and methodological constraints that are characteristic of naturalism. This is unproblematic to the extent that these constraints are inadequately defended. Notably, they seem to have no credible defence when applied to the analogous field of pure mathematics. To ban from mathematical research a priori considerations or ontological commitments beyond the foreseeable domain of natural science would abolish the discipline. This would be unwise. Unconstrained mathematics efforts yield some of our clearest and most secure knowledge.

It seems similarly unwise to burden epistemological inquiry with naturalistic constraints. Much non-naturalistic epistemological work is relevantly similar in method and subject matter to mathematical work that yields knowledge. Although established results are rare in epistemology, the example of mathematics shows that we have good reason not to blame the paucity of such success on a violation of naturalistic constraints.

See also: Epistemology and ethics

EARL CONEE

References and further reading

Alston, W. (1985) 'Concepts of Epistemic Justification', *The Monist* 68 (1): 57-89. (Argues against theories of

justification that appeal to intellectual duty.)
Ayer, A.J. (1956) *The Problem of Knowledge*, London: Macmillan.(Employs the evaluative primitive of having a right to be sure in an analysis of knowledge. See especially chapter 1.)
Bacon, F. (1620) *Novum Organum*, trans. J. Spedding, R.L. Ellis and D.D. Heath as *The New Organon*, ed. Fulton Anderson, Indianapolis, IN: Bobbs-Merrill, 1960. (Advocates an early form of modern scientific methodology.)
BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.(Widely discussed presentation of a coherentist theory of justification, where justification is said to be ultimately determined by a sort of intellectual responsibility.)
Chisholm, R. (1966) *Theory of Knowledge*, New York: Prentice Hall, 2nd edn, 1977.(Classic effort in analytical epistemology, one that makes irreducible use of a notion of intellectual duty.)
Descartes, R. (1701) *Rules for the Direction of the Mind*, trans. and ed. E.S. Haldane and G.R.T. Ross, London: Cambridge University Press, 1967.(Recommends rules for discovering the truth with clarity and distinctness.)
Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(Contains an extended exposition and defence of the partially advisory discipline of epistemics.)
Kornblith, H. (1983) 'Justified Belief and Epistemically Responsible Action', *The Philosophical Review* 92: 33-48.(Explains epistemic justification for belief in terms of responsible conduct.)
Locke, J. (1689) *Essay Concerning Human Understanding*, ed. A.C. Fraser, New York: Dover, 1959.(Contains a classic statement of empiricist evaluative epistemology.)
Plantinga, A. (1993) *Warrant: The Current Debate*, New York: Oxford University Press.(Contains a useful survey and critique of literature on epistemic justification, and includes a brief sketch of Plantinga's proper functionalist theory of warrant.)
Stich, S. (1990) *The Fragmentation of Reason*, Cambridge, MA: MIT Press. (Attacks analytic epistemology and advocates a pragmatic sort of advisory epistemology.)

Norms, legal

A legal norm sets a standard of behaviour. As a norm, it thus can remain in existence even though it is broken. Norms can be distinguished from causal laws which need to be reinterpreted if an exception is found. Linguistic signals help us determine what the norm is. Thus 'ought', 'must', 'shall', 'have to', 'right', 'wrong', 'good', 'bad', and so on, characteristically belong to the statement of norms, whereas words like 'is', 'are', 'were', 'will be' 'possible', 'impossible' tend to show descriptive rules. These linguistic signals reflect a difference, they do not constitute it. There are many counterexamples: thus 'swimming is forbidden' and 'we ought to be at the col now' express normativity and description respectively.

Whatever is for someone a standard for their conduct is normative for them. One might say that the idea stems from the notion of measurement. That we 'run the rule' over someone or 'get the measure of them' stems from the idea of measuring, of imposing a standard on them or on oneself.

Where does the legal norm stem from? There are two main views: the practice theory holds that norms are expressions or articulations of people's behaviour; the interpretive theory holds that norms are not connected to behaviour in the way the practice theory holds but are the means whereby we make sense of such behaviour. But the connection between the two groups is much closer than appears at first sight.

1 Normativity and description

The law prohibiting murder seems both different from, and more mysterious than, the law of gravity. Why? One can break the former and it will still be valid and in existence but we still have to follow it. There is still something that constrains us, that binds us. The fact that the murder rate has gone up will not be taken as evidence that it does not exist. However, if objects started to float away into space then the law of gravity would have to be revised, an exception to the scientific law must modify it. The law cannot be broken.

We can explain this by looking at the difference between 'doing something as a rule' and 'having a rule about something', as Hart posed it (1961). Take the behaviour of a group of cars at a traffic intersection. We could say that they stop at the red traffic lights 'as a rule'. This would imply that there is a certain regularity of behaviour but that if there is a departure from it this will be merely something unusual, a departure from the average. Scientific laws stem from this sort of observation. They generalize into a rule what normally happens, what people do 'as a rule'. This will take the form of some causal explanation about the nature of the world. But we can give another explanation. We may say that the cars' behaviour is governed by the Road Traffic Act. There is a difference between that law and the causal one we observed earlier. The former is provisional. If we find that cars do not stop at red lights any more that means that our causal law is invalid and we have to amend it. Whereas the fact that cars do not stop at red traffic lights does not make the Road Traffic Act invalid. We ought to obey even if we do not. We are not looking at the generalization of 'as a rule' behaviour. Rather we are looking at a standard of behaviour, departure from which would not just be seen as unusual but as a reason for criticism. Here there is a prescription as to what behaviour is required. In the first case there was not. The former is a normative law while the causal law is descriptive.

We now turn to asking what the legal norm is and how we can recognize it. This is closely tied up with where normativity itself comes from. The question as to whether something is a religious norm, a moral norm or a legal norm is mainly a formal question. Thus, as we will see below, the basic question when looking at the legal norm is to understand its normativity.

2 The practice theory

We may first of all think of the norm as stemming from simple iteration; that is, what is done once will be done again, and will through repetition induce expectation, in this way producing normativity. Thus, *per* Hayek (1969), normativity comes from custom which is generalized into abstract and general rules. This is the best way because it reflects that what people should have as standards of behaviour is what over a long period of time they have done. What is important here is the iteration. Something happening over and over again is what imparts normativity. The legal norm is the abstraction of that.

Hayek's theory is a modern version of older theories that stressed custom as the source of law. John Austin thought that custom missed out the element of the will: the legal norm was a species of command. The norm is the imposition of the will of one person, who sets the standard of behaviour, on another. Thus Austin (1971) gives a definition of law (the legal norm) as the command of a sovereign backed by a sanction. Laws are thus a species of commands, directed by sovereigns to their subjects. They are an exercise of political will or power and necessarily involve the imposition of sanctions in the case of disobedience. But, behind obedience to this will, there re-enter descriptive regularities of custom and usage in the shape of what Austin calls habit.

Austin defines the sovereign as the 'political superior' towards whom the population exhibits a 'habit of obedience'. As a political superior the sovereign's power and authority is based in social practices and in the social fact of habitual obedience - the social source of law. The normativity of law, one might say, comes ultimately from the barrel of the gun which people habitually obey (see Sovereignty §2).

H.L.A. Hart is more complicated. He wants to reject the command theory but still claim that normativity comes from social practice. For him norms are not merely commands but can be seen as rules. The key is in his famous distinction between being obliged and having an obligation. Austin's version of the legal norm misses the idea of obligation. It can be best likened to the gunman who demands money from you. You are obliged to give it to him, but only as long as he is pointing the gun at you. The legal norm lasts after the gunman has gone. It is something more than merely a habit of obedience to some commander. Habits refer only to the external regularities of behaviour. By contrast there is an internal aspect in the case of rules of conduct shared by members of a social group. This depends upon the attitude (a critical reflective attitude) of insiders of the group. For them a certain regular pattern of behaviour is not merely observable regularity. It is a standard for the criticism of conduct which contravenes the pattern in question. This is expressed in the idea of obligation. This comes both from people's behaviour and from their attitude. To say that someone has an obligation implies two statements, a cognitive statement that there are observed regularities and a volitional one that people have the critically reflective attitude to that regularity.

The above are examples of a general way of seeing the legal norm that can be called the practice theory; the norm must be inextricably connected to real action in the world. Normativity then stems from people's practices. But this idea has also been criticized. One can always intelligibly ask of a practice whether one ought to follow it and therefore the norm cannot be constituted by the practice. For Hart this does not make sense. If there is a critically reflective practice it would be unintelligible for someone within that to ask whether they ought to follow it. It would be like asking if the standard metre bar in Paris (by which the metre used to be defined) is really a metre. However, for Kelsen (1960) it would make sense because, in following this practice, we assume that is how a metre ought to be defined. And this explains why for Kelsen the *Grundnorm* (the norm from which all other norms in the system get validity) has to be presupposed to be valid while the equivalent for Hart (the ultimate rule of recognition) just is.

3 The norm as interpretation

For Kelsen, then, the legal norm is the meaning of an act of will. It is not the police waving the car down but the meaning of the police's act as saying that 'you ought, by the authority vested in me by the Road Traffic Act, to stop' that is the legal norm. The content of the norm must not be confused with its physical expression. But behaviour still has some importance in the sense that one can only derive a legal norm in the context of efficacious behaviour. It only makes sense to understand the physical activities that constitute the parliamentary process as law-making behaviour when people by and large do treat them as such and obey them.

In separating the norm more and more from behaviour in the physical world, we can see the legal norm as a standard of judgment. Thus Dworkin (1986) attacks the rule-based theory of law and sees principles and latterly rights as being the key to the legal norm (see Dworkin, R.). But where do they come from? In his interpretivist turn we see the return of practice. For the rights are our interpretation of the community's practice. We can thus say that they are the meaning of the community's practice. Now this sounds remarkably like Kelsen's version of the legal norm. For it is the meaning of the practices that is important and that is what gives us the rights. If however we think of the practices as constituting the rights then we see something more like Hart, even though Dworkin, ironically, started his professional life by attacking Hart. The meaning of the practices is locked into the practices that we do and we cannot go beyond them; the legal norm (in the form of rights) comes from our interpretation of

them. For him too it makes no sense for someone inside the practice to question its validity.

This shows the difficulties of making a rigid separation between the normative and the descriptive, at least in the way that the normative is generated. Because, even for Kelsen, it makes no sense to generate a meaning except in the context of by-and-large efficacy. We have a sort of spectrum then. The norm is custom or habit and is thus fully expressed in the practice or, more complicatedly, in a critically reflective practice. Kelsen separates the critically reflective from the practice in order to separate the meaning of the norm from the act of will that generates it. But that separation does not work totally because we can, as Dworkin does, construct the hermeneutic circle and see the meaning as constituting the practice (see Common law).

That one cannot get totally away from practice can also be seen in another critic of the practice theory, Joseph Raz (1979). For him the legal norm is a species of reason for action, specifically an exclusionary or content-independent reason for action. We can explain it in this way. Someone is telephoned late at night by a friend and offered the chance of a speculative but potentially highly lucrative investment. The only drawback is that they have to make up their mind immediately. Knowing he is a bit drunk and tired, and thus not trusting his judgement, what should he do? He could make the decisions on the balance of all the reasons *pro* and all the reasons *contra* including the reason that he is tired and a little drunk. He could, on the other hand, say that he makes it a rule not to take investment decisions late at night. The former way of doing it would be to make the judgment on the balance of reasons. The latter would use the rule ‘do not take investment decisions late at night’ as a reason for excluding consideration of the investment decision on the balance of reasons, even though that decision might be to make the investment. His decision would operate at a second-order level in respect to the reasons of substance as to whether to take the investment decision or not. It would thus be an exclusionary reason or a content-independent reason. We would have a reason for not thinking about the substance of the case because of this second-order reason.

But where does that reason come from? From where is, in Raz’s terms, the ‘legal point of view’ generated? In the end we can find the generation of these only in the institutions of law as they are. We thus see that the standard criticism of natural law theories that they conflate the descriptive with the normative is, in some sense, replicated among other, more positivistically inclined, legal theorists. They both face similar problems - where is normativity generated from?

John Finnis (1980) in his natural law theory sees the legal norm, on Razian lines, as an exclusionary reason. But it is a reason for action because it instantiates the objective values of the universe, more specifically the values of practical reason and human sociability, and it is these that generate legal norms. But how do we know these values? It is not our drives which are the goods. The good of life is not to be reduced to our sexual drive. We are not animals. Rather it is the way in which the drives are transformed through our intelligent understanding that makes them basic values. For example, our drive to survive is not what the good of life is. Survival is not a basic good. Rather it is our understanding of this which enables us to see that we should not survive at all costs and that is not what the good of life means. The problems here are the same as they are in positivist theories.

For Finnis legal norms instantiate moral norms and are *determinationes* of ultimately self-evident values. We can then say that most theories of the legal norm see it as in some way connected to the things that people do. The ’ought’ comes in some way from causal regularities in the world, as bearer, constituter or condition of the norm. So we can see that the norm comes in the end from people’s behaviour or their understanding of it. We ought to do X because in some way we *do* do it. And this appears true even in apparently metaphysical accounts such as those of the natural law. Normativity then is something connected to what people do and not something mysterious at all (see Natural law §3).

But for some there was still a mystery to be explained. The template or measure way of looking at the norm was not the best. More important were metaphors of chain and yoke. What made us do things? What yoked us to rules and legal norms? In Judaism we speak of the ‘yoke’ of the law. This seems a metaphysical question that for the Scandinavian legal realists remained unanswered (see Legal realism §1). Hence they answered it in a naturalistic way: by reducing the norm to our psychological feelings of constraint and being bound, which stemmed from an earlier, more magical age. Thus the magic was reduced to psychological states; the Freudian theory of normativity, as it were.

See also: Law, philosophy of; Legal positivism

ZENON BAŃKOWSKI

References and further reading

Austin, J.L. (1971) *The Province of Jurisprudence Determined*, London: Weidenfeld and Nicolson.(A statement of the law as command theory.)

Dworkin, R. (1986) *Law's Empire*, London: Fontana.(An interpretivist view of law.)

Finnis, J. (1980) *Natural Law and Natural Rights*, Oxford: Clarendon Press.(A natural law theory.)

Hägerström, A. (1953) *An Enquiry into the Nature of Law and Morals*, Stockholm: Almquist and Wiksell.(A psychologistic account from a strong Scandinavian realist perspective.)

Hart, H.L.A. (1961) *The Concept of Law*, 2nd edn, with postscript, P. Bullock and J. Raz (eds), Oxford: Clarendon Press, 1994.(The *locus classicus* in English of the practice theory.)

Hayek, F. (1969) *Law, Legislation and Liberty*, London: Routledge & Kegan Paul.(A customary theory of law.)

Kelsen, H. (1960) *Reine Rechtslehre*, 2nd edn, Vienna: Deuticke; trans. M. Knight, *Pure Theory of Law*, Berkeley, CA: University of California Press, 1967. (The main continental theory of the norm as a scheme of interpretation.)

Raz, J. (1979) *The Authority of Law*, Oxford: Clarendon Press.(A statement of the theory of exclusionary reasons.)

Ross, A. (1968) *Directives and Norms*, London: Routledge & Kegan Paul.(An acute positivist account, from a mild Scandinavian realist perspective.)

Schauer, F. (1991) *Playing by the Rules*, Oxford: Clarendon Press.(An acute positivist account, from a mild Scandinavian realist perspective.)

Shiner, R. (1992) *Norm and Nature*, Oxford: Clarendon Press.(A general account with anti-positivist leanings.)

Norris, John (1657-1711)

John Norris was a philosopher in the Platonic tradition of the seventeenth century. His philosophy combines elements from both the French and English aspects of that tradition: he was an admirer of Henry More and was the leading English disciple of Malebranche, whose philosophy he did much to popularize in England. A churchman by profession, much of his writing is concerned with the practical application of divinity. Central to Norris' thought is his theory that the proper and immediate object of both human knowledge and human love is God, who is identified with truth. Thus necessary truths are known directly in God and, conversely, to know eternal and necessary truth is to know God.

John Norris was born at Collingbourne Kingston in Wiltshire, England. He was educated at Winchester College and Exeter College, Oxford, becoming fellow of All Souls College in 1680. In 1689, he was appointed rector of Newton St Loe. In 1692, through the good offices of Lady Masham and John Locke, he was appointed vicar of Bemerton near Salisbury, where he remained for the rest of his life. Norris' high regard for the Cambridge Platonists is evident from his brief correspondence with Henry More, published as *Letters Philosophical and Moral* and appended to his 1688 *The Theory and Regulation of Love* (see Cambridge Platonism). It is likely that Norris' early Platonist leanings combined with Augustinian strands in his theology to predispose him towards the philosophy of Malebranche. Norris himself denied that his discovery of Malebranche changed his philosophical standpoint (Acworth 1979: 64). None the less the impact of Malebranche is both clearly evident and duly acknowledged in his mature writings. But there were important aspects of Malebranche's philosophy which Norris did not accept, notably Malebranche's account of the distribution of Grace. Nor did he change his view that God acts out of pure benevolence in dealing with created things. It is perhaps truer to say, therefore, that Norris' reading of Malebranche enabled him to articulate his own analogous views more fully and more precisely.

His early writings (later published in *Miscellanies* in 1687) anticipate some of the main themes of his later, Malebranchian phase. For example, his essay 'Of the Advantages of Thinking' argues that ideas and necessary and eternal truths are in God, and are immediately present to the human mind as the proper object of human knowledge. Norris' concern in this essay is with proving the existence of God rather than with epistemology. The arguments are more fully stated in *Reason and Religion* (1689), which propounds Malebranche's theory of the 'vision in God', and are most fully developed in *An Essay towards the Theory of the Ideal or Intelligible World* (1701-4), which he had originally conceived while at Oxford as a *Metaphysica Platonica*. In this, his most substantial book, Norris argues for the existence of an ideal world of intelligible and unchanging entities which serve as the archetypes of the natural world. This ideal world exists in God as the 'omniform Essence' of God. Only the ideal world can be truly known. Since the mind cannot know matter directly, we have no certain knowledge of the physical world. Our knowledge of physical nature derives not from perception of physical bodies, but from our seeing the essences of things in God. It is this Malebranchean idealism which underlies Norris' critique of Locke in his *Cursory Reflections upon a Book Called an Essay Concerning Human Understanding* (1690), where he criticizes Locke for deriving the idea of God from sense knowledge.

Just as God is the proper object of human knowledge, according to Norris, so also, is God the proper object of human love. In *The Theory and Regulation of Love* (1688) Norris distinguishes two kinds of love: love of God and love of things. The former he calls love of concupiscence or desire, that is 'a motion of the Soul towards good'. The latter he calls love of benevolence, or disinterested love. The former is irresistible, uniting the soul with the good, which it lacks, that is, with God. Love of benevolence includes self-love and charity, and is the love most perfectly exemplified by God. Later, in his *Discourse concerning the Measure of Divine Love* (1693) he introduced an important modification to his account of human love of God, invoking Malebranche's occasionalist account of sensation. The proper object of desire is that which causes us pleasure, including pleasurable sensations. Since sensations are sense impressions planted in the mind by God on the occasion of the presence of corporeal objects, God is the cause of our pleasure. Therefore, as the true cause of our pleasure, God is the proper object of the love of concupiscence.

Norris' importance as a philosopher derives from his continuing the Platonic tradition of seventeenth-century English philosophy into the eighteenth century, and his popularizing the philosophy of Malebranche in England.

He also played a role as a mediator of philosophy to a growing lay readership through his books and his association with the journal *The Athenian Mercury*. By writing accessibly on philosophical themes, Norris was probably more influential than can be specifically documented, though particular mention might be made of John Wesley and Arthur Collier. He is also notable for the encouragement he gave to women interested in philosophy, among them Mary Chudleigh, Catharine Cockburn and Mary Astell.

SARAH HUTTON

List of works

Norris, J. (1687) *A Collection of Miscellanies*, London.(Early essays, including 'Of the Advantages of Thinking. A Metaphysical Essay towards a Demonstration of a God from the Steady and Immutable Nature of Truth'.)

Norris, J. (1688) *The Theory and Regulation of Love, a Moral Essay*, Oxford.(His brief correspondence with Henry More was appended to this work, published as *Letters Philosophical and Moral*.)

Norris, J. (1689) *Reason and Religion*, London.(A Malebranchian discussion of divine and human nature.)

Norris, J. (1690) *Cursory Reflections upon a Book called, An Essay Concerning Human Understanding, in Christian Blessedness*, London.(A critique of Locke's *Essay*.)

Norris, J. (1690) *Reflections on the Conduct of Human Life*, London.(Dedicated to Lady Masham: emphasizes the pious goal of philosophy.)

Norris, J. (1693) *Discourse concerning the Measure of Divine Love*, in *Practical Discourses upon Several Divine Subjects*, London, 4 vols, 1691-8.

Norris, J. (1697) *An Account of Reason and Faith*, London.(Reply to Toland's *Christianity not mysterious* (1696).)

Norris, J. (1701-4) *An Essay towards the Theory of the Ideal or Intelligible World*, 2 parts, London; repr. Hildesheim: Olms, 1974; repr. New York: Garland, 1978. (Metaphysics strongly influenced by Malebranche.)

Norris, J. (1708) *A Philosophical Discourse concerning the Natural Immortality of the Soul*, London.(On the immortality of the soul.)

Norris, J. and Astell M. (1695) *Letters Concerning the Love of God*, London. (Correspondence with Mary Astell.)

References and further reading

Acworth, R. (1977) 'Malebranche and his Heirs', *Journal of the History of Ideas* 38: 673-6.(On Norris and his influence.)

Acworth, R. (1979) *The Philosophy of John Norris of Bemerton (1657-1712)*, Hildesheim: Olms.(A comprehensive study of Norris' life, theology and philosophy.)

MacKinnon, F.I. (1910) *The Philosophy of John Norris of Bemerton*, Baltimore, MD: Psychological Monographs of the Philosophical Review.(On Norris' philosophy.)

McCracken, C.J. (1983) *Malebranche and British Philosophy*, Oxford: Clarendon Press.(Norris as the principal conduit for Malebranchism in Britain.)

Nous

Commonly translated as 'mind' or 'intellect', the Greek word nous *is a key term in the philosophies of Plato, Aristotle and Plotinus. What gives* nous *its special significance there is not primarily its dictionary meaning - other nouns in Greek can also signify the mind - but the value attributed to its activity and to the metaphysical status of things that are 'noetic' (intelligible and incorporeal) as distinct from being perceptible and corporeal. In Plato's later dialogues, and more systematically in Aristotle and Plotinus,* nous *is not only the highest activity of the human soul but also the divine and transcendent principle of cosmic order.*

In its pre-philosophical usage *nous* is only one among a number of terms for mind. It is chiefly distinguished from these other words by its tendency to signify 'intelligent' activity - realizing, understanding, planning, visualizing - rather than mental processes more generally, including the emotions.

The earliest Greek philosophers traded on this usage of the term. Heraclitus complained that 'much learning does not teach *nous*'. In Parmenides, the cognate verb *noein* and other related words (*noēma*, 'thought' and *noētos*, 'thinkable') are crucial to his argument. Drawing an absolute distinction between 'opinion' (*doxa*) based upon sense perception and 'truth' (*alētheia*), Parmenides argues that 'noetic' activity, properly speaking, is indissolubly linked to true discourse, valid reasoning and the cognition of reality. What is not true cannot be spoken or 'thought', a rule that applies not only to 'nothing' but also to the illusory data of sense perception. The reality that Parmenides' *nous* deduces and apprehends is unqualified 'being', homogeneous and invariant in time and place.

Anaxagoras (§4), with a cosmology strongly influenced by Parmenides, adopted *nous* as the controlling principle of the universe. Making *nous* quite separate from everything else, he characterized it as 'the finest and purest of all things, which has all knowledge about everything and the greatest power'. *Nous* causes the primordial mixture of other things to rotate and separate into distinct beings.

In the dialogues of Plato the treatment of *nous* was powerfully influenced by these antecedent conceptions. In his *Phaedo*, Socrates favours the idea that *nous* has organized the universe in the best possible way (an idea, he suggests, that Anaxagoras failed to carry through). This conception of *nous* was fully developed in other Platonic dialogues including the *Timaeus*, where it is figuratively expressed in the teleological thinking of the world's divine manufacturer (demiurge). In the *Republic*, the three great images of sun, divided line and cave are ways of distinguishing levels of reality and modes of cognition. Common to all three images is a distinction between the visible world of 'unknowable' phenomena and the 'noetic' world of stable and intelligible Forms. *Noēsis* - the highest activity of the soul's rational component - has cognition of the Forms as its objective, which it pursues by seeking understanding that is unhypothetical and absolutely secure. In ethical and psychological contexts Plato also uses *nous* as a term for the soul's 'rational component', with meanings that may be as broad as 'mind' in everyday English.

Although Plato's special uses of *nous* left their mark on Aristotle, the latter arrived at systematic ideas concerning *nous* as the distinctive faculty of the human soul (see Aristotle §19). In Aristotle's general model of the soul, psychic functions are realizations of bodily potentialities. *Nous*, by contrast, 'has no actual existence before it thinks', and it has no corresponding organ as 'perception' has in being the function of eye, ear, and so on. Like Plato, Aristotle links *nous* to the thinking of incorporeal 'forms' - the definable essences of things; but in contrast to Plato's independently existing 'Forms', those of Aristotle only become actual 'thought objects' in being thought since *nous* is identical in its actuality to what it thinks. For Aristotle, in contrast again with Plato, *nous* can only perform its activity with the help of data provided by 'imagination' (*phantasia*), which is the soul's capacity to represent sensory information, and it functions as the agent not only of theoretical activity but also of purposive action in everyday life.

In a notoriously obscure chapter (III 5) of his work *On the Soul*, Aristotle distinguishes *nous* as 'a capacity to become everything' from *nous* as 'a capacity to make everything', in the way that light makes potential colours actual. This 'active' *nous*, called 'immortal', has often been identified with the Aristotelian Unmoved Mover, whose life is 'a thinking of thinking' (see Aristotle §16). But Aristotle probably regarded human thought as being godlike rather than as being a product of the Unmoved Mover, who exists as an eternally transcendent thinker.

For Plotinus (§4), *nous* comprises 'primary reality', the domain of intelligence and intelligible beings. He construes this domain as an 'emanation' from the ineffable One, the ultimate principle of everything. Taken universally, *nous* corresponds more or less to a syncretism of Plato's Forms with Aristotle's Unmoved Mover. Everlastingly contemplating the One, *nous* is construed as an equivalence between thought thinking itself and intelligible beings as the only true thinkables. The activity of *nous* 'overflows' into 'soul', the principle of embodied life. As a lower level of reality, soul can only think things by treating them successively and separately. Human beings live primarily at the level of 'soul', but they also, by virtue of their immortal and 'undescended' self, have access to identification with *nous* and thereby to a mode of being in which thinker and thought are completely unified. In this transcendent condition, the mind is reality itself.

Stoics and Epicureans tend to use other words for the mind, probably because as rigorous physicalists they found *nous* too strongly tinged by Platonic metaphysics.

See also: Alexander of Aphrodisias §2; Marcus Aurelius §2; Neoplatonism §3; Numenius §2; Platonism, Early and Middle §5; Proclus §5; Psychē; Ptolemy; Theophrastus §3

A.A. LONG

References and further reading

Aristotle (*c.* mid fourth century BC) *On the Soul*, trans. H. Lawson-Tancred, Harmondsworth: Penguin, 1986.(The main treatment of *nous* is found in book III.)

Everson, S. (ed.) (1991) *Companions to Ancient Thought 2: Psychology*, Cambridge: Cambridge University Press. (Includes chapters on Plato, Aristotle and Plotinus.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(Greek text, with translation and commentary, of Heraclitus, Parmenides and Anaxagoras.)

Plato (*c.*380s BC) *Phaedo*, trans. D. Gallop, Oxford: Clarendon Press, 1975.(Includes notes; see *Phaedo* 97-9 for the cosmic role of *nous*.)

Plato (*c.*370s BC) *Republic*, trans. R. Waterfield, Oxford: Oxford University Press, 1993. (Plato's main discussion of *nous* is found in books VI-VII.)

Plato (*c.*360s BC) *Timaeus*, trans. H.D.P. Lee, Harmondsworth: Penguin, 1963.(Plato's seminal account of the interplay of *nous* and necessity in the formation of the world.)

Plotinus (*c.* AD 250-66) *Enneads*, V 1-9 trans. A.H. Armstrong, *Plotinus*, vol. 5, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1984. (Greek text with facing translation.)

Nozick, Robert (1938-)

Although Robert Nozick has published on an enormous range of topics, he is best known as a political philosopher, and especially for his powerful and entertaining statement of libertarianism. In Anarchy, State, and Utopia (1974), Nozick presents an image of a fully voluntary society, in which people cooperate only on terms which violate no one's rights.

Nozick's other major contributions to philosophy include an analysis of knowledge, and an accompanying response to scepticism, an account of personal identity and contributions to decision theory and the theory of rationality.

1 Introduction

Robert Nozick was born in Brooklyn and studied at Columbia and Princeton Universities. He has taught at Harvard since 1969, and was appointed Arthur Kingsley Porter Professor of Philosophy in 1985.

Although he first came to prominence with two papers published in 1969, 'Newcomb's Problem and Two Principles of Choice' and 'Coercion', Nozick's first book (his defence of libertarianism), *Anarchy, State, and Utopia* (1974), is considered his most important work to date. This was followed by *Philosophical Explanations* (1981), The Examined Life (1989) (in which he remarks that he no longer considers himself to be a libertarian) and *The Nature of Rationality* (1993).

Nozick's eclecticism is accompanied by a particularly engaging writing style. He regrets the increasingly technical turn taken by recent intellectual work and has done much to make his own research accessible to a general readership without compromising its philosophical content. His openness appears in another way also: Nozick is often the first to point out the difficulties with his own position, believing that 'there is room for words on subjects other than last words' (1974: xii). Yet, in his search to break new ground he has rarely returned to a subject to answer critics or to fill in the gaps he admits are exposed.

2 *Anarchy, State, and Utopia*

Anarchy, State, and Utopia begins with the words: 'Individuals have rights, and there are things no person or group may do to them (without violating their rights)' (1974: ix). His conclusions fall into three parts: a defence of the minimal state; a theory of economic justice; and a utopian vision of society.

Nozick claims - in opposition to the anarchist - that the existence of a minimal state (the 'night-watchman' state of classical liberalism) is consistent with individual (negative) natural rights to life, liberty and property. To defend this claim he uses an 'invisible hand' argument to show how the state could emerge.

In Locke's state of nature (see Locke, J. §10), individuals have a natural right of self-defence and a right to punish those who violate their rights. Nozick argues that such people would rationally band together in groups to enforce their claims of justice. Commercial protection agencies would follow and, Nozick argues, such agencies would merge or join federations until a single 'dominant protection agency' exists. This yields what Nozick terms an ultra-minimal state, for not all individuals need join. However, Nozick believes that the dominant protection agency has the right to prohibit individuals from exercising their natural right to punish, provided it compensates them for this by offering protection services. Thus, ultimately the protection agency would satisfy two conditions Nozick believes sufficient to constitute a state: it (virtually) monopolizes legitimate coercion within its territory, and it offers protection to almost all. By showing that a minimal state could come into existence without violating rights, Nozick believes he has demonstrated that *at least* a minimal state is justified.

The centrepiece of Nozick's libertarianism, however, is his theory of distributive justice. Nozick argues that the more-than-minimal state - a welfare state, for example - would violate people's rights. Nozick's own theory of justice - the entitlement theory - falls into three parts: principles of justice in the initial acquisition of property; in transfer; and in rectification, which simply concerns how to rectify violations of the first two principles (see Justice §5; Property §3). The principle of justice in acquisition is not fully stated, although Nozick remarks that it must include one element taken from Locke: an individual's appropriation must not worsen any third party's condition. Nozick instead gives most attention to the principle of justice in transfer, which essentially states that a transfer is

just if, and only if, it is voluntary. His contention is that only his theory properly respects liberty. All others, he claims, are defeated by his notorious ‘Wilt Chamberlain’ example.

In this example, Nozick claims that any ‘pattern’ of holdings - each according to their needs, for example - will inevitably be destroyed by the free and voluntary trading and gift-giving behaviour of individuals. The only way of enforcing a pattern involves ‘constantly interfering’ in individuals’ lives: either by prohibiting certain transfers or by intrusively redistributing property. Even a system of taxation and welfare payments is rejected by Nozick on the grounds that it forces some people to labour for the benefit of others, without choice or reward. Accordingly, taxation is ‘on a par with forced labour’. Hence, any state more extensive than the minimal state will violate individual rights to liberty: ‘The socialist society would have to forbid capitalist acts between consenting adults’ (1974: 163).

The third part of Nozick’s project is to show that the minimal state is ‘inspiring as well as right’. Here he makes it clear that libertarianism is intended to provide a set of background rights and duties for society, but people may enter into whatever voluntary arrangements they wish. Thus Nozick argues that a group may set up any type of community they wish, provided they have the resources and do not coerce others to join them. Thus libertarianism is viewed as a ‘framework for utopia’ in which individuals can act out their own model of utopia, in company with like-minded others (see Libertarianism §§1, 3).

Nozick’s views have generated an enormous critical literature. Critics claim that the entire weight of his position stands on the justification of the initial acquisition of individual property rights. And on this point Nozick is particularly weak, doing so little to state and defend his principle of justice in acquisition. Critics have also doubted that Nozick can justify even the minimal state, given the very strong assumptions of individual freedom with which he starts. Nozick has not published a response to these criticisms.

3 *Philosophical Explanations*

Nozick’s second book, *Philosophical Explanations*, ranges unconfined over topics in metaphysics, epistemology and the philosophy of value. While here he makes important contributions to the issues of personal identity and the foundations of ethics, his treatments of knowledge and scepticism have attracted the greatest attention.

What is it to know something, as distinct, for example, from guessing correctly? The traditional analysis of knowledge asserts that knowledge is justified true belief. Yet this seems to have been refuted by ‘Gettier examples’: once we appreciate that some beliefs, while justified, are nevertheless false, it is a short step to realize that a justified belief might be true by accident or fluke (see Gettier problems).

To avoid such problems Nozick presents the ‘truth-tracking’ theory. This is an externalist theory, and Nozick’s own approach is a variety of reliabilism (see Reliabilism), in which, broadly, one has knowledge if one’s true beliefs are generated by a reliable mechanism. Nozick claims that, if ‘*S* knows that *p*’, then four conditions are individually necessary and jointly sufficient:

(1) *p* is true.
(2) *S* believes that *p*.
(3) If *p* were not true then *S* would not believe it.
(4) If *p* were true then *S* would believe it.

Conditional statements (3) and (4) replace the appeal to justification in the traditional account. These are to be interpreted as subjunctive or counterfactual conditionals rather than as material implications. Using possible worlds semantics for counterfactual conditionals (see Counterfactual conditionals §3), the third condition can be read as: ‘In close possible worlds where *p* is not true, *S* does not believe *p*’. If your belief is true by accident, there will be a close possible world in which you still have that belief, even though it is false. The belief fails therefore to meet the third condition, and so does not count as knowledge.

We can now see why Nozick calls his theory the ‘truth-tracking’ theory: knowledge requires not just that you believe the truth, but that your belief ‘tracks’ the truth among close possible worlds. Together the third and fourth conditions require that, not only do you believe the truth, but in reasonably similar situations where it is false you would not believe it, and where it is true you would.

Nozick's analysis gives him a reply to the sceptic about knowledge. If we grant that you cannot know that you are not now a brain in a vat on Alpha Centauri, then it seems to follow that you cannot know anything else: for example, that you are now on Earth, reading. Nozick's reply is to point out that the sceptic's argument presupposes that 'knowledge is closed under known logical implication', and that his own analysis of knowledge shows that this assumption is false.

The closure principle is a relatively simple idea. If you know that *p*, and that *p* logically entails *q*, then according to the closure principle it follows that you know that *q*. The sceptic uses this principle (in reverse) to argue that we have no knowledge. If I am reading on Earth, then it logically follows (and I know that it follows) that I am not a brain in a vat on Alpha Centauri. Hence, according to the closure principle, if I know that I am reading on Earth, then I know that I am not a brain in a vat. But, the sceptic claims, as I do not know that, I do not know that I am reading on Earth. Indeed, I cannot know anything known to be inconsistent with my being a brain in a vat on Alpha Centauri.

The closure principle is central to the sceptic's case. But on Nozick's analysis the principle is false. Nozick points out that whether or not an individual knows a proposition depends on how their beliefs vary over a set of close possible worlds. But the possible worlds to take into account differ according to the proposition under consideration. Hence, I can know that *p*, and know that *p* entails that *q* without knowing that *q*, and the sceptic's argument falls.

While an ingenious response, critics argue that Nozick's analysis of knowledge should be rejected. For it has the peculiar consequence that I can know a conjunction '*p* and *q*' without knowing one of the conjuncts *q* for the set of possible worlds in which '*p* and *q*' is true is not identical to the set of possible worlds in which '*q*' is true. Although this is a consequence Nozick points out himself, for many it is a *reductio ad absurdum* of his approach.

4 *The Examined Life* and *The Nature of Rationality*

The Examined Life, Nozick's third book - a book about 'living and what is important in life' (1989: 11) - has not made much impact on the philosophical world. However, his fourth, *The Nature of Rationality* (1993), marks a return to issues of decision theory and rationality, and so contributes to ongoing debates within the analytic tradition.

Nozick was the first to present Newcomb's problem to the philosophical world, and his discussion has remained a classic work in decision theory, emphasizing the distinction between evidential and causal decision theory. In *The Nature of Rationality* he introduces a new idea: symbolic utility. An action or decision may be symbolic - expressive of an emotion or attitude, for example - and so may have value not so much in its effects, but by its standing as a symbol. To illustrate, Nozick points out that for some people minimum wage legislation may have value as a way of symbolizing the idea of helping the poor, even if it turns out to be ineffective as a policy. Acting rationally, on Nozick's view, is a matter of 'maximizing decision-value', which is a weighted sum of causal, evidential and symbolic utility (see Decision and game theory).

See also: Knowledge, concept of; Personal identity; Rational choice theory; Scepticism

JONATHAN WOLFF

List of works

Nozick, R. (1969a) 'Newcomb's Problem and Two Principles of Choice', in N. Rescher (ed.) *Essays in Honour of Carl G. Hempel*, Reidel: Dordrecht.(First philosophical discussion of Newcomb's problem.)

Nozick, R. (1969b) 'Coercion', in S. Morgenbesser, P. Suppes and M. White *Philosophy, Science and Method*, New York: St Martin's Press.(Influential analysis of the concept of coercion.)

Nozick, R. (1974) *Anarchy, State, and Utopia*, Oxford: Blackwell.(Powerful statement of libertarianism.)

Nozick, R. (1981) *Philosophical Explanations*, Oxford: Oxford University Press.(Wide-ranging discussion of topics in metaphysics, epistemology and ethics.)

Nozick, R. (1989) *The Examined Life*, New York: Simon & Schuster.(A book about 'living and what is important in life'.)

Nozick, R. (1993) *The Nature of Rationality*, Princeton, NJ: Princeton University Press.(Presents a theory of rational action.)

References and further reading

Luper-Foy, S. (ed.) (1987) *The Possibility of Knowledge*, Totowa, NJ: Rowman & Littlefield.(Contains critical discussions of Nozick's views on knowledge and scepticism.)

Paul, J. (ed.) (1982) *Reading Nozick*, Oxford: Blackwell.(Collection of essays discussing *Anarchy, State, and Utopia*.)

Wolff, J. (1991) *Robert Nozick*, Cambridge: Polity Press.(Critical analysis of Nozick's libertarianism.)

Numbers

Numbers are, in general, mathematical entities whose function is to express the size, order or magnitude of something or other. Historically, starting from the most basic kind of number, the positive integers (1, 2, 3,...), which appear in the earliest written records, the notion of number has been generalized and extended in several different directions - often in the face of considerable opposition.

Other than the positive integers, the most venerable are the rational numbers (fractions), which were known to the Egyptians and Mesopotamians. The discovery, by Pythagorean mathematicians, that there are lengths that cannot be expressed as fractions occasioned the introduction of irrational numbers, such as the square root of 2, though the Greeks managed only a geometric understanding of these. The number zero was recognized, first in Indian mathematics, by the seventh century; the use of negative numbers evolved after this time; and complex numbers, such as the square root of -1*, appeared first at the end of the Middle Ages. Infinitesimal numbers were developed by the founders of the calculus, Newton and Leibniz, in the seventeenth century (and were later to disappear from mathematics - for a time); and infinite numbers (ordinals and cardinals) were introduced by the founder of modern set theory, Cantor, in the nineteenth century.*

The introductions of three of these kinds of number, in particular, occasioned crises in the foundations of mathematics. The first (concerning irrational numbers) was finally resolved in the nineteenth century by the work of Cauchy and Weierstrass. The second (concerning infinitesimals) was also resolved then, by the work of Weierstrass and Dedekind. The third (concerning infinite numbers), which involves paradoxes such as Russell's, still awaits a convincing solution.

It is seemingly impossible to give a rigorous definition of what it is to be a number. The closest one can get is a family-resemblance notion, with very ill-defined boundaries.

1 Natural numbers

All human societies would appear to have some form of counting, however limited, both for size - cardinality (one, two, three,…) - and for order - ordinality (first, second, third,…). The recognition that an infinitude of positive integers could be used for these purposes was already in place in the earliest civilization from which we have written records, that of Egypt *circa* the fourth millennium BC. The investigation of basic arithmetic operations (addition, multiplication and so on) was also well under way then. This was continued in the next great Middle Eastern civilization, that of Mesopotamia (Babylon), *circa* the second millennium BC. The investigations were greatly facilitated by the Mesopotamian invention of place-notation, that is, the idea that a single numeral can represent different quantities depending on where it occurs in a string of numerals. Thus, for example, in decimal notation, the first '2' of '2,121' represents two thousands and the second represents two tens. (The Mesopotamians actually preferred a sexagesimal to a decimal base.)

The *natural* numbers comprise the positive integers together with zero. The idea of zero as a number was not to be found in either of these civilizations, although the Mesopotamians did sometimes leave a gap where we would now write a zero. (Thus ' 2 2' might represent 202 or 2,020, and so on.) Though the origin of the idea is uncertain, zero was certainly being used by the seventh century AD by Indian mathematicians, who had also developed place-notation and used a decimal base. They therefore possessed, in effect, our modern number system (though the symbols they used were different). This system was taken up in the Arabian civilization around the eighth century, and thence passed into Europe around the twelfth.

2 Rational and irrational numbers

Positive integers quantify wholes. Once one starts to measure, it becomes clear that a way is required to quantify parts. *Rational* numbers, as ratios of natural numbers (with non-zero denominator), clearly serve this function. These were known to the Egyptians, at least in the form of unit fractions (of the form $1/n$) and, in a more sophisticated and general form, to the Mesopotamians. Decimal fractions (of the form 0.23) appear to have been developed by Arabic mathematicians, and by the time we reach Stevin in late sixteenth-century Europe, the modern computational system for rational numbers is essentially in place.

The creation of modern mathematics is usually reckoned to have occurred in Ancient Greece in the second half of the first millennium BC. What was distinctive about the Greek approach was that mathematics was freed from its practical roots and took on a purely theoretical form. In particular, the central concern of Greek mathematics was proof, especially proofs in geometry. As far as arithmetic goes, the first part of this epoch was dominated by Pythagoreanism, according to which everything can be explained in terms of the positive integers. (It follows that rational numbers are of little theoretical importance: given two lengths to compare, if one chooses one's unit of measurement small enough, the lengths will be integral.)

The discovery that the Pythagorean assumption was wrong came as a distinct shock. Most simply, the diagonal of a unit square (whose length is the square root of two) is not commensurable with its sides. The proof of this (one of the first *reductio* arguments in the history of mathematics) is now celebrated. It is not known who discovered it, though it was probably one of the later Pythagoreans, *c*.400 BC. At any rate, it inaugurated the first of three crises in the foundations of mathematics that have been associated with the introduction of a new kind of number.

The new numbers in this case are *irrational* numbers (that is, numbers which cannot be written as a ratio of integers - written as decimals they are infinite and non-repeating), such as the square root of two (and pi (π), although the irrationality of this was not proved until the eighteenth century by Lambert). They are required, together with rational numbers, for general quantification of length: in fact, the only conception of such magnitudes available to the Greeks was as geometric lengths. Given this representation, some geometric account of arithmetic operations had to be given. The hardest of these is division, which was solved by Eudoxus (*c*.360 BC) in his theory of proportions, given in book 5 of Euclid's *Elements*. But even for Eudoxus, ratios were not entities in their own right. Operations which treat irrational numbers in the form of surds (expressed in ways such as $\sqrt{2}$, $\sqrt[3]{71}$), as bona fide entities, can be found in Indian mathematics and, in particular, in the seventh-century mathematician Brahmagupta. Some use of surds was made by Arabic mathematicians, but their use did not become common until the sixteenth century (for example, in the work of Stevin and Cardan). A clear statement to the effect that irrational numbers are first-class numerical citizens can be found in Newton, in the seventeenth century; however, an adequate understanding of what this citizenship amounted to had to wait until the nineteenth century.

3 Negative and complex numbers

Subtraction had been familiar to mathematicians since the earliest times. The idea that a sensible quantity might itself be negative, however - and what the sense of it might be - took a long time to catch on. *Negative* numbers, the rules for operating with them, and their use as roots of equations appear in Brahmagupta, and were consequently known to Arabic mathematicians - though they did not use them as roots of equations. The first time they appear in this way in European mathematics is in the work of fifteenth-century mathematician Nicholas of Chuquet (who, however, did not allow zero to be a root!). Even Cardan, a century later, who employed negative roots extensively, called them 'fictional'. The idea that negative numbers might correspond to direction reversal is to be found in the early seventeenth-century mathematician, Girard; but, even after the invention of analytic geometry by Descartes a little later, many mathematicians simply ignored the negative parts of the plane. And even as late as the middle of the eighteenth century some textbook writers were still disputing the claim that the product of a pair of negative numbers is positive.

Once the use of negative numbers became common mathematical practice, the thought that there might be a new kind of number to provide for their square roots was not far behind. If we call all of the numbers mentioned so far *real* numbers, and write the square root of -1 as i (following Euler), the product of i with any real number is called an *imaginary* number. More generally, the sum of a real number and an imaginary number ($a + \mathrm{i}b$) is called a *complex* number. One of the earliest occurrences of complex numbers in mathematics is in the work of Cardan, who observed, as we would now put it, that a cubic equation with a real root may also have complex roots, though Cardan himself regarded this observation as useless. A few years later, Bombelli articulated the use of complex numbers (in the same context) much further. Opposition to them did not then cease, however; for example, their use was frowned upon by Newton. But by the time of, and especially in the work of, Euler, in the eighteenth century, complex numbers had been shown to have numerous uses, both as the solutions of equations and elsewhere - for example, in important relationships between trigonometric functions. And by the time that Gauss, in the early nineteenth century, showed that every polynomial equation with complex coefficients has complex roots, complex numbers were well entrenched. With the development of electromagnetic theory later in the

century, complex numbers were shown to be capable of representing even physical quantities (for example, impedance).

The problem of what sense to make of complex numbers was also solved by Gauss and later contemporaries, such as Argand. The complex number $x + \mathrm{i}y$ could be thought of as the point (x, y) in two-dimensional Euclidean space (now usually called the Argand plane), with the arithmetic operations on complex numbers defined in a suitable way. This prompted the idea that points in a higher-dimensional space could also be thought of as numbers of a certain kind. Defining a suitable notion of multiplication for points in a three-dimensional space turned out to be impossible, but the problem was solved for four dimensions by the Irish mathematician William Hamilton later in the nineteenth century. Hamilton defined a class of numbers known as 'quaternions', of the form $x + \mathrm{i}y + \mathrm{j}z + \mathrm{k}w$, where $\mathrm{ijk} = \mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1$. The properties of quaternions were investigated for some years, but no real use was found for them, so the investigations lapsed. By this time, the observation that a somewhat different set of operations on points in Euclidean space might be fruitful had been made, and vector algebra was born. (If one thinks of complex numbers as two-dimensional vectors, then complex addition and vector addition are the same thing, but complex multiplication and the various vector multiplications are distinct.)

4 Infinitesimal numbers

The numbers discussed so far have all been finite. But numbers of some kinds are infinite - or their intuitive inverse; infinitesimal. It is sometimes (dubiously) claimed that infinitesimals can be found in Greek mathematics. What certainly can be found, from about the third century BC on, is a method called 'the method of exhaustion', (probably) invented by Eudoxus and developed by Archimedes. The main purpose of this method is the computation of areas and volumes of non-rectilinear geometric figures, and requires their approximation by rectilinear figures. For example, a circle is approximated by a regular polygon inscribed within it. As the length of the sides of the polygon decreases (and the number of sides increases) the approximation gets closer. By the late sixteenth century we find mathematicians such as Stevin and Wallis suggesting, in effect, that if we make the sides of the polygon infinitesimally small, the approximation will be exact - or, at least, will differ only infinitesimally from the true value. Hence, a new kind of number, *infinitesimals*, had to be recognized, to quantify infinitesimal lengths.

The use of infinitesimals in essentially this way was developed by Newton and Leibniz in the next century as the foundation of modern calculus (see Analysis, philosophical issues in §1). Though their inventions were independent, and used quite different notations, the essential ideas were the same. For example, to determine the gradient of the curve $y = x^2$ at the point $x = a$, we consider a point on the curve an infinitesimal distance *d* away along the *x*-axis. The slope of the line joining these two points is

$$\frac{(a+d)^2 - a^2}{d} = \frac{2ad + d^2}{d}$$
$$= 2a + d$$

but since *d* is infinitesimally small, we can disregard it. Hence the gradient is 2*a*.

The power of infinitesimal methods was so great, in both pure and applied mathematics, that they soon became entrenched. However, many people - most notably Berkeley - were severely critical. In particular, infinitesimals had at the same time to be both non-zero (since one divided by them) and zero (since one ignored them). This dilemma inaugurated the second foundational crisis in mathematics associated with the introduction of a new kind of number. But this one would have to wait only 200 years for a solution.

5 Transfinite numbers

It is clear that there are infinite quantities in mathematics, and hence that there must be a size that these quantities can be, namely, infinite. But for most of the history of mathematics, infinitude was thought of as mathematically indeterminate, and hence there was little to be said about the quantity 'infinity'. (One does, however, sometimes find mathematicians - even of the stature of Euler - taking it as a definite quantity, in particular, as the reciprocal of zero.) Even in the heyday of the infinitesimal calculus, infinite numbers played little part (though Leibniz certainly held that the reciprocals of distinct infinitesimals were distinct infinite numbers). The recognition of different orders of infinity, and consequently of the sense of infinite numbers, therefore came as another highly contentious

idea.

This recognition is almost entirely due to Georg Cantor in the late nineteenth century. He realized that, given some iterable operation (for example, forming the set of topological limit points of some set of points), it made sense to consider the result of having applied the operation an infinite number of times, and then to apply the operation again, and again…indefinitely (see Cantor, G. §1). To quantify this ordering, Cantor introduced a new class of infinite numbers, *transfinite ordinals*, the least of which he denoted by a lower-case omega, ω (see Set theory §2). He also defined arithmetic operations on these numbers which generalize the operations on finite numbers, but which have rather unusual properties. For example, addition is not commutative: $\omega + 1 \neq 1 + \omega$.

Having generalized the notion of order into the infinite, Cantor next generalized the notion of size. To do this he needed a criterion of size that would work (even) in infinite domains. He adopted the criterion (also suggested by Frege (§8)) that two collections have the same size just if their members can be put into one-to-one correspondence (that is, each member of either collection corresponds to exactly one member of the other). He observed that adding members to an infinite set may not increase its size. (This fact had been noted by some previous mathematicians - for example, Galileo - who thought it so absurd that they rejected infinity as a quantity.) In another of the most famous *reductio* arguments in the history of mathematics, Cantor also established that for any infinite collection, there is one of greater size. This is now called 'Cantor's theorem' (see Cantor's theorem). To quantify the different sizes Cantor introduced a new kind of infinite number, *transfinite cardinals* (see Set theory §3), denoted by the Hebrew letter aleph ($\aleph$) with a subscript ($\aleph_0$ being the smallest, the size of the natural numbers); and defined arithmetic operations on these which, again, generalize the finite case. These operations also have striking properties: for example, double any aleph and you get what you started with.

It should be noted that natural numbers can be thought of indifferently as ordinals or cardinals. This is because, canonically, we count the size of a finite collection by ordering, and its cardinality is n just if the last object in the ordering is the nth. However, transfinite ordinals and cardinals are quite distinct since, as just observed, if one adds further objects to an ordered set, the size may remain the same. Transfinite ordinal and cardinal arithmetics are also quite distinct. Cardinal addition, for example, is commutative, where ordinal addition is not. It is worth noting that there are very simple questions concerning basic transfinite arithmetic operations that are still unanswered (see Continuum hypothesis).

Cantor's introduction of transfinite numbers into mathematics did not entail the use of infinitesimals. This is because the unusual properties of (cardinal and ordinal) multiplication allow no obvious sense to be given to the notion of division, and so not to reciprocation. However, the use of infinite numbers brought its own problems. In particular, the totality of all ordinal numbers is an ordered collection and hence must have an ordinal, which must be greater than, and so distinct from, all ordinals. Similarly, the totality of all objects must have the largest possible cardinal, but, by Cantor's theorem, there is a larger. These paradoxes are now called the Burali-Forti paradox and Cantor's paradox, respectively; Russell's paradox, concerning the set of all sets that are not members of themselves, is a stripped-down version of Cantor's (see Paradoxes of set and property §4). They heralded the third crisis in the foundations of mathematics associated with the introduction of a new kind of number; a crisis which, unlike the first two, is as yet not satisfactorily resolved. These paradoxes fuelled the rejection of Cantor's ideas by some, including Kroneker; but despite this, they were absorbed into orthodox mathematics within about fifty years.

6 Weierstrass and Dedekind

The final resolution of the first two crises we met (concerning irrationals and infinitesimals) took place in the nineteenth century, a period when mathematicians became particularly concerned with the rigorous foundations of their subject. In this context, the work of Weierstrass and Dedekind is particularly significant.

Early in the nineteenth century, the notion of a 'limit' appeared in Cauchy's formulation of the calculus. His method differed from that set out in §4 above as follows: instead of taking d to be some infinitesimal quantity, we let it be a finite quantity, and then consider the limit of what happens when d approaches zero (the limit being a quantity that may be approached as close as we please, though never, perhaps, attained). Despite the fact that Cauchy possessed the notion of a limit, he mixed both infinitesimal and limit terminology, and it was left to Weierstrass, later in the century, to replace all appeals to infinitesimals by appeals to limits. At this point infinitesimal numbers disappeared from mathematics (though they would return, as we shall see).

Weierstrass also gave the first modern account of negative numbers, defining them as signed reals, that is, pairs whose first members are reals and whose second members are 'sign bits' ('$+$' or '$-$'), subject to suitable operations.

A contemporary of Weierstrass, Tannery, gave the first modern account of rational numbers. An 'equivalence relation' is a relation *R* that is reflexive (*xRx*), symmetric (if *xRy* then *yRx*) and transitive (if *xRy* and *yRz* then *xRz*). Given an equivalence relation on a domain of objects, an 'equivalence class' is the set of all those things in the domain related to some fixed object. Tannery defined the rationals as equivalence classes of pairs of natural numbers (the second of which is non-zero), under the equivalence relation '$\sim$' (tilde), defined as follows.

$\langle m, n\rangle \sim \langle r, s\rangle$ iff $m.s = r.n$

The problem of irrational numbers was finally solved independently by Weierstrass, Cantor and Dedekind, who gave different but equivalent constructions of real numbers. Weierstrass' is in terms of infinite decimal expansions; Cantor's is in terms of convergent infinite sequences of rationals. Dedekind's construction is the simplest: consider any splitting of the rational numbers (which can be taken to include the integers, since the integer *n* may be identified with the rational number $n/1$) into two non-empty disjoint collections *L* ('left') and *R* ('right'), such that anything less than a member of *L* is in *L*, and anything greater than a member of *R* is in *R*. This pair is now called a 'Dedekind cut' (or 'section'). Real numbers (including irrational numbers) can be thought of as such sections (or just one of their parts).

Dedekind also gave the first axiom system for the natural numbers, in terms of an initial number and the successor operation ($+1$). In modern form, the axioms are as follows:

(1) 0 is a number.
(2) The successor of any number is a number.
(3) 0 is the successor of no number.
(4) Any two numbers with the same successor are the same.
(5) If 0 has some property, and the successor of any number with that property also has it, then all numbers have it.

These axioms, together with the recursive definitions for addition and multiplication (also given by Dedekind), are now usually named after Peano, who gave a formalized version a few years later.

7 Set-theoretic reduction

The drive for rigour in the foundations of number theory reached its height in the reduction of all numbers to sets, due to the logicists Frege, Russell and Whitehead at the turn of the twentieth century (see Logicism). The key to this was a set-theoretic definition of cardinal and ordinal numbers. The cardinal numbers were defined as equivalence classes of sets under the equivalence relation 'can be put into one-to-one correspondence with'. (For Russell and Whitehead, this was restricted to sets of fixed type, in an attempt to avoid the paradoxes of infinite number - see Theory of types.) A 'well-ordering' on a set is an ordering such that every non-empty subset has a least member in the ordering (see Set theory §2). Two well-ordered sets are 'order-isomorphic' if they can be put into a one-to-one correspondence that preserves the ordering. The ordinal numbers were defined as the equivalence classes of well-ordered sets (of a given type) under the equivalence relation of order-isomorphism. Cardinal and ordinal arithmetic operations were defined in an appropriate fashion.

Given the ordinals, other numbers could then be defined in a relatively straightforward way. Natural numbers are the finite ordinals (which can be shown to satisfy the Peano axioms - see §6 above); rational numbers can be defined by the Tannery construction; real numbers as Dedekind sections; negative numbers as Weierstrassian pairs; complex numbers can be defined as pairs of signed reals, thought of as points on the Argand plane. In each case, arithmetic operations can be defined in natural ways.

Since the Frege/Russell reduction, several other non-equivalent, but equally good, definitions of the cardinal and ordinal numbers have been discovered (and in virtue of this, the claim that these numbers just *are* certain sets is difficult to maintain). The most elegant of these is due to von Neumann. According to this, each ordinal is simply the collection of all smaller ordinals. (Thus, zero is simply the empty set, the least infinite ordinal is the set of natural numbers, and so on.) Cardinal numbers are identified with 'initial' ordinals, that is, least ordinals of each

size.

8 Developments from logic

Work in mathematical logic in the twentieth century has provided several notable developments bearing on numbers. Three are particularly important. The first was the proof by Gödel in 1931 that the Peano axioms, and all other consistent axiom systems for arithmetic, are incomplete, in the sense that there are truths of arithmetic that cannot be proved from the axioms - at least if the underlying logic is first-order (see Gödel's theorems). The axioms are complete if the underlying logic is second-order and the induction principle (5) is formulated as a second-order axiom and not just a first-order schema; but second-order logic is not itself axiomatizable (see Second- and higher-order logics §1). This raises profound questions about the nature of both numbers and our knowledge thereof, that fall outside the bounds of this entry.

The second development concerns the paradoxes surrounding transfinite numbers. The orthodox view that has emerged this century is that embedded in Zermelo-Fraenkel set theory (ZF). According to this, there just is no totality of all ordinals, all sets or other 'large' collections, and so the question of their size does not arise. Although this account provides enough set theory for most mathematics (though not all: category theory appears to require large sets of just this kind), it can hardly be said to be conceptually adequate. For example, standard logic defines the sense of a quantifier in terms of the domain (totality) over which it ranges. It is therefore unclear what the sense of the quantifiers of ZF is, if, as it claims, there is no such totality. (See Priest 1987.)

The third development is due to Robinson and is called 'nonstandard analysis' (see Analysis, nonstandard). As was proved originally by Löwenheim and Skolem, (first-order) theories of number have nonstandard models (see Löwenheim-Skolem theorems and nonstandard models). In particular, any theory of the reals will have such models. Robinson showed that in all of these models, there are non-zero numbers that are smaller than any real number: infinitesimals. Using these, he demonstrated that the reasoning of the infinitesimal calculus (which is much more intuitive than limit reasoning) can be interpreted in a perfectly consistent manner. Hence, infinitesimals have been rehabilitated as perfectly good numbers.

9 Number in general

The preceding review of the development of the notion of number naturally prompts the question of what a number is. One might interpret this as the question of whether numbers are Platonic objects, mental constructions, or nothing more than mystified numerals. This is a central issue in the philosophy of mathematics.

Alternatively, in virtue of the plethora of kinds of numbers we have seen, one might interpret the question as asking what makes entities of certain kinds, but not others, numbers. Beyond the rather vague characterization with which this article began, it seems difficult to give a general characterization of number. The most fundamental numbers (both historically and conceptually), the natural numbers, measured size (or order), were subject to distinctive operations (such as addition) and could be the roots of equations. Each of these central features has played a role in generating new kinds of numbers (different concerns being dominant on different occasions). The result is a collection of entities which are related by family resemblance (as observed by Wittgenstein in §67 of *Philosophical Investigations*), though the boundaries of the family seem somewhat arbitrary. It is difficult to see why, for example, complex numbers (or quaternions) should be called numbers, but not vectors or numerical matrices; both of these share the central features of natural numbers.

This conclusion is reinforced by recent work by Conway (1976). He gives a (transfinite) recursive construction that generalizes both the Dedekind construction of the reals and the von Neumann construction of ordinals. Essentially, a number is any pair, $\langle L, R\rangle$, such that all the members of *L* and *R* are numbers, and every member of *R* is greater than or equal to ($\geq$) every member of *L* (see Analysis, philosophical issues in §2). ' $\geq$' and the arithmetic operations are also defined in a natural recursive manner. The construction generates virtually all the numbers we have met in this article, including infinitesimals, but excluding, notably, the complex numbers (and if cardinals are to be identified with initial ordinals, a non-uniform definition of arithmetic operations is necessary). Moreover, the construction generates many novel numbers, for example, notably, numbers obtained by applying the full range of real-number operations to infinite numbers, for example, $\omega - 1$, $\sqrt{\omega}$, which make no sense on the usual understanding. Moreover, a simple generalization of the construction (dropping the ordering condition on

L and *R*), produces even more number-like objects (which Conway calls ‘games’, because, in a certain sense, they code the strategic possibilities in a two-person game).

Just conceivably, a unifying account of number might eventually be found, but in the meantime the emergence of new kinds of numbers seems likely. For example, there are nonstandard inconsistent models of arithmetic which contain inconsistent numbers (natural numbers with inconsistent properties). These have some notable applications. For example, some of them can be shown to provide solutions for arbitrary sets of simultaneous linear equations. (See Mortensen 1995.) And just as the existence of nonstandard models of analysis made infinitesimals legitimate, so might these legitimize the notion of an inconsistent number.

See also: Analysis, philosophical issues in; Antirealism in the philosophy of mathematics; Arithmetic, philosophical issues in; Logical and mathematical terms, glossary of; Realism in the philosophy of mathematics; Set theory, different systems of

GRAHAM PRIEST

References and further reading

Bell, J.L. and Machover, M. (1977) *A Course in Mathematical Logic*, Amsterdam: North Holland. (Chapter 10 gives a clear account of the modern set-theoretic reduction of cardinal and ordinal numbers.)

Benacerraf, P. (1965) ‘What Numbers Could Not Be’, *Philosophical Review* 74: 47-73; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.(A standard critique of set-theoretic reductionism.)

Boyer, C.B. and Merzbach, U.C. (1968) *A History of Mathematics*, New York and London: Wiley & Sons; 2nd edn, 1989, repr. 1991.(An excellent reference on the history of mathematics, recently updated.)

Cajori, F. (1919) *A History of Mathematics*, London: Macmillan; 5th edn, New York: Chelsea, 1991. (A classic on the history of mathematics. Now a bit dated, but still good value.)

Conway, J.H. (1976) *On Numbers and Games*, London: Academic Press.(Provides a unified construction of many different kinds of number. Cited in §9.)

Crossley, J.N. (1980) *The Emergence of Number*, Singapore: World Scientific.(A very readable history of number.)

Dedekind, R. (1963) *Essays on the Theory of Numbers*, New York: Dover.(Includes English translations of two of the most important essays on numbers ever written.)

Lakatos, I. (1978) ‘Cauchy and the Continuum’, in *Collected Papers*, vol. 2, *Mathematics, Science and Epistemology*, Cambridge: Cambridge University Press, ch. 3.(A perceptive analysis of the infinitesimal calculus, its downfall and its relationship to nonstandard analysis.)

Mortensen, C. (1995) *Inconsistent Mathematics*, Dordrecht: Kluwer.(Chapters 2 and 8 include discussions of inconsistent numbers and their use in solving inconsistent sets of linear equations. Cited in §9.)

Nechaev, I.V. (1987) ‘Number’, in M. Hazewinkel (ed.) *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet Mathematical Encyclopaedia*, Dordrecht: Kluwer.(A clear but more technical account of number.)

Priest, G. (1987) *In Contradiction*, The Hague: Martinus Nijhoff.(Chapter 2 includes a discussion of problems of ZF. Cited in §8.)

Wittgenstein, L.J.J. (1953) *Philosophical Investigations*, ed. G.E.M. Anscombe and R. Rhees, trans. G.E.M. Anscombe, Oxford: Blackwell.(Numbers illustrate Wittgenstein’s notion of family resemblance. Cited in §9.)

Numenius (*fl. c.* mid 2nd century AD)

Numenius was a Platonist philosopher. He came from Apamea (Syria) and wrote in Greek. His work - now lost - is usually considered Neo-Pythagorean in tendency, and exercised a major influence on the emergence of Neoplatonism in the third century. A radical dualist, he postulated the twin principles of god - a transcendent and changeless intellect, equated with the Good of Plato's Republic - and matter, identified as the Pythagorean Indefinite Dyad: god is good, matter evil. In addition to this supreme god, he added at a secondary level a creator-god, one of whose aspects is the world-soul, itself further distinguished into a good and an evil world-soul. He had a strong interest in Oriental wisdom, especially Judaic, and famously called Plato 'Moses speaking Attic'.

1 Life, work and influence

Nothing is known of Numenius' life, but he can be dated with reasonable accuracy by the fact that he is attested as the teacher of one Harpocration, who was also influenced by the Athenian Platonist Atticus, who in turn flourished in the AD 170s. He is often mentioned in conjunction with a 'companion', Cronius, who was presumably associated with his school, and who may possibly be the addressee of Lucian's treatise on Peregrinus.

Of his works none has survived, but some extracts of his dialogue *On the Good* are preserved by Eusebius in his *Preparation for the Gospel*, as also are some considerable passages from a lively polemical work, *On the Apostasy ['Diastasis'] of the Academics from Plato*, which helps to clarify Numenius' own position, while providing some useful data on the New Academy. Alongside this, we know of the works *On the Indestructibility of the Soul* and *On the Secret Doctrines of Plato*, treatises *On Numbers* and *On Place*, and a work called *Epops*, or 'The Hoopoe', which probably embodies a pun on *epopteia* (mystical vision). We also have an extended account of his doctrine on matter preserved by the late Roman commentator on Plato's *Timaeus*, Calcidius, who may well be more extensively indebted to him than he acknowledges.

His philosophical importance is considerable. He was a major influence, through the mediation of Ammonius Saccas (not to be confused with Ammonius, son of Hermias), on the father of Neoplatonism Plotinus and his followers Amelius and Porphyry, as well as the Christian theologian Origen (see Neoplatonism §1; Porphyry §4; Origen). His Pythagoreanism consists of presenting Plato as a disciple of Pythagoras (see, for example, fragments 7, 24.57), although without derogating from Plato's greatness (as was done by more extreme Pythagoreans, such as Moderatus of Gades).

Numenius was much interested in the wisdom of the East and in comparative religion. He attracted the interest of Church Fathers by his references to Jahveh, Moses and even Jesus (fr. 1). Indeed, he described Plato as 'Moses speaking Attic' (fr. 8), which seems to imply an acceptance of something like Philo's wholesale allegorization of the Pentateuch (see Philo of Alexandria §1). There has been speculation that he was himself of Jewish stock - his hospitality to the Jewish tradition is certainly notable - but this is not a necessary inference. Numenius may simply be reflecting the syncretistic religious and philosophical milieu in which he lived.

2 Metaphysics

Numenius' views on ethics and logic are not known (although his ethical stance may be assumed to be austere), so we may confine ourselves to his metaphysics and psychology. He is at odds with previous Pythagoreans in maintaining a radical dualism between the first principles of god (the Monad, the Good) and matter (the Dyad), instead of subordinating the material Dyad to the all-generating Monad, as is done by his Pythagoreanizing predecessors from Eudorus through Moderatus to Nicomachus of Gerasa. Numenius' dualism allies him rather with Plutarch and Atticus, and leads him, like them, to postulate an evil world-soul, derived from a reading of Plato (*Laws* X) to balance the beneficent world-soul (see Plutarch of Chaeronea§§3-4).

Numenius proffered a system of three levels of spiritual reality: a primal god (the Good, or the Father), who is almost supra-intellectual; a secondary, creator-god (the demiurge of Plato's *Timaeus*); and a world-soul. In this he anticipates to some extent Plotinus, although he was more strongly dualist than Plotinus in his attitude to the physical world and matter, and does not take the radical step of declaring his first principle to be above intellect - although his intellect is an intellect *at rest*, as opposed to the secondary intellect, which is *in motion* (fr. 15). Numenius does seem to have felt the awkwardness of attributing both intellectual and creative activity to a first

principle which is utterly transcendent and completely unitary (in fragment 12, he is insistent that the first god is ‘inactive in respect of all activity’), but he was no more willing than his predecessors to declare it to be ‘beyond being *and* intellection’. Indeed, mindful of Plato (*Sophist* 248e) he is even unwilling to deny it a form of motion, speaking of the ‘rest’ (*stasis*) of the first god as ‘an innate motion’ - ancestor of the ’spiritual motion’ of intellect in later Neoplatonism.

The distinction between his second and third gods is rather subtle, and has led to confusion. In fragment 11 he declares that the second and third gods are ’one’, but a division results from the demiurgic activity of the second god with regard to matter. This difference splits the god into two aspects, similar to the ’transcendent’ and ’immanent’ aspects of the *logos* of god in the philosophies of Philo of Alexandria and Plutarch (at least in the latter’s treatise *On Isis and Osiris*). The ’lower’ aspect of the demiurgic intellect may be viewed as a ’third god’, and constitutes a rational world-soul, which bestows form on the chaos of matter to create the physical world at the price of suffering multiplicity itself.

Numenius is, like any later Platonist, concerned with basing his doctrine on suitable Platonic proof-texts. His doctrine of three gods he derives primarily from the Platonic *Second Letter*, which he would have naturally taken as genuine, but he derives confirmation also from a well-known passage of the *Timaeus*, where he takes the ‘essential living being’ (*auto ho esti zōon*) as the primal god (equated in turn with the ’Good’ of the *Republic*). This essential living being the demiurgic intellect contemplates, thus bringing into being the world of Forms, while also ‘discursively intelligizing’ (*dienoēthē*) the physical world, thus constituting itself, or a lower projection of itself, as the world-soul (fr. 22). The fact that the *Timaeus* myth presents the demiurge as creating the world-soul may have led Numenius to describe his third god as a ‘creation’ (*poiēma*), as Proclus alleges (fr. 21) it is certainly an active principle.

Adumbrated in this fragment is his doctrine of *proschrēsis* (‘calling in the help of… ’) which, although obscure, may anticipate Neoplatonic doctrines of the interaction between *hypostases* (see Plotinus §3). Each level of being (the primal god, the demiurge) performs its proper role only by *proschrēsis* of the level of being below it. This serves to link up the entire spiritual world into a dynamic whole. It may be behind the rather bombastic imagery of ’grandfather, son, and grandson’, which Proclus criticizes (fr. 21), and which Numenius seems to have used to characterize the relations between the three principles.

It is in his doctrine of matter (Calcidius, fr. 52) that Numenius’ dualism emerges most strongly. Matter, as the Pythagorean Indefinite Dyad, is coeval with the primal god, and is a positively evil force. Numenius is scornful of his predecessors’ attempts to derive the Dyad from the Monad by one formula or another (fr. 52.15-). He discerns the postulation of a maleficent, ‘material’ soul in Plato *Laws* (X 896e), which, like Plutarch before him, he sees as represented also by the ‘disorderly cause’ of the *Timaeus* (see Plutarch §§3-4). This can be brought to order by the demiurgic intellect, but never completely neutralized.

This ’evil’ soul is represented in each individual also (frs 43-4), although how exactly its functions are distinguished from those of the rational soul is less than clear. It is not a simple distinction between rational and irrational soul, since Numenius also envisages the rational soul as picking up a series of ’garments’ as it descends through the heavenly spheres to incarnation, some of which are irrational. But later evidence from Porphyry and Iamblichus indicates that Numenius’ psychology was regarded as distinctive and strongly dualist. Finally, Numenius, like Middle Platonists in general, had no objection to postulating transmigration into animals if a soul became excessively corrupt.

See also: Neo-Pythagoreanism; Platonism, Early and Middle §§1, 4

JOHN DILLON

References and further reading

Beutler, R. (1940) ‘Noumenios’, *Realencyclopädie der klassischen Altertumswissenschaft*, suppl. 7: 664-78(A useful general account, now superseded in large part by Frede (1987).)

Dillon, J.M. (1977) *The Middle Platonists*, London: Duckworth.(Pages 361-79 offer an introductory account of Numenius.)

Dodds, E.R. (1960) ’Numenius and Ammonius’, *Entretiens Fondation Hardt* 5: 3-32.(Good discussion of the problem of Numenius’ influence on Ammonius Saccas.)

Frede, M. (1987) 'Numenius', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36 (2) 1,034-75.(An excellent discussion of problematic aspects of Numenius' thought.)
Numenius (*fl. c.* mid 2nd century AD) Fragments, in E. des Places (ed.) *Numénius*, Paris: Les Belles Lettres, 1973. (The standard edition with French translation.)

Nursing ethics

Nursing ethics may be defined simply in relation to what nurses do that doctors and others do not characteristically do; or in relation to the nursing perspective on any issues in health care and medicine. More radically, it claims to employ a distinctive conceptual framework, regarding care, rather than cure, as fundamental. Nursing ethics concerns itself with the relationship between 'carer' and 'cared for' and the meanings embedded in that relationship. It is the moral exploration of an illness or disability as a personal life crisis rather than an instance of a biomedical generalization.

The divergence of medical and nursing ethics took its cue from a distinction between curing and caring. Curing, involving diagnosis, treatment and prognosis, was taken to be characteristic of medicine (with surgery as the paradigm), while caring, involving preparation for treatment, maintenance, and recuperation, was characteristic of nursing (with, perhaps, palliative or domiciliary care as paradigmatic). These were soon taken to be more than just two aspects of physical management of patients but conceptually distinct kinds of activity, one directed to a 'distanced' biology-based approach to dysfunctional specimens, or even parts of specimens, of *Homo sapiens*, and the other a humanities-based intimate approach to individual persons with unique life crises. Whereas this distinction is not sustainable as an absolute it does provide enlightening analytic comparisons of emphasis or tendency.

The distinction has continued to develop as a philosophical one and has merged with gender debates. Nursing ethics has taken up the notion of 'care' formulated by Carol Gilligan (1982) and Nel Noddings (1984) (see Feminist ethics §1). The debate is largely an engagement with the 'principalism' founded by Beauchamp and Childress (1979) (see Medical ethics §6). Some nursing ethicists have developed the idea that the 'ethics of care', rooted in a feminine response to life, is a superior alternative to a masculine 'ethics of justice'. Certainly the overwhelming majority of nurses are female, but while some take this as evidence of an advantageous feminine moral attitude others have thought it reflects only political contingencies.

Some insist that rationalistic, abstract and liberal 'impartiality' and 'distance' lead to indifference and unresponsiveness in modern medicine, and expound a contrasting 'partialist' nursing ethics of context, personal engagement and unique human relationships (see Impartiality). Nursing theorists influenced by Gilligan, such as Fry and Watson, take up her suggestion that whereas a principles approach seeks agreement, the caring approach seeks mutual understanding through an exploration of the meaning of a personal crisis (illness, disability, infirmity). Gadow (1980) has a distinctively nursing notion of care in her concept of 'existential advocacy': a true nursing relationship involves helping patients make sense of their life experiences so as to regain some control, clarifying the patients' values so as authentically to exercise their freedom of self-determination. The emphasis on the 'carer/cared for' relationship has led one stream of nursing ethics, influenced by Benner, into phenomenological methods.

Other writers have attempted to integrate principles-based approaches with 'caring relationship' approaches. Beauchamp and Childress, admitting that principles may be insufficient for health care ethics, have responded with a clarification of 'mutual interdependence' and 'emotional response' in the 'ethics of care' and still regard the notion of caring as too undeveloped to be integrated coherently. One empirical study finds no significant difference in the distribution of 'impartialist' and 'partialist' modes of thinking among doctors and nurses and suggests the modes really operate together but at different levels.

The question of a difference between medical and nursing ethics, whether a sharp one or merely one of emphasis and tendency, is carried into questions about which health care issues nurses identify as ethically problematic and how they approach them. Nurse ethicists often do, and *qua* nurses arguably should, select issues which concern nursing activity (the assessment, planning, implementation and evaluation of care) as opposed to medical activity (diagnosis, treatment, prognosis). Such issues would be found in, for example, health promotion and preventative nursing, continuing care of the elderly, and residential and hospice care. Given the intimate and ongoing nature of the nursing process (as opposed to the transitory medical intervention) ethical issues are identified in everyday and routine issues, such as incontinence and toileting, falls among the elderly, and choice of clothing and sleep periods.

Ethically aware nurses may more readily see the human, personal, cultural and social aspects of any issue, whether medical or nursing, such as patient self-esteem and privacy, pain alleviation and comfort, empowering and encouraging self-care, security and general wellbeing. Ethical issues in nursing research may relate in content to clinical nursing (such as wound or bed sore management) or may, in form, approach issues through meanings and relationships employing the methods of ethnography or ‘action theory’.

The considerations above also apply to ethical issues in the organization and professional development of nursing. The themes of ‘nurse autonomy’ and ‘advocacy’ are particularly strong. The first concerns the degree to which nurses can exercise moral choice as employees and as the subordinates of doctors. Thus, professional and contractual obligations may come into conflict, for example, over informing the patient in their interest about a matter the employer may regard as commercially confidential (see Professional ethics); and medical judgment and nursing judgment may conflict, for example, over when to discharge a patient. In these contexts the rights and wrongs of ‘blowing the whistle’ may arise. This may even be entailed by the demand that the nurse act as ‘advocate’ of the patient. The interpretation of this notion is controversial. While it requires that the nurse plead the cause of the patient, it is not always clear whether this is to be given the ‘banal’ interpretation that the patient by definition is someone who needs another to do what they would do if able, or the ‘radical’ interpretation that the nurse must challenge and speak up to defend the patient against any obstacle whatsoever to their autonomy (see Autonomy, ethical).

‘Nursing ethics’ may be taken as a generic term to cover the concerns of all the professions allied to medicine. In midwifery ethical issues emerge from maternal choice, the professional autonomy of the midwife *vis-à-vis* obstetrics, and ‘natural childbirth’. In physiotherapy there are questions around delimiting potentially coercive or unavoidably uncomfortable or even futile procedures, such as chest physiotherapy on terminally ill patients.

See also: Applied ethics; Bioethics; Help and beneficence

GEOFFREY HUNT

References and further reading

Beauchamp, T.L. and Childress, J.F. (1979) *Principles of Biomedical Ethics*, Oxford: Oxford University Press; 4th edn, 1994.(Definitive work of ‘principalism’, criticized by many nursing ethicists.)

Benner, P. (ed.) (1994) *Interpretive Phenomenology: Embodiment, Caring and Ethics in Health and Illness*, London: Sage.(A phenomenological approach to nursing ethics. Not easy reading.)

Fry, F.T. (1989) ‘Toward a Theory of Nursing Ethics’, *Advances in Nursing Science* 11 (4): 9-22.(Argues for a moral-point-of-view theory with caring as a fundamental value.)

Gadow, S. (1980) ‘Existential Advocacy: Philosophical Foundations of Nursing’, in S.F. Spiker and S. Gadow (eds) *Nursing Images and Ideals: Opening Dialogue with the Humanities*, New York: Springer.(Draws on existentialist assumptions.)

Gilligan, C. (1982) *In a Different Voice: Psychological Theory and Women’s Development*, Cambridge, MA: Harvard University Press.(Very strong influence on nursing ethics.)

Hunt, G. (1994) *Ethical Issues in Nursing*, London: Routledge.(Attempts to identify obstacles to good practice rather than formulate a theory of good practice as such.)

Noddings, N. (1984) *Caring: A Feminine Approach to Ethics and Moral Education*, Berkeley, CA: University of California Press.(Strong influence on nursing ethics.)

Rickard, M., Kuhse, H. and Singer, P. (1996) ‘Caring and Justice: A Study of Two Approaches to Health Care Ethics’, *Nursing Ethics* 3 (3): 212-23.(An empirical study of approaches to ethical decision-making in nurses and doctors.)

Watson, J. (1988) *Nursing: Human Science and Human Care: A Theory of Nursing*, New York: National League for Nursing.(An exploration of the meanings embedded in caring.)

Nursing Ethics: An International Journal for Healthcare Professionals.(Six issues a year, covering all aspects of nursing ethics and philosophy, with some law.)

Nyāya-Vaiśeṣika

*The Nyāya school of philosophy developed out of the ancient Indian tradition of debate; its name, often translated as 'logic', relates to its original and primary concern with the method (*nyāya*) of proof. The fully fledged classical school presents its interests in a list of sixteen categories of debate, of which the first two are central: the means of valid cognition (perception, inference, analogy and verbal testimony) and the soteriologically relevant objects of valid cognition (self, body, senses, sense objects, cognition, and so on). The latter reflect an early philosophy of nature added to an original eristic-dialectic tradition. On the whole, classical Nyāya adopts, affirms and further develops, next to its epistemology and logic, the ontology of Vaiśeṣika. The soteriological relevance of the school is grounded in the claim that adequate knowledge of the sixteen categories, aided by contemplation, yogic exercises and philosophical debate, leads to release from rebirth. Vaiśeṣika, on the other hand, is a philosophy of nature most concerned with the comprehensive enumeration and identification of all distinct and irreducible world constituents, aiming to provide a real basis for all cognitive and linguistic acts. This endeavour for distinction (*viśeṣa*) may well account for the school's name. Into the atomistic and mechanistic worldview of Vaiśeṣika a soteriology and orthodox ethics are fitted, but not without tensions; still later the notion of a supreme god, whose function is at first mainly regulative but later expanded to the creation of the world, is introduced. In the classical period the Vaiśeṣika philosophy of nature, including the highly developed doctrine of causality, is cast into a rigorous system of six, later seven, categories (substance, quality, motion, universal, particularity, inherence, nonexistence). Nyāya epistemology increasingly influences that of Vaiśeṣika.*

The interaction and mutual influences between Nyāya and Vaiśeṣika finally led to the formation of what may be styled a syncretistic school, called Nyāya-Vaiśeṣika in modern scholarly publications. This step, facilitated by the common religious affiliation to Śaivism, occurs with Udayana (eleventh century), who commented on texts of both schools. Subsequently, numerous syncretistic manuals attained high popularity. Udayana also inaugurated the period of Navya-Nyāya, 'New Logic', which developed and refined sophisticated methods of philosophical analysis.

1 Sources

The basic Vaiśeṣika text is the *Vaiśeṣikasūtra*, ascribed to the legendary founder of the school, Kaṇāda ('grain-eater'; also Kaṇabhuj, and Kāśyapa by clan name). This compilation of some 400 concise and often enigmatic mnemonic sentences contains materials from several developmental stages of the system; substantial parts of it may have been in existence already during the first two centuries AD. Old commentaries are referred to in later works, but have for the most part been lost. Mainstream classical Vaiśeṣika largely goes back to the *Padārthadharmasaōgraha (Compendium of Properties of the Categories)* - also called *Praśastapādabhāṣya (Praśastapāda's Commentary)* - of Praśastapāda (early sixth century), the great systematizer and scholiast of the school. Another compendium from approximately the same period, the innovative *Treatise on the Ten Categories* by Candramati (surviving in Chinese translation), had no lasting impact. The major works of the classical period are commentaries on the *Padārthadharmasaōgraha*: the *Vyomavatī(The Spacious [Commentary])* by Vyomaśiva (ninth century), the *Nyāyakandalī(The White-Flowered Tree of Nyāya)* by Śrīdhara (late tenth century) and the *Kiraṇāvalī(A Sequence of Rays)* by Udayana (early eleventh century). A commentary on the *Vaiśeṣikasūtra* by Candrānanda (probably ninth or tenth century) was published only in 1961; although strongly influenced by the *Padārthadharmasaōgraha*, it preserves a more genuine version of the text than the until then standard commentary by Śaōkara Miśra, the *Vaiśeṣikasūtra-Upaskāra (The Ornament of the Vaiśeṣikasūtra; fifteenth century)*.

At the basis of Nyāya is the *Nyāyasūtra*, ascribed to Akṣapāda. Its earliest preserved commentary, Pakṣilasvāmin Vātsyāyana's *Nyāyabhāṣya (A Commentary on Nyāya)*, dates from the second half of the fifth century. Its subcommentary, the *Nyāyavārttika (A Supplementary Commentary on Nyāya)* by Uddyotakara Bhāradvāja (sixth century), represents a defence and renewal of Nyāya in response to Buddhist logic. Of many commentaries on these two texts, only Vācaspati Miśra I's voluminous *Nyāyavārttikatātparyaṭīkā (A Commentary on the True Intention of the Nyāyavārttika*; tenth century) has survived. It is commented upon by Udayana in his *Nyāyavārttikatātparyapariśuddhi (The Clarification of the True Intention of the Nyāyavārttika)*. Important independently composed treatises of the classical period are Jayanta Bhaṭṭa's *Nyāyamañjarī(A Cluster of Blossoms of Nyāya*; late ninth century) and Bhāsarvajña's *Nyāyasāra (The Essence of Nyāya)*, with the author's own

extensive commentary, *Nyāyabhūṣaṇa (The Ornament of Nyāya*; tenth century). Udayana's *Ātmatattvaviveka (The Discernment of the True Nature/Reality of the Self)* mainly defends the notion of a permanent self against Buddhist criticism, whereas his *Nyāyakusumāñjali (A Handful of Flowers of Nyāya)* establishes the existence of a supreme god. Of Navya-Nyāya works, Gaōgeśa's monumental *Tattvacintāmaṇi (The Jewel of Reflection on the True Nature [of Things]/Reality*; fourteenth century) deserves special mention (see Gautama, Akṣapāda; Vātsyāyana; Uddyotakara; Udayana; Gaōgeśa).

2 Origin and orientation of Vaiśeṣika

The beginnings and development of Vaiśeṣika prior to its classical form are still largely obscure. The most valuable and knowledgeable attempt so far at a comprehensive 'archaeology' of Vaiśeṣika is the brilliant, though necessarily speculative and sometimes intuitive reconstruction of the school's formation by Erich Frauwallner. About thirty years after the publication of his *Geschichte der indischen Philosophie (History of Indian Philosophy)*, volume II (1956), Vaiśeṣika studies experienced a renaissance, and scholars turned again to his stimulating reconstruction. Some developments traced by him have been reconsidered; his evaluation of the original orientation of the school has also become the subject of debate. According to Frauwallner, early Vaiśeṣika is an outstanding example of a philosophy of nature that strives systematically to understand and explain the world out of a pure philosophical urge for knowledge, without mythological, theological or soteriological interests. Within the broader tradition of philosophy of nature in which Vaiśeṣika arose, Frauwallner also locates the materialist schools (see Materialism, Indian school of). Other philosophical schools belonging to that tradition, however, notably Jainism, Nyāya and Mīmāṃsā, combined philosophy of nature with soteriological and ethical teachings. Frauwallner assumes that Vaiśeṣika too, during its development, was influenced by the religious and moralistic tendencies of contemporary philosophies and eventually added a soteriology, still later a theology, to its teachings. These developments he perceives in stark contrast to the school's original spirit and considers them to have had negative consequences for it. Some modern Indian interpreters, notably Anantalal Thakur (1969), also stress the purely scientific attitude of the *Vaiśeṣikasūtra*.

The other extreme lies in ascribing an original and primary soteriological concern to early Vaiśeṣika (for example, Madeleine Biardeau) or a general religious orientation by claiming its main purpose to be the defence of the Vedas and of orthodox ritualism and ethics against heterodox challenges (Jan Houben 1994). Because sources for the school's early history are lacking, it is safer to assume a middle stand, leaning towards Frauwallner's side. The basic orientation of early and even classical Vaiśeṣika is certainly towards cosmology, the systematic and comprehensive explanation of the world, its origin and the processes, natural, psychological, and so forth, occurring through the interaction of its constituents. Some rudimentary soteriology, however, may have been attached already to the early teachings, even if it was an extraneous element of minor importance. The very acceptance of the notion of an individual soul would necessitate in its historical context some ideas about the soul's involvement in rebirth and its possible release from it.

A similar situation can be assumed for those teachings in the epic *Mahābhārata* that have a pronounced character of philosophies of nature. Their predominant topics are the origin and constitution of the world, centring around the concept of the five 'great elements', earth, water, fire/light, wind/air and space/'ether'. The elements are respectively characterized by firmness, moistness, and so on, and, among other qualities, by the 'special qualities' that are the five sense objects, namely, smell, flavour, form, touch and sound, which are sometimes conceived as the emanations of the elements. Various qualities to be subsumed under these five 'special qualities' are enumerated in long and heterogenous lists. All things of this world, including the body and the five senses, are made up of the five elements. The generally strong interest in the empirical human being is manifest in theories about the formation of the body, its constitution by the elements and its internal processes, such as digestion. Attention is also paid to the explanation of cognitive processes. The five senses are recognized as different from their seats, that is, the sense organs; according to the principle 'like grasps like', they grasp the five 'special' elementary qualities and consist of the element which possesses the respective quality. An internal organ or sense (*manas*) functions in non-sensory cognitive processes and rules over the activities of the external senses. Further entities in the world are the individual souls, the sentient and active subjects within the bodies. They are responsible for life, and due to involvement with the sense objects and nescience they become the subjects of transmigration. Time and the cardinal directions appear also as important factors in the world.

It is within the milieu of these teachings that one can situate the beginning of Vaiśeṣika. During its formation they were further developed and refined; in particular, the entities perceived as primary constituents of the world were systematized, some of them subsumed elsewhere, some accorded independent status, and some eliminated, towards an ordered and complete enumeration of irreducible world constituents.

3 Origin and formation of Nyāya

The roots of the Nyāya school are in the ancient Indian tradition of debate (*vāda*). In the Vedic period, priests already held verbal contests as part of ritual, to discover and appropriately formulate the enigmatic true nature of the world and of the rituals themselves. Subsequently, the Upaniṣads refer to religious-philosophical debates at the assemblies of local rulers. The winners were rewarded not only with recognition and fame, but also with material goods and royal patronage. Thus rules of debate had to be fixed, and guidelines were compiled to assure an orderly and fair procedure. They also provided criteria to determine sound and unsound arguments and thus winners and losers. These early guidelines developed into manuals which covered heterogenous points, such as the characteristics of the opponent and composition of the audience with respect to psychological tactics, types of debate, manner of speech, the parts of an argument and their order, means of valid cognition employed in an argument, types of faulty reasons, unfair criticism, and various points of defeat. The central position, however, was occupied increasingly by consideration of the method (*nyāya*) of proof. Epic references to numerous ‘models of method’ (*nyāyatantra*) pronounced by debaters may point to this tradition.

The tradition of debate was closely related to the branch of knowledge or ‘science’ concerned with principled and reasoned reflection (*ānvīkṣikīvidyā*), or even led to its development. Further designations related to this ‘science’ of argumentative corroboration or disproof also refer to its special concern and general method, which, like the term ‘nyāya’, is often rendered as ‘logic’: ‘the science of turning around (in examination)’ (*tarkavidyā*, *tarkaśāstra*), ‘the science of (formal and logical) coherence’ (*yuktiśāstra*) and ‘the science of reasons’ (*hetuvidyā*, *hetuśāstra*).

At first the manuals of debate and reasoned examination may have been formulated without presupposing specific metaphysics, and therefore could be used by different religious and nascent philosophical traditions. However, the central concern with the method of proof led to increasing reflection on the means of valid cognition and logical issues, and eventually certain traditions of debate became associated with certain philosophical outlooks. Further, in the case of Nyāya, the reflection on the *objects* of valid cognition, one of the categories of debate, initiated the inclusion of an early philosophy of nature, similar to the one from which Vaiśeṣika developed, into an originally eristic-dialectic tradition. With the addition of a plain soteriology, the process of the formation of a separate philosophical school was concluded.

4 Ontology: the three older categories

The *Vaiśeṣikasūtra* presents a developed philosophy of nature clad in a rigorous doctrine of categories (*padārtha*, literally ‘meaning of a word’). The six categories fall into two triads, of which the first may historically precede the second: (I) substance, (II) quality and (III) motion; (IV) universal, (V) particularity and (VI) inherence. Further categories, namely potentiality and nonpotentiality, existence (the highest universal) and nonexistence, were introduced by Candramati, but not generally accepted. Nonexistence (*abhāva*), though, increasingly attracted philosophical interest and was finally adopted as a seventh category in the eleventh century (see Negative facts in classical Indian philosophy §5). As elaborated by Wilhelm Halbfass (1992), common to everything comprised by the six classical or seven postclassical categories are the three abstract attributes of ‘is-ness’ (*astitva*), that is, objectivity or factual identifiability, knowability and nameability. This points to the basic endeavour of Vaiśeṣika, the enumeration and identification of what there is in order to provide a real basis for all cognitive and linguistic acts. However, only substances, qualities and motions exist in the fullest sense, that is, as concrete ‘objects’ or ‘things’ (*artha*), because the highest universal, existence (*sattā*, literally ‘being-ness’, ‘real-ness’; or *bhāva*, literally ‘becoming’, but used synonymously), inheres in them. Universals, particulars, inherence and nonexistence are also ontologically distinct and irreducible world constituents; in contrast to the first three categories, they can be substrata only of abstract properties (*dharma*), but not of further ontologically founded entities, including the universal ‘existence’ - that is, they can function as logical subjects (*dharmin*) only.

(I) The category of substance (*dravya*) comprises nine items. The first four are (1-4) the four elements, earth,

water, fire and wind, which exist in the form of eternal atoms and impermanent aggregates. The fifth element, (5) 'ether' (*ākāśa*), functions primarily as the eternal substratum of sound; it is inferred by elimination from the very existence of the quality sound, which requires a substantial support. Because sound can arise everywhere, 'ether' is considered as one and omnipresent. In such qualities, including eternity, 'ether' resembles time, space and the individual souls or selves. (6) Time (*kāla*) is conceived as a timeless, static and inactive substance responsible for our notions of relative position in time and the qualities of temporal remoteness and proximity in limited and noneternal substances. Similarly, (7) space (*diś*) accounts for our notions of relative spatial position and remoteness and proximity in space. The ubiquitous (8) selves (*ātman*) are restricted to individual bodies as the places of their experiences and activities only because their atom-sized and eternal (9) 'internal organs' (*manas*) are connected with a body. However, the self may have originally been considered as having limited dimension, that is, as being contained within the body, and thus capable of motion and physical activity.

The impermanent substances, the objects of everyday experience, are wholes (*avayavin*) consisting of a single element each, with the possible admixture of further elements. These wholes constitute something new over and above their parts (*avayava*), a claim contested especially by the Buddhists. According to the classical theory of causality, the smallest whole is constituted by two conjoined atoms; the first visible whole is the sun mote or triad, composed of three dyads. All further substances are produced by various numbers of larger, visible parts.

(II) The qualities (*guṇa*) are unrepeatable quality individuals, particularizations in time and space of abstract qualities such as colour or flavour. They are ontologically distinct from the substances in which they inhere, but dependent on them as their substrata. Caused in the whole substance by the qualities of its parts, they are destroyed with its destruction. Some qualities inhere in only one substance, such as temperature, colour, flavour and smell, the 'special qualities' of the physical elements. They accumulate in the eternal atoms of wind, fire, water and earth, where they too are eternal: wind possesses only (tepid) temperature, fire (hot) temperature and (bright white) colour, and so on. Other qualities inhere in more than one substance simultaneously, for example, the 'common qualities' conjunction and separation, important factors in causation, and number (above one). The list of seventeen qualities in the *Vaiśeṣikasūtra* further enumerates the 'common qualities' dimension (ranging from atomic size to large size to infinite size, that is, ubiquity), separateness, and remoteness and proximity (in space and time). Further 'special qualities', namely of the self, are cognition, pleasure and pain, desire and aversion, and effort, all impermanent. It is by effort that the omnipresent self, incapable of movement itself, initiates mechanistically conceived psychological and physical processes. To these seventeen seven more are added by Praśastapāda's time: weight, fluidity, viscidity, disposition (comprising momentum, elasticity, and impressions responsible for recollection left in the self by cognitions), merit (*dharma*) and demerit (*adharma*), and sound, the 'special quality' of 'ether'.

It is not observed that, for example, colours possess further colours. Moreover, if they did, this would result in an infinite regress. Motion, on the other hand, is only observed with corporeal things. Therefore, all qualities are devoid of further qualities or motions.

(III) The five items in the category of motion (*karman*) go back to a classification of human actions: throwing upwards and downwards, bending and stretching, and going. Subsequently all other types of motion, inhering in atoms, the 'internal organs' and impermanent substances, are subsumed under 'going'; motion itself is momentary and inheres in only one substance at a time. It plays an essential role in the atomistic-mechanistic world view of Vaiśeṣika, for example, in causation, to effect the conjunction between substances to produce new wholes (see Causation, Indian theories of §5), and in the mechanistic explanation of natural processes, macrocosmic (for example, meteorological) and microcosmic (for example, digestion, birth and death, cognition), often effected by 'the invisible' (*adṛṣṭa*), a force also at work in the ritual and ethical context (see §7) and later identified with merit and demerit inhering in the self. A complete chapter of the *Vaiśeṣikasūtra* is devoted to mechanics in the broadest sense, although a science of mechanics, as in ancient Greece, did not evolve.

Like qualities, motions are ontologically separate from, but dependent on, supporting substances; they are also devoid of qualities or further motions.

5 Ontology: the younger categories

(IV) Universals (*sāmānya*, literally 'commonness', 'sameness'/'similarity') inhere in substances, qualities and

motions; they are the hypostasized generic properties responsible for the connection of various entities with the same concept or designation and for their conception as belonging to the same group. Contrary to qualities, they are one, eternal and repeatable; for example, the one universal 'cow-ness', characterized as having horns, a dewlap, a tail, and so on, resides fully and simultaneously in every cow alive and is unaffected by the demise of an individual cow. Although dependent on substances, qualities and motions as their substrata, universals provide them with their generic character and identity. The highest universal is 'being-ness' or existence, which is common to substances, and so on; the lower or specific universals (*sāmānyaviśeṣa*), such as 'substance-ness', 'colour-ness' and 'hot-ness', effect not only a notion of recurrence and inclusion of their substrata in one class, but also of their exclusion from entities belonging to other classes. Special attention is devoted to the differentiation of real universals from mere abstract properties (*dharma*); by Udayana's time, criteria have been developed to discern real from unreal universals (see Universals, Indian theories of §2).

(V) The elusive category particularity (*viśeṣa*) or ultimate particularity (*antyaviśeṣa*), as clarified by Halbfass (1992), comprises, according to Praśastapāda, the 'factors of irreducible individual identity'. These particularities assert timeless, unchangeable generic and numeric identity for permanent entities, notably the individual atoms, selves and 'internal organs'; they allow accomplished Yogis to discern individual atoms of earth with their special powers of perception, for example. Particularities also inhere in the unitary substances 'ether', time and space. For Candramati, they alone possess particularity, inasmuch as their generic identity cannot be asserted by specific universals: there are no other entities in their class to be included.

(VI) Inherence (*samavāya*, literally 'concourse') is the unitary eternal entity responsible for the connection of distinct entities which are in a relationship of support and supported and cannot exist separately. It causes the notion that something is 'here', that is, resides in something else; for example, a product or whole, such as a cow, inheres in its causes or parts, such as horns, dewlap, tail, and so on. The entities belonging to different categories are joined by way of this dependent residence to form the actual complex things of this world. At the base are the substances; due to inherence the universals existence, substance-ness and, for example, cow-ness reside in a substance, as do its qualities, such as white colour, and motions, that is to say going. White colour and going are again related to existence and their respective generic characters by inherence.

The ontology of classical Nyāya closely resembles that of Vaiśeṣika. This is due to their similar background regarding philosophy of nature (see §3). The historically later chapters of the *Nyāyasūtra* basically presuppose and assert classical Vaiśeṣika ontology in their discussions of the means and (soteriologically relevant) objects of valid cognition. The Vaiśeṣika categoriology, however, was fully adopted and further developed starting with Uddyotakara only. The lasting points of difference are mainly of a scholastic nature; a major one concerns the mode of colour change in fired pots. According to Vaiśeṣika, the fire destroys the pot up to its atoms, which lose their original dark colour due to the heat; subsequently the heat causes a new colour in them, and thanks to 'the invisible' a similar pot is built up again, with a new reddish colour. Nyāya, on the other hand, allows tiny interstices in the structure of the pot into which the fire can intrude; its heat then causes a colour change of the *whole* pot, which remains structurally intact and the same throughout the process. The independently minded thinker Bhāsarvajña argued for revisions of the categories, but without lasting impact. In his innovative *Padārthatattvanirūpaṇa (Examination of the True Nature/Reality of the Categories)*, the sixteenth-century Navya-Naiyāyika Raghunātha Śiromaṇi eliminates 'ether', time and space, for example, as separate categories by ascribing their roles to God, and introduces new categories such as the moment and potency.

6 Epistemology, logic and dialectics

The very composition of the *Nyāyasūtra* demonstrates the traditionally central interest of Nyāya in these areas. Among the sixteen categories of debate in Nyāya, the means or instruments of valid cognition (*pramāṇa*) occupy the first place. The larger number of *pramāṇas* current at the time is reduced to four: (1) perception (*pratyakṣa*), (2) inference (*anumāna*), (3) analogy or comparison (*upamāna*), and (4) verbal testimony (*śabda*). In Vaiśeṣika, the *pramāṇas* are treated under the category of quality because cognition is a 'special quality' of the self. Cognition is divided into knowledge and nescience; knowledge includes perception and inference as *pramāṇas*, as well as recollection and the intuitional cognition of omniscient seers. Analogy and verbal testimony are subsumed under inference. Both schools distinguish between, for example, perception as a means and the resulting knowledge (*pramā*); in the controversy about the intrinsic or extrinsic validity of the resulting knowledge they

adopt the latter position. They further hold that the *pramāṇas*, although of different natures and modes of functioning, can converge on the same object, contrary to the view that their objects are exclusive (see Epistemology, Indian schools of §2; Knowledge, Indian views of §3).

(1) *Perception* (as *pramā*) arises from a contact between self, sense, 'internal organ' (*manas*) and object (= perception as *pramāṇa*). The atom-sized, swiftly moving 'internal organ' serves as the regulating link between the omnipresent self and the elementary-materially conceived senses (see §2); this explains the denial of the simultaneity of different sense perceptions and cognitions in general. The senses either move out to their objects, as do the visual rays, or are approached by them. A differentiation into six types of contact was necessitated by the introduction of Vaiśeṣika categoriology to account for the perception of, say, colour, which as a quality cannot be the substratum of a further quality conjunction (see Sense perception, Indian views of).

Large substances that possess colour are perceptible by vision and touch as something over and above their qualities. Some thinkers even maintained that their perception without any qualities is sometimes possible. Generally they appear with other substances, qualities and motion as their qualificands; due to the perception of their specific universals they appear qualified as similar to others of their class. At first, only the 'special qualities' of the self were assumed to be perceptible by a specific contact between the 'internal organ' and the self; from them its existence had to be inferred. However, several thinkers also claimed its self-perception through this contact. The other perceptible qualities and motion are perceived only when inhering in perceptible substances; this perception is further dependent on perception of their specific universals, whereas the perception of universals, lacking further universals, merely follows that of their substrata. Inherence is only inferable in Vaiśeṣika, whereas in Nyāya it is perceived as qualifying its relata, like nonexistence.

Unlike Dignāga, certain Naiyāyikas (the Ācāryas, commentators on the *Nyāyavārttika*) maintained that perception can be conceptual, that is, that language/concepts in the form of impressions may be part of its causal complex and may influence its content. In doing so, they gave up the realistic principle, defended by the Vyākhyātṛs (commentators on the *Nyāyabhāṣya*), that the external and internal objects of perception have to be equivalent.

(2) *Inference* depends on perception and is threefold: (a) 'as before', (b) 'as the rest' and (c) 'observed from similarity'. The *Nyāyasūtra* provides no further explanations here, and all the commentaries seem unreliable. Albrecht Wezler (1969) has argued convincingly that parallel passages in early Buddhist texts preserve the appropriate examples: (a) refers to recognition by some special mark, (b) to inductive generalization and (c) to analogical reasoning. Vātsyāyana provides two alternative explanations for each type, none of which fits the *sūtra*'s terminology smoothly. For instance, (b) is explained as inference either from effect to cause or by elimination.

Perhaps the original intention of the *sūtra* was lost because inference was just one among many topics in manuals of debate (see §3) and numerous attempts to characterize it were current; for example, the *Vaiśeṣikasūtra* refers to four situations in which entities can function as sound reasons or signs to infer others: A is connected (to B), A is related by inherence (to B), A is related by inherence to the same thing (as B), and A is in contradiction (to B). As a *pramāṇa*, inference appears in the *Nyāyasūtra* in a category different from the members of proof. Only after Vasubandhu merged inference and proof did the former become central in Indian philosophy. Subsequently, the two are distinguished in Nyāya and Vaiśeṣika only in terms of purpose, as 'inference for oneself' and 'inference for another'. The old notions of inference are largely forgotten, and the structure of proof becomes basically that of inference, dominated from Praśastapāda and Uddyotakara onwards by the concept of three characteristics (trairūpya) of a valid reason and refined by Trilocana's concept of an essential relation between probans and probandum (see Inference, Indian theories of §§4-6).

The members of proof are proposition, reason, illustration (by similar or dissimilar examples), application and conclusion. For example:

Sound is impermanent,
because it is produced,
like a pot (which is produced and impermanent).
Sound is like that (that is, produced).
Therefore, it is impermanent.

A dissimilar example would be the self, which is not produced, and is permanent.

For Nyāya, there are twenty-four types of sophistries with which to attack such an argument unfairly, and twenty-two ways of losing a wrangle (‘points of defeat’), of which the use of fallacious reasons is most important philosophically. Fallacious reasons appear also as a separate category, which consists of the deviating or inconclusive reason, the reason contradicting a tenet, the reason which is the same as the original issue, the reason which is the same as the probandum, and the reason which has passed the proper time (for enunciation). The *Vaiśeṣikasūtra* also mentions two fallacious reasons (‘unestablished nonstatements of a reason’), the unreal/untrue and the doubtful. Subsequently, all of them are reinterpreted, sometimes drastically, and refined within the framework of *trairūpya*. The fallacious reasons of Vaiśeṣika are equated with the first two of Nyāya (in reverse order), and Praśastapāda adds a third type, the reason which does not lead to any decision; it subsumes the too specific reason, a subtype of the Nyāya inconclusive reason. The last two items in Nyāya, in their updated forms, are adopted also in later Vaiśeṣika.

(3) By *analogy* one knows the connection of a certain designation with an object. Analogy as *pramāṇa* is hardly ever used in philosophical discussion; it is employed very narrowly and should not be confused with cognition of similarity, which is obtained by perception and inference.

(4) *Verbal testimony* is defined as instruction by a reliable person and includes everyday linguistic communication and scriptures. Correct reasoning cannot be in conflict with the Vedic scriptures; yet their argumentative corroboration, even validation, is proclaimed to be the central task of Nyāya. In practice, however, Nyāya is more concerned with the nature of articulated sound *per se*, its duration, and the relationship between words and their referents. On all these issues Nyāya held its own views and strongly opposed not only Buddhists, but also Mīmāṃsakas and Sāōkhyas (see Testimony in Indian philosophy).

7 Soteriology and ethics

Both schools claim that knowledge of their categories leads to release (*apavarga*). In Nyāya, Vātsyāyana narrowed this knowledge to adequate understanding of the twelve soteriologically relevant objects of valid cognition; appearing as a separate category, they indicate the more harmonious integration of soteriology in early Nyāya than in Vaiśeṣika (see §§2-3). The *Nyāyasūtra* proclaims that with the absence of wrong understanding (which absence is caused by adequate understanding of the categories), the three faults (*doṣa*) - craving, aversion and delusion - disappear, and subsequently the linguistic, mental and physical activity instigated by them ceases; the cessation of activity means that a result of activity will not arise and thus not effect a new life for it to be experienced by the acting subject. Consequently - suffering being found nowhere but in life - no suffering will occur; release, which is defined as final and complete liberation from suffering, is attained. According to the *Vaiśeṣikasūtra*, liberation (*mokṣa*) technically consists in the absence of all conjunctions concerning the self and involving it in pleasure and pain, and in the nonmanifestation of a new body for it. In liberation, all ‘special qualities’ of the self vanish. This idea of liberation evoked strong criticism, especially from Vedāntists, who compared it to the undesirable state of an unconscious stone.

Although the faults or moral defilements (*kleśa*) may be natural to the self, release is possible, just as atoms of earth can lose their natural black colour when a pot is fired (see §5). Furthermore, because the faults are conditioned by wishful expectative thinking relating back to previous experience, their naturalness is doubtful, and they can be wiped out by elimination of this conditioning through adequate understanding. Adequate understanding is initiated by a preceptor and attained by contemplation and yogic exercises repeatedly practised in quiet secluded spots, accompanied by perfection of the self with the help of general and specific rules of restraint (*yama*, *niyama*). The grasp on knowledge has to be trained incessantly; here friendly debate with knowledgeable disciples, teachers, fellow students and others striving for salvation is recommended. Even wrangle and dispute involving mere criticism of the other party, the other two types of debate, can serve to protect and strengthen the ascertainment of the true nature of things. Unlike the *Nyāyasūtra*, the *Vaiśeṣikasūtra* proclaims that orthodox, ritual-dominated merit (*dharma*) accomplishes uplift in subsequent lives and liberation. The elusive sixth chapter treats orthodox ethical and ritual conduct, specifying means that effect a certain purpose or result which is called ‘the invisible’ (*adṛṣṭa*; see §4) and that are conducive to uplift. However, in contrast to the initial proclamation, the activity of merit and demerit is stated as the cause of the self’s involvement in rebirth, which highlights the complex and problematic integration of soteriology and ethics. The notion of a supreme god, whose function

within the atomistic-mechanistic worldview of Vaiśeṣika is mainly regulative, is found first in Praśastapāda's work.

See also: Hindu philosophy; Ontology in Indian philosophy

ELI FRANCO
KARIN PREISENDANZ

References and further reading

Bhaduri, S. (1947) *Studies in Nyāya-Vaiśeṣika Metaphysics*, Bhandarkar Oriental Series 5, Poona: Bhandarkar Oriental Research Institute.(Surveys and explicates basic issues regarding the category of substance - the doctrine of elements, time and space, atomism and causality.)

Bronkhorst, J. (ed.) (1994) *Proceedings of the Panel on Early Vaiśeṣika, Hong Kong, Asiatische Studien/Études Asiatiques* 48 (2).(Wide-ranging collection of essays by specialists on various aspects and problems of early Vaiśeṣika.)

Candramati (early 6th century) *Treatise on the Ten Categories*, trans. H. Ui, *The Vaiśeshika Philosophy according to the Daśapadārtha-Śāstra*, ed. F.W. Thomas, London: Royal Asiatic Society, 1917; repr. Chowkhamba Sanskrit Studies 22, Varanasi: The Chowkhamba Sanskrit Series Office, 1962.(A study of Vaiśeṣika and Nyāya doctrines as reported in early Buddhist literature is prefixed to an annotated translation of Candramati's treatise on the ten categories; also contains the Chinese text.)

Chemparathy, G. (1983) *L'autorité du Veda selon les Nyāya-Vaiśeṣikas*, Conférences et travaux 2, Louvain-la-Neuve: Centre d'histoire des religions.(Examines the attitude especially of Nyāya towards the Vedic scriptures and socio-religious orthodoxy.)

Faddegon, B. (1918) *The Vaiçeṣika-System, Described with the Help of the Oldest Texts*, Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam, Afdeeling Letterkunde, Nieuwe Reeks XVIII 2, Amsterdam: Johannes Müller.(The most comprehensive classic on all aspects of Vaiśeṣika; contains copious translations from the *Vaiśeṣikasūtra*, detailed discussions of the *Padārthadharmasaōgraha* and important extracts from Śrīdhara's *Nyāyakandalī*. Although written before the discovery of Candrānanda's commentary, it is still a valuable and thoughtful exposition.)

Frauwallner, E. (1955) 'Candramati und sein *Daśapadārthaśāstra*', in O. Spies (ed.) *Studia Indologica*, Bonner Orientalische Studien, Neue Serie 3, Bonn: Selbstverlag des Orientalischen Seminars der Universität Bonn, 65-85; repr. in G. Oberhammer and E. Steinkellner (eds) *Kleine Schriften*, Wiesbaden: Franz Steiner, 1982, 202-22. (Evaluates Candramati's innovative treatise and places it in its historical context.)

Frauwallner, E. (1956) *Geschichte der indischen Philosophie (History of Indian Philosophy)*, vol. 2, Reihe Wort und Antwort 6/II, Salzburg: Otto Müller Verlag, 15-250.(A most valuable reconstruction of the origins and development of Vaiśeṣika, with an exposition of the earliest philosophy of nature in the Nyāya school; also a clear, detailed and reliable presentation of the classical Vaiśeṣika system, including the first stages of scholasticism.)

Guha, D.C. (1968) *Navya Nyāya System of Logic*, Delhi: Motilal Banarsidass, 2nd rev. edn, 1979.(Explains and exemplifies the basic terminology and techniques of Navya-Nyāya philosophical analysis.)

Halbfass, W. (1991) *Tradition and Reflection: Explorations in Indian Thought*, Albany, NY: State University of New York Press.(Chapter 9, 'Competing Causalities: Karma, Vedic Rituals, and the Natural World', explores the concept of *adṛṣṭa* and its ambiguous position between philosophy of nature and ethics in its historical setting.)

Halbfass, W. (1992) *On Being and What There Is: Classical Vaiśeṣika and the History of Indian Ontology*, Albany, NY: State University of New York Press.(Clearly and fascinatingly written philosophical inquiry into the main issues of Vaiśeṣika ontology; insightful and extensively documented. Contains translations of key passages on Vaiśeṣika ontology.)

Hattori, M. (1972) 'Praśastapāda and Dignāga: A Note on the Development of the Vaiśeṣika Theory of *Anumāna*', *Wiener Zeitschrift für die Kunde Südasiens* 16: 169- 80.(On the relationship of Praśastapāda's theory of inference and fallacious reasons to early Buddhist logic.)

Houben, J.E.M. (1994) 'Liberation and Natural Philosophy in Early Vaiśeṣika: Some Methodological Problems', in J. Bronkhorst (ed.) *Proceedings of the Panel on Early Vaiśeṣika, Hong Kong, Asiatische Studien/Études Asiatiques* 48 (2): 711-48. (Reviews the controversial standpoints on the original orientation of Vaiśeṣika and argues for a new methodological approach to the problem.)

Isaacson, H. (1993) 'Yogic perception (*yogipratyakṣa*) in early Vaiśeṣika', *Studien zur Indologie und Iranistik* 18: 139-60.(Shows the development of the theory of yogic perception within the *Vaiśeṣikasūtra* up to

Praśastapāda.)
Jayanta Bhaṭṭa (late 9th century) *Nyāyamañjarī*, trans. J.V. Bhattacharyya, *Jayanta Bhaṭṭa's Nyāya-Mañjarī[The Compendium of Indian Speculative Logic]*, vol. 1, Delhi: Motilal Banarsidass, 1978.(Being an English translation of a Bengali translation of the Sanskrit original, this is often more of a vague paraphrase than an accurate translation.)
Joshi, L.V. (1986) *A Critical Study of the Pratyakṣa Pariccheda of Bhāsarvajña's Nyāyabhūṣaṇa*, Research Work Publication Series 22, Ahmedabad: School of Languages, Gujarat University.(A detailed annotated paraphrase of the section on perception in one of the most interesting Nyāya texts; Bhāsarvajña takes up the most important theories of his time and discusses them in an independently minded manner.)
Matilal, B.K. (1986) *Perception: An Essay on Classical Indian Theories of Knowledge*, Oxford: Clarendon Press.(Wide-ranging collection of analytically and comparatively orientated essays on the means of valid cognition in general, the distinction between knowledge and illusion, and the relationship of perception to the external world; special emphasis is put on Nyāya views.)
Matilal, B.K. (1990) *Logic, Language and Reality: Indian Philosophy and Contemporary Issues*, Delhi: Motilal Banarsidass, esp. chaps 1-2.(Chapter 1 examines basic issues of early and classical Nyāya, such as the categories of debate, the definition of inference, and fallacious reasons, and chapter 2 discusses topics such as empty terms, universal concomitance in inference, and the role of definition in Navya-Nyāya. The author endeavours to reassess these and other issues in the setting of contemporary Western philosophy.)
Mesquita, R. (1995) 'Der *Apavarga*-Begriff bei Śrīdhara: Eine vedāntische Erlösungslehre?' (Śrīdhara's Concept of Apavarga: A Vedāntic Theory of Release?), in M. Juntunen, W.L. Smith and C. Suneson (eds) *Sauhṛdyamaōgalam: Studies in Honour of Siegfried Lienhard on his 70th Birthday*, Stockholm: The Assocation of Oriental Studies, 215-58.(Detailed examination of Śrīdhara's exceptional position in Nyāya-Vaiśeṣika regarding the nature of the liberated self.)
Narain, H. (1976) *Evolution of the Nyāya-Vaiśeṣika Categoriology*, vol. 1, *Early Nyāya-Vaiśeṣika Categoriology*, Varanasi: Bharati Prakashan.(Starting from discussions of the concept of category in both schools, the book examines the possible sources of the categories and attempts to reconstruct their early development; written with a comparative perspective.)
Nenninger, C. (1992) *Aus gutem Grund: Praśastapādas anumāna-Lehre und die drei Bedingungen des logischen Grundes (On Good Grounds: Praśastapāda's Theory of Anumāna and the Three Conditions for Logical Grounds)*, Philosophia Indica: Einsichten-Ansichten 1, Reinbek: Dr Inge Wezler.(Translation and analytical study of the section on inference in the *Padārthadharmasaōgraha*.)
Nyāyasūtra (*c.*400 AD), trans. G. Jha, *The Nyāya-Sūtras of Gautama with the Bhāṣya of Vātsyāyana and the Vārtika of Uddyotakara*, *Indian Thought* 4-11, 1912-19; repr. Delhi: Motilal Banarsidass, 4 vols, 1984.(Complete translation together with the first preserved commentary and its subcommentary; contains profuse notes from further commentaries. However, the translation of the basic text is heavily influenced by the commentaries.)
Oberhammer, G. (1963) 'Ein Beitrag zu den Vāda-Traditionen Indiens' (A Contribution on the Indian Vāda Traditions), *Wiener Zeitschrift für die Kunde Süd- und Südostasiens* 7: 63-103.(On the background of Nyāya within the larger tradition of debate.)
Oberhammer, G. (1984) *Wahrheit und Transzendenz: Ein Beitrag zur Spiritualität des Nyāya (Truth and Transcendence: A Contribution on Nyāya Spirituality)*, österreichische Akademie der Wissenschaften, phil.-hist. Klasse, Sitzungsberichte 424, Veröffentlichungen der Kommission für Sprachen und Kulturen Südasiens 18, Vienna: Verlag der österreichischen Akademie der Wissenschaften.(The only monograph dedicated exclusively to soteriology as well as meditative and yogic practices in Nyāya; it utilizes Vātsyāyana's Nyāyabhāṣya and Bhāsarvajña's Nyāyabhūṣaṇa.)
Oberhammer, G. (1991) *Terminologie der frühen philosophischen Scholastik in Indien: Ein Begriffswörterbuch zur altindischen Dialektik, Erkenntnislehre und Methodologie (Terminology of the Early Indian Philosophical Scholastics: A Technical Dictionary of Ancient Indian Dialectics, Epistemology and Methodology)*, vol. 1, *A-I*, österreichische Akademie der Wissenschaften, phil.-hist. Klasse, Denkschriften 223, Beiträge zur Kultur- und Geistesgeschichte Asiens 9, Vienna: Verlag der österreichischen Akademie der Wissenschaften.(Alphabetically arranged articles on the technical terms of dialectics, epistemology and methodology, mainly in works up to the sixth century; the articles study the terms from the philological-historical point of view.)
Oberhammer, G., Prets, E. and Prandstetter, J. (1996) *Terminologie der frühen philosophischen Scholastik in Indien: Ein Begriffswörterbuch zur altindischen Dialektik, Erkenntnislehre und Methodologie (Terminology of the Early Indian Philosophical Scholastics: A Technical Dictionary of Ancient Indian Dialectics, Epistemology*

and Methodology), vol. 2, *U-Pū*, österreichische Akademie der Wissenschaften, phil.-hist. Klasse, Denkschriften 248, Beiträge zur Kultur- und Geistesgeschichte Asiens 17, Vienna: Verlag der österreichischen Akademie der Wissenschaften.(See Oberhammer 1991.)

Perry, B.M. (1997) 'Early Nyāya and Hindu Orthodoxy: *ānvīkṣikī* and *adhikāra*', in E. Franco and K. Preisendanz (eds) *Beyond Orientalism: The Work of Wilhelm Halbfass and its Impact on Indian and Cross-Cultural Studies*, PoznańStudies in the Philosophy of the Sciences and the Humanities 59, Amsterdam/Atlanta, GA: Editions Rodopi, 449-70.(Study of the four basic Nyāya commentaries on the self-understanding of Nyāya with regard to Vedic orthodoxy and on the relationship between reasoning and orthodox religious authority; also addresses the issue of who is qualified to study Nyāya.)

Potter, K.H. (ed.) (1977) *Encyclopedia of Indian Philosophies*, vol. 2, *Indian Metaphysics and Epistemology: The Tradition of Nyāya-Vaiśeṣika up to Gaōgeśa*, Princeton, NJ: Princeton University Press, and Delhi: Motilal Banarsidass. (Introductory survey and detailed summaries of individual works intended primarily for philosophers.)

Potter, K.H. and Bhattacharyya, S. (eds) (1993) *Encyclopedia of Indian Philosophies*, vol. 6, *Indian Philosophical Analysis: Nyāya-Vaiśeṣika from Gaōgeśa to Raghunātha Śiromaṇi*, Princeton, NJ: Princeton University Press, and Delhi: Motilal Banarsidass. (Introductory survey and detailed summaries of individual works intended primarily for philosophers.)

Praśastāpada (early 6th century) *Padārthadharmasaōgraha (Compendium of Properties of the Categories)*, trans. G. Jha, *Padārthadharmasaōgraha of Praśastapāda, with the Nyāyakandalīof Śrīdhara, Pandit*, new series, 25-37, 1903-15; repr. Chaukhambha Oriental Studies 4, Varanasi/New Delhi: Chaukhambha Orientalia, 1982. (A fundamental Vaiśeṣika text.)

Preisendanz, K. (1989) 'On *ātmendriyamanorthasannikarṣa* and the Nyāya-Vaiśeṣika Theory of Vision', *Berliner Indologische Studien* 4/5: 141-213.(Discusses contact between self, senses, 'internal organ' and object as part of the definition of perception; also contains a study of Nyāya-Vaiśeṣika optics in comparison with ancient Greek optics.)

Raghunātha Śiromaṇi (15th-16th century) *Padārthatattvanirūpaṇa (Examination of the True Nature/Reality of the Categories)*, trans. K.H. Potter, *The Padārthatattvanirūpaṇam of Raghunātha Śiromaṇi*, Harvard Yenching Studies Series 17, Cambridge, MA: Harvard University Press, 1957.(Translation of a Navya-Nyāya attempt to revise radically the traditional Vaiśeṣika categories.)

Ramanujam, P.S. (1979) *A Study of Vaiśeṣika Philosophy with Special Reference to Vyomaśivācārya*, University of Mysore Other Publication 68, Mysore: University of Mysore.(A clear and reliable paraphrase of large portions of the *Vyomavatī*.)

Schmithausen, L. (1965) *Maṇḍanamiśra's Vibhramaviveka, mit einer Studie zur Entwicklung der indischen Irrtumslehre (Maṇḍanamiśra's Vibhramaviveka, with a Study of the Development of the Indian Theory of Error)*, österreichische Akademie der Wissenschaften, phil.-hist. Klasse, Sitzungsberichte 247, 1, Veröffentlichungen der Kommission für Sprachen und Kulturen Südasiens 2, Vienna: Verlag der österreichischen Akademie der Wissenschaften.(Hidden as an appendix is the most remarkable study of theories of error in classical Indian philosophy. Of special importance to Nyāya studies is the reconstruction of the respective positions of the Ācārya and Vyākhyātṛ schools, and of the relationship between error and conceptual construction.)

Shastri, D.N. (1964) *Critique of Indian Realism*, Agra: Agra University; repr. as *The Philosophy of Nyāya-Vaiśeṣika and its Conflict with the Buddhist Dignāga School*, Delhi: Bharatiya Vidya Prakashan, 1976. (Exceptionally lively presentation of the major ontological and epistemological issues in the controversy between Nyāya-Vaiśeṣika realism and its Buddhist opponents.)

Thakur, A.L. (1969) '*adṛṣṭa* and *dharma* in the Vaiśeṣika Philosophy', *Ṛtam* 1 (1): 51-8.(Critical examination of some key passages in the *Vaiśeṣikasūtra* to demonstrate the original purely scientific orientation of Vaiśeṣika.)

Vaiśeṣikasūtra (*c.* 2nd-3rd century AD), trans. and ed. N. Sinha, *The Vaiśeṣika Sûtras of Kaṇâda with the Commentary of Śaōkara Miśra and Extracts from the Gloss of Jayanârâyaṇa together with Notes from the Commentary of Chandrakânta and an Introduction by the Translator*, Allahabad: Panini Office, 1911; repr. Sacred Books of the Hindus 20, Delhi: S.N. Publications, 1986. (Translation of the fifteenth-century commentary *Vaiśeṣikasūtra-Upaskāra*, including the *sūtra*-text transmitted in it; the translation of the *sūtras* follows the interpretation of this rather late commentary by Śaōkara Miśra.)

Vaiśeṣikasūtra (*c.* 2nd-3rd century AD), trans. M. Nozawa, 'The Vaiśeṣikasūtra with Candrānanda's Commentary (1)', *Numazu College of Technology Research Annual* 27: 97-116, 1992.(A translation of the first two chapters with parallel Sanskrit materials. Candrānanda's commentary contains a more original version of the *sūtra*-text than Śaōkara Miśra's, and the interpretation, although clearly post-Praśastapādan, reflects an earlier stage of

the system.)
Vattanky, J. (1993) *Development of Nyāya Theism*, New Delhi: Intercultural Publications.(Historical exposition of the Nyāya-Vaiśeṣika arguments for the existence of God and of the controversy with Buddhist philosophers.)
Viśvanātha Pañcānana (17th century) *Bhāṣāpariccheda* and auto-commentary *Siddhāntamuktāvalī*, trans. Mādhavānanda, *Bhāṣā-Pariccheda with Siddhānta-Muktāvalīby Viśvanātha Nyāya-Pañcānana*, Calcutta: Advaita Ashrama, 1954. (Translation of the most popular syncretistic manual.)
Wada, T. (1990) *Invariable Concomitance in Navya-Nyāya*, Sri Garib Dass Oriental Series 101, Delhi: Sri Satguru Publications.(A clear introduction to basic Navya-Nyāya concepts, with translations and explanations of important passages on the invariable concomitance between probans and probandum in inference.)
Wezler, A. (1969) 'Die "dreifache" Schlussfolgerung in *Nyāyasūtra* 1.1.5' (The 'Threefold' Inference in *Nyāyasūtra* 1.1.5), *Indo-Iranian Journal* 11: 190-211. (Clarification of the most probable original meaning of the *Nyāyasūtra*'s definition of inference, by comparison with early Buddhist texts.)
Wezler, A. (1982) 'Remarks on the Definition of "*yoga*" in the *Vaiśeṣikasūtra*', in L.A. Hercus *et al*. (eds) *Indological and Buddhist Studies: Volume in Honour of Professor J.W. de Jong on his Sixtieth Birthday*, Canberra: Faculty of Asian Studies, 643-86. (Argues that the sections on yoga, liberation and yogic perception were not part of the original *Vaiśeṣikasūtra*, but half-hearted concessions to contemporary philosophical tendencies.)

Nygren, Anders (1890-1978)

Nygren hoped to recover the uniqueness of Christianity from the impurities introduced by the attempts of nineteenth-century liberal theology to free it from metaphysical speculation and confessional dogmatism. He aimed to do this by grounding all religion in an analytic philosophy of religion which would enable him to stress the objective character of Christian theology in contrast to the arbitrariness of confessional theology. He achieved international influence by his claim in Agapē and Eros (1930-6) that the uniqueness of Christianity is love in the sense of agapē, *as opposed to Platonic* eros.

Like many of his theological contemporaries in Sweden, Nygren sought to show the legitimacy of theology as a university subject. While a professor at Lund (1924-48), he argued that theology could withstand positivistic criticism of the sort made by Axel Hägerström (§5), whose attack on Hegelian-inspired perspectives sought to include theology in its anti-metaphysical cleansing of Swedish philosophy. Nygren, like Hägerström, accepted Kant's criticism of a rational theology based on the traditional proofs of God's existence, but rejected Kant's reduction of religion to ethics (see Kant, I. §11). Like Schleiermacher (§7), he insisted that religion is a distinctive perspective, as are the perspectives of reason, practice and ascetic value. But Nygren does not base religion on a specific religious experience, which would make it primarily subjective. Rather, religions are an obvious historical fact. Nygren viewed philosophy in a Kantian fashion, as a critical study of the necessary conditions of meaning. Theology as a university subject could be justified as a historical critical study of the meaning of Christianity, and so could be utterly objective, intersubjective and testable (unlike confessional theology, which rests on revelation alone, as does the neo-orthodoxy of Barth and Brunner).

By the time of Nygren's last major work, *Meaning and Method* (1972), philosophy had taken a linguistic turn, with which he felt at home. Drawing upon Wittgenstein's *Philosophical Investigations*, he was able to develop his philosophy of religion in its idiom. He stressed context as crucial to the meaning of language, and held that philosophy could clarify the differences and relations between basic contexts of meaning. He also added what he called 'the context of contexts', consisting of four great questions which in *Agapē and Eros* he had called 'fundamental motifs'. He claimed that these larger contexts are simply given in human life, and summarized them as the question of truth, which is institutionalized as science; the question of good and evil, and right and wrong, institutionalized as law and ethical traditions; the question of the beautiful, institutionalized in art; and the question of the eternal, institutionalized as religion.

Nygren argued that philosophy of religion is the specification of the fundamental presuppositions that concern human encounters with the eternal; this for him is not an abstraction, such as timelessness, but that which breaks into time and enables us to see that life is not subject merely to the conditions of finitude. He did not perform this task, but indicated four elements in the existential confrontation with the eternal, which occur wherever there is religion: an unveiling, a disquieting judgment, a reconciliation, and a vital fellowship. These elements are merely formal and are specified only from a knowledge of the content of actual religions, but they enable one to specify the logical place of religion in relation to other fundamental perspectives, and to clarify confusions in religion that result from mixing contexts.

Nygren maintained that the task of theology is the study of concrete religions. He called his method 'motif research' and supposed it applicable to all historical research. One could discover the basic notion around which all the historical phenomena revolved, and which provided the context of meaning for those phenomena. The answer each religion gives to the fundamental question of the eternal is its motif. For Judaism it is law, for Christianity it is *agapē* and for Hellenic civilization it is *eros*.

Nygren contrasted the distinctive Christian conception of love, which he called *agapē*, with the Platonic notion, which he called *eros* (see Love §2). All *eros* is egocentric and acquisitive, whereas *agapē* springs wholly from within and takes no account of worth. The former is human, the latter divine. Nygren claimed that, as fundamental motifs, these should not be mixed, but that historically they had been, above all by Augustine. Only with the Reformation was the original New Testament notion of *agapē* restored. However, this thesis has been countered, and Nygren's characterization of both *agapē* and *eros* challenged (Burnaby 1938; D'Arcy 1954; Markus 1955).

DIOGENES ALLEN

List of works

Nygren, A. (1930-6) *Agapē and Eros*, London: Society for Promoting Christian Knowledge.(Contrasts Christian and human love.)

Nygren, A. (1949) *Commentary on Romans*, Philadelphia, PA: Muhlenberg.(A contribution to a recovery of Martin Luther's teachings.)

Nygren, A. (1972) *Meaning and Method*, Philadelphia, PA: Fortress.(A systematic account of Nygren's integration of theology and philosophy of religion.)

References and further reading

Burnaby, J. (1938) *Amor Dei*, London: Hodder & Stoughton.(A careful defence of Augustine against Nygren.)

D'Arcy, M.C. (1954) *The Mind and Heart of Love*, London: Faber & Faber, 2nd edn.(A Roman Catholic response to Nygren's view of *agapē* and *eros*.)

Hall, T. (1978) *Anders Nygren*, Waco, TX: Word.(A general introduction.)

Kegley, C.W. (ed.) (1970) *The Philosophy and Theology of Anders Nygren*, Carbondale, IL: Southern Illinois University Press.(A collection of essays.)

Markus, R.A. (1955) 'The Dialectics of Eros in Plato's *Symposium*', *Downside Review* 233: 219-30.(A criticism of Nygren's characterization of *eros*.)

Wittgenstein, L. (1953) *Philosophical Investigations*, ed. G.E.M. Anscombe and R. Rhees, trans. G.E.M. Anscombe, Oxford: Blackwell.(Wittgenstein's influential account of language.)

Oakeshott, Michael Joseph (1901-90)

Although Michael Oakeshott was in his own time a lone figure in a philosophical world dominated by Oxford analysis, he has come to be recognized as the most notable British political philosopher of the twentieth century. He is best known for his view that political activity is neither purely empirical nor the application of ideas, but 'the pursuit of intimations'. His image of culture as a conversation between different kinds of understanding has been widely accepted. Oakeshott first became celebrated in attacking what he called 'rationalism' in the Cambridge Journal (which he edited) in 1947. Oakeshott's rationalist is a restless political meddler who believes that politics is putting ideas into effect. The fullest statement of his political philosophy is On Human Conduct (1975), in which the modern state is understood as a tension between civil and enterprise association. Exploring the idea of a civil association is perhaps Oakeshott's most notable contribution to political philosophy.

1 Modes of experience

Michael Oakeshott was born at Chelsford, Kent, and attended St George's School, Harpenden. He studied history at Gonville and Caius College, Cambridge, and became a fellow after two years spent visiting German universities and as a schoolmaster. During the Second World War he served in the British army. In 1952 he succeeded Harold Laski as professor of political science at the London School of Economics, a post from which he retired in 1969.

The early Oakeshott was confidently classified as an idealist on the basis of *Experience and its Modes* (1933) and some contemporaries early on were tempted to pigeonhole and ignore him. Susan Stebbing reviewing the work in *Mind* thought that it merely reworked Bradley. A.J. Ayer said of Oakeshott's conservatism that 'it's all in Burke'. One reviewer as late as *On Human Conduct* (1975) merely assimilated him to Wittgenstein. All of this was to miss a highly original philosophical intelligence which explored the idea of the state more deeply than any of his contemporaries.

In *The Voice of Poetry in the Conversation of Mankind* (1959) Oakeshott provided the best short synoptic view of his philosophy, and also launched a famous image: human life is an endless conversation between different 'voices'. Practice, science, history and poetry are voices, while philosophy is 'the impulse to study the quality and style of each voice', and to reflect upon the relationship of one voice to another. The philosopher's reflections, however, make no specific contribution to this conversation, for only those with a specific conceptual point of view have anything to contribute. Philosophy is in particular distinct from practice, and can make no direct contribution to it, though Oakeshott occasionally allows that it may have such a virtue as 'hindering irrelevance'.

In *Experience and its Modes*, however, Oakeshott had appeared to give philosophy a rather grander role: it was the pursuit of 'what is ultimately satisfactory in experience'. The modes, by contrast, were forms of experience limited or 'arrested' by their logical dependence upon constitutive assumptions. Practice, for example, is experience understood in terms of desire and frustration - *sub specie voluntatis*. Science cultivates a mechanical world understood in terms of quantity. History is the world understood in terms of its pastness. Philosophy explores these conditionalities in the search for a concrete understanding of experience as released from the abstract or fragmentary character of the modes. It is both what is ultimately satisfactory in experience, and also recognized as a somewhat perverse enterprise: most of the time, men understandably prefer 'the empty kisses of abstraction'.

Oakeshott is evidently hard to classify. A consistent allegiance in his work is to a Platonic conception of philosophy as exploring the conditionalities of experience in order to discover its postulates - but Oakeshott's philosopher has no timeless realm to contemplate. Again, he seems to reveal the grand vision of the philosophical vocation found in Hegel, yet he combines this vision with an underlabouring insistence on philosophy's irrelevance to the world which links him to British empiricism, especially Hobbes. What is certainly clear is that he took philosophy to be the activity of thinking through the conditions of some human activity, purging it of contingency and merging its conclusions and its presuppositions in an ideal character revealing the logic of our understanding. His intellectual life consisted in successive inquiries of this kind - the first published in *Experience and its Modes*, a second in *The Voice of Poetry in the Conversation of Mankind* (where the fundamentals are revisited in order to accommodate art to the system of modes) and finally and most magisterially in the first two essays of *On Human Conduct*. Let us first, however, consider this same intellectual exercise performed on a smaller scale in essays written around 1946. They are to be found in *Religion, Politics and the Moral Life* (1993).

2 Epistemology

The basic error about political philosophy, he argued in 'The Concept of a Philosophy of Politics', was to think that it consisted merely in the *application* of some already established philosophy to the field of politics. Oakeshott's objection to this procedure is that 'application' is altogether too mechanical a process to have any place in philosophy. Oakeshott denied that philosophy could provide anything with foundations - one of the doctrines which has led Rorty and others to assimilate him to modern forms of pragmatism (see Pragmatism). Nor did it provide criteria of action, for practice could never achieve the coherence of a philosophical concept.

Oakeshott's ontology is sceptical without being in any way solipsistic. The world is composed of images, or 'intelligibles'. A thing is what it is imagined to be, but might be any number of other things. Concepts such as 'thing' or 'fact' are merely verdicts or judgments about some object of experience which always appears to the understanding in one specific guise or other, never as any kind of basic datum. Oakeshott frequently reverts to the doctrine of the *Meno* in which understanding always begins with something already in some degree understood, and moves on to other understandings, without ever achieving finality. Although he contrasts practical skill with rationalist notions of technology, Oakeshott rejects any notion of a basic level of common sense (frequently identified with practice) to which all controversies can be assimilated. The aim of *Experience and its Modes* is to show the impossibility of assimilating one mode to another, and drawing the conclusion that irrevelance, or *ignoratio elenchi*, is the most crippling form of error. In *The Voice of Poetry* he complains that the voice of practice is becoming a bore in the modern world because it is becoming strident and tries to dominate the conversation. Part of the reason for this is that practice and science can seem to be mutually assimilable because they have some names in common - such as 'fact'.

History also suffered from the kind of uncriticial monism which took historical facts to be the same kind of logical entity as facts in science and practice. A major thread in Oakeshott's career was his lifelong devotion to thinking out the minimal epistemological conditions of historical understanding. One effect was to banish the idea of cause from the field. He recognized, of course, that historians often affirmed causal relationships, but he interpreted this as a rhetoric of emphasis within a process in which the aim was to compose the circumstantial understanding available from the records so as to answer to the question: what happened? Only this could provide the intelligibility allowing us to understand why it happened. Such intelligibility was the most we might expect from history, but it was also the only manner in which human life itself might be explained.

3 Political philosophy

In his last major work, Oakeshott typically started all over again. *On Human Conduct* begins with a long essay on theorizing human conduct. Whereas nature may be understood in terms of process, the human world is self-conscious and procedural. The basic confusion of the social sciences has been a failure to clarify the distinction between, as it were, a wink and a blink.

The 'engagement' of understanding is, he tells us, a continuous, self-moved, critical enterprise of theorizing. It emerges out of 'misty intimations of intelligibility'. We distinguish 'goings on' in terms of some characteristic; identification is specification in terms of an ideal character composed of characteristics. At this point in Oakeshott's phenomemology of understanding a more coherent system of thought, which he calls a 'conditional platform of understanding' has emerged: every 'going on' is what it is in respect of being understood in terms of an ideal character specified as a composition of characteristics. Wittgenstein took a similar view in construing ideas in terms of 'family resemblances'. A theorist may, however, treat an identity not as a verdict to be accepted, but as an invitation to make such an identity more intelligible, and this can only be done by seeking to understand it in terms of its conditionality.

The voices and modes of Oakeshott's earlier formulations have here given way to a more pluralistic scheme. The theorizer may 'fish' for the understandings available from these conditional platforms of understanding. Yet the irony of all theorizing, Oakeshott remarks, is its propensity to generate, not an understanding, but a not-yet-understood. Thinking thus leads the theorizer towards higher platforms of understanding concerned with exploring the postulates of that activity itself.

Turning to human conduct, Oakeshott finds that it is a set of 'goings on' which exhibit intelligence. We necessarily

act in terms of some specific practice, and a practice is a language of self-disclosure. Inseparable from action in the moral dimension is what Oakeshott calls 'self-enactment' in which conduct reveals the self to itself and to others.

In the second essay, Oakeshott analyses the state, though the word itself is kept at bay (see State, the). Feeling that the modern vocabulary of politics had been seriously corrupted, he reinstates Latin terms (such as *civitas*, *lex* and *respublica*) as technicalities. The task is to distinguish the 'civil condition'. It certainly is not to be found in those transactions of reciprocal desiring which belong to the social rather than the civil world. A second 'categorially distinct mode of association' is revealed in 'agents related in the joint pursuit of some imagined and wished-for common satisfaction' (1975a: 112): what Oakeshott called 'enterprise association'. The state as an enterprise association is a view almost irresistibly invoked by the familiar question: what is the end or purpose of the state? But enterprises are voluntary, while the state is compulsory.

There does remain, however, a distinct form of association corresponding to the civil condition. It is 'relationship in terms of the conditions of a practice'. The *cives* of a civil association relate to one another in nothing else but a recognition of the authority of rules, in respect of which they are all equal. They are thus not the comrades of an enterprise association, but much more like the speakers of a language, whose intelligibility derives from conforming to its rules. This is a formal relationship in that the rules do not specify any particular conduct; they are merely 'adverbial qualifications' of whatever a member might choose to do. But although formal, these rules constitute a practice which is moral in the basic sense of not being instrumental to the enjoyement of specific satisfactions.

Here then is an analysis of the forms of human association so intricate that Oakeshott created a complex vocabulary and style in which to express it. It leaves any attempt at synopsis far behind.

See also: Political philosophy, history of

KENNETH MINOGUE

List of works

Oakeshott, M.J. (1933) *Experience and its Modes*, Cambridge: Cambridge University Press.(A lucid and witty account of Oakeshott's sceptical idealism.)

Oakeshott, M.J. (1939) *The Social and Political Doctrines of Contemporary Europe*, Cambridge: Cambridge University Press.(An anthology of writings on the doctrines current in the 1930s.)

Oakeshott, M.J. (1959) *The Voice of Poetry in the Conversation of Mankind*, London: Bowes & Bowes.(A short essay on aesthetics which recapitulates his entire philosophical position.)

Oakeshott, M.J. (1962) *Rationalism in Politics*, London: Methuen; expanded edn, ed. T. Fuller, Indianapolis, IN: Liberty Press, 1991.(Oakeshott's final and most systematic account of political philosophy.)

Oakeshott, M.J. (1975a) *On Human Conduct*, Oxford: Clarendon Press.(Oakeshott's last major work of political philosophy, exploring the relation between civil and enterprise association.)

Oakeshott, M.J. (1975b) *Hobbes on Civil Association*, Oxford: Blackwell.(Collects Oakeshott's essays on Hobbes, and notably the magisterial introduction to *Leviathan*.)

Oakeshott, M.J. (1983) *On History and Other Essays*, Oxford: Blackwell.(Essays elaborating a complete account of the logic of historical inquiry.)

Oakeshott, M.J. (1993a) *Religion, Politics and the Moral Life*, ed. T. Fuller, London: Yale University Press. (Notable for revealing the overriding concern with religion of the early Oakeshott.)

Oakeshott, M.J. (1993b) *Morality and Politics in Modern Europe: The Harvard Lectures*, ed. S.R. Letwin, London: Yale University Press.(A short and lively account of modern European politics in lectures given at Harvard in 1958.)

Oakeshott, M.J. (1996) *The Politics of Faith and the Politics of Scepticism*, ed. T. Fuller, London: Yale University Press.(A previously unpublished manuscript dealing in a philosophical manner with the way Oakeshott understood politics around 1950.)

Norman, J. (ed.) (1993) *The Achievement of Michael Oakeshott*, London: Duckworth.(Contains a full bibliography of Oakeshott's works by John Liddington.)

References and further reading

Greenleaf, W.H. (1966) *Oakeshott's Philosophical Politics*, London: Longman.(An early attempt to bring

Oakeshott's politics to a focus.)
Franco, P. (1990) *The Political Philosophy of Michael Oakeshott*, London: Yale University Press.(A full-length study of Oakeshott and his context.)
Grant, R. (1990) *Thinkers of Our Time: Oakeshott*, London: Claridge Press.(An excellent short account with some attention to Oakeshott's life and circumstances.)

Objectivity

Objectivity is one of the central concepts of metaphysics. Philosophers distinguish between objectivity and agreement: 'Ice-cream tastes nice' is not objective merely because there is widespread agreement that it is true. But if objectivity is not mere agreement, what is it? We often think that some sorts of claim are less objective than others, so that a different metaphysical account is required of each. For example, ethical claims are often held to be less objective than claims about the shapes of middle-sized physical objects: 'Murder is wrong' is held to be less objective than 'The table is square'. Philosophers disagree about how to capture intuitive differences in objectivity. Those known as non-cognitivists say that ethical claims are not, strictly speaking, even apt to be true or false; they do not record facts but, rather, express some desire or inclination on the part of the speaker. Others, dubbed subjectivists, say that ethical statements are in some sense about human desires or inclinations. Unlike the non-cognitivist, the subjectivist views ethical claims as truth-apt, but as being true in virtue of facts about human desires or inclinations. Some philosophers, referred to as anti-realists, disagree with both non-cognitivism and subjectivism, and attempt to find different ways of denying objectivity. Quietists, on the other hand, think that there are no interesting ways of distinguishing discourses in point of objective status.

1 Non-cognitivism and subjectivism

Intuitively, an ethical claim such as 'It is right to help those in distress' is less objective than a claim such as 'The table is square'. What does this intuition amount to? Non-cognitivists, such as Ayer (1936), Blackburn (1984) and Gibbard (1990), focus on the semantic function of the two claims. 'The table is square' has the function of asserting that a fact obtains in the world. If that fact obtains - if the table is square - then the claim is true; if not, it is false. Thus 'The table is square' is apt to be true or false: it is truth-apt. On the other hand, the non-cognitivist views 'It is right to help those in distress' as having a different semantic function: despite appearances, its function does not consist in asserting that a fact obtains in the world. It is not truth-apt; rather, it expresses an inclination, desire, or some other non-cognitive attitude of the speaker, in the same way that 'Hurrah!' or 'Boo!' merely expresses a favourable or unfavourable attitude. This is why some versions of non-cognitivism are called 'expressivism'. Ethical claims are less objective than claims like 'The table is square' since the latter are genuine truth-apt assertions, whereas the former are not. Non-cognitivists have viewed ethical claims as not objective in this sense for various reasons: because of widespread disagreement over moral matters, because moral facts seem 'queer' or 'odd', and because of the normative character of moral discourse.

There are problems with this way of capturing differences in objective status. Consider the inference:

1 If lying is wrong, then getting little brother to lie is wrong.
2 Lying is wrong.
3 Getting little brother to lie is wrong.

According to non-cognitivism, 'Lying is wrong' and 'Getting little brother to lie is wrong' are not truth-apt. But the inference is intuitively *valid*. And to say that an inference is valid is to say that the truth of the premises guarantees the truth of the conclusion. How can non-cognitivism explain the validity of the inference, given that it denies that (2) and (3) are even apt to be true? Relatedly, the non-cognitivist has a difficulty with (1). What sense can be made of a conditional whose antecedent and consequent are not truth-apt? Non-cognitivism has had difficulty in finding satisfactory answers to these questions (Hale 1993).

Subjectivists take a different tack. They admit that 'It is right to help those in distress' is truth-apt, that its sincere utterance makes a genuine assertion. They thus avoid the sort of difficulty outlined above. But if an ethical claim genuinely asserts that a fact obtains, what sort of fact is it? According to the subjectivist, it is simply a fact about the person making the claim: when I say 'It is right to help those in distress' I am asserting that I desire to help them. Ethical claims are truth-apt, but the facts they assert to exist are facts about us, our desires, inclinations or subjective states. When I say 'It is right to help those in distress' I literally *mean* 'I desire to help those in distress'. This contrasts with claims like 'The table is square' which, though truth-apt, are not analysable in terms of facts about human subjectivity. In this sense, 'The table is square' is the more objective.

Subjectivism has difficulty in accounting for moral disagreement. If Jones says 'It is right to help those in

distress' and Smith says 'It is wrong to help those in distress', they are disagreeing with each other, indeed contradicting one another. But if Jones means 'I (Jones) desire to help the unfortunate', whereas Smith means 'I (Smith) do not desire to help the unfortunate', there is no contradiction. So in what sense do they disagree?

Subjectivism may try to deal with this by focusing on a wider range of desires than those possessed by the individual who utters the statement. 'It is right to help those in distress' literally means 'Most people desire to help the unfortunate', so that its content is spelled out in terms of intersubjective agreement in desires. But this also faces problems. If it were true, it would be a contradiction to say that helping the unfortunate is right even though most people do not desire to do it. But it is not a contradiction, so even this wider subjectivist analysis must be flawed.

2 Anti-realist views of objectivity

Suppose we admit that two types of claim are truth-apt and yet do not have an anthropocentric subject matter. Do we have to view them as equally objective? Anti-realists, such as Wright (1986, 1992) and Dummett (1978), think not. They try to show how differences in objectivity can be captured, without adopting either non-cognitivism or subjectivism. How can we distinguish between 'It is right to help those in distress' and 'The table is square' if we allow that both are truth-apt, and that neither is susceptible to a subjectivist analysis? One way would be to develop an analogue of the distinction between primary and secondary qualities, the locus classicus for which is Locke (1689) (see Primary-secondary distinction; Secondary qualities). The standard example of a secondary quality is colour, while shape is usually held to be primary. What is involved in the claim that redness, for example, is a secondary quality? The idea is that there is a close relationship between facts about certain of our subjective states and statements like 'The mailbox is red'. In order to explain this relationship, we need to explain the notion of a best judgment. A judgment that an object is red is best when it is made in conditions that are optimally good for appraising whether or not the mailbox is red. These conditions could be specified very roughly as those that obtain out of doors, in the shade, at lunch time on a lightly overcast summer afternoon. To say that redness is a secondary quality is to say that our best judgments concerning the redness of objects determine the extension of the concept red. The extension of a concept is the class of things to which that concept can correctly be applied. So our best judgments about whether objects are red *determine* which class of objects the concept red can properly be applied to. In contrast, a primary quality like squareness is one whose extension is determined independently of facts about best judgments concerning squareness. Best judgments in this case merely detect an independently determined extension.

This way of saying how 'The mailbox is red' is less objective than 'The table is square' is different from the subjectivist analysis. To say that redness is secondary is not to say that 'The mailbox is red' literally means that 'If conditions were best, we would judge that the mailbox is red'. The claim is the weaker one, that our best judgments determine whether the mailbox falls in the extension of red. The proposal that redness is a secondary quality can therefore avoid the problems associated with subjectivism. And since it does not deny that 'The mailbox is red' is truth-apt, it avoids the problems associated with non-cognitivism. One way, then, to claim that ethical claims are less objective than claims like 'The table is square' is to argue that ethical qualities such as right, wrong, good and bad are secondary rather than primary.

According to the anti-realist, there can be more than one way of saying that one type of claim is less objective than another. Even if we cannot show that ethical qualities are secondary, there might be other ways of contrasting ethical claims with different claims. One way concerns what Wright (1986) calls 'objectivity of truth'. This notion of objectivity features prominently in Dummett's discussions of realism. To say that a class of statements exhibits objectivity of truth is to say that '[they] may be fully intelligible to us even though resolving their truth-values may defeat our cognitive powers (even when idealized)' (Wright 1986: 5). Their truth might be 'evidence-transcendent': they are determinately either true or false, even if we are in principle incapable of citing evidence for or against them. Take Goldbach's Conjecture (that every even number greater than two is the sum of two primes), or some claim about past happenings in some far distant galaxy. We understand these claims, but we are in principle incapable of resolving their truth-values. This gives us one way of claiming that mathematical or cosmological claims are more objective than claims about morals or comedy. There is little temptation to think that claims about the moral status of an action, or about the comic quality of a joke, may in principle transcend our best cognitive efforts. There are no difficulties akin to those affecting non-cognitivism and subjectivism: there is no

denial of truth-aptitude, nor an assignment of an anthropocentric subject matter.

3 Objectivity of meaning and quietism

The main danger for the anti-realist story about objectivity comes from Wittgenstein's rule-following considerations (RFC). According to Wright, the RFC endanger the objectivity of meaning. This is the view that 'the meaning of a statement is a real constraint, to which we are bound… by contract, and to which verdicts about its truth-value may objectively conform, quite independently of our considered opinion on the matter' (Wright 1986: 5). The meaning of a statement imposes requirements on what counts as correct use of the statement, which determine what uses are correct and incorrect independently of any opinions we may subsequently form. The problem is that if the RFC destroy the idea that meanings are objective in this sense, they thereby threaten the various ways in which anti-realists attempt to draw comparisons concerning objectivity. For example, Wright believes that objectivity of truth implies objectivity of meaning. If the RFC force us to reject objectivity of meaning, they also force us to reject objectivity of truth. And if no discourses possess objectivity of truth, appealing to failure of objectivity of truth will be useless for drawing comparisons between discourses. Likewise, since the truth of any statement is a function of its meaning together with facts about the world, rejection of objectivity of meaning may entail that all qualities are secondary. The possibility of appealing to the primary-secondary distinction in order to draw a contrast will be endangered. In short, the RFC seem to threaten us with quietism about objectivity: the view that no principled, metaphysically interesting contrasts concerning objectivity can be drawn.

Anti-realists try to find ways of avoiding quietism, while retaining their interpretation of the RFC (Wright 1992: Ch. 6). Note that some philosophers, such as McDowell (1995), think that there is a different way in which quietism about objectivity can open up. McDowell does not view the RFC as threatening the objectivity of meaning. So the relevance of the primary-secondary distinction and evidence-transcendent truth is not threatened in the direct manner envisaged in the previous paragraph. The distinctions the anti-realist wishes to draw can still be drawn: instead, what the RFC threaten is the idea that there is any interesting metaphysical *point* to be made by appealing to the distinctions in the first place. For example, it might be argued that the thought that the anti-realist primary-secondary distinction is of metaphysical relevance depends upon a conception of detecting or tracking facts that the RFC display to be untenable.

See also: Projectivism; Realism and antirealism

ALEXANDER MILLER

References and Further Reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz.(Mentioned in §1 above. Chapter 6 contains a classic statement of non-cognitivism about ethics.)

Blackburn, S. (1984) *Spreading the Word*, Oxford: Oxford University Press. (Mentioned in §1 above. Chapters 5 and 6 develop a modern version of non-cognitivism.)

Boghossian, P. (1989) 'The Rule-Following Considerations', *Mind* 98 (392): 507-49. (Comprehensive survey of the literature on the rule-following considerations.)

Dummett, M. (1978) *Truth and Other Enigmas*, London: Duckworth.(Mentioned in §2 above. Influential anti-realist book which contains much discussion about the objectivity of truth.)

Gibbard, A. (1990) *Wise Choices, Apt Feelings*, Cambridge, MA: Harvard University Press.(Mentioned in §1 above. Highly sophisticated development of non-cognitivism.)

Hale, B. (1993) 'Can there be a logic of attitudes?', in J. Haldane and C. Wright (eds) *Reality, Representation and Projection*, Oxford: Oxford University Press. (Mentioned in §1 above. Excellent survey of the difficulties faced by non-cognitivism.)

Hume, D. (1739-40) *A Treatise on Human Nature*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1968.(Book III contains the classic statement of subjectivism in ethics.)

Leiter, B. (ed.) (1997) *Objectivity in Law and Morals*, Cambridge: Cambridge University Press.(Collection of articles on the nature and role of objectivity in law and morality.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Oxford University Press, 1975.(Bk 1, ch 8 contains the historical precursor of the primary-secondary quality distinction mentioned in §2 above.)

McDowell, J. (1995) *Mind and World*, Oxford: Oxford University Press. (Mentioned in §3 above. Statement of

quietism about objectivity.)
Smith, M. (1994) *The Moral Problem*, Oxford: Blackwell.(Excellent study of the problems surrounding the objectivity of morals.)
Wright, C. (1986) *Realism, Meaning, and Truth*, Oxford: Blackwell.(Mentioned in §§2 and 3. Develops anti-realist views on objectivity. The introduction provides an excellent survey: probably the best place to start a study of objectivity.)
Wright, C. (1992) *Truth and Objectivity*, Cambridge, MA: Harvard University Press. (Mentioned in §§2 and 3. Further development of anti-realism. Chapter 6 contains a useful discussion of quietism.)

Obligation, political

The problem of political obligation has been one of the central concerns of political philosophy throughout the history of the subject. Political obligations are the moral obligations of citizens to support and comply with the requirements of their political authorities; and the problem of political obligation is that of understanding why (or if) citizens in various kinds of states are bound by such obligations. Most theorists conservatively assume that typical citizens in reasonably just states are in fact bound by these obligations. They take the problem to be that of advancing an account of the ground(s) or justification(s) of political obligation that is consistent with affirming widespread obligations. Other theorists, however, anarchists prominent among them, do not accept the conservative assumption, leaving open the possibility that the best theory of political obligation may entail that few, if any, citizens in actual states have political obligations.

Much of the modern debate about political obligation consists of attempts either to defend or to move beyond the alleged defects of voluntarist theories. Voluntarists maintain that only our own voluntary acts (such as freely consenting to the authority of our governments) can bind us to obedience. Because actual political societies appear not to be voluntary associations, however, voluntarism seems unable to satisfy conservative theoretical ambitions. Some individualists turn as a result to nonvoluntarist theories of political obligation, attempting to ground obligations in the receipt by citizens of the benefits governments supply or in the moral quality of their political institutions. Others reject individualism altogether, defending communitarian theories that base our political obligations in our social and political roles or identities. Individualist anarchists reject instead the conservative ambitions of such theories, embracing a voluntarism which entails that most citizens simply have no political obligations.

1 The problem

The problem of political obligation has been central to political philosophy from the earliest recorded philosophical texts to the most recent. While the actual term ‘political obligation’ is of relatively recent vintage, arguments as old as those of Plato’s early dialogue, *Crito* (*c*.395-87 BC), are clearly intended to address some version of the problem. Political obligations, as these are commonly understood, are the moral obligations (or duties) of citizens to support their countries, governments or political and legal institutions. They are usually taken to include at least the obligation to obey the laws of the land (or the just laws) and are often thought to include also further obligations of loyalty and ‘good citizenship’.

The problem of political obligation, then, is that of understanding the force and nature of these moral bonds and, most importantly, of understanding when (or if) and for what reasons (if any) the members of various kinds of states are subject to them. Peoples’ political obligations constitute the core of the moral relationship that exists between them and their polities, and they are thus closely related to such corresponding concepts as the legitimacy or *de jure* authority of the state (see Authority; Legitimacy). The question here is not: under what institutional requirements, backed by force, do we stand? Rather, political philosophers ask whether in addition to (or conceivably because of) the threat of institutional coercion, blind political habits and emotional grounds for obedience, there are also strong *moral* reasons for supporting and complying with the demands of our laws, governments or states.

The problem of political obligation has usually been understood conservatively. So understood, it is the problem of finding a sound justification for our intuitive conviction that most citizens (or at least those in certain familiar kinds of states) do in fact have political obligations. Conservative theorists will thus be centrally concerned with providing a suitably ‘general’ account of political obligation, one that shows why all citizens of good states have the obligations we suppose they have. But the problem can also be understood without any conservative commitments, so that an account’s lack of generality is not necessarily a defect. On this understanding, the theorist’s job is simply to give as full as possible an account of political obligation, without any special concern for justifying our pre-theoretical beliefs about the subject. Thus, an anarchist theory (which denied the existence of any political obligations) might on this latter understanding still constitute a successful (that is, valid, nondefective) theory of political obligation.

Theories of political obligation can be (roughly) divided into four groups, the most basic division being that

between ‘communitarian’ and ‘individualist’ theories. Communitarians typically maintain that our very identities are constituted in part by our roles (such as ‘membership’) in political society and that our political obligations are tied conceptually to, and follow trivially from, these roles. Individualists maintain, by contrast, that we should not in this way think of ourselves as essentially political beings; instead, our political obligations rest not on our institutional roles, but on contingent relations between political associations and ourselves (such as our consent to them or our receipt of benefits from them).

Individualist theories of political obligation can be further divided into two classes. ‘Voluntarists’ ground people’s political obligations in their personal performance of politically significant voluntary acts, such as promises or contracts to obey, consent to authority, or free acceptance of benefits from a political scheme. ‘Nonvoluntarist’ (individualist) theories hold that no such voluntary performance is necessary for political obligation. Simple nonvoluntary receipt of benefits may bind us, for instance, or the moral qualities of a government may bring it under more general moral duties that require us to give it our support.

The fourth group consists of the anarchists. ‘Anarchist’ theories deny altogether the existence of political obligations. While the inspiration for an anarchist view can itself be either individualist or communitarian, anarchism rejects the conservative assumptions of the standard theories in both categories.

None of these four approaches to the problem of political obligation is new. Indeed, three of them are suggested in various passages in Plato’s *Crito* (see Plato §6). There, Socrates argues first that he must obey the state’s commands because it has made him what he is (50d-e), suggesting a communitarian derivation of obligation from his politically-constituted identity. But he later contends as well (in a more individualist fashion) both that his obligations stem from a tacit promise to obey (51e-53a) and that the care and benefits provided for him by the state bind him to obedience (51c-e) (suggesting, respectively, voluntarist and nonvoluntarist positions).

2 Communitarian theories

Drawing their inspiration from Plato, Aristotle, Hegel and Wittgenstein, communitarians argue that our plans and purposes, our values, and thus essential aspects of our very identities, are given us by the roles we play within linguistic, social and political communities (see Community and communitarianism). It is as a result misleading to think (as individualists do) of our moral relation to the state as somehow optional or contingent. Citizen and state are not like unrelated contractors in economic negotiations, as voluntarists seem to maintain. Nor do our political obligations rest on externally derived moral duties, as nonvoluntarists claim. Rather we have obligations to obey the rules of our communities because this is part of what it means to be members of those communities. Who we are in our social contexts tells us what obligations we have. To ask for any further explanation of political obligation would be to ask the unintelligible question: why should our lives be regulated by what makes us what we are (Green 1882). The proper account of political obligation looks to justifications that are internal to our practices, not external to them.

Communitarians whose sympathies are Wittgensteinian typically advance some version of a conceptual argument: that political obligation is conceptually tied to membership in a particular political society, and that membership is not voluntary, optional or contingent upon external justifiability (Pitkin 1965-6; Horton 1992). Comparisons are often drawn between political obligations and family obligations (which are similarly nonvoluntary) or obligations to friends and colleagues (which we often ‘acquire’ rather than choose, and which we continue to have even when we do not want them). Lest we think these analogies appeal only to nonliberal theorists, it should be noted that they have motivated at least one well-known political liberal to offer an account of political obligation as an ‘associative’ or ‘communal’ obligation (Dworkin 1986).

Communitarians whose sympathies are more Aristotelian or Hegelian tend to modify or add to these conceptual arguments. Political community, they argue, is essential to human flourishing and the development of human moral capacities, such as agency or autonomy or self-conscious valuation. As such, not only are our political communities owed obedience and support (as long as they encourage this development); we have also an obligation to belong to and facilitate political communities (Taylor 1979). On this view, because our political relations contribute essentially to our identities as moral agents and autonomous choosers, these relations cannot themselves be thought of as freely chosen or as dependent on moral principles that bind us independent of our political roles. While feminist political philosophy has not focused centrally on the problem of political obligation,

most feminist accounts are also broadly communitarian or 'contextual' in character (Hirschmann 1992).

3 Voluntarist theories

Individualist theorists deny that we are essentially political beings, that our political obligations are a simple function of our identities as socially constituted persons (Green 1988). The political realm is a contingent, nonessential aspect of human life (even if it is also a quite typical feature of it), and our unchosen social and political roles cannot simply be assumed justifiably to define our moral responsibilities. This denial of the communitarian's 'political naturalism' is most strongly stated (or assumed) by voluntarists. The classical individualist theories of political obligation were mostly voluntarist in character, and nearly every voluntarist theory prior to the twentieth century was some variant of a consent or contract theory of political obligation (see Consent; Contractarianism §5). The terms of the modern debate about political consent were set most clearly by Locke (see Locke, J. §10). According to Lockean consent theory, political obligations are grounded in the personal consent of individual members to the authority of their government or political society. This consent can be either express (as in direct agreements or contracts to obey, oaths of allegiance, and so on) or tacit (indirectly given by other behaviour that signifies consent). But voluntary, intentional consent of some sort is necessary for political obligation; government without popular consent is tyranny (Locke 1690). Consent theories tend to focus on tacit (implicit, indirect) consent, given the apparent paucity of express consenters in modern political communities. Favourite candidates for acts of tacit consent on which our political obligations might rest include continuing to reside in a state one is free to leave, freely taking benefits from the state, voting in democratic elections and accepting adult membership in a state.

The intuitive appeal of consent theory (in its respect for each person's free choices) is considerable and its history is long and distinguished. But the theory has throughout that history been plagued by fundamental complaints that it is not in fact applicable to real political life. Real political societies are not voluntary associations and real citizens seldom give even tacit consent. Indeed, all of the acts alleged to constitute tacit consent to government are typically performed without any intention to consent to government authority at all; and they are often performed unfreely, simply because of the high cost of alternatives, such as emigration (Hume 1739-40). But if morally binding consent must be intentional and voluntary, such facts seem to force us to the conclusion that few citizens of actual states count as even tacit consenters and that consent theory cannot adequately account for the political obligations we believe these citizens to have (Simmons 1979).

Consent theorists have responded by specifying further conditions that must be satisfied if 'government by consent' is to be achieved (Beran 1987) and by insisting that genuine, binding consent is only given by full involvement in the political life of a participatory democracy (Pateman 1979). These responses, of course, involve to a certain extent giving up conservative ambitions in thinking about political obligation. But a more conservative move within the voluntarist camp has been to surrender instead the idea of consent as the paradigm ground of political obligation. Thus, fairness theories of political obligation maintain that our obligations are owed as reciprocation for benefits accepted from the workings of our cooperative legal and political institutions (Hart 1955). Consent to these institutions is not necessary for being obligated to support them and abide by their rules. Rather, it is enough that we freely accept the benefits that flow from the cooperative sacrifices of others; for to do so while refusing to do our own parts in the scheme would be to take unfair advantage of those who cooperated in good faith.

Fairness theories, however, have also been subjected to steady contemporary criticism. One prominent early proponent of a fairness account has since rejected it (for reasons followed by many others), arguing that citizens in actual political societies seldom freely accept the benefits their societies provide, but instead merely receive them without real choice (Rawls, J. §1). And others have challenged the portrayal of modern political communities as large-scale cooperative schemes, sufficiently like small-scale cooperative enterprises to give rise to similar obligations of fairness (Simmons 1979). As a result of these criticisms, many individualist theorists have concluded that voluntarist accounts of political obligation are unpromising, and they have turned instead to nonvoluntarist theories.

4 Nonvoluntarist theories

The distance from voluntarist to nonvoluntarist (individualist) theories of political obligation can seem at first

glance quite small. There are, for instance, nonvoluntarist versions of the fairness theory, according to which it is not our free acceptance of the benefits of government that grounds our obligations, but rather our (possibly nonvoluntary) receipt of certain cooperatively produced goods that are of great importance to both individuals and societies (such as the benefits of law enforcement and national defence) (Klosko 1992). And there are also nonvoluntarist gratitude theories, which view political obligation as a special case of the obligation to reciprocate in an appropriately grateful fashion for benefits conferred on us by others (Walker 1988). Such nonvoluntarist benefit (or reciprocation) theories obviously have much in common with voluntarist fairness theory. But the theoretical distance of such accounts from voluntarism is in fact considerable. For political obligations, instead of resting on what individuals choose to do (as in voluntarism), are now taken to rest on what merely happens to those individuals and on the virtues of the institutional arrangements under which they live. This is to locate the moral heart of political relationships in a quite different area than that identified by the voluntarist.

The distance from voluntarism is similarly deceptive in the case of hypothetical contractarian accounts of political obligation. Our obligations, on this approach, are determined not by our personal consent to (or contracts with) our political authorities, but by whether we (or some suitably described, more rational or more neutral version of us) would have agreed to be subject to such authorities in an initial choice situation (Rawls 1971; Pitkin 1965-6). Hypothetical contractarianism, because it centrally utilizes the idea of contract or consent, may at first seem to be just a development of or a variation on voluntarist consent theory (and its advocates often present it as such). In fact, however, hypothetical theories are no longer concentrating on individual choice or on specific transactions between citizen and state, but instead on the quality of the political institutions in question. Hypothetical contractarians ask whether our laws or governments are sufficiently just or good to have been consented to in advance by rational parties, in an initial specification of their terms of social cooperation. What matters here is not choice or individual history, but the nature and quality of government (see Contractarianism §5).

This emphasis on quality of government is also present in utilitarian theories of political obligation, despite their constant opposition to contractarian views. According to utilitarians, our political obligations are based in the (direct or indirect) utility of support for and compliance with government (Hare 1976). Because obedience generally promotes social happiness, it is typically obligatory. But, of course, obedience only promotes social happiness if the specific laws or government in question are well-framed, utility-producing devices; our political obligations are thus derived reasonably directly from determinations of governmental quality.

The well-known contemporary objections to utilitarian moral theory (Rawls 1971) are only one reason why critics have found unconvincing the utilitarian account of political obligation. Like all nonvoluntarist (and communitarian) theories, utilitarians accept birth and benefaction within certain kinds of political communities as sufficient to justify political bonds. Critics worry that this misses the transactional, bilateral character of legitimate cooperative relationships, which in political cases should involve citizen and state jointly committing themselves to acceptable arrangements.

5 Anarchist theories

Anarchism comes in many forms (see Anarchism). Variants range from communist to libertarian (from Karl Marx to Lysander Spooner). Some anarchists deny the very possibility of a legitimate state, while others deny only the legitimacy of existing states. Some urge the destruction of existing states, others only selective disobedience to them. But all forms of anarchism are united in rejecting the conservative assumption that most citizens have political obligations.

It is illuminating to recall that much of the force of communitarian, nonvoluntarist fairness and voluntarist fairness theories of political obligation derives from the perceived failure of consent theory. Consent theory has been attacked as inapplicable to real political societies and alternative theories have been preferred for their superior ability to explain our obligations. But it is important to see that this attack rests squarely on a conservative approach to the problem of political obligation. If the conservative assumption is abandoned, consent theory no longer appears defective. Rather, it can be taken to specify the true grounds of political obligation, grounds that are simply not satisfied in actual or possible states. Voluntarist anarchism thus re-emerges as an interesting theoretical possibility.

Classical anarchism (of both communitarian and individualist varieties) recommended the abolition of the state.

Late twentieth-century philosophical anarchism merely denies the existence of (widespread) political obligation, usually on voluntarist grounds, without making any revolutionary practical recommendations. Some philosophical anarchists have argued on a priori grounds that the authority of the state is inconsistent with individual autonomy (Wolff 1970). Others have argued only that existing states fail to satisfy the voluntarist requirements for political obligation (Simmons 1979). Both forms, however, have been attacked for failing to appreciate the overall force of multiple contributing grounds for political obligation (Gans 1992).

See also: Civil disobedience

A. JOHN SIMMONS

References and further reading

All of the below involve detailed and sometimes subtle argument, but all are accessible also to readers without formal training in philosohy.

Beran, H. (1987) *The Consent Theory of Political Obligation*, London: Croom Helm.(The most complete contemporary defence of consent theory.)

Dworkin, R. (1986) *Law's Empire*, Cambridge, MA: Harvard University Press.(Chapter 6 offers a defence of communal obligations.)

Gans, C. (1992) *Philosophical Anarchism and Political Disobedience*, Cambridge: Cambridge University Press. (Attack on philosophical anarchism.)

Green, L. (1988) *The Authority of the State*, Oxford: Oxford University Press.(Good general discussion of problems of authority and obligation.)

Green, T.H. (1882) *Lectures on the Principles of Political Obligation*, Ann Arbor, MI: University of Michigan Press, 1967.(Classic statement of a communitarian approach to political obligation.)

Hare, R.M. (1976) 'Political Obligation', in T. Honderich (ed.) *Social Ends and Political Means*, London: Routledge & Kegan Paul.(Utilitarian account of political obligation.)

Hart, H.L.A. (1955) 'Are There Any Natural Rights?', *Philosophical Review* 64 (2): 175-91.(First clear statement of the fairness account of political obligation.)

Hirschmann, N. (1992) *Rethinking Obligation*, Ithaca, NY: Cornell University Press.(A feminist critique and reorientation of political obligation theory.)

Horton, J. (1992) *Political Obligation*, Atlantic Highlands, NJ: Humanities Press.(The most thorough recent discussion of all aspects of the problem; comprehensive bibliography.)

Hume, D. (1739-40) *A Treatise on Human Nature*, ed. P.H. Nidditch, Oxford: Oxford University Press, 2nd edn, 1978.(Influential critique of consent theory and a defence of broadly utilitarian view is given in book III, part II, chapters VII-X.)

Klosko, G. (1992) *The Principle of Fairness and Political Obligation*, Lanham, MD: Rowman & Littlefield.(A nonvoluntarist fairness theory.)

Locke, J. (1690) *The Second Treatise of Government*, in *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1960.(Classic statement of the consent theory of political obligation, centred in chapters 7 and 8.)

Pateman, C. (1979) *The Problem of Political Obligation*, Berkeley, CA: University of California Press.(Discusses the history of the problem and defends a Rousseauian approach.)

Pitkin, H. (1965-6) 'Obligation and Consent, I and II', *American Political Science Review*, 59 (4): 990-9 and 60 (1): 39-52.(Defends both a Wittgensteinian, communitarian approach and a version of hypothetical contractarianism.)

Plato (*c*.395-87 BC) *Crito*, in *Plato: The Collected Dialogues*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1961.(The first recorded discussion of the problem of political obligation.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(The most important hypothetical contractarian account of political obligation is given in chapter 6.)

Simmons, A.J. (1979) *Moral Principles and Political Obligations*, Princeton, NJ: Princeton University Press. (Discusses individualist theories of political obligation and defends philosophical anarchism.)

Taylor, C. (1979) 'Atomism', in A. Kontos (ed.) *Powers, Possessions, and Freedom*, Toronto, Ont.: University of Toronto Press.(Defends a Hegelian approach and an 'obligation to belong'.)

Walker, A.D.M. (1988) 'Political Obligation and the Argument from Gratitude', *Philosophy and Public Affairs* 17 (3): 191-211.(Develops a gratitude theory of political obligation.)

Wolff, R.P. (1970) *In Defense of Anarchism*, New York: Harper & Row.(A defence of an a priori version of philosophical anarchism.)

Observation

Observation is of undeniable importance in the empirical sciences. As the source of information from the world itself, observation has the role of both motivating and testing theories. Playing this role requires more than just opening our eyes and letting nature act upon us. It requires a careful attention to the information conveyed from the world so that an observation is meaningful. Scientific observation, in other words, is more than a physical act of sensation; it must be an epistemic act as well, with sufficient meaning and credibility to contribute to knowledge. A report of an observation, therefore, must be more than a 'Yes, I see'. It must describe just what is seen, 'I see that ________'.

This obligation to make observation relevant to theory suggests that there is an essential influence of background theories on the observations themselves. The theories we believe or wish to test tell us which observations to make. And describing the results of observations, that is, bringing out their informational content, will always be done in the language of the conceptual and theoretical system already in place. For these reasons, observation is said to be indelibly theory-laden. And the influence of background beliefs is even greater in cases of indirect observation where machines, like microscopes and particle detectors, are used to produce images of the objects of observation. Here, the reliability of the machines, and hence the credibility of the observation, must be based on a theoretical understanding of the interactions that are the links in the chain of information.

The influence of theory on observation is often seen as a threat to the objectivity of the process of testing and verification of theories, and hence of science in general. If theories are allowed to, indeed required to, select their own evidence and then to give meaning and credibility to the observations, the testing process seems to be unavoidably circular and self-serving. Observation that is theory-laden would guarantee success. But a look at the history of science shows that it does not. There are plenty of cases of observations that are used to disconfirm theories or at least undermine the theorist's confidence. Perhaps there is a kind of observation that is not influenced by scientific theory and can serve as a common, objective source of information to put theories to a rigorous and meaningful test. Or perhaps all scientific observation does bear the influence of background scientific theories, but not necessarily of the theory the observation is being used to test. This independence between the theories that support an observation and the theory for which the observation serves as evidence can break the circle in the process of testing and perhaps restore objectivity.

1 The role of observation

Empirical sciences, whether they are natural sciences like biology, geology and astronomy, or social sciences like psychology and anthropology, make claims that are contingent on what is going on in the world. Non-empirical enterprises like mathematics and logic aspire to what must be true of the world and to what things are possible in the world. But the goal of science is between these two extremes; it is to describe what is in fact true of the world and what does in fact happen. This requires some contact with the actual world and some information from the particular objects of interest themselves. This is the role of observation.

Observations in science are essential both to stimulate the formation of theories and to regulate the testing of theories. In the role as stimulus, observation of perplexing or surprising events will prompt questions of why these things happen, and thereby lead to theoretical explanations. Recurring phenomena call for an understanding of the principles of organization and unification, the theories, to make sense of the observations. And the observation of similarities among phenomena, as Newton noticed similarities in the events of a falling apple and the orbit of the moon, suggest the formation of a unifying theoretical account. In these ways and more, observation can initiate and guide the preparation of theory.

A proposed theory, a hypothesis, can then be tested by being used to make predictions of what will be observed under specified conditions. Actually doing the observations is then an essential part of testing and verifying the theory. Various models of this process of testing and confirmation have been proposed and advocated, differing in the nature of the relation between theory and observation (see Confirmation theory). These models are all alike though in that they require some relation, some contact between theory and observation.

The importance of observation in science is of course because of the supposition that it is the source of information

from the world itself - unlike theory, that begins as information from the minds of scientists. Observation is intersubjective, public and accessible to anyone willing to have a look. This is the reason to expect that observation will provide a source of agreement, a common ground of information to arbitrate disagreements between differing theoretical accounts. Observation must be the source of facts, if anything is, and must provide the objective foundations of science, if anything can.

Whether observation can fulfill of these weighty expectations is itself an issue of disagreement among scientists and philosophers. Understanding the role of observation in science and its potential to be objective or foundational requires a look at the details of what observation must be like in order to contribute to scientific knowledge, and the details of what it is in fact like in the activities of actual science. The questions to ask are these: can observation be a source of agreement to decide between differing theories? Does it in fact play this role in science? And is observation a source of objective information about the world?

2 Informational content

To play the role as stimulus and evidence for scientific theories, observations must fulfil some basic, informational requirements. To serve as evidence, that is, to be a meaningful influence on a theory, an observation must be *of* something that is relevant to the theory. Scientific observation must be informative in the sense of being an observation *that* something is the case. It must be more than just a physical act of sensation; it must be an epistemic act with sufficient meaning and credibility to contribute to knowledge. Credibility is achieved only with some assurance that the observation has been done with care and precision, and in conditions that are conducive to accuracy. And a meaningful observation must be one described in a language that is relevant to the theoretical knowledge it is meant to challenge, change or re-enforce.

An effective way to attend to the informational content of an observation and to its epistemic status is to focus on observation reports rather than on the physical act of perception. The useful question to ask of an observational contact with the world is not simply, ‘do you see it?’, but the more revealing, ‘what do you see?’ or ‘do you see that this is that?’ Answers to these questions, that require more than a nod, give the full description of the observation. These observation reports are the most basic empirical data in science. Even if the report is private, a report to oneself, a useful observation must be presented in an informational form, observing, for example, that the sky is blue, because undescribed sensations can have no impact on theory. Undescribed sensations cannot serve as evidence in science. They neither motivate nor test theories.

This emphasis on language motivates positivists to attempt a neat dichotomy between the observational and theoretical vocabularies of science (see Logical positivism §4). Whether or not this sort of distinction can be made, and whether it has any significance for the structure of scientific knowledge, will be discussed in §6, below. The analysis of observation language, though, rather than simply the physical act of observation, accurately acknowledges that observation in science is more than simply opening one’s eyes and having a look. It is an important and careful activity whose epistemic value depends on what is observed and in what way. And these are assessable only of a meaningfully described observation, one with assertive informational content.

3 Theory-laden observation

The informational requirements as described above lead directly to the concern that observation in science is essentially theory-laden. Observation that is relevant to background knowledge will necessarily bear an influence from background knowledge. If this is true, that observation is influenced by theory, then observation loses its natural innocence. The activity of observation is not a passive acceptance of ready-made information (in the sense of ready-described and ready-justified) given by the objects of interest.

There are three distinct ways that theories can influence observations and that what we see can be dependent on what we think.

For one, background knowledge and the theories we have in mind are the source of authority in deciding what observations to make. Gathering data in science is neither comprehensive nor random, and it is a theoretical understanding of the world that directs our attention to what is important to observe and measure. This is not to suggest that advocates of theories seek out and report only the observations they are confident will reflect favourably on their theory. There is a difference between selecting the important tests and selecting rigged tests,

and while the persuasion of theory can determine which observations to perform, it cannot determine how they will turn out. Only the act of observing can do that.

Another way to think of the theoretical impact in selecting observations is to note that what we observe of the world is always just a small sample of the whole thing, and it is important to be sure it is a representative sample. To be representative a sample must be relevantly diverse, and issues of relevance must always be settled under the influence of some theoretical understanding or other. Testing the effects of some drug, for example, may require observations of people of various ages, weights, medical histories, and so on. But the sample does not have to be representative of different hair colours or preferences in flavours of ice cream, because our understanding of physiology does not regard these as relevant to the effects of drugs. Thus a good sample is one that has been done in light of current background knowledge.

A second influence that theories can have on observation is in assessing the credibility and reliability of the observation report. Scientific observations are acceptable only if they are carefully done. Conditions must be correct, in order to reduce the chances of distortion, and the propriety of conditions is often prescribed by theory. This is particularly apparent in the case of machine-aided observations (see §7). A microscope, for example, is an acceptable tool in science because there is good reason to think that the image produced is an accurate view of the specimen. The good reason is based in theories of optics describing how the microscope works. This theoretical support is rarely explicit in the use of microscopes or the claims of microscopic observation, but it is a source of credibility that must be available in the scientific community. Scientists do not accept the images produced by just any machine, only those they know to be reliable. And even when no machines complicate the observation, the pragmatics of acceptability of observation reports are still leveraged by theory and background knowledge. If an observation report threatens a well-entrenched theory, it is common to look first for a mistake in the observation. This use of theory to edit observation is simply the manifestation of the high standards expected of scientific evidence.

The third aspect of complicity between background knowledge and observation is in giving meaning to the evidence. An observation must be of something that is relevant to theory and so it must be described in a language that makes contact with theory. The observation must be fitted into the system of theoretical concepts. The necessary promotion from a merely physical act of sensation to an epistemic act of observation takes place under the authority of the theories in our background knowledge. To a physicist, for example, the streaks in a cloud chamber are indicative of the passage of subatomic particles. The streaks mean particles, and it is only at this informative level that the observation functions as evidence, useful for making an impact on theory. And it is only under the influence of theory that the description as particle tracks is possible.

Taken together, these three ways in which theories guide observations lead to the notion of ‘theory-laden observation’ (see Duhem, P.M.M.; Hanson, N.R.). The claim is that this is an essential aspect of science and scientific evidence, and not simply a lazy or suspicious way that science gets done by conniving scientists. To say that observation is theory-laden is to focus Kant’s more general claim, that experience without concepts is blind, on the case of science.

4 Using theory-laden observation

Theory-laden observation has serious implications for scientific method and the role of objectivity in science. If all observation bears the imprint of some theory or other, then the results of observation are not pure information from the objects in the world themselves, nor will they necessarily be data that are shared and of universal agreement among scientists. Persons with different theoretical backgrounds may well judge different observations as important and credible, and because of this, hold their theories to different standards. They may even give different meanings to the same perceptual event, and thus draw different conclusions from seemingly the same tests. This poses an immediate threat to objectivity in the testing of theories. If observations cannot provide a common standard against which to test competing theories, those theories may be incommensurable, leaving us with no reasonable way to rate one more likely to be true than the other (see Incommensurability; Underdetermination).

It would be a mistake to think that theory-laden observation is a guarantee that empirical testing will always come out positive and in support of the theories being tested. The history of science is full of examples in which observations went against theory. An interesting case is the solar neutrino problem. According to theory, the sun,

like all other stars, is fuelled by nuclear reactions occurring at its core, reactions that produce neutrinos. The neutrinos ought to be detectable on the earth. The observation of neutrinos is enormously complicated and indirect, and it is profoundly influenced by theory. Nonetheless, the results of observations have been bad news for the theory being tested, the claim of nuclear reactions in the sun. Not nearly as many neutrinos as predicted have been observed. Thus, complicity between theory and observation does not necessarily compromise the testing process in the sense of guaranteeing a supportive outcome.

It is important to confront the threat to objectivity in the light of the details of each particular case of observation in science. A general analysis of the nature of observation and evidence may be enough to show that there is some theory complicity in scientific observation, but beyond that we should ask specifically which theories are involved in any particular case. Perhaps the theories that influence an observation are independent of the particular theory the observation is meant to test. In such a case, there is no circularity in the testing process and hence no threat to objectivity. A look at specific cases might also show that the requisite conceptual ingredient in observation does not rely on the theories that we take to be the currency of science but only on some specifically perceptual background knowledge. If the background knowledge is simply common sense, then the worry about theory-ladening is of little concern to science. Observation in this case can provide data of common consent, useful for settling issues of dispute between competing theories.

Studying actual cases of observation in science can also show whether theories supply a tacit influence over observations, or if observers and those who make use of observation reports are aware of the theoretical contribution in an explicit interpretation of perception. If there is this awareness, the conceptual or theoretical influences can perhaps be abstracted away to reveal a perceptual core. The question will then become whether or not these purely perceptual data have sufficient informational content to be useful as scientific evidence.

5 Indirect observation

Many observations in science are physically indirect in the sense that the observer does not interact immediately with the object itself. Intermediate interactions are required to convey information of the object to an observer. The most obvious cases of indirect observation involve tools that enhance our powers of perception, from a simple magnifier, to a compound telescope, to an enormous bubble chamber. Using such machines, scientists report the observation of objects such as the moons of Jupiter, alpha particles, or even events in the universe just seconds after the Big Bang. There are also cases of indirect observation that do not involve instruments, cases in which the object of observation has left a natural imprint on what can be directly perceived. In this sense, fossils are an image of life in the past, and striations in bedrock show the movement of glaciers (see Geology, philosophy of §1).

All of these cases of indirect observation have in common that the information from the object of interest is mediated by a more-or-less lengthy causal chain from object to scientist. The multiple interactions can impose a physical influence on indirect observation, much as theoretical complicity adds a conceptual influence. And there is an obvious correlation between the two in that more physical complication in an observation is likely to require more theory to keep track of the flow of information.

The image produced in an indirect observation may, in some cases, not even resemble the object that is allegedly observed. Audible clicks of a Geiger counter could be the final stage in the observation of a beta-decay event or the passage of a cosmic ray. The length of a column of mercury in a thermometer allows us to observe the temperature of some other object. But this may be stretching the concept of observation too thin, having come so far from the original expectation of a source of manifest and intersubjective information from the objects themselves. Such highly indirect observation that demands overt interpretation draws the question: do these events even count as being observation?

Answering this sort of question is complicated by the realization that indirectness of observation is a matter of degree, depending on the length and complication of the chain of interactions from the object of observation to the final image. This realization clarifies the important questions to ask of indirect observation. Is the degree of physical removal of the object epistemically significant in the sense that a longer journey makes the information less reliable and makes the event somehow less of an observation? Is there a cut-off, a way to mark an indirect link between observer and specimen as too long or too complicated to count as an observation?

6 Observability

A distinction between what can and what cannot be observed is crucial to an empiricist philosophy of science. Observability is epistemically significant insofar as observation is the foundation of testing and verification of scientific claims (see Demarcation problem §5). The theme of empiricism is that scientists are justified in believing that what they claim about observables is true, but not justified in what they claim about unobservables. There is no denying that the stuff of the world behaves *as if* it is composed of atoms, but that is no reason to believe that it really *is* composed of atoms (see Fictionalism; Scientific realism and antirealism §3). Only by observing can we have reason to believe, and so only of what is observable can we hope to have justified belief.

Empiricism incurs an obligation to supply the criteria for distinguishing between observables and unobservables. This might seem as easy as opening our eyes and seeing what we can see, but observability cannot be decided on the basis of observation, since what is observable is not always observed. Furthermore, what has not been observed is not necessarily unobservable. In fact, anything that is worth looking for has the status of being unobserved but observable. Planets around distant stars, for example, are worth looking for because, being relevantly similar to the planets we see in our solar system (for example, being large and solid), they are observable, though none has been observed. The famous Michelson-Morley experiment to measure the earth's motion through the electromagnetic ether was definitive in its results only by considering the ether to be in some indirect way observable. Since it was observable but not observed, the conclusion was that the ether simply did not exist. Had it been judged to be unobservable there would have been no surprise at the null results. In all cases there will be some antecedent conceptual reason to think that an object is observable, in at least an indirect way, and hence worth looking for.

Observability is not an intrinsic feature of the object or event in question. It is a relational property, observable-to-us, and so it is subject to change if the human capacity to perceive or to go places in the universe to have a look changes. This underscores the fact that observability is intended to have *epistemic*, not *ontological* significance. The empiricist is not claiming that what is unobservable does not exist. Rather, what is unobservable is such that we have no justification in claiming that it exists or that it does not. Justification can of course change, as observability can change.

Any significant distinction between what counts as observable and what does not must acknowledge the complication of the act of observation itself. Observability cannot be merely a function of the perceptual limitations of human observers, since observation must be more than merely a physical perceptual event; it must be an informative epistemic event. Nor can observability depend on being independent of theory if all observation is to some degree influenced by the observer's theoretical background.

There is a potential confusion in this issue of observability, as to whether it applies to the entities of science or to the language of science. It is in the best interest of empiricism to focus on the observability of the entities themselves rather than on the language, since how we describe something will always be influenced by our conceptual and theoretical background, but whether or not we see something is just a fact of our perceptual prowess. A child may not be able to describe the planet Venus as a planet or as Venus in particular, but they can surely see it when directed to the proper location in the sky. This strategy of deciding observability of things, independent of their description in language, threatens the informational requirements of observation. Undescribed observations cannot function as evidence and cannot contribute to scientific knowledge. That is why children and others who lack sufficient training in the conceptual background of scientists cannot step right up and do science, no matter how good their eyesight. Noting the need for informative observation, the issue of observability is going to have to involve some aspects of language and its theoretical associations.

Drawing a distinction between observables and unobservables is further complicated by the cases of indirect observation. It would be difficult to argue that any intermediate instrumentation whatsoever disqualifies an image from being a genuine observation, since viewing the world through a pane of glass or eye-glasses is surely observing the world. But there is little difference between this and using a magnifying glass, and so if what can be seen through a window is observable then so is what can be seen through a magnifier. And this begins a continuum of indirectness with no clear break between what ought to count as observable and what ought not. If the single lens of a magnifying glass enhances our powers of observation and expands the domain of what is observable, then why not the same for a microscope or a telescope, both of which are just a series of magnifiers stacked together? Why not an electron microscope, which does with electrons just as an optical microscope does with light? And so

on. The challenge is to draw a line between what counts as observation and what does not in a way that is not arbitrary and that bears the epistemic significance required of empiricism.

7 Reliable observation

One can acknowledge that there is no sharp boundary between what is observable and what is not, yet still maintain that observability is of great epistemic importance. There are clear cases, after all, of things that we can observe, that the sky is blue for example, and equally clear cases of things that we cannot, that the 'up' quark has an electric charge of $+2 \setminus 3$. The colour of the sky is clearly observable, while the electric charge of a quark is not. And we have a higher level of justification for claims about the colour of the sky than we do for claims about the charge of a quark, precisely because of the difference in their observability. Similarly, in cases of machine-assisted observation such as a cloud chamber, there is a clear sense in which we observe the streaks in the chamber and from these we infer the presence of particles and subatomic events. The features of the streaks are more immediate to our perception, and for this reason, claims about the streaks are more reliable than are claims about the particles. Thus, observability, though not a sharp demarcation, is nonetheless significant in evaluating the justification for scientific claims.

At this point it may be advisable to consult science itself to understand differences of observability and their significance to the status of scientific entities and statements. This does not mean simply adopting the language of science and counting as observable whatever scientists say they observe. It requires an understanding of each scientific case of observation to see what, according to the scientific description, is going on. Science can provide a description of the causal chain and the flow of information from a putative object of observation to the observer, and can in each case reveal which theories are involved in giving meaning and credibility to the observation. A scientific account of the physical indirectness of an observation will detail not only the length of the interaction chain, but the reliability of the passage of information through each link. Understanding the interactions themselves shows that in each case there is some information that can and some information that cannot be conveyed. The physics of quarks, for example, indicates that while such properties as electric charge and mass can be discovered through interaction with a quark, the colour (not the normal visual colour but the specifically quantum-chromodynamical property given the whimsical name 'colour') can not. There is no interaction that reveals quark colour, and so this property is unobservable in principle, where it is a principle of science, not, philosophy. No amount of indirectness can sanction a claim to observe the colour of a quark.

The scientific account will also reveal the theoretical influence on an observation. Consulting science on the matter of observability will show exactly which theories are required to describe the physical indirectness of an observation and to give meaning to what is seen. This allows an assessment of the reliability of the theoretical support as well as its independence from the theories the observation is being used to test.

Given the complication of the issue of observability, and the need to consult the science of each case, the central question must be revised to be amenable to degrees of observability and to the specifics of each case. Of a scientific entity or some information in the world, it is not productive to ask simply whether it is observable or not. Better, more useful, to ask for the whole story, namely, how is it observed, or how is the information acquired? With the details explicit, the epistemic status of the information can be evaluated.

See also: Experiment; Information theory; Measurement, theory of; Theories, scientific; Theory and observation in social sciences

PETER KOSSO

References and further reading

Bogen, J. and Woodward, J. (1983) 'Saving the Phenomena', *Philosophical Review* 97: 302-52.(Introduces a distinction between data and phenomena that emphasizes that observation concerns data as evidence for claiming about phenomena, while phenomena are the proper object of explanation.)

Fodor, J. (1984) 'Observation Reconsidered', *Philosophy of Science* 51: 23-43.(Readable and relevant to the material of §§3, 5; arguing that observation is not theory-laden in any important way.)

Fraassen, B. van (1980) *The Scientific Image*, Oxford: Oxford University Press.(An empiricist account of the distinction between observable and unobservable entities.)

Hanson, N.R. (1958) *Patterns of Discovery*, Cambridge: Cambridge University Press; repr. 1961.(The origin of

the phrase 'theory-laden observation'.)

Hacking, I. (1983) *Representing and Intervening*, Cambridge: Cambridge University Press.(Chapter 10, 'Observation' and chapter 11, 'Microscopes', discuss the role of theories in observation, in particular the role of optical theories in the use of microscopes.)

Kosso, P. (1992) *Reading the Book of Nature,* Cambridge: Cambridge University Press.(Chapter 6 gives a clear overview of observation in science, and chapter 9 makes the case for independence between a theory being tested and the theories that influence observation.)

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press; 2nd edn, 1970. (An extremely influential book that includes the influence of theory on observation and the implications for incommensurability.)

Maxwell, G. (1962) 'The Ontological Status of Theoretical Entities', in H. Feigl and G. Maxwell (eds) *Minnesota Studies in the Philosophy of Science,* vol. 3, Minneapolis, MN: University of Minnesota Press.(Relevant to §5 in that it describes the continuum of indirectness of observation.)

Shapere, D. (1982) 'The Concept of Observation in Science and Philosophy', *Philosophy of Science* 49: 485-525. (Using science to describe observation as the flow of information.)

Woodward, J. (1989) 'Data and Phenomena', *Synthèse* 79: 393-472.(A detailed elaboration on Bogen and Woodward 1983.)

Occasionalism

Occasionalism is often thought of primarily as a rather desperate solution to the problem of mind-body interaction. Mind and body, it maintains, do not in fact causally affect each other at all; rather, it is God who causes bodily movements to occur 'on the occasion of' appropriate mental states (for example, volitions), and who causes mental states, such as sensations, on the occasion of the corresponding bodily states (for example, sensory stimulation).

This characterization, while correct so far as it goes, is seriously incomplete. Occasionalists have seen the lack of real causal influence between mind and body as merely a special case of the more general truth that no two created beings ever causally affect each other. The one and only 'true cause' is God, with created beings serving as the occasions for his causal and creative activity, but never as causes in their own right. (The one possible exception to this is that created agents may themselves bring about their own acts of will; this is necessary if they are to be in any sense free agents.) Occasionalism has always been held primarily for religious reasons, in order to give God the honour due to him as the Lord and ruler of the universe. It has never, however, been a majority view among philosophical theists.

1 Medieval occasionalism

The first thinker clearly to articulate an occasionalist position was the Muslim theologian al-Ghazali. He wrote in defence of orthodox Islam against the philosophers al-Farabi (§2) and Ibn Sina (§5), both of whom propounded emanationist systems based on a combination of Aristotelianism and Neoplatonism (see Creation and conservation, religious doctrine of §3; Aristotelianism in Islamic philosophy §2; Neoplatonism in Islamic philosophy §§2-3). One of al-Ghazali's fundamental objections to the emanationist scheme was that it was necessitarian and denied God the freedom due to him as creator and as the author of miracles. (For al-Ghazali, as for other orthodox Sunnis, human freedom was not a major concern.) Over against this, he affirmed the teachings of orthodox Ash'arite theology, according to which all natural beings are completely inert and the true and sole agent in nature is God (see Ash'ariyya and Mu'tazila). Thus, there is no necessity in nature itself which could constrain or limit God's omnipotent will. Unlike the Ash'arites, however, al-Ghazali presents a philosophical argument for this position. The only form of necessity he recognizes is logical necessity, and he has little difficulty in showing that causes do not logically necessitate their effects. The relation between what we take to be causes and effects is merely one of correlation and is purely contingent; the one real, productive cause of all things is God. In addition to this philosophical argument based on a critique of causality, al-Ghazali appeals directly to the theological dogma of God's absolute omnipotence. Indeed, al-Ghazali's vision of God is such that God is virtually the only true existent: 'There is no other being *with* Him, for Him to be greater than it…. [N]one has being save through His Face - so that His Face alone is' (quoted in Fakhry 1958: 72).

Al-Ghazali's rejection of philosophy was disputed by Ibn Rushd, but his views became widely accepted throughout Sunni Islam. His occasionalism was sharply criticized and rejected by both Maimonides (§4) and Aquinas, and never enjoyed wide support during the Christian Middle Ages. It was, however, embraced by the late medieval Ockhamist philosophers Pierre d'Ailly and Gabriel Biel.

2 Early modern occasionalism

Occasionalism in modern philosophy took its 'occasion' from the problems over causality that arose in Descartes' philosophy (see Descartes, R. §§8, 11, 13). (Not only the problem of mind-body interaction, however; causal interaction among bodies was also problematic for Descartes.) Descartes himself had tendencies towards occasionalism, and this view was adopted by the Cartesian philosophers Louis de La Forge and Géraud de Cordemoy. Arnold Geulincx may have been the first to state an important principle that is also found, in modified form, as a premise in arguments concerning causation given by Malebranche and Hume. The principle states that something cannot be done unless there is knowledge of how it is done; more specifically, people do not do what they do not know how to do. This quickly disposes of the contention that the mind causes bodily movements; in order for the mind to do this, we should have to know how to control the physiological processes involved in making such movements - the firing of neurons, for instance. Clearly, however, we do not know this. The principle also excludes causal efficacy for corporeal bodies, since these bodies obviously lack knowledge of how their

supposed effects are produced.

Nicolas Malebranche (§4) was clearly the most significant of the modern occasionalists. Basing his argument on the supposed impossibility of mind-body interaction, he writes that 'There is no real relation between one body and another, between one mind and another. In a word, no created thing can act upon another by an activity which is its own' (1688: 4.11). He argues, like Geulincx, that in order for our minds to be able to move our bodies we should have to understand, in full anatomical detail, how this is done. He also argues that God's causality pre-empts, as it were, any possible causality among creatures. No creature can exist unless God, by 'continuous creation', wills that it should exist. But in willing that a chair (for example) should exist, God must will that it exist in some particular place, in some particular state of motion or rest, and so on. Clearly, no created power (assuming there to be such) can cause the chair to be at rest or in motion unless God wills that it be so. But if God does will the chair to move, then necessarily it moves; there is nothing left to be done by any created agent. Thus, 'God communicates His power to created beings only because He has made their modifications the occasional causes… which determine the activity of His volitions in consequence of the general laws which He has prescribed to Himself' (1688: 7.10).

Occasionalism did not persist for long as a widely accepted view. But the influence of Malebranche was considerable, and is still felt today through Hume, many of whose sceptical arguments concerning causation derive from Malebranche.

3 The case for occasionalism

Is occasionalism simply a historical curiosity, or does it have the potential to become a viable position for present-day theists? Leaving aside Geulincx's far-from-evident principle, there are two principal arguments for occasionalism, based respectively on the critique of natural causality and on the pre-emptive causality of God. The argument from the critique of natural causality proceeds as we have already seen in al-Ghazali and Malebranche. Various criticisms are employed (many familiar to us through Hume) to argue that causality as we experience it in nature is not the necessary, genuinely productive relationship that we ordinarily take it to be. Natural causation having thus been disposed of, God is invoked as the 'true cause' for whose efficacious acts the various states of created beings are mere occasions. This argument, however, is dialectically weak. The occasionalist begins (as we all must) from our ordinary intuitions about causality developed through our ordinary, everyday causal interactions with the world. These intuitions are then undermined through the critique of causality. But then the occasionalist invokes those very same intuitions in order to point to God as the truly efficacious and productive cause of all that occurs; causation by God, however, is not subjected to the kind of critical analysis that has been applied to natural causation. Clearly, this procedure is inconsistent. If the critique of causality is as effective as occasionalists think, then (as Hume rightly saw) divine causation is, if anything, worse off than natural causation in consequence.

In view of this, the main burden of support for occasionalism needs to be borne by the argument for the pre-emptive causality of God, also employed by both al-Ghazali and Malebranche. (The argument of Malebranche based on continuous creation is a special case of this.) Classical theists are in agreement that God's causal activity must be seen as universal and pervasive throughout the created world. To suppose that there is some part of created reality in which God's activity is not involved is just to make that part of the creation independent of God, which detracts from divine dignity. Any attempt to circumvent this problem by designating different aspects of causation which pertain respectively to God and to creatures runs up against the complaint that the aspect assigned to creatures is removed from the sphere of God's activity, thus diminishing the honour due to God. Medieval Aristotelians tried to meet this challenge by the theory of dual causation, which claims to give full value both to the divine 'first cause' and to creaturely 'second causes'. This theory was spelled out in their doctrines of divine conservation and divine concurrence with creaturely action, but no detailed consensus was arrived at (see Creation and conservation, religious doctrine of §5). Contemporary philosophical theology has barely begun to address the issue, so it could fairly be urged that on this point occasionalism presents theistic philosophy with a challenge it has not yet met.

4 Occasionalism and immaterialism

Whether this argument provides effective support for occasionalism depends on whether occasionalism is itself a tenable view. It can be argued on both epistemological and metaphysical grounds that the material substance

posited by occasionalists is redundant and should be eliminated, leading to a Berkeleyan immaterialism (see Berkeley, G. §§3, 6-7). Epistemologically, occasionalism holds that material substances make no causal contribution towards our perceptions of objects. Occasionalists hold, to be sure, that light waves (for example) travel from observed bodies to our eyes, that nerve impulses connect the eyes with the brain, and that consequent to this we experience visual images. But there is no real causal connection between any two of these stages; in particular, the visual images are not produced by the brain (or even by the immaterial mind), but rather by God 'on the occasion' of the brain's being in an appropriate state. This means, however, that we have no direct evidence for the existence of the supposed intermediate stages in the process, and it is hard to avoid Berkeley's conclusion that it would be simpler for God to omit them.

A similar conclusion can be reached through metaphysical considerations. Occasionalism holds that created substances make no active causal contribution to subsequent states of the world. But then the occasionalist is faced with the question of whether created substances are endowed with causal powers at all. If they are, then God must somehow intervene to prevent the powers from being exercised - and indeed, the provision to created things of powers that are never exercised and are not intended by God to be exercised would seem highly unreasonable. So the occasionalist must deny that created beings possess causal powers. But what, if anything, might created substances consist in, if they lack such powers entirely?

Apparently the occasionalist answer is that while created substances lack active powers they do possess some passive powers; hence al-Ghazali's view of material beings as inert, and Malebranche's willingness to attribute movability and impenetrability (but not active power) to material objects. So as Freddoso suggests, 'God would supply all the active causal power in nature, while material substances would receive and channel God's causal influence as patients' (1988: 111). The difficulty with this is that the active/passive distinction, as we make it, cannot give occasionalism what it needs here. We do have some grip on the difference between an object's acting on another or being acted upon by another. (In many cases this would coincide with the direction of energy transfer.) But the very same molecular structure which makes a billiard ball impenetrable by another billiard ball also enables it, under appropriate conditions, to propel the other billiard ball in a desired direction. The active/passive distinction just does not cut deep enough to play the role it is cast in by occasionalism. But if material substances have neither active nor passive powers, then it is difficult to see the point of their existing - indeed, it is difficult to see what could be meant by claiming that they exist.

The correct conclusion would seem to be that the logic of occasionalism leads to Berkeleyan immaterialism, and that it is in this form, if any, that it may be able to survive as a viable contemporary option. A more attractive option, though slightly more remote from classical occasionalism, can perhaps be found in certain anti-realist interpretations of natural science (see Scientific realism and antirealism). Though diverging from the letter of both occasionalism and Berkeleyanism, these anti-realist views are akin to them in spirit in their desire to uphold science as an empirical enterprise while avoiding the mechanistic materialism which threatens on a realistic interpretation of scientific theories.

See also: Causality and necessity in Islamic thought; Edwards, J. §3; Miracles; Religion and science

WILLIAM HASKER

References and further reading

Berkeley, G. (1710, 1713) *A Treatise Concerning the Principles of Human Knowledge* and *Three Dialogues between Hylas and Philonous*, La Salle, IL: Open Court, 1986.(The basic texts for Berkeley's 'immaterialism'.)

Creery, W.E. (ed.) (1991) *George Berkeley: Critical Assessments*, London: Routledge & Kegan Paul.(Important collection of recent discussions of Berkeley's philosophy.)

Fakhry, M. (1958) *Islamic Occasionalism and its Critique by Averroes and Aquinas*, London: Allen & Unwin. (Sets out al-Ghazali's occasionalism in its historical context, and develops the Thomistic answer to occasionalism.)

Freddoso, A.J. (1988) 'Medieval Aristotelianism and the Case against Secondary Causation in Nature', in T.V. Morris (ed.) *Divine and Human Action: Essays in the Metaphysics of Theism*, Ithaca, NY: Cornell University Press, 74-118.(A penetrating study of the metaphysical aspects of occasionalism.)

al-Ghazali (1095) *al-Ghazali's Tahafut al-falasifah: The Incoherence of the Philosophers*, trans. S.A. Kamali, Lahore: Pakistan Philosophical Congress, 1963. (Problem 17 is the basic source for al-Ghazali's occasionalism.)

McCann, H.J. and Kvanvig, J.L. (1991) 'The Occasionalist Proselytizer: A Modified Catechism', *Philosophical Perspectives* 5: 587-615.(A contemporary defence of occasionalism; argues that occasionalism does not deprive laws of nature of their explanatory force.)

Malebranche, N. (1688) *Dialogues on Metaphysics and On Religion*, trans. M. Ginsberg, London: Allen & Unwin, 1923.(An excellent introduction to Malebranche's occasionalism.)

Malebranche, N. (1674-5) *The Search after Truth and Elucidations of the Search after Truth*, trans. T.M. Lennon and P.J. Olscamp, Columbus, OH: Ohio State University Press, 1980.(The basic source for the philosophy of Malebranche.)

Nadler, S. (ed.) (1993) *Causation in Early Modern Philosophy: Cartesianism, Occasionalism, and Preestablished Harmony*, University Park, PA: Pennsylvania State University Press.(Several of the articles discuss early modern occasionalism.)

Ogyū Sorai (1666-1728)

Ogyū Sorai (1666-1728) was one of the greatest, most erudite and most Sinocentric kogaku, *or 'Ancient Learning', philosophers of Tokugawa Japan. Sorai's call for a return to the most ancient philosophical classics of the Chinese tradition, the Six Classics, voiced the logical conclusion of* kogaku *tendencies. However, Sorai's ideas also inspired* kokugaku, *or 'National Learning', a literary movement advocating a return to the ancient writings of Japan which most purely expressed the Japanese soul prior to its distortion by Chinese philosophy.*

Sorai was born in Edo (now Tokyo), the shogun's capital. From 1679-90 Sorai's father, a *samurai*-physician who served the shogunate, was exiled to rural Kazusa (now in Chiba Prefecture) for reasons which are still unclear. In exile, Sorai's father supervised his son's study of the neo-Confucianism of Zhu Xi, a post-Buddhist, metaphysical form of ancient Confucianism (see Neo-Confucian philosophy). After returning to Edo, Sorai gained attention for his thorough grasp of Zhu Xi's teachings. From 1696 to 1709 he served Yanagisawa Yoshiyasu, grand chamberlain to the Shogun Tsunayoshi (1646-1709), as a scholar. Sorai's ideas informed the decision of Tsunayoshi's government in the 1703 'forty-seven *rōnin*' incident. He argued that the *rōnin* were praiseworthy for fulfilling a private morality, that of duty to their lord, but they also had broken public law by assassinating their lord's enemy, one of the shogun's officials. Rather than humiliate them with inflicted punishment, Sorai advised that the *rōnin* be allowed to commit ritual suicide (see Bushi philosophy §3).

After Tsunayoshi's death in 1709, Sorai became a private scholar. His first publication, a Sino-Japanese dictionary entitled *Yakubun sentei (Expedients for Translating Chinese)*, earned him modest fame in 1711. His next work, the *Ken'en zuihitsu (Reed Garden Miscellany)*, published in 1714, attacked the *kogaku* ideas of Itō Jinsai's *Gomō jigi (The Meanings of Terms in the Analects and the Mengzi)* from a neo-Confucian perspective. Sorai's attacks on Itō Jinsai are odd because earlier Sorai had written him expressing admiration for his *Gomō jigi*. Itō Jinsai never replied, leading to speculation that Sorai wrote the *Ken'en zuihitsu* out of resentment (see Itō Jinsai).

By 1717, Sorai had rejected neo-Confucianism in favour of his own brand of *kogaku* (ancient learning) thought. His masterworks, the *Bendō (Discerning the Meaning of the Way)* and *Bemmei (Discerning the Meanings of Philosophical Terms)*, called for a revival of the ancient Chinese philosophy of the Six Classics. They also attacked both Itō Jinsai's revival of the ancient teachings of Confucius and Zhu Xi's neo-Confucianism for indulging in fanciful subjectivism rather than objective study of ancient philosophical language (see Zhu Xi).

Sorai boasted that his philosophy would revive the Way of the ancient sage-kings (see Confucian philosophy, Chinese). That Way embodied the foundations of civilization, including the rites, music, penal laws and institutions of government. In Sorai's view, the sages earned their status because they invented the Way *ex nihilo*. Sorai thus denied the neo-Confucian claim that the Way existed both in the natural world and in human nature. He believed that there remained only one task for humanity, to follow the ancient Way reverently: neither rulers nor philosophers were to tamper with or deviate from it.

Unlike Zhu Xi, who claimed that sagehood was attainable via investigating things, Sorai alleged that it was impossible for men to become sages. While neo-Confucians insisted that human nature was universally good, Sorai denied that there was any moral value inherent in it. Rather, people were just born with various capacities and talents which could be realized by faithfully following the Way of the sages. When everyone perfected their talents, stability would reign.

While Confucius taught that *jin* (*ren* in Chinese), or humaneness, was a virtue that everyone should embody (see Confucian philosophy, Chinese §5), Sorai interpreted *jin* as the virtue of the great sages which enabled them to provide stability for the world through their Way. Religiously, Sorai insisted that people believe without question in ghosts and spirits, because the ancient sages made such belief part of their Way. Neo-Confucians, however, interpreted ghosts and spirits as the activities of *yin* and *yang*, thus defusing many superstitions about them (see Yin-yang).

During his final years, Sorai provided the Shogun Yoshimune (1684-1751) with solutions for the social and economic ills plaguing the state. His *Taiheisaku (Plan for an Age of Great Peace)* and *Seidan (Discourses on Political Economy)* advised that urban-based *samurai* be relocated in rural areas away from the vices of the

pleasure quarters. Sorai thought that the shogunate could thus provide for stability by creating, like the ancient sages, institutions which would rightly order human and natural resources.

See also: Bushi philosophy; Chinese Classics; Confucian philosophy, Chinese; Confucian philosophy, Japanese; Itō Jinsai; Neo-Confucian philosophy; Zhu Xi

JOHN ALLEN TUCKER

List of works

Ogyū Sorai (1666-1728) *Ogyū Sorai zenshū (Complete Works)*, ed. Yoshikawa Kōjirō and Maruyama Masao, Tokyo: Misuzu shobō, 1974, 20 vols.(The definitive edition of Sorai's Complete Works.)

Ogyū Sorai (1714) *Ken'en zuihitsu (Reed Garden Miscellany)*, in Yoshikawa Kōjirō and Maruyama Masao (eds) *Ogyū Sorai zenshū (Complete Works)*, Tokyo: Misuzu shobō, 1974, vol. 5.(The work in which Ogyū attacked the work of Itō Jinsai.)

Ogyū Sorai (1717) *Bendō (Discerning the Meaning of the Way)*, in Yoshikawa Kōjirō *et al.* (eds) *Ogyū Sorai*, Nihon shisō taikei vol. 36, Tokyo: Iwanami shoten; trans. O. Lidin, *Ogyū Sorai: Distinguishing the Way (Bendō)*, Tokyo: Monumenta Nipponica Monographs, 1970.(Lidin's edition is the only English translation of the Bendō.)

Ogyū Sorai (1717) *Bemmei (Discerning the Meanings of Philosophical Terms)*, in Yoshikawa Kōjirō*et al.* (eds) *Ogyū Sorai*, Nihon shisō taikei vol. 36, Tokyo: Iwanami shoten.(Calls for a revival of the ancient Chinese philosophy of the Six Classics.)

Ogyū Sorai (1727) *Taiheisaku (Plan for an Age of Great Peace)*, in Yoshikawa Kōjirō *et al.* (eds) *Ogyū Sorai*, Nihon shisō taikei vol. 36, Tokyo: Iwanami shoten.(Work on political economy.)

Ogyū Sorai (1727) *Seidan (Discourses on Political Economy)*, in Yoshikawa Kōjirō *et al.* (eds) *Ogyū Sorai*, Nihon shisō taikei vol. 36, Tokyo: Iwanami shoten.(Work on political economy.)

Ogyū Sorai (*c.*1717) *Gakusoku (Instructions for Students)*, in Yoshikawa Kōjirō and Maruyama Masao (eds) *Ogyū Sorai zenshū (Complete Works)*, Tokyo: Misuzu shobō, 1974, vol. 1; trans. R. Minear in 'Ogyū Sorai's Instructions of Students: A Translation and Commentary', *Harvard Journal of Asiatic Studies* 36, 1976: 5-81. (The only translation of Sorai's *Gakusoku* in English. Accompanied by a lengthy, thoughtful essay.)

References and further reading

de Bary, W.T. and Bloom, I. (eds) (1979) *Principle and Practicality: Essays in Neo-Confucianism and Practical Learning*, New York: Columbia University Press.(An anthology including sophisticated essays, many pertaining to Sorai, by leading Japanese and American scholars.)

Bellah, R.N. (1978) 'Baigan and Sorai: Continuities and Discontinuities in Eighteenth-Century Japanese Thought', in Tetsuo Najita and I. Scheiner (eds) *Japanese Thought in the Tokugawa Period, 1600-1868*, Chicago, IL: University of Chicago Press.(Reveals how Sorai and Baigan represent antithetical poles in eighteenth-century Japanese thought.)

Bitō Masahide (1978) 'Ogyū Sorai and the Distinguishing Features of Japanese Confucianism', in Tetsuo Najita and I. Scheiner (eds) *Japanese Thought in the Tokugawa Period, 1600-1868*, Chicago, IL: University of Chicago Press.(Discusses the distinctively Japanese elements in Sorai's understanding of the Chinese notion of heaven.)

Imanaka Kanshi (1966) *Soraigaku kisoteki kenkyū (A Basic Study of Sorai's Thought)*, Tokyo: Yoshikawa kōbunkan.(A major, post-Maruyama monograph on Sorai. Traces most of Sorai's ideas to Chinese sources.)

Lidin, O. (1973) *The Life of Ogyū Sorai*, Lund: Studentlitteratur.(The only biography of Sorai in English.)

Maruyama Masao (1974) *Studies in the Intellectual History of Tokugawa Japan*, trans. Mikiso Hane, Princeton, NJ: Princeton University Press.(An English translation of a seminal, though somewhat dated, work written in the early 1940s.)

McEwan, J. (1962) *The Political Writings of Ogyū Sorai*, Cambridge: Cambridge University Press.(Features numerous translations from *Sorai's Seidan* and *Taiheisaku*.)

Sakai Naoki (1992) *Voices of the Past: The Status of Language in Eighteenth-Century Japanese Discourse*, Ithaca, NY: Cornell University Press.(Argues that from the late seventeenth century, Japanese discourse began reconceptualizing the world in new and diverse ways, all of which had political consequences.)

Yamashita, S.H. (1984) 'Nature and Artifice in the Writings of Ogyū Sorai', in P. Nosco (ed.) *Confucianism and Tokugawa Culture*, Princeton, NJ: Princeton University Press.(Criticizes Maruyama's claim that Sorai's philosophy broke with the naturalistic thought of Zhu Xi.)

Yamashita, S.H. (1994) *Master Sorai's Responsals: An Annotated Translation of Sorai sensei tōmonsho*,

Honolulu, HI: University of Hawaii Press.(A brief translation and study of Sorai’s political philosophy.)

Yoshikawa Kōjirō (1983) *Jinsai, Sorai, Norinaga: Three Classical Philologists in Mid-Tokugawa Japan*, Tokyo: Tōhō gakkai.(Includes an English translation of Yoshikawa’s seminal ‘Sorai gakuan’ (An Introductory Study of Sorai).)

Yoshikawa Kōjirō, Maruyama Masao, Nishida Taichirō and Tsuji Tatsuya (eds) (1973) *Ogyū Sorai*, Nihon shisō taikei vol. 36, Tokyo: Iwanami shoten. (Contains extensively annotated Japanese texts, in *kambun* and *bungo*, of Sorai’s main works, the *Bendō, Bemmei*, *Gakusoku*, *Seidan*, *Taiheisaku* and selections from the *Sorai shū*, plus analytic essays by Yoshikawa, Tsuji and Maruyama.)

Oken, Lorenz (1779-1851)

In the early nineteenth century, Oken was one of several German scientists who developed views about the metaphysical presuppositions of science, promoted by Kant and especially by Schelling in order to forge links between their scientific investigations and the prevailing Romantic style of thought. Oken's particular concern was with biology, where he introduced bold taxonomic principles drawing on analogies with mathematical polarities and with our sensory and emotional capacities.

Lorenz Oken, or Okenfuss, was born in Bolsbach, in the Baden region of southern Germany, on 1 August 1779. His academic talents were recognized early, the local pastor and a neighbouring Franciscan Gymnasium providing the foundations of a classical education. Before graduating in 1804 from Freiburg University with a medical degree, he had become an enthusiastic supporter of the metaphysical speculations associated with *Naturphilosophie*, a movement endorsed by the early German Romantic thinkers which sought to uncover the general presuppositions of science, and had published an outline of his intended contributions (see Naturphilosophie). In 1807 he was called on Goethe's recommendation to a chair of medicine at Jena University. His *Lehrbuch des Systems der Naturphilosophie (Textbook of Systematic Philosophy of Nature)* was published shortly after his arrival. He remained at Jena, becoming Professor of Philosophy in 1812, but was dismissed in 1819 following his allegedly subversive political activities as editor of *Isis*, a periodical he founded in 1817 with the aim of providing comprehensive information about scientific and other matters to the general reader. The periodical was largely written byOken himself and it rapidly became widely read, partly for the outspoken liberal politics it promoted. He subsequently held teaching posts in medicine and physiology at several universities, and was instrumental in founding the influential *Gesellschaft Deutscher Naturforscher und Ärzte* (German Society for Scientists and Doctors) in 1822. He ended his career as Rector of Zürich University and died in Zürich on 11 August 1851.

Like other post-Kantian thinkers, Oken believed that the structures in terms of which we understand the natural world are supplied, not by that world, but by the mind. Kant himself had initiated investigations of the metaphysical foundations of science which were intended to reveal these structures, and Oken's contemporaries, such as Hegel and Schelling, responded with their own versions of a philosophy of nature (see Hegel, G.W. §7; Schelling, F.W. §1). In Oken's writings we find little of the logical analysis that these philosophers employed in describing these mental structures; instead, his abstract and generalized language conveyed in a philosophical idiom the Romantic conviction that imagination and feeling should play their part in scientific understanding. Thus, he took the applicability of mathematics to nature as a foundation for ideas about the spiritual and material, the eternal and the temporal, which also found expression in contemporary German Romantic poetry and art. For if nature is constructed by the self, though in opposition to the self, then as a whole it corresponds to mathematical infinity, because the self constructs it as unlimited in space, time and scope. And the origin of nature (conceptually, if not historically) corresponds to zero because the self constructs nature as originating from nothing, albeit with infinite potential. These constructs of an infinite nature and a primeval nothing are in polar opposition to each other, and the interaction between them generates our constructions and understandings of nature, just as the interaction between our conceptions of zero and infinity generates our constructions and understandings of natural numbers. In common with philosophers writing in the Romantic idiom of irony, contradiction and unrestrained feeling, Oken subjected language to enormous pressure. The resulting syntactical intensity mystified his readers but it was integral to the fervour of his thought, including his thought about nature.

In the physical sciences, his thinking was close to that of Schelling. Indeed, the two men were in contact during the first decade of the nineteenth century, when Oken was developing his philosophy of nature. Both understood optical, electrical and magnetic phenomena in terms of polarities of light and darkness, attraction and repulsion, these polarities being manifestations of an abstract dialectic of productivity and product. Oken's tone when writing of these matters was one of enthusiastic endorsement rather than critical consideration, and his aphoristic style encouraged him to affirm rather than defend his conclusions. He did not, therefore, provide the elucidation of the difficulties of this philosophy that his contemporaries might have sought or that modern readers might require.

The biological and medical sciences were, he found, fruitful fields in which to apply the concepts and analogies

suggested by his philosophy. They were, too, the sciences in which his expertise was recognized, for he was renowned as a meticulous descriptive anatomist. But the conventionality of his practical science stands in sharp contrast to the boldness of his theoretical speculations. So, for example, in taxonomy he used the idea of a single series of increasingly complex organisms, beginning with primal organisms, or *'Ur-thiere'*, and ending with human beings. Just as the comparable sequence of natural numbers displays patterns which are repeated within segments of the sequence, so Oken sought repeating patterns within his taxonomic series. In his earliest writings he had used the language of sensory and emotional capacities, often metaphorically, to describe these patterns, and he found imaginative ways of providing evidence of their applicability. However, if such patterns were not produced by perceptions of nature but by conceptions of our minds, it is not surprising that he was sometimes casual in confirming the details they implied.

Oken has a place in the history of biology and medicine for his experimental contributions to comparative anatomy, and for some of his taxonomic ideas which are said to anticipate certain modern views. His aim of bringing a Romantic style of thinking into his science led him to explore *Naturphilosophie*, and, in the history of philosophy, his significance is bound up with that of the distinctive coupling of speculative science with idealistic metaphysics. Certainly, Oken's philosophy served a need among his contemporaries, but it was his pedagogic and investigative skills, rather than his philosophical acumen, that earned their admiration.

See also: Romanticism, German

BARRY GOWER

List of works

Oken, L. (1804) *Uebersicht des Grundriss des Systems der Naturphilosophie (Survey of a Plan for Systematic Philosophy of Nature)*, Frankfurt am Main: P.W. Eicheberg. (A brief outline of Oken's characteristic approach to philosophy of nature.)

Oken, L. (1805) *Die Zeugung (Reproduction)*, Bamberg and Wirzburg: J.A. Goebhardt. (A speculative physiological essay on the generation of organisms from primordial 'infusoria'.)

Oken, L. (1808) *Erste Ideen zur Theorie des Lichts, der Finsterniss, der Farben and der Wärme (Preliminary Ideas Towards a Theory of Light, Darkness, Colour and Heat)*, Jena: Friedrich Frommann.(Explores the anti-Newtonian implications of Oken's philosophy of nature for a rapidly changing part of physics.)

Oken, L. (1809-11) *Lehrbuch des Systems der Naturphilosophie (Textbook of Systematic Philosophy of Nature)*, Jena: Friedrich Frommann, 3 vols; 2nd edn, 1831; 3rd edn, Zürich, F. Schulthess, 1843; trans. A. Tulk, *Elements of Physiophilosophy*, London, Ray Society, 1847.(A widely-read exposition ofOken's metaphysical philosophy of nature.)

Oken, L. (1813-26) *Lehrbuch der Naturgeschichte (Textbook of Natural History)*, Leipzig and Jena: C.H. Reclam, 5 vols.(A treatise encompassing experimental findings in zoology, biology and mineralogy within Oken's philosophy of nature.)

Oken, L. (1833-41) *Allgemeine Naturgeschichte für alle Stände (General Natural History for all Circumstances)*, Stuttgart: Hoffman, 13 vols.(Oken's final and most comprehensive study of natural history.)

References and further reading

Ecker, A. (1880) *Lorenz Oken. Eine biographische Skizze*, Stuttgart; trans. A. Tulk, *Lorenz Oken: A Biographical Sketch,* London: Kegan Paul, Trench & Co., 1883.(A flattering portrait by the son of one of Oken's colleagues; contains only meagre information aboutOken's philosophy.)

Gould, S.J. (1984) 'The Rule of Five', *Natural History* 93 (10): 14-23.(Explains Oken's numerical system for explaining natural order, comparing it with other pre-Darwinian systems.)

Mullen, P.C. (1977) 'The Romantic as Scientist: Lorenz Oken', *Studies in Romanticism* 16: 381-99.(A recent account of Oken's life with some details about his philosophical ideas.)

Olivecrona, Karl (1897-1980)

Olivecrona was a Swedish jurist of the 'realist' school, and from 1933 Professor of Procedural Law in the University of Lund. He regards law as a body of 'independent imperatives' effective in bringing about certain patterns of behaviour in society. The reality of law is psychological and behavioural, not dependent on some metaphysical world of norms.

Olivecrona's theory is non-voluntarist: there is no individual or group 'lawmaker' whose will constitutes law. Legal rules are not strictly imperatives. Yet in setting patterns of behaviour they have an expressive and emotive function and are regarded by the addressees as if they were imperatives.

Many independent imperatives have a performative form: they purport to create ideal power-relations, rights, duties and the like, and only indirectly to command action or abstention in a physical sense. Such performatives prompt beliefs in valid law and rights, but these beliefs do not (cannot) refer to any fact. Yet they have social functions. They direct human conduct, facilitate concise law-making, and even convey some vague information, for example, that the 'owner' of an estate in the usual course of things has a kind of control over it. The most important function of criminal law is to promote the conviction that crime leads to punishment (see Legal concepts §4).

Olivecrona in his later work pursued the history of ideas about rights and related concepts. In this, he contributed to understanding the crucial role of the idea of *suum* in the natural law theories of, among others, Grotius and Pufendorf. This idea is at the heart of ideas both of property and of personal rights to non-interference by others.

Olivecrona was a disciple of Axel Hägerström (§4), according to whom all knowledge concerns reality in time and space and the concept of value is inconsistent with reality; hence 'binding' law is an illusion. He has been grouped with Alf Ross and Vilhelm Lundstedt as a 'Scandinavian realist' (see Legal realism §1), their 'realism' descending from Hägerström's anti-idealist materialism. Their work wrestles with significant problems concerning normative concepts such as 'right' and 'duty'. If these do not have material counterparts, but function in essentially action-guiding ways, what theory of language and of reality properly captures their sense? If philosophy peels off the air of magic and mystery surrounding them, can they continue to guide action effectively?

Olivecrona's work deserves credit for its clarity in posing the fundamental problems of legal philosophy. He gave a precise and comprehensive reconstruction of the conceptual apparatus of law consistent with his anti-metaphysical framework (see Legal positivism §5). Yet such a reconstructionism can be criticized as too remote from the internal point of view of lawyers and inadequate to capture legal reasoning and legal thought in the context of justification (see Legal reasoning and interpretation). If juristic opinions are merely emotive and expressive, how can they be discussed intersubjectively? How can lawyers honestly fulfil their beneficial social functions (rightly recognized by Olivecrona) if they regard the 'valid law' as mere illusion?

See also: Law, philosophy of

ALEKSANDER PECZENIK

List of works

Olivecrona, K. (1939) *Law as Fact*, Copenhagen: Einar Munksgaard.(Basic 'realist' statement, flatly denying the existence of any 'world of norms'.)

Olivecrona, K. (1951) 'Realism and Idealism: Some Reflections on the Cardinal Point in Legal Philosophy', *New York University Law Review* 26: 120-31.

Olivecrona, K. (1971) *Law as Fact*, London: Stevens, 2nd edn.(Totally revised work, drawing much more on modern linguistic-analytical theory, especially that of J.L. Austin.)

References and further reading

Arnholm, C.J. (1962) 'Olivecrona on Legal Rights: Reflections on the Concept of Rights', *Scandinavian Studies in Law* 6: 11-33.(Explanation and critique of Olivecrona's views, drawing attention to the difference between his views and those of US realists.)

Bjarup, J. (1978) *Skandinavischer Realismus (Scandinavian Realism)*, Freiburg/Munich: Alber.(Critical

introduction to the works of Hägerström, Lundstedt, Olivecrona and Ross.)
Lewis, J.U. (1970) ‘Karl Olivecrona: "Factual Realism" and the Reasons for Obeying a Law’, *University of British Columbia Law Review* 5: 281-305.
MacCormack, G.D. (1970) ‘Scandinavian Realism’, *Juridical Review* 82: 33-54. (Interesting critique of the psychologism to be found in Olivecrona and in Alf Ross.)
Ruiter, D.W.P. (1993) *Institutional Legal Facts*, Dordrecht: Kluwer.(Chapter 1, on ‘Legal Validity’, criticizes and presents an alternative to Olivecrona’s performative-based account of validity.)

Olivi, Peter John (1247/8-98)

Condemned repeatedly by religious authorities, Peter John Olivi is one of scholasticism's most original and colourful figures. Although better known for his involvement in social and political debates within the Franciscan order, Olivi also took up the leading epistemological and metaphysical concerns of his day. His outright scorn for Aristotle and cautious rapport with Augustine combine to produce an exciting, insightful body of philosophical work.

Olivi studied in Paris without becoming a master of theology and spent much of his adult life teaching at various Franciscan houses in southern France. Although he produced a wide-ranging body of work, most of his philosophical views are contained in his question-commentary on the second book of Peter Lombard's *Sentences* (see Lombard, P.) This lengthy treatise takes up a wide range of philosophical problems such as the nature of matter, form, substance, quantity, the soul, causation, motion and the problems of universals and individuation. Although imbued with the Aristotelian terminology and perspectives of his contemporaries, Olivi feels free to reject this influence when it suits him. He characteristically remarks of Aristotle, 'his authority and that of any infidel and idolator is nothing to me - especially as regards matters that belong to the Christian faith or are very near to it' (*Quaestiones in secundum librum Sententiarum* q.16 ad 6).

Many of Olivi's most interesting and influential arguments concern human volition and cognition. His views on these topics are often strikingly similar to those of his Franciscan successors John Duns Scotus and William of Ockham. Olivi opposes any account from which it would follow that our will is constrained in its choices or determined to select one alternative over another. By his standards, human freedom requires that persons be able, at the same time at which they choose one thing, to choose instead its opposite; in other words that 'someone does one thing in such a way that at the same time he could have done the other while ceasing from the first - and vice versa' (*Quaestiones* q.57 ad 4). Olivi thinks that if free will is to be preserved, this counterfactual must hold not just until the moment of decision, but also at the moment of decision itself. To explain this difficult claim - that at the same time at which I did one thing I could have done the other - Olivi invokes the notion of natural priority (*Quaestiones* q.57 ad 10). Even though it seems as if I could not choose to cross the street at the very moment that I choose not to cross the street, I have an ability to choose otherwise which is naturally prior to but temporally simultaneous with the choice that I do make.

Olivi needs to make these subtle conceptual distinctions in order to reject any passivity or determinism on the part of will. In particular he wants to deny that the will is determined to choose in accordance with intellect's counsel. Intellect's role in decision-making is to present to the will various alternatives, presenting them as good in various respects. The will's own freedom, he holds, is a completely sufficient cause for its choosing one of those alternatives over another. Hence 'if one is asked why [a volition] ceases…it is fully sufficient to reply by saying that it was completely and sufficiently able to cease' (*Quaestiones* q.57 ad 5). Olivi claims that demanding any further explanation leads to the destruction of free will.

It is a peculiar feature of Olivi's thought that he sees a close connection between taking the will to be passive (and hence not free) and taking the cognitive powers to be passive. He says that it is the doctrine of the passivity of cognition that 'above all things moves many to believe that our will is totally passive' (*Quaestiones* q.58 ad 14). Olivi doesn't explain why anyone would link these two issues, but his concern on this point leads him to append to his discussion of will an extended attack on the standard medieval Aristotelian theory of human cognition. His central claim is that the human cognitive powers (intellect and senses) are not passive. To maintain this claim, Olivi reverses the usual causal order. We do not cognize in virtue of *receiving* external impressions; rather, our cognitive powers somehow reach out directly to their objects in the world. While Olivi rejects the ancient account in which vision takes place through a physical extromission, he allows what he calls a 'virtual extension' of the cognitive powers to the external world (*Quaestiones* q.26).

Although Olivi discusses the notion of virtual extension at length, it is never quite clear what he has in mind. However, the motivation for this position is clear enough. In part, he wants to reject any sort of passivity on the part of our cognitive powers. Further impetus comes from his claim that 'no body can directly move the soul's sensory powers, and much less the intellectual ones' (*Quodlibeta* I.4). In place of a direct causal connection, Olivi

postulates an obscurely explained ‘natural link’ (*colligantia naturalis*) between mind and body (*Quaestiones* q.111). In the standard scholastic account, on the other hand, external objects produce forms (known as sensible and intelligible species) in our cognitive faculties. Olivi denies not only that these species are produced by external objects, but also that such species occur at all: there is no internal representation of objects in the external world beyond the act of cognition itself. This view became common in the fourteenth century (in the works of Ockham, for example), but Olivi seems to have been the first to hold it. His most interesting arguments are epistemological. If there were such species, he says, we ‘would be diverted from seeing the object rather than led toward seeing it’ (*Quaestiones* q.74). On this basis he raises sceptical problems concerning our knowledge of the external world. His theory of virtual extension, in contrast, allows a direct realist account of cognition.

Olivi’s rejection of sensible and intelligible species was condemned by the Franciscan authorities, along with many of his other views. It seems clear that these philosophical claims drew ecclesiastical condemnation primarily because of his involvement in the controversial Franciscan spiritual movement. In any case, all of Olivi’s writings were twice condemned and ordered to be burned. For this reason, at least, his work received less philosophical attention than it deserves. Enough of his writings have survived, however, so that this neglect need not persist.

See also: Augustinianism; Gerard of Odo; William of Ockham

ROBERT PASNAU

List of works

There are no published translations of Olivi’s philosophical works.

Olivi, Peter John (*c.*1282) *Quaestiones in secundum librum Sententiarum (Questions on the Second Book of the Sentences)*, Bibliotheca Franciscana Scholastica 4-6, ed. B. Jansen, Quaracchi: Collegii S. Bonaventurae, 1922-6, 3 vols.(This edition contains Olivi’s most important philosophical writings. it should be distinguished from his unedited *Sentences* commentary (1287-8).)

Olivi, Peter John (*c.*1285) *Quaestiones logicales (Logical Questions)*, ed. S. Brown, *Traditio* 42, 1986: 335-88.(A fairly technical set of questions on reference and the relation between species and genus.)

Olivi, Peter John (*c.*1285) *Tractatus de verbo (Treatise on the Mental Word)*, ed. R. Pasnau in *Essays in Honor of Fr. Gedeon Gál*, St Bonaventure, NY: Franciscan Institute Press, 1997.(Edits the philosophically interesting beginning of Olivi’s commentary on the Gospel of St John.)

Olivi, Peter John (*c.*1287?) *Quodlibeta*, Venice, 1509.(There are no surviving manuscripts of these quodlibetal questions, making this early edition invaluable. Unfortunately, it has never been reprinted, and existing copies are very rare.)

References and further reading

Bérubé, C. (1976) ‘Olivi, critique de Bonaventure et d’Henri de Gand’ (Olivi’s Critique of Bonaventure and Henry of Ghent), in *Studies Honoring Ignatius Charles Brady*, St Bonaventure, NY: Franciscan Institute. (Discusses Olivi’s interesting and ambivalent treatment of the Augustinian theory of divine illumination.)

Bettoni, E. (1959) *Le dottrine filosofiche di Pier di Giovanni Olivi (The Philosophical Doctrine of Peter John Olivi)*, Milan: Società Editrice ‘Vita e Pensiero’.(A general survey of Olivi’s philosophy.)

Burr, D. (1971) ‘Petrus Ioannis Olivi and the Philosophers’, *Franciscan Studies* 31: 41-71.(Studies Olivi’s antipathy to non-Christian philosophy.)

Burr, D. (1976) ‘The Persecution of Peter Olivi’, *Transactions of the American Philosophical Society* 66: 3-98. (The most detailed study of Olivi’s life.)

Burr, D. (1989) *Olivi and Franciscan Poverty: The Origins of the Usus Pauper Controversy*, Philadelphia, PA: University of Pennsylvania Press.(Up-to-date biographical details, and a thorough discussion of Olivi’s role in controversies within the Franciscan order.)

Dumont, S. (1995) ‘The Origin of Scotus’s Theory of Synchronic Dependency’, *Modern Schoolman* 72: 149-67. (Establishes that Scotus’ libertarian theory of will has its origins in Olivi. Comments by S. MacDonald question the significance of the Olivi/Scotus position.)

Gieben, S. (1968) ‘Bibliographia Oliviana (1885-1967)’, *Collectanea Franciscana* 38: 167-95.(A very thorough guide.)

Jansen, B. (1921) *Die Erkenntnislehre Olivis (Olivi’s Epistemology)*, Berlin: Dümmlers.(The most complete discussion of Olivi’s epistemology and philosophy of mind.)

Pacetti, D. (1954) ‘Introduzione’, in *Quaestiones quatuor de domina*, Quaracchi: Collegii S. Bonaventurae.

(Catalogues Olivi’s work, giving its printed and manuscript sources.)
Pasnau, R. (1997) *Theories of Cognition in the Later Middle Ages*, Cambridge: Cambridge University Press. (Discusses at length Olivi’s novel views on mental representation and cognitive activity.)
Putallaz, F.-X. (1991) *La connaissance de soi au XIII*[e] *siècle. De Matthieu d’Aquasparta à Thierry de Freiberg (Self-knowledge in the Thirteenth Century: Matthew of Aquasparta and Dietrich of Freiberg)*, Paris: Vrin.(The chapter on Olivi ranges widely over his theories about the soul and knowledge.)
Schneider, T. (1973) *Die Einheit des Menschen (The Unity of Human Beings)*, Beiträge zur Geschicte der Philosophie und Theologie des Mittelalters, NF 8, Münster: Aschendorff.(A study of events leading up to the Council of Vienna in 1312, which declared it heretical to deny that the intellective soul, *per se* and essentially, is the form of the body. Schneider devotes a chapter to Olivi’s views, which were at the centre of the controversy.)
Simoncioli, F. (1956) *Il problema della libertà umana in Pietro di Giovanni Olivi e Pietro de Trabibus (The Problem of Human Freedom in Peter John Olivi and Peter of Trabes)*, Milan: Società Editrice ‘Vita e Pensiero’.(Contains some useful texts as well as a discussion of Olivi on free will.)

Oman, John Wood (1860-1939)

A central theme of John Wood Oman's writings is the possibility and actuality of knowledge that is not gained through science. He rejects as too simplistic the mechanistic view of the world. His belief in God rests not on the arguments of natural theology, but on the force and content of religious experience. The source of religion is to be found in our sense of the supernatural, from which stems also our moral dependence on God.

Born in the Orkney Islands in 1860, John Wood Oman was educated at the universities of Edinburgh and Heidelberg and served as a pastor in northern England for seventeen years. He then taught for a further twenty-eight years at the seminary of the English Presbyterian Church at Westminster College, Cambridge. He translated Friedrich Schleiermacher's *Lectures to Cultured Despisers of Religion* into English, and his emphasis on the centrality of religious experience was influenced by Schleiermacher's theology (see Schleiermacher, F. §7). He died in 1939.

Oman held religion to be an actual experience of an actual environment. It involves awareness of the holy, a direct sense or feeling of the supernatural. It also involves awareness of the sacred and a valuation of the supernatural as possessing absolute worth (see Religious experience §1). The distinction between natural and supernatural is coextensive with the distinction between relative value and absolute value. None the less, Oman held, religion's core is not this sense or its accompanying valuation, but their object; its emphasis is not on experiences or feelings, but on what elicits them.

In *The Natural and the Supernatural* (1931), Oman finds the source of religion in a sense of the supernatural. The supernatural, he says, is a kind of environment with its own sanctions. Experience of this environment provides humans with a degree of transcendence of their natural environment. Contact with any environment includes perception of meaning, and perception of meaning typically involves interpretation as well as recognition. Hence there are differences in how the actual environment is conceived, and consequent differences in how it is experienced.

In primitive religion, the supernatural is conceived animistically. In polytheism, it is thought of as distinct divinities each ruling over a part of nature. In cosmic pantheism, it is viewed as an organic whole with integrally related parts. In idealistic pantheism, it is thought of as a spiritual whole without matter. Thus contacting animistic spirits, placating nature's rulers, finding unity with the cosmos, and escaping illusion, are seen as means of salvation in these religions. Different notions of salvation correlate with diverse conceptions of the supernatural.

In monotheism, the supernatural is conceived as a personal God. Priestly, legalistic religions divide the world into the secular and the sacred; prophetic religions find the supernatural through the natural. Those who follow priestly religions seek the supernatural through ceremonies, and those who embrace prophetic religions see their life as worship and their duties as God's commands. For priestly religions, daily awareness of one's mundane life contains a sense of abiding purpose, and the sense of the supernatural includes an awareness of something of unconditional worth that lays an absolute claim upon our devotion and obedience. Refusal to worship and obey divides one from the supernatural environment in which one's true hope lies.

For all of its diversity, Oman contends, religion has two main branches: 'pantheism and absorption into the One…monotheism and victory over the many' (1931: 407). The former focuses on escaping from the world, the latter endeavours to create a better world. Oman views prophetic monotheism, as manifested in the Old Testament prophets and in the teachings of Jesus, as the highest religion.

In *Grace and Personality* (1919), Oman discusses an issue fundamental to prophetic religion:

> The essential quality of a moral person is moral independence and an ideal person would be of absolute moral independence. But the essential quality of a religious person is to depend on God; and he must be as absolutely dependent as a moral person must be absolutely independent.
>
> (1919: 58)

For prophetic religions, every good is given by divine courtesy and none is due to human effort. From the moral point of view, nothing that is morally good can be deterministically produced. Oman suggests that grace involves

God's giving us sufficient insight to see that what God wills is coextensive with what realizes our humanity; grace thus is not coercive but persuasive. Morality is not an arbitrary law, externally imposed, but a pattern for realizing our nature as persons.

Oman emphasized the importance of religious and moral experience in theology and philosophy of religion while also insisting on consistency and coherence. He was interested in religious traditions other than his own, and willing to learn from them. He did not suppose that this required that he reject his own tradition, think all experience equal, embrace relativism, or abandon efforts at rational assessments of religious doctrines.

KEITH E. YANDELL

List of works

Oman, J.W. (1902) *Vision and Authority*, London: Hodder & Stoughton.(An inquiry into the nature and object of Churchly authority.)

Oman, J.W. (1906) *The Problem of Faith and Freedom in the Last Two Centuries*, New York: A.J. Armstrong & Son.(Historical and critical discussion of faith and freedom in Anglo-European history over two centuries.)

Oman, J.W. (1919) *Grace and Personality*, Cambridge: Cambridge University Press.(A discussion of human freedom and responsibility, and divine sovereignty and grace.)

Oman, J.W. (1931) *The Natural and the Supernatural*, Cambridge: Cambridge University Press.(Offers an account of the relations between the natural and the supernatural with reference to the history of religions.)

References and further reading

Bevons, S.B. (1992) *John Oman and his Doctrine of God*, Cambridge: Cambridge University Press.(Discussion of Oman's particular monotheistic emphases.)

Elliot-Binns, J.E. (1956) *English Thought 1860-1900: The Theological Aspect*, London: Longmans, Green & Company.(Discusses the historical, philosophical, theological and social setting of English theology in the indicated period.)

Healey, F.G. (1965) *Religion and Reality: The Theology of John Oman*, Edinburgh: Oliver & Boyd.(General discussion of Oman's theological perspective.)

Horton, W.M. *Contemporary English Theology*, London: SCM Press.(Treats pre- and post-First-World-War English theology from a perceptive American perspective.)

Langford, T. (1969) *In Search of Foundations: English Theology 1900-1920*, Nashville, TN: Abington Press. (Presentation of English theological thought in its cultural and philosophical context; good bibliography.)

Mozley, J.K. (1951) *Some Tendencies in British Theology*, London: SPCK. (Discussion of British theology from *Lux Mundi* (1889) to the thought of Oman.)

Omnipotence

Traditional theism understands God to be the greatest being possible. According to the traditional conception, God possesses certain great-making properties or perfections, including necessary existence, omniscience, perfect goodness, and omnipotence. Philosophical reflection upon the notion of omnipotence raises many puzzles and apparent paradoxes. Could an omnipotent agent create a stone so massive that that agent could not move it? It might seem that however this question is answered, it turns out that, paradoxically, an omnipotent agent is not truly all-powerful. Could such an agent have the power to create or overturn necessary truths of logic and mathematics? Could an agent of this kind bring about or alter the past? Is the notion of an omnipotent agent other than God an intelligible one? Could two omnipotent agents exist at the same time? If there are states of affairs which an omnipotent agent is powerless to bring about, then how is the notion of omnipotence to be intelligibly defined? Yet if the notion of omnipotence is unintelligible, then traditional theism must be false. Another obstacle to traditional theism arises if it is impossible for God to be both perfectly good, and omnipotent. If an omnipotent God is powerless to do evil, then how can God be omnipotent?

1 What omnipotence means

The intelligibility of the divine attributes, considered individually and in combination, is a central concern among philosophers of religion. Among the attributes of the God of traditional theism is omnipotence. Some philosophers have tried to understand omnipotence in terms of the power to perform some suitably comprehensive set of tasks. Examples of tasks are to create a stone too massive for oneself to lift, to do evil, or to make oneself non-omnipotent. Experience has proved, however, that such an approach to the analysis of omnipotence is fruitless. More promising is the effort of some philosophers (such as Rosenkrantz and Hoffman 1980a, Flint and Freddoso 1983, and Wierenga 1989) to analyse omnipotence in terms of the power to bring about a suitably defined set of states of affairs (where a state of affairs is a propositional entity, which either obtains or fails to obtain).

Two senses of 'omnipotence' come to mind. First, there is the notion of having it within one's power to bring about any state of affairs whatsoever, including impossible and necessary states of affairs. Descartes, in the *Meditations* (1641), appears to have had a notion close to this one. But, as Aquinas in his *Summa theologiae* (1266-73) and Maimonides in his *Guide for the Perplexed* (*c*.1190) recognized, it is not possible for an agent to bring about an impossible state of affairs (for example, there being a colourless red object), since if it were, it would be possible for an impossible state of affairs to obtain, which is a manifest contradiction. Nor is it possible for an agent to bring about a necessary state of affairs (for example, any red object's being coloured). If it is possible for an agent, *a*, to bring about a necessary state of affairs, *s*, then possibly, (1) *a* brings about *s*, and (2) if *a* had not acted, then *s* would not have obtained. But a necessary state of affairs obtains whether or not anyone acts, so (2) is false. Therefore, it is not possible for an agent to bring about a necessary state of affairs. The first sense of omnipotence is consequently unintelligible.

A second sense of 'omnipotence' is that of maximal power. In this sense, the overall power of an omnipotent being could not be exceeded by any being. It does not follow that a being with maximal power can bring about every state of affairs, since doing so is impossible. Nor does it follow that a being with maximal power can bring about everything that any other being can bring about. From the fact that *a* can bring about *s*, and *b* cannot, it does not follow that *a* is overall more powerful than *b*, for it might be that *b* can bring about some thing or things that *a* cannot. This is the only sense of omnipotence that may be coherent. Thus, if a putative analysis of the concept of omnipotence is offered in defence of its intelligibility, then it should be an attempt to analyse the concept of maximal power.

Power is not the same thing as ability. Power is ability plus opportunity. A being who had maximal ability but was prevented by circumstances from exercising it would not be omnipotent. Necessarily, nothing prevents an omnipotent being from exercising its powers when it wills to do so.

Some have thought that there could be two (or more) omnipotent agents, a_1 and a_2, at the same time, t. If this were possible, then it could happen that at t, a_1 strives to move a button (while remaining omnipotent), and at t, a_2 strives to keep that button still (while remaining omnipotent). In a case of this kind, the button would be unaffected

(as to its motion or rest) by either one of them. Thus, in such a case, a_1 would at *t* lack the power to move the button, and a_2 would at *t* lack the power to keep the button still. Since it is absurd to suppose that an omnipotent being lacks the power to move a button or the power to keep it still, neither a_1 nor a_2 is omnipotent. Hence, there could not be two omnipotent beings at the same time.

Could there be a contingently omnipotent being? This seems possible on first thought, but the following argument seems to show the opposite. A being cannot have any perfection unless it has every perfection. This, together with the assumptions that a contingently omnipotent being has a perfection, and that an essentially omnipotent being has a perfection (namely essential omnipotence) which a contingently omnipotent being lacks, implies that contingent omnipotence is impossible.

In addition, assuming that traditional theism is true, God has necessary existence and is essentially omnipotent. Since there could not be more than one omnipotent being, it follows that there could not be a contingently omnipotent being (see Necessary being).

However, the first argument against the possibility of contingent omnipotence is incorrect. For example, a perfect triangle need not have the perfections of omniscience or omnipotence. Thus, it is false that a being cannot have a perfection unless it has every perfection.

The second argument against the possibility of contingent omnipotence presupposes traditional theism. But in analysing omnipotence, it is preferable to remain neutral about whether theism is true or false. In that case, omnipotence will not be assumed to be attributable only to the God of traditional theism or only to an essentially omnipotent being.

2 The paradox of the stone

One challenge to the coherence of the notion of omnipotence is based on the so-called paradox or riddle of the stone. Can an omnipotent agent, *a*, bring it about that there is a stone of some mass, *m*, which *a* cannot move? If the answer is ‘yes’, then there is a state of affairs that *a* cannot bring about, namely (i) that a stone of mass *m* moves. On the other hand, if the answer is ‘no’, then there is a different state of affairs that *a* cannot bring about, namely (ii) that there is a stone of mass *m*, which *a* cannot move. Thus, it might appear that omnipotence is paradoxical. But the appearance of paradox can be overcome if either (i) or (ii) is impossible, for, as has been seen, an omnipotent agent need not be able to bring about an impossible state of affairs. If, on the other hand, both (i) and (ii) are possible, then it is possible for some omnipotent agent to bring it about that both (i) and (ii) obtain, though at different times, and so there is a second solution to the paradox.

The first solution comes into play when there is an *essentially* omnipotent agent, a_1. The state of affairs of a_1’s being not omnipotent is impossible. Therefore, a_1 cannot bring it about that a_1 is not omnipotent. Since, necessarily, an omnipotent agent can move any stone, no matter how massive, state of affairs (ii) is impossible.

The second solution presupposes the only other alternative: that there is a *contingently* omnipotent agent, a_2. In this case, a_2’s being non-omnipotent is a possible state of affairs; and presumably, it *is* possible for a_2 to bring it about that a_2 is non-omnipotent. So a_2 can create and move a stone of mass *m* when omnipotent, and *then* bring it about that a_2 is not omnipotent and is unable to move that stone of mass *m*. Thus, a_2 is able to bring about both (i) and (ii), but only if they obtain at different times.

3 Things that an omnipotent agent is powerless to do

At this point, it might be thought that an analysis of omnipotence could be simply stated as the power to bring it about that any contingent state of affairs obtains. However, the following contingent states of affairs show that this simple analysis fails:

(a) A stone rolled.
(b) A stone rolls at *t* (where *t* is a past time).
(c) Socrates walks for the first time.
(d) The Nile floods an odd number of times less than four.

(e) A ball rolls and no omnipotent agent ever exists.
(f) Jones freely decides to walk.

(a) is a past state of affairs. Since it is impossible for any agent to have power over the past (this is the so-called ‘necessity of the past’), no agent, not even an omnipotent one, can bring about (a). Similarly, although (b) can be brought about prior to *t*, the necessity of the past implies that even an omnipotent agent cannot bring about (b) after *t*. In the case of (c), prior to Socrates’ first walk, an omnipotent agent can bring about (c). But once Socrates has walked, even an omnipotent agent cannot bring about (c). With respect to (d), before the Nile’s third flooding, an omnipotent agent can bring about (d), whereas after the Nile’s third flooding, even an omnipotent agent is powerless to bring about (d). (e) presents a special problem. While it is clear that (e) could not be brought about by an omnipotent agent, some philosophers (such as Hoffman and Rosenkrantz 1988) argue that it is possible for a non-omnipotent agent to bring about (e) by rolling a ball when, in fact, no omnipotent agent ever exists. (Recall that a maximally powerful being need not be able to bring about every state of affairs that any other being can.) Finally, an omnipotent agent other than Jones cannot bring about (f) if the libertarian theory of free will is correct, but apparently a non-omnipotent agent - namely Jones - can bring about (f).

Thus, an adequate analysis of omnipotence should not require an omnipotent agent to have the power to bring about (a), (b), (c), (d), (e) or (f), assuming in the case of (f) that libertarianism is correct. Given the seemingly wide disparity among contingent states of affairs (a)-(f), it might be conjectured that there is no analysis of omnipotence that can deal adequately with all of them while capturing the intuitive scope of power involved in omnipotence. Yet the following observations set the stage for just such an analysis.

For an adequate analysis of omnipotence, identifying some distinctive features of (a)-(f) might enable one to state an analysis of this kind. First, as (a) is not possibly brought about by any agent, an omnipotent agent should not be required to have the power to bring about a state of affairs unless it is possible that some agent brings about that state of affairs. Second, although (b) and (c) are possibly brought about by some agent, they are not repeatable - in other words, it is not possible that they obtain, then fail to obtain, and then obtain once more. (Note that the conditions that do not require an omnipotent agent to be able to bring about (a) and (b) also do not require that agent to be able to bring about impossible or necessary states of affairs.) Third, while (d) is repeatable, it is not unrestrictedly repeatable - it cannot obtain, then fail to obtain, then obtain again, and so forth throughout all of time. Fourth, even if (e) is unrestrictedly repeatable, it is a complex state of affairs, namely a conjunctive state of affairs, whose second conjunct is not repeatable. Reflection on the nature of repeatability and its relation to the power to bring about states of affairs leads to the conclusion that an omnipotent agent ought not to be required to have the power to bring about either a state of affairs which is not unrestrictedly repeatable or a conjunctive state of affairs one of whose conjuncts is not unrestrictedly repeatable. Finally, although (f) is unrestrictedly repeatable, (f) is identifiable with or analysable as a conjunctive state of affairs having three conjuncts, whose second conjunct is not possibly brought about by anyone. The conjunctive state of affairs in question can be informally expressed as follows: (I) Jones decides to walk, and (II) there is no *antecedent* sufficient causal condition of Jones’ deciding to walk, and (III) there is no concurrent sufficient causal condition of Jones’ deciding to walk. Because, as we have explained, it is impossible for an agent to have power over the past, the state of affairs which is the second conjunct of this conjunctive state of affairs is not possibly brought about by anyone. Hence, an omnipotent being should not be required to have the power to bring about a state of affairs which is identifiable with or analysable as a conjunctive state of affairs one of whose conjuncts is not possibly brought about by anyone.

4 An analysis of omnipotence

Based on insights gleaned from the preceding discussion, omnipotence can be analysed in three steps as follows (where ‘ $=_{\mathrm{df.}}$ ’ means ‘is defined as’):

Step 1. The period of time *t* is a sufficient interval for $s =_{\mathrm{df.}} s$ is a state of affairs such that: it is possible that *s* obtains at a time-period which has the duration of *t*.

Step 2. A state of affairs *s* is unrestrictedly repeatable $=_{\mathrm{df.}}$ s is possibly such that:
$(\forall n)(\exists t_1)(\exists t_2)(\exists t_3)\ldots(\exists t_n)(t_1 < t_2 < t_3 < \ldots < t_n$ are periods of time which are *sufficient intervals* for *s*, & *s* obtains at t_1, *s* does not obtain at t_2, *s* obtains at $t_3, \ldots,$ *s* obtains at $t_n \equiv n$ is odd).

In the preceding definition, *n* ranges over all natural numbers, and $t_1, \ldots, t_n$ are non-overlapping. Also, it is assumed for the purposes of this definition that either it is possible that time has no beginning, or it is possible that time has no end.

Step 3. (D1) *x* is omnipotent at *t* $=_{\text{df.}}$ $(\forall s)$ (it is possible for some agent to bring about *s* $\rightarrow$ at *t*, *x* has it within his power to bring about *s*).

In (D1), *x* ranges over agents, and *s* over states of affairs which satisfy the following condition:

(C1) (Ci) *s* is unrestrictedly repeatable, and of the form 'in *n* minutes, *p*', & (*p* is a complex state of affairs $\rightarrow$ each of the components of *p* is unrestrictedly repeatable & possibly brought about by someone), or (Cii) *p* is of the form '*q* forever after', where *q* is a state of affairs which satisfies (Ci).

Two technical explanations concerning (C1) are in order. First, *n* ranges over real numbers, and *p* is not itself equivalent to a state of affairs of the form 'in *n* minutes, *r*', where *n* is not equal to zero. Second, by a *complex* state of affairs is meant a state of affairs which is either constructible out of other states of affairs by use of the logical apparatus of first-order logic enriched with whatever modalities one chooses to employ, or else analysable (in the sense of a philosophical analysis) into a state of affairs which is so constructible. Accordingly, a *component* or *part* of a complex state of affairs *s* is one of those states of affairs out of which *s*, or an analysis of *s*, is constructed.

Happily, (D1) does not require an omnipotent agent to be able to bring about impossible or necessary states of affairs, or states of affairs such as (a)-(f). Moreover, (D1) does not imply that an omnipotent agent lacks any powers which an agent of this kind ought to possess. This is because an agent's bringing about a state of affairs can always be 'cashed out' in terms of that agent's bringing about an unrestrictedly repeatable state of affairs which satisfies the antecedent of the definiens of (D1). In other words, necessarily, for any state of affairs *s*, if an agent *a* brings about *s*, then either *s* is an unrestrictedly repeatable state of affairs which satisfies the antecedent of the definiens of (D1), or if not, then *a* brings about *s* by bringing about *q*, where *q* is an unrestrictedly repeatable state of affairs which satisfies the antecedent of the definiens of (D1). For example, an omnipotent agent can bring about the state of affairs 'that in one minute, Socrates walks for the first time' by bringing about the state of affairs 'that in one minute, Socrates walks', when this walk is Socrates' first. And while the former state of affairs is a non-repeatable one which (D1) does not require an omnipotent agent to be able to bring about, the latter state of affairs is an unrestrictedly repeatable state of affairs which (D1) does require an omnipotent agent to be able to bring about.

5 On the compatibility of God's omnipotence and perfect goodness

Some philosophers have thought that omnipotence is incompatible with certain other attributes. In particular, it has been argued that traditional theism, which posits a necessary, essentially omnipotent, essentially omniscient and essentially perfectly good deity, is faced with this problem. The charge has been that such a deity lacks the power to bring about evil, while non-omnipotent (and morally imperfect) beings such as ourselves possess this power. The exact form of the problem varies depending on what exactly the relation between God and evil is supposed to be. Suppose that if God exists, then this is a best possible world. Then if God exists, there could not be an evil unless it is necessary for some greater good, and any state of affairs containing evil incompatible with there being a maximally good world is impossible. Such a state of affairs could not be brought about by any agent whatsoever. In that case, any evil that any person brings about (moral evil) or that occurs in nature (natural evil) is necessary for some greater good. Assume that God exists and that, for instance, some person, *S*, brings it about that an evil, *e*, exists. Then, given our assumptions, since *S*'s bringing it about that *e* exists is necessary for some greater good, it is compatible with God's perfect goodness that God brings it about that *S* brings it about that *e* exists. On the other hand, if God exists and some or all evil is metaphysically incompatible with this being a maximally good world, then neither God nor anybody else has the power to bring about that evil. The main idea here is that if God's nature places moral restrictions on the nature of the universe and on what God can bring about, then it also places parallel restrictions on what any other agents can bring about. So, for example, if there being suffering in the universe is necessary for this being a maximally good world, then God cannot create a (maximally good) world without the presence of suffering of the appropriate sort (see Evil, problem of).

Other theists, for example, Alvin Plantinga, do not hold that God's existence implies the existence of a maximally

good world, but do hold that God seeks to create as good a world as God can (see Plantinga 1974). They allow that there may be evil that is unnecessary for any greater good that outweighs it. An evil of this sort involves free decisions of agents other than God which God does not prevent but which these other agents can prevent. Plantinga argues that God does no wrong in permitting an evil of this kind because God is powerless to bring about a crucial good - the existence of free human agents - without there being such an evil. Alternatively, it might be argued that God does no wrong in this sort of case because God does not know how to do better, since such knowledge is impossible. However, as an omnipotent God is *not* required to have power over free decisions of agents other than God, it can be seen that on these views, God's omnipotence and perfect goodness can be logically reconciled to the extent indicated along the lines taken above in discussing the view that God's existence requires a maximally good world. Of course, nothing that has been said here decides the issue of how much, if any, evil is compatible with the existence of the traditional God.

See also: Freedom, divine; God, concepts of; Goodness, perfect §§2-3; Monotheism; Process theism

JOSHUA HOFFMAN
GARY ROSENKRANTZ

References and further reading

Aquinas, T. (1266-73) *Summa theologica*, Chicago, IL: Great Books, Encyclopedia Britannica, 1952, vol. 1.(A readily available complete edition with a useful bibliography; see particularly Ia, 25, 3.)

Descartes, R. (1641) *Meditations on First Philosophy*, trans. J. Cottingham, R. Stoothoff and D. Murdoch, in *The Philosophical Writings of Descartes*, vol. 2, Cambridge: Cambridge University Press, 1984.(The standard edition; see particularly meditation 1.)

Flint, T. and Freddoso, A. (1983) 'Maximal Power', in A. Freddoso (ed.) *The Existence and Nature of God*, Notre Dame, IN: University of Notre Dame Press, 81-113.(An attempt to analyse omnipotence.)

Hoffman, J. (1979) 'Can God Do Evil?', *Southern Journal of Philosophy* 17 (2): 213-20.(Discusses the question of the coherence of divine omnipotence and perfect goodness.)

Hoffman, J. and Rosenkrantz, G. (1988) 'Omnipotence Redux', *Philosophy and Phenomenological Research* 49 (2): 283-301.(Critically surveys several analyses of omnipotence in the recent literature.)

Maimonides (*c.*1190) *Guide for the Perplexed*, trans. M. Friedlander, London: George Routledge & Sons, 1904. (See part I, chapter 15. This edition is readily available in paperback published by Dover Books.)

Pike, N. (1969) 'Omnipotence and God's Ability to Sin', *American Philosophical Quarterly* 6 (3): 208-16.(Argues that divine omnipotence and perfect goodness are incompatible.)

Plantinga, A. (1974) *God, Freedom, and Evil*, New York: Harper & Row.(Important free-will defence of theism against the problem of evil.)

Rosenkrantz, G. and Hoffman, J. (1980a) 'What an Omnipotent Agent Can Do', *International Journal for Philosophy of Religion* 11 (1): 1-19.(The original version of the analysis of omnipotence presented in the text above.)

Rosenkrantz, G. and Hoffman, J. (1980b) 'The Omnipotence Paradox, Modality, and Time', *Southern Journal of Philosophy* 18 (4): 473-9.(Discusses the problem of the stone; contains a useful bibliography concerning this issue.)

Wierenga, E. (1989) *The Nature of God*, Ithaca, NY: Cornell University Press. (Wide-ranging discussion of recent work on omnipotence with a defence of the author's analysis; see chapter 1.)

Omnipresence

Western Scripture and religious experience find God present everywhere. Western thinkers make sense of this as their concepts of God dictate. Pantheists hold that God's being everywhere is every bit of matter's being a part or an aspect of God. Panentheists say that as God is the soul of the universe, God's being everywhere is his enlivening the whole universe as souls enliven bodies. But most theists reject these views, as most think that if God is perfect, he cannot be, be made of, or be embodied in a flawed and material universe. Most theists think God intrinsically spaceless, that is, able to exist even if no space exists. Still, theists argue that God's knowledge of and power over creation make him present within it without occupying space or being embodied in matter. Some add that God is present in space not just by power and knowledge but in his very being. These try to explain a spaceless God's presence in space by likening it to the presence of a universal attribute like hardness. Hardness is not spread over hard surfaces, occupying them by having parts of itself in parts of them. Each part of a hard surface is hard. So all of hardness is in each part of a hard surface. So too, theists say, God is not spread out over space, filling parts of it with parts of himself. Rather, all of God is wholly present at each point in space and in each spatial thing.

1 Introduction

Western Scripture vividly states that God is present everywhere. The Psalmist asks:

> Where can I go from your Spirit?
>
> Where can I flee from your presence?
>
> If I go up to the heavens, you are there;
>
> If I make my bed in the depths, you are there.
>
> If I rise on the wings of the dawn, if I settle on the far side of the sea,
>
> even there your hand will guide me, your right hand will hold me fast.

(Psalms 139: 7-10)

Theists fill out the concept of God mainly by drawing inferences from God's being perfect (see God, concepts of §§2-6). Such thinking also backs the claim that God is omnipresent: a God from whose presence one could escape would be less perfect than a God inescapable.

If time and space are really one, God is omnipresent in time as in space. But most have thought of omnipresence only in relation to space, so this entry treats only space. We are present in each volume of space which includes us: the space of this room, and also this galaxy. But we are primarily present only in places we fill (I am in the galaxy because my body fills a place which is in the galaxy) and immediately present only to places directly next to our places. So if God is perfectly present to every place, he either fills or is next to each place. If nothing can exclude God, God fills not just places but the things which occupy them.

2 Omnipresence and embodiment

Pantheism is the doctrine that all material things compose or manifest one thing, which is God (see Pantheism). If all matter is part or an aspect of God, then God is wherever any material object is, that is, wherever God has a part or aspect. But the whole of God is not where any one part of God is, and God is not literally everywhere if there are true vacua, though he permeates all space if (say) some part of God is present in each volume of space, however small. Treatments of omnipresence should distinguish its nature (what it is to be omnipresent), cause (what makes God omnipresent) and manner (the kind of presence God has everywhere). For pantheism, then, God's omnipresence (if any) is the way space contains God; it rests on God's having every object as a part or aspect, and God is present as material objects are: parts of God in parts of a location, and the whole of God in the whole. Spinoza (§§2-4) (if a pantheist) is an exception, claiming that though all extension is an attribute of God, and all bodies modes of God, God has no parts or at least is indivisible (*Ethics* I, props 12-13).

For Stoic panentheists, 'God' named the spirit whose body is the cosmos, and also the living cosmos that spirit

enlivens. So Stoics thought that God permeates the cosmos as a soul permeates its body, sustaining the cosmos as soul sustains body. The Stoic world-soul was a fiery material stuff. So Stoics held that the stuff of God occupies space and shares spaces with all other material stuffs, which are not parts of God quite as the cosmos' fiery parts are. The Stoic theory of mixture allowed this, claiming that all ingredients of a mixture are present in each place the mixture occupies. For Stoicism, the cosmos existed amid a surrounding vacuum, and so though omnipresent in the cosmos, God was not literally present in every place. For Stoics, a finite space circumscribes God: the cosmos' dimensions are God's. To Stoics, then, God's omnipresence in the cosmos is the way he enlivens it; it rests on the world's embodying God, and God is in the world as soul in body (see Stoicism §§3-5).

One's account of omnipresence depends largely on whether one holds that:

(UB) the universe is God's body.

Why accept (UB)? If one finds pure physicalism a persuasive metaphysic, (UB) gives God a place within it. It surely is at least easier to understand the claim that God acts if God has some sort of body; this makes his actions more like those of persons we know. The Stoics thought that only material things *can* act. Tertullian agreed, and inferred that God is a body (*Against Praxeas*, ch. 7); the Stoics inferred that God must be not just material but the most perfect material thing, the cosmos (see God, concepts of §3). Swinburne argues that embodiment comes in degrees, and that the way God knows and acts in the universe embodies him in it to some extent. That is one's body, Swinburne reasons, which one can control or move directly as a 'basic action', and about whose states one knows immediately (1993: 104-6). Hartshorne (1964: 200) makes the same case. Sarot (1992) argues that God has emotions which require having a body. Some feminists (for example, McFague 1987: 69-77) argue that its ethical and spiritual consequences favour (UB).

None of this is compelling. Bodiless agency, mental life and feelings are conceivable; recent philosophy of mind helps by stressing the *functional* cores of our concepts of thought, action and emotion (Alston 1989: 64-80). Further, we make God more conceivable only by moulding our idea of God more in our own image. Some degree of this is inevitable, but it is wise to minimize it if Anselm of Canterbury (§4) is right that God's greatness puts him beyond our full conceptual grasp (*Proslogion*, ch. 15). If Swinburne and Hartshorne offer more than argument by analogy, they premise that necessarily, if *x* can move *y* directly or know about *y* immediately, then *x* is embodied in *y*. This is false if telekinesis and telepathy are logically possible. For it is not plausible that while a telepath reads my thoughts or a telekinetic moves my limbs, my body houses two persons. Finally, ethical and spiritual intuitions differ greatly, and so are frail grounds for metaphysical conclusions.

Why has mainstream Western theism so wholly rejected divine embodiment? If God is the most perfect possible being, God cannot exist only as embodied in part of the universe. For if he did, a being greater than him would be possible, namely one with just his knowledge and power, but the whole universe as a body. But even (UB) renders God physically finite if (as physicists now believe) the universe is finite. If he is physically finite, God has a surpassable value and importance; were he/the universe bigger, he might well be better.

If (UB) is true, the universe is all living matter; it is one large animal. Plato (§16) was willing to say this (*Timaeus*, 30b-c), but few wish to follow him. Again, the universe includes many items whose inclusion in God seems beneath God's dignity. If the universe is God's body, flaws in the universe are flaws in God's body. If my body is flawed, *I* am flawed; we say that *I* have a bad back, not merely that my body has one. Thus flaws in the universe are flaws in God; and if there is in fact a problem of natural evil, the universe is flawed (see Evil, problem of §§1, 6). Again, if part of my body is ill, *I* am ill. But then when parts of God's body are ill, as when humans have colds, God himself is sick. Being able to be ill, however, is an imperfection.

We can damage material things. If these are parts of God's body, then we can damage God, since to damage parts of someone's body is to damage that person. We can also literally push God (or parts of God) around. Such divine vulnerability seems unbecoming. It is why the Christian doctrine of the Incarnation shocked the Greeks and provoked Chalcedonian Christology as an antidote (see Incarnation and Christology §1).

If (UB) is true, then either necessarily God exists only if embodied, or God has a body only contingently. On the first alternative, God exists contingently (Swinburne), or the universe exists necessarily (Hartshorne); both claims are highly debatable. If God exists only if embodied, the universe *is* God; it constitutes him. But if the universe is God, it would still be God if it contained one atom less. If so, it would still be God without yet another atom, and

so on, to the eventual conclusion that a universe consisting of a single atom would still be God. But a single atom is certainly destructible; hence, God is possibly destructible, and so destructible now. The universe can be caused to exist; it is quite possible that the Big Bang had a cause. So if God does not exist save as embodied in a universe, God can be caused to exist - and cannot have created the universe *ex nihilo*. Further, if we are parts of God, our freedom is a lack of divine self-control, another imperfection in God.

Finally, many scientists and philosophers think there can be other universes, not spatially or temporally connected with ours (see Cosmology §3). Suppose that there is a second universe. Then if God is our universe, he cannot have made the other. His bringing it to be, an act in our time and space, would connect that universe to ours, and so the two would not be spatiotemporally discrete after all. Nor can God have both universes as bodies. If God's thoughts in one universe are spatiotemporally connected to his thoughts in the other, then God's thoughts connect the universes, making one of what we supposed were two. But one being cannot have two spatiotemporally unconnected lifelong streams of thought. So if there can be two universes and God must be embodied, there can be a universe God did not make, and there is a possible contingent object, a second universe, that God cannot create. Further, if a universe can embody a God, then if there can be two universes, arguably there can be a second God.

The view that God can exist unembodied but is contingently embodied may escape some of the difficulties raised. But its price may be emptying (UB) of most of its content: what does (UB) mean if the objections above do not apply, and how does it differ from traditional unembodied theism?

3 Omnipresence and divine spatiality

Many rabbis called God the universe's place (*makom*). St Paul said of God 'in him we live and move and have our being' (Acts 17: 27-8). Picking up this theme, Anselm declaimed to God that 'You are not in space… but all things are in you. For you are not contained by anything but rather you contain all else' (*Proslogion*, ch. 19). For Anselm, then, God's omnipresence is his way to contain all spatial things; Anselm *may* take this containment spatially (Leftow 1989).

For the Cambridge Platonist More (§5), all substance, as such, is extended, and infinite space is in fact God's infinite extension. Bradwardine and Oresme had moved this way in the fourteenth century; Spinoza too took infinite extension to be a divine attribute (*Ethics* I, prop. 15, proof). Newton (who read More) held that God is not space, but is present everywhere, and by being everywhere makes space exist. For Newton, God, by being spatial, establishes space: space is absolute in the sense of not being reducible to *creatures'* spatial relations, but reduces to or depends on God's spatiality. For Berkeley (§§6-7), to be is to be perceived by God, and so all space is divine perceptual space, a space 'within God's mind'. Newton was not a Berkeleyan divine phenomenalist, but his claim that infinite space is God's *sensorium* also seems to make all space God's perceptual space.

For these thinkers, even if God is immaterial, he is spatially extended or space is somehow literally in God. God's omnipresence is the way God spatially includes space and creatures; it rests on God's being, or being the foundation of, space, and God is present as space is. For all save Berkeley, the whole of God is not present at any point or to any creature.

4 The omnipresence of a spaceless God

In the West, Platonism first took God to be immaterial and intrinsically spaceless (that is, able to exist even if no space exists). So Platonism first discussed a spaceless God's presence in space. Plotinus (§§3-5), whose views were most important historically, took being in to be a variety of depending on. Spatially contained objects depend on their containers. So for Plotinus, spatial containment is one kind of being in. But it is not the only kind (*Enneads* VI, 4, 2). Plotinus held that soul is not in body but body in soul, and God is not in the world, but the world in God. For God to be omnipresent, he thought, is for God to 'contain' space and the world, that is, for the whole universe to depend on God. The availability since Plotinus of non-spatial uses of 'in' and 'contain' cautions us in reading the apparently spatial language of later treatments of omnipresence.

Plotinus thought God present everywhere as a whole, because each part of space causally depends on the whole of God. He took God to be whole in the whole of space and whole in each part of space, rather than having parts occupying parts of space. Plotinus noted that even if an object controlled or conserved has parts, its conserver need not have parts, for the whole controlling power is related as a whole to each part of the controlled object, rather

than part by part. He used analogies to suggest that being whole in the whole and in each part is conceivable even apart from his causal account of being in. Light seems to be everywhere without being divided into parts; the sound of a spoken word seems to be present as a whole in an entire area and also in each part of that area, since in each part one can hear the whole sound. Plotinus also argued that an immaterial soul is present as a whole in its whole body and in each part of its body, claiming that only this can explain how the soul feels pain in all parts of the body. For Plotinus, then, to be omnipresent is to contain the universe; God's omnipresence rests on God's causal relation to the universe; and God's presence is causal, not spatial, and not by parts.

Augustine (§§7-8) brought Plotinus' views into Christian thought, taking over Plotinus' general account of omnipresence, his argument about the soul's presence in the body, and his analogies of light and sound. Again, like Plotinus, Augustine models God's presence on that of abstract entities. The whole of a universal such as blueness is present at each point in a blue surface; so, too, the whole of God is at each point of space. Making a blue area larger does not make it bluer; being bluer is a matter of intensity. So, too, God's presence varies intensively, not extensively. As wholly present in any space or thing, God need not be more present in a large area than in a small, but God is more present in a saint than in a sinner. A universal such as blueness is in space without having parts covering parts of space (the parts of a blue surface are parts of the surface but not of blueness); so, too, God. Blueness does not have as dimensions (limits of its spatial extent) the spatial dimensions of its location(s). For wherever it is located, blueness is whole (each point in a blue surface is fully blue) and yet also present as a whole outside that place (at other places, or outside space altogether); again, the same is true of God. The abstract-entity comparison suggests that an immaterial God can fill a place without being contained by it, excluding other things from it, or being intrinsically spatial or physical. Perfect-being theology finds presence as a whole everywhere a more perfect way to be present, making God more fully present at each place.

Augustine took God's creating and sustaining all things to provide a further sense (like Plotinus') in which he is present in them; Aquinas made the Aristotelian roots of this thinking plain (*Summa theologiae* Ia, 8, 1). Aristotle insists that agents move only by direct contact with what they move (*Physics* VII, 2). Aquinas (among others) reads this as implying that a 'contact of power' links any immaterial agent and its effects (*Summa theologiae* Ia, 8, 2 *ad*. 1; 52, 1). If two things touch (that is, are in contact), they are in some sense at the same place. On the doctrine of divine simplicity Aquinas (§9) and other medieval thinkers shared, God is identical to God's power (*Summa contra gentiles* II, 8). So if God's power is in some way at the place of its effects, so is God. Later medieval philosophers (such as William of Ockham §9) increasingly questioned the doctrine of contact action. If action at a distance is possible, God's acting at a place does not entail that he is in any way present there. But a more basic point is that 'contact of power' seems just a metaphor. Whether or not things can act at a distance, immaterial things in no way literally touch material things.

Aquinas inherited from Lombard's *Sentences* (I, d. 37) the formula that God is in all things by power, presence and essence. Aquinas took God's indwelling by presence to be his having full, immediate knowledge of all things, and so took God's knowledge of spatial things to provide yet another sense in which God is present to them (*Summa theologiae* Ia, 8, 3). Perhaps Aquinas' thought is that if all things are present to God's knowledge, then as 'present to' is a symmetric relation, it follows that God's knowledge is present to all things, and given divine simplicity, God is identical to God's knowledge (*Summa contra gentiles* I, 45).

For Augustine and Anselm (*Monologion*, chaps 13-14), God is omnipresent because and in the sense that he creates and rules all. For Karl Barth, God rules all because he is omnipresent: 'all divine sovereignty… is based on the fact that for God nothing exists which is… not near' (1957: 461). For Aquinas, God is omnipresent because and in the sense that he knows all things immediately. For Newton, God knows because he is present:

> The *sensorium* of animals (is) that place to which the sensitive substance is present and into which the sensible images of things are carried… that there they may be perceived by their immediate presence to that substance… there is a Being, incorporeal… intelligent, omnipresent, who in infinite space, as (if) in his *sensorium*, sees the things themselves… and comprehends them wholly by their immediate presence to himself.
> (*Optics*, query 28; appended to Leibniz and Clarke [1717] 1956)

If God's omnipresence has some truly spatial meaning, it has an independent content which lets it explain other divine attributes. If it has none, the doctrine of omnipresence reduces to a metaphor for God's omniscience and creatorhood. This reduction may not be what Scriptural authors intended. Jeremiah declaims:

‘Can anyone hide in secret places so that I cannot see him?’ declares the Lord.

‘Do I not fill heaven and earth?’ declares the Lord.

(Jeremiah 23: 24)

The suggestion seems to be that God’s knowledge rests on his omnipresence, not vice versa. Again, the religious (and even more the mystical) sense of God’s presence seems to be not just that God knows us or creates us, but that God is in some sense here with us. So something closer to literal spatial location seems needed. Spaceless-God accounts do not provide it, save by the abstract-entity analogy.

See also: Creation and conservation, religious doctrine of §5; Occasionalism

BRIAN LEFTOW

References and further reading

Alston, W. (1989) *Divine Nature and Human Language*, Ithaca, NY: Cornell University Press.(Important work on the concept of a bodiless God’s action in the world. Clear and careful.)

Anselm (1076) *Monologion*, in J. Hopkins and H. Richardson (eds and trans) *Anselm of Canterbury*, Toronto, Ont.: Mellen, 1974, vol. 1.(Explores God’s relation to space at length.)

Anselm (1077-8) *Proslogion*, in J. Hopkins and H. Richardson (eds and trans) *Anselm of Canterbury*, Toronto, Ont.: Mellen, 1974, vol. 1.(Stresses God’s ‘containing’ space.)

Aquinas, T. (1259-65) *Summa contra gentiles*, trans. A. Pegis (bk 1) and J. Anderson (bk 2), Notre Dame, IN: University of Notre Dame Press, 1975.(Clear, careful statement of classical theism.)

Aquinas, T. (1266-73) *Summa theologiae*, New York: Benziger Brothers, 1948.(Book 1, question 8 treats God’s relation to space.)

Aristotle (mid 4th century BC) *Physics*, trans. R. Hardie and R. Gaye, in R. McKeon (ed.) *The Basic Works of Aristotle*, New York: Random House, 1941. (Gives the view of causal agency supposed in many treatments of omnipresence.)

Augustine (417) ‘Letter 187’, in *Letters 165-203*, trans. W. Parsons, The Fathers of the Church, vol. 30, ed. J. DeFerrari, New York: Fathers of the Church, Inc., 1955. (His most thorough treatment of God’s relation to space.)

Barth, K. (1957) *Church Dogmatics* II: 1, ed. and trans. G.W. Bromiley and T.F. Torrance, Edinburgh: T. & T. Clark.(Treats central divine attributes including omnipresence. Rambling but interesting.)

Brom, L. van den (1993) *Divine Presence in the World*, Kampen: Kok Pharos. (Idiosyncratic argument for a spatial God; wide in scope.)

Hartshorne, C. (1964) *Man’s Vision of God*, Hamden, CT: Archon.(Freewheeling case for and depiction of an embodied God.)

Jantzen, G. (1984) *God’s World, God’s Body*, Philadelphia, PA: Westminster. (Clear exposition and defence of the claim that the universe is God’s body.)

Koyré, A. (1957) *From the Closed World to the Infinite Universe*, Baltimore, MD: Johns Hopkins Press.(Valuable study of interaction between views of divine omnipresence and developing views of space.)

Leftow, B. (1989) ‘Anselm on Omnipresence’, *The New Scholasticism* 63 (3): 326-57.(Reads Anselm’s talk of God containing the world spatially.)

Leibniz, G.W. and Clarke, S. (1717) *The Leibniz-Clarke Correspondence*, ed. H.G. Alexander, Manchester: Manchester University Press, 1956.(Includes relevant extracts from Newton.)

McFague, S. (1987) *Models of God*, Philadelphia, PA: Fortress.(Examines the ethical and spiritual advantages of taking the universe to be God’s body.)

Plato (*c*.366-360 BC) *Timaeus*, trans. B. Jowett, in E. Hamilton and H. Cairns (eds) *Collected Dialogues of Plato*, Princeton, NJ: Princeton University Press, 1961. (His cosmology, depicting the material world as ensouled and alive.)

Plotinus (*c*.250-66) *Enneads* VI, trans. A.H. Armstrong, Cambridge, MA: Loeb Classical Library, Harvard University Press, 1988.(Seminal early treatment of spaceless beings’ presence in space.)

Sarot, M. (1992) *God, Passibility and Corporeality*, Kampen: Kok Pharos.(Case for divine embodiment, with encyclopedic references to literature.)

Spinoza, B. (1676) *Ethics*, in E. Curley (trans.) *The Collected Works of Spinoza*, Princeton, NJ: Princeton University Press, 1985, vol. 1.(Book 1 gives his theory of a spatial God.)

Swinburne, R. (1993) *The Coherence of Theism*, revised edn, New York: Oxford University Press.(Clear

treatment of many issues in philosophical theology; discusses omnipresence, disembodied agency and partial divine embodiment.)

Tertullian (*c*.213) *Against Praxeas*, in A. Roberts and J. Donaldson (eds) *The Ante-Nicene Fathers*, Peabody, MA: Hendrickson, 1994, vol. 3.(Second-century Christian polemic; treats God as embodied.)

Wainwright, W. (1987) 'God's World, God's Body', in T. Morris (ed.) *The Concept of God*, New York: Oxford University Press, 72-87.(Fine, clear discussion of pros and cons of divine embodiment.)

Omniscience

The concept of omniscience has received great attention in the history of Western philosophy, principally because of its connections with the Western religious tradition, which views God as perfect in all respects, including as a knower. Omniscience has often been understood as knowledge of all true propositions, and though several objections to any simple propositional account of omniscience have been offered, many philosophers continue to endorse such an analysis. Advocates of divine omniscience have discussed many problems connected with both the extent of omniscience and the relation between this property and other alleged divine attributes. Three such issues are: Can an omniscient being properly be viewed as immutable? Would an omniscient being have knowledge of the future, and is such knowledge consistent with our future actions' being genuinely free? And should omniscience be thought of as including middle knowledge? That is, would an omniscient being know (but have no control over) what other free beings would in fact freely do if placed in various different situations?

1 Omniscience and perfection

The concept of omniscience, though of considerable philosophical interest in itself, has attracted enormous attention throughout the history of philosophy primarily because of its theological connections. Most Western theists have embraced a picture of God as a being perfect in all respects, and this commitment has naturally led philosophers in this tradition to investigate the notion of noetic perfection. Such perfection is plausibly thought to entail omniscience - knowledge of everything. Hence the view that God is omniscient has been commonplace in Western monotheism.

Of course, to endorse this claim of divine omniscience is to open the door to several obvious questions. How exactly is omniscience to be understood? Of what is an omniscient being knowledgeable? And are there not reasons (related to time and human freedom) to suspect that omniscience might be a more limited perfection than one might initially have thought?

2 The analysis of omniscience

At first glance, the concept of omniscience, unlike many other divine attributes, would seem to be open to a very simple and direct analysis: a being is omniscient if and only if it knows all true propositions. Many philosophers would contend that such an analysis is indeed fully adequate. Still, questions have been raised.

Some, following Aquinas, have denied that perfect knowledge is properly thought of as propositional in structure. Our need to carve up our acquaintance with reality via discrete, complex propositions, it is argued, is a sign of our imperfection. A truly perfect knower would have no need for propositional knowledge, but instead would possess a simple and immediate intuitive awareness of reality. Many of the advocates of this position see it as flowing from the notion of divine simplicity, an attribute which has received renewed attention in recent years (see Simplicity, divine). Most philosophers, though, are reluctant to ascribe absolute simplicity to God; not coincidentally, perhaps, the view that God's knowledge is wholly non-propositional is also a minority view.

However, the analysis of omniscience offered above can also be questioned even by those who view divine knowledge as propositional. Assuming that God knows all true propositions, one might argue, his knowledge would still be imperfect if he knows *only* propositions. Three types of knowledge which supposedly go beyond the propositional are of particular interest here.

The first is *de re* knowledge, knowledge of things rather than knowledge of propositions. *De re* knowledge often appears to be present when the corresponding propositional, or *de dicto,* belief is not. For example, if JoAnn mistakenly believes that my pet iguana Willard is an alligator, then she might know *de re* of my pet iguana that he is green, even though the *de dicto* belief that my pet iguana is green is one that she rejects. Though the distinction between *de re* knowledge and propositional knowledge seems reasonable, this distinction may not require any emendation of the analysis of omniscience, since it is not clear that a being who knows all true propositions (including such truths as that Willard is an iguana) would be lacking any *de re* knowledge.

A second and similar objection to the propositional analysis of omniscience focuses on *de se* knowledge, knowledge of oneself. Such knowledge, one might contend, can be present even in the absence of the

corresponding *de dicto* knowledge. If Anastasia, hospitalized with amnesia and unaware of her name, reads an old newspaper account of her early life, she may well know *of herself* that she is in hospital, even though she does not believe that Anastasia is in hospital. Furthermore, one might question whether anyone distinct from Anastasia *could* possess the exact *de se* knowledge she has. Some defenders of the propositional analysis attempt to deal with this *de se* objection in a manner similar to the *de re* one: complete propositional knowledge, they argue, rules out any absence of knowledge which is held *de se*. Such responses call upon us to view Anastasia's *de se* knowledge as knowledge of a special type of proposition (perhaps involving her individual essence or haecceity), one which someone distinct from Anastasia could know, and one which anyone who knows all true propositions would know. Since many would question the existence of such special propositions, though, this type of response is quite controversial (for further discussion, see Wierenga 1989).

The third objection to the propositional analysis suggests that there is a distinctive qualitative dimension to certain kinds of knowledge which is ignored by that analysis. For example, even the most exhaustive propositional knowledge about chocolate leaves out a crucial element in a complete knowledge of chocolate - namely, knowing what chocolate tastes like. A being who knew all true propositions about chocolate, but who lacked actual acquaintance with its taste, could not be said to have perfect knowledge. Various responses to this objection are available. Some defenders of the propositional analysis flatly deny that experience of the sort suggested is required for one's knowledge of chocolate to be complete; others agree that we would expect a perfect being to possess such qualitative awareness, but deny that it should be thought of as knowledge in anything more than an analogical sense.

Though the three objections discussed here pose significant problems for the propositional analysis, problems sufficient to have led some philosophers to complicate or abandon it, none is clearly decisive, and many advocates of the propositional view remain. Whatever one's verdict on this matter, though, one who embraces the belief that God is omniscient faces many difficult questions concerning both the scope of omniscience and the compatibility of divine omniscience with other attributes commonly ascribed to God.

One such question is how a God who has perfect knowledge of our ever-changing world could himself be utterly immutable. Since immutability itself is often viewed as a perfection, and since it seems to be entailed by two other attributes (simplicity and eternity) frequently deemed essential to a perfect being, many theists are strongly inclined to view God as unchanging (see Immutability). Yet complete knowledge seems to require change on the part of the knower whenever there is change on the part of that which is known. In the face of this objection, some theists have abandoned the view that God is utterly immutable; they suggest that so long as God is stable in his powers, constant in his providential care for his creatures, resolute in his intentions, and so on, we need not be concerned by trivial changes in his noetic activity. Other theists have denied that the argument forces us to jettison immutability. For example, advocates of divine eternity have maintained that God's eternal perspective would allow him (in a sense) simultaneously to view events that are not simultaneous from our perspective; hence, his knowledge of all such events need imply no change in him (see Eternity). Though fraught with the potential for (and, as critics see it, the actuality of) paradox, this view has been developed with ingenuity by several writers (see especially Stump and Kretzmann 1981).

In addition to the question concerning immutability, two other significant issues connected with omniscience have received much critical attention. First, does an omniscient being know what we *will* freely do in the future circumstances which will in fact obtain? Second, does an omniscient being know what we *would* freely do were other circumstances to obtain?

3 Omniscience, foreknowledge and freedom

If God is all-knowing, it seems reasonable to assume that he knows not only what did happen and what is happening, but also what will happen. But how could comprehensive divine foreknowledge be compatible with the contingency of what is foreknown? In particular, how could our actions be free, be such that we could have done otherwise, if an infallible God knew in advance what we would do? This question has vexed some of the greatest minds in Western history, from the Greek tragedians (especially Sophocles) to such Christian luminaries as Augustine (§§6, 10), Aquinas and Jonathan Edwards (§2); in contemporary philosophy of religion, it remains perhaps the most-discussed problem connected with the attribute of omniscience.

Though the problem has been explicated in many different ways, perhaps its strongest form can be presented as follows. Suppose God knows that I will perform a certain action tomorrow - for example:

(1) God knows that I will go horseback riding tomorrow.

Clearly, this cannot be knowledge that God has recently acquired; if his omniscience is (so to speak) a stable feature, it must be knowledge he has always possessed. So:

(2) God knew last year that I will go riding tomorrow.

But (2) tells us about what was true in the past, and we generally think that what happened in the past is necessary - not, perhaps, logically necessary, but necessary in the sense of being fixed, settled and beyond anyone’s control. In this sense of necessity (which is variously called temporal necessity, accidental necessity and necessity *per accidens*), it follows from (2) that:

(3) It is necessary that God knew that I will go riding tomorrow.

Of course, no one can know something which is false. Indeed, in the case of God, traditional theists insist that God cannot so much as believe something which is false. So from (3) it follows that:

(4) If God knew that I will go riding tomorrow, then I will go riding tomorrow.

Now, (3) tells us that a certain truth is beyond anyone’s control, while (4) spells out a logical consequence of that truth. But surely what follows logically from something which is necessary must itself be necessary in that same sense. If (3) and (4) are beyond dispute, then, it follows that we are also saddled with:

(5) It is necessary that I will go riding tomorrow.

And since what I do by necessity is not something over which I hold the reins of power, it follows that my act of riding tomorrow is beyond my control, and hence will not be a free action. As this argument form could evidently be unharnessed from our equine example, the conclusion to which we are led is that no action can be free if God possesses foreknowledge.

Responses to this argument are many and varied; virtually no step is beyond dispute. Some have reacted by accepting the conclusion that foreknowledge rules out freedom and consequently rejecting the claim that our actions are free. Others who embrace the incompatibilist conclusion (the thesis that foreknowledge and freedom are incompatible) react in the opposite way: instead of denying freedom, they deny foreknowledge, and so deny that a premise such as (1) is ever true of a free action. Rather than denying that God is omniscient, many such incompatibilists suggest that omniscience does not require foreknowledge. Just as omnipotence does not entail the power to do what cannot be done (see Omnipotence §§1-3), so, they argue, omniscience does not include the power to know what cannot be known; since foreknowledge of free actions has been shown by this argument to be impossible, we ought not expect an omniscient being to have such foreknowledge (see Hasker 1989). Other incompatibilists, especially those favourably disposed towards the notion of divine eternity, criticize the argument’s blatantly putting God in the temporal realm. No divine cognitive activity, they insist, takes place in time, and hence the use of ordinary tensed propositions to refer to such activity is out of place. For such incompatibilists, (1) and especially (2) ought to be rejected.

Compatibilist rejoinders (that is, responses by those who see foreknowledge and freedom as compatible) come in two basic forms. The first of these is to deny that (3) follows from (2). Many facts about the past - those commonly called hard facts about the past - are exclusively about the past, and these do have the kind of necessity at issue in this discussion. But (2) is not exclusively about the past; it is about the future as much as it is about the past, and hence qualifies as only a soft fact about the past. Unlike hard facts about the past, soft facts need not be seen as necessary; hence, (3) does not follow from (2), and the argument dissolves.

This type of response, which draws its inspiration from William of Ockham, is often referred to as the Ockhamist response. An amazing amount of literature on this position has been generated and a large number of variations on it proposed. Various ways of understanding the notion of temporal or accidental necessity have been offered, and numerous means (often of great technical complexity) of distinguishing hard from soft facts have been formulated

(see Plantinga 1986). Some commentators have even suggested that the hard/soft distinction may be too coarse for the problem at hand: even if (2) turned out to be a soft fact about the past, it might well have a hard 'core' - such as the fact that God *believed* last year that I will ride tomorrow - which is sufficient for an analogue of the incompatibilist argument to succeed.

The second basic form of compatibilist response is to deny that (5) follows from (3) and (4). More generally, such a response denies that accidental necessity is closed under entailment or logical implication; that is, it suggests that an accidentally necessary proposition can entail a proposition that is not accidentally necessary. Such a response, which was suggested by the sixteenth-century Spanish theologian Luis de Molina (§§1-3) and has come to be known as the Molinist response, seems most plausible when accidental necessity is understood as applying to those truths which are beyond our ability to cause to be false (see Molinism). So understood, (3) seems clearly true (at least as long as we grant that we lack causal power over the past), as does (4), but (5) does not clearly follow. If we already knew that I had no power to cause my *not* riding tomorrow, the inference might be uncontroversial; but, of course, my causal impotence with regard to the future is what the argument is trying to prove and hence can hardly be assumed in this context.

It has been suggested that whether an Ockhamist or a Molinist response to the argument is more appropriate depends upon how accidental necessity is analysed. Under some interpretations of necessity (ones which equate the accidentally necessary with that which I cannot cause to be false), (3) will follow from (2), but (5) will not follow from (3) and (4). Under other interpretations (ones which equate the accidentally necessary with that which would be true no matter what I might do), (5) will be derivable from (3) and (4), but (3) will not follow from (2). On this view, the incompatibilist argument derives whatever plausibility it has from equivocation.

4 Omniscience and middle knowledge

Closely related to the issues discussed in the preceding section is the question of middle knowledge. Molinists have claimed that between God's natural knowledge (of necessary truths not under his control) and his free knowledge (of contingent truths under his control) is a third category of divine knowledge called middle knowledge - knowledge of contingent truths not under God's control. The most significant components of middle knowledge, according to Molinists, are the counterfactuals of creaturely freedom. These are conditional statements specifying, for any creature God might create and any situation in which that creature might be placed and left free, what it would freely do if placed in that situation. Molinists insist that, while there are truths of this sort, God would have no control over them because such control would conflict with creatures' being free in the full-fledged libertarian sense.

As Molinists see it, counterfactuals of creaturely freedom help us to understand how it is that God knows the future. Prior to any creative decision on his part, God knows via middle knowledge how any set of creatures he might create would freely respond to any creative action he might perform. Using this knowledge, God can decide which beings he will create in which situations. Given this decision, he can know immediately what free actions will in fact be performed. Assuming that all other facts about the future result either from necessary truths or from other divine decisions (for example, regarding which causal laws to establish and when, if ever, to intervene in their normal operations), it follows that God's foreknowledge would be not only complete, but also in no sense dependent upon his observing or being affected by that which he foreknew.

If God possesses middle knowledge, though, its value to him far exceeds its ability to produce foreknowledge. For central to the tradition is the claim that God not only knows what will happen, but directs the course of events; he is a sovereign, provident deity, not an all-knowing but passive witness to earthly occurrences. Middle knowledge gives God the means to exercise such providence in a very strong sense, for it allows him a clear way to incorporate even the free actions of his creatures in a complete and specific providential plan.

The advantages middle knowledge would offer in constructing a coherent picture of an omniscient God are widely, though not universally, acknowledged. Also recognized is the fact that there is a strong *prima facie* case for thinking that an omniscient being would know the sort of truths which compose middle knowledge. On the surface, it surely looks as though counterfactuals of the sort at issue - such as 'If I were to go to the doctor tomorrow, the nurse would take my blood pressure' - have a truth value, one which we often know and upon which we base our actions. Even when we do not know whether such a conditional is true or false, our natural assumption is that there

is a fact of the matter here, a fact which we would like to know and which any omniscient being would have to know.

Despite its theoretical fecundity and initial plausibility, middle knowledge has been the subject of enormous controversy, and several powerful arguments have been marshalled against it. Some critics, following the Thomistic attack initiated by Molina's contemporary Báñez, have charged that the Molinist position is founded on a view of freedom which is metaphysically and theologically misguided. God's sovereignty, they have contended, would be compromised were he powerless in the face of contingent truths about his creatures. A careful analysis of the notion of freedom, they suggest, allows us to see the counterfactuals in question as subject to divine control.

A more common line of attack among contemporary philosophers of religion grants Molinism its libertarian picture of freedom, but insists that this picture requires us to deny the possibility of true counterfactuals of creaturely freedom. For what, it is asked, could ground such a truth? If the action in question is truly free, no element of the situation could necessitate its performance; nor, in many cases, will there be any activity in the past, present or future of the actual world which can be seen as grounding the counterfactual, for most such conditionals will refer to creatures and/or situations which are never actual. There are truths, such critics grant, about what creatures *might* do if placed in certain circumstances, and perhaps even truths about what they *probably* would do, but truths of this sort fall far short of providing God with the kind of foreknowledge and providential control championed by the Molinist. While some critics view the apparent failure of Molinism with regret, others see it as a cause for celebration; a God who needs and is willing to take risks in interacting with his creatures is actually preferable, they suggest, to the all-controlling, manipulative God of the Molinists.

Molinists have offered thoughtful responses to such critics. Abandoning the libertarian picture of freedom, they have argued, is philosophically indefensible and theologically dangerous, in so far as it would render God ultimately responsible for evil. Some Molinists have responded to the grounding objection by questioning either the meaning or the truth of the claim that all propositions need grounds; others have suggested that counterfactuals of creaturely freedom do indeed have grounds, though (as with propositions about the past or the future, or about what is metaphysically possible) we might not find such grounds at the present time or in the actual world. Finally, Molinists have charged that the risk-taking, playing-the-odds picture embraced by some of their critics gives us a God who is little more than 'the bookie than which none greater can be conceived', a deity far inferior to that required by the Western religious tradition.

See also: God, concepts of; Providence §§3-4

THOMAS P. FLINT

References and further reading

Adams, M. (1967) 'Is the Existence of God a "Hard" Fact?', *Philosophical Review* 76 (October): 492-503.(A response to the Pike article cited below; offers a version of the Ockhamist position noted in §3.)

Alston, W. (1986) 'Does God Have Beliefs?', *Religious Studies* 22 (3-4): 287-306. (A clear, accessible discussion of the non-propositional picture of divine knowledge noted in §2 and the implications of such a picture for problems connected to immutability and foreknowledge.)

Fischer, J. (1983) 'Freedom and Foreknowledge', *Philosophical Review* 92 (January): 67-79.(Prominent attack on the type of Ockhamism defended by Adams in the essay noted above.)

Flint, T.P. (1990) 'Hasker's *God, Time and Knowledge*', *Philosophical Studies* 60 (1): 103-15.(An attempt to respond to the anti-Molinist argument offered in the Hasker book listed below.)

Freddoso, A. (1988) 'Introduction', in A. Freddoso (trans. and ed.) *On Divine Foreknowledge: Part IV of the Concordia*, Ithaca, NY: Cornell University Press, 1-81.(Excellent detailed presentation of the Molinist positions described in §3 and §4, along with summaries of and responses to major criticisms of Molinism.)

Hasker, W. (1989) *God, Time and Knowledge*, Ithaca, NY: Cornell University Press. (Challenging and sophisticated arguments against Molinism, along with a defence of the notion of a risk-taking God noted in §4.)

Kretzmann, N. (1966) 'Omniscience and Immutability', *The Journal of Philosophy* 63 (14): 409-21.(Influential presentation of the argument mentioned in §2 for the incompatibility of omniscience and immutability.)

Pike, N. (1965) 'Divine Omniscience and Voluntary Action', *Philosophical Review* 74 (1): 27-46. (Much-discussed version of the argument for freedom's incompatibility with divine omniscience.)

Plantinga, A. (1986) 'On Ockham's Way Out', *Faith and Philosophy* 3 (3): 235-69. (Careful and sophisticated discussion of the Ockhamist position discussed in §3.)

Stump, E. and Kretzmann, N. (1981) 'Eternity', *The Journal of Philosophy* 78 (8): 429-58.(Controversial but intriguing defence of the oft-maligned thesis of divine eternity.)

Stump, E. and Kretzmann, N. (1985) 'Absolute Simplicity', *Faith and Philosophy* 2 (4): 353-82.(The pre-eminent contemporary articulation of and argument for the claim that simplicity ought to be included among the divine attributes.)

Wierenga, E. (1989) *The Nature of God*, Ithaca, NY: Cornell University Press. (Technical but very rewarding examination of most of the issues discussed in this entry.)

Ontological commitment

A person may believe in the existence of God, or numbers or ghosts. Such beliefs may be asserted, perhaps in a theory. Assertions of the existence of specific entities or kinds of entities are the intuitive source of the notion of ontological commitment, for it is natural to think of a person who makes such an assertion as being 'committed' to an 'ontology' that includes such entities. So ontological commitment appears to be a relation *that holds between persons or existence assertions (including theories), on the one hand, and specific entities or kinds of entities (or ontologies), on the other.*

Ontological commitment is thus a very rich notion - one in which logical, metaphysical, linguistic and epistemic elements are intermingled. The main philosophical problem concerning commitment is whether there is a precise criterion *for detecting commitments in accordance with intuition. It once seemed extremely important to find a criterion, for it promised to serve as a vital tool in the comparative assessment of theories. Many different criteria have been proposed and a variety of problems have beset these efforts. W.V. Quine has been the central figure in the discussion and we will consider two of his formulations below.*

Many important philosophical topics are closely connected with ontological commitment. These include: the nature of theories and their interpretation; interpretations of quantification; the nature of kinds; the question of the existence of merely possible entities; extensionality and intensionality; the general question of the nature of modality; and the significance of Occam's razor.

1 Fundamental problems

The intuitive source of the notion of ontological commitment is people asserting things to exist. It may thus be thought of as a *binary relation* between persons or assertions of existence and the entities or kinds of entities they assert to exist. The main philosophical problem concerning commitment is whether there is a precise criterion for detecting commitments in accordance with intuition. Anyone who seeks a criterion of commitment must first decide what sorts of entities may properly stand in the positions of the relation. Unfortunately there are at least *five* significant candidates for occupying its first position: persons; acts of uttering (or inscribing); utterances (or inscriptions), that is, the actual 'tokens' produced; sentences uttered (or inscribed), that is, the 'types' of which tokens are produced; and propositions (which on some accounts are often expressed on occasions of utterance or inscription). And there is a similarly diverse range of candidates for the second relatum. So there is a daunting array of seemingly different ways to see commitment as a binary relation.

An encouraging thought about this proliferation is that it may not ultimately matter which approach one takes because, given any plausible choice, the relation may be thought of as holding *derivatively* between pairs of the unchosen sorts in a more or less natural way. It may also be that moving to a ternary (and so on) treatment would help absorb some of the complexity, but commitment is typically viewed as a binary relation, and we will follow that precedent here.

Another difficulty is that our intuitive talk of commitment is crucially ambiguous. Sometimes we speak as if a person who says there are whales is (thereby) committed to the existence of mammals, sometimes not. So, there is a sense of commitment that 'tracks entailments' and one that does not. Both senses are surely legitimate, but we should decide which we are discussing at the outset. If we distinguish thus between '*prima facie*' and 'all-in' commitment, then it is fair to say that the discussion provoked by W.V. Quine mainly concerns the former. This is not surprising given his famous antipathy to essential properties and analytic truth, as well as his explicit claims that the notion of commitment belongs to the 'theory of reference' (and not to the 'theory of meaning'). But we will see that some of Quine's formulations can seem more suitable for assessing all-in commitment than *prima facie*.

Quine's stated criteria are sometimes meant to reveal commitment to *specific* entities, sometimes to *kinds* of entities. We will focus on the latter. Here are two representative examples:

> (1) to say that a given existential quantification presupposes objects of a given kind is to say simply that the open sentence which follows the quantifier is true of some objects of that kind and none not of that kind.

(Quine 1953a: 131)

> (2) entities of a given sort are assumed by a theory if and only if some of them must be counted among the values of the variables in order that the statements affirmed in the theory be true.

(Quine 1953b: 103)

There is a serious problem of interpretation for any criterion intended to apply to theories. (This includes (1), for Quine's mention of open sentences following quantifiers reveals that he is thinking of certain sorts of formal theory.) The problem is that we use 'theory' in at least three different ways. First, we attribute 'informal theories' to people on the basis of their natural-language claims. Then we 'represent' these informal theories formally. Here the canonical notion is that of an (axiomatic) first-order theory along with an 'interpretation', which consists of a non-empty domain (the 'range' of quantifiable variables) plus an assignment of 'extensions' based on that domain to the nonlogical symbols of the theory. Note that we have already spoken of the purely syntactic component as a 'theory', but we also regard the syntax plus the extensional semantics as an 'interpreted theory'. So an unqualified use of 'theory' could mean any of three different things.

Now it seems almost nonsensical to think of a purely syntactic object (such as an uninterpreted sentence or theory) as ontologically committed to anything (except perhaps the syntactic elements that comprise it). Unfortunately this is the most literal construal of (1). Another construal makes the interpreted theory (or sentence, and so on) bear the commitment. This makes much better sense, but it may not make for a much better criterion. What (1) does under this understanding is, in effect, to define a certain metalinguistic relation term, 'presupposes' (see Presupposition). The relation this term expresses is evidently ternary - it relates interpreted closed sentences, objects (from the domain) and (somehow) *kinds*. The definition utilizes the notions of a formula with one free variable, the satisfaction relation (understood as relating objects and formulas - Quine's 'true of'), objects and (somehow) kinds. So (1) offers a new piece of terminology in the metatheory of interpreted first-order theories, enhanced by whatever is needed to talk about kinds in whatever way is intended. To this extent it is harmless enough, but it is not clear what it has to do with intuitive commitment. Notice that on its basis we may distinguish two sorts of objects in the domain of a theory: those that are of kinds 'presupposed' by at least one existential sentence; and, as may occur, those that are not. But any theory with both sorts of entities is already committed to all of them in a far more basic way since they are all essential components of that theory. No theory without them is the same theory.

It may therefore seem best to view (1) as directed to the case of informal theories, and the same may be said of (2). The intuitive idea is perhaps that the 'ontology' of an interpreted formal theory is 'transparent': It simply consists of the members of its domain of interpretation (or a readily distinguishable subset of them). So one might hope to reduce the hazy question of the commitments of informal theories to the transparent case of their formal representatives. But if the commitments of a given informal theory are unclear, then it would seem that the choice of its formal representative must also be unclear. So it appears that any criterion conceived in this way cannot work: to apply it we must first find a representing theory, but we cannot do this unless we already know the informal commitments.

The search for a criterion to apply to *theories* may thus seem fated to produce either near nonsense, triviality or practical futility, depending on what is meant by 'theory'. This would not be welcomed by Quine, especially since the original idea was to find a neutral means for assessing something not always clearly in view. It was to help us compare theories, sometimes with something like Occam's razor in mind. (Of course there may be a criterion for informal theories that makes no similar appeal to formal representatives.)

2 Problems of detail

We now consider some problems that arise from the details of (1) and (2). A serious defect of (1) is that it offers no means of assessing informal theories that make false existence claims. Since there are no unicorns, there is no interpreted theory the domain of which contains unicorns, and so there is none with unicorns available to satisfy relevant open sentences.

(2) offers an improvement over (1) in just this respect. For it may be seen as ruling that there *must* be some unicorns in the domain of *any* interpreted theory that would be a suitable representative of the informal theory, and

there may be a way of understanding this that makes no intuitive commitment to unicorns.

Now let us return to the matter of whales and mammals. It is clear that if the informal claim that there are whales receives a suitable (interpreted) first-order formulation, then a relevant open sentence will be true of some mammals and not true of any non-mammals. So (1) finds the informal theory committed to mammals as well as to whales. Whether (2) has the same consequence evidently depends on whether whales *must* be mammals, not (as in (1)) on whether they actually are mammals. So a philosopher who accepts (2) and, like Quine, denies that any entities have any non-trivial essential (or necessary) properties may claim that the distinction between *prima facie* and all-in commitment is a distinction without a difference (see Essentialism).

But this means there is a further difficulty in trying to evaluate Quine's formulations in the style of (2): they yield different results depending on the background philosophical views of the person applying them. When Quine applies (2) to an informal theory of whales, he concludes that it is not committed to mammals. But when an essentialist applies it, the conclusion is the opposite.

Formulations such as (1) and (2) reveal the influence (and importance) of a period that featured intense work in the areas of extensional formal semantics and the axiomatization of both scientific and philosophical theories. Such criteria are fairly characterized as seeing existential quantification as the source (and signal) of ontological commitment. They may be seen as differing elaborations of the theme expressed in Quine's famous slogan, 'To *be* is to be a value of a variable' (1939). Quine would later adopt a somewhat different perspective on the matter, the perspective of 'ontological relativity', which we will mention below. But there remain two very general criticisms of efforts in the quantificational spirit that merit separate discussion.

3 Two further problems

For a time the influence of Tarskian formal semantics was so great that it must have been difficult to imagine that any essentially different treatment of quantification could be defensible. But there is a coherent alternative called the substitutional interpretation. On this interpretation the semantics of a first-order language includes a non-empty set of terms of that language called the 'substitution class'. An existential quantification is counted true if and only if there is a member of the substitution class such that the result of removing the quantifier and replacing all resulting free occurrences of its variable by that member is true. So the truth-conditions for the quantifiers make no appeal to a domain of (potentially) extra-linguistic entities. Depending on how truth-values are assigned to atomic sentences - and one possibility is by sheer stipulation - a given substitutionally interpreted theory may have no intuitive ontological commitment beyond the constituents of the language itself, the truth-values, and perhaps some simple sets (including functions) of these entities.

The usefulness of substitutional quantification for representing informal theories may be debated (see Quantifiers, substitutional and objectual). One may ask whether the concept of truth embodied in a given substitutional theory appropriately matches the one implicit in a given informal theory. If the intuitive verdict is that the informal theory has non-trivial extra-linguistic commitment, then a substitutional theory lacking such commitment will be metaphysically inadequate (though it might get the truth-values right). But there remains the simple fact that the substitutional interpretation is perfectly coherent. From this it appears to follow that the mere assertion of an existential sentence does not provide a sufficient condition for non-trivial ontological commitment. So an adequate criterion of commitment should not presuppose that it does. The very existence and coherence of the substitutional interpretation thus creates a real difficulty for 'quantificational' criteria in the spirit of (1) or (2).

It has often been claimed, for one reason or another, that commitment is an 'intensional' notion (see Intensional entities §1). The main reason why (2) improves upon (1) is that it contains the word 'must'. But this suggests that the improvement is gained only at what Quine must see as great cost, namely, by exploiting the notion of *necessity*, which belongs to the theory of meaning. For it is natural to interpret the condition stated in (2) as meaning that it is necessary that if the statements affirmed in the theory are true, then objects of the required sort are values of its variables.

Others have found intensionality not in an appeal to necessity but in the (claimed) 'referential opacity' of the commitment relation itself, that is, in the resistance of its second position both to the substitution of coreferential expressions and to existential generalization. Commitment to Twain does not entail commitment to Clemens; nor does commitment to unicorns (*or* horses) entail there being entities to which one is committed. (Note that a

plausible claim of opacity can only be made concerning *prima facie* commitment.)

Here is a further (but related) reason to suspect intensionality. Suppose we try to state a criterion in which the condition expresses a binary relation between persons or theories (and so on) and something else. What are the plausible candidates for the second relatum? Consider first the theory that simply says there are horses. Well, there *are* horses, but the theory is not committed to any of them (either individually or collectively). One natural thought is that the second relatum is the property (or kind), being a horse. Accordingly, to say that the theory is committed to horses would mean that it cannot be true unless the property is instantiated. But properties are intensional entities. (In the case of commitment to specific entities, 'possible objects', including *merely* possible objects, are natural candidates analogous to properties.) There are other possibilities, but it has been argued that any adequate relational treatment must be intensional.

Yet another approach sees theories not as linguistic or quasi-linguistic entities, whether formal or informal, but rather as propositions (or sets of propositions). Then a theory is committed to, say, unicorns if it entails the proposition that unicorns exist. This approach is intensional because propositions are intensional entities. (It also appeals to entailment, which is usually (though not universally) thought to rest on necessity.)

So the suspicion of intensionality may arise for a number of reasons. But the intensionality always seems to flow from one of two somewhat related sources. The first is the classical problem of explaining how we can either assert or deny the existence of things that in fact do not exist. The other is that we are treating *prima facie* rather than all-in commitment. The (claimed) intensionality of commitment is not in itself disturbing, but it was vexing for Quine because of his general aversion to the intensional.

4 Translation and triviality

The possibility that an adequate criterion of commitment would somehow be 'trivial' arose in §1. Remarkably enough, it also surfaced in Quine in the wake of his important doctrines of 'ontological relativity' and the 'inscrutability of reference'. Details of these positions cannot be discussed here, but their effect on the matter of a criterion should be mentioned.

The key effect is to cast the question of a criterion as one concerning *translation*. Theories are stated in languages. An ontological criterion must also be stated in a language. For our purposes, the language of a criterion may be viewed as a technically enhanced version of an appropriate fragment of English. Call such a language E^*. Now it seems clear that if the question whether a theory T is committed to entities of kind K can be settled at all, then it can be settled as follows: simply translate the assertions (or theorems) of T into E^* and see whether the results include 'There exist objects of kind K' (where 'There exist' is understood referentially).

For reasons flowing from the aforementioned doctrines, there is a sense in which Quine does not regard this as 'settling' the question of T's ontology, because he thinks 'There exist objects of kind K', in E^*, is itself ontologically indeterminate. But he also thinks that this is an indeterminacy that cannot be removed and should not be lamented.

Any such 'translational' criterion can easily seem trivial. For either we already know the commitments of T, and so have no need for it, or else we do not, and so cannot apply it. Yet it seems perfectly natural and correct independently of the Quinean doctrines and, in particular, independently of their implication that it does not settle the question of a theory's commitment.

See also: Ontology; Free logics, philosophical issues in; Logical and mathematical terms, glossary of

MICHAEL JUBIEN

References and further reading

The following items all involve complex argumentation and at least some technicality. Many contain valuable further references.

Cartwright, R.L. (1954) 'Ontology and the Theory of Meaning', *Philosophy of Science* 21: 316-25.(The classic statement of the idea that ontological commitment involves the notion of necessity, as discussed in §3.)

Dunn, J.M. and Belnap, N.D. (1968) 'The Substitution Interpretation of the Quantifiers', *Noûs* 2: 179-85.(A discussion and technical presentation of substitutional semantics, as discussed in §3.)

Jubien, M. (1972) ‘The Intensionality of Ontological Commitment’, *Noûs* 6: 378-87.(Argues that any adequate criterion of commitment that treats the notion as a relation must be intensional, as mentioned in §3.)

Kripke, S.A. (1976) ‘Is there a Problem about Substitutional Quantification?’, in G. Evans and J. McDowell (eds) *Truth and Meaning*, Oxford: Clarendon Press, 325-419.(Argues that attempts to see substitutional quantification as covertly carrying the commitments of referential quantification are incorrect. Very detailed and technical in places. See the present §3.)

Marcus, R.B. (1972) ‘Quantification and Ontology’, *Noûs* 6: 240-50; repr. in *Modalities*, Oxford and New York: Oxford University Press, 1993, 75-87.(A classic statement of the ontological neutrality of substitutional quantification, as discussed in §3.)

Marcus, R.B. (1978) ‘Nominalism and the Substitutional Quantifier’, *Monist* 61: 351-62; repr. in *Modalities*, Oxford and New York: Oxford University Press, 1993, 111-24.(Explores the congeniality of substitutional quantification for various forms of nominalism.)

Quine, W.V. (1939) ‘A Logistical Approach to the Ontological Problem’, presented at the fifth International Congress for the Unity of Science, Cambridge, MA; repr. in *The Ways of Paradox and Other Essays*, New York: Random House, 1966, 64-9.(An early discussion of naming, reference and ontology. The source of ‘To *be* is to be a value of a variable’ mentioned in §2.)

Quine, W.V. (1953a) ‘Notes on the Theory of Reference’, in *From a Logical Point of View*, Cambridge, MA and London: Harvard University Press, 1961, 130-8.(The source of criterion (1), discussed in §§1-3, and of an explicit claim that commitment belongs to the theory of reference.)

Quine, W.V. (1953b) ‘Logic and the Reification of Universals’, in *From a Logical Point of View*, Cambridge, MA and London: Harvard University Press, 1961, 102-29.(The source of criterion (2), discussed in the present §§1-3. Contains an explicit claim of the need for a criterion of commitment.)

Quine, W.V. (1968) ‘Ontological Relativity’, *Journal of Philosophy* 65: 185-212; repr. in *Ontological Relativity and Other Essays*, New York and London: Columbia University Press, 1969, 26-68.(An early statement of the doctrine of ontological relativity and related matters, as mentioned in §4.)

Quine, W.V. (1981) ‘Things and Their Place in Theories’, in *Theories and Things*, Cambridge, MA and London: Belknap Press, 1-23.(More on ontological relativity, including the idea that the question of the ontology of a theory is fundamentally a question concerning translation. See the present §4.)

Scheffler, I. and Chomsky, N. (1958-9) ‘What is Said to Be’, *Proceedings of the Aristotelian Society* 69: 215-24. (The classic statement of the idea that the commitment context is referentially opaque; discussed in §3.)

Schwartz, R. (1993) ‘On "What is Said to Be"’, *Synthese* 94: 43-54.(A contemporary attempt to capture a concept of commitment using only extensional resources. See the present §§3, 4.)

Ontology

The word 'ontology' is used to refer to philosophical investigation of existence, or being. Such investigation may be directed towards the concept of being, asking what 'being' means, or what it is for something to exist; it may also (or instead) be concerned with the question 'what exists?', or 'what general sorts of thing are there?' It is common to speak of a philosopher's ontology, meaning the kinds of thing they take to exist, or the ontology of a theory, meaning the things that would have to exist for that theory to be true.

1 Existence and being

Since so many central debates of philosophy concern what types of things exist, scrutiny of the arguments used in them was bound to lead to investigation of the concept of existence and its logic. The most famous case of this kind is the Ontological Argument for the existence of God , and one of the most famous moves in the ensuing debate is Kant's claim that existence is not a property or predicate of existing things: 'cats exist' clearly tells us something, but it does not tell us of things which, in addition to being furry, feline and fleet of foot, have the further property of existence (see God, arguments for the existence of §§2-3; Existence).

This point, with which modern logic agrees, may teach us to formulate 'cats exist' as 'there are things which are cats' - in which it doesn't even look as if existence is functioning as a predicate. But many will think that this does not take us very far. It does not tell us how to describe the difference between a world in which cats exist and one in which they do not - other than by repeating the formula 'there are cats'; it will not advance our understanding of what it is for something to exist. Nor will it touch the somewhat dizzying question 'why does anything exist, rather than nothing?'.

Some philosophers, however, have taken an interest in existence or 'being' (as it tends to appear in their works or in the English translations), not because it appears as part of so many philosophical claims but because they take it for the central concept of philosophy. The most prominent examples are, in antiquity, Parmenides, and in the twentieth century Martin Heidegger.

2 What is there?

In its characteristically philosophical form, this is not a question of detail (for example, are there mammoths?) but about the most general *kinds* of thing: are there universals, or only particulars? - is there mind or spirit, or is there only matter? - is there anything that exists without being in space and time? Thus the debate on the first of these questions between Platonists and nominalists, or on the second between idealists and materialists, might in each case be described as a difference of opinion about the correct ontology. So might the conflict over whether values are objective aspects of reality, or rather 'in the eye of the beholder', a matter of how we react to things rather than the things themselves (see Emotivism; Projectivism).

The questions 'What kinds of thing *ultimately* exist?' or '… *really* exist?' or '… exist *in themselves*?' are even more characteristically philosophical forms of the general ontological question. To understand what usually lies behind these additional terms one needs a grasp of (1) the concept of a *reduction* and (2) the distinction between appearances and things-in-themselves.

Reduction. Berkeley famously claimed that material objects were just collections of 'ideas' (see Berkeley, G. §3). He did not mean that there were no chairs or tables, but that such things did not have a material, non-mental component; what really existed was all mental. There are spirits and their ideas, and we speak of chairs and so on when the latter occur in familiar, stable groupings. In modern terminology, he was claiming that material objects can be *reduced to* ideas. There are other common examples. In political discourse we often speak of what a particular state has done - but without having to suppose that there are such things as states distinct from the individual people who compose them. A once-popular thesis about the nature of mind was that there is nothing but bodies and their behaviour, and that words apparently naming mental states and happenings are just convenient ways of indicating types of behaviour (see Behaviourism, analytic; Reductionism in the philosophy of mind).

Things-in-themselves. We may distinguish between the way a thing appears, which will depend partly on the faculties and situation of whoever is perceiving it, and the way it is, independently of how anyone perceives it. The

latter is the thing-in-itself. The terminology was made instantly famous by Kant, arguing that space and time (and therefore everything in space or time) were merely the way in which a non-spatiotemporal reality, things-in-themselves, appeared to humans (see Kant §5).

3 Ontological commitment

The notion of ontological commitment has come to prominence in the second half of the twentieth century, mainly through the work of Quine (see Quine, W.V. §5). On Quine's view the right guide to what exists is science, so that our best guide to what exists is our best current scientific theory: what exists is what acceptance of that theory commits us to.

But what is that? How do we determine what existents the acceptance of a given theory commits us to? Quine proposes a criterion, often summarized in the famous slogan 'to be is to be the value of a variable'. We are to see what types of thing are quantified over when the theory is stated in canonical form with predicate calculus as the underlying logic; the theory's ontological commitment is precisely to things of those types. This line of thought has given rise to much discussion (see Ontological commitment).

See also: Abstract objects; Being; Idealism; Ontology in Indian philosophy; Realism and antirealism; Universals; Value, ontological status of

EDWARD CRAIG

References and further reading

Heidegger, M. (1953) *An Introduction to Metaphysics*, trans. R. Manheim, New Haven, CT: Yale University Press, 1959.(Heidegger's understanding of being. Readers new to Heidegger may not find this as introductory as the title had led them to hope.)

Kant, I. (1781, 1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1929.(For the famous passage about existence not being a predicate, A598/B626-A601/B629; for the distinction between appearances and things-in-themselves a good example is A45/B62-A46/B63.)

Nozick, R. (1981) *Philosophical Explanations*, Oxford: Oxford University Press.(Chapter 2, 'Why is there something rather than nothing?' is at times a little bewildering, but good fun - for the more experienced reader.)

Quine, W.V. (1948) 'On What There is', in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1953.(Classic paper on existence and ontological commitment. First published in the *Review of Metaphysics*.)

Williams, C.J.F. (1981) *What is Existence?*, Oxford: Clarendon Press.(Extensive discussion of all the issues arising out of the 'logic and language' - rather than the 'deep metaphysics' approach to existence.)

Ontology in Indian philosophy

All Indian philosophical traditions are deeply engaged with ontology, the study of being, since clarity about the nature of reality is at the heart of three intimately connected goals: knowledge, proper conduct and liberation from the continued suffering that is part of all human existence. The formulation of a list of ontological categories, a classification of reality by division into several fundamental objective kinds, however, is less widespread. There is little room for a doctrine of distinct, if related, ontological categories in a philosophical school that takes reality as one, even less if that one lies beyond description. If the phenomenal world is but illusory appearance, as, for example, in the Vedānta of Śaōkara, then a determination of kinds of entities does not recommend itself as a means to adequate analysis of the world. Even the Sāōkhya tradition's realism reduces the world to an evolution from two fundamental entities, spirit and matter. Categories make sense within the context of a pluralistic realism, an analysis of the world that finds it to be composed of a multiplicity of real entities. Such a view is found to some extent in Jaina philosophy, but is primarily defended and developed in the Nyāya-Vaiśeṣika school.

The Nyāya-Vaiśeṣika categories are seven: substance, quality, motion, universal, particular, inherence and not-being. While all are understood as real entities and objects of knowledge, substance is most fundamental as each of the others in some way depends on substance. Substances are nine: earth, water, fire, air, ether, time, space, self and mind. The first four are atomic: they may combine to form macroscopic substance, such as a clay pot, but in incomposite form they are indestructible atoms, as are the last two. Ether, time and space, likewise indestructible, are unitary and pervade all. In its irreducible parts, all substance is eternal; every composite whole is a destructible substance.

A relation of containment, called inherence, structures the categories. The qualities, actions and universals by which we might characterize a pot inhere in it. They are distinct entities from the pot, yet cannot exist apart from their underlying substrate. Composite substances like a pot are also contained in their parts by inherence, but the smallest parts, eternal substances, exist independently as receptacles that contain nothing. A whole, greater than the sum of its parts, is said to inhere in the parts while the parts are the inherence cause of the whole.

Eternal substance, the ultimate substrate of all, is a bare particular. An entity that is nothing but a receptacle for other entities, it furnishes criteria for separability and individuality, but cannot be defined in itself apart from others. This aspect of the concept of substance leads later Nyāya-Vaiśeṣika into extensive analysis of relations and negation.

1 Categories

While ontological speculation lies at the heart of all schools of classical Indian thought, a great portion of the tradition is not conducive to a doctrine of ontological categories. Categories provide a means to analyse the world according to the distinct kinds of entities that are the ingredients of all reality. The tradition growing out of the philosophical treatment of the Vedas in the Upaniṣads, in which all questions of ontology press towards the transcendent and the being of the phenomenal world is doubted and rejected as an illusion, ought to have limited use for such a tool. The arguments of Buddhist thinkers taking all phenomena as the only existence, a flux with no underlying reality, also leave no place for a classification of discrete kinds of realities. A theory of categories develops not in an idealism or nominalism, but presupposes some kind of pluralistic realism.

The most extensive system of ontological categories in classical Indian thought was worked out over the course of several hundred years by the Nyāya-Vaiśeṣika school. Traditionally this begins with the *Vaiśeṣikasūtra* of the legendary Kāṇāda, followed by the commentary of Vātsyāyana (fifth century AD) and the work of Praśastapāda (sixth century). The resulting most widely accepted list contains seven categories, the first three of which each comprise a standard list of members, as follows: (1) substance, which includes earth, water, fire, air, ether, time, space, mind and self; (2) quality, which includes colour, taste, smell, touch, number, measure, separateness, conjunction, disjunction, priority, posteriority, cognitions, pleasure, pain, desire, aversion, volition, gravity, liquidity, viscidity, disposition, merit, demerit and sound; (3) action, which includes throwing upwards, throwing downwards, contraction, expansion and going; (4) universal; (5) particularity; (6) inherence; (7) absence or nonexistence.

This list was appropriated in part by other schools, notably Jaina thinkers, who took substance as an all-inclusive category that is further subdivided and concurred to an extent with Vaiśeṣika in positing real, individual souls. The Jaina tradition could be viewed as an expansion of Sāōkhya dualism: in addition to conscious but formless reality and nonconscious material reality, it recognizes a nonconscious but formless reality (see Jaina philosophy §1; Sāōkhya §2). This is a doctrine of substance, classifying three kinds of permanent realities, and not a doctrine classifying appearances that are nothing in themselves, but are derivative manifestations of a transcendent reality. In addition to conscious reality, souls, and nonconscious matter, atoms and their aggregations, there are four kinds of nonconscious yet immaterial substance, each a medium for change. These are the media of motion, rest, space and time. All reality is identified with substance, which is seen as unifying permanence and impermanence. This all-inclusive category of substance is designed to explain how an entity can remain a real, self-same individual throughout change. Mīmāṃsāalso makes use of the notion of substance. In general, however, where categories appear in Indian thought outside the Nyāya-Vaiśeṣika tradition, they are ancillary tools, not the root of ontology. Such is the view of Vedāntin compendia of philosophy - short works that classify and summarize other schools - in considering Nyāya-Vaiśeṣika early and primitive, containing useful instruments for analysis but no insight into truth.

In the philosophical vocabulary of European languages, ‘category’ is a derivative of the Greek word first appropriated for philosophical jargon by Aristotle (§7). *Kategoria*, from the verb *kategorein*, literally ‘to speak in the marketplace’, entered the language of jurisprudence meaning ‘to accuse’, that is, to say something about someone. Aristotle adopted it for a classification of the different ways to say things about other things, a classification of kinds of predicates. A philosophical claim lies behind Aristotle’s method. If language is thought reflecting what there is, then the structure of language is the key to the structure of reality. The Indian term for ‘category’, *padārtha*, the object or meaning (*artha*) of a word (*pada*), suggests a related notion. Though it is not in the Vaiśeṣika texts, some Indian grammarians drew a connection between the first three categories, substance, quality and action, and the distinction between noun, adjective and verb. Certain Vaiśeṣika arguments do rely explicitly on word meaning, claiming that words could have no meaning if universals were not real entities.

All seven categories are all mentioned in the *Vaiśeṣikasūtra*, although the word we translate as ‘category’, *padārtha*, occurs only once, in a disputed verse where it designates the first six. The term does not become clearly established until Praśastapāda’s commentary on the *Vaiśeṣikasūtra*, the *Padārthadharmasaōgraha (Compendium of Properties of the Categories)*. The categories are not only the meaning or object of words; they represent a classification of all being, a definitive listing of the various kinds of things that exist. Each category represents a distinct kind of entity.

The reality of each category rests on three aspects: individual unity, causality and independence. (1) Whatever is a real entity is a separate individual, knowable as such, even if it cannot exist apart from something else. (2) Each category of thing is known as a cause: substance causes substance, quality quality, action action, universal and particularity cause cognition, particularity in an ultimate sense causes the individuality of eternal substances, and inherence causes subsistence in. (3) Most individual realities are dependent on or contained in something else, with the chain of dependence going back to independent eternal substances.

This does not yet give much clue as to what each category is. That task requires patience, since each category, while separate, is largely defined through relations to other categories.

2 Relations

Despite the emphasis on the particularity (*viśeṣa*) of the individual, the source of the name Vaiśeṣika, none of the categories can be adequately grasped without reference to others. Śrīdhara (tenth century) writes that the categories ‘become objects and properties with reference to one another’ (*Nyāyakandalī,* comment to III.1). In Vaiśeṣika all relations obtain between two entities. Most significant to understanding the structure of the categories is the relation called inherence. Itself a category, inherence informs the nature of all categories. There are two central aspects to its definition. It is a permanent relation, in contrast to the relations of conjunction and disjunction. The conjunction of two objects - two pebbles in contact, say - can soon be dissolved. Inherence, on the other hand, is a relation that persists until one or both of the relata are destroyed. Furthermore, inherence is explained by the metaphor of containment. In an inherence relation, one of the individuals is always contained by the other, which is its container. The contained is said to inhere in the container.

The relation of contained to container is the primary example of a relation of superstratum to substrate. (It is not, however, the only one. Other relations, including conjunction when one object is on top of another, are also of this general sort; see §9.) Common synonyms for the term ‘substrate’ (*ādhāra*) include ‘locus’ (*adhikaraṇa*) and ‘abode’ (*āśraya*).

3 Substance

The realism of the entire Nyāya-Vaiśeṣika system is rooted in the category of substance (*dravya*). It is the one category that comprises a plurality of independently existing individuals. Some of the others are dependent on substance as their substrate (quality, motion, ultimate particularity), some are dependent on cognition (universal, particularity), and inherence is a relation of other categories. Unlike a number of the categories, for which a variety of terms have been attempted as translations, *dravya* has been rendered virtually without exception as ‘substance’, from the Latin translation for the first of Aristotle’s categories. Beyond its technical usage in Vaiśeṣika, *dravya* can mean in Sanskrit any thing or object, a material or ingredient, thus medicine, also wealth or money, or, more broadly, any object of possession.

Substance is the basic stuff of which all things material and immaterial are made; it is the ultimate possessor. The inherence relation of containment manifests the ontological priority of substance. Qualities, actions, universals and particularities exist only as contained in something else. Substance is also self-productive, meaning that only substances can cause the existence of other substances. Conversely, substances can be contained only in other substances. When a substance comes into being, it is produced from the unification of previously existing substances, as when two clay halves are joined to form a pot. The pot is said to inhere in the halves; the parts contain the whole. The parts of a substance can of course be contained in simpler parts, but because this chain of containment cannot continue *ad infinitum*, there must be simplest substances. Since a substance exists as contained by being produced from other substances and the simplest atoms are contained in nothing simpler, the latter are eternal. Composite substances are dependent and transient, ultimate atoms independent and permanent.

Since impermanent substances arise through the physical union of prior substances, only material substances can become and pass away. Physical substances have an eternal form as well: atoms of earth, water, fire and air are many and indestructible. The material atoms are, in a sense, contained in something other, although this is not considered inherence. Ether, space and time are three substances that are said to be one, not many, eternal, all-pervasive and a receptacle for corporeal things. The soul or self (*ātman*), of which there are many, is immaterial and therefore eternal. Mind, countless atoms of which populate the world, is distinct from soul. It is understood as an unconscious instrument that transmits perception to the soul. Since this occurs through contact and mind possesses action (motion), it is said to be material, although it never combines to produce composite substances as the other physical substances do; it is material but not ‘tangible’. Mind possesses qualities, such as number, dimension and separateness, which are attributed to both material and immaterial substances, but Praśastapāda also gives it priority and posteriority, which are elsewhere said to inhere only in material substances, as well as ability, which is otherwise thought to inhere only in immaterial substances.

Not only mind, but the other kinds of substance are distinguished from one another by the qualities that can inhere in them. Consciousness and other mental attributes are not identical to soul, but are qualities that the soul can possess. The physical substances together with ether are distinguished in that each is the receptacle for a particular kind of quality, associated with a particular sense. Earth is the substrate for smell, water for taste and natural fluidity, fire for colour, air for touch, and ether for sound. Where a quality is associated with a specific substance, it is the mark (*lakṣaṇa*) of that substance. The mark is a cognitive mark, allowing us to recognize substances, but is not an essence (*svarūpa*) identical with the nature of the substance (see §9).

Although ultimate, eternal substances are the only truly independent individuals, all instances of every category are interpreted as real, separate entities, even if dependent. Even nonexistences are entities, though having no positive nature of their own. Caused substances are wholes, born with the conjunction of the substances that make up their parts. (The parts, not the conjunction, are interpreted as the cause of the whole; see §7.) Despite the dependence on its parts, the whole is no less an entity unto itself than are its parts, the parts of its parts, or the ultimate atoms. Just as every quality or action in a substance is logically separable and unique, and therefore a real entity, so too are composite substances real individuals.

Against the traditions that hold the phenomenal world to be an illusory manifestation of a nondual transcendent reality, Vaiśeṣika ontology sides with theories of causality called *asatkāryavāda*, which deny the pre-existence of an effect in its cause (see Causation, Indian theories of §§1, 5). Each caused thing is a new thing. In this, there is agreement with the Buddhist schools, the major philosophical opponents of the classical Nyāya-Vaiśeṣika thinkers. They clash where the Buddhists assign the novelty of all effects to the momentariness of all existence, while a system of ontological categories like the Vaiśeṣika system is a means to justify permanent identity of individual entities through time (see Momentariness, Buddhist doctrine of). Substances, taken as substrates of qualities and other categories, permit an explanation of how a thing can remain one identical individual while undergoing change. A rock rolls down a hill, but is the same rock; the motion inheres in the rock but is not identical to it. A soul can acquire knowledge it did not always have or experience coming and passing pleasures and pains, all while remaining the same individual.

Nevertheless, the permanence of individuals must be qualified. Nothing exists if not in a substance, and this substrate of all that really exists exists either as an atom or as a unity composed of atoms. The atoms are uncreated and indestructible, but any change in the composition of a caused substance produces a new substance. If a clay pot is chipped, or even if a few molecules are brushed from its surface, the substance that was the original pot is destroyed as surely as if smashed with a sledgehammer.

This is reflected in the Vaiśeṣika treatment of particularity or individuality. Individuality is the category designating what distinguishes individuals as separate entities. It inheres in substance, yet does not inhere in any composite substance such as ‘this sherd’. A pure and simple individuality can reside only in an ultimate substance, an atom. It is not needed to verify the separate existence of a composite substance, or a quality or action, because the individuality of a whole and of an attribute is demonstrated by containment in a substrate, and therefore by the relation of inherence. At the lowest level, however, a substance can inhere in nothing. Here, indeed, individuality is an indispensable concept. For atoms and other eternal substances are the necessary building blocks of the individuality of all other things. As the ultimate abodes of individuality, the atoms must be distinct from all they contain. Therefore, no attribute of an atom can serve to show what it is as distinct from the next atom. As ultimate, the atoms are a plurality of indistinguishables, unable to be distinguished by themselves, but only by the necessary inherence of individualities.

The attempt to identify substance leads Nyāya-Vaiśeṣika into a dilemma. Pure independence is what it is to be real. An individual quality or an action is one because it is in a substance; substances that are whole are one, because the parts they are in are substances; the smallest parts, the atomic substances, are one because they are independent from all else. As has been seen, this adds up to the ontological priority of eternal substances, given that to be is to be a separable individual. But by the same token it becomes very difficult to state what a given atom is - it cannot be identified with any of its predicates, for it is their substrate and an individual distinct from them - and therefore difficult to say how the given atom is in itself distinct from the next given atom. What is the substance apart from what it has? There seems to be no way to say, and yet there must be if knowledge as distinct from error is knowledge of the real. Consequently some thinkers in the Nyāya-Vaiśeṣika tradition are led to posit obscure ways of directly perceiving a substance apart from whatever inheres in it. The cow barely visible in the shadows at night yet distinctly a cow is said to be a perception of the substance by itself, because its attributes, specifically its true colour, are not perceived. Such arguments, unsurprisingly, were ridiculed by rival schools. In the same vein, Śrīdhara affirms that yogins of developed skill directly intuit individual atoms. If the positive determination of individuals as separate is tenuous at best, one is left with distinguishing entities negatively by reference to relations with other individuals. As Nyāya-Vaiśeṣika thought develops, negative relations take a place of increasing importance (see §10).

4 Qualities and their inherence in substance

Quality (*guṇa*) and action or motion (*karman*) are the two categories in addition to substance that are considered objects (*artha*). Praśastapāda also expresses this by saying that ‘beingness’ applies to all categories, but ‘being’ as a general term only to substance, quality and motion, while the remaining three have ‘self-being’, their being in themselves (*svabhāva*; at Praśastapāda’s time, nonexistence was not yet an acknowledged category).

Like actions, qualities exist as individual realities, not independently but inhering in substances. Some are causes of other qualities and each contains the universal ‘qualityness’. The number of qualities specifically listed varies

slightly in different texts. Praśastapāda lists those occurring in material substances as colour, taste, smell, touch, priority, posteriority, gravity, fluidity, viscidity and speed. Ten qualities inhere in immaterial substances only: intellect, pleasure, pain, desire, aversion, effort, virtue, vice, ability and sound. Those remaining occur in both: number, dimension, separateness, conjunction and disjunction.

A number of the qualities are commonly called aids to perception. Though not identical to substance or the essence of substance, the qualities allow us to distinguish substances as separate kinds and as separate individuals (see §§3, 9). Qualities (as well as actions) have another important function with respect to the discernment of individual entities: they allow us to identify real and enduring individuals. Substances can change with respect to quality or action and yet remain the same substance. This was a central aspect of Aristotle's categories: substance provided a means to explain stability within change. The degree to which the Vaiśeṣika categories provide this is, however, in comparison with Aristotle, severely limited. The restrictions on identity through change are evident in the ways qualities act as causes and appear as effects.

Certain qualities are uncaused. These are the qualities that are 'marks', such as the smell of earth, which are coeval with permanent substance. Caused qualities are produced in one of five ways: by antecedents like qualities, conjunction, disjunction, action or by 'cooking' (*pācaka*), which is defined as hot touch or contact with fire. Action causes the qualities of conjunction, disjunction and speed. Conjunction and disjunction are among the several qualities (along with number and separateness when greater than one) that are relations and are therefore said to inhere in more than one substance at a time, thus not actually distinguishing individuals. Disjunction is said to cause only sound and secondary disjunction.

Of the three remaining causal modes, conjunction in two forms accounts for identity of changing substance. A conjunction is the contact of substances, without those substances thereby forming one (although conjunction occurs wherever a composite substance is produced). As a cause of qualities, conjunction other than cooking (cooking being conjunction of fire with another substance) is primarily associated with the immaterial qualities, that is to say, the attributes of soul. Understanding, pleasure, pain and other mental qualities can come to be contained in the soul or leave the soul without altering the real identity of the individual soul.

Cooking is in fact a specific kind of conjunction. Praśastapāda does not give a list of qualities caused by cooking, as he does for the other species of qualitative cause. Yet it seems clear that cooking is efficacious wherever a material substance undergoes change of quality without being destroyed. Praśastapāda mentions cooking as a cause explicitly only in connection with one substance, earth. Whether other substances can undergo qualitative change by hot touch is left ambiguous. Also, while a change in the odour or taste of earth is mentioned as a possible effect of cooking, colour is the most frequent example. Thus, a clay jar fired is the same substance as the greenware, though its colour has been changed.

'Cooking' may strike the reader as odd, yet it has a crucial explanatory function in the categories. Each quality is one entity, just as much as each substance is one. Containment in the substance, however, guarantees the unity of the quality. Whenever there is a change in the underlying substance, a new and distinct quality is born, and, apart from the few instances of 'cooking', any change in quality is a change in underlying substance (see §6).

5 Action

Action (*karman*, also translated 'motion'), the third category and the last recognized as an 'object', is significant in the causal mechanism of conjunctions and disjunctions, two qualities that, in turn, are central in the explanation of qualitative change. The actions are attributes that exist only momentarily and only in material substances. Divided into five kinds - throwing upwards, throwing downwards, contraction, expansion and motion (*gamana*, also translated 'going') - actions may arise from gravity, fluidity, conjunction or effort. The first three qualities productive of actions all reside in material substances. The quality of effort, volitional dispositions of the soul, establishes the link of action with consciousness. The category nevertheless does not overlap fully with a usual notion of agency, being always rooted in material reality. Thus all action occurs as some kind of physical motion. The divisions within this category were always a subject of debate even within the Nyāya-Vaiśeṣika tradition. If all actions are movements of material substances, then why not take the first four simply as various instances of the fifth, motion? Each of the two pairs of opposites is, unlike motion, associated with a definite direction. It was also asserted that these directed actions may be produced by conscious effort, unlike motion, which may occur in

the absence of consciousness. Thus the kinds of action serve to distinguish intentionally caused movement from blind events.

Just as the list of actions was disputed, so too was the status of action as a category distinct from quality (notably by Raghunātha in the sixteenth century). In fact the category of action furnishes the basis for the changing object. Although not all qualities are permanent or coeval with their substrates (some change while the underlying substance retains identity), all are nevertheless static. In the firing of a clay jar, one colour replaces another. When conjunction, a quality closely associated with actions, occurs, it is one unitary object that inheres in two substances. Action, on the other hand, is dynamic, and, while inhering in substances just like qualities, is momentary. Unlike qualities, which can produce other qualities, an action never produces another action. So in a chain reaction of ricocheting balls, to take an example, a conjunction of two substances produces an action, which terminates when the next contact is made, which conjunction in turn produces another action, and so forth.

6 Universal and particularity

The universal (*sāmānya*) and particularity (*viśeṣa*, also translated 'individuator' or 'describer') inhere in other entities and allow us to group and distinguish them. A quality is distinguished in name from a universal by the addition of an abstract suffix to designate the universal. 'Red' is a quality, but 'redness' is a universal. The strict separation of qualities and motions from the substances that contain them makes universals indispensable. Qualities and actions are individual - the red contained in one substance is by its inherence a separate thing from the red contained in another - and consequently qualities (as well as actions and substances) cannot be grouped or related in virtue of what they are in themselves. It is another side of the dilemma with substance: it is something other than its attributes, yet it is not possible to state what it is other than its attributes.

This dilemma is the threat of nominalism, and to some extent the two categories of universal and particularity ward off nominalistic objections to the realism of the categories. Such objections arose in the extensive debates with Buddhist thinkers, who argued that all universals were mental constructs at most (see Nominalism, Buddhist doctrine of). The Buddhist claim in fact does not appear incompatible with the ambiguous assertion of the *Vaiśeṣikasūtra* that both universal and particularity are 'with regard to the understanding' (*buddhyapekṣa*). Later commentaries said that the universal is the cause of the cognition of inclusion, and the particularity the cause of the cognition of difference, while insisting that these categories, like the others, are real in themselves, not a mental product.

The universal may be defined as that which inheres in a plurality of objects and brings about the idea of itself in any one. The universal pervades all its objects, having the same form in all in which it inheres. In this it is distinct from qualities. Thus the blue of a blue substance is one individual, coextensive with and momentarily posterior to its underlying substrate. (The quality is produced momentarily after the substance is produced, since it is caused by the substance; to make it simultaneous would risk rendering the quality and substance identical.) It inheres in only one thing, and is itself the locus of inherence. The universal blueness, however, as opposed to the quality blue, inheres as one in the many distinct and individual qualities that are all correctly called blue. The inclusiveness of the universal is said to distinguish it from its substrate. Thus, for example, substanceness is not identical with substance, since the latter is known by exclusion, by reference to what it is not. A further conclusion drawn is the eternality of universals, since, as independent from their substrates, universals are not produced by their substrates and so do not go out of existence with the destruction of the substrate.

Unlike the quality that it is in, nothing is in the universal. Since it is not a locus of inherence, the universal (like the particularity and the category of inherence) is not an object. Even with the commitment to the reality of universals, there is a sense in which they do not exist. One reason for this is a regress problem. The most general of universals, being, inheres in all objects, that is to say in all substances, qualities and actions. If 'being' were to inhere in universals as well, and universals could therefore be substrates of inherence, then 'being' itself would be an object, requiring a yet more general kind of 'being' to inhere in it, and so forth *ad infinitum*.

Universals not only allow us to include, but also to distinguish. 'Blueness' makes it possible to group all blue entities as one in respect of colour, but it also distinguishes from other colours. For this reason the universals are divided into 'higher' and 'lower'. Of the higher there is only one, 'being', since it is the only universal that is purely inclusive, all objects being its substrates. All other universals allow discrimination, and so cause both the

cognition of inclusion as well as the cognition of particularity. All universals more specific than ‘being’ are therefore called ‘particular universals’ (*sāmānyaviśeṣa*). Any cognition that involves inclusion due to similarity of form, however, is properly called a universal, not a particularity.

Particularities, the cause of the cognition of distinction, like universals inhere eternally in substrates without themselves being inhered in. Unlike universals, however, each particularity inheres in only one substrate. Since they are eternal, and since they have only one substrate, that substrate can only be an eternal substance: atoms, ether, time, space, soul or mind. The atomic substances are characterized by nothing in themselves that allows their discrimination, and therefore particularity is ascribed as the cause of knowing atoms to be distinct, a distinction perceptible to the yogin alone.

Despite the seemingly unclear ontological status of universal and particularity, they are crucial concepts in the Vaiśeṣika system of categories. As has been seen, there is nothing about a substance itself, which is presented as a bare receptacle, a harbour for attributes, to allow us to say what the substances, the foundations of reality, are in themselves. This function falls largely to these categories, and is defended on the grounds that, apart from the reality of universals, there would be no unity of individuals, and consequently no object denoted by words, which would entail the vacuity of all categories.

7 Inherence

Inherence, perhaps the least accessible of the categories, is nevertheless essential both to the organization of the Nyāya-Vaiśeṣika categories and to the ontological priority of substance. This category represents a special kind of relation. It is the permanent and indissoluble connection of two entities which are knowably distinct. The relation of inherence also indicates an order of dependence, being a relation of container to contained. In virtue of the permanence and the order, inherence is distinguished from the other kinds of relation, the pair ‘conjunction’ and ‘disjunction’, which are classified as qualities.

The categories could be (and to a large extent are by the eleventh-century philosopher Udayana) defined simply in terms of inherence relations. Substance can inhere in substance, and is inhered in by qualities, actions and universals. Eternal substances are distinguished because they are also inhered in by a particularity. Qualities inhere in substances and have qualityness inhering in them; actions inhere in one substance at a time and have actionness inhering in them. Universals inhere in a plurality of entities but have nothing inhering in them. Likewise, nothing inheres in particularities, which, however, inhere in only a single substance. The relation inherence itself inheres in nothing and has nothing inhering in it.

In the order of inherence relations among the categories, the contained provides unity and identity, the container the reality of the single individual entity. Universals are not themselves objects, but are contained in objects of all sorts, actions, qualities and substances; actions and qualities are objects, but always contained in substances; substances, if contained at all, are contained in another substance. The nesting of contained and container comes to an end with the atomic, eternal substances.

The category of substance provides the foundation for the Nyāya-Vaiśeṣika quest for permanence and self-sameness, the continuity underlying all change. A quality must be a quality of something, an action must have an agent, and a word, to have meaning, must refer to something.

The inherence of substance in substance is always a whole-part relation, the whole inhering in the parts. As the whole is itself a unitary substance, the relation implies that the whole is something more than the aggregate of its parts. In support of this contention, the inherence of substance in substance is considered a special case of causality. The whole inheres in the parts and the parts are the inherence cause of the whole. The cloth inheres in the threads and the threads are the inherence cause of the cloth. The inherence of cloth in threads is not the same as the conjunction of the woven threads, on the grounds that conjunction, a temporary relation, occurs when action joins two previously separately existing substances, whereas the cloth, as contained, cannot exist separately from the threads. Thus, an aggregate of material substances does not make the whole aggregate one substance, but the inherence of the whole in the parts does. The clay pot is one substance, a whole that inheres in its parts, other substances, ultimately earth atoms; but two pots touching remain a distinct pair. A living organism is one substance made of innumerable substances, yet remains distinct from the ground it stands on. Classical Vaiśeṣika says little beyond distinguishing inherence from conjunction to maintain that wholes are other than aggregates.

Other contentions make the distinction difficult to uphold, as the slightest change in a composite substance, brushing a few molecules off the pot surface, transforms it into an entirely different whole, an utterly new substance. Inherence, while rendering substances the foundation of a pluralistic realism as the containers that are not contained, also obscures what the individual unity of the substances consists in.

8 Parallels to the relation of inherence

It is instructive to compare the relation of inherence with a partial analogue in Aristotle's *Categories*. The analysis of predication shows for Aristotle a relation of containment that points to an ultimate substratum. A universal term, such as 'horse', can act both as subject and predicate in a sentence. But 'this horse', the term denoting an individual animal, can function only as subject. Grammatically, it is an ultimate subject of predication, and indicates a substance. Entities, which are denoted with words, may be said of something, present in something, both said of and present in, or neither said of nor present in. The term 'this horse' is an example of the last, and therefore indicates a substance, which, in the strictest sense, can neither be said of nor is present in anything else.

A tool for analysis, these relations do not constitute separate categories for Aristotle, unlike the Vaiśeṣika category of inherence. Moreover, in early Nyāya-Vaiśeṣika there is nothing analogous to the Aristotelian 'said of' relation. The difference is revealing for understanding the structure and organization of the Vaiśeṣika categories. Members of the Aristotelian category 'quality' (of which Aristotle gives no definitive listing) can be both 'said of' and 'present in', although not in the same sense. In the universal statement 'Roses are red', the quality 'red' is said of the subject 'rose' and the quality is real (for other red substances exist) even though the red rose in my vase withers and turns brown. But it may also be said that the rose in my hand is red in the sense that red is 'present in' it, meaning a particular red, a red that is no longer when a petal falls. Aristotle asserts the universality of qualities along with their dependence on ontologically fundamental substance: red exists without this red thing, but not without red things; and this red does not exist apart from this red thing.

The manner in which a quality may be conceived as present in a substance in Aristotle is analogous to the Vaiśeṣika understanding of qualities, which are in all cases individual and coextensive with the substances in which they inhere. Thus the Vaiśeṣika literature goes to lengths to defend the reality of a colour such as 'variegated'. Where a composite material substance does not have one single shade throughout (the fabric of a Hawaiian shirt, for example), its colour must nevertheless be one. The quality is one not by being what it is, but by being contained. Consequently any change in quality - with the exception of the cases distinguished as qualitative change due to 'cooking' (see §4) - is due to an underlying change in the containing substance. In Aristotle, where one petal falls the rose still is, although altered, and that particular red is no more; in Vaiśeṣika, both the substance itself and its qualities have been succeeded by new entities.

Therefore it might be said that in the Vaiśeṣika categories the identity of form is a separate consideration from the unity of an individual. This also marks a telling difference from Aristotelian ontology, where the individual substance is not only one as a substrate, but is one because it is a 'this', an essence. (Although some of the Vaiśeṣika substances are distinguished by qualities peculiar to that kind, this 'mark' is never the equivalent of the Greek notion of essence.) Nothing in classical Vaiśeṣika is analogous to the Aristotelian notion of a quality being said of, not present in, a substance. Vaiśeṣika does acknowledge a real unity of form to qualities of different individuals. This must be explicable by means of the inherence relation, and thus the universals, such as substanceness, qualityness, redness, are recognized as a distinct category. These too are present by containment. Redness inheres, not in substances, but in the reds, which in turn inhere in substances. When a person smells the rose, inhaling pollen in the process, the original composite substance has been destroyed and replaced by a new one, yet we will continue, correctly, to call it a rose.

9 Later developments to inherence and relations

Where Aristotle grounded the individuality of substances in an essence, which led him to wrestle in the *Metaphysics* with maintaining the ontological priority of substance to form, the Nyāya-Vaiśeṣika categories were careful to keep substance separate from quality and universal. This led to another philosophical difficulty, the difficulty of saying what a substance in itself is, to give content to the assertion that a substance apart from all other categories has a *svarūpa*, a nature in itself. We have seen this in a couple of respects. It is difficult to justify the distinction of a whole from an aggregate. And the ultimate substances are reduced to bare particulars, atoms,

which in themselves cannot be distinguished, requiring the introduction of particularity as a category. The term *svarūpa*, self-nature, was often connected more closely with the discussion of universals as they qualify substances, not the substances in themselves.

Navya-Nyāya, in the Nyāya-Vaiśeṣika tradition the 'new' school inspired by Udayana and beginning with Gaōgeśa (fourteenth century), significantly expands the analysis of relations, in general associating relations more intimately with the relata. Thus inherence was traditionally asserted to be one, on the grounds that it is always the same relation, a permanent relation of contained to container. This item of doctrine becomes a matter of debate, with some asserting that inherences are many and equal in number to the pairs of entities related by inherence.

The kinds of relation were also rethought. In the earlier period, relations were explained primarily by inherence, a separate category, as well as conjunction and disjunction, which are qualities. Navya-Nyāya, holding that the individual may also be related to other entities by its own nature, locates some relations not in separate categories, but in the individuals. This *svarūpasambandha* ('self-linking connector', 'peculiar relation') comes to have many varieties, but is especially important as the relation between a particular universal, that is, a universal less abstract than 'being' that therefore distinguishes what it qualifies from other things, and the entity the universal qualifies. For Praśastapāda, short of yogic knowledge of a substance as its very self, the knowledge of substances and other individuals is had by perception of an individual's own nature (*svarūpa*), which is the perception of a particular universal (*sāmānyaviśeṣa*) inhering in the individual. In Navya-Nyāya, the universal form, while still of a different category from the individual it qualifies, is more closely associated with the individual by making the relation of redness to red and of Bobness to Bob not inherence but *svarūpasambandha*. The form is still separate, but its relation to the individual is located in the individual itself, no longer in a third category. In the face of ultimate substances, bare and seemingly indefinable particulars, the particular relation marks an attempt to make definition possible by a more elaborate mechanism of specifying a substrate's relation to other entities.

The other major use of the *svarūpasambandha* relation is in explaining negations, in the relation of a substrate to something nonexistent, as in the proposition 'There is no pot on the mat'. This relation involves the final category.

10 Nonexistence or absence

The seventh category is of later origin than the first six. Discussed in the *Vaiśeṣikasūtra*, although not classified as a category, it is given little attention by Praśastapāda. It is acknowledged as a separate category by later writers and accepted without question in the subsequent Navya-Nyāya tradition. With the acceptance of nonexistence as a category, the structure of the categories is threefold. Substance, quality and action are all *satta*, reality, the name of the highest universal, and are all designated objects. The first three as well as universal, particularity and inherence are all designated *bhāva*, real, existing or present, in distinction to the seventh, *abhāva*, nonexistence or absence. Analysis of negation comes to be a central concern of Navya-Nyāya, including such questions as whether an absence is perceptible and whether the absence of an absence is positive or negative, that is, whether it is itself an absence. The basic division of nonexistence or absence was into four: prior absence is the nonexistence of an entity before it has been produced, posterior absence is its nonexistence after destruction, absolute absence is its nonexistence at all times and mutual absence is the nonexistence of one thing in another, such as the mutual absence of fire in water. The assertion of a mutual absence is equivalent to an assertion of nonidentity.

Any assertion of nonexistence is taken as an implicit reference to some positive entity, called the *pratiyogi* (counterpositive), which could be present in a substrate. Ultimate atomic substances appeared in themselves to be bare receptacles. Even with the addition of particular relations grounded in the individual substrate, to say something about an entity in itself is to say something about its relation to something else. Consequently a negative statement demands that the absence be a distinct entity, even though having no *svarūpa*, no nature of its own. 'There is no pot on the mat' is an assertion that, like any other assertion about the mat, places the mat in relation to something other, namely an absence (see Negative facts in classical Indian philosophy).

11 The nature of the self

A theory of categories is intended to lay out the structure of reality, affording knowledge of what is. For philosophy in India, a person seeks knowledge of what is, not to satisfy human curiosity, but to fulfil the highest

possible goal, commonly described as liberation of the soul. Traditionally, philosophical knowledge is one means to that end. The realism of the categories yields at the same time an account of that highest state of soul.

Part of that account, much disputed by other schools of thought, is the reality of soul (rejected by the Buddhists) and the plurality of individual souls (rejected by most other major schools). While material substances are defined and their existence is inferred by their attributes, it is still argued that perception of material substances themselves as distinct from all inhering qualities is possible. The establishment of the soul is more problematic, because it is asserted from the outset that the soul is an imperceptible substance. Therefore it is knowable only by inference. The existence of the self is inferred primarily from perception and agency. There must be a perceiver to whom the perception belongs, an agent who acts. Beyond this argument, the soul is inferred on the basis of the understanding of the nature of qualities: since no quality is independent - all exist only inhering in a substance that can contain them - and since we perceive certain qualities that are not encountered in material substances, namely, cognition, pleasure, pain and so forth, the self must exist as their substrate. There are no marks of soul in the sense that smell is the mark of earth. Certain qualities can exist in soul alone - pleasure, cognition and others - and all can also be absent. No qualities are coeval with soul. More clearly than for any other substance, soul is a container, not identifiable with and separate from any of its contents. To seek out and grasp the structure of reality is therefore to grasp the self. Liberation is ‘the subsistence of the self in its own pristine condition, marked by the cessation of all the specific qualities pertaining to it’ (Śrīdhara, *Nyāyakandalī*, to *Padārthadharmasaōgraha* VI, x; trans. G. Jha, 1915). Paradoxically, the simple picture drawn of the soul epitomizes the entire system of categories in its intricate complexity.

See also: Categories; Matter, Indian conceptions of §2; Nyāya-Vaiśeṣika; Ontology; Universals, Indian theories of

DAVID AMBUEL

References and further reading

Faddegon, B. (1918) *The Vaicesika System*, Amsterdam: Akademie van Wetenschappen.(An early descriptive survey with summaries and translations of some texts.)

Halbfass, W. (1992) *On Being and What there Is*, Albany, NY: State University of New York Press.(An excellent, informed, very readable account of classical Vaiśeṣika.)

Ingalls, D. (1951) *Materials for the Study of Navya-Nyāya Logic*, Boston, MA: Harvard University Press. (Inevitably technical, but a good brief introduction to Navya-Nyāya.)

Matilal, B.K. (1977) *Nyāya-Vaiśeṣika*, vol. 6.2 of *A History of Indian Literature*, ed. J. Gonda, Wiesbaden: Harrassowitz.(A good scholarly survey.)

Nyāyasūtra (*c*.400 AD), trans. G. Jha, *Gautama's Nyāyasūtras with Vātsyāyana Bhāṣya*, Poona Oriental Series 59, Poona: Oriental Book Agency, 1939.(One of the foundational works in the Nyāya-Vaiśeṣika tradition; for further reference, Potter 1977, and Potter and Bhattacharyya 1993 are excellent sources.)

Potter, K.H. (ed.) (1977) *Encyclopedia of Indian Philosophies*, vol. 2, *Indian Metaphysics and Epistemology: The Tradition of Nyāya-Vaiśeṣika up to Gaōgeśa*, Princeton, NJ: Princeton University Press.(A good scholarly survey.)

Potter, K.H. and Bhattacharyya, S. (eds) (1993) *Encyclopedia of Indian Philosophies*, vol. 6, *Indian Philosophical Analysis: Nyāya-Vaiśeṣika from Gaōgeśa to Raghunātha Śiromaṇi*, Princeton, NJ: Princeton University Press. (Similar in structure to Potter 1977, covering a later period.)

Praśastāpada (6th century) *Padārthadharmasaōgraha (Compendium of Properties of the Categories)*, trans. G. Jha, *Padārthadharmasaōgraha* of Praśastapāda, with the *Nyāyakandalī* of Śrīdhara, Chaukhambha Oriental Series 4, Varanasi: Chaukhambha Orientalia, 1915.(One of the foundational works in the Nyāya-Vaiśeṣika tradition; for further reference, Potter 1977, and Potter and Bhattacharyya 1993 are excellent sources.)

Vaiśeṣikasūtra (*c*. 2nd-3rd century AD), trans. and ed. N. Sinha, *The Vaiśeṣika Sûtras of Kaṇâda with the Commentary of Śaōkara Miśra*, Sacred Books of the Hindus 6, Allahabad: Pāṇini Office, 1911.(One of the foundational works in the Nyāya-Vaiśeṣika tradition; for further reference, Potter 1977, and Potter and Bhattacharyya 1993 are excellent sources.)

Opera, aesthetics of

Opera, which may be defined as a dramatic action set in large part to music, is an inherently unstable art form, more so than any other. It has been characteristic of its practitioners and critics to call it periodically to order, in idioms which vary but carry much the same message: the music exists to further the drama. This has often been taken to be a matter of settling the priority of two elements: music and text. But in fact three are involved: music, text and plot (or action).

Opera began very abruptly in Northern Italy at the end of the sixteenth century, partly as the result of discussions about its possibility. To begin with, familiar Greek myths were employed, set in the vernacular, with simple accompaniments so that every word could be heard. This led to pre-eminence for the singers and for spectacle. After each wave of excess - vocal prowess, dance interludes, stilted plots and texts, then once again, in the nineteenth century, empty display, and later gargantuan orchestras - there was a movement of revolt. Philosophers rarely took part in these aesthetic disputes, most of them being uninterested in music, and possibly more relevantly, being uninterested in any subject which can only be studied in historical terms. But it is fruitless to think about opera apart from its manifestations; every great operatic composer makes his own treaty between the potentially warring elements, Wagner being the most passionate propagandist for his own conception. In the twentieth century the aesthetics of opera have become pluralistic, as has, to an unprecedented degree, the form itself. The perpetual danger is that opera should degenerate into entertainment, and it is always the same message that recalls it to its original function - one which most spectators and listeners are happy to ignore: opera is a form of drama.

1 The constituents of opera, and a promising beginning

An opera is a drama which is in large part sung and is accompanied by an orchestra. Hence one can see that it is not an art form in which the elements are likely to remain in a stable relationship. The history of theorizing about opera, which naturally runs alongside the history of opera itself, is one of constant complaint and propaganda for a return to the ideals of the founders of the form - though as musical and other means develop, it is taken for granted that they will be realized in different idioms.

Although drama with musical accompaniment was no new thing when the theorists of opera were holding their debates in Florence towards the end of the sixteenth century, it had never been discussed in a systematic way. The Florentine theorists wished to return to something approaching the original performance-conditions of Greek tragedy; they were typical of Renaissance propagandists, both in seeking a validation for their theories by claiming that they were a return to antiquity and in having only a vague idea of what that would be. The Camerata, as they called themselves, were in fact concerned to bring a new art form into existence in the light of the possibilities of staging large-scale spectacles at North Italian courts - an art form in which actors with fine singing voices could perform versions of classical myths with whatever instrumental forces happened to be available. It may well be the only major kind of art for which an agenda has been worked out in advance of any actual examples.

As a notoriously liquid language, Italian possessed an in-built possibility of lyricism which was lavishly exploited by the first great master of opera, Monteverdi (1567-1643), whose *Orfeo* got the form off to a remarkably impressive start. It is sparsely accompanied, so that every word can be clearly understood. The elements of opera - singing, orchestral music, text and drama - are in the accord that the Camerata had promulgated. Most of Monteverdi’s subsequent operas are lost, but his last two, dating from the early 1640s, survive; though they are both different from *Orfeo*, and from one another, they maintain the principles and thus the balance which he had exemplified from the beginning. But having begun as a court entertainment, opera rapidly became one that appealed to the masses. It would be absurd to maintain that its popularity led to its decline. That would probably have occurred in any case, but what took over were an ever-greater lavishness in stage spectacle, longer and less relevant dance interludes, and a tendency on the part of the vocal soloists to demonstrate their often prodigious technique. Within a century of its birth, opera had already become decadent.

2 Aesthetics and history

The aesthetics of any art form, if they are to serve any useful purpose, are necessarily closely linked to the form’s

history; this is especially the case with the performing arts. The great manifestos which punctuate opera's history have all been devoted to the same task: that of recalling opera to the high ideals with which it began, and analysing the factors which have led to its departure from them. How large, for example, can the orchestra become without rendering the words that the singers are producing incomprehensible or even inaudible? How can one hope to understand a singer who produces an elaborate sequence of notes on a single syllable? The old tug of war between pleasure and instruction is nowhere more acute than in opera, instruction favouring simplicity of means, pleasure various kinds of display. Many of opera's greatest practitioners and all its leading theorists have been people who conceived of it as playing an important role in society, and of performances of significant works as serving to unite the audience in a celebration of shared communal values. But the temptations to indulge in one or another mode of virtuosity have repeatedly proved too strong.

A more deep-seated reason for the instability of opera lies in its relationship with the purely instrumental music of its time. For its earliest practitioners this was hardly a problem, since instrumental music as such was not yet a going concern. There is nothing inMonteverdi's operas that we can recognize as large-scale musical organization. Music remains subordinate to drama to a degree which ears familiar with Mozart orWagner can find frustrating. But, in the eighteenth century, before purely instrumental music was fully under way as an autonomous art, composers of opera were developing the aria, in which a character who was previously getting through a lot of musical prose, as it may be termed, takes their time to explore an emotion at greater length, to a melody which is appropriate to that emotion. The effect of this was to hold up the action, so that opera tended, in its second century, to become increasingly static.

This stalemate could not be broken until instrumental music became much more diverse in its modes of organization than it had been before the birth of sonata form, in the latter part of the eighteenth century. However much discussion of opera there might have been, it could not attain an adequate degree of detail and sophistication until music itself had advanced beyond a dependence on dance forms and counterpoint. For debate about the relationship between text, drama and music depends on what are envisaged as the possibilities of music. The expression of conflicting emotions only becomes feasible when contrasting themes are brought into relationship with one another and their conflicts are resolved in terms of more or less large-scale harmonic organization. That is what happened in the sonatas, symphonies and chamber works ofHaydn and Mozart, and what therefore enabled Mozart to write his operatic masterpieces, which dwarf all those that preceded him. By contrast Gluck (1714-87), the co-author of the Preface to his opera *Alceste*, one of the most famous of all insistences on the primacy of drama, gives the impression of advance in his 'reform operas' only by returning, in his own way, to a pre-Handelian simplicity. He made a virtue of necessity by avoiding all musical elaboration, thus conveying a unique sense of directness and dramatic relevance.

3 Mozart: sublime complexity

Mozart (1756-91) is the only great composer who is equally distinguished in operatic and instrumental works. His greatness in each fed that in the other. He had no need to theorize about opera, because he was able to take over into his operas the procedures that he employed above all in his sublime series of piano concertos. In his mature operatic masterpieces the glories, above all, are the huge ensembles. Though, strictly speaking, they are not in sonata form, they give the listener the firmest sense of a musical argument which underpins and indeed absorbs into itself the text and action, creating a seamless whole.

It is perhaps these works, more than any others, which led Joseph Kerman, in his influential *Opera as Drama* (1956), to claim that, in relation to opera, music serves the same function that poetry serves in poetic drama; above all in Shakespeare. That cannot be quite right. If one wishes, one can read libretti by themselves (not recommended). But one cannot read Shakespeare's plays without the poetry. The most one can do along those lines is paraphrase it. Shakespeare does his thinking in the language which he commits to paper. But in opera, even when the librettist and the composer are the same person, there are the words and there is the music, and one can admire or deplore the setting of the former to the latter. What precisely follows from this it is hard to say. Certainly the effect in the finest of operas is one of indissolubility. But there remains a crucial difference between the intensification of effect which music can achieve, and the absolutely singular effect of poetic drama. That is one reason why there is so much more criticism of poetic drama than of opera.Mozart's Don Giovanni is perhaps as complex a figure as Shakespeare's Hamlet, but though there are vast amounts of commentary on both these

characters, that on Hamlet will always far outstrip that on the Don. Poetry seems to challenge us to explain its implications, since if it is as great asShakespeare's we will never complete the task. Music, at least when it functions as it does in Mozart's operas, deepens a meaning which the text and action provide, and can add ambiguities and complexities, but within more circumscribed limits, it seems. Thus even in the case of a work as fascinating in its levels of operation as *Così fan tutte*, there is always the text to refer to, and the text is rather straightforward. In one way, the music itself provides most of the criticism of the drama that we need, at the same time as it creates it.

4 The trauma of Wagner

As with nearly all supreme creative figures, Mozart was both an inspiration and a depressant to his successors. In the presence of his works, discussion of the aesthetics of opera seems superfluous. But once more, degeneration set in, and once more in Italy, the home of opera. The writers of so-called *bel canto* opera were largely unconcerned with theory, and in any case the culture industry of which they were a part gave them no time for anything but the composition of their next work. But it was the dominance of singers which again led to drama yielding to music. Meanwhile, in Paris, the demand for lavish spectacle reached a new pitch of vulgarity, perfectly catered for by the white elephants of Meyerbeer (1791-1864). 'Effects without causes' was Wagner's description of his operas. It is Wagner (1813-83) who provided the next set of fundamental arguments about the nature and function of opera. Having himself set off on a course which was internationally influenced, with Meyerbeer as one of the contributors to his development, he was led by a series of historical events (many of which were unconnected with his art) to think with a depth and sustained energy about opera which has no parallel in the history of any of the arts, and which has set the agenda for operatic aesthetics ever since.

Wagner's dissatisfaction with opera was part of a wider revulsion from the whole society in which he lived. As a participant in the Dresden uprising of 1848-9 he was impelled both to think through the whole set of relations between politics and art, and to develop the concept of a total work of art, which would mirror the reintegration of man and society, and without which he could see only misery ahead. He saw in myth the fundamental truths about the human condition, but it was myth reinterpreted from a psychological standpoint. This could only be realized by the writing of libretti (he always wrote his own) of a complexity and density previously unknown, and the relegation to a subordinate place of music, which would nevertheless possess a degree of sophistication comparable to that of Beethoven's instrumental works. The aim was something far more ambitious than any operatic composer, indeed any artist, had entertained before: the transformation of consciousness by overwhelming communal experience.

The enormous body of theorizing whichWagner produced after the failure of the uprising is characterized by generalizations of Hegelian proportions, and expressed in language no less forbidding. But when he was introduced to the writings of Schopenhauer in 1854, he revised his views about the importance of music in relationship to the other elements of music drama, moving to a position where he saw it as the main conveyor of the drama. His celebrated later statement, that his dramas were 'deeds of music made manifest', sums up his new position. There has been an enormous amount of debate about the connections between Wagner's theories and his practice, but the overwhelming force of the latter and the ambitiousness of the former have guaranteed that his views and his art have been a major focus of aesthetic discussion, and not only of opera. Operatic theory and practice have been polarized, for the last 140 years, into pro- and anti-Wagnerian positions. The most strenuous objectors have been of two kinds: those who feel that his claims for art, and for music drama in particular, are preposterously inflated, and who have therefore advocated a return to some kind of pre-Wagnerian aesthetic (Stravinsky being the most articulate case); the second group have seen opera as an instrument of social change only on the condition that it stresses its own artificiality and leads its spectators to think about issues in a highly conscious way. Pre-eminent among this school have been the Marxists Brecht and Weill.

5 Aesthetic pluralism

Few general conclusions can be drawn from studying the history of opera and its commentators. Each operatic composer of genius has worked out the relationship between text, drama and music in his own way. Since operas which possess a fine text but contain no distinguished music do not survive, and some operas survive which possess a wretched plot and text but magnificent music, it is often claimed that opera is fundamentally a musical, not a dramatic form. The fact that one can enjoy an opera while having only a vague idea of what is being sung or

what the action is seems to point in the same direction. These claims leave open the view that the operas which matter most are those in which drama is articulated by music, and that those in which only the music counts are less artistically significant.

Operas are popular for different reasons, and many opera lovers, or fans, are more interested in hearing superb voices show off than in anything else. The distinction between entertainment and art seems to be more laxly applied here than in the case of most other art forms. The battle for opera as a serious form, which ranks at its best with the greatest art forms, is one that is constantly having to be re-fought; it can be no coincidence that it is always conducted in terms of opera as drama.

See also: Music, aesthetics of; Performing art §§1-3

MICHAEL TANNER

References and further reading

Donington, R. (1991) *Opera and its Symbols*, New Haven, CT: Yale University Press.(A highly original history of opera and its aesthetics in terms of recurring symbols.)

Goldmann, A. and Sprinchorn, E. (eds) (1964) *Wagner on Music and Drama*, New York: E.P. Dutton.(An excellent selection of Wagner’s prose, unfortunately in terrible translation; but no other is available.)

Kerman, J. (1956) *Opera as Drama*, New York: Oxford University Press; 2nd edn, 1989. (The most influential account of opera to have been published this century.)

Kivy, P. (1988) *Osmin’s Rage: Philosophical Reflections on Opera, Drama and Text*, Princeton, NJ: Princeton University Press.(An historical account of opera from Monteverdi to Mozart, substituting for ‘opera as drama’ a notion of ‘drama made music’.)

Strunk, O. (ed.) (1950) *Source Readings in Music History*, New York: W.W. Norton.(Contains all the seminal texts of the history of operatic aesthetics up to the end of the nineteenth century, except for inadequate representation of Wagner.)

Williams, B. (1992) ‘The Nature of Opera’, in *The New Grove Dictionary of Opera*, London: Macmillan, vol. 3. (A brilliant short account of the basic issues in operatic aesthetics and practice.)

Operationalism

'Operationalism', coined by the physicist Percy W. Bridgman (1927), has come to designate a loosely connected body of similar but conflicting views about how scientific theories or concepts are connected to reality or observation via various measurement and other procedures. Examples of an operation would be the procedure of laying a standard yardstick along the edge of a surface to measure length or using psychometric tests to measure sexual orientation. In the 1920-50s different versions of operationalism were produced by, amongst others: Bridgman, who was concerned with the ontology of basic units in physics; behaviourists such as E.C. Tolman, S.S. Stevens, who were concerned with the measurement of intervening variables or hypothetical constructs not accessible to direct observation, as well as B.F. Skinner, who sought to eliminate such nonobservables; and positivistic philosophers of science who were analysing the meaning of terms in scientific language. Conflation of their different operationalist philosophies has led to a great deal of nonsense about operational definition, methodology of observation and experiment, and the meaning of scientific concepts. Operationalist doctrines were most influential in the social sciences, and today the primary legacy is the practice of operationally defining abstract social science concepts as measurable variables.

1 Bridgman

Nineteenth-century physics extended the scope of Newtonian mechanics to include radiant heat, sound, electricity and magnetism, and so on, by modelling these phenomena as waves through a rarefied fluid-like ether. The 1887 Michelson-Morely interferometer experiments failed to detect a suitable ether, thereby laying the basis for the replacement of Newtonian mechanics by relativity theory. P.W. Bridgman viewed the Einsteinian revolution as challenging the assumptions on which classical experimental physics was based, and thus saw special relativity theory as calling into question all absolute physical principles and the existence of absolute physical matter. What, then, could be the subject matter experimental physics studied? Perceiving Einstein as having built relativistic physics out of various basic or elementary operations such as were used to measure relative simultaneity, he concluded that these basic operations not only connected theory to experiment but themselves were what experimental science studied. The heart of Bridgman's operationalism thus was an ontological view about the basic constituents of empirical reality. When he said, 'we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations' he was making the claim that concepts ultimately referred to constituent basic operations and nothing more. He thus rejected the idea that there were any *things* which the basic measurement procedures measured. In later works he embraced the incipient subjectivism of this position.

Bridgman insisted that each operation defines a separate concept, by which he meant that the referents of a given concept were precisely the mix of operations used in the specific operational definition. Critics charged that concepts like *temperature* admit of different measurement procedures, and so it was a mistake to equate the *meaning* of temperature with any particular procedure for measuring temperature. The extension of the term 'temperature' encompasses things measured by different procedures, but extension is not the sense of 'meaning' in Bridgman's dictum that the meaning of a concept is the set of operations; for his position denied the existence of things which could be the extension of a concept. Bridgman readily acknowledged that there could be connections among operationally defined concepts, and that terms like 'temperature' encompassed such connected operationally defined concepts. The basis of these connections was theory. For example, the various distinct concepts lumped under the rubric 'length' are 'connected' by the Lorentz transformation group from relativity theory - a kind of 'fusion' of concepts (see Relativity theory, philosophical significance of §1). Not all concepts in theories were definable by operations of physical measurement or experimental manipulation and design, and Bridgman allowed that 'paper and pencil' operations could be used to operationally define them.

For Bridgman operational analysis was a means for explicating and validating scientific concepts by disassembling theories and determining what was objective (rooted in basic operations). Non-objective theory content led to pseudo-problems. However, operational definition was not intended by Bridgman to be constitutive of concepts, a formula for doing science, the meaning link between formal and factual propositions, or a means of eliminating subjectivity from scientific knowledge. He increasingly found positivistic and behaviourist 'operationalist' philosophies that attempted to do so uncongenial. Although he had extended exchanges with operationalists like

Stevens and Skinner, these eventually led to total estrangement.

2 Logical positivism

Logical positivists, too, focused on the implications of the Einsteinian revolution: where Newtonian mechanics went wrong was postulation of the fictitious ether, which, referring to nothing, was said to be an empty or meaningless concept. Utilizing resources of the new symbolic logic they sought to develop a precise language for doing science in which concepts referring to fictitious entities were precluded. Things were ontologically real precisely if statements about them could be verified. In their reconstructed language for science, vocabulary terms were divided into the *logical*, the *observational* and the *theoretical*. Logical portions of language were reduced to symbolic logic and hence were tautological. Observational terms referred to directly experienceable things and so observation-language assertions were epistemically and ontologically non-problematic. Unrestricted introduction of theoretical terms led to the postulation of fictitious entities. To avoid this, legitimate theoretical terms had to be suitably defined by a 'dictionary' of 'correspondence rules' that gave observational content to theoretical assertions. Until correspondence rules are added, the only meanings theoretical terms have are those 'implicitly defined' by the formal connections in theoretical laws (see Theories, scientific §§1-2).

Initially the correspondence rules were required to be *explicit definitions* that reduced theoretical terms to abbreviations for complex observable conditions according to Schlick's dictum that the meaning of a statement is its method of verification (see Meaning and verification §2). Explicit definitions are stipulative definitions, hence analytic truths. Positivists assimilated Bridgman's dicta that the meaning of a concept is its corresponding set of operations and that each operational procedure is identified with a distinct concept to explicit definition since basic operations could be described using observation terms. Such conflation fundamentally distorts Bridgman's position. Positivistic operational definitions are describing procedures for measuring the very physical things Bridgman denied, and operational definitions are being put forward both as a method for doing science and as connecting the formal to the factual.

Later positivists came to see that some terms in theoretical physics could not be identified with observable conditions. Taking an increasingly realist position on theoretical terms they became concerned that, say, temperature could be measured via different operational procedures, so they loosened restrictions on admissible theoretical terms by admitting a plurality of *reduction sentences* which were operational procedures *partially defining* the terms by specifying observable meaning relative to a specific experimental context. Different operational definitions of the same term allowed for different procedures and contexts. Reduction sentences assert contingent relations between apparatus, theoretical entities, and their observable manifestation as the output of detectors or measuring devices. They are empirically true or false, and thus do not specify meaning analytically like stipulative explicit definitions. What connects the various correspondence rules partially defining 'temperature' is the contingent fact that all these temperature procedures are measuring the same physical *thing*.

Positivists continued to conflate their position with Bridgman's, now likening operational definitions to reduction sentences - which led them to criticize Bridgman's view that each operational procedure defined a distinct concept. They also criticized Bridgman for his growing emphasis on the subjectivity he saw growing out of the denial of any absolute physical reality and his view that the objectivity of science rested in what scientists *do* when they operationalize.

Later the positivists came to see that some terms (such as the Ψ function in quantum mechanics) could not be individually defined even by reduction sentences. This led to the weaker requirement that theoretical terms did not have to be individually operationally defined so long as they occurred essentially in the theory and the theory itself had observable consequences specified by an *interpretive system* of correspondence rules linking the theoretical and observational components. Positivists were never able satisfactorily to specify requirements for interpretive systems that demarcated scientific from nonscientific systems. Trading on earlier conflations, in moving to interpretive systems some positivists saw themselves as rejecting Bridgman's operationalism.

3 Behaviourisms

Although 'operationalism' was introduced to psychology around 1930, S.S. Stevens brought it to prominence. He argued that scientific disputes are rooted in disagreements over the application of concepts. Since the application

of concepts is rooted in concrete operations, agreement on the operations that should be associated with a concept would eliminate previously unresolvable disagreements. If there are disagreements in the application of a concept these can be resolved in an 'operational regress' in which applicability is settled by iterated appeal to less problematic concepts.

Operations consist in gross physical behaviours (say, laying a yardstick along a surface) and a result (number at the point of edge congruence). *Denoting* is the most fundamental behaviour on which all operations depend. *Results* are the outcomes of discrimination, which 'is the fundamental operation of all science'. Science thus depends on observer characteristics and so psychology becomes 'the propaedeutic science', providing data on which operational procedures depend. But psychology *qua* science itself must be operationalist.

Although operations rest in scientists' discriminatory capabilities, Stevens rejected any notion of a sensory given or subjective basis for science. Operations *are* behaviours and the basic operation is 'to react discriminatively'. Science thus is 'public'. Measurement scales are constructed out of complexes of operations which determine their arithmetical properties.

Around 1933-4, Stevens encountered Bridgman's operationalism and logical positivism, and in subsequent writings allied his independently developed 'operationalist' philosophy with their positions despite substantive disagreements over their reliance on subjective experience or the sensory given. It is unclear whether Stevens subscribed to the view that concepts are synonymous with their associated operations, but he was unsympathetic towards what he understood to be positivism's 'linguistic operationalism'.

E.C. Tolman pioneered an 'operationalist' approach to mental phenomena in the 1920s, although he did not adopt the term until 1936. Mentalistic concepts required external criteria of application. Behaviour is a complex function of stimulus conditions, hereditary makeup, past training and appetite or aversion. *Intervening variables* represent an organism's readiness to respond to stimuli and are to be operationally defined by means of standard experiments. One measures independent and dependent variables. Intervening variables are specified by holding all but one independent variable constant then determining the curve expressing the functional relationship between the selected independent and the dependent variable. Intervening variables thus required defining experiments. As Carnap was making the switch from explicit definitions to reduction sentences, Tolman spent some time in Vienna (1933-4), and his characterizations increasingly resembled Carnap's hypothetical constructs defined by reduction sentences. Whereas positivists viewed the operations linguistically, Tolman continued to construe them psychologically.

Stevens' articulation of the operationalist position provided focus for discussions of methodological issues surrounding behaviourism. Clark Hull (1940) had been an early advocate of the geometric method in psychology and quickly allied himself with axiomatic themes in positivism. Although he made little contribution to operationalism himself, his student Spence collaborated with the positivist Gustav Bergmann to produce a positivistic explicit definition version of operationalism (1941). B.F. Skinner was an early advocate of an explicit definition version of operationalism which held that since explicit definitions rendered theoretical terms (including intervening variables) eliminable, there was no justification for allowing theoretical terms in science. Hempel (1958) presents the positivistic rejoinder, stressing the importance of systematization and integration which theoretical terms allow.

Criticisms charged that operationalism was trivial or was a minor part of science, needlessly proliferated concepts, ultimately leaves psychological concepts undefined, and makes experimental error meaningless since there is no standard against which operations can be validated. The last charge was rebutted later by development of evaluative notions such as concurrent and construct validity.

4 The operationalist legacy

Although positivism had limited influence in the development of operationalist philosophies, it was an active focus of comparison and dissemination for disparate operationalist views. Scientific practice in psychology and sociology became increasingly behaviourist and operationalist in a manner that minimized fundamental differences. It became commonplace that theoretical terms should be operationally defined, but what that demanded became increasingly unclear.

MacCorquodale and Meehl (1948) crystallized issues when they construed intervening variables as those definable by explicit definitions and hypothetical constructs as those only definable via reduction sentences. Today it is commonplace that operational definitions are required when 'remote' quantitative variables are invoked. Yet exactly what operational definitions come to remains unresolved due to the conflations noted above. Disputes over the adequacy of operational definitions are frequently dismissed as 'mere semantic disputes' - a move appropriate only for stipulative explicit definitions - when they turn on whether the operational definitions correctly capture empirical connections between hypothetical-construct theoretical terms and various observable states of independent and dependent variables.

Nevertheless operational definitions remain an adequacy requirement for social science research.

See also: Behaviourism in the social sciences; Behaviourism, analytic; Logical positivism; Scientific method

FREDERICK SUPPE

References and further reading

Bergmann, G. and Spence, K.W. (1941) 'Operationalism and Theory in Psychology', *Psychological Review* 48: 1-14; repr. in M.H. Marx (ed.) *Psychological Theory: Contemporary Readings*, New York: Macmillan, 1951. (Presents an explicit definition positivistic version of operationalism.)

Blumberg, A. and Feigl, H. (1931) 'Logical Positivism, a New Movement in European Philosophy', *Journal of Philosophy* 28: 281-96.(Sets forth positivistic program and identifies Bridgman's operational definitions with their implicit definitions.)

Boring, E. *et al*. (1945) 'Symposium on Operationalism', *Psychological Review* 52 (Sept.): 241-94.(Papers by Boring, Bridgman, Feigl, Israel, Pratt, Skinner; organized in response to Israel and Goldstein attack.)

Bridgman, P.W. (1927) *The Logic of Modern Physics*, New York: Macmillan.(Original statement of his operationalism.)

Bridgman, P.W. (1936) *The Nature of Physical Theory*, Princeton, NJ: Princeton University Press; repr. *Philosophical Writings of Percy Williams Bridgman*, New York: Arno Press, 1980.(In this work Bridgman pursues the subjectivist tendencies in his operationalism.)

Bridgman, P.W. (1950) *Reflections of a Physicist*, New York: Philosophical Library; 2nd edn, 1955; repr. Arno Press, New York, 1980.(The 2nd edition of 1955 contains ten papers additional to the original of 1950. Contains reprints of nearly all his philosophical and nontechnical papers on operationalism.)

Bridgman, P.W. (1952) *The Nature of Some of Our Physical Concepts*, New York: Philosophical Library; repr. *Philosophical Writings of Percy Williams Bridgman*, New York: Arno Press, 1980.(Later restatement of his operationalist philosophy similar to his 1936.)

Carnap, R. (1936-7) 'Testability and Meaning', *Philosophy of Science* 3: 420-66, 4: 1-40.(Classic positivistic criticism of explicit definition operationalization and articulation of reduction sentences approach.)

Carnap, R. (1956) 'Methodological Character of Theoretical Concepts', in H. Feigl and M. Scriven (eds) *Minnesota Studies in the Philosophy of Science*, Minneapolis, MN: University of Minnesota Press, vol. 1, 33-76.(Positivistic interpretative system account of operationalization which has extended discussion of operationalizing psychological concepts.)

Chronbach, L. and Meehl, P.H. (1955) 'Construct Validity in Psychological Tests', *Psychological Bulletin* 52: 281-302.(Classic articulation of validity notions for evaluating operational definitions.)

Frank, P. (ed.) (1961) *The Validation of Scientific Theories*, New York: Collier.(A retrospective symposium on operationalism.)

Hardcastle, G. (1995) 'S.S. Stevens' Philosophy of Science', *Philosophy of Science* 62: 404-24.(Assesses relations of Steven's operationalism to other behaviourisms and positivism.)

Hempel, C.G.(1952) *Fundamentals of Concept Formation in Empirical Science*, Chicago, IL: University of Chicago Press.(Positivistic account of operationalization that criticizes reduction sentence approach and introduces interpretative sentences.)

Hempel, C.G. (1954) 'A Logical Appraisal of Operationalism', *Scientific Monthly* 79: 215-20; repr. in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press, 1965: 123-33. (Positivistic critique of operationalism that conflates different versions.)

Hempel, C.G. (1958) 'The Theoretician's Dilemma', in H. Feigl, M. Scriven and G. Maxwell (eds) *Minnesota Studies in the Philosophy of Science*, Minneapolis, MN: University of Minnesota Press, vol. 2, 37-98; repr. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press, 1965. (Positivistic rejoinder to Skinner on necessity of theoretical terms.)

Hull, C.L. *et al.* (1940) *Mathematico-Deductive Theory of Rote Learning: A Study in Scientific Methodology*, New Haven, CT: Yale University Press.(Unsuccessful attempt to do psychology the positivistic way using symbolic logic.)

Israel, H. and Goldstein, B. (1944) 'Operationalism in Psychology', *Psychological Review* 51: 177-88.(Scathing attack on operationalism that prompts Boring to organize the 1945 'Symposium on Operationalism'.)

Koch, S. (ed.) (1959) 'Epilogue', *Psychology: A Study of a Science*, New York: McGraw-Hill, vol. 3, 729-88.(One of his many critical attacks on behaviourist operationalisms.)

MacCorquodale, K. and Meehl, P.E. (1948) 'Hypothetical Constructs and Intervening Variables', *Psychological Review* 55: 95-107; repr. in H. Feigl and W. Sellars (eds) *Readings in the Philosophy of Science*, New York: Appleton-Century-Crofts, 1953.(Classic differentiation of intervening variables and hypothetical constructs on the basis of operational definition form.)

Moyer, A.E. (1991) 'P.W. Bridgman's Operational Perspective on Physics', *Studies in the History and Philosophy of Science* 22: 237-58, 373-97.(Historical account. 'Part I: Origins and Development' and 'Part II: Refinements, Publications, and Receptions'.)

Rogers, T.B. (1989) 'Operationalism in Psychology: A Discussion of Contextual Antecedents and an Historical Interpretation of its Longevity', *Journal of the History of the Behavioral Sciences* 23: 139-53.(Historical account.)

Schlick, M. (1936) 'Meaning and Verification', *Philosophical Review* 45: 339-69; repr. in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, 1949.(Argues the meaning of a statement is its method of verification.)

Skinner, B.F. (1961) *Cumulative Record*, New York: Appleton-Century-Crofts.(Enlarged edition that contains reprintings of his papers detailing his operationalism and rejection of theories.)

Smith, L.D. (1986) *Behaviorism and Logical Positivism: A Reassessment of the Alliance*, Stanford, CA: Stanford University Press.(Comprehensive history arguing that behaviourist operationalism has minimal debts to positivism.)

Stevens, S.S. (1935) 'The Operational Basis of Psychology', *American Journal of Psychology* 47: 323-30. (Programmatic statement of his 'operationism'.)

Stevens, S.S. (1935) 'The Operational Definition of Psychological Concepts', *Psychological Review* 42: 517-26. (Basic statement of his 'operationalism'.)

Stevens, S.S. (1936) 'Psychology: The Propaedeutic Science', *Philosophy of Science* 3: 90-103.(Stevens 'fast-moving' summary of the operationalist point of view written for philosophers of science.)

Stevens, S.S. (1939) 'Psychology and the Science of Science', *Psychological Bulletin* 36: 221-63.(An influential review article.)

Suppe, F. (1977) 'The Search for Philosophic Understanding of Scientific Theories', in F. Suppe (ed.) *The Structure of Scientific Theories*, Urbana, IL: University of Illinois Press, 2nd edn, 3-243.(Contains a comprehensive discussion of positivistic views on theories; pages 17-27 contains a detailed discussion of their changing operationalist views.)

Tolman, E.C. (1932) *Purposive Behavior in Animals and Men*, New York: Century.(Introduces intervening variables method but does not call it operationalist.)

Tolman, E.C. (1966) *Behavior and Psychological Man: Essays in Motivation and Learning*, Berkeley, CA: University of California Press.(Reprinting of many of his papers on operationalism and behaviourism.)

Waters, R.H. and Pennington, A. (1938) 'Operationalism in Psychology', *Psychological Review* 45: 414-23. (Ambivalent critique raising the measurement error problem.)

Walter, M.L. (1992) *Science and Cultural Crisis: An Intellectual Biography of Percy Williams Bridgman (1882-1961)*, Stanford, CA: Stanford University Press.(Part III is a comprehensive account of Bridgman's operationalism and his interactions with positivists and behaviourists.)

Optics

Optics as physics concerned with the manipulation and study of light and, more recently, the general study of electromagnetic radiation, has a history back to ancient Egypt, and systematic study to classical Greece. But physics has proved better able to manipulate light than to explain its fundamental nature.

'Geometrical optics' treats light as a bundle of discrete rays, tracing their rectilinear paths reflected from surfaces and refracted through transparent media. 'Physical optics' treats light as a wave. It explains the dispersion of white light into spectral colours, the bands and colour patterns of diffraction phenomena, and aspects of the absorption and scattering of light. Characterizing the way in which the physical aspects of light become the perceptual aspects of shape and colour joins physics, physiology and philosophy in the perennial question of the correspondence of our perceptions to the physical world itself.

A modern view of light describes it in terms of massless particulate photons. This 'quantum optics' treats the absorption and emission of light by matter; providing precise knowledge of matter's inner structure, and the technology of lasers. Philosophically, quantum optics has led to the fundamental question: what is light? What is this natural entity which is created and destroyed in a particle-like way and yet propagates through space - and lenses, holes and slits - in a wave-like way? Experiments in which individual photons interfere with themselves make it hard to think of them as having unique paths. Experiments involving the correlation of photon properties threaten attempts to describe photons as having individual properties and interacting only locally.

1 History

Optical technology goes back at least three millennia. Mirrors made of polished copper and bronze were in use in ancient Egypt. The earliest known theories of light were held by the Greeks, who recognized the rectilinear propagation of light and stated the law of reflection. Lens combinations, in the form of the refracting telescope and microscope, were developed in the sixteenth century, and used notably by Galileo (see Galilei, G. §2). But the historical debate about light relevant to current discussions of its nature began in the late seventeenth century.

Isaac Newton was the chief champion of the corpuscular account of light, holding that the different colours of light he saw in his dispersal of white light by means of a prism were associated with different corpuscles (see Newton, I. §2). His view stood against the undulatory view championed by Christiaan Huygens.

In Huygens' view, light was solely the result of rapidly (but finitely) propagating longitudinal undulations in ethereal matter, undulations which slowed down in denser media (in contrast to Newton's more rapidly moving corpuscles). Huygens was able to derive the laws of reflection and refraction, to explain the double refraction of certain crystals, and to recognize in it a manifestation of a kind of double character to light itself, a 'two-sidedness' as Newton described it, the phenomenon we now call polarization. The wave model was much improved by Thomas Young and Augustin Fresnel in the early nineteenth century. They introduced the revolutionary idea that ethereal undulations could interfere negatively as well as positively, thereby accounting for the diffraction patterns (bands of light and dark) arising from obstacles and apertures.

The peculiar two-sidedness of light remained a problem until Young's suggestion that ethereal undulations were transverse to the ray direction. Polarization was then simply a manifestation of the undulations of the ether in perpendicular directions.

This largely qualitative model of light propagation was rendered firmly quantitative by the work of James Clerk Maxwell in the mid-nineteenth century. Maxwell's concern was with the nature of electrical and magnetic phenomena, and he was able to distill the essence of these phenomena into a very small set of mathematical expressions (see Electrodynamics). From these equations, Maxwell was able to show that an electromagnetic disturbance could propagate as a transverse wave, with a speed determined by basic, and independently measurable, electric and magnetic properties of the medium. When Maxwell substituted known values of these properties into his velocity expression, the number he obtained was the same as that empirically determined (by Fizeau and Foucault) for the speed of light, roughly $186,000$ miles per second. Light appeared to be precisely an electromagnetic disturbance propagated in the form of transverse waves through the ether.

Mathematically, with Maxwell's equations, the modern model of physical optics is complete. But conceptually, a great step had yet to be taken. For though Maxwell's mathematical analysis did not require any physical model of the ether, he remained firmly in the grip of it as a material underpinning. Yet as much as he was unable to imagine an electromagnetic undulation without an underlying undulatory medium, Maxwell nonetheless recognized the electromagnetic field itself as a repository of energy. In a vision with roots in the ideas of Michael Faraday, Maxwell saw the space surrounding electrified and magnetic bodies as suffused with energy. Rather than electromagnetic energy flowing through conductors with the flow of current, it was seen as flowing around and into them from the surrounding immaterial field. This view of the electromagnetic field as an autonomous entity, and light as an undulation in the field itself, is the field theory of physical optics (see Field theory, classical).

2 Classical optics

These two views of light - the remnants of the corpuscular model in ray or geometrical optics and the electromagnetic wave model in physical optics - together define what is called classical optics. Almost all the optical phenomena comprising the world of visual experience, and the many ubiquitous optical apparatus which enhance this world, are generally understood in terms of classical optics.

Geometrical optics traces the rectilinear paths of light reflected from surfaces and refracted through transparent media in an optical system. Such a system, besides mirrors and lenses, may contain prisms, apertures, beamsplitters, gratings, filters, polarizers, diffusers, even fibre bundles. Each ray from an object point is traced through the system to a single point on an image plane. All these points make up the geometrical image of the object as formed by the system.

Where the wave character of light becomes important is in the realm of physical optics. Physical optics treats the entire spectrum of electromagnetic radiation, from radio waves of kilometre wavelengths, through microwaves, infrared, visible and ultraviolet light, to X-rays and finally gamma rays at a thousand-millionth of a millimetre. Though vastly different processes are involved in the emission and absorption of electromagnetic radiation over this range, its propagation through space - and through lenses, holes and slits - is treated in a perfectly unified way by Maxwell's equations and their interpretation; descriptive of an autonomous, energy-containing, yet immaterial field. Polarization, interference, diffraction, are all variations on the basic field-wave property of superposition: the linear combination of the perpendicularly oriented transverse components of multiple waves. Together with aspects of the absorption and scattering of light, they fully account for the shapes and colours we perceive in the world. How our visual universe is created from the processes of physical optics is another matter, however.

3 Vision

Our eyes are highly complex compound lens systems with electrochemical receptors. Place these eyes in the world of physical optics and one does have the means for accounting for virtually all the visual appearance of the world, though it is surprising how many and various are the physico-optical properties of objects that contribute to our perception of them. Still, though the image of the world on our retinas must contain all the information required for seeing, the relationship between that information and the content of our visual field is far from straightforward.

The resolution of detail permitted by our lens-receptor system, understood passively as a camera, is far less than what we actually resolve perceptually. We have blind spots and blood-vessel shadows on our retinas, chromatic aberration and scattering in our lenses, quivering and drifting induced by eyeball musculature, yet we see no vestiges of these in our perceptual world. Clearly, much visual processing goes on in the brain. Unfortunately, the further one moves from the visual receptors themselves and into the activity of the visual cortex of the brain, the less certain our understanding is.

The intimate involvement of the brain in seeing inevitably raises the philosophical issue of the correspondence of the visual field and the real world. Is, for instance, colour a genuine physical property, a combination of physical properties, or supervenient on some set of physical properties? An objectivist would answer the question in the affirmative. A subjectivist might counter that colours are only properties of physical objects in the sense of being dispositions of those objects to affect us visually in the way they do. As our understanding of neurophysiology increases, this debate becomes ever more subtle (see Colour, theories of).

4 Light and relativity theory

Nineteenth-century concerns with the anomalous physical properties of the luminiferous ether, and generations of efforts to detect it, were swept aside conceptually in the early twentieth century by Albert Einstein's special theory of relativity (see Einstein, A. §2). (Michelson and Morley's interferometer experiment in 1887, popularly regarded as definitive, established any velocity of the earth with respect to an ether at less than $4,700$ metres per second. Results from the 1980s limit this to 0.05 metres per second. Manipulation of light continues to have its own dynamic, independent of conceptual revolutions.) In Einstein's theory, the velocity of light is completely independent of the relative velocity of the source or any observer. (This assumption of Einstein's, in turn, has now been tested for relative velocities more than 90 per cent of light speed.) In a spatial analogy, this invariant velocity is like a far object that remains always at the same distance, no matter how one moves.

Conceptually, this anomalous property of light falls right out of the modern formulation of Einstein's theory as a spacetime theory (see Spacetime §2). The intrinsic geometry of nature is four-dimensional, and our experience of it is as three-dimensional space and one-dimensional time, thus our notion of simultaneity and our determinations of velocity, are a function of our relation to this intrinsic geometry, our state of motion. In this formulation, light has a special place, traversing a special set of intrinsic straightest paths in (generally curved) spacetime. It is our locally-variable velocity determinations that produce the universal invariant velocity of light (see Relativity theory, philosophical significance of §§1-2). Its invariance is enshrined in our current choice of world standards: the velocity of light was in 1983 set at its best measured value of 299,792,458 metres per second and the international standard of length was adjusted accordingly. All measured distances must now be thought of as light seconds.

5 Light and quantum mechanics

Einstein is closely associated with the other great twentieth-century revolution in our understanding of light, the quantum or (as it is now called) photon theory (see Einstein, A. §4). In proposing that light is transmitted by free quanta that retain their particulate character as they travel through space, Einstein joined Max Planck in imagining that the exchange of energy between radiation and matter could take place only discontinuously, in energy quanta related to the frequency of the radiation. The myriad interference effects that wave optics explained so readily were to be accounted for by inter-quanta interactions. Niels Bohr's treatment of the spectrum of the hydrogen atom added a crucial third component to this early quantum theory. In Bohr's treatment, atomic electron orbits of only certain energies were allowed. The frequency of radiation emitted and absorbed by the atom reflected the transitions between these discrete electron energy states, with the energy-frequency relation that of Planck.

Bohr's central idea has led to the vast field of modern spectroscopy, in which the emission and absorption of radiation by matter is analysed for astoundingly precise knowledge of molecular, atomic, nuclear and subnuclear structure. (The laser - a coherent, tightly focused beam of light created by the stimulated de-excitation of a population of atoms all excited to precisely the same state - has been one product, and essential tool, of modern spectroscopy.)

But the problem Einstein confronted with his hopes for an inter-quanta interaction theory of light interference now looms larger than ever. For it is now possible to manipulate experimentally single photons of light, and interference effects persist. The classic experiment involves placing a half-silvered mirror, or other beamsplitter, downstream from a light source, creating dual optical paths. (A half-silvered mirror is one in which the metallic coating is too thin to be opaque, so light will be both reflected and transmitted through it.) The presence of single photons is confirmed by correlating detection events in each path: perfect anticorrelation is unambiguous evidence for single photons. Yet when path detectors are replaced by mirrors, and the paths recombined, an interference pattern of individual photon impacts gradually accumulates. This occurs even when (effectively) the choice of mirrors or detectors is made after the light passes through the beamsplitter. It seems difficult to say, prior to the downstream choice of detectors or mirrors, that the light followed a single path (detectors) or both paths (mirrors). Yet surely any sense of unique energetic history would require that circumstance to be decided at the beamsplitter.

Another classic experiment is based on yet another proposal made by Einstein (see Einstein, A. §5) and developed by John Bell in the 1960s (see Bell's theorem). The proposal has to do with multi-component quantum systems, spatially separated after being prepared in such a way that the properties of the individual components are tightly correlated. The results from the separate detections of component properties are predicted to display correlations

that are inconsistent with the two-fold view that the properties established at preparation are retained throughout the individual path histories, and that the measured values reflect only local interactions with the detectors. The experiment involves the preparation from a single light source of pairs of photons with correlated polarizations. Travelling different optical paths, the streams of photons display individual randomly oriented polarizations, which nevertheless upon coincident comparison exhibit the unusual correlations. It seems difficult in this case to maintain the view that the photons can indeed be prepared with determinate properties and then causally separated.

In these experiments the interpretive problems of light merge with the general interpretive problems of the quantum theory (see Quantum mechanics, interpretation of). Our great sophistication in manipulating light places it at the experimental centre of this most profound contemporary controversy.

ROGER JONES

References and further reading

Hardin, C.L. (1988) *Color for Philosophers*, Indianapolis, IN: Hackett Publishing Company.(A good source for the discussion of §3.)

Hecht, E. and Zajac, A. (1974) *Optics*, Reading, MA: Addison-Wesley.(A good undergraduate text, accessibly written. Strong on manipulations of all kinds, from Maxwellian electrodynamics in §2 to lasers in §5.)

Ronchi, V. (1939) *Storia della luce*, Bologna: Zanicelli; trans. and revised by V. Barocas, *The Nature of Light: An Historical Survey*, Cambridge, MA: Harvard University Press, 1970.(The standard historical survey through the eighteenth century.)

Zajonc, A. (1993) *Catching the Light: The Entwined History of Light and Mind*, New York: Bantam.(An unusual history, with quantum optics (§5) through 1991, written by a physicist.)

Ordinal logics

By an ordinal logic is meant any uniform effective means of associating a logic (that is, an effectively generated formal system) with each effective ordinal representation. This notion was first introduced and studied by Alan Turing in 1939 as a means to overcome the incompleteness of sufficiently strong consistent formal systems, established by Kurt Gödel in 1931.

The first ordinal logic to consider, in view of Gödel's results, would be that obtained by iterating into the constructive transfinite the process of adjoining to each system the formal statement expressing its consistency. For that ordinal logic, Turing obtained a completeness result for the class of true statements of the form that all natural numbers have a given effectively decidable property. However, he also showed that any ordinal logic (such as this) which is strictly increasing with increasing ordinal representation cannot have the property of invariance: in general, different representations of the same ordinal will have different sets of theorems attached to them. This makes the choice of representation a crucial one, and without a clear rationale as to how that is to be made, the notion of ordinal logic becomes problematic for its intended use.

Research on ordinal logics lapsed until the late 1950s, when it was taken up again for more systematic development. Besides leading to improvements of Turing's results in various respects (both positive and negative), the newer research turned to restrictions of ordinal logics by an autonomy (or 'boot-strap') condition which limits the choice of ordinal representations admitted, by requiring their recognition as such in advance.

1 Incompleteness, consistency statements and reflection principles

The point of departure for the study of ordinal logics is Kurt Gödel's second incompleteness theorem, also referred to as the theorem on unprovability of consistency (see Gödel's theorems). With each effectively generated formal system *S* containing a modicum of arithmetic is associated the arithmetical statement Con_S of consistency of *S*. If *S* is consistent then Con_S is not provable in *S*, and so the system S' obtained by adjoining Con_S as an axiom to *S* is stronger than *S*. It may, in fact, happen that S' is inconsistent, but if *S* satisfies a condition slightly stronger than consistency (to be described below) so also does S'. Let $\mathcal{C}$ be the class of effectively generated formal systems satisfying this strengthened consistency condition. Thus by beginning with any system S_0 in $\mathcal{C}$ and forming, in succession, $S_1 = S_0', S_2 = S_1', \ldots\}$, we obtain more and more complete systems in $\mathcal{C}$. As will be explained in §2, this procedure can also be extended into the transfinite, since the union of any effectively generated sequence of systems in $\mathcal{C}$ is again in $\mathcal{C}$. In this section, matters concerning formal systems and extension principles such as those of Gödelian type will be explained as needed to make these notions precise. Certain preliminaries from recursion theory and arithmetical definability theory are necessary for this purpose (see Computability theory; Recursion-theoretic hierarchies).

The variables $x, y, z, \ldots$ are here taken to range over the set $N = \{0, 1, 2, \ldots\}$ of natural numbers. For $P(x)$ a property of natural numbers, $\{x \colon P(x)\}$ denotes the set of all *x* in N for which $P(x)$ holds. The class $\sum_1^0$ in the arithmetic hierarchy consists of all sets of the form $\{x \colon \exists y R(x, y)\}$ where *R* is primitive recursive; $\sum_1^0$ is coextensive with the class of recursively enumerable (RE) sets, i.e. those which are effectively generated according to Church's thesis. There is a uniform $\sum_1^0$ enumeration $W_e = \{x \colon yT(e, x, y)\}$ of all $\sum_1^0$ sets for $e = 0, 1, 2, \ldots$, where *T* is a suitable primitive recursive relation; *e* is said to be an *index* of an effectively generated set *W* if $W = W_e$. The only other classes in the arithmetic hierarchy to which reference will be made below are the $\prod_1^0$ sets $\{x \colon \forall R(x, y)\}$ and the $\prod_2^0$ sets $\{x \colon \forall y \exists z R(x, y, z)\}$. (See Recursion-theoretic hierarchies for more information about the $\prod$ and $\sum$ classifications.)

For simplicity, on the formal side we shall deal only with systems whose language is that of first order (or elementary) number theory. Primitive recursive relations may be taken as basic, or defined (by a method due to Gödel) in terms of the operations of addition and multiplication. Either way, we can then speak of formulas in $\sum_1^0$, $\prod_1^0$ and $\prod_2^0$ form (among others). All notions of truth and definability are referred to the standard model, in which the variables are taken to range over *N* and the basic operations and relations receive their intended interpretation.

To carry notions of effectiveness over to the syntactic side, one uses the association with each formula *A* of its

Gödel number, denoted $\#A$. Then the above enumeration of $\sum_1^0$ sets induces an enumeration of sets of formulas, $S[e] = \{A\colon \#A \in W_e\}$; a set *S* of formulas is said to be effectively generated (or RE) if $S = S[e]$ for some *e*, and in that case *e* is called an *index* or *presentation* of *S*. Effective operations on such sets are explained in terms of operations on their indices. Moreover, effective sequences $S_n(n \in N)$ of RE sets of formulas can be considered as given by (an index *f* of) a total recursive function $\{f\}$ with $S_n = S[\{f\}(n)]$ for each *n*. Associated effectively with each such *f* is an index of $\cup_n S_n$ as an RE set.

Gödel's incompleteness theorems apply to effectively generated extensions of a formal system containing a modicum of arithmetic. Most familiar for the latter purpose is the system *PA* of Peano Arithmetic (Peano's axioms in first-order form), though much weaker systems suffice. Thus in the following, only RE sets *S* of formulas will be considered for which *S* includes *PA*. Then *A* is provable from *S* if *A* is a consequence of *S* by means of the axioms and rules of the first-order predicate calculus with identity. For simplicity, we assume that the non-logical axioms of *S* are given as closed formulas, that is, sentences. *S* is *consistent* if it does not prove the sentence $0 \neq 0$, and is 1-*consistent* if each $\sum_1^0$ sentence provable in *S* is true. *S* is called ω-*consistent* if there is *no* formula $A(x)$ for which we have both $\exists x A(x)$ and $\neg A(\overline{n})$ provable in *S* for each *n*. (The notion of ω-consistency was introduced by Gödel in 1931 to show that if *G* is the sentence which expresses of itself that it is not provable in *S* then $\neg G$ is not provable in *S*; it was observed later that 1-consistency suffices for this conclusion.) Finally, *S* is said to be *true* if each sentence of *S* is true. Note that truth implies ω-consistency which implies 1-consistency which, in turn, implies consistency. We write $S_1 S_2$ if every *A* provable in S_1 is provable in S_2, and $S_1 \equiv S_2$ if $S_1 \subseteq S_2$ and $S_2 \subseteq S_1$. S_2 is said to be stronger than S_1 if $S_1 \subseteq S_2$ but some *A* is provable in S_2 which is not provable in S_1.

With each index or presentation e of an effectively generated system *S* is associated a $\sum_1^0$ formula $Prov_S(x)$ which expresses that *x* is the number of a formula provable from *S*. Then the $\prod_1^0$ sentence $Con_S = \neg Prov_S(\#A_0)$, where A_0 is $(0 \neq 0)$, formally expresses the consistency of *S*. (To be more precise, we should write $Prov_e(x)$ and Con_e for these formulas, for each index *e* of *S*.) According to Gödel's second incompleteness theorem, if *S* is consistent (and includes *PA*) then *S* does not prove Con_S. By hypothesis, Con_S is true in this case, but it may be that $S' = S \cup \{Con_S\}$ is inconsistent; however, if *S* is 1-consistent then so also is S'.

For familiar systems *S*, we come to accept their consistency by informally recognizing that each axiom of *S* is true, hence that each sentence *A* provable from *S* is true. The formal statement of the latter recognition can be given in the language of arithmetic without the truth predicate, simply as

(*) $Prov_S(\#A) \rightarrow A$

For, by Tarski's truth criterion (see Tarski's definition of truth), we have $Tr(\#A) \leftrightarrow A$. The collection of all statements of the form (*) is called the *(local) reflection principle* (Rfn_S) for *S*. Note that the instance $Prov_S(\#A_0) \rightarrow A_0$, with A_0 as above, is equivalent to $\neg Prov_S(\#A_0)$, that is, to Con_S. In other words, the consistency of *S* is a consequence of Rfn_S. As with the addition of Con_S to *S*, 1-consistency is preserved under the expansion of *S* by all instances of Rfn_S for which *A* is in $\prod_2^0$ form. However, we must assume a stronger hypothesis for the full extension, namely: if *S* is true so also is $S \cup Rfn_S$. In the following, we shall be concerned with the extensions of systems given by $S' = S \cup \{Con_S\}$ and $S' = S \cup Rfn_S$, among others. In each of these cases, an index of S' is found effectively from one for *S*.

2 Ordinal logics; Turing's results

Passages from *S* to S' for effectively generated *S*, such as described in the preceding section, can be iterated effectively into the transfinite by using closure under effective unions at limit stages. For this one needs a notion of a system of recursive (or 'constructive') ordinal notations. The first such system was introduced in 1936 by Alonzo Church and Stephen C. Kleene using representations of ordinals by certain terms in Church's lambda calculus (see Lambda calculus). Several variants of this system were subsequently introduced which are all equivalent in the sense that the same ordinals are represented by the different systems. One of these was given by Kleene (1938) in more current recursion-theoretic terms as a subset *O* of *N* together with a partial ordering relation $<_O$ between elements of *O* and a map $a \mapsto \mid a \mid$ from *O* into the ordinals, with the property that if $a <_O b$ then $\mid a \mid < \mid b \mid$.

This representation is not in general unique, that is, the same ordinal α may have many $a \in O$ with $\mid a \mid = \alpha$ (if there are any such $a \in O$ at all). The relation $<_O$ is tree-like in the sense that for each $b \in O$, $\{a\colon a <_O b\}$ is linearly ordered by $<_O$.

The system *O* is also equipped with: (i) a unique element 0_O with $\mid 0_O \mid = 0$, (ii) a recursive operation sc_O from *O* to *O* such that for $a' = sc_O(a)$, $\mid a' \mid = \mid a \mid +1$ and (iii) a recursive operation lim_O from indices of recursive increasing sequences of members of *O* to *O*, such that whenever $\{f\}(n) = a_n$ is in *O* and $a_n <_O a_{n+1}$ for all *n* then $\mid lim_O(f) \mid = lim_n \mid a_n \mid$. Finally, *O* is the least set satisfying the recursive closure conditions (i) - (iii). An ordinal α is said to be recursive if $\alpha = \mid a \mid$ for some $a \in O$; in a sense, these are the effectively countable ordinals. The least non-recursive ordinal is denoted ω_1^{CK}, i.e. the least effectively uncountable ordinal (in the sense of Church and Kleene).

Turing's notion of ordinal logic in his seminal 1939 paper was framed in terms of the lambda calculus; moreover, his notion of a logic was idiosyncratic in that he only considered sets of $\prod_2^0$ formulas. Basically, for Turing, an ordinal logic is any effective association of a logic in his sense with each constructive ordinal notation. The following is an adaptation of Turing's notions to the more current recursion-theoretic framework together with the treatment of logics as formal systems in the sense of §1, which respects all of his results.

Thus, by an *ordinal logic* is meant any partial recursive *E* which associates with each $a \in O$ an index $E(a)$ of an effectively generated formal system S_a, so that $S_a = S[E(a)]$. In the rest of this section S_a is assumed to be a set of sentences in the language of arithmetic with $PA \subseteq S_a$. The logic given by *E* is denoted by $\langle S_a^E \rangle_{a \in O}$, and the superscript '*E*' is suppressed when there is no ambiguity. An ordinal logic is said to be *invariant* (*up through* α) if for all *a*, $b \in O$ with $\mid a \mid = \mid b \mid$ and $\mid b \mid \leq \alpha$) we have $S_a \equiv S_b$; it is said to be *increasing* (*strictly increasing, up through* α) if whenever $a <_O b$ then $S_a \subseteq S_b$ (S_b is stronger than S_a, for $\mid b \mid \leq \alpha$). The ordinal logic is said to be *continuous* if whenever *f* represents an increasing sequence in *O*, then $S_{lim_O(f)} = \cup_n S_{\{f\}(n)}$. Finally, it is said to be *complete* for a class *T* of sentences if for each *A* in *T* there exists $a \in O$ with *A* provable in S_a.

Of special interest are ordinal logics obtained in the following way. We suppose given a partial recursive function (called a *succession principle*) which associates with each index of an effectively generated system *S* an index of a system S' extending *S*. Then given any initially specified system S_0 containing *PA*, one can construct an ordinal logic $\langle S_a \rangle_{a \in O}$ such that:

(i) $S_{0_O} = S_0$, (ii) $S_{sc_O(a)} = S'_a$ for each $a \in O$, and (iii) the ordinal logic is continuous. Then given any notion of correctness $\mathcal{C}$ of systems which applies to S_0 and which is closed under the operation $S \mapsto S'$ and under effective unions, each S_a will belong to $\mathcal{C}$. This holds for the collection $\mathcal{C}$ of 1-consistent systems when $S' = S \cup \{Con_S\}$ and for the collection of true systems when $S' = S \cup Rfn_S$. An ordinal logic constructed satisfying (i)-(iii) is said to be based on the succession principle $S \mapsto S'$; it is strictly increasing if S' is always stronger than *S*, for *S* in $\mathcal{C}$.

For the ordinal logic based on the succession principle $S' = S \cup Rfn_S$, Turing obtained completeness for the class *T* of true $\prod_1^0$ sentences. He asked whether one can obtain completeness for the class of true $\prod_2^0$ sentences; this was answered in the negative by Feferman in 1962 (see §3). In any case, Turing's $\prod_1^0$ completeness result is disappointing (as he himself thought) due to the following general result which is obtained: an ordinal logic cannot be strictly increasing and invariant. Thus a true $\prod_1^0$ sentence provable in some S_a need not be provable in an S_b with $\mid b \mid = \mid a \mid$, and so the completeness result is sensitive to which ordinal notations are used.

Closer inspection of Turing's arguments revealed the following. First of all, his completeness result for true $\prod_1^0$ sentences already holds for the ordinal logic based on the succession principle $S' = S \cup \{Con_S\}$. For this, one associates with each primitive recursive predicate $R(x)$ a number a_R such that a_R is in *O* just in case $\forall x R(x)$ is true, and is such that $\mid a_R \mid = \omega + 1$ and S_{a_R} proves $\forall x R(x)$ when that holds. The notation a_R is not 'natural', unlike our usual notation for $\omega + 1$; it is simply fabricated to do this job. Similarly, an ordinal logic can't be strictly increasing and invariant up through $\omega + 1$; this is best possible for continuous ordinal logics, since one has unique notations in *O* for the natural numbers, hence invariance up through ω.

The original aim for ordinal logics was to obtain an association of effective formal systems S_α with ordinals α in

such a way that $S_{\alpha+1} = S'_{\alpha}$ and $S_{\alpha} = \cup_{\beta<\alpha} S_{\beta}$ for limit α. The definition of ordinal logics as assignments $a \mapsto S^E_a$ for $a \in O$ was necessitated in order to make the association effective. Unless one has some canonical way of associating with each recursive ordinal α an $a \in O$ with $\mid a \mid = \alpha$, the notion of ordinal logic is problematic for its original purpose.

3 Subsequent work

Turing did no further work on ordinal logics after 1939, and the subject went into limbo until it was taken up again in the late 1950s by Georg Kreisel and by Feferman. Part of Feferman's work (1962) consisted in recasting Turing's work from the lambda calculus in ordinary recursion-theoretic terms and in sharpening his results, essentially as described in the preceding section. Finally, attention was restricted to continuous ordinal logics based on a given effective succession principle $S \mapsto S'$; these were called *transfinite recursive progressions of axiomatic theories*. The new work in 1962 was for a recursive progression based on the so-called *uniform reflection principle for a logic* S (RFN_S) which consists of all sentences of the form

(**) $\forall x Prov_S(\#A(\overline{x})) \rightarrow \forall x A(x)$

where for each *n*, $\overline{n}$ is the numeral denoting *n*. Thus (RFN_S) for $A(x)$ expresses that if *S* proves $A(\overline{n})$ for each *n* then $\forall x A(x)$ is true, and so it is a kind of formalized ω-rule. The main result of the author's 1962 work was that for the recursive progression based on the uniform reflection principle, we have completeness for *all* true arithmetical sentences. Moreover, one obtains completeness along suitable paths *P* through *O* (i.e. where *P* is a subset of *O* linearly ordered by $<_O$ which contains a representative for each recursive ordinal). Namely, it was shown that there is a path *P* through *O* (and recursive in *O*) such that for each true arithmetical sentence *A*, there exists $a \in P$ with *A* provable in S_a. Now this completeness result is sensitive to the choice of path because of Turing's non-invariance result, and so the crucial question to be raised is: *what is a natural path P through O*? As of the time of writing, there is still no convincing answer to this question. Familiar natural systems of ordinal representation all close off at a recursive ordinal. One feature of these is that they are defined inductively; furthermore, sets generated by arithmetical inductive definitions are in the class $\prod^1_1$ in the analytic hierarchy (as is the set *O*). (See Recursion-theoretic hierarchies for more information about the analytic hierarchy.) It was shown by Feferman and Spector in 1962 that $\prod^1_1$ paths through *O* exist, but that if *P* is any such path then $\cup_{a\in P} S_a$ is incomplete for true $\prod^0_1$ sentences. Hence if the answer to the question about natural paths requires them to be inductively defined, one will have a very strong incompleteness result for recursive progressions restricted to natural ordinal notations.

4 Autonomous progressions of theories and reflective concepts of proof

In order to characterize informal concepts of proof embodying some kind of stepwise reflective character such as that of *finitist proof* or *predicative proof*, Georg Kreisel suggested in 1958 the use of certain ordinal logics restricted by an *autonomy* (or 'boot-strap') condition. The restriction consists in allowing one to advance to an S_a for $a \in O$ only if it has been proved in some previously accepted $S_b(b <_O a$ that *a* is in *O*. Formally, this requires an extension of the language of arithmetic to admit predicates expressing well-foundedness in one way or another. Not so incidentally, imposition of the autonomy condition partially circumvents the problem of which ordinal notations are to be used in an ordinal logic. However, the purposes are substantially different, since autonomous progressions are incomplete for true $\prod^0_1$ sentences and break off at a recursive ordinal.

For the concept of predicative provability in analysis, Kreisel proposed consideration of an autonomous progression of ramified analytic systems. In 1964, Kurt Schütte and Feferman independently obtained a characterization of the least non-predicatively provable ordinal under this proposal. In a 1991 article, Feferman introduced a notion of *reflective closure* of a system *S* which does not require any of the (problematic) machinery of ordinal logics and showed that the reflective closure of *PA* has exactly the same arithmetical consequences as the autonomous progression of ramified systems. It is proposed there that the reflective closure of a system *S* contains everything one ought to accept if one has accepted the basic notions and principles of *S*.

See also: Logical and mathematical terms, glossary of

SOLOMON FEFERMAN

References and further reading

Davis, M. (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Hewlett, NY: Raven Press.(Includes a reprinting of Turing's 1939 paper on ordinal logics, among other fundamental papers by Gödel, Church, Kleene, Turing and Post.)

Feferman, S. (1962) 'Transfinite Recursive Progressions of Axiomatic Theories', *Journal of Symbolic Logic* 27: 259-316.(Re-examination, technical improvement and extension of Turing's work on ordinal logics as described in §3.)

Feferman, S. (1964) 'Systems of Predicative Analysis', *Journal of Symbolic Logic* 29: 1-30.(Use of autonomous progressions of theories for the characterization of predicate provability in analysis, referred to in §4.)

Feferman, S. (1988) 'Turing in the Land of O(z)', in R. Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press, 113-47.(A readable account of how Turing came to work on ordinal logics, together with a semi-technical exposition of his basic results and of subsequent work.)

Feferman, S. (1991) 'Reflecting on Incompleteness', *Journal of Symbolic Logic* 56: 1-49. (Introduces the concept of reflective closure and shows its equivalence with the autonomous progression for predicative provability; referred to in §4.)

Feferman, S. and Spector, C. (1962) 'Incompleteness Along Paths in Progressions of Theories', *Journal of Symbolic Logic* 27: 383-90.(The incompleteness results described in §3.)

Gödel, K. (1931) 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38: 173-98; trans. 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 592-617.(Gödel's fundamental paper on incompleteness of sufficiently strong formal systems; reprinted in the original and in English translation in Gödel 1986.)

Gödel, K. (1986) *Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort, New York and Oxford: Oxford University Press.

Hodges, A. (1993) *Alan Turing: The Enigma*, London: Burnett Books; New York: Simon & Schuster.(A very readable, absorbing and informative biography of Turing with a description of all his work.)

Kleene, S.C. (1938) 'On Notation for Ordinal Numbers', *Journal of Symbolic Logic* 3: 150-5.(Kleene's system *O* of recursive ordinal notations described in §2.)

Kreisel, G. (1960) 'Ordinal Logics and the Characterization of Informal Concepts of Proof', *Proceedings of the International Congress of Mathematicians, 14-21 August 1958*, Cambridge: Cambridge University Press, 289-99.(Proposed the use of autonomous ordinal logics for the characterization of finitist proof and suggested a similar use for predicative proof; referred to in §4.)

Kreisel, G. (1970) 'Principles of Proof and Ordinals Implicit in Given Concepts', in J. Myhill, R.E. Vesley and A. Kino (eds) *Intuitionism and Proof Theory*, Amsterdam: North Holland, 489-516.(Discussion of the work on autonomous ordinal logics for finitist and predicative provability.)

Turing, A.M. (1939) 'Systems of Logic Based on Ordinals', *Proceedings of the London Mathematical Society*, series 2, 45: 161-228; repr. in M. Davis (ed.) *The Undecidable*, Hewlett, NY: Raven Press, 1965, 155-222. (Turing's fundamental paper on ordinal logics whose notions and results are described in §2.)

Ordinary language philosophy

Ordinary language philosophy is a method of doing philosophy, rather than a set of doctrines. It is diverse in its methods and attitudes. It belongs to the general category of analytic philosophy, which has as its principal goal the analysis of concepts rather than the construction of a metaphysical system or the articulation of insights about the human condition. The method is to use features of certain words in ordinary or non-philosophical contexts as an aid to doing philosophy. The uses in non-philosophical contexts are taken to be paradigmatic; it is in them that meaning lives and moves and has its being. All ordinary language philosophers agree that classical philosophy suffered from an inadequate methodology that accounts for the lack of progress. But proponents of the method do not agree about whether philosophical problems are solved or dissolved; that is, they do not agree about whether philosophical problems are genuine problems for which there are solutions or whether they are merely pseudo-problems, which can at best be diagnosed.

1 The justification of the method

Ordinary language philosophy flourished between 1940 and 1965. It was practised most vigorously at the University of Oxford although it had distinguished practitioners in places as remote from each other as Nebraska in the USA and Adelaide in Australia. There was no comparable movement in Europe, although the most famous philosopher associated with ordinary language philosophy, Ludwig Wittgenstein, was born and brought up in Austria. Ordinary language philosophy evolved out of logical positivism. Although these two shared a negative opinion of traditional philosophy and promoted the virtues of clarity and precision, their differences are more striking. While logical positivism was scientistic and anti-metaphysical, ordinary language philosophy is anti-scientistic and neutral about metaphysics. It is anti-scientistic because it holds that science is a technical discipline that logically depends upon nontechnical concepts. The argument for this position is that science depends upon technical terms; every technical term must be defined either in words that have ordinary meanings or in other technical terms. If the latter, then these terms themselves must eventually be defined in words that have ordinary meanings since technical terms inherently require an explication in ideas that are already understood. Hence all science depends upon a pre-theoretical understanding of the world. As J.L. Austin said, ordinary language is not the last word, but it is the first ([1961] 1970: 185). Ordinary language philosophy is neutral about metaphysics: some practitioners think that metaphysics can be done and others think that it cannot. There is a similar split over whether such traditional conflicts as realism versus nominalism and free will versus determinism have a resolution.

Behind ordinary language philosophy lies an idea that goes back to Socrates, namely that since language is the expression of thought, studying language is a good way to study the concepts that people have. In order to understand what a cause is, or what truth or knowledge is, it makes sense to look at how people use the words ‘cause’, ‘truth’ and ‘knowledge’. This requires looking closely at situations in which such words can be and are used, and when they cannot be and are not, in order to understand the scope and limits of the concepts they express. Such studies invite the question of the general conditions under which the concepts apply; and that connects the method with traditional philosophical concerns. A presupposition of the method is the anti-Cartesian belief that a person’s mental life is not transparent and that people do not have direct access to their own concepts.

Proponents of the method also believed that most philosophically interesting concepts are more complex and rich than usually depicted and that the surest, if not the only, way to uncover these features is to study the actual deployment of them in a variety of linguistic contexts. One of the great failings of other types of philosophy is therefore supposed to be oversimplification. Austin quipped that oversimplification was not so much the occupational hazard of philosophers as it was their occupation. He thought that philosophers concentrated on one or two favoured uses of a word, when the word itself had a much more complex meaning, which was best revealed by investigating the full spectrum of its uses. Austin resisted the idea that there should be one or two categories into which every kind of object should fit. ‘Why, if there are nineteen of any thing, is it not philosophy?’ he asked. There is another philosophical reason for studying ordinary language. Philosophers care to keep separate things separate. This requires an adequate set of distinctions. Austin believed that ordinary language contained many more distinctions than a philosopher could think up sitting in an armchair, both because ordinary language had to function efficiently for people in a vast array of situations and also because it had evolved over a very long period

of time. For the most part, useful distinctions would survive; useless ones would disappear. Wittgenstein thought that one of the big problems with classical philosophy was that philosophical language did not do any real work: ‘philosophical problems arise when language *goes on holiday*’ (1953: §38; original emphasis).

This helps to explain why Austin wrote a paper about ‘Three Ways of Spilling Ink’ ([1961] 1979). Although the choice of spilling ink was intended to be playful and deflating, a wry allusion to the philosophical enterprise itself, it had a serious purpose. The three ways - deliberately, intentionally and on purpose - are three dimensions that are relevant to the issue of assigning responsibility for an action, clearly a philosophical topic; but previous philosophers had not differentiated between them. Through a series of clever examples, Austin illustrated that there are subtle but important differences between acting deliberately, intentionally and on purpose. There is one further reason why Austin chose to study ways of spilling ink: he wanted a non-momentous, not previously studied action in order to avoid the distortions that come from preconceptions. The distinctions operating within ordinary language are not polluted by philosophical prejudices.

2 Paradigm case arguments

One of the most intriguing and seemingly powerful tactics devised by ordinary language philosophers is the ‘paradigm case argument’. Such arguments were typically deployed to refute scepticism, either in general or with respect to specific items, to show that in (at least some of) speech, some determinate knowledge is presupposed. For example, in order to be able to doubt either that people have hands or that something specific is a hand, the word ‘hand’ must have a meaning; and this meaning is or can be taught by pointing to something and saying that it is a hand. In such a situation, there is no room for doubting that that thing is a hand since it is paradigmatically the kind of situation in which the word is defined. Consequently, the ability to doubt that there are hands presupposes that one knows or did know that something is or was a hand. Thus is scepticism refuted. Variations on this argument were developed to prove that human beings know that they have free will, that the past is real, that some things are red, and that induction and deduction are cogent ways of reasoning. Wittgenstein’s remark, ‘It is part of the grammar of the word "chair" that *this* is what we call "to sit on a chair"’ (1958: 24; original emphasis), has sometimes been taken as at least hinting at the strategy of paradigm case arguments.

However, there are general objections to paradigm case arguments. Even if there are situations in which people say ‘This is a hand (or red)’ in a paradigmatic case of hand-naming (or red-attributing), it does not follow that the intended object of reference is a hand (or red). If the speaker is being deceived by a Cartesian demon or is the subject of a neurosurgeon’s experiment, then what the speaker says may be false, and they may have no way of knowing that it is false. From the fact that a person believes that they are viewing an object under ideal conditions, it does not follow that they are.

Another objection is that the arguments seem to be too strong in the sense that they can be applied to prove that certain things exist that we do not believe exist. For example, one way to explain the meaning of the word ‘witch’, which historically may actually have been used, would have been to point to some woman, probably an old, poor, friendless and unattractive one, and to say ‘That is a witch’. For that community of speakers, that woman would have been a paradigm case of a witch. But very few people today would want to hold that she was in fact a witch. Although proponents of paradigm case arguments wanted to restrict the words that could be taught by paradigm examples, it is dubious that they discovered a principled way of doing this.

3 Criticisms

Ordinary language philosophy tended to offend the unconverted. It often gave the impression of being trite or vacuous; of being concerned merely with verbal issues or semantics. Explicit treatments of how many ways there were to spill ink, or the difference between shooting a donkey accidentally or by mistake did not endear critics, who thought the examples highlighted the triviality of the enterprise and demeaned the profession of philosophy. One philosopher complained, ‘I don’t care how truck drivers talk’, and another that he did not care how Oxford dons talked. This covers the spectrum of one dimension of criticism.

A more general objection is that ordinary language philosophy talks only about words and not things; it seems to glory in verbal disputes and questions of semantics. Austin would have disagreed:

> one thing needs specially emphasizing to counter misunderstandings. When we examine what we should say

> when, what words we should use in what situations, we are looking again not *merely* at words… but also at the realities we use the words to talk about: we are using a sharpened awareness of words to sharpen our perception of, though not as the final arbiter of, the phenomena. For this reason, I think it might be better to use, for this way of doing philosophy,… 'linguistic phenomenology', only that is rather a mouthful.
>
> ([1961] 1970: 182; original emphasis)

Another complaint targeted the Oxonian ideal that philosophers should speak clearly and precisely; it was summarized in the slogan 'Clarity is not enough'. But the ordinary language philosophers never said it was. Clarity was one ideal, preached as a reaction against the German and British idealism of the nineteenth century; not the only one.

The two strongest criticisms were that neither the methodology of ordinary language philosophy nor its results were systematic. These criticisms are independent. One could have a systematic way of studying ordinary language, which would yield the result that ordinary concepts lack the kind of structure that is typically counted as systematized knowledge. And one could hold that systematic results are obtainable using a non-systematic procedure. It is not easy to evaluate these criticisms, and ordinary language philosophers did not agree about the proper response.

4 Systematic methodology and results

Far from being moved by the charges that they did not have a systematic methodology and did not seek systematic results, Wittgenstein and other ordinary language philosophers, including O.K. Bouwsma (1965), would have vaunted them. They thought that one of the chief pitfalls of philosophy was the belief in and desire for systematization, a defect attributable to the desire to make philosophy a science. Wittgenstein always held that philosophy is not like science; it is either higher or lower. Unlike science, which always aims at change in the name of progress, ordinary language and the thinking that it expresses are fine as they are and do not need improvement: 'Philosophy may in no way interfere with the actual use of language, it can in the end only describe it. For it cannot give it any foundation either. It leaves everything as it is' (1953: §124). Behind this idea is the belief, shared by Wittgenstein and G.E. Moore, that ordinary language is the expression of a common sense view of the world that cannot be wrong in its basic components.

Other ordinary language philosophers would deny that they apply no systematic method. Austin adumbrated a methodology explicitly in several essays, notably 'A Plea for Excuses' ([1961] 1970), and the method was refined by Avrum Stroll (1973). In any case, it may be unfair to require a strict methodology of ordinary language philosophy since the idea that scientists proceed according to a strict methodology is dubious.

Even if there were no systematic methodology, that would not preclude an orderly investigation of a phenomenon; one dictated by the phenomenon itself. P.F. Strawson treated the foundations of logic in an orderly way, in the course of arguing that the project of formal logic to codify inferences in general must fail because 'ordinary language has no exact logic'. Although Strawson does not employ a systematic methodology in a technical sense, he is not averse to systematic results. In *Individuals* (1959), he presents what he takes to be the general conceptual scheme by which humans conceive all of reality.

Neither proceeding with a fixed methodology nor proceeding systematically determines whether one's results will be systematic. Stroll used a systematic methodology to study the concept of surfaces (1988); and one of his central results is that there is no one concept of surface, but many, which sometimes intermingle. This result is part of a broader claim that humans categorize the world in a piecemeal fashion. Thus Stroll would deny the fairness of the claim that ordinary language philosophy has no methodology and would hold that the application of his method reveals that there is no single, neat system that humans use to think about surfaces. And yet Stroll does not eschew generalization. He formulates theorems, as general as the subject allows, about various concepts of surfaces, each of which has limited applications.

The most lasting results of ordinary language philosophy have been in the philosophy of language. While there is a difference between ordinary language philosophy, which tries to solve philosophical problems by studying bits of linguistic usage, and the philosophy of language, which tries to give a general account of what language is and how it functions, interest in one is often accompanied by interest in the other. Austin and Grice devised powerful

theories of speech acts and conversation, respectively, complete with technical vocabularies (see Speech acts). John Searle and Zeno Vendler, who were trained by these philosophers and practised ordinary language philosophy early in their careers, have also produced lasting work in the philosophy of language.

At the opposite end from searching for systematic results is the practice of deconstruction. Although it is usually associated with continental philosophy, a good case can be made for applying the term to some ordinary language philosophy. Deconstruction is the tactic of showing that some philosophical distinction is hopelessly flawed, usually because the characterizations of the terms are incoherent. In *Sense and Sensibilia* (1962), Austin showed that the distinction material/immaterial was bogus and led to mistakes. In *Surfaces*, Stroll showed how the apparently proper distinction direct/indirect does not have general application. In particular, this distinction almost never makes sense in application to perception, and its misapplication has led to serious mistakes in the theory of perception.

A renewed interest in metaphysics, partially attributable to Strawson's book *Individuals*, the development of powerful logical theories that could be used to describe natural languages, and a renewed interest in scientific treatments of the mind contributed to the decline of ordinary language philosophy in the late 1960s. The techniques and insights of ordinary language philosophy were no longer conspicuous in philosophy during the 1990s, but some of them had been incorporated into it.

See also: Ordinary language philosophy, school of

A.P. MARTINICH

References and further reading

Austin, J.L. (1961) *Philosophical Papers*, ed. J.O. Urmson and G.J. Warnock, London: Oxford University Press, 2nd edn, 1970, 3rd edn, 1979.(Collected essays of the most prominent Oxford practitioner of ordinary language philosophy.)

Austin, J.L. (1962) *Sense and Sensibilia*, New York: Oxford University Press; repr. 1979. (An attack on some standard philosophical treatments of perception in the spirit of ordinary language philosophy.)

Bouwsma, O.K. (1965) *Philosophical Essays*, Lincoln, NE: University of Nebraska Press; repr. 1982.(A collection of Bouwsma's distinctive way of doing philosophy.)

Caton, C. (ed.) (1963) *Philosophy and Ordinary Language*, Urbana, IL: University of Illinois Press.(Classic essays in ordinary language philosophy.)

Donnellan, K.S. (1967) 'Paradigm-Case Arguments', in *The Encyclopedia of Philosophy*, ed. P. Edwards, New York: Macmillan, vol. 6, 39-44.(Includes a good bibliography.)

Flew, A. (ed.) (1965) *Logic and Language* (first and second series), Garden City, NY: Anchor Books; repr. Oxford: Blackwell, first series, 1978; second series, 1973. (Classic essays in ordinary language philosophy.)

Flew, A. (ed.) (1966) *Essays in Conceptual Analysis*, London: Macmillan.(Standard examples of ordinary language philosophy.)

Grice, H.P. (1958) 'Postwar Oxford Philosophy', repr. in *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, 1989.(A defence of the method of ordinary language philosophy.)

Martinich, A.P. (1991) 'Analytic Phenomenological Deconstruction', in *Certainty and Surface in Epistemology and Philosophical Method*, ed. A.P. Martinich and M. White, Lewiston, NY: Mellen, 165-84.(Relates ordinary language philosophy to late twentieth-century continental philosophy.)

Ryle, G. (1971) *Collected Papers*, London: Hutchinson, 2 vols.(Ryle's essays range more broadly than ordinary language analysis.)

Strawson, P.F. (1952) *Introduction to Logical Theory*, London: Methuen.(An ordinary language analysis of the foundations of logic.)

Strawson, P.F. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen; new edn, London: Routledge, 1990.(A descriptive metaphysics by an ordinary language philosopher.)

Stroll, A. (1973) 'Linguistic Clusters and the Problem of Universals', *Dialectica* 27: 219-59.(The most complete and cogent statement of the method of ordinary language analysis.)

Stroll, A. (1988) *Surfaces*, Minneapolis, MN: University of Minnesota Press.(A powerful application of the method to an issue relevant to both metaphysics and epistemology.)

Stroll, A. (1994) *Moore and Wittgenstein on Certainty*, New York: Oxford University Press.(Explores the connection between ordinary language and common sense beliefs.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell, 2nd edn, 1958; repr. 1973.(The most

famous work of ordinary language philosophy.)
Wittgenstein, L. (1958) *The Blue and Brown Books*, Oxford: Blackwell.(A work in which Wittgenstein seems to hint at the strategy of paradigm case arguments.)

Ordinary language philosophy, school of

The label 'ordinary language philosophy' was more often used by the enemies than by the alleged practitioners of what it was intended to designate. It was supposed to identify a certain kind of philosophy that flourished, mainly in Britain and therein mainly in Oxford, for twenty years or so, roughly after 1945. Its enemies found it convenient to group the objects of their hostility under a single name, while the practitioners thus aimed at were more conscious of divergences among themselves, and of the actual paucity of shared philosophical doctrine; they might have admitted to being a 'group' perhaps, but scarcely a 'school'. The sharp hostility which this group aroused was of two quite different sorts. On the one hand, among certain (usually older) philosophers and more commonly among the serious-minded public, it was labelled as philistine, subversive, parochial and even deliberately trivial; on the other hand, some philosophers (for instance, Russell, Popper and Ayer), while ready enough to concede the importance in philosophy of language, saw a concern with ordinary language in particular as a silly aberration, or even as a perversion and betrayal of modern work in the subject.

How, then, did 'ordinary language' come in? It was partly a matter of style. Those taken to belong to the school were consciously hostile to the lofty, loose rhetoric of old-fashioned idealism; also to the 'deep' paradoxes and mystery-mongering of their continental contemporaries; but also to any kind of academic jargon and neologism, to technical terms and aspirations to 'scientific' professionalism. They preferred to use, not necessarily without wit or elegance, ordinary language. (Here G.E. Moore was an important predecessor.) Besides style, however, there were also relevant doctrines, though less generally shared. Wittgenstein, perhaps the most revered philosopher of the period, went so far as to suggest that philosophical problems in general actually consisted in, or arose from, distortions and misunderstandings of ordinary language, a 'clear view' of which would accomplish their dissolution; many agreed that there was some truth in this, though probably not the whole truth. Then it was widely held that ordinary language was inevitably fundamental to all our intellectual endeavours- it must be what one starts from, supplying the familiar background and terms in which technical sophistications have to be introduced and understood; it was therefore not to be neglected or carelessly handled. Again it was urged, notably by J.L. Austin, that our inherited everyday language is, at least in many areas, a long-evolved, complex and subtle instrument, careful scrutiny of which could be expected to be at least a helpful beginning in the pursuit of philosophical clarity. It was probably this modest claim- overstated and even caricatured by its detractors- which was most frequently supposed to be the credo of ordinary language philosophers. It was important that Russell - like, indeed, Wittgenstein when composing his Tractatus Logico-Philosophicus (1922) - firmly believed, on the contrary, that ordinary language was the mere primitive, confused and confusing surface beneath which theorists were to seek the proper forms of both language and logic.

1 The 'ordinary language' label

It is not at all easy to discern the historical origin of the label 'ordinary language'. Other philosophical schools do not present this difficulty; pragmatists, for instance, or logical positivists consciously, even proudly, identified themselves as such. But being an ordinary language philosopher seems always to have been something of which one was accused, rather than something which one claimed. The philosophers so designated were a roughly recognizable group, but they would certainly have objected to the suggestion that they formed a 'school'. That suggestion seems to have originated among those seeking to set up a convenient target to attack; those attacked would mostly not have claimed the label for themselves, and indeed would probably have objected, not disingenuously, that it was unclear to them what it was supposed to mean. We must first ask what this group (rather than school) comprised, and why it was regarded in some quarters with such remarkable hostility.

After anticipatory stirrings in the pre-war years, the philosophical 'tendency' in question flourished from 1945 for roughly the next twenty years, being confined almost wholly to the sphere of philosophy in English, and within that sphere chiefly, if only through weight of numbers, to philosophy in Oxford. 'Oxford philosophy' was, in fact, an alternative label which had some currency at the time. Here the position of Wittgenstein is a perturbing factor. Among 'Oxford philosophers' he was, well before the publication of *Philosophical Investigations* in 1953, the most esteemed and influential of contemporaries; on the other hand he lived and worked, somewhat reclusively, in Cambridge rather than Oxford, and also (less trivially) himself regarded Oxford as 'a philosophical desert', the meagre fruits of which were to him utterly distasteful. The explanation of this odd state of affairs is partly that

Wittgenstein was temperamentally incapable of attachment to any group of which he was not the acknowledged leader; and partly that, while ‘Oxford philosophy’ was certainly closely akin in substance to Wittgenstein’s later work, its characteristic cool, ironic urbanity of manner was odious to him, wholly lacking in the *Sturm und Drang* which really serious philosophizing entailed. Thus it came about that, while Wittgenstein was always conspicuous among those arraigned as ‘ordinary language philosophers’, he himself would furiously have disclaimed any kinship with the other targets of that critical fire.

Why did ordinary language philosophy incur such hostility? There were two quite distinct reasons. The first, put succinctly, was that in certain quarters it was confusedly regarded as a species of, or a descendant of, logical positivism (see Logical positivism). Old-fashioned philosophers, and many high-minded members of the general reading public, had of course been outraged - as it was intended that they should be - by the brisk dismissiveness of, for example, A.J. Ayer’s *Language, Truth and Logic* of 1936; and it was easily though wrongly supposed for many years thereafter that ‘young’ or ‘modern’ English-language philosophers were all more or less of the same deplorable party - iconoclastic, sceptical, irreverent, subversive of true seriousness - all logical positivists. More significant, however, was the hostility of some leading figures from well within the professional circles of philosophy itself: Russell perhaps most conspicuously, but also, for example, Popper and, perhaps less vehemently, Ayer. This will be explored further in §3 below; but the basic charge here was that ‘Oxford’ or ‘ordinary language’ philosophy was an aberrant, trivializing perversion of good philosophical practice, substituting, in place of honest theorizing and argument, pedantic scrutiny of intrinsically uninteresting detail. The target-figure here seems to have been a most unlikely-looking amalgam of Wittgenstein and J.L. Austin - the former cast as the heresiarch, perhaps, the latter as the pedant. The most furious onslaught was made, with Russell’s blessing, in Ernest Gellner’s *Words and Things* (1959).

2 General character of the ‘school’

To a more than trivial extent, membership of the group in question was a matter of style. There was a conscious hostility to the lofty, rather loose rhetorical manner of, for example, its idealist predecessors; this did not preclude - as the writings of, say, Gilbert Ryle sufficiently illustrate - elegance and wit, but it ruled out eloquence and uplift. (Here, once more, Wittgenstein is a perturbing factor. Though his presence broods, so to speak, over the group, the style of his own writings is absolutely idiosyncratic, personal to himself, seemingly owing nothing to any predecessor, and unlike that of any contemporaries, even of his own disciples.) There was an even more emphatic distaste for the riddle-spinning, paradox-delighting ‘deep’ discourse of most contemporary continental philosophers, with whom, indeed, any kind of academic communication was neither sought nor, probably, practicable. (In this particular distaste there was, I think, an element of moral disapproval; it was felt that the weird, mind-boggling pronouncements of some continental sages were not only unprofitable but largely bogus - an intellectual fraud.) So far, the doggedly plain-man manner of G.E. Moore was a significant influence; nor, so far, does the ‘ordinary language’ style much diverge from that of logical positivism (see Moore, G.E. §3). There is, however, a further point to be made. It was characteristic of logical positivism that, while certainly hostile to philosophical uplift and eloquence, it regarded mathematics and the physical sciences with special reverence; and this tended in philosophical writings towards a liking for the highly regimented handling of batteries of specially introduced technical terminology. This too the ‘ordinary language’ school typically avoided. This was not a matter of dogma: Austin, for example, was entirely happy to employ invented technical terms in his work in the philosophy of language when he felt a clear need for them; it was rather that members of the group did not particularly like philosophical technicality or neologism, suspecting that it was usually unnecessary and often did more harm than good. A large part of the reason for applying the ‘ordinary language’ label was that it was in entirely - and deliberately - ordinary language that those so identified almost always wrote and debated.

A point of more philosophical substance was that members of the ‘ordinary language’ group were unfriendly, at the very least, towards the idea of philosophical ‘theories’. In practice this took the form of a deliberate abstention from the pursuit of generality, of wide-ranging explanations or justifications, of purportedly systematic examination of whole families of concepts; as Bernard Williams has observed, in the relevant period ‘"Piecemeal" was a term of praise.’ Limited issues were to be addressed strictly one at a time, particular concepts brought, as it were, individually under the microscope, and each thoroughly dealt with before going on to another. It was often remarked that the typical ‘Oxford’ philosopher of the post-war years recorded his researches in discrete papers and articles rather than in whole books; and it is worth noting that as early as 1933 the periodical *Analysis* was

launched with the specific aim of promoting and publishing papers far shorter, far more tightly circumscribed in intention than the usual offerings to established philosophical journals.

Diverse motives were at work here. In the case of Wittgenstein himself and his closest followers, theory was rejected as being fundamentally out of place in, inappropriate to, philosophy; it was held that the problems of philosophy were essentially not the sort of perplexities for which theory-construction could offer any solution - even, perhaps, that to talk of 'solution' was itself to put the case in false terms. This must be further considered in what follows. For Austin, however, and probably for many others, the preference for 'piecemeal' working was a simpler matter of prudently limited ambition, of not biting off more than one could analytically chew. It was felt - in this following Moore - that much traditional philosophy had come to grief through the attempt to settle far too much far too quickly; impatient to get to the great work of theory-construction, philosophers had too often simply neglected or hastily over-simplified or carelessly misrepresented the relevant detail, with the inevitable consequence that the theories turned out to be chimerical and often laughably fragile. It was not that theory was held to be intrinsically out of place; it was simply that, in the current state of philosophical affairs, it was premature and over-ambitious. What was felt to be needed at the time was patient spadework - walking, even plodding if necessary, before trying to run. It was probably in this respect that the ordinary language group was most clearly not to be identified with logical positivism; for the latter, after all, was itself a highly ambitious theory, and fundamentally a theory of meaning, which was indeed supposed capable of settling, rigorously and in fairly short order, almost everything. The first chapter of *Language, Truth and Logic* is entitled 'The Elimination of Metaphysics'; the last is entitled 'Solutions of Outstanding Philosophical Disputes'. It was Austin in particular who regarded such sweeping claims as not only injudicious and unjustified, but as rather disreputable.

3 Philosophical tenets

In dealing more directly with the distinctive philosophical tenets of the ordinary language group, one must begin with the later views of Wittgenstein - not, as has been noted in §1 above, because he would have claimed or conceivably tolerated classification as a member of that group, but because first, he was a pervasive, even dominant influence on those who were its members, and second, in his later work 'ordinary' language was accorded a particularly clear and fundamental place.

The story, briefly, is this. In the early work that issued in the *Tractatus Logico-Philosophicus* of 1922, Wittgenstein had sought - then following quite closely lines laid down by Russell - to work out and exhibit the essence of all language, its really fundamental character and mode of working; and this he found in the fundamental relation, on which all meaning depended, between what he called elementary propositions and atomic facts, a relation which he called 'picturing'. But of course the propositions that occur in the actual use of language are, he held, never themselves elementary propositions. So what are they? The answer seemed to lie ready to hand: they are - they could only be - truth-functional compounds of elementary propositions, constructions, in accordance with the principles of Russellian logic, from the necessary though nearly undetectable basic material. It was, of course, to be not only admitted but emphatically affirmed, that languages in their ordinary use do not look at all like this; no one, if asked simply to observe a language and describe it, would arrive at anything resembling this description. But the reason for this was held to be that the surface of language, its easily-visible features, were radically confusing and misleading as to its essential character; the theorist must seek to find, beneath or behind mere surface vagueness and superfluities, the essential structure that - according to theory - must be there (see Russell, B. §9; Wittgenstein, L.J.J. §§3-5).

Briefly, over the next ten years or so, Wittgenstein came to two conclusions: first, that the theory of the *Tractatus* was not in fact the rigorous unveiling of fundamental truth from behind misleading surface phenomena, but was rather the perverse foisting upon the real world of artificial and fantastic theoretical preconceptions; and second, crucially, that this was the general character of philosophical theorizing, and the genesis of all philosophical problems and perplexities. 'Philosophical theories are a product of the imagination, and they offer us simple, but seemingly profound pictures, which blind us to the actual complexities of language. The new philosophy is an organized resistance to this enchantment, and its method is always to bring us back to the linguistic phenomena, with which we are perfectly familiar, but which we cannot keep in focus when we philosophize in the old way' (Pears 1971: 16). Or in Wittgenstein's own words: 'What *we* do is to bring words back from their metaphysical to their everyday usage'; and again: 'We must do away with all *explanation,* and description alone must take its

place' (1953: §116, 109). This is 'ordinary language' philosophy in the fairly strong and precise sense that it holds that a clear view, a plain description, an uncorrupted grasp of ordinary language will exorcize, so to speak, philosophical perplexity, which arises when familiar concepts are, at the supposed behest of 'theory', crudely mishandled: 'the idea is that in philosophy to theorize is to falsify, and the facts about language are offered as a corrective' (Pears 1971: 38). This thesis is very close to that of Wittgenstein's near-contemporary, Gilbert Ryle, as set out in his 'Systematically Misleading Expressions' of 1932 and developed in his later description of philosophy as, essentially, the rectification of 'category mistakes' (see Ryle, G. §1). Almost all members of the ordinary language group accepted the thesis, more or less fervently, as an illuminating contribution to the handling of some of the problems of philosophy, though few would have ventured to maintain, as Wittgenstein at least appeared to do, that every philosophical problem had this linguistic origin and was amenable to this linguistic treatment. Austin in particular was always explicitly opposed to any suggestion that 'philosophical' problems were of any single sort, or that any simple strategy or method would be appropriate to every case.

A different point may be made by reference again to Ryle. In discussing what he calls the 'very misleading phrase "ordinary language",' he says, 'I do think that a philosopher's arguments, when couched in technical terms, tend either to elude one or in fact to go agley. It's much easier to catch a philosopher out, including oneself, if he is not talking in technical terms with which you are unfamiliar, and the most important thing about a philosopher's arguments is that it should be as easy as possible for other people, and especially for himself, to catch him out if he can be caught out' (Magee 1971: 111). This fairly modest idea, which surely owes a good deal to the practice, if not to any stated doctrine, of G.E. Moore, was very widely accepted. It would not have been held that there was no place in philosophy, ever, for invented technical terminology, but it certainly was held that such terminology should always be regarded warily, even with suspicion. Even, for instance, the fairly harmless-looking term of art 'sense-datum' was held to be a potentially damaging stumbling-block in the philosophy of perception (see Austin's *Sense and Sensibilia* (1962)).

A rather different, but again not particularly controversial tenet was that 'acquiring the use of technical vocabularies can only take place against a background of ordinary language…ordinary language is the basis on which technical language is erected to begin with' (Caton 1963: ix). Roughly, technical terms are defined, some no doubt in other technical terms, but ultimately in non-technical, everyday terms. Unless we are clear and careful in our handling of those terms, hoped-for technical refinements will be liable to lapse into confusion.

One should bring in here a related but, for a time, far more ambitious thesis, which became known in the literature as 'the paradigm case argument'. This was the claim that in the general interests of lucidity one should not only accurately describe the ordinary use of expressions; one can also appeal to the fact that an expression has an ordinary use as a cogent argument in itself. If an expression is in established, ordinary use, then its meaning will be settled (and learned) by reference to the standard or paradigm cases in which it is applied by speakers of the language; and if it really is in ordinary use, there must be such cases. But this general truism may seem to have highly important implications. The verb 'know', for example, is in common use; we can all recognize 'paradigmatic' cases in which a person would be said to know something. If so, that is what knowledge is, and in those cases there is knowledge. One need say no more in refutation of scepticism. Or consider the case of denial of freedom of the will. It will not - cannot - be denied that there are paradigmatic, 'meaning-illustrating' circumstances in which people are ordinarily said to do things freely, of their own free will. If so, that is what free will is, and there *is* free will. To deny that there is free will, while conceding, as of course one must, that certain instances actually occur which are standard examples of what 'acting freely' ordinarily means, is not merely to take up a position which no sensible person would accept: it is to propound a thesis which could not be true, which seems indeed self-contradictory.

One can see why this form of argument was, for a time, highly appealing. It was, in the preferred style of the period, satisfactorily deflationary. It was both simple and, at least, plausible. It seemed that ordinary language, at this point at least, could do real philosophical work. Nevertheless, though the paradigm case argument was accorded much powder and shot by such vehement critics as Ernest Gellner (1959: 30-7), it probably never had many wholehearted supporters. It was not only that it seemed altogether too quick, even slick, uncomfortably suggestive of some philosophical sleight-of-hand. It had more serious deficiencies. In many cases it seems not really to engage with the problem at issue. In the case of free will, for example, even if it were taken as establishing that there are - must be - 'paradigmatic' instances of persons acting freely, perhaps not much is

achieved, for the perplexed philosopher is probably not inclined to deny that that is so; his worry is that he cannot see how it could be so, given other things that we wish to preserve, such as scientific principles; and on that the quick argument sheds no light at all. But no doubt the most serious deficiency was always that the quick inference from ‘ordinary use’ to the existence of actual instances was highly insecure. That the use of an expression is current and well-established in the language does not show that such use is, in all or even any respects, justified and legitimate. There was at one time, and no doubt still is, a meaning and ordinary use for the words ‘witch’ and ‘leprechaun’, but we are not obliged to conclude that there are or were witches or leprechauns. The reality of miracles is not deducible from the currency in English of the word ‘miracle’.

Less ambitious and less controversial was the view, associated particularly with J.L. Austin, that ordinary language can sometimes be a valuable philosophical resource (see Austin, J.L. §2). At least in areas of general, perennial human concern - in perception, for example, or human action and behaviour - natural languages will be found to have evolved highly complex and often subtle (though familiar) vocabularies and idioms, so devised as to mark genuine distinctions that have been found to be of importance. If so, then it will surely be prudent for philosophical discussion in such areas at least to begin with careful and comprehensive scrutiny of the relevant ‘ordinary’ vocabulary, the distinctions and connections found there being probably ‘more sound…and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our armchairs of an afternoon - the most favoured alternative method’ (Austin 1960: 182).

It was probably this view of Austin’s which was most commonly taken to be the central characteristic credo of the ordinary language group, and not wholly unfairly - it was a view very widely accepted both in theory and practice. Critics, however, too often fanned the flames of their hostility by grossly overstating the claim that was being put forward. It was suggested that Austin was peddling a comprehensive panacea for all philosophical ills - suggesting that exhaustive examination of ordinary language would be, for the resolution of every philosophical problem, the only fruitful and necessary procedure. But this was a crude caricature. Austin was always insistent that the problems assembled under the name ‘philosophy’ were of very many quite different sorts, needing quite different treatments. What he said, and many agreed with him, was that, in suitable cases, scrutiny of the resources of ordinary language would probably be one good way to begin.

Finally, however, it may be that what most firmly united the members of this ‘school’ or group was the simple belief that ordinary everyday language was interesting, deserving of attention, worth investigating. A famous illustration of this would be the contrast between Russell’s article ‘On Denoting’ of 1905 and Strawson’s ‘On Referring’ of 1950. Both address the same issue - certain problems raised by the use of so-called definite descriptions, such as ‘the author of *Ulysses*’. But whereas Russell glances rather perfunctorily at what he took here to be the muddled primitive practices of ordinary language and hurries on to the construction of his own Theory of Descriptions, Strawson finds our own everyday practices of referring to things both interesting and important, and indeed such that, if accurately described, they reveal as unnecessary the revisionary formalization which Russell attempted. There is a somewhat similar contrast between Ayer’s *The Foundations of Empirical Knowledge* (1940) and Austin’s *Sense and Sensibilia*. Ayer’s real interest is in epistemology and the theory called phenomenalism; ‘what we say’ in the field of perception seems to him fairly trivial (see Ayer, A.J. §3). Austin, by contrast, finds our everyday language on perceptual matters to be quite rich, complex and subtle, and again (he perhaps hints) such that, if we accurately describe it, we may in the end see no need for the invention of new terminology or the continuation of theories. I would mention here also A.R. White’s *Modal Thinking* (1975), a book which deliberately eschews the technicalities and problems of formal modal logic in favour of detailed examination of the use and interconnections of such everyday expressions as ‘possible’, ‘can’ and ‘may’.

4 Later developments

It was perhaps inevitable that the philosophical ‘frame of mind’ described here should in time not so much wither away as evolve into something different. Essentially, what proved unstable in the ordinary language outlook was the rejection of theory, the deliberate abstention from generalization in favour of the ‘piecemeal’, case-by-case scrutiny of discrete particulars. The desire to generalize, to unify, to understand, to find patterns and connections and system seems too profoundly characteristic of philosophy to be long suppressed. Furthermore, the question came increasingly to be raised: why should it be suppressed? It had been held, notably by Austin and in practice by many others, that generalization and theory-construction was dangerously over-ambitious and premature; but

surely it could not be premature for ever; and ambition need not be wholly renounced. Wittgenstein, indeed, had gone further. He had rejected all theory, all 'explanation', as not merely premature but as radically misconceived and inappropriate to philosophy, but that depended on his own view of what philosophical problems are and what philosophy is - a view which, as remarked above, was never very widely accepted as the whole truth. Thus philosophers came increasingly to ask (perhaps some had never really ceased to ask): is small-scale piecemeal description really enough? Why should one not try to explain, to interconnect, to paint a general picture? Why, in short, should one not have a theory? Such a progression seems particularly clear in the work of P.F. Strawson (see Strawson, P.F. §§2-8). We have, in 'On Referring', the careful delineation of part of the use of everyday language, its use in referring to individual items. But this led on, in *Individuals* (1959), to the theory that a certain class of items - substantial spatiotemporal individuals - provides the basic targets of reference and subjects of predication. And in due course the further questions of why that should be so and whether it might have been otherwise led into the quite self-consciously quasi-Kantian theorizing of *The Bounds of Sense* (1966). But that is merely one good example of what occurred very generally - namely, from about 1960 onwards, a massive resurgence of the readiness to theorize, to seek not merely to observe but to understand, not merely to describe but to explain. The work of the 'ordinary language school' will perhaps have been justified if the products of this resurgence are found to be less crude, less obscure, less hastily fabricated, and less fragile than have been all too many philosophical theories of the past.

See also: Ordinary language philosophy

GEOFFREY WARNOCK

References and further reading

Austin, J.L. (1962) *Sense and Sensibilia*, Oxford: Clarendon Press.(Austin uses an analysis of our ordinary language of perception to criticize the 'sense-datum' theory of Ayer (1940).)

Austin, J.L. (1960) *Philosophical Papers*, Oxford: Oxford University Press, 3rd edn, 1979. (The paper 'A Plea for Excuses' (pages 175-204) offers unusually explicit 'ordinary language' doctrine.)

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz.(A brief and brilliant exposition of logical positivism.)

Ayer, A.J. (1940) *Foundations of Empirical Knowledge*, London: Macmillan.(The main target of Austin's *Sense and Sensibilia*.)

Caton, C.E. (ed.) (1963) *Philosophy and Ordinary Language*, Urbana, IL: University of Illinois.(Contains a number of representative articles, including 'Ordinary Language' by Ryle, and Strawson's 'On Referring'. It also has a good bibliography.)

Flew, A. (ed.) (1965) *Logic and Language*, New York: Doubleday.(A useful collection in one volume of two anthologies of representative papers, originally published by Blackwell in 1951 and 1953.)

Gellner, E. (1959) *Words and Things*, London: Gollancz.(A fiercely polemical, even intemperate attack on 'linguistic philosophy'. Russell's strongly supportive introduction is historically interesting.)

Magee, B. (ed.) (1971) *Modern British Philosophy*, London: Secker & Warburg.(A series of dialogues both with and about contemporary philosophers, originally done for radio; clear and useful, and with a valuable appendix of suggested reading.)

Magee, B. (ed.) (1978) *Men of Ideas*, London: BBC.(A collection of conversations, originally televised, that with Bernard Williams (pages 134-49) being particularly relevant and useful.)

Passmore, J. (1957) *A Hundred Years of Philosophy*, London: Duckworth.(Contains an instructive chapter on 'Wittgenstein and Ordinary Language Philosophy' (pages 425-58).)

Pears, D. (1971) *Wittgenstein*, London: Collins, Modern Masters series.(An admirable brief account.)

Russell, B. (1905) 'On Denoting', *Mind* new series 14: 479-93; repr. in B. Russell, *Logic and Knowledge*, London: Allen & Unwin, 1956, 41-56. (Contains Russell's 'theory of descriptions', initially regarded as a classic piece of linguistic analysis, but later the target of Strawson (1950).)

Russell, B. (1959) *My Philosophical Development*, London: Routledge, 1993.(The chapter, 'Some replies to criticism' is fiercely critical of 'ordinary language philosophy'.)

Ryle, G. (1971) *Collected Papers 2*, London: Hutchinson. ('Systematically Misleading Expressions' (1932), referred to in §3, is on pages 39-62. See also 'Categories', 170-84.)

Strawson, P.F. (1950) 'On Referring', *Mind* new series 59; repr. in P.F. Strawson, *Logico-Linguistic Papers*, London: Methuen, 1971, 1-27.(Famous ordinary-language critique of Russell (1905).)

Strawson, P.F. (1959) *Individuals*, London: Methuen.(Strawson here moves beyond the analysis of ordinary

language to ‘descriptive metaphysics’.)
Strawson, P.F. (1966) *The Bounds of Sense*, London: Methuen.(A critical discussion of Kant’s masterpiece. The connections between Strawson’s works above are usefully discussed in B. Magee, *Modern British Philosophy*, 117-24.)
Urmson, J.O. (1956) *Philosophical Analysis*, Oxford: Clarendon Press.(Part III, ‘The Beginnings of Contemporary Philosophy’ is relevant; the book is also one of Russell’s targets in *My Philosophical Development*.)
White, A.R. (1975) *Modal Thinking*, Oxford, Blackwell.(An unusually single-minded example of the ‘ordinary language’ method at work.)
Wittgenstein, L.J.J. (1922) *Tractatus Logico-Philosophicus*, London: Routledge & Kegan Paul, 1961.(A new translation from the German of this desperately difficult work.)
Wittgenstein, L.J.J. (1953) *Philosophical Investigations*, Oxford: Blackwell.(Not an easy work, but essential reading.)

Oresme, Nicole (*c.*1325-82)

Nicole Oresme, a French thinker active in the third quarter of the fourteenth century, occupies an important position in late medieval natural philosophy. He was especially notable for his mathematical approach, in which he represented the intensities of qualities and of speeds by geometrical straight lines, which allowed them to be 'plotted' in principle against both distance and time. He held that the shapes of the resulting graphs would then have explanatory force in the manner of ancient atomism, but, like the latter, his doctrine had a weak empirical basis. His graphical representations of speed have been compared to those later given by Galileo, but there are no grounds for positing influence. He was prominent in developing a particular mathematical language of ratios, which had earlier been used by Thomas Bradwardine to propose a 'law' relating speeds to forces and resistances, and Oresme likewise applied the language to cosmological and physical questions. He was a firm opponent of much of astrology and of magic, and to this end he employed both naturalistic and sceptical arguments. He gave many strong arguments in favour of a daily rotation of the earth, but finally concluded that it was at rest: his gambit had primarily a sceptical and fideistic purpose.

1 Life

Originating from Normandy, Oresme was educated at the University of Paris, where in 1348 he entered the College of Navarre (a royal foundation), and in 1356 became master of its theological class and hence Grand Master of the whole college. Soon afterwards he came into close contact with the royal court, and was in a more or less formal sense a tutor to the Dauphin, who in 1364 ascended the throne as Charles V. Oresme's association with the court remained through a variety of ecclesiastical appointments, culminating in that of the bishopric of Lisieux in 1377.

Oresme's writings are grounded in traditional Aristotelian scholasticism, but informed by an acute mathematical consciousness and coloured to an extent by incipient humanistic currents. The result is an elegant and urbane scepticism, which is nevertheless powered by a strong desire for scientific knowledge and, although Oresme's works cover a wide range of philosophical topics, his main importance lies in issues affecting cosmology and natural philosophy.

2 Mathematics and mathematical physics

Oresme was capable of seeing the world in strongly geometrical terms, to such an extent that he has often been compared with Descartes and his famous claim that his physics was nothing but geometry. Oresme, however, was not so strongly reductionist, and his geometrical picture is richer and more complicated. A central idea was that qualities could be represented at any point by straight lines erected perpendicularly to the subject that they informed (in the case of a linear or two-dimensional subject), proportional in length to the intensity of the quality. In the case of a solid subject, this would ideally have given rise to a four-dimensional configuration for each of its qualities, but because the space available to him was dimensionally limited Oresme conceived it as being conflated by superposition into three dimensions. Oresme then maintained that the shape of the configuration would have explanatory power, so that, for example, a quality that displayed a saw-like 'graph' would be especially pungent. Oresme was very excited by the potential explanatory capabilities of his 'imagination', and explicitly compared his doctrine to ancient atomism (see Atomism, ancient). Nevertheless, it must be admitted that the empirical basis was weak: Oresme did not specify how the differing intensities of qualities should physically be measured, and rather mutedly admitted that 'the ratio of intensities is not so properly nor so easily attained by the senses as the ratio of extensions' (*De configurationibus*).

By Oresme's time there was a growing tendency, especially through the influence of the so-called Oxford Calculators, to treat speeds by analogy with qualities (see Oxford Calculators). Within the Aristotelian tradition, motions and their speeds were essentially holistic, with the focus being on the whole motion of the whole body from its beginning to its end. Using Oresme's mathematizing approach, this gave rise to a view of the speed of a body's motion as a sort of five-dimensional object, with three spatial dimensions (those of the subject), one temporal dimension and one dimension of intensity of speed. If we focus on the motion of a point, this reduces the dimensions to two, with intensity of speed being plotted against time. This allows for comparison of different speeds: for instance, a speed whose intensity increases uniformly through time can, by equation of areas, be shown

to be equivalent to a uniform speed with the intensity that the non-uniform speed had at half-time. This is close to an essential theorem of Galileo in his *Two New Sciences*, and even the diagrams are reminiscent of each other (see Galilei, Galileo). This has given rise to much discussion of the possible influences of fourteenth-century discussions on later mechanics. The issue is still open, but probably actual influence did not extend much beyond the idea (obvious to us, but not always so) of representing instantaneous velocities or intensities of speeds by straight lines proportional in length to their magnitudes.

Problems of measure did not end with the representation of physical quantities by straight lines: these themselves had to be measured, and in the Middle Ages in this context the language of ratios (not usually regarded as being themselves numbers or fractions) was especially important. This rested on the intuition that a meaningful numerical expression of the quantity of a straight line could only be given (assuming that one did not regard the line as being made up of a certain number of points) by comparison with another line. Sometimes this was achieved tacitly by the assumption of one straight line to act as unit, but in most theoretical discussions two or more lines were explicitly brought into play, so that we could give an expression of the mutual measure of *A* to *B* by saying, for instance, that they were as 19 to 13. This was unproblematic for rational ratios, but not when it was a question of expressing the ratio of two incommensurable quantities. In Book V of the *Elements*, Euclid had given a sophisticated criterion (usually ascribed to Eudoxus) for equality between ratios in general (see Mechanics, Aristotelian), but this was often misunderstood and little used in the Middle Ages, and in any case did not give a quantitative expression of the size of the measure.

A more interesting development came from the musical tradition, and depended on a peculiarity of mathematical language. This tradition depended heavily on ratios, and allowed for their combination so that, for instance, combining the ratios of 7:5 and 5:4 gave rise to the ratio of 7:4. This was often regarded as addition, whereas we should be more inclined to think in terms of the multiplication of the corresponding fractions. Nevertheless the addition view gave rise to a perfectly coherent theory, which easily led by extension to the multiplication and division of ratios by numbers, and in turn to the possibility of numerical expressions of the sizes of at least some irrational ratios. There are some hints of this in Robert Grosseteste, but in the early fourteenth century Thomas Bradwardine developed the theory more explicitly, showing for instance how the ratio of the diagonal of a square to its side, which we should say was as $\sqrt{2} : 1$, could be numerically expressed or ‘denominated’ as ‘half of the double ratio’. Bradwardine did not provide much systematic development of the pure mathematical theory of denominating ratios, but this task was undertaken by Oresme in his *Algorismus proportionum (Algorism of Ratios)* and *De proportionibus proportionum (On Ratios of Ratios)*. He clearly showed how ratios partook of the nature of continuous quantity and could be multiplied by any fraction, but also that there were still probably irrational ratios that were not commensurable with any rational ratio, and hence were not susceptible of denomination by the new techniques.

Bradwardine made use of his language of ratios to give expression to a new law of motion, which has sometimes been regarded as radically anti-Aristotelian. In fact, given the addition view of the composition of ratios, it could (at the time) be regarded as a perfectly natural interpretation of Aristotle and his commentator, Averroes (see Ibn Rushd). The law was adopted by Oresme, who expressed it as: ‘Speed follows the ratio of the power of the mover to the thing moved, or its resistance’ (*De proportionibus proportionum*). It is thus syntactically very simple but, if translated into modern mathematical language, it demands the use of a logarithmic or exponential function, which does violence to the context of its origin. Oresme, who with his very systematic approach to ratios could focus clearly on ratios between ratios, was concerned to show that, given any two unknown ratios (rational or irrational), it was probable that they were incommensurable with each other and, if so, then given any two speeds, it was probable that they were incommensurable with each other. This result was a major feature in his attacks on many aspects of contemporary astrology.

3 Astrology and magic; naturalism and scepticism

Oresme was a prominent opponent of both astrology and magic, and some of his criticisms were clearly intended for the ears of members of the French royal court. To this end one of his anti-astrological tracts was written in French, and although like all contemporary Western scholars, Oresme’s customary literary language was Latin, he was also a pioneer in writing philosophical treatises (including some Aristotelian translations) in the vernacular. With the possible exception of his *Questio contra divinatores horoscopios (Question Against Horoscopic*

Divinators), Oresme did not deny the reality of astrological influences, but he did contest the ability of men to make predictions from them except in the most general of ways, and he warned against the moral dangers of attempting to do so. One of his most sophisticated, although possibly least effective, arguments was that from the probable incommensurability of motions noted above. If some of the heavenly motions were incommensurable with each other, then the planets could never return simultaneously to exactly the same positions, and so any precise occurrence of a Great Year, implying a cyclical universe, was impossible. Oresme concluded his most developed treatise on the subject, the *De commensurabilitate vel incommensurabilitate motuum celestium (On the Commensurability or Incommensurability of the Heavenly Motions)* with a dream in which Apollo presided over a debate between Arithmetic and Geometry as to whether there is such incommensurability, but the dream vanished before Apollo was able to pronounce a definitive judgement. Thus, although Oresme had argued for the probability of incommensurability, he emphasised that even here there is no certainty.

Oresme morally condemned magic if it were indeed employing evil demons, but was also keen to show that many effects often accounted magical were in fact produced naturalistically. Ironically, astrology had often been used to give such explanations, but Oresme necessarily employed a variety of other strategies. Among these were appeals to perceptual errors, trickery and the large effects of small causes, but especially notable was the configuration doctrine discussed above, which could in principle account for a host of occult virtues and happenings that were otherwise inexplicable. Puzzlingly, however, Oresme makes little if any appeal to this doctrine in a group of works later than the *Tractatus de configurationibus qualitatum et motuum (Treatise on Configurations of Qualities and Motions)* and assigned by manuscript evidence to 1370. This has led some scholars to contest the dating, but perhaps more plausibly we should see it as evidence of an almost playfully sceptical tendency, revealed especially in Oresme’s later work: to elaborate in considerable detail an elegant and imaginative schema, but to conclude that in the end it may be no more than imaginary. A particularly renowned example of this (together with other examples less well-known) occurs in his *Livre du ciel et du monde (Book on the Heavens and the World)* of 1377. In the course of his commmentary on the French translation of Aristotle’s *On the Heavens* which he presents there, Oresme questions the standard view of a stationary Earth and presents very strong arguments in favour of its diurnal rotation, before famously concluding that:

> Everyone holds and I believe that [the heaven] is thus moved and not the Earth, ‘for God has established the orb of the Earth, which shall not be moved’, notwithstanding the reasons to the contrary, for they are persuasions that do not conclude evidently. But considering all that has been said, one could believe that the Earth is thus moved and not the heaven, and the contrary is not evident; and yet at first glance this seems as much or more against natural reason as all or many of the articles of our faith. And so what I have said in this way for diversion can be valuable for confuting and reproving those who would impugn our faith with reasonings.
>
> (*Livre du ciel et du monde*)

In this way, an enlightened but sceptical scientific attitude buttressed religious fideism.

See also: Bradwardine, T.; Mechanics, Aristotelian; Natural philosophy, medieval

GEORGE MOLLAND

List of works

Oresme, Nicole (1351-5?) *Algorismus proportionum (Algorism of Ratios)*, ed. M. Curtze, *Der Algorismus Proportionum des Nicolaus Oresme*, Berlin: S. Calvary & Co., 1868. (Latin text of the mathematical work on the algorithm of ratios. An English translation of Part I can be found in E. Grant, ‘Part I of Nicole Oresme’s *Algorismus proportionum*’, *Isis* 56, 1965: 327-41, reprinted in E. Grant, *Studies in Medieval Science and Natural Philosophy*, London: Variorum Reprints, 1981.)

Oresme, Nicole (1351-5?) *De proportionibus proportionum (On Ratios of Ratios)*, ed. E. Grant, *De proportionibus proportionum and Ad pauca respicientes*, Madison, WI: University of Wisconsin Press, 1966. (Latin text and English translation of a treatise concerning ratios and motions.)

Oresme, Nicole (1351-5?) *Ad pauca respicientes (To Those Who Reflect On a Few Things)*, ed. E. Grant, *De proportionibus proportionum and Ad pauca respicientes*, Madison, WI: University of Wisconsin Press, 1966. (Latin text and English translation of a treatise concerning celestial motions and the question of the Great Year.)

Oresme, Nicole (1351-5?) *De commensurabilitate vel incommensurabilitate motuum celestium (On the Commensurability or Incommensurability of the Heavenly Motions)*, ed. E. Grant, *Nicole Oresme and the Kinematics of Circular Motion: Tractatus de commensurabilitate vel incommensurabilitate motuum celi*, Madison, WI: University of Wisconsin Press, 1971.(Latin text and English translation of Oresme's most mature work on the question of the commensurability or incommensurability of the heavenly motions.)

Oresme, Nicole (1351-5?) *Quaestiones super Geometriam Euclidis (Questions on Euclid's Geometry)*, ed. H.L.L. Busard, Leiden: Brill, 1961.(Latin text and English paraphrase of questions loosely related to Euclid's Elements.)

Oresme, Nicole (1351-5?) *Tractatus de configurationibus qualitatum et motuum (Treatise on Configurations of Qualities and Motions)*, ed. M. Clagett, *Nicole Oresme and the Medieval Geometry of Qualities and Motions: A Treatise on the Uniformity and Difformity of Intensities known as Tractatus de configurationibus qualitatum et motuum*, Madison, WI: University of Wisconsin Press, 1968.(Contains Latin text and English translation of Oresme's principal work on his doctrine of configurations. Clagett also suggests a provisional chronological ordering of Oresme's writings.)

Oresme, Nicole (1360?) *Livre de divinacions (Book on Divinations)*, ed. G.W. Coopland, *Nicole Oresme and the Astrologers: A Study of his Livre de Divinacions*, Liverpool: Liverpool University Press, 1952.(Contains French text and English translation of one of Oresme's anti-astrological works as well as the Latin text of another earlier work.)

Oresme, Nicole (1370?) *Questio contra divinatores horoscopios (Question Against Horoscopic Divinators)*, ed. S. Caroti, *Archives d'histoire doctrinale et littéraire du moyen age* 43, 1976: 201-310.(Edition of Latin text of an anti-astrological question, the first part of a four-part work, probably to be dated to 1370.)

Oresme, Nicole (1370?) *De causis mirabilium (On the Marvels of Nature)*, ed. B. Hansen, *Nicole Oresme and the Marvels of Nature: A Study of his De causis mirabilium with Critical Edition, Translation and Commentary*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1985.(Edition of the second part of the four-part work just noted. It concentrates on naturalistic explanations of marvellous effects: the title is not Oresme's own.)

Oresme, Nicole (1372-4) *Le Livre de Politiques d'Aristote (The Book on the Politics of Aristotle)*, ed. A.D. Menut, *Transactions of the American Philosophical Society*, new series 60, Part 6, Philadelphia: American Philosophical Society, 1970.(French translation with commentary on Aristotle's *Politics*.)

Oresme, Nicole (1377) *Le livre du ciel et du monde (Book on the Heavens and the World)*, ed. A.D. Menut and A.J. Denomy, Madison, WI: University of Wisconsin Press.(Edition and English translation of Oresme's French translation of and commentary on Aristotle's *On the Heavens*.)

References and further reading

Babbitt, S.M. (1985) *Oresme's Livre de Politiques and the France of Charles V*, *Transactions of the American Philosophical Society* 75, Part 1, Philadelphia: American Philosophical Society.(Study of Oresme's commentary on Aristotle's *Politics* in its historical context.)

Caroti, S. (1979) 'La Critica contro l'Astrologia di Nicole Oresme e la sua Influenza nel Medioevo e nel Rinascimento' (Nicole Oresme's Criticism of Astrology and its Influence in the Middle Ages and Renaissance), *Atti della Accademi Nazionale, Classe di Scienze Morali* 238: 543-685.(Thorough study of Oresme's anti-astrological writings and their influence.)

Caroti, S. (1984) '*Mirabilia*, Scetticismo, e Filosofia della Natura nei *Quodlibeta* di Nicole Oresme' (Marvels, Scepticism and Philosophy of Nature in Nicole Oresme's *Quodlibets*), *Annali dell'Istituto e Museo di Storia della Scienza di Firenze* 9: 3-20. (Oresme's attempts to naturalise apparently magical phenomena; sees signs of the influence of the *De configuationibus* in the works dated 1370.)

Caroti, S. (1993) 'Oresme on Motion (*Questiones super Physicam* III, 1-7)', *Vivarium* 31: 8-36.(Oresme's conception of the ontological status of motion as being that of a mode or condition of the thing moved.)

Durand, D.B. (1941) 'Nicole Oresme and the Mediaeval Origins of Modern Science', *Speculum* 16: 167-85. (Classic early discussion of Oresme's place in the development of science.)

Grant, E. (1962) 'Late Medieval Thought, Copernicus, and the Scientific Revolution', *Journal of the History of Ideas* 23: 197-220.(A still very useful study, which contrasts Oresme and Copernicus on the motion of the Earth.)

Grant, E. (1981) *Studies in Medieval Science and Natural Philosophy*, London: Variorum Reprints.(Several of the chapters relate to Oresme.)

Grant, E. (1993) 'Jean Buridan and Nicole Oresme on Natural Knowledge', *Vivarium* 31: 84-105.(The differing positions of Buridan and Oresme on the status of natural philosophy *vis-à-vis* theology.)

Menut, A.D. (1966) 'A Provisional Bibliography of Oresme's Writings', *Mediaeval Studies* 28: 279-99.(A

supplementary note can be found in *Mediaeval Studies* 31, 1969: 346-7.)
Molland, A.G. (1974) 'Nicole Oresme and Scientific Progress', *Miscellanea Mediaevalia* 9: 206-20.(Oresme's ambivalent attitude to the possibility of making progress in natural philosophy.)
Quillet, J. (ed.) (1990) *Autour de Nicole Oresme (Concerning Nicole Oresme)*, Paris: Vrin.(Useful collection of essays, including P. Souffrin on instantaneous velocity, S. Caroti on the philosophy of nature in the 1370 collection, M. Lejbowicz on the anti-astrological writings and J. Quillet on happiness.)
Souffrin, P. and Segonds, A.P. (eds) (1988) *Nicolas Oresme: Tradition et Innovation chez un Intellectuel du XIV Siècle (Nicole Oresme: Tradition and Innovation in a Fourteenth-Century Intellectual)*, Padua: Programma e 1+1, and Paris: Les Belles Lettres.(Collection of essays, including E. Grant on knowledge and truth, H. Hugonnard-Roche on argumentation, J. Meusnier on probabilism, G. Molland on holism and mathematical science, J.D. North on physics and cosmology and A.P. Youschkevitch on mathematics.)
Sylla, E.D. (1984) 'Compounding Ratios: Bradwardine, Oresme and the First Edition of Newton's *Principia*', in E. Mendelsohn (ed.) *Transformation and Tradition in the Sciences*, Cambridge: Cambridge University Press, 11-43.(Oresme's work on ratios as part of a continuing tradition.)

Orientalism and Islamic philosophy

Orientalism is the concept that there is something very special and different about the thought of those living in the East, which can be discovered through the methods of scholarship current in the West. It is a reflection of the relationship of imperial and intellectual domination of a West which feels it is superior to an 'inferior' East. This often results in an understanding of Islamic philosophy which sees the latter as essentially unoriginal, derivative and of only historical interest. While orientalists have produced interesting and important work, most fail to appreciate the independent status of the material which they analyse.

1 The notion of Orientalism

Orientalism is the branch of scholarship that uses traditional Western methods as a means of understanding and gathering knowledge pertaining to the Orient. The term was also used by Edward Said (1978) to elucidate his own challenge to the validity of such methods.

On the one hand, Orientalism has given us much of what we know about the Oriental world at large. Late nineteenth-century authors are especially worthy of consideration for their contributions to an understanding of foreign cultures and peoples. On the other hand, however, several problems arise from the attitudes and methods used in traditional Orientalist discourse, which in turn has had an impact - often negative - upon Western consciousness. This influences and distorts the framework through which the West approaches the Orient in general and Islam in particular.

The Orient encompasses a far greater area than simply that of the Arabs and the Muslim community; exotic images from India, China, Japan and Korea are conjured up in the minds of Western people when they think of the Orient. However, Orientalism has had a particular impact on the study and understanding of Islamic philosophy. Many scholars' understanding of Islamic philosophy is, 'that Islamic civilization as we know it would simply not have existed without the Greek heritage' (Rosenthal 1975: 14).

2 Orientalism as a political doctrine

Orientalism has several different but interrelated meanings. In its general sense, it describes the way in which the West looks at the Orient in order to understand it within the context of Western experience. More specifically, Orientalism is a categorical approach by Western scholars as an attempt to form a collective body of knowledge about the Orient. Included in this enterprise is the study of Eastern philosophies, history, religion, culture, language and social structures. To understand the effect of Orientalism on Islamic philosophy, however, we need first to understand it as a political doctrine; from its inception, Orientalism was primarily political and secondarily cultural and philosophical.

Orientalism as a system of the knowledge began in the late seventeenth century, reaching its zenith in the late nineteenth and early twentieth centuries. During this period Britain and France, and later the USA, were involved in the struggle for Western domination. This Anglo-French-American experience was essentially imperialistic. Implicit in the Orientalist attitude, therefore, is the belief that the Orient had passed its golden age as the West was being 'born', and was thus in decline. This view of the Orient as backward and barbaric led to interpretations which resembled more closely what scholars wished to believe, rather than what actually was the case. In philosophy, this attitude led to the belief that the entire system of Islamic philosophy was based upon the Greek inheritance; this in turn led to the belief that Islamic philosophers were not good Muslims, as philosophy and religion apparently could not be reconciled.

Along with this attitude, Orientalism also played an active role in advancing Western interests in the East. The pursuit of knowledge of the Orient was often not an end in itself. The study of Islamic philosophy merely confirmed many Western scholars in their belief of the superiority of their own culture.

3 The Orientalist approach to Islam

From the beginning, Orientalists have viewed Islam in two ways. First, as it had borrowed liberally from Abrahamic (Judaeo-Christian) traditions, Islam was considered to be a crude parody of Christianity. Second, Islam

was looked upon as an alien menace which historically had enormous military and political success throughout the world, and consequently was a threat to Western civilization.

In Orientalism, Islam first had to be placed within the realm of Western understanding with respect to Christian concepts rather than regarded on its own terms. One way of accomplishing this was to make analogies between Christian religions and Islam. The obvious parallel is the one which some Orientalists draw between Muhammad and Christ. Since Christ is central to the Christian faith, Westerners assumed that Muhammad holds the same place in Islam. This misconception helped to popularize the use of the name ‘Mohammedanism’, a term highly offensive to Muslims. The Christ analogy also served to reinforce the notion that Muhammad was nothing more than an ‘impostor’ and a pale version of the Christian Messiah.

Islam also provided a provocation to the West in another respect. From the time of the Arab conquests in the seventh century to the ascendancy of the Ottoman Empire, Islam itself posed a formidable challenge to the Christian world. Islamic empires - Arab, Ottoman or those in Spain and North Africa - had quite effectively challenged and, at least for short periods of time, dominated Christian Europe.

4 Orientalism and Islamic philosophy

The first wave of Muslim conquest in AD 632- 4 secured for the Arab Muslims the strongly Hellenized territories of Syria and Egypt along with the western part of the Sassanian Persian empire. At first the new conquerors may have been suspicious about the culture of classical antiquity, as both religion and language separated the Arab Muslims from the vanquished peoples. However, the former overcame their anxiety remarkably quickly and began instead a cultural conquest and assimilation of ancient knowledge.

Philosophers such as Ibn Sina, al-Kindi, al-Farabi and Ibn Rushd all interpreted the Islamic inheritance of classical philosophy and attempted an assimilation of it into mainstream Islam in their writings (see Aristotelianism in Islamic philosophy; Neoplatonism in Islamic philosophy; Platonism in Islamic philosophy). The classification of the sciences, in encyclopedic proportions, was carried out by the likes of al-Khwarizmi in his *Mafatih al-‘ulum*, and by a group of scholars in the ninth and tenth centuries who called themselves the ‘Brethren of Purity’ (Ikhwan al-Safa’). The attitude of these and other philosophers was one of acceptance and the transferring of intellectual history, rather than an adoption of so-called ‘foreign’ ideas. Many of these ideas were foreign indeed, as we do not see any mention of them in the Qur’an and earlier traditional works, but the underlying theme was the gaining and spread of knowledge, *‘ilm*. Gaining of knowledge is obligatory according to the tradition of Muhammad, the prophet of Islam, who said: ‘Seeking knowledge (*‘ilm*) is obligatory upon all Muslims, men and women.’

This approach is clearly brought out in the epistles of the Ikhwan al-Safa’ (see Ikhwan al-Safa’). During a lengthy debate in these epistles, between humans and animals in the court of the king of the *jinn*, examples from history are given in which the conquering nations took over, translated into their language and made their own the knowledge of vanquished peoples. Solomon is cited as a classic example:

> Our sciences and the sciences of all the nations are [acquired] one from the other. If this was not the case, from where did the Persians get astrology, astronomy and observatories? Did not they take it from the Indians? If it was not for Solomon, where did the Israelites get the sciences…. He took it from the kings of all nations when he conquered their territories and transformed them to Hebrew…
>
> (Rasa’il Ikhwan al-Safa’ 2: 288)

This transcendental open-mindedness has been misrepresented many times by Western scholars, who are more interested in finding something new in the Islamic sciences than in attempting to understand the transmission of the corpus of human knowledge from one people to another. The Orientalist outlook mentioned above appears clearly when scholars such as Walzer and Rosenthal expend much effort in finding faults within the Islamic philosophical system, rather than using their impressive abilities to develop a better understanding of the amalgamation and legacy that has been left by the Islamic philosophers. In Walzer’s important work, *Greek into Arabic* (Walzer 1962), the attitude is that everything the Arab philosophers had to say was ‘borrowed’ from the Greeks. Even when we are not able to find the source, according to Walzer, it can be assumed safely that the original Greek source is no longer extant.

This imaginary dichotomy between philosophy and theology, as assumed by the Orientalists, has led to a severe

crippling of the understanding of the achievements of the Islamic philosophers. This attitude itself has its roots in the initial reaction to the impact of Greek philosophy upon Christianity. Orientalist scholars assume that the Muslims felt exactly the same as the early Christians did about the conflict between philosophy and theology. Of course there were debates between Islamic philosophers and theologians, but in the Arab and Islamic milieu the attitude was rather different from that in the Christian world; in the former, the philosophers were simply taking what they assumed was their legitimate inheritance from the corpus of human knowledge.

In conclusion, these notions about the Orient eventually ‘created’ the Orient. The ‘Orient’ (East) is, in fact, only East from an European perspective: it is a relative, not absolute, term. More precisely, the ‘Orient’ is whatever the Orientalists say it is: it is a series of abstractions based upon Western-generated ideas rather than upon Oriental realities. Islamic philosophy needs to be studied as more than just a reflection of Greek ideas; it needs to be considered as an area of thought which came into contact with a variety of different cultures, and which developed out of these and out of a meditation on the Islamic sciences themselves into an independent and original form of philosophical thought.

See also: Greek philosophy: impact on Islamic philosophy; Illuminationist philosophy; Islam, concept of philosophy in; Islamic philosophy, modern; Islamic theology; Mystical philosophy in Islam

UBAI NOORUDDIN

References and further reading

Corbin, H. (1993) *History of Islamic Philosophy*, trans. L. Sherrard, London: Kegan Paul International.(One of the very few approaches to Islamic philosophy which is not especially orientalist.)

Ikhwan al-Safa’ (1956) *Rasa’il Ikhwan al-Safa’ (Epistles of the Brethren of Purity)*, Cairo.(The letters of the Brethren of Purity, which provide evidence of their ‘humanistic’ views.)

Jalal al-Azm, S. (1982) ‘Orientalism and Orientalism in Reverse’, *Khamsin* 8: 5-27. (Interesting description of how some Islamic writers go in for Orientalism in reverse, arguing that only Muslims can understand Islamic culture.)

Leaman, O. (1996) ‘Orientalism and Islamic Philosophy’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, 1143-8.(A critique of persistent Orientalist attitudes to the study of Islamic philosophy.)

Rosenthal, F. (1975) *The Classical Heritage in Islam*, London: Routledge & Kegan Paul.(Useful account of the links between the classical world and Islam.)

Said, E. (1978) *Orientalism*, London: Routledge & Kegan Paul.(A key work on the subject of Orientalism by the writer who really started the controversy.)

Walzer, R. (1962) *Greek into Arabic*, Oxford: Cassirer.(Excellent on the links between Greek culture and Islamic philosophy.)

Origen (*c.*185-*c.*254)

An ascetic Christian, prodigious scholar and dedicated teacher, Origen devoted his life to exploring God's revelation. Much of his work takes the form of commentaries on Scripture. He argued that Scripture has three levels: the literal, the moral and the spiritual. The literal level veils the others, and we need God's help to find the divine mysteries behind the veil. His commentaries directly or indirectly influenced the practice of exegesis throughout the patristic period and the Middle Ages.

Origen used his spiritual exegesis, as well as arguments, concepts and models drawn from philosophy, to tackle the theological problems of his day: the compatibility of providence and freedom, the relation of the Father, Son and Holy Spirit to each other and to rational creatures, the problem of evil, and the origin and destiny of the soul. He is famous - or infamous - for arguing that the souls of angels, demons and human beings enjoyed a previous heavenly existence, but that they sinned and fell. God created the world to punish and remedy their faults.

1 Life and works

Origen was born to Christian parents in Alexandria. When he was seventeen, his father was martyred in the persecutions of Septimius Severus. Origen, the oldest child, supported his family by teaching grammar. Shortly afterwards, Bishop Demetrius asked Origen to take charge of the catechetical school in Alexandria. In that time of persecution, such a charge meant preparing catechumens not only for baptism, but for martyrdom as well. Accordingly, in addition to teaching, Origen tended to the martyrs at their trials and in prison, and even accompanied them to their executions. He himself lived an ascetic life, sleeping and eating very little, devoting himself to teaching, study and pastoral work with an uncompromising intensity.

He studied philosophy in Alexandria, perhaps with Ammonius Saccas, who taught Plotinus. He read widely in Greek philosophy, but also studied Jewish writers, in particular Philo of Alexandria, and was conversant with rabbinical traditions. In Alexandria, he founded his own school of advanced study. However, at some time between AD 230 and 233, Origen fell from Demetrius' good graces and was banished from the church of Alexandria. Origen then moved to Palestine, where he enjoyed strong support from the bishops, and founded another school of advanced study. When the persecution of Christians resumed under the Emperor Decius, Origen was arrested, imprisoned and tortured. When the persecutions ended, he was released. He died shortly thereafter.

Origen wrote a prodigious number of works, some scholarly, some polemical, but all with the pastoral aim of helping Christians to avoid error and deepen their knowledge of moral and spiritual truth. In *Kata Kelsou (Against Celsus; known in its Latin translation as Contra Celsum)*, he defends Christianity against charges of irrationality levelled by the pagan philosopher Celsus. Origen's *Peri archōn (On First Principles)* is the first work of Christian systematic theology; however, acutely aware of his own fallibility, he proposes the elements of his system very tentatively. Origen also preached and wrote commentaries on nearly every book of Scripture, although most of these works have not survived and some exist only in Latin translations. It is impossible to overestimate the importance of these commentaries, whose methodology and content had a profound influence on Basil of Caesarea, Gregory Nazianzen and Gregory of Nyssa in the East, and on Jerome and Ambrose in the West (see Patristic philosophy). Through these writers, Origen influenced the practice of exegesis throughout the Middle Ages.

2 Revelation

In Origen's view, Christians can come to understand moral and spiritual truths through study of Scripture. The Holy Spirit revealed these truths to the apostles and prophets, who recorded them in their writings for the benefit of the faithful. However, to express these truths plainly would risk allowing the enemies of the Church to ridicule them. Consequently, the authors veiled these truths in narratives and legal ordinances which constitute the literal level of Scripture. To find the spiritual level, Origen uses typological and allegorical exegesis. Typological exegesis, practised before Origen by Justin and Irenaeus as well as by the authors of the New Testament, interprets passages in the Old Testament as foreshadowing something in the gospel, whether historical or spiritual. Allegorical interpretation, practised before Origen by Philo of Alexandria, understands every detail of Scripture to have a moral or spiritual significance.

Finding the spiritual meaning of Scripture requires more than philological skill. As Origen sees it, one must first lead the spiritual life, a life of devotion to Christ in prayerful and diligent study of the Scriptures. As one becomes more like Christ and participates in him, one becomes worthy to be taught the mysteries of the spiritual gospel hidden by the literal gospel. Christ himself must lift the veil from the letter. Origen supposes that Jews, Gnostics and literalists misread the Scriptures because they do not listen to Christ's explanation, available inwardly to those who lead the spiritual life. In his commentaries and homilies, Origen often prays for Christ's help, or asks his listeners to pray for him, so that Christ might lift yet more of the veil from the letter. This process will continue in the next life, where Christ will teach the blessed, face to face, more and more of the spiritual mysteries (see Revelation).

3 Theology

The prophets and apostles taught in plain terms only a small part of the spiritual gospel, including the most important truths: that God is one and the creator of the world; that his Son took a human body without resigning his divinity, and that he suffered, died and rose from the dead; that the Holy Spirit inspired the saints of both Testaments; that the rational soul is free and will be rewarded or punished according to its deserts. The apostles left the explanation and clarification of these truths to others who have received the gift of wisdom from the Holy Spirit. Origen's theology consists largely of what he takes to be this explanation and clarification. Acutely aware of the immensity of this task and of his own limitations, Origen often presents his views tentatively. With equal intensity and humility, he combs the Scriptures to investigate the relationship among the persons of the Trinity, God's motives in creation; the incarnation and resurrection, and the origin and destiny of human beings, angels and demons. For these investigations, Origen borrows concepts, models and arguments from various philosophical schools, including the Stoic, Peripatetic and, in particular, the Platonic schools (see Neoplatonism; Platonism, Early and Middle). Though he finds philosophical tools and speculations useful in the service of Christianity, he warns his students to give their allegiance to God and not to any philosophical school. Too often, giving allegiance to one school means closing one's mind to other ways of thinking.

In Origen's theology, God is a Trinity of three beings or hypostases: the Father, the Son and the Holy Spirit, coeternal, incorporeal and uncreated. Each of these three beings has its own attributes and role to play in the created world and in the life of the Trinity itself. The Father is absolutely simple, immutable and impassible. He receives nothing from any source, but is himself the source of all being and all goodness. The Son, begotten by the Father, bears the titles Wisdom, Word, Truth and Life, which characterize him as he is in himself, and about one hundred further titles which characterize his relationship to the created world, such as Redeemer, Good Shepherd and High Priest. The Holy Spirit sanctifies rational creatures, conferring spiritual gifts on them. The Spirit receives being and goodness from the Father, and wisdom and intelligence from the Son.

Origen's admirers and detractors alike express worries about his subordination of the Son and Holy Spirit to the Father. Only the Father is God in his own right; the Son and Holy Spirit merely participate in his Godhead, as they participate in his being and goodness. However, Origen never asserts that the Son and Holy Spirit are God or good to a lesser degree than the Father, only that they receive their Godhead and goodness from him.

The Father creates the world and everything in it *ex nihilo*, not from pre-existent matter. Origen sometimes proposes that the Father creates not just one world but a succession of worlds, though that series itself has a beginning in time. His denial of the eternity of this series makes him vulnerable to the criticism that God was idle before the creation. However, to think God idle is both inconceivable and impious. God so overflows with goodness and creative power that it is impossible for him not to exercise this power and to manifest his goodness. Recalling the verse from Psalms 104 (103), 'In Wisdom hast thou made them all', Origen speculates that in addition to the temporal creation, there is also an eternal creation: The Father eternally creates in the Son, whose chief title is Wisdom, the plans and patterns of the temporal world as an expression of his power and goodness. In the Son, then, exist the genera and species of all things, and perhaps the patterns for individuals as well. The Son, therefore, provides the link between the One and the many (see Trinity).

4 Rational creatures

Origen develops much of his thought about rational creatures by reflection on the problem of evil and the goodness of God (see Evil, problem of). He was impressed by the great inequalities of life: inequalities of power, privilege

and opportunity, of flourishing and suffering. Why are some rational creatures angels, some demons and others human beings? Why are some human beings born into power and privilege and others not? Why are some Christians martyred for their beliefs? The Gnostics claim, Origen writes, that some souls are by nature good, others evil, and still others indifferent, and one's lot in life depends on the nature of one's soul. Otherwise one's lot would be a matter of blind luck (see Gnosticism). Origen rejects this Gnostic view. God's providence governs every individual's life, so the inequalities of life cannot be due to luck. Moreover, because he is good and just, God would not introduce inequality arbitrarily and, therefore, unjustly.

Origen proposes that rational creatures enjoyed a pre-existent life, united to God in heavenly contemplation. God created them entirely equal and sinless. However, because they were rational, their actions were up to them; and because God created them out of nothing, they were mutable. Although Christ's human soul retained its innocence, other souls - perhaps all others - fell from their original union with God. Origen sometimes suggests that they grew satiated and lost interest in their contemplation of God. God responded to their fall by creating the world and assigning the fallen creatures a place in it according to their deserts. Those who distanced themselves from God the least became angels; those who distanced themselves the most became demons; and those whose fall was neither great nor small became human beings.

Human life, then, is God's punishment for sins committed in the pre-existence. The particular circumstances in which human beings find themselves - their advantages and disadvantages - are assigned to them according to their deserts. However, Origen finds God's punishment remedial as well as retributive. God is a doctor who sends us bitter medicine only to cure us. Origen also likens God to a loving tutor whose discipline aims at teaching the pupil a lesson. He puts us to the test because he wants us to face the secrets of our hearts: it is only through temptation that we see what we are really like. If we do not learn our lessons in this lifetime, then God does not abandon us, but continues the plan for our salvation into our next life (see Salvation).

From Paul's teaching that, at the end, God will be all in all, Origen formulates his account of the *apocatastasis*, or restoration, when rational creatures regain the purity of their original state and God fills their minds. In *Peri archōn*, Origen suggests that the *apocatastasis* will be universal, so that all rational creatures, even the Devil, will eventually turn to God. However, he has no settled view on this subject. In a letter to friends in Alexandria, he explicitly denies that the Devil will be saved, and his homilies on Jeremiah suggest that some human beings too will suffer eternal punishment.

See also: Patristic philosophy; Revelation; Trinity

JEFFREY HAUSE

List of works

Origen (*c.*185-*c.*254) Works, in *Origenes werke*, Die Griechischen Christlichen Schriftsteller des ersten drei Jahrhunderte, Berlin: Akademie Verlag, 1899-1913.(Editions of the works of Origen, including all the individual works cited below.)

Origen (218-38) *Ōrigenous tōn eis to kata Iōannēn (Commentary on the Gospel according to John)*, ed. E. Preuschen in *Origenes werke*, Leipzig: Hinrichs, 1903; trans. R.E. Heine, Washington, DC: Catholic University of America Press, 1989.(Meticulous effort to uncover the spiritual meaning of the most profound and complex gospel. An important source for Origen's theology of the Trinity and exegetical methodology.)

Origen (220-30) *Peri archōn (On First Principles; in Latin, De principiis)*, ed. P. Koetschau in *Origenes werke*, Leipzig: Hinrichs, 1913; trans. G.W. Butterworth, *On First Principles*, New York: Harper & Row, 1966.(A systematic treatment of God, rational creatures and their return to God. The original Greek manuscript survives only in fragments; the Latin translation, *De principiis*, is complete.)

Origen (233) *Peri euchēs (On Prayer)*, ed. P. Koetschau in *Origenes werke*, Leipzig: Hinrichs, 1899; trans. J.E.L. Oulton in *Alexandrian Christianity*, London: SCM Press, 1954.(A discussion of prayer and its role in Christian life; contains an interpretation of the Lord's Prayer.)

Origen (235) *Eis marturion protrepticos (Exhortation to Martyrdom)*, ed. P. Koetschau in *Origenes werke*, Leipzig: Hinrichs, 1899; trans. H. Chadwick in *Alexandrian Christianity*, London: SCM Press, 1954.(Intended to encourage, strengthen and instruct Christians awaiting martyrdom.)

Origen (241-44) Homilies on Jeremiah, ed. E. Klosterman, *Jeremiahomilien*, in *Origenes werke*, Leipzig: Hinrichs, 1901. (Homilies exposing Jeremiah, containing attacks against adversaries of Christianity and discussions of resurrection and punishment in the afterlife.)

Origen (248) *Kata Kelsou (Against Celsus; in Latin, Contra Celsus)*, ed. P. Koetschau in *Origenes werke*, Leipzig: Hinrichs, 1899; trans. H. Chadwick, Cambridge: Cambridge University Press, 1965.(A defence of Christianity against pagan attacks, together with counterattacks against paganism.)

References and further reading

Clark, E.A. (1992) *The Origenist Controversy: The Cultural Construction of an Early Christian Debate*, Princeton: Princeton University Press.(Thorough discussion of the controversies over Origenist thought in the late fourth and early fifth centuries.)

Crouzel, H. (1989) *Origen*, trans. A.S. Worrall, San Francisco: Harper & Row.(An exploration of select aspects of Origen's thought.)

Daniélou, J. (1955) *Origen*, trans. W. Mitchell, London: Sheed & Ward.(Useful and readable introduction to Origen's thought.)

Harl, M. (1958) *Origène et la fonction révélatrice du Verbe Incarné (Origen and the Revelatory Role of the Incarnate Word)*, Paris: Patristica Sorbonensia, Éditions du Seuil.(Thorough discussion of the incarnate Christ's role as revealer of God.)

Orphism

Orphism, a speculative trend within Greek religion, claimed the mythical singer Orpheus as founder and prophet. In changing forms, it is in evidence from the 6th century BC *to the end of antiquity. Hexameter poems were attributed to Orpheus, especially a theogony about the origin of gods, world and mankind, and sectarian groups led an 'Orphic life', practising mystery cults supposedly founded by Orpheus. The main goal was salvation of the soul from evil, traced to 'ancient guilt', with a view to a blessed existence after death; this usually included the doctrine of transmigration*

Myths about Orpheus, son of the Muse Calliope and a Thracian called Oiagros, are in evidence from the sixth century BC and current in the fifth. Orpheus, a lyre-player and singer, is said to have taken part in the expedition of the Argonauts. He enchants even animals with his music. He descends to the netherworld to get back his wife Eurydice, an attempt that fails. Although torn to pieces by Thracian bacchants, his severed head continues to sing and to give oracles.

A theogony is attributed to Orpheus, the key document of which is a fragmentary papyrus book, unearthed at Derveni in 1962, containing quotations and allegorical commentary on Orpheus' text (see Most and Laks 1977). The theogony begins with 'Night', from which a series of 'kings' of gods springs: Uranos (Heaven), Cronos and Zeus. Zeus, after a strange incident of 'swallowing' a 'phallus (?)', brings forth the other gods from himself and proceeds methodically to establish the world. Then, by incest with his 'Mother', he begets a further offspring, Persephone, leading towards the creation of mankind.

The Derveni text ends here. Plato alludes to Orpheus' theogony, which comes to a stop 'in the sixth generation' (Dionysus as the father of man?). He attests that, according to the teachings of Orphics (*Orphikoi*), the soul is incarcerated in the body during this life to be punished for certain unspeakable crimes. The corresponding myth is traceable through much-debated allusions in older texts, but becomes explicit in later sources: mankind rose from the soot of the Titans, burnt by the lightning of Zeus after they had treacherously killed and eaten the child Dionysus (also called Zagreus), son of Persephone and Zeus; hence the sufferings, the warning against falling back into the 'old Titanic nature', and the yearning to get rid of the 'unlawful forefathers', 'to reach the end of the cycle, and to take breath from evil'. The means towards this are ritual purifications and the adoption of a 'pure' life, the 'Orphic life', characterized above all by vegetarianism. There are special myths about the judgment of the dead and punishments or blissful existence in the netherworld, and a belief in transmigration, with repeated chances of personal guilt and atonement in successive lives. Transmigration, together with certain taboos, constitutes an overlap with Pythagoreanism (§3). This makes it impossible to decide exactly what role Orphic literature and teaching played in the emergence of the concept of a personal 'soul', or *psychē* (see Psychē), independent of the body and antithetical to it, before Plato's philosophical restatement of this theory (see Plato §§10-13).

Other evidence is the gold lamellae found in tombs of southern Italy, Thessaly, Lesbos and Crete (see Pugliese Carratelli 1993 for a partial translation of the evidence). Dated from the late fifth to the second century BC, these bear instructions for the dead on how to pursue their way towards bliss in the beyond. The texts differ in length and in specific details, but still form an interlocking whole. As they refer to 'bacchants and initiates' (Hipponion, southern Italy) and to the 'release' wrought by Bacchios, that is, Dionysus (Pelinna, Thessaly), they clearly presuppose rites of Dionysiac or 'Bacchic' mysteries, spread by travelling initiation priests. The initiate may proclaim his identity as 'son of Earth and starry Heaven'; there is the prospect of Elysian life, after drinking from the 'lake of memory'; there also occurs the declaration 'you have become god instead of a human'; in addition, there are suggestive but enigmatic formulae such as 'kid I fell into the milk'. The reference to transmigration is opaque, and that to Orpheus is not explicit, but Orphism remains the most probable context.

Orphics, in connection with Dionysus and with dualistic speculation about 'soul-body' and 'life-death-life', also are attested at Olbia in the fifth century BC. Instructions for practical mysteries are found in an Egyptian papyrus text from the third century BC (see Kern 1963: fr. 31).

In Hellenistic times, the theogony of Orpheus went through various conflicting versions and expansions. Oriental influence can be discerned in the striking predilection for grotesque mixtures of images. 'Unageing Time'

(*Chronos ageraos*), for example, appears in the form of a three-headed snake, with the heads of a bull, a god and a lion. The main text had 'Time' (Chronos) for the primeval god, hatching an 'Egg' from which the world was fashioned. The Egg then broke in half to release a mysterious god of light, Phanes, also called Erikepaios (an unexplained) and identified with Dionysus. At a later stage, Zeus swallowed Phanes and everything else in order to recreate the whole of the world from himself; here one can trace Stoic concepts and terms about the 'sympathy' of the universe (see Posidonius §5). This theogony became a reference text for Neoplatonic philosophers until the end of paganism, from Porphyry to Proclus and Damascius. A revelation of polytheism more serious than those of Homer and Hesiod, it was adapted to Platonic ontology by allegorical interpretation. There is also a collection of 'Hymns of Orpheus', stemming probably from a sectarian community in Asia Minor, which testifies to the ongoing practice of Orphic cults in the imperial age (see Quandt 1952).

By contrast, long before, in the third century BC, a 'Testament of Orpheus' had been fabricated. Reflecting contacts with Judaism, it sets out to counter the Orphic theogony. In it, Orpheus revokes polytheism and proclaims the one god, creator of the universe. This was used by both Jews and Christians to rebuke paganism (see Riedweg 1993 for a detailed discussion).

WALTER BURKERT

References and further reading

Borgeaud, Ph. (ed.) (1991) *Orphisme et Orphée*, Geneva: Droz.(Takes account of recent discoveries.)

Brisson, L. (1995) *Orphée et l'Orphisme dans l'Antiquité Gréco-Romaine*, Aldershot: Variorum.(Scholarly articles on special problems.)

'Der Orphische Papyrus von Derveni' (Derveni Papyrus), *Zeitschrift für Papyrologie und Epigraphik* 47 (1982), appendix 1-12.(Preliminary edition of the Derveni Papyrus.)

Kern, O. (ed.) *Orphicorum Fragmenta*, Berlin: Weidmann, 2nd edn, 1982.(Collection of fragments, outdated but not yet replaced; fragment 31 is the Egyptian papyrus text.)

Most, G.W. and Laks, A. (eds) (1997) *Studies on the Derveni Papyrus*, Oxford: Clarendon Press.(Additions to the text, English translation, interpretative studies.)

Nilsson, M.P. (1952) 'Early Orphism and Kindred Religious Movements', in Nilsson, *Opuscula Selecta*, Lund: Gleerup, vol. 2, 628-83.(A solid survey of earlier work.)

Pugliese Carratelli, G. (ed.) (1993) *Le lamine d'oro 'Orfiche'*, Milan: Scheiwiller.(Texts of the gold plates; still incomplete.)

Quandt, G. (1952) *Orphei Hymni (Hymns of Orpheus)*, Berlin: Weidmann.(Standard edition of the Greek text.)

Riedweg, C. (ed.) (1993) *Jüdisch-hellenistische Imitation eines orphischen Hieros Logos (Jewish-Hellenistic Imitation of a 'Sacred Discourse' of Orpheus)*, Tübingen: Narr.(Contains the Testament of Orpheus.)

West, M.L. (1983) *The Orphic Poems*, Oxford: Clarendon Press.(Concentrates mainly on the Derveni theogony and further Orphic theogonies, including the evidence from Olbia.)

Ortega y Gasset, José (1883-1955)

The Spanish philosopher Ortega borrowed themes from early twentieth-century German philosophy and applied them with new breadth and urgency to his own context. Calling his philosophy 'vital reason' or 'ratiovitalism', he employed it initially to deal with the problem of Spanish decadence and later with European cultural issues, such as abstract art and the mass revolt against moral and intellectual excellence. Vital reason is more a method for coping with concrete historical problems than a system of universal principles. But the more disciplined the method became, the deeper Ortega delved into Western history to solve the theoretical and practical dilemmas facing the twentieth century.

1 Life

Typifying liberal Spanish philosophers of the late nineteenth and early twentieth centuries, José Ortega y Gasset philosophized in order to reverse Spain's historical decline. Born into two publishing families, he achieved eminence as a philosopher-journalist, an educator, and in 1931 as a politician. Postdoctoral studies in Germany acquainted him with Neo-Kantian philosophers of culture Georg Simmel and Hermann Cohen (see Neo-Kantianism). From Cohen's follower Paul Natorp, Ortega learned of Husserl's phenomenology, which he introduced into Spain in 1913 in newspaper articles, three years after obtaining the Chair in Metaphysics at the University of Madrid (see Husserl, E.; Phenomenological movement). His first book, *Meditaciones del Quijote (Meditations on Quixote)* (1914), employs phenomenological descriptions in essays on the salvation of Spanish culture by means of Cervantes' insights. Among journals Ortega founded, the most distinguished, *Revista de Occidente (Review of the West),* begun in 1924, informed Spaniards of innovations in philosophy, anthropology, physics and art. Ortega phenomenologically analysed avant-garde art in his renowned *La deshumanización del arte (The Dehumanization of Art)* (1925). He sensed significant progress being made in saving Spanish civilization through rigorous concepts.

Yet the following eleven years offered little cause for optimism. In 1929, intervention in university affairs by the dictator Miguel Primo de Rivera prompted Ortega to resign his Chair. Renting theatres, he offered a Bergson-style public course, *Čiquest;Qué es filosofía? (What Is Philosophy?)* (1929-30), in which he unveiled his mature principles of human existence. In 1930 Ortega's best-known book appeared, *La rebelión de las masas (The Revolt of the Masses)*, launching a philosophical attack on an over-egalitarian age for aggrandizing mediocrities such as Primo and Mussolini. During the Second Spanish Republic (1931-6), Ortega founded a party of intellectuals. While an elected Member of the Spanish Parliament, he deplored extremism of right and left, only to abandon politics himself in 1932 and leave the country in 1936, before civil war erupted.

In 1945 he returned to Franco's Spain, but the hostility of Church and State drove him to lecture abroad whenever possible in his final years. In his richest, most rigorous work, the posthumous *La idea de principio en Leibniz y la evolución de la teoría deductiva (The Idea of Principle in Leibniz and the Evolution of Deductive Theory)* (1946-50), he rebuts hostile clerics. To Spain and Spanish America he bequeathed an elegant and precise philosophical vocabulary, along with a number of outstanding students, notably Xavier Zubiri and Juan David García Bacca.

2 Relations between reason and life

The problem of the relationship between reason and life lies at the root of Ortega's thinking. Both terms vary in meaning as his method of 'vital reason' matures, the variations stemming from his dialogues with contemporaries. He corresponded with the older philosopher Miguel de Unamuno, who, with Kierkegaard and Liberal Protestant theologians, conceived authentic existence as the struggle between reason and life: life affirms immortality after death; reason denies it. But Ortega dodged religious issues to concentrate on the philosophy of culture. In Georg Simmel he found an opposition posited between life as process and life as cultural principle, with authentic culture a synthesis of both. Hence in *El tema de nuestro tiempo (The Modern Theme)* (1923), Ortega wrote, 'Life should be cultured, but culture must be vital'. Simmel's notion of life as process, a continual becoming, drew Ortega close to Nietzsche, Bergson and the Darwinists (see Vitalism). Yet Ortega found a more fruitful idea in Cohen, who held that, to be oneself, it is necessary to use logic, such as the induction and deduction employed in mathematical physics. Ortega embraced the conception of life as a search for identity, but qualified Cohen's emphasis on

scientific logic: for Ortega, the problem of personal identity stands in the foreground, and the solution of physico-mathematical logic only at midground. Therefore, from 1910, Ortega's point of departure became the idea of life as a problem of self-authentication, but without Cohen's faith in the constructs of scientific reason as a solution to it.

Ironically, it was Cohen's fervent follower Natorp who drew Ortega's attention to other forms of reasoning, by alerting him to the strengths and weaknesses of Husserlian phenomenology. Before 1907, Husserl had stressed the cognitive value of the direct contemplation of essences called 'intuition' (*Anschauung*): I can 'intuit' an apple, the colour red, a circle, a mythical fruit. From 1907 on, however, Husserl stressed that 'pure' intuition requires self-conscious reflection, and Natorp argued that reflection is a mental construct, mediating between mind and object and thereby preventing the pure intuition of essences. Natorp's objection would recur from 1914 to 1950 in all Ortega's discussions of Husserl. However, Ortega incorporated into his own thought a modified version of the unmediated intuition of the pre-1907 Husserl. 'Vital reason' is most concisely formulated in the phrase, 'I am myself and my circumstance'; according to Ortega, I am a self-aware interaction with my concrete surroundings, a repertory of possibilities which my intuitive reason helps me realize. The word 'circumstance', as employed in 1914, also connotes Ortega's homeland. So, unlike Unamuno, who perceives an irreconcilable clash between reason and life, Ortega applies Husserlian intuitive reason to resolve problems of national life or 'circumstance'.

Ortega's thinking on life and reason deepened in 1929. Reading the work of Husserl's student Heidegger in 1928 helped Ortega to organize his earlier philosophy into the four principles first formulated in *What Is Philosophy?*: (1) my life is the 'radical reality', the framework in which all other realities (including reason) appear; (2) my life is a self-conscious problem of individual realization; (3) my problem amounts to deciding among specific possibilities for self-authentication; (4) the plurality of these possibilities defines my freedom in life; their quantitative finiteness, my limitation. But, rejecting Heidegger's ontological jargon, the mature Ortega adopted the life-centred lexicon of Wilhelm Dilthey, whose philosophy of history Ortega read thoroughly in 1928. Thus emerged the last step in the evolution of 'vital reason', its absorption of what Dilthey called 'historical reason'. Ortega's third basic principle defines life as decision among possibilities, but what determines the decision, according to Dilthey, is a generational worldview (*Weltanschauung*), a system of subrational beliefs. Narrating the beliefs guiding great decisions can solve cultural problems.

What, therefore, is 'vital reason' as method? First, the posing of a current philosophical problem; second, phenomenological reduction of the problem; finally, narration of historical causes of the problem so as to solve it in practice. The structure of Ortega's most concise book illustrates this - *Historia como sistema (History as System)* (1935). Here he poses the problem of the loss of faith in scientific reason, because it is felt to be incapable of defining human being. How then should one reorient humankind in the universe? Phenomenological reduction prescribes the suspension of historical forejudgments, ascent to universals, descent from universals to essences, and description of essences. So Ortega suspends the historical forejudgments, received from the natural sciences, that humans have a nature, a changeless substance knowable with scientific logic. He also dismisses ideas of humans as body or psyche, since he regards such ideas as naturalistic. Instead, he finds that the 'radical reality' of each human being is precisely the life of that individual, and he describes that existence by enumerating its main aspects: its self-consciously problematic quality, the capacity for making decisions from among possibilities, and the limitations on these possibilities. Since historical reason can explain the decision-making, history is the discipline which will orient human beings in the universe by narratively clarifying their beliefs, defining them and instructing them in past errors to avoid such errors in the future. In proof, Ortega narrates the history of the West, in which belief in God gave way to belief in nature (whose study is natural science), and once scientism faltered, there arose belief in nothing but disillusionment, the science of which is history.

3 Works

Ortega's production exceeds the dozen volumes of over 500 pages each in the *Alianza* edition of his *Obras Completas*. Like Simmel and Max Scheler he addresses a vast thematic spectrum; let us summarize, first, his most translated work, *The Revolt of the Masses*, and, second, his *Leibniz* book, the one most deserving of philosophical attention.

Like all his major writings, *The Revolt of the Masses* begins with the posing of a current issue: why do the masses demand space in cultural sites once reserved for the moral and intellectual elite? Ortega dismisses as historical

forejudgments all political and economic explanations. Instead, he accepts what for him is a self-evident universal: an idea of society as dynamic interaction between moral and intellectual minorities and multitudes. Out of these two pure human types - elites and masses - Ortega derives a definition of the then-current mass type: a hybrid, it combines the great potential of the elite with the inertia of the mass type of bygone eras. Though powerful, the masses of the 1930s create nothing and even rebel against the elite, which does in fact create. The revolt evinces cultural strengths - the rise in the quality of life, the adventure of living in a dangerous era, the growth of personal possibilities - but even greater weaknesses - mass aggressiveness in public affairs, opposition to liberalism, obliviousness to history, overspecialization, excessive reliance on the state, and historical demoralization. Ortega proposes as a panacea the formation of a 'United States of Europe'.

From 1946 to 1950, he penned his Leibniz book in response to intellectually rebellious 'mass-men' - Spanish and Mexican Neo-Thomists who wrongheadedly criticize his mental style as metaphorical, overly dependent on sensory evidence and lacking stable principles. In 1946, on the tercentennial of Leibniz's birth, he posed a central problem: why, after formulating many philosophical principles, did Leibniz treat them so lightly? The implication was that the playful Ortega was in some way akin to the rigorous Leibniz, and this conclusion emerges along with the corollary that Ortega's foes themselves lack rigour and principles. Using 'historical reason', he shows that Aristotle, the father of Neo-Thomism, displayed excessive dependence on the senses, inexactness and inconsistency in handling principles. Ortega divides all philosophers into the disciplined, principled variety, based on intuition and not the senses (Plato, Descartes, Leibniz, Husserl, Dilthey and Ortega) and the less 'principled' type (Aristotle, the Stoics, the Schoolmen, Kierkegaard, Ortega's rivals Unamuno and Heidegger, and his detractors, the Neo-Thomists). 'Aristotelians' tend towards sombreness, anguish, inimical to genuine philosophy, while the Presocratics and Plato likened philosophizing to riddle-solving or participating in sports and games. This attitude, Ortega feels, confirms his distinction between rational ideas and subrational beliefs; humans act upon ideas or mere theories without ultimate seriousness, while beliefs act upon humans and are taken seriously. Therefore Ortega's apparent non-seriousness, like Leibniz's, does not negate his discipline. He confesses early opposition to Unamuno's tragic sense of life, the clash between reason and life, by speaking of a 'sportive and festive sense of existence', which implied an intellectual sportsman's endeavour to harmonize life and reason. 'Vital reason' is no more deficient in principles than the philosophies of Plato, Husserl and Dilthey. Ortega, one of the twentieth century's greatest essayists, deserves recognition as a brilliant transitional philosopher, bridging Husserl's phenomenology and Heidegger's ontology of existence with his own four principles.

See also: Spain, philosophy in

NELSON R. ORRINGER

List of works

Ortega y Gasset, J. (1914) *Meditaciones del Quijote*, Madrid: Residencia de Estudiantes; trans. E. Rugg and D. Marín, *Meditations on Quixote*, New York, 1961.(Ortega's first book, intended to be one of a series of ten - not completed. It contains Ortega's ideas on 'radical reality', and the seminal phrase 'I am myself and my circumstances'.)

Ortega y Gasset, J. (1923) *El tema de nuestro tiempo*, Madrid: Calpe; trans. J. Cleugh, *The Modern Theme*, New York, 1933; repr. 1961.(Ortega's exposition of perspectivism and vitalism.)

Ortega y Gasset, J. (1925) *La deshumanización del arte e ideas sobre la novela*, Madrid: Revista de Occidente; trans. H. Weyl, *The Dehumanization of Art: Ideas on the Novel*, Princeton, NJ, 1948.(Ortega's best-known works of aesthetics, combining an appraisal of the state of art in the early twentieth-century Europe with a characteristically energetic exposition of how aesthetic experience functions.)

Ortega y Gasset, J. (1929-30) *Čiquest;Qué es filosofía?*, Madrid: Revista de Occidente, 1958; trans. M. Adams, *What Is Philosophy?*, New York, 1960.(A course of eleven lectures given by Ortega in 1929, of note because closure of the university meant that the last ten were given in a theatre. A good example of Ortega as a communicator.)

Ortega y Gasset, J. (1930) *La rebelión de las masas*, Madrid: Revista de Occidente; trans. A. Kerrigan, *The Revolt of the Masses*, Notre Dame, IN, 1986.(Radically elitist analysis of the threat posed to society, culture and democracy by the masses.)

Ortega y Gasset, J. (1935) *Historia como sistema y Concordia y libertad*, Madrid: Revista de Occidente; trans. H. Weyl and W. Atkinson as 'History as System', in *Philosophy and History*, Oxford, 1941; also as *History as a System and Other Essays toward a Philosophy of History*, Greenwood, CT, 1961.(Ortega formally

conceptualizes the idea of historical reason as narrative reason.)

Ortega y Gasset, J. (1946-83) *Obras Completas (Complete Works)*, Madrid: Revista de Occidente, vols 1 and 2, 1946; vols 3-6, 1947; vol. 7, 1961; vols 8 and 9, 1962; vols 10 and 11, 1969; vols 1-11 repr. and vol. 12, Madrid: Alianza Editorial, 1983.(Standard edition of the author's complete works.)

Ortega y Gasset, J. (1947-50) *La idea del principio en Leibniz y la evolución de la teoría deductiva*, Madrid: Revista de Occidente; trans. M. Adams, *The Idea of Principle in Leibniz and the Evolution of Deductive Theory*, New York, 1971.(Essay in which Ortega contrasts philosophers according to principle, and focuses his own playful style of 'vital reason', or 'reason from life's point of view'.)

References and further reading

Cerezo Galán, P. (1984) *La voluntad de aventura. Aproximación crítica al pensamiento de Ortega*, Barcelona: Ariel.(Insightful on Ortega's early reactions to Husserl and Heidegger.)

Donoso, A. and Raley, H.C. (1986) *José Ortega y Gasset: A Bibliography of Secondary Sources*, Bowling Green, OH: Philosophy Documentation Center.(The most complete bibliography of research about Ortega, alphabetized by authors and themes.)

Orringer, N.R. (1984) 'La crítica de Ortega a Aristóteles y sus fuentes', *Cuadernos Salmantinos de Filosofía* (Salamanca) 11: 557-98; repr. in M. Durán, *Ortega, hoy*, Veracruz, 1985, 173-224.(Contributor's detailed source study of Ortega's *Leibniz* book.)

Orringer, N.R. (1984) 'Ortega y Gasset's Critique of Method', *Comparative Criticism* 6: 135-54.(Presents the arguments of §2 above.)

Ouimette, V. (1982) *José Ortega y Gasset*, Boston, MA: Twayne.(Concise biography and brief but clear synthesis of Ortega's philosophy.)

Oswald, James (1703-93)

James Oswald, Scottish theological writer, used the philosophy of 'common sense' to try to found religious and moral conviction on principles that were impervious to scepticism. In a long running controversy over Church discipline, he defended the right of individual parishes to choose their ministers, seeing the prevailing system of patronage as favouring the advocates of a fashionable kind of civility that was too tolerant of scepticism and intellectual innovation, and too indifferent to the Church's traditional concerns with public and private morality. Though never part of the Aberdeen philosophical community, Oswald corresponded with Reid, and late in life collaborated with him in charitable work for the sons of clergy.

James Oswald was born at Dunnet in Caithness, where he succeeded his father as parish minister in 1726. He moved to Methven, Perthshire, in 1750, ministering there till 1783. He was Moderator of the General Assembly of the Church of Scotland in 1765, when he was awarded an honorary doctorate at Glasgow. As early as 1732, reporting to the Caithness Presbytery on a colleague's conduct, Oswald cited 'the common sense of mankind' to determine the binding character of promises. In *Submission and Obedience* (1753) he used the same authority to justify seeing the institutions of government as part of the divine plan for the world, arguing that the Church's mission could only be accomplished on the assumption that it too operated according to comparable principles of good order and government. From these derive obligations to duty, both civil and religious, which are 'obvious to the Understanding of Men'. This is reminiscent of George Turnbull's contention in his *Philosophical Enquiry* (1726, published 1731) that whatever is fit and binding in individuals' relations with one another and with God can be known by common sense (see Turnbull, G.). Turnbull and Oswald were products of the same divinity school at Edinburgh; they also shared a view of miracles as law-governed events.

'Common sense', meaning a direct deliverance of our common human reason or cognitive faculty, was cited by Oswald to justify practices the abandonment of which would undo social institutions, affront conscience or undermine religion. He subsequently extended it to cover all beliefs a challenge to which could be considered destructive of sense or sanity. Wherever the self-evidence of common-sense judgments was impugned, some form of incoherence (social or intellectual) would ensue. The commonest way of seeming to impugn them was to allow that they needed, or were even capable of, reasoned justification. On the contrary, they were held to constitute the primary truths which all reasoned justification must already assume (see Common sense school).

Oswald was unperturbed at the diversity of principles to which he accorded common-sense status, and in *An Appeal to Common Sense* (1766-72) resisted any attempt to give a precise analysis of this status or rest it on more formal criteria. The thread holding his primary truths together is that they had virtually all been targets of Hume's scepticism (see Hume, D.). Oswald commended Hume for seeing they are beyond the scope of reasoning, but he equated this too narrowly with strict demonstration, and argued against Hume that neither custom nor sentiment, both of which he considered highly fallible resources and the roots of bigotry and folly, could account for the transparent certainty of these unreasoned truths. Some degree of mental maturity, intelligent observation, even repeated experience, may be necessary to our initial appreciation of them, but thereafter their 'self'-evidence exceeds any other evidence we could cite for them. Oswald also criticized Francis Hutcheson and Henry Home for placing moral perception in feeling rather than cognitive judgment, but he agreed that moral training proceeds more by engaging the heart than the intellect (see Hutcheson, F.; Home, H. (Lord Kames)).

Oswald's primary truths fluctuate between the particular and the universal. They include beliefs with regard to our own existence and continued identity; the existence, identity and relative placings of objects around us and their motions, powers and causal operations expressed in the laws of nature; the existence of other sentient life with powers of action and traits of character; mathematical axioms; the being and providence of God; the real distinction between vice and virtue and its enforcement through punishment and reward; and the existence of obligations to God, others and ourselves. These are topics too central to this life and the next to be left to the rarified argumentation of the learned, and Oswald supposed it inherently ridiculous that the intelligent multitude should remain in suspense on them. A certain amount of unacknowledged chauvinism entered into Oswald's characterization of this multitude, whom he tended to see reflected in the persons of property and judgment who formed the mainstay of the Church of Scotland.

The laws of nature, for Oswald, are not formal, mathematical laws, but such relatively nebulous principles as that bodies sink by virtue of their weight, or that fire has the power to consume combustible materials. He means that the gravity of bodies and the power of fire are too much facts of life to be subject to caveats of inductive logic. In arguing that God's existence is evident to common sense, he wishes to avoid the metaphysical nicety of a priori demonstration and the inductive hazards of argument by analogy. To detect design, even on a limited scale, in inanimate nature is no different in principle from recognizing it among animals and humans, which in turn is no different (notwithstanding Hume's argument to the contrary) from actually experiencing the powers of things in their manifestation; and once Oswald has reached this point he develops his argument beyond the strict requirements of common sense philosophy, into a general theodicy.

Oswald's *Appeal* was sympathetically received in *The Monthly Review*, and by the Dissenter William Enfield in a tract directed to Joseph Priestley (see Priestley, J.). Philip Skelton commended Oswald's argumentative skill but thought he undervalued revelation. Favourable notices in *Göttingische Anzeigen* led to a German translation (1774), but the popularity of Oswald's work was shortlived. Priestley (1774) condemned his declamatory rhetoric, his restricted understanding of 'reason', and his dogmatic use of 'common sense' as an arbitrary block to legitimate enquiry. The translator of Claude Buffier's *First Truths* (1780) (see Buffier, C.) thought Oswald grossly over-extended the scope of common sense, and that he failed to see how often he was describing what were implicitly reasoned judgments.

See also: Enlightenment, Scottish

M.A. STEWART

List of Works

Oswald, J. (1753) *Some Thoughts Relating to that Submission and Obedience Due to the Authority and Decisions of the Supreme Judicature of the Church. Communicated in Two Letters from One Clergyman to Another*, Edinburgh: Wright.(Issued anonymously.)

Oswald, J. (1766) *A Sermon, Preached at the Opening of the General Assembly of the Church of Scotland, May 22. 1766: To which are annexed, Letters on Some Points of Importance Contained in the Sermon*, Edinburgh: Kincaid and Bell.

Oswald, J. (1766-72) *An Appeal to Common Sense in behalf of Religion*, vol. 1, Edinburgh: Kincaid and Bell; vol. 2, Edinburgh: Kincaid and Creech; German translation by F.E. Wilmsen, *Appelation an den gemeinen Menschenverstand zum Vortheil der Religion*, Leipzig, 1774, 2 vols.(Issued anonymously.)

Oswald, J. (1767) *Letters concerning the Present State of the Church of Scotland, and the Consequent Danger to Religion and Learning, from the Arbitrary and Unconstitutional Exercise of the Law of Patronage*, Edinburgh: Gray.(Issued anonymously.)

Oswald, J. (1770) *The Divine Efficacy of the Gospel Dispensation: A Sermon Preached before the Society in Scotland for Propagating Christian Knowledge*, Edinburgh: Kincaid and Bell.

Oswald, J. (1774) *Six Sermons on the General Judgment*, Edinburgh: Gray.

Oswald, J. (1791) *Hypocrisy Detestable and Dangerous: Four Sermons*, Glasgow: Duncan and Chapman.(Issued anonymously.)

References and further reading

Anon. (1780) *'Preface' to First Truths, and the Origin of our Opinions, Explained: With an Enquiry into the Sentiments of Modern Philosophers relative to our Primary Ideas of Things: Translated from the French of Père Buffier, to which is prefixed a Detection of the Plagiarism, Concealment, and Ingratitude of the Doctors Reid, Beattie, and Oswald*, London: Johnson.

Ardley, G. (1980) *The Common Sense Philosophy of James Oswald*, Aberdeen: Aberdeen University Press.(A polemical, somewhat partisan exposition of Oswald's *Appeal to Common Sense*.)

Cooper, J. (1948) *James Oswald (1703-93) and the Application of the Common Sense Philosophy to Religion*, unpublished dissertation, University of Edinburgh.(A slightly sentimentalized intellectual biography, offering a convenient survey of Oswald's principal writings, and original research on his early career.)

Enfield, W. (1770) 'A Second Letter to the Reverend Dr Priestley', appended to *Remarks on Several Late Publications Relative to the Dissenters: In a Letter to Dr Priestley*, London: Bladon.(Issued anonymously.)

Kuehn, M. (1987) *Scottish Common Sense in Germany, 1768-1800*, Kingston, Ont., and Montreal: McGill-Queen's University Press, ch. 3.(Examines the reception of Oswald's philosophy.)

Laurie, H. (1902) *Scottish Philosophy in its National Development*, Glasgow: Maclehose, ch. 9.(Includes a hostile appraisal of Oswald's work.)

Priestley, J. (1774) *An Examination of Dr Reid's Inquiry into the Human Mind on the Principles of Common Sense, Dr Beattie's Essay on the Nature and Immutability of Truth, and Dr Oswald's Appeal to Common Sense in behalf of Religion*, London: Johnson; repr. New York: Garland Publishing, 1978.

Sher, R.B. (1985) *Church and University in the Scottish Enlightenment*, Edinburgh: Edinburgh University Press. (Discusses Oswald's role in Church politics.)

Skelton, P. (1784) 'Some thoughts on Common Sense, Intended for the More Intelligent Reader' in *Works*, vol. 6, Dublin: privately published.

Turnbull, G. (1731) *A Philosophical Enquiry Concerning the Connexion Betwixt the Doctrines and Miracles of Jesus Christ*, London: Willock.(Published under the pseudonym Philanthropos.)

Other Minds

It has traditionally been thought that the problem of other minds is epistemological: how is it that we know other people have thoughts, experiences and emotions? After all, we have no direct knowledge that this is so. We observe their behaviour and their bodies, not their thoughts, experiences and emotions. The task is seen as being to uncover the justification for our belief in other minds. It has also been thought that there is a conceptual problem: how can we manage to have any conception of mental states other than our own? It is noteworthy that there is as yet no standard view on either of these problems. One answer to the traditional (epistemological) problem has been the analogical inference to other minds, appealing to the many similarities existing between ourselves and others. This answer, though it is no longer in general favour among philosophers, still has its defenders. Probably the favoured solution is to view other minds as logically on a par with the unobservable, theoretical entities of science. That other people have experiences, like us, is seen as the best explanation of their behaviour.

1 What generates the two problems of other minds?

The epistemological problem arises from two facts. We lack direct knowledge of the mental states of other people. We have such knowledge of at least some of our own mental states. Put the two claims together and we have the traditional problem of other minds. The relevant asymmetry, between our own case and that of others, turns on the question of direct knowledge, not observation. Being able to observe the mental states of others would not enable us to avoid the problem. What would be needed would be the ability to observe those mental states as the mental states of others. They would have to come labelled. The situation would only then be symmetrical. We would have the direct knowledge we lack.

The same asymmetry generates the conceptual problem. How can we have the concept of other people's experiences given that we have direct knowledge only of our own experiences, labelled as our own? Once again, the problem is not that we cannot observe the pains of others. What would be needed would be observing such pains as, indeed, the pains of others.

There has been comparatively little discussion of the conceptual problem, and no more will be said here about the conceptual problem of other minds, other than to note that solving it would not, other than controversially at best, remove the epistemological problem. That problem will be hereinafter the one to be discussed.

2 Who has the epistemological problem?

The heroic way of avoiding the problem of other minds is to deny that the claimed asymmetry, between ourselves and others, holds. Some have done so by insisting that we have direct knowledge of the mental states of others, though this has generally been seen as implausible. However, there seems to be an important strand of thinking within feminist theory that would endorse this rejection of our asymmetry (see Feminism and psychoanalysis). Similarly, continental European philosophy has commonly taken the view that other people are needed for us to acquire our own sense of ourselves as persons (see Alterity and identity, postmodern theories of). So our sense of others goes before our sense of self. That would seem to demand some capacity to know about others before one knows about oneself. The asymmetry would, presumably, be reversed, and the problem of other minds give way to the problem of our own minds.

Given that the asymmetry is accepted, and thus the traditional understanding of the problem of other minds, it has been almost a commonplace to believe that only a traditional dualist view of the mind produces a difficult problem of other minds (see Dualism). Though generally theories of mind accept the asymmetry, not all are thought to have a difficult problem. Behaviourism (see Behaviourism, analytic) is a theory of mind that either is thought to have no problem at all, or, if it does, nevertheless has no difficulty in solving the problem. There is no special problem about knowing about the behaviour and behavioural dispositions of another.

Functionalism is another theory of mind that is thought not to have a difficult problem. Mental states are viewed as internal states of the organism, regulating its responses to its surroundings. The various mental states are differentiated by their various roles, and they have no other features relevant to their being the mental states they are (see Functionalism). It is then claimed to be straightforward, faced with the behaviour of others, to infer that

such internal states exist, in the appropriate relations to the behaviour. Eliminative materialists have been seen as not having any problem. If there are no minds, then there are no other minds, therefore there is no problem of other minds (see Eliminativism).

However, it has been argued that all theories of mind leave the other minds problem intact, and difficult. Theories of mind are, indeed, theories of mind, all minds, one's own and others. Given that they are to be true of all minds, including other minds, they cannot, it has been argued, be used to show that there are, indeed, other minds. It is unacceptable to argue that, say, functionalism is true, and then use this to solve the other minds problem, since it cannot be known to be true, to hold of minds in general, unless it holds of other minds. How could that be known without the other minds problem having been, somehow, solved?

In conjunction with this line of argument, it has been pointed out that a theory of mind, embracing all minds, needs to embrace its propounder's mind as well as other minds. So a crucial part of the evidence (generally implicit) for a theory of mind will be that the theory fits the propounder's experience. That is the only way direct evidence in its favour can be obtained. So the propounder's experience is crucial. It has been argued that there is no escaping dependence on that experience and, as we shall see, any such dependence has been seen to be a fatal weakness, whenever it exists, in any would-be justification for belief in other minds.

3 The analogical inference to other minds

The traditional solution to the problem of other minds has been the analogical inference to other minds. Other people behave like me in similar circumstances and have the same physico-chemical composition. When I burn myself it hurts and I cry out and wince. When other people are burned they cry out and wince. I can thus infer that they are in pain too. More generally, others are very like me. I know I have beliefs, experiences and emotions. So I am entitled, given how like me they are, to infer that other people also have beliefs, experiences and emotions.

This traditional analogical inference to other minds is now generally incorporated in a hypothetic inference (scientific inference, inference to the best explanation) to other minds. That there are alternative hypotheses about others which have to be ruled out requires that the argument take the form of a hypothetic inference (see Inference to the best explanation). But the appeal to one's own case, and to similarities, remains crucial in this analogical/hypothetic inference.

That, indeed, one's own case is crucial, gives rise to the classical, continuing objection to the analogical inference to other minds, that it is a generalization from one case. Such generalizations are almost invariably unsound. Though there have been attempts to put the analogical inference to other minds in a form that avoids this objection, it is generally accepted that such attempts have not succeeded.

However, its supporters argue that, even so, the analogical/hypothetic inference remains a sound inference, despite its dependence on one case. More than one case is needed where what is at issue is the question of a causal link between events. Where it can be known from one case that there is such a causal link, that one case will then be enough. It is claimed that the relevant causal link, involving mental states, can be known to hold from one's own case. Though it has traditionally been insisted that the relevant causal link is between mental states and behaviour, it has been urged that the relevant causal link needs to be between brain states and mental states if the analogical/hypothetic inference is to be defensible.

The other classical objection to the analogical inference to other minds has been that its conclusion was impossible to check, not just in fact, but in principle. This feature no longer seems to be seen as having any epistemological relevance.

4 Other minds as theoretical entities

This is probably, among Anglo-American philosophers, the favoured solution to the problem of other minds. The justification is in the form of a hypothetic inference. That others have mental states is hypothesized to account for how they behave. However, one proceeds purely from the outside. No evidence gathered from one's own case is used to support this hypothesis. The one case objection thought widely to be fatal to the analogical inference is, crucially, avoided, so it is widely believed.

It is generally considered that treating other minds in this way, as theoretical entities, will succeed if one has a

functionalist (or some such) view of the mind. The two seem made for each other. However, the argument in §1, that no theory of mind has an advantage over any other in supporting belief in other minds would, if successful, apply to this particular attempt to avoid dependence on one's own case.

It has been argued, conversely, that unless one enlists the help of a functionalist (or some such) theory of the mind, treating other minds as theoretical entities will not succeed. A traditional view of mental states, allowing that they have intrinsic content - in particular, phenomenal properties, such as the hurtfulness of pain (see Qualia) - cannot be supported by this method. Treating the mental states of others as theoretical entities, it is argued, will not provide those states with the needed intrinsic properties. That content can only be filled in by an appeal to one's own case.

5 Criteria and other minds

Treating other minds as theoretical entities, though an alternative to the analogical inference, takes the form of a hypothetic inference. Criterialists, by contrast, have sought to avoid the one case problem by eschewing any form of inference. They have insisted that the link between behaviour and mental states is neither entailment (as in behaviourism) nor an inductive inference. The link is claimed to be conceptual and such links are characterized as criterial. Behaviour is a criterion for the presence of mental states. It has been claimed by some that such a non-inferential connection is required if we are to have any concept of the experiences of others.

An example of such a claimed non-inferential link would be the claim that itching is conceptually linked to scratching, not merely contingently correlated. Our concept of itching is that it disposes one to scratch. It is further claimed that scratching is, thereby, evidence of itching, given that itching disposes one to scratch.

That there are such conceptual links has been widely argued and widely denied. That, if there are, they would provide a sufficient basis for belief in the mental states of other people has been, if anything, even more vigorously contested. The thrust of the attack on the use of criteria to support belief in other minds seems to be that such conceptual links fail to bridge the gap between observed behaviour and the unobserved inner states to which they are conceptually linked. If there is no entailment directly from the one to the other, and there is no appeal to some form of inductive inference, it is argued that we are left with the gap. The gap cannot be crossed by *fiat*, as it were.

One way of understanding what has been called the attitudinal approach to other minds, is to see it as going beyond other uses of criteria in insisting that our conception of other human figures is that they are souls, have experiences. That is how we perceive them. It is as immediate as that, preceding any belief. This criterial view seems, however, to inherit the criterial gap. However we conceive of reality, our conceptions might be mistaken. There are attitudes to things and people which, though more immediate and deeper than inferential belief, are, nevertheless, mistaken (racist and sexist attitudes might be given as instances).

6 Private language and other minds

It has been widely accepted that language is a public phenomenon. Some have insisted that it is essentially public. One way of understanding this claim is classically associated with Wittgenstein. A language that is necessarily private, that is, such that only one person can understand it, is (logically) impossible (see Private language argument).

The connection with the problem of other minds has been controversial. However, it seems clear that the connection exists in the reason given for the impossibility of such a private language. That reason bears directly on the analogical inference to other minds.

A necessarily private language is claimed to be impossible because a language has to be, in principle, subject to checking by someone other than an individual user of the language. Generally, a user of the analogical inference to other minds is in breach of this principle. They have insisted that each of us knows what psychological terms mean (at any rate, some of them) from our own case and only from our own case. Their usage would not then be, in principle, one that could be checked for consistency. Functionalists, by contrast, make no such claim, and it should be noted that the connection from private language to other minds depends on the use of terms whose meaning would be private in the relevant sense. The connection is not directly to any particular argument to other minds.

The argument that a check is needed (in principle, so Robinson Crusoe is not in trouble) has generally been that, in

its absence, no distinction can be made between its seeming to the language user that their usage is consistent, and its being so. They have only their impression to go on. The issue has been vigorously contested.

That a private language is impossible has not generally been used explicitly as an argument for other minds. Nor could it be. After all, it is only in principle that a language is to be checkable by other people. An important, though indirect role, however, could be seen as its supporting the criterialist insistence on the need for a conceptual connection between inner states and publicly observable states (see Criteria).

See also: Mind, child's theory of

ALEC HYSLOP

References and further reading

These are all non-technical and manageable with reasonable effort.

Buford, T.O. (ed.) (1970) *Essays on Other Minds*, Chicago, IL: University of Illinois Press.(A useful, varied collection, including some classical items, particularly two by N. Malcolm expounding Wittgenstein.)

Hill, C.S. (1991) *Sensations: a Defense of Type Materialism*, New York: Cambridge University Press.(Chapter 9, 'Knowledge of other minds', discusses the conceptual problem of other minds, and also defends a version of the analogical inference to other minds.)

Hyslop, A. (1994) *Other Minds*, Dordrecht: Kluwer.(An extended defence of the analytical/hypothetic inference to other minds.)

McDowell, J. (1982) 'Criteria, Defeasibility, and Knowledge', *Proceedings of the British Academy*, 68: 455-79. (Hostile to criteria but exploring the issues at a fundamental level.)

McGinn, C. (1984) 'What is the Problem of Other Minds?', *Proceedings of the Aristotelian Society*, supplementary vol. 58: 119-37.(A discussion of the conceptual problem of other minds.)

Pargetter, R. (1984) 'The Scientific Inference to Other Minds', *Australasian Journal of Philosophy* 62: 158-63.(A good (favourable) account of other minds as theoretical entities.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(The classical source for criterialism and private language.)

Otto, Rudolf (1869-1937)

Rudolf Otto, an early and leading student of religious experience, was a devout Christian thinker (part theologian, part philosopher, part phenomenologist of religious experience) who was strongly influenced by the philosophy of Immanuel Kant. He held that numinous experience - experience of the uncanny that is strongest and most important in cases in which it seems to its subject to be experience of God - is unique in kind. Such experience of God, he held, occurred in both Semitic and South Asian monotheistic traditions. Recognizing the intellectual or doctrinal content of numinous experience, but influenced by Kant's thesis that knowledge-giving concepts cannot refer beyond possible objects of sensory experience, Otto tried to remain faithful to both numinous experience and Kantian philosophy by talking about 'ideograms' that express the content of numinous experience but, allegedly at least, are not concepts.

Rudolf Otto was born into a devout Lutheran family in Peine, Hanover. Between 1888 and 1891 he studied theology in Erlangen, Göttingen and Bavaria. In 1891 he returned to Göttingen, where he became acquainted with Ritschlian theology. He received his licentiate in theology there for a study entitled 'Spirit and Word According to Luther', and was appointed a *Privatdozent* in 1899. He became Professor of Systematic Theology in 1904, a post which he also held in Breslau (1914-17), and in Marburg from 1917 until his retirement in 1929. He travelled widely in Egypt, Palestine, Greece, North Africa, India and the Far East. He learned Sanskrit and was an early figure in the effort to make South Asian and Indian thought accessible to a European audience. Otto was a pioneer in the history and phenomenology of religion, and the author of profound studies of mysticism, Śaōkara, Meister Eckhart, India's religions and the kingdom of God. The present entry will focus on his view of religious language and his account of numinous experience.

Otto, influenced by Kant, denied that concepts apply to God. But he also asserted the fundamental importance of doctrines about God and contended that it is possible for human beings to experience God. Kant had claimed that no concept could apply, in such a way as to produce knowledge of anything, to anything that cannot somehow be sensed. The application of concepts beyond the range of objects of possible sensory experiences, he held, yields no knowledge. Since God is not a candidate for being literally sensed, the concept 'God' can have no knowledge-providing application. Otto, working in Kant's shadow, proposed to substitute 'numinous ideograms' for concepts. God is inaccessible to reason, but, Otto held, can be felt in numinous experience, where one is encountered by God. This enables one to form an ideogram of God. Otto's examples of ideograms are 'the Void' and 'the Wholly Other', and he wrote that the necessity for ideogramic expression and the ineffability of the numinous experience 'teaches us the independence of the positive content of the experience from the implications of its overt positive expression' ([1917] 1923: 34). His acceptance of a Kantian perspective, his contention that numinous experience is incapable of full conceptualization, his view that God has a conceptually accessible aspect and a conceptually inaccessible aspect, and his attraction to negative theology blended together to yield talk about ideograms. None of this led Otto to conceive of God as irrational or to say that contradictions can be true of God. None the less, the force of 'ideogram' is suspiciously like that of 'concept', applied to something that is not an object of possible sense experience.

Otto is most famous for his emphasis on numinous experience, which ranges from a vague sense of the uncanny to explicit putative experiences of God. These latter, stronger versions of numinous experience form our primary grounds for thinking that there is a God. They are, Otto believed, genuine encounters with God, the powerful content of which leaves no doubt as to the reality of the experience's object. To have a numinous experience is to experience (at least apparently) a being who is majestic, awesome, living, holy and overwhelming, in whose presence one realizes one's unholiness, dependence and comparative insignificance. What is most important for Otto is the being that is experienced; one's reactions to it are but corollaries to the being's own attributes. Further, the object of numinous experience possesses ultimate and unsurpassable worth. A human being is passive, not active, in such experiences. Otto believed that such experiences guarantee their own veridicality, needing no help from natural theology or philosophical argument.

Kant claimed that moral experience is not reducible to any other sort of experience, and that morality has an indefeasible validity all its own. Otto agreed about morality and made the same claim regarding religion and

numinous experience. Thus he stressed that numinous experiences, while in some ways resembling other experiences, are none the less unique in their content. Further, they are inexplicable in natural terms. Hence religion has a genuine object and theology a distinct subject matter (see Religious experience).

However, it is not clear that even if numinous experience is indeed unique in kind that this yields veridicality; on one popular view of secondary qualities, each is unique in type and secondary qualities are mind-dependent. Nor is it clear that numinous experience is not veridical even if it is not unique in content. But be that as it may, Otto's emphasis on the centrality of numinous experience in monotheistic traditions has had a strong influence on the philosophy and phenomenology of religion.

KEITH E. YANDELL

List of works

Otto, R. (1907) *Naturalismus und Religion*, Tübingen: Mohr; trans. J.A. Thomson and M.H. Thomson, *Naturalism and Religion*, London: Williams & Norgate, 1907.(Here Otto endeavours to keep what seems right in naturalism without eliminating religion.)

Otto, R. (1909) *Kantisch-Fries'sche Religionsphilosophie*, Tübingen: Mohr; trans. E.B. Dicker, *The Philosophy of Religion Based on Fries and Kant*, London: Williams & Norgate, 1931.(Otto's philosophy of religion discussed in terms of the perspectives that most influenced it.)

Otto, R. (1917) *Das Heilige*, Munich: Beck, 25th edn, 1936; trans. J.W. Harvey, *The Idea of the Holy*, Oxford: Oxford University Press, 1923.(Otto's *magnum opus*, in which he develops his account of numinous experience.)

Otto, R. (1926) *West-Östliche Mystik*, Gotha: Klotz; trans. B.L. Bracey and R.C. Payne, *Mysticism East and West*, New York: Macmillan, 1932.(Interesting and controversial comparison of Śaōkara and Eckhart.)

Otto, R. (1930) *India's Religion of Grace and Christianity Compared*, trans. F.H. Foster, London: SCM Press. (Argues that Hindu monotheism includes a theology of divine grace.)

Otto, R. (1931) *Religious Essays*, trans. B. Lunn, London: Oxford University Press. (Essays supplementary to *The Idea of the Holy*.)

Otto, R. (1938) *The Kingdom of God and the Son of Man*, Grand Rapids, MI: Zondervan Press.(Otto's interpretation of the kingdom of God.)

References and further reading

Almond, P.C. (1984) *Rudolph Otto: An Introduction to His Philosophical Theology*, Chapel Hill, NC: University of North Carolina Press.(Good, accessible discussion of Otto's thought and influence.)

Davidson, R.F. (1947) *Rudolph Otto's Interpretation of Religion*, Princeton, NJ: Princeton University Press. (Discussion of Otto's notion of the numinous as religiously central.)

Yandell, K.E. (1971) *Basic Issues in the Philosophy of Religion*, Boston, MA: Allyn & Bacon.(Discusses Otto's account of religious experience; see chapter two.)

Overton, Richard (d. *circa* 1665)

Overton was one of the leading figures of the radical Leveller movement in England in the 1640s. He fought for the equality of all men before the law and for complete religious and political toleration, often by appealing to notions such as the social contract and the natural law. In metaphysics he denied that the soul is a separate immaterial and immortal substance, arguing that immortality is not achieved until the resurrection. His views on the soul may have influenced Milton.

Overton's dates are unknown. An Englishman, he may have spent some time in the Netherlands and appears to have matriculated in Queens' College, Cambridge in 1631. He was most active from the early 1640s to the mid-1650s, and was arrested and imprisoned several times for political reasons. He probably died in the mid-1660s.

In numerous pamphlets, some of which were published under the pseudonym 'Martin Marpriest', Overton expressed his political views about religious toleration, democratic republicanism, the importance and dignity of the individual's property, and the equality of all men before the law. Often this was done in a vivid humorous and satirical style. While Overton appealed to notions such as the social contract and the natural law, he did not present any detailed philosophical analysis of these concepts. His most important pamphlets include: *The Araignement of Mr. Persecution* (1645) which contains his ideas on toleration; *A Remonstrance of Many Thousand Citizens* (1646a) which calls for the abandonment of monarchy, arguing that all power rests with the people; *An Arrow against all Tyrants* (1646b) which emphasizes the political rights of the individual; and *An Appeale from the Degenerate Representative Body of the Commons of England Assembled at Westminster* (1647) which upholds the principles of 'right reason' that in some circumstances justify a direct appeal to the people. Also important are his editorials for the weekly newspaper *The Moderate* (1648-9). It has been argued that Overton played an important role in the development of seventeenth-century 'possessive individualism' according to which the individual is naturally the owner of their own person. However, the nature of Overton's individualism is still a matter of dispute among scholars.

Overton's most philosophical work is *Mans Mortalitie* (1643) (a revised version of the book first appeared in 1655 under the title *Man wholly Mortal*). Overton's denial of an immaterial and immortal soul, and his view that all hope for a future life is in the resurrection, link his work to doctrines in early Christian thought. Some of Overton's arguments are similar to those of Pomponazzi in *On the Immortality of the Soul* (1516). Other influences include Socinianism, George Wither's introduction and translation of Nemesis' *The Nature of Man* (1636), and Henry Woolnor's *The True Originall of the Soule* (1642) (see Socinianism).

According to Overton, the belief in an immortal soul is a 'heathenish invention' which later made its way into Christian dogma. He attempts to argue for his position both through an interpretation of the Bible and through philosophical reasoning. His arguments are very much couched in scholastic terminology, for example, 'form and matter' and 'potentiality and actuality'. To a large extent, his strategy consists in pointing out absurdities in the immortalist position.

Overton's main philosophical argument against the natural immortality of the soul is based on his view that what is called 'the soul' is not an immaterial substance: since man is wholly a compound material being, he cannot be intrinsically indestructable or immortal. As in his political writings, Overton in *Mans Mortalitie* emphasizes the rationality of man, but for him, rationality does not entail immateriality. Rather, he argues that rationality is a function of man's material being. The difference between man and beast is said to be only a matter of degree. Man is an animal whose (physical) parts are endowed with several faculties which make him a 'living Rationall Creature'. All of man's faculties depend on the body and are thus mortal: this is true of faculties which men have in common with beasts as well as of those which are distinctive of man (that is, reason, consideration, science and so on). The faculties exist in the human body as accidents exist in a subject. It is only through the body that the faculties constitute a unitary human self; their 'unite substance' is in the body. If it were possible for man's faculties to exist independently of the body, we would have to assume as many souls as there are faculties. Thus we would be led to the absurd conclusion that each faculty constitutes a separate soul and that there is no unitary self to which the various faculties belong.

Overton infers from the essential physical nature of man that (1) man is not by nature indestructible or immortal and that (2) man's immortality is a gift of God received at the resurrection. Thus, Overton does not deny immortality altogether, but he makes the afterlife dependent on the resurrection. While he rejects the notion of an immaterial human soul, it is not clear that his 'materialism' extends universally. What he says about the soul is at least consistent with the view that there are non-human immaterial spirits (for example, angels). But he assigns to God a material dwelling-place in subscribing to the belief that the sun is the seat of divinity.

Mans Mortalitie aroused considerable controversy. It was attacked as heretical in the House of Commons in 1644, and it came under attack in print by Guy Holland in *The Prerogative of Man* (1645), Alexander Ross in *The Philosophicall Touchstone* (1645), Thomas Edwards in *Gangraena* (1646), and several others. On the positive side, Overton's views on the soul are sometimes said to have influenced Milton; but the matter is controversial.

See also: Soul, nature and immortality of the

UDO THIEL

List of works

Overton, R. (1643) *Mans Mortalitie*; ed. H. Fisch, Liverpool: Liverpool University Press, 1968.(A reprint of the 1643 edition, with textual notes and a brief introduction by the editor.The most impotant differences between this and the revised 1655 edition are noted.)

Overton, R. (1645) *The Araignement of Mr. Persecution*; repr. in W. Haller *Tracts on Liberty in the Puritan Revolution 1638-1647*, vol. 3, New York: Columbia University Press, 1934.(Argues against religious uniformity.)

Overton, R. (1646a) *A Remonstrance of Many Thousand Citizens*; repr. in W. Haller *Tracts on Liberty in the Puritan Revolution 1638-1647*, vol. 3, New York: Columbia University Press, 1934.(Expresses Overton's republicanism; defends the equality of all men before the law.)

Overton, R. (1646b) *An Arrow against all Tyrants*; repr. with prefatory note, University of Exeter: The Rota, 1976.(Written in the form of a letter to 'Henry Martin, a Member of the House of Commons'; attempts to justify men's natural rights.)

Overton, R. (1647a) *An Appeale from the Degenerate Representative Body of the Commons of England Assembled at Westminster*, London.(Addressed to the New Model Army which is 'now the only formal and visible head that is left unto the people'. Argues that a direct appeal to the people can be justified in some circumstances.)

Overton, R. (1647b) *The Commoners Complaint*; repr. in W. Haller *Tracts on Liberty in the Puritan Revolution 1638-1647*, vol. 3, New York: Columbia University Press, 1934. (Recounts the misfortunes of Overton and his wife in order to illustrate the suppression of 'the just rights and freedoms of the Commons of England'.)

References and further reading

Burns, N.T. (1972) *Christian Mortalism from Tyndale to Milton*, Cambridge, MA: Harvard University Press. (Contains a chapter 'Overton, Milton and Hobbes'.)

Edwards, T. (1646) *Gangraena, or, A Catalogue and Discovery of many of the Errours, Heresies, Blasphemies and pernicious Practices of the Sectaries of this Time*; repr. with intro., The University of Exeter: The Rota, 1977.(A polemic against the Levellers and Overton's mortalism.)

Frank, J. (1955) *The Levellers. A History of the Writings of Three Seventeenth-Century Social Democrats: John Lilburne, Richard Overton, William Walwyn*, Cambridge, MA: Harvard University Press.(Discusses Overton's political pamphlets.)

Gimelfarb-Brack, M. (1979) *Liberté, égalité, fraternité, justice! La vie et l'oeuvre de Richard Overton, Niveleur*, Berne: Lang.(The only monograph devoted to Overton, with a detailed biographical account and a good bibliography.)

Holland, G. (1645) *The Prerogative of Man: Or, his Soules Immortality*, Oxford.(A critique of Overton's *Mans Mortalitie*, rejecting both Overton's interpretation of the Bible and his philosophical reasoning. According to Holland, the article of the resurrection 'proves nothing against the perpetuity of the Soule'.)

Macpherson, C.B. (1962) *The Political Theory of Possessive Individualism. Hobbes to Locke*, Oxford: Clarendon Press.(Chapter III.4 deals with Overton as a 'possessive individualist'.)

Pomponazzi, P. (1516) *De immortalitate animae*, trans. W.H. Hay, II as *On the Immortality of the Soul*, in E. Cassirer, P.O. Kristeller and J.H. Randall, Jr (eds) *The Renaissance Philosophy of Man*, Chicago, IL: University of Chicago Press, 1948.(A reliable English translation of Pomponanzzi's best known treatise, by

which Overton may have been influenced.)

Ross, A. (1645) *The Philosophicall Touch-Stone*, London.(Attempts to prove ‘the immortality of mans soule briefly, but sufficiently’. Overton’s arguments are described as ‘the weak Fortifications of a late Amsterdam Ingeneer, patronizing the Soules Mortality’.)

Watner, C. (1980) ‘"Come what, come will": Richard Overton, Libertarian Leveller’, *The Journal of Libertarian Studies* 4: 405-32.(On Overton’s pamphleteering career and the libertarian aspects of the Leveller movement.)

Wither, G. (1636) *The Nature of Man. A learned and useful Tract, written in Greek by Nemesius,... Englished by Geo. Wither*.(Nemesius, Bishop of Emesa, wrote *De natura hominis* in *c*.400. Wither’s introduction bears resemblances to Overton’s views.)

Woolnor, H. (1641) *The True Originall of the Soule*, London.(Argues that the soul is produced ‘neither by creation nor by propagation’, but by a ‘certain meane way between both’. The soul consists of ‘matter of a higher kind’.)

Wootton, D. (1991) ‘Leveller Democracy and the Puritan Revolution’, in *The Cambridge History of Political Thought 1450-1700*, ed. J.H. Burns and M. Goldie, Cambridge: Cambridge University Press, 412-42.(A general discussion of the Leveller movement, with references to Overton and to Edwards’ *Gangraena*.)

Owen, Gwilym Ellis Lane (1922-82)

G.E.L. Owen led the reorientation in ancient philosophy that began in the 1950s in Britain and North America. He approached the texts with a profound knowledge of classical scholarship, but also as an analytic philosopher, understanding them as conceptual investigations of live philosophical interest. Concerned primarily with the logic of argumentation, philosophy of science and metaphysics, he wrote influential articles on Parmenides, Plato and Aristotle. Equally important were his classes at Oxford (1953-66), at Harvard (1966-73) and finally at Cambridge, in which he constantly developed and tested his ideas and methods.

1 Logic and metaphysics in Plato and Parmenides

After an interruption for war service, G.E.L. Owen completed his BA at Corpus Christi College, Oxford, and stayed on for the new graduate degree of B.Phil., which he took in 1950, the second year it was awarded. His B.Phil. thesis, written under the supervision of Gilbert Ryle, was a bold and far-reaching study of logic and metaphysics in the 'later dialogues' (as under his influence these particular works came to be called) of Plato: *Parmenides*, *Theaetetus*, *Sophist*, *Statesman* and *Philebus*; in it he laid the groundwork for the series of remarkable graduate seminars on Plato that he later gave at Oxford, Harvard and Cambridge. It led to his first article, 'The Place of the *Timaeus* in Plato's Dialogues' (1953), in which he set himself the task of overturning the scholarly consensus in regarding the *Timaeus* as the culmination of Plato's work in metaphysics. Timaeus majestically portrays the construction of the physical world by a god looking to separated Platonic Forms as models ('paradigms') of the qualities and substances to be imposed upon the transient 'flux' of physical reality. This seemed to Owen to rest on metaphysical theses, familiar from such 'middle-period' dialogues as *Phaedo* and *Republic*, that the more sophisticated 'later' dialogues had overthrown. So it was rather in the logical explorations of the 'later' dialogues that the high point of Platonic metaphysics was reached. In Owen's view, philosophically well-motivated ambitions had led Plato in the middle dialogues to linguistic and conceptual abuses in postulating the 'being', 'unity' and 'stability' of Forms and describing the character of physical objects as multiplicities that 'become' or 'appear' but 'are' not anything. Disturbed by the resulting logical difficulties and paradoxes, he undertook in the 'later' dialogues a careful examination of the logic of these and related concepts. He discovered that kind of philosophical analysis to be inherently rewarding, and hoped thereby to replace the earlier theory of Forms with a conceptually more adequate version. This early work of Owen's was carried forward later in brilliant papers on the analysis of being and not-being in the *Sophist* and on the antinomies about unity developed at length in the *Parmenides*. His work on the logical and metaphysical analyses of the 'later' dialogues fundamentally altered, and deepened, the study of these works - even if his arguments did not win most scholars over to dating the *Timaeus* early, close to the *Republic*.

Owen was the first University Lecturer at Oxford, and subsequently the first Professor, appointed specifically for Ancient Philosophy (1953); later he went to Harvard to establish a graduate programme there in Classics and Philosophy. He did massive research for his year-long lecture-courses at Oxford, attended also by graduate students, on the Presocratics. Only his interpretations of Parmenides and Zeno of elea were ever published. For Owen, Parmenides was a 'philosophical pioneer of the first water', not the cosmologist in the tradition of Anaximander that conventional scholarship saw. He stood self-consciously aside from that tradition, using independent deductions to work out the characteristics that, by force of logic, reality must possess. He argued that anything that can be thought of or spoken about must 'be' or 'exist', and that what 'is not' or 'does not exist' is literally unthinkable. Since change, differentiation and plurality of any kind would logically require the being of what 'is not', reality must be a single, timeless, spatially continuous, internally undifferentiated, unchanging entity. Reality as it appears to us, and as the earlier cosmologists represented it, is essentially diverse and changing. But Parmenides' 'much-contested proof' reveals how seductive conventions of human thought are which, despite their logical appearance, rest on logically unacceptable foundations.

2 Dialectic, physics and metaphysics in Aristotle

More than half of Owen's published papers, and a similar proportion of his teaching, dealt with Aristotle. Here again, he was primarily interested in the logic of arguments. In a reversal of Werner Jaeger's then-dominant picture of Aristotle's philosophical development, Owen showed that Aristotle's early work in logic had led him

while still a member of Plato’s Academy to reject Plato’s theory of Forms. Only much later did independent study lead to a sort of ‘Platonism’ in his metaphysical thought. Central to Owen’s work on Aristotle was his account of the role of ‘dialectic’, presented in ‘*Tithenai ta phainomena*’ (1961): this title refers to the process of ‘setting down’ the phenomena needing to be sorted out in order to reach a general theory in some area of inquiry. As Owen noticed, even in a subject like physics the ‘phenomena’ for Aristotle are mostly not results of observation, but certain opinions *about* what is observed: matters of common or customary opinion, or ordinary linguistic usage, or the views of the ‘wise’ - respected philosophers or other experts. A ‘dialectical’ inquiry works through these ‘phenomena’ so as to discover and then resolve tensions among them by making appropriate distinctions: the ‘true’ theory is the one that can claim maximum support from the phenomena once so reconstructed. For Owen, such ‘dialectical’ reasoning was a paradigm of good philosophical method. He examined it in several contexts in Aristotle’s *Physics*, especially the theories of time and motion, and at several stages of Aristotle’s engagement with ‘first philosophy’, or metaphysics.

Owen’s students, and others who came under his influence in formative years, spearheaded a wide expansion of work on ancient philosophy. Their research includes logic, metaphysics and philosophy of science, but also ethics, political philosophy, philosophy of mind and other fields, and it extends into the post-Classical periods of Hellenistic, Roman and later ancient philosophy.

JOHN M. COOPER

List of works

Owen’s unpublished seminar notes and papers are deposited in the Classics Faculty Library, University of Cambridge.

Owen, G.E.L. (1986) *Logic, Science and Dialectic*, ed. M.C. Nussbaum, London: Duckworth.(A collection of all Owen’s published scholarly papers, with the exception of six book reviews and one commentary.)

References and further reading

Schofield, M. and Nussbaum, M.C. (1982) *Language and Logos*, Cambridge: Cambridge University Press. (Studies in ancient Greek philosophy by fifteen scholars, former students and others indebted to the stimulus of discussion with him, presented to Owen on his sixtieth birthday.)

Oxford Calculators

'Oxford Calculators' is a modern label for a group of thinkers at Oxford in the mid-fourteenth century, whose approach to problems was noticed in the immediately succeeding centuries because of their tendency to solve by 'calculations' all sorts of problems previously addressed by other methods. If for example the question was, what must a monk do to obey the precept of his abbot to pray night and day, a 'calculator' might immediately rephrase the question to ask whether there is a minimum time spent in prayer that would be sufficient to fulfil the abbot's precept, or a maximum time spent that would be insufficient to fulfil the precept. Or, if grace was supposed to be both what enables a Christian to act meritoriously and a reward for having so acted, then a calculator might ask whether the degree of grace correlated with a meritorious act occurs at the moment of the meritorious act, before the act when the decision to act is being made, or after the act when the reward of increased grace is given. If a body was hot at one end but cold at the other, then a calculator might ask not whether it is to be labelled hot or cold, but how *hot it is as a whole. Finally, if it was asked whether a heavy body acts as a whole or as the sum of its parts, then a calculator might take the case of a long thin rod falling through a tunnel pierced through the centre of the earth and attempt to calculate how the rod's velocity would decrease as parts of the rod passed the center of the cosmos, if it acted as the sum of its parts.*

*Of these four questions, the last two were asked by Richard Swineshead, a mid-fourteenth century fellow of Merton College, Oxford, whose Liber calculationum (Book of Calculations) led to his being given the name 'Calculator'. By association with Richard Swineshead, other Oxford masters including Thomas Bradwardine, Richard Kilvington, William Heytesbury, Roger Swineshead and John Dumbleton have been labelled the 'Oxford Calculators'. Their work contains a distinctive combination of logical and quantitative techniques, which results from the fact that it was often utilized in disputations on sophismata (*de sophismatibus*). This same group of thinkers, with emphasis on their mathematical rather than logical work, has been called the 'Merton School', because many but not all of the Calculators were associated with Merton College, Oxford. Besides calculatory works, the same authors wrote works in which calculatory techniques are not so prominent, including commentaries on Aristotle, mathematical compendia and commentaries on Peter Lombard's Sentences.*

1 Richard Swineshead's *Liber calculationum*

The central book of the Oxford calculatory tradition, the *Liber calculationum (Book of Calculations)* focuses on applications of mathematical techniques to natural philosophical problems (Murdoch and Sylla 1976). Of the sixteen treatises of the *Liber calculationum*, the first four deal with the intension and remission of forms or the increase and decrease of qualities such as heat or whiteness. First, the issue of the intension and remission of forms is raised in the abstract. Should the degree of heat, for instance, be measured by nearness to the maximum degree of heat or by the distance from zero degree (Clagett 1950)? Second, questions concerning the intension and remission of forms are raised concerning a body nonuniformly qualified with heat or some other quality. Should it be said to be as hot as its maximum degree or as hot as its mean degree? Third, how should one characterize the combined degree of an element having two unequal qualities, such as heat and moisture: is the combined degree as intense as the mean between its degree of heat and its degree of moisture or is it, for example, only as intense as the more remiss quality? Swineshead concludes that such an element is as intense as the geometric mean between its two qualities, so that if one quality is in the second degree and the other in the eighth, then the element as a whole should be said to be in the fourth degree.

Having debated and determined such questions about the measures of qualities, the *Liber calculationum* turns next to measurements of quantities of bodies, first to degrees of density or rarity and then to measures of velocity of augmentation and diminution. Treatise VII deals with the problem of reaction. If a hot body heats the cold body next to it, will the cold body simultaneously react and cool the hot body? If, for action to occur, the agent must have more power than what it acts on, how can each of the two neighbouring bodies have sufficient power to act on the other at the same time? Treatise VIII consequently asks directly how the power of an agent should be measured. Is it, for instance, by its intensity, so that a body hot in the fourth degree would have a power in the fourth degree? Or does power depend on the 'multitude of form', taken to depend not only upon the intensity and extension of the form, but also upon the density of the matter in which it inheres? Also, how is the 'difficulty of action' to be measured? Is it by the ratio between the power of the agent and the resistance or is it, as Swineshead

concludes, simply by the power of the agent acting to its utmost?

Treatise X deals with maxima and minima and Treatise XI with the problem, mentioned at the start of this article, of a thin rod moving through the centre of the earth (Hoskin and Molland 1966). Treatises XII and XIII deal with illumination and the action of light sources. Treatises XIV and XV deal with problems of quantifying local motion with respect to its causes, first explaining in the abstract how velocity varies when the ratio of force to resistance varies, assuming the truth of Thomas Bradwardine's dynamical function (see Bradwardine, T.), and second in the concrete, assuming that resistance is encountered by an agent as it moves through a medium of varying resistance. Finally, Treatise XVI of the *Liber calculationum* deals with the rates at which a body takes on the maximum degree of a quality; as, for instance, a heating agent warms a body next to it, first assimilating the parts of the body nearest to the agent and then parts farther away.

As can be seen from this brief survey of the contents of the *Liber calculationum*, the Oxford 'calculations' in their core manifestation concerned every technique that had been developed up to that time for introducing quantification into natural philosophy, plus some new ones (see Natural philosophy, medieval). The intension and remission of forms was a type of quantification previously most highly developed in determining the degrees of compound medicines. Measurements of illumination had been developed within the science of optics or perspective and possibly also in astrology, where each planet was supposed to exert an influence on earth depending upon its position in the heavens. Ways of relating forces, resistances and velocities of local motion had received a thorough treatment in Thomas Bradwardine's *De proportionibus velocitatum in motibus (On the Ratios of Velocities in Motion)*, completed in 1328. The 'calculator' focused his attention only on physical questions, but the theory of intension and remission of forms had been and continued to be applied by others to the intension and remission of charity or grace in the human soul, and calculations could also be applied in theology.

2 Calculations and disputations on sophismata

What gave 'calculations' their distinctive style was the fact that they were connected not only to natural philosophy or theology, but also to the disputations on sophismata (*de sophismatibus*) prominent at Oxford in this period (Sylla 1982). As exercises occurring at the culmination of the undergraduate curriculum, disputations on sophismata were devoted to the explication of propositions that are at first glance counterintuitive, such as 'Infinite are finite' (*Infinita sunt finita*) or 'Socrates is whiter than Plato begins to be white'. In an oral disputation on a sophisma, students or masters competed against each other in proposing and arguing for and against possible resolutions. In an earlier period, disputations *de sophismatibus* called upon the disputants to use especially what they knew of logical or grammatical theory. As time went on, natural philosophy and mathematics were also called upon, thus resulting in the so-called calculations.

What disputations *de sophismatibus* were like in earlier fourteenth-century Oxford, before calculations became prominent, can be seen in Richard Kilvington's *Sophismata*, which includes the sophisma mentioned above, 'Socrates is whiter than Plato begins to be white' (Kretzmann and Kretzmann 1990) (see Kilvington, R.). Assuming the case that Socrates has some degree of whiteness and that Plato begins to intensify in whiteness starting from zero, the sophisma sentence seems to apply, but not without problems. Suppose that the comparative term 'whiter' is taken to be measured by the ratio of the more intense quality to the less intense, so that if Socrates is white in the second degree and Plato in the first degree, then Socrates will be said to be twice as white as Plato. If Plato has zero whiteness at t0 and subsequently begins to intensify in whiteness, then when one considers the ratio between the whiteness of Socrates and that of Plato, it will get larger towards the beginning of Plato's intensification, tending to infinity as Plato's whiteness tends to zero. It seems to follow that immediately after Plato begins to intensify in whiteness, Socrates will be infinitely whiter than Plato, but this seems to be contradictory, it might be argued, because Socrates is only finitely white. Is something wrong with the analysis? Confronted with such sophismata, Oxford scholars concluded that it might be useful to study systematically how any and all things should be quantified.

William Heytesbury's *Sophismata* and *Regulae solvendi sophismata (Rules for Solving Sophismata)* exemplify the forms that Oxford disputations on sophismata took after the introduction of calculations (Wilson 1956). Heytesbury divides his rules for solving sophismata according to set techniques, referred to by historians as 'analytical languages' or 'conceptual algorithms' (Murdoch 1969, 1975). There are purely logical techniques, such as ways of solving 'insolubilia', or self-falsifying propositions, such as 'I am telling a lie'. Other techniques

analyze sophismata in terms of first and last instants, maxima and minima, and measures of motion in the categories of local motion, alteration, and augmentation and diminution. The same sophisma sentence could be analysed on different occasions applied to different hypothetical situations (‘cases’) and using different analytical languages. Thus the sophisma sentence ‘Socrates is whiter than Plato begins to be white’ might be analysed by considering that motion of alteration is a so-called successive entity extrinsically limited at its beginning and its end; there is no first instant of alteration but only a last instant before the alteration begins, and no last instant of alteration but only the first instant at which the final degree has been induced.

On another occasion the same sophisma sentence could be analyzed using the language of maxima and minima, pointing out that as Plato moves from zero whiteness the intension of whiteness increases continuously, so that there is no minimum degree of whiteness gained, but rather smaller and smaller degrees *ad infinitum* down to zero. When the implication was drawn from the case that if relative intensities are measured by the ratio they have to each other, then Socrates will be infinitely whiter than Plato immediately after he begins to be white, then the sophisma sentence could be analyzed using distinctions about senses of the words ‘immediate’ and ‘infinite’. Sometimes ‘infinite’ is used to mean a ‘potential infinite’, and sometimes an ‘actual infinite’. Thus the integers are potentially infinite because, given any integer, one can find a higher integer; but the integers are not actually infinite because there is no single infinite number. Heytesbury’s *Sophismata* demonstrates in practice the many rules that he lays out in the *Regulae solvendi sophismata*.

3 Calculations and measures of motion

The Oxford Calculators’ belief in the potential power of mathematics to solve problems of sophismata seems to have been inspired by the success of Thomas Bradwardine’s *De proportionibus velocitatum in motibus* (Crosby 1955) (see Bradwardine, T.). Bradwardine had shown how neatly the Euclidian theory of ratios could be used to clarify the relations of forces, resistances and velocities in motions as discussed with less success in Aristotle’s *Physics*. Bradwardine’s function, relating forces, resistances and velocities, was said to measure motion with respect to its cause (*tanquam penes causam*), while measures of the distances traversed or intensities or sizes gained were said to be measures with respect to effect (*tanquam penes effectum*). Discussions of alternative ways of measuring local motion, alteration, and augmentation and diminution *tanquam penes effectum* are common in the works of the Oxford Calculators.

4 Calculations and logic

These mathematical considerations were introduced into a context that had already been highly developed through the application of the new logic, especially the theories of supposition and syncategoremata (see Language, medieval theories of; Logic, medieval). The problematic meaning of a sophisma could be shown to result from a confusion of compounded and divided senses, as in the sentence ‘Something white can be black’. The word ‘infinite’ was said to have a syncategorematic sense, corresponding to the potential infinite, as well as a categorematic sense, corresponding to the actual infinite. The sophisma sentence ‘Infinite are finite’ was explained by saying that when ‘infinite’ appears at the beginning of a sentence, it is to be taken syncategorematically. There is not one thing that is infinite, but many things which together are infinite. In this way the calculators could conclude that there is no one infinite number, but all the numbers together are infinite, meaning something like the modern statement that the set of all integers is infinite.

5 *De incipit et desinit*; maxima and minima

Walter Burley, in his work on first and last instants and in his works on intension and remission of forms, had developed techniques that were also used as analytical languages by the later Calculators. As described above, the technique of first and last instants or of analyzing the use of the verbs ‘it begins’ and ‘it ceases’ (*incipit et desinit*), distinguished whether a permanent form (such as whiteness) or a successive entity (such as the motion of traversing a distance) has a first instant of being or a last instant of non-being. Using the analytical language of maxima and minima, it might be asked whether there is a maximum weight that Socrates can carry or a minimum weight too heavy for him. Again, propositions or sophismata would be analyzed using supposition theory to see which alternative was the correct one.

6 Calculations and natural philosophy

After their development, the techniques of the Oxford Calculators were also applied not only to solving explicitly labelled sophismata, but also to dealing with the normal questions of natural philosophy (see Natural philosophy, medieval). Roger Swineshead, author of the *Obligationes* and *Insolubilia* (and thought to be a distinct person from Richard Swineshead, despite a confusion of names in the manuscripts), made such an application in his *De motibus naturalibus (On Natural Motions)*, alternating sections dealing with motion in the categories of place, quality and quantity with sections explicitly devoted to measures of motion, often proposing sophismata that could be developed using these measures. Thus Roger Swineshead proposed that a body which gains a degree uniformly throughout is altered indivisibly more than a body which in the end is non-uniformly qualified with the given degree as the minimum uniform degree that does not qualify it.

Calculatory techniques were also used by John Dumbleton, author of a large *Summa logicae et philosophiae naturalis (Summa of Logic and Natural Philosophy)* which combined calculation with something resembling a commentary on the natural works of Aristotle. Thus the techniques of the Oxford Calculators were seen as useful in the ordinary business of natural philosophy, although they gained their special flavor by frequent use in disputations *de sophismatibus*.

7 Calculations and theology

Shortly after they were put to use in the disputations *de sophismatibus*, the techniques of the Oxford Calculators became very popular at Paris and elsewhere, often in combination with the nominalistic analyses associated with William of Ockham. Bachelors of theology commenting on Peter Lombard's *Sentences* frequently used calculatory techniques to address theological problems. In many cases, calculatory techniques indeed helped to resolve technical questions such as those concerned with degrees of grace or with first and last instants (for example, whether there is a first instant of salvation or a last instant of sinfulness). While some commentators applied the calculatory techniques wherever possible, others deplored the invasion of such quibbling technicalities into sacred subjects (Murdoch 1978).

See also: Natural philosophy, medieval

EDITH DUDLEY SYLLA

References and further reading

Clagett, M. (1950) 'Richard Swineshead and Late Medieval Physics', *Osiris* 9: 131-61.(An analysis especially of the first treatise of the *Liber calculationum*.)

Clagett, M. (1959) *The Science of Mechanics in the Middle Ages*, Madison WI: University of Wisconsin Press. (Text, translations and commentaries. Provides the context, within medieval mechanics, of the work of the Oxford Calculators.)

Crosby, H.L. (ed. and trans.) (1955) *Thomas of Bradwardine: His Tractatus de proportionibus*, Madison, WI: University of Wisconsin Press.(The treatise that most inspired the mathematical side of the Calculators' work.)

Gilbert, N. (1976) 'Richard de Bury and the "Quires of Yesterday's Sophisms"' in E.P. Mahoney (ed.) *Philosophy and Humanism: Renaissance Essays in Honor of Paul Oskar Kristeller*, New York: Columbia University Press. (Describes the impact of the Oxford calculations at Paris and some of the negative reaction engendered there.)

Hoskin, M. and Molland, A.G. (1966) 'Swineshead on Falling Bodies: An Example of Fourteenth Century Physics', *British Journal for the History of Science* 3: 150-82. (Edition of Treatise XI of Swineshead's *Calculationes* dealing with motion of a rod through the centre of the earth.)

Kretzmann, N. and Kretzmann, B. (1990) *The Sophismata of Richard Kilvington*, Oxford: Oxford University Press and Cambridge: Cambridge University Press.(Best place to start for an understanding of how sophismata were solved in practice.)

Lewis, C. (1980) *The Merton Tradition and Kinematics in Late Sixteenth and Early Seventeenth Century Italy*, Padua: Antenore.(Investigates the extent of the Calculators' influence in Italy.)

Longeway, J. (1984) *William Heytesbury On Maxima and Minima: Chapter 5 of Rules for Solving Sophismata, with an anonymous fourteenth-century discussion*, Dordrecht: Reidel.(Contains an English translation of Heytesbury on maxima and minima, plus the Latin text of the anonymous discussion.)

Maier, A. (1952) *An der Grenze von Scholastik und Naturwissenschaft (On the Border between Scholasticism and Natural Science)*, 2nd edn, Rome: Edizioni di Storia e Letteratura.(See III.1. 'Die Calculationes des 14. Jahrhunderts'.)

Maier, A. (1966) *Die Vorläufer Galileis im 14. Jahrhundert (The Precursors of Galileo in the Fourteenth*

Century), 2nd edn, Rome: Edizioni di Storia e Letteratura.(See especially, 'Der Funktionsbegriff in der Physik des 14. Jahrhundert' (The Concept of Function in the Physics of the Fourteenth Century), 81-110.)

Molland, A.G. (1968) 'The Geometrical Background to the "Merton School"', *British Journal for the History of Science* 4: 108-25.(Explains background of Bradwardine's function.)

Murdoch, J.E. (1969) 'Mathesis in Philosophiam Scholasticam Introducta: The Rise and Development of the Application of Mathematics in Fourteenth-Century Philosophy and Theology', in *Actes du Quatrième Congrès International de Philosophie Médiévale*, Montreal, Que.: Institut d'Études Médiévales, 215-54.(How quantification entered medieval natural philosophy.)

Murdoch, J.E. (1975) 'From Social into Intellectual Factors: An Aspect of the Unitary Character of Medieval Learning', in J.E. Murdoch and E.D. Sylla (eds) *The Cultural Context of Medieval Learning*, Dordrecht: Reidel, 271-339.(Describes in some detail the conceptual algorithms or analytical languages used throughout much of fourteenth-century philosophy and theology.)

Murdoch, J.E. (1978) *Subtilitates Anglicanae in Fourteenth-Century Paris: John Mirecourt and Peter Ceffons*, New York: New York Academy of Sciences.(Describes the vogue for Oxford calculatory ideas in Parisian theology.)

Murdoch, J.E. and Sylla, E.D. (1976) 'Swineshead (Swyneshed, Suicet, etc.), Richard', in C.C. Gillispie (ed.) *Dictionary of Scientific Biography*, vol. 13, New York: Charles Scribner's Sons, 184-213.(Extended description of the *Book of Calculations*, treatise by treatise.)

Sylla, E. (1971) 'Medieval Quantifications of Qualities: The "Merton School"', *Archive for the History of the Exact Sciences* 8: 9-39.(Describes the ways that the Calculators quantified qualities as contrasted with Nicole Oresme.)

Sylla, E. (1973) 'Medieval Concepts of the Latitude of Forms: The "Oxford Calculators"', *Archives d'histoire doctrinale et littéraire du moyen age* 40: 223-83.(The work of the Calculators as related to latitudes, degrees of quality and intension of forms.)

Sylla, E. (1982) 'The Oxford Calculators', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 540-63.(Addresses especially the connection of the work of the calculators to Oxford undergraduate disputations on sophismata.) -

Sylla, E. (1986) 'Galileo and the Oxford *Calculatores*: Analytical Languages and the Mean-Speed Theorem for Accelerated Motion', in W.A. Wallace (ed.) *Reinterpreting Galileo*, Washington, DC: Catholic University of America Press, 53-108.(Addresses the question of whether Galileo used the work of the Oxford Calculators, and especially the Merton mean-speed theorem, in his kinematics.)

Sylla, E. (1987) 'The Oxford Calculators in Context', *Science in Context* 1: 257-79. (How the work of the Calculators was related to the academic milieu in which it was carried out.)

Sylla, E. (1991) *The Oxford Calculators and the Mathematics of Motion, 1320-1350*, New York: Garland.(Ph.D. dissertation, Harvard University, 1970; contains chapters on the physics of motion in the categories of quality and place, on measurement by proportionalities and latitudes, and on the wider calculatory tradition, as well as extensive outlines in Latin of several key calculatory works.)

Weisheipl, J.A. (1959) 'The Place of John Dumbleton in the Merton School', *Isis* 50: 439-54.(How Dumbleton's *Summa* relates to the work of the other Calculators.)

Weisheipl, J.A. (1968) 'Ockham and Some Mertonians', *Mediaeval Studies* 30: 163-213. (Overview of the work of the Calculators in comparison to Ockham.)

Weisheipl, J.A. (1969) 'Repertorium Mertonense', *Mediaeval Studies* 331: 174-224.(Most thorough available catalogue of works, with manuscript locations, of the Oxford Calculators.)

Wilson, C. (1956) *William Heytesbury: Medieval Logic and the Rise of Mathematical Physics*, Madison, WI: University of Wisconsin Press.(Decribes the contents of Heytesbury's *Rules for Solving Sophismata* and related works.)

Paine, Thomas (1737-1809)

Thomas Paine, born in Norfolk, England, spent his early years as an undistinguished artisan and later excise officer. In 1774 he emigrated to America and settled in Philadelphia where he became a journalist and essayist. His Common Sense (1776) and sixteen essays on The Crisis (1776-83) were stunning examples of political propaganda and theorizing. In the late 1780s, in Europe, Paine wrote The Rights of Man (1791-2) and attacked the English political system. During the French Revolution he was a Girondin in the French Convention and wrote The Age of Reason (1794, 1796), savagely criticizing Christianity. He died in New York in 1809, an important figure in the sweep of the revolutionary politics in America, England, and France at the end of the eighteenth century.

1 Background and republican political theorizing

Thomas Paine was born on 29 January 1737 in Thetford, in Norfolk, England, where he spent the first thirty-seven years of his life, a poor, unsuccessful, and unknown artisan. In desperation he emigrated to America in 1774 and rapidly received recognition as an intellectual leader of the American Revolution.

The juxtaposition of Paine's early career and later fame has attracted wide attention. In periods of conservative stability Paine has been reviled as an extreme radical on the fringes of the American Revolution, a demagogic drunkard not fit to be associated with the Founding Fathers of the United States. In periods of ferment commentators have hailed Paine as an authentic democrat of the people, whose thought and practice can be a model for the left.

Paine's natural talents and the more open social structure of America obviously contributed to the astonishing improvement in his fortunes by the mid-1770s. But the colonial uprising was mainly middle-class in its orientation, led by the political gentry. In this social milieu Paine was something of an outsider. His background and personality made him unpredictable and never fully acceptable. He himself was never at ease in the upper reaches of the society he was helping to create, although he gloried in the recognition and status he had achieved.

Paine's success was to some extent a product of his marginal status. He had a shrewd understanding of the actual political world and was zealously committed to government by and for the masses. But he was no more an erudite thinker than the other American founders, nor were his political notions unconventional in their context. None the less, he had the rare gift - always presumed to be the produce of his working life - of being able to write for the populace. Paine understood what the people wanted or needed to hear better than any of the other Founding Fathers. All of his tracts from the seventies to the nineties hit the nerve of popular debate while simultaneously advancing the democratic ideas that the Americans had taken from John Locke and the English 'Commonwealthmen'.

Paine was interested in applied science and absorbed popular expressions of Newtonian physics (see Newton, I.). A series of laws governed the natural world. They were open to the understanding of all and could be put to work for human benefit. Paine was a 'rationalist' in an era termed the American Enlightenment, believing that fundamental truths could be arrived at by careful investigation and reflection on experience. Thus, he rejected Christianity, as depending on unreasonable beliefs (for example, the Virgin birth), and delighted in pointing out inconsistencies in the Bible. Instead, the harmonies of the natural world led one to conclude that a benevolent Deity had created an ordered cosmos. This 'clockwork' God did not, however, intervene directly in human life but had imprinted humankind with moral impulses that enabled people to achieve happiness on earth once they had rationally considered their own best interests.

This set of metaphysical and moral beliefs led to a grasp of political life that was indebted to Locke and crucial to Paine's philosophical politics. The creator allocated 'natural rights' to all people - or at least all men - equally. Differences in station were justified only because of variations in talent, industry, and frugality. In the order of things men should be provided with equal opportunities and an equal voice in determining how society should be governed. Society was the natural outcome of the human condition - the relationships that derived from human beings living together, raising their young, and co-operating in common endeavour. These joint implicit responsibilities embodied the social contract. On the other side, government was a negative instrument only

necessary to the extent that matters went awry; its powers were limited and, in the natural course of things, largely unnecessary.

What had gone wrong in America to impel the nation's leaders to revolution? Paine and others like him embraced the views of some dissenters in England earlier in the eighteenth century. These Commonwealthmen had developed a critique of English society that supposed the monetary power of the Crown to have destroyed balanced government. The Commonwealthmen looked back to a time of imagined civic virtue and forward to a restoration of the natural scheme of governance.

At the end of the eighteenth century, Paine popularized and extended this view in America, leavening it with Lockean individualism. This was the 'republicanism' of the colonies, which combined a constellation of historical, political, and philosophical ideas. The opponents of the Commonwealthmen, Paine said, had triumphed in England, corrupted its society, and wished to place a yoke of tyranny on the colonies. Paine, in contrast, urged the liberty of individual conscience, religious toleration, resistance to tyrants, and reform of the legislature. He denounced hereditary political institutions and desired to lodge power with 'the people', no matter how restrictedly they were defined.

2 Propaganda against England

Paine elaborated these views in the 1770s as the most inflammatory and didactic exhorter of the Revolutionary forces. *Common Sense* (1776) presented a justification for the Revolution, and the first *Crisis* paper (1776) is sometimes credited with keeping the Revolutionary army intact during the early fighting. In the 1780s Paine joined the Federalists in supporting a new constitution. He believed that a powerful central authority would promote the commercial and territorial expansion he desired, and he saw the Federalists as safeguarding *laissez-faire* economics in a democratic empire.

When Paine returned to Europe in the late 1780s, however, he focused again on theorizing about revolution, especially the one occurring in France and its potential for human liberation. At the same time he hoped to change the contours of English politics. Although first welcomed by men like Edmund Burke, who had sympathized with the American colonies, Paine became disgusted with the possibilities for renewal of the English system. *The Rights of Man* (1791-2) explicitly attacked Burke's ideas and defended the French Revolution. Paine was contemptuous of the claims of tradition and custom that led to commitment to a hereditary monarchy. The book's praise of French politics also signalled a break with the increasingly conservative American Federalists who came to fear the radicalism of the Continent.

The Rights of Man additionally suggested ways in which a republic could effect the rudiments of a welfare state, and hinted at the potential for redistributive economic justice. This concern was later fleshed out in *Agrarian Justice* (1797). But even in 1792 Paine had gone too far for English authorities and was indicted for sedition. He fled to Paris in September. The events of the 1790s and his writings in that decade have ever after made him a hero to the English left.

3 Role in the French Revolution

In France Paine combined the political theorizing and revolutionary action that were characteristic of his career. But when the Girondins lost control of the Convention and the French Terror began, Paine was endangered. Isolated by his ignorance of the language, he refused to alter his now conservative ideas and was taken to the prison where he assumed he would be put to death. In 1794, however, he was released and lived quietly with the American Ambassador to France, James Monroe.

In *The Age of Reason* (1794), written during this time in France, Paine returned to his religious concerns. He articulated the premises of Enlightenment Deism (see Deism) but also bitterly assaulted Christianity and its biblical basis. Too conservative in France, his religious views proved too hostile to established institutions when he returned to America in 1802. He was vilified in many quarters as a representative of the Anti-Christ and lived out his last years in penury and public disgrace.

See also: American philosophy in the 18th and 19th centuries; Franklin, B.; Jefferson, T.

BRUCE KUKLICK

List of works

Paine, T. (1947) *Complete Writings of Thomas Paine*, ed. P.S. Foner, New York: Library of America, 2 vols. (*Four Letters on Interesting Subjects* (1776) and *Old Truths and Established Facts: Being an Answer to a Very New Pamphlet Indeed* (1792), both probably written by Paine, are not included.)

References and further reading

Ayer, A.J. (1988) *Thomas Paine*, Chicago, IL: University of Chicago Press.(Notable because the author was a famous twentieth-century philosophical and humanist writer.)

Claeys, G. (1989) *Thomas Paine: Social and Political Thought*, London: Routledge.(Concentrating on *The Rights of Man*, this study places the tract in the context of Paine's earlier writings and evolving thoughts.)

Fast, H. (1943) *Citizen Tom Paine*, New York: Grove Atlantic.(A novel with an excellent psychological characterization.)

Foner, E. (1976) *Tom Paine and Revolutionary America*, Oxford: Oxford University Press.(Excellent study.)

Paley, William (1743-1805)

William Paley, theologian and moral philosopher, expressed and codified the views and arguments of orthodox Christianity and the conservative moral and political thought of eighteenth-century England. Paley says that his works form a unified system based on natural religion. Like others during this period, Paley thought that reason alone, unaided by revelation, would establish many Christian theses. He is confident that a scientific understanding of nature will support the claim that God is the author of nature. Paley belongs to the anti-deist tradition that holds that revelation supplements natural religion. The most important revelation is God's assurance of an afterlife in which the virtuous are rewarded and the vicious are punished. Natural and revealed religion, in turn, provide the foundation for morality. God's will determines what is right and his power to reward and punish us in the afterlife provide the moral sanctions. On the whole, Paley is concerned with sustaining Christian faith, and ensuring that people known what their duties are and do them.

1 Life and writings

Paley was born in Peterborough, England. He entered Christ's College Cambridge in 1759 and graduated in 1763 with highest distinction. After teaching for a few years in a small school, he returned to Cambridge where he taught moral philosophy, divinity and Greek testament for nine years. His lectures were noted for their clarity and lucid organization. He left Cambridge when he married, obtaining the first of various positions in the Church of England, the highest as archdeacon of Carlisle.

Paley wrote four books. In the first, *The Principles of Moral and Political Philosophy* (1785), he defends theological utilitarianism. In his *Horae Paulinae: Or The Truth of the Scripture History of St Paul Evinced* (1790), he defends the authenticity of the New Testament writings, in particular the scriptural history of St Paul. In *A View of the Evidences of Christianity* (1794), he defends revelation and argues for the veracity of Christian miracles. In *Natural Theology* (1802), his best known work today, Paley develops a detailed statement of the argument from design. Except for his *Horae Pauline*, his writings enjoyed a huge success and went through numerous reprintings. His *Principles* was adopted as a textbook at Cambridge in 1786 and his *Evidences* in 1822, and both were used as textbooks in England and the United States until well into the mid-1800s. His portrait hangs in Christ's College along with those of Milton and Darwin. His works, however, have by and large faded into obscurity.

2 Theological and apologetical writings

Paley addresses his *Natural Theology* to Christian readers with the hope that it may help to sustain their belief in God in times of doubt and that they may come to see the world as a 'temple' and life as 'an act of adoration'. The argument of the *Natural Theology* has two parts. In both parts he defends the argument from design against some traditional objections (see God, arguments for the existence of). First, Paley wants to establish that there are strong rational grounds for belief in the existence of God. He argues that the natural order, like a watch, was clearly designed for a purpose. As in the case of the watch, we should infer that the universe had an intelligent creator. Because each example alone is sufficient to support the conclusion that there is an intelligent author of nature, Paley says that by enumerating examples he is adding to the proof. He offers many detailed examples to support the claim that the various parts of nature are like machines exhibiting design and purpose. Most of his examples are from anatomy, and his favourite is the eye. Its intricate structure parallels that of a telescope and both obviously serve a purpose. Paley considers but dismisses a number of alternative explanations for the purposefulness of the natural world. For example, he argues that nature's intricate structure is so striking that it could not be a product of the chance coming together of atoms.

Second, Paley argues that a scientific understanding of nature helps to establish the traditional attributes of a Christian God. God is a person, since, as a contriver, he must have a mind. The fact that nature exhibits a uniform plan suggests his unity, although Paley acknowledges that this only shows a 'unity of counsel'. As cause of nature, God's power and intelligence must be adequate to their effect, yet the terms 'omniscient' and 'omnipotent', he admits, are only convenient superlatives. Paley devotes considerable effort to attempting to establish God's benevolence, arguing that the contrivances of nature are beneficial and that it is 'a happy world after all'. In addition, God gave us and other animals the capacity to experience pleasure, even though, he assumed, pleasure is biologically unnecessary. But Paley concedes that natural religion cannot by itself solve the problem of evil (see

Evil, problem of).

Paley addresses the credibility of christian revelation in his *Evidences*. He assumes that God chose to reveal his will to the early Christians by performing miracles. The credibility of revelation thus turns on whether there are good reasons for believing that these miracles occurred. But David Hume (1777) had argued that it is contrary to experience that miracles should be true, but consistent with experience that testimony should be false (see Hume, D. §2; Miracles). So there are two issues: should we believe in the occurrence of miracles, and is the testimony of the witnesses credible? Paley takes up these issues one by one. What tells against miracles, Hume thought, is the fact that the laws of nature work uniformly. Paley counters that if we have independent proof that there is a God who created the universe - including the laws of nature - for special purposes, it is not improbable that God may have occasionally intervened in the natural order. In response to Hume's charge that the witnesses to miracles are not credible, Paley replies that the suffering of the early converts demonstrates their sincerity. Both Christian and non-Christian sources attest to their ordeals. Moreover, the New Testament writings are authentic, neither forged nor altered during their transmission. Further support for their authenticity is given in *Horae Paulinae*. The many undesigned coincidences between the Acts of the Apostles and Paul's letters, Paley argues, show that neither is fraudulent. In the *Evidences* Paley also responds to Hume's complaint that the miracle stories of various religions undermine each other's credibility, arguing that those of other religions may be dismissed on a number of grounds, including delusion and exaggeration.

3 Ethical and political writings

Paley insists that natural and revealed religion are the true foundation for morality, chastising those moralists who decline to bring in scriptural authorities. He also complains that most earlier theorists did not pay sufficient attention to working out what in specific instances our duties are, and getting people to do them. The key elements of Paley's theory are summed up in his three-part definition of virtue: 'the doing of good to mankind, in obedience to the will of God, and for the sake of everlasting happiness' (1785: 47). Paley is a theological voluntarist: God's will ultimately determines what is right (see Voluntarism). Since God is benevolent and wills our happiness, the rule we should use in deciding what to do is the utility principle. Right actions are those which promote the general happiness. Although we are basically self-interested creatures, God's power to reward and punish us in an afterlife make it in our interest to promote the happiness of others. Paley thinks that moral obligation is an unproblematic notion. It is like the obligation of a soldier to obey his officer or a servant his master, namely, being 'urged by a violent motive resulting from the command of another' (1785: 66). As Paley acknowledges, the only difference between prudential and moral obligation on this view is that the sanctions attaching to prudence are confined to this world whereas moral sanctions look to the 'world to come' (1785: 71).

Most of the *Principles* consists in a detailed discussion of our duties to God, others and ourselves. The last section contains his political theory. In most cases Paley defends the status quo. He does not use the utility principle as a tool of reform as did the later utilitarians. He argues that it is consistent with God's will that people own property and that the positive law of each country should regulate its distribution. He also argues that inequalities in wealth are advantageous to society, increasing productivity, and hence the general happiness. Like other utilitarians, Paley objects to the social contract theory. He traces the origin of government to paternal authority and the need for strong leaders in times of war. Citizens continue to obey out of habit and self-interest. The reason people ought to obey government is that God wills the general happiness and civil society promotes that end. In his discussion of more topical issues, Paley, as might be expected, defends the status quo; even so, throughout his adult life he vehemently opposed the slave trade and actively sought to end it.

See also: Deism; Providence; Utilitarianism

CHARLOTTE R. BROWN

List of works

There is no modern edition of Paley's works, but his works are readily available in many libraries. Selections from his *Natural Theology* are to be found in numerous introductory philosophy textbooks and in anthologies of classical readings in philosophy of religion.

Paley, W. (1785) *The Principles of Moral and Political Philosophy*, Dublin: Exshaw, White.(Based on lectures Paley gave while teaching at Cambridge, its success is said to have moved Bentham to publish his *Introduction*

to the Principles of Morals and Legislation. The opening sections contain his objections to the moral sense theory and a defence of theological utilitarianism. The rest of the book consists in a detailed discussion of our particular duties to God, others and ourselves, as well as Paley's political theory.)

Paley, W. (1790) *Horae Paulinae: Or The Truth of the Scripture History of St Paul, Evinced by a Comparison of the Epistles which bear his Name with the Acts of the Apostles, and with one another*, London and Dublin.(This is a work in Christian apologetics and is the least philosophically interesting of Paley's writings. He defends the authenticity of both the Acts of the Apostles and Paul's thirteen letters.)

Paley, W. (1794) *A View of the Evidences of Christianity. In three parts*, London and Dublin.(A work in Christian apologetics, Paley defends the credibility of Christian miracles and so Christian revelation against Hume's objections. He thinks this task is important because revelation is our only assurance of an afterlife in which the good are rewarded and the evil are punished.)

Paley, W. (1802) *Natural Theology: Or, Evidences of the Existence and Attributes of the Deity, collected from the Appearances of Nature*, London: Faulder and Philadelphia, PA: John Morgan.(The most widely read of Paley's works, it contains a detailed restatement of the argument from design. Although Hume's *Dialogues on Natural Religion* (1779) consists in a sustained and devastating attack on the argument from design, Paley does not reply to these objections.)

References and further reading

Barker, E. (1948) 'Paley and His Political Philosophy', in *Traditions of Civility*, Cambridge: Cambridge University Press.(A sympathetic discussion of Paley's political theory.)

Clark, M. (1974) *Paley*, Toronto, Ont.: University of Toronto Press.(A general introduction to Paley's thought with a detailed biography.)

Hume, D. (1777) 'Of Miracles', in *An Enquiry Concerning Human Understanding*, ed. L.A. Selby-Bigge, revised P.H. Nidditch, Oxford: Clarendon Press, 1988.(Argues that the power of testimony to establish the credibility of miracles and thereby to establish the authority of a religious system.)

LeMahieu, D. (1976) *The Mind of William Paley*, Lincoln, NE: University of Nebraska Press.(A general, thematic study of Paley's theology and philosophy.)

Raphael, D. (1991) *British Moralists: 1650-1800*, Indianapolis, IN: Hackett Publishing Company, vol. 2.(Contains selections of the works of many of the British Moralists of the seventeenth and eighteenth centuries. Includes a short selection of some of the theoretically more interesting parts of Paley's moral theory.)

Schneewind, J. (1977) 'The Early Utilitarians', in *Sidgwick's Ethics and Victorian Moral Philosophy*, Oxford: Clarendon Press.(A clear summary of Paley's utilitarianism which places it in the context of other utilitarian writers of the period.)

Schneewind, J. (1990) *Moral Philosophy from Montaigne to Kant*, Cambridge: Cambridge University Press, vol. 2.(Contains selections of the works of many seventeenth- and eighteenth-century moral philosophers from both sides of the British Channel. In addition to an informative general introduction, each individual selection has a separate introduction, annotation and bibliography. Contains a slightly more comprehensive selection of the theoretically interesting parts of Paley's moral theory, including his objections to the moral sense theory and his account of rights.)

Panaetius (*c*.185-*c*.110 BC)

Panaetius, a Greek philosopher from Rhodes, brought new vitality to Stoicism in the second century BC by shifting the focus of its ethical theory from the idealized sage to the practical problems of ordinary people. Working a century after Chrysippus had systematized Stoicism, Panaetius is often labelled the founder of 'Middle Stoicism' for defending new and generally more moderate positions on several issues. Because none of his writings survive, his influence is hard to gauge precisely and easily underrated. But his impact, especially in Rome where he was closely associated with many in the ruling elite, was profound. His emphasis on public service and the obligations imposed by power and high station probably helped shape the ideology of Roman imperialism. Through Cicero, whose writings preserve many of his ideas, he had lasting influence, especially on early modern moral and political thought.

1 Life and works

Panaetius, the eldest son of an eminent family on the commercially important Greek island of Rhodes, studied at the great library in Pergamum before coming to Athens in the 150s BC. There he studied under Diogenes of Babylon, the leading Stoic of the time. Visiting Rome in the 140s BC, he stimulated intense interest in Greek philosophy in general and Stoicism in particular among the so-called 'Scipionic circle' - a loose-knit group of politically conservative but culturally philhellenic nobles centred on Scipio Aemilianus (then the dominant figure in Rome) and idealized in Cicero's dialogue *On the Republic*. He accompanied Scipio as an informal advisor on a major diplomatic tour of Egypt, Asia Minor and Greece in 140-138 BC, then returned to Athens. Around 130 BC he succeeded Antipater of Tarsus as head of the Stoic school in Athens and taught there until his death about twenty years later. The most notable of his numerous students was Posidonius.

Famously learned, Panaetius wrote on geography, astronomy, musical theory and philology as well as philosophical subjects. But none of his books survives, and while extant works by many other writers are indebted to his work, a shortage of explicit citations makes it impossible to reconstruct his thought in detail. We are best informed about his treatise *On Duty*, which Cicero (§2) followed closely in writing *On Duties* (see §2). Other works included *On Providence* (used by Cicero in *On the Nature of the Gods*), *On Equanimity* (a major source for essays on the same theme by Seneca and Plutarch), a work on enduring pain and another on government. Living at a time of renewed interest in earlier philosophy, Panaetius especially admired Aristotle (probably for his lost dialogues and not the extant treatises) and above all Plato, whom he styled 'the Homer of philosophers' and quoted often. He also discussed the life and thought of Socrates and his companions, and the authenticity of dialogues ascribed to Plato and others.

Panaetius arrived in Athens when the Stoic system elaborated by Chrysippus and his successors faced acute criticism from rival schools of philosophy, and, although his most original and influential work was in ethics, he proposed modifying Stoic positions in other areas as well. He joined Academic sceptics in marshalling arguments against the validity of divination, which earlier Stoics had considered a reliable science and an important sign of providence; against astrology, for example, he cited the remoteness of the stars, diversity in the lives and characters of twins, and the manifest influence of heredity and environment. In cosmology, he vigorously defended the Stoic conception of the cosmos as directed by an immanent providential deity but abandoned the two closely related doctrines of cosmic conflagrations and eternal recurrence (whereby the universe is periodically reduced to undifferentiated fire, then regenerated identically in an everlasting cycle) (see Stoicism §5) in favour of the Aristotelian thesis that the cosmos is everlasting. In philosophy of mind, he simplified the Stoic model of the soul by restricting it to cognitive and appetitive functions. He also argued that inherited psychic traits show that the human soul is mortal, on the ground that inheritance entails birth and anything born must die.

2 Practical ethics

Panaetius' most enduring legacy is in ethics, where he significantly broadened the appeal of Stoicism by formulating a theory of obligation rich in implications for all social ranks and especially for the conduct of public affairs. Although his writings are lost, much of his theory survives in Cicero's *On Duties*, which was closely modelled on his own treatise *On Duty*. There Panaetius focused on two questions: what actions does virtue require, and what actions are advantageous? (Cicero chides him for neglecting what to do when the two conflict, but then

justifies the omission by arguing that conflicts are illusory because transgressing virtue is never really advantageous.) Downplaying the gap between doing what is right and doing it from perfectly virtuous motives - encapsulated in the Stoic distinction between 'duties' or 'appropriate acts' (*kathēkonta*) and 'correct acts' (*katorthēmata*) (see Stoicism §§15-16) - Panaetius emphasized practical precepts over abstract principles, and, in order to help people of imperfect character make 'progress' toward virtue, he constructed systematic rules of conduct for the proper pursuit and use of things that Stoicism maintained have only instrumental value, such as wealth, power and prestige.

Panaetius based his proposals on the Stoic account of moral development through a natural process called 'familiarization' or 'affiliation' (*oikeiōsis*) (see Hierocles; Stoicism §14). But he refined this theory by tracing all four cardinal virtues back to infant roots: wisdom to an innate desire for knowledge, justice to the social impulses inherent in rationality, courage to an impetus to excel, and temperance to an instinct for self-preservation. Accordingly he reformulated the Stoic acount of the human good - originally defined as 'living virtuously' and standardly explicated as 'living in agreement with nature' - as 'living by our naturally given propensities'. This richer picture of human nature served in part to answer charges that the Stoic ideal of virtue was neither natural nor attainable: by ascribing 'seeds' of virtue to everyone, Panaetius highlighted the continuity between normal human tendencies and the perfect virtue of a sage. But his account also justified assigning greater moral worth to ordinary people and their actions; and in his treatment of the several virtues he paid little attention to how a sage would behave and a great deal to what everyone else should do.

In charting the duties associated with each virtue, Panaetius gave new prominence to positive obligations involving the wellbeing of society and its members. Justice, for instance, he extended to include not only prohibitions against causing harm to others but also duties to avert or remedy harm through generosity and public service; and courage, which earlier Stoics had associated primarily with scorning danger and enduring pain, he analysed as a species of 'magnanimity' (literally 'greatness of soul'), which he argued finds its finest expression in constructive civil leadership, not military triumph. A similar emphasis on effective action over purity of motives appears in his account of temperance, where he shifted the focus from personal to social welfare by introducing a highly original analysis of 'propriety' (*prepon*, Latin *decorum*). Exploiting analogies with the theatre, he classified duties under four kinds of 'roles' (Latin *personae*): some duties apply universally because imposed by (1) human nature, but many vary individually because imposed by (2) individual capacities and temperament, (3) station and circumstance, or (4) our own deliberate choice. One aim of this schema was to guide deliberation by articulating the variables affecting what different people should do. But since propriety is achieved by balancing the obligations imposed by one's several roles, the schema also gave new weight to social and ethical diversity and showed more clearly how to achieve moral 'progress'.

Panaetius' account of propriety reflects his rejection of the asceticism and iconoclasm early Stoics had adopted from the Cynics. But his concessions to the more humane values of civility and urbanity also marked a major expansion of the scope of obligation and gave Stoic ethics a new and distinctly political orientation. While focused on personal conduct, his ethical theory articulates a paradigm of public service that imposes substantial social and political obligations on people of talent, wealth and influence, but reduces demands on those less favoured by birth or fortune. The result is a model of paternalism which Cicero and others exploited in order to justify the aristocratic institutions of Rome. Panaetius, whose life spanned the period when the formerly autonomous Greek states were reduced to Roman provinces, thus provided Romans with a framework for justifying imperialism in general, and Roman rule in particular, as a relation of mutual advantage that could foster peace, prosperity and political stability.

See also: Aretē; Eudaimonia

STEPHEN A. WHITE

References and further reading

Alesse, F. (1994) *Panezio di Rodi e la tradizione stoica (Panaetius of Rhodes and the Stoic Tradition)*, Naples: Bibliopolis.(Meticulous study of Panaetius' innovations in ethics, psychology and cosmology.)

Cicero, M.T. (late 44 BC) *On Duties (De officiis)*, trans. M. Griffin and E.M. Atkins, Cambridge: Cambridge University Press, 1991.(The first two thirds of this work are based directly on Panaetius' lost treatise *On Duty*; the remainder purports to complete his unfinished work.)

Dawson, D. (1992) *Cities of the Gods: Communist Utopias in Greek Thought*, Oxford: Oxford University Press. (Chapter 5 argues that Panaetius dramatically shifted Stoic political and social thought in conservative directions.)

Gill, C. (1988) 'Panaetius on the Virtue of Being Yourself', in A. Bulloch, E.S. Gruen, A.A. Long and A.F. Stewart (eds) *Images and Ideologies: Self-Definition in the Hellenistic World*, Berkeley, CA: University of California Press, 330-53.(Clear general account of Panaetius' major innovations in ethics.)

Panaetius (*c*.185-*c*.110 BC) Fragments, in *Panaetii Rhodii Fragmenta (Fragments of Panaetius of Rhodes)*, ed. M. van Straaten, Leiden: Brill, 3rd edn, 1962. (Standard collection of Greek and Latin texts on Panaetius' life and thought; excludes substantial material not explicitly ascribed to him but reasonably believed to derive from his work.)

Pan-Africanism

Pan-Africanism covers a wide range of intellectual positions which share the assumption of some common cultural or political projects for both Africans and people of African descent. The political project is the unification of all Africans into a single African state, sometimes thought of as providing a homeland for the return of those in the African diaspora. More vaguely, many self-identified pan-Africanists have aimed to pursue projects of solidarity - some political, some literary or artistic - in Africa or the African diaspora. The Pan-Africanist movement was founded in the nineteenth century by intellectuals of African descent in the Caribbean and North America, who saw themselves as belonging to a single negro race. As a result the Africa of pan-Africanism has sometimes been limited to those regions of sub-Saharan Africa largely inhabited by darker-skinned peoples, thus excluding those lighter-skinned north Africans, most of whom speak Arabic as a first language.

In the twentieth century this racialized understanding of African identity has been challenged by many of the African intellectuals who took over the movement's leadership in the period after the Second World War. Founders of the Organization of African Unity, such as Gamal Abdel Nasser of Egypt and Kwame Nkrumah of Ghana had a notion of Africa that was continental. However, the movement's intellectual roots lie firmly in the racial understanding of Africa in the thought of the African-American and Afro-Caribbean intellectuals who founded it.

Pan-Africanism began as a movement in the diaspora among the descendants of the slave populations of the New World and spread to Africa itself. As a result the forms of solidarity it articulated aimed to challenge anti-black racism on two fronts: racial domination in the diaspora and racialized colonial domination in the African continent. The movement's fissures have occurred where these two clearly distinguishable projects have pulled it in different directions.

1 First formulations: W.E.B. Du Bois

The term *pan-Africanism* was coined around the time of the first pan-African congress in London in 1900. However, the idea of a form of solidarity of the negro race appeared in the nineteenth century. Like eastern European pan-slavism and the forms of nationalism that created modern Germany and Italy, these early forms of solidarity reflected a philosophical tradition derived from Herder in which peoples were the central participants in the creation of world history, expressed themselves in literature and folk culture and sought political expression in nation-states. The first black intellectual to produce a fully theorized account was W.E.B. Du Bois (1868-1963) in his paper 'The Conservation of Races' (1897) which was the second occasional paper to be published by the American Negro Academy and made use of the term *pan-negroism*. Du Bois studied with William James as an undergraduate at Harvard and went on to do graduate work at the University of Berlin. He was, therefore, familiar with the intellectual traditions of modern European nationalism and their philosophical underpinnings.

In 'The Conservation of Races' (1897) Du Bois argued that 'the history of the world is the history, not of individuals, but of groups, not of nations, but of races' (1897: 76). (He mentions Slavs, Teutons and the Romance race, indicating that, like so many other Western intellectuals of his day, he thought of real nations as races.) He argued that the differences between races were 'spiritual, psychical differences - undoubtedly based on the physical, but infinitely transcending them' (1897: 77). Also, he insisted (in a manner reminiscent of Herder) that each race was 'striving, each in its own way, to develop for civilization its particular message' (1897: 78).

The problem for pan-negroism thus framed was how negro people were to deliver their message.Du Bois was clear that African-Americans (whom he called the 'advance guard of the negro people' (1897: 79)) were to play the leading role in that task. He thought so because, although their ancestors had arrived in the New World as slaves, some like himself had been exposed to the best education and the highest forms of knowledge.

2 Nineteenth-century origins

Although Du Bois's formulation owed much to the intellectual ferment of European nationalism, it also had roots in the thought of a number of earlier African-American thinkers, whose work is best understood in the context of the broad nineteenth-century history of anti-slavery, or *abolitionist* thought. The focus of attention for all the major black thinkers in the New World in the early nineteenth century was the abolition of slavery and the slave trade.

As it was widely believed that racial antagonism between blacks and whites was inevitable (a common view explicitly held by US presidents Jefferson and Lincoln) one major preoccupation of some abolitionists was to find territories that could be inhabited by freed blacks. The colony of Sierra Leone was created in the late eighteenth century by UK abolitionists as a home for freed blacks and the black poor of the UK. The American Colonization Society played a similar role in the creation of Liberia in the 1820s. Other schemes were proposed to colonize parts of Latin America, the Caribbean and the US western frontier.

All of these schemes presupposed that Africans and their descendants in the New World belonged naturally together in a political community separated from other peoples. There were significant voices raised in protest against this assumption, notably that of the American ex-slave and abolitionist Frederick Douglass. They were joined by many others after the USA formally recognized the citizenship of people of African descent in the post-Civil War amendments to the US constitution. In the first half of the nineteenth century the majority view among white and black intellectuals was that a home was needed for the negroes if they were to be free.

Perhaps the three most important black intellectual ancestors of pan-Africanism were Martin R. Delany (1812-85), Alexander Crummell (1822-98) and Edward Wilmot Blyden (1832-1912). Martin Delany was born in southern USA, but his family moved to Pennsylvania during his youth. He began a medical education at Harvard, but was forced to leave by the refusal of white students to work alongside him. Delany's contributions to the prehistory of pan-Africanism began with a sense of a profound connection with Africa. He was proud that he was a 'full-blooded negro' and he named his children after Toussaint L'Ouverture, the black leader of the Haitian revolution, Ramses after the Egyptian Pharaoh and Alexandre Dumas after the French novelist who had some African ancestry. Delany was also a powerful voice for black emigration from the USA. He argued in *The Condition, Elevation, Emigration and Destiny of the Colored People of the United States* (1852) that only in a country without white people could black people flourish. In that early work, Delany refrained from making the obvious suggestion that blacks should travel to Africa. This was not because he was against the idea but because he and other leaders of the re-emigration movement believed that most African-Americans, convinced by anti-negro propaganda, were likely to see Africa as inhospitable. In his *Official Report of the Niger Valley Exploration Party* (1861), written after he had been to Africa, he wrote of the continent as 'our fatherland' and argued that its regeneration required the development of a 'national character'. He proposed the formula, 'Africa for the African race and black men to rule them' (Geiss 1974: 165), which is one of the earliest formulations of a pan-Africanist principle.

Alexander Crummell was born in New York and educated in the UK at Cambridge University, the first African-American to attend. He was also an ordained Anglican clergyman and the first African-American intellectual to spend a significant amount of time in Liberia. When Delany visited that country in 1859, he met Crummell who had been there by then for two decades. In *The Future of Africa: Being Addresses* (1862), a collection of essays and lectures written while he was in Liberia, Crummell developed a vision of Africa as the motherland of the negro race. In 'The English Language in Liberia' (1860a), based on a lecture given on Liberian independence day, he argued that African-Americans exiled in slavery to the New World, by divine providence had been given 'at least this one item of compensation, namely, the possession of the Anglo-Saxon tongue'. Similarly he argued for the providential nature of the transmission of Christianity to negro slaves, believing that it was the duty of 'free coloured men' in the USA to work for the Christianization of their ancestral continent. In the essay 'The Relations and Duties of Free Colored Men in America to Africa' (1860b), he expressed with great clarity the underlying racial basis of his understanding of negro identity, defining a race as 'a compact, homogeneous population of one blood ancestry and lineage' and arguing that each race had certain 'determinate proclivities' which manifested themselves in the behaviour of its members. Crummell, with Blyden, was one of the founders of Liberia College, later the University of Liberia. Unlike Blyden he did not become a permanent resident of Liberia. He returned rather to the USA where he continued to argue for the importance of an engagement with Africa on the part of blacks in the African diaspora. Crummell was the leading spirit in the foundation of the American Negro Academy and was present at the meeting at which Du Bois first read 'The Conservation of Races' (1897). He was a significant influence on Du Bois, who included an essay about Crummell in his influential publication *The Souls of Black Folk* (1903).

Edward Wilmot Blyden was born in the West Indies. He travelled to Liberia in 1850 under the auspices of the American Colonization Society. He became a citizen of that country for the rest of his life. Like Crummell, he was

a priest. Blyden was a polyglot scholar whose essays included quotations in the original languages from Alighieri, Dante, Virgil and Saint-Hilaire. He studied Arabic in order to teach it at Liberia College and later became the Liberian ambassador to Queen Victoria. In *Christianity, Islam and the Negro Race* (1887), first delivered as a lecture to the American Colonization Society in 1883, he too expressed the conviction that underlies Du Bois's first explicit formulation of pan-Africanism: 'Among the conclusions to which study and research are conducting philosophers, none is clearer than this - that each of the races of mankind has a specific character and specific work' (1887: 94).

Both Blyden and Crummell had little respect for the traditional cultures of Africa. They shared the conviction that the task of Christianized blacks of the diaspora owed a responsibility towards the conversion of their African cousins. Blyden argued that one could not deduce from Africa's 'state of barbarism' any innate deficiency in the negro. He observed that 'there is not a single mental or moral deficiency now existing among Africans - not a single practice now indulged in by them - to which we cannot find a parallel in the past history of Europe' (1887: 58). He was convinced that the negro race had an intrinsic affinity with religion and that the diffusion of Islam in west Africa had laid the groundwork for conversion to Christianity.

3 The pan-African congresses and the Garvey movement

If the intellectual ancestry of pan-Africanism is in the work of Du Bois, Delany, Crummell and Blyden, its institutional history begins with Sylvester Williams, a London barrister born in Trinidad. In 1897 he planned to bring together the 'African race' around the world. In July 1900, after a preliminary conference in 1899, such a gathering took place in London. There were four African representatives - one each from Ethiopia, Sierra Leone, Liberia and the Gold Coast colony - and around twelve from the USA, including Du Bois; eleven representatives came from the West Indies and five from London.

The conference opened with the aim of allowing black people to discuss the condition of the black race around the world. In 1921 Du Bois and others organized a second congress which met in three sessions in London, Brussels and Paris. This time there were representatives from Francophone and Lusophone Africa. A final declaration was issued which insisted on the equality of the races, the diffusion of democracy and the development of political institutions in the colonies. It also urged the return of negroes to their own countries and attention under the rubric of the League of Nations both to race relations in the industrialized world (the 'negro problem') and the condition of workers in the colonies. A third congress occurred in London in 1923 and continued in Lisbon. This second stage of the congress was more an opportunity for Du Bois to talk to some people from the Portuguese colonies on his way from London to Liberia where he was the official representative of the USA to preside over the installation of the Liberian president.

The Pan-African Congress movement disappeared until the fifth congress in Manchester in 1945 at which the continent took on a key role, as opposed to the diaspora. From there Du Bois's contribution was overshadowed by Kwame Nkrumah, who was to be Ghana's first prime minister. Du Bois was the only significant African-American present.

During the period between the First and Second World Wars when the Pan-African Congress movement was most active, sentiment received a substantial practical boost from the growth of the Garvey movement led by Marcus Garvey. He was a Jamaican immigrant to the USA, who developed his Universal Negro Improvement Association into the largest black movement in the African diaspora. The slogan of the movement was 'Back to Africa' and Garvey planned a shipping line for the purpose, although relatively few members of the organization actually left the New World for the Old. Garvey's commitment to racial pride and to the celebration of black historical achievement and his concern to link the diaspora to the continent make him an important figure in the movement's history.

4 George Padmore

George Padmore (1902-59), a West Indian-born intellectual played an important role in planning the congress in 1945. Born Malcolm Nurse, Padmore was a Trinidadian who spent time in the USA studying at Columbia University and Fisk, a black university which Du Bois also attended. Padmore worked for the Communist Party as an organizer among students at Howard University, the black university in Washington, DC. Later he spent time in

Germany and Russia, where he became the head of the Negro Bureau of the Red International of Labour Unions in 1930. In the next few years he worked for communist organizations in Austria and Germany, moving to London in 1935. From then until his death in 1959 he was the leading theorist of pan-Africanism and a close friend and adviser of Kwame Nkrumah. His *Pan-Africanism or Communism? The Coming Struggle for Africa* (1956) is probably the most important statement of his position.

5 Legacy

In the period after the Second World War African intellectuals were preoccupied with the question of independence. Once independence was attained, pan-Africanism became an ideology through which relations among the newly independent states could be thought about. Pan-Africanist rhetoric continues to be important in the language of the Organization of African Unity founded in 1963. In that same period, black intellectuals in North America were taken up with questions of civil rights. There were always resonances between these two projects - Du Bois was involved in both throughout his long life and died a citizen of Ghana. African diplomats sought to get civil rights questions raised in the forum of the United Nations, but pan-Africanism took philosophical form in the period leading up to Padmore’s work. Its major theoretical works are those of Padmore and Du Bois.

K. ANTHONY APPIAH

References and further reading

Blyden, E.W. (1887) *Christianity, Islam and the Negro Race*, repr. Edinburgh: Edinburgh University Press, 1967. (Blyden’s major work.)

Brotz, H. (1994) *African-American Social and Political Thought*, New York: Basic Books.(Contains significant excerpts from the works of all the major figures of the movements, with annotations and brief biographies.)

Clarke, J.H. (1974) *Marcus Garvey and the Vision of Africa*, New York: Vintage.(A spirited biography of Garvey.)

Crummell, A. (1860a) ‘The English Language in Liberia’, in *The Future of Africa: Being Addresses*, repr. Detroit, MI: Negro History Press, 1969.(A lecture given on the occasion of Liberian independence day.)

Crummell, A. (1860b) ‘The Relations and Duties of Free Colored Men in America to Africa’, in *The Future of Africa: Being Addresses*, repr. Detroit, MI: Negro History Press, 1969: 215-81.(Written from Liberia in the form of a letter to an African-American correspondent.)

Crummell, A. (1862) *The Future of Africa: Being Addresses*, repr. Detroit, MI: Negro History Press, 1969.(A collection of Crummell’s major writings from Liberia.)

Delany, M.R. (1852) *The Condition, Elevation, Emigration and Destiny of the Colored People of the United States*, repr. in *Search for a Place: Black Separatism and Africa, 1860*, Ann Arbor, MI: University of Michigan Press, 1969; Salem, NH: Ayer, 1988.(Delany’s first defence of African-American emigration.)

Delany, M.R. (1861) *Official Report of the Niger Valley Exploration Party*, repr. in *Search for a Place: Black Separatism and Africa, 1860*, Ann Arbor, MI: University of Michigan Press, 1969.(The report Delany wrote urging emigration to Africa after a visit there.)

Du Bois, W.E.B. (1897) ‘The Conservation Of Races’, repr. in S. Foner (ed.) *W.E.B. Du Bois Speaks. Speeches and Addresses 1890-1919*, New York: Pathfinders Press, 1970: 73-85.(A key early statement of Du Bois’s racial vision.)

Du Bois, W.E.B. (1940) *Dusk of Dawn: An Essay Toward an Autobiography of a Race Concept*, repr. Millwood, NY: Kraus-Thomson Organization Limited, 1975.(An historical reflection on Du Bois’s changing understanding of race, published in his seventy-third year.)

Du Bois, W.E.B. (1903) *The Souls of Black Folk*, repr. New York: Penguin, 1996. (Du Bois’s major statement on US black culture.)

Garvey, M. (1923-6) *Philosophy and Opinions of Marcus Garvey*, repr. New York: Atheneum, 2 vols, 1992.(A compilation of Garvey’s writings exploring many of the issues central to pan-Africanism.)

Geiss, I. (1974) *The Pan-African Movement: A History of Pan-Africanism in America, Europe and Africa*, New York: Africana Publishing Company; London: Methuen.(A solid scholarly survey of key figures in the movement and of its antecedents.)

Griffith, C.E. (1975) *The African Dream: Martin R. Delany and the Emergence of Pan-African Thought* University Park, PN: Pennsylvania State University Press.

Hooker, J.R. (1967) *Black Revolutionary: George Padmore’s Path from Communism to Pan-Africanism*, New York: Praeger.

Lewis, D.L. (1993) *W.E.B. DuBois - biography of a race, 1868-1919*, New York, NY: Henry Holt.
Lynch, H.R. (1967) *Edward Wilmot Blyden: Pan-Negro Patriot 1832-1912*, London: Oxford University Press.(A fine biography of Blyden.)
Moses, W.J. (1978) *The Golden Age of Black Nationalism: 1850-1925*, Hamden, CT: Archon Books.(An elegant cultural history exploring the life and work of figures such as Crummell, Blyden, Delany and Du Bois.)
Moses, W.J. (1989) *Alexander Crummell: a Study of Civilization and Discontent*, New York, NY: Oxford University Press.
Padmore, G. (1953) *The Gold Coast Revolution: the Struggle of an African People from Slavery to Freedom*, London: D. Dobson.
Padmore, G. (1956) *Pan-Africanism or Communism? The Coming Struggle for Africa*, repr. Garden City, NY: Doubleday, 1971.(Padmore’s major theoretical statement of pan-Africanist principle.)
Williams, M.W. (1992) *Pan-Africanism: An Annotated Bibliography*, Pasadena, CA: Salem Press.

Panpsychism

Panpsychism is the thesis that physical nature is composed of individuals each of which is to some degree sentient. It is somewhat akin to hylozoism, but in place of the thesis of the pervasiveness of life in nature substitutes the pervasiveness of sentience, experience or, in a broad sense, consciousness. There are two distinct grounds on which panpsychism has been based. Some see it as the best explanation of the emergence of consciousness in the universe to say that it is, in fact, universally present, and that the high-level consciousness of humans and animals is the product of special patterns of that low-level consciousness or feeling which is universally present. The other ground on which panpsychism is argued for is that ordinary knowledge of the physical world is only of its structure and sensory effects on us, and that the most likely inner content which fills out this structure and produces these experiences is a system of patterns of sentient experience of a low level.

1 The nature of panpsychism

Through prejudice and misunderstanding, panpsychism is often thought a somewhat fanciful doctrine; thus commentators often try to save some admired master from association with it. Consequently, an uncontroversial list of panpsychists is problematic. Especially debated is the case of Spinoza, and there is some argument too over Whitehead, though certainly many process philosophers working in the Whitehead and Hartshorne tradition are panpsychists. Other thinkers either committed, or strongly inclined, to panpsychism include Gustav Fechner, R.H. Lotze, Friedrich Paulsen (d.1908) William James, Josiah Royce, C.H. Waddington (d.1975) and Charles Hartshorne. Leibniz, Schopenhauer and Bergson advanced positions akin to panpsychism.

Panpsychism is expressed somewhat variously in virtue of differing usages of such words as ‘consciousness’, ‘sentience’, ‘feeling’ and ‘experience’. Here we shall use the word ‘consciousness’ taken in a very broad sense. (Readers may mentally substitute ‘sentience’, if they prefer, provided they do not understand this purely behaviourally.) Any individual such that there is a truth *as to what it is like to be it* (in general or at some particular moment) is conscious, and its consciousness is what that truth concerns. I may have limited power to grasp what it is like to be you, but I cannot seriously doubt that there is something there to be right or wrong about, as sensible people think also true about animals (except perhaps the very simplest). But if I try to imagine what it is like to be this table here, everyone will agree that there is nothing there to be imagined.

The point is not that one’s conscious states must be like something, must have a character; that is true of everything. But things do divide, in common opinion, into those such that there is and those such that there is not *something that it is like to be them* (though this expression is only an idiomatic pointer to something it requires a certain sophisticated obtuseness to be unable to identify).

The paradigm panpsychist maintains that each of the ultimate units of the physical world (whether particles, events or even mutually influencing fields) out of which all other physical things are made, are conscious in this sense. Of course, they are not self conscious, or thoughtful, but each has some dumb feeling of its own existence and of its exchange of influence with other things. It does not follow, they will rightly insist, that every physical thing is conscious. Many things, such as sticks and stones, made of these ultimate units are not so. Thus while a stone will be composed of conscious units without itself being conscious, a waking human brain will both itself be conscious and be composed of what is conscious, perhaps at two levels (for example, the neurons in the brain may be individually conscious and also made up, like everything else, of ultimate physical units which are so). Hence panpsychism, as such, leaves open the question as to which individuals above the minimal scale are sentient and which not, though particular systems of panpsychism may have their own suggestions as to how this can be decided. (Some think an element of behavioural spontaneity is its chief external mark.)

2 Arguments for panpsychism

The first argument for panpsychism is that it can ground the best account of how something so apparently novel as consciousness could have arisen within a physical world whose development has otherwise been simply a re-arrangement of the homogeneous. One theory based on this argument is that the experience of those non-ultimate units in nature which are conscious are literally composed of the experiences of their ultimate parts. Thus my experience consists in the experiences of my neurons (and those of the experiences of their ultimate

parts) which unite in a way in which the experiences of the parts of non-conscious things do not. Thus nothing essentially novel has come into the universe with human or animal consciousness, only new solidifications of the consciousness (sometimes called 'mind dust') already pervasive in the universe.

This is not a very satisfactory view. For consciousness seems to exist only in distinct individual units or 'centres', and it seems doubtful that these can combine to make more comprehensive ones. Even if this is not in principle impossible (and some panpsychists have reasons for thinking it possible in principle), introspection of our own consciousness hardly suggests that any of its components have a distinct sense of their own being.

However, there are various ways in which the theory may be made more promising. Perhaps it is a law of mental nature that when conscious individuals form a system of a certain type, that system becomes conscious in its own right so that its behaviour is due to a combination of the mental states of its parts and its mental state as a whole, without the latter strictly being composed of the former. If so, the emergence of high-level consciousness like ours is due to laws concerning the 'charge' of consciousness associated with matter in general rather than the result of its purely physical character, and thereby seems more intelligible.

The panpsychism we have considered so far implies no particular view of the nature of physical reality. It simply holds that each ultimate unit of the physical world has a certain 'charge' of sentience, which is additional to its physical characteristics, and that in certain circumstances more complex units of nature receive their own individual 'charge' of sentience too.

The second argument for panpsychism favours the different conclusion that consciousness is the real 'stuff' of the physical universe (though this may be further reinforced by the first argument). It starts with the claim that, metaphysics apart, we only know the structure of physical things as such (and of the physical world in general) and their sensory effects on centres of consciousness like ourselves - identified as the consciousness pertaining to certain complex physical things similarly specified - not the content in which that structure is realized concretely. Thus our knowledge of physical reality, so far as it goes beyond characterizations of things simply as the cause of certain sensations in ourselves, is rather like the kind of knowledge of a piece of music which someone born deaf might have from a musical education based entirely on the study of musical scores, such as could lead them to play (perhaps somewhat lifelessly) compositions on the piano without having any idea of the specific quality of heard sound in general or of the particular sounds currently being produced.

It being acknowledged that (metaphysics apart) we only know the structure of physical reality (or only the structure and the experiences it produces in us), a speculative mind will wonder whether its qualitative nature must remain entirely unknown. They will reflect that there is one kind of thing, after all, of which we do know the inherent quality, namely our own consciousness and, by inference, also the consciousness of other humans and to a limited extent animals. In short, we know the generic nature of consciousness and a good deal about the specific forms it can take. Moreover, we are incapable even of conceiving in any genuinely full way any thing more than a mere abstract structure (needing to be embodied in something more concrete to exist) which is not a form or a content of consciousness. (This is an essentially idealist claim which cannot be examined here: see Idealism §2.) Could this be because what we call the generic essence of consciousness is in fact the generic essence of all possible fully concrete reality? If so, the reality which produces 'perceptual' experiences in us, and the structure of which science (and less precisely, common sense) aspires to formulate and control, must somehow be composed of consciousness. This is at least a hypothesis worth exploring as the only alternative to saying that matter is unknowable in its inner essence, and as likely also to cast light on the mind-body or mind-brain relationship.

3 Consciousness in panpsychism

But how can consciousness have that kind of structure? The simplest hypothesis is that there are innumerable interacting centres of consciousness which are the inner being of nature's ultimate physical units; that there are definite lines of possible influence between them, either connecting them immediately or through 'intervening' ones; and that the 'geometry' of these lines is more or less adequately represented by what we conceive of as their spatial or spatiotemporal relations (or those of complexes including them), while a description of any particular physical process is thus a purely structural account - supplemented by an indication of how it is liable to affect our own perceptual experiences - of the way in which these possible influences have become actual.

However, there are various alternative paths which an attempt to understand the world panpsychistically may take.

In particular, there are alternative accounts of how the over-all or dominant consciousness of a human being (such as is calling itself 'I' when one speaks) or animal fits into the scheme. The hypothesis just described suggests that the laws which govern the interaction of such a dominant centre of consciousness with the lower-level centres of consciousness which constitute its body, and via that with other things, are distinct from those which govern those interactions between lower-level centres of consciousness which constitute more 'purely physical' processes both within its body and in nature at large. This has a certain kinship to dualism since, though mind and matter are ultimately the same kind of 'stuff', fresh laws of interaction apply where centres of consciousness of a higher level than those present throughout nature come into operation (see Dualism §5). Also it is likely to conceive the spatial location of such a dominant centre of consciousness as more diffuse than that of the physical units corresponding to the centres which it dominates. (In principle, it can recognize wholes other than animals with a similarly dominant consciousness diffusely existing within them.)

However, other forms of panpsychism try to fit our consciousness into a world of interacting centres of consciousness whose structural description remains solely that of universal physics. These have some kinship with 'double aspect' conceptions of mind and brain. Various further versions of panpsychism are possible too, all sharing the claimed advantage of conceiving the mind-brain relation as a relation between things which are of the same generic kind. Actually, panpsychists have tended to develop their theories on the basis of an event or process ontology rather than that of individual continuants which endure through a certain length of time, so that for them the natural world in its inner being consists of innumerable streams of interacting experience rather than of interacting sentient continuants. Panpsychism is also sometimes associated with some form of absolute idealism according to which all things are included in one all-embracing consciousness in a manner which displays itself as their containment in a single spatiotemporal system.

T.L.S. SPRIGGE

References and further reading

Hartshorne, C. (1950) 'Panpsychism' in V. Ferm (ed.) *A History of Philosophical Systems*, New York: Philosophical Library, 442-53.(A clear statement of Hartshorne's panpsychism.)

Hartshorne, C. (1962) *The Logic of Perfection*, La Salle, IL: Open Court.(Chapters 7 and 8 offer a statement of Hartshorne's process philosophy, including his panpsychism - what he prefers to call his 'psychicalism'.)

James, W. (1890) 'The Mind-Stuff Theory', in *The Principles of Psychology*, vol. 1, ch. 4, London: Macmillan; Cambridge, MA: Harvard University Press, 1981. (Critical examination of the view that animal and human consciousness is a totality composed of little bits of sentience pervasive in the natural world.)

James, W. (1911) 'Novelty and Causation: The Perceptual View', in *Some Problems of Philosophy*, ch. 13, New York: Longmans, Green & Co.(This appears as ch. 9 in the scholarly Harvard University Press edition of 1979. A late work of James in which he seems finally to adopt the panpsychism with which he had toyed throughout his life.)

Nagel, T. (1974) 'Panpsychism', *Philosophical Review* 53 (October); repr. in *Mortal Questions*, Cambridge: Cambridge University Press, 1979.(An influential philosopher who sees some point in panpsychism, even if he does not endorse it.)

Paulsen, F. (1892) *Introduction to Philosophy*, 2nd US edn, trans. F. Thilly, intro. W. James, New York: Holt & Company, 1930.(A fascinating discussion of a range of philosophical problems, many of which the author thinks are best solved by the view that the inner essence of what appears to us as physical is pervasively mental.)

Royce, J. (1892) *The Spirit of Modern Philosophy*, New York: W.S. Norton & Co., 1967.(In chapter 12 Royce distinguishes between the world of description and the world of acquaintance. The second, which is mental, is the inner essence of the former, which is the world as viewed by science.)

Sprigge, T.L.S. (1983) *The Vindication of Absolute Idealism*, Edinburgh: Edinburgh University Press.(The first half of this book provides a defence of panpsychism.)

Whitehead, A.N. (1929) *Process and Reality*, corrected edn, New York: Free Press, 1978.(The main statement of Whitehead's process philosophy which many, but not all readers, understand as committed to panpsychism.)

Pan-Slavism

In B.H. Sumner's words: 'Since Pan-Slavism was in general not so much an organized policy, or even a creed, but rather an attitude of mind and feeling, it was at the time correspondingly difficult to gauge its power, just as it is now to analyse its different elements' (Sumner 1937). Logically and philosophically weak, and in fact usually deficient in any kind of intellectual structure, this identification with, preference for, and emphasis on the Slavs has, nevertheless, been a presence in European (and to a much lesser extent world) history ever since its emergence in early nineteenth century.

1 Origins

While modern Pan-Slavism had some interesting predecessors, such as the seventeenth-century Croat and Roman Catholic priest Juraj Križanić who spent many years in Russia and advocated Slav unity under the Muscovite tsar, its own emergence should be dated from the first part of the nineteenth century. It was produced by two related intellectual and cultural phenomena: the new Romantic ideology centred on organicism and, increasingly, historicism; and the establishment of modern philology. The first provided the fundamental worldview for at least the early stages of Pan-Slavism; the second pinpointed the object of allegiance and devotion, the organism to which one belonged, namely Slavdom as attested by the use of a Slav language. Some pioneers of new thought, notably Herder, not only provided the necessary framework for Pan-Slavism, but also themselves dealt favourably with the Slavs. Ironically, that new thought, although many-sided, rich, and eventually as extensive as Western civilization itself, could also be regarded as most especially German.

2 Pan-Slavism in central and eastern Europe

Very much in the spirit of the age, Pan-Slav ideas eventually influenced Slav intellectuals wherever they could be found, including the small and isolated Sorbian communities in Saxony. Even the briefest account must mention developments in four large areas: the Czech and Slovak territories of the Habsburg Empire; Illyrian or South Slav lands (the latter being another crucial linguistic designation, although in fact it omitted Bulgaria), both inside the Habsburg state as in the cases of the Croats and the Slovenes, and outside it as in the case of most of the Serbs; divided Poland; and Russia.

In this climate of ill-defined hopes and dreams nationalism took hold of two young Lutheran Slovaks who may be regarded as the fathers of early Pan-Slavism, Jan Kollár (1793-1852) and Pavel Josef Šafařík (1795-1861). Kollár became its first poet, Safarík its first scholar. Kollár's inspiration came directly from his student years at the University of Jena, 1817-19, and represented both a faithful reproduction of the new German Romantic nationalism and its transformation into a Slav one, in opposition to the German. The two pioneers were followed by numerous other enthusiasts including the indefatigable writer Ludevít Stúr (1815-56) and the historian Frantisek Palacký (1798-1876), who 'construed Czech history as a struggle between the peace-loving, naturally democratic Slavs and the bellicose and aristocratic Germans' (Kohn 1953: 24). Different kinds of Slavs constituted almost half of the population of the Habsburg Empire, and Slav issues were bound to be prominent in its politics, especially in the age of nationalism. The revolutionary years of 1848 and 1849 witnessed such developments as a revolution in Prague, as well as in Vienna and in Budapest, an abortive Pan-Slav congress in Prague, and Croat support of the Habsburgs in their war against the Hungarians. The settlement of 1867 transformed the state into a dual entity, Austrian-German and Hungarian, increasing Slav dissatisfaction and also suggesting further change into a triune body, the Slavs finally becoming equal to the two dominant nationalities. But other possibilities were pre-empted by the First World War and the resulting collapse of the Habsburg Empire. That war, of course, was significantly related, both in its immediate origins and in its more fundamental causes, to Slav, even Pan-Slav, issues.

Backward in terms of education and participation in general European culture, in fact many of them oriented for centuries more towards Istanbul than towards Vienna, the South Slavs too began to be affected by Pan-Slavism. Ljudevit Gaj (1809-72) was the moving spirit of this Illyrism, later Yugoslavism (South Slavism). A Croat, educated in Graz and Budapest, influenced by Kollár and active especially in Zagreb, Gaj produced a *Short Outline of Croat-Slovene Orthography* which did much to establish the modern unified Serbo-Croatian literary language, which the author used to publish the first Croatian newspaper and even to create the Croatian national

anthem. Croat-centred Yugoslavism was further developed by the Roman Catholic bishop Josip Juraj Strossmayer (1815-1905) and other Zagreb intellectuals. But other South Slavs were less responsive. The Serbs, following Vuk StefanovićKaradžić (1787-1867), writer and language reformer, and other pioneers, stressed their own nationalism and looked at contemporary Serbia as the Piedmont of the South Slavs, who were to be brought together in the great Serbia of the future. Even the representatives of the small Slovene people, animated by their poet France Prešeren (1800-49), emphasized Slovene cultural distinctiveness and considered themselves loyal Austro-Slavs. As to the Bulgarian awakening, promoted especially by Iurii Hutsa (1802-39), better known after his change of name as Iurii Ivanovich Venelin, it consisted mainly in a certain emancipation from the Greek ecclesiastical and intellectual domination and a bid for a mighty Bulgaria sponsored by Russia.

Whereas South Slavs, and some other Slavs as well, can be described as ‘awakening’ in the first decades of the nineteenth century, Poland was then experiencing a literary and intellectual renaissance of world significance. The disparity makes all comparisons inadequate. Such thinkers as Stanislaw Staszic (1755-1826), Josef Marie Hoene-Wroński (1778-1853), August Cieszkowski (1814-94), Bronislaw Trentowski (1808-69) and the historian Joachim Lelewel (1786-1861), together with such great writers and poets as Adam Mickiewicz (1789-1855), Juljusz Slowacki (1809-49) and Zygmunt Krasiński (1812-59), focused explosively their Slav, and indeed Pan-Slav, themes on the historic and contemporary tragedy of Poland and in particular on its relations with Russia. Suggested remedies ranged from wise leadership of Slavdom by the Russian tsar to hopes for a revolutionary and democratic Russia united with a similar Poland, and to still more unreal messianic and apocalyptic visions - often accompanied by some of the most brilliant denunciations and criticisms of Russia and the Russians ever made. Moreover, Polish messianic thought easily burst the boundaries of Slavdom, presenting Poland as the redeemer of the world, ’Christ of the nations’.

3 Pan-Slavism in Russia

In contrast to Habsburg Slavs or Poles, Russians (or at least Great Russians, not to prejudge the complex and evolving issues of Ukrainian and Belorussian nationalisms) did not live in a state dominated by other nationalities. Nor was the Russian Empire a marginal Balkan principality. Indeed Russia stood out as the only Slav world power, the beacon of expectation and hope for many non-Russian as well as Russian Pan-Slavs. Yet, in spite of its impressively advantageous position, Russian Pan-Slavism both in its nature and its historical role bore a striking resemblance to its counterparts west and south of the Russian borders. Derived from German Romanticism, ’Slav’ themes were ably developed by a number of Russian intellectuals, meshing well with a certain sympathy for the Slavs (not the Poles, however), in particular as against Turkey and later the Germanic empires, present in the broader Russian public. Yet Pan-Slav intentions were usually at cross-purposes with government policy, and Russian Pan-Slavism remained characteristically a matter of sentiment - and at times a bugbear to ill-informed foreigners - but not an effective force.

The Slavophiles, who constructed the most comprehensive and the most creative Russian Romantic ideology, are of special interest for the ’Slav’ theme (see Slavophilism). The leading members of the group, all of them landlords and gentlemen-scholars of broad culture and many intellectual interests, included Aleksei Stepanovich Khomiakov (1804-60), who applied himself to everything from theology and world history to medicine and technical inventions, Ivan Vasil’evich Kireevskii (1806-56) who has been called the philosopher of the movement, his brother Pëtr Vasil’evich Kireevskii (1808-56) who collected folk songs and left very little behind him in writing, Konstantin Sergeevich Aksakov (1817-60), a specialist in Russian history and language, his brother Ivan Sergeevich Aksakov (1823-86), later prominent as a publicist and a Pan-Slav, and Iurii Fëdorovich Samarin (1819-76), who was to have a significant part in the emancipation of the serfs and wrote especially on certain religious and philosophical topics, on the problem of the borderlands of the empire, and on the issue of reform in Russia. This informal group, gathering in the salons and homes of Moscow, flourished in the 1840s and 1850s until the death of the Kireevskii brothers in 1856 and of Khomiakov and Konstantin Aksakov in 1860.

Slavophilism expressed a fundamental vision of integration, peace and harmony among human beings. On the religious plane it produced Khomiakov’s concept of *sobornost’*, an association in love, freedom, and truth of believers, which Khomiakov considered the essence of Orthodoxy. Historically, so the Slavophiles asserted, a similar harmonious integration of individuals could be found in the social life of the Slavs, notably in the peasant commune - described as a ’moral choir’ by Konstantin Aksakov - and in such other ancient Russian institutions as

the *zemskii sobor* (a kind of Muscovite estates general to advise the tsar). Again, the family represented the principle of integration in love, and the same spirit could pervade other associations of human beings. As against love, freedom and cooperation stood the world of rationalism, necessity and compulsion. It too existed on many planes, from the religious and metaphysical to that of everyday life. Thus it manifested itself in the Roman Catholic Church - which had chosen rationalism and authority in preference to love and freedom and had seceded from Orthodox Christendom - and, through the Catholic Church, in Protestantism and in the entire civilization of the West. Moreover, Peter the Great introduced the principles of rationalism, legalism, and compulsion into Russia, where they proceeded to destroy or stunt the harmonious native development and to seduce the educated public (see Enlightenment, Russian §1). The Russian future lay in a return to native principles, in overcoming the Western disease. After being cured, Russia would take its message of harmony and salvation to the discordant and dying West. The all-embracing Slavophile dichotomy represented - as pointed out by F. Stepun and others - the basic romantic contrast between the romantic ideal and the Age of Reason. As to Russian Pan-Slavism, the Slavophiles should be considered its important predecessors rather than its fully-fledged adherents. The Slavs never occupied a central position in their teaching, and in fact had only a minor and superficial connection with Khomiakov's theology and none at all with Ivan Kireevskii's philosophy. Yet other members of the group, and even the same Khomiakov, developed popular 'Slav' themes in prose and poetry, glorifying their imaginary ancestors and brothers, and providing one of the most decisive and far-reaching criticisms of the West available. There was a certain logic to Ivan Aksakov's becoming, later in the century, a leader of a genuine Pan-Slav movement. (Professor S. Lukashevich even suggested that he did so after he concluded that a titanic war between the Slavs and the Germans would shake Russia to its very depths and thus bring about the return to the old Russian spirit and principles and the end of the 'Western', Petrine period of Russian history - a crucially important transformation, unaccounted for in the original Slavophile theory.)

'Slav' and even Pan-Slav ideas were also propounded by certain adherents to the government ideology which came to be known as Official Nationality, such as Professors Mikhail Petrovich Pogodin (1800-75) and Stepan Petrovich Shevyrev (1806-64), although never by Tsar Nicholas I himself or his immediate assistants. Other enthusiasts included two great literary figures, the poet Fëdor Ivanovich Tiutchev (1803-73) and, later, the novelist Fëdor Mikhailovich Dostoevskii (1821-81), as well as a number of lesser writers. Russian Pan-Slavism even acquired a remarkable supporter on the extreme left: Mikhail Aleksandrovich Bakunin (1814-76), 'founder of nihilism and apostle of anarchy', who believed in the great revolutionary potential of the Slavs and participated in the Pan-Slav congress in Prague in 1848 and some other revolutionary events of those years.

The Russian Pan-Slav movement of the second half of the nineteenth century is usually traced to the Moscow Slav Benevolent Committee founded in January 1858 and to its later sections or more autonomous organizations in such cities as St Petersburg, Kiev and Odessa. The committees devoted themselves to cultural and philanthropic enterprises such as scholarships for Balkan students in Russia and some financial aid to schools and churches in the Orthodox parts of the Ottoman and Habsburg Empires. They also served as a fulcrum for Pan-Slav thought, with special emphasis on the Russian language as the obvious common language for the Pan-Slav world. A Pan-Slav congress was held in Moscow in 1867, with eighty-one Slav guests from abroad, but it produced no tangible results, the question of Poland remaining the most intractable divisive issue. Russian Pan-Slavism experienced a great revival during the Balkan Wars of the late 1870s as thousands of Russian volunteers flocked to fight for the Bulgarians and the Serbs against the Turks before Russia itself joined the fray. But if the Treaty of San Stefano of 1878 could be considered the high point for Russian Slav sentiments and Pan-Slavism, its remaking that same year at the Congress of Berlin came as a sharp disappointment. Slav and Pan-Slav sentiments also rose occasionally later in connection with major events in the Balkans. The First World War itself became inevitable when Russia came to the aid of Serbia against Austria-Hungary. Bulgaria, however, chose to fight on the opposite side, Habsburg Slavs provided manpower for the Habsburg armies, although some of them went over to the Russians, while Poland remained a problem for the Russian Empire until its dying day.

Together with world politics, the intellectual climate also changed. Whereas Pan-Slavism stemmed from Romanticism and developed for decades within the Romantic framework, in the second half of the nineteenth century it came to be influenced and even dominated by new trends of thought: 'realistic' (as in *realpolitik*), pragmatic and at the same time scientistic, that is, applying (or rather misapplying) scientific categories to human society and history. Danilevskii's *Rossiia i Evropa*, published in the periodical press from 1869 and as a book first

in 1871, and frequently referred to as the Bible of Pan-Slavism, may be considered the epitome of the new approach. An able natural scientist, Nikolai Iakovlevich Danilevskii (1822-85) concluded that universal history and culture were delusions and that the only reality consisted in the existence and evolution of entirely separate linguistic-ethnographic entities. He discovered about a dozen such entities as well as some peoples who did not properly belong to history, because they offered either nothing or only destruction. Five laws were basic to the historical process: (1) language identified race or family of peoples; (2) for its particular civilization to develop that entity had to have political independence; (3) the basic principles of each entity were unique and could not be transmitted to another cultural-historical type; (4) to obtain the full richness of development each entity should be a federation of states and not absorbed by one of them; (5) as in the case of perennial monocarpic plants the period of growth of these cultural-historical types was indefinite, but the florescence and fruition of each occurred briefly and only once. The next step in history was to be the replacement of the Romano-Germanic, or European, by the Slav type, led, of course, by Russia - for in the visions of Ivan Aksakov and other Pan-Slavs a titanic war loomed on the horizon. Once successful, the Slavs could well institute the first society harmoniously synthesizing all four main aspects of human activity: the religious, the cultural, the political and the socioeconomic (it is especially in this last respect that the earlier cultural-historical types had proven so far deficient). In the meantime only Slav success mattered, not chimeras of universal civilization, universal standards or humanity itself.

4 Historical character and significance

At least so far Pan-Slavism has proven abortive as it failed in its expressed aim of uniting the Slavs, let alone of inaugurating a new Slav age of European and world history. The fortuitous circumstance that all Slavs found themselves, after the Second World War, within the confines of the Soviet Union or its East European satellites was but a vicious travesty of the Pan-Slav ideal, because that ideal could not even be expressed in the new regime, except in a highly circumscribed manner and in subservience to quite different beliefs of the new rulers. Nor did communist unity survive.

Reasons for the failure of Pan-Slavism are not difficult to find. A destabilizing and destructive doctrine, it had to challenge continuously the establishment, or rather establishments. Hence Austro-Slavism, which reflected both the restrictions on Pan-Slavism in the Habsburg state and the willingness of many Slavs to accommodate themselves to these restrictions at the expense of broader Pan-Slav vistas. In Russia the Pan-Slavs never captured the government and could only very occasionally provide valuable support for it or throw significant obstacles in its path. Old religions were often at cross purposes with Pan-Slavism. Thus the Russian emphasis on Orthodoxy or the Polish on Roman Catholicism frequently alienated and even antagonized Slavs who adhered to other creeds. Moreover, Pan-Slavism proved to be only one of the secular beliefs in modern European history, and not the most powerful one. Its overwhelming rival has been nationalism. The two are remarkably similar in their irrationality, romantic origins, linguistic emphasis and organicism. However, Romantic nationalism demands allegiance to the so-to-speak primary, ’national’ group, while Pan-Slavism and its counterparts direct it to the larger, secondary group, ’tribal’, more broadly ethnic or ’racial’. And it is to the primary group that the allegiance has been going. Thus Bulgaria fought a whole series of wars with Serbia; Ukrainian nationalism rose in opposition to Poland and Russia; Russian-Polish relations hardly need comment; even more telling may be the collapse of Yugoslavia (not Pan-Slav, to be sure, but nevertheless a secondary formation, for it aimed to bring at least the South Slavs together) in Croatian and Serbian massacres.

See also: South Slavs, philosophy of

NICHOLAS V. RIASANOVSKY

References and further reading

Danilevskii, N. Ia. (1869) *Rossiia i Evropa (Russia and Europe)*, New York: Johnson Reprint Corporation, 1966. (The so-called Bible of Pan-Slavism.)

Fadner, F. (1961) *Seventy Years of Pan-Slavism in Russia, Karazin to Danilevskii, 1800-1870*, Washington, DC: Georgetown University Press.(A good antidote for those readers who might imagine that this entry exhausts the subject.)

Kohn, H. (1953) *Pan-Slavism: Its History and Ideology*, Notre Dame, IN: University of Notre Dame Press. (Provides a rich and varied picture. Professor Kohn also published several other books related to the topic of this article.)

Petrovich, M.H. (1956) *The Emergence of Russian Panslavism, 1856-1870*, New York: Columbia University Press.(Professor Petrovich wrote with a certain lustre as well as with a marvellous sense of humour.)

Riasanovsky, N.V. (1952) *Russia and the West in the Teaching of the Slavophiles: A Study of Romantic Ideology*, Cambridge, MA: Harvard University Press.(Contains a useful bibliography.)

Riasanovsky, N.V. (1959) *Nicholas I and Official Nationality in Russia, 1825-1855*, Berkeley and Los Angeles, CA: University of California Press.

Riasanovsky, N.V. (1992) *The Emergence of Romanticism*, New York and Oxford: Oxford University Press.

Sumner, B.H. (1937) *Russia and the Balkans, 1870-1880*, Oxford: Clarendon Press.(Some crucial problems presented by a primarily diplomatic, rather than intellectual, historian.)

Walicki, A. (1982) *Philosophy and Romantic Nationalism: The Case of Poland*, Oxford: Oxford University Press. (An excellent study. Other relevant books by Professor Walicki are also very highly recommended.)

Pantheism

Pantheism contrasts with monotheism (there is one God), polytheism (there are many gods), deism (God created the world in such a way that it is capable of existing and operating on its own, which God then allows it to do) and panentheism (in God there is a primordial and unchanging nature, and a consequent nature that changes and develops). Etymologically, pantheism is the view that Deity and Cosmos are identical. Theologically, it embraces divine immanence while rejecting divine transcendence. If atheism is the denial that anything is divine, pantheism is not atheism; if atheism is the claim that there is no Creator, Providence, transcendent Deity, or personal God, pantheism is atheistic.

Spinoza, perhaps the paradigm figure for pantheism, was described by some as 'a God-intoxicated man' and by others as an atheist. On his account, only God or Nature *exists, a single, necessarily existing substance whose modes and qualities exhaust reality. Conceivable equally properly as physical or as mental, God or Nature is no proper object of worship, creates nothing, grants freedom to none, hears no prayer, and does not act in history. Personal immortality, on Spinoza's view, not only does not occur, but is logically impossible. It is one thing to value nature so highly that one calls it a divinity, another to believe in God in any monotheistic sense.*

This much said, it must be admitted that 'pantheism' is not easy to define precisely. As conceived here, pantheism need not be a variety of materialism, and if it is materialistic it includes a high view of the worth of matter. Yet 'pantheism' has served as a term of abuse, and as another term for 'atheism' and 'materialism' and 'deism', terms bearing quite different senses.

1 God and world

Within a pantheistic perspective, God is not conceived as transcending the universe and so is not thought of as Creator or Providence, cause of the world though distinct from it, an agent who affects history, or a comprehending hearer of prayer and receiver of worship. There is nothing not part of the world on which the world might depend. Monotheism means by 'world', roughly, 'whatever exists besides God' and this makes possible its claim that the world depends for its existence on God while God exists independently. Pantheism means by 'world' simply 'all there is' with nothing left over, and hence rejects any Creator/creature distinction. From a monotheistic standpoint, this is atheism, however valuable the pantheist may conceive everything, or each thing, to be (see Atheism §1).

Nor, in contrast with polytheism, does pantheism view nature as coming under the control of various beings, each with limited capacities and knowledge. Polytheistic deities are gods and goddesses of limited capacities, but they are typically conceived as possessing a transcendence of nature incompatible with pantheism. Further, pantheism is a radical type of monism, holding both that there is, strictly, only one thing, however many and diverse its interlocking elements, and one fundamental kind of thing (see Monism). Thus not only the distinction between Creator and creation, in which the former transcends the latter, but such distinctions as soul and body, concrete and abstract, and immaterial and material, seen as identifications of kinds of things such that anything belonging to one kind cannot belong to the other, are rejected.

2 Varieties of pantheism

Historical views often identified as pantheistic include Hesiod's *Theogony* and Stoicism, with its doctrine that God is a rational spirit who, shapeless himself, makes himself into all things, and its view of God as the tension or tendency to coherence that holds together the various sorts of things that there are, each sort itself defined in terms of its degree of internal tension or coherence (see Hesiod; Stoicism §§3-5). Xenophanes (§3) arguably viewed the cosmos as a living, conscious, divine unity. The Upaniṣads (800 BC and later) contain passages in which Brahman is represented as claiming identity with all sorts of things. The pantheistic and polytheistic tendencies expressed in the Upaniṣads are countered by other passages that offer a more monotheistic view and a tendency in Hinduism to henotheism. Advaita Vedānta Hinduism, represented classically by Śaōkara, holds that there exists only qualityless Brahman, a view similar to Schelling's (see Schelling, F.W.J. von §2; Vedānta). Johannes Scottus Eriugena (§3) held that anything exists only in so far as it participates in the essence of God, but he also insisted that no term applicable to creatures can be predicated of God. Fichte (§§6-7) thought of the world as the material through which

the Ego achieves its moral work, and identified the divine with a moral order that is composed of Ego and world. Giordano Bruno (§5) held God to be distinct from finite individuals only by including them within God's being, and followed Hermetic writings in making the unity of the All in the One a basic theme of his own thought (see Hermetism). Meister Eckhart, Boehme and Hegel (§§3-5) are not infrequently read pantheistically. Benedict de Spinoza's philosophy is a particularly fully developed and influential version of pantheism (see Spinoza, B. de §§2-4).

3 Two routes to pantheism

One might arrive at a pantheistic view of things by first believing that a self-conscious ontologically independent being exists, and then coming to think that if *B* depends for its very existence on *A* then *B* is part of *A* or *B* is in some other manner not a distinct being from *A*. Then one would have started as a monotheist and ended as a pantheist. This process will be helped along if one also thinks that if *B* depends for its very existence on *A*, then for any quality *Q* that *B* has, *B* has *Q* only because *A* causes *B* to have *Q*. Monotheism plus determinism plus the view that dependence rules out distinctness will yield pantheism. This is one route.

A different possible route goes in the opposite direction, as it were. If one believes that everything that exists depends every moment for its existence on something else, that everything that exists is valuable by virtue of simply existing, and that dependence rules out distinctness, then one will presumably conclude that there exists but one organic thing and (at least if one believes that valuable parts guarantee a valuable whole or else that the whole itself is valuable just by virtue of its existing) that this one thing is valuable. If one thinks very highly of its value, one may express one's view by saying that it is divine, using 'divine' in some such sense as 'possesses the highest possible value'. It is perhaps in this attitude that pantheism contrasts with atheism, in so far as there is a contrast.

4 Spinoza: God or Nature

Either route will leave one in the position of thinking that there is one whole composed of intimately related and interdependent parts. One rather precise way of putting this doctrine was developed by Spinoza. Spinoza's complex views, classically expressed in his *Ethics* (1677) come to something like the following. *God or Nature* is the only thing that exists. It is a substance - a possessor of states and properties and not itself a state or property. It has logically necessary existence, and each item that exists that is not identical to God or Nature is a mode or state of it, or a characteristic of it. Causal connections are properly understand as matters of, or at least as somehow isomorphic with, logical entailments. God or Nature is described as 'cause of itself', but this is not intended literally. It is Spinoza's way, not of saying that God or Nature both does not exist (so that its existence can be caused) and does exist (in order to do the causing), but that it is logically impossible that God or Nature should not exist or should depend for existence on anything.

Spinoza thinks that fatalism is true - any true proposition is necessarily true and any false proposition is necessarily false. Thus nothing that does exist could possibly have failed to exist, and nothing that does not exist could possibly have existed. Thus, while Leibniz and a great many other philosophers have talked about and thought in terms of possible worlds, believing that there is an infinite number of ways things might have been, for Spinoza there is not a multiplicity of possible worlds. There is exactly one way that things might have been, and that is the way things are; possibility is necessarily coextensive with actuality. Any characteristic that any state of God or Nature has is one that it could not possibly not have had, and any characteristic that any state of God or Nature lacks is one that it could not possibly not have lacked. Hence true natural science should be conceived of as analogous to the geometrician Euclid's *Elements*, which itself is conceived of as consisting of necessarily true axioms and necessarily true theorems that necessarily follow from the axioms. True physics is a fully developed deductive (every axiomatic truth is included and every derivable theorem has been derived) system whose axioms are necessarily true, and physical reality is isomorphic with true physics. Psychology, the science of mental states, is also a fully developed deductive system whose axioms are necessarily true.

God or Nature can be conceived under the attribute of extension, as what exactly corresponds to the true physics. God or Nature can also be conceived under the attribute of thought, as what exactly corresponds to the true psychology. Each way of conceiving it is entirely correct and leaves nothing out, save the fact that there is another equally exhaustive way of correctly describing and explaining things. Along with the various attributes and their corresponding theories, there is a set of notions - substance, attribute, mode, cause, finite, infinite, and the like -

that are involved in one's conceiving anything under any attribute, and thus that are in principle core concepts in any theory.

Indeed, Spinoza thinks that while we are able to conceive of only two attributes of God or Nature, there is an infinite number of attributes, to each of which in principle corresponds an exhaustive description of everything and an explanation of everything it is logically possible to explain, and each of which has its own core concept that plays within it the same role that extension and thought play in the theories they are central to. While 'being a body' and 'being a mind', then, are kind-defining features for Spinoza, he denies that they define ontologically incompatible kinds. Rather, he holds, they define exhaustive accounts of things, neither of which can be translated into the other, and each of which gives an account of the same thing as the other; necessarily connotatively incommensurate, they are necessarily denotatively identical. The resulting view is neither materialism, which would require that the true physics be the whole story, nor idealism, which would require that the true psychology be the whole story.

5 Consequences

Several consequences follow from a pantheistic perspective:

(1) *Regarding freedom*. A person is *categorically free* regarding an action if and only if, under prevailing conditions, they have the power both to perform and to refrain from performing it, and, whichever they do, they do not make false a law of logic or of nature, or some truth about the past. No one can possess categorical freedom if pantheism is true. This consequence is especially clear for Spinoza's version; his epigram is that freedom is only ignorance of causes. Human freedom presupposes logically possible alternatives to what is done, and according to Spinoza there are none. Divine freedom is a matter of none of the characteristics of God or Nature following from any essence but God or Nature's own (see Freedom, divine §§1-2). What is valued for humans is an understanding of the nature of things, joined with equanimity in accepting truths as necessary truths. In other varieties of pantheism, too, possessing freedom requires a separateness and autonomy hard to reconcile with pantheistic metaphysics. A categorically free agent is independent in the sense that they can so act as to determine some of their own properties in a manner not already determined by the situations and properties of other things. The unity that pantheism posits arises from everything depending for its existence and its properties on the existence and properties of other things (or on everything else depending for its existence and properties on one thing). States of a substance, parts of an organism, are not categorically free agents, even if the substance or organism is all-encompassing.

(2) *Regarding immortality*. Another consequence concerns individual immortality. In Spinoza's version, individual survival is logically impossible. Martha's body is a state or mode of God or Nature conceived under the attribute of extension. It is identical to what can also be described as Martha's mind or soul, which is the same mode as Martha's body, though now considered under the attribute of thought. Martha's body ultimately dies and disintegrates; hence it necessarily dies and disintegrates. Since Martha's mind or soul is identical to Martha's body, Martha's mind or soul ceases to exist when Martha's body ceases to exist. Spinozistic 'immortality' is a matter of what Martha knew corresponding to a part of what God or Nature conceived under the attribute of thought knows, and of Martha's material remains being composed of physical stuff that is part of what makes up God or Nature conceived under the attribute of extension. Spinoza is thus committed to the problematic view that necessarily, and hence either eternally or everlastingly, true propositions can entail propositions that specify particular temporal limitations, such as 'Martha exists from the date of her birth to the date of her death, neither before the one nor after the other'. 'The one substance there is necessarily exists' must, regarding its finite modes, entail propositions having the form 'Mode M exists only for a while'.

Even in pantheism's non-Spinozistic varieties, individual persons are conceived as so related to both other persons and to things that the very existential dependence and causal interaction that constitutes or yields their unity (that makes things *pan*theistic) militates against any ontologically deep individuality at any time. The at least potential separateness of a person from their body and natural environment, of the sort that doctrines of individual survival typically presuppose, is inconsistent with a pantheistic conception of things (see Soul, nature and immortality of the §2).

(3) *Regarding error and evil*. It is Spinoza's view that all error is simply confusion; what is false is

self-contradictory and thus, he believes, cannot be thought. The closest, then, it is possible to come to error is that one only confusedly understands the truth. There can be no such thing as false belief. Similarly, there is no room for genuine evil. What is, necessarily is, and the distinctions we make between good and evil have no purchase on a world in which it is logically impossible that anything be different.

Even if there are logically contingent truths, if everything is thought of as divine there is pressure to deny that anyone actually believes erroneously or that anything is genuinely evil (intellectual error itself may be seen as one variety of evil). Admittedly, dependence and lack of clarity are not perfections, but a pantheist can claim that they (in contrast to error and evil) are inevitable concomitants of being less than the whole. Thus while God or Nature cannot be evil, and God or Nature is all there is, a mode may seem defective considered by itself, but were it considered in the full context of its necessary existence as part of a system of modes and hence as part of God or Nature, its apparent defectiveness would vanish. Neither can there be sin, and so there can be no need for divine forgiveness (see Evil §3).

Spinoza’s statement of pantheism allows one to see what it entails. *Mutatis mutandis*, similar conclusions regarding freedom, personal immortality, and error and evil are likely to follow from other versions of pantheism.

See also: Deism; God, concepts of

KEITH E. YANDELL

References and further reading

Flint, R. (1877) *Theism*, Edinburgh and London: William Blackwood & Sons.(Baird Lecture for 1876; polytheism, pantheism and theism are compared in chapter 2.)

Flint, R. (1879) *Anti-Theistic Theories*, Edinburgh and London: William Blackwood & Sons.(Baird Lecture for 1877, with discussion of the history and content of pantheism in chapters 9 and 10; see also the appendices.)

Orr, J. (1897) *The Christian View of God and the World*, New York: Charles Scribner’s Sons.(The Kerr Lectures for 1890-1; critical of pantheism, with discussion of its nature and history.)

Spinoza, B. de (1677) *Ethics Demonstrated in Geometrical Order and Divided into Five Parts*, trans. E. Curley, *The Collected Works of Spinoza*, vol. 1, Princeton, NJ: Princeton University Press, 1985.(Spinoza’s *magnum opus* in metaphysics and epistemology.)

Thomas, G.F. (1965) *Religious Philosophies of the West*, New York: Charles Scribner’s Sons.(A critical analysis of religious thinkers from Plato to Tillich; see chapter 7.)

Thomas, G.F. (1970) *Philosophy and Religious Belief*, New York: Charles Scribner’s Sons.(A systematic sequel to the historically ordered previous volume; see chapter 4.)

Paracelsus (Philippus Aureolus Theophrastus Bombastus von Hohenheim) (1493-1541)

Paracelsus (pseudonym of Theophrastus Bombastus von Hohenheim) was an itinerant Swiss surgeon and physician who formulated a new philosophy of medicine based on a combination of chemistry, Neoplatonism and the occult, all within a Christian framework. His works, usually in German rather than Latin, were mostly published after his death. His importance for medical practice lay in his insistence on observation and experiment, and his use of chemical methods for preparing drugs. He rejected Galen's explanation of disease as an imbalance of humours, along with the traditional doctrine of the four elements. He saw the human being as a microcosm that reflected the structure and elements of the macrocosm, thus presenting a unified view of human beings and a universe in which everything was interconnected and full of vital powers. Paracelsian chemical medicine was very popular in the late sixteenth and seventeenth centuries, largely due to its presentation as part of a general theory.

1 Life and works

Theophrastus Bombastus von Hohenheim (sometimes given the additional names Philippus Aureolus) was born in Einsiedeln, a small Swiss town near Zürich. His pseudonym, Paracelsus, first recorded in 1529, may mean 'greater than Celsus' (a Roman medical writer). When his father, a medical doctor interested in alchemy, moved to Villach in Austria in 1502, Paracelsus came into contact with mining technology and the study of metals. He worked as an apprentice in the mines near Schwaz, but otherwise little is known of his education. He may have studied with Johannes Trithemius (1462-1516), author of many works on magic. He probably studied medicine at the University of Ferrara (1513-16), though there is no documentary evidence that he received a degree. Although medical training had a practical component, the focus was academic, involving study of standard Latin translations of Hippocrates (see Hippocratic medicine), Galen and Avicenna (see Galen; Ibn Sina). He then worked as an army surgeon during his travels across Europe from 1517 to 1524. Surgery was primarily a practical discipline, and much of the surgical literature was in the vernacular. The relationship between academically trained physicians and surgeons was an uneasy one, and Paracelsus persistently complained that he was recognized as a surgeon, not as a physician. From 1524 to 1525 Paracelsus was in Salzburg, but had to leave in a hurry, possibly because of involvement in a peasant rebellion. Subsequently he visited a number of spas, and developed an interest in the minerals found in spa waters. In 1526 he was in Strasbourg where he made some useful contacts leading to his appointment (March 1527) as municipal physician in Basle. This post carried with it the right to lecture, and led to a major clash with the university, which had not been consulted. Paracelsus announced that he was not going to teach the works of Hippocrates and Galen, and he burned the Canon of Avicenna in public. Furthermore, he gave his lectures in Swiss-German, not Latin. After verbally attacking a magistrate for not giving him full support in a case he brought for non-payment of fees (a frequent problem), he was forced to flee from Basle early in 1528. He now began a second set of journeys in Switzerland, Austria, Bohemia and Southern Germany, which are much better documented than his earlier journeys. In 1529, he visited Nuremberg where he studied and wrote on syphilis. He gave a very accurate description of the disease, but his recommendations for treatment were highly controversial. In 1533 he spent time in the mining areas of Appenzell, studying the diseases of miners, and giving the first written account of an occupational disease. He also continued his studies of spa waters. He claimed that he effected many marvellous medical cures, but seems never to have had much financial success. He was invited to Salzburg by the bishop, but died there in 1541, leaving few possessions.

Paracelsus was a prolific writer, though few of his works were published during his lifetime. Those that were, tended to be about prophecy and astrological forecasts. A few works require special mention. In his *Paragranum* (1529-30), he argued that medicine should be based on the four pillars of natural philosophy, astronomy, alchemy and virtue, by which he meant the individual powers of doctors, patients, herbs and metals. In his *Volumen medicinae paramirum* (*c.*1520) and his *Opus paramirum* (1531) he discussed basic medical doctrines. His main work on surgery, *Grosse Wundarznei*, was written in 1536 and printed immediately. His greatest and most comprehensive work, *Astronomia Magna*, was written 1537-8 and published in 1571. The first collected edition of his works dates from 1589-91, and was followed by other editions and by Latin translations. His works were widely diffused during the seventeenth century.

2 Microcosm and macrocosm

Paracelsus attacked medical orthodoxy in the name of Christian philosophy. He set out to find a new basis for medicine, and indeed for the study of the whole human being, by reverting to the ancient wisdom of Neoplatonism, enriched by elements from Hermeticism (or Hermetism), alchemy, astrology and magic. His favourite contemporary source was Marsilio Ficino (§2). He presented the resulting mixture in a biblical framework, emphasizing the necessity of Christian faith for those who sought the power of controlling natural forces, especially in healing the sick. He also emphasized the old doctrine of the two Books, the Book of Nature and the Book of Revelation (namely, the Bible), from both of which humans must learn to read.

Paracelsus' central doctrine concerned the relationship between the greater world, the macrocosm, and the lesser world, the microcosm, or human being. The elements making up the human being are those found in the macrocosm, and their structural relations are the same. In particular, the human reflects physical reality through the corporeal body which is perishable, living or spiritual reality through the astral body which gives life but is also perishable, and purely spiritual reality through the immortal soul which will take on an eternal body. Conversely, just as the human being is alive, so the macrocosm is permeated by life, and even metals grow. Furthermore, the entire universe is characterized by an elaborate series of correspondences. Plants and minerals contain powers or virtues which capture those of celestial bodies, and these powers in turn relate to states of the human body.

These doctrines have both epistemological and practical implications. Because the human being reflects the macrocosm, knowledge of external objects can be reached through knowledge of internal states and relations, particularly those involving the astral body. Knowledge is also bound up with power. Although human beings are endowed with wisdom, and hence are not governed by the stars, knowledge of the stars is particularly important in practice. Astrology enables us to know the powers, relations and influences of celestial bodies, and natural magic shows us how to control them.

Paracelsus' natural philosophy included a new doctrine of elements. He rejected the traditional four elements of earth, air, fire and water, accepting them only as visible composites. In their place he proposed sulphur and mercury, both found in Arabic alchemical sources, and salt, which he added himself. These elements should not be confused with the composites we normally call sulphur, salt and mercury. As an element, sulphur represents a principle of organization, mercury a principle of activity, and salt a principle directing matter towards a solid state. Nor should they be thought of as basic particles, for they have different qualities in different objects.

3 Medicine and chemistry

Paracelsus' rejection of the traditional four elements had a direct effect on his medical theory. Traditional medicine explained illness in terms of the four humours or bodily fluids to which the four elements gave rise: blood, phlegm, choler (or red or yellow bile) and black bile (or melancholy). Illness was taken to be a lack of balance between the four humours, a balance restored through diet and herbal medicine. Emotions, too, were linked to the four humours, and people were sanguine, phlegmatic, choleric or melancholic according to which humour predominated. These internal, mechanistic explanations were dismissed by Paracelsus. For him, diseases are distinct from one another, and are often related to chemical imbalances in specific organs that result from an interaction with external agents. These agents could be minerals, or they could be 'astral poisons' transmitted by the air. Astral influences were particularly important in explaining mental illness. Diseases are combated through the plants and metals whose 'signatures' show them to have the appropriate virtues or powers. There is always a precise correspondence between a disease and its antidote (see Medicine, philosophy of; Hippocratic medicine).

Paracelsus brought his practical experience to bear on the problem of treating diseases. Through his studies of alchemy and mining he had learned much about metals, and metallurgical techniques. He developed new laboratory methods, such as the concentration of alcohol by freezing. He used chemical techniques to derive extracts from traditional herbs, but more importantly he employed such substances as mercury, arsenic and antinomy in his treatments. Unfortunately, while his observations of diseases were remarkably exact, it is unlikely that his ability to cure them surpassed that of traditional practitioners.

Paracelsus' real achievement was his unified approach to chemistry which brought together alchemical, metallurgical and pharmaceutical techniques. Indeed, he came to see the whole world and its creation in terms of chemical transformation and separation.

4 Influence

From the 1550s onwards, as his works were edited and translated, Paracelsus' reputation began to grow. Paracelsian doctors were appointed to various European courts. Van Helmont was influenced by him, and Newton possessed a major edition of his works. At the same time, his philosophy was attacked by such figures as Mersenne and Gassendi, and there were disputes, often complex and bitter, between Galenists and Paracelsians in faculties of medicine.

Paracelsianism is particularly significant for the history of modern science. In the words of Charles Webster, 'the first major confrontation of the Scientific Revolution was between Paracelsus and Galen, rather than between Copernicus and Ptolemy' (1982: 3-4). The reasons for Paracelsus' popularity had to do with his presentation of a mystical, vitalistic philosophy of nature within a Christian framework. The continued attention paid to Paracelsus, and to Neoplatonism and the occult in general, show that a non-mechanistic, non-quantitative, non-mathematical approach still made sense, even to those who, like Newton, were responsible for advances in the new science.

See also: Alchemy §5; Hermetism; Hippocratic medicine; Medicine, philosophy of; Religion and science

E.J. ASHWORTH

List of works

Paracelsus (*c*.1520-1541a) *The hermetic and alchemical writings of Aureolus Philippus Theophrastus Bombast*, trans. A.E. Waite, London: J. Elliott, 1894; repr. Berkeley, CA: Shambhala and New York: Random House, 1976.(Good translations of selected works on alchemy and hermetic medicine.)

Paracelsus (*c*.1520-1541b) *Sämtliche Werke. I. Abteilung. Medizinische, naturwissenschaftliche und philosophische Schriften (Collected Works. Division I: Writings on Medicine, Natural Science and Philosophy)*, ed. K. Sudhoff, vols 6-9, Munich: O.W. Barth; vols 1-5, 10-14, Munich and Berlin: R. Oldenburg, 1922-33.(Standard critical edition.)

Paracelsus (*c*.1520-1541c) *Sämtliche Werke. II. Abteilung. Die theologischen und religionswissenschaftlichen Schriften (Collected Works. Division II: Writings on Theology and Religion)*, vol. 1, ed. W. Matthiessen, Munich: O.W. Barth, 1923; vols 2-7 and supplementary volume, ed. K. Goldammer *et al.*, Wiesbaden: Franz Steiner, 1955-73.(Standard critical edition.)

Paracelsus (*c*.1520) *Volumen medicinae paramirum*, trans. K.F. Leidecker, *Bulletin of the History of Medicine*, supplementary vol. 11, Baltimore, MD: Johns Hopkins University Press, 1949.(Discussion of basic medical doctrines. The original text is found in *Sämtliche Werke. I. Abteilung*, vol. 1.)

Paracelsus (1529-30) *Paragranum*, in K. Sudhoff (ed.) *Sämtliche Werke. I. Abteilung. Medizinische, naturwissenschaftliche und philosophische Schriften*, vol. 8, Munich: O.W. Barth, 1922-33.(Argues that medicine should be based on the individual powers of doctors, patients, herbs and metals.)

Paracelsus (1531) *Opus paramirum*, in K. Sudhoff (ed.) *Sämtliche Werke. I. Abteilung. Medizinische, naturwissenschaftliche und philosophische Schriften*, vol. 9, Munich: O.W. Barth, 1922-33.(Discussion of basic medical doctrines.)

Paracelsus (1536) *Grosse Wundarznei*, in K. Sudhoff (ed.) *Sämtliche Werke. I. Abteilung. Medizinische, naturwissenschaftliche und philosophische Schriften*, vol. 10, Munich and Berlin: R. Oldenburg, 1922-33.(Paracelsus' major work on surgery.)

Paracelsus (1571) *Astronomia Magna*, in K. Sudhoff (ed.) *Sämtliche Werke. I. Abteilung. Medizinische, naturwissenschaftliche und philosophische Schriften*, vol. 12, Munich and Berlin: R. Oldenburg, 1922-33.(Paracelsus' most comprehensive work. Although written in 1537-8, *Astronomia Magna* was only published posthumously.)

Paracelsus (1951) *Selected Writings*, ed. J. Jacobi, trans. N. Guterman, New York: Pantheon; 2nd edn, 1958.(A thematically organized series of translations based on the critical edition; useful glossary and bibliography.)

References and further reading

Allers, R. (1944) 'Microcosmus from Anaximandros to Paracelsus', *Traditio* 2: 319-407.(Scholarly study of a key notion.)

Debus, A.G. (1977) *The Chemical Philosophy: Paracelsian Science and Medicine in the Sixteenth and Seventeenth Centuries*, New York: Science History Publications.(A standard two-volume work on Paracelsus and his influence.)

Debus, A.G. (1978) *Man and Nature in the Renaissance*, Cambridge: Cambridge University Press.(A good short

introduction to Renaissance science and medicine, with full bibliography.)

Debus, A.G. (1991) *The French Paracelsians: The Chemical Challenge to Medical and Scientific Tradition in Early Modern France*, Cambridge: Cambridge University Press.(About the influence of Paracelsus.)

Pagel, W. (1982) *Paracelsus: An Introduction to Philosophical Medicine in the Era of the Renaissance*, revised 2nd edn, Basle and New York: Karger.(The standard account of Paracelsus' life and works.)

Pagel, W. (1985) *Religion and Neoplatonism in Renaissance Medicine*, London: Variorum.(A collection of papers most of which are devoted to Paracelsus.)

Siraisi, N.G. (1990) *Medieval and Early Renaissance Medicine: An Introduction to Knowledge and Practice*, Chicago, IL, and London: University of Chicago Press. (Very readable, full of fascinating details and illustrations.)

Vickers, B. (1984) 'Introduction', in *Occult and scientific mentalities in the Renaissance*, Cambridge: Cambridge University Press, 1-55.(Places Paracelsus in context.)

Webster, C. (1982) *From Paracelsus to Newton: Magic and the Making of Modern Science*, Cambridge: Cambridge University Press.(Referred to in §4. A readable, short discussion of the relation between occult studies and early modern science.)

Paraconsistent logic

A logic is paraconsistent if it does not validate the principle that from a pair of contradictory sentences, A *and* $\sim A$*, everything follows, as most orthodox logics do. If a theory has a paraconsistent underlying logic, it may be inconsistent without being trivial (that is, entailing everything). Sustained work in formal paraconsistent logics started in the early 1960s. A major motivating thought was that there are important naturally occurring inconsistent but non-trivial theories. Some logicians have gone further and claimed that some of these theories may be true. By the mid-1970s, details of the semantics and proof-theories of many paraconsistent logics were well understood. More recent research has focused on the applications of these logics and on their philosophical underpinnings and implications.*

The idea that a contradiction implies everything (*ex contradictione quodlibet* - ECQ) has always been a contentious one in logic. Despite this, formal paraconsistent logics, which do not validate ECQ, are creatures of the twentieth century. The earliest ones were constructed in Russia by N.A. Vasil'ev *c*.1912 - an Aristotelian logic - and I.E. Orlov in 1929 - a relevance logic (see Anderson, Belnap and Dunn 1992). However, these had no impact at the time.

Work on formal paraconsistent logics did not begin in earnest until after the Second World War. Since then they have been proposed independently by many logicians, the earliest notable ones being S. Jaśkowski (in Poland) in 1948, F.G. Asenjo (Argentina) *c*.1954, N.C.A. da Costa (Brazil) *c*.1958 and T.J. Smiley(the UK) in 1959. Work on relevance logic by A.R. Anderson and N.D. Belnap (in the USA) also started in the late 1950s (see Relevance logic and entailment), and the specifically paraconsistent aspects of relevance logic were developed by R. Routley and others (in Australia) in the late 1960s and 1970s. (Note that a paraconsistent logic need not be relevant.) Since then, work on formal paraconsistent logic has continued apace in many places, but most notably in Brazil (under the leadership of da Costa) and Australia. (Some of the original work is rather difficult to come by or of a rather preliminary nature. The best access points in the literature are: Jaśkowski (1969); Asenjo (1966); da Costa (1974); Anderson and Belnap (1975); and Routley (1977).)

A major motivation behind the construction of formal paraconsistent logics has always been the idea that in many contexts we may have information that is inconsistent, but from which we want to draw conclusions in a controlled way. (The term 'paraconsistent' was coined by M. Quesada at the third Latin American Symposium on Mathematical Logic in 1976, to indicate just this.) Examples that are frequently appealed to are: evidence provided by different witnesses, constitutions and other legal documents, various scientific theories, numerous philosophical theories, and information in a computer database. In such contexts, even though the data are incorrect, if we are stuck with them (as we may well be), then the logic had better be paraconsistent. Moreover, inconsistent scientific theories, even if they are not correct, may still be useful, or good approximations to the truth.

Some paraconsistent logicians have claimed that inconsistent theories may actually be true (see, for example, Priest 1987). The view that some contradictions or contradictory theories are true is called 'dialeth(e)ism' - from 'dialetheia' (a term coined by Priest and Routley in 1982), which means a true statement of the form $A \,\&\, \sim A$. The most commonly cited examples of dialetheias are the paradoxes of self-reference, such as Russell's and the liar paradox (see Paradoxes of set and property; Semantic paradoxes and theories of truth). The failure to obtain consensus on any consistent account of the paradoxes gives this suggestion its appeal. Other suggested examples of dialetheias include: statements about objects on the borderline of some vague predicate; moral dilemmas; and dialectical contradictions in the tradition of Hegel and Marx.

Some approaches to paraconsistent logic, such as Smiley's, obtain a suitable inference relation by starting with the classical one and filtering out ECQ and other undesirables. Such logics typically give up the transitivity of entailment. A more common approach is to specify a notion of entailment semantically, defined in terms of truth-preservation over a class of interpretations. For this approach, it is necessary to have a mechanism whereby contradictory sentences may simultaneously hold in an interpretation. For this reason, one may think of paraconsistent logic as a kind of dual of intuitionist logic (see Intuitionism), the former violating the law of non-contradiction, the latter violating the law of the excluded middle. (Though it is quite possible to have a paraconsistent logic in which $\sim (A \,\&\, \sim A)$ is semantically valid.) Paraconsistent logics of this kind

characteristically invalidate the Disjunctive Syllogism (DS): $A \vee B$ and $\sim A$ entail *B*. For both premises may be true in virtue of the properties of *A*, while *B* is not.

Various techniques have been proposed to achieve the required end. Jaśkowski's is to interpret 'true' as 'true in some possible world or other'. Da Costa's is to give up the truth-functionality of negation, so that if *A* is true, $\sim A$ may be either true or false. Routley's is to treat negation as an intensional operator, so that $\sim A$ is true at a world *w* if *A* is false at some associated world, w^*. Asenjo's suggestion (which can also be harnessed in the semantics of relevance logics) is to allow sentences to take a non-classical truth value, which may be thought of as *both true and false*, and which is a fixed point for negation.

By the mid-1970s the semantics and proof theories of many paraconsistent logics were well developed. More recently, much work has gone into their applications, both technical and philosophical. The technical applications all involve the investigation of inconsistent theories. Notable results in this area include a proof that naïve set theory with (an unrestricted comprehension axiom and) a suitable underlying paraconsistent logic is non-trivial (Brady 1989), and a proof (by R. Meyer) that there is a (consistent!) arithmetic that can prove its own non-triviality. The techniques involved in the latter involve the construction of models of full first-order arithmetic, many of which are finite (see Meyer and Mortensen 1984). It is also possible to construct non-trivial theories that contain both self-reference and epistemic operators, and that are semantically closed, in the sense of Tarski (see Priest 1991a).

The philosophical applications are more diffuse. Traditionally, consistency was thought to be the cornerstone of many important philosophical notions (for example, truth, rationality), but the viability of paraconsistent logic throws down a challenge to any such claim. In the light of this, a reappraisal both of such notions and of the significance of results that turn on consistency, for example, Gödel's incompleteness theorems, is called for and is in train.

Many would be prepared to concede the possibility of a limited use for paraconsistent logic, for example, as an inference engine for a computational database. The major criticism as far as this goes is that a paraconsistent logic is too weak to permit useful inference. In particular, a number of writers have argued that the DS is essential to most practical inference. In the author's opinion, this sort of objection carries little weight. For a start, the logical resources of the programming language PROLOG are validated in most paraconsistent logics when suitably interpreted. More generally, most paraconsistent logicians have been prepared to accept the usability of principles such as the DS in consistent ('normal'?) contexts. This idea has motivated the construction of formal non-monotonic logics in which the DS is a default inference (see Batens 1989; Priest 1991b).

Much of the criticism of paraconsistent logic has fallen on dialetheism. One of the chief general criticisms is that the paraconsistent semantics for negation fails to capture (true) negation, but this is difficult to make stick for the many-valued semantics. Some have argued that if contradictions are not to be rejected as logically unacceptable then logic is ruined as an instrument of rational criticism, but this objection trades on a confusion of what is possible according to formal logic and what is rationally possible. A third important general criticism is that if one is to accept *A* one must rationally reject $\sim A$. However, dialetheism aside, there are situations where we seem to have little rational option but to accept a contradiction (for example, the paradox of the preface). Most of the application-specific criticisms have tried to undercut a dialetheic solution to the self-referential paradoxes by arguing that paraconsistent solutions to these collapse into triviality - or, at least, fail to do so only by *ad hoc* manoeuvres of a kind that make consistent purported solutions so unsatisfactory. To date, such arguments have not hit the mark squarely. (On much of the above, see Smiley and Priest (1993).)

The possibility of violations of the law of non-contradiction was widely canvassed by pre-Aristotelian philosophers. Aristotle attacked this possibility in *Metaphysics* (1005b8-1009a5). His arguments are distinctly dubious, as was shown by Łukasiewicz (1971), but his authority has determined subsequent orthodoxy. Only a few (notably Hegel) have challenged it. Even fewer have produced a sustained defence of the law. The technical viability of paraconsistent logic and dialetheism therefore raises a challenge to contemporary philosophy of no little significance.

See also: Logical and mathematical terms, glossary of

GRAHAM PRIEST

References and further reading

Many of the works cited in §1 are in rather inaccessible places, not in English or of a preliminary nature. I have often therefore cited, instead, later papers which do not have the same features. All the works cited contain both formal and non-formal material, the relative amounts varying.

Anderson, A.R. and Belnap, N.D. (1975) *Entailment*, vol. 1, Princeton, NJ: Princeton University Press. (A summary of the early work on relevance logic.)

Anderson, A.R., Belnap, N.D. and Dunn, J.M. (1992) *Entailment*, vol. 2, Princeton, NJ: Princeton University Press.(Includes references to Vasil'ev and Orlov.)

Asenjo, F.G. (1966) 'A Calculus of Antinomies', *Notre Dame Journal of Formal Logic* 7: 103-6.(A report of Asenjo's early work on paraconsistent logic.)

Batens, D. (1989) 'Dynamic Dialectical Logics', in G. Priest, R. Routley and J. Norman (eds) *Paraconsistent Logic*, Munich: Philosophia, ch. 6.(The first non-monotonic paraconsistent logic.)

Brady, R.T. (1989) 'The Non-Triviality of Dialectical Set Theory', in G. Priest, R. Routley and J. Norman (eds) *Paraconsistent Logic*, Munich: Philosophia, ch. 16.(Includes a proof that naïve set theory with a suitable underlying paraconsistent logic is non-trivial.)

Costa, N.C.A. da (1974) 'On the Theory of Inconsistent Formal Systems', *Notre Dame Journal of Formal Logic* 15: 497-510.(A report of da Costa's early work on paraconsistent logic.)

Costa, N.C.A. da and Marconi, D. (1989) 'An Overview of Paraconsistent Logic in the 80s', *Journal of Non-Classical Logic* 6: 5-32.(Particularly helpful on work in South America.)

Jaśkowski, S. (1969) 'Propositional Calculus for Contradictory Deductive Systems', *Studia Logica* 24: 143-57. (An English translation of the earliest influential paper on paraconsistent logic.)

Łukasiewicz, J. (1971) 'The Law of Contradiction in Aristotle', *Review of Metaphysics* 24: 485-509. (An English translation of Łukasiewicz's critique of Aristotle, on the law of non-contradiction.)

Meyer, R.K. and Mortensen, C. (1984) 'Inconsistent Models for Relevant Arithmetics', *Journal of Symbolic Logic* 49: 917-29.(On models of inconsistent arithmetics.)

Priest, G. (1987) *In Contradiction*, The Hague: Kluwer.(A detailed defence of dialetheism.)

Priest, G. (1989) 'Dialectic and Dialetheic', *Science and Society* 53: 388- 415.(A discussion of Hegel, Marx and dialetheism.)

Priest, G. (1991a) 'Intensional Paradoxes', *Notre Dame Journal of Formal Logic* 32: 193-211.(The construction of non-trivial theories containing both self-reference and epistemic operators.)

Priest, G. (1991b) 'Minimally Inconsistent LP', *Studia Logica* 50: 321-31.(Consistency as a default assumption.)

Priest, G. (1984) 'Paraconsistent Logic', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, vol. 2, Dordrecht: Reidel, 2nd edn, forthcoming.(The best reference work on paraconsistent logic currently available. It has a thorough overview of the technical aspects, with some philosophical comment and references to the literature.)

Priest, G., Routley, R. and Norman, J. (eds) (1989) *Paraconsistent Logic: Essays on the Inconsistent*, Munich: Philosophia.(This standard reference work includes many interesting papers, a bibliography up to the mid-1980s and editorial essays which go into the topics discussed here at much greater length.)

Routley, R. (1977) 'Ultralogic as Universal?', *Relevance Logic Newsletter* 2: 50-90, 138-75; repr. as an appendix in *Exploring Meinong's Jungle, and Beyond*, Canberra: Research School of Social Sciences, Australian National University, 1980.(A statement of Routley's earlier views on paraconsistency.)

Smiley, T.J. (1959) 'Entailment and Deducibility', *Proceedings of the Aristotelian Society* 59: 233-54. (The earliest filter paraconsistent logic.)

Smiley, T.J. and Priest, G. (1993) 'Can Contradictions be True?', *Proceedings of the Aristotelian Society*, supplementary vol. 67: 17-54.(A debate concerning a number of objections to dialetheism.)

Paradoxes of set and property

Emerging around 1900, the paradoxes of set and property have greatly influenced logic and generated a vast literature. A distinction due to Ramsey in 1926 separates them into two categories: the logical paradoxes and the semantic paradoxes. The logical paradoxes use notions such as set or cardinal number, while the semantic paradoxes employ semantic concepts such as truth or definability. Both often involve self-reference.

The best known logical paradox is Russell's paradox concerning the set S of all sets x such that x is not a member of x. Russell's paradox asks: is S a member of itself? A moment's reflection shows that S is a member of itself if and only if S is not a member of itself - a contradiction.

Russell found this paradox by analysing the paradox of the largest cardinal. The set U of all sets has the largest cardinal number, since every set is a subset of U. But there is a cardinal number greater than that of any given set M, namely the cardinal of the power set, or set of all subsets, of M. Thus the cardinal of the power set of U is greater than that of U, a contradiction. (The paradox of the largest ordinal, discussed below, is similar in structure.)

Among the semantic paradoxes, the best known is the liar paradox, found by the ancient Greeks. A man says that he is lying. Is what he says true or false? Again, either conclusion leads to its opposite. Although this paradox was debated in medieval Europe, its modern interest stems from Russell, who placed it in the context of a whole series of paradoxes, including his own.

1 General comments

Etymologically, a paradox is something 'against' ('para') '[common] opinion' ('dox'). Nowadays it means a claim that seems absurd but has an argument to sustain it. A paradox appears 'paradoxical' when one is uncertain which premise to abandon.

The paradoxes about sets involve the principle of comprehension, which states that, for any concept, there is a class of all those objects for which the concept is true. (We use the terms 'class' and 'set' interchangeably, thus allowing each author discussed below to keep his own terminology.) Historically, the paradoxes about sets are related to the traditional antinomies of the infinite (as expressed, for example, by Galileo), to Zeno's paradoxes and to Kant's antinomies (see §4 below).

2 Burali-Forti: how not to discover a paradox

The paradox of the largest ordinal, often called Burali-Forti's paradox, is as follows. The set of all ordinals is well-ordered and so has an ordinal number, Ω, which is the largest ordinal. But for any ordinal α, there is a larger ordinal $\alpha + 1$. Hence $\Omega + 1$ is larger than Ω, a contradiction.

The early history of this paradox is a comedy of errors. In 1897 Cesare Burali-Forti *almost* discovered it. He failed to do so because he misunderstood Cantor's definition that a linearly ordered set is 'well-ordered' if (a) it contains a first element, (b) every element with a successor has an immediate successor and (c) every set of elements with a successor has an immediate successor. Accidentally omitting (c) from the definition, Burali-Forti introduced a new notion of 'perfectly ordered class' and gave for such classes an argument close to the paradox. But the conclusion he drew was that the order types of such classes are not linearly ordered. He did not think he had discovered a paradox, since nothing that he wrote challenged previous results. Later that year, he realized that he had misconstrued Cantor's definition, but saw no contradiction between his work and Cantor's, since his concerned perfectly ordered classes, not ordinals. For five years, there was no discussion in print about Burali-Forti's article. (This has often been mistakenly denied, as when van Heijenoort stated that Burali-Forti's article 'immediately aroused the interest of the mathematical world' (1967: 104).)

In 1902, after Louis Couturat alerted him to Burali-Forti's work, Russell asserted (wrongly) in print that Burali-Forti claimed that the ordinals are not linearly ordered. Here was the beginning of a genuine paradox, for Russell applied Burali-Forti's argument to the ordinals. Russell's solution was to claim that the class of all ordinals has no ordinal since it is not well-ordered. (Later he rejected this solution.)

In 1903 Russell gave the first clear statement of the paradox of the largest ordinal in his book *The Principles of Mathematics*. Thus Russell was the key figure in its emergence. Alonzo Church erroneously claimed that ‘it was through Burali-Forti’s paper that there first came to general attention the threat to the foundations of mathematics that is constituted by the antinomies’ (1971: 235). It was not Burali-Forti but Russell who first pointed out this threat in print.

3 Cantor: side-stepping the paradoxes

In 1883 Cantor had introduced infinite ordinal numbers (see Set theory; Cantor’s theorem). He distinguished sharply between finite ordinals, infinite ordinals and what he called the ‘absolutely infinite’, which included the set of all ordinals. In 1897, when Hilbert asked whether every set can be well-ordered, Cantor answered yes. Cantor’s argument indirectly included the paradox of the largest ordinal and explicitly relied on the absolutely infinite. Yet he did not regard this as paradoxical, since he had previously made the relevant distinction between the infinite and the absolutely infinite. (E.H. Moore independently found this ordinal paradox and wrote about it to Cantor in an unpublished letter of 1898.)

Replying to Hilbert, Cantor made the notion of the absolutely infinite more precise. In print two years earlier, he had defined a set as a ‘gathering together’ of distinct objects. Now he called a set ‘completed’ if it is possible, without contradiction, to think of all its members as gathered together into a whole; any other set, such as that of all ordinals, was ‘absolutely infinite’. In his Paris lecture of 1900, Hilbert affirmed that each of Cantor’s infinite cardinal numbers is consistent, but denied that there exists a set of all cardinals. Nevertheless, Hilbert gave no indication that set theory was threatened.

In 1899 Cantor communicated to Dedekind the proof that every set can be well-ordered (see Cantor 1991). Once more, the paradox of the largest ordinal was involved when Cantor showed, as part of the proof, that the collection of all ordinals is absolutely infinite. Now Cantor distinguished between what he called consistent collections and inconsistent collections; only the former were sets. (A ‘consistent’ collection was what he had earlier called ‘completed’.) Writing to Hilbert shortly afterwards, Cantor saw Dedekind’s foundation as threatened, since Dedekind assumed that every well-defined collection was consistent. Cantor had pointed out to Dedekind that the collection of everything thinkable, which Dedekind had used to prove the existence of an infinite set, was inconsistent.

4 Russell: the paradoxes emerge

The paradoxes were not discussed in print until 1903, when they appeared in Russell’s *Principles of Mathematics*. But Russell had found the paradox of the largest cardinal before he learned of Burali-Forti’s article or of Cantor’s unpublished thoughts on the subject. Russell’s own paradox emerged in May 1901 from simplifying his paradox of the largest cardinal.

But the roots of Russell’s paradox go much deeper. In 1896 he had rejected Cantor’s infinite ordinals as ‘impossible and self-contradictory’. During that period Russell proposed various antinomies in the spirit of Kant and Hegel, including the ‘contradiction of relativity’. In 1898 he characterized mathematics as follows:

> One pervading contradiction occurs almost… universally. This is the contradiction of a difference between two terms, without a difference in the conceptions applicable to them. I shall call it the contradiction of relativity. This, with addition and the manifold, appear to define the realm of Mathematics.
>
> (1898: 166)

By 1899, Russell was no longer an idealist but a Platonist. Although he ceased propounding antinomies in the style of Kant, he proposed one that originated with Leibniz. ‘Mathematical ideas’, he wrote in his 1899-1900 draft of the *Principles*, ‘are almost all infected with one great contradiction. This is the contradiction of infinity. All antinomies,… so far as they are valid at all, will be found reducible to the antinomy of infinite number’ (1900: 70).

What was this antinomy of infinite number? In 1899 Russell gave it the following form, while distinguishing between

> (a) all numbers, (b) the greatest number, (c) the last number. Observe that (b)… means the number applying to the greatest collection. All three are commonly called infinite number, and imply an antinomy, since their being

> can be both proved and disproved. (a) the most fundamental: There are many numbers, therefore there is a number of numbers. If this be *N*, $N + 1$ is also a number, therefore there is no number of numbers.
>
> (1899: 265)

Russell's antinomy of infinite number has the same formal structure as the paradox of the largest cardinal, which he formulated only in late 1900. That paradox can be stated in the following way: the class of all cardinal numbers has a cardinal number; but there is another class which has a larger cardinal number (by Cantor's theorem, the set of subsets of a set with cardinal *N* has a cardinal larger than *N*); hence there is no cardinal number of the class of all cardinal numbers. Likewise, the antinomy of infinite number has the same formal structure as his later paradox of the largest ordinal.

In November 1900 Russell redrafted part of the *Principles*. The following passage from that draft manuscript contains the earliest version of his paradox of the largest cardinal, which is closely connected with the antinomy of infinite number.

> There is a certain difficulty in regard to the number of numbers, or the number of individuals or of classes. Numbers, individuals, and classes, each form a perfectly definite class, and… every class must have a number. Now the number of individuals must be the absolute maximum of numbers, since every other class is a proper part of this one…. But Cantor has given two proofs… that there is no greatest number.

Soon Russell informed Couturat that Cantor's proofs were erroneous when applied to the class of all classes. Couturat doubted the existence of that class. But Russell vigorously defended it: 'If you grant that there is a contradiction in this concept, then the infinite always remains contradictory' (Moore and Garciadiego 1981: 327).

In May 1901 Russell discovered his paradox by applying Cantor's proof that there is no greatest cardinal to the class of all classes. During that month he wrote the earliest surviving manuscript containing his paradox. It was not expressed in terms of classes but of predicates:

> We saw that some predicates can be predicated of themselves. Consider now those… of which this is not the case…. But there is no predicate which attaches to all of them and to no other terms. For this predicate will either be predicable or not predicable of itself.
>
> (1993: 195)

Either case implies the opposite, hence a contradiction. Thus there is no predicate predicable of all and only those predicates not predicable of themselves. Russell concluded that the principle of comprehension was false. A year later, he wrote to Frege about it. While Frege saw that the paradox for predicates did not arise in his system, he was devastated by the class form of Russell's paradox. Frege abandoned the principle of comprehension, which seemed to him the only possible logical foundation for arithmetic.

In 1903 Russell published Russell's paradox, together with the paradoxes of the largest cardinal and the largest ordinal, in the *Principles*. By that date, one or more of these three paradoxes was also known to Dedekind and Hilbert (thanks to Cantor), to Husserl (thanks to Zermelo; see §5 below), to Peano (thanks to Russell) and to E.H. Moore (independently, in a letter to Cantor), but none of them published such a result. Russell's and Frege's reactions to the paradoxes, which they treated as a major challenge, stand in striking contrast to the early reactions of the others mentioned.

Russell's book also generalized his paradox: if *R* is a relation, suppose there is some *w* such that, for any *x*, *x* has the relation *R* to *w* if and only if *x* does not have the relation *R* to *x*; substituting *w* for *x* leads to a contradiction; hence there is no such *w* (1903: 102). (Russell's paradox results from substituting membership for *R*.) Later Quine expressed Russell's generalization as follows: nothing can have a relation to all and only those things that do not have it to themselves.

At the end of this book Russell tentatively presented his earliest version of the theory of types as a solution to the paradoxes. Then '*x* is a member of *x*' was meaningless, since if *x* is a member of *u*, then *x* and *u* had to belong to different types (see §7 below).

5 Hilbert, Bernstein and Grelling: German reaction to the paradoxes

Cantor was not alone in recognizing, before paradoxes were published, that there is no set of all sets. In his 1890 book on the algebra of logic, Ernst Schröder rejected such a universal set, arguing that it leads to a contradiction. Reviewing Schröder's book, Edmund Husserl rejected both Schröder's argument and his conclusion, since both resulted from a failure to distinguish between being a member of a set and being a subset. In 1902 Ernst Zermelo informed Husserl that Schröder was correct in his conclusion, although mistaken in his argument. Zermelo showed that a set which contains each of its subsets as a member is self-contradictory. He did so by stating Russell's paradox, which he had found by 1900, before Russell. Zermelo concluded that there is no set of all sets.

In 1903 Hilbert wrote to Frege that Zermelo had discovered Russell's paradox three years earlier, after Hilbert had informed Zermelo of other paradoxes which Hilbert had found a year before that. (In an 1899 letter to Frege, Hilbert had observed that there is no set of all cardinal numbers, but gave no hint that this was a threatening discovery.) Influenced by Frege rather than Russell, Hilbert first published on the paradoxes in 1905. He rejected the principle of comprehension, regarding it as the source of the paradoxes. These showed that traditional logic did not satisfy the demands of set theory; they would be solved only by an axiomatic development of logic. That year he devoted a long series of lectures to the problem without coming to a solution. Although one often reads that he did not concern himself with logic during the next decade, in fact he repeatedly lectured on the paradoxes. He only reached a solution in 1917 when he adopted Russell's theory of types (see §7 below).

In Germany the first published discussion about the paradox of the largest ordinal occurred in 1905. Felix Bernstein used this paradox while attempting to refute Zermelo's proof (1904) that every set can be well-ordered, arguing that there is indeed a set *W* of all ordinals (Moore 1982: chap. 2). But in France, Jacques Hadamard rejected Bernstein's argument: 'It is the very existence of the set *W* that generates a contradiction…. One has the right to form a set only with previously existing objects, and it is easily seen that the definition of *W* presupposes the opposite' (Hadamard in Moore and Garciadiego 1981: 339).

Several neo-Friesian philosophers commented on the paradoxes, including Alexander Rüstow (1910), in his book on the liar paradox, and Gerhard Hessenberg (1906), who found the notion of set unclear. Hessenberg argued that Russell's paradox was not dangerous for mathematicians, since they did not use a universal set, but that the paradox of the largest ordinal was dangerous.

In 1908 the neo-Friesian philosophers Kurt Grelling and Leonard Nelson discussed those paradoxes, including the predicate form of Russell's paradox. Grelling formulated a related paradox, later called Grelling's paradox: a word is called autological if it applies to itself, and otherwise is called heterological. Thus 'short' is autological and 'long' is heterological. Now 'heterological' is either autological or heterological. But either possibility leads to the opposite.

6 Berry, König and Richard: paradoxes of definability

The first paradoxes of definability emerged during 1904-5 from three distinct sources. Richard's paradox, due to Jules Richard, was the first to be published. Conceived as a response to Hadamard's comments on the paradox of the largest ordinal (see §5 above), but directed at the continuum of real numbers, Richard's paradox was the following: consider in decimal form all the real numbers that can be defined by a finite number of words. These numbers form a denumerable set *E*, whose members are $a_1, a_2, a_3, \ldots$. We define a real number *N* not in *E* by letting the *n*th decimal place of *N* be one more than the *n*th place of a_n unless the *n*th place is 8 or 9 (in which case the *n*th place of *N* is 1). Since *N* differs from the *n*th place of a_n for each *n*, then *N* is not in *E*. But *N* is defined by a finite number of words and so is in *E*, a contradiction.

In 1906 Henri Poincaré stressed the importance of Richard's paradox, and found in it a single explanation for all the paradoxes: objects defined by a vicious circle must be avoided. In Richard's paradox, there is a vicious circle since *N* is defined by using *E* but is itself in *E*. If one restricted *E* to those numbers defined without reference to *E*, there would be no contradiction. Yet Poincaré viewed the actual infinite, which he rejected, as being equally the source of the paradoxes.

Although it was the first definability paradox to be published, Richard's was not the first to be discovered. That honour goes to George Berry of Oxford University. In December 1904 Berry sent Russell the following paradox: consider the least ordinal not definable in a finite number of words; this ordinal is itself defined in a finite number of words, a contradiction.

This paradox, which Russell published in 1906 in response to Poincaré's comments on Richard's paradox, is known by the name he gave it: the paradox of the least indefinable ordinal. At that time Russell modified this paradox to give one that did not involve the actual infinite and that he called Berry's paradox, although Berry did not formulate it: the least integer not nameable in fewer than nineteen syllables. But this integer has just been named in eighteen syllables, a contradiction.

A fourth paradox of definability, called the Zermelo-König paradox, was found in 1905 by Julius König, who published it not as a paradox but to refute Zermelo's claim that the continuum of real numbers can be well-ordered. König observed that the set of real numbers which are definable in a finite number of words is denumerable. Since the continuum is not denumerable, there are real numbers *x* which cannot be defined in a finite number of words. If the continuum can be well-ordered, then there is a least such *x*, a contradiction. Hence the continuum cannot be well-ordered.

Once Richard's paradox was seen to be similar to König's result, it became clear that they concerned definability and had nothing to do with well-ordering the continuum.

7 Russell: solving the paradoxes by type theory

In 1905 Russell discussed three kinds of solutions to the paradoxes: zigzag theories, theories of limitation of size, and no-classes theories. In a zigzag theory, a property $\phi(x)$ determines the class of all *x* satisfying $\phi(x)$ when $\phi(x)$ is sufficiently simple; the difficulty with such a theory, Russell noted, was in specifying what made $\phi(x)$ sufficiently simple. A theory of limitation of size, such as that of Cantor or later Zermelo, would state that a property $\phi(x)$ determines the class of all *x* satisfying $\phi(x)$ if the class is 'limited' in size; the difficulty here was to specify how large an ordinal exists. Finally, a no-classes theory would dispense with assuming the existence of classes and operate directly with properties; although such a theory would preserve classical analysis and geometry, it was unclear how much set theory would survive.

Russell also gave a general form to all paradoxes about classes: suppose that there is a property ϕ and a function *f* such that if ϕ is satisfied by all members of the class *u*, then $f(u)$ satisfies ϕ and is not a member of *u*; the assumption that there is a class *w* of all *x* satisfying ϕ and that $f(w)$ exists leads to the contradiction that $f(w)$ has and does not have the property ϕ. In Russell's paradox, ϕ is '*x* is not a member of *x*' and $f(u)$ is *u*; in the paradox of the largest ordinal, ϕ is '*x* is an ordinal' and $f(u)$ is the ordinal number of *u*.

In the *Principles* (1903) Russell had tentatively presented his earliest version of the theory of types as a solution to the paradoxes. But during 1903 Russell became dissatisfied with his theory of types, and in 1904 adopted a zigzag theory. In 1905 and 1906 he preferred a no-classes theory. But the boundaries between his different theories were fuzzy. In 1906 he stressed that his no-classes theory contained types and was close to his 1902 theory of types.

The emergence of definability paradoxes coincided with a shift in Russell's views. Earlier, he had primarily investigated paradoxes involving classes. In 1906 he was increasingly occupied with the liar paradox. In that context he formulated the vicious circle principle, which any proposed solution to the paradoxes must satisfy: whatever involves a bound variable must not be a possible value of that variable. While accepting Poincaré's analysis that vicious circles caused the paradoxes, Russell vigorously rejected Poincaré's claim that the actual infinite caused them too.

In 1908 Russell published his *ramified* theory of types, which included much of the no-classes theory. Whereas in 1900 Russell had vigorously defended the class of all classes, his theory of types explicitly rejected it (as well as the notions of all propositions, all relations, all definitions and all ordinals). The heart of the theory was the distinction between orders and types, where orders concerned definitions, and the central assumption was his dubious axiom of reducibility, which was used to solve the semantic paradoxes.

Russell's theory was not well received at first, and was rejected by Poincaré in part because of the axiom of reducibility. The *simple* theory of types, which dispensed with this axiom and with definability, was due to Leon Chwistek and Ramsey in the 1920s (see Theory of types). In the simple theory, there was no concern with the semantic paradoxes, which could not even be formulated in it.

8 Zermelo: solving the paradoxes by axiomatizing set theory

In 1908 Zermelo published his axioms for set theory, without reformulating logic or mentioning the semantic paradoxes. To understand Zermelo's motivation, we must consider two articles he published that year in the same journal, finished within two weeks of each other, and which referred to each other. The second article contained his axiomatization, while the first defended his theorem that every set can be well-ordered. The first also gave a new proof of his theorem, based explicitly on axioms from his second article. The two articles form a single unity, revealing that his axiomatization was motivated as much by a concern to secure his theorem as by the need to solve the paradoxes (see Moore 1978).

The key to Zermelo's axiomatization was his axiom of separation, which restricted the principle of comprehension to subsets of any previously existing set. To preserve set theory, he included axioms which generated sets from previously existing sets, especially the power set and union axioms (see Set theory, different systems of).

But Zermelo's approach seemed arbitrary until he introduced the cumulative type hierarchy of sets in 1930. This hierarchy began with the empty set as the lowest level, and any level *S* was followed by a next level, the power set of *S*. The levels were indexed by ordinals, and when the index was a limit ordinal, the level was the union of all previous levels. The hierarchy of levels was cumulative since a member of any level belonged to all following levels, by contrast with Russell's types, which were disjoint.

Today Zermelo's hierarchy, as embodied in the Zermelo-Fraenkel axioms formulated in first-order logic, is generally accepted by mathematicians as the proper solution to the logical paradoxes. Among philosophers, the semantic paradoxes still lack such a generally accepted solution.

9 Brouwer, Weyl and Skolem: attacking set theory with paradoxes

Opponents of set theory repeatedly used the paradoxes as a weapon against it. This began with Poincaré in 1906 and continued with Brouwer in 1907 (see Intuitionism). Both saw intuition as central, the axiomatization of logic as suspect, and the natural numbers as fundamental. Both engaged in later attacks on axiomatization and set theory. A decade later Hermann Weyl continued the attack on analysis and set theory, using Grelling's and Richard's paradoxes.

In 1923 Thoralf Skolem used a different paradox to argue that set-theoretic concepts are relative, not absolute. Skolem's paradox states that there is a countable model for axiomatic set theory, although there exist uncountable sets. The solution distinguishes between levels. There exists some set *M* which is uncountable within the model, where there is no function mapping *M* onto the natural numbers; but outside the model there is such a function, and so, from outside, all the sets in the model are seen as countable.

10 Zermelo, von Neumann and Gödel: turning paradoxes into theorems

From a mathematical perspective, what is most striking is how paradoxes have repeatedly been turned into theorems. We consider four examples.

For Zermelo, Russell's paradox became the theorem of his axiomatic set theory that there is no set of all sets (1908).

For von Neumann, the paradox of the largest ordinal became the theorem of his axiomatic set theory (1925) that every set can be well-ordered. The proof was based on the axiom that every proper class (and no set) can be mapped one-to-one onto the class of all sets. Thus proper classes, which were 'too big' to be sets, were rehabilitated by no longer being considered to be contradictory, as they had been for Cantor and Zermelo. In effect, von Neumann made precise Cantor's distinction between consistent and inconsistent collections.

Gödel's incompleteness theorem, which states that any sufficiently rich formal system contains propositions that cannot be proved or refuted in the system (see Gödel's theorems), was obtained, in part, by applying Richard's paradox to the notion of provability (1931).

Finally, Tarski's theorem on the indefinability of truth (1933) had a close connection with the liar paradox. Tarski saw the universality of natural languages as the source of the semantic paradoxes, since this universality allows truth to be defined in such languages (see Tarski's definition of truth).

See also: Logical and mathematical terms, glossary of; Semantic paradoxes and theories of truth

GREGORY H. MOORE

References and further reading

Benthem, J.F. van (1978) 'Four Paradoxes', *Journal of Philosophical Logic* 7: 49-72.(Analyses the interrelations between the paradox of the largest cardinal, Russell's paradox, Curry's paradox about self-reference without negation, and Löb's version of the liar paradox.)

Beth, E.W. (1964) *The Foundations of Mathematics*, Amsterdam: North Holland.(Part 6 has an older, but still useful, summary of the paradoxes.)

Cantor, G. (1991) *Briefe (Letters)*, ed. H. Meschkowski and W. Nilson, Berlin: Springer.(Includes Cantor's 1899 letters to Dedekind and 1897-1900 letters to Hilbert relating to the paradoxes.)

Church, A. (1971) 'Logic, History of, IV: Modern Logic', in *Encyclopaedia Britannica* vol. 14: 231-7. (A general discussion, including the paradoxes.)

Fraenkel, A.A., Bar-Hillel, Y. and Levy, A. (1973) *Foundations of Set Theory*, Amsterdam: North Holland.(A careful analysis of the various solutions of the logical paradoxes, including axiomatic set theory, the theory of types, and metamathematics.)

Garciadiego, A. (1985) 'The Emergence of the Non-Logical Paradoxes of the Theory of Sets, 1903- 1908', *Historia Mathematica* 12: 337-51.(The early history of the semantic paradoxes.)

Grelling, K. and Nelson, L. (1908) 'Bemerkungen zu den Paradoxieen von Russell und Burali-Forti' (Remarks on the Paradoxes of Russell and Burali-Forti), *Abhandlungen der Fries'schen Schule*, new series, 2 (3): 301-34. (Grelling's paradox.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press.(Translates relevant articles by Burali-Forti, Gödel, König, Richard, von Neumann and Zermelo. Includes Russell's 1902 letter announcing his paradox to Frege and mentioning that he had previously sent the paradox to Peano but received no reply.)

Hessenberg, G. (1906) *Grundbegriffe der Mengenlehre (Fundamental Concepts of Set Theory)*, Göttingen: Vandenhoeck & Ruprecht.(The earliest textbook of set theory.)

Hughes, P. and Brecht, G. (1975) *Vicious Circles and Infinity: A Panoply of Paradoxes*, Garden City, NY: Doubleday.(A popular anthology of many paradoxes.)

Moore, G.H. (1978) 'The Origins of Zermelo's Axiomatization of Set Theory', *Journal of Philosophical Logic* 7: 307-29.(The role of the axiom of choice and the paradoxes in axiomatization.)

Moore, G.H. (1982) *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*, New York: Springer. (The second chapter indicates how the paradoxes were intertwined with Zermelo's proof that every set can be well-ordered.)

Moore, G.H. (1988) 'The Roots of Russell's Paradox', *Russell: The Journal of the Bertrand Russell Archives*, new series, 8: 146-56.(The connection between Russell's paradox and his antinomy of infinite number.)

Moore, G.H. and Garciadiego, A. (1981) 'Burali-Forti's Paradox: A Reappraisal of its Origins', *Historia Mathematica* 8: 319-50.(Detailed historical treatment of this paradox.)

Peckhaus, V. (1990) *Hilbertprogramm und kritische Philosophie: das Göttinger Modell interdisziplinarer Zusammenheit zwischen Mathematik und Philosophie (Hilbert's Programme and Critical Philosophy)*, Göttingen: Vandenhoeck & Ruprecht.(Includes reaction to the paradoxes by the neo-Friesian school.)

Rang, B. and Thomas, W. (1981) 'Zermelo's Discovery of the ''Russell Paradox''', *Historia Mathematica* 8: 15-22.(How Zermelo found Russell's paradox, prior to Russell, and discussed it with Husserl.)

Russell, B.A.W. (1898) 'An Analysis of Mathematical Reasoning, Being an Inquiry into the Subject Matter, the Fundamental Conceptions, and the Necessary Postulates of Mathematics', in *The Collected Papers of Bertrand Russell*, vol. 2, *Philosophical Papers 1896-99*, ed. N. Griffin and A.C. Lewis, London: Routledge, 1990, 155-242.(Includes the contradiction of relativity.)

Russell, B.A.W. (1899) 'The Fundamental Ideas and Axioms of Mathematics', in *The Collected Papers of Bertrand Russell*, vol. 2, *Philosophical Papers 1896-99*, ed. N. Griffin and A.C. Lewis, London: Routledge, 1990, 261-305.(Includes the antinomy of infinite number.)

Russell, B.A.W. (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992.(First publication of Russell's paradox, the paradox of the largest cardinal and the paradox of the largest ordinal.)

Russell, B.A.W. (1906) 'Les paradoxes de la logique', *Revue de métaphysique et de morale* 14: 627-50; trans. 'On ''Insolubilia'' and their Solution by Symbolic Logic', in *Essays in Analysis*, ed. D. Lackey, London: Allen & Unwin, 1973.(Analysis of the paradoxes, including the semantic ones.)

Russell, B.A.W. (1993) *The Collected Papers of Bertrand Russell*, vol. 3, *Toward the 'Principles of*

Mathematics' 1900-02, ed. G.H. Moore, London and New York: Routledge.(See 'The Principles of Mathematics (Draft of 1899-1900)' (3-180) on the antinomy of infinite number; and 'Part I of the *Principles*, Draft of 1901' (181-208) for the earliest extant version of Russell's paradox.)

Rüstow, A. (1910) *Der Lügner (The Liar)*, Leipzig: Teubner.(The liar paradox.)

Zermelo, E. (1904) 'Beweis, daß jede Menge wohlgeordnet werden kann', *Mathematische Annalen* 59: 514-16; trans. S. Bauer-Mengelberg, 'Proof that Every Set can be Well-Ordered', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 139-41.(Zermelo's proof of the well-ordering theorem.)

Paradoxes, epistemic

The four primary epistemic paradoxes are the lottery, preface, knowability, and surprise examination paradoxes. The lottery paradox begins by imagining a fair lottery with a thousand tickets in it. Each ticket is so unlikely to win that we are justified in believing that it will lose. So we can infer that no ticket will win. Yet we know that some ticket will win. In the preface paradox, authors are justified in believing everything in their books. Some preface their book by claiming that, given human frailty, they are sure that errors remain. But then they justifiably believe both that everything in the book is true, and that something in it is false.

The knowability paradox results from accepting that some truths are not known, and that any truth is knowable. Since the first claim is a truth, it must be knowable. From these claims it follows that it is possible that there is some particular truth that is known to be true and known not to be true.

The final paradox concerns an announcement of a surprise test next week. A Friday test, since it can be predicted on Thursday evening, will not be a surprise yet, if the test cannot be on Friday, it cannot be on Thursday either. For if it has not been given by Wednesday night, and it cannot be a surprise on Friday, it will not be a surprise on Thursday. Similar reasoning rules out all other days of the week as well; hence, no surprise test can occur next week. On Wednesday, the teacher gives a test, and the students are taken completely by surprise.

1 Lottery and preface paradoxes

The lottery paradox - first developed in Kyburg (1961) - and the preface paradox - originally formulated in Makinson (1965) - have a similar structure, although some (for example, Pollock (1986)) hold that they have different solutions. Each hinges on a conflict between a rule of acceptance, a condition on the transfer of warrant, and an axiom about warrant.

Rule of acceptance: There is some threshold short of certainty where acceptance of a claim is warranted or justified.
Transfer condition: A set of warranted claims is closed under deduction. That is, a set of warranted claims includes all the deductive consequences of that set.
Warrant axiom: It is not possible to be warranted in believing *p* and, for the same time and the same individual, be warranted in believing not *p*.

One standard approach is to find some fault with the transfer condition. Denials of the transfer condition sometimes result from explicit consideration of these paradoxes, sometimes from more general considerations within the theory of knowledge. For example, Kyburg addresses the lottery paradox explicitly and holds that it relies on the faulty conjunction principle, the principle according to which a person is warranted in believing a conjunction *p&q* if they are warranted in believing *p* and warranted in believing *q*. In this way, outright contradictions are thought to be avoidable in the set of warranted beliefs for a person at a time, while still allowing that the set can be inconsistent, that is, be such as to deductively imply a contradiction. Other epistemologists (for example, Nozick (1981)) develop general theories of knowledge to answer the Gettier problem in which the transfer condition is abandoned (see Gettier problems; Deductive closure principle).

To accept such a solution may require abandoning a coherence theory of justification, for a commonly proposed minimal condition for coherence is logical consistency. If this approach is the best available for solving these paradoxes, the paradoxes will have taught us that coherence theories of justification must be abandoned (see Knowledge and justification, coherence theory of §3).

A point of importance concerning those who abandon the transfer condition explicitly to solve the paradoxes is that abandoning the conjunction principle alone will not do the trick: the paradoxes can be generated without employing any conjunction principle. In the lottery paradox, one can infer from individual premises that each particular ticket will lose, together with the knowledge of exactly how many tickets there are in the lottery, that no ticket will win. This claim can be deduced without using the conjunction principle, yet contradicts the knowledge that some ticket will win, thereby violating the warrant axiom according to which contradictory warranted claims are impossible.

Moreover, a fully satisfactory solution to these paradoxes that denies the deductive closure condition would need to offer an alternative principle in its place. For it is obvious that some of our beliefs are warranted precisely because we deduce them from other things we know.

A different approach is to question the rule of acceptance. There are two approaches that could be taken here. The first is simply to succumb to the paradox, granting that no acceptance is ever warranted when a claim is less than certain. This solution, however, seems unduly restrictive. For the small chance of illusion, hallucination and other types of misperception indicate that our ordinary perceptual beliefs are less than certain.

Alternatively, one may be suspicious of the idea of acceptance or belief, on Bayesian grounds. Bayesian epistemology arises within the context of an application of a subjective interpretation of probability theory to epistemological issues, and begins by noting that we believe some things more strongly than others (we believe that 2 + 2 = 4, and that there is at present no life on Mars, but we are much more confident of the first than the second) (see Probability theory and epistemology). So, perhaps, the concept of belief in ordinary language is a coarse-grained way of talking about certain mental states that, strictly speaking, only come in degrees. Once such a viewpoint is accepted, two lines emerge as to how to treat the ordinary concept of belief. First, one might hold that there is no such thing as belief *simpliciter*, but rather there are only *degrees of belief*, and hence that the acceptance condition is ill-formed since it is formulated using the concept of belief *simpliciter*. Another alternative is to think that there is some way of understanding the ordinary concept of belief in terms of degrees of belief, using the lottery and preface paradoxes to show that such a definition cannot proceed in terms of some threshold of degree of belief above which the ordinary concept of belief applies (Kaplan 1981). The task for this kind of Bayesian view is to give an account of the ordinary concept of belief which, in conjunction with the transfer condition, does not violate the warrant axiom. In either case, the Bayesian approach gives some reason to be suspicious of the acceptance condition, and thereby suggests that the paradoxes might best be dealt with by abandoning or altering that condition.

2 Knowability paradox

The knowability paradox derives from work by Fitch (1963) on value concepts. The paradox depends on two claims:

(1) Everything is knowable - that is, for all propositions *p*, if *p* is true, then it is possible that it is known (by someone at some time) that *p* is true.
(2) Some things are not known - that is, there exists some proposition *p*, such that *p* is true, and it is not known that *p* is true.

We then substitute an instance of (2) - a claim of the form '*q* is true and it is not known that *q* is true' - as the value for *p* in the first claim. Since we are committed to the truth of this instance of (2), the antecedent of (1) is true when this instance of (2) is the value for *p* in (1). We thus deduce that it is possible that a certain conjunction is known, namely that *q* is true and that it is not known that *q* is true. Since to know a conjunction is to know that each of its conjuncts is true, we can infer that it is possible that it is known that *q* is true and that it is known that it is not known that *q* is true. However, since knowledge implies truth, the latter piece of knowledge implies that it is not known that *q* is true. So we thereby deduce that it is possible that it is both known and not known that *q* is true.

The crucial features of this argument are the distribution principle (that is, that knowledge of a conjunction implies knowledge of each of the conjuncts) and the factive character of knowledge (that is, that knowing *p* implies *p*). The remainder of the argument relies on no principles of inference beyond those of first-order quantification theory with modal operators. So the paradox has much broader implications than merely those concerning knowledge. Fitch (1963), for example, used the paradox to argue that verificationism, the thesis that all truths are knowable, entails a very silly form of verificationism according to which all truths are known (see Meaning and verification §6). The broader implications of the paradox, almost completely unaddressed in the literature, concern versions of antirealism according to which truth is an epistemic notion (for example, truth is empirical confirmation in the long run, or verification by an ideal scientific community, or what is warrantedly assertible, and so on). Such antirealisms fall within the scope of the paradox, for these epistemic notions distribute over conjunction every bit as much as knowledge does, and their theory of truth makes these epistemic notions factives. So, the paradox can

be formulated in terms of the claims that some truths are not justified or verified in this special antirealist way, and that all truths are epistemic in the favoured antirealist way.

The paradox could be resolved in a way that retains both claims (1) and (2) above if knowledge fails to distribute over conjunction. The paradox requires moving from knowledge of a conjunction to knowledge of each of the conjuncts, that is, from K(A&B) to K(A) & K(B). Some theories of knowledge deny such a principle (for example, Nozick (1981)); however, these denials are not supported by independent argument but are rather only implications of the general theory. Because of this feature, the denial can easily appear to be a defect of the theory rather than a virtue. In any case, denying the principle requires a further explanation of the difference between cases where such distribution is acceptable and those where it is not, for it is clear that one can come to know, for example, that the George owns a car by learning that George owns a car and a truck.

3 The surprise examination paradox

The surprise examination paradox first appeared in print in D.J. O'Connor (1948). It originated earlier when a Swedish mathematician, Lennart Ekbom, discussed at Ostermalms College a difficulty he had noticed with an announcement by the Swedish Broadcasting Company during the Second World War. The announcement said that a civil-defence exercise was to be held during a particular week. No one was to know in advance which particular day of the week the exercise would be conducted. Ekbom noticed that the unexpectedness of the exercise was problematic, which forms the core of the surprise examination paradox. This paradox appears in many guises and under many names, including among others the 'prediction' and 'hangman' paradoxes. All have essentially the same form as that represented by the surprise examination.

Sceptical approaches to this paradox deny that the announcement of a surprise examination warrants any beliefs about the future. Thus, Quine (1953) maintains that even a surprise examination announced only one day in advance would not be paradoxical, for such an examination would be a surprise as long as one could not know in advance that the examination would be given tomorrow.

Non-sceptical approaches to the paradox grant that one can know on the basis of the announcement that an examination will take place. For such approaches, an examination given the last day of the week would thereby not be a surprise, contrary to the Quinean approach above. Such approaches, in order to resolve the paradox, must find something wrong with the announcement itself (Shaw 1958; Kaplan and Montague 1960) or with the inferences by the students (Meltzer 1964), for the paradox ends by having an examination occur that surprises the students. A standard approach following the first line is to show that the teacher's statement is self-referentially incoherent, as in the Liar paradox where 'This statement is false' is incoherent. For example, the announcement might mean or imply 'There will be a surprise examination next week, and there being no surprise examination is deducible from this entire statement'.

Such approaches have been rejected for violating a coherence requirement on the announcement (Bosch 1972), according to which it is plainly obvious that the announcement is coherent and hence that any adequate resolution of the paradox must locate an error in the students' reasoning.

More recent discussion has focused on different varieties of the paradox, and whether formal approaches to the paradox suffice to resolve it. In spite of being originally castigated as 'rather frivolous' (O'Connor 1948), this paradox has attracted by far the most attention of the epistemic paradoxes (nearly one hundred articles have been written on it, compared to only a handful for the others). Moreover, hardly anything uncontroversial is to be found regarding the paradox, including how to interpret the announcement, how to resolve the paradox, whether there is one or several paradoxes involved, whether the paradox is simply a variant of other well-known paradoxes, and what conditions a proper resolution of the paradox must satisfy. Some would hold that it is the deepest of the paradoxes; in any case, the attention it has received shows that it is far from frivolous.

See also: Epistemic logic; Scepticism

JONATHAN L. KVANVIG

References and further reading

Binkley, R. (1968) 'The Surprise Examination in Modal Logic', *Journal of Philosophy* 65: 127-36.(Develops a sceptical approach to the paradox interpreted as involving modal operators.)

Bosch, J. (1972) ‘The Examination Paradox and Formal Prediction’, *Logique et Analyse* 15: 505-25.(Argues that the paradox must find a flaw in the students’ reasoning rather than in the announcement itself.)
Edgington, D. (1985) ‘The paradox of Knowability’, *Mind* 95: 557-68.(Attempts to develop a variation of verificationism immune to the knowability paradox.)
Fitch, F. (1963) ‘A Logical Analysis of Some Value Concepts’, *The Journal of Symbolic Logic* 28 (1963): 135-42. (The original source for the central argument of the knowability paradox.)
Kaplan, M. (1981) ‘A Bayesian Theory of Acceptance’, *The Journal of Philosophy* 78: 305-30.(Develops an account of acceptance in terms of degrees of belief that avoids the lottery paradox.)
Kaplan, D. and Montague, R. (1960) ‘A Paradox Regained’, *Notre Dame Journal of Formal Logic* 1: 79-90. (Searches for a truly paradoxical reading of the announcement in the surprise examination paradox, and finds one that makes the announcement analogous to a ‘liar’ statement.)
Kyburg, H. (1961) *‘Conjunctivitis’, Probability and the Logic of Rational Belief*, Middletown, CT: Wesleyan University Press.(The original source of the lottery paradox.)
Makinson, D.C. (1965) ‘The Paradox of the Preface’, *Analysis* 25: 205-7.(The original source of the preface paradox.)
Meltzer, B. (1964) ‘The Third Possibility’, *Mind* 73: 430-3.(Argues that the students’ reasoning is at fault for assuming the law of excluded middle.)
Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press.(A view on which knowledge is not closed under deduction, and knowledge does not distribute over conjunction.)
O’Connor, D.J. (1948) ‘Pragmatic Paradoxes’, *Mind* 57: 358-9.(The original published version of the surprise examination paradox.)
Pollock, J. (1986) ‘The Paradox of the Preface’, *Philosophy of Science* 53: 246-58. (Presents a solution to the preface paradox that arises out of artificial intelligence work on defeasible reasoning.)
Quine, W.V. (1953) ‘On a So-Called Paradox’, *Mind* 62: 65-7.(Argues that the announcement could be coherent even if given only one day in advance, thereby proposing a sceptical solution to the surprise examination paradox.)
Shaw, R. (1958) ‘The Paradox of the Unexpected Examination’, *Mind* 67: 382-4. (Claims that the solution to the paradox is found in distinguishing two readings of the announcement, only one of which can be used to deduce that no such examination can be given and which is self-referentially incoherent.)
Williamson, T. (1993) ‘Verificationism and Non-Distributive Knowledge’, *Australasian Journal of Philosophy* 71: 78-86.(Argues that knowledge must distribute over conjunction.)

Paranormal phenomena

The term 'paranormal phenomena' refers to the class of anomalous events studied within the field of parapsychology. Parapsychology's principal areas of investigation are extrasensory perception (ESP), psychokinesis (PK), and cases suggesting that personal consciousness survives the death of one's body. ESP phenomena are apparent instances of anomalous transfer of information. They divide into telepathy, clairvoyance and precognition. PK phenomena are forms, roughly speaking, of apparent mind-over-matter. Survival research deals primarily with cases of ostensible reincarnation and mediumship (or 'channelling').

The data of parapsychology raise a number of deep philosophical issues. Cases suggesting survival challenge materialist theories of the mind, and (according to some) provide good evidence for Cartesian dualism. ESP and PK challenge assumptions about the nature and temporal direction of causal relations, and also suggest the intimidating possibility that we have direct access to and influence on the thoughts and bodily states of others.

1 Varieties of paranormal phenomena

Paranormal phenomena, the phenomena investigated in parapsychology, are often called 'psi' phenomena. The term 'psi', an abbreviation for 'psychic', is considered by many to be more theoretically neutral than 'paranormal', 'psychic' or 'parapsychological': it seems to avoid issues about what should count as normal, and appears not to beg the question as to whether the phenomena under consideration fall naturally within the domain of psychology or whether they are explainable in physical terms.

Although the boundaries of parapsychology are not rigidly defined, it is clear that the field is concerned with anomalous phenomena falling into three principal areas of investigation. The first is the evidence for extrasensory perception (ESP). The study of ESP focuses on two sorts of phenomena in particular. 'Telepathy' is the influence of one person's mental states on those of another, other than by the usual or known means, and 'clairvoyance' is the similarly anomalous influence of a physical state of affairs on a person's mental states. For example, it would be a case of telepathy if a person's thoughts were directly caused by someone at a remote location thinking about something similar. It would be clairvoyance if a person's thoughts about a plane crash were directly caused by a distant plane crash. Although popular treatments of ESP suggest that telepathy and clairvoyance are types of anomalous knowledge (of remote states of affairs), philosophers and parapsychologists have long recognized that the evidence for ESP often suggests nothing more than anomalous types of causal interaction, without any sort of robust cognition deserving to be called 'knowledge'. For example, suppose that a burning house in another city caused one to have nothing more than - possibly incongruous - thoughts about fire or burning houses. If that occurs without the subject knowing that the house (or some house) is burning, it would be a manifestation of what C.D. Broad (1962) called 'clairvoyant interaction' rather than 'clairvoyant cognition'.

The second area of parapsychological investigation is psychokinesis (PK), often called 'mind-over-matter'. The study of PK concerns the apparent ability to produce physical effects independently of familiar or recognized sorts of intermediate causal links. These effects include the ostensible movement of remote objects, materializations (the apparently instantaneous production of matter), and apports (the apparently instantaneous relocation of an object). In cases of so-called 'macro-PK', as in poltergeist disturbances and mediumistic phenomena such as table levitations, the effects are produced on objects visible to the naked eye. By contrast, present-day laboratory investigations of PK study what is called 'micro-PK', the production of statistically significant nonrandom behaviour in normally random microscopic processes such as radioactive decay and thermal noise.

Although cases of apparent precognition should be classified as types of telepathy or clairvoyance, they raise distinctive issues about causality. That is because they suggest that an event (such as a plane crash) can causally influence an *earlier* state of affairs (say, a precognitive dream of the plane crash) (see Causation §7; Time travel §1). Of course, that interpretation would be resisted by those who are opposed to backward causation. In that case, if apparent precognitions cannot be explained away as mere coincidences or in terms of normal or familiar processes, two options remain. First, one could explain precognition as a form of unconscious inference based on contemporaneous information acquired by ESP. For example, the precognizer might have learned through ESP about the *present* state of the plane or mental state(s) of the passengers or crew, and that information might have surfaced to conscious awareness in a dream or premonition. Second, one might interpret precognition as a type of

PK, as telepathic influence, or both, by which the precognizer brings about the events apparently precognized.

The third area of parapsychological investigation deals with evidence suggesting the survival of personal consciousness following bodily death. Survival research focuses on cases of ostensible reincarnation and mental mediumship (that is, 'channelling' of information from an apparently deceased communicator). The evidence for survival consists primarily of anomalous knowledge of two sorts. First, some individuals demonstrate knowledge of facts or information which they presumably had no normal opportunity to acquire, and which concerns obscure or intimate details of some deceased person's life. Second, some individuals display skills or abilities - for example, speaking a language or playing a musical instrument - which they had no normal opportunity to acquire or develop, but which were associated with a deceased person either apparently communicating through a medium or apparently reincarnating in (or possessing) the person displaying the anomalous skills or abilities.

2 The challenge to materialism

Popular writings (and some philosophical works) on parapsychology often assert that the existence of paranormal phenomena would be evidence against materialist theories of the mental, and in favour of dualism (see Dualism; Materialism in the philosophy of mind). But that claim is somewhat contentious and in need of clarification.

It is important to note that there are at least two major forms of dualism. The first is Cartesian (or substance) dualism, according to which nature consists of two distinct kinds of stuff: mental (which is unextended and thinking) and physical (which is extended and unthinking). A somewhat weaker form of dualism is a level-of-description dualism, according to which nature may be described by means of either mentalistic or physicalistic vocabularies that are not entirely inter-translatable. For example, partisans of this weaker form of dualism might assert that nature consists of only physical stuff, but that the physical sciences have significant descriptive or explanatory limitations - say, in connection with the domain of mental states. (And that position can assume various forms. See, for example, Anomalous monism; Epiphenomenalism). Or, one could argue that nature is not inherently either mental or physical, but that different aspects of nature can be characterized relative to different vocabularies, or at different levels of description (see Neutral monism). Proponents of these two sorts of position could be thought of as substance-monists but level-of-description dualists. They illustrate how one can hold that mental phenomena generally are not adequately described or explained in physical terms, without claiming that the mind is a distinct kind of substance or thing. One might even say that 'the mind' is merely a general term for the class of mental events (or a certain aspect of what persons and some other organisms do), just as 'the weather' is a general term for the class of meteorological events (or a certain aspect of planetary phenomena).

Now initially, at least, it would appear that the phenomena of ESP and PK present no more of a threat to materialist theories of the mental than do everyday cognitive phenomena such as memory and volition. In fact, one can (at least hope to) handle the former just as many (at least hope to) handle the latter. The strategies would be the same in both cases. For example, one might concede that we cannot now specify the physical mechanisms underlying psychological phenomena generally, but then argue that somewhere down the road, a more sophisticated physical science will accommodate all such currently intractable phenomena. Or, one might attempt to specify underlying mechanisms or processes for the mental phenomena in question (for example, as many have tried in the case of memory and as some parapsychologists have attempted for at least certain psi phenomena). There are also many others who contend that mental phenomena generally cannot be explained with respect to underlying physical mechanisms or processes. One would think that from their perspective, the existence of ESP and PK pose problems of explanation no deeper than those posed by normal mental processes. Psi phenomena would be inexplicable in physical terms, but not in virtue of any of their features that are distinctively paranormal. As far as the physical sciences are concerned, they would be as intractable as the phenomena of memory and volition, and for the same reasons. Hence, although the forms of ESP and PK might support a level-of-description dualism, it would not be in virtue of their paranormality.

Some might argue that out-of-body experiences (OBEs) would support a stronger, Cartesian dualism. For example, suppose that a subject reports on a state of affairs visible only from the spot allegedly visited while ostensibly out of the body (there is some experimental and anecdotal evidence for this). And suppose that at the time of the experience, physical perturbations are detected at the spot where one's perceptual perspective seems to be, even though there is no object visibly occupying that position (there is preliminary experimental evidence for this as

well). Would that show that the mind is a thing that can leave and function physically apart from the body? In principle, one could argue instead that OBEs are nothing but imagery-rich clairvoyance, and that the physical disturbances can be explained as a form of accompanying PK.

Probably the only sort of parapsychological evidence that could clearly help support Cartesian dualism would be the evidence for survival. If it could be shown conclusively that consciousness survives the death and decomposition of the body, we might decide that this state of incorporeal existence can only be explained in terms of a mental substance. On the other hand, evidence of survival might equally be used to argue for a form of idealism, or a pluralistic worldview with an inventory of substance-kinds of at least three.

3 Problems of causality

Some might think that the phenomena of ESP and PK pose problems about causality because they would be examples of action-at-a-distance. But in fact, that issue worries few scientists and philosophers. Most would reject the claim that all causality is by physical contact (as in, billiard-ball causality), and some would, in any case, hold out hope for specifying the as yet undiscovered physical processes that allegedly fill in the apparent spatial causal gaps.

The most vexing problems about causality posed by the evidence concerns temporal direction. We have already noted that apparent precognition suggests that precognized events cause (earlier) precognitive experiences (although other explanatory options remain). In recent years, additional *prima facie* evidence for retro-causation has been provided by some provocative micro-PK experiments, designed and originally conducted by physicist Helmut Schmidt, and later replicated in other laboratories. These are experiments using pre-recorded random targets.

The general structure of such an experiment is as follows. Suppose that on day one a binary random number generator (that is, an electronic coin-flipper) is automatically activated (in the absence of anyone present) to record sequences of heads and tails onto audio cassettes, heads in the right channel and tails in the left. Suppose twenty such cassettes are recorded, and suppose also that a duplicate record of heads and tails is simultaneously produced on paper punch tape for the purpose of a permanent record. Then on day two, half of the cassettes are randomly selected to be test tapes, leaving the other cassettes designated as control tapes. At this point no one knows the contents of any of the tapes. On day three, the ten test tapes are played back to a subject who thinks he is taking a normal PK test with spontaneously generated targets. A computer counts the numbers of heads and tails (with a duplicate record once again being made on punch tape), and it turns out that the test tapes contain a statistically significant excess of heads over tails (and also that the effect size is similar to that demonstrated by subjects when all targets are being generated in real time). Then the control tapes are examined for the first time, and it turns out that they contain only chance levels of heads and tails (and also that both they and the test tapes match the punch-tape record for day one). But the only difference between the test and control tapes is that the test tapes were played for a subject making a PK effort. So it appears that the subject’s effort on day three biased the random generator on day one to produce an excess of heads only for the test tapes (whose selection as test tapes, recall, was not decided until later). In some follow-up experiments, it appeared that repeated playback of the pre-recorded targets to the subject (without the subject’s knowing this) seemed to increase the hit rate and improve overall PK scores. Schmidt called this an ‘addition effect’.

Once again, other explanatory options remain, even if one rules out explanations in terms of normal or familiar processes. For example, some have suggested that the results of these experiments could be explained in terms of clockwise (and presumably unconscious) psychic functioning, apparently on the part of the experimenter. Although some regard this sort of apparent ‘super-psi’ as antecedently incredible in its scope and refinement, others argue that it is not substantially different from what has been documented anecdotally (especially in the best poltergeist cases and cases of physical mediumship).

What many find most unsettling about the evidence for ESP and PK is that if these phenomena occur, then presumably they do not occur only in the lab, or when parapsychologists set out to look for them. They might be occurring surreptitiously or inconspicuously much of the time in everyday contexts and even in normal scientific research. In fact, reliable evidence for ESP and PK seems to force us to entertain seriously a ’magical’ worldview usually associated only with so-called primitive cultures, and according to which we have direct and intimate

access to and influence on the thoughts and bodily states of others.

See also: Experiment; Parapsychology; Reincarnation

STEPHEN E. BRAUDE

References and further reading

Almeder, R. (1992) *Death and Personal Survival: The Evidence for Life After Death*, Lanham, MD: Rowman & Littlefield.(A useful presentation and sympathetic evaluation of the evidence for postmortem survival. Its critiques of sceptical dismissals of the evidence are especially good.)

Braude, S.E. (1979) *ESP and Psychokinesis: A Philosophical Examination*, Philadelphia, PA: Temple University Press.(Gives an overview of some of the best experimental evidence in parapsychology and considers its philosophical implications. Also examines related theoretical issues in the philosophy of mind, and contains additional chapters on synchronicity and the meaning of ‘paranormal’.)

Braude, S.E. (1986) *The Limits of Influence: Psychokinesis and the Philosophy of Science*, New York & London: Routledge & Kegan Paul; revised edn, Lanham, MD: University Press of America, 1997.(A presentation, defence and philosophical assessment of the evidence for large-scale psychokinesis (PK), focusing on nineteenth- and twentieth-century cases of physical mediumship. Also contains detailed examinations of apparitions and precognition, including a discussion of retro-causation.)

Braude, S.E. (1987) ‘Psi and Our Picture of the World’, *Inquiry* 30: 277-94.(An overview of the implications of the evidence of parapsychology for prevailing scientific theories, the debate between dualists and monists, and traditional views of causality.)

Braude, S.E. (1992) ‘Survival or Super-Psi?’ and ‘Reply to Stevenson’, *Journal of Scientific Exploration* 6: 127-44, 151-6.(An examination of the difficulties in distinguishing evidence of postmortem survival from evidence of psychic functioning among the living. Argues that the latter hypothesis is more interesting, subtle and plausible than its critics usually suppose.)

Broad, C.D. (1962) *Lectures on Psychical Research*, London: Routledge & Kegan Paul.(A classic, careful, comprehensive and philosophically penetrating examination of both experimental and anecdotal evidence, marred only by extended discussion of some now-discredited card-guessing tests. Excellent discussions of apparitions, mediumship and out-of-body experiences.)

Penelhum, T. (1970) *Survival and Disembodied Existence*, London: Routledge & Kegan Paul.(A study of some difficulties in making sense of the notion of bodiless existence and personal identity.)

Scriven, M. (1976) ‘Explanations of the Supernatural’, in S.C. Thakur (ed.) *Philosophy and Psychical Research*, New York: Humanities, 181-94.(Useful discussion of the constantly shifting and slippery meaning of the terms ‘supernatural’ and ‘materialism’, as well as the obstacles to rejecting psi phenomena in virtue of their *prima facie* incompatibility with received science.)

Parapsychology

Tales of dreams that come true, 'mind over matter' and other such oddities are both familiar and old. Parapsychology investigates such things, attempting to use scientific and, especially, experimental methods to investigate whether and in what circumstances humans can glean information without using ordinary perceptual means (that is, by extrasensory perception or ESP) or can alter the physical environment simply by willing it (that is, by psychokinesis). Such phenomena, if they exist, are often grouped under the heading psi(-phenomena).

It is sometimes claimed that parapsychology presents a challenge to physicalism. However, ostensible psi-phenomena are known through their physical effects and are studied within parapsychology by ordinary scientific methods. In fact, models intended to explain psi-phenomena by known physical processes have been seriously discussed within parapsychology. In any case, parapsychology alone could not show that psi-phenomena have no physical explanation; that is a judgment for physics itself. More important, physicalism is a very broad doctrine that should not simply be equated with the requirement that everything be explained within physics. It is very unclear what we would gain by denying that psi-phenomena are physical.

1 Origins and method

Parapsychology has its roots in nineteenth-century investigations of spiritualism. Although parapsychologists still occasionally study spontaneous phenomena such as hauntings and individual so-called 'gifted' subjects who appear to have strong psychic abilities, the dominant tone of the field in the twentieth century was set by the American researcher Joseph B. Rhine, who opened his laboratory at Duke University in 1930. Rhine is responsible for such terms as 'extrasensory perception' (ESP) and 'psychokinesis'. He believed that the best way to investigate parapsychological phenomena was by the rigorous use of statistical methods, and he typically worked with ordinary subjects in experiments that involved simple choices among clearly distinguishable targets (for example, cards with a limited number of distinct symbols). He also emphasized questions about the psychological correlates of psychic ability and the factors that might inhibit or facilitate it. Rhine's work became the model for most parapsychological research thereafter.

The flavour of parapsychological research can be gleaned from an important example: the so-called *ganzfeld* (from the German for 'whole field') studies. In these experiments, the subject is put into a state of relaxation and mild sensory deprivation - a state that is hypothesized to be conducive to psi. Then an experimenter in a separate room concentrates on one of four randomly selected pictures. Finally, a second experimenter, who does not know which picture was viewed, shows copies of all four to the subject, who is asked to choose the one he thinks was the target. By chance, subjects should be right 25% of the time. The actual success rate is closer to 33%. This is a large effect by social science standards. If you could find a casino with these sorts of odds, you would win $130 for every $100 you bet. Across the many *ganzfeld* studies that have been performed up to the writing of this entry, it has been estimated that the observed effect would occur by chance roughly one time in a billion.

2 Parapsychology and physicalism

Clearly there are different sorts of questions one could ask about such experiments. Some have to do with the design of the experiment. Particularly important will be questions about procedures for preventing fraud and, more generally, what we might term 'sensory leakage' - cues detectable by the familiar five senses. Other questions have to do with statistical analysis, replication and randomization. Questions about randomization are particularly subtle and are the subject of some ongoing debate, but even some of parapsychology's harsher critics agree that there is no simple way to dismiss the *ganzfeld* experiments. However, beyond these methodological issues are questions about just how results of this sort ought to be interpreted assuming they stand up to scrutiny. In particular, do they undermine physicalism?

Rhine (1954), writing before the *ganzfeld* experiments, believed that the answer is yes. On the basis of his own work, he believed that psi-phenomena are not subject to shielding, do not diminish with distance and, in fact, cannot be explained by appeal to normal physical mechanisms. He speculated that, instead, they involve some sort of non-physical 'psi energy' that displays its effects by being converted into ordinary physical energy. Rhine welcomed this conclusion because he saw physicalism and what he took to be its attendant determinism as a threat

to freedom and responsibility (see Determinism and indeterminism; Dualism; Materialism in the philosophy of mind).

It is hard to see how Rhine's hypothesis of 'psi energy' could address his own concerns. It may be that freedom and responsibility require deliberation and other intentional phenomena to have a certain autonomy from the non-psychological (see Free will). This might argue against reductionism or eliminativism. But the connection between this claim and the existence of some supposedly non-physical means of propagating information or influence - 'psi energy' - is obscure at best. How would the ability to move things or know things with the help of 'psi-energy' add to one's moral freedom? Indeed, what would it have to do with deliberation and choice at all? Moreover, it is not at all clear how Rhine's hypothesis would render psi non-physical. Positing a new form of energy as part of the explanation for manifestly physical effects would seem to be a proposal for an addendum to physics itself.

In fact, nothing in the results of parapsychology provides clear reasons to think that parapsychological phenomena, if genuine, violate normal physical principles. This is not least because physicists have not given any sustained attention to the question of how such things might be explained. Here it is worth remembering that the history of science provides numerous examples of phenomena (magnetism, for instance) that once were thought beyond the pale of physics but which are now understood as unquestionably physical. But there is a more general point: the mere fact that we do not have an explanation of a phenomenon in terms of fundamental physics has never really been taken as a reason to count the phenomenon as non-physical. For example, thermodynamic phenomena counted as physical before statistical mechanics attempted to provide micro-accounts of such things as temperature and entropy (see Thermodynamics). Physics may never have much to say about economics, but by itself this hardly shows that physicalism is false, since physicalism is a much broader doctrine than reductionism and kindred notions (see Reduction, problems of; Reductionism in the philosophy of mind).

It might be thought that such examples fail to do justice to the worry. Whatever the details of a microscopic account of thermodynamic phenomena, the universality of the principles of thermodynamics makes it virtually impossible to imagine what could be meant by classing them as non-physical. And while economics may group things in ways that that make reduction to physics implausible (try to give a physical definition of 'money') nothing in economics suggests any obvious conflict with physics. Psi-phenomena seem very different. They seem anomalous, and anomalous in a particular way: they look essentially like action at a distance - information or influence travels we know not how from one point to another. That is the sort of thing for which we expect a mechanism or medium and which we also expect to satisfy a variety of physical principles such as conservation laws.

Suppose we found a replicable psi-phenomenon for which no mechanism was forthcoming and which persistently refused to mesh smoothly with the laws of physics. What then?

Our sense that such things as ESP and psychokinesis ought to involve a mechanism or medium is presumably part of a larger sense that they ought to be explainable, and that a satisfactory explanation would posit some process connecting the distant events. As it turns out, however, nature is not innocent of apparently unmediated connections. Within quantum mechanics there are strong correlations among distant events that seem to defy explanation by appeal to any reasonable mechanism (see Bell's theorem). While it must be stressed that these phenomena do not permit the sending of signals, they should at least caution us not to put too much stock in our intuition that all connections must be explained by appeal to a mediating process. It is debatable whether these quantum mechanical connections are *explained* in quantum theory - or anywhere else - at all. Still, there is a difference. Quantum mechanics is a fundamental theory and at that level we might expect explanation or perhaps even intuitive intelligibility to come to an end. Parapsychology deals with the macroscopic level, and that is not where we expect to bump up against the limits of explanation.

It is not clear how much weight to give such intuitions about the macroscopic. Once again, quantum mechanics is instructive. Given the fundamental laws of quantum mechanics, it is deeply puzzling that measurements, which are macroscopic events, actually have results at all (see Quantum measurement problem). Needless to say, attempts to solve the problem, whether by conceptual innovation or physical speculation, continue, but there is no guarantee that they will produce satisfactory results.

We might characterize physicalism in terms of some idealized final physics. If brute exceptions to the principles of that ideal physics had to be carved out to accommodate parapsychological phenomena, then it might be held that physicalism, construed as a doctrine about the hegemony of physics, was false. This would mean that principles and laws of physics would lack the sort of universality sometimes claimed for them, and it would mean that natural phenomena are less unified than we might have hoped.

However, just such things have already been argued by some philosophers of science on grounds quite independent of parapsychology. It has been argued that physics has no special rule in enumerating the kinds of things there are nor in generating the laws they obey (see Dupre 1993). It has been argued that the laws of physics themselves do not fit together in any smooth way. And it has been argued that the very idea of physical laws as features of nature rather than just features of the models we use is deeply suspect (see Laws, natural).

We have been asking whether parapsychological phenomena might conflict with physicalism. The answer is a resounding ‘Perhaps’. But suppose, hypothetically, that the answer is ‘Yes’. What then? Are such doctrines as Cartesian dualism (see Dualism) forced upon us? Must we regard parapsychological phenomena as supernatural? Have we discovered the limits of science itself?

The only sensible answer to each of these questions is ‘No!’ Nature may well be highly variegated, but positing a non-spatial mental substance is as uninformative in light of parapsychology as it is without it. Parapsychological phenomena are peculiar if genuine, but their manifestations are changes in the physical make-up or the belief states of natural things, including such things as ourselves. As it stands, the very existence of psi remains controversial and so any claims about psi, as well as any claims based on it, remain answerable to checking by means that rely ultimately on the familiar five senses. If, in some improbable future, psychic phenomena became so common and so robust that we could rely on them as we rely on more familiar means of gaining knowledge or altering the landscape, this would mean that there are more ways of coming to know and change the world than we had previously thought, but it is hard to see how this would demonstrate any limitations on science; if anything, it would augment its resources.

3 Conclusion

A healthy scepticism about parapsychological claims is perfectly reasonable. Psi-phenomena would, indeed, violate various common-sense expectations and parapsychology is a field in which fraud and wishful thinking have been special problems. Furthermore, there is something odd about a field whose subject matter is largely defined negatively. At the same time, the best of parapsychological research applies the same standards that hold in the rest of science. Psi-phenomena *could* precipitate a scientific revolution, but so might any number of other things. As things stand, the results are modest and the implications unclear. For now the appropriate response would seem to be to wait and see.

See also: Paranormal phenomena

ALLEN STAIRS

References and further reading

Bem, D.J. and Honorton, C. (1984) ‘Does Psi Exist? Replicable Evidence for an Anomalous Process of Information Transfer’, *Psychological Bulletin* 115: 4-19.(Honorton developed the ganzfeld technique. Following Bem and Honorton’s article is a critique by psychologist Ray Hyman and a reply by Bem.)

Broughton, R.S. (1991) *Parapsychology: The Controversial Science*, New York: Ballentine Books.(An introduction to parapsychology intended for a general audience.)

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Oxford University Press.(Cartwright argues that laws in physics are neither as important for explanation nor as unified as has often been argued.)

Dupre, J. (1993) *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*, Cambridge, MA: Harvard University Press.(A defence of what the author refers to as ‘ontological pluralism’, including an extensive critique of reductionism and, more generally, of the view that physics has a special primacy among the sciences.)

Edge, H.L., Morris, R.L., Palmer, J. and Rush, J.H. (1986) *Foundations of Parapsycholgy: Exploring the Boundaries of Human Capability*, Boston, MA, London and Henley: Routledge & Kegan Paul.(The authors are all active researchers in parapsychology. The essays deal with historical, methodological and experimental

issues.)
Flew, A. (ed.) (1987) *Readings in the Philosophical Problems of Parapsychology*, Buffalo, NY: Prometheus Books.(A useful collection of classic and recent articles on parapsychology. The essay by Irving Thalberg, ‘Are Paranormal Events Non-Physical?’, is a useful discussion of parapsychology and physicalism.)
Kurtz, P. (ed.) (1985) *A Skeptic’s Handbook of Parapsychology*, Buffalo, NY: Prometheus Books.(As the title suggests, a collection of articles mostly sceptical in tone, though with a few essays by partisans of psi.)
Rhine, J.B. (1954) ‘The Science of Non-Physical Nature’, *Journal of Philosophy* 51: 801-10.(Rhine offers his reasons for believing that psi-phenomena cannot be understood within the framework of physics. Excerpts from this essay are reprinted in Flew (1984).)
Teller, P. (1984) ‘A Poor Man’s Guide to Supervenience and Determination’, *The Southern Journal of Philosophy*, Spindell conference supplement 22: 137-62.(A review of the literature on supervenience, which is an important concept in certain versions of non-reductive materialism. Teller goes on to argue that materialism is such a general thesis that it is virtually a necessary truth.)

Pareto principle

A social state is said to be Pareto-efficient when there is no feasible alternative to it in which at least one individual is better off while no individual is worse off. The Pareto principle tells us to move from Pareto-inefficient to Pareto-efficient states. Suppose a large basket of fruit is shared among a group in some way or another - one apple, two peaches, a dozen cherries each, for instance. If the fruit can be exchanged so that at least some people get more enjoyment from what they have, and no one gets less, the Pareto principle instructs us to do so; indeed, it instructs us to carry on exchanging until no more improvements of this kind are possible.

The Pareto principle gets its name from the Italian economist Vilfredo Pareto (1848-1923), who showed that when a competitive market reaches an equilibrium, the outcome is Pareto-efficient, or a 'Pareto-optimum' (Pareto 1906). Pareto believed that valid interpersonal comparisons of utility are impossible, and so rejected the utilitarian view that a society is efficient when it maximizes aggregate utility. But each person could judge whether they were better off in state S1 or state S2, and so Pareto concluded that if everyone preferred S2 social welfare must be greater in S2, whereas if some preferred S1 and some S2, no such judgment could be made.

The Pareto principle has a weaker and a stronger form. The weaker holds that we should choose S2 in preference to S1 when everybody judges that they are better off in S2; the stronger holds that we should choose S2 when nobody judges that S1 is better, and at least one person judges that S2 is better. The weaker version requires that everyone gain if the change is to be a Pareto improvement; the stronger version requires only that some should gain and nobody should lose.

For many economists the Pareto principle seems to be self-evidently rational. How can we settle for a state of affairs when there is a feasible alternative that improves the position of some but worsens that of no one? The problem is rather that, seen as a criterion of efficiency, the principle is still too weak. In the real world there are likely to be many Pareto-efficient states to choose between, but here the principle gives us no guidance. If in moving from S1 to S2 many badly-off people gain but a few well-off people lose, we might think there is a gain in overall efficiency. But we cannot make such a judgment using the original Pareto principle alone (see Economics and ethics §3).

Philosophers have been less certain that the principle is a requirement of practical rationality. Two kinds of questions have been raised. Suppose that S1 represents a state of affairs that is distributively just by some criterion - for instance, each person enjoys an equal level of welfare. If in S2 some people are better off than in S1 but the criterion of justice is violated - welfare levels are no longer equal - it may not be irrational to choose Pareto-inefficient S1 over Pareto-efficient S2.

Second, Sen (1982) has argued that the Pareto principle may conflict with the liberal principle that each person should decide matters falling in their private domain, such as which position they sleep in. If people have strong 'meddlesome preferences' - that is, they are deeply concerned about what goes on in other people's personal domains - there will be cases in which the Pareto principle selects a social state that the liberal principle would exclude. This issue has now generated a substantial body of literature suggesting different ways of avoiding Sen's paradox.

DAVID MILLER

References and further reading

Buchanan, A. (1985) *Ethics, Efficiency, and the Market*, Oxford: Clarendon Press, ch. 1.(A critical defence of the Paretian notion of efficiency.)

Cohen, G.A. (1995) 'The Pareto Argument for Inequality', *Social Philosophy and Policy* 12 (1): 160-85.(Attacks Rawls' use of the Pareto principle to justify economic inequalities.)

Little, I.M.D. (1957) *A Critique of Welfare Economics*, Oxford: Oxford University Press, 2nd edn, chaps 6-7.(A critical review of welfare economists' attempts to extend the Pareto principle.)

Pareto, V. (1906) *Manual of Political Economy*, trans. A.S. Schwier, London: Macmillan, 1972.(The original source of the Pareto principle.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, §§12-13.(Rawls invokes the

Paretian criterion of efficiency, and argues that the difference principle gives us a way of choosing between the many possible Pareto-efficient social distributions on the grounds of justice.)

Sen. A. (1982) *Choice, Welfare and Measurement*, Oxford: Blackwell, chaps 13-14.(‘The Impossibility of a Paretian Liberal’ contains the original statement of Sen’s argument. ‘Liberty, Unanimity and Rights’ develops it through a review of the literature provoked by the first article.)

Parmenides (early to mid 5th century BC)

Parmenides of Elea, a revolutionary and enigmatic Greek philosophical poet, was the earliest defender of Eleatic metaphysics. He argued for the essential homogeneity and changelessness of being, rejecting as spurious the world's apparent variation over space and time. His one poem, whose first half largely survives, opens with the allegory of an intellectual journey by which Parmenides has succeeded in standing back from the empirical world. He learns, from the mouth of an unnamed goddess, a dramatically new perspective on being. The goddess's disquisition, which fills the remainder of the poem, is divided into two parts; the Way of Truth and the Way of Seeming.

The Way of Truth is the earliest known passage of sustained argument in Western philosophy. First a purportedly exhaustive choice is offered between two 'paths' - that of being, and that of not-being. Next the not-being path is closed off: the predicate expression '... is not' could never be supplied with a subject, since only that-which-is can be spoken of and thought of. Nor, on pain of self-contradiction, can a third path be entertained, one which would conflate being with not-being - despite the fact that just such a path is implicit in the ordinary human acceptance of an empirical world bearing a variety of shifting predicates. All references, open or covert, to not-being must be outlawed. Only '... is' (or perhaps '... is... ') can be coherently said of anything.

The next move is to seek the characteristics of that-which-is. The total exclusion of not-being leaves us with something radically unlike the empirical world. It must lack generation, destruction, change, distinct parts, movement and an asymmetric shape, all of which would require some not-being to occur. That-which-is must, in short, be a changeless and undifferentiated sphere.

In the second part of the poem the goddess offers a cosmology - a physical explanation of the very world which the first half of the poem has banished as incoherent. This is based on a pair of ultimate principles or elements, the one light and fiery, the other heavy and dark. It is presented as conveying the 'opinions of mortals'. It is deceitful, but the goddess nevertheless recommends learning it, 'so that no opinion of mortals may outstrip you'.

The motive for the radical split between the two halves of the poem has been much debated in modern times. In antiquity the Way of Truth was taken by some as a challenge to the notion of change, which physics must answer, by others as the statement of a profound metaphysical truth, while the Way of Seeming was widely treated as in some sense Parmenides' own bona fide physical system.

1 Life and work

Parmenides lived and taught and, if the ancient tradition is reliable, framed legislation at Elea (modern Velia), a Greek city in southern Italy. His most eminent pupil was Zeno of Elea, author of the celebrated motion paradoxes. Plato in his *Parmenides* describes a visit by Parmenides and Zeno to Athens, usually thought to be around 450 BC, when he says Parmenides was aged about 65, Zeno about 40. He implies that Parmenides' work had been published a good deal earlier, since Zeno had defended it against its critics in his youth, perhaps around 470 BC. It could in fact have appeared as early as the 490s. In any case, the visit Plato describes is probably fictional, and the chronological details are open to suspicion (see Socratic dialogues §1). The only completely safe chronological assertion is that Parmenides wrote before Zeno, Anaxagoras and Empedocles.

About 150 lines survive from his poem, which was his sole published work. Most of these (around 107) belong to its first part, the proem plus the Way of Truth, which appears largely complete. The second part, the Way of Seeming, must originally have been at least as long. The numbered ordering of the fragments, established by editors, seems mainly reliable, but the original position of a few (especially fragments 3 and 5) remains controversial.

The language, which constantly exploits echoes of Homeric epic, is opaque and densely metaphorical. It will be impossible in what follows to do justice to many important questions of nuance. A further complication is that Parmenides is compelled to borrow for his arguments the very language of negation, change and differentiation which his conclusions will ultimately outlaw.

The Way of Truth is methodically argued. Although the opening describes Parmenides' journey to a goddess who

undertakes his enlightenment, this symbolizes less a religious revelation or appeal to mere authority than his own hard-won arrival at a god's-eye view, his intellectual odyssey of distancing himself from the familiar temporal world: he travels to the gates at which the paths of night and day meet, a mythically inspired vantage point which erases the alternation of night and day, together, we may suppose, with all the associated temporal and spatial distinctions. When the goddess proceeds to enlighten him, it is with a coordinated set of arguments. Her most favoured form of argument starts from the conclusion: *c*, because *b*, because *a*.

2 The three paths

An initial choice offered is between two paths. Later a third path, although not even represented as a formal possibility, has to be mentioned and blocked off because despite its incoherence it is the route which ordinary people actually try to take.

The first two paths are: '(It) is' (*esti*) and '(It) is not' (*ouk esti*). It turns out that you can only coherently say the former, and this in the later moves will radically restrict what can be truly said of that-which-is. But what is the status of this 'is'? In particular, does it have a subject and/or a predicate?

Unlike English 'is', its Greek equivalent *esti* can constitute a grammatically complete sentence, '(It) is', even when as here no explicit subject is supplied. At this initial stage, when we are still finding out what restrictions must be imposed on the character of whatever 'is', the failure to specify any subject term, even 'it' or 'something', looks deliberate. What, if anything, can stand as the subject of 'is' must not be in any way prejudged.

The Greek use of 'be' does not fall neatly into a complete or existential use ('*x* is') and an incomplete or copulative one ('*x* is *y*'). To be is, primarily, to be *something*, but this may be specified ('*x* is red', '*x* is a book') or left unspecified ('*x* is'). To a Greek ear these are not distinct senses; they also tend to shade into the so-called 'veridical' use, where 'being' means 'being the case', since the 'being' in question is viewed as once again copulative, *x*'s being *y*. It has long been disputed whether Parmenides' use of 'be' is existential, copulative, veridical or one 'fused' out of these. While some kind of fused usage looks likeliest, no confusion need arise from thinking of his poem as primarily an inquiry into what there is. At least, his description of that-which-is as ungenerated, continuous, immobile etc. suggests an object rather than (as on a narrowly veridical reading) a truth - indeed, an object which directly and successfully competes with the phenomenal world for the status of what-there-is.

There are few uncontroversial points of interpretation, but the opening argument can be paraphrased, with explanatory glosses, along the following lines.

There are two conceivable paths [that is, ways forward]*: to say '... is', and to say '... is not'* [these are actually presented as '… necessarily is' and '… necessarily is not'; contingency is assumed to involve an illicit conflation of being with not-being, and is therefore held over as a third path, see below]. *But saying '... is not' is not a way forward that you could ever actually explore* [that is, '… is not' is a predicate which can never be successfully supplied with a subject, by either (a) thinking of the subject or (b) naming it]. *This is because (a) you could not know that-which-is-not* [that is, you could not know what it is, since that-which-is-not is nothing; therefore you could not pick it out in thought as a subject], *(b) nor could you speak of it* [the nonexistent is not available for referring to]. *Besides, anything that you can speak of and think of* [that is, pick out as a subject term] *must be* [and is therefore automatically debarred from serving as subject of '… is not']. *This is because it is, at least, possible for it to be* [being conceivable, it at least *could* exist], *whereas for a 'nothing' to be is impossible* [there could never exist a non-existent thing]. *In short, '... is not' is a path that you could never travel* [an expression with no application] (frs 2, 6.1-3).

Likewise you can discount a third path, 'the one wandered by know-nothing two-headed mortals' [when they place their trust in the senses and accept a world of contingency, change and diversity]. *They 'consider being and not-being the same and not the same'* [since phenomenal things possess predicates only in certain respects, at certain times, etc., every case of being is also a case of not being; this is self-contradictory - see rule 2 in §3 - and/or entails the now outlawed not-being]. *It is a back-turning path. Do not out of habit follow it, relying on your senses, but judge by reason my refutation* [of it] (frs 6.4-9, 7).

3 The description of that-which-is

What follows is a journey down the one path that remains negotiable, the path of '… is' (fr. 8.1-49). Now that the inquiry has been focused exclusively on that-which-is, the goddess is ready to tell Parmenides *what* it is, and proceeds to enumerate its predicates. These, or perhaps the ensuing arguments for them, she calls the 'signposts' along the route.

On this path there are many signposts that it is [1] *unborn and unperishing*, [2] *a unique whole*, [3] *unshaken* [= unmoved], [4] *perfect/complete/balanced* [the text is disputed here, but §8 will tend to favour *atalanton*, 'balanced'] (fr. 8.2-4).

She immediately adds, '*Nor* was *it once, nor* will *it be, since it* is *now*, *one*, *continuous*.' It is controversial whether this is a fifth item on the list of signposts and, if so, where it is proved. It may be safer to regard it as her parenthetical justification of the preceding present tense '… it *is* unborn' etc., explaining that for an altogether undifferentiated entity a past and future cannot be distinguished from its present. It has been held that the move introduces the notion of timeless being. On another interpretation, 'now' signals that it still occupies time, but with the *passage* of time abolished - possibly foreshadowing the later notion of eternity. (On time, see also §7.)

Analogous to the question of time is that of space. Some would take the goddess's description of that-which-is to make it an altogether non-spatial entity, others a space occupier albeit without spatial distinctions. Is it an entity as innocent of spatio-temporal being as, say, the number 2, or something which occupies space and time but is nevertheless not subject to spatial and temporal distinctions? The following analysis will favour the latter, which not only is the more straightforward reading of the text but also allows the individual arguments to work better (especially those for predicates 3 and 4).

The favourite objection to a literal spatial reading has been that that-which-is will then prove to be a finite sphere, in which case it will have to be surrounded by vacuum or not-being, in contravention of Parmenides' own ban. However, this difficulty arises only given the starting assumption that space is infinite. That assumption became standard in Ionian philosophy (see Anaximander §2; Melissus §3), but is much less evident in the western Greek tradition to which Parmenides belongs: Empedocles, like Plato and Aristotle later, regarded the universe as a finite sphere bounded by the heaven, with no space, empty or otherwise, beyond it. The very notion of space as a self-subsistent entity with its own dimensions, slow to emerge in ancient thought, was probably unknown to Parmenides. Provided Parmenides' sphere is envisaged from within, like the familiar world with its apparently spherical sky, not from outside like an orange, the supposition of space beyond need not impose itself. Arguably, his spherical being is a radical redescription of our own spherical world, in which no feature beyond the shape itself survives.

At fragment 8.6-49 the goddess defends in turn each of the predicates of that-which-is. The arguments lean on two implicit rules.

Rule 1: no proposition is true if it implies that, for any *x*, '*x* is not' is, was or will be true. This draws directly on the preceding argument (see §2).

Rule 2: there are no half-truths (that is, no proposition is both true and false; no question can be answered 'Yes and no'). This may sound innocuous, but is used lethally to outlaw all qualified truths and ensure that nothing can possess a given predicate in one way (at one time, in one respect, etc.) but not in another, as is required by the variable world believed in by two-headed mortals. One might think that even in that world temporally or otherwise qualified propositions would obey rule 2, for example, that 'My egg is warm at 9 am' and 'My egg is not warm at 2 pm' could both be true without qualification. But Parmenides would insist that the unqualified proposition 'My egg is warm' is another which must obey the rule, and not come out true in a way, false in a way.

The arguments proceed as follows.

4 The denial of generation and destruction

It cannot [taken as a whole] have had a beginning, since (a) that would have been from previous not-being [rule 1], *and (b) there could have been nothing, prior to its generation, to make it come into being precisely when it did, rather than sooner* [a celebrated anticipation of the Principle of Sufficient Reason] (fr. 8.6-10).

Likewise it must be totally or not at all [rule 2; that is, it cannot be generated piecemeal]: *no additional bits can*

come into being, since that too would be from previous not-being [rule 1] (fr. 8.11-14).

Thus both generation and [by parity of reasoning] *destruction are excluded* [In both wholesale and piecemeal destruction something comes not to be, contravening rule 1. But there is no obvious analogue to (b) above, the Principle of Sufficient Reason: in that-which-is there might well, for all we know at this stage of the argument, be sufficient reason for destruction to occur later rather than sooner, for example, progressive decay.] (fr. 8.13-21).

5 Continuity

It is also indivisible [or perhaps 'undivided']. *This is because it is completely homogeneous, its continuity not disrupted by varying degrees of being* [This would entail corresponding degrees of not-being, contrary to rule 1. Since it cannot to any extent not-be at one point what it is at another, there is nothing true of any individual part to distinguish it from any other part. So talk of 'parts' becomes empty.] (fr. 8.22-5).

6 Immobility

'Motionless, in the limits of mighty bonds', (a) 'it is unstarting and unstopping' [it neither starts off nor comes to a halt], *because coming-to-be and perishing have already been refuted* [and starting is the coming-to-be of motion, stopping is the perishing of motion]. *And (b) it stays just where it is, held by Necessity 'in the bonds of a limit which imprisons it all round'* [that is, that-which-is, taken as a whole, does not move anywhere, because it entirely fills its own spatial boundary, leaving itself no room to move]. [And it must have this boundary] *'because it is not fitting for that-which-is to be incomplete'* [absence of a definite boundary being a form of incompleteness]. *This in turn is because 'it is not lacking: if it were, it would lack everything'* [and therefore not exist at all; by rule 2, it cannot be both lacking and not lacking] (fr. 8.26-33).

This argument is often read as one against change in general; but the language strongly suggests that motion is the target. (Nothing in it specifically excludes the perpetual rotational movement of a sphere, but would this be conceivable in an entity with no distinct parts (see §5)?) Its 'boundary' argument works better for motion than for other forms of change. Parmenides' general objection to change may be located not here but in his arguments against generation (see §4), where the denial of piecemeal generation could well include that of new properties.

7 Monism

So far the goddess has defended predicates 1, 2 and 3 (see §3). The next lines, fragment 8.34-41, are difficult and controversial. They appear to interrupt the sequence of arguments, with the proof of predicate 4 not beginning till line 42. Some regard them as somehow part of that proof, others as a summary of results so far, yet others as a digression against empiricism. A still likelier explanation is that before she can embark on her final proof, concerning the shape of that-which-is, the goddess must pause to establish formally its *singularity*. She has already demonstrated that it itself forms an undivided whole, but she must still show that neither (a) thought, nor (b) time, nor (c) the plural objects of the phenomenal world, can be anything over and above it. The following paraphrase suggests how.

(a) *Thinking is identical with that which prompts thought* [that is, its object, being]. '*For in what has been said* [that is, the preceding arguments] *you will not find thinking separate from being*' [It has been seen that thought and being always go together, since not-being is unthinkable. There are no grounds for distinguishing the thinking subject from the object thought. Thinking is being and being is thinking. See also fragment 3 (of uncertain location) - *'For to think and to be are the same'* - which has often been given other, somewhat strained, translations to avoid this admittedly very difficult notion.] (fr. 8.34-8).

(b) *'Neither is there, nor will there be, time* [for the reading of the text, see Coxon (1986)] *over and above being, since Fate has bound it down to be the whole* [hence spatially all-inclusive, so that there can be no external measure of time] *and unmoved* [so that there can be no internal measure of time].'

(c) *'Therefore* [since that-which-is is all-inclusive and motionless] *it* [that-which-is itself] *has been named all the things which mortals have posited, believing them to be real' and to undergo changes of all kinds* [That is, when ordinary people talk about empirical objects, properties etc. they are speaking, not (impossibly) of nothing, but of that-which-is, even if they are radically misdescribing it. So the fact that ordinary talk is about something does not mean that there is something over and above that-which-is for it to be about.] (fr. 8.38-41).

Whether or not posterity was right to proclaim Parmenides the champion of the One against the Many, he was undoubtedly a monist. But Monism as such was not new. All his major forerunners, from Thales to Heraclitus, had believed everything to be ultimately analysable as manifestations of one thing, namely a single underlying stuff. Parmenides was the first *eliminative* monist. Everything is still one, but in a way which, instead of accounting for plurality, eliminates it.

8 Symmetry

'But since there is an outermost limit [see §6]*, it is bounded on all sides, like the mass of a well-rounded ball, being equally balanced on all sides from the centre* [that is, it is spherical in shape; many interpreters make it resemble a sphere in some other way than shape, for example, in perfection or uniformity, which risks underplaying this geometrically precise description, and makes the grounds which follow less apposite]. *For it must not be bigger or smaller here than there* [it cannot be asymmetrical, as it would have to be if not a sphere]. *For there is no not-being to prevent it reaching the same distance'* [if it were cut off short in one direction, that would mean its not-being beyond that point, in contravention of rule 1], *nor are there degrees of being* [it cannot be asymmetrical by thinning out in places], *because it is, as a whole, 'immune'* [to depletion], *having precisely equal being throughout, up to its limits* (fr. 8.42-9).

This ends the Way of Truth. Can the goddess really be telling Parmenides that that-which-is is literally spherical, without jeopardizing its partlessness? In a sphere you can distinguish hemispheres, segments and other parts. True, but at least the sphere is the one solid that you *can* think of as a whole without distinguishing parts (contrast, for example, a cube, which must be thought of with eight corners etc.). And the upshot of the poem so far, with its attack on human perspectives, has been to persuade Parmenides not to try to enforce any such distinctions. See also fragment 4 (of uncertain location): *'Gaze in thought equally upon absent things as firmly present. For thought will not split off that-which-is from clinging to that-which-is, scattered or gathered everyhow everywhere in the world.'*

9 The Way of Seeming

The goddess now turns to the 'opinions of mortals', and sets out, unargued, an analysis of the phenomenal world in terms of the combination of two opposite 'forms' or elements, one light and fiery, the other heavy and dark. The surviving fragments are scanty, but the cosmology includes the following: a role for a creative goddess, a detailed description of the heavens as a set of concentric bands, an embryology and a physiology of human thought.

The Way of Seeming has been promised by the goddess from the outset: *'You must find out everything - both the unshaken heart of well-rounded truth, and the opinions of mortals. In them there is no true trust, but you must learn them too anyway…'* (fr. 1.28-31). At its opening she calls it 'deceitful', if also 'plausible'. She adds that Parmenides must learn it *'so that no opinion of mortals may outstrip [or overtake] you'* (fr. 8.61), apparently meaning that the cosmology will be the best of its kind, able to compete successfully with those already offered by others. And there can be no doubt that this part of the poem is, for whatever reason, making an original contribution to cosmology. It even incorporates two major astronomical discoveries: that the Morning Star is identical to the Evening Star, and that the moon's light comes from the sun. But if all cosmology is false, what is the point of even entering such a contest?

Where Parmenides' major predecessors had been material monists, Parmenides offers his own distinctively dualist scheme. So his analogous move from one entity in the Way of Truth to two in the Way of Seeming is anything but casual. Indeed, the arithmetical point is precisely what the goddess emphasizes: mortals *'have made up their minds to name* two *forms, of which they should not name* one' (fr. 8.53-4). This, although sometimes taken to mean that they should not name *so much as* one, or even not *merely* one, is most easily read as saying that precisely one out of the two is a mistake. That itself, if correct, might be understood as purely numerical: any total of two is one too many. Or it can be taken specifically: one of these two forms is identical with that-which-is, but the other should not have been added - in which case it becomes tempting (and Aristotle among others was tempted) to assimilate the illicit second element to what in the Way of Truth was called not-being. As to which element is which, it is usually the fiery element that is assumed to correspond to that-which-is, although Popper (1992) has advocated the dark heavy one, in which the fiery one produces the mere semblance of change, just as the moon, in reality solidly spherical, was known by Parmenides to appear to wax and wane through the play of light on it.

The question remains how the cosmology relates to the Way of Truth. Some have held that it itself contains a measure of truth, others on the contrary that it is there to expose the patent falsity of any departure from strict monism. Of these, the former sits ill with the goddess's contrast between mortal opinions and truth, while the latter conflicts with the fact that the cosmology, rather than being *patently* false, is highly 'plausible'. A more attractive possibility is that she wants to show how surprisingly similar the deceptive empirical world is to the stark Way of Truth. The whole range of cosmic phenomena can be generated by allowing the intrusion of just one additional item - by starting out with two instead of one. This hypothesis makes good sense of the often noticed fact that the cosmological descriptions in the Way of Seeming pointedly echo the language of the Way of Truth. For example, the description of the *'encircling heaven'* as *'bound down by Necessity to hold the limits of the stars'* (fr. 10) recalls that of that-which-is as held motionless by Necessity in the bonds of a limit (see §6).

Thus interpreted, the Way of Seeming does not pretend to vindicate phenomena, but it does tackle the biggest headache faced by any convert of Parmenides: how can human experience have got things quite so catastrophically wrong? The answer is that the step from appearance to reality is much smaller than it may initially look.

This admittedly still does not begin to explain how the error of mortals ever occurred. If Parmenides is right, there are no separate thinking subjects. All thought is that-which-is thinking itself. How it could find room to misconceive itself is a question on which Parmenides leaves us to puzzle.

10 Influence

Parmenides marks a watershed in Presocratic philosophy. In the next generation he remained the senior voice of Eleaticism, perceived as champion of the One against the Many. His One was defended by Zeno of Elea and Melissus, while those who wished to vindicate cosmic plurality and change felt obliged to respond to his challenge. Empedocles, Anaxagoras (§2), Leucippus and Democritus framed their theories in terms which conceded as much as possible to his rejections of literal generation and annihilation and of division. On the other hand, in this they were swayed less by Parmenides' own primary argument, the inadmissibility of '… is not', which won few converts, than by an older and less controversial premise, the impossibility of generation *ex nihilo*.

Plato (§16), in his *Sophist*, did take up puzzles raised by Parmenides' rejection of not-being. But it was in recognition of Parmenides' rigorous methodology, not his doctrines, that in his later work Plato expressed a degree of allegiance to the Eleatic tradition, letting Parmenides or an unnamed follower of his eclipse Socrates as principal speaker in three major dialogues.

Plato's own declared veneration of Parmenides is exceeded by that of the Neoplatonists, who treat his 'one being' as an integral part of their own metaphysical hierarchy. Among his modern admirers, perhaps the most prominent is Heidegger, for whom Parmenides was the last great thinker before the 'forgetting of being' which has blighted the subsequent history of metaphysics.

See also: Being; Melissus; Presocratic philosophy

DAVID SEDLEY

References and further reading

There are good English translations of all or most of Parmenides in Austin (1986), Barnes (1979, 1986), Coxon (1986), Gallop (1984), Guthrie (1962-78), Hussey (1972) and Kirk, Raven and Schofield (1983). With Parmenides more than most writers, any translation is an interpretation, and it is prudent always to compare two or more.

Austin, S. (1986) *Parmenides. Being, Bounds, and Logic*, New Haven, CT: Yale University Press.(In-depth study of Parmenides' logic.)

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge & Kegan Paul, chaps 9-11.(An unconventional but exhilarating analysis)

Barnes, J. (1987) *Early Greek Philosophy*, Harmondsworth: Penguin.(Chapter 9 is a translation of Parmenides' fragments, complete with the contexts in which they are preserved.)

Coxon, A.H. (1986) *The Fragments of Parmenides*, Assen: Van Gorcum.(An important edition, informative on the text and secondary tradition)

Furley, D. (1973) 'Notes on Parmenides', in E.N. Lee, A.P.D. Mourelatos and R. Rorty (eds) *Exegesis and Argument*, Assen: Van Gorcum, 1-15; repr. in D. Furley, *Cosmic Problems*, Cambridge: Cambridge University

Press, 1989, 27-37.(Incisive commentary on a series of crucial passages.)
Furth, M. (1968) 'Elements of Eleatic Ontology', *Journal of the History of Philosophy* 6: 111-32; repr. in A.P.D. Mourelatos, *The Pre-Socratics*, Garden City, NY: Doubleday, 1989, 241-70.(Classic exposition of the 'fused' sense of 'be' in Parmenides.)
Gallop, D. (1984) *Parmenides of Elea*, Toronto, Ont.: University of Toronto Press.(Contains text, translation and notes; a basic and clear exposition, founded principally on the interpretation of Owen (1960).)
Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols.(The most detailed and comprehensive English-language history of early Greek thought; the treatment of Parmenides, in volume 2 pages 1-80, is a helpful guide to the scholarship.)
Heidegger, M. (1954) 'Moira (Parmenides VIII, 34-41)', in *Early Greek Thinking*, San Francisco, CA: Harper & Row, 1975, 79-101.(Heidegger's interpretation and appreciation of Parmenides on the relation of thought to being, in English translation.)
Hussey, E. (1972) *The Presocratics*, London: Duckworth.(Chapter 5 contains a lucid short account of Parmenides.)
Kahn, C.H. (1969) 'The Thesis of Parmenides', *Review of Metaphysics* 22: 700-24.(Leading exposition of the 'veridical' sense of 'be' in Parmenides.)
Kahn, C.H. (1988) ' Being In Parmenides and Plato', *La Parola del Passato* 43: 237-61.(Further refines Kahn's 1969 interpretation.)
Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(A valuable survey of Presocratic philosophy, including texts and translations; Parmenides occupies chapter 8.)
Long, A.A. (1963) 'The Principles Of Parmenides' Cosmogony', *Phronesis* 8: 90-107; repr. in R.E. Allen and D.J. Furley (eds), *Studies in Presocratic Philosophy*, London: Routledge & Kegan Paul, 1975, vol. 2, 82-101. (Important study of the Way of Seeming.)
Mackenzie, M.M. (1982) 'Parmenides' Dilemma', *Phronesis* 27: 1-12.(One of the most philosophically interesting discussions of Parmenides.)
Mourelatos, A.P.D. (1970) *The Route of Parmenides*, New Haven, CT: Yale University Press. (A subtle overall reading of Parmenides, sensitive to his place in the poetic as well as the philosophical tradition.)
Owen, G.E.L. (1960) 'Eleatic Questions', *Classical Quarterly* 10: 84-102; repr. in R.E. Allen and D.J. Furley (eds), *Studies in Presocratic Philosophy*, London: Routledge & Kegan Paul, 1975, vol. 2, 48-81; repr. in G.E.L. Owen, *Logic, Science and Dialectic*, London: Duckworth, 1986, 3-26.(The seminal modern study of Parmenides as philosopher.)
Parmenides (early to mid 5th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker*, Berlin: Weidemann, 6th edn, 1952, vol. 1, 217-46.(The standard collection of the ancient sources; includes Greek texts with translations in German.)
Popper, K.R. (1992) 'How the Moon Might Shed Some of her Light Upon the Two Ways of Parmenides', *Classical Quarterly* 42: 12-19.(Radical proposal concerning the relation between the two halves of the poem.)
Reinhardt, K. (1916) 'The Relation Between the Two Parts of Parmenides' Poem', trans. in A.P.D. Mourelatos, *The Pre-Socratics*, Garden City, NY: Doubleday, 293-311.(Classic study of the problem.)
Seidel, G.S. (1964) *Martin Heidegger and the Pre-Socratics*, Lincoln, NE: University of Nebraska Press.(Explains the fundamental importance of Parmenides to Heidegger's conception of metaphysics.)
Sorabji, R. (1983) *Time, Creation and the Continuum*, London: Duckworth.(Chapter 8 deals with Parmenides' treatment of time.)
Stokes, M.C. (1971) *One and Many in Presocratic Philosophy*, Washington, DC: Center for Hellenic Studies. (Locates Parmenides in a historical context.)

Particulars

Particulars are to be understood by contrasting them with universals, that term being used to comprise both properties and relations. Often the term 'individuals' is used interchangeably with 'particulars', though some restrict the term 'individuals' to those particulars whose existence has more than momentary duration.

It is sometimes taken as a distinctive feature of particulars that they cannot be in more than one place at a time, whereas universals are capable of being wholly present in more than one place at a given time: if you have a white thing here and a white thing there, then you have two particulars but only one property. This way of distinguishing between particulars and universals may help us to focus on apt paradigm cases of each, but arguably this does not get us to the heart of the matter. On the one hand, some think it is possible, at least in principle, for a magician, or Pythagoras, or a time traveller, or a subatomic particle to be in two places at once, even though each is a particular. On the other hand, some think that there are properties which could not possibly be manifested in two different places at the same time, and yet which nonetheless are universals: think, for instance, of the divine property of absolute perfection, or of the conjunction of all intrinsic properties of a Leibnizian monad (or possible world); or of Judas' property of simply being Judas.

Particulars are things which have properties and which stand in relations - particulars 'instantiate' properties and relations. By itself, however, this does not distinguish particulars from universals since universals, too, are naturally thought to have properties and to stand in relations. What distinguishes particulars is the fact that, while a particular instantiates properties and relations, nothing instantiates a particular. Universals both 'have' (properties and relations) and are 'had'; particulars 'have' but are not 'had'. Since a particular is not instantiated by another thing, it is sometimes said to exist 'in itself', whereas a universal exists 'in' something else. For this reason, the term 'particular' is related to the term 'substance', which is traditionally used to mean something capable of independent existence.

1 What is at stake

In some of his dialogues, Plato seems to assume that nothing is real, knowable or of value unless it is permanent (see Plato §15). The particular things we experience through the senses are all impermanent; so Plato infers that these particulars are unreal, unknowable and of no value. The things which are real, knowable and of value are the moral, political, aesthetic and mathematical ideals which individual people, social groups and works of art may aspire towards but never really embody perfectly - the Forms.

In an ancient Buddhist text, the existence of common-sense particulars is called into question (Horner 1963: 34-8; Eliot 1910: 668-72). Suppose, for example, that you reach out and touch something, saying 'This is a chariot'. The ancient sage calls attention to the fact that what you are touching is, more strictly, a wheel, and more strictly still, a hub, spoke or rim, and of course only the surface of that. So there is really no such thing as a chariot; and the same goes for other particulars, including persons. A person is like a flame, or an eddy in a river - there is nothing which persists through the passage of time. Particulars are only shadows cast by quirks of Indo-European syntax, with no more substantiality than the 'It' in 'It is raining'.

In contrast, for Descartes there is nothing more certain than the existence of the self: his philosophy rests on the famous inference 'I think, therefore I am'. Western liberal democracies place great store, at least in theory, on the importance of the individual. In the arts, a conception of individual artistic genius has held us in a tight grip. And modern mathematical logic rests heavily on names and variables which are interpreted as picking out individuals from a domain of discourse. In a great many ways, individuals, or particulars, are central to our thinking, not only about this world of impermanence but also about the timeless truths of mathematics. Many philosophers are therefore diametrically opposed to the Platonic or Buddhist deflation of particulars.

2 Bare particulars

Aristotle agreed with Plato that knowledge - science - is concerned with things which are necessary and hence permanent. Yet for Aristotle this did not exclude knowledge of particulars. Admittedly, each particular is subject to generation and corruption; nonetheless, each particular does have certain essential properties in virtue of which it belongs to a species and genus. And although individuals are transient, species are permanent. Individuals can be

subjects of knowledge, thanks to their possession of essential properties (see Aristotle §§14-15).

Aristotelian essentialism is often disavowed by philosophers but, arguably, essentialist assumptions are pervasive in our thinking at both a common-sense and a philosophical level. The essentialist holds that each particular has certain properties which it cannot lose except by ceasing to exist, and which it could not have lacked from the start except by never having come into being. The contrary doctrine is sometimes called the doctrine of 'bare particulars' - that particulars do not have essential properties. Although essentialism is regularly disparaged by philosophers, so is the contrary doctrine of bare particulars. The rejection of both together would require a radical elimination of the very idea of any but merely verbal necessities or possibilities, as in Quine (1953).

3 Bundles of properties

Consider a particular which has several essential properties and several accidental properties. There is room to consider a theory which identifies the particular with the collection of essential properties, and there is room to consider a theory which identifies the particular with the collection of all its properties. On consideration, however, it can be seen that a particular cannot be identified with any collection which includes an accidental property. An accidental property is something which, by definition, the particular could have lacked. But the bundle of properties which includes that accidental property would not have been the bundle that it is if it had not included that property.

Consider, therefore, the theory which identifies a particular with the collection of its essential properties. An objection can be raised if we show that there can be several particulars which share exactly the same essential properties. Certainly Aristotle thought that different members of the same species share the very same essential properties. For Aristotle, therefore, what distinguishes particulars of the same species is not their 'form', but the 'matter' on which this form is impressed.

Hence if we are to identify a particular with a collection of essential properties, we must suppose that distinct particulars cannot ever share exactly the same essential properties. Each particular must have some property which no other particular could have. This is what is called an individual essence. One way of thinking of an individual essence is by construing it to be an unshareable conjunction of severally-shareable properties as, for example, Leibniz's monads. An alternative would be to suppose that for each particular there is a simple essential property called a 'haecceity' (from Latin, meaning 'thisness') which that particular alone possesses and which nothing else could have had either instead of or as well as it. This notion was articulated by the Islamic philosopher Avicenna (see Ibn Sina §4), and was taken up by some of the scholastic philosophers in Europe, such as Duns Scotus in the thirteenth century, but has been roundly denounced by many since that time.

4 Properties particularized

There is thus a line of argument which leads from the stereotypically empiricist idea that a thing is a bundle of properties to the distinctly non-empiricist doctrine of haecceities or individual essences. An empiricist can avoid inconsistency by particularizing properties.

For the sake of illustration, suppose coldness to be a typical property of a thing. According to some philosophers, then, the coldness of one block of ice is one thing, and the coldness of another block of ice is quite another thing. It is not just that the second block of ice is a second instance of a cold thing but, rather, the coldness of the second block of ice is a second instance of coldness. Whenever something is cold, there exists something which we may call its coldness: this may be called a 'property instance' or (following D.C. Williams) a 'trope'. According to the trope theory, a particular can be taken to be a 'bundle of properties' provided we take each of these properties to be an abstract particular.

The trope theory is a contender for, but has not attracted, majority support. Some of the Scholastic nominalists, like Ockham, can be read as having articulated such a view, and the view has been revived by a series of philosophers in the twentieth century: a classic exploration is found in Nelson Goodman's *The Structure of Appearance* (1951) (see Nominalism §§2, 4).

5 Particulars as aggregates

Emergentism is the doctrine that something new emerges or comes into being when materials come together and

take the form of something with a kind of unity and persistence across time, as for instance a living being. Such a thing is constituted by a plurality of parts, but is something distinct from just the aggregate of those parts.

The distinction between a thing and the aggregate of its parts can be drawn by reference to their logical independence: you could have had either one without the other. Consider a particular ship, say the ship of Theseus. This ship is constituted by planks of wood. But the planks could have existed without ever constituting a ship; and the ship could have been constituted by different planks. Hence the ship is not the same thing as the aggregate of the planks which constitute it.

This conception of an individual as something other than just the aggregate of its parts has been bedevilled by a puzzle, famously described by Hobbes (Molesworth 1837-45 (4): 135; the example traces back to Plutarch's 'Life of Theseus' §§22-3). Suppose that, over time, worn planks on the ship of Theseus are replaced by new ones, and a working ship is maintained in continuous existence; meanwhile, the worn planks which are taken away are gradually reassembled in just the way they were arranged in the original ship. One principle leads us to say that the working ship with new planks is the ship of Theseus; another principle leads us to say that the reassembled ship with worn planks is the ship of Theseus. Yet an individual ship cannot be in two places at once: which of them is the ship of Theseus? This ship puzzle can be paralleled by puzzles of personal identity across time (see Personal identity).

Objections of this sort have been raised against the theory that there are particulars which have parts, but which are not just the aggregates of those parts (the contrary theory takes all particulars to be just the aggregates of their parts). Goodman and Quine (1947) have been very influential in promoting the idea that particulars are best understood through the theory of the part-whole relation, or mereology (see Mereology). Furthermore, according to Goodman, Quine and others, an individual which persists through some interval of time has distinct parts existing at each distinct time in that interval. Among the things which exist at one time is a timeslice which exists only at that time; among the things which exist at another time is a different timeslice; the thing which exists across several distinct times is just an aggregate of distinct timeslices.

Hobbes' puzzle question, 'Which is the ship of Theseus?', is then diagnosed as turning upon a merely verbal vagueness. There are two distinct aggregates which are broadly ship-shape, and it is mere semantic indeterminacy which makes us unsure as to which of them should be called 'the ship of Theseus'. Similar reasoning applies equally to persons.

6 The state of the art

Strawson (1959) draws a distinction between 'revisionary' and 'descriptive' metaphysics. In revisionary metaphysics, a theory is advanced to replace our ordinary language and thinking. In descriptive metaphysics, a description is given of the deepest presuppositions of our existing theories. Using these criteria, the broadly Aristotelian theory does seem to qualify as a relatively descriptive theory, by contrast with Plato's more revisionary claims that particulars are unreal, unknowable and of no value. The Aristotelian theory is also more descriptive than theories which deny that particulars have essential properties, theories which take particulars to be bundles of tropes and theories which treat all particulars as merely the aggregates of the spatial and temporal parts out of which they are constituted. The more revisionary theories may be right; but the more conservative theories still have a lot of life left in them. The time is not ripe for a consensus to emerge on the nature of particulars.

See also: Identity; Substance; Universals

JOHN C. BIGELOW

References and further reading

Adams, R.H. (1979) 'Primitive Thisness and Primitive Identity', *Journal of Philosophy* 76: 5-26.(A difficult but seminal article on particulars as irreducible to shareable properties.)

Armstrong, D.M. (1980) 'Identity Through Time', in P. van Inwagen (ed.) *Time and Cause*, Dordrecht: Reidel, 67-78.(An accessible defence of particulars as aggregates of timeslices, as discussed above in §5.)

Armstrong, D.M. (1989) *Universals: An Opinionated Introduction*, Boulder, CO, San Francisco, CA and London: Westview Press.(An introductory exposition of a late twentieth century resuscitation of a roughly Aristotelian theory distinguishing particulars from universals.)

Campbell, K. (1989) *Abstract Particulars*, Oxford: Blackwell.(An accessible defence of properties particularized,

as discussed in above in §4. Useful for further references.)
Eliot, C.W. (ed.) (1910) *Sacred Writings*, vol. 2, in *The Harvard Classics*, New York: Collier, vol. 45.(Contains source material translated from the Milindapañha and from the Visuddhimagga, mentioned in §1 above.)
Goodman, N. (1951) *The Structure of Appearance*, Cambridge, MA: Harvard University Press.(A difficult but important landmark on particulars as bundles of properties or as aggregates of timeslices, and of properties particularized, and of other ways of construing particulars.)
Goodman, N. and Quine, W.V. (1947) 'Steps Toward a Constructive Nominalism', *Journal of Symbolic Logic* 12: 105-22.(A difficult but important work defending nominalism, and also laying the groundwork for mereology or the theory of aggregates, relevant to §5 above.)
Horner, I.B. (ed. and trans.) (1963) *Milinda's Questions*, *Sacred Texts of the Buddhists*, Oxford, Pali Text Society, vol. 22.(Source material on the Buddhists' treatment of persisting individuals as mere aggregates of momentary things in 'Questions of King Milinda and Nāgasena' mentioned above in §1.)
Lewis, D.K. (1976) 'Survival and Identity', in A.O. Rorty (ed.) *The Identities of Persons*, Berkeley, CA: University of California Press, 17-40.(A difficult but clear exposition of a theory of persons as aggregates of timeslices.)
Molesworth, W. (ed.) (1837-45) *Elements of Philosophy*, vol. 4 of *The English Works of Thomas Hobbes of Malmesbury*, London: J. Bohn.(Part 2, chapter 2: 135 contains a famous passage on the ship of Theseus, mentioned in §5 above.)
Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press.(An accessible and entertaining exposition of a theory of persons as aggregates of timeslices, and of supposed consequences that this might have for moral theory.)
Quine, W.V. (1953) 'Three Grades of Modal Involvement', *Proceedings of the XI International Congress of Philosophy* 14: 65-81; repr. in *The Ways of Paradox and Other Essays*, New York: Random House, 1966, 155-74.(A classic attack on Aristotelian or any other sort of essentialism.)
Strawson, P.F. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen.(A transcendental justification of a roughly Kantian sort, for a common-sense theory of individuals as irreducible to properties.)
Williams, D.C. (1966) *The Principles of Empirical Realism*, Springfield, IL: Charles Thomas.(A classic defence of a theory of properties particularized.)

Partiinost’

Partiinost’ *(Russian for partyness, often translated as party-mindedness, partisanship or party spirit) was long the controlling principle of Soviet Marxism. Though commonly identified with thought control,* partiinost’ *originally signified social analysis of thought joined with moral judgment, an ancient combination that can work against the powers that be as well as for them. Lenin’s version changed from revolt to thought control after his party came to power in 1917, but especially after Stalin’s ‘revolution from above’ twelve years later. In 1950 Stalin began a restriction of* partiinost’ *by declaring ‘science’ separate from ‘ideology’. Such reform accelerated after his death in 1953, but slowed down from the mid-1960s to the late 1980s. Then a new burst of reform set off the collapse of the Soviet system and of* partiinost’*, though the problems that engendered it - the entanglement of group interests and claims of truth - persist.*

The self-serving nature of group beliefs has often been exposed. Xenophanes on divinity (‘If oxen could paint, their gods would be oxen’), Thrasymachus on justice (‘the interest of the stronger’) and Thomas More on ‘a conspiracy of the rich, who pursue their own aggrandizement under the name of the commonwealth’, are famous examples. Hobbes distinguished between ideas that involve interests - such as justice - and ideas that ‘cross no man’s ambition, profit, or lust’, such as the sum of angles in a triangle. He added however: ‘If [that sum] had been… contrary to the interest of men that have dominion… it would have been, if not disputed, yet by the burning of all books of geometry, suppressed, as far as he whom it concerned was able.’ Thus his sociology of ideas included a bellicose determinism: a ruling party is compelled to suppress thought that threatens its interests. He had seen parties or ‘factions’ bring civil war; the only remedy was for the sovereign to suppress them.

‘Party’ has not always meant factiousness. Shakespeare imagined the War of the Roses brought to an end by those who ‘dare maintain the party of the truth’, and the philosophers of the Enlightenment have been praised as ‘the party of humanity’. The emergence of modern parties, as special-interest groups engaged in routine contests for power, made identification of a party with the interests of humanity seem threadbare rhetoric, hardly concealing the sordid realities of partisan self-seeking. That may be one reason why *Das Kommunistische Manifest (The Communist Manifesto)* of 1848 declared that ‘the Communists do not form a separate party opposed to other working-class parties’.

Marx and Engels were political isolates without a significant party, not only in 1848 but through most of their lives, in part because they admired the scholar who seeks all-human truth rather than the ‘party man’ who serves special interests. Belief in disinterested inquiry mingles in Marx’s writing with insistence that ideology is unavoidable when class interests are entangled with claims of reason. Only the proletarian overthrow of class rule will bring the end of class bias in ideas. In that march towards truth the manifesto of 1848 says that Communists do not ‘form a separate party’, but they are ‘the most advanced and resolute section… of the proletariat, [who] have over the great mass… the advantage of clearly understanding the line of march, the conditions and the ultimate general results of the proletarian movement’. Thus Marx in 1848 laid out a rationalization of *partiinost’* for Communist Party-states of the twentieth century.

But that *is* a rationalization, which ignores contrary elements in Marx’s thought and in its uses, both for liberal pluralism by Marxists in Germany and for one-party tyranny by Marxists in Russia or in China and other ‘underdeveloped’ places. Liberal pluralism has emerged mostly in ‘developed’ lands, often through protracted experience of civil war, while Hobbesian choices between civil war and autocracy have attended the spread of party politics to places perceived as ‘backward’, in acute need of drastic ‘development’.

In nineteenth-century Russia talk of *grazhdanstvennost’*, citizenship, as something to be created through organized struggle, foreshadowed *partiinost’*. Chernyshevskii opened ‘Antropologicheskii printsip v filosofii’ (The Anthropological Principle in Philosophy) (1860) with a declaration that ‘every philosopher has been a representative of one of the political parties struggling in his time for predominance over the society to which the philosopher has belonged’. Even P.L. Lavrov, the Russian radical who exalted the ‘critically thinking individual’, enjoined such a thinker to serve ‘the people’ through a party, ‘to devote his energies to this party and to be guided by its advice’. ‘Party’ here signified commitment to a dream of an ideal system, which provoked splits as readily as organizations among fellow dreamers. Party discipline, often preached, was an additional impetus to splitting.

Trotsky's famous warning against Lenin's version of such discipline - that it would replace the working class by the Party, the Party by the Central Committee, the Central Committee by the dictator - could come true only with the accession to state power (see Lenin, V.; Trotsky, L.).

In 1895, when Lenin first used the word *partiinost'*, Russia was an autocracy without political parties. He was criticizing a fellow Marxist for showing the inevitability of capitalism but not its evil, though objectivity requires condemnation along with explanation. 'Materialism includes, so to speak, *partiinost'*, obliging one in any judgment of an occurrence to take directly and openly the viewpoint of a definite social group'. Like Chernyshevskii and Lavrov, Lenin was pointing to a fellowship of belief rather than a concrete political organization, and that was still his usage in *Materializm i ėmpiriokrititsizm (Materialism and Empiriocriticism)* (1909). By that time a Marxist party had been formed, and had split into Bolsheviks and Mensheviks, but Lenin's book assigned all philosophers to two 'parties' of a much broader kind: idealists and materialists, who had been serving the interests of rulers and ruled, of faith and science, since ancient times.

With that crude class analysis of philosophical schools Lenin justified the expulsion from the Bolshevik organization of comrades who rejected the dialectical materialism that he had learned from Plekhanov, a Menshevik in party affiliation. Lunacharskii, one of the expelled Bolsheviks, challenged Lenin to acknowledge his break with a rule laid down by Bebel, a German Marxist leader:

> 'We have no dogma and therefore cannot have heretics'. Let us grant that Bebel was mistaken, that both heresy and its condemnation are a possible phenomenon in Social Democracy. But don't we have the right to expect that in our Party free thought will have at least the same guarantees as are given to it by the Catholic Church, which prides itself on its intolerance?
>
> (Joravsky 1961: 39)

Lenin ignored that demand for precision and due process; *partiinost'* was casually invoked, never carefully analysed, even after 1917, when his party won power and suppressed rival parties.

At first the Central Committee limited its efforts to control intellectual life. Its ideological bureaucracy correlated ideas with the interests of social classes rather than parties in the literal sense, and did not try to enforce strict uniformity on thinkers. Such laxity was assailed in a 'discussion' following Stalin's complaint, in 1929, that 'theorists' were not adequately serving 'practicians', whose 'chief' (*vozhd'*) he was. Henceforth he would be 'chief' of 'theorists' as well, philosophers included. 'The basic lesson of the philosophical discussion', as an ideological official (Kol'man) explained to mathematicians, was to end 'all efforts of any theory, of any scholarly discipline, to conceive itself as an autonomous, independent discipline… isolated from Party guidance'.

Philosophy became commentary on speeches of Stalin or decrees of the Central Committee, while ideological bureaucrats complained that the commentary was timid and unoriginal, and that it was subversive to limit *partiinost'* by attempting careful distinctions between science and ideology. The 'primacy of practice' required philosophers to dig up quotes from 'the classics of Marxism-Leninism' showing that the changing intuitions of Party leaders were justified by theory as well as practice. In short, an anti-intellectual, authoritarian pragmatism was the heart of Stalinist *partiinost'*. One of its effects was an ironic inversion of the social analysis of thought compacted with moral judgment, which had once been the meaningful core of the concept. Ferocious attacks on individualistic thinkers regenerated admiration of the type, while servants of 'the people's party' came to be regarded as immoral lickspittles, saying nothing of substance.

The stultifying effects on intellectual life, including the creative expertise demanded by practical leaders, brought Stalin to call a retreat. In 1950 he announced that 'no science [*nauka*] can develop and prosper without the clash of opinions, without freedom of criticism', a 'generally recognized rule that has been ignored and violated in the worst way'. Of course he blamed 'petty despots' in charge of particular disciplines, not the Central Committee or the concept of *partiinost'*, but he opened the way to restriction of the concept by declaring that 'science' or 'scholarship' (*nauka*) was not part of ideology. Efforts to spell out such restriction were feeble until Khrushchev began 'de-Stalinization' after Stalin's death in 1953. Even then Soviet philosophers were much more timid than their Polish and Yugoslav comrades. Kolakowski was especially forceful in showing the absurdities of 'institutional Marxism', which insisted that the intuitions of Party chiefs were infallible products of a scientific ideology.

The ouster of Khrushchev in 1964 checked further debate until the late 1980s, when another reformist party chief, Gorbachev, began a campaign for ‘openness’ and ‘restructuring’. That led so swiftly to systemic collapse that serious discussion of *partiinost’* hardly began. It became instantly extinct, remembered only to be denounced or renounced by the minority of Russian thinkers still sympathetic to Marxism in some form. Serious inquiry in the sociology of knowledge, including the complex interdependence of group interests and claims of truth, is barely beginning to address the record of *partiinost’*, whether in Russia or in other Communist countries.

See also: Ideology; Marxist Philosophy, Russian and Soviet

DAVID JORAVSKY

References and further reading

Chernyshevskii, N.G. (1860) ’Antropologicheskii printsip v filosofii’ (The Anthropological Principle in Philosophy), Sovremennik.(Major article by the founding father of Russian populism. Republished many times, for instance, in *Polnoe sobranie sochinenii*, vol. 7.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.(See especially chapters 11 and 22.)

Joravsky, D. (1961) *Soviet Marxism and Natural Science 1917-1932*, New York: Columbia University Press and London: Routledge.(See especially ’Lenin and the Partyness of Philosophy’ and ’The Great Break, 1929-1932’.)

Joravsky, D. (1983) ‘The Stalinist Mentality and the Higher Learning’, in *Slavic Review*, 575-600.(The bizarre patterns of *partiinost’* in various scholarly disciplines.)

Kolakowski, L. (1964) *Der Mensch ohne Alternative: von der Möglichkeit und Unmöglichkeit Marxist zu sein (Man Without Alternative: On the Possibility and the Impossibilty of Being Marxist)*, Munich: Piper.(Exposé of the absurdities in *partiinost’*, by the eminent Polish philosopher.)

Kol’man, E. (1931) *Na bor’bu za materialisticheskuiu dialektiku v matematike (In Struggle for Materialist Dialectics in Mathematics)*, Moscow: Gos. nauch.-tekh. izd.(An ideological official on *partiinost’*. At the end of a long life he ‘repented’.)

Lavrov, P.L. (1884) ‘Sotsial’naia revoliutsiia i zadachi nravstvennosti’ (Social Revolution and the Tasks of Morality), in *Vestnik narodnoi voli*; repr. in *Filosofiia i sotsiologiia (Philosophy and Sociology)*, Moscow: Mysl’ 1965.

Lenin, V.I. (1909) *Materializm i ėmpiriokrititsizm*, in *Polnoe sobranie sochinenii*, Moscow: Gospolitizdat, 1958-69, 5th edn, vol. 18; *Materialism and Empiriocriticism*, in *Collected Works*, Moscow: Progress and London: Lawrence & Wishart, 4th English edn, vol.14.(Also, see *Partiinost’* in the subject index of the collection.)

Marx, K. and Engels, F. (1848) *Das Kommunistische Manifest*, trans. *The Communist Manifesto*, in D. Fernbach (ed.) *The Revolutions of 1848*, Harmondsworth: Penguin, 1973, 62-98.

Stalin, I.V. (1950) ’Marksizm i voprosy iazykoznaniia’ (Marxism and Questions of Linguistics) in *Sochineniia (Works)*, vol. 3, Stanford: Hoover Institute, 1967; trans. *Marxism and Linguistics*, New York, International Publishers, 1951.(The 1950 declaration on freedom of criticism, widely translated.)

Pascal, Blaise (1623-62)

Blaise Pascal was a mathematical prodigy who numbered among his early achievements an essay on conic sections and the invention of a calculating machine. In his early twenties he engaged in the vigorous European debate about the vacuum, undertaking, or causing to be undertaken, a series of experiments which helped to refute the traditional view that nature abhors a vacuum and setting out clearly the methodology of the new science. In 1646 he came under the influence of Jansenism; this he seems to have rejected for a short time in the early 1650s, but he then underwent a profound spiritual experience which transformed his life and drew him into close association with leading Jansenists, with whom he collaborated in producing the polemical Lettres provinciales *(1656-7). At the same time he planned to write an apology for the Christian religion, but ill-health so affected his final years that this only survives in the fragmentary form of the* Pensées *(1670). He made significant contributions to mathematics, especially in the fields of geometry, number theory and probability theory, and he also helped to describe the* 'esprit géométrique' *which characterized the new science of the 1650s. He argued that geometry was superior to logic in that it could provide not only demonstrative procedures but also axioms from which to work; and he set down appropriate rules of argument. His religious writings were published shortly after his death; many attempts have been made to reconstruct the apology which they encapsulate. It seems most likely that this would have fallen into two parts, the first setting out the wretchedness of humans without God, the second demonstrating the truth of Christianity and the felicity of the religious life. Humans are portrayed in Augustinian terms as corrupt, vapid creatures, prey to their passions and the delusions of imagination; but they are also shown to possess greatness through their reason and self-awareness, which can bring them to recognize that Christianity alone has represented their predicament accurately, and that they should turn to religion, even if initially they lack the instinctive faith which is the hallmark of the saved. In the 'wager' fragment, Pascal employs his mathematical insights to revivify an old apologetic argument (that it is wiser to bet on God existing rather than on his not existing) and to link it to an existential imperative (that we all are obliged to choose between these alternatives). The adroit interplay between scepticism, rationalism and faith of the first part is succeeded by a second part which argues the veracity of Christianity from Biblical interpretation, prophecies and miracles. Pascal concedes that this cannot carry absolute conviction; but he insists that the rejection of such arguments is caused not by man's rational powers but by his corrupt passions. Pascal's* Pensées *are written for the most part in terse aphoristic form; he aspired to a style that was so accessible that the reader would believe he was experiencing as his own the thoughts that he read. Although Pascal said at the end of his life that he considered his mathematical pursuits a quite separate enterprise from his religious writings, a common epistemology can be found in both, together with a scientific outlook which Pascal saw as superior to the philosophical alternatives of his day.*

1 Early life and mathematical works

Blaise Pascal was born in Clermont in Auvergne, the son of a government official who was also an enthusiast for the new mathematical learning. With his family, he moved to Paris in 1631, and stayed there until 1638, when his father was forced to flee because of his public opposition to an aspect of Richelieu's fiscal policy. Blaise was educated privately by his father, who wanted him to be fully conversant with Greek and Latin before introducing him to mathematics: but his prodigy son worked out for himself the principles of geometry as far as the thirty-second proposition of Book I of Euclid at the age of twelve. It was also during this Parisian period that he was able to attend the mathematical academy of Father Marin Mersenne who was actively engaged both in European scientific and philosophical circles and in the religious controversy surrounding the new scientific ideas, as Pascal himself was also to become engaged. It is unclear how well-grounded Pascal was in the classics and in traditional Aristotelian logic and physics, and whether his unconventional education contributed to the originality that he showed in his later career.

After his father had been pardoned by Cardinal de Richelieu for his dissent and given the office of royal tax commissioner for the province of Haute Normandie, Pascal went with him to Rouen in 1640; in the same year his first mathematical publication, the *Essai pour les Coniques* (Essay on conical sections), appeared. Two years later he invented a calculating machine, which he had originally conceived to help his father with his tax work; this remarkable achievement was far in advance of the industrial skills required to produce it, although various versions were made and demonstrated to scientific colleagues, prominent politicians and members of the aristocracy.

2 The debate over the vacuum

Between 1646 and 1648, Pascal became embroiled in the fierce European debate concerning the existence of the vacuum. Torricelli's experiment with a barometer, which involved placing a tube of mercury upside down in a bowl of mercury, had been made public in France by Mersenne in 1644, and had given rise to many competing interpretations. Nearly all of these had recourse to the notion of atmospheric pressure as an explanation, and there was general agreement that the space at the top of the tube contained some kind of rarefied and invisible matter, which was consistent with the Aristotelian adage *natura abhorret vacuum*. In 1647, Pascal published his *Expériences nouvelles touchant le vide (Experiments on the vacuum)*; this was the summary of a series of experiments he conducted with Pierre Petit, using variously sized and shaped tubes and different liquids. Through them he was able to determine the quantity of water and mercury that could be supported by air pressure and the size a siphon had to be in order to function. He also set out in this summary the reasons why there was no rarefied and invisible matter occupying the space above the column of liquid supported in the barometer, but did not feel able yet to affirm the existence of a vacuum. He was challenged in his conclusions by Father Étienne Noel, the Jesuit Rector of the Collège de Clermont in Paris and a proponent of traditional Aristotelian physics; Pascal set out in his reply what are now taken to be the basic principles governing the application of scientific judgment and method. At the same time, he wrote to his brother-in-law Florin Périer to ask him to undertake the experiment of carrying a barometer up a mountain (the Puy-de-Dôme), the results of which showed that the level in the column of mercury varied with height. Pascal confirmed this himself on a church tower in Paris, and published the findings in 1648 in his *Récit de la grande expérience de l'équilibre des liqueurs (Account of the great experiment on equilibrium in liquids)*; he concluded that experiment, not authority, governed physics, that his experiments had shown that nature has no horror of a vacuum, and that air pressure accounts for all the effects associated with said imaginary horror. These experimental writings played an important role in discrediting Aristotelian and scholastic scientific ideas (see Aristotle; Aristotelianism, medieval; Aristotelianism, Renaissance).

3 Pascal and Jansenism

Pascal's father suffered an accident in 1646, which brought him into contact with a priest sympathetic to the ideas of Cornelis Jansen (1585-1638), the bishop of Ypres. Pascal also encountered Jansenist ideas on grace and piety, and was deeply affected by them, as were other members of his family. After her father's death in 1651, Blaise's sister Jacqueline became a nun at the convent at Port-Royal, which was the centre of Jansenist doctrine and religious practice in France (see Port-Royal). Pascal opposed her vocation strongly, and indeed for two years thereafter led a life very different from hers, consorting with free-thinkers, gamblers and libertines of fashionable Parisian society; but on the night of 23 November 1654 he underwent a profound spiritual experience which altered his life irrevocably.

This 'nuit de feu' was such that Pascal sewed the record that he made of it at the time of the event itself as a permanent memento into his clothes, where it was found at his death. The document records an experience of conversion; not an intellectual experience, but one which persuaded Pascal of the superiority of instinctive belief. (This conviction was strengthened two years later, when his niece was cured miraculously at Port-Royal of an apparently incurable fistula.) As a result of his experience, Pascal went to Port-Royal-des-Champs for a two-week retreat in 1655. There he met Isaac Le Maistre de Saci, a Jansenist theologian, with whom he had a debate, an account of which was published in 1720 as *Entretien avec M. de Saci (A conversation with M. de Saci)*. This debate indicates not only that Pascal had conceived of an apology of the Christian religion in terms which would speak most powerfully to the very libertines and gamblers of the Parisian society which he had just forsaken, but also that he felt that one of the most powerful voices with which he had to contend was that of Michel de Montaigne, the gentleman philosopher who had championed the cause of scepticism in the later sixteenth century. It is also clear that he felt the need to reassure those who had been shaken in their faith by new scientific developments, especially those in astronomy. Among Pascal's immediate sources of inspiration was Antoine Arnauld, the leading Jansenist theologian and philosopher, who was at that time on the point of being condemned by the Sorbonne for his religious views. Together with Arnauld and Pierre Nicole, another prominent Jansenist, Pascal composed the *Lettres provinciales (Provincial letters)*, a series of eighteen letters published in 1656-7. These constitute a scathing attack on the moral and theological views of the Jesuits, who were the most vociferous opponents of both Arnauld and Jansenism. In this debate the Jesuits were somewhat unfairly represented as a religious faction which engaged in deliberate deception for political ends and sacrificed doctrine to morals, and the

Jansenists were for their part depicted as crypto-Calvinists whose interpretation of St Augustine was both erroneous and heretical. Pascal and his co-authors tried vigorously to rebut the charge of heresy levelled at Jansen's writings while still acknowledging the authority of the Church that as Roman Catholics they were bound to accept. As it transpired, the debate was won *de facto* by the Jesuits; the *Lettres provinciales* were placed on the Index of Forbidden Books in 1657, and the Jansenist movement itself was condemned by the Pope shortly after. But the wit of the letters and their dazzling display of satire and irony have ensured that the judgment of posterity has been accorded to Pascal's side, at least in literary terms. For all that, the school at Port-Royal was closed in 1661, and the remaining *solitaires* and nuns were forced to sign a document to mark their submission to the Roman Catholic Church; this prompted Pascal to write the *Écrit sur la signature du formulaire (Tract on the signing of the formulary)*, urging Port-Royalists not to sign. In it he reiterated the defence of Jansenism which is found in the *Lettres provinciales*, distinguishing between fact and faith or law in matters of religion. Councils, Fathers of the Church and popes are infallible on matters of faith, but not on matters of fact, and Pascal contends that the identification of Jansenism with heresy is a matter of fact, not law. The *Écrit* was never published because Nicole and Arnauld contradicted Pascal's advice. After a stormy altercation with them, Pascal renounced all further engagement in religious controversy. Instead he devoted his final months of life to the poor of the part of Paris in which he lived, not only making over all his worldly goods to them, but also organizing what was in effect the world's first omnibus service, which carried passengers from one part of Paris to another for a fixed fare. Throughout his short life, he had never enjoyed long periods of good health; he died on 19 August 1662 after a protracted and painful illness, and his final days were marked by his own deep piety and his desire to fulfil his religious duties to the last.

4 Works posthumously published

Although Pascal was well known to the European scientific community in his lifetime, few of his works were published. His last foray into the world of mathematics - the solution to a set of problems concerning the nature of the cycloid - was in fact circulated anonymously as a competition in 1658, although its author was sufficiently well-known for the Dutch mathematician and physicist Christiaan Huyghens (1629-95) to write to Pascal about it in the following year. After Pascal's death some mathematicians (including Leibniz) had access to his scientific papers, but these did not all appear in print until much later. Pascal's religious writings, however, were posthumously edited by his family and friends; the *Pensées* and other short works appeared in 1670, and the *Entretien avec M. de Saci* in 1728. The original edition of the *Pensées* was both an abridgment and a reworking of Pascal's papers, only part of which had been put in order by him; since the mid-nineteenth century various attempts have been made to reconstruct the manuscript as it was left by Pascal, culminating in Louis Lafuma's edition of 1952 which is now taken to be standard. More recently still Pol Ernst has been able not only to reconstruct the pages which were cut up by Pascal himself when he decided to arrange the fragments into thematic groups, but also to establish the date of composition of the major part of the project (1656-8). This recent scholarship has permitted a (somewhat conjectural) chronological ordering of the *Pensées* to be published and has opened up new possibilities for their interpretation.

5 Mathematical philosophy

Pascal himself said in a letter to Pierre Fermat written in 1660 that he felt that his religious writings had little connection with his scientific and mathematical work. His outlook, however, was deeply influenced by what he conceived to be a new way of looking at the world inspired by geometry, and most commentators would agree that all his writings are impregnated with it. He himself made strictly mathematical contributions to number theory, geometry and probability theory, but he also involved himself in the wider polemic about the status of science in his day. In his letter to Father Noel, he set out the prerequisites of sound scientific methodology, laid down the rules for making affirmative or negative scientific judgments (through axioms and apodictic demonstration), and for establishing or disproving hypotheses about the physical world, which in his view could never be more than provisional. In the same letter, Pascal referred to the rival claims of authority (in this case, the authority of Aristotle) and scientific demonstration; this topic is more fully developed in the *Préface sur le traité du vide (Preface to a treatise on the vacuum)* (1651?), in which Pascal shows that experiment and correct reasoning should govern the sciences, and that authority and historical example have no place in them. His view of science is very much a progressive one; as more and more experiments are undertaken with more sophisticated instruments, previously accepted hypotheses are supplanted by newer ones. Thus the hypothesis of occult qualities or powers,

which was postulated to explain what lies beyond sensory perception, and the Aristotelian distinction between act and potency, should be replaced if experimentation can show that they are inadequate according to Pascal's rules. Natural causation and phenomena are unchanging; human attempts to understand them are relative to the historical moment at which the attempts are made. In a striking image, Pascal refers to the successive generations of scientists as a single person in a perpetual state of existence and development. Thus, when we disagree with scientists of the past, we are not contradicting them, since by applying the principle of charity we would have to agree that we would have understood the world in their way had we lived in their times with their resources; and it follows also that they would have agreed with us today for the same reasons. (It is worth noting that Pascal is willing to believe that past scientists acted and wrote in good faith, whereas he refused to concede that his Jesuit opponents did so.)

De l'Esprit géométrique et de l'art de persuader (On the spirit of geometry and the art of persuasion) (1657-8?) is a yet more sophisticated presentation of the new scientific outlook. Pascal begins by conceding that definitions in geometry are nominal and not real, and that what are taken for axioms are intuitive perceptions which can neither be demonstrated nor reasonably be doubted. The four terms which he identifies in this way are number, space, movement and time. All share the property of being infinitely divisible and infinitely extensible. This insight is counter-intuitive to those who conceive knowledge as finite but, unless it can be grasped, then the geometric spirit itself cannot be comprehended. Pascal is not claiming that man's capacity for knowledge is unlimited; merely that the immediate information of his senses and his reason have to be transcended if scientific advances are to be made. Geometry emerges from this as superior to logic, in that it can both provide axioms and engage in demonstration, whereas logic can only do the latter.

The second part of the work is devoted to the thorny problem of persuasion; here the will comes into question as the path through which human assent to a given argument is to be obtained. Even here, however, a method or a set of rules are supplied for the correct conduct of an argument. Terms must be given clear definitions, axioms must be incontrovertible and must all be explicit, and conclusions should be checked by substituting definitions for the terms used. Pascal's discussion of scientific method is therefore distinct from Bacon's negative use of induction although, like Bacon, he conceives of science in evolutionary terms; nor does it evince Descartes' greater reliance on the resources of human reason; it harnesses the arguments of the sceptics, but escapes from their epistemological dilemma by positing intuitive truths which do not come to us from the exercise of our intellect (see Bacon, F. §§2-6; Descartes, R. §§2-3).

6 Theology and the human condition

In 1658, Pascal gave an account of his planned apology for the Christian religion to his friends at Port-Royal, which is probably the reason for his cutting up his sheets of reflections and notes and arranging them into bundles. It is far from certain, however, that all the fragments which survive were written as part of this one project, or that the project was anything like complete and fixed in Pascal's mind. Not only are sections of the text in dialogue, with none of the voices clearly identified, it also seems likely that the grammatical first person who appears in a number of other fragments does not refer in all cases to Pascal himself. Such interpretative problems have not deterred past editors from partial or total reconstructions of the apology. The Lafuma edition, which is less interventionist, begins with a section devoted to the proposed organization of the work, which makes its general character clear. In the first part, the human condition was to be described, and shown to be wretched; in the second, human felicity with God and the truth of the Christian religion were to be demonstrated. In a series of sections entitled 'vanity', 'misery', 'ennui', 'causes of effects', 'greatness', 'contradictions', 'distraction', 'philosophers', Pascal sets out a description in implicitly Augustinian terms of human experience of the world. The arbitrariness and injustice of human political institutions, the vapidity of human pastimes, the false notions of social hierarchy, and the wilful flight of humans from confronting the primordial questions of their existence are all memorably expressed, often in terse aphoristic form. Notable is the very low assessment of the moral nature of humanity, whose self is said to be hateful. Pascal then turns to the paradox that in the human's very wretchedness there lie the seeds of greatness. Although the passions and imagination pervert and oppress them, humans also possess reason and self-awareness, and even can attain to certain knowledge about their environment through their intuitive grasp of geometric axioms. Such intuition, which Pascal sites in the heart, is also where true faith in God is to be found; the function of his apology is to persuade his reader rationally that the Christian religion is true because its description of the human condition is accurate, but it cannot do more than predispose the reader to

receive the gratuitous divine gift of faith. In this sense it is a superfluous enterprise; and although in his *Écrits sur la Grâce (Writings on grace)* (1657-8?) he tries to come to terms with the Tridentine proposition that the just are able of themselves to obey the Commandments, Pascal seems to concede the inefficacy of instilling rational conviction alone at various points in the *Pensées*.

As he does elsewhere, Pascal here sets up powerful binary oppositions between philosophies and exploits their reciprocal failings and strengths. Scepticism is modest in its claims, yet impotent and negative; dogmatism (roughly, rationalistic neo-Stoicism) is presumptuous and yet has some purchase on the real world. Philosophy is therefore incapable both of knowing humanity and of taking action. Pascal finds a way beyond this impasse by exploiting his insight about the infinitely great and the infinitely small in a novel manner: humans are everything with respect to nothing, and nothing with respect to everything (God and the cosmos). They are also imbued with a desperate desire for certainty, and cannot suspend their judgment indefinitely, because they are subject to an existential imperative: ‘you must take on the bet, for you are in the game’ (Pascal 1670: 550 [L418]). This is the context of the famous wager argument, which had been used before Pascal by other apologists. Either God exists, or he does not: if human life is vain and wretched (as Pascal believes he has demonstrated), humans have nothing to lose by betting on the next life (that is submitting to the Christian religion), for they have lost nothing in the case of God’s nonexistence, and gained everything if he does exist. Some have seen this argument as purely rhetorical; others have accused it of being a case of *petitio principii* (the worthlessness of life is presupposed in the conclusion that we have nothing to lose in sacrificing the allegedly vain pleasures of this world); yet others see in it an ingenious example of decision theory *avant la lettre*. Whatever interpretation is given to it, it can be seen to be particularly apposite to the potential constituency of libertines and gamblers whom Pascal was intending to address.

Pascal’s apology is therefore not simply fideistic, postulating the two truths of faith and reason; it employs both faith and reason, as well as doubt, to achieve certainty. As one *pensée* has it ‘one must doubt in the right way, assert in the right way, and submit in the right way’ (1670: 523 [L170]). Descartes tried to demonstrate too much, and relied on his *cogito* to gain knowledge of an infinite and hence incomprehensible God; Montaigne doubted too much, and did not agonize enough about his passive acceptance of ignorance and hedonism; only Pascal (and his mentor St Augustine) used reason to affirm, to negate and to recognize its own limitations appropriately. In the second part of his apology, Pascal planned to use arguments from Biblical history, prophecies, miracles, and above all else the interpretation of Holy Writ to present the case for the truth of Christianity; but he conceded that such proofs were not absolutely convincing, although they were sufficient to secure the consent of those who read them free from the perverting effects of their own corrupt passions.

In the *Pensées*, as in the *Esprit géométrique*, Pascal is at pains to stress the importance of style. The great art, which he identified in Epictetus, in Montaigne’s essays as well as in his own writings, was to write in such a way that those reading would think that they could have written the text themselves. As he shrewdly put it: ‘it is not in Montaigne, but in me that I find all that I find in his text’ (1670: 591 [L689]). Knowledge has to be made desirable, and also to be made accessible; the writer is its undetectable mediator; truth, as he had asseverated in the letter to Father Noel, is itself not a historical or personal phenomenon. The fortunes of his own writings do not seem to bear this out. In the late seventeenth century, he was the champion of those who embraced Augustine’s gloomy vision of corrupt human nature; in the eighteenth century, he was reviled for the same reason by moralists of a more optimistic bent; in the twentieth century he was caused to doubt the existence of God in the early part, a Kierkegaardian existentialist in the middle, and is now appreciated as a rhetorician and precursor of probability theory. The sustained interest shown in his work is a testament to its linguistic brilliance, its brevity and its clear vision of an epistemology embracing both the seen and the unseen, the intuited and the perceived, and the natural and the supernatural, all grounded in an impressively coherent outlook derived from his particular conception of geometry.

See also: Decision and game theory

IAN MacLEAN

List of works

Pascal, B. (1963) *Oeuvres complètes (Complete works)*, ed. L. Lafuma, Paris: Seuil.(The standard edition which includes the correspondence.)

Pascal, B. (1640) *Essai pour les Coniques (Essay on conical sections)*.(Published as a poster.)
Pascal, B. (1647) *Expériences nouvelles touchant le vide (Experiments on the vacuum)*, Paris: Margat.
Pascal, B. (1648) *Récit de la grande expérience de l'équilibre des liqueurs (Account of the great experiment on equilibrium in liquids)*, Paris: Savreux.
Pascal, B. (1651?) *Préface sur le traité du vide (Preface to a treatise on the vacuum)*.(Unpublished manuscript.)
Pascal, B. (1656-7) *Lettres provinciales*, trans. A.J. Krailsheimer as *The Provincial Letters*, Harmondsworth: Penguin, 1967; trans. T. M'Crie *et al*. in *The Provincial Letters: Pensées: Scientific Treatises*, Chicago, IL and London: Encyclopaedia Britannica, 1990.(The French first editions of the various letters are without imprint: both translations are good.)
Pascal, B. (1657-8?) *De l'Esprit géométrique et de l'art de persuader (On the spirit of geometry and the art of persuasion)*.
Pascal, B. (1657-8?) *Écrits sur la Grâce (Writings on grace)*.
Pascal, B. (1661) *Écrit sur la signature du formulaire (Tract on the signing of the formulary)*.(Unpublished manuscript.)
Pascal, B. (1670) *Pensées*, Paris: Port-Royal; seconde copie repr. ed. P. Sellier, Paris: Mercure de France, 1991; trans. A.J. Krailsheimer, Harmondsworth: Penguin, 1995.
Pascal, B. (1728) *Entretien avec M. de Saci (A conversation with M. de Saci)*, in P. Desmolets (ed.) *Continuation des Mémoires de littérature et d'histoire*, Paris: Tronchai.
Pascal, B. (1990) *The Provincial Letters: Pensées: Scientific Treatises*, trans. T. M'Crie *et al*., Chicago, IL and London: Encyclopaedia Britannica.
Pascal, B. (1992) *Discours sur la religion et sur quelques autres sujets (A Discourse on religion and other subjects)*, ed. E. Martineau, Paris: Fayard/Colin.(An edition based on the supposed chronological composition of the *Pensées*.)

References and further reading

Carraud, V. (1992) *Pascal et la philosophie (Pascal and philosophy)*, Paris: PUF.(Comprehensive, accessible and orthodox account.)
Davidson, H.M. (1979) *The origins of certainty: means and meaning in Pascal's Pensées*, Chicago, IL and London: University of Chicago Press.(Relates Pascal to developments in seventeenth-century epistemology.)
Descotes, D. (1993) *L'argumentation chez Pascal (Argumentation in Pascal)*, Paris: PUF.(Highly technical account of Pascal's logic, using nonstandard notation.)
Ernst, P. (1990) *Géologie et stratigraphie des Pensées de Pascal (Geology and stratigraphy of the Pensees of Pascal)*, Paris: Thesis of the University of Paris IV, 4 vols.(Mentioned in §4 above.)
Goldmann, L. (1959) *Le dieu caché. Études sur la vision tragique dans les Pensées de Pascal et dans le théâtre de Racine*, Paris: Gallimard; trans. P. Thody as *The hidden God: a study of tragic vision in the Pensées of Pascal and the tragedies of Racine*, London, Routledge & Kegan Paul, 1964.(A Marxist reading of Pascal, relating him to the worldview of his class.)
Hacking, I. (1975) *The emergence of probability: a philosophical study of early ideas about probability, induction and statistical inference*, Cambridge: Cambridge University Press.(Contains the claim that the wager argument is an early example of decision theory.)
Krailsheimer, A. (1980) *Pascal*, Oxford: Past Masters.(A good general introduction.)
McKenna, A. (1990) *De Pascal à Voltaire (From Pascal to Voltaire)*, Oxford: Voltaire Foundation, 2 vols.(Thorough study of the early reception of Pascal, and of the influences on his thought.)
Marin, L. (1975) *La critique du discours: sur la logique de Port-Royal et les Pensées de Pascal (The critique of discourse: on the Logic of Port-Royal and the Pensées of Pascal)*, Paris: Minuit.(A thought-provoking account of Jansenist philosophy.)
Mesnard, J. (1967) *Pascal*, Paris: Hatier.(A very good introduction in French.)
Mesnard, J. (1976) *Les Pensées de Pascal (The Pensées of Pascal)*, Paris: Société d'édition d'énseignement supérieur.(Good nontechnical account of the religious and literary aspects of the *Pensées*.)
Mesnard, J. *et al.* (eds) (1979) *Méthodes chez Pascal (Methods in Pascal)*, Paris: PUF.(Contains good papers on all aspects of Pascal's writing.)
Metzer, S.E. (1967) *Discourses of the fall: a study of Pascal's Pensées*, Berkeley, CA: University of California Press.(Good on religious aspects of Pascal's thought.)
Parish, R. (1989) *Pascal's Lettres provinciales: a study in polemic*, Oxford: Oxford University Press.(Best available account of the *Lettres provinciales*.)
Sellier, P. (1970) *Pascal et saint Augustin (Pascal and St Augustine)*, Paris: Colin.(Excellent study of Pascal's

debt to Saint Augustine.)

Thirouin, L. (1991) *Le hasard et ses régles: le modèle du jeu dans la pensée de Pascal (Chance and its rules: the model of gambling and play in Pascal's thought)*, Paris: Vrin.(Very good on the relationship of Pascal's thought to mathematics.)

Passmore, John Arthur (1914-)

John Passmore was born in New South Wales and studied at the University of Sydney. He taught there before moving to Otago in New Zealand and then to the Australian National University. He is perhaps best known for A Hundred Years of Philosophy which has been widely recognized as a major feat of philosophical scholarship. He has contributed widely to topics in the history of philosophy, philosophy of education, philosophy of science and philosophy of the environment. He is one of the pioneers of what has come to be called applied philosophy.

Passmore was born in Manly, New South Wales. He was an early student of the influential Scottish-born philosopher John Andersonat the University of Sydney. He taught at Sydney before going to the Chair of Philosophy at the University of Otago in 1950. In 1955 he moved to the Institute of Advanced Studies at the Australian National University where he was Reader and then Professor of Philosophy. He has held a number of senior visiting appointments in universities outside Australia and was Tanner Lecturer at Cambridge University in 1980.

For much of his career Passmore was something of a lone voice in Australian philosophy. His interests were more historical and applied than was usual in Australia in the 1950s and 1960s; indeed his work was probably better known during this period in Britain and America than in Australia. This changed with an increase in interest in the history of philosophy and especially in the contributions philosophers can make to issues of general concern that started in the 1970s, and many of his writings, particularly those on the philosophy of the environment, have become a focal point of discussion in Australia as well as overseas. He has an unusual ability to write argumentatively rigorous and philosophically sophisticated prose that readers not professionally involved in philosophy find challenging but not intimidating. In consequence his influence has extended well beyond philosophy departments. He has played an important role in showing the world at large how philosophers of an essentially analytical bent can make major contributions on subjects of general concern. He must be regarded as one of the pioneers of applied philosophy.

Passmore's best known book, especially among professional philosophers, is *A Hundred Years of Philosophy*, first published in 1957. It is widely recognized as a remarkable work of scholarship that displays an encyclopedic knowledge of the major philosophical movements of the preceding one hundred years combined with an intimidating understanding of their historical origins. It established his reputation in the history of philosophy. It has been a goldmine for teachers and students of philosophy, and a major first port of call for non-philosophers.

It is, however, his writings on the philosophy of the environment that have been the most influential in the sense of setting the agenda (see Environmental ethics; Green political philosophy). In *Man's Responsibility for Nature* he argues that we need as a matter of urgency to change our attitudes to nature, that we cannot go on living, as we have been, as predators on the biosphere. At the same time he argues vigorously against the view that we should see our present predicament as a sign that we should abandon the Western scientific and rationalist tradition and embrace some form of irrationalism or mysticism. The book is both an important contribution to environmental philosophy and an attack on the scientific irrationalism that is prevalent in some sections of the environment movement. In particular, he argues that the view that it is the Western tradition *per se* that is responsible for our present predicament is based on a misreading of and a selective attention to the complex historical facts. Rather, Passmore argues, we should use the tools that this tradition has provided us with to find our way out of our present predicament. A distinctive feature of the argument of the book, and indeed of much of Passmore's work, is the way it is set within a historical setting. He seeks to tell us how we got to where we are before telling us what we should do about it.

In *Philosophical Reasoning* a number of distinctive styles of reasoning in philosophy are identified and subjected to critical scrutiny. The book was published in the early 1960s, a time during which it was commonly claimed that these distinctive styles of argument, recently uncovered, served to dissolve many traditional philosophical problems. The enthusiasm with which philosophers embrace a new style of argumentation is not always matched by their readiness to examine its credentials, and this book played an important, if sobering, role.

See also: Australia, philosophy in

FRANK JACKSON

List of works

Passmore, J.A. (1951) *Ralph Cudworth*, Cambridge: Cambridge University Press.
(1952) *Hume's Intentions*, Cambridge: Cambridge University Press.
(1957) *A Hundred Years of Philosophy*, London: Duckworth.
(1961) *Philosophical Reasoning*, London: Duckworth.
(1970) *The Perfectibility of Man*, London: Duckworth.
(1974) *Man's Responsibility for Nature*, London: Duckworth.
(1985) *Recent Philosophers*, London: Duckworth.
(1991) *Serious Art*, London: Duckworth.

References and further reading

Brown, R. and Rollins, C.D. (1969) *Contemporary Philosophy in Australia*, London: Allen & Unwin.(Articles on sundry philosophical topics by Australians, with an introduction by Alan Donagan on the state of Australian philosophy and its background.)

Patañjali (*c*. 2nd century BC)

The grammarian Patañjali lived in the second century BC*, before the appearance of the classical systems of Indian philosophy. The aspects of his thought that we would call philosophical are concerned primarily with questions of meaning and meaning-bearers in language.*

1 Patañjali and philosophy

Patañjali is the author of the *Mahābhāṣya (Great Commentary)*, which comments on Pāṇini's famous grammar called *Aṣṭādhyāyī* (*c*. fourth century BC) and on Kātyāyana's *Vārttikas*; the latter have only survived as part of the *Mahābhāṣya (Great Commentary)* (*c*. second century BC). Patañjali is one of the few authors of early India whose approximate date is probably known. He lived around the middle of the second century BC in the north of India during the incursions into the country of the Graeco-Bactrian king Menander, to which he refers. This early date is confirmed by the absence of influence from the classical systems of Indian philosophy on his thought, with the possible exception of Sarvāstivāda Buddhism.

The *Mahābhāṣya* is not primarily a philosophical treatise. It addresses some philosophical questions in its introduction (whose title is *Paspaśāhnika*) while commenting upon certain grammatical rules (*sūtra*) of Pāṇini. Interestingly, Patañjali (who often sides in this respect with Kātyāyana) does not always share Pāṇini's point of view.

2 The linguistic units that have meaning

Rule 1.2.45 of Pāṇini's grammar states that 'what is meaningful, but is not verbal root or affix, is nominal stem' (*arthavad adhātur apratyayaḥ prātipadikam*). If we use the term 'stem' to refer to both verbal roots and nominal stems, this *sūtra* makes clear that for Pāṇini only stems and affixes really have meaning. Combinations of stems and affixes, that is, words and sentences, have at best meanings that are derived from those of the constituent stems and affixes. For Patañjali the situation is the exact opposite of this: words and sentences, not stems and affixes, are meaningful. This is clear from his discussion of rule 1.2.45 where he observes that this *sūtra* would assign the designation 'nominal stem' (*prātipadika*) to words and sentences. This contingency is avoided with the help of some far-fetched and unconvincing arguments. Next Patañjali turns to the question of how stems and affixes can be thought to have meaning, which he considers problematic, unlike Pāṇini. In Patañjali's opinion the nominal stem *vṛkṣa* on its own expresses no meaning. Cardona (1967-8) delineates the following solution offered by Patañjali: the method of concurrent occurrence of meaning and linguistic unit (*anvaya*) and absence of these two (*vyatireka*) shows that *vṛkṣa*, which is the common part of *vṛkṣas*, or 'one tree' and *vṛkṣau*, or 'two trees', must have the meaning, 'tree'. In this way, Patañjali derives the meaning of the stem *vṛkṣa* from the 'real' meanings 'one tree' and 'two trees'. The pair *pacati*, 'he cooks' and *paṭhati*, 'he recites', he similarly points out (*sūtra*1.3.1) allows us to assign a meaning to the common part -*ati*. Speaking generally, the meanings of stems and affixes are derived from the meanings of complete words.

Patañjali's deviation from Pāṇini is most easily explained by the hypothesis that he was influenced by the Sarvāstivāda Buddhists, who had reified phonemes and words (and perhaps sentences), but not stems and affixes, into existing 'real' elements (*dharma*). This is confirmed by the fact that Patañjali, too, accepts phonemes and words, but not stems and affixes, as independently (and eternally) existing entities. In connection with these independently existing phonemes and words, Patañjali uses several times the expression *sphoṭa*, which plays a major role in the discussions of later grammarians. Unlike them, Patañjali does not look upon the *sphoṭa* as a meaning-bearer, according to Joshi (1967). He speaks, for example, of the *sphoṭa* of individual phonemes. This *sphoṭa* is different from the sound (*dhvani*) which manifests it. The word or phoneme (*śabda*) is the *sphoṭa* and the sound is a property of the word or phoneme. If one person speaks slowly, another quickly, the *sphoṭa* is the same, only the manifesting sound is different (see Language, Indian theories of §§1, 3).

3 Do phonemes have meaning?

Although reticent with regard to the meaningfulness of stems and affixes, Patañjali pays great attention in his second chapter (*āhnika*) to the question of whether individual phonemes have meaning. He argues that certain stems and affixes consist of just one phoneme. The verbal root *i*, for example, means 'go'. Certain words which are

identical but for one phoneme, express different meanings: *kūpa* means ‘well’, *sūpa* ‘soup’, *yūpa* ‘sacrificial post’. Also, the removal of one phoneme can change the meaning: *vṛkṣa* means ‘tree’, *ṛkṣa* ‘bear’. Finally, if phonemes had no meaning, collections of phonemes could have no meaning either; one hundred blind people cannot see more than what one blind person can see. Patañjali subsequently rejects these arguments. Sounds normally have no meaning, because they can be modified, elided, or change position in a grammatical derivation. Since only collections of phonemes have meaning, nothing can be concluded from sets of similar words like *kūpa*, *sūpa* and *yūpa*, nor indeed from the pair *vṛkṣa* and *ṛkṣa*. Collections of phonemes can have features which the constituent phonemes do not possess: a chariot, too, can perform functions which its parts cannot.

4 The meaning of a word

What is the meaning of a word according to Patañjali? In the second *āhnika* four kinds of words are distinguished according to what they refer to: words that refer to a genus (*jātiśabda*), those that refer to a quality (*guṇaśabda*), those that refer to an action (*kriyāśabda*) and arbitrary proper nouns (*yadṛcchāśabda*). The first three of these clearly designate nouns, adjectives and verbs respectively. But in the first *āhnika* Patañjali enumerates (twice over) the following four classes of words: nominal words (*nāman*), verbs (*ākhyāta*), preverbs (*upasarga*) and particles (*nipāta*). This enumeration is taken from Yāska’s *Nirukta* (*c*. third century BC [1967: I.1]), where the meanings of these different types of words are elaborately discussed. Nominal words (nouns and adjectives), for example, are there described as ‘having *entity* as their predominant notion’ (*sattvapradhāna*), verbs as ‘having *being* as their predominant notion’ (*bhāvapradhāna*) (Kahrs 1986: 117). Patañjali was probably aware of the meanings assigned to words there.

Elsewhere (especially in *sūtra* 1.2.64) Patañjali distinguishes two possible meanings of words: the form (*ākṛti*) or the individual object (*dravya*). The former of these two positions is associated with the name of Vājapyāyana. Vyāḍi, on the other hand, held that words denote individual objects. Here, it seems, the discussion concerns itself with nouns primarily. In fact, there is reason to believe that for Patañjali ‘form’ (*ākṛti*) and ‘genus’ (*jāti*) were synonyms. Patañjali himself held that both form and individual object constitute the meaning of words.

See also: Meaning, Indian theories of §§1-2

JOHANNES BRONKHORST

List of works

Patañjali (*c*. 2nd century BC) *(Vyākaraṇa-) Mahābhāṣya*, ed. F. Kielhorn, 1880-5, 3 vols; 3rd edn, ed. K.V. Abhyankar, Poona: Bhandarkar Oriental Research Institute, 1972.(Comments on Pāṇini’s *Aṣṭādhyāyī*.)

Patañjali (*c*. 2nd century BC) *Mahābhāṣya*, trans. S.D. Joshi and J.A.F. Roodbergen, Poona, 1968.(Eleven volumes have appeared between 1968 and 1990. They continue to be published.)

References and further reading

Bronkhorst, J. (1987) *Three Problems Pertaining to the Mahābhāṣya*, Poona: Bhandarkar Oriental Research Institute.(Especially chapter 3, ‘The *Mahābhāṣya* and the development of Indian philosophy’.)

Cardona, G. (1967-8) ‘*Anvaya* and *vyatireka* in Indian grammar’, *Adyar Library Bulletin* 31-2: 313-52.(Discusses the use of this pair of conceptual tools.)

Cardona, G. (1980) *Pāṇini: a Survey of Research*, Delhi: Motilal Banarsidass.(A full bibliographical breakdown of editions and translations of the *Mahābhāṣya*.)

Joshi, S.D. (1967) *The Sphoṭanirṇaya of Kauṇḍa Bhaṭṭa*, Poona: University of Poona.(The introduction contains a discussion of Patañjali’s views on the nature of words.)

Kahrs, E. (ed.) (1986) ‘Durga on *bhāva*’, in *Kalyāṇamitrārāgaṇam: Essays in Honour of Nils Simonsson*, Oslo: Norwegian University Press, 115-44.(Discusses the possible interpretations of *sattvapradhāna* and *bhāvapradhāna*.)

Pāṇini (*c*. 4th century BC) *Aṣṭādhyāyī*, ed. and trans. O. Böhtlingk, Leipzig, 1887; repr. Hildesheim and New York: Georg Olms, 1977.(Pāṇini’s famous grammar book.)

Yāska (*c*. 3rd century) *Nirukta*, ed. and trans. L. Sarup, Delhi: Motilal Banarsidass, 1967.(The meanings of different types of words are discussed in detail.)

Paternalism

Restriction of people's liberty of action is paternalistic when it is imposed for the good of those whose liberty is restricted and against their will. The argument in favour of paternalism is that, if one can prevent people from harming themselves, there is no reason not to do so. Versions of the ethical creed of liberalism tend to oppose paternalism. One argument is that as a practical matter the policy of permitting paternalism tends to do more harm than good in the long run, or at least less good than a strict refusal to countenance paternalism would achieve. Another argument appeals to a right of autonomy which paternalism is held to violate whether or not its consequences on the whole are undesirable. Paternalist advocacy can be 'hard' or 'soft'; soft paternalism is the doctrine that paternalism can only be justifiable when the individual action that is being restricted was not chosen in a substantially voluntary way.

1 The concept of paternalism

One behaves paternalistically if one treats an adult as though one were a parent dealing with a child. One's behaviour shows concern for the welfare of the person and a presumption that one's judgment about what will promote it is superior. The paradigm of paternalism, and the focus of most philosophical discussion of it, is restriction of people's liberty against their will for their own good.

Whether a restriction of liberty is paternalistic or not depends on its rationale. If most people want a law requiring that they wear seat-belts when riding in cars for their own protection, and we enact a law in order to cater to this desire, enforcing this policy restricts people's liberty, but not against their will. In its application to the minority who do not want to be under this requirement, the law is still not paternalistic - though it may be unfair - if its rationale is administrative convenience and not the aim of restricting people's freedom against their will for their own good.

2 Utilitarianism for and against paternalism

The case for paternalism is simple. If we can prevent people from harming themselves, why not do it? According to act utilitarianism, one ought always to do whatever would produce the most good (utility) for people (and perhaps other sentient creatures) (see Utilitarianism). Whenever restricting someone's liberty to prevent harm to that very person is the utility-maximizing act, then act utilitarianism requires us to do it. Utilitarianism imposes stricter requirements of benevolence than most people accept (see Help and beneficence). One might hold that the better the cost-to-benefit ratio of a paternalistic act, the more obligatory the performance of the act. Or one might hold that so long as a paternalist imposition would do more good than harm, it is permissible even if not morally required.

In *On Liberty* (1859), a classic statement of a liberal utilitarian antipaternalism, John Stuart Mill does not dispute the theoretical possibility that in particular circumstances restricting someone's liberty for their own good might be the best thing to do from a utilitarian standpoint. Mill nevertheless proposes a liberty principle, which asserts that in modern societies the liberty of a sane and cognitively competent adult should never be restricted except to prevent harm to other persons who do not consent to this involvement (see Mill, J.S. §12).

Among its several implications, Mill's liberty principle forbids paternalistic restriction of liberty. Mill argues that we should adhere to this principle because in the long run, as a practical matter, rigid adherence to this rule would better promote utility than any alternative policy including the more flexible rule that forbids paternalism except in those particular situations in which paternalism happens to be utility-maximizing. Given that humans, including the humans who would administer any paternalist rule, are imperfectly rational, imperfectly well-informed, and not always disposed to be moral, the attempt by society to follow the liberty principle would produce more human happiness than any other course.

Mill's pragmatic defence of the liberty principle appeals to the likelihood that individuals will know their own interests better than those who are tempted to impose paternalistically on them, so such restriction will often be misguided. He also sketches a broad account of human nature and of what is needed to promote human fulfilment. According to Mill, humans vary in their natures. Given this diversity, no one or a few modes of life will be suitable for everyone. Moreover, our individual natures are not transparent to ourselves. One must learn about one's own

individual nature in order to develop a reasonable plan of life. This process of individual self-discovery requires wide freedom to try out different modes of life and to observe the results of other people's experiments in living. As we learn about the possibilities inherent in our nature, we come to value some above others, and work to mould our characters so as to conform to a self-chosen ideal. In short, Mill believes we need to develop our individuality to have a good chance of happiness, and to develop individuality we need wide individual freedom.

Leaving aside the merits of Mill's bold psychological speculation, one notes that his argument from the need for individuality to a rule against paternalism is incomplete. Even if we need wide freedom to develop our individual natures, it might well be best if people are stopped from the most disastrous and stupid self-harming activities.

A second criticism of Mill's argument is that, even if his claim that refraining from paternalistic imposition would maximize utility in the long run were correct, it might still be morally desirable to embrace some paternalism to achieve fair distribution of utility. A strict no-paternalism policy will tend to produce good outcomes for prudent individuals who are skilled at making and executing personal decisions and less good outcomes for individuals whose prudential skills are poor. On the average the imprudent bad choosers will be living worse lives. Even if strict no-paternalism happened to be utility-maximizing, instituting some paternalism might be better for worse-off persons, and required by fairness norms, especially if one thinks that imprudence is often nonblameworthy.

3 The right to autonomy; hard and soft paternalism

A more direct argument for the rejection of paternalism would dispense with utilitarian calculations and simply posit a moral right to live one's own life as one chooses so long as one does not thereby harm others in wrongful ways. On this view each individual has a right to autonomy, a right to govern their own life. Neither other individuals nor society as a whole has the authority to restrict an individual's liberty so long as the person's conduct is at worst self-harming and does not violate the rights of others.

The right to autonomy just described is stronger than a right against paternalistic imposition, because restriction of someone's liberty of action is paternalistic only if done for the sake of the individual who is coerced (or others who consent voluntarily to share involvement in the activity), and one might restrict an individual's liberty when it is conceded that the individual is not wrongfully harming others, in circumstances such that restricting the person's liberty would benefit others.

The right to autonomy as so far characterized is little more than a slogan pending further clarification of what its components might mean. One idea that calls for clarification is the idea that I have the right to act according to my own choices in self-affecting matters. But some restriction of people's choices does not seem to involve any overriding of their considered judgment and will. If Gertrude drinks a cup of poison, thinking it is harmless wine, she does not choose to bring about her death voluntarily. If Ulysses is driven temporarily insane by the Sirens' song, restricting his liberty for his own good is not really done against his will, because the will that chooses on the basis of insane impulse is not his authentic will. The ideal of voluntary choice encompasses many disparate disabilities. The more one chooses voluntarily, the more one is not choosing in ignorance of material matters of fact, not choosing on the basis of an error in reasoning, not choosing under conditions that tend to cloud judgment, not choosing under threat of coercion or duress, not insane or lacking basic cognitive capacities, and so on. Writers have distinguished 'hard' and 'soft' advocacy of paternalism. Soft paternalism is the doctrine that paternalism can only be justifiable when the individual choice that is restricted fails to be substantially voluntary - to meet a threshold level of voluntariness. The hard paternalist position holds that paternalism can be justifiable even if the individual action that is restricted is substantially voluntarily chosen.

One might interpret the right of autonomy as prohibiting paternalistic restriction of liberty only if the individual choice that is restricted is substantially voluntary. Or one might adopt a sliding scale: the more voluntary the choice, the stronger the moral presumption against the legitimacy of paternalism. Other views are possible - for example, one might hold that the legitimacy of paternalism depends both on the degree of voluntariness of the choice that is to be restricted and on the severity and danger of self-harm which the chosen conduct threatens to bring about.

See also: Confucian philosophy, Chinese §5; Consent; Freedom and liberty; Law and ritual in Chinese philosophy; Liberalism; Respect for persons

RICHARD ARNESON

References and further reading

Buchanan, A.E. and Brock, D. (1989) *Deciding for Others: The Ethics of Surrogate Decision Making*, Cambridge: Cambridge University Press, esp. 29-47.(Wide-ranging discussion of the treatment of the incompetent.)

Dworkin, G. (1972) ‘Paternalism’, *Monist* 56: 64-84.(Criticizes Mill on paternalism.)

Feinberg, J. (1986) *The Moral Limits of the Criminal Law*, vol. 2, *Harm to Self*, Oxford: Oxford University Press. (Most comprehensive recent treatment of the issue of paternalistic restriction of liberty.)

Hill, T.E., Jr (1991) ‘Autonomy and Benevolent Lies’, in *Autonomy and Self-Respect*, Cambridge: Cambridge University Press, 25-42.(Argues that some types of paternalistic manipulation violate autonomy and human dignity.)

Mill, J.S. (1859) *On Liberty*, Indianapolis, IN: Hackett Publishing Company, 1978. (Classic statement of a liberal utilitarian argument against paternalism in the context of a broad defence of individual freedom.)

Nozick, R. (1974) *Anarchy, State and Utopia*, New York: Basic Books.(Rejects paternalism by arguing for libertarian individual rights.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Oxford University Press, esp. 307-21.(Discusses issues of personal identity as they bear on the question of the rationality of caring now about the wellbeing of future stages of oneself, or of what one takes to be oneself.)

Patočka, Jan (1907-77)

*Patočka was a Czech philosopher, one of the last pupils of Husserl. From the mid 1930s he developed his own approach to philosophical problems of the life-world (*Lebenswelt*), its structures and human activities in it. After 1945 he expanded this theme to incorporate other phenomenological problems (movement, freedom, aesthetics). He also paid much attention to the history of philosophy since antiquity, particularly to Comenius. After 1972, when the Communist system deprived him of his academic position, he elaborated his philosophical concept of History in his 'Heretical Essays'. He was one of the authors of the political manifesto of the Czechoslovak political opposition, Charter 77, and his name became a symbol of moral resistance against totalitarian power.*

Patočka (born in Turnov, died in Prague) was a leading Czech philosopher of the twentieth century. After studying in Prague, Paris, Berlin and Freiburg, where he was a pupil of Husserl and Heidegger, he was awarded a Chair at Prague's Charles University as a result of his work *The life-world as a philosophical problem*. At this time he was secretary of the Czech-German group 'Cercle philosophique de Prague', and helped with the preparation of the International Philsophical Congress in Prague in 1934. During the Nazi occupation of Czechoslovakia (1939-45) universities were closed, and Patočka taught at a grammar school. After the war he returned to his university post but was dismissed for political reasons in 1949. He could, however, carry on his scientific work at the Academy of Sciences. At the time of the 'Prague Spring' (1968) he was reinstated at the University, but had to leave again in 1972. Subsequently, he organized illegal lectures and seminars, joined an opposition movement against the Communist regime and was co-author of the political manifesto, Charter 77. When he suddenly died after a series of police interrogations, his name became a symbol of moral resistance against totalitarian power, and of the fight for human rights.

Patočka follows the humanistic tradition of Czech thinking which starts with the founder of the Czech Reformation, Jan Hus, and his predecessors; he demands that philosophy should lead us from daily banality to full humanity. The philosopher is defined as one who has broken through the cycle of demand and consumption, who has elevated themselves above the world towards permanent values. Patočka links this mission of philosophy with an analysis of the crisis of modern rationality as described by Husserl. He seeks a way out of this crisis in the adherence to the life-world. This is the elementary world of humans, the pre-reflexive and prescientific world, but a world in which we can find the sense of our existence. Patočka does not understand this world in the sense of Husserl's transcendental subjectivity, but emphasizes its existential, ontological dimension, its autonomy and non-reducibility; he calls his approach 'a subjective phenomenology'. Humans control this world (and are controlled by it) by three different kinds of movement. First, the movement of acceptance, 'taking root', that is finding one's place in the world. Second, the movement of defence, the labour which keeps humans going in the world and secures their basic needs. Third, the movement of truth, transcendence, which humans use to step over matters given to them directly and with which they reach the world as a whole. Patočka was alone for a long time in developing the problems of the life-world. Only from the end of the 1950s were these problems dealt with more frequently in phenomenological writing. The international philosophical public first saw the fundamentals of Patočka's work in the 1976 French translation of his *The life-world as a philosophical problem*.

Another of Patočka's contributions to phenomenological problems concerns the philosophy of history. Husserl himself only touched on the question of history in his *Crisis in the European Sciences*, where he understands the problem of history in a narrow sense as the problem of the *notion* of the life-world. He does not take this problem as far as a concrete human life in the world, in society and in history (see Husserl, E. §12). Patočka attempted to do that in his work *Heretical Essays on the Philosophy of History*, written in the early 1970s at a time when he was not allowed to publish in his home country. This book, which outlined the foundations for Patočka's political activities, was therefore distributed illegally. (A French translation appeared in 1981, with a prologue by Ricoeur and an epilogue by Jakobson.) According to Patočka, history starts when people realize the problems of their hitherto matter-of-course life, when the meaning of their life as understood until then is shattered and they discover the possibility of a free life. This happened first in the Greek *polis*, origin of politics, history and philosophy, whose inter-relationship became the legacy of European people. The loss of the previous purpose of life was the beginning of a search for a deeper, freer and more demanding purpose, but not for the final purpose - history is not something conclusive. This is why one cannot take the attitude of a disinterested spectator towards history; each

has to feel responsible for it. From this follows the need to influence general matters and at the same time to 'look after one's soul'. Care for the soul is unthinkable without care for the commune. Human activity is a prerequisite for the movement of history.

Throughout his life, Patočka also studied the history of philosophy, understanding it as a stimulus for live philosophizing. In this spirit he wrote studies of Socrates, Plato, Aristotle and numerous works on Comenius. It is mainly to Patočka's credit that Comenius ceased to be seen as just a brilliant educator: Patočka proved that Comenius's ideas on education were based on an original philosophical concept, so that his outstanding position in the history of European philosophy has also to be recognized.

See also: Czech Republic, philosophy in

Translated by G.R.F. Bursa
JOSEF ZUMR

List of works

Patočka, J. (1936) *Le monde naturel comme un problème philosophique (The life-world as a philosophical problem)*, trans. J. Danëk, Amsterdam: Martinus Nijhoff, 1976.(The translation which first brought Patočka's fundamental ideas to an international audience.)

Patočka, J. (1976-7) 'Wars of the 20th Century and the 20th Century as War', *Telos* 30: 116-26.(The closing chapter of the book 'Heretical Essays on the Philosophy of History'.)

Patočka, J. (1977) 'The Husserlian Doctrine of Eidetic Intuition and its Recent Critics', in F. Elliston and P. McCormick (eds) *Husserl, Exposition and appraisals*, Notre Dame, IN, and London: University of Notre Dame Press, 150-9.(An interpretation of one of the central notions of Husserl's philosophy.)

Patočka, J. (1987-92) *Ausgewählte Schriften (Selected Writings)*, Stuttgart: Klein-Cotta, 5 vols.(The most important works of Patočka in German translation, including writings on culture, philosophy of history, phenomenology and Czech issues.)

References and further reading

Bloss, S., Stróżewski, W. and Zumr, J. (eds) (1995) *Intentionalität, Werte, Kunst - Husserl, Ingarden, Patočka (Intentionality, Values, Art - Husserl, Ingarden, Patočka)*, Prague: Filosofia.(Essays on the three phenomenologists by German, Polish and Czech authors in both English and German.)

Horak, P. and Zumr, J. (eds) (1992) *La Responsabilité/Responsibility*, Entretiens de Prague, Institut International de Philosophie, 1990, Prague: L'Institut de Philosophie de l'Academie Tchecoslovaque des Sciences.(A session in honour of Patočka, with the texts of lectures delivered by R. Klibansky, P. Ricoeur and L. Hejdánek, and a bibliography of translations of Patočka's writings into European languages.)

Kohak, E. (1989) *Jan Patočka: Philosophy and Selected Writings*, Chicago, IL: University of Chicago Press.(A study with English translations of some important works by Patočka.)

Tassin, E. and Richir, M. (eds) (1992) *Jan Patočka, philosophie, phénoménologie, politique (Philosophy, phenomenology, politics)*, Grenoble: Jérôme Milou.(Essays in French on Patočka by French, German and Czech authors.)

Patristic philosophy

*Early Christian writers used terminology and ideas drawn from Graeco-Roman philosophical literature in their theological writings, and some early Christians also engaged in more formal philosophical reflection. The term 'patristic philosophy' covers all of these activities by the 'fathers' (*patres*) of the Church. The literature of nascent Christianity thus contains many concepts drawn from Graeco-Roman philosophy, and this early use of classical ideas by prominent Christians provided an authoritative sanction for subsequent philosophical discussion and elaboration.*

Early Christians were drawn to philosophy for many reasons. Philosophy held a pre-eminent place in the culture of the late Hellenistic and Roman world. Its schools provided training in logical rigour, systematic accounts of the cosmos and directions on how to lead a good and happy life. While philosophical movements of the period, such as Neoplatonism or Stoicism, varied widely in their doctrines, most presented accounts of reality that included some representation of the divine. These rationally articulated accounts established the theological and ethical discourse of Graeco-Roman culture. As such, philosophy had a natural appeal to Hellenistic Jewish and early Christian thinkers. It provided a ready language in which to refine ideas about the God of the ancient Hebrew scriptures, and to elaborate the trinitarian God of Christianity. It also helped to bring conceptual coherence to the ideas found in the scriptures of both religions. Finally, it provided the common intellectual discourse that those communities required in order to present their central tenets to the majority culture of the Roman empire.

To a considerable extent, the notion of 'philosophy' suggested to the ancients a way of life as much as an intellectual discipline. This too drew Christians to the teachings of the philosophers. While there were doctrines and prescriptions of behaviour specific to the major schools, philosophers in general tended to advocate an ethically reflective and usually rather ascetic life, one which conjoined intellectual with moral discipline. This ethical austerity was prized by early Christians as an allied phenomenon within Graeco-Roman culture to which they could appeal in debates about the character of their new movement. The tacit validation that philosophy offered to the Christian movement was thus multifaceted, and, while it was sometimes thought to be associated with unacceptable aspects of pagan religious culture, philosophy provided some educated Christians with a subtle social warrant for their new life and beliefs.

It should be noted that ancient Christianity was itself a complex movement. Like Graeco-Roman philosophy, Christianity included a broad spectrum of beliefs and practices. Thus those early Christians who developed their beliefs with reference to philosophy endorsed a wide range of metaphysical and ethical doctrines, ranging from materialism to extreme transcendentalism, from asceticism to spiritual libertinism. Yet, while diversity is evident, it is also true that the Christian movement came to develop a rough set of central beliefs and some early forms of community organization associated with those beliefs. This incipient 'orthodoxy' came to value some sorts of philosophy, especially Platonism, which seemed best suited to its theological agenda. This tacit alliance with Platonism was fraught with ambiguity and uncertainty, and it was never a reciprocal relationship. Nonetheless, in the second and third centuries a type of Christian philosophical theology emerged which owed much to the Platonic school and became increasingly dominant among orthodox Christian authors. It was this trajectory that defined the character of patristic philosophy.

Early Christian thought had its origins in Hellenistic Judaism, and its initial character was defined by the dominant patterns of that tradition. This early phase extended through the first half of the second century AD, as Christianity began to define its distinctive themes associated with the nature and historical mission of Jesus Christ. Throughout the second century, Christianity became increasingly a movement made up of gentile converts; some of these new members had educations that had included philosophy and a few were even trained as philosophers. Thus Christian thought began to show increased contact with the Graeco-Roman philosophical schools, a trend no doubt reinforced by the critical need for Christians - as a proscribed religious minority - to defend their theology, ritual practices and ethics in the face of cultural and legal hostility.

This so-called 'age of the apologists' lasted throughout the second and third centuries, until Christianity began to enjoy toleration early in the fourth century. However, it would be a mistake to consider Christian philosophical thought in that period as primarily directed towards the surrounding pagan society. In many respects philosophy, as the intellectual discourse of Graeco-Roman culture, offered gentile Christians a means to clarify, articulate and

assimilate the tenets of their new faith. This process of intellectual appropriation appears to have been of considerable personal importance to many Graeco-Roman converts. Christian philosophical theology helped them to recover ideas familiar from their school training and to find unfamiliar concepts defended with the rigour much prized within Graeco-Roman culture.

After Christianity became a licit religion in the fourth century, philosophical activity among Christians expanded. The task of theological self-articulation became increasingly significant as Christianity grew in the fourth and fifth centuries towards majority status within the Empire, with imperial support. In this later period the range and sophistication of Christian thought increased significantly, due in part to the influence of pagan Neoplatonism, a movement that included a number of the finest philosophers active since the classical period of Plato and Aristotle. Later patristic philosophy had a defining influence upon medieval Christian thought through such figures as Augustine and Dionysius the pseudo-Areopagite, establishing both the conceptual foundations and the authoritative warrant for the scholasticism of the Latin West and Greek East.

1 Influences

Judaism in the time of Jesus of Nazareth was the heir of centuries of reflection on the unique existence of God, some of which was partly conditioned by Greek philosophical speculation. The book of Proverbs in the Old Testament emphasizes the one God who is the source of all wisdom and who produced the world (Proverbs 3: 19; 8: 22-31). Here wisdom begins to appear in a hypostatized role, as a quasi-independent power which serves to instantiate and express the hidden purposes of God (Proverbs 8: 1-31). Wisdom is the first product of God, and as such is a mediator between God and humanity. In the *Wisdom of Solomon*, an intertestamental work written between 100 BC and AD 50, wisdom (*sophia*) is presented as a distinct power that orders the cosmos and reveals the moral structure that conforms to that divine production. Wisdom is praised for her power and exalted as the source of prophecy and human sanctity. These texts indicate a characteristic aspect of Hellenistic Jewish theology: exaltation of a single, universal God in cosmological and ethical terms combined with discussion of a secondary principle that serves as an instrument of divine production and moral revelation.

Philo of Alexandria, the pre-eminent Jewish philosopher of the period, was preoccupied with the articulation of biblical theology in terms of Greek philosophy. His treatises, written in Greek, had a defining influence on early Christian philosophy, both in method and content. The works of Philo present philosophical interpretations of the Hebrew Bible, which he read in a Greek translation (the Septuagint), using strategies of allegorical exegesis. His philosophical sources are characteristic of the Hellenistic period, a mixture of Stoic, Peripatetic, Pythagorean and Platonic conceptions in which the transcendentalism of the Pythagoreans and Platonists was predominant (see Platonism, Early and Middle; Peripatetics; Pythagoreanism; Stoicism). For Philo, the patriarchs, especially Moses, were philosophers; it was to Moses that God revealed the transcendent archetypes upon which creation was based. These paradigms - understood as Platonic forms - constituted both the inherent patterns of order within the cosmos and the transcendent powers which generated it. Collectively, they were described as the Logos or divine Word, the ‘place’ of the forms. Philo used numerous ways to express this notion of an intermediate level of reality; sometimes the Logos seems to be only an aspect of God, in his role of relating to his products, but at other times the Logos seems to be a secondary entity that emerges from God and is used as an instrument to structure the cosmos. Other names or figures are also used, such as Wisdom (as we have seen), Goodness or Sovereignty; each represents an independent aspect of the divine nature. These ‘principles’ or ‘powers’, however understood ontologically, function to mediate between the primordial aspect of God and the cosmos that emerged from him.

One aspect of Philo’s portrayal of God bears particular mention: his unwillingness at times to attribute descriptive terms to God, particularly those that suggest direct divine interaction with the cosmos. Here Philo seems not only to be building upon traditional Hebraic reluctance to employ a proper name for God, but also to be drawing upon Pythagorean and Platonic theology, which proscribed ascribing predicates to the ‘One’ or the ‘Good’. Powers such as Logos or Sophia appear as the pre-eminent representations of the hidden and unsearchable divine nature, the self-revelation of God in metaphysical terms. But He Who Is, the God of Exodus, remained hidden in his august transcendence. This model of philosophical theology, with its interplay between a largely ineffable God and intermediate divine representations, would have a major effect on subsequent Christian thought, suggesting a philosophical approach to reflecting on the nature of Christ (see God, concepts of).

The use of philosophical discourse in Hellenistic Jewish thought was part of a natural strategy of cultural

association, drawing Judaism into discussion with ideas that were well-established within its host society. But it represents as well an internal development away from more archaic anthropomorphic conceptions of God, a development common to the Mediterranean religions of the age. Philo's use of Graeco-Roman philosophy is an important instance of this anti-anthropomorphism. God is a remote divine being who, if he can be described at all, can be approximated by metaphysical terms such as 'One', 'Good' or 'Monad'. It was God's Logos or Sophia that acted as a demiurge to craft the cosmos and now controls the world according to God's good purposes.

2 Beginnings of Christian thought

Some of this thinking seems to have influenced the earliest Christian writers of the New Testament, especially Paul and John. Although there is no firm consensus among scholars on the precise details of this relation, some patterns are evident. In Romans 1: 19-21, creation is understood to be an act of divine revelation, universal in its scope. The one God must, for Paul, be more than the guardian of an individual people, but also the source and judge of all humanity (Romans 3). The one God directs all things providentially for the good. This idea of a single God whose universal revelation is creation can also be found in speeches attributed to Paul in Acts (14 and 17).

It is in explaining the nature of Jesus Christ that the letters attributed to Paul clearly evince aspects of Hellenistic Jewish wisdom theology. Colossians 1: 15-23 presents a pre-existent cosmic power who is the image of the invisible God and the entity first born from God. This being is the creative power, by whom all things came into being, whether earthly or heavenly. The cosmic Christ is thus ontologically central: 'all things were created by him, and for him, and he is before all things, and by him all things consist' (Colossians 1: 16-17). In Paul's theology it is this power that manifested itself in Jesus of Nazareth, who died on the cross in order to reconcile a fallen world to himself and thus to God. The exact import of such Pauline passages is difficult to assess, since their conceptual background can only be approximately recovered. The same is true of the splendid prologue to the Gospel of St John. Similar prepositional language is used there to indicate the causal and ontological significance of a unique entity. The Word or Logos was *with* God in the beginning and serves as the power *through* which all things were made. As such, the Logos is the principle of life and light, and a power manifested historically in the human person of Jesus of Nazareth. This iteration of a cosmic mediator in Pauline and Johannine texts became the basis of subsequent Christian ontology, although these interpretations would differ significantly.

In the early second century, we find evidence of a developing consensus among Christian authors about God's metaphysical properties. Ignatius, the early second-century bishop of Antioch, writes about Christ as a timeless, invisible, intangible, and impassible being who shed these characteristics in order to become human (*Epistle to Polycarp* 3: 2). In the mid-century *Apology* of Aristides, God is presented in philosophical terms as the unmoved mover, the ruler of the universe and an eternal being without beginning or end. As such, he is understood to be ungenerated, immutable, free from need or defect, immobile and exempt from partition or division. These metaphysical epithets were commonplace among pagan philosophers of the period, and their adoption by early Christian apologists suggests a common theological discourse to which the new movement was making explicit appeal (see God, concepts of).

The effort to articulate a first principle which was transcendent of the cosmos led many Graeco-Roman thinkers, particularly Pythagoreans and Platonists, to reject the ascription of many characteristics, such as mutability or mobility, to the deity. This 'negative theology' came sometimes to regard all predication to be inadequate, so that even notions such as perfection or immateriality seemed too restrictive (see Negative theology). This tendency - mediated through Jewish thought - is also to be found in second-century Christian thinkers, especially some loosely classed as 'gnostics', who believed that they possessed a special, higher knowledge (*gnosis*) of the ineffable God beyond the divinity revealed in the sacred scriptures (see Gnosticism). An example is Basilides, a gnostic teacher of the mid-second century, who adopted a line of thinking whose roots go back at least as far as Plato's Academy. He held that the first principle could not be described by analogy with anything within the cosmos, but was best thought of as 'without being'. This completely transcendent power cannot be understood from the perspective of the universe; indeed, even calling it 'ineffable' is suspect. Similar views are common among other second- and third-century gnostic Christian authors. Their theology usually contrasted a hidden first principle, to whom special access is required, with lesser but misguided divine powers, whose presence is more easily attested.

However, this transcendentalism was not the whole story in early Christian thought. Some Christians were

influenced by more materialistic Roman philosophical schools, and their thinking was sharply opposed to the strongly transcendentalist tendencies of the gnostics. Perhaps the best example is Tertullian, a Roman jurist who converted to Christianity and campaigned vigorously for an understanding of the scriptures which was corporealistic, along the lines of Stoicism and Roman medical theory. For Tertullian, God was a spirit *sui generis*, a being understood in an attenuated materialist way, as a force which was invisible but nonetheless part of the cosmos. Similarly, the human soul was a corporeal entity, transmitted in the act of conception and immortal only through divine intervention. Such thinking was constructed in opposition to the developing transcendentalism of the early Christian Platonists, and it underscores the fact that the scriptural record could well have been read through a very different philosophical prism. But, as it happened, Tertullian's materialistic theology became a dead end, as the affinity of Platonism and Christianity was successfully argued by Christian philosophers in the second century.

3 Early Christian Platonism

By the latter half of the second century AD, Christianity had become largely a gentile movement, with a complex political and cultural relationship to the Roman Empire. Philosophical reflection was an important component in the developing self-articulation of the movement. A good example is the second-century *Apology* of Athenagoras of Athens, who wrote eloquently to the Stoic Emperor Marcus Aurelius in defence of Christianity, arguing that in rejecting the pagan gods Christians were not atheists, but believers in a single first principle. This cultural polemic is supported by his philosophical treatise *On the Resurrection of the Dead*. Here he addresses the composite nature of the human person, and argues that the unity of body and soul must be restored after death through resurrection of the body if human beings are to fulfil the eternal destiny intended by their divine creator. These arguments have an Aristotelian flavour, indicating a sophisticated and flexible use of philosophical sources. Similarly the three books entitled *Ad Autolycum* by Theophilos of Antioch, written about AD 181, contain arguments cast in philosophical terms. Written by a Christian convert to a pagan friend, these treatises argue for the uncreated nature of God, who is immutable but creates through his Logos. This Word exists both within God and also as the external expression of God through which he expressed himself in the act of creation.

The most important of these early apologists was Justin Martyr, who was executed *circa* AD 165. Justin was trained as a philosopher, having tried the Stoic, Peripatetic, Pythagorean and Platonic schools before converting to Christianity. His extant writings include two *Apologies* and a *Dialogue with Trypho the Jew*. Justin spent time as an itinerant Christian philosopher and even wore the philosopher's cloak (*pallium*) before settling in Rome, where he founded a school. Justin's approach to Graeco-Roman philosophical and religious culture was complex. In his view, the ancient Greek thinkers had access to the Old Testament and learned from it. Moreover, the divine Logos, the Wisdom of God, was the rational foundation of both the universe and the human soul, so that a natural knowledge of God was accessible to pagan philosophers. For Justin, all who think and act according to reason do so by participating in the Logos. Abraham and Socrates were Christians in anticipation of Christ, the incarnate Logos. Justin thus used philosophy to present a broad defence of Christianity, one that not only employed specific arguments in defence of scriptural conceptions, but also discovered a wide-ranging historical scheme to conjoin both the Hebraic and Hellenic cultures within Christianity.

The central locus of Christian philosophical activity in the second century was Alexandria, a major cultural centre of the Empire. There we find a Christian school, a succession of Christian teachers and pupils whose activities seem to parallel that of pagan philosophers. Clement of Alexandria is the first of the Alexandrian Christian teachers about whom much is known, although he himself refers to several earlier figures, especially Pantaenus, a former Stoic. Clement was probably a convert; he was born in mid-century and died around AD 215.

His principal work, the *Stromateis (Miscellanies)*, is a compendium of ethical and theological observations on a wide range of topics. Clement's goal seems to have been the presentation of a broadly educated, cultured form of Christianity. Its adepts would be Christian gnostics; unlike others who claimed that name, they would be both members of the developing orthodox movement and truly enlightened thinkers. Their life was described by Clement in ethical terms drawn from Stoic, Platonic and Aristotelian literature, characterized by an effort to approximate the stability and impassibility of God's own being. Pagan philosophy was significant in Clement's view since it represented a parallel phenomenon, the result of the illumination of the Greeks by the divine Logos. This notion of the natural knowledge of God allowed Christians like Clement to treat philosophy as an anticipation

of Christianity, a path to a common fund of knowledge completed by the revelations found in the Jewish and Christian scriptures.

In reading scripture, Clement owed much to Philo’s allegorical method of exegesis. Through the latter, Clement was able to discover the gnostic, or deeper, meaning of texts whose origins and character seemed remote from cultured Graeco-Roman sensibilities. This more advanced reading, grounded in philosophy, separated the simple believer from the mature or gnostic Christian, whose wisdom was enriched by this joint foundation in natural and revealed truth. Clement’s capacious model of Christian wisdom, with its compatibilist understanding of the Graeco-Roman and Hebraic traditions, helped to establish an enduring Christian approach to philosophy and, more generally, to the classical heritage.

Clement’s Logos theology had another feature which followed Philo’s thought, namely his emphasis on the unknowability of God. This ‘negative’ theology was, as we have seen, an important theological strategy in the period. The concept of a primordial but uncharacterizable godhead revealed through a lower power or aspect was widely accepted by Pythagoreans, Platonists and many heterodox Christian gnostics. But Clement, like Philo before him, put this negative theology to a positive use. The ultimate God was understood as beyond finite description, having no conceptually adequate predicates; but God was not thereby cut off from the world, nor was the cosmos at the mercy of ignorant or malevolent intermediate powers. For Clement, this hidden God, alluded to by the philosophers, was made manifest in his Word, the Logos, the transcendent power which produced the world and then became present in the person of Jesus of Nazareth. It is this fine balance between the hidden and the apparent that characterizes Clement’s theology, setting his views into sharp contrast both with heterodox gnosticism and with more affirmative or literalist approaches to theology within early Christianity. This difference, between apophatic theology (the theology of *apophasis* or denial) and kataphatic theology (that of *kataphasis* or assertion) thus came to the surface within orthodox Christianity through the catechetical school at Alexandria.

4 Origen

The greatest theologian in the era before Christianity’s legitimation in the fourth century was Origen of Alexandria, a product of the Alexandrian Christian school. Origen was born into a Christian family about AD 185; his father was martyred, a fate which Origen himself also suffered, dying subsequent to torture in the Decian persecution around AD 254-5. An intense reader of scripture from his youth, Origen studied both in Christian schools and in the pagan philosophical school of Ammonius Saccas, the teacher of Plotinus. Origen’s many writings encompass apologetics (for example, *Contra Celsum (Against Celsus)*), biblical commentary and systematic treatises (*De principiis (On First Principles)*). Much has been lost, although the substantial works that remain give ample evidence of vast scriptural erudition combined with a bold command of contemporary Platonic thought. Origen may fairly be said to be the most innovative Christian thinker of the pre-Nicene era, the most brilliant and important Christian intellectual before Augustine.

Unlike Clement, Origen generally avoided negative theology, perhaps because of its popularity among the heterodox gnostic thinkers whom he opposed. His theology focused upon the self-diffusion of God, a single divinity that produced a finite and intelligible image. This image - the Word - is the collective world of the Platonic intelligibles and the foundation of the lower, created cosmos. The Word is eternally generated by the Father, so that God is eternally productive of his perfect image. As such, the Logos appears to hold an intermediate status in the structure of reality, the link between God and the cosmos. Matter was also attributed by Origen to God, who was thus the only foundation of reality.

Origen also made human freedom central. Souls were seen as rational beings created with free choice. On Origen’s Platonic exegesis of Genesis, souls that turned away from God descended to lower levels of reality. Freedom of choice produced psychic precipitation, the falling of rational souls into their present corporeal state. This primordial loss of perfection resulted in a cosmic system of distinct worlds. Each degree of reality was attuned to the level of the soul’s descent. Souls only slightly separated from God could continue to engage in everlasting contemplation; Origen identified these with the stars. Others were entombed in human bodies and are subject to death. The demons are souls who chose the most extreme removal from God. Thus the cosmos was the direct manifestation of a fundamentally moral phenomenon; reality is the expression of the ethical disposition of its inhabitants. No other Christian philosopher in the patristic period linked moral freedom so directly with ontology.

Since evil is, in this theory, the direct result of the soul's choice, the natural evils of our world and mortality itself are epiphenomena of moral evil (see Evil, problem of). Death is thus punitive, and suffering educative. The cosmos is a vast penal colony designed for the rehabilitation of souls. Origen seems moreover to have held that all souls would eventually be successfully reformed. This universalism presents cosmic history as a providential process which will eventuate in a restoration of perfection, 'that God may be all in all'. Whether this reformation of souls precludes for Origen an additional, subsequent misuse of freedom is unclear. One corollary of Origen's universalism, that the devil would be saved, was not lost on his contemporaries; it contributed to the controversy which surrounded Origen's thought throughout late antiquity.

5 Later patristic philosophy: Gregory of Nyssa, Marius Victorinus, Ambrose

In AD 313, the imperial Edict of Milan initiated a new era of religious legitimacy for Christianity; shortly thereafter, the great ecumenical Council of Nicaea (AD 324-5) began the process of defining orthodoxy with imperial support. Christian philosophy was influenced by both events. No longer was there a need for political polemics in defence of the faith, although the effort to present Christianity to educated pagans continued unabated. Now, internal efforts to define and explain the doctrines of orthodoxy came to the fore.

These developments are particularly evident in the works of three Cappadocian bishops, Gregory Nazianzen (*circa* AD 330-90), Basil of Caesarea and Gregory of Nyssa (*circa* AD 330-90). Of these, the most philosophically acute was Gregory of Nyssa, whose theology is much influenced by Origen and by post-Plotinian Platonism (see Neoplatonism). Unlike Origen, Gregory did not ground his thought exclusively on a model of free will and psychic precipitation. Rather, divine creation is the dominant image and the soul but a creature conditioned by God's providential intention. That eternal plan involved materiality and the body. Thus the fall constituted a desecration of human nature, but it did not generate the mingling of soul with body. The misery of human life can be overcome only by restoration of our original state through God's intervention, through Christ. Philosophy can guide the soul to recognition of its condition, but the soul can be saved only by God's activity. Philosophy alone can never be sufficient for salvation.

Gregory's thought put a renewed emphasis on the conceptual transcendence of God, so that the divine Father was presented along Plotinian lines as being beyond all predicative descriptions. In Gregory's case, this view was based upon a sharp recognition of divine infinity. The created soul has, as its future course, the eternal process of the contemplation of God's infinitude; unification of the human soul with the one God can never be wholly consummated.

Gregory of Nyssa and the other Cappadocians are evidence of a great emergence of Christian Platonism in the Greek world of the late fourth century. This same post-Nicene flowering of philosophical theology also occurred in the western portions of the Empire. Here the influence of the Greek philosophical schools was somewhat less direct, in part because of linguistic difficulties, but it was still felt. Marius Victorinus was a Roman rhetorician much influenced by the works of the Plotinian school; he translated some of these into Latin and wrote treatises with a pronounced Neoplatonic influence in defence of Nicene orthodoxy. For Victorinus, Platonism seemed a conceptual resource and philosophical ally of Christian orthodoxy. This is also true to an extent of Ambrose, the great fourth-century bishop of Milan, who was concerned not only with re-drafting Christian theology in a Platonic idiom accessible to cultured pagans, but also with presenting Christian asceticism and theology as a successful rival to Platonic philosophy. His strategy had internal resonance within Christianity as well, allowing orthodoxy to assert its superiority to Arianism, Manicheism and other forms of Christianity (see Manicheism). His greatest success was the conversion (from Manicheism) of the North African rhetorician Augustine, whom he baptized in AD 387. It was the preaching of Ambrose, together with his treatises, that led Augustine to study 'the books of the Platonists' which Marius Victorinus had translated. This proved a spiritually volatile mixture, sending Augustine into an orthodox Christian trajectory and a life of asceticism.

6 Later patristic philosophy: Augustine, Boethius, Pseudo-Dionysius

Like Gregory of Nyssa and Ambrose, Augustine was determined to construct a coherent Christian theology which could stand as a compelling alternative to the still active and prestigious pagan tradition. It is well to remember that Julian the Apostate, the recidivist emperor who sought to restore a paganism through the revitalization of polytheistic cult and Neoplatonic theology, ruled during Augustine's youth (AD 361-3). Augustine's thought might

well be considered as a fresh response to the same set of desiderata as had moved Gregory and Ambrose: problems internal to Christianity and external ideas which were hostile to orthodoxy. Perhaps because he read little Greek and had an informal philosophical education, Augustine's thought is less a Christianized Platonism than it is a systematic development of Christian theology, using Platonic epistemology and metaphysics. Often Augustine seems intent on answering classical philosophical questions in novel, Christian ways.

Nowhere is this clearer than in the *Confessiones (Confessions)*, where Augustine carefully asserts his own personal success at Plotinian contemplation in the vision at Ostia, only to conclude that pagan philosophy, while providing the soul with epistemic access to God, was inadequate as a means of assuring continuing salvific association with God. Augustine presents this theme as an autobiographical refutation both of Manichean materialism, now confounded by the soul's transcendental vision through contemplation, and Platonism, found wanting in its prideful over-estimation of the soul's natural proximity to the eternal and the divine. These opponents are the background to the brilliantly original account of memory and time in Books X and XI of the *Confessiones*, where the phenomenology of the fallen, embodied, and temporally constrained soul is explored. This account of the human soul constructing a lapsarian self and so constituting time as the medium of its collective anxiety and loss had a decided impact on the development of the introspective consciousness of the West and on the framing of a philosophical account of Genesis.

Throughout the vast corpus of his later writings, Augustine frequently makes use of philosophical notions, usually in contexts whose primary focus is theological. In the *De civitate Dei (City of God)*, however, he returns to the question of the value of philosophy, especially Platonism. In this sustained argument, he makes plain his admiration for the intellectual utility of philosophy, while rejecting the philosophers' pretence to salvific efficacy and excoriating their continuing acceptance of polytheistic cult. It is interesting to note that Augustine views Platonism as a natural approximation to Christianity, achieved through reason but without the benefit of revelation. Yet the Platonists were, in the end, spiritually incoherent, accepting a first principle, the Good or One, while countenancing polytheistic worship.

While Augustine was in many respects a *sui generis* thinker with a penchant for original reflection on philosophical topics, there were other Christian thinkers in late antiquity whose works were more clearly bound by philosophical conventions. Boethius is an outstanding example. An active member of the Roman senatorial circle during the Ostrogothic period in Italy, Boethius undertook the revival of technical philosophy at a time of its marked decline in the western provinces. He translated and commented on some of Aristotle's logical works, and wrote a series of five theological treatises on the Christological debates of his time. Most important for medieval and renaissance readers, he completed the *De consolatione philosophiae (Consolation of Philosophy)*, a protreptical work on the value of the philosophic life, while a political prisoner before his execution. In most respects, Boethius was a conventional Christian Platonist, although the Christian element can be found only in his theological works (see Platonism, medieval). His efforts to articulate and defend the congruity of divine foreknowledge and human freedom were critical to later medieval scholasticism, as were his commentaries on Aristotle. His pupil Cassiodorus continued aspects of the encyclopedic legacy of Boethius, concentrating his efforts on articulating the seven traditional liberal arts (grammar, dialectic, rhetoric, arithmetic, geometry, music and astronomy). He also wrote a treatise on the soul, arguing that it is transcendent of materiality, but distinct from God in its capacity for evil. As is the case with his teacher Boethius, Cassiodorus is significant primarily for his role in transmitting classical philosophy to the medieval age (see Encyclopedists).

At about the same time (early sixth century) in the Greek East there surfaced a collection of treatises under the name of Dionysius the Areopagite (St Paul's Athenian convert in Acts 17). The identity of the actual author has been the subject of extended speculation and remains undetermined, although the works came to be accepted as genuine in both the Byzantine East and later in the Carolingian West (see Pseudo-Dionysius). These treatises represent the high-water mark of compatibilism between Christianity and late Platonism. They centre on the absolute predicative transcendence of God. The soul must grasp the total darkness of divinity as well as the vast manifestation of its hidden nature in the cosmos that emerged from it. The whole of reality is a theophany, an expression of the One who cannot be known *per se*. The levels of reality that are manifest to us constitute grades of being, each part of a pattern of divine self-presentation and return. The concept of God, while undefinable, is not privative or nugatory but can be partially understood as characterizing the source of reality.

This theology represents a Christian redaction of the pagan Platonism of the fifth century. While assuredly Christian in its theological terminology, its basic ontology is intelligible only when read against the metaphysics of the pagan Neoplatonists. This was one possible line for Christian intellectual development. However, there were others who resisted such assimilation. Chief among them was John Philoponus, who was active in the Greek East during the first half of the sixth century. Philoponus was sharply opposed to pagan Neoplatonism, especially that of Proclus. He criticized Aristotelian and Neoplatonic claims about the eternity of the cosmos, and favoured a model of temporal creation. Even if the world were beginningless, it must be seen as a contingent system that depended for its existence upon a transcendent source. In his view, the eternalism of the pagan cosmologists occluded this central metaphysical point (see Eternity of the world, medieval views of). Moreover, the commitment of many pagan Platonists to a theory of necessary cosmic emanation from the One further confused the issue. In consequence, Philoponus attempted to clarify the fundamental ontological dependence of a contingent cosmos upon its divine creator. He is a representative of a vigorous Christian philosophical movement in late antiquity which sought to sharpen the lines of conceptual demarcation between Christianity and pagan Platonism, while using common methods of philosophical argumentation to achieve that end.

The closing of the pagan Platonic academy in Athens was ordered by the Emperor Justinian in 529. Subsequently philosophy continued in a variety of Christian philosophical schools, each committed to different, rival theologies, including monophysites, Chalcedonians and Nestorians. It was on this foundation that later Byzantine Christian philosophy and early Islamic philosophy were able to develop, transmitting the thought of both the classical and the patristic traditions into the medieval period.

See also: Augustine; Boethius, A.M.S.; Byzantine philosophy; Clement of Alexandria; Encyclopedists; Gnosticism; Hellenistic philosophy; Manicheism; Marius Victorinus; Neoplatonism; Origen; Pelagianism; Platonism, Early and Middle; Tertullian, Q.S.F.

JOHN PETER KENNEY

References and further reading

Armstrong, A.H. (ed.) (1966) *The Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press.(The major review of philosophy in late antiquity.)

Augustine (397-401) *De civitate Dei (The City of God)*, trans. J. O’Meara, Harmondsworth: Penguin, 1972. (Influential work of Augustine.)

Augustine (413-27) *Confessiones (Confessions)*, trans. R.S. Pine-Coffin, Harmondsworth: Penguin, 1961; trans. H. Chadwick, Oxford: Oxford University Press, 1991.(Influential work of Augustine.)

Chadwick, H. (1981) *Boethius: The Consolations of Music, Logic, Theology and Philosophy*, Oxford: Clarendon Press.(A study of Boethius and his age.)

Hadot, P. (1995) *Philosophy as a Way of Life: Spiritual Exercises from Socrates to Foucault*, trans. M. Chase, Oxford: Blackwell, 1995.(Essays on ancient philosophy as a spiritual discipline.)

Origen (220-30) *Peri archōn (De principiis) (On First Principles)*, ed. P. Koetschau, Leipzig: Hinrichs; trans. G.W. Butterworth, *On First Principles*, New York: Harper & Row, 1966.(Edition and translation of Origen’s work.)

Origen (220-30) *Contra Celsum (Against Celsus)*, ed. P. Koetschau in *Origenes werke*, Leipzig: Hinrichs; trans. H. Chadwick, Cambridge: Cambridge University Press, 1965.(Edition and translation of Origen’s work.)

Osborne, E. (1993) *The Emergence of Christian Theology*, Cambridge: Cambridge University Press.(Examines pre-Nicene Christian thought.)

Pelikan, J. (1993) *Christianity and Classical Culture: The Metamorphosis of Natural Theology in the Christian Encounter with Hellenism*, New Haven, CT: Yale University Press.(A study of Cappadocian natural theology.)

Quasten, J. (1950-60) *Patrology*, Utrecht/Antwerp: Spectrum Publishers.(Comprehensive review of early Christian literature.)

Rist, J.M. (1994) *Augustine: Ancient Thought Baptized*, Cambridge: Cambridge University Press.(Augustine’s thought reviewed in reference to later classical philosophy.)

Stead, C. (1994) *Philosophy in Christian Antiquity*, Cambridge: Cambridge University Press.(A short study of later pagan and early Christian thought.)

Wolfson, H.A. (1956) *The Philosophy Of The Church Fathers*, Cambridge, MA: Harvard University Press.(Dated, but comprehensive.)

Patrizi da Cherso, Francesco (1529-97)

*Francesco Patrizi was an Italian humanist and anti-Aristotelian who took up a newly-founded chair of Platonic philosophy at Ferrara in 1578, the first such chair in Europe. Through his various writings he contributed to poetic theory, rhetoric, and historiography, as well as to military history and hydraulics. His two most influential works were his Discussiones Peripateticae (1581) and his Nova de universis philosophia (New Philosophy of Universes) (1591). Patrizi cast doubt on the authenticity of many of the works attributed to Aristotle, and argued that Aristotle's philosophy was incompatible with Christianity. He believed it should be replaced with his own synthesis of Platonism, Neoplatonism, and Hermeticism (or Hermetism). Patrizi saw light as the basic metaphysical principle, and interpreted the universe in terms of the diffusion of light (*lumen*) from God, the primary light (*prima lux*). His most influential doctrine concerned space, which he argued to be infinite, three-dimensional, and distinct from the bodies it contained.*

1 Life and works

Francesco Patrizi (Franciscus Patritius) was born in Cherso, near Trieste. He studied in Venice and Ingolstadt, before going to Padua where he studied and taught private pupils for a number of years (1547-54). It was at this leading Aristotelian institution that he was introduced to Marsilio Ficino's *Theologia platonica (Platonic Theology)* (see Ficino §3), and became a Platonist. He spent some years away from the university, serving as secretary and administrator to various Venetian noblemen, and travelling extensively in France, Spain and Cyprus (where he perfected his knowledge of Greek). His return to the academic world marked the belated institutional recognition of Platonic philosophy. Shortly after Francesco de'Vieri (Verino) had begun teaching Platonic philosophy at Pisa in 1576, Patrizi took up a newly-founded chair of Platonic philosophy at Ferrara in 1578. In 1592 he was called to Rome by Pope Clement VIII and became professor of Platonic philosophy at the Sapienza. He still held this post at his death, despite ecclesiastical censure of his writings by the Roman Congregation of the Index, which ordered changes in his *Nova de universis philosophia (New Philosophy of Universes)*.

Patrizi's humanist and literary interests are shown in his works on poetry, rhetoric and history. He wrote poetry himself, one of his earliest writings was a discourse on types of poetic inspiration (*Discorso della diversità dei furori poetici*) (1553), and he later entered into a dispute with the poet Torquato Tasso about poetics. In his *Della poetica* (1586) he attacked the Aristotelian theory of poetry as mere imitation, and appealed to the marvellous as an aesthetic category (see Aristotle §29). The poet, who is 'maker of the marvellous', operates on the soul, putting it in tune with cosmic harmony. His views about the relationship between words and things also appear in his earlier work on rhetoric (*Della retorica*) (1562), where he claimed that search for rhetorical ornamentation had corrupted the original Adamic language which gave direct access to things (see Language, Renaissance philosophy of §2). In his work on history (*Della historia*) (1560), which had a great influence on later historiography, he attacked Ciceronian orthodoxy (see Cicero §2). He wanted to separate history from rhetoric, seeing it as an autonomous discipline with its own intellectual justification based on a grasp of causes and effects.

His other works included discussions of love, unpublished during his lifetime, in which he blended Neoplatonism with a naturalistic materialism. In another vein entirely, he wrote on military history, and was also the author of treatises on hydraulics, stemming from the conflict between Ferrara and Bologna over the silting of the river Po. Surviving correspondence shows his links with various intellectual circles.

His two most important works were the *Discussiones peripateticae* (*Discussionum peripateticarum tomi quattuor*) where we find his main attack on Aristotelianism, and the *Nova de universis philosophia* which presents his light-metaphysics and his philosophy of nature.

2 Anti-Aristotelianism

The prime source for Patrizi's attack on Aristotle is his *Discussiones peripateticae* published at Basle in 1581, though the first volume had been issued separately at Venice in 1571. This volume contained a biography of Aristotle which highlighted his personal shortcomings, such as stealing the ideas of his predecessors, and a history of Peripatetic philosophy. The three additional books discussed Aristotle's sources, the discord between Plato and Aristotle, and the defects of particular Aristotelian doctrines. Patrizi cast doubt on the authenticity of many of the

works attributed to Aristotle, and he foreshadowed modern commentators by arguing that Aristotle's *Metaphysics* was a series of fragmentary treatments of diverse topics rather than an integrated whole (see Aristotle §11).

Patrizi's main philosophical quarrel with Aristotle was that his philosophy was incompatible with Christianity, especially with respect to the providence and omnipotence of God. In his Dedicatory Epistle to *Nova de universis philosophia*, he blamed the study of Aristotle for the fact that 'Common men laugh indiscriminately at Philosophers with this saying, which is now a commonplace: this man is a philosopher, he does not believe in God' (Henry 1979: 559-60). The remedy for this situation was to be found in Patrizi's own philosophical synthesis of Platonism, Neoplatonism and Hermeticism (also known as Hermetism).

In order to understand Patrizi's synthesis, we must recognize his belief, taken from Ficino (see Ficino, M. §2), in an ancient theology (*prisca theologia*) handed down to sixteenth-century thinkers by a long and continuous tradition which began with Moses or even Noah, and which had the work of Plato as its ancient culmination. Both Plato himself, whom Patrizi presented as a genuinely systematic thinker, and the various works of the Hermetic tradition, were fitted into a Neoplatonic framework (see Neoplatonism).

Patrizi's attachment to non-Aristotelian sources is shown by his textual work. He translated John Philoponus' commentary on Aristotle's *Metaphysics* (1583) (see Philoponus §2) and Proclus' *Elements of Theology and Physics* (1583) (see Proclus §§2, 4, 9). One of the works he appended to his *Nova de universis philosophia* was his Latin translation of the *Theology of Aristotle*, a ninth-century Arabic compilation of extracts from Plotinus (§4) which Patrizi described as 'the mystic philosophy of the Egyptians dictated by Plato and taken down by Aristotle'. He also added an edition and Latin translations of the Greek *Corpus Hermeticum* and some other Hermetica, including the Latin *Asclepius*, all probably dating from between AD 100 and 300, as well as the Chaldaean Oracles, a work of the second century AD, which he attributed to Zoroaster (see Chaldaean Oracles). Contemporaries such as Teodoro Angelucci in his *Exercitationes* (1585) were already beginning to cast doubt on the historicity of the *Corpus Hermeticum*, when viewed as the work of the mythical Egyptian sage, Hermes Trismegistus, and as part of the *prisca theologia* handed on by Orpheus to the ancient Greeks, but Patrizi insisted on its authenticity and importance. In his Dedicatory Epistle he wrote: 'Surely Hermes' little book on piety and philosophy contains more philosophy than all of Aristotle's work' (Henry 1979: 553). He also claimed that Hermes was a little earlier than Moses, and had spoken more clearly of the Trinity (see Hermetism).

3 Light-metaphysics

Patrizi's main philosophical work, *Nova de universis philosophia* (Ferrara, 1591; Venice, 1593), is divided into four parts. Part One is called 'Panaugia' (all-splendour), a word borrowed from Philo Judaeus (see Philo of Alexandria). In the Platonic tradition, light had always been seen as a symbol of goodness and as the closest analogy to God in the physical world, and had early been turned into a metaphysical or ontological principle. Patrizi presented God as the primary light (*prima lux*), who produces light (*lumen*), first in his Son, then in all other incorporeal creatures. Finally, God produces all corporeal creatures via the diffusion of light in space. The roots of Patrizi's view can be found, not just in Plotinus and Proclus, but also in the medieval Jewish philosopher Ibn Gabirol and the Christian Robert Grosseteste (§3) (see Illumination).

Part Two is called 'Panarchia' (all-principle) and discusses the first principles that stem from God, who is One but who is also called the One-all. The highest principles are unity, essence, life and intelligence. The four lowest principles are nature, quality (see Ficino, M. §3), form and body. The intermediate principle is soul, which is midway between spirit and matter. In his discussion of this hierarchy of being, Patrizi was mainly concerned with the higher principles and the nature of God.

Part Three is called 'Pampsychia' (all-soul), and discusses all types of soul, from the irrational souls of plants to the world soul which animates the entire cosmos.

The fourth part is called 'Pancosmia' (all-world), and it is here that we find Patrizi's views on mathematics and natural science.

4 Philosophy of nature

Patrizi's view of space, which he called a 'corporeal incorporeal', is fundamental to his philosophy of nature.

Space has to be corporeal because it is extended or three-dimensional; it has to be incorporeal because it lacks both resistance and density, and can serve as a container for fully corporeal things. A similar analysis can be applied to light, which Patrizi also called a ‘corporeal incorporeal’; both space and light can then be seen as borderline entities between the realms of matter and spirit. Indeed, the analysis can be extended to soul, which also falls between the realms of matter and spirit (see §3), because soul is diffused throughout the physical world without itself being a physical thing.

Patrizi argued that space is prior to ordinary corporeal objects both metaphysically and temporally, and that it is a condition of the existence of other things. He was clearly influenced here by Plato’s *Timaeus* (see Plato §16), but his presentation was strengthened by his systematic, principled, rejection of Aristotle’s conclusions about place. Aristotle was confused because he had no notion of an incorporeal corporeal and did not see that what distinguishes space from body is not merely dimensionality but resistance.

Patrizi’s theory of space led him to new views of the void. Most notably, he called for an actually infinite void outside the physical world. Betraying some confusion (since an infinite space will have no centre), he added that it must be round, with the rotating earth as its centre. He also accepted an inter-particulate void, made of the naturally existing spaces between corpuscular bodies. Finally, like Telesio, he took stock arguments about the possibility of a naturally occurring intra-mundane void or vacuum, and used them against Aristotle’s denial of such a vacuum. He included reference to empirical data such as the operation of bellows and water clocks and what happens to closed containers of frozen water, but no actual experiment seems to have been involved. Theoretical reasoning predominated, and he did not really go beyond knowledge already available from ancient and Arabic sources. Both here and in his general discussion of space, Patrizi can be seen as preparing the way for modern science, but his role should not be overemphasized.

Another part of Patrizi’s philosophy of nature involved replacing Aristotle’s four basic qualities, hot, cold, wet and dry, with the four principles space, light (*lumen*), heat and *fluor* (fluidity, flux), which combine to form a hierarchy of mixed bodies. Particularly important here is the claim that *fluor* offers resistance to *lumen*. A corollary of this list of basic principles which were not hard or solid, was Patrizi’s claim that the heavens were fluid, and that the heavenly bodies were not fixed in hard, solid spheres as had commonly been thought. It is important to note that while Patrizi’s contemporary Tycho Brahe had astronomical arguments for the same view, Patrizi’s arguments were basically metaphysical.

A cautious assessment is also appropriate when one considers Patrizi’s praise of mathematics, especially geometry, for this seems more closely linked with his belief in an intelligible world and ideal forms than with any belief that mathematics could provide a tool for developing philosophy or for explaining the physical world.

Patrizi seems to have been more influential than the other so-called philosophers of nature, Bruno, Campanella and Telesio, and he was widely read in the seventeenth century. His theory of light was attacked by Mersenne (§3), while his views on space may have been handed on to Newton (§4) through the Cambridge Platonist, Henry More (§5), as well as Gassendi (§4).

See also: Hermetism; Neoplatonism; Platonism, Renaissance; Telesio, B.

E.J. ASHWORTH

List of works

Patrizi da Cherso, F. (1553) *Discorso della diversità dei furori poetici (Discourse on Types of Poetic Inspiration)*, Venice: G. Griffio.(Included in a volume with a series of other works, starting with *La città felice*.)

Patrizi da Cherso, F. (1560) *Della historia (On History)*, Venice: A. Arrivabene.

Patrizi da Cherso, F. (1562) *Della retorica (On Rhetoric)*, Venice: F. Senese.

Patrizi da Cherso, F. (1529-97) *L’amorosa filosofia (The Philosophy of Love)*, ed. J.C. Nelson, Florence: Istituto Nazionale di Studi sul Rinascimento, 1963.

Patrizi da Cherso, F. (1581) *Discussionum peripateticarum tomi quattuor (Peripatetic Discussions in Four Books)*, Basle.(On the limits of Aristotelianism as a basis for Christian theology.)

Patrizi da Cherso, F. (1586) *Della poetica (On the Art of Poetry)*, ed. D. Aguzzi Barbagli, Florence: Istituto Nazionale di Studi sul Rinascimento, 1969-71, 3 vols.

Patrizi da Cherso, F. (1591) *Nova de universis philosophia (New Philosophy of Universes)*, Ferrara.(The Venice 1593 edition was published after Patrizi had been criticized by the Congregation of the Index.)

Patrizi da Cherso, F. (1970) *Emendatio in libros suos Novae philosophiae (Amendments to the Books on the New Philosophy)*, in P.O. Kristeller (ed.) *Rinascimento* 10: 215-18.

Patrizi da Cherso, F. (1975) *Lettere ed opuscoli inediti (Unpublished Letters and Short Works)*, ed. D. Aguzzi Barbagli, Florence: Istituto Nazionale di Studi sul Rinascimento.(The opuscula include a dialogue on love, a discussion of the order of Plato's dialogues, and works on hydraulics.)

References and further reading

Antonaci, A. (1984) *Ricerche sul neoplatonismo del Rinascimento: Francesco Patrizi da Cherso (Investigations of Neoplatonism in the Renaissance)*, Galatina (Lecce): Salentina.(About the *Discussiones* and their background.)

Henry, J. (1979) 'Francesco Patrizi da Cherso's concept of space and its later influence', *Annals of Science* 36: 549-75.(Referred to in §2. A detailed discussion of Patrizi's role in the development of concepts of void space and an infinite universe.)

Kristeller, P.O. (1964) *Eight Philosophers of the Italian Renaissance*, Stanford, CA: Stanford University Press. (Contains a readable introduction to Patrizi's thought.)

Purnell, F. (1976) 'Francesco Patrizi and the critics of Hermes Trismegistus', *Journal of Medieval and Renaissance Studies* 6: 155-78.(Patrizi's response to criticisms of the Hermetic writings.)

Schmitt, C.B. (1967) 'Experimental Evidence for and against a Void: The Sixteenth-Century Arguments', *Isis* 63: 352-66; repr. in C.B. Schmitt, *Studies in Renaissance Philosophy and Science*, London: Variorum, 1981. (Useful material on Patrizi's philosophy of nature.)

Wilmott, M.J. (1985) '"Aristoteles exotericus, acroamaticus, mysticus": two interpretations of the typological classification of the "Corpus Aristotelicum" by Francesco Patrizi da Cherso', *Nouvelles de la république des lettres* 2: 67-95. (Patrizi's re-evaluation of Aristotle's texts.)

Paul of Venice (1369/72-1429)

Like other teachers in fifteenth-century Italian universities, Paul of Venice focused on logic and natural philosophy in an undergraduate programme directed toward the education of medical students. Despite Paul's theological training and important position in the order of Augustinian friars, nearly all his works are non-theological. His prolific writings popularized the achievements of Oxford logic and Parisian physics in a framework derived from Aristotle and Averroes. As a philosopher he is best known for his Averroist position on the human soul, and for his moderate realism with respect to universals.

Paul of Venice (Paolo Nicoletti Veneto, born Udine 1369/1372, died Padua 1429) was active in three spheres. As an Augustinian friar, he served as Provincial of his Order. As a diplomat, he served as an ambassador of the Venetian Republic to Germany, Hungary and Poland. His most important career, however, was as a university teacher and author of textbooks. He taught briefly at various Italian universities, but his most constant association was with the University of Padua. He studied at the Augustinian *studium* in Padua both before and after his theological studies in Oxford (1390-3), and he taught at the University itself for many years. His academic concerns are reflected in the nature and number of his writings.

By far the most popular of all his works was the early *Logica parva*, which survives in over eighty manuscripts and was still being printed at the end of the sixteenth century. It shows the clear influence of his stay in England, for it takes the form of the textbooks then in use at Oxford and Cambridge. It begins with a tract called *Summule*, which summarizes standard medieval material on terms, propositions and arguments, and then offers a series of short treatments of supposition theory, consequences, the proof of terms, obligations and insolubles (see Logic, medieval). Paul uses sophismata, or puzzle-cases, to illuminate logical points, a standard technique in English logic texts. One of his less popular logic texts, the *Sophismata*, is entirely devoted to such puzzles, and they also form a large part of his *Logica magna*. The latter work, immensely long, was little read in Paul's time, but in the twentieth century a good deal of attention has been paid to it because it summarizes much of the logical work of the second half of the fourteenth century. Here as elsewhere, Paul often copied his sources word for word, and his own contribution was largely limited to modifying or expanding the views of others. Thus his solution to the problem of insolubles, which differs from the one he gives in the *Logica parva*, is a modification of Roger Swyneshed's solution (see Language, medieval theories of; Oxford Calculators).

The most popular of Paul's logic works after the *Logica parva* is a commentary on Aristotle's *Posterior Analytics* which, as a discussion of scientific reasoning, supplements his equally popular *Summa philosophie naturalis (Summa of Natural Philosophy)*. Each of these works survives in over fifty manuscripts and a number of printed editions. The *Summa* begins with a discussion of physics, and ends with short treatises on the soul and on metaphysics. In the latter, Paul uses material on the latitudes of forms drawn from the Oxford Calculators and the Parisian John of Ripa to explain the hierarchy of being. He argued that, contrary to the traditional theological view, God is not properly regarded as the measure of creatures since he is infinitely distant from them. Instead, finite things are ordered in the hierarchy of being in accordance with their distance from the bottom of the hierarchy, the zero grade of being.

One of Paul's chief philosophical concerns was the nature of universals, an issue he discussed in various places, including the *Summa philosophie naturalis (Summa of Natural Philosophy)*. His intention was to reconcile the 'modern' view, that knowledge begins with the cognition of singulars, with a realist view of universals that he saw as more consonant with Aristotle and Averroes (see Ibn Rushd). He argued that universals have real existence independent of the work of the intellect, but that their existence apart from singular things is only potential. His realist view of universals affected his attitude to Aristotle's categories, which he saw as reflecting the ontological structure of reality, and it is noteworthy that a large part of the discussion of metaphysics in the *Summa* is devoted to the categories (see Universals).

Another of Paul's concerns in the *Summa* was the nature of the human soul. He adopted the Averroist thesis that the possible intellect is one for all human beings, but argued that this one soul was individuated by becoming the substantial form of individual human beings. As a result, my knowledge can indeed be different from your knowledge. Only in his later commentary on Aristotle's *De anima* (the *Scriptum super libros De anima*) did Paul

worry about the compatibility of what he took to be the correct interpretation of Aristotle and Averroes with Christian doctrine. This commentary also reveals that Paul had changed his mind about the agent intellect. In the *Summa* he argued that it was the same as the possible intellect, but now he argued that it was to be identified with God. A further feature of Paul's Averroism was his acceptance of the thesis that the sensitive soul which we have in common with animals and the intellective soul which characterizes human beings are two separate souls (see Averroism).

The tension in Paul's works between the positions of Aristotle and Averroes and the rather different results of fourteenth-century Oxford and Paris is well illustrated by his changing position on impetus theory. In his *Summa*, he gave an account of projectile motion in terms of the impetus the projected object receives from its projector, a doctrine that he took largely from the Parisians John Buridan and Albert of Saxony. In his commentary on Aristotle's *Physics*, which was written in 1409, he reverted to the Aristotelian view that natural motion depends on the continued activity of a mover (see Natural philosophy, medieval). Despite Paul's often-changing opinions and his frequent lack of originality, his work was very influential in Italy during much of the fifteenth and sixteenth centuries.

See also: Aristotelianism, medieval; Aristotelianism, Renaissance; Averroism; Logic, medieval; Natural philosophy, medieval

E.J. ASHWORTH

List of works

Note: a number of Paul's works, including the *Scriptum super libros De anima* and the *Sophismata*, are not available in any modern edition; see Bottin (1983) for a list of early printed editions.

Paul of Venice [Paulus Venetus] (1395-6) *Logica parva*, trans. A.R. Perreiah, Munich, Vienna: Philosophia Verlag, 1984.(Translation and thematic analysis of Paul's most popular logical work; unreliable on the historical background.)

Paul of Venice [Paulus Venetus] (1397-8) *Logica magna*, ed. and trans. in a series of volumes, Oxford: Oxford University Press for the British Academy, 1978-.(Several volumes, some without much scholarly apparatus, have appeared so far.)

Paul of Venice [Paulus Venetus] (1397-8) *Logica magna (Tractatus de suppositionibus)*, ed. and trans. A.R. Perreiah, St Bonaventure, NY: The Franciscan Institute, 1971.(Contains part of Paul's discussion of supposition theory: the omissions are not indicated.)

Paul of Venice [Paulus Venetus] (1408) *Summa philosophie naturalis (Summa of Natural Philosophy)*, Hildesheim: Georg Olms, 1974.(Facsimile reprint of 1503 edition; difficult to read. A summary of natural philosophy, including *On the Soul* and *Metaphysics*.)

References and further reading

Bottin, F. (1983) 'Logica e filosofia naturale nelle opere di Paolo Veneto' ('Logic and Natural Philosophy in the Works of Paul of Venice'), in A. Poppi (ed.) *Scienza e Filosofia all'Università di Padova nel Quattrocento*, Trieste: Lint, 85-124.(A useful discussion of major themes; gives an accurate list of Paul's works and their early printed editions.)

Conti, A. (1990) *Johannes Sharpe: Quaestio super Universalia*, Florence: Olschki. (Contains a partial edition of Paul's Question on universals, together with discussion and references.)

Conti, A. (1996) *Esistenza e Verità: Forme e Strutture del Reale in Paolo Veneto e nel Pensiero Filosofico del Tardo Medioevo (Existence and Truth: Forms and Structures of Reality in Paul of Venice and the Philosophical Thought of the Late Middle Ages)*, Rome: Instituto Storico Italiano per il Medio Evo.(The first full study of Paul's metaphysics, with special emphasis on being and existence, universals and singulars, and problems of individuation.)

Olivieri, L. (ed.) (1983) *Aristotelismo Veneto e Scienza Moderna (Venetian Aristotelianism and Modern Science)*, Padua: Antenore. (Contains articles on different aspects of Paul's work.)

Perreiah, A.R. (1986) *Paul of Venice: A Bibliographical Guide*, Bowling Green, OH: Philosophy Documentation Center.(A bibliography of manuscripts only; also includes a biography of Paul and a contentious discussion of the authorship of the *Logica magna*.)

Pecham, John (*c.*1230-92)

John Pecham, an English Franciscan, taught at Paris and Oxford, and died as Archbishop of Canterbury. His philosophical career represents a concentrated effort to defend the traditional views of Augustine and Anselm (among other theologians) against what was perceived as a growing tendency toward heterodox Aristotelianism, exemplified in such doctrines as the eternity of the world, a single intellect for all humankind and a divinity that had no knowledge of individual beings.

Born in Patcham in Sussex, England, Pecham was educated at the Benedictine monastery at Lewes and joined the Franciscans at Oxford during the 1250s. After continuing his education at Oxford, he was sent to Paris in the 1260s to complete his theological studies and in 1270 became regent master in Paris in the Franciscan chair of theology. Sometime after 1271 he returned to Oxford, where he became the Franciscan regent master of theology. He held that position until 1275, when he was elected minister provincial of the Franciscans in England, and in 1277 he was appointed lecturer to the papal curia. In 1279 he was named Archbishop of Canterbury, an office he held until his death.

As a student in the late 1260s, Pecham had undoubtedly heard the sermons of Bonaventure, who had alerted his listeners to the growing threat of unorthodox Aristotelianism. If Pecham was not directly involved in compiling the list of thirteen errors condemned in 1270 by Etienne Tempier, Bishop of Paris, he certainly would have agreed that the condemned propositions were erroneous. Together with William de la Mare, Pecham was one of the first Franciscans to oppose Thomas Aquinas, whose opinions were viewed as compromising Christian doctrine and too deferential to Aristotle and Averroes (see Ibn Rushd). This tendency continued in the followers of Pecham, namely Matthew of Aquasparta, Roger Marston, Bartholomew of Bologna, William of Falagar and, later, Vital du Four.

A true follower of Bonaventure, Pecham shows a fundamental allegiance to Augustine while accommodating the philosophy of Aristotle where possible. In the critically edited texts of the surviving works, Pecham shows little interest in logic or metaphysics. His writings reveal a preoccupation with the theory of knowledge, philosophical psychology, natural philosophy and science. In his theory of knowledge, Pecham supported Augustinian divine illumination with regard to first principles, claiming that the human intellect needed the ‘eternal reasons’ for the certitude of intellectual and moral first principles, although not for their contents (see Augustinianism). In contract to Aquinas, Pecham held that the human (as well as the angelic and divine) intellect had a direct knowledge of singulars. According to Aquinas’ interpretation of the Aristotelian thesis that the senses know the singular and the intellect the universal, the intellect must reflect on phantasms, products of the imagination, which receives its images from the senses, and then abstract the universal from the phantasms. In response, Pecham argues that the intellect abstracts either knowingly or unknowingly. If knowingly, then it had direct knowledge of the singular in the first place; if unknowingly, then how can it be called intellectual knowledge at all?

In addition to a set of disputed questions on the soul (*Quaestiones de anima*), Pecham also wrote a treatise on the soul (*Tractatus de anima*). In the former, he dismisses traducianism, which would have the human soul come from the divine substance, or from the bodies of the parents, or develop from a sensitive soul. Nor were all human souls created at the beginning of time, as Origen had held. Rather, each soul is created directly by God and infused into the body. Pecham vigorously defends the immortality of the soul, claiming that this can be demonstrated by seven irrefutable arguments (see Soul, nature and immortality of the). At the same time, he is opposed to multiple souls in the human being. There is only one (intellective) human soul, which, however, encapsulates the vegetative and sensitive functions as grades of a single intellectual ‘form’. He strenuously attacks as heretical Averroes’ denial that each human being had its own rational soul. Such an opinion jeopardized immortality and rendered the statement ‘I understand’ impossible, as Aquinas had pointed out.

The powers of the soul, though multiple, are not really distinct from one another or from the soul. The vegetative grade of the form is distinguished into nutritive, augmentative and generative. The sensitive grade has motive and apprehensive powers, the latter being distinguished into external (the five senses) and internal powers, the latter including the ‘common sense’, the imagination, the estimative (determining what is friendly or hostile) and the memory. The intellectual soul likewise has apprehensive and motive powers, the apprehensive being the agent intellect, the ‘possible’ intellect and the intellectual memory. The rational appetite or will comprises concupiscible

and irascible powers whereby it seeks the good and flees the harmful, powers which it shares with the sensitive grade. Freedom is not a separate power of the will; nevertheless, the will is so free that it can withhold consent in the face of the dictates of the (practical) intellect. This virtual containment of the various powers in the one intellectual soul anticipates, it would seem, the formal distinction of Duns Scotus.

In natural philosophy, Pecham opposes Aquinas on several issues. In Aristotle's view, every composite substance was made up of matter and form. Aquinas had held that prime matter, as the basis for substantial change, was pure potentiality. Pecham held that it was a positive identity, essentially and really distinct, so that by his divine power God could create prime matter distinct from any form whatever (*Quodlibeta quatuor* IV q.1). Aquinas likewise had held that there was but one substantial form in the human being uniting the soul to the body. Pecham responded with his theory of multiple grades, vegetative and sensitive, which persisted as (substantial) components of the human composite (*Quodlibeta quatuor* IV q.25). Thus the bodily, vegetative and sensitive 'forms' are not successively 'corrupted out' by the advent of the higher forms, as Aquinas would have it, but remain as grades of the higher form. Pecham may have been the first to introduce the grades theory as a refinement of 'tri-animism' or the plurality of substantial forms. While his treatise on this subject is apparently lost (Douie 1952: 280 n. 2), the salient points of his theory undoubtedly survive in his faithful disciple Roger Marston's *Quodlibet* II q. 22 (see Natural philosophy, medieval). Pecham likewise rejected Aquinas' view on the eternity of the world. The latter held that although *de facto* the world was created in time and there is nothing theologically or philosophically repugnant to the world's being created from all eternity (Brady 1974; Bukowski 1979). Like Bonaventure, Pecham believed that creation from all eternity was fundamentally contradictory (see Eternity of the world, medieval views of).

Pecham wrote a treatise on optics as well as a textbook on optics, the *Perspectiva communis*, and also a treatise on the spheres, a set of four disputed questions on the stars and a tract on mystical numbers (*De numeris misticis*). In addition he wrote numerous treatises on Franciscan spirituality (Brady 1974; Teetaert 1933) including a tract on evangelical poverty. Much of Pecham's work survives only in manuscripts, namely a considerable number of disputed questions and his commentary on the *Sentences* of Peter Lombard, the last three books of which appear to have been lost. A comprehensive assessment of his thought must await the critical edition of these works.

See also: Aristotelianism, medieval; Augustinianism; Bonaventure; Eternity of the world, medieval views of; Soul, nature and immortality of the

GIRARD J. ETZKORN

List of works

Pecham, J. (1269-73) *De numeris misticis (On Mystical Numbers)*, ed. B. Hughes, *Archivum Franciscanum Historicum* 78, 1985: 3-28, 333-83.(Edition of Pecham's tract on mystical numbers.)

Pecham, J. (1277-9) *Perspectiva communis*, ed. D. Lindberg, *John Pecham and the Science of Optics*, Madison, WI: University of Wisconsin Press, 1970.(Includes brief biographical sketch in English and extensive bibliography.)

Pecham, J. (1269-73) *Quaestio disputata 'De aeternitate mundi' (A Disputed Question on the Eternity of the World)*, ed. O. Argerami, *Patristica et Mediaevalia* 1, 1975: 82-100.(Edition of Pecham's disputed question on eternity.)

Pecham, J. (1269-73) *Quaestiones de aeternitate mundi (Questions on the Eternity of the World)*, ed. I. Brady, 'John Pecham and the Background of Aquinas's *De aeternitate mundi*, in *St. Thomas Aquinas (1274-1974): Commemorative Studies*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, vol. 2, 11-71; ed. and trans. V. Potter, New York: Fordham University Press, 1993.(Further questions on the eternity of the world.)

Pecham, J. (1269-73) *Quaestiones de anima (Questions on the Soul)*, ed. H. Spettmann, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters 19, 5, 1918.(Also includes *Quaestiones de beatitudine corporis et animae* and *Quaestiones de anima* excerpted from Pecham's commentary on Book 1 of Lombard's *Sentences*.)

Pecham, J. (1269-73) *Quodlibeta quatuor (Four Quodlibets)*, ed. G. Etzkorn and F. Delorme, Bibliotheca Franciscana Scholastica Medii Aevi XXV, Grottaferrata: Collegio S. Bonaventura, 1989.(Includes copious cross-references to parallel treatment in Pecham and near contemporaries.)

Pecham, J. (1269-73) *Summa de esse et essentia (Summa on Existence and Essence)*, ed. F. Delorme, Rome: Studi Franciscani, 1928.(Edition of this metaphysical treatise.)

Pecham, J. (1269-73) *Tractatus de anima (Treatise on the Soul)*, ed. G. Melani, Studi Francescani, vol. 1,

Florence: Edizioni Studi Francescani, 1948.(The immortality and powers of the soul.)
Pecham, J. (1269-73) *Tractatus de perspectiva (Treatise on Optics)*, ed. D. Lindberg, St Bonaventure, NY: Franciscan Institute Publications, 1972.(Edition of treatise on optics.)
Pecham, J. (1279-92) *Registrum epistolarum Fratris Ioannis Peckham (Register of Letters of Brother John Pecham)*, ed. C.T. Martin, 3 vols. in Rerum Britannicarum Scriptores, London, 1788.(Edition of Pecham's collected correspondence.)

References and further reading

Brady, I. (1974a) 'Jean de Pecham', *Dictionnaire de Spiritualité* 8: 645-9. (Concentrates on Pecham's spiritual writings.)
Brady, I. (1974b) 'John Pecham and the Background of Aquinas's De aeternitate mundi', *St. Thomas Aquinas (1274-1974): Commemorative Studies*, vol. 2: Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 11-71. (Study plus text of Pecham's *Disputed Questions on the Eternity of the World*.)
Bukowski, T. (1979) 'J. Pecham, T. Aquinas et al. on the Eternity of the World', *Recherches de Théologie Ancienne et Médiévale* 46: 216-21.(Disputes Brady's claim that Aquinas directed his *De aeternitate* treatise against Pecham.)
Callebaut, A. (1925) 'Jean Peckham O.F.M. et l'augustinisme', *Archivum Franciscanum Historicum* 18: 441-72. (Pecham's place in Augustinian thought.)
Crowley, T. (1951) 'John Peckham O.F.M. Archbishop of Canterbury: Versus the New Aristotelianism', *Bulletin of the John Rylands Library* 33: 242-55.(On Pecham's role as an opponent of Aquinas and his followers.)
Doucet, V. (1933) 'Notulae bibliographicae de quibusdam operibus Fr. Ioannis Pecham, O.F.M.', *Antonianum* 8: 307-28, 425-59.(Best study of Pecham's works and manuscripts.)
Douie, D. (1952) *Archbishop Pecham*, Oxford: Clarendon Press.(Focuses on Pecham's career as archbishop.)
Ehrle, F. (1889) 'J. Peckham über den Kampf des Augustinismus und Aristotelismus in der zweiten Hälfte des 13. Jahrhunderts' (J. Pecham and the Debate over Augustinianism and Aristotelianism in the Second Half of the Thirteenth Century), *Zeitschrift für katholische Theologie* 13: 172-93.(Discusses Pecham's own role in the debate.)
Etzkorn, G. (1989) 'John Pecham, O.F.M.: A Career of Controversy', in E.B. King *et al.* (eds) *Monks, Nuns, and Friars in Mediaeval Society*, Sewanee, TN: The Press of the University of the South, 71-82.(Pecham opposing the secular masters and Aquinas.)
Lindberg, D. (1965) 'The "Perspectiva Communis" of John Pecham: Its Influence, Sources and Content', *Archives Internationales d'Histoire des Sciences* 18: 37-53.
Lindberg, D. (1968) 'Bacon, Witelo, and Pecham, the Problem of Influence', *XII Congres Internationale d'Histoire des Sciences* 111-A: 103-7.(Article discussing Pecham's influence.)
Spettmann, H. (1919) 'Die Psychologie des Johannes Pecham', Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters 20: 1-102.(On Pecham's epistemology.)
Teetaert, A. (1933) 'Pecham, Jean', *Dictionnaire de Théologie Catholique* 12: 100-14.(In French; most comprehensive and thorough.)
Wielockx, R. (1985) 'Jean de Pecham', *Catholicisme* X: 1005-7.(Short survey.)

Peirce, Charles Sanders (1839-1914)

Peirce was an American philosopher, probably best known as the founder of pragmatism and for his influence upon later pragmatists such as William James andJohn Dewey. Personal and professional difficulties interfered with his attempts to publish a statement of his overall philosophical position, but, as the texts have become more accessible, it has become clear that he was a much more wide-ranging and important thinker than his popular reputation suggests.

He claimed that his pragmatism was the philosophical outlook of an experimentalist, of someone with experience of laboratory work. His account of science was vigorously anti-Cartesian: Descartes was criticized for requiring an unreal 'pretend' doubt, and for adopting an individualist approach to knowledge which was at odds with scientific practice. 'Inquiry' is a cooperative activity, whereby fallible investigators progress towards the truth, replacing real doubts by settled beliefs which may subsequently be revised. In 'The Fixation of Belief' *(1877), he compared different methods for carrying out inquiries, arguing that only the 'method of science' can be self-consciously adopted. This method makes the 'realist' assumption that there are real objects, existing independently of us, whose nature will be discovered if we investigate them for long enough and well enough.*

Peirce's 'pragmatist principle' was a rule for clarifying concepts and hypotheses that guide scientific investigations. In the spirit of laboratory practice, we can completely clarify the content of a hypothesis by listing the experiential consequences we would expect our actions to have if it were true: if an object is fragile, and we were to drop it, we would probably see it break. If this is correct, propositions of a priori metaphysics are meaningless. Peirce applied his principle to explain truth in terms of the eventual agreement of responsible inquirers: a proposition is true if it would be accepted eventually by anyone who inquired into it. His detailed investigations of inductive reasoning and statistical inference attempted to explain how this convergence of opinion was achieved.

Taken together with his important contributions to formal logic and the foundations of mathematics, this verificationism encouraged early readers to interpret Peirce's work as an anticipation of twentieth-century logical positivism. The interpretation is supported by the fact that he tried to ground his logic in a systematic account of meaning and reference. Much of his most original work concerned semiotic, the general theory of signs, which provided a novel framework for understanding of language, thought and all other kinds of representation. Peirce hoped to show that his views about science, truth and pragmatism were all consequences of his semiotic. Doubts about the positivistic reading emerge, however, when we note his insistence that pragmatism could be plausible only to someone who accepted a distinctive form of metaphysical realism. And his later attempts to defend his views of science and meaning bring to the surface views which would be unacceptable to an anti-metaphysical empiricist.

From the beginning, Peirce was a systematic philosopher whose work on logic was an attempt to correct and develop Kant's philosophical vision. When his views were set out in systematic order, positions came to the surface which, he held, were required by his work on logic. These include the theory of categories which had long provided the foundations for his work on signs: all elements of reality, thought and experience can be classified into simple monadic phenomena, dyadic relations and triadic relations. Peirce called these Firstness, Secondness and Thirdness. He also spoke of them as quality, reaction and mediation, and he insisted that the error of various forms of empiricism and nominalism was the denial that mediation (or Thirdness) was an irreducible element of our experience. Peirce's 'synechism' insisted on the importance for philosophy and science of hypotheses involving continuity, which he identified as 'ultimate mediation'. This emphasis upon continuities in thought and nature was supposed to ground his realism. Furthermore, his epistemological work came to focus increasingly upon the requirements for rational self-control, for our ability to control our inquiries in accordance with norms whose validity we can acknowledge. This required a theory of norms which would explain our attachment to the search for truth and fill out the details of that concept. After 1900, Peirce began to develop such an account, claiming that logic must be grounded in ethics and aesthetics.

Although pragmatism eliminated a priori speculation about the nature of reality, it need not rule out metaphysics that uses the scientific method. From the 1880s, Peirce looked for a system of scientific metaphysics that would fill important gaps in his defence of the method of science. This led to the development of an evolutionary cosmology,

an account of how the world of existent objects and scientific laws evolved out of a chaos of possibilities through an evolutionary process. His 'tychism' insisted that chance was an ineliminable component of reality, but he argued that the universe was becoming more governed by laws or habits through time. Rejecting both physicalism and dualism, he defended what he called a form of 'Objective Idealism': matter was said to be a form of 'effete mind'.

1 Life and works

Peirce's father, Benjamin Peirce, was one of the most respected American mathematicians of the nineteenth century. As well as holding chairs in mathematics and astronomy at Harvard, he was instrumental in establishing the American Academy of Arts and Sciences and served as Superintendent of the United States Coast Survey. Charles, whose remarkable intellectual abilities were soon recognized, was brought up in Cambridge, Massachusetts, in a house often visited by the leading philosophical and scientific figures of the time. Indulged and encouraged, his abilities were evident even in his most youthful writings. His undergraduate career at Harvard (1855-9) was undistinguished, but he was subsequently the first student to graduate *summa cum laude* in chemistry from the University's new Lawrence Scientific School.

By the late 1860s, Peirce's future looked bright. From 1861 he had worked as an aide to the Coast Survey. He lectured on logic at Harvard several times between 1865 and 1869. By 1867, his first important series of philosophical papers was appearing and he was elected to the American Academy of Arts and Sciences. In 1872 he founded a 'Metaphysical Club' where he discussed philosophical issues with kindred spirits, including William James, Chauncey Wright, Francis Abbot, and Oliver Wendell Holmes Jr. Although some friends already feared that Peirce's independence of mind and his difficult and wild personality would prove obstacles to academic success, his future seemed assured when, in 1879, he obtained a lectureship in logic in the new Graduate School at Johns Hopkins University. At Johns Hopkins, Peirce and a small group of talented students made important contributions to the logic of relations and the theory of probabilistic reasoning, and (independently of Frege) Peirce and his student O.H. Mitchell introduced quantifiers into logic (see Predicate calculus).

In the mid-1880s Peirce's professional life fell apart. His post at Johns Hopkins was suddenly terminated due to his personal irregularities, and his unreliability soon led to the end of his employment with the Coast Survey. Retreating to rural Pennsylvania, Peirce built a house where he lived with his second wife, often in desperate poverty, until his death in 1914. He was not wholly cut off from the academic world. His friendWilliam James arranged for him to give occasional series of lectures in Cambridge, Massachusetts - although he faced the widespread suspicion that he would be a bad moral influence upon the young students of Harvard. He continued to publish and he wrote many reviews for *The Nation*; and he produced eighty-thousand manuscript pages which are a major source for scholars trying to understand his later philosophical position. In 1903 he entered into an important correspondence with Victoria Lady Welby, an English scholar whose work in 'Significs' suggested a common interest with Peirce the semiotician. Lady Welby introduced Peirce's work to I.A. Richards and C.K. Ogden, whose account of them, *The Meaning of Meaning*, helped bring them to the attention of the wider philosophical community.

A finished statement of Peirce's philosophical system was never completed, but his published writings and manuscripts are extensive and range widely. During the late 1860s, a series of papers appeared, in the *Journal of Speculative Philosophy* and elsewhere, which summarized his early ideas on cognition, inference, logic, signs and the theory of categories. Attempts to incorporate this material into a textbook on logic in the early 1870s were thwarted, in part because of intractable problems about reference. However, material intended for this text surfaced, in a less systematic form, in Peirce's most famous series of papers, *Illustrations of the Logic of Science*, which appeared in the *Popular Science Monthly* in 1877-9. This included his best-known papers, 'The Fixation of Belief' and 'How to Make Our Ideas Clear', the standard source for his pragmatism. The most important work from the early 1880s was in formal logic, and *Studies in Logic*, a book of essays by Peirce and his students at Johns Hopkins, contained the Peircean version of the logic of quantification - which provided the key to solving his problems about reference.

The second half of the 1880s saw Peirce's interest turn to metaphysics and the revision of his theory of categories. An important series of five papers on metaphysical topics appeared in *The Monist* in 1891-3. Several attempts to collect his earlier papers in book form and to write a logic text survive from the following decade, including the

important *Grand Logic* of 1893. Constantly seeking funding to support various grandiose publishing ventures, Peirce made a very useful application to the Carnegie Foundation for support to write thirty-six ‘memoirs’ summing up his philosophical position in 1902. Although unsuccessful, the surviving drafts of the application cast invaluable light upon the structure of his thought. Also useful are lectures delivered at or around Harvard during this period. A fascinating series dealing with logical and metaphysical topics was given to the Cambridge Conferences in 1898 and has been published as *Reasoning and the Logic of Things*. After William James had made pragmatism famous, Peirce devoted much energy to displaying the superiority of his version of the doctrine, and he valuably seized the opportunity when invited to lecture on pragmatism at Harvard in 1903. He returned to the task of proving pragmatism in an incomplete series of papers published in *The Monist* a few years later. And it was around this time, in 1908, that he offered his ‘Neglected Argument for the Reality of God’: his view of God was somewhat pantheist, and he shared his father’s view of scientific research as an attempt to read God’s ‘great poem’, holding that scientific observation was a kind of religious experience.

2 Inquiry and the fixation of belief

In his writings from the 1860s and 1870s, Peirce explicitly questioned Cartesian assumptions about cognition. The papers in the *Journal of Speculative Philosophy* challenged the claim that we have introspective knowledge of our own mental states and disputed the foundationalist view that we have ‘intuitive’ knowledge - knowledge which is not logically dependent on prior opinions. The influence of the common-sense tradition emerges when Peirce criticizes Descartes’ use of the method of doubt. Cartesian doubt is impossible: it will be ‘mere self-deception and not real doubt’. Peirce enjoins us not to pretend to doubt in philosophy what we do not doubt in our hearts. Cartesians find the ultimate test of certainty in the individual consciousness, while Peirce urges us to follow the successful sciences in trusting to the critical conversation of a community of inquirers. And philosophy should follow the sciences in trusting the ‘multitude and variety of its arguments rather to the conclusiveness of any one’. Our experience of the growth of knowledge suggests that although we are fallible, we can make progress as members of a community of investigators. Descartes and his followers have provided no reason for us not to emulate this in philosophy. Peirce’s work on logic and epistemology attempted to vindicate this fallibilist but optimistic view of scientific inquiry. In the 1860s, he tried to do this by arguing that all thoughts are signs and all mental action is inference, and then explaining the grounds of our trust in the forms of reasoning that are used in science. We shall focus here on the arguments he employed in the following decade.

In ‘The Fixation of Belief’ (1877), Peirce is concerned with ‘guiding principles of reasoning’. These are propositions which formulate rules of inference: an inference is good if its guiding principle is true. Habits of reasoning express such guiding principles and almost any proposition could express one: to the proposition that all humans are mortal corresponds the rule that on hearing of the humanity of someone, their mortality may be inferred. However, Peirce is concerned with the special class of logical rules, of propositions that are ‘absolutely essential as guiding principles’ and he tells us that these are ‘necessarily taken for granted in asking whether a certain conclusion follows from certain premises’. The papers in the *Illustrations of the Logic of Science* attempt to draw out facts and principles which ‘are deduced from the assumptions which are involved in the logical question’, and which we must ‘already know before we can have any clear conception of reasoning at all’. He is hunting for the presuppositions of inquiry and with the fundamental norms which govern our participation in it.

The paper begins by characterizing inquiry which, we are told, starts with the posing of a question, with a real doubt, and concludes with the settled acceptance of an answer to the question, with belief. Both belief and doubt are characterized in a functionalist manner: beliefs are settled states which ‘guide our desires and shape our actions’; doubt is an unsettled state whose only effect on action is to provoke inquiry directed at its elimination. A genuine doubt is required to motivate inquiry (contrary to Cartesian doctrine) and once the doubt has been eliminated, inquiry comes to an end: ‘the sole object of inquiry is the settlement of opinion’. Since to believe something is always to believe it to be true, it is empty to say that our aim in inquiry is to arrive at the truth.

What methods should we use in order to settle opinions or fix beliefs? Peirce considers four: three prove unsatisfactory and the method of science is endorsed. It is the only method which is consistent with the presuppositions of inquiry, so the norms it provides are ‘absolutely essential’. Were we to adopt the method of tenacity, we would seize on any answer to our question, doing all that was required to ignore or resist anything that might persuade us to abandon it. The ‘social impulse’ prevents us from making this our method: our certainty will

inevitably be disturbed when we encounter people who hold other opinions and it is inevitable that we will do so. Proponents of the method of authority would allow a monarch or religious leader to choose an answer for an entire community and would control sources of available information to prevent our certainties being shaken. But this too will fail: no one could control opinions on *every* question, and we would eventually face problems through encountering people who are not subject to our chosen authority. The failure of these methods shows that we cannot live with a method which ignores the social dimension of inquiry or one that allows the correctness of an answer to be determined by the will of some individual. The a priori method meets these conditions: we are to accept the answer to our question which is ‘agreeable to reason’, the answer which seems most plausible. But this fails because the correctness of an answer remains a subjective matter: the method would be likely to make truth a matter of fashion.

The method of science introduces objectivity. It requires us to carry out inquiries in accordance with norms which reflect the following fundamental hypothesis:

> There are real things, whose characters are entirely independent of our opinions about them; those realities affect our sense according to regular laws, and, though our sensations are as different as our relations to the objects, yet, by taking advantage of the laws of perception, we can ascertain by reasoning how things really are; and any man, if he have sufficient experience and reason enough about it, will be led to the one true conclusion.
>
> (1992-4: 120)

Peirce’s arguments in favour of this method are rather unclear, but he strongly suggests that it alone is in harmony with ‘the logical question’. It is the only method ‘which presents any distinction of a right and wrong way’.

> The feeling which gives rise to any method of fixing belief is a dissatisfaction at two repugnant propositions. But here already is a vague concession that there is some *one* thing to which a proposition should conform. Nobody, therefore, can really doubt that there are realities…
>
> (1992-4: 120)

But Peirce also insists that everyone uses this method about ‘a great many things’, abandoning it only when unsure how to do so, and points to the triumphs of the scientific tradition as further recommendation. However the method of science has so far been formulated sketchily, and we should now examine Peirce’s attempts to fill in some of the details.

3 Pragmatism

Although Peirce did not use the word in print until more than twenty years later, ‘How to Make Our Ideas Clear’ (1878a) introduces the rule for achieving complete clarity about the contents of concepts, propositions and hypotheses which was later called ‘pragmatism’. The rule helps us to carry out scientific investigations in a responsible, self-controlled manner, but Peirce’s principle also has important applications within logic and philosophy. First, it is used to clarify concepts like reality and probability, which are fundamental to his understanding of the method of science. Second, its account of the meaning of propositions helps to explain how the method of science can indeed show anyone ‘the one true conclusion’ to an investigation. And third, it is needed to demonstrate that there are no important propositions - and none on which science depends - whose truth values cannot be established using the method of science. Peirce must show, for example, that all ‘ontological metaphysics’ is ‘gibberish’ (see Meaning and verification).

Peirce contrasts three grades of clarity in our apprehension of a concept or proposition. I possess the first when I unthinkingly apply the concept in my experience, and the second involves possession of an abstract definition. For example, we can define reality as ‘that which is not whatever we happen to think it’, but is unaffected by what we may think of it. These notions of clarity, familiar from rationalists’ talk of ‘clear and distinct ideas’, can appear sufficient only in the light of a discredited logic and philosophy. The third ‘pragmatist’ grade, Peirce later stressed, accords with experimentalist approaches to inquiry and laboratory experience. His formulation of his method for achieving this third grade of clarity is:

> Consider what effects, which might conceivably have practical bearing, we conceive the object of our conception to have. Then our conception of those effects is the whole of our conception of the object.

(1992-4: 132)

His examples, and his later formulations, clarify this obscure statement. For example:

> To say that a body is heavy simply means that, in the absence of opposing force, it will fall. This (neglecting certain specifications of how it will fall, etc., which exist in the mind of the physicist who uses the word) is evidently the whole conception of weight.

(1992-4: 133)

From the proposition that the body is heavy, I can derive conditional propositions predicting the experiential results of different actions: if I reduce the force acting to support the heavy body, it will fall. Pragmatism holds that I can provide a *complete* clarification of a concept by listing such conditional propositions: they tell me what effects my actions will have if the concept applies to a specified object. As Peirce insisted, this is an ‘experimentalist’s’ view of the content of propositions.

‘How to Make Our Ideas Clear’ uses the principle to clarify the concept of reality. Listing different methods for investigating the velocity of light, Peirce notes that, while users of each may initially produce different results, ‘as each perfects his method and his processes, the results will move steadily together towards a destined centre’. It is typical of scientific inquiry that ‘no modification of the point of view taken, no selection of other facts for study, no natural bent of mind even, can enable a man to escape the predestinate opinion’. Peirce’s clarification of the concept of reality is thus: ‘The opinion which is fated to be ultimately agreed to by all who investigate is what we mean by the truth, and the object represented in this opinion is the real’. Truth is defined by reference to long-run convergence of opinion among responsible inquirers: it is not independent of ‘thought in general’ but it is independent of what any individual may think at any particular time.

Another important application of Peirce’s principle is his account of probability, presented in the third paper of the series, ‘The Doctrine of Chances’ (1878b) (see Probability, interpretations of). Urging that we clarify the idea of probability by examining ‘what real and sensible difference there is between one degree of probability and another’, he follows Locke in identifying probability as a property of inferences. Deductively valid arguments are those which belong ‘to a genus of arguments all constructed in the same way, and such that, when their premises are real facts, their conclusions are so also. If the argument is demonstrative, then this is always so; if it is only probable, then it is for the most part so’ (1992-4: 146). The probability of a mode of argument is thus ‘the proportion of cases in which it carries truth with it’. A probability statement formulates a guiding principle for such an inference. If the probability of a tossed coin landing with heads uppermost is 0.48, then a claim is made about inferences from the premise that this coin was fairly tossed to the conclusion that it lands with heads uppermost: in the long run, such inferences will be successful around 48 per cent of the time. Pragmatism supports a frequency account of probability.

> As we go on drawing inference after inference of the given kind, during the first ten or hundred cases the ratio of successes may be expected to show considerable fluctuations; but when we come to the thousands and millions, these fluctuations become less and less; and if we continue for long enough, the ratio will approximate towards a fixed limit.

(1992-4: 146)

4 Induction and the method of science

So far, the method of science has been defined in terms of its underlying presupposition: for any question that we investigate, there is a correct answer which we are fated to reach if we carry out our investigations long enough and well enough. A more detailed account is needed of the scientific method and of how this convergence in opinion is to be secured. What inferences and methods should we employ, and how do they lead us to the truth?

Like other nineteenth-century thinkers, Peirce emphasized the self-correcting character of inductive reasoning: in the long run, error is sure to be eliminated and whatever survives testing will be the truth. Peirce’s development of this idea exploited an analysis of statistical sampling (‘quantitative induction’) which attempted to establish that successive sampling of a population is guaranteed eventually to discover the proportion of the population having a particular character. This was supplemented by the claim that all inductive inference can be modelled on statistical

sampling.

Suppose we wish to know the proportion of white beans in large bag of beans. Drawing out a handful, we find that 75 per cent of these beans are white. An inductive argument would then tentatively conclude that 75 per cent of the beans in the whole bag were white. So long as the handful is sufficiently large, and the beans in the bag are evenly distributed, the probability is high that the proportion of white beans in the bag as a whole is fairly close to 75 per cent. Moreover, if it is not, further sampling will arrive at a figure that is more accurate than the original one. The mathematical details of Peirce’s approach need not delay us, although we should note that this is in harmony with his pragmatist clarification of the concept of probability: if the probability of a bean drawn from the bag being white is 0.75, then as we draw more and more handfuls, we expect to find that 75 per cent of those that we draw are indeed white.

But quantitative induction forms only a small part of scientific inference. The experimental testing of theories involves an apparently different form of inference which is referred to as ‘hypothesis’ in Peirce’s earlier writings, and later as ‘qualitative induction’. But Peirce’s pragmatism reveals an analogy between quantitative and qualitative induction. For pragmatism holds that there is nothing to the content of a hypothesis apart from a set of conditional predictions, expectations about the experiential consequences we should expect our actions to have were the theory to be true. We are not in a position to test every one of those conditional predictions, although we can see that, if all were confirmed, the hypothesis would be true. When a hypothesis is tested empirically, then, we sample those predictions, and we infer from the success of the sample to the reliability of the ‘population’ of predictions forming the content of the hypothesis. The probability calculus is not applicable in these cases, of course, but as long as the pragmatist principle clarifies the full meaning of the hypothesis, the parallel cannot be denied.

There is an important difference between quantitative and qualitative induction. When subsequent sampling conflicts with my estimate that 75 per cent of the beans are white, it does so by introducing a revised estimate of the proportion: eventually, continued sampling will arrive at the correct figure. When evidence conflicts with a hypothesis in the course of qualitative induction, it does not introduce a replacement hypothesis. We depend upon our ability to think of promising hypotheses and to choose the right ones for empirical test. These matters are addressed in Peirce’s account of ‘abduction’. Much of the logic of abduction investigates methodological rules to be followed in deciding which hypotheses to test: we should prefer those that can be tested economically and whose falsity would be quickly exposed; we should also favour those that seem simple, cohere with our metaphysical views, or appeal to explanatory mechanisms similar to those that have been successful elsewhere in science. Many of these matters concern the economics of research: we plan our investigations in order to maximize progress and minimize expenditure of time, money and effort.

5 Abduction and critical common-sensism

More fundamental than this ‘economics of research’, however, is the ‘primary abduction’ on which all scientific activity depends, the presupposition or hope that we have a capacity for guessing right. We have a ‘natural instinct for the truth’: ‘the human mind is akin to the truth in the sense that in a finite number of guesses it will light upon the correct hypothesis’ (1931-58: 7, 139). Unless we can be confident that we shall think of a good hypothesis fairly soon, the motivation to carry out scientific research would weaken. Of course, this provides no reason for believing that the primary abduction offers, at best, grounds for hoping that it is correct. The history of science helps by showing that ‘it has seldom been necessary to try more than two or three hypotheses made by clear genius before the right one was found’.

Peirce’s account of the growth of scientific knowledge involves one further complexity. When first introduced, a hypothesis offers a vague picture which is incomplete and incorrect in many of its details: at best, we hope that some precise revision of it can be defended. Inductive testing will guide us in making revisions, in rejecting some adjustments and developing others:

> The familiar kinetical theory of gases illustrates this well. It began with a number of spheres almost infinitesimally small occasionally colliding. It was afterward so modified that the forces between the spheres, instead of merely separating them, were mainly attractive, that the molecules were not spheres, but systems, and that the part of space within which their motions are free is appreciably less than the entire volume of the

gas. There was no new hypothetical element in these modifications.

(1931-58: 7, 135-6)

This background to abduction will also contain beliefs and inferences drawn from common sense. Such beliefs and inferences are ‘acritical’: they cannot be subjected to critical evaluation and when asked to defend them we can say only ‘everything counts for them and nothing counts against them’. This slowly evolving body of opinions is indubitable and can almost be seen as instinctive. It is certain largely because common-sense beliefs are invariably vague: any proposed precise version of them would be fallible, but its defeat would not demonstrate that another way of making them precise cannot be found. General ideas drawn from analytical mechanics provide a common-sense framework against which research in physics develops, and the assumptions about beliefs, desires and meanings reflected in decision theory and microeconomics articulate the common-sense background to the social and human sciences. Peirce’s ‘critical common-sensism’ emphasizes the role of such instinctive certainties in grounding our policy of reflective critical inquiry. We should try to doubt such inferences and propositions in order to satisfy ourselves that they are indeed part of common sense, but, of course, we should not deceive ourselves into thinking that they are indeed dubitable on disreputable Cartesian grounds.

Although common-sense beliefs and vague theoretical pictures are hard to criticize, Peirce described himself as a ‘contrite fallibilist’: we are aware that most of our theories require revision and we anticipate that any of them might have to be abandoned (see Fallibilism). We look on them as currently accepted opinions, acknowledging that accepting them now is the best means to eventual progress towards the truth. In Peirce’s later writings, he expresses this fallibilism by denying that we should ever *believe* our current scientific results, and arguing that pure scientific research should be divorced from any concern with the useful applications which its results may have. Ironically, the method of science does not provide a means to the short-term fixation of belief.

Belief, according to Peirce, must result from instinct or sentiment, from the action of common sense, and he was scathing about those who trust theoretical reflection in connection with ‘vital’ matters. When facing important moral decisions or agonizing over the existence of God, we should spurn reflective deliberation, trusting instead our moral and religious sentiments, the instinctive manifestations of common-sense wisdom.

6 Architectonic and self-control

From around 1900, Peirce insisted that his work in logic and epistemology should be embedded in a systematic framework: a proof of pragmatism and a defence of induction were required, and his search for these led to the attempt to ground logic in aesthetics and ethics and to his development of phenomenology as a fundamental branch of philosophy. Why was he dissatisfied with his earlier defence of his views?

First (a minor point), he came to deny that we could establish that something is true by showing that it is a presupposition of inquiry: at best, that warrants our *hoping* that it is true. Hence his views risk refutation through a demonstration that what scientists hope to be true is not in fact so. Second, his pragmatism claimed that all scientific propositions can be fully elucidated using the pragmatic principle: there is nothing to their content apart from conditional predictions. Scientific work is shaped by explanatory ideals which guide us in establishing a *system* of knowledge. If concepts, such as continuity, which have a role in setting these ideals cannot be elucidated using the pragmatist principle, then Peirce’s pragmatism would be on shaky ground. He needed to show that such concepts were consistent with pragmatism - a conclusion his hero Kant would have denied. Moreover, he aspired to an argument for pragmatism which would convince even those who antecedently believed that they understood concepts that it would rule out of court. A particular threat concerned Peirce’s realism. He wished to take natural necessity seriously, and to claim that pragmatism was committed to the objectivity of ‘would-bes’ - to statements about what *would* occur in various possible situations (see §9 below). However, it is commonly assumed that the kind of verificationism that his position represents cannot make sense of the objectivity of ‘would-bes’. Many were convinced by Hume and others that verificationism required nominalism.

Finally, his argument for the superiority of the method of science claimed that the ‘social impulse’ acted against methods such as tenacity and authority. If this is a *psychological* claim about human inquirers, it is hard to see how it establishes the *normative* claim that only the method of science ought to be employed. Moreover, we have noted that since the method of science only promises to settle belief stably in the long run, requiring us to identify our interests with those of the wider community, we are required to use a different method (trusting to instinct) when

we try to settle vital questions. So, it seems, we do use an alternative method much of the time. Or if we use the method of science, confident that we can trust its results in connection with everyday matters, there is a question about the right with which we do so. And why is it rational for us to devote our lives to the cooperative pursuit of long-run stable settlement of beliefs? How can the search for the truth about reality be a good goal for us to adopt? It seems that Peirce must embed his views in a general theory of rationality, in a general theory of how we can order our lives in a self-controlled manner. Otherwise, his defence of the method of science begs too many questions.

By the late 1890s, Peirce was emphasizing that philosophy should have an architectonic character. It should be systematic and guided by a plan of the structure of knowledge as a whole. Thus he relied upon an explicit classification of the sciences, especially the philosophical sciences which can be schematized thus:

(1) Mathematics
(2) Phenomenology (or ‘Phaneroscopy’)
(3) Normative science:
(a) Aesthetics
(b) Ethics
(c) Logic:
(i) Speculative grammar
(ii) Critic
(iii) Methodeutic
(4) Metaphysics

Mathematics, our ‘practice of necessary reasoning’, needs no foundations, but provides techniques used by phenomenology to obtain a theory of categories from reflection upon all that appears to us in any way. Using these categories, the first two normative sciences provide an account of the rationality of ends: they identify what it is possible to admire unconditionally, and what it is possible to adopt as an unconditional end for conduct. Logic then investigates the norms governing inquiries: the first branch provides a systematic account of sign interpretation and reference; the second classifies arguments and explains their validity; and the third states and defends the methodological principles that guide inquiries. Peirce’s theory of truth is defended within his logic. Metaphysics develops a general scientific account of ‘the most general features of reality and real objects’. Together, these philosophical sciences explain the possibility of rational, self-controlled inquiry.

7 Categories, phenomenology and normative science

Throughout his career, Peirce’s philosophical ideas depended upon a system of categories, a set of universal conceptions which could be used to classify anything that could be experienced, thought about or imagined (see Categories). Growing out of his criticisms of Kant’s theory of categories during the 1860s, the system was subject to considerable development and refinement until the final decade of his life and had a fundamental role in grounding his other philosophical views. Although his terminology varied, he most commonly referred to these categories as ‘Firstness’, ‘Secondness’ and ‘Thirdness’. It is easiest to understand them by following Peirce himself in seeing how they are reflected in the logical forms of a ‘perfectly exact, systematic and analytic language in which all reasoning could be expressed and be reduced to formal rules’. Such a language might contain expressions like ‘… is red’ which express properties of a single subject as well as relational expressions such as the two place ‘…is taller than…’ and the three place ‘…gives…to…’. Peirce insisted that an adequate language would contain expressions of all these three kinds, and also that it need contain no expressions for relations between four or more objects. Concepts or phenomena are classified as forms of Firstness, Secondness and Thirdness according to whether they are expressed by general expressions which take one, two or three subjects. Firstness is what it is independently of anything else, Secondness involves what is in relation to something else, and Thirdness is manifested when something mediates between two others.

Peirce’s important 1867 paper ‘On a New List of Categories’ contained the first published version of the theory. It appeared before Peirce had worked much on the logic of relations and made no use of the terminology of Firstness, Secondness and Thirdness. However it was self-consciously Kantian in defending a system of categories by discussing what was required to reduce the manifold of sense to the unity of a proposition: Peirce spoke of

reducing the manifold of Substance to the unity of Being. We do so, he urged, by ascribing a quality to the substance, by classifying it or describing it. But how can we do that? Since 'we know a quality only by means of its contrast or similarity with another', we can ascribe qualities to things only because we can make relational judgments of similarity and difference. And we can make judgments of similarity only because we can carry out 'comparisons': we can interpret one object as a representation of another. So making sense of things by ascribing qualities to them depends upon our ability to make relational judgments and to work with the three-place relation of representation. All three categories are required for even the simplest kind of judgment. Peirce soon came to reject the assumption that quality (monadic characteristics) had a special role in unifying the manifold: the logic of relations led to the overthrow of Aristotelian logic and taught that relational judgments were just as fundamental as these simple monadic ones. Although this led to a reformulation of his theory of categories, he continued to use the argument of this early paper. He claimed that it embodied his 'one contribution to philosophy'.

With the development of the logic of relations, Peirce exploited his 'remarkable theorem' that it was impossible to define triadic relations in terms of simpler ones, whereas it was always possible to provide such definitions for relations with four or more relata. Within standard systems of the logic of relations, this does not hold: it is possible to reduce triadic relations to dyadic ones. However, recent scholarship suggests that Peirce's claim holds for his own systems of logic and that, for the purposes of constructing a system of categories, the Peircean systems are more perspicuous than those encountered in standard logic texts. Since this argument uses mathematics to defend the categories and mathematics was the foundational discipline without Peirce's architectonic structure, Peirce's continued reliance upon this argument is not surprising.

However, from the late 1880s new quasi-empirical arguments were introduced, culminating after 1900 in a phenomenological defence of them. 'A Guess at the Riddle' (1887-8) traced the 'triad' through logic and semiotics, in metaphysics, psychology, physiology and biology, and in physics and theology, thereby producing a 'long list' of categories. This investigation culminated in the development of Peirce's evolutionary metaphysics (see §10 below). Although not intended as a proof of the theory of categories, it was relevant to establishing a doctrine fundamental to Peirce's realism: Thirdness is found in the physical and biological realms as well as among mental and linguistic phenomena.

From 1903, Peirce undertook to defend his categories by looking 'directly upon the universal phenomenon, that is upon all that in any way appears, whether as fact or as fiction'. With the aid of special analytical techniques, his search again discovered the three fundamental categories, Firstness, Secondness and Thirdness. Since he remained convinced by the early 'logical' arguments, there is a question of why this further defence was required. The answer lies in Peirce's architectonic, his account of how the philosophical sciences all hang together. If the categories are to be used in logic and the other normative sciences, they must be defended in a more fundamental discipline. Since formal logic is a branch of mathematics, this may not seem problematic, but the logical argument for the categories required Peirce to show that a particular mathematical formalism could indeed be used as a system of formal logic: formal logic is, after all, a branch of *applied* mathematics. By showing that a formalism which incorporates the categories is adequate to describe 'all that in any way appears, whether as fact or as fiction', Peirce can be confident in relying upon his categories when he turns to the study of our attempts to discover the facts in logic.

Peirce's work in normative science seeks a theoretical account of what we can admire unconditionally (aesthetics), adopt as an ultimate end (ethics), or adopt as a fundamental goal for inquiry (logic). Thought experiments are employed: we consider different possible objects of experiments or imagine living with a particular aim in various counterfactual circumstances. In each case, the good consists in a manifold of phenomena each with their own qualitative character of Firstness, but standing in dyadic relations, reacting against each other. But this Secondness is mediated: the whole exhibits a coherence or unity which we can apprehend. Beauty, goodness and truth involve a kind of organic unity which excites our admiration, motivates our conduct and satisfies our inquiries. The method of science ensures progress towards a coherent body of opinions which brings our actions and experiences into harmony.

8 Signs and interpretations

Peirce's semiotic provides a wholly general theory of meaning and representation. An important move in his critique of the Cartesian conception of mind during the 1860s was that all thoughts and experiences are signs. In

later years, he used these semiotic conceptions to develop a sophisticated account of language - particularly of the language used by a 'scientific intelligence'. His philosophical account of science and mathematics focused upon the role of science interpretation in both, and important problems in the ontology of mathematics were answered by reference to the special character of the signs or representations used within that discipline. And the grounds of logic or deductive validity were traced to features of meaning and the sign relation. It is not surprising that Peirce described logic as 'only another name for *semiotic*, the quasi-necessary or formal doctrine of signs'. There are few of Peirce's writings that are not concerned with questions of meaning and signification.

His work exploits the fundamental insight that signification is a form of Thirdness. The sign relation is irreducibly triadic, and the third element in this relation is interpretation. A name denotes an object or a sentence stands for a state of affairs by virtue of being interpreted in subsequent thought, speech or action as doing so. The interpretant, which is itself a sign with the same object, thus mediates between the original sign and its object. Thus my understanding of a report that an animal before me is a cat can be variously manifested in my explicitly thinking or saying that this is what the utterance means, in my inferring that the animal probably likes milk, in my offering it cat food, or in my showing surprise when in suddenly barks. Unless the sign has the capacity to produce such interpretant thoughts (the meaning of which in turn depends upon how they are interpreted) it would have no significance. Since a thought too is a sign, the content of a thought (or any other mental event) is determined by how it is interpreted and developed in subsequent thought. In general, the content of a thought or utterance depends upon its effects.

Much of Peirce's writings about signs was directed at constructing a classification of types of signs: he sought an exhaustive classification of signs, objects and interpretants. One motivation for this was Peirce's desire to defend the pragmatist principle by showing that it would enable us to be reflectively aware of all those features of how a sign should be interpreted that were relevant to our scientific purposes. He argued that the 'ultimate logical interpretant' of a sign was a habit of expectation which could be fully described by application of the pragmatist principle. Some of the classifications are of more general interest, offering important insights into thought, language and other forms of representation. And his writings on signs contain valuable contributions to the understanding of mental phenomena such as emotions, sensory experiences and religious experience as well as interesting discussions of proper names, vagueness, conditionals, modality, quantification, force and content and other live issues in the philosophy of language.

The question of how we are guided in arriving at interpretations of signs introduces the best known of Peirce's classifications, that between icon, index and symbol. Roughly, interpretation of an iconic sign is grounded in a resemblance between the sign and its object, interpretation of an index exploits a 'real existential relation' between the two, and we can defend our interpretation of a symbol by reference to an established practice of so interpreting it. Photographs and maps are icons, signposts and pointing fingers are indices, and it is easy to see that natural language has a strong symbolic component. The utterance of a symbolic sign is always a *replica* of the symbol (a token of the type), and its interpretation appeals to the interpretations offered of other replicas or tokens. Peirce denied that there were 'pure' icons or indices. Maps and signposts are 'hypo-icons' and 'hypo-indices', having a strong conventional component. The rules governing the use of maps do not fix the interpretation of particular features of the map unaided: they rather guide us in how to use it as an icon. Similarly, although we are able to understand pointing fingers only because we are masters of the practice of doing so, the rules of this practice do not explicitly lay down what a particular finger is pointing at. Rather, they guide us in how to interpret the finger as a conventional index.

This classification exploits Peirce's system of categories. Interpretation of an index depends upon a dyadic existential relation between sign and object: it exploits properties the sign would lack if its object did not exist. When we understand a symbol, we appeal to triadic features of the sign, such as how its replicas have been interpreted in the past. And an understanding of icons exploits common features of sign and object which each could have, even if the other had not existed. In his 1885 paper, 'On the Algebra of Logic: a Contribution to the Philosophy of Notation', Peirce announced that an adequate language for descriptive or scientific purposes must contain signs of all three kinds. Indeed, any proposition must have symbolic, iconic and indexical components. Indices such as names, demonstrative expressions, pronouns and quantifier expressions are required if the proposition is to speak of any external things at all: the universe of discourse must be specified indexically. Symbols are the only general signs, and generality is essential to reasoning: we must be able to recognize that

premises and conclusions contain tokens of the same word or phrase type, replicas of the same symbol.

The value of iconic signs is that we can exploit their similarity with their objects, learning more about the latter by noting features of the icon. We can make discoveries about the terrain by making measurements on a map, or learn more about a building by examining a photograph. A proposition provides a sort of diagram of a state of affairs, mirroring its logical structure. Inference exploits this abstract resemblance: we learn more about the world by making substitutions in propositions in the light of other information that is to hand and observing the result of our ‘experiment’. This is clearest when we think of formal logic, which, like all of mathematics, constructs iconic representations of the structure of propositions and arguments which enable us to investigate the validity of arguments. Unless descriptive propositions were icons and shared abstract structural properties with their objects, Peirce argued, we could make no sense of how reasoning can provide us with new knowledge. The formal character of reasoning shows that we interpret propositions as icons. So any language is a device for constructing conventional, logical diagrams which are tied down to concrete, existing things through the use of indices.

After 1900, Peirce sought a proof for his pragmatism. He needed to show that no aspect of the meaning of a proposition, nothing relevant to its cognitive functioning, was omitted when we clarify it using the pragmatist principle. One strategy was to develop his theory of signs and show that interpreting a proposition in the light of pragmatism produced its most explicit (‘ultimate’) logical interpretant. Another strategy was to construct a complete classification of argument forms and to show that we needed no information missed by a pragmatist elucidation in order to employ a proposition in reflective, controlled argument. An important tool here was the system of existential graphs (see Logic machines and diagrams). A system of formal logic analogous to later natural deduction systems, the graphs promised ‘moving pictures of thought’ which would exhibit the structure of all possible forms of argument. As well as a complete system of propositional and predicate calculus, the graphs promised an account of sophisticated reasonings involving modality, continuity and abstraction. If all reasoning is reflected in the system of existential graphs, and if the pragmatic principle is true for all concepts whose conceptual role is modelled within them, then the pragmatist is vindicated. Unfortunately there is little evidence that Peirce completed the graphs to his satisfaction.

9 Realism

Peirce described himself as a realist, rejecting nominalism and epistemic idealism as the source of most philosophical ills (see Nominalism; Realism and antirealism). This can be surprising because pragmatism is often understood as an antirealist, verificationist doctrine.

It is useful to distinguish three different themes in his realism. The first is a response to the problem of universals. Peirce argues that it is a mistake to understand realism about ‘generals’ or universals as a claim about the existence of a particular realm of curious abstract objects. Realists are committed to the objectivity of propositions concerning whether this or that object is a horse; they are not committed to the existence of horseness as a special particular. This view was most clearly expressed in an 1870 review of a new edition of Berkeley’s works where Peirce distinguishes two conceptions of reality, nominalist and realist. They differ in how they develop the uncontroversial claim that the real is ‘that which is not whatever we happen to think it, but is unaffected by what we may think of it’. Nominalists note that our thoughts are caused by sensations which are, in turn, caused by something outside the mind, and identify the real with these efficient causes of our sensations and thoughts. This picture of reality, which is closely allied to the representative theory of perception and the correspondence theory of truth, makes problematic the question of whether anything in reality corresponds to general conceptions such as ‘horse’ or ‘man’ and encourages the sort of mistrust of our conceptual interpretation of sensory input that makes Cartesian strategies attractive. Peirce favours the realist conception which notes that ‘to every question there is a true answer, a final conclusion, to which the opinion of every man is constantly gravitating’. Reality is the ‘final cause’ of inquiry, the answer to a question that we are ‘fated’ to reach if we inquire long enough and well enough.

This view of reality is reminiscent of Kant’s ‘empirical realism’, and encourages realism about generals because it is objective whether (for example) Pegasus was a horse. The account of reality obtained from the pragmatist principle supports this form of ‘realism’. However, since truth is explained in terms of the convergence of opinion, there is no conceptual gap between truth and what efficient, responsible inquirers would agree upon. The position thus appears to be in tension with the realist view that truth is not determined by what we take it to be. Moreover, a realist will presumably hold that there are many truths which inquiry will never uncover, all trace of them having

been lost. In the 1870s, Peirce suggested that this was a verbal matter of whether a diamond which was created and destroyed without being exposed to any sort of test was hard, but he later insisted that this was a mistake which his realism enabled him to avoid.

A second theme in Peirce's realism responded to this difficulty. Peirce's realism became more extreme in the 1880s when he defended objective modalities. Natural laws manifest real, natural necessity and there are facts concerning 'would-bes' - subjunctive conditionals concern what *would* occur in various counterfactual (but real) possibilities (see Laws, natural; Counterfactuals). The conditionals employed when we clarify a concept using the pragmatist principle are expressed in the subjunctive mood: they concern what would happen (or what would have happened) had certain actions been carried out. So in later work, Peirce's frequency theory of probability is transformed into a propensity view: the probability of an argument is the proportion of cases in which it *would* transmit truth from premises to conclusion.

Returning in 1906 to the example of the diamond which is destroyed before being tested for hardness, Peirce noted that science teaches that diamonds have in common a molecular structure which accounts for their hardness. We know that if it had been tested, it would have proved hard. The hardness of individual diamonds is not ontologically prior to the fact that diamonds (in general) are hard. This form of realism is connected to the reality of Thirdness. Since we directly experience Thirdness, we are directly aware of necessity or mediation. Peirce's synechism turns the question of realism into the question whether there is real continuity, real mediation.

The third theme reflects Peirce's answer to problems about reference which he faced during the 1870s: we directly perceive external things, referring to them demonstratively in perceptual judgments (see Reference). This allows for the possibility of reference to external objects (including theoretical entities) while being aware that some questions about their character have answers which will never be discovered by us. The irreducibility of Secondness is relevant here: we are aware of external things as reacting against us, as 'other'; and this is reflected in the fact that reference to external things is fundamentally indexical, involving demonstrative expressions in perceptual judgments. Theoretical entities are known to exist because they too possess Secondness - they react with other existing things.

10 Scientific metaphysics and evolutionary cosmology

In *The Monist* (1891-3), Peirce published a series of five papers which fit poorly with the anti-metaphysical flavour of his pragmatist writings. Developing themes that he had been exploring since 1883-4, they contained an evolutionary cosmology, which, he suggested, involved a form of objective idealism, and they argued, against Darwin, that this process of evolution should be seen as, in a sense, purposive: 'agapastic' evolution was driven by evolutionary love, and the universe is a vast mind perfecting itself through time (see Idealism). Some commentators, like Thomas Goudge (1950), attributed this work to a philosophical split personality: the naturalistic, tough-minded, anti-metaphysical pragmatist betrayed a taste for transcendental speculation which could not be reconciled with the rest of his work. Peirce himself was clearly sensitive to the need to show that the different elements of his philosophical vision formed a unity. His cosmology attempted to understand the most general features of our view of the world using the method of science. Metaphysics was empirical, differing from the special sciences only in its generality and in its avoidance of sophisticated experiment: its data was drawn from everyday facts whose very familiarity is likely to hide them from our notice.

He also insisted that his metaphysics was required by his work on logic and the method of science: self-controlled reasoning, employing the scientific method, was possible only if a Peircean metaphysics is possible too. In 'The Fixation of Belief', Peirce sought presuppositions of inquiry. He soon saw that showing that something was a presupposition of inquiry (or of a particular inquiry) did not establish its truth: at best, we are warranted in *hoping* that it is true. We must hope that there is a reality, independent of us, of which we can obtain knowledge, that the questions we investigate have answers, that we possess the freedom needed for logical self-control, that we can tell the difference between a good argument and a bad one, that we have the instincts posited in the 'primary abduction' and so on. We must look to metaphysics for the account of mind and reality that explains the truth of these hopes or regulative ideas. If no such account is possible, rational self-control is an illusion.

According to Peirce, the fundamental rule of logic is: 'Don't block the road of inquiry.' We should never assume that any question lacks an answer or that any regularity lacks an explanation: we should always hope that an

explanation is to be found. This led him, in 1883 and later, to set himself the task of explaining the fact that there are general laws. Nominalists take there to be brute regularities which have no explanation, and physical atomism appears to require this nominalistic picture. If all law and regularity can be explained, nominalism must be rejected and Peirce is committed to the rejection of metaphysical or physical atomism. Peirce's answer is to construct an evolutionary metaphysics which traces the evolution of law out of pure possibility (of Thirdness out of Firstness). And this evolutionary story enjoins that the universe contains pure chance (Peirce's 'tychism'): a generalizing tendency reinforces chance regularities and the universe becomes steadily more regular and ordered. Just as we acquire more and more habits as our hypotheses are confirmed and habits of expectation are reinforced, so the universe itself becomes more and more 'hidebound with habits' (see Determinism and indeterminism).

The analogy just offered is important for understanding Peirce's evolutionary story: our ideas of inference, purpose and the growth of knowledge - ideas linked to the practice of sign interpretation - provide the model we must use in making sense of the physical world. Problems of interaction show that we require a monistic system of metaphysics, and physicalism can make no sense of consciousness and other mental phenomena. Hence Peirce's view that the entire universe is a vast mind: laws are analogous to habits, and the ways in which laws determine events have to be understood by appeal to teleology and final causation. Physical events differ from ordinary mental events in being less flexible, being more 'hidebound with habits'. We arrive at a form of idealism which, although it does not make reality depend upon the contents of human minds or on its being known, still holds that our fundamental explanatory categories are those most familiar from ordinary psychological explanation. The evolution of law is not the mechanical and meaningless process described by Darwin. Peirce's evolution is 'agapastic': evolutionary love is manifested in the ways in which the universe becomes steadily more perfect. Against this background, it is no surprise that in 1908 Peirce published a 'Neglected Argument for the Reality of God'. He believed that the Universe was a vast mind, and that even those who explicitly defend atheism are likely to have a natural belief in this pantheistic god. It is manifested in a confidence that the Universe is steadily becoming more perfect, and in our sense that through contributing to science we contribute to 'the process of creation', to the growing rationality of the Universe.

See also: Doubt; Empiricism; Hegelianism §5; Induction, epistemic issues in; Logic in the 19th century; Pragmatism §1; Scepticism; Science, 19th century philosophy of; Scientific realism and antirealism; Semiotics; Truth, pragmatic theory of §1

CHRISTOPHER HOOKWAY

List of works

Although Peirce published extensively, he did not publish any major philosophical treatises summarizing his position. The following list samples some of the more important pieces, but inevitably it omits much that is significant: any list of 'major works', particularly after 1890, will be controversial and arbitrary. This material is all readily available in the first two collections, listed below.

Peirce, C.S. (1931-58) *Collected Papers of Charles Sanders Peirce*, ed. C. Hartshorne and P. Weiss (vols 1-6) and A. Burks (vols 7-8), Cambridge, MA: Harvard University Press, 8 vols.(A selection of papers and excerpts from manuscripts which was responsible for introducing Peirce's work to a wide audience. Since the material is organized thematically rather than chronologically and the texts are not all reliable, this edition is being superseded by *The Writings of Charles S. Peirce*.)

Peirce, C.S. (1982-) *The Writings of Charles S. Peirce: A Chronological Edition*, ed. M. Fisch, C. Kloesel, E. Moore, N. Houser *et al.*, Bloomington, IN: Indiana University Press.(A reliable and extensive selection of Peirce's published and unpublished writings, organized chronologically, and expected to run to thirty volumes. An indispensable tool for research, it is probably less useful for the new reader. The introductions to the different volumes provide an invaluable intellectual biography of Peirce.)

Peirce, C.S. (1992-4) *The Essential Peirce*, ed. N. Houser, and C. Kloesel, Bloomington, IN: Indiana University Press.(An excellent two-volume selection which contains reliable texts of all of Peirce's most important published and unpublished works.)

Peirce, C.S. (1867) 'On a New List of Categories', *Proceedings of the American Academy of Arts and Sciences* 7: 287-98.(A classic but difficult paper, containing the first published statement of Peirce's theory of categories. Reprinted as chap. 1, vol. 1 of Houser and Kloesel (eds) *The Essential Peirce*.)The first three of the following papers contain an attack on Cartesian approaches to philosophy, introduce Peirce's claim that all thought is in signs, and develop an account of truth and cognition on this basis. They are reprinted as chapters 2 to 4 of

Houser and Kloesel (eds) *The Essential Peirce*.
Peirce, C.S. (1868a) 'Questions Concerning Certain Faculties Claimed for Man', *Journal of Speculative Philosophy* 2: 103-14;(Challenges the Cartesian assumption that we can trust introspection as a source of information about the mind and denies that there are any 'intuitions', first premises for reasoning which are not shaped by earlier thoughts.)
Peirce, C.S. (1868b) 'Some Consequences of Four Incapacities', *Journal of Speculative Philosophy* 2: 140-57; (The first published statement of Peirce's semiotic - his claim that all thought is in signs - and his ideas about reality.)
Peirce, C.S. (1868c) 'On a New List of Categories', in Houser and Kloesel (eds) *The Essential Peirce*, Bloomington, IN: Indiana University Press, 1-10.(The first statement of Peirce's post-Kantian categorial scheme.)
Peirce, C.S. (1869) 'Grounds of Validity of the Laws of Logic: Further Consequences of Four Incapacities', *Journal of Speculative Philosophy* 2: 193-208.(Draws out the implications of the two previous papers for the explanation of the validity of inductive and deductive reasoning.)
Peirce, C.S. (1870) 'Fraser's *The Works of George Berkeley*', *North American Review* 113: 449-72.(A lengthy review which contains important discussions of realism versus nominalism (Houser and Kloesel 1992-4) and Peirce's conception of truth. Chapter 5 of this series of six papers is based on Peirce's attempt to write a logic text during the early 1870s. Although they are readable and extremely influential, the role of Peirce's theories of signs and categories is not made explicit. Available in chapters 7 to 12 of Houser and Kloesel.)
Peirce, C.S. (1877) 'The Fixation of Belief', *Popular Science Monthly* 12: 1-15.(Classic statement of Peirce's theory of inquiry, defending the method of science as a method for replacing doubt by settled belief.)
Peirce, C.S. (1878a) 'How to Make Our Ideas Clear', *Popular Science Monthly* 12: 286-302. (Introduces the pragmatist principle for clarifying concepts and ideas and uses it to explain the meaning of Reality.)
Peirce, C.S. (1878b) 'The Doctrine of Chances', *Popular Science Monthly* 12: 604-15(Uses the pragmatist principle to clarify our understanding of statements of probability.)
Peirce, C.S. (1878c) 'The Probability of Induction', *Popular Science Monthly* 12: 705-18. (Develops his view of probability and links it to an account of the success of inductive reasoning.)
Peirce, C.S. (1878d) 'The Order of Nature', *Popular Science Monthly* 13: 203-17.(Attacks Mill's defence of inductive reasoning and argues that the success of science requires that human beings possess 'innate ideas': due to natural selection, they are predisposed to arrive at the correct theories of the world.)
Peirce, C.S. (1878e) 'Deduction, Induction and Hypothesis', *Popular Science Monthly* 13: 470-82.(A formal and historical classification of the kinds of reasoning used in science.)
Peirce, C.S. (1887-8) 'A Guess at the Riddle', manuscript, 1887-8.(This manuscript elaborates Peirce's system of categories through showing how they are manifested in different fields of knowledge. Chapter 19 of Houser and Kloesel (1992-4).)Although occasionally hard and obscure, the first of the following five papers present the metaphysical system which Peirce had been developing throughout the 1880s. Reprinted as chapters 21 to 25 of Houser and Klousel (1992-4).
Peirce, C.S. (1891) 'The Architecture of Theories', *The Monist* 1: 161-76.(Discusses the proper approach to developing a system of metaphysics and describes the fundamental concepts to be used. His evolutionary cosmology is sketched and the importance of chance and continuity for an adequate philosophical system is emphasized.)
Peirce, C.S. (1892a) 'The Doctrine of Necessity Examined', *The Monist* 2: 321-37. (Criticizes the most common arguments for determinism and defends his 'tychism', the doctrine that the universe must contain absolute chance.)
Peirce, C.S. (1892b) 'The Law of Mind', *The Monist* 2: 533-59.(Introduces Peirce's 'synechism', the use of the idea of continuity in philosophy. The paper discusses the mathematical analysis of continuity as well as emphasizing its importance for understanding the mind.)
Peirce, C.S. (1892c) 'Man's Glassy Essence', *The Monist* 3: 1-22.(Applies Peirce's theory of categories and his synechism to the understanding of matter and the relations between mind and body: contains the suggestion that 'matter is effete mind'.)
Peirce, C.S. (1892d) 'Evolutionary Love', *The Monist* 3: 176-200.(Discusses the mechanisms of evolution, finally defending 'agapism', the doctrine that love is operative in the evolution of the universe.)
Peirce, C.S. (1898) *Reasoning and the Logic of Things*, ed. K.L. Ketner, Cambridge, MA: Harvard University Press.(An important series of lectures which provides a useful introduction to Peirce's work in logic and philosophy. As well as an important discussion of theory and practice, they contain clear introductions to some of his logical theories and a useful discussion of causation and continuity. Hilary Putnam's valuable

introduction contains useful information about Peirce's work on the mathematics of continuity.)
Peirce, C.S. (1903) *Pragmatism*, in *Collected Papers of Charles Sanders Peirce*, ed. C. Hartshorne and P. Weiss (vols 1-6), Cambridge, MA: Harvard University Press, 1931-58, vol. 5, 13-131.(Lectures to Harvard University philosophy department, which attempt to state and defend Peirce's pragmatism. They elaborate his views on phenomenology and the categories; logic, ethics and aesthetics as normative sciences; abduction and perceptual judgments; and realism. An important, though sometimes obscure source.)The first three of the following papers were part of an uncompleted series intended to 'prove' pragmatism. Manuscript drafts of later papers in the series survive.
Peirce, C.S. (1905a) 'What Pragmatism Is', *The Monist* 15: 161-81.(Clarifies Peirce's pragmatism, relating it to laboratory practice and christening it 'pragmaticism' to distinguish it fromWilliam James' position.)
Peirce, C.S. (1905b) 'Issues of Pragmaticism', *The Monist* 15: 481-99.(Contains important discussions of Peirce's mature realism and his critical common-sensism.)
Peirce, C.S. (1906) 'Prolegomena to an Apology for Pragmaticism', *The Monist* 16: 492-546.(An introduction to Peirce's 'existential graphs', a system of formal logic which is to have a fundamental role in the unfinished proof of pragmatism.)
Peirce, C.S. (1908) 'A Neglected Argument for the Reality of God', *Hibbert Journal* 7: 90-112.(An interesting argument for God's reality which is intended to fit Peirce's account of the method of science.)
Eisele, C. (ed.) (1976) *The New Elements of Mathematics*, The Hague: Mouton.(A five-volume anthology of Peirce's contributions to mathematics and mathematical logic.)
Hardwick, C. (ed.) (1977) *Semiotic and Significs*, Bloomington, IN: Indiana University Press.(This contains Peirce's correspondence with Lady Welby and is an important source for his work on signs.)

References and further reading

The secondary literature on Peirce is now enormous and the list below omits much that is interesting and important. The most important source of journal articles is the quarterly *Transactions of the Charles S. Peirce Society*.

Brent, J. (1993) *Charles Sanders Peirce: A Life*, Bloomington, IN: Indiana University Press.(A fascinating study of Peirce's career and of his troubled personal life.)
Burch, R. (1991) *A Peircean Reduction Thesis*, Lubbock, TX: Texas Tech University Press.(An important study which develops Peirce's formal logic and his use of it to defend his categories.)
Colapietro, V. (1989) *Peirce's Approach to the Self*, Buffalo, NY: State University of New York Press.(Discusses Peirce's philosophy of mind and subjectivity from a perspective that takes his semiotic very seriously.)
Fisch, M. (1986) *Peirce, Semiotic and Pragmatism*, Bloomington, IN: Indiana University Press.(A collection of papers dealing with biographical and philosophical matters by a distinguished scholar whose influence on the development of Peirce studies has been unparalleled.)
Goudge, T. (1950) *The Thought of C.S. Peirce*, Toronto: University of Toronto Press.(Influential early monograph which argued that the tough-minded, naturalistic, empiricist aspect of Peirce's work could not be reconciled with the speculative, 'transcendental' side of his character.)
Hausman, C. (1993) *Charles S. Peirce's Evolutionary Philosophy*, Cambridge: Cambridge University Press.(This book emphasizes the systematic structure of Peirce's work, studying his pragmatism, his theory of categories and his account of continuity and places special emphasis upon his metaphysical writings.)
Hookway, C. (1985) *Peirce*, London: Routledge & Kegan Paul.(A general account of Peirce's work which develops the interpretations offered in this entry.)
Misak, C. (1991) *Truth and the End of Inquiry*, Oxford: Oxford University Press.(A book which concentrates on Peirce's theory of inquiry and defends a Peircean conception of truth.)
Murphey, M. (1961) *The Development of Peirce's Philosophy*, Cambridge, MA: Harvard University Press; reprinted 1993, Indianapolis, IN: Hackett Publishing Company.(Influential and important book which established the importance of studying the development of Peirce's work, claiming that he defended four distinct systems.)
Ogden, C.K. and Richards, I.A. (1923) *The Meaning of Meaning* , London: Kegan Paul.(Influential discussion of meaning and literature containing an appendix on Peirce's work in semiotic.)
Rescher, N. (1978) *Peirce's Philosophy of Science*, Notre Dame, IN: University of Notre Dame Press.(Guide to Peirce's philosophy of science placing interesting emphasis upon his ideas about the economics of research.)
Robin, R. (ed.) (1967) *Annotated Catalogue of the Papers of Charles S. Peirce*, Amherst, MA: University of Massachusetts Press.(An essential guide to Peirce's many important manuscripts in the Houghton Library at Harvard.)

Skagestad, P. (1981) *The Road of Inquiry*, New York: Columbia University Press. (Argues for the relevance of Peirce's philosophy of science for current debates over realism.)

Welby, V. (1903) *What is Meaning?*, London: Macmillan.(Contains some of Welby's work in 'Significs' and reviewed by Peirce in *The Nation*.)

Pelagianism

Pelagius, a Christian layman, was active around AD *400. The thesis chiefly associated with his name is that (i) human beings have it in their own power to avoid sin and achieve righteousness. Critics objected that this derogates from human dependence on the grace of God. Pelagius did not deny that the power to avoid sin is itself a gift of God, an* enabling *grace; but he was understood to deny the need for* cooperative *grace, divine aid in using the power rightly, or at least to assert that (ii) such aid is a reward for human effort, and so not an act of grace. Later thinkers who held that God's aid, though not a reward, goes only to those who do make an effort, were accused of believing that (iii) there is no need of* prevenient *grace in causing the effort in the first place. So Pelagianism is a tendency to magnify human powers: its defenders saw it as a (frightening) challenge to humans, its detractors as an insult to God. It was hard without Pelagianism to find a place for free will, or with it for original sin.*

1 Pelagius

Pelagius lived from around 360 to after 418, came from Britain, moved to Rome as a young man, and was later active in Palestine. Rome in the Christianized empire of the 380s had become an administrative backwater (the Western capital having moved to Milan) in which families with glorious names cultivated the memory of their pagan past. Even so, Christian adherence was beginning to be prudent and fashionable; and Pelagius, finding that the new religion often sat lightly on these people, formed - or more likely joined - a puritan group which propagated Christianity as a complete change of life. He insisted that each time you confront a choice you have the possibility of making the right one; and (a stronger claim) that you have the possibility of making the right choice always - the possibility of perfection, sinlessness. The bonds of habit do not excuse. God has fashioned commandments that are suitable to our condition. No one should complain 'like insolent and incompetent servants, "It's too much, it's too difficult, we're only human, there's a limit to what a person can cope with"' (*Ad Demetriadem* 16).

This power to act aright comes by God's grace, that is, as a favour, and in particular through Christ, who 'delivered us from the wrath of judgment by remitting sins, and by his teaching and example' (*Expositiones*, on 1 Thessalonians 1: 10). Human perfection is not an immunity to sin, let alone the antinomian licence to do anything whatever blamelessly; on the contrary, one must keep 'running' (*Ad Demetriadem* 27). Under pressure from critics, Pelagius was willing to 'denounce anyone who believes or says that the grace of God by which "Christ came into this world to save sinners" [1 Timothy 1: 15] is not necessary in each hour, in each moment, in each act that we do' (reported in Augustine, *De gratia Christi* 2.2). But right action is always possible; as he said in a phrase gently mocked by Augustine, 'We are unable to be unable to be unsinners' (*De natura*, quoted in Augustine, *De natura et gratia* 49.57).

2 The reaction

Neither in this nor his other opinions was Pelagius conscious of innovation, but he came under attack from two redoubtable controversialists, Jerome (*c*.342-420) and Augustine (§§6, 13). Jerome, unfairly equating sinlessness with Stoic 'apathy' - emancipation from appetites such as hunger - accused (perhaps counter-accused) him of being tainted with the views of the earlier Church Father Origen, who had lately come under suspicion of heresy. By this time Jerome and Pelagius were neighbours in Palestine, where Pelagius had fled before the impending sack of Rome by the Visigoths in 410. Passing through Carthage in his flight, Pelagius had left there a colleague, Coelestius, whose views on baptism came to Augustine's ears along the African coast at Hippo. Pelagius in Palestine was accused of heresy in 415, but acquitted. The Africans appealed, and in 418 an imperial rescript banned the expression of support for Pelagians and exiled their leaders from Rome, whose bishop Zosimus, after wavering, issued a (lost) ecumenical letter of condemnation. From that time Pelagianism has been a heresy.

The nature of a heresy is defined by the decrees of the Church, not the opinions of its supposed adherents. The Pelagian heresy consists mainly of two parts, both of which Augustine attempted (in his voluminous extant writings on the affair) to fasten on the members of Pelagius' circle and to refute: restriction of the scope of grace, and denial of original sin. On grace, a council held at Carthage in 418 condemned three graded errors: that God's grace is of no avail in aiding people not to sin; that it gives such aid but only by instructing and is not the means of

having the motive and the strength not to sin; and that it aids obedience to God's law but only by making obedience easier (Denzinger and Schoenmetzer 1967: 225-7). On original sin the same council declared it heretical to deny either that Adam's sin is transmitted to his progeny or that infants who die unbaptized will suffer in hell (Denzinger and Schoenmetzer: 223-4). Pelagians had argued that since only avoidable deeds are sinful, infants must be sinless and their baptism cannot be for the remission of sins.

3 Massilians

Under these blows the Pelagian movement lost impetus. But in Augustine's last years a moderated form of it arose which even after his death continued to trouble his follower Prosper of Aquitaine (*c*.390-*c*.463). People were saying, Prosper anxiously reported to Augustine, that 'all effort is abolished, all virtues destroyed, if God's arrangements precede [in the old sense, 'prevent'] human wills, and under the name of predestination a kind of fatal necessity is introduced' (*Epistula* 225.3). Pelagius had held that predestination 'is the same as foreknowledge' (*Expositiones*, on Romans 8: 29). But if it is causative, 'preventing' the will, how can the will be free? A number of monastic communities felt the same way, especially in southern Gaul. According to John Cassian of Marseilles (Massilia): 'We acknowledge that human efforts cannot secure it [perfection] by themselves without the aid of God; on the other hand we insist that only people who work and sweat receive God's mercy and grace' (*De institutis* 12.14) Even if one cannot succeed without help, one must be able to try on one's own: anything else would impose fatalism. It was left to later centuries to worry over Augustine's unsatisfactory response to this problem; meanwhile the defiance of these 'Massilians' was eventually condemned at a council at Orange in 529.

4 Ockhamists

In Western Christian controversy, accusations of Pelagianism have never died out; but we must now skip to the fourteenth century. William of Ockham wrote: 'As to Pelagius' error, I say that his view was that everyone has the power, from what is purely natural to him, to avoid every sin, actual and original, and to merit eternal life as his due [*de condigno*]' (*Dubitationes addititie* AA-BB; Leff 1975: 494 n.208). Ockham is applying a distinction foreign to Pelagius' lifetime, between 'condign' and 'congruous' merit. You merit something *de condigno* when it is your due, *de congruo* when it is merely fitting or 'meet' (article 13 of the 39 Articles of the Church of England) (see Sanctification §2). He ascribes to Pelagius the view that human beings before grace already have the power to *earn* salvation. The ascription is unhistorical, but Pelagius does seem committed, as the Massilians also were committed, to the weaker thesis that fallen human nature suffices to make God's cooperative aid a fitting reward for good human effort; to 'do what is in you' is a 'congruous merit'. This was Ockham's own position (in his language); so that it is not surprising that an attack on him and others by his Oxford contemporary Thomas Bradwardine (§3) should have been entitled *De causa Dei contra Pelagium*. Ockham's special contribution had been to maintain that even the fittingness of reward for good effort depends on God's valuing the effort; God might have preferred a will such as Satan's (see William of Ockham §10). Nevertheless, grace is a response; even if not subsequent to the good will in time, it is conditional on it, and thus not prevenient in the sense demanded by the council of Orange. Bradwardine (and Luther, who later concurred) had reason for his label.

5 Grace and free will

During the sixteenth and seventeenth centuries interest in Pelagianism was kept alive by the perceived difficulty for its opponents of reconciling grace with free will. Erasmus, unfriendly to Augustine, defended free will against Luther. According to Calvin, every champion of free will is an enemy of prevenient grace, and so guilty of the Massilian error which was now called semi-Pelagianism. But two ways out were sought by others. The heretical Roman Catholic bishop Jansen, rabid Augustinian, argued that spontaneity is enough for freedom: fallen man is subject to 'the necessity of sinning' (1640, 'De statu naturae lapsae' 3.11), but free will is compatible with necessity because everything is in our power which 'happens when we will' (1640, 'De gratia' 6.5 264a B). Arminius, inspirer of the breakaway Remonstrant movement among Dutch Calvinists, preferred a solution that had meanwhile been urged by Spanish Jesuits, notably in Molina's brilliant *Concordia*: although free will excludes necessity, causes - even divine ones - do not necessitate, so that prevenient grace is not 'irresistible' (1629: 4.122b; 1825 (I): 600) even when 'efficacious' (see Molina, L. de §§1-3). What had turned out to be the major conceptual tangle in the Pelagian question was thereafter bequeathed, without its identification of God's will as

first cause, to the attention of secular philosophy.

See also: Grace; Justification, religious; Predestination; Sin §2

CHRISTOPHER KIRWAN

References and further reading

Abercrombie, N. (1936) *The Origins of Jansenism*, Oxford: Clarendon.(Surveys orthodox anti-Pelagianism in the Roman Catholic tradition, and includes valuable summaries of Molina's *Concordia* and Jansen's *Augustinus*.)

Arminius (Harmens, J.) (1629) *Declarationes*, in *Opera theologica*, Louvain; trans. J. Nichols, *The Works of James Arminius*, London: Longman, 1825.(Rejected strict Calvinism.)

Augustine (418) *De gratia Christi et de peccato originali (On the Grace of Christ and on Original Sin)*, Corpus Scriptorum Ecclesiasticorum Latinorum, vol. 42, Vienna: Tempsky, 1866-.(Contains excerpts from Pelagius' treatise on free will, *De libero arbitrio* (*c*.415); a translation can be found in Schaff (1886-9).)

Augustine (413-15) *De natura et gratia (On Nature and Grace)*, Corpus Scriptorum Ecclesiasticorum Latinorum, vol. 60, Vienna: Tempsky, 1866-.(Contains excerpts from Pelagius' treatise on nature, *De natura* (414); a translation can be found in Schaff (1886-9).)

Augustine (429-30) *Opus imperfectum (Work Unfinished)*, Patrologiae Cursus Completus, series Latina, vol. 45, ed. J.P. Migne, Paris, 1844-55.(Contains long excerpts from the treatise *Ad Florum* (To Florus, *c*.426) by Pelagius' follower Julian of Eclanum.)

Bonner, G.I. (1987) *God's Decree and Man's Destiny*, London: Variorum.(Chapter 11 is valuable on Pelagianism in Pelagius' day.)

Bradwardine, T. (*c*.1344) *De causa Dei contra Pelagium*, ed. H. Savile, London, 1618.(Attack on Ockham.)

Cassian, J. (*c*.415) *De institutis (On Modes of Life)*, Corpus Scriptorum Ecclesiasticorum Latinorum, vol. 17, Vienna: Tempsky, 1866-; trans. E.C.S. Gibson, in P. Schaff and H. Wace (eds) *A Select Library of the Nicene and Post-Nicene Fathers of the Christian Church*, series 2, Oxford: Parker, and New York: The Christian Literature Co., 1890-1900, vol. 11.(A source of what was later called semi-Pelagianism.)

Denzinger, H. and Schoenmetzer, A. (1967) *Enchiridion symbolorum, definitionum et declarationum de rebus fidei et morum (Handbook on the Creeds, etc.)*, 34th edn, Barcelona, Freiburg im Breisgau, Rome, New York: Herder; an earlier edn trans. R.J. Deferrari, *The Sources of Catholic Dogma*, St Louis, MO and London: Herder, 1957. (Contains all the Roman Catholic condemnations of Pelagianism.)

Evans, R.F. (1968) *Pelagius: Inquiries and Reappraisals*, London: A. & C. Black.(A careful examination of the evidence for Pelagius' place in the thought of his time.)

Ferguson, J. (1957) *Pelagius: an Historical and Theological Study*, Cambridge: Heffer. (Remains the most readable attempt at a full appraisal.)

Jansen, C.O. (1640) *Augustinus*, Rouen: Berthelin, 1652.(Magisterial exposition of Augustine on grace, containing the first serious history of the Pelagian heresy.)

Leff, G. (1957) *Bradwardine and the Pelagians: A Study of his 'De Causa Dei' and its Opponents*, Cambridge: Cambridge University Press.(An account of the Ockhamists.)

Leff, G. (1975) *William of Ockham: The Metamorphosis of Scholastic Discourse*, Manchester: Manchester University Press, and Totowa, NJ: Rowman & Littlefield. (Useful for references to Ockham's opinions.)

Molina, L. de (1588) *Liberi arbitrii cum gratiae donis, divina praescientia, providentia, praedestinatione et reprobatione concordia (Concord of Free Will with Grace, etc.)*, ed. J. Rabeneck, Madrid: Society of Jesus, 1953.(Elaborate and undervalued examination of problems surrounding free will.)

Passmore, J. (1970) *The Perfectibility of Man*, London: Duckworth.(Chapter 5 comments on the perennial tension between Pelagianism and Augustinianism in Christianity.)

Pelagius (406-9) *Expositiones xiii epistularum Pauli*, ed. A. Souter, *Pelagius' Expositions of Thirteen Epistles of St Paul*, Texts and Studies 9, part 2, Cambridge: Cambridge University Press, 1926; partly trans. T.S. de Bruyn, *Pelagius' Commentary on St Paul's Epistle to the Romans*, Oxford: Oxford University Press, 1993. (Anonymous, but certainly by Pelagius, though so much corrupted in transmission that his comments on theologically sensitive matters are often hard to recover.)

Pelagius (413) *Ad Demetriadem*, Patrologiae Cursus Completus, series Latina, vol. 30: 15-45, ed. J.P. Migne, Paris, 1844-55; also vol. 33, 1099-1120.(Besides this letter to Demetrias and the *Expositiones* above, a varying number of other surviving letters and short treatises have been attributed to Pelagius. See also Augustine above.)

Prosper of Aquitaine (*c*.428) no. 225 of Augustine's *Epistulae*, Corpus Scriptorum Ecclesiasticorum Latinorum, vol. 57, Vienna: Tempsky, 1866-; trans. P. de Letter, in J. Quasten *et al*. (eds), *Ancient Christian Writers: the*

Works of the Fathers in Translation, vol. 32, Westminster, MD: Newman Press, and London: Longmans, Green & Co., 1946-.(Champion of Augustinianism.)

Rees, B.R. (1988) *Pelagius: A Reluctant Heretic*, Woodbridge: Boydell Press. (Embodies recent scholarship.)

Rupp, E.G. and Watson, P.S. (eds) (1969) *Luther and Erasmus: Free Will and Salvation*, in J. Baillie, J.T. McNeill and H.P. van Dusen (eds) Library of Christian Classics, London: SCM Press, and Philadelphia, PA: Westminster Press, 1953-.(A translation by A.N. Marlow and B. Drewery of Erasmus' *De libero arbitrio (On Free Will)* and Luther's *De servo arbitrio (On the Bondage of the Will)*, with introduction.)

Schaff, P. (ed.) (1886-9) *A Select Library of the Nicene and Post-Nicene Fathers of the Christian Church*, series 1, repr. Edinburgh: T. & T. Clark; Grand Rapids, MI: Eerdmans, 1971-80, vol. 5.(Contains translations of the works by Augustine cited above.)

William of Ockham (*c*.1340) 'Dubitationes addititie', appendix to *Super iv libros sententiarum reportatio*, in Gregg (ed.) *Opera plurima*, vol. 4.(A small part of Ockham's grand philosophical construction.)

Perception

Sense perception is the use of our senses to acquire information about the world around us and to become acquainted with objects, events, and their features. Traditionally, there are taken to be five senses: sight, touch, hearing, smell and taste.

Philosophical debate about perception is ancient. Much debate focuses on the contrast between appearance and reality. We can misperceive objects and be misled about their nature, as well as perceive them to be the way that they are: you could misperceive the shape of the page before you, for example. Also, on occasion, it may seem to us as if we are perceiving, when we do not perceive at all, but only suffer hallucinations.

Illusions and hallucinations present problems for a theory of knowledge: if our senses can mislead us, how are we to know that things are as they appear, unless we already know that our senses are presenting things as they are? But the concern in the study of perception is primarily to explain how we can both perceive and misperceive how things are in the world around us. Some philosophers have answered this by supposing that our perception of material objects is mediated by an awareness of mind-dependent entities or qualities: typically called sense-data, ideas or impressions. These intermediaries allegedly act as surrogates or representatives for external objects: when they represent aright, we perceive; when they mislead, we misperceive.

An alternative is to suppose that perceiving is analogous to belief or judgment: just as judgment or belief can be true or false, so states of being appeared to may be correct or incorrect. This approach seeks to avoid intermediary objects between the perceiver and the external objects of perception, while still taking proper account of the possibility of illusion and hallucination. Both responses contrast with that of philosophers who deny that illusions and hallucinations have anything to tell us about the nature of perceiving proper, and hold to a form of naïve, or direct, realism.

The account of perception one favours has a bearing on one's views of other aspects of the mind and world: the nature and existence of secondary qualities, such as colours and tastes; the possibility of giving an account of the mind as part of a purely physical, natural world; how one should answer scepticism concerning our knowledge of the external world.

1 Perception, objects, appearance and illusion

We perceive objects, features of objects, and events: I can see a lilac bush, notice the texture of a piece of velvet, hear an explosion. We can also perceive facts, that things are a certain way: I may see that there are three empty coffee cups in my office, or smell that the milk has gone off. When we perceive things, they appear a certain way to us and we come to acquire knowledge concerning them, but one can also perceive facts without perceiving any object in particular: scanning the horizon you might see that there is nothing at all in your vicinity; you have perceived that something is the case, but there is no object that you have seen.

Objects can vary in their appearance: how something looks to you depends on the point of view from which you see it; how things appear to you depends on the conditions under which you perceive, for example whether the lighting is good; and they depend on your powers of perception, for example how much you can smell or taste may depend on whether you have a cold. The same object may appear differently to different observers. Furthermore, you may not only perceive different aspects of an object from others, you may misperceive it, or suffer an illusion concerning it.

One misperceives something where one does perceive it, but it appears differently to one from how it really is. Misperceptions occur in different ways: one might mistake a clump of grass for a rabbit; in odd lighting the walls of a room may look peach, even when they are really off-white; one may be subject to illusions due to disease. Other examples of illusion happen in normal conditions of perception to all perceivers. Here is one example, common in psychology text books, called the Ponzo illusion:

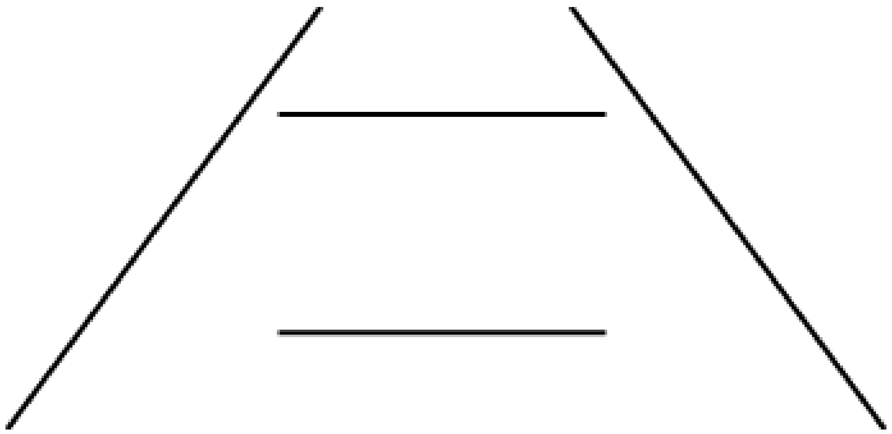

The two horizontal lines are in fact the same length, even though the top one looks longer. Psychologists and neuroscientists are adept at causing other illusions as well: in one alarming example, subjects are made to feel as if their noses are growing back into their heads.

It is also possible to bring about hallucinations. These are cases in which it seems to subjects as if they are really seeing or feeling something, when in fact they are seeing nothing at all. They are most commonly caused by psychological disturbance, but it is also possible to bring them about through appropriate stimulation of the brain: apparent sightings of lights, called phosphenes, arise from stimulation of the visual cortex.

Philosophers tend to draw the same conclusion from all these of kinds of example: it is possible for things to seem a certain way to you, even if they are not that way. Just as one cannot determine from the object of perception how things appear to a perceiver, so one cannot determine from how things appear to a perceiver what the object of perception is, or even whether there is such an object. This implies that an account of perception involves two distinct tasks: on the one hand, it needs to explain what perceptual experience is, the state of mind when things sensorily appear a certain way to one; on the other, it needs to explain what it is for such an experience to be a genuine perception of some object, as opposed to a mere illusion or hallucination.

Under what conditions is a perceptual experience the perception of a given object? One might think that how things appear must match the object in some respect, but there seems to be no particular respect in which I must perceive correctly in order to perceive an object: one can misperceive the colour, shape, location, taste, smell or texture of things. Even if how things appear to me matches perfectly the scene before me, I do not necessarily perceive the scene: if a scientist has induced a visual hallucination in me as of an orange, merely placing an orange in front of my blindfolded eyes will not thereby make me see it. In response to such examples of veridical hallucination, some philosophers have suggested that one can perceive an object only if it causes one's perception of it. This condition is not sufficient by itself to distinguish veridical hallucinations from perceptions: the orange before my blindfolded eyes may be resting on the switch of the machine inducing the hallucination. Various attempts have been made to refine the causal condition in order to give a complete account of the difference between hallucination and perception, but none as yet seems entirely satisfactory in explaining our ability to discern a difference between the two.

The basic task for any theory of perceptual experience is to explain the following. In perceiving, things appear to you a certain way. For example, when you see a red patch on the wall in front of you, it can look to you as if there is a red patch there. In such a case, where you perceive things to be the way that they are, a case of veridical perception, a description of what your experience is like - that is, of how things appear to you - is also a true description of what you can see. In this case, it seems as if a description of your experience is also a description of the things in the world which you can perceive. However, you could have an experience just the same for you even if you were having an illusion, and not seeing something red, or a hallucination and not seeing anything at all. In either case, you would be inclined to give the same description of your experience as in the veridical case, but in neither would the description be true of what was before you in the world, since no red patch would be there. So, how can it be correct to describe your experience in terms of objects in the world, when the description can be true when applied to your experience, but false when applied to the world? This is the question which lies behind the so-called argument from illusion, and different theories of perception answer it in different ways.

2 Sense-datum theories of perception

Some philosophers claim that we are not aware, or anyway not immediately aware, of objects in the world around us, but only of things which depend on the mind for their existence and nature. Suppose that there is no red patch on the wall before me, but that a neuroscientist has so affected my brain that it looks to me as if there is a red patch before me. One might think that it could only appear to me as if some red thing was before me, if there actually was something red of which I was aware. Since the neuroscientist can give me the hallucination just by affecting my brain in the right way, it seems as if bringing about the experience was sufficient for there to be such a red patch for me to be aware of. At least in the case of hallucination, I would be aware of an entity which depended for its existence on my awareness of it; philosophers have used the term sense-datum, for such entities (see Sense-data).

The hallucination could be just the same for me as the experience I would have were I actually looking at a physical red patch on the wall. And the neuroscientist could be bringing this about by stimulating my visual cortex in just the way it would be stimulated if I was looking at the wall. So, it may seem plausible to suppose that my hallucination is of the same sort as the experience I would have were I genuinely perceiving. If the hallucinatory experience is of a kind which involves being aware of a sense-datum, then even in the case of veridical perception I will be aware of such a mind-dependent entity. Hence we arrive at the sense-datum theory of experience: that we are aware of mind-dependent entities in all perceptual experience.

One criticism of this line of thought is that it involves a form of fallacious reasoning. If I believe that there are fifteen elephants in the next room, it does not follow that there are fifteen elephants in the next room which I believe to be there - after all, my belief may be false - so it would be a mistake to accept the latter claim on the basis of the former. Critics suggest that in the above claim, when one moves from the uncontroversial claim that it looks to me as if there is a red patch before me, to the conclusion that there is a red patch which looks to be before me, just such a mistake in reasoning has been made. However defenders of sense-datum theories of perception would agree that the claim does not follow simply as a matter of logic. Rather, they claim that we can only offer an adequate explanation of what experience is like, if we accept that the latter claim is true when the former is.

How good an explanation of what experience is like does this offer? The theory is committed to claiming that when things appear a certain way to you, then some sense-datum actually is that way. Suppose it seems to you as if there is a rabbit in the field before one, is one then aware of some mind-dependent bunny? In order to resist this conclusion, the sense-datum theorist needs to restrict the range of qualities which can strictly be apparent to us; namely to only those which sense-data might plausibly be thought to have. In the case of vision, this has traditionally been restricted to colour, shape and size (there is some dispute over whether visual data are located in a two-dimensional or a three-dimensional visual space). Hence, it could not strictly look (in the relevant sense) as if a rabbit was before one, only as if something rabbit-shaped were there.

If appearances are to be accounted for solely in terms of mind-dependent entities, what connection holds between experience and the objects of perception, such as the rabbit? Representative theories of perception typically hold that material objects are the indirect or mediate objects of perception in virtue of reliably causing our experiences of sense-data. I can perceive the rabbit in front of me, because the visual experience of something rabbit-shaped is caused by the rabbit. One traditional objection to representative theories of perception is that they lead to scepticism concerning the external world (see Scepticism). In response, some philosophers adopt an alternative view of the connection between objects of perception and experience, taking physical objects to be no more than constructions out of mind-dependent entities or experiences (see Phenomenalism; Idealism §2). Few philosophers find this position acceptable. Must a sense-datum theorist choose between scepticism and idealism? One may deny that representative theories of perception introduce any special sceptical problems. Whether this is so or not, depends on whether sense-data should be thought of as cutting us off from the external world, as a veil of perception, rather than providing us with our only access to the world.

Are the metaphysical implications of the sense-datum theory acceptable? It requires us to accept that in addition to such familiar things as rocks and chairs, and discoveries such as black holes and neutrinos, there are mind-dependent objects which come into and go out of existence as each person has an experience. Philosophers have complained that such entities must be inherently mysterious and that we can discern no readily agreed method of determining when one has the same or a different sense-datum. Furthermore, their existence would rule

out explaining how the mind comes to be part of a purely physical, natural world (see Materialism in the philosophy of mind).

Some philosophers have thought that these worries would be lessened if we eliminated inner objects of awareness, and appealed instead simply to ways in which one senses. According to this view, when one senses a red patch, one should not suppose that there has to be some object, a patch which one senses, but rather that one senses redly or in a red manner. One might compare this to singing a lullaby: we need not suppose that there is some thing which the singer sings, rather that they have sung in a certain, quieting manner. Where the sense-datum theory posits inner objects and qualities in order to explain the character of perceptual experience, this adverbial theory of experience appeals just to the manners or qualities of experiencing, sometimes called qualia, subjective qualities or sensational properties of experience (see Mental states, adverbial theory of).

Critics of the adverbial theories have pointed out that we make a distinction between on the one hand sensing a red square and green triangle, and on the other sensing a green square and a red triangle. It is difficult to explain this difference without appealing to the idea that an object is both square and red in the one case, and square and green in the other. There seems to be no sense in which a state of mind, or a sensing, could literally be square or round.

There is a further objection to be made from how we describe our experiences which applies to both adverbial and sense-datum theories. Various philosophers have denied that introspection of one’s visual experience reveals the mind-independent objects which one perceives, and the features which one perceives them to have. Our experience of the world is diaphanous or transparent: introspection of it takes one through to the objects and features in the world, much as in staring at a pane of glass head on, one’s attention is taken through to what lies beyond it.

The objection has two aspects to it. On the one hand, there is a negative claim that no introspective evidence can be found for the existence of mind dependent objects or qualities. Since such things are introduced as just that of which one is aware, one might expect that if there are such things, they should just be obvious through introspection. It is not evident that this negative claim is correct: opponents have claimed that evidence for them is provided by the example of after-images; or in the way that a nearer object can fill up more of one’s field of view than a more distant object, even though both look the same size; or in our awareness of the fact that we are seeing the shape of an object rather than feeling it. Whether such examples really do demonstrate the existence of subjective qualities is a matter of further dispute, but the other element of the transparency objection can be made without having to settle that matter. It points out a positive element of the character of experience, as revealed through introspection, namely that mind-independent objects of perception are there in your experience. An account of experience which appeals solely to sense-data and qualia would not predict that these mind-independent elements should be part of our experience.

One reply to this is to deny that perceptual experiences really do have this diaphanous character. So, it has been suggested that we describe an experience as of a cube merely in order to indicate the typical cause of such experiences. But this suggestion is implausible. We do sometimes pick out experiences indirectly by reference to their typical cause, as when we call a certain distinctive pain a nettle sting. In this case it does not seem as if the term nettle should be applied directly to a description of what the pain is like. But one may claim that visual experience is not like that, the description of experience as of something red, or square, or hard involves applying those terms directly to what one senses.

The objection is more pressing against adverbial theories than sense-datum theories. We describe our visual and tactual experiences in terms of shape properties. As noted above, a sense-datum theorist will dispute whether we also have to describe our experiences in terms of other qualities which could only be the properties of material objects. However, even if the objection is not decisive, it has been influential as a motivation for an alternative account of perceptual experience.

3 Intentional theories of perception

If it can seem to me that there is a red patch on the wall in front of me, when nothing physical is there, and the patch in question seems to be a part of the objective world external to me, then perhaps it can seem to me as if something is so, without anything actually having to be that way. We are all familiar with the fact that one may think that something is the case, when it is not so, as when one believes England will win the World Cup at cricket, and one may think about something, even if it does not really exist, as when small children hope that Santa

Claus will visit their house that night. An intentional theory of perception claims that the case of perceptual experience is parallel to these examples: perceptual experience is intentional, and allows for incorrectness in its content, and non-existence of its objects in just the way that beliefs and judgments may do so (see Intentionality).

When it appears to one as if there is a red patch before one, one's experience represents the world as containing such a patch. If there is no patch there, how one's experience represents the world to be is incorrect, and one's experience is illusory. If the object of one's experience is non-existent, like Santa Claus, then one's experience is a hallucination and not the perception of anything. For the intentional approach, experience is ascribed a representational content which may be correct or incorrect; the experience is veridical, or illusory depending on the correctness of this content. What the experience is like for the subject, as revealed through reflection on it, is to be explained by that representational content. It is important to note that the intentional theory does not introduce the representational content of experience as an object of awareness, in the way that a sense-datum theory introduces sense-data as objects of awareness. Rather, the intentional theory claims that, in the case of perceiving, the objects of awareness are the mind-independent objects that one perceives, and which one's experience represents to be present. In the case of hallucination, there are no objects of awareness, but the character of experience is just the same as that in the case of perception: in both cases it is to be explained by the representational content that the experience has. So the intentional theory offers an account not of the objects of awareness, but rather the mode in which we can be aware of objects in the world.

How like belief could experience be? One view is that we can identify experience with the acquiring of beliefs, or with dispositions to acquire belief. A problem for this suggestion is that one can disbelieve one's senses. Those familiar with the Ponzo illusion do not believe that the top line is longer, but it still looks longer. A belief theory of experience may respond to this example by claiming that you are still disposed to believe that the top line is longer, it is just that other beliefs you have prevent you forming that belief. But this response is not really satisfactory: when you look at the illusion, you then have an experience which could have led to the belief in question; no mere disposition to acquire a belief can be identified with an occurrent state of mind which interacts with one's other mental states.

This suggests that experiencing is not believing, but one might still claim that both states have the same representational content. That claim has also been denied. For the conditions needed for thinking something may be different from the conditions for having certain kinds of experience: it seems conceivable that we can have the same experiences as infant humans or as other animals, but they cannot share the same thoughts as us; our beliefs are expressible linguistically, and rest on sophisticated conceptual abilities which neither infants nor other animals need have. One may also claim that the character of experience is too rich to be encompassed by any one set of categories, or set of concepts that one can bring to bear on it. For these, and related reasons, some philosophers have claimed that the content of experience is different from that of thought or belief in not being purely conceptual: experience has a non-conceptual content as well (see Content, non-conceptual).

Is this sufficient to explain the character of experience, and the differences between experiencing things to be a certain way and merely believing or thinking them to be that way? Many philosophers have thought not. They have insisted that there must be some form of non-representational quality to experience in addition to any representational content it may have (see Qualia). The issue here returns us to the first part of the transparency objection to sense-datum theories: it is an open question whether there are evidently subjective, or qualitative aspects of experience other than how the world is presented to us. If one accepts that there are, then there is more agreement than one might initially have supposed between traditional sense-datum theories, which supplement awareness of sense-data with interpretation, and alternatively intentional theories, which supplement representational content with subjective qualities.

4 Naïve realism and disjunctive theories of appearance

Sense-datum theories of perception appeal to illusions and hallucinations in arguing for the existence of inner objects of perception, and intentional theories of perception appeal to illusions to justify the view that experience is representational. But there is a strand of philosophical criticism which denies that one can draw any conclusion about perception from cases of illusion and hallucination. One motive for such criticism is an endorsement of a certain kind of naïve, or direct, realism concerning perception.

Defenders of sense-datum theories claim that the best account of experience commits one to thinking that one can sense a quality only if there actually is an instance of that quality which one senses - the argument from illusion then leads them to suppose that such qualities must be mind-dependent; defenders of intentional theories claim that the best account of experience commits one to thinking that experience is directed at, or of, mind-independent objects - the example of illusion leads them to claim that one's experience merely represents these things. A naïve realist might endorse both the initial claim of the sense-datum theorist and that of the intentional theorist, while trying to resist their conclusions: on this view, when one perceives, one is aware of some mind-independent objects and their features, and such objects and features must actually be there for one to have such an experience. This cannot be true of cases where there is no appropriate physical object of perception, as in hallucination, so the naïve realist must claim that the account applies only to cases of perception. Hence they must deny that the state of mind one is in when one perceives something is of a sort which could have occurred even were one having an illusion or hallucination. They must claim that when it appears to one as if there is a red patch there, then either there is a red patch which is apparent to one, or it is merely as if there were such a red patch apparent to one: nothing more need be in common between the situations, such as an inner object of which one is aware, or a representational content. The view, which we can call a disjunctive theory of appearances, claims that perceptual experience does not form a common kind of state across perception, illusion and hallucination.

The view has been thought objectionable for a number of reasons. First, as all sides agree, for any perception one has, one could have a matching illusion or hallucination which one could not distinguish from the perception. According to the disjunctive view, there is a genuine difference here, but how can there be a difference in the conscious state of mind which the subject is unable to detect? The disjunctive view is committed to claiming that we can be misled about the kind of conscious state we are in; indeed one might sum up the view as claiming that illusions and hallucinations mislead one not only about objects in the world, but about themselves as well.

Second, the same brain activity as can bring about perceptions can bring about hallucinations. Furthermore, any physical outcome a case of perceiving can produce, for example kicking a ball, could equally well be caused by a matching hallucination. So the disjunctive view must claim that there is a real difference between experience which is a perceiving and one which is a hallucination even though there is no causal difference between them. But many philosophers claim that there can only be a real difference between things where they differ in their causal powers. It remains a matter of dispute whether the consequences of the view are unacceptable or the objections without answer - a similar set of arguments attends debates about the nature of content and thought (see Content: wide and narrow).

See also: Bodily sensations; Empiricism §3; Perception, epistemic issues in; Secondary qualities; Vision

M.G.F. MARTIN

References and further reading

Austin, J.L. (1962) *Sense and Sensibilia*, Oxford: Clarendon Press.(Posthumously published lectures on perception which contain a scathing attack on the argument from illusion and sense-data. Hugely influential in its time, and aspects of it continue to be significant.)

Burnyeat, M. (1979) 'Conflicting Appearances', *Proceedings of the British Academy*, 65: 69-111.(A provocative discussion of the argument from illusion from ancient Greek debate to early twentieth-century discussions of sense-data.)

Crane, T. (ed.) (1992) *The Contents of Experience*, Cambridge: Cambridge University Press.(Various useful articles some defending intentional theories of perception, some forms of subjectivism. The introduction to the collection is also a useful survey of the topic.)

Cornman, J. (1975) *Perception, Common Sense and Science*, New Haven, CT: Yale University Press.(A very comprehensive presentation of the debate and its implications for metaphysics and theory of knowledge.)

Dancy, J. (ed.) (1988) *Perceptual Knowledge*, Oxford: Oxford University Press.(A collection of papers on knowledge and perception, which reprints some of the most important papers on perception in recent years. In particular: P.F. Strawson, 'Perception and its Objects', is a good introduction to the subject; D.M. Armstrong, 'Perception and Belief', presents a belief-analysis of perception while F. Dretske, 'Sensation and Perception', attacks Armstrong, and introduces the idea of perception containing 'analogue information'; H.P. Grice, 'The Causal Theory of Perception' defends the causal theory of object perception, and David Lewis, 'Veridical Hallucination and Prosthetic Vision' offers an account of what it is to see the scene before one, rather than to

see an object, in terms which refine the causal theory; Paul Snowdon, 'Perception, Vision and Causation' presents a disjunctive theory of appearances as a part of an attack on Grice's causal theory; John McDowell, 'Criteria, Defeasibility and Knowledge' presents a disjunctive theory of appearance as a general account of perception.)

Gibson, J.J. (1966) *The Senses considered as Perceptual Systems*, London: Allen & Unwin. (Controversial psychologist, defender of ecological or direct theory of perception and critic of the role of representation in psychological theories, still influential among some philosophers.)

Grice, H.P. (1962) 'Some Remarks about the Senses', in R. Butler (ed.) *Analytic Philosophy*, 1st Series, Oxford: Blackwell, 133-53.(A good discussion of the differences between the senses.)

Harman, G. (1989) 'The Intrinsic Quality of Experience', in J. Tomberlin (ed.) *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*, Atascadero, CA: Ridgeview, 31-52.(Defends an intentional theory of perception, using the transparency argument, denies that there are any non-representational features of experience, and criticizes Peacocke's arguments in Peacocke (1983).)

Jackson, F. (1977) *Perception: A Representative Theory*, Cambridge: Cambridge University Press.(A very clear and vigorous defence of a sense-datum theory of experience; contains very useful criticisms of belief-analyses in ch. 2 and adverbial theories in ch. 3; as well as responses to objections from scepticism to representative theories of perception in ch. 5.)

Marr, D. (1982) *Vision*, New York: W.H. Freeman.(Presents a general cognitive psychological view of vision, equally influential among psychologists and philosophers.)

O'Shaughnessy, B. (1989) 'The Sense of Touch', *Australasian Journal of Philosophy* 67: 37-58. (Interesting discussion of the sense of touch and the contrasts between touch and sight.)

Peacocke, C.A.B. (1983) *Sense and Content*, Oxford: Clarendon Press, ch. 1.(Defends the view that experience has both representational content and certain non-representational features which Peacocke calls sensational properties.)

Peacocke, C.A.B. (1992) *A Study of Concepts*, Cambridge, MA: MIT Press, ch. 3.(Develops a sophisticated variant of the intentional theory involving types of non-conceptual content variously labelled, 'scenario content' and 'proto-propositional content'; modifies the views presented in Peacocke (1983).)

Perkins, M. (1983) *Sensing the World*, Indianapolis, IN: Hackett Publishing Co.(Defends an account of the representative theory of perception somewhat like Jackson's, except its form is closer to that of an adverbialist theory rather than a sense-datum one. Separate chapters discuss each sense modality in turn.)

Searle, J. (1983) *Intentionality*, Cambridge: Cambridge University Press, ch. 2.(A defence of an intentional theory of perception which criticizes both sense-datum theories and Austin's views.)

Swartz, R. (ed.) (1965) *Perceiving, Sensing and Knowing*, Berkeley, CA: California University Press.(A large collection of articles and extracts from books on perception, now almost entirely out of date: a useful source for early- and mid-twentieth-century views on perception.)

Yolton, J. (1984) *Perceptual Acquaintance from Descartes to Reid*, Minneapolis, MN: Minnesota University Press.(A discussion of the history of philosophy of perception which questions the popular attribution of representative theories of perception to various early modern philosophers, including Descartes and Locke.)

Perception, epistemic issues in

We learn about the world through our five senses: by seeing, hearing, smelling, tasting and feeling. Sense perception is a primary means by which we acquire knowledge of contingent matters of fact. We can also acquire such knowledge by, for instance, conscious reasoning and through the written and spoken testimony of others; but knowledge so acquired is derivative, in that it must be based, ultimately, on knowledge arrived at in more primary ways, such as by sense perception.

We can perceive something without acquiring any knowledge about it; for knowledge requires belief, and we can perceive something without having any beliefs about it. Viewing any but the most simple visual scenes we see many things we form no beliefs about. However, when we perceive something, we are acquainted with it by its sensorially appearing (looking, sounding, smelling and so on) some way to us. For we see something if and only if it looks some way to us, hear something if and only if it sounds some way to us, and so on. When, based on how they appear, we form true beliefs about things we perceive, the beliefs sometimes count as knowledge.

Often the way something appears is the way it is. The red, round tomato looks red and round; the sour milk tastes sour. But the senses are fallible. Sometimes the way something appears is different from the way it is. Appearances can fail to match reality, as happens to various extents in cases of illusion. There are, for instance, optical illusions (straight sticks look bent at the water line) and psychological ones (despite being exactly the same length, the Müller-Lyer arrows drawings look different in length). In such cases, looks are misleading.

The ever-present logical possibility of illusion makes beliefs acquired by perception fallible: there is no absolute guarantee that they are true. But that does not prevent them from sometimes counting as knowledge - albeit fallible knowledge. Recognitional abilities enable us to obtain knowledge about things from how they perceptually appear. Sense perception thus acquaints us with things in a way that contributes to positioning us to acquire knowledge about them. The central epistemic issues about sense perception concern its role in so positioning us.

1 Perceptual knowledge is perceiving-that

Sense perception plays at least some role in the acquisition of virtually all our empirical knowledge. Are there any cases of factual knowledge that deserve the label ‘perceptual knowledge’? Sensorially perceiving that *P* is a way of knowing that *P* (in part) by perceiving, and for this reason is fairly standardly taken to deserve that label. Thus it is held that one has visual knowledge that *P* if and only if one sees that *P*, auditory knowledge that *P* if and only if one hears that *P*, and so on: one might see that the light has turned green; hear that the piano is out of tune; smell that the milk is sour; taste that there is garlic in the food; or one might feel the tomato is over-ripe.

We speak of seeing-whether, seeing-when, seeing-where, seeing-how, seeing-who and seeing-which. These all involve seeing-that. For example, one sees whether there is a chair in the room if and only if one sees that there is a chair in the room or one sees that there is not a chair in the room. One sees where the corkscrew is if one sees, for instance, that it is on the table; and one sees who is at the door if one sees that it is Tom. (These claims generalize to the other senses.)

When one sees, hears, smells, tastes or feels that *P*, one believes that *P*, and *P* is true. What distinguishes these ways of perceiving that *P* is the perceptual means by which the belief is acquired or sustained - whether it is by seeing and so on - and the perceptual means of acquisition of at least some of the evidence on which the belief is based.

2 The primary-secondary perceiving-that distinction

It is possible for one to perceive that *A* is (has/would) *F* by perceiving that *B* is *G*, when *A* is distinct from *B* and/or *F* is distinct from *G*. One might, for instance, hear that the alarm clock is ringing by hearing that a ringing sound is coming from the bedroom; smell that the toast is burning by smelling that a burnt odour is coming from the toaster; taste that the wine is from southeastern Australia by tasting that it has such-and-such a taste; or see that the branch would support one’s weight by seeing that it has a certain thickness. In such cases, perceiving that *B* is *G* is part of the constitutive means by which one perceives that *A* is *F*: the way one perceives that *A* is *F*, how one perceives that *A* is *F* is, in part, by perceiving that *B* is *G*. Moreover, that *B* is *G* - something one perceives to be the case - is

part of the evidence on which one's belief that *A* is *F* is based.

Fred Dretske (1995) has drawn a useful technical distinction between primary and secondary cases of perceiving-that *A* is (has/would) *F*. One secondarily perceives that *A* is *F* if and only if one perceives that *A* is *F* by perceiving that *B* is *G*, where *A* is distinct from *B* and/or *F* is distinct from *G*; one primarily perceives that *A* is *F* if and only if one perceives that *A* is *F*, but not secondarily.

A striking feature of secondarily perceiving that *A* is *F* is that it does not require perceiving *A*. One might see that the fuel tank is empty by seeing that the gauge reads empty, and thus without seeing the fuel tank itself. In contrast, when one primarily perceives that *A* is *F*, one does so at least partly by means of perceiving *A*; moreover, one must perceive of *A* that it is *A* and *F*. This condition is met in all cases of perceiving that *A* is *F* - whether primarily or secondarily - partly by perceiving *A*. If one sees the car is touching the kerb (in part) by seeing the car, then one sees of it that it is a car and touching the kerb.

3 Primary perceiving-that and sense experience

A related distinction can be drawn between primary and secondary appears-as-if states. Some *A primarily appears* to one as if it is *A* and *F* if and only if it appears to one as if it is *A* and *F*, but not by appearing to one as if it is *B* and *G*, for any *B* distinct from *A* or any *F* distinct from *G*. *A* secondarily appears to one as if it is *A* and *F* if and only if it appears to one as if it is *A* and *F*, but not in a primary way. When one secondarily perceives that *A* is *F* partly by means of *A*'s appearing to one as if it is *A* and *F*, *A* secondarily appears to one as if it is *A* and *F*. When one primarily perceives that *A* is *F*, *A* primarily appears to one as if it is *A* and *F*. Thus, if one primarily hears that a sound is coming from the left, then a sound coming from the left sounds to one as if it is a sound and coming from the left.

G.N. Vesey (1971) labels appears-as-if states 'epistemic appearances'. Like beliefs, such states have propositional contents and involve the exercise of concepts. However, as Frank Jackson (1977) argues, they are not beliefs or even inclinations to believe. One Müller-Lyer arrow may visually look to me as if it is longer than the other without my believing the one is longer than the other, or even being inclined to so believe. Familiar with the illusion, I may be certain they are of the same length.

Roderick Chisholm (1957) labels the use of sensory-appearance words in such locutions as '*A* appears (looks/sounds, and so on) as if it is *F* and 'It appears as if *A* is *F* their 'epistemic use'. He has distinguished such uses from their phenomenal use, their use in locutions of the form '*A* appears to *S*', where what fills in the blank are terms for sensory properties - properties such as redness, squareness, loudness, saltiness, putridness, and the like - or terms for properties wholly composed of sensory properties. (He also distinguishes their use in locutions such as '*A* looks like an *F* to *S*', labelling it the 'comparative use'.) The phenomenal use of an appearance verb is not equivalent to its epistemic use or any restricted range of its epistemic use. For while *A* cannot appear to one as if it is *F* unless one has the concept of *F*, *A* can phenomenally appear *F* to one without one's having the concept.

When A primarily appears to one as if it is *A* and *F*, *A* phenomenally appears *F* to one. Thus, one is in an epistemic appearance state only if one is undergoing a sense experience. But how else, exactly, epistemic appearances relate to sense experiences is an empirical question that bears on such issues as the functional architecture of our minds, and whether naturalistic accounts of intentionality and sensory consciousness are possible - issues that go to the heart of the mind-body problem. On one currently leading proposal - developed by Dretske, Alvin Goldman, Christopher Peacocke and others - which promises to integrate with current information-processing psychological models of percepts and concepts, sense experiences have a kind of non-conceptual representational content, and states such as epistemic appearances conceptually utilize aspects of their contents. On this picture, sense experiences by which we perceive (rather than hallucinate) discriminate certain sorts of environmental states of affairs within the range of our sense organs. We acquire perceptual recognitional concepts - either through a combination of our genetic endowment and normal processes of maturation or through perception-based learning - and in the process come to undergo epistemic appearances that discriminate states of affairs partly by means of their discrimination by our sense experiences. In acquiring perceptual recognitional abilities, we thus build on our endowed perceptual discriminative capacities in a way that enables us to acquire knowledge about things we perceive from how they phenomenally appear to us.

4 Perceiving-that and the interaction of beliefs

When one perceives that *A* is *F* by perceiving that *B* is *G*, one believes that *A* is *F*, believes that *B* is *G*, and the first belief state is based, in part, on the second. The basing relation involves causal dependence: the basing belief must contribute to causing the other belief or, if the believer already possesses that belief, to causally sustaining it.

Secondary perceiving-that does not require that the perceiver should consciously infer that *A* is *F* from the fact that *B* is *G*. Such conscious inferences are atypical - people in the process of acquiring perceptual recognitional abilities by learning make them (think of learning to read) - but once the abilities are fully acquired, the belief that *A* is *F* is typically formed without the mediation of conscious inference. Indeed, one can see that *A* is *F* by seeing that *B* is *G* without either consciously thinking that *A* is *F* or that *B* is *G*. Such conscious thought episodes are usually absent in cases of perceiving-that.

As Dretske (1995) notes, to perceive that *A* is *F* by perceiving that *B* is *G*, *B*'s being *G* must indicate, in the circumstances in question, that *A* is *F*. The one must indicate the other in the way in which, in certain circumstances, that there is smoke indicates that there is fire. The nature of indication remains a topic of discussion. But on one view if, in certain circumstances *C*, *Q* indicates that *P*, then, it is at least highly probable that if *C* and *Q*, then *P*.

Philosophers are divided over whether in cases of secondary perceiving-that, one must know that *B*'s being *G* indicates (in the circumstances) that *A* is *F*; and even over whether one must be justified in believing that, that is, believe it for reasons that would justify one in believing it. Proponents of strong internalism maintain that in cases of secondarily perceiving that *A* is *F*, one's reasons for believing that *A* is *F* must include that *B*'s now being *G* indicates that *A* is *F*, and that this must itself be held for reasons that suffice for one to know it. Proponents of weak internalism require only that it be held for reasons that suffice for one to be justified in believing it. Proponents of externalism maintain that even the requirement that the perceiver should be justified in believing the relevant indication relation holds grossly over-intellectualizes perceptual knowledge. Externalists charge that internalists conflate the issue of whether people know with whether they know that they do, and/or whether they are justified in believing or claiming that they do. Beliefs that count as perceptual knowledge, they maintain, are warranted by the reliability of their source. Internalists counter that this unacceptably makes perceptual knowledge a matter of luck (see Internalism and externalism in epistemology).

Nevertheless, to perceive that *A* is *F* by perceiving that *B* is *G*, the perceiver must regard *B*'s being *G* as indicating, in the circumstances, that *A* is *F*. That *B*'s being *G* so indicates that *A* is *F* is a theoretical hypothesis about the relationship between *B*'s being *G* and *A*'s being *F*. Since secondary perceiving-that requires a background belief in such a hypothesis, secondary perceiving-that is theory-ladened.

Natural indication plays a role in primary perceiving-that as well. To primarily perceive that *A* is *F*, its basically appearing to one as if *A* is *F* must indicate, in the circumstances, that *A* is *F*. It is controversial, however, whether to primarily perceive that *A* is *F*, one must believe that its appearing to one as if *A* is *F* indicates, in the circumstances, that *A* is *F*. Internalists maintain that such a belief is required and must be at least justified. Externalists typically deny that even such a belief is required; the requirement, they maintain, over-intellectualizes primary perceiving-that. If such a belief is required, perceivers who lack beliefs about how things appear to them could not be perceivers-that. Thus, animals and very young children (who lack the concept of experience) would never perceive that anything is the case. (That is not to deny they ever perceive anything. Perceiving things does not require perceiving-that.)

Even if there is no such belief requirement on primary perceiving-that, primary perceiving-that is, nevertheless, in a sense, theory-ladened: it is dependent on background beliefs. Whether one would, in any situation, believe that *A* is *F* will depend, at least to some extent, on what else one would believe in that situation. As Wilfred Sellars (1968) argues, every mode of belief-acquisition involves, to some extent, the interaction of beliefs (see Sellars, W. §3).

5 Perceptual knowledge is fallible

Proponents of sense-data theory maintain that when *A* is a kind of sense datum, and *F* is a sensory property, its

primarily appearing to one as if *A* is *F* provides one with conclusive reason to believe that *A* is *F*. Sense data cannot be hallucinated, and cannot appear to have sensory properties they lack. Thus, in such cases, if it primarily appears to one as if *A* is *F*, then *A* is *F*. There is no appearance-reality gap here. Sense data are essential components of sense experiences; the way they appear is an aspect of the experience itself. To have a colour-experience, for instance, is just to immediately see or visually sense a coloured visual sense datum. (This is an act-object theory of sense experience.) Sense data are, on these grounds, claimed to be essentially private: only you can perceive yours and only I can perceive mine.

Some sense-data theorists defend an empiricist thesis according to which such primary epistemic appearance states provide an infallible foundation for all empirical knowledge (see Foundationalism §1; Empiricism §4). Introspection affords an infallible way of knowing whether it primarily looks to one as if *A* is *F* (see Introspection §§1-3). One can know a priori that if it primarily looks to one as if *A* is *F*, then *A* is *F*. So, relying only on introspection and a priori reasoning, one can know both what it is that primarily appears to one to be the case, and that it is the case. Such theorists acknowledge that on this indirect realist view, there is a logical gap between states of one’s sense data having sensory qualities and extra-experiential states of affairs involving material objects and physical events. In response to the charge that their view renders the physical world itself unknowable - hidden behind a veil of sensory appearance - they hold that beliefs about such extra-experiential matters can be arrived at by theorizing about the causal factors responsible for our perceiving sense data.

There are two main lines of argument for the existence of sense data. One is the argument from illusion, demonstrated as follows:

(P1) If something looks *F* to one, then one sees something *F*
(P2) In cases of illusion, something looks *F*, but nothing physical that one sees is *F*
(C) Therefore, in cases of illusion, one sees something nonphysical

The argument is standardly supplemented by an appeal to a continuity-of-experience principle to the effect that since our non-illusory experiences are phenomenologically like our illusory ones, if we see nonphysical things when undergoing illusory experiences, we do so also when undergoing non-illusory ones. This principle has been plausibly challenged by J.L. Austin (1962) (see Austin, J.L. §2). But to focus on the argument from illusion itself, it is valid only if ‘see’ is used relationally throughout - used in such a way that if *S* sees *X*, then there is an *X* such that *S* sees it. For the phenomenal use of ‘see’ has no existential import. (A victim of delirium tremens may phenomenally see a pink elephant, even though there is no pink elephant that they see.) If ‘see’ is used phenomenally in (P1), the premise is unobjectionable; but the argument is invalid. However, if ‘see’ is used relationally, then (P1) renders the argument question-begging.

The second line of defence is that nothing physical could be a primary bearer of sensory properties. No compelling case for this has been made, however. According to secondary-quality views of sensory properties, they are dispositions to appear phenomenally in certain ways; for example, redness is the disposition to look red. According to primary-quality views, sensory properties are experience-independent properties; on one version, they are the physical bases for dispositions to appear phenomenally in certain ways; so, for instance, the basis for a physical surface’s disposition to look red might be its having a certain sort of disposition to affect light (for example, a certain sort of spectral reflectance distribution). Both sorts of theories admit of many non-equivalent versions (see Secondary qualities §1; Colour, theories of §3). A secondary-quality theory need not apply that there are sense data. And if any primary-quality theory of sensory properties is correct, then the second line of defence of sense data fails.

The first and foremost problem with sense-data theory is that there is no good reason to believe that there are any sense data, and compelling reason to believe there are none. Philosophers in the Wittgensteinian tradition have challenged the very coherence of the notion of an essentially private entity. (Some of the literature on Wittgenstein’s argument against the possibility of an essentially private language bears on this.) But even if the notion is coherent, the problems for sense-data theory remain legion. Sense data are wholly immaterial. So, unless physics cannot be trusted, they lack causal powers. How, then, do sense experiences influence epistemic appearances? If sense experiences are partly constitutive of epistemic appearances rather than causes of them, such aspects of epistemic appearances could exert no causal influence on beliefs. How, then, could we form beliefs about our sense data?

Intuitive support for sense-data theory comes from the fact that we speak of seeing such things as red after-images. When we see a red after-image, there is nothing physical that we see that is red. But in these cases we see something red only in the phenomenal sense, not in the relational sense. Seeing a red after-image just consists in having a certain sort of visual experience - as of, say, a red circle - acquired in a certain way. Experiences are states or events, and thus are not themselves coloured. Moreover, while we relationally perceive by having sense experiences, we do not perceive sense experiences themselves.

Coloured areas, sounds, odours and tastes, which are among the usual candidates for sense data, arguably meet some of the conditions which sense-data theorists maintain they meet. For example, they play a primary role in perceiving material objects. We can say that a subject basically perceives *A* by undergoing a sense experience *E* if and only if the subject perceives *A* by undergoing *E*, but not by perceiving something else *B* by undergoing *E*. Arguably, sounds are basically heard, odours basically smelled, tastes basically tasted; and, if ‘colour’ is used in an extended sense to include brightness properties (for example, glaring, glowing and so on) as well as chromatic colours (red, green and so on) and achromatic colours (white, black, and shades of grey), what are basically seen are coloured surfaces and volumes.

As David Sanford (1976) and others argue, coloured surfaces and the rest are, however, not sense data. They are publicly observable: you and I can both see the brown surface of the table top, hear the loud sound of the gunshot, or smell the putrid odour of the cheese behind the refrigerator. The brown volume of the Coca-Cola can be photographed; sounds can be recorded. Sounds and odours travel in physical space: objects emit, produce or give off sounds and odours. On a direct realist view, coloured surfaces and the rest are not constituents of sense experiences; rather, they are in physical space before our sense organs and among the causes of our sense experiences.

Coloured surfaces and the rest are, moreover, arguably the primary bearer of sensory properties. A material object is loud only if the sound of it is loud, putrid only if the odour of it is putrid, red only if some surface or volume of it is red. On a secondary-quality theory, they are the primary bearers of the relevant dispositions; on a primary-quality theory, they possess physical bases for the dispositions.

The thesis that coloured surfaces and the rest are basically perceived and primary bearers of sensory properties offers no support to infallible foundationalism. We can have illusory experiences of them and even hallucinate them. A sufferer of delirium tremens might hallucinate the pink surface of an elephant, a sufferer of command-hallucinations the sound of a voice issuing commands. According to intentional theories of sense experience, the first has a visual experience as of a pink surface, the second an auditory experience as of sounds. These experiences have contents. But nothing pink is seen; no sound is heard.

See also: Fallibilism; Sense-data

BRIAN P. McLAUGHLIN

References and further reading

Alston, W. (1986) ‘Internalism and externalism in epistemology’, in *Epistemic Justification*, Ithaca, NY: Cornell University Press, 185-226.(Careful treatment of the varieties of internalism and externalism.)

Armstrong, D.M. (1973) *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.(Offers a causal, reliabilist account of perceptual knowledge.)

Austin, J.L. (1962) *Sense and Sensibilia*, London: Oxford University Press.(Criticizes A.J. Ayer’s sense-data theory.)

Ayer, A.J. (1940) *The Foundations of Empirical Knowledge*, London: Macmillan. (Defends an empiricist theory of the foundations of empirical knowledge which postulates sense-data.)

Chisholm, R.M. (1957) *Perceiving: A Philosophical Study*, Ithaca, NY: Cornell University Press.(Examines the role of epistemic justification in perceptual knowledge; distinguishes the epistemic, comparative and phenomenal uses of appearance words; and attempts to say how sense experiences must be related to objects we perceive by having them.)

Crane, T. (ed.) (1994) *The Contents of Experience: Essays on Perception*, Cambridge: Cambridge University Press.(Contains an essay on the adverbial theory of sense experience - the theory that sense experiences are ways of being sensorially appeared to; an essay on the disjunctive theory of sense experience - a theory according to which there is no experiential state common to perception and hallucination - and contains several

essays on the intentional theory.)
Dancy, J. (ed.) (1988) *Perceptual Knowledge*, Oxford: Oxford University Press. (Contains several seminal papers on epistemic issues in perception.)
Dretske, F.I. (1969) *Seeing and Knowing*, London: Routledge & Kegan Paul, Chicago, IL: University of Chicago Press.(Examines the relational use of 'see'; defends the view that seeing things does not require seeing-that; offers an analysis of seeing things, an analysis of seeing-that, and distinguishes primary from secondary perceiving-that.)
Dretske, F.I. (1980) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press/Bradford Books.(Offers an information-theoretic account of the contents of percepts and of concepts, and of perceiving things.)
Dretske, F.I. (1995) 'Perceptual Knowledge', in J. Dancy and E. Sosa (eds) *A Companion to Epistemology*, Oxford: Blackwell.(Contains the account of the distinction between primary and secondary knowledge presented in §2 above which, along with §4, relies heavily on this essay.)
Goldman, A. (1976) 'Perceptual Knowledge and Discrimination', *Journal of Philosophy* 73 (20): 771-91.(Offers an analysis of perceptual knowledge as involving discriminative capacities, and argues that perceptual knowledge has a pragmatic, contextual dimension. What counts as adequate evidence in a typical country setting to see of something that it is a barn may fail to be adequate evidence if one is in 'phoney barn country', an area peppered with papier mâché barn facsimiles.)
Grice, H.P. (1961) 'The Causal Theory of Perception', *Proceedings of the Aristotelian Society* supplementary vol. 35: 121-68.(Account of perceiving things in terms of causal relations between sense experiences and what is perceived; argues that looks-as-if states are always involved in perception, even though we do not typically use 'looks as if' locutions in such situations since they pragmatically imply doubt-and-denial conditions.)
Hilbert, D. (1987) *Color and Color Perception*, Stanford, CA: Centre for the Study of Language and Information, Stanford University.(Defends a primary-quality theory of colour, according to which surfaces colours are spectral reflectance distributions.)
Jackson, F. (1977) *Perception: A Representative Theory*, Cambridge: Cambridge University Press.(Sophisticated defence of sense-data theory, arguing that only visual sense data could be the primary bearers of colour.)
Jackson, F. and Pargatter, R. (1987) 'An Objectivist's Guide to Subjectivism About Color', *Revue Internationale de Philosophie* 160: 127-41.(Defence of a primary-quality theory of colours as the physical bases of dispositions to look coloured.)
Johnson, M. (1992) 'How to Speak of Colors', *Philosophical Studies* 68: 221-63.(Defends a dispositional theory of colour.)
Lewis, D. (1980) 'Veridical Hallucination and Prosthetic Vision', *Australasian Journal of Philosophy* 58: 239-49. (Analyses the capacity to see in terms of a wide pattern of counterfactual dependence between possible scenes before one's eyes and matching visual experiences.)
McDowell, J. (1994) *Mind and World*, Cambridge, MA: Harvard University Press.(Contains a chapter in which it is argued that coherentism is false since experiences can rationally constrain beliefs, and a chapter in which it is argued that experiences have only conceptual content.)
McLaughlin, B.P. (1984) 'Perception, Causation, and Supervenience', *Midwest Studies in Philosophy* 9: 569-91. (Distinguishes basic from non-basic perception of things using the by-relation as in §5 above; argues that sounds, odours, tastes and coloured surfaces are physical; and argues that since non-basic perception has a pragmatic dimension, the problem of what kind of causal connection pairs perceived objects with sense experiences by which they are perceived should be restricted to basic perception.)
Moore, G.E. (1922) 'The Nature and Reality of Objects of Perception', in *Philosophical Studies*, London: Routledge & Kegan Paul, 1960, 31-86.(Contains an early version of sense-data theory.)
Peacocke, C. (1992) 'Scenarios, Concepts, and Perception', in T. Crane (ed.) *The Contents of Experience: Essays on Perception*, Cambridge, Cambridge University Press, 1994, 105-35.(Develops several related notions of non-conceptual contents of visual experiences.)
Perkins, M. (1984) *Sensing the World*, London: Routledge & Kegan Paul.(Contains a detailed, empirically well-informed discussion of sounds, odours, tastes and coloured surfaces; argues that we are directly sensorially conscious of such things as the colours of surfaces, the loudness of sounds, the saltiness of tastes and the putridness of odours.)
Pitcher, G. (1971) *A Theory of Perception*, Princeton, NJ: Princeton University Press. (Offers a doxastic theory of sense experience according to which sense experiences are suppressed inclinations to believe.)
Price, H.H. (1954) *Perception*, London: Methuen.(Defends sense-data theory.)
Sanford, D. (1976) 'The Primary Objects of Perception', *Mind* 85: 189-208.(Argues that sounds are primary objects of hearing, odours are primary objects of smelling, tastes are primary objects of taste, coloured

expanses are primary objects of seeing, and that theses primary objects of perception are physical existents.)

Sellars, W. (1968) *Science and Metaphysic*s, London: Routledge & Kegan Paul. (Contains seminal essays on epistemic issues in perception. See especially ‘Empiricism and the Philosophy of Mind’, which contains a discussion of the Myth of the Given.)

Strawson, P.F. (1974) ‘Causation in Perception’, in *Freedom and Resentment and Other Essays*, London: Methuen, 76-89.(Discusses the role of causation in the notion of perception, and presents an intentional theory of sense experience.)

Urmson, J.O. (1968) ‘The objects of the five sense’, *Proceedings of the British Academy* 54: 117-31.(Argues that sounds and odours are not sense data, but also argues that nothing in vision plays a role analogous to the roles played by sounds in hearing and by odours in smelling.)

Vesey, G.N.A. (1971) *Perception*, London: Macmillan.(Introduces the term ‘epistemic appearance’, and discusses the phenomenal and relational uses of perception verbs.)

Perfectionism

Perfectionism is a moral theory according to which certain states or activities of human beings, such as knowledge, achievement and artistic creation, are good apart from any pleasure or happiness they bring, and what is morally right is what most promotes these human 'excellences' or 'perfections'. Some versions of perfectionism hold that the good consists, at bottom, in the development of properties central to human nature, so that if knowledge and achievement are good, it is because they realize aspects of human nature. With or without this view, perfectionisms can differ about what in particular is good, for example, about the relative merits of knowing and doing. The most plausible versions of perfectionism affirm both self-regarding duties to seek the excellences in one's own life and other-regarding duties to promote them in other people. Some critics argue that the latter duties, when applied to political questions, are hostile to liberty and equality, but certain versions of perfectionism endorse liberty and equality. Perfectionist ideas can also figure in a pluralist morality where they are weighed against other, competing moral ideas.

1 Broad versus narrow perfectionism

Despite being ignored by English-speaking philosophers for much of the twentieth century, perfectionism is one of the leading moralities of the Western tradition. It is wholly or in part the morality of, among others, Plato, Aristotle, Aquinas, Leibniz, Hegel, Marx, Nietzsche, Brentano, T.H. Green, F.H. Bradley, G.E. Moore and W.D. Ross.

Defined broadly, perfectionism is a teleological morality with an objective theory of the human good (see Good, theories of the; Teleogical ethics; Welfare). As a teleological morality, it is centred on claims about the good and characterizes right action in terms of the good, for example, as that which will result in the most good possible (see Right and good). Its structure, therefore, is similar to that of utilitarianism (see Utilitarianism). But whereas utilitarians traditionally characterize the good subjectively, as consisting in pleasure, the fulfilment of desires, or 'happiness', perfectionists value states or activities of humans other than happiness (see Happiness; Hedonism). They hold that, for example, knowledge, achievement, and aesthetic appreciation are good apart from any pleasure or satisfaction they bring. Their presence in a life makes that life better independently of how much they are wanted or enjoyed, and their absence impoverishes it even if it is not a source of regret.

Defined more narrowly, perfectionism is based on an objective theory of a special kind. It holds that the good consists at bottom in developing one's 'nature', or realizing a 'true self' (see Human nature §1): certain properties are central to one's identity, and one's good consists in developing these properties to a high degree. In some versions of this narrower perfectionism the relevant identity belongs to one as an individual, but more commonly it is shared with all human beings, so that, for example, the properties distinctive of or essential to humans determine the human good. If knowledge and achievement are good, on this view, it is because they realize aspects of human nature.

Narrow perfectionism has often been accompanied by dubious non-moral views, for example, that developing human nature is each human's 'function' or purpose, or that premises about human nature entail conclusions about the good. But it is best considered apart from these views, as making only the moral claim that a being's good is determined by its nature. If this claim can be sustained, narrow perfectionism has two attractive features. Its central ideal of developing human nature is intuitively appealing in itself, or at least has been found so by philosophers from Aristotle to Marx and Nietzsche. And this ideal offers to unify what might otherwise be an unordered list of objective goods. But narrow perfectionism faces serious objections. It must provide a more precise definition of human nature that both retains the intrinsic appeal of the narrow ideal and includes only properties that in themselves seem worth developing. This dual test is not passed by the commonest definitions of human nature, since humans have both distinctive properties (such as killing things for fun) and essential properties (such as occupying space) that do not seem morally valuable. Indeed, many contemporary philosophers doubt whether, especially given the findings of evolutionary biology, any true theory of human nature can ground plausible claims about value (see Evolution and ethics). If they are right, narrow perfectionism is no longer a viable option. The most plausible morality of this kind is not a narrow but a merely broad perfectionism, one that makes objective claims about the good without grounding them in human nature.

2 Theoretical and practical perfections

Different versions of perfectionism can differ in their particular claims about the good. (Among narrow perfectionisms, these differences rest on deeper disagreements about human nature.) Some perfectionists, notably Plato, Aristotle and Aquinas, take the highest human good to be the development of theoretical reason, found especially in the contemplation of knowledge. Others value most, or only certain, practical excellences of acting on the world. Marx equates the human good with productive labour, especially in cooperation with others, while Nietzsche values large-scale, world-transforming creativity. A third view, found in Bradley, values theoretical and practical perfections equally and characterizes them in parallel ways. On this view the best knowledge is explanatorily integrated, with general principles explaining more particular known truths. The greatest practical goods involve a similar integration, with many subordinate ends achieved as means to a single overarching end. This grants perfectionist value to any life that is highly unified, and to particular activities that are complex and difficult.

More specific perfectionist claims find intrinsic goodness in political activity, either by government leaders or democratic participants; in love, friendship and other interpersonal relations; and in the creation or appreciation of artistic beauty. There can also be intrinsic perfectionist evils, such as false belief and failure. A more difficult question is whether moral virtue should be a perfectionist good and vice a perfectionist evil. This connects to the larger issue of how perfectionism accounts for other-regarding duties.

3 Perfectionism and other-regarding duties

Perfectionism gives a central place to self-regarding duties. Whereas some philosophers confine moral evaluation to acts that affect other people, perfectionists affirm moral duties to seek goods such as knowledge and achievement in one’s own life. To be a complete morality, however, perfectionism must also capture other-regarding duties such as the duties not to kill, to relieve starvation, and so on.

Some perfectionists attempt this within a formal structure that is egoistic, so each human’s ultimate duty is only to seek their own perfection. They do so by claiming that virtuous action, including the exercise of other-regarding virtues such as justice and beneficence, is part of the agent’s good (see Egoism and altruism §4; Virtue ethics; Virtues and vices §3). But this egoistic account of other-regarding duties is problematic. In a merely broad perfectionism, one can simply add virtue to a list of objective goods. But especially if virtue is given special weight against other goods, it is hard to see how the resulting morality differs significantly from a non-perfectionist one that simply tells agents to act justly and beneficently without tying this to claims about their good. Some egoistic perfectionists try to derive the goodness of virtue within narrow perfectionism. Thus, some argue that human nature consists in rationality, and that exercising rationality requires acting virtuously. But if there is a sense of rationality in which rationality is plausibly part of human nature, for example, is plausibly essential to humans, this rationality does not seem specially tied to virtuous action. It can be exercised not only in just and beneficent acts but equally in ones that ignore others’ needs (such as philosophizing while others starve) or positively hurt them.

A more promising approach is to give perfectionism a universalistic structure, so each human’s ultimate duty is to promote the greatest perfection of all, with others’ good counting as much as one’s own (see Universalism in ethics). This yields other-regarding duties, though on a distinctive basis: the reason we ought to relieve starvation and not kill is to promote others’ excellence. And, given a universalistic structure, perfectionism can go on to treat virtue as good. It can hold, as Brentano, Moore and Ross have done, that whenever something is good, loving it for itself, that is, desiring, pursuing, or taking pleasure in it for itself, is also good. Then desiring and pursuing the objective good of others, which for perfectionists constitutes virtue, is itself an objective good. This account of virtue is an attractive addition to broad perfectionism, but it presupposes that each agent should already care about the perfection of all. Given a structure that already grounds other-regarding duties, perfectionism can add that fulfilling these duties virtuously promotes one aspect of the agent’s good. It is harder for it, while remaining distinctively perfectionist, to derive other-regarding duties from a concern only with the agent’s perfection.

4 Perfectionism and politics

Applied to political questions, perfectionism uses the same teleological standard as for evaluating acts by individuals: the best government, leader or law is the one that most promotes perfection. Though this is a classical

view of politics, some critics worry that it threatens the values of liberty and equality (see Equality; Freedom and liberty). Will a state committed to objective goods not force citizens into activities it deems valuable, thereby violating individual liberty? Instead of distributing resources equally, will it not concentrate them on a small elite with the greatest talents for excellence? Some perfectionisms, for example, those of Plato, Aristotle and Nietzsche, substantiate these worries, but others, notably Marx's and Green's, endorse liberty and equality. They do so because of specific features of their structure and theories of value.

Perfectionism can most simply value liberty by making the free choice of one's life-activities itself an objective good (see Autonomy, ethical). This perfectionist valuing of choice lies behind the classical liberalisms of Wilhelm von Humboldt and J.S. Mill. In addition, perfectionism can hold, as Green did, that the value of any activity depends crucially on the attitude with which it is done, and that state coercion not only cannot produce good attitudes but is likely to discourage them. These arguments leave room for non-coercive state action to promote perfection, such as subsidizing good activities or in other ways encouraging them.

Some perfectionisms are anti-egalitarian because of their structure. Thus, Nietzsche makes the goal of individual and state action the greatest perfection, not of all, but of the few best individuals. The most attractive perfectionisms, however, give equal weight to gains in perfection by all people. And they will support roughly equal distributions of resources if they can make the following empirical claims: that people's overall talents for perfection do not differ immensely, and that many highly valuable perfections require only modest resources. Some theories of the good, such as Plato's and Aristotle's, make these empirical claims unlikely, but others, such as Green's, support them. If perfection is largely a matter of people's attitudes, then many inexpensive activities can have value, and each person has many chances to have some worthwhile talent. There is further support for equality if the perfections of different people enhance each other. If each person's development requires a similar development for others, as Marx held, then all will do best in a society where all have the resources for excellence. In short, though some versions of perfectionism do deny liberty and equality, others, arguably the most plausible, endorse them.

5 Perfectionism in pluralism

Perfectionist ideas can figure not only in a pure morality but also in a pluralist one where they are weighed against competing ideas about, for example, utility or rights (see Moral pluralism). One possibility is a teleological pluralism with a duty to promote not only perfectionist goods but also utilitarian ones such as pleasure. J.S. Mill's theory, with its distinction between higher and lower pleasures, fits this mixed teleological pattern (see Mill, J.S. §9). Another pluralist morality combines a teleological duty to promote what is good with non-teleological duties about killing, truth-telling and the like, as Ross's theory does. In these pluralist contexts perfectionist ideas make a distinctive contribution, grounding both self-regarding and other-regarding duties to pursue objective goods such as knowledge and achievement. The central perfectionist thought is that these states are good independently of any connection with happiness and that a prime moral duty is to promote them in oneself and in others.

See also: Aretē; Consequentialism; Eudaimonia; Self-realization

THOMAS HURKA

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis: Hackett Publishing Company, 1985.(Immensely influential version of narrow perfectionism.)

Attfield, R. (1987) *A Theory of Value and Obligation*, London: Croom Helm. (Narrow perfectionism extended beyond humans to all living organisms.)

Hurka, T. (1993) *Perfectionism*, New York: Oxford University Press.(Survey of narrow and broad perfectionism.)

Marx, K. (1844) *Economic and Philosophical Manuscripts*, in K. Marx and F. Engels, *Collected Works*, vol. 3, New York: International Publishers, 1975.(Outlines human perfection as productive, cooperative activity.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press, ch. 6.(Expounds an elaborate objective theory of the good.)

Nietzsche, F. (1886) *Jenseits von Gut und Böse*, trans. W. Kaufmann, *Beyond Good and Evil*, New York: Vintage Books, 1966.(Brazenly anti-egalitarian perfectionism.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, section 50.(Contains a liberal-egalitarian critique of perfectionism.)

Raz, J. (1986) *The Morality of Freedom*, Oxford: Clarendon Press.(Perfectionism in political philosophy, with autonomy as a central value.)

Performatives

There are certain things one can do just by saying what one is doing. This is possible if one uses a verb that names the very sort of act one is performing. Thus one can thank someone by saying 'Thank you', fire someone by saying 'You're fired' and apologize by saying 'I apologize'. These are examples of 'explicit performative utterances', statements in form but not in fact. Or so thought their discoverer, J.L. Austin, who contrasted them with 'constatives'. Their distinctive self-referential character might suggest that their force requires special explanation, but it is arguable that performativity can be explained by the general theory of speech acts.

In *How to Do Things with Words* (1962), J.L. Austin challenges the common philosophical assumption that indicative sentences are necessarily devices for making statements. Just as nonindicative sentences are marked grammatically for nonassertive use (interrogative sentences for asking questions and imperative sentences for requesting or ordering), certain *indicative* sentences are also so marked. Austin's paradigm is any sentence beginning with 'I' followed by an illocutionary verb (see Speech acts §1), such as 'promise', 'apologize' or 'request', in the simple present tense and active voice. One can make a promise by uttering the words 'I promise to go', but not by uttering 'I promised to go' or 'She promises to go'. The first-person plural can be performative too, as in 'We apologize… ', and so can the second-person passive, as in 'You're fired'. The word 'hereby' may be used before the performative verb to indicate that the utterance in which it occurs is the vehicle of the performance of the act in question.

Austin contended that these 'explicit performative utterances' are, unlike 'constatives' (statements, predictions, hypotheses, and so on), neither true nor false. In saying 'I promise to go' one is making a promise, not stating that one is making it. A performative promise is not, and does not involve, the statement that one is promising. It is an act of a distinctive sort, the very sort (promising) named by the performative verb. And, according to Austin, making explicit what one is doing is not describing what one is doing or stating that one is doing it.

Now it is also possible to promise without doing so explicitly, without using the performative verb, and this raises the question of whether there is a theoretically important difference between promising explicitly and doing it inexplicitly. A superficial difference is that in uttering the words 'I promise to go' the speaker is saying that he is promising to go and that this, what he is saying, is assessable as true or false. It is true just in case he is doing what he says he is doing - that is, promising to go. In general, in making an explicit performative utterance the speaker is saying what he is doing - and is thereby doing it. Does this mean that performativity requires a theoretically special explanation?

One suggestion is that performativity is a matter of linguistic meaning. Perhaps there is a special semantic property of performativity, so that it is part of the meaning of words like 'promise', 'apologize' and 'request' that one can perform an act of that very sort by uttering a performative sentence containing that verb. One problem with this suggestion is that it implausibly entails that such verbs are systematically ambiguous. For a performative sentence can be used literally but nonperformatively - for example, to report some habitual act. For instance, one might say 'I apologize… ' to describe typical situations in which one apologizes. Moreover, it seems that even if verbs like 'promise', 'apologize' and 'request' were never used performatively, they would still mean just what they mean in fact. Imagine a community of users of a language just like English in which there is no practice of using such verbs performatively. When people there perform acts of the relevant sorts, they always do so, just as we sometimes do, without using performative verbs, such as making promises by saying 'I will definitely… ', giving apologies by saying 'I'm sorry', and issuing requests by using imperative sentences. In this hypothetical community the verbs 'promise', 'apologize' and 'request' would seem to have the same meanings that they in fact have in English, applying, respectively, to acts of promising, apologizing and requesting. The only relevant difference would be that such acts are not performed by means of the performative form. It seems, then, that in our community, where they are sometimes performed in this way, performativity is not a matter of meaning.

It might be suggested that a special sort of convention is required for uses of the performative form to count as promises, apologies, requests, and so on. Explaining this by appealing to convention is gratuitous, however, for performativity is but a special case of a more general phenomenon. There are all sorts of other forms of words which are standardly used to perform speech acts of types not predictable from their semantic content, such as 'It

would be nice if you… ' to request, 'Why don't you… ?' to advise, 'Do you know… ?' to ask for information, 'I'm sorry' to apologize, and 'I wouldn't do that' to warn. In particular, there are hedged and embedded performatives, such as 'I can assure you… ', 'I must inform you… ', 'I would like to invite you… ', and 'I am pleased to be able to offer you… ', utterances to which the alleged conventions for simple performative forms could not apply. Could such conventions be suitably generalized? The variety of linguistic forms standardly used for the indirect performance of such speech acts seems too open-ended to be explained by any convention (or set of conventions) that is supposed to specify just those linguistic forms whose utterance counts as the performance of an act of the relevant sort.

An alternative explanation is needed. In general, speech acts are acts of communication, whose success in that regard requires the audience to identify the speaker's intention (see Speech acts §2). As Bach and Harnish (1979) have argued, what is special about the use of standardized forms of words, such as those illustrated above, is not that they provide the precedent for any conventions. Rather, the precedent provided by standardization serves to streamline the inference required on the part of the audience.

Convention seems relevant to performativity only in certain institutional contexts, where a specific form of words is designated, and often required, for the performance of an act of a certain sort. This is true of those performative utterances involved in, for example, adjourning a meeting, sentencing a criminal, or christening a ship. However, ordinary performative utterances are not bound to particular institutional situations. Like most speech acts, they are acts of communication and succeed not by conformity to convention but by recognition of intention (see Communication and intention).

See also: Austin, J.L.; Searle, J.R.; Speech acts

KENT BACH

References and further reading

Austin, J.L. (1962) *How to Do Things with Words*, Cambridge, MA: Harvard University Press.(Formulates the distinction between performative and constative utterances and proposes a convention-based account of their successful felicitous performance.)

Bach, K. and Harnish, R.M. (1979) *Linguistic Communication and Speech Acts*, Cambridge, MA: MIT Press. (Chapter 10 argues that performativity can be explained inferentially, without appealing to any special meanings or conventions connected with performative verbs.)

Bach, K. and Harnish, R.M. (1992) 'How Performatives Really Work', *Linguistics and Philosophy* 15: 93-110. (Critically examines Searle's conventionalist approach and defends the inferentialist account.)

Recanati, F. (1987) *Meaning and Force: The Pragmatics of Performative Utterances*, Cambridge: Cambridge University Press.(Surveys the literature on performatives and presents an inferentialist account of how they work.)

Searle, J.R. (1989) 'How Performatives Work', *Linguistics and Philosophy* 12: 535-58.(Critically examines the inferentialist approach and defends a conventionalist account.)

Peripatetics

The title 'Peripatetics' designates followers of the philosophical tradition founded by Aristotle: at first those who continued his inquiries, and in the Roman period those who interpreted and commented on his writings. The distinctive Peripatetic tradition was eventually absorbed into Neoplatonism. The adjective 'Peripatetic' is often used as an equivalent of 'Aristotelian'. Peripatetic doctrines were marked by a rejection of the extreme views characteristic of Stoicism.

'Peripatetic' was the name applied to the Aristotelian tradition in philosophy, the name allegedly deriving from a habit of walking about (*peripatein*) while teaching but more probably from teaching in a colonnade or *peripatos*. The independent history of the tradition in antiquity falls into three principal phases. In the sixty years after the death of Aristotle in 322 BC, his followers, and especially the heads of the school (see Theophrastus; Strato), continued his philosophical and scientific inquiries into a wide variety of topics. The school was not concerned to recommend a way of life - or perhaps saw the pursuit of disinterested theoretical inquiry as itself a way of life. Its failure to provide the moral guidance that other contemporary schools were offering (see Epicureanism; Hellenistic philosophy; Stoicism) may have contributed to its subsequent decline; and from the middle of the third to the middle of the first century BC there were few Peripatetics of importance - Critolaus, who accompanied Carneades §1 on the embassy to Rome in 155 BC, is the chief exception - and interest in Aristotle took the form of summary accounts of his doctrines (like that preserved for us by Diogenes Laertius), often reflecting later preoccupations.

This type of interest was continued, but on a more informed basis, by scholars such as Arius Didymus and Nicolaus of Damascus after the event which is commonly thought to have begun the third phase of the Peripatetic tradition, Andronicus of Rhodes' revival, in the mid-first century BC, of direct study of the 'esoteric' or previously unpublished treatises which we possess today. We need not believe the story, in Plutarch and Strabo, that these texts had been hidden in a cellar in Asia Minor until the first century BC. But with Andronicus there began a sequence of commentaries which extended from Andronicus himself to the end of a distinctive Peripatetic tradition and even beyond. There was a corresponding decline of interest in the 'exoteric' or published works, now lost to us.

The commentators concentrated on the aspects of Aristotle's treatises which are 'philosophical' in a modern sense, neglecting the biological, rhetorical and political writings. Important figures were Andronicus' pupil Boethus, Aspasius, whose commentary on the *Nicomachean Ethics* is the earliest of the commentaries we possess except through later reports, and above all, Alexander of Aphrodisias. The commentators sought to reconcile discrepancies between different Aristotelian texts and thus furthered the construction of Aristotelianism as a distinctive system of doctrine. After Alexander those who undertook the study of Aristotelian texts and produced commentaries on them were, with the exception of Themistius, Neoplatonists who regarded Aristotle's philosophy as an introduction to that of Plato (see Aristotle Commentators; Neoplatonism §2).

In philosophical doctrine the distinctive feature of the Peripatetic tradition is moderation. In metaphysics members of the school advocated a this-worldly realism which rejected both Platonic transcendent Forms and Stoic nominalism. In physics and theology a recurring theme is that the universe - regarded with Aristotle as everlasting - is a system in which the same degree of organization cannot be expected at every level. In ethics, although individual formulations varied, Peripatetics followed Aristotle himself both in advocating moderation with regard to the emotions (*metriopatheia*) rather than the Stoic rejection of them (*apatheia*), and in regarding external goods as having some influence on human happiness even though virtue itself was much more important.

See also: Aristotle §30

R.W. SHARPLES

References and further reading

For works by individual Peripatetics, see Carneades, Diogenes Laertius, Plutarch, Strato and Theophrastus. For Aspasius and Themistius, see Aristotle Commentators.

Barnes, J. (1997) 'Roman Aristotle', in J. Barnes and M. Griffin (eds) *Philosophia Togata II: Plato and Aristotle at Rome*, Oxford: Clarendon Press, 1-69.(Important challenge to the long-standing consensus that Andronicus

produced the first edition of the Aristotelian corpus as we know it.)

Gottschalk, H.B. (1987) 'Aristotelian Philosophy in the Roman World', in H. Temporini and W. Haase (eds) *Aufstieg und Niedergang der römischen Welt*, Berlin: de Gruyter, II 36: 2, 1079-1174.(A detailed and clear account of the Peripatetic tradition from Andronicus to the generation before Alexander of Aphrodisias.)

Moraux, P. (1973-) *Der Aristotelismus bei den Griechen (Aristotelianism among the Greeks)*, Berlin: de Gruyter, 3 vols.(The definitive treatment of the Peripatetic tradition in the Roman period. Volume 2 was published in 1984, volume 3 is forthcoming.)

Sharples, R.W. (1997, forthcoming) 'The Peripatetic School', in D.J. Furley (ed.) Routledge History of Philosophy, vol. 2, *Aristotle to Augustine*, London: Routledge. (A general account of the history of the school, with bibliography.)

Wehrli, F. (1969-78) *Die Schule des Aristoteles (The Aristotelian School)*, Basle: Schwabe, 2nd edn.(Collected fragments and testimonia of Peripatetics of the Hellenistic period, with German commentary.)

Personal identity

What is it to be the same person today as one was in the past, or will be in the future? How are we to describe cases in which (as we might put it) one person becomes two? What, if anything, do the answers to such questions show about the rationality of the importance we attach to personal identity? Is identity really the justifier of the special concern which we have for ourselves in the future? These are the concerns of this entry.

In order to answer the question about the persistence-conditions of persons we must indulge in some thought experiments. Only thus can we tease apart the strands that compose our concept of personal identity, and thereby come to appreciate the relative importance of each strand. There are plausible arguments against attempts to see the relation of personal identity as constitutively determined by the physical relations of same body, or same brain. I can survive with a new body, and a new brain. But it does not follow; nor is it true, that a person's identity over time can be analysed exclusively in terms of psychological relations (relations of memory, belief, character, and so on). To the contrary, the most plausible view appears to be a mixed view, according to which personal identity has to be understood in terms of both physical and psychological relations. This is the view which can be extracted from our core (that is, minimally controversial) set of common-sense beliefs about personal identity.

The possibility of the fission of persons- the possibility that, for example, a person's brain hemispheres might be divided and transplanted into two new bodies - shows that the mixed view has to incorporate a non-branching or uniqueness clause in its analysis. The concept of personal identity, contrary to what we might first be inclined to believe, is an extrinsic concept (that is, whether a given person exists can depend upon the existence of another, causally unrelated, person).

Some philosophers have recently tried to forge an important connection between theories of personal identity and value theory (ethics and rationality). The possibility of such a connection had not previously been investigated in any detail. It has been argued that, on the correct theory of personal identity, it is not identity that matters but the preservation of psychological relations such as memory and character. These relations can hold between one earlier person and two or more later persons. They can also hold to varying degrees (for example, I can acquire a more or less different character over a period of years). This view of what matters has implications for certain theories of punishment. A now reformed criminal may deserve less or no punishment for the crimes of their earlier criminal self. Discussions of personal identity have also provided a new perspective on the debate between utilitarianism and its critics.

1 Criteria of personal identity

What is it to be a person? What is it for a person at one time to be identical to some person at a later time? Although the two questions are obviously related, my concern in this entry will be with the second question. (For more on the first question, see Persons; Mind, bundle theory of.) However, I assume this much about what it is to be a person: a person is a rational and self-conscious being, with a (more or less) unified mental life. There are indeed cases (multiple personality, split-brain patients, and so on) in which the apparent lack of mental unity casts doubt on whether a single person occupies a given body (see Split brains). But such cases are exceptional. A normal person is a mentally unified individual. The central question of personal identity is the question of what distinguishes the sorts of changes we mentally unified individuals can survive from the sorts of changes which constitute our death.

On one very familiar view (associated with Plato, Descartes and the Christian tradition) a person can survive bodily death. Bodily death is not the sort of change which constitutes personal death. On this view, a person is an immaterial (that is, non-spatial) soul, only contingently attached to a physical body (see Soul, nature and immortality of the; Soul in Islamic philosophy). This view has few philosophical adherents today. It is fraught with metaphysical and epistemological difficulties. (For example: how can an immaterial soul interact with the material world? How can I know that you have a soul?) In what follows I simply assume, without further argument, that our continued existence is not the continued existence of an immaterial soul.

I do not have to deny that in some possible worlds, there are persons who are immaterial souls; but ours is not such a world. And our concern here is with the conditions for the identity over time of actual (human) persons. Once we

have given up the immaterialist view of ourselves, we can say the following. A person is a psychophysical entity, which is essentially physically embodied. That is a person (a typical adult human, for example) consists of a biological organism (a human body), with a control centre (the brain) that supports their mental life. Persons are essentially mental, and essentially physically embodied. But this is not the end of puzzles about personal identity; it is just the beginning.

When we judge that a friend before us now is identical to the friend we saw yesterday, we typically make this judgment of personal identity under optimal conditions. In such a case, our friend today is physically continuous with our friend yesterday (they possess the very same brain and body). And our friend today is also psychologically continuous with our friend yesterday (they possess the very same beliefs, character, desires, memories, and so on, with only very slight changes). In this case, our identity judgment is true in virtue of the obtaining of both physical and psychological continuities. The puzzle of personal identity is: which continuity (if any) is the more important or central to our concept of personal identity? Evidently, reflection on the paradigm case just described will not help us to answer that question. We need to consider thought experiments where the continuities come apart.

There are three broad accounts or criteria of personal identity over time: the physical criterion, the psychological criterion, and the mixed criterion. These criteria do not purport just to offer quite general ways of telling or of finding out who is who. They also purport to specify what the identity of persons over time consists in: what it is to be the same person over time. According to the physical criterion, the identity of a person over time consists in the obtaining of some relation of physical continuity (typically either bodily continuity or brain continuity). On this view, to be the same person is to be the same living biological object (whether body or brain).

According to the psychological criterion, the identity of a person over time consists in the obtaining of relations of psychological continuity (overlapping memory chains, or memory together with the retention of other psychological features such as well-entrenched beliefs, character, basic desires, and so on). The psychological criterion splits into a narrow version and a wide version. According to the narrow version, the cause of the psychological continuity must be normal (that is, the continued existence of one's brain) if it is to preserve personal identity; according to the wide version, any cause will suffice (normal or abnormal). (See Parfit 1984 ch. 10 for more on this distinction.) Sub-versions of the wide and narrow versions differ over the question of whether any one psychological relation is privileged with respect to identity preservation. (For example, John Locke thought that memory was such a privileged relation.)

Each of the physical and psychological criteria divides into many different versions. The distinctive claim of the mixed criterion is that no version of either the physical or psychological criterion is correct. The best account of a person's identity over time will make reference to both physical and psychological continuities.

I now want to examine in more detail different versions of the physical and psychological criteria. My conclusion will be that all familiar versions of these criteria are open to objection, and that we should accept the mixed criterion. The mixed criterion best captures our core (that is, minimally controversial) beliefs about personal identity.

We can begin with the physical criterion. As noted, this criterion divides into two criteria: the bodily criterion and the brain criterion.

2 Physical criteria

Physical criteria: the bodily criterion. According to the bodily criterion, person A at time t_1 is identical to person B at t_2 if and only if A and B have the same body (that is, they are bodily continuous). Note that A and B can truly be said to have the same body, even though the body at the later time has no matter in common with the body at the earlier time (see Continuants). In such a case, however, the replacement of matter must be gradual, and the new matter must be functionally absorbed into the living body. This is how it is in the life of a normal human being.

The bodily criterion accords with most of our ordinary judgments of personal identity. However, there appear to be logically possible cases in which the deliverances of the bodily criterion conflict with our considered judgments. The particular case I have in mind is that of brain transplantation. Such transplants are, of course, technologically impossible at present; but that is hardly relevant. The speculations of philosophers are not confined to what is

technologically possible.

Sydney Shoemaker was the first to introduce such cases into the philosophical literature. He wrote:

> It is now possible to transplant certain organs… it is at least conceivable…that a human body could continue to function normally if its brain were replaced by one taken from another human body.…Two men, a Mr. Brown and a Mr. Robinson, had been operated on for brain tumors, and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's head and Brown's brain, eventually regains consciousness. Let us call the latter 'Brownson'.…When asked his name he automatically replies 'Brown'. He recognizes Brown's wife and family…, and is able to describe in detail events in Brown's life…of Robinson's life he evidences no knowledge at all.
>
> (Shoemaker 1963: 23-4)

We can suppose, in addition, that Brown and Robinson are physically very similar, and that their bodies are equally suited for the realization of particular dispositions or abilities (for example, playing the piano, or hang-gliding).

The description of this case which commands almost universal assent is that Brown is the same person as Brownson. Virtually no one thinks that the correct description is: Robinson acquires a new brain. Receiving a new skull and a new body seems to be just a limiting case of receiving a new heart, new lungs, new legs, and so on. If Brown is the same person as Brownson, and yet Brownson's body is not the same body as Brown's body, then it follows that the bodily criterion is false.

Physical criteria: the brain criterion. In the light of this example, it would be natural for a defender of the physical criterion to move to the brain criterion: A at t_1 is the same person as B at t_2 if and only if A and B possess the same brain. But is this a plausible criterion of personal identity? I think not. The following scenario is conceivable. Imagine that robotics and brain science have advanced to such a stage that it is possible to construct a silicon brain which supports the very same kind of mental life as that supported by a flesh-and-blood human brain. Imagine also that parts of a human brain (say, a cancerous part) can be replaced by silicon chips which subserve the very same mental functions as the damaged brain tissue.

Suppose that the whole of my brain gradually becomes cancerous. As soon as the surgeons detect a cancerous part, they replace it with silicon chips. My mental life continues as before - the same beliefs, memories, character, and so on, are preserved. Eventually, the surgeons replace all my biological brain with a silicon brain. Since my mental life, and physical appearance and abilities, are unaffected by this replacement, we have no hesitation in judging that I have survived the operation. The procedure preserves personal identity. But is this judgment of personal identity consistent with the brain criterion? The answer to this question depends on whether my (later) silicon brain is deemed to be identical to my (earlier) human brain.

It is plausible to suppose that, if an object (such as a heart, brain or liver) is biological, then that very object is essentially biological. That is, for example, my flesh-and-blood brain could not have been anything but a biological entity. This essentialist thesis is consistent with the view that the function of any given biological object (a human heart, for example) could, in principle, be carried out by a non-biological object (a mechanical pump, say). Hence, I am happy to concede that my later silicon brain is indeed a brain; but it is not remotely plausible to think that it is the same brain as my earlier human brain. Rather, the effect of all the tissue removals and bionic insertions in my skull is to destroy one brain and replace it with another.

Our brain example shows that the sort of matter or stuff with which we replace an object's removed parts can affect the overall identity of that object, even if continuity of form or function is preserved. My (earlier) human brain is not identical to my (later) silicon brain. Yet I survived the operation. Hence, the brain criterion is false.

However, there is a deeper worry about the tenability of the brain criterion. Why did we move to the brain criterion, in response to counterexamples to the bodily criterion? Was it because the human brain is a three-pound pinkish-grey spongy organ that occupies human skulls? No, we moved to the brain criterion because of what the human brain does, namely, supports directly our mental life. It is surely because of its mind-supporting function

that we are inclined to single out the brain as the seat of personal identity. Consequently, we should not see our identity over time as tied necessarily to the continued existence of the human brain we presently have. What matters most is that our stream of mental life continues to be supported by some physical object, not that it continues to be supported by the very same biological organ.

3 Psychological and mixed criteria

I take the bionic brain example to undermine not just the brain criterion, but also the narrow version of the psychological criterion - I survive with a bionic brain, yet the cause of my psychological continuity is abnormal. Hence, it might be thought, the combined effect of these conclusions is to push us towards the wide version of the psychological criterion. Indeed some readers may have wondered why I made no appeal in the first place to examples such as Bernard Williams's brain-state transfer device or Derek Parfit's teletransporter in order to establish the wide version of the psychological criterion.

What are these examples? First, Williams describes a device which can wipe a brain clean, while recording all the information stored in the brain. This machine can then reprogramme another clean brain. The machine can thus preserve psychological continuity in the absence of any physical continuity. Second, Parfit has made much use of the *Star Trek* fantasy of teletransportation. In this scenario, a physical and psychological blueprint is made of a person. The person is then painlessly destroyed. The blueprint is transmitted to another location where, out of different matter, an exact physical and psychological replica of the original person is made.

Some philosophers have claimed that, in both these examples, the original person is identical to the later person, and so concluded that physical continuity is not necessary for personal identity over time. Other philosophers have taken the opposite view, and concluded that physical continuity is necessary for personal identity over time. But these are judgments in which theory is being allowed to dictate intuitions. In truth, there simply is no general agreement about whether the original is the same person as the replica, in cases in which there is psychological continuity but no physical continuity.

Further, there appears to be a decisive objection to the wide version of the psychological criterion - the criterion which holds that personal identity over time consists in psychological continuity, however that continuity is caused. (A defender of the wide version would agree with my verdicts about the counterexamples to the body and brain criteria, and would also regard the normal operation of the brain-state transfer device and the teletransporter as identity-preserving.) Imagine that I step into the teletransporter booth. My psychophysical blueprint is constructed, and sent to another location, where a replica is created. Unfortunately, the machine malfunctions and fails to destroy me. I step out of the booth, intrinsically no different from when I went in. In this case, we have no hestitation in judging that I continue to exist in the same body, and therefore that the replica is not me. But both me-later and my replica stand to me-earlier in the relation of psychological continuity. If the cause of that continuity is deemed to be irrelevant to personal identity, as in the wide version of the psychological criterion, then it ought to be the case that both later candidates have an equal claim to be me. Yet, as we have seen, we strongly believe that I am identical to the later person who is physically and psychologically continuous with me. Consequently, the wide version of the psychological criterion cannot be correct.

I conclude that the best account of personal identity over time is provided by the mixed criterion. We have seen that neither continuity of body nor brain (nor, by extension, the continuity of any other human organ) is a necessary condition for personal identity over time. But we should not conclude from this that psychological continuity, whatever its cause, is sufficient for personal identity over time. As just noted, that thesis does not accord with our intuitions. The most consistent and plausible view that can be recovered from our core set of common-sense judgments appears to be the following: psychological continuity is necessary for personal identity over time; a sufficient condition of personal identity over time is not psychological continuity with any cause, but psychological continuity with a cause that is either normal or continuous with the normal cause (this is why I continue to exist with a bionic brain). One might well ask: why do we have just this concept of personal identity and not some other (such as the wide version of the psychological criterion)? Here, as with other conceptual analyses, there may be no non-trivial answer to this question.

Our discussion thus far has made a certain simplification. The counterexample to the wide version of the psychological criterion exploited the fact that relations of psychological continuity are not logically one-one. In

that example, one of the streams of psychological continuity did not have a normal cause. However, it is logically possible for a person at one time to be psychologically continuous with two or more later persons, even when both streams of psychological continuity have their normal cause (that is, the continued existence of the brain hemispheres). But the relation of identity is logically one-one: I cannot be identical to two distinct people. It seems, therefore, that the sufficient condition for personal identity endorsed in the previous paragraph will have to be modified, unless either such branching is impossible or the possibility of branching can be redescribed so that it does not conflict with our sufficient condition. The problem raised by the possibility of branching continuities is known as the problem of fission.

4 Fission of persons

As we have seen, much work in personal identity has made use of various thought experiments or imaginary scenarios. The method of thought experiments in personal identity has recently been subject to criticism. It has been claimed that we should not take our intuitions about thought experiments as guides to philosophical truth, since such intuitions may be prejudiced and unreliable. These criticisms are, I think, misplaced. For one thing, such criticisms ignore the frequent and legitimate use of thought experiments in virtually all traditional areas of philosophy (for example, in theories of knowledge and in ethics). Why is their use in discussions of personal identity singled out for criticism? Second, and more important, thought experiments can be useful in understanding the structure of a concept and the relative importance of its different strands, provided that there is general agreement about the best description of the thought-experiment. There is such general agreement about the counterexamples to the body and brain criteria. Some philosophers have tried to gain mileage from thought experiments in the absence of such general agreement - for example, the case of teletransportation discussed above. But it would be unwarranted to infer from the existence of such abuses that thought experiments can never perform any useful function in discussions of personal identity (see Thought experiments).

One thought experiment which has been much discussed in recent years, and which does not fall into the teletransporter category of thought experiments, is that of the fission or division of persons. This thought experiment is interesting because it shows us something about the nature or metaphysics of personal identity, about what it is to be the same person over time.

Fission is a situation in which one thing splits into two (or more) things. Fission does occur in nature (amoebae, for example). Fission of persons, of course, does not occur - but it might. We can devise a thought experiment to flesh out this possibility. Consider a person, Arnold. Like us, Arnold has a mental life that is crucially dependent upon the normal functioning of his brain. Arnold also has a property which most of us do not have, but might have done: each of his brain hemispheres support the very same mental functions. If one of Arnold’s hemispheres developed a tumour, that hemisphere could simply be removed and Arnold’s mental life would be unaffected.

Suppose that Arnold’s body develops cancer. The surgeons cannot save his body, but they can remove Arnold’s brain and transplant both hemispheres into two brainless bodies, cloned from Arnold’s body many years ago. Arnold agrees to this and the operation is successfully carried out. The fission of Arnold has taken place. We now have two people - call them Lefty and Righty - both of whom are psychologically continuous with Arnold (same character, beliefs, apparent memories of a common past, and so on). They are also physically similar to Arnold, and each contain a hemisphere from Arnold’s brain. There is both physical and psychological continuity linking Arnold with Lefty and with Righty. Suppose also that Lefty and Righty are in different rooms in the hospital, and exercise no causal influence on each other. How should this case be described - who is who? A number of responses have been suggested, which we shall examine in turn.

(1) ‘The case is not really possible, so we can say nothing about it and learn nothing from it.’ This view is implausible. Hemisphere transplants may be technologically impossible; but they are surely not logically impossible. Indeed, hemisphere transplants, like other organ transplants, appear to be nomologically possible (that is, consistent with the laws of nature). If so, such transplants are also logically possible. Response (1) is not a serious contender.

(2) ‘Arnold has survived the operation, and is one or the other of Lefty or Righty.’ Immediately after fission, Lefty and Righty are physically and psychologically indistinguishable. Both stand in the same psychological and physical relations to Arnold. They both believe that they are Arnold. According to response (2), one is right and

the other wrong.

Response (2) is implausible for two reasons. First, since Lefty and Righty are symmetrically related to Arnold in respect of physical and psychological continuities, the claim that, for example, Arnold is Lefty, can only be sustained on something like the Cartesian view of persons. If we think of the person as an immaterial ego that typically underlies streams of psychological life, we can suppose that Arnold's ego pops into the left-hand stream of consciousness, leaving the right-hand stream ego-less or with a new ego. As noted at the beginning of §1, this view of persons is bizarre. The postulation of such an ego is idle, and conflicts with both science and common sense.

Second, the metaphysical absurdity of the Cartesian view has an epistemic counterpart. According to response (2), when Arnold divides, he survives in one of the two streams. So either Arnold is Lefty or Arnold is Righty. But how can we know which? From the third-person point of view, we have no reason to make one identification rather than the other. Nor is appeal to the first-person perspective of any help: both Lefty and Righty take themselves to be Arnold. Nothing in either stream of consciousness will reveal to its bearer that it is Arnold. So if, for example, Arnold is Lefty, this truth will be absolutely unknowable. There may be no incoherence in the idea of unknowable truths, but we should be suspicious of any theory of personal identity which implies that truths about who is who can be, in principle, unknowable. For these reasons, we should reject response (2).

(3) 'Arnold survives fission as both Lefty and Righty.' There are three ways in which we can understand this response. According to the first way, Lefty and Righty are sub-personal components of a single person, Arnold. According to the second way, Arnold is identical to both Lefty and Righty (hence, Lefty is Righty). According to the third way, Lefty and Righty together compose Arnold (so that two persons are parts of one larger person, just as Scotland and England are parts of one larger country).

These views are hard to believe. It seems plain common sense that Lefty and Righty are both persons (not sub-personal entities), and that they are numerically distinct. Lefty and Righty both satisfy the normal physical and psychological criteria for personhood. They qualify as persons. And they are two. They may be exactly alike immediately after fission, but exact similarity does not imply numerical identity. (Two red billiard balls may be exactly similar, yet numerically distinct.) Further, they will soon begin to differ, mentally and physically, so that it would be intolerable to regard them as anything but distinct persons.

According to the remaining version of response (3), Arnold exists after fission composed of Lefty and Righty, now regarded as persons in their own right. This is sheer madness. The postulation of Arnold's existence in this circumstance (in addition to that of Lefty and Righty) does no work whatsoever. It is completely idle. Second, can we really make any sense of the idea that one person might be composed of two separate persons? That one person might be composed of two bodies and two minds? To be a single person is to possess a unified mental life. (This is why we are sometimes reluctant to regard a split-brain patient as constituting a single person.) Yet, supposedly, after fission Arnold is permanently composed of two unconnected spheres of consciousness. How could they possibly constitute a single person? If Lefty believes that Clinton will win the next election, and Righty believes that he will not, does Arnold believe that Clinton will both win and lose the next election? Such problems multiply. It seems that all ways of understanding response (3) skewer our concept of a person to such an extent that they cannot be taken seriously.

5 Fission of persons (cont.)

(4) 'The case of Arnold's fission has been misdescribed. Lefty and Righty exist prior to fission, but only become spatially separate after fission.' This theory also has different versions. Some philosophers think that only Lefty and Righty occupy the pre-fission body, and that the name Arnold is ambiguous. Others think that three people (Arnold, Lefty and Righty) occupy the pre-fission body, but that only Lefty and Righty survive fission. The differences between these versions of the theory will not concern us.

This theory is very strange. It involves a tremendous distortion of our concept of a person to suppose that more than one person occupies the pre-fission body. Surely to one body and a unified mind, there corresponds only one person? However, the strangeness of response (4) may depend on one's general metaphysics. In particular, the degree of strangeness may depend on whether we accept a three- or four-dimensional view of continuants such as persons.

On the three-dimensional view of persons, persons are wholly present at all times at which they exist (much as a universal, such as redness, is said to be wholly present in each of its instantiations). On this view, persons are extended only in space, not in time, and have no temporal parts. On the four-dimensional view of persons, persons are four-dimensional entities spread out in space and time. Persons have temporal parts as well as spatial parts. Hence, at any given time, say 1993, only a part of me is in existence, just as only a part of me exists in the spatial region demarcated by my left foot. (See Time for more on the contrast between three- and four-dimensional views.)

On the three-dimensional view of persons, response (4) is not just strange but barely intelligible. Consider a time just prior to fission. On this view, two wholly present persons (entities of the same kind) occupy exactly the same space at the same time. This ought to be as hard to understand as the claim that there are two instantiations of redness in some uniformly coloured red billiard ball. On the four-dimensional view, however, Lefty and Righty are distinct persons who, prior to fission, share a common temporal part. It ought to be no more remarkable for two persons to share a common temporal part than for two persons (Siamese twins) to share a common spatial part.

Suppose, for present purposes, that we accept the four-dimensional view. Response (4) is still counter-intuitive. It is implausible to hold that two persons (Lefty and Righty) share a common temporal segment in the absence of any psychological disunity. We should be loath to give up the principle that to each psychologically unified temporal segment there corresponds just one person. Second, there is the problem of how we are to account for the coherence and unity of the I-thoughts associated with the locus of reflective mental life that occupies the pre-fission body. How can there be such unity if two persons occupy that body? These objections to response (4) may not be decisive, but they do show that the multiple occupancy view is problematic, and we should avoid it if we can.

(5) ‘What’s the problem? When Arnold divides into Lefty and Righty, Arnold ceases to exist (one thing cannot be two). Lefty and Righty then come into existence, and are numerically distinct, though initially very similar, persons.’ This is the response I favour. When Arnold divides, there are two equally good candidates for identity with him. Since they are equally good, and since one thing cannot be two things, Arnold is identical to neither. And since there is no one else with whom we could plausibly identify Arnold, Arnold no longer exists. This response respects the logic of identity, and does not violate our concept of a person by supposing that one (post-fission) person is composed of two persons or that more than one person occupies the pre-fission body. This is a victory for common sense!

Indeed it is so. But it is important to realize that, in embracing response (5), we are committing ourselves to a quite particular conception of the identity over time of persons. On this view, Arnold is not Lefty. Why is this true? The reason given is not: because Arnold and Lefty do not have the same body, or because Arnold and Lefty do not have the same (whole) brain. (These would anyway be bad reasons - see §1.) The reason is that one thing cannot be two. Arnold is not Lefty because Righty also exists. Whether Arnold continues to exist depends upon whether he has one continuer or two. Since Lefty and Righty are causally isolated from each other, this implies that the identity over time of a person can be determined by extrinsic factors. Theories that allow for such extrinsicness are sometimes called best-candidate theories of personal identity. According to these theories, B at t_2 is the same person as A at t_1 if and only if there is no better or equally good candidate at t_2 for identity with A at t_1. If there are two equally good candidates, neither is A.

Are such theories, and hence response (5), acceptable? Some philosophers have thought not, but for bad reasons. It has been thought that best-candidate theories violate the widely accepted thesis that identity sentences are, if true, necessarily true and, if false, necessarily false. Is not the upshot of response (5) precisely that Arnold is not Lefty, but that had Righty not existed (had the surgeon accidently dropped the right hemisphere, for example), Arnold would have been Lefty? Here we have to be careful. The widely accepted thesis is that identity sentences containing only rigid singular terms (that is, terms which do not shift their reference across possible worlds) are, if true, necessarily true and, if false, necessarily false. We can read the term Lefty as rigid or as non-rigid. If it is non-rigid (perhaps abbreviating the definite description ‘the person who happens to occupy the left-hand branch’), then it is true that, had Righty not existed, Arnold would have been Lefty. But this result is consistent with the necessity of identity and distinctness. If Lefty is rigid, then the best-candidate theorist, if he is to respect the necessity of identity, must deny that Arnold would have been Lefty if Righty had not existed. If Righty had not

existed, Arnold would then have occupied the left-hand branch, but that person (namely, Arnold) is not Lefty. Lefty doesn't exist in the nearest world in which Righty does not exist, though an exact duplicate of Lefty - twin Lefty - exists there.

Best-candidate theories do not violate the necessity of identity. However, they do have consequences that might be thought objectionable. Consider again the world in which Arnold divides into Lefty and Righty. According to the best-candidate theory, Lefty can truly say, 'Thank goodness Righty exists, otherwise I wouldn't have existed'. Given that Lefty and Righty exert no causal influence on each other, such dependency is apt to seem mysterious.

These consequences are not objectionable. They simply illustrate the fact that properties like being occupied by Lefty (where Lefty is understood to be rigid) are extrinsic properties of bodies. That is, whether the left-hand body has the property of being occupied by Lefty, rather than by Twin Lefty, is fixed by an extrinsic factor (namely, the existence or non-existence of Righty). But this is not counter-intuitive. The property being occupied by Lefty is not a causal property of a body. In contrast with properties of shape and weight, and so on, this identity-involving property does not contribute to the causal powers of any body in which it inheres. (The causal powers of the left-hand body are unaffected by whether Lefty or Twin Lefty is the occupant.) It is typical of a non-causal property that its possession by an object may depend upon what happens to other objects which exercise no causal influence on it. For example, the property of being a widow is not a causal property and, unsurprisingly, whether a woman is a widow may depend upon what happens to someone who, at the relevant time, exercises no causal influence on her. Response (5) teaches us that identity-involving properties (like being occupied by Lefty) are also extrinsic. This is not a counter-example, merely a consequence.

The best-candidate theory provides the most satisfying response to the case of fission. It also reveals something important about our concept of personal identity: its structure is that of an extrinsic concept. This result may be surprising, but it is not objectionable. If we combine this result with the central claim of the last section, we arrive at the following modified sufficient condition for personal identity over time: A at t_1 is identical to B at t_2 if A stands to B in the relation of psychological continuity with a cause that is either normal or continuous with the normal cause, and there is no better or equally good candidate at t_2 for identity with A at t_1.

6 Value theory

In this final section, I investigate the implications (if any) of the metaphysics of personal identity for value theory. A contemporary philosopher, Derek Parfit, is the most well-known advocate of such implications: he argues that the most plausible metaphysics of persons yields radical conclusions for ethics and rationality (value theory).

It has, of course, long been acknowledged that there is a link between theories of persons and value theory. For example, a religious person's belief that we are immaterial souls will obviously bear on their view of the morality of abortion and euthanasia. However, in this case, the value of persons is not called into question; what is in question is simply the extension of the concept person. The intent of Parfit's project is far more subversive: it is to undermine the significance we currently attach to personal identity and distinctness. Whether or not it is ultimately successful, it is important to recognize the form or shape of this project.

The central feature of Parfit's value theory is the thesis that personal identity is not, in itself, an important relation. It is various psychological relations which matter, relations which are concomitants of personal identity in the normal case (but not, for example, in the case of fission). According to this theory, it would be irrational of me strongly to prefer my own continued existence to death by fission.

I will be concerned with arguments for the thesis that personal identity is not what matters. This thesis has two strands. One strand is that personal identity over time is unimportant; the other strand is that personal identity at a time is unimportant. (These are the diachronic and synchronic strands, respectively.) The thesis that the identity of a person over time is unimportant has been taken to undermine the self-interest theory of rationality, and has implications for the tenability of trans-temporal moral notions such as compensation, responsibility and personal commitment. The thesis that the identity and distinctness of persons at a time is unimportant has been thought to lend support to utilitarianism.

The thesis that personal identity over time is unimportant implies that pure self-interested concern is irrational. That is, it is irrational for me to be especially concerned about the fate of some future person just because that

person is me. It follows that the self-interest theory of rationality is false. According to this theory, which has dominated so much thinking about rationality (see Rationality, practical) there is only one future person that it is supremely rational for me to benefit: the future person identical to me. Since the self-interest theory places immense weight on a relation which has no rational significance, this theory cannot be correct.

Further, if we do not believe that personal identity over time is important, this may change our attitude to punishment, compensation and commitment. Consider a case where there are only weak psychological connections between different stages of the same life. (For example, a one-time criminal may now be completely reformed, with a new and more respectable set of desires and beliefs.) On the present view, the grounds are thereby diminished for holding the later self responsible for the crimes of the earlier self, or for compensating the later self for burdens imposed on the earlier self, or for regarding earlier commitments as binding on the later self. The truth that the earlier person is the later person is too superficial or unimportant to support the opposite view.

The thesis that the identity and distinctness of persons at a time is unimportant has been taken to imply that the fact of the 'separateness of persons' is not 'deep', and that less weight should be assigned to distributive principles. The synchronic thesis thus supports (in part) the utilitarian doctrine that no weight should be assigned to distributive principles: we should simply aim to maximize the net sum of benefits over burdens, whatever their distribution (see Utilitarianism).

These are radical claims. They are all underwritten by the thesis that personal identity is not what matters. What are the arguments for this thesis? I shall discuss one argument for the diachronic thesis (the argument from fission), and one argument for the synchronic thesis (the argument from reductionism).

Recall our earlier discussion of fission. I argued that the most plausible description of fission is that the pre-fission person is numerically distinct from the post-fission offshoots. This constitutes the first premise of the argument from fission (which I will present in the first person): (1) I am not identical to either of my fission products. The second premise is this: (2) fission is not as bad as ordinary death. This premise is taken to imply a third: (3) my relation to my offshoots contains what matters. The first and third premises jointly imply that personal identity is not what matters.

This is an interesting argument, which has had many adherents in recent years. But there is a problem with it. The problem concerns the move from the second to the third premise. The second premise is certainly true: the prospect of fission is not as bad as that of ordinary death. What grounds this premise, and what exhausts its true content, is simply that presented with a choice between those two options, virtually everyone would choose fission. Such a choice is both explicable and reasonable: after fission, unlike after ordinary death, there will be people who can complete many of my projects, look after my family, and so on. However, if the third premise is grounded in the second, the claim that my relation to my offshoots contains what matters merely reflects the innocuous truth that fission is preferable to ordinary death. This robs the argument of any radical import. Its conclusion does nothing to undermine the rationality of self-interest, or the rationality of strongly preferring continued existence to both fission and ordinary death. (This argument is not improved if it is merely asserted that fission is just as good as ordinary survival. What is the argument for this claim?)

What of the argument from reductionism? Reductionism is the view that a description of reality which refers to bodies and experiences, but omits reference to persons, can be complete. It would leave nothing out (see Persons §3). The argument from reductionism attempts to show that, if reductionism is true, the fact of the separateness of persons (the fact that you and I are distinct persons, for example) is not deep or significant, and hence less weight should be assigned to distributive principles.

The argument can be presented as follows. Suppose that reductionism is true: reality can be completely described without reference to persons. If such a complete and impersonal description is possible, how can the boundaries between persons be important? Failing an answer to this question, the argument from reductionism concludes that the boundaries between persons are not morally significant.

The validity of this argument turns on the truth of the general principle that if reality can be completely described without referring to Fs, then the boundaries between Fs cannot be of any importance. Both the interpretation and plausibility of this principle are unclear. A more definite worry is that reasons may be presented for dissatisfaction with the argument's premise, reductionism about persons. (In particular: can our mental life really be completely

described in impersonal or identity-neutral terms?) Unless those objections can be met, we should reject the argument from reductionism.

The central arguments for the thesis that identity is not what matters are both open to dispute. The failure of these arguments emphasizes how difficult it is to undermine the importance I attach to the fact that such-and-such a person tomorrow is me, and to the fact that you are not me. Unless other arguments are forthcoming, we can continue reasonably to believe that personal identity is important, and to endorse the traditional views in ethics and rationality which that belief supports.

See also: Alterity and identity, postmodern theories of; Consciousness §1; Morality and identity; Subject, postmodern critique of; Reincarnation §§4-6

BRIAN GARRETT

References and further reading

Descartes, R. (1641) *Meditations*, trans. E.S. Haldane and G.R.T. Ross, Cambridge: Cambridge University Press, 1969.(Famous argument for the distinctness of mind and body, and for the immateriality of the soul.)

Garrett, B.J. (1990) 'Personal Identity and Extrinsicness', *Philosophical Studies* 59: 177-94.

Garrett, B.J. (1992) 'Persons and Values', *Philosophical Quarterly* 42: 337-45.(These articles expand on the material in §§2 and 3.)

Hume, D. (1739) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1978, 251-63.(Classic statement of a sceptical position about the unity of personal identity.)

Lewis, D. (1976) 'Survival and Identity', in A.O. Rorty (ed.) *The Identities of Persons*, Los Angeles: University of California Press.(Lewis defends the multiple occupancy thesis criticized in §3.)

Locke, J. (1690) *An Essay Concerning Human Understanding*, ed. J. Yolton, London: Dent, 1961, bk II ch. xxvii.(A subtle discussion of persons, thinking substances, and consciousness. Locke is a defender of a memory version of the Psychological Criterion of personal identity (see §1).)

Noonan, H. (1989) *Personal Identity*, London: Routledge & Kegan Paul.(First-rate historical and contemporary survey.)

Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press, ch. 2.(Exciting discussion of the 'best-candidate' theory.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Oxford University Press, Part 3.(An excellent introduction to many of the themes of this entry.)

Shoemaker, S. (1963) *Self-Knowledge and Self-Identity*, Ithaca, NY: Cornell University Press.(Influential discussion of issues concerning criteria of personal identity over time.)

Unger, P. (1992) *Persons, Consciousness and Value*, New York: Oxford University Press.(Thorough and imaginative contemporary discussion. Defence of the physical criterion of personal identity.)

Wiggins, D. (1980) *Sameness and Substance*, Oxford: Blackwell, ch. 6.(Contains doubts about the possibility of the fission of persons.)

Williams, B. (1973) 'The Self and the Future', in Williams, B. (ed.) *Problems of the Self*, Cambridge: Cambridge University Press.(Discussion of one famous thought experiment which seems to support the psychological criterion when described in one way, and to support the physical criterion when described in another way.)

Personalism

Personalism is the thesis that only persons (self-conscious agents) and their states and characteristics exist, and that reality consists of a society of interacting persons. Typically, a personalist will hold that finite persons depend for their existence and continuance on God, who is the Supreme Person, having intelligence and volition. Personalists are usually idealists in metaphysics and construct their theories of knowledge by inference from the data of self-awareness. They tend to be nonutilitarian in ethics and to place ultimate value in the person as a free, self-conscious, moral agent, rather than in either mental states or in apersonal states of affairs. Typically, holding that a good God will not allow what has intrinsic value to lose existence, they believe in personal survival of death.

The term 'personalism', even as a term for philosophical systems, has myriad uses. There is said to be, for example, atheistic personalism (as in the case of McTaggart, famous for embracing both atheism and the immortality of the soul), absolute idealistic personalism (Hegel, Royce, Calkins), and theistic personalism (Bowne, Brightman, Bertocci). Leibniz and Berkeley are seen as early personalists; both were theists and idealists. Kant, while not strictly a personalist, was influential in personalism's history. In particular, B.P. Bowne (1847-1910) borrowed freely from Kant, while refusing to accept a Kantian transcendentalism in which our basic concepts or categories apply in a knowledge-giving way only to appearances and not to reality. R.H. Lotze made personality and value central to his worldview, and was a European precursor of American personalism.

1 Criteria for personalism

Some negative criteria may be helpful. Presumably one is ruled out as a serious candidate for being a personalist if one supposes that persons are made up out of some more ultimate constituents on which they depend, or if one supposes that something else has as great value as or greater value than self-conscious agents, or if one supposes that there is neither a Supreme Person nor any other person who exists independently of nonpersons. Presumably, again, a personalist does not think that persons have proper parts - that persons can themselves be components of other persons. To put things positively, for the personalist, persons are simple (noncomposite), of an intrinsic worth not equalled by nonpersons, and at least one of their number is ontologically ultimate.

Kant, Berkeley, Lotze and Bergson were major philosophical influences on Borden Parker Bowne (1847-1910), the principal founder of an American school of philosophy standardly called 'personalism'. This is the central sense of the term, at least in so far as it is applied to American philosophy. Bowne's philosophy is idealistic, theistic and (of course) personalistic. He finds in human self-consciousness, and especially religious and moral experience, clues to the nature of things. In his view, beginning with direct awareness is not only the safest, but the only proper method in epistemology and the only route to a correct metaphysic. Persons are self-conscious agents. In the self-awareness of persons lies the basis of knowledge and the proximate source of continuity through change, a personal God being the ultimate source of all finite being:

> Now when we consider our life at all critically, we come upon two facts. First, we have thoughts and feelings and volitions which are inalienably our own. We also have a measure of self-control, or the power of self-direction. Here, then, we find in our experience a certain selfhood and a relative independence. This fact constitutes our personality. The second fact is that we cannot regard ourselves as self-sufficient and independent in any absolute sense. And a further fact is that we cannot interpret our life without admitting both of these facts.
>
> (Bowne 1908: 280-1)

2 The personalist worldview

Personalism, beginning with these facts, develops a worldview that begins with immediate, self-conscious experience and interprets not only the life of the individual but the world at large in personalistic terms. This involves the claim that the basic categories or fundamental concepts of our thought should be understood in terms applicable to persons and their experiences.

The notions of substance, unity, identity and change find their home in personal unity and identity. Each person,

aware of themselves as having fleeting experiences and as existing throughout those experiences, is acquainted with a self-conscious intelligence that experiences change and hence endures. What unifies our experiences at any time is that all are states belonging to a single mind, and what provides experiential continuity over time is that mind's continued existence as a subject of experience. Our self-awareness, then, gives us knowledge of our status as mental substances or selves; memory plus self-awareness grants us knowledge of our own identity and continuity.

The notions of contingency, dependence and freedom receive similar treatment. Each person is aware of thinking and acting, affecting their environment and guiding much of the course of their thought and conduct; they are also aware of things that intrude themselves on their attention, impede their will, and externally aid or thwart their wellbeing. Thus there is consciousness both of a degree of independence, in which one somewhat transcends one's environment, and a degree of dependence, in which one's life and wellbeing are dependent on what one encounters. We are aware of being able to act freely within a large but not unlimited framework. This experiential complexity is an accurate clue to our existence as distinct but dependent creatures, individuals who are finite responsible agents.

The category of causality also receives personalistic analysis in connection with Bowne's idealism. This view, in part inherited from Berkeley (§§3, 6-8), consists in holding that there are no mind-independent extended things; what ordinary perception consists in, and what science's laws range over, is human experiences with perceptual contents. There are then, in Bowne's view, no physical objects, only perceptual experiences, and science does not require the existence of mind-independent things. In addition, our perceptual experiences succeed one another, but nothing in our awareness of them informs us that a preceding perceptual experience brings about or produces its successor; causality, in so far as perception reveals it or science discusses it, is but a matter of sequential regularities. Science records perceptual regularities but does not recount genuine causal connections.

None of this is intended to dismiss science or to deny its achievements. Bowne here echoes Berkeley's view that, in effect, while it is clear that the laws of physics are true, it is not at all clear that what they are true of is mind-independent. The equations of basic physics presumably are correct; what they range over is not a matter for science to decide. Direct awareness, however, does decide it, and what they are true of is perceptual experiences. Whatever the merits of this view of science, it was neither Berkeley's nor Bowne's intention to reject scientific data. Bowne was quite content to accept evolutionary theory, and viewed it (in an idealist version) as telling us much about the process by which God brought about finite persons in an environment suitable for moral growth and religious awareness.

None the less, the fact that science is possible is, suggests Bowne, highly significant. It is not logically impossible that our experiences should be random and our attempts to make sense of our perceptual environment - to predict, understand, explore and exploit it - regularly fail. That our perceptual environment is composed of experiences with perceptual contents rather than extended mind-independent objects in no way precludes this unfortunate scenario as a live option; the perceptual world might have been simply incoherent. Since we do not produce our perceptual experiences, we do not cause their orderliness.

The concepts of space and time also receive personalistic treatment. If all there is to what we refer to as extended physical objects is the contents of our perceptual experiences, then we have no reasons to suppose that space itself has mind-independent existence. Space is nothing beyond relationships between human experiences with perceptual content. Since the mind itself is not identical to its perceptual experiences or reducible to the contents of those experiences, minds are not in space.

More surprising, perhaps, is Bowne's contention that the mind itself is not in time. Time, too, he claims, is but a relationship between human experiences (perceptual or not), and since the mind is what *has* experiences, it is not included in relationships that hold only among its experiences. Hence it is not in time. (Presumably Kant's views particularly influenced Bowne here.)

3 Idealism and theism

In the case of each of the basic concepts discussed above, Bowne contends, we find no reason to suppose that their role in knowledge resides in ranging over things that are neither minds nor mind-dependent. Neither do we find reason to think that what they apply to in knowledge-giving ways is something less or other than what is actually

there to be known.

The possibility of science, Bowne holds, requires explanation. So does our existence as enduring individuals. Further, our existence, not only as selves, but as agents who partially transcend their environments also requires explanation.

Bowne contends that the only causation we experience is the exercise of our own powers as agents; here, he holds, we are aware of something being brought about or produced by something else, not merely of one kind of thing temporally following some thing of another kind in a way typical of members of those kinds. Since we have experiences that we did not cause that reflect opinions we do not hold, values we do not accept, or plans we did not make, and since we generally represent others as aiding or opposing us, and agreeing or disagreeing with us, it is most reasonable to ascribe these to minds like ourselves. The idealist's ground for believing in minds other than one's own is that experiences one has but did not cause are to be explained by one's mind being affected by another mind - mental telepathy (or its close cousin).

Further, if we stick to what our experience suggests, then we will explain the existence of mental substances which enjoy continuity of existence and regularity in sense perception by a mind whose effects these things are - a God who has created us as real but limited agents. In this way, among others, Bowne's idealism and his monotheism mesh. Berkeley had suggested that those perceptual contents that we refer to as providing our natural environment are best viewed on the analogy of a divine language - as experiences that God causes and which manifest a constant structure that renders our experience manageable, that is, as a case of sustained benevolent mental telepathy (or its close cousin). Bowne follows Berkeley's example.

Similarly, we will explain the existence of finite minds by the creative action of an infinite mind whose existence does not itself need explanation, and as we find that our actions, in so far as they are rational, stem from our purposes, we may take it that this mind intended to produce creatures such as ourselves. Thus we are led to believe in a personal God possessed of intelligence and will and (since it is good that persons be created) moral goodness. Theism provides a single explanation that traces each of these facts to the same source. Of course, it is logically possible that these inferences may have led us to false conclusions, but they are eminently reasonable inferences - more so than any competitors; and we have no good reason to think the conclusions in question to be false.

Since, Bowne argues, no mind is or can be a part of any other mind - necessarily each is distinct from every other - the notion of a single cosmic mind of which each finite mind is a piece or a partial manifestation is incoherent. Further, our independence, limited though real, argues against any such view, as do the facts that we have false beliefs and perform wrong actions. So we may dismiss any notion of a 'world mind' of which we are modes.

Similarly, since we plainly are aware of our own existence and experiences, we may dismiss any suggestion that our self-awareness is somehow illusory; for that matter, illusions too require self-conscious subjects. Hence we may also dismiss any notion of there being a qualityless reality underlying the one we experience, and the view that what exists is utterly unlike what we experience.

At this point, personalism's metaphysics links up with its ethics. Persons as self-conscious agents possessed of freedom have inherent worth; they themselves - not certain of their more pleasant or more contemplative states - are the locus of ultimate value. Since there is a morally good God who will not allow intrinsic personal worth to be destroyed, there is individual immortality.

4 Howison and Brightman

George Holmes Howison (1834-1916) was another American personalist. A philosopher and mathematician, he graduated from Marietta College (Ohio) and taught mathematics at Washington University (St Louis), where he joined the St Louis Philosophical Society. He also taught philosophy at the Massachusetts Institute of Technology, Harvard Divinity School, and the Concord School of Philosophy, moving in 1884 to the University of California to organize a philosophy department there.

In sharp contrast to Bowne, Howison thought that a person who depended for existence on God could not be a free moral agent. Hence he posited self-determining persons, each of whom seeks in their own way to fulfil themselves as an individual. Persons are not created and have no beginning to their existence. In so far as one seeks fulfilment properly, one does so in ways that do not infringe on others doing likewise. God, on his account, serves as an

example of the moral ideal - a sort of Aristotelian final cause rather than an efficient cause, a standard of perfection but not a creator. Howison views evolution not as having produced a society of interacting minds each seeking fulfilment, but as the product of that seeking.

Bowne's successor at Boston University was another personalist, Edgar Sheffield Brightman (1884-1953). A student of Bowne's, he taught at Nebraska Wesleyan University (1912-15), Wesleyan University (1915-1919), and in 1919 at Boston University, where in 1925 he became the first Borden Parker Bowne Professor of Philosophy. Perhaps his best-known suggestion is that the divine nature contains a 'given' which God's will neither created nor approves. This, Brightman held, is the primary source of evil. God is constantly gaining control over the given, and thus growing as a person, but complete victory is not in sight. While distinctive, this perspective is typical of neither monotheism nor personalism. Bowne's own approach to evil appealed to human freedom and the conditions under which the growth of finite minds to moral maturity is possible.

Personalism as a school has not achieved the fame or influence of pragmatism (another American philosophical perspective). It is most accurately seen as a type of theistic idealism that does not purport to offer a new philosophical perspective. In its articulation by Bowne, and later Brightman and Bertocci (Brightman's successor at Boston), its philosophical perspective is at least as clear and cogent as its contemporary competitors.

See also: Idealism

KEITH E. YANDELL

References and further reading

Bertocci, P. (1952) *An Introduction to the Philosophy of Religion*, New York: Prentice Hall.(An introductory study by Brightman's successor that focuses on the teleological argument.)

Bertocci, P. (1970) *The Person God Is*, London: Allen & Unwin. (A defence of monotheism.)

Bowne, B.P. (1908) *Personalism*, Boston, MA: Houghton Mifflin.(A popular presentation of the personalist viewpoint, based on the N.W. Harris Lectures of 1907 at Northwestern University; see the entry on Bowne for a list of his works.)

Brightman, E.S. (1940) *A Philosophy of Religion*, New York: Prentice Hall.(An introductory discussion by Bowne's successor of most of the issues in philosophy of religion.)

Cunningham, G.W. (1953) *The Idealistic Argument in Recent British and American Philosophy*, Century; Westport, CT: Greenwood Press, 1969.(Discusses a century of Anglo-American idealist thought (1808-1910) in expository and critical terms; includes a chapter on Bowne.)

Deats, P. and Robb, C. (eds) (1986) *The Boston Personalist Tradition in Philosophy, Social Ethics, and Theology*, Macon, GA: Mercer University Press.(Discussion of personalism's roots, trunk and branches by thinkers associated with the tradition.)

Flewelling, R.T. (1915) *Personalism and the Problems of Philosophy*, New York: Methodist Book Concern.(A sympathetic discussion of personalism that concentrates on Bowne's position, which it sets in philosophical context.)

Howison, G.H. (1934) *George Holmes Howison, Philosopher and Teacher*, ed. J.W. Buckingham and G.W. Stratton, Berkeley, CA: University of California Press. (Collection of Howison's works.)

Persons

We are all persons. But what are persons? This question is central to philosophy and virtually every major philosopher has offered an answer to it. For two thousand years many philosophers in the Western tradition believed that we were immaterial souls or Egos, only contingently attached to our bodies. The most well-known advocates of this view were Plato and Descartes. Few philosophers accept this view now, largely because it is thought to face a number of intractable metaphysical and epistemological problems (for example: how can an immaterial soul or mind interact with the material world? How can I know that you have a soul?). The recoil from Cartesianism has been in three different directions. One direction (the animalist) emphasizes the fact that persons are human beings, evolved animals of a certain sort. A second direction (the reductionist) is represented by David Hume: the self or person is not a Cartesian entity, it is a 'bundle of perceptions'. Finally, there is a theory of persons influenced by the views of John Locke, according to which persons are neither essentially animals nor reducible to their bodies or experiences.

1 What is a person?

We can all tell persons apart from other kinds of thing, such as kangaroos, mountain ranges and billiard tables. We are also very good at re-identifying the same person at different times. These are cognitive skills we all possess and exercise effortlessly all the time. So we all know that Madonna is a person but the Empire State Building is not, and the 1990 Governor of Arkansas is the same person as the 1997 US President. But can we say anything at a general or abstract level about the grounds of these uncontroversial truths?

Whatever else might be required to be a person, it seems clear that a person is a mental being. A person is the type of entity that possesses a mind. (The mind does not have always to be conscious: a sleeping or comatose person is still a person.) We can be more specific. Not just any mental being is a person. My cat is a mental being - it can feel pain or hunger, for example - yet it is not a person. So a person is (at least) a being that possesses a particular sort of mind. A person does not just have sensations of pain and pleasure, but also intentional, world-directed, mental states like belief, desire, fear and hope.

Indeed, persons possess a range of particularly sophisticated mental states, including self-reflective mental states. I am capable of having not just the belief that it is raining but also beliefs about myself. These are not just beliefs about someone who happens to be me (as when I think 'the person born on such-and-such a day is Scottish', referring to myself but forgetting that *I* am that person). They are fully self-conscious beliefs about myself, the sort of beliefs I have when I say 'I remember that it snowed last Christmas' or 'I intend to holiday in Fiji when this term is over'. Persons are self-conscious mental beings. This observation is confirmed when we reflect on how much of what matters in our mental life and social interactions presupposes the self-consciousness of ourselves and others.

This view of persons was expounded clearly by John Locke in the seventeenth century (see Locke, J. §8). Locke wrote that a person is 'a thinking, intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places' (1689, 2 (27): 9). In this definition, Locke specifies some of the elements that comprise our concept of self-consciousness. In particular, he cites thinking, intelligence, reason, reflection and the ability to engage in tensed first-person judgements. These features he holds to be constitutive of our concept of a person. We now have the more abstract and general account we were seeking of why Madonna is a person and the Empire State Building is not. In addition, we also have a more abstract and general account of the grounds for our ordinary judgements of personal identity over time. On this view, the reason why the 1990 Governor of Arkansas is the same person as the 1997 US President is because a single stream of consciousness (continuity of character, beliefs intentions and so on) uniquely links the 1990 Governor with the 1997 President (see Personal identity §1).

An important feature of Locke's account is that his definition of a person in no way presupposes the answer to a more traditional question: namely, to what ontological category do persons belong? That is, are persons immaterial (non-spatial) souls only contingently attached to their bodies (as Plato and Descartes believed)? Are persons wholly material beings? If so, are persons necessarily animals of a certain sort, or might there be robot persons? On Locke's view, these questions cannot be answered simply by examination of our concept of a person.

Reflection on what it is to be a person cannot alone decide the question of whether we are immaterial souls, animals or bionic beings. In order to decide that question we need to investigate ourselves empirically, albeit in a fairly undemanding way. On the strongest version of Locke's view, not only is it not true a priori that all persons are animals, it is not necessary either. If all persons are human beings, this is just a contingent truth.

2 Persons and human beings

We must be clear about what is and what is not controversial in Locke's view. It is not controversial that the features Locke cited (reason, reflection and so on) are central to our concept of a person. What has seemed controversial to some is the stronger claim that truths about the ontological status of persons (for example, that all persons are animals) are contingent. This stronger claim is denied by animalists. They hold that all persons *must* be animals. They do not hold that all persons have to be human beings - chimpanzees and dolphins, for example, could qualify as persons if their behaviour revealed a suitably impressive mental life.

This move away from the Lockean position is apt to seem unpersuasive. To start with, animalists can hardly be presenting their definition as an a priori conceptual truth. People who believe in the actual existence of non-animal persons (God, angels, souls and so on), or people who believe in the possibility of robot persons, do not appear to be committing any conceptual error. Further, the truth that I am a human being (or, more generally, an animal) seems resolutely a posteriori. Animalists must therefore conceive of their definition as a necessary a posteriori truth. And its source must lie in the further supposed truth that animality is a necessary a posteriori constraint on possession of the mental life which is characteristic of persons. But what reason do we have to believe this? As far as I can see, we have none.

More tellingly, the most plausible description of certain thought experiments counts against the animalist position. I can make perfectly good sense of the possibility that I might gradually become an entirely bionic being. My various bodily and brain parts may gradually be replaced by functionally identical bionic parts. Provided that the changes preserve my continuing mental life, abilities and appearance, we would have little hesitation in saying that I survived a process of total replacement. Yet, at the end of the process, I am not an animal of any sort. Consequently, it cannot be true that all persons must be animals. Hence, we have good reason to think that the truth that all persons are animals is merely contingent. In addition, we also have good reason to think that the relation between a person and the human being they share their matter with is not that of numerical identity. If I can survive the destruction of the human being with whom I am presently spatially coincident, I cannot be identical to that human being. The Lockean view has been vindicated.

3 Reductionism

The issue of reductionism is perhaps the most complex topic that any theory of persons must address. In order to assess reductionism, we must be clear about what the position involves, what reasons there are to believe in it, and what arguments there are against it. Reductionism is an ontological view of persons. We can begin to understand the view by contrasting it with two different ontological views of persons. One view is Cartesian: persons are immaterial substances which are only contingently linked to their bodies (see Descartes, R. §8; Dualism). The other view, which we might call the Intermediate view, holds that persons are psychophysical substances, which are necessarily embodied. This is a version of the Lockean view defended in the previous section.

Unlike defenders of the latter views, the reductionist holds that the ontological status of persons is secondary or derivative. The primary elements of our ontology are mental states (pains, 'thinkings', 'rememberings' and so on) and physical bodies. Reductionism has received expression in a number of ways. The most common is that a person is nothing 'over and above' their body and mental life. Alternatively, facts about persons and personal identity are 'verbal'. It is a verbal issue whether we choose to call two people at different times 'the same person', just as it is a verbal decision whether to call seasickness 'pain'. However, these suggestions are rather vague, and need to be sharpened. The key thought is that we can completely describe all mental and physical events impersonally, without reference to persons. The experiences that make up a person's life, on this view, are not essentially events in the life of a person, any more than the bricks that make up a house are essentially parts of a house.

Thus understood, why should we believe in reductionism about persons? One reason is that reductionism has been

thought to be the only serious alternative to Cartesianism. But as we have seen, this overlooks the possibility of the Intermediate view. It is worth emphasizing that there is no instability in the Intermediate view. This view implies that no reductive account of persons is available. This would be problematic only if all genuine concepts and relations must admit of reductive analyses. That condition is simply a dogma.

It might be thought that another reason favouring reductionism is Hume's famous remark about the unencounterability of the subject in introspective experience. Hume wrote that ' when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never catch myself at any time without a perception, and never can observe anything but the perception' (Hume 1739 (1976): 252). However, Hume's failure to encounter himself in experience is a fact about the phenomenology of inner reflection. It does not imply that the self does not exist or that we can completely describe all mental life without reference to persons.

Reductionism thus lacks any compelling motivation. In addition, the view faces a number of difficulties. First, reductionists have a problem explaining why we find it so hard to make anything of the idea of an unowned or subjectless experience. A particular toothache, for example, has to be *had* by someone. Experiences require subjects. But there is a deeper problem for the reductionist. The mental life of a human being does not, of course, simply consist in experiences of pleasure and pain. The mental life of a normal human includes mental states with complex contents, for example, intending to visit grandmother next month, or remembering how Alice Springs looked at dawn last year. The contents of these mental states appear to presuppose personal identity. I can only be said to remember, from the inside, *my own* experiences; I can intend only that *I* do such-and-such. But in that case there seems to be no prospect of a complete and impersonal description of these mental states - personal identity is built into them.

The standard manoeuvre in the face of this objection is to invoke the concept of, for example, quasi-memory. This concept is stipulated to be like ordinary memory in all relevant respects, except that quasi-memory does not presuppose personal identity. I can quasi-remember, from the inside, someone else's experiences. For example, as the result of a brain-graft, I may come to have a quasi-memory of standing in front of the Taj Mahal (even though I know that I have never been to India). Since quasi-memory does not presuppose personal identity, the reductionist can re-describe the psychological life of a person without using terms, such as memory and intention, which presuppose personal identity.

The crucial question is whether the notion of quasi-memory can serve the reductionist's purposes. Is the concept of quasi-memory intelligible independently of the concept of memory? Is the content of a quasi-memory really identity-neutral? Quasi-memories are *illusions* of memory; so the content of a quasi-memory is not identity-neutral, it is merely illusory in respect of its identity-involving content. This parallels what we are inclined to say about the contents of perceptual illusions. Consequently, despite its ingenuity, the retreat to quasi-memory cannot stave off a fundamental objection to reductionism (McDowell 1993: §6). These objections to reductionism, only gestured at here, are powerful. The onus is upon the reductionist to reply to them.

See also: Mind, bundle theory of; Reductionism in the philosophy of mind

BRIAN GARRETT

References and further reading

Cassam, Q. (1992) 'Reductionism and First-Person Thinking', in D. Charles and K. Lennon (eds) *Reduction, Explanation, and Realism*, Oxford: Clarendon Press. (Another useful critique of reductionism.)

Cassam, Q. (1993) 'Parfit on Persons', *Proceedings of the Aristotelian Society* 1993 (1): 17-37.(Useful critique of reductionism.)

Evans, G. (1982) *The Varieties of Reference*, Oxford: Oxford University Press. (Groundwork for the discussion in Section 6 of McDowell. 6.6 and 7.5 are especially relevant.)

Garrett, B.J. (1990) 'Persons and Human Beings', *Logos* 11: 47-56.(More detailed discussion of the issues raised in §2 and §3 above.)

Garrett, B.J. (1991) 'Personal Identity and Reductionism', *Philosophy and Phenomenological Research* 51 (2): 361-73.(As with the 1990 article, a more detailed examination of the issues discussed in §2 and §3.)

Hume, D. (1739) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1978.(Classic statement of the 'bundle theory' of the self outlined above in §3.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. J. Yolton, London: Dent, 1961.(A subtle discussion of persons, thinking substances, and consciousness. Book II, chapter 27 contains the relevant passages, as cited in §1 above.)

McDowell, J.H. (1993) 'Reductionism and the First Person', in J. Dancy (ed.) *Reading Parfit*, Oxford: Blackwell. (A sophisticated attack on reductionism.)

Peter of Auvergne (d. 1304)

Peter of Auvergne, a thirteenth-century Parisian master, wrote extensively on logic, natural philosophy and theology. His thought progresses from modism in logic to an independent synthesis of Aristotelian philosophy along the lines begun by Thomas Aquinas, culminating in a theology reconciling the ideas of his teachers Henry of Ghent and Godfrey of Fontaines. His reputation has been based largely on his association with Aquinas, but recent investigations have shown the independence of his thought.

Peter of Auvergne was a secular master at the University of Paris in the thirteenth century. He was born at Crocq (Crocy) in south-central France, appointed rector of the university on 7 March 1275, awarded the bishopric of Clermont on 21 January 1302, and died 25 September 1304.

Peter's many works allow us to witness his intellectual development. He is credited with writing commentaries on all of Aristotle's works, but some of these have been lost and some surviving works are of doubtful attribution. His earliest commentaries, on the *logica vetus* (old logic), are typical of modistic logic, characterized by a confidence that grammatical analysis can reveal a link between reality, signification and understanding (see Language, medieval theories of; Logic, medieval). His early commentaries are influenced also by the recent assimilation of the *Posterior Analytics*, which motivated the attempt to establish an Aristotelian science based upon universal and necessary propositions. Peter was an early practitioner of the question-commentary in the Faculty of Arts at Paris, although he combines older traditions of textual analysis with his promotion of the modistic programme. The *Posterior Analytics* commentary attributed to him shows a considerable degree of complexity, as do the undoubtedly authentic *Sophismata*. Peter's logical works influenced Simon of Faversham, Durandus of Auvergne and Radulphus Brito.

As a young Master of Arts at Paris, Peter was appointed rector of the university to help settle the conflict between the followers of Siger of Brabant and those of Alberic of Rheims, which indicates his conciliatory personality as well as his considerable reputation among his contemporaries. His fame was sufficiently great to motivate the false attribution to him of works by Albert the Great, Thomas Aquinas, an anonymous imitator (the *Anonymus Matritensis*) and, following the condemnation at Oxford in 1270, works by Siger of Brabant.

However, Peter's greatest fame comes from his association with Thomas Aquinas. The extent of their personal contact is not known, but Peter was one of the earliest influenced by Aquinas' natural philosophy and theology and, most importantly, he completed Aquinas' unfinished commentary on Aristotle's *On the Heavens*. His work was also adapted to complete Aquinas on the *Meteora* and the *Politics*. Five other works of his have been printed among the works of Aquinas. Nonetheless, much modern scholarship has been devoted to showing that Peter frequently disagreed with Aquinas.

Many commentaries on natural philosophy belong to Peter's next stage of development and, if all are authentic, Peter mastered the works of Aristotle to an extent unparalleled since Albert the Great. These works, like medieval commentaries on the *Parva naturalia* in general, have not yet attracted much investigation. The commentary on the *Physics* once attributed to Siger of Brabant, now plausibly attributed to Peter by William Dunphy (1953), belongs to this period; it addresses incipiently many themes of Peter's later *Quodlibets* (see Natural philosophy, medieval).

Peter's commentary on the *Metaphysics* shows an author concerned not merely with understanding Aristotle's text, but also with formulating an Aristotelian metaphysics which accommodates theology as well as explaining the world. Because several Peters from Auvergne flourished at the same period, doubts have been raised about whether the author of the *Metaphysics* commentary is the same as the logician of the 1270s and the theologian of the *Quodlibets*, but scholars now generally accept the single attribution. It may be noted that both the commentaries on the *Metaphysics* and on Porphyry's *Isagōgē* attributed to Peter use the unusual expression 'determination of matter' (*signatio materiae*) in discussing individuation; and *Metaphysics* V q. 32 argues in exactly the same way as does the commentary on the *Categories* q. 50, that identity is a purely mind-dependent relation.

Peter's commentary on the *Ethics* is typical of its milieu; it often follows Aquinas, although in

question-commentary form. His writings on the *Politics* (a question-commentary and a literal commentary adapted to complete Aquinas' commentary) influenced later writers, particularly his views that a virtuous multitude has a right to rule and that laws should serve the many rather than the elite few.

Peter came late to theology, not becoming a master of theology until the end of the 1290s. His theology has been seen as part of a continuing evolution away from the ideas of Thomas Aquinas, because of the influence of his teachers Henry of Ghent and Godfrey of Fontaines; Peter occasionally refuses to take sides in conflicts between the ideas of these two masters. Peter holds that theology is a separate science; he accepts Godfrey's criticism of the view of that theology is a science subalternate to the science of happiness. As a theologian, Peter revises his theory of individuation, accepting Godfrey's view that form is the principle of individuation. Following the impetus of Bishop Tempier's condemnation of 1277, which criticized limitations of God, Peter emphasizes the distinction between God's absolute power and God's ordained power, and insists that God in his absolute power can do anything which does not involve a contradiction; but he holds that the creation of an actual infinity of things involves such a contradiction. Peter defends moderate Aristotelian positions against threats posed by the ideas of Averroes (see Ibn Rushd), such as divine ignorance of particular things or an eternal world, although Peter (like Aquinas) accepts the possibility of an eternal world with its concomitant infinity of souls (see Averroism). At times Peter subscribes to Neoplatonism, in particular the positing of an emanation from the First Cause and a hierarchy of being (see Platonism, medieval).

Only a few questions from Peter's commentary on the *Sentences* seem to have survived. Polemical comments in the *Sentences* commentary by John Quidort (see John of Paris) suggest that Peter held that immaterial beings (such as God and angels) produce effects directly by the intellect, rather than via the will, and that many powers of the soul result from divine illumination (Muller (ed.) 1961).

Peter of Auvergne's long career shows us a thirteenth-century Parisian master of more than ordinary productivity and influence. In outline, his thought progresses from modism in logic to an independent Aristotelian synthesis along the lines begun by Aquinas, culminating in a theology reconciling the ideas of his famous teachers. However, our complete assessment of his contribution is still clouded by uncertain attributions and unedited texts.

See also: Aquinas, T.; Aristotelianism, medieval; Logic, medieval

ROBERT ANDREWS

List of works

Peter of Auvergne (*c*.1270s) *Sophismata*.(Sophisma III is edited and discussed in S. Ebbesen (1988) 'Concrete Accidental Terms: Late 13th Century Debates about Problems Relating to Such Terms as "album"', in N. Kretzmann (ed.) *Meaning and Inference in Medieval Philosophy: Studies in Memory of Jan Pinborg*, Dordrecht: Reidel.)

Peter of Auvergne (*c*.1275) *Quaestiones super Praedicamentis (Questions on the Categories)*, ed. R. Andrews, 'Petrus de Alvernia: Quaestiones super Praedicamentis. An Edition', *Cahiers de l'institut du moyen-âge grec et latin* 55, 1987: 3-84.(Edition of Peter's questions on Aristotle's *Categories*.)

Peter of Auvergne (*c*.1275) *Quaestiones super librum Perihermenias (Questions on De interpretatione)*.(An edition of q. 12 is to be found in S. Ebbesen (1986) 'Termini accidentales concreti: Texts from the Late Thirteenth Century', *Cahiers de l'Institut du Moyen-âge grec et latin* 53: 37-150.)

Peter of Auvergne (1277-83) *Quaestiones super librum Ethicorum (Questions on Aristotle's Nicomachean Ethics)*.(An edition of what survives of this commentary is in A.J. Celano, 'Peter of Auvergne's Questions on Books I and II of the *Ethica Nicomachea*: A Study and Critical Edition', *Mediaeval Studies* 48, 1986: 1-110.)

Peter of Auvergne (*c*1280s) *Quaestiones in Metaphysicam (Questions on the Metaphysics)*.(An edition of ten *quaestiones* appears in A. Monahan, 'Peter of Auvergne's *Quaestiones in Metaphysicam*', in J.R. O'Donnell (ed.) *Nine Mediaeval Thinkers*, Studies and Texts 1, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1955. An edition of *Quaestiones* V q. 18 is to be found in S. Ebbesen, *'Termini accidentales concreti: Texts from the Late Thirteenth Century'*, *Cahiers de l'institut du moyen-âge grec et latin* 53, 1986: 37-150. An edition of *Quaestiones* XII qq. 6a-d is to be found in W. Dunphy, 'The 'Quinque viae' and Some Parisian Professors of Philosophy', in *St. Thomas Aquinas 1274-1974: Commemorative Studies*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1974, vol. II.)

Peter of Auvergne (1296-1300) *Quodlibeta*.(Published editions include Quodlibet I q. 21 (1296) and Quodlibet V qq. 9-10 (1300), G. Cannizzo, 'I 'Quodlibeta' di Pietro d'Auvergne', *Rivista di Filosofia neo-scolastica* 56,

1964-5: 486-500, 605-48, and 57: 67-89; Quodlibet II q. 5 (1297), E. Hocedez, 'Une question inédite de Pierre d'Auvergne sur l'individuation', *Revue néo-scolastique de philosophie* 36, 1934: 335-86.)

Peter of Auvergne (before 1304) *Quaestiones super librum Posteriorum (Questions on the Posterior Analytics)*, selections edited in S. Ebbesen and J. Pinborg, 'Studies in the Logical Writings Attributed to Boethius de Dacia', 1970, and J. Pinborg, 'A New MS of the Questions on the *Posteriora Analytica* Attributed to Petrus de Alvernia (Clm. 8005) with the Transcription of Some Questions Related to Problems of Meaning', *Cahiers de l'institut du moyen-âge grec et latin* 3, 1973: 1-54; 10: 48-62. (Edition of questions 23-6 and 31-3, possibly by Peter of Auvergne, of unestablished date.)

Peter of Auvergne (before 1304) *Quaestiones super librum Physicorum (Questions on the Physics)*.(An edition of Books I-IV of a work formerly attributed to Siger of Brabant but now believed to be Peter of Auvergne's Questions on the *Physics*, can be found in P. Delhaye (ed.) *Siger de Brabant 'Questions sur la Physique d'Aristote'*, Les Philosophes Belges XV, Louvain: Institut superieur de philosophie, 1941.)

References and further reading

Dunphy, W. (1953) 'The Similarities Between Certain Questions of Peter of Auvergne's *Commentary on the Metaphysics* and the Anonymous *Commentary on the Physics* Attributed to Siger of Brabant', *Mediaeval Studies* 15: 159-68.(Parallel passages suggest that the *Physics* commentary may be by Peter of Auvergne.)

Hocedez, E. (1930) 'La théologie de Pierre d'Auvergne' (The Theology of Peter of Auvergne), *Gregorianum* 11: 526-55. (Analysis of the theological material of the *Quodlibets*.)

Hocedez, E. (1933) 'La vie et les oeuvres de Pierre d'Auvergne' ('The Life and Works of Peter of Auvergne'), *Gregorianum* 14: 3-36. (Biography and list of works.)

Hocedez, E. (1935) 'La philosophie des Quodlibets de Pierre d'Auvergne' ('The Philosophy of the Quodlibets of Peter of Auvergne'), in A. Long *et al*. (eds) *Aus der Geisteswelt des Mittelalters. Studien und Texte Martin Grabmann zur Vollendung des 60. Lebensjahres von Freunden und Schülern gewidmet*, Beiträge zur Geschichte der Philosophie des Mittelalters Suppl. 3.2, Munich: Aschendorff. (Philosophical aspects of the *Quodlibets*.)

Lohr, C.H. (1972) 'Medieval Latin Aristotle Commentaries, Authors: Narcissus-Richardus', *Traditio* 28: 334-46. (Lists manuscripts of Aristotelian commentaries.)

Muller, J.-P. (ed.) (1961) *Jean de Paris (Quidort) O.P. 'Commentaire sur les 'Sentences'*, Rome: Herder. (Passages in Quidort's *Commentary on the Sentences* may give hints about the doctrines of Peter of Auvergne's lost *Commentary on the Sentences*.)

Roensch, F.J. (1964) *Early Thomistic School*, Dubuque, IA: The Priory Press, 92-8. (Summary of life and works.)

White, K. (1990) 'St. Thomas Aquinas and the Prologue to Peter of Auvergne's *Quaestiones super De sensu et sensato*', *Documenti e studi* I (2): 427-56.(Influence of Aquinas on Peter's Questions on Aristotle's *On the Senses*.)

Peter of Spain (*c.*1205-77)

For hundreds of years, a number of works in philosophical psychology, medicine and logic have been attributed to a single thirteenth-century author known as Peter of Spain. According to the latest research, however, there were actually two authors of that name. One, who later became Pope John XXI, wrote on medicine and on the soul. In his writings on the soul, this Peter argued for an independent material form that prepared the body for the soul, its substantial form, and he took an Augustinian view of God's role in illuminating the intellect in ordinary cognition. The second Peter of Spain was a Spanish Dominican who wrote the logical works. He is chiefly known for his Tractatus, an influential textbook of logic written before 1250, expounding both traditional topics and the theory of supposition, concerning reference in sentential contexts.

The first Peter was born in Lisbon around 1205, of noble parentage. He completed the liberal arts course at Paris in the 1220s and then studied medicine, probably in southern France, becoming a professor of medicine at Siena in 1245. Peter wrote his medical works, including the famous *Thesaurus pauperum (A Treasury for the Poor)* and his treatises on the soul in the 1240s and 1250s, as well as an exposition of the works of Pseudo-Dionysus. From 1250 Peter lived mostly in Lisbon, enjoying various preferments. In 1272 he became physician to Pope Gregory X, and in 1273 Archbishop of Brage and Cardinal of Frascati. He was elected pope in 1276. As Pope John XXI, he issued two Bulls in 1277 to Etienne Tempier, the Bishop of Paris, ordering the examination of teaching at the University of Paris. Tempier's condemnation of 219 heretical propositions in theology and natural philosophy emphasized the absolute power of God, contributing to the general fourteenth-century stance that the laws of nature are contingent on God's will and reflect no strict impossibilities (see Aristotelianism, medieval). In May 1277 the roof of the pope's private study collapsed and he died of his injuries.

In his treatises on the soul, Peter reveals himself a conservative Augustinian of the stamp of Robert Grosseteste. Pursuing the question why the soul, a unity in itself and the form of the body, should give rise to many organs, he points out that the soul has many powers, and needs different organs to realize these powers in matter. Thus there are sense organs for the sensitive power, muscles for the power of self-movement and so on. He identifies two agent intellects, one within the individual human soul, which abstracts forms as they occur in nature from sensible particulars, and the other God himself, who must illuminate the abstracted form in the soul to impart scientific knowledge. Scientific knowledge presupposes a cognition of the form prior to its realization in the material world, from which we can see why things that have the form must be as they are if they are to realize the form in matter. Thus the soul can understand itself only by understanding its powers as prior to their material implementation in the body, so that it can see what shape the body must take to be the body of such a form. Just as the eyes need light to see, so the intellect needs God's illumination to make natural objects intelligible to it. The soul's intellectual powers are not a product of natural processes, but of God himself, while other powers could be produced purely naturally (see Soul, nature and immortality of the).

The second Peter was most noted for his *Tractatus*, summarizing the logic of his time in twelve tracts, concerning not only the traditional Aristotelian topics but also the theory of supposition, a logical doctrine first developed in the twelfth century (see Language, medieval theories of; Logic, medieval). Peter claims that the signification of a term belongs to it considered in abstraction from the term's occurrence in a particular proposition. Thus, the term 'man' always signifies the universal *man*. However, an occurrence of this term may 'supposit' for different things in different sentences, for instance, for a specified or unspecified individual, or several individuals, as in 'a man is running' 'that man is running' and 'all men are running'. For though the 'natural supposition' of a term is always the same - namely, all past, present and future individuals that fall under the term's signification - its 'accidental supposition' varies with context. Peter assumed that a term, all by itself, has a natural capacity to refer rooted in its signification, and that its use in a sentence may restrict that natural supposition in various ways but in no way accounts for it. All this coordinates well with realism, which explains how universals could be conceived independently of reference to individuals, and takes reference to individuals to depend on a prior grasp of universals (see Realism and anti-realism; Universals). In his *Syncategoremata (On Syncategorematic Words)*, Peter insists that the copula has no function other than joining the predicate and subject terms, which each have their own signification and reference independently of their sentential context. Peter's views on these matters are at odds with those of William of Sherwood, whose characteristically English approach is more friendly to

nominalism (see Nominalism). His *Tractatus* and *Syncategoremata* were written in the 1230s in southern France, at Toulouse or Montpellier, from which the earliest medieval commentaries on these works stem, including those of Robert Anglicus and William Arnaldi.

See also: Augustinianism; Language, medieval theories of; Logic, medieval; Universals

JOHN LONGEWAY

List of works

Peter of Spain [Pope John XXI] (before 1258) *Expositio libri De anima, De morte et vida, et De causis longitudinis et brevitatis vitae (Explanation of [Aristotle's] On the Soul, On Death and Life and On the Length and Brevity of Life)*, ed. M. Alonso, *Obras filosóficas*, vol. 3, Madrid: Instituto de Filosofico 'Luis Vives', 1952.(Brief commentaries on Aristotle's works.)

Peter of Spain [Pope John XXI] (before 1258) *In De anima (Commentary on [Aristotle's] On the Soul)*, ed. M. Alonso, *Obras filosóficas*, vol. 2, Madrid: Instituto de Filosofico 'Luis Vives', 1944.(A more extended commentary on Aristotle's *On the Soul*.)

Peter of Spain [Pope John XXI] (before 1258) *Scientia libri De anima (The Science of the Book On the Soul)*, ed. M. Alonso, *Obras filosóficas*, vol. 1, Madrid: Instituto de Filosofico 'Luis Vives', 1941.(An exposition of Peter's own views on the soul, closely informed by his study of medicine.)

Peter of Spain [Pope John XXI] (before 1258) *Expositio super libros de Beato Dionisio (Exposition of the Books of the Blessed Dionysius the Areopagite)*, ed. M. Alonso, *Exposiçao sobre os livros do Beato Dionisio Areopagita*, Lisbon: Instituto de alta cultura, 1957.(A theological work, expositing some Neoplatonic mystical treatises bearing on our knowledge of God which were, at the time, semi-canonical because of their attribution to Dionysius the Areopagite, who was supposed to have been converted by St Paul.)

Peter of Spain [Pope John XXI] (1272) *Thesaurus pauperum (A Treasury for the Poor)*, Antwerp: T. Martin, 1947.(A brief medical handbook for poor students who could not afford the more massive volumes from which its information is extracted.)

Peter of Spain [writer on logic] (*c.*1230) *Tractatus* (Treatises), ed. L.M. De Rijk, Assen: Van Gorcum, 1972; trans. J.P. Mullally, *The Summulae Logicales of Peter of Spain*, Notre Dame, IN: Notre Dame University Press, 1945; trans. F.P. Dineen, *Language in Dispute*, Philadelphia, PA: Benjamins, 1990; selections translated in N. Kretzmann and E. Stump, *The Cambridge Translations of Medieval Philosophical Texts: Volume One: Logic and the Philosophy of Language*, Cambridge: Cambridge University Press, 1988.(De Rijk's introduction is important. The translations in Kretzmann and Stump are the best.)

Peter of Spain [writer on logic] (*c.*1230) *Syncategoremata* (On Syncategorematic Words), ed. L.M. De Rijk, trans. J. Spruyt, Leiden: Brill, 1992; trans. J.P. Mullaly, *Peter of Spain: Tractatus Syncategorematum and Selected Anonymous Treatises*, Milwaukee, WI: Marquette University Press, 1964. (Concerns words which do not themselves have signification, but alter the meanings of terms with signification in the sentences in which they occur.)

Peter of Spain [writer on logic] (*c.*1230) *Tractatus exponibilium (Treatise on Exponible Propositions)*, ed. and trans. J. Spruyt, *Peter of Spain on Composition and Negation*, Nijmegen: Ingenium, 1990.(Part of the *Syncategoremata*. Concerns sentences with a misleading grammatical form, usually involving a concealed negation.)

References and further reading

De Libera, A. (1982) 'The Oxford and Paris traditions in Logic', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(Places Peter's logical work in the context of his time.)

d'Ors, A. (1997) 'Petrus Hispanus O.P., Auctor Summularum', *Vivarium* 25: 21-71.(Presents the evidence that there were in fact two writers known as Peter of Spain.)

Longeway, J. (1992) 'Peter of Spain', in J. Hackett (ed.) *Dictionary of Literary Biography*, vol. 115, *Medieval Philosophers*, Detroit: Bruccoli Clark Layman. (Contains a full bibliography.)

Petrarca, Francesco (1304-74)

With Dante and Boccaccio, Petrarca (known as Petrarch) made the fourteenth century the most memorable in Italian literature. He was also the first great humanist of the Italian Renaissance. He brilliantly and self-consciously exemplified humanism's classical, rhetorical, literary, and historical interests; with him the movement came of age. He was a proponent not only of classical Rome ('What else is history,' he once asked, 'than the praise of Rome?'), but also of contemporary Rome, constantly calling for the popes at Avignon to return to their proper See and urging restoration of Rome as the seat of the Empire, even if that meant supporting the visionary Roman revolutionary Cola di Rienzo. He came to see himself also as a moral philosopher. His ethical interests were closely tied to his cultural interests and personal situation as a lay moralist (though technically he was a cleric). His outlook and method differed from that of contemporary Aristotelians, whom he attacked on a broad cultural front, sounding many of the themes that would become common in subsequent conflicts between humanism and scholasticism in the Renaissance.

1 Life and poetry

Petrarca, known in the English world as Petrarch, was born in Arezzo. The son of a Florentine exile, he left Tuscany with his family in 1312 to become part of the sizeable Italian colony around the papal court of Avignon in Provence, where his father found employment as a notary (see Humanism, Renaissance §1). In 1316, young Francesco was sent to study law at the University of Montpellier and then, in 1320, to the University of Bologna. Petrarca did not want to be a lawyer, and when his father died in 1326 he returned home to live off his inheritance. Around 1330, as his inheritance ran out, Petrarca decided upon a clerical career. He only took minor orders, but the ecclesiastical benefices he accumulated allowed him to live a comfortable life. In 1337 he had a son, Giovanni, and in 1343 a daughter, Francesca, who proved to be a great comfort to him in his old age.

About 1337, after returning from his first visit to Rome, Petrarca had finally had his fill of Avignon and purchased a country home in nearby Vaucluse. He enjoyed his rural retreat immensely, and it was there that he began many of his most important projects. Nonetheless, in 1341, he left Vaucluse for a wider stage: Petrarca had arranged a ceremony for himself in Rome, at which he would be crowned poet laureate. (He believed it to be the first such event since antiquity, but in fact Albertino Mussato had been crowned poet laureate at Padua in 1315.) Wanting the sanction of major political authority, he arranged to have himself 'examined' by King Robert of Naples before proceeding on to the elaborate coronation ceremony in Rome. He also made sure that the Roman authorities conferred upon him all the rights and privileges of a university master in the two disciplines of poetry and history. He then went to Parma to enjoy the patronage of its lord, Azzo da Correggio. He returned to Vaucluse in the spring of 1342.

For the next decade Petrarca frequently returned to Italy for lengthy stays, and in the summer of 1353 finally settled there. To the dismay of his Florentine friends, he accepted the hospitality of the Visconti tyrants of Milan. He continued to travel, once even to Prague on a Visconti embassy, but he maintained his residence in Milan until 1362, when he decided that he preferred to live in Venice. The Republic provided him with the Palazzo Molin on the Riva degli Schiavoni in return for the bequest of his books, which the Venetians (vainly) hoped to make the core of a public library. In the agreement, doubtless in accordance with Petrarca's desires, the Venetian Senate called him a 'moral philosopher'. In 1368, however, Petrarca changed residence yet again, this time moving to Padua and the protection of its tyrant Francesco da Carrara. He died in his country home at Arquà near Padua.

Petrarca had earned the laurel crown in Rome as a Latin poet. Early in his career, he wrote metrical letters (eventually published as the *Epistolae Metricae*); he also wrote Latin eclogues (collected as the *Bucolicum Carmen*) and composed *Africa*, the first Latin epic written since antiquity. Over his career Petrarca wrote more poetry in Latin than in Italian (he referred to his Italian poems as 'mere trifles' in later life), but in the high Renaissance Petrarca's Latin style no longer found favour: the Europe-wide movement of 'Petrarchism' exclusively involved the imitation of his vernacular poetry. Most of Petrarca's surviving Italian poems were the work of his maturity, and Petrarca was refining his major collection of Italian poems (the *Canzoniere* or *Rime Sparse*) until his death. Today his Italian poems are the only part of his writings widely admired in the original.

Although he wrote sonnets on other themes (including the grand patriotic hymn 'Italia mia'), the vast majority of

Petrarca's Italian lyrics concern his love of Laura. He first saw her in Avignon on 6 April 1327, and although she rejected him, Laura remained the abiding love of his life. Inspired by her, Petrarca composed (in the vernacular) the collection of *Trionfi (Triumphs)*, which began with the *Triumphus Cupidinus (Triumph of Love)* in the 1330s. The *Triumphus Pudicitiae (Triumph of Chastity)* (written in the 1340s) celebrated Laura's honour and purity, which Petrarca saw as elevating him as well. After Laura died, he wrote the *Triumphus Mortis (Triumph of Death)* and subsequently the *Triumphus Famae (Triumph of Fame)*, *Triumphus Temporis (Triumph of Time)* and *Triumphus Eternitatis (Triumph of Eternity)*, finishing the collection with the hope of seeing Laura forever in heaven.

2 Ethical writings

At his coronation as poet laureate, Petrarca had made sure that he was also recognized as a historian. He had in fact begun in 1337 a series of biographies of classical, predominately Roman, heroes, called *De Viris Illustribus (On Illustrious Men)*. But by the time of his move to Venice in 1362, he wished also to be considered a moral philosopher. One of his first efforts as a moral philosopher was the unwieldy *Rerum Memorandarum Libri (Books on Matters that Ought to be Remembered)*, begun about 1343 and never completed. Scholars have deduced that it was supposed to be a work on the four cardinal virtues (prudence, justice, courage and temperance) and perhaps the corresponding vices. In fact it is a compilation of examples and anecdotes illustrating the different aspects of prudence.

Far more successful and interesting is the dialogue *Secretum (The Secret)*, first written in the 1340s, when Petrarca was undergoing a mid-life crisis, and subsequently revised. Petrarca has St Augustine examine why he, Petrarca, was suffering from depression (called by the name of one of the deadly sins, *accidia*). With his youth behind him and death waiting just over the horizon, Petrarca still could not shake free of earthly desires, most specifically, of his love of Laura and his desire for fame. The Augustine of the dialogue is quite unhistorical. He is a voluntarist who blames Petrarca's problems on insufficient will power (though towards the end he tells Petrarca to call on God to make good the failings of his will). The dialogue ends with Petrarca understanding better his predicament, but not really having overcome it. Petrarca was in fact a great admirer of Augustine as the master psychologist of the *Confessiones (Confessions)*.

No less personal was the treatise *De Vita Solitaria (On the Solitary Life)*, begun at Vaucluse in 1346. This work has strong religious overtones, but was in fact a lengthy defence of Petrarca's very secular and comfortable life away from the madding crowd and free of commitments that might limit his freedom to write and study. The next year, after visiting his brother Gherardo, a monk at the Carthusian monastery of Montrieux, he more legitimately made many of the same arguments in a paean to monastic life which he entitled *De Otio Religioso (On Religious Leisure)*.

His most 'professional' and largest work as a moral philosopher was the *De Remediis Utriusque Fortunae (On the Remedies of Good and Bad Fortune)*. Inspired by the Stoic ideal of the wise man who does not let the vagaries of fortune affect his inner peace, Petrarca started on his *Remedies* about 1354 and finished in 1366. Though cast in the form of a dialogue, *Remedies* was in essence a recipe book for a life of moderation. If you enjoyed anything, Petrarca told you why you should not be too elated by it. For instance, he explained that the enjoyment of wrestling was senseless for the spectator and insane for the participant. He also offered advice to encourage people who were facing misfortunes: for example, he argued that those who had been blinded were at least spared the many painful sights that life visits upon us. In all, he detailed 122 kinds of good fortune and 132 of bad fortune and their remedies. If some of his remedies strike us as ineffective, it is at least clear that he was striving to be practical and of genuine use to ordinary people.

In his ethical pronouncements Petrarca tended to favour Stoic formulas, but operationally, in his own life, he lived as an Aristotelian, desiring an adequate supply of worldly goods, friends, leisure, moderate pleasures, and the support of the powerful so that he could satisfy to the fullest his creative and intellectual impulses. Nonetheless, he was generally unsympathetic to Aristotle, while usually speaking well of Plato. On the authority of Augustine, he believed that Plato's doctrines harmonized well with Christianity. He especially attacked as opponents of Christianity Aristotle's Arab commentator Averroes (see Ibn Rushd) and contemporary Aristotelians who agreed with Averroes.

These tendencies stand out in one of his last works, the invective *De Sui Ipsius et Multorum Ignorantia (On His Own Ignorance and That of Many Others)*, started in 1367 and completed in 1371. Four Venetians favouring the Aristotelian scientific tradition had pronounced Petrarca a good man but an ignorant one, after discovering at a dinner party in 1366 that he could not answer certain questions in zoology. In response, Petrarca said he would not take account of the classical eloquence for which he was famed, but which his critics disdained; he feigned Socratic ignorance for his own self and attacked as inane the vaunted scientific knowledge of his critics, avowed the superiority of Christian faith over any philosophical wisdom, and while asserting Plato's superiority over Aristotle, confessed that his philosophical guides were the eloquent Latin authors Cicero and Seneca.

3 Humanist culture

As the intellectual weakness of this invective reveals, Petrarca's stance was more cultural than philosophical. In contrast with the legal, scientific and philosophical bias of contemporary learned culture, he proposed a learning dominated by rhetorical, literary, historical and classicist interests. His largest body of writing is his letters, which, inspired by his discovery of Cicero's *Familiar Letters* at Verona in 1345, he gathered, revised and published in several different collections. The first collection was his *Epistolae Familiares (Familiar Letters)*; then came his *Rerum senilium libri (Books of Matters of Old Age)*; he brought together some polemical letters in a collection called *Liber sine nomine (Book Without Name)*; and finally there are various letters of his that he never incorporated into any collection. Addressed to prelates, secular rulers, friends and enemies, Petrarca's letters spread the new culture of humanism as effectively as any of his writings. Since he did not merely study his favourite authors, but conversed with them as well, he even addressed some letters to Cicero and other worthies of antiquity. He also cultivated the other most preferred humanist rhetorical genre, the oration, several times making speeches on behalf of the Visconti (see Humanism, Renaissance §2).

In sum, apart from his vernacular poetry, Petrarca left to posterity not so much a philosophy as a classicizing language, a set of literary genres, a constellation of literary and historical interests, and a body of newly discovered classical texts which gave emerging Renaissance humanism self-awareness and a cultural agenda. He also gave Renaissance humanism its first hero.

See also: Alighieri, Dante; Humanism, Renaissance §§2, 4

JOHN MONFASANI

List of works

Petrarca, F. [Petrarch] (1304-74) *Opere latine (Works in Latin)*, ed. A. Bufano, Turin: Unione tipografico-editrice torinese, 1975, 2 vols.(Excellent anthology; includes original text with facing Italian translation for the Latin.)

Petrarca, F. [Petrarch] (1304-74) *Canzoniere* or *Rime Sparse*, trans. R.M. Durling, *Petrarch's Lyric Poems: The Rime Sparse and Other Lyrics*, Cambridge, MA: Harvard University Press, 1976.(English translation of Petrarca's vernacular poetry.)

Petrarca, F. [Petrarch] (begun 1330s) *Trionfi*, in F. Neri *et al.* (eds) *Rime, Trionfi e poesie latine*, Milan and Naples: Ricciardi, 1948; trans. E.H. Wilkins, *The Triumphs of Petrarch*, Chicago, IL: University of Chicago Press, 1962.(One of Petrarca's major collections of vernacular poetry. The Neri anthology includes the original text; Wilkins is an English translation.)

Petrarca, F. [Petrarch] (begun 1337) *De Viris Illustribus (On Illustrious Men)*, ed. G. Martellotti, Florence: Sansoni, 1964.(Modern critical edition of Petrarca's series of biographies of classical, predominately Roman, heroes.)

Petrarca, F. [Petrarch] (begun 1340s) *Africa*, trans. T.G. Bergin and A.S. Wilson, New Haven, CT, and London: Yale University Press, 1977.(English translation of the first Latin epic since antiquity. The poem is now more valued for historical than literary reasons; Petrarca did not publish it in his lifetime.)

Petrarca, F. [Petrarch] (begun 1340s) *Petrarch's Bucolicum Carmen*, trans. T.G. Bergin, New Haven, CT, and London: Yale University Press, 1974.(English translation of Petrarca's Latin eclogues.)

Petrarca, F. [Petrarch] (begun 1340s) *Secretum*, in D. Martellotti *et al.* (eds) *Prose*, Milan and Naples: Ricciardi, 1955; trans. W.H. Draper, *Petrarch's Secret*, London: Chatto & Windus, 1901.(Dialogue with Augustine, in which the saint takes a voluntarist approach to solving Petrarca's mid-life anxieties; first written in the 1340s but subsequently revised. Martellotti gives the original text with facing Italian translation of the Latin; Draper is a translation into English.)

Petrarca, F. [Petrarch] (begun *c.*1343) *Rerum Memorandarum Libri (Books on Matters that Ought to be Remembered)*, ed. G. Billanovich, Florence: Sansoni, 1943.(Modern critical edition of one of Petrarca's first works of moral philosophy; it was never finished. A compilation of examples and anecdotes illustrating different aspects of prudence.)

Petrarca, F. [Petrarch] (after 1345) *Epistolae Familiares*, ed. V. Rossi and U. Bosco, Florence: Sansoni, 1933-42, 4 vols; trans. A.S. Bernardo, *Letters on Familiar Matters*, vol. 1, Albany, NY: State University of New York Press, vol. 2, Baltimore, MD: Johns Hopkins University Press, 1975-82.(Petrarca's first collection of letters. Rossi and Bosco is a modern critical edition; Bernardo a translation into English.)

Petrarca, F. [Petrarch] (after 1345) *Rerum senilium libri (Books of Matters of Old Age)*, in *Opera Omnia*, Ridgewood, NJ: Gregg Press, 1965, vol. 2, 812-1070; trans. A.S. Bernardo *et al.*, *Letters of Old Age*, Baltimore, MD: Johns Hopkins University Press, 1975-82, 2 vols.(Petrarca's second collection of letters. The Gregg Press *Opera* is a reprint of the three-volume Basle edition of 1554; Bernardo provides an English translation.)

Petrarca, F. [Petrarch] (after 1345) *Liber sine nomine (Book Without Name)*, ed. P. Piur, *Petrarcas 'Buch ohne Name' und die päpstliche Kurie, ein Beitrag zur Geistesgeschichte der Frührenaissance*, Halle: Niemayer, 1925.(Collection of Petrarca's polemical letters.)

Petrarca, F. [Petrarch] (begun 1346) *De Vita Solitaria (On the Solitary Life)*, in D. Martellotti *et al.* (eds) *Prose*, Milan and Naples: Ricciardi, 1955.(A lengthy defence of Petrarca's comfortable life of study and writing at Vaucluse. Original text with facing Italian translation of the Latin.)

Petrarca, F. [Petrarch] (*c.*1354-66) *De Remediis Utriusque Fortunae*, in *Opera Omnia*, Ridgewood, NJ: Gregg Press, 1965, vol. 1, 1-254; trans. C.H. Rawski, *Petrarch's Remedies for Fortune Fair and Foul*, Indianopolis, IN: Indiana University Press, 1991, 5 vols.(Inspired by Stoic ideals, Petrarca's largest work as a moral philosopher; attempts to offer practical help for ordinary people. The Gregg Press *Opera* is a reprint of the three-volume Basle edition of 1554; the translation includes bibliographical references, indexes and a commentary by Rawski.)

Petrarca, F. [Petrarch] (1371) *De Sui Ipsius et Multorum Ignorantia*, trans. H. Nachod, *On His Own Ignorance and That of Many Others*, in E. Cassirer *et al.* (eds) *The Renaissance Philosophy of Man*, Chicago, IL: University of Chicago, 1948.(Begun in 1367 and completed in 1371. Polemical attack on medieval definitions of scientific knowledge and defence of the superiority of the Christian faith over philosophical wisdom; Petrarca asserts Plato's superiority over Aristotle, and acknowledges his philosophical debt to Cicero and Seneca.)

References and further reading

Billanovich, G. (1951) 'Petrarch and the Textual Tradition of Livy', *Journal of the Warburg and Courtauld Institutes* 14: 137-208.(Important for Petrarca's classical scholarship.)

Foster, K. (1984) *Petrarch: Poet and Humanist*, Edinburgh: Edinburgh University Press.(Good on Petrarca's poetry.)

Kristeller, P.O. (1964) *Eight Philosophers of the Italian Renaissance*, Stanford, CA: Stanford University Press, ch. 1.(The first chapter treats Petrarca's philosophical ideas.)

Kristeller, P.O. (1983) *'Petrarcas Stellung in der Geschichte der Gelehrsamkeit' (Petrarca's Place in the History of Learning)*, in K.W. Hempfer and E. Straub (eds) *Italien in Humanismus und Renaissance: Festschrift für Erich Loos um 70 Geburtstag*, Wiesbaden: Steiner.(A synthesis of much useful material.)

Mann, N. (1984) *Petrarch*, Oxford: Oxford University Press.(An elegant appreciation.)

Nolhac, P. de (1907) *Pétrarque et l'humanisme*, Paris: Librairie Champion, 2nd edn, 2 vols.(Old but important.)

Rico, F. (1974) *Vida u obra de Petrarca: I. Lectura del Secretum (Life and Work of Petrarca. Vol. 1: Reading the Secretum)*, Padua: Antenore. (A fundamental study.)

Wilkins, E.H. (1961) *Life of Petrarch*, Chicago, IL: University of Chicago Press.(The best biography in English by a master of Petrarchan scholarship.)

Petrażycki, Leon (1867-1931)

Leon Petrażycki was a strikingly original philosopher of law whose theoretical system ranged over moral and legal philosophy, theory of science, psychology and sociology. The aim of his philosophy of law is 'Legal Policy', that is, social engineering through law directed at the 'social ideal of love'.

A Polish nobleman, born in the county of Vitebsk (in present-day Belarus), Petrażycki was educated at the Russian high school in Vitebsk and studied medicine and law at the Russian Imperial University in Kiev. From 1890 to 1906, he held a scholarship in Berlin, and from 1898 to 1917 was Professor in History of Legal Philosophy at St Petersburg University. From 1919 to 1931 he held a chair in sociology at the University of Warsaw. He published in German, Russian and Polish.

Petrażycki's ontology is materialist in a broad sense. Values do not exist. Material objects and psychical phenomena do exist objectively, yet causality is an axiom of theoretical reason. His epistemology is moderately realist: science employs empirical observation, induction and deduction, to construct an analogue of reality. Practical reason is founded on the social postulate of love. His methodology demands that theories be neither 'jumping' (too wide: saying about a whole class what is true only for part of it) nor 'lame' (too narrow: saying of part alone what is true of the whole).

His psychology is based on a new classification of psychical phenomena. Beyond feeling, knowing and willing, there exist 'impulsions'. They are passive-active: for example, feeling hunger, one strives for food. Moral impulsions consist of the 'feeling' of one's duty together with an active readiness to do it. Legal impulsions are the only phenomena around which to build an adequate legal theory. So sound legal theory is psychological. Studying law is studying legal impulsions. These are imperatively attributive, that is, concerning one's own duties intertwined with another's rights or vice versa. These are studied by introspection in one's own case, only analogically in others'.

Law is explained in evolutionary terms. A person whose rights are not respected retaliates. The legal psyche is thus aggressive (while the moral psyche, lacking the element of rights, is peaceful). Legal rules and legal concepts adapt themselves unconsciously to achieving unity among legal impulsions of different people. Positive law, and also the law of illegal organizations, laws of games, and so on, are to be studied subject to the same criteria of theoretical adequacy. The impulsions of positive law are attached to intersubjectively cognizable 'normative facts', in the case of 'official law' such as legislation or custom. Legal rules are projections of shared legal impulsions. Doctrinal law makes law more uniform and better adapted to its pacific social function.

Petrażycki accounted for the evolution of law comprehensively since Roman times, and anticipated later theories in many fields, for example, in relation to post-decisional dissonance, the cognitive role of language, and the hypothetico-deductive method in science. Yet legal institutions cannot be fully described in individualistic-psychological terms, and lame theories, not being false, are in many cases better than none at all.

See also: Law, philosophy of; Wróblewski, J.

ALEKSANDER PECZENIK

List of works

Petrażycki, L. (1955) *Law and Morality*, trans. H.W. Babb, 20th Century Legal Philosophy Series, vol. 7, Cambridge, MA: Harvard University Press.(This is the only version available in English of Petrażycki's work; it gives an account of all his main ideas.)

References and further reading

Olivecrona, K. (1948) 'Is a Sociological Explanation of Law Possible?', *Theoria* 14: 167-207.(Petrażycki's theory is taken as a significant attempt to elaborate a non-voluntaristic concept of law in socio-psychological terms.)

Timasheff, N.S. (1947) 'Petrażycki's Philosophy of Law', in P. Sayre (ed.) *Interpretations of Modern Legal Philosophies: Essays in Honour of Roscoe Pound*, Cambridge, MA: Harvard University Press, 748-69.(A spirited attempt to reconcile Petrażycki's work to the philosophical concerns of the post-war world.)

Phenomenalism

On its most common interpretation, phenomenalism maintains that statements asserting the existence of physical objects are equivalent in meaning to statements describing sensations. More specifically, the phenomenalist claims that to say that a physical object exists is to say that someone would have certain sequences of sensations were they to have certain others. For example, to say that there is something round and red behind me might be to say, in part, that if I were to have the visual, tactile and kinaesthetic (movement) sensations of turning my head I would seem to see something round and red. If I were to have the sensations of seeming to reach out and touch that thing, those sensations would be followed by the familiar tactile sensations associated with touching something round.

Rather than talk about the meanings of statements, phenomenalists might hold that the fact that something red and round exists just is the fact that a subject would have certain sequences of sensations following certain others. The phenomenalist's primary motivation is a desire to avoid scepticism with respect to the physical world. Because many philosophers tied the meaningfulness of statements to their being potentially verifiable, some phenomenalists further argued that it is only by reducing claims about the physical world to claims about possible sensations that we can preserve the very intelligibility of talk about the physical world.

There are very few contemporary philosophers who embrace phenomenalism. Many reject the foundationalist epistemological framework which makes it so difficult to avoid scepticism without phenomenalism. But the historical rejection of the view had more to do with the difficulty of carrying out the promised programme of translation.

1 Why phenomenalism?

Philosophers who embraced phenomenalism typically did so in the attempt to defeat scepticism with respect to the physical world. On one historically prominent and intuitively plausible model of our perceptual relations to the physical world, physical objects are different from, but causally responsible for, sensations which allow us to know those objects. Our sensations are the end product of a long causal chain that (we believe) involves our sense organs and ultimately our brain states. In veridical experience, our sensations in some sense represent the objects we take to be their cause. In illusion and hallucinatory experiences, our sensations fail to represent their external causes (see Perception).

This view of perception and the closely related view of the nature of our knowledge of physical objects are grist for the sceptic's mill. It seems easy enough to establish that the subjective, fleeting sensations that cause us to believe in the existence of physical objects never guarantee the truth of what we believe. The representative realist takes sensations to be the *effects* of physical objects, effects that could, in principle, be produced some other way (by artificially stimulating the relevant region of the brain, for example). But if our only access to the physical world is through sensations, then how could we ever establish sensations as a reliable indicator of the existence of physical objects? The most common way we have of establishing one thing as a sign of something else is to establish a correlation between the two in past experience. But we could never correlate a sensation with a physical object different from a sensation if all we have to go on ultimately are sensations. We can establish connections between sensations, but never between sensations and their alleged causes.

The phenomenalist seeks to rescue common sense from the sceptical implications of representative realism. The root of scepticism lies in a mistaken conception of the physical world, the phenomenalist argues. We must avoid the sceptic's attempt to drive a wedge between sensations and objects, by finding a conceptual link between the existence of physical objects and the occurrence of sensations. But we must do so without making the existence of physical objects depend on the occurrence of any *actual* sensations. It seems obvious that the very concept of a physical object is the concept of an object whose existence is independent of minds and their actual experiences. The table I see now can continue to exist even if I leave the room and there is no-one else around to perceive it. If the concept of a physical object is the concept of something that can exist even when it is unsensed, we cannot identify physical objects with sensations. But, the phenomenalist argues, that does not prevent us from analysing the meaning of physical-object statements into more complex statements describing sensations. We should view statements asserting the existence of physical objects as equivalent in meaning to statements describing 'possible' sensations. More precisely we should take physical object statements to be making conditional

assertions about the sensations someone *would* have *were* that person to have certain other sensations. Subjunctive (counterfactual) conditionals of this sort are made true by connections between sensations, and there seems to be no insurmountable difficulty in establishing correlations between past sensations of the sort necessary to establish causal connections between patterns of sensations (assuming, of course, that we will be able to deal with other sceptical problems concerning knowledge of the past and induction) (see Counterfactual conditionals).

Phenomenalism is compatible with a number of different views about the nature both of sensations and of the selves that have sensations. Many of the historically important phenomenalists were sense-datum theorists who analysed a visual sensation, for example, in terms of a subject's being directly aware of a two-dimensional object existing outside of physical space and exemplifying phenomenal colour and shape (see Sense-data). But there is nothing to stop a phenomenalist from embracing a so-called adverbial or appearing theory which identifies sensations with the mind's exemplifying certain non-relational properties (see Mental states, adverbial theory of). Just as phenomenalists may disagree among themselves over the nature of sensations, so too they may disagree over the analysis of that which has sensations. A phenomenalist can take the subject of sensations to be a mind of the sort posited by Berkeley, or can attempt to reduce subjects themselves to 'bundles' of actual and possible sensations (see Mind, bundle theory of).

Because early statements of phenomenalism were often not clear, it is difficult to make any definite claims about the historical origins of the view. There are certainly passages in which the idealist, Berkeley, seemed to suggest the conditional analysis of physical objects in trying to explain how his view that only minds and ideas (which include sensations) exist is compatible with common sense. In one passage he seems to argue that the story of a creation that takes place before the existence of finite minds should be understood in terms of truths about what finite minds would have experienced had they existed ([1713] 1954: 100). A more promising candidate for being the first phenomenalist might be John Stuart Mill (1865). Mill identified physical objects with the 'permanent possibility of sensations' and explained this in terms of the sensations one would have under certain conditions (see Mill, J.S. §6).

2 Criticisms

For their enterprise to be successful, it is necessary that phenomenalists actually produce plausible 'translations' of statements about the physical world into statements employing purely 'phenomenal' language - that is, statements referring to nothing other than minds and their sensations. Any reference to physical objects will 'contaminate' the subjunctive conditionals employed by the phenomenalist and will defeat the whole purpose of the analysis which is to reduce fully talk about the physical world to allegedly less problematic talk about sensations. Historically, the most influential objections to phenomenalism focused on the alleged impossibility of plausibly completing a phenomenalistic analysis. The clearest of these objections was advanced by Roderick Chisholm (1948) and has become known as 'the argument from perceptual relativity'. The phenomenalist, C.I. Lewis, argued (1946) that part of what it means to claim that there is a doorknob in front of me and to the left is that if I were to seem to see the doorknob and seem to be initiating a certain grasping motion, then in all probability the feeling of contacting a doorknob would follow. Chisholm points out that if the conditional really were part of what the existence of a doorknob involved, then it would be impossible for the doorknob to be there without the conditional being true. But it is easy to imagine a situation in which, though the doorknob is there, I would not be able to have any tactile sensations of it. My hand might be anaesthetized; then, even though the doorknob is there, I would not feel anything upon seeming to reach out to touch it. The proposition that there is a doorknob there does not entail anything about what sensations I would have were I to have others, because the sensations I would have always depend on facts about the internal and external *physical* conditions of perception (facts about the lighting conditions, sense organs, and so on). We cannot, however, revise the conditionals we employ in a phenomenalistic analysis to make reference to standard or normal physical conditions of perception (the argument goes) without referring to the physical world and thus defeating the goal of providing an analysis that refers to nothing other than connections between sensations.

Another closely related objection to phenomenalism argues that contingent subjunctive conditionals of the sort employed by phenomenalists in their analyses presuppose *lawful* regularities between sensations. But again, because sensations are always causally dependent on physical conditions, there simply are no lawful regularities involving sensations alone.

The above objections are often presented by philosophers sympathetic to the general philosophical framework presupposed by phenomenalists. Certainly many contemporary philosophers reject many of the assumptions that motivated phenomenalists. Thus, for example, many contemporary epistemologists reject the kind of foundationalism that led phenomenalists to worry that a resolution of sceptical problems was impossible if we separate physical objects from sensations. A good many of these epistemologists embrace some version of externalism which allows one to know facts about the physical world simply by having one's beliefs caused in the appropriate way (see Internalism and externalism in epistemology). Still other contemporary philosophers reject the conception of philosophical analysis as an attempt to specify necessary and sufficient conditions for the truth of propositions. Again, many of these will accept an externalist or causal theory of *meaning* which makes the meaning of expressions largely an empirical question inaccessible to a priori analysis. Such views of meaning will either reject, or significantly reinterpret, the verifiability criterion of meaning that drove so many phenomenalists (see Meaning and verification).

3 Responses

It is impossible in this context to evaluate challenges to the phenomenalist's basic philosophical presuppositions. But is there any response that the phenomenalist might make to the argument from perceptual relativity? The key to such a response would seem to involve 'protecting' the antecedents of the subjunctive conditionals employed by the phenomenalist. That something round is before me now entails only that if I were to seem to reach out and grasp the object *under certain conditions*, I would seem to feel something round. The phenomenalist's critics claim that one cannot specify the relevant conditions without introducing reference to the physical world, thereby defeating the whole purpose of the analysis. But is that true? A phenomenalist might argue that one can introduce into one's phenomenalistic analysis a clause about normal or standard conditions whose purpose is to *denote* those conditions (*whatever* they are) that normally (defined statistically) accompany certain sequences of sensations. The normal conditions clause might include reference to other facts about what sensations would follow others, facts about 'things-in-themselves' (construed as entities whose intrinsic nature will always remain a complete mystery), facts about the intentions of a God and so on. While the introduction of such a clause would move one away from 'pure' versions of phenomenalism by allowing into one's analysis expressions that might denote things other than minds and sensations, it might be argued that the modification is epistemically harmless, for we are always justified in believing, other things being equal, that conditions are normal, that is to say that conditions are as they usually are.

Once one modifies the analysis this much, however, it is not altogether clear why one should not move all the way to a version of what might be called a causal theory of objects, one that resembles classical phenomenalism much more than representative realism. On this 'phenomenalistic' causal theory, to assert the existence of a physical object is simply to assert the existence of a thing (whatever its intrinsic character might be) that has the power to produce certain sensations and that would produce certain patterns of sensations following an initial sequence of sensations *when conditions are normal*. This 'causal' theory looks a great deal like phenomenalism - indeed it may be another natural way of construing Mill's reference to physical objects as the permanent possibilities of sensations. Unlike traditional phenomenalism, however, this theory might have an easier time analysing physical-object statements which assert only the existence of something (somewhere, sometime) having certain physical characteristics. Such statements are a real problem for traditional phenomenalism, for when the physical-object statement fails to describe a physical object in relation to potential perceivers of those objects, it is not clear whose sensations we can plausibly refer to in our subjunctive conditionals. Notice, however, that the kind of causal theory sketched above faces identical problems to classical phenomenalism when it comes to specifying *subjunctively* the relevant powers to produce sensations. If this is right, there may not be as huge a gap as was historically supposed between phenomenalism and at least some causal theories of the physical world.

See also: Empiricism; Idealism; Perception, epistemic issues in

RICHARD FUMERTON

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, New York: Dover, 2nd edn, 1946. (Contains a vigorous defence of logical positivism and a phenomenalistic analysis of physical-object propositions in terms of actual and possible sense-data.)

Berkeley, G. (1713) *Three Dialogues Between Hylas and Philonous*, ed. C.M. Turbayne, Indianapolis, IN: Bobbs-Merrill, 1954.(Attacks representative realism and defends the view that physical objects can be reduced to 'ideas' - where 'ideas' include sensations. There are passages in which he seems to defend the phenomenalistic view that one should appeal to counterfactuals describing possible sensations in the analysis of physical objects.)

Chisholm, R. (1948) 'The Problem of Empiricism', *Journal of Philosophy* 45 (1948): 412-17.(Contains what is probably the single most influential objection to phenomenalism, presented in a remarkably clear and straightforward manner.)

Firth, R. (1950) 'Radical Empiricism and Perceptual Relativity', *Philosophical Review* 59 (1950): 164-83, 319-31.(A vigorous defence of phenomenalism against the argument from perceptual relativity. Firth's defence rests on certain general considerations about meaning, and the article, while clearly written, will be accessible only to those with a strong philosophical background.)

Lewis, C.I. (1946) *An Analysis of Knowledge and Valuation*, La Salle, IL: Open Court. (Despite his reservations about the word, Lewis in this book presents what I take to be the most sophisticated and detailed defence of phenomenalism. The book is accessible to relatively unsophisticated philosophy students provided that they are willing to spend some time studying it.)

Mill, J.S. (1865) *An Examination of Sir William Hamilton's Philosophy*, London: Longmans, Green, 1889.(Mill's identification of physical objects with the permanent possibility of sensations and his explication of what he means using subjunctive conditionals describing sensations makes him a good candidate for one of the earliest phenomenalists.)

Phenomenological movement

The phenomenological movement is a century-old international movement in philosophy that has penetrated most of the cultural disciplines, especially psychiatry and sociology. It began in Germany with the early work of Edmund Husserl, and spread to the rest of Europe, the Americas and Asia. In contrast with a school, a movement does not have a body of doctrine to which all participants agree; rather, there is a broad approach that tends to be shared. The phenomenological approach has at least four components.

First, phenomenologists tend to oppose naturalism. Naturalism includes behaviourism in psychology and positivism in social sciences and philosophy, and is a worldview based on the methods of the natural sciences. In contrast, phenomenologists tend to focus on the socio-historical or cultural lifeworld and to oppose all kinds of reductionism. Second, they tend to oppose speculative thinking and preoccupation with language, urging instead knowledge based on 'intuiting' or the 'seeing' of the matters themselves that thought is about. Third, they urge a technique of reflecting on processes within conscious life (or human existence) that emphasizes how such processes are directed at (or 'intentive to') objects and, correlatively, upon these objects as they present themselves or, in other words, as they are intended to. And fourth, phenomenologists tend to use analysis or explication as well as the seeing of the matters reflected upon to produce descriptions or interpretations both in particular and in universal or 'eidetic' terms. In addition, phenomenologists also tend to debate the feasibility of Husserl's procedure of transcendental epoché *or 'bracketing' and the project of transcendental first philosophy it serves, most phenomenology not being transcendental.*

Beyond these widely shared components of method, phenomenologists tend to belong to one or another of four intercommunicating and sometimes overlapping tendencies. These tendencies are 'realistic phenomenology', which emphasizes the seeing and describing of universal essences; 'constitutive phenomenology', which emphasizes accounting for objects in terms of the consciousness of them; 'existential phenomenology', which emphasizes aspects of human existence within the world; and 'hermeneutical phenomenology', which emphasizes the role of interpretation in all spheres of life. All tendencies go back to the early work of Husserl, but the existential and hermeneutical tendencies are also deeply influenced by the early work of Martin Heidegger. Other leading figures are Nicolai Hartmann, Roman Ingarden, Adolf Reinach and Max Scheler in realistic phenomenology, Dorion Cairns, Aron Gurwitsch and Alfred Schutz in constitutive phenomenology, Hannah Arendt, Jean-Paul Sartre, Maurice Merleau-Ponty, and Simone de Beauvoir in existential phenomenology, and Hans-Georg Gadamer and Paul Ricoeur in hermeneutical phenomenology.

1 Matrix and origins

Phenomenology began in the reflections of Edmund Husserl during the mid-1890s, but some find forerunners as far back as Plato and Aristotle. There are immediate anticipations in the work of four figures, not all of whom influenced all phenomenologists. In *Essai sur les données immediates de la conscience (Time and Free Will)* (1889), Henri Bergson offered a concrete and qualitative description of conscious life with an emphasis on how it flows and how abiding geometrized objects are constructed. He did not influence Husserl, but sits in the background for Roman Ingarden, KitarōNishida and French phenomenology. In *Psychologie vom empirischen Standpunkt (Psychology from an Empirical Standpoint)* (1874), Husserl's teacher Franz Brentano urged the priority of a descriptive over an explanatory psychology of psychical phenomena, which he distinguished from physical phenomena by their *Intentionalität* (intentionality) or directedness at immanent contents. Husserl eventually opposed his teacher's immanentism, denying that physical objects have an 'inexistence' in intentional acts; developed a richer classification of mental phenomena; and came to call his work phenomenology rather than descriptive psychology.

Wilhelm Dilthey similarly called for a descriptive psychology, held that it would be fundamental among the human sciences (*Geisteswissenschaften*), and described the construction of the historical world in such sciences. Husserl's interest in these matters took some time to be recognized; the debt Martin Heidegger owed to Dilthey has been more easily appreciated. Finally, in *Principles of Psychology* (1890) William James was also concerned to describe what he called the stream of thought, including believing and willing, and his distinction between 'the object of thought' and 'the topic of thought' resembles that between 'the object as it is intended' and 'the object

that is intended'. James was read by Husserl and some later phenomenologists.

After studying mathematics and astronomy at Leipzig, pursuing mathematics at Berlin, and hearing Brentano lecture at Vienna, Husserl took his doctorate in mathematics at Vienna under Leo Königsberger. He then habilitated under Brentano's disciple Carl Stumpf at Halle in 1887 and taught there as *Privatdozent* until 1901, when he became an Extraordinarius at Göttingen. He became Ordinarius at Freiburg in 1916, retired in 1928 and died in 1938.

Husserl's Berlin teacher in mathematics, Carl Weierstraß, encouraged the quest for absolutely secure foundations within mathematics, but Husserl went beyond mathematics to seek grounds for all the sciences. His first work, *Philosophie der Arithmetik* (1891), attempted to account for the concept of number by relating it, in the manner of Brentano's psychology, to the mental operation of counting. This work in descriptive psychology was soon contested by Gottlob Frege as being psychologistic - psychologism being the doctrine, prominently defended by John Stuart Mill and his followers in Germany, that empirical psychology is the fundamental philosophical discipline and that because concepts and propositions are mental contents, logic is a branch of psychology and logical laws are empirical psychological laws. The myth later arose that Frege helped Husserl overcome his psychologism, but close study by J.N. Mohanty, Karl Schuhmann and others of correspondence and minor writings has shown that Husserl took this step in 1894 for other reasons.

Early in his second major work, *Logische Untersuchungen (Logical Investigations)* (1900-1) - the work that actually launched the phenomenological movement - Husserl contended that logic is not fundamentally an art based on the facts of mental life, but instead fundamentally contains 'pure logic' as a theoretical science of ideal logical forms that are not themselves parts of conscious life. Later in this work he then reflected on the correlative psychical processes in which logical forms are intended, provoking accusations of a relapse into psychologism from those for whom any reference to conscious processes is anathema. What Husserl's less extreme anti-psychologism forbade was in fact the reduction of logical structures to real intentional processes.

The key doctrine in the latter part of the *Investigations* proposes that just as there can be fulfilment of empty intentions of sensuous objects, for example, when we see or hear the same matters as had merely been conceived of previously, there can also be fulfilment of empty categorial intentions by categorial intuition, that is, a non-sensuous seeing of how predication takes form. Propositional truth is accomplished when a formerly empty predicative judgment is brought into coincidence with a predicatively formed state of affairs. Husserl could then call for a return '*zu den Sachen selbst*', best rendered as 'to the matters themselves', that is to say, a return from the blind manipulation of symbols to an insightful approach to the corresponding states of affairs, which include the matters themselves of concern to formal logic. This injunction was soon generalized beyond the theory of logic, formal ontology and the theory of parts and wholes to regions of all sorts; the phenomenological movement then ensued.

The four successively emerging, intercommunicating and sometimes overlapping tendencies within the phenomenological movement thus far all stem from the so-called 'descriptive' phenomenology of the first edition of the *Logical Investigations*.

2 Realistic phenomenology

Immediately after the publication of the *Logical Investigations*, Johannes Daubert persuaded a group of fellow students of Theodore Lipps at Munich to abandon Lipps' psychologism and accept the *Logical Investigations* as their philosophical bible. Many of these students, including Adolf Reinach, soon went to Husserl at Göttingen. The phrase 'phenomenological movement' first arose in this group, and realistic phenomenology became a distinct tendency within it only when Husserl developed the so-called 'transcendental turn' that its members did not accept. Daubert published nothing in his lifetime. Alexander Pfänder's *Phänomenologie des Wollens (Phenomenology of Willing and Motivation)* (1900) is retrospectively seen as the earliest major document of realistic phenomenology, and he and Reinach, as well as Max Scheler and Moritz Geiger, led the first generation of realistic phenomenologists. In 1913, together with Husserl, they began editing the *Jahrbuch für Philosophie und phänomenologische Forschung*, the quasi-official organ of the movement. The second generation includes Theodor Celms, Hedwig Conrad-Martius, Dietrich von Hildebrand, Roman Ingarden, Aurel Kolnai, Edith Stein and Kurt Stavenhagen.

In order to gain a systematic body of a priori knowledge on a wide range of matters, this first tendency emphasizes Husserl's eidetic method. Eidetic method involves suspending belief in any actual facts with which one begins, feigning variations of the matter at issue, and then grasping the invariant or universal essence that the facts, fantasies and any 'thought experiments' exemplify or instantiate. Since there is always already a vague and tacit acquaintance with essences, eidetic method is a procedure of clarification and description: a method of discovery, not invention. Terms and relations of possibility, compossibility, necessity and contingency by virtue of which facts are intelligible are thus disclosed. Husserl importantly distinguished formalizing universalization, which yields formal ontology, from generalizing universalization, which yields taxonomies. Mistakes in employing this method - its use in accounting for itself included - can be made, but are also in principle discoverable and corrigible by means of it.

The realistic phenomenologists maintained a metaphysical realism of universals and particulars. Geiger contributed to aesthetics and the a priori foundations of geometry. In *Der Formalismus in der Ethik und die materiale Wertethik (Formalism in Ethics and Non-Formal Ethics of Values)* (1913, 1916) Scheler objected to Kant's ethical formalism and advocated a value-realism in ethics. Reinach analysed accusing, commanding, promising, questioning and other social speech acts in 'Die apriorischen Grundlagen des bürgerlichen Rechts' ('The Apriori Foundation of the Civil Law') (1913), thereby contributing to the philosophy of law as well as the human sciences. Conrad-Martius and Stavenhagen contributed to the philosophy of religion, and Stein is now recognized not only for reflections on empathy and the human sciences, but also, through lectures from around 1930 collected posthumously under the title *Die Frau* (1959), for contributions to feminism. Hartmann showed the influence of Husserl and Scheler in his rejection of Neo-Kantianism and his central reliance on eidetic method in *Grundzüge einer Metaphysik der Erkenntnis (Outline of a Metaphysics of Knowledge)* (1921) and *Ethik* (1925). Ingarden, chiefly known for *Das literarische Kunstwerk (The Literary Work of Art)* (1931), carried realistic phenomenology to Poland. Gustav Shpet introduced phenomenology into Russia.

Herbert Spiegelberg, a student of Pfänder, later wrote the monumental *The Phenomenological Movement* (1960) as well as descriptively oriented studies; more recently, Karl Schuhmann has functioned as the historian of realistic phenomenology. Barry Smith and David Woodruff Smith lead the efforts to connect current Anglo-American analytical philosophy not only with Brentano and related Austrian philosophy, but also with realistic phenomenology.

3 Constitutive phenomenology

The founding text of constitutive phenomenology is the first book of Husserl's *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie (Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy)* (1913). Posthumous works have made it clear that Husserl's transcendental constitutive phenomenology began by 1906 and is broader than the books published in his lifetime seem to show. Even during his lifetime he recognized in 'Nachwort zu meinen *Ideen*… ' ('Author's Preface to the English edition of *Ideas*… ') (1930) a 'constitutive phenomenology of the natural attitude', also called a 'mundane' or 'worldly' phenomenology, that amounts chiefly to phenomenological psychology. Much in the realistic, existential and hermeneutical tendencies can be seen as convergent with this mundane constitutive phenomenology. Nevertheless, the aim of most constitutive phenomenology is transcendental.

Constitutive phenomenology emphasizes processes within conscious life as they are intentive of objects, but it also reflects correlatively on the objects as intended in such processes. Constitutive phenomenology is specified by its concern with constitution. To analyse the 'constitution' of a matter is definitely not to distinguish the components of which it is composed, but rather to describe the syntheses of intentive processes in conscious life with which it correlates as an intentional object. The expression was taken from Kantianism, but is not confined to operations of conceptually structuring objects. There is pre-predicative experience in which objects are constituted as perceived, valued, willed and so on, but not yet formed into states of affairs.

Between *Logical Investigations* and *Ideas I*, Husserl published 'Philosophie als strenge Wissenschaft' (1911). It shows how his concern had broadened from the formal sciences of logic and mathematics to include the natural as well as the human or cultural sciences. Thus what he seeks in *Ideas I* are subjective conditions for the possibility of science of all kinds. His transcendentalism, however, differs from those of Kant and others in so far as he holds that the conscious life in which the world and worldly sciences could be subjectively grounded is itself the object

of reflective observation, eidetic intuition and description.

The opening part of *Ideas I* is devoted to eidetic method. It has unfortunately led to confusion between the eidetic and the transcendental methods, between going from 'facts' to 'essences' and going from conscious life in the world to conscious life as transcendental. Conscious life in its non-worldly or transcendental status is the same life that is originally encountered in the 'natural' or naively world-accepting attitude of the zoological and cultural sciences as well as in everyday life; but if the world is not to be grounded in part of itself, part of this worldly or natural (and realistic) attitude needs to be reduced to a transcendentally reflective attitude, and the conscious life, then reflectively thematized, needs to have its 'being in the world' placed in suspense. This is accomplished through transcendental phenomenological bracketing or *epoché*, a species of suspended judgment focused on the spatial, temporal and causal relations of conscious life with the rest of the world. In this attitude the world can be seen as an object intended to by non-worldly conscious life that, in the technical signification of the word, 'constitutes' it. 'Constitution' refers to the ways in which types of objects correlate with types of conscious processes. Husserl went on then to assert that conscious life has a more fundamental being than its being in the world, which not even all other transcendental phenomenologists accept.

In its middle parts *Ideas I* describes the natural attitude and transcendental *epoché*, offering detailed analyses of the parallel structures of the 'noema' or object as it is intended to and the 'noesis' or intentive conscious process in which objects are constituted. It also discusses how conscious life has an inner time in which each conscious process is 'protentive' to later and 'retrotentive' to earlier processes; how there is an 'I' who can engage in the processes strictly called acts; how sensuous 'stuff' is formed in perception; and how objects have characteristics as believed in, valued and willed, as well as modes of appearance and manners of givenness, including clarity and distinctness in recollection and imagination as well as in perception and 'eideation'.

The last part of *Ideas I* is devoted to rational justification. The theory of reason is the culmination of transcendental phenomenology. Justification for Husserl comes from the seeing, intuiting or 'evidencing' of the matters themselves. There is adequate and inadequate evidence, and apodictic and assertoric evidence, and such can directly and indirectly justify not only believing, but also valuing and willing; there is then epistemological, axiological and practical reason.

The second book of Husserl's *Ideas*, chiefly composed in 1912-15 and devoted to the natural and human sciences, was also worked on by his assistants Edith Stein and Ludwig Landgrebe; although it was not published until 1952, it was known in manuscript to Martin Heidegger before *Sein und Zeit (Being and Time)* (1927) and to Maurice Merleau-Ponty before *Phénoménologie de la perception (Phenomenology of Perception)* (1945). Husserl returned to the formal sciences in *Formale und transzendentale Logik (Formal and Transcendental Logic)* (1929).

At Freiburg during the 1920s and in retirement until his death in 1938, Husserl went beyond the 'static phenomenology' that uses eidetic method to disclose types of possible objects and consciousness. 'Genetic phenomenology', as he termed it, seeks to elucidate how active syntheses have origins in passive syntheses, a search that emphasizes time in individual life but also extends into history, intersubjectivity, the genesis of the lifeworld, and the teleology of conscious life in what he more broadly calls 'generative phenomenology'. These are all central issues in the later Husserl and in the background of his last work, *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie (The Crisis of European Sciences and Transcendental Phenomenology)* (1936).

Husserl's thought was eclipsed in Germany during the Nazi period, but continued to be developed after the Second World War by a number of figures, three of whom can be mentioned. First, in *Théorie du champ de la conscience (The Field of Consciousness)* (1957), Aron Gurwitsch draws upon Gestalt psychology to revise Husserl's accounts of the 'I' and attention, denying the need for the former as organizer and asserting the inherent organization of the field of consciousness into theme, thematic field, and margin. Second, in *Collected Papers* (1962-96), Alfred Schutz reflects from the standpoint of the 'constitutive phenomenology of the natural attitude' on the everyday common-sense constitution of the socio-cultural lifeworld and on how the cultural sciences - economics and sociology in particular - can know aspects of it. And, third, in *Zur Kritik der hermeneutischen Vernunft (The Critique of Hermeneutical Reason)* (1972), Thomas Seebohm returns to the traditional methodical hermeneutics as interpretation and critique of texts and traces that was pursued in Friedrich Schleiermacher, Augustus Boeckh and Wilhelm Dilthey, and seeks a transcendental phenomenological grounding for it.

It has been easy on the basis of the publications of his lifetime to caricature the mature Husserl as a modern-day (but nonrepresentationalist) Cartesian for whom disembodied and situationless intellects reflect upon the forms of their own thinking and have great difficulties knowing and interacting with one another. Closer study shows, however, that places for philosophy of the cultural as well as the natural sciences, for value theory and ethics, and for embodiment, empathy and communal life are sketched in those very same publications, although only developed in lectures and manuscripts, many of which have been published posthumously in *Husserliana* (see Phenomenology, epistemic issues in).

4 Existential phenomenology

Existential phenomenology is not structured by the complex concern for reason and the theory of science so prominent in constitutive phenomenology. Existential phenomenology draws ultimately upon the mundane reflective-descriptive spirit of the *Logical Investigations* as well upon the intensified interest in the 1920s and 1930s in Nietzsche and Kierkegaard, the latter urging a new signification for the word ‘existence’. The immediate occasion, however, is a misconstrual of Martin Heidegger’s *Being and Time*. This incomplete masterpiece is actually not devoted to human existence but rather ‘fundamental ontology’.

The old word ‘ontology’ had been revived by Husserl to name eidetic accounts of objects and their regions; realistic phenomenologists continue that usage, and Husserl investigates the regional ontologies of nature, body, psyche and culture in *Ideas II*. Attempting to radicalize constitutive phenomenology, Heidegger’s work is ‘ontology’ because it explicates the Being of beings (*Sein der Seienden*) and ‘fundamental’ because it seeks grounds beyond the mundane regional ontologies recognized by Husserl. The work contains an ‘existential analytic’ of human being or ‘Dasein’, not for a philosophical anthropology but as a means to this fundamental ontology.

Dasein - also translated as ‘existence’ or, in the early French translations of Henri Corbin, *réalité humaine* - is the being where the world is disclosed and the being whose mode of being is to understand Being, to bring it and related matters to light through seeing rather than constructing, and to find words for such matters. Dasein, Heidegger says, is being-in-the-world. This is not the world referred to in the positive sciences that Husserl emphasized even in *Die Krisis*, but rather the world as a set of everyday concerns and purposes, the world in which equipment is used and talk goes on. Dasein finds itself thrown into a situation not of its choosing; it is concerned with the future; it is for the most part distracted; and, deep down, it is anxious before its most extreme possibility - its own nothingness. But Dasein can heed the call of its own inmost possibility to live authentically and resolutely. Such terms were also used by Heidegger to support National Socialism during the 1930s, but they disappeared from his writing after the war, when he completed his turning (*Kehre*) from the oblique approach through Dasein to the direct thinking of Being.

Being was always Heidegger’s central issue. The third division of Part I of *Sein und Zeit* was to have gone beyond Dasein to show how the meaning of Being is time, but that division was not written, which made it even easier to construe the analytic of Dasein as philosophical anthropology, a construal that Heidegger emphatically challenged in his *Brief über den ‘Humanismus’ (Letter on Humanism)* (1947).

Hannah Arendt was influenced by Karl Jaspers as well as Heidegger during the 1920s, and is thus arguably the first existential phenomenologist, even though her contributions to political theory and problems of ethnicity, such as *The Origins of Totalitarianism* (1951), only appeared after the war. It is also arguable that existential phenomenology appears in Japan with Miki Kiyoshi’s *Pasukaru niokeru Ningen no Kenkyu (A Study of Man in Pascal)* (1926) and Kuki Shuzou’s *Iki no Kouzou (The Structure of Iki)* (1930). Chiefly, however, the existential tendency developed in France during the 1930s. The early Emmanuel Levinas interpreted Husserl and Heidegger together and helped introduce phenomenology into France and overall has more in common with the existential than with the other tendencies. Gabriel Marcel reflected upon fidelity, having, hope, promising and so on; opposed intellectualism and ‘objectivity’; and emphasized the embodiment, finitude, sensuousness and situatedness of existence in the world. His chief interest, however, independently paralleling Heidegger, is in Being as the ground of existence. Like Arendt, Sartre and Beauvoir, Marcel was not a professional academic and often wrote for general audiences.

The background influences on phenomenology in France in the 1930s also included Scheler, the rediscovered early

and humanistic Marx, and especially Hegel as presented by Jean Wahl and Alexandre Kojéve, who both argued for extensive convergencies between the phenomenologies of Hegel and Husserl. The issues of finitude, freedom, history, negation and individual and group conflict became prominent for Beauvoir, Merleau-Ponty and Sartre, who led the classic period of French phenomenology.

Jean-Paul Sartre studied Husserl and began to write on the ego, imagination and emotion in the mid-1930s, soon also studying Scheler and Heidegger. His *L'être et le néant (Being and Nothingness)* appeared in 1943. Sartre's approach relies on reflection upon, and eidetic description of, types of intentionality and objects as they present themselves, and he produced concrete analyses of many matters, for instance, historicity, authenticity, situation and especially individual freedom, which, for him, is the source of meaning and value. Human reality is what it has chosen to be, for existence precedes essence. In later work Sartre's emphasis on freely choosing individuals declined, he became doubtful about phenomenology's ability to explain historical conditioning, and turned to the writings of Marx.

Maurice Merleau-Ponty did pursue an academic career and chiefly wrote for fellow academics. He found many insights in science - especially cultural or human science, psychology in particular - but, like other existential phenomenologists, he opposed objectification and categorization and emphasized the ambiguous, concrete, contingent and particular. Against the early Sartre, he considered human freedom to be limited by its situation. In his main work, *Phenomenology of Perception*, he was concerned not with pure consciousness, but rather with human existence as embodied perception (or behaviour) in the world, and with how what is perceived has, for subjects, inherent structures of the sort described in Gestalt psychology and Aron Gurwitsch's work (see Gestalt psychology). While the body can be objectified in science, it is originally lived as subjective, and art, language, the other, politics, sexuality, space and so on, are to be analysed in relation to it. Merleau-Ponty died having composed only part of his body-focused ontology.

Simone de Beauvoir, often too closely associated philosophically with Sartre, is the third leading French existential phenomenologist. She likens existential conversion, a suspending of will in order to grasp the conditions of one's life, to Husserl's transcendental *epoché*; appreciates Heidegger's concern with the future, but finds change rather than being-toward-death central; and accepts from Merleau-Ponty that the human body is historical, denying, however, that a woman is her reproductive or sex-object body. For her, phenomenology is centrally concerned with friendship, as her autobiography and letters show, and also with age, class, ethnicity, gender, oppression and liberation. Beauvoir inspired the second wave of feminism with *Le deuxième sexe (The Second Sex)* (1949), opposing the myth by which categories defined in contrast with male categories are imposed on women; analysing the lived experience of meaning in feminine being-in-the-world; and urging that females are not born but become women (see Feminism).

Existential phenomenology spread widely from France. It was also extended to the human sciences, beginning with work in The Netherlands and Flanders during the 1950s, and has been represented in the United States by Maurice Natanson and in a structural version by Bernhard Waldenfels in Germany. It was eventually eclipsed by structuralism in France, but became central to the vast expansion of phenomenology in the United States that began in the 1960s - where the relevance of existential phenomenology for new problematics, such as feminism, is increasingly recognized (see Existentialism; Existentialist ethics; Existentialist theology).

5 Hermeneutical phenomenology

According to the *Logical Investigations*, perception, recollection, imagination and so on have a sense or meaning prior to expressions in propositional form, and some consider this insight an anticipation of hermeneutical phenomenology. This fourth tendency also begins - and without the 'existential' interpretation - in Heidegger's *Being and Time*. Traditional hermeneutics was chiefly interpretation of texts, but now all experience is seen as being affected by language and interpretation. Heidegger interprets 'phenomenology' as the *logos* of the *phainomenon*, these words being construed, respectively, as what makes matters manifest and what is made manifest, the latter including undisclosed as well as disclosed aspects. The phenomena of authenticity, death, care and above all Being itself are thus interpretable. Dasein always already has some understanding of Being that can be refined through philosophical interpretation, although the truth thus won conceals as well as reveals. The analysis of Dasein that so influenced existential phenomenology is actually a hermeneutics of Dasein that seeks to bring out hidden aspects. This includes self-interpretations, which refer back to earlier generations and are thus

historical. The hermeneutical approach, especially to texts, continues in Heidegger's later work, although the word hermeneutics does not.

The first phenomenological interpretation beyond Heidegger is Hans-Georg Gadamer's *Platons dialektische Ethik (Plato's Dialectical Ethics)* (1931) (see Gadamer, H.-G.). It accepts Heidegger's notion of revealing and concealing truth as well as his focus on the active participant in life rather than the scientific observer. Its emphasis on the ethical aspects of the openness of one interlocutor to another is continued in his quite influential *Wahrheit und Methode (Truth and Method)* (1960). Not exclusively focused on texts, this is a 'philosophical hermeneutics' that is concerned with the general theory of understanding, that defends tradition and authority against Enlightenment attacks, and that is urged for use in law, literature and theology. It remains phenomenological in its use of Husserl's notion of intentionality and theory of perception to oppose naturalism and relativism. Interpretations even of the unthought in a text seek to be fulfilled by the matters themselves referred to by the text.

Paul Ricoeur studied Marcel and Jaspers as well as Husserl and Heidegger while interned as a prisoner of war. In 1950 he published the translation of Husserl's *Ideen I*, which he began in the camp; he subsequently played a central role in advancing French Husserl scholarship and regularly contended that his own evolving position was compatible with Husserl's original inspiration, whereby meaning is transcendent of conscious life. The first expression of his own thought converged with realistic phenomenology in employing eidetic method to analyse the voluntary and involuntary (he had also studied Pfänder's work). His concerns then with freedom, the other and evil converged with existential phenomenology, and, finally, he has focused on understanding as requiring texts or text-like structures and he thus joined hermeneutical phenomenology. He interprets not only religious symbols, for example, of the creation, but also the unconscious of psychoanalysis in works such as *Le conflit des interpretations (The Conflict of Interpretations)* (1969a). More recent work interprets metaphor, time, narrative, the 'same' or 'self' and the other.

Gadamer and Ricoeur have promoted hermeneutical phenomenology quite actively and this tendency has been strong in the United States. Calvin O. Schrag has contributed to the philosophy of language in *Experience and Being* (1969), Don Ihde to the philosophy of technology in *Technics and Praxis* (1979), Graeme Nicholson to the philosophy of perception in *Seeing and Reading* (1984), and then there is Patrick Heelan's *Space-Perception and the Philosophy of Science* (1983) and Joseph J. Kockelmans' *Ideas for a Hermeneutic Phenomenology of Natural Science* (1993). Other philosophical work has been done in aesthetics, ethics, history, language, law, literature, politics and religion. Hermeneutical phenomenology has extensively influenced not only the philosophy of the human sciences but the human sciences themselves.

Lest the debates within the movement and the structure of this essay give the impression that phenomenology went off in four separate ways, it must be emphasized that hermeneutical phenomenology draws nearly as much on existential and constitutive phenomenology as on the early Heidegger; that, while highly original, existential phenomenology is conscious of its central inspiration in Husserl and Scheler as well as Heidegger; that there has been from the outset extensive mutual borrowing as well as criticism between constitutive and realistic phenomenology; and that, by virtue of his effort reflectively to analyse and describe the matters themselves of conscious life and what is, in manifold ways, intended in it, Husserl is, as Ricoeur has said, not the whole of phenomenology, but he is 'more or less its centre' (see Hermeneutics).

6 Prospects

The phenomenological movement began with Husserl's *Logical Investigations*. History has shown that Wilhelm Dilthey was correct to proclaim this work 'epochal'. Besides the tendencies sketched above, a shifting geographical focus can be noted. This focus was in Germany until 1933, then shifted to France until about 1960; after that, while inspiration came from both Germany and France, the largest part of phenomenologists have come from the United States. Other enduring national traditions of phenomenology began in Japan, Russia and Spain before the First World War; arose in Australia, Czechoslovakia, Hungary, Italy, Korea, The Netherlands and Flanders, Poland and Yugoslavia, as well as the United States and France between the wars; and emerged after the Second World War in Canada, China, Great Britain, India, Portugal, Scandinavia and South Africa. By the 1980s, genuinely international (and not just transatlantic) conferences and other forms of collaboration were intensifying; now that generations-deep underground tendencies have begun surfacing after the end of the Cold War, it is all the

more likely that this trend will continue. It may turn out that the German, French and American periods of the phenomenological movement began to be succeeded in the early 1990s by an international period in which there are many centres.

As a century-old, world-wide, still growing and increasingly multidisciplinary movement, phenomenology is arguably the central movement in twentieth-century philosophy, and its vitality and momentum should carry it far into the twenty-first century.

See also: Phenomenology in latin america; Phenomenology of religion

LESTER EMBREE

References and further reading

Arendt, H. (1951) *The Origins of Totalitarianism*, New York: Harcourt Brace Jovanovich; 2nd edn, enlarged, 1958.(A mature work in which history and ethnicity are emphasized as aspects of communal human existence.)

Beauvoir, S. de (1949) *Le deuxième sexe*, Paris: Gallimard; trans. H.M. Parshley, *The Second Sex*, New York: Alfred A. Knopf, 1953.(The founding text for the second stage of feminism, which emphasizes not just formal equality but gender-specific cultural attitudes that need to be reconciled.)

Becker, O. (1923) 'Beiträge zur phänomenologischen Begründung der Geometrie und ihrer physikalischen Anwendungen' ('Contributions to the phenomenological grounding of geometry and its applications in physics'), *Jahrbuch für Philosophie und phänomenologische Forschung* 7.(As the title indicates, deals with geometry and its applications.)

Bergson, H.-L. (1889) *Essai sur les données immediates de la conscience*, Paris: Alcan; trans. F.L. Pogson, *Time and Free Will*, London: Macmillan, 1910.(Anticipates the phenomenological focus on inner time, the time of the stream of conscious life, in contrast with the time coupled with space in which trees grow.)

Brentano, F. (1874) *Psychologie vom empirischen Standpunkt*, ed. O. Kraus, Hamburg: Felix Meiner; trans. A.C. Rancurello, D.B. Terrell and L.L. McAlister and ed. L.L. McAlister, *Psychology from an Empirical Standpoint*, London, Routledge, 1969; 2nd edn, 1995.(Fundamental text of the descriptive psychology from which phenomenology arose through the rejection of immanentism.)

Caputo, J. (1987) *Radical Hermeneutics*, Bloomington, IN: Indiana University Press.(Example of recent hermeneutical phenomenology.)

Derrida, J. (1990) *Le probléme de la génèse dans la philosophie de Husserl (The problem of genesis in Husserl's philosophy)*, Paris: Presses Universitaires de France.(Beginning of this author's challenging interpretations of Husserl.)

Dilthey, W. (1985-) *Selected Works*, ed. R.A. Makkreel and F. Rodi, Princeton, NJ: Princeton University Press, 6 vols.(Central author in relation to whose work phenomenologists reflect on the cultural sciences.)

Embree, L. *et al.* (1996) *The Encyclopedia of Phenomenology*, Dordrecht: Kluwer.(Reviews the whole movement and can be used to approach particular figures, national traditions, and tendencies in it.)

Gadamer, H.-G. (1931) 'Platos dialektische Ethik: Phänomenologische Interpretationen sum Philebos', in *Platos dialektische Ethik*, Hamburg: Felix Meiner, 1968; trans. *Plato's Dialectical Ethics*, New Haven, CT: Yale University, 1991.(Arguably the first document of hermeneutical phenomenology beyond Heidegger.)

Gadamer, H.-G. (1960) *Wahrheit und Methode*, Tübingen: J.C.B. Mohr; trans. *Truth and Method*, London: Sheed & Ward, 1975.(This author's masterpiece.)

Gurwitsch, A. (1957) *The Field of Consciousness*, Pittsburgh, PA: Duquesne University Press, 1964; trans. M. Butor, *Théorie du champ de la conscience*, Bruges and Paris: Desclée de Brouwer.(Creative development of constitutive phenomenology incorporating descriptive results from Gestalt psychology.)

Hartmann, N. (1921) *Grundzüge einer Metaphysik der Erkenntnis (Outline of a Metaphysics of Knowledge)*, Berlin: de Gruyter.(Metaphysical development from Husserl and realistic phenomenology.)

Hartmann, N. (1925) *Ethik*, Berlin: de Gruyter; trans. S. Coit, *Ethics*, London: George Allen & Unwin, 1932. (Major contribution to phenomenological ethics.)

Heelan, P. (1983) *Space-Perception and the Philosophy of Science*, Berkeley, CA: University of California Press. (Recent hermeneutical contribution.)

Heidegger, M. (1927) 'Sein und Zeit', *Jahrbuch für Phänomenologie und phänomenologische Forschung* 8; trans. J. Macquarrie and E. Robinson, *Being and Time*, London: SCM Press, 1962.(Incomplete masterpiece of second most important figure in phenomenology and source of hermeneutical as well as existential phenomenology.)

Heidegger, M. (1947) *Platons Lehre von der Wahrheit. Mit einem Brief über den 'Humanismus'*, Bern: Francke; trans. F.A. Capuzzo and J.G. Gray, 'Letter on Humanism', in M. Heidegger, *Basic Writings*, ed. D.F. Krell,

New York: Harper & Row, 1986, 193-242.(Text in which Heidegger rejects existential interpretation of *Sein und Zeit*.)

Husserl, E. (1891) *Philosophie der Arithmetik*, ed. L. Eley, The Hague: Martinus Nijhoff, 1970.(Attempts to account for the concept of number by relating it to the mental operation of counting.)

Husserl, E. (1900-1) *Logische Untersuchungen*, Halle: Max Niemeyer; trans. J.N. Findlay, *Logical Investigations*, London: Routledge & Kegan Paul, 1970.(Epoch-making source of the phenomenological movement in which the reflective-descriptive approach was radicalized and extended beyond the original epistemology of logic and mathematics.)

Husserl, E. (1911) 'Philosophie als strenge Wissenschaft', *Logos* 1 (1910-11) 289-341; trans. as 'Philosophy as rigorous science', in *Husserl: Shorter Works*, ed. P. McCormick and F.A. Elliston, Notre Dame, IN: University of Notre Dame Press, 1981, 166-97.(Shows how Husserl's concern broadened from logic and mathematics to include the natural sciences.)

Husserl, E. (1913) *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, Erstes Buch*, in *Husserliana* vol. 3, ed. W. Biemel, The Hague: Martinus Nijhoff, 1976; trans. F. Kersten, *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, First Book*, The Hague: Martinus Nijhoff, 1983.(Founding text of Husserl's mature constitutive phenomenology.)

Husserl, E. (1913) *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, Zweites Buch*, in *Husserliana* vol. 4, ed. M. Biemel, The Hague: Martinus Nijhoff, 1952; trans. R. Rojcewicz and A. Schuwer, *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, Second Book*, The Hague: Martinus Nijhoff, 1989.(Shows Husserl's theory of the lived body and interest in the cultural as well as the natural sciences.)

Husserl, E. (1929) 'Formale und transzendentale Logik: Versuch einer Kritik der logischen Vernunft', *Jahrbuch für Philosophie und phänomenologischeForschung* 10: v-xii; 1-298; trans. D. Cairns, *Formal and Transcendental Logic*, The Hague: Martinus Nijhoff, 1969.(Husserl's mature philosphy of logic and mathematics.)

Husserl, E. (1930) 'Nachwort zu meinen *Ideen…* '; trans. as 'Author's Preface to the English edition of *Ideas…* ', in *Husserl: Shorter Works*, ed. P. McCormick and F.A. Elliston, Notre Dame, IN: University of Notre Dame Press, 1981, 43-53.(Recognizes a 'mundane' or 'worldly' phenomenology, that amounts chiefly to phenomenological psychology.)

Husserl, E. (1931) *Cartesianische Meditationen*, in *Husserliana* vol. 1, ed. S. Strasser, The Hague: Martinus Nijhoff, 1960; trans. G. Pfeiffer and E. Levinas, *Méditations Cartésiennes*, Paris: Vrin, 1931; trans. D. Cairns, *Cartesian Meditations*, The Hague: Martinus Nijhoff, 1960.(Short, long available, but quite difficult atttempt by Husserl to summarize his position.)

Husserl, E. (1936) *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*, in *Husserliana* vol. 4, ed. W. Biemel, The Hague: Martinus Nijhoff, 1954; trans. D. Carr, *The Crisis of European Sciences and Transcendental Phenomenology*, Evanston, IL: Northwestern University Press, 1972.(Husserl's last work, this text shows an historically oriented approach to natural science, emphasizes the lifeworld, and the relation of psychology to other cultural sciences with transcendental phenomenology.)

Ihde, D. (1979) *Technics and Praxis*, Dordrecht: Reidel.(The beginning of this author's rich phenomenological philosophy of technology.)

Ingarden, R.W. (1931) *Das literarische Kunstwerk*, Tübingen: Max Niemeyer; trans. G.G. Grabowicz, *The Literary Work of Art*, Evanston, IL: Northwestern University Press, 1973.(Major contribution to phenomenological aesthetics.)

James, W. (1890) *Principles of Psychology*, New York: Henry Holt.(Anticipation of phenomenology recognized by Husserl, Gurwitsch, Schutz and others.)

Kuki, S. (1930) *Iki no Kouzou (The structure of Iki)*, Tokyo: Iwanami-shoten.(Arguably an early contribution to existential phenomenology.)

Kockelmans, J.J. (1993) *Ideas for a Hermeneutic Phenomenology of Natural Science*, Dordrecht: Kluwer.(Culminating contribution to hermeneutical phenomenology.)

Merleau-Ponty, M. (1945) *Phénoménologie de la perception*, Paris: Gallimard; trans. C. Smith, *Phenomenology of Perception*, London: Routledge & Kegan Paul, 1962.(Major text of this author and of existential phenomenology.)

Miki, K. (1926) *Pasukaru niokeru Ningen no Kenkyu (A Study of Man in Pascal)*, Tokyo: Iwanami-shoten.(Probably the first manifestation of existential phenomenology, but long unrecognized in the West.)

Mohanty, J.N. (1982) *Husserl and Frege*, Bloomington, IN: Indiana University Press.(Studies connections between these two tradition founders and dispels myth that Husserl overcame his psychologism under Frege's

influence.)
Mohanty, J.N. (1989) *Transcendental Phenomenology: An Analytic Account*, Oxford: Blackwell.(Excellent recent presentation.)
Mulligan, K. (ed.) (1987) *Speech Act and Sachverhalt: Reinach and the Foundations of Realist Phenomenology*, The Hague: Martinus Nijhoff.(As the title indicates, discusses the foundations of realist phenomenology.)
Nicholson, G. (1984) *Seeing and Reading*, Atlantic Highlands, NJ: Humanities Press International.(Hermeneutical phenomenology of perception.)
Pfänder, A. (1900) *Phänomenologie des Wollens*, Leipzig: Johann Ambrosius Barth; trans. H. Spiegelberg, in *Phenomenology of Willing and Motivation*, Evanston, IL: Northwestern University Press, 1967.(Important early text of realistic phenomenology.)
Reinach, A. (1913) ‘Die apriorischen Grundlagen des bürgerlichen Rechts’, in *Sämtliche Werke*, ed. K. Schuhmann and B. Smith; trans. ‘The Apriori Foundation of the Civil Law’, *Alethia* (1982) 3: 1-142.(As the title indicates, deals with the apriori foundations of civil law.)
Ricoeur, P. (1969a) *Le conflit des interprétations*. *Essais d’hermeneutique*, Paris: Seuil; ed. D. Ihde; trans. W. Domingo *et al*., *The Conflict of Interpretations: Essays in Hermeneutics*, Evanston, IL: Northwestern University Press, 1974.(Mature expression of author’s hermeneutical phenomenology.)
Ricoeur, P. (1969b) *Hermeneutics and the Human Sciences*, New York: Cambridge University Press. (Hermeneutical phenomenology of human or cultural sciences.)
Sartre, J.-P. (1943) *L’être et le néant*, Paris: Gallimard; trans. H.E. Barnes, *Being and Nothingness*, New York: Philosophical Library, 1956.(Major source of existential phenomenological ontology.)
Scheler, M. (1913, 1916) *Der Formalismus in der Ethik und die Materiale Wertethik*, Bern: A. Francke; trans. M.S. Frings and R.L. Funk, *Formalism in Ethics and Non-Formal Ethics of Values*, Evanston, IL: Northwestern University Press, 1973.(Major ethical position from standpoint of realistic phenomenology.)
Schrag, C.O. (1969) *Experience and Being*, Evanston, IL: Northwestern University Press.(Early expression of hermeneutical phenomenology in United States.)
Schutz, A. (1962-96) *Collected Papers*, Dordrecht: Kluwer Academic Publishers.(Highly influential source for constitutive phenomenology of the natural attitude and phenomenology of the social type of natural sciences.)
Seebohm, T. (1972) *Zur Kritik der hermeneutischen Vernunft (The Critique of Hermeneutical Reason)*, Bonn: Bouvier.(Transcendental phenomenological grounding of the method of the historical human sciences.)
Sokolowski, R. (1974) *Husserlian Meditations*, Evanston, IL: Northwestern University Press.(Fine example of Husserlian research.)
Spiegelberg, H. (1982) *The Phenomenological Movement*, 3rd edn, revised and enlarged with the collaboration of K. Schuhmann, The Hague: Kluwer Academic Publishers.(Classic study only superceded by Embree’s 1996 *Encylopedia*.)
Stein, E. (1959) *Die Frau*, *Ihre Aufgabe nach Natur und Gnade*, Louvain: E. Nauwelaerts; trans. F.M. Oben, *Essays on Women*, in *Collected Works*, vol. 2, Washington, DC: ICS Publications, 1987.(Late recognized inception of phenomenological investigations of gender.)

Phenomenology in Latin America

The Latin American struggle against the positivism of the nineteenth century was the primordial endeavour of the founders of Latin American thought, such as José Enrique Rodó (1872-1917), José Vasconcelos (1882-1959), Alejandro Korn (1860-1936), Carlos Vaz Ferreira (1871-1958), Alejandro Deústua (1849-1945), Enrique Molina (1871-1956) and Antonio Caso (1883-1946). These thinkers fought to win their philosophical freedom in a battle against continuing such European currents as Neo-Kantianism and the existential-phenomenological movement, on the one hand and on the other, developing a philosophy that was purely Latin American.

In all of this the influence of the Spanish philosopher José Ortega y Gasset (1883-1955) was fundamental. Ortega y Gasset was mainly interested in the problems of history and culture. His famous apothegm, 'I am myself and my circumstance' (1947: 322), was the justification for a phenomenological movement axiological in nature, as well as being the point of departure for the affirmation of Latin American circumstance.

Few have cultivated the Husserlian style of phenomenology in Latin America outside the classroom. Latin American thinkers have preferred to apply the phenomenological method to the different fields of knowledge. In particular, those dealing with sociocultural aspects, literary criticism, socio-economic structures and juridical axiology. In all works related to axiology the thought of the philosophers Nicolai Hartmann (1882-1950) and Max Scheler (1874-1928) were enormously influential.

Many of those originally inclined towards phenomenology drifted from the 1970s onwards, towards analytic philosophy, philosophy of science and logical neopositivism.

1 Phenomenology and literature

Theoretical Latin American writers in the field of literary criticism are interested especially in the ontology of the literary work. This interest, which includes the task of discovering and/or constructing Latin American subjectivity and identity in relation to their specific and differing surroundings, focuses on the basic principles of phenomenological epistemology. They base their own epistemology on the epistemological thinking of early Husserl. This phase of the philosophy of Husserl (§3), also known as neorealism, states that reality can be grasped in the essences as revealed in consciousness.

Particularly important is the application to literature of the Husserlian formula the mutual implication of subject and object and his notion of transcendental subjectivity. Graciela Maturo (1928-), for example, uses these ideas to achieve her goal of explaining (and to construct by explanation) the peculiarity of Latin American subjectivity. This critical approach to Latin American literature is justified because, as in other countries, especially France, Spain, Germany and Italy, the literary work of art is used to reveal aesthetic, ethical or metaphysical philosophical intuitions and preoccupations. Maturo makes this point in her writings and the phenomenological method helps her to describe the literary phenomenon on a particularly interesting continent. In doing so, she joins other writers and thinkers who have tried to accomplish the construction of a Latin American identity, which is separate from, although influenced by Europe (see Marginality; Cultural identity).

2 Relevant thinkers: Argentina

Carlos Astrada (1894-1970) was professor at the University of La Plata and Buenos Aires. His background was in Neo-Kantianism under the teachings of Scheler, Husserl and Heidegger in Freiburg (see Neo-Kantianism). He first oriented his work towards existential phenomenology (1936), but later returned to Kant and moved on to Marx, then Hegel (1967). Astrada interpreted authentic existential freedom more as social liberty than as individual freedom. He contributed interesting works in the area of phenomenology.

Arturo Cambours Ocampo applied the phenomenological method to literary creation in his book, *Lenguaje y creación: notas para una fenomenología del estilo literario (Language and Creation: Notes Towards a Phenomenology of Literary Style)* (1970). He showed the usefulness of the phenomenological method for studying the language of literature in *Literatura y estilo: notas para una estética del escritor y su experiencia fenomenólogica (Literature and Style: Notes Towards an Aesthetic of the Writer and their Phenomenological Experience)* (1985). His aim was to establish the aesthetics upon which the literary style is founded.

Raúl Castagnino is a literary critic with a phenomenological orientation. He is a member of the phenomenological movement which is literary in nature and has many followers in Latin America and elsewhere. He became known especially for *El análisis literario: introducción metodológica a una estilística integral (Literary Analysis: Methodological introduction to an Integral Stylistics)* (1953).

Carlos Cossio (1903-) was mostly preoccupied with the philosophy of law. He managed to achieve an original synthesis between phenomenology and existentialism, asserting that the object of law is human behaviour: the norms of behaviour were not themselves laws.

Juan Luis Guerrero (1899-1957) forged his academic career as professor of ethics at the University of Buenos Aires and of aesthetics at the University of La Plata. Influenced by Max Scheler, Guerrero's conception was centred around the gnoseological aspect of individual subjectivity, on the ethics of the human personality and on aesthetics.

J.A. Madile's primary interest was the phenomenologcial nature of the epistemological problem. The influence of Nicolai Hartmann is clearly evident in his work.

Graciela Maturo (1928-) followed a current of thought which was also intensively cultivated in Europe to apply her phenomenologcial knowledge and reflections in the field of literature. She considered phenomenology the best methodology and the most appropriate approach to unravel the elements characteristic of Latin American literature and literary criticism, as she states in *Fenomenología, creación y crítica: sujeto y mundo en la novela latinoamericana (Phenomenology, Creation, and Criticism: Subject and World in the Latin American novel)* (1989). She also discusses the interaction of Latin American versus European identities. In *Imagen y expresión: hermeneútica y teoría literaria desde America Latina (Image and Expression: Hermeneutics and Literary Theory from Latin America)* (1991), several authors contribute to the influence of phenomenological-existentialist hermeneutics, principally cultivated in France.

3 Relevant thinkers: Brazil

Creso Coimbra was interested in Latin American identity. He applied phenomenology to the problems of culture in his work *Fenomenología da cultura brasileira (Phenomenology of Brazilian Culture)* (1972).

Julio Fragata was a follower of Husserl and attempted to establish a Husserlian phenomenology as the basis of all philosophical enterprise, as illustrated in *A fenomenologia de Husserl como fundamento da filosofia (The Phenomenology of Husserl as the Foundation of Philosophy)* (1959). Subsequently in *Problemas da fenomenologia do Husserl (Problems in the Phenomenology of Husserl)* (1962), he analysed the contrast between the early and later works of Husserl.

Alfredo Naffah Neto (1947-) witnessed both state terrorism and the work of the country's death squads. He successfully applied the phenomenological method to the historical situation of the last dictatorships of Brazil (1985).

4 Relevant thinkers: Colombia, Costa Rica, Chile

In Colombia Danilo Cruz Vélez was interested in the problems of the philosophy of culture and philosophical anthropology. The influence of Scheler and Heidegger were in evidence in his central thesis (1970) which focused on the concept of philosophy as a perpetual beginning, without any presuppositions.

In Costa Rica Helio Gallardo (1942-) was involved with the trend of searching for a philosophy deeply rooted in Latin American social problems. He implemented phenomenological analysis to understand those problems. On the occasion of the quincentenary of the encounter of cultures in the Americas, he wrote *500 años: fenomenología del mestizo: violencia y resistencia (500 Years: Phenomenology of the Mestizo: Violence and Resistance)* (1993).

In Chile Mario Ciudad was influenced by Henri-Louis Bergson, as was the entire first generation of the opponents of logical neopositivism. His work, *Bergson y Husserl: diversidad en la coincidencia (Bergson and Husserl: Diversity in Coincidence)* (1960) showed an interest in the phenomenology of Husserl and tried to find points of coincidence with the philosophy of life of Bergson.

5 Relevant thinkers: Mexico

Antonio Caso (1883-1946) became well known through a series of lectures in which he refuted the wave of positivism at that time. He introduced new philosophical currents, emphasizing the philosophy of Bergson and Husserl. The latter enabled him to broaden the idea of positivism. He worked on the development of his own thought inspired by his intuitions with respect to essence and existence. His distinctiveness derived from his pluralist concept of reality and his interpretation of philosophy as a kind of synthesis of reflections about nature and culture. For Caso, philosophy is an integratory superscience in which existence and thought are concordant.

Carmen Cecilia Hernández de Ragoña utilized phenomenological methods in *Fenomenología histórica de las religiones (Historical Phenomenology of Religions)* (1975) to analyse other fields, cultivating especially the phenomenology of religion.

Samuel Ramos (1897-1959), a disciple of Antonio Caso, was influenced by Ortega y Gasset and Scheler. He was concerned with humanity generally, starting with the Mexican (1934). He attempted the affirmation of freedom with personality by overcoming the conditioning imposed by the circumstances of civilization.

Arturo Rivas Saínz applied the phenomenological method to the poetic phenomenon with clarity, insight and a great command of metaphysical intuition (1950).

6 Relevant thinkers: Peru, Puerto Rico, Uruguay, Venezuela

Peruvian, Francisco Miró Quesada Cantuarias (1918-) wrote extensively in the areas of the phenomenology of logic and the philosophy of science (1941), both of which he deemed rigorous enough to resolve the problems of humankind. He never abandoned his preoccupation with humanity's problems in the ethical-political sphere. His compatriot José Tamayo Herrera was interested in mythopoetic activity - the ability to create myths. His ideas feature in *Fenomenología de la creación poética (Phenomenology of the Poetic Creation)* (1963).

In Puerto Rico, Charles Rosario, with a focus on poetry and philosophy, applied his knowledge of phenomenological philosophy to contemporary Puerto Rican society (1965). From 1898 there was a convergence of two contrasting points of view which have remained unresolved. These are the Puerto Rican's identity with Spanish and Anglo-Saxon cultures, both of which are based on the economic and political power of the USA and the need to reaffirm a distinct Puerto Rican cultural identity through literature and thought. Rosario uses the phenomenological method which describes the social, moral, literary and other phenomena to explain the traumatic change from an agricultural society to an industrial one. The change produced a crisis for Puerto Ricans in the perception of their own identity, especially in comparison with other Latin American countries.

In Uruguay Homero Altesor in his book, *Fenomenología del cuerpo: anatomía filosófica (Phenomenology of the Body: Philosophical Anatomy)* (1986) applied phenomenology to the analysis of the different parts of the human body as a function of the total harmony of humankind.

Juan Llambías de Azevedo (1907-1972) based juridical axiology on the phenomenology of Scheler and Hartmann, as well as Husserl, to establish the reality of rights as an objective value independent from any subjective judgment. Llambías de Azevedo (1940) explained a prolix phenomenology of the acts of conscience that allow the constitution of the object of the law. The investigation of the essence of the object of the law drew him, after Nicolai Hartmann, towards identifying an aporetics of law. The most important aspect of this aporetic was the relationship between the content of law and values. In addition to any subjective dimension of the judgment of values Llambías de Azevedo rejected any relativism with respect to the subjective dimension of it. Following Scheler, he submits that the supreme value is the person, who is the ultimate foundation of the law and of the values of justice and equality.

In Venezuela Ernesto Mayz Vallenilla (1925-) wrote several works interpreting and expounding the works of Dilthey, Husserl and Heidegger (1956; 1960). His thesis was based on the conception of temporality which starts from nothingness. His thought is within the tradition of the propounding of Americanist thought. The main part of his work has compared and contrasted Heidegger and Marx.

See also: Existentialist thought in Latin America; Phenomenological movement

MARÍA TERESA BERTELLONI

References and further reading

Aguinsky de Iribarne, E. (1971) *Fenomenología y ontología jurídica*, Buenes Aires: Ediciones Pannedille.

Altesor, H. (1986) *Fenomenología del cuerpo: anatomía filosófica (Phenomenology of the Body: Philosophical Anatomy)*, Montevideo: Imprenta Mimeográfica Pesce.(Phenomenology applied to different parts of the human body.)

Astrada, C. (1936) *Idealismo fenomenológico y metafísica existencial (Phenomenological Idealism and Existential Metaphysics)*, Buenos Aires: Imprenta de la Universidad.(Existential phenomenology.)

Astrada, C. (1967) *Fenomenología y praxis (Phenomenology and Praxis)*, Buenos Aires: Ediciones Siglo Veinte. (Astrada discusses the practice of phenomenology.)

Barreto, O. (1992) 'Fenomenología de la religiosidad mapuche' (Phenomenology of Mapuche Religiosity), *Documentario Patagónico (Patagonian Documentary)*, Buenos Aires: Centro Salesiano de Estudios 'San Juan Bosco'.(A series from the Salesian Historical Archive of northern Patagonia.)

Bormida, M. (1976) *Etnología y fenomenología: ideas acerca de una hermenéutica del extrañamiento (Ethnology and Phenomenology: Ideas About the Hermeneutics of Estrangement)*, Buenos Aires: Ediciones Cervantes. (Part of the trend of affirmation of Hispanic US thought.)

Cambours Ocampo, A. (1970) *Lenguaje y creación: notas para una fenomenología del estilo literario (Language and Creation: Notes Towards a Phenomenology of Literary Style)*, Buenos Aires: Ediciones Cervantes.(An application of phenomenology to literary creation.)

Cambours Ocampo, A. (1985) *Literatura y estilo: notas para una estética del escritor y su experiencia fenomenólogica (Literature and Style: Notes Towards an Aesthetic of the Writer and their Phenomenological Experience)*, Buenos Aires: Ediciones Marymar.(The text includes a bibliography of critical texts.)

Carvalho da Costa, V. (1983) *Umbanda: os seres superiores e os orixas/santos: un estudio sobre a fenomenologia do sincretismo umbandistico na perspectiva da teologia catolica*, São Paulo: Ediciones Loyola.

Caso, A. (1914) *La filosofía de la intención (The Philosophy of Intuition)*.

Caso, A. (1933) *El concepto de la historia universal y la filosofía de los valores (The Concept of Universal History and the Philosophy of Values)*, Mexico: Ediciones Botas.(The writer introduces new philosophical currents.)

Caso, A. (1934) *La filosofía de Husserl (The Philosophy of Husserl)*, Mexico: Imprenta Mundial.(A discussion of the phenomenologist.)

Caso, A. (1941) *Positivismo, neopositivismo y fenomenología (Positivism, Neopositivism and Phenomenology)*, Mexico: Centro de Estudios Filosóficos de la Facultad de Filosofía y Letras.(The author broadens the idea of positivism, developing ideas inspired by his intuitions with respect to essence and existence.)

Castagnino, R.H. (1953) *El análisis literario: introducción metodológica a una estilística integral (Literary Analysis: Methodological introduction to an Integral Stylistics)*, Buenos Aires: Editorial Nova.(A stylistic analysis from a phenomenological perspective.)

Castagnino, R.H. (1980) *Fenomenología de lo poético (Phenomenology of the Poetic)*.

Ciudad, M. (1960) *Bergson y Husserl: diversidad en la coincidencia (Bergson and Husserl: Diversity in Coincidence)*, Santiago de Chile: Anales de la Universidad de Chile.(The author compares the two philosophers.)

Coimbra, C. (1972) *Fenomenología da cultura brasileira (Phenomenology of Brazilian Culture)*, São Paulo: LISA.(Phenomenology applied to Brazilian cultural problems.)

Cruz Vélez, D. (1948) *Nueva imagen del hombre en la cultura (New Image of Man in Culture)*.

Cruz Vélez, D. (1962) *Problemas da fenomenologia do Husserl (Problems in the Phenomenology of Husserl)*, Braga: Livraria Cruz.(Compares early and later works of Husserl.)

Cruz Vélez, D. (1970) *Filosofía sin supuestos: de Husserl a Heidegger (Philosophy without Presuppositions: from Husserl to Heidegger)*, Buenos Aires: Editorial Sudamericana.(Influenced by Scheler and Heidegger.)

Cruz Vélez, D. (1977) *Aproximaciones a la filosofía (Approaches to Philosophy)*.

Cullen, C.A. (1978) *Fenomenología de la crisis moral: sabiduría de la experiencia de los pueblos*, Buenes Aires: Ediciones Castañeda.

Derisi, O.N. (1968) *El Último Heidegger; approximaciones y diferencias entre la fenomenología existencial de M. Heidegger y la ontología de Santo Tomás*, Buenos Aires: Editorial Universitaria de Buebos Aires.

Fragata, J. (1959) *A fenomenologia de Husserl como fundamento da filosofia (The Phenomenology of Husserl as the Foundation of Philosophy)*, Braga: Livraria Cruz.(An attempt to establish Husserlian phenomenology as the basis of philosophy.)

Gallardo, H. (1993) *500 años: fenomenología del mestizo: violencia y resistencia (500 Years: Phenomenolgy of*

the Mestizo: Violence and Resistance), San José: Editorial Departamento Ecuménico de Investigaciones.(A search for a philosophy rooted in Latin American social problems.)

Guerrero, J.L. (1983) *Determinaciones de los valores morales: materiales para la constitución de una axiología general como fundamento de la ética axiológica (Determination of Moral Values: Materials for the Construction of a General Axiology as Basis of Axiological Ethics)*, 2nd edn, 1983.

Hernández de Ragoña, C.C. (1975) *Fenomenología histórica de las religiones (Historical Phenomenology of Religions)*, Mexico: Editorial Jus.(Interesting analysis of the phenomenology of religion.)

Llambías de Azevedo, J. (1940) *Eidética y aporética del derecho: prolegómenos a la filosofía del derecho (Eidetics and Aporetics of Law: Prolegomena to the Philosophy of Law)*, Buenos Aires: Espasa-Calpe. (Juridical axiology based on the phenomenology of Scheler and Hartmann.)

Maturo, G. (1989) *Fenomenología, creación y crítica: sujeto y mundo en la novela latinoamericana (Phenomenology, Creation and Criticism: Subject and World in the Latin American novel)*, Buenos Aires: F. García Cambeiro.(Phenomenology applied to Latin American literature and literary criticism.)

Maturo, G. (1991) *Imagen y expresión: hermeneútica y teoría literaria desde America Latina (Image and Expression: Hermeneutics and Literary Theory from Latin America)*, Buenos Aires: F. Garcia Cambeiro.(A literary theoretical approach to hermeneutics.)

Mayz Vallenilla, E. (1956) *Fenomenología del conocimiento: el problema de la constitución del objeto en la filosofía de Husserl (Phenomenology of Knowledge: the Problem of the Composition of the Object in the Philosophy of Husserl)*, Caracas: Facultad de Humanidades y Educación, Universidad Central de Venezuela. (Examines the works of Husserl.)

Mayz Vallenilla, E. (1959) *Realidad ontológica de un ser latino-americano (Ontological Reality of a Latin American Being)*.

Mayz Vallenilla, E. (1960) *Ontología del conocimiento (Towards a New Humanism)*, Caracas: Facultad de Humanidades y Educación, Universidad Central de Venezuela.(A new approach to humanism.)

Millas, J. (1943) *Idea de la individualidad (The Idea of Individuality)*, Santiago: Prensas de la Universidad de Chile.

Millas, J. (1949) *Goethe y el espíritu del Fausto (Goethe and the Spirit of Faust)*.

Millas, J. (1956) 'Ortega y la responsibilidad de la inteligencia' (Ortega and the Responsibility of Intelligence), Serie Negra, *Filosofía* 2, Santiago: Anales de la Universidad de Chile.(A discussion of Ortega y Gasset's work.)

Miró Quesada Cantuarias, F. (1941) *Sentido del movimiento fenomenológico (The Meaning of the Phenomenological Movement)*, Lima: Librería e imprenta D. Miranda.(A comprehensive study that includes the phenomenology of Husserl, Nicolai Hartmann and Heidegger.)

Naffah Neto, A. (1985) *Poder, vida e morte em situação de tortura: esboço de uma fenomenologia do terror (Power, Life and Death in a Situation of Torture: Draft of a Phenomenology of Terror)*, São Paulo: Hucitec. (Phenomenology applied to the historical situation of Brazil.)

Ortega y Gasset, J. (1947) *Meditaciones del Quijote: Obras Completas*, vol. 1, Madrid: Revista de Occidente, 6th edn, 1963, trans. E. Rugg and D. Marín, *Meditations on Quixote*, New York: W.W. Norton, 1961.(An example of a work by one of the founders of phenomenology.)

Pacini, D. (1969) *Crise filosofica do seculo actual: ensaio sobre a fenomenología do entendimiento do conhecimiento e do sentimento*, Rio de Janeiro: Ediciones Distribuidora Record.

Portilla, J. (1985) *Fenomenología del relajo y otros ensayos*, Mexico: Fondo de Cultura Económica.

Ramos, M.L. (1969) *Fenomenología da obra literaria*, Rio de Janeiro: Ediciones Forense.

Ramos, S. (1934) *El perfil del hombre y la cultura en México,* Mexico: Imprenta Mundial, trans. P.G. Earle, intro. T.B. Irving, *Profile of Man and Culture in Mexico*, Austin, TX: University of Texas Press, 1962.(Discusses humanity, beginning with the Mexican.)

Ramos, S. (1940) *Hacia un nuevo humanismo (Towards a New Humanism)*.

Ribeiro de Moura, C.A. (1989) *Critica da razao na fenomenología*, São Paulo: Ediciones Nova Estela.

Riega, A.T. de la (1973) *Conocimiento, violencia y culpa: la fenomenología frente al haber vital en la diferencia (Knowledge, violence and Guilt: Phenomenology in the Face of Life)*, Buenos Aires: Editorial Paidós.

Rivas Saínz, A. (1950) *Fenomenología de lo poético: notas de asedio (Phenomenology of the Poetic: Notes of Siege)*, Mexico: Tezontle.(Phenomenology of poetics.)

Rosario, C. (1965) *Primer encuentro (Phenomenology of Everyday Life: Puerto Rico becomes a Mass Society)*, San Juan: Ateneo Puertorriqueño.(Phenomenology is used to explain the change from an agricultural to an industrial society.)

Tamayo Herrera, J. (1963) *Fenomenología de la creación poética (Phenomenology of the Poetic Creation)*,

Lima: Empresa Editora América.(Phenomenology applied to the ability to create myths.)

Phenomenology of religion

The phenomenology of religion is a descriptive approach to the philosophy of religion. Instead of debating whether certain religious beliefs are true, it asks the question 'What is religion?' It seeks to deepen our understanding of the religious life by asking what (if anything) the phenomena we normally take to be religious have in common that distinguishes them from art, ethics, magic or science. Since the search for what is common presupposes difference and brings to light an astonishing array of divergent beliefs and practices, the quest for the essence of religion unfolds quite naturally into questions of typology 'What are the most illuminating ways of classifying religious differences?'

Sometimes the phenomenology of religion is motivated by a desire for quasi-scientific objectivity, combined with at least a soft scepticism about metaphysical speculation; if we cannot decisively resolve the metaphysical mysteries of life, people with this approach argue, at least we can give an unbiased description of those interpretations of the world we normally designate religious. At other times the phenomenology of religion has a more existential orientation: whether or not our arguments can settle questions about the ultimate shape of being, we have to choose our own mode of being-in-the-world; and if we are to decide intelligently whether or not to be religious, we need to be as clear as we can about what it means to be religious - ineluctable uncertainty may make faith something of a leap, but the leap need not be blind.

1 Descriptive philosophy of religion

Typically, the philosophy of religion is a normative enterprise, reflecting on the truth of religious beliefs. Thus proofs of the existence (and nature) of God are offered in the attempt to establish the truth of various theistic claims, while the existence of evil is offered as evidence against such claims. Both sides of this argument attempt to show that the other's arguments do not decisively establish the conclusion put forth, and the discussion often turns from the question of truth to the question of rationality. If God's existence cannot be decisively proved or disproved, is it rational to believe in God? But this, too, is a normative question.

By contrast, the phenomenology of religion brackets or temporarily sets aside such questions as to whether we are obliged or permitted or forbidden by the standards of truth and rationality to hold various religious beliefs. Instead, its point of departure is the fact that religion is an observable phenomenon of human life, and its task is to help us better understand what religion is by giving descriptive analyses of that aspect of human experience. In this regard it strongly resembles the philosophy of science and the philosophy of art, the tasks of which are neither to praise nor to bury science and art, but to give us deeper insight into the structures and functions of these widespread human activities. In all three cases, the observability of the phenomena is no guarantee of agreement when it comes to descriptive analysis.

By asking the question 'What is religion?', the phenomenology of religion expands the subject matter of the philosophy of religion in two ways. First, since religion is as much a matter of practice as of belief, it abandons the assumption that the philosophy of religion is primarily concerned with locating religious belief on some hierarchical scale of cognitive acts inspired by Plato's divided line. Instead, it focuses attention on religion as a language game that is a fully fledged form of life or mode of being-in-the-world.

Second, in seeking the essence, the common nature, or at least the family resemblances of a wide variety of phenomena usually taken to be religious, it requires a multicultural approach not usually found in the normative philosophy of religion, which often restricts itself to issues arising out of Jewish and Christian monotheism. Data for philosophical reflection emerge from work in comparative religion or the history of religions, such as Frazer's *The Golden Bough* or Eliade's *Patterns in Comparative Religion*.

As the search for the essence or common core of religion encounters divergences, numerous and often sharp, among religious beliefs and practices, a new task arises: typology. Now the question 'What is religion?' is supplemented by the question 'What differences among religious phenomena are most fundamental, and what kinds of classification are most illuminating?' It often turns out that historically distinct traditions, such as Judaism, Christianity, Islam, Hinduism, Buddhism, Daoism and so forth, can be replaced for such purposes by categories that cut across these traditions. Among these categories are Eliade's distinction of cosmos from history,

James' distinction of once-born from twice-born and Bergson's distinction of static/closed from dynamic/open (see Eliade, M. §2; Bergson, H.-L. §8.)

2 Scientific and existential phenomenology of religion

Historically speaking, the phenomenology of religion is a post-Kantian affair, and one strand has Kant himself and Husserl as its mentors. From Kant (§8) comes scepticism regarding metaphysical disputes about the way the world really is, along with the notion that we can nevertheless set forth the basic structures of our experience of the world. By making the phenomenal rather than the noumenal our subject matter, we can retain the goal of a philosophy that is scientific in the sense of giving us objective knowledge about something. Husserl's ideal of philosophy as rigorous science seeks to preserve this moment in Kant from lapsing into naturalism or psychologism. At the same time, his accounts of transcending the natural attitude by moving from fact to essence and by bracketing questions about the reality of the world as distinct from the mode of its givenness in experience (the *epoche*) seek to give methodological rigour to the descriptive tasks of postmetaphysical philosophy (see Husserl, E.).

A quite different motivation underlies the existential orientation in the phenomenology of religion. This strand presupposes both the hermeneutical and the existential turns phenomenology was to take away from Husserl's conception of a rigorously scientific philosophy; and it replaces Kant and Husserl with Heidegger and Kierkegaard as mentors for the phenomenology of religion. The hermeneutical critique, building on the late work of Husserl, becomes fully explicit in Heidegger (§§2-3) and Gadamer (§§3-4). All our cognitions are interpretations whose own possibility presupposes our immersion in language, in culture and in history in such ways that the ideal of a purely scientific understanding, free from all particular perspectives and presuppositions, is a chimera. No phenomenology can have as its goal the preservation of scientific objectivism in a postmetaphysical context.

What then can the goal be? It is at this point that Kierkegaard (§§4-6) is able to provide an existential answer. Not only does reflection always emerge from within the concrete situatedness of human existence; it has as its goal the clarification of the possibilities that confront us so that we can choose responsibly what kind of lives to live, especially since we must choose without the epistemic guarantees we would like.

As a descriptive approach to the philosophy of religion, the phenomenology of religion is more than just an alternative to the normative mode. It is always, at least implicitly, a critique of the latter. In its scientific mode, it extends the Kantian judgment that theoretical reason cannot settle metaphysical disputes to reason in all its modes, and views traditional philosophy of religion as engaging in speculation that is unwarranted because undecidable. While the existential mode often shares this view, its critique grows primarily out of its appreciation of Pascal's distinction between the God of the philosophers and the God of Abraham, Isaac and Jacob. It fears that the debates over natural theology too quickly become a theoretician's luxury in a realm of abstraction that loses touch both with the living God and with the concrete concerns and inescapable choices of existing individuals.

Both kinds of phenomenology of religion assume a kind of Socratic ignorance, a denial that we can attain final certainty about ultimate reality; and both place the Socratic question 'What is religion?' in this context. Sharing the Socratic preoccupation with the question about the way we should live, the existential mode transforms the objective question 'What is religion?' into the subjective question 'What would it mean to be religious as distinct from being irreligious?'

These two modes of phenomenology of religion are ideal types. Some important descriptions of the religious experience proceed with hardly a pause for methodological reflection; others proceed on the basis of historical, psychological, sociological or even literary self-understandings; still others, with clear philosophical intent, describe their project in ways that do not correspond exactly to these descriptions of the scientific and existential modes. But these tendencies can be perceived, in varying degrees, across a vast body of literature, all of which can be read as contributing to this philosophical project. We read descriptions of the religious life phenomenologically, whatever their genre, when we look for meanings (or essences) rather than facts; when we view these meanings as living, as what happens rather than what happened; and, if we move to the existential mode, when we consider these meanings as live options, as possibilities for our own existence.

3 The sacred

Phenomenology is a kind of empiricism; it addresses itself to the modes of our experience. But this does not keep it from making the world (the object of experience) its theme in various ways. Similarly, the phenomenology of religion is not compelled to focus on the subjective side of religion (for example, faith) as distinct from its objective side (for example, God). Indeed, if anything, the tendency has been to focus precisely on the ‘object’ of religious experience.

However, this is doubly problematic. First, the concept of God or gods is not appropriate for such religious ‘objects’ as the Buddha nature, which is at once the being of all things and *anatta*, no-self or no-substance; the *mana* of the Melanesians, and its many dynamistic correlates; or even the spirits of animistic religion. Second, even in contexts that speak freely about God, the claim is widespread that the highest experience of God or Brahman or the Buddha nature and so forth is beyond the structure of subject-object experience. Hence the quotation marks when speaking of the ‘object’ of religion.

The terms ‘sacred’ and ‘holy’ have come to serve as the generic names for the ‘object’ of religion. The ways in which these notions are spelled out provide the first answer to the question ‘What is religion?’ or ‘What is the common feature in virtue of which we call this very diverse set of phenomena religious?’ Two classic accounts highlight possibilities for both convergence and divergence.

Geerardus van der Leeuw gives a triadic structure to his classic work *Religion in Essence and Manifestation: A Study in Phenomenology* (1933): the object of religion, the subject of religion, and object and subject in their reciprocal operation. The object of religion, which he calls the sacred, is above all else power. But this power is not just any agency. Van der Leeuw stresses its remoteness. However frequently one encounters it, it never becomes usual or familiar, but remains a ‘*highly exceptional* and *extremely dangerous "Other"*’ ([1933] 1963 (1): 24). Accordingly, encounters with the sacred are accompanied by amazement, fear, and especially awe.

Van der Leeuw stresses that in its ‘primitive’ - that is, dynamistic and animistic - forms, this notion of sacred power lacks two features that often accompany it in the religions of the world’s ‘high’ civilizations. The first of these is a principle of unity in terms of which it is possible to think of the universe as an ordered and integrated cosmos. The second is a moral link between power and something like justice so that ‘the ground of the world may be trusted’ ([1933] 1963 (1): 31). In its generic sense, then, the sacred power ‘remains merely dynamic, and not in the slightest degree ethical or "spiritual"’ ([1933] 1963 (1): 28).

Rudolf Otto’s earlier study, *The Idea of the Holy* (1917), challenges this latter claim. Seeking to elaborate on what Schleiermacher (§7) called the feeling of absolute dependence, he defines the holy or the numinous as the *mysterium tremendum et fascinans* (overwhelming and fascinating mystery). To speak of the holy as mysterious is to focus explicitly on that in the ‘object’ of religion which is non-rational or ineffable in the sense of exceeding our conceptual apprehension. To call it the *mysterium tremendum* is to accentuate the awe-fulness of the holy, its ability to evoke fear and dread. This is the aspect of the sacred that allows it to be designated the ‘Wholly Other’. Finally, to describe it as *fascinans* is to find it to be ‘uniquely attractive and *fascinating*’ ([1917] 1958: 31). If the wrath of God is an expression of tremendousness, the mercy of God is an expression of the fascinating aspect of the holy. We are naturally ambivalent before the holy, drawn to it and repelled by it at the same time (see Otto, R.).

But for Otto, this complex ‘object’ of experience is not just a matter of power but also of value. Before the numinous I experience not only the limits of my power but also the limits of my worth. The notions of sin and defilement that emerge go beyond the notions of guilt and remorse derived from morality, just as holiness is not reducible to goodness; however, the value categories of religion are intimately related to those of morality. Over against van der Leeuw, Otto leaves us with two important and probably interrelated questions: Is the dialectical tension between attraction and repulsion generic to religion, or only a feature of some species? And is there an essential link between religion and morality, or is the supreme and unique value of the holy a specific rather than a generic feature of religion?

One approach to this latter question involves the sharp distinction between religion and magic. If the sacred is a matter of value-free power, then it entails no normative restraints against making it a means to human ends. The only barriers are technical, and the sacred power is on a par with atomic power: can we figure out safe and reliable procedures for putting this power at our own disposal? On Otto’s view that the holy is a category of value, we

should say at this point that we are no longer dealing with religion but with magic. On van der Leeuw’s view, by contrast, we should speak here of instrumental religion, recognizing that phenomena we normally designate as religious sometimes have this means-end structure constrained by no values but our own purposes.

See also: Existentialist theology; Mysticism, nature of; Religious experience

MEROLD WESTPHAL

References and further reading

Bergson, H.-L. (1932) *Les Deux Sources de la morale et de la religion*, Paris: F. Alcan; trans. R.A. Audra and C. Brereton, *The Two Sources of Morality and Religion*, Garden City, NY: Doubleday & Company, 1935.(A dyadic typology.)

Eliade, M. (1949) *Le mythe de l’éternel retour: archétypes et répétition*, Paris: Gallimard; trans. W.R. Trask, *The Myth of the Eternal Return*, Princeton, NJ: Princeton University Press, 1954.(A dyadic typology.)

Eliade, M. (1949) *Traité d’histoire des religions*, Paris: Payot; trans. R. Sheed, *Patterns in Comparative Religion*, New York: New American Library, 1958.(A rich, thematic study of a wide variety of phenomena.)

Frazer, Sir J.G. (1922) *The Golden Bough: A Study in Magic and Religion*, New York: Macmillan, abridged edn, 1960.(An encyclopedic source of materials.)

James, W. (1902) *The Varieties of Religious Experience*, Cambridge, MA: Harvard University Press, 1985.(A classic study that includes a dyadic typology along with studies of conversion, saintliness and mysticism.)

Leeuw, G. van der (1933) *Phänomenologie der Religion*, Tübingen: J.C.B. Mohr (Paul Siebeck); trans. J.E. Turner, *Religion in Essence and Manifestation: A Study in Phenomenology*, New York: Harper & Row, 1963. (First major study to call itself a phenomenology of religion.)

Otto, R. (1917) *Das Heilige*, Gotha: L. Klotz; trans. J.W. Harvey, *The Idea of the Holy*, New York: Oxford University Press, 1958.(Classic source of descriptions of the holy as ‘Wholly Other’ and as *mysterium tremendum et fascinans*.)

Ricoeur, P. (1960) *La symbolique du mal*, Paris, Aubier; trans. E. Buchanan, *The Symbolism of Evil*, New York: Harper & Row, 1967.(Rich study of the symbolic and mythological dimensions of religion, with a fourfold typology.)

Westphal, M. (1984) *God, Guilt, and Death: An Existential Phenomenology of Religion*, Bloomington, IN: Indiana University Press.(A comprehensive, introductory study that includes discussion of methodology and the essence of religion along with a triadic typology.)

Phenomenology, epistemic issues in

Phenomenology is not a unified doctrine. Its main proponents - Husserl, Heidegger, Sartre and Merleau-Ponty - interpret it differently. However, it is possible to present a broad characterization of what they share. Phenomenology is a method of philosophical investigation which results in a radical ontological revision of Cartesian Dualism. It has implications for epistemology: the claim is that, when the foundations of empirical knowledge in perception and action are properly characterized, traditional forms of scepticism and standard attempts to justify knowledge are undermined.

Phenomenological method purports to be descriptive and presuppositionless. First one adopts a reflective attitude towards one's experience of the world by putting aside assumptions about the world's existence and character. Second, one seeks to describe particular, concrete phenomena. Phenomena are not contents *of the mind; they all involve an experiencing subject and an experienced object. Phenomenological description aims to make explicit essential features implicit in the 'lived-world' - the world as we act in it prior to any theorizing about it. The phenomenological method reveals that practical knowledge is prior to propositional knowledge - knowing* that *arises from knowing* how.

The key thesis of phenomenology, drawn from Brentano, is that consciousness is intentional, that is, directed onto objects. Phenomenologists interpret this to mean that subjects and objects are essentially interrelated, a fact which any adequate account of subjects and objects must preserve. Phenomenological accounts of subjects emphasize action and the body; accounts of objects emphasize the significance they have for us.

The aim to be presuppositionless involves scrutinizing scientific and philosophical theories (Galileo, Locke and Kant are especially challenged). Phenomenology launches a radical critique of modern philosophy as overinfluenced by the findings of the natural sciences. In particular, epistemology has adopted from science its characterization of the basic data of experience.

The influence of phenomenology on the analytic tradition has been negligible. The influence on the Continental tradition has been greater. The phenomenological critique of modern science and philosophy has influenced postmodern thought which interprets the modernist worldview as having the status of master narrative rather than truth. Postmodern thought also criticizes the positive phenomenological claim that there are essential features of the lived-world.

1 Phenomenological reduction

The first phenomenological move is the phenomenological reduction, also called by Husserl 'bracketing' or 'the epoché'. The move involves distancing oneself from one's everyday 'immersion' in the ordinary practical activities of life, adopting a reflective standpoint upon one's experience of the world. This is taken to be the necessary standpoint from which to engage in genuine philosophical enquiry, one which phenomenologists criticize other philosophers for failing to adopt. The philosophical standpoint is radically different from the 'natural' attitude of common sense and of scientific enquiry. The crucial difference is that, in the natural attitude, one assumes unquestioningly that the world exists. The philosophical attitude, in contrast, puts aside - brackets - this assumption.

The philosophical standpoint after the reduction is differently characterized by transcendental phenomenology (Husserl) and existential phenomenology (Heidegger, Sartre, Merleau-Ponty) (see Phenomenological movement). Husserl believed that it was a 'transcendental' standpoint, 'outside' the natural world. Existential phenomenologists believed that such a standpoint is neither attainable nor necessary. The human standpoint is essentially *in* the world. The reduction is only partial: one cannot put aside all one's existential assumptions at the same time. However reflective, the philosopher still takes for granted the existence of the armchair.

This first phenomenological move can appear puzzling: What is this reflective standpoint? Why is it essential for philosophical investigation? How can one put aside one's belief that the world exists? It is important to note, however, that there are clear similarities between phenomenological reduction and certain familiar, traditional moves made in articulating scepticism about the external world. The sceptic, in challenging the claim that we can know that the external world exists, is clearly requiring us to suspend our everyday judgment that it does (see

Scepticism).

The point of the phenomenological reduction is to bring to our attention the realm of phenomena, the ‘lived-world’, the world as we experience it. What exactly are phenomena? What is normally wanted in response to this question is an account of how phenomena relate to objects in the external world: are they real features of these objects or are they merely features of experiencing subjects? It is important to see that this question is supposed to have been ruled out by the phenomenological reduction. We are no longer assuming that there *is* an external world. Of course, following the reduction, there still *appears* to be a world. One aim of phenomenology is to explore how the world appears with a view to finding the experiential basis for, and meaning of, our belief that the world exists independently of our experience of it. Phenomena are our experiences of the world; we must, at this stage, remain neutral concerning the ontological status of that ‘world’.

2 Presuppositionless description

The second phenomenological move is to describe phenomena without presuppositions. This is a difficult instruction to follow. Description normally has some purpose; what is the phenomenological purpose? One central phenomenological purpose is to expose and avoid certain philosophical presuppositions. One such presupposition involves the notion of sensation which has had a central role in philosophy since Descartes, and in particular the idea that sensations are the basic atoms of experience. This thesis, it is claimed, has blinded philosophers to the true character of experience; phenomenological description can avoid this theoretical assumption.

The first phenomenological discovery, following the reduction, is that ‘nothing has changed’. The world of objects does not disappear or dissolve into an airy realm of mere contents of consciousness. Nor does one’s experience transform into bare colour patches, uninterpreted sounds and so on. To characterize experience in this way is a ‘mistake’ which traditional epistemology has made.

A natural objection to raise at this stage is that no-one claims that uninterpreted sense data are what we are conscious of: they are rather to be construed as the unconscious building blocks of conscious experience. The phenomenological reply to that criticism is to ask what, if not conscious experience, is the basis for the philosophical belief in sensations?

The diagnosis is as follows: modern philosophy has taken its account of the physical world from physics which focuses exclusively on the primary qualities of objects. Philosophers, notably Locke, set themselves the task of reconciling this account of the physical world with human experience (see Locke, J. §2). The thesis is that objects as construed by physics have powers to affect experiencing subjects. The notion of sensation as the basic data of perception is derived, not from experience, but from scientific findings, in particular scientific accounts of human sense organs: given what our sense organs are like, this is what the basic sensory data *must* be like. Sensations are theoretical entities of psychology. Philosophical accounts of perception and empirical knowledge, empiricist and rationalist alike, have endorsed this same notion of sensation.

The phenomenological criticisms of these philosophical theses are, first, that while theoretical entities may have a place in the natural sciences, they have no legitimate place in philosophical enquiry. Physics offers a highly abstract conception of objects. This has been extremely successful in making predictions about the world and facilitating manipulations of it; but philosophy and science have different aims. Philosophy should not take scientific realism for granted. The thesis that objects are properly defined in terms of their primary properties is another philosophical thesis which phenomenological description should not presuppose. Second, sensations are theoretical entities of a school of psychology which fails on its own terms to give an adequate account of perception. But, even if the notion of sensation had been a successful one in scientific psychology, that would be no reason for philosophy to adopt it. It is not just the notion of sensation but also the psychologists’ conception of the mind as having contents which is challenged by phenomenology. This conception of the mind is another assumption which phenomenological description needs to avoid. Third, the notion of sensation has given rise to problems in epistemology which are insoluble because that very notion of sensation as mental contents caused by external objects always allows the sceptical challenge to be put, since there is no way of testing the causal claim. That the relation between experience and objects is a causal, contingent one is another presupposition which phenomenological description is to avoid. Moreover, even if knowledge of the external world could be justified by being shown to be based on sensations, this would not justify *our* claims to knowledge since sensations are not the

basic data of our perceptions. Nothing in the phenomena supports this notion of sensation.

According to phenomenology, the philosophical thesis that sensations are the basic units of experience is not only false; it is indicative of a misconception of the philosophical enterprise. It is not the task of philosophy to seek to reconcile the findings of science with experience, but to explore the experiential basis of beliefs, including scientific ones. Phenomenological description aims to reveal, not the constitutive parts - the 'atoms' - of a phenomenon, but the greater whole, the 'synthesis' which makes it the phenomenon it is. Careful scrutiny from the reflective standpoint, can reveal this 'synthesis' because all its details are 'implicit' in the phenomenon. In the natural attitude, one notices only the explicit features.

How does one set about phenomenological description? First one selects a philosophically significant topic, a 'theme', for example, perceiving physical objects. Next, one selects a phenomenon - for example, seeing a table. What is implicit in this phenomenon? What are the surrounding features which make it the phenomenon it is? Vision is explicitly involved. Vision and other senses are implicitly involved: if we see something as a table, we have a range of expectations concerning how it would appear from other angles, to the touch, to hearing, and so on. Support for the claim that these expectations are implicit in the phenomenon lies in the fact that we would be surprised if they were not realized in the appropriate circumstances. If we explore these expectations further, we find that implicit in seeing the table are certain motor skills: we could move around the table, bump into it and so on. Also involved in perceiving the table is that one knows what sorts of uses tables are put to, which typically involves having engaged in those practices. The significance of the table as something to be sat at, worked on or eaten from is implicit in seeing it and refers back to practices of sitting, working and eating. These are all features of the perceiving subject.

Phenomenological description also reveals implicit features of the object perceived, the wider 'synthesis' surrounding the way the table is involved in this phenomenon. The table appears as figure against a ground. The background contributes to the phenomenon - one would miss it if it were not there; but it is not the focus of one's attention. The table presents itself as a unified object, persisting unchanging as one walks around it. The table looks heavy, smooth, dusty, bare, uncluttered - colours and shapes are not the only visible features. The table appears as an artefact, manufactured from raw materials by someone using other artefacts; it was made for someone's use, in a certain sort of social setting. These are the sorts of features which phenomenological description reveals.

3 Uncovering essence

Such descriptions of specific 'concrete' phenomena, it is claimed, bring to light the richness and variety of the lived-world in contrast to the abstractions in terms of which scientific theory represents the world. The philosophical significance is twofold. First, such descriptions, it is claimed, have a critical role: they show that the theoretical presuppositions which were originally set aside are false: the descriptions serve as counterexamples to any philosophical theory which endorses the presuppositions. Second, they have a positive role: by describing particular phenomena, one can uncover essences, essential features of the lived-world.

One essential feature of phenomena is the intentionality of consciousness: consciousness is directed upon objects (see Intentionality). The phenomenological interpretation of this, what attention to the phenomena reveals, is that the relation between consciousness and its objects is an internal one. As such, it is not amenable to causal explanation. All consciousness is essentially consciousness of objects and all objects are essentially, if not explicitly, objects of consciousness. Idealist and realist claims to the contrary ignore this essential feature implicit in all phenomena.

An essential feature of perception is that it presupposes practical skills. All perceptual knowledge essentially depends on knowing how. Propositional knowledge is always 'situated' in practice. Philosophical accounts of the foundations of knowledge have often ignored this essentially practical foundation.

An essential feature of objects is their practical significance for us: they resist or assist our efforts to use them and live with them. Objects of theoretical knowledge are essentially also objects of our concern. Philosophical accounts of the character of objects in the external world have ignored this essential basis of theory in practice. Further, objects of the lived-world are not discrete entities, standing in causal, mechanical, contingent relations; they have internal relations with each other. Objects are essentially parts of complex wholes.

This move from description to essence looks problematic since it appears to involve a move from particular claims to universal and necessary ones. Phenomenologists differ in the way they purport to uncover essences. For transcendental phenomenology, essences can be detected only from the reflective standpoint, the standpoint of the transcendental ego. Essences have their origin in the transcendental ego: experience is structured as it is because the ego is structured as it is. Husserl believed that transcendental phenomenology leads to transcendental idealism (see Idealism §§3, 7). For existential phenomenology, there is no such transcendental ego, no transcendental grounding of the essential features of experience. The reflecting subject, the subject that is practising phenomenology and finding essential features of the lived world, is still an existential subject, living in the world.

As a test for what is essential, Husserl uses the criterion that we cannot imagine it otherwise. So, for example, we cannot imagine a perception of a physical object which did not involve ‘horizons’, expectations concerning other possible perceptions of it; we cannot imagine believing that a solid physical object exists without also knowing that there are possible circumstances - waking from a dream, its fading into thin air - which would show that belief to be false.

Merleau-Ponty looks for essential features of the perception of objects in natural capacities which are temporally prior to perceiving objects. He appeals to psychological data (drawn especially from Gestalt psychology) concerning, for example, the way babies focus their eyes on objects of interest and comfort while quite unable to focus on other less significant objects in their visual field (see Gestalt psychology). Such examples, he claims, show that natural responses to significant surroundings are temporally prior to the perception of objects. From this temporal priority, Merleau-Ponty believes, follows a logical priority: there are necessary prerequisites of learning or perceiving, features implicit in experience without which we would not have the experience we do.

Husserl and Merleau-Ponty believed that, by uncovering essences, phenomenology could avert the ‘crisis’ threatened by the prevalence of positivism. Heidegger and Sartre introduce an explicitly moral dimension into the search for essences. Both believe that one can reveal essences by stripping away all that is inauthentic in one’s dealings with the world. For Sartre, the search will reveal that all essence is dependent on human interests, human projections of non-being onto the world of Being; for Heidegger the end point of the search will be that Being will simply reveal itself as that with which we are ‘primordially’ connected (see Being §4). Analytic philosophers have problems with that ‘Being’. A sympathetic reading of these phenomenologists would be that they are seeking to characterize, not a state which philosophical enquiry enables one to achieve, but rather the process of philosophical enquiry. One may never reach the point at which all one’s beliefs are true, all one’s values as they ought to be. There may be no sense to attach to such a possibility, nothing which would count as recognizing it to obtain. One can, none the less, strive against false beliefs and bogus values. On this reading, Heidegger and Sartre may appear misguided in their attempts to say the unsayable; but we can see them as putting forward recognizable philosophical ideals.

4 Phenomenological ontology

Phenomenology shares with Cartesian dualism the belief that subjects and objects are radically different kinds of thing. They are not, however, logically distinct, nor are they the kinds of thing Descartes took them to be (see Descartes, R. §8). Subjects are essentially situated in the world, and the world of objects is essentially our home. Philosophical accounts of conscious subjects and objects in the world must never lose sight of this essential interrelatedness. The transcendental ego might seem to constitute an exception to this; but the claim of transcendental phenomenology is that the transcendental standpoint is just one standpoint which the ego can take and, from that standpoint, it can reflect upon itself as an empirical ego essentially situated in the world.

The phenomenological account of the conscious subject is of a viewpoint on the world, an ‘openness’ to the world. Further, since perceiving the world essentially involves acting in and on the world, the conscious subject is essentially an active and, so, embodied subject. Our bodies are not mechanisms in which consciousness happens to reside; they are the possessors of habits and skills without which action - and therefore perception - would not be possible. All phenomenologists emphasize the importance of the body, but Merleau-Ponty produces the most systematic account. Bodily skills and habits - what he terms ‘bodily intentionality’ - are in many respects exercized ‘pre-consciously’: conscious attention to habitual engagement with one’s familiar surroundings can, notoriously, disrupt performance. Subjectivity is essentially bodily. The subject of consciousness depends on the

‘body-subject’. The body, with its skills and habits, cannot operate except on appropriate instruments, in a suitable habitat, in a world of objects amenable to its skills - miming is a special art.

The phenomenological account of objects is that they present themselves as objects of our gaze, to be wondered at. We perceive them, not as bearers of purely factual primary and secondary properties, but as having all manner of significance, ambiguities and indeterminacies. Sartre writes with approval that Husserl has restored to objects their horrors and their charm. For phenomenology, these - usually deemed at best tertiary properties of objects - are the properties which make them possible objects for us, that without which we could never have discovered their primary and secondary properties.

Underlying the perceptual properties of objects are their practicality: they are, in Heidegger’s terms, ‘ready-to-hand’ prior to being ‘present-at-hand’. As such, they presuppose wider wholes which give the specific instruments a meaningful context. Ultimately, the context is the whole world, and it is a communal world - an essential backdrop for all subjects.

In summary, subjects are essentially bodily, with skills and habits which require instruments and habitats to develop and come into play. The intentionality of consciousness depends on bodily intentionality. Objects compel, invite or resist our actions. Our ability to view them with a neutral scientific gaze involves our ignoring the very features which enable them to be objects for us.

Modern philosophy, empiricist and rationalist, it is claimed, has ignored these fundamental interactions which are presupposed by any propositional knowledge of the world. Deprived of this practical underpinning, propositional knowledge is open to sceptical challenge in a way in which practical knowledge, knowing how to do things, is not. Any notion of a pure conscious subject or an independently existing object is the result of a process of abstraction from the lived-world. To proceed to worry about how the one can interact with or have knowledge of the other is to forget the starting point in practice from which the process of abstraction proceeds, a starting point which, according to phenomenology, is implicit in all phenomena and can be read off from them from a suitably reflective standpoint.

A natural objection to raise against phenomenology is that it gives an account only of intentional objects, objects of consciousness, and not of ‘real’ objects in the external world. The phenomenological rejoinder to this is firstly, that intentional objects are not to be construed as mental contents, inner effects or representations of an external world. Second, phenomenology would press the objector to articulate the distinction between intentional and ‘real’ object. Phenomenology explores the basis in experience of claims concerning the existence of ‘real’ objects; any account of ‘real’ objects which does not so relate them to objects of experience is strictly nonsense. On the phenomenological account of our relations with the external world, we find that our experience is mostly just as it should be to support our claims that the world exists and that we know that it does.

See also: Brentano, F.C.; Consciousness; Introspection, psychology of; Knowledge, Tacit Perception, epistemic issues in

JANE HOWARTH

References and further reading

Brentano, F. (1874) *Psychologie vom empirischen Standpunkt*, trans. A.C. Rancurello, D.B. Terrell and L.L.McAlister as *Psychology from an Empirical Standpoint*, London: Routledge & Kegan Paul, 1973. (Complex discussion of psychology. This translation is reliable and readable.)

Cooper, D.E. (1996) *Heidegger*, London: The Claridge Press.(Short and highly readable introduction to the work of Heidegger. Chapter 3 is especially relevant.)

Dreyfus, H.L. (1979) *What Computers Can’t Do: The Limits of Artificial Intelligence*, New York: Harper & Row. (Phenomenological critique of artificial intelligence.)

Dreyfus, H.L. (1982) *Husserl, Intentionality and Cognitive Science*, Cambridge, MA: MIT Press. (Applications of phenomenology to recent developments in philosophy.)

Dreyfus, H.L. (1991) *Being-in-the-World: A Commentary on Heidegger’s ‘Being and Time’, Division I,* Cambridge, MA: MIT Press.(Clear and helpful commentary.)

Hammond, M., Howarth, J. and Keat, R. (1991) *Understanding Phenomenology*, Oxford: Blackwell.(Clear and readable introduction to phenomenology through analyses of writings by Husserl, Merleau-Ponty and Sartre.)

Heidegger, M. (1927) *Sein und Zeit*, trans. J. Macquarrie and E. Robinson as *Being and Time*, Oxford: Blackwell, 1962.(Difficult and often obscure. The translation is conscientious, with helpful notes and index.)

Heidegger, M. (1993) *Basic Writings*, trans. D. Farrell Krell, London: Routledge.(Selection of Heidegger's works from 1927 until his death. The later works promise a depth which is challenging if hard to fathom. Contains a general introduction, and brief introductions to each section.)

Husserl, E. (1931) *Cartesianische Meditationem*, trans. D. Cairns as *Cartesian Meditations*, The Hague: Martinus Nijhoff, 1977.(Short but demanding text outlining the phenomenological method, with many technicalities and few examples. The translation is mainly faithful, with many notes.)

Husserl, E. (1936) *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*, trans. D. Carr as *The Crisis of European Sciences and Transcendental Phenomenology*, Evanston, IL: Northwestern University Press, 1970. (A longer and more readable work than *Cartesian Meditations*. The translation contains a helpful introduction and notes.)

Langer, M.M. (1989) *Merleau-Ponty's 'Phenomenology of Perception': A Guide and Commentary*, London: Macmillan.(Detailed exposition of Merleau-Ponty's text.)

Merleau-Ponty, M. (1945) *Phénoménologie de la Perception*, trans. C. Smith as *Phenomenology of Perception*, London: Routledge & Kegan Paul, 1981.(Combines detailed criticism with engaging description. The translation replaces a detailed and useful table of contents with an index.)

Sartre, J.-P. (1943) *L'Être et le Néant*, trans. H. Barnes as *Being and Nothingness*, London: Methuen, 1958. (Combines at times tortuous argument with flowing and peceptive descriptions. This translation has been criticized for inaccuracies, but the introduction by Mary Warnock is helpful as, in parts, is the Key to Special Terminology.)

Philip the Chancellor (1160/85-1236)

Philip occupies a pivotal place in the development of medieval philosophy. He is among the very first in the Latin West to have a fairly complete picture of both the newly available natural philosophy and metaphysics of Aristotle and the work of the great Muslim thinkers, Avicenna and Averroes. His Summa de bono, composed sometime between 1225 and 1236, shows the broadening of philosophical interests and the growth of philosophical sophistication that accompanied reflection on these new materials. Philip's Summa had a major impact on subsequent thirteenth-century thinkers, particularly Albert the Great, whose own Summa de bono is closely modelled on that of Philip.

Philip the Chancellor (Philippus Cancellarius) studied at Paris and became Master of Theology there. He was named chancellor of Notre Dame in 1217 and held that position until his death. As the chancellor, Philip was charged with supervising teaching at the University of Paris, and in that capacity he was a key player in several of the university's formative crises. In addition to *Summa de bono*, his works include various (unpublished) *quaestiones*, some four hundred sermons (still unpublished), a homiletics manual on the Psalms and some highly regarded poetry.

Philip's *Summa de bono* is a massive summary of theological and philosophical matters, roughly similar in size to Peter Lombard's *Sentences* and William of Auxerre's *Summa aurea*. However, although Philip's project is in the tradition of these earlier works, it is strikingly different. He takes the notion of the good (*bonum*) as the point of departure and organizing principle for his *Summa*. After a brief introductory discussion which includes three questions on the metaphysics of goodness and four on the highest good and the way created good flows from it, he devotes the entire work to created goods. He offers no extended treatment of God, the Trinity, Incarnation or the sacraments, topics that dominate Lombard's *Sentences* (see Lombard, P.)

The main parts of Philip's *Summa* correspond roughly to the divisions marked by his theoretical classification of the main kinds of created good. The first part is devoted to the good creatures possess by virtue of their nature (*bonum naturae*). Under this heading he discusses the nature of purely intellectual creatures (angels), purely corporeal natures (things created on the first five days of creation), and creatures composed of both intellectual and corporeal natures (human beings). The short second part is devoted to the good possessed by actions by virtue of their type (*bonum in genere*). The third and last part deals with the good that belongs to a creature because of a perfection it has by virtue of both its own nature and activity and the activity of God. Philip calls this the good associated with grace (*bonum gratiae*) and includes under it the topics of grace and virtue. In this third part of *Summa de bono*, Philip gives only passing attention to questions having to do with grace, devoting most of it (roughly half the entire *Summa*) to a detailed treatment of the theological and moral virtues. This part of the *Summa* significantly advances the project, begun by William of Auxerre, of elaborating a comprehensive, systematic moral theory.

Philip's classification of created goods identifies two other sorts of created good which do not figure in the plan of the *Summa*: the good that belongs to an action because of features of it less general than its type - namely, its end and circumstances (*bonum a fine, bonum a circumstantia*) - and the good associated with a perfection that is a pure gift of God - namely, glory (*bonum gloriae*). It is unclear whether Philip intended to include discussions of these goods and, hence, whether he left the *Summa* unfinished. *Summa de bono* is best known for its opening questions, which may constitute the first systematic treatise on the transcendentals. In his prologue, Philip proposes to begin the *Summa* with an inquiry into the relation of good (*bonum*) to being, one and true (*ens*, *unum*, *verum*) because these are the most general principles and because ignorance of the nature of principles can lead one to shipwreck in matters of faith. Philip's position in these questions is seminal. Drawing on suggestions from the newly available Muslim tradition of commentary on Aristotle and reflecting a new, sophisticated familiarity with the concepts of Aristotelian metaphysics, Philip's account centres around the thesis that the transcendentals are extensionally equivalent, although they are intensionally or conceptually distinct. The exposition and defence of this thesis is the defining characteristic of the thirteenth-century discussion of the transcendentals, which includes the work of Alexander of Hales, Albert the Great, Aquinas and Bonaventure.

See also: Albert the Great; Aristotelianism, medieval; William of Auxerre

SCOTT MacDONALD

List of works

Philip the Chancellor (1225-36) *Summa de bono*, ed. N. Wicki, *Philippi Cancellarii Summa de bono*, Berne: Francke.(The traditional dating of *Summa de bono* places it in the period from 1228 to 1236, but Wicki argues for an earlier date, 1225-8.)

References and further reading

MacDonald, S. (1992) 'Goodness as Transcendental: The Early Thirteenth-Century Recovery of an Aristotelian Idea', *Topoi* 11: 173-86.(An examination of *Summa de bono*'s opening questions on the metaphysics of goodness and their place in the development of thirteenth-century thought.)

Wicki, N. (1985) *Philippi Cancellarii Summa de bono* vol. 1, Berne: Francke, especially ch. 1, 'Vie de Philippe le Chancelier' and ch. 3, 'Données de la tradition manuscrite et problèmes d'histoire littéraire'.(Succinct, authoritative account of Philip's life and work and a useful survey of the important secondary literature.)

Philo of Alexandria (*c.*15 BC-*c.* AD 50)

Philo of Alexandria is the leading representative of Hellenistic-Jewish thought. Despite an unwavering loyalty to the religious and cultural traditions of his Jewish community, he was also strongly attracted to Greek philosophy, in which he received a thorough training. His copious writings - in Greek - are primarily exegetical, expounding the books of Moses. This reflects his apologetic strategy of presenting the Jewish lawgiverMoses as the sage and philosopher par excellence, *recipient of divine inspiration, but not at the expense of his human rational faculties. In his commentaries Philo makes extensive use of the allegorical method earlier developed by the Stoics. Of contemporary philosophical movements, Philo is most strongly attracted to Platonism. His method is basically eclectic, but with a clear rationale focused on the figure of Moses.*

Philo's thought is strongly theocentric. God is conceived in terms of being. God's essence is unreachable for human knowledge (negative theology), but his existence should be patent to all (natural theology). Knowledge of God is attained through his powers and, above all, through his Logos *('Word' or 'Reason'), by means of which he stands in relation to what comes after him. In his doctrine of creation Philo leans heavily on Platonist conceptions drawn from reflection on Plato's Timaeus. The conception of a creation* ex nihilo *('from nothing') is not yet consciously worked out. Philo's doctrine of human nature favours the two anthropological texts in Genesis 1-2, interpreting creation 'according to the image' in relation to the human intellect. With regard to ethics, both Stoic concepts and peculiarly Jewish themes emerge in Philo's beliefs. Ethical ideals are prominent in the allegorical interpretation of the biblical patriarchs.*

Philo's influence was almost totally confined to the Christian tradition, which preserved his writings. He was unknown to medieval Jewish thinkers such as Maimonides.

1 Life and works

The thought of Philo of Alexandria is very much the product of his combined Alexandrian and Jewish background. Biographical details are very scarce, but we know that he was born into a prominent and extremely wealthy Jewish family (Josephus, *Antiquities* XVIII 258). In AD 39 he was chosen as leader of a delegation that sailed to Rome to protest to the emperor Gaius Caligula against the pogrom suffered by the Jewish community of Alexandria at the hands of the local populace. The interview with the emperor went badly, but did not cost him his life (see the unintentionally hilarious report in *Embassy to Gaius* 349-67). This incident is often connected with a passage at the beginning of *On the Special Laws*, where Philo complains that involvement in political affairs distracts him from more serious pursuits, and looks back wistfully to the time when he could fully devote himself to the life of philosophy.

From his writings it is apparent that Philo received a thorough training in Greek philosophy, although we have no idea from whom he received it (it has been suggested he may have had house tutors). His knowledge of the Platonic and Stoic traditions is especially thorough. His writings thus provide important, although imprecise, information on contemporary philosophical developments in Alexandria. Philo's extensive use of Plato's *Timaeus* and other dialogues reflects the beginnings of the Middle Platonist movement associated especially with the figure of Eudorus (see Platonism, Early and Middle §1). He is also the first writer to cite the so-called 'ten modes' of the Neo-Pyrrhonist Aenesidemus, who was active in Alexandria a generation before his birth.

Philo was a prolific writer, and the majority of his works appear to have survived. These amount to exactly fifty treatises in all. They can best be divided into three separate groups:

(1) There are five purely *philosophical* treatises, discussing subjects such as the eternity of the cosmos, divine providence, and the rationality of animals. These offer much valuable material on Greek philosophy, and it is only occasionally possible to discern that their author was Jewish.
(2) In addition there are four *historical/apologetic* works, in which Philo defends Judaism against contemporary attacks.
(3) The remaining forty-one works are all *commentaries* on the books of Moses.

These are best sub-divided into four categories:

(a) Two books on the life of Moses, introducing the writer of the holy books to a broad public, both Jewish and Greek.
(b) Ten books on the Exposition of the law, first describing the creation of the world (Genesis 1-3) and the lives of the patriarchs Abraham and Joseph, then expounding the Decalogue, the Special Laws and other subjects related to Mosaic legislation.
(c) The Allegorical commentary in twenty-one books, giving an astonishingly complex running commentary on Genesis 2-17, in which the chief exegetical method used is allegorization of the text.
(d) Eight books of Questions and answers on Genesis and Exodus, in which questions are posed on most verses of the biblical text, and both literal and allegorical answers are given in a consecutive but formally unconnected way.

The great majority of Philo’s writings thus concentrate on the five books of Moses (or Pentateuch), which he read in the Greek translation of the Septuagint. The direction of Philo’s loyalties is very clear. He wishes to defend the cultural and religious heritage of his people in their minority position in Alexandrian society. His chief apologetic strategy is to focus in on the figure of the great sage Moses, exploiting the considerable prestige accorded to barbarian sages such as Zoroaster and the Egyptian priests in contemporary Greek philosophy. Moses is the great sage and paradigm for righteous and holy living. His writings, if appropriately read, are an inexhaustible fount of wisdom.

It may seem, however, that Philo is facing an impossible task. Only a few passages in the Pentateuch, such as the creation account and the stories of Moses on the mountain, offer much scope for philosophical elucidation. His solution is to exploit the method of allegorical interpretation, invented and developed by earlier Greek philosophers (including Stoics) in order to defend the authority of Homer. Like Homer, Moses allegorized: that is to say, he said one thing but meant another. The meaning behind his words has to be uncovered. Moses tells the story of the patriarch Abraham (and Philo would never deny that this progenitor of the Jewish race existed), but the deeper meaning of the text reveals the story of the human soul on the path to perfection and felicity (see §6). Philo is one of the great masters of the allegorical method. In practice it gives him the scope he needs to introduce Greek philosophical doctrines and so develop his own philosophical views by means of a presentation of Mosaic thought.

2 Philosophical position

Understandably, Philo shows no loyalty to any particular school of Greek philosophy. None can match the prestige enjoyed by the school of Moses. From this viewpoint Philo’s method may be considered thoroughly eclectic, but with a clear rationale. It is apparent from his commentaries, however, that his philosophical sympathies lie with Platonism. He is particularly attracted to the revival of transcendentalism undertaken in the Middle Platonist movement. Stoic doctrines are found especially in the area of ethics (see §6). Other themes reflect the influence of Jewish thought, and so offer interesting contrasts with ideas in Greek philosophy.

3 Epistemology

Philo makes no attempt to present a systematic theory of knowledge, but there is at least one epistemological question that he cannot avoid: How was it possible for Moses to attain the pinnacle of human wisdom? Moses is regarded as a prophet, exalted beyond other human beings. Philo distinguishes between two types of prophecy, both of which have antecedents in Greek reflection on the subject. In the first kind, which enables the prophet to predict the future, the prophet is empowered to ‘stand outside’ himself and through divine possession become an instrument of the divine voice speaking through him. This may be called *ecstatic* prophecy. In the second type the prophet is also inspired by God, but remains in full possession of his rational abilities, which allows his mind to contemplate the nature of reality in its fullness. This may be termed *noetic* prophecy. The prophet is one who ‘through virtue rather than birth has advanced to the service of the truly Existent,…since he has within him a noetic sun and shadowless beams of light, which give him the clear apprehension of things invisible to sense but perceptible to the mind’ (*On the Special Laws* 4.192). Philo’s Platonist assumptions emerge clearly here, even if the term ‘apprehension’ (*katalēpsis*) has a Stoic background. It is above all this second type of prophecy, exemplified by Moses’ vision of the *paradeigma* (a Platonic technical term) of the tabernacle and its contents on Mount Sinai (*Exodus* 25: 9), that enables him to be the sage *par excellence* and the author of the sacred books.

4 Theology

As a result of his Jewish background Philo's thought is resolutely theocentric. God is conceptualized primarily in terms of true being. The link to Platonism is apparent, but for Philo the source is above all God's words to Moses in Exodus 3:14, 'I am he who IS'. It is typical of Philo that he alternates speaking of God in personal and impersonal terms: that is, between 'he who is' and 'that which is'.

Crucial to Philo's theology is the distinction between God's *existence* and his *essence*. Through observation and experience of the natural world and particularly of their own intellectual powers, human beings can without any difficulty conclude that God exists and that he is creator of the universe. However, gaining knowledge of God's essence is beyond the reach of human conceptual abilities. Not even the great Moses, although he made many requests, was granted this privilege. God is unknowable in his essence, and thus also unnameable, uncircumscribable and unutterable: that is, there is no name or description that can give accurate expression to his essential nature. Philo thus has a negative and a positive theology, rather similar to what one finds in the Middle Platonist handbook of Alcinous.

God is thus utterly transcendent. At the same time, however, he stands in close relation to the cosmos as its creator and provident maintainer (see §5). Philo argues that the two chief names for God refer to his powers (*dynameis*). The name *theos* (God) indicates his creative power (from the root *tithēmi*, 'I set in place'); the name *kyrios* (Lord) indicates his ruling power. More famously the figure of God's Logos (usually rendered 'Word', but 'Reason' is better) is invoked in order to explain God's relation to what is other than him. Philo's Logos doctrine is complex and certainly not always philosophically consistent. This is due to its double origin in both Jewish thought (especially the 'and God said' in Genesis 1) and Greek thought (especially Stoicism (§3)) (see Logos). The Logos has an immanent aspect, indicating the divine presence in the created realm. It also has a transcendent aspect, and is sometimes identified with the intelligible world. The chief difficulty posed by Philo's doctrine of the Logos is the following. Sometimes Philo speaks of the Logos as if it were simply an aspect of the divine nature, namely that aspect which is accessible to human thought precisely because it is related (in a model/image relation) to that which follows it. At other times the Logos is treated as a *hypostasis*, that is, a self-subsistent theological entity that is at least to some degree independent of God himself. As a Jew, however, Philo refuses to accept the Platonist solution of a hierarchy of divine principles at different ontological levels (see Neoplatonism §3).

5 Doctrine of creation

A second pillar of Philo's thought is the doctrine of creation. Philo is utterly convinced that visible material reality has been created by God, and has nothing but scorn for the minority opinion that it was the result of chance or spontaneous development. Furthermore, because God has created the universe, he will also take care of it through the action of divine providence. Philo agrees with Plato against the Stoics that, although the universe had a beginning, it will not be subject to total destruction. This, he argues, was already Moses' view, as shown by Genesis 8: 22 (see *On the Indestructibility of the Cosmos* 1-19).

In his treatise *On the Creation of the Cosmos According to Moses*, Philo presents his views on creation in greater detail. The cosmos was created in six days. This does not mean that God needed a length of time in which to complete his work. In fact, everything came into being simultaneously, because time commenced with the cosmos itself. The details of Philo's interpretation are strongly influenced by contemporary interpretations of Plato's *Timaeus*. Remarkably, the first day of creation is taken to describe the creation of the intelligible cosmos, which serves as model for the visible cosmos and is equated with the Logos (see §4). If one would wish to summarize the doctrine of creation, Philo writes, 'one might say that the noetic cosmos [of day one] is none other than the Logos of God while he [that is God] is engaged in his creative task' (*On the Creation of the Cosmos* 24). It is noteworthy that Philo does not dissociate God from the creative task and attribute it solely to his Logos, because this would surely endanger the conviction of a unique God.

Philo is less than clear on the origin and ontological status of matter, and especially on the question whether it is created by God or has an independent existence before creation. He does not discuss the question in his commentary on the creation account (Genesis 1: 2 is taken to refer to the intelligible world), and elsewhere he appears to vacillate. It is safest to conclude that he remains prisoner of the axiom almost universally present in Greek philosophy that nothing can come out of nothing, and so is unable to face the full consequences of a doctrine of creation *ex nihilo* such as was developed in early Christian thought (in reaction to Gnosticism).

Philo's cosmology follows the dominant Platonic-Aristotelian model. The cosmos is a work of great beauty and order, produced by a good and gracious creator. It is not autonomous, and certainly should not be worshipped. This was the mistake made - in terms of Philo's allegorical scheme - by the Chaldaeans, whom Abraham left behind in his search for the true God (*On the Life of Abraham* 68-71).

6 Doctrine of human nature, ethics

Philo's doctrine of human nature, the third pillar of his thought, is based primarily on the two texts in the creation account, Genesis 1: 26-27 and 2: 7, both of which are interpreted in terms of the Greek philosophical ideal of human reason. In the former text human beings, male and female, are created 'according to the image (*eikōn*) of God', which is taken to mean that they resemble God not in terms of the body or the lower soul with its passions, but through the rational soul or the mind, which is also their immortal part. In the latter text the human body is 'inbreathed' by God's spirit (*pneuma*). This creative act is also taken to refer to the formation of the human rational faculty. Following basically Platonist lines, Philo regards the human goal (*telos*) as 'assimilation (*homoiōsis*) unto God'. This can be accomplished because of the human image relation to God: that is, it can be accomplished through the powers of the intellect. It is in gaining knowledge of God that humans become like God. This was the nature of Moses' paradigmatic quest. What might seem to be the ultimate goal, however, the knowledge of God's essence, is unreachable, because then assimilation would become identity, which is impossible on account of the gulf separating creator and creature (an instructive contrast can be drawn with Plotinus' doctrine of the union with the One; see Plotinus §3).

In the formulation of his ethical ideals Philo extracts much from Stoicism, but places it in a basically Platonist framework. The journey of the soul involves various stages. It begins with the struggle against the passions resulting from association with the body. As learners advance, they develop the exercise of reason and embark on the path of the virtues. The goal to be attained is the life of perfection, the life lived by the wise person (*sophos*). The wise person is characterized by a freedom from all passion (*apatheia*), not in the sense of having no emotions whatsoever, but because irrational passions have been converted into rational emotional states (*eupatheiai*). However, the ideal of the wise person as represented by Moses is lofty and seldom attained. For many the patriarchs are more accessible symbols of the level a human being can attain. They represent three aspects of the quest for perfection. Abraham is the learner, Isaac is the man with natural aptitude for the quest, Jacob is the practiser who never yields in his struggle to reach the goal. The quest for perfection and the ideal of the wise person is the fourth and final pillar of Philo's thought.

Philo's ethics are not entirely Greek. In his treatment of the virtues Jewish themes can be detected. Philo is more positive towards the feelings of repentance and pity than we find in Greek ethics, and tends to regard piety as the greatest of the virtues. Foreign to Greek philosophical thought also is his emphasis on the nothingness (*oudeneia*) of human beings before God's face, a theme which anticipates the role of humility in Christian ethics.

7 Influence

Philo's influence on the course of philosophical thought was limited to the Christian tradition. The church fathers preserved his works because his method of using Greek philosophical themes to explain scripture appealed to them. Attempts by Wolfson (1947, 1968) and others to demonstrate Philo's influence on later Platonist thought have not been successful. Medieval Jewish philosophers did not know Philo, and Jewish interest in him did not revive until the sixteenth century. From the viewpoint of the history of ideas Philo is interesting above all because in his writings the traditions of Greek philosophy and Judaeo-Christian thought converge for the first time.

DAVID T. RUNIA

List of works

The names of Philo's fifty treatises cannot be included here (but see the sub-division given in §1). A list with the conventional English and Latin titles can be found at the beginning of Loeb Classical Library edition. The chronology of individual writings cannot be determined with any certainty.

Philo (20-50 AD) *Philo*, trans. F.H. Colson and G.H. Whitaker, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1929-62, 10 vols, plus 2 suppl. vols, trans. R. Marcus.(The standard English translation, with Greek text; badly in need of revision.)

Philo (20-50 AD) *The Works of Philo Complete and Unabridged*, London: Bohn's Classical Library, 1854-5; repr.,

trans. C.D. Yonge, Peabody, MA: Hendrickson, 1993.(A single volume edition with modern numbering; unreliable and incomplete.)

Philo (20-50 AD) *On Providence*, vols I and II in M. Hadas Lebel (ed.) *Les Œuvres de Philon d'Alexandrie*, vol. 35, *De providentia I et II*, Paris: Les Éditions du Cerf, 1973.(Survives only in Armenian; not included in the Loeb Classical Library edition.)

Philo (20-50 AD) *On Animals*, ed. A. Terian, *Philonis Alexandrini de Animalibus: The Armenian Text with an Introduction, Translation and Commentary*, Chico, CA: Scholars Press, 1981.(Survives only in Armenian; not included in the Loeb Classical Library edition.)

Philo (20-50 AD) *The Contemplative Life*, trans. D. Winston, *Philo of Alexandria: The Contemplative Life, The Giants and Selections*, New York and Toronto, Ont.: Paulist Press, 1981.(The best anthology of Philo's writings, containing many of his more interesting philosophical passages.)

Philo (20-50 AD) *The Giants*, trans. D. Winston, *Philo of Alexandria: The Contemplative Life, The Giants and Selections*, New York and Toronto, Ont.: Paulist Press, 1981.(The best anthology of Philo's writings, containing many of his more interesting philosophical passages.)

References and further reading

Baer, R.A. (1970) *Philo's Use of the Categories Male and Female*, Leiden: Brill.(Important work on the male-female relation in the interpretation of the creation account and the doctrine of human nature.)

Dillon, J.M. (1977) *The Middle Platonists: 80 B.C. to A.D. 220*, London: Duckworth, 139-81. (Excellent on Philo's Platonism.)

Mansfeld, J. (1988) 'Philosophy in the Service of Scripture: Philo's Exegetical Strategies', in J.M. Dillon and A.A. Long (eds) *The Question of 'Eclecticism'*, Berkeley, CA: University of California Press, 70-102.(On Philo's rationale in using philosophical ideas.)

Radice, R. and Runia D.T. (1988) *Philo of Alexandria: An Annotated Bibliography 1937-1986*, Leiden: Brill. (Comprehensive and heavily indexed guide to the scholarly literature, continued yearly in the journal *The Studia Philonica Annual*.)

Runia, D.T. (1986) *Philo of Alexandria and the Timaeus of Plato*, Leiden: Brill.(Standard treatment of Philo's extensive exploitation of Plato's cosmological dialogue.)

Runia, D.T. (1990) *Exegesis and Philosophy: Studies on Philo of Alexandria*, London: Variorum. (Collection of studies on exegetical method and philosophical content.)

Runia, D.T. (1993) *Philo in Early Christian Literature*, Assen: Van Gorcum and Minneapolis, MN: Fortress.(On Philo's reception and preservation in the Christian tradition.)

Tobin, T.H. (1983) *The Creation of Man: Philo and the History of Interpretation*, Washington, DC: Catholic Biblical Association of America.(Strong on Platonist background; emphasizes the differing traditions used by Philo.)

Winston, D. (1985) *Logos and Mystical Theology in Philo of Alexandria*, Cincinatti, IL: Hebrew Union College Press.(Best compact introduction to Philo's theology.)

Winston, D. (1990) 'Judaism and Hellenism: Hidden Tensions in Philo's Thought', *The Studia Philonica Annual* 2: 1-19.(Important observations on Philo's theory of prophecy and ethics.)

Wolfson, H.A. (1947) *Philo, Foundations of Religious Philosophy in Judaism, Christianity and Islam*, Cambridge, MA: Harvard University Press, 4th edn, 1968, 2 vols.(Magnificent synthesis, but too systematic and too Jewish a view of Philo.)

Philo of Larissa (*c*.159-*c*.83 BC)

Philo, head of the Academy from 110 to 88 BC, likened philosophy to medicine. No doubt he was a conscientious therapist himself; but we know little enough about his methods and practices. For most of his life he seems to have been a happy sceptic, and an unremarkable one. But towards the end of his career he introduced - or was deemed to have introduced - startling innovations into the Academy: in particular, by rejecting or modifying the reigning definition of knowledge he was able to separate himself from the scepticism of his school and to rewrite its history.

1 Life and thought

Philo came from Larissa in Thessaly. As a boy he was taught philosophy by a certain Callicles, a friend of Carneades; so that by the time he moved to Athens - perhaps in 134 BC - he was already acquainted with Academic philosophy (see Academy). For some years he continued his studies with Clitomachus, whom he eventually succeeded as scholarch in 110 BC. He remained in office for two decades. Then, in 88 BC, during the turmoil of the Mithridatic War, he left Athens and travelled to Rome. He resumed his teaching, impressing Cicero and Cicero's friends. Some five years later he died.

Our evidence for Philo's thought is meagre. The most substantial text develops a parallel between philosophy and medicine. Doctors must first persuade their patients to accept their own treatment and to reject the advice of rival practitioners; then they must eliminate the causes of disease and implant the causes of health; next they must examine the nature of health, which is the goal of their art; fourth, they must produce comprehensive prescriptions for diet and regimen; and finally - for life is brief and patients are busy - they must prepare a summary brochure on 'healthcare for all'.

So it is also with philosophers. First, a protreptic discourse commending virtue and refuting its detractors; then the elimination of false opinions, which infect the organs of the soul, and the introduction of true opinions; third, an account of the ultimate aim of philosophizing, namely 'happiness' (see Eudaimonia); then practical prescriptions, comprising an ethics and a political philosophy; and finally handbooks for the hurried.

The conception of philosophy as mental therapy was neither unique to Philo nor original to him, but his is the most expansive version of the thing which any ancient text presents. We know little about the particular forms which his therapy took: one notice indicates that he did not altogether abjure the soft pleasures of the Epicureans; another informs us that he taught rhetoric (but perhaps separately from philosophy); otherwise the chief points of interest lie in epistemology.

2 Epistemology

For most of his career Philo of Larissa appeared as an unexceptional philosopher. He continued the Academic technique of arguing 'on both sides' (see Carneades §1); and he professed some form of scepticism. In short, 'he enhanced the teachings of Clitomachus' (Numenius, in Eusebius, *Preparation for the Gospel* XIV 9.1) without adding anything new to them. And then, near the end of his life, he produced work which scandalized his old colleagues, who could hardly believe that it came from his pen.

The innovation is depicted by Cicero as sudden and shocking, and dated to Philo's Roman period. Other texts hint at a gradual realization rather than a flash of enlightenment - and to some extent Philo seems to have been anticipated by Carneades' pupil Metrodorus. However that may be, the scandalous innovation appears to have had two aspects, one of them philosophical, the other historical.

Philo urged that

> as far as the Stoic criterion, i.e. apprehensive appearance (*phantasia katalēptikē*), is concerned, things are inapprehensible; but as far as the nature of the things themselves is concerned, they are apprehensible.
>
> (Sextus Empiricus, *Outlines of Pyrrhonism* I 235)

If things are apprehensible in themselves but not according to the Stoic criterion, then presumably the Stoic criterion is false: it must constitute, or be based upon, a mistaken conception of what knowledge is. And Cicero

reports that Philo 'weakens and destroys' the Stoic account of apprehension. In particular, he denied the crucial element in the Stoic account, namely the claim that an apprehensive appearance must be such that it *could not* have come from any source other than the source it in fact came from (*Academics* II 18) (see Stoicism §12).

We are not told why Philo rejected this claim: no philosophical arguments on the matter have survived. But we do know one result of his having done so: scepticism could now be avoided - for the Academic arguments for scepticism had fastened upon the very clause which Philo decided to reject (see Arcesilaus §2; Carneades §4).

Looking out of my window I receive an appearance of a magpie on the lawn. I thereby know that there is a magpie on the lawn provided that the appearance is an apprehensive appearance. According to the standard definition, the appearance will only be apprehensive if it *could not* have been caused by anything other than a magpie on the lawn. But surely it *could* have been caused by any number of other things - by a jay on the lawn or by an ingenious hologram. And something like this seems to hold not only of my magpie but of all appearances: hence scepticism. But once we reject the requirement that an apprehensive appearance *could not* have been otherwise caused, the sceptical argument collapses. Knowledge does not require that the magpie appearance *could not* have been caused by a jay - but only that it *was not* caused by a jay. Since the appearance was *in fact* appropriately caused, it is apprehensive; and I thereby know that there is a magpie on the lawn.

Such, or so it seems, was the philosophical aspect of Philo's innovation. It would be silly to shower praise on him, for we do not know why he rejected the standard account of knowledge, or what he offered in its place, or how he replied to the objections which his innovation excited. But his innovation was admirable: the requirement that apprehensive appearances *could not* have been otherwise caused is implausibly strong.

Philo was now able to look afresh at the history of the 'sceptical' Academy. Arcesilaus and Carneades, he saw, had not really been attacking *knowledge* or arguing for *scepticism*; rather, they had been attacking a false conception of what knowledge was. Perhaps they themselves supposed that they were attacking the very citadel of knowledge; but in point of fact they were tilting at an antiquated windmill. Hence the 'sceptical' Academy had never been sceptical: its heroes had never cast any doubt on the possibility of knowledge.

No need, therefore, to speak of an Old and a New Academy. The dogmatism of Plato's original school had presupposed the sane, Philonian, account of knowledge - an account which the Peripatetics had inherited; and the 'New' Academics had implicitly countenanced Platonic dogmatism. Hence Philo 'denied that there were two Academies' (Cicero, *Academics* I 13).

Philo's innovations were roughly greeted. Some Academicians maintained that his abandonment of the standard definition of knowledge was folly - and self-destructive folly to boot, inasmuch as he would now be dragged back to the very scepticism which he had hoped to escape (see Antiochus §2). Others accused Philo of misunderstanding the achievement of the New Academy and of betraying his heritage. The controversy was heated; but we learn no details. Nor does Philo seem to have had much effect. After him, it is true, sceptical Platonists were hard to find, but there is no reason to believe that this was due to the arguments of Philo.

JONATHAN BARNES

References and further reading

Barnes, J. (1989) 'Antiochus of Ascalon', in M. Griffin and J. Barnes (eds) *Philosophia Togata I: Essays on Philosophy and Roman Society*, Oxford: Clarendon Press.(Includes an account of Philo's innovations.)

Cicero, M.T. (early 45 BC) *Academics*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933.(The primary source for Philo's dispute with Antiochus.)

Mette, H.-J. (1986/7) 'Philon von Larisa und Antiochus von Askalon', *Lustrum* 28/9: 9-63.(Collects the ancient testimonies on Philo.)

Tarrant, H. (1985) *Scepticism or Platonism? The Philosophy of the Fourth Academy*, Cambridge: Cambridge University Press.(An account of the philosophical orientation of Philo's Academy.)

Philo the Dialectician (late 4th-early 3rd centuries BC)

A member of the Dialectical school, Philo was a Greek philosopher whose claim to fame is twofold. First, he maintained that one proposition implies another if and only if either the former is false or the latter is true, however irrelevant the two propositions may be to one another. Second, he maintained that some things are necessarily prevented from happening and are nevertheless possible: the water around them prevents shells at the bottom of the ocean from ever being actually perceived; nevertheless, such things are perceptible. As often happened with Dialecticians' ideas, these ideas of Philo elicited (and perhaps were meant to elicit) notoriety rather than agreement.

1 Life, works and reputation

Philo's birthplace is unknown: the belief that it was Megara is a misconception based on the further misconception that he belonged to the Megarian school. In fact, Philo was a Dialectician (see Dialectical school). He was a pupil of the eminent Dialectician Diodorus Cronus at the same time as Zeno of Citium. In spite of his great admiration for both Diodorus and Philo, Zeno left to found the Stoic School (see Stoicism). Philo stayed. He wrote a no longer extant dialogue called the *Menexenus*; its characters included the five daughters of Diodorus, who were themselves all Dialecticians. He also wrote *On Meanings* (or perhaps *On Sign-Inferences*) and *On Forms of Argument*; these works too have been lost, as have the works which the prominent Stoic Chrysippus wrote in reply. Philo had the Dialectician's characteristic interest in paradoxes, and his name is associated with many of them. The only topics on which we have any details of his views are implication and modality.

2 What follows from what?

Hellenistic logicians generally, and rightly, agreed that no falsehood ever follows from a truth: for example, if it is daylight but I am not talking, then 'I am talking' does not follow from 'It is day'. In consequence, a necessary condition for one proposition to follow from another is that either the former is true or the latter is false. Then, as now, logicians agreed that this condition is necessary; then, as now, logicians disagreed over what other conditions are necessary (see Diodorus Cronus §3). Philo's contribution was to maintain that there are no other necessary conditions, and that this condition is therefore not only necessary but also sufficient. In short, according to Philo, one proposition follows from another when and only when either the former is true or the latter is false.

Already in antiquity, this definition was known to have some bizarre consequences. One is that whenever it is day but I am not talking, the proposition 'I am talking' does not follow from the proposition 'It is day'; it does however follow on all other occasions; that is, it follows throughout the night, and it follows also whenever I am talking in the daytime. Thus one proposition can follow from another even though the pair are, to all appearances, entirely irrelevant to one another. Another bizarre consequence known to antiquity is that 'It is day' follows from 'It is night' during the daytime, but not otherwise. Thus even where two propositions are relevant to one another in such a way that each evidently implies the other's falsehood, one of them will also imply the other's truth.

In modern times, people have sometimes given Philonian accounts of conditionals: a conditional is true, on such an account, if and only if either the antecedent is false or the consequent is true. No modern logician however has given such an account of when an argument is valid. Philo did. Like other Hellenistic logicians, he will have meant his definition of 'following from' to apply simultaneously both to the relation between consequent and antecedent in a true conditional and to the relation between conclusion and premises in a valid argument. According to all Hellenistic logicians, the conditional 'If it is day, it is night' is true if and only if the argument 'It is day; so it is night' is valid; and hence, according to Philo, that argument is valid throughout the night and invalid during the day.

It is not known why Philo adopted so strange an idea. His idea is however fairly easily inferred from a couple of assumptions which were widely current among Hellenistic logicians. One of those assumptions runs: validity is only one among several features which our reasonings should possess. That the conclusion follows from the premises is one thing; that the premises are acknowledged as true is another; that the conclusion is not already acknowledged as true is a third; and so on. The other assumption runs: each of these features contributes in a different way to turning our reasonings into decent proofs; and the distinctive contribution of validity is to ensure

that we will not reach false conclusions unless we have started from false premises. This distinctive contribution has been made if the conclusion follows from the premises in the manner defined by Philo. Hence why should not Philo's definition of 'following from' be an adequate definition of validity?

3 Possibility

Diodorus Cronus defined the possible as that which either is or will be true (see Diodorus Cronus §4). Philo devised examples which, if acceptable, refute Diodorus' definition. The examples ascribed to Philo have a common pattern. A typical example runs: a log in mid-Atlantic is not and never will be burning; this is because the log is 'necessarily prevented' from burning by the water that surrounds it; nevertheless, the log can burn; this is because the subject *log* is 'appropriate' to the predicate *burns*, and because 'the bare appropriateness of the subject' to a predicate is enough to make it, 'in itself and as far as its own nature is concerned', capable of having the predicate. Now only if things are possible does it make sense to talk about preventing them: thus since asbestos cannot burn, we need not immerse asbestos in the Atlantic, or take any other precautions, to prevent it from burning. Hence we cannot easily dismiss Philo's examples of possibilities that neither are nor ever will be actual. The very fact that they are prevented from happening indicates, we might say, that they are genuine possibilities.

It is not only Diodorus' definition that Philo's examples threaten. These examples threaten also the more widely accepted thesis that a thing cannot do what it necessarily does not do. For according to Philo's examples, things can act in ways in which they are necessarily prevented from acting; and yet it seems that they necessarily do not do what they are necessarily prevented from doing.

Perhaps Philo's purpose in devising his examples was simply to embarrass the consensus equating the impossible with the necessarily false. Later sources however tell us that, besides his examples, Philo also gave a definition of possibility. According to Philoponus: Philo says that the possible is that which is actualized or can be actualized, even though it never is actualized; as when we say that the shell at the bottom of the ocean is perceptible (*On Aristotle's Prior Analytics* 169).

And according to Boethius, Philo gave systematic definitions of a range of modal concepts:

> Philo says that the possible is that which, by the intrinsic nature of the proposition, would receive truth, as when I say that today I will reread the *Bucolics* of Theocritus: this, if nothing external prevents it, can in itself be truly asserted. This very same Philo defines the necessary in the same way, as that which, since it is true in itself, never could be receptive of falsehood. The non-necessary moreover he defines as that which in itself could receive falsehood: and the impossible as that which in accordance with its intrinsic nature never could receive truth.
>
> (*De Interpretatione* II 234)

The definition in Philoponus moves around too tight a circle to be illuminating. Less unhelpful are the definitions in Boethius. The definitions of the necessary and the non-necessary are not circular at all. Moreover, the clauses about 'intrinsic nature' mean that even the definitions of the possible and the impossible are not as tightly circular as the definition in Philoponus. Nevertheless, these definitions leave something to be desired. Where our subject is *log* and our predicate is *burns*, we have some independent understanding of talk about the subject's 'nature', and its 'appropriateness' to the predicate; but where our 'subject' is the proposition *I will today reread the Bucolics of Theocritus* and our 'predicate' is *true*, such talk is hardly comprehensible save as a long-winded way of saying that the proposition is possible. It is therefore difficult to avoid the suspicion that these definitions were devised not by Philo himself, but by later scholars drawing on his examples of unrealized possibilities. This suspicion is confirmed by the way that the definitions in Boethius invoke as unproblematic the equivalence between the impossible and the necessarily false that Philo's own treatment of his examples had apparently undermined.

See also: Consequence, conceptions of; Logic, ancient §5; Modal logic; Modal logic, philosophical issues in

NICHOLAS DENYER

References and further reading

Döring, K. (1972) *Die Megariker (The Megarics)*, Amsterdam: Grüner.(Collection of ancient sources, together with some German commentary.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae (The Fragments of Socrates and the Socratics)*, Naples: Bibliopolis.(Volume 1 pages 414-37 contains a complete collection of the ancient evidence about Philo.)

Kneale, W. and Kneale, M. (1962) *The Development of Logic*, Oxford: Clarendon Press. (Chapter 3, sections 1, 2 and 3, discusses Philo.)

Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press. (Contains some of the main sources, together with English commentaries and translations.)

Philodemus (*c*.110-*c*.40 BC)

Philodemus of Gadara, a Greek epigrammatic poet, was also an influential Epicurean philosopher. Scrolls containing many of his works, buried by the eruption of Vesuvius in AD 79, have been partially recovered and deciphered. Their themes include epistemology, theology, ethics, philosophical history, poetics, rhetoric and music. He energetically defends Epicureanism against other philosophies, and his own interpretation of Epicureanism against rival factions. Although not a notably original thinker, Philodemus became highly regarded in educated Roman circles.

Philodemus was born in Gadara, southeast of the Sea of Galilee, and died in Italy. He studied at Athens under Zeno of Sidon, then head of the Epicurean school, to whom he retained a lifelong loyalty. Moving to Italy probably in the 70s, he became a friend and philosophical adviser to L. Calpurnius Piso Caesoninus, the father-in-law of Julius Caesar. He seems to have spent some of his life at Herculaneum, probably in his patron's villa, where an Epicurean library consisting largely of his works was excavated in the 1750s. He was known and highly respected by Cicero, and seems to have numbered the poets Virgil and Horace among his circle.

Before the recovery of his philosophical library, Philodemus was known mainly as the author of some elegant epigrams preserved in the Palatine Anthology. Most of them are on the theme of love, but among the others are an invitation to a celebration in honour of Epicurus' birthday, poems about modest feasts and a prayer for a safe journey. The carbonized papyrus scrolls containing his philosophical treatises, because of the adverse physical conditions in which they survived, all lack at least their opening part, and most of them are altogether very poorly preserved. It has nevertheless been possible to gain from them considerable insight into Philodemus' working methods and attitudes to philosophical history and orthodoxy, and his views on the liberal arts. While some of his works may be independent compositions, others, such as *On Signs* and *On Frankness*, are little more than his own transcripts of the lectures he had heard from Zeno in Athens.

Philodemus' wide interest in rhetoric, music and poetry, although at first sight surprising in an Epicurean, is probably not untypical of the school in his day. With regard to rhetoric, he cites chapter and verse from his Epicurean authorities to attack as heterodox the austere rival wing of the school which regarded all rhetoric as philosophically unacceptable. In *On Music*, he criticizes the belief that music influences character: being irrational, it cannot influence the morally all-important rational soul, and it gains its educational power only when combined with language. His *On Poems*, as well as being a major source on Hellenistic poetics, is of particular interest for its development of the thesis that only the combination of word and thought can make a poem a good one. *On the Good King According to Homer* carries a covert political message for Philodemus' Roman audience, praising the mild Homeric king Alcinous as a just ruler who brings prosperity.

Later Philodemus turned to such ethical themes as death and anger, as well as to more practical topics like household management. His *On Frankness* is a work of especial interest, exploring its topic in the specifically Epicurean context of teacher-pupil relationships, with special emphasis on the therapeutic aspect of Epicurean philosophy as medicine for the soul. In *On Anger*, Philodemus seeks a middle position between the Stoic aim of eradicating passion and the Peripatetic ideal of 'moderate emotion': some anger is natural and inevitable, but it should always be controlled by reason to the greatest possible extent (see Peripatetics; Stoicism §19).

On Signs is an epistemological treatise, recording debates between Epicureans (including Zeno of Sidon) and critics who are probably Stoics. The Epicurean method of inference based on similarity, whose main species are (1) induction and (2) analogical reasoning from the visible to the invisible, is defended as both sound and scientifically fundamental (see Epicureanism §7). Two other works, *On the Gods* and *On Piety*, explore the nature of divinity and defend Epicurus' theology against critics who considered it impious. Philodemus' reverence for his Epicurean authorities finds a more concrete embodiment in his biographical studies of early Epicureans through their collected correspondence. He also gained some recognition for his historical work *Compendium of the Philosophers*, from which his histories of two schools, the Academy and the Stoa, partially survive among the papyri.

There is little evidence that Philodemus was interested in physics. His main fields were ethics, theology and the

liberal arts. He may be read as on the one hand a school loyalist, deeply concerned with Epicurean traditions and their correct interpretation, and on the other hand an ambassador mediating these traditions to a largely Roman audience.

MICHAEL ERLER

List of works

Few of Philodemus' treatises are at present available in usable English translations, although much valuable work has been done since 1970 on improving the Greek texts. The following are the more important or accessible among them. They cannot be individually dated, but were probably written in the order given below. Further editions of Philodemus' works will be found in the series *La scuola di Epicuro*, general editor M. Gigante, Naples: Bibliopolis, and English translations in the forthcoming series D. Blank, R. Janko, and D. Obbink (eds) *Philodemus' Aesthetic Works*, Oxford: Oxford University Press.

Philodemus (*c*.80-40 BC) *Rhetorica*, in 7 or more books, in H.M. Hubbell 'The *Rhetorica* of Philodemus', *Transactions of the Connecticut Academy of Arts and Sciences* 23 (1920): 243-382; Books I-II trans. D. Blank, Oxford: Oxford University Press, forthcoming.(The former is not a translation but an exegetical paraphrase)

Philodemus (*c*.80-40 BC) *On Piety*, ed. D. Obbink, *Philodemus on Piety Part I*, Oxford: Oxford University Press, 1996.(Epicurean views about piety; exposition of erroneous beliefs; includes text and translation of the philosophical part.)

Philodemus (*c*.80-40 BC) *On the Good King According to Homer*, ed. T. Dorandi, *Filodemo. Il buon re secondo Omero*, Naples: Bibliopolis, 1982.(Text, Italian translation and commentary; how to derive a benefit from the poetry of Homer.)

Philodemus (*c*.80-40 BC) *History of the Academy*, ed. T. Dorandi, *Filodemo. Storia dei filosofi. Platone e l'Academia*, Naples: Bibliopolis, 1991.(Text, Italian translation and commentary; about Academics from Plato to Aristus.)

Philodemus (*c*.80-40 BC) *On Anger*, ed. G. Indelli, *Filodemo. L'ira*, Naples: Bibliopolis, 1988.(Text, Italian translation and notes; on the Epicurean view of anger and how to deal with this passion.)

Philodemus (*c*.80-40 BC) *On Frankness*, ed. A. Olivieri, *Philodemi peri parrēsias libellus*, Leipzig: Teubner, 1914; trans. D. Kanstan, D. Clay, C. Glad., J. Thorn and J. Ware, Atlanta, GA: Scholars Press, 1998.(The former is Greek text only; the latter includes an emended version of Olivieri's text, with facing translation, introduction and notes.)

Philodemus (*c*.80-40 BC) *On Signs*, P. De Lacy and E. De Lacy, *Philodemus, On Methods of Inference*, Naples: Bibliopolis, 2nd edn, 1978.(Text, English translation and notes; on Epicurean scientific method.)

References and further reading

Annas, J. (1989) 'Epicurean Emotions', *Greek, Roman and Byzantine Studies* 30: 145-64.(Exposition of Philodemus' *On Anger*.)

Asmis, E. (1990) 'Philodemus' Epicureanism', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36: 4, 2369-406.(Useful survey article.)

Erler, M. (1994) 'Epikur; Die Schule Epikurs; Lukrez' (Epicurus; The School of Epicurus; Lucretius), in H. Flashar (ed.), *Die Philosophie der Antike, Band 4: Die Hellenistische Philosophie*, Basle: Schwabe, 289-362. (Informative about Philodemus, with extensive bibliography.)

Gigante, M. (1983) *Ricerche Filodemee (Research on Philodemus)*, Naples: Macchiaroli, 2nd edn.(Pioneering essays on Philodemus.)

Gigante, M. (1990) *Filodemo in Italia*, Florence: Le Monnier; trans. D. Obbink, *Philodemus in Italy. The Books from Herculaneum*, Ann Arbor, MI: University of Michigan Press, 1995.(Historical evaluation of Philodemus.)

Nussbaum, M. (1994) *The Therapy of Desire*, Princeton, NJ: Princeton University Press.(Makes extensive use of Philodemus' *On Anger* and *On Frankness* in reconstructing Epicurean moral therapy.)

Obbink, D. (ed.) (1995) *Philodemus and Poetry*, Oxford: Oxford University Press. (Important collection of essays, with extensive bibliography.)

Sedley, D. (1989) 'Philosophical Allegiance in the Greco-Roman World', in M.T. Griffin and J. Barnes (eds) *Philosophia Togata. Essays on Philosophy and Roman Society*, Oxford: Clarendon Press, 97-119.(Examines Philodemus' attitude to his school's scriptures.)

Philolaus (*c.*470-380/9 BC)

The Greek philosopher Philolaus of Croton, a contemporary of Democritus and Socrates, was a pre-eminent Pythagorean. His book counts as the first written treatise in the history of Pythagoreanism. Surviving in fragments, it constitutes an important source for our knowledge of fifth-century Pythagoreanism and supplements the picture given by Aristotle of Pythagorean doctrine.

Like earlier Presocratics Philolaus sought to furnish a comprehensive cosmology. Arguing from logical propositions, he posited two pre-existing principles: 'unlimited things' and 'limiting things'. United by harmony these two principles account for the formation of the cosmos and its phenomena. Since Philolaus also invokes number as an all-powerful explanatory concept, it is likely that he associated his first principles and the things originating from them with numbers. The emphasis on harmony and number accords with early Pythagoreanism.

Philolaus also wrote on musical theory and astronomy. A noteworthy feature of his astronomy is the displacement of the earth from the centre of the cosmos by fire, pictured as the 'hearth' of the universe. The fragments further attest Philolaus' interest in embryology, the causes of diseases, and physiology combined with psychological functions.

It was not unusual for early Greek philosophers to treat such a wide variety of topics. The distinctive elements of the thought of Philolaus are the logical arguments evinced in the fragments and the epistemological role of number for understanding the structure of reality.

1 Life

It is generally believed that Philolaus was a native of Croton in southern Italy where Pythagoras had established a community of disciples, although Philolaus would have been too young to have known the master himself. He is also associated with Tarentum and Thebes, cities he is likely to have sought out after the anti-Pythagorean revolts in Croton around 450 BC. That he spent time in Thebes is certain from a report in Plato's *Phaedo* (61d) stating that Philolaus lectured in that city. This must have occurred before 399 BC, the dramatic date of the *Phaedo*. Plato himself, during his travels to Sicily around 388 BC, may have met Philolaus, but this remains speculative.

Ancient sources relate that Philolaus published three books of Pythagoras which Plato sought to buy at great expense. Although these books have been shown to belong to later pseudo-Pythagorean literature, the story at least indicates that Philolaus contributed to the dissemination of Pythagorean ideas and that Plato was interested in Philolaus. Indeed, Plato was charged in antiquity with having copied his *Timaeus* from a book by Philolaus. Bearing the standard Presocratic title, *On Nature*, Philolaus' text represents the first attested written work of Pythagoreanism. With Philolaus, Pythagoreanism passes from an oral to a written tradition. A core of the fragments that survive from Philolaus' book have been established as genuine. These, as well as the testimonies of other ancient authors about Philolaus, form the basis for reconstructing his philosophy.

2 First principles

The Presocratic tradition in Greek philosophy, for all its diversity, has in common the central quest to understand, explain and, if possible define the essential nature (*physis*) of the world. Philolaus emerges from that tradition. The very beginning of his book sets out clearly the object of his philosophical concern: 'The nature in the cosmos was joined together from unlimiteds and limiters, both the whole cosmos and everything in it' (fr. 1). Like other Presocratics Philolaus seeks to explain the nature that makes up this world-order, and from the outset he states his explanatory principles: unlimiteds and limiters (the Greek literally translates as 'unlimited and limiting things'). The existence of both is justified on logical grounds: since the cosmos cannot be composed solely of limiters nor solely of unlimiteds, it must therefore be composed of both, as is evidenced by the phenomena of the world (fr. 2). If all things were wholly limited or wholly unlimited, the multiplicity and diversity that characterizes the world would be non-existent. (Here and in fragment 6, where the principles are said to be 'neither alike nor related', Philolaus may be polemicizing against the Eleatic arguments of Parmenides (§§4-8) and Melissus (§§2-4), who by postulating a single, uniform and eternally stable being denied all plurality, diversity and change.) Moreover, the possibility of human knowledge depends upon the existence of *both* principles, for 'there will not at all be anything that knows if all things were unlimited' (fr. 3).

Because Philolaus does not supply examples for ‘unlimiteds and limiters’, scholars have conjectured empty spaces and atoms, the even (or unlimited) and odd (or limited) elements of number or material components and their defining shapes, but none of these hypotheses receives unqualified support from the fragments. Possibly Philolaus never intended to define these principles more narrowly, given the epistemological scepticism he expresses at the beginning of fragment 6:

> The being of things, being eternal, and nature itself admit of divine and not human knowledge except that none of the things that exist and are known by us could have come to be if the being of the things from which the cosmos is composed did not already exist, [the being] of both limiters and unlimiteds.

The eternal and essential nature of things, including that of limiters and unlimiteds, ultimately lies beyond human understanding. Limiters and unlimiteds as universal, pre-existing principles have at this pre-cosmic stage an indefinable, abstract quality. But when they are joined in the actual process of forming the cosmos and the things in it, their manifest operation renders them in some sense concrete and knowable. On the best evidence of the fragments, the unlimited and limiting principles and the things composed of them become known through number.

3 Cosmogony and number

‘The cosmos is one’ (fr. 17). Limiters and unlimiteds alone, however, do not suffice to account for the unity of the universe. As opposing principles they need to be joined in order to produce a cosmos, a term that in Greek expresses beauty, order and structure. In a continuation of fragment 6, Philolaus reveals what links the unlimited and limiting principles:

> But since these principles pre-existed and were neither alike nor related it would have been impossible for them to be ordered (*kosmēthēnai*), if harmony (*harmonia*), in whatever way it came to be, had not come upon them. Now like things and related things did not require in addition any harmony, but things unlike and not related… such things must be closely linked by harmony, if they are going to be contained in an order (*en kosmōi*).

With the introduction of *harmonia* (literally, a ‘fitting together’), Philolaus evidences a continuity with those Presocratics who emphasized a bonding force in the universe (see Empedocles §3; Heraclitus §3), and especially with the Pythagoreans for whom *harmonia* was a central tenet (see Pythagoreanism §2). Philolaus is not concerned with fathoming the origin of harmony but simply postulates it as necessary to explain the manifest coherence of things.

While the cosmological significance of harmony as an ordering force is implied throughout fragment 6, its specific application to the generation of the cosmos is made in the next fragment: ‘The first thing fitted together, the One, in the middle of the sphere is called the hearth’ (fr. 7). In Philolaus’ astronomy (see §4) the universe is conceived of as a sphere (the uniform nature of a sphere in which the notions of ‘up’ and ‘down’ are relative is stressed in fragment 17). For the concept of a spherical cosmos Philolaus may have been indebted to Parmenides. But breaking with the geocentric view of the Presocratics Philolaus locates in the middle of the sphere the element of fire, here called the hearth (in Greek domestic and religious culture the burning hearth occupied the centre of households and temples). As the centre of a sphere was considered its ‘origin’, so fire functions as the origin of the cosmos, which develops from the centre outwards. The primacy of fire (which also figures significantly in the philosophy of Heraclitus and Hippasus the Pythagorean), its status as the ‘first thing’ in the cosmos, finds a telling correspondence in Philolaus’ biology, according to which our bodies at the embryonic stage are solely constituted out of the ‘hot’. It was common in Greek thought, especially among medical writers with whom Philolaus shows some acquaintance, to make analogies between the life of humans (or animals) and the larger cosmic organism; in Philolaus, the heat or fire that has such a definitive role in biological generation is likewise the primary element in the formation of the cosmos, the starting point of all generation.

As the central fire is itself a result of harmony, it must then also be a compound of unlimited and limiting things, but the extant fragments again do not specify what these are. The product, however, is clearly a unity, described as ‘the One’. Of the different interpretations concerning the One in fragment 7, the view taken in this entry is that it refers to the number one. In fragment 4 Philolaus states emphatically that ‘all things that are known have number; for it is impossible for anything whatsoever to be understood or known without it’. With this assertion Philolaus places himself squarely in the Pythagorean tradition of number philosophy, although it remains questionable

whether he equated things with numbers as some Pythagoreans apparently did. Without question, however, number plays a crucial epistemological role for Philolaus: things can only be understood by means of number. Philolaus breaks down number into 'two proper kinds, odd and even, and third from a mixture of both of these, even-odd' (fr. 5); of the two proper kinds there are 'many forms, which each thing itself signifies'. This means that individual things in the world are signified by the many forms of number, which are the even and odd natural numbers, while it is implied that there are not many forms of the 'even-odd'; possibly only one. That form of the even-odd would be the One, since the Pythagoreans considered the One to be both odd and even in so far as it originates the number series that must consist of both odd and even numbers. It fits Philolaus' system that he should assign the number one to the 'first thing fitted together'. For things to 'have number' means that they can be known by their mathematical structure. The central fire, a harmonization of limiters and unlimiteds, is thus comprehended by a number that is itself a mixture of unlimited and limited - in numerical terms, of even and odd (the Pythagoreans put even on a par with unlimited and odd with limit). Just as the One stands at the beginning of the number series, so the central fire is the starting point for the development of the universe that, taken as a whole, will also be one (fr. 17), while the constituent bodies of the universe are signified by the numbers deriving from the One. In assigning the One to the central fire Philolaus intimates from the outset the close connection between things and their informing numbers that will remain a permanent feature of the cosmos.

The surviving fragments do not explain how the cosmogonical process continued, but a clue may be taken from passages in Aristotle that appear to be based partly on Philolaus' book. Aristotle reports that immediately after the construction of the Pythagorean One, the unlimited is drawn in and limited by the limit. From this and other Aristotelian accounts the picture emerges of the cosmos as a living organism that breathes in various unlimited stuffs from 'outside the heaven', the locale of all that is infinite. The particular unlimiteds are time, breath and the void (emptiness). Their interaction with the One, which is here given a limiting function, produces an orderly world where bodies (and numbers, spatially conceived) are separated by the interjection of empty space (see Democritus §2) and time is divided into measurable units. Some of this information probably derives from other Pythagorean sources, but that Philolaus subscribed to the general concept of a breathing universe originating from the One is suggested again by his embryology: the living animal, being hot, immediately upon birth draws in cold air from the outside and discharges it again, thus starting the regulated process of respiration (A27). This is paralleled on the cosmic level where the central fire (the hot) draws in and imposes limits on breath (the cold) or the void, which are unlimiteds from outside the universe.

4 Astronomy

The most striking feature of Philolaus' astronomy, as noted in §3, is the location of fire in the centre of the spherical cosmos. It is the hearth of the universe (this is not the sun, which is but one of the orbiting planets - heliocentrism first appears with Aristarchus in the third century BC). Philolaus' departure from the usual geocentric world-view of the Greeks appears to be motivated by Homer's description of Tartarus 'as far below Hades as heaven is from earth' (*Iliad* VIII 16). Although Homer does not associate Tartarus with fire, the ancient Pythagorean idea of a fire in the middle of the earth as well as the volcanic fires of southern Italy would have led Philolaus to postulate a sole central fire, identified with the Tartarus of Greek myth (hence also called the 'prison' or 'defence tower of Zeus' in the Philolaic-Pythagorean tradition). The central fire forms the natural counterpart to the fiery envelope that surrounds the spherical universe at its uppermost boundary (A16). In between and revolving around the central fire, apparently at decreasing velocities, are ten astronomical bodies in the following order: counter-earth; earth; moon; sun; the five planets (in no specified order); and the fixed stars. The counter-earth was not simply introduced, as Aristotle charges, to bring the number of all the planets up to the perfect Pythagorean ten. Rather, Philolaus seems to have posited it as a kind of earth in reverse, the world of the dead, and identified it, again according to the Homeric model, with the invisible Hades of Greek myth. Because it could not be seen, he theorized that the earth in its orbit rotates only once on its axis (this itself was to prove a fruitful idea in the history of astronomy), thus keeping its inhabited side turned away from the centre and rendering the counter-earth invisible. There are also indications that Philolaus shared the belief that the earth moved in an inclined orbit (ellipsis) to account for the different altitudes of the sun and the change of seasons. Other astronomical theories held by Philolaus deal with the periodic destruction of the earth, or parts of it, by fire and water (A18), the nature of the sun (which is like a glass reflecting the light and heat of the central fire) (A19) and the inhabitants of the moon (A20).

Philolaus' astronomy thus combines elements of inventive theorizing based on myth (the central fire, the counter-earth) with empirical data concerning celestial phenomena (night and day, the four seasons, etc.) to yield an orderly cosmos. Obviously, his astronomy must be appreciated within the context of the fifth century BC in which the progressive thought of the day was an amalgam of elementary scientific observations, philosophical preconceptions and mythological remnants.

5 Miscellanea

The discovery, attributed to Pythagoras, that the concordant intervals that make musical harmony possible could be expressed by certain numerical ratios, found a cosmic application in Pythagoreanism: 'the whole universe is harmony and number' (see Pythagoreanism §2). It is not surprising, therefore, that Philolaus, after describing the cosmic role of *harmonia* (fr. 6), constructs a diatonic scale to account for 'the size of harmony', incorporating the basic Pythagorean ratios of octave ($2:1$), the fifth ($3:2$) and the fourth ($4:3$) (fr. 6a). One significance of Philolaus' musical theory is that it may lie at the basis of Plato's account of the ratios comprising the structure of the world-soul (*Timaeus* 34b).

Philolaus also called the human soul a harmony (A23). Some scholars have argued that this isolated testimony is based on an inference from Plato's *Phaedo*, where the soul as a harmony of opposites is presented as a Pythagorean teaching, and is therefore spurious. But in the light of the pre-eminent role that Philolaus assigns to harmony, it is reasonable to assume that he would have brought this universally important concept to bear on any particular discussion of the soul.

Further interests of Philolaus concerned biology and medicine, but it is not known to what extent he wrote about these topics. Of his views on embryology and disease we have a single report (A27). Philolaus believed that the embryo is innately hot, because both sperm and the womb in which it is deposited are hot. Immediately at birth, 'the hot' inhales cold air from the outside and emits it again 'like a debt'; the process of respiration is presented as necessary for the cooling of the body. Diseases originate from bile, blood, and phlegm.

Philolaus posited four physiological principles, classified as follows:

> The head is the seat of intellect, the heart of life and sensation, the navel of rooting and first growth, the genitals of the sowing of seed and generation. The brain is the origin of man, the heart the origin of animals, the navel the origin of plants, the genitals the origin of all (living things). For all things both flourish and grow from seed.
>
> (fr. 13)

By locating intellect in the head, Philolaus differs from those Presocratics who associated thinking with the heart (for example, Empedocles). Another noteworthy detail is that Philolaus distinguishes between thought and sensation; this distinction was to be developed and clarified by Plato, who in the *Timaeus* also located psychic functions in different parts of the body. Overall, Philolaus' fourfold division of faculties, neatly applied in a hierarchical fashion to humans, animals, plants and all organisms, manifests the Pythagorean, and generally Presocratic, impulse to detect the order and structure by which the multiplicity of the world's phenomena can be comprehended.

See also: Archytas; Neo-Pythagoreanism; Presocratic philosophy

HERMANN S. SCHIBLI

References and further reading

Huffman, C.A. (1993) *Philolaus of Croton: Pythagorean and Presocratic*, Cambridge: Cambridge University Press.(Contains all the authentic as well as the spurious fragments and testimonia, with English translations, commentaries and readable interpretative essays; also includes a comprehensive bibliography. Argues that the One in fragment 7 does not refer to number but rather to a paradigmatic unity as found in much of Presocratic thought.)

Kahn, C.H. (1974) 'Pythagorean Philosophy before Plato', in A.P.D. Mourelatos (ed.) *The Pre-Socratics*, Garden City, New York: Doubleday.(A useful introductory essay that emphasizes Philolaus as a primary source for our knowledge of early Pythagoreanism.)

Kingsley, P. (1995) *Ancient Philosophy, Mystery and Magic: Empedocles and Pythagorean Tradition*, Oxford:

Oxford University Press.(An important book which throws new light on the study of Pythagoreanism; includes two chapters exploring the mythical background to Philolaus' notion of the central fire.)

Philolaus (*c*.470-380s BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidmann, 6th edn, 1951, vol. 1, 398-419.(The standard collection of the ancient sources both fragments and testimonia, the latter designated by 'A'; includes Greek texts with translations in German.)

Schibli, H.S. (1996) 'On "The One" in Philolaus, Fragment 7', *Classical Quarterly* 46.1: 114-30.(Argues on the basis mainly of Aristotle that Philolaus' 'One' is best understood as number.)

Philoponus (*c.* AD 490-*c.*570)

John Philoponus, also known as John the Grammarian or John of Alexandria, was a Christian philosopher, scientist and theologian. Philoponus' life and work are closely connected to the city of Alexandria and its famous Neoplatonic school. In the sixth century, this traditional centre of pagan Greek learning became increasingly insular, located as it was at the heart of an almost entirely Christian community. The intense philosophical incompatibilities between pagan and Christian beliefs come to the surface in Philoponus' work.

His œuvre comprised at least forty items on such diverse subjects such as grammar, logic, mathematics, physics, psychology, cosmology, astronomy, theology and church politics; even medical treatises have been attributed to him. A substantial body of his work has come down to us, but some treatises are known only indirectly through quotations or translations. Philoponus' fame rests predominantly on the fact that he initiated the liberation of natural philosophy from the straitjacket of Aristotelianism, through his non-polemical commentaries on Aristotle as well as his theological treatises deserve to be appreciated in their own right.

Philoponus' intellectual career began as a pupil of the Neoplatonic philosopher Ammonius, son of Hermeas, who had been taught by Proclus and was head of the school at Alexandria. Some of his commentaries profess to be based on Ammonius' lectures, but others give more room to Philoponus' own ideas. Eventually, he transformed the usual format of apologetic commentary into open criticism of fundamental Aristotelian-Neoplatonic doctrines, most prominently the tenet of the eternity of the world. This renegade approach to philosophical tradition, as well as the conclusions of his arguments, antagonized Philoponus' pagan colleagues; they may have compelled him to abandon his philosophical career. Philoponus devoted the second half of his life to influencing the theological debates of his time; the orthodox clergy condemned him posthumously as a heretic, because of his Aristotelian interpretation of the trinitarian dogma, which led him to enunciate three separate godheads (tritheism).

The style of Philoponus' writing is often circuitous and rarely entertaining. However, he combines an almost pedantic rigour of argument and exposition with a remarkable freedom of spirit, which allows him to cast off the fetters of authority, be they philosophical or theological. Although his mode of thinking betrays a strong Aristotelian influence, it also displays a certain doctrinal affinity to Plato, stripped of the ballast of Neoplatonic interpretation. His works were translated into Arabic, Latin and Syriac, and he influenced later thinkers such as Bonaventure, Gersonides, Buridan, Oresme and Galileo.

1 The grammarian and commentator

In order to appreciate the magnitude of Philoponus' achievement as a philosopher, one has to consider briefly the background. One common way for philosophers in late antiquity to develop and communicate their ideas was the composition of commentaries on classical philosophical authors such as Plato and Aristotle. Generations of commentaries were written from the time of Alexander of Aphrodisias (early third century AD) to the seventh century and later, forming a massive tradition. Many of these works evolved out of the school practice of the Neoplatonists (see Neoplatonism §1) and conformed to several requirements: (1) the commentaries consisted of an extremely detailed oral exegesis of a philosophical text, designed for the benefit of students; (2) not uncommonly, pupils were responsible for taking verbatim notes of their teachers' instruction, thus turning the lecture into a book; (3) each commentary constituted a step within a substantial curriculum of philosophical training which began with Aristotle and aimed at progressing towards Plato - a commentator was expected to demonstrate the agreement between Plato and Aristotle, indeed, to show forth the harmony among all ancient philosophers including Homer; and (4) philosophy in this way was viewed as education as much as pagan religious exercise by which human beings could perfect their intellect and character with a view to 'becoming godlike as far as possible'.

After a basic education in grammar Philoponus, who was probably born into a Christian family, embarked on this curriculum. As a pupil of Ammonius, son of Hermeas, he assumed responsibility for writing down the lectures of the school's ailing head: the commentaries on Aristotle's *On Generation and Corruption*, *On the Soul*, *Prior Analytics* and *Posterior Analytics* explicitly state in their titles that they are 'based on Ammonius' seminars'. The commentaries on the *Categories*, the *Physics* and the *Meteorology* do not make such claims, which suggests, in conjunction with other evidence, that Philoponus was teaching these courses himself. Presumably succeeded Ammonius at some point, although he never held the chair of philosophy but remained *grammatikos*, professor of

philology. After the death of Ammonius, the school's leadership seems to have passed into the hands of the mathematician Eutocius and then on to the philosopher Olympiodorus, who was pagan.

In many ways, Philoponus' commentary on the *Categories* may be regarded as typical of its kind. The exegesis of Aristotle's text is prefaced by a general clarification of matters of interest to the student reading philosophy for the first time. Beginners were first instructed in Aristotle's logical works, arranged according to the principle of increasing complexity: the *Categories*, dealing with simple utterances, were followed by the *De interpretatione* (on propositions), the *Prior Analytics* (on syllogisms) and the *Posterior Analytics* (on proof). Discussing the problem of the subject matter (*skopós*) of the *Categories* (see Aristotle §7), Philoponus says that the *Categories* are about simple utterances (*phōnai*) signifying simple things (*pragmata*) by means of simple concepts.

As opposed to the Stoics, Neoplatonists followed Alexander of Aphrodisias in regarding logic not as a separate philosophical discipline but as philosophy's tool, its *Órganon*. In their commentaries on the *Prior Analytics*, both Philoponus and Ammonius attempt to reconcile the two positions.

The contemporary Athenian Neoplatonist Simplicius, who spent much time repudiating Philoponian ideas, doubted Philoponus' competence as a logician, and modern scholars tend to agree with him. Historians of logic, however, acknowledge that Philoponus was the first to render a satisfactory definition of the syllogism which states that the major premise includes the predicate term of the conclusion, the minor premise the subject term *On Aristotle's Prior Analytics* 67). In Philoponus, one also encounters for the first time the schematic diagram facilitating the construction of valid syllogisms, later termed by the schoolmen *pons asinorum* (*On Aristotle's Prior Analytics* 274).

The commentary on Aristotle's *On the Soul* is perhaps the earliest commentary to contain passages in which Philoponus abandons proper exegesis in order to modify or criticize Aristotelian doctrine, a tendency even more conspicuous in the *Physics* commentary, which can be dated to 517. It is convenient to distinguish between two kinds of criticism: substantial modification of Aristotelian ideas on the one hand, outright rejection on the other. Philoponus' commentary *On the Soul* presents a good example of the first kind of criticism. In *On the Soul* (II 7) Aristotle understood light as incorporeal and described its appearance as an instantaneous transition from the potentiality (*dýnamis*) of a medium to be transparent to the actuality (*enérgeia*) of transparency. The description fails to account both for the laws of optics and for the phenomenon that the region below the moon is warmed by the light of a celestial body, the sun. Philoponus proceeds to interpret the term *enérgeia* not as an actual state but as an 'incorporeal activity' which is capable of warming bodies, just as the soul is in the case of animals. Due to this shift of terminology, light is now understood as dynamic. It is possible to trace Philoponus' further development of the idea: in the *Meteorology* commentary, which belongs to a stage when Philoponus had already abandoned the assumption of an immutable celestial element (*aithēr*), he argues that light and heat may be best explained as consequences of the nature of the sun, which is fire. Heat is generated when the rays emanating from the sun are refracted and warm the air through friction.

2 The grammarian and commentator (cont.)

Philoponus' commentary *On Aristotle's Physics* contains an array of examples of innovative and critical commentary. One of his most celebrated achievements is the theory of impetus, which is commonly regarded as a decisive step from an Aristotelian dynamics towards a modern theory based on the notion of inertia. Concepts akin to impetus theory appear in earlier writers such as Hipparchus and Synesius, but Philoponus never mentions them. As far as one can tell from the text (*On Aristotle's Physics* 639-42), his point of departure is an unsatisfactory Aristotelian answer to a problem that was to puzzle scientists for centuries: why does, for example, an arrow continue to fly after it has left the bowstring? Since Aristotle supposed that (1) whenever there is motion there must be something which imparts the motion, and (2) mover and moved must be in contact, he was led to conclude that the air displaced in front of the projectile somehow rushes round it and pushes from behind, thus propelling the projectile along. This theory was still in vogue among Aristotelians of the sixteenth century, despite the fact that a thousand years earlier Philoponus had demolished it. He proposed instead that a projectile moves on account of a kinetic force impressed on it by the mover, which exhausts itself in the course of the movement. Philoponus compares this 'incorporeal motive *enérgeia*', as he calls it, to the activity earlier attributed to colour and light.

Once projectile motion was understood in terms of an impetus in this way, it became possible to reassess the role

of the medium: far from being responsible for the continuation of a projectile's motion, it in fact impedes it (*On Aristotle's Physics* 681). Against Aristotle Philoponus concludes that there is in fact nothing to prevent one from imagining motion taking place through a void. As regards the natural motion of bodies falling through a medium, Aristotle's verdict that the speed is proportional to the weight of the moving bodies and indirectly proportional to the density of the medium is disproved by an experiment similar to that which Galileo was to carry out centuries later (*On Aristotle's Physics* 682-4).

Philoponus' impetus theory ties in with wider criticisms of Aristotelian principles of physics. Aristotle had dismissed the notion of a void as an incoherent logical impossibility. Philoponus concedes that in nature empty spaces never become actual, but he insists that a clear conception of the void is not only coherent but also necessary if one wants to explain movement in a plenum (the place-swapping of bodies presupposes that there is empty space to be filled; *On Aristotle's Physics* 693-4) or such phenomena as the force of the vacuum (571-2). Philoponus' elaborate defence of the void (675-94) is closely related to his conceptions of place and space (*On Aristotle's Physics* 557-85). Aristotle defined the place of a body as the inner surface of the body that contains it (*Physics* IV 4); Philoponus replies that place ought to be conceived as the three-dimensional extension equal to the determinate size of the body, that is, its volume. Likewise, space is indeterminate pure three-dimensional extension devoid of body. It is not infinite, however, but coextensive with the size of the universe.

Philoponus' discussion of matter dovetails well with this conception of space. In the commentary *On Aristotle's Physics* (687-8) he argues in a vein similar to Aristotle in *Metaphysics* (VII 3). In abstracting all qualities from body Aristotle arrived at prime matter, which the Neoplatonists later defined as incorporeal and formless. Philoponus, in contrast, arrives at something he calls 'corporeal extension', which is a composite of prime matter and indeterminate quantity and must not be confused with space. His argument here may still be regarded as elaboration on Aristotle. However, in book XI of the polemical treatise against Proclus (see §2), he rejects the Aristotelian-Neoplatonic conception of prime matter and lets the most fundamental level of his ontology be constituted by 'the three-dimensional', indeterminate corporeal extension. In this he claims to be following the Stoics; this lowest ontological level is reminiscent of the Cartesian *res extensa*, although Descartes would not permit the distinction between space and corporeal extension (see Stoicism §3; Descartes, R. §11).

In order to rebuff the likely objection that the three-dimensional cannot be the most fundamental level of being because extension, belonging to the category of quantity, is an *accident* and requires the assumption of an underlying *subject*, Philoponus argues that extension is in fact an essential and inseparable differentia of the three-dimensional, just like heat in fire or whiteness in snow. Thus quantity is constitutive of body as such, which amounts to a promotion of quantity to the category of substance. There are indications that Philoponus would have modified Aristotle's scheme of the *Categories* if he had revised his early commentary on that treatise.

The incomplete commentary on the *Meteorology* may well be the last commentary Philoponus wrote on Aristotle. It is worth noting how the lectures are presented, with an air of aloofness and, at times, deliberate vagueness. In several places, especially when he has to comment on the nature and movement of the heavens, Philoponus breaks off and refers the student to previously published work: what Philoponus *really* has to say about the text seems no longer appropriate for the classroom. He has successfully shaken off the weight of Aristotle's and anyone else's authority, and far from demonstrating the harmony among philosophers he himself contributes to their dissent. The commentator has turned into a critic.

3 The critic

In the fifth century, the Athenian Neoplatonist Proclus wrote a defence of the pagan Greek belief in the eternity of the world. His aim was to show that Christian creationism was intellectually untenable. Eighteen arguments were loosely grouped around the myth of Plato's *Timaeus*, which, according to Proclus, was consistent with an eternalist reading.

In 529, when the Emperor Justinian suppressed pagan philosophy in Athens, Philoponus published a reply entitled *On the Eternity of the World against Proclus*. The book is something like an anti-commentary to the Proclean arguments. Combing through the text Philoponus repudiates literally every single point made by Proclus. Although his efforts are evidently motivated by faith, he keeps Biblical theology out of his polemic, attempting to refute Proclus solely within the framework of Platonism. The *Timaeus* is read as a genuine account of creation, and a

fresh analysis of the processes of generation and corruption even renders viable an idea which Greek philosophers never allowed: creation out of nothing. Even if it were true that creation out of nothing never occurs in nature, Philoponus argues, God is surely more powerful a creator than nature and therefore capable of it.

Philoponus' battle against eternalism may be divided into three different stages. The first, the work against Proclus, is followed by a second and even more provocative publication, *On the Eternity of the World against Aristotle*. This work was published *c.*530-4 and involved a close scrutiny of the first chapters of Aristotle's *On the Heavens* (on the theory of aether) and the eighth book of the *Physics* (on the eternity of motion and time). The third stage is represented by one, perhaps two non-polemical treatises in which arguments against eternity and for creation were arranged in systematic order.

Like the polemic against Proclus, that against Aristotle is mainly devoted to removing obstacles for the creationist. If Aristotle were right about the existence of an immutable fifth element (*aithēr*) in the celestial region, and if he were right about motion and time being eternal, any belief in creation would surely be unwarranted. Philoponus succeeds in pointing to numerous contradictions, inconsistencies, fallacies and improbable assumptions in Aristotle's philosophy of nature. Dissecting the text in an unprecedented way, he consistently refutes Aristotle's arguments and so paves the way for demonstrative arguments for non-eternity. One such argument (fr. 132) reported by Simplicius (*On Aristotle's Physics* 1178.7-1179.26) relies on three premises: (1) if the existence of something requires the pre-existence of something else, then it will not come to be without the prior existence of the other; (2) an infinite number cannot exist in actuality, nor be traversed in counting, nor be increased; and (3) something cannot come into being if its existence requires the pre-existence of an infinite number of other things, one arising out of the other. From these not un-Aristotelian premises, Philoponus deduces that the conception of a temporally infinite universe, understood as a successive causal chain, is impossible. Furthermore, given that the spheres have different periods of revolution, the assumption of their motion being eternal would lead to infinity being increased, even multiplied, which Aristotle thought absurd.

The non-polemical treatises exploit, among others, Aristotle's argument that an infinite *dýnamis* cannot reside in a finite body (*Physics* VIII 10); Philoponus infers that since the universe is finite, it cannot have the *dýnamis* to exist for an infinite time. As in the case of the theory of light, this argument involves a shift of meaning. In the context of Aristotle's prime mover argument, *dýnamis* meant 'kinetic force'; Philoponus, however, uses the word in the sense of 'existential capacity' or 'fitness'.

By the end of the 530s Philoponus seems to have stopped producing philosophical works; for reasons on which one can only speculate, his career as a philosopher was over.

4 The theologian

Since there is no evidence that John Philoponus belonged to a clerical order, it is difficult to picture his professional life as a theologian. He published his theological treatises under the same name as before, *Iōánnēs grammatikós*, but it seems hard to imagine him as a professor of grammar. Likewise, his nickname *philóponos* (lover of work) is probably best interpreted as an acknowledgement of his literary productivity, not as an indication of his status as one of the Christian brethren in Egypt who called themselves by that name.

Perhaps some fifteen years after his attack on Aristotle, Philoponus published the Creation story, *On the Creation of the World (De opificio mundi)*, his only theological work extant in Greek. While discussing the biblical text Philoponus frequently refers to philosophers such as Aristotle, Plato and Ptolemy as well as to Basil the Great, whose own treatise on the Creation served as his inspiration (see Creation and conservation, religious doctrine of). The treatise has received some attention from historians of science, because Philoponus suggests at one point (I 12) that the movement of the heavens could be explained by a 'motive force' impressed on the celestial bodies by God at the time of creation. Philoponus had discussed impetus theory for the first time in the context of forced motion (see §2); now he applies the theory to the universe at large. Significantly, Philoponus compares the rotation imparted to the celestial bodies to the rectilinear movements of the elements as well as to the movements of animals: they are all brought about by the creator's divine impetus. Due to this suggestion Philoponus is often credited with having envisaged, for the first time, a unified dynamic theory, striving to give the same kind of explanation for phenomena which Aristotle had to explain by various different principles.

Judging from the fragmentary evidence, Philoponus' later theological treatises are characterized by a curious

mixture of Christian piety and Aristotelian philosophy. On the eve of the fifth Council (Constantinople 553), Philoponus stepped forward as a partisan of monophysite Christology which, in the course of the century, had become increasingly influential in the eastern part of the empire. The monophysites, who emphasized the divinity of Christ, were scandalized by the conjunction of Christological formulas enunciated at the Council of Chalcedon in 451. There Christ was confessed to be: (1) consubstantial (*homooúsios*) with the Father; (2) consubstantial with humans; (3) one person and one *hypóstasis* ('existence'); and (4) 'discernible in two natures', a phrase referring to the '*unitatem personae in utraque natura intelligendam*' endorsed by the Bishop of Rome and familiar to theologians of the Latin West. Although propositions (1), (2) and (3) were no longer controversial, proposition (4) could be read as an unholy concession to those who regarded Jesus merely as a divinely inspired man. In addition, how could it be sensible to predicate two entirely different natures of a single entity?

In the *Arbiter* Philoponus takes the view that the phrase 'in two natures' ought to be abandoned. His main strategy is to argue that in this context the meaning of the terms 'nature' and '*hypóstasis*' is virtually identical. 'Nature' has two senses, one general, one particular - we can speak of the nature of man in general or of the nature of this individual man. When one speaks of the unification of two natures in Christ, the reference cannot be to the universal natures of godhead and manhood (otherwise it would be true to say that the Father and the Spirit have also become man, since the universal nature of 'godhead' applies to them as it does to the Son). Hence, the reference must be to the particular nature of the Logos and the particular nature of Jesus the man. But 'particular nature' is virtually the same as *hypóstasis*, which Philoponus uses also as a synonym of 'person' and 'individual'. Since it is agreed that Christ is one person and *hypóstasis*, he must consequently be of one nature, albeit a complex one which combines and preserves the properties of both god and man: 'one composite nature' (*mía phýsis sýnthetos*).

Besides monophysitism Philoponus' name is associated with the precarious doctrine of tritheism. However, there is an important difference: whereas monophysitism was a reputable and powerful movement among Eastern theologians, tritheism is merely a hostile label attached to those who were not afraid to render the mystery of the Trinity intelligible by drawing some clear distinctions. Already in the *Arbiter* it became clear that Philoponus associated with the term *hypóstasis* something like 'primary substance' in the Aristotelian sense (see above). The fragments of the very late treatise *On the Trinity* confirm this: since *hypostasis* is certainly not an accident of divinity, it must be the case that the three *hypostases* of the Trinity are three divine substances with distinct properties. Only on this assumption, too, is it reasonable to speak of consubstantiality: if there were only *one* divine substance, what could it mean to say that it is consubstantial with itself? When the Cappadocians spoke of the Trinity as 'one substance (*ousia*), three *hypostases*', they were not, Philoponus argues, enunciating a fourth primary substance, but using 'substance' in the secondary, abstract sense. In accordance with Aristotle Philoponus goes on to claim that universals exist only in the human mind, thus reducing monotheism apparently to a mere concept. In order to refute any reminiscences of pagan polytheism Philoponus points out that, unlike the gods of the pagans, the three divinities of the Trinity are all of the same divine nature (in the universal sense).

Philoponus had followers, but this kind of trinitarian philosophy was unacceptable to anyone not committed to Aristotle's ontology. And so, the so-called tritheists faced immediate and severe criticism; Philoponus himself was condemned (posthumously) by the Council of Constantinople in 680/1. It is often stated that Philoponus was one of the first thinkers to reconcile Aristotelian philosophy with Christianity; his contemporaries probably took a different view. Pagan philosophers certainly abhorred the way in which Philoponus used Christianity as a vantage point from which to disturb the harmony of the wise, and Christian theologians could not but castigate his dubious attempt at comprehending a spiritual mystery by making it conform to Aristotle.

5 Influence

In addition to inciting the disapproval of his contemporaries, both pagan and Christian, Philoponus' immediate influence may have been considerable among fellow monophysites in Egypt, but his condemnation as a heretic in 681 made the further proliferation of his theological ideas impossible. Likewise, his arch enemy Simplicius, also a pupil of Ammonius, son of Hermeas, submitted Philoponus' anti-eternalism almost immediately to thundering polemic which resounded through the centuries with the effect that later thinkers like Thomas Aquinas and Zabarella underestimated Philoponus in general. Although his work was studied by the Arabs, who referred to him as Yahya an-Nahwi or Yahya al-'Asqalani, they too tended to side with Aristotle. However, the arguments against

eternity eventually persuaded Bonaventure and Gersonides, and impetus theory was reaffirmed by Buridan and his pupil Oresme. In the sixteenth century, the first editions and numerous translations (into Latin) of Philoponus' commentaries and the treatise against Proclus began to appear in print. In particular his criticism of Aristotle in *On Aristotle's Physics* was widely discussed and persuaded such diverse thinkers as Giovanni Pico della Mirandola and, for different reasons, Galileo Galilei.

See also: Aristotle Commentators

CHRISTIAN WILDBERG

List of works

John Philoponus (*c*.510-15) *On Words with Different Meanings in Virtue of a Difference of Accent (De vocabulis quae diversum significatum exhibent secundum differentiam accentus)*, ed. L.W. Daly, American Philosophical Society Memoirs 151, Philadelphia, PA: American Philosophical Society, 1983.(A philosophically unimportant work on Greek homonyms.)

John Philoponus (*c*.510-15) *On Aristotle's On Generation and Corruption*, ed. H. Vitelli, *Commentaria in Aristotelem Graeca* XIV 2, Berlin: Reimer, 1897.(A commentary based on Ammonius' seminars containing virtually no criticism of Aristotle.)

John Philoponus (*c*.510-15) *On Aristotle's On the Soul*, ed. M. Hayduck, *Commentaria in Aristotelem Graeca* XV, Berlin: Reimer, 1897. (Probably the ideas are substantially Ammonius'. The authenticity of book III is disputed, because a Latin version attributed to Philoponus differs from the text transmitted in Greek (see Verbeke 1966).)

John Philoponus (*c*.512-17) *On Aristotle's Categories*, ed. A. Busse, *Commentaria in Aristotelem Graeca* XIII 1, Berlin: Reimer, 1898.(A *Categories* commentary typical of the school of Ammonius.)

John Philoponus (*c*.512-17) *On Aristotle's Prior Analytics*, ed. M. Wallies, *Commentaria in Aristotelem Graeca* XIII 2, Berlin: Reimer, 1905.(Purports to be based on Ammonius' seminars.)

John Philoponus (*c*.515-20) *On Aristotle's Posterior Analytics*, ed. M. Wallies, *Commentaria in Aristotelem Graeca* XIII 3, Berlin: Reimer, 1909.(Also purporting to be based on Ammonius, but there are signs of a later revision.)

John Philoponus (517) *On Aristotle's Physics*, ed. H. Vitelli, *Commentaria in Aristotelem Graeca* XVI- XVII, Berlin: Reimer, 1887-88; trans. A.R. Lacey, *Philoponus, On Aristotle's Physics 2*, London: Duckworth, 1993; trans. M. Edwards, *Philoponus, On Aristotle's Physics 3*, London: Duckworth, 1994; trans. P. Lettinck, *Philoponus, On Aristotle's Physics 5 to 8*, London: Duckworth, 1994.(This work represents the starting point for Philoponus' independent intellectual development; it includes important digressions on the notions of "place" and "void".)

John Philoponus (*c*.520-40) *On the Use and Construction of the Astrolabe*, ed. H. Hase, Bonn: E. Weber, 1839, and Rheinisches Museum für Philologie 6: 127-71, 1839; repr. and trans. A.P. Segonds, *Jean Philopon, traité de l'astrolabe*, Paris: Librairie Alain Brieux, 1981; trans. H.W. Green in R.T. Gunther (ed.) *The Astrolabes of the World*, Oxford: Holland Press, 1932, 2 vols; repr. London: Holland Press, 1976, 61-81.(The oldest extant Greek treatise on the astrolabe.)

John Philoponus (529) *On the Eternity of the World against Proclus (De aeternitate mundi contra Proclum)*, ed. H. Rabe, Leipzig: B.G. Teubner, 1899.(Philoponus' first major polemical treatise.)

John Philoponus (*c*.530-4) *On the Eternity of the World against Aristotle (De aeternitate mundi contra Aristotelem)*, trans. C. Wildberg, *Philoponus, Against Aristotle on the Eternity of the World*, London: Duckworth, 1987.(Not extant, this work consists of reconstructed fragments.)

John Philoponus (*c*.530-5) *On Aristotle's Meteorology*, ed. M. Hayduck, *Commentaria in Aristotelem Graeca* XIV 1, Berlin: Reimer, 1901.(This set of classroom lectures exemplifies how open criticism of Aristotle could at the same time be signalled and suspended.)

John Philoponus (*c*.530-5) *On the Contingency of the World (De contingentia mundi)*, trans. S. Pines, 'An Arabic summary of a lost work of John Philoponus', *Israel Oriental Studies* 2 (1972): 320-52; for similar excerpts in Simplicius, see D. Furley and C. Wildberg, *Philoponus, Corollaries on Place and Void with Simplicius, Against Philoponus on the Eternity of the World*, London: Duckworth, 1991, 95-141.(A short, non-polemical anti-eternalist treatise.)

John Philoponus (*c*.530-40) *On Nicomachus' Introduction to Arithmetic*, ed. R. Hoche, parts I/II, Wesel: Bagel, 1864, 1865, part III, Berlin: Calvary, 1867.(The only mathematical work attributed to Philoponus.)

John Philoponus (*c*.546-9) *On the Creation of the World (De opificio mundi)*, ed. W. Reichardt, Leipzig: Teubner, 1897.(The date of composition originally proposed by the editor appears more likely now than the

frequently suggested 557-60.)

John Philoponus (*c.*552) *Arbiter (Diaetetes)*, trans. A. Šanda, *Opuscula monophysitica Ioannis Philoponi*, Beirut: Typographia Catholica PP.Soc.Jesu, 1930; extracts trans. W. Böhm, *Johannes Philoponos, Grammatikos von Alexandrien*, Munich, Paderborn and Vienna: Schöningh, 1967, 414-29.(A philosophical justification of monophysitism; not extant in Greek, the former is Syriac text with Latin translation, the latter a German translation.)

John Philoponus (567) *On the Trinity (De trinitate)*, trans. A. Van Roey, 'Les fragments trithéites de Jean Philopon', *Orientalia Lovaniensia Periodica* 11 (1980): 135-63.(The main source for a reconstruction of Philoponus' trinitarian doctrine; not extant so Syriac fragments with Latin translation.)

References and further reading

Haas, F.A.J. de (1996) *John Philoponus on Matter. Towards a Metaphysics of Creation*, Leiden: Brill.(Offers a lucid discussion of Philoponus' conception of matter, its roots in the whole Neoplatonic-Aristotelian tradition, and its significance for Philoponus' doctrine of creation.)

Hainthaler, T. (1990) 'Johannes Philoponus, Philosoph und Theolog in Alexandria' (John Philoponus, Alexandrian Philosopher and Theologian), in A. Grillmeier (ed.) *Jesus der Christus im Glauben der Kirche, Band 2/4: Die Kirche von Alexandrien mit Nubien und Äthiopien nach 451*, Freiburg, Basle and Vienna: Herder, 109-49.(Particularly valuable as a survey of Philoponus' theological ideas.)

Lee, T.-S. (1984) *Die griechische Tradition der aristotelischen Syllogistik in der Spätantike (The Greek Tradition of Aristotelian Syllogistic in Late Antiquity)*, Göttingen: Vandenhoeck & Ruprecht.(A comparative evaluation of Alexander's, Ammonius' and Philoponus' commentaries on the *Prior Analytics*.)

Sambursky, S. (1962) *The Physical World of Late Antiquity*, London: Routledge & Kegan Paul, 154-75.(Standard work on late ancient theories concerning space, time, matter and mechanics.)

Scholten, C. (1996) *Antike Naturphilosophie und christliche Kosmologie in der Schrift De opificio mundi des Johannes Philoponos (Ancient Philosophy of Nature and Christian Cosmology in John Philoponus' Treatise De opificio mundi)*, Berlin: de Gruyter.(A careful analysis of the natural philosophy behind key concepts employed in Philiponus' commentary on the Biblical Creation story.)

Sorabji, R.R.K. (1983) *Time, Creation and the Continuum*, London: Duckworth.(Pages 193- 231 contain a lucid discussion of Philoponus' anti-eternalist infinity arguments.)

Sorabji, R.R.K. (ed.) (1987) *Philoponus and the Rejection of Aristotelian Science*, London: Duckworth. (A collection of twelve articles on Philoponus' physical theories, his theology and historical influence; extensive bibliography.)

Verbeke, G. (ed.) (1966) *Jean Philopon, Commentaire sur le de anima d'Aristote, traduction de Guillaume de Moerbeke (John Philoponus, Commentary on Aristotle's On the Soul*, translated by William of Moerbeke), *Corpus Latinum Commentariorum in Aristotelem Graecorum* III, Paris: Éditions Béatrice-Nauwelaerts; parts thereof trans. W. Charlton, *Philoponus, On Aristotle on the Intellect (de Anima 3.4-8)*, London: Duckworth, 1991. (This Latin commentary differs substantially from the Greek commentary *On the Soul*.)

Verbeke, G. (1985) 'Levels of Human Thinking in Philoponus', in C. Laga, J.A. Munitz and L. van Rompay (eds) *After Chalcedon: Studies in Theology and Church History*, Leuven: Peeters, 451-70.

Verrycken, K. (1990) 'The Development of Philoponus' Thought and its Chronology', in R.R.K. Sorabji (ed.) *Aristotle Transformed*, London: Duckworth, 233-74.(A clear exposition of the main ideas that mark Philoponus' intellectual development; the whole volume is an excellent guide to philosophy in late antiquity, with extensive bibliography.)

Wildberg, C. (1988) *John Philoponus' Criticism of Aristotle's Theory of Aether*, Berlin: de Gruyter.(An analysis of books I-V of *On the Eternity of the World Against Aristotle*.)

Wolff, M. (1971) *Fallgesetz und Massebegriff (The Law of Falling Bodies and the Concept of Mass)*, Berlin: de Gruyter.(A discussion of Philoponus' dynamics and theory of matter.)

Photography, aesthetics of

Claims that photography is aesthetically different from and, in many versions of the argument, inferior to the arts of painting and drawing have taken various forms: that photography is a mechanical process and therefore not an artistic medium; that it is severely limited in its capacity to express the thoughts and emotions of the artist; that its inability to register more than an instantaneous 'slice' of events restricts its representational capacity; that it is not a representational medium at all. Some of these arguments are thoroughly mistaken, while others have an interesting core of truth that will emerge only after some clarification. Central to this clarification is an account of the precise sense in which photography is mechanical, and an explication of our intuition that a photograph puts us 'in touch with' its subject in a way that a painting or drawing cannot. Both these ideas need to be separated from the mistaken view that it is the nature of photography to provide images that are superlatively faithful to the objects they represent. Throughout this entry we shall consider only the case of photographs of a relatively unmanipulated kind, ignoring darkroom techniques that blur the distinction between photography and painting.

1 The supposed limitations of photography

Photography and the traditional forms of image-making - painting and drawing - differ in important ways. Intuitively, a painting can be 'of' something that does not exist (though this idea is hard to make precise in an acceptable way); a photograph can only be of something that actually transmitted light into the mechanism. A photograph can depict more than the artist intended to depict, such as when magnification displays evidence of a crime; a painting records what the artist thought the scene looked like, and the grain of discriminability is set by the artist's own visual acuity and manual skill. Certain kinds of subject seem to have more force or cause more offence when represented photographically; a photograph of martyrdom would be difficult to appreciate aesthetically, whatever its compositional merits. It has not, however, been easy to explain these judgments within a unified theory of the nature and limitations of photography.

One kind of argument for the artistic inferiority of photography depends on claims about the limitations of what photography can convey - limitations which favour the particular, the instantaneous and the visible or surface features of things over the general, the progressive and the hidden aspects, and especially over the underlying emotional and motivational themes. Most of these arguments apply also to painting, an art against which they are rarely directed. It is true that painting can explicitly present a narrative by placing noncontemporaneous events within the same frame, while a photograph of the 'pure' or unmanipulated kind cannot. But even paintings that confine themselves to the depiction of a single moment can draw on beliefs and traditions external to the work to provide narrative force. Photographs can also achieve narrative status by appeal to such sources. It has also been argued that the 'mechanical' nature of the photographic method does not allow significant scope for expressive or intentional qualities. In exactly what sense photography is mechanical is a question we shall examine in detail later on, but any acceptable explication of that notion will surely have to accommodate the fact that some photographs do express beliefs, evaluations and emotions through the evident choices the photographer has made concerning such things as lighting, depth of field, exposure and subject. It may be true that there are fewer choices to be made in determining the visible appearance of a photograph than in determining the visible appearance of a painting, where every brush stroke is, in principle, a matter of independent decision. It does not follow that photography is a less expressive medium than painting. Sometimes the restrictions that an artistic form imposes on its practitioners give a special significance to the choices they are able to make within the constraints imposed. Arguments to the effect that photography is, by its nature, artistically limited are most charitably understood as reactions against the excessively ambitious claims sometimes made on behalf of photography: that it makes representational painting redundant, that it has an unrivalled capacity for documentation, or that it represents the world to us in an undistorted and uninterpreted form.

The claim that photography faithfully reproduces appearance, in the sense that photographs of things are or can be superlative likenesses of them, needs to be distinguished from another thesis with which it is sometimes confused: that photography mechanically reproduces the appearances of things. Both these claims have been used to advance a further claim: that photography gives us access to things in the real world rather than providing, as painting does, mere representations of those things. Let us call these claims the likeness thesis, the mechanicity thesis and the reality thesis respectively. Critics of the reality thesis have attacked both the mechanicity thesis and the likeness

thesis. Against the likeness thesis I believe they have been successful, having shown that there is nothing about photography as a medium that makes its products essentially more faithful than painting to visible appearances. Against the mechanicity thesis, however, the critics have been less successful, their efforts having been directed against its naive and implausible versions. But while the mechanicity thesis stands, it provides little support for the reality thesis, as we shall see.

2 Photography as mechanism

Claims that photography is a mechanical medium are common; clear expositions of the thesis are rare. Perhaps the clearest has been offered by Kendall Walton (1984).Compare a painting and a photograph of a rose. In both cases there is a dependence between the appearance of the rose and the appearance of the picture; if the rose had been different in various ways, the picture would have been different, and so, correspondingly, would the viewer's experience of looking at the picture. But there is a difference between the painting and the photograph: the patterns of dependence that hold in the case of the painting hold in virtue of facts about the painter's mental state. If the rose had had more petals, the painting would have shown more - but only because the painter would have mentally registered the greater number and incorporated it into the painting. In testing for dependence in the case of painting, if we vary the appearance of the rose, but hold fixed the painter's beliefs about its appearance, the appearance of the painting does not vary. Things are different with the photograph; holding fixed the photographer's mental state leaves the dependence between scene and photograph intact. Had the rose had more petals, the photograph would have shown more, regardless of whether the photographer noticed their number. A hallucinating painter will paint what they think is there; a photograph will show what is there. Photography is mechanical in that the relevant counterfactuals are independent, in the sense just described, of mental states. Here we shall concentrate on this version of the claim that photography is mechanical.

Having explicated the mechanicity thesis, let us now make the reality thesis comparably plausible and precise. Some versions of the reality thesis seem to assert obvious falsehoods. André Bazin's claim (1967) that 'the photographic image is the object itself' is hard to reconcile with the fact that the Washington Monument and a photograph of it are utterly distinct and possibly very distant from each other in space and time. A better explication is provided once again by Kendall Walton, according to whom photographs are 'transparent': seeing a photograph of X is a case of *really seeing* X, just as seeing X through a window or through a telescope and seeing X's reflection in a mirror are such cases. On the other hand, seeing a painting or drawing of X is not a case of seeing X. Photographs are 'aids to vision', whereas 'handmade' images are ways of representing. Roger Scruton (1981) has likened photography to the artful arrangement of a frame around a street scene; what we see inside the frame may be pleasing, but it is not a representation. It is the same, he says, with the photograph. Scruton is best understood as asserting, not that the photograph of the street is, or contains, the street, but that seeing the photograph is a way of seeing the street. This way of understanding the reality thesis does not require the photograph to be identified with or even overlap with its subject; we may see someone in a mirror without having to suppose that the mirror is part of the person. Nor does it require spatiotemporal proximity between photograph and subject; we now see stars that ceased to exist millions of years ago. It is this version of the reality thesis which I shall consider. I shall call it the transparency thesis.

Let us see how, on Walton's construal, the mechanicity thesis supports the transparency thesis. Recall that the patterns of dependence holding between the rose and the photograph of the rose hold also between the rose and my experience of seeing the photograph of the rose, and that these dependencies hold independently of the beliefs of the photographer. Let us say that in that case my experience of the photograph is 'mechanically dependent' on the state of the rose. By contrast, my experience of looking at the painting of the rose is 'intentionally dependent' on the state of the rose; if we imagine the appearance of the rose to vary but hold fixed the painter's mental states, my visual experience of seeing the painting does not vary. Now consider the case where I directly see a rose in front of my eyes, through a window, through a lens or in a mirror. All these are cases of mechanical dependence; they are also cases of genuinely seeing the rose. Seeing the photograph of the rose shares this mechanical dependence, but seeing the painting of one does not. So seeing the photograph of the rose is to count as genuinely seeing the rose, but seeing the painting is not.

This nexus of views about the mechanicity and transparency of photography has the advantage of not denying certain facts about photography which are sometimes cited in favour of the view that there is not, after all, any

principled difference between photography and other representational media like painting: for example, that people typically choose the subject, location and lighting conditions of the photographs they take, and that these choices can all be made with more or less skill, aesthetic sensibility and expressive effect. While true, these considerations are largely irrelevant to the thesis that photographs are mechanical and transparent. When we look at something with the naked eye, how it looks to us depends on lighting conditions and on spatial relations between the viewer and the object viewed, and we can choose to look at this rather than at that. None of this impugns the claim that seeing things with the naked eye is in a certain sense a mechanical process: that when we look in a certain direction it is not up to us or to anyone else to decide what we see; what we see is determined 'mechanically' by what is in front of our eyes and not by decisions or preferences. Nor is it an objection to the claims of mechanicity and transparency that photographs do not always look very much like - and never look exactly like - the things they are photographs of. A photograph may grossly distort the appearance of a man by being created with the help of a distorting lens. But you can look at a man through a distorting lens, thereby seeing him in a grossly distorted way: you are still seeing the man. Nor need the advocate of mechanicity and transparency deny that skills are exercised in the production of photographs, and that photographs so produced may exhibit these skills.

Does the mechanicity of photography provide a convincing argument for its transparency? Imagine two indiscernible clocks, A and B, where mechanical connection ensures that the time told by the hands of A causally determines the time told by those of B. If I am looking at clock B, then it is true of my current visual state that, if the hands of A told a different time, my current visual awareness of B would have been correspondingly different. And these counterfactuals, we may assume, are independent of the intentions of agents in the sense described above. But my seeing B is not also a case of my seeing A. So mechanicity is not sufficient for transparency.

Walton grants the insufficiency of mechanicity for transparency; there might, he concedes, be a machine which examines objects and churns out linguistic descriptions of them in such a way that there is mechanical dependency between the object and the descriptions produced; but reading the descriptions would not be a case of seeing, or of perceiving in any way, the thing described. What stops reading the description from being a case of seeing the object described is, Walton claims, the lack of relevant similarities between the experience of reading and the experience of seeing the object itself. Here we are close to the assertion that mechanicity and likeness (in some relevant respect) jointly constitute a sufficient condition for transparency. But that is not so: the experience of seeing clock B from our earlier example is, in every relevant respect, very much like the experience of seeing clock A, but seeing clock B is not a case of seeing clock A. This is not to say that the transparency thesis is false, but merely that it is not supported by the mechanicity thesis.

See also: Film, aesthetics of; Semiotics

GREGORY CURRIE

References and further reading

Batkin, N. (1991) 'Paul Strand's Photographs in *Camera Work*', *Midwest Studies in Philosophy* 16, ed. P. French, T. Uehling and H. Wettstein, Notre Dame, IN: Notre Dame University Press, 1991.(Combines philosophical analysis of photographic representation with critical reflection on an important moment in the history of photography.)

Bazin, A. (1967) 'The Ontology of the Photographic Image', in *What is Cinema?*, vol. 1, Berkeley, CA, and Los Angeles, CA: University of California Press, a translation of selected essays from *Qu-est-ce que le cinéma?* 4 vols, 1958-65, Paris: Éditions du Cerf.(Advocates a version of the reality thesis.)

Currie, G. (1991) 'Photography, Painting and Perception', *Journal of Aesthetics and Art Criticism* 49: 23-9. (Argues that the mechanicity thesis does not support the reality thesis.)

Scruton, R. (1981) 'Photography and Representation', *Critical Inquiry* 7: 577-603; repr. in *The Aesthetic Understanding*, New York: Methuen, 1983.(Argues that photography is a representationally and expressively impoverished form.)

Snyder, J. and Allen, N.W. (1976) 'Photography, Vision and Representation', *Critical Inquiry* 2: 143-69.(Argues against both the likeness thesis and the mechanicity thesis, without distinguishing between them very clearly.)

Walton, K. (1984) 'Transparent Pictures. On the Nature of Photographic Realism', *Critical Inquiry* 11: 246-77. (Argues that the mechanicity thesis supports the reality thesis, while importantly clarifying both.)

Physis and nomos

In the fifth and fourth centuries BC *a vigorous debate arose in Greece centred on the terms* physis *(nature) and* nomos *(law or custom). It became the first ethical debate in Western philosophy. Is justice simply a matter of obeying the laws, or does it have some basis in nature? If the laws conflict with my natural needs and desires, why should I submit to them? Is society itself 'natural', and what difference might the answer make to our evaluation of it? Both* nomos *and* physis *had their supporters, while some tried to dissolve the antithesis altogether.*

Physis has two principal meanings. It can refer to the cosmic order as a whole, or to the prime constituent(s) of that order: in these contexts it can often be translated as 'reality'. It can also refer to the intrinsic characteristics of a thing, especially a living thing, or to the thing's growth towards these characteristics. This dynamic aspect of *physis* can seem to give it a prescriptive as well as a descriptive force: it is *good* that things grow towards their mature state and achieving it constitutes their flourishing.

Nomos (plural: *nomoi*) can signify both the unwritten customs and the written laws of a society, and sometimes also an unwritten universal law of divine origin. Even when it refers to a custom, its force is always prescriptive: it indicates not merely a practice, but what the majority accepts as the right practice. However, increasing travel and historical researches heightened awareness of how particular *nomoi* differ between cultures and over time, and the consequent sense of their transience prompted growing speculation about their authority.

The debate was entered by playwrights, orators and historians as well as philosophers. In Sophocles' *Antigone* (450-60), the particular laws of the state are opposed to the universal ordinances (*nomima*) of the gods. Normally, however, the contrast was between the *nomoi* of a particular state and the perceived requirements of human nature - a theme so common that sliding from what is right in law to what is right by nature and vice versa became a standard debating move (see Aristotle, *Sophistical Refutations* 173a7-16). One leading supporter of *physis* was Antiphon, who argued that the claims of the laws are generally antithetical to those of human nature, and that the individual's interests will be best served by following the latter when legal punishment can be avoided, especially given the inevitability of nature's reprisals if flouted, compared to the uncertain outcome of a broken law.

Amongst other champions of *physis*, Callicles in Plato's *Gorgias* makes a radical distinction between conventional justice and 'natural justice'. It is a 'law of nature' that the naturally superior should seize political control; they should also freely indulge their physical desires. Such behaviour is only termed unjust and intemperate by a convention of the weak majority. The notion of a 'law of nature' also occurs in Thucydides' Melian dialogue (V 105.1-3), where the Athenians claim that it is simply an empirical fact that the superior power can do what it wishes; considerations of justice are only relevant among equals. Thrasymachus, however, in Plato *Republic* I, holds that justice is the invention of those in power: they make the laws in their own interests and call obedience to these laws 'justice'. The prudent will on the contrary further their own interests, acting unjustly where necessary. This position is stated in terms of the *nomos*/*physis* debate at *Republic* 359c.

Underlying all these differing positions is a view of humanity as naturally in conflict. The same premise, however, was also employed by supporters of *nomos*, such as Critias and Protagoras (§§2-3). Both argue that without *nomos* life would be chaotic, brutish and short: humans cannot survive unless they live in societies, and societies cannot exist without an agreed moral code. Protagoras, however, seems to offset people's natural aggression with their innate capacity to absorb social training. Democritus (§5) further advocates that people internalize the *nomoi* of society so that these become the *nomos* of the soul.

Protagoras' position has links with the social contract theories prevalent during this period (see, for example, the celebrated position put forward by Glaucon in Plato's *Republic* 358e-359b). In Plato's *Crito*, Socrates argues that each citizen, by freely electing to remain in their city and receive the benefits of its protection, has made an implicit contract to abide by its *nomoi*. Perhaps the strongest defence of *nomos*, however, comes with Plato and Aristotle, who in different ways seek to dissolve the *nomos*/*physis* antithesis by claiming that receptivity to *nomoi* is an essential aspect of human *physis*.

The debate had important political implications. Each term was used to serve both oligarchic and democratic causes. Traditionally *physis* had been employed by aristocrats to support the notion of a superior class naturally

fitted to rule, but by the late fifth century BC appeals to *physis* were also being used to question the social hierarchy. Antiphon claimed that there was no natural distinction between the high- and low-born, or between Greek and non-Greek, and in the mid fourth century Alcidamas attacked as unnatural the institution of slavery. But the defenders of these social distinctions also appealed to *physis*, as Aristotle does regarding slavery in *Politics* book I.

The liberal deployment of *physis* was associated with the democratic belief that human excellence (see Aretē) could be taught through immersion in the *nomoi* of the state, rather than being acquired through aristocratic inheritance. The supremacy of written *nomoi* was also upheld by democratic supporters, while oligarchs increasingly claimed backing from the controversial unwritten *nomoi*.

Issues of epistemology and language were also embraced. Democritus claimed that sensible properties exist only by convention, while in Plato's *Cratylus* it is considered whether there is a 'natural name' for each thing, or whether this too is a matter of *nomos* (see Language, ancient philosophy of). Nor were the gods above scrutiny: a fragment of the lost play *Sisyphus* (possibly by Critias) argues that the gods are a political invention to promote social control.

The debate raises tough philosophical problems. Is the derivation of an ethical prescription from appeals to nature legitimate? (See Naturalism in ethics.) Some scholars hold that the teleological connotations of *physis* make such a derivation at least plausible. Again, can human nature even be distinguished from *nomos*, either conceptually or in practice?

Despite such difficulties, the debate initiated an important tradition in ethical and political thought which required that perceived human needs and desires at least be taken into consideration, even if they are ultimately to be transcended.

ANGELA HOBBS

References and further reading

Classen, C.J. (ed.) (1976) *Sophistik*, Darmstadt: Wissenschaftliche Buchgesellschaft.(A useful collection of German and English articles.)

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, vol. 3, Cambridge: Cambridge University Press; part of vol. 3 repr. as *The Sophists*, Cambridge: Cambridge University Press, 1971.(A thoughtful and scholarly exploration; see especially pages 55-134.)

Heinimann, F. (1945) *Nomos und Physis*, Basle: Reinhardt.(A classic study in German which places the debate in its intellectual and historical context.)

Kerferd, G.B. (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press. (Clear and stimulating; the *nomos*/*physis* debate is discussed particularly in chapter 10.)

McKirahan, R.D. (1994) *Philosophy Before Socrates*, Indianapolis, IN: Hackett.(A helpful collection of translated texts with a lucid commentary; for the *nomos*/*physis* debate see chapter 19.)

Piaget, Jean (1896-1980)

The Swiss psychologist Jean Piaget was the founder of the field we now call cognitive development. His own term for the discipline was 'genetic epistemology', reflecting his deep philosophical concerns. Among Piaget's most enduring contributions were his remarkably robust and surprising observations of children. Time after time, in a strikingly wide variety of domains, and at every age from birth to adolescence, he discovered that children understood the world in very different ways from adults.

But Piaget was really only interested in children because he thought they exemplified basic epistemological processes. By studying children we could discover how biological organisms acquire knowledge of the world around them. The principles of genetic epistemology could then be applied to other creatures, from molluscs to physicists. Piaget's other enduring legacy is the idea that apparently foundational kinds of knowledge were neither given innately nor directly derived from experience. Rather, knowledge was constructed as a result of the complex interplay between organisms and their environment. Piaget saw this view as an alternative to both classical rationalism and empiricism.

1 The stages of development

Piaget described development as a series of wide-ranging stages. At each stage children had quite different basic logical and representational capacities, which underpinned all their behaviour. Infants initially arrived in the world with nothing but motor reflexes and sensory systems. During infancy they elaborated these basic capacities into complex contingent instructions for behaviour, which related actions and sensations. Piaget called these 'sensori-motor schemas'. Infants could make substantial progress even within this apparently limited framework. They learned a great deal about objects, space, action and causality. Nevertheless, there were, quite literally, no representations of the world in infancy, and no conception of a world independent of the infant's experience. There were only instructions for action.

Piaget defended this position with examples of apparently peculiar and irrational behaviour in infancy. One set of examples involved infants' attempts to obtain interesting objects. Six month-old infants behaved as if objects that disappeared under a cloth no longer existed. If they were in the midst of reaching for an object and it was hidden under a cloth, they would give up their attempt to get it. Later, they learned that pulling the cloth would make the object reappear. But if the object was hidden under one cloth and then under another they would continue to search under the original cloth. Similarly, 9-month-olds who learned to pull a blanket towards them to get an object on top of the blanket would continue to do so even when the object was placed to one side of the blanket.

Children only developed the capacity to represent the external world symbolically at about 18 months. This new kind of representation was reflected in the development of language, pretend play and deferred imitation. These symbolic representations also allowed children to understand that there was a world independent of their experience of it. And they enabled children to solve problems 'in their heads', without actually having to act on the world.

These representations, however, initially had little logical or causal structure. They were 'pre-operational'. As a result preschool children were unable to perform even simple logical and causal reasoning. When they were asked to explain natural phenomena they resorted to animistic explanations, or to explanations in terms of their own desires, and they might reverse cause and effect. For example, a 3-year-old child who was asked why it got dark at night might reply that it got dark because we need to sleep or because you could not see then. Similarly, preschool children were unable to understand even simple logical principles such as the law of transitivity. If they were told that stick *A* was longer than stick *B* and stick *B* than stick *C*, they might still deny that *A* was longer than *C*. Finally, these very young children showed little understanding of hierarchical relationships, such as class-inclusion. If you showed them an array of four red flowers and two white ones, they would report that there were more red flowers than flowers.

At about 6, children did begin to use causal and logical reasoning. These kinds of reasoning were, however, initially closely tied to the perceptual appearances of objects. They were 'concrete operational'. The most famous examples of this involved 'conservation'. In a conservation task the properties of objects remain the same in spite

of transformations of the object's appearance. In a classic experiment, Piaget showed children a tall, thin glass of water, poured the water into a short, wide glass and asked children whether there was the same amount of water, more, or less. Seven-year-olds consistently said that there was less water in the short glass, even after the water was poured back and forth from glass to glass. The children were unable to use information about relations among objects and transformations of objects to override perceptual information about the features of objects.

Piaget thought concrete operations were reflected in the child's understanding of psychological states as well as physical objects. Just as school-age children were unable to coordinate different perceptual appearances, they were unable to coordinate their own experiences and those of others (see Mind, child's theory of). They were 'egocentric'. If 7-year-olds were shown a complex scene from one perspective, they could not project how the scene would look to another person at a different viewpoint. Similarly, in moral reasoning children calculated harm by looking at the actual amount of damage that was caused, without considering intention or motivation (see Moral development). Only at adolescence, in the stage of formal operations, did children become able to use logical and scientific reasoning in a way that was fully abstracted from the details of their own experience.

2 The mechanisms of development

Piaget rejected both the classical developmental mechanisms of nativist views, such as maturation, and those of empiricist views, such as association or reinforcement. Moreover, he also rejected, or at least underplayed, the influence of language and social interaction. Instead he proposed three very general mechanisms of development, all involving an interaction between the knowledge the child had already developed and new information from the outside world. Piaget called his overall view of development 'constructivism'.

Assimilation. The child adapted and interpreted information from the world to fit his existing schemas. The manifestations of assimilation ran from a newborn who generalized his sucking reflex from a nipple to a rattle, to a preschool child who used blocks as dolls in her pretend play, to a school-age child in a conservation task who simply misreported counter-evidence to his claims about the water.

Accommodation. The opposite effect took place: the child was forced to adapt his schemas to fit new information from the world. Again these effects could vary from an infant who had to change his style of sucking to accommodate the unyielding rattle, to a preschool child who imitated her mother's phone conversation in an initially uncomprehending way, to a child in a conservation task who at least temporarily admitted that the same water was in both glasses.

Equilibration. This really represented a kind of balance between assimilation and accommodation. When the child's representations and the evidence from the world matched, the process of accommodation and assimilation would end and the child could at least temporarily settle on a particular kind of representation. Piaget saw cognitive development as a highly dynamic process, constantly balancing representations that had already been constructed with new input from the outside world.

Moreover, the child's active attempts to interact with the world were the basic motor driving this dynamic process. Assimilation and accommodation could work only if the child were engaged with the world, usually by physically acting on it. Piaget sometimes illustrated the mechanisms with the biological example of an animal literally assimilating part of its environment by eating it, and literally accommodating the food by physically changing its body as a result.

3 Piaget's influence

Much of Piaget's most significant work was produced in the 1920s and 1930s in Geneva. But Piaget, like the Gestalt psychologists, remained largely unknown in the United States during the long reign of behaviourism (see Behaviourism, methodological and scientific; Gestalt psychology). However, a few Piagetians, like Gestaltists, came to America after the war, and kept the tradition alive in small enclaves in liberal arts colleges. In the 1960s Piaget was rediscovered in the United States and fully translated into English. His rediscovery was itself both a consequence and a cause of the 'cognitive revolution' in psychology. Piaget was one shining example of how you could construct an account of the mind that was mentalistic, but that also emphasized the biological foundations of psychology. Unlike earlier psychologists such as Brentano and Wundt, but like Chomsky and the early computationalists, he detached the idea of mentalistic explanation from phenomenology (see Mind, computational

theories of; Unconscious mental states).

Moreover, Piaget's work led to the recognition that developmental questions were of broad interest and importance in psychology (until the 1960s much developmental psychology was located in home economics departments). Piaget also had a strong influence on education, particularly on 'progressive' educational theory and practice.

4 After Piaget

Where does the theory stand now? Most of Piaget's actual observations, often based only on pen and paper notes of a few children, have held up remarkably well. On the other hand, new techniques have given us a much broader additional database. Most often, the new data suggest that children are more competent and sophisticated in their understanding of the world than Piaget supposed. These new observations have led most cognitive developmentalists to reject the details of Piagetian theory (see Cognition, infant; Cognitive development).

In particular, the idea of broad-ranging stages of development has come under increasing attack. Instead, cognitive development appears to be quite specific to particular domains of knowledge. For example, the infant who fails to understand object permanence may show a sophisticated and clearly representational understanding of object movement or of human action. The same child who gives an animistic or pre-causal explanation of the fact that it gets dark at night will give an entirely accurate causal account of how their tricycle works. The preschool child who is egocentric in Piaget's experiments can tell you that someone on the other side of a screen from them will not be able to see what they themselves see. Even newborn infants show a rich understanding of some domains.

Similarly, it appears that Piaget overestimated the importance of action in cognitive development, particularly in infancy. Very young infants show signs of reasoning and learning about objects, well before they can act on them. On the other hand, Piaget probably underestimated the importance of social interaction and language in cognitive development. Much of the new research suggests that children are extremely sensitive to social information and are tuned in to the social world from an early age. More generally, Piaget's account of assimilation and accommodation as the basic mechanisms of 'constructivism' now seems too vague.

Some developmentalists have seen these problems as a reason for rejecting not only the detailed theory but the project of 'constructivism' - of a middle way between classical nativism and empiricism - altogether. Many have returned to one version or another of the classical philosophical alternatives. Modularity theories, for example, are a variant of classical rationalism (see Modularity of mind). Connectionist or dynamic systems theories, or some version of social constructivism, are a variant of classical empiricism (see Connectionism; Empiricism). However, most developmentalists would probably prefer to revise Piaget's general view of development rather than to replace it. Perhaps the clearest contemporary theoretical legacy of Piaget is the idea that cognitive development is the result of the same mechanisms that lead to theory change in science (see Scientific method; Kuhn, T.). Testing or confirming the theory by experiment might be a modern version of assimilation, while revising the theory in the light of counter-evidence would be more like accommodation. Piaget's most enduring legacy of all, however, was the idea that the grand questions of 'genetic epistemology', questions that date back to Socrates, could be answered by paying attention to the small details of the daily lives of our children.

ALISON GOPNIK

List of works

Piaget, J. (1926) *The Language and Thought of the Child*, London: Kegan Paul, Trench, Trubner & Co; New York: Harcourt Brace.

Piaget, J. (1929) *The Child's Conception of the World*, London: Kegan Paul, Trench, Trubner & Co; New York: Harcourt, Brace.

Piaget, J. (1930) *The Child's Conception of Physical Causality*, London: Kegan Paul, Trench, Trubner & Co.

Piaget, J. (1948) *The Moral Judgement of the Child*, trans. M. Gabain, Glencoe, IL: Free Press.

Piaget, J. (1952) *The Child's Conception of Number*, trans. C. Gattegno and F.M. Hodgson, London: Routledge & Kegan Paul.

Piaget, J. *et al.* (1952) *Judgment and Reasoning in the Child*, New York: Humanities Press.

Piaget, J. (1952) *The Origins of Intelligence in Children*, trans. M. Cook, New York: International Universities Press.

Piaget, J. (1954) *The Construction of Reality in the Child*, trans. M. Cook, New York: Basic Books.

Piaget, J. and Inhelder, B. (1956) *The Child's Conception of Space*, trans. F.J. Langdon and J.L. Lunzer, London: Routledge & Kegan Paul.
Piaget, J. (1957) *Logic and Psychology*, with an introduction on Piaget's logic by W. Mays, New York: Basic Books.
Piaget, J., Inhelder, B. and Szeminska, A. (1960) *The Child's Conception of Geometry*, trans. E.A. Lunzer, London: Routledge & Kegan Paul.
Piaget, J. (1962) *Play, Dreams, and Imitation in Childhood*, trans. C. Gattegno and F.M. Hodgson, New York: Norton.
Piaget, J. (1967) *The Psychology of Intelligence*, trans. M. Piercy and D.E. Berlyne, London: Routledge & Kegan Paul.
Piaget, J. (1968) *On the Development of Memory and Identity*, trans. E. Duckworth, Worcester, MA: Clark University Press.
Piaget, J. and Inhelder, B. (1969) *The Psychology of the Child*, trans. H. Weaver, New York: Basic Books.
Piaget, J. (1969) *The Mechanisms of Perception*, trans. G.N. Seagrim, New York: Basic Books.
Piaget, J. (1970) *The Child's Conception of Time*, trans. A.J. Pomerans, New York: Basic Books.
Piaget, J. (1970) *The Child's Conception of Movement and Speed*, trans. G.E.T. Holloway and M.J. Mackenzie, New York: Basic Books.
Piaget, J. (1970) *Genetic Epistemology*, trans. E. Duckworth, New York: Columbia University Press.
Piaget, J. (1970) *Science of Education and the Psychology of the Child*, trans. D. Coltman, New York: Orion Press.
Piaget, J. (1970) *Structuralism*, trans. and ed. C. Maschler, New York: Basic Books.
Piaget, J. (1971) *Psychology and Epistemology*, trans. A. Rosin, New York: Grossman.
Piaget, J. (1971) *Insights and Illusions of Philosophy*, trans. W. Mays, New York: World Publishing.
Piaget, J. (1972) *The Principles of Genetic Epistemology*, trans. W. Mays, London: Routledge & Kegan Paul.
Piaget, J. (1973) *The Child and Reality: Problems of Genetic Psychology*, trans. A. Rosin, New York: Grossman.
Piaget, J. and Inhelder, B. (1973) *Memory and Intelligence*, trans. A.J. Pomerans, New York: Basic Books.
Piaget, J. (1974) *Understanding Causality*, trans. D. Miles and M. Miles, New York: Norton.
Piaget, J. and Inhelder, B. (1974) *The Child's Construction of Quantities: Conservation and Atomism*, trans. A.J. Pomerans, London: Routledge & Kegan Paul.
Piaget, J. and Inhelder, B. (1975) *The Origin of the Idea of Chance in Children*, trans. L. Leake, Jr., P. Burrell and H.D. Fishbein, New York: Norton.
Piaget, J. (1975) *Biology and Knowledge: An Essay on the Relations between Organic Regulations and Cognitive Processes*, trans. B. Walsh, Chicago, IL: University of Chicago Press.
Piaget, J. (1976) *The Grasp of Consciousness: Action and Concept in the Young Child*, trans. S. Wedgwood, Cambridge, MA: Harvard University Press.
Piaget, J. (1977) *The Development of Thought: Equilibration of Cognitive Structures*, trans. A. Rosin, New York: Viking Press.
Piaget, J. (1978) *Success and Understanding*, trans. A.J. Pomerans, Cambridge, MA: Harvard University Press.
Piaget, J. (1978) *Behavior and Evolution*, trans. D. Nicholson-Smith, New York: Pantheon Books.
Piaget, J. (1980) *Adaptation and Intelligence: Organic Selection and Phenocopy*, trans. S. Ames, with foreword by T.A. Brown, Chicago, IL: University of Chicago Press.
Piaget, J. (1980) *Experiments in Contradiction*, trans. D. Coltman, Chicago, IL: University of Chicago Press.
Piaget, J. and Ackermann-Valladao, E. (1987) *Possibility and Necessity*, trans. H. Feider, Minneapolis, MN: University of Minnesota Press.
Piaget, J. and Garcia, R. (1989) *Psychogenesis and the History of Science*, trans. H. Feider, New York: Columbia University Press.
Piaget, J. and Garcia, R. (1991) *Toward a Logic of Meanings*, ed. P.M. Davidson and J. Easley, Hillsdale, NJ: Erlbaum.

References and further reading

Flavell, J. (1968) *The Developmental Psychology of Jean Piaget*, Princeton, NJ: Van Nostrand.(One of the first and still the best account of Piaget's psychological theory. Piaget himself has a Kantian style that can be very rough-going. Flavell is the best way to get a clear sense of the general view.)
Flavell, J., Miller, P. and Miller, S. (1993) *Cognitive Development*, Englewood Cliffs, NJ: Prentice Hall, 3rd edn. (A contemporary account of cognitive development post-Piaget, written at a quite introductory level.)

Pico della Mirandola, Giovanni (1463-94)

Giovanni Pico della Mirandola, today the best known of Renaissance philosophers, was a child prodigy and gentleman scholar who studied humanities, Aristotelianism and Platonism with the greatest teachers of his day. He claimed to have mastered, by the age of twenty-four, all known theological systems, Christian and non-Christian, from Moses to his own time. He was the first important Christian student of the Jewish mystical theology known as Kabbalah. The purpose of Pico's philosophical and theological studies was to produce a grand synthesis of religious wisdom which would both deepen understanding of Christian truth and also serve as an apologetic weapon against non-Christians. This was the project outlined in Pico's most famous work, De dignitate hominis (On the Dignity of Man) (1486), and further illuminated by his Conclusiones (1486) and Apologia (1487). As part of this larger project, Pico planned to write a concord of Plato and Aristotle, of which only a fragment, the treatise De ente et uno (On Being and the One) (1491), was ever finished. Although he proposed to found a new theological school based on an esoteric reading of all theologies past and present, he did not believe that these theologies were the same in substance, differing only in expression. He insisted on the differences between Platonism and Christianity, while holding that every major theological tradition did contain some elements of truth.

In addition to other, non-philosophical works, Pico wrote the Commento (1486), a commentary on a Neoplatonic poem that in effect constituted a critique of Marsilio Ficino's most famous work, the dialogue De amore (On Love) (1469). He criticized Ficino as too literary and defended the use of precise technical language in philosophy. Pico used Neoplatonic metaphysics to rediscover the 'secret mysteries' of pagan theology (though he sometimes criticized the reliability of the Neoplatonists as guides to Plato's thought) and offered a fresh interpretation of the metaphysics of love based on his own reading of Platonic sources, seeing human erotic love as a psychological process distinct from cosmic love.

1 Life and works

Giovanni Pico della Mirandola was born in the small north Italian principality of Mirandola in 1463. Though he eventually succeeded an older brother as Count of Mirandola and Concordia, he was originally intended by his parents for an ecclesiastical career. With this end in view he was sent in 1477 to the University of Bologna to study canon law. In 1479 he left Bologna without taking his degree and went to Ferrara, where he studied with the famous humanist schoolmaster Battista Guarino. By 1480 he was in Padua hearing lectures on Aristotelian philosophy and studying privately with the Jewish Averroist Elijah Delmedigo (see Averroism, Jewish). He also at this time formed a lasting friendship with the great Venetian scholar and humanist Ermolao Barbaro. Barbaro tried to interest Pico in his own project of freeing Aristotle from the 'corrupt' interpretations of medieval Arabic and Latin commentators with the help of Aristotle's ancient Greek commentators. In 1482 Pico wrote to Marsilio Ficino asking for a copy of his recent *Theologia platonica (Platonic Theology)* (1474). He evidently found the work interesting, for in 1484 he came to Florence to sit at Ficino's feet and master the new Platonic sources that Ficino was making available to the Latin world. In 1485 Pico went to Paris to study medieval scholastic philosophy, returning to Florence a year later, where he began to study Hebrew and Arabic with the Jewish Kabbalist Flavio Mithridates. After a scandalous episode in which he attempted to abduct the wife of an Aretine tax official, Pico settled in Perugia where he continued his Hebrew and Kabbalist studies under the tutelage of several Jewish teachers.

Pico's education thus exposed him to practically every sort of philosophical subculture that high Renaissance Europe could offer: humanism, Paduan Averroism, the theological traditions of the great religious orders (particularly Thomism), and the more esoteric traditions of Florentine Neoplatonism and medieval Jewish Kabbalah. Pico mastered each of these traditions without himself being mastered by any of them. Despite his enviable knowledge of Greek, Arabic and Hebrew, and his ability to write elegant humanist Latin, Pico never endorsed the shallow humanist tendency to dismiss scholastic philosophy because of its neglect of good Latinity and its unconcern with the study of Greek philosophy in the original. In a famous letter to Ermolao Barbaro (1485), Pico praised the great medieval philosophers and defended the right of philosophers to use their own precise technical language. Pico was equally discriminating in his use of Ficinian Platonism. As he wrote in another letter to Barbaro (echoing Cicero), he came to his studies with Ficino 'not as a deserter' from Aristotle

'but as a spy' (*explorator*). His own doctrines, though clearly indebted to Ficino as well as to ancient Neoplatonists, were at the same time distinctive, and Pico did not hesitate to criticize even the most authoritative ancient sources, such as Plotinus and Proclus, when he believed that they had misread Plato (see Platonism, Renaissance).

Pico's career as a philosophical writer began in 1486, when he composed the original draft of his *Commento sopra una canzona de amore composta da Girolamo Benivieni secondo la mente et opinione de' Platonici (Commentary on a poem about love composed by Girolamo Benivieni, according to the understanding and opinion of the Platonists)*, known as the *Commento* for short. This work contained extracts from three other works, never published and now presumably lost: a commentary on Plato's *Symposium*, a poetic theology (a reconstruction of 'ancient theology' based on his reading of the pagan myths) and a treatise on love. In 1486 Pico also conceived the idea of holding a great public disputation (or 'council' as he referred to it) in Rome in which he would reveal a new theological system based on a profound study of all philosophies and theologies past and present. In aid of this project he published the *Conclusiones*, a list of 900 theses drawn from Christian, Jewish, Muslim, Platonic, Aristotelian, Hermetist, Orphic, Zoroastrian, Kabbalist and other sources. He announced that he would defend his theses publicly early in 1487, and summoned to the disputation the Pope, the College of Cardinals, and all philosophers and theologians in Italy, offering to pay the travelling costs of any challengers. He also composed, in December 1486, the so-called *De dignitate hominis (Oration on the Dignity of Man)* which was intended as an introductory speech to open the 'council'. This was to become Pico's most famous work.

On 20 February 1487 Pope Innocent VIII, at the instigation of some theologians at his court, ordered Pico's disputation suspended and appointed a commission of inquiry which in due course condemned thirteen of Pico's theses. Pico responded by publishing his *Apologia* in defence of the condemned theses. In a bull dated 4 August 1487 but not (it seems) published until November, Innocent finally condemned all 900 theses *en bloc* and ordered Pico arrested for heresy. By this time Pico had escaped to France, where he was detained by the French king at the Pope's request. The philosopher was eventually allowed to return to Florence (1488) to live under the protection of Lorenzo de'Medici, who had intervened on his behalf with the Pope. Apart from brief visits to Ferrara, Pico spent the last six years of his life at Lorenzo's 'court', attempting to carry forward his project of making a concord of all theologies. In 1489 he published *Heptaplus de opere sex dierum Geneseos (On the Sevenfold Narration of the Six Days of Genesis)*, an attempt to turn the first chapter of Genesis into a textbook of Greek cosmology through the liberal use of allegory. The short tract *De ente et uno (On Being and the One)* was completed in 1491; it was part of a much longer work demonstrating the concord of Plato and Aristotle which was left unfinished at Pico's death in 1494 and is now lost. After the death of Lorenzo de'Medici, Pico came briefly under the influence of the religious reformer Savonarola (see Scepticism, Renaissance §3), and it is to this part of his life that Pico's biblical commentaries and shorter religious and devotional writings belong. Pico also composed a long tract against astrology, an exposition of Kabbalah and a number of letters and poems. Here we shall be concerned only with the philosophical portion of Pico's literary production.

2 The theory of love

Pico's independence as a thinker and his synthetic abilities are well illustrated by his first major work of philosophy, the *Commento*. This was a philosophical commentary in the manner of Dante's *Convivio* on some verses by Pico's friend, Girolamo Benivieni, a gentleman-poet who like Pico had been a disciple of Ficino. Benivieni had composed a *canzone*, 'Amor dalle cui', in emulation of the famous *canzone* 'Donna me prega' written about 1285 by Guido Cavalcanti (whose descendent, Giovanni Cavalcanti, happened to be Ficino's 'Platonic lover'). Guido Cavalcanti's poem had interpreted the phenomena of human love using the Aristotelian psychology and physics just then coming into vogue in Florence; Benivieni, two centuries later, tried to use Ficino's new Platonic metaphysics to describe divine love. His *canzone* was in fact based on five chapters (II.5, 7; V.4; VI.4, 7) of Ficino's most popular work, *De amore (On Love)* (1469) (see Ficino, M. §4). Pico's *Commento* on the poem had two chief aims. One aim was to rediscover the 'secret mysteries' of pagan theology using Neoplatonic metaphysics as a key: in the manner of Proclus and Pletho, Pico argued for identifying pagan divinities with various metaphysical entities and principles (Venus for example with Beauty itself, Uranus with God, Neptune with Mind, Vulcan with the Demiurge of the *Timaeus*, and so forth). Pico's second aim was to supplement and correct Ficino's account of the metaphysics of love with a fresh interpretation based on Pico's own reading of Platonic sources. The original draft of the *Commento*, indeed, was in essence a sharp and often personal

critique of *De amore*. (In the draft finally published in 1519, the attacks on Ficino were edited out by mutual friends.)

Pico criticized Ficino on numerous points, both methodological and doctrinal. Pico found Ficino's approach too literary, faulting it for its failure to define terms and employ precise scholastic terminology. Doctrinally, Pico insisted against Ficino that the contemplative intellect while in the body does not enjoy union with God himself, but only with the First Mind, the highest unity in the Plotinian hypostasis of *nous* (see Plotinus §3). Moreover, while Ficino had built his conception of Platonic love on the metaphysical identity of *philia* and *eros*, friendship and love, Pico asserts that the *eros* of which Plato spoke in the *Symposium* is a special sort of metaphysical power, distinct from friendship, from God's love for his creatures and from other unitive powers in the cosmos (see Love). In the same way, Pico argues for a sharp distinction, both metaphysical and moral, between physical and intellectual love, against Ficino's view that the two 'Venuses' formed a continuum, a unified system of cosmic dynamics. Against Ficino, again, Pico adopts an anti-Plotinian definition of Beauty, seeing it as a visual or intelligible harmony rather than an emanation from an Idea of Beauty belonging to *nous*. Finally, Pico rejects Ficino's attempt to integrate human love into a cosmic cycle of divine creative love descending from God down to creatures and linked back to God by the desire of creation for its creator. Pico, instead, sees love as a psychological process distinct from cosmic love. Divine love is a desire excited by intelligible beauty, and Pico assimilates the excitation of divine love in the rational soul to the process of cognition. In this way Pico transforms Aristotle's analysis of cognition in *On the Soul* (see Aristotle §17), as well as elements taken from Plotinus and Avicenna (see Ibn Sina), into an esoteric key to the speech of Diotima in the *Symposium*.

Despite the vehemence of Pico's disagreements with Ficino, one may not assume that the *Commento* represents Pico's considered views on any question. As the subtitle indicates, the work is intended as a kind of thought experiment or mental game. 'According to the understanding and opinion of the Platonists' means 'what a Platonist would or ought to say about love', not necessarily what Pico himself believed. This is a broadly Averroist approach to Plato in that Pico distinguishes explicitly between 'Christian truth' and Platonic 'understanding and opinion', and it was designed, like much of Pico's work, to infuriate his teacher Ficino. The aim of Ficino's scholarly and philosophical work was always to show the harmony of Plato with Christianity; his chief opponents were Aristotelian Averroists who in Ficino's view were deepening the divide between religion and philosophy (see Averroism). Pico's *Commento* insists at numerous points on the differences between Platonism and Christianity. He asserts that the Platonists believed in the eternal recurrence of the cosmos; in the pre-existence of the soul; in the possibility of embodied souls having direct knowledge of God; in the existence of intermediate creators between God the Supreme Creator and the cosmos. All of these doctrines since the twelfth century had subjected Platonists to charges of unorthodoxy. But since Pico in the *Commento* was writing under the persona of a Platonist, we are not licensed to assume that he had necessarily, at this stage of his career, embraced unorthodox beliefs, nor that his thought went through a development from an early Platonism to a mature Aristotelianism. Where Pico speaks *in propria persona* his works display great internal consistency.

3 The Roman disputation

Within a month of circulating the first draft of his *Commento*, indeed, Pico had published his own 900 theses and had begun drafting an oration meant to introduce them to the learned world. The oration is in effect an apology for a layman's devoting himself to theology and an answer to critics of Pico's proposed disputation. It belongs to the genre of the academic prolusion, a speech meant to introduce a university course of lectures or a formal disputation. Like most academic prolusions, it is divided into two parts, the first part praising the subject-matter, the second offering an overview of the material to be covered. Pico praises his subject-matter, theology, arguing that human beings fulfil the highest potentialities of their nature by contemplating divine things. Humans have a protean nature which permits them to live like beasts or angels, but they have a moral duty to live the highest kind of life of that they can. In order to achieve an angelic life of contemplation and love of the divine, one must pass through the stages of mystical education as set forth by Pseudo-Dionysius the Areopagite: purification, illumination, perfection. Following Byzantine tradition, Pico assimilates these stages to the four parts of philosophy, arranged hierarchically from the lowest (moral philosophy and dialectic), through natural philosophy, to the highest (metaphysics or theology) (see Byzantine philosophy). Using allegory he then proceeds to find these same stages in a variety of theological traditions, including the Mosaic, Pauline, Platonic, Zoroastrian and Pythagorean. Like Plato, Pico believes that the illumination acquired at the apex of metaphysical consciousness

will bring order and peace to the individual soul and unity among all rational beings who share this highest form of consciousness (see Illumination).

Pico's speech was given the title *On the Dignity of Man* by its sixteenth-century editors, and this has led to much misleading commentary on the part of modern scholars. The posthumous title seems to have been derived from the first few pages of the text, where Pico states (in the form of a myth) that the excellence of the human being does not derive from its central place in the Great Chain of Being (as Ficino had maintained), but precisely from its ability to choose its own place in the order of nature. Themselves standing outside the order of nature, human beings can shape their own essence; they can degrade themselves to live like beasts, or can raise themselves to the angelic order. Modern interpreters have seen this passage as embodying a distinctively Renaissance anthropology or even as the forerunner of more modern conceptions of human nature. It has been implausibly maintained that Pico's conception of the human being - placed in the centre of the world as an observer but outside the hierarchy of fixed essences - anticipates Descartes' theory of mind; or that Pico's description of the plasticity of human nature and the impact of moral choices upon metaphysical dignity can be linked to the existentialist doctrine that existence precedes essence. It should nevertheless be recognized that, despite its modern ring, much of Pico's anthropology is derivative from Boethius and Greek patristic sources, particularly from the Platonisms of Gregory of Nyssa (see Patristic philosophy) and Origen. Pico chose not to advertise his debt, probably because the orthodoxy of both Gregory and Origen was suspect. In any case, Pico's account of the 'chameleon' nature of human beings very likely represents an attempt to understand in a Christian sense the Pythagorean and Platonic doctrines of palingenesis or the transmigration of souls. Pico's statement that humans stand outside the triple hierarchy of being, occupying (as he says in the *Heptaplus*) a 'fourth world', may not amount to more than a dramatic restatement of the familiar twelfth-century doctrine of the human being as a microcosm of the universe.

The second part of the *Oration* explains why Pico wanted to defend so many theses from so broad a range of sources both Christian and non-Christian. Disputations in theology faculties of the period typically involved a few dozen theses, mostly from Christian authorities. In Italy theological disputation was normally confined to licensed members of religious orders. So Pico's performance, though not unprecedented, was highly unusual for a layman who lacked a university degree. Pico argued that the method and aims of his inquiry required him to cover a vast field. Theologians who belonged to developed school traditions could confine themselves to the clarification of details. But he was proposing to found a new theological school altogether, and a school based on an esoteric reading of all theologies past and present. He could not therefore be expected to limit himself to a few points of disputation. The same procedure, he observed, was followed by Aristotle, who always began a new inquiry by reviewing all previous opinions. Pico went on to give a *tour d'horizon* of the 900 theses, trying especially to show why the Latin and Arabic theological traditions ought to be enriched by the study of ancient theologies such as Platonism, Hermeticism, Zoroastrianism, Orphism and Kabbalism (see Ficino, M. §2). Pico took particular pride in his efforts to study Kabbalah from a Christian perspective; in the revival of 'ancient theology' this was one area where he had clearly moved beyond the work of Ficino. Pico assured his audience that, rightly understood, this esoteric theological tradition handed down from the time of Moses would only strengthen Christian efforts to convert the Jews.

Pico's published *Conclusiones* and his *Apologia* cast further light on his goals as a philosopher and theologian. Pico's aim is not to show that all theologies agreed in all respects. He was not an eclectic of the ancient type who believed that all philosophers agreed in substance, disagreeing only about words. He held that every major theological tradition contained elements of truth and that these elements could be combined into a grand synthesis that was at the same time compatible with orthodox Christianity. Like Nicholas of Cusa, he believed that the understanding previous theologians had of the sources of their own theological traditions was not necessarily correct. In fact his aim was to construct a new theology using material from existing historical theologies as building blocks. His new theology would be superior to existing theologies because it would give a richer understanding of Christian truths. This is not to imply that Pico was interested in directly challenging Christian dogmas. But he did wish (like many humanists) to move Christian theological speculation beyond the narrow circle of traditional authorities represented by Aristotle, the Bible and the *Sentences* of Peter Lombard. Situating Christian theology in the spectrum of world theologies would show both the superiority of Christian truth and the presence of Christian truths in all religions. Pico's ecumenism is therefore a militant ecumenism. At the same time, Pico hoped that recovering the esoteric truths of other religious traditions would make the mysteries of Christianity

more intelligible.

Pico's concordism might then be read as a kind of Thomistic project to incorporate non-Christian wisdom into a Christian speculative framework. But there is a critical difference between Thomas Aquinas and Pico which is best grasped by noting what is left out of Pico's theology. Pico has no ecclesiology, no soteriology in the traditional sense, and no doctrine of grace. Instead, Pico emphasizes the radical freedom of human beings in this life to achieve angelic levels of illumination and moral dignity through the exercise of powers belonging to human nature *qua* human. The neglect of grace is typical of the humanist theologies of the period, and reflects the widespread contempt for the Church's practice of reinforcing its wealth and position by marketing spiritual favours. That Pico attempted to propagate his new theology among the highest authorities in Christendom thus becomes a highly significant fact. It suggests that he shared some of the aims, if not the strategies, of the Protestant reformers of the next generation. It implies that the theological movement started by Ficino and Pico was not the esoteric and politically quietist movement it is sometimes represented as being, but had transformative aims as radical as those of Girolamo Savonarola and Martin Luther.

4 The concord of Plato and Aristotle

After Pico's first brash attempt to convert the Roman church to his new theology had failed, he devoted his energies to working out two pieces of his larger concordist project: the concord of Christianity and Kabbalism and the concord of Plato and Aristotle. The former task was taken up first, in *Heptaplus de opere sex dierum Geneseos (On the Sevenfold Narration of the Six Days of Genesis)*. Here Pico, like the Platonists of the twelfth century, attempted to use a non-Christian theology as a key to the creation stories in the first chapter of Genesis (see Chartres, School of). Unlike the hexaemeral treatises of the earlier Platonists, however, Pico maintained that Genesis 1 was an esoteric text. Moses had deliberately implanted in it secret cosmological doctrines that might be abstracted by an interpreter learned in Kabbalist lore. Pico's stated purpose in wresting cosmological mysteries from the Bible was to show that Moses knew what the best philosophers knew. He could in this way show that there was no gap between reason and revelation. But whatever Pico's stated aims, his view of Genesis had radical implications. To employ a modern parallel, it is as though a contemporary exegete of the Bible should maintain that there was not only no conflict between Genesis and evolutionary biology, but that Moses had intentionally concealed evolutionary teachings under the guise of Genesis' account of creation. This radical claim, combined with Pico's circular hermeneutical method, in effect turns the literal sense of Genesis into a myth and subordinates traditional Christian teachings to the teachings of philosophers and non-Christian theologians.

Of Pico's other project of his last years, the concord of Plato and Aristotle, we have only the short tractate *On Being and the One*. The purpose of this work was to criticize one of the cardinal theological doctrines of Ficino and the ancient Neoplatonists, namely the radical transcendence of the One 'beyond being'. On this issue Ficino followed Plotinus and Proclus, who had in turn based themselves on a particular reading of Plato's *Parmenides* (see Plato). The Neoplatonists regarded that dialogue as an esoteric work containing the most sublime mysteries of Platonic theology. Against Ficino, Pico maintained that *esse ipsum* (being itself) and the One were 'convertible' or identical in extension, and that both were distinct from *ens* (participated being). This conception was borrowed from Thomas Aquinas; Pico's originality lay in claiming that it was shared by both Plato and Aristotle. To argue the point, Pico was obliged to challenge two assumptions that had gone largely unexamined for a millennium. The first, characteristic of ancient Neoplatonists but found also among Greek patristic and Byzantine interpreters, as well as Latin theologians such as Augustine, was that Plato was superior to Aristotle in theology (see Aristotelianism, medieval; Platonism, medieval). Pico maintained that Aristotle was Plato's equal as a theologian. In effect he includes Aristotle among the ancient theologians. The second assumption was that the Neoplatonists were reliable guides to the thought of Plato. This too Pico denied. Pico revived the Academic sceptical interpretation of the *Parmenides*, which regarded the dialogue as an eristic work, a dialectical exercise containing no positive doctrines. Unlike the Academic sceptics, however, Pico did not interpret all the dialogues of Plato this way. While in his view the *Parmenides* was a purely eristic work, other dialogues of Plato, particularly the *Sophist* and *Philebus*, could be used to recover Plato's own views on metaphysical questions and to correct the false readings of the *platonici*.

Pico's interpretation of Plato thus represents a step towards the modern distinction between Plato and the Neoplatonists, even if his view of the authentic Plato is very different from that of modern scholarship. His

exegetical project is very much a Renaissance project in its attempt to distinguish the original source from later glosses and to privilege the former over the latter. Pico's concordism also reflects the situation of Renaissance philosophy in another respect. In the central Middle Ages, philosophical study in the universities had been carefully regulated by official curricula, textbooks and university statutes. Theology was even more closely controlled. By the fifteenth century, thanks largely to the humanist movement, philosophers were being introduced to an ever-widening range of philosophical and theological opinion, often in far less regulated circumstances. The bewildering variety of intellectual choices and the greater freedom of study (if not of expression) posed a major challenge to philosophers of the later fifteenth and sixteenth centuries. Some philosophers responded by embracing scepticism; others by advocating fideism (see Scepticism, Renaissance); still others - ultimately the most influential group - by exploring new methodologies for adjudicating between incommensurable systems of thought. Pico's concordism represents a fourth alternative. If many of Pico's hermeneutical techniques, especially allegory, now seem little more than primitive forms of secondary elaboration, the dream of a great theological system that would reconcile in itself trans-historical and trans-cultural truths taken from all forms of religious wisdom was one that would have a considerable fortune down to the present day.

See also: Aristotelianism, Renaissance; Ficino, M.; Hermetism; Kabbalah; Platonism, Renaissance

JAMES HANKINS

List of works

Pico della Mirandola, G. (1463-94) *Opera omnia (Collected Works)*, Basle: Heinrich Petri, 1572; repr. Turin: Bottega d'Erasmo, 1971.(Volume two of the 1971 edition contains photo-reprints of all works of Pico not contained in the original 1572 edition.)

Pico della Mirandola, G. (1485) 'Letter to Ermolao Barbaro (3 June 1485)', Latin text in *Opera omnia*, Turin: Bottega d'Erasmo, 1971, 351-8; trans. Q. Breen, 'Giovanni Pico della Mirandola on the conflict of philosophy and rhetoric', *Journal of the History of Ideas* 13 (1952): 325-426.(Referred to in §1. Pico defends technical scholastic terminology against humanist critics. The second letter to Barbaro mentioned in §1 can be found in volume 1, page 368, of the 1557 *Opera omnia* (printed in Basle; repr. Hildesheim: Olms, 1969).)

Pico della Mirandola, G. (1486) *Commento*, in E. Garin (ed.) *De dignitate hominis, Heptaplus, De ente et uno e scritti vari*, Florence: Vallecchi, 1942; trans. S. Jayne, *Commentary on a Canzone of Benivieni*, New York: Lang, 1984.(Pico's version of Platonic love; discussed in §2. Jayne's English translation has a useful introduction and textual improvements.)

Pico della Mirandola, G. (1486) *Conclusiones sive theses DCCCC (Nine Hundred Conclusions or Theses)*, ed. B. Kieszkowski, Geneva: Droz, 1973.(Referred to in §§1 and 3. A list of theses for debate which summarizes Pico's concordist philosophy.)

Pico della Mirandola, G. (1486) *De dignitate hominis*, in E. Garin (ed.) *De dignitate hominis, Heptaplus, De ente et uno e scritti vari*, Florence: Vallecchi, 1942; trans. C.G. Wallis, P.J.W. Miller and D. Carmichael, *Pico della Mirandola: On the Dignity of Man, On Being and the One, Heptaplus*, Indianapolis, IN, and New York: Library of Liberal Arts, 1965; ed. E. Garin, *La cultura filosofica del Rinascimento italiano*, Florence: Sansoni, 1961, 231-40; trans. E.L. Forbes, *On the Dignity of Man*, in E. Cassirer, P.O. Kristeller and J.H. Randall (eds) *The Renaissance Philosophy of Man*, Chicago, IL, and London: University of Chicago Press, 1948.(Referred to in §§1 and 3. An overview of Pico's concordism. Garin 1961 gives the original version of the text.)

Pico della Mirandola, G. (1487) *Apologia*, in *Opera omnia*, Turin: Bottega d'Erasmo, 1971, vol. 1, 114-240. (Referred to in §3. Pico defends himself against the charge of heresy.)

Pico della Mirandola, G. (1489) *Heptaplus de opere sex dierum Geneseos (On the Sevenfold Narration of the Six Days of Genesis)*, in E. Garin (ed.) *De dignitate hominis, Heptaplus, De ente et uno e scritti vari*, Florence: Vallecchi, 1942; trans. C.G. Wallis, P.J.W. Miller and D. Carmichael, *Pico della Mirandola: On the Dignity of Man, On Being and the One, Heptaplus*, Indianapolis, IN, and New York: Library of Liberal Arts, 1965. (Referred to in §§1 and 4. Harmonizes Christianity and Kabbalah on the interpretation of Genesis.)

Pico della Mirandola, G. (1491) *De ente et uno*, in E. Garin (ed.) *De dignitate hominis, Heptaplus, De ente et uno e scritti vari*, Florence: Vallecchi, 1942; trans. C.G. Wallis, P.J.W. Miller and D. Carmichael, *Pico della Mirandola: On the Dignity of Man, On Being and the One, Heptaplus*, Indianapolis, IN, and New York: Library of Liberal Arts, 1965.(Referred to in §§1 and 4. Harmonizes Plato and Aristotle on the nature of God.)

References and further reading

Allen, M.J.B. (1986) 'The Second Ficino-Pico Controversy: Parmenidean Poetry, Eristic and the One', in G.C.

Garfagnini (ed.) *Marsilio Ficino e il ritorno di Platone: Studi e documenti*, Florence: Olschki, 417-55; repr. in M.J.B. Allen, *Plato's Third Eye: Studies in Marsilio Ficino's Metaphysics and its Sources*, London: Variorum, 1995.(Pico and Ficino's encounter over the interpretation of the *Parmenides*; see §4.)

Aranci, G., De Marco, P. and Verdon, T. (1994) 'Teologie a Firenze nell'età di Giovanni Pico della Mirandola' (Theologies in Florence during the Age of Giovanni Pico della Mirandola), special issue of *Homo vivens* 5 (2). (Recent work on Pico from the quincentenary celebrations.)

Craven, W.G. (1981) *Giovanni Pico della Mirandola, Symbol of His Age: Modern Interpretations of a Renaissance Philosopher*, Geneva: Droz.(Useful antidote to the secondary literature; better on destruction than reconstruction.)

Di Napoli, G. (1965) *Giovanni Pico della Mirandola e la problematica dottrinale del suo tempo (Giovanni Pico della Mirandola and the Doctrinal Issues of His Time)*, Rome: Desclé.(Pico's relationship to late Quattrocento scholasticism.)

Ficino, M. (1469) *De amore*, trans. S. Jayne, *Commentary on Plato's Symposium On Love*, Dallas, TX: Spring Books, 1985.(Referred to in §2. Gives a useful introduction describing the influence of *De amore* on Renaissance literature.)

Ficino, M. (1474) *Theologia platonica de immortalitate animae (Platonic Theology, On the Immortality of the Soul)*, ed. and trans. R. Marcel, *Théologie platonicienne de l'immortalité des âmes*, Paris: Les Belles Lettres, 1964-79, 3 vols.(Referred to in §1. Edition of the Latin text with a French translation.)

Garin, E. (1937) *Giovanni Pico della Mirandola, Vita e dottrina (Life and Doctrines)*, Florence: Le Monnier. (Standard account of Pico's life; interpretation of works now largely superseded.)

Lubac, H. de (1974) *Pic de la Mirandole: Études et discussions (Studies and Discussions)*, Paris: Aubier Montaigne.(Pico and his patristic sources.)

Reinhardt, H. (1991) *Freiheit zu Gott: Der Grundgedanke des Systematikers Giovanni Pico della Mirandola (Freedom towards God: The Basic Thought of the Systematic Thinker Giovanni Pico della Mirandola)*, Weinheim: VCH Acta humaniora.(Pico as forerunner of Vatican II.)

Roulier, F. (1989) *Jean Pic de Mirandole (1463-1494), humaniste, philosophe et théologien (Humanist, Philosopher and Theologian)*, Geneva: Slatkine. (Comprehensive monograph on Pico.)

Viti, P. (ed.) (1994) *Pico, Poliziano, e l'umanesimo di fine Quattrocento (Pico, Poliziano and Humanism at the End of the 15th Century)*, Florence: Olschki.(Contains important new biographical material on Pico.)

Wirszubski, C. (1989) *Pico della Mirandola's Encounter with Jewish Mysticism*, Cambridge, MA: Harvard University Press.(Pico and Kabbalism.)

Pietism

'Pietism' refers to a Protestant reform movement, arising in the late 1600s in Lutheran Germany, which turned away from contests over theological and dogmatic identity in Protestant confessionalism and urged renewed attention to questions of personal piety and devotion. As such, it has only the most tenuous historical connections to the Christocentric piety of the devotio moderna *or the northern humanist piety of Erasmus or Zwingli. It found its first major voice in P.J. Spener and A.H. Francke, and established its principal centres of influence at the state university at Halle in 1691 and the Moravian community at Herrnhut in 1722. Pietism found followers and allies in the European Reformed churches, in the Church of England (especially through the example of John and Charles Wesley and through the Moravian exile community in England), and in Britain's English-speaking colonies. In the colonies, pietism not only found Lutheran and Reformed colonial hosts, but also saw in New England Puritanism a movement of similar aspirations. Pietism's impact on the spirituality of western Europe and America was clearly felt in the eighteenth-century Protestant Awakenings, and continues to have an influence in the shape of Anglo-American evangelicalism.*

1 Definition

Pietism may be defined very broadly as a reawakening in western Europe (and Europe's American colonies) of the most intense and aggressive forms of devotional Christianity from about 1650 to 1750. In this broad sense, it can include Jansenism in Roman Catholic France, the Non-Jurors, the eighteenth-century Moravians, and the English and American Methodists who followed John and Charles Wesley. But in its most specific use, the term 'pietism' was first used in 1674 to refer to the followers of the German Lutheran spiritual reformer Philip Jakob Spener, and it is from this particular movement within Lutheranism that 'pietism' is generally understood to derive its historical identity.

Pietism can be characterized as a Christian and Protestant movement which located the essence of Christianity more in the quality of an individual's personal feeling of relationship to God than in a rational assent to the various forms of Christian theology or in submission to any Christian rites of identity, such as baptism or the Eucharist. It stands apart from its medieval, mystical and humanist predecessors in its biblicism, its persistently Protestant forms, and the extraordinary energies it devoted to promoting popular awakenings and evangelism. Similarly, pietism's preoccupation with 'personalizing' Christian belief was never a matter of total withdrawal from conventional Christian forms, but rather a matter of emphasis in the priority and use of those forms, since not even the most radical pietist groups completely repudiated the Christian sacraments or questioned the validity of organized churches. But the pietists did subordinate these considerations to those of personal devotion to God. This allowed them to take a remarkably tolerant and irenic stance towards questions of confessional debate within Protestantism, since such questions had little significance for the pietists except in so far as they promoted the individual's personal devotion to God.

But irenicism across confessional boundaries stopped short when pietists addressed the boundary between those who had experienced a renewed dedication of their spiritual lives to God and those who either had not or did not care to do so. Pietists drew a deep antithetical line between those who had experienced a 'new birth' or 'second birth' (*Wiedergeburt*) and those with a purely 'notional' Christianity, who had never experienced contrition and repentance. On the one hand, this drove them to proselytization and missionary programmes among nominal Christian and non-Christian populations in Europe, America and Asia, but on the other, it drove them to place so much distance between 'notional' Christianity and their own personal practice that many pietists came close to a demand for moral perfection. In the case of John Wesley, perfection was actually promoted as an attainable goal (see Sanctification §3).

2 History

The foundations for pietism lie in the first wave of Protestantism, which cultivated a strong sense of personal Christian spirituality alongside its concern with revising Catholic dogma. Much of that devotionalism was submerged, however, during the waves of religious warfare that swept over Europe as a result of the Reformation, and lasted, with intermissions, until 1648. The need clearly to identify who Protestantism's friends and enemies were fostered a century of theological system-building and confession-writing (*Verconfessionalisierung*) in

Germany. But by the end of the Thirty Years' War, with the German lands devastated by the marauding criss-cross of Protestant and Catholic armies, no absolute victor had emerged to legitimize any of the Protestant confessions, and many Lutherans felt betrayed by the failure of their state churches and princes to deliver either Protestant triumph or reform of public life. After 1680, these anxieties were heightened by a renewed Catholic determination to undo the Westphalian compromises of 1648 within Catholic-ruled domains.

Spener was pastor and senior of the ministerium in Frankfurt. He became convinced that the state Lutheran churches had failed the German people and had descended into theological irrelevance and doctrinal trivia. Disgusted, he published *Pia Desideria: or, Heartfelt Desires for a God-Pleasing Improvement of the True Protestant Church* (1675) as his blast against the prevailing lifelessness of the Lutheran churches, and in it he set out six proposals for reviving the German churches:

(1) rededicate the energies of the churches to the study of the Bible as the object of devotional meditation and spirituality;
(2) break down the wall of separation between clergy and laity;
(3) emphasize practical, rather than theological or intellectual, Christianity, in terms of service and worship;
(4) abandon religious argument and controversy with other churches or sects;
(5) reorganize the training of future ministers in the German universities so as to stress the formation of spiritual lives rather than doctrine;
(6) revitalize preaching with view to promoting the people's edification.

In 1686, Spener was appointed official preacher to the court of the Duke of Saxony. There, he met and influenced a young lecturer at the Saxon university at Leipzig, August Hermann Francke. As a theological student, Francke acknowledged that he had grasped confessional Lutheranism 'only in my reason and in my thought' and it was not until he had experienced repentance (*Bußkampf*) that 'all sadness and unrest of my heart was taken away at once, and I was immediately overwhelmed as with a stream of joy so that with full joy I praised and gave honor to God who had shown me such great grace.' Francke 'arose a completely different person' and became Spener's friend and disciple, and when in 1691 the Elector Frederick William III of Prussia promoted the *Ritterakademie* at Halle to the rank of university, Francke was appointed a professor and quickly turned Halle into an engine of pietist recruitment and evangelism. Still another centre of pietist energy sprang up in 1722 under the leadership of Nikolaus Ludwig Graf von Zinzendorf. Zinzendorf was the eldest son of a noble Saxon family, but his parents (and especially the grandmother who was mostly responsible for his upbringing) had been greatly influenced by Spener. His estates at Berthelsdorf lay directly across the track of Moravian Protestant refugees, fleeing forced conversion by the Jesuits under imperial Habsburg rule. Zinzendorf, hearing of the plight of these refugees, turned Berthelsdorf into a refuge for the Moravians and helped organize a community there named 'Herrnhut' ('the Lord watches over'), with a view to turning the entire effort into an experiment in organized pietism.

Spener's criticism of the state Lutheran churches in Germany made the pietists an object of suspicion for Lutheran confessionalists. The use of field meetings, and the encouragement of 'children's testimony' (testimony to religious experience by children in public adult meetings) among Silesian pietists disturbed Lutheran leaders, and the pietists' absorption in personal piety led to charges of indifference to, and even outright departure from, the Lutheran confessional standards. In 1734, Zinzendorf himself was charged with straying from the letter of Lutheran theology; in 1736, he was banished from Saxony. The resurgence of Catholic attempts to impose religious uniformity in France and the Austrian domains also provided an opportunity for Catholic authorities to abrogate tacit agreements with Lutheran minorities in their lands, and in November 1731, the pietist Lutherans of Salzburg were expelled by the archbishop and began a march across Europe that ended only when Prussia opened its doors to approximately 20,000 of the exiles.

Zinzendorf, borrowing money from sympathetic Dutch bankers, moved across Protestant Europe to promote local pietist movements in Württemburg (where he alternately dominated and antagonized F.C. Oetinger and J.A. Bengel) and Switzerland (where the major pietist converts among the state Reformed clergy were Samuel Schumacher and Samuel Lutz in Berne). He was also able to carry the Moravian example abroad, founding new *Herrnhuts* at Marienborn and Herrnhaag in 1736 and 1738, gaining official recognition for the Moravians in England, sponsoring Moravian missions to the West Indies, and planting a Moravian colony in British North America. It was in Georgia that a party of Moravian immigrants confronted the young John Wesley, then an

emissary of the Church of England, and when Wesley returned to London in 1738, it was in a Moravian society-meeting in London's Aldersgate Street that he experienced the vivid personal conversion that led to the founding of Methodism.

Pietism also found already existing allies among the English-speaking Puritans in the British colonies of New England. Spener had drawn significant encouragement from the writings of the English Puritans Lewis Bayly, John Bunyan and Richard Baxter, and the expatriate William Ames wielded great influence on the Dutch pietists Willem Brackel and Abraham Hellenbroeck. Although by the early 1700s much of Puritanism's fervour had settled back on its lees in Britain and New England, the American Puritan minister Cotton Mather insisted in 1717 that 'the *American Puritanism*' was 'so much of a Peece with the Frederician Pietism, that if it were possible…to be transferr'd unto our Friends in the lower *Saxony*, it would find some Acceptance' (Lovelace 1979: 33). The Puritan philosopher-theologian Jonathan Edwards (§1), whose reading list included most of the prominent German pietist titles, was moved by the example of the New Jersey Dutch pietist T.J. Freylinghuysen, the Methodist George Whitefield, and the Scots-Irish pietist Gilbert Tennent to promote (and then write a series of sympathetic rationalizations for) the colonial religious revivals known as the 'Great Awakening' (1739-44). By the time of the deaths of Edwards in 1758 and Zinzendorf in 1760, pietism had taken up residence in all of the confessional Protestant churches of Europe, and had established missions or colonies in Greenland, Scandinavia, Egypt, Ceylon and the Caribbean.

3 Continuing influence

Many accounts of pietist history suggest a communal harmony that many of the pietists actually lacked: Zinzendorf could be tyrannical and bullying, and the resentment of Francke's successor at Halle, Gotthilf August, at Zinzendorf's high-handedness led to a cooling of relations between Halle and Herrnhut. Wesley, similarly, was dictatorial and self-centred, breaking ties over personal and theological issues with his talented associate Whitefield and with his Welsh disciples Howell Harris and Daniel Rowland. As a result, pietism never achieved any measure of hegemony within any major Protestant confessional grouping in the 1700s. Only among separatist movements such as the United Brethren and the Methodists did pietism become the dominant form of theological identity, and even then, by the time the Methodists separated from the Church of England after Wesley's death, the movement had become an expression of an idiosyncratic Wesleyanism rather than the larger spirit of pietism.

Part of that failure was due to the ferocious resistance offered to pietism by confessional theologians. Lutheran confessionalists were concerned that the pietist preoccupation with personal experience would obscure the forensic status of justification, which according to Lutheran dogma was established not by the individual's repentance, but by the death of Jesus Christ (see Justification, religious §4). Calvinist theologians, even those in New England otherwise inclined to sympathize with them, were troubled by the pietists' willingness to redefine the atonement of Christ in general terms in order to heighten the possibility of repentance and redemption for all, which appeared in Calvinist eyes to diminish the actual worth of the atonement as a real means of redemption (see Atonement §4). Wesley in particular repudiated the moderate Calvinism of the Church of England's Thirty-Nine Articles of Religion for an evangelicalized Arminianism which converted the free-will rationalism of Jacobus Arminius into an engine for freely chosen and freely demanded conversion. And, as in the example of both Halle and the Methodists, the high demands of pietism for Christian conduct could easily yield to legalistic requirements for conformity to mere behavioural codes.

However, many other criticisms of the pietists as ascetic, anti-cultural or anti-intellectual have less justification. Except for the Moravians, pietist movements were open associations with little inclination towards communal withdrawal from the world. Francke, the most talented organizer among the pietist leaders, created one of the largest public charities in Europe with the Halle Orphanage in 1695, and his example was imitated (though on a lesser scale) by both Wesley and Whitefield. The musical culture of the Moravians, even in their colonial settlements, was of a very high order, and pietist hymnody, especially from Charles Wesley and Gerhard Tersteegen, has been absorbed into the Church music of many of their confessional critics. If the pietists were, at first appearance, hostile to any learning that lacked personal piety, they also shared a basic similarity with the Enlightenment in their refusal to accept the testimony of past authority for truth and the demand that truth be grounded in demonstration and experience. The pietist revolt against formalism would be re-echoed in the Christian Romantics, especially in Friedrich Schleiermacher (§§1, 7) (who has been rightly called the executor of

pietism's last will and testament) and Søren Kierkegaard (§4). The pietist influence mediated in America by the 'Great Awakening' had a long-term influence in American pragmatic philosophy and has been the dominant force in the shaping of modern American and British evangelicalism.

ALLEN C. GUELZO

References and further reading

Beyreuther, E. (1978) *Geschichte des Pietismus* (History of Pietism), Stuttgart: Steinkopf.(General history and critique of pietism.)

Erb, P.C. (ed.) (1983) *Pietists: Selected Writings*, New York: Paulist Press. (Anthology of major pietist writings, including excerpts from Spener, Francke, Tersteegen and Zinzendorf, with introduction.)

Heppe, H. (1879) *Geschichte des Pietismus und der Mystick in der Reformiter Kirche der Niederlande* (History of Pietism and Mysticism in the Dutch Reformed Church), Leiden: Brill.(History of pietist influence on Dutch Calvinism.)

Lovelace, R.F. (1979) *The American Pietism of Cotton Mather: Origins of American Evangelicalism*, Grand Rapids, MI: Eerdmans.(Stresses the connections between American Puritanism and Continental pietism.)

Nagy, P. (1976) 'Pragmatism and American Pietism', *Transactions of the Charles Sanders Peirce Society* 12 (1): 165-81.(Connects the concentration of pietism upon the active and therapeutic influence of religion with the shape of American pragmatism.)

Rack, H.D. (1989) *Reasonable Enthusiast: John Wesley and the Rise of Methodism*, London: Epworth. (Intellectual biography of Wesley, with detailed treatment of his interaction with the Moravians.)

Spener, P.J. (1675) *Pia Desideria: or Heartfelt Desires for a God-Pleasing Improvement of the True Protestant Church*, trans. T.G. Tappert, Philadelphia, PA: Fortress, 1964.(Standard English edition of this work, with introduction.)

Stoeffler, F.E. (ed.) (1976) *Continental Pietism and Early American Christianity*, Grand Rapids, MI: Eerdmans. (Essays on the influence of pietism on various early American religious communities, including German Lutheran, Dutch Reformed and Moravians.)

Stoeffler, F.E. (1965) *The Rise of Evangelical Pietism*, Leiden: Brill.(Surveys the background to the rise of pietism and concludes with a survey of Spener.)

Ward, W.R. (1992) *The Protestant Evangelical Awakening*, Cambridge: Cambridge University Press.(Overview of eighteenth-century evangelical revivals in Central Europe and America, with special attention to the role played by pietism and especially by Zinzendorf.)

Planck, Max Karl Ernst Ludwig (1858-1947)

Planck was a German theoretical physicist and leader of the German physics community in the first half of the twentieth century. Famous for his introduction of the quantum hypothesis in physics, Planck was also a prolific writer on popular-scientific and philosophical topics. Even more so than his younger contemporary Albert Einstein, Planck was well-known in his day for his defence of a realist conception of science and his explicit criticism of the positivism of Ernst Mach and the Vienna Circle.

Max Planck was born in Kiel and after 1867 moved with his family to Munich. Planck studied physics with Philipp von Jolly at Munich and with Herman von Helmholtz and Gustav Kirchhoff at Berlin, taking his doctorate in Munich in 1879. After working as a *Privatdozent* in Munich, Planck was named *Außerordentlicher Professor* in Kiel in 1885. Most of his career was lived in Berlin, where he was called as Kirchhoff's successor in 1889, being named *Ordentlicher Professor* in 1892. During his Berlin years, Planck became perhaps the most influential figure in the German physics community, serving, among other roles, as secretary of the Prussian Academy of Sciences from 1912 until 1938 and as President of the Kaiser-Wilhelm-Gesellschaft from 1930 to 1935 and again briefly after the Second World War, before it was renamed the Max-Planck-Gesellschaft in his honour in 1946. Among Planck's students in Berlin was Moritz Schlick, who completed a dissertation under Planck in 1904. From 1914 until 1933, Albert Einstein was one of his closest colleagues; Erwin Schrödinger was his successor. After retirement in 1928, Planck continued to be quite active, writing especially on broader popular and philosophical topics, and struggling after 1933 to preserve some of the dignity and integrity of German physics under Hitler. Planck was a political conservative and a defender of the highest ideals of German culture. For all of Planck's prominence, his life was dogged by tragedy, including the death of his first wife Marie, in 1909, the deaths of his twin daughters Grete and Emma, in childbirth in 1917 and 1919, the death of his elder son Karl, from war wounds in 1918, the destruction by bombing in 1944 of his home in Berlin-Grünewald, along with all of the personal papers and possessions it contained, and the execution of his younger son Erwin, in 1945, for complicity in the plot to assassinate Hitler.

Planck's early work concerned primarily thermodynamics and efforts to clarify the concept of entropy - this in the tradition of Rudolf Clausius and the mechanical theory of heat. One early controversy involved his defence of the mechanical worldview against energeticist critics of atomism and mechanism, including Wilhelm Ostwald and Georg Helm; another involved his defence of the second law of thermodynamics as a deterministic law against Ludwig Boltzmann's statistical interpretation of the entropy law (see Thermodynamics). Planck's most famous contribution to physics, for which he earned the Nobel Prize in 1918, was his introduction of the quantum hypothesis. The idea of the quantum first appeared in the formula for the frequency distribution of the energy of black-body radiation as a function of temperature that Planck derived late in 1900. The Planck formula contained two constants. One is the constant, h, later recognized as representing the quantum of action and named the Planck constant. Planck originally thought that it was only the exchange of energy between matter and radiation that was quantized. Einstein first proposed in 1905 that free radiant energy itself exists in a quantized form as photons. Planck's analysis of how the second constant, k, the Boltzmann constant, entered the derivation of the distribution formula yielded a theoretical determination of the value of Avogadro's number and hence confirmation of the atomistic hypothesis. Planck was an early and enthusiastic supporter of the theory of relativity (see Relativity theory, philosophical significance of), but to the end of his life remained sceptical of the Copenhagen interpretation of quantum mechanics associated with Niels Bohr and Werner Heisenberg (see Quantum mechanics, interpretation of).

Planck's commitment to atomism led him into a well-known debate with Ernst Mach over realism and positivism (Planck 1909, 1910; Mach 1910) (see Scientific realism and antirealism; Logical positivism). In his lecture, 'Die Einheit des physikalischen Weltbildes' (1909), Planck argued against Mach that ' however little we know about their more specific properties, the atoms are no more and no less real than the heavenly bodies', adding that what sustained the 'great masters' of natural science - Copernicus, Kepler, Newton, Huygens and Faraday - in their struggles against 'received opinion' and 'higher authorities' was not the 'economical point of view' but their 'rock-solid…belief in the reality of their world picture'. With the rise of the Vienna Circle and logical empiricism in the later 1920s (see Vienna Circle), Planck returned to this theme. It was sufficiently important for him to make

it the focus of his first major address to the Kaiser-Wilhelm-Gesellschaft after being named its President. *Positivismus und reale Außenwelt* (1930) represents Planck's most thoroughgoing critique of positivism. According to Planck, when positivism is carried to its logical conclusion, it ' denies the concept and the necessity of an objective physics, that is to say, one independent of the individuality of the researcher'. An objective physics is achieved only through a 'step into the metaphysical', with the hypothesis 'that our experiences do not themselves constitute the physical world, that they instead only give us information about another world that stands behind them and is independent of us, in other words, that a real external world exists'.

DON HOWARD

List of works

Planck, M. (1909) 'Die Einheit des physikalischen Weltbildes' ('The Unity of the Physical World Picture'), *Physikalische Zeitschrift* 10: 62-75; repr. *Die Einheit des physikalischen Weltbildes*, Leipzig: S. Hirzel; repr. in Planck 1949, 28-51; trans. in Planck 1925, 1-41.(Planck's first critique of Machian positivism.)

Planck, M. (1910) 'Zur Machschen Theorie der physikalischen Erkenntnis. Eine Erwiderung' ('On Mach's Theory of Physical Knowledge: A Reply'), *Physikalische Zeitschrift* 11: 1,186-90.(Planck's rejoiner to Mach's reply.)

Planck, M. (1920) *Die Entstehung und bisherige Entwicklung der Quantentheorie*, Leipzig: J.A. Barth; trans. H.T. Clarke and L. Silberstein, *The Origin and Development of the Quantum Theory*, Oxford: Clarendon Press, 1922.(Nobel Prize address.)

Planck, M. (1925) *A Survey of Physics: A Collection of Lectures and Essays*, trans. R. Jones and D.H. Williams, London: Methuen; repr. *A Survey of Physical Theory*, New York: Dover, 1960.(An early collection of Planck's popular and philosophical essays in English translation.)

Planck, M. (1930) *Positivismus und reale Außenwelt (Positivism and the Real External World)*, Leipzig: Akademische Verlagsgesellschaft; repr. in Planck 1949, 228-45; trans. in Planck 1932, 64-106.(Planck's mature critique of positivism.)

Planck, M. (1932) *Where Is Science Going?*, trans. J. Murphy, New York: W.W. Norton; repr. Woodbridge, CT: Ox Bow Press, 1981.(A second collection of Planck's popular and philosophical writings in translation.)

Planck, M. (1936) *The Philosophy of Physics*, trans. H.W. Johnston, New York: W.W. Norton.(The third and final English-language collection of Planck's popular and philosophical essays.)

Planck, M. (1948) *Wissenschaftliche Selbstbiographie*, Leipzig: J.A. Barth; repr. with additional documents bearing upon its origin, Halle/Saale: Deutsche Akademie der Naturforscher Leopoldina, 1990; trans. F. Gaynor, *Scientific Autobiography*, New York: Philosophical Library, 1949; repr. New York: Greenwood Press, 1968. (Planck's scientific autobiography.)

Planck, M. (1949) *Vorträge und Erinnerungen (Lectures and Reminiscences)*, Stuttgart: S. Hirzel, 5th edn; 10th edn, Darmstadt: Wissenschaftliche Buchgesellschaft, 1965.(The most complete collection of Planck's popular and philosophical writings. Contains an expanded version of *Wege zur physikalischen Erkenntnis* (Paths to Physical Knowledge), Leipzig: S. Hirzel, 1944.)

Planck, M. (1958) *Physikalische Abhandlungen und Vorträge. Aus Anlaß seines 100. Geburtstages (23. April 1958) (Physical Papers and Lectures. On the Occasion of His 100th Birthday)*, Braunschweig: Friedrich Vieweg und Sohn.(Planck's collected scientific papers.)

References and further reading

Heilbron, J.L. (1986) *The Dilemmas of an Upright Man: Max Planck as Spokesman for German Science*, Berkeley, CA: University of California Press.(Definitive recent biography focusing on the personal and institutional setting of Planck's work.)

Hermann, A. (1973) *Max Planck in Selbstzeugnissen und Bilddokumenten (Max Planck in His Own Words and Photographic Records)*, Reinbeck: Rowohlt.(An accessible biographical collection.)

Kangro, H. (1976) *Early History of Planck's Radiation Law*, London: Taylor & Francis.(A detailed study of Planck's derivation of the energy distribution law for black-body radiation.)

Klein, M.J. (1962) 'Max Planck and the Beginnings of the Quantum Theory', *Archive for History of Exact Sciences* 1: 459-79.(Klein's three papers constitute one of the definitive studies of Planck's work on the radiation law.)

Klein, M.J. (1963) 'Planck, Entropy, and Quanta, 1901-1906', *The Natural Philosopher* 1: 83-108.(See 1962.)

Klein, M.J. (1966) 'Thermodynamics and Quanta in Planck's Work', *Physics Today* 19 (11): 23-32.(See 1962.)

Kretzschmar, H. (1967) *Max Planck als Philosoph (Max Planck as a Philosopher)*, Munich and Basel: Ernst

Reinhardt.(The most recent extended study of Planck's philosophy, informative even though it goes too far out of its way to dispute materialist, Marxist-Leninist interpretations of Planck's philosophy.)

Kuhn, T.S. (1978) *Black-body Theory and the Quantum Discontinuity 1894-1912*, Oxford: Clarendon Press. (Along with Kangro and Klein, the other major historical study of Planck's work on the quantum hypothesis.)

Lowood, H. (1977) *Max Planck: A Bibliography of His Non-technical Writings*, Berkeley, CA: Office for the History of Science and Technology, University of California.(Quite detailed and comprehensive.)

Mach, E. (1910) 'Die Leitgedanken meiner naturwissenschaftlichen Erkenntnislehre und ihre Aufnahme durch die Zeitgenossen' (The Main Principles of My Theory of Scientific Knowledge and its Reception by My Contemporaries), *Physikalische Zeitschrift* 11: 599-606.(Mach's reply to Planck's criticism in 'Die Einheit des physikalischen Weltbildes'.)

Vogel, H. (1961) *Zum philosophischen Wirken Max Plancks. Seine Kritik am Positivismus (On the Philosophical Work of Max Planck: His Critique of Positivism)*, Berlin: Akademie-Verlag.(A helpful study even though it attempts to impose a materialist, Marxist-Leninist interpretation on Planck's philosophy.)

Platform Sutra

The Platform Sutra is the single most important work of early Chinese Chan Buddhism, perhaps of the entire Chan/Sôn/Zen tradition. It purports to contain the teachings of the Sixth Patriarch Huineng (638-713), whom it celebrates as an illiterate but enlightened sage. The centrepiece is an exchange of verses attributed to Shenxiu (606?-706) and Huineng, generally taken to represent, respectively, a gradual or progressive self-cultivation leading to perfect enlightenment and a sudden or subitist style of practice in which enlightenment is attained all at once. The Platform Sutra was actually composed around 780, more than a century after the events it describes. Its reportage is demonstrably inaccurate and its traditional interpretation problematic, but this does not alter the profound mythopoeic importance of the text.

The events described in the *Platform Sutra* supposedly took place at the Chan community led by Hongren (601-74), described here as the Fifth Patriarch. Needing a successor, he instructed his students each to compose and submit a verse describing their understanding of Buddhism. The only one to do so was Shenxiu, depicted here as the earnest and well-educated but as yet unenlightened senior student of the community, who after much consternation wrote the following:

The body is the *bodhi* tree.

The mind is like a bright mirror's stand.

At all times we must strive to polish it

and must not let dust collect.

Hongren praised this verse in public and instructed his students to recite and practice according to it, but in private he told Shenxiu that it was insufficient. Shortly thereafter, Huineng, who was working in the rice threshing room, heard Shenxiu's verse being recited by an acolyte. He immediately understood that its author was still unenlightened and offered his own response:

Bodhi originally has no tree.

The bright mirror also has no stand.

Fundamentally there is not a single thing.

Where could dust arise?

Hongren rejected this verse in public, but in private he bestowed both a final teaching and the status of Sixth Patriarch on Huineng, gave the still unordained layman his robe and bowl as proof of the transmission, and told him to leave the monastery. Huineng spent the next sixteen years living with hunters in the mountains before he presented himself at a Buddhist temple, was recognized as the Sixth Patriarch, and received ordination as a Buddhist monk.

It is the dust-polishing of the first verse that has been interpreted since at least the time of Zongmi (780-841) onward as an expression of gradualism, while the negations contained in the second verse are taken as evidence of subitism. The third line of the second verse, in fact, has been described by the modern Zen popularizer D.T. Suzuki as 'a bomb thrown into the camp of Shen-hsiu [Shenxiu] and his predecessors' (Suzuki 1949: 22). However, these interpretations, the religious identity of the Northern and Southern schools and the nature of the relationship between them are habitually misunderstood in popular writings on Chan.

First, note that there is no explicit reference to gradual improvement in 'Shenxiu's' verse; in fact, Shenxiu and his followers taught a doctrine of the constant and perfect practice of the *bodhisattva*. In the Chinese Buddhist context of the eighth century and beyond, gradualism was considered too elementary, limited by the dualistic conceptualization of enlightenment as a goal that should be reached through progressive effort. In fact, the historical Shenxiu's teachings are expressed in dualistic terms, but they are not gradualistic. Second, where is the subitism of 'Huineng's' verse? This verse cannot stand by itself, but is dependent on the first. If the original author of these verses understood Shenxiu's teachings, then the intent would have been to suggest that Huineng's were

similar in content but more advanced in that they went beyond conceptual or dualistic positions. This explanation of the meaning of the two verses is supported by the following details: (1) Shenxiu and Huineng were never at Hongren's side at the same time, so that the story is fictional; (2) the famous third line above is not contained in the earliest version of the text (from Dunhuang), which actually attributes two slightly different verses to Huineng; and (3) the terms used in both verses - including the third line cited here - were derived from Northern school sources (McRae 1986: 1-6, 235-38). The doctrines of the *Platform Sutra* are similar to those of the Oxhead school of early Chan, which made extensive use of a thesis-antithesis-synthesis form of doctrinal exposition.

Nevertheless, the populist image of Huineng as illiterate sage was profoundly important to the development of Chinese Chan, in that it provided a framework by which to emphasize the primacy of the actual experience of enlightenment over all other factors. By recognizing Huineng as Sixth Patriarch, the Chan tradition showed that - in theory, at least - it would not hesitate to recognize someone so déclassé as an illiterate woodcutter from the far south. Thus, in a fashion akin to the legendary acceptance of Shun as sage emperor, every subsequent selection of a successor by any teacher within Chan - no matter how well-placed the candidate might be in social, political, or economic terms - could similarly be ascribed to purely religious factors. The paradoxical impact of the Huineng story was thus strong ideological support for the developing lineage schema of Chan, in which masters were identified within specific networks of patriarchal relationship.

In addition, from the advent of the *Platform Sutra* the Chan school maintained a consistent preference for sudden over gradual, as well as an equal reluctance to use dualistic doctrinal formulations in teaching. This combination resulted in the development of a new type of religious practice: encounter dialogue, the sometimes spontaneous repartee between master and student that formed the new locus of spiritual cultivation. Huineng had exemplified an intuitive style of teaching that responded to the needs of individual students, and this image was transformed with time into the more dynamic persona of monks such as Linji (d. 866 or 867), who forced his students through their spiritual paces with shouts and fists. The texts of later periods of Chan are stylistically quite different from the *Platform Sutra*, which thus marked the end of early Chan even as it helped lay the groundwork for later developments.

See also: Buddhist philosophy, Chinese; Linji; Self-cultivation in Chinese philosophy; Zongmi

JOHN R. McRAE

References and further reading

McRae, J.R. (1986) *The Northern School and the Formation of Early Ch'an Buddhism*, Honolulu, HI: University of Hawaii Press.(A reappraisal of early Chan, including the *Platform Sutra*.)

Suzuki, D.T. (1949) *The Zen Doctrine of No-mind*, London: Rider.(A largely outmoded but widely known interpretation of Chan.)

Yampolsky, P.B. (ed. and trans.) (1967) *The Platform Sutra of the Sixth Patriarch*, New York: Columbia University Press.(The most widely used translation of the Dunhuang text, with a long introduction on the history of the text.)

Plato (427-347 BC)

Plato was an Athenian Greek of aristocratic family, active as a philosopher in the first half of the fourth century BC. He was a devoted follower of Socrates, as his writings make abundantly plain. Nearly all are philosophical dialogues - often works of dazzling literary sophistication - in which Socrates takes centre stage. Socrates is usually a charismatic figure who outshines a whole succession of lesser interlocutors, from sophists, politicians and generals to docile teenagers. The most powerfully realistic fictions among the dialogues, such as Protagoras and Symposium, recreate a lost world of exuberant intellectual self-confidence in an Athens not yet torn apart by civil strife or reduced by defeat in the Peloponnesian War.

Some of Plato's earliest writings were evidently composed in an attempt to defend Socrates and his philosophical mission against the misunderstanding and prejudice which - in the view of his friends - had brought about his prosecution and death. Most notable of these are Apology, which purports to reproduce the speeches Socrates gave at his trial, and Gorgias, a long and impassioned debate over the choice between a philosophical and a political life. Several early dialogues pit Socrates against practitioners of rival disciplines, whether rhetoric (as in Gorgias) or sophistic education (Protagoras) or expertise in religion (Euthyphro), and were clearly designed as invitations to philosophy as well as warnings against the pretensions of the alternatives. Apologetic and protreptic concerns are seldom entirely absent from any Platonic dialogue in which Socrates is protagonist, but in others among the early works the emphasis falls more heavily upon his ethical philosophy in its own right. For example, Laches (on courage) and Charmides (on moderation) explore these topics in characteristic Socratic style, relying mostly on his method of elenchus (refutation), although Plato seems by no means committed to a Socratic intellectualist analysis of the virtues as forms of knowledge. That analysis is in fact examined in these dialogues (as also, for example, in Hippias Minor).

In dialogues of Plato's middle period like Meno, Symposium and Phaedo a rather different Socrates is presented. He gives voice to positive positions on a much wider range of topics: not just ethics, but metaphysics and epistemology and psychology too. And he is portrayed as recommending a new and constructive instrument of inquiry borrowed from mathematics, the method of hypothesis. While there are continuities between Plato's early and middle period versions of Socrates, it is clear that an evolution has occurred. Plato is no longer a Socratic, not even a critical and original Socratic: he has turned Socrates into a Platonist.

The two major theories that make up Platonism are the theory of Forms and the doctrine of the immortality of the soul. The notion of a Form is articulated with the aid of conceptual resources drawn from Eleatic philosophy. The ultimate object of a philosopher's search for knowledge is a kind of being that is quite unlike the familiar objects of the phenomenal world: something eternal and changeless, eminently and exclusively whatever - beautiful or just or equal - it is, not qualified in time or place or relation or respect. An account of the Form of Beautiful will explain what it is for something to be beautiful, and indeed other things are caused to be beautiful by their participation in the Beautiful. The middle period dialogues never put forward any proof of the existence of Forms. The theory is usually presented as a basic assumption to which the interlocutors agree to subscribe. Plato seems to treat it as a very general high-level hypothesis which provides the framework within which other questions can be explored, including the immortality of the soul. According to Phaedo, such a hypothesis will only stand if its consequences are consistent with other relevant truths; according to Republic its validity must ultimately be assured by its coherence with the unhypothetical first principle constituted by specification of the Good.

The Pythagorean doctrine of the immortality of the soul, by contrast, is something for which Plato presents explicit proofs whenever he introduces it into discussion. It presupposes the dualist idea that soul and body are intrinsically distinct substances, which coexist during our life, but separate again at death. Its first appearance is in Meno, where it is invoked in explanation of how we acquire a priori knowledge of mathematical truths. Socrates is represented as insisting that nobody imparts such truths to us as information: we work them out for ourselves, by recollecting them from within, where they must have lain untapped as latent memory throughout our lives. But innate forgotten knowledge presupposes a time before the soul entered the body, when it was in full conscious possession of truth. Phaedo holds out the promise that the souls of philosophers who devote their lives to the pursuit of wisdom will upon death be wholly freed from the constraints and contaminations of the body, and achieve pure knowledge of the Forms once again.

Republic, Plato's greatest work, also belongs to this major constructive period of his philosophizing. It gives the epistemology and metaphysics of Forms a key role in political philosophy. The ideally just city (or some approximation to it), and the communist institutions which control the life of its elite governing class, could only become a practical possibility if philosophers were to acquire political power or rulers to engage sincerely and adequately in philosophy. This is because a philosopher-ruler whose emotions have been properly trained and disciplined by Plato's reforming educational programme, and whose mind has been prepared for abstract thought about Forms by rigorous and comprehensive study of mathematics, is the only person with the knowledge and virtue necessary for producing harmony in society. Understanding of Forms, and above all of the Good, keystone of the system of Forms, is thus the essential prerequisite of political order.

It remains disputed how far Plato's vision of a good society ruled by philosopher-statesmen (of both sexes) was ever really conceived as a blueprint for practical implementation. Much of his writing suggests a deep pessimism about the prospects for human happiness. The most potent image in Republic is the analogy of the cave, which depicts ordinary humanity as so shackled by illusions several times removed from the illumination of truth that only radical moral and intellectual conversion could redeem us. And its theory of the human psyche is no less dark: the opposing desires of reason, emotion and appetite render it all too liable to the internal conflict which constitutes moral disease.

While Republic is for modern readers the central text in Plato's œuvre*, throughout much of antiquity and the medieval period Timaeus was the dialogue by which he was best known. In this late work Plato offers an account of the creation of an ordered universe by a divine craftsman, who invests pre-existing matter with every form of life and intelligence by the application of harmonious mathematical ratios. This is claimed to be only a 'likely story', the best explanation we can infer for phenomena which have none of the unchangeable permanence of the Forms. None the less Timaeus is the only work among post-Republic dialogues, apart from a highly-charged myth in Phaedrus, in which Plato was again to communicate the comprehensive vision expressed in the Platonism of the middle period dialogues.*

Many of these dialogues are however remarkable contributions to philosophy, and none more so than the self-critical Parmenides. Here the mature Parmenides is represented as mounting a powerful set of challenges to the logical coherence of the theory of Forms. He urges not abandonment of the theory, but much harder work in the practice of dialectical argument if the challenges are to be met. Other pioneering explorations were in epistemology (Theaetetus) and philosophical logic (Sophist). Theaetetus mounts a powerful attack on Protagoras' relativist theory of truth, before grappling with puzzles about false belief and problems with the perennially attractive idea that knowledge is a complex built out of unknowable simples. Sophist engages with the Parmenidean paradox that what is not cannot be spoken or thought about. It forges fundamental distinctions between identity and predication and between subject and predicate in its attempt to rescue meaningful discourse from the absurdities of the paradox.

In his sixties Plato made two visits to the court of Dionysius II in Sicily, apparently with some hopes of exercising a beneficial influence on the young despot. Both attempts were abysmal failures. But they did not deter Plato from writing extensively on politics in his last years. Statesman explores the practical knowledge the expert statesman must command. It was followed by the longest, even if not the liveliest, work he ever wrote, the twelve books of Laws, perhaps still unfinished at his death.

1 Life

Evidence about Plato's life is *prima facie* plentiful. As well as several ancient biographies, notably that contained in book III of Diogenes Laertius' *Lives of the Philosophers*, we possess a collection of thirteen letters which purport to have been written by Plato. Unfortunately the biographies present what has been aptly characterized as 'a medley of anecdotes, reverential, malicious, or frivolous, but always piquant'. As for the letters, no scholar thinks them all authentic, and some judge that none are.

From the biographies it is safe enough to accept some salient points. Plato was born of an aristocratic Athenian family. He was brother to Glaucon and Adimantus, Socrates' main interlocutors in the *Republic*; his relatives included Critias and Charmides, members of the bloody junta which seized power in Athens at the end of the Peloponnesian War. He became one of the followers of Socrates, after whose execution he withdrew with others

among them to the neighbouring city of Megara. His travels included a visit to the court of Dionysius in Sicily. On returning from Sicily to Athens he began teaching in a gymnasium outside the city, called the Academy.

The *Seventh Letter*, longest and most interesting of the collection of letters, gives a good deal of probably trustworthy information, whether or not it was written by Plato himself. It begins with an account of his growing disenchantment with Athenian politics in early manhood and of his decision against a political career. This is prefatory to a sketch of the visit to Dionysius in Syracuse, which is followed by an elaborate self-justifying explanation of why and how, despite his decision, Plato later became entangled in political intrigue in Sicily, once the young Dionysius II had succeeded to his father's throne. There were two separate visits to the younger Dionysius: one (*c*.366 BC) is represented as undertaken at the behest of Dion, nephew of Dionysius I, in the hope of converting him into a philosopher-ruler; the other (*c*.360 BC) was according to the author an attempt to mediate between Dionysius and Dion, now in exile and out of favour. Both ventures were humiliating failures.

Of more interest for the history of philosophy is Plato's activity in the Academy. We should not conceive, as scholars once did, that he established a formal philosophical school, with its own property and institutional structures. Although he acquired a house and garden in the vicinity, where communal meals were probably taken, much of his philosophical teaching and conversation may well have been conducted in the public space of the gymnasium itself. Some sense of the Academy's distinctive style may be gleaned from evidence of the contemporaneous writings of the philosophical associates he attracted, notably his nephew Speusippus, Xenocrates, Aristotle and the mathematician Eudoxus. Discussion of Plato's metaphysical ideas figured prominently in these; but orthodoxy was not expected, to judge from their philosophical disagreements with him and with each other. Aristotle's early *Topics* suggests that an important role was played by formal disputation about philosophical theses.

From the educational programme of the *Republic* one might have guessed that Plato would have attached importance to the teaching of mathematics as a preparation for philosophy, but we have better evidence for his encouraging research in it. While he was not an original mathematician himself, good sources tell us that he formulated problems for others to solve: for example, what uniform motions will account for the apparent behaviour of the planets. Otherwise there is little reliable information on what was taught in the Academy: not much can be inferred from the burlesque of comic playwrights. Since almost certainly no fees were charged, most of those who came to listen to Plato (from all over the Greek world) must have been aristocrats. Some are known to have entered politics or to have advised princes, particularly on constitutional reform. But the Academy had no political mission of its own. Indeed the rhetorician Isocrates, head of a rival school and admittedly not an unbiased witness, dismissed the abstract disciplines favoured by the Academy for their uselessness in the real world.

2 Writings

Thrasyllus, astrologer to the emperor Tiberius, is the unlikely source of the arrangement of Platonic writings adopted in the manuscript tradition which preserves them. For his edition of Plato he grouped them into tetralogies, reminiscent of the trilogies produced in Athenian tragic theatre. These were organized according to an architectonic scheme constructed on principles that are now only partially apparent, but certainly had nothing to do with chronology of composition. His arrangement began with a quartet 'designed to show what the life of the philosopher is like' (Diogenes Laertius, III 57): *Euthyphro*, or 'On Piety', classified as a 'peirastic' or elenctic dialogue (see Socrates §§3-4), which is a species of one of his two main genres, the dialogue of inquiry; *Apology*, *Crito* and *Phaedo* are all regarded as specimens of exposition, his other main genre, or more specifically as specimens of ethics. These four works are all concerned in one way or another with the trial and death of Socrates.

There followed a group consisting of *Cratylus*, or 'On the Correctness of Names', *Theaetetus*, or 'On Knowledge', *Sophist* and *Politicus* (often Anglicized as *Statesman*). Plato himself indicates that the last three of this set are to be read together. They contain some of his most mature and challenging work in epistemology, metaphysics and philosophical methodology. In this they resemble *Parmenides*, with its famous critique of the theory of Forms, the first of the next tetralogy, which was completed by three major dialogues all reckoned 'ethical' by Thrasyllus: *Philebus*, an examination of pleasure, *Symposium* and *Phaedrus*, both brilliant literary divertissements which explore the nature of love.

A much slighter quartet came next: two dialogues entitled *Alcibiades*, plus *Hipparchus* and *Rivals*. None of these,

with the disputed exception of the first *Alcibiades*, is thought by modern scholarship to be authentic Plato. They were followed by *Theages*, a short piece now generally reckoned spurious, *Charmides*, *Laches*, *Lysis*. These three works are generally regarded by modern scholars as Socratic dialogues: that is, designed to exhibit the distinctive method and ethical preoccupations of the historical Socrates, at least as Plato understood him, not to develop Plato's own philosophy. Thrasyllus would agree with the latter point, since he made them dialogues of inquiry: *Laches* and *Lysis* 'maieutic', in which the character 'Socrates' attempts as intellectual midwife to assist his interlocutors to articulate and work out their own ideas on courage and friendship respectively; *Charmides* elenctic, with the interlocutors Charmides and Critias and their attempts to say what moderation is put to the test of cross-examination, something Thrasyllus interestingly distinguished from philosophical midwifery.

The next group consisted of *Euthydemus*, *Protagoras*, *Gorgias*, *Meno*, important works in which modern scholarship finds analysis and further elaboration by Plato of the Socratic conception of virtue. The first three present a Socrates in argumentative conflict with sophists of different sorts (see Sophists), so it is understandable that under the general heading 'competitive' Thrasyllus characterized *Euthydemus* and *Gorgias* as dialogues of refutation, and *Protagoras* as a dialogue of display - presumably because Protagoras and Socrates are each portrayed as intent on showing off their debating skills. *Meno*, on the other hand, is labelled an elenctic work. It was followed by the seventh tetralogy: *Hippias Major* and *Hippias Minor*, two very different dialogues (of refutation, according to Thrasyllus), both featuring the sophist of that name; *Ion*, a curious piece on poetic performance; and *Menexenus*, a still more curious parody of a funeral oration, put in the mouth of Pericles' mistress Aspasia.

For the last two tetralogies Thrasyllus reserved some of Plato's major writings. The eighth contained the very brief (and conceivably spurious) *Clitophon*, in which a minor character from the *Republic* plays variations on themes in the *Republic*, the second dialogue in the group, and generally regarded nowadays as Plato's greatest work. This quartet was completed by *Timaeus* and its unfinished sequel *Critias*, no doubt because these dialogues represent themselves as pursuing further the discussions of the *Republic*. The pre-Copernican mathematical cosmology of *Timaeus* no longer attracts readers as it did throughout antiquity, and particularly in the Middle Ages, when the dialogue was for a period the only part of Plato's *œuvre* known to the Latin West. Finally, the ninth tetralogy began with the short *Minos*, a spurious dialogue taking up issues in the massive *Laws*, Plato's longest and probably latest work, which was put next in the group. Then followed *Epinomis*, an appendix to *Laws* already attributed to one of Plato's pupils in antiquity (Philip of Opous, according to a report in Diogenes Laertius, III 37). Last were placed the *Letters*, briefly discussed above.

3 Authenticity and chronology

Thrasyllus rejected from the canon a variety of minor pieces, some of which still survive through the manuscript tradition. Modern judgment concurs with the ancient verdict against them. It also questions or rejects some he thought genuinely Platonic. But we can be fairly sure that we still possess everything Plato wrote for publication.

Attempting to determine the authenticity or inauthenticity of ancient writings is a hazardous business. Egregious historical errors or anachronisms suffice to condemn a work, but except perhaps for the *Eighth Letter*, this criterion gets no purchase on the Platonic corpus. Stylistic analysis of various kinds can show a piece of writing to be untypical of an author's *œuvre*, without thereby demonstrating its inauthenticity: *Parmenides* is a notable example of this. Most of Plato's major dialogues are in fact attested as his by Aristotle. The difficult cases are short pieces such as *Theages* and *Clitophon*, and, most interestingly, three more extended works: the *Seventh Letter*, *Alcibiades I* and *Hippias Major*. Opinion remains divided on them. Some scholars detect crude or sometimes brilliant pastiche of Plato's style; a parasitic relationship with undoubtedly genuine dialogues; a philosophical crassness or a misunderstanding of Platonic positions which betrays the forger's hand. Yet why should Plato not for some particular purpose recapitulate or elaborate things he has said elsewhere? And perhaps he did sometimes write more coarsely or didactically or long-windedly than usual. Such assessments are inevitably matters of judgment, on which intelligent and informed readers will legitimately differ.

Prospects for an absolute chronology of Plato's writings are dim. There are no more than two or three references to datable contemporaneous events in the entire corpus (leaving aside the *Letters*). Relative chronology is another matter. Some dialogues refer back to others. A number of instances have been mentioned already, but we can add a clear reminiscence of *Meno* in *Phaedo* (72e-73b), and of *Parmenides* in both *Theaetetus* (183e-184a) and *Sophist*

(217c). According to one ancient tradition *Laws* was unfinished at Plato’s death, and Aristotle informs us that it was written after *Republic* (*Politics* 1264b24- 7), to which it appears to allude (see, for example, *Laws* 739a-e). Attempts have sometimes been made to find evidence, whether internal or external, for the existence of early versions of works we possess in different form (see for example Thesleff 1982). One example is the suggestion that Aristophanes’ comedy *Ecclesiazousae* or *Assembly of Women* (388 BC) was parodying an early version of book V of *Republic*. But while the idea that Plato may have revised some of his writings is plausible, concrete instances in which such revision is plainly the best explanation of the phenomena are hard to find. Even if they were not, it is unlikely that the consequences for relative chronology would be clear.

For over a century hopes for a general relative chronology of Plato’s writings have been pinned on the practice of stylistic analysis. This was pioneered by Lewis Campbell in his edition of *Sophist* and *Politicus*, published in 1867. His great achievement was to isolate a group of dialogues which have in common a number of features (added to by subsequent investigators) that set them apart from all the rest. *Timaeus*, *Critias*, *Sophist*, *Politicus*, *Philebus* and *Laws* turn out to share among other things a common technical vocabulary; a preference for certain particles, conjunctions, adverbs and other qualifiers over alternatives favoured in other dialogues; distinctive prose rhythms; and the deliberate attempt to avoid the combination of a vowel at the end of one word followed by another vowel at the beginning of the next. Since there are good independent reasons for taking *Laws* to be Plato’s last work, Campbell’s sextet is very likely the product of his latest phase of philosophical activity. Application of the same stylistic tests to the Platonic corpus as a whole, notably by Constantin Ritter (1888), established *Republic*, *Theaetetus* and *Phaedrus* as dialogues which show significantly more of the features most strongly represented in the late sextet than any others. There is general agreement that they must be among the works whose composition immediately precedes that of the *Laws* group, always allowing that *Republic* must have taken several years to finish, and that parts of it may have been written earlier and subsequently revised. *Parmenides* is ordinarily included with these three, although mostly on non-stylistic grounds.

Since Campbell’s time there have been repeated attempts by stylometrists to divide the remaining dialogues into groups, and to establish sequences within groups. The heyday of this activity was in the late nineteenth and early twentieth centuries. Since the 1950s there has been a revival in stylistic study, with the use of increasingly sophisticated statistical techniques and the resources of the computer and the database. Secure results have proved elusive. Most scholars would be happy to date *Phaedo*, *Symposium* and *Cratylus* to a middle period of Plato’s literary and philosophical work which may be regarded as achieving its culmination in *Republic*. But while this dating is sometimes supported by appeal to stylistic evidence, that evidence is in truth indecisive: the hypothesis of a middle period group of dialogues really rests on their philosophical affinities with *Republic* and their general literary character. The same can be said *mutatis mutandis* of attempts to identify a group assigned to Plato’s early period.

The cohesiveness of Campbell’s late group has not gone unchallenged. For example, in 1953 G.E.L. Owen mounted what for a while seemed to some a successful attack on his dating of *Timaeus* and *Critias*, on the ground that these dialogues belong philosophically in Plato’s middle period. Broadly speaking, however, stylistic studies have helped to establish an agreed chronological framework within which most debates about philosophical interpretation now take place. This is not to say however that there is unanimity either about the way Plato’s thought developed or about the importance of the notion of development for understanding his philosophical project or projects in the dialogues.

4 The Platonic dialogue

Who invented the philosophical dialogue, and what literary models might have inspired the invention, are not matters on which we have solid information. We do know that several of Socrates’ followers composed what Aristotle calls *Sōkratikoi logoi*, discourses portraying Socrates in fictitious conversations (see Socratic dialogues). The only examples which survive intact besides Plato’s are by Xenophon, probably not one of the earliest practitioners of the genre.

One major reason for the production of this literature was the desire to defend Socrates against the charges of irreligion and corrupting young people made at his trial and subsequently in Athenian pamphleteering, as well as the implicit charge of guilt by association with a succession of oligarchic politicians. Thus his devotion to the unstable and treacherous Alcibiades was variously portrayed in, for example, the first of the *Alcibiades* dialogues

ascribed to Plato and the now fragmentary *Alcibiades* of Aeschines of Sphettos, but both emphasized the gulf between Alcibiades' self-conceit and resistance to education and Socrates' disinterested concern for his moral wellbeing. The same general purpose informed the publication of versions of Socrates' speech (his 'apology') before the court by Plato, Xenophon and perhaps others. Writing designed to clear Socrates' name was doubtless a particular feature of the decade or so following 399 BC, although it clearly went on long after that, as in Xenophon's *Memorabilia* (see Xenophon §2). After starting in a rather different vein *Gorgias* turns into Plato's longest and angriest dialogue of this kind. Socrates is made to present himself as the only true politician in Athens, since he is the one person who can give a truly rational account of his conduct towards others and accordingly command the requisite political skill, which is to make the citizens good. But he foresees no chance of acquittal by a court of jurors seeking only gratification from their leaders.

Placing Socrates in opposition to Alcibiades is a way of defending him. Arranging a confrontation between a sophist (Protagoras or Hippias) or a rhetorician (Gorgias) or a religious expert (Euthyphro) or a Homeric recitalist (Ion) and Socrates is a way of exposing their intellectual pretensions, and in most cases their moral shallowness, while celebrating his wit, irony and penetration and permitting his distinctive ethical positions and ethical method to unfold before the reader's eyes. The elenchus (see Socrates §§3-4) is by no means the only mode of argument Socrates is represented as using in these fictional encounters. Plato particularly enjoys allowing him to exploit the various rhetorical forms favoured by his interlocutors. But it is easy to see why the dialogue must have seemed to Plato the ideal instrument not only for commemorating like Xenophon Socrates' style of conversation, but more importantly for exhibiting the logical structure and dynamic of the elenchus, and its power in Socrates' hands to demolish the characteristic intellectual postures of those against whom it is deployed.

In these dialogues of confrontation Socrates seldom succeeds in humbling his interlocutors into a frank recognition that they do not know what they thought they knew: the official purpose - simultaneously intellectual and moral - of the elenchus. It would not have been convincing to have him begin to convert historical figures with well-known intellectual positions. The main thing registered by their fictional counterparts is a sense of being manipulated into self-contradiction. In any case, the constructive response to the extraordinary figure of Socrates which Plato really wants to elicit is that of the reader. We have to suppose that, as conversion to philosophy was for Plato scarcely distinguishable from his response to Socrates (devotion to the man, surrender to the spell of his charisma, strenuous intellectual engagement with his thought and the questions he was constantly pursuing), so he conceived that the point of writing philosophy must be to make Socrates charismatic for his readers - to move us to similar devotion and enterprise. In short, the dialogues constitute simultaneously an invitation to philosophy and a critique of its intellectual rivals.

Whatever Plato's other accomplishments or failures as a writer and thinker, one project in which he unquestionably succeeds is in creating a Socrates who gets under the reader's skin (see Socrates §7). Plato has a genius for portrayal of character: the 'arrogant self-effacement' of Socrates' persona; the irony at once sincere and insincere; the intellectual slipperiness in service of moral paradox; the nobility of the martyr who loses everything but saves his own soul, and of the hero who stands firm on the battlefield or in face of threats by the authorities; relentless rationality and almost impregnable self-control somehow cohabiting with susceptibility to beautiful young men and their erotic charm. Also important is the ingenious variety of perspectives from which we see Socrates talking and interacting with others. Sometimes he is made to speak to us direct (for example, *Apology*, *Gorgias*). Sometimes Plato invites us to share complicity in a knowing narrative Socrates tells of his own performance (as in *Charmides*, *Protagoras*). Sometimes someone else is represented as recalling an unforgettably emotional occasion when Socrates dominated a whole roomful of people, as in the most powerfully dramatic dialogues of all, *Phaedo* and *Symposium*. Here we have the illusion that Socrates somehow remains himself even though the ideas advanced in them must go beyond anything that the historical Socrates (or at any rate the agnostic Socrates of *Apology*) would have claimed about the soul and its immortality or about the good and the beautiful.

5 The problem of writing

It might seem strange that an original philosopher of Plato's power and stature should be content, outside the *Letters* if some of them are by him, never to talk directly to the reader, but only through the medium of narrative or dramatic fiction, even granted the pleasure he plainly takes in exhibiting his mastery of that medium. This will become less mysterious if we reflect further on Socrates and Socratic questioning. At any rate by the time of the

Meno, Plato was wanting to suggest that the elenchus presupposes that understanding is not something one person can transmit in any straightforward way to another, but something which has to be worked out for oneself and recovered from within by recollection. The suggestion is made by means of an example from mathematics, where it is transparently true that seeing the answer to a problem is something that nobody else can do for us, even if Socrates' questions can prompt us to it. The moral we are to draw is that in pressing his interlocutors on what they say they believe, Socrates is merely an intellectual midwife assisting them to articulate for themselves a more coherent and deeply considered set of views, which will ideally constitute the truth.

The Platonic dialogue can be interpreted as an attempt to create a relationship between author and reader analogous to that between Socrates and his interlocutors. Given that that relationship is to be construed in the way indicated in *Meno*, the point of a dialogue will be like that of the elenchus: not to teach readers the truth (it is strictly speaking unteachable), but to provoke and guide them into working at discovering it for themselves. Most of the dialogues of Campbell's late sextet are admittedly more didactic than one would expect on this view of the dialogue, and it is significant that except in *Philebus* Socrates is no longer the main speaker. Yet even here use of the dialogue form can be taken as symbolizing that responsibility for an active philosophical engagement with what Plato has written rests with the reader, as the difficulty and in some cases the methodological preoccupations of most of these works confirms.

In a much discussed passage at the end of *Phaedrus* (275-8), Socrates is made to speak of the limitations of the written word. It can answer no questions, it cannot choose its readers, it gets misunderstood with no means of correcting misunderstanding. Its one worthwhile function is to remind those who know of what they know. By contrast with this dead discourse live speech can defend itself, and will be uttered or not as appropriate to the potential audience. The only serious use of words is achieved when speech, not writing, is employed by dialecticians to sow seeds of knowledge in the soul of the learner. If they commit their thoughts to writing they do so as play (*paidia*). The *Seventh Letter* (341-2) makes related remarks about the writing of philosophy; and at various points in, for example, *Republic*, *Timaeus* and *Laws*, the discussions in which the interlocutors are engaged are described as play, not to be taken seriously.

Interpreters have often taken these written remarks about writing with the utmost seriousness. In particular the Tübingen school of Platonic scholarship has connected them with references, especially in Aristotle, to unwritten doctrines of Plato. They have proposed that the fundamental principles of his philosophy are not worked out in the dialogues at all, but were reserved for oral discussions in the Academy, and have to be reconstructed by us from evidence about the unwritten doctrines. But this evidence is suspect where voluble and elusive when apparently more reliable. There are two star exhibits. First, according to the fourth century BC music theorist Aristoxenus, Aristotle used to tell of how when Plato lectured on the good he surprised and disappointed his listeners by talking mostly about mathematics (*Harmonics II*, 30.16-31.3). Second, at one point in the *Physics* (209b13-6) Aristotle refers to Plato's 'so-called unwritten teachings'; and the Aristotelian commentators report that Aristotle and other members of the Academy elsewhere wrote more about them. Plato's key idea was taken to be the postulation of the One and the great and the small, or 'indefinite dyad' as principles of all things, including Forms. In his *Metaphysics* (I.6) Aristotle seems to imply that in this theory the Forms were construed in some sense as numbers. It remains obscure and a subject of inconclusive scholarly debate how far the theory was worked out, and what weight we should attach to it in comparison to the metaphysical explorations of the dialogues of Plato's middle and late periods (see for example Ross 1951, Gaiser 1968, Guthrie 1978, Gaiser 1980, Burnyeat 1987).

The general issue of how far we can ascribe to Plato things said by interlocutors (principally Socrates) in his dialogues is something which exercises many readers. The position taken in this entry will be that no single or simple view of the matter is tenable: sometimes, for example, Plato uses the dialogue form to work through a problem which is vexing him; sometimes to recommend a set of ideas to us; sometimes to play teasingly with ideas or positions or methodologies without implying much in the way of commitment; and frequently to suggest to us ways we should or should not ourselves try to philosophize. As for the Tübingen school, we may agree with them that when it introduces the Form of the Good the *Republic* itself indicates that readers are being offered only conjectures and images, not the thorough dialectical discussion necessary for proper understanding. But the notions of seriousness and play are less straightforward than they allow. Playing with ideas - that is, trying them out and developing them to see what might work and what will not - is the way new insights in philosophy and science are often discovered. When we meet it in Plato's dialogues it usually seems fun without being frivolous.

Nor should we forget that the Platonic dialogue represents itself as a spoken conversation. It seems hard to resist the thought that we are thereby invited to treat his dialogues not as writing so much as an attempt to transcend the limitations of writing. Perhaps the idea is that they can achieve the success of living speech if treated not as texts to be interpreted (despite Plato's irresistible urge to produce texts devised precisely to elicit attempts at interpretation), but as stimuli to questions we must put principally to ourselves, or as seeds which may one day grow into philosophy in our souls.

6 Early works

There is widespread scholarly agreement that the following are among Plato's earliest writings: *Apology*, *Crito*, *Ion*, *Hippias Minor*, *Laches* and *Charmides*. *Apology*, as we have noted, best fits into the context of the decade following Socrates' death, and so does *Crito*, which explores the question why he did not try to escape from the condemned cell; the others are all short treatments of questions to do with virtue and knowledge, or in the case of *Ion*, with expertise (*technē*), and all are relatively simple in literary structure. The brief *Euthyphro* and the much longer *Protagoras* and *Gorgias* (with which *Menexenus* is often associated) are usually seen as having many affinities with these, and so are put at least fairly early, although here anticipations of the style or content of the mature middle-period dialogues have also been detected. The connections in thought between *Lysis*, *Euthydemus* and *Hippias Major* and middle-period Plato may be argued to be stronger still, even though there remain clear similarities with the dialogues generally accepted as early. We do not know whether Plato wrote or published anything before Socrates' death; *Menexenus* cannot be earlier than 386 BC, *Ion* might be datable to around 394-391 BC, but otherwise we can only guess.

All those listed above fall under the commonly used description 'Socratic dialogues', because they are seen as preoccupied with the thought of the historical Socrates as Plato understood him, in contrast with writings of the middle period, where 'Socrates' often seems to become a vehicle for exploring a more wide-ranging set of ideas (see Socrates §2). In the Socratic dialogues discussion is confined almost exclusively to ethical questions, or problems about the scope and credentials of expertise: metaphysics and epistemology and speculation about the nature and powers of the soul are for the most part notable by their absence. Use of the elenchus is prominent in them as it is not, for example, in *Republic* (apart from book I, sometimes regarded as an early work subsequently reused as a preface to the main body of the dialogue). The hypothesis that philosophizing in this style was the hallmark of the historical Socrates is broadly consistent with what we are given to understand about him by Xenophon, Aristotle and Plato's *Apology* - which is usually thought to be particularly authoritative evidence, whether or not it is a faithful representation of what Socrates really said at his trial.

How historical the historical Socrates of the hypothesis actually is we shall never know. The conjecture that many of the Socratic dialogues are early works is likewise only a guess, which gets no secure support from stylometric evidence. None the less the story of Plato's literary and philosophical development to which it points makes such excellent sense that it has effectively driven all rival theories from the field. The placing of individual dialogues within that story remains a matter for controversy; and doubts persist over how far interpretation of Plato is illuminated or obstructed by acceptance of any developmental pattern. With these provisos, the account which follows assumes the existence of a group of early Socratic dialogues in the sense explained.

The convenience of the description 'Socratic dialogues' should not generate the expectation of a single literary or philosophical enterprise in these writings. It conceals considerable variety, for example as between works devoted to articulating and defending the philosophical life and works which problematize Socratic thought as much as they exhibit its attractions. This distinction is not an exhaustive one, but provides useful categories for thinking about some of the key productions of Plato's early period.

7 Apologetic writings

Moral, or indeed existential, choice, to use an anachronistic expression, is the insistent focus of *Apology*. God has appointed Socrates, as he represents it to his judges, to live the philosophical life, putting himself and others under constant examination. The consistency of his commitment to this mission requires him now to face death rather than abandon his practice of philosophy, as he supposes for the sake of argument the court might require him to do. For confronted with the choice between disobeying God (that is, giving up philosophy) and disobeying human dictate (that is, refusing to do so), he can only take the latter option. What governs his choice is justice:

> It is a mistake to think that a man worth anything at all should make petty calculations about the risk of living or dying. There is only one thing for him to consider when he acts: whether he is doing right or wrong, whether he is doing what a good man or a bad man would do.
>
> (*Apology* 28b)

Whether death is or is not a bad thing Socrates says he does not know. He does know that behaving wrongly and disobeying one's moral superior - whether divine or human - is bad and shameful. The demands of justice, as his conscience (or 'divine sign') interpreted them, had earlier led him to choose the life of a private citizen, conversing only with individuals, rather than the political life: for justice and survival in politics are incompatible. When he did carry out the public obligations of a citizen and temporarily held office, justice again compelled him to choose the dangerous and unpopular course of resisting a proposal that was politically expedient but contrary to the law. As for those with whom he talked philosophy, they too faced a choice: whether to make their main concern possessions and the body, or virtue and the soul; that is, what belongs to oneself, or oneself. And now the judges too must choose and determine what is just as their oath requires of them.

Crito and *Gorgias* continue the theme in different ways. *Crito* has often been found difficult to reconcile with *Apology* when it argues on various grounds (paternalistic and quasi-contractual) that citizens must always obey the law, unless they can persuade it that it is in the wrong. Hence, since the law requires that Socrates submit to the punishment prescribed by the court, he must accept the sentence of death pronounced on him. The higher authority of divine command stressed in *Apology* seems to have been forgotten. Once again, however, the whole argument turns on appeal to justice and to the choices it dictates: we must heed the truth about it, not what popular opinion says; we must decide whether or not we believe the radical Socratic proposition that retaliation against injury or injustice is never right (see Socrates §4). *Gorgias*, one of the longest of all the dialogues, ranges over a wide territory, but at its heart is the presentation of a choice. Socrates addresses Callicles, in whose rhetoric Nietzsche saw an anticipation of his ideal of the superman:

> You see that the subject of our arguments - and on what subject should a person of even small intelligence be more serious? - is this: what kind of life should we live? The life which you are now urging upon me, behaving as a *man* should: speaking in the assembly and practising rhetoric and engaging in politics in your present style? Or the life of philosophy?
>
> (*Gorgias* 500c)

The dialogue devotes enormous energy to arguing that only philosophy, not rhetoric, can equip us with a true expertise which will give us real power, that is power to achieve what we want: the real not the apparent good. Only philosophy can articulate a rational and reliable conception of happiness - which turns out to depend on justice.

8 *Laches* and *Charmides*

Contrast the works outlined in §7 with *Laches* and *Charmides*, which were very likely conceived as a pair, the one an inquiry into courage, the other into *sōphrosynē* or moderation. Both engage in fairly elaborate scene setting quite absent from *Crito* and *Gorgias*. In both there is concern with the relation between theory and practice, which is worked out more emphatically in *Laches*, more elusively in *Charmides*. For example, in *Laches* Socrates is portrayed both as master of argument about courage, and as an exemplar of the virtue in action - literally by reference to his conduct in the retreat from Delium early in the Peloponnesian War, metaphorically by his persistence in dialectic, to which his observations on the need for perseverance in inquiry draw attention.

A particularly interesting feature of these dialogues is their play with duality. Socrates confronts a *pair* of main interlocutors who clearly fulfil complementary roles. We hear first the views of the more sympathetic members of the two pairs: the general Laches, whom Socrates identifies as his partner in argument, and the young aristocrat Charmides, to whom he is attracted. Each displays behavioural traits associated with the virtue under discussion, and each initially offers a definition in behavioural terms, later revised in favour of a dispositional analysis: courage is construed as a sort of endurance of soul, *sōphrosynē* as modesty. After these accounts are subjected to elenchus and refuted, the other members of the pairs propose intellectualist definitions: according to Nicias (also a general), courage is knowledge of what inspires fear or confidence, while Critias identifies *sōphrosynē* with self-knowledge.

Broad hints are given that the real author of these latter definitions is Socrates himself; and in *Protagoras* he is made to press Protagoras into accepting the same definition of courage. There are also hints that, as understood by their proponents here, this intellectualism is no more than sophistic cleverness, and that neither possesses the virtue he claims to understand. Both are refuted by further Socratic elenchus, and in each case the argument points to the difficulty of achieving an intellectualist account which is not effectively a definition of virtue in general as the simple knowledge of good and bad. *Laches* explicitly raises the methodological issue of whether one should try to investigate the parts of virtue in order to understand the whole or vice versa (here there are clear connections with the main argument of *Protagoras*).

Aristotle was in no doubt that Socrates 'thought all the virtues were forms of knowledge' (*Eudemian Ethics* 1216b6); and many moves in the early dialogues depend on the assumption that if you know what is good you will *be* good (see Socrates §5). But *Laches* and *Charmides* present this Socratic belief as problematical. Not only is there the problem of specifying a unique content for the knowledge with which any particular virtue is to be identified. There is also the difficulty that any purely intellectual specification of what a virtue is makes no reference to the dispositions Charmides and Laches mention and (like Socrates) exemplify. In raising this difficulty Plato is already adumbrating the need for a more complex moral psychology than Socrates', if only to do justice to how Socrates *lived*. If the viewpoints of Laches and Nicias are combined we are not far from the account of courage in *Republic*, as the virtue of the spirited part of the soul, which 'preserves through pains and pleasures the injunctions of reason concerning what is and is not fearful' (442b).

9 Other dialogues of inquiry

In *Protagoras* it is Socrates himself who works out and defends the theory that knowledge is sufficient for virtuous action and that different virtues are different forms of that knowledge (see Aretē). He does not here play the role of critic of the theory, nor are there other interlocutors who might suggest alternative perceptions: indeed Protagoras, as partner not adversary in the key argument, is represented as accepting the key premise that (as he puts it) 'wisdom and knowledge are the most powerful forces governing human affairs' (352c-d). It would be a mistake to think that Plato found one and the same view problematic when he wrote *Laches* and *Charmides* but unproblematic when he wrote *Protagoras*, and to construct a chronological hypothesis to cope with the contradiction. *Protagoras* is simply a different sort of dialogue: it displays Socratic dialectic at work from a stance of some detachment, without raising questions about it. *Protagoras* is an entirely different kind of work from *Gorgias*, too: the one all urbane sparring, the latter a deadly serious confrontation between philosophy and political ambition. *Gorgias* unquestionably attacks hedonism, *Protagoras* argues for it, to obtain a suitable premise for defending the intellectualist paradox that nobody does wrong willingly, but leaves Socrates' own commitment to the premise at best ambiguous (see Socrates §6). Incommensurabilities of this kind make it unwise to attempt a relative chronology of the two dialogues on the basis of apparent incompatibilities in the positions of their two Socrates.

Space does not permit discussion of *Ion*, or of *Hippias Minor*, in which Socrates is made to tease us with the paradox - derived from his equation of virtue and knowledge - that someone who *did* do wrong knowingly and intentionally would be better than someone who did it unintentionally through ignorance. Interpretation of *Euthyphro* remains irredeemably controversial. Its logical ingenuity is admired, and the dialogue is celebrated for its invention of one of the great philosophical questions about religion: either we should do right because god tells us to do so, which robs us of moral autonomy, or because it is right god tells us to do it, which makes the will of god morally redundant.

Something more needs to be said about *Lysis* and *Euthydemus* (which share a key minor character in Ctesippus, and are heavy with the same highly charged erotic atmosphere) and *Hippias Major*. They all present Socrates engaging in extended question and answer sessions, although only in *Hippias* is this an elenchus with real bite: in the other dialogues his principal interlocutors are boys with no considered positions of their own inviting refutation. All end in total failure to achieve positive results. All make great formal play with dualities of various kinds. Unusually ingenious literary devices characterize the three works, ranging from the introduction of an *alter ego* for Socrates in *Hippias* to disruption of the argument of the main dialogue by its 'framing' dialogue in *Euthydemus*, at a point where the discussion is clearly either anticipating or recalling the central books of *Republic*. All seem to be principally preoccupied with dialectical method (admittedly a concern in every dialogue). Thus *Hippias* is a study in definitional procedure, applied to the case of the fine or beautiful, *Lysis* a study in thesis

and antithesis paralleled in Plato's *œuvre* only by *Parmenides*, and *Euthydemus* an exhibition of the contrast between 'eristic', that is, purely combative sophistical argument, demonstrated by the brothers Euthydemus and Dionysodorus, and no less playful philosophical questioning that similarly but differently ties itself in knots. It is the sole member of the trio which could be said with much conviction to engage - once more quizzically - with the thought of the historical Socrates about knowledge and virtue. But its introduction of ideas from *Republic* makes it hard to rank among the early writings of Plato. Similarly, in *Lysis* and *Hippias Major* there are echoes or pre-echoes of the theory of Forms and some of the causal questions associated with it. We may conclude that these ingenious philosophical exercises - 'gymnastic' pieces, to use the vocabulary of *Parmenides* - might well belong to Plato's middle period.

10 The introduction of Platonism

Needless to say, no explicit Platonic directive survives encouraging us to read *Meno*, *Symposium* and *Phaedo* together. But there are compelling reasons for believing that Plato conceived them as a group in which *Meno* and *Symposium* prepare the way for *Phaedo*. In brief, in *Meno* Plato introduces his readers to the non-Socratic theory of the immortality of the soul and a new hypothetical method of inquiry, while *Symposium* presents for the first time the non-Socratic idea of a Platonic Form, in the context of a notion of philosophy as desire for wisdom. It is only in *Phaedo* that all these new ideas are welded together into a single complex theory incorporating epistemology, psychology, metaphysics and methodology, and constituting the distinctive philosophical position known to the world as Platonism.

Meno and *Symposium* share two features which indicate Plato's intention that they should be seen as a pair, performing the same kind of introductory functions, despite enormous differences for example in dialogue form, scale and literary complexity. First, both are heavily and specifically foreshadowed in *Protagoras*, which should accordingly be reckoned one of the latest of Plato's early writings. At the end of *Protagoras* (361c) Socrates is made to say that he would like to follow up the inconclusive conversation of the dialogue with another attempt to define what virtue is, and to consider again whether or not it can be taught. This is exactly the task undertaken in *Meno*. Similarly, not only are all the *dramatis personae* of *Symposium* except Aristophanes already assembled in *Protagoras*, but at one point Socrates is represented as offering the company some marginally relevant advice on how to conduct a drinking party - which corresponds exactly to what happens at the party in *Symposium* (347c-348a).

Second, both *Meno* and *Symposium* are exceedingly careful not to make Socrates himself a committed proponent either of the immortality of the soul or of the theory of Forms. These doctrines are ascribed respectively to 'priests and priestesses' (*Meno*) and to one priestess, Diotima, in particular (*Symposium*); in *Meno* Socrates says he will not vouch for the truth of the doctrine of immortality, in *Symposium* he records Diotima's doubts as to whether he is capable of initiation into the mysteries (a metaphor also used of mathematics in *Meno*) which culminate in a vision of the Form of the Beautiful. In *Symposium* these warning signs are reinforced by the extraordinary form of the dialogue: the sequence of conversations and speeches it purports to record are nested inside a Chinese box of framing conversations, represented as occurring some years later and with participants who confess to inexact memory of what they heard.

Phaedo for its part presupposes *Meno* and *Symposium*. At 72e-73b *Meno*'s argument for the immortality of the soul is explicitly recalled, while the Form of Beauty is regularly mentioned at the head of the lists of the 'much talked about' Forms which *Phaedo* introduces from time to time (for example, 75c, 77a, 100b). It is as though Plato relies upon our memory of the much fuller characterization of what it is to be a Form supplied in *Symposium*. Unlike *Meno* and *Symposium*, *Phaedo* represents Socrates himself as committed to Platonist positions, but takes advantage of the dramatic context - a discussion with friends as he waits for the hemlock to take effect - and makes him claim prophetic knowledge for himself like a dying swan (84e-85b). The suggestion is presumably that Platonism is a natural development of Socrates' philosophy even if it goes far beyond ideas about knowledge and virtue and the imperatives of the philosophical life to which he is restricted in the early dialogues.

11 *Meno*

Meno is a dialogue of the simplest form and structure. It consists of a conversation between Socrates and Meno, a young Thessalian nobleman under the spell of the rhetorician Gorgias, interrupted only by a passage in which

Socrates quizzes Meno's slave, and then later by a brief intervention in the proceedings on the part of Anytus, Meno's host and one of Socrates' accusers at his trial. The dialogue divides into three sections: an unsuccessful attempt to define what virtue is, which makes the formal requirements of a good definition its chief focus; a demonstration in the face of Meno's doubts that successful inquiry is none the less possible in principle; and an investigation into the secondary question of whether virtue can be taught, pursued initially by use of a method of hypothesis borrowed from mathematics. Although the ethical subject matter of the discussion is thoroughly Socratic, the character and extent of its preoccupation with methodology and (in the second section) epistemology and psychology are not. Nor is *Meno*'s use of mathematical procedures to cast light on philosophical method; this is not confined to the third section. Definitions of the mathematical notion of shape are used in the first section to illustrate for example the principle that a definition should be couched in terms that the interlocutor agrees are already known. And the demonstration of an elenchus with a positive outcome which occupies the second is achieved with a geometrical example.

It looks as though Plato has come to see in the analogy with mathematics hope for more constructive results in philosophy than the Socratic elenchus generally achieved in earlier dialogues. This is a moral which the second and third sections of *Meno* make particularly inviting to draw. In the second Socrates is represented as setting Meno's untutored slave boy a geometrical problem (to determine the length of the side of a square twice the size of a given square) and scrutinizing his answers by the usual elenctic method. The boy begins by thinking he has the answer. After a couple of mistaken attempts at it he is persuaded of his ignorance. So far so Socratic. But then with the help of a further construction he works out the right answer, and so achieves true opinion, which it is suggested could be converted into knowledge if he were to go through the exercise often. The tacit implication is that if elenchus can reach a successful outcome in mathematics, it ought to be capable of it in ethics too.

None the less direct engagement with the original problem of what virtue is is abandoned, and the discussion turns to the issue of its teachability, and to the method of hypothesis. Here the idea is that instead of investigating the truth of proposition *p* directly 'you hit upon another proposition *h* ('the hypothesis'), such that *p* is true if and only if *h* is true, and then investigate the truth of *h*, undertaking to determine what would follow (quite apart from *p*) if *h* were true and, alternatively, if it were false' (Gregory Vlastos' formulation (1991)). After illustrating this procedure with an exceedingly obscure geometrical example, Socrates makes a lucid application of it to the ethical problem before them, and offers the Socratic thesis that virtue is knowledge as the hypothesis from which the teachability of virtue can be derived. The subsequent examination of this hypothesis comes to conclusions commentators have found frustratingly ambiguous. But the survival and development of the hypothetical method in *Phaedo* and *Republic* are enough to show Plato's conviction of its philosophical potential.

The slave boy episode is originally introduced by Socrates as a proof of something much more than the possibility of successful inquiry. The suggestion is that the best explanation of that possibility is provided by the doctrine of the immortality of the soul, a Pythagorean belief which makes the first of its many appearances in Plato's dialogues in *Meno* (see Psychē; Pythagoras §2; Pythagoreanism §3). More specifically, the idea as Socrates presents it is that the soul pre-exists the body, in a condition involving conscious possession of knowledge. On entry into the body it forgets what it knows, although it retains it as latent memory. Discovery of the sort of a priori knowledge characteristic of mathematics and (as Plato supposes) ethics is a matter of recollecting latent memory. This is just what happens to the slave boy: Socrates does not impart knowledge to him; he works it out for himself by recovering it from within. Once again, although the Socrates of *Meno* does not in the end subscribe to belief in learning as recollection of innate knowledge, it is embraced without equivocation in *Phaedo*, as also in the later *Phaedrus*. But *what* exactly is recollected? *Phaedo* will say: knowledge of Forms. *Meno* by contrast offers no clues. The introduction of the theory of Forms is reserved for *Symposium*.

12 *Symposium*

Symposium has the widest appeal of all Plato's writings. No work of ancient Greek prose fiction can match its compulsive readability. Plato moves through a rich variety of registers, from knockabout comedy and literary parody to passages of disturbing fantasy or visionary elevation, culminating in a multiply paradoxical declaration of love for Socrates put in the mouth of a drunken Alcibiades. Love (*erōs*) is the theme of the succession of *encōmia* or eulogies delivered at the drinking party (*symposion*) hosted by the playwright Agathon: not sublimated 'Platonic' love between the sexes, but the homoerotic passion of a mature man for a younger or indeed a teenager.

This continues until Aristophanes (one of the guests) and Socrates broaden and transform the discussion. Socrates' speech, which is a sort of anti-eulogy, develops a general theory of desire and its relation to beauty, and it is in this context that the idea of an eternal and changeless Form makes its first unequivocal appearance in Plato's *œuvre*. Thus Plato first declares himself a metaphysician not in a work devoted to philosophical argument, but in a highly rhetorical piece of writing, albeit one in which fashionable models of rhetoric are subverted.

Love and beauty are first connected in some of the earlier *encōmia*, and notably in Agathon's claim that among the gods 'Love is the happiest of them all, for he is the most beautiful and best' (195a). This thesis is subjected to elenchus by Socrates in the one argumentative section of the dialogue. Agathon is obliged to accept that love and desire are necessarily love and desire *for something*, namely, something they are in need of. Following his concession Socrates argues that beauty is not what love possesses but precisely the thing it is in need of. This argument constitutes the key move in the philosophy of the dialogue, which Plato elaborates in various ways through the medium of Diotima, the probably fictitious priestess from whom Socrates is made to claim he learned the art of love in which he has earlier (177d) claimed expertise. First she tells a myth representing Love as the offspring of poverty and resource, and so - according to her interpretation - occupying the dissatisfied intermediate position between ignorance and wisdom which characterizes philosophy: hence presumably the explanation of Socrates' claim to be an expert in love, since the pursuit of wisdom turns out to be the truest expression of love. Then she spells out the theoretical basis for this intellectualist construction of what love is. The theory has rightly been said to combine 'a psychology that is strictly or loosely Socratic with a metaphysics that is wholly Platonic' (Price 1995).

This psychology holds that a person who desires something wants not so much the beautiful as the good, or more precisely happiness conceived as permanent possession of the good. Love is a particular species of desire, which occurs when perception of beauty makes us want to reproduce. (Socrates is made to express bafflement at this point: presumably an authorial device for indicating that Diotima's line of thought is now moving beyond anything Plato considered strictly Socratic.) Diotima goes on to explain that reproduction is the way mortal animals pursue immortality, interpreted in its turn in terms of the longing for permanent possession of good with which she has just identified desire. Other animals and many humans are content with physical reproduction, but humans are capable of *mental* creation when inspired by a beautiful body, and still more by a beautiful soul or personality. This is how the activities of poets and legislators and the virtuous are to be understood.

Perhaps Plato thought these ideas, although no longer Socratic, provided a convincing explanation of the drive which powered Socrates' philosophical activity in general, and made him spend so much time with beautiful young men in particular. However that may be, in what follows he has Diotima speak of greater mysteries which 'I do not know whether you [that is, Socrates] would be able to approach'. These are the subject of a lyrical account of how a true lover moves step by step from preoccupation with the beauty of a single beloved, to appreciating that there is one and the same beauty in all bodies and so loving them all, and then to seeing and loving beauty in souls or personalities and all manner of mental creations, until he 'turns to the great sea of beauty, and gazing upon this gives birth to many gloriously beautiful ideas and theories, in unstinting love of wisdom [that is, philosophy]' (210d). The final moment of illumination arrives when the philosopher-lover grasps the Beautiful itself, an experience described as the fulfilment of all earlier exertions. Unlike other manifestations of beauty the Form of the Beautiful is something eternal, whose beauty is not qualified in place or time or relation or respect. It is just the one sort of thing it is, all on its own, whereas other things that are subject to change and decay are beautiful by participation in the Form. Only someone who has looked upon it will be capable of giving birth not to images of virtue (presumably the ideas and theories mentioned a little earlier), but to virtue itself, and so achieving immortality so far as any human can.

It is striking that the doctrine of the immortality of the soul forms no part of Diotima's argument. If we assume the scholarly consensus that *Symposium* postdates *Meno*, this poses something of a puzzle. One solution might be to suppose that, although *Meno* presents the doctrine, Plato is himself not yet wholly convinced of its truth, and so gives it no role in his account of the desire for immortality in *Symposium*. This solution might claim support from the fact that *Phaedo* takes upon itself the task of arguing the case for the immortality of the soul much more strenuously than in *Meno*, and in particular offers a much more careful and elaborate version of the argument from recollection. Additionally or alternatively, we may note that when Plato presents the doctrine of the immortality of the soul in the dialogues, he always treats it as something requiring explicit proof, unlike the theory of Forms,

which generally figures as a hypothesis recommending itself by its explanatory power or its ability to meet the requirements of Plato’s epistemology. Since Diotima’s discourse is not constructed as argument but as the explication of an idea, it is not the sort of context which would readily accommodate the kind of demonstration Plato apparently thought imperative for discussion of the immortality of the soul.

13 *Phaedo*

The departure point for *Phaedo*’s consideration of the fate of the soul after death is very close to that idea of love as desire for wisdom which Diotima offers at the start of her speech in *Symposium*. For Socrates starts with the pursuit of wisdom, which he claims is really a preparation for death. This is because it consists of an attempt to escape the restrictions of the body so far as is possible, and to purify the soul from preoccupation with the senses and physical desires so that it can think about truth, and in particular about the Forms, which are accessible not to sense perception but only to thought. Pure knowledge of anything would actually require complete freedom from the body. So given that death is the separation of soul from body, the wisdom philosophers desire will be attainable in full only when they are dead. Hence for a philosopher death is no evil to be feared, but something for which the whole of life has been a training. The unbearably powerful death scene at the end of the dialogue presents Socrates as someone whose serenity and cheerfulness at the end bear witness to the truth of this valuation.

Symposium implied that a long process of intellectual and emotional reorientation was required if someone was to achieve a grasp of the Form of Beauty. *Phaedo* has sometimes been thought to take a different view: interpreters may read its argument about recollecting Forms as concerned with the general activity of concept formation in which we all engage early in life. In fact the passage restricts recollection of Forms to philosophers, and suggests that the knowledge they recover is not the basic ability to deploy concepts (which Plato seems in this period to think a function of sense experience), but hard-won philosophical understanding of what it is to be beautiful or good or just. The interlocutors voice the fear that once Socrates is dead there will be nobody left in possession of that knowledge; and the claim that pure knowledge of Forms is possible only after death coheres with the *Symposium* account very well, implying as it does that the path to philosophical enlightenment is not just long but a journey which cannot be completed in this life.

The proposal that the soul continues to exist apart from the body after death is immediately challenged by Socrates’ interlocutors. Much of the rest of *Phaedo* is taken up with a sequence of arguments defending that proposal and the further contention that the soul is immortal, pre-existing the body and surviving its demise for ever. The longest and most ambitious of these arguments is the last of the set. It consists in an application of the method of hypothesis, which is explained again in a more elaborate version than that presented in *Meno*. The hypothesis chosen is the theory of Forms, or rather the idea that Forms function as explanations or causes of phenomena: beautiful things are beautiful by virtue of the Beautiful, large things large by virtue of the Large, and so on. Socrates is made to represent his reliance on this apparently uninformative or ’safe and simple’ notion of causation as a position he has arrived at only after earlier intellectual disappointments: first with the inadequacies of Presocratic material causes, then with the failure of Anaxagoras’ promise of a teleological explanation of why things are as they are (see Anaxagoras §4).

He soon goes on to argue however that the hypothesis can be used to generate a more sophisticated model of causation. Instead of proposing merely that (for example) hot things are hot by virtue of the Hot, we may legitimately venture the more specific explanation: ‘Hot things are hot by virtue of fire’, provided that it is true that wherever fire exists, it always heats things in its vicinity, being itself hot and never cold. After elaborating this point Socrates is ready to apply the model to the case of life and soul. By parity of reasoning, we may assert that living things are alive not just in virtue of life, but in virtue of soul, given that wherever soul exists it makes things it occupies alive, being itself alive and never dead. From this assertion there appears to follow the conclusion whose derivation is the object of the exercise: if soul is always alive and never dead, it must be immortal (that is, incapable of death) and so imperishable.

Phaedo, like *Republic*, ends with a sombre myth of last judgment and reincarnation, designed primarily to drive home the moral implications of Plato’s distinctive version of soul-body dualism. It reminds us of the Pythagorean origins of the doctrine of the immortality of the soul. Yet the Platonism of *Phaedo* owes a great deal also to the metaphysics of Parmenides. Both here and in *Symposium* the characterization of Forms as simple eternal beings,

accessible only to thought, not the senses, and the contrast both dialogues make with the changing and contradictory world of phenomena, are couched in terms borrowed from Parmenides and the Eleatic tradition which he inaugurated. Platonism can accordingly be seen as the product of an attempt to understand a fundamentally Socratic conception of philosophy and the philosophical life in the light of reflection on these two powerful Presocratic traditions of thought, using the new methodological resources made available by geometry.

14 *Republic*

Republic is misleadingly titled. The Greek name of the dialogue is *Politeia*, which is the standard word for constitution or ordering of the political structure: ’political order’ would give a better sense of what Plato has in mind. There is a further and deeper complication. Once you start reading the dialogue you find that it is primarily an inquiry into justice, conceived as a virtue or moral excellence of individual persons. The philosophical task it undertakes is the project of showing that justice so conceived is in the best interests of the just person, even if it brings nothing ordinarily recognizable as happiness or success, or indeed (as with the sentence of death passed on Socrates) quite the opposite. Thus *Republic* carries forward the thinking about justice begun in early dialogues such as *Apology*, *Crito* and *Gorgias*. Why, then, the title’s suggestion that it is a work of political rather than moral philosophy?

One way of answering this question is to attend to the formal structure of *Republic*. After book I, an inconclusive Socratic dialogue which none the less introduces, particularly in the conversation with Thrasymachus, many of the themes pursued in the rest of the work, the interlocutors agree to take an indirect approach to the problem of individual justice: they will consider the nature of justice and injustice in the *polis*, that is the (city-)state, in the hope that it will provide an illuminating analogy. Books II-IV spell out the class structure required in a ’good city’. It is suggested that in such a state political justice consists in the social harmony achieved when each class (economic, military, governing) performs its own and only its own function. This model is then applied to the individual soul (see Psychē). Justice and happiness for an individual are secured when each of the parts of the soul (appetite, emotion, reason) performs the role it should in mutual harmony. In working out the idea of psychic harmony, Plato formulates a conception of the complexity of psychological motivation, and of the structure of mental conflict, which leaves the simplicities of Socratic intellectualism far behind, and one which has reminded interpreters of Freudian theory, particularly in books VIII-IX. Here he examines different forms of *un*just political order (notably oligarchy, democracy and tyranny) and corresponding conditions of order, or rather increasing *dis*order, in the soul.

Political theory therefore plays a large part in the argument of the dialogue, even though the ultimate focus is the moral health of the soul, as is confirmed by the conclusion of book IX. Socrates suggests that it may not matter whether we can actually establish a truly just political order, provided we use the idea of it as a paradigm for founding a just city within our own selves.

This account of *Republic* omits the central books V-VII. These explore the notion of political order much further than is necessary for the purposes of inquiry into individual justice. This is where Plato develops the notion of a communistic governing class, involving the recruitment of talented women as well as men, the abolition of the family, and institution of a centrally controlled eugenic breeding programme. And it is where, in order to meet the problem of how the idea of the just city he has been elaborating might ever be put into practice, he has Socrates introduce philosopher-rulers:

> Unless either philosophers rule in our cities or those whom we now call rulers and potentates engage genuinely and adequately in philosophy, and political power and philosophy coincide, there is no end, my dear Glaucon, to troubles for our cities, nor I think for the human race.
>
> (*Republic* 473c-d)

What Plato perhaps has most in mind when he makes Socrates speak of ’troubles’ is as well as civil war the corruption he sees in all existing societies. As he acknowledges, this makes the emergence of an upright philosopher-ruler an improbability - and incidentally leaves highly questionable the prospects of anyone but a Socrates developing moral order within the soul when society without is infected with moral disorder.

Here we touch on another broadly political preoccupation of *Republic*, worked out at various places in the

dialogue. It offers among other things a radical critique of Greek cultural norms. This is highlighted in the censorship of Homer proposed in books II and III, and in the onslaught on the poets, particularly the dramatists, in book X, and in their expulsion from the ideal city. But these are only the more memorable episodes in a systematic attack on Greek beliefs about gods, heroes and the departed, on the ethical assumptions underlying music, dance and gymnastics (see Mimēsis), and again erotic courtship, and on medical and judicial practice. *Republic* substitutes its own austere state educational programme, initially focused on the training of the emotions, but subsequently (in books VI and VII) on mathematics and philosophy. Plato sees no hope for society or the human race without a wholesale reorientation, fostered by an absolute political authority, of all the ideals on which we set our hearts and minds.

Republic itself is written in such a way as to require the reader to be continually broadening perspectives on the huge range of concerns it embraces, from the banalities of its opening conversation between Socrates and the aged Cephalus to its Platonist explication of the very notion of philosophy in the epistemology and metaphysics of books V-VII. At the apex of the whole work Plato sets his presentation of the Form of the Good, as the ultimate goal of the understanding that philosophy pursues by use of the hypothetical method. The dialogue offers a symbol of its own progress in the potent symbol of the cave. We are like prisoners chained underground, who can see only shadows of images flickering on the wall. What we need is release from our mental shackles, and a conversion which will enable us gradually to clamber out into the world above and the sunlight. For then, by a sequence of painful reorientations, we may be able to grasp the Good and understand how it explains all that there is.

15 Critical dialogues

Parmenides is that rare phenomenon in philosophy: a self-critique. Plato here makes his own theory of Forms the subject of a penetrating scrutiny which today continues to command admiration for its ingenuity and insight. *Theaetetus* (datable to soon after 369 BC) also reverts to Plato’s critical manner. It applies an enriched variant of the Socratic elenchus to a sequence of attempts to define knowledge. The confidence of *Phaedo* and *Republic* that Platonist philosophers are in possession of knowledge and can articulate what it consists in is nowhere in evidence, except in a rhetorical digression from the main argument. Methodological preoccupations are dominant in both works. *Parmenides* suggests that to defend the Forms against its critique, one would need to be much more practised in argument than is their proponent in this dialogue (a young Socrates fictively encountering a 65-year old Parmenides and a middle-aged Zeno). And it sets out a specimen of the sort of exercise required, running to many pages of purely abstract reasoning modelled partly on the paradoxes of Zeno of Elea, partly on Parmenides’ deductions in the Way of Truth (see Parmenides §§3-8). *Theaetetus* likewise presents itself, initially more or less explicitly, later implicitly, as a model of how to go about testing a theory without sophistry and with due sympathy. While the conclusions achieved by this ’midwifery’ - as Socrates here calls it - are as devastatingly negative as in the early dialogues, we learn much more philosophy along the way. Many readers find *Theaetetus* the most consistently rewarding of all the dialogues.

A sketch of the principal concerns of the two dialogues will bring out their radical character. *Parmenides* raises two main questions about Forms. First, are there Forms corresponding to every kind of predicate? Not just *one* and *large*, or *beautiful* and *just*, familiar from the middle period dialogues, but *man* and *fire*, or even *hair* and *dirt*? Socrates is represented as unclear about the issue. Second, the idea that other things we call for example ’large’ or ’just’ are related to the Form in question by participation is examined in a succession of arguments which seek to show that, however Forms or the participation relation are construed, logical absurdities of one kind or another result. The most intriguing of these has been known since Aristotle as the Third Man: if large things are large in virtue of something distinct from them, namely the Form of Large, then the Large itself and the other large things will be large in virtue of *another* Form of Large - and so *ad infinitum*.

Theaetetus devotes much of its space to considering the proposal that knowledge is nothing but sense perception, or rather to developing and examining two theories with which that proposal is taken to be equivalent: the view of Protagoras (§3) that truth is relative, since ’man is the measure of all things’, and that of Heraclitus that everything is in flux, here considered primarily in application to the nature of sense perception. The dialogue is home to some of Plato’s most memorable arguments and analogies. For example, Protagoreanism is attacked by the brilliant (although perhaps flawed) self-refutation argument: if man is the measure of *all* things, then the doctrine of the relativity of truth is itself true only in so far as it is believed to be true; but since people in general believe it to be

false, it must be false. The next section of *Theaetetus* worries about the coherence of the concept of false belief. Here the soul is compared to a wax tablet, with false belief construed as a mismatch between current perceptions and those inscribed on the tablet, or again to an aviary, where false belief is an unsuccessful attempt to catch the right bird (that is, piece of knowledge). In the final section the interlocutors explore the suggestion that knowledge must involve the sort of complexity that can be expressed in a *logos* or statement. Socrates' 'dream' that such knowledge must be built out of unknowable simples fascinated Wittgenstein (§5), who saw in it an anticipation of the theory of his *Tractatus*.

Are we to infer that in opening or reopening questions of this kind Plato indicates that he is himself in a real quandary about knowledge and the Forms? Or is his main target philosophical complacency in his readers, as needing to be reminded that no position is worth much if it cannot be defended in strenuous argument? Certainly in the other two dialogues grouped here with *Parmenides* and *Theaetetus* the theory of Forms is again in evidence, presented as a view the author is commending to the reader's intellectual sympathies. *Cratylus* is a work whose closest philosophical connections are with *Theaetetus*, although its relative date among the dialogues is disputed. It is a pioneering debate between rival theories of what makes a word for a thing the right word for it: convention, or as Cratylus holds, a natural appropriateness - sound somehow mirroring essence (see Language, ancient philosophy of §2). Underlying Cratylus' position is an obscurely motivated commitment to the truth of Heracliteanism (see Cratylus). For present purposes what is of interest is the final page of the dialogue, which takes the theory of Forms as premise for an argument showing that the idea of an absolutely universal Heraclitean flux is unsustainable. As for *Phaedrus*, it contains one of the most elevated passages of prose about the Forms that Plato ever wrote.

The context is an exemplary rhetorical exercise in which *Symposium*'s treatment of the philosophical lover's attraction to beauty is reworked in the light of Republic's tripartition of the soul. Subsequently Plato has Socrates dismiss the speech as 'play', useful only for the methodological morals about rhetorical procedure we happen to be able to derive from it - together with a preceding denunciation of love by Socrates, capping one by his interlocutor Phaedrus - if we are dialecticians. This comment has led some readers to conjecture that Phaedrus accordingly marks Plato's formal leave-taking of the theory of Forms: in retrospect he sees it more as rhetoric than as philosophy or dialectic, which will henceforward confine itself to something apparently less inspiring - the patient, thorough, comprehensive study of similarities and differences. Yet *Phaedrus* is pre-eminently a dialogue written not to disclose its author's mind, but to make demands on the sophisticated reader's. Perhaps Socrates' great speech on the philosophical lover is 'play' not absolutely, but only relative to the controlling and unifying preoccupation of the dialogue, which is to work through a fresh examination of rhetoric, going beyond Gorgias in explaining how it can be a genuine form of expertise, based on knowledge of truth and variously geared to the various psychological types to which oratory addresses itself. We might speculate that Plato writes the speech as he does precisely because he thinks or hopes many of his readers will be of a type persuadable to the philosophical life by its vision of the soul's desire for the Beautiful.

16 Later dialogues

The theory of Forms also figures prominently in *Timaeus*. *Timaeus* is Plato's one venture into physical theory, and appropriately has in the Italian Greek Timaeus someone other than Socrates as main speaker. It is presented as an introduction to the story of Atlantis, allegedly an island power defeated by the prehistoric Athenians, and mentioned only by Plato among classical Greek authors. The conflict between Atlantis and Athens was to be the subject of *Critias*, conceived as a dialogue that would demonstrate the political philosophy of *Republic* in practice. But *Critias* was never completed, so *Timaeus* stands as an independent work.

The argument of *Timaeus* is based on the premise that the universe is not eternal but created - although debate has raged from antiquity onwards whether this means created in time, or timelessly dependent on a first cause. From the order and beauty of the universe Plato infers a good creator or craftsman (*dēmiourgos*), working on pre-existing materials (with their own random but necessary motions) from an eternal blueprint encoding life and intelligence: namely, the Form of Animal. The greater part of *Timaeus* consists in an account of how first the universe (conceived of as a living creature), then humans are designed from the blueprint for the best. Much use is made of mathematical models, for example for the movements of the heavenly bodies and the atomistic construction of the four elements. The account is presented as inevitably only a 'likely story', incapable of the

irrefutable truths of metaphysics.

There is no more austere or profound work of metaphysics in Plato’s *œuvre* than *Sophist*. Like many of the post-*Republic* dialogues it is ’professional’ philosophy, probably written primarily for Plato’s students and associates in the Academy. The style of *Sophist* and the remaining works to be discussed is syntactically tortuous and overloaded with abstraction and periphrasis; they are altogether lacking in literary graces or dramatic properties which might commend them to a wider readership. *Sophist*’s main speaker is a stranger from Elea, symbolizing the Parmenidean provenance of the problem at the heart of the long central section of the dialogue: how is it possible to speak of what is not (see Parmenides §2)? This puzzle is applied for example both to the unreality of images and to falsehood, understood as what is not the case. The solution Plato offers required some revolutionary moves in philosophical logic, such as the explicit differentiation of identity from predication, and the idea that subject and predicate play different roles in the syntax of the sentence. These innovations and their bearing on analysis of the verb ’to be’ have made *Sophist* the subject of some of the most challenging writing on Plato in the twentieth century.

The companion dialogue *Politicus* or *Statesman* addresses more squarely than *Republic* did the practical as distinct from the theoretical knowledge of the ideal statesman. Its contribution to this topic consists of three major claims. First is the rejection of the sovereignty of law. Plato has nothing against law as a convenient but imprecise rule of thumb in the hands of an expert statesman, provided it does not prevent him using his expertise. Making law sovereign, on the other hand, would be like preferring strict adherence to a handbook of navigation or a medical textbook to the judgment of the expert seafarer or doctor. If you have no such expert available, a constitution based on adherence to law is better than lawlessness, but that is not saying much. What law cannot do that expert rulers can and must is judge the *kairos*: discern the right and the wrong ’moment’ to undertake a great enterprise of state. This proposition follows from the second of Plato’s key claims, which is represented as one true of all practical arts: real expertise consists not of measuring larger and smaller, but in determining the norm between excess and defect - a notion which we ordinarily think more Aristotelian than Platonic (see Aristotle §22), although it recurs in a different guise in *Philebus*. Finally, Plato thinks we shall only get our thinking straight on this as on any matter if we find the right - usually homely - model. *Statesman* makes the statesman a sort of weaver. There are two strands to the analogy. First, like weaving statesmanship calls upon many subordinate skills. Its job is not to be doing things itself, but to control all the subordinate functions of government, and by its concern for the laws and every other aspect of the city weave all together. Second, the opposing temperaments of the citizens are what most need weaving together if civil strife is to be avoided, and (as in *Republic*) expert rulers will use education and eugenics to that end.

Statesman shares themes with both *Philebus* and *Laws*. *Philebus* is the one late dialogue in which Socrates is principal speaker, as befits its ethical topic: the question whether pleasure or understanding is the good, or at least the more important ingredient in the good life. After so much insistence in middle-period dialogues on the Form as a unity distinct from the plurality of the phenomena, it comes as a shock to find Socrates stressing at the outset that there is no merit in reiterating that pleasure or understanding is a unity. The skill resides in being able to determine what and how many forms of understanding and pleasure there are. What *Philebus* goes on to offer next is a model for thinking about how any complex structure is produced, whether a piece of music or the universe itself. It requires an intelligent cause creating a mixture by imposing limit and proportion on something indeterminate. This requirement already indicates the main lines of the answer to our problem, at any rate, if it is accepted that pleasure is intrinsically indeterminate. Clearly intelligence and understanding will be shaping forces in the good life, but pleasures are only admissible if suitably controlled. At the adjudication at the end of the dialogue, this is just the result we get. The majority of the many forms of pleasure defined and examined in the course of the dialogue are rejected. They do not satisfy the criteria of measure and proportion which are the marks of the good.

17 *Laws*

The vast *Laws* is in its way the most extraordinary of all Plato’s later writings, not for its inspiration (which flags) but for its evidence of tireless fascination with things political. Its relation to *Republic* and *Statesman* has been much debated. What is clear is that Plato is legislating - through the last eight of its twelve long books - for a second best to the ideal state and ideal statesman of *Republic*, with greater zeal than *Statesman* might have led one to expect. Is this because he has lost faith in those ideals, which still seemed alive in *Statesman* at least as ideals?

That view is in danger of overlooking *Republic*'s own indication that it would be wrong to expect in practice anything but an approximation of the ideal.

Philosophers do not often read *Laws*. But book X presents Plato's natural theology, as the background to laws dealing with atheists. And perhaps the most interesting proposal in the dialogue concerns the very idea of legislation. It is the notion of a 'prelude' to a law, which is the attempt the legislator should make to *persuade* citizens of the necessity of the prescriptions of the law itself. Here is a theme which relates interestingly to conceptions of reason, necessity and persuasion found in several other dialogues, notably *Republic* and *Timaeus*.

18 Plato's influence

Plato's influence pervades much of subsequent Western literature and thought. Aristotle was among those who came to listen to him in the 'school' he founded in the Academy; and a great deal of Aristotle's work is conceived in explicit or implicit response to Plato. Other philosophical traditions flourished after Aristotle's time in the last centuries BC, and the Academy of the period read Plato through sceptical spectacles (see Arcesilaus). But from the first century AD onwards Platonism in various forms, often syncretistic, became the dominant philosophy of the Roman Empire (see Platonism, Early and Middle), especially with the rise of Neoplatonism in late antiquity (see Neoplatonism). Some of the Fathers of the early Greek Church articulated their theologies in Platonist terms; and through Augustine in particular Plato shaped, for example, the Western Church's conception of time and eternity (see Patristic philosophy). A Neoplatonist version of him prevailed among the Arabs (see Platonism in Islamic philosophy).

With the translation of Plato into Latin in the high Middle Ages (see Platonism, medieval) and the revival of Greek studies in the Renaissance, Platonism (again in a Neoplatonic guise) once more gripped the minds of learned thinkers in the West, for example at the Medici court in fifteenth century Florence (see Platonism, Renaissance). But none of the great philosophers of the modern era has been a Platonist, even if Plato was an important presence in the thought of a Leibniz or a Hegel or a Russell. Probably he has never been studied more intensively than in the late twentieth century. Thanks to the availability of cheap translations in every major language and to his position as the first great philosopher in the Western canon, he figures in most introductory courses offered every year to tens of thousands of students throughout the developed world.

MALCOLM SCHOFIELD

List of works

Plato (390s-347 BC) *Platonis Opera*, ed. J. Burnet, Oxford: Clarendon Press, 1900-7, 5 vols.(The classic Oxford Classical Texts edition of the Greek text of all Plato's works, including those frequently regarded as spurious or dubious; follows the Thrasyllan organization of the corpus adopted in the manuscript tradition.)

Plato (390s-347 BC) *Platonis Opera*, vol. 1, *Euthyphro, Apologia Socratis, Crito, Phaedo, Cratylus, Theaetetus, Sophistes, Politicus*, ed. E.A. Duke, W.F. Hicken, W.S.M. Nicoll, D.B. Robinson and J.C.G. Strachan, Oxford: Clarendon Press, 1995.(The first volume of a new Oxford Classical Text, designed to replace Burnet's 1900-7 edition.)

Plato (390s-347 BC) *The Collected Dialogues of Plato including the Letters*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1961.(Useful one-volume collection of translations by various hands; excludes a few of the dialogues generally regarded as spurious.)

Plato (390s-347 BC) *Plato: Complete Works*, ed. J.M. Cooper, Indianapolis, IN: Hackett, 1997.(The most complete one-volume collection, with translations by various hands; includes *Spuria* and *Dubia*, introductions and bibliography; many dialogues are also available separately in paperback editions.) Any chronological order for Plato's writings is speculative. That given below reflects the view of Plato taken in this entry. The list includes all the works generally agreed to be authentic, and one or two that may be inauthentic.

Plato (*c*.395-387 BC) *Ion*, trans. and ed. T.J. Saunders, in *Plato: Early Socratic Dialogues*, Harmondsworth: Penguin, 1987.(Includes bibliography.)

Plato (*c*.395-387 BC) *Hippias Minor*, trans. R. Waterfield, in T.J. Saunders (ed.) *Plato: Early Socratic Dialogues*, Harmondsworth: Penguin, 1987.(Includes bibliography.)

Plato (*c*.395-387 BC) *Crito*, ed. J. Burnet, in *Plato's Euthyphro, Apology of Socrates, and Crito*, Oxford: Clarendon Press, 1924; trans. R.E. Allen, *Socrates and Legal Obligation*, Minneapolis, MN: University of Minnesota Press, 1970.(The former contains notes on the Greek text, the latter notes and essays.)

Plato (*c*.395-387 BC) *Apology*, ed. J. Burnet, in *Plato's Euthyphro, Apology of Socrates, and Crito*, Oxford:

Clarendon Press, 1924; trans. R.E. Allen, *Socrates and Legal Obligation*, Minneapolis, MN: University of Minnesota Press, 1970.(The former contains notes on the Greek text, the latter notes and essays.)

Plato (*c*.395-387 BC) *Gorgias*, ed. E.R. Dodds, *Plato Gorgias*, Oxford: Clarendon Press, 1959; trans. T.H. Irwin, *Plato: Gorgias*, Oxford: Clarendon Press, 2nd edn, 1995; trans. R. Waterfield, *Plato Gorgias*, Oxford and New York: Oxford University Press, 1994.(A major work of scholarship, Dodds includes introduction, summaries and full commentary on the Greek text; Irwin and Waterfield include bibliographies.)

Plato (*c*.386-380 BC) *Menexenus*, trans. and ed. R.G. Bury, in *Plato: Timaeus, Critias, Cleitophon, Menexenus, Epistles*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1929. (Greek text with facing English translation.)

Plato (*c*.386-380 BC) *Euthyphro*, ed. J. Burnet, in *Plato's Euthyphro, Apology of Socrates, and Crito*, Oxford: Clarendon Press, 1924; trans. R.E. Allen, *Plato's 'Euthyphro' and the Earlier Theory of Forms*, London: Routledge & Kegan Paul, 1970.(The former contains notes on the Greek text, the latter commentary and essays.)

Plato (*c*.386-380 BC) *Laches*, trans. I. Lane, in T.J. Saunders (ed.) *Plato: Early Socratic Dialogues*, Harmondsworth: Penguin, 1987.(Includes bibliography.)

Plato (*c*.386-380 BC) *Charmides*, trans. D. Watt, in T.J. Saunders (ed.) *Plato: Early Socratic Dialogues*, Harmondsworth: Penguin, 1987.(Includes bibliography.)

Plato (*c*.386-380 BC) *Protagoras*, trans. B. Jowett, revised M. Ostwald, *Plato Protagoras*, Indianapolis, IN: Bobbs-Merrill, 1956; trans. C.C.W. Taylor, *Plato: Protagoras*, Oxford: Clarendon Press, 2nd edn, 1991. (Ostwald includes a classic essay by G. Vlastos as introduction; Taylor includes notes and bibliography.)

Plato (*c*.386-380 BC) *Meno*, ed. R.S. Bluck, *Plato's Meno*, Cambridge: Cambridge University Press; ed. and trans. R.W. Sharples, *Plato: Meno*, Warminster: Aris & Phillips, 1985; trans. J.M. Day, *Plato's Meno in Focus*, London and New York: Routledge, 1994.(Bluck includes introduction and commentary on the Greek text; Sharples includes notes and bibliography; Day has both introduction and bibliography as well as essays by various hands.)

Plato (*c*.386-380 BC) *Symposium*, ed. R.G. Bury, *The Symposium of Plato*, Cambridge: Cambridge University Press, 2nd edn, 1932; ed. K.J. Dover, *Plato: Symposium*, Cambridge: Cambridge University Press, 1980; trans. A. Nehamas and P. Woodruff, *Plato: Symposium*, Indianapolis, IN: Hackett, 1989; trans. R. Waterfield, *Plato Symposium*, Oxford and New York: Oxford University Press, 1994. (Both Bury and Dover include introductions and notes on the Greek text; also Nehamas and Woodruff, and Waterfield, include introduction and bibliography.)

Plato (*c*.386-380 BC) *Phaedo*, ed. C.J. Rowe, *Plato: Phaedo*, Cambridge: Cambridge University Press, 1995; trans. R. Hackforth, *Plato's Phaedo*, Cambridge: Cambridge University Press, 1955; trans. D. Gallop, *Plato Phaedo*, Oxford and New York: Oxford University Press, 1993. (Rowe includes introduction, notes on the Greek text and bibliography; Hackforth offers commentary; Gallop includes notes and bibliography.)

Plato (*c*.380-367 BC) *Hippias Major*, ed. D. Tarrant, *The Hippias Major attributed to Plato*, Cambridge: Cambridge University Press, 1928; trans. P. Woodruff, *Plato: Hippias Major*, Oxford: Blackwell, 1982.(The former offers Greek text with introduction and commentary; the latter includes introduction, essays and bibliography.)

Plato (*c*.380-367 BC) *Republic*, ed. J. Adam, revised. D.A. Rees, *The Republic of Plato*, Cambridge: Cambridge University Press, 1963, 2 vols; trans. P. Shorey, *Plato: Republic*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1930, 2 vols; trans. A.D. Lindsay, revised T.H. Irwin, *Plato: Republic*, London: Dent, 1992.(Adam includes Greek text with notes; the Loeb edition has Greek text with facing English version; Lindsay includes introduction and bibliography.)

Plato (*c*.380-367 BC) *Cratylus*, ed. L. Méridier, *Platon: Cratyle*, Budé series, Paris: Les Belles Lettres, 3rd edn, 1961; trans. H.N. Fowler, in *Plato: Cratylus, Parmenides, Greater Hippias, Lesser Hippias*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1926.(Méridier offers Greek text with facing French translation, introduction and notes; Fowler has Greek text with facing English translation.)

Plato (*c*.380-367 BC) *Euthydemus*, ed. E.H. Gifford, *The Euthydemus of Plato*, Oxford: Clarendon Press, 1905; trans. R.K. Sprague, *Plato, Euthydemus*, Indianapolis, IN: Bobbs-Merrill, 1985; trans. R. Waterfield, in T.J. Saunders (ed.) *Plato: Early Socratic Dialogues*, Harmondsworth: Penguin, 1987.(Gifford includes Greek text with notes; Sprague and Waterfield both include bibliographies.)

Plato (*c*.380-367 BC) *Lysis*, trans. D. Bolotin, *Plato's Dialogue on Friendship*, Ithaca, NY: Cornell University Press, 1979; trans. D. Watt, in T.J. Saunders (ed.) *Plato: Early Socratic Dialogues*, Harmondsworth: Penguin, 1987.(Both include bibliographies.)

Plato (*c*.380-367 BC) *Parmenides*, trans. F.M. Cornford, *Plato and Parmenides*, London: Routledge & Kegan Paul,

1939; trans. M.L. Gill and P. Ryan, *Plato: Parmenides*, Indianapolis, IN: Hackett, 1996.(Cornford supplies a running commentary; Gill and Ryan include substantial introductory essay and bibliography.)

Plato (*c*.380-367 BC) *Theaetetus*, ed. L. Campbell, *The Theaetetus of Plato*, Oxford: Clarendon Press, 1883; trans. J.H. McDowell, *Plato: Theaetetus*, Oxford: Clarendon Press, 1973; trans. M.J. Levett, revised M. Burnyeat, *The Theaetetus of Plato*, Indianapolis, IN: Hackett, 1990.(Campbell includes Greek text and notes; Burnyeat includes bibliography and book-length introductory essay of classic status.)

Plato (*c*.366-360 BC) *Phaedrus*, ed. and trans. C.J. Rowe, *Plato: Phaedrus*, Warminster: Aris & Phillips, 1986; trans. A. Nehamas and P. Woodruff, *Plato: Phaedrus*, Indianapolis, IN: Hackett, 1995.(Both includes bibliographies; the latter also includes an introduction.)

Plato (*c*.366-360 BC) *Timaeus*, ed. A. Rivaud, *Platon: Timée, Critias*, Budé series, Paris: Les Belles Lettres, 1925; trans. F.M. Cornford, *Plato's Cosmology*, London: Routledge & Kegan Paul, 1937.(The former has Greek text with facing French translation, introduction and notes; the latter includes commentary of classic status.)

Plato (*c*.366-360 BC) *Critias*, ed. C. Gill, *Plato: The Atlantis Story*, Bristol: Bristol Classical Press, 1980; trans. H.D.P. Lee, *Plato: Timaeus and Critias*, Harmondsworth: Penguin, 1971.(Gill includes introduction and commentary.)

Plato (*c*.366-360 BC) *Sophist*, ed. L. Campbell, *The Sophistes and Politicus of Plato*, Oxford: Clarendon, 1867; trans. F.M. Cornford, *Plato's Theory of Knowledge*, London: Routledge & Kegan Paul, 1935; trans. N.P. White, *Plato: The Sophist*, Indianapolis, IN: Hackett, 1993.(Campbell's is the classic edition, especially for the introduction on Plato's late style, and includes notes on the Greek text; White includes introduction and bibliography, Cornford a running commentary.)

Plato (*c*.366-360 BC) *Statesman (Politicus)*, ed. L. Campbell, *The Sophistes and Politicus of Plato*, Oxford: Clarendon Press, 1867; trans. J.B. Skemp, *Plato: The Statesman*, London: Routledge & Kegan Paul, 1952; ed. and trans. C.J. Rowe, *Plato: Statesman*, Warminster: Aris & Phillips, 1995.(Campbell's is the classic edition, especially for the introduction on Plato's late style; Skemp offers a substantial introduction; Rowe includes bibliography and notes.)

Plato (*c*.360-347 BC) *Philebus*, ed. R.G. Bury, *The Philebus of Plato*, Cambridge: Cambridge University Press, 1897; trans. R. Hackforth, *Plato's Examination of Pleasure*, Cambridge: Cambridge University Press, 1945; trans. J.C.B. Gosling, *Plato: Philebus*, Oxford: Clarendon Press, 1975; trans. D. Frede, *Plato: Philebus*, Indianapolis, IN: Hackett, 1993.(Bury offers notes on the Greek text, Hackforth a commentary, Gosling and Frede include substantial introductions and bibliography.)

Plato (*c*.360-347 BC) *Seventh Letter*, ed. R.S. Bluck, *Plato's Seventh and Eighth Letters*, Cambridge: Cambridge University Press, 1947; trans. W. Hamilton, *Plato: Phaedrus and Letters VII and VIII*, Harmondsworth: Penguin, 1973.(Bluck includes Greek text with notes.)

Plato (*c*.360-347 BC) *Laws*, ed. E.B. England, *The Laws of Plato*, Manchester: Manchester University Press, 1921, 2 vols; trans. T.J. Saunders, *Plato: Laws*, Harmondsworth: Penguin, 1970. (England includes Greek text with notes; Saunders includes introduction, summaries and bibliography.)

References and further reading

Allen, R.E. (1965) *Studies in Plato's Metaphysics*, London: Routledge & Kegan Paul.(A collection of mostly seminal essays by various hands.)

Aristoxenus (late 4th century BC) *Harmonics*, trans H.S. Macran, *The Harmonics of Aristoxenus*, Oxford: Clarendon Press, 1902.(Contains Greek text and English translation, with introduction and notes.)

Brandwood, L. (1990) *The Chronology of Plato's Dialogues*, Cambridge: Cambridge University Press.(A sober critical history of the stylometric study of Plato, including assessments of the work of Campbell and Ritter.)

Burnyeat, M.F. (1987) 'Platonism and mathematics: a prelude to discussion', in A. Graeser (ed.) *Mathematics and Metaphysics in Aristotle*, Bern and Stuttgart: Paul Haupt Verlag, 213-40.(A challenging re-evaluation of Plato's treatment of the epistemological and ontological status of mathematics, including a new approach to the 'unwritten doctrines'.)

Campbell, L. (1867) *The Sophistes and Politicus of Plato*, Oxford: Clarendon Press.(Includes a pioneering introductory essay on Plato's late style.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius: Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(Greek text with facing English translation; book III contains Diogenes' life of Plato.)

Friedländer, P. (1958, 1964, 1969) *Plato*, trans. H. Meyerhoff, London: Routledge & Kegan Paul, 3 vols.(An account of the dialogues particularly recommended for its treatment of the philosophical significance of their

literary characteristics.)
Gaiser, K. (1968) *Platon’s ungeschriebene Lehre (Plato’s Unwritten Teaching)*, Stuttgart: Ernst Klett, 2nd edn. (An authoritative statement of the Tübingen interpretation of Plato.)
Gaiser, K. (1980) ‘Plato’s enigmatic Lecture "On The Good"’, *Phronesis* 25: 5-37.(A lucid restatement in English of the Tübingen interpretation.)
Grote, G. (1867) *Plato and the Other Companions of Sokrates*, London: John Murray, 2nd edn, 3 vols.(An unrivalled account of the dialogues by the greatest nineteenth-century Plato scholar.)
Grube, G. (1980) *Plato’s Thought*, London: Athlone Press.(Accessible introductory account of Plato’s thought, with new introduction and bibliography by D. Zeyl.)
Guthrie, W.K.C. (1975, 1978) *A History of Greek Philosophy*, vol. 4, *Plato, The Man and his Dialogues: Earlier Period*, vol. 5, *The Later Plato and the Academy*, Cambridge: Cambridge University Press.(The Plato volumes of the most detailed and comprehensive English-language account of Greek philosophy; indispensable for bibliography and general orientation; volume 5 includes an assessment of the Tübingen school’s interpretation of Plato.)
Irwin, T.H. (1995) *Plato’s Ethics*, Oxford: Oxford University Press.(A major philosophical study.)
Kahn, C.H. (1996) *Plato and The Socratic Dialogue*, Cambridge: Cambridge University Press.(An important study questioning developmental assumptions in standard accounts of the chronology of the dialogues.)
Kraut, R. (1992) *The Cambridge Companion to Plato*, Cambridge: Cambridge University Press.(Useful collection of essays by different authors on many aspects of Plato’s work; includes extensive bibliography.)
Owen, G.E.L. (1953) ’The Place of the *Timaeus* in Plato’s Dialogues’, *Classical Quarterly* 3: 79-95; repr. in R.E. Allen, *Studies in Plato’s Metaphysics*, London: Routledge & Kegan Paul, 1965, 313-38; repr. in G.E.L. Owen, *Logic, Science and Dialectic*, London: Duckworth, 1986, 65-84.(Controversial attempt to interpret *Timaeus* as a middle-period dialogue.)
Price, A.W. (1995) *Mental Conflict*, London: Routledge.(An exploration of the theories of mind devised by Greek philosophers to account for psychological conflict.)
Popper, K.R. (1966) *The Open Society and its Enemies*, vol. 1, *The Spell of Plato*, London: Routledge & Kegan Paul, 5th edn.(The most influential work on Plato in the twentieth century; attacks Plato’s political thought as totalitarian.)
Ritter, C. (1888) *Untersuchungen über Platon: die Echtheit und Chronologie der Platonischen Schriften (Researches into Plato: The Authenticity and Chronology of the Platonic Writings)*, Stuttgart: Kohlhammer. (The best example of a sustained use of stylistic criteria to determine questions of chronology and authenticity.)
Ross, W.D. (1951) *Plato’s Theory of Ideas*, Oxford: Clarendon Press.(Particularly valuable for its review of the evidence for Plato’s ‘unwritten doctrines’.)
Scott, D. (1995) *Recollection and Experience: Plato’s Theory of Learning and its Successors*, Cambridge: Cambridge University Press.(Important for its exegesis of Platonic ‘recollection’ and of the philosophical tradition it inaugurated.)
Thesleff, H. (1982) *Studies in Platonic Chronology*, Helsinki: Societas Scientiarum Fennica.(Speculations premised on the assumption that many dialogues are revisions of earlier versions.)
Vlastos, G. (1981) *Platonic Studies*, Princeton, NJ: Princeton University Press, 2nd edn. (Penetrating essays by a leading scholar.)
Vlastos, G. (1991) *Socrates: Ironist and Moral Philosopher*, Cambridge: Cambridge University Press.(A study of the Socrates of the early dialogues.)

Platonism in Islamic philosophy

*Plato seems to have been more an icon and an inspiration than an authentic source for Islamic philosophers. So far as is known, the only works available to them in Arabic translation were the Laws, the Sophist, the Timaeus and the Republic. His name was often invoked as a sage and an exemplar of that wisdom available to humankind among the Greeks before the revelation of the Qur'an. This in itself could represent a kind of affront to orthodox Islam, which tended to view the human situation before the Qur'an's 'coming down' as one of pervasive ignorance (*jahaliyya*). However, the rise of humanist culture in Baghdad during the ninth and tenth centuries* AD, *which involved Syriac Christian translators, presupposed a gradual acceptance of Greek wisdom in which Plato figured paradigmatically, even though far fewer of his works were made available in translation than those of Aristotle.*

Plato's influence on Islamic philosophy can be observed most clearly in ethics and political philosophy, given the works available to Islamic thinkers. However, his role lay more in creating an environment hospitable to philosophical reflection than in contributing to the formation of specific philosophical doctrines (where the influence of Aristotle was stronger) (see Aristotelianism in Islamic philosophy). He was referred to as 'the sublime and divine Plato', no doubt because his writings seemed to lead one more directly than any other Greek philosopher to reflect on human actions in the light of transcendent goals. At this level of inspiration, collections of sayings attributed to Plato, notably on the adverse relation of knowledge to wealth and power, helped to set a stage on which philosophy could play a propaedeutic role for Muslims seeking truth as they followed the 'straight path' laid out in the Qur'an. At the same time 'philosophy' so practised could present itself as an encompassing way of life, so competing with observant Islam. Here a discussion inspired by Plato regarding the relative weight of logic and grammar is relevant, since Arabic had tended to legislate semantic conflicts by recourse to grammar, while Greek philosophical texts (themselves originating in another language) extolled logic as a norm for rational discourse, transcending the peculiarities of a single tongue and the grammar proper to it. This potential conflict came to the fore in considering the qualities required for a just ruler of a Muslim polity, specifically regarding the relative merits of 'prophecy' (the generic Islamic term for the deliverances of revelation) and philosophical reason.

The *locus classicus* for such considerations is Plato's *Republic*, which offered an ideal paradigm for a just ruler that was adopted in lieu of Aristotle's more legislative treatment in the *Politics* - the only text of Aristotle's not translated into Arabic. Al-Farabi's treatise on the 'perfect state' (*al-Madina al-fadila (The Virtuous City)*) presents a didactic Neoplatonic version of Plato's *Republic*, one in which 'the Good' is transmuted into 'the First' in such manner that the ordering proper to cosmos and the microcosmic ideal polity emanates from the ever-fruitful One (see al-Farabi §§2, 4). Al-Farabi states unequivocally that philosophical reason outstrips prophecy as a requisite for the wise and just ruler, but the pattern established in his treatise was able to be adapted by those who weighed their relative merits otherwise. What was severely contested, however, was the relevance of Plato's ideal scenario (or its adaptation by al-Farabi) to the actual ruling of an Islamic polity. Rulers themselves took issue with it, speaking from experience, as did intellectuals (such as al-'Amiri) who assimilated Plato's lofty philosophical ideals to Sufi ascetic practices. For such as these, Plato's dictum that philosophers are prevented from attaining wisdom by the mores of the city in which they live spoke more directly to their experience.

Plato's teaching on the human soul as 'an incorporeal substance that moves the body' seemed to offer a philosophical teaching conducive to Islam, even though Ibn Sina's way of adopting this teaching would put 'philosophy' in conflict with Qur'anic faith in resurrection of the body (see Ibn Sina §6; al-Ghazali §3). Ethical thinkers like Ibn Miskawayh adopted Plato's tripartite division of the soul, however, in elaborating an ethical teaching relating Islam to a wider humanist culture, relying on extant sayings which quoted Plato: 'whoever rules his reason is called wise; whoever rules his anger is called courageous; and whoever rules his passion is called temperate.' The influence of sayings of this sort would permit a wise ruler like Ibn al-'Amid to say that he considered himself a 'member of the following [*shi'a*] of Plato, Socrates and Aristotle'. In this manner, Plato contributed an anthropology to Islamic thought which could be used to elaborate the 'straight path' of the Qur'an as well as bring it into contact with a wider humanist civilization.

See also: Aristotelianism in Islamic philosophy; al-Farabi; Greek philosophy: impact on Islamic philosophy; Neoplatonism; Plato; Platonism, Early and Middle

DAVID BURRELL

References and further reading

Alon, I. (1991) *Socrates in Medieval Arabic Literature*, Leiden: Brill.(Interesting account of the different ways in which the figure of Socrates was used in the history of Islamic philosophy.)

Badawi, ‘A. (1968) *La transmission de la philosophie grecque au monde arabe (The Transmission of Greek Philosophy in the Arab World)*, Paris: Vrin.(Detail of texts of Plato available to Arabs.)

al-Farabi (*c*.870-950) *al-Madina al-fadila (The Virtuous City)*, trans. R. Walzer, *Al-Farabi on the Perfect State*, Oxford: Clarendon Press, 1985.(Transmutation of Plato’s *Republic*.)

Kraemer, J. (1979) ‘Alfarabi’s *Virtuous City* and Maimonides’ Foundations of the Law’, in M. Kister, S. Pines and S. Shaked (eds) *Studia Orientalia Memoriae D.H. Baneth Dedicata*, Jerusalem: Magnes.(Analysis of al-Farabi’s transmutation of the *Republic*.)

Kraemer, J. (1986a) *Humanism in the Renaissance of Islam*, Leiden: Brill.(Describes the cultural revival in Baghdad during the Buyid age.)

Kraemer, J. (1986b) *Philosophy in the Renaissance of Islam*, Leiden: Brill.(Abu Sulayman al-Sijistani and his philosophical environment.)

Ibn Miskawayh (940-1030) *Tadhib al-akhlaq (Cultivation of Morals)*, ed. C. Zurayk, Beirut, 1967; trans. C. Zurayk, *The Refinement of Character*, Beirut, 1968. (A summary of Ibn Miskawayh’s ethical system.)

Peters, F. (1996) ‘The Greek and Syriac Background’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 3, 40-51.(Discussion of some of the important features of Greek and Syriac culture as sources of Islamic philosophy.)

Rosenthal, F. (1940) ‘On the Knowledge of Plato’s Philosophy in the Islamic World’, *Islamic Culture* 14: 398-402.(Traces some key sayings of Plato.)

(1975) *The Classical Heritage in Islam*, trans. E. and J. Marmorstein, London: Routledge.(A rich description of Plato’s heritage in the Islamic world.)

Walzer, R. (1960) ‘Aflatun’, in *Encyclopedia of Islam*, Leiden: Brill, vol. 1: 234-5. (Survey of Platonic texts available to Arabs.)

Walzer, R. (1962) ‘Platonism in Islamic Philosophy’, in *Greek into Arabic: Essays in Islamic Philosophy*, Columbia, SC: University of South Carolina Press.(Reprint of lecture detailing four aspects of Plato’s influence.)

Platonism, Early and Middle

Platonism is the body of doctrine developed in the school founded by Plato, both before and (especially) after his death in 347 BC*. The first phase, usually known as 'Early Platonism' or the 'Early Academy', ran until the 260s* BC*, and is represented above all by the work of Plato's first three successors, Speusippus, Xenocrates and Polemo. After an interval of nearly two centuries during which the Academy became anti-doctrinal in tendency, doctrinal Platonism re-emerged in the early first century* BC *with Antiochus, whose school the 'Old Academy' claimed to be a revival of authentic Platonism, although its self-presentation was largely in the terminology forged by the Stoics. The phase from Antiochus to Numenius is conventionally known as Middle Platonism, and prepared the ground for the emergence of Neoplatonism in the work of Plotinus. Its leading figures are Antiochus, Eudorus, Plutarch of Chaeronea, Atticus, Alcinous, Albinus, Calvenus Taurus and Numenius. Its influence is also visible in major contemporary thinkers like the Jewish exegete Philo of Alexandria and the doctor Galen.*

Like Neoplatonism, Early and Middle Platonism were founded on a very close reading of the text of Plato, especially the Timaeus*, often facilitated by commentaries, and further supplemented by knowledge of his 'unwritten doctrines'. However, Early and Middle Platonists did not develop nearly so elaborate a metaphysics as the Neoplatonists, and there was a much greater concentration on ethics. Most Middle Platonists regarded Aristotle as an ally, and incorporated significant parts of his thought into Platonism, especially in ethics and logic. Some were Neo-Pythagorean in tendency, and most claimed in some sense to be able to trace Platonic thought back to Pythagoras.*

The Platonists developed the dualism of the One (an active, defining principle) and the Indefinite Dyad (an indeterminate, material principle), bequeathed by Plato, especially through his oral teachings. These eventually emerged as, respectively, God and matter, supplemented by the Platonic Forms, which Middle Platonists typically identified with God's thoughts. The world-soul was distinguished from the demiurge or creator, who was in turn either distinguished from or collapsed into the primary divinity, a supreme intellect. Some, notably Plutarch, postulated in addition a counterbalancing evil world-soul. As regards the human soul, Plato's division of it into a rational plus two irrational parts was maintained, along with his doctrine of transmigration. There was also an increasing focus on the intermediary role played by daemons in the functioning of the world.

*In ethics, most Middle Platonists came to effect an assimilation between Plato's and Aristotle's views. All agreed with Plato and Aristotle, against the Stoics, that as well as moral there are also non-moral goods, such as health and wealth. While there was a consensus that the latter are not necessary for happiness, some, notably Antiochus, defended the view that non-moral goods are indispensable, at least to supreme happiness. In addition, Aristotle's doctrine that virtue lies in a 'mean' became a central feature of Platonist ethics. As for the 'goal' or 'end' (*telos*) of life, this came from the first century* BC *onwards, perhaps starting with Eudorus, to be specified by Platonists as 'likeness to God'. Finally, the issue of determinism was, in the wake of Hellenistic philosophy, recognized as important by Middle Platonists, who defended the existence of free will.*

Platonism never developed its own logic, but adopted Peripatetic logic, including both syllogistic and the theory of categories, both of which, it was claimed, had been anticipated by Plato.

1 Historical outline

Plato was succeeded in the headship of the Academy (see Academy) by his nephew Speusippus, whose most distinctive contributions to Platonism were developments of Plato's 'unwritten doctrines', based on a mathematicizing metaphysics. Speusippus' own successor, Xenocrates, did much more to further the formalization of Plato's thought into a system. He was succeeded in turn by Polemo, whose own contributions were mainly in ethics (see Ariston of Chios §2), although he is said still to have maintained the Platonic system as a whole: Polemo's contemporary Crantor was the first Platonic commentator. Soon after Polemo's death, the Academy became the sceptical 'New Academy', and remained so for nearly two centuries. Only in the early first century BC was a serious attempt made, by Antiochus of Ascalon, to recreate doctrinal Platonism under the title 'Old Academy', although his version was perceived as uncomfortably close to Stoicism. Antiochus, like most Middle Platonists in his wake (the main exceptions being Eudorus and Atticus), regarded Aristotle as himself a Platonist, whose ideas could, with some caution, be incorporated into Platonism.

Antiochus strongly rejected the claim of the sceptical Academics to be true to the Platonic tradition, although some later Platonists, notably Plutarch of Chaeronea, were prepared to accept the contribution of the New Academy to the overall development of Platonism, and to use the arguments of the New Academy (as, indeed, was Antiochus himself on occasion) to combat the Stoics. It was perhaps around this time that a tradition grew up which held that the New Academics had ‘dogmatized in secret’, using scepticism merely as a device for confounding their opponents, particularly the Stoics, and revealing positive Platonic doctrines to trusted disciples after years of initiation. In general, however, the scepticism of the New Academy came to be regarded as an aberration in Platonism.

The Athenian ’Academy’ does not seem to have survived into the later part of the first century BC (we hear of only two other heads, the first of whom was Antiochus’ brother, Aristus). Thereafter there was no central Platonic school, and the centre of activity in Platonism shifted for a while to Alexandria, where we find the shadowy figure of Eudorus propounding a more transcendental, Pythagorean-influenced Platonism. He apparently wrote a commentary on Plato’s *Timaeus*, as well as a critical one on Aristotle’s *Categories*. He also seems to have written on the first principles of the Pythagoreans, expounding his own doctrine in the process. Perhaps close in date to Eudorus is the anonymous commentator on Plato’s *Theaetetus*, part of whose commentary survives on papyrus.

The next known representative of Platonism, in the early first century AD, is the emperor Tiberius’ court philosopher Thrasyllus, who came from Alexandria, and produced what was to become the definitive edition of Plato’s works (see Plato §§2-3). In addition, something can be learnt of contemporary Platonism from the works of the Jewish philosopher Philo of Alexandria. The scene then moves back to Athens, in the latter half of the first century, with Plutarch’s teacher Ammonius, who was prominent in civic life there. Plutarch himself did not continue Ammonius’ school - it is doubtful whether anyone did - but set up a circle of his own in his home town Chaeronea.

Plutarch is an interesting mixture of influences. On the one hand he supported an almost Persian degree of dualism in his interpretation of Plato, which involves, among other things, a literal interpretation of Plato’s creation myth of the *Timaeus*, and belief in an independent, irrational world-soul. On the other he fostered a hospitality to the tradition of the New Academy, which sets him at odds with the tradition of both Antiochus and Eudorus, although this did not make him a sceptic to any serious extent. In ethics, Plutarch inclined to Aristotelian moderation rather than Stoic asceticism.

In the next generation, we find in Athens Calvenus Taurus from Beirut, who ran a school of much the same simple nature as that of Ammonius (although we cannot judge whether he is in a strict sense Ammonius’ successor). Taurus was, like Plutarch, Aristotelian in his ethics, but he did not share Plutarch’s views on the temporal creation of the world. He was perhaps succeeded in Athens by Atticus, who may also be the first incumbent of the Chair of Platonic Philosophy which the emperor Marcus Aurelius (§1) set up in Athens in AD 176. Atticus, however, agreed with Plutarch, against Taurus, on the question of the temporal creation of the world, and showed, in the one (polemical) work of his of which we have any fragments, a strong aversion to the Aristotelian tradition, with a corresponding hospitality to the Stoics. We know of a pupil of his, Harpocration of Argos, who seems, however, also to be influenced by the Neo-Pythagorean tradition emanating from Numenius.

Neo-Pythagoreanism is a significant force in Middle Platonism (see Neo-Pythagoreanism). Although most of its adherents wished to be regarded as a distinct sect, it seems more accurate to regard it as a phenomenon emanating from the Early Academy, and specifically from the tendency of Speusippus and Xenocrates to project their versions of Platonism back onto Pythagoras. Perhaps its most substantial figures are Moderatus of Gades, who repudiated the Platonists and claimed to be restoring the true doctrines of Pythagoras, and Numenius of Apamea (Syria), the ’grandfather’ of much of Neoplatonism, noted for his knowledge of, and hospitality towards, oriental philosophical and religious systems, including Judaism. Plotinus seems more influenced by Numenius than by any other Middle Platonist.

Other influential Platonists of the second century included Gaius and his pupil Albinus, the teacher of Galen. We also have two important handbooks of Platonism of the time: one, the *Didaskalikos* or ’Manual of Platonism’ by Alcinous (whom scholars long wrongly identified with Albinus), the other, *On Plato and his Doctrine*, composed by the distinguished north African rhetorician Apuleius of Madaura. Both are invaluable summaries of Platonic doctrine of the early centuries AD.

2 Plato's legacy

Plato bequeathed to his successors not so much a well-rounded and coherent philosophical system as an immensely fertile and provocative series of problems. Certainly there were distinctive doctrines that one can identify as Platonic - the theory of Forms or the doctrine of the (tri)partition of the soul - but even these, when examined, are seen to be fluid and open-ended. There was also a persistent tradition about 'unwritten doctrines' of Plato's, in particular, the derivation of everything from just two principles, the 'One' and the 'Indefinite Dyad'. Such views are better seen as working hypotheses, to be taken up and developed at will by his successors.

Plato's immediate successors were not obedient disciples, but independent thinkers who, to paraphrase the words of the most independent of them, Aristotle, 'loved Plato, but loved truth more'. Nor, so far as we can see, did Plato demand or encourage unquestioning acceptance of his beliefs. The whole rationale of composing dialogues rather than formal treatises, as well as the process of dialectic, indicates the opposite. It is in fact a profound mystery how Plato presented, or 'published', the dialogues, and what relation they were meant to have to the on-going work of the school. In the case of at least one dialogue, the *Timaeus*, one of the most influential in later times, and in view of the controversy that sprang up immediately after Plato's death as to its correct interpretation, it would seem that the master himself declined to give an authoritative exegesis of it. If, among Plato's own pupils, Aristotle could maintain that Plato taught the temporal creation of the world (*On the Heaven* I 12, II 2) while Speusippus and Xenocrates could deny this, claiming that Plato only presented such a scenario 'for instructional purposes', it cannot be the case that Plato himself made his intentions clear.

If then the dialogues are not the sole repository of Platonic doctrine, what is? There is, in fact, considerable secondary evidence as to the views entertained by Plato on certain basic questions. A good deal of this is preserved only in late authorities, such as the commentators on Aristotle (particularly Alexander and Simplicius) - although such information is sometimes explicitly referred back to contemporaries of Plato, such as Hermodorus. The greatest single source is however Aristotle himself, and Aristotle had been an associate of Plato's for twenty years. It has been claimed that Aristotle's reports of Platonic doctrines not present in Plato's dialogues are based on repeated misunderstandings of the content of the dialogues (Cherniss 1945); this view is too extreme, but we must always remember that Aristotle, although he doubtless knew what he was talking about in each instance, is nearly always giving a polemical and allusive, rather than a scholarly and systematic, account of what he knows. Nevertheless, by looking at his evidence with a properly critical eye, one can deduce a body of doctrine which is, in broad outline at least, coherent and reasonable, and fits in well with the developments attributed to Speusippus and Xenocrates.

3 Early Platonist metaphysics

To begin with first principles, it seems clear that Plato, in his later years at least, had become increasingly attracted by the philosophical possibilities of Pythagoreanism: that is to say, the postulation of a mathematical model for the world (see Pythagoreanism §2). Mathematics seems to have been a discipline much worked on in the Academy, and the insights derived from this (amplified by the researches of his mathematician colleagues such as Eudoxus, Menaechmus and Theaetetus) drove Plato progressively to certain general conclusions. He arrived at a system which involved a pair of opposed first principles, and a triple division of levels of being, the latter doctrine giving a vital central and mediating role to the soul, both world-soul and individual soul. Reflections of these basic doctrines can be seen in such mature dialogues as the *Republic*, *Philebus*, *Timaeus* and *Laws*, but they can not be deduced from the dialogues alone.

As first principles Plato established the One and the Indefinite Dyad. The One is an active principle, imposing 'limit' (*peras*) on the limitlessness (*apeiron*) of the opposite principle. The Dyad is regarded as a sort of duality (also termed by Plato 'the great-and-small'), infinitely extensible or divisible, being simultaneously infinitely large and infinitely small. The influence of the Dyad is to be seen all through nature in the phenomena of continuous magnitudes, and in excess and defect which have continually to be checked by the imposition of the correct measure. This process has an ethical aspect also, since, anticipating a well-known doctrine of Aristotle, the virtues are to be seen as correct measures ('means') between extremes of excess and deficiency on a continuum (see §8).

The Indefinite Dyad is primarily the basic unlimitedness or 'otherness' on which the One acts, but it is also the irrational aspect of the soul, and again the substrate of the physical world, corresponding to Plato's 'receptacle' of

becoming. Plato notoriously does not have a doctrine of ‘matter’ (*hylē*) as such, in the Aristotelian sense, but he does address the question of the nature of such ’substratum’ as there may be to the action of the various active principles in his system, namely God, the Forms and the soul. He addresses this particularly in the *Timaeus*. For him, the ‘receptacle’ (*hypodochē*) or ’place’ (*chōra*), introduces into the Forms, as they are projected forth from the intelligible realm through the mediation of soul, an element of indefinite multiplicity and imperfection which it is beyond the power of God to overcome completely.

By acting upon the Dyad, ’limiting’ it, the One generates the Form-numbers. At this point, however, the evidence becomes confused. It seems that Plato came finally to view the Forms as numbers, or mathematical entities of some sort. A special importance is attached by him, as it was by the Pythagoreans, to the ’primal numbers’, one, two, three and four, and their sum total, ten (the decad). The first four numbers seem to be in some way inherent in the One, and come to actuality in the process of the initial limitation of the Dyad. In *Metaphysics* XIII 7, Aristotle attacks the Platonist doctrine of Form-numbers, which seems to describe a process whereby numbers are generated by, first, the Dyad producing the number two, by doubling the One, and then producing other numbers by adding to two and to each successive number either the One or itself. The whole process remains obscure however, as perhaps it was in Plato’s own mind. From this action of the primal numbers upon the Dyad, and its reaction on them, all other Form-numbers are generated.

These numbers are what the Platonic Forms have become. But how many of them are there? The numbers up to ten certainly hold some sort of distinctively basic position, but the multiplicity of physical phenomena requires that the basic numbers combine with each other in some way to produce compound numbers, which can stand as the formulas for physical phenomena. There must, in fact, be a hierarchy among the Forms. Within the decad, the first four numbers (the *tetraktys*) came to play a major part in Plato’s cosmology, as they did in those of his successors. They are the principles providing the link between the absolute unity of the One and the three-dimensional physical multiplicity around us. These four also have a geometrical aspect, although how that is linked to their essential nature is not clear in Plato (in Speusippus’ system, certainly, the geometricals are a separate level of being). None the less, one is also the point, two the line, three the plane and four the solid - the last two being triangle and pyramid respectively. (For early Platonist innovations on this scheme, see Speusippus §2; Xenocrates §2.)

4 Middle Platonist metaphysics

Middle Platonists were always in the position of oscillating between the poles of attraction constituted by Stoicism and Peripateticism (see Stoicism; Peripatetics), although they added to the mixture of these influences a strong commitment to a transcendent supreme principle, and a non-material, intelligible world apart from this one, which stands as a paradigm for it. Provided that one accepted this basic position, a fairly wide range of views on ethical, physical and logical questions was acceptable, there being no central body or institution to prescribe any strict degree of orthodoxy.

By the first century BC Platonism had established a system of three principles: God, matter and the system of Forms. The fact that the Forms were viewed as the contents of God’s mind does not seem to have conflicted with their status as an independent principle, the paradigmatic cause of the world. This was presumably on the ground that the object of God’s thought was also an objective reality, true being.

Throughout the Middle Platonist period, a dominant theme is the nature and activity of the supreme principle, or God. Some later Platonists preserved the opposition of the Old Academy of Monad, or One, and Dyad, although they varied the relationship that they postulated between the two: Antiochus seems simply to have accepted the Stoic pair of an active and a passive principle (the Platonic Forms being subordinated to the Stoic *logos* or cosmic ‘reason’ (see Stoicism §§3, 5; Logos §1), but Eudorus, while re-establishing the Academic and Pythagorean Monad and Dyad, placed above them both a supreme One, possibly drawing some inspiration here from the metaphysical scheme of Plato’s *Philebus*. Plutarch returned to the basic duality, but he and his follower Atticus granted the Dyad rather more independence than would be traditional for Platonism, because of their commitment to a pre-cosmic ’evil’, or disorderly principle. As for the Forms, Plutarch and Atticus saw them as thoughts of God, although Atticus was accused by Porphyry later of maintaining that they are outside the divine intellect, which probably only means that he wished to assert their objective reality.

For Alcinous and Apuleius, God is dominant and matter simply passive, not even attaining to actuality, while the Forms are, once again, God’s thoughts. In the Neo-Pythagoreanism of Numenius, however, a radical dualism seems to be asserting itself, although partially held in check by the influence of mainstream Platonism. Since it was from Numenius more than from any other variety of Platonism that Plotinus derived his inspiration, this tendency to dualism later persisted within his thought.

It is important also to notice the remarkable system propounded by the Neo-Pythagorean Moderatus of Gades at the end of the first century AD, since it seems to anticipate to some extent Plotinus’ system of *hypostases*. According to Porphyry, Moderatus attributed to Plato a system according to which ‘the first One is above being and all essence, while the second One - which is the ‘truly existent’ and the object of intellection - is the Forms; the third, which is the realm of the soul, participates in the One and the Forms, while the lowest nature, that which comes after it, that of the sense-world, does not even participate, but receives order by reflection from those others’. Moderatus sees an appropriate level of matter manifesting itself at each of these levels.

5 Psychology

In early Platonist metaphysics, it may be only at the level of soul that the four basic numbers assume their geometrical aspect. The composition of the soul in the *Timaeus* is certainly intimately bound up with the doctrine of the four dimensions (see §3) and the relations of proportion between them. From the soul, the four dimensions are projected upon matter, in the form of combinations of basic triangles, to form the four elements, fire, air, water and earth. Thus an uncompromisingly mathematical model of the universe is laid down.

The doctrine of the soul is a central feature of Plato’s system, but it is not without its obscurities. The form which the Form-numbers take on in the soul, and the manner in which they are projected upon matter, are problems to which one can find no satisfactory answer in surviving documents. These problems relate primarily to the world-soul, but the individual soul is of the same composition, although its relation to the world-soul is obscured by figurative language in the *Timaeus* (41d). (For one development of Plato’s account, see Xenocrates §2.) In the individual soul, if we may believe Aristotle (*On the Soul* I 2), the four dimensions of point, line, plane and solid correspond to four levels of cognition: intuitive knowledge (*nous*); scientific knowledge (*epistēmē*); opinion (*doxa*); and sense-perception (*aisthēsis*). This fourfold division seems to be alluded to in Plato (*Laws* X, 894a), and the Line simile of the *Republic* (VI, 509d-) may in some ways adumbrate it.

The questions of the survival of the individual soul, and of its periodic reincarnation, are not settled definitively in early Platonism. Plato certainly argues for the immortality of the soul, but does this include the two lower, irrational elements of the soul, its ‘spirited’ and ‘appetitive’ parts? These are called ’mortal’ in the *Timaeus*, but the *Phaedrus* suggests on the contrary that they too might be immortal. Again, is the process of transmigration into animals to be taken seriously, as a passage in the *Phaedrus* (249b) might suggest, or can human souls only pass into other human bodies? These questions were answered variously by later Platonists. The soul is certainly the linch-pin of the universe for Plato, mediating as it does between the intelligible and sensible realms of existence, and containing in itself the elements or principles of both.

Speusippus and Xenocrates are reported to have considered the entire tripartite soul immortal, but no details of their arguments are reported. The most developed Middle Platonist descendant of their view is in chapter 25 of Alcinous’ *Didaskalikos*. The discarnate soul has three parts: the ‘discriminative’ or ‘cognitive’, the ‘impulsive’ or ‘dispositional’, and the ‘appropriative’. These seem to be three irreducible aspects of the soul, when viewed in dissociation from specific bodily organs. Only during incarnation do they become respectively the ‘rational’ (or ‘calculating’), the ‘spirited’ and the ‘appetitive’ parts, harnessed to the requirements of incarnate life.

6 Theology

Part of the process of tidying up loose ends in Plato’s thought which seems to have taken place in the Early Academy involved distinguishing between the Good of the *Republic* (and perhaps the ‘Monad’ from the ‘unwritten doctrines’) and the creator or ‘demiurge’ of the *Timaeus*. The latter could not straightforwardly be regarded as a supreme deity, since he turns to another entity, the paradigm, or ’essential living being’, which is not of his making. There are thus the conditions here for a distinction between a primary and secondary God, the first totally transcendent, and no more than the ultimate condition of all being (in something of the same way as

Aristotle's 'unmoved mover'), the second an active creator.

How this matter was handled in the Early Academic period is less than clear, however. Both Speusippus and Xenocrates recognized hierarchies of being, but they do not seem to have made this particular contrast. Middle Platonism does come to recognize besides the first principles two intermediate and mediating entities which are to some extent in competition with one another: namely, the demiurge and the world-soul. If the *Timaeus* was not taken as literally creationist, the demiurge could properly disappear, and blend into the supreme divine intellect, becoming its active, causative aspect. This is the path followed in Philo of Alexandria's *On the Creation of the World* and Alcinous' *Didaskalikos* (ch. 10). For others, however, the demiurge comes to be seen as a subordinate God, sometimes, as by Atticus (fr. 8), assimilated to the Stoic *logos*, but sometimes, as by Numenius (frs 15 and 16), presented as a distinct creator-God, subordinate to a supreme, static, self-contemplating entity, derived from the Good of the *Republic*. As for the world-soul, it starts out as the entity whose creation by the demiurge is described in the *Timaeus*, but traces appear, in such thinkers as Philo and Plutarch, of a rather more august figure, which almost seems to reflect the Speusippean principle of multiplicity, a figure not positively evil, but simply responsible for multiplicity, and thus for all creation. In Philo this figure appears as *sophia*, or God's 'wisdom', an entity interchangeable with the Dyad. Elsewhere, as in Alcinous' *Didaskalikos* (chaps 10 and 14) and in Plutarch's *On the Creation of Soul in the Timaeus*, the world-soul is depicted as an irrational entity, requiring 'awakening' by God, who is the demiurgic intellect.

7 Daemons

The question of daemons, or intermediate beings, is something which Plato leaves vague, but was of considerable importance for his successors. Plato's chief exposition of the role of daemons occurs in *Symposium* (202e), where they are presented as messengers or intermediaries between gods and humans, hence themselves less than perfectly divine. Daemons of similar status occur in the myth in Plato's *Statesman*, assisting the god in his administration of the world (271d-e), and the so-called 'young gods' of the *Timaeus* (who are at least partially to be identified with the planetary and star-gods) perform similar tasks for the demiurge.

In time two theories emerged on the nature of daemons, one 'static', the other 'dynamic'. Both these theories recognize a place for 'evil' daemons, either as retributive agents of God, or as souls undergoing punishment in the universe who are filled with malice towards human beings. Xenocrates is the earliest Platonist who seems to have developed a theory of daemons: he held the ‘static’ view - which he expressed in geometrical terms - that they are a distinct and permanent level of being. His daemons sound like permanent fixtures in the universe, although the question of their relationship with disembodied souls is unclear in the evidence available to us. The later, alternative theory, represented by such figures as Plutarch and Apuleius, is a ‘dynamic’ one according to which daemons are in fact souls, on their way either up or down the scale of being towards complete purification, and thus divinization, in the sun, or conversely towards renewed embodiment on the earth (see Plutarch of Chaeronea §5).

The Middle Platonist cosmos was in fact filled with subordinate, intermediate beings, chiefly daemons, but also in later times with such entities as heroes and angels. Heroes and angels, the latter possibly non-Hellenic in origin (they are found in the Chaldaean Oracles, frs 137-8; see Chaldaean Oracles), were certainly accepted by Neoplatonic times into the Platonic universe. Heroes are more respectable, but the distinction between them and daemons in the Middle Platonic period is not quite clear. The Stoic Posidonius wrote a treatise *On Heroes and Daemons*, but it is lost. One distinction might be that heroes are souls formerly embodied, but this distinction assumes a permanent class of unembodied souls, which is only acceptable on the 'static' theory. Whatever the differences in detail, however, it is common ground for all Platonists that between God and humans there must be a host of intermediaries, in order that God may not be contaminated or disturbed by too close an involvement with matter.

8 Ethics

Many of the main concerns of later Platonist ethics already appeared in the Early Academy. Definitions of happiness were produced by Speusippus, Xenocrates and Polemo. Polemo's formula, 'life in accordance with nature', made him an important forerunner of Stoicism (see Stoicism §17). At the very beginning of the Middle Platonist era, Polemo's definition of happiness, now identified formally as the ‘end’ or ‘goal’ (*telos*) of life, was

revived by Antiochus. When we turn to Alexandrian Platonism, however, in the person of Eudorus, we find that the Antiochian definition, influenced as it is by Stoic thought, had been abandoned in favour of a more spiritual, perhaps more truly Platonic, ideal of 'likeness of God' (*homoiōsis theōi*), derived from a famous passage of Plato's *Theaetetus* (176b), and from the *Timaeus* (90a-d). This formula remained the distinctive Platonic definition of the *telos* ever afterward.

A second key issue was whether virtue is sufficient for happiness. Polemo championed the view that, although virtue is sufficient for happiness, there are also non-moral goods. This goes back to Plato, and especially to his distinction in *Laws* I, 631b-c, between higher, or 'divine' goods - that is, goods of the soul, the virtues - and lower or 'human' goods - that is, goods of the body such as health or beauty, and external goods such as good fame or wealth. This passage was constantly cited in later Platonist debate. By Antiochus' time the battle lines had already been clearly drawn between the Peripatetics and the Stoics, and Platonists had to make their choice. Antiochus, although Stoicizing in so much else, sided here with the Peripatetics and the Early Academy in declaring that for complete happiness all three classes of good were required in some measure: goods of the body and external goods as well as the virtues. In this he was opposed by Eudorus, who sided with the Stoics, declaring that the two inferior classes of good could not count as an integral part of happiness, or the *telos*. This was not the end of the argument however. All through the period the two alternatives secured adherents: Plutarch and Taurus agreed with Antiochus and the Peripatos, Alcinous and Atticus with the Stoa.

A frequent further feature of Middle Platonist ethics, well represented in Apuleius and Alcinous, is the incorporation of the Aristotelian definition of virtue as based on a 'mean' (see Aristotle §22). There are in fact sufficient indications that Plato (see especially *Statesman* 283e-285b) had regarded the virtues as intermediate states between the two extremes of 'too much' and 'too little', with justice (symbolized by the Pythagorean *tetraktys* (see §3)) as the force holding the universe together - a metaphysical concept as well as an ethical one.

A question with considerable consequences for ethics, although it was seen rather as a part of physics, is that of free will and necessity, with which is intertwined that of God's providence. Before Epicurus and the Stoics (Chrysippus in particular) had stated the problem of determinism in its starkest form (see Epicureanism §12; Stoicism §§20-1), the question had not, it seems, been one of great urgency. Plato presents it only in mythical form (*Republic* X, especially 617d-619a, *Phaedrus*), but he does, like Aristotle (§20), maintain a belief in personal freedom of choice. Aristotle treats the suggestion that there is no such thing as freedom of choice as a mere sophistic paradox (*Nicomachean Ethics* III). Xenocrates did write an essay on fate (*heimarmenē*), but its contents are unknown. For the Middle Platonists, the problem of free will and necessity could not be dismissed so easily, and they found little help in Plato or Aristotle, although they did make appeal to key passages of both. Philo of Alexandria, in his *On Providence*, gives us the first defence of the Platonist position, which asserts both freedom of the will and the existence of providence with more vigour than logical force. Plutarch also touches on the theme repeatedly, although his most serious discussions of the subject have not survived: the essay *On Fate* surviving under his name is certainly not by him, though it is of great interest for Middle Platonic theory. Alcinous and Apuleius both contribute short discussions of the topic. In general, however, the Middle Platonists, while producing many scholastic formulations, fail to solve the problem, and bequeathed it in all its complexity to Plotinus.

9 Logic

In the area of logic, we must distinguish between what we can discern of genuinely Early Academic logic and what ultimately became accepted in the Platonic school, the latter being nothing less than Aristotelian logic in its entirety and including developments by Aristotle's pupils Theophrastus and Eudemus. The evidence points to Plato's having bequeathed to the Academy the system of division (*diairesis*), as first outlined in the *Phaedrus* (265d-), a system which acquired also a cosmogonic aspect. It is the soul's business to bring order out of chaos by making the right 'divisions', hitting the right means and harmonies. There is no evidence that Plato himself developed anything as elaborate as the Aristotelian syllogistic and system of categories, although Aristotle's 'ten categories' could be seen as little more than an elaboration of the two basic categories of 'absolute' (that is, substance) and 'relative' adopted by the Early Academy. Plato himself seems to have distinguished the basic categories of 'absolute' (*kath' hauto*) and 'other-relative' (*pros hetera*), the latter subdivided into 'opposite-relative' (*pros enantia*) and 'relative' proper (*pros ti*), which in its turn was divided into 'definite' and

'indefinite'. This much was reported by Plato's own pupil Hermodorus (quoted by Simplicius, *On Aristotle's Physics* 247.30-; Hermodorus, fr. 7), undoubtedly representing the practice of the Academy in his day, not the evidence of the dialogues.

In the field of logic, the primary achievement of the Middle Platonic period was to appropriate Aristotelian logic, together with the developments attributable to Theophrastus and Eudemus, for Platonism. Theophrastus and Eudemus had developed the system of hypothetical syllogisms which had been more or less ignored by Aristotle. They distinguished between 'pure' and 'mixed' hypotheticals, the latter being those adopted later by the Stoics (although the Stoics seem to have been the first to develop a logic of propositions rather than terms) (see Logic, ancient). In so far, then, as the Middle Platonists borrowed from Stoic logic (which they also did), they doubtless felt justified in this by finding it largely prefigured in Theophrastus. We find this synthesis exhibited in Alcinous' *Didaskalikos* (ch. 6), as well as in Apuleius' *De interpretatione* and Galen's *Introduction to Logic*. In the *Didaskalikos*, both categorical and hypothetical syllogisms are discerned in Plato's dialogues, particularly *Parmenides*. Plato is further credited there with knowledge of the ten Aristotelian categories, which Alcinous discerns once again in the *Parmenides*, while Plutarch (*On the Creation of Soul* 1023e) sees them operating in *Timaeus* (37a-b). The anonymous commentator on Plato's *Theaetetus*, who finds the categories in the *Theaetetus*, also makes extensive use of Aristotle's *Topics* as a handbook for the analysis of Platonic arguments.

Along with this tendency towards synthesis, however, there existed a tradition within the anti- Aristotelian wing of Platonism (Eudorus, Lucius, Nicostratus and Atticus), criticizing Aristotle's *Categories* (although not, it seems, the rest of his syllogistic). The anti-Aristotelians challenged Aristotle's originality (he was accused of stealing the idea of the ten categories from the Pythagorean Archytas), his completeness and the validity of some of his distinctions. Also, taking him to be making a division of reality rather than of language, they asserted that his distinctions were suitable to the sensible world but not to the intelligible.

Despite all this activity, it cannot be said that the Middle Platonists added much that is valuable to the science of logic. There are a few lost works of Plutarch, such as *A Reply to Chrysippus on the First Consequent*, *A Lecture on the Ten Categories*, *A Discourse on Hypothesis*, and *On Tautology*, which sound interesting, but there is little reason to suppose, on the basis of Plutarch's surviving works, that they contributed much of basic importance (see further Plutarch of Chaeronea §6).

See also: Calcidius; Celsus; Galen; Neoplatonism §1

JOHN DILLON

References and further reading

Editions of texts and fragments for individuals in the Platonic tradition - Alcinous, Antiochus, Apuleius, Calcidius, Celsus, Galen, Numenius, Philo of Alexandria, Plato, Plutarch of Chaeronea, Speusippus and Xenocrates - can be found in the entries under their names. Some others are included in the following list.

Albinus (mid 2nd century AD) *Prologue*, ed. O. Nüsser, *Albins Prolog und die Dialogtheorie des Platonismus*, Stuttgart: Teubner, 1991.(A study of Albinus within the context of Platonist school instruction.)

Anon. (*c*.50 BC-AD 150) *On Plato's Theaetetus*, ed. G. Bastianini and D. Sedley, *Corpus dei papiri filosofici*, part 3, Florence: Olschki, 1995, 227-562.(Partially preserved on papyrus, the earliest surviving commentary on Plato.)

Atticus (early 2nd century AD) Fragments, in E. des Places (ed.), *Atticus, Fragments*, Paris: Les Belles Lettres, 1977.(The standard collection, with French translation and notes.)

Calvenus Taurus (mid 2nd century AD) Fragments, in M.-L. Lakmann (ed.) *Der Platoniker Tauros in der Darstellung des Aulus Gellius (The Platonist Tauro as presented by Aulus Gellius)*, Leiden and New York: Brill, 1994.(Full discussion of all references to Taurus' teaching activities.)

Cherniss, H. (1945) *The Riddle of the Early Academy*, Berkeley and Los Angeles, CA: University of California Press.(Includes useful material on early Platonist metaphysics; over-sceptical about the 'unwritten doctrines' of Plato.)

Crantor (early 3rd century BC) Fragments in H.-J. Mette (ed.) 'Zwei Akademiker heute: Krantor von Soloi und Arkesilaos von Pitane' (Two Academics in Contemporary Scholarship: Crantor of Solus and Arcesilaus of Pitane), *Lustrum* 26 (1984): 7-94.(With German commentary.)

Deitz, L. (1987) 'Bibliographie du platonisme impérial antérieur à Plotin' (Bibliography of imperial Platonism before Plotinus), in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de

Gruyter, II 36: 1, 124-82.(The fullest Middle Platonist bibliography currently available.)

Dillon, J. (1977), *The Middle Platonists*, London: Duckworth.(A full introduction to Middle Platonism; has much on Early Platonism also.)

Dodds, E.R. (1928) 'The Parmenides of Plato and the Origin of the Neoplatonic One', *Classical Quarterly* 22: 129-42.(Seminal contribution on Moderatus.)

Dörrie, H. and Baltes, M. (1987-96), *Der Platonismus in der Antike (Platonism in Antiquity)*, Stuttgart and Bad Cannstatt: Frommann-Holzboog, vols 1-4.(Monumental selection of sources with rich commentary.)

Gersh, S. (1986) *Middle Platonism and Neoplatonism, The Latin Tradition*, Notre Dame, IN: University of Notre Dame Press.(Useful accounts of Antiochus, Apuleius and Calcidius.)

Glucker, J. (1978) *Antiochus and the Late Academy*, Göttingen: Vandenhoeck & Ruprecht. (Important rewriting of the history of the Platonic 'school'.)

Harpocration (*c*. AD 175) ed. J. Dillon, 'Harpocration's *Commentary on Plato*: Fragments of a Middle Platonic Commentary', *California Studies in Classical Antiquity* 4 (1974): 125-46; repr. in J. Dillon (ed.) *The Golden Chain*, Aldershot: Variorum, 1990.(Collection of fragments, with translation and commentary.)

Hermodorus (mid 4th century BC) Fragments, in M. Isnardi Parente (ed.) *Senocrate - Ermodoro, frammenti*, Naples: Bibliopolis, 1981.(Fragments with Italian translation and notes.)

Krämer, H.-J. (1967) *Der Ursprung der Geistmetaphysik (The Origin of the Metaphysics of Mind)*, Amsterdam: Grüner.(Stimulating but speculative; good on Xenocrates.)

Merki, H. (1952) *Homoiōsis Theōi: Von der Platonischen Angleichung an Gott zu Gottähnlichkeit bei Gregor von Nyssa (Likeness to God: From the Platonic Assimilation to God to Likeness to God in Gregory of Nyssa)*, Freiburg: Paulusverlag.(The only full-length study of the Platonist *telos*.)

Moraux, P. (1973, 1984) *Der Aristotelismus bei den Griechen (Greek Aristotelianism)*, vols 1-2, Berlin and New York: de Gruyter.(Important on the Aristotelian elements in Middle Platonism.)

Plato (mid 4th century BC) Unwritten Doctrines in K. Gaiser, *Platons ungeschriebene Lehre (Plato's Unwritten Doctrine)*, Stuttgart: Klett, 1963.(The testimonia are collected at the end of the volume.)

Plato (*c*.366-360 BC) *Timaeus*, trans. F.M. Cornford, *Plato's Cosmology*, London: Routledge & Kegan Paul, 1933. (Plato's hugely influential work on cosmology.)

Polemo (late 4th-early 3rd century BC) Fragments, ed. M. Gigante, 'Polemonis Academici Fragmenta', *Rendiconti dell'Accademia Archeologica di Napoli* 51 (1976): 91-144.(Fragments in Greek and Latin only.)

Tarrant, H. (1993) *Thrasyllan Platonism*, Ithaca, NY, and London: Cornell University Press. (A reconstruction of Thrasyllus' contribution to the Platonist tradition.)

Whittaker, J. (1987) 'Platonic Philosophy in the Early Centuries of the Empire', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36: 1, 81-123.(A most useful survey, but somewhat concentrated on Alcinous.)

Zintzen, C. (ed.) (1981) *Der Mittelplatonismus (Middle Platonism)*, Darmstadt: Wissenschaftliche Buchgesellschaft.(Reprints a number of classic papers on Middle Platonism.)

Platonism, medieval

Medieval Platonism includes the medieval biographical tradition, the transmission of the dialogues, a general outlook spanning commitment to extramental ideas, intellectualism in cognition, emphasis on self-knowledge as the source of philosophizing, and employment of the dialogue form. Platonism permeated the philosophy of the Church Fathers, the writings of Anselm and Abelard, the twelfth-century renaissance, the Italian Renaissance and the northern renaissance. Indeed the mathematical treatment of nature, which inspired the birth of modern science in the works of Kepler and Galileo, stems in part from late medieval Pythagorean Platonism.

*The term 'Platonism' is of seventeenth-century origin. Medieval authors spoke not of Platonism but rather of Plato and of Platonists (*platonici*), applying the term 'Platonist' to an extreme extramental realism about universals, or a commitment to the extramental existence of the Ideas. Thus John of Salisbury characterized Bernard of Chartres as 'the foremost Platonist of our time' in regard to his theory of ideas. For Aquinas, Platonists hold an overly intellectualist account of human knowledge, ignoring the mediation of the senses. In general, medieval writers agreed with Cassiodorus' maxim,* Plato theologus, Aristoteles logicus. *Plato was primarily a theologian, an expert on the divine, eternal, immaterial and intelligible realm, a classifier of the orders of angelic and demonic beings, whereas Aristotle was primarily a logician and classifier of the forms of argument.*

Medieval Platonism combines elements drawn from Middle Platonism and Neoplatonism. It generally assumes a dualistic opposition of the divine and temporal worlds, with the sensible world patterned on unchanging immaterial forms, often expressed as numbers. It also affirms the soul's immortality and direct knowledge of intelligible truths, combined with a suspicion of the mortal body and a distrust of the evidence of the senses. Neoplatonists sympathized with Porphyry's aim (in his lost De harmonia Platonis et Aristotelis) of harmonizing Plato with Aristotle. A Platonic outlook (largely inspired by the Timaeus) dominates the early Middle Ages from the sixth to twelfth centuries, whereas the thirteenth and fourteenth centuries, the age of scholasticism, witnessed an explosion in the knowledge of Aristotelian texts, often transmitted through Arabic intermediaries. The new interest in Aristotle was such that, although the Timaeus was widely lectured on during the twelfth and early thirteenth centuries, by 1255 it was no longer required reading at the University of Paris. Interest in Plato re-emerged in the Italian Renaissance with the availability of genuine works of Plato, Plotinus and Proclus. Nevertheless, through Pseudo-Dionysius in particular, Platonism reverberates in many thirteenth-century authors, especially in theology.

1 Platonism and Christianity

Platonism's medieval popularity is related to its overall religious outlook. The early Christian Fathers, especially Clement of Alexandria, Origen, Gregory of Nyssa, Basil, Ambrose and Augustine, regarded Platonism as closer to Christianity than other ancient pagan philosophies, and readily grasped it as the vehicle for articulating the Christian message (see Patristic philosophy). Plato was praised for anticipating Christianity by recognizing the existence of a unique, transcendent, benevolent deity who freely created the world. Plato also taught the doctrine of a created immortal, rational human soul, made in the image of God (*Theaetetus* 176b), and even prescribed a way of salvation. Plato was also an expert on the nature of the divine intelligences. Eriugena, for example, thought of Plato as one who taught the nature of the world soul and of angels (*Periphyseon* III.732d). Indeed some, such as Peter Abelard and John of Salisbury, even discovered hints of the Trinity in Plato (Abelard, *Theologia Christiana* I.68; John of Salisbury, *Policraticus* VII.5).

Augustine's Platonism was hugely influential on medieval philosophy, and Book VIII of his *De civitate Dei (City of God)* is a convenient source book of ancient philosophical ideas. Augustine portrays Platonism as a systematic philosophy focused on unity, truth and goodness. In his early *De vera religione (On True Religion)*, Augustine claimed one need only change a few words to see how closely Plato resembled Christianity (*De vera religione* IV.7), and in *De civitate Dei*, Plato is portrayed as the philosopher closest to Christianity; for example, Plato had defined philosophy as the love of God (VIII.11). According to the *Confessiones*, Augustine's conversion to Christianity had been influenced by his reading 'books of the Platonists' (*libri platonicorum*), most likely Marius Victorinus' translations of Plotinus and Porphyry. These texts convinced Augustine that truth was incorporeal and that God was eternal, unchanging and the cause of all things, parallelling truths revealed in St Paul's epistles.

Augustine was deeply influenced by the Christian Neoplatonism of Victorinus: *Confessiones* VII.11 reproduces Victorinus' doctrine of the various levels of reality (see Marius Victorinus). Also according to Augustine, Plato's understanding of God as true being mirrored the Biblical definition of God as 'I Am Who Am' (Exodus 3:14).

Among the most important Augustinian texts for medieval readers were the *De genesi ad litteram (Literal Commentary on Genesis)* and *De doctrina Christiana (On Christian Doctrine)* which provided the medieval world with a semiotics and scriptural hermeneutics, influencing Cassiodorus, Rhabanus Maurus, Hugh of St Victor, Peter Lombard, Bonaventure, Grosseteste and Erasmus. *De doctrina Christiana* offered a formidable scriptural justification for Christian appropriation of pagan thought: just as the Jews fleeing captivity had borne off the spoils of the Egyptians, so too the Christians could make use of the pagan heritage to teach morality and religion. The work reiterates Ambrose's claim that Plato's wisdom came directly from the Prophet Jeremiah. Augustine later recognized that Plato lived about a century after Jeremiah, but he continued to entertain the notion that Plato could have learned about the Bible from contact with holy men in Egypt, and medieval philosophers, including Abelard and Ficino, carried on this tradition. However, the Christian Fathers often suspected Plato because of his commitment to the soul's pre-existence and transmigration, his polytheism and his silence on the incarnation (interpreted as an innocent ignorance by Petrarch and Ficino).

Platonism persisted in theological discussions on the nature of the divinity. Neoplatonic writers from Eriugena to Nicholas of Cusa thought of God as both beyond being (*superessentialis*) and yet the form of all created beings (*forma omnium*). Eriugena calls God the 'form of forms' (*forma formarum*); for Thierry of Chartres, God is 'the form of being' (*forma essendi*). Following the condemnations of Amaury of Bène and David of Dinant in 1210, neo-Aristotelian philosophers criticized these formulations as leading to pantheism. Thus Aquinas developed his distinction between the divine being and the individual being of each thing. However, both Meister Eckhart and Nicholas of Cusa later reapplied the formula *forma omnium* to God.

Aside from theology and cosmology, Platonism was evident in medieval epistemology, ethics and social and political thought. The Platonic emphasis on certain knowledge over opinion, on intellectual knowledge as opposed to the unreliable offerings of the senses, entered into the Middle Ages through Augustine. The Platonic doctrine of recollection continued in the Augustinian form of illuminationism (for example, in Bonaventure), whereby the mind is said to know by being illuminated from within (either by a natural or a divine light or by a combination of both) (see Augustinianism). Plotinus' identification of evil with privation and non-being was repeated by Augustine and Aquinas. In the medieval period, Platonism in mathematics, the view that mathematical entities (such as numbers and classes) exist separately in their own right, took the form of a defence of the reality of universals as real things (*res*) against the nominalist position that universals were merely words (*voces*) (see Buridan, J.; Universals).

2 The sources of medieval Platonism

Until the fifteenth century, the only Platonic texts available in the Latin west were part of the *Timaeus* (17a-53b) and, from the mid-twelfth century, the *Meno* and *Phaedo*. Medieval Platonism was largely indirect, filtered through the writings of the Christian Fathers, especially Augustine. Gregory of Nyssa's Platonism influenced Eriugena. Aspects of Platonism (for example, the theory of ideas) were also transmitted through Latin writers including Cicero, Seneca, Martianus Capella, Aulus Gellius, Macrobius, Boethius, Cassiodorus and Isidore (see Encyclopedists, medieval). Medieval discussions of Platonic ideas were based not on Platonic dialogues (for example, Aquinas shows no evidence of having read the *Meno*), but primarily on Augustine's discussion of ideas in his *De diversis quaestionibus LXXXIII (On Eighty-Three Different Questions)*, Question 46, which itself drew on Cicero's *Academics* I.19. For Augustine, the Platonic Ideas (*ideae*) were really divine paradigms in the mind of God. Augustine distinguished the divine ideas from the *logoi* or *rationes* of things, created forms which guaranteed the continuity of the species through time, a version of the Stoic seminal reasons discussed chiefly in his *De genesi ad litteram*. Similarly, medieval people learned of Platonic arguments for the immortality of the soul from Cicero's *Tusculan Disputations*.

The *Timaeus* was the only Platonic dialogue widely circulating through the whole medieval period, available in the fourth-century Latin translation of Calcidius. Cicero's earlier translation (*De universo*) was almost unknown in the early Middle Ages (although Augustine expressly cites it in *De civitate Dei* XIII.16). Also in wide circulation in the Middle Ages was Calcidius' extensive, eclectic *Commentary on the Timaeus*, mingling elements of Middle

Platonism (inspired by Numenius) and Porphyrian Neoplatonism. It popularized a Middle Platonist view of Plato for the Middle Ages, in which the cosmos is constructed from three principles: God (*deus*), form (*exemplum*) and matter (*silva*). The first principle, God, is the Supreme Good, cause of all and end of all, beyond substance and nature and above all intellect. God is also characterized as free will and as providence. The second principle is matter, understood as neither sensible nor intelligible, lacking all form. The divine mind informs matter as soul vivifies body. The third principle is form, the intelligible world of the Ideas, understood as God's thoughts. The world soul, understood as made by God, is a kind of second mind. Calcidius was frequently glossed in the twelfth century in particular; of interest to medieval writers was his discussion of the four elements and his number speculation. Interest in numbers was justified by Scripture, and buttressed by Augustine in *De musica* and by Boethius' *De arithmetica*, and by other texts that communicated Pythagorean Platonism to the West (for example, Martianus Capella's *De nuptiis Philologiae et Mercurii (The Marriage of Philology and Mercury))*. Calcidius' *Commentary* includes references to the argument in *Phaedrus* that the soul is self-moving, the comparison in *Theaetetus* of the mind with a wax tablet and the *Republic*'s comparison of the Form of the Good with the sun.

Macrobius' *Commentarius in somnium Scipionis (Commentary on the Dream of Scipio)* was another influential source of Porphyrian Neoplatonism, especially important for twelfth-century philosophers such as William of Conches. Macrobius focuses on the final section of Cicero's *On the Republic*, his version of Plato's *Republic*, which is known as 'Scipio's Dream' (*somnium Scipionis*). Here a dream is recounted which is reminiscent of the Myth of Er, providing an account of the destiny of human souls emphasizing the need to live a life of virtue and hold the body in contempt. Macrobius' allegorical interpretation offers a typically Neoplatonic cosmology including the three hypostases, One, Mind and Soul. The account of the world soul includes a discussion of the nature of the self-mover. Macrobius gives an account of the procession of the soul from God down through the Homeric 'golden chain' of beings (see Encyclopedists, medieval §3). Martianus Capella's fourth-century allegorical compendium of the Liberal Arts, the *De nuptiis Philologiae et Mercurii*, also conveyed Platonic sentiments (see Encyclopedists, medieval §4), as did Cassiodorus' *Institutiones (Institutions)* and Isidore of Seville's *Etymologiae (Etymologies)* (see Encyclopedists §§6-7). These works were influential from the ninth century to the twelfth century in particular.

The most influential work of Boethius was his *De consolatione philosophiae (Consolation of Philosophy)*, which transmitted a Stoicized Platonism to the medieval world. This work presents the Platonic view that the soul can become forgetful of itself through immersion in the affairs of the body, but that it can recover its essential rational nature and attain to the vision of God. Several poems in the *De consolatione philosophiae* transmit in condensed form a Platonic cosmology, especially III metrum 9, '*O qui perpetua mundum ratione gubernas*' (You, who in perpetual order, govern the universe), which was widely commented on from the ninth century onwards (for example, by Remigius of Auxerre).

Also from the ninth century, medieval authors were exposed to another blend of Platonism deriving from Proclus and emphasizing the transcendence of the divine, to the extent that the divine is better described as non-being than as being. These Christian texts, purporting to be written by Dionysius, St Paul's convert at Athens, and thus as ancient and authentic as the Gospels themselves, were in reality pious forgeries produced by a sixth-century Christian follower of Proclus. Pseudo-Dionysius' *De divinis nominibus (The Divine Names)* examines scriptural and philosophical appellations for the divine and argues that they all fail to fully express the nature of the highest being, who is nameless and beyond all names. Names are really processions from the divinity and do not reach the divinity itself. Negations, in fact, express the nature of the divine more accurately than affirmations. This theme is expressed even more radically in the *De mystica theologia (Mystical Theology)*, which had enormous influence on the later medieval mystical tradition, transmitting to the Latin West the Platonism of the *Parmenides* in the form of negative theology. Pseudo-Dionysius had an enormous influence on Albert the Great, Aquinas, Bonaventure and Grosseteste among others, particularly through his concept of the self-diffusion of the good (*bonum diffusivum sui*), his principle that all things have being through being one, and his notion that the being of all things is the 'above being' of the divinity (*esse omnium est superesse divinitatis*) (see Pseudo-Dionysius).

Both Boethius and Pseudo-Dionysius contributed to the development of medieval Platonism by continuing to emphasize the primacy, transcendence and unspeakability of the one, good God. Pseudo-Dionysius, however, following Proclus, formalized the system of hierarchical levels postulated to exist between the divine One and the formless nothing, chiefly in two books: *De coelesti hierarchia (The Celestial Hierarchy)* and *De ecclesiastica*

hierarchia (The Ecclesiastical Hierarchy). Influenced by Pseudo-Dionysius' Proclean formulations, medieval Platonists thought of reality as a series of ontological levels which proceed from the One right down to the nebulous realm of formless matter.

Johannes Scottus Eriugena, an Irishman who resided at the Carolingian court, produced in the ninth century the first synthesis of the Platonism of Augustine and Pseudo-Dionysius. Eriugena's Platonism is indirect, through Greek Christian Platonists, Basil and Gregory of Nyssa as well as Pseudo-Dionysius. Though aware of the theological differences between the Augustinian and Dionysian traditions, he regarded them as different expressions of the one truth. Eriugena's dialogue *Periphyseon (On the Division of Nature)* developed a Neoplatonic cosmological system which synthesized Dionysian and Augustinian Platonism. All things proceed from and return to the One in an eternal cosmic cycle (*exitus-reditus*), which is at the same time God's self-articulation. The spatio-temporal world which appears solid and corporeal is really an incorporeal world of qualities which emanates from the primordial causes, which are eternal but created ideas in the mind of God. Eriugena boldly identifies Augustine's primary causes with Pseudo-Dionysius' divine willings, thus synthesizing eastern and western interpretations of the Platonic ideas.

All things must return to their source, and the divine ideas will be reunited in God. Human souls are originally one with the One, but in their outgoing they become shrouded in appearances, generating the corporeal body. Each will also return to be one in the Logos, though each soul will remain at the level dictated by the level of its intellectual contemplation. Eriugena follows Gregory of Nyssa in claiming that corporeal body is merely an illusion produced by the commingling of incorporeal qualities, and that the division of the sexes is a consequence of the Fall. The unformed matter from which God creates is really God's own hidden, transcendent nature. Although Eriugena refers to the world soul (*Periphyseon* I.476c), drawing on Macrobius and Virgil, it does not play a significant part in his system but perhaps is to be identified with the Holy Spirit, as in later twelfth-century Platonism.

3 Platonism in the debates about universals

The medieval debate over the ontological status of universals (signified by general terms such as 'animal' or 'man') re-enacts the dispute between Plato and Aristotle over the nature of forms. This problem emerges in Porphyry's *Isagōgē(Introduction)* where, introducing Aristotle's *Categories*, he raises a number of questions while commenting on Aristotle's *De interpretatione (Peri hermeneias)* I 16a 3-18, including whether universals had real existence or were only creations in the mind; if they were real, whether they were corporeal or incorporeal; and if incorporeal, whether they existed apart from sensible things or were contained in them (see Porphyry). Boethius, in his commentaries on Porphyry, suggested an answer to the question which attempted to reconcile Plato with Aristotle. For Boethius, following Aristotle, a universal is by its nature predicable of many things, and hence is not an individual and cannot be a sensible substance. However, neither is it a merely empty category since it is indeed the thought of something. The species is the result of the mind's abstracting from what is given in the senses. The universal is in itself immaterial, but it subsists in material beings and does not exist apart from things, as Plato thought (see Universals).

Medieval interpretations of Boethius led to the dispute between nominalism (identified with Roscelin of Compiègne) and realism (identified with William of Champeaux and Bernard of Chartres). William of Champeaux, a teacher at the Cathedral School of Notre-Dame in Paris, known chiefly through his student and critic Peter Abelard, was an extreme realist. According to John of Salisbury, Bernard of Chartres was also a Platonist about universals, holding that aside from the ideas in God there were created forms (*formae nativae*) that existed apart from God and from the mind (Dutton 1991: 70-96). According to Abelard, William of Champeaux held the view that a universal exists in common in each of the instances; thus man exists identically in Socrates and in Plato. In order to be identical in each instance, it must be identical in itself and to a thing existing apart. Roscelin, on the other hand, maintained that universals were mere names, utterances, 'vocal breath' (*flatus vocis*) as Anselm termed them (see Roscelin of Compiègne).

Peter Abelard, who had been taught by both Roscelin and William, attempted to mediate between these extremes, especially in his *Logica ingredientibus (Logic for Beginners)*. Abelard, who regarded Plato as the 'greatest of philosophers' (*maximus philosophorum*), held a modified Platonism about universals, whereby the universals are held to be forms existing in the mind of God (*conceptio Dei*). He ridiculed the nominalist position that species were merely words, but he was also critical of the view that species were things. Species were predicated of things,

and only words, not things, can be predicated of things. The problem is to decide what ontological status these *predicamenta* (substantial categories) have. For Abelard, a thing is always concrete and individual. When several things, such as humans, share a common nature, what they share is 'being human' (*esse hominem*). This is not itself a thing but neither is it merely a name; it is a concept founded in the thing but not existing in the same mode as the thing. Abelard's interest is in the meaning which is expressed by a proposition, the *dictum* (what is said), which is neither identical with the words nor with the thing but is something intermediate.

4 Twelfth-century Platonism

In the eleventh century, Platonism, mediated through Augustine's works, is evident in Anselm, most notably in the *Monologion* and particularly in this work's acceptance of the existence of the forms and the self-existent highest good. The twelfth century saw a Platonic renaissance, centred mainly in the cathedral schools of Chartres and St Victor (see Chartres, School of; Hugh of St Victor), and was characterized by cosmological speculation inspired by the *Timaeus*, combined with Boethius, Macrobius and Martianus Capella. The challenge was to produce a complete scientific picture in conformity with Genesis from the fragment of Plato's natural philosophy which was known.

From Philo and the Middle Platonists onwards, the parallels between Plato's cosmology in the *Timaeus* and the account of creation in Genesis provided the opportunity for Platonic commentaries on the work of the six days (the *Hexaemera*). *Timaeus* 41c was interpreted as teaching that the world is created by the will of God who is 'Father of all'. Abelard deduced from the *Timaeus* that God - the most perfect being - had created the most perfect world, a doctrine which was revived by Leibniz in the seventeenth century. For twelfth-century authors, Plato's literary method of exposition was similar to Christian parable; Plato taught using fables and symbols (*integumenta*, or 'coverings') which the commentator must interpret.

Using Calcidius' commentary on the *Timaeus* (see §2), twelfth-century Platonists developed an account of the world in terms of the four elements and in terms of complex number symbolisms. Among commentators on the *Timaeus*, perhaps the most Platonic were Bernard of Chartres, Thierry of Chartres and William of Conches. These writers stress the relation between macrocosm and microcosm, and harmony between the divine and created spheres. William of Conches, who probably taught at Chartres, is the most important of the twelfth-century Platonists, and his *Timaeus* commentary is the most extensive medieval commentary on that dialogue. For William, the *Timaeus* is a unified theological work displaying the beneficence of the creator. He also commented on Boethius' *De consolatione philosophiae* and on Macrobius, as well as composing two systematic works, *Philosophia mundi* and a revised version entitled the *Dragmaticon*, set in the form of a dialogue between the Duke of Normandy and the Philosopher. Many Pythagorean elements and much number symbolism was associated with the articulation of cosmology in the twelfth century. Thierry of Chartres in his *De sex dierum operibus (Concerning the Works of the Six Days)* sees creation as an articulation of unity into plurality, following the suggestion of Calcidius.

William, also following Calcidius, sees the *Timaeus* as a work of natural justice showing how God creates and governs the world. God has established an unvarying natural law which is discoverable at the heart of things. For him, God creates the intellectual realm and allows other causes (such as stars) to govern the lower world, thus proposing a doctrine of mediated creation at variance with Augustine's single-act view (see Natural philosophy, medieval). The Chartres school followed Bernard of Chartres in positing a level of created forms between God and sensible reality, influenced by Eriugena's primordial causes and Augustine's seminal reasons (see §2). Seeking to reconcile Plato and Aristotle, as John of Salisbury reports, Bernard of Chartres posited intermediaries between God and created things. These native forms (*formae nativae*) link the eternal archetypes to passive matter (*hyle*) (see Chartres, School of).

William's account of creation discusses the role of the four elements in detail. Like others (for example, Adelard of Bath), William saw himself as expanding on the teaching of Plato: 'It is not my intention to expound here the words of Plato, but to set down here the view of natural scientists [*physici*] concerning substances; but even if I have not expounded Plato's words, I have said all that he said about elements, and more' (*Dragmaticon*, quoted in Dronke 1988: 309). He attempts to define the elements and addresses the question as to whether they are perceptible by the senses and corporeal and whether the division of matter ends with these indivisibles (atoms). William takes the view that the four elements are corporeal, unchanging substances which, however, are only

found in combination. The elements then are corporeal but are actually grasped by intellect since they are too small to be perceived by the senses on their own. Though they are unchangeable, they are created. God first made the four elements from nothing and then everything else out of the four elements, except the soul of man, which God made directly.

A major challenge to Christianizing Plato's cosmology was to interpret the role played by the Platonic Demiurge (see Plato §16). Christian Platonists were initially quick to identify the Demiurge with the Logos, the Second Person of the Trinity. This allowed them to make a further identification between the Holy Spirit and the world soul (*anima mundi*), which in the *Timaeus* enlivens the material cosmos. William of Conches initially, in his commentary on Macrobius, quite boldly identified the world soul of the *Timaeus* with the Holy Spirit, as Abelard was alleged to have done. The Council of Sens had condemned the identification, attributing it to Abelard. William then appears to have grown more cautious, simply offering a number of different views in his *Philosophia* Book One (the world soul is the Holy Spirit, or a natural force implanted in things by God, or a certain incorporeal substance in bodies) and making no reference to the world soul in his *Dragmaticon*.

Bernard Silvestris, in his partly versified, allegorical account of the creation, the *Cosmographia*, makes use of many Platonic ideas from the *Timaeus*, including that of a world soul personified as Endelichia (who also appears in Martianus Capella and Cicero), but in a manner quite different from William of Conches. Bernard has a world of ideas (Noys) and a domain of unformed matter (personified as Silva - Calcidius' term for 'matter'). Gradually Noys imposes order on Silva until the whole world has been made. The sensible world imitates the intelligible; man is a microcosm of the macrocosm. Bernard also saw Plato as beginning with two principles: *unitas et diversum*, unity and diversity (see Bernard of Tours). William of Conches explicitly connects Plato with Pythagoras, and argues that since number possesses the highest perfection, nothing can exist without number. Another Platonic cosmology in versified form was Alan of Lille's *De planctu naturae (The Lament of Nature)*, a dialogue between the poet and Nature, which was influenced by Bernard Silvestris.

5 Platonism in the thirteenth century

Thirteenth-century knowledge of Plato drew on the usual sources in Augustine and Latin writers, but also gained new insights into Plato from criticisms contained in the rediscovered works of Aristotle and his Arabic commentators, especially Averroes (see Ibn Rushd). Rather stiff literal translations of the *Meno* and *Phaedo* were produced by Henricus Aristippus in the 1150s; although listed in the library of the Sorbonne after 1271, these texts were not much studied and had little influence. Similarly, William of Moerbeke's translation of part of the *Parmenides* with Proclus' commentary also had little influence until popularized by Nicholas of Cusa. Moerbeke also translated Proclus' *Elements of Theology* which was available to Thomas Aquinas. Moerbeke's follower, the Flemish encyclopedist Henry Bate, was one of the first to be able to discern the difference between Plato's own texts and the later Platonism of Proclus. In general, however, Platonism in the thirteenth century survived mainly in the universities' theology faculties, as the arts faculty syllabuses were gradually reorganized to accommodate the new Aristotelianism.

One of the most influential texts for thirteenth-century philosophers was the *Sentences* of Peter Lombard. Lombard stated that Plato had three principles to explain the cosmos: matter, forms and the divine artificer, whereas Aristotle had only two: matter and species. This passage was regularly commented on to clarify whether Plato and Aristotle accepted the doctrine of creation, and whether they thought that creation was compatible with the beginninglessness of the world (as Aquinas held). In his commentary on the *Sentences*, Albert the Great acknowledged that Plato had posited a world of forms that existed independently of the mind of God. Albert's outlook was strongly influenced by Neoplatonism and no doubt helped to shape the Platonism in the thought of his student, Thomas Aquinas.

Though Aquinas is the great exponent of the new Aristotelianism, adopting Aristotle's criticisms of the univocal understanding of the good in Plato and other criticisms of the existence of the Ideas, yet he remains quite Platonic in other domains, for example, in his account of participation (how created things participate in being and receive the gift of *esse* from the divine being) which was strongly influenced by Pseudo-Dionysius. Aquinas sides with Aristotle against Platonism, which he sees as a doctrine that overstressed the mind's intellectual capacities, claiming that humans could know immaterial forms directly without mediation of the senses. In *Summa contra gentiles* I.13.10, in his discussion of the argument for the existence of God from motion, Aquinas explicitly

discusses the difference between the Platonic and Aristotelian conceptions of the nature of motion, drawing on *Phaedrus* 245c. His source, however, is not directly Plato but more probably the tradition stemming from Calcidius and Macrobius.

Robert Grosseteste translated and commented on Pseudo-Dionysius. His *De luce(On Light)* offers a typically Neoplatonic cosmology and metaphysics of light. Grosseteste's account of the soul weds Aristotelian naturalism with a Neoplatonic account of the higher principles of intellect and reason.

Platonism in the thirteenth century is often associated with members of the Franciscan order and with a mathematical approach to the understanding of nature. Richard Rufus of Cornwall defended Plato's theory of ideas against Aristotle's criticisms, and Bonaventure's *Itinerarium mentis in Deum (Journey of the Mind to God)* is thoroughly Platonic (see Bonaventure). Bonaventure accepted a form of Augustinian illuminationism, a Christianized version of Platonic recollection (see Augustinianism). For Bonaventure, sensible things are traces of divine things. In typically Platonic terms, Bonaventure's sixth step in the mind's advancement towards God refers to the Good beyond being.

The later thirteenth century saw a re-emergence of an Averroist Platonism, particularly with regard to the knowledge of separate substances. Siger of Brabant, for example, took a more Platonic line than Aquinas had done in arguing for the soul's direct knowledge of separate intelligible substances. Henry of Ghent's doctrine of the separate being of essences (*esse essentiae*) was also considered Platonist (see Henry of Ghent). Meister Eckhart held Neoplatonic, Proclean-inspired theories of the nature of the soul, and his affirmation of an uncreated part of the soul parallels the doctrine of the undescended part of the soul in pagan Neoplatonism (see Neoplatonism). The German Dominican writers Dietrich of Freiberg and Berthold of Moosburg are more openly favourable to Plato and Proclus, with Berthold writing a commentary on Proclus' *Elements of Theology*.

Nicholas of Cusa appears to have introduced William of Moerbeke's Latin translation of the *Parmenides* to a medieval audience, and he was also familiar with Proclus' *Commentary on the Parmenides*. Nicholas developed a strongly Neoplatonic account of the nature of the divine being who so transcends and reconciles all oppositions as to be called the 'coincidence of opposites' (*coincidentia oppositorum*), echoing Eriugena's view of God as 'the opposite of opposites' (*oppositio oppositorum*). Drawing on the twelfth-century hermetic text *The Book of the Twenty-Four Philosophers*, Nicholas represents God as an infinite sphere whose centre is everywhere and whose circumference is nowhere. Elsewhere, for example in *De li non aliud (On the Not-Other)*, he develops the immanence and transcendence of God. In *De docta ignorantia (On Learned Ignorance)*, Nicholas expresses the Platonic view that forms or notions exist separately from the things of which they are the forms, rejecting Aristotle's criticisms as shallow misunderstandings. Here Nicholas follows Eriugena and Thierry of Chartres in calling God 'form of all' (*forma omnium*), 'form of being' (*forma essendi*) and 'form of forms' (*forma formarum*). For Nicholas, all forms exist as one in God but 'contractedly' in created things.

See also: Aristotelianism, medieval; Averroism; Chartres, School of; Gilbert of Poitiers; Medieval philosophy; Neoplatonism; Platonism in Islamic philosophy; Platonism, Early and Middle; Platonism, Renaissance; Pseudo-Dionysius

DERMOT MORAN

References and further reading

Aertsen, J. (1992) 'The Platonic Tendency of Thomism and the Foundations of Aquinas' Philosophy', *Medioevo* 18: 53-70.(A recent critical assessment of the impact of Platonism on Thomas.)

Alain of Lille (*c.*1120-1203) *De planctu naturae (The Lament of Nature)*, ed. N. Häring, *Studi Medievali* ser. 3, 19, 1978: 797-879.(Critical edition of Latin text only.)

Augustine (401-14) *De Genesi ad litteram (The Literal Commentary on Genesis)*, trans. J. Taylor, New York: Newman Press, 1982, 2 vols.(First full-length English translation of this hugely influential commentary on Genesis. An indispensable sourcebook of Augustinian ideas on creation, the seminal principles and the nature of man.)

Augustine (413-27) *De civitate Dei (The City of God)*, trans. H. Bettenson, Harmondsworth: Penguin, 1984. (English translation of Augustine's mature discussion of the Christian goal of creating a heavenly kingdom in the light of the difficulties of the earthly kingdom of the Roman Empire.)

Beierwaltes, W. (ed.) (1969) *Platonismus in der Philosophie des Mittelalters (Platonism in Medieval*

Philosophy), Darmstadt: Wissenschaftliche Buchgesellschaft. (A collection of important articles in German on the influence of Platonism in the Middle Ages, including the ground-breaking 1916 article by Clemens Baeumker, 'Der Platonismus im Mittelalter'.)

Beierwaltes, W. (1990) 'Eriugena's Platonism', *Hermathena* 149: 53-72.(Authoritative article seeking to define Eriugena's Platonism, arguing that Eriugena's Platonism is indirect, transmitted through theological authorities.)

Bernard of Chartres (*c*.1110-*c*.1125) *Glosae super Platonem (Glosses on Plato)*, ed. P.E. Dutton, *The Glosae super Platonem of Bernard of Chartres*, Toronto, Ont.: Pontifical Institute of Medieval Studies, 1991.(Contains critical text of the commentary on *Timaeus* with an excellent introduction in English on the nature of the commentary, with a discussion of Bernard's concept of native forms (*formae nativae*).)

Chenu, M.-D. (1968) *Nature, Man and Society in the Twelfth Century*, trans. J. Taylor and L.K. Little, Chicago, IL: University of Chicago Press.(Scholarly but readable account of the twelfth-century renaissance.)

Cherniss, H.F. (1930) *The Platonism of Gregory of Nyssa*, Berkeley, CA: University of California Press. (The standard account of the Platonism of the Cappadocian Father who strongly influenced the formation of Christian Platonism.)

Copenhaver, B.P. and Schmitt, C.B. (1992) 'Platonism', in B.P. Copenhaver (ed.) *Renaissance Philosophy*, Oxford: Oxford University Press, 127-95.(Excellent survey of the nature of Platonism in the Renaissance.)

Dronke, P. (1974) *Fabula: Explorations into the Uses of Myth in Medieval Platonism*, Leiden: Brill.(Good discussion of medieval understanding of Platonic allegory, the world soul and the cosmic egg.)

Dronke, P. (ed.) (1988) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press.(Indispensable. Contains detailed accounts of philosophical developments in the twelfth century, especially the chapters 'The Platonic Inheritance' by T. Gregory, 'Philosophy, Cosmology and the Twelfth-century Renaissance' by W. Wetherbee and 'Thierry of Chartres' by Dronke.)

Dutton, P.E. (1984) 'The Uncovering of the *Glosae super Platonem* of Bernard of Chartres', *Medieval Studies* 46: 192-221.(Discusses Dutton's reasons for attributing the twelfth-century commentary on the *Timaeus* to Bernard of Chartres.)

Eriugena, Johannes Scottus (*c*.867) *Periphyseon (On the Division of Nature)*, trans. I.-P. Sheldon-Williams, revised by J.J. O'Meara, Montreal: Bellarmin; Washington, DC: Dumbarton Oaks, 1987.(Complete English translation of the *Periphyseon*.)

Garin, E. (1955) 'Ricerche sulle traduzioni di Platone nella prima metà del sec. XV' (Research on the Platonic Tradition in the First Half of the Fifteenth Century), in *Medioevo e Rinascimento: Studi in onore di Bruno Nardi*, Florence.(Garin is the classic source for the study of the influence of Plato in the late medieval and Renaissance period.)

Garin, E. (1958) *Studi sul platonismo medievale (Studies in Medieval Platonism)*, Florence: Le Monnier. (Important scholarly research on Platonism.)

Gersh, S. (1986) *Middle Platonism and Neoplatonism: The Latin Tradition*, Notre Dame, IN: University of Notre Dame Press, 2 vols.(Well-documented, authoritative survey of the Latin Platonic tradition including Cicero, Apuleius, Calcidius, Macrobius, Martianus Capella and Boethius.)

Gibson, M. (1969) 'The Study of the *Timaeus* in the Eleventh and Twelfth Centuries', *Pensiamento* 25: 183-94. (Useful critical discussion on the influence of Calcidius on twelfth-century writers.)

Gregory, T. (1955) *Anima Mundi (World Soul)*, Florence.(Definitive study in Italian of the medieval tradition of the world soul.)

Gregory, T. (1958) *Platonismo medievale: studi e ricerchi (Medieval Platonism: Studies and Research)*, Rome: Instituto Storico Italiano per il Medio Evo.(One of the most important scholarly studies of medieval Platonism.)

Gregory, T. (1974) 'Abélard et Platon' (Abelard and Plato), in E.M. Buytaert (ed.) *Peter Abelard*, Louvain.(A study of Abelard's Platonism.)

Hankins, J. (1990) *Platonism in the Italian Renaissance*, Leiden: Brill, 2 vols.(Well-documented, up-to-date critical study of Renaissance Platonism.)

Haren, M. (1992) *Medieval Thought: The Western Intellectual Tradition from Antiquity to the Thirteenth Century*, 2nd edn, London: Macmillan.(Excellent, extremely readable survey of medieval intellectual history.)

Henle, R.J. (1956) *St. Thomas Aquinas and Platonism*, The Hague: Nijhoff.(Collects together with useful comments Thomas' references to Plato throughout his works. Texts are given in Latin only.)

Klibansky, R. (1982) *The Continuity of the Platonic Tradition during the Middle Ages*, London: The Warburg Institute; Millwood, NY: Kraus International Publications.(Re-issue of 1939 edition with supplementary chapters. The classic study on the transmission of Platonism in the Middle Ages focusing on the knowledge of the original texts and the history of their translation.)

Knowles, D. (1988) *The Evolution of Medieval Thought*, 2nd edn, London: Longman.(Reliable historical survey of medieval thought.)

Little, A. (1950) *The Platonic Heritage of Thomism*, Dublin: Golden Eagle Press. (Survey of Platonic influence on Thomas.)

Macrobius (probably *c*.430) *Commentary on the Dream of Scipio*, trans. W.H. Stahl, New York: Columbia University Press, 1952.(Edition with introduction and notes. Extremely important source of Platonic ideas, especially from ninth to the twelfth centuries.)

Martianus Capella (perhaps *c*.470) *De nuptiis Philologiae et Mercurii (The Marriage of Philology and Mercury)*, trans. W.H. Stahl, R. Johnson and E.L. Burge in *Martianus Capella and the Seven Liberal Arts*, vol. 2, New York: Columbia University Press, 1977.(Complete English translation of this sprawling commentary on all aspects of knowledge, divided into the seven liberal arts. The section on 'Dialectic' is an important compendium of philosophical ideas from diverse sources.)

McEvoy, J. (1982) *The Philosophy of Robert Grosseteste*, Oxford: Clarendon Press. (Comprehensive study of the philosophy of Grosseteste, emphasising the continuity between his scientific and theological interests, with an excellent discussion of light metaphysics.)

Nicholas of Cusa (1440) *De doctora ignorantia (On Learned Ignorance)*, ed. and trans. J. Hopkins, *Nicholas of Cusa on Learned Ignorance: A Translation and Appraisal of De Docta Ignorantia*, 2nd edn, Minneapolis, MN: Banning, 1985. (English translation with commentary of an important fifteenth-century text which develops a negative theological approach.)

Southern, R.W. (1970) 'Humanism and the School of Chartres', in R.W. Southern (ed.) *Medieval Humanism and Other Studies*, Oxford: Blackwell, 61-85.(The classical account of twelfth-century humanism, revised in Southern (1979).)

Southern, R.W. (1979) *Platonism, Scholastic Method and the School of Chartres*, Reading: University of Reading.(Important study which updates and revises Southern's earlier account of twelfth-century humanism, taking issue with the traditional view of a Chartres 'school'.)

Southern, R.W. (1982) 'The Schools of Paris and the School of Chartres', in R.L. Benson, G. Constable and C.D. Lanham (eds), *Renaissance and Renewal in the Twelfth Century*, Cambridge, MA: Harvard University Press, 113-37.(Critical reconsideration of the extent of the so-called School of Chartres.)

Spade, P.V. (ed.) (1994) *Five Texts on the Medieval Problem of Universals: Porphyry, Boethius, Abelard, Duns Scotus, Ockham*, Indianapolis, IN: Hackett.(An excellent selection of the key texts in the medieval debate on universals.)

Steel, C. (1990) 'Plato latinus', in J. Hamesse and M. Fattori (eds) *Rencontres de cultures dans la philosophie médiévale. Traductions et traducteurs de l'antiquité tardive au XIVe siècle. Actes du Colloque international de Cassino*, Louvain-La-Neuve: Institut d'Études Médiévales, 301-16.(Up-to-date, authoritative survey of the Latin tradition of Plato.)

Wetherbee, W. (1972) *Platonism and Poetry in the Twelfth Century*, Princeton, NJ: Princeton University Press. (Excellent on intellectual and literary background to twelfth-century poetry, especially Alain of Lille.)

William of Conches (*c*.1130) *Glosae super Platonem (Commentaries on Plato)*, ed. E. Jeauneau, Paris: Vrin, 1965.(William's glosses on the *Timaeus*.)

Platonism, Renaissance

Though it never successfully challenged the dominance of Aristotelian school philosophy, the revival of Plato and Platonism was an important phenomenon in the philosophical life of the Renaissance and contributed much to the new, more pluralistic philosophical climate of the fifteenth and sixteenth centuries. Medieval philosophers had had access only to a few works by Plato himself, and, while the indirect influence of the Platonic tradition was pervasive, few if any Western medieval philosophers identified themselves as Platonists. In the Renaissance, by contrast, Western thinkers had access to the complete corpus of Plato's works as well as to the works of Plotinus and many late ancient Platonists; there was also a small but influential group of thinkers who identified themselves as Christian Platonists. In the fifteenth century, the most important of these were to be found in the circles of Cardinal Bessarion (1403-72) in Rome and of Marsilio Ficino (1433-99) in Florence. Platonic themes were also central to the philosophies of Nicholas of Cusa (1401-64) and Giovanni Pico della Mirandola (1463-94), the two most powerful and original thinkers of the Quattrocento. While the dominant interpretation of the Platonic dialogues throughout the Renaissance remained Neoplatonic, there was also a minority tradition that revived the sceptical interpretation of the dialogues that had been characteristic of the early Hellenistic Academy.

In the sixteenth century Platonism became a kind of 'countercultural' phenomenon, and Plato came to be an important authority for scientists and cosmologists who wished to challenge the Aristotelian mainstream: men like Copernicus, Giordano Bruno, Francesco Patrizi and Galileo. Nevertheless, the Platonic dialogues were rarely taught in the humanistic schools of fifteenth-century Italy. Plato was first established as an important school author in the sixteenth century, first at the University of Paris and later in German universities. In Italy chairs of Platonic philosophy began to be established for the first time in the 1570s. Though the hegemony of Aristotelianism was in the end broken by the new philosophy of the seventeenth century, Plato's authority did much to loosen the grip of Aristotle on the teaching of natural philosophy in the universities of late Renaissance Europe.

1 The revival of Plato

'More men praise Aristotle; greater men, Plato.' So wrote Petrarch (see Petrarca, F. §2) around 1368 in the invective *De sui ipsius et multorum ignorantia (On His Own Ignorance and That of Many Others)* (completed in 1371). Petrarch's opinion came to be representative of the humanist movement in general during the Renaissance (see Humanism, Renaissance). The humanists deplored the extent to which Aristotle dominated the philosophical curricula of contemporary universities. They knew from their classical reading that the ancients, both pagan and Christian, had considered Plato to be the greatest of philosophers. They also knew that the Byzantines, their schoolmasters in matters Greek, had regarded the study of Aristotle's writings as propaedeutic to that of the 'higher', more sublime philosophy of Plato. The neglect of Plato thus became for humanists a standard part of their polemic against scholasticism's narrowness as an educational programme. The humanists, like their idol Cicero, wished to see the educated elites of Renaissance Italy introduced to the full range of ancient philosophy. Hence the humanist project to restore the educational values of classical antiquity brought about the revival of Platonic studies in the early Renaissance.

The study of Plato had not been entirely neglected in the Middle Ages (see Platonism, medieval). Medieval readers had inherited from antiquity two partial versions of the *Timaeus* made, respectively, by Cicero and Calcidius, and Calcidius' rendering, accompanied by his commentary, had been among the most widely-studied texts of the twelfth century. Henricus Aristippus in the mid-twelfth century had translated the *Phaedo* and the *Meno*, and around 1280 Aquinas' translator, the Dominican William of Moerbeke, translated Proclus' commentary on the *Parmenides*, which included much of Plato's text in the lemmata. Much was also known indirectly of Plato's writings via the Latin philosophical tradition, especially Cicero, Apuleius, Macrobius and Augustine, as well as through Avicenna and other Arabic philosophers. The Latin Aristotle, too, was important for the medieval knowledge of Plato; the arguments of the *platonici* in favour of metaphysical realism frequently cited by scholastic philosophers came mostly from this source.

Still, the Italian humanists in the course of a single century did far more to make Plato available in Latin than ten centuries of Western medieval scholarship had done. Leonardo Bruni (see Aristotelianism, Renaissance §7;

Humanism, Renaissance §§5, 7), Uberto Decembrio and Cencio de'Rustici, all pupils of the émigré Greek Manuel Chrysoloras (d. 1415), translated among them some ten dialogues, including the *Gorgias*, *Crito*, *Phaedo*, *Apology* and *Republic*. In the following generation the *Republic* was translated twice more, by Uberto Decembrio's son Pier Candido and by the Sicilian Antonio Cassarino. The Milanese humanist Francesco Filelfo translated the *Euthyphro* and some of the *Letters*, while in Rome the papal secretary Rinuccio Aretino rendered the *Crito*, *Euthyphro* and *Axiochus*. George of Trebizond, a humanist and papal secretary from Venetian Crete, turned the *Laws*, *Epinomis* and *Parmenides* into Latin. In 1462 Pietro Balbi, a protégé of Nicholas of Cusa and Bessarion, translated Proclus' *Platonic Theology*. In Florence, Lorenzo Lippi da Colle, a neo-Latin poet and protégé of Piero de'Medici, translated the *Ion*. The translation activity of the humanists culminated in the work of Marsilio Ficino, who in 1484 published the first complete Latin version of Plato's works and in 1496 published a series of annotations and commentaries on the major dialogues. Ficino also published in Latin the complete works of Plotinus and a number of works by minor ancient and Byzantine Platonists. So while in 1400 the Latin world possessed only two complete and two partial versions of the Platonic dialogues and a developed commentary tradition only on the *Timaeus*, by the end of the fifteenth century it had available in Latin all thirty-six dialogues of the Thrasyllan canon, the pseudo-Platonic *Definitions* and *Halcyon*, three of the six *Spuria*, and a large and sophisticated body of ancient and modern commentaries and other aids to the study of Plato's texts.

The progress of Platonic studies continued in the sixteenth century. Ficino's *Platonis opera* in Latin was revised three times (1532, 1557 and 1592) and there were two fresh Latin versions of the complete works, those of Janus Cornarius (1561) and Jean de Serre (1578). The *editio princeps* of Plato in Greek, edited by Marcus Musurus, was published by Aldus in 1513. It was followed by the major recensions of Simon Grynaeus (1534), Marcus Hopper (1556) and Henri Estienne (1578). The latter edition remained standard until the early nineteenth century. The trickle of vernacular translations of Plato that began in the fifteenth century turned into a flood in the sixteenth, though the first complete edition of Plato's works in a vernacular language was Dardi Bembo's Italian version published in 1601.

Renaissance scholarship on the text of Plato was not always accompanied by an understanding of Platonic thought. Before the time of Bessarion and Ficino humanist translators commonly had a very imperfect grasp of the philosophical arguments and doctrines they were endeavouring to translate. Since they were eager for Plato to be accepted in the Christian world, it was common for them to bowdlerize (especially passages alluding to homosexuality) and Christianize (especially references to polytheism). After the middle of the Quattrocento, however, Renaissance humanist scholars and Platonists largely accepted the guidance of the ancient Neoplatonic tradition in interpreting the works of Plato. Thanks largely to Cardinal Bessarion's *In calumniatorem Platonis (Against Plato's Calumniator)* (1469) and Ficino's commentaries on Plato, Neoplatonic interpretations of Plato dominated the later Renaissance and continued to find followers in the early modern period, down to Thomas Taylor and Samuel Taylor Coleridge in the later eighteenth century. Despite the dominance of the Neoplatonic Plato, however, there was from the beginning of the Renaissance a minority tradition of interpreting Plato. In antiquity itself the theological Plato of the Neoplatonists had been balanced by the sceptical Plato of the New Academy. In Cicero and in Augustine's *Contra Academicos (Against the Academics)* Renaissance humanists met another Plato, a Plato without positive doctrines of his own, an ironic and playful doubter who liked to disguise his own opinions while presenting all points of view in inconclusive dialogues. One can find traces of this sceptical Plato in many Renaissance writers, from Petrarch (see Petrarca, F.) and Leonardo Bruni in the early Renaissance to Melanchthon, Nizolius, Ramus and Montaigne in the sixteenth century. Yet only rarely do those who view Plato as a sceptic work out their interpretation in detailed readings of the dialogues themselves. The most famous, indeed almost the only example of this, is Giovanni Pico della Mirandola's reading of the *Parmenides* as a dialectical display, an interpretation which explicitly challenged Ficino's theological reading (see Pico della Mirandola, G. §2).

In the sixteenth century Renaissance interpreters of Plato began to move away from these traditional alternatives. Although the credit for distinguishing between Plato's original doctrine and that of his Neoplatonic interpreters is often given to Friedrich Schleiermacher in his 1804 edition of Plato, in fact the distinction has a long prehistory in Renaissance and early modern scholarship, going back to Jean de Serre's introduction to the Stephanus edition (1578) of Plato. The Protestant reformers were highly suspicious of rationalistic theologies in general and attempts to combine pagan philosophy with Christianity in particular. They were as hostile to Ficino's attempt to base

Christian theology on Plato as they were to the Aristotelian theology of the scholastics. Protestant scholars who wanted to save Plato from the Florentine Platonists declared either that he should be read purely as literature, laying aside his theological views, or, like Jean de Serre, that his true philosophy needed to be cleansed of 'les gloses des Platoniciens'. Although de Serre's own attempt to reconstruct Plato's original philosophical views was not impressive, the principle he enunciated allowed later interpreters such as Isaac Vossius, Leibniz, and Jakob Brucker (author of the first critical history of philosophy), to shake off the Neoplatonic interpretation of Plato.

2 Renaissance anti-Platonism

The virulence of the opposition to the Platonic revival explains much about the character of that revival and about the approaches to Plato characteristic of Renaissance Platonists. Though Aristotle had been successfully domesticated in European universities during the thirteenth century (see Aristotelianism, medieval §4), this circumstance did little to smooth the path for Plato in the fifteenth. The attempt to revive the Platonic texts during the early Renaissance was opposed by a number of authoritative figures such as the Dominican reformer Cardinal Giovanni Dominici, St Antoninus (the Archbishop of Florence), Pope Paul II and the famous preacher Girolamo Savonarola. Scholastic opponents of Plato in the sixteenth and seventeenth centuries repeated the same criticisms over and over again. But criticism of Plato was not confined to scholastics. A number of humanists and patrons of the humanities also had reservations about renewing the study of Plato. Indeed, Plato's chief critic in the fifteenth century was George of Trebizond, a humanist and papal secretary, whose *Comparatio philosophorum Aristotelis et Platonis (Comparison of the Philosophers Aristotle and Plato)* (1458; printed in 1523 as *Comparationes philosophorum Aristotelis et Platonis*) constituted a compendium of Renaissance anti-Platonism. Trebizond's arguments presented a real challenge to the Platonic revival. The humanist movement had battened on the promise that the recovery of antiquity would mean the recovery of sound moral doctrine and brilliant examples of virtue. Plato represented a threat to this humanist project on several fronts.

The criticisms may be reduced to three. The first was the charge that Plato's teaching was unsystematic and therefore pedagogically useless. Critics complained that his doctrine was too obscure, being hidden under the personae of interlocutors who contradicted each other. Socratic dialogue was considered far more difficult to follow than the Ciceronian dialogues popular among the humanists. At the same time Plato, unlike Aristotle, had neither divided his works systematically by subject matter nor adapted them to the understanding of beginners. As a result, it was difficult for teachers to lecture on Plato's writings in the painstaking, line-by-line fashion employed in Renaissance schools and universities, and it was difficult for students to memorize and retain his doctrine. The second and more serious set of charges focused on the moral deficiencies of the dialogues. Most notorious were the passages in Book V of the *Republic* where Plato has Socrates advocate a system of common ownership of women, children and goods; abortion; female military service; and other doctrines. Criticism of Platonic communism had a long history, from Aristotle himself to Lactantius, Jerome and Gratian's *Decretum* (the most popular medieval textbook of church law), and the doctrine had been routinely condemned in the Middle Ages, first by canon lawyers and later by scholastic commentators on *Politics* II. In the Renaissance, however, humanists through their translations gave direct access to the text of Plato and so revived the storm of disapprobation. Plato's critics declared that any philosopher who promoted such doctrines should not be read in schools. Other passages caused different kinds of trouble. Dialogues such as the *Lysis*, *Charmides* and *Symposium*, depicting scenes of homosexual gallantry, caused grave offence. It was difficult to understand why Plato had spent the first two books of the *Laws* arguing for the value of drunkenness in moral education. And in the *Laws* as in other dialogues Plato gave unequivocal approval to the worship of the gods of the Greek Pantheon - gods whom the Fathers of the Church had identified as demons. Plato even admitted that Socrates had lived his life in obedience to a demon.

This leads to the third group of charges against Plato: that his theological views were incompatible with Christian truth. The humanists, quoting a famous passage in Augustine's *De civitate Dei (The City of God)*, had argued that Plato's belief in individual immortality and creation made his theology closer to Christianity than Aristotle's (see Augustine §8). Plato's critics replied that, whatever his merits as a theologian, they were outweighed by his defects. They attacked his heterodox views on the pre-existence and transmigration of souls. They noted that, even if Plato had believed in creation, he had not believed in creation *ex nihilo*; in the *Timaeus* it seemed that the 'receptacle' (or 'prime matter', as it was called by Renaissance interpreters) was already in being at the moment of creation. Other critics believed that the subordinationist architecture of Plato's intelligible cosmos (as explicated

by later Platonists) had encouraged the Arian heresies of the fourth century and was therefore dangerous. Though there were similar theological problems with the reception of Aristotle, Plato's critics pointed out that, in the case of Aristotle, it was possible to separate the unorthodox parts from the useful parts. If his *Metaphysics* and On the Soul were dangerous, they could be omitted or read only after suitable formation. The student could benefit from the Organon or *On Generation and Corruption*, where logic and physical questions were dealt with (it seemed) in isolation from theology. But in the case of Plato, this could not be done. The immortality proofs of the *Phaedo* and *Meno* were mixed up hopelessly with the doctrine of pre-existence. The myth of creation in the *Timaeus* was inseparable from passages that appeared to posit the eternity of unformed matter. The refutation of atheism in *Laws* X was advanced in support of a system of polytheistic cultus. Socrates' description of the ascent to the form of Beauty in the *Symposium* is made among the drunkenness and disordered passions of a pagan dining party.

The advocates of Plato did their best to protect him. As has been noted, many humanist translators bowdlerized obscure or unorthodox passages in their translations. Pier Candido Decembrio employed an elaborate system of arguments, prefaces, glosses and chapter headings in his translation of the *Republic* to give the impression that the work was a systematic treatise on politics and not the welter of obscure and multidirectional argument it appeared to be at first sight. Later Quattrocento interpreters of Plato had more sophisticated ways of defending Plato. The pagan Platonists of late antiquity, in rivalry with Christian theologians of the time, had converted the Neoplatonism of Plotinus into a systematic theology and had developed elaborate hermeneutical methods designed to educe this theology from the dialogues of Plato. The hermeneutics of the later Platonists, especially Proclus, were taken over by Bessarion and Ficino and applied anew to Plato's works. The technique of allegory was particularly useful in dealing with offensive or heterodox passages. Certainly, any fifteenth-century reader who had wanted to understand the dialogues of Plato in depth would naturally have resorted to the late ancient commentators on Plato simply because they were among the few available guides to his thought. But the recovery of the Neoplatonic interpretation of Plato was also driven by the need to protect him from the criticism of contemporaries.

3 Cardinal Bessarion and the Roman academy

The interest of the popes and the papal court in Plato goes back to the very beginnings of the Platonic revival, when Bruni dedicated his translation of the *Phaedo* to Innocent VII in 1405. A fair proportion of the early humanist translations of Plato can be associated with Roman cardinals or with the papal court, and about a quarter of the approximately 350 manuscripts containing these versions are of Roman provenance. The correspondence of Roman humanists in the first half of the Quattrocento is filled with references to the opinions and sayings of Plato. In this period, the interest of the Roman curia in Plato, reflecting its devotion to humanist ideals, was mostly literary, moralistic and apologetic. From Cicero, Seneca and other Latin sources the Romans had become curious about the figure of Socrates as a model of virtuous conduct. They were delighted to find confirmation of Christian truths in Plato's eschatological myths and in his proofs of the immortality of the soul. They were attracted by Plato's ability to place his literary skill at the service of moral and religious lessons.

With the Council of Ferrara/Florence in 1438-9, however, there was a decisive shift in interest towards Plato's cosmology, his mathematical physics and particularly his psychology and theology. There were two principal reasons for what might be called the 'metaphysical turn' in the reception of Plato. The first was the presence at the Council of the Greek philosopher Gemistos (who later changed his name, rather mysteriously, to Pletho). The Council of Ferrara/Florence had been called in an attempt to put an end to the schism between the Greek and Latin churches, and Pletho, among the most learned men of the Greek world, had been taken along by the Greek delegation as a theological adviser. Though Pletho was later accused of paganism by the conservative cleric George Scholarios (later to become the Orthodox Patriarch Gennadios) as well as by the equally conservative Roman convert, George of Trebizond, it is unclear whether he was really a pagan or rather an extreme syncretist who combined Orthodox Christianity with other forms of religious wisdom, including Platonism. Platonism, which Pletho referred to as 'the Hellenic theology', was in any case not to be identified with the 'vulgar' ancient polytheism of myth and cult, but was an esoteric religious wisdom known to the better minds among the ancient pagans. The key to understanding this hidden wisdom for Pletho was the *Platonic Theology* of Proclus, which Pletho saw as a kind of *summa* or systematic treatment of the Platonic dialogues. Pletho believed himself to have prophetic powers and predicted that within a short time Christianity and Islam would disappear, to be replaced by a new, purified Hellenic theology based on Plato.

The paganizing and prophetic sides of Pletho's thought, however, remained hidden at the time of the Council. During the Council Pletho attracted attention chiefly for his lectures, later published under the title *De differentiis Platonis et Aristotelis (On the Differences between Plato and Aristotle)* (1439). These were attended by many Westerners, who thus received their first initiation into the Neoplatonic interpretation of Plato. They were intended to illustrate two main themes: the superiority of Plato to Aristotle in terms of their relative compatibility with Christianity, and the basic concord of Plato with Aristotle.

These themes had considerable resonance at the Council. The council's purpose was to reconcile the theology of the Latins, whose speculative development was based largely on Aristotelian logic and natural science, with the theology of the Greeks, heavily indebted to the form of late Platonism associated with Pseudo-Dionysius the Areopagite. To the Byzantine cleric Bessarion (who in 1439 was created a cardinal of the Roman church) Platonism seemed to offer a formula for reconciling the two traditions. Ancient Platonists had believed that there existed beyond the multiplicity of ordinary sense-experience radical unities in the cosmos, and that these unities were cognizable only in a heightened state of awareness called *noesis*. Since language was based on sense-experience, it was inadequate for representing the radical unities of the divine realm. The discursive reasoning and dogmatic definitions of the Latins were thus inappropriate tools for theology. Platonism, regarding dogmas, liturgical forms and creeds as inadequate approximations or symbols of inexpressible truths, could thus provide the basis for a theology of concord between Greeks and Latins. Both Greek and Latin creeds could be true, even if they seemed opposed from the viewpoint of discursive reasoning. Bessarion used this conception to convince the Greeks that they were not abandoning Orthodoxy by acceding to Latin liturgical practices and Latin formulations of the creed. The belief that a theology of concord could be built upon Platonic foundations proved a durable one. It was responsible for much of Platonism's popularity in the dawning age of religious controversy, while Aristotelian theologies were frequently blamed for rigidity and dogmatism. Platonic concordism was an important element in the philosophical development of Nicholas of Cusa, whose famous treatise *De docta ignorantia (On Learned Ignorance)* (1440) grew out of reflection on the theological issues of the Council.

Following the Council there arose a dispute among the intellectuals of the émigré Greek colony in Rome whose headquarters was Bessarion's palace library or 'academy'. The dispute lasted for thirty years. At issue was whether Platonism or Aristotelianism formed a better speculative framework for Christian theology. Behind this issue lay the even more charged question of whether and how the Byzantine theological tradition might be integrated with the Latin. This 'Plato-Aristotle Controversy', as it is called, culminated in George of Trebizond's *Comparatio philosophorum Aristotelis et Platonis* (1458) and Bessarion's reply, *In calumniatorem Platonis* (drafted 1459, printed 1469). The aim of George's work was to show the total incompatibility of Plato (and by implication, the 'corrupt' traditions of Byzantine theology) with the 'pure' Aristotelian theologies of the Latin church. Bessarion's *In calumniatorem* in Books I, III and IV consisted of a point-by-point refutation of George's *Comparatio*, while Book V sought to discredit George by attacking his translation of Plato's *Laws*. Book II provided a summary of Platonic theology, based on the Greek Church Fathers and on late ancient Platonists, particularly Proclus. In addition to introducing the Latin West to the late ancient and Byzantine view of Plato's theology, Bessarion's work made clear for the first time to Western scholars the degree to which the theological tradition of ancient Greek Christianity had been indebted to Platonism. The publication of his *In calumniatorem* coincided with the beginnings of the Renaissance revival of the Greek Fathers, and the latter did much to increase the popularity of Platonism in the later fifteenth and sixteenth century.

4 The Platonism of the Florentine school

Florentine humanists had been interested in Plato from the late fourteenth century, when Coluccio Salutati had acquired copies of Calcidius and of Henricus Aristippus' translation of the *Phaedo*. The interest continued in the early fifteenth century with Leonardo Bruni (see Aristotelianism, Renaissance §7; Humanism, Renaissance §§5, 7), who dedicated two of his seven translations of Plato to Cosimo de'Medici, the great banker, politician and patron of arts and letters. The Medici over several generations played a key role in the revival of Plato in Florence. Cosimo de'Medici is said to have heard Pletho's lectures during the Council of Florence, and it was probably from Pletho that Cosimo purchased the manuscript of Plato's dialogues he later gave to Marsilio Ficino to serve as a basis for his translation. In 1462 Cosimo gave Ficino a house in Florence and a farm in the countryside near his own villa of Careggi. Cosimo's grandson, Lorenzo 'the Magnificent', continued to support Ficino by acquiring

several benefices for him and paying the printing costs of some of his scholarly works. Ficino had other patrons besides the Medici, and the Medici supported philosophers of many other schools besides Ficinian Platonism, but by the early sixteenth century the Medici had acquired a reputation for special devotion to Plato. In 1513, the *editio princeps* of Plato's works in Greek was dedicated to the Medici pope, Leo X, whose father Lorenzo had received the dedication of Ficino's Latin translation of Plato thirty years earlier.

Ficino and Cristoforo Landino were the two central figures in the Florentine school of Platonism. Landino was a professor of rhetoric and Latin literature at the University of Florence who gave Platonizing interpretations of various literary works, most famously Virgil's *Aeneid* and Dante's *Divina Commedia (Divine Comedy)*. Landino also composed the *Disputationes Camaldulenses (Camaldulensian Disputations)* (1472), a brilliant depiction of Lorenzo's entourage engaged in philosophical discussions about the immortality of the soul. These dialogues show better than any other source the character of Florentine Platonism during the 1470s. Marsilio Ficino, one of the greatest scholars of the century, was only briefly a professor at the University; he kept instead a private *gymnasium* or *academy* where he educated numerous members of the Florentine elite. He also gave public lectures on Plato and Plotinus and maintained a large philosophical correspondence. He translated all the works of Plato and Plotinus in addition to those of a number of other ancient Neoplatonists, and wrote a major work amassing arguments to prove the immortality of the soul. Besides Ficino, Landino and their pupils, it is hard to find other important intellectuals in the Florence of Lorenzo de'Medici who were deeply influenced by Platonism. Johannes Argyropoulos, while lecturing at the University of Florence, introduced Platonic elements into his explications of Aristotle. Angelo Poliziano came briefly under Ficino's spell during the mid-1470s, when he produced his Platonizing *Stanze per la giostra (Stanzas for the Tournament)* and an unfinished translation of the *Charmides*. The astronomer Lorenzo Bonincontri, the humanist Bartolomeo Scala and the philosopher Giovanni Pico della Mirandola all made eclectic use of Ficinian Neoplatonism. Most important of all, from the point of view of spreading interest in Platonism, was Lorenzo de'Medici himself, whose Italian poetry frequently contained themes, images and conceits borrowed from Ficino's writings.

While the chief theme of Bessarion's Platonism was the integration of Greek and Latin theological traditions, Ficino's Platonism was broader and more polymorphous in its interests. Ficino considered his main role to be that of a theological reformer. He wanted to restore Platonism to its ancient function as the handmaid of Christian theology, replacing the failed Aristotelian syntheses of Christianity and Aristotle that had characterized the medieval period. His chief concern as a philosopher was establishing a rational basis for the doctrine of the immortality of the soul, the key issue separating 'secularizing' from 'Christianizing' Aristotelians in Italian universities of the Renaissance. But Ficino and his followers were also interested in using the truths of Platonism to explore many other subjects, including astrology, magic, medicine, literary theory, musical theory and the theory of love.

With the fall of the Medici in 1494 and the coming of Savonarola, Ficinian Platonism was temporarily eclipsed, and many of his followers became Savonarolans. In the sixteenth century, interest in Ficino's work revived and his Platonism exerted a broad cultural influence. Especially popular was Ficino's 'philosophy of love', which provided European poets in the Petrarchan tradition with a new repertory of conceits and images. It is nevertheless difficult to find professional philosophers who could be called Ficinian Platonists. In Florence, Francesco Cattani da Diacceto continued Ficino's work during the first Medici Restoration, but disagreed with him on many points, preferring the concordism of Pico. Later in the century Francesco de'Vieri the Younger attempted another revival of Ficinian Platonism. In Rome, Ficino's most important epigone was the Augustinian Giles of Viterbo, a preacher and theologian of the papal court whose theology is thought to have influenced the programmes of Michelangelo's decoration of the Sistine Chapel ceiling and Raphael's frescos in the papal *Stanze*. The definition of the immortality of the soul as an official Catholic dogma by the Fifth Lateran Council (1512-7) may be partly credited to Giles' influence and to that of Ficino's former pupil, Pope Leo X. Ficino's studies of magic and the occult were appropriated by figures as diverse as Symphorien Champier, Agrippa von Nettesheim, Paracelsus and Giordano Bruno. Ficino's belief in an 'ancient theology' (see Ficino, M. §2) remained influential, even after the *Corpus Hermeticum* was exposed as a forgery by the Protestant scholar Isaac Casaubon in the early seventeenth century. Finally, Ficino's contention that Plato's philosophy was more compatible with Christian theology than Aristotle's continued to find numerous champions among both humanists and scholastics for more than a century.

5 Plato in humanist schools and in universities

Despite the intellectual energies expended on the Platonic revival of the fifteenth century, it was a long time before the dialogues of Plato were included regularly in the curricula of Renaissance schools and universities. Manuel Chrysoloras, the Byzantine émigré who taught the first generation of Italian humanists to learn Greek, seems to have read the *Gorgias* and the *Republic* privately with some of his pupils. There is some evidence that the famous humanist schoolmasters Guarino Veronese and Gasparino Barzizza used Chrysoloras' Latin translation of the *Republic* in private tutorials with certain students. George of Trebizond reports having heard lectures on the *Gorgias* in the school of another famous schoolmaster, Vittorino da Feltre, presumably in the 1420s, and the Byzantine Theodore Gaza, who had been an assistant to Vittorino, lectured on the same text at the University of Ferrara in 1446. In Battista Guarino's educational tract, *De ordine docendi et studendi (A Plan of Teaching and Study)* (1456) - a work said to capture the educational practice of Guarino Veronese - the reading of Plato and Aristotle is recommended as the crown of a long programme of humanistic studies, but it is not known how often, if ever, the full programme was carried out. In 1459 another Byzantine, Johannes Argyropoulos, is said to have lectured privately in Florence on Plato's *Meno*. Ficino gave public lectures on Plato's *Philebus*, Plotinus' *Enneads*, and no doubt other texts as well, but these lectures were probably not part of any school or university course of study.

In the 1490s the study of Plato spread to Northern Europe with the lectures of Paulus Niavis (Paul Schneevogel) on the pseudo-Platonic *Epistulae (Letters)* and *Erastes* (in Ficino's translations) at the University of Leipzig. From the first decade of the sixteenth century until the 1580s there was considerable interest in Plato among humanistically inclined masters at the University of Paris and particularly at the Collège Royale. A number of famous French classical scholars such as Jérôme Aleander, Denys Lambin, Adrien Turnèbe and Marc-Antoine Muret are known to have lectured on a wide variety of texts. Most of these texts were chosen from among shorter and easier dialogues of interest to students of literature, such as the *Crito*, *Apology*, *Hipparchus*, *Minos*, *Theages*, *Euthyphro*, *Cratylus*, Book I of the *Republic*, Socrates' speech from the *Symposium*, and the pseudo-Platonic *Epistulae*. Turnèbe was rather more ambitious, giving courses on the *Timaeus* and *Phaedo*, among other texts. Petrus Ramus used Plato as a stick with which to beat Aristotle in his famous anti-Peripatetic lectures, and encouraged the study of Plato by making his own translation of the pseudo-Platonic *Epistulae*. It was in part Ramus' influence that led to a renewed interest in Plato in German *gymnasia* of the late sixteenth century. The reformer and humanist Johannes Sturm had lectured on the *Gorgias* in Strasbourg as early as the 1540s, but in the 1580s and 1590s there were schools in Cologne, Nuremberg, Rostock and Strasbourg where students could study the *Apology*, *Gorgias*, *Theaetetus*, *Phaedrus* and *Republic*.

With the exception of Ficino's lectures almost all of this study took place in humanist schools or as a result of university lecture courses given by humanists. The emphasis was mostly on grammar and on providing general moral and cultural formation. Scholastic philosophers for a long time did not lecture on Plato nor make him the subject of their professional attentions; Pico's proposal to defend Platonic theses in 1487, had it taken place, would have been the first known instance of a scholastic disputation involving Platonic texts (see Pico della Mirandola, G. §1). This began to change at the end of the fifteenth century when in 1497 Niccolò Leonico Tomeo (1456-1531) was appointed to lecture on the Greek text of Aristotle and Plato at the University of Padua. The philosopher and magician Henricus Cornelius Agrippa von Nettesheim (1486-1535), too, is known to have lectured on the *Symposium*, either at Pavia or Turin, in the second decade of the sixteenth century. Yet it was only in the 1570s that professional philosophers began to teach Plato in the universities with any regularity. Special chairs for the teaching of Platonic philosophy were established in several Italian universities. Most important was the chair set up at Pisa by the Medici grand dukes, who saw themselves as continuing the historic commitment of the Medici to Platonic philosophy. The chair was occupied for half a century by four professors, Francesco de'Vieri the Younger (appointed 1576), Jacopo Mazzoni (1588-97), Carlo Tommasi da Cortona (*c*.1597-1606), and Cosimo Boscagli (d. 1621). Francesco Patrizi da Cherso (1529-97), the greatest expert on Plato of the late Renaissance, was for thirteen years (1578-92) the occupant of a special chair of Platonic philosophy at the University of Ferrara; in 1592 he moved to the Sapienza in Rome where he taught Plato until his death. There was also an unsuccessful attempt to establish a chair in Platonic philosophy at the University of Bologna. All the 'Platonic professors' were serious students of Plato's philosophy who lectured extensively on the text of Plato in Latin, the *Timaeus*, *Republic* and *Phaedo* being special favourites.

6 Plato and the new cosmologies

The introduction of Plato into the philosophy faculties of Italian universities was one of several signals that the hegemony of Aristotelian philosophy was weakening. The Plato chairs were founded in the belief that his philosophy was more compatible with Christianity than Aristotle's; Patrizi promised Pope Gregory XIV that replacing Aristotle with Plato in the universities would purify Catholic theology and put an end to the Protestant schism. Claims such as these revived the interest of philosophers in comparing Plato and Aristotle. There was already by the middle of the sixteenth century a flourishing literature on this topic. A spectrum of opinion emerged, ranging from Francesco Patrizi's view, in his *Discussiones peripateticae (Peripatetic Discussions)* (1581), that Plato was superior at all points to Aristotle, to the opposite view, maintained by Aristotelian scholastics such as Girolamo Borro (one of Galileo's teachers) in his *De peripatetica docendi atque discendi methodo (On the Peripatetic Method of Teaching and Learning)* (1584). Somewhere in the middle were Francesco de'Vieri the Younger and Jacopo Mazzoni, who, following what had by then become the traditional Florentine position, argued for the concord of Plato and Aristotle and for the harmony of both philosophers with Christian theology.

The search for a more adequate philosophical basis for Christian theology led ultimately to the new science and philosophy of the seventeenth century, and Renaissance Platonism had a significant role to play in this development. The 'new cosmologists' of the late sixteenth century frequently invoked Plato as a counter-authority when attacking Aristotelian science - even men such as Telesio, Campanella and Galileo, who otherwise made little use of Platonism in their science. Sometimes Galileo, when trying to conceal the shameful novelty of his science, would pretend to be reviving a lost, pre-Aristotelian cosmology - an apologetic strategy that recalls Ficino's belief in an 'ancient theology'. There were in fact some striking similarities between the new science and Platonic science, the latter reconstructed (sometimes tendentiously) from the *Timaeus* and other dialogues. Already in the 1540s Copernicus had been encouraged by the symbolic role of the sun in Platonic metaphysics (as mediated by Ficino's *De sole*) to embrace the heliocentric hypothesis. Heliocentrism remained linked to Neoplatonism as late as the time of Bruno and Kepler. In the late sixteenth century other elements of Platonic science appealed. Plato was credited with a simplified elemental theory which by excluding quintessence undermined the physical basis of the two-sphere universe. The light-metaphysics of Patrizi and Bruno, an inheritance from Proclus and the pseudo-Dionysius, has been considered a forerunner of Galileo's distinction between primary and secondary qualities. Giordano Bruno's belief in infinite worlds was taken from Proclus via Nicholas of Cusa. Proclus was also an important source for Patrizi's rather different, more mathematical argument for the infinity of the cosmos. The new cosmologists moreover professed to find in the *Timaeus* authority for their non-Aristotelian concepts of gravity, time and space, as well as for the 'mathematization of the cosmos' that begins with the generation of Patrizi and Galileo. Though it was ultimately the new science of Galileo and Newton that put an end to the dominance of Aristotelian science, the 'countercultural' influence of Plato was important in loosening the allegiance of Renaissance natural science to its medieval past.

See also: Ficino, M.; Humanism, Renaissance; Patrizi da Cherso, F.; Pico della Mirandola, G.; Plato; Platonism, Early and Middle; Platonism in Islamic philosophy; Platonism, medieval

JAMES HANKINS

References and further reading

Bessarion (1469) *In calumniatorem Platonis (Against Plato's Calumniator)*, in L. Mohler (ed.) *Kardinal Bessarion als Theologe, Humanist und Staatsmann*, vol. 2, Paderborn and Berlin: Schoningh, 1923-42; repr. Aalen: Scientia, 1967.(Referred to in §§1-3. Cardinal Bessarion's defence of Plato against the criticisms of George of Trebizond.)

Copenhaver, B.P. and Schmitt, C.B. (1992) *Renaissance Philosophy*, Oxford: Oxford University Press.(A clear and reliable introductory textbook.)

George of Trebizond (1458) *Comparationes philosophorum Aristotelis et Platonis (Comparisons of the Philosophers Aristotle and Plato)*, Venice: Iacobus Pentius, 1523; repr. Frankfurt am Main: Minerva, 1965. (George's famous attack on Plato; referred to in §§2-3. The manuscript version of this work (published in 1458) was entitled *Comparatio philosophorum Aristotelis et Platonis*.)

Hankins, J. (1990) *Plato in the Italian Renaissance*, London, Leiden and Copenhagen: Brill, 2 vols.(A detailed account of the reception of Plato in the fifteenth century.)

Kraye, J. (1993) 'The Philosophy of the Italian Renaissance', in G.H.R. Parkinson (ed.) *Routledge History of Philosophy, volume IV: The Renaissance and Seventeenth-century Rationalism*, London and New York:

Routledge, 16-69, esp. 26-37.(A short, accessible overview with full bibliographies.)
Kristeller, P.O. (1956) *Studies in Renaissance Thought and Letters*, Rome: Edizioni di Storia e Letteratura. (Important for the context of Ficino and for the later influence of Ficinian Platonism.)
Kristeller, P.O. (1964) *Eight Philosophers of the Italian Renaissance*, Stanford, CA: Stanford University Press.(A readable introduction to the philosophy of Ficino, Pico, Bruno and Patrizi among others.)
Kristeller, P.O. (1979) *Renaissance Thought and Its Sources*, New York: Columbia University Press.(Contains a masterly overview of key themes in Renaissance philosophy with special attention to Platonism.)
Landino, C. (1472) *Disputationes Camaldulenses (Camaldulensian Disputations)*, ed. P. Lohe, Florence: Sansoni, 1980.(Referred to in §4. Imaginary conversations on Platonic themes between Lorenzo de'Medici and members of his circle.)
Nicholas of Cusa (1440) *De docta ignorantia*, ed. E. Hoffmann and R. Klibansky, Leipzig: Meiner, 1932; trans. J. Hopkins, *Nicholas of Cusa on Learned Ignorance: A Translation and Appraisal of De Docta Ignorantia*, Minneapolis, MN: Arthur J. Banning Press, 2nd edn, 1985.(Referred to in §4. Platonic paradoxes about the knowledge of God.)
Patrizi da Cherso, F. (1581) *Discussionum peripateticarum tomi quattuor (Peripatetic Discussions in Four Books)*, Basle.(Referred to in §6. On the limits of Aristotelianism as a basis for Christian theology.)
Petrarca, F. [Petrarch] (1371) *Le Traité 'De sui ipsius et multorum ignorantia'*, ed. L.M. Capelli, Paris: Honoré Champion, 1906; trans. H. Nachod, *On His Own Ignorance and That of Many Others*, in E. Cassirer *et al.* (eds) *The Renaissance Philosophy of Man*, Chicago, IL: University of Chicago Press, 1948.(Critique of scholastic Aristotelianism and pioneer text of humanist Platonism; referred to in §1.)
Pletho [Georgios Gemistos] (1439) *De differentiis Platonis et Aristotelis (On the Differences between Plato and Aristotle)*, ed. B. Lagarde, 'Le *De differentiis* de Pléthon d'après l'autographe de la Marcienne', in *Byzantion* 43 (1973): 312-43; trans. C. Montague, in C.M. Woodhouse, *Gemistus Plethon: The Last of the Hellenes*, Oxford: Clarendon Press, 1986, ch. 11.(Referred to in §3. Argues for the superiority of Plato to Aristotle.)
Schmitt, C.B. (1976) 'L'introduction de la philosophie platonicienne dans l'enseignement des universités à la Renaissance' (The Introduction of Platonic Philosophy in the Teaching of Renaissance Universities), in *Platon et Aristote à la Renaissance*, Paris: Vrin, 93-104.(Collects evidence for the study of Plato in Italian universities of the late Renaissance.)
Yates, F.A. (1964) *Giordano Bruno and the Hermetic Tradition*, London: Routledge & Kegan Paul.(Classic work on the impact of the Ficinian magical tradition upon the development of modern science.)

Pleasure

From Plato and beyond, pleasure has been thought to be a basic, and sometimes the only basic, reason for doing anything. Since there are many forms that pleasure can take and many individual views of what pleasure consists in, much attention has been given to how pleasures may be distinguished, what their motivational and moral significance might be, and whether there may not be some objective determination of them, whether some may be good or bad, or some better as pleasures than others.

But first there is the question of what pleasure is. It has been variously thought to be a state of mind like distress only of the opposite polarity; merely the absence or cessation of or freedom from pain; a kind of quiescence like contentment; or the experiencing of bodily sensations which, unlike sensations of pain, one does not want to stop. We also identify and class together particular sources of pleasure and call them pleasures of the table, company, sex, conversation, solitude, competition, contemplation or athletic pleasures. In this sense there may be some pleasures which we do not enjoy. But most generally pleasure is what we feel and take when we do enjoy something. This raises the questions of what is encompassed by 'something', what it is to enjoy anything, and the extent to which theories of pleasure can accommodate both our passivity and activity in pleasure. The most influential theories have been those of Plato, Aristotle and empiricists such as Hume and Bentham.

1 Logically proper objects of enjoyment

Elizabeth Anscombe observes that there is both enjoyment of substances: activities themselves, or things, or happenings, or existence itself; and enjoyment of facts: that one made the visit, that one is talking to the wittiest person in the room, that the sun is shining again. Both kinds may be illusory, the apparent fact not a fact, the activity not actually occurring, which calls for qualification in terms of what one thinks to be so as possibly opposed to what is so. The significance of these distinctions is that it may be true that I am enjoying what I take to be the fact that I am having a holiday while it is false that I am enjoying the holiday itself; and the first may be true even though I am mistaken in thinking that I am having a holiday: I may be on call and about to be summoned.

This gives us one interpretation of the distinction between true and false pleasures: my holiday would be truly pleasant if it is true both that I am enjoying the holiday and the fact that I am having a holiday and it is true that I am having a holiday. If any one of these is false it might reasonably be thought that my holiday could not be truly pleasant. But this is not the only interpretation of this distinction.

2 Theories of pleasure

Plato thought that pleasure is the replenishment or filling of a natural lack (for example, *Republic* IX) or, better, in a move to include something more active, the perception of the replenishment of a natural lack (*Philebus*). So eating is a pleasure if one is hungry and one is attending to and enjoying eating as satisfying one's hunger. What one enjoys is always some process which one perceives as restoring a deficiency that has occurred and one has become aware of. Pleasure is not caused by these processes or the perception of them; they are the pleasures in question. So according to this theory the proper object of enjoyment is a process which is perceived to restore the body's silence or remove the clamour of the mind.

This raises the question of how natural lacks may be identified. Even if it is true that one lacks what one desires, it does not follow that all lacks so identified are natural. And while this theory may accommodate plausibly bodily desires and pleasures it seems not easily to deal with intellectual ones, though the examples of curiosity and the pleasures of education may lessen these doubts. Posing greater difficulties are the pleasures of anticipation and memory, and aesthetic pleasures. These are undoubtedly pleasures but only in a strained sense do they imply lacks. Plato was aware of these problems, and partly because of them came in the *Philebus* to see that there was no theory that covered all pleasures.

Freud's theory of pleasure can be regarded as a mirror image of Plato's. Both are homoeostatic theories in that pleasure is the return to rest, though Freud sees that not in replenishment but in discharge of affect at the point where the person's interest is invested.

Aristotle in *Nicomachean Ethics* appears to offer two theories: pleasure is the unimpeded activity of a natural

faculty (VII); pleasure completes an activity by supervening upon it ‘as the bloom of youth does on those in the flower of their age’ (X). Whether or not these are two theories, at least they agree in that the proper object of enjoyment is not a process as Plato thought, but an activity. Aristotle’s distinction is that a process is not complete until it reaches its end whereas an activity has no end in this sense and is complete in itself. One cannot enjoy convalescing because one is not yet well, though one can enjoy such health as one has at any one time. As we might say, activities concern our purposes rather than our ends. If we purpose to be calm and reasonable we do not cease our activity because it has been once successful, whereas our end of being in Cambridge in the spring is extinguished once it is achieved. Pleasure is in achievement, completion, perfection, whether the achievement is continuous or episodic.

G.E.L. Owen has argued that there are two theories but they are about different things. In book VII Aristotle is concerned with the objects of enjoyment; in book X he has moved to the question of what enjoying contributes to the enjoyed activity, in the realization, perhaps, of the need to embrace our passivity. He answers that pleasure is a perfection added to an activity which is already complete in itself. It seems on this view, however, that Aristotle has two sometimes incompatible determinations of the proper object of enjoyment: the unimpeded activity itself, perfect as the prime of youth is perfect; and the perfection of that activity, the bloom on the cheek of those in the prime of youth. It is as if one is in the first case enjoying the unimpeded listening to a Schubert sonata and in the second enjoying the perfection of one’s listening. The second may be part of one’s enjoyment of the first but it need not be. Perhaps the solution is that one’s enjoyment enhances one’s listening and the more exquisite one’s enjoyment the more perfect one’s listening. In that case Owen’s two theories do not remain distinct and neither, as a good consequence, does our activity and passivity in pleasure.

Henry Sidgwick in *The Methods of Ethics* written towards the end of the nineteenth century observed that Aristotle’s theory in one form or another was still current. His interpretation of it is that every normal sense perception or rational activity has its corresponding pleasure and that the most perfect, the exercise of an unimpaired capacity on the best object, is the most pleasant. The pleasure follows the activity immediately giving it a kind of finish like the bloom of youth.

Empiricist theories of pleasure have three characteristics: first, pleasures are quantifiable; second, pleasures either are bodily sensations (see Bodily sensations) localized like those that occur when one warms one’s hands before a fire or nonlocalized like the physical wellbeing after exercise, or are qualities or impressions of such sensations; and third, there is an extensional relation between the pleasure and its object which may or may not be a part of its cause. It is a serious question whether any empiricist from Locke and Hume through Bentham and John Stuart Mill to the pioneers in psychological theory are consistent empiricists about pleasure.

Pleasure comes in amounts which can be increased or decreased and calculated or measured and quantitatively compared. Bentham distinguishes five variables which are involved in determining the total amount of pleasure given by a particular experience: intensity, duration, fecundity (possibilities of further pleasures), purity (absence of mixture with pain) and extent. This requires that pleasures be identifiable occurrences like sensations, or an immediately recognizable quality of such occurrences, which have the right kind of properties for quantification.

So pleasures are caused by other occurrences and the relation between the pleasure experienced and the object of it is quite contingent (or extensional). I may be enjoying the wine I am savouring but why I am enjoying it may have nothing to do with the wine itself. I may have been conditioned or otherwise caused to enjoy whatever is put in front of me. Even when the wine itself is the occasion of my enjoyment of it, it could have been otherwise. That is, empiricist theories of pleasure, while being able to accommodate talk of objects of enjoyment, do not determine what those objects are. There are in theory no proper objects and there is no activity in our enjoyment, only process.

3 Morally proper objects of enjoyment

For classical utilitarian and hedonistic theories (see Hedonism; Utilitarianism) the only morally proper object is pleasure and the absence of pain. But such singularity at the highest level requires quotidian diversity. The most common way of dealing with the question of whether some objects of enjoyment are morally better than others is to distinguish between higher and lower pleasures. For example, John Stuart Mill in the second chapter of *Utilitarianism* writes that some kinds of pleasures are more desirable and more valuable than others and so differ

in quality as well as in quantity. 'It is better to be a human being dissatisified than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied.' He follows Plato who says (*Republic* IX) that the greatest pleasures are those associated with the exercise of our highest faculties (reason) and that apart from their intensity mental pleasures are preferable to any amount of physical pleasure.

Clearly this is tendentious for all that it might appeal to us. Those competent to judge between pleasures doubtless must have experience of them but there would appear to be no place for the idea of expertise in comparative judgments of them. A better test would be whether the pursuit of a particular pleasure is self-defeating.

Aristotle's approach looks more promising. Pleasures, being so closely associated with the activities to which they add perfection, can be compared and judged by evaluating the activities themselves, which one does by considering the lives in which they naturally occur. The bad man's pleasures are not the good man's pleasures since the good man chooses not to live the kind of life the bad man enjoys. Julia Annas argues that Aristotle must therefore reject hedonism and any form of utilitarianism that includes hedonism as a part because for him pleasure is not a single end specifiable independently of different chosen activities and the lives they partly constitute. This yields another interpretation of the distinction between true and false pleasures: true pleasures are those of a good life and false pleasures are not. This shifts the question to what a good life is, but that question has a clearer focus.

4 The content of enjoyment

This question does not concern the objects of enjoyment but what it is to enjoy anything. Empiricist theories construe enjoyment as the having of bodily sensations or episodes like sensations: thrills, transports, raptures or convulsions; the list is Gilbert Ryle's. It is Ryle's argument that this construal of pleasure, though natural to the pioneers in psychological theory, is seriously mistaken. Their programme, he believed, was to construct a dynamic theory of human conduct in which such things as wantings and likings are processes, the mental counterparts to the pressures and attractions of mechanical theory. Psychic motions could be calculated if the durations and intensities of pleasures and desires were measurable and construed on the model of the parallelogram of forces.

But, he argues, pleasure or enjoyment are not processes of any sort. My enjoyment of a game, for example, is not some occurrence that accompanies the game itself, that could distract my attention from the game, that can occur by itself without the game, be described as processes are, as fast or slow, as having some location within the body image, as being, indeed, pleasant or not! These absurdities flow from the mistaken thought that the role of the concept of pleasure is the precise counterpart of the concept of pain. Ryle shows that pleasure is not a sensation as pain mostly is and so not an affective process. There are ways of construing pleasure and pain that will keep them as counterparts, but that will succeed only in making both pleasure and pain not sensations themselves but attitudes to sensations; and there are ways of keeping pains as sensations and not attitudes to sensations, but then they would no longer be counterparts of pleasures. Are there sensations of pleasure rather than pleasurable sensations of warmth, touch, motion, and the like? There seem to be some obvious examples but if there are they are hardly representative of everything that we enjoy and we should do better to avoid misleading assimilations entirely.

This returns us to the necessity of thinking of pleasure as something we take, to our being active and not merely passive in enjoyment. The activity need not be vigorous and may be merely the activity of the imagination or of paying attention or of attending finely to what is happening or to what we are engaged in doing. There are resonances here with Aristotle, and Ryle comes close. He speaks of our liking and disliking as special qualities of the interest we have in an activity, and of that interest itself as a special quality of the activity in question.

This does not imply that we do not feel anything when we enjoy something, only that our feeling enjoyment is not usefully interpreted as being something like a bodily sensation, localised or not. Neither does it follow that we cannot compare our pleasures and discuss their merits and demerits, or contrast them with our pains and miseries. We do, but not by counting and calculating. Contemporary calculations of welfare functions have little to do with pleasure as here reviewed. Unless we take pleasure in what we do and experience, we should get none; but if we feel nothing, it is not pleasure that we take.

GRAEME MARSHALL

References and further reading

Annas, J. (1980) ‘Aristotle on Pleasure and Goodness’, in A. Oksenberg Rorty (ed.) *Essays on Aristotle’s Ethics*, Berkeley and Los Angeles, CA: University of California Press, 285-99.(A very good collection of essays on Aristotle’s views on a number of related topics.)

Anscombe, G.E.M. (1967) ‘On the Grammar of "enjoy"’, repr. in *The Collected Philosophical Papers of G.E.M. Anscombe*, vol. 2, Oxford: Blackwell, 1981, 94-100.(Intricate argument but no technicality.)

Aristotle (*c*. mid 4th century BC) *The Ethics of Aristotle*, trans. J.A.K. Thomson, Harmondsworth: Penguin, 1955.(A classic of Western philosophy.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, repr. London: The Athlone Press, 1980, esp. ch. V.(A classic text.)

Freud, S. (1920) *Beyond the Pleasure Principle*; repr. in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, trans. and ed. J. Strachey, London: Hogarth Press, vol. 18.(Characteristically inventive and contentious.)

Gibbs, B. (1986) ‘Higher and Lower Pleasures’, *Philosophy* 61: 31-59.(A readable and erudite discussion of the subject.)

Gosling, J.C.B. (1975) *Plato, Philebus*, Oxford: Clarendon Press.(The principal commentary on the dialogue.)

Gosling, J.C.B. and Taylor, C.C.W. (eds) (1982) *The Greeks on Pleasure*, Oxford: Clarendon Press.(A scholarly and informative discussion of an immense topic.)

Hume, D. (1739/40) *A Treatise on Human Nature*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1888, esp. Book I, part IV, section II; Book II, part I, section I; Book III, part I, sections I and II.(A classic text.)

Mill, J.S. (1861) *Utilitarianism*, in *Collected Works*, vol. X, Toronto, Ont: University of Toronto Press.(A classic text.)

Owen, G.E.L. (1972) ‘Aristotelian Pleasures’, *Proceedings of the Aristotelian Society*, 1972, 135-52.(A challenging interpretation.)

Plato (*c*.380-367 BC) *Republic*, trans. H.D.P. Lee, Harmondsworth: Penguin, 1955.(A classic of Western philosophy.)

Ryle, G. (1954) ‘Pleasure’, in *Dilemmas*, Cambridge: Cambridge University Press, 54-67.(An entertaining, deflationary, essay.)

Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan, 1962, esp. Bk II, ch. VI.(A classic text.)

Plekhanov, Georgii Valentinovich (1857-1918)

Known as 'the Father of Russian Marxism', Plekhanov was the chief popularizer and interpreter of Marxism in Russia in the 1880s. His interest in the philosophical aspects of Marxism made him influential outside as well as inside Russia. He was a prolific writer, and dealt with several aspects of Marxist thought.

Plekhanov was an important figure in the Russian revolutionary movement. He was a founder member of the Russian Social Democratic Party, and a leading figure in its Menshevik wing after it split into Bolsheviks and Mensheviks in 1903. As a politician, Plekhanov was constantly involved in polemics with political and ideological opponents. Most of his theoretical works are to some degree polemical, and it was the conflicts among Russian revolutionary groups that shaped Plekhanov's interpretation of Marx's thought.

A basic feature of this interpretation was that Russia's historical development was like that of Western European countries, and would pass through a capitalist phase before progressing to socialism. Accordingly, Plekhanov gave prominence to those of Marx's writings which could be presented in a deterministic way. Plekhanov insisted that Marxism was a materialist doctrine (as opposed to an idealist one) and as such recognized the primacy of matter in all spheres of existence.

Plekhanov was in many ways an innovator, being the writer who first coined the term 'dialectical materialism', and who drew attention to the Hegelian origins of Marx's system. His writings were quickly translated into several European languages. His interpretation of Marxism was much admired by Lenin, and was to form the basis of the official ideology of the Soviet Union. The conception of Marxism that Plekhanov propounded continues to exercise a profound influence on conceptions of Marxism throughout the world.

1 Life

Georgii Valentinovich Plekhanov was born in 1857 in the village of Gudalovka in the Tambov province in central Russia. His family belonged to the minor land-owning nobility. From 1868 Plekhanov attended the military academy in Voronezh, and with the intention of following a military career, in 1874 he enrolled in the St Petersburg Mining Institute.

It was while studying in St Petersburg that Plekhanov first came into contact with representatives of the Russian revolutionary movement, including S.M. Stepniak-Kravchinskii, P.B. Akselrod and L.G. Deich, who belonged to the group 'Land and Liberty'.

The members of Land and Liberty held that the function of revolutionaries was to articulate and strive to achieve the concrete aspirations of the common people. They believed that they should avoid projecting onto the people the desires of the educated intelligentsia. The term to describe a revolutionary of this kind was 'Narodnik' (Populist). In the 1870s Russian revolutionaries inspired by Bakunin, Lavrov or Marx all considered themselves to be 'Narodniki'. The slogan which Plekhanov held best to embody the Narodnik principle was: 'The liberation of the workers is the task of the workers themselves'. These were the opening words of the First International's constitution, and had been written by Marx.

In the 1870s there were three main currents of opinion in the Russian revolutionary movement: (1) the followers of Bakunin, who believed that life in the village communities had made the Russian peasants natural communists, and that it only required agitation by the revolutionaries to stir them into rebellion against the state; (2) the followers of Lavrov did not agree that Russian peasants were socialists by nature, and thought that they ought to be educated by the intelligentsia, to bring them up to the same level of consciousness as the intellectual elite in society; (3) the disciples of Tkachëv, by way of contrast, were of the opinion that there was no need to prepare the common people in any way for revolution; it was simply a matter of seizing control of the state by a well-organized conspiracy, and then using the state machinery to change society in a way beneficial to the common people.

At the start of his revolutionary career Plekhanov became a Bakuninist. As such, he was hostile to the state, and thought that the only worthwhile change would be a social one that would come from below. Like most Russian revolutionaries of the period Plekhanov considered political activity a waste of effort. He believed that Marx was in agreement with Bakunin on this point, since he understood Marx to have argued that the economic and social

base brought about changes in the political superstructure and not the other way round.

In December 1876 Plekhanov organized and took part in Russia's first revolutionary demonstration, held at the Kazan Cathedral in St Petersburg. Several participants were arrested, but Plekhanov himself escaped, and began his life as a professional revolutionary. The following year he went abroad, spending several months in Paris, where he met Lavrov, and in Berlin, where he made contact with the German Social Democrats.

On his return to Russia in 1877 Plekhanov conducted revolutionary agitation in Saratov, recruiting local workers and students to Land and Liberty. After a brush with the police, he returned to the capital and joined the editorial board of the journal *Zemlia i volia (Land and Liberty)*, to which he contributed the article 'The Law of the Economic Development of Society and the Tasks of Socialism in Russia'. In it he argued that since the communal system of agriculture was still prevalent in Russia, socialism might be established there directly, without the necessity for capitalist development to run its course.

Consistent with the belief that social change came from the economic base, Land and Liberty maintained that terrorism was incapable of bringing about social transformation, but that it might be used as a reprisal against the authorities. This position began to change following the wave of popular sympathy aroused by the trial and acquittal of Vera Zasulich in 1878 for her assassination attempt on the governor of St Petersburg. This gave an impetus to the use of terror as a political weapon. In 1879 Land and Liberty split to give rise to two new organizations: 'People's Will', which favoured the use of terror, and 'Black Repartition' led by Plekhanov, which adhered to Narodnik principles and eschewed any type of political action.

In order to avoid arrest, Plekhanov left Russia in 1880, expecting to be able to return shortly. In fact he remained in exile for thirty-seven years, until the Revolution of 1917. In Paris Plekhanov translated 'The Communist Manifesto' into Russian as one of a series of socialist works to be published by himself in association with Lavrov. Marx and Engels supplied a foreword in which they admitted the possibility that the peasant commune might be the basis for communist development in Russia. By 1881, however, Plekhanov had come to the conclusion that Russia had now embarked on the path of capitalist development, and that all other paths were now closed to it. Plekhanov considered his change of view to be an adoption of Marxism, but it was a viewpoint out of step with Marx's own thinking.

2 Life (cont.)

Plekhanov's Black Repartition group was overshadowed by People's Will, especially after the latter group organized the assassination of Alexander II in 1881. In 1883 Plekhanov launched a new group, the 'Liberation of Labour', which, he claimed, espoused the doctrines of Marx and Engels. He elaborated the new group's standpoint in a work entitled *Sotsializm i politicheskaia bor'ba (Socialism and the Political Struggle)*, published in October 1883. The book admitted that Black Repartition had been wrong to reject political action, but condemned People's Will for having gone too far in the opposite direction, and adopting the Blanquist tactics of Tkachëv. The Liberation of Labour, Plekhanov argued, was the optimum synthesis of the two extremes represented by Black Repartition and People's Will.

Lev Tikhomirov replied to Plekhanov's book on behalf of People's Will, pointing out that his organization's standpoint and tactics had been grossly misrepresented by Plekhanov. In response Plekhanov published the book *Nashi raznoglasiia (Our Differences)* (1884). This argued that the adherents of People's Will subscribed to a quasi-Slavophile notion that Russian development was unlike that of Western Europe, and that Russia could avoid a capitalist stage. This quasi-Slavophile doctrine Plekhanov dubbed 'Narodism', changing completely the usage of the term. He traced the origins of this Narodism to Herzen and Chernyshevskii, claiming that everyone who believed that Russia might avoid capitalist development belonged to the Narodnik current. This presentation of Russian intellectual history was hotly contested at the time, especially by those Plekhanov classed as Narodniks, but it later gained wide acceptance.

In the late 1880s Plekhanov studied Hegel's philosophy, and made use of this in his polemical essay against Tikhomirov, 'Novyi zashchitnik samoderzhaviia' (A New Champion of Autocracy) (1889) and the article 'Zu Hegel's sechzigstena' (On the Sixtieth Anniversary of Hegel's Death) published in German in 1891. In this latter work Plekhanov first gave currency to the term 'dialectical materialism' to characterize Marx's philosophical method. Plekhanov elaborated the concept of 'dialectical materialism' at length in *K voprosu o razvitii*

monisticheskogo vzgliada na istoriiu (The Development of the Monist View of History) (1895), in which he drew on his studies of classical German Philosophy and French materialism of the eighteenth century. Plekhanov believed that Marx's dialectical materialism was a synthesis of elements taken from both of these schools - metaphysical materialism and dialectical idealism (see Dialectical materialism §§1-2).

In this work Plekhanov also accused his Narodnik opponents of 'subjectivism', which he traced back to Lavrov's *Istoricheskie pis'ma (Historical Letters)* (1870). He tried to find writings by Marx which would also condemn 'subjectivism', and interpreted Marx's contributions to *The Holy Family* (1845) in this light.

At the turn of the century Plekhanov and the other members of Liberation of Labour became involved in a polemic with a new generation of social democrats in Russia, who believed that the efforts of revolutionaries should be devoted less to abstract theorizing and more to catering for the concrete needs of the workers' movement. Plekhanov labelled social democrats of this type 'Economists', and enlisted Lenin's help in a campaign to discredit them.

At the Second Congress of the Russian Social Democratic Labour Party in 1903 Plekhanov supported Lenin and the Bolsheviks against the Menshevik wing. By the end of the year, however, he had changed affiliations and sided with the Mensheviks. In later years he was constantly to vacillate between the two wings of the party. In the years following the 1905 Revolution, when a 'Liquidationist' current appeared within Menshevism arguing that in the new conditions there was no longer any need for an underground party organization, Plekhanov joined with the Bolsheviks in denouncing this as heresy. Between 1907 and 1910 he also cooperated with Lenin in a campaign against the current of Russian socialist philosophy represented by Aleksandr Bogdanov, his main publication in this respect being the series of open letters with the general title 'Materialismus Militans'.

Plekhanov's accord with the Bolsheviks was destroyed after the outbreak of the First World War. Whereas the left wing of Russian Social Democracy condemned the war, Plekhanov became an ardent advocate of an Entente victory over the Germans. He maintained this position even after he returned to Petrograd in April 1917, contributing pro-war articles to the Menshevik newspaper *Edinstvo (Unity)*.

The main theoretical work written by Plekhanov after *The Development of the Monist View of History* was *Osnovnye voprosy marksizma (Fundamental Questions of Marxism)*, which appeared in 1908. This, however, mainly restated ideas which had been formulated in earlier writings. In 1909 the Moscow publishing house Mir commissioned Plekhanov to write a history of Russian social thought in the nineteenth century. This project became Plekhanov's main literary endeavour for the rest of his life. When he died, on 30 May 1918, however, only three out of a projected seven volumes had been completed.

3 Plekhanov's Marxism

Plekhanov became an extremely influential interpreter of Marx's ideas because he wrote at an early and crucial stage in their propagation. He was attracted to the scheme of historical determinism associated with Marx's name in the early 1880s when both Marx and Engels were still alive, and when relatively few of Marx's works had been published (see Marx, K. §8).

Moreover, Marx died in 1883 without leaving behind him a completed system. Originally he had intended to deduce the categories of political economy in the same manner as Hegel had deduced the categories of philosophy, and in doing so show that the world market was contained in the concept of capital. The reasoning was that when capital - externalized human social nature - became a world system and reached its point of culmination, it would collapse and give way to socialism. By the 1870s, however, Marx had come to realize that the circulation and reproduction of capital would not of itself erode existing economic and social structures. In studying the economic processes in Russia he formed the opinion that capitalist development might be forestalled there and socialism established on the basis of the peasant commune.

Although Plekhanov was well aware of Marx's views on the possibilities of Russian development, he did not try to amplify or propagate them. Instead, he attributed to Marx the doctrine that all countries passed through the same historical stages, and that capitalist development would necessarily take place in Russia.

Plekhanov was not a dispassionate student of Marx's ideas. He approached them as a polemicist, emphasizing some aspects of Marx's work and ignoring others as the stance he took in a given polemic required. The kind of

philosophical questions Plekhanov took up were invariably those which had a bearing on the factional disputes in which he was involved.

The dispute on the efficacy of political action impelled Plekhanov to examine the relationship between historical determinism and the freedom of human action. This was to be a recurrent theme in his writings. Plekhanov argued that freedom and necessity were interconnected ’dialectically’, that they formed the kind of unity, a freedom-necessity, that Schelling had proposed in his philosophy of identity (see Schelling, F.W.J. §2). He also compared the interrelationship between freedom and necessity to Spinoza’s Substance, which led him to define Marxism as a variety of Spinoza’s philosophy. Plekhanov also claimed that the freedom-necessity relationship had been satisfactorily formulated by Hegel when he stated that freedom implied knowledge of the laws of nature and submission to them. The primacy given to necessity in this formulation allowed Plekhanov to assert that the course of history was determined, not by human volition, but by the material productive forces.

In keeping with this conception of human history Plekhanov repeatedly stressed that human nature was not a constant, but an eternally changing product of the historical process. He claimed that Schelling, Hegel and Marx had all subscribed to this view. He was only able to find a single (ambiguous) passage in Marx to support this contention.

Plekhanov emphasized that ’dialectics’ as he understood it differed substantially from simple evolution. He quoted Hegel to the effect that in both nature and human history transformations of quantitative into qualitative differences took place not gradually, but by ’leaps’. These leaps Plekhanov considered to be characteristic of the dialectical method that Marx and Engels had taken over from Hegel. In fact, in the passage cited Hegel was concerned to show that creation and destruction ought to be accounted for in terms of the development of the Concept, and that in changes from quantity to quality and *vice versa* the element of unpredictability - the ’cunning of the Concept’ - operated at even the most basic levels of existence. Hegel’s actual argument was lost on Plekhanov, who had only a superficial understanding of Hegel’s system.

Plekhanov drew a firm distinction between idealist and materialist philosophies, asserting that only the latter were compatible with Marxism. He rejected any suggestion that human knowledge of the material world was conditioned in some way by the constitution of the human mind. He insisted that things-in-themselves were material objects and these directly or indirectly acted on the human senses to produce sensations. Materialism, as Plekhanov understood it, recognized the material nature of things, while idealism denied this materiality. This was an argument taken up by Lenin in his book *Materializm i ėmpiriokrititzism (Materialism and Empiriocriticism)* (1909) (see Lenin, V.I. §1). The implication was that Marx’s philosophical conceptions were derived from the eighteenth-century French materialist tradition rather than from the line of inquiry that Kant had initiated and to which Fichte, Schelling and Hegel had, in their turn, contributed.

4 Significance and evaluation of Plekhanov’s work

Plekhanov was able to achieve prominence as an authority on the philosophical foundations of Marx’s thought because by the end of the nineteenth century knowledge of the intellectual context in which Marx had grown up had become extremely rare. As a result, his standing was high not only in Russia, but also throughout Europe. It was largely in the interpretation given to them by Plekhanov that Marx’s ideas were received by the wider public. This is indicated by the fact that Plekhanov’s term ’dialectical materialism’ as a synonym for ’Marxism’ gained general currency.

Dialectical materialism, as Plekhanov had elaborated it, formed the philosophical foundation for the official ideology of the Soviet state. In this form it was widely propagated throughout the world, and was the presentation in which most people were introduced to Marx’s thought. Even the influential Marxist philosopher Georg Lukács took Plekhanov’s interpretation of Marxism as the starting point for his own conception of Marx’s ideas. Lukács regarded Plekhanov and Engels as commentators on Marx of approximately equally high stature.

Plekhanov came to Marx’s system when that system was incomplete and was in the process of modification by its author. It therefore lacked an obvious internal cohesion. Plekhanov exacerbated this situation by preserving the historical determinist aspect which Marx had rejected, and refusing to recognize, as Marx had done, human social nature as the driving force behind the historical process. As a result, Plekhanov propounded a ’Marxism’ that was internally inconsistent and contained the almost mystical implication that somehow ’matter’ would, in the course

of human history, create a socialist society by 'leaps'.

Moreover, Plekhanov's insistence for polemical reasons on interpreting Marx's disagreements with the Bauer brothers in 1844-5 as a campaign against 'subjectivism' gave a misleading impression of the character of Marx's system and how it had emerged. Marx was in fact concerned to combat the common perception that philosophical abstractions had an existence apart from the concrete material they generalized. This was the sense in which he considered himself a 'materialist'. The attempt to eliminate abstractions and other forms of 'reflection' linked Marx to the main current of German philosophy. Plekhanov classed the latter as 'idealist' and would not countenance any debt by Marx to it.

The polemical tactic of assigning all previous Russian socialist currents to the category of Narodism distorted the intellectual history of Russia. It gave the impression that Narodism and Marxism were two distinct phases in the development of Russian socialist thought, and that Narodism was characterized by 'subjectivism', a Slavophile belief in Russia's uniqueness, and a conviction that capitalism could not take root in Russia. Narodism in this sense was a creation of Plekhanov's, but because it formed an integral part of Soviet ideology, it was incorporated into Soviet historical writing.

See also: Marxist philosophy, Russian and Soviet §§1-2

JAMES D. WHITE

List of works

Plekhanov, G.V. (1923-7) *Sochineniia (Works)*, ed. D. Riazanov, Moscow: Gosizdat, 24 vols.(The fullest edition.)

Plekhanov, G.V. (1956-8) *G.V. Plekhanov: Izbrannye filosofskie proizvedeniia*, Moscow: Gosizdat, 5 vols; trans. B. Trifonov, *G. Plekhanov: Selected Philosophical Works*, Moscow: Progress, 1974-81, 5 vols.(This edition contains most of Plekhanov's main theoretical writings.)

Plekhanov, G.V. (1883) *Sotsializm i politicheskaia bor'ba*, Geneva: Tipografiia gruppy 'Osvobozhdenie Truda'; *Socialism and the Political Struggle*, in *Selected Philosophical Works*, Moscow: Progress, 1974-81, vol. 1, 49-106.

Plekhanov, G.V. (1884) *Nashi raznoglasiia*, Geneva: Tipografiia gruppy 'Osvobozhdenie Truda'; *Our Differences*, in *Selected Philosophical Works*, Moscow: Progress, 1974-81, vol. 1, 107-352.

Plekhanov, G.V. (1895) *K voprosu o razvitii monisticheskogo vzgliada na istoriiu*, St Petersburg: Skorokhodova; *The Development of the Monist View of History*, in *Selected Philosophical Works*, Moscow: Progress, 1974-81, vol. 1, 480-737.

Plekhanov, G.V. (1908) *Osnovnye voprosy marksizma*, St Petersburg: Slovo; *Fundamental Questions of Marxism*, in *Selected Philosophical Works*, Moscow: Progress, 1974-81, vol. 3, 117-83.

Plekhanov, G.V. (1910) 'Materialismus Militans' in *Ot oborony k napadeniiu (From Defence to Attack)*, St Petersburg; *Selected Philosophical Works*, Moscow: Progress, 1974-81, vol. 3, 188-283.(A series of open letters against the current of Russian socialist philosophy represented by Aleksandr Bogdanov.)

Plekhanov, G.V. (1914-16) *Istoriia russkoi obshchestvennoi mysli (History of Russian Social Thought)*, trans. M. Niqueux and M. Godot, *Histoire de la pensée sociale russe*, preface by J. Scherrer, Paris: Institut d'Études Slaves, 1984.(The 1984 edition is a translation into French of Plekhanov's 'Introduction' to this unfinished work.)

References and further reading

Baron, S. (1962) 'Between Marx and Lenin: George Plekhanov', in L. Labedz (ed.) *Revisionism: Essays on the History of Marxist Ideas*, London: Allen & Unwin.(This short biographical essay concentrates more on Plekhanov's political views than his philosophical ideas.)

Baron, S. (1963) *Plekhanov: The Father of Russian Marxism*, Stanford, CA: Stanford University Press.(The fullest biography of Plekhanov in English. Baron describes Plekhanov as 'one of the most sensitive and creative of Marx's disciples', perhaps not appreciating the extent to which Plekhanov distorted Marx's system.)

Kolakowski, L. (1978) *Main Currents of Marxism*, Oxford: Oxford University Press, vol. 2, ch. 14.(A perceptive analysis of Plekhanov's philosophical development.)

Walicki, A. (1969) *The Controversy over Capitalism: Studies in the Social Philosophy of the Russian Populists*, Oxford: Clarendon Press.(First published in 1965 as an introduction to a Polish collection of 'Narodnik' writings on social philosophy. This collection was an expanded version of a previously published Soviet work

of the same type. Walicki's essay elaborates on the Soviet interpretation, originating with Plekhanov, that the 'subjectivist' Narodnik doctrines gave way to the 'objectivist' Marxist current.)

Walicki, A. (1980) *A History of Russian Thought from the Enlightenment to Marxism*, Oxford: Clarendon Press. (This repeats the version first advanced in the 1965 introductory essay.)

Plotinus (AD 204/5-70)

Plotinus was the founder of Neoplatonism, the dominant philosophical movement of the Graeco-Roman world in late antiquity, and the most significant thinker of the movement. He is sometimes described as the last great pagan philosopher.His writings, the so called Enneads, are preserved as whole. While an earnest follower of Plato, he reveals other philosophical influences as well, in particular those of Aristotle and Stoicism. Plotinus developed a metaphysics of intelligible causes of the sensible world and the human soul. The ultimate cause of everything is 'the One' or 'the Good'. It is absolutely simple and cannot be grasped by thought or given any positive determination. The One has as its external act the universal mind or 'Intellect'. The Intellect's thoughts are the Platonic Forms, the eternal and unchanging paradigms of which sensible things are imperfect images. This thinking of the forms is Intellect's internal activity. Its external act is a level of cosmic soul, which produces the sensible realm and gives life to the embodied organisms in it. Soul is thus the lowest intelligible cause that immediately is immediately in contact with the sensible realm. Plotinus, however, insists that the soul retains its intelligible character such as nonspatiality and unchangeability through its dealings with the sensible. Thus he is an ardent soul-body dualist. Human beings stand on the border between the realms: through their bodily life they belong to the sensible, but the human soul has its roots in the intelligible realm. Plotinus sees philosophy as the vehicle of the soul's return to its intelligible roots. While standing firmly in the tradition of Greek rationalism and being a philosopher of unusual abilities himself, Plotinus shares some of the spirit of the religious salvation movements characteristic of his epoch.

1 Life and writings

We possess a fairly reliable account of Plotinus' life and writings by Porphyry, his student and editor. Porphyry composed a biography, *On the Life of Plotinus and the Order of his Books*, which prefaced his posthumous edition of Plotinus' writings. At the age of twenty eight Plotinus began his philosophical studies in Alexandria under a certain Ammonius (often called Ammonius Saccas, and not to be confused with Ammonius the teacher of Plutarch of Chaeronea (§1) or with Ammonius, son of Hermeas) and studied with him for several years. After making a futile attempt at a journey to the East in order to acquaint himself with the philosophy of Persia and India, he settled in Rome at the age of forty. He established a school in Rome and stayed there except during his final illness.

The extant corpus of Plotinus' writings is one of the largest we have of any ancient philosopher, and we probably possess everything he wrote. His works are treatises, written in Greek, that grew out of discussions in his school, and vary greatly in length and scope. Porphyry arranged the treatises according to subject matter into six 'enneads' - six sets of nine treatises. In order to arrive at this division he had to split some treatises. Conventionally, references to the *Enneads* are often given only in numbers: 'V 3(49).2, 14-16', for instance, means '5th Ennead, 3rd treatise (which is number 49 on Porphyry's chronological list of Plotinus' writings), chapter 2, lines 14 to 16'.

2 General characteristics

Plotinus' thought grows out of a tradition of so-called Middle Platonism (see Platonism, Early and Middle), which developed during the first two centuries AD. The Middle Platonists interpreted Plato dogmatically: that is, unlike the Academic sceptics, they took Plato to hold definite views. Their interests lie primarily in metaphysics and psychology, broadly understood. Even if they show obvious influence from other schools of thought - many Aristotelian and Stoic notions had become a part of the language of philosophy itself - Middle Platonists believe that the fundamental philosophical truths are to be found in Plato. Plotinus shares these general characteristics.

It is sometimes said that Plotinus is a system builder who never reveals his whole system in an organized way, and that the system must be inferred from bits and pieces here and there in his writings. Another common dictum is that every one of his treatises presupposes all the rest and the whole system. There is something to these claims. However, even if behind Plotinus' writings there lies a more organized comprehensive view than meets the eye, his mind is not that of the rigid system builder. His philosophical genius rather consists in the exceptional sensitivity and depth of thought with which he addresses difficulties inherent in the Platonic tradition.

3 The One and the hierarchy of principles

Plotinus works with a fundamental dichotomy between the intelligible (*to noēton*) and the sensible (*to aisthēton*). The intelligible world is the realm of the real (in the sense of existing independently, being by virtue of itself), and it is unchanging and non-spatial. The sensible world, by contrast, is an unreal, changeable image of the intelligible, expressed in spatial extension. Within both these realms there are further divisions, so that the result is a hierarchical ontology. At the apex of the hierarchy is the One (also referred to as the Good), then, in descending order, are Intellect and Soul (see Nous; Psychē). These three levels, the One, Intellect and Soul, are often called *hypostases* and their names are customarily capitalized. Particular souls - of people, animals and the soul of the sensible world itself - also belong to the intelligible order but are in close contact with the sensible realm. The hierarchy continues in the sensible world, where organisms, forms in matter (that is, sensible qualities), inanimate bodies and matter constitute the main stages. As is evident from the non-spatial nature of the intelligible world, such terms as 'hierarchy', 'above' and 'below' in the present context are not to be understood spatially but as indicating ontological priority.

The One, Intellect, Soul, and, with certain qualifications to be explained below, matter too, are explanatory postulates of the kind the Greek philosophers called principles (*archai*) (see Archē). Plotinus takes a strong realist position with regard to his principles: not only do they exist, they are more real, exist in some fuller sense, than that which they were meant to explain. Inorganic bodies, organisms and their functions, and human consciousness are phenomena to be accounted for in terms of the principles.

Considerations such as the following help explain the plurality of stages in the Plotinian hierarchy. Once a distinction has been made between what is to be accounted for and a principle that explains it, questions may arise about the principle itself: the principle itself may turn out to have features that stand in need of an explanation. A further principle must then be assumed to account for the first one. The process of seeking further principles continues until a principle that needs no further explanation is reached, a principle about which no further questions can be asked. For Plotinus this ultimate principle is the One.

Plotinus generally holds that the principles themselves must have the features they explain. For instance Soul, which is the principle of life, is itself alive. Moreover, the principles ideally have these features in such a way that it is pointless to ask why they have them. The principle possesses in and of itself what other things possess as a derived and contingent feature and hence one that requires explanation. Plotinus frequently expresses this by saying of a principle that it is such and such in itself (*en heautōi*), whereas other things have the same feature as in another (*en allōi*).

Unity is a central concept in Plotinus' doctrine: each stage of the hierarchy has a characteristic kind of unity, with the One as the absolutely simple ultimate principle which is the cause of all other unity that there is, and thereby, the cause of everything else whatsoever. To be is to be one thing, to be unified, and the more unified something is the more of a being it is. The most striking feature of the world of everyday experience is in fact the unity of it as a whole and of individual objects, especially living things, in it. The organization, regularity and beauty that are evident in the world of everyday experience, all of which may be said to express its unity, cannot be explained in terms of its constituent parts. The latter are what is unified and their unity is an imposed feature which must come from elsewhere. The unity revealed in the sensible world is in Plotinus' view far from perfect but it gives the sensible world the reality it has. The same may be said of our experiences of ourselves: introspection will show that the human soul has a more perfect kind of unity than anything pertaining to the body, although even the soul does not have unity of itself (IV 2(4).2; IV 7(2).6-7). Thus, the world of everyday experience, both the external world and our mental life, point beyond themselves to a higher level of reality which is its principle.

This process of going upwards from everyday phenomena obviously draws on Plato's dialectic as described for instance in the *Symposium* and the *Republic*. There are many instances of such a procedure in the *Enneads*, the most famous being Plotinus' first treatise, *On Beauty* (I.6911)), where he builds on Diotima's speech in the *Symposium*. This treatise has been influential in art, especially during the Renaissance. The ascent from the beauty of corporeal things to the Beautiful itself is, as one would expect, interpreted in terms of the Plotinian hierarchy and general doctrine of intellectual ascent (see §6). There is a noteworthy deviation from Plato's views of the arts as expressed in the *Republic* in that for Plotinus art does not imitate nature but operates parallel with nature (I 6.3; V 8(31).1). Thus, the artist uses the intelligible world directly and gives it an expression in the sensible world.

Ignoring the intermediate stages for the moment, the self-sufficiency of the principles with regard to the features they explain, together with the claim that whatever else there is presupposes unity, leads to the highest principle, the One, which is both absolutely simple and unique (V 4(7).1). The doctrine of the One, even if adumbrated by the tradition before Plotinus, is probably his most significant innovation. Some of his Middle Platonist predecessors believed in a simple first principle, but like Aristotle they thought that this simple principle was an intellect of some sort. In Plotinus' view, by contrast, Intellect involves plurality: there is plurality in thought because there is at least a conceptual distinction between the thought and its object, and what is thought is in any case varied (see, for example, V 3(49).10).

The One is unique and involves no variation or limitation. From this it follows that the One cannot be given any positive characterization. It cannot be grasped by thought or known in its true nature since any thought of it distorts in so far as the thought is bound to be composite (V 5(32).6; VI 9(9).4). It would even be inappropriate to say of the One that it is, or that it is one inasmuch as such expressions indicate something unified rather than the absolutely simple nature that gives unity to whatever is unified (VI 9.5). Nevertheless, it is possible to approach the One and even become one with it in a kind of non-cognitive union, a 'vision' which defies all description (VI 9.8-11). It is as a result of this doctrine of a union with the ultimate principle, a union which transcends conceptualization, that Plotinus has been called a mystic. Without wishing to deny the significance of this union with the One, it must be said that it does not play a major role in his writings. In spite of the fact that the One is ineffable, Plotinus manages to say a great deal about it. This is however not a real inconsistency: the ineffability of the One is the thesis that one cannot predicate anything of it since any predication would imply the One's composition. This does not mean one cannot talk about its relations to other things and in general about its role in the ontology.

Even if there are precedents in Plotinus' tradition for a supreme formal principle, most of his predecessors would postulate in addition other ultimate principles. In Plotinus everything derives from the One, even if the lower levels in the hierarchy function in fact as principles of multiplicity. In this sense Plotinus is an ardent monist.

4 Intellect, Soul and the sensible world

The divine Intellect, the stage after the One, is also the realm of the (Platonic) Forms and of the real or primary being. The identification of the realm of the Forms with primary being is straightforward: by definition each Form is eminently and of itself that which it causes in others. But this primary being is also the universal Intellect: the Forms exist as the thoughts of Intellect, whose thinking of the Forms is described as self-thought, and its knowledge of them as a kind of self-knowledge (see especiallyV 5, V 8(31) and V 3). Of several arguments Plotinus advances for the claim that the Forms are internal to the Intellect, the philosophically most interesting one is an argument to the effect that if the Forms are outside the Intellect, its knowledge of them must be acquired; but the Forms are the standards of judgment and if the Intellect does not possess these standards previously it will lack the necessary means of recognizing the impression of each Form for what it is. So, if the Intellect does not essentially contain the Forms as its thoughts, its knowledge becomes problematic: an unacceptable conclusion since it is agreed that the universal divine Intellect has supreme knowledge. Plotinus' identification of primary being with a divine Intellect implies that there is a level of reality where knowledge and being, epistemology and ontology, coincide. This Plotinus takes to be a necessary condition of the possibility of knowledge.

As mentioned above, Intellect is characterized by a greater unity than the sensible world. This is first of all brought out by the fact that Intellect is non-spatial and non-temporal and hence free from the dispersion that has to do with space and time. Second, the part-whole relations in Intellect are such that not only does the whole, which is more than the sum of its parts, contain its parts, the whole is also implicit in each of the parts (see, for example, VI 2(43).20). Third, there is no real distinction between subject and attribute at the level of Intellect. Instead Plotinus posits intellectual substance and its activity (*energeia*), which is identical with the substance. Much of this doctrine about the relationships between items on the level of Intellect is founded on interpretations and suggestions in Plato's late dialogues. Plotinus takes the five 'greatest kinds' of the *Sophist* - being, sameness, difference, motion and rest - as the highest genera of his ontology. Each of these is distinct, but nevertheless presupposes and is interwoven with all the others. As a whole they constitute the Intellect or the intelligible substance. Particular Forms are generated from them.

The integrity of Intellect implies that Intellect's thought is different from ordinary discursive thought: Intellect

grasps its objects and all their relations in an atemporal intuition of the whole, employing neither inferences nor words; the vehicles of its thought are the very things themselves, the prototypes and causes of which all other things, whether natural phenomena or lower modes of human thought, are inferior manifestations.

Soul is the level below Intellect. On account of the multiplicity of its functions, Soul is in some ways the most complex of the Plotinian *hypostases* and conceptually the least unified one. The historical sources of Plotinus' notion of Soul are primarily Plato's dialogues, above all the *Timaeus*, but Plotinus' psychology also reveals strong Aristotelian and Stoic influences.

Plotinus makes certain distinctions within the psychic realm. There is the *hypostasis* Soul, which remains in the intelligible realm, and there are the World-Soul and the souls of individuals, where the latter two are on the same level (IV 3(27).1-8). Within the two latter types of soul, Plotinus further distinguishes between a higher and a lower soul, corresponding to a distinction between soul which operates directly through a body and soul which does not (this distinction coincides with the distinction between rational and non-rational soul). Soul is the intelligible level which is responsible for the sensible world. The lower soul, sometimes referred to as nature (*physis*), produces matter itself, inorganic bodies and ordinary living things, including the sensible cosmos itself, which according Plotinus is a supreme organism (IV 4(28).33).

Plotinus holds that all souls are one, that all souls are identical with the *hypostasis* Soul (and by implication with one another). The Neoplatonists after Porphyry rejected this doctrine but Plotinus maintains it consistently and attaches considerable importance to it (IV 9(8); VI 4 (22).4; IV 3.1-8). Such a doctrine seems to be implied by the combination of two Plotinian doctrines that we have just mentioned: the Soul's membership of the intelligible realm (or the realm of real being) and the integrity of that realm.

It has been mentioned that Intellect is outside space and time. In *On Eternity and Time*, Plotinus states his views on time. The treatise contains interesting and powerful criticisms of the views of Aristotle, the Stoics and the Epicureans. As is often the case, Plotinus' own views are a development of Plato's *Timaeus*. He defines eternity as 'the life which belongs to that which exists and is in being, all together and full, completely without extension or interval' (III 7 (45). 3, 36-8) and time as 'the life of soul in the movement of passage from one mode of life to another' (III 7.11, 43-5). Thus, time comes in at the level of Soul as the 'image of eternity'. This means that Soul, in producing the sensible world, unfolds in successive stages what at the level above is present all together and without temporal interval.

Certain difficulties arise precisely on account of Soul's close relationship with the sensible. In the first place, how can Soul cause, administer and ensoul the extended sensible world without thereby coming to share in its extended nature? How can Soul operate in different parts of extension without being divisible into spatially distinct parts itself? If it is divided, its intelligible status will be lost or at least seriously threatened. This difficulty is increased by the fact that according to common and deeply ingrained opinion Soul is present in the bodies it ensouls. Plotinus was deeply disturbed by these and other puzzles having to do with Soul's relationship with the sensible realm as is demonstrated by the fact that he returns to the questions repeatedly. Despite brave attempts, it is questionable whether he succeeds in giving a satisfactory account of them.

One solution Plotinus frequently suggests and argues for, mainly from facts about the unity of consciousness in sensation, is that the Soul is present as a whole at every point of the body it ensouls. In this way it can be at different places without being divided. Its being present as a whole in different parts of space shows its different ontological status from that of bodies which have numerically distinct spatial parts (see, for example, IV 2 (4).2). Another account, however, presents Soul as not present in body at all, but rather the reverse - body as present in Soul. Body is in Soul in the same way that bodies may be said to be in light or in heat: they thereby become illuminated or warm without (in Plotinus' view) dividing or affecting the source of light or heat in any way. Similarly, bodies become ensouled, alive, by virtue of the presence of Soul. These ideas are explored in VI 4 (22) and 5 (23) (constituting a single treatise), which contain what is perhaps Plotinus' subtlest account of the relation between the sensible and the intelligible. Here he attempts to explain the apparent presence of the intelligible in the sensible in terms of its having effects in the sensible without being locally present there.

At the bottom of the Plotinian hierarchy lies matter. Plotinian matter is like the One in that it permits no positive characterization, but this is for exactly the opposite reasons. The One is, one might say, so full, so perfect, that it

eludes any positive description. Matter, on the contrary, is such on account of its utter privation, or lack of being: it is sheer potentiality. Matter in Plotinus' system is the receptacle of immanent bodily forms, such as colours, shapes and sizes. Physical objects, bodies, are composites of matter and such immanent forms (VI 3(44).15). Matter itself is not subject to change but underlies change: as Forms come and go matter remains unaffected (III 6(26)). It is as such imperceptible, but reason convinces us of its existence as a purely negatively characterized substrate of forms. As what underlies the forms of bodies, it might be tempting to identify matter with space or with mass. Plotinus nevertheless rejects this, maintaining that the three-dimensionality of space presupposes local determination and that mass contains form, whereas matter is totally indeterminate and without form (II 4(12).8-12). Nevertheless, matter is the principle of spatial extension in that the dispersion characteristic of space is due to matter. Matter is an explanatory principle in the sense that it is necessary for explaining plurality, although it is not a principle of being in Plotinus' sense.

In the treatise I 8(51), Plotinus discusses at length questions concerning evil, a topic also brought up in many other treatises. The intelligible world is perfect and totally self-sufficient. While the sensible world is not, it is a reflection of the former and contains nothing which does not have its origins there. It is therefore puzzling how evil can arise. Plotinus argues that evil as such does exist and he identifies it with matter, understood as lack of all form, or non-being. This is absolute evil. Other things are evil in a relative way in so far as they have a share in matter.

5 Causation and critique of Aristotle

In the preceding account of the Plotinian principles mention has been made of the relationships between the *hypostases*. It remains, however, to address this topic generally. Plotinus inherits from Plato two ways of describing the relationship: the language of participation and the language of model (*paradeigma*), imitation and image (*eikōn*, *eidōlon*). A lower level participates in a higher one, and thereby comes to have the character of the latter, or it imitates the higher to the same effect. Both these ways of describing the relation see it from the viewpoint of the effect. Plotinus and the other Neoplatonists have in addition ways of describing the relation in terms of the causal agency of the higher level. This is what is commonly called emanation, although Plotinus' language is quite varied here. He often simply uses expressions such as 'to make' (*poiein*) and 'to proceed' (*proienai*) for the activity of a higher realm. He also frequently uses the analogies of the sun and the light it radiates, fire and heat and the like, to illustrate how a higher *hypostasis* generates a lower, and occasionally he uses the metaphors originating in language about water (for example, 'to flow out'). He is well aware of the fact that these are metaphors that must not be understood too literally. The term 'emanation' may mislead in so far as it suggests that the cause spreads itself out. Plotinus, on the contrary, consistently maintains that the cause always remains unaffected and loses nothing by giving away.

In Plotinus there is sometimes an explicit and often an implicit distinction drawn between 'internal' and 'external activity' (*energeia*) (see, for example, V 3.12; V 4(7).2). This distinction runs through every Plotinian cause down to Soul and is crucial for an understanding of causation in the system. Keeping in mind what has been said about the identity of a substance with its activity, the internal activity will be the same as the thing itself. In terms of the light analogy the internal act is analogous to whatever the source of light, considered in itself and as a source of light, is doing. The external act is this same entity considered as operating in something else, causing the brightness on the wall for instance.

The matter is more complicated still: there is not only a process from the cause, but also a reversion (*epistrophē*) of the produced towards its source without which the external act is incomplete. Since there is no pre-existent material principle, a recipient of form, what becomes informed must come from the informing cause. Thus, the outgoing aspect of a given level in the hierarchy functions as a material principle for the level below, the returning aspect as the informing of that material principle. Thus, reversion is equivalent to imitation. The product is an image or expression of the original, an effect, which nevertheless is not cut off from its cause, because the effect depends on the activity of the cause.

All activity except the internal activity of the One is a form of thinking or contemplation (*theōria*) (III 8(30)). (Even the One's activity appears to be mental activity of a sort, although it is not 'thinking' (see, for example, V 4(7).2; VI 8(39).16.) This holds even of the activity of the lower phase of the World Soul, nature, which creates the sensible forms in matter: all action (*praxis*) and production (*poiēsis*) have contemplation as their goal;

production is the result of weak or imperfect thinking. We may visualize the system as a hierarchy where each stratum has an external activity, which in an attempt to grasp in thought its source creates an image of it, revealing it as more 'unfolded', less unified. So in a way the same items exist on every level: the One is everything there is, but in such a unified form that no distinctions are to be found. Likewise, Intellect and Soul, and finally the physical world contain everything there is. Only at the very lowest levels, that of matter and immanent sensible forms, is there no generation, which is another way of saying that we have reached the bottom.

Plotinus wrote some treatises on technical philosophical subjects such as potentiality and actuality (II 5(25)) and substance and quality (II 6(17)). The treatise on matter (II 4 (12)) may also be said to belong to this group. The most extensive among such treatises, however, is VI 1(42)-3(44), which Porphyry gave the title 'On the Kinds of Being'. In VI 1 Plotinus critically discusses Aristotle's *Categories* (see Aristotle §7; Categories §1), which he takes to be about the genera of being and, more briefly, the so-called Stoic categories (see Stoicism §6). He presents in VI 2 his own account of the genera of intelligible being, which for him is the only real being, and in VI 3 a revised version of Aristotle's categories doctrine is offered as an account to hold for sensible things. Plotinus' criticisms of Aristotle's category doctrine are founded on the assumption that it is about the genera of being. His main critical points are: (1) Aristotle's *Categories* fail to give a place to intelligible beings and hence cannot present a universal doctrine of the genera of being - it cannot be supposed to hold for both sensibles and intelligibles since there is no common genus for both and no discourse covering both at once; (2) each of the categories described by Aristotle fails to be one genus.

6 Human beings

A noteworthy feature of Plotinus' psychology is his use of Aristotelian machinery to defend what is unmistakably Platonic dualism. For instance he uses the Aristotelian distinctions between rational, perceptive and vegetative soul much more than the tripartition of Plato's *Republic* (see Psychē). He employs the Aristotelian notions of power (*dynamis*) and act (*energeia*), and sense perception is described very much in Aristotelian terms as the reception of the form (*eidos*) of the object perceived (see Aristotle §18). However, he never slavishly follows Aristotle and the reader should be prepared for some modifications even where Plotinus sounds most Aristotelian.

Plotinus identifies human beings with their higher soul, reason. The soul, being essentially a member of the intelligible realm, is distinct from the body and survives it. It has a counterpart in Intellect which Plotinus sometimes describes as the real human being and real self. As a result of communion with the body and through it with the sensible world, we may also identify ourselves with the body and the sensible. Thus, human beings stand on the border between two worlds, the sensible and the intelligible, and may incline towards and identify themselves with either one. For those who choose the intelligible life, philosophy (dialectic) is the tool of purification and ascent. As noted previously, however, it is possible to ascend beyond the level of philosophy and arrive at a mystical reunion with the source of all, the One. In contrast with the post-Porphyrian Neoplatonists, who maintained theurgy as an alternative, Plotinus stands firmly with classical Greek rationalism in holding that philosophical training and contemplation are the means by which we can ascend to the intelligible realm.

Plotinus' account of sense perception is an interesting example of how he can be original while relying on tradition. Sense perception is the soul's recognition of something in the external sensible world. The soul alone only knows intelligibles and not sensibles. If it is to come to know an external physical object it must somehow appropriate that object. On the other hand, action of a lower level on a higher is generally ruled out and a genuine affection of the soul is impossible because the soul is not subject to change. Plotinus proposes as a solution that what is affected from the outside is an ensouled sense organ, not the soul itself. The affection of the sense organ is not the perception itself however, but something like a mere preconceptual sensation. The perception properly speaking belongs to the soul. It is a judgment (*krisis*) or reception of the form of the external object without its matter. This judgment does not constitute a genuine change in the soul for it is an actualization of a power already present. In formulating this problem Plotinus' dualism becomes sharper, and in some respects closer to Cartesian dualism, than anything found in Plato or previous ancient thinkers. Plotinus contrasts sense perception as a form of cognition with Intellect's thought, which is the paradigm and source of all other forms of cognition. Sense perception is in fact a mode of thought but it is obscure. This is because the senses do not grasp the 'things themselves', the thoughts on the level of Intellect, but mere images. Since they are images they also fail to reveal the grounds of their being and necessary connections.

As the preceding account may suggest, Plotinus sees the goal of human life in the soul's liberation from the body and from concerns with the sensible realm and identification with the unchanging intelligible world. This is in outline the doctrine of Plato's middle dialogues. There are noteworthy elaborations, however. Plato holds that the soul's ability to know the Forms shows its kinship with them (*Phaedo* 79c-e). Plotinus agrees and presents a doctrine about the nature of this kinship which is left unclear in Plato. For as we have seen the whole realm of Forms is for Plotinus the thought of Intellect. The human soul has a counterpart in Intellect, a partial mind which in fact is the true self and is that on which the soul depends. This has two interesting consequences for the doctrine of spiritual ascent: (1) the soul's ascent may be correctly described as the search after oneself and, if successful, as true self-knowledge, as fully becoming what one essentially is; (2) on account of Plotinus' doctrine about the interconnectedness of Intellect as a whole, gaining this self-knowledge and self-identity would also involve gaining knowledge of the realm of Forms as a whole.

Plotinus' views on classical Greek ethical topics such as virtues (see Aretē) and happiness (see Eudaimonia) are determined by his general position that intellectual life is the true life and proper goal of human beings. He devotes one treatise, I 2(19), to the virtues. The suggestion in Plato's *Theaetetus* (176a-b) that the virtues assimilate us with the divine is central to his views on them, and serves as his point of departure. The question arises how to reconcile this doctrine with the doctrine of the four cardinal virtues in the *Republic*. Plotinus distinguishes between political virtues, purgative virtues and the paradigms of the virtues at the level of Intellect. These form a hierarchy of virtues. The function of the political virtues (the lowest grade) is to give order to the desires. It is not clear, however, how these can be said to assimilate us to god (Intellect), for the divine does not have any desires that must be ordered and hence cannot possess the political virtues. Plotinus' answer is that although god does not possess the political virtues, there is something in god answering to them and from which they are derived. Furthermore, the similarity that holds between an image and the original is not reciprocal. Thus, the political virtues may be images of something belonging to the divine without the divine possessing the political virtues.

There are two treatises dealing with happiness or wellbeing: I 4(46) and I 5(36). In the former treatise Plotinus rests his own position on Platonic and Aristotelian doctrines, while criticizing the Epicureans and Stoics. He rejects the view that happiness consists at all in pleasure, a sensation of a particular sort: one can be happy without being aware of it. He also rejects the Stoic account of happiness as rational life. His own position is that happiness applies to life as such, not to a certain sort of life. There is a supremely perfect and self-sufficient life, that of the Intellect, upon which every other sort of life depends. Happiness pertains primarily to this perfect life that is in need of no external good. Since all other kinds of life are reflections of this one, all living beings are capable of at least a reflection of happiness according to the kind of life they have. Human beings are capable however of attaining the perfect kind of life, that of Intellect. In the latter treatise Plotinus holds with the Stoics that none of the so-called 'external evils' can deprive a happy person of their happiness and that none of the so-called 'goods' pertaining to the sensible world are necessary for human happiness (see Stoicism §§15-17). In I 5 he argues that the length of a person's life is not relevant to happiness. This is because happiness, consisting in a good life, is the life of Intellect and this life is not dispersed in time but is in eternity, which here means outside time.

7 Influence

Plotinus is one of the most influential of ancient philosophers. He shaped the outlook of the later pagan Neoplatonic tradition, including such thinkers as Porphyry and Proclus, and he left clear traces in Christian thinkers such as Gregory of Nyssa (see Patristic philosophy §5), Augustine and Boethius. Since all these were extremely influential in their own right, Plotinus has had great indirect impact. He clearly played a significant role in preparing for medieval philosophical theology. A forged extract from the *Enneads* was known in the Islamic world as the *Theology of Aristotle*. The supposed Aristotelian origin of this text helped the fusion of Aristotelianism and Neoplatonism that characterizes much of Arabic philosophy. Neoplatonism saw a revival in Europe during the Renaissance. A Latin translation of the *Enneads* by Marsilio Ficino first appeared in 1492 and gained wide distribution. Plotinus exerted considerable direct influence on many sixteenth-and seventeenth-century intellectuals. Even if the popularity of Neoplatonism and Plotinus receded in the seventeenth century, many individual thinkers since have read and been influenced by Plotinus, for instance Berkeley, Schelling and Bergson.

See also: Neoplatonism

EYJÓLFUR KJALAR EMILSSON

List of works

Plotinus (*c.*AD 205-66) *Enneads*, trans. A.H. Armstrong, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1966-88.(Based on the authoritative editions of P. Henry and H.-R. Schwyzer, with minor modifications; Greek text with English translation.)

References and further reading

Armstrong, A.H. (1940) *The Architecture of the Intelligible Universe in the Philosophy of Plotinus*, Cambridge: Cambridge University Press.(A classic study.)

Atkinson, M. (1983) *Plotinus: Ennead V.1 On the Three Principal Hypostases*, Oxford: Oxford University Press. (Very valuable commentary on a central treatise; abundant in references.)

Blumenthal, H.J. (1971) *Plotinus' Psychology*, The Hague: Martinus Nijhoff.(A detailed pioneering work.)

Blumenthal, H.J. (1987) 'Plotinus in the Light of Twenty Years' Scholarship, 1951-1971', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36.(Good bibliography.)

Bussanich, J. (1988) *Philosophia Antiqua*, vol. 49, *The One and its Relation to Intellect in Plotinus: A Commentary on Selected Texts*, Leiden: Brill.(Treats a central and important aspect of Plotinus' thought.)

Corrigan, K. and O'Cleirigh, P. (1987) 'The Course of Plotinian Scholarship from 1971 to 1986', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 1.(Good bibliography.)

Emilsson, E.K. (1988) *Plotinus on Sense-Perception: A Philosophical Study*, Cambridge: Cambridge University Press.(Analytic approach to Plotinus' philosophy of mind.)

Gerson, L.P. (1996) *Plotinus*, London: Routledge.(Extensive philosophical discussion of most aspects of Plotinus' thought.)

Lloyd, A.C. (1990) *The Anatomy of Neoplatonism*, Oxford: Oxford University Press.(A philosophically penetrating study, rewarding even if not an easy read.)

O'Meara, D.J. (1993) *Plotinus: An Introduction to the Enneads*, Oxford: Oxford University Press.(Highly recommendable introduction to Plotinus; contains a good bibliography.)

Rist, J.M. (1967) *Plotinus: The Road to Reality*, Cambridge: Cambridge University Press. (Essays on various Plotinian topics.)

Schroeder, F.M. (1992) *Form and Transformation: A study in the Philosophy of Plotinus*, Montreal: McGill University Press.(Contains several sensitively written essays on special topics.)

Pluralism

'Pluralism' is a broad term, applicable to any doctrine which maintains that there are ultimately many things, or many kinds of thing; in both these senses it is opposed to 'monism'. Its commonest use in late twentieth-century philosophy is to describe views which recognize many sets of equally correct beliefs or evaluative standards; and in this sense it is akin to 'relativism'. Societies are sometimes called 'pluralistic', meaning that they incorporate a variety of ways of life, moral standards and religions; one who sees this not as undesirable confusion but a proper state of things, espouses pluralism.

The principle famously known as Ockham's Razor - that we should posit no more entities than our purposes demand - expresses a disinclination to ontological pluralism, though not of course a promise that we can always dispense with it (see Ockham, W. §2; Durandus of St Pourçain). Kant believed such a disinclination to be a basic trait of human reason, which naturally seeks as much unity as possible in its view of reality (see Kant §8)). A similar evaluation emerges in Leibniz's view that God's creation contains a maximum of individual variety falling under a minimum of general laws.

Ironically, Kant's own philosophy sowed the seeds of pluralism in another sense, at least equally important and conceptually rather more difficult: that there may be many ways of interpreting the world and reacting to it, very different from each other yet equally justified. Kant did not himself take that view; but by making the world we experience something that we construct, and which therefore depends on the nature of our own mental workings, he invited the thought that at other times, or in other cultures, others might construct it differently.

At one level, Hegel accepted the invitation, declaring the forms of human thought and experience to be historically variable. But this was no true pluralism, since he held the variation to be progressive, and progressing towards the point at which consciousness will see things as they really are. So later ways of thinking, though still inadequate, are less inadequate than earlier ones, and Hegel's friendliness to pluralism is only provisional - behind it lies a firm commitment to the view that ultimately there is just one true way to see reality.

Monotheists are almost certain to be unfriendly to pluralism, since they will typically hold that there is a uniquely true set of beliefs and a uniquely right set of moral standards - those of the deity - so that any *de facto* pluralism is just an effect of human incapacity. Hence significant support for any thoroughgoing pluralism acquired real momentum only in the nineteenth century as thinkers, following the lead of D.F. Strauss and Feuerbach, began to view God as a projection of the human mind, or to abandon theism altogether. Nietzsche's 'perspectivalism' - truth exists only as truth from a particular perspective, and the Truth (capital T) does not exist at all - is one of its products (see Nietzsche, F. §4).

That conclusion was certainly connected with Nietzsche's view that belief and the closely related concept of truth are in fact instruments for advancing the believer's purposes ('Truth is that kind of error without which a species cannot live'); and a variant of this view is central to the pragmatism of William James. It leads very naturally to pluralism. For we are quite familiar with the idea that different individuals or groups may have different purposes and needs, as may the same individual or group at different times and in different circumstances. So if responsiveness to needs and purposes is what validates belief and/or determines truth, these may differ as well. Then we find ourselves with a plurality of truths, and a resulting tendency to hold that 'true', when properly used, must really mean 'true *for* S' (some person or group) and that where the relevant values of S differ, two 'truths' are not necessarily commensurable with each other (see Relativism).

One should distinguish between a true pluralism and the toleration of minority opinions or codes of behaviour. Toleration of deviant opinion is quite compatible with believing that there is just one set of truths. Non-interference can still be argued for on the grounds that the existence of certain false opinions and certain forms of bad behaviour are lesser evils than what would be needed to suppress them; or, as J.S. Mill (1859) argued, because we may be more likely to light on the truth if a variety of opinions are represented in our debates - a view one might call *methodological* pluralism. However, in going on to praise individuality as an essential component of human well-being, and to argue the case for allowing a private sphere immune from legal and social pressure, Mill was advocating a limited but genuine pluralism of life-styles. A concern of liberal political theory

has been to understand the arrangements needed to permit and support such pluralism while maintaining a stable society.

This is no purely theoretical exercise, for there are also social and political influences favouring certain types of pluralism. Many societies are now culturally and religiously diverse; and 'postcolonial' attitudes make for a far more egalitarian and less judgmental approach to other peoples' practices and beliefs. Sustaining such an attitude, where different cultures find themselves in close contact with each other, indeed for many purposes intermingled, is not without its problems; and it is not clear how far religious believers can sincerely regard other religions in so tolerant a way without in effect giving up their own.

See also: Anthropology, philosophy of; Citizenship; Monism; Multiculturalism; Postmodernism; Pragmatism; Relativism; Religious pluralism; Toleration

EDWARD CRAIG

References and further reading

James, W. (1904) 'Humanism and Truth', *Mind* 13 (October): 457-75; repr. in *Pragmatism* and *The Meaning of Truth*, ed. A.J. Ayer, Cambridge, MA: Harvard University Press, 1978.(Readable statement of the pragmatic view of the nature of truth, from which a pluralism naturally emerges.)

Mill, J.S. (1859) *On Liberty*, London: Dent, 1972.(See chapter 2 for what is called above 'methodological pluralism', chapter 3 for something more like pluralism proper: the value of individuality.)

Nietzsche, F. (1883-8) *The Will to Power*, trans. W. Kaufmann and R.J. Hollingdale, New York: Vintage, 1968. (Sections 553-69 exemplify the pluralistic 'perspectivalism' widely found in Nietzsche's writings.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(Argues for principles and a polity which citizens may reasonably endorse in spite of having diverse or pluralist substantive values.)

Plutarch of Chaeronea (*c.* AD 45-*c.*120)

The Greek biographer and philosopher Plutarch of Chaeronea is the greatest Greek literary figure of the first century AD. He is properly called Plutarch of Chaeronea, to distinguish him from the minor fourth-century AD Platonist Plutarch of Athens. His fame rests not so much on his contributions to philosophy as on those to history and biography. Indeed, despite the survival of a large body of philosophical and semi-philosophical writings known under the collective title Moralia, most of his more technical philosophical treatises have perished. Nevertheless, his importance for our understanding of the development of Middle Platonism is great. Plutarch is a reasonably orthodox Platonist (in so far as that expression has any meaning), although his Platonism has some distinctive features. Against Antiochus, he accepts the sceptical New Academy as part of the Platonic tradition, but he also exhibits a degree of cosmic dualism (postulating a pre-cosmic evil soul) which goes rather beyond the Platonist norm. It is misleading, however, to oppose him to a supposed tradition of 'school-Platonism'.

*As the ethical 'end' or 'goal' (*telos*), he adopts the normal later Platonist one of 'likeness to god'. In ethics, as in logic, he tends to favour Aristotelianism rather than Stoicism (advocating, for example, moderation of the passions rather than their extirpation, and appropriating as authentically Platonic the Aristotelian categories and syllogistic). A tendency to favour New Academic scepticism seems indicated in the titles of some of his lost works, but is not very evident in the surviving ones.*

As first principles, he postulates a pair consisting of God - who is one, the Good, and really existent - and the Platonic-Pythagorean Indefinite Dyad, which is a principle of multiplicity, and ultimately material. As secondary principles, he seems to adopt a logos*, or active reason-principle of god, although the evidence for this is not copious, and a world-soul, which is essentially irrational but desirous of 'impregnation' with reason by the* logos*. There is also in the universe, however, an active principle of disorder, which can never be entirely mastered by the divinity. A hint of Persian dualism seems to enter here into Plutarch's thought. Such a system is derivable above all from his essay On Isis and Osiris, which may not be entirely typical. Other distinctive features include a tendency to triadic divisions of the universe, a developed demonology, and an interesting combination of Aristotelian and Stoic logics.*

1 Life and works

Plutarch of Chaeronea is the dominant figure in Platonism, and in Greek literature as a whole, in the latter part of the first century AD, and the first quarter of the second (see Platonism, Early and Middle). Born probably around AD 45 (he presents himself as a beginning student of philosophy in Athens in AD 66/7), he lived until at least AD 119, when he is reported to have been appointed Procurator of Achaea (presumably a largely honorary post) by the emperor Hadrian. He came from the small town of Chaeronea in Boeotia, where his family had been prominent for many generations, and this connection remained important to him throughout his life, as also did his connection with Delphi, where he was elected in later life a priest of Apollo. In Athens he studied under a Platonist named Ammonius (not to be confused with Ammonius, son of Hermeas), of Egyptian origin, who became a prominent citizen of Athens, and to whom Plutarch seems to have owed a degree of cosmic dualism which colours his Platonism. In later life, after undertaking some travel, notably to Rome, he settled down in Chaeronea, presiding over a 'mini-Academy' of his friends and relations, and taking an active part in local politics, including those of Delphi. Most of his dialogues are set in either Chaeronea or Delphi.

A vast body of work by Plutarch has come down to us, much of it, however, non-philosophical - most notably his 'parallel lives' of famous Greeks and Romans - although philosophical reflections may be found anywhere in his work. On the other hand, many of his more technical philosophical works, some with tantalizing titles, are lost (discarded, no doubt, in favour of Neoplatonist treatments of the same subjects, making a full evaluation of his philosophical position more difficult. He was, however, a regular (one hesitates to say 'orthodox') Platonist. It would be wrong to regard him as a maverick, or an 'eclectic'.

His chief surviving works of philosophical interest are the following: *On Isis and Osiris*, *On the E at Delphi*, *On the Oracles at Delphi*, *On the Obsolescence of Oracles*, *Is Virtue Teachable?*, *On Moral Virtue*, *On Delays in Divine Punishment*, *On the Daemon of Socrates*, *On the Face in the Moon*, *Problems in Plato*, *On the Creation of the Soul in the Timaeus*. Interesting things are said elsewhere, scattered, for example, through the biographies, in

the nine books of *Table Talk*, or in the many other essays which, although Plutarch would have considered them ’philosophical’, we would tend to regard as merely ’literary’. We also have a number of treatises of a polemical nature, directed against other schools, notably the Stoics and the Epicureans. Against the Stoics are: *On the Contradictions of the Stoics* and *On Common Conceptions [koinai ennoiai], against the Stoics*; and against the Epicureans: *That one cannot live happily following Epicurus*, *Against Colotes* (Colotes had attacked the implicit scepticism of Plato and others) and *The Doctrine ’Live in Obscurity’* [*lathe biōsas*, the basic Epicurean injunction].

This is a vast corpus, and certain variations of doctrine undoubtedly emerge, but basically Plutarch can be seen as a consistent Platonist, well within the parameters of what might be considered ’orthodoxy’ - bearing in mind that there was no canon of orthodoxy, and that a wide latitude of doctrine was acceptable within the tradition.

2 Ethics

As regards the ‘end’ or ’goal’ (*telos*) of life, for Plutarch, as for all later Platonists after Antiochus, the supreme object of human endeavour is ’likeness to god’ (as opposed to the Stoic ‘conformity to nature’, adopted by Antiochus (see Platonism, Early and Middle §8). This is achieved through the exercise of the virtues, but particularly through the acquisition of theoretical wisdom (*sophia*), rather than practical wisdom. We find this well expressed in a passage of the dialogue *On Delays in Divine Punishment* (550d), where Plutarch himself is the speaker. Here, in fact, the *telos* of conformity to nature is artfully subsumed under the ’higher’ one of likeness to god by the statement that ’Nature kindled vision in us’, so that the soul should come to a knowledge of god.

On the subject of virtue and happiness, Plutarch inclines on the whole to the more ’broad-minded’ ethical position of Antiochus, as against the Stoic-Pythagorean asceticism (see Neo-Pythagoreanism) observable in such figures as Eudorus and Philo of Alexandria. Significantly, his terminology in this area is Aristotelian rather than Stoic. In his essay *On Moral Virtue* we have a useful statement of his ethical theory. It takes the form of an attack on the Stoic position that the soul is unitary, and that there is no such thing as a distinct irrational part. Moral virtue is to be distinguished from theoretical virtue, in that it has emotion (*pathos*) as its subject matter and reason as its form, reason acting to harmonize the essentially disorderly passions, and thus producing a set of virtues as ’means’. This is really Peripatetic theory (see Aristotle §22) overlaid with Pythagorean influences.

Peripatetic too is Plutarch’s position that all three types of good - spiritual, bodily and external - make their contribution to happiness (*On Common Conceptions* 1060c-), aligning him once again with Antiochus against Eudorus. In this passage he attacks Chrysippus for not admitting bodily and external goods as an essential component of happiness, although nature commends them to us (see Stoicism §§15-6).

3 Metaphysics - first principles

Plutarch’s view of the supreme principle, or God, is just what one would expect of a Platonist: God is real being (*to ontōs on*), eternal, unchanging, non-composite, one (*On the E at Delphi* 392e). He knows all things and directs all things (*On Isis and Osiris* 351d, 382b), thus exercising providence (*pronoia*). He is also ’the object of striving for all nature’ (*On the Face of the Moon* 944e), and thus is aligned with the Aristotelian prime mover (see Aristotle §16). To this ’Monad’ is opposed an ’Indefinite Dyad’, as ’the element underlying all formlessness and disorder’ (*On the Obsolescence of Oracles* 428f; *On Isis and Osiris* 369c). This is not simply matter, but a more actively disorderly principle which acts on matter, in opposition to the Monad, to produce the differentiation and multiplicity of beings. It is this principle which Plutarch sees as active in the pre-cosmic chaos described in Plato’s *Timaeus*. The interaction of these two primal forces generates, first, number (and so the system of Forms) and then the physical world. In Plutarch’s emphasis on the Dyad (and on the existence of an ’evil’ world-soul) there is discernible an element of dualism which is more than Platonic, although its derivation is unknown (see also his comparison of Monad and Dyad with Ormuzd and Ahriman the warring supreme divinities of Zoroastrian religion, at *On Isis and Osiris* 369e).

The first principle creates, and relates to, the world through suitable intermediaries, primary among which is the *logos* (cosmic ‘reason’) (see Logos §1). Originally a Stoic concept, *logos* appears in a Platonist context already in the thought of Philo of Alexandria. It has been doubted whether Plutarch really maintained a *logos* doctrine, but the evidence that he did seems sufficient. In a passage of *On Isis and Osiris* (373a-b), for example, two aspects, or moments, of the god Osiris are distinguished as his soul and his body. His soul is ’eternal and indestructible’,

whereas his body is repeatedly torn asunder by Typhon (the Indefinite Dyad) and is constantly being reassembled by Isis (the world-soul). The ’body’ of Osiris, as it turns out, is the *logos* in its immanent aspect, in matter, whereas his soul is the transcendent *logos*, as the sum total (and organizing principle) of the Forms in the mind of god.

As for the status of the Forms as thoughts of God (the standard Middle Platonic view), we may refer to the passage in the essay *On Delays in Divine Punishment* mentioned in §2, where we see God himself, as the totality of the Forms, presented as the model (*paradeigma*) for the physical world, and for mankind in particular. This seems based on an interpretation of Plato’s *Timaeus*, according to which god (as the creator or ‘demiurge’) is his own *paradeigma* (instead of looking to it as something external). Unfortunately, we lack Plutarch’s essays *Where are the Forms?* and *How does Matter Participate in the Forms? It Makes the Primary Bodies*, but we can derive some idea of their contents both from this passage and from the title of the latter essay. Presumably what Plutarch means by ’it makes the primary bodies’ is that matter (with which Plutarch has no hesitation in identifying Plato’s ’receptacle’ in *Timaeus*), on being ‘irradiated’ by form, constitutes the basic triangles of which the four elements are composed.

Besides the general theory of Forms, Plutarch produces in *On the Obsolescence of Oracles* a remarkable doctrine connecting the five ’greatest kinds’ of Plato’s *Sophist* with the basic constituents of the physical world, namely, the four elements plus the physical world as a whole. He is here taking the five basic molecules described by Plato in the *Timaeus* (53c-56c) - of which the cube represents earth, the pyramid fire, the octahedron air and the eikosihedron water - and relating them respectively to rest, motion, identity and difference, while the dodecahedron, which represents the physical world as a whole, corresponds to being. Whether this is an original contribution of Plutarch’s is not clear, but it anticipates interestingly the use of the ’greatest genera’ by Plotinus as a set of ’categories’ for the intelligible world (see Plotinus §4).

The celebrated ’dualism’ of Plutarch comes out in his description of the Forms as ‘seized by the element of disorder and confusion which comes down from the region above’ (*On Isis and Osiris* 373a). This seems to imply not just the rather negative unruly material principle of the *Timaeus*, but a positive force, a ’maleficent soul’, which has at some stage itself already broken away from the intelligible realm. We seem thus to be brought close to Gnostic beliefs, but Plutarch claims the authority of Plato for this. In *Laws* (X 896d), Plato seems, at least, to postulate, in opposition to the beneficent world-soul, another ’of the opposite capacity’, which is responsible for all irrational motion in the universe. Anything which is soul is also alive and self-moving, so that a considerable step has been taken beyond the inanimate principle of ’otherness’ or ’necessity’ postulated in the *Timaeus*.

4 Psychology

This ’negative’ or ’evil’ principle in the universe, however, must be distinguished from the world-soul proper, which, although essentially irrational, is a ’positive’ force, in the sense that it strives for ordering (or in mythical terms, ’impregnation’) by the *logos*. Plutarch’s position on this, however, is somewhat complex. In *On Isis and Osiris*, Isis is equated with the ’receptacle’ of the *Timaeus* and with matter, as well as with wisdom, and takes on very much the same role as does wisdom in the earlier system of the Platonizing Jew Philo of Alexandria, suggesting a tendency in Alexandrian Platonism (if we may connect Plutarch with that through Ammonius) to identify at least the positive aspect of matter with the world-soul, and to connect them both with the Pythagorean/Early Academic Dyad, which is not properly an ’evil’ principle, as is the ‘Seth-Typhon’ figure in Plutarch’s system. This amalgam produces an entity which is on the one hand ’fallen’ and imperfect (although filled with longing for perfection), on the other the cause of our creation and the vehicle by which we can come to know God. In his treatise *On the Creation of the Soul in the Timaeus*, Plutarch presents us with the remarkable image of a ‘slumbering soul’, which is periodically roused to rationality by the ’better element’ (the *logos*). This seems to be inspired by a non-literal interpretation of the myth of Plato’s *Statesman*, since it becomes plain from what follows (1027a) that the process of arousal is actually not periodic but continuous. We have therefore an essentially irrational world-soul, completely interwoven with a rational soul or *logos*, thus maintaining a constant cosmic tension.

The belief in an eternal, independent irrational soul naturally has consequences for Plutarch’s views on the composition both of the world-soul and of the individual soul, as described, in particular, in the *Timaeus*. Indeed, Plutarch devotes to this question a whole treatise, *On the Creation of the Soul in the Timaeus*, constituting a

detailed exposition of his views, in the form of an extended commentary on the *Timaeus*. The specific problem concerns what we are to understand by 'the substance which is divisible about bodies' in the description of the elements of the soul. Plutarch criticizes his Platonist predecessors, such as Xenocrates §2 and Crantor, for not recognizing that this 'divisible substance' must in fact be the irrational 'disorderly' soul that animates the pre-cosmic chaos (*On the Creation of the Soul* 1014b). So this disorderly element which Plato calls 'necessity' (*Timaeus* 48a, 56c, 68e) cannot be taken as something characterless, such as matter, but must be an active principle, the disorderly or 'maleficent' soul. We must distinguish, he argues, between matter, as described in *Timaeus* (50e), which is bereft of all quality and power of its own, and this refractory element in the universe, which has distinct powers and qualities. He adduces Plato (*Philebus* 24a and *Statesman* 273b) to fortify his position - which depends, however, on the premise of the unity of Plato's thought.

A noteworthy doctrine of Plutarch's is his strong distinction between the human soul and the human mind. Of course, such a distinction had always been present in some form in philosophic speculation, and perhaps most notably in Aristotle's *On Soul*, but Plutarch seems to give it a special twist. In *On the Face in the Moon*, we find the assertion that mind is just as distinct from soul as soul is from body. We may note here, apart from the unequivocal tripartition of the human being, a notion of three terms and two conjunctions - body and soul producing the irrational soul (*to alogon*), soul and mind producing the reason (*logos*). The rational soul, then, as our centre of consciousness, is not to be seen as mind in its pure state, but rather as consisting in the conjunction of pure mind with something lower and essentially irrational. This doctrine takes on a further twist in *On the Daemon of Socrates*, where we find the mind presented as in a way *external* to the body, presiding over it as its 'daemon', while different individuals have very different relations to their highest element, some allowing it to float freely 'above' them, others dragging it down into contamination with bodily concerns. This owes something, perhaps, to Aristotle's *On the Soul* III 5, but also to Plato's *Timaeus*. The latter (90a) refers to 'the most authoritative element of our soul' as a daemon, which God has given to each one of us, and which raises us up from earth towards our kindred in the heavens. Nevertheless, Plutarch's doctrine is interestingly distinctive, finding analogues in certain treatises of the *Corpus Hermeticum*, Treatises I and X in particular.

This threefold division of the individual has its equivalent on the cosmic level, in the form of a threefold division of the universe, such as we find in *On the Face in the Moon*, where Plutarch bases his theory on one of Xenocrates' concerning the combination of three 'densities' or 'tunings' of matter with the elements fire, air and earth respectively. Plutarch also proposed another, more complex, theory, which can be found in *On the Daemon of Socrates*. Here, the Monad, mind and nature (which could be taken as the irrational soul) are presented as three 'links' holding together four levels of being. How seriously Plutarch took, or means us to take, these constructions is uncertain, but they do serve as evidence for the existence of interesting lines of speculation. In particular, the identification in the latter passage of the highest principle as a Monad superior to a mind finds something of a parallel in the theological chapter 10 of the *Didaskalikos* of Alcinous.

5 Daemonology

The more transcendent the supreme deity becomes, the more he stands in need of other beings to mediate between him and the material world, over which, in Platonism, he always exercises a general supervision (*pronoia*). The *logos* fulfils this function, but Plutarch, like all other Platonists, also postulates a daemonic level of being, and his theorizing about these entities is more elaborate than is found elsewhere. For Plutarch, both god's transcendence and his providential care for the world must be preserved, and the universe can tolerate no sharp divisions or sudden transitions. The basic inspiration for this concept of daemons in linking together the incompatible extremes of the universe is a famous passage in Plato's *Symposium* (202e), but the details were obviously worked out further by his follower Xenocrates, who is a major source for Plutarch.

In the dialogue *On the Obsolescence of Oracles*, for example, we find a geometrical elaboration of the basic Platonic doctrine attributed to Xenocrates, and commended by the speaker, according to which daemons are compared to isosceles triangles, median between the equilateral of the gods and the scalene, representing humanity. Earlier Plutarch's adoption of Xenocrates' tripartition of the world in *On the Face in the Moon* was noted. Here the moon, which served there as the focus of the median realm, and as the proper place of souls (and indeed as the symbol of the world-soul), is now made to serve also as the proper place of daemons, which is suitable to their mediating status.

Plutarch, like Xenocrates, recognizes the existence of both 'good' and 'evil' daemons, and in this connection we must distinguish a complexity, and possible incoherence, in his account of them. According to one view, daemons are souls in the process of either ascent from or descent into incarnation; according to another, they are permanent members of a spiritual hierarchy. Plutarch seems to entertain, at various points, both what one might term a 'static' and a 'dynamic' theory of daemons. A good statement of the latter view occurs in *On the Obsolescence of Oracles*, which describes a constant process of transmutation of human souls into daemons and 'heroes' (a distinct category in later Platonism), and even occasionally (as in the case of Heracles) into gods. There may well be a degree of incoherence in Plutarch's theory. It is really not clear whether some souls remain permanently disembodied (to constitute either good or evil daemons), or whether all are subject to the process of ascent and descent.

It remains to speak of the personal or guardian daemon, the most notable example of which, for later Platonists, was the 'daemon' of Socrates. In Plutarch's essay *On the Daemon of Socrates*, we find a vivid description of how the guardian daemon can assist an individual. These guardians are, it seems, all souls who have been through incarnation themselves, and have now earned their release from the cycle of rebirth; they are not souls that have never suffered incarnation. Only noble souls are qualified to serve as guardians.

In connection with his theory of the guardian daemon, Plutarch indulges in some interesting speculations about the mode of contact between the daemonic and the human mind, and between mind and matter in general (*On the Daemon of Socrates* 588f-589b). This is a problem which had not bothered Platonists, from Plato onwards, as much as it should have, and Plutarch does not really address the true difficulties (he simply speaks with wonder of the 'slight impulse' by which the mind sets the bones and sinews in motion), but at least he raises the question. Only Plotinus, later, really tries to address the problem of the mode of interaction between the material and the immaterial (for example, in *Enneads* III 6(26).1-5).

6 Logic and epistemology

Technical logic was not among Plutarch's more vital concerns, and there is in his extant works not much to indicate which system he followed, although there is no reason to think that it differed much from that set out somewhat later in Alcinous' *Didaskalikos* (ch. 6), which is itself a basic account of Middle Platonic logic. Plutarch did compose a *Lecture on the Ten Categories*, now lost, of which one would gladly have known the contents. Did he remain true to Aristotle, or did he, perhaps, adopt some of the Pythagoreanizing 'corrections' introduced by Eudorus, such as the placing of quality before quantity? We also find in the extant list of Plutarch's works two other treatises whose titles sound logical, *A Reply to Chrysippus on the First Consequent* and *On Tautology*, as well as some works on rhetoric, which was counted as a part of logic. Apart from revealing a tendency to attack the Stoics, which is what we should expect from Plutarch, these titles do not tell us very much.

That Plutarch regarded Plato as being already in possession of the so-called Aristotelian logic is indicated in *On the Creation of the Soul*, which is a commentary on *Timaeus* 37a-b, a passage in which most of the Aristotelian categories might be discerned, if viewed with the eye of an adherent. Sceptics might object to the equating of the examination of identity and difference with quantity and quality, and the equating of a thing's *existence* in a certain sense with its *activity*. A loyal Platonist could, however, argue that any distinguishing of a thing from other things must involve a statement of its quality and quantity (as indeed Aristotle himself notes in the *Categories* (11a15)), and that, on Plato's own principle, as laid down in the *Sophist* (248c), existence means acting and being acted upon. Furthermore, the Aristotelian categories 'being-in-a-position' and 'having' can conceivably be viewed as latent in the latter part of the passage. Plato, then, in Plutarch's view, knows the categories, but is not concerned with enumerating them anywhere as baldly as did his famous pupil. On the basis of these categories, Plutarch declares, Plato constructs his epistemology, according to which the soul, when operating on the level of sense-perception, produces opinions which are fine and true, and when operating on the level of pure reason, intelligence and knowledge. These true opinions and beliefs, a Platonist would claim, are naturally expressed in syllogistic form. Alcinous, in *Didaskalikos* ch. 6, discerns various syllogisms in the first and second hypotheses of the *Parmenides*. Plutarch would doubtless have agreed with him, but we lack any evidence on this. In another passage, however, from the polemical treatise *Against Colotes* (1115d-), Plutarch appears to identify Plato's discussion of non-being as 'otherness' in the *Sophist* (255a-258e, especially 257b) as providing a stimulus to both Aristotelian and Stoic logic. Aristotle and the Stoics, the argument runs, seized on the concepts of 'not-being' and

'otherness' discussed in the *Sophist*, divested them of their metaphysical implications, and constructed purely logical systems out of them.

In projecting the Aristotelian and Stoic logical systems back into Plato, Plutarch is in agreement, as has been said, with what we find in Alcinous, and also among other near-contemporaries, such as Apuleius and Galen. There is no sign in his extant works of the disposition to attack Aristotelian logic, in particular the categories, that is evident in Eudorus before him, or in his own follower Atticus. On the other hand we do find, in *On the E at Delphi*, an interesting commendation of the Stoic hypothetical syllogism as the basic component of human reasoning, as stated by Plutarch's friend Theon. He seems to rank it as superior to the Aristotelian, but this may simply be a consequence of the context, which concerns the elucidation of the Delphic E, here interpreted as 'if' (*ei*).

JOHN DILLON

List of works

Plutarch (late 1st-early 2nd century AD) *Lives*, ed. B. Perrin *et al.*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1914-26, 11 vols.(Greek text with facing translation: comparative pairs of biographies of major figures from Greek and Roman history.)

Plutarch (late 1st-early 2nd century AD) *Moralia*, ed. F.C. Babbitt *et al.*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1927-76, 16 vols.(Greek text with facing translation: essays, speeches and treatises on a wide range of issues, largely philosophical. Of particular importance are volumes 5, 7 and 12-13.)

References and further reading

Alcinous (*c*. 2nd century AD) *Didaskalikos tōn Platōnos dogmatōn*, in J. Dillon (ed.) *Alcinous, The Handbook of Platonism*, Oxford: Clarendon Press, 1993.(English translation, with introduction and philosophical commentary.)

Babut, D. (1969) *Plutarque et le stoicisme*, Paris: Presses Universitaires de France.(Very useful on Plutarch's relation to Stoicism.)

Brenk, F.E. (1977) *In Mist Apparelled: Religious Themes in Plutarch's Moralia and Lives*, Leiden: Brill.(A good discussion, particularly on daemonology.)

Brenk, F.E. (1987) 'The Religious Spirit of Plutarch', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36: 1, 248-349.(Updating of previous work.)

Dillon, J. (1977) *The Middle Platonists*, London: Duckworth.(Chapter 4 is an introductory account of Plutarch.)

Froidefond, C. (1987) 'Plutarque et le Platonisme', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36: 1, 184-233.(A good survey, but, disputably, sees *On Isis and Osiris* as the culminating expression of Plutarch's metaphysics.)

Jones, R.M. (1916) *The Platonism of Plutarch*, Menasha, WI: Banta; repr. New York: Garland, 1980. (Oversimplified and out of date, but still of some value.)

Plato (*c*.366-360 BC) *Timaeus*, trans. F.M. Cornford, *Plato's Cosmology*, London: Routledge & Kegan Paul, 1937. (Plato's hugely influential work on cosmology.)

Thévenaz, P. (1938) *L'Ame du monde, le devenir, et la matière chez Plutarque (World-Soul, Generation and Matter in Plutarch*), Paris: Les Belles Lettres.(Includes a translation of the first part of the *On the Creation of the Soul*.)

Ziegler, K. (1951), 'Plutarchos von Chaironeia', in Pauly-Wissowa (ed.) *Realencyclopädie der Klassischen Altertumswissenschaft*, 21: 1, cols 636-962.(An excellent general survey; also published separately.)

Pneuma

Pneuma*, ’spirit’, derives from the Greek verb* pneo*, which indicates blowing or breathing. Since breathing is necessary for life and consciousness,* pneuma *came to denote not only wind and breath but various vital functions, including sensation and thought, and was understood by some philosophers as a cosmological principle. It became especially important in Stoicism, which explained the world in terms of matter and the rational structure exhibited in all its forms; this is established by rhythmical variations in the* tonos *or ‘tension’ of the* pneuma*.*

In Hebrew tradition, where Greek was used, pneuma *stood for life, consciousness, and for invisible conscious agents, angels or demons. In Christian thought it denotes divine inspiration, in particular the Holy Spirit acknowledged as a divine Person. At John 4:24 it is used, unusually, to describe God himself.*

1 Early Greek philosophy and medicine

Anaximenes (§1), the third in the traditional list of philosophers, regarded air as the basic substance of the universe, which it penetrated and controlled like breath in the human body; it was thus divine, or the source of divinity. Other forms of matter derive from it by condensation (wind, water, earth) or rarefaction (fire). In most accounts it is named as air, *aēr*, but *pneuma* is once found, and is a likely alternative.

Diogenes of Apollonia (§§2-3) also regarded air as the universal source, an imperishable and intelligent divine being which prescribes the orderly structure of the world: ’All things see and hear and have intelligence from this same (thing)’. It constitutes the soul (see Psychē) of all living creatures; and an elaborate description of the veins in the human body probably served to show how it was distributed. Theophrastus summarizes Diogenes’ views of sensation and emotion; apparently the air within us assumes various qualities which it transmits to the brain. There is no explicit reference to *pneuma*; but in fragment 4 ’men and other animals live by the air which they breathe’ (*anapneonta*).

Meanwhile Empedocles (§6) had also taught that air is diffused through the body (fr. 100), but without giving it a primary role in sensation; rather, each of the four elements is perceived by the corresponding element within us (fr. 109); but these are mixed together in the blood, so that consciousness resides in the heart (fr. 105). Thereafter physicians centred on the island of Cos continued to regard the brain as the central directive organ, while the Sicilian school located it in the heart, from which *pneuma* proceeds along with the blood.

Plato uses *pneuma* to denote both wind and breath. At *Theaetetus* 152b people’s differing perceptions of the same wind (*pneuma*) as warm and cold are used to illustrate the subjectivity of sense perception. At *Phaedo* 70a, 77e, the theory of a breath-soul is gently derided. The physiology of the *Timaeus* probably reflects the teaching of the Sicilian school conveyed by Plato’s contemporary Philistion, particularly in the description of diseases; 82e refers to noxious *pneuma* in the veins; 78-9 discusses the physics of respiration, and choking sensations are mentioned at 84e and 91e, but there is no general attempt to connect *pneuma* with consciousness or muscular control.

Sicilian influence can also be found in Aristotle’s contemporary Diocles of Carystus and in Aristotle himself. Both distinguish two species of *pneuma*: the air which is breathed in and regulates the temperature of the body, and a psychic *pneuma* regarded as the seat of vital warmth which resides in the heart, once again identified as the seat of consciousness (see Strato §5).

2 The Stoics

The Stoics gave the concept of *pneuma* its most comprehensive development (see Stoicism §3). It was described as a blend of air and fire, the two active elements; but an alternative theory conceived of it as the substance underlying the four elements, although its fineness and rapid movement are still emphasized. *Pneuma* varies in its *tonos* or intensity; and rhythmical variations of *tonos* can exhibit structural organization and thus express reason itself, as we now know the human voice does. Rationality is thus present at various levels; in the consistency of physical substances, in the self-preservation of plants and animals and in human reason, itself identified as a physical system, the ’directive principle’, *hēgemonikon*, located in the heart, assuming the old theory of the arteries as air-channels by which pneumatic messages are accepted and delivered. Sight as well as hearing was explained by impulses of *pneuma*.

Human reason itself is a lower analogue of the cosmic reason which penetrates every level of being such that the cosmos is compared to a living organism whose parts exhibit orderly coherence. In its physical aspect that reason is described as 'constructive fire' which generates the four elements and the orderly structure of the world, and persists when this structure is resolved back into it in a universal conflagration, from which a new and similar world will evolve. In contrast, the human reason can survive death, but not persist indefinitely.

The Stoics acknowledged the traditional Greek cults, sometimes identifying the cosmic reason ('intelligent *pneuma*') with Zeus and explaining the other gods as physical elements (for example, Hera as air, *aēr*) (see Stoicism §5). Since the world existed for the benefit of rational creatures, especially mankind, some later Stoics conceived Zeus in theistic terms as kindly and protective. Less perceptive critics, including many Christians, caricatured the Stoics as mere pantheists who identified God with the material universe.

Among medical writers Erasistratus distinguished between the vital *pneuma* centred in the heart and the psychic *pneuma* residing in the brain. A similar doctrine was taught by Galen, who is otherwise critical of Erasistratus. Neither kind of *pneuma* is identified with the soul; the psychic *pneuma* is its primary instrument. (For the Pneumist school of medicine see Hellenistic medical epistemology §1.)

3 Jewish and Christian thought

Stoic influence appears in the biblical book of Wisdom (7: 22), which represents the divine wisdom as a distinct personality endowed with intelligent spirit (*pneuma noeron*). In Philo of Alexandria (§4) Platonic ideas predominate, and the Logos assumes more importance; in *Creation* 1.24 it is identified with the 'intelligible world' (see Logos §2). *Pneuma* however reproduces two distinct words employed in Israelite tradition; *ruach* is a god-given life-giving principle which assumes personal characteristics only when located in a human body; *nephesh* denotes this principle as personalized. Philo used *pneuma* to denote both wind and breath, which (1) permeates the whole body, (2) is despatched through the windpipe for speaking, (3) is the material of the *hēgemonikon*, (4) is identified with divine inspiration conveying truth and knowledge. Philo does not, like the Stoics, use *pneuma* to denote God himself, although at *Giants* 27 it is given comparable epithets.

Pneuma is freely used in the New Testament, usually with little philosophical reference; but St Paul introduces a triad of 'spirit, soul and body' (1 *Thessalonians* 5: 23), perhaps a modification of the Platonic intellect, soul, body, suggested by the use of *pneuma* for *ruach*. Valentinians and other Gnostics used this Pauline triad in a doctrine of three classes of people: the Gnostic elect, ordinary Christians and pagans, all of whom had appropriate destinies. Other Gnostics (Sethians and Basilides) used *pneuma* to denote a principle intermediate between light and darkness, or between the highest reality and gross matter.

New Testament usage includes the commonplace 'wind' and 'breath', the latter also regarded as a life-giving force surrendered at death. Its use by St John (4: 24) to denote God himself is unusual; but it commonly denotes discarnate personal agents, both beneficent and malign, as well as the human personality or its various moods and affections, including divine inspirations, which are called, or originate from, the 'Holy Spirit' (or 'Spirit of God'), seen in the divine presence manifested in Jesus and present derivatively in his disciples and the Church at large. The Holy Spirit is linked with God and Christ in a trinitarian confession, at first informally (2 Corinthians 13: 14, *c.* AD 55), but soon in a credal formula (Matthew 28: 19, perhaps *c.* AD 100, but recalling established usage).

The Apologists and other second-century Christians adopted this trinitarian confession, but gave the Spirit little philosophical consideration, while emphasizing the divine Logos, as in Philo's Platonism, as God's intermediary. *Pneuma* however sometimes denotes the divine element in Christ, who is 'God according to the (his?) *pneuma*'. Origen (§3) carefully discusses all three divine Persons and distinguishes their functions, once confining the Spirit's influence to the Church, whereas the Logos instructs all rational beings in obedience to the Father. This strongly hierarchical Trinity slightly resembles the triadic theology developing in Platonic circles; Eusebius could quote Numenius with approval. But the terminology is quite different; and in Plotinus (§4) the third hypostasis, Soul, corresponds, if at all, with the Christian Logos; and *pneuma* is used, *inter alia*, for a principle which unites Soul with matter.

The Creed of Nicaea (AD 325), gave the Spirit only formal recognition; but a revival of Origenistic ideas soon encouraged a more concrete presentation: and *c.* AD 360 a controversy led to recognition of the full and equal divinity of the three divine Persons. Thereafter the Spirit's distinctive impact was recognized in devotional

literature more than in the Church's formal theology.

See also: Qi

CHRISTOPHER STEAD

References and further reading

Bauer, W. (1957) *A Greek-English Lexicon of the New Testament*, trans. W.F. Arndt and F.W. Gingrich, Chicago, IL: University of Chicago Press, Πνευμα (*pneuma*), 680-5; repr. 1979, 674-8.(Good brief analysis with some notes on non-Christian usage; requires some knowledge of Greek.)

Burton, E.W. (1918) *Spirit, Soul and Flesh*, Chicago, IL: University of Chicago Press.(Useful source book from Aeschylus to the New Testament; Greek texts with English translations.)

Kittel, G.F. (ed.) (1968) *Theological Dictionary of the New Testament*, Πνευμα (*pneuma*) and Πνευματικ[oacgr]ς , Grand Rapids, MI: Eerdmans, vol. 6, 334-451.(Full scale survey, including *pneuma* in the Greek world, besides biblical and Gnostic usage; English translation from the German.)

Sandbach, F.H. (1975) *The Stoics*, London: Chatto & Windus.(Well written introductory account; consult the index entry *pneuma*.)

Verbeke, G. (1974) *Pneuma*, in Ritter, J., *Historisches Wörterbuch der Philosophie*, Basle: Schwabe, vol. 3, Geist, II, cols 157-62.(Authoritative brief study in German.)

Poetry

Though poetry today seems a relatively marginal topic in philosophy, it was crucial for philosophy's own initial self-definition. In ancient Greece, poetry was revered as the authoritative expression of sacred myth and traditional wisdom. With Socrates and Plato, philosophy began by distinguishing itself from poetry as a new, superior form of knowledge which could provide better guidance for life and even superior pleasure. Just as the sophists were attacked for relativism and deception, so were poets stridently criticized for irrationality and falsehood. For Plato, not only did poetry stem from and appeal to the emotional, unreasoning aspects of human nature; it was also far removed from truth, being only an imitation of our world of appearances which itself was but an imitation of the real world of ideas or forms. He therefore insisted that poets be banished from his ideal state because they threatened its proper governance by reason and philosophy.

Subsequent philosophy of poetry has been devoted to overcoming Plato's condemnatory theory, while tending to confirm philosophy's superiority. This task, begun by Aristotle, was for a long time pursued primarily under Plato's general model of poetry (and indeed all art) as imitation or mimesis*. The main strategy here was to argue that what poetry imitates or represents is more than mere superficial appearance, but rather general essences or the ideas themselves. For such theories, poetry's relation to truth is crucial. Other theories were later developed that preferred to define and justify poetry in terms of formal properties or expression, or its distinctively beneficial effects on its audience. These strategies became increasingly influential from the time of Romanticism, but can be traced back to more ancient sources.*

The vast majority of theories follow Plato in treating poetry as a distinct domain, separate from and subordinate to philosophy. But since Romanticism, some have argued for the essential unity of these two enterprises. Great philosophy is here seen as the poetic creation of new ways of thinking and new forms of language, while the role of poetry as uniting and gathering things together so that the truth and presence of being shines forth.

1 The attack on poetry

As Plato remarked, 'There is an ancient quarrel between philosophy and poetry.' Philosophy developed in Athens, largely through a struggle for intellectual supremacy fought between the sophists and the artists. Poetry was the prime artistic enemy since it best expressed the sacred wisdom of tradition and lacked the banausic character of plastic art.Socrates andPlato aggressively defined philosophy in sharp contrast to poetry, claiming that it provided superior wisdom and more sublime joys of contemplation. The critique of poetry that attacked its irrationality and ignorance reached its virulent peak in Book X of Plato's *Republic*, but can already be found in earlier Socratic dialogues such as *Ion* and *Apology*. Poets (and their rhapsodes or performers, of which Ion is one) do not really know how and of what they speak; they have no genuinely teachable knowledge or skill, but must rely on the divine madness of inspiration, with which they in turn infect and mislead their audiences. Poets not only lack the knowledge to explain their compositions, but they have no real specialist knowledge of the topics or arts they describe; and they mislead the public by suggesting that they are wise men (*sophoi*) and have such knowledge. Philosophy, in contrast, is said to provide not only a superior kind of truth, but, as we see in the *Symposium*, greater beauty and higher pleasures of desire. If poetry affords a view of beautiful objects, philosophy offers the rapt contemplation of the more perfect forms on which these beauties are based, culminating in the vision of the form of Beauty itself.

Plato's *Republic* paradoxically affirms poetry's importance and its dangerous lack of value. Books II and III assign poetry a central role, though closely censored and controlled, in the education of the state's guardians (censorship being essentially limited to impieties and falsehoods about the gods and to the imitation of weak behaviour and low characters). But Book X banishes all poetry from the state except for 'hymns to the gods and praises of famous men'.

Plato offers many arguments here. Since poetry is merely a representation of the appearance of the objects in our world, which is itself a world of appearances that imitates the real world of ideal forms, poetry is metaphysically inferior - a mere appearance of an appearance. For similar reasons, it is also epistemologically inferior. As an imitation of an imitation, poetry is 'thrice removed from the truth', and its teachings are therefore unreliable and misleading, far worse than the opinions of artisans, let alone the knowledge of philosophers which reaches the

forms themselves. Poetry is also psychologically suspect because, in appealing to the senses and emotions rather than to 'the calculating and rational principle in the soul', it is based on and stimulates 'an inferior part of the soul'.

This intimate link to our lower nature of senses and passions provides Plato with his most damning argument against poetry: its insidious moral corruption. Poetry perverts the soul by strengthening its lower part and undermining its rule by reason - feeding 'the passions instead of drying them up, she lets them rule'. Since it is neither easy nor amusing to imitate calm wisdom, poetry relies on a display of strong emotions that noble people would be loath to display in reality but that they assume can be safely enjoyed when fictionally represented in another (for example, indulging one's pity for another's imagined grief and fear). But, Plato argues, since reason's guard is lowered while sentiment is raised, 'the contagion must pass from others to themselves'. Here we find a theory of poetry's irrational contagion that is psychologically more subtle than that offered in *Ion*. We also find the spur to Aristotle's defence of poetry through the idea of catharsis.

2 Defence of poetry as imitation

While Aristotle's *Poetics* defines poetry as imitation, thus distinguishing it from mere writing in verse, his account of imitation is broader and more positive thanPlato's. He insists on imitation's cognitive value and rejects the Platonic metaphysics that located general forms in a transcendental world of ideas. By their very nature, humans imitate more than all other animals and learn first through imitation. The pleasures of imitation or representation are essentially cognitive. Since we are pleased by representations of things that in themselves would displease us (for example, corpses and hated animals), such pleasure must come from the recognition or inference of the thing represented. For 'the act of learning is not only most pleasant to philosophers' but to others as well, even if they achieve less of it. Here we already see a recurrent pattern in Aristotle in which the defence of poetry none the less implies and reinforces the superiority of philosophy.

Aristotle thought that poetry's imitations diverged according to the poet's natural dispositions. 'Noble actions of noble heroes' were represented in tragedy and epic, while 'Comedy is an imitation of baser men'. Epic differs from tragedy in having narrative (rather than dramatic) form, a single meter, and no limit in length of action. Aristotle devotes most attention to tragedy, arguing that it is the higher form, since it is more concentrated, effective and unified. Its analysis in terms of six parts (most important of which are plot and character) initiates the formalist approach to poetry, which we shall consider later.

Despite its different modes, poetic imitation was united, for Aristotle, in representing general, essential truths (based on 'the laws of probability or necessity') rather than mere superficial appearances or concrete particularities. 'Poetry, therefore, is more philosophical and more significant than history, for poetry is more concerned with the universal, and history more with the individual.' Such praise may answer Plato's critique of imitation, but only by implicitly affirming philosophy's superiority. In any case, the idea of poetry as imitation of general truths was influentially revived in the eighteenth century by Samuel Johnson and applied to painting by Joshua Reynolds. Aristotle offers yet another way of defending imitation: in tragedy, as in portrait painting, representations should be ameliorative, making their noble subjects look 'better than they are'. Tragedy, moreover, is specifically defended against the charge that it corrupts the soul by inspiring passions of pity and fear. To protect the soul from their harmful lurking influence, such passions need to be exorcized through poetic arousal and harmlessly poetic, cathartic discharge.

The Neoplatonist Plotinus offered a strategy that could accept Plato's transcendental metaphysics and still affirm the value of poetry as imitation (see Plotinus §3). The poet, he argued, is able to circumvent the sensual objects of material nature so as to imitate 'the Ideas from which Nature itself derives'. Employing this strategy, later thinkers could argue that poetry was one of our highest achievements and forms of truth. The idea that poetry could represent transcendent ideas was maintained by Italian Neoplatonists in the sixteenth century and played a central role in the German Romanticism of Novalis and F.W.J. Schelling, for whom 'poetry is a representation of the absolute' (1859); the English poet Shelley similarly theorized that the'poet participates in the eternal, the infinite, and the one' (1824).

Like his colleague Schelling, Hegel saw art's function as one of representing ideas (see Hegel, G.W.F. §8); he therefore claimed poetry's superiority to all the other arts, on the grounds that it was better suited to represent

ideas and ideality since it relied less on material media. ‘Poetry is…the universal art of the mind’, which ‘is not fettered…to an externally sensuous material’ (1835-8). But such praise, like Aristotle’s, still subordinated poetry to the higher value of philosophy, which is altogether free of sensuous shape.

3 Artefact theories

Perhaps Aristotle’s most successful defence of poetry was not to redeem the value of *mimesis*, but to suggest a very different way of conceiving poetry - as a formal artefact that can be appreciated in terms of its own composition and not simply by what it represents. His *Poetics* is mostly an analysis of the form of tragedy in terms of its six parts (plot, character, diction, thought, spectacle and melody), showing how they should function to constitute the work as an effective, unified whole. Here Aristotle introduces his influential notion of organic unity, which requires that the object have ‘a proper magnitude’ and ‘a proper arrangement of the component elements’. The plot (the most important part) must be so tightly unified that ‘if any one part is transposed or removed, the whole will be disordered and disturbed’.

Aristotle’s conception of poetry as a rational activity of constructing objects (a making or *poiesis* involving knowledge and ‘true reasoning’) serves to counterPlato’s attack on poetry as irrational inspiration that contaminates the soul and corrupts action. But it also serves to minimize poetry’s ethical import and render it inferior to *praxis* or action. While poetry’s making has its end outside itself and its maker (its end and value being in the object made), action has its end both in itself and in its agent, who is affected by how they act, though not allegedly by what they make (*Nicomachean Ethics* VI).

Aristotle’s view of poetry as ‘making’ later developed into the idea that the poet does not so much imitate nature as create a separate imaginative world. Adumbrated in Sidney’s *An Apology for Poetry* (1583) and in some works of the Italian renaissance, this view achieved considerable influence and elaboration in the movements of Romanticism and ‘art for art’s sake’, which, in seeing poetry as a second creation, tended to regard it as an end in itself and to deify its creator. The early Romantic Karl Philip Moritz (1788) insisted that a poem (or any authentic work of art) creates a coherent world of its own and has its entire value and purpose in itself. This idea of poetry’s intrinsic order and value (and its external purposelessness) was reinforced by Kant’s account of the purposeful purposelessness of beauty and byEdgar Allen Poe’s advocacy of the ‘poem written solely for the poem’s sake’, a notion that inspired theorists of ‘art for art’s sake’ (1850). Their position is expressed in A.C. Bradley’s claim that the nature of ‘poetry is to be not a part, nor yet a copy, of the real world…but to be a world by itself, independent, complete, autonomous’ (1909).

Though it may distance itself from Romanticism, poetic theory of the twentieth century has mostly shared the view of poetry as a well-organized formal object to be examined and valued on its own linguistic terms. In continental theory, this approach is reflected both in the formalist and structuralist schools, while in Anglo-American criticism it is seen in T.S. Eliot’s early objectivism and in the New Critical views he inspired. In aesthetics it is best represented by Monroe Beardsley’s principles of independence and autonomy, which assert that literary works are individual, self-sufficient entities whose meaning and value are independent of the works’ genesis and reception (Beardsley 1973). Poetry’s gain in autonomy as ‘object’ seems, here again, to be paid for by isolating it from the real world of practice.

4 Expression theories

Another influential theory of poetry conceives it as the expression of the poet’s mind and feelings. Such expression theories can be traced back to Longinus, who, in defining sublimity, emphasized the poet’s greatness of soul and ‘vehement and inspired passion’ (*On the Sublime*). But in modern times they begin with Giambattista Vico (see Vico, G. §4), who linked poetry with myth, religion and language itself as the primal, creative way in which human imaginative thought gives form, meaning and expression to the reality and experience it encounters (Vico 1744). The idea of poetry as the natural expression of our formative, organic and synthetic imagination is developed byColeridge and endorsed by Shelley, who defines poetry as ‘the expression of the imagination connate with the origin of man’ (1824).Croce’s theory of art as formative intuition-expression also emerges from this line of thought, while other expressionist theories instead emphasize the expression of emotions (see Croce, B. §2; Artistic expression §4). Thus Wordsworth (1800) and Mill (1833) respectively define poetry as ‘the spontaneous overflow of powerful feelings’ and ‘feeling, confessing itself, to itself, in moments of solitude’. In contrast to

Mill's suggestion of the privacy of poetic expression, Tolstoy (1896) advocated that art should not merely express but communicate emotion to the audience. I.A. Richards (1924) likewise stresses the importance of poetry's effect on the readers' emotions and attitudes, though not on their intellectual beliefs. By viewing poetry as 'emotive language', he sought to free it from the demand on referential language that it be assessed in terms of literal or scientific truth, and so to free it from condemnation as absurd falsehood.

5 Use theories

Another group of theories defines poetry by its characteristic uses. If Aristotle helped define tragedy by its cathartic effect,Horace more widely and influentially defined poetry's aims as pleasure and useful instruction, which could be optimally combined. This idea, elaborated bySidney andSamuel Johnson (who thought it the writer's duty to better the world), is endorsed in some form by many modern humanists. Differences, however, typically arise both as to the nature of this pleasure (for example, how sensual, formal or cognitive it is) and as to the kind of edification poetry offers (for example, didactic presentation of alleged truths or merely an educational sensitizing of emotions). Matthew Arnold (1881) assigned poetry the necessary function of providing a consoling, meaningful interpretation of life, after science had destroyed the consolation of religion. Many other uses have been given to poetry, though Sartre (1948) defines it in contrast to prose by its non-utility. T.S. Eliot (1933), who increasingly emphasized poetry's functions of pleasure, edification, the expansion and refinement of language and (through it) of feeling and thought, also came to realize that poetry was a social product, whose functions and very forms change over history as a result of changing social factors and needs. This raises the doubt, voiced most loudly by analytic aestheticians, such as Morris Weitz (1955), that an open genre like poetry cannot be defined in terms of any permanent essence or function. Perhaps it is then best explained not by ahistorical definition but by rich historical narrative.

6 The unity of poetry and philosophy

While most follow Plato in viewing poetry as distinct from philosophy, several important Romantic and post-Romantic thinkers have advocated their synthesis or fundamental unity. The basic strategy is to construe poetry very broadly as the basic expression of our formative, synthetic imagination and then to argue that all creative philosophical work involves such imaginative construction in presenting a unifying vision of the world. Thus Hegel (as a young Romantic, before he developed his mature philosophy) affirmed that the philosopher needs as much aesthetic power as the poet, and Coleridge insisted that one cannot be 'a great poet without being at the same time a profound philosopher' (1817). As Schlegel defined the Romantic programme, 'poetry and philosophy should be made one' in a continuing effort to shape the world creatively through integrative imaginative interpretation (Schlegel 1797-1800).

With his striking view of the world as an artistic text created by imaginative interpretation, Friedrich Nietzsche too identifies the philosopher with the poet, since both try to reinterpret and so reshape the world in accordance with their will and values. Late Heidegger (1971) links poetry with philosophy in a different way. Real philosophy or thinking is poetic not because it wilfully reinterprets the world in a novel representation, but because it respectfully and concretely unveils or illuminates the world's obscured truth of being. Poetry is 'the saying of the unconcealedness of what is'. Moreover, since truth's revelation depends on the creation of 'a clearing' or world where it can occur and be seen, poetry founds truth by gathering objects together into a world and providing a sense of that world's unity. Here, as in earlier Romantic theories, poetry is broadly conceived as a formative force common to all art and not merely poetry in the standard sense.Heidegger calls this narrower genre 'poesy' and regards it as privileged since it exists in language, the medium of truth which identifies and links beings. The strategy of seeing poetry as philosophy by seeing the latter in terms of language and insisting on the essentially poetic nature of language can be traced back through the Romantics to Vico.

The contemporary American philosophers Richard Rorty (1989) and Stanley Cavell (1988) are among those who advocate a unity of philosophy and poetry. Rorty's glorification of the philosopher as 'strong poet' is basically Nietzschean, but, to safeguard liberal democracy, he insists that the philosopher's wilful, radically innovative poetic reconstructions be confined to the private sphere of self-creation. Cavell, more in the tradition of English Romanticism, thinks poetry and philosophy can redeem each other and the world, through a heightened recognition and celebration of ordinary experience. Recognizing their Romantic roots, these contemporaries turn to poetry as a model because they think that conceiving philosophy under the scientific model has made it

increasingly technical, but of extremely limited value for what they see as its central tasks of interpreting our lives and perfecting ourselves.

See also: Lessing, G.E. §2; Structuralism in literary theory §3

RICHARD M. SHUSTERMAN

References and further reading

Abrams, M.H. (1953) *The Mirror and The Lamp: Romantic Theory and the Critical Tradition*, New York: Oxford University Press.(An excellent study of the varieties of poetic theory under the fourfold classification of theories into imitative, pragmatic, expressive and objective, similar to the one employed here. Useful bibliography.)

Adams, H. (ed.) (1992) *Critical Theory Since Plato*, New York: Harcourt Brace, 2nd edn.(A compendious anthology of aesthetic writings from most of the major figures in Western thought, prefaced by very brief but helpful introductions and bibliographies.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. and notes T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985.(Aristotle's ethical theory.)

Aristotle (*c*. mid 4th century BC) *Poetics*, trans. R. Janko, Indianapolis, IN: Hackett Publishing Company, 1987.(Aristotle's famous examination of the nature of poetry and tragedy; referred to in §3 above.)

Arnold, M. (1881) *The English Poets*, London: Ward.(Its 'Introduction' expresses Arnold's view discussed in §5.)

Beardsley, M. (1973) *The Possibility of Criticism*, Detroit, MI: Wayne State University Press, 1973.(A clear and accessible expression of the philosophy of New Criticism.)

Bradley, A.C. (1909) *Oxford Lectures on Poetry*, London: Macmillan.(Contains Bradley's famous lecture 'Poetry for Poetry's Sake'.)

Cavell, S. (1988) *In Quest of the Ordinary*, Chicago, IL: University of Chicago Press.(A case for the importance of uniting philosophy and literature for the recognition of the ordinary.)

Coleridge, S.T. (1817) *Biographia Literaria*, Princeton, NJ: Princeton University Press, 1983.(A well-edited edition of this richly suggestive but often difficult work.)

Croce, B. (1909) *Aesthetic*, trans. D. Ainslie, London: Macmillan, 1922.(The first volume ofCroce's Philosophy of Spirit; selections in Adams (1992).)

Eliot, T.S. (1933) *The Use of Poetry and the Use of Criticism*, London: Faber & Faber, 1964.(A very readable mix of English literary history and theory based on Eliot's Norton lectures at Harvard.)

Hegel, G.W.F. (1835-8) *The Philosophy of Fine Art*, trans. F.P.B. Osmaston, London: G. Bell & Sons, 1920.(A difficult four-volume work, with a very useful introduction.)

Heidegger, M. (1971) *Poetry, Language, Thought*, trans. A. Hofstadter. New York: Harper & & Row.(A collection of several later essays by Heidegger on the relations of poetry and language to philosophy.)

Johnson, S. (1759) *Rasselas*, ed. J.P. Hardy. Oxford: Oxford University Press, 1988; also inA.T. Hazen (ed.) *The Works of Samuel Johnson*, New Haven, CT: Yale University Press, 1958.(Johnson's novel parallelsVoltaire's *Candide*, musing on human malevolence and vanity. Selections in Adams (1992).)

Lacoue-Labarthe, P. and Nancy, J.-L. (1988) *The Literary Absolute*, New York: State University of New York Press.(A study of the German Romantic theories of Schlegel, Schelling and Novalis.)

Longinus (1st century AD) *On the Sublime*, trans. W.R. Roberts. *Longinus on the Sublime*, Cambridge: Cambridge University Press, 1899.(Longinus' account of the sublime, primarily in relation to literature. Reprinted in Adams (1992).)

Moritz, K.P. (1788) 'Über die bildende Nachahmung des Schoenen', in J. von Hans (ed.) *Schriften zur Ästhetik und Poetik*, Tübingen: M. Niemeyer, 1962.(Hailed by many as the first important formulation of the aesthetics of Romanticism.)

Mill, J.S. (1833) 'What is Poetry?', *Monthly Repository* January 1833; also in H. Adams (ed.) *Critical Theory Since Plato*, New York: Harcourt Brace, 2nd edn, 1992.(First part of the essay later published as 'Thoughts on Poetry and its Varieties' - an early attempt by Mill to provide a theory of poetry and art, strongly influenced by Thomas Carlyle.)

Nietzsche, F. (1873) 'On Truth and Lies in a Nonmoral Sense', in D. Brazeale (ed. and trans.) *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870s*, Atlantic Highlands, NJ: Humanities Press, 1979.(Selections in Adams (1992).)

Plato (*c*.395-387BC) *Apology*, trans. H. Tredennic. Harmondsworth: Penguin, 1954; reprinted in Plato, *The Collected Dialogues including Letters*, ed. E. Hamilton and H. Cairns, Oxford: Oxford University Press, 1975.(Plato's account of the trial ofSocrates.)

Plato (*c*.395-387 BC) *Ion*, trans T.J. Saunders, Harmondsworth: Penguin, 1987.(A short, accessible dialogue dealing with expertise and poetic performance. This edition includes a bibliography.)
Plato (*c*.386-380 BC) *The Symposium of Plato*, ed. R.G. Bury. Cambridge: Cambridge University Press, 2nd edn, 1932; trans. A. Nehamas and P. Woodruff, *Plato: Symposium*, Indianapolis, IN: Hackett, 1989.(Plato's most widely read dialogue, set in a drinking party hosted by the playwright Agathon and dealing with themes of love and beauty.)
Plato (*c*.380-367 BC) *Republic*, trans. P. Shorey. Cambridge, MA: Loeb Classical Library, Harvard University Press, 1930.(Plato's greatest work, an inquiry into justice and the ideal political order.)
Plotinus (*c*.AD 250) *Enneads*, trans. S. MacKenna, revised by B.S. Page, London: Faber & Faber, 1956.(Plotinus' applies his Platonic concept of the hierarchy of phenomena to the realm of the aesthetic, building on Diotima's speech in the *Symposium*. See especially *Enneads* 5.8.)
Poe, E.A. (1850) 'The Poetic Principle', inH. Adams (ed.) *Critical Theory Since Plato*, New York: Harcourt Brace, 2nd edn, 1992.(Edgar Allen Poe's thesis on the nature of poetry.)
Richards, I.A. (1924) *Principles of Literary Criticism*, London: Routledge.(A very influential work for the New Criticism.)
Rorty, R. (1989) *Contingency, Irony and Solidarity*, Cambridge: Cambridge University Press.(Rorty's account of the philosopher as strong poet.)
Sartre, J.-P. (1948) *What is Literature?*, trans. D. Caute. New York: The Philosophical Library, 1949.(Draws an interesting distinction between poetry and prose literature in terms of its disengagement from utility.)
Schelling, F.W.J. (1859) *The Philosophy of Art*, trans. D.W. Stott. Minneapolis, MN: University of Minnesota Press, 1989.(Presents a highly complex and systematic theory of Romantic, idealist aesthetics that was influential in German Romanticism.)
Schlegel, F. (1797-1800) 'Lyceum Fragments and Athenaeum Fragments', in E. Behler and H. Eichner (eds) *Kritische Schriften und Fragmente*, 1988.(A brief selection can be found in Adams (1992).)
Shelley, P.B. (1824) 'A Defence of Poetry', in J. Shawcross (ed.) *Literary and Philosophical Criticism*, 1909.(Classic text of English Romanticism in which Shelley makes exalted claims about the importance of poetry and the stature of poets; reprinted in Adams (1992).)
Shusterman, R.M.(1992) *Pragmatist Aesthetics*, Oxford: Blackwell.(Contains an expanded account of the philosophical strategy of compartmentalizing poetry, discussed in §§1-3, as well as a critical account of Rorty's theory of the strong poet.)
Shusterman, R.M. (1997) *Practicing Philosophy: Pragmatism and the Practical Life*, London: Routledge.(Examines the resurgence of the idea of philosophy as a poetics of life in twentieth-century philosophy, while also noting some of its pre-Romantic roots.)
Sidney, Sir P. (1583) *An Apology for Poetry*, in L. Soens (ed.) *Sir Philip Sidney's Defence of Poesy*, Lincoln, NB: University of Nebraska Press, 1970. (A stirring defence of poetry's value and of its impunity from falsehood, since it does not truly assert propositions; reprinted in Adams (1992).)
Tolstoi, L. (1896) *What is Art?*, trans. A. Maude. New York: Bobbs-Merrill, 1962.(An extremely important formulation of the expression-communication theory of art; selections in Adams (1992).)
Vico, G. (1744) *The New Science*, trans. T.G. Bergin and M.H. Fisch. Ithaca, NY: Cornell University Press, 1968.(A very rich, influential work and perhaps the first to argue that language and thought are essentially poetic.)
Weitz, M. (1955) 'The Role of Theory in Aesthetics', *Journal of Aesthetics and Art Criticism* 16: 27-55.(An influential analytic essay on the problem of defining art and its genres.)
Wordsworth, W. (1800) 'Preface to Second Edition of the Lyrical Ballads', in *The Prose Works of William Wordsworth*, Oxford: Clarendon Press, 1974.(An influential classic of English Romantic criticism that provides the essential theory of Wordsworth's poetry; reprinted in Adams (1992).)

Poincaré, Jules Henri (1854-1912)

Although primarily a mathematician, Henri Poincaré wrote and lectured extensively on astronomy, theoretical physics, philosophy of science and philosophy of mathematics at the turn of the century. In philosophy, Poincaré is famous for the conventionalist thesis that we may choose either Euclidean or non-Euclidean geometry in physics, claiming that space is neither Euclidean nor non-Euclidean and that geometry is neither true nor false. However, Poincaré's conventionalism was not global, as some have claimed. Poincaré held that only geometry and perhaps a few principles of mechanics are conventional, and argued that science does discover truth, despite a conventional element.

Poincaré followed new developments in mathematics and physics closely and was involved in discussion of the foundations of mathematics and in the development of the theory of relativity. He was an important transitional figure in both of these areas, sometimes seeming ahead of his time and sometimes seeming very traditional. Perhaps because of the breadth of his views or because of the way in which philosophers focused on issues or small pieces of his work rather than on accurate history, interpretations of Poincaré vary greatly. Frequently cited by the logical positivists as a precursor, and widely discussed in the philosophy of science and the philosophy of mathematics, Poincaré's writings have had a strong impact on English-language philosophy.

1 Life and work

Henri Poincaré was born on 29 April 1854 in Nancy, France. His father was a physician and his cousin Raymond became president of the Third Republic during the First World War. Poincaré received his education at the Lycée of Nancy, and at the École Polytechnique and the École des Mines in Paris. He worked briefly as an engineer before obtaining his doctorate in mathematics in 1879, and after teaching for a short time at the University of Caen, in 1881 became a professor at the University of Paris, where he taught mathematical physics until his death on 17 July 1912. His mathematical work was extremely wide-ranging, and he is considered to be the last person able to make major contributions in all areas of mathematics. He also made important contributions to astronomy and theoretical physics, as well as to philosophy of science and philosophy of mathematics.

Poincaré's early mathematical work was in function theory, and before the age of thirty he became famous for his discovery of automorphic functions of one complex variable ('Fuchsian' functions). Poincaré's description of his discovery while entering a bus is quoted widely in discussions of scientific discovery: ' At the moment when I put my foot on the step the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry' (1908, [1913b] 1982: 387-8).

Poincaré was awarded first prize for a contribution to a competition on the n-body problem, leading to his celebrated work in celestial mechanics (*Œuvres*, vols 7, 8). It was in this context that Poincaré proved his recurrence theorem that was influential in the development of chaos theory (see Chaos theory). He then worked in algebra and number theory, and eventually founded algebraic topology, a combinatorial theory of n-dimensional figures (*Œuvres*, vol. 6). Poincaré's lectures in mathematical physics at the University of Paris discussed a wide range of problems in turn-of-the-century physics, which he recognized to be a period of transition or even crisis (*Œuvres*, vols 9, 10). He also developed his conventionalism, claiming that there is no true metric geometry of space and that the choice of Euclidean or non-Euclidean geometry in physics is a matter of convenience (see Space §3; Geometry, philosophical issues in). There has been controversy over whether or not his contribution to the theory of relativity can be said to have anticipated Einstein's theory. Near the end of his life Poincaré became involved in discussion over the foundations of mathematics and developed an original solution to the semantic and set-theoretical paradoxes (see Paradoxes of set and property §9).

Poincaré published over 500 scientific papers and over thirty books. He was elected to the Académie des Sciences, the Légion d'Honneur and the Académie Française in France, as well as numerous foreign scientific and philosophical societies, and was awarded many prizes and honours. Interpreted as a total radical conventionalist, as an ultra-traditionalist or as inconsistent, Poincaré has been the source of continual discussion and comment by philosophers.

2 General philosophy of science

In 1890s France Poincaré had to face a popular climate decidedly hostile to science. While French philosophers turned away from Comte's positivism and Taine's materialist determinism towards the idealism of Henri Bergson, Poincaré became a popularizer of science and defended it against the attacks in vogue in the popular press, such as Tolstoi's argument that science is useless because it cannot tell us how to live, and the charge that science is 'bankrupt'. On the other hand, the long success of experimental science and the overturning of principles that were once held to be certain led Poincaré to think that rationalism was dead. The hypothetical method is the central feature of Poincaré's philosophy of science, and he considered it to be a middle ground between the dogmatic claims of the rationalist old guard and the contemporary anti-science views. A large motivation for his philosophical views seems to have been his desire to prove that science is cumulative and that the efforts of scientists in the past had not 'been sterile and vain', despite the great changes that were coming about in physics and mathematics (1905, [1913b] 1982: 280).

Poincaré distinguished three kinds of hypothesis: 'natural', 'indifferent' and 'real generalizations'. Natural hypotheses are first principles that are assumed a priori to be true. 'Indifferent hypotheses' is Poincaré's name for 'mechanical models' and refers to the nineteenth-century British tradition (of Faraday and Maxwell) of making a representation in a physical sense that provides a mechanical explanation of unobserved processes. Following a traditional form of usage (not one which Poincaré employed, however) we can refer to 'real generalizations' as fundamental laws, which should be contrasted with phenomenological laws that are taken directly from experience (see Laws, natural §1).

Poincaré is often read as an instrumentalist, one who holds that sciences should aim at giving an account of what is observable, not at explaining phenomena in terms of unobservable entities and processes (see Scientific realism and antirealism §1). Poincaré's view that the (metric) geometry of space is a matter of convention is often thought to be part of an instrumental view of all theoretical entities in science including space. However, Poincaré insisted that scientific theory is (or can be) true. In fact, Poincaré was involved in a dispute with Duhem over whether or not fundamental laws carry a truth-value. Duhem claimed that fundamental laws are neither true nor false because they are always approximate, a view that Poincaré rejected. Furthermore, while Poincaré explicitly denies the existence of atoms in *Science and Hypothesis* (1902a), arguing that the atomic hypothesis was useful but not the only possible explanation for chemical and other phenomena, he accepted their reality after he learned of the work of Perrin (Nye 1976), which shows that Poincaré had no general injunction against theoretical entities. In at least one case, Poincaré also accepted the refutation of a fundamental principle of physics, what he calls a 'natural hypothesis', by empirical theory. When he accepted the quantum theory as presented in the first Solvay conference in 1911, Poincaré gave up the hypothesis of continuity (McCormmach 1967).

3 Geometric conventionalism

Poincaré argued that we have no intuition of distance and therefore no a priori method of choosing an external standard to define distance. Furthermore, Poincaré argued that there is no spatial feature by which an external standard can be determined empirically and, therefore, alternative metrics are possible. Since statements about geometry cannot be determined a priori and cannot be determined empirically, they have a special status as conventions. Poincaré's geometric conventionalism has often been analysed in the context of later discussions of spacetime theories. The most widely promulgated argument for conventionalism is epistemological, and Poincaré's conventionalism has been thought by most interpreters to be based on an epistemological argument.

Poincaré presented two arguments that have been interpreted as defending conventionalism with epistemological arguments. In his famous parable of what measurement would be like in a non-Euclidean world, he introduces temperature as a 'distorting influence' on our measuring devices, which was taken to be an argument that one can always introduce a new hypothesis to account for data that contradicts one's theory. In other words, it has often been thought that Poincaré was using a version of the Duhem-Quine thesis (see Underdetermination §3) to defend conventionalism. However, the parable was part of an argument to prove that a non-Euclidean world is imaginable. Since a conventionalist must refute a priori arguments that space is Euclidean (such as Kant's), this argument plays an important role in the debate over conventionalism, but it may not be the whole story. Another of Poincaré's arguments that has been interpreted as epistemological concerns the definitions of primitive terms in physics. Lobatchevskii proposed using astronomical measurements to decide between Euclidean and non-Euclidean

geometry. In a passage that has been the subject of much controversy, Poincaré argued that the most such measurements can tell us is what ratios obtain between certain physical objects, in this case light rays. If one takes light rays to be straight lines, one is committed to a particular metric geometry by the results of empirical testing, but one could associate straight lines with some other physical object.

The problem with the epistemological interpretation of Poincaré's geometric conventionalism is that Duhemian underdetermination arguments establish his thesis too well. If we interpret Poincaré as using such an argument to defend his thesis of the conventionality of metric, it becomes difficult to understand how he could use such arguments in the case of metric only. If Poincaré did not deny the truth of physics in general, conventionality of metric cannot simply be a special case of a global underdetermination of theory by evidence.

An alternative defence of conventionalism bases geometric conventionalism on a fully relational theory of space. Poincaré went even further than current Einsteinian theories in rejecting Newtonian absolute space and held a purely relational view: there is no physical space and all of mechanics should be described in terms of the relations of physical objects (see Space §§2-3). These relations between physical objects are expressed in geometric terms but, according to Poincaré, any metric properties expressed in these relations are artifacts of our description. He thought that pure relationalism is empirically confirmed and that even acceleration and rotation can be treated as relative motions. Poincaré's physics is very programmatic, despite the fact that he followed developments in theoretical physics very closely and himself developed ideas for a theory of relativity that was close to that of Lorentz (Darrigol 1995). However, neither the special theory of relativity nor the general theory of relativity are relational in a sense that will support the conventionality of metric since relationalist theories of space cannot account for all spatial features without adding enough structure to determine a congruence standard (see Relativity theory, philosophical significance of).

Given that arguments for relationalism fail, one may agree with Sklar that the only way left to argue for conventionalism is the global underdetermination argument (Sklar 1974). While the epistemological interpretation of conventionalism provides a strong defence, it is not compatible with Poincaré's general views. The misrepresentation of Poincaré's philosophical views has made him look inconsistent, and inconsistency is arguably worse than holding a view that has been disconfirmed by empirical theory, which is all that is implied by interpreting Poincaré as basing his conventionalism on a relational theory of space.

4 Semantic conventionalism

The idea that conventionalism can be defended by changing the explicit definitions of terms has led to a semantic interpretation of Poincaré's views, especially his claim that Euclidean and non-Euclidean geometries are translatable. Poincaré often says that geometry is the language of physics and that we can translate a physical theory that uses Euclidean geometry into one that uses non-Euclidean geometry in much the same way as we might translate a French text on physics into German (Stump 1991: 640). Poincaré's claim that Euclidean and non-Euclidean geometries are translatable has generally been thought to be based on his introduction of a model to prove the consistency of Lobachevskian geometry and to be equivalent to a claim that Euclidean and non-Euclidean geometries are logically isomorphic axiomatic systems, but this interpretation is anachronistic because Poincaré never developed a formal conception of scientific theories. Interpreting conventionalism as part of a general semantic theory leads to the charge that it is trivial (trivial semantic conventionalism). For example, we could defend Poincaré's claim that physical bodies can be described as if they were in a strongly curved Lobachevskian space if we were to redefine 'rigidity' and postulate universal forces as distorting influences, but Poincaré specifically rejected such arguments when he distanced himself from É. Le Roy's interpretation of his conventionalism. Poincaré clearly intends his conventionalism to be more than a trivial change in notation.

Poincaré's translation thesis has a mathematical rather than a meta-mathematical basis. The mathematical basis of Poincaré's translation thesis is that the underlying manifolds of Euclidean and Lobachevskian geometries are homeomorphic (topologically equivalent). Assuming as Poincaré did that metric relations are not factual, it follows that we can rewrite a physical theory using Euclidean or Lobachevskian geometry to express the same facts. Because of his desire to prove that science is cumulative, Poincaré wrongly predicted that Euclidean geometry will always remain the preferred geometry in physics. He felt he could do so because his conventionalism would allow him to be able to rewrite in a Euclidean framework any experiment that seems to prove that the world is Lobachevskian.

5 Philosophy of mathematics

In the last years of his life, Poincaré became involved in discussions over the foundations of mathematics (1908). Concerning pure geometry, Poincaré holds the formalist-sounding views that we have no preaxiomatic understanding of geometric primitives, that rigour demands that we eliminate all appeals to intuition in geometry, and that pure (metric) geometry is neither true nor false. However, he takes quite a traditional position on arithmetic, holding that the axioms of arithmetic are synthetic a priori truths, that the notion of whole number is irreducible, and that we have a special intuitive knowledge of the fundamental principles of arithmetic - mathematical induction and the continuum. Poincaré takes an anti-Kantian view in geometry, since he accepts the consistency of non-Euclidean geometries and denies that we have any intuition of geometry, but in arithmetic Poincaré explicitly defends Kant's view that a special intuition of number is necessary against the claim of the logicists that arithmetic can be reduced to logic (see Kant, I. §5; Logicism). (Despite his explicit defence of Kant, Poincaré's idea of arithmetical intuition cannot be the same as Kant's, since he had argued earlier that we have no intuition of time, as well as no intuition of space.)

Poincaré criticized Russell's logicist programme by arguing that it is impossible to remove intuition from arithmetic. He argued that any proof of the principle of mathematical induction must rely on principles that cannot be purely logical, a critique of logicism that turned out to be correct. He also correctly predicted that it would be impossible to prove that arithmetic is consistent if one followed Hilbert's point of view in defining numbers by a system of axioms. His criticism of Hilbert and of Peano's axiomatization of arithmetic as not leading to true definitions of numbers may seem surprising, since Poincaré had praised Hilbert's work on geometry in a long review and even had adopted some of Hilbert's methods (Poincaré 1902b, 1904) (see Hilbert's programme and formalism §2).

Although Poincaré became involved in disputes over the foundations of mathematics mostly as a critic, he also developed a positive solution to the semantic and set-theoretical paradoxes in his discussion of impredicative definitions, which make use of a set of which the object to be defined is a member. He was one of the first to discuss a vicious circle principle and argued that the paradoxes can be solved if viciously circular impredicative definitions are not allowed. It is generally acknowledged that Poincaré's philosophy of mathematics is broadly speaking intuitionistic (seeIntuitionism §1). Besides his support of a Kantian view of arithmetic, Poincaré rejected the actual infinities of a realist interpretation of set theory. However, Folina (1992) has argued persuasively that Poincaré staked out a position between classical Platonism and intuitionism in the philosophy of mathematics by arguing that Poincaré's conception of the continuum is more classical than intuitionist.

See also: Conventionalism; Logic in the 19th century

DAVID J. STUMP

List of works

Poincaré, H. (1902a) *La Science et L'Hypothése*; Paris: Flammarion, 1968.(Philosophical discussions of mathematics and physics, notably including his views on the continuum and geometric conventionalism.)

Poincaré, H. (1902b) 'Les Fondements de la Géométrie', *Bulletin des Sciences mathématiques*, 2nd series, 26: 249-72; repr. *Œuvres*, 1915-56, vol. 11, 92-113.(A review of Hilbert's *Foundations of Geometry* and a development of topological ideas.)

Poincaré, H. (1904) *Rapport sur les Travaux de M. Hilbert, professeur à l'Université de Goettingen, Présentés au troisième du prix Lobatchefsky*, Kasan: Imperial University Press.(A longer version of 1902b.)

Poincaré, H. (1905) *La Valeur de la Science*; Paris: Flammarion, 1970.(Philosophical discussions of mathematics and physics, notably including views on the number of dimensions of space and a clarification of his conventionalism.)

Poincaré, H. (1908) *Science et Methode*; Paris: Flammarion, 1920.(Philosophical discussions of mathematics and physics, notably including a discussion of logicism and of the set-theoretical paradoxes.)

Poincaré, H. (1913a.) *Dernières Pensées*; Paris: Flammarion, 1963.(Posthumous collection of philosophical and popular scientific writings. Covers the infinite, ether theories, quantum theory, dimensions, and the axioms of geometry.)

Poincaré, H. (1913b) *The Foundations of Science: Science and Hypothesis, The Value of Science, Science and Method*; Washington, DC: University Press of America, 1982.(Single volume-edition of his collected

philosophical and popular scientific papers. Contains translations of 1902a, 1905 and 1908.)
Poincaré, H. (1913c) *Mathematics and Science: Last Essays*; New York: Dover, 1963.(Translation of 1913a)
Poincaré, H. (1916-56) *Œuvres*, Paris: Gauthier-Villars.(Collected mathematical and scientific writings.)

References and further reading

Darrigol, O. (1995) 'Henri Poincaré's Criticism of Fin de Siècle Electrodynamics', *Studies in History and Philosophy of Modern Physics* 26: 1-44.(See §3.)
Folina, J. (1992) *Poincaré and the Philosophy of Mathematics*, London: Macmillan.(A study of Poincaré's philosophy of mathematics.)
Friedman, M. (1983) *Foundations of Space-Time Theories: Relativistic Physics and Philosophy of Science*, Princeton, NJ: Princeton University Press.(A study of spacetime theories, including criticism of conventionalism.)
Giedymin, J. (1982) *Science and Convention: Essays on Henri Poincaré's Philosophy of Science and the Conventionalist Tradition*, Oxford: Pergamon Press.(A collection of papers on conventionalism, with an especially useful chapter on the controversy over Poincaré's contribution to the theory of relativity.)
Giedymin, J. (1991) 'Geometrical and Physical Conventionalism of Henri Poincaré in Epistemological Formulation', *Studies in History and Philosophy of Science* 22: 1-22.(A defence of the epistemological interpretation of Poincaré's conventionalism.)
Greffe, J.L., Heinzmann, G. and Lorenz, K. (eds) (1996) *Henri Poincaré: Science et philosophie, Science and Philosophy, Wissenschaft und Philosophie, International Congress Nancy, France 1994*, Paris: A. Blanchard, and Berlin: Akademie Verlag.(A large collection of current work on Poincaré, stemming from the 1994 Poincaré Congress.)
Grünbaum, A. (1968) *Geometry and Chronometry in Philosophical Perspective*, Minneapolis, MN: University of Minnesota Press.(Classic non-epistemological defence of spacetime conventionalism.)
Hadamard, J.S. (1922) The Early Scientific Work of Henri Poincaré, *Rice Institute Pamphlet* IX: 111-83.(See 1933.)
Hadamard, J.S. (1933) The Later Scientific Work of Henri Poincaré, *Rice Institute Pamphlet* XX: 1-86. (Extensive analysis of Poincaré's scientific work.)
Heinzmann, G. (1996) 'Actes du Congrès Henri Poincaré 1994', *Philosophia Scientiae* 1 (2-4).(Three special issues of the journal, containing work stemming from the 1994 Poincaré Congress.)
Holton, G. (1974) *The Thematic Origins of Scientific Thought: Kepler to Einstein*, Cambridge, MA: Harvard University Press.(Contrasts the scientific styles of Poincaré and Einstein.)
McCormmach, R. (1967) 'Henri Poincaré and the Quantum Theory', *Isis* 58: 37-55.(Study of Poincaré's acceptance of the early quantum theory.)
Miller, A.I. (1984) *Imagery in Scientific Thought: Creating 20th Century Physics*, Boston, MA: Birkhäuser.(Study of scientific creativity, with Poincaré and Einstein as the main examples.)
Mooij, J.J.A. (1966) *La Philosophie des Mathematiques de Henri Poincaré*, Paris: Gauthier-Villars.(The classic study of Poincaré's philosophy of mathematics.)
Nye, M.J. (1976) 'The Nineteenth-Century Atomic Debates and the Dilemma of an "Indifferent Hypothesis"', *Studies in the History and Philosophy of Science* 7: 245-68.(Includes a study of Poincaré's acceptance of the existence of atoms.)
Pap, A. (1946, 1968) *The A Priori in Physical Theory*, New York: Russell & Russell.(An influential interpretation of Poincaré's conventionalism as a linguistic theory of the a priori.)
Reichenbach, H. (1928) *Philosophie der Raum-Zeit-Lehre*, Berlin and Leipzig: de Gruyter; trans. M. Reichenbach, *The Philosophy of Space and Time*, New York: Dover, 1957.(Classic statement of the epistemological argument for geometric conventionalism.)
Rey, A. (1907) *La Théorie de la physique chez les physiciens contemporains*, Paris: Alcan.(A classic study of Poincaré's geometric conventionalism.)
Rougier, L.A.P. (1920) *La Philosophie géométrique de Henri Poincaré*, Paris: Alcan.(A classic study of Poincaré's geometric conventionalism.)
Sklar, L. (1974) *Space, Time, and Spacetime*, Berkeley, CA: University of California Press.(A classic presentation of spacetime theories, including criticism of conventionalism as in §3.)
Stump, D.J. (1989) 'Henri Poincaré's Philosophy of Science', *Studies in History and Philosophy of Science* 20: 335-63.(Elaboration of the material in §§2 and 3 above.)
Stump, D.J. (1991) 'Poincaré's Thesis of the Translatability of Euclidean and non-Euclidean Geometries', *Nous* 25: 639-57.(Elaboration of the material in §4 above)

Torretti, R. (1978, 1984) *Philosophy of Geometry from Riemann to Poincaré*, Dordrecht: Reidel.(Detailed presentation of the development of geometry, including an interpretation of Poincaré's conventionalism.)

Poland, philosophy in

Philosophy in Poland has developed largely along the same lines as its Western European counterpart. Yet it also has many aspects which are peculiar to itself. Historically, the founding of the University of Cracow in 1364 marks the formal beginning of Polish philosophy as an academic discipline: prior to this, philosophy was taught at numerous smaller schools, and many Poles were educated abroad, which accounts for the early influence of Western scholars and literature.

In the medieval period, philosophy in Poland followed four chronologically successive currents of thought: the via moderna, which attached itself to the nominalism of Ockham and his disciples; the via communis, which sought to find a compromise between the old ways and these new ideas; the via antiqua, which marked a return to earlier philosophical trends; and a period of early humanism. The thought of Aristotle became dominant during the fifteenth century, as was the case at practically all universities of Central and Western Europe, and although this prevailed until the eighteenth century, philosophy did not remain stagnant - variations were numerous (including Protestant Aristotelianism). The prominence of political thought in the sixteenth century reflects the fact that Poland developed a new constitutional order at this time, the 'democracy of nobles' (the nobility accounted for about ten per cent of the total population). Nicholas Copernicus, prominent in modern astronomy and natural science, played a fundamental role in the development of philosophy during this period.

The eighteenth-century Polish Enlightenment was shaped mainly by the clergy and hence was initially Christian in outlook. A more radical Enlightenment programme was propagated at a later stage. The following century saw the loss of Polish independence, and Polish thinkers were more prominent in exile than in their own country. At home, this coincided with a period of Romanticism and mystical philosophy ('Messianism'), with influences of Kant and particularly of Hegel. The end of the nineteenth century saw a variety of old and new philosophical orientations, ranging from medieval thought to positivism and Marxism, while 1895 saw the beginning of the Lwów School of philosophy which was to become prominent in the twentieth century.

After Polish independence in 1918, logic and methodology flourished under the influence of the Lwów School. However, elements of a variety of other Western schools of thought were also present, including that of British analytical philosophy. After the Second World War, administrative strictures were imposed in order to give prominence to Marxism. A certain liberalization took place after 1956, but its effects were dampened by a highly intrusive censorship. Despite this, philosophy in Poland continued to build upon the pre-Communist trends of Thomism and phenomenology, and to incorporate the new modes of thought emerging in the West. Since 1989-90, Marxism has lost its politico-administrative supports and censorship has disappeared, so that contemporary philosophy in Poland is entering a new phase of development.

1 The medieval period

Between the tenth and fifteenth centuries, philosophy in Poland developed in a manner akin to its Western European counterpart. It shared the same religious and metaphysical base, the same scholarly resources and the same school structure. Close contact was maintained with Western scholars, and so the influences on the development of philosophy were similar, facilitated by the widespread use of Latin. Nevertheless, Polish philosophy at this time also had its own defined peculiarities, especially in the fields of Practicism (and an accompanying interest in socio-political problems and in 'devotio moderna' - Mateusz z Krakowa (d. 1410), Jakub z Paradyża (d. *c*.1464) and others), an almost-universally accepted Concordism, an exceptional flourishing of Buridanism and a methodological emphasis on the separation of philosophy from theology (see Buridan, J.). Conciliarism was also prevalent. Several other disciplines flourished in Poland at this time, including law (which developed a *jus gentium* long before Grotius), science (in which Copernicus was educated) and historico-philological thought which gave birth to Polish humanism.

Poland's conversion to Christianity in 966 marked the beginning of the culture which bred this philosophy and learning. Schools possessing libraries with philosophical works were founded during the tenth to twelfth centuries, in conjunction with cathedrals and religious houses, and scholarly literature was imported from abroad. The first Polish chronicles were written by Gall. Anonim in the eleventh to twelfth centuries, and by the beginning of the fourteenth century, schools in Poland numbered in excess of 100. The liberal arts were widely taught and many

Poles were sent abroad to be educated, studying in Paris, Bologna, Montpelier and Padua.

Scholars of international stature emerged during the twelfth and thirteenth centuries, including Wincenty Kadłubek (d. 1223), the chronicler, lawyer and philosopher; Marcin Polak (d. 1278-9), author of the famous papal and imperial chronicles; Peregryn z Opola (d. after 1333), whose sermons were disseminated throughout Europe; and Witelo (d. *c*1290), whose philosophy, while concerned with anthropology, optics, demonology and the metaphysics of light, remained under Platonic and Neoplatonic influences while drawing from Chalcidius, Avicenna (see Ibn Sina), Alhazen (see Ibn Hazm) and Algazel (see al-Ghazali), among others. In the first half of the fourteenth century, many other philosophers and theologians with Polish connections received widespread recognition in ethics, anthropology, physics, metaphysics and logic. All came under the decisive influence of Aristotle and all assigned a subordinate role to philosophy in relation to theology.

The foundation of the University of Cracow by King Casimir the Great in 1364 marked the watershed in the history of medieval philosophy in Poland. Following its reform under the royal Jagiełło family (Jadwiga and Władysław) in 1400, the University became an important European centre of learning, and by the beginning of the fifteenth century, the so-called *via moderna* (associated with the nominalism of William of Ockham and his followers) developed dynamically there (see Nominalism. By the second half of that century, the search for compromise between *the via antiqua* and the *via moderna* gave birth to the so-called *via communis* (the 'common path'), following the rules of Concordism. Precursors of that path include Pihotr Wysz (d. 1414), a distinguished philosopher whose commentaries on the works of Aristotle have recently been rediscovered.

Among those who favoured the *via communis* position were the Cracow lawyers Stanisław z Skarbimierza (d. 1431) and Paweł z Włodkowic (d. after 1434). The latter's doctrines on just war theory and on political and religious freedom were a novelty in Europe and influenced subsequent Polish toleration during the period of the religious wars in the sixteenth and seventeenth centuries. Other advocates of the *via communis* include the two most important Polish philosophers of the first half of the fifteenth century, Pawełz Worczyna (d. *c*.1430) and Benedykt Hesse (d. 1456). Paweł's work in ethics is characterized by Practicism, while Hesse's commentaries on the works of Aristotle maintain a spirit of Buridanism modified by the influence of Londorius. The new physics developed by Hesse departs from Aristotle and introduces new concepts into natural science, such as the theory of impulse. In his great commentary on the Gospel of Matthew, Hesse addresses problems of ethics, asceticism, law and economics, resolving them in a spirit of pragmatism and practicality.

The second half of the fifteenth century also saw a brief revival of Neoplatonism, Thomism, Scotism, Averroism and Augustinianism. By the end of the century, however, explicit humanistic tendencies began to appear; among its chief advocates was the most famous of Polish scholars, Nicholas Copernicus.

2 The modern period

Owing to lively contacts with Italian centres of learning, the beginnings of Renaissance humanism were evident in Poland as early as the first half of the fifteenth century (see Humanism, Renaissance). Early humanism was characterized primarily by its literary character. The writings of Aristotle continued to be the basic texts in philosophical education in Poland. From the beginning of the sixteenth century, however, a renewed interest in ancient texts became widespread, and teaching and commentaries on Aristotle began to change noticeably. At the University of Cracow, Renaissance Aristotelianism prevailed under the influence of Jan Schilling (d. 1518) and Grzegorz za Stawiszyna (d. 1540) who both sought to explain Aristotle using Aristotle himself rather than scholastic commentary.

Humanism brought a philosophical method different from that of scholasticism, based on a broad range of ancient texts. Philosophy was now primarily regarded as an instrument for perfecting individuals as citizens and thus for perfecting society. The works of Erasmus of Rotterdam played a significant role in this process. However, the eclecticism of Renaissance philosophy became prevalent at the University of Cracow from the end of the second half of the sixteenth century (this model also dominated in the first fifty years of the influential Zamość Academy from its foundation in 1595). The curriculum began to incorporate a broader spectrum of ancient writings, including Plato, the Stoics, the Neoplatonists and St Augustine. Cicero was especially popular.

While particular emphasis was placed on practical disciplines, the most interesting results were achieved in logic. In his *Commentariorum Artis Dialecticae* (1563), for example, Jakub Górski (d. 1585) examines Stoic dialectics in

conjunction with Aristotle's logic, closely connecting logic and rhetoric. He discusses proofs with the help of probability arguments, limiting the discussion, however, to the most well-established proofs. Adam Burski (d. 1611) offers a rich anthology of ancient texts from the history of logic in his *Dialectica Ciceronis* (1604). It expounds on ancient dialectics, comparing and contrasting the logic of the Stoics with that of Aristotle.

The spirit of the Reformation, both positive and negative, prevailed in the extensive social and political writings which started in the 1540s. These were characterized by the fusing of religious and the political ideas. Republican and democratic currents of thought were decisively dominant, and the ideology of an absolute monarchy had few or no representatives in sixteenth-century Poland. Of the numerous political writers at this time, two expressed their views in the context of broader philosophical theories. Andrzej Frycz Modrzewski (d. *circa* 1572) was the author of the foundational work *De Republica Emendanda* (1551-4) which contained proposals for the reform of social customs, the law, the church, education and the army. Although a proponent of strong monarchy, he subordinated royal power to the rule of law, advocating that the king should be chosen by representatives of all classes. He proposed the creation of a criminal code which was to be binding on all and stressed the values of Christianity, while striving for union between Protestantism and Roman Catholicism. The resulting united Polish Church he envisaged as dependent on the king and on an ecumenical council which would exercise supreme power. Modrzewski's works were translated into several languages. In contrast to this his contemporary, Stanisław Orzechowski (d. 1566), identified Polish national traditions with Catholicism and stressed the importance of the clergy and the superiority of their way of life, thus inaugurating the Counter-Reformation in Polish political writing.

Poland was not unaffected by the rebirth of Thomism (and, to a lesser extent, Scotist philosophy) throughout Europe in the second half of the sixteenth century which was closely connected to the post-Tridentine reform movement in the Roman Catholic Church. At the turn of the seventeenth century, the Dominican and the Jesuit models of Thomistic philosophy as a commentary on Aristotle had become prevalent all over Europe. The former model, cultivated at the University of Cracow and the Zamość Academy, stressed the necessity of remaining faithful to the teachings of Thomas Aquinas, while the latter followed Suárez in emphasizing the need to update them, and was endorsed at the Jesuit-run Vilnius Academy. The Franciscans, meanwhile, philosophized in the spirit of John Duns Scotus.

A new dimension was added with the advent of Protestant Aristotelianism - Protestants did not have their own university in Poland, although they operated five schools where the level and range of philosophy taught was not fundamentally different from their university counterparts. (The Czech philosopher Comenius taught at one such school in Leszno from 1628 to 1656.) Political philosophy in the seventeenth century generally followed the republican-democratic trends of the previous century. However, absolute monarchy was advocated in some quarters.

The first half of the eighteenth century saw a continuation of the Scholastic Aristotelianism of the Dominicans and opposition to modern theories by figures such as Jerzy Gengel (d. 1730) and Adam Malczewski (d. 1754). By mid-century, however, characteristic perspectives of Enlightenment philosophy (the so-called *philosophia reccentiorum*) had penetrated Poland, primarily from Germany and Italy. This connected certain threads of Aristotelian philosophy with selections from Descartes, Gassendi, Locke, Newton and Leibniz. Epistemological and methodological concerns were primarily taken up in logic, limiting and sometimes even eliminating formal logic. Questions about world view (*Weltanschauung*) were prominent in metaphysics, while the teaching of natural law developed in ethics (see Natural law). The duty of taking part in social and political life was emphasized, as well as the essential role of Christianity in forming an authentic moral culture. Aristotelian philosophy of nature was limited or even abandoned in favour of the modern natural sciences. As the evolution and development of the Polish Enlightenment was carried out by the Catholic clergy, especially the religious orders and to a lesser extent the Protestant clergy, the typical anti-religious tendencies did not feature initially. By the end of the century, however, French and English influences had seeped through to inspire a more radical Enlightenment programme of rationalism, empiricism and naturalism (see Enlightenment, Continental).

3 The nineteenth and twentieth centuries

Until 1830, influences of the French Enlightenment, German Kantianism and the Scottish philosophy of 'common sense' continued to be the main traits identifiable in Polish philosophy (see Neo-Kantianism; Common Sense

School). Precursors of the Messianic philosophy which would dominate between the uprisings of November 1830 and January 1863 also appeared, marked by certain supra-national traits connected with Romanticism and post-Kantian German Idealism (see German idealism; Romanticism, German). These also possessed certain national characteristics that were part of Poland's religious and cultural traditions, and especially of its political and social situation as a nation deprived of statehood and sovereignty. General features of this Messianism include spiritualism, a highly personalized notion of the relationship between God and humans, and an orientation towards action and change - especially moral transformation of the person and the nation.

Most prominent at the time were Józef Maria Hoene-Wroński (d. 1854), who propounded a decidedly rationalistic philosophy, and August Cieszkowski (d. 1894) who represented the non-rationalistic side. Both systems, which had Neoplatonist orientations, were richer than Hegelianism by virtue of their theories of action and of the future. This was also the period of the so-called 'Catholic Philosophy' which, in Poland, often came closer to fideism and either was subject to Hegelianism or opposed to it. (Outside Poland, this philosophy came to be connected to neo-scholasticism.) Jewish philosophy developed in Poland during this period, influenced by Enlightenment rationalism and German Idealism. Its most prominent representatives included Mendel Lewin Satanower and Nachman Krochmal.

Positivism (as well as variations on Neo-Kantianism which came close to positivism) became widespread after 1863. This occurred not just as a result of the influence of Comte, J.S. Mill, Spencer and Bain, as well as scientism and Darwin's theory of evolution (see Darwin, C.): a domestic pre-positivistic tradition with roots in both the eighteenth century and Poland's politico-social conditions also entered the picture. Representatives of positivism during this period include Michał Wiszniewski (d. 1865) who also sympathized with the Scottish school (though he was not without Kantian influences). An English translation of his *Charaktery rozumów ludzkich (Sketches and Characters or the Natural History of Human Intellects)* was published in London in 1853. Within a broader understanding of positivism, Polish contributions are evident in the fields of scholarly research on mediumism (Investigations of problems concerning hypnosis and parapsychology - Julian Ochorowicz, d. 1935) and the methodology of medicine (Władysław Biegański, d. 1919). Representatives of other philosophical orientations of the late nineteenth and early twentieth centuries include Henryk Struve (d. 1912), an eclectic post-Hegelian scholar, and Wincenty Lutosławski (d. 1954), a proponent of the 'national philosophy' who was influential not so much for his spiritualistic metaphysics as his *Origin and Growth of Plato's Logic* (1897), a work on the chronology of Plato's corpus.

A new era of development began in Polish philosophy at the end of the eighteenth century, largely due to the 1895 appointment of Kazimierz Twardowski to the chair of philosophy at the University of Lwów, and to the establishment of the first specialized philosophical journal in Poland, *Przegląd Filozoficzny*, in Warsaw in 1897-8. The Lwów School (subsequently called the Lwów-Warsaw School) founded by Twardowski emphasized the use of detailed analysis, clear and precise philosophizing and inspired research in logic and methodology. (Notwithstanding certain common figures, one should distinguish the Lwów-Warsaw School of Philosophy from the Warsaw School of logic - which arose after 1918 - and the Warsaw School of mathematics.)

Many of Twardowski's students became prominent in their own right in the fields of philosophy, logic and psychology. Leopold Blaustein (d. 1944) was the author of many original works on descriptive psychology and was a pioneer in psychology pertaining to film and radio. Kazimierz Ajdukiewicz (d. 1963) produced important results in theory of language and logical analysis of the theoretical problems of cognition. His works, especially those published in *Erkenntnis* and in the Polish *Studia Philosophica* (along with Tarski's famous 1933 paper 'Pojęcie prawdy w językach nauk dedukcyjnych' - translated as 'The Concept of Truth in the Languages of the Deductive Sciences') influenced the development of the Vienna Circle and gave birth to logical semantics (see Tarski, A.; Vienna Circle). Czeżowski (d. 1981), another student of Twardowski, turned from formal logic to general theory of science and treated logic as learning in the general form of a science. Admitting various forms of immediate knowledge, he was a probabilist in terms of evaluating knowledge in relation to reality. Zawirski (d. 1948) also became prominent: he produced the book *L'évolution de la notion du temps* as well as various works in the philosophy of physics, multivalent and intuitionist logic and was a founder of causal logic (see Polish logic).

Other philosophical orientations were evident in interwar Poland. Empirical-critical thought was promoted by thinkers such as Władysław Heinrich (d. 1957), Narcyz Łubnicki (d. 1988) and Bolesław J. Gawecki (d. 1984).

Joachim Metallman (d. *c.*1942) was primarily interested in epistemology and methodology of the natural sciences. Leon Chwistek (d. 1944) pursued formal logic, searching for the logical system most compatible with nominalism which could simultaneously serve as a basis for metaphysics. He was also an artist and art theoretician. Władysław Tatarkiewicz (d. 1980) was distinguished in the history of philosophy, the history of aesthetics and art history. The style of philosophizing in his works on axiology is evocative of G.E. Moore and similar authors in analytic philosophy (see Axiology).

Benedykt Bornstein occupies a unique position in this era through his metaphysical speculation, connected as it was with logic and mathematics (known as geometric logic or 'topologic'). He linked pan-rationalism with accentuation of the foundational role of intellectual intuition. His last publications include *Geometric Logic* (1939), *Teoria absolutu. Metafizyka jako nauka ścisła (The Theory of the Absolute. Metaphysics as Science)* (1948). Roman Ingarden (d. 1970) promoted an objectivist version of phenomenology, thereby providing an alternative to the Lwów-Warsaw School and to minimalist analytical-positivist thought.

Notwithstanding the presence of several active representatives, Thomism in Poland before the Second World War possessed an eclectic and derivative character compared to Thomism abroad. The dominant influence came from the Louvain School, from whence came the founder and first rector of the Catholic University of Lublin (KUL), Idzi Radziszewski. Two important initiatives deserve note. First, the international project 'Corpus Philosophorum Medii Aevii' was founded at this time by Konstanty Michalski (d. 1947), a distinguished historian of philosophy and pioneer of the Thomistic philosophy of history. Second, on the occasion of the third Polish Philosophical Congress in Cracow in 1936, the project of the renewal of Thomistic philosophy and theology was formulated by what came to be known as the 'Cracow Circle'. This group of four opted to employ contemporary tools of logic. Their initiative was a novelty, even beyond Polish Thomism, and their aims were in part accepted at KUL after 1945.

Despite the loss of many promising academics and students during the Second World War, philosophy in Poland has been significantly active since that time. Having strengthened its political position, the ruling communist party sought to subordinate the whole culture to its Marxist ideology. However, despite many pressures and restrictions, the Communists never quite succeeded in eliminating non-Marxist orientations in Polish intellectual life. The traditions of the Lwów-Warsaw School continued in logic and in the methodology of the sciences. The objectivist phenomenology of Ingarden continued, at least in aesthetics and art theory, while at KUL, the Lublin School of existential Thomism developed. Polish thought passed through periods of Stalinism, Marxism, the Solidarity phase of 1980-1 and into the final breakthrough of 1989-90, which ended the domination of the Communist party.

4 Contemporary Polish philosophy

At the present time, philosophy in Poland represents a wide range of schools and perspectives. Phenomenology, Thomism, various schools of positivism and post-positivism, Marxism and the philosophy of dialogue (encounter) are all represented. The influences of Augustinianism, Hegelianism, hermeneutics and process philosophy can also be noted (see Hermeneutics; Process philosophy).

Phenomenology owes its position in Poland especially to Roman Ingarden, who continued to teach and whose thought flourished in Cracow after the Second World War. Notables among his students include Władysław Stróżewski, who connected phenomenology first with existential Thomism and later with certain elements of Platonism and Hegelianism in a 'dialectical phenomenology', and Józef Tischner who, in the end, departed far from the thought of his mentor. In phenomenological circles one also finds Karol Wojtyła (later Pope John Paul II), whose personalism combines the philosophy of being in the existentialist Thomistic spirit with the philosophy of subject in the spirit of classical phenomenology.

The primary centre of Thomistic philosophy is at the Faculty of Philosophy at KUL, and this philosophical orientation also prevails in the Faculty of Christian Philosophy at the Academy of Catholic Theology in Warsaw (Mieczysław A. Krąpiec, Stefan Swieżawski, Mieczysław Gogacz). Existential Thomism, borrowing from Gilson and Maritain while emphasizing the importance of the theory of being, supplies the intellectual framework, with the more traditional school of Thomism declining in prominence in recent years. At the Pontifical Academy of Theology (PAT) in Cracow, the philosophy of dialogue has been promulgated, while Andrzej Nowicki formulated a Marxist version of the philosophy of encounter ('inkontrologia'). PAT, under the banner of philosophy in

science, fosters a programme of interdisciplinary research (based upon identification and resolution of philosophical problems in the context of particular sciences) employing ideas from Platonism and from process philosophy. Since the death of Izydora Dąmbska in 1983, the Lwów-Warsaw School has been less prominent than in former years.

Marxism in Poland, although primitive, doctrinaire and Stalinist after the Second World War, later differentiated itself into various orientations. These have included humanist (invoking the young Marx), scientific (appealing to Engels), moderate, Leninist/revisionist, open Marxism (invoking modern scientific methodology and certain post-positivistic trends), and those who promoted a programme of eclectic 'universalism'. 'Individualistic' philosophers include Andrzej Grzegorczyk (a logician whose philosophy connects elements of scientism, existentialism and a Christian world view) and Jerzy Szymura (who defended a version of Hegelianism close to that of Bradley). Prominent Poles living abroad include Józef M. Bocheński (d. 1995) (following a Thomistic period, he later identified himself as an analytical philosopher who cherished the Aristotelian tradition), and Leszek Kołakowski who has changed from Marxist to critic of Marxism and whose current views can be characterized as a post-Hegelian, post-Marxist historicism and sociologism with a positivistic version of rationalism.

See also: Aristotelianism in the 17th century; Enlightenment, Continental; Enlightenment, Jewish; Humanism, Renaissance; Krochmal, N.; Marxist philosophy, Russian and Soviet; Thomism

JAN CZERKAWSKI
ANTONI B. STĘPIEŃ
STANISŁAW WIELGUS

References and further reading

Birkenmajer, A. (1972) 'Histoire de sciences et de l'enseignement en Pologne', *Studia Copernicana* 4 (X): 819. (A history of the sciences and of general education in Poland.)

Bornstein, B. (1939) *Geometric Logic*, Warsaw: Wolna Wszechina Polska. (Mentioned in §3 above. An interesting parallel between the structure of space and the structure of thought.)

Bornstein, B. (1948) *Teoria absolutu. Metafizyka jako nauka ścisła (The theory of the absolute. Metaphysics as science)*, Łódź: Łódzkie Towarzystwo Naukowe.(An attempt to treat geometric structure of space as a representation of the structure of being; mentioned in §3 above.)

Chudy, W. (1990) 'Filozofia polska po II wojnie światowej' (Polish philosophy after the Second World War), *Studia Philosophiae Christianae* 26 (1): 129-41. (Survey of Polish philosophy in the period 1945-90.)

Czerkawski, J. (1992) *Humanizm i scholastyka (Humanism and scholasticism)*, Lublin: RW KUL.(Studies in the history of philosophical culture in Poland in the sixteenth and seventeenth centuries.)

Domański, J., Szczucki, L., Ogonowski, Z. (eds) (1978-9) *700 lat myśli polskiej (Seven-hundred years of Polish thought)*, 5 vols, Warsaw: PWN.(Includes a selection of classical texts and essays by Polish scholars from the thirteenth to the nineteenth centuries, excepting the eighteenth.)

Domański, J., Szczucki, L., Ogonowski, Z. (1989) *Zarys dziejów filozofii w Polsce; wieki XIII-XVII (A short history of philosophy in Poland, 13th-17th centuries)*, Warsaw: PWN.(This work contains an extensive bibliography of classical texts and essays by contemporary scholars.)

Gogacz, M. (1978) 'Tomizm w polskich środowiskach uniwersyteckich XX wieku' (Thomism in Polish university circles in the 20th century), in *Studia z dziejów myśli świetego Tomasza z Akwinu*, Lublin: TN KUL, 335-50. (Survey of twentieth-century Thomist philosophy in Polish academia.)

Hinz, H., Sikora, A. (eds) (1964) *Polska myśl filozoficzna. Oświecenie. Romantyzm (Polish philosophical thought: The Enlightenment and Romanticism)*, Warsaw: PWN.(An anthology of writings by eighteenth- and nineteenth-century Polish philosophers,with a bibliography of their works.)

Jordan, Z.J. (1963) *Philosophy and Ideology*, Dodrecht: Martinus Nijhoff.(Traces the development of philosophy and marxism in Poland since the Second World War.)

Kadler, A. (ed.) (1955, 1960, 1971) *Bibliografia filozofii polskiej (Bibliography of Polish philosophy)*, 3 vols, Warsaw: PWN.(Vol. 1 (1750-1830); Vol. 2 (1831-64); Vol. 3 (1865-95). More volumes in preparation. Contains valuable bibliographic material for the student interested in more advanced study.)

Kuksewicz, Z. (ed.) (1975-88) *Dzieje filozofii średniowiecznej w Polsce (History of medieval philosophy in Poland)*, 10 vols, Wrocław: Ossolineum.(Six authors describe medieval Polish logic, methodology of science, philosophy of being, philosophy of nature, philosophy of man - theory of soul, epistemology, moral philosophy, beginnings of humanism in Poland, beginnings of social sciences - and methodological principles.)

Lutosławski, W. (1897) *Origin and Growth of Plato's Logic*, London: Longmans, Green.(Famous work on the chronology of Plato's corpus. Mentioned in §3 above.)

Modrzewski, A.F. (1551) *De Republica Emendanda*, Cracow.(Contained proposals for extensive social reform. Mentioned in §2 above.)

Ogonowski, Z. (1992) *Filozofia polityczna w Polsce w XVII wieku i tradycje demokracji europejskiej (Political philosophy in Poland in the 17th century and the traditions of European democracy)*, Warsaw: PAN.(An anthology of texts by seventeenth-century Polish philosophers.)

Skolimowski, H. (1967) *Polish Analytical Philosophy*, London and New York: Routledge & Kegan Paul.(Review of the main features.)

Stępień, A.B. (1966) 'Fenomenologia w Polsce (w powojennym dwudziestoleciu)', *Studia Philosophiae Christianae* 2 (1): 29-47.(Phenomenology in Poland in the twenty years after the Second World War.)

Stępień, A.B. (1968) 'La filosofia tomastica nella Polonia contemporanea', *Sapienza* 1968: 509-28.(Survey of Thomistic philosophy in Poland, written during the 1960s.)

Stępień, A.B. (1975) 'La filosofia nella Polonia contemporanea', *Rivista di Filosofia Neo-scolastic*a 67: 288-99. (Review of philosophy in contemporary Poland.)

Szaniawski, K. (ed.) (1988) *The Vienna Circle and the Lwów-Warsaw School*, Dodrecht: Kluwer Academic Publishers.(Vol. 28 in the Nijhoff International Philosophy series. Not for the beginner.)

Szczucki, L. (ed.) (1978) Nauczanie filozofii w Polsce w XV-XVIII wieku (Teaching philosophy in Poland in the 15th -18th centuries), Wrocław: Ossolineum.(A selection of writings about the teaching of philosophy in Cracow, Zamość, Gdansk and Torun.)

Wielgus, S. (1990) *Obca literatura biblijna w Średniowiecznej Polsce (Foreign biblical literature in medieval Poland)*, Lublin: Redakcja Wydawnictw KUL.(A collection of foreign, Western biblical writings which were present in medieval Polish schools and in the University of Cracow in the fifteenth century.)

Wielgus, S. (1991) 'Die Theorie des Menschen in den Werken Krakauer Theologen aus der zweiten Hälfte des XV Jahrhunderts' in B. Mojsisch and O. Pluta (eds) *Historia Philosophiae Medii Aevi*, vol. 1, Amsterdam and Philadelphia, PA: B.R. Grüner, 1047-64.(An exposition of the philosophical theory of man, based on the main Polish biblical commentaries written in the second half of the fifteenth century.)

Wielgus, S. (1992) *Średniowieczna łacińskojęzyczna biblistyka polska (Polish medieval biblical science in Latin)*, Lublin: Redakcja Wydawnictw KUL.(Reviews the scientific production of medieval Polish scientists who were taking up theology and philosophy.)

Wielgus, S. (1995) 'Średniowieczna filozofia w Polsce(Polish medieval philosophy)', in *Z badań nad średniowieczem*, Lublin: Redakcja Wydawnictw KUL. (A brief history of medieval Polish philosophy with a very extensive bibliography of Polish medieval texts and essays by contemporary scholars.)

Wiszniewski, M. (1837) *Charaktery rozumów ludzkich* (*Sketches and characters or: The natural history of human intellects*), Cracow, 1842.(Mentioned in §3 above. The translation was published by Saunders & Otley in London (1853) under the pseudonym James Whitecross.)

Woleński, J. (1989) *Logic and Philosophy in the Lwów-Warsaw School*, Dodrecht: Kluwer Academic Publishers. (This is the English version of the first extensive monograph on the Lwów-Warsaw School.)

Zawirski (1936) *L'évolution de la notion du temps*, Cracow: PAU.(Mentioned in §3 above. An outline of the history of various conceptions of time, with analytical and critical commentaries.)

Polanyi, Michael (1891-1976)

Michael Polanyi was almost unique among philosophers in not only fully acknowledging but in arguing from the tacit dimensions of our knowledge which concern the many things which we know but cannot state nor even identify. He argued that our knowledge is a tacit, personal integration of subsidiary clues into a focal whole, and he elaborated this structure of knowing into a corresponding ontology and cosmology of a world of comprehensive entities and actions which are integrations of lower levels into higher ones. Polanyi used these accounts of knowing and being to argue against the 'critical' demands for impersonal, wholly objective and fully explicit knowledge, against reductionist attempts to explain higher levels in terms of lower ones, and to defend the freedom of scientific research and a free society generally.

1 Life

Michael Polanyi was born in Budapest into a distinguished Jewish family. He trained as a doctor but was primarily interested in research in chemistry. He was also active in support of liberal movements in Hungary but was sceptical of socialism. Opposed to the Bolshevik government in 1919 and distrusted by the Horthy regime which overthrew it, he left Budapest in 1919 and pursued a career in chemistry in Germany. In 1933 he emigrated to England and became Professor of Physical Chemistry at the University of Manchester. He made several visits to the Soviet Union, where a meeting in 1935 with Bukharin alerted him to Marxist proposals for the central planning of scientific research. He was already writing on economics and trying to articulate his ideas of a free society.

In 1948 he gave up his distinguished career in chemistry and was appointed to a non-teaching chair in Social Science in order to prepare his Gifford Lectures for 1951-2, which later became *Personal Knowledge* (1958), his most important book. In 1959 he became a Senior Research Fellow at Merton College, Oxford. Linguistic analysis, which was prevalent there, gave little support to his style of philosophy, but he attracted more attention in the USA, where he gave several series of lectures, published as *The Tacit Dimension* (1966) and *Meaning* (1975).

His philosophical work, in content and style, was never merely academic. Replete with examples, especially from natural science, it always addressed itself beyond purely philosophical questions to the social and political problems of the twentieth century. Every one of his books begins with some aspect of our present discontents, works back to their presuppositions, articulates a constructive alternative, and then indicates the new directions which thought, science and society should take.

2 Tacit integration

Polanyi is best known for his statement from *The Tacit Dimension*, 'We know more than we can tell' (1966: 4). From *Personal Knowledge* onwards, he gives many examples of such tacit knowledge. But his unique contribution to philosophy is his account of knowing as a tacit integration of subsidiary clues into a focal whole. Whereas phenomenology stresses that consciousness is intentional in that it is always consciousness *of* something, Polanyi shows that mind has a double intentionality: A attends from B to C (see Intentionality; Phenomenological movement). We attend *from* the subsidiary clues and to the focal whole. For example, a blind man using a stick does not pay attention to the impact of his stick upon his hand, but to the tip of the stick and what that tells him about what it touches. The sensations in the palm of his hand are subsidiary clues which he integrates into a focal awareness of the path and obstacles in front of him. If we switch our attention to the subsidiary clues, then sooner or latter we shall lose sight of the focal object to which we were attending. We attend from words to their meanings, and by repeatedly pronouncing a word we forget what it means. We can regain its meaning only by using it. This relation of subsidiary details and clues to the focal whole is a functional one. We use the clues or details to apprehend or perform the focal whole. Subsidiary awareness is therefore more than merely subconscious or fringe awareness. Furthermore, what we attend to can itself become a subsidiary clue used in attending to something else, as when learning to play the piano we first attend to our fingers, and then from them to the music to be played.

Polanyi shows that these integrations pervade all our knowledge, and not just perception. They control our use of language, which can never be rendered exact and precise. Because we always attend from some set of subsidiary clues, we can never make our knowledge wholly focal. Therefore we can never make it wholly explicit and

articulate, nor subject it to a complete critical scrutiny. It must always rest upon the tacit and therefore acritical foundations of personal judgment and commitment. The objectivist ideal of a fully tested, detached, impersonal and 'objective' knowledge cannot be met. But knowledge is 'personal' and not 'subjective'. It is attention away from ourselves to a world that anchors our commitments. Reality outruns our conceptions of it and shows itself in ways that we cannot anticipate (see Tacit knowledge).

We use as subsidiary clues details of the object known, its context and ourselves. The object known is itself a comprehensive entity, or complex performance, in which details on a lower level are integrated by the principles of operation of the next higher level which determine the boundary conditions left open by the operational principles of the lower level. In attending from its subsidiary details, we re-perform, as it were, its own internal integration. Our own bodies are primarily what we use and so attend *from*, in knowing and acting within the world. We indwell them. Likewise, by extension, we indwell the objects that we know. As the blind man incorporates his stick into his body by indwelling it, so we incorporate our acquired knowledge, ideas and frameworks into our minds by indwelling them. Polanyi bridges the Cartesian gap between the self and an 'external world' with the indwelling of tacit integration (see Dualism).

Polanyi described his philosophy as a 'post-critical' and 'fiduciary' one, which aims to enable us explicitly to hold beliefs which we know might be false. As against the 'critical' attempt to apply the method of doubt and to reject whatever cannot be proved, it seeks to articulate the ultimate beliefs which are presupposed by our proximate ones and practices. The premises of science, for example, are embodied and tacitly known in the practice of scientific research. From results which we already accept as true and methods which we already hold to be valid we can try to articulate the principles and beliefs which they presuppose. We cannot start with the latter and use them to justify the former (see Merleau-Ponty, M.).

3 Applications

Polanyi applied his account of tacit integration in knowing and being to many topics, such as the relation between language and meaning, the structure of a work of art, the refutation of various reductionisms, the relation of body and mind, the nature of discovery, and the role of tradition in conveying tacit knowledge. His distinctive philosophy of science emphasizes the roles of the personal factors of belief, passionate engagement, imagination and authority, as against supposedly exact and impersonal observations and procedures, and denies the conventional distinction between 'discovery' and 'justification'. For what science aims at is truth, and not just trivial truths, but scientifically valuable ones. No random generation of problems and hypotheses can guide it, only the trained but tacit sense of the scientist for what is significant, probable and discoverable by himself with the means at his disposal. For example, it is a matter of judgment as to whether discrepant observations can be ignored as inevitable error, left to be cleared up later, or taken to require a revision of current theory (see Scientific method).

In his political and economic theory Polanyi argued that 'objectivist' assumptions have undermined our ability to uphold our personal commitments to intangible ideals. Totalitarian systems, such as Marxism, take advantage of this to discredit moral ideals and freedom, and disguise, in supposedly objective predictions of the future course of history, their own calls for centralized power to attempt a total renewal of society. Scepticism is no longer sufficient to combat destructive ideologies. A free society is one that is dedicated to transcendent ideals such as truth, justice and charity and not one that is 'open' to all ideologies. Its freedom is principally the 'positive' one to serve those ideals. But science, the economy and other activities of human life cannot be centrally planned. They manifest the spontaneous order of the mutual adjustment of individual elements.

Polanyi argued against the dichotomy of fact and value. The activities of life and mind have their own internal standards. To know them, we must grasp and apply those standards and recognize relevant entities and events as 'achievements', realizations of those standards, which cannot be neutrally described nor be apprehended by detached observation.

R.T. ALLEN

List of works

Polanyi, M. (1940) *The Contempt of Freedom*, London: Watts; repr. New York: Arno Press, 1975.(Four essays on the Soviet economy, the Marxist threat to science, and collectivist planning.)

Polanyi, M. (1945) *Full Employment and Free Trade*, Cambridge: Cambridge University Press; 2nd edn, 1948. (Polanyi's exposition of Keynes' theory.)
Polanyi, M. (1946) *Science, Faith and Society*, London: Oxford University Press; 2nd edn, with new preface, Chicago, IL: University of Chicago Press, 1964.(Science, as self-dedication in faith to the discovery of truth, provides a model for a free society.)
Polanyi, M. (1951) *The Logic of Liberty*, London: Routledge.(Ten papers on freedom in science, the economy and society, and the logical impossibility of central planning.)
Polanyi, M. (1958) *Personal Knowledge*, London: Routledge.(Polanyo's *magnum opus*, expounding and elaborating his central ideas on the personal, tacit and fiduciary roots of knowledge.)
Polanyi, M. (1959) *The Study of Man*, London: Routledge.(An elaboration of Polanyi's account of man as a responsible knower and agent.)
Polanyi, M. (1966) *The Tacit Dimension*, London: Routledge.(The structure of tacit integration and its implications.)
Polanyi, M. (1969) *Knowing and Being*, ed. M. Grene, London: Routledge.(Fourteen articles elaborating various aspects of Polanyi's philosophy.)
Polanyi, M. and Prosch, H. (1975) *Meaning*, Chicago, IL: University of Chicago Press. (Polanyi's final statement of his philosophy, focusing on the notion of meaning, in perception, language, art, myth, religion and society.)
Polanyi, M. (1996) *Society, Economic, Philosophy: Selected Papers by Michael Polanyi*, ed. R.T. Allen, New Brunswick, NJ: Transaction Publishers.(Twenty-five of Polanyi's articles not reprinted in his books, with full bibliography and summaries of all the articles not selected.)

References and further reading

Allen, R.T. (1990) *Polanyi*, London: Claridge Press.(A general introduction).
Allen, R.T. (1992) *Transcendence and Immanence in the Philosophy of Michael Polanyi and Christian Theism*, Lewiston, NY: Mellen for Rutherford House (Edinburgh).(Explores the theological applications of Polanyi's philosophy.)
Crewdson, J. (1994) *Christian Doctrine in the Light of Michael Polanyi's Theory of Personal Knowledge: A Personalist Theology*, Lewiston, NY: Mellen.(A theological development of Polanyi's philosophy.)
Gelwick, R. (1977) *The Way of Discovery*, New York: Oxford University Press.(A general introduction.)
Langford, T. and Poteat, W. (eds) (1968) *Intellect and Hope*, Durham, NC: Duke University Press.(Studies of particular aspects of Polanyi's work. Includes a bibliography up to 1967.)
Poteat, W. (1985) *Polanyian Meditations*, Durham, NC: Duke University Press. (Applications and elaborations of Polanyi's epistemology.)
Prosch, H. (1986) *Michael Polanyi: A Critical Exposition*, New York: State University of New York Press.(A more advanced introduction.)
Sanders, A.F. (1988) *Michael Polanyi's Post-Critical Epistemology: A Reconstruction of Some Aspects of 'Tacit Knowing'*, Amsterdam: Rodolfi.(Compares Polanyi with other contemporary philosophers.)
Scott, D. (1985) *Everyman Revived*, Lewes: The Book Guild.(A more informal introduction.)

Polish logic

The term 'Polish logic' was coined by McCall to signal the important contributions to modern logic by logicians from Poland between the wars. There were several centres of research, of which the Warsaw school, which grew out of the earlier Lwów-Warsaw philosophical movement, was the most significant. Its development was closely connected with the Warsaw school of mathematics, which gave it its characteristic mathematical bent.

Polish logic took as its point of departure the main trends in logical research of the time and it has influenced both subsequent logical research and subsequent work in the Western analytic tradition of philosophy. Its chief contributions were: (1) an enrichment of existing logical theory (including work on Boolean algebras, the sentential calculus, set theory, the theory of types); (2) new logical theories (for example, Leśniewski's systems, Łukasiewicz's many-valued logics, Tarski's theory of truth, theory of the consequence operation and the calculus of systems); (3) new methods and tools as well as improvements of existing methods (for example, the matrix method of constructing sentential calculi, axiomatizability of logical matrices, algebraic and topological interpretations of deductive systems, permutation models for set theory, the application of quantifier elimination to decidability and definability problems); and (4) the application of formal methods to the study of the history of logic, resulting in a new understanding of the logics of Aristotle, the Stoics and the medievals.

1 Historical introduction

In 1918, after 120 years of subjugation by foreign powers, Poland reappeared on the map of Europe as a free and independent nation-state. In 1939 the outbreak of the Second World War reversed these brief gains, bringing down the curtain on the Second Republic and inflicting massive damage on its economy, culture and learning. The main centres of logic research between these dates were Cracow, Lwów, Poznan, Wilno and, most importantly, Warsaw. The Warsaw school of logic, founded by Łukasiewicz and Leśniewski, began as an intellectual offshoot of the Lwów-Warsaw philosophical movement, but quickly eclipsed it in both quantity and quality of research.

The development of the Warsaw school of logic was closely connected with that of the Warsaw school of mathematics. In 1918 one could already speak of Warsaw as a fairly strong centre of research in set theory and topology under the direction of Janiszewski, Mazurkiewicz and Sierpiński. The two schools shared organizational structures, swapped directors and collaborated on many academic initiatives. This accelerated the development of both schools and deepened their research in key areas. In later years Kuratowski, Lindenbaum, Tarski and finally Mostowski made significant contributions to both schools.

2 Łukasiewicz: sentential calculi and history of logic

Łukasiewicz's more philosophical research culminated in his study of the principle of contradiction in Aristotle (1910), which was an important influence on the development of logic in Poland. After his discovery of a 'non-Aristotelian' many-valued logic (1918-20; see Łukasiewicz, J. §3), he turned to the study of the logic of propositions - in particular, to the sentential calculus, which he believed to be the simplest deductive discipline central to all scientific reasoning, and around which he built a research programme which served as a focus for research work on propositional logics throughout Poland (see Łukasiewicz 1941).

Łukasiewicz's programme can be seen as divided into two periods. The first, up to 1930, is characterized by conceptual and foundational achievements (see Łukasiewicz and Tarski 1930; Łukasiewicz 1929).

Notable among the conceptual achievements are a definition of a recursive grammar for propositional languages (including propositional quantifiers); the now famous bracket-free notation known as Polish notation; a precise definition of the syntactic operations of substitution and replacement; a new axiom system for classical propositional logic together with a new completeness proof for this system; formulations of several systems of many-valued logic (see Many-valued logics, philosophical issues in); a rigorous distinction between considerations of syntax and proof theory, on the one hand, and metalogical and semantic considerations on the other; and the invention of the matrix method of constructing sentential calculi.

Among the foundational achievements are various technical results in the search for single-sentence axiomatizations (notably by Tarski), Wajsberg's discovery of a criterion for finite axiomatizability, Lindenbaum's

theorem on the existence of adequate matrices for arbitrary systems of propositional logic closed under substitution, and pioneering studies of partial calculi (for example, the pure implicational and the pure equivalential fragments of classical sentential logic) and of sentential calculi extended by propositional quantifiers and variable functors of higher order.

The second period, from 1931 on, is characterized by the expansion of the programme to cover a greater number of propositional calculi, and its enrichment with new methods drawn from algebra and topology.

Noteworthy results of this period include a complete algebraic semantics for Lewis' S5 (see Wajsberg 1933), Jaśkowski's invention - independently of Gentzen - of the idea of a system of natural deduction (see Jaśkowski 1934; Natural deduction, tableau and sequent systems §1), a complete algebraic semantics for intuitionistic logic (Jaśkowski 1936), a topological interpretation of certain sentential calculi, notably the intuitionist variant (Tarski 1956), and a proof of the separability of the intuitionist connectives (Wajsberg 1938).

Still, a great deal of research in this period continued to be devoted to finite logical matrices. Wajsberg (1935), Sobociński (1936) and Słupecki (1939a) worked on solutions to the axiomatizability problem for many-valued logics. Słupecki (1939b) found a criterion for functional completeness of finite algebras.

Łukasiewicz also founded a programme of historical research whose goals were both a thorough exegesis and an analytical reconstruction of ancient and medieval logic. The source of this programme was his own book on Aristotle (1910), which distinguished ontological, logical and psychological formulations of the principle of contradiction. Łukasiewicz himself traced the historical roots of propositional logic to Stoic dialectic and countered prevailing views of the Middle Ages by shedding new light on the importance of the achievements of Petrus Hispanus and Duns Scotus (1934); he provided an axiomatic reformulation of Aristotle's syllogistic (1929, 1939, 1951; Logic, ancient §7). In the course of these investigations, Łukasiewicz introduced a new formal idea: the concept of a 'rejection system' - a Hilbert-type system for disproving rather than proving formulas - which he intended as a syntactic counterpart to the semantic device of refutation. This idea was subsequently elaborated by Słupecki (1949).

Under Łukasiewicz's influence a number of scholars outside Warsaw undertook related research. In Wilno, Czeżowski (1936) studied the treatment of modal sentences in Aristotle and Korcik (1937) provided an interpretation of some of Aristotle's writings in terms of a theory of the conversion of assertoric sentences. Bocheński (1938) wrote a history of the treatment of modal sentences in ancient and medieval times. In Poznan, Jordan (1937) studied the use of the axiomatic method from Plato and Eudoxus to Euclid's *Elements*. In Cracow, Salamucha analysed Saint Thomas' *ex motu* proof of the existence of God (1934), showed that William of Ockham had anticipated large parts of classical propositional calculus (1935) and wrote a scholarly treatise on medieval antinomies (1937).

Łukasiewicz's programme revealed that ancient and medieval logicians had considered 'formal' problems similar in character to some of the problems addressed by contemporary formal logic. This prompted a new interest in the subject and eventually led to a critical re-evaluation of large parts of the history of logic (see Scholz 1931; Łukasiewicz 1951; Mates 1953; Bocheński 1951, 1956; Kotarbiński 1957).

3 Leśniewski: foundations of mathematics

The main goal of Leśniewski was to provide a foundation for mathematics that would yield a real understanding of the antinomies and not simply avoid them. He did not accept the principles and assumptions of Cantorian set theory, regarding them as artificial and counterintuitive. He found the notion of an empty set an affront to reason. He was suspicious of *Principia Mathematica* on the grounds that its formalism was careless, ambiguous and did not pay sufficient attention to the use/mention distinction. He was an uncompromising critic of formalism for formalism's sake, and sought to ground mathematics in extensionalism, constructive nominalism and two-valued logic, insisting that only these principles would assure that it expressed meaningful, true sentences about the real world. He was, in the words of Tarski, the first person to become fully aware that a language which contains its own semantics and within which the usual logical laws hold must inevitably fall prey to liar-type paradoxes and hence be inconsistent.

The cornerstones of Leśniewski's work are *Mereology* (1927-31, outlined in 1916), *Ontology* (1920), his theory of

semantic categories (1929) and *Protothetic* (1929) (see Leśniewski 1929). The notion of a semantic category, due originally to Husserl, plays a role in Leśniewski's construction of formal theories similar to that played by the notion of type in *Principia Mathematica*. Ajdukiewicz later reformulated the theory of semantic categories in the form of a calculus (see Ajdukiewicz, K. §5; Ajdukiewicz 1978, article 7).

Constructive nominalism had important metalogical consequences for Leśniewski, confronting him with problems of how to describe a system, its language, axioms and rules (especially rules of definition). Notoriously, he clung to the belief that at any particular moment a logical system contains only as many theorems as have been proved up to that point in time. This led him to an extreme constructivism in the theory of definitions. Definitions in his system were 'creative', that is, after introducing definitions of new terms, the system was, in his sense, stronger than before. His definitions can thus not be regarded as mere abbreviational conveniences.

Tarski, in his 1924 doctoral dissertation under Leśniewski, investigated a fragment of *Protothetic* which might be called the second-order propositional calculus. He showed that in this system all sentential functors can be defined in terms of the equivalence connective and universal quantifier; and that one can prove counterparts to all the basic metalogical laws for classical propositional calculus. These results supported Leśniewski's conviction that metatheoretical considerations outside the system were dispensable.

Mereology influenced Tarski in his papers on solid geometry and Boolean algebras, and also found application in Woodger and Tarski's work on the axiomatic foundations of biology (see Woodger 1937). Sobociński (1934) published successive simplifications of the axiom system for *Ontology*. Kotarbiński's philosophy of reism, and his classic *Elements* (1929), owe much to the ideas of Leśniewski.

Leśniewski's influence spread chiefly through lectures and seminars. He was unquestionably one of the founding fathers of logic in Poland between the wars. Only after the Second World War, through some of his students' publications and a monograph by Luschei (1962), did the wider community take an interest in his achievements (see Srzednicki, Rickey and Czelakowski 1984). Some notes from his Warsaw lectures were collected and published in 1988.

4 Tarski: semantics and metamathematics

The role of Tarski in the development of Polish logic was in every way exceptional. Up to 1939 he published about sixty research papers covering various aspects of logic, set theory, measure theory (including the famous Banach-Tarski theorem on the 'paradoxical' decomposition of the sphere), axiomatic and structural aspects of Boolean algebras, truth theory and other areas of semantics (see Tarski 1956).

Tarski's contributions in pure logic were outstanding, and his work in formal or 'scientific' semantics and metamathematics helped shape Polish logic. His signal achievement was to establish that, for languages with specifiable structure, semantic notions such as satisfiability, truth, consequence and definability could be treated mathematically, and that this was fruitful (see Tarski's definition of truth; Consequence, conceptions of; Model theory).

A metamathematical theme was present in Polish logic from the early 1920s, beginning with Ajdukiewicz (1921). Tarski's efforts allowed investigations of the sentential calculi to be carried out within an explicitly metamathematical framework. Tarski generalized this framework to a mathematical theory of two primitive concepts (sentence and consequence) which he called 'the methodology of the deductive sciences'. Within this theory he was able to provide a conceptual apparatus for investigating deductive systems. (Lindenbaum contributed much to this work, including the widely known Lindenbaum maximality lemma.) Tarski later enlarged his metamathematics to encompass the calculus of systems and the concepts of ω-consistency and ω-completeness. These contributions were of a foundational nature, explicitly concerned with metatheory *per se*, and consciously framed in a meta-metalanguage. (See Tarski 1956.)

One of Tarski's best-known discoveries is the 1930 theorem on the eliminability of quantifiers for the theory of real closed fields, together with its consequences for the decidability of the arithmetic of real numbers and elementary geometry, and the definability of sets of reals. To this category also belongs Tarski and Mostowski's work of the late 1930s on the decidability of the theory of well-orderings.

A recurring motif in Tarski's work was the idea of algebraization. He believed that by abstracting an algebraic

calculus from a theory, the tools developed for the theory could be extended to other disciplines. This idea exerted a far-reaching influence on the subsequent course of logical investigations in Poland and elsewhere.

Largely through Tarski's work on truth, Polish logic also influenced philosophy. He spoke out against allowing one's repertoire of methods to be constrained by philosophical 'isms' such as formalism, logicism, constructivism or intuitionism (though his own scientism was exempt), and urged free use of all the resources of set theory, including transfinite techniques. He never accepted Wittgenstein's thesis that there is only one language of philosophical and scientific discourse, and persuaded some members of the Vienna Circle (most notably Carnap) to set it aside as well (see Tarski 1992).

5 Other work

Four recurring themes characterized the work of the Warsaw set theorists: (1) the axiom of choice, its equivalents and various weakenings, and its role within and beyond set theory; (2) the continuum hypothesis, its role in mathematics and its relation to the axiom of choice; (3) the problem of the independence of the axiom of choice from other principles of set theory; and (4) the arithmetic of cardinal and ordinal numbers (see Axiom of choice; Set theory).

Sierpiński's paper (1918) was the first major contribution to (1). He argued that before an informed decision could be made concerning whether to adopt the axiom of choice, one should first investigate its deductive strength. In 1924, Tarski discovered that several statements in the arithmetic of infinite cardinals were equivalent to the axiom of choice and also systematized the theory of finite sets and showed that it was necessary to assume the axiom of choice to prove the equivalence of various definitions of finiteness (see Tarski 1986: vol. 1, 39-48, 65-117). Lindenbaum was the first to pose the question 'What is the logical relation between the continuum hypothesis and the axiom of choice?'. These and related results were summarized in a 1926 paper by Lindenbaum and Tarski (see Tarski 1986: vol. 1, 171-204), where the two authors famously stated, without proof, that the generalized continuum hypothesis implies the axiom of choice (188, theorem 94). In 1934, Sierpiński published a compendium of principles known to be consequences of the generalized continuum hypothesis, but it was not until 1947 that he was able to supply the missing proof (1976: 485-8).

In 1938 Mostowski investigated the independence of definitions of finiteness in a system of type theory by means of the method of relativization of quantifiers (see Mostowski 1979: vol. 2, 18-67). During the years 1935-8 he and Lindenbaum also refined and extended Fraenkel's permutation models to investigate the independence of the axiom of choice from the ordering principle and various weaker forms of the axiom (vol. 1, 290-338; vol. 2, 70-4).

In Cracow, Sleszyński published a survey of the development of formal logic from Aristotle to Russell and Whitehead. Especially valuable is his treatment of nineteenth-century mathematical logic (1925-9). W. Wilkosz wrote textbooks on modern set theory and axiomatic foundations of number theories.

Chwistek developed a system for the foundation of logic and mathematics. Starting from several improvements of the ramified theory of types (see Theory of types), he arrived at a purely constructivist theory without the axioms of reducibility and extensionality and without impredicative definitions (1924-5). Later, with his pupils J. Herzberg and W. Hetper, he outlined a formal theory of expressions called 'rational metamathematics' (1935). Chwistek's work in logic was strongly influenced by his philosophical beliefs, which set him apart from the mainstream of Polish logic (1961-3).

Pepis (1937, 1938) investigated the decision problem for the predicate calculus; more specifically, he dealt with the problem of reducing the question of the satisfiability of a logical formula from the general case to that of a formula having a special normal form.

6 Post-war developments

The Second World War caused the deaths of many logicians, including Lindenbaum and his wife J. Hosiasson-Lindenbaum (who worked on inductive logic), Herzberg, Hetper, Pepis, Salamucha, Z. Schmierer and Wajsberg. Others were forced to emigrate: Bocheński, H. Hiż, Jordan, Łukasiewicz, H. Mehlberg, Sobociński and Tarski. Manuscripts were lost or destroyed, among them Łukasiewicz's manuscript for a book on syllogistic, Mostowski's notes on the proof of quantifier elimination for well-orderings, and Leśniewski's *Nachlaß*. Galley

proofs of a new journal, *Collectanea Logica*, were destroyed in 1939 and the journal never appeared.

Post-war Polish logic picked up where pre-war work had left off. In mathematical logic and the foundations of mathematics, the main figure was Mostowski (1979), whose contributions were wide-ranging, along with A. Grzegorczyk, A. Erenfeucht, J. Łoś, C. Ryll-Nardzewski and R. Suszko. Łośis best known for his 1954 notion of categoricity-in-power, his 1955 lemma on ultraproducts and generally for his work on model theory (parts of which were carried out in collaboration with Suszko). Ryll-Nardzewski contributed to model theory (omitting types and ω-categoricity) and arithmetic (showing first-order Peano arithmetic is not finitely axiomatizable). Together, Łośand Ryll-Nardzewski shed light on the logical importance of Tychonoff's compactness theorem and the prime ideal theorem. Joint papers by Grzegorczyk, Mostowski and Ryll-Nardzewski (see Mostowski 1979: vol. 1, 371-89; vol. 2, 427-31) are important for our understanding of natural numbers and ω-logic. Mostowski's expository essays and surveys have influenced the wider philosophical community (see Mostowski 1952, 1955, 1965).

Tarski's ideas of the algebraization of logic and of metamathematics were elaborated in numerous papers by Rasiowa and Sikorski and culminated in a classic work, *The Mathematics of Metamathematics* (1963).

In 1958 Łośand Suszko proposed applying Tarski's theory of the consequence operation to investigations of abstract sentential calculi. This became a research paradigm of post-war Polish logic, drawing on concepts and tools from contemporary model theory and universal algebra. Its principal contributors were J. Czelakowski, W. Dziobak, W.A. Pogorzelski, T. Prucnal, Suszko, P. Wojtylak, R. Wójcicki and A. Wroński. Wójcicki (1987) summarizes the most important results of this programme.

The following logical journals were established in Poland after the Second World War: *Studia Logica* (in 1953), *Bulletin of the Section of Logic* (in 1972), *Reports on Mathematical Logic* (in 1973) and *Logic and Logical Philosophy* (in 1993). An account of post-war Polish logic can be found in Wójcicki (1997) and that of the foundations of mathematics in Marek (1977).

7 Polish notation

In Polish logical notation all operators are placed before the variables over which they have scope.

Np	Negation ($\neg$)
Kpq	Conjunction ($\wedge$)
Apq	Weak or inclusive disjunction ($\vee$)
Jpq	Strong or exclusive disjunction, or non-equivalence ($\not\equiv$)
Cpq	Conditional ($\rightarrow$)
Epq	Biconditional ($\leftrightarrow$)
Dpq	Sheffer stroke or non-conjunction ($\mid$)
Πpq	Universal quantification ($\forall$) (of q with respect to the sentential variable p)
Σpq	Existential quantification ($\exists$) (of q with respect to the sentential variable p)
ΠxFx	Universal quantification (of Fx with respect to the individual variable x)
ΣxFx	Existential quantification (of Fx with respect to the individual variable x)

Because we know for each operator whether it is one-place or two-place, we know how many of the following letters fall within its scope. So parentheses are not required to avoid ambiguity or to show association. For example, $p \rightarrow (q \rightarrow r)$ is written *CpCqr* and $(p \rightarrow q) \rightarrow r$ is *CCpqr*; and

$((p \wedge q) \rightarrow \neg r) \rightarrow (r \rightarrow (\neg p \vee \neg q))$

is *CCKpqNrCrANpNq*.

The parenthesis-free notation, later called Polish notation, was invented by Łukasiewicz in 1924 and published (1929) in application to the sentential calculus, the sentential calculus with quantifiers and Aristotle's syllogistic. (See also Łukasiewicz and Tarski 1930: 39, note 2; 54-.) The notation is very convenient in theoretical considerations; sometimes it is also used in textbooks (see, for example, Prior 1955).

See also: Logical and mathematical terms, glossary of

JAN ZYGMUNT

References and further reading

Ajdukiewicz, K. (1921) *Z metodologii nauk dedukcyjnych*, Lwów: Nakładem Polskiego Towarzystwa Filozoficznego; trans. J. Giedymin, 'From the Methodology of Deductive Sciences', *Studia Logica* 19: 9-46, 1966.(Referred to in §4.)

Ajdukiewicz, K. (1978) *The Scientific-World Perspective and other Essays 1931-1963*, ed. J. Giedymin, Dordrecht: Reidel.(Twenty-two papers in English translation with an introduction by the editor comparing Ajdukiewicz's radical conventionalism to the conventionalism of Poincaré and LeRoy; referred to in §3.)

Bocheński, I.M. (1938) *Z historii zdańmodalnych (From the History of Modal Sentences)*, Lwów: Wydawnictwo OO. Domnikanów.(Referred to in §2.)

Bocheński, I.M. (1951) *Ancient Formal Logic*, Amsterdam: North Holland.(Referred to in §2.)

Bocheński, I.M. (1956) *Formale Logik*, Freiburg and Munich: Alber; trans. I. Thomas, *A History of Formal Logic*, Notre Dame, IN: University of Notre Dame Press, 1961; repr. New York: Chelsea, 1970. (Referred to in §2.)

Chwistek, L. (1924-5) 'The Theory of Constructive Types (Principles of Logic and Mathematics)', *Annales de la Société Polonaise de Mathématique* 2: 9-48, 3: 91-141.(Referred to in §5.)

Chwistek, L. (1935) *Granice nauki: zarys logiki i metodologii nauk ścislych*, Warsaw and Lwów: Książnica Atlas; revised edn trans. H.C. Brodie and A.P. Coleman, *The Limits of Science: Outline of Logic and of the Methodology of the Exact Sciences*, London: Routledge & Kegan Paul, 1948.(Referred to in §5.)

Chwistek, L. (1961-3) *Pisma filozoficzne i logiczne (Philosophical and Logical Writings)*, ed. K. Pasenkiewicz, Warsaw: PWN, 2 vols.(Almost all of Chwistek's works written in Polish: in volume 1, his philosophical, methodological and aesthetic writings; in volume 2, his logical and mathematical ones (together with *Granice nauki*). Referred to in §5.)

Czeżowski, T. (1936) 'Arystotelesa teoria zdańmodalnych' (Aristotle's Theory of Modal Propositions), *Przegląd Filozoficzny* 39 (3): 232-41.(Referred to in §2.)

Jaśkowski, S. (1934) *On the Rules of Supposition in Formal Logic*, Warsaw: Nakładem Seminarium Filozoficznego Wydzialu Matematyczno-Przyrodniczego Uniwersytetu Warszawskiego; repr. with notational changes in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967. (Referred to in §2.)

Jaśkowski, S. (1936) 'Recherches sur le système de la logique intuitioniste', in *Actes du Congrès International de Philosophie Scientifique, Paris 1935*, vol. 6, *Actualités Scientifiques et Industrielles*, Paris: Hermann, 58-61; trans. in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967. (Referred to in §2. For a detailed elaboration of Jaśkowski's semantics for intuitionist logic see Surma (1971).)

Jordan, Z. (1937) *O matematycznych podstawach systemu Platona: z historii racjonalizmu (Mathematical Foundations of Plato's System: On the History of Rationalism)*, with a French summary, Poznan: Nakładem Poznańskiego Towarzystwa PrzyjaciółNauk.(Referred to in §2.)

Jordan, Z. (1945) *The Development of Mathematical Logic and of Logical Positivism in Poland Between the Two Wars*, Oxford: Oxford University Press; partly repr. in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967, 346-406.(Succinct and readable, but without technical details; lacks historical perspective.)

Korcik, A. (1937) *Teoria konwersji zdańasertorycznych u Aristotelesa w świetle teorii dedukcji: studium historyczno-krytyczne (The Theory of Conversion of Assertoric Propositions in Aristotle in the Light of the Theory of Deduction: Critical and Historical Study)*, Wilno: Wydawnictwo Studiów Teologicznych.(Referred to in §2.)

Kotarbiński, T. (1929) *Elementy teorii poznania, logiki formalnej i metodologii nauk (Elements of the Theory of Knowledge, Formal Logic and the Methodology of Sciences)*, Lwów: Ossolineum; 2nd revised edn, Warsaw: PWN, 1961; 2nd edn trans. O. Wojtasiewicz, *Gnosiology: The Scientific Approach to the Theory of Knowledge*, Oxford: Pergamon Press, and Wrocław: Ossolineum, 1966. (Referred to in §3.)

Kotarbiński, T. (1957) *Wykłady z dziejów logiki (Lectures on the History of Logic)*, Wrocław and Łódź: Ossolineum; 2nd edn, Warsaw: PWN, 1985.(Referred to in §2.)

Leśniewski, S. (1988) *S. Leśniewski's Lecture Notes in Logic*, ed. J.T.J. Srzednicki and Z. Stachniak, Dordrecht: Kluwer.(Notes from lectures and seminars given by Leśniewski in the period 1919-39. They cover topics ranging from his own systems to Peano arithmetic and Whitehead's theory of events. Referred to in §3.)

Leśniewski, S. (1992) *Collected Works*, ed. S.J. Surma, J.T. Srzednicki and D.I. Barnet, with an annotated bibliography by V.F. Rickey, Warsaw: PWN, and Dordrecht: Kluwer, 2 vols.(English translation of all Leśniewski's published works. Flawed in some respects; see a review in *History and Philosophy of Logic* 15: 227-35, 1994.)

Łukasiewicz, J. (1910) *O zasadzie sprzeczności u Arystotelesa: studyum krytyczne (On the Principle of Contradiction in Aristotle: A Critical Study)*, Cracow: Polska Akademia Umiejętności; 2nd edn with an introduction and notes by J. Woleński, Warsaw: PWN, 1987.(Referred to in §2.)

Łukasiewicz, J. (1929) *Elementy logiki matematycznej*, ed. M. Presburger, Warsaw: PWN; 2nd revised and abridged edn, ed. J. Słupecki, Warsaw: PWN; trans. *Elements of Mathematical Logic*, Oxford: Pergamon Press, and New York: Macmillan, 1963.(Referred to in §2.)

Łukasiewicz, J. (1934) 'Z historii logiki zdań', *Przegląd Filozoficzny* 37: 417-37; German trans. by the author, 'Zur Geschichte der Aussagenlogik', *Erkenntnis* 5: 111-31, 1935; trans. On the History of the Logic of Propositions, in *Selected Works*, ed. L. Borkowski, Amsterdam: North Holland, 1970; and in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967.(According to H. Scholz, the most interesting thirty pages ever written on the history of logic. Referred to in §2.)

Łukasiewicz, J. (1939) 'O sylogistyce Arystotelesa' (On Aristotle's Syllogistic), *Sprawozdania Polskiej Akademii Umiejętności* 44: 220-7.(Referred to in §2.)

Łukasiewicz, J. (1941) 'Die Logik und das Grundlagenproblem', *Les entretiens de Zürich sur les fondements et la méthode de sciences mathématiques, 6-9 décembre 1938*, 82-100; trans. 'Logic and the Problem of the Foundations of Mathematics', in *Selected Works*, ed. L. Borkowski, Amsterdam: North Holland, 1970. (Referred to in §2.)

Łukasiewicz, J. (1951) *Aristotle's Syllogistic from the Standpoint of Modern Formal Logic*, Oxford: Oxford University Press; 2nd enlarged edn, 1957.(Referred to in §2.)

Łukasiewicz, J. (1970) *Selected Works*, ed. L. Borkowski, Amsterdam: North Holland.(Anthology of all important papers from the entire creative work of Łukasiewicz.)

Łukasiewicz, J. and Tarski, A. (1930) 'Untersuchungen über den Aussagenkalkül', *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie* (Classe 3) 23: 30-50; trans. 'Investigations into the Sentential Calculus', in J. Łukasiewicz, *Selected Works*, ed. L. Borkowski, Amsterdam: North Holland, 1970; and in A. Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press, 1956.(Referred to in §2.)

Luschei, E.C. (1962) *The Logical Systems of Leśniewski*, Amsterdam: North Holland.(The only book which describes the whole of Leśniewski's scientific achievements; includes a large chapter on terminological explanations and directives. An invaluable and unique guide for those who want to read the originals, but it is not a systematic exposition of Leśniewski's systems. Referred to in §3.)

Marek, W. (1977) 'The Foundations of Mathematics in Poland after World War II', in R.O. Gandy and J.M.E. Hyland (eds) *Logic Colloquium '76*, Amsterdam: North Holland, 129-38.(Referred to in §6.)

Mates, B. (1953) *Stoic Logic*, Berkeley and Los Angeles, CA: University of California Press.(Referred to in §2.)

McCall, S. (ed.) (1967) *Polish Logic 1920-1939*, Oxford: Clarendon Press.(Classic anthology of papers by Ajdukiewicz, Chwistek, Jaśkowski, Leśniewski, Łukasiewicz, Słupecki and Wajsberg, with notes by Kotarbiński on the development of formal logic in Poland in the years 1900-39; includes part of Jordan (1945).)

Mostowski, A. (1952) *Sentences Undecidable in Formalized Arithmetic: An Exposition of the Theory of Kurt Gödel*, Amsterdam: North Holland.(Referred to in §6.)

Mostowski, A. (1965) 'Thirty Years of Foundational Studies: Lectures on the Development of Mathematical Logic and the Study of the Foundations of Mathematics in 1930-1964', *Acta Philosophica Fennica* 17: 1-180; repr. New York: Barnes & Noble, 1966; and in *Foundational Studies, vol. 1*, Amsterdam: North Holland, 1979. (History of mathematical logic from Gödel to P.J. Cohen, written from a Polish point of view; referred to in §6.)

Mostowski, A. (1979) *Foundational Studies*, Amsterdam: North Holland, 2 vols.(An anthology of the most important of Mostowski's works from 1937-76; includes a complete bibliography and five essays summarizing his work on decidability, recursion and hierarchy theory, set theory, model theory, logical calculi and arithmetic. Volume 1 may be of interest for its survey and expository papers. Referred to in §§5-6.)

Mostowski, A. *et al*. (1955) ‘The Present State of Investigations on the Foundations of Mathematics’, *Rozprawy Matematyczne (Dissertationes Mathematicae)* 9: 1-48.(Referred to in §6.)

Pearce, D. and Woleński, J. (eds) (1988) *Logischer Rationalismus: Schriften der Lemberg-Warschauer Schule (Logical Rationalism: Writings of the Lwów-Warsaw School)*, Frankfurt: Athenaeum.(Shorter papers, in German, by Ajdukiewicz, J. Hosiasson-Lindenbaum, M. Kokoszyńska, Kotarbiński, Łukasiewicz and Tarski.)

Pepis, J. (1937) *O zagadnieniu rozstrzygalności w zakresie węższego rachunku funkcyjnego (On the Decision Problem in the Domain of the Restricted Functional Calculus)*, with a German summary, Lwów: Archiwum Towarzystwa Naukowego we Lwowie, §3, vol. 7, fasc. 8, 1-172 ([465]-[616]).(Referred to in §5.)

Pepis, J. (1938) ‘Untersuchungen über das Entscheidungsproblem der mathematischen Logik’ (Investigations into the Decision Problem of Mathematical Logic), *Fundamenta Mathematicae* 30: 257-348.(Referred to in §5.)

Prior, A. (1955) *Formal Logic*, Oxford: Clarendon Press; 2nd edn with corrections, 1962.(An example of the use of Polish notation in a textbook.)

Rasiowa, H. and Sikorski, R. (1963) *The Mathematics of Metamathematics*, Warsaw: PWN.(Referred to in §6.)

Salamucha, J. (1934) ‘Dowód ‘‘ex motu’’ na istnienie Boga: analiza logiczna argumentacji św. Tomasza z Akwinu’, *Collectanea Theologica* 15: 53-92; trans. ‘The Proof *ex motu* for the Existence of God: Logical Analysis of St. Thomas’ Arguments’, *The New Scholasticism* 32: 334-72, 1958. (Referred to in §2.)

Salamucha, J. (1935) ‘Logika zdańu Wilhelma Ockhama’ (Propositional Logic in William of Ockham), *Przegląd Filozoficzny* 38 (3): 208-39.(Referred to in §2.)

Salamucha, J. (1937) ‘Pojawienie sięzagadnieńantynomialnych na gruncie logiki średniowiecznej’ (The Appearance of the Problem of Antinomies in Medieval Logic), *Przegląd Filozoficzny* 40 (1): 68-89, 40 (3): 320-43.(Referred to in §2.)

Scholz, H. (1931) *Geschichte der Logik*, Berlin; repr. as *Abriß der Geschichte der Logik*, Freiburg: Alber, 1959; trans. K.F. Leidecker, *Concise History of Logic*, New York: Philosophical Library, 1961.(Referred to in §2.)

Sierpiński, W. (1918) ‘L’ axiome de M. Zermelo et son rôle dans théorie des ensembles et l’analyse’ (Zermelo’s Axiom and its Role in Set Theory and Analysis), *Bulletin de l’Académie des Sciences de Cracovie, Classe des Sciences Mathématiques*, série A, 97-152; repr. in *Oeuvres Choisies*, vol. 2, *Théorie des Ensembles et ses Applications, Travaux des Années 1908-1929*, Warsaw: PWN-Editions Scientifiques de Pologne, 1975. (Referred to in §5.)

Sierpiński, W. (1934) *Hypothèse du Continu (The Continuum Hypothesis)*, Monografie Matematyczne, vol. 4, Warsaw: Grasiński.(Referred to in §5.)

Sierpiński, W. (1976) *Oeuvres Choisies (Selected Works)*, vol. 3, *Théorie des Ensembles et ses Applications, Travaux des Années 1930-1966 (Set Theory and its Applications: Papers from 1930 to 1966)*, Warsaw: PWN-Editions Scientifiques de Pologne.(Referred to in §5.)

Sleszyński, J. (1925-9) *Teoria dowodu. Podług wykładów uniwersyteckich Prof. Dr Jana Sleszyńskiego opracowałS.K. Zaremba (The Theory of Proof. Edited by S.K. Zaremba from the University Lectures of Professor Dr. Jan Sleszyński)*, Cracow: Nakładem Kółka Matematyczno-Fizycznego Uczniów Uniwersytetu Jagiellońskiego, 2 vols.(Referred to in §5.)

Słupecki, J. (1939a) ‘Dowód aksjomatyzowalności pełnych systemów wielowartościowych rachunku zdań’, *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie* (Classe 3) 32: 110-28; trans. ‘A Proof of the Axiomatizability of Full Systems of Many-Valued Sentential Calculus’, *Studia Logica* 29: 155-68, 1971.(Referred to in §2.)

Słupecki, J. (1939b) ‘Kryterium pełności wielowartościowych systemów logiki zdań’, *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie* (Classe 3) 32: 102-9; trans. ‘A Criterion of Fullness for Many-Valued Systems of Propositional Logic’, *Studia Logica* 30: 153-7, 1972. (Referred to in §2.)

Słupecki, J. (1949) ‘On Aristotelian Syllogistic’, *Studia Philosophica* (Poznań) 4: 275-300.(Referred to in §2.)

Sobociński, B. (1934) ‘O kolejnych uproszczeniach aksjomatyki ‘‘ontologii’’ prof. S. Leśniewskiego’, in *Fragmenty Filozoficzne. Księga Pamiątkowa ku Uczczeniu Piętnastolecia Pracy Nauczycielskiej w Uniwersytecie Warszawskim Profesora Tadeusza Kotarbińskiego*, Warsaw: Nakładem Uczniów, 144-60; trans. ‘Successive Simplifications of the Axiom-System of Leśniewski’s Ontology’, in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967.(Referred to in §3.)

Sobociński, B. (1936) ‘Aksjomatyzacja pewnych wielowartościowych systemów teorii dedukcji’ (Axiomatization of Certain Many-Valued Systems of the Theory of Deduction), *Roczniki Prac Naukowych Zrzeszenia Asystentów Uniwersytetu Józefa Piłsudskiego w Warszawie* 1: 399-419. (Referred to in §2.)

Srzednicki, J.T.J., Rickey, V.F. and Czelakowski, J. (eds) (1984) *Leśniewski’s Systems: Ontology and Mereology*, The Hague: Martinus Nijhoff, and Wrocław: Ossolineum.(Twelve representative papers on ontology and mereology by Leśniewski’s pupils, Z. Kruszewski, C. Lejewski, J. Słupecki, B. Sobociński; and

their pupils, J. Canty, R. Clay and B. Iwanuś. Referred to in §3.)

Surma, S.J. (1971) 'Jaśkowski's Matrix Criterion for the Intuitionistic Propositional Calculus', *Universitas Iagiellonica Acta Scientiarum Litterarumque*, vol. 265, *Schedae Logicae* 6: 21-54.(See Jaśkowski 1936.)

Surma, S.J. (1977) 'On the Work and Influence of Stanisław Leśniewski', in R.O. Gandy and J.M.E. Hyland (eds) *Logic Colloquium '76*, Amsterdam: North Holland, 191-220.(A biography and annotated bibliography of Leśniewski, with an outline of his logical systems.)

Surma, S.J. (1979) 'On the Origin and Subsequent Application of the Concept of Lindenbaum Algebra', in L.J. Cohen, J. Łoś, H. Pfeiffer and K.-H. Podewski (eds) *Logic, Methodology and Philosophy of Science*, Amsterdam: North Holland, vol. 6, 719-34.(Rather technical exposition of the subject described in the title, with a full bibliography of Lindenbaum's works.)

Surma, S.J. (1987) 'The Logical Work of Mordchaj Wajsberg', in J. Srzednicki (ed.) *Initiatives in Logic*, Dordrecht: Martinus Nijhoff, 101-15.(An account of Wajsberg's life and his logical research.)

Szumilewicz-Lachman, I. and Cohen, R.S. (ed.) (1994) *Zygmunt Zawirski: His Life and Work. With Selected Writings on Time, Logic and Methodology of Science*, trans. F. Lachman, Dordrecht: Kluwer.(Valuable especially for Zawirski's two papers, 'The Genesis and Development of Intuitionistic Logic' and 'The Evolution of the Notion of Time'; also includes essays by Szumilewicz-Lachman on the Lwów-Warsaw school and the Vienna Circle.)

Tarski, A. (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press; repr. and ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(Referred to in §§2, 4. For a careful presentation of the content of the book see the review by W.A. Pogorzelski and S.J. Surma, *Journal of Symbolic Logic* 34 (1): 99-106, 1969.)

Tarski, A. (1986) *The Collected Papers*, ed. S.R. Givant and R.N. McKenzie, Basle: Birkhäuser, 4 vols. (Reprints of all of Tarski's writings excluding his books; unfortunately original pagination has not been kept.)

Tarski, A. (1992) 'Alfred Tarski: Drei Briefe an Otto Neurath', ed. R. Haller, trans. J. Tarski, *Grazer Philosophische Studien* 43: 1-32.(Referred to in §4.)

Wajsberg, M. (1933) 'Ein erweiterter Klassenkalkül', *Monatshefte für Mathematik und Physik* 40: 113-26; trans. 'An Extended Class Calculus', in *Logical Works*, ed. S.J. Surma, Wrocław: Ossolineum, 1977.(Referred to in §2.)

Wajsberg, M. (1935) 'Beiträge zum Metaaussagenkalkül I', *Monatshefte für Mathematik und Physik* 42: 221- 42; trans. 'Contributions to the Meta-Calculus of Propositions', in *Logical Works*, ed. S.J. Surma, Wrocław: Ossolineum, 1977.(Referred to in §2.)

Wajsberg, M. (1938) 'Untersuchungen über den Aussagenkalkül von A. Heyting', *Wiadomości Matematyczne* 46: 45-101; trans. 'On A. Heyting's Propositional Calculus', in *Logical Works*, ed. S.J. Surma, Wrocław: Ossolineum, 1977.(Referred to in §2.)

Wójcicki, R. (1987) *Theory of Sentential Calculi. An Introduction*, Dordrecht: Reidel.(Referred to in §6.)

Wójcicki, R. (1997) 'The Postwar Panorama of Logic in Poland', in M.L. dalla Chiara *et al.* (eds) *Logic and Scientific Methods*, Dordrecht: Kluwer, 497-508.(Referred to in §6.)

Woleński, J. (1989) *Logic and Philosophy in the Lvov-Warsaw School*, Dordrecht: Kluwer.(An authoritative and comprehensive treatise with an excellent bibliography.)

Woodger, J.H. (1937) *The Axiomatic Method in Biology*, with appendix by A. Tarski, Cambridge: Cambridge University Press, and New York: Macmillan.(Referred to in §3.)

Political philosophy

Political philosophy can be defined as philosophical reflection on how best to arrange our collective life - our political institutions and our social practices, such as our economic system and our pattern of family life. (Sometimes a distinction is made between *political* and *social* philosophy, but I shall use ‘political philosophy’ in a broad sense to include both.) Political philosophers seek to establish basic principles that will, for instance, justify a particular form of state, show that individuals have certain inalienable rights, or tell us how a society’s material resources should be shared among its members. This usually involves analysing and interpreting ideas like freedom, justice, authority and democracy and then applying them in a critical way to the social and political institutions that currently exist. Some political philosophers have tried primarily to justify the prevailing arrangements of their society; others have painted pictures of an ideal state or an ideal social world that is very different from anything we have so far experienced (see Utopianism).

Political philosophy has been practised for as long as human beings have regarded their collective arrangements not as immutable and part of the natural order but as potentially open to change, and therefore as standing in need of philosophical justification. It can be found in many different cultures, and has taken a wide variety of forms. There are two reasons for this diversity. First, the methods and approaches used by political philosophers reflect the general philosophical tendencies of their epoch. Developments in epistemology and ethics, for instance, alter the assumptions on which political philosophy can proceed. But second, the political philosopher’s agenda is largely set by the pressing political issues of the day. In medieval Europe, for instance, the proper relationship between Church and State became a central issue in political philosophy; in the early modern period the main argument was between defenders of absolutism and those who sought to justify a limited, constitutional state. In the nineteenth century, the social question - the question of how an industrial society should organize its economy and its welfare system - came to the fore. When we study the history of political philosophy, therefore, we find that alongside some perennial questions - how can one person ever justifiably claim the authority to govern another person, for instance? - there are some big changes: in the issues addressed, in the language used to address them, and in the underlying premises on which the political philosopher rests his or her argument. (For the development of the Western tradition of political philosophy, see political philosophy, history of; for other traditions, see Political philosophy in classical Islam; Political philosophy, Indian; African philosophy, Anglophone; Marxism, Chinese; Bushi philosophy; Shōtoku constitution; Sunzi; Marxist thought in Latin America.)

One question that immediately arises is whether the principles that political philosophers establish are to be regarded as having universal validity, or whether they should be seen as expressing the assumptions and the values of a particular political community. This question about the scope and status of political philosophy has been fiercely debated in recent years (see Political philosophy, nature of). It is closely connected to a question about human nature (see Human nature). In order to justify a set of collective arrangements, a political philosophy must say something about the nature of human beings, about their needs, their capacities, about whether they are mainly selfish or mainly altruistic, and so forth. But can we discover common traits in human beings everywhere, or are people’s characters predominantly shaped by the particular culture they belong to?

If we examine the main works of political philosophy in past centuries, they can be divided roughly into two categories. On the one hand there are those produced by philosophers elaborating general philosophical systems, whose political philosophy flows out of and forms an integral part of those systems. Leading philosophers who have made substantial contributions to political thought include Plato, Aristotle, Augustine, Aquinas, Hobbes, Locke, Hume, Hegel and J.S. Mill. On the other hand there are social and political thinkers whose contribution to philosophy as a whole has had little lasting significance, but who have made influential contributions to political philosophy specifically. In this category we may include Cicero, Marsilius of Padua, Machiavelli, Grotius, Rousseau, Bentham, Fichte and Marx. Two important figures whose work reflects non-Western influences are Ibn Khaldhun and Kauṭilya. Among the most important twentieth-century political thinkers are Arendt, Berlin, Dewey, Foucault, Gandhi, Gramsci, Habermas, Hayek, Oakeshott, Rawls, Sartre and Taylor.

1 Political institutions and ideologies

What are the issues that, historically and today, have most exercised political philosophers? To begin with, there is a set of questions about how political institutions should be arranged. Today we would think of this as an enquiry into the best form of state, though we should note that the state itself is a particular kind of political arrangement of relatively recent origin - for most of their history human beings have not been governed by states (see State, the). Since all states claim Authority over their subjects, two fundamental issues are the very meaning of authority, and the criteria by which we can judge forms of political rule legitimate (see Legitimacy; Contractarianism; General will; Power; Tradition and traditionalism). Connected to this is the issue of whether individual subjects have a moral obligation to obey the laws of their state (see Obligation, political), and of the circumstances under which politically-inspired disobedience is justifiable (see Civil disobedience; Revolution). Next there is a series of questions about the form that the state should take: whether authority should be absolute or constitutionally limited (see Absolutism; Constitutionalism); whether its structure should be unitary or federal (see Federalism and confederalism); whether it should be democratically controlled, and if so by what means (see Democracy; Representation, political). Finally here there is the question of whether any general limits can be set to the authority of the state - whether there are areas of individual freedom or privacy that the state must never invade on any pretext (see Law, limits of; Freedom of speech; Coercion; Property; Slavery), and whether there are subjects such as religious doctrine on which the state must adopt a strictly neutral posture (see Neutrality, political; Toleration).

Beyond the question of how the state itself should be constituted lies the question of the general principles that should guide its decisions. What values should inform economic and social policy for instance? Part of the political philosopher’s task is to examine ideas that are often appealed to in political argument but whose meaning remains obscure, so that they can be used by politicians from rival camps to justify radically contrasting policies. Political philosophers try to give a clear and coherent account of notions such as Equality, Freedom and liberty, Justice, Needs and interests, Public interest, Rights and Welfare. And they also try to determine whether these ideas are consistent with, or conflict with, one another - whether, for instance, equality and liberty are competing values, or whether a society might be both free and equal at once.

Further questions arise about the principles that should guide one state in its dealings with other states. May states legitimately pursue what they regard as their national interests, or are they bound to recognize ethical obligations towards one another (see Development ethics)? More widely, should we be seeking a cosmopolitan alternative under which principles of justice would be applied at global level? (see International relations, philosophy of; Justice, international). When, if ever, are states justified in going to war with each other? (See War and peace, philosophy of.)

Over about the last two centuries, political debate has most often been conducted within the general frameworks supplied by rival ideologies. We can think of an ideology as a set of beliefs about the social and political world which simultaneously makes sense of what is going on, and guides our practical responses to it (see Ideology). Ideologies are often rather loosely structured, so that two people who are both conservatives, say, may reach quite different conclusions about some concrete issue of policy. Nevertheless they seem to be indispensable as simplifying devices for thinking about a political world of ever-increasing complexity.

No political philosopher can break free entirely from the grip of ideology, but political philosophy must involve a more critical scrutiny of the intellectual links that hold ideologies together, and a bringing to light of the unstated assumptions that underpin them. The most influential of these ideologies have been Liberalism, Conservatism, Socialism, nationalism (see Nation and nationalism; Eurasian movement; Pan-Slavism; Zionism; Pan-Africanism) and Marxism (see Marxism, Western; Marxist philosophy, Russian and Soviet; Marxism, Chinese). Other ideologies are of lesser political significance, either because they have drawn fewer adherents or because they have been influential over a shorter period of time: these include Anarchism, Communism, Fascism, Libertarianism, Republicanism, Social democracy and Totalitarianism.

2 Contemporary political philosophy

The last quarter of the twentieth century has seen a powerful revival of political philosophy, which in Western societies at least has mostly been conducted within a broadly liberal framework. Other ideologies have been outflanked: Marxism has gone into a rapid decline, and conservatism and socialism have survived only by taking on board large portions of liberalism. Some have claimed that the main rival to liberalism is now

communitarianism (see Community and communitarianism); however on closer inspection the so-called liberal-communitarian debate can be seen to be less a debate about liberalism itself than about the precise status and form that a liberal political philosophy should take - whether, for example, it should claim universal validity, or should present itself simply as an interpretation of the political culture of the Western liberal democracies. The vitality of political philosophy is not to be explained by the emergence of a new ideological revival to liberalism, but by the fact that a new set of political issues has arisen whose resolution will stretch the intellectual resources of liberalism to the limit.

What are these issues? The first is the issue of social justice, which in one form or another has dominated political philosophy for much of the century. Most of the many liberal theories of justice on offer have had a broadly egalitarian flavour, demanding at least the partial offsetting of the economic and social inequalities thrown up by an unfettered market economy (see Market, ethics of the; Justice; Rawls, J.; Dworkin, R.; though for dissenting views see Hayek, F.A. von; Nozick, R.). These theories rested on the assumption that social and economic policy could be pursued largely within the borders of a self-contained political community, sheltered from the world market. This assumption has become increasingly questionable, and it presents liberals with the following dilemma: if the pursuit of social justice is integral to liberalism, how can this be now be reconciled with individual freedoms to move, communicate, work, and trade across state boundaries?

The second issue is posed by feminism, and especially the feminist challenge to the conventional liberal distinction between public and private spheres (see Feminist political philosophy). In many respects feminism and liberalism are natural allies, but when feminists argue for fundamental changes in the way men and women conduct their personal relationships, or advocate affirmative action policies for employment that seems to contravene firmly-entrenched liberal principles of desert and merit, they pose major challenges to liberal political philosophy (see Desert and merit).

Third, there is a set of issues arising from what we might call the new politics of cultural identity. Many groups in contemporary societies now demand that political institutions should be altered to reflect and express their distinctive cultures; these include, on the one hand, nationalist groups asserting that political boundaries should be redrawn to give them a greater measure of self-determination, and on the other cultural minorities whose complaint is that public institutions fail to show equal respect for those attributes that distinguish them from the majority (for instance their language or religion) (see Nation and nationalism; Multiculturalism; Postcolonialism). These demands once again collide with long-established liberal beliefs that the state should be culturally neutral, that citizens should receive equal treatment under the law, and that rights belong to individuals, not groups (see Citizenship; Affirmative action; Discrimination). It remains to be seen whether liberalism is sufficiently flexible to incorporate such demands.

Finally, liberalism is challenged by the environmental movement, whose adherents claim that liberal political principles cannot successfully address urgent environmental concerns, and more fundamentally that the liberal image of the self-sufficient, self-directing individual is at odds with the ecological picture of humanity's subordinate place in the system of nature as a whole (see Green political philosophy; Environmental ethics). Liberalism, it is said, is too firmly wedded to the market economy and to consumption as the means of achieving personal well-being, to be able to embrace the radical policies needed to avoid environmental disaster.

None of these problems is capable of easy solution, and we can say with some confidence that political philosophy will continue to flourish even in a world in which the sharp ideological divisions of the mid-twentieth century no longer exist. We may also expect a renewal of non-Western traditions of political philosophy as free intellectual enquiry revives in those countries where for half a century or more it has been suppressed by the state. Political questions that have concerned philosophers for two millennia or more will be tackled using new languages and new techniques, while the ever-accelerating pace of technological and social change will generate new problems whose solution we can barely begin to anticipate.

See also: Alienation; Civil society; Consent; Corruption; Critical theory; Culture; Economics and ethics; Evolution and ethics; Historicism; Liberalism, Russian; Pareto principle; Population and ethics; Postmodernism and political philosophy; Violence; Work, philosophy of; Partiinost'; Anti-Semitism; Religion and political philosophy; Family, ethics and the; Paternalism; Race, theories of; Sovereignty; Cultural identity; Law, philosophy of

DAVID MILLER

References and further reading

Goodin, R. and Pettit, P. (1993) *A Companion to Contemporary Political Philosophy*, Oxford: Blackwell.(A comprehensive guide to current thinking in political philosophy, this book focuses especially on normative issues.)

Hampsher-Monk, I.W. (1992) *A History of Modern Political Thought, Major Political Thinkers from Hobbes to Marx*, Oxford: Blackwell.(Contextualized studies of the political theories of Hobbes, Locke, Hume, Rousseau, The Federalist, Burke, Bentham, Mill, Hegel and Marx.)

Kymlicka, W. (1990) *Contemporary Political Philosophy*, Oxford: Oxford University Press.(Each chapter covers a major school of contemporary political thought.)

Wolff, J. (1996) *An Introduction to Political Philosophy*, Oxford: Oxford University Press.(An accessible general introduction which discusses the arguments of both historical and contemporary political philosophers.)

Political philosophy in classical Islam

Political philosophy in Islam is the application of Greek political theorizing upon an understanding of Muhammad's revelation as legislative in intent. In lieu of Aristotle's Politics, unknown in medieval Islam, Plato's political philosophy assumed the primary role in an explanation of the nature and purpose of the Islamic state. Al-Farabi conceived of the prophet as a latter day philosopher-king, Ibn Bajja and Ibn Tufayl took their cue from Socrates' fate and cautioned the philosopher against the possibility of successfully engaging in a philosophical mission to the vulgar masses, and Ibn Rushd presented philosophy as a duty enjoined by the law upon those able to philosophize.

1 Background

Two principal facts have formed political philosophy in Islam: first, the revelation of Muhammad, and second, the absence of Aristotle's *Politics*, whether by intention or historical circumstance, from the canon of texts translated from Greek and Syriac in ninth-century Baghdad (see Greek philosophy: impact on Islamic philosophy §2). Muhammad's revelation is of course foundational for Islam itself. From the vantage point of political theorizing, however, the revelation accounts for the perceived divinity of the state (the caliphate) which Muhammad is taken to have founded. It further accounts for the state's legal basis, with the state established on and grounded in obedience to a set of divine injunctions and sanctions. Muhammad's revelation, then, was perceived by political philosophers in Islam as providing an opportunity to understand and clarify the nature of the perfect state here and now, not in some distant future when human nature might be transformed. For Islamic political philosophers, the divine law (*shari'a*) revealed to Muhammad was a necessary and sufficient condition for bringing about human felicity.

The second fact noted above, the absence of Aristotle's *Politics* from the Arabic philosophical corpus, is of capital importance in understanding Islamic political philosophy. As late as the sixth century AH (twelfth century AD), this work was apparently unavailable (see Rosenthal's introduction to *Averroes' Commentary on Plato's Republic* (1969: 22)). The absence of the *Politics* meant that Plato and Platonic political philosophy, as found in the *Republic* and the *Laws*, became paradigmatic (see Platonism in Islamic philosophy). The implications of this are enormous, for Aristotle's critique of Platonic political philosophy from the standpoint of his own non-idealist philosophical anthropology was not to become part of the tradition of political philosophy in Islam. Instead, what we find in the Islamic political philosophers are variations on standard Platonic themes, pre-eminently the notion of the prophet as an analogue to the Platonic philosopher-king, the ambiguous role of philosophy in the practical political sphere, and the deep division between an elite and the vulgar masses (see Plato §14). These themes are especially prominent in the founder of political philosophy in Islam, al-Farabi, and it is to him that one must first turn.

2 al-Farabi

For al-Farabi (§4), 'the idea of the philosopher, supreme ruler, prince, legislator, and Imam is but a single idea. No matter which one of these words you take, if you proceed to look at what each of them signifies among the majority of those who speak our language, you will find that they all finally agree by signifying one and the same idea' (*Tahsil al-sa'ada*: 43-4). These remarks, not anomalous in the Farabian corpus, indicate the convergence, indeed identity, of theoretical and practical concerns. These dual concerns, reminiscent of the Platonic notion of the philosopher-king, are together present in the Farabian notion of the prophet. For al-Farabi, the Prophet (Muhammad) is to be understood as a divinely inspired legislator, offering a perfect way of life and a community in which to flourish.

The successors to the Prophet - caliphs and imams - preserve this original beneficence by so ruling that each group of believers is offered a measure of truth commensurate with its particular capacity to comprehend it. This latter stratification of the social and political framework is again reminiscent of Plato and his division of the perfect state into intellectually disparate groups. For al-Farabi, the relevant division is between those who are able to ground their belief philosophically and those who are not, these latter being the 'simple' believers. Religion is an imitation of philosophy in this scheme, the former presenting the truth in a non-theoretical, non-abstract way, in pictorial terms replete with parables and stories. Viewed thus, al-Farabi's political teaching may well be seen as interpreting

the historical state founded by Muhammad along the lines of Platonic utopian theorizing. However, such Platonically-inspired intellectual elitism is offered not as a blueprint for political reform in a distant future, but rather as a fair description of Muhammad’s constitution and the state he founded. In its own way, Farabian political philosophy is a defence of Islam.

3 Ibn Bajja and Ibn Tufayl

In turning to Ibn Bajja and Ibn Tufayl, one turns not only from East to West, from Baghdad to Andalusia, but one also notes a marked change of emphasis away from that aspect of al-Farabi’s political Platonism which identified the prophet with a philosopher-king who rules a state founded on and governed by a divine law. For Ibn Bajja and Ibn Tufayl, philosophy (theoretical insight) and practical politics do not mix. They are quite incommensurable. As Leaman puts it: ‘ibn Bajja’s problem is how the philosopher in the imperfect state should relate to society’ (Leaman 1980: 110). The emphasis here must be upon the notion of ‘the imperfect state’. Courtiers both, Ibn Bajja and Ibn Tufayl must have taken a hard look at the political scene around them and drawn the relevant (pessimistic) conclusions.

In his *Tadbir al-mutawahhid (The Governance of the Solitary)*, Ibn Bajja addresses himself to the *nawabit* (weeds) in imperfect societies. These *nawabit* are the nonconformists in the societies they inhabit. They do not share the common goals and aspirations. Simply put, they are the philosophers in an imperfect world, and for Ibn Bajja the focus is how best to secure their happiness and safety. Ibn Bajja’s ‘realism’ leads him to fasten upon that strand of the Platonic political philosophical tradition which is the underside of al-Farabi’s political Platonism, the non-utopian strand. The problem for the philosopher in the midst of an imperfect society is to achieve happiness while avoiding dirty hands, not befouling himself (and philosophy) with the hopes and desires of the masses. The *nawabit* dwell amongst people, but do not find perfection in their midst. Thus they must live in isolation, dissociating themselves from ‘those whose end is corporeal [and] those whose end is the spirituality that is adulterated with corporeality’ (*Tadbir al-mutawahhid*: 78). The background here is no less Platonic for its being non-utopian and apolitical. For Plato, Athens and her citizenry stand guilty by their execution of Socrates, the philosopher. The lesson to be drawn is clear for Ibn Bajja as he forcefully denies any easy commensurability between philosophy and politics, between theory and practice in imperfect societies. For Ibn Bajja, the *nawabit* exist in spite of the society they inhabit (as in *Republic* VI 496d-e), and thus owe it nothing.

Ibn Tufayl’s allegorical tale, *Hayy ibn Yaqzan (The Living Son of the Vigilant)*, is not at odds with the ‘realism’ of his predecessor Ibn Bajja. Indeed, the story may well be read as a vivid elaboration of the latter’s thought about the incommensurability of philosophy and politics in an imperfect world and the resulting isolation which the philosopher must seek if he is ever to achieve felicity. The whole thrust of Ibn Tufayl’s allegory is the lesson that Hayy, the protagonist, finally learns painfully, that the philosopher will be unable to communicate successfully the deepest truths (the illuminative mysteries) as discovered by himself to mankind at large. Hayy’s strenuous efforts to instruct, born of compassion for and inexperience of humanity, are shown to be useless as even the best among men ‘recoiled in horror from his ideas and closed their minds…. [And] the more he taught, the more repugnance they felt’ (*Hayy ibn Yaqzan*: 150). Worse, Hayy comes to realize that not only are his pedagogical efforts nugatory, but they are also counterproductive, for they tend to undercut the very beliefs which the simple (non-philosophical) believers hold. The tale ends with Hayy, dispirited, returning to the isolated island whence he came.

Plato’s pessimism about the possible transformation of the empirical realm is manifest in the works and teachings of Ibn Bajja and Ibn Tufayl. The philosopher-king, the prophet, has no greater a foothold among his contemporaries than did Abraham amongst the idolaters or Socrates amongst the Athenians. The mass of mankind cannot become philosophers, and to attempt such a transformation is both dangerous for the philosopher and bad public policy. Both Ibn Bajja and Ibn Tufayl may be understood as cautioning would-be prophets (philosophers) of the perils inherent in their mission. Correlatively, they can be read as urging a rather conservative (traditional) status quo political agenda, in which stability and order prevail and are grounded in a punctilious observance of the law (*Hayy ibn Yaqzan*: 153-4). If society cannot be remade from the ground up, then perhaps no innovation whatever should be tolerated, lest anarchy and antinomianism prevail.

4 Ibn Rushd

Ibn Rushd (Averroes) wrote commentaries on both Plato's *Republic* and Aristotle's *Nicomachean Ethics*. For Ibn Rushd, the *Nicomachean Ethics* provides the theoretical substructure for the practical sciences, while the *Republic* (in lieu of Aristotle's *Politics*) provides the practical, political foundations for the attainment of human felicity, the goal of the practical sciences. So conceived, Plato's *Republic* provides a kind of blueprint for the best political order. In this framework, Ibn Rushd resuscitates the Farabian emphasis on the active role the philosopher should play in the political arena. Though he is aware of the precarious position of the philosopher, this activist focus should be understood in contrast to that of his Andalusian predecessors, Ibn Bajja and Ibn Tufayl, and their focus upon the necessity for the philosopher to maintain an apolitical stand.

Ibn Rushd thus shares with al-Farabi the view that philosophy has political implications. He underscores this dramatically by grounding the very study of philosophy in the law. The first question of his famous *Fasl al-maqal (Decisive Treatise on the Connection between Religion and Philosophy)* is: 'What is the attitude of the law to philosophy?' Ibn Rushd's answer is that the law commands the study of philosophy for those capable of it. There is thus a duty to philosophize. It should be noted, however, that the study of philosophy is no idle enterprise for Ibn Rushd, for given the legal injunction upon those capable of studying philosophy, a practical political obligation follows. The law revealed by the Prophet is a divine law, given to insure the well being of the entire community. If then the law enjoins philosophy, the philosopher is obliged to 'use' his wisdom for the benefit of all, in just the same way that the Farabian prophet (or Platonic philosopher-king) uses his wisdom, in a manner commensurate with the audience in question. Inasmuch as only the philosopher has an insight into the truth in a 'straight', undiluted way, without mediation of the senses, only he can interpret the law in an appropriate manner. Only the philosopher can allegorize when necessary, for only he knows the grounds upon which the allegory resides.

In his own way, Ibn Rushd provides an important defence of philosophy in a society governed by lawyers and judges like himself. Lest politics and political philosophy be handed over to those who (merely) apply the law, Ibn Rushd argues forcefully for a practical political philosophy which probes the foundations and guiding principles of the law. This, then, is arguably the greatest achievement of political philosophy in Islam, to conceive of a society grounded in obedience to a divine law as itself the manifestation of a coherent and theoretically defensible structure.

See also: al-Farabi; Greek philosophy: impact on Islamic philosophy; Ibn Bajja; Ibn Rushd; Ibn Tufayl; Law, Islam philosophy of; Platonism in Islamic philosophy; Political philosophy, history of; Political philosophy, nature of

DANIEL H. FRANK

References and further reading

Butterworth, C.E. (ed.) (1992) *The Political Aspects of Islamic Philosophy*, Cambridge, MA: Harvard University Press.(An up-to-date collection of essays on, among others, al-Kindi, al-Razi, al-Farabi, Ibn Sina, Ibn Bajja, Ibn Tufayl and Ibn Rushd.)

Daiber, H. (1996) 'Political Philosophy', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 50, 841-85.(Account of some of the main political thinkers from the classical period.)

al-Farabi (*c*. 870-950) *Tahsil al-sa'ada (The Attainment of Happiness)*, trans. M. Mahdi, *Al-Farabi's Philosophy of Plato and Aristotle*, Ithaca, NY: Cornell University Press, 1962.(Translation, with introduction and notes, of *Tahsil al-sa'ada* and the précis of the philosophy of Plato and Aristotle.)

Frank, D. (1996) 'Ethics', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 55, 959-68.(The ethics of the classical period, with particular reference to their links with society.)

Galston, M. (1990) *Politics and Excellence: The Political Philosophy of Al-Farabi*, Princeton, NJ: Princeton University Press.(Al-Farabi's political philosophy against the backdrop of the Greek philosophical tradition.)

Ibn Rushd (*c*.1174) Commentary on Plato's *Republic*, ed. and trans. E.I.J. Rosenthal, *Averroes' Commentary on Plato's Republic*, Cambridge: Cambridge University Press, 1969; trans. R. Lerner, *Averroes on Plato's Republic*, Ithaca, NY: Cornell University Press, 1974.(Rosenthal is an edition of extant Hebrew text, with introduction, translation and notes. Lerner is a translation with introduction and notes, in part critical of Rosenthal.)

Ibn Tufayl (before 1185) *Hayy ibn Yaqzan (The Living Son of the Vigilant)*, trans. L.E. Goodman, *Ibn Tufayl's Hayy ibn Yaqzan*, Los Angeles, CA: Gee Tee Bee, 1983. (A felicitous translation of the tale, with introduction

and notes.)
Leaman, O. (1980) ‘Ibn Bajja on Society and Philosophy’, *Der Islam* 57 (1): 109-19.(Ibn Bajja’s political philosophy.)
Leaman, O. (1988) *Averroes and His Philosophy*, 2nd edn, Richmond: Curzon.(A recent study of Ibn Rushd’s metaphysics, practical philosophy, and theory of language and truth.)
Lerner, R. and Mahdi, M. (eds) (1963) *Medieval Political Philosophy: A Sourcebook*, Ithaca, NY: Cornell University Press.(Important collection of primary texts in translation, including al-Farabi’s *al-Siyasa al-madaniyya (The Political Regime)* and *Tahsil al-sa‘ada*, Ibn Bajja’s *Tadbir al-mutawahhid*, Ibn Tufayl’s *Hayy ibn Yaqzan* and Ibn Rushd’s *Fasl al-maqal*.)
Mahdi, M. (1991) ‘Philosophy and Political Thought: Reflections and Comparisons’, *Arabic Sciences and Philosophy* 1: 9-29.(A fine overview of the nature of political philosophy in Islam.)
Rosenthal, E.I.J. (1962) *Political Thought in Medieval Islam*, Cambridge: Cambridge University Press.(The most comprehensive account in English. Part II is entitled ‘The Platonic Legacy’.)

Political philosophy, history of

The history of political philosophy attempts to yield a connected account of past speculation on the character of human association at its most inclusive level. 'History' or 'philosophy' may be stressed depending on whether the organizing principle is the temporal sequence or conceptual framework of political thought. Anglophone work has increasingly been organized around distinctive political 'languages' defined by specific vocabularies, syntaxes and problems, for example, classical republicanism, Roman law, natural law, utilitarianism. Chronologically it has been usual to observe divisions between ancient, medieval, Renaissance, early modern and modern periods of study.

Ancient Greece is the source of the earliest political reflection, with a continuous history in the West. Here reflection on the nature and proper organization of political community stimulated inquiry into the difference between nature and convention, the public and the domestic realm, the distinctive character of political rule, the relationship between political life and philosophy, the identity of justice, and the taxonomy of state-forms - as well as a more sociological investigation of the stability and decline of political regimes.

Greek political vocabulary was adapted to existing Roman republican practice (by Polybius and Cicero for example), which soon gave way to an imperial constitution stressing peace, order and unity. Rome thus generated two contrasting political ideals - that of the virtuous active republican citizen, and that of the unified empire governed by Roman law. Together with questions about the causes of its own rise and decline, Rome thus provided political values and historical material for subsequent philosophical and historical reflection.

Christianity undermined the pagan autonomy of politics in the name of a higher, transcendent ideal. However, it adapted much of Greek rationalism and the political vocabulary of classical culture in elaborating a creed and an institutional form. In turn it lent legitimacy to imperial and royal officeholders of Rome and barbarian successor-kingdoms.

Medieval political philosophy was characteristically preoccupied with the relationship between pope and king, church and regnum, *but philosophy as a discipline was subordinated to theology. This was challenged by the rediscovery of Aristotle's self-sufficiently secular political ideal, a challenge met for a while by Aquinas' synthesis. However, the autonomy of secular politics was continually reasserted by a sequence of writers - Bartolus of Sassoferrato, Marsilius of Padua, Bruni and Machiavelli - who revived and reformulated classical republicanism using both Roman law and new Renaissance techniques and insights.*

The Reformation, although initially politically quiescent, gave rise to new conflicts between secular and sacred rule. In particular, radical claims about the responsibility of all believers for their own salvation fed through in various ways into more individualistic political philosophies. In early modern Europe, using the strikingly new (and originally Catholic) vocabulary of natural right, Hugo Grotius aspired to provide a common secular basis for a shared political morality, on the basis of individual rights derived from a universal right of self-preservation. This was widely explored by seventeenth and eighteenth century thinkers, notably Hobbes and Locke, and culminated politically in the American and French Revolutions. In the aftermath of the French Revolution, language of natural rights was rejected both by conservative thinkers, such as Burke, and by a new, utilitarian radicalism largely forged by Bentham.

Attempts to grasp the political character of economic transformations and Empire in early modern Europe resulted in a growing engagement with the essentially historical character of politics, the dynamic of which republican discourse was particularly well suited to exploring. Avoiding the loss of liberty which the acquisition of Empire had seemed to entail in Rome involved rethinking possible patterns of politico-economic development, providing a new definition of liberty which stressed personal and economic over political freedom, and proposing that impersonal institutional devices could replace virtuous motives in guaranteeing political liberty and stability. Such possibilities were explored by Montesquieu and Constant in France, Hume and Smith in Britain and 'Publius' (Madison, Hamilton and Jay) in America. They were rejected outright by Rousseau, for whom only the active citizen could guarantee rights, civic or civil.

The French Revolution was not only an event in which political philosophy played an important if hotly contested

role; it also, like the rise and fall of Rome, provided a central topic for subsequent political reflection. The character of modernity, the nature of revolution, the relationship of political ideas to political action, the strength or weakness of rationalism as an informing principle, the viability and desirability of the Revolutionary ideals of liberty, equality and fraternity, all became topics of philosophical speculation by post-Revolutionary thinkers such as Constant, Cabet, de Tocqueville, Burke, de Maistre, Saint-Simon, Owen and Coleridge, as well as a later generation including Comte, Carlyle and Marx.

In contrast to his predecessors' use of Lockean psychology and the conditioning effects of experience and association to understand the processes of socio-economic change, Kant's postulation of the transcendent self initiated a new vocabulary of idealism. This culminated in Hegel's attempt to show how philosophical and historical (including political) change could be understood as the development and realization of a trans-historical consciousness or Geist, *seeking to overcome internal tensions through a process of projection and transcendence.*

The notion that human self-understanding and practices are to be understood historically immensely influenced subsequent political thinking, being central to the ideas of Marx, Nietzsche and Freud (as well as shaping many of J.S. Mill's modifications of classical utilitarianism). All three of the former owed insights to Hegel's claims about the crucial and emblematic character of the master-slave struggle. However, while for Hegel and Marx the slave's insights represent the transition to a higher form of consciousness - mediated in Marx's case by a class revolution - for Nietzsche (despairingly) and Freud (resignedly) repression was a constitutive and self-perpetuating feature of modern politics.

While nineteenth century political thought was preoccupied with the historical conditioning of political sensibilities, Freud's discovery of the unconscious was accompanied by the emergence of a mass, irrationalist politics, characteristic of the twentieth century, and more suited to sociological than philosophical analysis. Nevertheless rationalist political theory, deriving from utilitarianism, and frequently drawing on (and contributing to) economic thought, remains the dominant accent in contemporary political philosophy.

1 Nature of the subject

At its broadest the history of political philosophy encompasses all recoverable philosophical reflection on the nature of politics. However, if more than a merely chronological presentation, it must give some account of the relationships obtaining over time between political philosophies. Such an account may emphasize the philosophical or the historical character of those relationships. Two extreme positions may be identified. At one extreme the history of political philosophy is conceived of as a series of attempts to articulate essentially timeless and universal truths about the character of human political association; the relationship between past political philosophers is understood as a shared and highly abstract trans-historical conversation. The other extreme stresses the constitutive character of contingent historico-political circumstance on the formulation of political philosophy, and hence upon its history. For these scholars the relationship between past political philosophers will be much mediated by their contemporaries' and predecessors' less philosophical works - sermons, tracts, polemic and the more symbolic forms expressed in ceremonial and institutional frameworks.

During the last thirty years the history of political philosophy, certainly in the Anglophone world, has been revolutionized by the second - historical - approach. This has affected content as well as methodology and has parallelled more general doubts about the possibility of characterizing philosophy as an activity abstracted from other social activities. The revolution involved rejecting the notions that 'ideas' could be identified as appropriate units of diachronic historical investigation (Skinner 1988), and that philosophically abstract political writings could, in themselves, comprise a history (Pocock 1971). More constructively, the revolution claimed that works of political philosophy were best - indeed perhaps only - to be fully understood, either individually or as part of a history, by relating them to their immediate linguistic and controversial context. The history of political philosophy was now to be understood as an account of linguistic and controversial innovation undertaken within relatively stable and discrete political languages - those of republicanism, natural jurisprudence or utility, for example.

While the historians of political philosophy address essentially historical questions, political philosophers have increasingly also used readings of the history of political philosophy to frame philosophical arguments. These have

particularly focused on the identity and supposed political and philosophical legacy of the Enlightenment (see Foucault, M.; Habermas, J.; MacIntyre, A.; Historicism).

2 Ancient: Greek

The history of political philosophy normally begins with the Greeks. Although philosophical reflection on politics was conducted in many early cultures, for example in China and ancient Mesopotamia, and although in some cases the texts produced (such as the Jewish Old Testament) subsequently influenced Western political thought, a continuous tradition of treating political questions at a high level of philosophical abstraction began in Greece of the fifth century BC.

Presocratics and Sophists: physis *and* nomos. The progress of Greek science, in seeking to provide agentless explanations of natural phenomena, created a bifurcated language of impersonal, natural laws contrasted with the conventional laws or rules governing human society (see Physis and nomos). In the rapidly changing, commercial, and increasingly egalitarian, society of fifth-century Athens this linguistic polarity was soon deployed in philosophical disagreements over the proper character of human political institutions. At least some of the Sophists - a heterogeneous group of professional teachers - taught that human laws were mere conventions and that nature sanctioned only prudential limits on self-interest (see Sophists). For Ancient Greeks virtue was an instrumental quality or skill that brought success - and success was a sufficient sign of virtue. In a democratic city-state, this was the art of persuasion, political or forensic oratory. But could mere success, regardless of means or content, count as virtue?

Socrates and Plato. Against these naturalistic, and in some cases morally opportunistic, arguments emerged the figure of Socrates. Although himself seen as a Sophist by moral conservatives and charged at his trial with ‘making the weaker argument appear the stronger’, Socrates, while a rationalist critic of conventional moral pieties (see *Euthyphro* and book I of the *Republic*), aspired to a transcendent political morality, and a metaphysical identity of the self which could be harmed only by moral turpitude. His acceptance of the death sentence, while believing it to have been wrongfully imposed, and his refusal to accept an offer of escape from friends, demonstrated his strong sense of political obligation, stemming from a contractual account of the relationship between state and citizen (see Socrates).

This is the intellectual setting of the *Republic* in which Socrates’ pupil - Plato, using his master’s persona to develop his own ideas - invokes the transcendent world of the Forms to rebut Thrasymachus’ claim that justice is merely whatever serves the self-interest of ‘the strongest’ (see Thrasymachus). While the particular institutional recommendations of the *Republic*, - equality of women, abolition of the family, communism - have clearly been marginal issues in Western political philosophy until relatively recently, the work identified what would become major and recurrent themes: the bearing of epistemology on political reality, cognitive and moral pre-eminence as a claim to privileged rule, the relationship between social class structure and individual psychology, the political role of education and the political dimensions of the processes of social change. Plato’s appeal to a rational, if veiled, truth, and his conception of the bad or imperfect as the corruption or privation of the true or good, proved congenial to the construction of a Christian philosophy.

Aristotle. Aristotle’s championship of man’s political nature also locates him in the Greek debate about the rival claims of nature and convention (see Aristotle). For Aristotle, nature, while not transcendent, was not simply empirically given, but involved a *telos*, an end or purpose which all natural entities possessed. For freemen, this included the achievement of civic life in a *polis* (a Greek city-state). Cultures - such as that of the Persians - which did not support such communities were unnatural and their inhabitants inferior. Nature was not only everywhere goal-directed, but internally complex and composed of diverse relations of rule and subordination. The *polis* for Aristotle was a complex entity comprising various subordinate forms of association (notably the family and the village) and containing a corresponding diversity of forms of rule. That form of association and rule which he identified as distinctively political obtained peculiarly and uniquely amongst equals (unlike most moderns Aristotle identified equality as a condition, not a presupposition, of political status). Apart from philosophical speculation itself, living in the *polis* allowed the citizen the fullest exercise and highest development of their faculties.

The categories Aristotle identified in his *Politics* provided an enduring vocabulary for political analysis down at

least to the emergence of modern political science in the nineteenth century. He devoted sustained attention to the concept of citizenship and produced a sixfold classification of constitutions - monarchy, tyranny, aristocracy, oligarchy, democracy, ochlocracy - according to whether rule was by the one, the few rich or the many poor, and whether it was conducted in the interests of the rulers themselves or the whole people. He also explored the social conditions which might render these types stable and the characteristics of mixed regimes, preferring - as most practicable - a hybrid aristocracy-democracy, or ‘polity’. Although Aristotle’s political writings were lost to Europe until their recovery in the thirteenth century, his work - together with Greek and Roman practice - was the source of a major European political tradition, republicanism.

Plato and Aristotle, like most Greeks of the classical period, saw the polis as the essential environment for the life of the fulfilled individual, and saw no unavoidable conflict between its demands and the proper aspirations of the individual.

Hellenistic schools. The aspiration to unite the languages of nature and citizenship was severely stressed by the breakdown of the city-state and the emergence of Macedonian and subsequently Roman imperialism. In the absence of a political structure in which to locate them, political categories in Hellenistic culture either disappeared from the analysis of man’s nature - which stressed either the individual’s natural sensations (see Epicureanism) or the withdrawal from social obligation (see Cynics) - or else they were projected onto a larger canvas, the cosmos itself, in which a shared rationality was claimed to generate universal citizenship and universal subjection to ‘natural law’ (see Natural law; Stoicism).

3 Ancient: Roman

Although the Greek observer Polybius applied Aristotelian categories to Roman history and gave an influential account of the dynamics of constitutional change, Roman republican political practice long predated theoretical reflection about it.

It was only with the imminent collapse of the republic that a theoretical synthesis of indigenous practice with Greek political language, influenced by Stoicism, was provided by Cicero, in his *On the Republic* and *Laws*. The text of the former was lost in late antiquity, but a number of important passages survived in quotation, including the definition of justice (‘that which gives to every man his due’) and of a commonwealth (‘the people’s affair; and the people is not every group of men, associated in any manner, but is the coming together of a considerable number of men who are united by a common agreement about law and rights and by the desire to participate in mutual advantages’) (Cicero, *On the Republic*: 129). In the *Laws*, he gave clear expression to the idea of a universal natural law transcending the positive laws of particular states. Cicero’s *Duties* became a hugely influential handbook of practical political morality while his work on rhetoric included a suggestive account of the passage of humans from a condition of nature into one of political society (see Cicero).

The establishment of a post-republican order under Augustus heralded a literary glorification of republican history and austere personal morality (see the work of Livy and Virgil), accompanied in practice by a consolidation of the various republican powers into the office of the princeps. The slide into imperial absolutism evoked Plutarch’s use of biography as a form of political reflection (*Lives of the Emperors*) and Tacitus’ influential essay on the oratorical political implications of the loss of public debate (*Dialogue on Oratory*) as well as a eulogy of the simple, virtuous and free Germans living outside the Empire (*Germania*).

Roman republicanism and law. Rome bequeathed to the West two potent political vocabularies and evaluative associations: one - going back to Aristotle - based on the active life of austere republican virtue and participatory citizenship (at least for the economically independent); and another - quintessentially Roman - based on a unified and authoritarian imperial system, essentially juridical and administrative. The judgment of the former on the latter was negative in the extreme; empire destroyed *virtù*, even civic and political life itself. On the other hand, the peace and prosperity normally prevailing under the unified Empire passed on another set of ideals. Roman law provided a vital political and jurisprudential resource for succeeding societies and political thinkers as well as comprising a mine of historical evidence. The sources were the compilations of the sixth century Emperor Justinian (collectively known as the *Corpus Iuris Civilis*). While these compilations had little directly to say about politics, and were neither systematically integrated nor metaphysically grounded, they were a rich source of examples and maxims, applied politically down to the late eighteenth century (see Roman law).

The incorporation of Christianity into the tradition of political theory both joined Judaic tradition with that of the Graeco-Romans and created a new language with a new agenda, and an authoritative, if hardly unequivocal, text (the Bible). The early Christians, expecting Christ's return, expressed an active contempt for earthly matters. Later recognition that this return was not imminent forced Christian thinkers to address the question of the Church's relationship to secular authorities - a question that became more complicated when the Roman Empire became Christianized.

Although the language of apostolic Christianity was thus originally separated from politics, it was, from early on, suffused with Greek philosophical influences. With the conversion of Constantine (AD 312) and the granting of toleration, however, and even more with Theodosius' adoption of Christianity as the public religion of the Empire (AD 393), Christian religious thought, Greek philosophy and Roman political thought and institutions were drawn together. Despite attempts by, for example, Tertullian, to reject Greek canons of rationality ('I believe because it is absurd'), and by Donatus and others to create a purified Church membership in, but independent of, the secular realm, the ideal that emerged was of the Church as a secular catholic body, including sinners and saints, committed to articulating a single and philosophically coherent orthodox creed (Nicaea), and establishing a *modus vivendi* with Christian secular powers. More threatening to Christian autonomy was the widespread tendency, particularly identified with Eusebius and Prudentius, and shared by Augustine in his early writings, to endow the Roman adoption of Christianity, and therefore Roman political institutions themselves, with providential and even eschatolological significance - to blend the universal Church and the universal empire, as images of the heavenly society, and even to see it as an intimation of the second coming of Christ.

The sack of Rome by its Gothic mercenaries under Alaric (AD 410) seemed to impugn not only this implicit Christian philosophy of history, but the power of the Christian God and the historical status of Christianity itself. Augustine's *The City of God* was composed to repudiate such implications, to distinguish both the Christian Church from Roman politics, and the idealized Heavenly City of the elect from the secular Church on Earth. As the title suggests, Augustine self-consciously drew on Graeco-Roman political language, systematically subverting its secular meaning and projecting it on to a Christian, transcendent realm. This rhetorical strategy had already been used by Augustine's teacher Ambrose, who gave a Christian meaning to the Ciceronian dictum that the foundation of justice was good faith, and applied Cicero's image of the body politic to the body of Christ as his church. Yet while Ambrose tended to continue to identify Christian emperors with Christian purposes, Augustine's most celebrated move was to repeat Cicero's definition of justice - 'which gives to every man what is his due' - and to argue that since the secular realm distracts man from union with his God, the secular city, even when Christian, can never be truly just.

Augustine applied to politics a Christianized Platonism already explored by Plotinus in order to avoid the dualist heresy of Manicheism (of which he had at one time been a devotee). Christian Platonism could account for evil, not (heretically) as a countervailing cosmic force, but simply as a privation of good, brought about by humans' own wrongful ordering of priorities. Since the Fall humans were congenitally incapable of correct evaluation of things without the intervention of divine grace. Political institutions and the peace they could bring were of value, but only to deal with outward and secular conflict; they were incapable of reaching men's souls. Man was sociable by nature and political through sin.

In 494 AD an authoritative statement of the relationship between Church and polity was provided by Pope Gelasius I, who articulated the doctrine of the two swords, spiritual (*auctoritas sacrata pontificum*) and secular (*regalis potestas*). Emperors were to submit to the Church's teachings on religious issues, clerics were to obey the laws. The emperor was to wield the secular sword at the command of the Church. This went beyond the Pauline injunction to leave to Caesar that which was Caesar's, and assumed a deeper identity between 'Christian' and 'Roman' and a division of labour - spiritual and political - amongst Roman Christians in pursuit of a shared end.

4 Medieval

The supervention of the Germanic kingdoms on an already fragmented Roman Empire and the subsequent centuries of only sporadic order did not terminate Roman political discourse or the periodic aspiration to reinstitute the universal Christianized Roman Empire. It did result, however, in the temporary loss of important works which might have sustained it (Aristotle's *Politics*; Cicero's *On the Republic*; Justinian's *Corpus*), and it did largely end systematic political philosophy. For many centuries the known 'texts' of political philosophy have been largely the

practices, oaths, institutions and proclamations of political actors, and as we have come to recognize linguistic acts we have, conversely, also recognized the textual qualities of acts.

While kingship was a characteristically Germanic institution, marginalized - at least in philosophical treatises - by the republican traditions of Greece and (republican) Rome, it became enmeshed in the language and concepts of Roman law, Christianity and universal empire, almost as soon as it appeared. The persistence of a diversity of rulers within Christendom, and the existence of a Church claiming universal spiritual jurisdiction, provided the parameters within which this language had to be redefined. Despite the subsequent myths of elective councils and of constitutional and popular kingship, early Germanic kings, especially those who seemed capable of restoring a Holy Roman Empire, often continued to claim, or had bestowed upon them, Roman imperial authority. Sometimes the language of imperial kingship itself was ascribed to the Church: Isidore of Seville saw the existence of a universal church as the body of Christ the King as nevertheless quite compatible with the existence of a diversity of secular kingdoms. The rulers of such kingdoms were, however, all to be thought of as providentially appointed, and were obliged to pursue Christian ends; where they did not, they were seen as just instruments of God’s wrath.

Philosophy, in medieval education, was in the service of and subordinated to theology. It attempted to establish coherence by synthesizing the deliverances of eminent authorities on disputed questions. The most striking achievement of this kind lay in the work of Thomas Aquinas. He synthesized not only much of the elaborate and diverse Christian tradition but also sought to incorporate into it the potentially threatening naturalistic political philosophy of Aristotle, newly recovered from the world of Arabic scholarship and translated by William of Moerbeke. Aquinas’ agreement with Aristotle’s claim that man had a natural end or *telos* in political or social life - however qualified by the further end in God to which only the Church could minister - marked a different and much more congenial view of the relationship between politics and Christian belief, and between man’s natural and religious ends, than Augustine’s. For Aquinas, ‘grace does not abolish nature but perfects it’. Aquinas offered an influential account of a nesting hierarchy of laws governing the world - eternal, divine, natural and human - and of the kind of reason governing each and their method of promulgation. Such a hierarchy held the potential, present in all juristic thought, for a critical appraisal of positive law, yet Aquinas, while expressing a prudentially grounded preference for a limited monarchy, tended, like many medieval thinkers, to look to local positive law in defining legitimate political institutions.

Within the language - and the practice - of law, a major tension existed between the Roman law doctrine locating the law in the will of the prince and both natural-jurisprudential and Germanic customary doctrine which implied the duty of the prince to legislate according to pre-existing universal, natural or customary standards, a tension which influenced much of the language of sovereignty and constitutionalism (see Roman law; Common law). This was expressed in two contrasting medieval beliefs about the source of political authority, which (to simplify) might be seen as ascending from the people or as descending from God (Ullman 1966). The fourteenth-century Jurist Bartolus of Sassoferrato and his pupil Baldus elaborated a justification of the popular source of sovereignty in city-states. Natural reason, whether through custom or the active consent of citizens, was, according to them, adequate to create an autonomous and self-sufficient commonwealth. Each thinker applied the Roman law terminology of corporations to both city-states and kingdoms. Political entities could thus be conceived of as *personae fictae*, with a continuous identity, distinct from the natural persons comprising them or their ruler. An even more strikingly methodologically individualist - even anarchistic - position was advanced by the nominalist William of Ockham.

Another language, deriving not from Roman law but from Germanic custom - the language of community - provided a variant of the view that power ascended. Medieval communities of various kinds were customary collectivities without necessarily being legally *personae*. Thus there was another language through which the idea of a collective unity could be articulated. Here, and in certain feudal structures, we find the origins of the idea of political representation.

The Church was also a collective entity, on one description a *corpus mysticum* identified with the body of Christ. Yet just as the kingdom had a transcendent identity as well as a physical one, so the Church needed a legal and political identity as well as a religious one. Conciliarist thinkers, notably Nicholas of Cusa, developed a view of the Church as superior to the pope and capable of representation through its council. Some, such as Pierre d’Ailly and Jean Gerson, drew a parallel between the secular model of the mixed constitution - king, aristocracy and

demos or people - and the structure of the Church, comprising pope, cardinals and council. (See Black 1970.)

5 Renaissance

From the end of the eleventh century, the emergence and growth of self-governing cities in northern Italy, their internal political relations, and their antagonistic relationship with the Holy Roman Emperor (who claimed allegiance from many of them), created two important needs: first, for the production of impressive and persuasive diplomatic and other kinds of ceremonial, public and commercial documents; second, for a language capable of articulating resistance to imperial authority. The first was met by the development of a sophisticated class of scribes, the *dictamini*, who applied their knowledge of classical rhetoric and literature increasingly to the development of distinctive civic genres - model orations, city chronicles and political advice books. The second was more problematic. Both Catholicism and Roman law, drawing on Rome's imperial rather than republican legacy, emphasized the importance of political unity: the lowest regions of Dante's *Inferno* were reserved for those who had destroyed the sources of either religious or political unity - Judas Iscariot, Brutus and Cassius. It was one of the ideological triumphs of the civic republican Renaissance to convert Brutus into a hero.

Both Marsilius of Padua and Bartolus of Sassoferrato used scholastic and juristic resources to elaborate *de facto* legitimations of the autonomy of city-states from both imperial and papal authority. Bartolus spelled out the claim that free cities were 'princes to themselves', and that this meant that they were free both to assign jurisdictions and to establish forms of internal government without higher approval (Skinner 1978). Marsilius' strikingly modernistic (and heretical) view not only presented the 'whole body of the people', legislating through their representatives, as the source of law, but subordinated ecclesiastical authority to that law. In this notion of disparate, individually omnicompetent political entities lay the origin of the conception of a world of states which supervened on the aspirant universality of Christendom. Marsilius and Bartolus both derived their views from a republican argument, although the autonomy of states was also consistent with a Roman law defence of the absolute ruler, now thought of as *imperator in regno suo*.

However, there was a language available to, and increasingly being recovered by, both the *dictatores* and the humanist scholars of classical literature which praised, indeed glorified, the active political life in the free polity. Although initially the *dictatores* focused purely on classical rhetorical technique and style, and humanists were uninterested in or even repelled by the political life of the Ancients, by the fourteenth century both were employing ancient models to revive classical republicanism. This celebrated the liberty secured by active citizenship and explored *topoi* - founding, location, laws, institutions, political culture and social order - which affected the freedom and survival of the republic, thereby reviving interest in the secular causal processes at work in historico-political change.

The culmination of this development was the writing of Niccolò Machiavelli. Appropriating in particularly innovative ways the characteristic genres of the new civil science - the advice book (*The Prince*, *The Art of War*), the classical textual commentary (*Discourses on…Livy*), the civic history (*History of Florence*) - Machiavelli repudiated not only Christian homiletics but many Ciceronian values (see Machiavelli, N.).

The major polarity in Machiavelli's analysis was between *virtù* - the energy and capacity to act effectively in the political realm - and *fortuna* - the personification of those unpredictable secular forces which control human affairs. Political action comprised the imposition of order (form) on an inchoate political nature (matter). In so presenting politics, Machiavelli scandalously presented man performing a divine role. In *The Prince* Machiavelli pays particular attention to the extreme situation of a ruler seeking power where there is either no pre-existing order, or one inimical to his rule, thus highlighting the role of and qualities comprising *virtù* in the absence of assistance from custom, divine grace or nature. Such extreme but politically paradigmatic circumstances required, he claimed, absolute political autonomy and an opportunistic subordination of moral norms, whether Christian or classical, to the requirements of political survival.

In the *Discourses* Machiavelli turned his attention to the problems of republican regimes. His focus on Roman history provoked a deeper analysis of the secular processes, the interplay between historical circumstance, laws, military experience and social structure which shaped the characters of the citizens and the survival of the state. This analysis was to prove immensely influential for later republicanism, both as ideology and as an emergent secular historico-sociological analysis.

6 The Reformation

While Machiavelli considered religion only as a factor in political sociology, the Protestant Reformation reinvigorated both religious ideals and religious jurisdiction as a source of political argument. Underlying Luther's rejection of indulgences, outward forms and the jurisdictional authority claimed by the Church, lay a more fundamental theological question - that of man's free will in gaining salvation (see Luther, M.). This was made explicit in his exchange with the Christian humanist Erasmus. Luther argued that humankind were saved through faith (not works) bestowed by God (not individual human will, however trained or disciplined). This theology not only undermined - as intended - the sacramental and clerical efficacy of the Church, it changed its identity from a catholic and secular administrative structure into a *congregatio fidelis*, with purely admonitory power. At the same time Luther distinguished sharply between man's spiritual and worldly existence. For secular purposes man possessed free will, and political obedience was required, not only to curb sin and for the purposes of natural peace, but because even tyrants represent God's will in the world and may not be actively resisted. German peasants in revolt blasphemed by 'making Christian freedom a physical thing' (Oakley 1991: 173). It was the mistake of the Anabaptists' 'radical reformation' to erode Luther's distinction between the spirit and the world, which led them to believe they could, by acts of political will, provoke moral and religious reformations.

Perhaps the most difficult problem for Reformation thinkers to articulate remained the question of resistance where the magistrate opposed the reformation required by God. Luther himself after 1530 acknowledged that where the constituted authority itself allowed an active right of resistance (for example, in canon and civil law), it was not sinful to resist if magistrates - even the emperor - acted outside the law. Two distinct arguments emerged. One was constitutional, and sought to legitimate resistance by identifying officeholders other than the emperor (or, later, the French or Spanish king) as 'powers that be [which] are ordained of God' who might exercise a right - indeed a duty - of resistance consistent with St Paul's authoritative injunction to obey established power (Hotman, Beza, Mornay). This could be done either by federalizing relations within the Empire, or dignifying what Calvinists called 'inferior magistrates' - variously identified with provincial or civic officials and buttressed by classical allusions to Spartan ephors and Roman tribunes. The other more radical argument licensed the resistance even of private individuals. According to it the failure of rulers to operate within the laws defining their office effectively reduced them to the status of private individuals against whose actions other individuals and collectivities possessed - under canon, civil, and even natural law - a right of active defence.

Reformation discourse was also radicalized by the language of covenant theology. Calvin presented holy history as a series of covenants between God and his people. Reformation churches were formed through covenants among their members, who had a duty to God to maintain the terms on which the community was instituted. The theory thus provided a radical individualist and contractual basis for political authority, obedience and resistance, which was, however, utterly theological in origin. The initial separation between world and spirit was thus overcome in the political language of the Reformation, which had on the one hand incorporated a secular theory of resistance to government and, on the other, formulated an individualistic account of the source of political authority in voluntary agreement.

Calvinist constitutional resistance theory was further elaborated in France by the minority Huguenots, anxious to appeal to a wider constituency against the emergence of an absolutism increasingly inclined to impose Catholic orthodoxy (Salmon 1987; Kelly 1970) and above all in the Spanish Netherlands, where the successful Dutch Revolt occasioned a massive pamphlet literature elaborating a language of civic independence, the true significance of which is only just coming to be acknowledged (Van Gelderen 1993).

7 Scepticism and toleration

From these contexts, too, emerged a range of arguments in favour of religious toleration (see Toleration). Some of these appealed to the essential similarity of all religions and the unimportance of dogma. Others deployed sceptical arguments, claiming that it was wrong to coerce other people over matters in which rational certainty was unattainable. Meanwhile, Michael de L'Hopital and Jean Bodin forwarded *politique* arguments which held that, whether truth in religious questions was attainable or not, toleration should be practised as a condition of civil peace.

One major consequence of internal religious war, in France, the Low Countries and - as is increasingly being

recognized - Germany, as well as later in England, was a revival of ancient Stoicism (see Lipsius, J.; Montaigne, M.E. de; Stoicism). This fuelled both an inner perseverance and constancy in the face of adversity and political acceptance of *de facto* powers. A growing moral agnosticism led politically to a renewed, sceptically-based insistence on the need to accept the established faith and political order. In this context there appeared too a controversial new political term: 'reason of state' - the 'knowledge by which…stable rule over people may be founded, preserved and extended…when such actions cannot be considered in the light of normal rules' (Botero in Burke 1991: 480).

Both the term 'state' and a vocabulary in which its imperatives could be presented as autonomous and irresistible were now available. They were first assembled by Jean Bodin, who used legal terminology to define sovereignty rather than inquiring into the constitutional practice of particular regimes or assembling it from distinct powers as, even in imperial Roman law, it always had been. Eschewing Roman law as a historically parochial phenomenon, Bodin sought to understand sovereignty analytically as a necessary feature of the state - 'an absolute and perpetual power over citizens and subjects' (see Sovereignty). All office in the state is held at the sovereign's pleasure. Moreover, the sovereign not only enforces the law, but makes it (without consultation if it pleases), and stands above it, bound only by the objective conditions that determine its own survival - a rule of succession, conformity to natural law and the law of God.

From this basis Bodin proceeded to analyse the natural and historical circumstances affecting the successful establishment of states. In thus admitting natural and historical considerations only as factors affecting success or failure in instituting an abstractly defined conception of state, Bodin cut through both the competing languages of legal constitutionalism and naturalism, while not quite closing the door to interpretations of God's purposes.

Annexing a Bodinian conception of sovereignty to a strict interpretation of the divine ordination of political power expressed in God's grant to Adam yielded the socially reverberant language of patriarchalism associated with Sir Robert Filmer, Locke's famous protagonist. The historico-philosophical (as opposed to sociological) interest of Filmer lay perhaps less in his patriarchalism than in the objections he was able to pose to the radical contract theory emerging amongst the Levellers and other participants in the English Civil War - objections that Locke later sought with some ingenuity to overcome. Filmer's counterargument drew upon theories of long standing - amongst Catholics, as he gleefully pointed out in *Patriarcha* - which grounded political power in a voluntary contract of individuals and derived from this principles both of resistance and of private property.

8 Early modern natural rights theory

Several Catholic publicists of the Reformation period had argued for constitutional limitations on political power, which was conceived to emanate from, and in some ultimate sense remain in, the community. John Major followed Gerson in articulating an unlimited subjective conception of 'right' as free disposal, drawing the corollary that kings can have no such right over their people. However, it was the Spanish neo-Thomists who began to elaborate the most distinctive and far-reaching natural rights language to combat a range of opponents, including Lutherans, Machiavellians, conciliarists and Catholic apologists of Iberian imperialism in America.

The revival of Thomism was an attempt to counter the heretical (Lutheran) view that humans were incapable of knowing God's law, and that dominion was founded only on (inscrutable - and therefore unquestionable) grace. Dominicans and Jesuits reasserted Aquinas' hierarchy of law, with Natural law as both a specification of divine and eternal law and a practicable criterion of the legitimacy of human, positive laws. Stretching Aquinas, they also claimed natural law was imprinted in, and even a habit of, the mind, available to all humans independent of revelation (see Molina, L. de; Soto, D. de; Suárez, F.). This enabled them to establish the legitimacy of secular institutions (and indeed that of the Church, as a historical entity with a tradition of doctrinal interpretation) by deriving them directly under natural law. This same argument also refuted the claims of Machiavelli, curiously allied to those of Luther, concerning the irresistibility of reason of state.

Although human community was natural, and although natural law was available to men in what became known as the 'natural condition', in which men were free and independent, political society was a voluntary human creation. Although the concept of the 'social contract' was not yet theoretically elaborated, and although Suárez emphasized the pre-existence of community to avoid already notorious difficulties involved in constituting it from individuals through contract, the notion of consent or agreement to the establishment of political rule - as one condition of its

legitimacy - is clearly present (see Vitoria, F. de).

One important application of neo-Thomism had been the claim made on behalf of King Ferdinand of Spain that American Indians - being evidently incapable of political society - must lack natural reason and so be natural slaves who might justifiably be captured and deprived of their possessions. This position was attacked from the 1530s onwards by members of the school of Salamanca using similar neo-Thomist language. Vitoria and Las Casas in particular cited the, by then, undoubted empirical evidence of Aztec and Inca capacity for civilized life, demonstrating that Indians had indeed used their natural reason to construct political institutions. They could not therefore be slaves, and their property and authority was legitimated under natural law.

This vocabulary of a state of nature, natural law and the establishment of political authority by a contract alienating natural rights was transformed from an identifiably Catholic position into the foremost language of the modernizing and Protestant communities of Europe within the space of a generation, largely by the Dutchman Hugo Grotius. Writing initially in the context of international legal controversy, where neither sceptical acceptance of positive authority nor appeal to shared values was productive, Grotius performed a series of moves which subsequent generations regarded as constitutive of modern natural law theory. First, following the Jesuits, he located the source of natural law in the will of the agent. Second, following ancient and revived stoicism, he made self-preservation the primary natural right: self-defence and a right of acquisition followed as consequences of this. Third, he emphasized the potentially naturalistic grounds of such a theory, which would hold even if (scandalously contrary to his own professed belief) there were no God. Perhaps even more importantly, he adopted a systematic and deductive mode of argument which self-consciously modelled itself on mathematics.

Immediately picked up by Selden and Hobbes in England, by Samuel Pufendorf in Germany and at the Swedish court, and by John Locke, working in England and The Netherlands (and aided by the popularizing work of the Frenchman Jean Barbeyrac), within two generations this enterprise had gained a huge pre-eminence in political-philosophical discourse. Taken literally, the idea of a comprehensive, aboriginal, subjective natural right assumed that legitimate political authority must have emerged through some consensual transfer of right from the individual. This might lead either to a historical search for plausible relics of such a contract in customary law (see Selden, J.) or to its identification with specific events - such as Magna Carta or practices such as the coronation oath. Alternatively, philosophers might try to reconstruct the terms of such a grant of authority a priori, on the grounds of what would have been reasonable or - in the case of most thinkers including Locke who retained a natural law constraint - legitimate to agree to. Some thinkers made the grant of right virtually unconditional, binding and irrevocable, establishing as it did the power to punish, which they tied logically or effectively to obligation; others saw it as highly conditional and revocable. Some, like Hobbes, considered there to be two contracts - a contract of union constituting society and a contract of subjection or government between the resulting society and its rulers (Hobbes, T. §7). Others denied the latter (see Rousseau, J.-J.) or the former (as had Vitoria and the Jesuits), or conflated the two (Hobbes). At its most radical, the criterion of natural right became a continuing condition of governmental legitimacy, and the failure to respect it presumptive grounds for resistance (as, for example, wuth the Levellers, Locke and Paine).

Hobbes' masterpiece, *Leviathan*, inserted a Bodinian idea of sovereignty into a synthesis of the language of natural right and contract, materialistic psychology and naturalistic epistemology. His sustained deductive argument, supposedly modelled on geometry, established a model for political philosophy which has lasted, and increased in fascination, until the present day.

A major issue for the new natural rights language, which played a central role in the increasingly commercial and imperial states of Europe, was that of the origin and moral limits of private property rights. A divine sanction for property rights was a logical casualty of the new secular basis of modern natural rights theory, and it was implausible to attach them directly to the natural condition. Arguing against the Royalist Robert Filmer, who identified all authority - political and economic - with the patriarchal monarch or male household head, John Locke distinguished the different occasions and mechanisms by which political and property rights emerged - a collective act of trust and individual labour, respectively. In so doing he gave a philosophical justification of some main aspects of the modern state, combining as it did centralized political authority with dispersed economic rights.

9 The political language of a modern commercial society

Hobbes, despite the fascination of his ideas, was widely repudiated for the atheism confidently supposed to underlie his egoism, naturalism and materialism. The powerful new view of natural right as unconstrained and subjective was seen to license a threateningly egoistic psychology - despite attempts to mollify it by ascriptions of natural sociability to humans (as in Pufendorf). Such worries were reinforced by the increasingly commercial character of society, analysis of which was a major preoccupation for eighteenth century thinkers. Bernard Mandeville's claim that the 'private vices' of individuals aggregatively produced 'public benefits' still seemed to offer - as he may have intended - only the alternatives of austerity or hypocrisy (see Mandeville, B.). His opponents sought to diffuse this opposition either by asserting an innate moral disposition or moral sense (see Moral sense theories; Hutcheson, F.), or by constructing some convincing account of how moral dispositions emerged from more primitive psychological traits, such as sympathy (see Hume, D.) or sociability (see Smith, A.).

The growing preoccupation with the social processes by which the moral sentiments - or more generally the 'opinion' on which politics was increasingly seen to rest - were constructed, transmitted and sustained revealed a tension in political philosophy between moral philosophy or natural jurisprudence and historical sociology. A republican tradition deriving ultimately from Machiavelli's speculations on the decline of Rome, but reinterpreted by Harrington and his English followers, and by Montesquieu in France, provided a helpful vocabulary in which to treat patterns of secular historical change. Yet while this republican language was a rich resource in theorizing long-term and impersonal socio-economic processes, it was preoccupied with decline and initially suspicious of wealth, commerce and refinement as destructive of the public virtue required in the commonwealth.

The emergence of a political theory built around a programme of 'politeness' in seventeenth century France and in Britain at the start of the eighteenth century by Shaftesbury, by Addison and Steele in their journal *The Spectator* and by Francis Hutcheson was one of a series of attempts to accommodate the polity's presumed need for political virtue with the development of a wider commercial society. The ultimate achievement of Adam Smith and David Hume was to explain how the progress of commerce, by encouraging rational patterns of thought, wider social interaction and refined appetites, led to the decline of irrational 'enthusiasms' -party political or religious - and encouraged the emergence of a polite society supportive of liberal, moderate politics (see Enthusiasm). Like Montesquieu they came to reject the simple contrast between the liberty of republics and the servility of monarchies, and described modern, commercial monarchies as supportive of a new kind of private liberty resting on structural features rather than on individual qualities. Both thinkers, however, recognized the precariousness of this balance, Smith pointing to the debilitating effects produced by the division of labour, and Hume to the danger of regimes bankrupting their economies through succumbing to the temptation offered by the new institutions of public credit. In general this analysis drew attention to the vital role of opinion in sustaining, or undermining, political rule, a preoccupation wholly in keeping with the undoubted fact of increasing popular political mobilization. Hume's programme for a more complaisant politics involved the deployment of sceptical and consequentialist arguments to displace opinions derived from metaphysical notions of right in favour of sober understandings of interest.

Belief in right was, nevertheless, the motivating force behind two modern Revolutions in America and France. While Locke, Blackstone and English common law maxims ('no taxation without representation') were invoked to justify resistance to English tyranny, American publicists increasingly drew on Montesquieu and Hume in seeking to establish a constitutional settlement in their new, extensive and federal republic. *The Federalist Papers* composed by 'Publius' (Jay, Madison and Hamilton) sought to adapt republican vocabulary to persuade a wide political public that the proposed constitution provided for a central government powerful enough to replace the defective Articles of Confederation, while leaving enough power to the states to assuage the worries of those who believed that republican liberty could not survive in large polities. Through its popular, yet state-by-state, ratification the American constitution comprised an uneasy tension between popular and constitutional sovereignty. The restraint of power, something of an eighteenth-century Anglophone preoccupation, was achieved in America through its division - between centre and periphery, between different government institutions (legislature, executive, judiciary) and through the staggering of elections over time.

In France, the preoccupation with 'opinion' took a much more didactic form through the activities of the *philosophes* particularly in the assembling and publication - principally by Denis Diderot and Jean d'Alembert - of *L'Encyclopédie*, a self-conscious attempt to reform the terminology and preoccupations of scientific, cultural and political language. Much influenced by a reading of Lockean psychology and politics, but with strong materialist

and agnostic strands, the *philosophe* movement conducted an elitist rather than a democratic intellectual guerrilla war in (and later from outside) the territory of French absolutist sovereignty. However, it was the anti-*philosophe* Jean-Jacques Rousseau who produced the most arresting political philosophy, synthesizing French absolutism and his native Genevan republicanism.

Rousseau fundamentally challenged the growing accommodation of inherited political languages to commercial modernity so evident in the writings of Hume, Montesquieu and Smith, and later 'Publius' and the *philosophes*. His *Discourses*, especially that on the *Origins of Inequality*, while recognizing the essentially sociohistorical sources of morality, asserted an intimate connection between the development of commercial civilization and that of dependency, inequality and loss of self. For him there was hardly space - conceptual or historical - for an account of a society advanced enough to be civil and moral without yet being divided and corrupt. In *Du contrat social (The Social Contract)*, however, he sought to present this possibility in the language of contract theory. Central to his argument was the absolute, yet inalienable, quality of sovereignty established by the contract constituting society and retained by the citizens in their collective identity as a sovereign body independent of government. Rousseau claimed thereby to have solved the paradox of political freedom, for despite subjection to a common sovereign, each as legislator 'remains as free as before'.

Rousseau's attempts to ensure constitutionality - through observation of due formal processes, the subjection of all to universally phrased laws and his insistence both that the sovereign body could never be represented, and that it might sometimes fail to express the legitimizing, but noumenous, 'general will' - were widely ignored by his followers during the Revolution, reaching their most ironic expression in Robespierre's claim to embody the latter in his person.

The French Revolution - like the rise and fall of Rome before it - defined a matrix of problems and linguistic perspectives for political thinkers of the succeeding centuries. These included the conception of sovereignty as popular will and the consequent dangers of popular mobilization, the identity and significance of the Revolution within world historical processes and the relationship between philosophical doctrine and politico-historical outcomes.

10 Response to the French Revolution

The dominant language of the Revolution was that of modern, secular, natural rights, which interpreted such rights as inherent properties of individuals, and which attempted to deduce the legitimacy of particular institutional arrangements from these abstract ethical principles. While this model of political philosophizing remains very much alive today (see Rawls, J.) its implication in the violence of the Revolution led to strong philosophical reactions.

Both features of natural rights were attacked by Edmund Burke, whose response has had an enduring appeal in the Anglophone and German-speaking worlds. *Reflections on the Revolution in France* stressed the inadequacy of human reason to comprehend the function and utility of many social institutions, beliefs and practices which have been built up over time through adaptation to local circumstance. Drawing on sceptical epistemology and common law jurisprudence, Burke privileged the collective knowledge embodied in customs and institutions over individual claims to knowledge, particularly those that derived from analytical or deductive reasoning. The Revolution was thus, at the level of ideas, an outbreak of irrational secular 'enthusiasm', parallel to religious movements of the past. Moreover its particular content - natural right - licensed the intrusion into politics of all human agents whether suitably educated or not. Important responses to Burke included those of Tom Paine, William Godwin and Mary Wollstonecraft, who, while rhetorically defending natural rights against Burke's attack, used a republican vocabulary of virtue both to open discussion of the political significance of the private arena and to establish a political role for women.

In France the aftermath of the Revolution produced attempts to divert the language of republican liberty from its ancient austere and civic-activist roots, to which Benjamin Constant famously sought to add an understanding of modern, commercial liberty. Alexis de Tocqueville, too, in *De la démocratie en Amérique (Democracy in America)*, provided a socio-historical analysis both of the *ancien régime* and of what he took to be the paradigm of emerging modern society. Meanwhile the rationalist, social-engineering aspect of the Enlightenment might be said to have lived on in the industrial utopianism of Fourier, Saint-Simon and his followers including Comte (see

Utopianism; Saint-Simon, C.-H. de R.; Comte, A.).

While the language of radical natural rights declined in the nineteenth century (later to be revived in the era of international charters and organizations) a new radical language - that of secular utilitarianism - was being consciously forged at this time by Jeremy Bentham and would be elaborated and promulgated by his many reforming disciples (see Bentham, J.; Utilitarianism). While general appeals to utility had been common in earlier epochs, particularly in Cicero and his followers, Bentham proposed something much more precise; namely, that the aggregate quanta of anticipated pleasure and pain should be the sole criterion of public policy. He also laid out exhaustively the dimensions by which the necessary calculations might be attempted. Resolutely dismissive of metaphysics - including inalienable natural rights (which he viewed as ‘nonsense on stilts’) - Bentham’s initial work was in jurisprudence, applying positivist criteria to Blackstone’s *Commentaries on the Laws of England*, but he gradually arrived at the view that only a democratic franchise could ensure the political will to implement the utilitarian legal code he sought. Influential in British and French nineteenth-century social reforms and in a variety of liberal nationalist movements from Greece to South America, Bentham’s version of utilitarianism nevertheless ran into philosophical difficulties, from which John Stuart Mill, the son of Bentham’s collaborator James Mill, sought to rescue it.

Mill’s reformulations involved acknowledging a qualitative dimension of pleasure, limiting utility’s potential for eroding moral rules - especially justice - by developing, in all but name, a distinction between rule- and act-utility, and placing the utilitarian assessment of political phenomena and institutions within an historical, developmental schema. In this way - and by a famous defence of the liberty needed to innovate - he sought to increase, as well as use, the ‘good qualities’ present in society, and to secure ‘utility in the largest sense, grounded on the permanent interests of a man as a progressive being’. This extension of utilitarian language nevertheless diminished its original programmatic policy appeal, as well as its original distance from the ideal-regarding criteria to which Mill’s successors, the ‘New Liberals’, increasingly turned.

11 Idealism and Marxism

Late nineteenth-century liberals in Britain (such as T.H. Green, Bernard Bosanquet and F.H. Bradley) and in Italy (Benedetto Croce) appealed to a revived idealist language which, following Kant, had developed through to mid-century in Germany, but, with a few significant exceptions (such as Coleridge) had not exported well (see Idealism). This language rejected the Lockean premise of an essentially passive mind, and instead emphasized, in varying degrees, the agency of the mind in organizing experience, and the essentially ideational character of reality. Powerfully articulated by Kant, the social and political implications of this were spelled out most fully by Hegel, who sought to synthesize history and philosophy - time being, he claimed, the dimension in which thought articulates itself through a series of projections which it does not initially recognize as its own product. In the social sphere of right (objective mind), Hegel sought to show that the essentially legal categories of right such as property, contract and punishment - no less than the institutions of the family, civil society and ultimately the state itself - are not obstacles to individual freedom. Instead they are logically necessary corollaries of the very idea of a free will which come into historical existence as these different aspects of consciousness separate themselves out. Working to an agenda set very much by the French Revolution, Hegel defined freedom as the historical movement of will from mere potential and contingency towards its differentiated concrete embodiment in institutions. Attempts, such as those of the Jacobins, to appeal directly to abstract standards - ‘the revolutionary will’ - therefore courted a destructive anarchy. For Hegel the will finds emancipation when it recognizes the differentiated structures of the modern state as its own product. Another, only questionably political, view would be taken by Nietzsche for whom self-creativity required the individual will to power to overcome all ‘given’ contexts, natural or historical; failure to do so would reinforce a ‘slave’ mentality of subjection (see Nietzsche, F.).

Hegel’s development from the *Phenomenology* to the elements of the *Encyclopedia of the Social Sciences* including the *Philosophy of Right*, represents an increasingly abstract - although always historicized - account of what was, in his initial writings, a deeply socially embedded history of consciousness. The political legacy of his philosophy was extremely ambiguous, since the claim that historical progress represents the working out of reason in the world could be read either as an endorsement of the status quo or as an exhortation to subject it to rational criticism. However, since Hegel’s conception of reason was supposed to be inscrutable to those acting on its behalf, the former reading is the more plausible.

Largely for political reasons, the battle over Hegel's legacy was fought in terms of the philosophy of religion. A series of commentators - David Strauss, Ludwig Feuerbach and the young Karl Marx - sought to interpret religion, which Hegel had described as 'truth in a veil', as a representation of the rationality or irrationality of the life of its own society. While Feuerbach saw Christianity as mystified anthropology and the incarnation of Christ as an allegory of the religious form assumed by humanism, he was unable to explain why humanism should have taken this form in the first place. In the early writings of Marx he argued that the sources of religion, as an alienated form of consciousness, could be traced to social deformations of man's essentially creative nature, which Marx located in the economy. Marx interpreted Hegel's historical schema of the alienation of mind from its own intellectual products as a mystified account of the alienation of humanity from its own material products. This entailed in turn a series of alienations - from creative activity (experienced as drudgery), from other humans (experienced as material factors or competitors), from nature (experienced as something to be subjected) and from men's own philosophical identity or 'species being' (as collective creators) (see Alienation).

The consequence of this alienated activity was a structure of alienated and alienating institutions, the primary political feature of which was the class dominance of those effectively controlling the means of subsistance. The state - on Hegel's account the 'presupposition of civil society' - was shown instead to be a product of civil society, and thus dependent on the balance of class power. In the words of the *Communist Manifesto*: 'the executive of the modern state is but a committee for managing the common affairs of the whole bourgeoisie'.

Although Marx arrived at his revolutionary position through an internal critique of idealism, and such language persisted in his writings, his actual analysis of historical processes, and of the particular form of society then prevailing, owed much to the Scottish historical sociologists and the political economy spawned by Smith's *The Wealth of Nations*. The conception of social evolution through stages characterized as different 'modes of subsistence', and the notion of the social production of ideas and institutions intrinsic to and functional for each stage, was vital to his historiography. Equally the development, from Smith to Ricardo (interpreted by Marx as a development in objective economic circumstance), of a terminology for analysing the creation and distribution of value, was vital to his critical economics and to his projections about pauperization, shifts in social structure, demand failure and the falling rate of profit, which he claimed would immobilize capitalism as a means of deploying resources and generating wealth.

The young Marx had criticized capitalism for denying the human essence - what he viewed as a collective rational creativity. The mature Marx self-consciously eschewed metaphysical ideals. His socialism was scientific in the sense of claiming to rest on an objective analysis, whereby competition amongst capitalists would gradually increase the division of labour and reduce the labour factor of production, simultaneously diminishing profits (which on Marx's analysis ordinarily represented surplus value created by labour), reducing demand and swelling the ranks of the unemployed and unskilled workers. Revolution would occur when profit rates were too low to generate investment despite the huge productive potential developed by capitalism. Production would have then to be organized socially - which only the dispossessed workers would be willing to do.

Classical Marxism retreated under the combination of theoretical developments in economics - principally marginalist analysis - and political developments in early twentieth-century western Europe, especially the advent of social democracy. Its Leninist version, which sought to apply Marxism to newly industrialising states as vulnerable points in capitalism conceived of as an international system, survived as an ideology of Soviet imperialism in eastern Europe and anti-imperialist movements elsewhere, but largely ceased to generate serious philosophical work (see Marxism, Chinese; Marxist philosophy, Russian and Soviet; Marxist thought in Latin America; Marxism, Western). Varieties of Marxism continued to inspire political analysis, however, notably that associated with the Frankfurt School (see Critical theory; Frankfurt School).

Most of the major languages of political theory have similarly survived to the present, albeit in changed or often hybrid form. Thus Rawls' synthesis of utilitarian and Kantian perspectives is enmeshed in a debate with a variety of communitarian approaches with debts to revived Aristotelian or republican language, and with secularized versions of natural rights language and contractarian theory (see Gauthier 1986; Contractarianism). Wholly new challenges to these adaptations of traditional languages are to be found in the voices of feminist theorists who have offered strikingly unfamiliar and rewarding perspectives on what were thought to be well understood historical texts, as well as a different agenda for contemporary political philosophy.

See also: al-Afghani; Bushi philosophy; Gandhi, M.K.; Guanzi; Huainanzi; Jia Yi; Kumazawa Banzan; Marsilius of Padua; Marxism, Chinese; Political philosophy, Indian; Political philosophy in classical Islam; Shōtoku constitution; Sunzi; William of Ockham; Yi Yulgok; Zionism

IAIN HAMPSHER-MONK

References and further reading

Black, A.J. (1970) *Monarchy and Community: Political Ideas in the Later Conciliar Controversy 1430-1450*, Cambridge: Cambridge University Press.

Burke, P. (1991) 'Tacitism, Scepticism, and Reason of State', in J. Burns and M. Goldie (eds) *The Cambridge History of Political Thought 1450-1700*, Cambridge: Cambridge University Press.

Burns, J. (ed.) (1988) *The Cambridge History of Political Thought, c.350-1450*, Cambridge: Cambridge University Press.

Burns, J. and Goldie, M. (eds) (1991) *The Cambridge History of Political Thought, 1450-1700*, Cambridge: Cambridge University Press.(A collection of essays by individual authors, with an extensive bibliography.)

Cicero (54-51 BC) *De republica (On the Republic)*, trans. G.H. Sabine and S.B. Smith, *On the Commonwealth*, Columbus, OH: Ohio State University Press, 1929.

Gauthier, D. (1986) *Morals by Agreement*, Oxford: Clarendon Press.

Hampsher-Monk, I.W. (1992) *A History of Modern Political Thought, Major Political Thinkers from Hobbes to Marx*, Oxford: Blackwell.(Contextualized studies of the political theories of Hobbes, Locke, Hume, Rousseau, The Federalist, Burke, Bentham, Mill, Hegel and Marx.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.

Kelly, D.R. (1970) *Foundations of Modern Historical Scholarship: Language, Law and History in the French Renaissance*, New York: Columbia University Press.

Machiavelli, N. (1531) *Discorsi sopra la prima deca di Tito Livio*, trans. L. Ricci, 'Discourses on the First Ten Books of Titus Livy', in *The Prince and The Discourses*, New York: Modern Library, 1950.

Machiavelli, N. (1532) *Il principe*, trans. L. Ricci, 'The Prince', in *The Prince and The Discourses*, New York: Modern Library, 1950.

Oakley, F. (1991) 'Christian Obedience and Authority, 1520-1550', in J. Burns and M. Goldie (eds) *The Cambridge History of Political Thought 1450-1700*, Cambridge: Cambridge University Press.

Plamenatz, J. (1993) *Man and Society*, ed. R. Wokler, Harlow: Longman, 3 vols.(Revised text of a classic study with historical contextual material excised from the original edition reinserted by Wokler.)

Pocock, J.G.A. (1971) 'Languages and Their Implications', in *Politics, Language and Time*, New York: Atheneum Press.

Pocock, J.G.A. (1975) *The Machiavellian Moment*, Princeton, NJ: Princeton University Press.(The development of the language of republicanism from Aristotle to the American Frontier, chiefly in Italy and England from the sixteenth to eighteenth centuries.)

Pocock, J.G.A., Schochet, G.J. and Schwoerer, L.G. (eds) (1993) *The Varieties of British Political Thought 1500-1800*, Cambridge: Cambridge University Press, in association with the Folger Institute, Washington, DC.

Salmon, J.H.M. (1987) *Renaissance and Revolt*, Cambridge: Cambridge University Press.

Skinner, Q. (1978) *The Foundations of Modern Political Thought*, Cambridge: Cambridge University Press, 2 vols.(The political languages of the Renaissance and Reformation.)

Skinner, Q. (1988) 'Meaning and Understanding in the History of Ideas', in J. Tully (ed.) *Meaning and Context: Quentin Skinner and his Critics*, Cambridge: Polity Press.

Ullman, W. (1966) *Principles of Government and Politics in the Middle Ages*, London: Methuen.

Van Gelderen, M. (1993) *The Dutch Revolt*, Cambridge: Cambridge University Press.

Wolin, S. (1960) *Politics and Vision*, Boston, MA: Little, Brown.(Idiosyncratic but insightful.)

Political philosophy, Indian

While Western political theory has been framed as the struggle between the state and the individual, Indian political philosophy has been more concerned with issues of self-liberation, morality and leadership. Until recently, with the advent of institutionalized or syndicated Hinduism, Indian society made a softer distinction between state and religion. Classical Indian political theory, as with Kauṭilya, centred on axioms on how to maintain and expand power. Kauṭilya argued that reason, the edicts of the king, and his own rules of governance, the Arthaśāstra, were as important for decision-making as the ancient religious treatises, which defined social structure and one's duty to family, caste and God. With the exception of the Arthaśāstra, politics was expressed through the ability not so much to govern as to define social and moral responsibility, what one could or could not do and who could oversee these rules.

Like all civilizations, India had periods of rule by accumulators of capital and traders, warriors and kings, and Brahmans and monks; there were also revolts by peasants. Still, philosophy was in the hands of the Brahmans, the priestly class. This philosophy was primarily not about artha *(economic gain) or* kāma *(pleasure), but about* dharma *(virtue) and* mokṣa *(liberation from the material world). The attainment of salvation, of release from the bonds of karma, was far more important than the relationship between the individual and the sovereign, as was the case in Western political philosophy.*

1 Historical background

While the European Enlightenment was considered the end of Church power and the beginning of secular power, of humanism and liberalism, there was no similar Enlightenment in India. This is not to say that there was no tension between king and Brahman, between state and religious authorities; indeed, Nicholas Dirks (1987) writes that the central conundrum in Indian history has been who should rule: the Brahman or the *kṣatriya* (warrior)? Kauṭilya focused more on coercive power and less on interpretive power. In contrast, the Buddha, committed to interpretive power, remained silent when questioned whether the *ātman* (the individual soul) and Brahman (the Cosmic Consciousness) exist. Understanding that a positive response would reify the self, returning its control to the Brahmanical class, he opened up the self and consciousness to more liminal spaces. Furthermore, recognizing that official state power was largely circumscribed by priestly power, he focused less on royal power and more on the spiritual community (*saōgha*) and on right livelihood as a way of social transformation.

Thus, instead of the search for a perfect society, of the linear march and inevitable victory of democracy and progress, the regulation and liberation of the self has been far more important. It is this self and its emancipation from personal and social history that has been the centre of Indian philosophy. Indian political thought, however, should not be seen as fundamentally despotic, with authoritarianism the rule. Indeed, P.R. Sarkar (1987) makes the controversial claim that democracy first originated with the Licchivis of Vaiśālī, a clan that flourished over 2,500 years ago. They developed a written constitution, abolished the monarchy and, through elections, formed an executive body.

However, while in Western political theory the assumption of man as evil led to the system of federalism, to checks and balances of power, in India neither evil nor good was assumed. *Vidyā* (introversion, leading to enlightenment or good) and *avidyā* (extroversion, leading to degeneration or evil) remain in constant struggle, both simultaneously present. Evil was explained as ignorance, as *māyā*, not as an embodied force. The ideal leader is the one who can lift the veil of untruth, transcend this duality (or at least minimize *avidyā*) and thus become the moral and spiritual leader. In contrast, Islamic political theory has been saddled with the problem of reattaining the perfect state achieved during the time of the Prophet (see Political philosophy in classical Islam). India did not have such a historical event; only Rāma Rājya, the mythical kingdom of Rāma, where food was abundant and all lived in peace, had utopian connotations. But it was only in the twentieth century that this imagined polity became a political platform, largely in response to the utopias of Marxism, liberalism and Islam. Still, for Indians it is not so much a secular state that is desired (one that is amoral, efficient and fair) but a pluralistic state that does not take sides with religions, thereby enabling authentic cultural plurality.

2 Temporal dimensions

Whereas Western political philosophy creates a division between the religious and the secular, between Church and State, Indian (and Chinese) philosophy, based less on monotheistic, highly structured religions, and more on direct intuitive experiences of the mysterious, has found this to be unnecessary. A different temporal orientation is also taken. Instead of a linear movement from ancient, to classical or feudal, to modern (scientific, rational, nation-state-oriented), or from religious, to philosophical, to scientific, as with Auguste Comte, there is a complex historical cycle with different temporal levels. At one level, there is the cosmic time of the stars and gods; at another level, there is the fourfold structure of student, householder, active citizen, and renunciate; at the societal level, there is the rise and fall of society, a process which involves the degeneration of righteousness from the golden, to the silver, to the copper and then to the iron. In the iron age, the *avatār*, or redeemer, brings all back to *dharma* and the golden era (*satya-yuga*) begins again, but with the silver around the corner. This was the promise of Kṛṣṇa to be reborn whenever virtue declines. At the level of the individual, Indian philosophy is the story of karma, of endless births, with escape only possible through enlightenment (see Karma and rebirth, Indian conceptions of). Instead of utopia pulling society forward, 'eupsychias' (ideal places of the mind) have been more prevalent. However, even as the self has been central, history has been dynastic, consisting of epics and myths, of ages and episodes, not of the workers' struggle, as recent Marxist Indian philosophers have argued; structure has been far more important than agency.

Traditional Indian historiography has framed the past as Hindu (Vedic), Muslim and Modern (British rule and then Independence). However, unlike the Western scheme there is no pretence that prior eras are stepping stones to the present, no overarching theory of scientific progress. This is not to deny the reformist urges throughout Indian history, and particularly the efforts of Raja Rammohun Roy (see Brahmo Samaj) and Aurobindo Ghose, but rather to argue that moral philosophy has been more central than the massive social engineering of Marxism or the amoral market mechanisms of liberal capitalism. Even when utopia has been imagined, such as Rāma Rājya, it has been done with the past in mind, not with the pull of the future.

3 The Indian episteme

Indian political theory should not and cannot be divorced from the Indian episteme, the boundaries of knowledge that contextualize what is knowable. The liberation of the self is the centrepiece of this system. Doing so frees one from karma and helps one to live in *dharma* and achieve *mokṣa*. Truth is considered not so much as accuracy or fidelity to the empirical but as therapy, as that which reduces suffering. In this sense, classical Indian philosophy is postmodern, with reality consisting of many levels and different traditions and philosophies touching these levels - true epistemological pluralism. Truth is thus 'both/and' instead of 'true and false'. This means that contradictions are tolerated. Moreover, there are many ways of knowing the world: devotion, reason, sense-inference, authority and intuition (see Knowledge, Indian views of). The physical universe and the mental universe are considered symmetrical: as above, so below; as inside, so outside. Individual and cosmos, body and mind, self and society are linked together. Politics and religion are thus not separated but linked. There is a holistic unity of discourse, with truth ultimately residing in the individual, even though Brahmans and rulers have attempted to have authority vested in caste and state. Philosophical systems are also grand, touching every level of existence from bathing to political economy, world order and cosmic meaning. Order and elegance are as important as accuracy in the Indian episteme. This episteme itself is not questioned, but is the given in Indian philosophy.

Unlike other epistemes, which undergo dramatic shifts, the Indian episteme is additive. New discourses are added, changing the episteme and Indianizing the new discourse. Modernity, of course, is threatening to force the classical pluralism of Tantra, Vedānta, Buddhism and Islam to fit into the straitjacket of nationalism, of one people, one guru or one god. This has led to a politics of structured religion, to the creation of a 'Hinduism', a religion linked to national identity. In this realist model, it is the state and its functionaries that are most important. Reality is not moral or spiritual, but about capturing power so as to ensure that the identity of and material gains for one's community or nation are maximized. The self is no longer devoted to selfless service (*sevā*), but to maximizing wealth and power. In this self, the Other is to be feared: the Other is the enemy, not to be embraced, as Gandhi argued, but used for personal gain. However, while in the short run Hindu nationalism may become hegemonic, its exclusionary tactics will most likely self-destruct. It will 'other' itself. In the long run, the ecology of cultures and traditions that is India will soften modernity's univocal claims, making it but one view among many.

4 *Varṇa*

The ‘Other’ has generally been more those of a different class than those of a different religion, since Indian religious and philosophical traditions are syncretic and universal. It is *varṇa* (social division; literally ‘colour’) that both orders society and creates structural problems for social justice and transformation. While *varṇa* for Orientalists is an all-encompassing social structure, others have argued that Indian society is no different from other historical societies, where agency has existed at times with individuals, at times with religious authorities, at times with traders and at times with the military. Caste is one category, but not the final category of analysis. Defenders of caste have argued that it is one among many ways of ordering the world - of engaging in politics. Politics is about the negotiation of power and meanings. While modernizers want to rid India of caste, they forget its relationship to order, with caste basically creating a permanent underclass, as in other societies. For the political theorist Rajni Kothari (1970), more important than *varṇa* is *jhat*, the myriad associations, relationships and lineages that are perceived as caste.

For Gandhi, the task was to make everyone into *śūdras* (peasants) to create a society with equal distribution. His was a devastating critique of modernity and instrumental rationality, not of technology, as commonly thought. For Gandhi, self-reliance is the road and the goal. One cannot distinguish between ends and means. This counters the classical position of the *Bhagavad Gītā*, where, in certain situations, violence is allowed, especially since at the level of the Absolute, victim and victor are united in destiny.

The legacies of Gandhi are many. Two are critical: self-reliance rather than instrumental rationality - that is, humans should work together to build small-scale communities - and nonviolence. In conflicts, it is nonviolence that is victorious since it touches the heart of the oppressor, thereby transforming the oppressor. Indeed, for Gandhi, history itself is the march from violence to nonviolence, from barbarism to civilization.

5 Structure, history and the cycle

The work of P.R. Sarkar (1921-90) is an alternative to the critical traditionalism of Gandhi and the reforming efforts of Aurobindo. Sarkar attempts to develop a political theory that has structure (historical patterns), agency (the role of individuals and leadership) and superagency (the role of divine intervention). For Sarkar, there are four types of power: the economic, or *vaiśyan*, the coercive/protective, or *kṣatriyan*, the normative/ideological, or Brahman, and the chaotic/disruptive, or *śūdra*. Sarkar derives these from the classical Indian social system of *varṇa*, but reinterprets them not as biological caste categories but as evolutionary, psychosocial paradigms.

These four types of power are related to four historical ages: the age of workers, the age of warriors, the age of intellectuals and the age of capitalists. At the end of the capitalist era, there is a workers’ revolution; this leads to a centralization of power, from which evolves an age of warriors, in which there is a centralized polity. Each age has it own contradictions, and by denying the other forms of power, each era naturally leads to the next. These are not ideal types, nor is Sarkar’s cycle a defence of current modernity, as with the historian Buddha Prakash, who argues (1958) that the current age is the Golden Era - rather than an *avatār*, independence and industrialism have awakened India after centuries of oppressive sleep.

Sarkar does not intend to end the cycle of history - it is not societal perfection he seeks - but he hopes to minimize the exploitation in each era through the development of a new type of leader, the *sadvipra*, one with a complete and pure mind. The *sadvipra* exists in a context not of state or individual, but of *samāj*, or family: a selfless family on a collective journey through life. Through such leadership, Sarkar envisages a new society committed to gender coordination, self-reliant cooperative economies, and *pramā*, a dynamic balance between physical, mental and spiritual potentials.

However, unlike classical political philosophy, where the sage exists *outside* traditional constructs of power (being neither Brahman nor *kṣatriya*), in Sarkar’s model, this new type of power - protective, service-based, innovative and interpretive - is placed at the centre of the wheel. Like Aurobindo with his yogocracy and Gandhi with his *satyāgrahī*, Sarkar develops the ideal of a modern yogi. This, however, is not the religionization of politics but its spiritualization, the aim that Aurobindo and Gandhi struggled to achieve.

But for Aurobindo, history is not structural but idealistic, a process in which the Godhead enters individuals and communities. Thus, following Hegel, nationalism becomes part of the manifestation of the soul, of the divine

drama. It becomes an extraordinary event that represents the will of Consciousness in creating a better human condition. For Aurobindo, God works not only through *avatārs* such as Kṛṣṇa, but also through nationalistic and other movements as well.

6 Postcolonial discourses

More recent Indian political theory has been concerned with creating a post-colonial self, with ridding India of the intimate enemy of the British history. This is not a history of dynasties or of large meta-narratives, but a history of the subaltern, of women, of the epistemologically oppressed, as developed by writers such as Ashis Nandy, Ranajit Guha, Vandana Shiva and Gayatri Spivak. This rewriting of India has been a search for a historically grounded India not overlaid with spiritual essences, for a politics of individuals and communities struggling to create a new self, an often localized identity that has meaning in its own mythological, cultural and economic context. Development, science and instrumental rationality are criticized as violent, disempowering of communities and dangerous to the environment.

Borrowing from postmodernist critiques of the construction of identity, the categories of 'Hinduism', 'caste', 'Mother India' and 'nation' are questioned. The politics of how these discourses are used by state power further to regulate, militarize and homogenize society are deconstructed with the intention of rescuing the plurality of truth that has typified Indian philosophy.

See also: Duty and virtue, Indian conceptions of; Political philosophy, history of; Political philosophy, nature of; Postcolonialism

SOHAIL INAYATULLAH

References and further reading

Copley, A. (1993) 'Indian Secularism Reconsidered: from Gandhi to Ayodhya', *Contemporary South Asia* 2 (1): 47-65.(Criticism of the new syndicated Hinduism.)

Dhawan, G. (1957) *The Political Philosophy of Mahatma Gandhi*, Ahmedabad: Navajivan Publishing House. (Excellent summary of Gandhi's life and thoughts; easy to follow.)

Dirks, N. (1987) *The Hollow Crown*, Cambridge: Cambridge University Press. (Difficult but rewarding. Explores power, culture and kingship in Indian history.)

El-Affendi, A. (1991) *Who Needs An Islamic State?*, London: Grey Seal Books. (Excellent analysis of Islamic political theory; easy to follow.)

Guha, R. and Spivak, G.C. (1988) *Selected Subaltern Studies*, New York: Oxford University Press.(Difficult but novel approach to politics and history.)

Inayatullah, S. (1997) 'Prabhat Rainjan Sarkar: Agency, Structure and Transcendence', in J. Galtung and S. Inayatullah (eds) *Macrohistory and Macrohistorians*, Westport, CT: Praeger, 132-40.(An easy introduction to Sarkar's thought - comparisons to other grand thinkers such as Ibn Khaldun, Karl Marx and Herbert Spencer are made.)

Inayatullah, S. (1997) *Situating Sarkar: Historical, Structural and Poststructural Inquiries*, Singapore: Ananda Marga Publications.(Explores Sarkar's political and social thought. Uses a critical Foucauldian analysis, also comparative essays on diverse thinkers such as Ssu-Ma Ch'ien, Montesquieu, Aurobindo and Gandhi.)

Inden, R. (1986) 'Orientalist Constructions of India', *Modern Asian Studies* 20 (3): 401-46.(Breakthrough article on the construction of the text of India.)

Kauṭilya (*c*.320 BC) *Arthaśāstra*, ed. and trans. T.N. Ramaswamy, *Essentials of Indian Statecraft: Kauṭilya's Arthashastra for Contemporary Readers*, London: Asia Publishing House, 1962.(An abridged version that is easier to get through than the original.)

Kothari, R. (1970) *Caste In Indian Politics*, New Delhi: Orient Longman. (Traditional but useful political-science approach to caste.)

Murty, K.S. (ed.) (1967) *Readings in Indian History, Philosophy and Politics*, London: Allen & Unwin.(Short essays by Aurobindo, Nehru, Kosambi, Majumdar, Panikkar, Munshi, Murti and others. Covers Marxist, idealistic, theistic, humanistic and other renderings of Indian political philosophy; an essential and easy read.)

Nandy, A. (1967) *Traditions, Tyranny and Utopias*, New Delhi: Oxford University Press.(Brilliant, difficult, but very important book that touches on politics, culture and history.)

Prakash, B. (1958) 'The Hindu Philosophy of History', *Journal of the History of Ideas* 16 (4): 494-505.(Novel rendering of Indian political history that mixes the classical and the modern.)

Said, E. (1979) *Orientalism*, New York: Vintage Books.(The classic work on how the Orient has been fabricated by the West; skilfully written.)

Sardar, Z. (ed.) (1997) *South Asia: Fifty Years On*, special issue of *Futures* 29 (10).(Essays on the future of south Asia by leading south Asian thinkers.)

Sarkar, P.R. (1987) *PROUT in a Nutshell*, trans. V. Avadhuta, Calcutta: Ananda Marga Publications, 16 vols. (Creative polemic, not technical, but written as *sūtras* rather than analysis. PROUT is Progressive Utilization Theory, an alternative global model of political economy, which Sarkar claims will succeed communism and capitalism.)

Shiva, V. (1988) 'Reductionist Science as Epistemological Violence', in A. Nandy (ed.) *Science, Hegemony and Violence: A Requiem for Modernity*, Delhi: Oxford University Press, 232-57.(Brilliant chapter by a leading environmental and feminist writer; difficult for those not accustomed to poststructural sensibilities.)

Thapar, R. (1966) *A History of India*, vol. 1, Baltimore, MD: Penguin.(Classical study of Indian history.)

Varma, V.P. (1960) *The Political Philosophy of Sri Aurobindo*, New York: Asia Publishing House.(Excellent explanation of the grand neo-Hegelian system of Aurobindo.)

Varma, V.P. (1974) *Studies in Hindu Political Thought and its Metaphysical Foundations*, Delhi: Motilal Banarsidass.(Located in the Indian episteme, a useful and thorough approach.)

Political philosophy, nature of

Political philosophy developed as a central aspect of philosophy generally in the world of ancient Greece, and the writings of Plato and Aristotle made a basic and still important contribution to the subject. Central to political philosophy has been a concern with the justification or criticism of general political arrangements such as democracy, oligarchy or kingship, and with the ways in which the sovereignty of the state is to be understood; with the relationship between the individual and the political order, and the nature of the individual's obligation to that order; with the coherence and identity of the political order from the point of view of the nation and groups within the nation, and with the role of culture, language and race as aspects of this; with the basis of different general political ideologies and standpoints such as conservatism, socialism and liberalism; and with the nature of the basic concepts such as state, individual, rights, community and justice in terms of which we understand and argue about politics. Because it is concerned with the justification and criticism of existing and possible forms of political organization a good deal of political philosophy is normative; it seeks to provide grounds for one particular conception of the right and the good in politics. In consequence many current controversies in political philosophy are methodological; they have to do with how (if at all) normative judgments about politics can be justified.

1 Universalism in political philosophy

A central question in the history of political philosophy has been the concept of the political itself, and therefore the range of actions which can and should be subject to law and coercive interference. Political philosophers have tried several argumentative strategies in their attempts to provide a solution. One such strategy, which has taken many different forms, has been to invoke a conception of human nature (see Human nature). If we could reach a good understanding of human nature and basic human needs, desires and purposes, we could then devise the appropriate set of political arrangements which would best meet them. Sometimes such conceptions of human nature have in turn been set against much wider philosophical theories about the place of human nature in the cosmos (as for example in Plato, Augustine and Aquinas); sometimes in relation to God specifically, as the creator who endowed persons with the nature they have (as for example in the writings of Calvin or, in a more limited way, Locke); sometimes in a view of the place of human beings in the natural order (as for example in Hobbes, or Aristotle via his writings on the relationship between man and the biological order); or in an account of how human needs and powers develop in history (as in Hegel and Marx). Equally, some conceptions of human nature as a basis for political organization have been much more limited and empirical in form (as for example in the utilitarian writings of Bentham, who took as a fact of life the propensity of human beings to seek pleasure and avoid pain). Despite all of these differences, we may find a similar strategy being adopted; namely, that in order to justify a particular set of political arrangements we have to understand the basic needs, drives, desires, interests and purposes of human beings and then argue from these to a proper understanding of the appropriate nature and role of politics in human life.

It follows from this view that the claims of political philosophy are both universal and foundationalist. They are universal in that most political philosophers have assumed that their claims about human nature and the concomitant understanding of politics, society and the individual are quite general in scope and not restricted to an understanding of human nature in the context of a particular culture and society. They are foundationalist in the sense that a conception of the good political order is taken to rest upon an objectively true account of human nature. In addition, political philosophy has frequently been regarded as teleological in character, directed towards goals set by the nature of what it is to be human and the kind of political organization which will fulfil human nature as we understand it.

2 The critique of political philosophy

This conception of political philosophy has met with fierce criticism during the twentieth century, to the extent that in the middle years of the century political philosophy as a normative discipline was regarded as dead. There were various reasons for this. One central reason has to do with the claimed relationship between philosophical prescriptions about the good society and its political form, and the basis for such judgments in accounts of human nature. Following the lead of David Hume, many empiricist philosophers have argued that there is a logical gap

between any account of human nature and the evaluative and prescriptive consequences which are thought to derive from this (see Fact/value distinction). On this view there cannot be anything in the conclusion of a valid argument which is not present in the premises. If a theory of human nature is supposed to be empirically grounded, that is to say testable or verifiable by empirical investigation, then such a set of empirical propositions about human nature cannot support evaluative conclusions about political organization. (This is sometimes known as the is/ought gap.) This claim has frequently gone alongside the claim that the theories of human nature with which political philosophers have operated have not been strictly empirical but rather have been constructed with a strong evaluative dimension built into them - that, for example, some human needs or human purposes have higher moral priority over others. On this view, the reasons why political philosophers have thought that their theories of human nature have supported evaluative conclusions is that the theories themselves have contained ineliminable evaluative elements. The upshot is that theories of human nature cannot provide a secure foundation for political philosophy, because either they are empirical theories which will not support normative conclusions or they are themselves normative and stand in need of justification.

This objection to the strategies followed by normative political philosophers has been reinforced by a growing sense of the subjectivity of values. Sometimes this is taken as a sociological observation - that it is as a matter of fact that the modern world is marked by the view that moral values are matters of subjective preference. If this is so neither normative political philosophy nor the normative understanding of human nature with which political philosophers have worked can be given an objective foundation. Moral judgments, whether they apply to theories of human nature encompassing accounts of human needs and purposes and the priorities between them, or directly to the nature and scope of politics, are matters of subjective preference.

Equally, attempts have been made by philosophers such as the logical positivists who flourished in the 1930s to provide an elaborate justification for understanding moral judgments in a subjective and emotive way (see Logical positivism; Emotivism). The positivists argued that apart from the truths of logic and mathematics, which are tautologies (that is to say, hold by definition), the only meaningful statements are those which are in principle empirically verifiable, as for example the propositions of natural science and everyday empirical observations. This led them to the view that moral judgments, including those made about the nature of politics, have no cognitive content, convey no information. Rather they embody or evince the emotional attitudes of those who make the judgments. In this sense moral judgments are irredeemably subjective.

In the same vein it was argued that political philosophers, particularly utilitarians such as Bentham and J.S. Mill, had committed what G.E. Moore called the naturalistic fallacy when they attempted to give normative conceptions such as good and bad an empirical definition in terms of pleasure and pain (see Moore, G.E. §1). These conceptions are not extensionally equivalent because it is always possible to say ‘this is pleasant, but is it good?’, a question which would make no sense if the words denoted the same thing. Again, this would mean that it is impossible to give an empirical basis to normative reasoning about politics.

These philosophical arguments about the limitation of empirical justification for normative conclusions reinforced the idea that normative judgments are subjective. If this is so then the project of normative political philosophy has to be abandoned. There can be no foundationalist or universalist basis for political judgments. If political philosophy has a role at all it is very modest and involves conceptual analysis - that is to say, clarification of the concepts that might be used in empirical political science with a view to purging such concepts of evaluative, and therefore subjective, overtones.

3 The communitarian response

Over the past thirty years, however, there has been a vigorous resurrection of political philosophy. How have political philosophers felt able to engage in a discipline which had looked nothing more than a pseudodiscipline, merely recording the consequences of the moral predilections of the theorists themselves?

Several factors have contributed to the revival. It has not proved possible to produce a watertight account of the principle of verifiability central to positivism, whose aim was to provide a clear criterion in terms of which the empirical and the non-empirical could be demarcated. Partly because of this, and partly under the influence of Wittgenstein’s later writing, there has been much more attention paid to the meaning that terms have in particular contexts (including the moral and political), than to any attempt to impose some kind of antecedent standard of

meaningfulness on language. This has freed political philosophy to consider, for example, the meaning which a term such as ‘justice’ might have in different discourses, contexts and communities, rather than trying to establish its meaning a priori (see, for instance, Walzer 1983).

Those who wish to link analysis in political philosophy to the interpretation of the communities and networks within which values and principle have their meaning have come to be called, in one contemporary usage of the term, ‘communitarians’ (see Community and communitarianism). This approach to political philosophy does not aim to justify values and principles in some foundationalist or universalistic sense, but rather to interpret (in a way that goes beyond surface description) the values which animate a particular society, or specific communities or discourses within it. It side-steps the problem of justification posed by the positivists and others who regard values as subjective. It eschews the idea of providing some kind of universalistic foundation for political values, rooted, for example, in a theory of human nature. It takes communities and the social meanings embodied in communities and other ways of life as basic and seeks to interpret them.

Another reason for the revival was that the underlabourer role to which positivism consigned political philosophy - providing central definitions of political concepts to be used in empirical political analysis - proved impossible. The evaluative component of political terms is not a peripheral and detachable feature, but is central to the meanings which these terms have in particular discourses, ideologies and communities. This perception is linked to the idea that basic social and political concepts are essentially contestable (Gallie 1956). This is the view that such concepts cannot be given some wholly neutral, value-free, experiential definition; rather, their meaning is bound to different ways in which different discourses and communities construe human purposes, human interests and human flourishing.

Recognizing this may lead us to two alternative conclusions. The first limits political philosophy to description and interpretation. Given that political concepts are essentially contestable, and that they are contested because their meaning is relative to a particular discourse or ideology, then all that political philosophy can do is describe the role which the concept plays in one discourse and explain how its meaning differs in others (see Ideology §4). On the other hand, if a political philosopher wishes to do more than trace how a political concept relates to a wider network of beliefs and values, and instead wishes to offer some basic justification for using that concept in a particular way, then that must involve privileging one way of thinking about human interests and purposes over another. This would take the subject back to its more traditional form, that is reflecting on human purposes and the nature of human flourishing and endorsing a set of political values which would facilitate this.

The problem here is how to ground a normative theory of politics in circumstances where people disagree about their conceptions of the good. In so far as this dilemma has been solved in contemporary political philosophy it has been in one of two ways. The first has been to put, as John Rawls (§4) argues: ‘the Right before the Good’ (Rawls 1971). That is to say, we can recognize that citizens differ about what they take to be good in human life, but nevertheless they have an interest in the justice of the political framework within which they pursue their own diverse goals. One can argue that it is possible to produce a compelling account of the rules of justice within a diverse society by making some rather minimal assumptions about certain basic human goods which all people are assumed to want more rather than less of, whatever else they may think is good in life; and then to invoke some general principles of rationality to decide upon rules governing the distribution of such ‘primary goods’. Theories of this kind aim to justify a political order by reasoning on the basis of slight assumptions about human goods, while recognizing the pervasive nature of moral diversity in relation to more specific human goals and purposes.

The other main approach to the problem of justification has been to emphasize the idea of dialogue in politics, and to explore the conditions for free and rational dialogue about basic political precepts. Here again is recognition of the problem of philosophical justification of principles in a situation of moral diversity, and the response is to ask under what conditions rational and noncoercive dialogues about political goals could take place. Examples of this sort of approach are to be found in the work of Ackerman (1980) and Habermas (1987).

Contemporary political philosophy does not have a settled nature. Following the challenge of positivism, it has become much more reflective about methodology, but as the different approaches of communitarians and foundationalists illustrate, there are still profound differences about the scope and ambition of political philosophy. Is it to be engaged in justification or interpretation? Is it concerned with universal values or the values of particular communities? Are the problems of moral diversity to be resolved by invoking minimal assumptions about human

nature and applying formal modes of reasoning to these, or by grappling with this diversity as it manifests itself in the different interpretations of values in a particular society?

See also: Political philosophy, history of

RAYMOND PLANT

References and further reading

Ackerman, B. (1980) *Social Justice in the Liberal State*, New Haven, CT: Yale University Press.(A book which seeks to produce an argument in favour of liberal political arrangements while making use of minimalist moral assumptions by utilizing neutral dialogue.)

Barry, B. (1965) *Political Argument*, London: Routledge & Kegan Paul.(A classical text in post-war political philosophy. An excellent example of conceptual analysis. Important also in its contrast between ‘want regarding’ and ‘ideal-regarding’ political philosophies.)

Barry, B. (1973) *The Liberal Theory of Justice*, Oxford: Clarendon Press.(A systematic analysis of J. Rawls’ *Theory of Justice* in both its methodological and substantive aspects.)

Berlin, I. (1958) *Four Essays on Liberty*, Oxford: Oxford University Press, 1969.(A contemporary classic adumbrating the distinction between negative and positive liberty.)

Gallie, W.B. (1956) ’Essentially Contested Concepts’, *Proceedings of the Aristotelian Society* 56.(A classic piece of analysis which shows that there cannot be a purely neutral analysis of political concepts.)

Galston, W. (1991) *Liberal Purposes*, Cambridge: Cambridge University Press.(A book from within the American Liberal tradition which shows that Liberalism has to draw upon some conception of the virtues.)

Gewirth, A. (1978) *Reason and Morality*, Chicago, IL: University of Chicago Press.(Probably the most thorough going defence of rights within the context of moral pluralism.)

Habermas, J. (1987) *The Philosophical Discourse of Modernity*, Cambridge: Polity Press.(One of Habermas’ many books which provides background to the points made in this entry.)

Kymlicka, W. (1989) *Liberalism, Community, and Culture*, Oxford: Clarendon Press.(Looks at the relationship between liberalism and ‘thicker’ ideas about community and culture.)

Plant, R. (1991) *Modern Political Thought*, Oxford: Blackwell.(An overview of many aspects of post-war political thought.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(*The* classic text of post-war political philosophy in which the strategy of putting the ‘Right’ before the ‘Good’ is developed in a very sophisticated way.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(Considers how public deliberation about constitutional arrangements should be conducted in a liberal society.)

Rorty, R. (1989) *Contingency, Irony, and Solidarity*, Cambridge: Cambridge University Press.(Develops ideas drawn from Wittgenstein and continental philosophy into a ‘postmodern’ approach to political philosophy.)

Walzer, M. (1983) *Spheres of Justice: A Defence of Pluralism and Equality*, New York: Basic Books; London and Oxford: Blackwell.(A challenging book which rejects the idea that there can be ‘thin’ moral conceptions which can be drawn upon to underpin principles to govern society.)

Pomponazzi, Pietro (1462-1525)

Pietro Pomponazzi was the leading Aristotelian philosopher in the first quarter of the sixteenth century. His treatise De immortalitate animae (On the Immortality of the Soul) (1516) argues that although faith teaches immortality, natural reason and Aristotelian principles cannot prove it. In De incantationibus (On Incantations) (first published in 1556), Pomponazzi attempts to demonstrate on rational grounds that all reported miraculous suspensions or reversals of natural laws can be explained by forces within nature itself. Separating faith and reason once again, Pomponazzi proclaims his belief in all canonical miracles of the Church. These arguments cast doubt on morality, for without an afterlife, humanity is deprived of rewards for virtue and punishment for evil; and nature itself appears to be governed by impersonal forces unconcerned with human affairs. However, morality is restored to the universe by the human powers of rational reflection which lead to the pursuit of virtue. Yet in De fato (On Fate) (first published in 1567), Pomponazzi challenges the very basis of his own ethical doctrine by arguing that all activity of insentient and sentient beings is directed to preordained ends by environmental factors. Unable to justify human freedom on rational grounds, he then seeks to re-establish it using arguments derived from Christian natural theology, thus reversing his usual separation of faith and reason.

1 Life and works

Pietro Pomponazzi (Petrus Pomponatius) was born in Mantua. His adult life was spent as a professor of natural philosophy at the Universities of Padua, Ferrara and Bologna, where he was admired by his students and paid well. Many of his unpublished lectures and questions survive because students, impressed by their intellectual rigour, copied and recopied them. Pomponazzi's most productive years were spent at the University of Bologna where he taught from 1511 until his death. During this period, he published the immortality treatises, *De immortalitate animae (On the Immortality of the Soul)* (1516), *Apologia* (1518) and *Defensorium* (1519), and wrote but did not publish *De incantationibus* and *De fato* (each written in 1520 but published posthumously in 1556 and 1567 respectively).

2 Immortality

Pomponazzi's most famous work, the treatise *On the Immortality of the Soul* (1516), encloses the philosophic argument in a conventional fideistic shell. At the beginning and end of the treatise, he remarks that the doctrine of immortality must be accepted as revealed truth, certain and indubitable; all contrary arguments are false (see Soul, nature and immortality of the §2). The philosophic core of the treatise, however, argues strongly for mortality on rational and Aristotelian grounds. The philosophic argument holds that only the intellect, the highest part of the soul, is capable of surviving the body. Proof of this survival must be found in an activity of the intellect that functions without bodily dependence. No such activity can be discovered, however, because the highest activity of the intellect, the attainment of universals in cognition, is always mediated by sense impressions. Thus even the highest power of intellection is always directly linked to and sustained by the sensitive powers which, in turn, are dependent on the vegetative power. All sources agree that the sensitive and vegetative powers weaken and perish, and hence the intellect, although immaterial in itself, must perish with them. The whole soul therefore dies with the body.

The philosophic doctrine of mortality thus stands in direct opposition to the religious doctrine of immortality. Is Pomponazzi proclaiming the simultaneous truth of two opposites? Some contemporary critics claimed that he asserted the notion of a 'double truth': the doctrine of immortality can be simultaneously true in theology and false in philosophy. However, there is no evidence to support such a claim. He held to only one truth but whether that was the truth of faith or philosophy is open to question. There is good reason to believe that Pomponazzi maintained the absolute truth of philosophy. In discussing the origin of the doctrine of immortality, he notes that religious lawmakers asserted immortality 'not caring for truth but only for righteousness'. Apparently, the truth the religious lawmakers 'do not care for' is the philosophic doctrine of mortality.

On Immortality provoked an immediate furore. Public attacks on Pomponazzi culminated in a warning by Pope Leo X demanding that he disavow the doctrine of mortality and bring his position in line with the decree of the Fifth Lateran Council (1513) which had declared the immortality of the soul to be a dogma of the Church. The intervention of Pomponazzi's powerful friends in the Church allowed him to continue teaching and to publish

replies to his critics. The attacks were accompanied by a spate of books written against Pomponazzi, who wrote two works in his own defence. In his *Apologia* (1518) he replied to Gaspar Contarini and in *Defensorium* (1519) he answered Agostino Nifo (§1). Although more elaborate and diffuse than the original treatise, these works do not alter Pomponazzi's basic position. A full-scale attack came from the prominent Dominican theologian Bartolomeo de Spina. He alone charged Pomponazzi with heresy and demanded an Inquisitorial trial. His call went unheeded. Spina wrote two books against Pomponazzi but the first one, *Tutela veritatis de immortalitate (The Whole Truth Concerning Immortality)* (1519), is probably the more significant. Noting Pomponazzi's assertion that immortality is the invention of religious lawmakers, Spina concludes that Pomponazzi's attitude is not that of a believing Christian.

This issue continues to divide Pomponazzi's interpreters. If, as Pomponazzi always says, revelation proceeds from supernatural truths beyond rational discourse, it follows that reason cannot invalidate any proposition of faith. This fideism, many scholars argue, is certainly compatible with faith so there is no reason to doubt Pomponazzi's sincerity (see Kristeller 1964). Others have reached a different conclusion (see Pine 1986). Citing Pomponazzi's constant identification of philosophy with truth and religions with social convention, they argue that he stressed the importance of religions as a force producing social cohesion but lacking in speculative veracity. The association of truth with philosophy is particularly striking in his idealization of the philosopher as one who is isolated, derided and persecuted for his lonely struggle towards virtue and truth.

3 Miracles

Pomponazzi's investigation of miracles, *De incantationibus*, was written in 1520 but published posthumously in 1556. The decision to leave the work unpublished probably stemmed from Pomponazzi's concern for his own safety after the immortality controversy.

The philosophical discussion of miracles, proceeding by rational argument alone, is again enclosed in a fideistic shell, professing belief in all canonical miracles of the Church. Yet the philosophical discussion discovers natural causes, first for contemporary events deemed miracles, then for biblical miracles, and finally for the miracles specifically related to the founding of all religions. Thus Pomponazzi concludes with a natural basis for religion itself, which now becomes the effect of the workings of the Intelligences, divine impersonal forces moving the heavens, rather than the fiat of a personal deity.

Many supposedly miraculous cures and visions produced by magicians and divine messengers undoubtedly occurred, says Pomponazzi. Yet they all have perfectly natural explanations. Naturalistic accounts are also applied to biblical events, casting doubt on the miracles of Moses, Elisha and Christ. After the discussion of particular miracles, religion itself is depicted as a product of the Intelligences and subject to the eternal cycles of birth, growth and decay. At its birth, a religion is favoured by the Intelligences. It is at that point that the heavens, directed by the Intelligences, produce 'sons of God', men possessing the ability to command nature. But as the natural process develops, every religion grows weaker and eventually dies. Christianity itself is dying, having reached the end of its life cycle.

The doctrine of a universe without miracles, says Pomponazzi, cannot be spread to the vulgar. They must still believe that a providential God, angels and demons are the causes of 'miraculous events'. But the philosophical elite knows that these are 'fictions' designed to lead men to virtue (see Miracles).

4 Morality and freedom

Pomponazzi's conception of morality is based on the free choice of the individual. That choice is free only when the passions are subdued and reason is ascendant. Since only the philosophic elite is capable of dominating the passions, they alone are truly free. Through reason they understand that virtue is its own reward and vice its own punishment. The masses, ruled by their passions, are subject to environmental and planetary forces eliminating their freedom.

The centrality of the concept of freedom in Pomponazzi's thought is underlined by his insistence in the last book of *De fato* that freedom is a metaphysical necessity based on the Great Chain of Being. For that chain of being requires a nature differentiated both from the gods and from the lower forms of life, all of whose natures are fixed and hence without freedom of choice. The universe, he insists, requires a being whose nature is a product of free

choice rather than a fixed essence. Human nature is created by individual choices rather than a divinely given essence. This theme, asserted in various forms throughout Pomponazzi's works, is in contradiction to the determinism he forcefully defends in the first book of *De fato*. Thus the battle between freedom and determinism is at the very core of Pomponazzi's thought and is never fully resolved.

The attack on freedom in Book 1 of *De fato* takes the form of a defence of Stoic determinism (see Stoicism §20). Every event, Pomponazzi argues, is directed by an external cause which guides its internal nature to a predetermined end. Although we deliberate over moral alternatives, our choices are already determined by natural forces directing our wills. Yet after subjecting the will to environmental determinants, Pomponazzi undertakes a new attempt to save human freedom within a framework of Christian theology. This represents a surprising reversal of method since all of Pomponazzi's previous works have insisted that theology is not a legitimate concern of the philosopher who always proceeds 'within natural limits'. By combining doctrines derived from Aquinas and Duns Scotus, Pomponazzi attempts to rescue the will from environmental determinants, although he recognizes that his solution is not entirely satisfactory (see Free will §2).

When Pomponazzi moves from environmental determinants to the attributes of God, the problem is still more complicated. Pomponazzi discusses divine omniscience in the specific form of divine foreknowledge. Divine foreknowledge apparently precludes human freedom for if God infallibly knows all events before they occur, human deliberation about alternatives cannot affect the outcome. Any denial of this theological determinism implies that we can choose an alternative unknown to God. This impugns God's infallibility, as we would then have the power to falsify his knowledge. Here Pomponazzi draws on the Boethian-Thomistic solution which attempted to reconcile divine foreknowledge and human freedom by designating God's knowledge as 'timeless' (see Omniscience §§3-4). This timelessness is a total simultaneity in which all events are eternally present to God without succession. There is no future mode in God's eternal knowledge, for that would make divine knowledge indeterminate and possibly false. Pomponazzi qualifies this traditional view in an original way by asserting that God knows what is future to us in two ways: determinately and indeterminately. When any future event becomes past or present within the temporal flow, God knows this event as determinate. Such events are 'beyond their causes', with all their potentialities actualized so that God knows them determinately; they correspond to some settled actuality. This mode of divine foreknowledge Pomponazzi calls the future-as-present or future-as-past. Events which have not yet occurred within the temporal flow are designated the future-as-future. These events are by logic and nature unknown and unknowable. God's knowledge of them is uncertain in the sense that such an event is contingent. God's infallible foreknowledge is not destroyed by this contingent mode because he still possesses all temporal modes within his 'timeless' eternity. As God himself decides to prescind from the determinate modes of the future, he limits his future knowledge to pure contingency, thus allowing for the free operation of the human will. As unusual as this position is when compared to Pomponazzi's other teachings, it seems dictated by his recognition of the need to provide a metaphysical ground for ethics.

Pomponazzi's arguments are marked by the realization and admission of difficulty and paradox. His conclusions are attained only after the most complete elaboration of all possibilities. It is his tenacity in the face of complexity and honesty in the confrontation of difficulty that makes Pomponazzi, in Kristeller's words, 'one of the most difficult but also one of the most sympathetic and interesting of the philosophers of the Aristotelian school, of the Renaissance and of every age' (1983: 23).

See also: Aristotelianism, Renaissance §5; Cajetan §4; Ficino, M. §3

MARTIN L. PINE

List of works

Pomponazzi, P. (before 1516) 'Pietro Pomponazzi: Utrum anima sit mortalis vel immortalis' (Whether the Soul is Mortal or Immortal), ed. W. Van Dooren, Naples: Nouvelles de la République des lettres, 1989.(A fine critical edition with an analysis of *Pomponatius in Libros de Anima (Pomponazzi's commentary On the Soul)*. Set forth just before the *De immortalitate*, Pomponazzi here elaborated most major aspects of his theory which is developed more decisively in 1516.)

Pomponazzi, P. (1516) *De immortalitate animae*, trans. W.H. Hay II, *On the Immortality of the Soul*, in E. Cassirer, P.O. Kristeller and J.H. Randall, Jr (eds) *The Renaissance Philosophy of Man*, Chicago, IL: University of Chicago Press, 1948.(A reliable English translation of Pomponazzi's best known treatise. It

provoked the immortality controversy, arguing for the philosophic doctrine of mortality while accepting immortality on faith alone. Excellent footnotes by Kristeller and a useful introductory essay by Randall.)

Pomponazzi, P. (1525) *Tractatus acutissimi, utilimi et mere peripatetici (The Most Insightful, Useful and Fully Peripatetic Treatises)*, Venice.(The only collected edition; contains Pomponazzi's early works on physics, the three immortality treatises - *De immortalitate animae* (1516), *Apologia* (1518), *Defensorium Auctoris* (1519) - and the late work *De nutritione et augmentatione (On Nutrition and Growth)* (1521), which treats in part the problem of materialism in regard to the soul. With the notable exception of G. Zanier (1991), *De nutritione* has been neglected by Pomponazzi scholars.)

Pomponazzi, P. (1556) *De naturalium effectuum causis sive de incantationibus (On the Natural Effects of Causes or on Incantations)*, Basle.(An attack on miracles which finds natural causes for all miracles and then reduces all religious miracles to the effect of heavenly forces.)

Pomponazzi, P. (1567) *De fato, de libero arbitrio, praedestinatione providentia dei libri V (Five Books on Fate, Free Will, the Providence of God and Predestination)*, ed. R. Le May, *Libri quinque de fato, de libero arbitrio, de praedestinatione*, Lugano: Thesauris Mundi, 1957.(The most complex of Pomponazzi's published works. Beginning by arguing for Stoic determinism, the treatise then seeks to rescue free will, using arguments from Christian natural theology, and develops a novel theory of divine activity to rescue human freedom in light of divine foreknowledge and predestination. May provides an excellent critical edition of *De fato* with full references to all of Pomponazzi's sources, the manuscript literature and secondary sources.)

Pomponazzi, P. (1462-1525) *Studi su Pietro Pomponazzi (Studies on Pietro Pomponazzi)*, ed. B. Nardi, Florence: Felice Le Monnier, 1965.(A very important collection of articles. Nardi traces Pomponazzi's evolving position on immortality, reason and revelation, resurrection and miracles, with very full quotations from Pomponazzi's unpublished lectures and questions. The classroom comments on religious doctrines are often more critical than in the published works.)

Pomponazzi, P. (1462-1525) *Corsi inediti dell' insegnamento padovano (Unpublished Courses from the Paduan Teaching)*, ed. A. Poppi, Padua: Antenore, vol. 1, 1966; vol. 2, 1970.(Critical editions of Pomponazzi's lectures, demonstrating that Pomponazzi's basic position on immortality was developed around 1504. Other important issues (such as the *regressus* and universals) are discussed.)

References and further reading

Di Napoli, G. (1963) *L'immortalità dell'anima nel Rinascimento (The Immortality of the Soul in the Renaissance)*, Turin.(A full discussion of the doctrine of immortality among the humanists, Platonists and Aristotelians of the Renaissance. A large portion of the book is devoted to Pomponazzi and the immortality controversy.)

Fiorentino, F. (1868) *Pietro Pomponazzi, Studi storici su la suola Bolognese e Padovana del secolo XVI con molti documenti inediti (A Historical Study of the Bolognese and Paduan School of the Sixteenth Century with Many Unpublished Documents)*, Florence: Successori Le Monnier.(A classic study. Fiorentino was the first to emphasize Pomponazzi's defence of the rational doctrine of mortality as opposed to medieval arguments for the indemonstrability of immortality. Very complete on the historical background of the doctrine of immortality in ancient and medieval eras, with a clear analysis of Pomponazzi's theory of the soul, and a less successful attempt to analyse *De incantationibus* and *De fato*.)

Kristeller, P.O. (1964) *Eight Philosophers of the Italian Renaissance*, Stanford, CA: Stanford University Press. (Referred to in §2. A series of essays on the major Italian philosophers of the Renaissance notable for its clarity and careful delineation of major ideas. The chapter on Pomponazzi presents a balanced and nicely articulated view of *De immortalitate animae*. The author argues for Pomponazzi's sincerity in regard to religion.)

Kristeller, P.O. (1983) *Aristotelismo e sincretismo nel pensiero di Pomponazzi (Aristotelianism and Syncretism in the Thought of Pomponazzi)*, Padua: Antenore. (Quoted in §4. A brief, brilliant essay showing the influence of Platonic, Neoplatonic and Stoic sources on Pomponazzi. Successfully demonstrates that Pomponazzi's epistemology is more empiricist that Aristotle's.)

Pine, M.L. (1986) *Pietro Pomponazzi: Radical Philosopher of the Renaissance*, Padua: Antenore.(Referred to in §2. The only work attempting a full analysis of Pomponazzi's philosophy since Fiorentino (1868). A full discussion of the evolution of the doctrine of immortality is presented, as well as a complete analysis of the immortality controversy. *De incantationibus* and *De fato* are thoroughly examined. An attempt is made to find textual basis for the insincerity of Pomponazzi's assertions of Christian faith.)

Tavuzzi, M. (1995) 'Silvestro da Prierio and the Pomponazzi Affair', *Renaissance and Reformation*, new series, 19 (2): 47-61.(A fine discussion shedding new light on the role of Spina in the immortality controversy. Spina was clearly encouraged by Silvestro da Prierio, a high Papal official who was vehemently opposed to Pomponazzi and claimed he had been misrepresented in Pomponazzi's *Apologia*.)

Zanier, G. (1991) ‘La biologia teoretica del pensiero pomponazziano’ (Biological Theory in Pomponazzi’s Thought), in D. Facca and G. Zanier, *Filosofia, Filologia, Biologia: Itinerai dell’Aristotelismo cinquecentesco (Philosophy, Philology, Biology: Paths of 16th-Century Aristotelianism)*, Rome, 105-30.(The only work devoted to an analysis of major themes in the neglected *De nutritione et augmentatione* (1521). An interesting discussion taking full account of *De incantationibus* and *De fato*. The argument demonstrates that Pomponazzi still maintains mortality in this late work.)

Popper, Karl Raimund (1902-94)

Popper belongs to a generation of Central European émigré scholars that profoundly influenced thought in the English-speaking countries in this century. His greatest contributions are in philosophy of science and in political and social philosophy. Popper's 'falsificationism' reverses the usual view that accumulated experience leads to scientific hypotheses; rather, freely conjectured hypotheses precede, and are tested against, experience. The hypotheses that survive the testing process constitute current scientific knowledge. His general epistemology, 'critical rationalism', commends the Socratic method of posing questions and critically discussing the answers offered to them. He considers knowledge in the traditional sense of certainty, or in the modern sense of justified true belief, to be unobtainable.

After the Anschluss, *Popper was stimulated by the problem of why democracies had succumbed to totalitarianism and applied his critical rationalism to political philosophy. Since we have no infallible ways of getting or maintaining good government, Plato's question 'Who should rule?' is misdirected. To advocate the rule of the best, the wise or the just invites tyranny disguised under those principles. By contrast, a prudently constructed open society constructs institutions to ensure that any regime can be ousted without violence, no matter what higher ends it proclaims itself to be seeking. Couched in the form of extended critiques of Plato and Platonism as well as of Marx and Marxism, Popper's political philosophy has had considerable influence in post-war Europe, East and West.*

1 Life and works

Born in Vienna in 1902, the youngest child of a barrister, Karl Raimund Popper was educated at the University of Vienna, where he studied mathematics, music, psychology, physics and philosophy. He taught in secondary school between 1930 and 1936. Apprehension about Nazism persuaded him to emigrate in 1937, to become lecturer in philosophy at Canterbury University College, Christchurch, New Zealand. In January 1946 he became Reader in Logic and Scientific Method at the London School of Economics, was promoted to professor in 1949, and retired from full-time teaching in 1969. Among many honours, he was knighted in 1965, elected Fellow of the Royal Society in 1976, and made a Companion of Honour in 1982.

After a specialized start in the philosophy of science, Popper revealed himself as a philosopher of wide reach, making contributions across the spectrum from Presocratic studies to modern logic, from politics to probability, and from the mind-body problem to the interpretation of quantum theory. With all of his books in print, and translated into many languages, Popper's is one of the most discussed philosophies of the century. Yet, he insisted, his ideas are systematically misunderstood and misrepresented; this led him to devote uncommon energy to issues of interpretation and commentary on his own work.

Popper published three major works between 1935 and 1945. The first,*Logik der Forschung* (1935), his theory of science, appeared in English as *The Logic of Scientific Discovery* only in 1959. The second, *The Poverty of Historicism* (1957), first appeared in 1944-5 and extended his theory of science to history and society, severely criticizing the notion of historical laws. The third, *The Open Society and Its Enemies* (1945), is a two-volume treatise on the philosophy of history, politics and society.

Popper's other principal works consist of two collections of major papers, *Conjectures and Refutations* (1962), and *Objective Knowledge* (1972); a Library of Living Philosophers volume (1974) containing an intellectual autobiography and a set of replies to his critics, the former appearing separately as *Unended Quest* (1976); a collaboration with Sir John Eccles on a study of the mind-body problem, *The Self and Its Brain* (1977); *Die Beiden Grundprobleme der Erkenntnistheorie (The Two Basic Problems of the Theory of Knowledge)* (1979), the extant fragment of the book he was writing before *Logik der Forschung* superseded it; and the long-delayed *Postscript to the Logic of Scientific Discovery* (1982-3), much of which dates from the period 1955-7. Most of these books have seen multiple editions, involving sometimes minor and sometimes major changes. Throughout his career Popper also produced many original papers on diverse topics, and lectured all over the world. His manuscripts and correspondence fill some 450 archive cartons at the Hoover Institution, Stanford University.

2 Theory of science

Two problems structure Popper's theory of science: he calls them 'the problem of induction' and 'the problem of demarcation'. The problem of induction can be formulated: what relation holds between theoretical knowledge and experience? The problem of demarcation can be formulated: what distinguishes science from metaphysics as well as from logic and mathematics?

The received answers to these problems are: we get knowledge from experience by means of induction, that is, by inferring universal theories from accumulations of particular facts; and the inductive method demarcates science from metaphysics as well as from logic and mathematics. However, Hume showed that inductive inferences are invalid, hence the *problem* of induction: either we get knowledge from experience by invalid means (irrationalism) or we do not get any at all (scepticism); and induction collapses as a demarcation criterion (see Demarcation problem §1; Induction, epistemic issues in).

In Part I of *The Logic of Scientific Discovery*, Popper's solutions to these two problems are set out and shown to converge: knowledge results when we accept statements describing experience that contradict and hence refute our hypotheses; thus a deductive rather than an inductive relation holds between theoretical knowledge and experience. Experience teaches us by correcting our errors. Only hypotheses *falsifiable* by experience should count as scientific. There is no need for the inductive leap that Hume thought illogical but unavoidable; and the Hobson's choice between irrationalism or scepticism is avoided. To the question, 'where do hypotheses come from, if not inductively from experience?', Popper answers, like Francis Bacon, that they come from our propensity to guess (see Bacon, F. §6); in any case they cannot come from observation alone because there is no observation without hypotheses. Hypotheses are both logically and psychologically prior to observation. We are theorizing all the time in order to navigate in the world, and our encounters with negative evidence are the bumps that deliver information about the shape of reality.

The Logic of Scientific Discovery is dialectical in style, dealing with the traditional alternatives and the objections to each idea as it goes along. It is remarkable how frequently critics rediscover objections set out and answered in the book. The commonest objection is that, just as no amount of experience will conclusively verify a statement, so no amount of experience will conclusively falsify it. To answer this objection Popper points to a logical asymmetry. A universal statement cannot be derived from or verified by singular statements, no matter how many are marshalled. It can, however, be contradicted by one singular statement. The logic of falsification is the issue; conclusiveness is a red herring. Another argument, to the effect that the force of falsifying evidence can always be evaded by *ad hoc* definition or simple refusal to countenance it, Popper finds insuperable. The way to proceed, he concludes, is to entrench falsifiability in a methodology.

For Popper, a methodology is a policy decision governing action and embodied in norms or 'methodological rules'. Our decisions concern which course of action will best foster our aims. Thus falsificationism is made into a supreme rule to the effect that the 'rules of scientific procedure must be designed in such a way that they do not protect any statement in science against falsification' ([1935] 1959: 54). The rule for causality is typical of the small number offered: 'we are not to abandon the search for universal laws and for a coherent theoretical system, nor ever to give up our attempts to explain causally any kind of event we can describe' ([1935] 1959: 61). A broad epistemological ambition is revealed when Popper generalizes: 'It might indeed be said that the majority of problems of theoretical philosophy, and the most interesting ones, can be reinterpreted…as problems of method' ([1935] 1959: 56) (see Scientific method §2).

Throughout *The Logic of Scientific Discovery*, Popper defines his position by debate and contrast with logical positivist positions regarding meaning, and with two traditional views regarding science, inductivism and the conventionalism of Poincaré and Duhem (see Conventionalism §1; Logical positivism §4). It is notable that, like the logical positivists, Popper expresses unbounded respect for science. Unlike them, he grants a constructive (historical) role to metaphysics in science, seen as directly descended from the earliest Greek speculations about the nature of the world. The demarcation between science and metaphysics is thus a matter for decision, not a discovery about the nature of things. Popper's attacks on central logical positivist contentions contributed to the demise of that movement.

Popper respects conventionalism as self-contained, defensible and most likely consistent. His objection is that it risks treating obsolete or floundering science as incontrovertible truth. Yet Popper is a conventionalist in one respect: methodology. As opposed to the 'methodological naturalism' of the logical positivists, who treat the

demarcation between science and metaphysics as a difference existing in the nature of things, or rather, in the nature of language, Popper is a ‘methodological conventionalist’, proposing rules that embody choices or decisions - which are in turn governed by aims. His demarcation should be judged, Popper maintains, by whether it proves fruitful in furthering the aims of discovering new ideas and new problems.

Part II of *The Logic of Scientific Discovery* consists of chapters on theories, falsifiability, the empirical basis, testability, simplicity, probability, quantum theory and corroboration. Each is an expansion, development and defence of the ideas briefly stated in Part I and parries a particular cluster of critical objections. The chapters on probability and corroboration, for example, deal at length with the objections that the pervasive probability statements of modern science are not falsifiable, and that they measure the strength of our inductive evidence. The chapters endeavour to show how probability statements can be falsified in relevant ways, and how they are better interpreted as statements of frequencies rather than as measures of inductive support.

3 Later ideas

The English translation of *The Logic of Scientific Discovery* is palimpsestic: while translating it Popper intercalated comments, glosses, developments and corrections in new footnotes and appendices, as well as drafting a supplementary work, the three-volume *Postscript* of 1982-3. Opinion differs over whether all this is fully consistent. A case in point is ‘The Aim of Science’ section of the *Postscript*, published already in 1957, which argues that science aims at satisfactory explanations. It centres around a historical example (Galileo, Kepler, Newton), showing how each theory superseded and explained its predecessor. Satisfactory explanation, in addition to being testable, must fulfil other conditions, making it a rather stronger aim than falsifiability, one that may or may not be the same as the aim of science articulated at the end of *The Logic of Scientific Discovery*, of discovering ‘new, deeper and more general problems’.

Certainly Popper acknowledges some changes of view. Since metaphysical dispute surrounded the concept of truth, he carefully avoided using it in *The Logic of Scientific Discovery*, making do with logical relations (implication, tautology, contradiction). Later, convinced by Tarski’s work, he made free use of the concept of truth and of getting nearer to the truth (verisimilitude). Again, his criticisms of conventionalism in *The Logic of Scientific Discovery* were methodological. In later years they were also openly metaphysical, as Popper espoused a robust realism and indeterminism (1982b, 1983b).

Throughout *The Logic of Scientific Discovery* there are Darwinian metaphors - the struggle for survival among theoretical systems, natural selection, fitness to survive - although, in the last pages of the book, the view that science is an instrument of biological adaptation is rejected. This Darwinian leitmotif became a controversial issue in Popper’s later work: did evolutionary biology yield to the same methodological analysis as physics? Were the central ideas of Darwin or of the modern synthesis falsifiable? To complicate matters, Popper changed his mind on this central question, viewing Darwinism as a historical hypothesis in *Objective Knowledge*, and as an unfalsifiable near tautology in *Unended Quest*. In contrast to his earlier view, Popper also began to advocate an evolutionary epistemology, that is, an attempt to explain the very existence of a truth-seeking science within the framework of natural selection, in effect to give a biological twist to Kant’s problem, ‘How is knowledge possible?’ His second Herbert Spencer Lecture (1975) treats both endosomatic and exosomatic adaptations as forms of knowledge. Biological considerations also weigh heavily in Popper’s part of *The Self and Its Brain*. Pitting himself against the reductionist materialism of most contemporary mind-body specialists, Popper there marshalled mainly indirect arguments for an interactive pluralism.

Reflections on biology seem to have been behind a bold new metaphysical initiative of 1967-8, especially the provocatively entitled ‘Epistemology Without a Knowing Subject’ (1972: ch. 3). Distinguishing the world of physical things from the world of mental things, Popper argued that objective knowledge is located in neither, but in ‘World 3’ - the world of humanly created objective contents of thought. Such intellectual products have an objective existence: theories, problems, problem-situations, theoretical situations and critical arguments have properties and logical interrelations that lack physical or mental analogues. Stored knowledge exists even if no living person retrieves it. Critics of World 3 find some of its consequences counterintuitive: for example, it contains not only all truths, but also all falsehoods, which thus have an equally objective ‘existence’.

In the late 1940s Popper published a particularly forceful and elegant system of natural deduction that is of

considerable interest, both intrinsically and because in it he views deductive logic as the *organon* of criticism (1947a, 1947b). He reports that while he has repaired some defects in it, he never brought it to completion. His technical attentions became focused on the theory of probability, to which he had already contributed in *The Logic of Scientific Discovery*. The result was a highly abstract axiomatic system that made no explicit assumptions about any logical relations among the elements on which probability is defined, and thereby established that probability is a genuine generalization of deducibility. The system is open to many new interpretations of probability statements. A particularly important one, superseding the frequency interpretation of *The Logic of Scientific Discovery*, views probabilities as measures of the 'propensities' of states of the world to develop one way rather than another. In the *Postscript* (1982a) and *A World of Propensities*, this view was developed into a striking new metaphysics (see Probability, interpretations of §4).

4 Democracy, society and individualism

In *Unended Quest*, Popper recounts how, as a politically conscious adolescent, he flirted with Communism. He was quickly disillusioned when he judged Communist actions irresponsible in leading to the deaths of some demonstrators. (Individual autonomy, responsibility, Socratic fallibilism and the obligation to reduce suffering are the keynotes of his scattered remarks about ethics.) In the 1920s he began a critique of Marx and Marxism, first tried out in a talk of 1935, 'The Poverty of Historicism'. 'Historicism' is Popper's name for the idea that there are inexorable laws of historical development: the demand that if natural science can predict eclipses then social science ought to be able to predict political revolutions. In a highly systematic way, Popper set out to show how both those who think the social sciences are not at all like the natural sciences (the 'anti-naturalists') and those who think the social sciences very like the natural sciences (the 'naturalists'), share the aim of predicting history. Both recommend the methodology of historicism, which he sees as impoverished and inclined to treat societies as wholes responding to the pressures of inchoate social forces (see Historicism).

Instead, Popper recommends 'methodological individualism': rules to the effect that the behaviour and actions of collectives should be explained by the behaviour of human individuals acting appropriately to the logic of their social situation as best they can and as best they see it. Not only was the alternative unfruitful, he argued, but the best social explanations of Plato and Marx were individualist. What look like holistic phenomena are to be explained as the 'unintended consequences' of such individual actions reverberating through the social set-up (see Holism and individualism in history and social science §1).

According to methodological individualism, social theories are tested not by historical predictions, which are little more than prophecies, Popper argues, but by attempts to invent institutions that correct social faults by social engineering. Man-made social institutions are hypotheses in action, he says. If we are to refute these hypotheses we need to avoid complicating matters with large-scale experiments, or too many at once, otherwise assessment will be impossible. For we must also reckon with the interference factor of 'the Oedipus effect', that is, the way in which a prediction about the future becomes an altering factor in the situation as human beings are aware of it, thus 'interfering' with the outcome.

The Open Society and Its Enemies was a 'truly unintended consequence' of an attempt to expand aspects of *The Poverty of Historicism* to satisfy puzzled friends. When it grew too large, Popper made it a separate work. On publication in 1945 it elevated him from academic obscurity to academic fame. He became a controversial and well-known public intellectual.

Described modestly as a 'critical introduction to the philosophy of politics and of history', *The Open Society* had become, in the seven years of its gestation, a major treatise on the intellectual and social ills of the time, offering an explanation of how totalitarianism had gained intellectual respectability and how purging post-war society of it would involve rethinking politics, education and social morality. Its title refers to two ideal-types used throughout. A closed society is one which takes a magical or tabooistic attitude to tradition and custom, which does not differentiate between nature and convention. An open society marks that difference and confronts its members with personal decisions and the opportunity to reflect rationally on them. Heraclitus, Aristotle and Hegel are briefly discussed, but the book's two intellectual anti-heroes are Plato in volume 1 and Marx in volume 2.

The volume entitled 'The Spell of Plato' answers two questions: first, why Plato espoused totalitarian ideas and second, why students of Plato have whitewashed that fact and beautified him. The answer to the first question, a

deeply sympathetic piece of writing, sketches a portrait of the young Plato contemplating with dismay the closed world of tribal Athens giving way to a more liberal and open society, with loss of social privilege and chaos arriving hand in hand. Popper insists on the brilliance of Plato's sociological analysis of the causes of change and of his proposals to arrest it and staunch the deterioration it brings. In answer to the second, Popper courts controversy by suggesting that Plato's intellectual followers, flattered by the role offered them, engaged in a long-running *trahison des clercs* by presenting as liberal and enlightened the doctrine of the philosopher king.

To expose the commitment of Plato and Platonists to totalitarianism, Popper had to clarify our ideals of a liberal and democratic social order and show how Plato indulges in persuasive definitions, while attempting to show that his totalitarian *Republic* is 'just'. (Popper also mounts a general attack on the idea that philosophy should seek out the essence of universal words such as justice, democracy and tyranny. He argues that natural science uses the methodology of nominalism, not of essentialism, and social science and philosophy would do well to follow suit.) Mindful of Plato's distaste for majority rule, as rule of the mob or rule of the worst, Popper carefully discusses tyranny, and concludes that the problem is not the question of what is popular, for certain kinds of tyrants are very popular and could easily be elected. So an open and liberal society is not to be identified with a popularly elected government. No more is it a matter of what is just, good or best, for none of these offers insurance against tyranny in their name. In line with his theory of science, and of knowledge generally, he proposes a *via negativa*. The issue is not what regime we want, but what to do about ones we do not want. The problem with tyranny is that the citizens have no peaceful way in which to rid themselves of it, should they want to. Popper proposes a now famous and generally endorsed criterion for democracy as that political system which permits the citizens to rid themselves of an unwanted government without the need to resort to violence. He exposes Plato's question, 'Who should rule?', and all similar discussions of sovereignty as subject to paradoxes because the question permits an inconsistency to develop between the statement designating the ruler (for example, the best or wisest should rule) and what the ruler commands (for example, the best or wisest may then tell us: obey the majority, or the powerful). Popper noticed that the question carries the authoritarian implication that whoever is so named is *entitled* to rule. He replaces them with the practical question 'how can we rid ourselves of bad governments without violence?', with its implication that rulers are on permanent parole. Popper's is a fundamentally pessimistic view that all governments are to one extent or the other incompetent and potentially criminal in their misbehaviour, and that only a political system which allows them to govern at the sufferance of citizens who can withdraw their support readily is one with more or less effective checks against abuse. Even so, the fallibility of our institutional hypotheses enjoin upon us an eternal vigilance.

The second volume, 'The High Tide of Prophecy, Hegel and Marx', argues that the prophetic tendency in Heraclitus and Plato produced in Hegel a damaging incoherence and charlatanry, and in Marx a project for the scientific study of society that, despite noble emancipating aims, foundered, especially among the followers, on the confusion of prediction with unscientific historical prophecy and hence on a fundamental misconstruction of scientific method. The chapters on Marx are among the most penetrating commentaries ever written on him, and are both sympathetic in their appreciation and unremitting in their criticism. Although Popper clearly regards Plato as the deeper thinker, he argues that Marx has much to teach us about how moral and emancipatory impulses can go awry. Marxists and the radical experimenters of the Soviet Union are judged harshly, as are all forms of nationalism.

Appearing in 1945 just after the end of the war, a war that had forged an alliance with the Soviet Union, the book antagonized many powerful intellectual interest groups. Platonists were taken aback to be accused of being apologists for totalitarian tendencies in Plato (even though Popper was not the first to point these out) and Marxists were equally affronted. The aftermath was strange. Although Popper was at first widely read and indignantly denounced, it later became bad form among Marxists and classicists to mention him by name. Yet, to an extraordinary extent, in the following decades his work set the agenda for apologetic Platonists, Hegelians and Marxists. In many cases a book or article makes most sense when seen as covertly engaged in trying to confute some point Popper made in *The Open Society*.

As a teacher of philosophy, Popper displayed ambivalence: philosophical problems emerge from science (1962: ch. 2), so the best preparation was an education in a first-order subject, preferably scientific. In the classroom, he had the charisma of one possessed by intellectual problems, thinking about them all the time. In lecture or seminar he could be intellectually fierce and confrontational, in personal encounter sweet and encouraging. He regularly

displayed an astonishingly quick intuitive grasp of the logic of any position presented to him, even from the most meagre of clues, and an eagerness to strengthen and elaborate on it before setting about criticizing it. He thus exemplified the values he advocated: intellectual seriousness, personal responsibility and disinterestedness, that is, doing justice to ideas regardless of their temporary embodiment. The failure of much critical commentary to meet these standards lies behind his complaint of misrepresentation. But there are also other reasons. If Popper is correct, not only is much in the traditional way of doing philosophy misdirected, but even the questions are wrongly put. Any attempt to map Popper's ideas into traditionally oriented discussions risks misrepresentation. The frequent practice of reconstructing Popper's philosophy timelessly, plucking materials from works published as far apart as fifty years, flies in the face of his emphasis on the structuring role of problems and problem-situations in all intellectual activity, particularly inquiry. To do justice to the originality and creativity of his work, scholarship needs in the first instance to respect its intellectual context of production.

See also: Carnap, R.; Discovery, logic of; Explanation in history and social science; Fallibilism; Inductive inference; Lakatos, I.; Liberalism; Meaning and verification; Natural deduction, tableau and sequent systems; Scientific realism and antirealism; Vienna Circle

IAN C. JARVIE

List of works

Popper, K.R. (1935) *Logik der Forschung (The Logic of Research)*, Vienna: Springer; trans. *The Logic of Scientific Discovery*, London: Hutchinson, 1959.(Crisp and incisive statement of his classic ideas on philosophy of science and of epistemology. The English translation contains additional material, as explained in §3.)

Popper, K.R. (1945) *The Open Society and Its Enemies*, London: Routledge.(A critical introduction to the philosophy of history and of politics that takes Plato, Hegel and Marx to task for their anti-democratic and anti-critical tendencies.)

Popper, K.R. (1947a) 'New Foundations for Logic', *Mind* 56: 193-235, with corrections in (1948) 57: 69-70.(A system of natural deduction.)

Popper, K.R. (1947b) 'Logic Without Assumptions', *Proceedings of the Aristotelian Society*, 48: 251-92. (Investigates the notion of deducibility and the foundations of logic in language.)

Popper, K.R. (1957) *The Poverty of Historicism*, London: Routledge.(States and rebuts the arguments for belief in inexorable laws of human destiny.)

Popper, K.R. (1962) *Conjectures and Refutations*, London: Routledge.(A wide-ranging collection of essays from the period 1940-62, including several that have become classics.)

Popper, K.R. (1972) *Objective Knowledge*, Oxford: Clarendon Press.(A collection of essays from 1948-72, many being extensions of and alterations to his original philosophy of science and epistemology.)

Popper, K.R. (1975) 'The Rationality of Scientific Revolutions', in R. Harré (ed.) *Problems of Scientific Revolutions*, Oxford: Clarendon Press, 72-101.(Popper's answer to Kuhn's irrationalist view of scientific revolutions.)

Popper, K.R. (1976) *Unended Quest*, London: Fontana.(Intellectual autobiography; approachable and written with considerable verve.)

Popper, K.R. and Eccles, J.C. (1977) *The Self and Its Brain*, Berlin: Springer.(Joins a fellow Nobelist in proposing an evolutionist and interactionist model of the mind-body problem.)

Popper, K.R. (1979) *Die Beiden Grundprobleme der Erkenntnistheorie (The Two Basic Problems of the Theory of Knowledge)*, Tübingen: Mohr.(All that survives of the book Popper was working on when he turned instead to *The Logic of Scientific Discovery*.)

Popper, K.R. (1982a) *Quantum Theory and the Schism in Physics*, London: Hutchinson. (Reflections on problems of the interpretation of quantum mechanics; part II of the *Postscript*.)

Popper, K.R. (1982b) *The Open Universe*, London: Hutchinson.(Statement of Popper's metaphysical indeterminism; part III of the *Postscript*.)

Popper, K.R. (1983a) *Realism and the Aim of Science*, London: Hutchinson.(Long commentary on many of the issues originally discussed in *The Logic of Scientific Discovery* and their reception; part I of the *Postscript*.)

Popper, K.R. and Miller, D. (1983b) 'A Proof of the Impossibility of Inductive Probability', *Nature* 302: 687-8. (Proof that evidence, far from supporting a hypothesis, undermines the statement of the evidence and hence the hypothesis itself.)

Popper, K.R. (1987) 'Why Probabilistic Support is Not Inductive', *Philosophical Transactions of the Royal*

Society of London, A321: 569-91.(Restatement and elaboration of Popper and Miller (1983).)
Popper, K.R. (1990) *A World of Propensities*, Bristol: Thoemmes.(The propensity interpretation of probability statements developed as a full-blown metaphysics.)
Popper, K.R. (1992) *In Search of a Better World*, London: Routledge.(Speeches and address to general audiences.)
Popper, K.R. (1994a) *Knowledge and the Body-Mind Problem*, London: Routledge.(Essays from his last thirty years on the mind-body problem and other matters.)
Popper, K.R. (1994b) *The Myth of the Framework*, London: Routledge.(Essays from his last thirty years on the philosophy of physical and social science.)
Schilpp, P.A. (ed.) (1974) *The Philosophy of Karl Popper*, La Salle, IL: Open Court. (Two volumes, containing Popper's intellectual autobiography, 33 critical essays and 'Replies to My Critics', as well as a complete bibliography up 1973.)
Adorno, T.W. *et al*. (1976) *The Positivist Dispute in German Sociology*, London: Heinemann.(Contains Popper's 'The Logic of the Social Sciences' and 'Revolution or Reform?'.)

References and further reading

Ackermann, R. (1976) *The Philosophy of Karl Popper*, Amherst, MA: University of Massachusetts Press.(A review and positive evaluation of Popper's philosophy of science.)
Ackermann, R. (1993) *A Philosopher's Apprentice*, Amsterdam: Rodopi.(An idiosyncratic reminiscence of studying with Popper.)
Albert, H. (1985) *Treatise on Critical Reason*, Princeton, NJ: Princeton University Press.(Critical rationalism as expounded by a leading German Popperian.)
Andersson, G. (1994) *Kuhn's, Lakatos's, and Feyerabend's Criticism of Critical Rationalism*, Leiden: E.J. Brill. (Critical rationalism defended against its major critics.)
Bartley, W.W., III (1962) *The Retreat to Commitment*, New York: Knopf.(An influential attempt to improve upon Popper's critical rationalism.)
Bartley, W.W., III (1976-82) 'Critical Study: The Philosophy of Karl Popper: Parts I, II and III', *Philosophia* 6: 463-94; 7: 676-716; 12: 121-221.(A sympathetic critical essay on the Schilpp volume (1974).)
Berkson, W. and Wettersten, J. (1984) *Learning from Error*, La Salle, IL: Open Court.(An investigation into the origin of Popper's ideas in his early psychological researches.)
Bunge, M. (ed.) (1964) *The Critical Approach to Science and Philosophy*, New York: Free Press.(The first collection of essays discussing Popper's ideas.)
Burke, T.E. (1983) *The Philosophy of Popper*, Manchester: Manchester University Press.(Critical commentary on a few selected topics. Popper is said to have set the agenda in a number of fields. Wittgenstein gets the worse of some comparative discussions.)
Gellner, E. (1985) 'Positivism Against Hegelianism', in *Relativism and the Social Sciences*, Cambridge University Press.(A robust critical review of the metaphysical issues between Popper and his idealist critics.)
Hacohen, M.H. (1993) 'The Making of the Open Society: Karl Popper, Philosophy and Politics in Interwar Vienna', unpublished Ph.D. dissertation, Columbia University. (A lengthy attempt to relate Popper's work to the Viennese context.)
Hattiangadi, J.N. (1978-9) 'The Structure of Problems, Parts I and II', in *Philosophy of the Social Sciences*, vol. 8, 345-65; vol. 9, 49-76.(Argues that solving problems is a sufficient aim for science and that the history of science is lines of debate.)
Jarvie, I.C. (1972) *Concepts and Society*, London: Routledge.(A Popperian philosophy of the social.)
Jarvie, I.C. and Shearmur, J. (eds) (1996-7) *Philosophy of the Social Sciences*, special issues, 26 (4) and 27 (1). (In celebration of the 50th anniversary of the publication of *The Open Society and its Enemies* containing eight papers on the origins, impact and contemporary relevance of Popper's great book.)
Johansson, I. (1975) *A Critique of Karl Popper's Methodology*, Stockholm: Akademiforlaget.(A vigorous assault on Popper's methodology.)
Leblanc, H. (1989) 'Popper's Formal Contributions to Probability Theory', in *Perspectives on Psychologism*, The Hague: Nijhoff, 341-67.(An important evaluation of Popper's technical contribution.)
Levinson, P. (ed.) (1982) *In Pursuit of Truth*, Atlantic Highlands, NJ: Humanities.(A second collective volume in Popper's honour.)
Magee, B. (1973) *Popper*, London: Fontana/Collins.(The first comprehensive exposition for the general reader.)
Miller, D. (1994) *Critical Rationalism: Restatement and Defence*, La Salle, IL: Open Court.(Sharp and forceful, including discussion of critical issues surrounding truth and probability.)

Munz, P. (1985) *Our Knowledge of the Growth of Knowledge*, London: Routledge. (Wittgenstein's theory of knowledge is criticized as relativist, Popper's evolutionary view is defended.)
O'Hear, A. (1980) *Karl Popper*, London: Routledge.(Critical account that treats Popper timelessly and within the received framework of academic philosophy and, unsurprisingly, vindicates Wittgenstein.)
O'Hear, A. (ed.) (1995) *Karl Popper: Philosophy and Problems*, Cambridge: Cambridge University Press.(Fifteen Royal Institute of Philosophy lectures that range widely over Popper's work.)
Pralong, S. and Jarvie, I.C. (eds) (1997) *Popper's Open Society After 50 Years*, London: Routledge.(The proceedings of a Prague conference originally organized by Ernest Gellner, to assess Popper's work in the context of the opening up of former Iron Curtain countries.)
Radnitzky, G. (1968) *Contemporary Schools of Metascience*, Goteborg: Akademiforlaget.(A survey of the field from a Popperian 'criticist' perspective.)
Schroeder-Heister, P. (1984) 'Popper's Theory of Deductive Inference and the Concept of a Logical Constant', *History and Philosophy of Logic* 5: 79-110. (Reconstructs Popper's ideas on logic so as to take account of criticism.)
Simkin, C. (1993) *Popper's Views on Natural and Social Science*, Leiden: Brill.(An accurate and mildly critical account by a colleague from the New Zealand days.)
Stove, D. (1981) *Popper and After. Four Modern Irrationalists*, Oxford: Pergamon. (A scathing, if less than persuasive, attack.)
Wettersten, J. (1992) *The Roots of Critical Rationalism*, Amsterdam: Rodopi.(Argues that tensions in Popper's work and among his followers go back to the philosophical and psychological traditions he was attempting to reconcile.)
Williams, D.E. (1989) *Truth, Hope and Power: The Thought of Karl Popper*, Toronto, Ont.: University of Toronto Press.(An ambitious commentary that perhaps overreaches its grasp.)
Wisdom, J.O. (1952) *Foundations of Inference in Natural Science*, London: Methuen. (One of the earliest and best expositions of Popper's ideas in English.)

Population and ethics

Ethical concern with population policies and with the issue of optimal population size is, generally speaking, a modern phenomenon. Although the first divine injunction of the Bible is 'be fertile and multiply', systematic theoretical interest in the normative aspects of demography has become associated largely with recent developments which have provided humanity with unprecedented control over population size, mainly through medical and economic means. Once the determination of the number of people in the world is no longer a natural given fact, but rather a matter of individual or social choice, it becomes subject to moral evaluation.

However, the extension of traditional ethical principles to issues of population policies is bedevilled by paradoxes. The principle of utility, the ideal of self-perfection, the idea of a contract as a basis for political legitimacy and social justice, the notion of natural or human rights, and the principle of respect for persons - all these presuppose the existence of human beings whose interests, welfare, rights and dignity are to be protected and promoted. But population policies deal with the creation of people and the decision concerning their number. They relate to the creation of the very conditions for the application of ethical principles.

1 'The population bomb'

The number of people now living is larger than the number who have ever lived in all past generations of humanity taken together. This striking fact of absolute numbers stands alongside a rate of growth which now doubles the population of the world every thirty-five years. Sharp awareness of the threat that these facts and trends present goes back to the early nineteenth century, primarily to the work of Thomas Malthus, who argued that the number of people multiplies in geometric progression while the resources which support their life can be expected to grow at most in arithmetic progression. The threat of overpopulation arising from this gap in rates of growth can be checked either by factors like death from starvation, illness and war, or by preventive measures such as controlling the number of births. Following Malthus, it is commonly believed that there is an ethical asymmetry between the two kinds of checks: the premature death of living people is morally worse than the non-creation of new people. Thus, a rise in life expectancy is generally considered a blessing, while an increase in births is treated as a menace. However, balancing world population is particularly problematic, since overpopulation is primarily due to the decrease in death rates rather than the increase in birth rates. Consequently, even if very restrictive population policies were implemented (such as zero population growth), many decades would pass before the number of births and the number of deaths balanced one other.

Concern with the danger of overpopulation is historically and analytically associated with two major normative issues, themselves interdependent: the just distribution of welfare across generations (usually referred to as the problem of savings); and responsibility for the environment (conservation). Thus, since the 1920s economists have attempted to fix optimum population size in economic terms, such as the highest income per capita across generations, and theories of justice (particularly since Rawls) have tackled the population question through the principles of justice applied intergenerationally (see Future generations, obligations to; Justice). Meanwhile, environmentalists have argued over the last few decades that the number of people in the world should be limited in terms of the carrying capacity of the planet and respect for the integrity of the natural environment, either for the sake of future people ('shallow ecology') or for its own sake ('deep ecology') (see Environmental ethics). However, one can take the opposite approach and rather than subordinate the issue of population to economic or ecological considerations, treat the existence and number of people as the foundation for both economic justice and environmental policies. Thus, for instance, one may take the existence of a growing population as a *given* fact which should dictate a new approach to the natural environment in which the wilderness will have a radically weakened status.

The metaphors of 'the population bomb' and 'population explosion' typically convey a morally negative evaluation of the increase in the number of human beings. However, beyond the empirical debate between the 'alarmist', neo-Malthusian view and its opponents who see no danger in population growth as such, a critical philosophical analysis of the issue raises the following questions. Is the size of world population an ethical issue at all? If it is, what are the normative criteria for determining the 'right' (or optimal) size? If such criteria can be formulated and justified, what are the ethical constraints in implementing suitable policies for the attainment of

this size?

2 Optimum population size

The question of optimal population size touches upon the deepest problems concerning the nature of value: is value ascribed impersonally to 'the world', or only to actual human beings whose very existence is presupposed by any evaluation? The first view (impersonalism) treats the population issue as a function of a value or an end considered in abstraction. The ideal number of people is that which would promote that end to the maximal degree. This could be average or total utility, the foundation of a Kantian 'kingdom of ends', or the establishment of a Marxian classless society. The second view (commonly called 'person-affecting') holds that moral evaluation applies only to existing human beings and hence cannot serve to assess the value of any particular number of human beings, indeed not even of the very existence of humanity. According to this view, 'good' is always 'good *for*', and consequently one cannot say that to be born (or not to be born) is a morally good (or bad) thing. Increasing or reducing the number of future people cannot in the person-affecting view be considered morally desirable (or obligatory) since the interests of no particular people are promoted or rights of particular persons respected either by bringing them into the world or by preventing them from coming into existence.

The two views face difficult problems, which have been extensively discussed by Derek Parfit (1984). Take first utilitarianism, which has been the most popular version of impersonalism in the context of population questions (see Utilitarianism). Total utilitarianism leads to the so-called 'repugnant conclusion', according to which the world would be overcrowded with people whose extremely low quality of life is offset by their huge number, thus achieving the highest value of total utility. Average utility fares no better, since it recommends a policy which on the one hand would make it obligatory to bring a new child into the world whose happiness would be a little above the current average, and on the other hand would prohibit the addition of extra people whose standard of living would be even slightly below that enjoyed by present people. Negative utilitarianism, which aims at the overall reduction of pain rather than the promotion of happiness, could lead to the absurd ideal of total abstention from procreation.

Turning from these impersonal criteria to the person-affecting view, we are confronted by the problem of identity. We cannot say that population policies serve the interests or the welfare of future people since they themselves directly and indirectly affect the identity of those people. Thus, if we decide on a restrictive population policy, for example, by introducing family planning programmes, the people who are actually going to be born are not the same as those who would have been born had we abstained from interfering in people's reproductive choices. The idea of an optimal population size is therefore logically incoherent. On the one hand, there is the indeterminacy of an ideal based on the combination of two variables; namely, 'maximum happiness to maximum people', which has been shown traditionally to have no unique 'solution', as it can be achieved equally by raising the amount of happiness or by raising the number of people. On the other hand, there is the incommensurability of the respective value of two populations consisting of different individuals; namely, the problem of identity, in which case the person-affecting view makes it impossible to compare the value of the lives of alternative sets of people.

Other philosophers have noted the special difficulties incurred in applying contract theory to the issue of population. These derive from the problem of representation in Rawls' 'original position': who takes part in the initial contract? All possible people (but can these be even theoretically identified)? All actual people (but their number is exactly the issue to be decided by the contract)? Representatives of each generation, or rather a representative of one particular generation, who does not know which generation it is? Contractors negotiating their coming into being, their very existence, is an idea which leads to interesting paradoxes that have been discussed primarily by John Rawls (1971) and his critics (see Contractarianism §7).

Our intuitions concerning optimal population size are also indeterminate. Beyond some clear upper and lower limits of over- and underpopulation, we do not have any firm beliefs as to the right balance between the quantity and quality of human lives. Furthermore, our preferences concerning numbers are typically 'adaptive', that is to say they are themselves formed and affected by the size of population to which we belong. We thus tend to project our own preferences on our descendants, assuming (without justification) that they would not want to live in a more crowded environment than we did. From this welfarist perspective, optimal population size is a myth and the population issue lies in principle beyond the scope of ethics. Yet, we can be sure that future generations are going to have some basic needs and wants which will be hard to fulfil if the rate of population growth remains

unchecked. This in most views provides a moral justification for control measures.

3 Means of implementation

Even if there are criteria for fixing a desired population size, all philosophers agree that there are constraints on their implementation. One fundamental problem belongs to distributive justice: how should the burden of achieving the demographic goal be distributed, both within the present generation and between generations? Unlike the abstract issue of optimal population size, which deals with future (possible) people, the implementation of population policies affects actual people and their interests. Thus, for instance, should the burden of curtailing growth fall on all classes equally (as in China where each family is allowed to have no more than one child) or proportionally to the actual (or desired) number of children in a family? Should poor families be compensated when reproductive restraint means that they have fewer working members (an acute problem in the Third World)? Should religious beliefs play a role in deciding differential criteria in the distribution of the burden of birth control? Furthermore, may rich nations make economic aid conditional upon the adoption of family planning programmes? These questions of distributive justice can be tackled in terms of the various theories of justice. But when raised on the intergenerational level they become elusive. The goal of a certain population size can be achieved either by an aggressive policy within, for example, two generations or by a milder one within four. The problem is that full equality in distribution of the burden would require a spread over an indefinite number of generations, which is absurd.

Population policies usually involve a gap between private rights and public interest, and hence require politically enforced coordination (so as to avoid free riders). This is Garrett Hardin’s view of the *Tragedy of the Commons* (1968), in which overgrazing can be prevented (for the benefit of everybody) only by state interference. But procreation is usually considered a private matter in which the law should not interfere. Positive means, like enforcing sterilization or legally limiting the number of children, are usually considered immoral. But indirect incentives (positive or negative) through taxation, free birth control or education are acceptable. There is also wide agreement today that a rise in the standard of living, as well as an improvement in the status of women in society and in the family, are both desirable goals in themselves and effective means of restraining population growth.

DAVID HEYD

References and further reading

Barry, B. (1977) ‘Justice Between Generations’, in P. Hacker and J. Raz (eds) *Law, Morality, and Society*, Oxford: Clarendon Press.(Deals with the problems of applying traditional theories of justice in the intergenerational sphere.)

Bayles, M. (1980) *Morality and Population Policy*, Tuscaloosa, AL: University of Alabama Press.(A comprehensive discussion of the ethics of population policies and their implementation.)

Dasgupta, P. (1987) ‘The Ethical Foundation of Population Policies’, in D. Gale Johnson and R.D. Lee (eds) *Population Growth and Economic Development*, Madison, WI: University of Wisconsin Press.(A sophisticated analysis by an economist of the issue of optimum population size.)

Hardin, G. (1968) ‘The Tragedy of the Commons’, *Science* 162 (December): 1243-8.(A classic presentation of the problem of coordination in population policies.)

Hartmann, B. (1987) *Reproductive Rights and Wrongs: The Global Politics of Population Control and Contraceptive Choice*, New York: Harper & Row.(A feminist point of view of the subject.)

Heyd, D. (1992) *Genethics: Moral Issues in the Creation of People*, Berkeley, CA: University of California Press. (A radically person-affecting view; see in particular chapter 5.)

Narveson, J. (1967) ‘Utilitarianism and New Generations’, *Mind* 76 (1): 62-72.(The first modern discussion of the problems of applying utilitarian principles to population issues.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press, part IV.(Extensive discussion of the paradoxes of population; includes a critique of the person-affecting view.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA.: Harvard University Press, §§44-5.(The application of the theory of justice to future generations, particularly in the issue of savings.)

Sikora, R.I. and Barry, B. (eds) (1979) *Obligations to Future Generations*, Philadelphia, PA: Temple University Press.(The best collection of essays on the subject.)

Singer, P. (1976) ‘A Utilitarian Population Principle’, in M. Bayles (ed.) *Ethics and Population*, Cambridge, MA: Schenkman.(A sophisticated person-affecting view.)

Pornography

There are three main questions about pornography. (1) How is pornography to be defined? Some definitions include the contention that it is morally wrong, while others define it neutrally in terms of its content and function. (2) Why is it wrong? Some accounts see the moral wrong of pornography in its tendency to corrupt individuals or to have detrimental effects on the morality of society; other accounts declare pornography to be objectionable only in so far as it causes physical harm to those involved in its production, or offence to unwilling observers. (3) Should pornography be restricted by law? Controversy here centres around whether the law should be used to discourage immorality, and whether the importance of free speech and individual autonomy are such as to rule out legislating against pornography. Here, the pornography debate raises very general questions about law and about autonomy in liberal societies.

1 What is pornography?

The question appears simple, but it admits of many conflicting answers. Etymologically, ‘pornography’ means the depiction of women as prostitutes, and although this may now appear an inappropriate definition, some feminists have appealed to it as part of an argument for the legal restriction of pornography. Thus Andrea Dworkin and Catharine MacKinnon (1988) have proposed that pornography be defined as ‘the graphic, sexually explicit subordination of women whether in pictures or in words’, and they go on to include within the definition representations in which women are presented dehumanized as sexual objects or commodities, or as whores by nature, or as experiencing sexual pleasure in being raped. The definition is contentious for a number of reasons. First, it seems unduly restrictive, since it excludes, for example, gay pornography and child pornography. Moreover, the definition is value-laden. In declaring that pornography definitionally involves the subordination of women, Dworkin and MacKinnon assume that it is morally wrong but this needs to be shown, not simply stated. For these reasons, some writers have proposed a ‘neutral’ definition of pornography, which aims to identify pornographic material by reference to its degree of sexual explicitness and to leave open the question whether such explicitness is morally objectionable or an appropriate object of legal restriction.

An example of this second approach is the 1979 *Report of the Committee on Obscenity and Film Censorship in England and Wales* (‘The Williams Report’). The Report uses the term ‘pornography’ to refer to ‘a book, verse, painting, photograph, film, or some such thing - what may in general be called a *representation*’, and proposes that ‘a pornographic representation is one that combines two features: it has a certain function or intention, to arouse its audience sexually, and also a certain content, explicit representations of sexual material (organs, postures, activity etc.). A work has to have both this function and this content to be a piece of pornography’ (8.2). This definition simply refers to the function and content of representations and leaves questions of morality and of legal restriction to be addressed separately.

However, this definition has also been criticized on the grounds that it does not accurately reflect current usage. Thus, the 1986 Meese Commission in the United States declared that ‘the appellation "pornography" is undoubtedly pejorative’, and objected to the Williams definition because it did not reflect this fact. Similarly the US Johnson Commission (1970) declined to use the term at all in its legal recommendations, arguing that it appeared to have ‘no legal significance and most often denoted subjective disapproval of material rather than its content or effect’.

There is, therefore, considerable controversy surrounding the definition of pornography: in ordinary usage it appears to be a pejorative term, which expresses subjective disapproval of sexually explicit material; but it is argued that to be useful in a legal context it must go beyond the merely subjective and make reference to content and function in a way which is, as far as possible, morally neutral.

2 What is wrong with pornography?

There is also considerable dispute about whether and why pornography is morally wrong, and about whether the nature of any wrongness is sufficient to justify legal restriction or prohibition. Some argue that pornography corrupts those who consume it. Indeed this claim is built into the current British law governing obscenity, which defines material as ‘obscene’ when it has a ‘tendency to deprave and corrupt’ those who are likely to read, see or

hear it. A similar understanding of the wrong of pornography is implicit in the Dworkin-MacKinnon Ordinance, which construes pornography as a purveyor of sexist images of women and a corrupting influence on society. For them, pornography acts as a kind of moral poison in the societal water supply, and can serve to legitimize contemptuous attitudes towards women, and possibly even sexual violence against them.

In both cases the charge against obscene material generally, and pornographic material in particular, is that it is morally damaging, either to the individuals who consume it or to society in general.

Others, however, eschew reference to moral harm, either because they are sceptical about the very notion, or because they feel that it should not be appealed to in a legal context. Thus, the Williams Report deems pornography to be wrong only in so far as its production involves the infliction of physical harm, or if it causes offence to unwitting and unwilling observers. Moral wrongness is not, in itself, any part of the case against pornography. Indeed, as was noted above, the Williams Report defines pornography simply in terms of sexual explicitness and makes no reference to its moral status.

3 What justifies legislating against pornography?

Not surprisingly, the different definitions of pornography, coupled with different accounts of why it may be objectionable, yield different proposals for legal action. Those who believe that pornography may be restricted because it is morally damaging, either to the consumer or to society more generally, are often accused of paternalism (see Paternalism). This is a serious charge because an important guiding principle of liberal democratic societies is that the state should be neutral between competing conceptions of the good. It should not attempt to impose on its citizens any particular, favoured understanding of the best way to live, nor should it prohibit or discourage actions simply on the grounds that they are believed to be morally undesirable (see Liberalism).

Connectedly, the principle of free speech is a very important one in liberal societies, and some have argued that restricting or suppressing pornography offends against that principle (see Freedom of speech). The importance of free speech is often explained by reference to individual autonomy: people can best become autonomous in societies where they are able to develop their own views and opinions, and that requires that they have access to a wide range of opinions and beliefs from which they themselves may choose (see Autonomy, ethical).

However, both the argument from neutrality and the argument from free speech have been questioned, most notably by feminists who claim that the free availability of pornography is itself something which undermines the neutrality of the state and denies freedom of speech to certain groups. Thus, it has been suggested that pornography is akin to racist speech: just as racist speech may undermine the self-respect of black people or minority groups, so pornography undermines the self-respect of women and makes it difficult for them to develop autonomy.

This feminist critique of pornography also questions the individualist assumptions of liberal political philosophy. On this understanding, the justification for legislating against pornography is not the moral harm or offence which it causes to specific individuals, but rather the damage which is done to the self-esteem of women as a group.

See also: Law, limits of; Liberalism; Sexuality, philosophy of

SUSAN MENDUS

References and further reading

Dworkin, A. (1981) *Pornography: Men Possessing Women*, New York: Putnam.(The most influential statement of important feminist objections to pornography.)

Dworkin, A. and MacKinnon, C. (1988) *Pornography and Civil Rights: A New Day for Women's Equality*, Minneapolis, MN: Organizing Against Pornography.(Contains a statement of the Model Ordinance governing pornography presented to the State of Minneapolis by Dworkin and MacKinnon.)

Dworkin, R. (1985) 'Do We Have a Right to Pornography?', in *A Matter of Principle*, Cambridge, MA: Harvard University Press.(A clearly written philosophical defence of the liberal commitment to free speech in the specific context of the pornography debate.)

Hawkins, G. and Zimring, F.E. (1988) *Pornography in a Free Society*, Cambridge: Cambridge University Press. (A comparative anaysis of three reports on pornography - the American Meese and Johnson Reports, and the British Williams Report.)

MacKinnon, C.A. (1994) *Only Words*, London: Harper Collins.(One of the most recent, and most controversial, statements of feminist objections to pornography.)

Williams, B. *et al*. (1979) *Report of the Committee on Obscenity and Film Censorship* ('The Williams Report'), Cmnd 7772, London: Her Majesty's Stationery Office; abridged version repr. as B. Williams (ed.) *Obscenity and Film Censorship*, Cambridge: Cambridge University Press, 1981.(A British Report which appeals to John Stuart Mill's harm principle in its recommendations for changes to the laws governing obscenity and film censorship.)

Porphyry (*c*.233-309 AD)

The late ancient philosopher Porphyry was one of the founders of Neoplatonism. He edited the teachings of Plotinus into the form in which they are now known, clarified them with insights of his own and established them in the thought of his time. But, in reaction to Plotinus, he also advanced the cause of Aristotle's philosophical logic. Indeed, Porphyry is responsible for the resurgence of interest in Aristotle, which continued to the Middle Ages and beyond. Because of Porphyry, later Greek philosophy recovered both its Platonic and its Aristotelian roots, and Neoplatonism aimed to combine inspired thought with academic precision.

He was a scholar of great learning, with interests ranging from literary criticism and history to religion. An example is his defence of vegetarianism, which anticipated the modern debate on ecological preservation. Humans and animals belong to the same family. Seeking to preserve life is a matter of extending philanthropy and respect to all living species, which are our natural siblings. Ideally we ought to display 'harmlessness' even towards plants, except that our bodies, being composite and mortal, need to consume something else for food. Thus we should be ever conscious of the destructive effect that our eating habits and consumerism have on the creation of which we are part, and should try to keep to a simple lifestyle.

Porphyry's attention to logic, metaphysics and all other topics was driven by his firm belief that reason exercised by pure mind leads to the true essence of things, the One God. Intellectual activity detaches the soul from passions and confusions, and concentrates its activity on the real things. Porphyry attacked Christianity and Gnosticism because he thought they appealed to the irrational. Mysteries and rituals are fitted for those who are unable to practise inward contemplation. Salvation comes to those leading the life of the philosopher-priest.

1 Life and philosophical conflicts

Porphyry was born in the ancient Phoenician port of Tyre. His parents were Syrian and he was originally named after his father Malkhos ('king'). He spent many years expanding his general knowledge, and learning the languages (including Hebrew) and religions of the eastern Roman Empire. He then travelled to Athens, one of the two main centres of teaching in the Empire, and became a mature student of the eminent scholar Longinus. According to the vivid descriptions typical of the age, Longinus was a 'living library and walking museum' and the academic's critical attention to detail, clarity of style and erudition left their permanent mark on the keen student. Longinus renamed him 'Porphyry', which in Greek means 'purple', the royal dye for which Tyre was famous.

At the age of thirty, Porphyry went to join Plotinus in Rome. His first impression of him was ambivalent: Plotinus was original but could not express himself clearly - 'The seminars were like conversations' and 'The logical coherence of the argument was not made obvious'. Longinus disagreed with Plotinus over the contents of reality, the Platonic Forms. Longinus considered them distinct from the thoughts about them. Plotinus taught that they are identical with thoughts in God's intellect. 'On the third try' Porphyry came to 'understand with difficulty' Plotinus' position, was persuaded and 'tried to rouse in the master the ambition to organize his doctrines' (*Life of Plotinus* 18).

Five years later Porphyry moved away from Rome to Sicily. His own retrospective explanation was that he suffered from depression to the point of contemplating suicide. Recent studies, however, have linked his decision to crises in Plotinus' circle. On the one side, Plotinus' closest spokesman, Amelius, turned increasingly to theurgy as the means to enlightenment, and moved away to Apamea (Syria) (see Iamblichus). On the other, Porphyry disagreed with Plotinus' denigration of Aristotle, especially of his categories of being. Porphyry had concluded that the chief enemy was the irrational appeal of Gnosticism and Christianity (see Gnosticism). Hellenic philosophies had to close ranks. During his stay in Sicily Porphyry composed a treatise on the 'harmony of Aristotle and Plato' and wrote the key works that marked the revival of Aristotelian studies. He wrote them in a teachable form, especially the *Introduction* (*Isagōgē*) to logic and a commentary *On Aristotle's Categories* in user-friendly question and answer format. He also wrote the unique philosophical study of vegetarianism, *On Abstinence from Animal Food* (it includes an invaluable history of arguments for and against). Equipped with his knowledge of textual criticism and languages, he published a critical appraisal of Christian doctrines and the Bible (not superseded until modern times), and proved a Zoroastrian book a forgery. Most importantly he kept in touch

with Plotinus, who entrusted him with revising his teachings.

Porphyry returned to Rome much later (in the early 280s) - well after Plotinus' solitary death in AD 270 from leprosy. In this period he married, continued his sojourns, wrote commentaries on Plato and treatises on ethical, philological and historical topics, and added to his already considerable reputation as the most lucid and learned scholar. Clarity and concision are indeed Porphyry's traits. He attracted the interest of Iamblichus, who may have visited him. However, he soon disagreed vigorously with Iamblichus' fondness for the supra-rational and they parted company in the 290s. By AD 301 or 305 Porphyry had completed the grand task of editing and publishing Plotinus' teachings in the form in which they are now known - the six *Enneads*. These he prefaced, in the conventional scholastic manner, with the master's biography. Interestingly, Plotinus is presented as the paragon of the true philosopher, who reaches enlightenment through contemplation, not through theurgy, in contrast to Iamblichus. During this latter period Porphyry was invited to the court of the eastern Roman emperor, but may have returned to Rome where he died.

2 Language and metaphysics

For Porphyry, Aristotle and Plato agree once their domains are demarcated. How are they in 'harmony'? Porphyry's treatise on this is lost but we can extrapolate from other works and fragments, mainly those on Aristotle's logic. First, Porphyry disagreed with Plotinus over the fundamental divisions of things, the categories (see Aristotle §7; Categories §1). What do we categorize? Our ways of talking (words)? Our ways of thinking (concepts)? Things that exist (beings)? Furthermore, when we form statements and propositions, are the predicates we use, linguistic expressions, concepts, classes, or realities? As Porphyry tells us in the commentary *On Aristotle's Categories*, the Stoics, several Middle Platonists and Plotinus (politely left unnamed in Porphyry's attack) took Aristotle's theory of categories to be about real beings, and rejected it. Platonists could never accept that only individual, material things constitute the ultimate substance. The Platonic separable Forms are the 'truly real', and as Plotinus asserted the true categories should follow from the 'five most general kinds' in Plato's *Sophist* (see Plotinus §4). Porphyry, in contrast, saw an inherent validity in Aristotle's analysis. The categories are indeed valid when limited to the fields of logic and semantics. We base our accounts and phrases on individual concrete 'substances': for example, this or that human, animal, house; we talk of physical qualities, places, and so on. Abstracted universals are constructs of personal minds, but are not real universals, the Forms. Thus, Aristotle's categories do not impinge on what may or not be independently real. In our ordinary experience we live in a world of 'signifying sounds' (*sēmantikai phōnai*). These enable us to communicate the concepts we acquire about the objects we sense. The fundamental divisions of our immediate common experience refer to 'expressions signifying objects' (*On Aristotle's Categories* 58.3-6).

Second, how well we communicate depends on how precise our definitions are. In the *Introduction* to Aristotle's logic (in the *Categories* and *Topics*), Porphyry argued that only five 'signifying terms' cover the range of predicates we use: genus (general type), species (specific), difference (what distinguishes a family of species from the rest within a genus), property (what is characteristic of a species) and accident. Things are thus defined by their 'properties in the strict sense' (*kyriōs idia*): for example, a horse is what neighs, and conversely, what neighs is a horse. Indeed, a thing is a 'bundle' (*athroismos*) of properties (compare Russell, B.A.W. §12). Porphyry made the scheme amenable to Platonists by interpreting difference and property (also genus and species) as potentialities for fulfilment, not actualities. Defining a human as a laughing animal refers only to the capacity, which is always present, not the act of laughing itself. This makes it possible to admit degrees of fulfilment: the 'more or less' approximation to the archetypal form.

Third, Porphyry supported the distinction between immanent and transcendent universals. This we see earlier in the Middle Platonist Alcinous. Universals in concrete things correspond to the forms as Aristotle saw them, that is, as compounds with matter. Universals that are separable from material manifestation correspond to the Platonic Forms (see Platonism, Early and Middle §4).

Fourth, Porphyry seems to have interpreted a key passage in Aristotle's *Metaphysics* (XII) in a way that supported the Neoplatonic One as the foundation of all existence (see Plotinus §3). Since Porphyry's commentary on the *Metaphysics* is lost, evidence may be reconstituted from references to him in a commentary on the *Categories* by Dexippus, a student of Iamblichus (see Hadot 1974). For Aristotle, the study of physical substances presupposes a pure, intelligible substance. For a Neoplatonist like Porphyry, Aristotle suggests a series of substances, from the

physical to the incorporeal. The latter can be described only by analogy with things perceived by the senses. Furthermore, the intelligible substance gives coherence to the series of substances only because it has itself received unity from a pure One.

The ’harmonization’ through demarcation seems to underlie Porphyry’s extensive discussion of incorporeals and corporeals in his *Sentences Leading to the Intelligibles*. Although it is intended as a concise introduction to Plotinus, it contains fresh ideas anticipating modern distinctions of mind/body. Soul, intellect and the One are classed together as incorporeal. They lack spatial extension, location and volume. Without corporeality, the categories and relations of our ordinary language, big-small, in-out, here-there etc., lose their normal reference. So to locate such incorporeals we resort to paradoxical statements: for example, ’everywhere and nowhere’. Bodies are not ’zoo-cages’ for minds, nor do souls ‘fall down’ to earth. Rather, incorporeal entities exercise their activity in a particular area, according to their ’disposition’. Incorporeals and bodies do not mix but form an asymmetrical ’assimilation’ and ’union’. For Porphyry, this is the reason why our world of experience ’remains much distant’ from reality (*Sentences* 35.30).

From another perspective, the Neoplatonic One emerges as the sole entity to possess its own foundation for existence. Everything else exists in relation to it. This, together with the telescoping of the metaphysical differences between soul and intellect may mean that Porphyry was a monist. On the other hand, in his *Sentences* we see a proliferation of elemental bodies of the soul. The controversial anonymous *Commentary on the Parmenides*, attributed to Porphyry or a follower, reinforces the importance of the One, but also offers first evidence of the threefold structure, being-life-intellect, which will typify the complex metaphysics of Iamblichean Neoplatonism. It may simply be the case that Porphyry contracted or expanded metaphysical layers depending on the distinction he was examining.

Porphyry extended his polymathy to religion. He wrote on daemons and is the first to quote the Chaldaean Oracles. Because these appear to be out of character for someone known for his rationalist attitude, it has been suggested that he was concerned with them in his earlier, pre-philosophical life. However, belief in the existence of such entities was commonplace in all strata of ancient society. What is more pertinent is his evaluation of such things. Judging from his attack on Iamblichus, Porphyry does not dismiss rituals and spirit-work but limits them to a purifying role (see §3). The divine cannot be reached in this way. Philosophical contemplation alone can activate the soul towards intellect and the One God.

3 Ethics

Porphyry treats ethics as applied metaphysics and epistemology. We garner his views from lines in the *Sentences* which refer to virtue, from the *Letter to Marcella* (Porphyry’s wife) and from *On Abstinence*. In all three the goal of moral life is ’becoming like God’. In the first two he establishes that intellectual contemplation is the surest path to God. In the third he fleshes out how a rational, just life preserves things and does no harm to them.

The best example of his ethics is in the opening remarks of his *Letter to Marcella*, where he reminds his wife, ’you should put aside irrational distress, which is an emotion, and not be led by trivial judgments, but remember the divine words by which you were initiated into upright philosophy. For a steadfast attention to these words will habitually prove itself in action. It is deeds that give proof of our doctrines, and we must live by our faith to bear faithful witness to our teachings’ (8.4-8).

The aim of ethics is to show the way of assimilating one’s self to the true being, the Neoplatonic One. Since intellect is the level closest to the divine One, the path to God is philosophical. One must first reach the state of ’impassibility’ (*apatheia*) (see Stoicism §9). The calm, rational mind separates from the passions that try to tear apart the unity of human life (on separation of soul, see Plato §13). A person’s living presence is not the body and therefore cannot be physically located. ‘Surely, you must not think of my self as something that can be touched and be incidental to the senses: my true self is remote from the body, without colour and shape, not to be touched by the hands but grasped only by the mind’ (*Letter to Marcella* 8.11-3).

For Porphyry, true priests are the ’wise’, those philosophers who practise what they teach:

> ‘Whoever practises wisdom practises knowledge of God, not by always praying and sacrificing, but by practising piety in the actions towards God… we must make ourselves pleasing to God, and consecrate our own

> disposition by making it similar to the immortal and blessed nature'. He also states: 'The wise soul joins with God and always beholds God and keeps his company…. It is not the utterance of the wise that is honourable to God, but deeds. A wise person honours God even silent…. The wise person, therefore, alone is a priest, alone is beloved by God, alone knows how to pray'.
>
> (17.1-6, 16.7-17.2)

The philosopher respects all life as expressions of the One true being. Thus the philosopher-priest is a pacifist towards all types of life. There should be no killing, even for food, and certainly no killing of animals. Furthermore, we should regret that we have to eat other life forms, even plants (*On Abstinence* III 27).

Following Plotinus, Porphyry identified four kinds of virtue: civil life, purification, contemplation and exemplary. Each is fit for an appropriate grade of life:

> 'There is a difference between the virtues of the citizen; those of the one who attempts to rise to contemplation, and who, on this account, is said to be contemplative; those of the one who does contemplate perfectly; and finally those of the pure Intelligence, which is completely separated from soul'.
>
> (*Sentences* 32)

Each virtue befits a level of awareness:

> 'The objects (*skopoi*), differ with the kind of virtues. The object of the civil virtues is to moderate our passions to conform withthe activities of human nature. That of the purificatory virtues is to detach the soul completely from the passions. That of the contemplative virtues is to apply the soul to intellectual activities, even to the extent of no longer having to think of the need of freeing oneself from the passions…. The exemplary virtues reside within intelligence'.
>
> (*Sentences* 32)

4 Literary theory

During his time with Longinus, Porphyry developed his skill for literary criticism - 'critic' (*kritikos*) was a title of professional recognition. What is the meaning of a literary account? How are we to understand it? Does it refer to concrete or symbolic events? The last question was the more problematic because it opens wide the field of interpretation and links literature to the representational arts. The question then arises whether the creation of an artist (*poiētēs*) reflects truth or somehow misleads. Put in a Platonic context, do artists (in literature, sculpture, drawing) use artistic images as representations and 'imitations' of intangible truths, or do they exercise their artistic ''licence' (*exousia*) without regard for truth, and so present distortions and falsehoods?

Porphyry showed his 'critical' skill both negatively (to prove that some Gnostic and Christian holy books were recent forgeries) and positively (notably the allegorizing of Homer's *Odyssey* and the Cave of Nymphs). Building on a tradition that can be traced to Numenius (a second-century AD Platonist, who anticipated Plotinus and Iamblichus), Porphyry interpreted the Homeric hero as the human soul. The hero's journey is an allegory for the journey of the soul through life with the attendant travails and terrors: the process of psychological internalization. Poetic descriptions of the lands, dwellings and temples the hero visits, and of the people and wondrous creature he encounters, can be interpreted on many levels according to the Neoplatonic levels of knowledge and being.

Works of literature and art are not just for entertainment or purging the emotions. They may point to truths about knowledge and being, just as language mediates reality. Thus literary theory (and aesthetics) becomes part of the philosopher's purview.

5 Influence

As the editor and publicist of the *Enneads*, Porphyry ensured the dissemination of Plotinian thought. Although he expanded the emerging Neoplatonic metaphysics, in this he remained in his master's shadow. This Plotinian-Porphyrian theological Neoplatonism can be seen in Augustine and his followers in the Latin West.

Porphyry shone independently, however, in his work on Aristotle and logic. He is the pivotal figure among the Aristotle Commentators, rescuing Aristotelian studies from total obscurity and securing their place in

post-Plotinian philosophy. It is this new, broader Neoplatonism we find in Boethius, and it is this which gave rise to medieval scholasticism and so-called Aristotelianism (see Aristotelianism, medieval; Aristotelianism, Renaissance). On the Greek side, both Porphyry and Aristotle became part of the Iamblichean Neoplatonism and were enshrined in the curriculum that lasted for centuries. In the Latin West, we may trace the influence of Porphyry's *Isagōgē* both to autonomous logical theory (see, for example, the work of Shyreswood and William of Ockham), and to the debate on nominalism and realism (see Abelard). We may also profitably reflect on Porphyry's contribution to the problem of sense and reference.

See also: Neoplatonism

LUCAS SIORVANES

List of works

Porphyry's works may be dated according to very broad eras of interest, the mainly Aristotelian, followed by the Platonic, then finally a work against Iamblichus (*Letter to Anebo*) and an edition of Plotinus' *Enneads*. Twenty or so of his eighty titles survive.

Porphyry (mid 3rd century AD) *Philosophy from the Chaldaean Oracles*, ed. G. Wolff, *De philosophia ex oraculis*, Berlin: Springer, 1856; repr. Hildesheim: Olms, 1962.(Greek text.)

Porphyry (mid 3rd century AD) *On Aristotle's Categories*, trans. S. Strange, London: Duckworth, 1992.(English translation; for Greek text, see *Introduction (Isagōgē)*.)

Porphyry (mid 3rd century AD) *Introduction (Isagōgē)* ed. A. Busse, *Isagoge et in Aristotelis categorias commentarium*, CAG 4.1, Berlin: Reimer, 1887; trans. E. Warren, *Porphyry the Phoenician: Isagōgē*, Toronto, Ont: Pontifical Institute of Medieval Studies, 1975.(English translation.)

Porphyry (mid 3rd century AD) *Cave of the Nymphs in the Odyssey*, Arethusa Monographs 1, ed. L.G. Westerink, Buffalo, NY: State University of New York Press, 1969; trans. R. Lamberton, *Porphyry on the Cave of the Nymphs*, Barrytown, NY: Station Hill, 1983.(Greek text with English translation.)

Porphyry (mid 3rd century AD) *The Homeric Questions*, ed. R. Schlunk, New York: P. Lang, 1993.(Greek text and translation.)

Porphyry (mid 3rd century AD) *Against the Christians*, ed. A. von Harnack, *Porphyrius. Gegen die Christen*, Berlin: Reimer, 1916; trans. R.J. Hoffmann, *Porphyry Against the Christians: The Literary Remains*, Frome: Prometheus Trust, 1994.(The former is Greek text, the latter an English translation of the fragments.)

Porphyry (late 3rd century AD) *Letter to Marcella*, ed. W. Pötcher, Leiden: Brill, 1969; trans. K. Wicker, *Porphyry the Philosopher, To Marcella*.(The former is Greek text and German translation, the latter an English translation.)

Porphyry (late 3rd century AD) *History of Philosophy*, ed. E. des Places and A. Segonds, *Porphyre. Vie de Pythagore, - Lettre à Marcella, - frg. de l' Historie de la philosophie*. Paris: Les Belles Lettres, 1982.(Parallel Greek text and French translation.)

Porphyry (late 3rd century AD) *On Abstinence from Animal Food*, ed. J. Bouffartigue and M. Pattilon, *Porphyre. De l'Abstinence*, Paris: Les Belles Lettres, 1977-94.(Parallel Greek text and French translation.)

Porphyry (late 3rd century AD) *Sentences Leading to the Intelligibles*, ed. E. Lamberz, *Porphyrius Sententiae ad Intelligibilia*, Leipzig: Teubner, 1975; trans. K. Guthrie, *Launching-Points to the Realm of Mind*, Grand Rapids, MI: Phanes Press, 1988.(The former is Greek text; the latter is an English translation of an earlier edition.)

Porphyry (late 3rd century AD) *Letter to Anebo*, ed. A.R. Sodano, *Porfirio. Lettera ad Anebo*, Naples: L'Arte Tipografica, 1958.(Greek text and Italian translation; "Anebo' was Porphyry's pseudonym for Iamblichus, so the Letter usually appears with editions of Iamblichus' *On the Mysteries*.)

Porphyry (mid to late 3rd century AD) *Select Works of Porphyry*, trans. T. Taylor, London, 1817; repr. Frome: Prometheus Trust, 1994.(Includes English translations and explanatory notes for *Abstinence from Animal Food*, *Cave of the Nymphs in the Odyssey* and *Auxiliaries to the Perception of Intelligible Natures*.)

Porphyry (mid 3rd to early 4th century AD) *Fragmenta*, ed. A. Smith, Stuttgart and Leipzig: Teubner, 1993. (Collection of surviving Greek and Latin fragments of Porphyry's works on Aristotle, Plato, metaphysics, science, history, religion and literature.)

Porphyry (early 4th century AD) *Life of Plotinus and The Enneads of Plotinus*, trans. A.H. Armstrong, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1966.(Parallel Greek text and English translation.)

References and further reading

Brisson, L. (ed.) and Centre National Recherche Scientifique (1990-2) *Porphyre. La vie de Plotin (Porphyry. The Life of Plotinus)*, Paris: Vrin.(Important articles on Porphyry in relation to Plotinus.)
Dombrowski, D. (1984) *The Philosophy of Vegetarianism*, Amherst, MA: University of Massachusetts Press. (Porphyry and the Platonic background on vegetarianism.)
Edwards, M.J. (1990) 'Porphyry and the Intelligible Triad', *Journal of Hellenic Studies* 110: 14-25.(On Porphyry's Neoplatonic metaphysics.)
Evangeliou, C. (1988) *Aristotle's Categories and Porphyry*, Leiden: Brill.(The logic and ontology of Porphyry's theory.)
Hadot, P. (1968) *Porphyre et Victorinus*, Paris: Études Augustiniennes.(Mainly on metaphysics and theology, and their relation to Christianity.)
Hadot, P. (1974) 'The Harmony of Plotinus and Aristotle according to Porphyry', in R. Sorabji (ed.) *Aristotle Transformed*, Ithaca, NY: Cornell University Press.(Further analysis on the harmony between Plato and Aristotle.)
Lamberton, R. (1986) *Homer the Theologian: Neoplatonist Allegorical Reading and the Growth of the Epic Tradition*, Berkeley and Los Angeles, CA: University of California Press. (Includes extensive study of Porphyry's literary criticism and theory of interpretation.)
Romano, F. (1985) *Porfirio di Tiro*, Catania: University of Catania.(Survey of Porphyry's philosophy.)
Smith, A. (1974) *Porphyry's Place in the Neoplatonic Tradition*. The Hague: Martinus Nijhoff.(Classic study of Porphyry in relation to Plotinus.)
Smith, A. (1987) 'Porphyrian Studies since 1913', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II 36:2, 719-73.(Survey of scholarship on Porphyry.)
Sorabji, R. (ed.) (1990) *Aristotle Transformed*, Ithaca, NY: Cornell University Press.(Articles on Porphyry, Aristotle and logic.)
Spade, V. (ed.) (1994) *Five Texts on the Mediaeval Problem of Universals: Porphyry, Boethius, Abelard, Duns Scotus, Ockham*, Indianapolis, IN: Hackett Publishing Company. (Translated texts showing a development of Porphyry's thought.)

Port-Royal

The reform of the abbey of Port-Royal-des-Champs in 1608 coincided with the vast movement of monastic reform which characterized the Counter-Reformation. In 1624 a second abbey was created, Port-Royal-de-Paris and, after some rivalry among spiritual leaders of the day, this was directed by Jean Duvergier de Hauranne, abbot of Saint-Cyran and one of the leaders of the parti dévot *hostile to Richelieu. In 1640 the abbot's friend Jansenius published the Augustinus, a* résumé *of Augustine's doctrine, of which five propositions on divine grace were condemned by the Vatican. Ecclesiastics in France were obliged to accept this condemnation, and the resistance of the nuns of Port-Royal brought about persecution, imprisonment and finally the destruction of the abbey in 1710. The intellectual history of the abbey extends far beyond theological quibblings, however. In 1626, Saint-Cyran defended Charron against the Jesuit Garasse and formed a firm alliance between the Pyrrhonism of Montaigne and the philosophy of Augustine. In the 1640s, Antoine Arnauld played an important role in the diffusion of Cartesianism, confirmed by the publication of the Logique de Port-Royal (1662). The 1670 publication of Pascal's Pensées can be interpreted as a symptom of the rivalry between Descartes and Gassendi in contemporary apologetics and, in the following years, the polemics between Arnauld and Malebranche played an important role in the definition of Christian rationalism.*

1 Rise and fall of Port-Royal

In June 1599, Jacqueline Arnauld was chosen as coadjutrix of the Cistercian abbey of Port-Royal-des-Champs, south of Paris near Versailles. On 5 July 1602, under the name of Mother Marie-Angélique de Sainte-Madeleine (Mother Angélique), she succeeded Jeanne de Boulehart as abbess. At that time, the crucial role that Port-Royal would play in the religious, political, philosophical and cultural world of seventeenth-century Europe could not have been anticipated.

A sermon preached by a visiting friar in March 1608 was the occasion of the spiritual conversion of Mother Angélique and, with the advice of Cistercian abbots and a number of spiritual directors from various orders, she introduced the strict rules of the Cistercians into the abbey of Port-Royal. The reform of Port-Royal thus coincided with the vast movement of monastic reform which characterized the Counter-Reformation: Port-Royal soon gained a considerable reputation and the reform spread to other monasteries of the Cistercian order. In May 1625, the nuns of Port-Royal-des-Champs were transferred to the new abbey of Port-Royal in Paris and, in July 1627, the community was transferred from the jurisdiction of the Cistercian order to that of the archbishop of Paris. After the creation of the Institut du Saint-Sacrement in 1633, the community became the subject of intense rivalry among the leading spiritual directors in Paris: the Oratorians around Bérulle, Sébastien Zamet (Bishop of Langres), Octave de Bellagarde (Bishop of Sens), and Jean Duvergier de Hauranne (Abbot of Saint-Cyran). The latter became the spiritual director of Port-Royal in 1635. He was a friend of Cornelius Jansenius (Bishop of Ypres), and their spiritual ambitions were based on a return to the authority of Augustine, to positive theology, and on opposition to the influence of the Jesuits in France. This had political implications, in so far as Saint-Cyran could be said to belong to the 'parti dévot', strongly opposed to Richelieu's foreign policy of alliance with Protestant nations to resist the pressure of the Habsburgs in Spain and Austria. Jansenius' pamphlet *Mars gallicus* (1635) is an indication of the political resistance to official foreign policy and, on the 14 May 1638, Saint-Cyran was arrested and imprisoned in the dungeon of Vincennes. He was released only after the death of Richelieu in 1643, and died himself a few months later.

Nevertheless, by this time, the movement of spiritual reform at Port-Royal had gained considerable impetus. In January 1637 a well-known lawyer, Antoine Le Maistre, retired from the professional world to seek spiritual guidance at Port-Royal-de-Paris; his example was followed by others and, in mid-1638, these 'Solitaires' withdrew to Port-Royal-des-Champs, accompanied by the school ('Petites Ecoles') of Port-Royal, recently founded in Paris. The great work by Jansenius, *Augustinus* (1640), aimed to give a faithful résumé of Augustinian theology and revived old quarrels on the nature and power of divine grace. On 6 March 1642, the papal bull *In eminenti* condemned the work of Jansenius and in 1643 and 1644 Antoine Arnauld produced his *De la Fréquente communion* and his defence of Jansenius. Port-Royal was henceforth to be known as the centre of 'Jansenist' (or, as Port-Royal theologians claim, 'Augustinian') theology. In 1649, Nicolas Cornet, 'syndic' of the Faculty of Theology of the Sorbonne, reduced Jansenius' doctrine to seven articles and submitted them to the Vatican for

examination; five of them were condemned in the papal bull *Cum occasione*, published on 31 May 1653, and this condemnation was renewed in the bull *Ad sacram* on 16 October 1656. This same year, a condemnation of Jansenist theology was demanded of all ecclesiastics. The resistance of the nuns of Port-Royal drew political persecution on the abbey, since Mazarin sought to exploit the Jansenist quarrel in his negotiations with the Vatican.

The year 1655 was important in the development of public opinion concerning theological affairs. The occasion that sparked events was the refusal of communion to the duke of Liancourt, a friend of Port-Royal, by Charles Picoté, a vicar of Saint-Sulpice, on the grounds that the duke had provided accommodation in his house for Toussaint Desmares and Amable de Bourzeis, both notorious Jansenists. Antoine Arnauld wrote two lengthy letters in defence of the duke, which had little effect, but the theologian then asked Pascal to try his hand: from January 1656 onwards, the *Provincial Letters* were an immediate success and contributed to the popular image of Jesuit casuistry and moral laxism (see Pascal, B.). Moreover, despite the expulsion of Antoine Arnauld from the Sorbonne, the Jansenist cause seemed to be under divine protection when, in March 1656, Pascal's niece, Marguerite Périer, was miraculously cured of a dangerous eye infection. Debates on the 'miracle de la Sainte-Epine' led Pascal to develop the notion of historical testimony and thus embark on the project of an apology, of which the *Pensées* provides the unfinished sketch.

But political interests proved too strong for Port-Royal. First Mazarin, then King Louis XIV, were each determined to eliminate this source of potential resistance to monarchical power. The abbey was forbidden to receive new nuns, the boarding school for girls was suppressed as were the 'Petites Ecoles', and the 'Solitaires' were expelled. The abbey was thus condemned to die and, despite the 'Peace of the Church' (a respite from persecution between September 1668 and April 1679), the exile of Antoine Arnauld and Pierre Nicole and the tight control exercised by the archbishop of Paris brought an end to Port-Royal. The last nuns were dispersed in 1709, and the abbey was destroyed the following year.

In 1713, the papal bull *Unigenitus*, condemning 101 propositions extracted from the book of *Réflexions morales* (1693) by Pasquier Quesnel, created new conflicts that lasted throughout the eighteenth century. Jansenism now left the moral and theological field and began a Parliamentary and political career.

2 Intellectual life

Much has been said (by L. Goldmann, in particular) of the Parliamentary links of the Jansenists and this sociological characteristic has been interpreted in terms of a 'tragic' philosophy attributed to Pascal and to Racine. Objections have been formulated by J. Orcibal and J. Mesnard. Beyond the extreme fragility of Goldmann's interpretation of 'Jansenist' authors, suffice it to say that all religious orders recruited their members from the Parliamentary milieu: Port-Royal was no exception. Furthermore, Port-Royal had many connections with the *noblesse d'épée*, as can be seen from the list of nuns, of girls at the boarding-school and pupils at the Petites Ecoles. A full sociological analysis of the world of Port-Royal has yet to be completed; the financial history of the abbey would also reveal many secrets. The Parisian group of friends cannot suffice to characterize the community, since a vast network of friends of Port-Royal was established throughout the country, concentrated in particular in the dioceses of sympathetic bishops at such as those at Angers, Alet, Beauvais, Rouen, Troyes, Châlons-sur-Marne, Montpellier and in Lorraine.

The intellectual history of the abbey is intricate. Many of the nuns themselves had considerable intellectual qualities and proved them in the archives which provided, in the following century, ample material for the historical defence of 'Jansenism'. Moreover, for the first time in Catholic theology, the resistance of the nuns to the demand for condemnation of Jansenius in 1656 was founded on the definition of the rights of the individual conscience: a major step towards the formulation, in the following century (and following the writing of Pierre Bayle), of a doctrine of religious toleration.

Among the friends of Port-Royal, cultural activity was extremely important. Apart from the vast theological productions of Saint-Cyran and Antoine Arnauld, the pedagogical activity of the Petites Ecoles gave rise to a number of works by Claude Lancelot, as well as to the *Grammaire générale et raisonnée* (1660) and the *Art de penser* (or *Logique de Port-Royal*) (1662). Arnauld d'Andilly produced influential translations of Biblical history and Patristics; Isaac Le Maistre de Sacy produced an important translation of the Bible, which reflects the

changing status of the layman in the Catholic church in the seventeenth century. Le Nain de Tillemont, a pupil of the Petites Ecoles, gained a great reputation as a historian. The influence of Jansenism on Racine, also a pupil at Port-Royal, has been much debated. Some critics discern an Augustinian influence in the works of Pascal, La Rochefoucauld, Nicole, Mme de La Fayette and Mme de Sévigné; La Fontaine was involved in the composition of a *Recueil de poésies chrétiennes et diverses* (1671), and the director of this volume, Brienne, was a close friend of Port-Royal and the probable author of the *Discours sur les passions de l'amour*. Augustinian moralism, which founded a portrait of human nature akin to that painted by Hobbes, was later to influence Bayle, Locke and Mandeville. Conversations of friends of Port-Royal have been recorded in the *Recueil de choses diverses* (1670-1). Outside literature, Philippe de Champaigne and his nephew (and adopted son) Jean-Baptiste were among the foremost painters of the day: Philippe de Champaigne's portraits of the theologians and nuns of Port-Royal are outstanding.

In the field of philosophy, Port-Royal was first characterized by the intervention in 1626 of Saint-Cyran in defence of Charron, the disciple of Montaigne, against the attack of the Jesuit Garasse. This can be interpreted as an alliance of Augustinian theology with Christian Pyrrhonism, and heralded the subsequent quarrels between rationalists and Pyrrhonists after the publication of Descartes' *Discours de la méthode* (1637) (see Descartes, R. §3; Pyrrhonism). The introduction of Cartesianism into Port-Royal opened a new era, sometimes called the 'second Port-Royal', strongly influenced by the authority of Antoine Arnauld, author of the fourth set of *Objections* to Descartes' *Meditations*. Arnauld was indeed extremely favourable to the 'new philosophy' of Descartes, regarded as a much needed replacement for the Thomist Aristotelian defence of religious faith (see Medieval philosophy). Charles Wallon de Beaupuis, a former pupil of Arnauld, was the director of the Petites Ecoles at Port-Royal. The duke of Luynes, a neighbour of the abbey in his château of Vaumurier, first translated into French Descartes' *Meditations* in 1647, the *Objections* being translated by Clerselier. Nicolas Fontaine gives an account of vivisection at Port-Royal, based on the Cartesian theory of animal-machines and, according to the testimony of his niece, Marguerite Périer, this was an aspect of Cartesian philosophy with which Pascal agreed. A manuscript of Louis Du Vaucel, published by G. Rodis-Lewis, testifies, however, that the adoption of Cartesianism was not unanimous at Port-Royal.

The publication of the *Logique de Port-Royal* (1662), with subsequent editions bringing a number of important additions, was a major event in the intellectual history of the seventeenth century. A first version was sent by Arnauld to Mme de Sablé in April 1660, and this copy may help to solve the vexed question of the role played by Nicole in the composition of the text. Furthermore, a note in the *Recueil de choses diverses* attributes a role in the composition to Le Bon, a teacher at the Petites Ecoles, who is supposed to have worked on notes by Descartes' disciple Clerselier. The work constitutes an attempt to present a synthesis of Augustinian, Cartesian and Pascalian thought, and exerted a strong influence on Malebranche and his recalcitrant Benedictine disciple, Robert Desgabets and, indeed, on the whole context of debate around the 'new philosophy'. This debate was particularly lively at the Oratoire, where André Martin composed, under the pseudonym of Ambrosius Victor, a systematic comparison of passages extracted from Augustine and Descartes (1667), and from which Bernard Lamy suffered exile after a series of Cartesian lectures (1675). It may be added that, whereas Arnauld remained faithful to Cartesianism throughout his life, strongly denouncing the sceptical work of Pierre-Daniel Huet, Nicole seems to have been lukewarm: he composed a stringent criticism of Desgabets' attempts to provide a detailed Cartesian account of transubstantiation, and his sympathy for Descartes' philosophy seems to have been limited to an adherence to the criterion of 'évidence'.

The publication of Pascal's *Pensées* (1670) was a new occasion for Cartesian thought to make itself felt in works emanating from Port-Royal. Recent critics interpret this edition as an attempt to disguise, by careful modifications of the text, a Gassendist critique of Descartes as Cartesian apologetics. Key terms in Pascal's vocabulary are marked by the influence of Gassendi, and yet Arnauld, who directed the editorial committee, managed to produce an apology compatible with Cartesianism. The success of the work was immediate, and it exerted an influence on subsequent philosophers, in particular on Malebranche, Bayle, Locke and even Hume.

Antoine Arnauld's philosophical controversies with Malebranche in themselves constituted an important chapter of intellectual life at Port-Royal. The controversies began in 1680 when Arnauld came across a manuscript version of Malebranche's *Traité de la nature et de la grâce*: despite Arnauld's last-minute attempt to discourage Malebranche, the treatise was printed and the Jansenist theologian promptly attacked the Oratorian's so-called

'Epicureanism' (see Epicureanism). A debate on the nature of pleasure and happiness ensued, in which Pierre Bayle played an important role by publishing detailed analyses in the *Nouvelles de la république des lettres* and finally coming down strongly in favour of Malebranche. In 1683 Arnauld's critique of Malebranche's *Recherche de la vérité* appeared in *Des Vraies et des fausses idées* (1683), which was to involve the two philosophers in a long and intense debate on the nature of ideas. Originally regarded as friend of Port-Royal, Malebranche became a bitter enemy of the Port-Royal theologians, denouncing them in the ironic formula 'personnes de piété' (see Arnauld, A. §3).

The paradoxically parallel influence of Epicurus, Montaigne and Augustine in seventeenth-century France put Port-Royal at the centre of debates on theological, ecclesiastical, moral, philosophical and literary questions. The Port-Royal theologians and their friends were quick to appeal to public opinion, and thus played a major part in the transformation of these debates from obscure, technical quibblings to major events in social and intellectual history.

See also: Logic in the 17th and 18th centuries; Moralistes

ANTONY McKENNA

References and further reading

Arnauld, A. (1641) *IVes Objections aux Méditations philosophiques de Descartes*, in Descartes, R. *Méditations métaphysiques (Metaphysical meditations)*, ed. F. Alquié, Paris: Garnier frères, 1967, vol. II: 632-704. (Arnauld's objections to Descartes' *Méditations*.)

Arnauld, A. (1643) *De la Fréquente Communion, où les sentiments des Pères, des papes et des Conciles, touchant l'usage des Sacrements de Pénitence et d'Eucharistie, sont fidèlement exposez (Of Frequent Communion)*, Paris.(First attack by Antoine Arnauld against 'laxist' morals of the Jesuits.)

Arnauld, A. (1683) *Des vraies et des fausses idées contre ce qu'enseigne l'auteur de la Recherche de la verité (Of true and false ideas)*, Cologne.(Arnauld's attack on Malebranche's epistemology.)

Arnauld, A. and Lancelot, C. (1660) *Grammaire générale et raisonnée, contenant les fondements de l'art de parler expliquez d'une manière claire et naturelle (General and analytical grammar)*, Paris.(Famous French grammar composed for the schools of Port-Royal.)

Arnauld, A. and Nicole, P. (1662) *La Logique ou l'art de penser*, ed. P. Clair and F. Girbal, Paris: Vrin, 1981; trans as *Logic or the Art of Thinking*, ed. J. Vance Buroker, Cambridge: Cambridge University Press, 1996. (The 1981 French version is a good critical edition of this major seventeenth century text. The manuscript mentioned in the text is in the Bibliothèque Nationale f.fr. 19915)

Bayle, P. (1684-8) *Nouvelles de la république des lettres*, Rotterdam.(Bayle's periodical work, consisting of book reviews; in 1685 he inserts an appendix to defend Malebranche against Arnauld.)

Descartes, R. (1637) *Discours de la Méthode (Discourse on method)*, ed. E. Gilson, Paris: Vrin, 1925. (Descartes' famous discourse inaugurating a philosophical revolution against Aristotelian scholastic tradition.)

(1680-5) *Discours sur les passions de l'amour (Discourse on the passions of love)*, ed. G. Brunet, Paris, 1959. (Pseudo-Pascalian text, now attributed to Brienne by J. Mesnard.)

Goldmann, L. (1955) *Le Dieu caché. Étude sur la vision tragique dans les Pensées de Pascal et dans le théâtre de Racine (Study of the tragic vision in Pascal's Pensées and the plays of Racine)*, Paris: Gallimard.(Controversial sociological interpretation, relating the 'tragic' vision of these two authors to the social situation of Parliamentary circles in the seventeenth century.)

Jansenius, C. (1635) *Alexandri Patricii Armacani, theologi, Mars Gallicus, seu de justitia armorum et foederum regis Galliae libri duo (The French God of War)*, Louvain.(Pro-Spanish pamphlet against Richelieu's policy of protestant alliance.)

Jansenius, C. (1640) *Cornelii Jansenii episcopi Iprensis Augustinus*, Louvain.(Exposition of the doctrine of Saint Augustine.)

Lafond, J. (1977) *La Rochefoucauld. Augustinisme et littérature*, Paris: Klincksieck, 3rd edn, 1986.(Major study in the perspective suggested by P. Sellier.)

Lafond, J. (ed.) (1992) *Les Moralistes du XVIIe siècle*, Paris: R. Laffont.(Excellent editions of major French moralists, including Pascal, La Rochefoucauld, La Bruyère and an excellent introduction on epigrammatic literature.)

Lafond, J. and Mesnard, J. (eds) (1984) *Images de La Rochefoucauld: Actes du Tricentenaire*, Paris: PUF. (Articles on the reception of La Rochefoucauld and on his relations with Port-Royal and the hôtel de

Liancourt.)
La Fontaine, J. de and Loménie de Brienne, H. de (eds) (1671) *Recueil de poésies chrétiennes et diverses (Collection of Christian and diverse poetry)*, Paris.(Important collection for the definition of Port-Royal esthetics, prefaced by Nicole.)
Lesaulnier, J. (1992) *Port-Royal insolite: édition critique du Recueil de choses diverses*, Paris, Klincksieck. (Critical edition of conversations in 1670-71, between friends of Port-Royal in the circle of Jean Deslyons, who was often to be found at the hôtel de Liancourt.)
Lesaulnier, J. and McKenna, A.N. (eds) (1998) *Dictionnaire de Port-Royal au XVIIe siècle*, Paris: Universitas; Oxford: The Voltaire Foundation.(Bio-bibliographical dictionary of over 2,500 characters related to Port-Royal; the introduction examines Jansenist theology and the importance of Port-Royal in seventeenth-century culture; offers detailed chronology and exhaustive chronological bibliography.)
Malebranche, N. (1680) *Traité de la nature et de la grâce*, Amsterdam: Elzevier.(One of an important series of works in which the Oratorian Malebranche proposes an original alliance between Augustine's doctrine of grace and his own version of Cartesian rationalism. Bayle highlights the publication of this work as the turning point in relations between Malebranche and Port-Royal.)
Martin, A. [Ambrosius Victor] (1667) *Philosophia christiana*, Paris.(Collection of works in which the Oratorian Martin proposes an alliance of Saint Augustine and Descartes.)
McKenna, A.N. (1986) 'La composition de la *Logique* de Port-Royal', *Revue philosophique* 1986: 183-206. (Details of current hypotheses.)
McKenna, A.N. (1990) *De Pascal à Voltaire: le rôle des Pensées de Pascal dans l'histoire des idées entre 1670 et 1734*, vols 276-7 of *Studies on Voltaire and the eighteenth century*, Oxford: Oxford University Press.(Lengthy study of Pascal's reception and influence on major philosophers and minor writers, including apologists and clandestine philosophers.)
Mesnard, J. (1965) *Pascal et les Roannez*, Paris: Desclée de Brouwer.(Detailed history of family relations, thus giving a clear picture of the Parisian circles within which Pascal moved.)
Mesnard, J. (1992) *La Culture du XVIIe siècle*, Paris: PUF.(Important collection of articles on all aspects of seventeenth-century literature.)
Mesnard, J. (1993) *Les Pensées de Pascal*, Paris: SEDES.(The best global commentary on the *Pensées*. Pages 309-30: refutation of Goldmann's interpretation of Pascal and Racine.)
Nadler, S. (1989) *Arnauld and the Cartesian Philosophy of Ideas*, Manchester: Manchester University Press. (Good study of Arnauld's philosophy which is readily comprehensible to the non-specialist.)
Nadler, S. (1992) *Malebranche and Ideas*, New York and Oxford, Oxford University Press. (Good study of Malebranche's polemical debate with Arnauld.)
Orcibal, J. (1947-8) *Les Origines du Jansénisme:* 1, *La Correspondance de Jansénius*; 2-3, *Jean Duvergier de Hauranne, abbé de Saint-Cyran et son temps*, Louvain-Paris.(Very detailed history of the career of Saint-Cyran and of his contacts in all social circles.)
Pascal, B. (1964-) *Œuvres complètes*, ed. J. Mesnard, Paris: Desclée de Brouwer. (Monumental, critical edition of all Pascal's works, including a commentary relating his writing to spiritual traditions and contemporary debate. Four volumes published as of 1995.)
Pascal (1992), special edition of *XVIIe siècle* 177.(Recent research on Pascal's *Pensées*.)
Quesnel, P. (1693) *Le Nouveau Testament en français avec des réflexions morales sur chaque verset, pour en rendre la Lecture plus utile et la Méditation plus aisée à ceux qui commencent à s'y appliquer (New Testament in French with moral reflections on each verse)*, Paris: Pralard.(Commentary on the New Testament, from which were extracted 101 propositions condemned in the bull *Unigenitus*.)
Rodis-Lewis, G. (1950) 'Augustinisme et cartésianisme à Port-Royal' in *Descartes et le cartésianisme hollandais*, Paris and Amsterdam.(Important article showing controversy at Port-Royal over Cartesian philosophy and including the objections of Louis Du Vaucel.)
Sellier, P. (1970) *Pascal et saint Augustin*, Paris: Armand Colin.(Pioneer work on Pascal's theological thought, concluding that he was very well acquainted with Augustine's works.)
Sellier, P. (1969) 'La Rochefoucauld, Pascal, saint Augustin', *Revue d'histoire littéraire de la France* 69: 551-75. (The first study to suggest a relation between La Rochefoucauld and Augustinian theology, a line of enquiry which has proved very useful.)

Posidonius (*c*.135-*c*.50 BC)

Posidonius of Apamea (Syria) was a Stoic philosopher and student of Panaetius. He taught in Rhodes. He combined a passion for detailed empirical research with a general commitment to the basic systematics of Stoic philosophy (which, however, he was willing to revise where necessary). As such he was probably the most 'scientific' of the Stoics. His wide-ranging investigations of all kinds of physical phenomena (especially in the areas of physical astronomy and meteorology) became particularly renowned, the best known case being his explanation of Atlantic tides as connected with the motions of the moon.

His most original philosophical contributions are to be located in the connected areas of psychology and ethics. Posidonius appears to have been committed to a slightly Platonizing version of Stoic psychology, according to which the passions are no longer regarded as a malfunctioning of the rational faculty, but as motions of the soul which take their origin in two separate irrational faculties (anger and appetite). This revised moral psychology is accompanied by some corresponding revisions in ethics such as the conception of moral education as the blunting of the motions of the irrational faculties.

1 Life and writings

Born in Apamea (Syria), Posidonius studied at Athens with the Stoics Antipater and Panaetius (that is, before 110 BC when the latter died). He settled in Rhodes where he founded a philosophical school of his own, and soon became one of the leading intellectual personalities of his time. He obtained Rhodian citizenship and held the highest political office (*prytanis*), Rhodes being at the time an independent ally of Rome. He represented Rhodes in an embassy to Rome where he negotiated with Marius in 87/86 BC. Earlier he had travelled in the western Mediterranean visiting Etruria, southern Gaul (where he studied the habits of the Celts), Spain (where he observed Atlantic tides at Cadiz) and North Africa. Back in Rhodes he continued his philosophical activities. Cicero attended his lectures in 77 BC. Pompey visited him in 66 and 62 BC, on his way to and back from his campaign against Mithridates. In 51 BC Posidonius revisited Rome on a second embassy. He died soon after.

None of Posidonius' writings survive, so we are dependent on fragments (quotations) and testimonia provided by other ancient writers. More than forty titles of works are attested, the most important ones being: *Protreptics*; *Physical Theory* (consisting of at least eight books); *On the Cosmos* (at least two books); *Meteorologica* (at least seven books); *On the Ocean*; *On the Gods* (at least five books); *On Fate* (at least two books); *On Divination* (five books); *On the Soul*; *Ethical Theory*; *On Passions*; *On Duty* (at least two books); *On the Criterion*; *Against Zeno of Sidon*; and *Histories* (fifty-two books).

The belief that Posidonius also wrote separate commentaries on some of Plato's works (*Timaeus*, *Phaedrus*, *Parmenides*) has been rather persistent, but is based on an unwarranted inference from the relevant ancient testimonies.

2 The Posidonian question

Modern scholarly interest in Posidonius as a philosopher in his own right goes back to the first collection of fragments, edited by J. Bake in 1810. This collection triggered rather than settled the question to what extent Posidonian material, philosophical as well as doxographical, can be found in various ancient sources. In the nineteenth and twentieth centuries opinions on what might count as the Posidonian 'corpus' have differed widely. Under the influence of German *Quellenforschung* Posidonius was credited with the intellectual authorship of many ideas in such different authors as Cicero, Varro, Manilius, Vergil and Seneca, usually on arbitrary grounds and without any explicit ascription in the passages at issue. The various pictures of Posidonius to which this 'pan-Posidonianism' gave rise usually exhibited strongly metaphysical or even mystical features. In the course of the twentieth century more sceptical approaches have gradually prevailed, and the sound methodological principle that every assessment of Posidonius' thought should be based on explicitly attested evidence was at the basis of the edition of the fragments and testimonia by Edelstein and Kidd (1972). The rival edition by Theiler (1982) is still rather indebted to the German pan-Posidonianist tradition. The account which follows is based on attested fragments only; references are to the fragments (fr.) and testimonia (T) as numbered in the edition of Edelstein and Kidd.

3 Logic

Posidonius followed Stoic orthodoxy in dividing philosophy into three parts: physics, logic and ethics. He was more explicit than others, however, in stressing the interconnection of these parts, likening philosophy as a whole to a living being with physics as blood and flesh, logic as the bones and sinews, and ethics as the soul (frs 87, 88). What little is preserved of his works on the bones and sinews of philosophy does not suggest any major departures from mainstream Stoic logic, although Posidonius may have been the first Stoic to examine the logical basis of relational syllogisms, that is, syllogisms making use of relations like ‘larger than’ or ‘equal to’ (fr. 191).

4 Physics

Also in the area of physics - which Posidonius claimed to be primarily concerned with ‘aetiology’, or finding causes (fr. 18, T85) - he subscribed to the Stoic orthodoxy in many respects. Yet a typical feature of Posidonian physics was the way he extended his aetiological investigations to the smallest details of the whole range of physical phenomena (thus inviting a comparison with Aristotle; see also T85). Indeed his explanations of the size and distance of the sun, its settings and risings, the moon, solar and lunar eclipses, comets, halos, rainbow, thunder, hail and winds became particularly renowned in later antiquity. Yet he regarded these various physical phenomena as in one way or another interconnected by cosmic sympathy (*sympatheia*) - a feature, incidentally, which some scholars have taken to be the exclusive hallmark of Posidonian physics, but which is in fact firmly rooted in previous Stoic philosophy. Of particular interest in this respect is his theory of the tides in the ocean as being connected with the daily, monthly and annual motions of the moon (frs 138 and 214-20). The conception of cosmic *sympatheia*, together with the identification of the active principle of the cosmos with ‘fate’ (fr. 103), also provides the philosophical basis for the practice of divination, which Posidonius, unlike his teacher Panaetius, zealously defended (frs 106-13).

On the other hand Posidonius did not hesitate to propose revisions of early Stoic physics where he thought they were required. Thus he probably rejected the early Stoic claim that the void surrounding the cosmos was infinite, arguing instead that it was just large enough to be able to receive the cosmos during its periodic conflagrations (fr. 97a-b) (see Stoicism §5). A more important change concerned psychology, where Posidonius followed Stoic orthodoxy in defining the soul as a ‘warm breath’ (fr. 139) and in claiming that it does not have irrational *parts*, located in different bodily organs, but sided with Plato (§14) and Aristotle (§§20, 22-3) against Chrysippus in maintaining that the soul has different *faculties* (frs 30-4, 142-6): a rational faculty, one irrational faculty responsible for anger and aggression, and yet another irrational faculty responsible for appetites (see Psychē).

5 Ethics

Posidonius made this moral psychology the basis of his theories concerning the nature of good and evil, the virtues and the end (frs 30, 38, 150a-b). He claimed that the origin of moral evil is not to be sought in the corrupting influence of external factors, as Chrysippus had thought, but in the emotional pull of the irrational faculties of the soul (frs 161, 169, 186). He accordingly rejected the Chrysippean conception of the passions as (false) judgments of the rational faculty of the soul (see Stoicism §19), arguing that it was unable to explain how emotions can arise and abate over time (fr. 165), that it would involve the counterintuitive consequence that animals and children, being irrational, do not have emotions (fr. 33), and that it left Chrysippus without a proper explanation of how emotions can arise in the first place (frs 34, 165). According to Posidonius, the passions are motions of one of the two irrational faculties of the soul (frs 151- 2), a conception which he claims to be firmly rooted in common experience (fr. 156), supported by the language used by the ancient poets as well as by ancient practices (fr. 164, 50-2), and sanctioned by the authority of Cleanthes (frs 32 and 166) and, possibly, Zeno of Citium (T91, 93).

Each of the three psychic faculties of the soul has its own natural object (a revision of the early Stoic theory of *oikeiōsis* or natural appropriation; see Stoicism §14): the two irrational faculties strive after power and pleasure respectively, whereas the rational faculty naturally seeks the moral good (frs 158, 160). Each faculty accordingly has its proper virtue. Only the virtue of the rational faculty may be termed ‘knowledge’ and can be taught. The virtues of the irrational faculties (that is, their restraint) are irrational and have to be acquired through habituation (fr. 31). Moral education thus becomes a kind of therapy: the rational faculty is to be given predominance and the movements of the irrational faculties should be blunted by habituation and good practices (fr. 169. 106-17).

The end (*telos*) is defined as 'to live contemplating the truth and order of all things together and helping in organizing it, in no way being led by the irrational part of the soul' (fr. 186; the misleading 'part' here may not represent Posidonius' original wording). Posidonius appears to have regarded this definition as a clarification of the early Stoic definition of the end as 'living in accordance with nature' (fr. 185) (see Stoicism §17).

Apart from his revisions in moral psychology, Posidonius appears to have remained faithful to the main tenets of Stoic ethics, such as those concerning the self-sufficiency of virtue or the moral value of wealth and health.

6 Mathematics, geography and history

Most of Posidonius' contributions to what we would call the 'special sciences' are to be classed under philosophy (physics), with the exception of his work in mathematics and (at least part of) his work in history and geography. According to Posidonius mathematics (including astronomy) does not belong to philosophy at all, being crucially different from physics in so far as it is merely concerned with the quantitative aspects of phenomena and hence unsuitable for finding the essence of things, that is, for aetiology (fr. 18, T85). Indeed he remained faithful to this principle in using mathematical constructs and methods for descriptive rather than explanatory purposes, for example, to estimate the circumference of the earth (fr. 202) or to construct an orrery (T86). Nevertheless he also studied mathematics as a subject in its own right, defending the method and principles of Euclidean geometry against Epicurean criticisms in his 'Against Zeno of Sidon' (fr. 47).

His geographical works and his *Histories* (a continuation of Polybius' historical work, covering the period from 145 to *c*.86 BC) constituted a 'mixed bag' of simply descriptive passages - such as those dealing with the length of the 'isthmus' of Gaul (fr. 248) or with the revolt of mining slaves in Attica *c*.100 BC (fr. 262) - and more philosophical elements. As an example of the latter we may note the theory of geographical zones (*klimata*), which is supposed to explain some of the ethnological characteristics of the people inhabiting various parts of the world (frs 49 and 169, 85-95). These characteristics in their turn were thought to explain certain historical actions of those people (see also fr. 272 on the Cimbri). Moreover, there is a more general link between historiography and ethics in so far as history as a whole is presented as a process of gradual decline caused by the preponderance of the irrational faculties of the soul (fr. 284).

7 Influence

Posidonius was highly esteemed in his own day, and his works were read and used by thinkers like Cicero. Yet the extent of his general philosophical influence is still a matter of debate and has certainly often been overestimated. Among later authors, such as Strabo, Cleomedes, and Alexander of Aphrodisias, he appears to have been primarily known for his detailed investigations in such special areas as meteorology and geography.

See also: Cleomedes; Panaetius; Stoicism §1

KEIMPE A. ALGRA

References and further reading

Bake, J. (1810) *Posidonii Rhodii Reliquiae Doctrinae*, Leiden: Haak.(The first collection of Posidonius' fragments.)

Dobson, J.F. (1918) 'The Posidonius Myth', *Classical Quarterly* 12: 179-95.(An influential critical assessment of 'pan-Posidonianism'.)

Edelstein, L. (1936) 'The Philosophical System of Posidonius', *American Journal of Philology* 57: 286-325.(A useful overview based on explicitly attested material only.)

Kidd, I.G. (1978) 'Philosophy and Science in Posidonius', *Antike und Abendland* 24: 7-15. (An investigation of Posidonius' conception of philosophy, mathematics and the liberal arts.)

Kidd, I.G. (1988), *Posidonius II: The Commentary*, Cambridge: Cambridge University Press, 2 vols.(Rich commentary, full of further references, on Edelstein and Kidd (1972).)

Laffranque, M. (1964) *Poseidonios d'Apamée: essai de mise au point (Posidonius of Apamea: an attempt at clarification)*, Paris: Presses Universitaires de France.(A good general introduction based on sound methodological principles; contains a useful overview of the history of Posidonian scholarship and of the Posidonian question.)

Posidonius (*c*.135-50 BC) Fragments, in L. Edelstein and I.G Kidd (ed.) *Posidonius I: The Fragments*, Cambridge: Cambridge University Press, 1972; 2nd edn, 1989(The modern standard edition of the fragments and

testimonia.)
Reinhardt, K. (1921) *Posidonius*, Munich: Beck.(A study which has been influential and important in so far as it criticizes earlier pan-Posidonianist approaches, but which is itself too liberal in ascribing unattested material to Posidonius on the basis of the fuzzy notion of the ’innere Form’.)
Steinmetz, P. (1994) ‘Poseidonios aus Apameia’, in H. Flashar (ed.) *Die Philosophie der Antike, Band 4: Die Hellenistische Philosophie*, Basle: Schwabe, 670-705.(A clear general survey, rather indebted to Reinhardt’s work, but useful for the bibliographical information it provides.)
Theiler, W. (1982) *Poseidonios: Die Fragmente*, Berlin and New York: de Gruyter, 2 vols. (Is too liberal in including non-attested material and should therefore be used with caution; the commentary, however, may sometimes be useful.)

Positivism in the social sciences

Positivism originated from separate movements in nineteenth-century social science and early twentieth-century philosophy. Key positivist ideas were that philosophy should be scientific, that metaphysical speculations are meaningless, that there is a universal and a priori scientific method, that a main function of philosophy is to analyse that method, that this basic scientific method is the same in both the natural and social sciences, that the various sciences should be reducible to physics, and that the theoretical parts of good science must be translatable into statements about observations. In the social sciences and the philosophy of the social sciences, positivism has supported the emphasis on quantitative data and precisely formulated theories, the doctrines of behaviourism, operationalism and methodological individualism, the doubts among philosophers that meaning and interpretation can be scientifically adequate, and an approach to the philosophy of social science that focuses on conceptual analysis rather than on the actual practice of social research. Influential criticisms have denied that scientific method is a priori or universal, that theories can or must be translatable into observational terms, and that reduction to physics is the way to unify the sciences. These criticisms have undercut the motivations for behaviourism and methodological individualism in the social sciences. They have also led many to conclude, somewhat implausibly, that any standards of good social science are merely matters of rhetorical persuasion and social convention.

1 Positivist doctrines

In contemporary debates ‘positivism’ refers to a broad attitude about science and philosophy that has its origins in the social theories of Comte de Saint-Simon and Auguste Comte and the philosophical doctrines of the logical positivists. The positivists often did not agree which doctrines were essential to their position or how those ideas should be interpreted, and the term ‘positivism’ has come to stand for a set of ideas that some positivists would not have fully endorsed. Though they advocated all the ideas discussed below in some form, this entry concentrates on the ideas currently associated with positivism, for it is those doctrines that have been most influential in debates about the social sciences.

Saint-Simon and Comte advanced three ideas typical of positivism: (1) that science is the highest form of knowledge and that philosophy thus must be scientific; (2) that there is one scientific method common to all the sciences; and (3) that metaphysical claims are pseudoscientific. The second of these ideas, frequently known as ‘naturalism’, is most directly relevant to the social sciences. It says that the social realm, as part of the natural world, can and must be studied by standard scientific methods. For Saint-Simon and Comte, this meant that laws governed the social world, and that the job of the social sciences was to identify those laws. They also believed that society should be run by an elite who could discern these laws and use them to solve social problems.

Subsequent positivists advocated these core ideas but with new interpretations and additions. The logical positivists shared the antimetaphysical sympathies of Comte and also wanted philosophy to be ‘scientific’; they did not conceive of it as an empirical discipline. Instead they assimilated philosophy more to the model of logic, where the logician’s task was to formulate the a priori rules of valid inference and explicate basic logical concepts. Philosophy became primarily clarifying and formulating the rules and concepts of good science, and the logical positivists were thus committed to the existence of a single, universal a priori scientific method (see Logical positivism).

The logical positivists attacked metaphysics by bringing in the empiricist tradition (one foreign to Saint-Simon and Comte). They argued that sensory knowledge was the most certain and that any concept not directly about sensory experience should be ‘translatable’ into observational concepts. Those concepts that could not be so translated, primarily metaphysical concepts, were rejected as meaningless. This claim also led to a very specific and long-persuasive picture of science: science is divided into two parts - the observational and the theoretical - and theoretical terms must be definable in observational terms.

From these ideas the positivists drew several conclusions relevant to the social sciences. Since scientific knowledge came from applying an a priori logic of science, good social science must be value free (see Value judgments in social science §2). The logical positivists also thought that good science was unified science, and promoting the unity of the different sciences was one of their chief goals. By ‘unity’ they understood various

things. They thought there was one scientific method common to all the sciences, including the social sciences. They also thought that the different sciences could and should be unified by being reduced to physics.

2 Influence on the philosophy and practice of the social sciences

Positivism strongly influenced both the practice of the social sciences and philosophical views about them. Naturalism, the view that good social science embodies the scientific method common to the natural sciences, is the dominant approach among practising social scientists (see Naturalism in social science §1). Operationalism, the idea that all theoretical terms must be specified by the operations that measure them, has been influential in the social sciences and is a variant of the positivist ideal of translating theoretical terms into empirical ones; the notion of revealed preference in economics - the idea that preferences of consumers should be identified with their observable consumption patterns - has similar roots (see Operationalism §2). Behaviourism in political science has similar origins, for again the idea is that only explanations based on directly observable factors are scientifically adequate.

One attempt to formulate the positivists’ logic of science was Karl Popper’s idea that good science is falsifiable and has undergone severe tests (1962). Many philosophers and social scientists have tried to use Popper’s ideas to evaluate the social sciences. Popper himself argued that both Marxism and Freudian psychology were pseudoscientific because they were non-falsifiable. Similar doubts have been raised about economics, because economists often fail to reject their theories in the face of conflicting data (see Popper, K.R.).

Positivist assumptions have also influenced the use of statistics in the social sciences. Much social research uses statistical methods to test hypotheses and implicitly treats statistical inference as a mechanical, purely logical process. Social research also often reports only correlations between variables, drawing no conclusions about causes. Both practices have positivist origins. The early developers of modern statistics were sympathetic to positivism, and interpreted their results as a concrete implementation of the positivists’ a priori scientific method or logic of science. The refusal to draw causal conclusions comes in part from the positivist rejection of metaphysics, for many positivists thought causation an obscure metaphysical concept that should be rejected in favour of lawful regularities between observables. This emphasis on the observable also found a sympathetic ear among economists. Friedman’s famous defence of economics against the charge that its models were highly unrealistic argued that only predictions count (1953); unrealistic theories did not matter if they predicted the observed data (see Economics, philosophy of §4).

Another influential idea was the positivists’ notion that the various sciences should be unified by reducing them to physics. This was generally interpreted as the claim that the sciences form a hierarchy, with sociology reducible to psychology, psychology to biology, and biology to physics. These reductions were to show that the laws and theories of each science were derivable from or special cases of the science ‘below’ it. This positivist ideal fits well with the tradition in the social sciences known as methodological individualism, a doctrine asserting that all social phenomena must be explained as the outcome of individual behaviour. Methodological individualism remains the official philosophy of many economists, and philosophers have defended it as well. It is important to note, however, that much social science from the start rejected individualism for a holism that claimed macrosociological processes are essential to explanation; Comte, for example, advocated such a view (see Holism and individualism in history and social science).

Positivism also played a major role in numerous other debates in the philosophy of the social sciences. The idea of *Verstehen*, namely, that the social sciences rely on a special kind of intuitive insight to grasp the meanings of social phenomena, was roundly rejected by philosophers on the grounds that such a process could not be put in terms of observable data. Arguments that any social science dealing with meanings is scientifically inadequate also have positivist roots. Quine (1960), for example, argued that linguistic meaning is inherently indeterminate and thus scientifically suspect, because we cannot capture meanings in observational or physical terms (see Quine, W.V. §2). Related doubts about a science of meaning turn on the fact that meanings involve intensionality - substituting terms with equivalent meanings does not always preserve the truth value of statements. This makes it difficult to turn social science theories into axiomatized formal systems, a requirement for good science again inherited from the positivists.

Finally, positivism has influenced in subtle ways how philosophers think philosophy of social science ought to be

done. Relevant positivist assumptions include the following ideas: that philosophers should analyse scientific concepts; that such analyses ought to be roughly in the form of a definition or necessary and sufficient conditions; that the cognitive content of science can be fully and best expressed by identifying the formal structure of its theories; and that the epistemic evaluation of science requires reference only to the formal relation between data and hypothesis. These ideas still influence how philosophers practise the philosophy of social science. For example, the long debate over functional explanation in the social sciences has largely been framed in terms of finding the necessary and sufficient conditions for having a function or existing for a purpose (see Functional explanation §3). Much debate in philosophy of economics has been over what are the fundamental postulates of neoclassical economics, implicitly assuming that there is one theory to capture and that formalizing theories is the best way to understand a science. Current assessments of the social sciences also often ignore the social, psychological and historical factors that influence the social sciences as well as the empirical practices of social research; assessments are made instead by appeal to broad conceptual considerations about, for example, the nature of human agency (‘free will’) or the nature of mental phenomena (‘having meaning’). This practice goes back to the positivists’ emphasis on logical explication.

3 Criticisms and contributions

Positivism has, of course, been widely criticized. Perhaps most heavily criticized was the positivist claim that theories must be translatable into observational terms. This claim was criticized on several grounds: (1) the theory/observation distinction is difficult to draw in any sharp way, (2) attempts to translate theoretical terms into observational ones often presupposed theoretical terms in the process of describing the observational data; theoretical terms can be applied in indefinitely many ways to observations, and (3) even if the theory/observation distinction could be drawn, every scientific test involves background theoretical assumptions, thus showing that observational evidence has no absolute epistemic status. These criticisms also lead to doubts about the positivist notion of a unified science. If theoretical terms cannot and need not be reduced to observational ones, then it seems unlikely that the special sciences (biology, psychology, sociology, and so on) are reducible to physics or that they must be to be good science.

These criticisms certainly undercut positivist doctrines in the social sciences. Behaviourism and operationalism lose much of their motivation if good natural science is not translatable into purely observational terms. Methodological individualism likewise is much less plausible once the reductionist picture of scientific unity is rejected. In fact, a detailed literature has shown that precisely the same problems that prevent any reduction of the theoretical to the observational confront attempts to reduce the social to the individual.

The positivist search for an a priori universal scientific method was also criticized. Doubts about a universal scientific method came because the actual practice and history of science seems to show diverse methods across disciplines and historical periods. Doubts about the a priori status of scientific method resulted from the general philosophical attack, stemming from Quine’s work, on foundationalism, which is roughly the idea that philosophy stands outside the sciences and can make a priori judgments about them. The holistic picture of theory evaluation mentioned above also suggested that abstract scientific methods far from exhaust the practice of science and that much substantive, domain-specific information is involved in their application.

These criticisms have many implications for the social sciences. Statistical inference is not a mechanical way to get at the truth from the data but instead depends on substantive background assumptions. Also, understanding and evaluating social science requires looking at a complex combination of practices and assumptions rather than focusing on formalized theories alone. Some social scientists and philosophers draw further and much more drastic conclusions. For example, it is widely claimed that the social sciences need not worry about being good science by natural science standards because there is no universal scientific method and because good science is only a matter of convention - good science is only a matter of what persuades or what powerful people in the scientific community accept. Many social scientists and philosophers now advocate these views and thus see naturalism as misguided.

These conclusion are hasty. It is true that at some level of description the different sciences use different methods. Yet it also true that the different specialities within the same science, for example, biochemistry and ecology within biology, also use different methods. To avoid the conclusion that there are as many methods as there are scientists, we might distinguish abstract from realized methods. Scientific virtues such as testability and objectivity

are abstractions, ones that are embodied in different practices in different scientific domains. Differences in specific methods across time and place need not threaten naturalism if they result from adapting the same abstract scientific method to the specific circumstances of a given science. For example, the idea that a good test involves ‘controls’ for spurious influences is common to much good science, though how such control is achieved varies greatly from area to area. So there may be a broad core of scientific rationality that the social sciences can and must achieve, though its actual embodiments will no doubt depend on features specific to the social domain. Naturalism in this form remains a serious position.

It is likewise misguided to conclude from the demise of positivism alone that good science is merely a matter of what persuades, or a social construction. If methodological judgments are a not true a priori, they might none the less be empirically warranted. Good methods are those that effectively promote scientific goals such as truth. Thus we can grant the criticisms of positivism and yet still judge that some inquiries are better than others because experience shows they are better able to produce the truth. In short, the positivists’ a priori universal method and an ‘anything goes’ approach are not the only two alternatives. However, if judgments about good science are both empirical and dependent on how scientific virtues are realized, philosophy of social science then must look carefully at the actual practice of social research and the distinction between philosophy and social science becomes blurred.

See also: Positivism, Russian

HAROLD KINCAID

References and further reading

Friedman, M. (1953) *Essays in Positive Economics*, Chicago, IL: University of Chicago Press.(Contains Friedman’s instrumentalist defence of economics.)

Gordon, S. (1991) *The History and Philosophy of Social Science*, London: Routledge. (A broad survey that includes positivism’s historical role in the social sciences.)

Hausman, D. (1992) *The Inexact and Separate Science of Economics*, Cambridge: Cambridge University Press.(A survey of issues in philosophy of economics with very clear discussions of positivism, Popper and post-positivist approaches to philosophy of science.)

Kincaid, H. (1996) *Philosophical Foundations of the Social Sciences: Analyzing Controversies in Social Research*, Cambridge: Cambridge University Press.(Discusses positivism and methodological individualism in the social sciences; defends the view of philosophy of social science suggested in §3.)

Laudan, L. (1987) ‘Progress or rationality? The prospects for normative naturalism’, *American Philosophical Quarterly* 24: 19-31.(Defends the idea mentioned in §3 that methodological judgments are empirical.)

McCloskey, D. (1985) *The Rhetoric of Economics*, Madison, WI: University of Wisconsin Press.(A sample of the literature arguing that the failure of positivism shows that judgements about good science are conventional.)

Passmore, J. (1957) *A Hundred Years of Philosophy*, Harmondsworth: Penguin. (Historical material on logical positivism.)

Popper, K. (1962) *Conjecture and Refutations*, New York: Basic Books.(Outlines Popper’s falsificationism and briefly applies it to Marx and Freud.)

Quine, W. van O. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Presents the arguments mentioned in §2 that there can be no science of interpretation.)

Suppe, F. (1977) *The Structure of Scientific Theories*, Urbana, IL: University of Illinois Press.(Surveys key criticisms of logical positivism.)

Positivism, Russian

Positivism in Russia was not a separate, well-defined philosophical school but, rather, a broad, multidisciplinary current of thought, characterized by a cult of 'positive science', commitment to scientific, empirical methods and rejection of the metaphysical tradition in philosophy. As a rule, Russian positivists sympathized with materialism, although distanced themselves from the metaphysical assumptions of materialist philosophy. Their philosophical aspirations were usually limited to the investigation of specific problems, but their optimistic belief in the power of science could push them in the direction of an all-embracing 'scientific philosophy'.

In accordance with the naturalistic evolutionism of the second half of the nineteenth century, scientific philosophy involved the conception of objective laws of development and biological and/or sociological relativism - both incompatible with the ethicist standpoint. Many Russian thinkers, especially the so-called 'subjective sociologists', tried to combine positivist scientism with ethicism, claiming an independent status for moral values. Their efforts, however, could be successful only at the expense of abandoning the scientific rigour demanded by consistent positivists. Hence their views were semi-positivist rather than positivist. Through undermining the arrogant self-confidence and monopolistic claims of the crude forms of positivistic scientism, they paved the way for the outspoken revolt against positivism at the end of the 1890s.

1 Dogmatic and critical positivism

The founder of positivism, Auguste Comte, became known in Russia as early as the late 1840s. His critique of metaphysics and his conception of the three stages of human development (theological, metaphysical and positive) influenced Russian radicals of the 1860s, especially D. Pisarev (see Russian Materialism: 'the 1860s'), and Russian populists of the 1870s, especially Mikhailovskii (see Lavrov, P.L.; Mikhailovskii, N.K.). For the latter Comte was more palatable than Herbert Spencer because he recognized the validity of a 'subjective method' in sociology and did not claim that a positivist approach to social philosophy was incompatible with value judgments.

The most consistent adherent of Comtian positivism in Russia was Grigory Vyrubov (1843-1913), active mostly in France, where he became a personal friend of Littré and a co-founder of *La Philosophie positive*, the chief organ of French positivism. His philosophical standpoint, expressed most clearly in his article 'Le certain et le probable, l'absolu et le relatif' (in the first volume of *La Philosophie positive*, 1867) is thoroughly dogmatic. Its basic proposition is that scientific knowledge can lay claim to absolute truth and that the opposition between the absolute and the relative is in itself relative, because anything real that can be translated into the language of scientific laws can be called absolute. Following Comte, Vyrubov dismissed not only classical ontology but also epistemology, for which he proposed to substitute the methodology of individual sciences. Hence he was very hostile to any sign of a renewal of interest in Kantian criticism (see Kant, I. §4).

A different type of positivism was represented by Vladimir Lesevich (1837-1905). He accused Vyrubov of a naïve belief that all problems could be solved by science and postulated that positivism should concentrate on elaborating its own theory of knowledge. This required a rapprochement between positivists and Neo-Kantians. Positivism, Lesevich maintained, should become a '*critical* philosophy of reality'. To achieve this, positivists should learn from Kantian epistemological criticism without accepting the possibility of a priori knowledge.

Lesevich expounded these views in the 1870s. A decade later he drew near to the empiriocriticism of Ernst Mach and became the chief Russian representative of the 'second positivism', emphasizing the importance of epistemological reflection. On the other hand, he argued that epistemology should cease to exist as a distinct specialized discipline separated from science. In his book *What is a Scientific Philosophy?* (1891) he advocated the view that the task of 'scientific philosophy' was to make science 'philosophical', instead of just making philosophy 'scientific'.

In his theory of the social sciences Lesevich tried to mediate between positivism and the populist 'subjective method'. Lavrov and Mikhailovskii, he argued, were right in their criticism of Spencer's 'objectivist' approach to social evolution. Epistemological reflection shows that only the abstract sciences can make use of purely objective methods; the applied sciences are directed towards activity, and therefore have to solve problems of value judgments and value implementation. Many arguments for this line of reasoning were to be found, of course, in the

works of the Neo-Kantians (see Neo-Kantianism, Russian §2).

2 Positivism and psychology

In the positivists' discussions on 'scientific philosophy' much attention was paid to psychology, which was often confused with epistemology. Vyrubov simply reduced epistemology to psychology, and even Lesevich was convinced that epistemology was largely based on psychological data.

The most extreme representative of positivistic 'psychologism' in the interpretation of the classical problems of philosophy was a professor at Moscow University, Matveii Troitskii (1835-99). His *Nemetskaia psikhologiia v tekushchem stoletii (German Psychology of the Current Century)*, (2 volumes, 1867), was a vehement attack on the philosophical bias of German psychology, written from the point of view of English empirical, associationist psychology. His chief work, entitled *The Science of the Spirit* (2 volumes, 1882), was an attempt to provide a psychological interpretation of logic.

A more sophisticated version of psychologism was advocated by a liberal historian and philosopher, Konstantin Kavelin (1818-85). His main philosophical work*, Zadachi psikhologii (The Tasks of Psychology)* (1872), was a cautious defence of German Idealism, through interpreting it as a disguised form of psychological investigation which concentrated on the active side of the human psyche. In his theory of the relation between mind and body Kavelin represented a standpoint halfway between psychophysical parallelism and interactionism. This position did not satisfy either the materialists or the consistent idealists and spiritualists. The eminent physiologist Ivan Sechnov attacked Zadachi psikhologii for the treatment of psychic processes as autonomous phenomena, while the Slavophile Iurii Samarin (see Slavophilism) accused Kavelin of exaggerating the soul's dependence on the body and external environment.

Nikolai Grot (1852-99) was initially a more militant and consistent positivist. He studied under Troitskii and his first works - his master's thesis on the psychology of sense perception and his doctorate on the psychological interpretation of logic - displayed a consistently associationist standpoint. He also tried to justify 'naïve realism' in epistemology and deterministic evolutionism in the theory of mind.

In 1886 Grot was appointed professor at Moscow University and shortly afterwards became president of the Moscow Psychological Society. In 1889 he founded the journal *Problems of Philosophy and Psychology*. This gave him an influential position in academic philosophical circles in Russia. In the first issue of the journal he published a programmatic essay on the need of elaborating a distinctively Russian philosophical culture. His own philosophical position had by then undergone a substantial change: he turned away from positivism and openly declared his conversion to metaphysics. This new philosophy of 'monodualism' was to be a synthesis of monism and dualism and a resolution of ontological antinomies in the idea of God. But it was not a return to the speculative constructs of German Idealism: in his article 'What is Metaphysics?' (1900) Grot took care to explain that he meant an inductive metaphysics, attempting to base itself on the data of inner experience and therefore closely linked to psychology. Towards the end of his life he tried to reconcile his new system with positivistic scientism. This was owing to his interest in the energetics of Wilhelm Ostwald, which revived his belief that natural sciences could provide a solution to metaphysical problems.

3 Positivism and sociology

The field in which Russian positivism made its most valuable contribution was sociology. The most philosophical among the positivist sociologists in Russia was Evgenii de Roberty (1843-1915), the follower of Comte, a friend of Vyrubov and a contributor to *La Philosophie positive*. He was one of the first representatives of self-conscious sociologism, that is, a point of view which maintains that social phenomena represent a specific property of organized matter, and therefore cannot be reduced to, or explained by, any wider class of better-known phenomena. Hence sociology should be a new fundamental discipline, completely independent of biology, psychology or economy. De Roberty developed this view in his major work in Russian, entitled *Sociology, Its Basic Task and Methodological Peculiarities* (1880).

Philosophical implications of this standpoint, elaborated mostly in de Roberty's works in French (such as *L'Inconnaissable*, 1889; *Sociologie d'action*, 1908; *Le Concept de la raison et les lois de l'univers*, 1912), included the view that an isolated individual cannot be a thinking being and therefore psychology ought to be

based on sociology, and not vice versa. Another implication was the rejection of the classical definition of truth: objective truth, de Roberty argued, cannot exist, since truth is only a matter of collective experience; the difference between objective and subjective boils down to the difference between individual experience that has been totally dominated by collective experience, and individual experience that has been subjected to collective experience only to a minor extent. Finally, de Roberty's conception of the social roots of knowledge was linked to a philosophical interpretation of sociality as a specific and supreme form of energy - a 'supra-organic' energy, arising out of the interaction of many minds.

A different version of the Comtian tradition was represented in Russia by the outstanding historian of legal institutions, Maksim Kovalevskii (1851-1916). He saw himself as a supporter of Comte and, at the same time, a disciple of Marx (by whom he was befriended). In sociology he was a typical evolutionist, convinced of the uniformity and universal applicability of the basic laws of social development (see his *Sociology*, 2 volumes, 1910). He defined progress as the strengthening of the bonds of human solidarity - the constant expansion of the 'environment of peaceful coexistence', from tribal unity through national patriotism to the solidarity of the whole human race.

In his political views Kovalevskii was a moderate liberal. The liberal belief in universal laws of social progress was questioned by populist sociologists such as Lavrov and Mikhailovskii, who wanted Russia to develop differently from the West. As a rule, this course involved a departure from positivist scientism. The most positivistic among the populist theorists was the historian Nikolai Kareev (1850-1931). In his theoretical works, which include *Osnovnye voprosy filosofii istorii (Fundamental Problems of the Philosophy of History)* (1883) and *Etiudy sotsiologicheskie i filosofskie (Sociological and Philosophical Studies)* (1895) among others, he recognized four types of knowledge: numenological knowledge, being outside the scope of positive science; phenomenological knowledge, dealing with the phenomena themselves; nomological knowledge, determining the general laws; and deontological knowledge which justifies our ideals. He conceived of history as a phenomenological science, dealing with individual events; the task of sociology was to formulate general laws of historical development. The existence of such laws, however, did not exclude, in Kareev's view, the importance of individual choices; hence the need to supplement sociology by deontological, ethical knowledge. In this way Kareev tried to reconcile the positivist (and Marxist) conception of the objective laws of historical development with populist 'subjective sociology'.

4 The revolt against positivism in social philosophy

At the turn of the century Russian philosophy and social thought underwent a radical reorientation, known as the revolt against positivism. A peculiarly important role in this revolt was played by a group of thinkers (Pëtr Struve, Nikolai Berdiaev, Sergei Bulgakov and a few others) who wanted to oppose positivistic scientism (as leaving no place for moral ideals) and, at the same time, to resist populist 'subjectivism' (as containing an element of relativism and thus depriving moral ideals of objective validity). They realized this double task by embracing transcendental (Neo-Kantian) idealism, and soon developed it into transcendent (metaphysical) idealism; the final stage was an openly religious philosophy (see Russian Religious-Philosophical Renaissance §5).

Despite defining their main adversary as positivism, the thinkers in question did not engage in polemics against classical Western positivists and tended to ignore the adherents of positivism in Russia. Instead, they concentrated on the critique of Marxism, as interpreted by the theorists of the Second International. It was so because most of them were former Marxists and, more importantly, because the entire Russian intelligentsia saw the last word of the positive science of society in the 'scientific socialism' of Marx, and not in the theories of Comte or Spencer. In his populist phase, before becoming 'the Father of the Russian Marxism', Georgii Plekhanov defined Marxism in the Comtian fashion: as representing the 'positive stage' in the development of social theory. As a Marxist, he defended the position of a rigorously scientistic objectivism, according to which ethical ideals must be derived from, and coincide with, the objective laws of historical necessity. Owing to this, the initial phase of the revolt against positivism in Russia took the form of a confrontation between deterministic Marxism and Kantian ethical idealism.

See also: Positivism in the social sciences

ANDRZEJ WALICKI

References and further reading

Hecker, J.F. (1916) *Studies in History, Economics and Public Law, vol. 67, Russian Sociology, A Contribution to the History of Sociological Thought and Theory*, New York: Columbia University Press.(Detailed analysis of Kareev's social philosophy and sociology.)

Shkurinov, P.S. (1980) *Pozitivizm v Rossii xix veka (Positivism in Russia in the Nineteenth Century)*, Moscow. (This relatively recent monograph on the philosophies of Russian positivism is the most comprehensive in Russian.)

Sorokin, P.A. (1928) *Contemporary Sociological Theories*, New York.(Detailed presentation of de Roberty's sociological theory.)

Walicki, A. (1980) *A History of Russian Thought from the Enlightenment to Marxism*, trans. H. Andrews-Rusiecka, Stanford, CA: Stanford University Press and Oxford: Clarendon Press.(General account in English.)

Zenkovsky, V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, 2 vols; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, London and New York: Routledge & Kegan Paul and Columbia University Press, 1953, ch. 12 and 25.(General account in English.)

Positivist thought in Latin America

Between 1850 and the 1920s European positivism became a major intellectual movement in Latin America. It asserted that all knowledge came from experience; that scientific thinking was the model for philosophizing and that the search for first causes or ultimate reason typical of religion and metaphysics, was an obsolete mode of thinking. Positivism set out to discover the most general features of experience. Its tasks were to use them to explain and predict phenomena, to develop a social science that would furnish objective grounds for moral choice and to help create the best society possible.

In Latin America positivism became a social philosophy which represented a cogent alternative to romanticism, eclecticism, Catholicism and traditional Hispanic values. It offered the prospect of a secular society in which the knowledge gained from science and industry would bring the benefits of order and progress. It assumed that the social sciences had the power to improve the human condition and it demanded political action. Three main currents were in evidence: autochthonous positivism indigenous to the region and concerned with local social and political issues, social positivism derived from Auguste Comte and stressing the historical nature of social change, and evolutionary positivism influenced by Herbert Spencer and asserting the biological nature of society.

Autochthonous positivism emerged in the 1830s from the influx of liberal ideas which followed the wars of independence fought on the US continent by those wishing to gain freedom from Spain. Urging an intellectual revolution, swift social change and material progress, autochthonous positivism paved the way for European positivism proper. Social positivism appeared around the 1850s and argued for the necessity of educational reforms to solve the continent's problems. It required participation in political life and became a radical force in spite of opposition from supporters of the status quo. By the 1880s evolutionary positivism had steered the movement in a conservative direction in support of laissez-faire *policies, individualism and gradual change.*

1 European positivism

Also called naturalism, scientism and social Darwinism, European positivism was first defined by Saint Simon as the philosophic application of scientific methods to human affairs. Auguste Comte systematically reshaped this insight into a method, a philosophy of science, a theory of development and a polity in *Cours de philosophie positive (Course of Positive Philosophy)* (1830-42).

Positive science does not acknowledge anything beyond the objective relations between phenomena given in experience. The method is to observe these relations and generalize them as coherent principles with which to explain reality. The scientific principles which result from these generalizations depend on the stage of development of the science. According to Comte's law of the three stages, the sciences, societies and the intellect pass through the mythical, metaphysical and positive stages, each representing a characteristic type of explanation of social structures and mental development. The sciences have progressed at different rates: the more general sciences, astronomy and physics, overcome the metaphysical stage earlier than the more specific sciences, such as chemistry or biology, which take longer. The last science to become positive was sociology. Its task was to solve the problems of attaining order and progress in society. In the *Système de politique positive (System of Positive Philosophy)* (1851-4) Comte describes the one society capable of achieving these goals, the sociocracy: a dictatorship of the proletariat and altruistic capitalists, under the moral tutelage of sociologist-priests of the Religion of Humanity.

In England, John Stuart Mill subscribed to positivism in early editions of *System of Logic* (1843) but recanted in *Auguste Comte and Positivism* (1865) because of the social and political implications of Comte's *Système de politique positive (System of Positive Philosophy)*. Herbert Spencer at first accepted Comte's stand on metaphysics and epistemology but eventually disagreed over the deeper nature of change in his essay *Reasons for Dissenting from the Philosophy of M. Auguste Comte* (1864). In *Social Statics* (1851) and later in *First Principles* (1862), Spencer developed an evolutionary principle that predated Darwin's formulation in 1859, which said that organisms are constantly challenged by the environment and other organisms. To survive they must adapt and pass on this adaptation to their offspring. This struggle for survival must produce stronger, more aggressive and cunning individuals, as only the fittest can survive. Like organisms, societies compete for limited resources. To survive they must adapt to internal or external changes. Complex societies which manage to survive, such as those

in Europe, are the fittest. The struggle leaves no room for altruism, a lesson not to be lost on political economy, since only policies that stimulate individual competition but do not interfere with natural development, can increase a society's chances of survival.

2 Latin American positivism

In Latin America, European positivism gained widespread attention as a scientific philosophy that offered solutions to problems which overburdened the continent, such as lack of freedom, political anarchy, venality and economic chaos. Social positivists pointed to the continent's lack of development. Their solution was to reach the positive stage through lay education. It was to de-emphasize rhetoric and literature and teach mathematics and science instead of theology. The resultant effect would be better citizens free from superstition or dogma, motivated by altruism and competent to use the social sciences to reorganize society and develop industry. Evolutionary positivists diagnosed society as sick as a consequence of inferior racial mixtures. Their solutions were to improve the racial stock and stimulate economic growth through European human and capital migration, encourage individual independence and adopt a *laissez-faire* policy.

Both positivist camps justified their views with evidence gathered from the sciences. Furthermore both envisaged that the future result lay in a technological industrial society with a structure and wealth similar to that of the more civilized European nations. Another reason for the favourable reception of European positivism was the presence of an autochthonous variety. The ideas that had given rise to positivism in Europe had been discussed in Latin America before the wars of independence. These ideas incited a rejection of traditional Hispanic values and Catholic thought and fostered an appreciation of science and its potential for guiding reform. Autochthonous positivism has been portrayed as a parallel development, although could also be one of historical opportunism. Indeed, social or evolutionary positivists called select figures autochthonous positivists to capitalize on their prestige and legitimize positivist claims historically.

For details about the progress of positivism in Latin America, the historical circumstances of individual nations have to be taken into account. The most interesting are Argentina, Mexico, Brazil and Chile.

3 Argentina

In Argentina the first expression of a philosophy akin to positivism is found in the works of Domingo Sarmiento and Juan Bautista Alberdi (see Argentina, philosophy in). Anticipating Spencer, Sarmiento (1845) pointed out the importance of ethnic and geographic factors in the formation of political character. For both philosophers humanity progresses from either a bucolic, or barbarian state to a complex urban culture, which is the road to civilization. Later in *Conflictos y armonías en las razas de América (Conflict and Harmony Among the Races in America)* (1883), Sarmiento adopted a racism justified by Spencerian evolutionism. He claimed that Latin America's lack of progress was due to racial inferiority deriving from the mixture of Indian, African and Mediterranean blood. He thought progress could only be made by improving the racial mixture through unrestricted immigration from northern European countries, such as Germany or Denmark. Alberdi's *Bases y puntos de partida para la organización política de la República Argentina (Foundations and Points of Departure of the Political Organization of the Argentine Republic)* (1852) foreshadowed arguments made by social positivists. In it he claimed that Argentina's social problems were the result of outdated ideas and traditions. He thought that reforms should be moral and educational and they should teach individuals self-reliance for self-rule. Alberdi joined Sarmiento's call for increased European immigration. They agreed that to govern was to populate, that is, to educate, improve and civilize. But in order to civilize through population a civilized population was needed in the form of European immigrants. Comtian positivism was introduced in 1872 into the Paraná Teachers College by Pedro Scalabrini and in 1910 at the National University of La Plata by J. Alfredo Ferreira. As positivism became so institutionalized, its influence was to prove long-lasting. It concerned itself mainly with scientific problems and educational improvements and voiced its proposals in journals, such as *La Escuela Positiva* (1894-9) and *El Positivismo: Órgano del Comité Positivista Argentino (Positivism: Organ of the Argentine Positivist Committee)* (1925-35), the latter the organ of the Comité Positivista Argentino.

Evolutionary positivism gained popularity among scientists at the University of Buenos Aires, such as psychiatrist José María Ramos y Mejía, palaeontologist Florentino Ameghino, sociologist Carlos Octavio Bunge and José Ingenieros, although they did not found a school. However, they did exercise considerable influence. An important

exponent was Ingenieros, a psychiatrist by training who also distinguished himself in sociology and philosophy. In *Principios de psicología (Principles of Psychology)* (1911), Ingenieros begins as a committed evolutionist, but admitted the need for improvement because he felt inductivism neglected the speculative aspect of science. As a solution, in his book *Proposiciones relativas al porvenir de la filosofía (Propositions About the Future of Philosophy)* (1918), he proposed an experiential metaphysics that could generate future scientific hypotheses. Ingenieros also found evolutionary ethics unable to account for human ideals and in *Hacia una moral sin dogmas (Towards a Moral Society Without Dogma)* (1917) he pursued an idealism that can only be justified in evolutionary terms.

4 Mexico

As early as 1833 José María Luis Mora anticipated positivism by defining it as a practical attitude opposed to the theoretical and espoused by those who sought newer educational forms to replace the old colonial dogmatism (see Mexico, philosophy in). In 'Revista política de las diversas administraciones que la República mexicana ha tenido hasta 1837' (Political Review of the Many Administrations the Republic has had Since 1837) (1837), Mora expressed dissatisfaction with the anarchy which followed independence and asserted that social progress would finally be made when Mexicans adopted a liberal attitude which guaranteed individual autonomy. The Mexican-American War (1845-8) and the French intervention (1863-7) reinforced this opinion and in 1867 President Benito Juárez called on Gabino Barreda, a student and follower of Comte, to reform education. The goal was an educated class that would lead the country out of economic stagnation and political chaos towards modernization and industrialization. Barreda believed that Mexico had undergone a political revolution and was ready for an intellectual one. This meant giving up Catholicism and liberalism as remnants of the theological and metaphysical stages for a new order akin to the positivist polity. According to Barreda, societies could progress only if they reached a consensus of opinion like that in the sciences. Thus, a scientific education was a necessary step towards progress.

Enactment of the public education law institutionalized Barreda's suggestions at the secondary level. A decade later a new class of intellectuals emerged who were trained along positivist lines. This generation, however, abandoned Comte for Spencer's evolutionism. They founded a political party, los Científicos, which advocated economic solutions and individual initiative over educational reforms. They supported the government in power from 1892-1910. The most interesting exponent of los Científicos was Justo Sierra. In *The Evolution of the Mexican People* (1900), Sierra uses the evolving organism analogy to argue that the cause of Mexico's illness was the sudden transition from a primitive to a militaristic society. The symptoms were revolutions and political atavisms. To restore health, what was needed was gradual evolutionary growth.

5 Brazil

Brazil's nineteenth-century intellectual development differed from the rest of Latin America (see Brazil, philosophy in). Colonized by Portugal, Brazil lacked universities and schools of philosophy. In 1821, without war or revolution, it became an independent monarchy. Social positivism was introduced by Luís Pereira Barreto and Benjamin Constant Botelho de Magalhães to science students around the 1870s at the Military School of Rio de Janeiro. Pereira Barreto eloquently listed in *As Três Filosofias (Three Philosophies)* (1874) the advantages of Comte's positivism and the need to liberate Brazil from religious dogma. Constant had become acquainted with Comte in 1857. He founded the Sociedade Positivista in 1871 and discussed positivism after 1873 in his science and mathematics lectures. Even the most orthodox aspects of social positivism were well received. Miguel Lemos and Raimundo Teixeira Mendes converted to Comte's Religion of Humanity. In 1881 they founded the Templo da Humanidade.

Evolutionary positivists formed a smaller yet influential group among lawyers at the Recife Law School. Its most important exponents were Tobias Barreto, author of *Estudos Alemães (German Studies)* (1882) and his students, Sílvio Romero and Clóvis Bevil qua, the latter one of the authors of Brazil's civil code. They criticized social positivists, particularly the orthodoxy, for suggesting abstract morality rather than concrete economic advances as remedies for Brazil's problems.

Antagonisms between social and evolutionary positivists worsened after 1888 when Constant led the overthrow of the monarchy. In shaping the new republican constitution, social positivists successfully argued for the separation

of Church and state, the abolition of slavery and the establishment of a lay educational system. Even the national flag contained the positivist motto *Ordem e Progresso* (Order and Progress). Their more radical proposals, such as the establishment of a sociocracy and making the Cult of Humanity the official state religion, were blocked by evolutionists and liberals at the national level, although they remained part of certain state constitutions like that of Rio Grande do Sul.

6 Chile

Chile produced several social positivists, but scarcely any evolutionists. The earliest were José Victorino Lastarria and Francisco Bilbao, both students of Venezuelan humanist Andrés Bello while he resided in Chile. Lastarria came across Comte's ideas in 1868, although he had anticipated them in 'Investigaciones sobre la influencia social de la conquista y del sistema colonial de los Españoles en Chile' (Investigation into the Social Influence of the Conquest of Spanish Colonialism in Chile) (1844) in which he held that all historic movement was towards freedom and progress and that the colonial past, full of errors and dogmas, had to be overcome. Bilbao also blamed the colonial past and Catholicism in particular for Chile's lack of development. In *Iniciativa de la América: idea de un congreso federal de las repúblicas (American Initiative: The Idea of a Federal Congress of Republics)* (1856), he suggested the creation of a continental federation of republics which would allow equality, freedom of worship, universal education and would unify the continent under an international law.

As in Brazil, an orthodoxy appeared, led by brothers, Jorge and Juan Enrique Lagarrigue, who had read Comte's works in the 1870s. An eloquent writer, Juan Enrique applied a positivist analysis to issues of national and international peace, the role of women, socialism, the right to strike and the duty of the priesthood of Humanity to mediate between workers and patricians. He attacked parliamentary governments as bankrupt because they rested on metaphysical foundations. This stand clashed with less orthodox positivists such as jurist Valentín Letelier. Letelier had become acquainted with Comte at the same time as the Lagarrigue brothers and admired the positivist programme, but could not accept the religious dogma or the absolutism of the *Système de politique positive (System of Positive Philosophy)* (1851-4). This difference was most evident during the conflict in 1891 between an executive backed by conservatives and a liberal congress. Letelier sided with the congress on liberal grounds while the Lagarrigue brothers argued for a dictatorship of the executive to represent another step towards sociocracy.

7 Positivism elsewhere in Latin America

In other countries positivism was not as widespread. Uruguayan evolutionism was as much a reaction to an idealist and spiritualist philosophy as to political chaos and moral corruption. Positivists tried to improve the nation's morality as well as their material wealth through economic, political and educational reforms. In Cuba, a Spanish colony until 1898, there were no supporters of social positivism, but Spencer had an advocate in Enrique José Varona. Eugenio María de Hostos of Puerto Rico, also developed a sociology along evolutionist lines. Ecuadorean Juan Montalvo, Peruvian Manuel González Prada and José Martí, for example, found positivism a convenient framework in which to cast moral and political ideas without being committed to the doctrine.

8 The decline

Positivism declined towards the end of the nineteenth century. Intellectually it had become stifled and internal squabbles depleted its ranks. Politically it had shifted from being liberal and reformist to become conservative and inactive. Policies were too vague or impossible to implement, or else they were based on insubstantial empirical evidence. Los Científicos fell catastrophically during Mexico's revolution in 1910. Positivism was found to be philosophically inadequate. Alejandro Korn and Francisco Romero eroded its credibility in Argentina. In Uruguay, the philosophies of Carlos Vaz Ferreira and the literary works of José Enrique Rodó offered better alternatives. In Mexico, José Vasconcelos and Antonio Caso pointed to the doctrinal poverty of positivism. Eventually, vitalism, Marxism, existentialism, Platonism and a general return to classical philosophy offered more articulate views. By the 1920s positivism was a term of censure. However, its influence proved subtly pervasive as elements of the doctrine were evident in the views of its strongest detractors. As a legacy it left a trust in technological change and a belief in popular education and economic prosperity.

See also: Anti-positivist thought in Latin America; Positivism in the social sciences

OSCAR R. MARTÍ

References and further reading

Alberdi, J.B. (1852) *Bases y puntos de partida para la organización política de la República Argentina (Foundations and Points of Departure of the Political Organization of the Argentine Republic)*, Buenos Aires: Imprenta Argentina; repr. Buenos Aires: Depalma, 1964.(An argument for improving the political organization of Argentina through European immigration.)

Ardao, A. (1971) 'Assimilation and transformation of Positivism in Latin America', in R.L. Woodward (ed.) *Positivism in Latin America*, Lexington, MA: Heath, 11-16.(An excellent account of the different modes of positivist thought throughout Latin America.)

Barreto, T. (1882) *Estudos Alemães (German Studies)*, Recife: Typografia Central, 2nd edn, Rio de Janeiro: Laemmert, 1891, repr. Rio de Janeiro: Editora Record, 1991.(An exposition of European philosophy and evolutionism.)

Bilbao, F. (1856) *Iniciativa de la América: idea de un congreso federal de las repúblicas (American Initiative: The Idea of a Federal Congress of Republics)*, Paris: D'Aubusson y Kugelmann. (The author suggests the creation of a continental federation of republics to unify the continent under an international law.)

Comte, A. (1830-42) *Cours de philosophie positive (Course of Positive Philosophy)*, Paris: Bachelier, repr. in A. Comte, *Œuvres d'Auguste Comte*, Paris: Editions Anthropos, 1968, vols 1-6.(The main formulation of Comte's positivism, grounded in the study of the philosophy and history of science.)

Comte, A. (1851-4) *Système de politique positive (System of Positive Philosophy)*, Paris: Librarie Mathias, repr. in A. Comte, *Œuvres d'Auguste Comte*, Paris: Editions Anthropos, 1968, vols 7-10.(An attempt to ground science and progress on a religious and political system.)

Crawford, W.R. (1963) *A Century of Latin American Thought*, revised edn, Cambridge, MA: Harvard University Press.(A classic work.)

Cruz C.J. (1964) *A History of Ideas in Brazil*, trans. S. Macedo, Berkeley, CA: University of California Press. (Good account of the development of positivism in Brazil.)

Davis, H.E. (1972) *Latin American Thought: A Historical Introduction*, New York: The Free Press.(A general historical treatment.)

Ingenieros, J. (1911) *Principios de psicología (Principles of Psychology)*, Buenos Aires, repr. in *Obras Completas*, vol. 6, Buenos Aires: Elmer Editor, 1957.(An important reduction of psychology to evolutionary biology.)

Ingenieros, J. (1917) *Hacia una moral sin dogmas (Towards a Moral Society Without Dogma)*, Buenos Aires: Losada, 1961.(An attempt to ground ethics on idealism and evolutionary theory.)

Ingenieros, J. (1918) *Proposiciones relativas al porvenir de la filosofía (Propositions About the Future of Philosophy)*, Buenos Aires: Losada, 1960.(A programme to define philosophy along scientific positivist lines.)

Jaksić, I. (1989) *Academic Rebels in Chile*, Albany, NY: State University of New York Press.(An excellent treatment of academic philosophy in Chile.)

Kilgore, W.S. (1969) 'The Development of Positivism in Latin America', *Interamerican Review of Bibliography* 19: 133-45.(A brief assessment of positivism.)

Lastarria, J.V. (1844) 'Investigaciones sobre la influencia social de la conquista y del sistema colonial de los Españoles en Chile' (Investigation into the Social Influence of the Conquest of Spanish Colonialism in Chile), in *Discursos Académicos (Academic Discourses)*, Santiago de Chile: Imprenta del Siglo, repr. in *Obras Completas de Don J.V. Lastarria*, vol. 7, Santiago de Chile: Imprenta Barcelona, 1906.(A negative assessment of the effects of Hispanic values in Chile.)

Lenzer, G. (ed.) (1975) *Auguste Comte and Positivism: The Essential Writings*, New York: Harper Torchbooks.(A good collection of Comte's writings).

Martí, O.R. (1983a) 'Gabino Barreda and Moral Philosophy', *Aztlán* 14: 373-403.(A more detailed treatment of Barreda's positivism and moral philosophy.)

Martí, O.R. (1983b) 'Auguste Comte and the positivist utopias', in E.D.S. Sullivan (ed.) *The Utopian Vision*, San Diego, CA: San Diego State University Press. (A treatment of Comte's utopian views and its worldwide influence.)

Martí, O.R. (1989) 'Sarmiento y el positivismo' (Sarmiento and Positivism), *Cuadernos Americanos (American Notebooks)* 13: 142-54.(An examination of the claim that Sarmiento held views reflecting autochthonous positivism, published in a distinguished journal on Iberian and Iberoamerican culture, literature, philosophy, politics and the arts.)

Mill, J.S. (1843) *A System of Logic, Ratiocinative and Inductive*, London: J.W. Parker; repr. New York: Longmans, Green, 1961.(A logic text that endorses the positivist programme of reducing deductive to inductive logic.)

Mill, J.S. (1865) *Auguste Comte and Positivism*, Ann Arbor, MI: Ann Arbor Paperbacks, 1973.(Mill's criticisms of Comte's philosophy.)

Mora, J.M.L. (1837) 'Revista política de las diversas administraciones que la República mexicana ha tenido hasta 1837' (Political Review of the Many Administrations the Republic has had since 1837), in *Obras Sueltas*, Paris: Librería de Rosa, repr. Mexico: Porrúa, 1963.(A criticism of political developments in Mexico.)

Peel, J.D.Y. (ed.) (1972) *Herbert Spencer On Social Evolution*, Chicago, IL: University of Chicago Press.(A collection of Spencer's most important writings.)

Pereira Barreto, L. (1874) *As Três Filosofias (Three Philosophies)*, São Paulo, repr. in E. Spencer Macial de Barros (ed.) *Obras filosóficas*, São Paulo: Editorial Grijalbo, 1967.(An important exposition of positivism in Brazil.)

El Positivismo: Órgano del Comité Positivista Argentino (Positivism: Organ of the Argentine Positivist Committee) (1925-35), Buenos Aires: Comité Positivista Argentino.(Argentinian journal which published articles with a Comtian positivist orientation.)

Revista Positiva, Filosófica, Literaria, Social y Política (Positivist, Philosophical, Literary, Social and Political Review) (1901-14), Mexico.(Popular journal in its day.)

Sarmiento, D.F. (1845) *Life in the Argentine Republic in the Days of the Tyrants; or, Civilization and Barbarism*, New York: Hafner, 1974.(An autochthonous positivist's classic in which he examines the social ills which affect Argentina.)

Sarmiento, D.F. (1883) *Conflictos y armonías en las razas de América (Conflict and Harmony Among the Races in America)*, Buenos Aires: S. Ostwald.(An attempt to trace Latin American shortcomings to a mixture of inferior races.)

Sierra, J. (1900) *The Political Evolution of the Mexican People*, Austin, TX: University of Texas Press.(A classic analysis of Mexican history by an important evolutionary positivist.)

Simon, W.M. (1963) *European Positivism in the Nineteenth Century: An Essay in Intellectual History*, Ithaca, NY: Cornell University Press.(A history of European positivism).

Spencer, H. (1851) *Social Statics*, London: John Chapman, repr. New York: Robert Schalkenbach Foundation, 1995.(An early formulation of sociology.)

Spencer, H. (1862) *First Principles*, London: Williams & Norgate, repr. in T.W. Hill (ed.), Westport, CT: Greenwood Press, 1976.(An important philosophical work that tries to structure philosophy along evolutionary principles.)

Spencer, H. (1864) *Reasons for Dissenting from the Philosophy of M. Auguste Comte and Other Essays*, Berkeley, CA: Glendessary Press, 1968.(Spencer's criticisms of Comte's philosophy.)

Woodward, R.L. (ed.) (1971) *Positivism in Latin America*, Lexington, MA: Heath.(An adequate anthology in English.)

Zea, L. (1968) *El Positivismo en México*, Mexico: Fondo de Cultura Económica; trans. J. Schulte, *Positivism in Mexico*, Austin, TX: University of Austin Press, 1974.(An excellent account of positivism in Mexico.)

Zea, L. (1981) *El Pensamiento Positivista Latinoamericano (Latin American Positivist Thought)*, Caracas: Biblioteca Ayacucho.(The best collection of Latin American positivist writings.)

Possible worlds

The concept of Possible worlds arises most naturally in the study of possibility and necessity. It is relatively uncontroversial that grass might have been red, or (to put the point another way) that there is a possible world in which grass is red. Though we do not normally take such talk of possible worlds literally, doing so has a surprisingly large number of benefits. Possible worlds enable us to analyse and help us understand a wide range of problematic and difficult concepts. Modality and modal logic, counterfactuals, propositions and properties are just some of the concepts illuminated by possible worlds.

Yet, for all this, possible worlds may raise more problems than they solve. What kinds of things are possible worlds? Are they merely our creations or do they exist independently of us? Are they concrete objects, like the actual world, containing flesh and blood people living in alternative realities, or are they abstract objects, like numbers, unlocated in space and time and with no causal powers? Indeed, since possible worlds are not the kind of thing we can ever visit, how could we even know that such things exist? These are but some of the difficult questions which must be faced by anyone who wishes to use possible worlds.

1 Pros and cons

Although known to Leibniz, it was not until the middle of the twentieth century that possible worlds came to occupy the attention of a significant number of philosophers and logicians (see Leibniz, G.W. §3). While this was initially due to the discovery that possible worlds permitted the formulation of a model-theoretic semantics for modal logic, it gradually became clear that possible worlds could be used to shed light on many other notions which had been regarded as problematic: intensional logic, counterfactuals, propositions and properties are but some of the areas illuminated by possible worlds. Indeed, some have even argued that the existence of possible worlds is a tenet of our ordinary common-sense beliefs. It is uncontroversially true that there are many ways the world could have been - but what are these many ways but possible worlds under another name?

However, although possible worlds may provide solutions to old problems, they raise new difficulties of their own. Since possible worlds are not the kind of thing with which we can causally interact, how do we know that there are such things, and how have we come to possess the kind of modal knowledge we credit ourselves with? Then there is the problem of transworld identity: how one and the same object can exist at different possible worlds. The problem arises for the following reason: not only is it possible that there might have been people who were twelve feet tall, Quine himself might have been twelve feet tall. In possible worlds terms, this means that there is a possible world at which Quine is twelve feet tall, apparently making Quine exist in more than one world.

When it comes to questions about their metaphysical nature - what kinds of things possible worlds are - there is very little agreement. It is agreed that possible worlds must be capable of representing many different ways the world could have been, and there can be no possible worlds in which grass is both green and red; but whether possible worlds are abstract or concrete, whether they exist independently of us or are our own creation, whether they have structure or are simples - are all contentious issues.

2 Extreme realism

For the extreme realist, all possible worlds are on a par: there is no distinction in kind between the actual and the possible. All that marks off the actual world from the infinitely many merely possible worlds is that the actual world is the world we happen to inhabit. Just as the actual world contains flesh and blood human beings, so some merely possible worlds contain flesh and blood human beings. Just as the actual world is made up of concrete spatio-temporally extended objects, so other possible worlds are made up of concrete spatiotemporally extended objects.

At first sight it appears that the extreme realist must solve the problem of transworld identity by having one and the same object literally appearing in more than one world. But there is an ingenious alternative. Quine might exist at a possible world w not because he is a part of w, but because w contains some other person (call him ‘Kwine’) who *resembles* Quine. If Kwine and Quine share certain properties, if they lead similar lives and have similar parents then Kwine is said to be Quine’s *counterpart*. It is possible for Quine to be twelve feet tall if there exists some world which contains a counterpart of Quine, and that counterpart is twelve feet tall.

Extreme modal realists use neither modal nor intensional concepts in their theory. Accordingly, they can use possible worlds to analyse non-circularly the concepts of necessity, of counterfactuals, of propositions and of properties. Since the number of different concepts taken as primitive is few, an advantage of this theory is that it is ideologically parsimonious.

Yet for all its strengths, extreme modal realism has few proponents. Although ideologically parsimonious, the theory is as ontologically unparsimonious as possible: for any kind of object there could be, be it a centaur, ghost or unicorn, there is some world which contains such an object. Moreover, the very idea that there is an infinite number of concrete possible worlds, all on an equal footing, simply defies belief. The view that there really are non-actual Centaurs, ghosts and unicorns goes against our strongly held common-sense beliefs.

3 Combinatorialism

The combinatorialist takes possible worlds as recombinations of the fundamental elements of the actual world. For any fundamental property *F* and any atomic object *a*, the combinatorialist maintains that it is possible that *a* instantiate *F*. *Any* collection of such states of affairs counts as a possible world.

The restriction to atomic objects and fundamental properties is needed to rule out impossible recombinations such as an object's being simultaneously both round and square, or red and green. It is a crucial part of the combinatorialist's idea that such a restriction succeed in ruling out all impossibilities. But there are reasons to think it does not. For there may be fundamental properties which are nevertheless mutually exclusive. For example, having a mass of one gram and having a mass of two grams may both be fundamental properties - but nothing can be simultaneously both one gram and two grams.

Not only are there recombinations which are not possible, there are possibilites which are not recombinations. For it is possible that there be things which do not actually exist, and possible that there have been different fundamental properties not instantiated in the actual world. Such possibilities cannot be constructed by recombining only actually existing objects and properties.

4 Possible worlds as novels

On this view, worlds are nothing more than consistent works of fiction which represent different possible states of affairs in the same way novels do. It is possible that grass is red if there is some consistent novel which contains the sentence 'grass is red', or some collection of sentences which entail that grass is red. Since novels can contain names for actually existing objects, any novel containing the sentence 'Quine is twelve feet tall' represents Quine as being twelve feet tall - thus dealing with the problem of transworld identity.

But though this theory neither inflates our ontology nor defies belief (there are few who doubt the existence of novels!), as it stands it simply fails to do justice to the facts of modality. For not any book can count as a possible world. For instance, any novel containing the two sentences 'Grass is red' and 'Grass is green' fails to represent a way in which the world *could* have been. We could avoid this problem by identifying possible worlds with those novels which *could have been true* - but there is a drawback. If the concept of possibility is being *used* in saying what possible worlds are, the concept of possibility cannot be analysed in terms of possible worlds on pain of circularity.

Not only are there novels which do not represent possible worlds, there are some possible worlds which are not represented by any novel. For there are infinitely many different lengths an object could be. Accordingly, there are an infinite number of different possible worlds. Yet there is only a finite number of novels. Moreover, there are merely possible worlds which contain as much complexity and detail as the actual world around us, yet we mortals have no hope of writing a novel describing such a world in such fine detail. Some possible worlds appear to elude our descriptive capabilities.

If we are to have any hope of reducing possible worlds to novels, then our novels cannot be the kinds of things we actually read or write. Instead, our novels must be identified with sets of sentences in some idealized abstract language, a language which, though not capable of being written or spoken, is capable of representing many more possible worlds. One way to implement this idea is to construct a language the terms and predicates of which are actual objects and properties and the sentences of which are set-theoretical constructions thereof. On this scheme,

the ordered pair $\langle F, a \rangle$ is interpreted as saying that *a* is *F*. Possible worlds can then be identified with consistent sets of sentences.

Though such a powerful language enables one to represent an infinite number of different possibilities, it is still a moot point whether *all* the possible worlds can be represented. Moreover, the increase in power comes at a price. For possible worlds are now identified with mathematical objects: sets of ordered n-tuples. Unlike novels, the existence of mathematical objects *is* a contentious issue.

Finally, some fear that this theory does not provide an objective theory of modality. For sentences have the meanings they do in virtue of the way *we* interpret them: had there been no humans, there would have been no meaningful sentences. Accordingly, possible worlds owe their existence to us. This is difficult to square with the view that whether or not something is possible is a matter which is independent of human thoughts and beliefs. ‘Possibly *P*’ cannot be equivalent to ‘There is a possible world in which *P*’ if the former is objectively true but the latter holds only subjectively.

5 Moderate realism

Instead of taking possible worlds as sets of sentences, perhaps possible worlds should be constructed out of the *meanings* of sentences. Perhaps possible worlds should be taken as consistent sets of *propositions*. Obviously, such entities must be understood as existing independently of our own thought and language if worlds are not to owe their existence to us. (Some variants of this theory replace propositions with *states of affairs* or *abstract properties*.)

This theory faces some familiar problems. As above, worlds are abstract. As above, since it is only those sets which *could* have been true which form the possible worlds, the theory cannot analyse possibility without circularity. Of yet greater concern is the fact that this theory takes the notion of proposition as primitive. Admitting such things as a new fundamental kind of object is ontologically costly, and some find the idea that there is an infinite number of abstract representations ‘out there’ every bit as ludicrous as extreme realism. Moreover, the notion of proposition had come under a lot of fire, and one of the more attractive features of possible worlds was that they promised to throw light upon these peculiar creatures. Yet how can this promise be fulfilled by a theory which itself takes this problematic notion as primitive?

See also: Abstract objects; Intensional entities; Intensional logics; Modal logic; Modal logic, philosophical issues in

JOSEPH MELIA

References and further reading

Armstrong, D.M. (1989) *A Combinatorial Theory of Possibility*, Cambridge: Cambridge University Press.(A defence of combinatorialism.)

Bricker, P. (1987) ‘Reducing Possible Worlds to Language’, *Philosophical Studies* 52: 331-55.(The author scrutinizes the position that possible worlds are novels but concludes that there are certain possibilities which lie forever beyond our descriptive resources.)

Lewis, D. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.(An advanced, yet clear and thorough examination of the metaphysics of possible worlds. The author makes a surprisingly powerful case for extreme realism.)

Loux, M.J. (1979) *The Possible and the Actual*, Ithaca, NY: Cornell University Press.(A comprehensive collection of essays covering all theories of possible worlds.)

Plantinga, A. (1989) *The Nature of Necessity*, New York: Clarendon Press.(Aside from a careful and clear defence of moderate realism, this book includes an interesting discussion of essential properties and uses possible worlds to reassess medieval theological arguments for the existence of God.)

Rosen, G. (1990) ‘Modal Fictionalism’, *Mind* 99 (395): 327-54.(Rather than identifying each possible world with a fiction, Rosen suggests that we view the theory that there are possible worlds as itself a fiction.)

Post, Emil Leon (1897-1954)

Emil Post was a pioneer in the theory of computation, which investigates the solution of problems by algorithmic methods. An algorithmic method is a finite set of precisely defined elementary directions for solving a problem in a finite number of steps. More specifically, Post was interested in the existence of algorithmic decision procedures that eventually give a yes or no answer to a problem. For instance, in his dissertation, Post introduced the truth-table method for deciding whether or not a formula of propositional logic is a tautology.

Post developed a notion of 'canonical systems' which was intended to encompass any algorithmic procedure for symbol manipulation. Using this notion, Post partially anticipated, in unpublished work, the results of Gödel, Church and Turing in the 1930s. This showed that many problems in logic and mathematics are algorithmically unsolvable. Post's ideas influenced later research in logic, computer theory, formal language theory and other areas.

1 Life and early work

Emil Leon Post was born in Augustow, Poland. He emigrated with his family in 1904 to New York City, where he lived for the rest of his life. He received a BS from City College in 1917. As a graduate student in mathematics at Columbia University, he studied Whitehead and Russell's *Principia Mathematica* with Cassius J. Keyser. He wrote his dissertation under Keyser's guidance and received his Ph.D. in 1920. He received a postgraduate Proctor Fellowship to Princeton for 1920-1, where he studied with the geometer and postulate theorist Oswald Veblen. To this period can be dated the birth of many of his most important ideas. He taught mostly in New York City high schools until 1935, when he was appointed to the faculty of City College where he remained until his death. During his adult life, Post struggled constantly with manic-depressive illness. To a great extent, this accounts for the incomplete development of many of his ideas. But he also delayed publication while attempting to work out the most general possible results.

In his dissertation (published in 1921) Post carried out one of the first metamathematical studies of a system of logic. This was the propositional logic of *Principia Mathematica*. In addition to proving the completeness, consistency and decidability of this system, he showed that it is complete in a sense that is now called 'Post completeness', that is, the addition to the axiom set of any formula not provable from it produces an inconsistent system. (P. Bernays had proved completeness and consistency in his *Habilitationsschrift* of 1918, but only published these results in 1926.)

Post also showed, going beyond Sheffer (1913), that each truth-function can be defined by a combination of the negation ($\sim$) and disjunction ($\vee$) truth-functions of *Principia Mathematica*. In a later monograph (1941), he showed that a set of truth-functions is truth-functionally complete, as $\{\sim, \vee\}$ is, if and only if for each of five types of truth-function at least one member of the set does not belong to that type (see Pelletier and Martin 1990). At the end of his dissertation, Post also showed how to develop an *m*-valued propositional logic, as opposed to the traditional logic with its two values of True and False (see Many-valued logics).

2 Canonical systems

In his year as a Proctor Fellow at Princeton (1920-1), Post's main research was into 'generated sets' of strings on a finite alphabet, that is, inductively defined classes of strings. Post studied three categories of such 'canonical systems'. The systems he labelled 'canonical system *C*' are now known as 'Post production systems'. This laid the foundations for the field that is now called the theory of formal languages (see Formal languages and systems). Each canonical system specifies a class of rule sets. These rule sets define languages, conceived as infinite sets of finite strings of characters on a finite alphabet. This approach is analogous to the specification of the class of recursive functions as those which result from a finite number of applications of a certain base group of number-theoretic functions. What is commonly called 'Church's thesis' is the proposition that the number-theoretic functions which can be computed by any means at all are exactly the ones that are the recursive functions. The converse is obvious (see Church's thesis). Analogously, we can say that Post's thesis is that any finitely specifiable language is generated by the rules of some canonical system. He showed that any set of strings that can be generated using one of his canonical systems can be generated in either of the other two systems. Post

was able to show that a decision procedure for his canonical systems furnishes a decision procedure for the first-order logic of *Principia Mathematica*. He sketched, but did not complete, a proof of the converse. This led him to conjecture that Hilbert's *Entscheidungsproblem*, that is, the decision problem for first-order logic, is unsolvable. This was more than ten years before Church's proof (1936) of its unsolvability (see Church's theorem and the decision problem).

3 Recursion theory

Post's canonical systems were later shown to be equivalent to other formulations of the class of computable procedures, such as lambda-definability, recursive functions and Turing machines. Using this equivalence, a set of formulas is called 'recursive' if there is a decision procedure for the set using any of the methods. It is called 'recursively enumerable' if there is a procedure using any of these methods which successively lists the elements of the set. Post's 'Recursively Enumerable Sets of Positive Integers and Their Decision Problems' (1944) is foundational for modern recursion theory. It was there that he first articulated the 'fundamental theorem' of recursion theory; that a set is recursive if and only if both the set and its complement are recursively enumerable.

In this paper, Post also shows that there is a recursively enumerable but undecidable set of positive integers for which the decision problem has the highest 'degree of unsolvability'. If there were a decision procedure for a set of this degree, there would also be a decision procedure for any other undecidable set. Post states the question, known as Post's Problem, of whether there are recursively enumerable but undecidable sets of positive integers which have lower than this maximum degree of unsolvability. The existence of sets of such lower degrees was shown independently in Muchnik (1956) and Friedberg (1957).

4 Other work

Post's methods gained the attention of mathematicians when he showed (1947) that there is no algorithmic method for solving the 'word problem of Thue' which dated from 1914. In this paper, he used the Turing machine formulation of computability. In the course of doing so, he provided a critique of details of Turing's paper. This is not surprising, since in 1936 he had independently formulated a notion of computation very much like that of Turing (see Turing machines).

In a diary entry Post wrote, 'I study Mathematics as a product of the human mind and not as absolute'. He believed that he was investigating a natural boundary between what he called the 'productive' and the 'creative' capacities of the human mind. For Post, the incompleteness of logical systems of proof showed that 'The logical process is essentially creative', that is, that the human capacity to know is not reducible to a system (see Appendix to Post 1965).

See also: Computability theory; Lambda calculus

MICHAEL SCANLAN

List of works

Post, E.L. (1993) *Solvability, Provability, Definability: The Collected Works of Emil L. Post*, ed. M. Davis, Basle: Birkhäuser.(This is a complete collection of work published in Post's lifetime, plus the posthumously published Post (1965). The editor's introduction is a comprehensive survey of Post's life and work.)

Post, E.L. (1921) 'Introduction to a General Theory of Elementary Propositions', *American Journal of Mathematics* 43: 163-85; repr. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematics, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 264-83.(Post's dissertation; requires introductory knowledge of mathematical logic.)

Post, E.L. (1936) 'Finite Combinatory Processes - Formulation I', *Journal of Symbolic Logic* 1: 103-5.(A conceptual account of the procedures now called Turing machines. Quite readable for anyone familiar with the latter.)

Post, E.L. (1941) *The Two-Valued Iterative Systems of Mathematical Logic*, Annals of Mathematics Studies, No. 5, Princeton, NJ: Princeton University Press.(Difficult even for specialists; should be read after Pelletier and Martin (1990).)

Post, E.L. (1944) 'Recursively Enumerable Sets of Positive Integers and Their Decision Problems', *Bulletin of the American Mathematical Society* 50: 284-316.(A beautiful paper; the most readable of Post's papers for anyone who has done an introductory mathematical logic course.)

Post, E.L. (1947) ‘Recursive Unsolvability of a Problem of Thue’, *Journal of Symbolic Logic* 12: 1-11. (Specialized; requires familiarity with Turing machines.)

Post, E.L. (1965) ‘Absolutely Unsolvable Problems and Relatively Undecidable Propositions - Account of an Anticipation’, in M. Davis (ed.) *The Undecidable*, Hewlett, NY: Raven, 1965, 338-433.(Post’s own account of the foundational significance of his canonical systems. A difficult but important paper.)

Post, E.L. (1990) ‘The Modern Paradoxes’, ed. I. Grattan-Guinness, *History and Philosophy of Logic* 11: 85-91. (This essay on Russell’s paradox is edited from Post’s manuscripts and is not included in the *Collected Works* (1993).)

References and further reading

Bernays, P. (1926) ‘Axiomatische Untersuchung des Aussagenkalküls der *Principia Mathematica*’ (Axiomatic Investigation of the Propositional Calculus of *Principia Mathematica*), *Mathematische Zeitschrift* 25: 305-20. (In addition to proving consistency and completeness of the propositional logic axioms of *Principia Mathematica*, about half of this article is devoted to independence proofs for those axioms.)

Church, A. (1936) ‘A Note on the *Entscheidungs-problem*’, *Journal of Symbolic Logic* 1: 40-1, 101-2. (Church’s classic article on the nonexistence of a decision procedure for first-order logic.)

Davis, M. (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Hewlett, NY: Raven.(This important collection includes papers by Gödel, Church, Turing, Rosser and Kleene, besides papers by Post. Together these provide a comprehensive survey of the heroic age of computability theory.)

Friedberg, R. (1957) ‘Two Recursively Enumerable Sets of Incomparable Degrees of Unsolvability’, *Proceedings of the National Academy of Sciences (USA)* 43: 236-8.(Independently of Muchnik (1956), Friedberg shows that there are degrees of recursive unsolvability. Uses the important priority method of proof.)

Grattan-Guinness, I. (1990) ‘The Manuscripts of Emil L. Post’, *History and Philosophy of Logic* 11: 77-83. (Describes the collection of Post manuscripts kept at the American Philosophical Society in Philadelphia.)

Muchnik, A.A. (1956) ‘Nerazreshimost’ problemy svodimosti teorri algoritmov’ (On the Unsolvability of the Problem of Reducibility in the Theory of Algorithms), *Doklady Akademii Nauk SSSR* 108: 194-7.(Gives the positive answer to Post’s Problem, that is, that there are degrees of reducibility for recursively enumerable, but non-recursive, sets.)

Pelletier, F.J. and Martin, M.N. (1990) ‘Post’s Functional Completeness Theorem’, *Notre Dame Journal of Formal Logic* 31: 462-75.(Gives an updated statement and proof of the main results of Post (1941).)

Sheffer, H.M. (1913) ‘A Set of Five Independent Postulates for Boolean Algebras, With Applications to Logical Constants’, *Transactions of the American Mathematical Society* 14: 481-8.(Shows that the single ‘Sheffer stroke’ operator can be used to define the two basic propositional operators, ‘ $\sim$’ and ‘ $\vee$’, in the propositional logic of *Principia Mathematica*.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(Post carried out one of the first metamathematical studies of a system of logic on the propositional logic of *Principia Mathematica*.)

Postcolonial philosophy of science

Are there laws of nature that today's modern sciences are ill-designed to discover? Does the universal use of these modern sciences require their value-neutrality, or are their social values and interests an important cause of their universality? What resources for scientific knowledge can other cultures' science projects provide?

Such questions are raised by recent postcolonial global histories that focus analyses on the role of European expansion in the advance of modern science and in the decline of other cultures' science traditions. These accounts challenge philosophers to re-evaluate unsuspected strengths in other scientific traditions and identify modern science's borrowings from them. They also identify European cultural features that have, for better and worse, constituted modern sciences and their representations of nature and thus seek to develop more realistic and useful accounts of the values, interests, methods, universality, objectivity and rationality of science.

1 European values and interests of modern science

Since the 1960s, it has been clear that social values and interests do and must shape the fabric of scientific cultures, their practices and knowledge claims. The great explanatory power of science turns out to have been expanded by historically distinctive values and interests. By the 1980s, the widespread dissemination of postcolonial histories - ones that report the continual exchanges and interminglings between cultures 'from the beginning' - led to accounts of how different cultures' sciences advance and then decline within such global relations. One main focus has been on the cognitive effects on modern sciences of the two-way causal relations between European expansion and the development of modern sciences in Europe, and a second on the benefits of nourishing cultural diversity in the resources brought to explaining nature's regularities. Selin (1997) is a valuable resource for this second project.

Joseph Needham's comparative studies of Chinese and European sciences were an important precursor to the new postcolonial science studies (1969). From such a perspective, he argued that European sciences were indeed superior at explanation, but not for every valuable scientific project, or always for the reasons that historians and philosophers assumed. Moreover, some of their successes, not just their failures, were due to their cultural values and interests. For example, Christian values and interests shaped the conception of 'laws of nature', though Christian belief in the heavenly crystal spheres blocked the development of astronomy. Thus religious beliefs had both positive and negative effects on the growth of scientific knowledge.

Attempting to steer a more effective path than Needham did between the older internalist and externalist histories of science, the comparative science studies of Needham's intellectual heirs try to show how modern science's representations of nature are not culturally neutral: they make the European presuppositions characteristic of their societies. For example, what we know about nature is restricted largely to what the sponsors of modern science have wanted to know. This information is embedded in models, metaphors and narratives for nature's processes that they found culturally and politically congenial. It also both enriched and limited what could be detected by the culture's distinctive ways of organizing scientific work.

Thus scholars such as Blaut, Goonatilake, Hess, McClellan and many of the contributors to Petitjean *et al.* (1992) and Harding (1993) argue that the modern sciences have produced accounts mainly of those aspects of nature that permitted upper and rising classes of Europeans to multiply and thrive, especially through the advance of their military and other expansionist projects. Scientists answered questions about how to improve European land and sea travel, to identify economically useful resources in the colonies, to mine gold, silver and other ores, to manufacture and farm for the benefit of Europeans living in Europe, the Americas, Africa and Asia, to improve their health and occasionally that of the workers who produced profit for them, and to defend the Europeans against their enemies. Expansion turned the world into a laboratory for the sciences in Europe. Nature remains *for* those who own it and have the means of extracting its resources, they argue. Such sciences have not been designed to explain such things as how to preserve fragile environments and scarce resources.

Moreover, it is argued that even if modern sciences bore no such cultural fingerprints, their commitment to value-neutrality would itself mark them as culturally distinctive. Most cultures value not neutrality but their own values and interests and as such are culturally identifiable. Surprisingly, it turns out that abstractness and formality

express distinctive cultural features, not the absence of any culture at all. Of course the cultural specificity of the neutrality ideal does not mark abstractness or formality as a defect, since most values and interests have the potential to enable us to ‘see’ nature in new ways. Instead, it raises questions about the conditions under which a commitment to neutrality advances the growth of knowledge.

2 Universality, objectivity, methods, rationality

Modern sciences are practised and admired around the world. Scientists come from every race, gender, class and culture, and the sciences’ formal languages permit communication (and agreement) between all. Moreover, the laws of nature that modern science articulates predict how nature’s forces will act on people indiscriminately, regardless of race, class, gender and culture. But such facts only support a weak claim to modern sciences’ universality. The familiar ‘strong universality’ thesis argues that the uniqueness of the universality of scientific claims (the ‘fact’ that no other scientific traditions have produced universal claims) is the consequence of an internal, epistemological condition: methods that eliminate cultural fingerprints from scientific representations of nature’s events and processes.

This uniqueness claim is challenged by postcolonial histories that show how the great explanatory power and appeal of modern sciences has been established in part as an empirical consequence of European expansion. Such accounts drive a wedge between the achievement of universality and the requirement of cultural neutrality. They suggest that other, culturally different, sciences could also generate empirically adequate laws of nature, ones that would be consistent with much of the evidence supporting the modern sciences’ representations of nature. They argue that modern sciences are not necessarily the best for discovering all of the laws of nature; they have intrinsic metaphysical and epistemological limitations for, like all others, they are shaped by the cultural interests they represent, by local discursive resources, and by culturally-preferred forms of scientific work. They question that the sciences and technologies that have worked best to advance some groups in the Northern hemisphere will work best to advance the rest of the world’s peoples, or, for that matter, to continue to sustain Northern elites.

Challenges to conventional conceptions of objectivity, method and rationality are also generated by these accounts - ones similar to those raised by gender theorists (see Gender and science; Feminist epistemology). Is the neutrality ideal that is so central to such concepts an obstacle to detecting the kinds of culture-wide values and interests in European sciences that are reported in the comparative studies? The postcolonial accounts started off thinking about modern sciences from the lives of people in cultures that have been disadvantaged by them. Are standpoint epistemologies thus required to explain the emergence of the postcolonial accounts in the first place, and to generate research methods capable of continuing to produce such more comprehensive and less distorted accounts? Should assessments of the rationality of modern sciences continue to be restricted to examining their technical, cognitive cores now that these are understood to be no longer separable from the cultural values and interests that permeate them?

Postcolonial science studies, emerging from both Northern and Southern science scholars, are part of the long tradition of criticism that has enabled the growth of knowledge. They challenge philosophers to consider an updated version of Thomas Kuhn’s famous question: should not single-stream histories of science be a source of phenomena to which theories of knowledge may legitimately be asked to apply?

See also: Objectivity; Rationality and cultural relativism; Scientific method

SANDRA G. HARDING

References and further reading

Blaut, J.M. (1993) *The Colonizer’s View of the World: Geographical Diffusionism and Eurocentric History*, New York: Guilford Press.(Shows historians’ problems with Eurocentric histories, which are the ones assumed by the conventional histories of science upon which philosophies of science draw. Explores relationship between flourishing European cultures and decline of others.)

Goonatilake, S. (1984) *Aborted Discovery: Science and Creativity in the Third World*, London: Zed Press.(Effects of European expansion on Asian and European sciences.)

Harding, S. (ed.) (1993) *The ’Racial’ Economy of Science: Toward a Democratic Future*, Bloomington, IN: Indiana University Press.(Contributors analyse diverse contexts for thinking about how costs and benefits of modern sciences have not always been democratically distributed. Section on ‘Objectivity, Method, and

Nature: Value-Neutral?' draws out central philosophical implications of such undemocratic distributions.)
Harding, S. (1998) *Is Science Multicultural? Postcolonialism, Feminism and Epistemology*, Bloomington, IN: Indiana University Press.(Shows central challenges to conventional philosophies of science and epistemologies that are raised by two decades of multicultural, postcolonial and feminist science studies, and how many of these challenges are to be found also in post-Kuhnian, 'post positivist', philosophy of science.)
Hess, D. (1994) *Science and Technology in a Multicultural World: The Cultural Politics of Facts and Artifacts*, New York: Columbia University Press.(How scientific 'facts and artifacts' are culturally constructed in valuable ways for different cultures in the past and present. Challenges conventional assumptions about immunization of scientific facts from cultural influences.)
McClellan, J.E. (1992) *Colonialism and Science: Saint Domingue in the Old Regime*, Baltimore, MD: Johns Hopkins University Press.(The intimate relations between the development of modern sciences in France and the needs of slave holders through a study of the archives of an ambitious and distinguished French scientific institute in the Caribbean. Challenges conventional understandings of the epistemological-political 'innocence' of 'scientific problems'.)
Needham, J. (1969) *The Grand Titration: Science and Society in East and West*, Toronto, Ont.: University of Toronto Press.(Comparisons and analyses of Chinese and Western sciences by the best-known (scientist-) historian of Chinese sciences. Identifies Eurocentric assessments of European and of Chinese sciences by leading histories of science - ones that philosophies of science assume.)
Petitjean, P., Jami, C. and Moulin, A.M. (eds) (1992) *Science and Empires: Historical Studies about Scientific Development and European Expansion*, Dordrecht: Kluwer.(Forty papers of the more than 100 presented at a UNESCO 1990 conference in which scholars from around the world critically examined the relationship between the flourishing of empires - especially the 500 year old European one up through the late twentieth century - and of scientific and technological traditions. Contributors consistently challenge Eurocentric assumptions about the causes of the successes of European sciences and technologies.)
Selin, H. (ed.) (1997) *Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures*, Dordrecht: Kluwer.(Over 600 scholarly articles review the histories and accomplishments of non-European cultures' scientific and technological traditions. Includes many philosophically focused essays.)

Postcolonialism

The term 'postcolonialism' is sometimes spelled with a hyphen - post-colonial - and sometimes without. There is no strict general practice, but the hyphenated version is often used to refer to the condition of life after the end of colonialism while the non-hyphenated version denotes the theory that attempts to make sense of this condition. The term is regularly used to denote both colonialism and imperialism even though these refer to different historical realities.

Like postmodernism and poststructuralism, postcolonialism designates a critical practice that is highly eclectic and difficult to define. It involves a studied engagement with the experience of colonialism and its past and present effects at the levels of material culture and of representation. Postcolonialism often involves the discussion of experiences such as those of slavery, migration, suppression and resistance, difference, race, gender, place and analysis of the responses to the discourses of imperial Europe, such as history, philosophy, anthropology and linguistics. Since conditions under imperialism and colonialism proper are as much the subject of postcolonialism as those coming after the historical end of colonialism, postcolonialism allows for a wide range of applications and a constant interplay between the sense of a historical transition, a cultural location and an epochal condition. Postcolonialism is seen to pertain as much to conditions of existence in former colonies as to conditions in diaspora. Both are frequently linked to the continuing power and authority of the West in the global political, economic and symbolic spheres and the ways in which resistance to, appropriation of and negotiation with the West's order are prosecuted. However the term is construed, there is as much focus on the discourse and ideology of colonialism as on the material effects of colonial subjugation. Because it has its source in past and continuing oppression, postcolonialism furthermore has affinities with multicultural, feminist, and gay and lesbian studies.

1 Seminal works

It is generally agreed that the single most influential work to define the purview of the term was Edward W. Said's *Orientalism* published in 1978. Drawing on the work of Michel Foucault, Said's main thesis was that the Western academic discipline of Orientalism was strictly speaking a means by which the Orient was produced as a figment of the Western imagination for consumption in the West and also as a means of subserving the ultimate project of imperial domination. Said's main ideas have been both criticized and extrapolated across various disciplines, and it is mainly thanks to his book that what is known as colonial discourse analysis gained coherence. Among studies inspired by his work have been those by Gauri Viswanathan (1990), looking at the introduction of the paradigm of English studies in India and the degree to which it attempted to shape local attitudes to Empire; Patrick Brantlinger (1988), which looks at the forms of imperial ideology as perceivable in the literary writings of the period 1830-1914; Martin Bernal (1987; 1991), which through a careful analysis of the sources of Greek civilization sought to show not only that this was heavily indebted to the influence of a black Egyptian civilization, but also the degree to which the discipline of classical studies in the eighteenth century attempted to obscure this contribution in the dominant disciplinary paradigms that were in use; and Valentin Mudimbe (1988) which showed that the notion of African systems of knowledge had always been governed by a Western form of knowledge.

Another text which is thought to be seminal in the field of postcolonial studies is by Bill Ashcroft *et al*., *The Empire Writes Back* (1989). This book was in many senses responsible for popularizing the term post-colonialism in literary studies, even though, as has been pointed out by Aijaz Ahmad (1995) among others, the term was first used by political scientists in the early 1970s to denote the condition of Third World political systems after decolonization. Before its publication much of what is described under the label of postcolonialism in literary studies was discussed under the rubric of contending terms such as 'New Literatures in English', 'International Literatures', 'Third World Literature' or 'Commonwealth Literature'. However, it was the notion of Commonwealth Literature that *The Empire Writes Back* was most intent on displacing. The writers of the book were of the view that the term showed too close and uncomfortable a relationship to metropolitan Britain, and the term 'post-colonial' was seen as a means of transcending this relationship. Another influential thesis they advanced was that the literature coming out of former colonies had the express purpose of 'writing back' to the empire, and by a variety of textual strategies, of subverting the dominant categories of representation that had circulated from the imperial centre. In this impulse they sought to correct something for which Said's book had been heavily criticized, namely his complete silence on the voices of resistance and/or complicity that could easily

be shown to have entered into a variety of relationships with Orientalist discourse. A third thesis of *The Empire Writes Back* was that postcolonial writing was also seriously contesting the domination of the English language by creating a variety of creolized and hybridized versions of the language. As a means of attending to this phenomenon, the authors proposed the term 'english' to designate this phenomenon; and even though their proposal for the use of this particular term was not readily taken up by other commentators on postcolonial writing, it is evident that the general note of resistance in their critical practice was taken up in numerous articles published in literary journals devoted to the field such as the *Journal of Commonwealth Literature*, *Callaloo*, *Wasafiri*, *Kunapipi* and *World Literature Written in English*, among others.

2 Problems with the term 'postcolonialism'

To gain a sense of the nature of some of the criticisms levelled against the term 'postcolonial', it is important to return to the question of its genesis. Though earlier it was noted it could be traced to the seminal publications of Said and Ashcroft *et al*., such a perspective may obscure the complex and somewhat composite genealogy of the theoretical practices denoted by the term. These can fruitfully be linked to what one critic has noted as the 'wide-ranging retrospect taken in the 1980s on the exclusionary forms of reason and universality composed by a Western modernity complicit with imperial expansion and colonialist rule' (Parry 1997: 4). This major retrospective glance was itself filiated to the broader criticisms of Western philosophy formulated by Jacques Derrida and the subject of wide discussion in Western universities since the early 1970s. In this regard, the work of Michel Foucault, as it described the discursive constitution of objects of knowledge in their insertion into hierarchies of power also proved influential and decisive. Behind all this, as has been noted by Sherry Ortner (1984), was the emergence of a linguistic paradigm in the humanities and social sciences from the 1960s onwards. But most directly significant for postcolonialism was the gradual and increasingly important research done in the areas of feminist, multicultural, minority and gay and lesbian studies. The impulse they shared in common with postcolonial studies was the desire to contest the centrality and authority of distinctive systems of domination, together contributing to the deciphering of systems of representation designed to validate institutional subordination and silence the voice of competitors. Quite often, work of a thorough, interdisciplinary nature was pursued, such as that by Gayatri Spivak (1996) in her linking of Marxism, feminism and eclectic cultural criticism.

One of the key debates in postcolonialism is to what extent the term can be used to designate studies undertaken in a variety of different disciplines. Thus, postcolonialism is seen by some of its practitioners as the paradigmatic anti-hegemonic theoretical orientation, with colonialism taken to be the archetypal and most brutal expression of hegemony; any theoretical tendency that sees itself as anti-hegemonic is then easily taken to be affiliated to postcolonialism. For others, however, it is its productive circularity that calls the postcolonial object of study into being, while allowing unreflexive notions of nationhood, race and identity, and even of colonialism itself, to be deployed without attention to the specificities of each discipline.

Another debate relates to the unresolved temporal transcendence signalled in the first part of the word, when in fact there is evidence everywhere that colonialism has not been fully transcended. For Anne McClintock (1992) the central flaw in the term lies in its implicit links with Enlightenment notions of progress, since the 'post' marks nothing but the march of history without really attending to the difficulties involved in such an implicit filiation to Enlightenment notions of progress. Other terms such as 'neo-colonialism' are thought to succeed better in capturing the sense of the continuing but disguised influence of the imperial metropolis, especially as the latter is refracted through different forms of globalism in the control of the financial and primary-produce markets and in the dominant modes of information gathering and dissemination more generally. In spite of these criticisms, however, the term continues to be used widely, partly because there seems to be no clear rival to mark the distinctive field, and partly also because, as Kwame Anthony Appiah (1994: 240) puts it, the term is seen as an important 'space clearing gesture'. What is more important, perhaps, is to note that postcolonialism is symptomatic of the perception of continuing injustice in the global political economy, the victims of which are seen to be formerly colonized people as well as their diasporic descendants.

3 Colonial discourse analysis and the question of historiography

One of the areas that Said's work greatly inspired was the re-visiting of the era of colonialism in order to explore the links between knowledge and power. An important extension of his work, however, took the direction of

showing that the colonial encounter affected both colonizer and colonized simultaneously, and that the formation of the colonizer's culture was heavily dependent upon the colonial encounter. The notion of an encounter between monolithic and radically opposed entities in history gave way to studies emphasizing the unstable nature of identity formation that affected both parties.

A major inflection of this kind of work has been the emphasis on a psychoanalysis of the colonial encounter and its effects. This view is most strongly associated with Homi Bhabha, whose *The Location of Culture* (1994) brings together a decade of essays devoted to this subject. This tendency was not really new in studies of colonialism. Notable studies psychoanalysing imperialism and colonialism had already been done for India by Ashis Nandy (1983) and for Africa by Franz Fanon (1952, 1961). The work of these writers was thought to be highly stimulating and to allow for very significant extrapolations to other geographical areas of the colonial encounter. Bhabha's main difference was in maintaining that resistance could be detected in the interstices of the unstable symbolic and discursive encounters between colonizer and colonized, and that the symbolic practices also revealed the unstable nature of colonial identity.

The nature of Bhabha's formulations is thought to have important implications for the study of colonialism. Of particular significance is the privileging of discourse in accounts of the colonial encounter and of resistance against it. Not only is this seen by mainstream historians as highly problematic in its avoidance of an engagement with the real and brutal processes of colonialism, it is also thought to privilege the reading practices of the critic themselves without paying adequate attention to the specific contours of relevant historical documents. Colonial discourse analysis may be said to have additional implications for the methodology of historical research and its writing. Notably, it regularly privileges metaphors and literary tropes and focuses on them as being of interpretive value in coming to terms with history. Furthermore, such metaphors and tropes are also seen to be determinate of historical process and are taken to be succinct expressions of the main qualities of colonial ideology. This intense focus on literary figures then allows for a different historiographic practice from that practised by traditional historians; there is less attention to the broad range of archival materials such as journals, diaries, official reports and the like; less atention to the comparative analysis of such documents in the attempt to arrive at a hierarchy of their value as evidence; and much more attention to the enunciatory modalities of different literary tropes and their manifestation in different historical documents. The apparently a-historical nature of such forms of analyses has frequently been pointed out by historians; but it is also evident that there is a readiness on the part of some to attempt a fusion of the two forms of historical enquiry, in the hope that the strengths of both can be drawn upon for a fuller understanding of colonialism.

A less contentious new historiographic practice is that applied from the early 1980s by members of the Subaltern Studies group in India. Taking as their main point of disagreement the dominant Orientalist, nationalist and Marxist historiographies that had dominated historical accounts of India, they sought to challenge these and, in their place, to produce a 'history from below', that would attend to the processes of the formation of a critical consciousness within the Indian peasantry during the height of colonialism and its aftermath. The Subaltern Studies project drew inspiration from the work of Antonio Gramsci, especially from his concept of hegemony and the place of the subaltern in the formation of national consciousness. Despite producing a wide-ranging number of studies, the Subaltern Studies project could not be said to have formed a 'school' as such. None the less, it represents the most coherent attempt in postcolonial studies to seize the practice of writing history for the purpose of producing knowledge about people who had hitherto been absent from such history.

4 Diaspora, hybridity and identity

The question of diasporas is also an important dimension of postcolonial studies. What is generally taken to be the high point of modern imperialism took off from about 1830, and has fruitfully been related to the formation of a new world order after the abolition of slavery. Slavery itself had ensured that large numbers of Africans had been moved to the West Indies and America to work on plantations and to sustain the global industrial economy of the period. Indian indentured labour also featured prominently in this mass movement of peoples across various geographical locations.

That first phase of diaspora, marked mainly by slavery, was later succeeded by another, less coherent, phase. This second phase was especially pronounced from about the end of the decolonization period in the 1960s, when for various economic and other reasons a large number of Asians and Africans moved to Europe and America. The

different phases of migration produced different relationships both to the countries of origin and to those of eventual sojourn. What is most significant, however, is that as generations of migrants settled in the West, an active diaspora consciousness began to be formed. Literature written by migrants is now an increasingly important part of postcolonial studies. Additionally, there have been serious attempts to discuss the complexities of cross-cultural existence and the ways in which this affects the individual psyche as well as the culture of the adopted country of residence. And in the United States, because the migrants who went there joined an already existing and active black population, the discussion of diaspora had to take account of a variety of often complex issues to do with identity.

Many of these issues have been steadily formulated by African-Americans themselves from as far back as the slavery period. W.E.B. Du Bois, author of *The Souls of Black Folk* (1903), formulates the notion of the split consciousness of black peoples, who continually have to see themselves in terms of the idea of an indigenous folk as well as in relation to their perception by the dominant categories of white America. And in a lecture series published as *Playing in the Dark* (1992), Toni Morrison revisits similar issues and asks what happens to the imagination of black authors who are conscious at some level of having to represent their own race to a race of readers that has always understood itself as universal, and has continually written minority races out of the account. The theme of identity formation for minority races has many resonances with Franz Fanon's ideas about the psychology of Africans under the brutal conditions imposed by French colonialism in Algeria. For others, the issue of identity in diaspora is seen more in the light of notions of hybridity, in which it is thought that the diasporic individual marries ideas from both country of origin and that of sojourn. The issues are vastly complicated as they are traced through different generations, and there are some who oppose the idea of hybridity, noting that it depends too much for its success on a concept of freedom which is ultimately class based. If postcolonialism seems to be a variable and problematic concept, to many its continuing use is sanctioned by the fact that there is no clear alternative that is able to speak to the complexities of modern existence while simultaneously taking account of the historic configuration of colonialism in a cross-disciplinary and fertile way.

See also: Postcolonial philosophy of science; Race, theories of

ATO QUAYSON

References and further reading

Ahmad, A. (1992) *In Theory: Classes, Nations, Literatures*, London: Verso, 1-20.(Highly focused critique of postcolonialism from a Marxist perspective; immensely readable and absolutely essential.)

Ahmad, A. (1995) 'The Politics of Literary Postcoloniality', *Race and Class* 36 (3). (Important discussion of some of the ideological underpinnings of postcolonialism; extends the discussion in §2.)

Appiah, A.K. (1994) *In My Father's House: Africa in the Philosophy of Culture*, London: Methuen.(Important discussion of key concepts and problems in Pan-Africanist thinking in its unreflexive dependence on biological ideas of race.)

Ashcroft, B., Griffiths, G. and Tiffin, H. (1989) *The Empire Writes Back: Theory and Practice in Post-Colonial Literatures*, London: Routledge.(Seminal introduction to literary postcolonialism; see §1.)

Ashcroft, B., Griffiths, G. and Tiffin, H. (1995) *The Post-Colonial Studies Reader*, London: Routledge. (Comprehensive reader; divided into various topical sub sections, it contains many of the seminal essays in the field up to the date of its publication.)

Bernal, M. (1987) *Black Athena*, vol. 1, London: Free Association Press.(Illuminating but controversial analysis of the discipline of classics, arguing for the influence of Black Egypt on Ancient Greece; see §1.)

Bernal, M. (1991) *Black Athena*, vol. 2, London: Free Association Press.(Continuation of the study of the black roots of Ancient Greek Civilization.)

Bhabha, H. (1992) 'Postcolonial Criticism', in S. Greenblatt and G. Dunn (eds) *Redrawing the Boundaries: The Transformation of English and American Literary Studies*, New York: Modern Languages Association of America.(Expands on the idea of the productive eclecticism of postcolonialism in §2.)

Bhabha, H. (1994) *The Location of Culture*, London: Routledge.(Seminal essays on the psychoanalyses of colonialism as noted in §3; very difficult for the beginner.)

Brantlinger, P. (1988) *Rule of Darkness: British Rule Literature and Imperialism, 1830-1914*, Ithaca, NY: Cornell University Press.(Account of imperial ideology as perceivable in the work of certain English writers; highly engaging and accessible; see §1.)

Du Bois, W.E.B. (1903) 'The Souls of Black Folk', in H.L. Gates and N.Y. McKay (eds) *The Norton Anthology of*

African American Literature, New York: Norton, 1997.(Seminal text of the literature of negro reconstruction and renaissance; fruitful links to be drawn with questions of diaspora and identity in postcolonial studies as noted in §4.)

Fanon, F. (1952) *Black Skin, White Masks*, trans C.L. Markham, New York: Grove Press, 1967.(Seminal work in postcolonial theory dealing with the question of subjectivity under brutal colonial domination; engaging and stimulating; see §§3, 4.)

Fanon, F. (1961) *The Wretched of the Earth*, trans. C. Farrington, Harmondsworth: Penguin, 1967.(Another highly regarded work by Fanon; the essay entitled 'On National Culture' is especially interesting.)

Gates, H.L., Jr. and McKay, N.Y. (1997) *The Norton Anthology of African American Literature*, New York: Norton.(Possibly the most important anthology of African American literature ever to be published.)

Guha, R. and Spivak, G.C. (1988) *Selected Subaltern Studies*, Oxford: Oxford University Press.(This special volume, containing a selection of papers by the Subaltern Studies group and with a foreword by Edward Said, gives a good indication of the range of the group's work.)

Kennedy, D. (1996) 'Imperial History and Post-Colonial Theory', *The Journal of Imperial and Commonwealth History* 24 (3): 345-63.(Represents a growing tendency by imperial historians to engage positively with aspects of colonial discourse analysis.)

Lefkowitz, M. (1996) *Not Out of Africa*, New York: Basic Books.(A point by point refutation of the central theses raised in Bernal's *Black Athena*.)

MacKenzie, J.M. (1995) *Orientalism: History, Theory and the Arts*, Manchester: Manchester University Press. (Delivers a sharp but measured critique of colonial discourse analysis and attempts to problematize the notion of Orientalism with a discussion of the fine arts and architecture in the imperial period.)

McClintock, A. (1992) 'Pitfalls of the Term Post-Colonialism', in P. Williams and L. Chrisman (eds) *Colonial Discourse and Post-Colonial Theory*, New York: Harvester Wheatsheaf, 1993, 291-304.(Good discussion of the temporal implications of the postcolonialism; see §2.)

Morrison, T. (1992) *Playing in the Dark: Whiteness and the Literary Imagination*, London: Picador, 1993. (Exploration of the degree to which the predominantly white American literary canon has always depended upon a subliminal social engagement with blackness even when this is silenced; useful questions for identity formation in the diaspora as noted in §4.)

Mudimbe, V.Y. (1988) *The Invention of Africa: Gnosis, Philosophy and the Order of Knowedge*, Bloomington, IN: Indiana University Press and London: James Currey.(Important contribution to the methodology of African studies, which argues that African systems of knowledge have so far been produced within an essentially Western epistemological framework; a bit difficult for the beginner; see §1.)

Nandy, A. (1983) *The Intimate Enemy: Loss and Recovery of Self under Colonialism*, Delhi and Oxford: Oxford University Press.(Study of the effects of colonialism on modern Indian psyche and the ways to transcend these problems; see §3.)

O'Hanlon, R. (1988) 'Recovering the Subject: *Subaltern Studies* and Histories of Resistance in Colonial South Asia', *Modern Asian Studies* 22 (1): 189-224.(Very good overview of the work of the Subaltern Studies group; see§3.)

Ortner, S.B. (1984) 'Theory in Anthropology since the Sixties', *Comparative Studies in Society and History* 26 (1): 126-66.(Discussion of the major alignments in theory in the social sciences and the humanities up to the early 1980s; see §2.)

Parry, B. (1997) 'The Postcolonial: Conceptual Category or Chimera?', *The Yearbook of English Studies* 27: 3-21. (Good discussion of the composite provenance of postcolonialism raised in §2; this special number of *YES* is devoted to 'The Politics of Postcolonial Criticism'.)

Prakash, G. (1990) 'Writing Post-Orientalist Histories of the Third World: Perspectives from Indian Historiography', *Comparative Studies in Society and History* 32 (1): 383-408.(Extended discussion of Indian historiography and the place of the Subaltern Studies project outlined in §3.)

Said, E.W. (1978) *Orientalism*, London: Chatto & Windus.(Seminal work taken to have inaugurated postcolonial studies; immensely readable; see §1.)

Said, E.W. (1993) *Culture and Imperialism*, London: Chatto & Windus.(Continues what was started in *Orientalism* but this time focuses on the literary sphere; also immensely readable with an excellent bibliography.)

Sekyi-Otu, A. (1996) *Fanon's Dialectic of Experience*, Cambridge, MA: Harvard University Press.(An exciting and engaged analysis of Fanon's work, which attends both to his intellectual debts and to the creative potential inherent in his work; excellent as an introduction to Fanon's debts to classical Marxist theory.)

Slemon, S. (1994) 'The Scramble for Post-Colonialism', in B. Ashcroft, G. Griffiths and H. Tiffin (eds) *The*

Post-Colonial Studies Reader, London: Routledge, 1995, 45-52.(Excellent overview of the applications of postcolonial theory; extends the account of the genealogy of the term raised in §3.)

Spivak, G.C. (1996) *The Spivak Reader*, ed. D. Landry and G. MacLean, London: Routledge.(Contains all Spivak's seminal essays; her work is very useful for exploring the intersection between feminism, Marxism and deconstruction; particularly important is 'Can the Subaltern Speak?'; a bit difficult for the beginner; see §2.)

Thomas, N. (1994) *Colonialism's Culture: Anthropology, Travel and Government*, Cambridge: Polity Press.(Good discussion of the varying and differential practices of colonialism that challenges some of the central tenets in the work of Homi Bhabha and others raised in §2.)

Viswanathan, G. (1990) *Masks of Conquest: Literary Studies and British Rule in India*, London: Faber & Faber. (Detailed discussion of the introduction of English literary studies into India and the ideological implications of this; see §1.)

Visweswaran, K. (1994) *Fictions of Feminist Ethnography*, Minneapolis, MN: Minnesota University Press. (Chapter 7 of this book offers a highly engaging discussion of notions of hybridity, contesting the term with a new one - hyphenation.)

Williams, P. and Chrisman, L. (1993) *Colonial Discourse and Post-Colonial Theory*, New York: Harvester Wheatsheaf.(Along with Ashcroft *et al.* (1995), a key teaching tool. Has an excellent general introduction pointing towards interesting areas of research in the field.)

Young, R. (1990) *White Mythologies: Writing History and the West*, London: Routledge.(Last three chapters offer stimulating discussion of the work of Edward Said, Homi Bhabha and Gayatri Spivak; significant for dealing with the question of the ideological implications of history writing in the West.)

Postmodern theology

The term 'postmodernism' is loosely used to designate a wide variety of cultural phenomena from architecture through literature and literary theory to philosophy. The immediate background of philosophical postmodernism is the French structuralism of Saussure, Lévi-Strauss, Lacan and Barthes. But like existentialism, it has roots that go back to the critique by Kierkegaard and Nietzsche of certain strong knowledge claims in the work of Plato, Descartes and Hegel. If the quest for absolute knowledge is the quest for meanings that are completely clear and for truths that are completely certain, and philosophy takes this quest as its essential goal, then postmodernism replaces Nietzsche's announcement of the death of God with an announcement of the end of philosophy.

This need not be construed as the death of God in a different vocabulary. The question of postmodern theology is the question of the nature of a discourse about deity that would not be tied to the metaphysical assumptions postmodern philosophy finds untenable. One candidate is the negative theology tradition of Pseudo-Dionysius and Meister Eckhart. It combines a vigorous denial of absolute knowledge with a theological import that goes beyond the critical negations of postmodern philosophy. A second possibility, the a/theology of Mark C. Taylor, seeks to find religious meaning beyond the simple opposition of theism and atheism, but without taking the mystical turn. Finally, Jean-Luc Marion seeks to free theological discourse from the horizon of all philosophical theories of being, including Heidegger's own postmodern analysis of being.

1 Key themes in postmodern philosophy

The themes of philosophical postmodernism whose bearing on philosophical theology is most direct are Heidegger's 'destruction' of the history of ontology and Derrida's 'deconstruction' of the metaphysics of presence. Heidegger's destruction, originally announced in *Being and Time* (1927), becomes a critique of 'the onto-theo-logical constitution of metaphysics' in *Identity and Difference* (1957). His point of reference is Aristotle's attempt to unify being, which is said in many ways. At the categorial level, it is substance that plays the role of unifying first principle for all the *pros hen* equivocals. But how is the world of actual substances itself to be unified? By a highest substance, God, who is universal by being first. In other words, the whole realm of beings can only be understood fully with reference to the highest being by whom their being is ordered (see Heidegger, M. §6).

Heidegger calls this confluence of the universality of ontology with the primacy of theology 'onto-theo-logy'. In a typically postmodern gesture, he treats this honoured principle, whose modern fulfilment he finds in Hegel, with undisguised disrespect. He remarks that in onto-theo-logy 'the deity can come into philosophy only insofar as philosophy, of its own accord and by its own nature, requires and determines that and how the deity enters into it' ([1957] 1969: 56). In other words, a tradition that stretches at least from Aristotle to Hegel uses God as a means to its own ends; anticipating the spirit of modern technology, it views even God as a resource in the service of its own will to power, the totalizing project of rendering the whole of being intelligible to human understanding. In transgressing this tradition, Heidegger invokes both a hermeneutics of finitude, which views such a project as hubris, and a hermeneutics of suspicion, which looks for hidden and disreputable motives beneath discourses generously decorated with piety.

Heidegger insists that his critique, so far from having its origin in atheism, can be seen as a form of religious protest. Echoing Pascal, he says of the god of philosophy as onto-theo-logy:

> Man can neither pray nor sacrifice to this god. Before the *causa sui* man can neither fall to his knees in awe nor can he play music and dance before this god.
>
> The god-less thinking which must abandon the god of philosophy, god as *causa sui*, is thus perhaps closer to the divine God. Here this means only: god-less thinking is more open to Him than onto-theo-logic would like to admit.
>
> (Heidegger [1957] 1969: 72)

Heidegger develops this critique explicitly against the Hegelian holism that would attain absolute knowledge by developing an all-inclusive categorial scheme, the Logic, and an all-inclusive theory of actual beings, the Philosophies of Nature and of Spirit. Derrida has his own critique of Hegel, but in this context the deconstruction

of the metaphysics of presence is best viewed as an assault on Cartesian foundationalism and its claim that, in a piecemeal manner and at the outset, we can achieve unambiguous meanings and final truths. It is just such meanings and truths that Derrida designates as presence, though the term ‘immediacy’ signifies much the same thing (see Derrida, J.).

What is present is both here and now, and Derrida’s notion of presence has both spatial and temporal ramifications. To be fully present to the meanings and truths necessary for even piecemeal absolute knowledge is to have found the transcendental signified, a meaning or a truth so self-contained as to require no reference to anything outside itself in semantic space and so finished as to require no reference to any clarification or validation subsequent to this moment in time. Deconstruction is not so much the assertion that we never reach the transcendental signified as it is the continuous showing of the spatial differences and temporal deferrals that undermine claims of total clarity and final certainty.

Derrida (1968) invents the term *différance* to stand for this union of difference and deferral in the critique of logocentrism, which can be defined as the claim that absolute knowledge is possible, that either prior to all linguistic mediation (as in the work of Husserl) or subsequent to the essential completion of that mediation (as in the work of Hegel), human thought stands face to face with being. The spatial part of the argument derives largely from structuralism, while the temporal part is derived from Husserl’s phenomenology of internal time consciousness for use, in typically *ad hominem* fashion, against Husserlian foundationalist claims of apodictic intuition.

For Heidegger and Derrida, as for Kierkegaard and Nietzsche, it is the temporal structure of human existence above all else that renders the transcendental signified a transcendental illusion for human thought. In this respect they are all Kantians, for whom it is foolish to claim the finality of eternity for human thought.

2 Postmodern theology as negative theology

The question of a postmodern theology that emerges is quite clear: what would a discourse look like that would still be theological but would vigilantly resist lapsing into metaphysics, which in this context signifies the confluence of onto-theo-logy, the metaphysics of presence, and logocentrism? Since Heidegger and Derrida are in agreement that we cannot just decide to leave metaphysics behind and be done with it, such a postmetaphysical theology will not be the triumphant freedom from all metaphysical tendencies, but the militant struggle against enslavement to them. Partly because of the mystical element in Heidegger’s thought, and partly because the negative character of Derridean deconstruction has from the outset suggested affinities with negative theology, the tradition of mystical theology that has Pseudo-Dionysius and Meister Eckhart among its leading figures is easily viewed as a paradigm of postmetaphysical theology.

Derrida’s answer to the question whether deconstruction is a kind of negative theology is an emphatic negative. Negative theology for Dionysius and Eckhart is part of the *via negativa*; it is a scepticism in the service of mysticism. In so far as deconstruction can be construed as a kind of scepticism, it serves no mystical project. It does not posit a ‘superessential’ deity, a God beyond being; it knows no nostalgia or hope for a pre- or postconceptual experience of pure presence to such a deity; and it does not address such a deity in the second person language (‘you’ or ‘thou’) of prayer or praise. Moreover, the mystical project, like its cousin, the logocentric project of Hegel, is committed to an overcoming of difference that betrays an allergic reaction to the very experience of otherness that deconstruction seeks to preserve.

Kevin Hart (1989) believes Derrida is right to reject the suggestion that deconstruction is a kind of negative theology. But he argues that negative theology is a deconstruction of positive theology, and in this way the paradigm of postmetaphysical theology. He agrees with Heidegger and Derrida that what the former calls the ‘step back’ out of metaphysics is an ongoing task. This means that negative theology will always be accompanied by positive theology, as it is in Dionysius and Eckhart.

But they have a distinctive account of the relationship between the two. For someone like Thomas Aquinas, positive theology is prior and negative theology is a supplemental check to see that our discourse is about God and not just our ideas of God. For Dionysius and Eckhart, negative theology is prior to all the statements of positive theology. At issue here is not the order of exposition, as if the question were what should go into Chapter One. The issue is whether the negative principle, the denial that our concepts could possibly be adequate to the divine

reality, is the first principle of theology.

3 Postmodern theology in a Nietzschean mode

The ‘postmodern a/theology’ of Mark C. Taylor is not a negative theology of the sort just described. Like Derridean deconstruction, a major source of its inspiration, it lacks the nostalgia and hope that mysticism shares with the positive theologies of theism. It has a strong Nietzschean bent and describes deconstruction as the hermeneutics of the death of God. But it describes itself as between belief and unbelief, because its response to Nietzsche’s announcement is neither the joyous, atheistic acceptance of God’s demise nor the angry or frightened, theistic rejection of it. This is not because it is indifferent to religious meaning, but because it finds both the simple affirmation and the simple denial of God’s reality to be too deeply enmeshed in the metaphysical thinking from which postmodernism seeks to extricate itself.

Taylor begins his a/theology (1984) with a deconstructive critique of four concepts that have been central to Judaic and Christian religion: God, self, history and book. In relation to tradition, this is an anti-theological gesture, but it purports to be of religious significance by opening up new possibilities for religious imagination. The positive rethinking of the four deconstructed themes is not described as reconstruction but as nomad thought in order to signify that it neither has nor seeks any fixed points of reference. Life is wandering and erring, and thus a maze; but this a/theology seeks for signs of ‘mazing grace’, as Taylor puts it.

The path from Nietzsche to Derrida leads all but inevitably through Heidegger. So it is not surprising to find an important Heideggerian theme in Taylor’s project, namely the attempt to find a mode of thought that is not enslaved to projects of mastery, control and domination. The task of a/theology is neither to get a handle on God, nor to find in God the key to our mastery of the world.

4 Postmodern theology in a Kierkegaardian mode

The postmetaphysical theology of Jean-Luc Marion owes no special debt to Kierkegaard; but it is written in a Kierkegaardian mode in the sense that it combines an uncompromising critique of the mainstream metaphysical traditions of the West with the desire to return to biblical religion. It points towards a dehellenized Christianity for those who still find themselves, or would like to find themselves, in the Church.

Marion (1982) stresses that the critique of metaphysics as onto-theo-logy is the critique of a certain kind of discourse. It does not entail either God’s unreality or the absence of a divine character that is God’s own, independent of our images and concepts of God. Its purpose is not to make the world safe for secularism, but to open the space for a new experience of God as love and as gift.

However, this space is to be found beyond the horizon of Being, and the Heideggerian notion that a critique of theology can represent a new openness to God is immediately supplemented with a critique of Heidegger. Marion affirms his identification of onto-theo-logy as the arrogant demand by philosophy that God enter its discourse on its terms and in its service. At least as strongly as Heidegger, he wants to free God from these constraints. But he insists that the ‘step back’ out of metaphysics is not a step back into Being. In other words, Heidegger’s attempt to think Being is more nearly a continuation of the onto-theo-logical constitution of metaphysics than a decisive break with it. The move away from the Athens of Plato and Aristotle needs to be in the direction of the biblical Jerusalem rather than the pre-Socratic Magna Graecia.

The major premise of Marion’s argument is the distinction between an idol and an icon. It concerns the ‘how’ rather than the ‘what’ of perception, which means that one and the same object could be an idol for one observer and an icon for another. An object of religious significance is an idol when it satisfies our perception, fulfils our intentions, freezes our gaze and brings it to rest. We have arrived and need not go further. An object is an icon when our gaze finds it necessary to ‘transpierce’ it in search of what exceeds it, when the invisible remains invisible in its visibility, when what is presented opens an abyss that we can never finish probing. While the idol limits the divine to the measure of the human gaze, the icon signifies an openness to that which exceeds every human measure. This analysis can be transferred from the sensible to the intelligible realm. Concepts, too, can serve as idols, not by virtue of their content (their ‘what’) but by virtue of their use (their ‘how’). Any theology that professes an adequation between its concepts and the divine reality is *ipso facto* idolatrous; this is the fatal flaw of onto-theo-logy.

Heidegger's thinking of Being breaks decisively with the Hegelian-Husserlian drive towards adequate concepts. But by making the human understanding of Being the horizon for any understanding of God, Heidegger also compels the divine to conform to the measure of human thought. After *Being and Time*, he understood the goal of philosophy to be to let things show themselves in and from themselves. Marion argues that the horizon of Being undermines this goal. The self-revelation of God as love and as gift shatters this and every horizon, for love and gift do not signify concepts that would be adequate to the divine reality and with which we might rest, but the excess of the divine reality to every attempt on our part to think it. For Marion, postmodern theology is the attempt to preserve this self-revelation from every philosophical theory of Being, modern or postmodern.

See also: Feminist theology; Negative theology; Postmodernism; Religion and epistemology

MEROLD WESTPHAL

References and further reading

Caputo, J.D. (1986) *The Mystical Element in Heidegger's Thought*, New York: Fordham University Press, 2nd revised edn.(Lucid study of the mystical side of the later Heidegger.)

Derrida, J. (1967) *La voix et le phénomène*, Paris: Presses Universitaires de France; trans. D. Allison, *Speech and Phenomena and Other Essays on Husserl's Theory of Signs*, Evanston, IL: Northwestern University Press, 1973.(Introduction to deconstruction as theory and practice through a detailed, deconstructive reading of Husserl.)

Derrida, J. (1968) 'Différance', in *Théorie d'ensemble*, coll. Tel Quel, Paris: Éditions du Seuil; trans. A. Bass, *Margins of Philosophy*, Chicago, IL: Chicago University Press, 1982.(Brief introduction to deconstruction, with a discussion of negative theology.)

Derrida, J. (1987) 'Comment ne pas parler: Dénégations', in *Psyché: Inventions de l'autre*, Paris: Galilée; trans. H. Coward and T. Foshay, 'How to Avoid Speaking: Denials', in H. Coward and T. Foshay (eds) *Derrida and Negative Theology*, Albany, NY: State University of New York Press, 1992.(Key essay on negative theology.)

Hart, K. (1989) *The Trespass of the Sign: Deconstruction, Theology and Philosophy*, Cambridge: Cambridge University Press.(Seminal study of postmodern theology as negative theology, with special reference to Heidegger and Derrida.)

Heidegger, M. (1927) *Sein und Zeit*, *Jahrbuch für Philosophie und phänomenologische Forschung* 8: 1-438; trans. J. Macquarrie and E. Robinson, *Being and Time*, New York: Harper & Row, 1962.(Major initial attempt of Heidegger to work free from the history of ontology; see especially §§6 and 8 for the 'destruction' of the history of ontology.)

Heidegger, M. (1957) *Identität und Differenz*, Pfullingen: Neske; trans. J. Stambaugh, *Identity and Difference*, New York: Harper & Row, 1969.(Important late Heideggerian account of the 'step back' from onto-theo-logical metaphysics.)

Lowe, W. (1993) *Theology and Difference: The Wound of Reason*, Bloomington, IN: Indiana University Press. (Derridean reflections on theology in relation to Karl Barth, among others.)

Marion, J.-L. (1982) *Dieu sans l'être: Hors-texte*, Paris: Librairie Arthème Fayard; trans. T.A. Carlson, *God Without Being*, Chicago, IL: University of Chicago Press.(A classic postmodern theology with a sustained critique of Heidegger's ontology.)

Taylor, M.C. (1984) *EЯRING: A Postmodern a/theology*, Chicago, IL: University of Chicago Press.(A search for religious meaning through the deconstruction of traditional concepts.)

Postmodernism

The term 'postmodernism' appears in a range of contexts, from academic essays to clothing advertisements in the New York Times. Its meaning differs with context to such an extent that it seems to function like Lévi-Strauss' 'floating signifier'(Derrida 1982: 290): not so much to express a value as to hold open a space for that which exceeds expression. This broad capacity of the term 'postmodernism' testifies to the scope of the cultural changes it attempts to compass.

Across a wide range of cultural activity there has been a sustained and multivalent challenge to various founding assumptions of Western European culture since at least the fifteenth century and in some cases since the fifth century BC: assumptions about structure and identity, about transcendence and particularity, about the nature of time and space. From physics to philosophy, from politics to art, the description of the world has changed in ways that upset some basic beliefs of modernity. For example, phenomenology seeks to collapse the dualistic distinction between subject and object; relativity physics shifts descriptive emphasis from reality to measurement; the arts move away from realism; and consensus politics confronts totalitarianism and genocide. These and related cultural events belong to seismic changes in the way we register the world and communicate with each other.

To grasp what is at stake in postmodernism it is necessary to think historically and broadly, in the kind of complex terms that inevitably involve multidisciplinary effort. This multilingual impetus, this bringing together of methods and ideas long segregated both in academic disciplines and in practical life, particularly characterizes postmodernism and largely accounts for such resistance as it generates.

Although diverse and eclectic, postmodernism can be recognized by two key assumptions. First, the assumption that there is no common denominator - in 'nature' or 'truth' or 'God' or 'the future' - that guarantees either the One-ness of the world or the possibility of neutral or objective thought. Second, the assumption that all human systems operate like language, being self-reflexive rather than referential systems - systems of differential function which are powerful but finite, and which construct and maintain meaning and value.

1 Historical context

'Postmodernism' is a historical term, indicating something that comes after modernity; so the definition of postmodernism varies depending on what is meant by 'modern'. When 'modern' refers to movements in the arts around the turn of the twentieth century - an efflorescence known as 'modernism' - then 'postmodern' refers to a fairly local phenomenon of the mid- to late-twentieth century (see Modernism). But when 'modern' is used in the historian's sense, to indicate what follows the medieval - that is, Renaissance culture and its sequels - then 'postmodern' refers to a more broadly distributed cultural phenomenon in European and US societies.

Given this latter use of the term 'modern', *post*modernism is what follows and transforms that particular Renaissance (some would prefer to call it Enlightenment) modernity. It is in this broader sense of modernity that the term 'postmodernism' takes on its full meaning. Here it signals a revisionary shift in the system of values and practices that have been broadly codified in European life over several centuries. Confusion between these two historical meanings of the term 'modern' skews discussion of the crucial philosophical, political and social issues at stake in postmodernism.

Such confusion certainly is not relieved by the ahistorical contributions of recent French philosophy which has taken the demystification of Western metaphysics back to Plato. Known as post-structuralism, this philosophy has been crucial to theoretical discussion of postmodernism, but not identical with it; the historical horizon of post-structuralism occupies a period so vast that it practically ceases to allow for history at all, producing discussions that blot out awareness of postmodernity as a historical event. Luce Irigaray and Jacques Derrida, for example, take on the foundations of Western philosophy and do not limit their critiques specifically to modern European culture (see Post-structuralism).

Modernity, defined from the vantage point of post-modernity, was a cultural epoch and 'episteme' (see §2) founded in a humanistic belief that the world is One. This belief, codified in centuries of realist art, representational politics and empirical science, is tantamount to the assertion that a common denominator can be found for all systems of belief and value: that the world is a unified field, explicable by a single explanatory

system. As this belief developed through increasingly secular and materialistic practices it became less secure in its claims to universal applicability. After the Renaissance the ‘totalizing’ claim to universal applicability was increasingly transferred from divinity to infinity: especially the infinity of space and time as they were radically reconstructed by Renaissance art and science. Postmodernism is the condition of coping without these absolute common denominators, especially without the neutral and homogeneous media of time and space which are the quintessential, field-unifying media of modernity.

Postmodernism specifically challenges the European culture that took its direction from the Renaissance, developed through the seventeenth century and the Enlightenment, and remains a common discourse for most citizens of Western democratic societies. In philosophy, in the arts, in science, in political theory and in sociology, postmodernism challenges the entire culture of realism, representation, humanism and empiricism. Postmodern critique thus goes to the very foundation of personal, social and institutional definition. Its challenges to knowledge and institutions are felt particularly in universities.

2 The role of language

Already well under way in the nineteenth century, the critique of Enlightenment rationalism found its postmodern turning point in the spreading influence of Ferdinand de Saussure, a Swiss teaching at Geneva whose lectures on linguistics (published in 1916 as *Cours de Linguistique Générale*) have become a keystone of postmodernism. Saussure’s deceptively simple idea was that the linguistic sign acts reflexively, not referentially. The word, he said, functions not by pointing at the world, but by specifying an entire system of meaning and value in which each word has its function. (Technically, Saussure’s word points to an idea, but that idea is itself linguistic; there is nothing prior to language.) To read or understand even the simplest linguistic sequence is to recognize difference; it is to perform an incalculably complex and continuous act of differentiation, which becomes more and more balanced and rich the more that linguistic sequence verges towards poetry or other complex usage.

In one way this is only common sense. Any speaker of more than one language knows the arbitrariness of the sign: what is ‘dog’ here, is ‘chien’ there. But the more languages one knows, the more obvious becomes the systemic value of any word, the more obvious the fact that it has no exact equivalent elsewhere, either in other languages or in the world. To take the simplest possible example, we understand the English word ‘dog’ not merely because conventionally we have associated it with a creature, but more complexly and largely because we differentiate it by composition and function from other words and functions (for example, from ‘dot’, ‘log’, ‘bog’, and from verbs, adverbs, conjunctions). What is being described in these rather dry terms is the language’s capacity for poetry: its capacity as a living language to provide its speakers with particular alphabets and lexicons of possibility, and to modify, even radically, the usages with which we constitute our worlds.

The postmodern moment, as Derrida says, is the moment ‘when language invaded the universal problematic’; this is the moment when it becomes clear that everything operates by such codes, that everything behaves like language. Body language, garment language, the silent expression of gesture, the layout of a city or a fashion magazine or a university: all these are complex, coded systems of meaning and value in which we function simultaneously in several, even many, at once. Even in humdrum activities we are expert well beyond our conscious measures. Language thus conceived is a model of organization that is both powerful and finite.

For describing systems of value that work like language, the terms *episteme* and *discourse* have emerged as useful. *Episteme* suggests the systemic nature of all knowledge (one can speak of the Western episteme); and *discourse* suggests the systemic nature of all practices (moral, social, domestic, political, reproductive, economic, intellectual). These two terms at least help the mind to find the fulcrum that allows thought to run in directions different from those inspired by terms like ‘reality’ or ‘nature’ (see §3).

Postmodernism differs from deconstruction, with which it is sometimes confused. Deconstruction is a methodology with agendas similar to some of postmodernism but with a much more limited capability. Deconstruction is a negative movement by which an interpreter of a code or sign-system (for example a novel or a psychoanalysis) looks for what is *not* present rather than what is present - looks for the points of crisis and breakdown in a system or a rationalization rather than its more obvious positivities. This methodology has the initial value of opening interpretation to complex reading, but it soon gets lost in circularity; the negative quality of its questioning often limits the creativity of the response. It is almost as though deconstruction is riveted on what it

has not got, operating on a kind of nostalgia for the referential view of language that postmodernism revises (see Deconstruction).

Postmodernism does not weep for referentiality. If the sign does not refer simply, but instead specifies a system of meaning and value, then interest lies in discovering what systems actually are in play and in seeing what different systems are capable of, whether they be literary texts, political movements or personal lives. This valence of postmodernism can be found not so much in the theoretical texts which have had such extensive recent attention, but in the creative work of artists and scientists who have in many cases anticipated the philosophical critiques of rationalism, and have gone well beyond them to locate their practical implications. Artists and film-makers like Magritte and Buñuel, postmodern novelists like Robbe-Grillet and Nabokov, post-Einsteinian scientists interested in quanta and chaos, feminists interested in new acts of personal and political attention, and architects who play with traditional conventions have explored the practical and material implications of postmodernism far more fully than have many of the more theoretical writers.

3 Challenge to the bases of consensus and representation

The view that all systems are self-contained and largely self-referential has a radical implication that either alarms or inspires - the implication that no system has any special purchase on Truth and, in fact, that it is impossible to establish a Truth. This implication goes beyond the recuperable relativism of the nineteenth century to unrecuperable difference in the twentieth. Where relative systems could still cohabit in the single world of modernity, postmodernity involves the recognition that, to a large extent, one's relative systems *construct* the world. In short, that the world is *not* One; that words like 'truth,' 'nature', 'reality' and even 'human' are weasel words because they imply, falsely, that an autonomous world of meaning and values exists, and that it *transcends* all finite and mutually exclusive human systems and somehow guarantees them. Postmodernism denies absolute status to any truth or nature or reality. The question always remains - what truth, which nature, whose reality?

While this post-structuralist critique of transcendence goes back to Plato, the postmodern version limits its ambit to the particular forms of transcendence made available in the Renaissance and Enlightenment. The discourse of modernity extends to infinity its neutral media of space and time and, in so doing, encourages us to forget finitude and to distribute our energy toward an infinite horizon. The discourse of postmodernity, on the other hand, treats time and space as dimensions of finite systems. This recognition of absolute and unmediateable finitude inspires reflexiveness because activity no longer can be referred to unchanging external absolutes. Such reflexiveness always remains experimental or improvisatory, and inaccessible to universal generalization.

The disappearance of transcendental reference creates four related crises: the crisis of 'the subject' (the irreducible individual, the one that is because it thinks or is conscious); the crisis of 'the object' (the 'things' - including the individual - that constitute a world that is single, not multiplied); the crisis of 'the sign' (the word that refers to the world thus constituted); and, consequently, the crisis of historicism (the temporal humanism that constitutes an uneasy unity in the world by formulating transcendence as The Future).

Postmodernism presents a new problematic of negotiation between finite systems of meaning and value where no transcendental reference is possible. This negotiation goes on all the time in complex ways but, in the discourse of modernity, neither philosophy, nor political science, nor indeed any science, nor even much art has attended to it. This new problematic is, in Craig Owens' words, 'how to conceive difference without opposition', and how to translate that problematic, as the Renaissance translated Christian humanism, into social and political terms.

Where modernity sought the single system, postmodernity plays with the elements of systems, combining them for limited agendas, using what is useful and leaving the rest, refusing responsibility for consistency within this or that totalized explanatory system. This 'bricolage' becomes a key value for postmodernism. What one wants to avoid at all costs is something without play, without slack, without the living capacity for movement; one wants play in the line, play in the structure, in the sense of flexibility and variability even to the point of reorganizing the structure.

By conceiving all particulars and practices as functions of systems rather than as semi-autonomous entities, postmodernism poses especially interesting problems for two agendas of modernity that are particularly valuable to twentieth-century social and political function: historicism and individualism. History as a single, universal system of human explanation depends upon a construction of temporality that belongs to humanism and the

Renaissance; its very notion of temporality is a kind of single-point perspective in time. Postmodernism puts history in the interesting position of considering its own historicity. The individual subject, sometimes appearing as the Cartesian *cogito*, does not, despite alarms, entirely disappear in postmodernism into systemic function. But individual identity and agency do have new definitions and functions when all practice is conceived differentially and systemically rather than naturally. Whatever is individual about a life, its unique and unrepeatable poetry, comes not from some 'natural' essence but from its particular specification of the complex discourse it inhabits.

Given the extent of the reformation implied by postmodernism, it is not surprising to find a flutter of reaction against it, not all of it informed and much of it governed by a desire to rescue particular epistemic investments from devaluation. The emphasis on linguistic reflexiveness, on the power of a system of signs to constitute meaning and value, has been taken by some commentators as an expression of inward-looking narcissism, a flight from 'reality' and, even, a threat to morality and order: it is as if the social, political and epistemic problems to which postmodernism responds were created by it, rather than by the entire culture of modernity. Postmodernism does not spell the end of meaning and value, still less the end of humanist meaning and value, but it does spell the end of certain hegemonies, especially those vested in what Alain Robbe-Grillet calls 'habitual humanism'. In any case, Saussure's idea of the reflexivity, rather than the referentiality of language has by no means yet been fully explored.

The postmodern critique, amid clatter and confusion, is only just under way. There is no responsible way to anticipate its full implications or trajectory. One can see, however, that postmodernism offers both a new freedom and a new constraint. The emphasis on the constructed nature of all knowledge and projects means that, because they have been invented, they can be changed; there is, morally or socially speaking, no 'nature' of things. On the other hand, the fact that with our languages we inherit so much of our beliefs and values ready-made means that we are much less original and autonomous than modernity suggested, and we express agency more locally, more collectively and less heroically than modernity allowed.

See also: Foucault, M.; Phenomenology, epistemic issues

ELIZABETH DEEDS ERMARTH

References and further reading

Derrida, J. (1966) 'La structure, le signe et le jeu dans le discours des sciences humains' ('Structure Sign and Play in the Discourse of the Human Sciences'), in *L'écriture et La Difference* (*Writing and Difference*), trans. A. Bass, Chicago, IL: University of Chicago Press, 1982.(A seminal essay based on Lévi-Strauss and Saussure; not for the beginner.)

Ermarth, E.D. (1992) *Sequel to History: Postmodernism and the Crisis of Representational Time*, Princeton, NJ: Princeton University Press.(Accessible interdisciplinary discussion of the main theoretical and historical issues of postmodernism; bibliography.)

Foucault, M. (1969) *L'Archaeologie du Savoir* (*The Archaeology of Knowledge*), trans. A.M. Sheridan Smith, New York: Pantheon; London: Tavistock, 1972.(Fairly accessible example of postmodern methodology from a key practitioner; English edition also includes the widely used 'Discourse on Language', Foucault's 1970 Inaugural Address at the College de France, originally published as 'L'ordre du discours', 1971.)

Harvey, D. (1992) *The Condition of Postmodernity: An Enquiry*, Oxford, Blackwell.(Sensible and wide-ranging cultural analysis.)

Hutcheon, L. (1989) *The Politics of Postmodernism*, London and New York: Routledge.(Thorough and intelligent introductory survey of discussion; bibliography.)

Lyotard, J.-F. (1979) *La condition postmoderne: rapport sur le savoir* (*The Postmodern Condition: A Report on Knowledge*), trans. G. Bennington and B. Massumi, Minneapolis, MN: University of Minnesota Press, 1984. (Standard, fairly accessible discussion of key issues.)

Lyotard, J.-F. (1983) *Le Différend* (*The Differend: Phrases in Dispute*), trans. G. van den Abbeele, Minneapolis, MN: University of Minnesota Press, 1988.(Standard, fairly accessible discussion of key issues.)

Owens, C. (1983) 'The Discourse of Others: Feminists and Postmodernism', in H. Foster (ed.) *The Anti-Aesthetic: Essays on Postmodern Culture*, Port Townsend, WA: Bay Press, 57-82.(A good essay in a good collection on postmodernism and culture.)

Robbe-Grillet, A. (1958) 'La nature, l'humanisme, et la tragédie' ('Nature, Humanism and Tragedy'), in *For a New Novel: Essays on Fiction*, trans. R. Howard, Evanston, IL: Northwestern University Press, 1989.(Key essay in a useful collection; accessible, if hit-and-run, style)

Saussure, F. de (1916) *Cours de Linguistique Générale* (*Course in General Linguistics*), trans. W. Baskin, New York: McGraw-Hill, 1959.(Seminal lectures published from students' notes, *c.*1906-11.)

Tyler, S. (1987) *The Unspeakable: Discourse, Dialogue, and Rhetoric in the Postmodern World*, Madison, WN: University of Wisconsin Press.(Accessible interdisciplinary discussion of postmodernism and language.)

Postmodernism and political philosophy

Just as there is much disagreement over both what is meant by 'postmodernism' and which thinkers fall under this rubric, so also is there disagreement over its implications for political philosophy. The claim of postmodernists that raises the most significant issues is that Western modernity's fundamental moral and political concepts function in such a way as to marginalize, denigrate and discipline 'others'; that is, categories of people who in some way are found not to measure up to prevailing criteria of rationality, normality and responsibility, and so on. The West's generally self-congratulatory attitude towards liberal democracy and its traditions obscures this dynamic. Postmodernism aims to disrupt this attitude, and its proponents typically see their efforts as crucial to a radicalization of democracy.

1 'Otherness' and critique

Postmodernism aims at exposing how, in modern, liberal democracies, the construction of political identity and the operationalization of basic values take place through the deployment of conceptual binaries such as we/them, responsible/irresponsible, rational/irrational, legitimate/illegitimate, normal/abnormal, and so on. Rather than focusing upon the first term of each pair, and thus encouraging a thoroughly affirmative image of the achievements of the West, postmodernists draw attention to the ways in which the boundary between the two terms is socially reproduced and policed. This process takes place within the general sway of what is seen as modernity's deep, but questionable, commitment to expanding our mastery of self, society and nature. This persistent commitment creates pressure for hardening boundaries, simultaneously fostering conformity among those who fall on the 'correct' side of the dividing lines, and marginalization and denigration for those on the 'wrong' side. The 'others' engendered by this dynamic may be racial, ethnic, sexual, national, or they may have no easily identifiable, external characteristics. They may simply not measure up in terms of some scale of normality. (Perhaps the classic postmodern critique in this vein is Foucault's *Discipline and Punish* (1977); see Foucault, M.) Given these concerns, it is easy to see affinities between postmodernism and other contemporary political initiatives, such as multiculturalism and feminism (see Feminist political philosophy §6; Multiculturalism).

Critics have often faulted postmodernists for the totalized quality of their critique. Both liberal democracies and far more authoritarian regimes are rendered into objects equally to be condemned: each creates and disciplines others, even if the precise mechanisms differ. This equation, critics have argued, robs political philosophy of any basis for making crucial normative distinctions between types of regimes.

Proponents respond that this argument misses their point. They seek only to deepen the critique of democracy, by sensitizing its supporters to the problem of otherness. But exactly what political orientation follows from such a perspective? Although there is a fair amount of uncertainty here, it is possible to sketch several alternatives.

2 Radicalizing democracy

Perhaps the most directly engendered political orientation stemming from postmodernism would be the stance of ongoing, impertinent dissent in which the goal is to highlight continually various modes by which the cognitive machinery and institutions of existing democracy spawn otherness. This would be a type of anarchism whose substantive orientation is directly parasitic upon the qualities of the system it attacks. No alternative political vision is offered. At its best, this stance might have the salutary effect of continually bringing to light inconspicuous phenomena of injustice. At its worst, it might become a sort of automatic propensity to disrupt, unconcerned about deleterious side effects.

Although ongoing criticism is part of what an enhancement of democracy would imply for many postmodernists, it is usually seen, on its own, as being too limited. What more affirmative vision does postmodernism offer? Sometimes there is allusion to what might be called an 'other politics', radically divorced from the present. In this vein, Derrida has spoken of 'the future of another law… lying beyond the totality of this present' (1985: 298); similarly, Foucault (1980) offered the idea of 'a new form of right'. The meaning of such appeals remains problematic since they have not been elaborated in any detail. Critics suspect that such a 'politics of the ineffable' is a rather obvious strategy for those who begin with such totalized critiques (McCarthy 1991).

Another way of rethinking politics would be in terms of some reorientation of liberal democracy's spirit or

institutions so as to make them more hospitable to otherness. One form this takes is a reconsideration of the prevailing understanding of democratic sovereignty and community as residing with ‘the people’, in whom final, unquestionable authority resides (see Democracy). This is a unifying, homogenizing concept that in fact has no reality in political processes and decisions, where, at most, segments of the population express their will. Postmodernists suggest that we reimagine democratic community in a more radically open way, as something never fully present, but rather always in essential relation to those who are absent from, or defeated or shunted aside in, any particular expression of political will. The bond of such a community would be acknowledged as a tenuous one, embodying as it does a more vivid realization of difference. It could be thought of as sustained by mutual ‘agnostic respect’, as Connolly (1991) has put it. In effect, this would mean that each individual or group would accept that the formation of its identity is always simultaneously the formation of that which it is *not* - its others. The resulting tension is inevitable; we thus do best to recognize and honour it as an existential condition. Ideally, democracy is the regime most congruent with this condition.

This sort of rethinking of the binary ‘us/them’ within a democratic community also has implications for how the external relations of polities are conceived. Postmodernism’s broad questioning of the constitution and policing of fundamental political dichotomies strikes directly at the sharp internal/external distinction that is at the philosophical heart of the predominant, state-centred view of international politics. Postmodernism here would invite us to consider more critically the way the presence of a state boundary is allowed to discount radically possible obligations to others who are defined as part of what is external or foreign. All of the foregoing suggestions for rethinking democracy could be categorized as components of something like a new philosophical ethos for democracy; that is, a new sensibility towards political life. But how would this new attitude towards otherness take shape institutionally? What institutions would be more hospitable to otherness and yet constrain it where appropriate? There does not appear to be much agreement on answers to such questions.

Consider, for example, the liberal value of tolerance (see Toleration). Postmodernism would presumably urge one not just to tolerate otherness, but rather celebrate it to some degree. But what would this mean in specific cases? Would one be more tolerant than liberals of pornography or radical groups that preach racism, or less tolerant? If one opted for the latter alternative, on the grounds that to do otherwise would express hostility to women or racial minorities, would that not also constitute a kind of intensification and enhanced policing of the bonds of community, an us/them?

Similar perplexities arise over the issue of what is called the liberal neutrality of the state. This traditional and central value of liberal politics holds that the state ought to be largely neutral *vis-à-vis* various sorts of groups; it should avoid using law to enhance or impede their cultural flourishing (see Liberalism; Neutrality, political). Postmodernism would seem to be at odds with this value. The claim is that liberal states are not in fact ever able to be completely neutral; formal neutrality merely allows culturally hegemonic groups to dominate and marginalize less powerful ones. On this view, a fairer democratic state would then be one that fiscally and institutionally tried to maintain a better balance. Of course, as with toleration, specific proposals to implement such ideas generate new quandaries. If the state is to subsidize the cultural and organizational life of some groups to a far greater degree than now, what will the criteria be for who gets funds and who does not? Will the distinction supported/unsupported carry its own invidiousness; and will it be worse than the present liberal alternative? (See Citizenship §3).

See also: Postmodernism

STEPHEN K. WHITE

References and further reading

Connolly, W. (1991) *Identity/Difference: Democratic Negotiations*, Ithaca, NY: Cornell University Press.(A careful attempt to engage a range of issues related to postmodernism and political philosophy.)

Critchley, S. (1992) ‘A Question of Politics: The Future of Deconstruction’, in *The Ethics of Deconstruction: Derrida and Levinas*, Oxford: Blackwell, 188-247.(An attempt to elucidate the idea of a postmodern democratic community.)

Derrida, J. (1985) ‘Racism’s Last Word’, trans. P. Kamuf, *Critical Inquiry* 12 (Autumn): 290-9.(Reflections on power and right from a postmodern perspective.)

Foucault, M. (1977) *Discipline and Punish*, trans. A. Sheridan, New York: Random House.(A classic model of

postmodern critique.)
Foucault, M. (1980) ‘Two Lectures’, in *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977*, ed. C. Gordon and trans. C. Gordon *et al.*, New York: Pantheon Books, 78-108.(Reflections on power and right from a postmodern perspective.)
Lyotard, J. and Thébaud, J.-L. (1985) *Just Gaming*, trans. W. Godzich, Minneapolis, MN: University of Minnesota Press.(An effort to articulate some postmodern criteria of fairness.)
McCarthy, T. (1991) ‘The Politics of the Ineffable: Derrida’s Deconstructionism’, in *Ideals and Illusions: On Reconstruction and Deconstruction in Contemporary Critical Theory*, Cambridge, MA: MIT Press.(Critique of a leading postmodern thinker.)
Taylor, C. (1984) ‘Foucault on Freedom and Truth’, *Political Theory* 12 (2): 152-83.(Critique of a leading postmodern thinker.)
White, S.K. (1991) *Political Theory and Postmodernism*, Cambridge: Cambridge University Press.(An overview of the controversy between postmodernists and their critics.)
Young, I. (1990) *Justice and the Politics of Difference*, Princeton, NJ: Princeton University Press.(An examination of the inadequacy of liberal approaches to justice for treating problems raised by radical critics such as postmodernists.)

Postmodernism, French critics of

French anti-postmodernism emerged with the generation of philosophers that came of age in the late 1970s and early 1980s and counts among its ranks some of the most visible and prolific young scholars in France. Unlike schools of thought such as phenomenology, existentialism or Marxism, French anti-postmodernism has no founding figure, central text or core doctrine; anti-postmodernism (a term seldom, if ever, used by the French) therefore is less a philosophical school than a characterization for a diverse group of thinkers who react against those trends that have dominated French intellectual life since the Second World War, especially Marxism, structuralism, existentialism and deconstruction. These trends, grouped together under the heading of 'postmodernism', are seen by anti-postmodernists as the last episodes in a failed intellectual adventure whose origins go back at least to the French Revolution.

Critical of nineteenth-century philosophy as having produced, on the one hand, totalizing, speculative philosophies such as those of Hegel and Marx, and, on the other hand, the anti-rationalism of Nietzsche and his postmodern scions, anti-postmodernists (or neo-moderns as the French prefer to say) represent a return to the concept of the individual and of history as the product of free human agency. Reaffirming the efficacy of public, rational discourse, they tend to be interested in political philosophy, taking democracy and its ideals as a model for raising and addressing philosophical issues. Pluralist in their outlook, they value the disciplinary structure of scholarly work; fields such as epistemology, theology, philosophy of science and the history of ideas which were neglected or marginalized by much postmodern thought have enjoyed renewed prestige and interest among the anti-postmodernists.

1 Historical background

Several main currents shaped French philosophy during the last half of the twentieth century; the tradition of German philosophy has been most important of all. From the nineteenth century, Hegel, Marx and Nietzsche played particularly prominent roles as did the thought of Husserl, Freud and Heidegger from the twentieth century. In addition to these, the early twentieth-century Swiss linguist Ferdinand de Saussure is credited with providing the inspiration for the emergence of structuralism during the 1950s. The 1960s and 1970s in France produced a distinctive mélange of German and structuralist philosophy that is often referred to as post-structuralist or postmodern philosophy. Though both postmodernists and anti-postmodernists would agree on these influences and the general stages of development, they disagree on their nature and significance.

After the Second World War two philosophical traditions were particularly prominent in France. Thanks to Alexandre Koyré and Alexandre Kojève during the 1930s and to Jean Hyppolite, Maurice Merleau-ponty and others during the post-war years, Hegelianism displaced neo-Kantianism as the leading school of thought (see Hegelianism). In particular, Hegel's conception of an all-encompassing dialectical history became an important resource for philosophical and political reflection during these turbulent decades. Merleau-Ponty, Emmanuel Levinas and Jean-Paul Sartre, among others, also introduced Husserl's phenomenology of consciousness and Heidegger's analysis of being and human existence to French thinkers. However, it was above all Sartre who, through his literary endeavours and political activism in addition to his philosophical works, popularized a synthesis of Marxism and existential phenomenology, an orientation that remained mainstream among the traditionally leftist French intelligentsia until the revolt of May 1968, after which its popularity waned. His existential Marxism was a humanism, that is, an ideology that envisaged an historical process of consciousness raising whereby individuals and humanity as a whole would be emancipated to their authentic mode of existence (see Existentialism).

While existential Marxism became increasingly popular from the 1930s to the mid-1960s, beginning in the 1950s there was a second important intellectual development. Inspired by Saussure's linguistics, several disciplines developed the method of analysis and theory of meaning that came to be known as structuralism (see Structuralism). Among them were the anthropology of Claude Lévi-strauss, the psychoanalysis of Jacques Lacan, and the literary theory of Roland Barthes. Of particular importance was the frame of reference for their analyses: they focused not on the 'lived experience' of an individual consciousness nor on the progressive emancipation of human nature, as did the existential Marxists, nor did they have a comprehensive ideology in the light of which the

significance of a given historical phenomenon could be adduced. For structuralists, the meaning of a word or, more generally, a ‘signifier’ is determined not by the intentions of the individual but by its differential relation to the system of signifiers that constitutes a given language; meanings are therefore relative to a specific linguistic culture rather than a product of individual human consciousness or of a universal, eschatological system such as Hegel’s. Therefore, in framing their investigations in terms of systemic structures and their instantiation in linguistic, social and psychological phenomena, the structuralists conceived of their method as scientific and objective in contradistinction to the subject-oriented, humanistic and ideological character of Hegelian, Marxist and phenomenological philosophies of consciousness.

Those philosophers educated in the 1950s and who emerged during the 1960s as dominant figures of postmodern philosophy were greatly influenced by structuralism in two particularly noteworthy respects. First, in privileging structure over process as well as the plurality of semiological systems over a universal system, they unequivocally rejected the cross-cultural and transhistorical philosophies of Marx and, especially, Hegel. Teleological history and its dialectical rationality came to be seen as an empty, neo-theological idealism divorced from facts. Second, and related to the first, the plurality of linguistic systems, and hence the plurality of possible meanings of a word or signifier, gave rise to a critique of the univocity of concepts. Above all, the concept of identity or of the self-identical subject so fundamental to philosophy since Descartes’ *cogito* came to be seen as a substanceless fiction posited by the Western European philosophical tradition (see Alterity and identity, postmodern theories of). The structuralist critique of Hegelian history and its basis in subjectivity eventually led to the reconfiguration of Marxism in the structuralist philosophy of Louis Althusser, who coined the term *anti-humanism* to describe his Marxism in which individuals were conceived of as sites in a general system of production and in which history was conceived of as a process with no subject, that is, as having no underlying unity or direction.

Georges Bataille, Gilles Deleuze and Pierre Klossowski were among those who introduced Nietzsche’s thought to France in the late 1950s and 1960s; this, coupled with the later Heidegger’s influential critique of technology and calculative reason, reinforced the anti-Hegelian elements of structuralism. Salient features of French Nietzscheanism were the critiques of identity, subjectivity and reason, the advocacy of the primordial status of language and the consequent centrality of ‘questions of interpretation’ in the place of ‘questions of fact’, and the rejection of progressivist or teleological conceptions of history. Discerning a complicity between the anthropocentrism of humanism and the metaphysics of modern philosophy inaugurated by Descartes and culminating in Kant and Hegel, postmodern philosophers influenced by Nietzsche - Heidegger, Jacques Lacan, Michel Foucault, Gilles Deleuze, Jacques Derrida, among others - proclaimed, each in their own way, ‘the death of man’, that is, the end of metaphysical humanism, the collapse of which defines the postmodern condition. In this respect they all, in one manner or another, affirm along with Nietzsche an ontological nihilism. Despite significant differences among them, they all find in Hegel the culmination of foundationalist modern philosophy as well as its foundering, and they all accept the uncertain, perspectival and ontologically nihilistic character of the contemporary age. By 1979, when Jean-François Lyotard published *La Condition Postmoderne* (*The Postmodern Condition*, 1984) which popularized the philosophical sense of ‘postmodernity’ and after the traditionally progressivist left had suffered political collapse in the election of 1978, postmodernism - Derrida’s deconstruction, Foucault’s archaeo-genealogy and Deleuze’s schizoanalysis, to name only the most prominent figures of postmodernism - had found a home in the leading universities and intellectual institutions of France (see Deconstruction; Postmodernism; Post-structuralism).

2 Points of contention

One of the most basic differences between postmodernists and anti-postmodernists is their divergent interpretations of history. Postmodernists hold that in some serious sense history and modernity have come to an end, be it with the collapse of Hegelianism, with the Nietzschean critique of metaphysics, or simply through exhaustion; their critiques and deconstructions are dedicated to driving home that the philosophical basis of progressive modernity has met its end. However, anti-postmodernists see postmodernism thereby as falling prey to a concept of historical necessity and therefore taking the collapse of speculative philosophy as determinative. Anti-postmodernists agree that the emancipatory project of Kant’s Enlightenment and the subsequent absolutization of historical reason in Hegel and Marx were ill-fated. But they do not see this as the final or inevitable conclusion of modern philosophy or historical progress. Rather, what they see ending is postmodernism, which for them is not the deconstructive revelation of a Hegelian error but its continuation. Holding that

postmodernism was too conditioned by the Enlightenment it sought to criticize, Luc Ferry proclaims that anti-postmodernists have grown beyond postmodernism and that ‘for the first time no doubt in the history of humanity, we are living in a time when this critique [of Enlightenment reason] has reached the minimal threshold of maturity’ (1995: 168).

If the metaphysics of the Enlightenment was a misadventure whose repercussions produced the terror of the French Revolution, the misology of Nietzsche and the postmodern critiques of subjectivity and reason, it nevertheless remains possible, according to anti-postmodernists, to discover other historical possibilities for thought and action than those that led to postmodernism. Accordingly, they are drawn to Alexis de Tocqueville rather than J.-J. Rousseau or Kant, to J.G. Fichte rather than Hegel or Marx and to Benjamin Constant rather than Nietzsche. Anti-postmodernists find in these and like thinkers not the death knell of history but indications of its infinitely open future horizon. For them, history is less the primary determinant (rational or otherwise) than it is a single, albeit important, element in human theoretical and practical endeavours.

A second and related basic issue at stake for anti-postmodernists also goes back to German Idealism. Although anti-postmodernists too are critical of the metaphysics of reason and its ideological excesses, they hold that the postmodern critique leads to a sterile nihilism. Vincent Descombes (1980: 182) sums up this complaint when he says for postmodernism there is:

> no *original*, the model for the copy is itself a copy … no *facts*, only interpretations and any interpretation is itself the interpretation of an older interpretation; there is no *meaning proper* to words, only figurative meanings … no *authentic version* of a text, there are only translations; *no truth*, only pastiche and parody.

And as Luc Ferry claims, this anti-humanism cripples the individual and leads to authoritarianism, either by giving rise to a panoply of master thinkers (such as Hegel, Marx, Nietzsche, Heidegger, Foucault, Derrida) who produce hermetic systems of oracular discourse that admit of no falsification, or by elevating non-human things, such as the natural environment, to a point where humanity is rendered insignificant if not odious (see Ecological philosophy §1; Næss, A.). According to Ferry, the challenge of anti-postmodernism is to develop a non-metaphysical humanism, that is, a conception of rationality and the subject that leads neither to the ideological excesses of the post-Enlightenment, nor to their debasement in postmodernism.

The anti-postmodernist emphasis on the individual, reason and their vision of the future as open to determination by human spontaneity indicates that freedom and the belief in the efficacy of rational discourse are central to anti-postmodernist thought and also reveals its predominantly political orientation. Often identifying with the seventeenth-century ‘quarrel of the ancients and the moderns’ in which modern, scientifically oriented philosophy rejected the theological, aesthetical and moralizing thought of the Renaissance, anti-postmodernists tend to be concerned with that open space in which communal thought and activity are constituted and which is supported by democratic processes. In this regard, the communicative demands of cooperative political activity valued by anti-postmodernism is clearly at odds with the effacing of the distinction between fact and fiction wrought by structuralism and extended by the Nietzschean, postmodern critique of reason.

The anti-postmodern appeal to common sense has capitalized both on some of the extravagances of postmodernism and on its relatively esoteric character. However, although their concerns are substantive and their criticisms often illuminating, anti-postmodernists often indulge in characterizations; more importantly, anti-postmodernists have yet to address adequately issues raised by postmodernism that touch on the viability of a positive conception of anti-postmodernism. Among these are the constitution of the conscious subject, the nature of language and its relevance for discourse and knowledge, the manner in which the past bears upon and influences the future, and the possibility and value of a non-metaphysical rationality. Its ability to address these and related matters will determine whether anti-postmodernism signals the epochal shift many of its members claim is at hand, or whether it proves to be only the most recent reactionary philosophical tantrum.

See also: Alterity and identity, postmodern theories of; Subject, postmodern critique of the

REGINALD LILLY

References and further reading

Descombes, V. (1980) *Modern French Philosophy*, trans. L. Scott-Fox and J.M. Harding, Cambridge: Cambridge

University Press.(One of the earliest and best overviews of French postmodern philosophy from an anti-postmodernist point of view.)
Ferry, L. (1995) *The New Ecological Order*, trans. C. Volk, Chicago, IL: University of Chicago Press, 1995.(A clearly written extension of anti-postmodernist thought applied to ecological movements.)
Ferry, L. and Renaut, A. (1990) *French Philosophy of the Sixties: An Essay on Antihumanism*, trans. M. Schanckenberg, Amherst, MA: University of Massachusetts Press.(A call to arms by two leading anti-postmodernists.)
Lilla, M. (ed.) (1994) *New French Thought: Political Philosophy*, Princeton, NJ: Princeton University.Press.(A representative selection of articles by anti-postmodernists focusing on political thought.)
*** Lyotard, J.-F.** (1979) *La Condition postmoderne: rapport sure le savoir*, Paris: Éditions de Minuit; trans. G. Bennington and B. Massumi, *The Postmodern Condition*, Minneapolis, MN: University of Minnesota Press, 1984.(A central text in postmodernism; somewhat challenging for the beginner.)
Nora, P., Gauchet, M. and Pomian, K. (eds) (1988) *Le Débat* 50 (May-August).(An important retrospective by one of the most important journals for anti-postmodernist thought; very clear for those who can read French.)
Pavel, T.G. (1989) ‘The Present Debate: News From France’, *Diacritics* 19 (1).(An excellent summary discussion of French philosophy in the 1980s.)

Post-structuralism

Post-structuralism is a late-twentieth-century development in philosophy and literary theory, particularly associated with the work of Jacques Derrida and his followers. It originated as a reaction against structuralism, which first emerged in Ferdinand de Saussure's work on linguistics. By the 1950s structuralism had been adapted in anthropology (Lévi-Strauss), psychoanalysis (Lacan) and literary theory (Barthes), and there were hopes that it could provide the framework for rigorous accounts in all areas of the human sciences.

Although structuralism was never formulated as a philosophical theory in its own right, its implicit theoretical basis was a kind of Cartesianism, but without the emphasis on subjectivity. It aimed, like Descartes, at a logically rigorous system of knowledge based on sharp explicit definitions of fundamental concepts. The difference was that, for structuralism, the system itself was absolute, with no grounding in subjectivity. Post-structuralist critiques of structuralism typically challenge the assumption that systems are self-sufficient structures and question the possibility of the precise definitions on which systems of knowledge must be based.

Derrida carries out his critique of structuralist systems by the technique of deconstruction. This is the process of showing, through close textual and conceptual analysis, how definitions of fundamental concepts (for example, presence versus absence, true versus false) are undermined by the very effort to formulate and employ them. Derrida's approach has particularly influenced literary theory and criticism in the USA. In addition, Richard Rorty, developing themes from pragmatism and recent analytic philosophy, has put forward a distinctively American version of post-structuralism.

1 Structuralism

In his lectures on linguistics, Ferdinand de Saussure proposed a view of language (*langue*) as a formal structure, defined by differences between systemic elements. According to Saussure, this structure is simultaneously present in and unites the two domains of thought and words. A given linguistic term (a sign) is the union of an idea or concept (the 'signified') and a physical word (the 'signifier'). A language is a complete system of such signs, which exists not as a separate substance but merely as the differentiating form that defines the specific structure of both signifiers (physical words) and signifieds (ideas). Saussure's view rejects the common-sense picture of the set of signifiers and the set of signifieds as independent givens, with the signifieds having meaning in their own right and the signifiers obtaining meaning entirely through their association with corresponding signifieds. Saussure denies this independence and instead maintains that signifiers and signifieds alike have meaning only in virtue of the formal structure (itself defined by differences between elements) that they share (see Structuralism in linguistics).

Saussure's structuralist approach was very successful within linguistics, where it was applied and extended by, among others, Jakobson and Troubetzkoy. By the 1950s the approach had been adapted in anthropology (Lévi-Strauss), psychoanalysis (Lacan), and literary theory (Barthes); and there were hopes that it could provide the framework for rigorous accounts in all areas of the human sciences. Three distinguishing features of this framework were: (1) a rejection of all idealist views of concepts and meanings as derived from the activity of consciousness; (2) an understanding of concepts and meanings as, instead, grounded in the structural relations among the elements of abstract systems; (3) an explication of such structural relations solely in terms of bipolar differences (for example, real/unreal, temporal/nontemporal, present/absent, male/female).

2 Post-structuralism: terminology

Post-structuralism is obviously closely tied to structuralism, but commentators have characterized the relationship in a variety of mutually inconsistent ways. Some writers make no distinction between structuralism and post-structuralism, applying the single term 'structuralist' to the entire range of thinkers from Saussure through to Derrida. More commonly, post-structuralism is distinguished as a separate development, but there is disagreement as to whether it is primarily a reaction against structuralism or an extension of it (as the term *Neostrukturalismus*, commonly used by Manfred Frank and other German commentators, suggests). Apart from matters of definition, there is even disagreement as to whether major figures such as Barthes, Lacan and Foucault are structuralists or post-structuralists.

Michel Foucault's book, *Les mots et les choses* (translated under the title *The Order of Things*) is an instructive example. In one sense it is quintessentially structuralist. The book first uncovers the fundamental epistemic systems (which Foucault calls 'epistemes') that underlie and delimit the subjective thought of particular eras. It then goes on to show how the apparent ultimacy of subjectivity is itself just the product of one contingent episteme, that of modernity, which is even now disappearing (the famous 'death of man'). Nevertheless, Foucault's essentially historical viewpoint in the work demonstrates the limitation of structuralism: its inability to give any account of the transitions from one system of thought to another. Foucault seems to have seen from the beginning that structuralism can not be historical, a fact that explains his constant insistence that he was not a structuralist, in spite of his obvious deployment of structuralist methods and concepts. So, although *Les mots et les choses* is a structuralist book, it at the same time makes clear the limits of structuralism and prepares the way for Foucault's later work on power and ethics which is distinctly post-structuralist (see Foucault, M.; Post-structuralism and the social sciences).

Despite these ambiguities and disagreements, the concept of post-structuralism is useful, if not essential, for understanding philosophy in France during the latter part of the twentieth century. One fruitful approach is to think of post-structuralism as a philosophical reaction to the structuralism that was such a powerful force during the 1960s in linguistics, psychology and the social sciences. It was neither a simple rejection or extension of structuralism but a series of philosophical reflections on the structuralist programme and achievement.

3 Two major post-structuralist theses

Although structuralism was never formulated as a philosophical theory in its own right, its implicit theoretical basis was, as noted above, a kind of Cartesianism without the subject. (Hence, the association of structuralism with the notion of the 'death of the subject'.) Post-structuralist critiques of structuralism are typically based on two fundamental theses: (1) that no system can be autonomous (self-sufficient) in the way that structuralism requires; and (2) that the defining dichotomies on which structuralist systems are based express distinctions that do not hold up under careful scrutiny.

The first thesis is not understood so as to support the traditional idealist view that systematic structures are dependent on the constitutive activities of subjects. Post-structuralists retain structuralism's elimination of the subject from any role as a foundation of reality or of our knowledge of it. But, in opposition to structuralism, they also reject any logical foundation for a system of thought (in, for example, its internal coherence). For post-structuralists, there is no foundation of any sort that can guarantee the validity or stability of any system of thought.

The second thesis is the key to post-structuralism's denial of the internal coherence of systems. The logical structure of a system requires that the applications of its concepts be unambiguously defined. (In the formalism of elementary number theory, for example, there must be no question as to whether a given number is odd or even.) As a result, the possibility of a systematic structure depends on the possibility of drawing sharp distinctions between complementary concepts such as odd/even, charged/uncharged, living/non-living, male/female and so on. Post-structuralist philosophers have been particularly concerned with the fundamental dichotomies (or oppositions) underlying structuralist theories in the human sciences. Saussure's linguistics, for example, is based on the distinction of the signifier from the signified; Lévi-Strauss's anthropology of myths employs oppositions such as raw/cooked, sun/moon and so on. In each case, post-structuralists have argued that the dichotomy has no absolute status because the alternatives it offers are neither exclusive nor exhaustive.

4 Derrida's critique of logocentrism

This sort of critique was extended to philosophy, particularly by Jacques Derrida, who finds Western philosophical thought pervaded by a network of oppositions - appearance/reality, false/true, opinion/knowledge, to cite just a few examples - that constitute what he calls the system of 'logocentrism'. This term derives from Derrida's conviction that at the root of Western philosophical thought is a fundamental distinction between speech (*logos*) and writing. Speech is privileged as the expression of what is immediate and present, the source, accordingly, of what is real, true and certain. Writing, on the other hand, is derogated as an inferior imitation of speech, the residue of speech that is no longer present and, therefore, the locus of appearance, deceptions and uncertainty. Plato's devaluation of writing in comparison with living dialogue is the most famous and influential example of this distinction. But

Derrida finds the distinction pervading Western philosophy and regards it as not just a preference for one form of communication over another but the basis for the entire set of hierarchical oppositions that characterize philosophical thought. Speech offers presence, truth, reality, whereas writing, a derivative presentation employed in the absence of living speech, inevitably misleads us into accepting illusions.

Derrida's critiques of the speech/writing opposition - and of all the hierarchical oppositions that attend it - proceed by what he calls the method of 'deconstruction' (see Deconstruction). This is the process of showing, through close textual and conceptual analysis, how such oppositions are contradicted by the very effort to formulate and employ them. Consider, for example, the opposition between presence and absence, which plays a fundamental role in Husserl's phenomenology (and many other philosophical contexts). Husserl requires a sharp distinction between what is immediately present to consciousness (and therefore entirely certain) and what is outside of consciousness (and therefore uncertain). But once Husserl undertakes a close analysis of the immediately present, he discovers that it is not instantaneous but includes its own temporal extension. The 'present', as a concrete experiential unit, involves both memory of the just-immediately-past (retention, in Husserl's terminology) and anticipation of the immediate future (protention). Thus, the past and the future, both paradigms of what is absent (not present), turn out to be integral parts of the present. Husserl's own account of the presence/absence opposition overturns it.

Deconstruction maintains that there is no stability in any of thought's fundamental oppositions. Their allegedly exclusive alternatives turn out to be inextricably connected; their implicit hierarchies perpetually reversible. As a result, there is an ineliminable gap between the intelligibility of a rational system and the reality it is trying to capture. Derrida expresses this gap through a variety of terms. He frequently speaks of *différance* (a deliberately misspelled homophone of the French *différence*) to emphasize, first, the difference between systematic structures and the objects (for example, experiences, events, texts) they try to make intelligible, and, second, the way in which efforts to make absolute distinctions are always deferred (another sense of the French *différer*) by the involvement of one polar opposite in the other. This latter phenomenon Derrida also discusses in terms of the 'trace' of its opposite always lingering at the heart of any polar term. He also employs the term 'dissemination' to refer to the way that objects of analysis slip through the conceptual net spread by any given system of intelligibility we devise for it.

5 Post-structuralism and literary theory

Thus far the discussion has focused on Derrida's deconstruction of the meaningful structures philosophers purport to find in reality and to express in their philosophical texts. But Derrida's approach is also readily applicable to literary texts (and the 'worlds' they create). This is because - like philosophical systems - poems, novels and other literary texts are typically thought to embody complete and coherent systems of meaning, which it is the task of literary criticism to extract. Although Derrida himself has dealt primarily with philosophical texts, his approach has been widely adopted by analysts of literature. (Of course, as should be expected, Derrida and his followers reject any sharp distinction between the philosophical and the literary.)

Traditional literary analysis has understood the meaning of a text as the expression of its author's mind; that is, as thoughts the author intended to convey in writing the text. The first stage of deconstructive criticism is the structuralist one of detaching meaning from authorial intention, locating it instead in the text itself as a linguistic structure. Roland Barthes, for example, showed how to analyse a text by Balzac entirely in terms of the formal codes it embodies, with no reference to what Balzac supposedly 'meant'. This structuralist move effects a 'death of the author' parallel to the anti-Cartesian 'death of the subject'. But the post-structuralists take the future step of denying a fixed meaning to even the autonomous text itself. It is not that a text lacks all meaning but that, on the contrary, it is the source of an endless proliferation of conflicting meanings. As deconstructionists delight in showing, any proposed privileged meaning of a text can be undermined by careful attention to the role in it of apparently marginal features. (For example, an orthodox Christian reading of Milton's *Paradise Lost* is deconstructed by a close study of certain details in its treatment of Satan.) There is no doubt, of course, that texts are often produced by authors trying to express what they think or feel. But what they write always goes beyond any authorial intention and in ways that can never be reduced to a coherent system of meaning.

The deconstructionist's point can also be understood as an undermining of the distinction between primary text and commentary. On the traditional view, a commentary is an effort to formulate as accurately as possible the

content (meaning) of the text. To the extent that it is successful, a commentary expresses nothing more and nothing less than this meaning. But for deconstructionists the meaning in question does not exist, and the commentary must be understood as nothing more than a free elaboration of themes suggested, but not required, by the text. Unable to be a secondary reflection, the commentary becomes as much an independent creation as the text itself.

6 Rorty's post-structuralist pragmatism

Richard Rorty's work is far removed, in both antecedents and style, from that of continental post-structuralists (see Rorty, R.). His critique of Cartesianism, derived more from Dewey than from Heidegger, is aimed at twentieth-century analytic philosophy rather than the structuralist human sciences; and his urbanely lucid prose contrasts sharply with the wilfully playful convolutions of Derrida and his followers. None the less, Rorty's analyses lead him to a critique of traditional philosophy very similar to that of the post-structuralists.

The focal point of Rorty's critique is the project (called foundationalism) of providing a philosophical grounding for all knowledge. Modern foundationalism originates with Descartes, but Rorty sees it as also the *leitmotif* of Descartes' successors, through Hume and Kant down to the logical positivists. Like Derrida, Rorty attacks traditional systematic thought by calling into question some of its key distinctions. Unlike Derrida, however, he does not carry out his attacks through close readings of classic texts but by deploying the results of recent analytic philosophy. He uses, for example, Quine's critique of the distinction between analytic and synthetic statements to argue that there are no foundational truths about the meaning of concepts. He appeals to Wilfrid Sellars's undermining of the distinction between theory and observation to reject empirical foundations of knowledge in interpretation-free sense data. He employs Donald Davidson's questioning of the distinction between the formal structure and the material content of a conceptual framework to reject Kantian attempts to ground knowledge in principles that define the framework of all possible thought.

In Rorty's view the upshot of these various critiques is to cut off every source of an ultimate philosophical foundation for our knowledge. Accordingly, he maintains, philosophy must give up its traditional claim to be the final court of appeal in disputes about truth. We have no alternative but to accept as true what we (the community of knowers) agree on. There is no appeal beyond the results of the 'conversation of mankind' so far as it has advanced to date. For us, there is no (upper-case) Truth justified by privileged insights and methods. There is only the mundane (lower-case) truth: what our interlocutors let us get away with saying.

It might seem that this rejection of foundationalism is a rejection of the entire tradition of Western philosophy since Plato. Rorty, however, distinguishes two styles of philosophy. First, there is *systematic philosophy*, the mainline of the Western tradition since Plato, which is defined by the foundationalist goal of ultimate justification. But, on the other hand, there is another enterprise, always marginal to the tradition, that Rorty calls *edifying philosophy*. Whereas systematic philosophers undertake elaborate and purportedly eternal constructions (which are always demolished by the next generation), edifying philosophers are content to shoot ironic barbs at the systematic thought of their day, exploding its pretensions and stimulating intriguing lines of counter-thought. The tradition of edifying thought can be traced back at least to the ancient Cynics and has been more recently represented by Kierkegaard, Nietzsche and the later Wittgenstein. Derrida's deconstructions are, on Rorty's view, a prime contemporary example of edifying philosophy.

Edifying philosophers, however, are philosophers only because they react against systematic philosophy. They do not differ from other sorts of cultural critics (novelists, literary theorists, social scientists) because of any distinctively philosophical method or viewpoint. If, in the wake of thinkers such as Derrida and Rorty, systematic philosophy is abandoned, philosophy will be too. The triumph of post-structuralism would, for better or worse, be the end of philosophy as we have known it.

See also: Postmodernism

GARY GUTTING

References and further reading

Culler, J. (1982) *On Deconstruction: Theory and Criticism after Structuralism*, Ithaca, NY: Cornell University Press.(Excellent discussion of Derrida's views and of their impact on literary criticism.)

Dews, P. (1987) *Logics of Disintegration: Poststructuralist Thought and the Claims of Critical Theory*, London:

Verso.(Analytic and critical essays on Derrida, Lacan, Foucault and Lyotard.)

Foucault, M. (1966) *Les mots et les choses: une archéologie des sciences humaines*, Paris: Gallimard; trans. A. Sheridan, *The Order of Things*, New York: Random House, 1970.(A critical history of the origins of the modern social sciences.)

Frank, M. (1989) *What Is Neostructuralism?*, trans. S. Wilke and R. Gray, Minneapolis, MN: University of Minnesota Press.(A detailed study and critique of the movement.)

Rorty, R. (1989) *Contingency, Irony, and Solidarity*, Cambridge: Cambridge University Press.(A collection of essays that formulates Rorty's position and comments on the views of other post-structuralists.)

Rorty, R. (1991) *Essays on Heidegger and Others*, Cambridge: Cambridge University Press.(Another collection of essays that formulates the author's position and comments on the views of other post-structuralists.)

Sturrock, J. (ed.) (1979) *Structuralism and Since*, Oxford: Oxford University Press.(A collection of introductory essays on the major figures.)

Post-structuralism in the social sciences

Structuralism was a twentieth-century approach in various social scientific disciplines (the 'human sciences') that promised to put them on a solid scientific basis. The origin and model of structuralism was Saussure's work in linguistics. Saussure's approach was later adapted in anthropology (by Lévi-Strauss), in psychoanalysis (by Lacan), and in literary theory (by Barthes). The core of structuralism was the treatment of distinctively human domains as formal structures in which meanings were constituted not by conscious subjects but by relations among the elements of a formal system.

Post-structuralism comprises a variety of reactions to structuralism, primarily by philosophers such as Derrida, Foucault, Deleuze, and Lyotard. The first stage of post-structuralist reflection on the social sciences, best represented by Foucault's discussion at the end of The Order of Things, accepted the structuralist approach and drew from it the philosophical consequences for our understanding of the social sciences. A later stage (including Foucault's Discipline and Punish and Lyotard's writings from the 1970s on) sought to transform our conception of the human sciences through a critique of structuralist presuppositions. In this later stage, post-structuralists continued to accept structuralism's elimination of the conscious subject but maintained that human existence could not be adequately understood without taking account of non-structural causal factors such as power (Foucault) and desire (Lyotard). Foucault argued for the inextricable tie between our knowledge of society and society's power structures. Lyotard maintained that Lacan's structuralist version of psychoanalysis ignored the way that desire corresponds to a reality that escapes the boundaries of any formal structure. He also developed a political and ethical stance based on the fundamental value of a plurality of desires.

1 Structuralism in the social sciences

Structuralism originated in Ferdinand de Saussure's linguistics, which treated language as a formal system. Saussure rejected the common, representationalist view of language as a set of physical objects (for example, sounds or marks) that have meaning only in the sense that they represent mental entities (ideas), which alone are intrinsically meaningful. On his picture, both signifiers (sounds or marks) and signifieds (ideas) have meaning simply in virtue of the formal structure that they share. This formal structure is the system of identities and differences that exist among linguistic elements (on one level, sounds; on another level, ideas). Saussure identified language (*langue*) with this formal structure (see Saussure, F. de).

Lévi-Strauss' cultural anthropology is a prime example of the extension of Saussure's view of language to the social sciences. He applies the basic approach of structural linguistics to anthropological phenomena such as kinship systems and myths. Lévi-Strauss describes such phenomena in terms of the signifier/signified distinction. Thus, for the case of kinship relations, the signifieds are the thought of a culture about family relationships (for example, about degrees of and respect due to such relationships, incest taboos), while the signifiers are the specific practices (customs, rituals, and so on) that 'express' these thoughts. As in the case of Saussurean linguistics, Lévi-Strauss' approach contrasts with a representationalist one. He does not, for example, see a society or culture as having certain ideas about kinship that are then implemented by practices corresponding to them (with the practices being material images of the society's self-understanding). Rather, both the ideas and the practices are specified by their shared formal structure, once again understood in terms of differences among elements within a system. We are, accordingly, able to understand a society's ideas and practices with no reference to the minds (thoughts or intentions) of subjects (see Lévi-Strauss, C.).

2 Foucault's archaeology of structuralism

Foucault's *The Order of Things* supports the structuralist elimination of a central role for subjectivity (see Foucault, M.). Using the term 'man' to denote subjectivity as the source and locus of mental representations of reality, he first argues that the category of 'man' was merely a contingent feature of modern (roughly, nineteenth-century) thought and as such has no essential role in accounts of human life and thought. In his concluding chapter, he further argues that the recent development of structuralist social sciences, particularly Lévi-Strauss' anthropology and Lacan's psychoanalysis, points the way to a mode of thought that does not rely on this category. Such sciences, he maintains, show how we could conceive of human reality without defining it in terms of its capacity for subjective representation. Foucault agrees that the modern social sciences - psychology

and sociology, for example - are, like philosophy since Kant, based on the primacy of subjectivity. He even maintains that these ‘sciences of man’ solve the problem (which had frustrated philosophers since Kant) of how man could be both the constituting source of the world’s meaning and, at the same time, just another natural object in the world.

In Foucault’s view, the human sciences solve this problem by introducing the notion of the unconscious mind. Operating solely on the level of consciousness, philosophers could provide no coherent account of how man could be both an object within the world, fully describable in term of the fundamental categories (life, labour, and language) of the human sciences, and at the same time the transcendental subject that constitutes the world and all its objects. But, Foucault maintains, the paradox disappears once we think of the world, including man, as constituted by an unconscious mind that is not itself an object in the world. However, he further maintains that the standard modern social sciences merely describe the results of the operations of the unconscious mind; they are not able to give a direct account of the nature of the unconscious and of the conditions of its possibility. Such an account is provided only by the structuralist social sciences: Lacan’s psychoanalysis and Lévi-Strauss’ cultural anthropology.

According to Foucault, the key difference between ordinary human sciences and the structuralist sciences is that, unlike the former, the latter are able to explain consciousness and its representations of the world in terms of more fundamental principles. Specifically, Lacan and Lévi-Strauss provide a description of (respectively) psychological and cultural structures that are not themselves representational but that explain the functioning of conscious representations.

Foucault notes that in one sense these structuralist accounts provide a foundation for the modern human sciences by explaining their basic category of man. But he also notes that, in so doing, they destroy the centrality of this category. The deepest understanding of human reality is no longer in terms of the subject and its representations but in terms of unconscious structural systems. Because they thus undermine the fundamental category of the human sciences, Foucault labels these structuralist disciplines ‘countersciences’.

3 Foucault’s genealogy of power

Despite its strong structuralist sympathies, *The Order of Things* implicitly raises serious questions about structuralism that adumbrate Foucault’s later post-structuralist viewpoint. The most serious issue concerns the status of Foucault’s own methodology in the book. For all his praise of structuralism’s surpassing modern thought and heralding the ‘death of man’, it is not at all clear that his own approach is structuralist. At root, this is because he writes as a historian, concerned with the temporal (diachronic) development of thought, while structuralist accounts can yield no more than synchronic time-slices of intellectual systems.

The Order of Things avoids this issue by restricting itself to questions about intellectual systems (*epistemes* in Foucault’s terminology) at a given time. But, as he makes clear at several points (most notably in his Foreword to the English translation), there is ultimately no avoiding the essential historical question of the causes of transitions from one *episteme* to another. When Foucault subsequently takes up the causal question (particularly in *Discipline and Punish* and in volume one of *The History of Sexuality*), his approach becomes explicitly post-structuralist.

At the heart of Foucault’s post-structuralism is his notion of power. He sees relations of power as the causal factors that lead from one *episteme* to another. Power as Foucault understands it has three key features, each of which puts it beyond the ordered systems of structuralism. First, power is productive. It not only expresses the repressive, exclusionary force of a system’s constraints but also creates new domains of knowledge and practice. Second, power is not located in any single control-centre; it is dispersed throughout the social system in innumerable local seats of power. These seats interact with one another but do not form a unified system. Third, although it is intimately related to systems of knowledge, power is more than the play of signifiers and signifieds within such systems. It is ultimately the formative action of one body on another.

Foucault’s explicit recognition of the relation of power to knowledge has important consequences for his understanding of the social sciences. These consequences are most explicitly developed by his study, in *Discipline and Punish*, of the connections between modern methods of social control (such as the prison) and the emergence of criminology and related social scientific disciplines.

Foucault presents *Discipline and Punish* as a genealogical history; that is, one that details the causal processes leading to significant changes in the history of thought. Genealogy concerns the connection between non-discursive practices and systems of discourse (bodies of knowledge). In this regard, Foucault's central claim is that there is an inextricable interrelation of knowledge (discourse) and power (expressed in non-discursive practices, in particular, the control of bodies).

In emphasizing the relation of knowledge and power, Foucault does not have in mind the standard Baconian idea, which sees knowledge as first existing as an autonomous achievement (pure science) that is later used as an instrument of action (technology). He maintains that knowledge simply does not exist in complete independence of power, that the deployment of knowledge and the deployment of power are simultaneous from the beginning. On the other hand, Foucault does not go so far as to identify knowledge with power; he does not, for example, make knowledge nothing more than an expression of social or political control. As he points out, if he thought there was no difference between knowledge and power, he would not have had to take such pains to discover the precise ways in which they are related to one another.

Foucault's positive view is that systems of knowledge, although expressing objective (and perhaps even universally valid) truth in their own right, are none the less always more or less closely tied to the regimes of power that exist within a given society. Conversely, regimes of power necessarily give rise to bodies of knowledge about the objects they control (but this knowledge may - in its objectivity - go beyond and even ultimately threaten the system of domination from which it arises).

In *Discipline and Punish*, Foucault considers the specific case of the connection between modern social scientific disciplines and the disciplinary practices used to control human bodies in the modern period. His primary example is the relation of the practice of imprisonment to criminology and other social scientific disciplines that deal with crime and punishment (for example, social psychology). But he treats imprisonment in the context of modern disciplinary practices in general (as employed in schools, factories, the military, and so on); and he shows how the prison served as a model and centre of diffusion for this whole range of disciplinary practices. These practices form the matrix from which spring many of the most important modern social sciences. In the first volume of his *The History of Sexuality*, Foucault argues further that other modern human sciences - for example, psychoanalysis - are tied to modern mechanisms for controlling sexual behaviour.

4 The unconscious and desire: from Lacan to Lyotard

Just as Foucault argues the limitations of structuralist accounts of the social in the face of the reality of power, so Jean-François Lyotard emphasizes the limits of structuralist psychology in the face of the reality of desire. In order to understand Lyotard's critique, it is necessary to place it in relation to the most important structuralist approach to psychology: Jacques Lacan's formulation of Freudian psychoanalysis.

The heart of Lacan's formulation is his claim that the unconscious mind has the structure of a language. Here he is understanding 'language' in Saussure's sense of a system of signs whose meanings are entirely defined by their roles in the system. As such, the unconscious has no essential connections to external objects and its desires (drives) refer to nothing beyond the semiological system from which they receive their meaning. Orthodox psychoanalysis (for example, Heinz Hartmann's ego psychology) regards the mature conscious ego and the 'objective' adult world to which it accommodates itself as the reality to which unconscious desires must be subordinated. Lacan, on the contrary, places this ego and its world in the domain of the 'Imaginary' and insists on its subordination to the 'Symbolic' (that is, the unconscious as an autonomous sign system). Lacan also allows for a domain of 'Reality', but he presents this as an unreachable limit of symbolic structures. Accordingly, desire is nothing more than a lack that can, in principle, never be fulfilled (see Lacan, J.).

In opposition to Lacan - and to many other structuralist and post-structuralist thinkers - Lyotard asserts the autonomy and primacy of the non-linguistic object. This is not to say that he revives the foundationalist claim (often associated with phenomenology) that the object is 'simply given' to the mind in an experience unformed by any linguistic categories. Lyotard agrees that there is no pre-linguistic experience. But, he maintains, it does not follow that the content of experience is exhausted by language. 'We can say that the tree is green, but this does not put the colour into the sentence' (1971: 52).

Lyotard expresses the independent reality of the object through his distinction between 'letter' and 'line'. By

‘letter’ Lyotard means a linguistic sign as understood by Saussure: an element in a system that has significance solely in virtue of its differences from other elements in a system. If, for example, a Magritte painting contains the written message, ‘This is not a pipe’, the letters used can have any physical shape at all (so long as those shapes allow us to distinguish a *t* from a *p*, and so on) and still convey the message. But, as lines in the painting, the precise shapes of the *p*, *t* and so on, are crucial. A thick red script will convey one aesthetic meaning, a thin green one another. For the *p* and *t* as letters, pre-linguistic content is irrelevant, but for them as lines it is crucial. For ‘This is not a pipe’ as read, there is no content irreducible to the linguistic system; but for ‘This is not a pipe’ as seen, there is. To use a parallel distinction of Lyotard’s, the expression as discourse is not reducible to the expression as figure.

Given this account, Lyotard is able to reject Lacan’s construal of desire as simply the lack of an object that can never be attained. Just as perception is fulfilled by an object with an intrinsic content that cannot be reduced to the linguistic structures of consciousness, so too desire can be fulfilled by an object irreducible to the linguistic structures of the unconscious. In this way, desire takes us beyond the boundaries of a purely structuralist understanding of the unconscious.

5 Lyotard’s politics of desire

Lyotard’s post-structuralist account of desire also has consequences for social and political thought. As Peter Dews (1987) has pointed out, psychological desire and political power must be thought of as necessary counterparts. Desire is precisely that which power constrains, and desire that which struggles against power. It should, therefore, be no surprise that desire is a fundamental category of Lyotard’s social and political thought. Specifically, Lyotard tried to develop a ‘libidinal politics’: a theoretical and practical standpoint based on the fundamental value of the flourishing of a plurality of diverse desires. Lyotard, like Foucault, saw knowledge and power as essentially connected, and maintained that ‘totalizing’ theories (for example, Marxism) claiming universal validity are sources of totalitarian social structures that destroy the plurality of desires.

Lyotard pays particular attention to what he sees as the inevitable conflict between justice and truth. We continually - and properly - make judgments that particular situations or actions are unjust. But we are also inclined to think that these judgments themselves require justification through derivation from a general account of the nature of human society. We think, that is, that particular prescriptions regarding justice must be justified by the truth of general theoretical descriptions. Lyotard, however, maintains that this appeal to general truth is itself an instance of injustice. The general description presents a total picture of society that excludes all alternative views as false and rejects as unacceptable desires based on these views. But, according to Lyotard, such an exclusion is inconsistent with the fundamental value of maintaining a plurality of desires. He holds, accordingly, that we should be content with ‘indeterminate’ judgments about particular instances of justice and injustice and eschew ‘determinate’ judgments intended to justify indeterminate judgments. (A judgment is determinate or indeterminate depending on whether or not it is ‘determined’ by a general theoretical principle.)

Lyotard also formulates his view of justice in terms of his concept of the ‘differend’. A differend is an incommensurability between different standpoints (language games, narratives - ultimately, desires). The sign of incommensurability is precisely the lack of any shared criteria for mediating the differences that define the differend. Lyotard holds that the goal of politics (and also of ethics and art) should be to produce and preserve differends; and, particularly, to protect them from the ‘totalitarian terror’ of exclusionary global truth-claims.

Lyotard’s work is in many ways the most thorough and the most radical development of a post-structuralist view of society and politics. There is no doubt that his rejection of social and political theory takes us far beyond the framework of traditional discussions in philosophy and the social sciences. There will be continuing disagreement as to whether this rejection is a bracing liberation or a debilitating confusion.

See also: Barthes, R.; Post-structuralism; Semiotics; Structuralism in social science

GARY GUTTING

References and further reading

Dews, P. (1987) *Logics of Disintegration: Poststructuralist Thought and the Claims of Critical Theory*, London: Verso.(Includes analytic and critical essays on Lacan, Foucault, and Lyotard.)

Foucault, M. (1966) *Les mots et les choses: une archéologie des sciences humaines*, Paris: Gallimard; trans. A. Sheridan, *The Order of Things*, New York: Random House, 1970.(A critical history of the origins of the modern social sciences.)

Foucault, M. (1975) *Surveiller et punir: naissance de la prison*, Paris: Gallimard; trans. A. Sheridan, *Discipline and Punish*, New York: Pantheon, 1977.(A geneological study of the modern prison and, more generally, of the disciplinary practices of modern society. The work introduces Foucault's influential views on knowledge and power.)

Foucault, M. (1976) *Histoire de la sexualité*, vol. 1: *la volonté de savoir*, Paris: Gallimard; trans. R. Hurley, *The History of Sexuality*, New York: Pantheon, 1978.(An introduction to a projected, but never completed, series of studies on aspects of modern sexuality.)

Lyotard, J.-F. (1971) *Discours, Figure*, Paris: Klincksieck.(A radical deconstructive approach to aesthetics which criticizes the structuralist model for language and especially the Lacanian interpretation of Freud.)

Lyotard, J.-F. (1983) *Le Différend*, Paris: Les Éditions de Minuit; trans. G. Van Den Abbeele, *The Differend*, Minneapolis, MN: University of Minnesota Press, 1988.(Defines the political in terms of the fundamental 'differends' (disputes) which cannot be equitably resolved for lack of a universal rule of judgment applicable to all of the arguments.)

Readings, B. (1991) *Introducing Lyotard*, London: Routledge.(A comprehensive survey of Lyotard's thought, emphasizing his work on art and politics.)

Potentiality, Indian theories of

*Indian philosophers wrote a great deal about potential (*śakti*) and capacity (*sāmarthya*); both of these words may also be translated as 'power' or 'force'. The Sanskrit word* śakti*, like the English word 'potential', derives from a modal verb meaning 'to be able', so one might also see the study of potential as being the study of ability. The principal issue about which Indians debated was that of where exactly potential or ability is located. If, for example, it can be said that a person has a potential to compose a poetic masterpiece, then it can be asked where exactly that potential is located. Is it located entirely within the potential poet, or is it distributed somehow between the poet and the circumstances in which the person functions? If one says it is located entirely within the person, then it can be asked why they produce the masterpiece only at one specific time in their life instead of earlier or later. In other words, it seems that if circumstances were required to enable the person to realize their potential, then the potentiality would belong as much to external circumstances as to the person. This may not seem an interesting question in the abstract, but it became of interest to philosophers in particular contexts. It was generally agreed, for example, that nothing happens without a prior potential that becomes actualized. If this is so, then the question naturally arises, to what did the potential belong out of the actualization of which the universe 'happened'? Other questions that Indian philosophers debated were: To whom does the potential to become wise belong? How is the potential of a supposedly eternal language actualized to convey meaning? Attempting to answer such questions was a preoccupation of philosophers of every Indian school.*

1 God's potential to create

Among the philosophical traditions in classical India, some were atheistic and others theistic. The issue of whether there existed a single creator for the entire universe was one that took on many dimensions as each side produced arguments and counter-arguments of increasing sophistication. Among the many dimensions that came up for discussion, one centred on the question of whether it made sense to say of a single creator that it had a potential (*śakti* or *sāmarthya*) to create, and if it did have such a potential, how that potential could be realized. If it could be shown that it made no sense to speak of a single creator having a potential to create something complex, argued the atheists, then the whole idea of creation through a single creator could be dismissed as unworthy of further consideration.

The principal strategy used by the atheists in questioning the creator's potential to create was to point out that those who spoke of a single creator, usually called the Supreme Lord (*parameśvara*) or some similar term, usually regarded the Lord as eternal. That is, the Lord was supposed never to have come into being (either from nothingness or from some other state of being) and was also supposed not to be liable to stop being. It was also generally accepted that nothing could be eternal unless it was also simple. That is, all eternal things were supposed to consist of exactly one part that was quite uniform in nature, so that the whole was not in any sense different from its unique part. Being indivisible, simple beings were generally supposed to be either omnipresent or infinitesimal. Those who posited the existence of a Supreme Lord usually argued that he uniformly pervades all of space at all times.

A question that naturally arises about an eternally simple ubiquitous being is how such a being is supposed to act. The way this question was pursued is illustrated by how it was approached by the Buddhist philosophers Dharmakīrti (seventh century) and Śāntarakṣita (eighth century). They argued that action, as normally understood, involves undergoing some change of state. This change of state must be either a change in how the being in question is related to things outside itself, or a change in how parts within the being are related. The first kind of change, consisting of a change in relation to external things, normally involves some kind of motion, that is, a change in relative location over time. To a being that is everywhere all the time, however, motion is clearly impossible. This leaves internal change as the only remaining kind of change that a simple being might undergo. In the case of complex objects, internal changes of state are easily accounted for by referring to the gain or loss of parts, such as attributes, or to relational changes among the parts. It is, however, difficult to imagine just what kind of change of state is possible for a being that has only one part. And therefore it would appear that a simple being is not constituted in such a way as to have the potential for any kind of self-motivated action.

A second argument that atheists used in arguing against a supreme creator's potential to act in any way at all

involved asking the question of where the supposed potential might reside. There appear to be two possibilities. It could be that the potential is contained entirely within the being that has it, or it could be that the potential lies outside the being. If it were argued that the potential to act is entirely resident within a being, then it could be asked what it is that triggers this potential into actuality. If the potential were self-activating, then it would have to be activating itself at all times, for otherwise it would require something outside itself to make it active at some times but inactive at others. But a potential that is perpetually active is not a potential at all; rather, it is an actuality. Even if one were to hold the view that the Supreme Lord is perpetually creating the universe, this would not explain why he creates the world slightly differently from one moment to the next, thereby giving the appearance of change to his creation. It is better, argued the atheists, to abandon the supposition that the potential to create is contained entirely with the Supreme Lord. This leaves the possibility that the Lord's potential to perform the action of creation comes from outside, just as an inanimate body's potential to be in motion comes from outside itself. In that case, argued the atheists, we should not be able to say that it is God who performs the action of creating the universe, but rather that it is that which is external to God that is performing the creative action, and that this external agent is merely using God as an instrument in the action being performed.

2 A complex being's potential to decompose

Whereas a simple being is not composed of parts and therefore lacks the potential to go out of existence through the process of decomposition, a complex or composite being is liable to decompose. It was agreed among classical Indian philosophers that complex objects are subject to destruction through decomposition. Where there was controversy was over the issue of how this decomposition comes about. In particular, there was controversy over whether the destruction of a complex thing requires a separate external cause to bring it about or whether complex things have an innate tendency to go out of existence spontaneously.

The position that complex things have an innate tendency to cease to exist was held by several Buddhist scholastics, such as Vasubandhu (fourth or fifth century), Dharmakīrti, and Śāntarakṣita. Taking as their point of departure the Buddha's observation that all composite things are impermanent, the Buddhist scholastics grew increasingly committed to the view that all composite things are momentary, that is, that all complex beings perish in the very moment in which they arise (see Momentariness, Buddhist doctrine of).

Among the arguments that were used to defend this doctrine, one was based on an analysis of potentiality. Vasubandhu argued that if a given complex being were capable of remaining unchanged for two consecutive moments, then there would be no reason for it ever to undergo change. Suppose that a being had a certain set of properties in one moment and then had exactly the same set of properties in the subsequent moment. One might ask whether among those properties the potential to perish was to be found. If not, then the complex being would have no potential to perish and therefore would be eternal. But even if the potential to perish were among the properties of the complex being, the question would still remain of how that potential is actualized. If the potential is self-activating, then it must activate itself in each moment of its existence; but if the potential to perish is realized in each moment of a thing's existence, it follows that the thing exists for no more than one moment.

This conclusion that things perish in the very moment in which they arise is quite contrary to how things are experienced. In our experience of things, it seems that complex objects endure for a period of time and then go out of existence as a result of being acted upon by some destructive agent. A piece of wood, for example, continues to exist unchanged until it comes into contact with fire, which is the cause of the wood's destruction. One might ask why we should not take our experiences of these events at face value. In addressing this question, Vasubandhu points out that in fact we do not directly experience the destruction of the wood by the fire. What really happens is that we see the wood, then we see the fire, and then we no longer see the wood. It may be tempting to explain this sequence of events by saying that the presence of the fire causes the disappearance of the wood, but it would be a mistake to claim that one actually observes the causal process of destruction at work. Indeed, argues Vasubandhu, it is impossible to observe fire destroying a piece of wood; if the fire and the wood are observed to exist at the same time, then it is obvious that the existence of the fire is not a cause of the absence of the wood, for the simple reason that the wood is not absent. On the other hand, if the wood is present in one moment and absent in the next, fire existing where the wood used to be, then one might say that the wood was replaced by flame. Seeing one thing replaced by another, however, is not the same as seeing the subsequent thing destroy the former thing. After considering various other dimensions of this problem, Vasubandhu finally concludes that the problem of what

causes the absence of a thing is one that arises only when one assumes that absence itself is an entity and therefore must have a cause. If, however, it is recognized that absence is not itself a thing, then it is clear that absence cannot have a cause, for only things have causes. Once it is recognized that there is no such thing as the cause of an absence, the question of how a potential cause of absence is actualized ceases to have any meaning.

3 A person's potential for cultivating virtue

The question of how a complex thing was supposed to become absent was not merely an issue for intellectual debate. An issue underlying this discussion was the question of whether certain vices, seen as responsible for keeping a person bound in the cycle of rebirth, required a special cause, such as human effort, to eliminate them. A representative account of the mechanism of karma and rebirth is that of Vasubandhu. According to his *Abhidharmakośa (Treasury of Abhidharma)*, a decision to act in a particular way has several consequences, one of which is a predisposition to act in a similar way in the future. If, for example, one acts with the intention of injuring another, then making the decision to injure others will be that much easier to make in the future. An action, in other words, creates a latent predisposition, which in turn is a factor in creating a future action. A being who is caught in the round of rebirths is captured in that cycle as a result of various kinds of latent predisposition, the most important being the predisposition to pursue sensual pleasures, to dislike sources of displeasure, to have a tenacious belief in a permanent self, and to hold on to dogmas that legitimate one's choices of action. These predispositions are all regarded as potentials. Therefore, the question arises as to how their potentiality is activated. Moreover, each of these potentials can be seen as a complex thing that therefore has the potential to decompose. Here the question is whether the potential to degenerate requires an external cause. Does it take work to rid oneself of a bad habit?

Vasubandhu's response to this question is that it does not require special effort to rid oneself of a habit; rather, it takes work to keep the habit intact. It takes effort to keep oneself habituated to seeking pleasures and holding onto dogmas. What takes no effort is to let go of these habits. If one stops making the effort necessary to reinforce a bad habit, it will eventually dissipate through its own potential to perish. For this reason, most of the specific precepts of the spiritual life are negative in character: avoid killing, avoid theft, avoid sexuality, avoid lying and avoid intoxicants. Moreover, the most effective forms of meditation consist in simply noting what mental states are present and making no effort either to pursue them or avoid them. If simply noted but not acted upon, one's urge to kill or to make a scathing remark will perish by itself without creating a karma. By not creating a karma that reinforces the habit to act malevolently, one eventually becomes free of the malevolent habit and of the views necessary to justify acting on those habits.

Several schools of Hindu religious philosophy, and one strand of the Yogācāra school of Buddhism, asserted the view that every sentient being possesses a potential for wisdom. This wisdom was seen as something positive and not merely as the absence of vices and misconceptions, and therefore the potential to make one wise must also be seen as positive. As in other discussions of potentiality, a question that arose was how this potential for wisdom can be activated. One of the answers was that the potential is in fact activated at every moment; in other words, every sentient being is indeed wise in its truest nature. The apparent lack of wisdom in some beings, or lapses of wisdom in beings who are normally wise, can then be accounted for not as an actual lack of wisdom but as an occultation of actual wisdom. In other words, a being's innately radiant, blissful and virtuous nature is concealed by the darkness of a transitory episode of ignorance or vice. This position, of course, gives rise to the question of how the ignorance or vice comes into being and to whom it belongs if not to the innately wise mind itself. Śaōkara's famous answer to this riddle was to reply that the question arises only from the perspective of ignorance, and since ignorance is ultimately illusory, there is not really a riddle here in need of an answer (see Śaōkara).

4 The potential of a sound to convey a meaning

From the earliest times, Indian theorists of language were intrigued with the potential or power that language has to evoke images and emotional responses in the mind of the hearer. It was generally recognized, however, that this potential cannot reside entirely within the sounds themselves, since a person unfamiliar with a language will not have appropriate intellectual and emotional responses to the sounds heard. Grammarians and poeticians both observed that a verbal performance, such as a poem, only had its effect when it was heard by a cultivated intellect; generally speaking, the more learned and sophisticated the intellect of the hearer, the more richness the poem yielded. This was said to be true of all language, but was especially important for the religious poetry of the Vedas.

A phrase of revealed Vedic poetry spoken in the context of a ritual performance was known as a *mantra*, and *mantras* were supposed to have a power not only to convey meanings, but to influence physical events. The power of a *mantra* was said to be at its maximum when pronounced by a priest who had a perfect understanding of the Vedic Sanskrit language. This fact was enough to make Buddhist and Jaina philosophers question whether the power of language was a feature of the sounds themselves, as Brahmanical thinkers claimed. Buddhists and Jainas were also dubious about the claim that Sanskrit had a power greater than any other language to evoke meanings and emotions, arguing instead that whatever language the hearer knows most intimately is most meaningful to the hearer; this would suggest that the potential resides not in the sounds but in the community of people who know a language.

One further controversy debated by Indian theorists of language had to do with how the potential to convey meaning was located in a sentence, and what role the individual syllables played. Some argued that each syllable has its own independent potential and that the sentence's potential to convey meaning is the sum of the potentials of the syllables. Others argued that the sentence as a whole has a potential, and the function of the syllables is merely to make that potential fully realized for the hearer. According to this view, even if only the first syllable of a sentence is spoken, the sentence as a whole still has its full meaning-bearing potential, but the hearer receives only a partial glimpse into what that meaning might be.

See also: Bhartṛhari; Buddhist philosophy, Indian; God, Indian conceptions of; Karma and rebirth, Indian conceptions of; Language, Indian theories of; Patañjali

RICHARD P. HAYES

References and further reading

Ingalls, D.H.H. (1953) 'Śaṃkara on the Question: Whose is *Avidyā*?', *Philosophy East and West* 3: 69-72.(A brief but intelligent exploration of the problem, mentioned in §3, of how an innately wise and blissful Brahman can be obscured by ignorance and to whom, given that Brahman is the essence of everything, this ignorance might belong.)

Jackson, R. (1993) *Is Enlightenment Possible?*, Ithaca, NY: Snow Lion.(Contains a discussion and translation of Dharmakīrti's arguments, mentioned in §1, concerning the possibility of divine action in general and God's potential to create the world in particular.)

Śāntarakṣita (eighth century) *Tattvasaōgraha (Grasping the Truths)*, trans. G. Jha, *The Tattvasaōgraha of Shāntarakṣita with the commentary of Kamalashīla*, Delhi: Motilal Banarsidass, 1st edn, 1939; repr. 1986, 2 vols.(An extensive collection of arguments produced by Buddhist thinkers on topics debated between them and non-Buddhist opponents. Volume 1, chapters 1-7 contains Buddhist arguments against God, Brahman and other notions of single, eternal creators. Chapter 8 deals with the questions of perishability discussed in §2. Volume 2, chapter 29 deals with the question of the potential of a word to convey meaning, discussed in §4.)

Vasubandhu (4th or 5th century) *Abhidharmakośa (Treasury of Abhidharma)*, trans. L. de la Vallée Poussin, *L'Abhidharmakośa de Vasubandhu*, Brussels: Institut belge des hautes études chinoises, 1971 (reprint), 6 vols. (Best translation available of this Buddhist classic. Volume 3 contains chapter 4 of the original work, which deals with karma. Here are contained the arguments about the causes of a thing's disappearance discussed in §2. The discussion of virtue in §3 is based on chapters 4 and 5 of the original text, found respectively in volumes 3 and 4 of the translation.)

Pothier, Robert Joseph (1699-1772)

Robert Joseph Pothier was one of the most influential of modern civilian jurists. At the end of a long period of rationalistic natural law thought, he produced a rational reconstruction of the civil law on the basis of which the transition from law as reason based on custom to law as product of the rational legislative will could be accomplished.

By his work in the systematization first of Roman law and subsequently of French customary law, Pothier cleared the way for the codification of French Law under Napoleon in the early nineteenth century. Born in Orléans of a family with a long tradition of service in the magistracy, he served most of his life as a magistrate in Orléans, becoming also a professor in the University of Orléans after publication of his first main work, which was a reconstitution of Justinian's *Digest* aimed at clarifying the underlying conceptual framework (see Justinian). This was the result of long and profound studies in Roman law (see Roman law §2). In each title he rearranged the materials, without further textual alteration, to reveal what he conceived to be the logical order underlying them.

Following upon this, he turned his attention to French law, which at that time was a patchwork, in the north based on various differing local customs according to different court districts, in the south on a partly modernized version of Justinianic civil law. He published a text on the customary law of Orléans, pointing out its variations from and similarities with other customs of different court jurisdictions in the north. Applying his rationalistic principles, Pothier then turned to a statement of the principles of law that in his view underlay both written and customary law in crucially important fields. His most notable contribution was his *Treatise on Obligations*, expounding general principles of contract-formation and contractual liability. This quickly achieved recognition inside and outside France, being translated into English in an American and a UK version in the early 1800s. It was followed by a series of works on more particular contracts, and then by work on matrimonial property and on marriage, interpreted as essentially a civil contractual obligation to which Church authorities also gave recognition, rather than as an essentially ecclesiastical or religiously-grounded institution.

In France, Pothier's influence was enormous, in that he provided a basis on which the compilers of a civil code could work when, in the circumstances of revolution, the idea of the rule of law was conceptualized as requiring abandonment of obscure customs expounded by judges, instead calling for their replacement with rationally ordered and duly published statutes (see Rule of law (Rechtsstaat) §1). In the common law world, then also engaged in a prodigious effort of rationalization, his work was scarcely less influential, as the common law struggled to produce out of materials based on pre-modern forms of action a rational law of contract fit for an industrializing and commercializing society. Thus did natural law ideas, such as Pothier's, create a rational structure of legal concepts that more positivistic successors could adopt.

See also: Law, philosophy of; Legal positivism; Natural law

NEIL MacCORMICK

List of works

Pothier, R.J. (1806) *A Treatise on the Law of Obligations or Contracts*, trans. W.D. Evans, London: Butterworth. (A well-known translation of the part of Pothier's work that was most influential in the English-speaking world, particularly in relation to the general theory of contracts.)

Pothier, R.J. (1861) *Œuvres de Pothier*, ed. M. Bugnet, Paris: Cosse et Marchal, Henri Plon. (A complete set of Pothier's main writings annotated and correlated to the Civil Code and to current (1861) French legislation. A useful indicator of his influence on the Napoleonic Code.)

Pothier, R.J. (1953) *A Treatise on the Contract of Letting and Hiring*, trans. G.A. Mulligan, Durban: Butterworths.(This is a more specialist work than (1806) above, but again gives insight into Pothier's theory of contracts.)

References and further reading

Smith, J.C. and Thomas, J.A.C. (1957) 'Pothier and the Three Dots', *Modern Law Review* 20: 38-49.(An essay on a problem of contemporary English law that reveals the historical influence of Pothier on the development of common law approaches to contract.)

Pound, Roscoe (1870-1964)

Roscoe Pound was a legal thinker who exercised profound influence in the USA in the first half of the twentieth century. The approach he advocated, under the banner 'sociological jurisprudence', was to present law as essentially an instrument for the pursuit and reconciliation of human interests.

Long-term Dean of the leading law school, Harvard, Pound played a prominent role in defining the dominant approach to understanding and teaching law for twentieth-century Americans. To understand any body of law, he argued, one must understand what interests it serves, and must be careful to discriminate different types of interest. Some interests are individual, like the interest of each person in their own life and health; some are social, like the interests of a society in controlling the spread of infectious diseases; others are public, like the interest of the state in the health and fitness of citizens eligible for military service. The tasks of law-making and of judging are tasks of balancing and harmonizing actually or potentially conflicting interests, with a view to maximizing interest-satisfaction at the lowest possible cost in frustration of other interests. Hence law can be considered 'a continually more efficacious social engineering'.

Pound's is thus an optimistic as well as an instrumentalist view of law. His upbringing and education in Nebraska, his early training as a botanist and his work on the flora of Nebraska, followed by experience as legal practitioner and pro tem judge in his home state, no doubt contributed to the progressivism exhibited in the sociological jurisprudence he developed as a professor called back east to the Harvard Law School. His scientific training is doubtless also reflected in the care he took over working out a complete taxonomy of the interests he could detect as being served by the law. Significant here also is his typology of 'individual', 'social' and 'public' interests, and his attempt to map the connections and differences between them. Also important is his insistence that in a process of balancing, like must be compared with like, and interests should be brought within the same type before balancing can be properly carried out. Recognition of some interests as irreducibly social or irreducibly public marks off his theory from classical utilitarianism, to which he has obvious affinities. In trying to develop a distinctive 'jurisprudence of interests', he had predecessors such as the German jurist Rudolf von Jhering.

Pound engaged in polemics with 'American realism' (see Legal realism §2), arguing for what he saw as a rational approach to legal studies, rather than exaggerating law's indeterminacies, as he thought the realists did. His own theory suffered at the hands of sociologists and others who challenged his belief in the essential harmony of interests despite class and other divisions in society. His vast five-volume statement of his jurisprudence is so ponderous as to be nowadays little read; his shorter, more introductory, writings retain far greater interest and vitality. He is a significant figure among 'legal instrumentalists' in US intellectual history, but has little direct influence at present.

See also: Law, philosophy of; Legal idealism; Public interest §5; Social theory and law

NEIL MacCORMICK

List of works

Pound, R. (1922) *An Introduction to the Philosophy of Law*, New Haven, CT: Yale University Press.(Well-written statement of Pound's basic position, containing discussions of, for example, processes of legal argumentation, that retain contemporary relevance and remain much-quoted.)

Pound, R. (1942) *Social Control through Law*, New Haven, CT: Yale University Press. (Mature statement of Pound's conception of law as a kind of 'social engineering', controlling society with a view to maximal satisfaction of interests, acknowledging also the role of morals, religion and education.)

Pound, R. (1959) *Jurisprudence*, St Paul, MN: West Publishing Co., 5 vols.(Compendious account of the whole range of Pound's views on jurisprudential issues, including his exhaustive survey and taxonomy of interests; lacking the vitality of his earlier work.)

References and further reading

Setaro, F.C. (1942) *A Bibliography of the Writings of Roscoe Pound*, Cambridge, MA: Harvard University Press. (This is a comprehensive bibliography up to 1942, but Pound lived long and remained productive, hence Straight's bibliography was necessary to complete the job.)

Straight, G.A. (1960) *A Bibliography of the Writings of Roscoe Pound: 1940-1960*, Cambridge, MA: Havard Law School Library. (This, together with the Setaro bibliography, lists the whole vast corpus, of which only the more philosophically interesting works are noted above.)

Summers, R.S. (1982) *Instrumentalism and American Legal Theory*, Ithaca, NY and London: Cornell University Press.(Outstanding account by a leading US jurist of the various strands in the school of thought to which Pound belonged, providing a clear view of Pound’s work in its context, and of his place in US thought.)

Power

The general notion of power involves the capacity to produce or prevent change. In social and political philosophy, narrower conceptions of power specify the nature of these changes. Social power is the capacity to affect the interests of agents. Normative power is the capacity to affect their normative relations, such as their rights or duties.

There are long-standing conceptual disputes about the nature of power. Some emphasize the role of actual or potential conflicts of interest in defining power relations, others take legitimacy and consensus about norms as the basis. There is also a dispute, involving the nature of social explanation, about whether power must rest with agents or with social structures or forces of some kind. The lack of consensus on these matters raises the overarching question of whether the concept of power, which seems to function as a descriptive term, in some way involves evaluation.

1 Social power and interests

In social and political theory, ’power’ is sometimes used wherever causal language is appropriate. We may thus speak of the power of incentives in the labour market or the power of socialization in child development. This may prompt the thought that power is everywhere and beyond escape. (Michel Foucault sometimes makes such promiscuous use of the term; see, for example, Foucault 1980.) But to identify social power with causal power is to abandon a helpful concept, for there are theoretical and practical interests in distinguishing among the causal processes in social life.

Many views delimit social power by reference to a significant capacity to affect people’s interests. (The question whether such power is better thought of as a resource or a relationship is somewhat artificial: it is both.) Hobbes (1651) said that power is one’s ’present means to obtain some apparent future good’, and thought the endless struggle for power an ineliminable part of human nature. The implicit reference to social conflict is made explicit in Max Weber’s definition of power as: ’the probability that one actor within a social relationship will be in a position to carry out his own will despite resistance, regardless of the basis on which this probability rests’ ([1922] 1968: 53).

What role do conflicts of interest have in power ascriptions? Steven Lukes (1974) has vigorously argued that since the most insidious and effective forms of power are those that affect people’s conceptions of their interests, power is more than the capacity to influence overt decision making, and more than the capacity to control which issues reach the public agenda; it also includes the capacity to influence how people perceive their own interests. Thus, power may be present even in the face of apparent consensus. To identify such power, however, requires a view of interests that is objective in the sense that it is independent of the actual beliefs and desires of the subjects (see Needs and interests).

Critics have wondered whether such a notion is very useful in descriptive and explanatory research (see §4). Perhaps a deeper worry, however, is whether ’objective interests’ can be defined independently of power, for it seems natural to say that objective interests are those interests people would have if they were not constrained by power relations (Kernohan 1989). Moreover, both Weber’s and Lukes’ theories seem guided by the idea that power affects people to their detriment. Does this mean that justified acts of paternalistic control cannot be exercises of power? This suggests that although social power involves the capacity to affect the interests of people, it is to be identified less by its actual or intended ends than by its means, such as coercion or manipulation. While we may allow that such means are *prima facie* detrimental to people’s interests, they need not be so when all things are considered.

2 Normative powers

Normative powers (for example, the power to promise or to command) involve the capacity to change normative relations: to create, extinguish or vary people’s rights and duties. The relevant relations may be a matter of law, morality or social convention.

Some maintain that all normative powers may be reduced to powers to impose or remove duties which may in turn

be reduced to direct or indirect threats of force. Each step of this reduction may be challenged - especially the second, which cannot account for the fact that people often recognize duties that they think neither will be nor should be enforced. Reductivist accounts, however, continue to be influential among sociologists who incline to a ‘realistic’ view of power.

The concept of normative power is important in legal philosophy, but some have also adopted it as the paradigm for political power in general. Hannah Arendt’s quirky definition makes sense in this light: ’Power corresponds to the human ability not just to act but to act in concert. Power is never the property of an individual; it belongs to a group and remains in existence only so long as the group keeps together’ (1970: 143). Her idea that power rests on accepted norms that constitute a group and that precede all instrumental reasoning about what the group should do is inspired by a particular account of what it is to be ’in power’ (see Arendt, H.). Similar thoughts underpin Talcott Parson’s idea that power is a ‘generalized medium of mobilizing commitments or obligation for effective collective action,’ a ’generalized facility or resource’ that rests on accepted norms of legitimacy (1967: 331). Focusing in these ways on norm-constituted capacities emphasizes consensus as opposed to conflict, which for Arendt is instead a matter of ’strength’ or ’violence’ and for Parsons ’force’ or ’coercion’. It also assimilates political power to authority, which is one of the most important normative powers (see Authority).

Since normative power affects people’s interests, it does fall under the general concept of social power. It is worth distinguishing, though, because of the way beliefs about legitimacy and norms mediate its effects. This partly explains why these theories make power relations seem less asymmetric than do Hobbesian or Weberian theories. To suppose that A has social power over B in respect of X often suggests that B does not have that same power over A in that same respect. But whether A and B mutually have the power to rule depends on the actual norms accepted in that society: these powers may be fully symmetrical.

3 Problems of agency

If power is a capacity to produce or prevent changes, in what does that capacity rest? Many assume that it must rest in individual or corporate agents. This is obvious in the case of normative power; but it is also implicit in the accounts of social power examined in §2, and explicit in Bertrand Russell’s (1938) definition of power as ’the production of intended effects’. Even if one may exercise power unintentionally (for example, recklessly), only things with the capacity for agency can exercise power at all.

Others hold that social power rests in non-agential things called social ’structures’ or ‘forces’. For the Marxist philosopher, Nicos Poulantzas, individual agents are but the locations or ’bearers’ of such forces, and power is ’the capacity of a social class to realize its specific objective interests’ (1973: 104). This view has been criticized by Ralph Miliband (1969, 1970), who holds that political agents retain at least a relative autonomy from class structures. At bottom is a deep dispute about the cogency of determinist explanations for social change.

Michel Foucault’s later writings suggest another non-agential view of power (Foucault 1978, 1980). First, he expands the relevant power structures to include realms of professional and scientific discourse. At its most extreme, this allows no distinction between claims to knowledge and bids for power. Second, such power creates agents: it is ’productive’ and not merely ’repressive’. On this view, naming, classifying or diagnosing people, literally creates criminals, the insane, homosexuals, and so on (see Foucault, M.).

It may be wondered whether this view is coherent. Whether or not the theory abstains from using the word ’knowledge’ or claiming truth for itself, it none the less purports to have some sort of warrant which, on its own terms, must be understood as a bid for power. This is pragmatically self-defeating. Second, the productive thesis elides a distinction between the literal production of agents and the production of a sense of identity on the part of pre-existing agents. (So while, for example, homosexuals may not have been created in the nineteenth century, an emergent gay identity may have been.) But the general notion that social power may affect not only one’s conception of one’s interests, as Lukes says, but also one’s conception of oneself as an agent capable of having interests, is an important deepening of that line of thought, and one independent of the more radical Nietzschean themes in Foucault’s work.

4 Power and values

Power seems to function as a descriptive term. In certain limited contexts (for example, in bargaining theory or

voting theory) it has proved possible to specify measures of power, and in the social sciences there are distinguished empirical studies of the bases of power. But in spite of this conceptual controversy persists, and not only over borderline cases, but also over what should be taken as paradigm cases of power. This may suggest that power is an ‘essentially contested concept’, disputes about which are fundamentally moral or political and which preclude resolution (see Political philosophy, nature of §3).

There are a number of sources of controversy here, and they need to be distinguished. First, referring as they do to causal capacities, power ascriptions inherit all the philosophical difficulties of causal language and, in particular, the difficulties of interpreting and evaluating counterfactual conditionals (see Counterfactuals; Causation). Second, as we have seen, accounts of social power also inherit disputes about the nature of human agency and interests (see §2). Third, power ascriptions are shaped by the practical point or purpose in making them. One influential view, advanced by Lukes (1974) and by William E. Connolly (1974), is that power is ascribed to fix moral responsibility: those with power are those who should change things. This raises further questions: does responsibility require intentional action, or are recklessness or even just negligence enough?

This thesis must, however, be framed with care. First, even if there is a link between power and responsibility, it does not follow that to say ’A has power over B’ is to make a moral judgment about A’s blameworthiness, for A may have a justification or excuse for acting or failing to act. Second, lacking the power to change something may itself be an excusing factor, on the principle that ought implies can. If that is so, then it must be possible to make some power ascriptions independently of the relevant judgments about responsibility.

Moral values are not, in any case, the only way in which we might delimit the field. Power ascriptions are part of a larger social theory, so there are also theoretical values such as simplicity, fecundity, coherence, and so on, to be considered. Moreover, although views about the nature and importance of certain human interests also guide concept formation, this need not involve any moral appraisal or prescription. To say that money traders have great social power is to think certain effects more important than others because of their links to central human interests; but it is not directly to assign moral praise or blame. We should not confuse the plausible claim that all descriptions are value-laden, with the stronger and dubious claim that all descriptions are morally fraught.

LESLIE GREEN

References and further reading

Arendt, H. (1970) ‘On Violence’, in *Crises of the Republic*, New York: Harcourt Brace.(Influenced by classical republican themes, this essay takes normative powers as the paradigm. See §2.)

Bachrach, P. and Baratz, M.S. (1962) ‘The Two Faces of Power’, *American Political Science Review* 56 (4): 947-52.(An important discussion of power as constituted by agenda-setting or background bias as well as by overt conflict.)

Connolly, W.E. (1974) *The Terms of Political Discourse*, Lexington, MA.: D.C. Heath.(In particular, see pages 10-44 on essentially contested concepts, and pages 86-137 on power and responsibility. See §4.)

Foucault, M. (1978) *The History of Sexuality*, vol. 1, *An Introduction*, trans. R. Hurley, New York: Vintage Books.(Pages 3-49 are particularly good on the ’repressive hypothesis’. See §3.)

Foucault, M. (1980) *Power/Knowledge: Selected Interviews and Other Writings, 1972-77*, ed. C. Gordon, New York: Pantheon Books.(Useful sampler. Power as ubiquitous in social life, with particular attention to its non-political and non-economic dimensions. See §§1, 3.)

Goldman, A. (1972) ’Toward a Theory of Social Power’, *Philosophical Studies* 23: 221-68.(A detailed analysis of both individual and collective power in terms of getting what one wants.)

Hobbes, T. (1651) *Leviathan*, ed. C.B. Macpherson, Harmondsworth: Penguin, 1968.(A conflict-based view of power based on an individualistic and egoistic moral psychology. See §1.)

Kernohan, A. (1989) ’Social Power and Human Agency’, *Journal of Philosophy* 86 (12): 712-25.(Argues that social power cannot be defined in terms of intentions, effects on interests or responsibility.)

Lukes, S. (1974) *Power: A Radical View*, London: Macmillan.(One of the most influential, and controversial, accounts. See §§1, 4.)

Miliband, R. (1969) *The State in Capitalist Society*, London: Weidenfeld & Nicolson.(A neo-Marxist study of forms of power in modern capitalism. See §3.)

Miliband, R. (1970) ’The Capitalist State: Reply to Nicos Poulantzas’, *New Left Review* 59 (January-February): 53-60.(A polemical reply to Poulantzas’ structural determinism. See §3.)

Morris, P. (1987) *Power: A Philosophical Analysis*, Manchester: Manchester University Press.(A good survey of

the philosophical issues.)
Parsons, T. (1967) *Sociological Theory and Modern Society*, New York: Free Press.(Essays defending and elaborating a consensus-based view of power. See §2.)
Poulantzas, N. (1973) *Political Power and Social Classes*, trans. T. O'Hagan, London: New Left Books.(A structuralist-Marxist theory of political power in capitalist states. See §3.)
Russell, B. (1938) *Power: A New Social Analysis*, London: Allen & Unwin, ch. 1.(Power as the key to social explanation. Surveys its forms, bases and the means of controlling it. See §3.)
Taylor, C. (1986) 'Foucault on Freedom and Truth', in D. Couzens Hoy (ed.) *Foucault: A Critical Reader*, Oxford: Blackwell, 69-102.(An important critical discussion, in a useful collection.)
Weber, M. (1922) *Economy and Society*, vol. 1., ed. G. Roth and C. Wittich, New York: Bedminster Press, 1968.(A major treatise on social and political economy, touching on a conflict-based view of power, related notions of authority and legitimacy, and important methodological issues. See §1.)
Wrong, D.H. (1979) *Power: Its Forms, Bases and Uses*, New York: Harper & Row.(An excellent survey from a sociological point of view.)

Practical reason and ethics

Practical reason is reasoning which is used to guide action, and is contrasted with theoretical reason, which is used to guide thinking. Sometimes 'practical reason' refers to any way of working out what to do; more usually it refers to proper or authoritative, hence reasoned, *ways of working out what to do.*

On many accounts practical reasoning is solely instrumental: it identifies ways of reaching certain results or ends, but has nothing to say about which ends should be pursued or which types of action are good or bad, obligatory or forbidden. Instrumental reasoning is important not only for ethics and politics, but for all activities, for example, in working out how to travel to a given destination.

Other accounts of practical reason insist that it is more than instrumental reasoning: it is concerned not only with working out how to achieve given ends, but with identifying the ethically important ends of human activity, or the ethically important norms or principles for human lives, and provides the basis for all ethical judgment.

No account of objective ethical values can be established without showing how we can come to know them, that is, without showing that some form of ethical cognitivism *is true. However, ethical cognitivism is not easy to establish. Either we must show that some sort of intuition or perception provides direct access to a realm of values; or we must show that practical reasoning provides less direct methods by which objective ethical claims can be established. So anybody who thinks that there are directly objective values, but doubts whether we can intuit them directly, must view a plausible account of practical reason as fundamental to philosophical ethics.*

1 Introduction

Most ethical positions or theories rely on one or more conceptions of practical reason, yet many fail to explicate, let alone to vindicate, the particular conceptions on which they rely. The only positions or theories which offer no account of practical reason are those which construe ethical claims either as noncognitive or as based directly on particular cognitions such as perceptions or intuitions that do not need to be linked to any reasoning process (see Analytic ethics; Emotivism; Intuitionism in ethics; Moral realism).

One pervasive disagreement among the proponents of various accounts of practical reason is between those who think that practical reasons should both justify and motivate action and those who think that they should justify, but need not motivate. The claim that reasons both justify and motivate is often labelled *internalism*, the thought being that anything that motivates (whether a desire, or a certain sort of belief, or some other internal state) must be internal to the agent who is motivated. The claim that reasons must justify but need not motivate is correspondingly termed *externalism* (see Moral motivation §§1-2). Broadly speaking, internalists think that externalists fail to show how reason can be practical, since they do not adequately address the question of motivation, and externalists think that internalists lose sight of the fact that practical reason must be reasoned, since they build an account of the contingencies of motivation into their account of reason. There are many versions both of internalism and of externalism.

Accounts of practical reasoning can be grouped under two very general headings. Many accounts of practical reason are *end-oriented* (also known as *teleological* or *consequentialist*): they seek to show how reason can select action (also attitudes and policies) that will contribute to certain ends or results (see Consequentialism; Teleological ethics). Many proponents of end-oriented practical reasoning think that practical reason can do no more. Since no objective account of the proper ends of human life can be found, we must settle for *subjective* accounts of human ends. We must agree with Hume that reason 'is, and ought only to be the slave of the passions' (Hume 1739-40: 415) (see Rationality, practical §1). However, there are other accounts of end-oriented practical reasoning which insist that reason can also identify certain *objective* ends, and so has the dual task of identifying the proper ends of action and guiding action towards those ends.

Yet other conceptions of practical reason do not view it as focused solely on ends and means to those ends. *Act-oriented* accounts of practical reason take it that action (also attitudes and policies) are guided by norms or principles, and that it is the task of practical reason to identify appropriate, reasoned norms and principles. Some norms and principles formulate very general ends, others point to quite specific acts: the Jesuit maxim 'For the greater glory of God' defines the end of an entire life; the homely advice 'Do not eat oysters unless there is a letter

"r" in the name of the month' offers limited and specific advice. Although some norms and principles define ends, norm- and principle-guided reasoning cannot be assimilated to end-oriented reasoning, which sees all practical reasoning as bearing on a means-ends complex. Act-oriented reasoning sees it as bearing on practical propositions (norms, rules, principles) and as guiding action without reference either to supposed objective ends or to the subjective ends of any agents.

Many accounts of act-oriented reasoning see it as norm-based: they take it that the fundamental ethical orientations (norms, categories, commitments, beliefs, senses of identity: hereafter *norms*) accepted in a society or tradition, or in an individual's life, provide the fundamental premises for practical reasoning. On such accounts, practical reasoning is internal to societies: it appeals to the norms that constitute the bedrock of certain lives or ways of life, so cannot coherently be brought into question by those who live these lives.

Other critical accounts of act-oriented ethical reasoning object that uncritical appeal to accepted norms cannot provide acceptable premises for practical reasoning; arguing from accepted norms is quite arbitrary, hence quite unreasoned. Any conclusions reached in this way will be relative to the assumed norms, so can support no more than one or another version of ethical relativism (see Moral relativism). Exponents of critical accounts of practical reason think that adequate act-oriented reasoning must offer reasons which do not presuppose any specific norms (or traditions, or identities), but rather must be public in the sense that they are relevant to an unrestricted audience, or, as Kant put it vividly in an early version of this thought, that they must 'address… the world at large' (1784: 57). The possibility of vindicating any critical, universal conception of act-oriented practical reason is questioned by proponents of other conceptions.

The search for a convincing account of practical reason raises issues fundamental not only to ethics and politics, but to a wide range of philosophical problems. They include the problem of seeing whether, and if so how, any account of practical reason can be vindicated without begging questions, the connection between practical and theoretical reason and the problem of showing how and how far various conceptions of practical reason can guide action, attitudes or policies.

2 End-oriented reasoning: reason is instrumental

Scepticism about the idea that reason has any objective ends of its own was put succinctly by Hume, who scoffed that ''tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger' (1739-40: 416), insisted that ''tis in vain to pretend, that morality is discover'd only by a deduction of reason' (457) and concluded that *ought* cannot be derived from *is* (469). If morality has no ends that are discovered by reason, then practical reason's only task is to show how the pursuit of the passions - of subjective ends - is to be organized effectively and efficiently. The central task of practical reason is the instrumental one of deploying our knowledge of causal relations to guide action: practical reason is simply an application of and derivative from aspects of theoretical reason; it needs no separate justification.

There is, however, a great deal to be said about the detailed operations of instrumental reasoning. In particular, if some metric can be found for the ends which are to be sought (for example, if there is a metric for the satisfaction of desires or preferences), and if the probabilities of achieving various ends can be quantified, then instrumental reasoning can be used to show which available actions maximize the satisfaction of preferences, and hence how given ends can be pursued efficiently. If this metric permits the desires or preferences of different persons to be measured using a common unit, practical reasoning may be able to guide social as well as individual decision-making and policy. Utilitarian ethical theory, game theory and a wide range of economic and social calculi have discussed numerous versions of subjective, end-oriented practical reasoning which are held to be appropriate for various contexts (see Rationality, practical; Rational choice theory; Utilitarianism.)

Merely instrumental accounts of practical reason have been criticized at least since Kant. The critics do not doubt that instrumental reasoning is a necessary aspect of practical reasoning, but they insist that it cannot be sufficient. All that it provides is an account of the use of empirical, and in particular causal, knowledge in pursuit of intrinsically arbitrary ends. Instrumental reasoners may show that it is necessary to break eggs if one wants to make an omelette, but they have nothing but preferences to cite as reasons for making or not making omelettes: all their conclusions are conditional on subjective ends. Instrumental reasoning in pursuit of individual preferences can support efficient egoism, or strategic thinking or certain limited and distinctively modern conceptions of

prudence: but this is all it can do (see Moral motivation §§4-6; Prudence §2).

Contemporary work on instrumental reasoning has addressed this criticism up to a point by insisting that practical reasoning requires not only that choices of action be instrumentally sound, but that they be based on well-ordered (for example, connected and transitive) preference-orderings. However, since these coherence conditions can be met by many different sets of preferences, the requirement of well-orderedness provides little further reasoned guidance. It does not add enough to an account of instrumental rationality to satisfy those who think that an account of practical reason should provide a more complete guide to reasoned choice, or to rebut the charge that instrumental reasoning gets no further than showing how intrinsically arbitrary (unreasoned) ends are to be pursued by rational means.

Some advocates of subjective end-oriented practical reasoning have held that desires and preferences, although subjective, provide an approximation to an objective account of the good. Some utilitarians, for example, have held that happiness is the sole good, that it is achieved by satisfying desires, and that therefore the optimal satisfaction of desires leading to maximal happiness produces the greatest good; they then claim that instrumental reasoning provides a sufficient account of practical reasoning for ethical as well as for egoistical, strategic and merely prudential purposes. Critics have countered that happiness is not the sole good, and moreover that it is not all of a sort, and cannot be aggregated, so that instrumental reasoning alone will not be enough to guide ethical choice. Others have objected that some sorts of happiness, such as happiness in the satisfaction of evil desires, or happiness produced by violating others' rights, does not contribute to the good at all, and conclude that instrumentally rational pursuit of happiness cannot be what ethics demands.

3 End-oriented reasoning: reason identifies objective ends

All of these difficulties would be resolved if reason could also identify objective ends. Instrumental reasoning is only one aspect of an older and more ambitious conception of end-oriented practical reasoning, which claims that reason both identifies the proper ends of action - the good - and can be used to steer action towards those ends. Such a position can be attributed to Plato, on whose account reason not merely can know its proper object, the [Form of the] Good, but strives for the Good. Far from being inert, reason is intrinsically active, has its own desires and its own end. In this picture there is no gap between theoretical reason, which guides right cognition, and practical reason, which guides right action. Human knowledge and desire have a common focus on the Good; *nous* and *eros* are aspects of one capacity, and the remaining practical problem is to align human life as that capacity directs (see Frede and Striker 1996).

The metaphysical and epistemological claims needed to support this more traditional vision of practical reason are hugely ambitious. Yet despite the difficulty of justifying objective conceptions of end-oriented practical reason, it has been widely accepted in varied forms. Aristotle, for example, took issue with Plato's unitary conception of the Good, and insisted that there are many goods, but held that they include the Good for man, which constitutes the proper end for human action (see Aristotle §§20-1; Eudaimonia). He offers a complex account of supplementary patterns of practical reasoning which can be used to identify action that contributes to the good for man. These include the *doctrine of the mean*, which purportedly offers a nonmechanistic way of selecting 'intermediate' action, so avoiding unacceptable extremes, and the so-called *practical syllogism*, by which conclusions about action (or possibly actions themselves) are inferred from general principles and claims about particular situations (see Anscombe 1957). Many later Neoplatonist and Christian thinkers also combine the ideas that the Good can be known by reason and that it is the proper end of human life. Plato's teleological view of reason has been widely if tacitly shared by countless other writers who neither share his metaphysical position nor provide any alternative, but who speak of certain ends as 'reasonable' without much by way of explanation, and often with far less by way of defence than Plato offers.

Of course, knowing what the Good is - what the true ends of human life are - is never enough to guide action. The effective pursuit of these ends also requires the use of reason to calculate which of various available acts contributes to those ends most effectively, as well as to identify any actions, attitudes or policies that are components of those ends. However, by itself this calculating, predominantly instrumental, side of end-oriented practical reason, as discussed in §2, cannot guide action without assuming subjective ends.

4 Act-oriented reasoning: reason appeals to norms

Other accounts of practical reason hold that it must guide action without reference either to subjective ends (because its conclusions would then be conditional on something wholly arbitrary) or to objective ends (because there is none). An adequate account of practical reason must bear more directly on action, and more specifically on the practical propositions (norms, rules, principles) which agents follow or embody in their lives. Just as theoretical reasoning prescribes ways of moving between elements that have syntactic and semantic structure, so too must practical reasoning. The central problem in using any act-oriented conception of practical reason for ethical purposes is to show *why* some but not other norms or principles, and so some but not other types of action and ends, attitudes and policies, are either good or ethically required.

Many forms of act-oriented practical reasoning maintain that the basis for distinguishing certain types of action from others can be found in the categories, beliefs and norms that form the constitutive elements of a society or of a sense of identity. These constitutive norms cannot be brought into question by anyone for whom they constitute the horizons of life and thought, so can provide basic premises for reasoning about what is good or bad, required or forbidden.

Once some fundamental norms have been identified, practical reasoning can be extended both by instrumental reasoning and by analysing the logical connections and implications between different norms (for example, between requirements, prohibitions and permissions) and the structure of systems of norms. At its most formal this sort of analysis draws on deontic logic; less formal conceptual investigations of types and systems of rules have also been undertaken, particularly by philosophers of law (see Deontic logic; Legal reasoning and interpretation) (see Raz 1975; Schauer 1991). However, since act-oriented practical reasoning does not provide any metric for ends it cannot use the maximizing patterns of practical reasoning which some subjective forms of end-oriented reasoning favour.

Norm-based practical reasoning is uncontroversially part and parcel of daily life, but philosophical argument that it is the basis for all ethical reasoning is highly controversial. Arguments that all ethical reasoning appeals to norms can be found in the work of Hegel, in historicist, communitarian and relativist writing, and (in more individualistic forms) in work by Wittgensteinians and by Bernard Williams.

Hegel's view that there is no gap between *is* and *ought* expresses not the implausible thought that whatever is accepted is acceptable, that (contrary to Hume) *ought* can be generally be derived from *is*, but the more profound view that all thinking and action must grow not out of abstract theories and principles but out of the deep structures of our actual situations (Hegel's term is *Sittlichkeit*, often translated as *ethical life*) (see Hegel, G.W.F. §8). The deep facts of our histories and lives are ones that we cannot 'go behind' and bring into question; rather they form the inescapable framework of our action, so constitute legitimate, indeed unavoidable, starting points for all practical, including ethical, reasoning.

Similar positions can be found in Wittgenstein and in certain Wittgensteinians. In Wittgenstein's assertion that 'there must be agreement… in judgments' (1953: paragraph 242) and in the widespread view that certain issues are 'not a matter for decision' (see Winch 1972), we see further versions of the thought that certain categories and norms form inescapable frameworks for life and thought, and that those who live within them lack any external vantage point from which to question or undermine them (see Wittgensteinian ethics). More individualistic versions of the same approach can be found in Bernard Williams' contention (1976) that certain identity-constituting personal projects and commitments are part of the framework which reasoning must assume, so cannot query (see Morality and identity §4).

Practical reasoning which appeals to constitutive norms, coupled with ordinary patterns of instrumental reasoning, has powerful means of guiding action at its disposal. Critics fear, however, that this power is bought too dearly. One alleged cost is that norm-based practical reasoning is supposedly conservative: it will always already have presupposed established (usually establishment) categories, norms, identities and commitments, and provides no vantage point from which they can be criticized. This criticism is rebutted by many proponents of norm-based reasoning who, like Hegel, point out that any constitutive norm, sense of identity or the like will be but part of a wider set of beliefs and norms, whose elements can be used to challenge, revise and renew one another (see MacIntyre 1981). Once norms and identities are seen in developmental, historical context there is no reason to suppose that practical reasoning that starts from those of a given place and time must be intrinsically conservative.

Perhaps a more worrying criticism of norm-based practical reasoning is that even if it is not intrinsically conservative, it is nevertheless unavoidably designed for those who have internalized a certain outlook and its categories and norms: it is insiders' reasoning. For outsiders, treating insiders' shared categories and norms, and the established practices and identities they support, as bedrock for practical reasoning lacks all justification, because it adopts arbitrary premises. Like instrumental reasoning, norm-based reasoning can at best reach conditional conclusions. Its advocates can only retort that there is no external vantage available from which these starting points can be rebutted or called into question.

5 Act-oriented reasoning: reason appeals to 'the world at large'

If act-oriented reasoning is to escape the confines and the criticisms which it incurs by treating socially specific categories and norms as the bedrock for ethical reasoning, then it must find some way of 'going behind' and criticizing these assumptions.

The classic version of a critical conception practical act-oriented practical reasoning was developed by Immanuel Kant, who held that reasoning should address 'the world at large' (that is, all reasoners) rather than the limited groups who share specific but intrinsically arbitrary norms and practices (see Kant 1784, 1786; O'Neill 1989 part I, 1992). If practical reasoning is to meet this standard, its first requirement must be the rejection of any principles which cannot be adopted by all, regardless of their social background, their accepted categories and norms, their established practices, their senses of identity or their desires (see Universalism in ethics §§1-2, 5). Kant summarized this requirement in the words: 'Act only on that maxim through which you can at the same time will that it should be a universal law' ([1785] 1903: 421). He claims that this principle of practical reason provides the supreme principle of morality, and should be called the 'categorical imperative' because it is the only way of reasoning practically which does not introduce arbitrary assumptions, and so the only one that can reach unconditional, namely categorical, conclusions. The component norms of social traditions and senses of identity may be scrutinized using the categorical imperative; if they cannot be willed as universal laws they must be rejected as unreasoned.

Kant's attempt to vindicate this critical account of practical reasoning is based on the thought that anything which deserves to be called reasoning must be something that can be given or received, exchanged or followed, among the widest 'public', that is universally, and correspondingly that anything which invokes the norms and beliefs of limited groups, let alone the favoured projects of individual lives, calls for rather than provides reasons. The authority of reason is simply the requirement of living by fundamental principles that are fit for universal use (in other words, that are *lawlike*). Since we know no intrinsic sources of authoritative standards - we have no account of the objectively good - the only meagre authority with which we are left is the injunction to reject principles which cannot be principles for all.

Once this fundamental use of practical reason has been used to identify certain core ethical principles or rules, critical practical reasoning can be extended by some of the moves used by instrumental and norm-based practical reasoning. Kant insists that instrumental reasoning is indispensable (although he cannot reinstate the metric assumptions or maximizing calculi advocated by those who hinge reasoning on subjective ends). He speaks of instrumental reasoning as guided by the principle of the 'hypothetical imperative', since by itself it can only licence conditional conclusions. His full account of practical reasoning also relies on the transitions between different modalities of required action (for example, between claims about obligations, permissions and prohibitions) that are used in norm-based reasoning.

There are many passages in which Kant indicates that the principles of theoretical and practical reason are fundamentally the same (such as [1785] 1903: 391). However in his case this appears to be the result neither of the derivation of practical from theoretical reason (as in instrumental reasoning) nor of the fusion of theory and practice (as in Platonism) but because his vindication of reason unites practical and theoretical reason (see Neimann 1994; O'Neill 1989 part II, 1992; Velkely 1989).

Kant's vindication of practical reason is controversial, and its adequacy as a guide to action is even more so. Since Hegel criticized Kant, many commentators have concluded that this stripped-down conception of practical reason is simply not enough to guide action, or alternatively that it will guide it in the 'wrong' direction. Some object that virtually any principle can be universally adopted and that this account of practical reason is not robust enough to

guide action; others object (with little plausibility) that the universality requirement makes rigidly uniform demands and leaves no room for the differentiation of action which human life requires.These classic (and incompatible) objections are respectively said to target empty ‘formalism’ and insensitive ‘rigourism’ (see Herman 1993; Hill 1992; O’Neill 1989 part II).

Kantian conceptions of practical reason have recently been taken up in two bodies of literature. Some writers have returned to Kant to look for more plausible interpretations of the procedures by which his conception of practical reason can guide action (see Herman 1993; Hill 1992; Korsgaard 1996; O’Neill 1989 part II) and to explicate his distinctive vindication of practical reason (see Neimann 1994; O’Neill 1989 part I, 1992; Velkely 1989). Others have offered a range of contemporary interpretations of the idea that the fundamental feature of practical reason is its capacity to be public (see Habermas 1995; Rawls 1993; McCarthy 1994).

6 Other aspects of practical reason

Few, if any, accounts of practical reason claim to offer a total guide for action (the exception may be certain utilitarian versions of subjective end-oriented reasoning, which supposedly reduce everything to calculation). Most insist that good practical reasoning must be linked with careful empirical reasoning, and that it will also need judgment to determine the specific way in which an end should be pursued or a norm or principle instantiated (see Moral Judgment). There is general acceptance of the point (emphasized by both Kant and Wittgenstein) that rules cannot provide instructions for their own application, and must be supplemented by judgment. Real life appeals to instrumental reasoning, to socially specific norms and to abstract principles are all inevitably indeterminate and must be augmented with judgment.

Reasoning practically is always a task rather than an automatic process, and there are therefore many ways in which it can fail. Some are cognitive failures (due, for example, to mistaken views about causal links or risks); others are closely linked to questions about motivation. Some central types of failure in practical reasoning have been the subject of extensive study (see Akrasia; Self-deception; Self-control).

See also: Examples in ethics; Good, theories of the; Moral education; Moral expertise

ONORA O’NEILL

References and further reading

Anscombe, G.E.M. (1957) ‘On Practical Reasoning’, in *Intention*, Oxford: Blackwell, 57-79; repr. in J. Raz (ed.) *Practical Reasoning*, Oxford: Oxford University Press, 1978, 33-45.(Discusses Aristotle’s practical syllogism.)

Audi, R. (1989) *Practical Reasoning*, London: Routledge.(Useful and accessible survey of different conceptions and patterns of practical reasoning; discusses Aristotle, Hume and Kant.)

Frede, M. and Striker, G. (eds) (1996) *Rationality in Greek Thought*, Oxford: Clarendon Press.(Scholarly papers on Greek views of reason.)

Habermas, J. (1995) *Between Facts and Norms*, Cambridge, MA: MIT Press. (Extensive discussion of practical - including ethical - reasoning.)

Herman, B. (1993) *The Practice of Moral Judgement*, Cambridge, MA: Harvard University Press.(Varied papers on Kantian approach to moral deliberation; discusses Williams’ internalist criticism of the approach.)

Hill, T.E., Jr (1992) *Dignity and Practical Reason*, Cambridge: Cambridge University Press.(Papers on Kant’s account of practical reason and ethics.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978.(Classic source for the view that there are no objective ends and that reason does not motivate but guides towards ends given by the passions.)

Kant, I. (1781/1787) *Critik der Reinen Vernunft*, trans. N. Kemp Smith, *Critique of Pure Reason*, London: Macmillan, 1973.(Kant’s considerations on the vindication of reason are mainly to be found in the prefaces and in the final part, titled *The Doctrine of Method*.)

Kant, I. (1784) *An Answer to the Question: ‘What is Enlightenment?’*, trans. B. Nisbet, in H. Reiss (ed.) *Kant: Political Writings*, Cambridge: Cambridge University Press, 2nd edn, 1992, 54-60.(Famous short essay on political philosophy, including material on the vindication of reason.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, in *Kants gesammelte Schriften*, ed. Königlichen Preußischen Akademie der Wissenschaften, Berlin: Reimer, vol. 4, 1903; trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New

York: Harper & Row, 1964.(References made to this work in the entry give the page number from the 1903 Berlin Akademie volume; these page numbers are included in the Paton translation. Classic short account of critical ethical reasoning.)

Kant, I. (1786) *What is Orientation in Thinking?*, trans. B. Nisbet, in H. Reiss (ed.) *Kant: Political Writings*, Cambridge: Cambridge University Press, 2nd edn, 1992, 237-49.(Short essay on thinking and philosophy of religion, including material on the vindication of reason.)

Korsgaard, C. (1996) *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.(Papers on Kant's account of practical reason and ethics.)

MacIntyre, A. (1981) *After Virtue: A Study In Moral Theory*, London: Duckworth. (An accessible defence of norm-based practical reasoning, and specifically of a Hegelianized reading of Aristotelian ethics. See ch. 5 for an account of deriving *ought* from *is*).

McCarthy, T. (1994) 'Kantian Constructivism and Reconstructivism: Rawls and Habermas in Dialogue', *Ethics* 105: 44-63.(Compares conceptions of practical reason used in the later works of Rawls and Habermas.)

Nagel, T. (1970) *The Possibility of Altruism*, Princeton, NJ: Princeton University Press.(Important account of prudential and ethical reasoning.)

Neimann, S. (1994) *The Unity of Reason: rereading Kant*, Oxford: Oxford University Press.(Argues for the unity of theoretical and of practical reason in Kant's philosophy).

O'Neill, O. (1989) *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge: Cambridge University Press.(Papers on Kant's vindication of reason and on patterns of ethical reasoning that start from the categorical imperative.)

O'Neill, O. (1992) 'Vindicating Reason', in *Kant: A Cambridge Commentary*, Cambridge: Cambridge University Press.(On Kant's vindication of a critical conception of reason.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press. (Reformulation of his earlier political philosophy basing it on a conception of public reason.)

Raz, J. (1975) *Practical Reason and Norms*, London: Hutchinson.(Careful account of conceptual links between different sorts of norms and systems of norms.)

Schauer, F. (1991) *Playing by the Rules: A Philosophical Examination of Rule-Based Decision Making*, Oxford: Clarendon Press.(Conceptual investigation of types and systems of rules.)

Velkeley, R. (1989) *Freedom and the End of Reason: On the Moral Foundation of Kant's Critical Philosophy*, Chicago, IL: University of Chicago Press.(Historically aware account of Kant's conception of reason and its vindication.)

Williams, B. (1976) 'Persons, Character and Morality', in A.O. Rorty (ed.) *The Identities of Persons*, Berkeley, CA: University of California Press, 197-216; repr. in *Moral Luck: Philosophical Papers 1973-80*, Cambridge: Cambridge University Press, 1981, 1-19.(Argues that reasons for action must be internal to the agent's own life.)

Winch, P. (1972) *Ethics and Action*, London: Routledge & Kegan Paul.(Deploys a Wittgensteinian conception of practical reason.)

Wittgenstein, L.J.J. (1953) *Philosophical Investigations*, Oxford: Blackwell, 88, paragraph 242.(This passage is usually read as a laconic formulation of the view that there are social limits to reason-giving.)

Pragmatics

Analytic philosophers have made lasting contributions to the scientific study of language. Semantics (the study of meaning) and pragmatics (the study of language in use) are two important areas of linguistic research which owe their shape to the groundwork done by philosophers.

Although the two disciplines are now conceived of as complementary, the philosophical movements out of which they grew were very much in competition. In the middle of the twentieth century, there were two opposing 'camps' within the analytic philosophy of language. The first - 'ideal language philosophy', as it was then called - was that of the pioneers, Frege, Russell and the logical positivists. They were, first and foremost, logicians studying formal languages and, through these formal languages, 'language' in general. Work in this tradition (especially that of Frege, Russell, Carnap, Tarski and later Montague) gave rise to contemporary formal semantics*, a very active discipline developed jointly by logicians, philosophers and grammarians. The other camp was that of so-called 'ordinary language philosophers', who thought important features of natural language were not revealed, but hidden, by the logical approach initiated by Frege and Russell. They advocated a more descriptive approach, and emphasized the 'pragmatic' nature of natural language as opposed to, for example, the 'language' of Principia Mathematica. Their own work (especially that of Austin, Strawson, Grice and the later Wittgenstein) gave rise to contemporary pragmatics, a discipline which (like formal semantics) has developed successfully within linguistics in the past thirty years.*

From the general conception put forward by ordinary language philosophers, four areas or topics of research emerged, which jointly constitute the core of pragmatics: speech acts; indexicality and context-sensitivity; non-truth-conditional aspects of meaning; and contextual implications. In the first half of this entry, we look at these topics from the point of view of ordinary language philosophy; the second half presents the contemporary picture. From the first point of view, pragmatics is seen as an alternative *to the truth-conditional approach to meaning associated with ideal language philosophy (and successfully pursued within formal semantics). From the second point of view, pragmatics merely* supplements *that approach.*

1 Pragmatics and ordinary language philosophy

The linguistic investigations undertaken by ordinary language philosophers in what was to become 'pragmatics' had been notably anticipated by various researchers belonging to other traditions (phenomenologists like Marty or Reinach, linguists like Bally or Gardiner, psychologists like Bühler, or anthropologists like Malinowski). However, what influenced ordinary language philosophers most was the conception of language advocated by 'ideal language philosophers', against which they reacted strongly (see Ordinary language philosophy).

Central in the ideal language tradition had been the equation of, or at least the close connection between, the meaning of a sentence and its truth-conditions. This truth-conditional approach to meaning, perpetuated by contemporary formal semantics, is one of the things which ordinary language philosophers found quite unpalatable (see Meaning and truth). Their own emphasis was on the distinction between 'language' and 'speech' (Gardiner 1932) or, equivalently, between 'sentence' and 'statement' (Austin [1950] 1971; Strawson [1950] 1971). It is the sentence (a unit of 'language') which has meaning, according to ordinary language philosophers; whereas it is the statement made by uttering the sentence in a particular context which has truth-conditions. The sentence itself does not have truth-conditions. Truth can only be predicated of sentences indirectly, via the connections between the sentence and the 'speech act' it can be used to perform. Rather than equating the meaning of a sentence with its alleged truth-conditions, some philosophers in the pragmatic tradition have suggested equating it with its speech act 'potential' (which may include, as a proper part, a certain truth-conditional potential; see Alston 1964: 37-9).

Suppose that we posit abstract objects, namely 'propositions', which have their truth-conditions essentially. Then the point made by ordinary language philosophers can be put as follows: sentences do not express propositions *in vacuo*, but only in the context of a speech act. Given that the same sentence can be used to make different speech acts with different contents, the 'proposition' which is the content of the speech act must be distinguished from the linguistic meaning of the sentence *qua* unit of the language ('sentence meaning') (see Propositions, sentences and statements). It must also be distinguished from the contextually determined meaning of a particular utterance of the sentence ('utterance meaning'), for the latter includes much more than merely the propositional content of the

speech act performed in uttering the sentence. Utterance meaning includes a rich 'non-truth-conditional' component: Besides the proposition it expresses, an utterance conveys indications concerning the type of speech act being performed, the attitudes of the speaker, the place of the utterance within the discourse, its presuppositions and so forth. Moreover there is a secondary layer of meaning which includes the 'contextual implications' of the speech act, and in particular what H.P. Grice called the 'conversational implicatures' of the utterance.

2 Speech acts

Speech act theory (see Austin 1975; Searle 1969) is concerned with communication: not communication in the narrow sense of transmission of information, but communication in a broader sense which includes the issuing of orders, the asking of questions, the making of apologies and promises, and so on. According to the theory, a speech act is more than merely the uttering of a grammatical sentence endowed with sense and reference. To speak is also to *do* something in a fairly strong sense: it is to perform what J.L. Austin called an 'illocutionary act'. In performing an illocutionary act, a speaker takes on a certain role and assigns a corresponding role to the hearer. By giving an order, speakers express the desire that their hearer follow a certain course of conduct and present themselves as having the requisite authority to oblige the hearer to follow the course of conduct in question, simply because it is their will. The social role taken on by the speaker who gives an order is embodied in the organizational notion of 'superior rank'. Austin stressed such institutional embodiments of illocutionary roles in order to show that language itself is a vast institution incorporating an array of conventional roles corresponding to the range of socially recognized illocutionary acts. From this point of view, assertion - the act of making a statement - is only one illocutionary act among many others.

Illocutionary acts have 'felicity conditions' (conditions which must be contextually satisfied for the illocutionary act to be successfully performed). Thus an assertion about an object 'presupposes' the existence of that object and is felicitous only if the object in question actually exists (Strawson [1950] 1971; Austin 1975). The study of felicity conditions is a central concern of speech act theory, along with the taxonomy of illocutionary acts. But the most central concern, perhaps, relates to the characterization of the very notion of an illocutionary act. Illocutionary acts are generally introduced ostensively, by examples, and they are distinguished both from the mere act of saying something ('locutionary act') and from the act of causing something to happen by saying something ('perlocutionary act', for example frightening, convincing and so on). The nature of the intermediate category of 'illocutionary acts' remains a matter of debate, however. The pioneers of speech act theory, Austin and Searle, advocated an institutional or conventional approach. In this framework the illocutionary acts performed in speech, like the acts that are performed in games (for example, 'winning a set' in tennis), are governed by rules and exist only against a background of conventions. But an alternative, 'intentionalist' view, originating from Grice [1957] 1989 and Strawson [1964] 1971, developed and is now the dominant trend in speech act theory (see §13) (see Speech acts).

3 Contextual implications

The notion of a contextual implication itself is a speech-act theoretic notion. If, besides the meaning or content of an utterance, there is another realm (namely that of the illocutionary act the utterance serves to perform), then along with the implications of what is said there will be a further set of implications derivable from the utterance (namely the implications of the illocutionary act itself). Some of these 'pragmatic' implications are fairly trivial. Thus, according to ordinary language philosophers, it is a rule of the language game of assertion that whoever asserts something believes what they say and has some evidence for it; even the liar, who does not obey this rule, has to pretend that they do if they want to participate in the game. This rule generates pragmatic implications: by asserting something and therefore engaging in the language game, the speaker 'implies' that they obey the rules of the game and, therefore, that they believe whatever they are asserting. The speaker cannot disavow these implications of their speech act without 'pragmatic contradiction'. A pragmatic contradiction is a conflict between what an utterance says and what it pragmatically implies. Thus Moore's famous paradoxical utterance, 'It is raining but I do not believe it', is not self-contradictory in the logical sense: the state of affairs it describes is logically possible (it might be raining without the speaker's knowing it). But the speaker's asserting that it rains implies that they believe it, and this contradicts the second part of the utterance. (The twin notions of pragmatic implication and pragmatic contradiction or 'pragmatic paradox' have been used to illuminate a variety of philosophical issues, including the nature of Descartes' *Cogito*.)

Less trivial are the contextual implications famously discussed by H.P. Grice (1989). According to Grice, the speaker making an utterance does not merely imply that they respect the rules of the language game; among the pragmatic implications of the utterance, we find a number of additional assumptions contextually required in order to maintain the supposition that the rules of the game are being observed. Suppose that I am asked whether I will go out; I reply: ‘It is raining’. As stated above, it is a rule of assertion that the assertor believes what they say and have some evidence for it. By virtue of this rule, my utterance implies that I believe that it is raining, and that I have some evidence for my assertion. Considered as an answer to a question, my utterance also implies that it provides the information requested by the addressee, for it is a rule of the Question-and-Answer game that the answerer must provide the requested piece of information. Now in order to maintain the supposition that the speaker’s utterance actually provides the requested information, additional premises are needed: for example, the assumption that the speaker will not go out if it rains. In conjunction with this contextual assumption, the utterance implies that the speaker will not go out, thereby providing a negative answer to the question. In so far as they serve to restore the utterance’s conformity to the rules of the game, the conclusion that the speaker will not go out and the contextual assumption through which it is derived are further pragmatic implications of the utterance. Grice called them ‘conversational implicatures’. Contrary to the more trivial pragmatic implications, they can be disavowed by the speaker without pragmatic contradiction (at least if there is another way of making the utterance compatible with the supposition that the rules of the game are being respected). This distinguishing feature of conversational implicatures is referred to as their ‘cancellability’. Implicatures which are not disavowed are legitimately taken as part of what the utterance communicates. They constitute a second layer of meaning, additional to what is literally said (see §12) (see Implicature).

4 Non-truth-conditional aspects of meaning

Like pragmatic implications, non-truth-conditional aspects of meaning are easy to account for if speech is considered as a rule-governed activity (Stenius 1967). What is the meaning of, for example, the imperative mood? Arguably, the sentences ‘You will go to the shop tomorrow at 8’, ‘Will you go to the shop tomorrow at 8?’ and ‘Go to the shop tomorrow at 8’ describe the same (sort of) state of affairs. The difference between them is pragmatic rather than descriptive: it relates to the type of illocutionary act being performed by the utterance. Thus the imperative mood indicates that the speaker, in uttering the sentence, performs an illocutionary act of a ‘directive’ type. (Such an act is governed by the rule that if the speaker performs a directive act with content *P*, the addressee is to make it the case that *P*.) To account for this ‘indication’, which does not belong to the utterance’s descriptive or propositional content, we can posit a rule or convention to the effect that the imperative mood is to be used only if one is performing a directive type of illocutionary act. This rule gives ‘conditions of use’ for the imperative mood. By virtue of this rule, a particular token of the imperative mood in an utterance u ‘indicates’ that a directive type of speech act is being performed by u. This (token-reflexive) indication conveyed by the token follows from the conditions of use that govern the type; these conditions of use constitute the linguistic meaning of the type (Recanati 1987: 15-7).

Pragmatic indications are a species of pragmatic implication: they are what the use of a particular expression pragmatically implies, by virtue of a certain condition of use conventionally associated with the expression. In contrast to more standard pragmatic implications, however, pragmatic indications are linguistically encoded, via the condition of use conventionally associated with the expression. Grice (1989) called such conventional pragmatic implications ‘conventional implicatures’, as opposed to ‘conversational implicatures’ (see Implicature §4). Whether they concern the type of the illocutionary act, as in the example I have given, or some other aspect of the context of utterance, pragmatic indications can always be accounted for in terms of conditions of use. They are ‘use-conditional’ aspects of meaning. Their exploration is one of the empirical tasks of semantics construed as the study of linguistic meaning under all its aspects (see §8 for an alternative construal of ‘semantics’).

5 Indexicals

Use-conditional meaning is not incompatible with descriptive content, in the sense that one and the same expression can be endowed with both. There are expressions which have a purely use-conditional meaning and do not contribute to truth-conditional content. Illocutionary markers like the imperative mood, or discourse particles such as ‘well’, ‘still’, ‘after all’, ‘anyway’, ‘therefore’, ‘alas’, ‘oh’ and so forth, fall into this category. Thus the following utterances have the same truth-conditional content, and are distinguished only by the pragmatic

indications they respectively convey:

Well, Peter did not show up
Still, Peter did not show up
After all, Peter did not show up
Therefore, Peter did not show up
Alas, Peter did not show up

But there are also expressions which have a two-layered meaning. Indexicals are a case in point. A 'rule of use' is clearly associated with indexicals: thus 'I' is governed by a convention of use (it is to be used to refer to the speaker). By virtue of this conventional rule, a use u of 'I' token-reflexively indicates that it refers to the speaker of u. But u also contributes to the utterance's truth-conditional content. 'I' being a directly referential expression, its truth-conditional contribution (its 'content') is its actual referent, not the rule of use which contextually determines the referent (Kaplan 1989: 481-563; Recanati 1993).

Besides the horizontal distinction between truth-conditional and non-truth-conditional aspects of meaning, we see that there is a vertical distinction between two levels of meaning for indexical expressions (Strawson [1950] 1971; Kaplan 1989: 481-563). At the first level - corresponding to the linguistic meaning of the expression-type - we find the rule of use conventionally associated with the expression. At the second level - corresponding to the context-dependent semantic value of the token - the rule of use determines the expression's 'content' (see Demonstratives and indexicals).

6 Levels of meaning

The two distinctions we have made, between truth-conditional and non-truth-conditional aspects of meaning on the one hand, and between levels of meaning on the other hand, should not be conflated (as they often have been). Despite appearances, they are orthogonal to each other. In the same way as the truth-conditional content of an indexical sentence is context-dependent and, therefore, belongs to the second level of meaning, the pragmatic indications conveyed by an expression governed by a rule of use also are context-dependent and belong to the second level of meaning. In other words, a distinction must be made between the rule of use (first level of meaning) and the pragmatic indications it contextually generates, in the same way as we distinguish between the rule of use and the truth-conditional content it contextually determines.

That pragmatic indications, though conventional, are context-dependent is shown by examples like (1):

(1) The weather is nice, but I have a lot of work.

The conjunction 'but' is governed by a certain condition of use which distinguishes it from 'and'. According to Ducrot (1972: 128-9), 'but' is to be used only if the following conditions are contextually satisfied:

(i) The first conjunct (*P*) supports a certain conclusion *r*;
(ii) The second conjuncts (*Q*) supports *not-r*;
(iii) *Q* is considered stronger than *P*, that is the whole utterance supports *not-r*.

Uttering (1) pragmatically implies that the conditions of use associated with 'but' are satisfied, that is that there is a conclusion *r* such that the first conjunct supports *r* and the second conjunct more strongly supports *not-r*. But the pragmatic implication conveyed by a particular utterance of (1) is much more specific. In context, the variable *r* is assigned a particular interpretation, for example 'we should go for a walk'. Example (1) therefore pragmatically implies something like the following: we should not go for a walk (because of all the work I have to do), despite the nice weather which suggests otherwise.

In so far as it is context-dependent and conveyed by the token, this pragmatic implication is to be located at the second level of meaning, alongside the content of indexicals. Even in a case where the pragmatic indication is fully conventional and not in need of contextual specification, it is conveyed by the token, not by the type. Thus a particular use u of the pronoun 'I' indicates that it (u) refers to the speaker of u. This token-reflexive indication is distinct from the rule of use, to the effect that for all *x*, if *x* is a token of 'I' it must be used to refer to the speaker of *x*.

The picture is further complicated by the Gricean distinction between what is literally said and what is non-literally or indirectly communicated. We end up with a three-fold distinction between the following layers of meaning:

First level. Rules of use ('character', in Kaplan's framework).
Second level. Truth-conditional content + pragmatic indications.
Third level. Conversational implicatures.

The need for a third level of meaning comes from the fact that the contextual process responsible for conversational implicatures (and non-conventional pragmatic implications in general) takes the second-level meaning of the utterance as input. When an expression is governed by a condition of use, using that expression pragmatically implies that the condition is satisfied. But conversational implicatures, in contrast to conventional implicatures, are not generated by virtue of a condition of use directly associated with a particular linguistic expression; they are normally generated by virtue of conversational norms that concern the content of utterances rather than the expressions which are used to convey that content. For example, a speaker should not say what he believes to be false (the 'maxim of quality' in Grice's terminology); as a result, saying that *P* pragmatically implies that the speaker believes that *P*. The generation of this pragmatic implication presupposes that the proposition expressed has been identified: from the fact that *the speaker has said that P*, together with the default assumption that the maxim of quality is respected, we can infer that *the speaker believes that P*. The implicature-generating process therefore deserves to be called a 'secondary pragmatic process' (Recanati 1993). There are three basic levels of meaning, with the context controlling the transition from the first to the second and from the second to the third. The proposition expressed by the utterance must first be contextually identified (primary pragmatic process) in order for the non-conventional pragmatic implications to be derived (secondary pragmatic process).

7 Open texture

For ordinary language philosophers, the truth-conditional or 'descriptive' content of an utterance is a property of the speech act, not a property of the sentence. A sentence only has truth-conditions in the context of a speech act. This is so not merely because of indexicality: the fact that the reference of some words depends on the context in a systematic way. Indexicality is only one form of context-dependence. There is another one, no less important, which affects the sense (the conditions of application) of words rather than their reference. According to Austin and Wittgenstein, words have clear conditions of application only against a background of 'normal circumstances' corresponding to the type of context in which the words were used in the past. There is no 'convention' to guide us as to whether or not a particular expression applies in some extraordinary situation. This is not because the meaning of the word is 'vague', but because the application of words ultimately depends on there being a sufficient similarity between the new situation of use and past situations. The relevant dimensions of similarity are not fixed once and for all; this is what generates 'open texture' (Waismann 1951). Ultimately, it is the context of utterance which determines which dimension of similarity is relevant, hence which conditions have to be satisfied for a given expression to apply (Travis 1975; 1981). It follows that the sense of ordinary descriptive words is context-dependent, like the reference of indexicals, though not quite in the same way. On this approach, which we may call 'contextualism', truth-conditions cannot be ascribed to sentence-types but only to utterances (Searle 1978; 1983).

Contextualism was a central tenet in the pragmatic conception of language developed by ordinary language philosophers (though some atypical ordinary language philosophers, like Grice, rejected it). This conception is, at bottom, a 'use theory of meaning'. Meaning is use, in the following sense: there is nothing more to meaning than use (Wittgenstein 1953). We are confronted with uses of words, and the meaning which those words acquire for us is only the sense we are able to *make* of those uses.

8 The semantics/pragmatics distinction

If much of contemporary pragmatics derives from the work of ordinary language philosophers, the name 'pragmatics' - contrasted with 'syntax' and 'semantics' - was coined by a philosopher in the ideal language tradition, Charles Morris (1938). The general conception associated with Morris' tripartite distinction has been very influential. On this conception (hereafter 'the traditional conception'), semantics and pragmatics are

complementary studies: semantics deals with meaning understood as representational content, whereas pragmatics deals with use. This view, still influential today, is at odds with the more radical conception developed within ordinary language philosophy and sketched in the previous sections. (According to the latter view, meaning cannot be divorced from use; semantics *is* pragmatics.) However, as we shall see, the border between semantics and pragmatics has become much fuzzier as the traditional conception, with its sharp contrast between the two disciplines, was modified in order to account for indexicality, pragmatic indications and related phenomena.

The traditional conception is commonly glossed in two different ways, corresponding to two different distinctions. The first distinction is that between 'meaning' (that is, representational content) and 'force', to be found both in Frege and Austin. (Often the term 'sense' is used instead of 'meaning', and the distinction referred to as the 'sense-force distinction'. Some commentators insist on distinguishing the use of 'sense' as the complement of 'force' from the use of 'sense' as the complement of 'reference'; see Sense and reference.) An utterance (1) represents a certain state of affairs, and (2) serves to perform a certain speech act. The semantics/pragmatics distinction is often expressed in terms of this distinction: semantics deals with representational content and studies the relations between words and the world, while pragmatics studies the relations between words and their users. The second distinction is that between sentence meaning and utterance meaning. Semantics is supposed to deal with the linguistic meaning of sentence-types, while pragmatics is concerned with the total significance of an utterance of the sentence by a particular speaker in a particular context. The traditional conception is indifferently expressed in terms of either distinction because, according to the traditional conception, the conventional meaning of the sentence-type *is* its representational content.

The problem with the traditional conception is precisely that it rests on the equation of the meaning of a sentence with its representational content. This equation cannot be accepted, for two reasons:

(1) Indexical sentences represent a specific state of affairs only in context; their representational content depends on some feature of their context of use, hence it cannot be equated with the context-independent meaning of the sentence-type. Because the truth-conditions of indexical sentences depend on their use, some authors have argued that the study of truth-conditions for such sentences belongs to pragmatics (Bar-Hillel 1954); it has even been suggested that pragmatics *is* the study of truth-conditions for indexical sentences (Montague 1968).
(2) As we have seen, there are use-conditional as well as truth-conditional aspects of the meaning of sentence-types. To account for them, it seems we must give up the purely truth-conditional conception of semantics and make room for a non-truth-conditional component within it. Alternatively, we can say that it is the business of 'pragmatics' to deal with some aspects of the linguistic meaning of sentence-types, namely those aspects which relate to use. Thus Gazdar (1979) defines pragmatics as 'semantics minus truth-conditions'.

Some philosophers have tried to defend the traditional conception by forcing use-conditional aspects of meaning into the mould of truth-conditional semantics. Take, for example, the imperative mood. One can use the pragmatic equivalence between the imperative 'Close the door!' and the 'explicit performative' 'I order you to close the door' (Austin 1975: 32) to support the claim that non-declarative sentences have a declarative paraphrase through which they can be given a truth-conditional analysis (Lewis 1970: 54-61). A number of similar attempts have been made to reduce use-conditional to truth-conditional aspects of meaning (for example, Davidson 1979). Despite these attempts, it is commonly acknowledged that not all aspects of linguistic meaning are truth-conditional. There are two components in the meaning of a sentence: a truth-conditional or descriptive component and a non-truth-conditional, pragmatic component. The pragmatic component of sentence meaning constrains the context of utterance: it is a 'procedural' component. As for the descriptive component of sentence meaning, it can no longer be equated with the truth-conditional content of the utterance (because of objection (1)), but it can still be construed as what *determines* that content with respect to a context of use.

This leaves us with two interpretations of the semantics/pragmatics distinction, both current in the contemporary literature. On one interpretation, dominant among philosophers, pragmatics deals with use - including use-conditional aspects of meaning (pragmatic indications, presuppositions and the like) - while semantics deals both with the descriptive component of sentence meaning and the truth-conditional content it determines (with respect to a context). The other interpretation, widespread among linguists, has it that semantics deals with conventional sentence meaning under all its aspects (including non-truth-conditional aspects), while pragmatics deals with use and the aspects of meaning which are contextual and use-dependent, that is, those that are conveyed

by the utterance but cannot be ascribed to the sentence-type (Katz 1977: 13- 22). (This category of ‘contextual meaning’ is somewhat heterogeneous; it includes both the semantic values of context-sensitive expressions, which are constitutive of the proposition literally expressed by the utterance, and other aspects of meaning which are not ‘literal’, conversational implicatures, for example.)

9 Context and propositional attitudes

The descriptive component of sentence meaning can be equated with a function from contexts to propositions (Stalnaker 1970; Kaplan 1989: 481-563). The ‘context’ is often construed as a package of various situational factors relevant to determining the semantic values of the context-sensitive constituents of the sentence. (See §10 for an alternative construal.) Thus the place of utterance, the identity of the participants in the speech episode and the time of utterance are among the factors on which the proposition expressed by an indexical sentence depends. It would be a mistake, however, to hold that only such ‘external’ (that is, non-intentional) features of the situation of utterance have a role to play in the determination of what is said. In many cases, what the speaker ‘has in mind’ is the relevant factor. Thus ‘John’s book’ can mean the book which John wrote, the book he bought, the book he is reading, and so forth. The sentence in which the expression occurs expresses a definite proposition only when a particular relation between John and a certain book has been contextually determined, but there is no ‘rule’ which enables the interpreter to determine the latter except that it must be the relation which the speaker ‘has in mind’. (This is in contrast with the case of ‘I’: as Barwise and Perry (1983: 33) pointed out, the reference of ‘I’ is fixed by the rule that ‘I’ refers to the speaker, irrespective of the speaker’s beliefs and intentions. Even if a speaker believes that they are Napoleon, their use of ‘I’ does not refer to Napoleon.) The same thing holds for demonstratives in general: contrary to the received opinion, the reference of a demonstrative is not the ‘demonstrated’ object, for there may be no accompanying demonstration; the reference, rather, is the object which the speaker has in mind and wishes to single out (Kaplan 1989: 565-614; Bach 1987).

It turns out that the context against which an utterance is interpreted includes factors like the intentions, expectations, beliefs and other propositional attitudes of the speaker and their audience. Especially important are the beliefs which are shared and ‘mutually known’ to be shared; they constitute a ‘common ground’ which can be exploited in discourse (see Stalnaker’s 1974 paper ‘Pragmatic Presuppositions’ in Davis 1991; Clark 1992). This introduces the topic of ‘presupposition’ which is generally mentioned, along with speech acts, indexicals and implicatures, as one of the central issues in pragmatics.

10 Presupposition

There is a basic sense in which ‘presupposing’ is a pragmatic attitude towards a proposition: that of ‘taking it for granted’. (One takes something for granted, for example, when one uses it as a hidden premise in an argument.) The ‘context’ is sometimes defined as a set of presuppositions in this sense, that is a set of propositions which are taken for granted at a given point in discourse (see Karttunen 1974, for example). Many authors think that in order to be part of the context, a proposition must be not only believed by the participants in the speech episode, but also believed to be believed, and so forth. Other authors find this ‘mutual belief’ requirement too strong (Smith 1982). Be that as it may, a more pressing question arises in connection with presuppositions. Beside the pragmatic notion of presupposition (where presupposing is something a *speech participant* does), is there also a purely semantic notion, where presupposing is something which a *sentence* does? For example, does the sentence ‘John stopped teaching undergraduates’ carry the presupposition that John used to teach undergraduates as part of its semantic, truth-conditional content?

The semantic notion of presupposition has been questioned on two grounds. First, it has been pointed out that presuppositions, like conversational implicatures, seem to be defeasible or cancellable. This might suggest that the basic, pragmatic sense is the only sense we can give to the notion of presupposition. *Sentences* do not have presuppositions; only the participants in a speech episode can presuppose something. This conclusion, however, seems too strong, for the conventional nature of presuppositions is manifest and well-documented. Arguably, what the defeasibility of presuppositions shows is not that presuppositions are non-conventional, but rather that they can be overridden if certain conditions are satisfied (Gazdar 1979).

More convincing is the claim that presuppositions, though part of the conventional meaning of the sentence, do not affect the truth-conditions of the utterance. Like pragmatic indications in general, the linguistic presuppositions

associated with certain expressions (such as the verb 'stop' in the example above) can be construed as conditions of use or constraints on the context (see Stalnaker's 1974 paper 'Pragmatic Presuppositions' in Davis 1991). The linguistic presupposition encoded by the verb 'stop' is a certain constraint on the context, namely the requirement that it contain a speaker with a certain pragmatic attitude (the attitude of 'presupposing') towards a certain proposition, or (if the context is directly construed as a set of propositions) the requirement that it contain a certain proposition, namely the proposition that John used to teach undergraduates. An utterance of 'John stopped teaching undergraduates' is 'appropriate' only in a context in which this constraint is satisfied. The constraint in question belongs to the non-truth-conditional component of sentence meaning: it does not affect the (truth-conditional) 'content' of the utterance (see Presupposition).

11 Interpretation and context-change

The rich, propositional notion of 'context' which features in discussions of presupposition is at the heart of contemporary pragmatics. As we have seen in connection with demonstratives and semantically indeterminate expressions, the context provides 'assumptions' concerning the speaker's intentions and expectations, which are used to determine the proposition expressed by the sentence. This process of determination is construed as fundamentally inferential and proposition-involving. (See §14 on the intentional-inferential approach.)

The propositional notion of context makes it possible to see the relation between context and content as two-way rather than one-way (Kamp 1985: 240). The proposition expressed, which depends on the context, itself changes the context. According to dynamic theories of discourse, the content of an assertion is normally fed into the context against which the next utterance will be interpreted (Karttunen 1974; Stalnaker 1978; Kamp 1985; Heim's 1988 paper 'On the Projection Problem for Presuppositions' in Davis 1991). The context of interpretation constantly changes - Stalnaker speaks of an 'everchanging context' - because it evolves as discourse proceeds. Thus it is possible for the context to shift in the middle of an utterance. This possibility accounts for a number of puzzling facts, including the defeasibility of presuppositions (see Discourse semantics).

If the proposition expressed by an utterance is normally fed into the context, the assumption that this proposition has been expressed *always* becomes part of the context as a result of the interpretation of the utterance. It is this assumption, together with the default assumption that the speaker respects the norms of conversation (plus various other assumptions included in the context), which make it possible to infer the conversational implicatures which enrich the overall meaning of the utterance. It follows that the contextual changes induced by an utterance by virtue of its expressing a certain proposition affect not only the interpretation of the utterances that follow, but equally the overall meaning of the very utterance responsible for the contextual change.

Another sort of context-change induced by an utterance has been described by David Lewis (1979). Sometimes the default assumption that the speaker respects the norms of conversation prevents the utterance from being interpreted with respect to the context at hand because, if it were so interpreted, it would violate the norms in question. This leads to a modification of the context in order to reach a more satisfactory interpretation. Thus if the utterance presupposes that *P*, and *P* is not part of the context at hand, it is introduced into the context in order to bring the utterance into conformity with the norms ('accommodation').

12 The strategic importance of conversational implicatures

If semantics and pragmatics both study the contextual determination of the proposition expressed (in so far as it depends both on the linguistic meaning of the sentence and the context), conversational implicatures fall within the sole domain of pragmatics, for they are not constrained by the linguistic meaning of the sentence in the way the proposition expressed is. Yet the theory of implicatures has important consequences for semantics. Thanks to Grice's theory, many intuitive aspects of meaning can be put into the 'pragmatic wastebasket' as implicatures, rather than treated as genuine data for semantics. Take, for example, the sentence '*P* or *Q*'. It can receive an inclusive or an exclusive interpretation. Instead of saying that 'or' is ambiguous in English, we may consider it as unambiguously inclusive, and account for the exclusive reading by saying that in some contexts the utterance conversationally implicates that *P* and *Q* are not both true. When there is such a conversational implicature, the overall meaning of the utterance is clearly exclusive, even though what is strictly and literally said corresponds to the logical formula '$P \vee Q$'. It is here that the complementary character of semantics and pragmatics is particularly manifest. Semantics is simplified because a lot of data can be explained away as 'implicatures' rather

than genuine aspects of the (literal) meaning of the utterance.

Grice's theory of implicatures has been extremely popular among semanticists precisely because it enables the theorist, when certain conditions are satisfied, to shift the burden of explanation from semantics to pragmatics. From this point of view, the most interesting notion is that of 'generalized' conversational implicatures (Grice 1989; Gazdar 1979; Levinson 1983). When a conversational implicature is generalized, that is generated by default, it tends to become intuitively indistinguishable from semantic content. Grice's theory has taught the semanticist not to take such 'semantic' intuitions at face value. Even if something seems to be part of the semantic content of an utterance, the possibility of accounting for it pragmatically must always be considered.

Grice's theory is important also because it has provided an influential argument against the contextualism professed by ordinary language philosophers. For example, Strawson had claimed that the truth-conditions of '*P* and *Q*' in English are contextually variable: the notion of temporal succession, or that of causal connection, or a number of other suggestions concerning the connection between the first and the second conjunct can enter into the interpretation of '*P* and *Q*', depending on the context (Strawson 1952: 81-2). 'They got married and had many children' means that they had children after getting married; 'Socrates drank the hemlock and died' means that he died as result of drinking the hemlock. Those aspects of the interpretation are very much context-sensitive; yet they affect the utterance's truth-conditions. The truth-conditions of '*P* and *Q*', therefore, are not fixed by a rigid rule, but depend on the context. Against this view, Grice has argued that the truth-conditions of '*P* and *Q*' *are* fixed and context independent. '*P* and *Q*' is true if and only if *P* and *Q* are both true. Thus 'They got married and had many children' would be true, even if they had the children before getting married. Certainly the utterance conveys the suggestion that the children came after the marriage, but this suggestion is nothing other than a conversational implicature, according to Grice (see Implicature §6). It does not affect the utterance's semantic content - its literal truth-conditions. Grice criticized his fellow ordinary language philosophers for confusing the truth-conditions of an utterance with its total significance. Though controversial (Travis 1985), this argument has been very popular, and it has played a major role in the subsequent downfall of ordinary language philosophy.

13 Communicative intentions

The pioneers of pragmatics (Malinowski and Austin, for example) used to insist on the social dimension of language as opposed to its cognitive or representational function. As pragmatics developed, however, it is the psychological dimension of language use that came to the forefront of discussions, in part as a result of Grice's work on meaning and communication.

Grice ([1957] 1989) defined a pragmatic notion of meaning: the notion of *someone* meaning something by a piece of behaviour (a gesture, an utterance and so forth). Grice's idea was that this pragmatic notion of meaning was basic and could be used to analyse the semantic notion, that is what it is for a linguistic *expression* to have meaning (see Grice, H.P. §3). Strawson soon pointed out that Grice's pragmatic notion of meaning could also be used to characterize the elusive notion of an illocutionary act (Strawson [1964] 1971). In §2, the view that illocutionary acts are essentially conventional acts (like the acts which owe their existence to the rules of a particular game) was mentioned. This conventionalist approach was dominant in speech act theory until Strawson established a bridge between Grice's theory of meaning and Austin's theory of illocutionary acts. Illocutionary acts, in the new framework, can be analysed in terms of the utterly non-mysterious notion of a 'perlocutionary act'.

A perlocutionary act consists in bringing about certain effects by an utterance. For example, by saying to you 'It is raining', I bring it about that you believe that it is raining. Now, according to the suggested analysis, to perform the illocutionary act of asserting that it is raining is (in part) to make manifest to the addressee one's intention to bring it about, by this utterance, that the addressee believes that it is raining. An illocutionary act therefore involves the manifestation of a corresponding perlocutionary intention. But there is a special twist which the suggested analysis inherits from Grice's original conception of meaning: the intention must be made manifest in a specially 'overt' manner. Not only must the speaker's intention to bring about a certain belief in the addressee be revealed by his utterance, but his intention to reveal it must also be revealed, and it must be revealed in the same overt manner. This characteristic (if puzzling) feature of overtness is often captured by considering the revealed intention itself as reflexive: a communicative intention, that is the type of intention whose manifestation constitutes the performance of an illocutionary act, is the intention to achieve a certain perlocutionary effect (for example, bringing about a certain belief in the addressee) via the addressee's recognition of this intention. Also

relevant to the characterization of overtness is the notion of ‘mutual knowledge’ (Lewis 1969; Schiffer 1972) which we have seen at work in the characterization of ‘contexts’ and ‘presuppositions’ (see §10) (see Communication and intention; Meaning and communication).

14 The intentional-inferential model

Even though the conventionalist approach to communication is still alive, the Grice-Strawson ‘intentionalist’ approach has gained wide currency in pragmatics. Typical in this respect are the neo-Gricean theories offered by Bach and Harnish (1979) and Sperber and Wilson (1986). They have put forward an inferential model of communication intended to supersede the ‘code model’ (or ‘message model’) that was inspired from Shannon and Weaver.

According to the code model, communication proceeds as follows: to communicate a certain content, the speaker encodes it into a sentence using the grammar of the language as a ‘code’ pairing contents and sentences (possibly with respect to a context of utterance); the interpreter, by virtue of their knowledge of the same grammar (and, perhaps, of the context), can decode the sentence and recover the intended content.

The alternative, inferential model of communication is very different. An utterance is seen as a meaningful action, one which provides interpreters with evidence concerning the agent’s intentions. What distinguishes communicative acts from other meaningful actions is what can be inferred from the evidence: a communicative act is an act which provides evidence of a certain ‘communicative intention’ on the part of the speaker. In other words, the speaker’s intention to communicate something is what explains their utterance, when considered as a piece of behaviour. From this point of view, the content of the communicative act - what is communicated - is the total content of the communicative intentions which can be inferred from it. Let us call this the utterance’s communicative meaning, distinct from the literal or conventional meaning of the sentence (determined by the code, that is the grammar). Understanding is essentially an inferential process in this framework, and the conventional meaning of the sentence provides only part of the evidence used in determining the communicative meaning of the utterance.

15 Pragmatics and modularity

A characteristic feature of recent work in the Gricean tradition has been the explicit employment of concepts from (and the intention to contribute to) cognitive science. For example, Fodor’s distinction between central thought processes and more specialized cognitive ‘modules’ has been found relevant to the characterization of the task of pragmatics (see Modularity). (See, for example, Kasher’s papers in Davis 1991 and Tsohatzidis 1995.)

In the inferential framework, comprehension involves not only a specifically linguistic competence, namely knowledge of ‘grammar’, but also general intelligence (that is, world knowledge together with inferential abilities). Contextual assumptions of various sorts, including assumptions about the speaker’s beliefs and expectations, play a crucial role not only in the inferential process which, according to Grice and his followers, underlies the generation of conversational implicatures, but also in the determination of the proposition literally expressed by the utterance (Carston 1988; Kempson 1988). Pragmatics, therefore, is concerned with the interaction between the language faculty and central thought processes in the task of linguistic comprehension (Sperber and Wilson 1986). It considers how the output of the linguistic module is centrally processed, that is processed against the interpreter’s complete belief system.

The major difficulty here is the ‘holistic’ character of the belief system (Fodor 1983). How is the total cognitive background restricted so as to yield a ‘context’ of manageable size? Sperber and Wilson’s ‘relevance theory’ addresses this crucial issue. Their view of pragmatics is grounded in a general theory of cognition as relevance-oriented. One of their central ideas is that the context in which an utterance is processed is not ‘given’ but ‘constructed’: it results from an active search driven by the overarching goal to maximize relevance.

Sperber and Wilson’s claim that there is no special faculty or module corresponding to pragmatics (since pragmatics studies the interaction between general intelligence and linguistic modules) must be qualified. As Sperber himself stressed in various places (for example, Sperber 1994), there is one special capacity which communicators must possess and which, arguably, poor communicators (for example, autistic children) do not possess: the capacity to ascribe complex propositional attitudes, such as higher-order intentions. (As we have seen,

understanding crucially involves a process of intention recognition, and the communicative intentions which must be recognized are complex, higher-order intentions.)

16 Cognitive science and contextualism

The intentional-inferential framework has been very influential in cognitive science. The Grice-Strawson interpretation of speech act theory has given rise to computational models of discourse where intention recognition plays a key role (see Grosz, Pollack and Sidner 1989; Cohen, Morgan and Pollack 1989). The idea that meaning is 'inferred' has also been taken very seriously. Many cognitive scientists believe that in interpretation the meaning of an utterance is arrived at by 'guesswork', rather than by decoding (Green 1989).

Typical in this respect are complex nominals (noun-noun compounds like 'finger cup' or adjective-noun compounds like 'philosophical kitchen'), which have received a good deal of attention. The semantic value which such expressions assume in context (unless they are idiomatic) is not predictable on a purely linguistic basis; it results from an act of 'sense creation'. Here again, what matters is what the speaker has in mind; the interpreter can only guess, using a variety of contextual clues. Such expressions whose sense is irreducibly contextual have been dubbed 'contextual expressions' (Clark 1992).

Contextual expressions are 'semantically indeterminate'. The extent of indeterminacy in language threatens the standard picture adhered to by most semanticists. According to that picture, 'what is said' is conventionally determined by the meaning of the sentence and the context. Then, and then only, is what the speaker really means 'inferred', if there is reason to think that what the speaker means is distinct from what is said (for example, if the speaker's literal contribution seems conversationally inappropriate and some further assumption is required in order to restore its conformity to conversational norms). As against this, both what is said and what is implied are seen as resulting from guesswork, because of semantic indeterminacy.

The fallback position for semanticists consists in holding that at least the meaning of the sentence-type is 'decoded' and can be read off the meanings of the constituents (as given in the lexicon) and the way they are put together. Decoding stops, and guesswork starts, only when we go from the meaning of the sentence to what the speaker means (this including both what is said and what is conversationally implied). But the fallback position itself has come under attack in the cognitive science literature. The very notion of the linguistic meaning of an expression-type has been questioned.

What is the linguistic meaning of a word? Is it possible to draw a line between 'pure' lexical knowledge and world knowledge? Arguably it is not (Langacker 1987: 154-66). For Langacker, the meaning of a word is a point of access into an essentially encyclopedic network. There is no distinction between 'dictionary' and 'encyclopedia'. A similar idea underlies Rumelhart's claim that interpretation is 'top down' from the bottom up (Rumelhart 1979). According to Rumelhart, what linguistic expressions do is evoke certain 'schemata' in memory. Those memorized schemata are part of our knowledge of the world. Interpretation, whether literal or non-literal, consists in finding evoked schemata which fit the situation the speaker seems to be talking about.

Langacker and Rumelhart reject not only the dictionary/encyclopedia distinction but also the literal/non-literal distinction as traditionally conceived. They insist that the same cognitive processes are involved in literal and (for example) metaphorical interpretations (see Gibbs 1994; Recanati 1995). A psychologist, Douglas Hintzman, has gone even farther, questioning the most basic distinction: that between the meaning of an expression-type and the contextual meaning of the token. Hintzman has developed a multiple-trace memory model in which the cognitive experiences associated with past tokens of a word interact with the present experience involving a new token of the word to yield the contextual 'meaning' of the latter (Hintzman 1986). Hintzman's model does not appeal to the notion of the literal meaning of the word-type. Words, as expression-types, do not have 'meanings' over and above the collection of token-experiences with which they are associated. The only meaning that words have is that which emerges in context.

Rejecting the three distinctions above - literal/non-literal, dictionary/encyclopedia and type-meaning/token-meaning - amounts to rejecting the semantics/pragmatics distinction. The 'eliminativist' approach to linguistic meaning developed within cognitive science constitutes a return to contextualism, the radical conception of pragmatics associated with ordinary language philosophy (§7). Pragmatics absorbs semantics in a contextualist framework.

See also: Strawson, P.F.; Wittgenstein, L.

FRANÇOIS RECANATI

References and further reading

Alston, W. (1964) *Philosophy of Language*, Englewood Cliffs, NJ: Prentice Hall.(A textbook in the ordinary language tradition.)

Atlas, J. (1992) *Philosophy without Ambiguity: A Logico-Linguistic Essay*, Oxford: Clarendon Press.(On semantic indeterminacy.)

Austin, J.L. (1971) *Philosophical Papers*, Oxford: Clarendon Press, 2nd edn.(A posthumous collection of Austin's papers.)

__ (1975) *How to Do Things with Words*, Oxford: Clarendon Press, 2nd edn.(Austin's lectures on speech act theory, delivered at Harvard University in 1955.)

Bach, K. (1987) *Thought and Reference*, Oxford: Clarendon Press.(Discusses a number of pragmatic issues.)

Bach, K. and Harnish, M. (1979) *Linguistic Communication and Speech Acts*, Cambridge, MA: MIT Press. (Speech act theory within an intentional-inferential framework.)

Bar-Hillel, Y. (1954) 'Indexical Expressions', in *Aspects of Language*, Jerusalem: Magnes Press, 1970, 69-88. (Classic paper by a disciple of Carnap who was influenced by ordinary language philosophy. The 1970 collection also contains Bar-Hillel's pioneering papers on indexicality, context-dependence and pragmatic implication.)

Baron-Cohen, S., Leslie, A. and Frith, U. (1985) 'Does the autistic child have a "theory of mind"?', *Cognition* 21: 37-46.(See §15.)

Barwise, J. and Perry, J. (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press/Bradford Books.(A semantic theory which takes context-sensitivity seriously.)

Blakemore, D. (1992) *Understanding Utterances: An Introduction to Pragmatics*, Oxford: Blackwell.(An introduction to Relevance Theory.)

Bühler, K. (1990) *Theory of Language*, trans. D. Goodwin, Amsterdam: Benjamins.(Originally published in 1934, still worth studying.)

Carston, R. (1988) 'Implicature, Explicature, and Truth-Theoretic Semantics', in R. Kempson (ed.) *Mental Representations: The Interface Between Language and Reality*, Cambridge: Cambridge University Press, 155-81.(On levels of meaning.)

Clark, H. (1992) *Arenas of Language Use*, Chicago, IL: University of Chicago Press/CSLI.(An important collection of papers by a pragmatically-oriented psychologist, most of them written in collaboration. Part 1 concerns the context as 'common ground'; part 3 deals with 'sense creation'.)

Cohen, P., Morgan, J.L. and Pollack, M. (eds) (1989) *Intentions in Communication*, Cambridge: MA: MIT Press.(Intention recognition and speech acts from a computational point of view.)

Cole, P. (ed.) (1978) *Syntax and Semantics 9: Pragmatics*, New York: Academic Press.(Includes papers by Kaplan, Stalnaker and Grice, among others.)

Cole, P. (ed.) (1981) *Radical Pragmatics*, New York: Academic Press.(Several papers on the interface between semantics and pragmatics.)

Cole, P. and Morgan, J.L. (eds) (1975) *Syntax and Semantics 3: Speech Acts*, New York: Academic Press.(An influential collection of papers, mostly on 'indirect speech acts'; includes Grice's third William James Lecture, on conversational implicatures.)

Davidson, D. (1979) 'Moods and Performances', in Margalit, A. (ed.) *Meaning and Use*, Dordrecht: Reidel, 1979, 9-20.(How to force pragmatic indications into the mould of truth-conditional semantics.)

Davis, S. (ed.) (1991) *Pragmatics: A Reader*, New York: Oxford University Press.(An inclusive collection, containing many influential papers.)

Ducrot, O. (1972) *Dire et ne pas dire: principes de sémantique linguistique (Saying and Not Saying: Principles of Linguistic Semantics)*, Paris: Hermann.(A speech-act theoretic approach to presupposition; contains insightful discussions of many pragmatic issues.)

Fodor, J. (1983) *The Modularity of Mind*, Cambridge, MA: MIT Press/Bradford Books.(Distinguishes between input systems and central systems in the human mind.)

Gardiner, A. (1932) *The Theory of Speech and Language*, Oxford: Clarendon Press.(A remarkable anticipation of speech act theory. Still worth reading.)

Gazdar, G. (1979) *Pragmatics: Implicature, Presupposition and Logical Form*, New York: Academic Press. (Rigorous work in the formal-semantic tradition.)

Gibbs, R. (1994) *The Poetics of Mind*, Cambridge: Cambridge University Press.(Argues that the same cognitive processes are at work in literal and non-literal communication.)

Green, G. (1989) *Pragmatics and Natural Language Understanding*, Hillsdale, NJ: Erlbaum.(An introduction to pragmatics for cognitive scientists.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge: Cambridge University Press.(Contains the William James Lectures and a collection of Grice's papers on the philosophy of language, including 'Meaning'.)

Grosz, B., Pollack, M. and Sidner, C. (1989) 'Discourse', in M. Posner (ed.) *Foundations of Cognitive Science*, Cambridge, MA: Bradford Books/MIT Press, 437-68.(A survey of computational pragmatics.)

Heim, I. (1988) 'On the Projection Problem for Presuppositions', in S. Davis (ed.) *Pragmatics: A Reader*, New York: Oxford University Press, 1991, 397-405.(Uses the notion of 'context-change potential'.)

Harnish, R.M. (1976) 'Logical Form and Implicature', in S. Davis (ed.) *Pragmatics: A Reader*, New York: Oxford University Press, 1991, 316-64.(How to shift the burden of explanation from semantics to pragmatics.)

Hintzman, D. (1986) '"Schema Abstraction" in a Multiple-Trace Memory Model', *Psychological Review* 93 (4): 411-28.(An exemplar model of categorization which has deep implications for semantics and pragmatics.)

Horn, L. (1989) *A Natural History of Negation*, Chicago, IL: University of Chicago Press.(The last three chapters of this monograph deal with the interplay between semantic and pragmatic factors in interpretation.)

Kamp, H. (1985) 'Context, Thought and Communication', *Proceedings of the Aristotelian Society* 85: 239-61.(An informal presentation of Discourse Representation Theory.)

Kaplan, D. (1989) 'Demonstratives' and 'Afterthoughts', in J. Almog, H. Wettstein and J. Perry (eds) *Themes from Kaplan*, New York: Oxford University Press, 481-563, 565-614.('Demonstratives' deals with the logic of indexicals, with 'Afterthoughts' providing further comments.)

Karttunen, L. (1974) 'Presupposition and Linguistic Context', *Theoretical Linguistics* 1: 181-94.(The third in an important series of papers by Karttunen on presuppositions.)

Kasher, A. (ed) (1997) *Pragmatics: Critical Assessment*, London: Routledge.(A new reader, very inclusive and containing postscripts by some of the authors.)

Katz, J. (1977) *Propositional Structure and Illocutionary Force*, New York: Crowell.(Semantics, pragmatics and speech act theory.)

Kempson, R. (1988) 'Grammar and Conversational Principles', in F. Newmeyer (ed.) *Linguistics: The Cambridge Survey*, vol. 2, 139-63.(Redraws the boundary between semantics and pragmatics.)

Langacker, R. (1987) *Foundations of Cognitive Grammar*, vol. 1, Stanford, CA: Stanford University Press.(A new paradigm, based on the rejection of traditional distinctions, such as syntax/semantics and semantics/pragmatics.)

Levinson, S. (1983) *Pragmatics*, Cambridge: Cambridge University Press.(The standard textbook.)

Levinson, S. (1988) *Generalised Conversational Implicatures and the Semantics/Pragmatics Interface*, unpublished typescript.(How pragmatics affect truth-conditions.)

Lewis, D. (1969) *Convention*, Cambridge, MA: Harvard University Press.(Introduces the notion of 'common knowledge'.)

Lewis, D. (1970) 'General Semantics', *Synthese* 22: 18-67.(Section 8 provides a truth-conditional treatment of non-declaratives.)

Lewis, D. (1973) *Counterfactuals*, Oxford: Blackwell.(Introduces 'salience' in connection with incomplete definite descriptions.)

Lewis, D. (1979) 'Scorekeeping in a Language-Game', *Journal of Philosophical Logic* 8: 339-59.(On 'accommodation' and context-change.)

Lewis, D. (1980) 'Index, Context, and Content', in S. Kanger and S. Öhman (eds) *Philosophy and Grammar*, Dordrecht: Reidel, 79-100.(Shows the need for a distinction between 'index' and 'context'.)

Lyons, J. (1977) *Semantics*, Cambridge: Cambridge University Press.(Volume 2 of this textbook contains three chapters on pragmatic issues.)

Malinowski, B. (1923) 'The Problem of Meaning in Primitive Languages', supplement to C. Ogden and I.A. Richards, *The Meaning of Meaning*, London: Routledge, 10th edn, 1949, 296-336.(Speech as a mode of social action.)

Margalit, A. (ed.) (1979) *Meaning and Use*, Dordrecht: Reidel.(A collection of papers dedicated to the memory of Bar-Hillel, a pioneer of pragmatics.)

Montague, R.M. (1968) 'Pragmatics', in R. Klibansky (ed.) *La Philosophie Contemporaine 1*, Florence: La Nuova Italia Editrice, 102-22.(A classic paper in formal semantics.)

Morris, C. (1938) *Foundations of the Theory of Signs*, Chicago, IL: University of Chicago Press.(Introduces the famous trichotomy between syntax, semantics and pragmatics. See §8.)

Nerlich, B. and Clarke, D. (1994) 'Language, action and context: Linguistic pragmatics in Europe and America (1800-1950)', *Journal of Pragmatics* 22: 439-63.(The pre-history of pragmatics.)
Nunberg, G. (1979) 'The Non-Uniqueness of Semantic Solutions: Polysemy', *Linguistics and Philosophy* 3: 143-84.(Influential paper on the pragmatics of reference.)
Recanati, F. (1987) *Meaning and Force: The Pragmatics of Performative Utterances*, Cambridge: Cambridge University Press.(A survey of pragmatics and speech act theory, with special attention to the debate between conventionalist and intentionalist approaches.)
Recanati, F. (1993) *Direct Reference: From Language to Thought*, Oxford: Blackwell.(The second part deals with the demarcation between semantics and pragmatics.)
Recanati, F. (1995, forthcoming) 'The alleged priority of literal interpretation', *Cognitive Science*, forthcoming. (Criticizes Grice's sequential model of non-literal interpretation.)
Reddy, M. (1979) 'The Conduit Metaphor', in A. Ortony (ed.) *Metaphor and Thought*, Cambridge: Cambridge University Press, 2nd edn, 1994, 164-201.(A classic paper which has influenced Lakoff's theory of metaphor.)
Rumelhart, D.E. (1979) 'Some problems with the notion of literal meaning', in A. Ortony (ed.) *Metaphor and Thought*, Cambridge: Cambridge University Press, 2nd edn, 1994, 71-82.(An iconoclastic paper. See §16.)
Schiffer, S. (1972) *Meaning*, Oxford: Clarendon Press.(An important work in the Gricean tradition, introducing the notion of 'mutual knowledge'.)
Searle, J. (1969) *Speech Acts*, Cambridge: Cambridge University Press.(A systematic theory of speech acts within a conventionalist framework.)
Searle, J. (1978) 'Literal Meaning', in *Expression and Meaning*, Cambridge: Cambridge University Press, 1979, 117-36.(A collection of influential papers on speech act theory. 'Literal Meaning' generalizes context-sensitivity: words have conditions of satisfaction only against a background of assumptions and practices.)
Searle, J. (1983) *Intentionality*, Cambridge: Cambridge University Press.(Chapters 5 and 6 bear on issues such as literal meaning, the 'background', and the role of speaker's intentions.)
Simon, H. (1994) 'Literary Criticism: A Cognitive Approach', in G. Güzeldere and S. Franchi (eds) '"Bridging the Gap": Where Cognitive Science Meets Literary Criticism', *Stanford Humanities Review*, supplementary vol. 4 (1): 1-26.(A pragmatic theory of meaning by one of the founding fathers of cognitive science.)
Smith, N. (ed.) (1982) *Mutual Knowledge*, London: Academic Press.(Contains an interesting debate between Clark and Carlson and Sperber and Wilson over mutual knowledge.)
Sperber, D. (1994) 'Understanding Verbal Understanding', in J. Khalfa (ed.) *What is Intelligence?*, Cambridge: Cambridge University Press, 179-98.(The role of meta-representations in communication.)
Sperber, D. and Wilson, D. (1986) *Relevance: Communication and Cognition*, Oxford: Blackwell.(A cognitively oriented theory in the Gricean tradition. See §§14-15.)
Stalnaker, R. (1970) 'Pragmatics', *Synthese* 22: 272-89.(A lucid and original survey.)
Stalnaker, R. (1974) 'Pragmatic Presuppositions', repr. in S. Davis (ed.) *Pragmatics: A Reader*, New York: Oxford University Press, 1991, 471-82.(See §10.)
Stalnaker, R. (1978) 'Assertion', in P. Cole (ed.) *Syntax and Semantics 9: Pragmatics*, New York: Academic Press, 1978, 315-22.(Sketches a theory of speech acts in terms of context-change.)
Stenius, E. (1967) 'Mood and Language-Game', *Synthese* 17: 254-74.(Language as a rule-governed activity. See §4.)
Strawson, P. (1950) 'On Referring', in *Logico-Linguistic Papers*, London: Methuen, 1971, 1-27.(On presupposition and the sentence/statement distinction. The collections contains Strawson's most important papers in the philosophy of language.)
Strawson, P. (1952) *Introduction to Logical Theory*, London: Methuen.(Insists on the differences between natural and artificial languages.)
Strawson, P. (1964) 'Intention and Convention in Speech Acts', in *Logico-Linguistic Papers*, London: Methuen, 1971, 149-69.(A classic of speech act theory. See §13.)
Travis, C. (1975) *Saying and Understanding*, Oxford: Blackwell.(Emphasizes the context-dependence of sense.)
Travis, C. (1981) *The True and the False: the Domain of the Pragmatic*, Amsterdam: Benjamins.(A case for contextualism.)
Travis, C. (1985) 'On What is Strictly Speaking True', *Canadian Journal of Philosophy* 15: 187-229.(Criticizes Grice's critique of contextualism.)
Tsohatzidis, S.L. (ed.) (1995) *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, London: Routledge.(A new collection containing original papers on pragmatics and speech act theory.)
Waismann, F. (1951) 'Verifiability', in A. Flew (ed.) *Logic and Language*, 1st series, Oxford: Blackwell, 17-44.

(On ‘open texture’, by a disciple of Wittgenstein.)
Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(A pragmatic approach to language.)

Pragmatism

Pragmatism is a philosophical tradition founded by three American philosophers: Charles Sanders Peirce, William James and John Dewey. Starting from Alexander Bain's definition of belief as a rule or habit of action, Peirce argued that the function of inquiry is not to represent reality, but rather to enable us to act more effectively. He was critical of the 'copy theory' of knowledge which had dominated philosophy since the time of Descartes, and especially of the idea of immediate, intuitive self-knowledge. He was also a prophet of the linguistic turn, one of the first philosophers to say that the ability to use signs is essential to thought.

Peirce's use of Bain was extended by James, whose The Principles of Psychology (1890) broke with the associationism of Locke and Hume. James went on, in Pragmatism (1907) to scandalize philosophers by saying that '"The true"… is only the expedient in our way of thinking'. James and Dewey both wanted to reconcile philosophy with Darwin by making human beings' pursuit of the true and the good continuous with the activities of the lower animals - cultural evolution with biological evolution. Dewey criticized the Cartesian notion of the self as a substance which existed prior to language and acculturation, and substituted an account of the self as a product of social practices (an account developed further by George Herbert Mead).

Dewey, whose primary interests were in cultural, educational and political reform rather than in specifically philosophical problems (problems which he thought usually needed to be dissolved rather than solved), developed the implications of pragmatism for ethics and social philosophy. His ideas were central to American intellectual life throughout the first half of the twentieth century.

All three of the founding pragmatists combined a naturalistic, Darwinian view of human beings with a deep distrust of the problems which philosophy had inherited from Descartes, Hume and Kant. They hoped to save philosophy from metaphysical idealism, but also to save moral and religious ideals from empiricist or positivist scepticism. Their naturalism has been combined with an anti-foundationalist, holist account of meaning by Willard van Orman Quine, Hilary Putnam and Donald Davidson - philosophers of language who are often seen as belonging to the pragmatist tradition. That tradition also has affinities with the work of Thomas Kuhn and the later work of Ludwig Wittgenstein.

1 Classical pragmatism

Charles Sanders Peirce, William James and John Dewey - often referred to as the three 'classical pragmatists' - had very different philosophical concerns. Except for their shared opposition to the correspondence theory of truth, and to 'copy theories' of knowledge, their doctrines do not overlap extensively (see Truth, pragmatic theory of). Although each knew and respected the other two, they did not think of themselves as belonging to an organized, disciplined philosophical movement. Peirce thought of himself as a disciple of Kant, improving on Kant's doctrine of categories and his conception of logic. A practising mathematician and laboratory scientist, he was more interested in these areas of culture than were James or Dewey. James took neither Kant nor Hegel very seriously, but was far more interested in religion than either Peirce or Dewey. Dewey, deeply influenced by Hegel, was fiercely anti-Kantian. Education and politics, rather than science or religion, were at the centre of his thought.

Peirce was a brilliant, cryptic and prolific polymath, whose writings are very difficult to piece together into a coherent system. He is now best known as a pioneer in the theory of signs, and for work in logic and semantics contemporaneous with, and partially paralleling, that of Frege. Peirce's account of inquiry as a matter of practical problem-solving was complemented by his criticisms of the Cartesian (and empiricist) idea of 'immediate knowledge', and of the project of building knowledge on self-evident foundations (of either a rationalist or empiricist kind).

Peirce protested against James' appropriation of his ideas, for complex reasons to do with his obscure and idiosyncratic doctrine of 'Scotistic realism' - the reality of universals, considered as potentialities or dispositions. Peirce was more sympathetic to metaphysical idealism than James, and found James' version of pragmatism simplistic and reductionist. James himself, however, thought of pragmatism as a way of avoiding reductionism of all kinds, and as a counsel of tolerance. Particularly in his famous essay 'The Will to Believe' (1896), he attempted to reconcile science and religion by viewing both as instruments useful for distinct, non-conflicting purposes.

Although he viewed many metaphysical and theological disputes as, at best, exhibitions of the diversity of human temperament, James hoped to construct an alternative to the anti-religious, science-worshipping positivism of his day. He approvingly cited Giovanni Papini's description of pragmatism as 'like a corridor in a hotel. Innumerable chambers open out of it. In one you may find a man writing an atheistic volume; in the next someone on his knees praying for faith; in a third a chemist investigating a body's properties … they all own the corridor, and all must pass through it'. His point was that attention to the implications of beliefs for practice offered the only way to communicate across divisions between temperaments, academic disciplines and philosophical schools.

Dewey, in his early period, tried to bring Hegel together with evangelical Christianity. Although references to Christianity almost disappear from his writings around 1900, in a 1903 essay on Emerson he still looked forward to the development of 'a philosophy which religion has no call to chide, and which knows its friendship with science and with art'. The anti-positivist strain in classical pragmatism was at least as strong as its anti-metaphysical strain, and so James and Dewey found themselves attacked simultaneously from the empiricist left and from the idealist right - by Bertrand Russell as well as by F.H. Bradley. Both critics thought of the pragmatists as fuzzy and jejune thinkers. This sort of criticism was repeated later in the century by the disciples of Carnap, most of whom dismissed the classical pragmatists as lacking in precision and argumentative rigour.

James wrote a few remarkable essays on ethics - notably 'The Moral Philosopher and the Moral Life' (1891), in which, echoing Mill's *Utilitarianism*, he says that every desire and need has a *prima facie* right to be fulfilled, and that only some competing desire or need can provide a reason to leave it unsatisfied. But neither James nor Peirce attempted any systematic discussion of moral or political philosophy. Dewey, however, wrote extensively in this area throughout his life - from *Outlines of a Critical Theory of Ethics* (1891) to *Human Nature and Conduct* (1922) and *Theory of Valuation* (1939).

Dewey urged that we make no sharp distinction between moral deliberation and proposals for change in sociopolitical institutions, or in education (the last being a topic on which he wrote extensively, in books which had considerable impact on educational practice in many countries). He saw changes in individual attitudes, in public policies and in strategies of acculturation as three interlinked aspects of the gradual development of freer and more democratic communities, and of the better sort of human being who would develop within such communities. All of Dewey's books are permeated by the typically nineteenth-century conviction that human history is the story of expanding human freedom and by the hope of substituting a less professionalized, more politically-oriented conception of the philosopher's task for the Platonic conception of the philosopher as 'spectator of time and eternity'.

In *Reconstruction in Philosophy* (1920) he wrote that 'under disguise of dealing with ultimate reality, philosophy has been occupied with the precious values embedded in social traditions …. has sprung from a clash of social ends and from a conflict of inherited institutions with incompatible contemporary tendencies'. For him, the task of future philosophy was not to achieve new solutions to traditional problems, but to clarify 'men's ideas as to the social and moral strifes of their own day'. This conception of philosophy, which developed out of Hegel's and resembled Marx's (see Hegel,G.W.F.; Marx, K.), isolated Dewey (particularly after the rise of analytic philosophy) from colleagues who thought of their discipline as the study of narrower and more precise questions - questions that had remained substantially unchanged throughout human history.

2 Pragmatism after the linguistic turn

Peirce was one of the first philosophers to emphasize the importance of signs. 'The word or the sign which man uses *is* the man himself,' he wrote, '… my language is the sum total of myself; for the man is the thought'. But, with the exception of C.I. Lewis and Charles Morris, philosophers did not take Peirce's work on signs very seriously. Indeed, for decades Peirce remained largely unread: he had never published a philosophical book, and most of his articles were collected and republished only in the 1930s.

By that time philosophy in the English-speaking world was already in the process of being transformed by admirers of Frege, notably Carnap and Russell. These philosophers accomplished what Gustav Bergmann was to baptize 'the linguistic turn' in philosophy. They thought that it would be more fruitful, more likely to yield clear and convincing results, if philosophers were to discuss the structure of language rather than, as Locke and Kant had, the structure of the mind or of experience. The early analytic philosophers, however, accompanied this turn

with a revival of the traditional empiricist idea that sense-perception provides foundations for empirical knowledge - an idea which, at the beginning of the century, the idealists and the classical pragmatists had united in rejecting. These philosophers also insisted on a strict distinction between conceptual questions (the analogue of Kant's 'transcendental' questions), now reinterpreted as questions about the meaning of linguistic expressions, and empirical questions of fact.

It was not until that distinction was questioned by Willard van Orman Quine in his groundbreaking 'Two Dogmas of Empiricism' (1951) that pragmatism was able once again to obtain a hearing (see Quine, W.V. §8). James and Dewey had been viewed during the heyday of logical positivism as having prefigured the logical positivist's verifiability criterion of empirical meaningfulness, but as unfortunately lacking the powerful analytic tools which the new logic had made available. However, Quine's suggestion that empirical observation of linguistic behaviour could not detect a difference between necessary, analytic truths and contingent, synthetic, yet unquestioned truths helped revive the pragmatists' combination of holism, anti-foundationalism and naturalism.

That suggestion was reinforced by other publications which were roughly simultaneous with Quine's. In *Philosophical Investigations* (1953), Ludwig Wittgenstein mocked the idea that logic is both 'something sublime' and the essence of philosophy, an idea which the younger Wittgenstein had shared with Russell (see Wittgenstein, L. §8). That book also reinvigorated the pragmatists' claim that most philosophical problems should be dissolved rather than solved. Wilfrid Sellars' 'Empiricism and the Philosophy of Mind' (1953) renewed both Peirce's assault on the idea of 'immediate experience' and his claim that the intentionality of the mental is derived from the intentionality of the linguistic, rather than conversely (see Sellars, W.). In America, this article had the same devastating effect on the notion of 'sense-datum', and thus on the empiricist roots of logical positivism, that J.L. Austin's work was simultaneously having in Britain (see Austin, J.L.). The work of Sellars and Austin conspired to deprive empiricism of the prestige which it had traditionally enjoyed in the Anglophone philosophical world.

Somewhat later, Thomas Kuhn's *The Structure of Scientific Revolutions* (1962) broke the grip of the positivist notion that natural science, because it offered paradigmatically rational methods and procedures, should be imitated by the rest of culture (see Kuhn, T.S.). The effect of these various anti-empiricist and anti-positivist writings was to make many post-positivistic analytic philosophers sympathetic to Dewey's suspicions of the Cartesian-Kantian problematic of modern philosophy. Hilary Putnam, the best-known contemporary philosopher to identify himself as a pragmatist, has written appreciatively about all three classical pragmatists, praising their refusal to distinguish 'the world as it is in itself' from the world as it appears in the light of human needs and interests.

On Putnam's account, ' the heart of pragmatism…was the insistence on the agent point of view. If we find that we must take a certain point of view, use a certain ''conceptual system", when we are engaged in practical activity…then we must not simultaneously advance the claim that it is not really the way things are in themselves' (1987). Putnam holds that our moral judgments are no more and no less 'objective' than our scientific theories, and no more and no less rationally adopted. He agrees with Dewey that the positivists' attempt to separate 'fact' from 'value' is as hopeless as their pre-Quinean attempt to separate 'fact' from 'language'.

Putnam has also come to the defence of the most notorious and controversial of the classical pragmatists' doctrines: the so-called 'pragmatist theory of truth'. Peirce said 'the opinion which is fated to be ultimately agreed to by all who investigate is what we mean by the truth, and the object represented in this opinion is the real'. Putnam has revived this idea, arguing that even if we cannot follow Peirce in *defining* 'true' as 'idealized rational assertibility', the latter notion is, as a regulative ideal, inseparable from an understanding of the concept of truth. He has criticized the correspondence theory of truth by arguing that any such correspondence of a belief to reality can only be to reality under a particular description, and that no such description is ontologically or epistemologically privileged. Putnam follows Nelson Goodman in saying that 'there is no one Way the World Is'.

3 Pragmatism as anti-representationalism

Putnam is chary, however, of endorsing James' claim that '"The true"…is only the expedient in the way of our thinking, as "the right" is only the expedient in our way of behaving'. That formulation was attacked by James' contemporaries as at worst an invitation to self-deception, and at best a confusion of truth with justifiability. Dewey tried to avoid the controversy by ceasing to use the word 'truth', and speaking instead of 'warranted

assertibility'. But this did not shield him from charges of confusion and inconsistency. Russell, reviewing Dewey, said that 'there is a profound instinct in me which is repelled by [Dewey's] instrumentalism: the instinct of contemplation, and of escape from one's own personality'. He and many other critics complained that pragmatism is unable to take account of the eternity and absoluteness of truth - of the fact that a sentence that contains no demonstratives is, if true, true in utter independence of changes in human needs or purposes. Putnam's treatment of truth is designed to avoid the appearance of relativism, and to escape such strictures as Russell's.

Despite its paradoxical air and its apparent relativism, however, James' claim does bring out pragmatism's strongest point: its refusal to countenance a discontinuity between human abilities and those of other animals. Pragmatists are committed to taking Darwin seriously. They grant that human beings are unique in the animal kingdom in having language, but they urge that language be understood as a tool rather than as a picture. A species' gradual development of language is as readily explicable in Darwinian terms as its gradual development of spears or pots, but it is harder to explain how a species could have acquired the ability to *represent* the universe - especially the universe as it really is (as opposed to how it is usefully described, relative to the particular needs of that species).

In a weak sense of 'represent', of course, an earthworm or a thermostat can be said to contain 'representations of the environment', since there are internal arrangements in both which are responsible for the reactions of each to certain stimuli. But it makes little sense to ask whether those representations are *accurate*. Philosophers who take epistemological scepticism seriously (as pragmatists do not) have employed a stronger sense of 'representation', one in which it does make sense to ask whether the way in which it best suits human purposes to describe the universe is an accurate representation of the universe as it is in itself (see Scepticism).

The idea that knowledge is accurate representation and the idea that reality has an intrinsic nature are inseparable, and pragmatists reject both. In rejecting these ideas pragmatists are rejecting the problematic of realism and antirealism - the question of whether there is or is not a 'matter of fact' about, for example, mathematics or ethics, whether beliefs in these areas are attempts to correspond to reality. Whatever may be said about truth, pragmatists insist, we cannot make sense of the notion of 'correspondence', nor of that of 'accurate representation of the way things are in themselves' (see Truth, correspondence theory of).

Donald Davidson is the philosopher of language whose work is most reminiscent of the classical pragmatists' attempts to be faithful to Darwin. Davidson has said that 'Beliefs are true or false, but they represent nothing. It is good to be rid of representations, and with them the correspondence theory of truth, for it is thinking that there are representations that engenders thoughts of relativism' (1989). He has argued that we need to get rid of what he calls 'the third dogma of empiricism', the distinction between the mind or language as organizing scheme, and something else (for example, the sensible manifold, the world) as organized content - the Kantian version of the dualism of subject and object (1974). In 'A Nice Derangement of Epitaphs' (1986), an attempt to radicalize and extend Quine's naturalistic approach to the study of linguistic behaviour, he has suggested that we 'erase the boundary between knowing a language and knowing our way about in the world generally', and that 'there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed'.

Davidson does not wish to be called a pragmatist, however, since he equates pragmatism with unfeasible attempts to reduce truth to some form of assertibility, thereby making it an epistemic concept, rather than a merely semantic one. Unlike Peirce and Putnam, Davidson thinks that we should treat 'true' as a primitive term, and should neither attempt to revitalize the correspondence theory of truth nor replace it with a better theory of truth. Davidson's strategy is summed up in his recommendation that we not say 'that truth is correspondence, coherence, warranted assertibility, ideally justified assertibility, what is accepted in the conversation of the right people, what science will end up maintaining, what explains the convergence on single theories in science, or the success of our ordinary beliefs' (1990). We should, he says in the same article, not offer an analysis of the meaning of 'true', but rather confine ourselves to describing 'the ultimate source of both objectivity and communication', namely, 'the triangle that, by relating speaker, interpreter and the world determines the contents of thought and speech'. The trouble with the correspondence theory, on Davidson's view, is that it cuts out the 'interpreter' side of the triangle, and treats truth as relation of 'matching' between speaker and world.

If one follows Davidson's advice, one can give up the pragmatist theory of truth without giving up the Darwinian naturalism which that theory was a paradoxical-sounding attempt to articulate. Such naturalism, however, entails

an abandonment of much of the problematic of contemporary philosophy. If truth is never the name of a relation (‘corresponding’, ‘representing’, ‘getting right’, ‘fitting’) which holds between sentences and non-sentences, there is no point in asking whether this relation holds for some true sentences (for example, perceptual reports or scientific theories) and not for others (for example, sentences about numbers or values). On this latter point, Putnam and Davidson are in agreement (see Truth, correspondence theory of).

Michael Dummett has suggested, plausibly, that the problematic of realism and antirealism is at the heart of the Western philosophical tradition (see Realism and antirealism). If he is right, and if Davidson is right in thinking that we should now abandon that problematic, then James’ and Dewey’s suggestions about how to end the traditional and seemingly sterile quarrels between materialists and idealists, positivists and metaphysicians, theists and atheists, science-worshippers and poetry-worshippers look more promising. The heart of both men’s pragmatism was not any particular doctrine about the nature of truth, of knowledge, or of value, but rather the hope that philosophy could renew itself by moving out from under traditional dualisms (subject-object, mind-world, theory-practice, morality-prudence) which recent science and recent social changes had, they believed, rendered obsolete.

The classical pragmatists saw themselves as responding to Darwin in the same way as the great philosophers of the seventeenth and eighteenth centuries had responded to Galileo and Newton. Philosophers such as Descartes, Locke and Kant attempted to accommodate old, precious, moral and spiritual aspirations to new scientific developments. James and Dewey thought that these attempts had been made obsolete by Darwin’s new account of the origin of our species, and that fresh attempts were needed. If one reads Quine’s and Davidson’s naturalization of semantics as a continuation of philosophy’s attempt to come to terms with Darwin, one can also read these two philosophers as continuing the larger enterprise which James and Dewey inaugurated.

4 Pragmatism and humanity’s self-image

By stepping back from its relation to traditional empiricism on the one hand and to the linguistic turn on the other, one can put pragmatism in a larger context. Much twentieth-century philosophy has been devoted to a criticism of the view, shared by Plato and Aristotle, that a capacity to know things as they really are is central to being human. Philosophers influenced by Nietzsche - notably Heidegger, Sartre and Derrida - have argued against the idea that cognition is the distinctively human capacity. Heidegger’s treatment of inquiry as a species of coping, in his discussion of *Vorhandenheit* in *Being and Time* (1927), has much in common with Dewey’s and Kuhn’s attempts to see scientific progress as problem-solving - as the overcoming of obstacles to the satisfaction of human needs, rather than as convergence towards a special, specifically cognitive, relation to reality. Both Dewey and Heidegger saw the Greek quest for certainty as debilitating. Neither granted the traditional assumption that, in addition to all the other needs human beings have, there is a need to know the truth (see Heidegger, M.).

Heidegger’s criticism of what he called ‘onto-theology’ - Western philosophy viewed as a series of attempts to find solace and support in the non-temporal - has much in common with Dewey’s criticism of what he called ‘intellectualism’. Both of these men saw the tradition which begins with Plato as a self-deceptive attempt to give the eternal priority over the temporal. So did Bergson and Whitehead, the founders of the tradition known as ‘process philosophy’, a tradition to which James (especially in his *Essays in Radical Empiricism*) made important contributions (see Process philosophy). This downgrading of the eternal is characteristic of a great deal of twentieth-century philosophy. It is found in James’ criticisms of Bradley, in Putnam’s criticism of Bernard Williams’ claim that we can use an ‘absolute conception of the world’ as a regulative ideal of inquiry, in Heidegger’s criticism of Husserl, and in Derrida’s criticism of Heidegger.

Downgrading eternity means downgrading both the idea of truth as eternal and the assumption that knowledge of eternal truth is the distinctively human activity. From a Davidsonian, as from a Deweyan, point of view, the only point of the doctrine that truth is eternal is to contrast truth with justification (which is obviously neither eternal nor absolute, because it is relative to the composition of the audience to which justification is offered, and thus to historical circumstance). But that contrast can be formulated without treating ‘truth’ as the name of a goal to be reached, or of an object to be admired. Davidson’s treatment of truth forbids us to think of inquiry as subject to a norm of acquiring true beliefs, in addition to the norm of providing adequate justification. There is no way to seek for truth apart from seeking for justification. Justification gets better as the community to which justification is offered becomes more sophisticated and complex, more aware of possible sources of evidence and more capable of

dreaming up imaginative new hypotheses and proposals. So pragmatists place the capacity to create complex and imaginative communities at the centre of their image of humanity, superseding the ability to know. Dewey and Putnam agree that the aim of inquiry is what Putnam calls 'human flourishing' - the kind of human life which is possible in free, democratic, tolerant, egalitarian societies. These are the societies in which the arts and the sciences proliferate and progress, and within which idiosyncrasy is tolerated.

The obvious difference between James, Dewey and Putnam on the one hand and Nietzsche, Heidegger and Foucault on the other - between the two most prominent sections of the twentieth-century revolt against the Greek self-image of humanity - is that these three Europeans do not share the Americans' enthusiasm for, and optimism about, liberal-democratic society. Nietzsche's, and the early Heidegger's, insistence on the resolute authenticity of the lonely individual, and their exaltation of will as opposed to intellect, are equally foreign to Dewey and to Putnam (though they have some echoes in certain passages of James). Rather than replace intellect by will, in the manner of Schopenhauer, pragmatists tend to replace knowledge by love, in the manner of Kierkegaard's contrast between Socrates and Christ (see Kierkegaard, S.A.).

For Dewey, the pragmatist who speculated most daringly, and developed the greatest historical self-consciousness, the glory of human beings is their ability to become citizens of a liberal-democratic society, of a community which constantly strives to see beyond its own limits - both with an eye to the inclusion of presently excluded or marginalized human beings and with respect to innovative intellectual and artistic initiatives. This is the capacity which most clearly sets us apart from other animals. It presupposes, of course, the capacity to use language, but for Dewey the point of having language, and therefore thought, was not to penetrate through the appearances to the true nature of reality, but rather to permit the social construction of new realities. For him, language was not a medium of representation, but a way of coordinating human activities so as to enlarge the range of human possibilities. These processes of coordination and enlargement, which make up cultural evolution, do not have a destined terminus called the Good or the True, any more than biological evolution has a destined terminus called The Ideal Life-Form. Dewey's imagery is always of proliferating novelty, rather than of convergence.

The naturalist strain in pragmatism, the attempt to come to terms with Darwin, is thus from a Deweyan point of view important mainly as a further strategy for shifting philosophers' attention from the problems of metaphysics and epistemology to the needs of democratic politics. Dewey once said that he agreed with Plato that politics was 'the science of the whole', a remark which summarized the following train of reasoning. Finding out what there is is a matter of finding out what descriptions of things will best fulfil our needs. Finding out what needs we should fulfil is a task for communal reflection about what human beings might become. Such cooperative inquiry into the possibilities of self-transcendence is best accomplished within a democratic society. So philosophers should stop asking about the nature of reality or of knowledge, and instead try to strengthen and improve the institutions of such societies by clarifying 'men's ideas as to the social and moral strifes of their own day'.

See also: Darwin, C.; Doubt; Empiricism; Logical positivism; Pragmatism in ethics; Scientific realism and antirealism

RICHARD RORTY

References and further reading

Ayer, A.J. (1968) *The Origins of Pragmatism: Studies in the Philosophy of Charles Sanders Peirce and William James*, San Francisco, CA: Freeman, Cooper & Co.(A somewhat unsympathetic examination, by a disciple of Carnap, of selected aspects of classical pragmatism.)

Brandom, R. (1994) *Making It Explicit: Reasoning, Representing and Discursive Commitment*, Cambridge, MA: Harvard University Press, 1994.(Chapter 1 sketches the similarities between the pragmatist view of meaning and Wittgenstein's. Chapter 5 contains a strikingly original reinterpretation of the point of pragmatist theories of truth, and the formulation of a new theory along similar lines.)

Davidson, D. (1986) 'A Nice Derangement of Epitaphs' in E. LePore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell, 433-46.(Attacks the idea that a language is a set of conventions, or something possessing an isolable structure. See the responses to this essay by Michael Dummett and Ian Hacking, in the same volume.)

Davidson, D. (1974) 'On the Very Idea of a Conceptual Scheme', repr. in *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1984.(Attacks 'the third, and perhaps last, dogma of empiricism: the distinction between scheme and content'.)

Davidson, D. (1989) 'The Myth of the Subjective' in M. Krausz (ed.) *Relativism: Interpretation and Confrontation*, Notre Dame, IN: University of Notre Dame Press, 159-72.(A criticism of the subject-object analysis of knowledge.)

Davidson, D. (1990) 'The Structure and Content of Truth', *Journal of Philosophy* 87: 279-28.(Argues that 'triangulation' between speaker, audience and world is required for intentionality.)

Dewey, J. (1882-1953) *The Early Works, The Middle Works, The Later Works*, Carbondale, IL: Southern Illinois University Press, 1969-90.(Dewey's view of the nature and function of philosophy is best summarized in *Reconstruction in Philosophy* (1920; vol. 12 of *The Middle Works*), his ethical views are most fully laid out in *Human Nature and Conduct* (1922; vol. 14 of *The Middle Works*) and his understanding of democracy in *The Public and Its Problems* (1927; vol. 2 of *The Later Works*).)

Dewey, J. (1903) 'Emerson - the Philosopher of Democracy', *The Middle Works*, vol. 3, 184-92.(This essay of 1909 describes Emerson as 'the first and as yet almost the only Christian of the Intellect'.)

Diggins, J.P. (1994) *The Promise of Pragmatism: Modernism and the Crisis of Knowledge and Authority*, Chicago, IL: University of Chicago Press.(A survey of uses of, and reactions to, the classical pragmatists' thought by twentieth-century social and cultural critics in the US.)

James, W. (1890) *The Principles of Psychology*, Cambridge, MA: Harvard University Press, 1981.(James said that this thousand-page treatise, originally published in 1890, 'rejects both the associationist and the spiritualist theories [of mental functioning]'.)

James, W. (1897) *The Will to Believe and Other Essays In Popular Philosophy*, Cambridge, MA: Harvard University Press, 1979.(The title essay (1896) defends the right to be religious in despite of science, and is James' best-known and most controversial publication. This volume also includes 'The Moral Philosopher and the Moral Life', which argues for 'the impossibility of an abstract system of ethics'.)

James, W. (1907, 1909) *Pragmatism and The Meaning of Truth*, Cambridge, MA: Harvard University Press, 1979; reprint of two books in one volume.(*Pragmatism* (1907) is the classic statement of James' overall philosophical outlook. The essays collected in *The Meaning of Truth* (1909) - of which 'Humanism and Truth' and 'The Pragmatist Theory of Truth and its Misunderstanders' are the most incisive - defend his pragmatic theory of truth against its critics.)

James, W. (1912) *Essays in Radical Empiricism*, Cambridge, MA: Harvard University Press, 1976.(Defends a kind of neutral monism, especially in the essays 'Does "Consciousness" Exist?' and 'A World of Pure Experience'.)

Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(Initiated the reaction against the attempt to isolate a 'scientific method' which differentiated science from other areas of culture.)

Lewis, C.I. (1923) 'A Pragmatic Conception of the A Priori', *Journal of Philosophy* 20: 169-77.(A very influential article, which attempted to blend classical pragmatism with Kant. Lewis acted as intermediary between classical pragmatism and logical empiricism.)

Murphy, J.P. (1990) *Pragmatism: From Peirce to Davidson*, Boulder, CO: Westview Press.(An introductory textbook. Murphy sees Quine and Davidson as continuing the pragmatist tradition. Contains a substantial bibliography.)

Okrent, M. (1988) *Heidegger's Pragmatism*, Ithaca, NY, and London: Cornell University Press.(Part One of this book discusses sections 12-24 of *Being and Time*, in which Heidegger's anti-Cartesian arguments parallel those of the classical pragmatists.)

Quine, W.V. (1951) 'Two Dogmas of Empiricism' in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1953.(A groundbreaking essay which initiated the 'post-positivistic' period of analytic philosophy.)

Peirce, C.S. (1934) *Pragmatism and Pragmaticism*, vol. 5 of *Collected Papers*, ed. C. Hartshorne and P. Weiss, Cambridge, MA: Harvard University Press.(See especially the anti-Cartesian polemic in 'Questions Concerning Certain Faculties Claimed for Man' (1868) and 'Some Consequences of Four Incapacities' (1868), and also the first explicit formulation of Peirce's pragmatism in 'How to Make our Ideas Clear' (l878). A new edition of Peirce's papers is now in preparation.)

Putnam, H. (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press.(Chapter 8, 'The impact of science on modern conceptions of rationality', is an important statement of pragmatism's criticisms of logical empiricism.)

Putnam, H. (1987) *The Many Faces of Pragmatism*, La Salle, IL: Open Court, 83.(Collects essays in which Putnam offers pragmatist solutions of various problems.)

Putnam, H. (1990) *Realism with a Human Face*, Cambridge, MA: Harvard University Press. (A collection of

essays, replying to critics of pragmatism, and relating Putnam's own work to James', Dewey's, and Goodman's.)

Putnam, H. (1992) *Renewing Philosophy*, Cambridge, MA: Harvard University Press. (Chapter 5, 'Bernard Williams and the absolute conception of the world' criticizes the notion of a description of reality which abstracts from human needs and interests.)

Rorty, R. (1991) *Objectivity, Relativism and Truth*, Cambridge: Cambridge University Press.(Part II of this book tries to fit Davidson's work into the pragmatist tradition by emphasizing his anti-representationalism.)

Saatkamp, H.J., Jr (ed.) (1995) *Rorty and the Pragmatists*, Nashville, TN: Vanderbilt University Press.(Includes five essays arguing that Rorty misreads and distorts both the classical pragmatists and Davidson, and Rorty's replies.)

Scheffler, I. (1974) *Four Pragmatists: A Critical Introduction to Peirce, James, Mead and Dewey*, London: Routledge & Kegan Paul.(Combines sympathetic presentation with detailed criticism.)

Sellars, W. (1953) 'Empiricism and the Philosophy of Mind', in *Science, Perception and Reality*, London: Routledge & Kegan Paul, 1956.(Argues against the 'Myth of the Given' and for the claim that 'all awareness is a linguistic affair'.)

Smith, J.E. (1984) *Purpose and Thought: The Meaning of Pragmatism*, Chicago, IL: University of Chicago Press. (Emphasizes the classical pragmatists' anti-atomistic conceptions of the nature of experience.)

Thayer, H.S. (1968) *Meaning and Action: A Critical History of Pragmatism*, Indianapolis, IN: Bobbs-Merrill.(The most comprehensive history, containing material on C.I. Lewis, F.C.S. Schiller, G.H. Mead, and others, as well as on Peirce, James and Dewey.)

West, C. (1989) *The American Evasion of Philosophy: A Genealogy of Pragmatism*, Madison, WI: University of Wisconsin Press.(Puts pragmatism in the context of American intellectual life, with special reference to Emerson and to leftist politics.)

White, M.G. (1954) *Towards Reunion in Philosophy*, Cambridge, MA: Harvard University Press.(Treats pragmatism as a corrective to logical positivism.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(Argues for a therapeutic approach to traditional philosophical problems, and against the Cartesian notion of immediate self-awareness.)

Pragmatism in ethics

Two components of the pragmatist outlook shape its ethical philosophy. It rejects certainty as a legitimate intellectual goal; this generates a nondogmatic attitude to moral precepts and principles. It holds, secondly, that thought (even that exercised in scientific inquiry) is essentially goal-directed in a way that makes the refinement of the control we exercise over how we act (for example, in drawing conclusions) integral to achieving any cognitive goal such as that of truth. This makes it possible to treat scientific inquiry as a model of how we might respond to moral problems and the reasonableness and impartiality required of a scientific inquirer as a paradigm of what may be expected in reaching moral judgments. This view of the nature of thought also inclines pragmatists to assess proposed solutions to moral conflicts in terms of consequences. But although human desires are taken as the raw material with which moral thinking must deal, it is not assumed that people's desires (what they take pleasure in) are fixed and can be used as a standard by which to assess consequences. Pragmatism is thus free to revert to a classical mode of thought (such as Aristotelianism) in which claims about human nature function as norms - a use which is made, for example, of the claim that humans are essentially social creatures.

1 James: Fallibilism

In 'Philosophical Conceptions and Practical Results' (1898), William James drew the attention of the public to a distinctive philosophic outlook which C.S. Peirce had begun to articulate in published articles twenty years earlier. Pragmatism, as Peirce had come to label his doctrines, located truth and reality at the limit of an indefinitely prolonged process of inquiry and insisted that nothing on this side of that limit could be known with absolute certainty - an argument known as 'fallibilism' (see Fallibilism). Pragmatism had the effect of undermining any claim to absolute and final cognitive authority.

James had already applied this attitude to 'ethical science' in an address, 'The Moral Philosopher and the Moral Life' (1891). Here James insisted that, as in 'physical science', we must be ready to revise from day to day the basis on which we measure and compare various goods and ills; 'no philosophy of ethics is possible in the old-fashioned absolutist sense of the term' (1891: 208). Ethical science could not in James' view aspire to delineate 'any abstract moral "nature of things" existing antecedently to the concrete thinkers themselves' (193). Good and ill were brought into the world by the feelings and opinions of conscious beings (190-1); obligation arose through the claims such beings made on one another (194); and in the best world imaginable '*every* demand' would be 'gratified as soon as made' (202; original emphasis).

As this is clearly not possible in the world in which we live, the task is to 'satisfy at all times *as many demands as we can*' (205; original emphasis). This task is carried out, however imperfectly, by following custom and judging by what is conventionally recognized as good. But although in cases of conflict there is a presumption in favour of what any given society has discovered to be a viable relative equilibrium - discoveries indeed 'quite analogous to those of science' (205) - it was clear to James 'that there is nothing final in any actually given equilibrium of human ideals' (206). "Rules are made for man, not man for rules" (James quoting T.H. Green, 206) and 'the *highest* ethical life… consists at all times in the breaking of rules which have grown too narrow for the actual case' (209; original emphasis).

To anyone who believes ethical principles are independent of human desire, perhaps even independent of God's will, this position would appear very radical. James however moved to keep his position within prevailing attitudes concerning the importance of religion to morality by declaring that God's standing as a conscious being with extraordinary attributes gave his demands a special importance in the project of 'satisfying as many demands as we can'. Even unbelievers have to recognize how belief in God inclines a person to adopt a 'strenuous mood' towards the realization of ideals; 'in a merely human world without a God, the appeal to our moral energy falls short of its maximal stimulating power' (212). James moreover hedged his radical stance on two sides. Ethical scepticism was not to be regarded as 'being one possible fruit of ethical philosophizing' (184) (see Moral scepticism). To give up the task of weaving conflicting demands into a more satisfactory system was to renounce moral philosophy. On the other hand it is incumbent on those who undertake philosophical reflection on moral principles not to apply a 'wayward personal standard' but to 'throw [their] own spontaneous ideals, even the dearest, impartially in with that total mass of ideals which are fairly to be judged' (199).

2 Peirce: Reasonableness

That we might impartially undertake the project of achieving 'ethical unity' - especially while acknowledging that the universe cannot be called upon to underwrite the hope that there is one correct outcome - is a notion that faces considerable resistance. It is particularly unappealing for anyone who holds that impartiality involves the kind of objectivity that consists in making thought conform to something antecedent to, and independent of, all thought and desire. The further suggestion that this project might have an intellectual standing comparable to that of natural science appears in this light (that of twentieth-century scientific realism) quite bizarre. In the nineteenth century, however, when it was not uncommon to regard the goal of natural scientists as consisting in an ideal refinement of (human) scientific thought rather than in the conformity of thought to an independent object, these suggestions appeared less strange (see Scientific realism and anti-realism §1).

It was in the context of such an outlook that C.S. Peirce first formulated his views on truth and reality. In an 1871 review of an edition of the works of Berkeley, Peirce distinguished two accounts of 'the real'. The first was a 'thing out[side] of the mind which directly influences sensation, and through sensation thought'; the second was what would be represented in thought if we were to prolong indefinitely the process of removing from our thought any 'arbitrary, accidental element, dependent on the limitations in circumstances, power and bent of the individual; an[y] element of error, in short' (1958: 8.12). As Peirce's thought developed he gave considerable attention to the problem of explaining how experience should constrain this development without capitulating to the first of the above views.

It might have been interesting had Peirce considered what would be involved in refining desires or demands (individuals' perceptions of their interests) through a process of inquiry, and how this project should be constrained by experience. Peirce, however, devoted little attention to the place of ethics in relation to other intellectual endeavours until the final period of his life, by which time James had made Peirce's brainchild, pragmatism, a topic of widespread and lively controversy. In papers written after 1900 (see 1958: 1.573-677) Peirce placed the values and motives that should inform inquiry in the context of wider claims, 'The only desirable object which is quite satisfactory in itself without any ulterior reason for desiring it, is the reasonable itself' (1958: 8.140), and he speculated that 'there is an energizing reasonableness that shapes phenomena in some sense… and has moulded the reason of man into something like its own image' (1966: 291). Peirce also stressed that his own primary field of endeavour, logic, which he conceived broadly as including scientific methodology, was a 'normative science'. This meant that its principles were to guide the activity of deliberate thinking (especially that in natural science) in the way that ethical principles guided action. 'I regard Logic as the Ethics of the Intellect' (1966: 415); 'logic is only an application of morality', he wrote to Lady Welby in 1908-9 (1966: 406).

3 Dewey: Consequentialism

Peirce's conception of thought as a form of human conduct was, along with fallibilism, a central feature of the outlook of those who identified themselves as pragmatists during its initial period. This conception moreover functioned within a general view of mental activity as essentially goal-directed. In his *Principles of Psychology* (1890), James had taken '*the pursuance of future ends and the choice of means for their attainment [to be] the mark and criterion of the presence of mentality in a phenomenon*' (8; original emphasis) and treated a concept in general as 'really nothing but a teleological instrument' (482). During the 1890s, under the acknowledged influence of James, John Dewey developed a position which he called 'instrumentalism' and which he was content to have identified as a form of pragmatism once there emerged a movement of that name.

Dewey's instrumentalism began with James' view, that representations (such as symbols, concepts and statements) should be treated as instruments and assessed relative to the purposes which they served, and developed from this a general account of problem-solving which characterized the aim of inquiry not in terms of discovering truth but in terms of unifying a discordant situation. The fullest version of this account is found in the sixth chapter of Dewey's *Logic: The Theory of Inquiry* (1938). This account of inquiry not only had the effect of insisting that theoretical questions be located in practical contexts, but was framed (some scholars believe primarily framed) with moral problems in mind. What James describes as the ideal resolution to a moral conflict, ('*Invent some manner* of realizing your own ideal which will also satisfy the alien demands' (1891: 205)) was for Dewey a better paradigm of the imperative of inquiry than 'Find out what is the case'.

It is important, however, to qualify this, for Dewey acknowledged to a greater extent than did James that the best resolution of a practical problem may lie in the modification of one's demands. Dewey stressed how much of moral education, as well as how much of the aim of moral reflection, consisted in making people mindful of the consequences of their conduct for themselves and for other people. The difference between 'reasonable and unreasonable desires and interests' (1939: 217) and that between the 'desirable' and the 'desired' (219) consists for Dewey in the extent to which a desire is informed by an understanding of the consequences of having it satisfied. Moral psychology commonly proceeds on the assumption that our desires are given and fixed, but 'nothing more contrary to common sense can be imagined than the notion that we are incapable of changing our desires and interests by means of learning what the consequence of acting upon them are' (218).

As a result of this outlook Dewey argued that moral thinking and thinking about values in general is structured exhaustively by the relations of means-ends and means-consequences; and he moved to subvert the widely accepted distinctions between final and instrumental values and intrinsic and extrinsic goods (1938: 214). No means adopted to achieve an end is so completely extrinsic to the experience of the people who use it that its contribution to their lives may be ignored. Nothing is so valuable in and of itself that the consequences of realizing it (including especially the consequences of whatever means are adopted towards its realization) do not need to be carefully weighed. The doctrine that there are purely final ends was for Dewey a projection of the belief that there are desires which may be taken as original and unalterable.

The claim that nothing should be regarded as a final end was intended by Dewey to prevent any component or aspect of human life from being sought in isolation from its bearing on the rest of experience. The notion of a final end also applies, however, to global features of the totality of human endeavours. One pattern of choices of means to realize our natural impulses might require more cooperative activities and move us towards greater social dependence on one another; another might require more self-reliance and move us towards less integrated ways of living with one another. Either of these global outcomes might appeal to people as final ends or intrinsic values. Questions that may be framed when the notion of final end is used in this sense can be avoided only if Dewey has some basis for thinking that a complete working out of all the means available to pursue our partially thought-out impulses, together with a complete working out of the consequences of each of the means we might adopt, will leave us with precisely one global option. Dewey did not argue explicitly for this conclusion but at the same time he did not hesitate to endorse the value of shared experience (1925: 157), and the ideal of community, as supreme (1927: 325-50).

4 Mead: Sociality

Important aspects of Dewey's philosophy were undeniably shaped by his relationship with his friend and sometime colleague George Herbert Mead. Mead developed a theory of the formation of the self and of rational thought which stressed the dependence of both on human sociality. What emerges in Mead's ethical writings is a more explicit acknowledgement than is made by Dewey that moral choices ultimately have to be assessed relative to a claim, at once factual and normative, about human nature. 'As human nature is essentially social in character, moral ends must be also social in their nature' (Mead 1934: 385). Dewey and Mead both wrote of the modification of desire as part of what was involved in developing a larger (broader) as opposed to a narrower self; 'the matter of selfishness is the setting-up of a narrow self over against a larger self' (1934: 388; compare Dewey and Tufts 1932: 302). But it was Mead who grounded the preference for the larger self in the claim about human nature. 'The escape from selfishness is not by the Kantian road of an emotional response to the abstract universal, but by the recognition of the genuinely social character of human nature' (1964: 220).

On this basis Dewey and Mead both identified themselves with movements for social reform, and Dewey achieved a degree of prominence in public life as a spokesman for a variety of causes on what was perceived as the left of American politics. As Dewey moved more into public life, logical positivism came to eclipse the influence of pragmatism in American universities. Positivism insinuated the view that truth and objectivity consist in the conformity of thought to an independent and antecedently existing reality (see Logical positivism §1). Ethical scepticism became a live philosophic option and was promoted via the claim that statements expressing moral judgments could be nothing more than expressions of subjective feeling (see Emotivism; Moral scepticism §3). Even after the vogue for positivism waned, pragmatism remained unfashionable until about 1980 when it enjoyed a modest revival under the influence of Richard Rorty. 'Neo-pragmatism' has yet to define its position on matters of

ethical theory, but its tendency to stress the values of individual and private, as opposed to communal and public, life suggests that its position is likely to represent a considerable departure at least from the pragmatism of Dewey and Mead.

See also: Axiology; Pragmatism

J.E. TILES

References and further reading

Dewey, J. (1925) *Experience and Nature*, Chicago, IL, and London: Open Court; repr. in *John Dewey: The Later Works 1925-1953*, ed. J.A. Boydston, Carbondale, IL: Southern Illinois University Press, 1981-90, vol. 1.(A mature comprehensive statement of Dewey's philosophy.)

Dewey, J. (1927) *The Public and Its Problems*, New York: Holt; repr. in *John Dewey: The Later Works 1925-1953*, ed. J.A. Boydston, Carbondale, IL: Southern Illinois University Press, 1981-90, vol. 2.(The central text for Dewey's political theory.)

Dewey, J. (1938) *Logic: The Theory of Inquiry*, New York: Holt; repr. in *John Dewey: The Later Works 1925-1953*, ed. J.A. Boydston, Carbondale, IL: Southern Illinois University Press, 1981-90, vol. 12.(Dewey's attempt to develop a logical theory where thought is conceived of as constrained by the need to resolve some kind of difficulty.)

Dewey, J. (1939) 'The Theory of Valuation', in *Encyclopedia of Unified Science*, vol. II, Chicago, IL: University of Chicago Press; repr. in *John Dewey: The Later Works 1925-1953*, ed. J.A. Boydston, Carbondale, IL: Southern Illinois University Press, 1981-90, vol. 13.(A monograph arguing against emotivism and similar positivist trends.)

Dewey, J. and Tufts, J. (1932) *Ethics*, New York: Holt; repr. in *John Dewey: The Later Works 1925-1953*, ed. J.A. Boydston, Carbondale, IL: Southern Illinois University Press, 1981-90, vol. 7.(A popular textbook which gives an account of the values of the moral life.)

Feffer, A. (1993) *Chicago Pragmatism and American Progressivism*, Ithaca, NY: Cornell University Press.(Places the thought of Dewey and Mead in its historical context.)

Gouinlock, J. (1972) *John Dewey's Philosophy of Value*, New York: Humanities Press.(Comprehensive philosophical background to Dewey's general ethical theory.)

James, W. (1890) *The Principles of Psychology*, vol. 1, New York: Dover, 1950. (The basis of the philosophical psychology used by James, Dewey and Mead.)

James, W. (1891) 'The Moral Philosopher and the Moral Life', in *The Will to Believe and Other Essays on Popular Philosophy*, New York: Dover, 1956, 184-215.(James' single most sustained treatment of ethics.)

James, W. (1898) 'Philosophical Conceptions and Practical Results', *University Chronicle*, Berkeley, CA: University of California Press; repr. in *Collected Essays and Reviews*, ed. R.B. Perry, New York: Longmans, Green, 1920, 406-37.(The lecture in which James, in discussing and applying Peirce's ideas, launched pragmatism as a philosophical movement.)

Joas, H. (1985) *G.H. Mead: A Contemporary Re-examination of his Thought*, trans R. Meyer, Cambridge, MA: MIT Press, ch. 6.(Argues that Mead's interest in intersubjectivity led him to emphasize objectivity and universality more than Dewey did.)

Mead, G.H. (1934) *Mind, Self and Society*, ed. C.W. Morris, Chicago, IL: University of Chicago Press.(A posthumous work on social psychology based on Mead's lectures including material on ethics.)

Mead, G.H. (1964) *Selected Writings: George Herbert Mead*, ed. A.J. Reck, Chicago, IL: University of Chicago Press.(Articles published by Mead, including several on ethics and social policy.)

Myers, G.E. (1986) *William James: His Life and Thought*, New Haven, CT: Yale University Press, ch. 13.(Deals with morality and includes James' social views.)

Peirce, C.S. (1958) *The Collected Papers of Charles Sanders Peirce*, ed. C. Harthorne and P. Weiss (1931-5, vols 1-6), revised by A.W. Burks (1958, vols 7-8), Cambridge, MA: Harvard University Press, vol. 8, paragraphs 7-38.(References as in this entry are standardly given by volume and paragraph number. This collection of Peirce's papers includes 'The Works of George Berkeley', his 1871 review of Berkeley.)

Peirce, C.S. (1966) *Charles S. Peirce: Selected Writings*, ed. P.P. Weiner, New York: Dover.(A representative selection of the Collected Papers with additional material such as correspondence with Lady Welby.)

Praise and blame

Praise and blame are philosophically interesting partly because, despite appearances, they are not simple opposites, but mainly because there are significant disagreements about whether, and when, they can be justified. The issue of justification connects praise and blame with some of philosophy's most central concerns: justice, desert and free will.

Disagreements about the justification of praise and blame tend to take two forms. In one the disagreement is about whether praise and blame can be justified without being deserved. Utilitarians, who argue that the rightness of praise and blame does not depend on desert, but on their contributing to the level of happiness, are opposed by those who believe that justice is of overriding value.

In its second form, the disagreement is about the essential requirements for deserving praise and blame. Among the conditions which have been proposed as essential are voluntariness (outlined originally by Aristotle), acting from the motive of duty (for praiseworthiness), and (usually in connection with blameworthiness) being free in a sense which is incompatible with determinism (the thesis that every event has a necessitating cause). Kant, who argued for both of the last two requirements, is a key figure in this debate.

1 The differences between praise and blame

Blame is a more complex phenomenon than praise. While the term 'praise' simply describes the conveying of a positive evaluation, 'blame' has a responsibility-attributing sense which is as much a part of its meaning as its evaluative component (see Responsibility).

Sometimes the responsibility-attributing sense predominates, as when a faulty carburettor is blamed for the malfunctioning of the car. But where persons are the objects, to blame is both to attribute responsibility and to censure. This does not mean that when inanimate things are blamed there is no element of evaluation: 'blame' always implies that something undesirable has occurred. But the extent to which blaming is an act of censure, as opposed (or in addition) to a judgment of responsibility is variable; whereas the extent to which praising is an act of commendation never varies, since it is just this.

Another difference is that while praise always consists in the voicing (either written or spoken) of approval, blame need not be voiced. One can blame oneself and others without saying so, because blame can consist in an attitude - for instance, the feeling of guilt which partly constitutes self-blame - or the silent passing of a judgment (as in 'She died without saying that she blamed him for the accident').

Because of these differences, blame (unlike criticism) can never simply be a negative version of praise. And although there is a considerable overlap in the constituency of objects which can be blamed and praised, there is not an exact coincidence either in the kinds of objects themselves, or the senses in which they can be blamed and praised.

2 Utilitarian versus desert theories of justified praise and blame

When philosophers disagree about the requirements for justified praising and blaming what concerns them is whether, and in what circumstances, these activities can be said to be morally right. And in considering this question they ignore some of the complexities mentioned above, to focus on those respects in which blame and praise *are* opposites - that is, to focus on blame as an act of censure and praise as an act of commendation.

There are broadly two kinds of dispute. The first is between utilitarians, who hold that the rightness of praising and blaming is to be determined by the contribution such activities make to the maximization of overall wellbeing, and those who hold that praising and blaming can only be right when deserved (see Justice; Utilitarianism). The second is over the question of what constitutes the deserving of blame and praise (see Desert and merit).

Utilitarianism regards the morality of praising and blaming, in the same way as it regards the morality of any act, in terms of the contribution these acts will make to the total level of happiness. So for the utilitarian praise and blame are morally justified if and only if they will help to increase overall happiness: they are seen as tools for encouraging beneficial (and discouraging harmful) behaviour. This means that the utilitarian could regard it as

right to blame people for things they had not done, if this was the most effective way of maximizing happiness. In other words, from the utilitarian point of view it could be right to blame someone who was (as one might say) blameless.

This view is anathema to those who regard justice as a value which should not be sacrificed for other values. To justice-valuers, the rightness of praising and blaming is determined by desert. But what is it to deserve praise or blame? In the moral context, it requires first of all that the agent should have done something morally right or wrong. This is agreed on by all non-utilitarians. It is also agreed that the act for which the agent deserves to be praised or blamed needs to have been voluntary, a view which can be traced at least as far back as Aristotle, who, as paraphrased by Ross, held that 'praise and blame attach to voluntary actions', that is, 'actions done (l) not under compulsion and (2) with knowledge of the circumstances' (*Nicomachean Ethics*) (see Aristotle §20; Will, the §1). The precise interpretation of each of these conditions of debate is a matter of debate, but for the purposes of this discussion, we can say that a voluntary act is one which has been performed knowingly and willingly.

3 Kant

But are knowledge and willingness enough? Kant (1785, 1788) held that praiseworthiness requires more: it also requires that the agent should be motivated by the thought of duty (see Duty; Kantian ethics). This suggests an asymmetry between the motivational requirements for praiseworthiness and those for blameworthiness. We do not believe that blameworthy agents must be inspired to act wrongly by the very thought of wrong-doing, as Satan was in Milton's *Paradise Lost* ('Evil be thou my Good'). Although we might regard such a motive as sufficient for blameworthiness, we do not regard it as necessary. It is considered enough, motivationally, for blameworthiness that agents were prepared to do what they knew to be wrong. But if Kant is right, it is not enough (motivationally) for praiseworthiness that they should merely have been prepared to do the right thing.

Many agree with Kant that praiseworthiness requires more than a willingness to do the right thing, but deny that praiseworthy acts must be motivated by the thought of duty, suggesting instead that such acts could have other admirable motives, such as sympathy, fellow-feeling, or a desire for justice. Kant would have agreed that these are admirable motives, but would have denied that they could be praiseworthy, because he held that praiseworthiness could be accorded only to actions which had moral worth and that such worth could not be based on any motives involving desires or other natural inclinations.

It has been argued that Kant's rejection of such motives as sources of moral worth is prompted by a concern for justice - a concern that something of such overriding value (as Kant thought moral worth to be) should not depend on the uneven and undeserved distribution of sympathetic and kindly inclinations, but should be available to all (see Impartiality). Kant's solution was to argue that the achievement of moral worth depends on the capacity to be moved by reason independently of desires and inclinations, a capacity which he appeared to think is possessed in equal measure by all adult human beings (apart from the insane) (see Self-control). Kant held that this capacity is manifested when agents are moved to act solely by the thought of duty.

4 Determinism and free will

Kant also believed (1781/1787) that blameworthiness required the ability to be moved to action independently of desires and inclinations. Here his worry was causal determinism. He held (1788) that desires are causally determining and as such inimical to the freedom required for moral accountability, in particular the freedom to have acted otherwise. The thesis that causal determinism is incompatible with such accountability (known as 'incompatibilism'), has been endorsed by many philosophers who, like Kant, are motivated at least partly by the belief that accountability requires the ability to have acted otherwise (see Free will). They argue as follows: if determinism is true, every event which has occurred, including the event of someone's acting, has had to occur; and since it would only be right to blame people for their actions if they could have acted otherwise, blame can never be justified. Kant, however, thought that blame could be justified, because he believed that the human capacity to be moved by reason (the same capacity he took to be the source of moral worth and which he thought enables us to act independently of our desires) in some way places us outside the causal order.

The latter belief has been subjected to much criticism, but many philosophers also reject the worry associated with determinism which led Kant to propose this as a solution. They argue that the essential requirement for moral

accountability is simply the ability to act as one wants to, an ability which is not threatened by determinism.

But incompatibilists are not comforted by this, for they have another worry, which is often confused with the anxiety about the ability to have acted otherwise, but which is in fact separable from it. This is the belief that the desires, preferences and inclinations which cause our actions have themselves been caused by factors for which we could not possibly be responsible. And, the incompatibilist asks, if we are not responsible for the causes of our motivational states, how can we deserve to be held responsible (to be blamed) for their results? This consideration also partly underlies Kant's insistence that blameworthiness cannot depend on desires. But his response to this worry (1793) was to claim that we have the ability to choose our moral natures; that the way we are morally is ultimately up to us, where 'up to us' means 'not dependent on any feature for which we are not completely responsible'. (Here in its fullest form is Kant's belief that in virtue of our reason, we can somehow transcend our natural, desiring, emotional selves - see Autonomy, ethical.)

The belief that we can desirelessly choose our natures (even our moral natures) has been dismissed as incoherent. But there is another model of freedom (known as 'agent-causation') which presents blameworthy agents as self-movers rather than self-choosers. The idea is that agents cause themselves to act in the light of their desires; they are not caused to act by their desires. This model too is thought to be problematic, but it is less easy to dismiss as incoherent. Both models are designed to meet the demand that there should be no contribution to an agent's actions for which the agent is not responsible.

The assumption that we cannot be morally responsible for our actions if we are not ultimately responsible for their causes (which motivates the above demand) has not received much direct attention. Those who find the assumption persuasive tend to treat it as something which does not require argument; those who do not find it persuasive tend either to ignore it, or to argue that our everyday practices of blame do not require justifications of this ultimate kind. But it could be argued that the assumption requires a more focused investigation, to determine whether it is possible either to provide a justification for it or to demonstrate that it is not tenable.

See also: Honour; Morality and emotions; Supererogation

MARTHA KLEIN

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by W.D. Ross, Oxford: Oxford University Press, 1975, III 1.(Provides a discussion of praise and blame and an analysis of voluntary actions.)

Kant, I. (1781/1787) *Critik der reinen Vernunft*, trans. N. Kemp Smith, *Critique of Pure Reason*, London: Macmillan, 1973, A528/B556-A558/B586.(Discusses the threat posed by causal determination to blameworthiness and the connection between reason and blame-deserving freedom.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(In the first chapter, Kant claims that moral worth belongs only to actions performed for the sake of duty. Ch. 2 touches on the connection between the human capacity for moral worth and the right of each human being to be treated with moral respect. In ch. 3, Kant argues for the connection between freedom, reason and morality.)

Kant, I. (1788) *Critik der practischen Vernunft*, trans. L.W. Beck, *Critique of Practical Reason*, Indianapolis, IN: Bobbs-Merrill, 1956, 99.(A very clear statement of Kant's worry that determinism would make it impossible for agents to act otherwise.)

Kant, I. (1793) *Die Religion innerhalb der Grenzen der blossen Vernunft*, trans. T.M. Greene and H.H. Hudson, *Religion Within the Limits of Reason Alone*, New York: Harper & Row, 1960, 40.(It is argued here that human beings can be held morally responsible for their actions only if they have freely chosen their moral natures.)

Klein, M. (1990) *Determinism, Blameworthiness and Deprivation*, Oxford: Oxford University Press.(Examines the free will/determinism debate and its relevance to blameworthiness. Appendix contains a commentary on an argument of Kant's which links morality and moral accountability with the existence of a desire-independent ability to act.)

Klein, M. (1990) 'Morality and Justice in Kant', *Ratio* 3 (1): 1-20.(A discussion of the connections between Kant's views on moral motivation and his concern for justice.)

Nozick, R. (1974) *Anarchy, State and Utopia*, New York: Basic Books; repr. Oxford: Blackwell, 1986, 125.(Here Nozick challenges an assumption which is similar (but not equivalent) to the assumption about ultimate

responsibility: that the foundations underlying desert must themselves be deserved all the way down.)

Sher, G. (1987) *Desert*, Princeton, NJ: Princeton University Press.(An interesting, clear and comprehensive analysis of the notion of desert, which challenges the view that if people do not deserve what makes their actions possible, [they cannot] deserve what they obtain through those actions.)

Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan; 7th edn, 1907; 7th edn repr. 1963, 428.(Presents the utilitarian view of praise.)

Strawson, P.F. (1974) ‘Freedom and Resentment’, in *Freedom and Resentment and other Essays*, London: Methuen, 1-25.(Considered a classic, this is an extremely influential defence of the view that our everyday practices of moral appraisal do not require justifications which involve the falsity of determinism.)

Williams, B. (1962) ‘The Idea of Equality’, in P. Laslett and W.G. Runciman (eds) *Politics, Philosophy and Society*, vol. 2, Oxford: Blackwell; repr.in *Problems of the Self: Philosophical Papers 1956-72*, Cambridge: Cambridge University Press, 1973: 230-49.(Contains fascinating comments about the relationship between ideas of equality and moral worth in Kant’s moral philosophy.)

Williams, B. (1971) ‘Morality and the Emotions’, in J. Casey (ed.) *Morality and Moral Reasoning*, London: Methuen, 1971; repr. in *Problems of the Self: Philosophical Papers 1956-72*, Cambridge: Cambridge University Press, 207-29.(Has inspired some of the discussion of Kant’s belief that justice could only be achieved if moral worth were available to all.)

Williams, B. (1976) ‘Moral Luck’, *Proceedings of the Aristotelian Society*, supplementary vol.50, 115-36; extensively revised version repr. in *Moral Luck: Philosophical Papers 1973-80*, Cambridge: Cambridge University Press, 1980, 21-39.(Another stimulating paper with some relevant remarks on Kant’s concerns about justice.)

Williams, B. (1985) *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press and London: Fontana, ch. 10.(Links the notion of blame which requires ultimate responsibility with what he calls ‘the morality system’ - a group of beliefs which he argues we would be better off without. Important, but not easy.)

Williams, B. (1989) ‘Internal Reasons and the Obscurity of Blame’, in *‘Making Sense of Humanity’ and other philosophical papers 1982-1993*, Cambridge: Cambridge University Press, 1995.(An interesting discussion of what Williams calls ‘focused blame’ and the way it works.)

Praxeology

Praxeology belongs to the pragmatic tradition and thus emphasizes that concepts - and the world - must be understood through and elucidated in terms of human activities and practices. Praxeology is not a school but rather a movement, originating in Denmark and Norway in the 1960s and 1970s, with internal debates and disagreements. Praxeologists stress that good conceptual and ontological analyses proceed by (and are presented in the form of) careful analyses of particular examples or cases. They emphasize the situatedness of the philosopher in the world. The inspiration comes primarily from the early Heidegger and the later Wittgenstein.

The term ‘praxeology’ means here a systematic philosophical analysis and account (*logos*) of human activities and practices (praxis). Praxeology in this sense has also been labelled ‘contextual pragmatics’ and, in its more Wittgensteinian form, ‘language game pragmatics’ (see Wittgenstein, L. §11). The ‘praxiology’ of Taduesz Kotarbiński (1965) and Lange (1963), which will not be treated here, is quite different, as it aims at a general scientific theory of efficient (rational) action; a similar idea is Ludwig von Mises’ ‘praxeology’(1949), a general theory of action in the form of a generalized political economy.

Praxeology started in Norway and Denmark and is still strongest in the Scandinavian philosophical community, though it has also had a considerable influence in the German speaking world. It began in the 1960s as a discursive, analytical practice in the circle around the Norwegian philosopher Jakob Meløe (1927-), which later, in the 1970s, came to be known as ‘praxeology’.

The most important philosophical task according to Meløe and later praxeologists is careful description and elucidation of human activities in their concrete settings. ‘The adequate *philosophers*’, Meløe (1970: 20) says, ‘are Wittgenstein, Heidegger and Austin, in that order’.

The praxeological movement is characterized by an integration of influences from various philosophical traditions, particularly phenomenology - stressing the life-world-pragmatism, hermeneutics and ‘linguistic phenomenology’ (see Austin, J.L.; Phenomenological movement; Pragmatism). There exists, however, a variety of analytical orientations, covering fundamental ontology and pure conceptual analysis, as well as political-philosophical cultural critique.

Certain common convictions about the world and the best way of doing philosophy define what it is to be a praxeologist:

(1) human insights are primarily located in human activities (practices);
(2) human actions are of various, overlapping kinds;
(3) the articulation of these insights calls for careful description and further elaboration and analysis of cases of the activities in question.

The first point expresses the pragmatist’s stance. The second is the expression of an anti-reductionist attitude: there are no final and universal schemes or action-types which apply to all possible human activities. What actions there are and how they are interrelated have to be brought to the fore through an analysis of particular cases.

Praxeology emphasizes the internal nature of the relation between an agent, the activity, the equipment and instruments, the material used or transformed and the context or ‘landscape’ in which the agent acts. From the point of view of praxeology an object *is* an object-as-used, an agent*is* a person-as-acting, the world *is* the world-as-we-act-and-live-in-it, and so forth. One source of inspiration here is Heidegger’s discussion about tools (*Zeug*) in *Sein und Zeit (Being and Time)* (1927) (see Heidegger, M. §3).

A knowing agent is seen as a constitutive element of an activity. Jakob Meløe says about somebody who is making shoes:

> He sees what he needs to see to do what he does, he masters what he needs to master to do what he does, he knows what he needs to know to do what he does, etc. I abbreviate all this to: he *knows* what he is doing - and let it stand as an axiom that *the agent* knows what he is doing.
>
> (1970: 23)

Furthermore, *in knowing this* agents also know *the world* they are acting in. Gunnar Skirbekk (1993: 153) calls the 'axiom' *the praxeological thesis*. This necessary and 'activity-constitutive' knowledge is to a large extent tacit (see Knowledge, tacit).

Praxeology is anti-sceptical, and as part of that anti-solipsistic, in the sense that praxeologists take it as their starting point that on the whole we know what we do in our life-world and that one person's knowledge is also accessible to others. Jakob Meløe writes:

> In a society with a complex division of work, we are all of us poor observers of the doings of most of our fellow men…. But also, there are no two jobs, or no two sets of practices, with no forms in common between them. And that is what makes our world one world.
>
> (1970: 26)

The third point above expresses the praxeological method as well as the contextualist stance of praxeology. From the point of view of praxeology an activity is always situated in a context and cannot be understood as the kind of activity it is unless it is understood as an *activity-in-that-context*.

The method of praxeology is the exhibition of what is constitutive of or essential to a certain activity through a minute analysis of *examples*. This amounts to an exhibition of what is 'inherent' or necessarily presupposed in an activity. The goal is to *show* this rather than to prove it.

The particularistic and contextualistic nature of praxeology has led to vivid debates, especially in Norway and in Germany (Böhler *et al.* 1986; Skirbekk 1993), between praxeologists and representatives of transcendental and universal pragmatics (Apel, Habermas and their followers). Praxeology seems, for example, to make very strong, in some sense universal, hermeneutic claims about communication and knowledge.

Meløe's statement about what makes our world one world is a good example. Some praxeologists deny that any such universal 'theoretical claim' have been made at all. What seems to be a general presupposition is said to exist only as analytical *practice*. Some praxeologists, however, accept and stress the existence of universal presuppositions (Skirbekk 1993).

See also: Hermeneutics

BENGT MOLANDER

References and further reading

Böhler, D., Nordenstam, T and Skirbekk, G. (eds) (1986) *Die Pragmatische Wende: Sprachspielpragmatik oder Transzendentalpragmatik? (Pragmatic Change: Language Game Pragmatics or Transcendental Pragmatics?)*, Frankfurt am Main: Suhrkamp.(Focuses on the debate between contextualists and universalists. See also, in particular, the paper by Böhler and the comments to it.)

Heidegger, M. (1927) *Sein und Zeit (Being and Time)*, 7th edn, Tübingen: Max Niemeyer, 1953.(In particular the discussion about tools and understanding, §§12-34.)

Kotarbiński, T. (1965) *Praxiology: An Introduction to the Sciences of Efficient Action*, Oxford: Pergamon Press and Warsaw: PWN-Polish Scientific Publishers.

Lange, O. (1963) *Political Economy*, vol. 1, *General Problems*, Oxford: Pergamon Press and New York: Macmillan.

Meløe, J. (1970) 'The Agent and his World', in G. Skirbekk (ed.) *Praxeology: An Anthology*, Oslo: Universitetsforlaget, 1983, 13-29.(The classic praxeological text).

Meløe, J. (1988) 'The Two Landscapes of Northern Norway', *Inquiry* 31: 387-401.(Shows Meløe as a mature praxeologist with very instructive examples.)

Mises, L. von (1949) *Human Action: A Treatise on Economics*, revised edn, New Haven, CT: Yale University Press, 1963.

Pahuus, M. (1983) 'On the Understanding of Tools and Articles for Use: A Comment on Jakob Meløe: The Agent and his World', in G. Skirbekk (ed.) *Praxeology: An Anthology*, Oslo: Universitetsforlaget, 1983, 30-7. (Showing the influence of Heidegger.)

Skirbekk, G. (ed.) (1983) *Praxeology: An Anthology*, Oslo: Universitetsforlaget.(The classical Scandinavian texts, also good as introductory reading.)

Skirbekk, G. (1993) *Rationality and Modernity: Essays in Philosophical Pragmatics*, Oslo: Scandinavian

University Press and Oxford: Oxford University Press.(Especially ‘Praxeological Reflections’ and ‘Contextual and Universal Pragmatics: Mutual Criticism of Praxeological and Transcendental Pragmatics’; good in bringing out the debate between contextualists and universalists.)

Prayer

The concept of prayer is now most commonly applied to any sort of communication which is addressed to God. That is, prayer is that activity in which believers take themselves to be speaking to God. One may ask God to do something (petitionary prayer), but that need not be the only sort of content that prayer may have. There are prayers in which one thanks God for something, and others in which one praises God and expresses one's adoration. A worshipper may also pray to express (or to make) a commitment to God, or to make a vow. Penitents pray to confess their sins, to express their repentance, and to ask for divine mercy and forgiveness. In general, any sort of speech-act which might be addressed by one human being to another could also be addressed to God, and thus be a prayer. Some such acts (such as, perhaps, commanding) might be thought inappropriate when addressed to God, but no doubt there can be inappropriate prayers. And some prayers may even be tentative and unsure about the existence of the addressee, prayers which might be thought of as beginning 'O God, if there is a God… '.

Some writers, principally from within a tradition of mysticism, also apply the notion of prayer in a somewhat broader sense - in, for example, expressions like 'prayer of quiet' and 'prayer of union'. Here 'prayer' seems to mean any intentional state - worship, adoration, enjoyment of the divine presence and love, and so forth - which the worshipper believes to be associated with a genuine contact with the divine, regardless of whether it contains an element of communication addressed to God.

In the sense of a communication addressed to the divine, prayer seems to fit best with the theistic religions, which construe God as a person, or as something like a person. Here the addressee is taken to be someone who is an appropriate recipient of a communicative act. The fit seems rather more awkward in those religions which construe the divine reality in impersonal terms. With reference to prayer in the theistic religions, a principal topic of philosophical interest involves the omniscience and benevolence of God - if he knows all my needs and desires, why inform him of them through prayer? And will he not satisfy all my needs regardless of whether I pray? If divine benevolence is conditional on prayer, it seems less than perfect. A response to the first question is to point out that not all speech-acts need be construed as conveying information; a response to the second is to argue that our having to ask for things on behalf of ourselves and others might make for a better world than if this were not the case. Another issue is the way in which God responds to prayer. Some argue that God responds through miracles; others suggest that God, knowing our future prayers, providentially created a world that would satisfy them - thus prayer causally influences earlier events.

1 Prayer and divine omniscience

The doctrine of divine omniscience, which is often a part of orthodox Christian theology, claims roughly that God knows every truth that is knowable. Most theologians have held that this includes all the truths about the future, including truths about the future acts of free creatures (see Omniscience). If God is omniscient in this way, then in prayer one does not inform God of something which he did not already know. Before one prays, God already knows what one wants, what one needs, what one will say in the prayer, what one will do, and so on. At this point, however, we must be careful. The divine foreknowledge may anticipate human action in the order of time, but it cannot substitute for that action, in the sense of making it irrelevant whether the action is actually done. For if the action were not done, then the divine knowledge would have been different from what it actually was. This applies to prayer as much as to anything else.

So perhaps God knows now that I will confess my sins tomorrow. But God knows this only if I will actually confess my sins tomorrow. If, when tomorrow comes, I do not confess, then God never knew that I would confess tomorrow. My action, so far from being made irrelevant by the foreknowledge, is instead essential to that foreknowledge.

It is true, however, that when in prayer I confess my sins, or express my desires and needs (or what I take to be my needs), and so on, I am not informing God of things which he did not already know. He already knew about those sins, he knew what I wanted, and what I thought I needed, and what I really needed. So there may be a problem about the point, the significance, of the practice of prayer in the light of the alleged omniscience of God.

Now it is true that the point of some human conversation is that of communicating information (or misinformation)

to a hearer who did not previously know it. But that is certainly not the only point of speech within a purely human context. It is not even the only point of speech in the declarative, indicative mood. Students who respond to a teacher's classroom question are often not informing the teacher of something the teacher does not already know. Nor, in such cases, do the students even suppose that this is what they are doing. They are attempting to demonstrate or to exhibit their own knowledge of the subject matter, not to extend the teacher's grasp of that subject matter.

Closer to much religious praying, when I have wronged someone, and I apologize to that person for what I have done, the apology usually includes a statement of what was done. But the hearer is often already painfully aware of what was done, and I know that this is the case. I am not trying to make the hearer aware of what was done. But the apology involves my acknowledging what I did, and I do that by saying what I did. My statement has a point, though its point is not that of conveying new information. The point of the confession is that I 'own up' to what I did, that I acknowledge the responsibility for it. And that is a prerequisite for, or perhaps a part of, a sincere apology. In a similar way, when I congratulate someone on an achievement, I normally say what the achievement was. It is not that I think that the hearer is unaware of the achievement. But the congratulation loses much of its punch if it is not conversationally pegged to the item that gives rise to it.

The fact is that human speech, even within a purely human context, is extremely varied and flexible. If we are to take seriously the practice of prayer, then we need to allow for at least a similar degree of flexibility in the human address to God.

2 Prayer and divine benevolence

The concern about divine benevolence, in connection with prayer, arises primarily with reference to petitionary prayer, in which the worshipper asks God to do something. I ask God to do something for me, or for someone else, something which I presumably think will be good for the recipient. But we may think that if God is fully benevolent, then he already (and independently of my prayer) wants to do what is good for me and for the rest of the world. Furthermore, because of his omniscience, he already knows (independently of my prayer) what would be good for all of us. Is there any way, therefore, in which my prayer could be relevant to whether God does the thing for which I ask?

If there is any truth at all in the claims that God exists and loves his creation and knows what all of us need, then indeed there must be many good things which God does for us without our asking him. Many of the things which are in fact necessary for my life, health and happiness are things of which I am completely unaware. Even if I spent all of my time praying and did nothing else, I could not pray for all of those things. So it seems implausible to think that God does only those good things for which we pray. And there may also be things for which we pray, and which God does, but which he would have done for us anyway because he is good and he loves us.

There may, however, be another class of things. These are the things that it would be good for God to do for us *if we ask him*, and it would not be good for God to do them otherwise. These would be divine acts whose goodness was constituted, in whole or in part, by the fact that we have prayed for them. Or, to look at it from the other direction, they would be good because they were divine responses to human requests. If there are things which belong in this category, then our praying for them is not irrelevant to whether God does them.

There seem to be analogous things within purely human contexts. No doubt parents do many good things for their children, regardless of whether the children ask them - before, indeed, the children are mature enough to ask. But if the children are to grow into a genuinely personal relation with their parents, then some of the initiative must shift to the other side. Making a request of another person, and then acknowledging the response with thanks, is part (of course, it is only a part) of growing into actuality as a person oneself. Petitionary prayer seems to be based on the conviction that human beings are invited into an analogous sort of personal interaction with the divine person.

The difficulty regarding prayer and divine benevolence may be more severe when we think of praying for good things - healing, for example - for someone else. It may seem puzzling that a good thing which God might do for another person might depend in any way on *my* requesting it. Here too, we must remember that if there is anything at all to the theistic view of the world, then God must be doing many good things for each of us regardless of whether anyone prays for them. But it may also be the case that part of the divine purpose for the world - a good purpose - is that we should grow into a genuine community, a community of mutual love. That would be a

community in which each of us does good things for the others, including the good of praying for the others, and each of us also receives with thanks this service from the others. It might be said that in such a community we all ‘take in each other’s washing’. Maybe in some deep way it is better that I do someone else’s washing, and they do mine, than for each to do our own. If so, then there may be things which, in the end, it would be good for God to do for me if someone else asks for them, and not otherwise.

So far the relation of prayer to the divine benevolence has been discussed as it might appear to someone primarily concerned to understand this aspect of religious faith. But these concerns may also arise in a more polemical context, one in which these problems are incorporated into some special version of the problem of evil (see Evil, problem of §§1, 6). In that context, some alleged difficulty about prayer would function in an argument which was intended to cast doubt on the existence of God, or on the divine benevolence or power, and such like. Initially at least, cases of ‘unanswered’ prayer - cases in which we have prayed for good things, and those good things have not happened - seem to be the most plausible cases on which to focus. And it seems that there are many such cases. Think, for example, of all those who have prayed for peace while wars continue to abound. On reflection, however, we may well come to think that there is nothing special in this version of the problem of evil. We can readily think of many good things (or things, at least, which seem to us to be good) which God has evidently not done in the world. If that fact is really incompatible with the existence or goodness of God, then theism is already in terrible difficulty, and adding a few more cases of good things for which we have prayed will probably make little difference to the logic of the situation.

If, on the other hand, we can think of something which blunts the force of the general form of this problem, then probably the same sort of thing can be said about unanswered prayers. If, for example, it is relevant and useful to point out that we may be mistaken about what really would be good for us or for the world, then it would be equally relevant to observe that we may make similar mistakes when we pray for what we think is good. So the case of unanswered prayer probably does not generate a significantly different version of the problem of evil.

Curiously, however, the case of ‘answered’ prayer might be a more promising candidate for generating a special and distinctive variety of that problem. For we may think that, in some cases at least, the good thing that was done would not have been done without the prayer, and we may then construe that possibility as incompatible with divine benevolence. We might think, therefore, that we have discovered a special and curious fact about the world, that of good things done in response to prayers, which cannot be fitted into a more or less orthodox theistic view of the world.

There is, of course, a very large literature about the problem of evil, and it continues to be a live topic of controversy in the philosophy of religion. It is useful to keep in mind, when engaging in this controversy, the difference between a theodicy and a defence (see Hick 1966 and Plantinga 1967). Roughly, a theodicy is an attempt to explain, from a theistic point of view, why there is evil in the world, why God permits it, and so on. That is a relatively ambitious project, and one of the requirements for succeeding in it is that the explanations we give should be true - that is, that we correctly identify the divine intentions and purposes in allowing evil.

A defence, on the other hand, is more modest. Defenders attempt merely to refute some anti-theistic argument from evil by showing that it is invalid or that it has some false premise. Defenders may choose not to propose any positive explanation of their own as being the true account of evil. Perhaps they will not even claim to know any such account. They may instead content themselves with pointing out that there are some possible explanations which are compatible with the facts of the world. If there are possibilities of that sort, then the facts of the world (including the evils) are not logically incompatible with the existence and benevolence of God.

It was suggested above that there may be things which it would be good for God to do in response to a prayer, and not otherwise. If we construe that as the actual explanation for the phenomenon of answered prayer, then it would constitute a partial theodicy, a theodicy for that particular curious fact about the world. But the mere logical possibility of such things, even if there actually are no such things, could serve as the basis for a defence. That is, the fact that such cases are possible would show that cases of answered prayer are not by themselves incompatible with the benevolence of God. It does not seem, therefore, that prayer contributes any serious special element to the problem of evil. If we can find a satisfactory theodicy or defence as a response to the more general versions of the problem of evil, then that line of response will probably cover the case of prayer as well. And if there is no satisfactory theodicy or defence for the more general versions, then there will be plenty of trouble for theism

regardless of what we say about prayer.

3 Prayer and the natural order of the world

The notion of God's doing something in the world has been prominent in the last few paragraphs, and that can hardly be avoided in discussions of petitionary prayer. For such prayers - 'Give us this day our daily bread', for example - characteristically ask God to do some more or less specific thing in the world. Perhaps the most purely 'spiritual' requests - 'Forgive us our sins', for example, or 'Receive us into the heavenly kingdom' - can be thought of as being divorced from the natural world and independent of what happens in it. But much petitionary prayer pretty clearly involves that world. There does not seem to be any plausible way of thinking about such prayer without also thinking about the possibility, or impossibility, of God's responding by way of some action in the natural world.

One such possible mode of divine response would be that of a miracle. Here a miracle will be construed as a direct and *de novo* causal or quasi-causal act of God which makes the natural world different from what it would have been, given only the other elements of the world and the ordinary causal laws and powers operative within it. This is close to the 'violation' concept which David Hume was trying to formalize in his celebrated definition, 'A miracle may be accurately defined, a transgression of a law of nature by a particular volition of the Deity, or by the interposition of some invisible agent' (1748, section X). Since Hume's time there has been a large amount of discussion of the possibility, actuality, significance, and so forth, of miracles in this sense (see, for example, Swinburne 1970). This topic is much too large to broach here, but it can be said that many participants in the practice of prayer consider this as one possible mode of divine response (see Hume, D. §2; Miracles).

Many theologians seem also to recognize another, non-miraculous, mode of response, which is sometimes called 'providential' (see, for example, Lewis 1947: Appendix B). The idea here is that the divine response to a petitionary prayer may consist of an event which happens in the 'ordinary' course of nature, through the operation of ordinary causes. So, for example, my recovery from an illness - prayed for by me and my friends - may happen by what seem to be the regular and ordinary effects of drugs, surgery, and so forth. But in what sense could the recovery then be a divine response to the prayer? Well, the course of the world may be thought to be determined by three sorts of factors - the laws of nature, the 'initial conditions' of the world, and the acts of free agents. (Some might want to include a fourth category, that of chance events. In this context, they can be handled in the same way as free acts.) Suppose we think also that God creates a world, in part, by creating laws of nature and initial conditions. He knows what future acts will take place in that world; in particular, he knows whether I and my friends will pray for my recovery from the illness. If then he arranges the laws of nature and the initial conditions in such a way that my recovery ensues, and if our prayers were a factor entering into the divine decision, then my recovery would be providential, and an answer to prayer, though it occurred in the ordinary way because of drugs and surgery.

This idea requires that we allow for the possibility of an event's playing a role in determining some previous event or state of affairs. So my prayer now is to be one of the factors that God considers in setting up the world 'in the beginning'. But this is probably not a special problem for prayer. For the most straightforward way of thinking about divine foreknowledge is to think that my action today determines something of what God knew yesterday and throughout all previous time. If I had done something different today then the divine knowledge would forever have been different - forever in the past as well as in the future.

Some prefer to resolve this problem by appealing to the idea that God is eternal, rather than, say, merely everlasting (see Eternity). And an eternal being is 'outside' time - for such a being there is no before or after. This, then, removes the 'fore' from the divine knowledge, and perhaps removes with it the feeling of a 'retrograde' causality in which my present act determines a previous state of knowledge in God. This appeal, of course, has its own difficulty; the idea of eternity is notoriously difficult to understand. But we have no guarantee that every truth is easy. Others (for example, Mavrodes 1984) prefer to accept the idea of temporally retrograde causality, or something like causality, in which future events play some role in shaping the world that comes before them. Perhaps the most interesting philosophical question in this connection is that of whether we can find any reason for a decisive rejection of this possibility.

See also: Deism; Process theism; Providence; Religion and science

GEORGE I. MAVRODES

References and further reading

Hasker, W. (1989) *God, Time, and Knowledge*, Ithaca, NY: Cornell University Press. (Chapter 7 includes a critical discussion of the idea of temporally retrograde causality.)

Hick, J. (1966) *Evil and the God of Love*, New York: Harper & Row.(This is an extensive discussion of several systems of theodicy.)

Hume, D. (1748) *An Enquiry Concerning Human Understanding*, ed. L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press, 1975.(Section X is probably the most famous philosophical discussion of miracles in Western philosophy, and the focus of many subsequent discussions.)

Lewis, C.S. (1947) *Miracles*, New York: Macmillan.(A general discussion of miracles. Appendix B discusses 'special providences' as distinct from miracles.)

Lewis, C.S. (1964) *Letters to Malcolm*, New York: Harcourt, Brace & World.(Highly readable, but careful, discussions of many aspects of the practice of prayer.)

Mavrodes, G.I. (1984) 'Is the Past Unpreventable?', *Faith and Philosophy* 1 (2): 131-45.(This is a further discussion and defence of the possibility of temporally retrograde causality.)

Plantinga, A. (1967) *God and Other Minds*, Ithaca, NY: Cornell University Press. (Chapters 5 and 6 constitute a careful development of a free-will defence (not a theodicy) in connection with the problem of evil.)

Stump, E. (1979) 'Petitionary Prayer', *American Philosophical Quarterly* 16 (2): 81-91.(Largely a discussion of prayer and divine benevolence, with considerable reference to the early and medieval philosophical theologians.)

Swinburne, R. (1970) *The Concept of Miracle*, London: Macmillan.(This is a defence of the possibility, and significance, of miracles in the Humean sense.)

St Teresa of Avila (1565) *The Autobiography of St. Teresa of Avila*, trans. E. Allison Peers, Garden City, NY: Image Books, 1960.(A fascinating mixture of autobiography, reflection and spiritual advice and teaching by one of the great mystical writers.)

Predestination

Predestination appears to be a religious or theological version of universal determinism, a version in which the final determining factor is the will or action of God. It is most often associated with the theological tradition of Calvinism, although some theologians outside the Calvinist tradition, or prior to it (for example, Augustine and Thomas Aquinas), profess similar doctrines. The idea of predestination also plays a role in some religions other than Christianity, perhaps most notably in Islam.

Sometimes the idea of predestination is formulated in a comparatively restricted way, being applied only to the manner in which the divine grace of salvation is said to be extended to some human beings and not to others. John Calvin, for example, writes:

> *We call predestination God's eternal decree, by which he compacted with himself what he willed to become of each man. For all are not created in equal condition; rather, eternal life is foreordained for some, eternal damnation for others. Therefore, as any man has been created to one or the other of these ends, we speak of him as predestined to life or to death.*
>
> *(Institutes, bk 3, ch. 21, sec. 5)*

At other times, however, the idea is applied more generally to the whole course of events in the world; whatever happens in the world is determined by the will of God. Philosophically, the most interesting aspects of the doctrine are not essentially linked with salvation. For instance, if God is the first cause of all that happens, how can people be said to have free will? One answer may be that people are free in so far as they act in accordance with their own motives and desires, even if these are determined by God. Another problem is that the doctrine seems to make God ultimately responsible for sin. A possible response here is to distinguish between actively causing something and passively allowing it to happen, and to say that God merely allows people to sin; it is then human agents who actively choose to sin and God is therefore not responsible.

1 What is predestination?

The idea of predestination belongs to a group of theological ideas which are closely related (if, indeed, they are not practically identical). Among them are foreordination (which is often used merely as a stylistic variant of predestination), divine election (the divine choosing of some persons to receive the gift of salvation), divine providence, the divine governance of the world, divine decrees and divine sovereignty. Perhaps a little more distantly related is the idea of divine foreknowledge (a special case, presumably, of divine omniscience). Sometimes one of these ideas is defined in terms of some of the others. But all of them seem to be special theological concepts, and this method of definition is not very useful to someone who is trying to make an initial entry into this set of ideas.

Calvin, however, in the passage quoted above, speaks of God as having 'compacted with himself what he willed to become of each man'. Similarly, the *Westminster Shorter Catechism* defines a divine decree as 'His eternal purpose according to the counsel of His will, whereby, for His own glory, He hath foreordained whatsoever comes to pass', a definition endorsed in the twentieth century by the prominent Calvinist theologian Louis Berkhof (1959: 102). Perhaps we can make a start, therefore, by thinking of predestination as something like a divine intention about how things should happen in the world. This intention belongs to eternity, or at least it predates the events to which it refers. Calvinist theologians usually add that this divine will is efficacious, in the sense that whatever is decreed, predestined, and so on, is certain to happen in just that way. This efficacy is sometimes said to be mediated through another divine act, the governance of the world.

Put in this way, the doctrine seems fairly straightforward, and makes a strong claim about the way in which the course of the world is related to the divine will. In response to an objection, however, many theologians introduce an important revision, which greatly weakens the idea. This revision will be considered in §3.

2 Sources of the doctrine

The doctrine of predestination appears to have four main sources:

(1) Certain biblical passages, most notably in the Pauline epistles, but also in some other places (Boettner (1948)

cites all of these).
(2) Speculation based on rather general considerations about the divine nature and activity - God's creation of the world, his power, his majesty, and so on. A classic example is that of Thomas Aquinas, who argues that 'since every agent acts for an end, the ordering of effects towards that end extends as far as the causality of the first agent extends'. But God is the First Cause, whose causality encompasses that of all created agents. And so 'it necessarily follows that all things, inasmuch as they participate in existence, must likewise be subject to divine providence' (*Summa theologiae* Ia, q.22, a.2).
(3) Speculation and argument based on a different special attribute of God, namely his foreknowledge of all future events. 'Foreknowledge implies certainty, and certainty implies foreordination' (Boettner 1948: 44).
(4) Empirical observations of religiously significant facts about the world - for example, that various people, because of the circumstances of their birth, their culture, and such like, have vastly different opportunities for hearing and understanding the Christian message, that there is a great disparity in the responses of people who do hear that message, and so on (see, for example, Calvin's *Institutes*, bk 3, ch. 21, sec. 1).

The second and third of these lines of argument, especially, would seem to support the most general and extensive versions of predestination, if they support any at all.

3 Objections and emendations

From early on in the history of the Church - apparently, indeed, from at least the time of the writing of the Pauline epistles - various objections have been raised against the doctrine of predestination. These include:

(1) That it is inconsistent with other passages in the New Testament, such as those which suggest that all human beings will eventually be saved (Boettner (1948) cites and discusses some of these passages, as well as some of the other lines of objection).
(2) That it is incompatible with the justice and/or the love of God, because it represents the divine choice and action as arbitrary and unfair.
(3) That it is incompatible with human moral responsibility, because it implies that there is no human free will (see Free will).
(4) That it impugns the character of God, by making God morally responsible for sin (see Omnipotence §5).

One of the most serious difficulties which faces a person who considers these objections, and possible replies to them, is that of getting a clear idea of just how strong a claim is made by the doctrine of predestination. The idea of predestination has the 'feel' of a causal, or quasi-causal, notion. Initially at least, we tend to think that it is being used to provide a causal explanation of certain facts in the world. Why did this person receive the Gospel when that one did not? Ultimately, at least, it was divine predestination, the divine decree, which had these results. Sometimes this sort of claim seems quite explicit:

> An effect conceived in posse only raises [*sic*] into actuality by virtue of an efficient cause or causes. When God was looking forward from the point of view of His original infinite prescience, there was but one cause, Himself. If any other cause or agent is ever to arise, it must be by God's agency. If effects are embraced in God's infinite prescience, which these other agents are to produce, still, in willing these other agents into existence, with infinite prescience, God did virtually will into existence, or purpose, all the effects of which they were to be efficients.
>
> (Dabney 1871 (1985): 212)

Often, as in this passage, the idea is developed by suggesting that God is the remote cause, the first cause, who determines all subsequent events by initiating a deterministic chain of causes and effects. This chain terminates in (in many cases) the actions of human agents, who are the proximate causes of events.

This sort of claim seems, of course, to generate immediately the third and fourth sort of objection mentioned above. With reference to free will, some writers (Luther seems to be a prominent example) seem content to accept the implication and to deny that there is any real free will. Others avail themselves of the idea sometimes called (in non-theological contexts) 'soft' determinism. They argue that God determines human acts by determining the motives, desires, and so forth, of human beings. Since humans then act in accord with their own desires and motives, they act 'freely', despite the fact that their acts are determined by a divine act from before the foundation

of the world.

Perhaps the other objection, that this view makes God responsible for sin because it makes him the cause of each sinful act, seems more troublesome to proponents of predestination. There is a tendency for writers on this topic to put forward very strong and extensive claims about predestination, and to offer arguments which, if they are sound, support these very strong claims. But later, under the pressure of objections, they fall back on versions of the doctrine that seem to be radically revised and weakened.

An example of such a revised version appears in the *Westminster Confession of Faith* (1646), which says that:

> God from all eternity did, by the most wise and holy counsel of his own will, freely and unchangeably ordain whatsoever comes to pass; yet so, as thereby neither is God the author of sin; nor is violence offered to the will of the creatures; nor is the liberty or contingency of second causes taken away, but rather established.
>
> (*Westminster Confession of Faith*, ch. 3)

This is evidently an attempt to address objections of the third and fourth types mentioned above. If we are to take it seriously, however, we may well be puzzled as to how we are now to understand the claim that God ordains all things. If God is not the author of sin, then it would seem that there are a great many events in the world of which God is not the author. And we may also wonder whether, if the world can have in it some sins of which God is not the author, it may not also have in it many other events of which God is not the author. Furthermore, what shall we say now of those arguments - from foreknowledge, for example - which would seem to include within the scope of the divine determination sins as well as anything else? If we take seriously the qualification that ‘neither is God the author of sin’, and so on, then what remains in our understanding of what divine foreordination amounts to?

In saying this, we may, on the other hand, be too careless and hasty in our understanding of the term ‘author’. Could God be the cause of everything without being the author of everything? If so, then we could retain the *Westminster Confession* qualification while still accepting the arguments for God’s universal causal responsibility. But in order to be relevant to the problem that is being addressed, authorship must be understood as generating moral responsibility for what one is the author of, while foreordination (or causality) without authorship must not generate such responsibility. What understanding of foreordination and authorship will satisfy these requirements?

One way of proceeding would be to construe foreordination as a very general sort of causal notion which includes at least two kinds of cause. One of them, authorship, is the sort that generates moral responsibility, and the other is the sort that does not. We have already mentioned two sorts of cause, remote and proximate. God was said to be the remote cause of everything, while (in many cases, at least) created agents were the more proximate causes. Unfortunately, however, this distinction does not have the right moral implications. Remote causes drain moral responsibility away from proximate causes. To construe authorship as proximate causality, and thus to say that God is not the proximate cause of sins, would have the opposite effect from that which the defender of predestination needs. The moral responsibility for sins would be transferred from the human agents to God, the remote cause.

A more promising strategy appeals to a different distinction. This is sometimes put as the difference between positive and negative causality, or between active and passive causality, or between causing something and merely allowing it to happen. This strategy is attractive because this distinction really does seem to have the required moral implications. The morality of doing an act yourself may be very different from that of merely allowing someone else to do a similar act, and actively causing something to happen may be morally different from merely allowing it to happen. So a nineteenth-century apologist for predestination says:

> To decree, is nothing more than to determine beforehand, or to foreordain; and, to resolve, or determine to do or permit anything, is to decree it in that sense…. That which is determined to be done, is decreed; and that which is determined to be permitted, is also decreed, when there is power to prevent it…. Now, in one or other of these ways, God ‘has foreordained whatsoever comes to pass.’ This, as you know, is the simple language of our catechism.
>
> (Smith 1854: 32-6)

And according to the Calvinist theologian Louis Berkhof, ‘the decree respecting sin is not an efficient but a permissive decree, or a decree to permit, in distinction from a decree to produce, sin by divine efficiency’ (1959:

108). Long before either of these, Thomas Aquinas had argued that reprobation (the predestination of some humans to damnation) was not the cause of the sinner's sin, although predestination was the cause of the elect's receiving grace. He says that predestination includes the will to confer grace, but of reprobation he says only that it 'includes the will to permit a person to fall into sin' (*Summa theologiae* Ia, q.23, a.3).

The crucial words in these accounts, of course, are 'permit', 'permitted', 'in one or other of these ways', and so on. What God permits may indeed be done by the free will of human beings, and perhaps it is then those human agents, and not God, who bear the moral responsibility for those acts. This really does undercut the objections that say that predestination makes God morally responsible for sin, or that it eliminates the free will of creatures.

But 'the simple language of our catechism' used what appear to be very strong notions - foreordination, predestination, divine election, and so on. One may be surprised to hear now that these apparently powerful claims about God can be satisfied merely by God's *allowing* other agents to act. At least, interpretations of this sort suggest that the doctrine of predestination may not be nearly as strong as it first appears, and that it is one which may be unusually prone to misapprehension.

See also: Eternity; Fatalism, Indian; Grace; Omniscience; Providence §§3-4; Reprobation; Salvation

GEORGE I. MAVRODES

References and further reading

Aquinas, T. (1266-73) *Summa theologica*, trans. Fathers of the English Dominican Province, London: Burns, Oates & Washbourne., 1920.(This is an extensive and extremely influential piece of philosophical theology by one of the major figures in medieval philosophy and theology.)

Berkhof, L. (1959) *Systematic Theology*, Grand Rapids, MI: Eerdmans, 4th edn.(A discussion of a standard range of theological topics by a significant American Calvinist theologian.)

Boettner, L. (1948) *The Reformed Doctrine of Predestination*, Grand Rapids, MI: Eerdmans, 6th edn.(This is a very extensive and useful discussion of this topic by a strong defender of the Calvinist tradition.)

Calvin, J. (1559) *Institutes of the Christian Religion*, trans. F.L. Battles, Philadelphia, PA: The Westminster Press, 1960.(This is the major theological work by one of the most prominent Protestant reformers of the sixteenth century.)

Dabney, R.L. (1871) *Lectures in Systematic Theology*, Grand Rapids, MI: Baker Book House, 1985.(This was originally published as *Syllabus and Notes of the Course of Systematic and Polemic Theology Taught in Union Theological Seminary, Virginia*; the Baker volume is reprinted from the revised edition of 1878.)

Smith, W.D. (1854) *What is Calvinism?* Philadelphia, PA: Presbyterian Board of Publication.(This is a highly polemical defence of Presbyterian Calvinism.)

(1646) *Westminster Confession of Faith*, in P. Schaff (ed.) *The Creeds of Christendom*, New York: Harper & Brothers, 1919, vol. 3, 676-703.(A confessional statement which has been very widely used in Presbyterian churches; see page 604 for the quotation in §3.)

(1648) *Westminster Shorter Catechism*, in P. Schaff (ed.) *The Creeds of Christendom*, New York: Harper & Brothers, 1919, vol. 3, 600-73.(The definition in §1 is from Question 7, on page 677.)

Predicate Calculus

The central feature of the 'predicate calculus' is a formal language designed to represent arguments involving generalizations. Aristotle had developed a formal logic of 'every', 'some' and 'no', but did not tackle the problem of combining these quantifiers (as they are now called) with relations, or with complex predicates, as in 'Everyone is F or G'. The Stoics had the beginnings of a sentential logic - the logic of the connectives 'and', 'or', 'if' and 'not' - which could later be adapted to deal also with complex predicates. The solution is not, however, a simple matter of pasting these two systems together. It involved four major steps. One was Boole's groundwork for a mathematical system of sentential logic by analogy with an algebra of truth and falsity. Another was De Morgan's move away from the traditional subject-predicate paradigm towards a relational one, so that relations of the form '__ loves __' could be treated naturally alongside predicates such as '__ is mortal'. The third was the invention by Peirce and Frege of the quantifier-and-variable notation. This solved the twin difficulties of adapting sentential logic to express complex predicates, as in $(\forall x)(Fx \vee Gx)$*, and linking the quantifiers unambiguously to the right clauses in multiply general sentences, for example, '* $(\forall x)(\exists y)$*x loves y' and '* $(\exists y)(\forall x)$*x loves y' as alternative renderings of 'Everyone loves someone'. Finally there was Frege's break with the traditional semantics of quantified terms. Instead of treating 'every man' as the subject of 'Every man is mortal', he treats it as a second-level predicate governing the first-level predicate 'is mortal'. At one stroke this obviates the futile search for an entity answering to 'everyone' and provides the rationale for disambiguating multiply general sentences.*

Frege and Peirce catered for quantification not merely over individuals but over properties, relations and functions of individuals, and Whitehead and Russell carried this further to cover quantification over relations of relations, and so on. These would now be called second-order or higher-order predicate calculi as opposed to the first-order predicate calculus - often simply called 'the' predicate calculus - with which this entry is primarily concerned.

One can give a mathematically precise characterization of the valid argument forms of the predicate calculus. That has become a kind of standard for demonstrating the logical validity of informal deductive arguments, and the semantics of the predicate calculus is the starting point for contemporary philosophy of language. Moreover, the analysis of argument that leads to and incorporates the predicate calculus is the most successful piece of philosophical analysis completed in modern times, and as such serves as a paradigm - for some to be emulated, for others to be avoided, but in one way or another central to the development of contemporary philosophy.

This entry concentrates on the central theme of the analysis of logical validity through an account of what it is for an argument 'form' of the predicate calculus to be valid.

1 Language of the predicate calculus

The notation used here is that of the Hilbert school. A commonly used alternative, derived from Peano and Russell, employs $\sim$ for negation, & for conjunction, $\supset$ for the conditional, $\equiv$ for the biconditional and $(\ )$ for the universal quantifier.

We build up the formulas of the predicate calculus from atomic formulas using the sentential connectives and quantifiers. To start with we need a stock of enough variables: $v_0, v_1, \ldots, v_n, \ldots$. The language may contain constant symbols and relation symbols. (One can also allow function symbols, which would be necessary, for example, for a standard treatment of arithmetic.) Let the constant symbols be $c_0, c_1, \ldots, c_n, \ldots$. Although the sentential connectives and quantifiers are often referred to as logical constants, in this entry, 'constant' always means a non-logical symbol serving as a name or potential name for an individual. Let the relation symbols be $R^n_0, R^n_1, \ldots, R^n_m, \ldots$. Those with a superscript of *n* are called *n*-placed and will be used for *n*-ary relations. One-placed or 'monadic' relation symbols may be called predicate symbols. It is convenient to bring the sentence letters of the sentential calculus (see §3) under this uniform notation by regarding them as zero-placed relation symbols. A 'term' is either a variable or a constant. (It will be assumed here that the language of the predicate calculus contains, as well as any non-logical relation symbols, the equality sign '=' used to denote the logical relation of identity. Some writers would call this 'predicate calculus with identity', keeping 'predicate calculus' for the version in which the equality sign is absent or in which identity is not given the special status of a logical

relation.)

An ‘atomic formula’ is an expression either of the form $s = t$, where *s* and *t* are terms, or of the form $R^n v_1 v_2 \ldots v_n$, where $v_1, \ldots, v_n$ are terms, for some *n*. For example, $R^2 vc$ might be used for ‘*v* is the daughter of Margaret Sanger’, where R^2 stands for the daughter relation and *c* stands for Margaret Sanger. (Here and elsewhere we shall omit the official apparatus of subscripts for the sake of simplicity of exposition.)

The constants and relation symbols function here as schematic letters that may receive varying interpretations, but there is nothing to block fixing the interpretations of some of them in a definite way, which results in an interpreted language of a sort more closely related to ordinary language. The system with only schematic letters, those without a fixed interpretation, is known as the ‘pure’ predicate calculus, while systems with symbols that have fixed interpretations are known as ‘applied’ predicate calculi. Frege considered only applied predicate calculi without any schematic letters. Hilbert introduced and emphasized the importance of the device of schematic letters.

If ϕ and ψ are formulas, then so are those expressions formed as indicated using the sentential connectives: $\neg\phi$, $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \to \psi$ and $\phi \leftrightarrow \psi$. If ϕ is a formula and *x* is a variable, then $(\forall x)\phi$ and $(\exists x)\phi$ are *quantified* formulas and the occurrence of ϕ is said to be the ‘scope’ of the quantifier. The symbols are formal counterparts of the English ‘not’, ‘and’, inclusive ‘or’, ‘if… then’, ‘if and only if’, ‘for all’, and ‘there exists at least one’, respectively. An occurrence of a variable *x* in a formula that is not in the scope of any quantifier over *x* (that is, a quantifier of the form $(\forall x)$ or $(\exists x)$) in the formula is said to be a ‘free’ occurrence, and a quantifier over the variable *x* in a formula is said to ‘bind’ every occurrence of *x* free in its scope. Such an occurrence (including the occurrence of *x* inside the quantifier itself) is then said to be ‘bound’ by the quantifier.

Some formulas are incomplete expressions, such as $R^2 vc$, which might be used for ‘*v* is a daughter of Margaret Sanger’, in which *v* is left free, while others are complete sentences and may be true or false on their own, such as $(\exists v)R^2 vc$, ‘Margaret Sanger has a daughter’ (literally, ‘There exists *v* such that *v* is a daughter of Margaret Sanger’). That distinction can be made precise as follows: a variable ‘occurs free’ in a formula ϕ just if any of its occurrences in the formula is free. A ‘sentence’ is a formula in which no variable occurs free. Thus, while the sentential connectives may on occasion indeed connect sentences, they may also connect formulas that are not sentences.

2 Logically valid arguments and interpretations

An argument is ‘valid’ if it is impossible that the premises be true and the conclusion false. Validity is the central feature of good deductive arguments, for if an argument is valid and the premises true, it follows that the conclusion must also be true. Thus, valid arguments are exactly those suitable for what may be the paradigm application of deductive arguments: establishing the truth of the conclusion on the basis of the truth of the premises. The conclusion of a valid argument is said to be a ‘tautological consequence’ of the premises.

There are other desirable features for an argument to have: to establish the truth of its conclusion, the premises must be true; and to be part of an explanation of its conclusion, its premises must in some sense be better known or understood than the conclusion, and, in particular, the argument must not be circular. We do not discuss such secondary features further here. For the most part, they are only requisites of arguments employed for specific purposes, unlike validity, which is a condition for acceptability of any deductive argument. The notion of validity relies on a notion of possibility that is surely in need of further analysis (but see below).

There are arguments whose validity is not purely logical. For example, ‘Caila is a female child, therefore Caila is a girl’ is valid because a female child is necessarily a girl, although that fact does not appear in the argument. This argument has the ‘form’ ‘ $F(c) \wedge C(c)$, therefore $G(c)$’. But so does ‘Caligula was a famous caesar, therefore Caligula was a god’, which is clearly not valid, since the premise is true and the conclusion false. In contrast, the argument ‘Everyone is mortal, therefore Socrates is mortal’ is itself valid and so is every other argument of the form ‘ $(\forall x)M(x)$, therefore $M(s)$’. Here the validity of the argument does not depend on anything beyond its form, and in particular not on its subject matter: the reasons for its validity are ‘topic neutral’. That has often been taken to be a criterion for something to be logical. (See Haack (1978) for a historical summary of views on what counts as logical, and Sher (1991) and Koslow (1992) for recent work on the subject.)

The primary business of logic is to characterize those arguments which are not only valid, but of a form such that

every argument of that form is also valid; briefly, those which are 'logically valid'. It is obvious that logical validity depends on a prior notion of form (see Logical form). The sentences of the pure predicate calculus are designed to serve as representatives of forms of sentences of natural language, and the notion that an argument is of a certain form can be made more precise using the idea of a 'natural-language interpretation'. For example, 'Caila is a female child, therefore Caila is a girl' is of the form '$F(c) \wedge C(c)$, therefore $G(c)$' under the interpretation that interprets '*c*' as 'Caila', '*F*' as 'is female', '*C*' as 'is a child' and '*G*' as 'is a girl'; but it is also of the form '*A*, therefore *B*' under the interpretation that interprets '*A*' as 'Caila is a female child' and '*B*' as 'Caila is a girl'.

The same argument may have many forms. In general, a natural-language interpretation of an argument form must specify a name or definite description for each constant and a predicate for each predicate symbol. A predicate amounts to little more than a sentence with a single blank in it: '__ is female'. Each *n*-placed relation symbol will similarly be interpreted using an *n*-ary 'sentence form' - a sentence with *n* blanks in it. A binary relation symbol might be interpreted using, for example, '__ is spicier than __'. One needs some mechanism for specifying which term following a relation symbol corresponds to which blank in the sentence form that is its interpretation, but such details will be omitted here.

It follows from the stipulations above that an argument is logically valid if and only if it is of a form that yields a valid argument under all natural-language interpretations. Such a form will be called a valid argument form, and the problem of determining which arguments are logically valid thus amounts to that of determining which argument *forms* are valid.

It is tempting to suppose that an argument form is valid just in case it has no natural-language interpretation that yields an argument with true premises and a false conclusion. That would eliminate the troublesome notion of possibility from the basis of logic. Indeed, Carnap (1934) made proposals along closely related lines. Unfortunately, such a proposal does not serve as an adequate criterion for validity: as Tarski (1936) pointed out, it might be that our language happened not to be rich enough to give rise to the requisite variety of interpretations, and thus the validity of an argument could depend on the richness of parts of the language that were not used in the argument in question. It can in fact be shown that the language of arithmetic is rich enough to yield all the interpretations one could need for the predicate calculus presented here (see Kleene 1952 and Hasenjäger 1953), but it would be question-begging to assume it at the present stage.

3 Sentential logic

Since validity is defined in terms of whether or not the premises of an argument form can be true while the conclusion is not, it is first necessary to specify under what circumstances a sentence is true. We assume every sentence has a 'truth-value' of either T or F, where F is anything other than true - which may include, depending on one's view, not only false, but also meaningless, undefined and perhaps other possibilities. (Any sentence that is not true will be referred to, briefly though perhaps inaccurately, as false. For calculi allowing more than these two truth-values, see Many-valued logics.)

It is illuminating to begin with the simple case of 'sentential logic' (also called the propositional calculus); that fragment of the predicate calculus without variables, constants or quantifiers, and in which one only allows zero-placed relation symbols. According to the definitions given in §2, an atomic sentence is formed by such a symbol followed by zero terms. It thus amounts to no more than a single 'sentence letter'. It is possible to specify the form of some arguments in sentential logic. For example, 'Jane Austen and Charlotte Brontë were novelists, therefore Jane Austen was a novelist' has the form '$A \wedge B$, therefore *A*'. But many arguments cannot be formalized in sentential logic so as to be valid. For example, 'Everyone is mortal, therefore Hannah Arendt is mortal', has the form '$(\forall v)R^1(v)$, therefore $R^1(c)$', but this is not a form available in sentential logic. The only form available in sentential logic is the invalid '*A*, therefore *B*'.

The truth-value of a sentence of sentential logic is determined by an assignment of truth-values to its constituent sentence letters, through the following inductive rules. The negation of a sentence is false if the sentence is true, and true if the sentence is false. The conjunction of two sentences is true if both are true, and false otherwise. The disjunction of two sentences is true if at least one is true, and false otherwise. A conditional $\phi \to \psi$ is true unless ϕ is true and ψ is false. A biconditional between two sentences is true if both sentences have the same truth-value

and false otherwise. The idea of treating the truth-value of complex sentences as a function of the truth-values of their components (see Compositionality) - the idea of 'truth-functional' connectives - dates back to the Stoics (see Logic, ancient §5). The treatment of negation, conjunction and disjunction as truth-functional is scarcely controversial, but the same cannot be said of the truth-functional treatment of the conditional and biconditional (see Counterfactual conditionals; Indicative conditionals).

The rules just given can be represented diagramatically in the following 'truth table':

ϕ	ψ	$\neg\phi$	$\phi\wedge\psi$	$\phi\vee\psi$	$\phi\rightarrow\psi$	$\phi\leftrightarrow\psi$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

Let us use the table to compute the truth-value of $((A\leftrightarrow B)\wedge A)\rightarrow(\neg B)$, given values for *A* and *B* of T and F, respectively. The truth-value of $A\leftrightarrow B$, according to the table, is F. But then, applying the rule for $\wedge$ to that and the given value for *A*, the truth-value for $(A\leftrightarrow B)\wedge A$ is also F. Since the truth-value of *B* is F, that of $\neg B$ is T. Finally, applying the rule for $\rightarrow$ to the last two facts, the truth-value of the whole sentence is T. This computation is represented by the following line of a truth table.

A	B	$A\leftrightarrow B$	$\neg B$	$(A\leftrightarrow B)\wedge A$	$((A\leftrightarrow B)\wedge A)\rightarrow(\neg B)$
T	F	F	T	F	T

A sentence that is assigned the value T on every line of its truth table is known as a 'tautology', a term introduced by Wittgenstein, who emphasized the importance of the method of truth tables.

Suppose Γ is a finite set of sentences and ϕ a sentence of sentential logic. It is then possible to construct a truth table that lists the truth-values of ϕ and all the sentences in Γ under all possible combinations of truth assignments to the sentence letters they contain. Now, suppose there is no line of that truth table in which all the sentences in Γ receive the truth-value T and the sentence ϕ receives the value F. Then the corresponding argument form 'Γ, therefore ϕ' is valid. For consider an arbitrary natural-language interpretation. The argument it yields must be valid - if not it would have to be possible for the premises to be true and the conclusion false. But that is not possible, by hypothesis. On the other hand, suppose that there is a line of the table on which all the sentences in Γ are true and ϕ is false. Then the argument form is not valid. For choose a line of the table of the type just assumed to exist. Consider a natural-language interpretation of the argument form that interprets each sentence letter by 'Snow is black' if it is assigned F on the selected line of the table, and by 'Snow is white' if it is assigned T on the selected line. The argument yielded by the interpretation just specified is one in which the premises are actually true and the conclusion is actually false. It is therefore not valid, and so the argument form is not valid.

This discussion demonstrates a remarkable fact: truth tables provide a precise criterion for whether or not an argument form in sentential logic is valid - it is valid just if there is no line of the table on which all the premises have value T and the conclusion has value F. (The troublesome notion of possibility has dropped out of the picture, since all that turned out to be relevant about possibility was the combinatorial list of possible ways for various simple sentences to be true or false.) Truth tables in fact constitute a mechanical decision procedure for whether an argument form of sentential logic is valid. For one can generate the table mechanically and then check it line by line.

Though truth tables provide a decision procedure in principle, it soon ceases to be a practically feasible one, since the number of lines of a truth table doubles with the addition of each sentence letter. The question of whether there is a more efficient procedure, in the sense that the difficulty of the computation grows as, for example, a polynomial function of the number of letters rather than an exponential one, is a version of perhaps the most important question in the theory of computational complexity (see Garey and Johnson 1979).

4 Expressive completeness

It might seem that the logic developed here is unduly limited in that it allows only five connectives. What about, for example, exclusive 'or', or the majority-of-three connective that holds among three sentences if and only if at

least two of them are true? The 'expressive completeness theorem' for sentential logic shows that the limitation is only an apparent one: every truth-functional connective, specified by any truth table whatever, is definable using those already introduced.

The proof of 'expressive completeness' (often called functional completeness) can be illustrated using the majority-of-three connective as an example. The sentence formed by applying it to *A*, *B* and *C* has the same truth table as

$$(A \wedge B \wedge \neg C) \vee (A \wedge \neg B \wedge C) \\ \vee (\neg A \wedge B \wedge C) \vee (A \wedge B \wedge C),$$

a disjunction whose components correspond to the four possible ways in which the majority-of-three connective creates a true sentence, these being read off the lines of the truth table. This example is in 'disjunctive normal form', that is, a disjunction of conjunctions of sentence letters and their negations. One can alternatively express an arbitrary truth table by a sentence in 'conjunctive normal form', that is, a conjunction of disjunctions of sentence letters and their negations. (The majority-of-three connective becomes $(A \vee B) \wedge (B \vee C) \wedge (C \vee A)$.) Any truth table can similarly be expressed using only 'and', 'or' and 'not'.

It follows that any set of connectives that can be used to define 'and', 'or' and 'not' is expressively complete. It is an elementary exercise to show in this way that negation plus any one of $\vee$, $\wedge$ or $\rightarrow$ is expressively complete. The binary connective that forms a true sentence if and only if both components are false ('not or') is also expressively complete, as is the one that forms a false sentence if and only if both components are true ('not and'). These last facts were first published by Sheffer (1913), and expressively complete sets of connectives were classified by Post (1941) and Wernick (1942).

5 Truth and logical consequence

The next task is to extend the account of the truth-conditions of sentences to the full predicate calculus. The truth-value of an atomic sentence is determined by whether the objects denoted by the constants stand in the relation denoted by the relation symbol. This can be expressed mathematically by means of an 'interpretation function' that associates an object with each relevant constant symbol and a set of ordered *n*-tuples of objects with each relevant *n*-placed relation symbol. (Though natural-language interpretations and interpretation functions play similar roles - they fix how the constant and relation symbols are interpreted - they are not the same: a natural-language interpretation takes each symbol to a suitable item of natural language, while an interpretation function takes each symbol to a suitable set-theoretic object.) For example, the relation on persons specified by '__ is shorter than __' could be symbolized using R^2, which would then be taken by the corresponding interpretation function to the set *S* of all ordered pairs of persons in which the first person is shorter than the second. If *c* were to be interpreted as G.E. Moore and *d* as Kareem Jabbar, then R^2cd would take on the value T precisely because the ordered pair $\langle$G.E. Moore, Kareem Jabbar$\rangle$ is in the set *S*. The present method goes beyond that of natural-language interpretations in an important respect: an interpretation function can take a relation symbol to an arbitrary collection of ordered *n*-tuples, even though it might not be specifiable in ordinary language.

In an applied predicate calculus, one selects certain constant and relation symbols to stand for certain objects and sets of *n*-tuples, reserving the rest to function as schematic letters. One way of introducing integer arithmetic, for example, involves fixing symbols for zero, 'less than' and so on. One can then add to the axioms of a logical system special nonlogical axioms involving those symbols and even restrict the language to eliminate schematic letters.

The truth-value of a universally quantified sentence, say $(\forall v)R^2cv$ (which, on the above interpretation, reads 'G.E. Moore is shorter than everyone'), would be easy to characterize if every relevant object were the interpretation of some constant symbol. It would then be true if all the following sentences were: $R^2cc_0, R^2cc_1, R^2cc_2, \ldots$. But there is no reason to suppose that every relevant object has a name. Consider 'All ravens are black' - surely not every raven has a name. And certainly, no one language could name all the uncountably many real numbers. Nonetheless, each relevant object could be given a name, and so the following works. First, specify a non-empty set to be the domain of the quantifiers (persons, ravens, real numbers,…). Second, specify a new constant that has not been employed in any of the sentences under consideration. For our

example, the domain is to be persons and the constant might be *d*. Then, continuing the example, $(\forall v)R^2cv$ is true if R^2cd is true no matter how the interpretation so far specified is expanded by interpreting the constant *d* as standing for a member of the domain.

It is now possible to give a precise definition of how the truth of sentences is determined. A 'language' (sometimes called a signature or similarity type) is a set of constant and relation symbols. (A language might better be called a vocabulary, but the terminology is standard.) A 'structure' (or 'interpretation') for a language $\mathcal{L}$ is an ordered pair with first member a non-empty collection - the 'universe' of the structure - and second member a suitable (in a sense defined immediately below) function with domain $\mathcal{L}$ - the 'interpretation function' of the structure. (The language is thus determined by the structure, being the domain of the interpretation function.) The interpretation function is required to map every constant symbol in $\mathcal{L}$ to a member of the universe, every sentence letter to a truth-value, every predicate symbol to a set, and every *n*-placed relation symbol, for *n* greater then 1, to a collection of ordered *n*-tuples of members of the universe.

When symbols with a fixed usage are in the language $\mathcal{L}$, one must require that the relevant objects be in the universe and that the interpretation function map such symbols to the correct objects and relations. We can avoid such complications by considering only languages that contain nothing but schematic letters. Arithmetic, for example, is then considered in terms not of a particular language but of a particular structure, and the axioms of arithmetic, because they will be written using schematic letters, will have interpretations and interpretation functions in addition to the main one.

If A and A' are structures with the same universe such that the language of A' includes the language $\mathcal{L}$ of A and possibly some other symbols, and if the interpretation functions of the two structures agree on $\mathcal{L}$, then A' is said to be an 'expansion' of A. When one structure is an expansion of another it differs only in that it may possibly expand the interpretation function to contain some additional symbols in its domain. Note that if *d* is some definite constant symbol not in $\mathcal{L}$, then for every member *a* of the universe of A there is an expansion in which the interpretation of *d* is *a*.

The truth-value of a sentence in a structure is defined by the following conditions on all structures A and sentences ϕ of the language of A:

(1) (a) If ϕ is of the form $c_i = c_j$, then the truth-value of ϕ in A is T if the interpretations of c_i and c_j in A (that is, the values of the interpretation function on c_i and c_j) are the same, and the truth-value is F otherwise.
(b) If ϕ is a sentence letter, its value in A is the truth-value to which it is mapped by the interpretation function.
(c) If ϕ is of the form $R^n c_1 \ldots c_n$ with $n > 0$, then the truth-value of ϕ in A is T if the *n*-tuple formed from the interpretations of $c_1, \ldots, c_n$ (in that order) is a member of the interpretation of R^n (which, recall, is a set of *n*-tuples) and it is F otherwise.
(2) If ϕ is a negation, conjunction, disjunction, conditional or biconditional, then its truth-value is determined from those of its constituents just as in sentential logic.
(3) If ϕ is of the form $(\forall x)\psi$, $\mathcal{L}$ is the language of A and $\mathcal{L}'$ is the language obtained from $\mathcal{L}$ by adding some definite constant symbol *d* not in $\mathcal{L}$, then the value of ϕ is T if the truth-value of the sentence $\psi(d)$ obtained from ψ by replacing every free occurrence of *x* by *d* is T in every expansion of A to a structure for $\mathcal{L}'$. The truth-value is F otherwise.
(4) If ϕ is of the form $(\exists x)\psi$, $\mathcal{L}$ is the language of A and $\mathcal{L}'$ is the language obtained from $\mathcal{L}$ by adding some definite constant symbol *d* not in $\mathcal{L}$, then the value of ϕ is T if there is at least one expansion of A to a structure for $\mathcal{L}'$, in which the truth-value of the sentence $\psi(d)$ obtained from ψ by replacing every free occurrence of *x* by *d* is T. It is F otherwise.

It is now possible to define, as Tarski did (1936), logical consequence and logical truth for the predicate calculus in a way parallel to the definitions of tautological consequence and tautology given above. Let the language $\mathcal{L}$ contain only schematic letters. Let Γ be a set of sentences of $\mathcal{L}$ and ϕ a sentence of $\mathcal{L}$. Then ϕ is a 'logical consequence' of Γ if there is no structure for $\mathcal{L}$ in which every sentence in Γ has value T and ϕ has value F; and ϕ is a 'logical truth' if it receives value T in every structure for $\mathcal{L}$. As in the case of sentential logic, one can establish that an argument formalized in the predicate calculus is logically valid just if the formalization of the conclusion is a logical consequence of the formalizations of the premises. In this characterization of logical

validity, the precise mathematical notion of a structure has replaced the troublesome notion of logical possibility.

For technical variants of the semantics presented here, including the idea of extending the language by giving the existing variables the role played here by the added constants, see Tarski (1936) and the textbooks by Mates (1965) and Tennant (1978). The customary semantics excludes the possibility of an empty universe or of constants that do not denote any individual, but both can be accommodated at the cost of some complication (see Free logics).

It is immediate from the definitions that in any structure, if *c* and *d* denote the same individual and ϕ is true, then so is the result of substituting *d* for any or all occurrences of *c* in ϕ. Likewise, relation symbols denoting the same set of ordered *n*-tuples, predicates denoting the same set, and sentences with the same truth-value, can all be substituted for one another without changing the truth-value. In short, the predicate calculus is 'extensional' through and through. Some natural-language forms are extensional, others not. Since 'Fibonacci' and 'Leonardo of Pisa' denote the same man and Fibonacci was a mathematician, it follows that Leonardo was a mathematician; but someone may know that Fibonacci was a mathematician without knowing that Leonardo was, so 'knows that' generates forms that are not extensional. The predicate calculus is not adequate to deal with such cases (see Intensional logics), but even within the extensional sphere, there is no counterpart for the predicate calculus of the expressive completeness theorem for sentential logic. It will in general be true that for a given structure there will be relations on the structure not definable in the language of the structure, structures indistinguishable by any sentence, and more classes of structures for a language than there are sentences of the language.

As for the definability of quantifiers, many are indeed definable with the help of the equality sign, including 'there are *n* things such that' for each finite number *n*. But the quantifier 'for finitely many' cannot be defined in the first-order predicate calculus. It requires the resources of second-order predicate calculus, in which the quantifiers range over relations of objects as well as over individual objects (see Second-and higher-order logics).

6 Formal deduction and decidability

As an alternative to the semantic approach adopted above, the inferential role of the various logical constants can be specified by codifying a system of rules for deduction and proof. The idea goes back to Aristotle, though the first fully rigorous system is that of Frege (1879), with his insistence that there be absolutely no room for doubt or disagreement whether a putative proof is in accordance with the rules or not. Some have gone further and taken purely syntactic rules for the manipulation of the logical constants to be constitutive of their meanings, and systems of this sort have been used to introduce intuitionistic logic as well as the 'classical' logic expounded here (see Intuitionism; Natural deduction, tableau and sequent systems).

There are many varieties of deductive system (see Formal languages and systems). The axiomatic variety is conceptually the simplest. A number of logical axioms are laid down, together with rules of inference of the form 'From $\phi_1, \ldots, \phi_n$ infer ψ' for suitably related ϕs and ψ. A common choice is to take the rule of *modus ponens*, 'From ϕ and $\phi \rightarrow \psi$ infer ψ', as the sole rule of inference, supplemented by numerous axioms to offset the absence of other rules. A formal deduction of a conclusion from a set of premises is defined as a finite sequence of sentences which ends with the conclusion and of which every member is either a logical axiom or a premise or else follows immediately from preceding members by a rule of inference. Formal proof is introduced as the limiting case of a deduction from the empty set of premises. A sentence is said to be 'deducible' from others in the system if there exists a formal deduction of it from them, and a sentence is a theorem if there exists a formal proof of it. Church (1956) gives a survey of axiomatic systems for sentential logic and for predicate logic. The latter generally invoke an extended deductive apparatus that covers formulas with free variables as well as sentences proper, but Quine (1955) presents an axiomatic system confined strictly to sentences.

Natural deduction systems include a more sophisticated kind of rule, designed to exploit assumptions - propositions assumed for argument's sake and treated as extra premises for as long as they are wanted. Proof by *reductio ad absurdum* is a good informal illustration of the idea: wishing to prove ϕ, one supposes for argument's sake that it is false and tries to derive a contradiction, whereupon one discards the assumption and infers that ϕ must be true. The definition of formal deduction now has to keep track of the introduction and subsequent discharge (discarding) of assumptions; see, for example, Lemmon (1965). This complication is, however, offset by there being fewer or even no axioms, and by the comparative naturalness and brevity of

deductions. (For this and some related kinds of deductive system see Natural deduction, tableau and sequent systems §1.)

To be adequate for its purpose, any deductive system must satisfy two conditions. It must be ‘sound’, in the sense that it does not license deductions whose conclusions do not follow logically from their premises. And it must be ‘complete’, that is, strong enough to ensure that every logical consequence of a set of sentences is deducible from them. When a system is both sound and complete, deducibility becomes equivalent to logical consequence and likewise theoremhood is equivalent to logical truth. Of the two conditions, completeness is nearly always much more difficult to establish. The first proof of the completeness of a deductive system for the predicate calculus was given by Gödel (1930; see also Model theory §4), while Henkin (1949) devised a novel method of proving completeness that has found applications far beyond its original one to the predicate calculus. Although second- and higher-order predicate calculi have significantly more expressive capacity than first-order, their great drawback is that there can never be a sound and complete deductive system for them (see Second- and higher-order logics §1). The dominant position of the first-order calculus is the result of its being rich enough to formalize a vast range of arguments from mathematics and elsewhere, yet simple enough for all its valid argument forms to be reducible to a few simple rules of inference.

If a sentence of the predicate calculus is a logical consequence of others, that fact can be shown by deriving it from them in a sound and complete deductive system. Although in practice this may require considerable ingenuity, in principle a machine can generate a list of successive finite sequences of sentences on which sooner or later the right one - by hypothesis there is a right one - will turn up. Nonetheless, the method of formal deduction does not provide a decision procedure for logical consequence. The fact that a deduction has not yet turned up means nothing. For there is in general no way of telling whether this is because no such deduction exists - that is, the conclusion in question is not a logical consequence of the premises - or whether it is merely that we have not tried hard or long enough. If we could find a counterexample - a structure in which the putative conclusion is false while the premises are true - that would indeed show that it is not a logical consequence of them. But the parallelism between truth tables and structures breaks down at this point. In sentential logic all one has to do is set out the relevant truth table and inspect its rows. There will, however, always be infinitely many relevant structures, not to mention structures with infinitely many objects. The analogue of setting out and inspecting the rows of a truth table thus involves infinitely many steps and cannot be carried out even in principle.

This strongly suggests the absence of a decision procedure for the predicate calculus, and Church has proved that none exists. However, positive results can be obtained in some restricted cases, the simplest and most important example being that there is a decision procedure for the monadic predicate calculus, that is, the fragment whose language includes predicate symbols but no *n*-placed relation symbols for $n > 1$. (See Church’s theorem and the decision problem.) For further results of this sort, see Dreben and Goldfarb (1979). Many interesting questions concerning feasibility and quantificational logic arise (see Börger 1989).

See also: Logical and mathematical terms, glossary of

SHAUGHAN LAVINE

References and further reading

Börger, E. (1989) *Computability, Complexity, Logic*, New York: North Holland.(A comparatively accessible treatment of the areas of research on the border between mathematical logic and computational complexity theory.)

Carnap, R. (1934) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *Logical Syntax of Language*, London: Kegan Paul, 1937.(An influential discussion of the relation between ordinary language and the sort of formal language introduced here, the notions of logical truth and consequence and much else. Carnap (1935) discusses many of the same ideas in a less technical way.)

Carnap, R. (1935) *Philosophy and Logical Syntax*, London: Kegan Paul, Trench, Trubner & Co.(A semi-popular account of Carnap’s attitude towards philosophy. A clear and useful example of one important sort of impact the predicate calculus has had on philosophers for much of the twentieth century.)

Carnap, R. (1942) *Introduction to Semantics*, Cambridge, MA: Harvard University Press.(An influential later discussion of the relation between ordinary language and the sort of formal language introduced here, the notions of logical truth and consequence, and much else.)

Church, A. (1956) *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press.(An influential,

exceptionally rigorous introduction to the predicate calculus, with a wealth of historical information.)

Dreben, B. and Goldfarb, W.D. (1979) *The Decision Problem: Solvable Classes of Quantificational Formulas*, Reading, MA: Addison-Wesley.(A detailed analysis of which fragments of the predicate calculus are simple enough have a decision procedure.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82.(The classic source for the predicate calculus, in the form of second-order predicate calculus.)

Garey, M.R. and Johnson, D.S. (1979) *Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$-Completeness*, San Francisco, CA: Freeman.(A very readable account of why truth tables are not a practical decision method, how the notion of a practical method can be characterized, and related problems of feasibility.)

Gödel, K. (1930) 'Die Vollständigkeit der Axiome des logischen Funktionenkalküls', *Monatshefte für Mathematik und Physik* 37: 349-60; trans. 'The Completeness of the Axioms of the Functional Calculus of Logic', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 582-91.(Cited in §5. The first proof of the completeness of an axiom system for the predicate calculus.)

Haack, S. (1978) *Philosophy of Logics*, New York: Cambridge University Press.(A thorough and engaging synoptic account of many of the issues raised in this entry and of alternatives and generalizations of the logic introduced here.)

Hasenjäger, G. (1953) 'Eine Bemerkung zu Henkins Beweis für die Vollständigkeit des Prädikatenkalküls der ersten Stufe', *Journal of Symbolic Logic* 18: 42-8.(Includes one of the first proofs that if a sentence of the predicate calculus has a model, it has one with domain the natural numbers and in which all the predicates are definable in the language of arithmetic.)

Henkin, L. (1949) 'The Completeness of the First-Order Functional Calculus', *Journal of Symbolic Logic* 14: 159-66.(Cited in §5.)

Hilbert, D. and Ackermann, W. (1928) *Grundzüge der theoretischen Logik*, Berlin: Springer, 2nd edn, 1938; trans. L.M. Hammond, G.G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, ed. R.E. Luce, New York: Chelsea, 1950.(A classic introduction to the predicate calculus, with a wealth of detail not available in later texts.)

Kleene, S.C. (1952) *Introduction to Metamathematics*, Amsterdam: North Holland, and New York: Van Nostrand. (Perhaps the most influential modern text on the mathematics of the predicate calculus.)

Koslow, A. (1992) *A Structuralist Theory of Logic*, New York: Cambridge University Press.(An insightful attempt to characterize what is logical in terms of an analysis of the consequence relation.)

Lemmon, E.J. (1965) *Beginning Logic*, London: Nelson.(Cited in §5. An exceptionally user-friendly exposition of a natural deduction system for the predicate calculus.)

Mates, B. (1965) *Elementary Logic*, New York: Oxford University Press.(Cited in §4.)

Peirce, C.S. (1931-58) *Collected Papers*, vol. 2, *Elements of Logic*, vol. 3, *Exact Logic*, ed. C. Hartshorne and P. Weiss, Cambridge, MA: Harvard University Press.(Peirce was one of the inventors of quantifiers, and he coined the term.)

Post, E.L. (1941) *The Two-Valued Iterative Systems of Mathematical Logic*, Princeton, NJ: Princeton University Press.(Includes an analysis of the complete sets of connectives for sentential logic.)

Quine, W.V. (1940) *Mathematical Logic*, Cambridge, MA: Harvard University Press, 1955.(Cited in §5.)

Sheffer, H.M. (1913) 'A Set of Five Independent Postulates for Boolean Algebras, With Application To Logical Constants', *Transactions of the American Mathematical Society* 14: 481-8.(The first publication of the fact that there is a single connective that is expressively complete for sentential logic.)

Sher, G. (1991) *The Bounds of Logic: A Generalized Viewpoint*, Cambridge, MA: MIT Press.(A thorough working out of the Tarskian idea that what counts as logical should be characterized in terms of what can be characterized using semantics of the sort described in §4.)

Tarski, A. (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956), 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(One of the most influential works of the century; includes the definition of truth presented here.)

Tarski, A. (1936) 'O pojęciu wynikania logicznego', *Przegląd Filozoficzny* 39: 58-68; trans. J.H. Woodger (1956), 'On the Concept of Logical Consequence', in *Logic, Semantics, Metamathematics: Papers from 1923*

to 1938, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(The definitive formulation of the semantic conception of consequence.)

Tennant, N. (1978) *Natural Logic*, Edinburgh: Edinburgh University Press.(Cited in §4.)

Wernick, W. (1942) 'Complete Sets of Logical Functions', *Transactions of the American Mathematical Society* 51: 117-32.(An analysis of the complete sets of connectives for sentential logic.)

Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press; 2nd edn, 1925; repr. 1994.(This is the work that established that, in a suitable sense, all of classical mathematics can be formalized within the higher-order predicate calculus.)

Predication

Some sentences have a very simple structure, consisting only of a part which serves to pick out a particular object and a part which says something about the object picked out. Expressions which can be used to say something about objects picked out are called predicates. Thus 'smokes' in 'Sam smokes' is a predicate. But 'predication' may refer either to the activity of predicating or to what is predicated. To understand either we need to know what predicates are and how they combine with other expressions.

Predicates, unlike proper names, can be negated. They combine with other expressions in ways described by categorial grammar: predicates are incomplete and are completed by other expressions, such as proper names. The word 'smokes' in 'Sam smokes' is called a monadic or 1-place predicate; predicates with two or more places are called relational predicates ('Sam loves Erna', '3 is between 2 and 4').

Since Frege it has been customary to hold that the incomplete or predicative parts of 'Sam smokes' and 'Sam is a smoker' are, respectively, 'smokes' and 'is a smoker' (see Frege, G. §§2-4). According to an older view, the incomplete part of 'Sam is a smoker' is the copula 'is', which is completed by two complete expressions, 'Sam' and 'a smoker' ('a' can be ignored as an accident of English). Sometimes called 'the two name theory of predication', this view allows two types of name - proper names and common nouns.

Is predication merely a matter of words? Are mental acts or the senses of words essential to it? For Frege, names and predicates are correlated with ideal senses, and the sentential wholes to which they belong with ideal thoughts or propositions (see Sense and reference §2; Frege 1892a). The sense or meaning of a proper name is sometimes called an individual concept, that of a predicate a concept (Frege himself used neither 'concept' nor 'meaning' in this way). It has been held that to use a predicate is to perform a mental act of predicating and that to use a proper name is to perform a mental act of naming or referring. Husserl and others appeal to both senses and mental acts and conceive of the latter as tokenings of the former. The unity of thoughts and of thinkings, on these views, resembles but is prior to the unity of the sentence: predicative senses require non-predicative senses; predicatings require referrings or acts of quantifying over individuals.

Predication occurs in the context of a variety of mental acts and linguistic actions: I may wonder or ask whether Sam smokes. So to predicate is not always to judge or assert. To assert that if Sam smokes he will smell is not thereby to assert that Sam smokes, although it is to predicate this of him in the context of entertaining or supposing the thought that Sam smokes.

Frege and Husserl pointed out that to deny that Sam smokes is just to assert that he does not smoke. Denying and judging are not on a par because 'not' belongs within the sentence, its sense to that of the sentence. There is, it is true, the phenomenon of polemic negation, as when the third word of the following sentence is stressed: 'Wittgenstein was not German'. But this is a pragmatic phenomenon.

What is it that is predicated? Just as names designate things, so too predicates are said to stand in a semantic relation (reference, signification) to one or more of the following: properties (attributes); what Frege (in his own technical sense) called 'unsaturated concepts'; sets; and relations between objects and possible worlds (see Semantics, possible worlds §9). Thus predication may be held to involve saying of Sam that he falls under the sense of the predicate 'smokes', exemplifies the property of being a smoker, or belongs to the set of smokers - or some combination of these. The semantic values of predicates and other parts of a sentence are sometimes said to form states of affairs or situations - which may or may not obtain - or facts.

The foregoing contains five simplifications and questionable presuppositions. First, does predication only occur within sentences (thoughts, thinkings)? If a builder shouts 'Slab!' to his assistant is he not producing a non-elliptic predication? Second, we may wonder, with Wittgenstein (1968), to whom the preceding example is due, whether there is in fact any uniform type of semantic relation between predicates, on the one hand, and properties and their ilk, on the other hand.

Third, the apparent (surface, grammatical) forms of predicates can diverge in different ways from their logical (deep) forms. To say that Pierre is tall is not to predicate the monadic property of being tall of him but to predicate,

of him and the average member of some reference class, the relational property of being taller than. Furthermore, if we accept the view (of Davidson; and of Parsons 1990) that many sentences involve quantification over events, then to say that Sam is smoking is to predicate of some event that it is an event of smoking and to predicate of it and Sam that Sam is the agent of it (see Adverbs §1).

Fourth, according to Ramsey (1990) there is no essential distinction, in 'Socrates is wise', between its subject and monadic predicate because the sentence expresses the same proposition as 'Wisdom is a characteristic of Socrates'. So no fundamental classification between individuals and properties can be based upon such a distinction.

Finally, if we take seriously the claim that predicates are incomplete, then we should say that in 'Sam smokes' the predicate is not 'smokes' but ' smokes', the gappy result of deleting the proper name from the sentence. Thus both 'Booth shot Lincoln' and 'Booth shot Booth' contain the predicate 'Booth shot '. But, as Geach indicates in his discussion of this example ('Quine on Classes and Properties', 1981: 222-5), the predicate here is not a physical part of the two sentences; it is a sentence pattern which 'Booth shot' helps to form. A related view is that such sentence patterns are abstract features of sentences, as melodies are of spoken sentences. But then we might wonder whether 'Booth ' is not just as much a sentence pattern as 'Booth shot '.

See also: Logical and mathematical terms, glossary of; Proper Names; Property theory; Strawson, P.F. §5

KEVIN MULLIGAN

References and further reading

Frege, G. (1891) *Funktion und Begriff*, Jena: Pohle; trans. P.T. Geach, 'Function and Concept', in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984, 137-56.(Classic account of the role of functions in predication.)

Frege, G. (1892a) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Meaning', in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984, 157-77.(The classic account of the sense and reference of sentences and their parts.)

Frege, G. (1892b) 'Über Begriff und Gegenstand', *Vierteljahrsschrift für wissenschaftliche Philosophie* 16: 192-205; trans. P.T. Geach, 'On Concept and Object', in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984, 182-94.(Classic account of what predicates and names signify.)

Geach, P.T. (1975) 'Names and Identity', in S. Guttenplan (ed.) *Mind and Language*, Oxford: Clarendon Press, 139-58.(Important treatment of the nature of predication, drawing on Frege.)

Geach, P.T. (1981) *Logic Matters*, Oxford: Blackwell.(See 'History of the Corruptions of Logic' for a historical sketch of the relations between the two name theory and the theory fully worked out by Frege.)

Gardies, J.-L. (1975) *Esquisse d'une grammaire pure*, Paris: Vrin; trans. K. Mulligan, *Rational Grammar*, Munich: Philosophia, 1985.(Useful guide to categorial grammar and its applications to natural languages.)

Parsons, T. (1990) *Events in the Semantics of English*, Cambridge MA: MIT Press.(Elegantly defends the controversial view that many sentences involve quantification over and predication of events and states.)

Ramsey, F. (1990) 'Universals', in *Philosophical Papers*, ed. H. Mellor, Cambridge: Cambridge University Press, 8-30.(Stimulating, critical examination of some ways of distinguishing between subjects and predicates and between particulars and properties.)

Simons, P. (1981) 'Unsaturatedness', *Grazer Philosophische Studien*, 14: 73-95.(Surveys different types of unsaturatedness and incompleteness and their relations to the notions of function, functor and predication.)

Strawson, P.F. (1974) *Subject and Predicate in Logic and Grammar*, London: Methuen.(Surveys fully most aspects of predication.)

Wittgenstein, L. (1968) *Philosophical Investigations*, Oxford: Blackwell.(Includes discussion of a series of examples that throw doubt on the plausibility of many common assumptions about sentences and semantic relations.)

Prescriptivism

Prescriptivism is a theory about moral statements. It claims that such statements contain an element of meaning which serves to prescribe or direct actions. The history of prescriptivism includes Socrates, Aristotle, Hume, Kant and Mill, and it has been influential also in recent times.

Moral statements also contain a factual or descriptive element. The descriptive element of morality differs between persons and cultures, but the prescriptive element remains constant.

Prescriptivism can allow for moral disagreement, and explain moral weakness. It can also explain better than other theories the rationality and objectivity of moral thinking.

1 Definition and history

Prescriptivism is a meta-ethical theory about the meaning of moral statements which holds that, in addition to any factual or descriptive meaning they may have, there is a prescriptive element in their meaning which is irreducible to any factual equivalent, but serves to prescribe or direct actions (see Analytic ethics §1; Moral judgment §1).

The prescriptivity of moral statements has been recognized from the earliest times by philosophers, though it is only in the present century that thinkers have begun to develop a clear account of the relation between the descriptive and prescriptive elements in their meaning. Socrates was in part a prescriptivist, as is evident from the trouble he had in explaining how one could think something good, and yet not pursue it (see especially Plato's *Protagoras* (358) and *Gorgias* (467)) (see Socrates §6). Aristotle, who took up Socrates' problem about weakness of will, showed clearly, by his handling of it, a prescriptivist strand in his theory (*Nicomachean Ethics* 1145b23) (see Akrasia). *Phronēsis* or practical wisdom, which tells us what we ought to do, is described by him as *epitāktikē* or prescriptive (*Nicomachean Ethics* 1143a8) (see Aristotle §23). The practical syllogisms which are the channel of this instruction all have prescriptive first premises (*Nicomachean Ethics* 1144a31, 1147a25; *On the Soul* 434a16; *On the Movement of Animals* 701a7); if they did not, they could not validly conclude in actions. He solved Socrates' problem in a way consistent with his prescriptivism (see Hare 1992a).

Hume also recognized the prescriptive element in the meaning of moral statements (see especially *A Treatise of Human Nature* (1739-40), where he wrongly claims that because of its lack of prescriptivity 'reason is perfectly inert') (see Hume, D. §4, 8); but his settled theory was subjectivist rather than prescriptivist (the two kinds of theory are to this day often confused). He says that 'when you pronounce an action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it' (1739-40). Moral statements are for him reports rather than expressions of feelings or desires.

Kant rescued reason from inertia, and was clearly a prescriptivist; moral facts play no significant part in his ethical theory, the basic constituents of which are maxims, that is to say, prescriptions, originating in the will, and expressed often in the imperative (the German *soll* is equivalent to an imperative) (see Kantian ethics). Echoing the language of Hume about 'is' and 'ought', J.S. Mill avowedly declared himself a prescriptivist by calling moral statements 'imperatives' (1843) (see Mill, J.S. §11). But modern descriptivist opponents of prescriptivism, who often have a blind spot for prescriptivity, have seldom noticed these premonitions of modern prescriptivism.

In recent times many philosophers have held prescriptivist views of more or less considered kinds (such as Carnap 1935: 18; Ayer 1936, 1949; Stevenson 1944) (see Ayer, A.J. §5; Carnap, R. §3; Stevenson, C.L.). These were all followers of emotivism, the view that moral statements get their meaning by expressing feelings or attitudes of approval or disapproval (see Emotivism). But prescriptivists do not have to be emotivists; if they are concerned for the rationality of moral thinking they will not be. It is possible to hold, with Kant, that moral maxims originate in the rational will, which he says 'is nothing but practical reason' ([1785] 1903: 412), and that, though their main function is not to state facts, the adoption of them can be a rational process. A recent version of prescriptivism put forward by R.M. Hare has, following Kant, claimed objectivity, in a sense carefully defined and distinguished from factuality, for moral statements (see Hare 1993).

2 Prescriptive and descriptive meaning

Prescriptive meaning is often likened generically to that of imperatives, which also prescribe actions; but careful

ethical prescriptivists emphasize also the differences, as well as the similarities, between moral statements and imperatives. The chief of these differences is that moral statements also have, unlike imperatives, a descriptive or factual content. This is linked with another feature of moral statements, known in the jargon as universalizability: the feature that a moral statement about one situation must apply to any other situation similar in all its universal properties (that is, those which can be described without reference to the individuals involved). That moral statements are universalizable is acknowledged by most thinkers, whether or not they are prescriptivists (see Universalism in ethics).

If moral statements are universalizable in this way, they will inevitably acquire a descriptive function as part of their meaning; for someone who makes such a statement must think that the moral properties of the situation about which it is made supervene on its nonmoral properties (the reasons for making it) in accordance with a principle applying to all identically similar situations (see Supervenience). Thus they will be taken to hold that there are facts that justify the statement, and it thus acquires a factual, descriptive content or meaning. However, since different people and cultures hold differing moral principles, this descriptive meaning (the truth conditions of the statement) may vary from one person or culture to another (see Moral relativism). It is the prescriptive meaning which remains constant: whatever kinds of behaviour different people or cultures prescribe, they are all prescribing them for all similar cases. This, as we shall see, is what gives moral reasoning its grip.

The relation between the prescriptive and descriptive elements in the meaning of moral statements has been much misunderstood. The prescriptive meaning is the function all normative and evaluative statements have of guiding our actions. This shows up in the fact that someone who makes one about his own proposed actions but does not act accordingly exposes himself to a charge of speaking insincerely, as also does someone who makes one about other people’s actions but does not will them (in the above broad sense) so to act.

The descriptive meaning is the standard or criterion or reason or principle in accordance with which the statement is made. For example, if I say, ‘You ought not to say that, because it would be a lie,’ I am applying the principle that one ought not to tell lies. The ‘because’-clause does not merely repeat the ‘ought’-statement; it adds a reason for it. A common mistake is to confuse the content of moral statements with the reasons for them. That this is a mistake is shown by the fact that different people might make the same moral statement for quite different reasons; they might disagree radically in their moral principles. They would then be giving the same prescription, but disagreeing in their reasons for it.

Likewise, two people who differ about a moral question, one saying, ‘He ought,’ and the other, ‘He ought not,’ are affirming and rejecting, respectively, the same moral statement; otherwise they would not be disagreeing.That is what is wrong with the kind of subjectivism mentioned above. The statement is the same because it contains the same universal prescription, which one of them accepts and the other rejects. The fact that they are disputing about the same statement, bound by the same logic, is what enables argument between them to begin, in order to seek a resolution of their dispute (see §3).

3 Prescriptivism defended

There are many confusions that have to be cleared up before we can understand the prescriptivist position. The first we have noticed already: prescriptivists hold that moral statements are expressions of volition (in a broad sense in which it covers Kant’s rational will and Aristotle’s *boulēsis* or rational desire). They do not hold that they are statements that the speaker in fact wills something. They are therefore not exposed to the charge that they make our duties depend on what we will, as if we could make something our duty by willing it. If that kind of subjectivism were correct, then if someone said that something was their duty they would mean that as a matter of fact they willed it; and if they really did will it, it would be impossible to refute them. Rather, prescriptivists hold that in making a moral statement we are expressing our rational will; and if someone else wills something different, the disagreement has to be resolved by reason, that is, by determining which prescription can be rationally willed in the light of all the nonmoral facts, and respecting the logical property of universalizability (see §2), which, as Kant again saw, affords a powerful weapon in moral argument.

Opponents of prescriptivism and in general of ethical internalism commonly object that it is possible for persons to make a moral statement but to have no disposition to act on it nor to desire others to act on it (see Moral motivation §§1-3). They might be weak-willed persons or akratic, or satanists, who do wrong acts just because

they are wrong, or amoralists who do not accept moral statements as guides to action (see Akrasia). Answers to these objections have been put forward by Hare (1992a, 1992b: 98, and 1981 respectively). They require careful distinctions between different senses in which one can think that something ought to be done, not all of which are prescriptive.

The prescriptive sense of words like ‘ought’ is their central but not their only sense, and the existence of other senses easily gives colour to objections of this sort. The fact that moral statements carry descriptive as well as prescriptive meaning also makes it easy for descriptivists, attending only to this descriptive meaning, to argue that the statements are purely descriptive, that is, that their meaning is determined by their truth conditions and by nothing else. Often such arguments are reinforced by appeals to the surface grammar of moral statements, which are commonly in the declarative or indicative mood (though ‘ought’-statements are an exception). But the fact that such statements do have descriptive meaning, determined by truth conditions, alongside their prescriptive meaning is quite enough to explain such phenomena. These, however, do not show that they lack prescriptive meaning, nor that this is not central.

Moral words fall into two classes, according to how tightly the two parts of their meaning are tied to them. The distinction is sometimes made in terms of a contrast between ‘thin’ and ‘thick’ moral concepts. There are first the primarily prescriptive (or primarily evaluative) words like ‘ought’, ‘wrong’ and ‘good’, whose prescriptive meaning is more firmly attached to them than their descriptive. If someone starts making some of their moral judgments by entirely new standards (they have become a pacifist, say, or a vegetarian), they can still go on using those words, in the same senses, to express and apply the new standards.

By contrast there are other words, the secondarily prescriptive, whose prescriptive meaning is tied to them more loosely than their descriptive. If the standards encapsulated in these words are discarded, as in principle they can be, the words have to be discarded with them (see Hare 1963: 189, footnote).

A common move of descriptivists is to claim that there is no real difference in prescriptivity between prescriptive and descriptive speech acts, because either can be used to guide action. To say that a man has frequently beaten his first wife before their divorce may provide an incentive not to become his second; yet, it is claimed, the statement that he indulged in wife-beating is all the same descriptive. But a lady such as ‘O’ who liked to be beaten might not call him a bad husband (Réage 1954: 17). Thus ‘wife-beater’ is only contingently a pejorative term, but ‘bad husband’ is necessarily so. That is the difference between inherently and merely contingently action-guiding expressions. Only the former are genuinely prescriptive.

In spite of appearances, prescriptivism has the advantage over descriptivist theories when it comes to establishing the rationality and objectivity of moral thinking. Objectivity may be defined as acceptability to all rational thinkers. All forms of descriptivism have to rely on appeals either to generally accepted moral convictions, which may vary from one culture to another or to the meanings of words (especially secondarily prescriptive words) which encapsulate such culturally relative values. Only prescriptivists can rely solely on the logic of moral statements, which can be invariant between cultures; they can ask, as Kant did, what one can will to be a universal law, applicable whoever occupies whatever role in a situation. Faced with this question, rational thinkers will agree on the answer (see Moral knowledge §3). They will thus come to agree on principles giving the truth-conditions of moral statements - principles which are acceptable to all rational thinkers and depend solely on logic and the nonmoral facts. This is the only way to secure objective, culture-invariant, truth for moral statements (Hare 1993).

See also: Logic of ethical discourse

R.M. HARE

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985.(Contains prescriptivist strands in his accounts of weakness of will, practical wisdom, and the practical syllogism.)

Aristotle (*c*. mid 4th century BC) *On the Soul*, trans. with notes by J.A. Smith, Oxford: Oxford University Press, 1931.(Discusses the prescriptive first premises of practical syllogisms.)

Aristotle (*c*. mid 4th century BC) *On the Movement of Animals*, trans. A.S.L. Farquarson, Oxford: Oxford

University Press, 1912.(Discusses the prescriptive first premises of practical syllogisms.)
Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz; 2nd edn, 1946, ch. 6.(Early and central discussion of emotivism.)
Ayer, A.J. (1949) ‘On the analysis of moral judgments’, *Horizon* 20; repr. in *Philosophical Essays*, London: Macmillan, 1965.(Important British advocate of emotivism.)
Carnap, R. (1935) *Philosophy and Logical Syntax*, London: Routledge & Kegan Paul.(Early imperativist.)
Hare, R.M. (1952) *The Language of Morals*, Oxford: Oxford University Press.
Hare, R.M. (1963) *Freedom and Reason*, Oxford: Oxford University Press.
Hare, R.M. (1981) *Moral Thinking*, Oxford: Oxford University Press.(The last three items set out the author’s prescriptivist position.)
Hare, R.M. (1992a) ‘Weakness of will’, in L. Becker (ed.) *Encyclopedia of Ethics*, New York: Garland Publishing Company.(Includes discussion of moral weakness.)
Hare, R.M. (1992b) *Essays on Religion and Education*, Oxford: Oxford University Press. (Includes papers on practical topics written from a prescriptivist point of view.)
Hare, R.M. (1993) ‘Objective Prescriptions’, in E. Villanueva (ed.) *Naturalism and Normativity*, Atascadero, CA: Ridgeview; also published in A.P. Griffiths (ed.) *Ethics* (Royal Institute of Philosophy Lectures, 1992/3), Cambridge: Cambridge University Press.(On prescriptivism and truth in ethics.)
Hare, R.M. (1997) *Sorting Out Ethics*, Oxford: Oxford University Press.(Includes critical discussion of various ethical theories from a prescriptivist point of view.)
Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 2nd edn, 1978, book 3, part I, section 1.(A classical attack on ethical realism.)
Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, in *Kants gesammelte Schriften*, ed. Königlichen Preußischen Akademie der Wissenschaften, Berlin: Reimer, vol. 4, 1903; trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(The reference made to this work in the entry gives the page number from the 1903 Berlin Akademie volume; the page number from the *original* German edition is 36. Both of these sets of page numbers are included in the margins of the Paton translation. This is the work of a prescriptivist who bases his ethics on ‘rational maxims’.)
Mill, J.S. (1843) ‘The Logic of Practice or Art’, in *System of Logic: Ratiocinative and Inductive*, London: Parker; repr. in *Collected Works of John Stuart Mill*, London: Routledge, vols 7 and 8, 1991, book 6, ch. 12.(Explains Mill’s prescriptivism.)
Plato (*c*.395-387 BC) *Gorgias*, trans. W.K.C. Guthrie, in E. Hamilton and H. Cairns (eds) *The Collected Dialogues including Letters*, Oxford: Oxford University Press, 1975.(Demonstrates prescriptivist elements in Socrates’ thought.)
(*c*.386-380 BC) *Protagoras*, trans. W.K.C. Guthrie, in E. Hamilton and H. Cairns (eds) *Plato, The Collected Dialogues including Letters*, Oxford: Oxford University Press, 1975.(Demonstrates prescriptivist elements in Socrates’ thought.)
Réage, P. (1954) *L’Histoire d’O*, Paris: Sceaux; cited in W. Young, *Eros Denied*, London: Weidenfeld, 1964, 241. (A well-known account of masochism.)
Stevenson, C.L. (1944) *Ethics and Language*, New Haven, CT: Yale University Press.(The fullest exposition of emotivism.)

Presocratic philosophy

The Presocratics were the first Western philosophers. The most celebrated are Thales, Anaximander, Pythagoras, Heraclitus, Parmenides, Zeno of Elea, Empedocles, Anaxagoras and Democritus. Active in Greece throughout the sixth and fifth centuries BC*, they concentrated on cosmogony and cosmology - the tasks of explaining the world's origin and order, without recourse to mythology.*

Socrates (469-399 BC) is perceived as marking a watershed in philosophy - a shift of focus from the origin and nature of the universe to human values. 'Presocratic' philosophy thus represents the era intellectually antecedent to Socrates, even though its exponents included contemporaries of his (some his juniors). Those thinkers contemporary with Socrates who shared his concentration on human values it is better not to call Presocratic (see Sophists). No complete Presocratic text survives. We have only later writers' quotations ('fragments'), summaries (see Doxography), criticisms, and so on, from which to glimpse the originals.

A mythological construction of the world was already integral to the earliest poetry familiar to the Greeks (*c*.700 BC) (see Hesiod; Homer). Philosophers' rationalizations of this picture concentrated initially on such questions as what the world's primeval stuff is, why the earth remains stable and, more generally, what made the world orderly. In time *kosmos* ('ordering'), came to mean 'world'. To explain cosmic order, biological, mechanical and even political models were developed. However, Presocratic philosophy was interested equally in the human soul and its destiny (see Psychē), and never altogether ignored human values. Another dominant issue was the possibility of human knowledge.

The main movements and phases were as follows. The three sixth-century BC Milesian philosophers, starting with Thales, were monists: each posited a single primeval stuff - for example, water, air. These came to function not just as the world's originative stuff but also perhaps as its enduring substrate (see Archē; Thales §2; Anaximander §§2-3; Anaximenes; Monism).

Pythagoreanism, although a secretive movement with cultic leanings, was highly influential throughout this era. Beyond a concern with the soul and survival, it promoted a mathematicizing approach to cosmology (see Pythagoras §1; Pythagoreanism §2; Orphism; Philolaus §§2-3).

Heraclitus (*c*.540-480 BC) kept the formal focus on the cosmos and the soul, but his approach was largely governed by metaphysical concerns, especially the paradoxical interdependence of opposites (see Heraclitus §3). Metaphysics and logic took centre stage soon after in the Eleatic movement, initiated by Parmenides and continued by Zeno of Elea and Melissus. Eleaticism was a radical critique of ordinary notions of being, defending instead a strict monism which outlawed all phenomenal distinctions and changes as illusory (see also Gorgias §3).

The later fifth-century BC cosmologists sought to deflect this Eleatic critique. Most were pluralists, positing more than one underlying element. Their contributions culminated in the atomic system of Democritus, which reduced all reality to atoms and void, conceding much to Eleaticism (see Atomism, ancient; Anaxagoras; Empedocles; Leucippus)

Presocratic philosophy is often seen as materialistic. However, no philosopher before Leucippus and Democritus reduced life and intelligence to something inanimate. Previous Presocratics considered these ineliminably present in things - either as already intrinsic to the primordial stuff(s) or, in Anaxagoras' system, by virtue of an irreducible dualism of mind and matter.

See also: Alcmaeon; Ancient Philosophy; Diogenes of Apollonia; Doxography; Epicharmus; Hippocratic medicine; Nous; Physis and nomos; Socrates; Xenophanes

DAVID SEDLEY

References and further reading

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge & Kegan Paul, chaps 9-11.(The most philosophically rewarding study of the Presocratics.)

Barnes, J. (1987) *Early Greek Philosophy*, Harmondsworth: Penguin.(Translation of the main sources, embedded

in the contexts in which they are preserved.)
Diels, H. and Kranz, W. (1952) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 3 vols.(The standard edition of the Presocratic philosophers and the Sophists; contains secondary testimonia, plus Greek texts of the fragments with translations in German.)
Furley, D.J. and Allen, R.E. (eds) (1975) *Studies in Presocratic Philosophy*, London: Routledge & Kegan Paul. (Two-volume collection of seminal papers.)
Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols. (Volumes 1 and 2 are the most detailed and comprehensive English-language history of Presocratic thought.)
Hussey, E. (1972) *The Presocratics*, London: Duckworth.(Accessible introductory account.)
Kirk, G.S. (1960) 'Popper on Science and the Presocratics', *Mind* 69: 318-39, repr. in D.J. Furley and R.E. Allen (eds), *Studies in Presocratic Philosophy*, London: Routledge & Kegan Paul, 1975, vol. 1, 154-77.(Reply to an earlier version of Popper (1963).)
Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(A valuable survey of Presocratic philosophy, including texts and translations.)
Long, A.A. (ed.) (forthcoming) *The Cambridge Companion to Early Greek Philosophy*, Cambridge: Cambridge University Press.(A series of essays on key philosophers and themes.)
McKirahan, R.D. (1994) *Philosophy before Socrates*, Indianapolis, IN: Hackett Publishing Company. (Comprehensive collection of sources in translation, with commentary.)
Mourelatos, A.P.D. (1974) *The Pre-Socratics*, Garden City, NY: Doubleday; repr. 1993. (Important collection of essays, with enlarged bibliography in the 1993 reprint.)
Popper, K.R. (1963) 'Back to the Presocratics', in *Conjectures and Refutations*, Routledge & Kegan Paul; repr. in D.J. Furley and R.E. Allen (eds), *Studies in Presocratic Philosophy*, London: Routledge & Kegan Paul, 1970, 1975, vol. 1, 130-53.(Tribute to the Presocratics for supposed anticipations of Popperian scientific method.)

Presupposition

There are various senses in which one statement may be said to 'presuppose' another, senses which are in permanent danger of being confused. Prominent among them are Strawsonian presupposition, a relation which obtains between statements when the falsity of one deprives the other of truth-value (for example, 'There was such a person as Kepler' is a Strawsonian presupposition of 'Kepler died in misery'); semantic presupposition, which obtains between a statement and a particular use of a sentence type, when the falsity of the statement means that that use will not after all constitute the making of a statement (for example, 'The name "Kepler" has a bearer' is a semantic presupposition of 'Kepler died in misery'); and pragmatic presupposition, a broader notion exemplified by the legitimate presumption that accepting or denying the statement 'Fred knows that the earth moves' means accepting 'The earth moves'.

1 Introduction

In most conversations a certain amount of knowledge is assumed by the parties. Conversing with people whom we know well, we can take it for granted that they know who we are and what we have recently been doing, along with much else. Ordinarily, 'presuppositions' are nothing other than these assumed parts of the conversational background.

The notions of presupposition which have been of greatest concern to linguists and philosophers, however, have been more narrowly defined (but see §4 below), typically being concerned with the way certain phrases or constructions indicate that the speaker is making a presupposition in the ordinary sense. So, if I affirm 'Wellington's victory at Waterloo was his greatest triumph', my form of words indicates that I am presupposing that Wellington was the victor at Waterloo and asserting that it was his greatest triumph. However, the commonly stressed opposition between presupposition and assertion is in some measure misleading. In asking 'Is Tony going to Salzburg again this summer?', I presuppose that he has been there before and inquire whether he intends to go there in the future; no assertion has been made.

This phenomenon may be closely related to another. If I affirm

(1) All John's children are asleep,

I presuppose that John has children and assert that they are asleep. In affirming

(2) If all John's children are asleep, we should not leave the house,

I do not make the same assertion, but I do still presuppose that John has children. It is this that linguists have in mind when they speak of presuppositions being easily 'inherited': they are often (but not always) retained even when the clauses that convey them are sententially embedded so as to lose their own assertive force. Much of the work done by linguists on this topic has been concerned with identifying constructions in which presuppositions do not survive, and to explain why inheritance is blocked (see various authors collected in Davies 1991).

Philosophers writing on the subject, by contrast, have largely been concerned to elucidate the nature of presupposition: to explain what one thing's presupposing another really amounts to; and to specify the kinds of item between which the relation of presupposition obtains.

2 Strawsonian presupposition

The writer most responsible for bringing the notion of presupposition to the attention of philosophers is P.F. Strawson, from whose work (1) above is drawn. Commenting on the situation of somebody who makes an assertion by uttering (1), Strawson remarks that

> he will not normally, or properly, say this, unless he believes that John has children…. But suppose…John has no children. Then is it true or false that all John's children are asleep? Either answer would seem to be misleading. But we are not compelled to give either answer. We can, and normally should, say that, since John has no children, the question does not arise.
>
> (1952: 173-4)

Strawson is eventually led to advance the following general explanation: ‘if [statement] *S′* is a necessary condition of the *truth or falsity* of [statement] *S*,…then *S presupposes S*Čprime;’ (1952: 175; original emphasis). So the statement that John has children is a ‘Strawsonian presupposition’ of the statement that all John’s children are asleep; for unless the former is true the latter is neither true nor false. Strawson’s talk of a ‘question [which] does not arise’ is liable to mislead. If John has no children, then the question ‘Is statement (1) true or false?’ may be raised; but the correct answer will be that it is neither.

To understand the notion of Strawsonian presupposition properly, it is necessary to understand what exactly a statement is. Strawson distinguishes between declarative ‘sentences’ - linguistic types the use of which on various occasions of utterance enables people to say things - and the ‘statements’ thereby made. It is statements, not sentences, that are true or false. Should a presupposition fail, there will be a ‘truth-value gap’: a statement will qualify neither as true nor false. An analogous notion applies to other kinds of speech act. So a yes/no question *Q* may be said to presuppose a statement *S* if the truth of *S* is a necessary condition of *Q*’s being answered by ‘yes’ or ‘no’ (see Strawson 1954).

The notion of Strawsonian presupposition has certainly found important philosophical application. In Strawson’s own writings it forms the basis of a counter to Russell’s suggestion that simple sentences involving ‘definite descriptions’ - that is, sentences of the form ‘The *F* is *G*’ - are not of subject-predicate form (see Descriptions). Russell held that such a sentence is best analysed as an existential sentence, ‘There is one and only one *F*, and it is *G*’, whereby the statement that the *F* is *G* straightforwardly entails the existence of an *F*. Having attacked Russell’s arguments in its favour, it was on precisely this score that Strawson tried to refute the theory itself: he insisted that the statement that the present king of France is bald presupposes (rather than entails) that there is a present king of France (see Descriptions §4).

In assessing this suggestion, note that when contemplating statements which are neither true nor false, Strawson requires ‘false’ to bear a sense other than ‘not true’. This means that, in the first place, the difference between Strawson and Russell is less than might at first have been supposed. Since Strawson accepts Russell’s account of the conditions under which a statement made by uttering ‘The *F* is *G*’ will be true, he agrees *ipso facto* with his account of the conditions under which such a statement is not true. They differ only over the conditions for such a statement to be false, Strawson deeming such statements to be neither true nor false in circumstances where Russell would deem them straightforwardly false.

In the second place, however, we are driven to wonder what Strawson means by calling a statement ‘false’, given that it is not a synonym for ‘not true’. It is reasonable to expect the attribution of falsity to a statement to be regulated by the principle that *S* is false just in case the negation of *S* is true, and this is a principle that Strawson’s treatment respects; indeed, it seems that he takes this principle to be explanatory of the notion of falsity. But then it becomes crucial to have a test for identifying the negation of a statement, especially for natural languages, where the syntax does not necessarily make this obvious. Since Strawson’s reluctance to classify ‘The king of France is bald’ as false when there is no king of France stems from a desire to reserve that attribution for the case in which there is a king of France and he is not bald, it is plain that Strawson regards the latter case as the proper negation of this statement. This, indeed, is what one might expect, given that Strawson is defending the view that ‘The *F* is *G*’ is a subject-predicate sentence. But the Russellian will want to draw a semantic distinction between ‘It is not the case that the king of France is bald’ and ‘The king of France is not bald’. This suggests that the issue between them ultimately turns, not upon questions concerning presupposition, but upon the proper treatment of negation (and other logical modifiers) in natural languages.

3 Semantic presupposition

A rather different notion of presupposition is suggested by Frege: ‘If anything is asserted, there is always an obvious presupposition that the simple or compound proper names used have a reference [*Bedeutung*]. If therefore one asserts "Kepler died in misery", there is a presupposition that the name "Kepler" designates something’ (1892: 40). In order to see the difference between this notion of presupposition and Strawson’s, it helps to make one assumption about statements; namely, that the same statement can be made in a variety of languages. Suppose, then, that there is a sequence of Chinese characters by the inscription of which it may be stated that Kepler died in misery. Such an inscription will certainly not contain the name ‘Kepler’, so it will not impede the making of that

statement, should the English (or German) name 'Kepler' lack a reference. But in that case,

(3) The name 'Kepler' has a reference

plainly is not a Strawsonian presupposition of

(4) Kepler died in misery,

that is, of the statement made by uttering (4). For the latter statement could be made by inscribing the Chinese characters instead, and the statement thereby made could have a truth-value even if (3) were false. What does qualify as a Strawsonian presupposition of (4) is the following statement:

(5) There was such a person as Kepler.

We need, then, to distinguish the new notion of presupposition, exemplified by (4) and (3), from that exemplified by (4) and (5).

In explicating this new notion, it helps to invoke again Strawson's distinction between a (declarative) sentence type - which is the bearer of meaning - and the statement made when the sentence is uttered or used - which is the bearer of truth-value. For with that distinction in mind, we see that the new notion is not a relation between statements, but a relation between a sentence - as used (with a particular meaning) upon a particular occasion - and a statement. Thus, the sentence 'You are drunk' (as used on a particular occasion) presupposes that the speaker is addressing somebody, in the sense that if not, then the use of that sentence on a particular occasion will not constitute the making of a statement. Generally, then:

A use of sentence σ in circumstance *c* presupposes statement *S* in so far as the truth of *S* is necessary for that use to qualify as the making of a statement.

When *S* and σ are so related, we may say that *S* is a 'semantic presupposition' of sentence σ, relative to *c*.

4 Pragmatic presupposition

Whatever the differences between them, Strawsonian and semantic presupposition are alike in depending upon the content of statements, or the meanings of sentence types. That is to say, they are alike in abstracting from the particular conversational purposes of the person uttering the sentence. In this respect they differ from another presupposition relation in which the actual speaker looms much larger. Following Stalnaker (1974), we might say that

A statement *S*′ is a 'pragmatic presupposition' of a statement *S* in so far as a hearer may reasonably infer that the speaker accepts *S*′ either from their acceptance of *S* or from their denial of it.

One salient difference between this explication and the definition of a Strawsonian presupposition is that the notions of truth and falsity have been replaced by those of acceptance and denial (acceptance of the negation). Another is the introduction of the notion of a hearer's reasonable inference - reasonable, that is, by the lights of a well-run conversation.

Whether one statement is a pragmatic presupposition of another depends crucially, then, upon what constitutes a well-run conversation, a notion philosophers have tried to explicate by formulating general 'conversational maxims', such as the rule (R) that one's hearer may assume that one is making the logically strongest relevant statement compatible with one's beliefs. To illustrate this, and to show how a pragmatic presupposition might obtain even when a Strawsonian presupposition does not, suppose that speakers are known to abide by (R), and consider the statements *S*: 'Fred knows that the earth moves' and *S*′: 'The earth moves'. *S*′ is not a Strawsonian presupposition of *S*: when *S*′ is false, we may conclude simply that Fred's belief is not in fact knowledge and that *S* itself is straightforwardly false. All the same, a strong case can be made for the contention that *S*′ is a pragmatic presupposition of *S*: I may reasonably infer that a speaker accepts *S*′, should I hear them either accepting or denying *S* (see Stalnaker 1974). The case of accepting *S* presents no difficulty; and a denial of *S* amounts to an assertion of 'Fred doesn't know that the earth moves', an assertion which I may suppose that the speaker would not have made had they accepted the stronger statement 'The earth does not move'.

See also: Descriptions; Implicature; Logical and mathematical terms, glossary of;Pragmatics

IAN RUMFITT

References and further reading

Davies, S. (1991) *Pragmatics: A Reader*, New York: Oxford University Press. (Includes many papers showing the nature of contemporary work on presupposition by linguists, including articles on the problem of 'inheritance' mentioned in §1.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. 'On Sense and Meaning', in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980. (Introduces the notion of semantic presupposition. The page numbers of the original publication appear in the margins.)

Stalnaker, R.C. (1974) 'Pragmatic Presupposition', repr. in S. Davies, *Pragmatics: A Reader*, New York: Oxford University Press, 1991.

Strawson, P.F. (1950) 'On Referring', *Mind* 59: 320-44; repr. in *Logico-Linguistic Papers*, London: Methuen, 1971. (A classic paper, using presuppositional considerations to attempt to refute Russell's theory of descriptions. Note that the term 'presupposition' is not used in the paper; it is called 'implication in the special sense'.)

Strawson, P.F. (1952) *Introduction to Logical Theory*, London: Methuen. (Defines (Strawsonian) 'presupposition'.)

Strawson, P.F. (1954) 'A Reply to Mr. Sellars', *Philosophical Review* 63: 216-31. (Extends presupposition beyond the assertoric.)

Price, Richard (1723-91)

Richard Price was a Welsh dissenting minister who contributed widely to philosophy and public life in latter-eighteenth-century Britain. The leading British ethical rationalist of the period, Price did much to establish intuitionistic and deontological traditions in ethics. He put forward searching criticisms of alternative empiricist conceptions, arguing that they could not account for morality's necessity and that they lacked an adequate theory of moral agency. More constructively, he argued that, contrary to the empiricists, all knowledge depends on the contribution of reason, and that rationalistic moral knowledge is no more problematic in principle than ordinary empirical knowledge. He also articulated a normative ethics of integrity that stressed the duty diligently to search out moral truth and then to act on the truth as one sees it.

As a political philosopher, Price made fundamental contributions through his doctrine of liberty as self-determination. The moral duty individuals have to determine themselves by their best moral judgment, Price believed, ultimately grounds the values of political liberty, independence, and democracy as well. Price's radicalism on these scores earned him the famous opposition of Edmund Burke. Nor was Price's sponsorship of these ideas simply theoretical. He was an important friend of the American Revolution, and his pamphlets analysing and defending it were taken seriously by proponents and opponents alike.

Price also did important work on the mathematical theory of probability and in proposing and instituting various social and economic reforms and practices upon its basis. He was instrumental in making Thomas Bayes' ideas about probability accessible to the learned world, and in making use of these and other probabilistic theories in developing insurance, self-help, and other financial schemes.

1 Life and works

Richard Price was born on 23 February 1723 at Tynton, Llangeinor, Glamorgan into a family that, on his father's side, had been involved in Welsh dissent for two generations. Much of Price's normative outlook can be understood against this background - his emphasis on piety, integrity, self-determination and freedom of conscience. Price's meta-ethics, on the other hand, was a reaction against Calvinist theological voluntarism of the sort represented by his dissenting minister father, Rice Price. As son Richard saw it, moral propositions claim by their very nature to represent an objective, necessary and non-arbitrary order. Voluntarism, whether secular or divine, must ultimately lead to moral scepticism (see Moral scepticism; Voluntarism). Consequently, Price believed that the moral core of religious dissent, to which he would be loyal throughout his life, could only be preserved by giving it a rationalist foundation. Together with articulating dissent's normative dimensions, this was Price's central philosophical project.

Price's great ethical work was *A Review of the Principal Questions and Difficulties in Morals* (1758). It is the repository of all his major ideas on ethics and the source of many of his political theories. The year *A Review* appeared, Price accepted a post at a Presbyterian church in Newington Green, near London. During the next fifteen years there, he published on an astonishing variety of topics. In 1764 and 1765 he edited Thomas Bayes' seminal writings on probability theory, publishing them in the *Transactions* of the Royal Society. *Four Dissertations*, a theological work, appeared in 1767. In 1770 and 1771 Price contributed three further papers to *Transactions*, one on astronomy and two on calculating life-expectancy and annuities, respectively. These last were incorporated into *Observations on Reversionary Payments* (1771), which established Price's considerable reputation in the field of insurance. It went through three editions, and part was expanded and published separately as *An Appeal to the Public on the Subject of the National Debt* in 1772.

Price's greatest fame, however, came from his political writings, beginning with *Observations on the Nature of Civil Liberty* (1776), a pamphlet that sketched his political philosophy and defended the American Revolution on its basis. Seven editions appeared during its first year of publication alone. Its major philosophical thesis was that states acquire legitimacy as trustees for consenting, self-determining moral agents, and have thereby a right to self-determination as representatives of their citizens. If, consequently, the American colonists wanted to govern themselves, that was both their right and duty. Price wrote several further works analysing and praising the American Revolution and one, *A Discourse on the Love of Our Country* (1789), celebrating the French Revolution, which drew Edmund Burke's ire in his *Reflections on the Revolution in France* (1790) (see Burke, E.).

Throughout this period of extraordinary creativity and prominence, Price remained active in communities of dissent. He resigned his post at Newington Green in 1783, but continued an active participant there, as he did in St Thomas' Square, Hackney, to which he removed in 1787. His final sermon at Hackney came just months before his death.

2 Meta-ethics

Price advanced a rationalist meta-ethics along with a critique of the moral sense theories of Francis Hutcheson and David Hume, turning against them arguments Hutcheson had used to criticize secular and theological voluntarism (see Hume, D.; Hutcheson, F.; Moral sense theories). Hutcheson (1725) had argued that moral good and evil are simple, irreducible ideas. Any reductionism, including voluntarism must fail, therefore, since it is one thing to say that I will suffer sanctions if I fail to act as commanded, but another to say that violation is morally wrong or evil and that sanctions will be deserved. Alternatively, if voluntarists hold that moral good consists in obeying authoritative commands, they then assume a background moral fact that can have no voluntarist explanation, namely, that these commands are ones that persons should obey.

Because he held an empiricist theory of ideas, Hutcheson concluded from this irreducibility that there is a distinctive moral sense which receives moral ideas. Price agreed that moral ideas are irreducible, but argued that attributing them to a contingent sense or sensibility makes them ultimately no less posited than if they were to depend on arbitrary command. According to Hutcheson, benevolence is morally good, not because it intrinsically or necessarily warrants approbation, but because beings with our moral sense (including God) happen to approve it.

There are several different elements to the Pricean critique of empiricism. First, Price argued that moral judgments make an implicit claim to *objectivity*, while any empiricist account must understand them ultimately to concern subjective experience. Second, moral judgments attribute properties based on the *intrinsic nature* of their objects. Third, they attribute these properties *eternally* and *necessarily*. Fourth and finally, moral judgments attribute properties independently of any *arbitrary* standard. When we attribute viciousness to dishonesty, for example, we implicitly hold that, by its very nature, dishonesty has a property that depends not on any response we happen to have, standard we arbitrarily apply, or any other contingent feature.

What misled the empiricists, Price believed, was their general theory of ideas. Price insisted that reason is itself a source of ideas. Were it not, we would lack a whole host of concepts that cannot possibly come from experience alone. Following Ralph Cudworth (whose *Treatise on Eternal and Immutable Morality* appeared posthumously in 1731, after Hutcheson's *Inquiry*), Price maintained that the notions of substance, duration, space, causation and necessity, among others, cannot come from sense experience, and so must be due to reason or the understanding (see Cudworth, R.). Nor should we doubt that reason is also the source of moral ideas. A rational basis for moral ideas, on the one hand, and any basis at all for the notions central to the science of nature, on the other, stand or fall together.

Price concluded that moral claims can be known by direct rational intuition. Some, the 'heads' of virtue, are self-evident. These are 'general principles', such as those enjoining veracity, fidelity, gratitude, justice and so on, that concern what it is right and wrong to do. Principles at this level of generality are bound to conflict in actual cases, however, and Price held that it is not self-evident what a person should do when they do. What is self-evident to reason is what moral considerations are relevant - what persons should and should not do, other things being equal.

Price's meta-ethics thus amounted to a metaphysical realism combined with a rational intuitionist epistemology (see Realism and antirealism). Moral propositions are like mathematical claims, Price thought. Their truth is determined by correspondence to facts about an objective, eternal, necessary order. And they are known through a rational perception of the same kind.

But how then is moral perception or knowledge related to motivation? One way this question arises is in connection with obligation, which Price's contemporaries tended to see as intrinsically tied to motive. Indeed some writers, rationalists *and* empiricists, maintained that obligation just *is* a motive, perhaps inescapable, perhaps distinctively moral. Price resisted this identification, holding that 'obligatory' is another name for the fundamental

irreducible moral notion. Obligation consists in the eternal and immutable moral features of the objective situation in which persons find themselves, not on their motivational state. At the same time, Price held that the *perception* of irreducible moral features *is* linked to motivation and, indeed, that 'it is not conceivable' that a person who perceives that an action ought to be done 'can remain uninfluenced, or want a motive' (1758: 186). But if motivation is no part of the perceived ethical facts, and if the faculty of perception is the understanding (the same faculty involved in the motivationally inert perception of mathematical facts), what explains this necessary motivation? What makes it, as Hume put it, 'universally forcible and obligatory'?

Price's reply was a doctrine of rationally necessary affect and motive, of which this was but one example. The only alternative to believing that moral properties are objective, eternal and immutable is to hold that nothing has a nature that determines its moral properties. This would be true no less of God than of other beings. But we can hardly deny, Price argues, that God is good simply by virtue of his nature. And this commits us to the conclusion that goodness is an eternal, objective and immutable property of a being with such a nature. Moreover, there are other aspects of God's nature that would seem equally essential and not merely contingently true of him. For example, it hardly seems that it could be merely contingently true that God is pleased by virtue or by the perception of beauty generally, or that he is benevolent, or that he desires what is morally good and just for his creatures. But if we accept that affects and motives of these sorts are essential to God, Price argues, we should also believe that they are essential to rational nature, rather than deriving from some contingent or 'implanted' motive or sense. It follows that being moved by the perception of ethical facts is intrinsic to the understanding.

3 Normative ethics

Just as Price rejected empiricist metaethics, so also did he deny the proto-utilitarian, normative ethics it was frequently used to support. The 'heads of virtue' are various and distinct, and although benevolence is one of these, the others are not reducible to it. Piety, veracity, justice and the other 'heads' are self-evidently intrinsically morally relevant, regardless of their actual or usual consequences.

Price distinguished between the virtue of actions and the virtue of agents. The former concerns whether, all things considered, an action is objectively right, irrespective of the motives that actually led or will lead to it. An action manifests the agent's virtue, on the other hand, only if it is done from the proper motives. Most importantly, Price held, actions manifest agent-virtue only when the agent 'acts from a consciousness of rectitude and with a regard to it as his rule and end' (1758: 184). Behind this thesis lies Price's doctrine that genuine agency entails self-government, and that this involves a person guiding themselves by their convictions concerning what they should do. If moral virtue is a virtue of agents, therefore, it must include this element.

For Price, the virtuous agent is a person of integrity, concerned to determine what they should do and, having determined that, to act as they think they should. The first duty of moral agents is diligently to determine the virtuous act, to see under what self-evident general principles the alternatives before them fall, and carefully to judge what they should do, all things considered. God has made human beings so that moral truth is tolerably accessible to them if they will but use their understanding attentively. Having made a conscientious judgment, Price held, a moral agent should then follow their conscience 'steadily and faithfully', even if it should turn out to have been mistaken.

Price distinguishes between the abstract and practical virtue of actions. The abstract virtue of an act is its moral quality, considered independently of the agent's motives and beliefs, including those concerning what the agent should do. This is what moral agents attempt to determine in deliberation: 'What is the right thing to do owing to the objective features of my practical context?' Practical virtue, by contrast, depends on agents' beliefs about the alternatives confronting them, including, importantly, their beliefs concerning the abstract virtue of these alternatives. This analytical distinction grounds an important Pricean normative doctrine, namely, that when moral agents have seriously deliberated, what they then should do is what they think they should do. Practical virtue is thus not just a moral quality an action would have if the objective features of alternatives confronting them were as the agent believed. If the agent makes a serious judgment that they should do something, then they really should do it. To act otherwise would be to compromise their integrity.

Consequently, while Price's rational intuitionism stresses eternal and immutable moral qualities actions have owing to their objective, intrinsic nature, these concern only abstract virtue, which is, however, of the first

importance in deliberation, as agents diligently seek to determine the abstract moral qualities of actions before them. But once deliberation is terminated and agents have formed considered beliefs, what they should then do is a matter of practical rather than abstract virtue. They should follow their conscientious judgment.

Coinciding in this normative doctrine are proto-Kantian theories of self-determining agency and character (see Kant, I.; Kantian ethics). Capacity to act on an independent judgment of what one ought to do is what makes a being a self-determining agent in the first place, and moral virtue consists in the excellent exercise of this capacity. Sometimes, indeed, Price seems to count any moral failure as a failure of agency, creating puzzles about how vice can ever be imputed to an agent.

4 Political philosophy

The idea that moral persons are autonomous and independent forms the basis of Price's political philosophy no less than of his ethics. He begins his *Observations on the Nature of Civil Liberty* with a discussion of four different kinds of liberty - physical liberty, moral liberty, religious liberty and civil liberty - arguing that the central idea running through them all is self-government. Physical and moral liberty are powers of self-determining agency, the latter, the power of following conscience. The third and fourth ideas, by contrast, are political rather than moral forms of liberty. A political order establishes religious liberty when each can exercise the form of religion they think best. And it establishes civil liberty when it is itself self-determining, governing itself 'without being subject to the impositions of any power' (1776: 3).

The central idea of Price's political philosophy is that the proper contrary of liberty is not hindrance, but slavery. A person is free when they can determine for themselves what they will do, and they are unfree when they are subject to arbitrary authority and control. The obstacles to moral liberty can be internal or external, political obstacles to 'liberty of the citizen' being primarily external. Citizens are unfree when they are subject to the will of others, most prominently, to the will of those who claim political authority over them. It is not enough for citizens to be free, moreover, that they face no actual obstacles to self-direction. Being subject to a master who can control at their pleasure is sufficient, even if the master chooses not to exercise control.

Price's central political tenet was thus a doctrine of the equal dignity of all persons as independent and autonomous. Legitimate political authority can derive only from the will of the people, therefore, and never once and for all but only continuously, subject to their continuing advice and consent. When numbers make direct democracy impossible, public officials function as representatives, their authority always deriving from and being answerable to, the will of the people. It follows also, Price argued, that a justified polity must ensure the equal liberty of all citizens. Citizens must be secured from threats to their liberty, whether posed by alien forces, by other citizens, or by government itself. Mastery by others deprives citizens, not just of valuable goods, but of their dignity as persons.

Price's enthusiasm for the American Revolution derived from his seeing it as the expression of these ideas. America provided the hope of a government rooted in equal dignity and so, Price thought, 'a new prospect in human affairs', 'beginning a new aera in the history of mankind' (1776-89: 17). To his American friends he recommended representative democracy, with legislators being frequently held to free elections, as well as a variety of measures designed to restrict entrenched power, including religious liberty and the abolition of primogeniture and all hereditary honours and titles. One thing that tempered his enthusiasm was the African slave trade (see Slavery) for which Price had nothing but scorn.

See also: Empiricism; Intuitionism in ethics; Moral realism; Rationalism; Utilitarianism

STEPHEN DARWALL

List of works

Price, R. (1758) *A Review of the Principal Questions and Difficulties in Morals*, London: Millar; ed. D.D. Raphael, Oxford: Clarendon Press, 1974.(Price's major work in ethics. This edition has a useful introduction.)

Price, R. (1764) 'An Essay towards solving a problem in the "Doctrine of Chances", by the late Rev. Mr. Bayes, F.R.S. communicated by Dr. Price, in a letter to John Canton, A.M., F.R.S.', *Philosophical transactions of the Royal Society* 1 (3): 370-418.(A presentation of what has come to be known as 'Bayes Theorem', which defines the relationship between the probability of a hypothesis given new evidence and the probabilities of this

hypothesis independently of the evidence, the evidence on the condition of the hypothesis obtaining, and the evidence itself. This result has significant importance for the theories of probability and rational decision.)

Price, R. (1765) "A Demonstration of the Second Rule in the Essay towards a solution of a problem in the Doctrine of Chances", *Philosophical Transactions of the Royal Society*, 1 (4): 296-325.(An improvement and development of the earlier paper.)

Price, R. (1767) *Four Dissertations*, London: Cadell.(Price's theology.)

Price, R. (1771) *Observations on Reversionary Payments*, London: Cadell and Davis.(Theoretical basis for life insurance.)

Price, R. (1772) *An Appeal to the Public on Subject of the National Debt*, London: Cadell.(On public finance.)

Price, R. (1776) *Observations on the Nature of Civil Liberty*, London: Cadell.(Price's major work of political philosophy.)

Price, R. (1776-89) *Political Writings*, ed. D.O. Thomas, Cambridge: Cambridge University Press, 1992.(A fine selection of Price's political philosophical writings includes selections from *Observations on the Nature of Civil Liberty* and *Observations on the Importance of the American Revolution*.)

Price, R. (1777) *Additional Observations on the Nature and Value of Civil Liberty and the War with America*, London: Cadell.(Further thoughts on politics and the American Revolution.)

Price, R. (1778) *A Free Discussion of the Doctrines of Materialism and Philosophical Necessity, in a Correspondence between Dr. Price and Dr. Priestley*, J. Johnson and T. Cadell, London.(Debate on free will and materialism.)

Price, R. (1784) *Observations on the Importance of the American Revolution and the Means of Making it a Benefit to the World*, London: Cadell.(Price's analysis of the American Revolution.)

Price, R. (1789) *A Discourse on the Love of Our Country*, London: Stafford and Cadell.(Includes Price's analysis of the French Revolution.)

References and further reading

Cudworth, R. (1731) *A Treatise on Eternal and Immutable Morality*.(A central rationalist work that much influenced Price.)

Hudson, W.D. (1970) *Reason and Right*, London: Macmillan.(A systematic discussion of Price's views in meta-ethics and normative ethics.)

Hutcheson, F. (1725) *An Inquiry into the Original of our Ideas of Beauty and Virtue; In Two Treatises*, London and Dublin: Darby.(A major empiricist work on which Price focused criticism.)

Thomas, D.O. (1977) *The Honest Mind: the Thought and Work of Richard Price*, Oxford: Clarendon Press.(The best introduction to Price's life and works - comprehensive and philosophically astute.)

Prichard, Harold Arthur (1871-1947)

One of the most influential Oxford philosophers of the twentieth century, Prichard was White's Professor of Moral Philosophy there from 1928 to 1937. His work combines epistemological realism and moral intuitionism. From 1906 onwards Prichard was active with the Oxford realists, who held, against idealists, that reality exists independently of mind, that knowledge is of reality, and that common-sense realism is correct. In ethics, he was the leader of the Oxford intuitionists who held, against utilitarianism, that common-sense morality is correct, its duties are known non-inferentially, and are an irreducible plurality of distinct kinds of act. His philosophical style displays concentration on specific problems, carefully using ordinary language to make precise distinctions in the absence of general theory. He influenced Oxford's next generation of Austin, Ryle, Hart and Berlin, who attended his classes and, occasionally, his 'philosophers' teas'.

1 Epistemological realism

Early in this century philosophy in Oxford was dominated by the opposition between idealists, such as T.H. Green, F.H. Bradley and Bernard Bosanquet, who held that mind constitutes reality, and the realists, such as J. Cook Wilson, H.W.B. Joseph, W.D. Ross and Prichard himself, who held that reality exists independently of mind (see Idealism). Prichard wrote his only book,*Kant's Theory of Knowledge* (1909), as a realist critique of Kant's transcendental analytic, and thereby developed the metaphysical and epistemological positions of the realist school. They are: (1) reality exists independently of mind; (2) if reality depends on mind then 'knowing is making', and mind makes reality. But (3) knowing is not making; knowing is being certain. Since we know when we know and know when we believe, as we show in choosing which word to use, it follows that knowledge is known in itself, while believing is known in relation to knowing (Urmson 1988).

Regarding perception, Prichard argues for a dualism of minds and objects and for common-sense realism. Prichard accuses Kant, and in other writings, Descartes, Locke, Berkeley, Hume and Russell, of wrongly hypostatizing 'appearances' when talking of how things themselves appear, thereby creating an intermediate entity variously called 'idea', 'representation', 'sensation', 'impression' or 'sense-datum'. These philosophers then quarrel over whether appearances copy reality or compose it. But they have given themselves an illusory problem.

2 Moral intuitionism

Prichard used the term 'intuitionist' to contrast his moral philosophy with that of the utilitarians. We can best approach his position by considering his greatest essay 'Does Moral Philosophy rest on a Mistake?' (1912). His affirmative response to this question is not modest, since the 'mistake' includes all moralists from Socrates to G.E. Moore. Their mistake is the assumption that there is something that all morally right actions have in common, from which their rightness can be inferred. According to Prichard, however, right actions possess no such quality. Thus moral knowledge is not inferential, but intuitive. One instance of this mistake is the belief that motives are the criterion of rightness. In discussing this view critically Prichard distinguishes two kinds of motive: 'moral', or acting from the conscientious desire to do what is right; and 'virtuous', or acting from an intrinsically good desire, such as compassion. Whichever kind one invokes, however, Prichard denies that it provides a criterion of right action, since right actions are sometimes done from bad motives. Utilitarianism and other consequentialist theories of duty provide another instance of the mistake. Prichard rejects traditional utilitarianism because it reduces all duties to the single duty of maximizing happiness, whereas in fact duties are essentially plural. He likewise rejects Moore's 'ideal' utilitarianism, which postulates a fundamental duty of maximizing intrinsically good outcomes of any kind, because it wrongly implies that there is a fixed hierarchy of types of intrinsically good outcomes (see Moore, G.E. §1). Thus Prichard rejects all consequentialist theories, and to the consequentialist claim that only consequences can resolve dilemmas rejoins that there is no single answer to dilemmas. Joseph, Prichard's fellow intuitionist, laments these conclusions, saying they 'leave moral duties in a heap' - as indeed they do. (see Consequentialism; Utilitarianism)

Despite his many criticisms of previous moral theories, Prichard never doubted that there is moral knowledge. How is it possible? Prichard's critics usually answer for him, 'We know by intuition', and suggest, by rolling their eyes and other such body language, that intuition is rapid, unhearing and irrational. But more may be learned by attempting a new approach which connects Prichard's realism and his 'intuitionism'. As a realist, Prichard

supported a common-sense realism which features common-sense knowledge of mind-independent objects. In morals, Prichard everywhere appeals to ‘ordinary, unreflective morality’. Thus, the parallel claim is that the acts that are really right are the moral duties of ordinary morality. This realism of common-sense morality is composed of the kinds of actions connected with promising, truth-telling, helping the helpless, returning what one has borrowed, paying debts, returning kindnesses and so on. The suggestion is that although the different kinds of ordinary obligations have different structures, a child can be taught to recognize them. When young, obedience may require the threat of punishment, but later, Smith, who has borrowed Brown’s electric drill, knows that he ought to return it because it is a loan and not a gift. From this basis, one can argue that intuition has a real object, namely the character of the kinds of moral action that are our ordinary obligations (see Intuitionism in ethics §2).

Once Prichard’s conception of intuitive moral knowledge is understood in this way, it becomes intelligible why he insists that before we know our duty we must consider an act’s circumstances (for instance, when telling the truth about a wayward, dying son to a grieving mother) and its consequences (for instance, when a surgeon discusses a gory operation over dinner). But has Prichard not eliminated consequences as a criterion of rightness? Apparently only as the single criterion, for he insists that no moral theory omits consequences. This insistence causes him to raise the question ‘What is the relationship between an action and its consequences?’ Logically, he cannot identify an action with its consequences or motives. He first suggests that an action is ‘setting oneself to effect a change’ and later he decides that ‘acting is willing’. Either way, Prichard clearly distinguishes ‘motive’ from ‘intended consequences’ and associates action with the latter. In turn, that permits him to argue that, depending on what will be taken for granted, we may include the intended consequences within the description of the act.

The remainder of Prichard’s moral philosophy is remarkable in its subtlety, but also for working upon problems implicit in ‘Does Moral Philosophy Rest on a Mistake?’ Thus, in ‘Duty and Interest’ ([1928] 1949) Prichard considers whether the sense of duty is a sufficient motive or if duty must be motivated by desire, and thereby enters the internalist versus externalist debate. In his lectures on T.H. Green, he states that philosophers have mistakenly sought one moral justification of political obligation, whereas different political states have different problems, requiring different justifications. Other implications are developed in ‘Duty And Ignorance Of Fact’ ([1932] 1949). In asking if duty depends upon ignorance of facts, Prichard divides the question with regards to common morality. Which is correct: subjective duty, which depends upon knowledge of facts, or objective duty, which identifies duty with what is independent of mind? I cannot have a duty to help my distant, ailing aunt if I do not know of her plight but, equally, I have a duty to help the helpless. This problem deadened his capacity to finish ‘a book on morals’. The essay ‘The Obligation To Keep A Promise’ ([*c*.1940] 1949) raises a problem akin to moral realism. First, however, Prichard is puzzled about the question ‘What is promising?’ ‘I promised’ is a statement that is either true or false. ‘I promise’ is not; it seems more like an action. But then promising, as doing, seems to create an obligation, an idea apparently foreign to Prichard’s thought. J.L. Austin admired Prichard, and their exchange of letters on promising as doing attests to Prichard’s influence on the formation of Austin’s ‘performative utterances’ (see Austin, J.L. §3; Performatives).

See also: Ordinary language philosophy, school of

JIM MacADAM

List of works

Prichard, H.A. (1906) ‘Appearances and Reality’, *Mind* 15: 223-9.(Prichard examines difficulties in two theories: ‘we know things only as they appear to us, and not as they are in themselves’ and ‘we know only "phenomena" or "appearances"’. Since both theories deny realism, Prichard finds both unacceptable.)

Prichard, H.A. (1907) ‘A Criticism of the Psychologists’ Treatment of Knowledge’, *Mind* 16: 27-53.(Prichard writes, ‘When… we perceive a chair or a tree, the direct object of the mind is the chair or tree, and not any mental modification referred to as a ‘perception’ or an image of it.)

Prichard, H.A. (1909) *Kant’s Theory of Knowledge*, Oxford: Clarendon Press.(Prichard provides a common-sense realist criticism of Kant’s Analytic. The argument is helpful to understanding Prichard’s own theory of knowledge and should be considered by students of Kant.)

Prichard, H.A. (1910) ‘Philosophic Pre-Copernicanism: An Answer’, *Mind* 19: 541-3. (Prichard defends naive realism against idealism: ‘Study [of idealism] appears to resemble measles in that all students of philosophy must undergo an attack of it’.)

Prichard, H.A. (1912) ‘Does Moral Philosophy Rest on a Mistake?’, in Pritchard 1949.(Prichard’s richest work: it

provides him with problems that he struggles with for the rest of his life.)
Prichard, H.A. (1919) ‘Professor John Cook Wilson’, *Mind* 28: 297-318.(In Prichard’s memoir of John Cook Wilson, he writes: ‘To speak of him dispassionately as a philosopher is difficult for one who, like the present writer, enjoyed uninterrupted intercourse with him since he first became his pupil some five and twenty years ago’.)
Prichard, H.A. (1928) ‘Mr Bertrand Russell’s *Outline of Philosophy*’, *Mind* 37: 265-82. (A pugnacious attack on Russell’s doctrines on knowledge and perception in *Outline of Philosophy*: ‘I should like to ask Mr Russell… whether there really is a word of truth in this view from beginning to end’.)
Prichard, H.A. (1949) *Moral Obligation, Essays And Lectures*, ed. and with a note by W.D. Ross, Oxford: Clarendon Press; revised and extended as *Moral Obligation and Duty and Interest, Essays and Lectures*, with intro. by J.O. Urmson, Oxford: Oxford University Press, 1968.(Contains all of Prichard’s published moral writings, including ‘Duty and Interest’ (1928), ‘Duty and Ignorance of Fact’ (1932) and ‘The Obligation to Keep a Promise’ (*c*.1940).)
Prichard, H.A. (1950) *Knowledge and Perception, Essays and Lectures*, ed. and with a preface by W.D. Ross, Oxford: Clarendon Press.(A collection containing Prichard’s most important nonmoral texts.)

References and further reading

Broad, C.D. (1950) ‘Critical Notice, Moral Obligation: Essays and Lectures by H.A. Prichard’, *Mind* 59: 55-6.(An insightful, sympathetic review by an able philosopher.)
Collingwood, R.G. (1939) ‘Minute Philosophers’, in *An Autobiography*, Oxford: Oxford University Press, ch. 3. (A biased account of the Oxford realists, but the book is well worth reading. Chapter 6, ‘The Decay of Realism’, is also relevant, since Collingwood had been a realist himself.)
Cook Wilson, J. (1929) *Statement and Inference, with other Philosophical Papers*, ed. A.S.L. Farquharson, Oxford: Clarendon Press, 2 vols; repr. 1969.(Writings of leader of Oxford realists, interesting in their own right, and containing letters to Prichard.)
Hornsby, J. (1980) *Actions*, London: Routledge & Kegan Paul.(Hornsby provides the most plausible interpretation of Prichard’s account of action.)
Joseph, H.W.B. (1931) *Some Problems In Ethics*, Oxford: Oxford University Press.(A discussion of Prichardian themes by Prichard’s fellow Oxford intuitionist and next-door neighbour.)
Mabbott, J.D. (1966) *An Introduction to Ethics*, London: Hutchinson University Library.(An easily read introduction to Prichard’s type of moral philosophy.)
Passmore, J. (1957) ‘Cook Wilson And Oxford Philosophy’, in *A Hundred Years of Philosophy*, London: Gerald Duckworth, ch. 10.(Passmore’s well-written history is the best introduction for the beginner.)
Price, H.H. (1947) ‘Harold Arthur Prichard, 1871-1947’, *Proceedings of the British Academy* 33: 331-50, Oxford: Oxford University Press.(Contains the best short account of Prichard’s philosophy.)
Prior, A.N. (1951) ‘The Virtue Of The Act And The Virtue Of The Agent’, *Philosophy*, 26: 121-30.(Prior contends convincingly that Prichard’s and Ross’s struggle to decide the correctness of subjective or objective duty was argued by Price and others of the British Moralists.)
Ross, W.D. (1930) *The Right And The Good*, Oxford: Clarendon Press.(Ross professes ‘my main obligation is to Professor H.A. Prichard’, but also acknowledges his debt to G.E. Moore.)
Ross, W.D. (1939) *Foundations of Ethics*, Oxford: Clarendon Press.(Both of his books contain insightful and influential arguments on the Oxford intuitionist position.)
Urmson, J. (1988) ‘Prichard and Knowing’, in J. Dancy, J. Moravcsik and C.C.W. Taylor (eds) *Human Agency: Language, Duty and Value*, Stanford, CA: Stanford University Press.(Contains the best discussion of Prichard’s account of knowledge.)
Walsh, W.H. (1981) ‘Kant’s Critique of Pure Reason: Commentators in English, 1875-1945’, *Journal of the History of Ideas* 42: 723-37.(Walsh does not praise Prichard as a Kant scholar, but recommends Prichard’s *Kant’s Theory of Knowledge* (1909) as impressive and as raising important questions regarding Kant’s philosophy.)

Priestley, Joseph (1733-1804)

A major figure of the British Enlightenment, Joseph Priestley is best known as a scientist and for his discovery of oxygen, though he was by profession a theologian, and also wrote on politics and education - more, indeed, than on science or metaphysics. His philosophical speculations were generally brought to support his theological arguments and were usually structured in a rhetorical rather than in a formal, systematic mode. He was a Unitarian in theology, an associationist, determinist and monist, with a curiously spiritualized materialism dependent, in part at least, on his scientific studies.

Joseph Priestley was born near Leeds, Yorkshire, England in 1733. He was self-educated in science and philosophy, reading on his own Locke's *Essay concerning Human Understanding* (1689) (see Locke, J.), Isaac Watts' *Logic: or the Right Use of Reason* (1725), and Willem Jacob Gravesande's Newtonian *Mathematical Elements of Natural Philosophy* (1720-1). At 19 he went to Daventry Academy to become a minister. Daventry's theological text was Philip Doddridge's *Course of Lectures on the Principle Subjects in Pneumatology, Ethics and Divinity* (printed 1763), from which Priestley learned of the Cambridge Platonists and of the Newtonian physico-theologians. He also read, at Daventry, David Hartley's *Observations on Man, His Frame, His Duty, and His Expectations* (1749), a work of greatest influence on him, converting him to associationism and to the doctrine of philosophical necessity (determinism) (see Hartley, D.).

Initially failing as a dissenting minister, Priestley became a successful teacher of languages and *belles-lettres* at Warrington Academy. There also he began his scientific career, writing his *History and Present State of Electricity* (1767), to which were added experiments and observations of his own. In 1767 Priestley became a Unitarian minister in Leeds. There he commenced his career as a religious polemicist, wrote an *Essay on the First Principles of Government* (1768), from which Jeremy Bentham derived his utilitarianism, and continued his scientific work. For the *History and Present State of Discoveries relating to Vision, Light and Colours* (1772), he adopted some ideas of Rudjer Josef Boscovic, which he would subsequently use for his justification of materialism. He also wrote his first paper there, published in the *Philosophical Transactions* (1772) of the Royal Society of London, on the chemistry of gases.

He left Leeds in 1773, to become librarian and companion to Lord Shelburne. During this period he did the major part of his pneumatic chemistry and the major part of his formal metaphysical writing. Leaving Shelburne in 1780, he settled as a minister in Birmingham. There he spent much time defending his scientific and metaphyical ideas and continuing his religio-political attacks on Church establishment - to the point, finally, of inciting a riot which drove him from Birmingham to London in 1791 and then, in 1794, into exile in the USA. There he continued the activities of his Birmingham years, until he died at Northumberland, Pennsylvania in 1804.

Priestley's explicit attention to philosophical writing began with *An Examination of Dr. Reid's Inquiry …On the Principles of Common Sense, &c.* (1775), attacking Scottish Common Sense Philosophy for its 'vain multiplication' of separate, arbitrary, instinctive principles to explain perceptions (see Common Sense School; Reid, T.). To disprove this naïve nativism, Priestley undertook an edition of selected parts of Hartley's *Observations on Man*, published as *Hartley's Theory of the Human Mind* (1775). Shorn of its specific references to religion and of the physiological doctrine of vibrations, this was, none the less, the avenue by which David Hartley's previously neglected *Observations on Man* was made known to the eighteenth and early nineteenth centuries.

Only in his first introductory essay to *Hartley's Theory of the Human Mind* does Priestley go beyond the *Observations on Man*, in which he was led into the most original of his metaphysical writing. In summarizing the physiological base suggested for associationism, Priestley asserted that Hartley's 'elementary' body mediating between mind and brain was unnecessary, and declared: 'I rather think that the whole of man is of some uniform composition, and that the property of perception…is the result of such an organical structure as that of the brain'.

This declaration provoked an outcry, prompting Priestley to write his *Disquisitions relating to Matter and Spirit* (1777). Into the *Disquisitions* Priestley poured the results of his thinking over the previous decades. 'Modern philosophical dualism' had no support from Scripture and led to the insolvable problem of interacting entities,

matter and spirit, defined so that there was nothing in common between them. It also took no heed of recent discoveries in electricity, optics and chemistry. Far from being solid, inert and impenetrable, matter could be defined as a ‘compages’ of geometrical points, surrounded by spheres of alternating repulsive and attractive force.

With this publication no defence was sufficient to escape from accusations of materialist atheism, and that charge was enhanced by the publication of Priestley’s *Doctrine of Philosophical Necessity Illustrated* (1777), in which he insisted that people’s actions were determined by laws of nature. Priestley’s arguments were ingenuous, emphasizing similarities between his doctrine and that of freewill: ‘All the liberty… that I say a man has not, is that of doing several things when all the previous circumstances… are precisely the same’. None the less, the doctrine of philosophical necessity was representative of the relation of cause and effect, from which Priestley derived his arguments for the being of God, and he insisted that will is determined by motive, and motive by state of mind and view of things. When these remain the same, people will always make the same choice and determination.

Attacked for his views, Priestley responded - in such books as *A Free Discussion of the Doctrines of Materialism and Philosophical Necessity* (1778) - that nothing could be known of matter but its properties; that his monism could be called ‘wholly spiritual’ as easily as ‘wholly material’; and that, by his description ‘matter is…resolved into nothing but the *divine agency*, exerted according to certain rules’.

These proposals were too radical, and his anti-establishment persona too obvious, for Priestley’s ideas to win any permanent acceptance in Britain. His achievements in all but science were soon forgotten, and Priestley has remained one of the neglected philosophical writers of his age.

ROBERT E. SCHOFIELD

List of works

Priestley, J. (1972) *The Theological and Miscellaneous Works of Joseph Priestley, LL.D. F.R.S. &c.*, ed. J.T. Rutt, New York: Kraus, 1972, 25 vols.(Contains all the philosophical works, but on the prefaces of the science books. Most of the latter, however, were reprinted separately by Kraus.)

References and further reading

Crook, R.E. (1966) *A Bibliography of Joseph Priestley 1733-1804*, London: Library Association.

Doddridge, P. (1763) *Course of Lectures on the Principle Subjects in Pneumatology, Ethics, and Divinity: with References to the most considerable Authors on each Subject*, ed. S. Clark, London: J.W. Clarke and R. Collins, W. Johnston, J. Richardson, S. Crowder and Co., T. Longman, B. Law, T. Field, and H. Payne and W. Cropley.(Doddridge may have given these lectures from the time of founding his Dissenting academy at Northampton in 1729. The lectures were used at Daventry Academy at least as late as 1775. They were printed for the first time in the edition cited here and were several times re-edited, with changes in the list of references, and republished (an early nineteenth-century American edition was published) but the edition of 1763 must be very like the form that Priestley used in 1750s.)

Gibbs, F.W. (1965) *Joseph Priestley: Adventurer in Science and Champion of Truth*, London: Nelson.(Standard full biography, but lacks complete documentation.)

Hartley, D. (1749) *Observations on Man, His Frame, His Duty, and His Expectations*, London: Leake & Fredrick; repr. Hildesheim: Olms, 1967.

Schofield, R.E. (1997) *The Enlightenment of Joseph Priestley: A Study of His Life and Works from 1733 to 1773,* University Park, PA: Pennsylvania State University Press.(A detailed biography, but only of the first forty years of Priestley’s life.)

Schofield, R.E. (1975) ‘Priestley, Joseph’, in *Dictionary of Scientific Biography* XI: 139-147, ed. C.C. Gillispie, New York: Scribner.

McEvoy, J.G. (1978-9) ‘Joseph Priestley, "aerial philosopher": Metaphysics and Methodology in Priestley’s thought 1772-1781’, *Ambix* XXV: 1-55, 93-116, 153-75; XXVI: 16-38.(A contrary view to that of Schofield.)

Primary-secondary distinction

The terminology of 'primary and secondary qualities' is taken from the writings of John Locke. It has come to express a position on the nature of sensory qualities - those which we attribute to physical objects as a result of the sensuous character of sensations they produce when they are perceived correctly by us. Since our senses can be differentiated from each other by the type of sensations they produce, sensory qualities are what Aristotle called 'proper sensibles' - those perceptible by one sense only. Colours, sounds, scents and tastes are always regarded as proper to their respective senses. What are the proper sensibles of touch, and whether there is similarly a single family of them, is a matter of controversy; but temperature at least is standardly regarded as proper to this sense. It is such sensory qualities that are candidates for being given the status of secondary qualities.

To regard sensory qualities as secondary is to hold that an object's possession of one is simply a matter of its being disposed to occasion a certain type of sensation when perceived; the object in itself possesses no sensuous character. Primary qualities, by contrast, are those which characterize the fundamental nature of the physical world as it is in itself. They are always taken to include geometrical attributes, and often some space-occupying feature; Locke's candidate for this latter was solidity. Although the terminology dates from the seventeenth century, this general doctrine goes back to the Greek atomists.

1 Locke's account

Although the distinction between primary and secondary qualities is of ancient lineage, the terminology became current through the writings of John Locke, who was consciously drawing on the work of his older contemporary Robert Boyle. Although a not unrelated distinction between primary and secondary attributes of matter is to be found in Medieval philosophy, when Boyle writes of 'secondary qualities, if I may so call them', he indicates that he took himself to be coining an expression to bear the sense that interests us. The distinction as it is drawn by Boyle and Locke is not one that centrally concerns the nature of sensory qualities or the relation between sensory states and physical reality, but rather the issue of what are fundamental and what are derivative attributes of material objects. This is clearly seen when one considers Locke's treatment of what are sometimes termed 'tertiary qualities' - such as a fire's power to melt wax. Such qualities are in fact termed 'secondary' by Locke; they differ from what are usually regarded as secondary qualities - that is, sensory qualities - only in that the latter are immediately and the former are mediately perceived. By 'secondary' Locke in fact means *derived*, *resultant* or *supervenient*.

In order best to understand this, we should bear in mind the hypothesis that Locke adopts as the most plausible account of the material world: namely, corpuscularianism, or what we would call atomism (although Boyle defines a corpuscle as a coalition of *minima naturalia* - which are the genuine atoms - so that a corpuscle is closer to our modern idea of a molecule) (see Atomism, ancient). According to this view, everything material either is an atom - a small, naturally indivisible, solid body - or is a complex body consisting entirely of a number of such atoms. In any such complex body the constituent atoms will be arranged in a certain way, the technical term for such an arrangement being 'texture'. By 'quality' Locke means any feature of an object in virtue of which it has the power to affect another object. In virtue of the texture possessed by a complex object, as well as, perhaps, in virtue of the constituent atoms moving in relation to one another in the body (for example, vibrating), that object will possess certain powers - powers that no noncomplex, nontextured object could possess. All such powers are secondary qualities, since they derive from, or are secondary to, the arrangement and motion of the atoms constituting a complex body.

Thus there is no question for Locke as to whether a single atom may possess any secondary quality; such a thesis is analytically false. The only genuine question in this area concerns which of our ideas are excited by such derivative powers. It is an immediate consequence of this that no secondary quality is a universal or, as Boyle and Locke put it, 'catholic' attribute of matter, since they derive from particular, contingent configurations of atoms in complex bodies. By contrast, primary qualities are non-derivative, fundamental features of matter, and hence are possessed by everything material as such - even individual, simple atoms. ('Hence', because such writers regarded the idea that a quality could be underived and yet non-catholic - by having a bare, brute presence in some bodies but not in others - as a nonstarter. Primary qualities are fundamental in that they are essential attributes of matter as

such.) Roundness, for example, can count as a primary quality even though not everything material is round, since it is but a determination of figure, which is possessed universally by body. Secondary qualities are immediately perceived when they are powers to occasion perceptual sensation in us, mediately perceived when they are powers to effect changes in other material objects, which latter changes alone we perceive.

One might object that the figure of a cricket ball ought to be a secondary quality since it results from an arrangement of the ball’s constituent corpuscles. However, as far as the perception of the ball’s shape is concerned, the cricket ball could just as well be a single, very large atom; the shape of the mass of configured atoms that is the ball is identical to the shape that a single atom may possess. We can see cricket balls and we cannot see individual atoms; however, the power of a cricket ball to appear round is not crucially dependent on the inner complexity of the ball, but simply on its size; atoms are simply too small to see. Another way of putting this introduces a second defining feature of the primary-secondary quality distinction: primary qualities excite ideas in us which *resemble* those qualities, whereas secondary qualities do not. With secondary qualities mediately perceived, this is obvious; we have no (pre-theoretical) idea of what it is in the fire that gives it the power to melt wax. Locke says that such qualities are generally agreed to be ‘barely powers’. By this he of course does not mean that these powers are brute and ungrounded: they are grounded in the complex texture and perhaps motion of the object’s constituent atoms. What he means is that they are barely powers as far as our ideas of them are concerned. Locke’s view is that we are really not much better off when it comes to secondary qualities immediately perceived. Certain complex bodies have the power to excite various sensory states in us; it is in virtue of the texture and motion of an object’s constituent atoms that it appears, for example, coloured or warm or scented to us. There is no colour or warmth or scent in objects distinct from the arrangement and perhaps motion of such objects’ constituent atoms. So an object looks coloured because of such atomic facts; whereas, in contrast, a round object standardly looks round because it is round. This is all that Locke’s much-criticized notion of resemblance means.

2 The standard account

Primary and secondary qualities are usually not spoken of today as they were by Boyle and Locke, George Berkeley being the principal instrument of change. In Berkeley we find two switches in terminology. First, the term ‘secondary quality’ is restricted to sensory qualities. Hence the central philosophical issue in this area becomes that of the relation between our senses and the nature of matter. Second, upholders of the primary-secondary quality distinction are said to hold that secondary qualities exist only in us - colours, scents and so on are merely sensations in the mind. ‘Secondary’ thus came to mean, or at least connote, ‘in the mind’ rather than ‘derivative’. As an interpretation of Locke, this is half right and half wrong. For Locke, secondary qualities are certainly in bodies themselves. However, colours and so forth are merely in the mind. This is because colour, for example, is not a secondary quality for Locke; colour exists only in us, or in idea. When he wishes to advert to the corresponding secondary quality, he will use a phrase such as ‘colour as it exists in bodies’. Because of all this, the precise meaning of ‘secondary’ is somewhat indeterminate in modern writings, although the general issue at stake is clear: do the sensuous features that we experience when we perceive physical objects inherently and irreducibly characterize the physical objects themselves? Does the world as it is in itself possess sensuous character?

3 Reasons for the distinction

Two sets of reasons exist for drawing our distinction: the first consists of reasons that are broadly empirical in nature, the second a priori. Under the first would fall even Locke’s observation that secondary qualities can be altered by a merely mechanical process: for example, pounding certain substances changes their colour. How can this be unless such qualities are wholly subsumable under a merely mechanical account of the world? More recent discoveries concerning the precise nature of sensory perception allow a more detailed identification of such qualities with scientifically specifiable states of bodies. If, for example, an object’s taste just is that in a body which gives rise under normal conditions to gustatory experiences when that body is placed in the mouth, and we discover that what thus causes such experiences is some chemical feature of that body, then taste must be identified, without remainder, with such a feature. By contrast, the best account available of why, for example, certain objects look round is that they are round.

A priori considerations must also be relevant, however, since roughly the same set of qualities have been given secondary status for over two thousand years. Certain qualities of bodies can be singled out a priori as being

sensory in a distinctive sense. It is not just that such qualities are perceptible; but also that we must enjoy a particular kind of experience in order to cognize them. It suffices to know what a circle is to know that it is a figure, all the points in the boundary of which are equidistant from a single point. To grasp this, no particular kind of perceptual experience is necessary. In order, however, to understand what colours, scents and so forth are, we must enjoy perceptual experiences in the relevant sense modalities. Sensory qualities are sense-specific in that cognition of them requires the enjoyment of sensation that is definitive of a specific sense modality. The issue of whether sensory qualities are secondary or not is the issue whether the very character of the sensation that we enjoy when we perceive a sensory quality is itself possessed by the object perceived. Is, say, the quale of redness which we experience when we see a red-looking object, inherently and irreducibly present in red objects, such that its presence by itself constitutes them as red?

Two distinct reasons may be offered for giving a negative answer to this. First, even if physical objects possessed such qualia, that fact would not constitute them having sensory qualities such as colours and scents. This is because it is possible that objects should possess such sensuous features in a manner that was irrelevant to the generation of perceptual experiences in us. Since, however, objects necessarily possess those sensory qualities they are discriminated as having by competent observers in normal circumstances, and since necessarily sensory qualities are the basis in objects of such discriminations, such a possibility shows that the possession by physical objects of such qualia does not logically suffice for possession of the respective sensory qualities: possession of such qualia does not *of itself* constitute the possession of any sensory quality. It remains, perhaps, a priori possible that such qualia should be contingently identifiable with sensory qualities - if they are indeed the physical basis of sensory discriminations; for a decision that this is in fact contingently false we need to turn to empirical considerations. Perhaps, however, certain high-level theoretical considerations, of a kind which seem to have motivated Locke, and which are not nicely categorized as either empirical or a priori, may count against even such a possibility: for example, the irremediably non-mechanical nature of such supposed physical qualia doubtless conflicts with an involvement in perceptual transactions with the physical world that perhaps must be at least partly mechanical in nature; certainly one of the seventeenth century's chief objection to the postulation of sensory qualities as 'real' was that they are ultimately unintelligible.

The second a priori consideration goes further and denies the very possibility of sensory qualia being merely physically realized. From the fact that sensory qualia determine sense modalities it is concluded that any such quale is an inherent feature of sensation. Nothing which is an inherent, qualitative characteristic of sensation could also be a feature of an insentient physical object, as sensory qualities are supposed to be. For example, for colour to be instantiated is for a sentient subject to be visually conscious is a certain manner. On this view there is as much chance of a mere physical object being sensuously coloured as there is of the pin that pricks you feeling pain.

Since the issue of secondary qualities concerns the nature of sensory qualities, whereas the issue of primary qualities concerns the fundamental nature of matter as material substance, philosophers can be and have been in substantial agreement over which qualities are secondary, while yet being in serious disagreement over the list of primary qualities. Although, perhaps, the issue of the secondary status of sensory qualities is amenable to a priori philosophical reflection, since the issue of primary qualities concerns nothing other than the fundamental concepts of physics, it is doubtless hubristic to expect any a priori settling of the question.

Finally, it is worth mentioning that the disjunction between primary and secondary qualities is not exhaustive - many qualities of physical objects are clearly neither physically fundamental nor sensory.

See also: Qualia; Secondary Qualities

A.D. SMITH

References and further reading

Alexander, P. (1985) *Ideas, Qualities and Corpuscles*, Cambridge: Cambridge University Press.(Detailed, scholarly treatment of the aspects of Locke's philosophy relevant to this topic.)

Boyle, R. (1666) *The Origin of Forms and Qualities According to the Corpuscular Philosophy*, repr. in *Selected Philosophical Papers of Robert Boyle*, ed. M.A. Stewart, Manchester: Manchester University Press, 1979. (Very readable account, and the source of Locke's ideas on this subject.)

Locke, J. (1689) *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Oxford University

Press, 1975.(The most famous and influential traditional treatment of the subject.)
McGinn, C. (1983) *The Subjective View*, Oxford: Clarendon Press.(One of the best modern accounts of the issues.)
Sellars, W. (1963) 'Philosophy and the Scientific Image of Man', in *Science, Perception and Reality*, London: Routledge & Kegan Paul.(One of the most influential modern science-based accounts of the distinction.)
Smith, A.D. (1990) 'Of Primary and Secondary Qualities', *Philosophical Review* 99 (2): 221-54.(A recent account dealing with both historical and substantive issues concerning the distinction.)

Prior, Arthur Norman (1914-69)

Prior is most often thought of as the creator of tense logic. (Tense logic examines operators such as 'It will be the case that' in the way that modal logic examines 'It must be the case that'.) But his first book was on ethics, and his views on metaphysical topics such as determinism, thinking, intentionality, change, events, the nature of time, existence, identity and truth are of central importance to philosophy. Using methods akin to Russell's in his Theory of Descriptions, he showed that times, events, facts, propositions and possible worlds were logical constructions. For example, we get rid of events by recognizing among other things that to say that the event of Caesar's crossing the Rubicon took place later than the event of Caesar's invading Britain is to say that it has been the case that both Caesar is crossing the Rubicon and it has been the case that Caesar is invading Britain. The title of the posthumous work, Worlds, Times and Selves (1977), indicates the breadth and depth of his thought. He is also fun to read. He died at the age of fifty-four, at the height of his powers.

1 Life

Arthur Norman Prior was born in New Zealand. He entered the University of Otago in 1932, taking his BA in philosophy in 1935 and his MA in 1937. He studied under J.N. Findlay, working with him through Prantl's history of logic. Findlay made him an assistant lecturer in 1937, but after a year he left the university. The next eight years were spent in a wandering fashion, in Italy, France and Britain, in the New Zealand Air Force and as a freelance journalist. Prior's first marriage ended in 1943. Later that year he married Mary Wilkinson, and this marriage brought him great happiness and a collaborator in every aspect of his work. His writings, at first mostly theological but becoming increasingly philosophical, attracted notice, and in 1946 he succeeded Karl Popper as lecturer at Canterbury University College at Christchurch in New Zealand, becoming professor in 1952. His first book, *Logic and the Basis of Ethics*, was published in 1949. By this time Prior's interests had become centred on logic, and in 1951 he submitted to his publishers a work which eventually saw the light as *Formal Logic* in 1955. In this, as in his subsequent works, he used the logical symbolism invented in Poland by Jan Łukasiewicz. Prior had close ties with Polish logicians, lecturing in Warsaw in 1961. He was John Locke Lecturer at Oxford University in 1956. In 1959 he returned to England to take up a chair at Manchester University. He also taught for periods in Chicago and Los Angeles. In 1966 he became a Fellow of Balliol College, Oxford. *Past, Present and Future* and *Papers on Time and Tense* were published in 1967 and 1968. Oxford gave him a Readership with effect from October 1969, but it was in that very month that his sudden death occurred, while he was on a visit to Norway. Various of his works were prepared for publication after his death, most notably *Objects of Thought* in 1971.

2 Tense logic

Prior became a player on the international philosophical scene with his delivery of the John Locke Lectures at Oxford in 1956, published in 1957 as *Time and Modality*. In these he explored analogies between the logic of necessity and possibility and the logic of past, present and future. He had earlier written an article on the Greek philosopher, Diodorus Cronus, who tried to reduce necessity to what is and always will be and possibility to what either is or at some time will be. This involved the use of tense operators, most importantly *Pp* and *Fp*, 'It has been the case that *p*' and 'It will be the case that *p*', having the same syntax as *Lp* and *Mp*, 'It is necessary that *p*' and 'It is possible that *p*'. Thus began a whole new branch of logic, 'tense logic', which for some years Prior developed single-handedly. *Past, Present and Future* and *Papers on Time and Tense* together traced the way the theory had grown and displayed the state of the art at that moment. As with all his logical work, systems were not created, proofs constructed or definitions given just for their own sake - for elegance or economy - but because philosophical issues were always being addressed. He believed that to regard tensed propositions as basic to the expression of facts about time was to give an affirmative answer to the metaphysical question whether or not, as it has become fashionable to say, 'time is tensed'. His article 'Thank Goodness That's Over' (1959) contains his best argument for this position. He found that different choices of axioms in tense logic could be used to characterize time as unending, cyclical, discrete, continuous, and so on. The problem of determinism could also be given expression in formal terms, and philosophical views about ceasing to exist and coming into existence prompted modifications of tense-logical systems. Prior believed that propositions expressing facts about time in terms of relations between events or between events and times were reducible to propositions where time determinations were expressible by means of operators on sentences. This itself has implications for 'ontological' issues: we do

not need to rank times or events as 'basic particulars'. Again, Prior's use of simple '*p*' as opposed to '*Pp*' or '*Fp*' to represent present-tensed propositions led naturally to a 'redundancy theory' of the present, parallel in some ways to the redundancy theory of truth. It is raining now and it is true that it is raining if and only if it is raining (see Tense and temporal logic; Truth, deflationary theories of).

3 Truth

The use of tense operators, like that of modal operators, is a challenge to extensionalism: the material equivalence of *p* and *q* is not sufficient to guarantee that of *Fp* and *Fq* any more than that of *Mp* and *Mq*. Prior was uninhibited by extensionalist qualms, and equally uninhibited in his use of quantifiers to bind variables of categories other than that of names. He was accordingly happy to use quantificational formulas such as 'for every *p*' or 'for some *q*', following here the precedent of Łesniewski's 'Protothetic' (see Łesniewski, S. §4). As well as the tense and modal operators, he made room for expressions like 'Robinson says that' or 'Jones thinks that' which form sentences out of sentences. The verbs of propositional attitude he described as predicates at one end and connectives at the other: to make 'said that' into a proposition I have to put something like 'Winnie-the-Pooh' in front of it and something like 'bees like honey' after it. Equipped in this way Prior is able to go ahead and paraphrase 'Jones believes whatever Robinson believes' by 'For every *p*, if Robinson believes that *p*, Jones believes that *p*'. He is thus able to give an analysis of 'true' in terms of quantifiers binding propositional variables. 'What Pooh said is true' means, roughly, 'For some *p*, Pooh said that *p* and *p*'. Quine and Geach had frequently drawn attention to the analogies between the use of bound name-variables in quantificational formulas and the use of pronouns in natural languages. Since bound propositional variables correspond to sentences as bound name-variables do to names, Prior talks of them as 'prosentences'. This expression, first used by Prior in his article on the correspondence theory of truth in the Encyclopaedia of Philosophy edited by Paul Edwards, has since been used by Dorothy Grover and others to give the name 'prosentential theory of truth' to Prior's account of truth (see Truth, deflationary theories of).

4 Logical truth

Prior had something to say not only about truth, but also about logical truth. He could make a profound point in a short paper apparently designed more as fun than as a serious contribution to philosophy. A perfect example of this is 'The Runabout Inference Ticket' (1960). Empiricists seek protection from the embarrassment caused them by our knowledge of logical truths by claiming that this sort of knowledge is due entirely to conventions of language. The necessary truth of propositions of the form 'If *p*, then either *p* or *q*' rests, on this view, on the fact that we have chosen to adopt conventions for the use of 'if' and 'or' which make it 'true by convention'. Prior described a new convention. We could introduce a word 'tonk' into our vocabulary. The conventions for using 'tonk' determine that from any proposition '*p*' it is possible to infer a corresponding proposition '*p* tonk *q*', and from '*p* tonk *q*' it is possible to infer '*q*'. Putting these together it would be legitimate on the basis of these conventions to infer any proposition from any other proposition. Every proposition of the form 'If *p* then *q*' would be a logical truth. Something has gone wrong: Prior leaves it to the reader to work out what that is.

5 Propositional attitudes and intentionality

When he died, Prior left a manuscript entitled *Objects of Thought*. The book has two parts, corresponding to the two senses of the phrase which is its title: 'What we think' and 'What we think of'. What we think are in a sense propositions, but Prior's understanding of 'proposition' is determined by his view, already described, of the syntax of sentences like 'Pooh thinks that bees like honey'. He regards propositions as 'logical constructions'. What this means is that sentences which have as their grammatical subjects expressions which seem to designate propositions mean no more than equivalent sentences in which no expression plays this role. Thus 'The proposition that Paris is in Scotland is false' means no more than 'It is not the case that Paris is in Scotland'. This is not saying of a strange object called a 'proposition' that it has the equally strange property of not being the case: it is saying something about Paris and Scotland. Similarly, in Prior's view, 'The proposition that Paris is the home of vice is believed by McTavish' does not ascribe the property of being believed by McTavish to some proposition, but asserts that the property of being the home of vice in the eyes of McTavish belongs to Paris. The deep grammar of phrases like 'McTavish believes that' is the same as the surface grammar of 'in the eyes of McTavish': both belong to the category of adverbial phrases.

If what we think are in Prior's view logical constructions, what we think of are common-or-garden objects, rather than 'intentional' objects. One important consideration which led Frege and others to the view that in 'McTavish believes that Paris is the home of vice' the word 'Paris' designates an intentional object is that the sentence may change its truth-value if we substitute for 'Paris' another name of the same city. This supposed phenomenon was called by Quine 'referential opacity'. Prior, in 'Is the Concept of Referential Opacity Really Necessary?' (1963), maintained that the change of truth-value would occur only if the expression substituted for 'Paris' was a definite description, for example, 'the capital of France', and only then if the definite description had what Russell called 'secondary occurrence'. Substitution of names for names, when both are being used as names, does not, in Prior's view, lead in any case to change of truth-value. Nevertheless Prior is not an extensionalist, since he allows that substitution of coextensional predicates and materially equivalent propositions cannot always be made *salva veritate*. But Prior found difficulty in deciding when a name really was being used as a name, when McTavish's thought was directly about Paris, the city. Like Russell, he came to the view that few of our thoughts are directly about their objects. These are issues he debates in the second part of *Objects of Thought* (see Propositional attitude statements).

6 Logic and ethics

Prior was at one time an editor of *The Journal of Symbolic Logic*, and his output in and contribution to the field of formal logic was considerable, though impossible to summarize in an entry of this length. At the same time, the early interest he showed in ethics was never extinguished. His work in this field, as in others, is notable for his learning in respect of the history of the subject, particularly, in this case, in the area of the seventeenth- and eighteenth-century British moralists. A good sample of his ability to put logic to work in the service of ethics is his essay 'The Autonomy of Ethics' (1960), reprinted in the collection, *Papers in Logic and Ethics* (1976a).

C.J.F. WILLIAMS

List of works

Prior, A.N. (1949) *Logic and the Basis of Ethics*, Oxford: Clarendon Press.(A logician's view of the so-called 'Naturalistic Fallacy' and its seventeenth- and eighteenth-century precursors.)

Prior, A.N. (1955) *Formal Logic*, Oxford: Clarendon Press.(A textbook providing an introduction to the propositional calculus, the predicate calculus, syllogistic and the logic of classes.)

Prior, A.N. (1957) *Time and Modality*, Oxford: Clarendon Press.(Prior's first attempt to apply the results of modal logic to a logic of tenses.)

Prior, A.N. (1959) 'Thank Goodness That's Over', *Philosophy* 34: 12-17; repr. in Prior, 1976a. (An argument for the indispensability of the 'A-series'.)

Prior, A.N. (1960) 'The Runabout Inference Ticket', *Analysis* 21: 38-9; repr. in Prior (1976a).(A *reductio ad absurdum* of the conventionalist account of logical truth.)

Prior, A.N. (1963) 'Is the Concept of Referential Opacity Really Necessary?', *Acta Philosophica Fennica* 16: 189-99. (This shows how distinctions of scope can provide an adequate explanation for failures of Leibniz's Law.)

Prior, A.N. (1967) *Past, Present and Future*, Oxford: Clarendon Press.(Together with the following volume, the main exposition of tense logic as Prior invented it.)

Prior, A.N. (1968) *Papers on Time and Tense*, Oxford: Clarendon Press.(With Prior (1967), the main exposition of tense logic.)

Prior, A.N. (1971) *Objects of Thought*, ed. P.T. Geach and A.J.P. Kenny, Oxford: Clarendon Press.(Prior's account of the nature of propositions and intentional objects.)

Prior, A.N. (1976a) *Papers in Logic and Ethics*, ed. P.T. Geach and A.J.P. Kenny, London: Duckworth.(A collection of Prior's most important papers not already contained in previous volumes. Includes 'The Autonomy of Ethics', 1960.)

Prior, A.N. (1976b) *The Logic of Propositions and Terms*, ed. P.T. Geach and A.J.P. Kenny, London: Duckworth. (Part of the original draft for the introduction to logic published earlier in part as *Formal Logic*.)

Prior, A.N. and Fine, K. (1977) *Worlds, Times and Selves*, London: Duckworth.(Further moves in the business of relativizing propositions to possible worlds, times and subjects of consciousness.)

References and further reading

Copeland, J. (ed.) (1996) *Logic and Reality: Essays in Applied Logic in Memory of Arthur Prior*, Oxford: Oxford

University Press.(Proceedings of a conference held in New Zealand in 1992 to commemorate Prior's work.)
Grover, D. (1992) *A Prosentential Theory of Truth*, Princeton, NJ: Princeton University Press.(A development of Prior's ideas on Truth.)
Kenny, A.J.P. (1970) 'Arthur Norman Prior', *Proceedings of the British Academy* 56: 320-49.(Discussion of Prior's life and work.)
Williams, C.J.F. (1973) 'Prior and Ontology', *Ratio* 15: 291-302.(A critical notice of Objects of Thought.)

Privacy

The distinction between private and public is both central to much legal and political thought and subject to serious challenge on philosophical, practical and political grounds by critics of the status quo. Privacy - the state of being withdrawn from the world, free from public attention, interference or intrusion - is a cherished social value that is being offered ever more protection. Increasingly, laws require people to respect the privacy of others: privacy is recognized as a fundamental right in international documents and national constitutions, and recent customs and social norms forbid intrusions that were once accepted. The concept of privacy is also widely abused: it has been used to justify private racial discrimination and state neglect of domestic violence, as well as social abdication of general economic welfare through laissez-faire *policies and the so-called privatization of social services. Critique of the public-private distinction is an important part of many critical theories, especially feminism and critical legal theory. These critics object that the public-private distinction is exaggerated, manipulable or incoherent.*

1 Privacy as a social value

The origin of privacy is controversial. Some maintain that the public-private distinction originated with the Greeks and that '[s]uch delineations could not have been made in the theocracies of the ancient Near East because in such cultures god-as-ruler permeates everything and no notion of the private is possible' (Wiltshire 1989: 8-9). Others argue that up to the middle of the sixteenth century 'privacy…was neither possible nor desired' (Stone 1977: 6), and that it first arose a few hundred years ago, as household size reduced and as houses began to be built with corridors and private spaces. Although it is unlikely that either of these assertions is fully correct, it also seems clear that different periods in history are likely to have had somewhat different concepts of privacy, just as different cultures today have different concepts.

Privacy is central to liberal thought - as a right the state guarantees to protect from interference by others or by the state itself. In Western democratic societies, privacy is generally seen as a state of being or a right enjoyed by an individual. Privacy is considered basic to a free and open society and crucial for individual development. It facilitates spontaneity and insulates the individual from social pressure to conform. It contributes to autonomy, creativity, the capacity to form human relationships and the development of personal responsibility. For this reason, liberal political philosophies - J.S. Mill's *On Liberty* (1859) is a prime example - nearly always seek to draw a line between that part of human behaviour which is private, and therefore not subject to legal or social control, and that part which is of public concern (see Law, limits of §4; Liberalism §§1-2).

Privacy is also extolled as essential to maintaining human relationships. Various collectivities, especially families, are sometimes said to require privacy or to be entitled to privacy. Unmarried heterosexual couples are often seen as deserving of a degree of privacy in order to be free to form a relationship. Lesbians, gay men and other sexual minorities have sought recognition of their relationships as alternative families entitled to privacy; sometimes this group-based approach has protected intimate relations more successfully than appeals to a notion of individual privacy.

2 Public protection of privacy

The concept of privacy as a legal right is of recent origin. In the USA it was introduced into the literature in 1890 by an influential law review article (Warren and Brandeis 1890). It received recognition as a constitutional right in the USA in 1965 in *Griswold* v. *Connecticut*, which established a privacy right for married couples to use birth control. This privacy right was found to attach to single people also and to protect an individual's access to abortion as well as to birth control (*Eisenstadt* v. *Baird* 1971; *Roe* v. *Wade* 1973). The right to privacy has also been recognized in major human rights documents such as the Universal Declaration of Human Rights.

Several competing state interests limit the protection given to privacy. Searches, seizures, involuntary blood and drug tests, and numerous other intrusions are allowed in the name of national security or crime detection. A major limitation upon privacy comes from competing rights, such as the freedom of speech and of the press (see Freedom of speech). In the USA, freedom of speech limits the protection given to 'public figures' against intrusions into privacy and limits defamation actions to cases in which the defamer acted 'maliciously' (*New York Times* v.

Sullivan 1964). Freedom of the press has been found to override the interests of a rape victim to remain anonymous, even when the victim reasonably fears that the publication of their name would cause them to be targeted for further abuse and violence (*Florida Star* v. *BJF, US* 1989).

Computer technology has heightened concern regarding the protection of privacy by radically increasing the possibilities for gaining information about others. Large quantities of information about a person can now be collected, stored, processed and retrieved at relatively low cost. Data protection laws are being developed and refined in many countries to deal with these issues. Computer bulletin boards raise privacy issues that are just beginning to emerge. The widespread use of computers by children has facilitated access to them by adult child molesters who may pose as children and eventually set up a rendezvous.

3 Abuses of the public-private distinction

The public-private distinction has been improperly used to justify social policies that maintain power relations and leave defenceless people exposed to conditions that, if recognized as a public concern, should seem unacceptable. For example, exploitation and abuses of employees long went unchecked and legislative efforts to protect workers were overturned by courts on the basis of *laissez-faire* policies grounded in large part on the notion that private property should not be subject to public control. Even today labour unions in many countries are restricted on the basis of a public-private distinction: many aspects of the work environment and decisions related to these concerns are removed from collective bargaining because such issues are said to involve private management discretion that should not be meddled with by the public.

Racism and discriminatory decisions based on prejudice were until recently defended as a private prerogative, not to be interfered with by the state. Even as US courts began to protect the rights of African-Americans, the public-private distinction was used to limit review of racial discrimination to instances in which ‘state action’ could be found. ‘Private’ discrimination was not forbidden by the Constitution. Notions of privacy continue to limit the scope legislators give to employment discrimination laws, often exempting employers in small, more private-seeming establishments. The British Race Relations Act (1965), which forbids discrimination by anyone serving ‘the public’, was found not to protect against discrimination by a so-called private club that chose members ‘by nomination and personal selection’ (*Race Relations Board* v. *Charter and Others*, UK 1973).

The supposed privacy of the family has shielded brutality and abuse from public scrutiny and redress. In many areas of the world a special exemption to the rape laws withdraws protection from women who are married to the man who forces sexual intercourse upon them, no matter how brutally he does so. Many criminal and civil laws against physical assault have similar exemptions when the victim is the wife or the child of the perpetrator. Where the laws do forbid violence within the family, notions of family privacy nevertheless often limit the effectiveness of the enforcement of these laws.

Of course, there are those who would defend these uses of the public-private distinction as properly protective of private property, personal choice and family privacy. One reason that privacy needs to be protected, they could argue, is that individual cases arise in which the urge to intervene will be great, yet overall in the long run, non-intervention is to be preferred.

A second kind of abuse of the public-private distinction occurs when there is slippage between the descriptive claim that something is either public or private and the normative claim that it should be treated as public or private. This slippage leads to circular arguments and to efforts to foreclose discussion of the advantages and disadvantages of the policy advocated.

4 Critiques of the public-private distinction

What some critics see as an abuse of the public-private system distinction, others see as a problem inherent in the distinction itself. Since everything is socially conditioned, there is no sphere that is wholly private - there is a public dimension to private affairs and vice versa. Moreover, the line that is drawn between public and private does not reflect a natural distinction but an ideological purpose. It mystifies inequalities in society by treating them as irrelevant to political equality. Marx, for example, saw the division between the private individual and the public citizen as part of the alienation of mankind under capitalism. The state was not the public bulwark against private oppression that liberal theorists supposed, but an agent of the ruling class.

Similar criticisms of *laissez-faire* capitalism were raised in the 1920s and 1930s by radical legal realists and others. The sphere of apparently free or private economic activity was dependent on legal rules created by the state, which thus structured and controlled power relations in society. Businessmen who called for an end to state regulation of the economy were inconsistent, because their own economic activity depended upon pro-business regulation.

The feminist critiques of the public-private distinction go further. Feminists point out that the public-private distinction is more than a distinction between the state and civil society, between the general interest and the particular interest, or between a person's public life as a citizen and private life as a self-regarding individual. It is also a distinction between men and women, between the public world of business and industry and the private, domestic world of the family. This gender association of private with women and public with men has important consequences for the role and status of women as well as for the particular meanings given to privacy (see Feminist political philosophy §§1, 7).

The legal realist critique of *laissez-faire* applies with equal force to arguments against state intervention in the family. The state structures domestic life through its family law provisions, as well as less noticeably through its policies regarding rape, employment discrimination, taxes, welfare benefits and childcare policies. Moreover, the public world of affairs - commerce, industry and so forth - is not separate from and opposed to the domestic world of the family; rather the two are inextricably interrelated (Olsen 1983). Carole Pateman has argued forcefully that the 'apparently impersonal, universal dichotomy between private and public' obscures the 'fact that patriarchalism is an essential, indeed constitutive, part of the theory and practice of liberalism' (Pateman 1983: 286)

Although many or most feminists might be said to accept privacy as such in that they, like others, appreciate solitude and would not want every aspect of their lives exposed to general public view, feminist criticism of the public-private distinction is more than just an objection to where the distinction is traditionally drawn and how it is used. Feminists challenge the way the line is drawn, the manner in which each term - public and private - is defined, and the basic dichotomous nature of the public-private distinction.

One way of understanding the gendered nature of privacy in Western thought is to consider that women are more closely associated with privacy but men are more assertive to claim a private realm for themselves. Women *are* privacy; they do not *have* privacy. Men as a group consider their interactions with women private, but as a group they also freely intrude upon the solitude of women, whether through street harassment or more gentle forms of imposition. This is a specific case of the general relationship between privacy and hierarchy. The hierarchically superior has privacy; the inferior does not. In order to have privacy, the inferior would have to exclude the superior in a way that is unnecessary for the superior.

FRANCES OLSEN

References and further reading

Benn, S.I. and Gaus, G.F. (eds) (1983) *Public and Private in Social Life*, London: Croom Helm. (Collection of essays examining the public-private distinction in several contexts.)

Dallmeyer, D.G. (ed) (1993) 'The Public/Private Distinction and its Impact on Women', in part 2, *Reconceiving Reality: Women and International Law*, Studies in Transnational Legal Policy no. 25, American Society of International Law, 93-169.(Discussion of feminist critiques of the public-private distinction in the context of international law.)

Lacey, N. (1993) 'Theory into Practice? Pornography and the Public/Private Dichotomy', in J. Conaghan and A. Bottomley (eds) *Feminist Theory and Legal Strategy*, Oxford: Blackwell; and special issue of *Journal of Law and Society*30 (93).(Contextualized feminist critique of the public-private distinction.)

Mill, J.S. (1859) 'On the Limits to the Authority of Society over the Individual', ch. 4, *On Liberty*. (Classic liberal presentation of public-private distinction.)

Olsen, F. (1983) 'The Family and the Market: A Study of Ideology and Legal Reform', *Harvard Law Review* 96: 1497-578. (Elaborate critique of the public-private dichotomy.)

Olsen, F. (1993) 'Constitutional Law: Feminist Critiques of the Public/Private Dichotomy', *Constitutional Commentary* 10: 319-27.(Discussion of feminist critiques of the public-private distinction in the context of constitutional law.)

Pateman, C. (1983) 'Feminist Critiques of the Public/Private Dichotomy', in S.I. Benn. and G.F. Gaus (eds) *Public and Private in Social Life*, London: Croom Helm.(Demonstrating the centrality of patriarchalism to liberalism in the context of feminist critique of the public-private dichotomy.)

Pateman, C. (1988) *The Sexual Contract*, Stanford, CA: Stanford University Press.(Rich expansion and development of the argument expounded in Pateman 1983.)
Pennock, J.R. and Chapman, J.W. (eds) (1971) *Nomos XIII: Privacy* New York: New York University Press. (Classic, if dated, collection of essays on privacy.)
Schoeman, F. (ed) (1984) *Philosophical Dimensions of Privacy: An Anthology*, Cambridge: Cambridge University Press.(Collection of essays on privacy from a philosophical perspective.)
Stone, L. (1977) *The Family, Sex and Marriage in England 1500-1800*.(Controversial history of family in England.)
Symposium (1982) 'The Public/Private Distinction', *University of Pennsylvania Law Review* 130: 1289-609. (Critiques and defences of the public-private distinction in the context of law.)
Wacks, R. (1993) *Privacy*, The International Library of Essays in Law and Legal Theory, Aldershot: Dartmouth Press, 2 vols.(Collection of essays on privacy from the legal perspective.)
Warren, S. and Brandeis, L. (1890) 'The Right to Privacy', *Harvard Law Review* 4: 193.(Classic argument in favour of recognition of the common-law right to privacy.)
Wiltshire, S.F. (1989) *Public and Private in Virgil's Aeneid*, Amherst, MA: University of Massachusetts Press. (Study of the classical origins of the public-private dichotomy.)

Private language argument

Ludwig Wittgenstein argued against the possibility of a private language in his 1953 book Philosophical Investigations, where the notion is outlined at §243: 'The words of this language are to refer to what can be known only to the speaker; to his immediate, private, sensations. So another cannot understand the language.' The idea attacked is thus of a language in principle incomprehensible to more than one person because the things which define its vocabulary are necessarily inaccessible to others; cases such as personal codes where the lack of common understanding could be remedied are hence irrelevant.

Wittgenstein's attack, now known as the private language argument (although just one of many considerations he deploys on the topic), is important because the possibility of a private language is arguably an unformulated presupposition of standard theory of knowledge, metaphysics and philosophy of mind from Descartes to much of the cognitive science of the late twentieth century.

The essence of the argument is simple. It is that a language in principle unintelligible to anyone but its user would necessarily be unintelligible to the user also, because no meanings could be established for its signs. But, because of the difficulty of Wittgenstein's text and the tendency of philosophers to read into it their own concerns and assumptions, there has been extensive and fundamental disagreement over the details, significance and even intended conclusion of the argument. Some, thinking it obvious that sensations are private, have supposed that the argument is meant to show that we cannot talk about them; some that it commits Wittgenstein to behaviourism; some that the argument, self-defeatingly, condemns public discourse as well; some that its conclusion is that language is necessarily social in a strong sense, that is, not merely potentially but actually. Much of the secondary (especially the older) literature is devoted to disputes over these matters.

An account of the argument by the influential American philosopher Saul Kripke has spurred a semi-autonomous discussion of it. But Kripke's version involves significant departures from the original and relies on unargued assumptions of a kind Wittgenstein rejected in his own treatment of the topic.

1 The significance of the argument

The idea that a language might be private is explicitly canvassed in the second of Bertrand Russell's lectures 'The Philosophy of Logical Atomism', and Wittgenstein's argument, though possibly directed against his own earlier views in *Philosophical Remarks*, may originate in the sustained criticism of Russell which informs much of his writing.

Immediately prior to the *Philosophical Investigations*' discussion of private language, Wittgenstein suggests that the existence of the rules governing the use of words and making communication possible depends on agreement in human behaviour. This agreement includes people's reacting in similar ways to similar training, and exemplifies contingent 'very general facts of nature' which make particular concepts and customs possible and useful. Thus, for instance, one can train most children to look at something by pointing at it (whereas dogs look only at one's hand), and this enables us to attach meaning to the gesture of pointing and derivatives like signposts. The immediate function of the private language argument is to show that the possibility of linguistic rules and concept formation in general depends upon the possibility of such agreement.

But it has a further, connected, function. Motivating Wittgenstein's discussions of both mathematics and psychology is hostility to metaphysical absolutes, to the idea that we can find the world as it really is in the sense that any other way of conceiving it must be wrong (compare Wittgenstein 1953: 230). Both numbers and sensations provide especially tempting cases for philosophers, who are inclined to imagine that these are objects which force their identities upon us, our classifications of them and the rules governing the uses of their names being dictated to us in advance by the phenomena themselves. Wittgenstein makes precisely analogous points in his treatment of the two cases. In both, the underlying confusion is about how the act of meaning determines the future application of a formula or name.

With numbers, one temptation is to confuse the mathematical sense of 'determine' in which, say, the formula $y = 2x$ determines the numerical value of y for a given value of x (in contrast with y (2x, which does not) with a causal sense in which a certain training in mathematics determines that normal people will always write the same value

for *y* given both the first formula and a value for *x* (in contrast with creatures for whom such training might produce a variety of outcomes). This confusion produces the illusion that the outcome of an actual properly conducted calculation is the inevitable result of the mathematical determining, as though the formula's meaning itself were shaping the course of events.

In the case of sensations, the parallel temptation is to suppose that their natures are self-intimating - I seem to feel what they are directly, so that I need only give a name to one of them and the rules for the subsequent use of that name will be fixed on the spot. Wittgenstein tries to show that this is an illusion, that even that apparently most self-intimating of all sensations, pain, derives its identity only from a sharable practice of expression, reaction and use of language. Now if, say, pain were to force its identity upon me in the way described, then the possibility of such a shared practice, which in turn depends upon those general background facts of nature, would be irrelevant to the concept of pain. That is, if the real nature of part of the world were revealed to me in a single mental act of naming, as the private linguist supposes, then all subsequent facts would necessarily be irrelevant and the name could be private. The private language argument concludes that they could not be irrelevant, that no names could be private, and that the notion of having the real nature of the world so revealed is confused.

Such confusions underlie a range of articulated philosophical ideas and theories, without themselves being so articulated. The argument thus attempts, not to refute any particular theory, but to remove the motivation for involvement in a range of seemingly independent tasks, problems and solutions. Here are some frequently cited examples.

A still very common idea, found in Locke, is that interpersonal spoken communication works by speakers' translation of their internal mental vocabularies into sounds followed by hearers' re-translation into their own internal vocabularies (see Locke, J. §5). Again, Descartes considered himself able to talk to himself about his experiences while claiming to be justified in saying that he does not know (or not until he has produced a reassuring philosophical argument) anything at all about an external world conceived as something independent of them. And he and others have thought: while I may be wrong in my judgments about the external world, I am infallibly correct if they are restricted to my sensations. Again, many philosophers have supposed there to be a problem of other minds, according to which I may reasonably doubt the legitimacy of applying, say, sensation-words to beings other than myself (see Other minds). In all such cases, the implication is that my language could in principle be private: for these problems and theories even to make sense, the ability to share must be irrelevant to meaning and it must be conceivable that I am confined to my own case. (This suggestion is controversial: the usual charge is that such philosophers are committed to supposing that there actually is a private language. But this is not obvious, and the argument would be just as significant if all they are committed to is the supposition that their presupposed internal language, for all the difference it would make to its usability for self-communing, could be private.) This is especially clear in the case of Descartes, who must hold it possible to identify one's experiences inwardly, that is, without using any resources supplied by one's embodiment in an independently existing world, such as the concepts acquired in a normal upbringing. How is this identification to be achieved? This is the question considered in the next section.

2 The nature of the argument

The private language argument is usually identified with §§256-71 of *Philosophical Investigations*, a tightly-knit discussion of sensations, with §258 being especially important. But this discussion cannot properly be detached from the earlier sections of the book, despite commentators having frequently done just this. As important as the already-noted connection with the lengthy prior discussion of rule-following is the treatment of ostensive definition.

The argument proper is preceded by a preliminary discussion (§§244-55) in which Wittgenstein distinguishes two senses of 'private' and argues that natural languages (English, for example) are private in neither. The question then arises, could anything be a private language? He approaches this question by considering how a private language might be arrived at. It cannot be reached via the actual language we speak, for that is not private, and the attempt to convert it by thought experiment into a private one by simply suspending all expression of sensation and imagining the speaker just to name a sensation, as it were in a vacuum, merely raises the question of what this is supposed to consist in and what it is for. But to give private language enthusiasts a run for their money, Wittgenstein imagines himself in the position of establishing a private language for the purpose of keeping a

record of his sensations.

But he faces a serious impediment to discussing the matter at all: that doing so requires the mention of actions like ostensive definition, concentrating the attention, speaking, writing, remembering, believing, and so on, in the very process of suggesting that none of these can really be done in the circumstances under consideration. (This must be remembered in reading what follows, which in strictness should be constantly disfigured with scare quotes.)

He considers the idea that I simply *associate* a sign, say ‘S’, with a sensation by concentrating my attention on the sensation and saying ‘S’ to myself (the private analogue of ostensive definition), and points out that if this is to be a genuine definition it must establish a persisting connection between sign ‘S’ and that sensation: ‘"I impress [the connection] on myself" can only mean: this process brings it about that I remember the connection right in the future (§258)’.

This single remark has caused much trouble. Many have thought ‘I remember the connection *right*’ means ‘I use ’S’ only when I have S.’ This had the argument resting on scepticism concerning memory, and provoked the criticism that memory’s fallibility is neither more nor less a problem for a private linguist than a public one, so that the argument threatens both or neither. Wittgenstein’s defenders retorted that fallibility is no problem only where mistakes can be corrected; in the private case, there can be no checking of memories, hence no chance of correction, and consequently talk of correctness is inappropriate. The critics responded by arguing either that such checking is possible, or that correctness does not require checkability.

This interplay of criticism and defence characterizes much of the commentary on the argument. But it is beside the point. Both sides mistakenly assume the connection in question between ‘S’ and the sensation to be a connection of *truth*, so that remembering the connection right is a matter of making the judgment that I am experiencing S only in the presence of S. But the connection Wittgenstein says must be remembered is a connection of *meaning*. It can now be seen that the argument makes no appeal to the fallibility of memory.

Imagine I am a private linguist. I have a sensation, and make the mark ‘S’ at the same time, as one might in an ordinary case introduce a sign by ostensive definition. Now suppose that later, I use ‘S’ in judging that I am again experiencing the same sensation. What do I mean by ‘S’ on this second occasion?

It cannot be that I mean the sensation I am now experiencing, for this collapses the distinction between meaning and truth and removes the status of factual assertion from my judgment that I am experiencing S - it becomes at best a fresh ostensive definition. (Thus I cannot say informatively that my shirt is orange and with the same utterance explain what ‘orange’ means by exhibiting my shirt.)

Can it be rather that I mean the sensation I named ‘S’ on the previous occasion? That would presuppose that I had indeed succeeded in giving my sensation the name ‘S’, but this cannot just be assumed: for there to have been an ostensive definition of ‘S’, a technique for the use of ‘S’ must have been established, one which leads to my using ‘S’ in the same way as before. This, though, is just what is in question. What would using the sign in the same way be here? The same way as what? Wittgenstein’s earlier discussion of ostensive definition showed that there can be such definitions only where a place for them is already prepared - just pointing and making noises does not establish what sort of thing is being pointed at, even in the public world, and as a private linguist I cannot even do that but at best can only concentrate my attention. ‘S’ must be the name of a *sensation*, and the fact that I have had a sensation and simultaneously inwardly muttered ‘S’ does not suffice to make ‘S’ the name of that sensation. (To name a sensation is to name a *kind*, but what kind? A sensation may be classified in indefinitely many ways: ache, sensation-in-leg, sensation-at-time-*t*,….) And if one thinks a private linguist could remember the meaning of ‘S’ by remembering rightly the past correlation of ‘S’ with a certain kind of sensation, one presupposes what needs establishing: that there was such an independent correlation to be remembered. Fallibility of memory, even of memory of meaning, is neither here nor there: the point is that there has to be the right sort of occurrence in the first place to be a candidate for being remembered; and if there is not, no memory is going to create it. If, alternatively, we do not suppose that there is something to be remembered which is independent of the memory, then ‘what seems right to me is right’, that is, there is nothing to be right or wrong about. (These points are made in Wittgenstein 1953: 258-68.)

The conclusion is that it is impossible for a private linguist to establish and maintain a rule for the use of an expression, so that meaning is unobtainable in a private language. Only operating in a world independent of one’s

impressions of it, in which one's operations are thus in principle available for scrutiny, can provide the possibility of real correlations of signs with objects and consistency in the usage of those signs.

3 Related considerations

The argument as given is embedded in a collection of arguments, observations and reminders on the connected topics of the relations between expression and description of experiences, between mind and behaviour, and between self-knowledge and knowledge of others, concentrated mainly in §§243-315 of *Philosophical Investigations* but spreading into other sections and other works. This whole collection is sometimes referred to as the private language argument. But the question of private language is present in several of Wittgenstein's even apparently unrelated discussions, such as those of seeing-as (seeing a triangle as resting on its base, for example), and kinaesthesis, where the only available descriptions of the experiences involved which do justice to their contents make direct reference to a sharable world of public objects.

Central to his thought here is the belief that no one argument will suffice to eliminate the commitment to private language, for it is the result of various powerful illusions. One of these is the idea that one can study the nature of mental phenomena by introspection, as though this gave us direct and unmediated access to the truth-condition for, say, 'My neck is itching', a condition to which only I am privy (see Introspection, psychology of). In contrast, Wittgenstein argues, the nature of mental phenomena is grasped not by introspection but by examining 'language and the actions into which it is woven' (what he calls the 'language-game'), a publicly available practice of using words: with pain, this involves crying, complaining, comforting, administering analgesics and so on (*Philosophical Investigations*, §7).

Another source of the idea of a private language is that, where thoughts and sensations are concerned, there is in the normal first-person case no gap between truthfulness and truth. Philosophers are inclined to interpret this as a special epistemic authority about one's own mind, whose contents are hidden from others. Wittgenstein's treatment of this has been one of the most disputed parts of his discussion; he has been interpreted as suggesting that the lack of first-person uncertainty in utterances like 'My leg hurts' derives from their being not genuine reports of inner occurrences but sophisticated groaning. It is now a short (and often-taken) step to infer that this involves a commitment to supposing that there is nothing behind the outer behaviour, and thus to a form of behaviourism (see Behaviourism, analytic). Such oscillation between privacy and behaviourism, as though the former were the sole alternative to the implausibility of the latter, is typical of the bind philosophers get into in this area, so that they cannot see that the blocking of this route to private language involves the attempt to undermine assumptions common to both. These include such ideas as that witnessing a person's acts is being confronted with neutral behavioural data from which mentality must be inferred or constructed, and that dealing with other human beings is always a matter of hypothesis rather than of attitude and response. But Wittgenstein's view is that these ideas misconstrue the metaphorical inner/outer contrast; his rejection of this construal is not a denial of the reality of the inner.

4 Kripke's Wittgenstein

Saul Kripke's account of Wittgenstein's treatment of rules and private language has generated much second-order discussion of a kind notably different from that of Wittgenstein himself. This is often carried on without apparent concern for the question of whether Wittgenstein's own arguments are captured by Kripke's interpretation. Thus the phrase 'private language argument' has acquired a further sense in which it refers to the arguments as given by Kripke.

Kripke's interpretation resembles that given here in its emphasis on the dependence of the discussion of private language on that of rule-following. But Kripke (1982: 68) says, 'The impossibility of private language emerges as a corollary of his sceptical solution of his own paradox', one stated at *Philosophical Investigations* §201: 'No course of action could be determined by a rule, because every course of action can be made out to accord with the rule.' Apparently overlooking Wittgenstein's reference to this paradox as involving a misunderstanding, Kripke takes it to be a profound sceptical problem about meaning. He formulates this problem in two ways.

The first formulation is this: there is no fact in which someone's meaning something consists. The absence of this fact, in Kripke's view, leads Wittgenstein to abandon the explanation of the meanings of statements like 'Smith

meant addition by ‘plus’’ in terms of truth-conditions; instead, they get explained in terms of assertibility-conditions, which involve actual (not merely potential) community agreement. (Hence the claim that this is a ‘sceptical solution’: Wittgenstein is supposed to concede to the sceptic the absence of truth-conditions for such statements.) This agreement, on Kripke’s account, legitimizes the assertion that Smith meant addition by ‘plus’ despite there having been no fact of the matter, and it of course rules out the possibility of private language immediately (see Private states and language §4). This first formulation, though, relies on the assumption that we have some idea of what a fact is, independent of a statement’s being true; and one of the fundamental lessons of *Philosophical Investigations* is that there is no such idea to be had, that the only route to the identification of facts is via the (often not easily discerned) uses of the expressions in which those facts are stated. These uses give us the truth-conditions.

The other formulation of the problem is that there is no nexus between someone’s meaning something and their subsequent behaviour, so that, for example, my grasp of the rule governing the use of ‘plus’ does not determine that I shall produce a unique answer for each of indefinitely many new additions in the future. The impression that something is missing here, though, is a result of just that kind of confusion about determination identified in §1 above.

Kripke’s account is of great intrinsic interest, but despite the acuteness of some of its observations, as an interpretation of the private language argument it is thus deficient, especially because of his unargued reliance on ideas which Wittgenstein himself argued against. This has not deterred philosophers from using it, as Norman Malcolm’s earlier account was used, as a way of avoiding direct confrontation with Wittgenstein’s own text.

See also: Consciousness; Criteria; Kripke, S.; Wittgenstein, L.

STEWART CANDLISH

References and further reading

Because of the great difficulty of the subject matter, none of this material is straightforward, but none is technical.

Boghossian, P.A. (1989) ‘The rule-following considerations’, *Mind* 98 (392): 507-49.(A critical survey of the literature concerning the Kripkean version of Wittgenstein’s arguments.)

Candlish, S. (1980) ‘The real private language argument’, *Philosophy* 55 (211): 85-94.(A fuller version of some of the points made in §2 of this entry.)

Canfield, J.V. (ed.) (1986) *The Philosophy of Wittgenstein*, vol. 9, *The Private Language Argument*, New York: Garland.(An extensive collection of papers exhibiting the variety of interpretations of the argument.)

Canfield, J.V. (1986) *The Philosophy of Wittgenstein*, vol. 10, *Logical Necessity and Rules*, New York: Garland. (Contains some papers on the Kripke version.)

Canfield, J.V. (1991) ‘Private language, *Philosophical Investigations* §258 and environs’, in R. Arrington and H.-J. Glock (eds) *Wittgenstein’s Philosophical Investigations: Text and Context*, London and New York: Routledge, 120-37.(A discussion of some common assumptions about how the argument works.)

Canfield, J.V. (1996) ’The Community View’, *The Philosophical Review* 105 (4): 469-88.(A treatment of the issue of whether the private language argument requires language to be necessarily and actually, or merely potentially, social; examines both original texts and principal interpretations.)

Hacker, P.M.S. (1990) *Wittgenstein: Meaning and Mind*, *An Analytical Commentary on the Philosophical Investigations*, vol. 3, Oxford: Blackwell, 1-286. (Part of a massively detailed study of *Philosophical Investigations*; a good place to find out about textual sources, connections, ramifications, significance.)

Jones, O.R. (ed.) (1971) *The Private Language Argument*, London: Macmillan.(A collection of some of the earlier influential discussions of the argument. Also includes Wittgenstein 1968. Not much overlap with Canfield 1986.)

Kenny, A. (1973) *Wittgenstein*, London: Allen Lane: The Penguin Press, ch. 10.(One of the clearest accounts of the argument. Its version inspires, but differs from, that given in §2 above.)

Kripke, S. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell.(The final published version of Kripke’s account.)

Malcolm, N. (1954) ’Wittgenstein’s Philosophical Investigations’, *The Philosophical Review* LXIII (4): 530-59. (Presents a highly influential early account of the private language argument.)

Winch, P. (1983) ‘Facts and superfacts’, *The Philosophical Quarterly* 33 (133): 398-404, repr. (slightly altered) in P. Winch *Trying to Make Sense*, Oxford: Blackwell, 54-63.(A clear and concise criticism of Kripke, from which some of the points in §4 above have been condensed.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell, 3rd edn 1967.(The seminal text. Spare and difficult without being obscure or technical. The central material is found in §§243-315, but the earlier discussions of ostension in §§1-36 and rule-following, §§142-242, are also drawn upon in the argument. See also §§316-97, and pages 220-4.)

Wittgenstein, L. (1968) ‘Notes for lectures on "private experience" and "sense-data"’, *The Philosophical Review* LXXVII: 275-320.(Also included in Jones 1971.)

Wittgenstein, L. (1975) *Philosophical Remarks*, Oxford: Blackwell.(A possible target of the private language argument.)

Private states and language

Something is 'private' if it can be known to one person only. Many have held that perceptions and bodily sensations are in this sense private, being knowable only by the person who experiences them. (You may know, it is often said, that we both call the same things 'green'; but whether they really look the same to me as they do to you, you have no means of telling.) Regarding the relation between private states and language two main questions have arisen:

(1) Could there be a 'private language', that is, a language in which a person communicates to themselves, or records for their own use, information about their own private states - this language being in principle incomprehensible to others, who do not know the nature of the events it is used to record. This question is primarily associated with Ludwig Wittgenstein.
(2) Can the nature of our private states affect the meaning of expressions in the public language, that is, the language we use for communicating with each other? Or must everything that affects the meaning of expressions in the public language be something which is itself public, and knowable in principle by anyone? Michael Dummett has argued that we must accept the second of these alternatives, and that this has far-reaching consequences in logic and metaphysics.

1 Epistemic privacy

The epistemically private is defined as whatever can in principle be known to one person only. The words 'in principle' are intended to exclude all cases in which something is in fact known to only one person, though it could be known to more. The only items that have generally been taken to be epistemically private in this very demanding sense are conscious mental states, primarily perceptual states and bodily sensations of a conscious subject, on the plausible grounds that nothing that a person can do can possibly give them a view of another's conscious states, but only of their bodily states and behaviour.

However, these grounds for taking such states to be epistemically private cannot by themselves decide the question whether there actually are any epistemically private items in the sense defined. Privacy, as we have seen, is defined in terms of what cannot be known to another, and from the fact that another cannot *experience* my headache it does not immediately follow that they cannot *know* what it is like. That must also depend on our understanding of knowledge, in particular on whether we can allow that what cannot be directly experienced by someone can nevertheless in some other way be known by them.

Discussion of this complex question is beyond the scope of this entry, in which it can only be remarked that some views of the nature of knowledge are not wholly unfavourable to the idea that the contents of another's consciousness may sometimes be known (see Reliabilism; Other minds). Here it will be assumed that there are epistemically private states of the type commonly suggested.

We have spoken above of *two* questions concerning private states and language, and it is important to realize that, although related, they must be considered separately if clarity is to be achieved. To see this, the reader should consider the following. An impressive argument (discussed in §§2-3) for denying that private states can play any part in the semantics of the language that we speak with each other begins from the premise that the nature of one's private states is unknown to anyone else, and concludes that if it in any way affected the meanings of our words we would not know what any other person meant, and so could not speak this language with mutual understanding - which is then taken to be absurd. As it stands this argument clearly has no application whatever to the first question, the possibility of a private language. For a private language is understood to be one in which no other speaker is involved, so that the fact that one's private states are unknowable to others is entirely beside the point. To reach the analogous conclusion it would have to be said that private states are unknowable not just to others but to their owners as well, and this much stronger claim clearly requires further and altogether different argument. Conversely, we shall also see (§4) that it is possible to argue against the feasibility of a private language using a line of thought that may not apply to the question about the public language.

2 Acquisition and manifestation

Two arguments now frequently offered for the conclusion that private states can play no role in determining the meaning of any expressions of our (public) language are known as the acquisition and manifestation arguments. Each assumes that, in so far as there exists a language in which we communicate with each other, the expressions of that language must be used in the same sense or meaning by all of us. The acquisition argument then asks how this state of affairs could come about; the manifestation argument asks how we could tell that it obtained.

Suppose that there are expressions of this public language whose meaning depends on the nature of certain of the private states of competent speakers. (Plausible examples are colour words - it might well be thought that what I mean by ’red’ depends at least in part on how certain things look to me - and words for sensations such as ’pain’.) Now consider the situation of language learners acquiring an understanding of one such expression. *Ex hypothesi* this means that they must come to use it in connection with private states of their own of the same kind as those already associated with it by competent speakers; but since these states, being private, are not accessible to the learners, how can they hit on the right meaning? Learning the language would call for an absurdly improbable series of lucky guesses.

The manifestation argument views the process from the other side: how are the teachers to tell that the learners have learnt their lesson correctly? How can the learners show that they have grasped (’manifest their grasp of’) the right meaning? For to grasp the right meaning is to connect with the expression particular private states rather than others; and because they are private that is something which the teachers are in principle unable to check. We could never know, in the case of these expressions, what any other speaker means by them.

We should note that neither of these arguments makes appeal to verificationism (see Meaning and verification) - both trade on the idea of what can be *known*, rather than of what is *meaningful*. The importance of this is not just that they avoid inheriting verificationism’s weaknesses; it is also that otherwise they could not, without deep suspicion of circularity, be used to lead to a conclusion not altogether unlike verificationism - as they in fact have (see §5).

3 Responses to these arguments

One response, hinted at in §1, would be to deny that we have any states which are in principle unknowable to others. In view of the history of such sceptical debates this seems unlikely to lead to anything decisive.

A possible reply to the acquisition argument is as follows. Perhaps human beings naturally assume that others, when in broadly similar circumstances, experience inner states similar to those which they themselves experience. If that is so, and if the assumption is in most cases correct, then the problem posed by the acquisition argument can be answered. For since on this hypothesis the learner’s private states resemble those of the teacher (when, for example, both are in the presence of the kind of object we call ’red’) and are believed by both parties to do so, there is no difficulty in seeing how the learner can come to associate with the word ’red’ private visual states which are of the same kind as those which the teacher associates with it.

But this does not settle the issue. For one thing, it makes no reply to the manifestation argument. For even if successful learning has occurred, how is anyone to tell? The teacher and the erstwhile learner have no way of knowing that they now mean the same. And if, as is often said, understanding someone is knowing what they mean, then they don’t understand each other - contrary to the hypothesis that they are speaking a mutually intelligible language.

It is not obvious, however, that understanding someone does require knowledge of their meaning, if knowledge is taken to be more than confident true belief, as it nearly always is (see Knowledge, concept of §2). For if your words express a certain thought, and I rightly and confidently believe that that is what they express, and so respond appropriately, it seems that everything we want of understanding has been achieved, and the additional demand for knowledge begins to look arbitrary.

Still the matter is not settled, for it may be replied that even if knowledge is not a necessary condition of understanding, nonetheless we *do* know what others mean by their words; therefore any theory which allows private states an essential role in semantics must be wrong, since it is committed to denying that we have such knowledge.

That commitment cannot be disowned, since it follows immediately from the definition of privacy, but it may be

less damaging than it at first looks. Perhaps the claim to knowledge in this case, in so far as it implies the capacity to give good reasons for what we claim to know, is indeed dubious. Such a view is controversial, but certainly not obviously untenable. The idea that some of our most basic beliefs are produced by a psychological mechanism which is non-rational, in the sense that its operations are quite independent of any capacity we may have to *give reasons* for the beliefs it produces, has a reputable history (see for instance Hume, D.). To enable it to act successfully any animal, including ourselves and our ancestors, needs a reliable and rapid way of forming beliefs about its environment; so our basic belief-forming mechanisms cannot rest on reasoning, which is too slow and uncertain. And if a belief was not formed by reasoning there is little ground for thinking that it must nonetheless be certifiable, retrospectively, by rational argument, since it is not clear what additional practical benefit that would confer.

It is arguable, therefore, that the appeal to the threat of scepticism yields no decisive proof of the inadmissibility of private states in the semantics of public language.

4 The approach via rule-following

The question may also be viewed from another direction. Recent discussion of Wittgenstein's writings about rules (see Meaning and rule-following) has lead some (eminently Saul Kripke 1982) to suggest that the notion that two items are of the same kind can only have meaning against the background of a consensus in usage. This principle gives rise to the following argument.

Signs have meaning only because they are used in accordance with rules. So if anything plays an essential role in determining the meaning of a sign it must be capable of playing this role consistently, that is, in the same way on the various occasions on which the sign is used. Therefore private states can play an essential role only if it can be said that the same kind of private state was experienced on different occasions. But that, according to the above principle, can be so only if there is a communal practice with regard to these items among a number of speakers. And there can be no such communal practice - precisely because they are private. It follows that they play no part in the determination of meaning.

The 'consensus-principle', once accepted, instantly defeats the possibility of a private language, if that be understood as a language which of necessity only one person can speak. But that does not mean that it immediately dismisses private states from any position in the semantics of public languages. For if one can have true beliefs about another person's private states (as perhaps one can, so long as the hypothesis advanced at the beginning of §3 remains unrefuted) then there can be such a thing as agreement about these states, and even disagreement, at least where insincerity is suspected. And then if this is held not to satisfy the demand for communal practice, it needs to be explained why the consensus required should need to be based on mutual knowledge rather than just confident belief - no easy task, since it has not proved easy convincingly to establish *any* form of the consensus principle.

5 Further consequences

This question has not normally been pursued for its own sake, but rather for that of its alleged consequences. M.A.E. Dummett, having concluded that a speaker's grasp of the meaning of a sentence must consist in their potential for outward, publicly accessible behaviour, tentatively suggests a further conclusion: it must consist in the capacity to recognize whatever state of affairs it is that makes the sentence true, in other words to tell that it is true, if and when it is. We can grasp no meaning, and so have no grasp of truth, beyond what we can in principle recognize as true; and it is this which is held to define antirealism and require the abandonment, characteristic of intuitionist mathematics, of the law of excluded middle as a generally valid logical principle.

The response to this train of thought must depend on the answers to a number of questions. Is it clear that the recognition of the fact that the truth conditions of a sentence obtain is the only (publicly accessible) way to demonstrate ('manifest') understanding of it? Is it even clear that the act of recognizing that certain conditions obtain *is* in the required sense publicly accessible? This cannot just be assumed. Recognition is an intensional state, since one can recognize the fact that *p* without recognizing the fact that *q*, even if these two facts always obtain together; we are therefore lead into the difficulties connected with such states, in particular the question whether they can be captured in purely physical behaviour (see Intensionality;Intentionality §2).

But neither should it be assumed that problems about meaning and understanding are easily resolved, or Dummett's conclusions easily avoided, if we are allowed the contrary assumption, that private states may be given a role in semantics. For that it is still too unclear what kinds of private state there are, and what their role could be, and hence what difference it makes if we admit them.

See also: Intuitionistic logic and antirealism; Private language argument

EDWARD CRAIG

References and further reading

All these items involve intricate argument but little or no technicality.

Craig, E.J. (1997) 'Meaning and Privacy', in *A Companion to the Philosophy of Language*, Oxford: Blackwell, 127-45.(Expansion of the material in this entry; also contains a section on the private language argument.)

Dummett, M.A.E. (1973) 'The Philosophical Basis of Intuitionistic Logic', in *Truth and Other Enigmas*, London: Duckworth, 1978, esp. 215-27.(Presents the arguments of §2 and suggests their connection with antirealism as discussed in §5.)

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell.(Expounds the considerations leading to the 'consensus' principle as mentioned in §4. A particularly fine piece of writing.)

Wright, C.J.G. (1987) 'Introduction', to *Realism, Meaning and Truth*, Oxford: Blackwell, esp. 13-23.(Expounds and discusses the arguments from acquisition and manifestation and connects them with antirealism.)

Probability theory and epistemology

The primary uses of probability in epistemology are to measure degrees of belief and to formulate conditions for rational belief and rational change of belief. The degree of belief a person has in a proposition A is a measure of their willingness to act on A to obtain satisfaction of their preferences. According to probabilistic epistemology, sometimes called 'Bayesian epistemology', an ideally rational person's degrees of belief satisfy the axioms of probability. For example, their degrees of belief in A and $-A$ *must sum to 1. The most important condition on changing degrees of belief given new evidence is called 'conditionalization'. According to this, upon acquiring evidence E a rational person will change their degree of belief assigned to A to the conditional probability of A given E. Roughly, this rule says that the change should be minimal while accommodating the new evidence. There are arguments, 'Dutch book arguments', that are claimed to demonstrate that failure to satisfy these conditions makes a person who acts on their degrees of belief liable to perform actions that necessarily frustrate their preferences. Radical Bayesian epistemologists claim that rationality is completely characterized by these conditions. A more moderate view is that Bayesian conditions should be supplemented by other conditions specifying rational degrees of belief.*

Support for Bayesian epistemology comes from the fact that various aspects of scientific method can be grounded in satisfaction of Bayesian conditions. Further, it can be shown that there is a close connection between having true belief as an instrumental goal and satisfaction of the Bayesian conditions.

Some critics of Bayesian epistemology reject the probabilistic conditions on rationality as unrealistic. They say that people do not have precise degrees of belief and even if they did it would not be possible in general to satisfy the conditions. Some go further and reject the conditions themselves. Others claim that the conditions are much too weak to capture rationality and that in fact almost any reasoning can be characterized so as to satisfy them. The extent to which Bayesian epistemology contributes to traditional epistemological concerns of characterizing knowledge and methods for obtaining knowledge is controversial.

1 Degrees of belief and probability

We ordinarily make a tripartite distinction among believing *Q*, believing not-*Q*, and suspending belief with respect to *Q*. Advocates of subjective probability think that beliefs come in more finely graded degrees. They further claim that rationality requires that degrees of belief satisfy the probability calculus and change in accord with a rule called 'conditionalization'. The classic sources for this view, Ramsey (1931), de Finetti (1972) and Savage (1954), argued that beliefs failing to conform to the probability calculus lead to actions that frustrate the believer's goals and for that reason a rational person will attempt to bring their degrees of belief into line with the probability calculus (see Ramsey, F.P. §2).

The degree to which someone believes that *Q* is manifested in their tendency to act on that belief. For example, if Arabella wants to get downtown by 5 p.m. and has a greater degree of belief that she will get downtown by 5 if she takes the train than if she takes the bus she will, other things being equal, take the train. One suggestion for measuring degrees of belief is in terms of the prices one is willing to pay for gambles. A simple gamble on *Q* is one in which $1 is paid if *Q* and nothing if not *Q*. Suppose that Arabella is willing to either buy or sell (in which case she is committed to pay the buyer of the gamble $1 if *Q* and nothing if not *Q*) a simple gamble on *Q* for $*x*, where $0 \leq x \leq 1$. In this case $*x* is the fair value of the gamble for Arabella and her degree of belief in *Q* is *x*. Advocates of subjective probability claim that an ideally rational person will assign degrees of belief to all propositions expressible in their language.

Nevertheless, measuring degrees of belief by willingness to pay for gambles is problematic. Experimental psychologists have accumulated evidence that seems to show that the value a person assigns to a gamble is dependent on how the gamble is described, on context, and is unstable (Kahneman, Slovic and Tversky 1982). Further, measuring degrees of belief in terms of gambles unrealistically assumes that the believer attaches no value (positive or negative) to gambling as such, that monetary gain corresponds to gains in utility, and that considering buying a gamble does not change one's degree of belief. These problems are not fatal since more sophisticated ways of measuring degrees of belief have been developed that do not depend on these assumptions. But even if there is a way of accurately measuring exact degrees of belief when they exist it is not plausible that we have exact

degrees of belief in most propositions. Most people would be unwilling to assign a fair value to a gamble. Usually the amount we would be willing to pay for a gamble is less than the amount at which we would sell it. These considerations have led to proposals for representing degree of belief in terms of intervals or more complicated structures (Levi 1980). For the most part these complications will be ignored here.

Degrees of belief that conform to the probability calculus are called 'subjective probabilities' since they are relative to believers. To satisfy the probability calculus a function P(*A*) (the probability of *A*) on a set of sentences (closed under negation and conjunction) must satisfy the consequences of the following axioms:

(1) $0 \leq \mathrm{P}(A) \leq 1$
(2) P $(A) = 1$ if *A* is logically true
(3) $\mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \mathbin{\&} B)$

Degrees of belief that satisfy the probability calculus are said to be 'coherent'. Psychologists have accumulated evidence that shows that people's degrees of belief are typically not coherent. For example, subjects in experimental studies frequently assign a higher probability to a conjunction $Q \mathbin{\&} R$ than they do to *Q* in violation of the probability calculus (Kahneman, Slovic and Tversky 1982) (see Rationality of Belief). The failure of beliefs to be coherent is not an objection to subjective probability since coherence is a normative notion. Advocates of subjective probability think coherence is a norm for degrees of belief in the way that logical consistency is a norm.

Among the more important arguments for the claim that rationality requires coherence is one based on 'the Dutch book' theorem. A Dutch book is a set of gambles such that one who takes the gambles suffers a net loss no matter what. For example, if someone has degrees of belief $\mathrm{P}(Q) = 0.6$ and $\mathrm{P}(-Q) = 0.6$ then they will exchange $.6 for a gamble on *Q* that pays $1 if *Q* and nothing otherwise, and $.6 for a gamble on $-Q$ which pays $1 if $-Q$ is true and nothing otherwise. They cannot win both, so no matter whether *Q* is true or false they will lose $.2. It can be proved that satisfaction of the probability calculus is necessary and sufficient for not being subject to this kind of Dutch book. This establishes that rationality requires coherence only if rationality requires immunity from Dutch books, which is not obvious.

2 Changing degrees of belief

Suppose that Arabella initially has degrees of belief P(*A*), P(*B*) and P($A \mathbin{\&} B$) and that $\mathrm{P}(A) \neq 0$ and then changes the degree of belief she assigns to *A* to $\mathrm{P}'(A) = 1$. Conditionalization says that she rationally ought to change the degree of belief she assigns to *B* to $\mathrm{P}'(B) = \mathrm{P}(B|A)$. P($B|A$) is the *conditional probability* of *B* on *A* and is equal to $\mathrm{P}(A \mathbin{\&} B)/\mathrm{P}(A)$. For example, if the initial degrees of belief are $\mathrm{P}(M) = 0.5$, $\mathrm{P}(W) = 0.4$, $\mathrm{P}(U) = 0.1$ (*M* is the proposition that Colonel Mustard committed the crime, *W* is the proposition that Mrs. White did it, and *U* that Professor Plum did it), then when Holmes discovers that Colonel Mustard is innocent, conditionalization prescribes that his new degrees of belief should be $\mathrm{P}'(W) = \mathrm{P}(W|-M) = 0.8$ and $\mathrm{P}'(U) = \mathrm{P}(U|-M) = 0.2$. These changes seem reasonable since they involve the least change in degrees of belief required to accommodate the new information while maintaining coherence.

An important rule for calculating conditional probabilities is Bayes' theorem (after Thomas Bayes, who used it to calculate the probability that the stars are distributed randomly). Suppose that someone assigns degree of belief P(*H*) (the prior probability of *H*) and conducts an experiment *E* with possible outcomes *e* and $-e$. Suppose also that they assign degrees of belief P($e|H$) and P($e|-H$). Then the conditional probability P($H|e$) is given by Bayes' theorem:

$$\mathrm{P}(H|e) = \frac{\mathrm{P}(H) \times \mathrm{P}(e|H)}{\mathrm{P}(H) \times \mathrm{P}(e|H) + \mathrm{P}(-H) \times \mathrm{P}(e|-H)}$$

To apply Bayes' theorem one needs prior probabilities P(*H*), P($-H$) and likelihoods P($e|H$), P($e|-H$). The likelihoods are usually thought of as deriving from statistical or theoretical information about probabilities. The prior probabilities are the believers' initial degrees of belief. Their status is controversial and has received much discussion some of which will be summarized later. As an example, suppose that a physician assigns the prior probability $\mathrm{P}(D) = 0.3$ to the hypothesis that a patient has a certain disease. Suppose that there is a test for the disease with outcomes *e* and $-e$ such that $\mathrm{P}(e|D) = 0.9$ and $\mathrm{P}(e|-D) = 0.2$. If the test is conducted and yields result *e* then according to the rule of conditionalization the physician's new degree of belief $\mathrm{P}'(D) = \mathrm{P}(D|e)$,

which by Bayes' theorem is 0.65.

Not all changes in degrees of belief are initiated by changes from an intermediate value to 1. Richard Jeffrey (1983) proposes a rule, 'Jeffrey Conditionalization' (JC), to accommodate such changes. If one changes the degree of belief assigned to *A* from P(*A*) to $\mathrm{P}'(A)$ then one should change the degree of belief assigned to *B* from P(*B*) to

(JC) $\mathrm{P}'(B) = \mathrm{P}'(A) \times \mathrm{P}(B|A) + \mathrm{P}'(-A) \times \mathrm{P}(B|-A)$

Conditionalization is the special case of JC in which $\mathrm{P}'(A) = 1$ (and $\mathrm{P}'(-A) = 0$). So, for example, if Arabella changes her degree of belief in *M* to 0.7, the degrees of belief she assigns to *W* and *U* should be 0.24 and 0.06. An analogue of Bayes' theorem can be derived to calculate how degrees of belief should change by JC.

A further suggested constraint on rational degrees of belief is a principle, called 'Reflection', that concerns attitudes towards future beliefs (van Fraassen 1984):

Reflection: $\mathrm{P}_t(Q|\mathrm{P}_{t+x}(Q) = b) = b$

According to Reflection, on the condition that a person assigns a degree of belief at time $t + x$ to *Q* of *b* they ought to assign a degree of belief at time *t* to *Q* of *b*. The principle expresses the idea that one should consider her future degrees of belief as being formed rationally on the basis of accurate evidence. Of course, if someone believes that for some reason the beliefs that she will have at $t + x$ are not formed rationally or accurately then it would seem that rationality may require departures from Reflection. For example, if one knew today that tomorrow one would be a subject in an hypnosis experiment and caused to believe that the earth is flat, that would provide no reason to believe today that the earth is flat.

David Lewis (1986) proposed another constraint on rational belief that connects beliefs about objective chances with degrees of belief.

(PP) $\mathrm{P}_t(Q|\mathrm{ch}_t(Q) = x \;\&\; E) = x,$

where $\mathrm{ch}_t(Q)$ is the chance or objective probability at time *t* of *Q*. Objective probabilities are objective features of reality independent of our beliefs. There is controversy concerning how to construe objective probabilities. Among the proposals are that they are irreducible, that they are hypothetical frequencies (van Fraassen 1980), that they are actual frequencies, and that they are determined by the best fitting theory of the world (Lewis 1994) (see Probability, interpretations of). In any case, PP says that the degree of belief one has at *t* in *Q* given the condition that the chance of *Q* at *t* is *x* and other admissible information *E* is itself *x*. To a first approximation *E* is admissible with respect to *Q* if *E* provides no information about how chance events at or after *t* will turn out. We apply PP when we assign a degree of belief of 0.5 to the event of the coin landing heads when flipped, given that we believe that the coin is fair. In situations where one does not know the objective chance of an event *e* but assigns degrees of belief to exclusive and exhaustive hypothesis concerning that chance, PP can be used to provide the conditional degrees of belief of *e* given the various hypotheses. In the example of Bayes' theorem described above, the degrees of belief P($e|D$) and P($e|-D$) may derive from applications of PP. By linking objective chances and degrees of belief, PP grounds a Bayesian epistemology of objective chances. Exactly why PP should determine degrees of belief in chance events has not received much discussion.

There are Dutch book arguments for conditionalization, JC, and Reflection. The Dutch book argument for JC establishes that if a person places a value on conditional gambles (a conditional gamble is one that is in effect only if a condition is satisfied), and changes their degrees of belief in conformity with any rule other than JC, they will be susceptible to a Dutch book by a bookie who knows their rule.

The Bayesian principles of rationality concerning belief are part of a more general account of rationality connecting belief, preference and decisions called 'Bayesian decision theory'. In Bayesian decision theory decisions are seen as analogous to gambles in which the consequences of a decision depend on possible states of nature. The degree to which a decision-maker values a consequence *c* is its 'utility' $u(c)$. If we assume that a person has sufficiently many degrees of belief then utilities can be measured in terms of choices among gambles. If someone prefers c_1 to anything else and prefers anything else to c_n and is indifferent between a gamble in which they obtain c_1 with probability *p*, and c_n with probability $1 - p$ (according to their degrees of belief) then

$u(c) = p$. The key principle of Bayesian decision theory is that one selects a decision from among the possible alternatives that maximizes expected utility. The expected utility of a decision is given by:

$u(d) = \Sigma \mathrm{P}(h_i) \times u(c_i)$.

The set $\{h_i\}$ is an exclusive and exhaustive set of states of nature, the P(h_i) are the decision-maker's degrees of belief, and the $u(c_i)$ are the utilities the decision-maker assigns to the consequence c_i of their doing *d* when h_i obtains (see Decision and game theory).

The view that rationality requires that beliefs cohere, that they change by conditionalization, and that decisions obey the maximum expected utility rule form the core of 'Bayesian epistemology' (Reflection and PP are attractive additions but have not received as much discussion). Bayesian epistemology does not provide traditional analyses of knowledge or justification but it does describe ideals of synchronic and diachronic rationality. Some advocates think of it as supplanting, and others as supplementing traditional epistemology.

One issue that arises within Bayesian epistemology involves the relationship between degrees of belief and other epistemological concepts. An issue that has received much discussion is acceptance. Epistemologists mean a number of different things by 'acceptance' although there seems to be agreement that someone who accepts a proposition is making a commitment to defend its reasonableness. Acceptance is usually thought of as rationally requiring closure under logical consequence and conjunction. If one accepts $h_1, \ldots, h_n$ then one rationally should accept every consequence of their conjunction. One proposal for connecting subjective probability and acceptance is rule R: one should accept *h* if $\mathrm{P}(h) > r$ (where *r* is a threshold value close to 1). But this suggestion immediately leads to 'the lottery paradox'. Suppose that $r = 0.99$ and that a fair lottery has 1,000 tickets. The acceptance rule implies that one should accept each statement $-hk_i$ that says that ticket k_i will not win (since $\mathrm{P}(-hk_i) = 0.999$). But one knows that one of the tickets will win, so one ends up accepting a contradictory statements. Various responses to the paradox have been proposed (see Paradoxes, epistemic §1).

3 Bayesian epistemology and scientific method

Support for the Bayesian approach to epistemology comes from the fact that various features of scientific method follow from Bayesian principles (Rosenkrantz 1977; Earman 1992) (see Confirmation theory; Inductive inference; Scientific method). For example, a natural measure of the extent to which evidence *e* confirms hypothesis *H* is $\mathrm{P}(H|e) - \mathrm{P}(H)$. It is easy to see from Bayes' theorem that the greater the ratio $\mathrm{P}(e|H)/\mathrm{P}(e|-H)$ the more *e* confirms *H*. The conditional probabilities P($e|H$) and P($e|-H$) often measure how well the two hypotheses explain *e*. So Bayes' theorem endorses the principle that evidence confirms a hypothesis to the extent that the hypothesis outperforms its competitors in explaining the evidence. A related consequence of Bayes' theorem is that the greater the difference between P($e|H$) and P($e|-H$) the more the result of the test confirms one or the other hypothesis. In other words, the more severe the test the more the confirmation.

The Bayesian approach also rationally motivates experimentation. The uncertainty relative to an inquirer of a set of hypotheses $\{h_k\}$ is measured by the entropy of the probability distribution over the h_k; entropy $\{h_k\} = \Sigma \mathrm{P}(h_k) \log \mathrm{P}(h_k)$. Entropy is greatest when all the h_k are assigned equal probability and least when one of the hypotheses is assigned a probability of 1 and the rest 0. An inquirer would like to reduce the entropy of their probability distribution since that brings them closer to certainty. It is a consequence of the probability calculus that any experiment E that has an outcome *e* such that $\mathrm{P}(e|h_i) \neq \mathrm{P}(e|h_k)$ for $i \neq k$ has a positive expected value of reducing entropy. If E has outcomes which differentiate among the h_k and if outcomes of repetitions of E are independent given the h_k then it follows that the inquirer can expect that as they continue to experiment, the entropy will continue to decrease to its minimum; that is that they will eventually assign a probability of 1 to one of the hypotheses. Further, if the probabilities P($e|h_k$) reflect the actual frequencies that would be obtained on independent repetitions of E then it can be shown that they can expect to assign a probability of 1 to the true hypotheses.

While most epistemologists accept true belief as the goal of rational inquiry, this claim has recently been criticized as mistaken or unjustified (Stich 1992). An interesting by-product of the Bayesian theory of experimentation can also be used to provide a justification of the instrumental value of acquiring true belief (Loewer 1995).

4 Criticisms of Bayesian epistemology

There is little doubt that aspects of scientific method can be illuminated by the Bayesian framework. But it is much less clear either that Bayesian epistemology provides satisfactory answers to traditional epistemological questions or that the Bayesian account of rationality is cogent. The claim that the Bayesian account of rationality is correct has been challenged by attacking the arguments that rationality requires coherence and conditionalization and by attacking the relevance of the Bayesian account of rationality.

The Dutch book argument has been criticized by arguing that mere susceptibility to Dutch books does not make one irrational. A person who thinks that their beliefs are not coherent can refuse to buy or sell gambles and thus avoid Dutch books. The Dutch book argument for conditionalization is even less persuasive. The argument establishes only that someone who conforms to a rule for changing beliefs must conform to conditionalization to avoid Dutch books. But it is not clear that rationality requires that belief-change conforms to a rule. In fact, it seems that at least on some occasions we change our beliefs without conforming to any rule and that it is rational to do so. For example, when one learns of a new theory to which one had not previously assigned a degree of belief, it may be appropriate to reassess certain other degrees of belief. But it is hard to see how there could be a plausible rule in this situation specifying how one should change those degrees of belief.

The most disturbing objection is that the Bayesian ideal is epistemically irrelevant to us because people have, at best, few exact degrees of belief. Further, it is practically impossible to make sure that one's degrees of belief are coherent or that they change by conditionalization. Bayesian epistemologists have attempted to deal with the first problem by changing their account to represent beliefs by sets of probability distributions. Coherence can be identified with a set of coherent degrees of belief. Whatever the plausibility of this suggestion for representing beliefs it makes the second problem even worse. This problem is that it will generally be impossible to determine whether an assignment of degrees to a collection of sentences is coherent. In fact there is no computationally effective method for achieving coherence in general since coherence requires assigning logically equivalent sentences the same degree, and never assigning a higher degree to *A* than to *B* if *A* logically implies *B*. There is no computationally effective method for determining in general whether or not *A* logically implies *B*.

The Bayesian response to these objections is to grant that, even if coherence and conditionalization are standards we cannot generally meet, they are ideals we should strive to approximate.

Whether or not conformity to Bayesian principles provides an epistemically valuable ideal, Bayesian epistemology is certainly incomplete as an account of rational belief. Specifically, there are at least three places at which it requires supplementation.

(1) If $\mathrm{P}(A) = 0$, then P($B|A$) is undefined, so conditionalization on propositions with probability 0 is undefined. But it seems plausible that sometimes a rational person may change their degree of belief in a proposition from 0 to 1 (or an intermediate value). There are some suggestions in the literature for how to extend conditionalization to this case.
(2) There is no constraint on the conditions under which propositions can initiate conditionalization or Jeffrey conditionalization. For example, someone's beliefs can conform to Bayesianism even though they change their probabilities on the basis of unreliable hunches. Rationality apparently requires restrictions on the changes that can initiate changes. Empiricism suggests that the inputs should be restricted to reliable observations. However, it seems dogmatic to restrict all changes to those initiated by observations. Sometimes reflection on a problem or thinking up a new hypotheses initiates changes in degrees of belief in ways that are considered rational. Bayesians have yet to produce a compelling account of rationally legitimate belief-changes.
(3) Coherence does not preclude assigning a high degree of belief to the claim that the earth is flat or taking the observation of green emeralds observed before time *t* as evidence for all emeralds being 'grue' (that is, green if observed before *t* and blue otherwise). This has been called 'the problem of the priors' since subjective Bayesian puts no constraints on the prior probabilities P(h_i) that enter as inputs into Bayes' theorem. Some Bayesians point to certain 'convergence theorems' to mitigate this problem (Savage 1954). According to one convergence theorem, even if experimenters begin with different probabilities concerning hypotheses $h_1, \ldots, h_n$ as long as these are non-zero and they agree on the likelihoods P($e_k|h_i$) of the outcomes e_k then, under very general conditions, repeated experimentation and conditionalization will result in the convergence of their degrees of belief concerning the h_i. However, this result is cold comfort since even approximate agreement may require an

unrealistically large amount of experimentation. Also, there is nothing in subjective Bayesianism that requires agreement on the likelihoods on which the convergence results depend.

Some Bayesians have suggested principles or considerations that constrain rational probability distributions. One way of doing this invokes 'principles of indifference'. The simplest indifference principle says that if one has no definite knowledge concerning a set of exclusive hypotheses, one ought to assign them equal probabilities. For example, if one had no knowledge concerning whether a ball to be drawn from an urn was red, *R*, or not red, $-R$, then one ought to assign the two hypotheses a probability of 0.5. The trouble with this proposal is that the principle of indifference is very sensitive to exactly how the hypotheses are described. If the problem were described in terms of the hypotheses *R*, *G* (the ball is green), and *B* (the ball is blue) then the principle of indifference entails probabilities of a third for each hypothesis. More sophisticated indifference principles have been proposed (Jaynes 1968) that give unambiguous probability assignments in certain situations. But the situations in which these principles apply are highly idealized and, in any case, it is not at all clear why the probabilities they assign are more rational than alternatives.

Another approach to constraining rational probability distributions appeals to logical probability. The idea is that considerations of logic determine conditional probabilities P($R|Q$); the degree to which *Q* probabilifies *R*. This proposal has been developed for simple formal languages. Carnap (1950) introduced a number of such measures (see Carnap, R. §5). For example, if a first-order language contains monadic predicates *F* and *G* and names $a, b, c, \dots$ then one measure assigns equal probability to each 'state description' $Fa \,\&\, Ga \,\&\, Fb \,\&\, Gb \,\&\, \dots$. One problem with this suggestion is that it precludes learning from experience since propositions concerning different individuals are probabilistically independent. A different suggestion assigns equal probability to each 'structure description' and then divides the probabilities equally among the state descriptions that satisfy the structure descriptions. The trouble with this idea is that it precludes assigning probabilities greater than 0 to generalizations. This can also be remedied (Hintikka 1966). However, no one has proposed a characterization of logical probability that is both applicable to ordinary beliefs and rationally compelling.

Another proposal for constraining rational probability distributions has received some discussion. It is that rational distributions are determined by consensus relative to a group (Lehrer and Wagner 1981). The idea is that a rational probability distribution relative to a group of 'experts' can be extracted by taking a weighted average - the weights representing group members' assessments of one another's reliabilities - of the degrees of belief of members of the group. This proposal has some awkward consequences. Specifically, all the members of the group can agree that two events are probabilistically independent, and yet the consensus distribution makes them dependent (Laddaga and Loewer 1985). The proposal is also too relativistic to represent rational priors. The problem of the priors is as unsolved today as it was when Bayes first proved his theorem.

See also: Learning; Common-sense reasoning, theories of

BARRY LOEWER

References and further reading

Carnap, R. (1950) *Logical Foundations of Probability*, Chicago, IL: University of Chicago Press.(Classic presentation of the logical concept of probability.)

Earman, J. (1992) *Bayes or Bust*, Cambridge, MA: MIT Press.(Excellent presentation of Bayesian inference in science.)

Finetti, B. de (1972) *Probability, Induction, and Statistics*, New York: Wiley. (Classic presentation of the subjectivist concept of probability.)

Fraassen, B. van (1980) *The Scientific Image*, Oxford: Oxford University Press.(Important work in the philosophy of science that argues for the view that scientists never have good reasons to believe claims about unobservables.)

Fraassen, B. van (1984) 'Belief and the Will', *Journal of Philosophy* 81 (5): 235-58.(Formulates and discusses the 'reflection principle' discussed in §2 above.)

Hintikka, J. (1966) 'A Two Dimensional Continuum of Inductive Methods', in J. Hintikka and P. Suppes (eds) *Aspects of Inductive Logic*, Amsterdam: North Holland. (Extends the logical concept of probability to first order languages.)

Jaynes, E.T. (1968) 'Prior Probabilities' I.E.E.E. *Transactions on Systems Science and Cybernetics* SSC-4:

227-44.(Sophisticated development of the principle of indifference for assigning prior probabilities.)
Jeffrey, R. (1983) *The Logic of Decision*, New York: McGraw-Hill.(Excellent introduction to Bayesian Decision theory.)
Kahneman, D., Slovic, P. and Tversky, A. (eds) (1982) *Judgement under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press. (Collection of articles concerning the psychology of probabilistic inference and decision reasoning.)
Kaplan, M. (1996) *Decision Theory as Philosophy*, New York: Cambridge University Press.(Discusses the role of acceptance in Bayesian epistemology.)
Laddaga, R. and Loewer, B. (1985) 'Destroying the Consensus', *Synthèse* 62: 79-95.(Criticizes Lehrer's proposal for forming consensus degrees of belief.)
Lehrer, K. and Wagner, C. (1981) *Rational Consensus in Science and Society*, Dordrecht: Reidel.(Develops an account of consensus degrees of belief.)
Levi, I. (1980) *The Enterprise of Knowledge*, Cambridge, MA: MIT Press. (Develops a quasi-Bayesian account of inference and decision. Includes the idea of representing belief as sets of probability distributions.)
Lewis, D. (1980) 'A Subjectivist's Guide to Objective Chance', in *Collected Papers*, New York: Oxford University Press, vol. 2, 1986.(Very significant article discussing the connection between subjective degrees of belief and objective probability.)
Lewis, D. (1994) 'Humean Supervenience Debugged', *Mind* 103: 473-90.(Follow up to Lewis' 1980 article.)
Loewer, B. (1995) 'The Value of Truth', in E. Villanueva (ed.) *Rationality and Truth*, Atascadero, CA: Ridgeview Publishing Co.(Provides a Bayesian justification for the value of true belief.)
Ramsey, F.P. (1931) 'Truth and Probability', in *Foundations of Mathematics and Other Essays*, London: Routledge & Kegan Paul.(Classic early presentation of subjective probability and utility.)
Rosenkrantz, R. (1977) *Inference, Method, and Decision*, Boston, MA: Reidel.(The best overall introduction to Bayesian philosophy of science and approaches to statistics.)
Savage, L.J. (1954) *Foundations of Statistics*, New York: Wiley.(Classic development of Bayesian foundations of statistics and decision theory.)
Stich, S. (1992) *Fragmentation of Reason*, Cambridge, MA: MIT Press.(Poses formidable objections to epistemology and includes a criticism of the claim that truth is valuable.)

Probability, interpretations of

The term 'probability' and its cognates occur frequently in both everyday and philosophical discourse. Unlike many other concepts, it is unprofitable to view 'probability' as having a unique meaning. Instead, there exist a number of distinct, albeit related, concepts, of which we here mention five: the classical or equiprobable view, the relative frequency view, the subjectivist or personalist view, the propensity view, and the logical probability view. None of these captures all of our legitimate uses of the term 'probability', which range from the clearly subjective, as in our assessment of the likelihood of one football team beating another, through the inferential, as when one set of sentences lends a degree of inductive support to another sentence, to the obviously objective, as in the physical chance of a radioactive atom decaying in the next minute. It is often said that what all these interpretations have in common is that they are all described by the same simple mathematical theory - 'the theory of probability' to be found in most elementary probability textbooks - and it has traditionally been the task of any interpretation to conform to that theory. But this saying does not hold up under closer examination, and it is better to consider each approach as dealing with a separate subject matter, the structure of which determines the structure of the appropriate calculus.

1 The project

The task of interpreting probability might be approached in three distinct ways. The first way is to see the project as one of providing an explicit definition of the term 'probability' or, more usually, of the predicate 'has a probability of value *p*'. The second is to provide operational content to this predicate; in other words, to provide a set of procedures by means of which various probability values can be measured or attributed. This second task may result only in the provision of sufficient conditions for attributing the predicate, and conversely a solution to the first task may be accomplished without the definition providing us with a way to measure probability values. At one time, when verificationism held sway, a failure to provide measurement criteria was considered fatal to a definition (see Logical positivism §4; Operationalism), but we should keep separate these two tasks. The third approach is to provide an implicit definition of the predicate 'has a probability of value *p*' by means of an axiomatized theory of probability, followed by an interpretation or model for that theory. That is, because the concept of probability is often considered to be mysterious and inaccessible to observation in a way that the concept of, say, blue, is not, the best and perhaps the only way of providing content to the concept is by constructing a detailed theory of probability and providing a model within which the theory is true. This third project consists in putting structural constraints on possible interpretations of a probability function, and it will not ordinarily result in an explicit definition. It will also not usually assign specific values to outcomes, except for certain extremal values, such as the certain or the impossible event.

2 The formal theory

The structure of elementary probability theory is often motivated by an appeal to familiar facts about relative frequencies. Suppose that a chance process, such as rolling a die, has *N* possible outcomes which form an outcome space $\Omega = \{1, 2, 3, 4, 5, 6\}$. By taking subsets of Ω, such as $\{2, 4, 6\}$, we have an event consisting in the die coming up even. Then take all possible subsets of Ωand this will form an algebra A of subsets of Ω, for example, a set of subsets of Ωclosed under complementation ('negation') and union ('disjunction'). Let us say that an event *E* in A occurs if the elementary outcome, such as '2', is in *E*. Then we can define a probability function P over A by letting $\mathrm{P}(E) =$ the number of times *E* occurs/the total number of repetitions of the chance process. This immediately gives us that

(1) $\mathrm{P}(\Omega) = 1$
(2) $\mathrm{P}(E) \geq 0$
(3) If two events *E* and *F* are mutually exclusive, then $\mathrm{P}(E \cup F) = \mathrm{P}(E) + \mathrm{P}(F)$.

For mathematical convenience, this elementary theory is usually extended by requiring countable additivity, wherein (3) is replaced by

(3′) $\mathrm{P}(\cup_{i=1}^{\infty} E_i) \sum_{i=1}^{\infty} \mathrm{P}(E_i)$ when the E_i are all mutually disjoint.

(The structure of A then needs to be more complex in ways that are inessential here.)

An important definition to add is that of conditional probability, $\mathrm{P}(A \mid B)$, the probability of *A* given *B*, defined as $\mathrm{P}(A \mid B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$. Two elements of A are *independent* if and only if $\mathrm{P}(A \mid B) = \mathrm{P}(A)$ or, equivalently, if $\mathrm{P}(A \cap B = \mathrm{P}(A) \times \mathrm{P}(B)$.

This abstract calculus is now no longer tied to the particular interpretation with which we began, and the algebra can be one of propositions rather than events, an ontology preferable for logical or subjective interpretations. Different accounts of probability can now be compared to this axiomatic theory.

3 Relative frequency theories

Here the ontology is one of event types, and the probability is explicitly defined either as the actual finite relative frequency as in §2 above, or as the limit value of the relative frequency when the total number of repetitions goes to infinity. This gives the value of the probability as an empirical property of the sequence or reference class of outcomes which generates the frequency, rather than absolutely. A naïve finite frequency interpretation is clearly unsatisfactory. Suppose a die is rolled a finite number of times *N*. Then no outcome, say a '6', can have a probability value more fine-grained than on a scale 0, $1/N$, $2/N$, …, l. Thus, after two throws, a '6' could only have a probability of 0, 0.5, or 1. This problem could be circumvented by requiring sufficiently many repetitions, where 'sufficiently many' would be a function of the number of possible outcomes and degree of precision required. A more standard remedy for this problem is to move to a limiting relative frequency interpretation, wherein the probability value is simply defined as the limit of m/N as $N \to \infty$, where *m* is the number of successes. This solves our first, definitional, task, but it produces a difficulty for the measurement project, for there is no guarantee that the value of the relative frequency after, say, 1,000 repetitions, will be the same as, or close to, the limiting value, because an unusual run of outcomes could occur initially. What one can do is appeal to the strong law of large numbers for binary valued outcomes, which asserts that if the repetitions of the experiment are independent and identically distributed (that is, the probability does not vary from repetition to repetition) then for every $\varepsilon > 0$, with probability one $\mid m/N - p \mid > \varepsilon$ only finitely often.

This provides assurance that in the long run, the probability that the limiting frequency will differ from the 'true' value of *p* is zero, but at the cost of introducing a second order probability which in turn needs to be interpreted. As an additional requirement, Richard von Mises (1957) correctly insisted that limiting frequency values can only be drawn from random sequences of data. Thus, if our die gave us the sequence of outcomes 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, …, the limiting frequency of a '2' would be 1/6, but quite clearly on every $(6N+2)\mathrm{nd}$ throw ($N = 0, 1, 2, \ldots$) the probability of '2' would be 1, and would be 0 on any other throw. This highlights what is known as the problem of single-case probabilities - how do we transfer a probability value from a class to a single outcome? A standard answer is that the appropriate frequency for an outcome type is one drawn from a random sequence. Attempts to solve this single-case problem have led to complex and fascinating theories of randomness and statistical relevance, each of which is of philosophical interest in its own right (Sklar 1993).

4 Propensity theories

We have seen above that relative frequencies are defined relative to a class of outcomes. If we focus on the fact that outcomes can be produced by a fixed set of generating conditions, then it is reasonable to attribute probabilities to a system in a physical context, especially when the system producing the outcomes is irreducibly indeterministic. This gives us a chance disposition or *propensity* to produce a given outcome. For example, the probability of decay within the next minute is a physical property of a radioactive atom, just as is its atomic weight. This view, which seems to have originated with C.S. Peirce, and was resurrected by Karl Popper (see Popper, K.R. §3), has been criticized as excessively metaphysical. This is an unfair criticism, for the propensity value can be measured empirically for a system with a fixed propensity by employing the strong law of large numbers, mentioned above, together with statistical estimation techniques, thus satisfying the operational criterion. Being a thoroughly ontological interpretation, propensity accounts do not provide an explicit reductionist definition of probability, for propensities are basic, often primitive, properties of the world. However, no satisfactory solution has yet been given for the third project of providing a propensity calculus. Nor, despite a number of efforts by Popper and others, has a satisfactory argument been provided that quantum probabilities are

obviously propensities in any detailed sense (see Causation; Quantum mechanics, interpretation of; Statistics).

5 Subjective probabilities

In contrast to the objective interpretations just described, probability has always had a close relation with degrees of rational belief. Within this tradition, one can measure such degrees of belief operationally by means of betting behaviour. Here the elements to which the probabilities attribute values are propositions or sentences rather than events. By means of an ingenious operational process, known as the Dutch book method, within which the probability value is defined in terms of the lowest odds at which the gambler will accept a bet, one can show that plausible constraints on rational behaviour are satisfied if and only if axioms (1)-(3) and the definition of conditional probability of §2 hold. The subjective probability assignments are then said to be coherent. Thus we have satisfied the second and third criteria cited earlier, although it should be noted that different individuals can assign widely different values to a contingent proposition and both be coherent. De Finetti (1937), however, held that (3′) was not true for subjective probabilities because placing an infinite number of bets makes no sense for human agents. The Dutch book method will give us what are called prior probabilities but, except for extremal values, these are not arrived at by an a priori process. Rather, they are the expression of an unarticulated amalgamation of background knowledge. Although there is some controversy over how best to revise beliefs in the light of empirical evidence, the traditional way to do this is via conditionalization, using Bayes’ theorem. If H_j is a hypothesis (say that the die is fair), *E* is empirical evidence (say data from throws) and P_0 is our prior subjective probability assignment, then

$$\mathrm{P}_1(H_j) = \mathrm{P}_0(H_j \mid E) = \mathrm{P}_0(E \mid H_j)\mathrm{P}_0(H_j) / \sum_i \mathrm{P}_0(E \mid H_i)\mathrm{P}_0(H_i)$$

where $\{H_i\}$ is the set of hypotheses under consideration and P_1 is the new, posterior, probability assignment. Provided that the members of $\{H_i\}$ are distinct hypotheses (only one can be true at any time) and that $\sum_i \mathrm{P}_0(H_i) = 1$ (the possibility that no member of $\{H_i\}$ is true is not entertained), $\sum_i \mathrm{P}_0(E \mid H_i)\mathrm{P}_0(H_i) = \mathrm{P}_0(E)$, which gives $\mathrm{P}_1(H_j) = \mathrm{P}_0(H_j \mid E)\mathrm{P}_0(H_j)/\mathrm{P}_0(E)$.

The above gives us a normative theory of how agents should distribute their degrees of belief, but there has always been a divergence between this normative Bayesianism and individuals’ actual degrees of belief. Investigations by psychologists and economists have revealed systematic differences from Bayesian prescriptions, even by agents well-versed in probabilistic reasoning.

6 Classical interpretations of probability

This approach allocates probability values by dividing the outcomes into equipossible cases, and then using a principle of indifference to give each equipossible case an equal probability. Thus, if one considers, on the basis of symmetry, that each side of a die is just as likely to come up as any other side, then the classical theory attributes a value of $1{:}\,6$ to each such outcome. This is not, contrary to many claims, always an a priori attribution, because it generally depends upon some specific empirical knowledge about symmetries of the system. The classical approach is unsatisfactory for a variety of reasons. First, it is limited to situations in which equipossible cases are available. Second, there are straightforward paradoxes associated with this approach, which allow different probability values to be attached to the same event. The simplest is one introduced by Bertrand. Suppose you have wine and water mixed, in a ratio somewhere between one part of water to one of wine, and two parts water to one of wine. Using an indifference principle on the ratio water:wine, we have that the probability of the ratio lying between 1 and 1.5 is $1{:}\,2$. Now consider the ratio wine:water. This can lie between $1{:}\,2$ and 1, and the water:wine ratio of $1{:}\,5$ is a wine:water ratio of $2{:}\,3$. The indifference principle then says that the probability of this ratio lying between 1 and $2{:}\,3$ has a probability of $2{:}\,3$. This is contradictory.

7 Logical interpretations and other approaches

The logical approach takes propositions as the objects to which probabilities are ascribed, and interprets a conditional probability as a logical relation giving a degree of inductive support from the conditioning sentence to the conditioned. Thus, $\mathrm{P}(H \mid E)$ is the degree of inductive support that the evidence statement *E* gives to the hypothesis *H*. To measure $\mathrm{P}(H \mid E)$, Carnap (1945) considered state descriptions. In the simplest case, one lists all the individuals $a_1 \ldots a_n$ in a world and all the predicates $F_1 \ldots F_r$. Then a state description is an attribution of

F_i or $\sim F_i$ to each individual a_j for all *i* and *j*, that is, a maximally consistent description of some possible world. If we now attribute, in an a priori way, a measure *m* on state descriptions (and this can be done in a number of ways), then the conditional logical probability is just $m(H \mathbin{\&} E)/m(E)$. The principal drawback to using state descriptions is that learning from experience is impossible, simply because all predicates are logically independent and hence observing an instance of the property it represents gives no information about instances of any other. For this reason, Carnap switched to structure descriptions, within which individuals are indistinguishable. Interestingly, the differences between Maxwell-Boltzmann statistics, Bose-Einstein, and Fermi-Dirac statistics in physics are representable as differences that depend upon which states are physically possible. Because it is an empirical fact which particles satisfy which statistics, this sheds considerable doubt upon the whole enterprise of making a priori probability attributions.

There are now well-entrenched theories of comparative probability within which numerical values are not assigned, but one outcome is simply considered to be at least as probable as another. Alternatively, rather than assigning specific values to a proposition, intervals of probability can be assigned to mirror our uncertainty about the correct value via upper and lower probabilities. The connections between subjective probability and objective chance are now also the subject of much interest.

See also: Carnap, R. §5; Confirmation theory; Decision and game theory; Field theory, quantum; Inductive inference; Probability theory and epistemology; Rational beliefs; Rational choice theory; Reichenbach, H. §3

PAUL HUMPHREYS

References and further reading

Carnap, R. (1945) ‘On Inductive Probability’, *Philosophy of Science* 12: 72-97.(An accessible introduction to logical probability by its primary advocate.)

Cohen, L.J. (1989) *An Introduction to the Philosophy of Probability and Induction*, Oxford: Clarendon Press. (Chapter 2 is a brief but useful critical survey of the interpretations discussed in this article.)

Feller, W. (1968) *An Introduction to Probability Theory and Its Applications*, New York: John Wiley & Sons. (Still the best introduction to the mathematical theory of probability. Volume 1 treats discrete outcome spaces; volume 2 (2nd edition) 1971 deals with continuous outcome spaces.)

Finetti, B. de (1937) ’Foresight: its Logical Laws, its Subjective Sources’; repr. in H. Kyburg and H. Smokler (eds) *Studies in Subjective Probability*, New York: John Wiley & Sons, 1964.(A classic source on subjective probability.)

Jeffrey, R. (1983) *The Logic of Decision*, Chicago, IL: University of Chicago Press, 2nd edn.(A systematic exposition of one variety of Bayesianism.)

Kahneman, D., Slovic, P. and Tversky, A. (eds) (1982) *Judgement under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.(A collection of papers by psychologists describing how individuals reason probabilistically. See §5.)

Lewis, D. (1980) ‘A Subjectivist’s Guide to Objective Chance’, in R. Jeffrey (ed.) *Studies in Inductive Logic and Probability*, vol. 2, Berkeley, CA: University of California Press.(Reprinted with important Postscripts in D. Lewis, *Philosophical Papers*, vol. 2, Oxford: Oxford University Press, 1986. This article details some connections between objective and subjective probabilities.)

Mises, R. von (1957) *Probability, Statistics, and Truth*, New York: Dover, 2nd edn. (A nontechnical treatment of a relative frequency approach.)

Popper, K. (1974) ‘Suppes’s Criticisms of the Propensity Interpretation of Probability and Quantum Mechanics’, in P.A. Schilpp (ed.) *The Philosophy of Karl Popper*, La Salle, IL: Open Court, 1,125-39.(A detailed treatment of a propensity theory. The article to which it is replying - P. Suppes, ‘Popper’s Analysis of Probability in Quantum Mechanics’, 760-74 - should be read too.)

Sklar, L. (1993) *Physics and Chance*, Cambridge: Cambridge University Press. (Chapter 3 is a readable overview of probability, with a section - §4 - on randomness.)

Process philosophy

In the broad sense, the term 'process philosophy' refers to all worldviews holding that process or becoming is more fundamental than unchanging being. For example, an anthology titled Philosophers of Process (1965) includes selections from Samuel Alexander, Henri Bergson, John Dewey, William James, Lloyd Morgan, Charles Peirce and Alfred North Whitehead, with an introduction by Charles Hartshorne. Some lists include Hegel and Heraclitus. The term has widely come to refer in particular, however, to the movement inaugurated by Whitehead and extended by Hartshorne. Here, process philosophy is treated in this narrower sense.

Philosophy's central task, process philosophers hold, is to develop a metaphysical cosmology that is self-consistent and adequate to all experienced facts. To be adequate, it cannot be based solely on the natural sciences, but must give equal weight to aesthetic, ethical and religious intuitions. Philosophy's chief importance, in fact, derives from its integration of science and religion into a rational scheme of thought. This integration is impossible, however, unless exaggerations on both sides are overcome. On the side of science, the main exaggerations involve 'scientific materialism' and the 'sensationalist' doctrine of perception. On the side of religion, the chief exaggeration has been the idea of divine omnipotence. Process philosophy replaces these ideas with a 'panexperientialist' ontology, a doctrine of perception in which nonsensory 'prehension' is fundamental, and a doctrine of divine power as persuasive rather than coercive.

1 Philosophy's central task

Process philosophy is based on the conviction that the central task of philosophy is to construct a cosmology in which all intuitions well-grounded in human experience can be reconciled. Whereas cosmologies were traditionally based on religious, ethical and aesthetic as well as scientific experiences, cosmology in the modern period has increasingly been based on science alone. Process philosophers find this modern cosmology, which can be called 'scientific materialism', inadequate to those human intuitions that are usually called aesthetic, ethical and religious and, more generally, to those 'commonsense' beliefs that we cannot help presupposing in practice - such as the belief that our thoughts and actions are not wholly determined by antecedent causes. Such beliefs, rather than being explained away, should provide the final criterion for philosophical thought. In enunciating this criterion, Whitehead (1929) and Hartshorne (1970) are adopting the pragmatic maxim of Peirce and James that, if an idea cannot be lived in practice, it should not be affirmed in theory. The worldview of scientific materialism is also held to be inadequate for science itself. Although this is most obvious in biology and psychology (see §3 below), it is true even for physics (Whitehead 1925).

Part and parcel of philosophy's task is its role as 'the critic of abstractions'. Because of the tendency for overstatement, the abstractions from the more specialized disciplines usually need to be reformulated before they can be integrated into a self-consistent cosmology (Whitehead 1925). Because the abstractions of the physical sciences have recently been dominant, the primary critical task now is 'to challenge the half-truths constituting the scientific first principles' (Whitehead [1929] 1978: 10). At the root of these half-truths is usually the 'fallacy of misplaced concreteness', in which an abstraction from something, useful for particular purposes, is identified with the concrete thing itself. This fallacy lies behind scientific materialism, according to which everything, including human experience, is to be explained in terms of the locomotion of bits of matter devoid of spontaneity, internal process and intrinsic value (Whitehead 1925). The suggested alternative is to reconceive the basic units of the world as processes (see Whitehead, A.N.).

2 Two kinds of process

Although 'process philosophy' (which Whitehead himself did not use) probably became the label for the school of thought he founded primarily because of the title of his main work, *Process and Reality* (1929), the term is apt: 'The reality is the process' and 'an actual entity is a process' (Whitehead [1925] 1967: 72; [1929] 1978: 41). Whereas all process philosophies in the broad sense could agree with these statements, it is the particular interpretation given to them that constitutes the distinctiveness of Whiteheadian process philosophy. Central to this distinctiveness is the twofold idea that the actual units comprising the universe are momentary 'occasions of experience' involving two kinds of process.

Partly through the influence of quantum physics, Whitehead conceived of the most fundamental units of the world, the most fully actual entities, not as enduring individuals but as momentary events. Enduring individuals, such as electrons, molecules, and minds, are 'temporally ordered societies' of these momentary events. The idea that actual entities are events with both spatial and temporal extensiveness is indicated by calling them 'actual occasions' (Whitehead [1929] 1978: 77). Their temporal extensiveness means that they cannot exist at an 'instant' (understood as a durationless slice of time). Rather, they constitute, as Bergson had suggested, a more or less brief duration (from perhaps less than a billionth of a second in subatomic events to perhaps a tenth of a second at the level of human experience).

This idea paves the way for recognizing two kinds of process: a process within an actual occasion, called 'concrescence' (because it involves moving from potentiality to concreteness), and a process between actual occasions, called 'transition'. These two kinds of process involve the two basic kinds of causation: 'efficient causation expresses the transition from actual entity to actual entity; and final causation expresses the internal process whereby the actual entity becomes itself'. Through this distinction Whitehead seeks to fulfil a central task of philosophy: 'to exhibit final and efficient causes in their proper relation to each other' ([1929] 1978: 150, 84). This proper relation is that every actual occasion begins by receiving efficient causation from prior actual occasions, completes itself by exercising final causation, understood as self-determination, and then exercises efficient causation upon following occasions. The temporal process involves a perpetual oscillation between efficient and final causation.

The two previous paragraphs provide two aspects of Whitehead's alternative to materialism's way of overcoming dualism in favour of a cosmology with only one type of actual entity. According to Cartesian dualism, minds were temporal but not spatial, while material bodies were spatially extended but essentially nontemporal (being able to exist at an instant). Whitehead's idea that all actual entities are spatiotemporal events overcomes that dualism. His idea that these events involve both concrescence and transition overcomes the further dualism between actual entities that can exert only efficient causation and those that can exercise self-determination. The central feature of Cartesianism, however, was the dualism between actual entities with experience and those without. This dualism is overcome through the rejection of 'vacuous actualities', meaning things that are fully actual and yet void of experience. This rejection is expressed positively by considering all actual occasions to be 'occasions of experience' (Whitehead [1929] 1978: 29, 167, 189). This doctrine means, with regard to the internal process of concrescence, that 'process is the becoming of experience' (Whitehead [1929] 1978: 166). The meaning is not that all actual entities are conscious - most are not - but that they have some degree of feeling.

Although 'panpsychism' is the customary name for philosophies of this sort, 'panexperientialism' is better for this particular version, partly because the term 'psyche', besides suggesting experience too sophisticated to attribute to atoms or even cells, also suggests that the ultimate units endure through time, rather than being momentary experiences. Another essential feature of process philosophy's version is that the 'pan', meaning 'all', does not refer to literally all things but only to all genuine individuals (see Panpsychism §1). This distinction is central to process philosophy's solution to the mind-body problem.

3 The mind-body problem

Panexperientialists, like materialists, consider insoluble the problem of dualistic interaction: How could mind and brain cells, understood as actualities of ontologically different types, interact? Materialism seeks to avoid this problem by thinking of the mind as somehow identical with the brain. However, besides still having the problem of how conscious experience could arise out of insentient neurons, materialism is also hard-pressed to explain the apparent unity and freedom of our experience. The move by eliminative materialists, denying that there is any experience, unity or freedom to explain, rejects in theory what is inevitably presupposed in practice. Whiteheadian process philosophy suggests, on the basis of its panexperientialism, a 'nondualistic interactionism' meant to avoid the problems of both dualism and materialism. With dualism, it distinguishes (numerically) between mind and brain. The distinct reality of the mind, as a temporally ordered society of very high-level occasions of experience, provides a locus for the unity of our experience and its power to exercise self-determination. But by rejecting dualism's assumption that the mind is ontologically different from the brain cells, panexperientialism removes the main obstacle to understanding how our experiences could interact with our brain cells. As Hartshorne puts it: 'cells can influence our human experiences because they have feelings that we can feel. To deal with the

influences of human experiences upon cells, one turns this around. *We* have feelings that *cells* can feel' (1962: 229).

The freedom of bodily action has also been a problem for materialists, who may admit that they cannot help presupposing this freedom while claiming that the scientific worldview has no room for it. One of the assumptions behind this claim is that the behaviour of subatomic particles is fully specified by the laws of physics. A second is that all wholes, including human beings, are analogous to rocks and billiard balls, so that all vertical causation must run upward, from the most elementary parts to the whole. In process philosophy's panexperientialist ontology, by contrast, all actual entities are internally constituted by their relations to other occasions. This view allows for the emergence of higher-level actual occasions, so that spatiotemporal societies of actual entities can be of two basic types: besides aggregational societies, such as rocks, there is what Hartshorne (1972) calls the 'compound individual', in which a society with the requisite complexity gives rise to a 'dominant' member, which can then exercise downward causation on the rest of the society. This downward causation is possible, furthermore, because the 'laws' of nature are really its most widespread habits, and because atoms and subatomic particles are open to the particular influences of the environment in which they find themselves (Whitehead [1938] 1966: 154-5;[1933] 1967: 41).

4 Perception and prehension

Another distinctive feature of Whiteheadian process philosophy is its challenge to the 'sensationalist' theory of perception, according to which all knowledge of the world beyond the mind comes through sensory perception (see Perception, epistemic issues in). More fundamental than sensory perception, suggests Whitehead (1925), is a nonsensory mode of perception, called 'prehension', which may or may not be conscious. One example (which we call 'memory') occurs when an occasion of experience directly perceives occasions in its own past. Another instance is the mind's direct reception of influences from its brain (which sensory perception presupposes). This direct prehension of other actualities, through which we know of the existence of the 'external world', is also called 'perception in the mode of causal efficacy', because it provides the experiential basis, denied by Hume, for our idea of causation as real influence.

This idea of nonsensory prehension is central to process philosophy. It is implicit in the idea of panexperientialism: because sensory perception can be attributed only to organisms with sensory organs, the idea that all actual entities have experience presupposes a more primitive mode of perceptual experience that can be generalized to all individuals whatsoever. This idea is also presupposed in the acceptance of aesthetic, ethical and religious experiences as genuine apprehensions. It is crucial, thereby, to the task through which philosophy 'attains its chief importance,' that of fusing science and religion 'into one rational scheme of thought' (Whitehead [1929] 1978: 15).

5 Reconciling science and religion

One side of this task of reconciling science and religion involves what has been discussed above - the replacement of the materialistic worldview, with which science has recently been associated, with panexperientialism, which allows religious and moral experience as well as freedom to be taken seriously. The other side of the task involves overcoming exaggerations from the religious side that conflict with necessary assumptions of science. Here the main exaggeration involves the idea of divine power. Whitehead and Hartshorne do believe that a metaphysical description of reality points to the necessity of a supreme agent to which the name 'God' can meaningfully be applied. (Arguments for the existence of God are developed much more fully by Hartshorne (1941, 1962) than by Whitehead.) But they strongly reject the traditional doctrine of divine power, according to which God, having created the world *ex nihilo*, can interrupt its basic causal processes - a doctrine that, besides creating an insuperable problem of evil, also conflicts with the assumption of scientific naturalism that no such interruptions can occur. Their alternative proposal is that the power of God is persuasive, not coercive (Whitehead 1929, 1933; Hartshorne 1984).

6 Developments in the movement

Although Whitehead was one of the first philosophers to be included in 'The Library of Living Philosophers' (Schilpp 1941), his philosophy was largely ignored in the decades subsequent to its articulation in the 1920s and

1930s, partly because of the turn to anti-metaphysical forms of philosophy. Another factor was that, even within circles still interested in developing a naturalistic cosmology, Whitehead's panexperientialism and affirmation of God were felt to exceed the limits of a proper naturalism, which was largely equated with materialism (see Materialism). The most prominent advocate of Whiteheadian process philosophy in the following decades, furthermore, was Hartshorne, whose focus on the idea of God, while creating interest in theological faculties, reinforced suspicions in philosophical circles. Around 1960, however, a spate of books on Whitehead's philosophy inaugurated a period of greater interest (Lawrence 1956; Leclerc 1958, 1961; Christian 1959; Lowe 1962; Sherburne 1961; Kline 1963). In 1971, a journal, *Process Studies*, was created for the purpose of furthering the study and development of process thinking. In 1991, a volume devoted to the philosophy of Hartshorne appeared in the 'Library of Living Philosophers' (Hahn 1991).

Whiteheadian process philosophy has exerted some influence in a number of branches of philosophy, such as the philosophies of science, education, and art. Its major influence thus far, however, has continued to be in the philosophy of religion (Cobb 1965, 1969; Frankenberry 1987; Griffin 1976, 1991; Ogden 1966), including discussions of the relation between science and religion in particular (Barbour 1966, 1990).

See also: Process theism; Processes

DAVID RAY GRIFFIN

References and further reading

Barbour, I. (1966) *Issues in Science and Religion*, Englewood Cliffs, NJ: Prentice Hall.(A widely used text written primarily from the perspective of process philosophy.)

Barbour, I. (1990) *Religion in an Age of Science*, San Francisco, CA: Harper & Row; London: SCM.(An updated replacement of the previous book, based on Gifford Lectures.)

Browning, D. (1965) *Philosophers of Process*, New York: Random House. (Readable selection of writings of a number of 'process philosophers' in the broad sense.)

Christian, W.A. (1959) *An Interpretation of Whitehead's Metaphysics*, New Haven, CT: Yale University Press. (Focusing on the idea of 'transcendence', this otherwise careful analysis is flawed by a misunderstanding of the 'perishing' of actual entities and thereby of the causal efficacy involved in 'transition'.)

Cobb, J.B. (1965) *A Christian Natural Theology: Based on the Thought of Alfred North Whitehead*, Philadelphia, PA: Westminster.(This book, which has become a standard, provides one of the most precise accounts of Whitehead's philosophy in the course of showing its relevance to religious issues.)

Cobb, J.B. (1969) *God and the World*, Philadelphia, PA: Westminster.(The first three chapters of this more popular presentation, based on a series of lectures, are especially recommended.)

Frankenberry, N. (1987) *Religion and Radical Empiricism*, Albany, NY: State University of New York Press. (Places the relevance of Whitehead in the context of the radical empiricism of William James.)

Griffin, D.R. (1976) *God, Power, and Evil: A Process Theodicy*, Philadelphia, PA: Westminster.(The first book-length treatment of theodicy from the perspective of Whiteheadian-Hartshornean philosophy.)

Griffin, D.R. (1991) *Evil Revisited: Responses and Reconsiderations*, Albany, NY: State University of New York Press.(Responses to critiques of previous book.)

Griffin, D.R. (1997) *Unsnarling the World-Knot: Consciousness, Freedom, and the Mind-Body Problem*, Los Angeles, CA: University of California Press.(A lengthy treatment of the material in §3 above, developing the Whiteheadian position in relation to recent efforts by dualists and especially materialists.)

Griffin, D.R., Cobb, J.B., Ford, M.P., Gunter, P.A.Y. and Ochs, P. (1993) *Founders of Constructive Postmodern Philosophy: Peirce, James, Bergson, Whitehead, and Hartshorne*, Albany, NY: State University of New York Press.(Five original essays plus an introduction discussing the relevance of what these five philosophers have in common, such as panexperientialism, to current discussions.)

Hahn, L.E. (ed.) (1991) *The Philosophy of Charles Hartshorne*, The Library of Living Philosophers vol. 20, La Salle, IL: Open Court.(Descriptive and critical essays about Hartshorne's philosophy with lengthy replies by Hartshorne plus an intellectual autobiography.)

Hartshorne, C. (1937) *Beyond Humanism: Essays in the New Philosophy of Nature*, Chicago, IL: Willet, Clark. (Collection of early essays arguing that nature is loveable in its parts and as a whole.)

Hartshorne, C. (1941) *Man's Vision of God and the Logic of Theism*, New York: Harper & Row.(Hartshorne's first attempt to apply the logic he had learned from C.I. Lewis and H.M. Sheffer to the philosophy of religion.)

Hartshorne, C. (1962) *The Logic of Perfection and Other Essays in Neoclassical Metaphysics*, La Salle, IL: Open Court.(The 'other essays' provide a very readable introduction to his philosophy of nature, freedom, and

religion.)
Hartshorne, C. (1970) *Creative Synthesis and Philosophic Method*, La Salle, IL: Open Court. (The best survey of Hartshorne's philosophy.)
Hartshorne, C. (1972) *Whitehead's Philosophy: Selected Essays, 1935-1970*, Lincoln, NB: University of Nebraska Press.(Essays explaining the respects in which Hartshorne agrees and disagrees with Whitehead's version of process philosophy.)
Hartshorne, C. (1984) *Omnipotence and Other Theological Mistakes*, Albany, NY: State University of New York Press.(Dealing entirely with religious issues, this is the easiest of Hartshorne's books.)
Kline, G.L. (ed.) (1963) *Alfred North Whitehead: Essays on His Philosophy*, Englewood Cliffs, NJ: Prentice Hall. (This still-helpful collection both reflected and helped spark the new interest in Whitehead's philosophy.)
Lawrence, N. (1956) *Whitehead's Philosophical Development: A Critical History of the Background of Process and Reality*, Berkeley, CA: University of California Press. (A careful tracing of the development up to, but not including, Whitehead's *magnum opus*.)
Leclerc, I. (1958) *Whitehead's Metaphysics: An Introductory Exposition*, New York: Macmillan.(Comparing the formative elements of Whitehead's system to Aristotle's four 'causes', this introduction is helpful except on efficient causation.)
Leclerc, I.(ed.) (1961) *The Relevance of Whitehead: Philosophical Essays in Commemoration of the Centenary of the Birth of Alfred North Whitehead*, New York: Macmillan; London: Allen & Unwin.(A very good collection of essays by major commentators.)
Lowe, V. (1962) *Understanding Whitehead*, Baltimore, MD: Johns Hopkins University Press.(In most respects still the best introduction.)
Ogden, S.M. (1966) *The Reality of God and Other Essays*, New York: Harper & Row.(A philosophically rigorous but accessible explication and application of Hartshornean theism.)
Schilpp, P.A. (1941) *The Philosophy of Alfred North Whitehead*, The Library of Living Philosophers vol. 3, New York: Tudor.(Descriptive and critical essays by many philosophers, including Lowe, Quine, R.W. Sellars, Hartshorne, Dewey and C.I. Lewis.)
Sherburne, D.W. (1961) *A Whiteheadian Aesthetic: Some Implications of Whitehead's Metaphysical Speculation*, New Haven, CT: Yale University Press.(A vigorous and in some respects controversial account.)
Whitehead, A.N. (1925) *Science and the Modern World*, New York: Free Press, 1967.(The first book of Whitehead's metaphysical period, it is essential for understanding his alternative to scientific materialism.)
Whitehead, A.N. (1926) *Religion in the Making*, Cleveland, OH: World, 1960.(Whitehead's first application of his metaphysical vision to the philosophy of religion.)
Whitehead, A.N. (1929) *Process and Reality: An Essay in Cosmology*, corrected edn, ed. D.R. Griffin and D.W. Sherburne, New York: Free Press, 1978.(His *magnum opus*, it contains both extremely illuminating and very difficult passages.)
Whitehead, A.N. (1933) *Adventures of Ideas*, New York: Free Press, 1967.(Besides being one of Whitehead's most readable books, it provides the best insight into his overall position, including his philosophy of culture.)
Whitehead, A.N. (1938) *Modes of Thought*, New York: Free Press, 1966.(His last and - along with *The Function of Reason* - his least technical book.)

Process theism

Process theism is a twentieth-century school of theological thought that offers a nonclassical understanding of the relationship between God and the world. Classical Christian theists maintain that God created the world out of nothing and that God not only can, but does, unilaterally intervene in earthly affairs. Process theists, in contrast, maintain that God and the basic material out of which the rest of reality is composed are coeternal. Moreover, process theists believe that all actual entities always possess some degree of self-determination. God, it is held, does present to every actual entity at every moment the best available course of action. And each entity does feel some compulsion to act in accordance with this divine lure. But process theists deny that God possesses the capacity to control unilaterally the activity of any entity. Thus, what occurs in relation to every aspect of reality involving a multiplicity of entities - for example, what happens in relation to every earthly state of affairs - is always a cooperative effort.

This understanding of the God-world relationship has significant theological implications. For instance, while classical Christians must attempt to explain why God does not unilaterally intervene more frequently to prevent horrific evils, process theists face no such challenge since the God of process thought cannot unilaterally control any earthly state of affairs. On the other hand, while most classical Christians maintain that God at times unilaterally intervenes in our world primarily because divine assistance has been requested, process theists naturally deny that God can be petitioned efficaciously in this sense since they believe that God is already influencing all aspects of reality to the greatest possible extent. Moreover, while most Christian theists believe that God will at some point in time unilaterally bring our current form of existence to an end, process theists maintain that the same co-creative process now in place will continue indefinitely.

Not everyone finds the process characterization of the God-world relationship convincing or appealing. But few deny that process theism has become a significant force in modern American theology.

1 Basic metaphysical tenets

The basic metaphysical roots of process theism are found in the work of Alfred North Whitehead (§4), with much of its explicit theological framework the result of the interpretation of Whitehead offered by Charles Hartshorne. While process theists differ among themselves on some issues, there is widespread agreement concerning the basic nature of reality.

Within the process metaphysic, the most fundamental constituents of reality are not ‘things’ or enduring substances. Rather, they are units of experience - usually called actual entities - that momentarily come into existence and then immediately perish. Each of these actual entities has a physical and mental component. As the entity comes into existence, the physical component takes account of (prehends) two things: its past - all that has gone before it; and God’s initial aim - that which God sees as the best possibility open to it, given the concrete situation. The entity also automatically feels some impulse to act in accordance with God’s initial aim. However, each actual entity has at least some power of self-determination, the function of its mental component. Thus, no actual entity is ever forced to do what God wants. It always has the power to choose its own subjective aim from among God’s initial aim and all the real possibilities its past has made available. Once its decision is made, once the entity has unified the data from its perspective, it perishes as an experiencing subject and remains only as part of the past for all subsequent actual entities.

Some sets of actual entities, however, have a unity of their own. In these societies of entities, each entity still inherits its past from all other entities. But the inheritance from the past members of its own society is dominant. A subatomic particle such as an electron is a good example of a basic entity of this sort. There is no enduring substance that can be identified as an electron. But there are societies of ‘electronic entities’ in which each largely repeats the form of the previous entity in its society. In fact, the carry-over is so great that any given society of electrons appears to behave as if it were a single entity. Accordingly, basic societies or aggregates such as these are often called ‘enduring individuals’.

Moreover, basic enduring individuals of a given type normally combine into more complex enduring individuals. Societies of subatomic particles form enduring atomic individuals, which themselves combine with other enduring

atomic individuals to form enduring molecular individuals, which in turn combine with other molecular individuals. In the higher-order animal realm, for example, various societies of enduring cellular individuals combine to form enduring multicellular animal bodies with a central nervous system. In some of these enduring individuals, the enduring nervous system gives rise to what we label the enduring human mind or soul.

In short, in process thought the whole is truly greater than the sum of its parts. What we perceive phenomenally as enduring individuals - for example, humans, dogs, cats and trees - are not just bits of inanimate and animate matter in a certain configuration. Such enduring individuals are really 'societies of societies' of varying complexity - with some possessing a unity of experience (humans, dogs, cats) and others not (rocks, stars, books). Moreover, each society possessing a unity of experience retains to some extent its own autonomous power of self-determination even as it combines with other societies to create more complex societies (enduring individuals). The more complex an enduring individual of this type, the greater the influence of its mental pole, that is, the more actual options it has open to it and the more conscious it is of such options. This explains why humans behave in much more creative ways than dogs, which in turn behave in much more innovative ways than plants.

Within process thought, God is not an exception to this metaphysical order. But while there is disagreement over whether God is an enduring individual or an actual entity, God clearly occupies a unique position in the process system. Only God takes account of (prehends) all that happens in the world. God alone experiences all that every entity experiences. Thus, although each actual entity exists only momentarily, its past is not lost. All that has occurred exists everlastingly as a unified whole in God's consciousness. It is in this sense that the whole universe is included in, and penetrated by, God - that the world is God's body (see God, concepts of §8).

God, however, also responds to what is experienced. God, in this capacity, is aware of all harmonized possibilities open to the world. It is on the basis of this knowledge that God continuously presents to every entity at every moment the optimum real possibilities open to it. Thus, while God is not an exception to the metaphysical rules which govern the process system, God is certainly the most significant individual within this system. God not only exerts more influence on the direction of the process than does any other enduring individual, God alone unifies this ontological process into a harmonized whole.

2 Comparison with classical Christian theism

Process theism is frequently compared to classical Christian theism by proponents and critics alike. Moreover, the two systems are in some ways similar. Proponents of both argue, in opposition to atheism or pure pantheism, for the existence of a God who is in some sense ontologically distinct from the rest of reality. Both believe that only a being who is morally perfect is worthy of worship. However, the differences between process theism and classical Christian theism are many and significant. The most crucial concerns God's power. Classical Christians have always held that God has the capacity to control everything (see Omnipotence §1). They differ significantly on the extent to which God actually exercises this power. In fact, some classical Christians believe God seldom chooses to intervene unilaterally. But even these theists - often called free-will theists - still uniformly agree that God *could* unilaterally control all aspects of reality. The fact that God often chooses not to do so is considered a self-imposed limitation.

However, within the process system the situation is quite different. Since all actual entities always possess some degree of self-determination, God could not unilaterally control any aspect of reality involving other entities, even if God so desired. In other words, within process thought, the fact that God never imposes the divine will on other entities is not a self-limitation. Rather, it is a metaphysical reality to which even God is subject. Persuasive power, as opposed to coercive power, is all that is available to God.

Another important difference surfaces in relation to the origin of our world. Classical Christian theists have normally held that the world is totally dependent on God in the sense that it was created *ex nihilo*. Although they differ considerably on the exact manner in which they believe this was accomplished, they maintain that there was once a time (logically speaking) when only God existed and that all else has been created, directly or indirectly, by divine decree (see Creation and conservation, religious doctrine of).

Process theists, of course, do not agree. Since they believe that God cannot unilaterally control the activity of any other entity and that there has always existed a plurality of actual creative entities, they maintain that reality has always been the result of co-creative, interdependent activity. It is probably true, most process theists admit, that

there was once a time when all actual entities other than God existed in a state of random, nonpurposeful activity. Moreover, the fact that we now have a vast array of highly complex, multistructured societies of actual entities is viewed as a tribute to the effectiveness of God's persuasive power. But God alone can guarantee nothing. God has always been dependent on the cooperation of all other self-determining entities to accomplish any creative goal.

Another fundamental classical Christian belief is that the world will reach a final and fixed state, a final *eschaton*. But process theists deny that there will ever be a time when the current process will stop. They deny, for example, that there can be a fixed state in which individuals experience eternal bliss or torment (see Eschatology §1; Feminist theology §2). The universe might at some point in time arrive at a state in which all actual entities are experiencing the maximum degree of harmonious, intense novelty, the highest good from the process perspective. But there could be no assurance that such a state would be long-lasting. A period of triviality or discord, the greatest evils, could quickly follow. What does in fact occur will always depend on the manner in which all other actual entities respond to God's influence.

Finally, many classical Christians have thought it important to maintain that God is omniscient in the sense that all that has occurred or will occur, including all that we will freely choose to do in the future, is open to the divine vision (see Omniscience). But process theists, not surprisingly, deny that God possesses such insight. God does know all that has occurred. However, since all reality is interdependent in the sense that what exists at every moment includes the self-determination of entities other than God, infallible knowledge of the future is impossible for any being, including God.

3 Comparison with classical Christian theism (cont.)

The significance of these doctrinal differences becomes most apparent when considered in relation to Christian practice.

Petitionary prayer. Both classical Christians and process theists acknowledge the importance and efficacy of prayer in which God is asked to respond to concerns. Those in both camps agree, for instance, that prayer of this type can affect the petitioner as well as any others who are aware of prayers being offered on their behalf. Some in both camps even believe that those who do not know that prayers are being offered for them may experience some direct effect (see Prayer).

But for classical Christians the primary value of petitionary prayer lies in the fact that it sometimes initiates unilateral divine activity that would not occur if such activity were not freely requested. That is, such prayer is deemed important primarily because it is believed that certain states of affairs that God can and would like to bring about will occur only if divine assistance is sought. However, process theists naturally deny that petitionary prayer can ever be efficacious in this sense. Since process theists believe that all entities possess some power of self-determination, they naturally deny that a request for assistance can ever initiate *unilateral* intervention in earthly affairs. And since God is already doing all that is possible to persuade each entity at each moment to make the optimum choice, it can never be claimed that any activity on our part could induce God to become more involved than would otherwise have been the case.

Divine guidance. While it is important to note that both classical Christians and process theists share their thoughts and concerns with God in prayer, it is equally important to note that those in both camps also want God's perspective shared with them. That is, both classical Christians and process theists want to know God's will. Moreover, those in both camps believe that God not only can, but does, share the divine will with us (see Providence).

However, classical Christians differ significantly from process theists on the question of how such knowledge is attained. While classical Christians believe that the divine perspective is something that God always desires to share with us, they hold that God at times does so only if such guidance is requested (only if it is consciously sought). On the other hand, process theists, as we have seen, believe that God always automatically presents to each of us at every moment the best available option. Furthermore, while most classical Christians believe that we can, at times, come to a fairly clear, *conscious* understanding of what God believes or how God would have us act if we are open to God and diligent in our search, process theists disagree. They believe that the manner in which the divine will is shared with all entities makes clear, conscious knowledge of that which God would have us do a very rare phenomenon.

Evil. All theists continue to face the following challenge: if God is omnipotent and perfectly good, why is there so much seemingly unnecessary pain and suffering in our world? Since classical Christians believe that God can unilaterally intervene in earthly affairs, they must maintain that God has decided to allow each instance of evil because such evil is either a necessary antecedent causal condition for the actualization of some creative goal (for example, God's desire that we learn patience) or an unavoidable by-product of some creative goal (for example, God's desire that we exercise freedom of choice). In short, classical Christians must maintain that God is powerful enough to rid the world of all evil, but possesses morally sufficient reasons for not doing so (see Evil, problem of).

Process theists, however, find this line of reasoning totally implausible. As they see it, a perfectly good being who could do more to rid the world of evil would surely do so. But the God of process theism, they are quick to point out, is not faced with this choice. God does deplore evil and is always attempting to persuade all other entities to help actualize the best possible states of affairs. For instance, God was horrified by the Holocaust and tried in every way possible to influence those responsible not to commit such atrocities. But since God cannot unilaterally control the self-determining activity of any other entity, God is never in a position to prevent any instance of evil such activity might produce. And therefore the fact that evil states of affairs arise (and sometimes even proliferate) does not count against God's goodness.

4 Critical discussion

Critics of process thought fall into three major categories. Some reject it because they believe its theological system to be incompatible with classical Christian doctrine. Pinnock (1987) has argued, for instance, that process theism must be rejected because process theists do not acknowledge the total transcendence of God, the full divinity of Jesus or the supernatural origin of Scripture. However, while this type of criticism may serve a useful function within the classical Christian system itself - namely, to dissuade those who may initially find the process system appealing - its value in the general debate over the adequacy of process thought is questionable. Critical contentions of this sort do make it clear that classical Christians differ from process theists on important points; but the reality of such differences is something process theists readily acknowledge.

Other critics argue that process theism must be rejected because some of its basic contentions are false or implausible. For example, Gruenler (1983) argues that the basic process metaphysic is incompatible with modern relativity theory. Others find the process explication of how the world as we know it came to be quite implausible. An even more common criticism of this type, though, is the contention by classical Christian theists that a being with the acknowledged limitations in power of the God of process theism is too weak to be considered worthy of worship.

In response to this latter claim, process theists maintain (as noted before) that a God who is worthy of worship must be perfectly good, but argue that only the God of process theism meets this requirement. The powerful, independent God of classical Christian theism, they grant, might be viewed as a convenient wish-fulfiller. But such a being cannot justifiably be viewed as a caring friend worthy of our loving respect. Moreover, they continue, given the amount of unnecessary evil in our world, only a God who could not do more to remove it can be considered morally praiseworthy (see Goodness, perfect).

Finally, some critics challenge the internal consistency of the process system. One of the most significant of these challenges centres around the question of whether the God of process theism would coerce (unilaterally control) other entities if this were possible (Basinger 1988). If process theists decide that God would do so, it then becomes difficult to see how they can justifiably maintain (as many do) that coercion is morally incompatible with divine perfection (as Ford 1978 says), or that persuasion is the greatest of all powers, the only power capable of worthwhile results (as Cobb 1976 says). On the other hand, if process theists maintain that God would not coerce even if possible, it then becomes difficult to understand how process theists can justifiably criticize the God of classical free-will theism for not coercing more often. Process theists, the critics in question allege, cannot have it both ways. In response, some process theists (for example, Griffin 1991) acknowledge that the God of process theism would coerce if possible, but deny that this is inconsistent with continuing to maintain that persuasive power, if properly understood, is morally superior to coercive power.

It is, of course, too early to assess the lasting impact of process thought. But this unique metaphysical system is clearly a significant force in American theology, and there is little reason to believe that it will not continue to

offer an important alternative to classical Christian theism for some time to come.

See also: Natural theology; Postmodern theology; Process philosophy

DAVID BASINGER

References and further reading

Basinger, D. (1988) *Divine Power in Process Theism: A Philosophical Critique*, Albany, NY: State University of New York Press.(An accessible challenge to the self-consistency of process thought.)

Cobb, J.B., Jr (1969) *God and the World*, Philadelphia, PA: Westminster Press.(An introduction to process thought.)

Cobb, J.B., Jr and Griffin, D.R. (1976) *Process Theology: An Introductory Exposition*, Philadelphia, PA: Westminster Press.(A very accessible introduction to process thought.)

Ford, L. (1978) *The Lure of God*, Philadelphia, PA: Fortress Press.(An accessible, sympathetic explication of process thought.)

Griffin, D.R. (1991) *Evil Revisited: Responses and Reconsiderations*, Albany, NY: State University of New York Press.(A rigorous philosophical defence of process thought.)

Gruenler, R. (1983) *The Inexhaustible God*, Grand Rapids, MI: Baker Book House. (An accessible evangelical critique of process thought.)

Hartshorne, C. (1948) *The Divine Relativity: A Social Conception of God*, New Haven, CT: Yale University Press.(A good introduction to Hartshorne's understanding of the interdependent relationship between God and the world.)

Neville, R. (1980) *Creativity and God*, New York: Seabury Press.(A rigorous metaphysical critique of process thought.)

Palin, D.A. (1994) *Probing the Foundations: A Study in Theistic Reconstruction*, Kampen: Kok Pharos.(An accessible introduction to, and defence of, process thought.)

Pinnock, C. (1987) 'Between Classical and Process Theism', in R. Nash (ed.) *Process Theology*, Grand Rapids, MI: Baker Book House.(An accessible critique of process theism.)

Whitehead, A.N. (1929) *Process and Reality*, New York: Macmillan.(Whitehead's most important work. Very difficult.)

Processes

A process is a course of change with a direction and internal order, where one stage leads on to the next. Processes can be physical (such as atomic decay), biological (such as the growth of living things), artificial (such as building a house) and social (such as carrying out a criminal investigation). Much of what is said about processes can be said about sequences of events. The concept of event, however, suggests a separate occurrence, whereas that of a process suggests something which is ongoing. There are matters, such as development in organisms, where to see what is happening as part of a process has an advantage over thinking of it as an event. Causes are generally spoken of as events, but the more dynamic concept of causal processes may get nearer to expressing the transition between cause and effect. Moreover, to explain something as a stage in a process can take account not only of what has happened in the past, but of what might happen in the future. This may (but need not) involve purpose; with organisms it involves development through functionally interrelated activities. In some social processes there can be a practical, moral significance in seeing a situation as a stage in a process, since this can encourage us to look to a further stage where something constructive might be brought out of what could otherwise be seen as simply an untoward event or an unhappy situation.

1 General nature and types of processes

Processes are systematic courses of change with an internal order and temporal direction. Their temporal direction consists in there being an absolute order in the stages by which they proceed. This need not imply that there is an absolute time apart from processes or events, and these might appear in a different order to observers in different frames of reference. But the order in which the stages of a process occur is irreversible. Most processes go towards completion at an end point, but this need not be so in all cases. There are also steady-state processes, periodic processes and staccato processes where the ongoing feature is one of sustaining a pattern.

A process has temporal parts. This is sometimes expressed by saying it ‘perdures’ - that is, it is not wholly present at any time. Substances which are wholly present at every time of their existence are said to ‘endure’. Processes mostly involve physical movement. Some, however, have an internal order not reducible to the direction of any movements that may accompany them. For instance, an intellectual process of argument proceeds through the participants seeing that one statement follows from another, or that it should be abandoned because of critical considerations raised in the course of the argument. An argument need not reach a conclusion - it can be broken off and resumed.

While all processes have an internal order and a direction, there are different kinds which, besides intellectual processes, include natural, artificial and social ones. Natural processes follow laws, as with the chemical reactions which are involved in the metabolism of food. Biological natural processes are clearly irreversible - digested food cannot be turned back into parts of plants and animals. However, some physical processes are said to be reversible, as when a rotating wheel is ‘put into reverse’. But here another process is set going whose stages run in the opposite direction, towards a state which is not identical to the originator of the process but is another of the same type.

Artificial processes are carried out teleologically by agents acting in natural conditions set by the available materials and the environment. Within these limits, the agents can sometimes vary the order in which what they do becomes a stage in a process. For instance, in constructing a house the foundations must be laid first and normally the walls built before the roof is put on. But the roof can be put on before the walls are built provided that there is a sufficient structure to support it. Natural processes can be interfered with through artificial processes, as in conducting experiments. The constituents of social processes are persons acting according to rules and conventions in environments set by natural and artificial processes. The upholding and administration of the Law is a collection of social processes with rules and conventions prescribing proper procedures (note that ‘process’ and ‘procedure’ have the same root). A legal trial is a social process - indeed ‘process’ is a technical term for legal proceedings.

Rules prescribe how certain social processes ought to be carried out. Customs and conventions show how a number of social processes are motivated by expectations as to how people are likely to behave according to the ways in which they have been socialized. The formal processes of law and government operate in a social

environment sustained by these informal processes. Generalizations about social processes (and also some natural ones) are often statistical (see Social laws).

2 Causation in processes

Artificial and social processes involve teleological causation where agents act purposively, but depend on efficient causation for the results of their actions (see Teleology). Natural processes in accordance with laws are not thought to involve teleology, though this is controversial in the case of some biological processes. In considering these, there is a distinction between causation which is internal to the process and that which consists of action on it from outside. W.E. Johnson (1934) used the term ‘immanent causation’ for the systematic changes due to co-variable and coordinated functioning in the process taken as a whole, and ‘transeunt causation’ for the action of separate elements on one another. Science aims to discover efficient causation between the elements of a system (for instance, the chemical actions in the process of metabolism). But the concept of immanent causation may still be needed for describing its functioning as a whole. The nourishing of the body is carried out through a great number of mutually supporting processes - the beating of the heart circulates the blood, and the heart will fail to beat unless the blood supply reaches it. There are also feed-back processes which come into effect to restore imbalances (for instance in keeping the body temperature within limits). These have been given the name ‘homeostasis’. They may not be teleological in the purposive sense, but are instances of immanent causation. Coordinated functioning in a changing system is the general character of what are called ‘organic processes’, as shown in the self-maintenance and development of living things.

3 Processes and events

Much of what is said philosophically about processes can also be said about events - for example, they have an internal order or direction, and an absolute order of before or after in their parts. What is pertinent here is how processes differ, if they do, from sequences of events. An event is an occurrence taken as a unit. It can be divided into temporal parts, each of which can then be taken as a shorter event. So the great storm in Britain in October 1989 can be taken as an event. It can be said that it was more severe at some times than at others, distinguishing its temporal parts as sub-events, such as the blowing of the gale from 1 a.m. to 1.15 a.m. It can then be said that the blowing of the gale at 1.10 a.m. caused the collapse of Mr Jones’ chimney (another event). Causation is often said to be a relation between events. It can also be spoken of in terms of processes: as the gale went on blowing even more fiercely, it caused the waves to go on rising higher and to flood the sea front. The event-form gives a sequence of occurrences; the process-form suggests something was going on. In this respect, to speak in terms of causal processes may come nearer to a dynamic description of what was happening.

Events are often expressed by participles forming gerunds (sometimes called verb nominalizations), which are parts of verbs construed as nouns and prefaced by ‘the’ or ‘a’. In the case of processes, the participle keeps its full verbal voice, perhaps in a continuous present. So in the phrase ‘the striking of the clock woke me up’, ‘the striking of the clock’ refers to an event, whereas in the phrase ‘I woke up while the clock was striking’, ‘while the clock was striking’ refers to a process.

4 Some metaphysical views

Some metaphysical views see reality as essentially in process - becoming, changing, passing away - or even claim that processes are all that exists. In that case there might be a philosophical description of the world expressed through verbs and adverbs. There would then be no things or persons, and this may raise a problem concerning agency. The subject of a proposition could be a dummy ‘it’; ‘Socrates is wise’ would become ‘It Socratizes wisely’ or ‘Socratizing wisely is going on’. Among the metaphysical religions, Buddhism has come nearest to such a view: in Buddhism there are no agents or substances, and the self is a stream of passing states. These are linked, however, in a causal process, or Karma (see Buddhist philosophy in India §1; Causation, Indian theories of).

A world of pure processes seems to need some structuring principle. The father of Western philosophies of process was Heraclitus. His philosophy has come to us in cryptic sayings, notably that ‘all things flow’ (*panta rei*). Yet this flux contains a principle, the Logos, which structures the flux by tensions of opposing forces, never coming to rest in a stable equilibrium. The contrasting position was taken by Parmenides, who so pressed the assertions that what is, is, and what is not, is not, as to deny the reality of change. This aroused serious discussion through the

arguments produced by Zeno of Elea. If everything not only is what it is, but is where it is, how does a moving body such as an arrow get from one position to another?

The problem of whether motion is a process of transition and not only occupancy of a succession of positions appears in a modern form in the so-called 'at-at' view given by Bertrand Russell. Motion is defined as *merely* (his italics) 'the occupation of different places at different times, subject to continuity' (1903: 473). Continuity here is density, and if space and time are dense continua this means that between any two positions or instants there will always be another. The 'at-at' view of motion states that a moving body will always be at the position appropriate to each instant. This, however, may not adequately meet Zeno's problem, which concerns motion as transition from one position to another. There is an analogous problem with the concept of change, defined as something having different properties at different times. This is perfectly true, but it loses the connotation of *changing* as being undergone by a subject.

The metaphysics of Aristotle contains a view of processes of becoming put in terms of the distinction between potentiality and actuality: things develop their potentiality towards full realization of their natures. This view of things as processes of becoming is circumscribed by his conception of a teleological development towards fixed essential natures. In spite of this limitation, Aristotle produced a subtle analytic treatment of the idea of a process of change. The word for this was *kinēsis*, as distinct from *metabolé* which was his general word for change. Our word 'kinetic' suggests movement, and *kinēsis* is best translated as 'process of change'. *Kinēsis* always contains a direction, a 'whence-whither' (*pothen poi*), but this need not be the direction of a movement in space; it can be the transition from one stage to another of a process of development. As a process of change *kinēsis* is continuous, though it may be broken off or interrupted.

Among other great philosophers, Hegel had a metaphysical system in which process played a central part (see Hegel, G.W.F. §4). Notably, world history was said to develop in a dialectical pattern analogous to a process of thought, where each stage expressed a partial aspect of an ultimate idea of fully conscious rationality. Marx re-interpreted the dialectical process not as a movement towards rationality, but as a movement of social change, where oppositions and tensions arise through conflicts between classes, at particular stages of technological development. The end state is conceived not as ideal rationality, but as a form of society in which classes, and thereby the conflicts which had driven the dialectical process, would be abolished (see Dialectical materialism). Views such as these have been castigated by Karl Popper (1957) as 'historicism', meaning by this that they impose a pattern on history. Nevertheless, such views have been constructive in encouraging the study of history as a complex of processes arising not only between individuals but out of relations between social institutions.

A philosophy of process which had a considerable influence in literature, for instance on Proust, was that of Bergson (see Bergson, H.-L. §4). Bergson saw reality as an open-ended evolutionary process, 'creative', as producing new forms of life. The word 'life' is in place here, because Bergson thought of evolution as driven by a living force, *élan vital*. Biological evolution is an integral part of this, in which we ourselves are participants. It is a temporal process in a strong sense - indeed Bergson spoke of reality as *durée*, real duration, continually creative of what is new. This is contrasted with time as measured by intellectual devices which represent a moving and changing reality as if it were a succession of states arrested at successive instants. Bergson held that the intellect was committed to this 'cinematographical' approach, and therefore failed to capture the moving and changing reality. There is, however, another mental capacity - 'intuition', through which we have an immediate experience of this reality in our own consciousness.

The later philosophy of Whitehead gives a metaphysics of a pluralistic universe, whose constituents are 'processes of becoming' (see Whitehead, A.N. §4). These are structured unities developing through their interrelations with each other. He used the generalized concept 'organism' for these unities. Their processes of becoming display reiterated patterns which can be given abstract formulations, especially through mathematics. Such intellectual formulations need not be considered distortions - Whitehead held that the world has a permanent basic structure which underlies them.

Whitehead's later philosophy is obscure, and it must be acknowledged that none of these metaphysical views, with the exception of that of Aristotle within his own terms, gives a precise critical discussion of what it is to be a process. Yet a number of matters such as growth, development, learning and producing creative work, are naturally thought of as processes, where understanding of what goes on at one stage is furthered by having regard to future

as well as to past stages.

See also: Causation; Change; Events; Process philosophy; Process Theism

DOROTHY EMMET

References and further reading

Aristotle (*c.* mid 4th century BC) *Physics*, III-IV, trans. and ed. E.H. Hussey, Oxford, Clarendon Press, 1983.(The main source for *kinēsis*, referred to in §4 above.)

Bergson, H. (1907) *Évolution Créatrice (Creative Evolution)*, trans. A. Mitchell, London: Macmillan, 1911.(The main source for the view outlined above in §4.)

Cannon, W.B. (1932) *The Wisdom of The Body*, Cambridge, MA, and London: Harvard University Press.(The original source for the concept of homeostasis mentioned in §2 above.)

Emmet, D. (1992) *The Passage of Nature*, London: Macmillan; Philadelphia, PA: Temple University.(A fuller treatment of the material in this entry.)

Hegel, G.W.F. (1832-45) *Lectures on the Philosophy of World History: Introduction: Reason in History*, trans. H.B. Nisbet, Cambridge: Cambridge University Press, 1975.(Especially relevant for the view of dialectical processes described in §4 above.)

Holz, H. and Wolf-Gazo, E. (eds) (1984) *Whitehead und der Prozeßbegriff*: *Whitehead and the Idea of Process*, Freiburg/Munich: Alber.(A collection of papers given to the First International Whitehead Symposium - English summaries of German and German of English papers.)

Johnson, W.E. (1934) *Logic*, Cambridge: Cambridge University Press.(Pt III, ch. 6, outlines the view of immanent causation referred to above in §2.)

Munsat, S. (1969) 'What is a Process?', *American Philosophical Quarterly* 6 (1): 69-87.(Differs from the view expressed in this entry in holding that all processes have an end state and that natural processes proceed in a way analogous to following a rule.)

Popper, K.R. (1957) *The Poverty of Historicism*, London: Routledge & Kegan Paul. (Popper's critique of theories of history, referred to in §4 above.)

Process Studies, School of Theology, Claremont, CA.(A journal largely concerned with views deriving from the later Whitehead. A number of articles are on the philosophy of religion.)

Russell, B.A.W. (1903) *The Principles of Mathematics*, 2nd edn, London: Allen & Unwin, 1937.(Outlines the view of motion referred to in §4 above.)

Sellars, W. (1981) 'Naturalism and Process', *The Monist* 64 (1): 37-65.(This takes up a suggestion in vol. 1 of C.D. Broad's *An Examination of McTaggart's Philosophy* (Cambridge: Cambridge University Press, 1935) that there could be a view of the world solely in terms of processes.)

Whitehead, A.N. (1929) *Process and Reality*, ed. D.R. Griffiths and D.W. Shelbourne, New York: Free Press; London: Collier Macmillan, 1978.(Outlines the metaphysics of processes of becoming referred to in §4 above. Other important works by Whitehead include *Science and the Modern World* (1926), *Adventures of Ideas* (1933) and *Modes of Thought* (1938), all published by Cambridge University Press.)

Proclus (*c.* AD 411-85)

The Greek Neoplatonist Proclus aimed to find a logical and metaphysical structure in which unity embraces but does not stifle diversity. He assumed the underlying unity of reality and of self, but was anxious to maintain the diversity of thought and existence. This led him to conceive of things as different species of one general whole, each part comprehending the rest but in its own specific, limited way. There are successive levels of awareness, thought and existence, ranging from that of ordinary experience, where we continuously have a passing understanding of the equally fleeting world, to that of ultimate unity, where we see the first principles and the total whole unqualified, as if by super-intuition. Reason, aided by imagination, and exercised by philosophy and science, elevates us towards that supreme state, which is at once the foundation of religious and of ethical values.

Proclus was interested in an integrated account of the nature of things. Asking questions about how and what we know, our perceptions and beliefs, prompts questions of the sort: what is the origin of knowledge? What is the nature of mind? Of the things we think and perceive? Of existence? For Proclus even questions about virtue, moral judgment and action, God, faith and salvation are all clarified by referring them to questions about their origins and nature. No subject escaped his attention, including the interpretation of poetic works by literary figures such as Homer, where Proclus saw language as a mediator to deeper truths. Philosophical inquiry leads to asking what order of reality substantiates things, whether in the field of mind, values, science or literature.

Proclus' system is complex, and he uses highly technical terms (most of which have their roots in Plato, Aristotle and Plotinus). But for him this is only appropriate to the richness of our conceptual world. We have complex and varied concepts about our surroundings and selves, not because humans necessarily have flamboyant fancies but because reality itself is complex. That we have glimpses of knowledge means that there is some unity between our mind and the objects of thought. Moreover, unity is essential to the identity of things, and without it they would be unintelligible, and conceptually unreal. The 'One' is a primitive absolute, and is fundamental to intelligibility and existence. Thinker, thoughts and realities are in some way one. Things are not disconnected but share layers of ever-increasing unity. Consequently questions about different kinds of being, knowledge, good, and so on, become questions about degrees.

Studying all this was not just a cerebral affair. To the Neoplatonist, understanding the scheme of things provided a guide on leading a good life and achieving what since Plato's days had been praised as the goal of human endeavour, 'true happiness' or eudaimonia*.*

1 Life

In his fifty-year career Proclus became the chief exponent of systematic Neoplatonism, the dominant philosophy of late antiquity. He was born in Byzantium, the capital of the eastern Roman Empire, but grew up in Lycia. The son of a successful advocate at the imperial court, he initially went to study law at Alexandria. However, while on a visit to Byzantium (associated with the foundation of a new university) he underwent a conversion to philosophy, which eventually led him to Athens at the age of nineteen. There Plutarch of Athens and Syrianus taught him Aristotle and Plato according to the Neoplatonic curriculum. He quickly rose to become the head of the prestigious Athenian School at the young age of twenty-five, until his death aged around seventy-five.

Proclus sought to practise what he taught. Biographical sketches and his poems reveal a man of passion. This he channelled into his academic work (teaching, seminars, writing), religious devotion and rigorous lifestyle (he was a vegetarian and abstemious). Friends, students and staff benefited regularly from his love and generosity, which compensated for his famous short temper and perfectionism. He was an inspiring figure and exercised his considerable authority politically both in and outside Athens. A proud defender of the 'Hellenic' values and religious multiformity, Proclus lived when Christianity became the state orthodoxy.

He was a prolific author, and we have evidence that he wrote more than forty titles. In keeping with the practice in the imperial Roman period, many of Proclus' treatises are extensive commentaries on classical greats. This does not mean they are not original: novel ideas and modifications or rejections of classical doctrines can be found in these commentaries, and in content there is little to distinguish them from monographs. For his general system the

main sources are the *Commentary on Plato's Parmenides*, *Platonic Theology* (synonymous with metaphysics), the study manual *Elements of Theology*, and for additional detail, the *Commentary on Plato's Timaeus*.

2 Metaphysics

Proclus' overall metaphysical standpoint he shares with his fellow Neoplatonists. He subscribes to realism, in that he accepts a reality independent of what we individuals think about it. But proper reality is not physical, because what we grasp with the five senses is episodic and often illusory. The things that exist are of the intellect - with the provision that such intellect and its ideas are not just personal but objective and universal (this can be traced to the revision of Aristotelian *nous* by the influential Alexander of Aphrodisias (§2)).

In his response to the problem of unity/diversity Proclus wants to steer between conceiving things as disparate and lumping them into an undifferentiated heap. He employs a rule of mixture that originally belonged to the theory of material stuffs in Anaxagoras (§3), but which near Plotinus' time had been extended to the domain of intellect and concepts: 'Everything is in everything, in a manner proper to each.' If things (conceptual and material) are ultimately one, they cannot have sharp boundaries to distinguish them, but must extend in some way 'all within all'. Similarity (and dissimilarity) supersede strict identity and difference (*Platonic Theology* III 7 S and W, VI 347-50 Portus; *Elements of Theology* proposition 108).

This metaphysical reason combines with the exhaustion of possible logical options under a heading *x* (that is to say, *x*; *x* and not-*x*; not-*x*) to yield the chains of 'intermediaries' for which Proclus is well known. Any two polar (directly opposite) terms have one or more intermediary, which is more or less similar to either pole. Although threefold leitmotifs are pre-eminent (for examples see below on participation), in fact as many intermediaries can be inserted as required (see Many-valued logics). However Proclus does not accept that everything we can think merits a niche in reality: only those that do not depend on a particular mind (*Commentary on Plato's Parmenides* col. 1054, 895-7) but can already stand on their own 'foundation' (*hypostasis*) (henceforth, some of these 'hypostatic' terms are indicated by capitalization). The task of the philosopher is to discover the true ones.

To make sense of diversity we group things with a common characteristic. But this raises the questions what is the status of such characteristic and what is its relation to the group and parts? Proclus distinguishes three senses of 'whole', and recognizes the full implication of saying that a whole is the sum of parts plus the cause of its unity (see also Aristotle, *Metaphysics* 1041b, and, especially, Plato, *Theaetetus* 204-5). The essential whole is a 'monad', a singularity that cannot be analysed: it is a 'whole prior to the parts' (*Elements of Theology* propositions 67-9). The characteristic group-type in its pure state is such an unanalysed whole. It stands as a prototype that qualifies and embraces all possible specific forms of it. The many parts of the whole express the prototype but unequally, because they have it in different conditions (for example, light itself, the light of the sun, firelight, glow-worm light). An attribute can thus be regarded as a family, a series, headed by its own prototype (for example, the monad of life), followed by the multitude of diverse forms where the attribute is found (for example, life of animal, life of plant).

In Platonic philosophy the relation of a 'Form' to its material particulars was often described as 'participation', a sharing-in, because each particular cannot be said to possess the entire Form. For the late Neoplatonist, participation extends to relations between conceptual entities also, so Proclus talks in general of what is 'participated in' and its 'participant'. Participation created many problems, not least, how can the Form be shared out and still maintain its integrity? By the whole-part theorem, the whole in itself is indeed a single thing, and is exempt from direct distribution into parts: it is 'imparticipable'. Thus, we arrive at the threesome imparticipable-participated-participant. Because Proclus is dealing with a metaphysical order where the attribute is more perfect and real than its subject (which is defined, and thus perfected, by it), he further distinguishes between a participated attribute that is complete and reified in itself (called 'self-hypostasized'), and a participated that must be always within a participant to accomplish this (such attributes are called 'image' and 'trace'). Between the fully transcendent ('imparticipable') and the fully immanent stands the sovereign immanent. This intermediary is the linchpin of Proclus' account. It refers to objective realities (for example, cosmic bodies, certain universal forces and properties, spirits) that are within reach of our apprehension but independent of us.

The constant 'flow' from unity to plurality, Proclus calls 'procession'. It leaves the source 'remaining' undiminished. The thing that 'proceeds' has its character diluted and altered by the conditions it is found in (for

example, Life itself becomes life of intellect, or life of plants). However, what ‘descends’ into plurality never severs its connection with its original, pure state, otherwise it would completely lose its definition (that is, the life of plant is still a life, not a plant-part). ‘Reversion’ is the tendency or movement to regain the lost pure definition (which is ‘remaining’ unaltered). These three are not discrete static states, but three distinguishable moments of a single dynamic process; practically everything undergoes all of them (with the One in a controversial position). For Proclus the three apply to every form, property or entity, not just to the conscious soul or cosmic strata of being (see Plotinus §3).

3 Levels

Things are composed of layers of qualities from corresponding degrees of reality. The various degrees are also modes because they modify any form, property or entity ‘in the manner proper to each’. (This is the philosophical reason for the high number of peculiar adverbials in the vocabulary of Proclus and later Neoplatonists.) The grades are bunched into broad levels (in addition to unity), typically: real existence, life, intellect, soul, physicality and body. (Principal classical sources are Plato’s *Sophist*, *Philebus*, *Parmenides*, *Timaeus*; see Plato §§15-16.) Each level has its ‘imparticipable monad’, followed by the multitude that is participated in: for example, Intellect and intellects. Moving from one level to another involves a substantive change, a ‘decrease’ (*hyphesis*) or an ‘elevation’ (*anagōgē*). Properties accumulate successively: real existence has unity, life has real existence and unity, intellect has life, real existence, and unity, and so on. So the ‘elevation’ also brings an analytical reduction to first principles.

Proclus arranges the levels of being by degree of perfection (completion) and by generality (*Platonic Theology* III 20-6), which is an inversion of Aristotle’s scheme of values. Soul is superior to body because it perfects it. Intellect perfects the soul, but is also more general because even animals have ‘a trace of cognition’. Life is superior to intellect because it is more general (even plants have it) and there cannot be a lifeless intellect. Being is, likewise, superior again because even inanimates have it. The rule - first by dint of being more general, derives from Proclus’ whole-part theorem (see also §6).

4 The One, the Good and the divine

The greatest unity, the ‘One’, transcends every possible attribution positive or negative, and so cannot be understood directly. It even transcends the very state of existence (see Plato, *Republic* 509b), so Proclus calls it ‘not being’ (*Elements of Theology* proposition 138), in the sense of prior to existent. This absolute metaphysical state coincides with that of value, since unqualified unity is the ultimate perfection (completion) desired by all. The One is identical to the Good and, as the supreme perfection, value and cause, is God. As fully imparticipable, God is transcendent, unknowable and ineffable. As the unity essential to everything that exists, the divine spark is immanent even in the lowliest material (see *Elements of Theology* proposition 145). However, Proclus carefully distinguishes divinity proper from deification by participation (for example, a divine body). Properly divine are only the One and the ‘self-hypostasized’ unities (see the discussion on *henads* given below), and by extension those entities predicated by such a unity (*Elements of Theology* proposition 114). The unknowable, supra-rational divine can be approached intellectually by stripping away attributions (known in Christian theology as the apophatic way to God), or by analogy, inferring what the One is like by its known effects. Ultimately individuals can reach the One by placing their faith in it with their own immanent ‘one’ during theurgical and mystical acts (see Chaldaean Oracles; Iamblichus).

But such extreme transcendence is almost self-defeating. How can the One and Good relate at all to existence? Proclus rejected Iamblichus’ two Ones (the fully transcendent and the causal). This still left the two fundamental problems (parallel to those faced, say, by modern big-bang cosmologists): How is existence produced from pre-existence? How does diversity arise from absolute simplicity? In answer to the first Proclus emphasizes the reified principles Limit and Unlimited (and Providence). In the second, he distinguishes the plurality of ‘ones’, the ‘henads’.

The One is essentially a limit, for there is nothing else except it. This also means that the One has no boundary imposed upon it (by something else): it is unlimited in potency. Limit and the Unlimited become the starting points of the chain of causation that produces the levels of being. Limit supplies definition and discreteness, and the Unlimited the ‘overflowing’ capacity to exist in continuity. The third factor is the providential activity of the One,

which reaches down to individual beings and perfects them with unity. Moreover, if the One is the prototype of oneness, then it must be the head of many participated 'ones', the 'henads'. A 'henad' is the very unity at the core of every existent, and there are as many henads as things that exist. Strictly the henads originate in the Limit, Unlimited and Providence, and so carry the seeds of differentiation at the root of the individual variety of things. Proclus distinguishes further the 'self-hypostasized henads' which belong to timeless or everlasting entities, the sort that tend to be worshipped as gods by diverse religions. The plain henads are the ones immanent in humans, animals, plants, minerals, and so on.

5 Intelligibles and soul

The highest level thought can reach is real existence, or pure being, the first broad category of the intelligibles. This is the object of thought (*noēton*) (the One is beyond thought). Here the contents of the One acquire their first layer of manifestation, in that they become 'really existent' and accessible to contemplation. Appropriately Proclus considers this to be the place of the universal pattern, the 'paradigm' (Plato, *Timaeus* 31a) and of eternity. Next he finds a middle intelligible level proper to pure Life, long before it is embodied in living beings. This stands for the capacity to multiply the contents of the whole. At the last intelligible level, the intellective (*noeron*), is Intellect itself (*nous*) (see Nous). From here and above we conceive things directly, as by intuition. The Intellect's essence is 'pure mind' (as embodied in the god Cronos in Plato, *Cratylus* 396b), whose contents are made distinct by the power of Intellect, identified with Rhea (*Cratylus* 402) and the 'Goddess' of the Chaldaean Oracles. The active exercise of Intellect results in the creative conception of things (*poiētikon*) (Plato, *Timaeus* 28c; Aristotle, *On the Soul* 430a12). Creative Intellect is God the Maker (Plato, *Timaeus* 29d-30c), who gives determinate form to the physical world. The Creator is well distinguished from the One - in sharp contrast with Christian doctrine. Intellect has polar outlooks: one contemplates the higher intelligibles and has intellective ideas and forms; the other is involved in producing time, soul and the physical things of ordinary experience.

Soul is what makes a body alive and 'is a substance between the really existent and becoming' (*Commentary on Plato's Timaeus* III 254.13-17). In the Platonic tradition soul straddles two domains: one is intelligible and beyond the bound of time; the other is physical and constrained by body, space and time. Proclus thus defines the soul's essence as atemporal but its activity as operating in time, because it cannot actualize all of its contents at once but has to unroll them onto the passing time. He distinguishes the 'imparticipable monad' of soul, which lies 'above the (physical) world' (*hyperkosmios*) and is associated with no body. The traditional Platonic world-soul (what moves the cosmos in the determinate way measured by science), although just as singular, now becomes a participated soul, whose body is the whole corporeal universe.

6 Body, matter and the One

The qualities characteristic of a body are said to be its nature (*physis*). In living beings nature is the instinctive, non-rational aspect of life, which is inseparable from the functions of body. It can be distinguished from soul proper (which is purposeful and separable) and from pure body (which is itself passive). However such distinctions are not always needed and so nature is often grouped with soul or body. An 'imparticipable' Nature is paradoxical (How can it be apart from a body?), which may explain why Proclus does not explicitly attribute to it the 'henads above and in the (physical) world' (*Platonic Theology* VI); although the link can be confirmed from diverse cross-references. Nature has the Necessity which determines physical behaviour. Indeed Nature's activity creates bodies directly, and seems also to be the monadic source of Body: there is no imparticipable body.

Body considered in itself is the last of things that exist in some sense. It is completely inert. Living beings move because of their vitality and inanimates according to their nature. For Proclus pure body is a quantity of a certain shape with three-dimensional extension. Matter, on the other hand, is not even a definite being, just the potential for something to be what it is.

At the low end of the metaphysical scale, entities become progressively less compound. There are living beings without a spark of intellect (plants), inanimates lack life, and matter lacks even determinate being. This is a result of the 'Proclean canon' that the more general and perfect the cause is, the further its extent of power. At the lower ranks, things receive attributes solely of the more general kind, not general *and* specific ones. Properties compound until the maximum number in the bundle is reached around the level of appearances that humans inhabit. From there, specific properties gradually diminish. So we arrive at simplicity in two different ways: in the superior sense

is the One; in the inferior is matter.

For Proclus this makes matter one continuous, universal power, and a direct extension of the One's Unlimited (in contrast to Plotinus' bisection of matter into intelligible and sensible). Matter is indeed to a degree good and of value. Like the One, it is most obscure and formless: the One is prior to intelligible definition, and matter beyond its reach. This does not mean, however, that Proclus considered material objects a short-cut to the One. 'Reversion' to the One is an 'elevation' through levels of being in the maximal sense, those with perfected properties.

7 Evil?

Proclus rejects completely the existence of evil as something absolute. He cannot find it in any of the levels of being, not even in matter which 'is in some way good'. Everything, including matter, has its roots in the One, which is Good. If there were a source for evil (Plutarch of Chaeronea (§3) blamed an evil world-soul, Gnostics the Maker, and Christians the devil) it would have to be 'beyond even the total lack of existence…further than the nothingness of non-existence' (*Commentary on Plato's Timaeus* I 374.14-17). What is commonly described as evil points to a relative weakness of the good: for example, when something is unnatural, or when it is bad for its purpose or deficient in some way (I 375, 381); in human actions, when the perpetrators are ignorant of what is best, and have a weak mind and soul (*On Evil* 50, 40-6).

Since evil is a parasitic and fictitious existence (*parhypostasis*) (*Platonic Theology* I 84-5), the only real choice open is the pursuit of the Good. This we fail to attain because of ignorance, inadequacies, selfish passions and a host of other restrictions. Freedom of action, therefore, means primarily to be free of such obstacles as inhibit or distort an individual's inherent desire to pursue the good.

8 Psychology

Psychē (usually translated as soul) is an alive, cognitive substance, as individual as the creature that has it. 'Mind' is a useful translation when contrasted with the perfect 'intellect' (*nous*), which stands above the partialities of soul. Every soul can cause spontaneous movement. The differences lie mainly in the way such movement is prompted and maintained by the soul's mental faculties. Animal *psychē* is simple. It is non-rational only, and concerned with sense-perception and appetites, although it still has a spark of intelligence. Human soul mixes rational and non-rational aspects (not parts because the soul itself is undivided). Thus, humans can lead a wide range of lives depending on how much they allow one aspect to control the other. Reflecting the complexity of its constitution, the human mind/soul has many faculties or 'powers'.

The non-rational side of the soul deals with input from the physical world. With sense-perception (*aisthēsis*) the soul receives the sensible qualities of material objects through the sensory organs. The diverse impressions, 'affections', are first organized into a unified impression. Next, with its opinion-forming faculty (*doxa*), the soul produces basic beliefs from the sensory evidence and any judgment about the source of these sense-impressions. Finally it visualizes things internally with its image-making faculty (*phantasia*).

The rational side of the soul deals with the outpourings from the intellect. Using its highest faculty, called *logos* (a word with many meanings (see Logos), but in this context signifying reason, definition and modelling), the individual mind has access to the 'inexhaustible supply' of ideas and inspirations. But being particular, the human mind admits them fragmented (not as intellective wholes). Soul has also its own concepts: '*psychē* was never a *tabula rasa* but is a tablet that has always been inscribed and is always writing itself and being written on by *nous*' (*Commentary on the First Book of Euclid's Elements* 16.8-10). Understanding is the result of the mental debate between successive arguments and propositions: that is, it is the result of 'discursive' reasoning. To compare the various concepts, the rational soul seems to employ its own imagination (*phantasia*), and sees representations of concepts as projections on the mind's screen from internal and external sources. By matching representations soul can complete or correct the impressions from the senses.

9 Soul and body

The individual soul, being particular, is incomplete and imperfect: it knows this and desires other things to complete it. Half-knowledge brings a sense of daring (*tolmē*), which results in the soul's total descent into body (in

contrast to Plotinus, who allowed a part to remain undescended). Once in the 'oyster shell' of the body it longs for proper completion and unification. From here the soul can be elevated, 'be saved', in three complementary ways, aided by the appropriate teachers and guiding spirits. With erotic love it strives to unite with the higher life (as in Plato's *Symposium*). With philosophical contemplation it reaches the intelligibles and speculates on the first principles. With theurgy (see Chaldaean Oracles; Iamblichus) it culminates in the leap of faith (*pistis*), which unites the soul's own 'one-henad' with the all-perfect One. The particular soul, through not knowing its place, can thus descend and ascend indefinitely through all the levels of thought and being (*Elements of Theology* proposition 206). This 'travel' is ontological when it involves the cycle of birth and death, and epistemological when the mind engages in lower or higher pursuits during a lifetime.

'Body' has several senses for Proclus. At its simplest it is pure space-extension, like a body of intangible light. Endowed with layers of qualities it becomes the physical body of our sensory experience. When soul descends into participation, it acquires a series of bodies, referred to as 'vehicles' (*ochēmata*). Every participated soul has at first a tenuous, 'luminous' body-vehicle (for example, the world-soul has universal space). However, any soul that descends further into the physical domain acquires a further vehicle composed of the four elements, fire, air, water, earth. Finally, those that incarnate on earth (for example, humans) acquire their particular fleshly body. So soul-types suit their degree of participation in body (consistent with the overall metaphysics). The vehicles played an important part in religion. After death the soul is purified by shedding its non-rational nature and the associated elemental vehicle. It is left free to ascend with its luminous vehicle.

Since the essence of human life resides in an immortal soul, which is separable from the physical body, Proclus adheres to the Pythagorean-Platonic doctrine of transmigration of soul (see Plato §11; Pythagoras §2; Pythagoreanism §3; Psychē). However, he does not believe that a soul can reincarnate backwards in the evolutionary scale towards full consciousness. References to reincarnations of humans in animals (Plato, *Timaeus* 42b-c), for Proclus, have sense if taken psychologically, not biologically: that is, a human can lead the life of a wolf (if unjust) or an ass (if gluttonous) but cannot be reincarnated as an actual wolf or ass.

10 Philosophy of science

Proclus' deep interest in the nature of reality includes the physical world and its scientific description by mathematics. He was familiar with practical and empirical science. In his main astronomical work, *Outline of Astronomical Hypotheses*, he describes the construction of a spherical astrolabe. He himself made some of the last reliable astronomical observations of antiquity (AD 475), and instilled this interest in Ammonius, son of Hermeas and Heliodorus.

Proclus' theory of knowledge is linked to his metaphysics of graded being, so instead of the division into rationalism and empiricism he has degrees of knowledge dependent mainly on the level of consciousness and awareness of the thinker (and on the kind of being the object of thought is). The clearest kind of knowledge is reserved for the non-discursive contemplation proper to the intelligibles, where realities are perceived directly and complete, as by intuition. Knowledge of the physical world falls below this, for it combines discursive reasoning with sensory perception.

What arises from sensation is contingent and frequently confused. It is an opinion (*doxa*). Discursive reasoning (*dianoia*) uses concepts to define things, and to correct sensory evidence (as in the apparent smallness of the sun due to the distance from the observer). Scientific knowledge results from the reasoning activity of the soul, which matches sensory-derived opinions with concepts, according to positive criteria for truth: syllogistically demonstrable proof, and correspondence to (metaphysical) reality.

Proclus' applications of his theory of scientific knowledge are most evident in his philosophy of astronomy, and specifically in his rejection of Ptolemaic cosmology (in which he stands unique until the Renaissance) (see Ptolemy). Two remarkable examples can be summarized. Proclus rejects the precession: not the precession of the equinoxes as we now understand the phenomenon, but Ptolemy's interpretation of it as a movement of all the fixed stars. For Proclus, such stars cannot precess because it is in their nature to be fixed, and those who claim that their observations fit the hypotheses fall foul of the false-false-true Aristotelian syllogism (for instance, premises: dogs are blue, blue things bark; conclusion: dogs bark). He denies that the planets are moved by nested heavenly spheres (a doctrine that dominated European astronomy until the sixteenth century) because the arguments for

them he finds to be conjectures, not necessary and demonstrable proofs, and because the celestial bodies are ontologically capable of moving by themselves in free space.

Proclus' philosophy of physics also marks a radical turn away from Aristotle on two key fronts: the elements of the physical world, and place as space. He rejects as *ad hoc* and superfluous Aristotle's fifth element for the heavens, the aether. The celestial bodies consist of the same (four) elements as the terrestrial. For instance, without a kind of fire the stars and the sun could not illuminate, and without a solid bit of earth-element they could not be opaque (as during eclipses). Employing his metaphysics of graded being, Proclus distinguishes the heavenly mode from the terrestrial: in the former, the elements are in their 'summits', for example, fire is purely illuminating, earth is purely tangible and opaque; in the latter, they are gross, for example, fire burns, earth is heavy. Furthermore, he develops a new account of the primary properties of the elements, which he derives from the size, shape and facility for motion of their particles.

The place a body occupies is not its boundary, as Aristotle stated. Proclus uniquely proposes that it must also be a kind of body, but without the mass that makes physical bodies resist each others' presence. It is three-dimensional space. By extension there is a cosmic space in which all the bodies of the universe are immersed. Space on its own, apart from the bodies that move in it, is like a body of light.

As to mathematics, Proclus rejects Aristotle's view that mathematical objects (numbers, points, lines, proportions and so on) are abstractions from physical things, and develops the Platonic argument that they exist independently of human experience, but without the inscrutability of Forms. Mathematicals have dimensions, occupy relative places, are many in number within each kind, and so share some of the features of physicality. They dwell in the mental essence of soul, but without dimension. Soul 'projects' them onto its imagination, as onto a screen, where they acquire their familiar dimensionality (*Commentary on the First Book of Euclid's Elements* 50-2). Thus mathematical knowledge is a mixture of reasoning and imagination. The mathematical study of bodies (mechanics, optics, astronomy) relies on holding particular images more than the study of shape in general (geometry). The study of numbers relies on it much less and therefore is closer to pure reasoning. Above all stands the study of rules and relations, of analysis and synthesis. Ultimately the objects of mathematics originate in the Limit-Unlimited of the One. Mathematics is a bridge between sensory perception and the most complete state of knowledge. Thus in education, mathematical study prepares the mind to apprehend the pure realities.

11 Influence

Through his students (for example, Ammonius) and writings, Proclus influenced later Greek philosophy at its two leading centres, Athens and Alexandria, until the end of antiquity in the seventh century. His metaphysical system was adopted by Pseudo-Dionysius the Areopagite for a Christian celestial hierarchy. In turn, this influenced both Byzantine thinkers (Maximus Confessor, John of Damascus) and those of the Latin West (see Eriugena, J.S.; Grosseteste, R.). Islamic theologians of the tenth century, such as the Ikhwan al Safa', were inspired by Proclus' emanationism and theory of mathematicals (see Ikhwan al Safa'; Neoplatonism in Islamic philosophy). Furthermore, Arab scholars produced a compilation of his *Elements of Theology* under an Aristotelian label. Thomas Aquinas (§7) was the first to detect that Proclus was the author of the doctrines in all these works. Aquinas' friend William of Moerbeke rendered the first Latin translations of Proclus' main works. A new wave of Proclus' direct influence arrives in eleventh- to fifteenth-century Byzantium, and in Humanist Europe, notably with Ficino and Nicholas of Cusa. From there, Proclus' philosophy and science can be found in Kepler, the Cambridge Platonists (see Cambridge Platonism), Spinoza, the English Romantics, Cousin, and in various philosophers of idealism, culminating with Hegel (see Idealism; Hegel, G.W.F.).

LUCAS SIORVANES

List of works

We cannot date Proclus' writings (many of which were based on lecture-courses) because he continued making additions after initial publication. The *Platonic Theology* seems to be his last major work. There are student's excerpts of the commentary on Plato's *Cratylus*, but commentaries on Plato's *Theaetetus*, *Philebus*, *Sophist*, *Phaedo*, *Phaedrus*, *Symposium* and on Aristotle's logic have not survived; on Plotinus there is only a tiny portion. In nineteenth-century London, Thomas Taylor, who influenced the Romantics, translated most of Proclus' works. However, the translations of modern, reliable editions of Greek texts are mainly in French or German (as are many

of the studies).
Proclus (mid 5th century AD) *The Elements of Physics* or *On Motion*, ed. A. Ritzenfeld, *Institutio physica*, Leipzig: Teubner, 1912; trans. T. Taylor, 1831.(The former is a parallel Greek text and German translation, the latter an English translation; the text has no Neoplatonic theorems but is a short study manual on Aristotle's *Physics* books VI and VII and *On the Heavens* book I.)
Proclus (mid 5th century AD) *The Elements of Theology*, ed. and trans. E.R. Dodds, Oxford: Clarendon Press, 2nd edn, 1963; repr. 1992.(Systematic study manual of Proclus' Neoplatonism.)
Proclus (mid 5th century AD) *On Providence*, ed. D. Isaac, *Dix problèmes concernant la providence*, Paris: Les Belles Lettres, 1977; trans. T. Taylor, 1833; repr. Frome: Prometheus Trust, 1998.(Isaac has parallel Greek text and French translation; Taylor's English translation also includes *On Fate*, and *On Evil*.)
Proclus (mid 5th century AD) *On Fate*, ed. D. Isaac, *Providence, Fatalité, Liberté* Paris: Les Belles Lettres, 1979. (Parallel Greek text and French translation.)
(mid 5th century AD) *On Evil*, ed. D. Isaac, *De l' existence du Mal*, Paris: Les Belles Lettres, 1982.(Parallel Greek text and French translation.)
Proclus (mid 5th century AD) *A Commentary on the First Book of Euclid's Elements*, ed. G. Friedlein, *In primum Euclidis elementorum librum commentarii*, Leipzig: Teubner, 1873; trans. G.R. Morrow, Princeton, NJ: Princeton University Press, 1970; repr. 1992.(Morrow is English translation of the 1873 text.)
Proclus (mid 5th century AD) *Commentary on Plato's Republic*, trans. A.J. Festugière, *Proclus Commentaire sur la République*, Paris: Vrin, 1970.(French translation.)
Proclus (mid 5th century AD) *Commentary on Plato's Timaeus*, ed. E. Diehl, *In Platonis Timaeum commentaria*, Leipzig: Teubner, 1903-6, 3 vols; repr. Amsterdam: Hakkert, 1965; trans. A.J. Festugière, *Proclus Commentaire sur le Timée*, Paris: Vrin, 1966-8, 5 vols; trans. T. Taylor, London, 1810; repr. Hastings: Chthonios Books, 1988; repr. Frome: Prometheus Trust, 1998.(Festugière provides a parallel Greek text and French translation; Taylor is a reprint of his 1810 English translation.)
Proclus (late 5th century AD) *Outline of the Astronomical Hypotheses*, ed. C. Manitius: *Hypotyposis astronomicarum positionum*, Leipzig: Teubner, 1909; repr. Stuttgart, 1974.(Parallel Greek text and German translation.)
(late 5th century AD) *Commentary on Plato's Alcibiades I*, trans. W.O. Neill, The Hague: Martinus Nijhoff, 1971; ed. A.Ph. Segonds, *Sur le premier Alcibiade de Platon*, Paris: Les Belles Lettres, 1985-6.(The former is an English translation, with commentary, the latter a parallel Greek text and French translation.)
Proclus (late 5th century AD) *Commentary on Plato's Parmenides*, ed. V. Cousin, *In Platonis Parmenidem*; Paris: Durand, 1864; repr. Hildersheim: Olms, 1961; trans. J.M. Dillon and G.R. Morrow, *Proclus' Commentary on Plato's Parmenides*, Princeton, NJ: Princeton University Press, 1987; repr. 1992.(Book 7 is preserved in a Latin rendering by W. Moerbeke, recovered by R. Klibansky and L. Labowsky, London 1953, published in a revised edition (ed.) C. Steel, *Commentaire sur le Parmenide de Platon*, 2 vols, Leuven: Leuven University Press, 1982-5; with introduction.)
Proclus (late 5th century AD) *Platonic Theology*, ed. H.D. Saffrey and L.G. Westerink, *Théologie platonicienne*, Paris: Les Belles Lettres, 1968-97; trans. T. Taylor, *The Platonic Theology in Six Books*, London, 1816; repr. New York: Selene Books, 1985-6.(The former is a parallel Greek text and French translation, the latter a reprint of Taylor's 1816 translation.).

References and further reading

Gersh, S. (1973) *Kinēsis Akinētos; A Study of Spiritual Motion in the Philosophy of Proclus*, Leiden: Brill. (Analysis of the central ideas of Proclus' metaphysics.)
Lloyd, A.C. (1967) 'Athenian and Alexandrian Neoplatonism', in A.H. Armstrong (ed.) *Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press, 302-25.(A concise, excellent introduction to Proclus.)
Lloyd, A.C. (1990) *The Anatomy of Neoplatonism*, Oxford: Clarendon Press.(The logical structure of Proclus' metaphysics examined in depth; for the advanced reader.)
Saffrey, H.D. and Pépin, J. (eds) (1987) *Proclus: lecteur et interprète des anciens (Proclus, Lecturer and Interpreter of the Ancients)*, Paris: CNRS.(Rich source for recent English, French and German articles on many aspects of Proclus' thought.)
Siorvanes, L. (1996) *Proclus: Neo-Platonic Philosophy and Science*, Edinburgh: Edinburgh University Press and New Haven, CT: Yale University Press.(Comprehensive account of Proclus, including analysis of his metaphysics and theory of knowledge, suitable for the newcomer.)

Prodicus (*fl.* late 5th century BC)

Prodicus was a Greek Sophist from the island of Ceos; he was active in Athens. He served his city as ambassador and also became prominent as a professional educator. He taught natural philosophy, ethics, and of course rhetoric, but he is best known as an authority on correct language, specializing in fine verbal distinctions. Prodicus' greatest influence was due to his naturalistic interpretation of the traditional Greek gods; as a result, he later figured on the short list of famous atheists.

Prodicus came to Athens on public business as representative of his native Ceos, and seems to have settled there permanently as a professional teacher, earning large sums from his course of instruction. By 423 BC he was famous enough to be mocked by Aristophanes in his comedy the *Clouds*, together with Socrates, as a teacher of natural philosophy (*Clouds* 360-1; A5). The Platonic Socrates claims to have heard Prodicus lecture on language, although he could only afford the cheap, one-drachma course, not the deluxe course for fifty drachmas (Plato, *Cratylus* 384b). And Socrates says that he frequently sends students to Prodicus who do not seem apt to profit from associating with himself (Plato, *Theaetetus* 151b). Prodicus is cited in Plato's dialogues as an authority on 'the correctness of names'. Plato has made him a figure of fun by representing him as insisting upon fine distinctions between near synonyms such as 'pleasure' and 'enjoyment', 'wanting' and 'desiring' (see, for example, Plato, *Protagoras* 337a-c, 340a).

Prodicus was author of a book entitled *Hōrai (Seasons)* from which Xenophon (*Memorabilia* II 1.21-34) has extracted a long quotation or paraphrase known as the 'Choice of Heracles'. In this text two allegorical figures, women named Virtue and Vice, present Heracles with alternative paths to happiness, one of which is short and easy, while the other is long and arduous (Xenophon, *Memorabilia*, II. 1.21-34; fr.2). The rhetoric is picturesque, the morality is predictable, but the allegory became extremely influential.

The most interesting philosophical contribution of Prodicus was a naturalistic theory of the origin of religion (fr.5). According to Sextus Empiricus, Prodicus claimed that 'the ancients considered that sun, moon, rivers and springs, and generally all things beneficial for our lives, are gods because of their usefulness'. Other reports suggest that the development of agriculture played an important role in Prodicus' explanation, and that names like Demeter and Dionysus were said to reflect the invention of bread and wine. According to several authors, Prodicus denied the existence of the traditional Greek gods. If this is correct, his theory would be more anthropological than allegorical.

Prodicus' explanation of the origin of religion apparently presupposes a more general account of the evolution of human culture from primitive beginnings, when the life of mankind was like that of beasts. This account in turn belongs in a larger cosmological context: the attempt of Ionian natural philosophy to explain how the current world order developed from a more primitive state of affairs 'in the beginning'. The rise of human civilization is the final chapter in Presocratic cosmogony, as we find it more fully reproduced in later authors such as Lucretius.

Prodicus' explanation of Demeter and Dionysus as inventors of bread and wine is echoed in Euripides' *Bacchae* (274-85), where it is used to justify the worship of Dionysus. Since we do not have Prodicus' own words, it is difficult to know how far his theory is compatible with the allegorical practice by which later philosophers preserved respect for the civic cult by reinterpreting the traditional gods in more scientific terms. Like the evolutionary account, the allegorical interpretation is older than Prodicus. Thus Metrodorus of Lampsacus and other followers of Anaxagoras will identify Zeus with intelligence (*nous*), Athena with art or craft (*technē*), Achilles with the sun, Hector with the moon, and so on (A4 and 6). Diogenes of Apollonia is reported to have identified Zeus with cosmic intelligence in the form of air (see Diogenes of Apollonia §2). It was this type of allegory that became conventional (Apollo as the sun, Artemis as the moon, Athena as wisdom, and so on) and was later developed systematically by Stoicism. According to some scholars, Prodicus belongs in this tradition.

If the report of atheism is correct, however, Prodicus was concerned only to explain and not to justify the worship of the gods. The relevant parallel would be what we find in the *Sisyphus* fragment (fr.25 variously attributed to Euripides and to Critias), where the invention of the gods is described as the device introduced by a clever man in order to enforce morality. A similar cultural prehistory must lie behind the fragment 30 of Democritus (compare

Democritus §4), that claims that 'a few learned people raised their hands on high and called the Air "Zeus"'. Prodicus' theory was apparently more elaborate. He seems to have distinguished two stages in the development of religion. In the earlier stage human beings began to worship the forces of nature on which they depended: sun, rivers, springs, and things useful for life. In the next stage, moving from the state of nature, they deified cultural heroes and heroines who invented agriculture and useful arts: Demeter, Dionysus, Athena and the like were divinized mortals, like Heracles and Asclepius. On this view, Prodicus is the author of the theory of religion later known as Euhemerism, which claims that the gods were originally human beings who were deified because of their great achievements.

See also: Sophists

CHARLES H. KAHN

References and further reading

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, vol. 3, Cambridge: Cambridge University Press; part of vol. 3 repr. as *The Sophists*, Cambridge: Cambridge University Press, 1971.(Full and scholarly account.)

Henrichs, A. (1975) 'Two Doxographical Notes: Democritus and Prodicus on Religion', *Harvard Studies in Classical Philology* 79: 93-123.(Discussion of key texts on Prodicus' theological thesis.)

Henrichs, A. (1984) 'The Sophists and Hellenistic Religion: Prodicus as the Spiritual Father of the Isis Aretologies', *Harvard Studies in Classical Philology* 88: 139-58.(Prodicus' theological thesis and its influence.)

Kerferd, G.B. (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press.(A briefer, more personal interpretation than Guthrie (1969).)

Plato (*c*.386-380 BC) *Protagoras*, trans. C.C.W. Taylor, Oxford: Oxford University Press, 2nd edn, 1991.(Shows Prodicus in conversation with Socrates, Protagoras and others. Includes a commentary; Prodicus appears in the gathering of famous Sophists.)

Prodicus (*fl.* late 5th century BC) 'Fragments', in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidmann, 6th edn, 1952, vol. 2, 380-.(The standard collection of the ancient sources in the original languages both fragments and testimonia, the latter designated with 'A'; includes Greek texts of the fragments with translation in German.)

Sprague, R.K. (ed.) (1972) *The Older Sophists*, Columbia, SC: University of South Carolina Press, 70-85.(Full English translation of the fragments and testimonia from Diels and Kranz (1952).)

Xenophon (*c*.360 BC), *Memorabilia*, trans. E.C. Marchant, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1923, II 1.21-34.(Greek text with English translation; Prodicus' 'Choice of Heracles' is at II 1.21-34.)

Professional ethics

Professional ethics is concerned with the values appropriate to certain kinds of occupational activity, such as medicine and law, which have been defined traditionally in terms of a body of knowledge and an ideal of service to the community; and in which individual professionals have a high degree of autonomy in their practice. The class of occupations aiming to achieve recognition as professions has increased to include, for example, nursing, while at the same time social and political developments have led to criticism of and challenge to the concepts of professions and professionalism. Problems in professional ethics include both regulation of the professional-client relationship and the role and status of professions in society. A central question for ethics is whether there are values or virtues specific to particular professions or whether the standards of ordinary morality are applicable.

1 What is a profession?

The term ‘profession’ is used in different senses. In a wide sense it simply means someone’s occupation; in a narrower sense it refers to a certain kind of activity, one carrying with it a certain status and associated with a particular ethic. Traditionally a profession has been marked out by a body of knowledge, mastery of which (at least partly) regulated entrance to its ranks; and by an ideal of service (see Airaksinen 1994). Since the body of knowledge had the potential to confer power, money and status, professionals were expected to use their skills for the benefit of the community. Those groups which have long been secure in their recognition as professions, the so-called liberal or learned professions such as medicine, divinity and the law, have also been characterized by a considerable degree of authority and autonomy in their practice. The epithet ‘liberal’ emphasized their suitability for the ‘free man’; ‘learned’ brought out their relationship to a body of knowledge. The fact that professionals are said to ‘practise’ rather than ‘work’ is itself significant in reinforcing their autonomy and status. Along with the autonomy of the individual professional, professional bodies have also been accorded a significant degree of autonomy in controlling both access to the profession and professional conduct.

Eliot Freidson (1994) outlines the problems of defining ‘profession’ in terms of a number of characteristics or traits and considers instead a definition in terms of process, that is, how particular occupational groups gain professional status. He argues, however, that this covertly defines a profession as an occupation that has gained professional status. Freidson’s view is that the concept of ‘profession’ is one that is tied to particular socio-cultural conditions, namely Anglo-American industrial nations which associate status with occupation, rather than with education, as is more common, he claims, in other European countries.

The traditional classification of professions has been subject to two contrasting trends: first, the attempt by some groups for recognition as professions or neo-professions; and second, challenge to the notion of professionalism either because of its conceptual inadequacy, as noted by Freidson, or on the grounds of its social consequences.

Nursing is a prominent example of an activity which has aimed at recognition as a profession, in its struggle to distance nurses from the image of handmaiden and to advance towards their acceptance as accountable professionals (see Nursing ethics); to discard the metaphor of loyalty and take up that of advocacy (see Winslow 1984). In such struggles, the concepts of professional autonomy and accountability play a part along with the identification of a body of knowledge.

Coinciding with a move on the part of some groups to become recognized as professions, there has been a challenge to the concepts of ‘profession’ and ‘professionalism’, and an attempt to replace the focus on the concept of profession with one of a set of competences, in other words, to concentrate on what people do and achieve rather than on their status. One reason for this has been that critical, reflective professionals, with autonomy over their practice, may be seen as a threat (see Williams 1996). A second reason is connected with the potential for professions to become self-serving elites (see Freidson 1994; Illich 1977). Yet a third is the increase in criticism if not litigiousness of more knowledgeable clients. Williams contrasts the professional approach and the competence-based approach in the following way: ‘Competent workers are content to be described as such, and trainers do not try to push them any further, while reflective practitioners aspire to research their own practice and agencies’ (Williams 1996: 8).

What is not in dispute between the two approaches is an attempt to distance what is required from incompetence.

J.K. Davis (1991), for example, has argued that for professionals it would not be sufficient that a client was satisfied, if the professionals themselves felt that the service was below standard. For the professional, however, it is more than simply doing a competent job: a worker becomes a professional by professing reasons for doing their work in a certain way.

2 Professional ethics: the problems

Problems of professional ethics fall into two broad categories, but both arise essentially from professional power. The first is concerned with the professional-client relationship, while the second relates to the role of professions and professionals in society as a whole.

Although an ideal of service is supposed to provide a safeguard to promote the use of professional expertise to help rather than harm, specialist knowledge, to which professionals have access and clients do not, does give power to the professional, and the client is thus placed in a vulnerable position. One caveat however is that the paradigm of a relationship between two individuals is inadequate because it overlooks those professions which do not conform to this pattern, such as teaching, which may be but commonly is not done on a one-to-one basis (see Langan 1991).

The most thorough discussion of the professional-client relationship has taken place in relation to the medical profession, with regard both to the feasibility of applying traditional ethical theories to the problems of medical practice and to the development of a set of principles specific to the context of the relationship between health care professional and patient, as in Beauchamp and Childress' principles of biomedical ethics (1979) - autonomy, beneficence, nonmaleficence and justice. A question about the use of power in the medical context would be whether practitioners should use their expertise to decide, beneficently, what is in the patient's interest, or provide information to enable the patient to take an autonomous decision (see Medical ethics §2).

The second category of problems is more concerned with the role and image of professionals in society. While it may be true that there has always been a tendency towards distrust of professionals, this has been exacerbated by social and political developments (see Pellegrino 1991). The trend towards client autonomy, attempts by government to curb the independence and privilege of professionals, and media criticism have all had their effect. Darryl Koehn (1994) adds to this list the assault of academic disciplines such as sociology and philosophy. The sociological critique has suggested that professions, rather than being essentially moral enterprises, are in fact effective monopolistic institutions and that the professed commitment to ethical ideals, rather than conferring legitimacy on the profession, is nothing more than ideology. Ivan Illich has termed the mid-twentieth century the age of 'disabling' professions: far from using their knowledge to serve, they have become forms of control, claiming the authority to determine human needs. 'Homes are transformed into hygienic apartments where one cannot be born, cannot be sick and cannot die decently' (Illich 1977: 27).

Philosophers have taken issue with a self-derived ethic which permits professionals to be guided by standards other than those of ordinary morality. 'Problems in professional ethics typically arise when the values dominant within particular professions come into conflict with other values in the course of practice. Professionals are likely to perceive these values as dominant where others may not' (Goldman 1992: 1018). The possibility of self-derivation is linked with the autonomy of professional bodies in determining standards of practice. A self-derived ethic might take one of two forms. In one form it is associated with the idea that there are certain ways of behaving appropriate to different roles, which diverge from those suited to people who do not fill that role. For example, it might be argued that a lawyer is under an obligation, arising out of the lawyer's role, to achieve the best result for a client even if that conflicts with what they believe as a private individual.

> I remember a custody case for a most disagreeable man whose wife had left him with the children. I remember reducing her to floods of tears in the witness box and I felt very badly about it because I thought she was a very nice woman, which she was, and her husband was a shit. On professional grounds I knew I had done a really good job. But as a father of young children I really thought that the right result hadn't been achieved.
>
> (Quoted in Chadwick 1991)

It is possible to interpret this quotation in different ways. It might be seen either as a clash between professional standards and 'ordinary' morality, or as a conflict between two different moral viewpoints: the view that one's duty is to see justice done and the view that one's duty is to the client and the court.

The second form in which a self-derived ethic might find expression is in a code of professional conduct or code of ethics. The possession of a code of professional conduct has been pivotal in debates about what constitutes a profession. Such a code can fulfil a variety of functions (see Chadwick 1992): offering a public statement of ideals and values; providing a disciplinary mechanism for a professional body; reassuring the public that the profession upholds certain standards; and educating members of the profession to 'think like' others in the group (see Davis, M. 1991).

The standards incorporated in a code may be either higher or lower than the standards of ordinary morality. Professionals have traditionally been prevented from doing things which people in other spheres of activity are permitted to do, such as advertising. This arises out of the purported commitment to serve first the interests of clients, rather than their own profit. On the other hand this same commitment can act as a shield to protect professionals from the criticism that they do things which would be frowned on in terms of ordinary morality, such as lying to clients or physically hurting them in order to promote some further end identifiable as being in the client's interests (see Hayry and Hayry 1994).

Criticism of a self-derived professional ethic, whether in the form of role ethics or a code of conduct, is based on arguments that if an action is morally right it should be susceptible of justification by the same moral arguments that apply to the behaviour of any other member of society - professionals should not require special ethical norms to be determined by themselves. For it is not clear how such norms could be justified if not by common moral principles (see Goldman 1992).

If this point is accepted, however, there are arguments in favour of having, if not a self-derived ethic, at least an ethic specific to a given professional group, in order to promote consistency among practitioners rather than leaving moral judgment to the fallible individual alone. Even if it is acknowledged that the individual practitioner has autonomy in their practice, the requirements of accountability and a set of defined standards has the result that the practitioner may literally be called to 'give an account' of themselves and their practice in terms that their professional peers will recognize (see Holdsworth 1994). This arguably affords more protection than the absence of professionally defined norms but does not constitute an argument for dispensing with accountability to society as well.

Sociological and philosophical criticism have constituted aspects of a phenomenon which might be called the challenge to the ethics of trust in the professions (see Trust). Traditionally the professional-client relationship has been regarded as a fiduciary one, in which the professional is worthy of trust and the client places trust in the professional. In the late 1980s and early 1990s, however, the ethics of trust was challenged. In the light of these challenges the question arises whether the appropriate response is to turn to an 'ethics of distrust' (see Pellegrino 1991). An ethic of distrust would proceed by attempting to regulate more closely the activities of professionals, by increased external monitoring and demands for accountability. According to one view this approach wins the day by default because the notion of an ethics of trust is not only difficult to sustain: it is actually incoherent.

Robert Veatch (1991) attacks what he sees as the three arguments supporting an ethic of trust: (1) that professionals serve the client's interest; (2) that professionals can present value-free facts to the client; and (3) that professionals should act on a set of virtues inherent in the profession. Veatch argues that modern professionals ought not to know what the client's interests really are - the most they can know is what the client's interests are in one particular area of life. Whereas medical professionals might be concerned primarily with promoting health, for example, health might not be the top priority for a patient (see Goldman 1992). Veatch also argues both that professionals cannot present value-free facts and that it is a serious mistake to think that any given profession is associated with one particular conception of virtue. For example, Talmudic and libertarian lawyers will be informed by different conceptions of the appropriate virtues.

A less pessimistic approach is to investigate the possibility of grounding trust in the professions anew. Koehn has argued that this is urgent because professions represent the mechanism chosen by Anglo-American morality for providing people with goods such as health and justice, and if professionals are not trustworthy, where are we to turn for help? For health and justice are not goods that are readily dispensed with. For Koehn, the challenge is to show not only that there are grounds for trust in the professions because they provide people with such goods, but also that they do not violate the requirements of ordinary morality.

3 Theoretical approaches: external and internal

How are the ethical questions to be addressed? The central issue seems to be the contrast between the external and internal perspectives. The external approach would be to take a normative theory, such as Kantianism or utilitarianism, and apply it to the issues arising in the professional-client encounter and the social role of professionals (see Kantian ethics; Utilitarianism). The alternative would be to examine whether, rather than applying such an external standard, there are values internal to particular professions. A feminist ethic of care could count as either external or internal according to how it was interpreted - whether, for example, an ethic of care provides a framework for addressing problems or whether it perceives care as a virtue inherent in certain professions (see Curzer 1993) (see Feminist ethics §1; Nursing ethics).

Providing an example of the application of an external theoretical framework, Richard Tur (1994) suggests that the standard conception of the lawyer's role arises from the political philosophy of liberalism with its neutrality concerning ideas of the good life, and is associated with the moral philosophy of utilitarianism (see Liberalism). The application of utilitarianism suggests both that the good lawyer is one who pursues the objectives of clients effectively, and that doing this under the adversarial system promotes the best interests of society as a whole. This is both because the adversarial system is justified in terms of facilitating the emergence of truth and because to allow individuals to pursue their own view of the good life will maximize social benefits.

The internal approach, on the other hand, might attempt to derive values internal to specific professions by examining the point of those professions. Rather than accepting them as Illich's 'dominant' professions that take it upon themselves to define human need, the question to ask is what pre-existing human need or value do and should they serve? This quest might take different forms. The identification of health and justice as goods that cannot readily be dispensed with, because they may be needed by vulnerable people, has been mentioned. Or there might be an argument for some intrinsic or 'transcendent' values embedded in a professional activity. In the case of law, Tur suggests innocence as an internal value that has implications for specific actions in practice - if innocence is non-negotiable then plea bargaining becomes unethical. Third, knowing the point of a practice such as a professional activity might point the way to virtues internal to the practice of that activity. The virtuous doctor and the virtuous lawyer will not need to apply an ethical theory such as engaging in a utilitarian calculation. As virtuous practitioners they will be able to see what is required in particular situations (see Virtue ethics; Virtues and vices).

On this view the virtues of a profession will be defined by reference to the meaning of the practice in question. This would facilitate an answer to the question, for example, of whether nursing ethics is distinct from medical ethics. The central issue that remains unresolved is whether the tension between a relativistic and a universalistic approach to professional ethics can be mediated by an appeal to the unity of the virtues in the moral life (see Meilander 1991) (see Moral relativism; Universalism in ethics).

See also: Applied ethics; Business ethics; Engineering and ethics; Journalism, ethics of; Medical ethics; Responsibilities of scientists and intellectuals; Technology and ethics

RUTH CHADWICK

References and further reading

Airaksinen, T. (1994) 'Service and science in professional life', in R.F. Chadwick (ed.) *Ethics and the Professions*, Aldershot: Avebury, 1-3.(On issues concerning the human genome project.)

Beauchamp, T.L. and Childress, J.F. (1979) *Principles of Biomedical Ethics*, New York: Oxford University Press; 4th edn, 1994.(Classic text on the four principles of biomedical ethics - autonomy, beneficence, nonmaleficence and justice.)

Chadwick, R. (1991) 'Is There a Difference Between Standards and Ethics?', *Journal of Advances in Health and Nursing Care*1: 75-89.(Discusses the difference between ethics and the standards defined by a profession.)

Chadwick, R. (1992) 'The function of corporate codes of ethics', in *Business Ethics: Contributing to Business Success*, Sheffield: Sheffield Business School, 79-87. (Overview of the functions of codes.)

Curzer, H.J. (1993) 'Is care a virtue for health care professionals?', *Journal of Medicine and Philosophy* 18 (1): 51-69.(A critique of care as a virtue.)

Davis, J.K. (1991) 'Professions,Trades and the Obligation to Inform', *Journal of Applied Philosophy* 8 (2): 167-76.(Distinguishes between professionals and workers in trades.)

Davis, M. (1991) ‘Thinking Like an Engineer: the place of a Code of Ethics in the Practice of a Profession’, *Philosophy and Public Affairs* 1 (20).(Analysis of a particular function of codes with reference to a specific example.)

Freidson, E. (1994) *Professionalism Reborn: Theory, Prophecy and Policy*, Oxford: Polity Press.(On the concept of profession and the future of professionalism.)

Goldman, A. (1992) ‘Professional Ethics’, in L.C. Becker (ed.) *Encyclopedia of Ethics*, Chicago, IL and London: St James Press, 1018-20.(Introductory article.)

Hayry, H. and Hayry, M. (1994) ‘The Nature and Role of Professional Codes in Modern Society’, in R.F. Chadwick (ed.) *Ethics and the Professions*, Aldershot: Avebury, 136-44.(Analysis of the role of codes.)

Holdsworth, D. (1994) ’Accountability: the Obligation to Lay Oneself Open to Criticism’, in R.F. Chadwick (ed.) *Ethics and the Professions*, Aldershot: Avebury, 58-87.(Analysis of the notion of accountability.)

Illich, I. (1977) *Disabling Professions*, London: Marion Boyars.(A critique of professional authority.)

Koehn, D. (1994) *The Ground of Professional Ethics*, London: Routledge.(Attempts to justify trust in the professions with reference to medicine, law and the ministry.)

Langan, J. (1991) ‘Professional Paradigms’, in E.D. Pellegrino, *et al*. (eds) *Ethics, Trust, and the Professions: Philosophical and Cultural Aspects*, Washington, DC: Georgetown University Press, 221-35.(Points out the limitations of the paradigm of the professional-client relationship as one between two individuals.)

Meilaender, G. (1991) ‘Are There Virtues Inherent in a Profession?’, in E.D. Pellegrino, *et al*. (eds) *Ethics, Trust, and the Professions: Philosophical and Cultural Aspects*, Washington, DC: Georgetown University Press, 139-55.(Compares views of professional ethics with reference to Oakeshott’s forms of moral life.)

Pellegrino, E.D. (1991) ‘Trust and Distrust in Professional Ethics’, in E.D. Pellegrino, *et al*. (eds) *Ethics, Trust, and the Professions: Philosophical and Cultural Aspects*, Washington, DC: Georgetown University Press, 69-85.(Explores the issues in the choice between an ethics of trust and an ethics of distrust in professional ethics.)

Tur, R.H.S. (1994) ‘Accountability and Lawyers’, in R.F. Chadwick (ed.) *Ethics and the Professions*, Aldershot: Avebury, 58-87.(Contrasts external and internal values with reference to law.)

Veatch, R. (1991) ‘Is Trust of Professionals a Coherent Concept?’, in E.D. Pellegrino, *et al*. (eds) *Ethics, Trust, and the Professions: Philosophical and Cultural Aspects*, Washington, DC: Georgetown University Press, 159-69.(Argues against an ethics of trust.)

Williams, B. (1996) *Freedom on Probation: a case study of the Home Office enforced changes to the University education and training of probation officers*. (Considers the challenge to the notion of professionalism with special reference to a specific case study.)

Winslow, G. (1984) ‘From Loyalty to Advocacy: A New Metaphor for Nursing’, *Hastings Center Report* 14 (3): 32-40.(Discussion specifically relevant to professionalization of nursing.)

Projectivism

'Projectivism' is used of philosophies that agree with Hume that 'the mind has a great propensity to spread itself on the world', that what is in fact an aspect of our own experience or of our own mental organization is treated as a feature of the objective order of things. Such philosophies distinguish between nature as it really is, and nature as we experience it as being. The way we experience it as being is thought of as partly a reflection or projection of our own natures. The projectivist might take as a motto the saying that beauty lies in the eye of the beholder, and seeks to develop the idea and explore its implications.

The theme is a constant in the arguments of the Greek sceptics, and becomes almost orthodox in the modern era. In Hume it is not only beauty that lies in the eye (or mind) of the beholder, but also virtue, and causation. In Kant the entire spatio-temporal order is not read from nature, but read into it as a reflection of the organization of our minds. In the twentieth century it has been especially non-cognitive and expressivist theories of ethics that have adopted the metaphor, it being fairly easy to see how we might externalize or project various sentiments and attitudes onto their objects. But causation, probability, necessity, the stances we take towards each other as persons, even the temporal order of events and the simplicity of scientific theory have also been candidates for projective treatment.

1 Mechanisms

Projective theories recognize the role of our own natures in shaping the judgments we make. They also promise useful metaphysical economies, by seeing us as responding to a world with less in it than is needed if we think that to every judgment there corresponds an independent property of things. They may therefore be motivated in a number of ways. On the side of metaphysics we may have concerns that if the real world is conceived in some special way, we introduce mysterious properties with equally mysterious relations to other properties. Or, we may have epistemological worries about the ways we know of the independent properties. On the side of the subject we might want to recognize that the judgments in question have special relations to action or motivation or emotion, and we might find this inexplicable if we see ourselves as simply cognizing an independent order of facts. Projectivism is therefore best seen as an explanatory theory, and accepting a projective theory itself manifests the commitment that some aspect of the world is not as independent of us as we might think.

Within the broad characterization there are many different accounts of the mechanisms of projection, of the consequences of a projective theory and of the method whereby we should assess whether a projective theory is appropriate in one area or another. Perhaps the most simple and forthright combination of doctrines would be that the mechanism is one of displacing, or relocating, what is in fact a feature of our experience, and making it into a feature of the world. The consequence is that we make a mistake, erroneously treating the world as containing features that it does not really contain (an error theory). The appropriate method is then to sift through the evidence that such a displacement actually occurs, and to decide whether we can live with the thought that some area of our judgment is thoroughly entangled in error. A theory of this form is found in the treatment of colour by Cartesian philosophers such as Malebranche and Arnauld. Colours are found only in the mind; the world itself is not coloured, but we naturally treat it as if it is. The answer to the question whether this makes God, who gives us this nature, a deceiver, is that it is good for us that we make the displacement, in the same way that it is good that we feel pains in our limbs: the mistake has a practical benefit. In more modern terms, colour vision is a useful adaptation to a world that is not coloured.

However, several features of this way of thinking may be questioned without departing from the projective model. We might deny that the mechanism is one of literal displacement. Perhaps the mind manages to make the world look as though some things have beauty, for example, without the really beautiful objects being in the mind itself; instead, some kind of pleasure or delight is objectified, it then appearing that there is a feature of things correlated with it. There would be no real or true bearer of beauty in the mental world but only the real, non-aesthetic properties of the object, the pleasure or delight that these arouse and the propensity we have to voice that pleasure by describing the object as beautiful. Theories of this tripartite kind are more commonly called projective in the contemporary idiom.

Historical examples include Spinoza's view that while everything is really necessary, we feel our own ignorance

when faced with the course of events and voice it by describing events as contingent. There is Hume's view that the real object of perception is the regular pattern of events, although faced with this we 'make no scruple' of foretelling one event upon the occurrence of another, and voice this by describing them as causally connected. There is equally Hume's view that, faced with some features of character such as temperance and cheerfulness, the mind feels a pleasure which it voices by deeming those features virtues. When we do this we 'gild or stain' objects with the 'colours borrowed from internal sentiment', and thereby 'erect in a manner a new creation' (see Hume, D.). In Kant, apart from what is said about space and time, there is his attitude to the judgment of beauty - that it essentially voices a felt delight, not prompted by a literal perception of beauty (see Kant, I.).

In the twentieth century, F.P. Ramsey and Bruno De Finetti held that talk of probabilities is a projection, not because events in the mental world are alone subject to probability, but because they saw such talk as voicing our degrees of confidence in response to observed frequencies of events. Also in Ramsey, to think of a regularity as a natural law is not to see it as somehow fixed by some special and mysterious locking device, but to give it a certain dignity as part of the system with which we meet the future. The emotive theory of ethics, designed to make a place for ethics despite its being neither empirical nor issuing in truths of logic, is a central example of the Humean approach (see Emotivism). Other examples occur in the works of the later Wittgenstein.

When something is described as nice, sublime, good or whatever, projective theories have traditionally found it important to locate the projected reaction, in the first instance, independently of describing it as 'whatever you feel'. If what you feel can only be identified as a tendency to make the judgment, then it is easy to think that no gain has been made. For if we are sincere, whenever we make a judgment we express a tendency to make the judgment. Seen this way there is an obligation on projectivists to identify the reaction in other terms: classic suggestions include a felt pleasure (lying behind judgments of beauty), a particular pressure on choice and action (behind ethical judgment), degrees of confidence (behind probabilities), willingness to rely on a regularity (behind seeing it as causal in nature) and so on.

2 Consequences

Each of the mechanisms outlined may suggest that some kind of error is involved. The stories bear some affinity to the familiar thought that an emotion may be incorrectly projected upon a thing, as when one nostalgically invests past events with an illusory glamour or perfection. However in the philosophical examples, it is not clear that we make any error when we project. We might think that we are deceived by our senses when we take the world as coloured, or by our passions when we take things to be beautiful, or by our degrees of confidence when we find events probable. But the alternative approach is that there is no question of error. All that is involved is a natural and desirable way of experiencing the world, and the fact that the categories involved have the projective explanation does not infect them with mistake. Discovering the mind-dependence of colours, for example, need not be presented as discovering that the world is not really coloured. It is just that the story behind the fact that it is coloured has its origin in the structure of our visual responses. To return to the analogy with the pain in the foot: it is not obviously an error to locate the pain in the foot (as it would be if the pain were really somewhere else, such as in the head).

If the metaphor of projection is dissociated from the idea of error, then it must be asked how much ordinary thought is compatible with the explanation. In ethics this question has proved especially controversial. To many writers it has seemed that the independent claim made on us by such things as moral obligation and duty becomes softened and lost if those categories are projections of attitudes or stances that we ourselves have taken up. The spectre arises that the explanation of the commitment in fact undermines it, so that there is a disharmony between ethics as it feels to the subject, and the philosophical appreciation of what it actually is. Others have argued that this is a mistake akin to worrying whether ethics is possible without God. There is no reason for the feeling of obligation to wane if it is revealed for what it is - a feeling - any more than there is reason for us to stop finding things funny when we recognize the subjective nature of that reaction. 'Quasi-realism' is a title invented to describe the attempt to reconcile projective theories with the appearances of independent (and even demanding) fact in such areas.

One focus for the kind of debate that arises is the notion of knowledge. It is frequently thought that a projective theory should deny that there is real knowledge to be had in an area. Thus a common title for the position is 'noncognitivism', and many projectivists have been content to accept it. However, if projectivism about colour

rejects the error theory, then it ought to be able to say not only that mail boxes in Britain are red, but that we know that they are. The force of saying this will be that there is no real possibility of an improved view of them, revealing that our original colour judgment was deficient - as it would be due to inadequate light, or a poor glimpse. Similarly in ethics, a projectivist may not only say that ingratitude is a vice, but that we know that ingratitude is a vice, again meaning that this judgment is secure enough for there to be no real chance of it being abandoned, whatever improvements in our moral sensibilities come about ('improvement' not meaning convergence on some antecedent truth, but simply better use of the virtues we already recognize: care, attention, imagination, sympathy). Once this is the landscape, it is hard to find simple ways of assessing projective theories.

3 A real debate?

Discontent may seize upon each of two dualisms, urging that projectivism magnifies the first, and the second is either underestimated or loses its identity as a distinct picture. The first is that of mind versus the world. In the case of colour, projectivism seems to depend upon this being a sharp division, with the issue being that of where the original place for colour is found - inside or outside, as it were. If projectivism in any area similarly depends upon separating how things are with one's mental life from how things are in the world, then it implies a highly controversial dualism, and may fall to a more unified picture of our lives in the world. This criticism may carry more weight in some areas than others. In the colour case, it may well be that proper attention to the way perception of colour, understood not as simply arousal of a sensation but as an engagement with the world, undermines the Cartesian story.

In the case of ethics, or modal judgment, or statements of law or probability, dualism does not seem to be the issue, since the non-representative nature of the basic responses is not a consequence of a split between mind and the world, but rather a consequence of distinguishing different aspects of the way we meet the world. This however introduces the second dualism, which is that of representation or description on one side, versus the reactions and responses that the projectivist needs to start with. Thus we have already touched on the apparent need for a projective theory to identify the reaction we project, at least initially, as something other than simply a judgment. But critics have frequently complained that the reactions cannot in fact be identified with the necessary precision without invoking the very vocabulary that needs to be avoided. For example, suppose we describe the state projected in making a positive ethical judgment as one of approval, the critic will say that approval is not a primitive feeling or sensation, but a highly complex state which is possible only to people already embedded in ethical practice. One reply is to say that we have enough grip of the primitive reactions (anger, guilt, willingness to succumb to pressure from others) in question to see our practices as based on them or developed out of them. Provided this rational reconstruction of the activity of moralizing is possible, a projectivist may grasp the nettle and put up with the apparent circularity: what is projected when we say that something is good may simply not be identifiable in any terms except those from the ethical family. Yet provided we can see how we got involved with this family, all may be well.

If the fundamental state of mind is not really one of making a judgment, then how can it properly acquire the trappings of judgment? A sharp form of this problem, originally aired by Frege, points to the numerous things we can do with judgments apart from voice them, and asks the projectivist to explain (see Frege, G.). For example, if saying that feminism is a good thing voices an attitude to feminism, what is said by 'if feminism is a good thing, we should employ more women' or 'feminism might be a good thing, but I am not sure', where no attitude is voiced? Although this problem has been attacked head-on, it is still often felt that a 'Fregean abyss' separates judgments with content from the projectivist's ingredients. On the other hand, if the abyss can be crossed, then one might worry about the point of starting with a non-representative response in the first place. Here again the projectivist need not give in, for even if our basic responses are essentially intentional, purporting to represent properties of things, nevertheless there is scope for different theories of how they evolve to be that way, and the basic notion of the objectification of a response may well remain the central ingredient in any such explanation.

See also: Relativism

SIMON BLACKBURN

References and further reading

Blackburn, S. (1993) *Essays in Quasi-Realism*, New York: Oxford University Press. (Chaps 5 and 6 locate some

of the issues and introduce the task for a quasi-realist theory based on projectivism.)

Gibbard, A. (1990) *Wise Choices, Apt Feelings*, Cambridge, MA: Harvard University Press.(The most impressive overall theory of normative judgment yet to appear.)

Hume, D. (1748) *Enquiry Concerning the Principles of Morals*, ed.L.A. Selby-Bigge, Oxford: Clarendon Press, 1978.(Eminently readable, and the classic introduction of the metaphor of gilding and staining the world. See especially the first Appendix.)

Kant, I. (1790) *The Critique of Judgement*, trans. J.C. Meredith, London: Oxford University Press, 1927.(Not so eminently readable: a sustained effort to explain how the objective order reflects the categories imposed by the mind.)

Ramsey, F.P. (1927) 'Facts and Propositions', in *Foundations - Essays in Philosophy, Logic, Mathematics and Economics*, ed. D.H. Mellor, London: Routledge & Kegan Paul, 1978.(The first modern philosopher to develop the idea that the issue is not truth, but the nature of judgment.)

Stevenson, C.L. (1945) *Ethics and Language*, New Haven, CT: Yale University Press. (An old but classic investigation of emotivism as a theory of ethics.)

Wittgenstein, L. (1956) *Remarks on the Foundations of Mathematics*, Oxford: Blackwell.(A disjointed but deep and impressive attempt to loosen the grip of realism about mathematical truth and necessity.)

Promising

Promising is often seen as a social practice with specific rules, determining when a promise has been made and requiring that duly made promises be kept. Accordingly, many philosophers have sought to explain the obligation to keep a promise by appealing to a duty to abide by such rules, whether because of the social benefits of the practice or because fairness requires one to abide by it. Others see breaking a promise as a direct wrong to the person whose expectations are disappointed.

Since one can be obliged to keep a promise even when some other course of action would produce greater happiness (or better consequences as measured by some other standard), the obligation to keep a promise cannot be derived simply from the beneficial consequences of doing so (see Consequentialism). In order to explain this obligation, many philosophers have been drawn to the idea that promising is a social practice whose rules, including the rule that promises must in general be kept, we are bound to obey. They have, however, offered different accounts of why we are so bound.

Hume (1739-40) argued that promises depend on the existence of a convention, although his account differed from most modern ones in emphasizing the *virtue* of fidelity rather than a duty or obligation (see Hume, D. §4). Hume called fidelity (or more generally the virtue of justice, of which fidelity is a central component) an 'artificial' virtue because he thought that the moral approval that an instance of fidelity excites derives from the usefulness of the general pattern of which it is a part rather than from the consequences of that particular action (which may in fact lead to more harm than good).

Practice-based accounts of promises have sometimes been seen as a way of reconciling utilitarian and apparently nonutilitarian intuitions (see Utilitarianism). Rawls, for example, argued in 'Two Concepts of Rules' (1955) that (1) promise making, like a move in a game, is an action that presupposes the rules of a certain practice, and (2) while the practice as a whole can be given a utilitarian justification, actions within it can be justified only in ways that its rules specify (see Rawls, J.). This means, in particular, that the fact that breaking a promise would produce more happiness than keeping it does not count as a justification for breaking it. Practice-based accounts typically involve two parts, corresponding to (1) and (2) above: an account of the social practice of promising and an explanation of the moral authority of this practice. The latter component need not, however, be utilitarian, and in *A Theory of Justice* (1971) Rawls himself offers a nonutilitarian account of this kind. The moral component in this account is the 'principle of fairness', which specifies that someone who voluntarily accepts the benefits of a just social practice is bound to do his or her part in turn, as the rules of that practice specify. It follows, assuming that the rules of the practice of promising are just, that those who help themselves to the benefits of this practice by using it to stabilize agreements are obliged to keep the promises they have made (except where those rules recognize justification for doing otherwise).

On this account, promising is taken to be a particular social institution, on a par with other institutions set up to provide and distribute various 'public goods'. All such institutions face the threat of 'free riding' by people who would like to get the benefits without bearing their share of the costs, and Rawls' appeal to the 'principle of fairness' construes breaking a promise as an instance of the more general wrong of 'free riding' in this way (see Decision and game theory).

But it is not obvious that the wrongs involved in breaking a promise, or making a deceitful promise, are of this general type. They seem more clearly to be wrongs *to* the person to whom the promise is made rather than to all those who contribute to keeping the institution going. Nor is it clear that promising is a social practice of this kind - one that, like a recycling scheme, is aimed at a certain social purpose, defined by certain (somewhat arbitrary) rules, and exists just in so far as the members of a certain group accept these rules as normative. When we try to decide, in a difficult case, whether someone has good reason for failing to fulfil a promise, it does not seem that we are addressing our minds to a kind of social fact, about what the rules of some particular practice are. Moreover, it seems that making a deceitful promise, or failing, without good reason, to keep a promise sincerely made, are wrongs of a kind that can also be committed in other ways without reference to promises or to any other 'practice'. There are many other ways to induce people to rely on us, thereby making it wrong to disappoint them.

This suggests that the wrongs involved in breaking a promise, or in making a deceitful one, belong to a broader class of wrongs having to do with our responsibility for the expectations we create. Practices of agreement-making, where they exist, are one mechanism for creating such expectations, but not the only one. A number of writers have offered ‘expectation-based’ accounts of promising along these lines. Difficulties faced by such accounts include the problem of explaining how promises can create expectations to begin with without appealing to a notion of obligation that does not arise from expectation, and the problem of explaining how promises can be binding in cases in which no expectation is in fact created (because, for example, the ‘promise’ that is offered is not believed).

There are obvious similarities between the idea of a promise and that of a legal contract, but there are also important differences. The law of contracts attempts to specify exactly the conditions that must be fulfilled in order for a valid contract to be made. As noted above, however, it is not clear that the morality of promises involves an informal institution with definite rules of this kind. The law of contracts, which deals chiefly with commercial transactions, is concerned primarily with questions of whether compensation is owed and how much, rather than with the obligation to specific performance. In the case of promises, however, our concern is almost wholly with the obligation to perform rather than with determining compensation.

See also: Logic of ethical discourse §3; Trust; Truthfulness

T.M. SCANLON

References and further reading

Anscombe, G.E.M. (1978) ‘Rules, Rights and Promises’, in *Ethics, Religion and Politics, Collected Philosophical Papers*, vol. 3, Minneapolis, MN: University of Minnesota Press.(Argues that the reason someone has to rely on a promise depends on a prior notion of obligation.)

Atiyah, P.S. (1981) *Promises, Morals and the Law*, Oxford: Clarendon Press.(A study of the morality of promises from the point of view of an expert on the law of contracts.)

Cavell, S. (1979) *The Claim of Reason*, New York: Oxford University Press, ch. 9. (Critical discussion of Rawls’ account of promising as a practice.)

Fried, C. (1981) *Contract as Promise*, Cambridge, MA: Harvard University Press. (An account of the law of contracts, seen as based on a prior moral idea of promises.)

Hume, David (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1960, book 3, part II, sections 1, 2, 5.(Classic statement of the view that promises depend on a convention.)

MacCormick, N. (1972) ‘Voluntary Obligations and Normative Powers I’, *Proceedings of the Aristotelian Society*, supplementary volume 46: 59-78.(An account of promises based on expectation and reliance.)

Rawls, J. (1955) ‘Two Concepts of Rules’, *Philosophical Review* 64: 3-32.(Argues that promising is an action defined by a social practice and is therefore governed by the rules of that practice.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Defends, in §52, the principle of fairness as the basis of the obligation to keep a promise.)

Raz, J. (1972) ‘Voluntary Obligations and Normative Powers II’, *Proceedings of the Aristotelian Society*, supplementary volume 46: 79-101.(Criticizes MacCormick’s view.)

Scanlon, T.M. (1990) ‘Promises and Practices’, *Philosophy & Public Affairs* 19: 199-226.(Criticizes the idea that promises depend on a social practice and defends an account based on expectation and reliance. Expands on points made in this entry.)

Searle, J. (1964) ‘How to Derive "Ought" from "Is"’, *Philosophical Review* 73: 43-58; repr. in W.D. Hudson (ed.) *The Is/Ought Question*, London: Macmillan, 1969.(A purported derivation of an ‘ought’ from an ‘is’ using the idea of promising as a social institution.)

Proof theory

Proof theory is a branch of mathematical logic founded by David Hilbert around 1920 to pursue Hilbert's programme. The problems addressed by the programme had already been formulated, in some sense, at the turn of the century, for example, in Hilbert's famous address to the First International Congress of Mathematicians in Paris. They were closely connected to the set-theoretic foundations for analysis investigated by Cantor and Dedekind - in particular, to difficulties with the unrestricted notion of system or set; they were also related to the philosophical conflict with Kronecker on the very nature of mathematics. At that time, the central issue for Hilbert was the 'consistency of sets' in Cantor's sense. Hilbert suggested that the existence of consistent sets, for example, the set of real numbers, could be secured by proving the consistency of a suitable, characterizing axiom system, but indicated only vaguely how to give such proofs model-theoretically. Four years later, Hilbert departed radically from these indications and proposed a novel way of attacking the consistency problem for theories. This approach required, first of all, a strict formalization of mathematics together with logic; then, the syntactic configurations of the joint formalism would be considered as mathematical objects; finally, mathematical arguments would be used to show that contradictory formulas cannot be derived by the logical rules.

This two-pronged approach of developing substantial parts of mathematics in formal theories (set theory, second-order arithmetic, finite type theory and still others) and of proving their consistency (or the consistency of significant sub-theories) was sharpened in lectures beginning in 1917 and then pursued systematically in the 1920s by Hilbert and a group of collaborators including Paul Bernays, Wilhelm Ackermann and John von Neumann. In particular, the formalizability of analysis in a second-order theory was verified by Hilbert in those very early lectures. So it was possible to focus on the second prong, namely to establish the consistency of 'arithmetic' (second-order number theory and set theory) by elementary mathematical, 'finitist' means. This part of the task proved to be much more recalcitrant than expected, and only limited results were obtained. That the limitation was inevitable was explained in 1931 by Gödel's theorems; indeed, they refuted the attempt to establish consistency on a finitist basis - as soon as it was realized that finitist considerations could be carried out in a small fragment of first-order arithmetic. This led to the formulation of a general reductive programme.

Gentzen and Gödel made the first contributions to this programme by establishing the consistency of classical first-order arithmetic - Peano arithmetic (PA) - relative to intuitionistic arithmetic - Heyting arithmetic. In 1936 Gentzen proved the consistency of PA relative to a quantifier-free theory of arithmetic that included transfinite recursion up to the first epsilon number, ε_0*; in his 1941 Yale lectures, Gödel proved the consistency of the same theory relative to a theory of computable functionals of finite type. These two fundamental theorems turned out to be most important for subsequent proof-theoretic work. Currently it is known how to analyse, in Gentzen's style, strong subsystems of second-order arithmetic and set theory. The first prong of proof-theoretic investigations, the actual formal development of parts of mathematics, has also been pursued - with a surprising result: the bulk of classical analysis can be developed in theories that are conservative over (fragments of) first-order arithmetic.*

1 Metamathematics

Proof theory is a branch of mathematical logic founded by David Hilbert around 1920 to pursue his programme in the foundations of arithmetic (see Hilbert's programme and formalism). Consistency is the crucial logical notion connected with Hilbert's investigations. In traditional, Aristotelian logic it was viewed as a semantic notion: two or more statements are consistent if they are simultaneously true under some interpretation. In modern logic there is a syntactic definition that fits complex theories since Frege's *Begriffsschrift* (1879): a set of statements is consistent with respect to a logical calculus if no statement of the form $P \& \neg P$ is derivable from the statements by rules of the calculus. If these definitions are equivalent for a logic we have a significant fact, as the equivalence amounts to the completeness (and soundness) of the logic's system of rules. The first such completeness theorem was obtained for sentential logic by Paul Bernays (1918) and, independently, by Emil Post (1921); the completeness of predicate logic was proved by Kurt Gödel in 1930. The crucial step in such proofs shows that *syntactic* consistency implies *semantic* consistency. The converse is established quite directly, but involves the notion of truth. Here we have located one central issue that has motivated proof-theoretic investigations: namely, to avoid the uncritical use of the broad concept of classical truth for infinite mathematical structures.

Cantor applied consistency in an informal way to sets. He distinguished, for example in a letter to Dedekind, a consistent from an inconsistent multiplicity: the latter is such ‘that the assumption that all of its elements ‘‘are together’’ leads to a contradiction’, whereas the elements of the former ‘can be thought of without contradiction as ‘‘being together’’’. Cantor had also conveyed these distinctions by letter to Hilbert in 1897. (See Purkert and Ilgauds (1987) for the text of the letters and Sieg (1990) for the wider historical context.) Hilbert pointed out, both implicitly (1900) and explicitly (1905), that Cantor had not given a rigorous criterion for distinguishing between consistent and inconsistent multiplicities. Hilbert suggested remedying the problem for analysis and giving the proof of the ‘existence of the totality of real numbers or - in the terminology of G. Cantor - the proof of the fact that the system of real numbers is a consistent (complete) set’ by establishing the consistency of an axiomatic characterization of the real numbers (1900). Indeed, he claimed that consistency could be established ‘by a suitable modification of familiar methods’. Hilbert’s own hints, partly in unpublished lecture notes, together with remarks by Bernays, make it plausible that he had a model-theoretic proof in mind. This problematic can be traced back to considerations in Dedekind’s work (1888) as explicated carefully in his letter to Keferstein, where he asked of his notion ‘simply infinite system’, ‘does such a system *exist* at all in the realms of our ideas? Without a logical proof of existence it would always remain doubtful whether the notion of such a system might not perhaps contain internal contradictions’. Clearly, Dedekind tried to establish consistency by exhibiting a suitable ‘logical’, that is, set-theoretic, model.

In 1904 Hilbert began to pursue a completely different strategy for giving consistency proofs. His new way of proceeding was still aimed at securing the existence of sets, but intended to exploit the formalizability of the theory at hand. Formalizations had to satisfy requirements stricter than those imposed on the structure of theories by the traditional (Euclidean) axiomatic-deductive method. The additional requirement was the regimentation of inferential steps in proofs: not only did axioms have to be given in advance, but the rules representing steps in mathematical arguments had to be taken from a predetermined list. To avoid a regression in the definition of proof and to achieve intersubjectivity at an absolutely minimal level, the rules had to be ‘formal’ or ‘mechanical’ and had to depend only on the syntactic form of statements. Thus, to exclude any ambiguity, a precise and effectively described language was also needed to formalize particular theories; the indications (Hilbert 1905) were very sketchy, however.

The general kind of requirements had been clear to Aristotle and were explicitly formulated by Leibniz; but only Frege, in his *Begriffsschrift*, presented - in addition to an expressively rich language with relations and quantifiers - an adequate logical calculus. Through the formalization of mathematical proofs Frege pursued a clear philosophical aim, namely to recognize the ‘epistemological nature’ of theorems. In the introduction to his *Grundgesetze der Arithmetik* (see Frege, G.), he wrote: ‘By insisting that the chains of inference do not have any gaps we succeed in bringing to light every axiom, assumption, hypothesis or whatever else you want to call it on which a proof rests; in this way we obtain a basis for judging the epistemological nature of the theorem.’ An epistemological analysis of theorems was also aimed for by Hilbert in his work on the foundations of geometry, but done quite differently, relying on a more traditional axiomatic presentation and pushing forward genuinely metamathematical investigations. For the emerging proof-theoretic work the formal aspect Frege had emphasized would be exploited by Hilbert in a distinctively novel way.

Hilbert repeatedly gave courses on the foundations of mathematics in the period 1904 to 1917; the spirit of these lectures is captured in his Zurich talk ‘Axiomatisches Denken’ (1918). The mathematical development and philosophical clarification of a new consistency programme began to be given only in Hilbert’s lectures on *Prinzipien der Mathematik* (Principles of Mathematics) presented in the winter term of 1917/1918. These lectures mark the beginning of the fruitful collaboration with Bernays, who supported Hilbert’s preparation in essential ways and also wrote careful notes. The notes of these and later lectures are amazing documents, as one finds in them for the first time a detailed modern presentation of the syntax and informal semantics of predicate logic and of finite type theories. The 1917/1918 notes served as the basis for Hilbert and Ackermann’s *Grundzüge der theoretischen Logik* (1928); indeed, with minor exceptions the notes contain all the material presented in the book. In lectures during the summer term of 1920 and the winter term of 1921/1922, proof theory and the ‘finitist consistency programme’ emerged.

2 Hilbert’s programme

For the purpose of the consistency programme, metamathematics was taken in Hilbert’s 1921/1922 lectures to be included in, if not coextensive with, the part of mathematics acceptable to constructivists such as Kronecker and Brouwer. The point of consistency proofs was no longer to guarantee the existence of sets, but to establish the instrumental usefulness of classical mathematical theories *T* - say, set theory - with respect to finitist mathematics. That focus rested on the observation that the statement formulating the consistency of *T* is equivalent to the reflection principle

$$Pr(a, `\sigma') \Rightarrow \sigma.$$

Here, *Pr* is the finitist proof predicate for *T*, σ a finitistically meaningful statement, and ‘σ’ its translation into the language of *T*. A finitist consistency proof for *T* would thus ensure that *T* is a reliable instrument for the proof of finitist statements. Other important metamathematical issues were the completeness and decidability of theories.

The formalizability of mathematics was obviously crucial for this proof-theoretic approach, and the programmatic goal was seen as a way to circumvent some philosophical issues, for example, concerning the nature of infinite sets in the case of set theory. For these reasons Hilbert’s philosophical position is (still) frequently equated with formalism in the sense of Frege’s articles ‘Über die Grundlagen der Geometrie’ (1903, 1906; see Frege, G.) and of Brouwer’s inaugural address ‘Intuitionism and Formalism’ (1913). Such a view is not completely unsupported by some of Hilbert’s polemical remarks during the 1920s, but on balance his philosophical views developed into a sophisticated instrumentalism, if that label is taken in Ernest Nagel’s judicious sense (see Nagel 1961). Hilbert’s is an instrumentalism emphasizing the contentual motivation of mathematical theories; that is perhaps most clearly expressed in the first chapter of Hilbert and Bernays (1934). A sustained philosophical analysis of proof-theoretic research in the context of broader issues in the philosophy of mathematics was provided by Bernays; his penetrating essays stretch over five decades (see Bernays 1976). Feferman (1988) and Sieg (1988) give complementary accounts; for altogether contrasting approaches, see Detlefsen (1986) and Simpson (1988).

The mathematical work is most remarkable for what it started, as it constitutes the beginnings of modern mathematical logic. Even before the work of Gödel and Gentzen it was rich in accomplishments: consider, as examples, Bernays’ completeness proof for sentential logic; the partial solutions to the decision problem (Hilbert’s *Entscheidungsproblem*) for predicate logic obtained by Behmann, Bernays, Schönfinkel and Herbrand; and the consistency proofs given by Ackermann, von Neumann and Herbrand. Taking for granted the broader conceptual clarifications and the focus on first-order logic, Herbrand’s ‘*théorème fondamentale*’ is perhaps the most significant result on purely logical grounds, but also because of its applicability in consistency proofs. Gödel pointed to its essence when he gave in a 1933 lecture the following formulation of Herbrand’s theorem: ‘If we take a theory which is constructive in the sense that each existence assertion made in the axioms is covered by a construction, and if we add to this theory the non-constructive notion of existence and all the logical rules concerning it, for example, the law of excluded middle, we shall never get into any contradiction’ (see Herbrand’s theorem).

The results obtained in the 1920s, including Herbrand’s theorem, were disappointing when measured against the hopes and ambitions of the Hilbert school: Ackermann, von Neumann and Herbrand had established essentially the consistency of first-order arithmetic with a very restricted principle of induction (for quantifier-free formulas). Actual limits on finitist considerations for consistency proofs had been reached; that became clear in 1931 through Gödel’s theorems and the realization that finitist proofs could be formalized in a weak fragment of number theory (see Gödel’s theorems). Initially, Gödel did not share the view on the limits of finitist reasoning, as is clear from the final remarks in his ‘Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I’ (On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems) (1931) and contemporaneous correspondence with von Neumann and Herbrand. But by December 1933, when he lectured in Cambridge, Massachusetts, he had changed his position: having isolated a ‘system A’ - essentially a version of primitive recursive arithmetic (PRA) - he made the following claim: ‘Now all the intuitionistic [that is, finitist] proofs complying with the requirements of the system A which have ever been constructed can easily be expressed in the system of classical analysis and even in the system of classical arithmetic, and there are reasons for believing that this will hold for any proof which one will ever be able to construct.’

The solvability of the *Entscheidungsproblem* had been made implausible by Gödel’s results, but the actual proof of unsolvability had to wait until 1936 for a conceptual clarification of ‘mechanical procedure’ or ‘algorithm’. Such a

clarification was achieved mainly through the work of Church and Turing (see Church's thesis; Computability theory). A precise notion of mechanical procedure was also needed to prove the incompleteness theorems for *general* 'formal' theories satisfying basic representability and derivability conditions; after all, Gödel had established limits only for (formalizations of) particular theories, such as the system of *Principia Mathematica* and then current axiomatic set theories. In his attempt to characterize a proper extension of the class of primitive recursive functions, Gödel introduced (in his Princeton lectures of 1934) the general recursive functions through an equational calculus (see Computability theory §2). The informal concept underlying Gödel's and also Church's approach - calculability of functions in a 'formal' calculus - was carefully analysed by Hilbert and Bernays (1939: supplement 2); they formulated recursiveness conditions for general deductive formalisms and showed that the number-theoretic functions whose values can be calculated in formalisms satisfying these conditions are exactly the general recursive ones.

The impact of the incompleteness theorems on Hilbert's programme was profound and yet limited. On the one hand, as remarked above, they pointed out definite limits of finitist considerations; on the other hand, they left open the possibility of modifying the programme to a *general reductive programme* that was no longer aiming for 'absolute' finitist consistency proofs, but rather for consistency proofs relative to 'appropriate' constructive theories. Before Gödel's results were known, Bernays (1930) had given a detailed and searching analysis of the philosophical aims of Hilbert's proof theory; in the *Postscriptum* published in 1976, Bernays expressed clearly what was lost due to the incompleteness theorems: the sharp distinction between what is intuitive and what is non-intuitive, a distinction that was basic for the proposed philosophical treatment of the problem of the infinite. Thus, it is this particular 'solution' to a philosophical problem that was shown to be impossible.

Work in proof theory continued with the explicit goal of achieving relative consistency proofs. Such work is in a venerable mathematical tradition, as the many examples of significant results show: for example, the consistency of non-Euclidean relative to Euclidean geometry; that of Euclidean geometry relative to analysis; the consistency of set theory with the axiom of choice relative to set theory without it; that of set theory with the negation of the axiom of choice relative to set theory. The mathematical significance of relative consistency proofs is often brought out by sharpening them to conservative extension results. Such results may ensure, for example, that the theories have the same class of provably total functions (see §6). However, the initial motivation for such arguments is most frequently philosophical: one wants to guarantee the coherence of the original theory on an epistemologically distinguished basis. One has to see the specific results that have been obtained in the pursuit of the general reductive programme from this perspective.

3 Mathematical work

The development of the general reductive programme is characterized by modifications of the two-pronged approach of Hilbert's original programme, namely (1) weakening the theories in which parts of mathematics are formalized, and (2) strengthening the theories in which the metamathematical considerations are carried out and, consequently, relative to which constructive consistency proofs can be given. The first modification reaches back to Weyl's *Das Kontinuum* (1918) and culminated in the 1970s, when it was realized that the classical results of mathematical analysis can be obtained in conservative extensions of first-order arithmetic. The second modification started with the work of Gödel and Gentzen in 1933, when they established independently the consistency of classical arithmetic relative to intuitionistic arithmetic; it led in the 1970s and 1980s to consistency proofs of subsystems of second-order arithmetic or, synonymously, subsystems of analysis relative to intuitionistic theories of constructive ordinals. Obviously, only a sketch of some main results can be attempted here.

Second-order arithmetic has been used for some time as a framework for the formal development of classical mathematical analysis; that is, the theory of the continuum set-theoretically described by Dedekind and Cantor. Because of this mathematical adequacy, second-order arithmetic and a variety of subsystems have been thoroughly investigated. The main set-theoretic principles for these systems are the comprehension axiom (CA)

$$(\exists X)(\forall y)(y \in X \Leftrightarrow S(y))$$

and the axiom of choice (AC) in the form

$$(\forall x)(\exists Y)S(x,Y) \Rightarrow (\exists Z)(\forall x)S(x,Z_x),$$

where *S* is in each case an arbitrary formula of the language and thus may contain set quantifiers; $y \in Z_x$ is defined to mean $\langle y, x \rangle \in Z$. These principles are impredicative, as the sets *X* and *Z* whose existence is postulated are in general characterized by reference to all sets of natural numbers. The induction principle is formulated either as a schema or as a second-order axiom

$$(\forall X)(0 \in X \;\&\; (\forall y)(y \in X \Rightarrow y' \in X) \\ \Rightarrow (\forall x) x \in X).$$

Theories are denoted by the name of their set-existence principle enclosed in parentheses; thus (CA) names full analysis. If $\restriction$ follows its name, a theory uses the second-order axiom to formalize induction. Two general results are of interest, as they show that second-order arithmetic has a certain robustness: (CA) is proof-theoretically equivalent to Zermelo-Fraenkel set theory without the power-set axiom; (AC) is conservative over (CA) for Π^1_4-formulas and properly stronger, as there is a Π^1_2-instance of AC not provable in (CA). In the presence of full CA, the theories with the induction schema, respectively, the second-order axiom of induction are equivalent, but they can be of strikingly different strength, when the set-existence principles are restricted; for example, $(\Pi^0_\infty\text{-CA})\restriction$ is conservative over Peano arithmetic (PA), whereas $(\Pi^0_\infty\text{-CA})$ proves the consistency of PA.

In 1917/1918 Hilbert and Bernays used ramified type theory with the axiom of reducibility to develop analysis; the published presentation (1939: supplement 4) is based on this early work, but employs full second-order arithmetic as the formal framework. They encouraged developments with restricted means already in 1920, writing of Brouwer and Weyl:

> The positive and fruitful part of the investigations into the foundations of mathematics carried out by these two researchers fits into the mould of the axiomatic method and is exactly in the spirit of this method. For one investigates here, how a part of analysis can be delimited by a certain narrower system of assumptions.
>
> (Hilbert and Bernays 1920: 34)

Subsystems of analysis are now mainly defined by restricting *S* in the set-existence schemata to particular classes of formulas. The set-theoretic demands can be reduced dramatically: Hilbert and Bernays' presentation (1939), for example, can be given quite readily in $(\Pi^1_1\text{-CA})\restriction$. Strictly mathematical work continued to accompany work on consistency proofs for subsystems; it had the aim of establishing the mathematical significance of subsystems and made use of work in the constructivist tradition.

By the mid-1970s, through final efforts of Feferman, Friedman and Takeuti, it was clear that classical analysis could be carried out in conservative extensions of number theory, for example, $(\Pi^0_\infty\text{-CA})\restriction$. In this context Friedman suggested pursuing a strategy familiar from investigations of the axiom of choice in set theory; namely, to establish the equivalence of certain set-existence principles with mathematical theorems. This theme is played with surprising variations in Friedman's and Simpson's work on subsystems and gave rise to the enterprise of 'reverse mathematics'. To mention just two, by now classic, examples: CA for arithmetic formulas is equivalent to the theorem that every bounded sequence of reals has a least upper bound and to König's lemma. Friedman introduced a second-order theory WKL_0 that extends primitive recursive arithmetic conservatively for Π^0_2-formulas; the theory is weak, but still provides a very good basis for developing parts of analysis and algebra. Simpson (1988) not only considers the development of mathematics in such a conservative extension of PRA a 'reductionist programme', but equates it with Hilbert's programme. One should recall, however, that Hilbert did not propose to redo all of mathematics in (a conservative extension of) PRA, but rather to justify - via finitist consistency proofs - the use of strong classical theories sufficient for the direct formalization of mathematical practice.

4 Logical tools

Hilbert's central idea for the metamathematical treatment of the consistency problem found its expression in what is known as the ε-calculus and the associated substitution method. Hilbert and Bernays (1939) present what was achieved in its terms; even Herbrand's (difficult) work was recast in terms of the ε-calculus. This tradition was kept up by Tait (1965) and more recently by Mints (1994). However, other logical tools turned out to be more useful for proof-theoretic investigations: Gentzen's sequent calculi (see Natural deduction, tableau and sequent systems §2) and Gödel's so-called *Dialectica* interpretation. In the latter, Gödel used computable functionals of

finite type to obtain a reduction of intuitionistic arithmetic (HA); joining this reduction with the consistency of PA relative to HA, a consistency proof relative to the system of these functionals was obtained. Influenced by the considerations of Hilbert (1926), Gödel presented this work in a lecture at Yale in 1941, but published it only in 1958 (see Gödel's theorems). Spector (1962) gave a consistency proof for full classical analysis using bar recursive functionals of finite type; this proof prompted a searching analysis in the *Stanford Report on the Foundations of Analysis* (1963).

Here the focus is on sequent calculi, as they have found the most extensive use and widest applicability. In the form given to them by Tait (1968), they allow the proof of finite sets of formulas built up from literals (atomic formulas or negations of such), conjunction, disjunction, universal and existential quantification and - depending on the theory - infinitary conjunction and disjunction. Thus, for just first-order logic the basic logical symbols are '$\wedge$', '$\vee$', '$\exists$', '$\forall$' and the rules of the calculi include the following ones, where Γ is used as a syntactic variable ranging over finite sets of formulas, Γ, ϕ stands for the union of Γ and the singleton ϕ, and $a \in P(\Gamma)$ means that the parameter *a* occurs in one of the formulas in Γ:

LA: $\quad \Gamma, \varphi, \neg\varphi, \quad \varphi$ atomic

$\wedge$: $$\frac{\Gamma, \varphi_0 \quad \Gamma, \varphi_1}{\Gamma, \varphi_0 \wedge \varphi_1}$$

$\vee_i$: $$\frac{\Gamma, \varphi_i}{\Gamma, \varphi_0 \vee \varphi_1}, \quad i = 0, 1$$

C: $$\frac{\Gamma, \varphi \quad \Gamma, \neg\varphi}{\Gamma}$$

$\forall$: $$\frac{\Gamma, \varphi a}{\Gamma, (\forall x)\varphi x} \quad a \notin P(\Gamma)$$

$\exists$: $$\frac{\Gamma, \varphi t}{\Gamma, (\exists x)\varphi x}$$

The crucial claim established by Gentzen was his '*Hauptsatz*' or cut-elimination theorem: every derivation in the logical system using the cut rule *C* can be transformed into a cut-free or normal derivation. Inspecting the rules, one notices that the premises of all the rules (except for *C*) contain only subformulas of formulas in the conclusion. Consequently, a normal derivation of a sequent Γ contains only subformulas of elements in Γ. This is the crucial subformula property of normal derivations, providing a bound on the complexity of formulas that can occur in a proof of Γ. In Gentzen (1935) this metamathematical fact is established and used to obtain in a most perspicuous way the most far-reaching consistency result that had then been obtained, namely Herbrand's.

For full first-order number theory PA this treatment was extended by Gentzen (1936) to a partial cut-elimination argument whose termination was established by quantifier-free transfinite induction up to the first epsilon number - $\mathrm{TI}(\varepsilon_0)$. Gentzen showed (1943) that this induction schema for every ordinal α less than ε_0 can be established in PA; this is the first 'ordinal analysis' of a formal theory. In the 1950s Lorenzen and, much more extensively, Schütte used infinitary extensions of Gentzen's finitary systems; in particular for the treatment of PA they used the so-called ω-rule that allows one to infer $\Gamma, (\forall x)\phi x$ from the premises $\Gamma, \phi n$ for each natural number *n* (see Non-constructive rules of inference §1). Though derivations are now infinite, PA-derivations can be embedded into finitistically described ones and can be transformed effectively into cut-free derivations; the natural ordinal length of these derivations is bounded by ε_0. Schütte extended these methods to treat systems of ramified analysis RA_α (of order α) and obtained, in particular, ordinal bounds on the length of normal derivations in terms of the Veblen hierarchy of ordinal functions. This work was used in 1963 by Feferman and Schütte independently to characterize the ordinal Γ_0 of predicative analysis, that is, the first ordinal α such that $\mathrm{TI}(\alpha)$ cannot be proved in RA_β for β less than α.

5 Reductive results

The mathematical work described in §3 was to a large extent inspired by the attempt to establish the significance of relative consistency proofs and/or to focus on manageable subsystems that might be a proper next target for such proofs. Most of the reductive results mentioned below can be established by the metamathematical tool of the sequent calculi, though some of the original proofs used different techniques.

The Gödel-Gentzen result for number theory can be extended to ramified analysis and also to obtain conservativeness for Π^0_2-formulas. This provides a relative consistency proof, as ramified systems with intuitionistic logic are certainly acceptable constructively (as long as the ordinals along which the systems are iterated are acceptable). In the early 1960s, partly through the study of predicativity, significant subsystems with *S* restricted to small classes of analytic formulas were isolated, among them $(\Sigma^1_1\text{-AC})$ and $(\Delta^1_1\text{-CA})$. Kreisel pointed out that

$(\Sigma^1_1\text{-AC}) \supseteq (\Delta^1_1\text{-CA})$.

Friedman showed that $(\Sigma^1_1\text{-AC})$ is conservative over $(\Delta^1_1\text{-CA})$ for Π^1_2-formulas; that the inclusion is proper was established later by Steel. Indeed, $(\Sigma^1_1\text{-AC})$ is conservative for Π^1_2-formulas over $(\Pi^1_0\text{-CA})_{<\varepsilon_0}$, a theory based on the transfinite iteration of the jump-operator and equivalent to ramified analysis of level less than ε_0. This result, due to Friedman, allowed the determination of the proof-theoretic ordinal of the systems, but it also showed - and this was quite unexpected - that these prima facie impredicative theories were predicative in the sense of Feferman and Schütte. Feferman (1964) and Kreisel (1968) provide excellent summaries.

Friedman's theorem for $(\Sigma^1_1\text{-AC})$ turned out to be a special case of a general result: $(\Sigma^1_{n+1}\text{-AC})$ is conservative over $(\Pi^1_n\text{-CA})_{<\varepsilon_0}$ for formulas in F_n, where F_0 is Π^1_2, F_1 is Π^1_3, and F_n is Π^1_4 for $n > 1$ (see Friedman 1970). As just explained, the case $n = 0$ was of special interest for the study of predicativity; the case $n = 1$ is also to be placed in the context of foundational investigations. The most immediate context is provided by Feferman (1970), who relates the systems $(\Pi^1_1\text{-CA})_{<\varepsilon_0}$ to the classical theory for (less than ε_0 times iterated) classes defined by generalized inductive definitions. Well-known examples are the classes O of constructive ordinals and W of recursive well-founded trees. Between them, Friedman and Feferman reduced the subsystems $(\Sigma^1_2\text{-AC})$ and $(\Delta^1_2\text{-CA})$ to the classical theory of the tree classes $\mathbf{W}_\nu$ with index ν less than ε_0. Feferman (1977) gives a detailed survey of mathematical and proof-theoretic work, including these last results.

Kreisel had introduced intuitionistic theories of iterated inductive definitions in the *Stanford Report* (1963); these theories were viewed as codifying constructive principles that might be used in consistency proofs for subsystems of analysis. Feferman's and Friedman's work described above established connections between subsystems of analysis and classical theories of inductive definitions, making a crucial step towards answering the major problem posed by Kreisel (1968): reduce $(\Sigma^1_2\text{-AC})$ to a constructive theory of inductive definitions. That would provide, as Kreisel put it, 'a solution to Hilbert's problem for the subsystem of analysis… $(\Sigma^1_2\text{-AC})$'. The classical theories for classes defined by generalized inductive definitions are reducible to intuitionistic theories for accessible such classes. This allowed, in particular, a satisfactory solution of Kreisel's open problem, as $(\Sigma^1_2\text{-AC})$ was shown to be reducible to $(\mathrm{ID})_{<\varepsilon_0}(\mathbf{O})$, the intuitionistic theory of constructive number classes with index less than ε_0. These and many further results were obtained by Buchholz *et al*. (1981); in particular, the ordinal analysis of the subsystems at hand. This work was influenced by earlier considerations of Howard, Tait and Takeuti.

Subsystems of set theory, in particular of admissible set theory, were used by Jäger and Pohlers in the early 1980s to provide a unifying approach to the investigations and, it was hoped, an avenue for analysing even stronger systems than those corresponding to the subsystems mentioned above (see Jäger 1986). This has indeed been a successful strategy and reveals, in the work of Rathjen and others, a deep connection between large cardinals and the constructive ordinals needed for the proof-theoretic investigation of such systems. Rathjen (1995) succeeded in analysing $(\Pi^1_2\text{-CA})$; it seems that the techniques developed for this case might allow the treatment of full analysis. These proof-theoretic investigations can no longer be motivated by the concern of 'securing' mathematical practice: the systems that are investigated are much stronger than needed for practice; the constructive ordinals used in the metamathematical theory are obtained in analogy to large cardinals in set theory.

The systems WKL_0, $(\Pi^0_\infty\text{-CA})\restriction$ and $(\Pi^1_1\text{-CA})\restriction$ have been recognized as significant for the formalization of

mathematical practice and are reducible to theories based on principles that are acceptable from constructive positions; after all, they are conservative for Π^0_2-formulas over PRA, HA and the intuitionistic theory for the finite constructive number classes $(\mathrm{ID})_{<\omega}(\mathbf{O})$. This provides a coherent perspective bringing out the complementary character of mathematical and metamathematical work that ultimately aims at relating significant parts of mathematical practice to distinctive foundational positions. But there is no obvious answer to the question 'What is the mathematical significance of those subsystems, when taken as vehicles for the formal axiomatic study of ordinary mathematics?'; similarly there is no obvious answer to the question 'What is the philosophical significance of the corresponding systems PRA, HA and $(\mathrm{ID})_{<\omega}(\mathbf{O})$, when taken as formal expressions of foundational positions?'. There is ample room for reflection on these questions; the work reported here and in the literature provides rich and crucial data (see Sieg 1990). For the philosophical reflection on the foundations of mathematics the investigations of subsystems of set theory provide additional, significant material: what are constructions that lead to 'accessible domains' and how is it that we recognize their associated laws?

6 Outlook

'Internal' mathematical and philosophical challenges of work in proof theory were sketched at the end of the previous section. However, the foundational goals of proof-theoretic investigations have been complemented over the last few decades by other important initiatives.

First, in the 1950s, Kreisel initiated work that was to exploit the gap between provability in particular formal theories and truth. That led, on the one hand, to 'global' characterizations of the provably total functions of theories and to related independence results. On the other hand, by attending 'locally' to mathematical details of proofs and by using proof-theoretic techniques, it led to explicit computational information of mathematical significance. This seems to have come to fruition through work by Luckhardt (1989) and Kohlenbach (1995).

Second, methods and results of mathematical logic, but in particular of proof theory, have played an increasing role in computer science. Clearly, there has also been significant and stimulating influence in the other direction. Various (type) systems - for example, Martin-Löf's, Girard's system F, and Feferman's systems of explicit mathematics - have been used for the presentation of proofs and computations, but also for describing transformations on them.

Third, there is a direct connection to the general topic of theorem proving; investigations here, when focusing on automated proof search, might reflect back into proof theory by providing data for a 'structural theory of (mathematical) proofs'. Such a structural proof theory would go beyond the representation of proofs in formal theories and articulate search heuristics expressing leading mathematical ideas for particular parts of mathematics. Saunders Mac Lane suggested such an extension of proof-theoretic investigations in his Göttingen dissertation of 1934, but it has been pursued only recently through computer implementation of appropriate search procedures.

See also: Hilbert's programme and formalism; Logical and mathematical terms, glossary of

WILFRIED SIEG

References and further reading

Bernays, P. (1918) 'Beiträge zur axiomatischen Behandlung des Logik-Kalküls' (Contributions to the Axiomatic Treatment of the Logical Calculus), unpublished *Habilitationsschrift*, Göttingen University; published in much revised form, 'Axiomatische Untersuchung des Aussagen-Kalküls der *Principia Mathematica*' (Axiomatic Investigation of the Propositional Calculus of *Principia Mathematica*), *Mathematische Zeitschrift* 25: 305-20, 1926.(The first proof of the completeness of a calculus for sentential logic.)

Bernays, P. (1930) 'Die Philosophie der Mathematik und die Hilbertsche Beweistheorie' (Philosophy of Mathematics and Hilbert's Proof Theory), *Blätter für Deutsche Philosophie* 4: 326-67; repr. in *Abhandlungen zur Philosophie der Mathematik (Essays on the Philosophy of Mathematics)*, Darmstadt: Wissenschaftliche Buchgesellschaft, 1976.(A searching analysis of the background for and the goals of Hilbert's proof theory.)

Bernays, P. (1976) *Abhandlungen zur Philosophie der Mathematik (Essays on the Philosophy of Mathematics)*, Darmstadt: Wissenschaftliche Buchgesellschaft.(Five decades of Bernays' essays.)

Brouwer, L.E.J. (1913) 'Intuitionism and Formalism', *Bulletin of the American Mathematical Society* 20: 81-96; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 77-89.(A formalism with which Hilbert's philosophical position is

frequently equated.)

Buchholz, W., Feferman, S., Pohlers, W. and Sieg, W. (1981) *Iterated Inductive Definitions and Subsystems of Analysis*, Berlin, Heidelberg and New York: Springer.(First proof-theoretic treatment of some impredicative subsystems of analysis.)

Davis, M. (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Hewlett, NY: Raven Press.(Anthology of the fundamental papers on the subject by Gödel, Church, Turing, Kleene, Rosser and Post.)

Dedekind, R. (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; trans. W.W. Beman (1901), 'The Nature and Meaning of Numbers', in *Essays on the Theory of Numbers*, New York: Dover, 1963.(*The* modern set-theoretic foundation of number theory.)

Detlefsen, M. (1986) *Hilbert's Program: An Essay on Mathematical Instrumentalism*, Boston, MA, and Dordrecht: Reidel.(A contrasting approach to those of Feferman (1988) and Sieg (1988).)

Feferman, S. (1964) 'Systems of Predicative Analysis', *Journal of Symbolic Logic* 29: 1-30.(An excellent summary of work described in §5.)

Feferman, S. (1970) 'Formal Theories for Transfinite Iterations of Generalized Inductive Definitions and Some Subsystems of Analysis', in J. Myhill, R.E. Vesley and A. Kino (eds) *Intuitionism and Proof Theory*, Amsterdam: North Holland.(See §5.)

Feferman, S. (1977) 'Theories of Finite Type Related to Mathematical Practice', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland.(A detailed survey of proof-theoretic work.)

Feferman, S. (1988) 'Hilbert's Program Relativized: Proof-Theoretical and Foundational Reductions', *Journal of Symbolic Logic* 53: 364-84.(A philosophical analysis of proof-theoretic research.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82.(This small booklet contains the most significant step from Aristotelian to modern logic.)

Friedman, H. (1970) 'Iterated Inductive Definitions and Σ^1_2-AC', in J. Myhill, R.E. Vesley and A. Kino (eds) *Intuitionism and Proof Theory*, Amsterdam: North Holland.(See §5.)

Gentzen, G. (1935) 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* 39: 176-210, 405-565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969; trans. 'Investigations into Logical Deduction', *American Philosophical Quarterly* 1: 288-306, 2: 204-18, 1964; repr. in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 68-131.(Gentzen's doctoral dissertation, including the calculi of natural deduction and sequents, and the consistency proof for arithmetic with quantifier-free complete induction.)

Gentzen, G. (1936) 'Die Widerspruchsfreiheit der reinen Zahlentheorie', *Mathematische Annalen* 112: 493- 565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1967; trans. 'The Consistency of Elementary Number Theory', in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 132-213.(See §4.)

Gentzen, G. (1943) 'Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie', *Mathematische Annalen* 119 (1): 140-61; trans. 'Provability and Nonprovability of Restricted Transfinite Induction in Elementary Number Theory', in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 287-308.(The first 'ordinal analysis' of a formal theory; see §4.)

Gödel, K. (1986-95) *Kurt Gödel: Collected Works*, ed. S. Feferman *et al.*, New York and Oxford: Oxford University Press, 3 vols.(Includes all the individual papers mentioned in the text.)

Hilbert, D. (1900) 'Über den Zahlbegriff' (On the Concept of Number), *Jahresbericht der Deutschen Mathematiker-Vereinigung* 8: 180-94; revised as appendix 6 of *Die Grundlagen der Geometrie*, Leipzig and Berlin: Teubner, 7th edn, 1930; trans. *The Foundations of Geometry*, Chicago, IL: Open Court, 2nd edn, 1971. (Hilbert's first paper addressing the foundations of arithmetic.)

Hilbert, D. (1905) 'Über die Grundlagen der Logik und der Arithmetik', in A. Krazer (ed.) *Verhandlungen des dritten internationalen Mathematiker-Kongreßes in Heidelberg vom 8. bis 13. August 1904*, Leipzig: Teubner, 174-85; trans. 'On the Foundations of Logic and Arithmetic', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 130-8. (Discusses the need for a greater formalization of proof theory.)

Hilbert, D. (1918) 'Axiomatisches Denken' (Axiomatic Thinking), *Mathematische Annalen* 78: 405-15; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer, 1935.(Captures the spirit of Hilbert's lecture courses on

the foundations of mathematics.)

Hilbert, D. (1920) ‘Probleme der mathematischen Logik’ (Problems of Mathematical Logic), lecture notes available in the Hilbert Nachlaß at the University of Göttingen, 46 pages.(Lecture notes from the summer term of 1920, written by N. Schönfinkel and P. Bernays.)

Hilbert, D. (1926) ‘Über das Unendliche’, *Mathematische Annalen* 95: 161-90; trans. ‘On the Infinite’, in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.(Fully developed formulation of the consistency programme and finitist standpoint; ‘proof’ of Cantor’s continuum theorem.)

Hilbert, D. and Ackermann, W. (1928) *Grundzüge der theoretischen Logik (Principles of Theoretical Logic)*, Berlin: Springer, 2nd edn, 1938; trans. L.M. Hammond, G.G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, ed. R.E. Luce, New York: Chelsea, 1950.(Based on Hilbert’s lectures in the winter term of 1917/1918.)

Hilbert, D. and Bernays, P. (1934, 1939) *Grundlagen der Mathematik (Foundations of Mathematics)*, Berlin: Springer, 2 vols; 2nd edn, 1968, 1970.(Classic work of mathematical logic; includes valuable statements of Hilbert’s conception of the axiomatic method and finitary reasoning.)

Jäger, G. (1986) *Theories for Admissible Sets: A Unifying Approach to Proof Theory*, Naples: Bibliopolis.(Uses subsystems of set theory to provide a unifying approach to the investigation of consistency proofs.)

Kohlenbach, U. (1995) *Real Growth in Standard Parts of Analysis*, *Habilitationsschrift*, Frankfurt University. (Develops work initiated by Kreisel, exploiting the gap between provability and truth; see §6.)

Kreisel, G. (ed.) (1963) *Stanford Report on the Foundations of Analysis*, unpublished seminar report, Stanford University.(Important papers (some published elsewhere) on subsystems of analysis that emerged from a seminar held at Stanford in the summer of 1963; the contributors are J. Harrison, W.A. Howard, G. Kreisel, R. Parikh and W.W. Tait.)

Kreisel, G. (1968) ‘A Survey of Proof Theory’, *Journal of Symbolic Logic* 33: 321-88.(An excellent summary of work described in §5.)

Luckhardt, H. (1989) ‘Herbrand Analysen zweier Beweise des Satzes von Roth: Polynomiale Anzahlschranken’ (Herbrand Analyses of Two Proofs of Roth’s Theorem: Polynomial Bounds), *Journal of Symbolic Logic* 54: 234-63.(Develops work initiated by Kreisel, exploiting the gap between provability and truth; see §6.)

Mints, G. (1994) ‘Gentzen-Type Systems and Hilbert’s Epsilon Substitution Method I’, in D. Prawitz, B. Skyrms and D. Westerståhl (eds) *Proceedings of the Ninth International Conference, Uppsala, Sweden, 1991*, vol. 9, *Logic, Methodology and Philosophy of Science*, Amsterdam: North Holland.(Continues use of the ε-calculus used by Hilbert; see §4.)

Nagel, E. (1961) *The Structure of Science - Problems in the Logic of Scientific Explanation*, New York: Harcourt Brace.(Hilbert’s views developed into instrumentalism in Nagel’s sense.)

Post, E. (1921) ‘Introduction to a General Theory of Elementary Propositions’, *American Journal of Mathematics* 43: 163-85; repr. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematics, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 264-83.(Proves the completeness of a calculus for sentential logic.)

Purkert, W. and Ilgauds, H.J. (1987) *Georg Cantor 1845-1918*, Basle: Birkhäuser.(A biography of Cantor that includes a number of very informative letters from Cantor to Hilbert from the late 1890s.)

Rathjen, M. (1995) ‘Recent Advances in Ordinal Analysis: $(\Pi^1_2\text{-CA})$ and Related Systems’, *Bulletin of Symbolic Logic* 1: 468-85.(Analysis of $(\Pi^1_2\text{-CA})$.)

Schwichtenberg, H. (1977) ‘Proof Theory: Some Applications of Cut-Elimination’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 867-95.(An elegant presentation of the proof-theoretic analysis of elementary number theory, PA.)

Sieg, W. (1988) ‘Hilbert’s Program Sixty Years Later’, *Journal of Symbolic Logic* 53: 338-48.(A philosophical analysis of proof-theoretic research.)

Sieg, W. (1990) ‘Relative Consistency and Accessible Domains’, *Synthèse* 84: 259-97.(Provides a historical and philosophical perspective on current proof-theoretic work.)

Simpson, S. (1988) ‘Partial Realization of Hilbert’s Program’, *Journal of Symbolic Logic* 53: 349-63.(A contrasting approach to those of Feferman (1988) and Sieg (1988).)

Spector, C. (1962) ‘Provably Recursive Functionals of Analysis: A Consistency Proof of Analysis by an Extension of Principles Formulated in Current Intuitionistic Mathematics’, in *Recursive Function Theory*, Proceedings of Symposia of Pure Mathematics, Providence, RI: American Mathematical Society.(A consistency proof for full classical analysis.)

Tait, W.W. (1965) 'The Substitution Method', *Journal of Symbolic Logic* 30: 175-92.(Continues use of the ε-calculus used by Hilbert; see §4.)

Tait, W.W. (1968) 'Normal Derivability in Classical Logic', in J. Barwise (ed.) *The Syntax and Semantics of Infinitary Languages*, New York: Springer, 204-36.(Introduces and uses an elegant and technically convenient formulation of sequent calculi for the investigation of classical logic and subsystems of analysis.)

Weyl, H. (1918) *Das Kontinuum: Kritische Untersuchungen über die Grundlagen der Analysis*, Leipzig: Veit; trans. T. Bole, *The Continuum: A Critical Examination of the Foundations of Analysis*, New York: Dover, 1994.(The first modification of Hilbert's original programme.)

Proper names

The Roman general Julius Caesar was assassinated on 14 March 44 BC by conspirators led by Brutus and Cassius. It is a remarkable fact that, in so informing or reminding the reader, the proper names 'Julius Caesar', 'Brutus' and 'Cassius' are used to refer to three people each of whom has been dead for about two thousand years. Our eyes could not be used to see any of them, nor our voices to talk to them, yet we can refer to them with our words.

The central philosophical issue about proper names is how this sort of thing is possible: what exactly is the mechanism by which the user of a name succeeds in referring with the name to its bearer? As the example indicates, whatever the mechanism is, it must be something that can relate the use of a name to its bearer even after the bearer has ceased to exist.

*In modern philosophy of language there are two main views about the nature of the mechanism. On one account, which originated with Frege, a use of a name expresses a conception or way of thinking of an object, and the name refers to whatever object fits, or best fits, that conception or way of thinking. Thus with 'Cassius', for example, I may associate the conception 'the conspirator whom Caesar suspected because of his size' (recalling a famous speech in Shakespeare's Julius Caesar). Conception theories are usually called 'sense' theories, after Frege's term '*Sinn*'. The other account is the 'historical chain' theory, due to Kripke and Geach. In Geach's words, 'for the use of a word as a proper name there must in the first instance e someone acquainted with the object named…. But…the use of a given name for a given object…can be handed on from one generation to another…Plato knew Socrates, and Aristotle knew Plato, and Theophrastus knew Aristotle, and so on in apostolic succession down to our own times. That is why we can legitimately use "Socrates" as a name the way we do' (1969-70: 288-9).*

1 Sense theories: Introduction

The idea that names express reference-determining conceptions or 'senses' originated with Frege (1892). But Frege was led to it not by wondering about the mechanism of reference, but rather about the difference in 'cognitive value' between two true identity statements of the respective forms $a = a$ and $a = b$ (for example, 'Clark Kent is Clark Kent' and 'Clark Kent is Superman'). Statements of the former sort, he says, are a priori, while the latter 'often contain very valuable extensions of our knowledge and cannot always be established a priori' ([1892] 1980: 56). Frege takes this to show that the propositions expressed by $a = a$ and $a = b$ cannot be the same even when *a* and *b* are names of the same person. Yet if the names do corefer, there is no difference in the references of the constituents of $a = a$ and $a = b$; moreover, these two sentences are assembled in the same way. So either they cannot express different propositions, or else - and this is the inference Frege drew - what determines the proposition a sentence *S* expresses cannot just have to do with the structure of *S* and the *references* of its constituent words and phrases (see Frege, G. §§3-4; Sense and reference §1).

Frege's proposal was that in addition to possessing a reference, any meaningful expression has a sense, the sense being a 'way of thinking' of the reference. The proposition a sentence expresses is determined by the senses of the words in it, not their references. Thus 'Clark Kent is Clark Kent' and 'Clark Kent is Superman' express different propositions because 'Clark Kent' and 'Superman' express different senses, senses which happen to be ways of thinking of the same reference. Taking the Superman fiction to be fact, the sense of 'Clark Kent' might be 'the mild-mannered reporter on *The Daily Planet* who has a crush on Lois Lane' while the sense of 'Superman' might be 'the blue-suited extraterrestrial who flies'. Here we specify different conceptions of the same individual in two definite descriptions (expressions of the form 'the so-and-so' - see Descriptions). The proposition that Superman is Clark Kent therefore has the content that the blue-suited extraterrestrial who flies is the mild-mannered reporter on *The Daily Planet* who has a crush on Lois Lane, which may indeed be, in Frege's phrase, an extension of our knowledge.

It was Frege's view that in an ideal language each name would have a fixed sense and reference for everyone; in an ordinary natural language, there are names that fail to refer, and users may fail to agree on the sense of a name for a specific individual, which Frege thought of as deficiencies of natural language. In a modification of Frege's views, Searle (1958) allowed that a name may be associated with a whole range of descriptions, different users using different ranges; the bearer of the name need not satisfy all the conditions mentioned in the descriptions, only 'a sufficient number', an intentionally vague condition.

Sense theories along such lines as these have been called ‘famous deeds’ sense theories. Famous deeds sense theories appear to have been conclusively refuted by Kripke (1972, 1980). Kripke objects to such sense theories both as they respond to Frege’s own puzzle about the difference between $a = a$ and $a = b$, and as they address our initial question about the mechanism of reference. Let us consider these two issues in turn.

2 Sense theories: Do names express senses?

According to Kripke, the propositions one expresses using names do not involve senses or ways of thinking expressed by the names, otherwise some such propositions would be both metaphysically necessary and a priori, which they are clearly not. For example, suppose the sense of ’Aristotle’ is ‘the pupil of Plato who tutored Alexander’. Then the sentence

(1) Aristotle was a pupil of Plato

would express the proposition with the content

(2) The pupil of Plato who tutored Alexander was a pupil of Plato.

But (2) is, in a certain sense, necessary, while (1) is not. There is no way things could have gone in which (a) a unique pupil of Plato who tutored Alexander exists and (b) that person was not a pupil of Plato. On the other hand, there are many ways things could have gone in which (a′) Aristotle exists but (b′) Aristotle was not a pupil of Plato (for example, he died young). In other words, granted that there is such a person as Aristotle, he may be a pupil of Plato or he may not be. But granted that there is such a person as the pupil of Plato who tutored Alexander, it follows that he is a pupil of Plato.

This example brings out Kripke’s famous distinction between ‘rigid’ and ‘non-rigid’ designators. In thinking about or describing other ways things could have gone (other possible worlds) we use a proper name such as ‘Aristotle’ consistently to denote the same person; this makes proper names rigid designators. But we use definite descriptions such as ‘the pupil of Plato who tutored Alexander’ differently. With respect to the actual world, this description picks out Aristotle; with respect to a possible world where someone else is the one and only pupil of Plato who tutored Alexander, the description picks out that other person, not Aristotle; and with respect to a possible world where either Plato or Alexander does not exist, the description fails to pick out anyone, even if Aristotle does exist. Hence such descriptions are non-rigid designators. Certain descriptions, such as ‘the positive square root of 9’, are as rigid as proper names, but this is on account of their subject matter, not their semantic role. It is because the typical famous deeds description is non-rigid that the contrast between pairs such as (1) and (2) *vis-à-vis* necessity and contingency arises (see Reference §2).

Just as (1) and (2) differ in modal status, they differ epistemically: bracketing the question of existence, (2) is, in a limited sense, knowable a priori, while (1) is not. That is, granted that there was such a person as Aristotle, it is a further, empirical question whether he was a pupil of Plato (there may be a controversy among historians about this). But granted that there was such a person as the pupil of Plato who tutored Alexander, it is not a further question, *a fortiori* not an empirical one, whether he was a pupil of Plato.

The same objections arise to Searle’s modified version of the sense theory. For example, no matter what range of descriptions we associate with a name, they will generate statements that are a priori in the manner of (2). If being ϕ logically implies being ϕ', then

(3) The thing which is *F*-and-*G* or *G*-and-*H* is *F*′ or *G*′ or *H*′

is essentially the same as (2), just more complicated. But if the predicates *F*, *G* and *H* encapsulate famous deeds, the corresponding

(4) *NN* is *F*′ or *G*′ or *H*′,

where ‘*NN*’ is the name with which ‘the thing which is *F* or *G* or *H*’ is associated, will be no more a priori than (1) (Kripke 1980).

3 Sense theories: Do senses determine reference?

Kripke (1972, 1980) demonstrates another flaw in famous deeds sense theories, namely, that they do not provide an adequate answer to the question about the mechanism of reference. This is because (1) the likely candidate for the sense of a name may pick out an object which is not in fact the name's bearer, or may fail to pick out anything, and (2) we can succeed in referring with a name even when we do not have a 'famous deeds' description associated with it.

Kripke illustrates (1) with two examples. If any description is associated with the name 'Gödel' it is 'the discoverer of the incompleteness of arithmetic'. Does this mean that 'Gödel' refers to that person? What if the theorem was actually proved by Schmidt, who died in strange circumstances, and Gödel got hold of Schmidt's work and represented it as his own? The very fact that we can understand this 'what if' shows that the reference of 'Gödel' is not fixed as whoever discovered the incompleteness of arithmetic. And though this example is fictional, there are similar actual cases: Peano's Axioms are not due to Peano; and Einstein was not the inventor of the atomic bomb (Kripke 1980). For a case where there are descriptions that do not pick out any object although the relevant name still refers, Kripke gives the example of the prophet Jonah, who really existed (according to the scholarly consensus), but whose career as described in the Bible is essentially fictitious. Again, the mere intelligibility of the claim 'Jonah was a historical person but everything uniquely identifying that the Bible says about him is fictitious' is enough to show that the reference of 'Jonah' is not fixed as the Hebrew prophet who was swallowed by a whale, or by any other condition, no matter how complicated or disjunctive, derived from the Book of Jonah.

As for (2), successful reference without associated (definite) descriptions, Kripke points out that most people can use the names 'Richard Feynman' and 'Murray Gell-Mann' to refer to those two people, but that, at best, all the typical person knows about either is that he is a famous physicist who won a Nobel Prize (this was before Feynman achieved popular fame for his role in the *Challenger* disaster inquiry). So we have difference in reference with no difference in associated descriptions; therefore associated descriptions are not at the heart of how proper name reference works.

Perhaps these examples only establish such a conclusion for descriptions that encapsulate famous deeds. Kripke considers some other approaches and concludes that they violate an important non-circularity condition: that candidate descriptions must not themselves embed the notion of reference in a way that cannot eventually be eliminated. For example, we might suggest that the reference of 'Socrates' is fixed as 'the man called "Socrates"'. Since 'called' just means 'referred to as', we do not explain how reference to Socrates is possible in this way: what we want to know is how Socrates *gets to be* the man referred to as 'Socrates'. So this goes nowhere as a proposal about the mechanism of reference, and Kripke plausibly argues that the same would be true for more complicated versions of the idea, for example, that 'Gödel' refers to the person to whom the proof of the incompleteness of arithmetic is commonly attributed. So these attempts at a non-famous-deeds description theory fail. However, in view of Kripke's critique of the famous deeds approach, it seems that if any sense theory is to work, it will have to be a non-famous-deeds account of some sort.

4 Historical chains

The main competing account of the mechanism of reference is the Geach-Kripke historical chain account, in which competence to refer with the name is transmitted across generations in the style adverted to in the quotation from Geach in this entry's prologue (1969-70). The historical chain account is sometimes called the 'causal' chain account, since it is held that the links in the chain are forged by transactions of a causal sort. In Kripke's own version, which he says is a 'picture' rather than a theory (1980: 97), a name is introduced into a community by some 'initial baptism' of an object with the name, and then the name is passed on from link to link, it being required at each step that the receiver of the name 'intend…to use it with the same reference as the [person] from whom he heard it' (1980: 96).

If we regard being told about, or otherwise hearing about, an object, as a way of becoming 'acquainted' with it, then the Geach-Kripke account instantiates an approach championed by Russell (1918), who made acquaintance with an object necessary for referring to it. However, Russell's notion of acquaintance was rather idiosyncratic: apart from my own sense-data, my self, universals and perhaps the present moment, I lack acquaintance with things, according to Russell. Since the ordinary names I use, ostensibly for other people and things, are not names of sense-data, Russell claimed that in a 'logical' sense, ordinary names are not 'proper' names. Rather, he

suggested, they are definite descriptions in disguise (see Russell, B.A.W. §9). For this reason, Kripke sometimes calls the view he opposes the 'Frege-Russell theory of names'. The point to bear in mind is that Frege and Russell had different accounts of how reference works (so there is no 'Frege-Russell theory of reference'), but because on Russell's account ordinary names do not really refer, the two philosophers end up saying similar-sounding things about such names.

Returning to Kripke's account of the mechanism of reference in terms of informational exchanges in which the intention to preserve reference is present, the obvious question is whether it is any improvement on circular description theories. After all, the notion of reference enters explicitly into Kripke's necessary condition for successful passing on of the name, and it also enters at the start of the chain, where some kind of demonstrative reference to the object being baptized is standardly made. To put the same question another way, if this account is explanatory, would there be anything wrong with a description theory which attributed to a name '*NN*' and a user of the name *U* the sense 'the object at the start of the chain of reference-preserving links by which I came into mastery of this name', in which 'I' refers to *U*?

One problem with this version of a description theory is that no ordinary speaker of a natural language associates any such description with a proper name, since the description embodies a philosophical theory, and it is difficult to see how one could justify claiming that the association is 'implicit'. Here there is a contrast with Kripke's proposal, which only requires speakers to have the intention to preserve the reference of the new name they have just learned, and it is surely plausible that speakers do have such an intention. Still, there is some sense in which the fundamental nature of the mechanism of reference is left unexplained. For instance, what exactly is it about an actual baptismal service, or about parents announcing their choice of name, that causes a new 'common currency' (see Kaplan 1990) name to be added to the language? Or is the request for further explanation here a demand for a 'reductive' account of reference in terms of non-semantic notions (see Reference §8), something perhaps impossible?

The causal chain picture is not without other difficulties, as Kripke mentions. For instance, straightforward application of the picture could lead to the conclusion that 'Santa Claus' is the name of a certain central European king, which seems wrong. Evans (1973) proposed a different causal account on which the 'causal source' of the information associated with the name determines whom it refers to, though in a possibly complex way. Evans observes that the reference of a name in a linguistic community can change over time if it was originally a name for *x* that was mistakenly but consistently misapplied to *y: y* would ultimately become the bearer of the name, being the dominant causal source of the information associated with it (unless, as Evans notes, younger members of the community defer in their use of the name towards those for whom it was once a name for *x*). Evans subsequently elaborated his account (1982: ch. 11).

5 Direct reference and Frege's puzzle

A striking feature of the Geach-Kripke picture is that it leaves no role for sense to play as that which determines reference. The only candidate for the 'meaning' of a name is therefore the name's reference itself. A theory which claims that the meaning of a name is just its reference, pure and simple, is often called a 'direct reference' theory, 'direct' signifying that reference is not mediated via sense. Such theories trace their origins through Russell back to Mill (1843); more recently, the idea is prominent in the writings of Marcus on reference (1961), and has subsequently been developed and defended in a sustained form by Salmon (1986) and Soames (1994).

The main problem which direct reference theories face is the puzzle about the difference between $a = a$ and $a = b$. The general 'failure of substitutivity' puzzle is the puzzle of how it is possible for the meanings of two sentences to differ if the sentences have the same structure and their corresponding parts have the same meaning. For example,

(5) It is self-evident to any rational thinker that if Superman exists, then Superman = Superman

seems true and 'Superman' and 'Clark Kent' have the same reference. Therefore, according to direct reference theory, (5) should have the same meaning as

(6) It is self-evident to any rational thinker that if Superman exists, then Superman = Clark Kent.

But at first sight, (6) does not even have the same truth-value as (5), never mind the same meaning.

Failure of substitutivity is handled straightforwardly on a sense theory of names which allows that coreferential names can have different senses. The claim would be that (not merely the reference but) the sense of ‘Superman’ enters into the truth-condition of (5), while in (6) the sense of ‘Clark Kent’ is also involved. There are various mechanisms which might be invoked here, Frege’s own being the simplest: in (5) and (6), according to Frege, the proper names refer to their *senses*, rather than to the person they normally refer to (see Propositional attitude statements §1; Sense and reference §5). But the trouble with any explanation of substitutivity failure for proper names that invokes senses is that sense theories of proper names have been so thoroughly battered by Kripke.

Direct reference theory seems to have less to work with to explain the semantic difference between (5) and (6). On perhaps the best known version of the theory, Salmon’s, it is simply denied that there is a semantic difference between (5) and (6). The appearance that substitution changes truth-value in cases such as (5) and (6) is explained as being due to pragmatic effects (see Pragmatics). Apparently, the only alternative for a direct reference theorist is to identify some assumption in the argument that substitution of coreferring names ought not to change meaning and deny that assumption. This strategy produces a possible target. We suggested that what makes substitutivity failure puzzling is that sentences with the same structure whose corresponding parts have the same meaning should themselves have the same meaning. But this presupposes that the meaning of the entire sentence is wholly determined by the meaning of its parts and their manner of composition (see Compositionality). However, the notion of structure alluded to in the phrase ‘manner of composition’ is one which is sensitive only to syntactic categories of expression, not to the identity of expressions. There is another, more ‘logical’ notion of structure, on which use of a different word, even one with the same meaning, can disrupt structure (Putnam 1954). In this logical sense, the structure of ‘if Superman exists, then Superman = Superman’ is ‘if *Et* then *t t*’, while ‘if Superman exists, then Superman = Clark Kent’ has the different structure ‘if *Et* then *t* = *t*Čast;’. If we hold that substitution is acceptable only when it does not change logical structure, then from (5) we can infer merely

(7) It is self-evident to any rational thinker that if Clark Kent exists, then Clark Kent = Clark Kent

which is presumably true if (5) is. Of course, this does not help with cases of substitutivity failure in which there is only one occurrence of the name being substituted (‘Lois believes that Superman is an extraterrestrial’), but some direct reference theorists have tried to develop a more general notion of structure for these cases (Taschek 1995).

It is the handling of substitutivity puzzles that is the deciding issue between sense theories and direct reference theories of proper names. Sense theorists need to find a viable account of the senses of names and direct reference theorists need to find a persuasive account either of why substitutivity fails in cases such as (5), or else of why it gives such a convincing appearance of doing so.

See also: De re/de dicto; Kripke, S.A.; Logical and mathematical terms, glossary of

GRAEME FORBES

References and further reading

Burge, T. (1973) ‘Reference and Proper Names’, *Journal of Philosophy* 70: 425-39. (Argues that, semantically, proper names should be understood as certain kinds of predicate.)

Deutsch, H. (1989) ‘On Direct Reference’, in J. Almog, J. Perry and H.K. Wettstein (eds) *Themes from Kaplan*, Oxford and New York: Oxford University Press, 167-95. (Applies a Kaplan-style approach to context-dependence to proper names.)

Devitt, M. (1981) *Designation*, New York: Columbia University Press.(Develops a causal theory of reference using the ‘language of thought’ hypothesis.)

Evans, G. (1973) ‘The Causal Theory of Names’, *Proceedings of the Aristotelian Society*, supplementary vol. 47: 187-208; repr. in *Collected Papers*, Oxford: Oxford University Press, 1985, 1-24.(A very accessible and insightful discussion of Kripke’s account.)

Evans, G. (1979) ‘Reference and Contingency’, *The Monist* 62: 161-84; repr. in *Collected Papers*, Oxford: Oxford University Press, 1985, 178-213.(Discusses a category of proper name not considered in this article, which Evans calls ‘descriptive’ names. These are names whose reference is explicitly fixed by a definite description, and which figure in Kripke-style examples of the contingent a priori.)

Evans, G. (1982) *The Varieties of Reference*, Oxford and New York: Oxford University Press.(Published after

Evans' death; some parts are unfinished. As well as Evans' final thoughts on proper names, includes influential discussion on Frege, Russell and the notion of an 'object-dependent' thought.)

Forbes, G. (1990) 'The Indispensability of *Sinn*', *Philosophical Review* 99: 535-63. (Develops a non-famous-deeds sense theory of names and applies it to substitutivity failure.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford and New York: Blackwell, 3rd edn, 1980.(Widely regarded as the paper which inaugurated modern philosophy of language.)

Geach, P.T. (1969-70) 'The Perils of Pauline', *Review of Metaphysics* 23: 287-300; repr. in *Logic Matters*, Berkeley, CA: University of California Press, 1972, 153-65. (States the historical chain view and discusses the use of names in intentional contexts.)

Jubien, M. (1993) *Ontology and the Fallacy of Reference*, Cambridge and New York: Cambridge University Press.(An unusual but interesting account of reference, based on 'revisionist' metaphysics.)

Kaplan, D. (1990) 'Words', *Proceedings of the Aristotelian Society*, supplementary vol. 64: 93-117.(Includes an original and accessible discussion of how names should be individuated.)

Kripke, S.A. (1972) 'Naming and Necessity', in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel, 252-355.(One of the most influential and widely discussed works in post-war philosophy.)

Kripke, S.A. (1979) 'A Puzzle About Belief', in A. Margalit (ed.) *Meaning and Use*, Dordrecht: Reidel. (Develops two famous cases - the London/Londres case and the Paderewski case - in defence of the thesis that substitution puzzles do not necessarily tell against the principle of substitutivity.)

Kripke, S.A. (1980) *Naming and Necessity*, Oxford and New York: Blackwell.(The book version of Kripke (1972).)

Marcus, R.B. (1961) 'Modalities and Intensional Languages', *Synthèse* 13: 303-22. (Includes an early statement of contemporary direct reference theory.)

Mill, J.S. (1843) *A System of Logic*, London: Parker; repr. in *The Collected Works of John Stuart Mill*, ed. J.M. Robson, London: Routledge, 1996, vol. 7.(Draws a famous distinction between the denotation and the connotation of a word and defends a metalinguistic analysis of identity sentences.)

Putnam, H. (1954) 'Synonymy and the Analysis of Belief Sentences', *Analysis* 14: 114-22.(Attacks compositionality assumptions in substitution puzzles.)

Recanati, F. (1993) *Direct Reference*, Oxford and New York: Blackwell. (Distinguishes a sense in which names are directly referential from two other semantically relevant senses in which they express modes of presentation. Also includes a good discussion of how 'what is said' by an utterance is determined.)

Russell, B.A.W. (1918) 'The Philosophy of Logical Atomism', *The Monist* 28: 495-527; 29: 32-63, 190-222, 345-80; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992.(One of many possible sources of Russell's views, which changed throughout his long life. Lectures II and VI are the most germane to this entry.)

Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press.(The most complete attempt by a direct reference theorist to account for the problem of (apparent) failure of substitutivity.)

Searle, J. (1958) 'Proper Names', *Mind* 67: 166-73.(A significant elaboration of broadly Fregean themes.)

Soames, S. (1994) 'Attitudes and Anaphora', in J.E. Tomberlin (ed.) *Philosophical Perspectives: 8 Logic and Language*, Atascadero, CA: Ridgeview, 251-72.(Within the context of direct reference theory, investigates the behaviour of anaphoric pronouns in propositional attitude ascriptions. Mostly nontechnical, but rigorous.)

Strawson, P.F. (1959) *Individuals*, London: Methuen.(Chapter 6 includes an elaborated Fregean account that Kripke often uses as a stalking horse.)

Taschek, W. (1995) 'Belief, Substitution, and Logical Structure', *Noûs* 29: 71-95. (Extends the approach of Putnam (1954) to account for a wider range of substitution puzzles.)

Property

Most of the great philosophers have expressed views on property, its justification and limits, and especially on the justification of having private property; generally, one must understand these views against the background of the economic and social conditions of their times. Notable theories include first possession (roughly, 'whoever gets his or her hands on it justifiably owns it'), labour ('whoever made it deserves to own it'), utility and/or efficiency ('allowing people to own things is the most effective way of running society') and personality ('owning property is necessary for personal development').

Few thinkers now defend the first possession theory but all the other three have their contemporary supporters. Some philosophers combine two or more theories into multi-principled or 'pluralist' justifications of property ownership. Many express concern about wide gaps between rich and poor and argue for constraints on inequalities in property holdings.

1 Concept of property

The concept of property is understood in two main ways. First, it applies to material things such as tools, houses and land. Second, it applies to bundles of 'rights'. Most lawyers and philosophers stress the second understanding over the first. The first is too narrow to accommodate intangibles such as copyrights, patents and trademarks. Moreover, for purposes of legal and philosophical analysis the second understanding is more useful.

Here the word 'rights' covers many normative modalities (see Rights §2). Following Hohfeld and Honoré, the package of rights called 'property' includes: claim rights to possess, use and receive income; powers to transfer, waive and exclude; a disability (a no-power) of others to force a sale; liberty rights to consume or destroy; and immunity from expropriation by the government. It is probably a vain enterprise to try to specify necessary and sufficient conditions for all and only those 'rights' that pertain to property rather than, say, to contract or tort. Yet it is a worthy undertaking to try to identify those 'rights' that seem most central to property or those rules that create such 'rights'.

The second understanding of property finds favour for many reasons. It applies to widely different cultures. It is useful in both legal and philosophical analysis. It can accommodate both 'will' and 'interest' theories of rights. It can capture both full ownership and limited property rights (such as easements), and both tangible and intangible property. It permits distinguishing among different sorts of property depending on the identity of the rightholder. Thus, a single person or a corporation has *private* property, a tribe has *communal* property and a government has *public* or *state* property.

This explanation of the concept of property is neutral with respect to which kinds of things can be the subject of property rights. Few would defend slavery, which is the most extreme form of property in the bodies or persons of others. More disputed are whether people can have property rights in the whole of their own bodies, in bodily parts for use in transplantation, in information, in cultural practices, in welfare payments or other forms of government largesse, or in seabed resources or objects in outer space.

The approach of Guido Calabresi and Douglas Melamed (1972) to the notion of property, which philosopher-economists and lawyer-economists often use, is not conceptually distinct from that of Hohfeld (1919) and Honoré (1961). For Calabresi and Melamed the basic idea is an 'entitlement'. An entitlement is, roughly, an interest that the law does or should protect. Decisions must be made as to which entitlements to protect and how to protect them. As to the latter decision, the law may use what Calabresi and Melamed call 'property rules', 'liability rules' and 'rules of inalienability'. These technical terms can be restated in Hohfeld's vocabulary. If a person's entitlement is protected by a property rule, then others have a disability (a no-power) with regard to obtaining the entitlement except at a price agreed to by its holder. If a person's entitlement is protected by a liability rule, then others have a disability with regard to obtaining or reducing the value of the entitlement unless they compensate its holder by an officially determined amount. If a person's entitlement is protected by a rule of inalienability, its holder has no power to transfer the entitlement to others. The real value of Calabresi and Melamed's approach rests in the light that it sheds on the integration of property and tort, on its application to pollution control, on its sensitivity to distributional as well as efficiency considerations, and on the choice between

civil and criminal sanctions for violations of property rights.

2 History of theorizing about property

Many of the great philosophers have offered views on property. Plato actually expresses two different views (see Plato §§14, 17). In the *Republic* he portrays an ideal society in which the rulers and auxiliaries have political power but almost no private property. Ordinary citizens possess private property, with limits on unequal distribution, but have almost no political power. In contrast, the *Laws* depicts a practical, second-best society. It favours individual private possession with underlying communal ownership. Regulations maintain a roughly equal distribution of property.

Aristotle advocates private rather than communal ownership on grounds that relate to the smooth functioning of a society and its economy. His *Politics* fails, however, to show how property should, or can, be 'in a certain sense common, but, as a general rule, private'.

Locke, at least in the *Second Treatise*, offers a labour theory of property (see Locke, J. §10). The interpretation of his theory is disputed. It is unclear how much the theory rests on 'mixing' one's labour with unowned things, or on barring the idle from taking the benefit of the labourer's pains, and the extent to which the needs of others and restrictions on spoilage limit the acquisition of property by labour. It is also disputed whether Locke's theory is proto-capitalist or stems from some conception of natural law.

Hume and Bentham are the first important utilitarian theorists of property. For Hume (1739-40), utility in the sense of common interest explains how private property arises (see Hume, D. §§4-5). It also justifies the general institution of private property and specific rules of property law. Bentham (1802) understands utility as the balance of pleasure over pain, and views property in terms of expectations (see Bentham, J. §2). The security of expectations, rather than equality in distribution, is for Bentham the weightier consideration in favour of private property. Nevertheless, Bentham is sharply critical of the English law of property of his day.

Kant and Hegel are also linked, although less closely and in more complicated ways than Hume and Bentham. For Kant, as for Locke, a form of private property can exist in the state of nature (see Kant, I. §10). But Kant has a social contract theory under which only society can give individual possession full normative significance as private property. Hegel regards Kantian private property as excessively individualistic and belonging to the domain of 'abstract right' (see Hegel, G.W.F. §8). A more sophisticated form of property exists in 'civil society' - roughly, the social correlate of *laissez-faire* capitalism. Only in the 'state' does private property emerge fully transformed; private property, unequally distributed, still exists, but is subject to heavier state regulation in the organic interest of all citizens.

Marx dismisses Hegel's dialectical defence of private property as so much claptrap. Marx believes that capitalist production, not unequal distribution, is the more serious problem, because 'capitalism' distorts human relationships. In particular, private property under capitalism involves 'alienation' - that is, a separation of persons from nature, the products of their labour, other human beings and even themselves (see Alienation §4). *Pace* Marx, it is unclear that all forms of private property must involve alienation, or that alienation will be absent from the mature communist society that he envisages.

3 Systematic justifications of private property

First possession and entitlement. Attempts to justify the acquisition of property by being the first person to possess it have won few converts. The general difficulty is to show why first possession should support full ownership rather than limited rights of use. Specific difficulties include articulating which acts of possession count, explaining how long they must continue, and identifying the item or area possessed. Suppose that someone claims title to an acre of farmland by standing in one place for an hour. Why should that person not have to farm rather than merely stand, to remain there for a year rather than just an hour, and to perform appropriate acts over the entire acre rather than only on an area two feet square? To many these difficulties have no ready solution. Still, those sympathetic to the rights of aboriginal peoples sometimes invoke first possession.

A related account is the libertarian entitlement theory of Robert Nozick (1974). Nozick's position is hazy on which acts are appropriate acts of acquisition under his principle of justice in acquisition. The literature contains sharp

attacks on his position. Among them are objections as to how any individual can, by unilateral action, impose moral duties on others to refrain from using certain resources, worries about inequalities of property holdings, and scepticism that anyone living today has morally valid property rights by transfer from some original acquirer. Nozick has not responded to these attacks. Other libertarian theories are either grounded in economics or based on strong conceptions of freedom and individual rights (see Libertarianism §3).

Labour. More promising are efforts to recast the labour theory in terms of the desert of the labourer. Lawrence Becker and Stephen Munzer offer somewhat different accounts of a labour-desert principle.

Becker (1977) holds that if a worker adds value to the lives of others in some morally permissible way and without being required to do so, that person deserves a fitting benefit. Property rights may be the most fitting benefit, and which benefit is 'most fitting' depends on purposes. However, Becker's theory leaves it unclear which purposes (for example, an attempt to gain property rights?) and whose purposes (the labourer's or those of other individuals?) are relevant. Also, he does not show why purpose should be the sole test of fittingness. Nevertheless, Becker rightly insists that losses inflicted by the labourer's work require some reduction in property rights or some offsetting compensation or taxation.

Munzer (1990) argues that if workers use their bodies to produce something or provide a service, then they have a *prima facie* claim to deserve property rights in the product or in wages. This claim is qualified by scarcity, by the needs and rights of others, and by some post-acquisition changes in situation. Restrictions on transfer may also apply. Moreover, since work is a social activity, a wage policy is in order that makes wages commensurate so far as possible with desert. The heavy qualifications that surround this version of the labour theory support moderate egalitarianism rather than the wide disparities in income and wealth that libertarian theories allow. Critics of Munzer's labour-desert principle have objected that its intellectual underpinnings are not sufficiently clear and that in both theory and practice no precise correlations exist between desert claims and property rights (see Desert and merit).

Utility and efficiency. Contemporary justifications of this sort owe a distant obligation to Hume and Bentham. These justifications often stress 'efficiency', which does not allow interpersonal comparison of individual preference satisfaction, over 'utility', which does allow them. Because of the invocation of efficiency, the most sophisticated advocates are frequently economists or academic lawyers influenced by economics rather than philosophers.

Utility and efficiency can, given some plausible assumptions about individuals' preferences, justify some public property as well as some rights of private property. They tend (again given certain assumptions about preferences) to justify moderate egalitarianism rather than highly unequal distributions of private property. However, more detailed information is needed to show how utility and efficiency can justify particular rules of property law or radical changes in existing property institutions. Perhaps the most influential theoretical result of applying efficiency to the law of nuisance bears the sobriquet 'Coase's Theorem' (Coase 1960). It holds that, under perfect competition and perfect information and with costless transactions, an efficient allocation of resources will result no matter what the decision of the courts concerning liability for damage.

Personality. Contrary to some reports, the personality theory of property did not die with Hegel and the British neo-Hegelians. The basic idea is that people need at least some private property in order to develop healthy character structures. The account of 'property for personhood' by Margaret Radin (1982) is probably the best known and most fully developed contemporary theory of this sort. She applies her theory to many practical legal problems, but some readers find the foundations of her account no clearer that those of Hegel. Her books also suggest that some things are so personal that they ought not to be property or 'commodities' at all.

Pluralist theories. The variety of possible justifications of property has suggested to some writers that the most plausible account of property is 'pluralist' - that is, contains two or more irreducible principles. Becker and Munzer explicitly embrace pluralist accounts. Other writers do so implicitly. Some critics dismiss such pluralism as eclectic. Other critics contend that if it is impossible to show that the principles never conflict, it is necessary to establish that conflict between principles is logically consistent and otherwise free from objection. Further, it may be harder to apply a pluralist theory rather than a unitary theory to practical problems. Still, those who favour a unitary theory - a theory with a single principle or at least a single supreme principle - must show how it can

accommodate the complexity of considered moral judgments concerning property. For anyone who shares Proudhon's worry that property is theft (1841), justification remains the key problem in the theory of property (see Proudhon, P.-J.).

4 Constraints on the distribution of property

Justice and equality. The liberal tradition in political theory, stemming particularly from Mill, often tries to limit inequalities of property holdings (see Mill, J.S. §§11-12). The limitation can take the form of a side constraint or, in a pluralist theory, of a separate principle of justice and equality. Few philosophers argue for strictly equal holdings. More modest is the 'difference principle' of John Rawls (1971; 1993) (see Rawls, J. §§1-2). Applied to property, this principle would hold that differences in holdings are justifiable only if they are to the greatest benefit of the least advantaged. A related view might concentrate on both the floor and the ceiling: Everyone should be ensured a minimum amount of property, and remaining inequalities, if any, should not undermine a fully human life for anyone in society.

Absence of exploitation. The radical tradition in political theory, traceable especially to Marx, stresses the need to eliminate exploitation and other differences in power that derive from unequal property holdings. Roughly, persons are exploited if others secure a benefit by using them as a tool or resource so as to cause them serious harm. Exploitation theory often concentrates as much on problems of economic production as it does on the real or imagined evils of unequal distributions of property. The 'Critical Legal Studies movement', in some of its forms, objects to exploitation and differences in power as they relate to property (see Critical legal studies). More rigorous accounts of exploitation come from the work of John E. Roemer (1982).

STEPHEN R. MUNZER

References and further reading

Aristotle (*c*. mid 4th century BC) *Politics*, ed. W.L. Newman, Oxford: Oxford University Press, 1887-1902, 4 vols; Books I-II trans. T.J. Saunders, Clarendon Aristotle Series, Oxford: Oxford University Press, 1995, esp. 1256,a,1-1259,a,35; 1262,b,38-1266,a,30.(Advocates private rather than communal ownership on grounds that relate to the smooth functioning of a society and its economy.)

Becker, L.C. (1977) *Property Rights: Philosophic Foundations*, London, Henley and Boston, MA: Routledge & Kegan Paul.(Admirably concise introduction.)

Bentham, J. (1802) *The Theory of Legislation*, ed. C.K. Ogden, London: Routledge & Kegan Paul, 1932, esp. 109-23, 148-98.(Classic utilitarian statement.)

Calabresi, G. and Melamed, A.D. (1972) 'Property rules, liability rules, and inalienability: one view of the cathedral', *Harvard Law Review* 85 (6): 1089-128.(Difficult but rewarding article.)

Coase, R.H. (1960) 'The problem of social cost', *Journal of Law & Economics* 3 (1): 1-44.(See §3 of this entry.)

Demsetz, H. (1967) 'Toward a theory of property rights', *American Economic Review* 57 (2): 347-59.(Influential account of efficiency and the origins of property.)

Harris, J.W. (1996) *Property and Justice*, Oxford: Clarendon Press.(Insightful study with many illustrations from English law.)

Hegel, G.W.F. (1821) *Philosophy of Right*, trans. T.M. Knox, Oxford: Clarendon Press, 1952.(§§41-71, 170-80, 182-256 and 257-360 explore property in the domain of 'abstract right', the family, 'civil society' and the 'state', respectively.)

Hohfeld, W.N. (1919) *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, West Port, CT: Greenwood Press, 1978. (Important but difficult reading.)

Honoré, A.M. (1961) 'Ownership', in *Oxford Essays in Jurisprudence (First Series)*, ed. A.G. Guest, Oxford: Clarendon Press, 107-47.(Lucid, accessible and influential essay.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 1978, 2nd edn. Bk III, Pt II, Sects II-IV.(Classic early utilitarian statement.)

Kant, I. (1797) *The Metaphysics of Morals*, trans. M. Gregor, Cambridge: Cambridge University Press, 1991.(The Rechtslehre, which is the first half of this book, is hard going but fundamental.)

Locke, J. (1689/90) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1960, 2nd edn, 1967, esp. *Second Treatise*, §§25-51.(Classic statement of the labour theory.)

Macpherson, C.B. (1962) *The Political Theory of Possessive Individualism: Hobbes to Locke*, Oxford: Clarendon Press.(Views Locke as a possessive individualist.)

Marx, K. (1843) *Critique of Hegel's Philosophy of Right*, ed. J. O'Malley, Cambridge: Cambridge University

Press, 1970, esp. 94-113.(See §2 of this entry.)
Marx, K. (1844) *Economic and Philosophic Manuscripts of 1844*, ed. D.J. Struik, New York: International Publishers, 1964.(See §2 of this entry.)
Mill, J.S. (1848) *Principles of Political Economy*, ed. W. Ashley, Fairfield, NJ: Kelley, 1976.(Excellent liberal statement on property rights and inheritance.)
Munzer, S.R. (1990) *A Theory of Property*, Cambridge: Cambridge University Press.(A pluralist theory with detailed applications to corporations, bequests and government takings of private property.)
Munzer, S.R. (1994) 'An uneasy case against property rights in body parts', *Social Philosophy & Policy* 11 (2): 259-86.(Discusses whether organs and other bodily parts should be the subject of property rights.)
Nozick, R. (1974) *Anarchy, State, and Utopia*, New York: Basic Books, 150-82.(A lively and prominent libertarian account.)
Penner, J.E. (1996) 'The "Bundle of Rights" Picture of Property', *UCLA Law Review* 43 (3): 711-820. (Sustained criticism of the picture.)
Penner, J.E. (1997) *The Idea of Property in Law*, Oxford: Clarendon Press.(Emphasizes the centrality of exclusion, the social use of property and 'things' to the concept of property.)
Plato (*c*.380-367 BC) *Republic*, trans. G.M. Grube, revised by C. Reeve, Indianapolis, IN: Hackett Publishing Company, 1992, 416d-423a.(A utopian theory.)
Plato (*c*.360-347 BC) *Laws*, trans. T.J. Saunders, *Plato: Laws*, Harmondsworth: Penguin, 1970, 740a-745a.(A practical, second-best theory.)
Proudhon, P.-J. (1841) *What is Property?*, trans. B.R. Tucker, New York: Dover, 1970.(Historically important questioning of the basis for private property.)
Radin, M.J. (1982) *'Property and personhood'*, *Stanford Law Review* 34 (5): 957-1015.(Uses Hegelian and non-Hegelian perspectives on persons to develop a communitarian theory of property.)
Radin, M.J. (1993) *Reinterpreting Property*, Chicago, IL, and London: University of Chicago Press.(Articles written between 1981 and 1992 that reveal a growing pragmatism.)
Radin, M.J. (1996) *Contested Commodities*, Cambridge, MA, and London: Harvard University Press.(Reworking of articles written between 1987 and 1995 that deal with 'commodification'.)
Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(The most notable contemporary account of justice.)
Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(Systematic rethinking and partial modification of the previous work.)
Reich, C. (1964) 'The new property', *Yale Law Journal* 73 (5): 733-87.(Perceptive treatment of government largess, property and liberty.)
Roemer, J.E. (1982) *A General Theory of Exploitation and Class*, Cambridge, MA: Cambridge University Press. (Exploitation theory.)
Ryan, A. (1984) *Property and Political Theory, Oxford: Blackwell*.(Elegantly written historical survey.)
Tully, J. (1980) *A Discourse on Property: John Locke and his Adversaries*, Cambridge: Cambridge University Press.(Views Locke as a natural lawyer.)
Waldron, J. (1988) *The Right to Private Property*, Oxford: Clarendon Press.(A brilliantly argued study that includes insightful discussions of Locke, Proudhon and Hegel.)

Property theory

Traditionally, a property theory is a theory of abstract entities that can be predicated of things. A theory of properties in this sense is a theory of predication - just as a theory of classes or sets is a theory of membership. In a formal theory of predication, properties are taken to correspond to some (or all) one-place predicate expressions. In addition to properties, it is usually assumed that there are n-ary relations that correspond to some (or all) n-place predicate expressions (for $n \geq 2$*). A theory of properties is then also a theory of relations.*

In this entry we shall use the traditional labels 'realism' and 'conceptualism' as a convenient way to classify theories. In natural *realism, where properties and relations are the physical, or natural, causal structures involved in the laws of nature, properties and relations correspond to only some predicate expressions, whereas in* logical *realism properties and relations are generally assumed to correspond to all predicate expressions.*

Not all theories of predication take properties and relations to be the universals that predicates stand for in their role as predicates. The universals of conceptualism, for example, are unsaturated concepts in the sense of cognitive capacities that are exercised (saturated) in thought and speech. Properties and relations in the sense of intensional Platonic objects may still correspond to predicate expressions, as they do in conceptual intensional realism, *but only indirectly as the intensional contents of the concepts that predicates stand for in their role as predicates. In that case, instead of properties and relations being what predicates stand for directly, they are what nominalized predicates denote as abstract singular terms. It is in this way that concepts - such as those that the predicate phrases 'is wise', 'is triangular' and 'is identical with' stand for - are distinguished from the properties and relations that are their intensional contents - such as those that are denoted by the abstract singular terms 'wisdom', 'triangularity' and 'identity', respectively. Once properties are represented by abstract singular terms, concepts can be predicated of them, and, in particular, a concept can be predicated of the property that is its intensional content. For example, the concept represented by 'is a property' can be predicated of the property denoted by the abstract noun phrase 'being a property', so that 'Being a property is a property' (or, 'The property of being a property is a property') becomes well-formed. In this way, however, we are confronted with Russell's paradox of (the property of) being a non-self-predicable property, which is the intensional content of the concept represented by 'is a non-self-predicable property'. That is, the property of being a non-self-predicable property both falls and does not fall under the concept of being a non-self-predicable property (and therefore both falls and does not fall under the concept of being self-predicable).*

1 Logical realism and conceptualism in second-order logic

Frege (1879) was among the first to formalize a theory of properties and relations. His theory amounts to a version of standard second-order predicate logic (see Second- and higher-order logics). By 'standard' here is meant an axiomatic formulation that contains among its valid theses all instances of the full impredicative comprehension principle (CP), in which bound predicate variables may occur, so that a property or relation can be specified in terms of a totality to which it belongs. Formally, CP is described as follows:

$$(\exists F)(\forall x_1)\ldots(\forall x_n)[F(x_1,\ldots,x_n) \leftrightarrow \phi],$$

where *F* is an *n*-place predicate variable not occurring free in ϕ and $x_1,\ldots,x_n$ are distinct individual variables.

By CP, every open formula ϕ containing n free individual variables is said to represent a property or relation, which, using the λ-abstraction operator (see Lambda calculus), may also be represented by a λ-abstract, $[\lambda x_1 \ldots x_n \phi]$, as a complex n-place predicate expression. (Here λ is an operator that binds *n* individual variables $x_1,\ldots,x_n$, and when affixed to a formula ϕ results in a complex predicate expression in which all of the occurrences of $x_1,\ldots,x_n$ in ϕ are bound. For example, the complex predicate of being objects *x* and *y* such that

$$F(x,y) \rightarrow (\exists G)[G(x,y) \wedge G(y,x)]$$

can be represented by the $\lambda\text{-abstract}$

$$\{\lambda xy(F(x,y) \rightarrow (\exists G)[G(x,y) \wedge G(y,x)]).)$$

The λ-abstract is taken as a valid substituend of the bound n-place predicate variables in universal instantiation

and to satisfy the principle of λ-conversion,

$$(\forall x_1)\ldots(\forall x_n)([\lambda x_1 \ldots x_n \phi](x_1, \ldots, x_n) \leftrightarrow \phi).$$

Because every open formula constructed from atomic formulas in terms of the logical connectives - including quantifiers for individual and predicate variables - is assumed to represent a property or relation in such a theory, the theory is said to amount to a version of ‘logical realism’ (regarding the existence of universals).

Not every property theory or formal theory of predication will validate CP; that is, not every open formula need represent a property or relation in such a theory. In standard *predicative* second-order logic, for example, formulas containing bound predicate variables are not allowed to occur in the comprehension principle - a constraint that is supposed to restrict the assumption of properties to those that can be represented by first-order formulas where predicate quantifiers do not occur. No λ-abstract in which predicate quantifiers occur will be well-formed in such a theory, and the principle of universal instantiation for predicate quantifiers will be restricted to formulas that contain no bound predicate variables. Formulas with predicate quantifiers can be used to specify properties and relations in *ramified* predicative second-order logic - but only as values of predicate variables of a higher ‘level’ in the ramified hierarchy, so that no property or relation can be specified in terms of a totality encompassing that property or relation. (See Church 1956: §58.) The motivation for such a restriction is unclear, however, if properties and relations are assumed to exist independently of language and thought. In fact, a substitutional interpretation of predicate quantifiers, with predicate variables having only first-order formulas as their substituends, can be given for standard predicative second-order logic - and a similar related interpretation can be given for its ramification - so that the restriction is equivalent to a nominalistic theory of predication, according to which there are no universals - that is, no properties or relations - beyond the predicate expressions of first-order languages. (See Cocchiarella 1986: ch. 1; 1989: §3.)

In a constructive conceptualist theory of predication, where concepts are understood as the rule-following cognitive capacities underlying the use of predicate expressions, the intent of such a restriction on predicate quantifiers is to preclude the possibility of forming impredicative concepts, that is, concepts that presuppose or conceptually involve a totality to which they belong (such as the concept of a least upper bound of a class of real numbers that has upper bounds). The predicative logic of ‘constructive conceptualism’ can be formulated as a nonstandard predicative second-order logic in which not all predicate expressions - now including all λ-abstracts, even those in which bound predicate variables occur - are assumed to represent a predicative concept. Such a logic is ‘free of existential presuppositions’ regarding predicate expressions; and, in particular, no presumption is made independently of meaning postulates as to which predicate constants stand for predicative concepts and which do not. Such a predicative logic can also be ramified, so that concepts can be formed in terms of quantification over concepts formed at a prior ‘stage’ of the ramified hierarchy; but each ‘stage’ of the ramification differs from the ‘levels’ of ramification in standard ramified second-order logic in being ‘free of existential presuppositions’ regarding predicate expressions. Such a nonstandard predicative logic can also be seen as part of a pattern of conceptual development in which impredicative concept-formation becomes possible, that is, as leading to the validation of CP in a ‘holistic’ conceptualism - if, as seems warranted given the history of mathematics, a pattern of reflective abstraction in which an idealized transition to a limit is acknowledged to be conceptually possible. (See Cocchiarella 1986: ch. 2; 1989: §§7-8.)

The nonstandard predicative logic of constructive conceptualism, and, similarly, the impredicative logic of holistic conceptualism, will contain a property theory if it is assumed - in a framework of conceptual intensional realism, for example - that properties and relations are intensional objects and that most (if not all) concepts will have such an intensional object as their intensional content. As intensional objects, properties and relations will then be constituents of propositions as the intensional contents of sentences. But, as already noted, as intensional objects, properties and relations will be the objects that nominalized predicates denote as abstract singular terms (a topic we will return to in §4 below).

2 Natural realism

There is another conception of properties and relations for which it is inappropriate to construe them as intensional objects, and in particular as constituents of propositions. On this alternative, properties and relations are components of states of affairs (which are part of the natural order, as opposed to propositions, which are part of

the intensional order), and in that regard they are the physical, or natural, properties and relations involved in the laws of nature. Such physical or natural properties and relations may correspond to some of our concepts or predicate expressions, but in general not every concept or predicate expression will even purport to represent, still less have, such a natural property or relation corresponding to it. No comprehension principle will be valid when the predicate quantifier is taken to refer to a natural property or relation, accordingly, and any assumption that a formula represents a natural property or relation will always amount at best to an empirical hypothesis of an applied scientific theory.

Traditionally (as, for example, in Aristotle's theory), natural properties were said to exist only *in re*, that is, only in the concrete objects that have such properties, which means that

$$(\forall^{N} F)(\exists x_1)\dots(\exists x_k)F(x_1,\dots,x_k),$$

where '$\forall^{N}$' is a universal quantifier that refers to natural properties and relations when affixed to predicate variables, is valid in such a theory. For this reason, the theory may be called a 'moderate realism'. This thesis is too strong, however, if such a theory is to allow for physical or natural properties and relations - such as those of certain transuranic elements - that *could* be realized in our world as a matter of a natural or causal possibility, even though in fact they never are. In a *modal* moderate realism, the thesis is weakened to

$$(\forall^{N} F)\Diamond^{c}(\exists^{e} x_1)\dots(\exists^{e} x_k)F(x_1,\dots,x_k),$$

where '$\Diamond^{c}$' is a modal operator for a causal or natural possibility, and '$\exists^{e}$' is an existential quantifier that refers only to existing (concrete) objects when affixed to individual variables - and where individual variables and singular terms are otherwise 'free of such existential presuppositions'. The 'concept' of (concrete) existence, defined as

$$E!(x) =_{\mathrm{df}} (\exists^{e} y)(x = y),$$

does not represent a natural property in such a theory - though an abstract property in the sense of an intensional object may be assumed as the intensional content of such a concept in conceptual intensional realism (which can be consistently combined with a natural realism in the general framework of 'conceptual realism'). (See Free logics for more on the logic of singular terms that are 'free of existential presuppositions'.)

3 Actualism and possibilism in modal logical realism

Existence is a property in 'modal logical realism', which is based on a logical or metaphysical notion of possibility and necessity (represented by '$\Diamond$' and '$\Box$'; see Modal logic), and which can be either actualist or possibilist. In *possibilist* modal logical realism, individual variables bound by '$\forall$' or '$\exists$' have possible, as well as actual, objects as their values (and abstract objects as well if such are assumed), and the statement that 'Some things do not exist', formulated as $(\exists x)\neg E!(x)$, is well-formed and meaningful. Predicate variables bound by these quantifiers refer to properties that might apply to nonexisting, as well as existing, objects - though some properties, such as $[\lambda x \neg E(x)]$, apply only to nonexisting objects, whereas others, such as $[\lambda x E(x)]$, apply only to existing (concrete) objects.

In *actualism*, only the actualist quantifiers '$\forall^{e}$' and '$\exists^{e}$' are allowed, so that individual variables bound by these quantifiers have only actual (existing, concrete) objects as their values, and predicate variables bound by these quantifiers have only 'existence-entailing' (e-)properties and (e-)relations as their values - and hence

$$(\forall^{e} F)\Box[F(x_1,\dots,x_n) \rightarrow E!(x_1) \wedge \dots \wedge E!(x_n)]$$

is valid in actualism. Unlike the situation in possibilism, not every open formula or λ-abstract represents an e-property or e-relation, which means that the comprehension principle for actualist modal logical realism must be suitably restricted. (See Cocchiarella 1989: §10.)

4 Nominalized predicates in property theories

In addition to the question of which open formulas or λ-abstracts can be said to represent properties or relations (and can be valid substituends of quantified predicate variables), property theories also differ on the question of whether or not properties and relations have an individual (or 'objectual') as well as a predicable nature, and in particular on whether or not they can also have properties and be 'logical subjects' of predication the way objects

in general are. In Frege's theory, for example, properties and relations are functions from objects to truth-values, where all functions, according to Frege, have an 'unsaturated' nature - a nature that categorially precludes them from being objects, all of which have a saturated nature. Even if predication is not explained in terms of functionality, however, properties and relations can still be construed as having an unsaturated nature that precludes them from being objects. In Wittgenstein's version of 'logical atomism' (see Wittgenstein 1922), for example, and in natural realism as well, material or natural properties and relations are the nexuses (or modes of configuration) of states of affairs, and as such they cannot also be objects configured in states of affairs (see Wittgenstein, L.; Logical atomism §1). Similarly, in conceptualism, predicable concepts, as rule-following cognitive structures underlying the use of predicate expressions, are unsaturated cognitive capacities whose exercise (or saturation) in thought and speech is what informs our speech and mental acts with a predicable nature - which means that concepts have only a predicable and not also an individual (or 'objectual') nature. This does not preclude the intensional content of concepts - which are also called properties and relations in the sense of the Platonist, rather than the Aristotelian, tradition - from being objects; namely, the abstract objects that nominalized predicates denote as abstract singular terms, as opposed to the unsaturated concepts that predicates stand for in their role as predicates. Frege held a view something like this in his version of logicism - except that, because of his commitment to an extensional logic, Frege took the objects denoted by abstract singular terms to be classes rather than intensional objects. The important point in all of these views is that even though the universals that predicate expressions stand for in a given formal theory of predication are not themselves objects (values of the individual variables), some (or all) of them may nevertheless have objects correlated with them as the denotata of the same predicate expressions nominalized as abstract singular terms.

The usual way of representing the nominalization of a simple or complex predicate expression as an abstract singular term is by simply deleting the parentheses and commas that are part of that expression's role as a predicate. Thus, whereas $F(x)$ and $R(x,y)$ are formulas in which F and R occur only as predicates, $G(F)$ and $G(R)$ are formulas in which they occur only as singular terms; and in $F(F)$ and $R(F,R)$, they occur first as predicates and then as singular terms. Similarly, where ϕ is a formula, $[\lambda x\phi]([\lambda x\phi])$ is a formula in which $[\lambda x\phi]$ occurs first as a predicate and then as a singular term. A stronger comprehension principle (CP^*_λ) can now be formulated in terms of identity,

$$(\exists F)([\lambda x_1 \dots x_n \phi] = F),$$

from which, by Leibniz's law (applied now even to abstract singular terms), CP is a consequence.

Bertrand Russell, like Frege, also maintained a version of logical realism; but, unlike Frege, Russell rejected the idea that properties and relations have an unsaturated nature that precludes them from being objects (even though he explained the unity of a proposition in terms of a relation occurring only as a relation and not also as an object). Unlike the situation for Frege, in other words, for Russell a nominalized predicate denotes the same property or relation as an abstract singular term that the predicate otherwise stands for in its role as a predicate. This difference is unimportant as far as Russell's paradox of predication is concerned, however, because the paradox, which can be derived from the following instance of CP (and hence from CP^*_λ) in which a nominalized predicate occurs as abstract singular terms,

$$(\exists F)(\forall x)[F(x) \leftrightarrow (\exists G)(x = G \wedge \neg G(x))],$$

applies to both Frege's and Russell's views of predicates and their nominalizations. Russell's way out of his paradox was the ramified theory of types, which involves grammatically categorizing predicate expressions not only in terms of the hierarchy of 'levels' of ramified second-order logic but also in terms of a 'vertical' hierarchy of orders in which predicates that take nominalized predicates as singular terms must always be of a higher order than the latter (see Theory of types).

Later authors, such as L. Chwistek and F. Ramsey, noted that ramification into levels (of the 'horizontal' hierarchy) is not needed to avoid logical paradoxes such as Russell's, and that a simple theory of types suffices instead - though others, such as A. Church, maintain that the ramified, 'horizontal' hierarchy is still needed for certain intensional or semantic paradoxes. R. Montague (1974) developed an intensional logic for natural language based on a version of simple type theory that was originally formulated by Church. Montague extended Church's theory by adding to it intension- and extension-forming operators (patterned after Frege's *Sinn*/*Bedeutung*

distinction; see Frege, G. §3), ‘ $\wedge$ ’ and ‘ $\vee$ ’, so that where ξ is a well-formed expression of the theory, $^\wedge\xi$ stands for the intension (or sense) of ξ, and $^\vee\xi$ stands for its extension (and hence $^{\vee\wedge}\xi = \xi$ is valid). Properties and relations of entities of types $t_1, \ldots, t_n$ are then taken as the intensions (or senses) of the classes of n-tuples of those types, and the predication of a property or relation (in intension) is defined as membership of the class that the property or relation has as its extension.

5 Conceptual intensional realism

There are other ways to avoid Russell’s paradox besides type theory. One way is to return to standard second-order logic with nominalized predicates as abstract singular terms and restrict λ-abstracts so that only those that can be homogeneously (h-)stratified from an external (metalinguistic) point of view are well-formed. (An expression is ‘h-stratified’ if it is possible to assign natural numbers to the predicates and singular terms occurring in it so that singular terms flanking the identity sign, as well as all of the subject expressions of any predicate expression occurring in it, are assigned the same number, with the predicate itself assigned the successor of that number.) This amounts to restricting the comprehension principle CP^*_λ so that even though the above instance of CP for Russell’s paradox is well-formed, it is not provable, because the λ-abstract

$$[\lambda x(\exists G)(x = G \wedge \neg G(x))]$$

is not h-stratified. The resulting system is equiconsistent with the simple theory of types.

Another way to avoid Russell’s paradox in the context of second-order logic with nominalized predicates is to allow all λ-abstracts, h-stratified or not, to be well-formed and to retain the full comprehension principle CP^*_λ, but allow the first-order part of the logic to be free of existential presuppositions regarding singular terms. The principle of λ-conversion must then be qualified as follows,

$$[\lambda x_1 \ldots x_n \phi](a_1, \ldots, a_n) \leftrightarrow$$
$$(\exists x_1) \ldots (\exists x_n)(a_1 = x_1 \wedge \ldots \wedge a_n = x_n \wedge \phi),$$

where no x_i is free in any a_j, for all i, j such that $1 \leq i \leq j \leq n$. Then, even though

$$(\exists F)([\lambda x(\exists G)(x = G \wedge \neg G(x))] = F)$$

is provable as an instance of CP^*_λ, that is, even though $[\lambda x(\exists G)(x = G \wedge \neg G(x))]$ stands for a predicable concept in its role as a predicate, all that follows by the argument for Russell’s paradox is that it cannot also denote an object in its role as an abstract singular term, because

$$\neg(\exists y)([\lambda x(\exists G)(x = G \wedge \neg G(x))] = y)$$

is also provable. In conceptual intensional realism, this result means that the attempt to ‘object-ify’ the content (in the sense at least of the truth-conditions) determined by the predicable concept represented by $[\lambda x(\exists G)(x = G \wedge \neg G(x))]$ as an intensional object (or property in the Platonist tradition) must fail. In a reconstruction of Frege’s extensional theory, it means that there is no extension (class) corresponding to the property in question. This is not possible in a reconstruction of Russell’s view, however, because for Russell every nominalized predicate denotes the same property that the predicate stands for. (See Russell 1903: §49.)

Finally, there are also strictly first-order property theories. Some are developed along the lines of Zermelo-Fraenkel set theory, with membership replaced by exemplification as a predication relation, and with the deletion at least of the axiom of extensionality (see Schock 1969), and sometimes also of the axiom of foundation (see Jubien 1989). (See Set theory, different systems of.) Some authors prefer to defer assuming specific axioms for predication and simply extend first-order logic to include nominalized predicates as abstract singular terms, leaving the choice open of how predication might be axiomatically developed as a relation regarding the denotata of such singular terms. (See Bealer and Mönnich 1989.) Others develop a first-order theory of properties based on assuming a similarity between revisionist theories of truth and predication as a first-order relation (see Turner 1987).

See also: Logical and mathematical terms, glossary of

NINO B. COCCHIARELLA

References and further reading

Bealer, G. (1982) *Quality and Concept*, Oxford: Clarendon Press.(Accessible to readers with an intermediate-level knowledge of logic; presents a theory of properties, relations and propositions as a first-order intensional logic.)

Bealer, G. and Mönnich, U. (1989) 'Property Theories', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 4, 133-251, 1994.(Argues for a first-order theory of properties, relations and propositions in which nominalized forms of complex predicates are included among the singular terms as 'intensional abstracts'.)

Chierchia, G. and Turner, R. (1988) 'Semantics and Property Theory', *Linguistics and Philosophy* 11: 261-302. (Develops a multi-sorted first-order theory of properties in which predication is analysed along the lines of a revisionist theory of truth.)

Chierchia, G., Partee, B. and Turner, R. (1989) *Properties, Types and Meaning*, vol. 1, Dordrecht: Kluwer.(This collection of papers is the result of a conference held in Amherst in 1986.)

Church, A. (1956) *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press.(An excellent text for the intermediate student; includes axiomatic and semantic formulations not only of propositional and first-order logic, but also of second-order logic, standard predicative and ramified second-order logic.)

Cocchiarella, N.B. (1986) *Logical Investigations of Predication Theory and the Problem of Universals*, Naples: Bibliopolis.(Formal theories of predication are semantically associated in this book with theories of universals for nominalism, logical and natural realism, and constructive and holistic conceptualism. The first part of this book, which can be read by students with an intermediate-level knowledge of logic, deals with predication theories as second-order logics; the second part extends those theories to second-order logic with nominalized predicates as abstract singular terms.)

Cocchiarella, N.B. (1989) 'Philosophical Perspectives on Formal Theories of Predication', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, vol. 4, Dordrecht: Reidel, 254-326.(Discusses in philosophical terms some of the formal theories of predication described in part 1 of Cocchiarella (1986).)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82.(Accessible to most students, although Frege's two-dimensional notation can be somewhat difficult at first; now recognized as a classic.)

Frege, G. (1952) *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980.(The essays 'On Concept and Object' and 'Function and Concept' present a good exposition of Frege's view of properties (which he called 'concepts') and relations as unsaturated functions from objects to truth-values; widely accessible.)

Jubien, M. (1989) 'On Properties and Property Theory', in G. Chierchia, B. Partee and R. Turner (eds) *Properties, Types and Meaning*, Dordrecht: Kluwer, vol. 1, 159-75.(A development of property theory as first-order theory based on a modified Zermelo-Fraenkel set theory.)

Montague, R. (1974) *Formal Philosophy: Selected Papers of Richard Montague*, ed. R.H. Thomason, New Haven, CT: Yale University Press.(See especially the essays 'Universal Grammar' and 'The Proper Treatment of Quantification in Ordinary English' for a development of Montague's intensional logic along the Church-Frege lines. 'Pragmatics and Intensional Logic' includes a formulation of intensional logic along more traditional Russellian lines as a second-order modal logic.)

Parsons, C. (1971) 'A Plea for Substitutional Quantification', *Journal of Philosophy* 68 (8): 231-7. (Argues for a substitutional - as opposed to a referential - interpretation of predicate quantifiers for standard predicative second-order logic. Instead of having properties and relations as their values, bound predicate variables have only open formulas with no predicate quantifiers as their substituends.)

Parsons, T. (1977) 'Type Theory and Ordinary Language', in S. Davis and M. Mithun (eds) *Linguistics, Philosophy, and Montague Grammar*, Austin, TX: University of Texas Press, 127-51. (This paper illustrates what English would be like if it obeyed the rules of a theory of types.)

Russell, B.A.W. (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992.(Includes an account of Russell's early view of properties, which Russell associated, but did not strictly identify, with propositional functions.)

Schock, R. (1969) *New Foundations for Concept Theory*, Uppsala: Library of Theoria, no. 12.(A first-order theory of properties as intensional objects.)

Turner, R. (1987) 'A Theory of Properties', *Journal of Symbolic Logic* 52: 455-72.(A first-order theory of properties based on an assumed similarity between predication and revisionist theories of truth.)

Wittgenstein, L.J.J. (1922) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1961.(The properties and relations of this classic of logical atomism are the nexuses, or modes of configuration, of atomic states of affairs, and as such cannot themselves be configured as objects in states of affairs.)

Prophecy

Most people associate prophecy with prognostication. However, an understanding of philosophical theories of prophecy requires that we recognize the full range of functions that a prophet may serve: oracle, cognizer of the divine, moral and social critic, teacher, political leader, legislator, miracle-worker.

A recurring and fundamental issue, dating to ancient times, is whether prophecy is to be explained naturally or supernaturally. The Muslim philosopher al-Farabi and the Jewish philosopher Maimonides exemplify the naturalist orientation. They understood prophecy as an imaginative 'imitation' or translation of scientific and philosophical truths. Their accounts emphasize not only the intellectual but also the political, legislative and educational functions of prophecy. The Muslim al-Ghazali and the Christian Thomas Aquinas illustrate the supernatural approach, albeit in greatly different ways.

The proposition that biblical prophetic experiences convey scientific and metaphysical knowledge was attacked in varied ways in the modern period. In religious traditions, prophets have been replaced as sources of religious knowledge by either mystics, or books and authoritative interpreters. However, theories of prophecy are linked to important issues about religious language, miracles, the nature of God, the ends of life, and the character of religion.

1 Functions of the prophet

The term 'prophecy' immediately conjures up the ideas of 'prognostication' and 'divination'. However, most philosophical accounts of prophecy, particularly since medieval times, presuppose a far broader conception of the prophet's function. The Greek word *prophetes*, from which the English derives, denotes one who 'speaks for' a god, who imparts divine messages; and in religious traditions, these messages are of various kinds.

While biblical prophets foretell events in the future, religious traditions also imply that they possess divinely imparted knowledge about events in the distant past (for example, creation, or the history of the Jews). According to medieval philosophers, they also possess special knowledge of the structure and workings of the cosmos. Moreover, they bring a moral message. Prophets also perform miracles. Finally, of course, they experience the divine presence in visions, auditions or dreams (see, for example, Isaiah 6). These different functions of the prophet constitute the 'data' that philosophical theories of prophecy have sought to explain.

2 Ancient theories: prophecy and veridical dreams

Philosophical speculation about prophecy began with ancient theories about divination and veridical dreams. Democritus asserted that in dreams, 'images' (*eidōla*) - possibly originating from statues of gods - affect the soul and divulge the future. Aristotle, during an early Platonic stage of his thought, maintained that the soul has a capacity for divination by virtue of its godlike nature. Later, in *On Divination in Sleep* (one of the *Short Natural Treatises*), he maintained that prophecy does not come from a god, since it is not 'the best and the wisest' who experience veridical dreams, but rather the melancholic. Aristotle explored various naturalistic accounts of this phenomenon. Ironically, the notion that prophecy need not occur to the best and the wisest, which Aristotle cited to support naturalism, later became associated with the supernatural view that God can turn whomever he wants into a prophet. An Arabic recension of the *Short Natural Treatises* created the false impression that Aristotle had traced veridical dreams to a cosmic intellect. This misconception influenced later developments among Muslim and Jewish philosophers.

3 Medieval theories: Muslim philosophers

Medieval philosophers debated the extent to which God directly intervenes in the world. Prophecy was implicated in this general debate. Is prophecy divinely bestowed, or is it a natural human attainment? Is prophetic knowledge continuous with everyday or scientific cognition, or is it of a special kind? What faculties must one possess to attain prophecy? Is special preparation required? What is the significance of prophetic imagery? What is meant by divine speech? Philosophers' answers to these questions were anchored in their general metaphysical and epistemological theories; the theory of prophecy is a window into medieval epistemology, which in turn is rooted in medieval metaphysics.

A vital figure in the development of medieval theories of prophecy was the Arabic philosopher al-Farabi. Although discrepancies exist among his various writings and he does not always use the term ‘prophecy’ where one might expect it, the common understanding of al-Farabi as a ‘Neoplatonized Aristotelian’ may be summarized as follows (see al-Farabi §2). Prophecy is attained by individuals of superior intellect by means of an ‘emanation’ from the ‘active intellect’. This emanation flows first to the individual’s passive (potential) intellect - thus actualizing the rational faculty - and then to the imaginative faculty. The imaginative faculty both receives and manipulates physical impressions. The prophet’s (perfected) imagination frames, or represents, scientific and metaphysical truths in sensory images. This enables the prophet to convey aspects of these truths to the untutored masses. Imagination - in concert with the ‘practical’ intellect, which has also received the overflow - is associated as well with lawgiving and political skill. Al-Farabi’s theory expresses his thesis that religion is an imaginative translation - or ‘imitation’- of scientific and metaphysical truth. The fusion of philosophy, law and politics in al-Farabi’s thought is modelled after the philosopher-ruler of Plato’s *Republic*. (In some places, we should note, al-Farabi calls the philosophical, nonimaginative reception of truth ‘revelation’ and ranks it above prophecy.) Veridical dreams and divination, as products of imagination, bear resemblance to prophecy.

While sharing al-Farabi’s naturalism, Ibn Sina (Avicenna) held that the active intellect can impart genuine theoretical knowledge even of forms that the individual did not acquire by abstraction (see Ibn Sina §§2-3, 5). Prophecy of this elevated type is due to a special ‘intuition’ or ‘insight’, and so Ibn Sina’s theory is sometimes characterized as mystical. Reflecting, like al-Farabi, an emphasis on the prophet’s political function, Ibn Sina believed that only prophets can create a bond between people, and he regarded the existence of prophetic lawgivers as a teleological provision of nature. He also developed a naturalistic account of how prophets work miracles: an excellent or noble soul may act upon physical matter, even as the soul acts upon one’s body.

Against al-Farabi and Ibn Sina, other Muslim thinkers refused to see prophecy as naturally acquired. The Ash‘arite al-Ghazali (§4) believed that God is the only true cause of events (see Occasionalism §1). Hence God directly conveys prophetic knowledge: he imparts it to the angels, who pass it on to the prophet. Al-Ghazali regarded naturalistic philosophical approaches to prophecy as contrary to Islam, though he fell into various contradictions (for example, he recognized that the natural faculties of the soul play *some* role in prophecy). At points he also denied the political function of the prophet. Al-Ghazali believed that sciences are possible only by virtue of divine assistance, thereby significantly expanding the range of prophetic phenomena in accord with his metaphysics of divine causation.

4 Medieval theories: Jewish and Christian philosophers

Al-Farabi’s theory of prophecy was adopted, with alterations and with the addition of motifs characteristic of Ibn Sina, by the Jewish philosopher Moses Maimonides (§5) (*Guide to the Perplexed* II, chaps 32-48). Maimonides believed prophecy to be a natural process: ‘an overflow’ (emanation) from God through the intermediation of the active intellect towards the rational faculty and thence the imaginative faculty. (He allowed for a ‘miraculous withholding’ of prophecy from a qualified individual, but then again any natural process is subject to miracles.) The prophets’ possession of imagination in combination with intellect sets them apart from philosophers, while their possessing philosophy in addition to imagination sets them off from soothsayers, diviners and statesmen. The prophecy of Moses is said by Maimonides to be an exception to his general account, but scholars have debated exactly where the difference lies.

Thus, Maimonides and al-Farabi take the object of prophetic knowledge to be natural science and metaphysics. While they recognize differing levels of prophecy, they require the prophet to have satisfied intellectual and not only moral prerequisites. Knowledge of the future depends on a combination of scientific knowledge and imagination. Both al-Farabi and Maimonides regard intellectual perfection, or perhaps intellectual-cum-political perfection, as the most noble end of human life; hence the highest type of prophet represents the most nearly perfect individual.

Maimonides adopts a striking method of biblical exegesis and a distinctive ontology of prophetic images. Prophetic language, as a product of imagination, is a figurative couching for scientific and religious claims and is utilized to communicate with the masses. Thus, Ezekiel’s account of the chariot conveys for Maimonides the metaphysics of Neoplatonized Aristotelianism (*Guide to the Perplexed* III, chaps 1-8). The prophets’ depiction of God in anthropomorphic terms likewise results from imagination (which Maimonides sometimes hints is needed

even to facilitate the prophet's own apprehension). The 'sensory' objects in prophetic dreams and visions are products of imagination, not extramental existents.

Alexander Altmann (1978) has observed that, because Islamic and Jewish philosophers believed that every natural event, including scientific and philosophical cognition, results from a cosmic system of emanations tracing ultimately to God, naturalism about prophecy and providence was more religiously palatable to them than to thinkers whose metaphysics distinguishes sharply between divine activity and the natural order. Nevertheless, medieval critics found naturalism religiously unsatisfying and its accompanying figurative construal of biblical language exegetically implausible. Questions also arose about how, in the absence of direct divine aid, a prophet could predict events with the kind of detail that infuses biblical prophecies, especially in the case of predictions that entail knowledge of 'future contingents', the future free actions of human beings. In addition, the capacity of the prophet to acquire truths that reason cannot - for example, the belief in creation *ex nihilo* (*Guide to the Perplexed* II, ch. 25) - is hard to explain if prophetic knowledge issues naturally from normal philosophical and scientific reasoning. (However, a naturalistic view might allow for a special direct or intuitive knowledge, as per Ibn Sina.)

Several Jewish philosophers before and after Maimonides, such as Judah Halevi and Hasdai Crescas are often thought to have adopted a supernatural view of prophecy. In truth, their theories are ambiguous. Isaac Abravanel's theory is perhaps a better example of a Jewish supernaturalist account of prophecy (see Isaac Abravanel §2).

Christian philosophical treatments of prophecy include Augustine's *De Genesi ad litteram (The Literal Meaning of Genesis)* XII and Aquinas' *Quaestiones disputatae de veritate (Disputed Questions on Truth)*, q.12. Aquinas, as distinct from the naturalists, held that prophecy is a *donum Dei*, a pure gift of God. Prophecy results from unmerited grace, and God might even bestow it on someone who seems unfit. This supernatural account is qualified in several ways. For example, Aquinas recognizes naturalistic explanations of veridical dreams and pagan prophecies; also, when God bestows prophecy on the unfit, he does so by improving the individual's intellectual and imaginative faculties, whence prophecy ensues naturally. (God's activity seems to be further mediated by angelic activity.) Aquinas ignores the political function of prophecy so salient in Islamic accounts and in Maimonides' theory. Altmann (1978) has related Aquinas' theory to his rejection of much of Ibn Sina's ontology.

5 Modern approaches

In the modern period, the epistemic value of prophetic utterances came under varied attacks. Turning Maimonides' emphasis on imagination against him, Benedict de Spinoza (§14) wrote that 'the prophets were endowed with unusually vivid imaginations, and not with unusually perfect minds'. Their scientific and philosophical beliefs merely reflected their own dispositions and prior opinions. Hence 'we are not at all bound to trust them in matters of intellect'. Spinoza's claims signalled the eclipse of the medieval position that prophecy delivers truth, as well as of its concomitant enterprise, the philosophical exegesis of Scripture (*Theologico-Political Treatise*: ch. 2).

The epistemological problem of authenticating prophecy had been well known to medieval philosophers, but subsequently the challenge to validate claims to revelation assumed a more antagonistic cast even on the part of believers. Thomas Hobbes attacked 'visions' as mere daydreams, and remarked famously that to say God has spoken to a person in a dream is 'no more than to say he dreamed that God spake to him'. In the absence of miracles, he said, we have no way to distinguish true revelation (1651: III, ch. 32). John Locke (§7), though himself a believer in divine revelation, held that 'firmness of persuasion is no proof that any proposition is from God' and insisted that a person must use reason 'to enable him to judge of his inspirations, whether they be of *divine* original or no' (1689: IV.xviii-xix).

Eminent thinkers, such as Isaac Newton (§7), accepted the truth of biblical prophecies, arguing that the fulfilment of some gave proof that the others will be fulfilled as well. Declarations that prophecies had been fulfilled were often impugned on the grounds that biblical prophecies were too obscure or ambiguous to be tested. Advancing a more general critique, David Hume (§2) concluded his classic argument against accepting reports of miracles by extending his distrust to reports of true prophecies: 'all prophecies [that is, true predictions not based on induction]

are real miracles' (1748, 1751: sect. X).

With time, various theorists assimilated or reduced prophecy to such categories as ecstasy (an identification already made by the ancients), psychosis, poetic inspiration, hallucination, fabrication and demagoguery. Anthropologists viewed biblical prophecy as a phenomenon with cognates in many cultures, thereby diminishing any special claims of the prophets of the Bible and Qur'an. Even in theological circles, a distinction eventually emerged between the experience of (divine) presence and the experience of content. The content of a prophecy, on this account, represents a subjective, culturally conditioned human response to the experience of presence. Although the medieval view that the prophet exercises imagination also implied prophetic subjectivity, the modern approach implied a denigration of prophetic claims to objective truth and stressed the ethical response rather than the purely cognitive.

These varied challenges to traditional views often went along with deep admiration for the ethical thrust of prophetic religion, and theories respectful of prophecy continued to appear within theological frameworks. However, the overall contrast with medieval times is dramatic.

6 Prophecy in contemporary religion and philosophy of religion

It is sometimes said that prophecy differs from mysticism in its emphasis on God's word and communication of it. Assuming this distinction, we might say that, in the major religions, the role of prophets has been overshadowed by two things: at the one pole, mystics and, at the other, as Hobbes long ago noted, books and authoritative interpreters. Mystics and texts have thus become the primary medium of revelation. Prophecy also figures little in twentieth-century analytically oriented treatments of religion, though there has been some discussion of prophecy in connection with the alleged incompatibility of divine foreknowledge and human free will (see Omniscience §§3-4). Concepts connected to prophecy - revelation, mystical experience, miracles, the paranormal - are explored extensively by philosophers, but usually with scant or no reference to prophecy *per se*. All this is surprising in so far as the claims of traditional theism ultimately rest on the thesis that God communicates with human beings.

See also: Chaldaean Oracles; Eternity; Mysticism, history of; Mysticism, nature of; Providence; Revelation

DAVID SHATZ

References and further reading

Altmann, A. (1978) 'Maimonides and Thomas Aquinas: Natural or Divine Prophecy?', *AJS Review* 3: 1-19.(This article gives more detail about the material concerning ancient philosophy, Maimonides and Aquinas discussed in §§2-4.)

Aquinas, T. (1256-9) *Disputed Questions on Truth*, Chicago, IL: Regnery, 1952. (Treats prophecy in the course of a discussion of predestination and foreknowledge.)

Aristotle (384-322 BC) *Short Natural Treatises* (*Parva naturalia*), ed. W.D. Ross, Oxford: Oxford University Press, 1955.(*On Divination in Sleep* presents a critical account of prophecy as being divinely inspired.)

Augustine (401-14) *De Genesi ad litteram*, trans. J.H. Taylor, *The Literal Meaning of Genesis*, New York: Newman Press, 2 vols, 1982.(Prophecy is discussed in Book XII, which is an influential account of different sorts of vision.)

Davidson, H. (1992) *Alfarabi, Avicenna and Averroes on Intellect: Their Cosmologies, Theories of the Active Intellect, and Theories of the Human Intellect*, New York: Oxford University Press.(A thorough, insightful account. See the index to find items about prophecy.)

al-Farabi (*c*.870-950) *Al-Farabi on the Perfect State: Abu Nasr al-Farabi's Mabadi' Ara Ahl al-Madina al-Fadila*, ed. and trans. R. Walzer, Oxford: Clarendon Press, 1985.(An edition of *The Principles of the Views of the Citizens of the Best State*. Relevant to §3, in conjunction with al-Farabi's *Aphorisms of the Statesman*.)

al-Ghazali (1058-1111) *Mishkat al-Anwar (The Niche of the Lights)*, trans. W.H.T. Gairdner, *Al-Ghazzali's Mishkat al-Anwar*, London: Asiatic Monographs Society, 19, 1924.(Described by Davidson as an 'inadequate' edition; should be read together with Davidson (1992: 134-44).)

Heschel, A.J. (1962) *The Prophets*, New York: Harper & Row.(A two-volume study. In volume 2, Heschel attempts to reply to modern accounts of prophecy and develop his own account. Not highly rigorous, but engaging and suggestive. Relevant to §5.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: University of Cambridge Press, 1991.(Provides grounds for assessing claims of revelation; see III, ch. 32.)

Huby, P. (1979) ‘The Paranormal in the Writings of Aristotle and his Circle’, *Apeiron* 13 (1): 53-62.(Argues that Aristotle shifted from a supernatural to a naturalistic account of prophecy.)

Hume, D. (1748, 1751) ‘On Miracles’, in *Enquiry Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 2nd edn, 1972.(Section X is Hume’s famous critique of miracles, which he extends to prophecy.)

Ibn Sina (first half of 11th century) *Kitab al-Najat*, trans. and ed. F. Rahman, *Avicenna’s Psychology*, Westport, CT: Hyperion Press, 1952, esp. 32-8.(Partial translation of *The Book of Salvation*.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. A. Fraser, Oxford: Oxford University Press, 1894; New York: Dover, 1959.(Locke emphasizes the role of reason in assessing revelation claims in Book IV.xviii-xix.)

Macy, J. (1985) ‘Prophecy in al-Farabi and Maimonides: The Imaginative and Rational Faculties’, in S. Pines and Y. Yovel (eds) *Maimonides and Philosophy*, Dordrecht: Martinus Nijhoff.(A fine analysis of comparisons and contrasts. Also contains a careful analysis of differences between four of al-Farabi’s works on prophecy.)

Maimonides, M. (*c*.1190) *The Guide to the Perplexed*, trans. S. Pines, Chicago, IL: University of Chicago Press, 1963.(Maimonides’ account of prophecy is in part II, chapters 32-48.)

Popkin, R. (1984) ‘Predicting, Prophesying, Divining and Foretelling from Nostradamus to Hume’, *History of European Ideas* 5 (2): 117-36.(An interesting analysis of how Hume undermined the enterprise of studying prophetic predictions.)

Rahman, F. (1958) *Prophecy in Islam*, Chicago, IL: University of Chicago Press. (Contains significant material, including extended quotations, on Islamic thinkers, especially al-Farabi, Ibn Sina and their ‘orthodox’ critics.)

Reines, A. (1970) *Maimonides and Abrabanel on Prophecy*, Cincinnati, OH: Hebrew Union College Press.(An arrangement and explanation of Isaac Abravanel’s critical comments on Maimonides’ *Guide*.)

Spinoza, B. de (1670) *A Theologico-Political Treatise*, trans. R.H.M. Elwes, vol. 1 of *The Chief Works of Benedict de Spinoza*, New York: Dover, 1951.(See especially chapter 2.)

Stump, E. and Kretzmann, N. (1991) ‘Prophecy, Past Truth, and Eternity’, in J. Tomberlin (ed.) *Philosophical Perspectives 5: Philosophy of Religion*, Atascadero, CA: Ridgeview, 395-424.(A discussion, responding to David Widerker, of the special problem that prophetic foreknowledge poses for free will: whereas divine knowledge might be said to be ‘outside time’, that knowledge comes to be ‘within time’ once God has communicated his knowledge to a human prophet. This article suggests a reply.)

Wierenga, E. (1991) ‘Prophecy, Freedom, and the Necessity of the Past’, in J. Tomberlin (ed.) *Philosophical Perspectives 5: Philosophy of Religion*, Atascadero, CA: Ridgeview, 425-45, esp. 433-41.(Attempts to resolve the prophetic foreknowledge-versus-human free will problem described in the Stump-Kretzmann article.)

Propositional attitude statements

Propositional attitude statements - statements about our beliefs, desires, hopes and fears - exhibit certain logical peculiarities. For example, in apparent violation of Leibniz's law of the indiscernibility of identicals, we cannot freely substitute expressions which designate the same object within such statements. According to Leibniz's law, every instance of the following scheme is valid:

a = b
*F(*a*)*
*Therefore, F(*b*)*

The validity of Leibniz's law seems beyond question. It says, in effect, that if an object has a certain property, then anything identical to that object also has that property. Valid instances abound. But consider the following apparently invalid instance:

(1) Hesperus is Phosphorus
(2) Hammurabi believed that Hesperus often rose in the evening
(3) Therefore, Hammurabi believed that 'Phosphorus' often rose in the evening.

If we take 'Hammurabi believed that…often rose in the evening' to serve as the predicate F and 'Hesperus' and 'Phosphorus' to be a *and* b *respectively, this argument appears to be an instance of Leibniz's law. Yet (3) apparently fails to follow from (1) and (2). Hammurabi believed that Hesperus and Phosphorus were two heavenly bodies not one. And he believed that Hesperus did, but that Phosphorus did not rise in the evening.*

We have derived a false conclusion from true premises and an apparently valid law. If that law is really valid, then our argument had better not be a genuine instance of the law. The tempting conclusion, widely accepted, is that we were wrong to construe propositional attitude statements as simple predications. We should not, that is, construe 'Hammurabi believed that…often rose in the evening' to be just a long predicate with the semantic function of attributing some property to the object commonly denoted by 'Hesperus' and 'Phosphorus'. But then the question arises: if attitude reports are not simple predications, what are they? Philosophers have disagreed sharply in their answers. Moreover, their disagreements are intimately connected to a wide range of deep issues about the nature of meaning and reference.

1 Frege on attitude statements

Gottlob Frege was among the first to offer systematic explanations of the form and content of propositional attitude statements (Frege 1892). Central to Frege's treatment of attitude statements is his distinction between sense and reference (see Sense and reference). Frege held that each significant linguistic expression plays two distinct but related semantic roles: it 'denotes' a reference and 'expresses' a sense. For example, the reference of a singular term like a proper name is typically some individual object. A sense, on the other hand, is a way of being given a reference. The sense of a proper name, for example, contains a condition, the unique satisfaction of which by an object *o* is necessary and sufficient to determine *o* as its reference. Perhaps the following captures the sense of 'Phosphorus':

o is Phosphorus just in case *o* is the last celestial object visible in the early morning sky just before sunrise.

The distinction between sense and reference holds not just for proper names but also for sentences. Frege took the reference of a complete declarative sentence to be its truth-value. A declarative sentence is thus a name of sorts - a name for a truth-value. So, for example, the reference of:

(4) Phosphorus can often be seen in the evening sky

is the True, as Frege called it. On the other hand, the sense of a complete declarative sentence is supposed to be a thought, according to Frege. The sense of (4) is the thought that Phosphorus can often be seen in the evening sky.

Frege held that that two expressions may denote the same reference while expressing distinct senses. 'Hesperus' and 'Phosphorus' denote the same object, but differ in sense, from Frege's viewpoint. Something like the

following satisfaction condition arguably captures the sense of ‘Hesperus’:

o is Hesperus just in case *o* is the first celestial object visible in the early evening sky after sunset.

Similarly, although sentence (4) and

(5) Hesperus can often be seen in the evening sky

denote the same reference - namely, the True - they express different thoughts. For the thought that Phosphorus can often be seen in the evening sky is, according to Frege, distinct from the thought that Hesperus can often be seen in the evening sky. One indicator of their distinctness is the possibility that a rational cognizer can simultaneously believe the one, while disbelieving the other. Hammurabi was just such a cognizer.

It is worth enumerating three essential characteristics of Fregean thoughts. First, thoughts are the primary bearers of truth-conditions. The very identity of a thought is constituted by its having the truth-conditions that it has. Sentences, by contrast, have their truth-conditions only derivatively, by being associated with a thought, in virtue of the conventions of a language. Second, thoughts are composite structured entities. A thought is a whole, composed of constituent parts. The parts of thoughts are themselves senses. For example, the thought that Hesperus is often visible in the evening has the sense of ‘Hesperus’ as a constituent part. Third, thoughts are the objects of the propositional attitudes. To believe that Hesperus is often visible in the evening is to stand in a certain relation - the believing relation - to the thought that Hesperus is often visible in the evening. To wonder whether Hesperus is often visible in the evening is to stand in a different relation - the wondering relation - to that same thought.

Frege endorses a principle of compositionality for both sense and reference (see Compositionality). He believes that both the reference and the sense of a complex expression are a function of the references and senses, respectively, of the constituent parts of that sentence (and the way those parts combine to form a whole). Two sentences which differ only by constituents which themselves have the same reference, should have the same truth-value. Similarly, two sentences which differ only by constituents which express the same sense, should themselves express the same thought. Now Frege is convinced that these compositionality principles are nowhere violated. But then the original quandary about attitude sentences remains.

Frege’s way out is to say that despite our initial assessment ‘Hesperus’ as it occurs in (2) and ‘Phosphorus’ as it occurs in (3) do not share a reference. Call the sense and reference of a term as it occurs in direct discourse its ‘customary’ sense and reference. Frege claims that when an expression is embedded within a that-clause, it undergoes a shift in both its sense and its reference. In particular, he holds that obliquely occurring expressions, as he called them, denote not their customary referents, but their customary senses. Moreover, an obliquely occurring term denotes a new sense: what he calls its ‘oblique’ or ‘indirect’ sense. An indirect sense is a mode of presentation not of a term’s customary reference, but of its customary sense. Thus in (2) ‘Hesperus’ refers not to the planet Venus but to the customary sense of ‘Hesperus’. Similarly in (3), ‘Phosphorus’ denotes the customary sense of ‘Phosphorus’. Indeed, a whole sentence undergoes a shift in sense and reference when it is joined with ‘that’ to form a clause. The that-clause in (2), for example, does not denote a truth-value, but the thought customarily expressed by ‘Hesperus often rose in the evening’ when it stands on its own. It denotes, that is, the thought that Hesperus often rose in the evening. And (2) as a whole says that Hammurabi stands in the believing relation to that thought. Similarly, the that-clause in (3) denotes the thought that Phosphorus often rose in the evening. And (3) attributes to Hammurabi the property of believing that thought.

Clearly, if Frege is correct, the invalidity of our original argument does not, after all, violate Leibniz’s law. Since the customary sense of ‘Hesperus’ is not identical to the customary sense of ‘Phosphorus’ our original argument turns out not to involve the substitution of co-referring terms.

2 Referential opacity

Philosophers have sometimes questioned the coherence of Fregean senses. W.V. Quine, for example, has argued that senses and related intensional notions are ‘creatures of darkness’ (Quine 1956). There is no saying, according to Quine, for arbitrary senses *x* and *y*, when *x* and *y* are the same sense and when they are two distinct senses (see Quine, W.V. §8; Radical translation and radical interpretation §§2-3). There can be no entities, he insists, where

there are no determinate criteria of identity. But if senses go, so too must Frege's account of propositional attitude statements.

Quine endorses Frege's negative conclusion that obliquely occurring terms do not play their customary referential roles. He offers, however, a markedly different diagnosis of that failure. Like Frege, Quine believes that where reference happens substitutivity reigns - that is, Leibniz's law holds. But unlike Frege, Quine takes the failure of substitutivity within attitude statements at face value. And he concludes, in effect, that reference cannot be happening within the context of propositional attitude statements or that-clauses generally. That-clauses, he claims, are 'referentially opaque'. Embed a term which elsewhere functions referentially within such a clause and substitution is no longer permissible, not because reference shifts, but because the embedded term is stripped of its referential function. Quine suggests, in fact, that it is a mistake to speak of 'embedding' a referring expression within a that-clause at all. He suggests that an expression like 'believed that Hesperus often rose in the evening' functions rather like a primitive one-place predicate with no 'logically germane' constituent parts. In that case, sentences like (2) are misleadingly spelled. A more revealing spelling would treat this predicate as a single, though very long word somewhat like: 'believed-that-Hesperus-often-rose-in-the-evening'. The presence of the string 'Hesperus' in this very long word is merely an accident of orthography. It would be no more correct to regard 'Hesperus' as a grammatical constituent of this very long word, than it would be to regard the word 'cat' as grammatical constituent of the word 'cattle'.

Referential opacity explains another logical peculiarity of that-clauses, according to Quine. Consider the law of existential generalization:

F(*a*)
Therefore, $(\exists x)\mathrm{F}(x)$

Informally, this scheme says that if *a* has F then there is something or other which has F. Existential generalization allows us to infer from the premise that 'John loves Mary' to the conclusion that 'John loves someone or other'. When we try to apply this evidently valid scheme within the context of a propositional attitude verb, we are led into apparent contradiction. Consider the following sentence:

(6) The president suspects that someone is a spy.

As it stands, (6) is subject to two different readings. On one reading, (6) says that the president 'suspects true', as we might put it, a certain statement or proposition - the statement or proposition that spying sometimes happens. This reading has been variously called the '*de dicto*', 'notional', or 'referentially opaque' reading (see De re/de dicto). There is a second reading of (6) which attributes to the president not just the generalized suspicion that spying sometimes happens, but the state of being 'suspicious of' some person or other. We might paraphrase this reading of (6) by:

(6b) The president suspects someone of being a spy

or by:

(6c) There is someone whom the president suspects of being a spy.

This reading is variously called the '*de re*', 'relational', or 'referentially transparent' reading. The two readings are not generally equivalent. It is possible to suspect that spying happens even when one has no suspicions about any particular person. The ambiguity in sentences like (6) is sometimes traced to an ambiguity of quantifier scope. On the so-called *de dicto* or opaque reading of (6), the quantifier has narrow scope relative to 'believes'. On the so-called *de re* or transparent reading, the quantifier has wide scope relative to 'believes'. As a first pass, we might represent the *de dicto* reading by (6d) and the *de re* reading by (6e):

(6d) The president suspects that $(\exists x)(x \text{ is a spy})$
(6e) $(\exists x)(\text{The president suspects that } x \text{ is a spy})$.

Quine has argued that constructions like (6e), in which a quantifier sitting outside the scope of the attitude verb 'reaches in' to bind a variable that lies within the scope of that attitude verb, are deeply problematic. We cannot,

he claims, 'quantify into' attitude contexts (Quine 1956). Or to put it in slightly different terms, propositional attitude verbs appear to block the interior reach of exterior quantifiers.

Suppose that the president has observed Stanley P. Young, a certain low-level White House aide, behaving in a rather furtive manner. He says with great conviction to the chief of White House security, 'Stanley P. Young is a spy.' That seems a sufficient basis for attributing to the president the belief that Stanley P. Young is a spy and thus a sufficient basis for taking this to be true:

(7) The president believes that Stanley P. Young is a spy.

If we construe 'The president believes…is a spy' as a complex predicate, then from (7) and the law of Existential Generalization, we should be able to infer:

(8) $(\exists x)(\text{The president believes that } x \text{ is a spy})$.

There may be a number people whom the president believes to be spies. Stanley P. Young is one such person. $(8')$ below would also seem to follow:

$(8')$ $(\exists x)(x = \text{Stanley P. Young \& the president believe that } x \text{ is a spy})$.

Now suppose that the president is subsequently introduced to one James Q. Money, a generous contributor to progressive causes and a man of stellar reputation. The president trusts Mr Money. When asked whether Money is a spy, the president denies even the possibility. That denial is grounds for inferring:

(9) The president believes that James Q. Money is not a spy.

From (9) and Existential Generalization we can infer:

(10) $(\exists x)(\text{The president believes that } x \text{ is not a spy})$.
$(10')$ $(\exists x)(x = \text{James Q. Money \& the president believe that } x \text{ is not a spy})$.

But unbeknownst to the president, James Q. Money leads a double life. He is none other than Stanley P. Young. Does the president believe or not believe James Q. Money, that is, Stanley P. Young, to be a spy? Intuition pulls in opposite directions. Since James Q. Money just is Stanley P. Young, we should, by Leibniz's law, be able to substitute 'Young' for 'Money' in $(10')$ without change of truth-value. Since this substitution happens outside the that-clause, it should be unaffected by the logical peculiarities of such clauses. Yet substitution yields:

$(10'')$ $(\exists x)(x = \text{Stanley P. Young \& the president believe that } x \text{ is not a spy})$.

Examples $(10'')$ and $(8')$ are of dubious consistency. No one can both be and not be a spy. So how can anyone consistently both believe a person to be a spy and believe that very person not to be a spy?

The answer lies in referential opacity. Because (7) and (9) are opaque, they do not ascribe to the president incompatible relations to Money, that is, Young. Indeed, opaquely construed, the Quinean insists, neither (7) nor (9) relates the president to Money, that is, Young at all. Recall again the Quinean view that when opaque contexts are perspicuously spelled the illusion that they contain referring expressions as proper constituents will have been dispelled. Indeed, opacity renders quantified sentences like $(10')$ and $(10'')$ utterly nonsensical. For proper respelling reveals that there is no genuine occurrence of an interior variable for the exterior quantifier to bind. An analogy with quotation - which Quine takes to be the referentially opaque context *par excellence* - makes the reason clear. Consider:

(11) '*x* is a spy' is an open sentence.

Enclosing an expression within quotation marks seals that expression off from the reach of external quantifiers. The quantifier below does not bind the quoted variable:

(12) (∃*x*)('*x* is a spy' is an open sentence).

What occurs within the quotation marks is really not a variable *in use*. It is rather a variable *being mentioned* (see

Use/mention distinction and quotation). Similarly, though what occurs in (10′) and (10″) bears a certain orthographic resemblance to a variable in use, it is not the real thing. For that reason we can no more quantify into propositional attitude contexts, Quine insists, than we can quantify into the context of quotation.

Quine's approach promises a way out of the threatened contradiction, but only by doing great violence to a number of potent intuitions. It seems frankly incredible that putatively referring terms embedded within a that-clause are mere grammatical illusions. Further, if propositional attitude statements lack logically germane constituents, there will be no accounting for certain systematic commonalities and differences among such statements. Compare and contrast, for example, (7) and (9) above with (13) and (14) below:

(13) The First Lady *doubts that* Young is a spy
(14) The vice-president *wonders whether* Young is a spy.

Example (13) seems to say that the First Lady doubts what the president believes, while (14) says that the vice-president has questions about what the First Lady and the president disagree about. Moreover, the president's belief, the First Lady's assurance and the vice-president's questions seem all to be about one and the same individual. Quine's approach abandons these intuitive judgments.

Quine is well aware of these costs. He believes, however, that we must pay the cost if we are to keep the creatures of darkness - Fregean senses and other intensional entities - at bay (see Intensional entities).

3 Innocence regained?

Donald Davidson has argued that if we could but regain our pre-Fregean semantic innocence:

> it would seem to us plainly incredible that the words 'The earth moves' uttered after the words 'Galileo said that' mean anything different, or refer to anything else, than is their wont when they come in other environments.
>
> (Davidson 1969: 172)

If Davidson is right then we can account for the logical peculiarities of attitude statements without positing Fregean reference shifts. Nor, if he is right, need we allow that otherwise referring terms are stripped of their referential functions when they occur in attitude contexts. Indeed, against Quine, Davidson holds that unless we have already been forced to introduce Fregean senses for other reasons, the peculiarities of attitude contexts provide no additional pressures to posit such entities. So if Davidson is right, Quine's drastic departures are not needed to keep intensions at bay.

Davidson defends what he calls a 'paratactic analysis' of the logical forms of indirect discourse, though his approach can be extended fairly directly to attitude ascriptions. Davidson's central claim is that a sentence like:

(15) Galileo said that the earth moves

is really a parataxis of two sentences. The first sentence of the parataxis in (15) ends with the word 'that'; the second starts with 'the earth'. By adding a bit of punctuation, we can represent (15) in a more perspicuous manner as:

(16) Galileo said that. The earth moves.

Call the first sentence the attribution sentence and the second the content sentence. In the attribution sentence of (16), the 'that' is a demonstrative pronoun which refers (on an occasion) to an utterance of the second sentence of (16).

Sentence (16) will be true, according to Davidson, just in case some utterance of Galileo's makes Galileo and the utterer of (16) 'samesayers'. Suppose that Galileo utters (17) and Smith utters (18):

(17) Eppur si muove.
(18) The earth moves.

In some sense, Galileo and Smith say the same thing. Moreover, if Smith utters (18) and Galileo utters (17) then

'the earth' in Smith's mouth and 'Eppur' in Galileo's mouth have the same reference.

Suppose that Smith wants to attribute Galileo's statement to Galileo. He can do so simply by saying the same thing as Galileo and then saying that he has just done so - as in the following scenario:

(19) Galileo: Eppur si muove.
(20) Smith: The earth moves. Galileo said that (too).

Here we have Galileo saying, in Italian, that the earth moves. We have Smith saying the same thing in English. And then we have Smith ascribing such a statement to Galileo. He does so by making a statement that turns himself and Galileo into samesayers (to use Davidson's phrase) and then stating that he has just done so. As matters stand, it remains unclear until Smith utters the second sentence of (20) just what the point of his initial utterance may be. But we can rectify that by reversing the order of the sentences as follows:

(21) Galileo said that. → The earth moves.

The arrow in (21) should be understood as a demonstration accompanying the demonstrative pronoun 'that'. The demonstrative refers to Smith's second utterance - the one by which Smith purports to turn himself and Galileo into samesayers. While Smith 'refers' to his own utterance, he implicitly 'quantifies over' Galileo's utterances. Smith says, in effect, that the demonstrated utterance of his and 'some utterance or other' of Galileo's make himself and Galileo samesayers.

Davidson's account yields a relatively straightforward explanation of the invalidity of inferences like the following:

(22) Galileo said that. → The earth moves.
(23) The earth is the third planet from the sun.
(24) Galileo said that. → The third planet from the sun moves.

There is no logical connection between the effect of substitution on the content sentence and the truth-values of the two attribution sentences. As Davidson puts it:

> There is no reason to predict, on grounds of form alone, any *particular* effect on the truth of [the attribution sentence] from a change in [the content sentence]. On the other hand, if the [content sentence] had been different in any way at all, [the attribution sentence] *might* have had a different truth-value, for the reference of the 'that' would have been changed.
>
> (Davidson 1969: 172)

Though Davidson's central idea is quite ingenious, there are both syntactic and semantic grounds for scepticism. For example, content clauses are not generally introduced by demonstratives. Neither 'whether' in 'Galileo wondered whether the earth moves' nor 'for… to' in 'I would prefer for you to leave' can plausibly be analysed as overt demonstratives. Moreover, demonstratives cannot in general be deleted, but the 'that' in '… said that… ' can. Compare the following:

(25) Davidson saw that man over there.
(26) Davidson said that the paratactic analysis is true.
(27) ?Davidson saw man over there.
(28) Davidson said the paratactic analysis is true.

It is also difficult to see how to extend the paratactic analysis to *de re* ascriptions. Consider (29):

(29) Every philosopher believes that he is wise.

Example (29) says, in effect, that every philosopher has a high opinion of himself. But if we interpret (29) along Davidsonian lines as:

(30) Every philosopher believes that. He is wise.

we get something rather different.

Finally it is clear that we need to know more about the samesaying relation. If the paratactic analysis is to overcome the need to introduce Fregean senses and the like, then samesaying cannot just be a matter of expressing the same Fregean sense. Davidson owes us an account of samesaying which eschews senses and other intensional entities altogether. Davidson is fully aware of this debt and labours mightily to pay it (Davidson 1969).

4 Structured complexes

A class of theories which has recently gained favour with a number of philosophers is what might be called the structured complex approach (see Intensional logic §5). Advocates of such theories maintain that attitude verbs express relations between agents and complex hierarchically structured entities of one sort or another. The core idea dates back to Carnap (1947). It has resurfaced, much modified, in Cresswell (1984), Segal (1989), Richard (1990), Higginbotham (1991) and Larson and Ludlow (1993). Different versions of this approach posit different structured complexes. Carnap posited hierarchically structured intensions. Others pair semantic values with something else: sometimes a sentence, sometimes a description of a sentence, sometimes a structured intension. Larson and Ludlow, for example, hold that that-clauses specify Interpreted Logical Forms (ILFs). An ILF is a kind of product of a phrase structure tree for a sentence and the sequence of semantic values for the semantically valued constituents of that sentence. On their approach, the sentence:

(31) Pierre admires London

will be associated with the ILF in Figure 1:

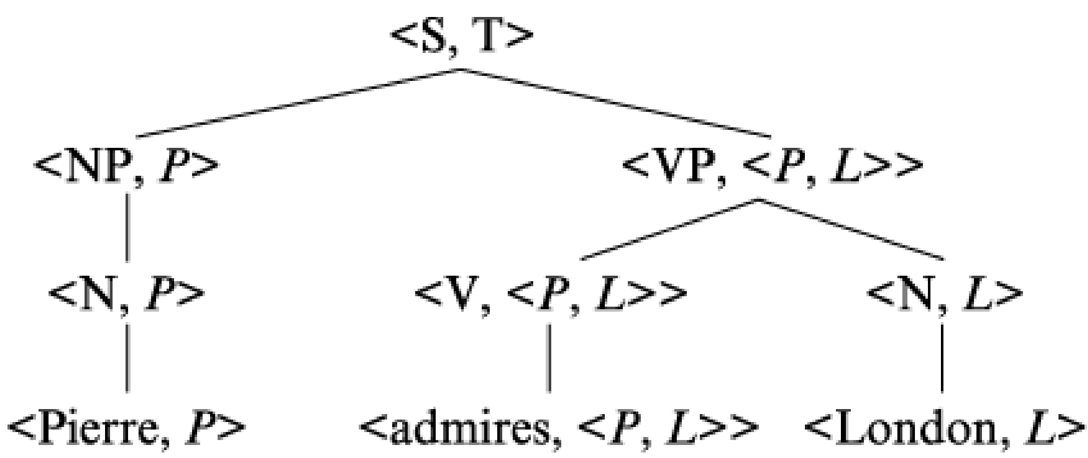

Figure 1

We need not worry about the exact formation rules that generate this ILF. The crucial point is that each node in this tree is labelled by a pair whose left member is either a lexical item (at the bottom-most nodes) or a phrase marker (at all other nodes) and whose right member is a semantic value. For example, the pair $\langle \text{Pierre}, P\rangle$ is the pair whose left member is the word 'Pierre' and whose right member is the reference of 'Pierre' (namely, Pierre himself). The pair $\langle \text{admires}, \langle P,\ L\rangle\rangle$ has left member 'admires' and a right member consisting of a pair whose members are Pierre and London. Such trees are intended to represent a sequence of semantic values and the order in which those semantic values are associated with the expressions whose values they are. Because lexical items (words themselves), and not just phrase makers, are constituents of ILFs, the following two sentences will have distinct ILFs:

(32) Hesperus rises in the evening.
(33) Phosphorus rises in the evening.

Nonetheless, (32) and (33) have, according to Larson and Ludlow, the same semantic content. That is, (32) and (33) have formally distinct but semantically equivalent ILFs. Armed with the notion of formally distinct but semantically equivalent ILFs, Larson and Ludlow are able to explain how it is possible for attitude statements like our original (2) and (3) to differ in truth-value. The thought is that an agent who stands in the believing relation to

the ILF of ‘Hesperus rises in the evening’ need not *ipso facto* stand in the believing relation to the ILF of ‘Phosphorus rises in the evening.’ Yet one who believes that Hesperus rises believes the very same semantic content as one who believes that Phosphorus rises.

Notice that Larson and Ludlow walk the middle ground between Davidson and Frege. Their approach is ontologically conservative like Davidson’s. For the structural descriptions and semantic values out of which ILFs are composed must already be introduced in to our ontology to account for the semantic and syntactic character of sentences other than those about propositional attitudes. But their explanation of failures of substitutivity is analogous to Frege’s own. They too appeal to the fact that, at least in attitude contexts, more is relevant to substitutivity than mere co-reference. That more is, however, not anything semantic. So we are not required, merely because of the peculiarities propositional attitude statements, to posit for each referring expression a second semantic role over and above its role of standing for a referent.

5 Hidden indexicality

Consider the following three theses:

(a) *The Direct Reference Thesis.* The sole semantic role of a proper name is to stand for its bearer.
(b) *Semantic Innocence.* A name plays the same semantic role when embedded within a that-clause as it plays when it occurs outside a that-clause.
(c) *The Fregean Intuition.* Propositional attitude statements which differ only by co-referring proper names may differ in truth-value.

Each of these theses has considerable independent plausibility. But they are not obviously consistent. Frege was moved to abandon both thesis (a) and (b) because he thought they could not be reconciled with (c) (see Proper names §5). More recently, Nathan Salmon (1986) has abandoned (c) out of a commitment to (a) and (b). He has insisted that it is merely a pragmatically generated illusion that sentences like (2) and (3) have different truth-values. He claims that strictly, literally speaking, one who believes that Hesperus rises *ipso facto* believes that Phosphorus rises. It is just that it is conversationally misleading to talk that way. That is because belief is not a two place relation between a believer and a proposition, but a three place relation between a believer, a proposition and what we might call a ‘way of believing’ that proposition. Propositions are never believed *simpliciter*, but always in some way or other, under some guise or other. Moreover, Salmon insists, a rational cognizer can believe a given proposition under one guise while failing to believe it under a different guise. But the crucial further point is that nothing in an attitude ascription semantically specifies the guise under which the proposition named by the that-clause is supposed to be believed. The guise simply does not enter into the strict literal truth-conditions of an attitude statement. Nonetheless, Salmon suggests, we can pragmatically suggest something about the guise, by using one expression rather than another semantically equivalent one in making an attribution. If we use ‘Hesperus’ we suggest something different about the guise than we would if were to use ‘Phosphorus’. That is why it is pragmatically misleading to use ‘Hesperus’ and ‘Phosphorus’ interchangeably in making attitude ascriptions.

By contrast, hidden indexicalists insist on all of our principles, and offer a strategy, with several variants, for reconciling them (see Schiffer 1979; Crimmins and Perry 1989; Richard 1990; Crimmins 1992). Proponents of the hidden indexical approach generally agree with the view that belief is a three-place, rather than merely a two-place relation. They even agree that that-clauses which differ only by co-referring proper names must refer to the same (singular) proposition. But they insist that attitude ascriptions do not just pragmatically suggest information about the guise under which the specified proposition is believed. Information about the guise is part of what is strictly, literally said by an attitude ascription. That is why, against Salmon, the Fregean intuition is here to stay.

Yet, given the assumption of semantic innocence, how could that-clauses which differ only by corresponding co-referring proper names specify distinct ways of believing? The answer, according to the hidden indexicalist, is that there is more to an attitude statement than meets either the eye or the ear. Attitude reports contain hidden or unarticulated indexical constituents. These unarticulated constituents are not reflected in the surface grammatical structure of the sentence, but in a context of utterance they somehow refer to a way of believing. Consequently, attitude sentences, in context, manage to express propositions with more explicit constituents than they themselves have. Such sentences can be compared, the hidden indexicalist says, to sentences like the following in which one term of a relation is suppressed, but can be supplied, in context, by competent producers and consumers:

It is raining.	(Where?)
I am ready.	(For what?)
That is tall.	(For a what?)
Smith believes that p.	(How?)

When one utters, ‘It is raining’ on a certain occasion, one strictly, literally says that it is raining in a certain place - typically the place of utterance - even though no explicit constituent of one’s utterance refers to that place. Just so, the hidden indexicalist maintains, when one utters a belief report on an occasion, one strictly, literally says that a certain proposition is believed in a certain way. And one does so even though no explicit constituent of one’s utterances refers to the way of believing. Now the crucial further claim has to be that although ‘Hesperus’ and ‘Phosphorus’ have the same semantic value and function innocently, even in that-clauses, nonetheless, exchanging one for the other can affect the way an unarticulated constituent gets hooked up to a reference as a function of context. And that can explain why they are not freely substitutable one for another.

Clearly, the hidden indexicalist owes us a detailed account of these matters. We need to be told what a way of believing is. We need to know how specified ways of believing vary as a function of facts about context and facts about articulated constituents. Fortunately, the best advocates of this approach have gone a long way towards answering such questions.

See also: Demonstratives and indexicals; Indirect discourse; Propositional attitudes

KENNETH A. TAYLOR

References and further reading

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press; enlarged 2nd edn, 1956. (Introduces the notion of an intensional isomorphism, the earliest version of the structured complex approach.)

Cresswell, M. (1984) *Structured Meanings*, Cambridge, MA: MIT Press.(Argues for structured intensions as objects of propositional attitudes.)

Crimmins, M. (1992) *Talk about Beliefs*, Cambridge, MA: MIT Press.(Defends a hidden indexical theory.)

Crimmins, M. and Perry, J. (1989) ‘The Prince and the Phone Booth: Reporting Puzzling Beliefs’, *Journal of Philosophy* 60: 435-45.(Defends a hidden indexical theory.)

Davidson, D. (1969) ‘On Saying That’, in D. Davidson and J. Hintikka (eds) *Words and Objections: Essays on the Work of W.V. Quine*, Dordrecht: Reidel, 158-74. (Articulates the paratactic analysis of indirect discourse.)

Davidson, D. (1967) ‘Truth and Meaning’, *Synthèse* 17: 304-23.(Articulates Davidson’s broader semantic programme.)

Devitt, M. (1996) *Coming to our Senses*, Cambridge: Cambridge University Press. (Defends a neo-Fregean account of attitude statements without positing Fregean senses; rejects hidden indexical approaches and direct reference.)

Fodor, J. (1989) ‘Substitution Arguments and the Individuation of Beliefs’, in G. Boolos (ed.) *Method, Reasons, and Language*, Cambridge: Cambridge University Press.(Construes modes of presentations as expressions in the language of thought, while holding a referentialist semantics.)

Frege, G. (1892) ‘Über Sinn und Bedeutung’ (On Sense and Reference), *Zeitschrift für Philosophie und philisophische Kritik* 100: 25-50.(A seminal work of Frege’s in which he argues for the distinction between the sense and reference of expressions and presents a number of important theses regarding the two notions. Translated as ‘On Sense and Meaning’ in P.T. Geach and M. Black (eds) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 3rd edn, 1980.)

Higginbotham, J. (1991) ‘Belief and Logical Form’, *Mind and Language* 6: 344-69. (Defends a version of the structured complex approach, drawing on the resources of contemporary syntactic theory. Relates this approach to Davidsonian extensional semantics.)

Larson, R. and Ludlow, P. (1993) ‘Interpreted Logical Forms’, *Synthèse* 95: 305-57.(Articulates and defends the ILF approach.)

Quine, W.V. (1956) ‘Quantifiers and the Propositional Attitudes’, repr. in *The Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press, revised and enlarged edn, 1976.(This widely reprinted article contains Quine’s argument against the permissibility of quantifying into opaque contexts.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: Harvard University Press. (Forcefully articulates Quine’s disdain for intensions as ‘creatures of darkness’.)

Recanati, F. (1993) *Direct Reference*, Oxford: Blackwell.(A full-scale defence of the direct reference thesis which articulates yet another way to reconcile direct reference, innocence and the Fregean intuition.)
Richard, M. (1990) *Propositional Attitudes: an Essay on Thoughts and How We Ascribe Them*, Cambridge: Cambridge University Press.(Argues that propositional attitude statements express relations to Russellian annotated matrices - a version of the structured complex approach; also defends the view that belief reports are indexical.)
Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press.(Denies the Fregean intuition and defends innocence and direct reference.)
Schiffer, S. (1979) 'Naming and Knowing', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press. (First appearance of the hidden indexical approach.)
Segal, G. (1989) 'A Preference for Sense and Reference', *Journal of Philosophy* 86: 73-89.(Another defence of a structured complex approach.)
Soames, S. (1987) 'Direct Reference, Propositional Attitudes, and Semantic Content', *Philosophical Topics* 15: 47-87.(Denies the Fregean intuition, while defending innocence and direct reference.)

Propositional attitudes

Examples of propositional attitudes include the belief that snow is white, the hope that Mt Rosea is twelve miles high, the desire that there should be snow at Christmas, the intention to go to the snow tomorrow, and the fear that one shall be killed in an avalanche. As these examples show, we can distinguish the kind of attitude - belief, desire, intention, fear and so on - from the content of the attitude - that snow is white, that there will be snow at Christmas, to go to the snow, and so forth. The term 'propositional attitudes' comes from Bertrand Russell and derives from the fact that we can think of the content of an attitude as the proposition the attitude is towards. It can be typically captured by a sentence prefixed by 'that', though sometimes at the cost of a certain linguistic awkwardness: it is more natural, for example, to talk of the intention to go to the snow rather than the intention that one go to the snow. The most frequently discussed kinds of propositional attitudes are belief, desire and intention, but there are countless others: hopes, fears, wishes, regrets, and so on.

Some sentences which contain the verbs of propositional attitude - believes, desires, intends, and so on - do not make ascriptions of propositional attitudes. For example: 'Wendy believes me', 'John fears this dog', and 'He intends no harm'. However, while these sentences are not, as they stand, ascriptions of propositional attitudes, it is arguable - though not all philosophers agree - that they can always be analysed as propositional attitude ascriptions. So, for example, Wendy believes me just in case there is some p such that Wendy believes that p because I tell her that p; John fears this dog just in case there is some X such that John fears that this dog will do X and so on.

Discussions of propositional attitudes typically focus on belief and desire, and, sometimes, intention, because of the central roles these attitudes play in the explanation of rational behaviour. For example: Mary's visit to the supermarket is explained by her desire to purchase some groceries, and her belief that she can purchase groceries at the supermarket; Bill's flicking the switch is explained by his desire to illuminate the room, and his belief that he can illuminate the room by flicking the switch; and so on. It is plausible - though not uncontroversial - to hold that rational behaviour can always be explained as the outcome of a suitable belief together with a suitable desire.

Some philosophers (examples are Grice and Schiffer) have used the propositional attitudes to explain facts about meaning. They hold that the meanings of sentences somehow derive from the contents of relevantly related beliefs and intentions. Roughly, what I mean by a sentence S is captured by the content of, say, the belief that I express by saying S.

One fundamental question which divides philosophers turns on the ontological status of the propositional attitudes and of their contents. It is clear that we make heavy use of propositional attitude ascriptions in explaining and interpreting the actions of ourselves and others. But should we think that in producing such ascriptions, we attempt to speak the truth - that is should we think that propositional attitude ascriptions are truth-apt - or should we see some other purpose, such as dramatic projection, in this usage? Or, even more radically, should we think that there is nothing but error and confusion - exposed by modern science and neurophysiology - in propositional attitude talk?

1 Elementary distinctions

Propositional attitude ascriptions standardly take the form 'X Fs that p', where 'X' denotes the subject of the attitude, 'F' is a verb of propositional attitude, and 'that p' gives the content of the attitude. So, for example, in the sentence 'John believes that snow is white', 'John' denotes the subject of the attitude, 'believes' is the verb of propositional attitude, and 'that snow is white' gives the content of the attitude.

Taking the form of these ascriptions at face value, it is natural to suggest that a propositional attitude ascription of the form 'X Fs that p' is true just in case X is in a propositional attitude state of type F - a F-state - which has the content that p. So, for example, 'John believes that snow is white' is true just in case John is in a belief-state which has the content that snow is white; 'Mary hopes that Sheila is happy' is true just in case Mary is in a hope-state which has the content that Sheila is happy; and so on. More generally, it is natural to divide a discussion of propositional attitudes into two parts: one part which focuses on the nature of propositional attitude states - the difference between hoping, fearing, believing, wishing, and so on - and one part which focuses on propositional

attitude contents - the difference between believing that snow is white, believing that snow is grey, believing that snow falls at night, and so forth.

It is plausible to think that belief and desire are the basic exemplars of two quite different kinds of attitudes, with characteristically different directions of fit. On the one hand, there are attitudes, like belief, which aim to fit the world - and, hence, which are important for theories of truth, impact of evidence, credence, and so on. And, on the other hand, there are attitudes, like desire, which aim to have the world fit them - and, hence, which are important for theories of value, virtue, wellbeing, and so on. Some philosophers think that all of the propositional attitudes can be explained in terms of beliefs and desires; they think that the apparent multiplicity of kinds of propositional attitudes is merely apparent - hope, for example, is a kind of desire about the future - and that direction of fit is the only fundamental dimension which needs to be considered in classifying the propositional attitudes.

2 Propositional attitude states

If beliefs, desires, intentions, and the like, are states of subjects, what kind of states are they? A plausible view is that they are functional states - states defined or determined or specified by their functional roles. According to functionalist theories, propositional attitude states are fully determined by their causal relations to one another, to perceptual inputs, and to behavioural outputs (see Functionalism). So, for example, what makes a certain state a state of believing that there are *Cheerios* in the cupboard turns on its being typically caused by packets of *Cheerios*, or more precisely, on the role it plays in the processing of perceptual input - say, guessing the contents of that visually presented cardboard box in the cupboard; on the way it leads to other states - say, believing that there is cereal in the cupboard; and on the way it combines with hunger and desire to eat some cereal to produce behaviour that leads to eating.

There are, however, two different ways of thinking of functional states: as role states or as realizer states. Realizer states are first-order states that stand in the functional roles; they are the particular states which play the roles. Role states are second-order states: the state of being in, some first-order state playing a certain kind of functional role. So, in the case of belief, say, there is the first-order realizer state (the state which actually fills the belief role - the state which mediates between perceptual inputs and behavioural responses in the way distinctive of belief); and there is the second-order role state (the state of being in some state or other which realizes the belief role - the state of being in a state that mediates in the way distinctive of belief between perceptual inputs and behavioural outputs). Should we think of belief in particular and propositional attitudes in general as realizer states or as role states?

The obvious reason for holding that propositional attitude states are realizer states rather than role states is that the propositional attitudes are centrally involved in the causation of behaviour, and it is the realizer states but not the role states that are centrally involved in the causation of behaviour. It is, for instance, the particular internal state which fills the role we associate with believing that snow is white that causes the behaviour distinctive of this belief - for example, uttering the words ‘Snow is white’.

There are, however, also considerations that point the other way, that favour holding, for example, that my belief that snow is white is the role state which is realized by some particular internal state of mine. In particular, it seems that taking propositional attitude states to be role states allows one to formulate psychological generalizations which one would miss if one took propositional attitudes to be realizer states. Consider, for example, the generalization: everyone who believes the world is going to end in ten minutes gets anxious. Suppose that there are actually M realizer states for the belief: $N_1, N_2, \ldots, N_M$. These states, we may suppose, each play the appropriate causally intermediate roles distinctive of the belief that the world is going to end in ten minutes in somewhat the way that transistors and valves, despite being very different, can both play the roles distinctive of amplification. If the belief state is identical with the realizer states, then the only way to generalize is to treat the realizer state as a disjunction $N_1 \vee N_2 \vee \ldots \vee N_M$. But this disjunction is ugly, and not suitable for the formulation of psychological law. The generalization can be saved by recasting it as one about role states, but then the only way the initial generalization can be a generalization about the belief that the world is going to end in ten minutes is if that belief is a role state. The general point is that the patterns we capture in psychological generalizations relate to role states, so the cost of holding that psychological states are realizer states is that psychological generalizations cease to relate psychological states.

Whether propositional attitude states are best thought of as role states or realizer states, it has seemed plausible to many philosophers that the realizer states are physical states of the brain - or, at any rate, states which supervene upon physical states of the brain (see Mind, identity theory of; Supervenience of the mental). This is put forward as a plausible scientific hypothesis. Only in the brain are there states of sufficient complexity to play the needed roles. It is, though, allowed as a logical possibility that there could have been other (perhaps even non-physical) realizer states which filled the same role.

Some philosophers, however, have worried about whether neural states could be the kinds of things which can be adverted to, even *inter alia*, in the explanation of rational action. The connections between neural states are merely causal, they argue - playing out the universal laws of physics, or of neuroscience - whereas the connections between states of propositional attitude are rational (and serve to rationalize behaviour). One belief makes another the rational one to have; what one believes and desires may make one action more rational than another; and so on (see Reasons and causes).

However, there are serious problems with this sort of objection. Would the cause of rationality be better served by having the attitudes occurring at random? And surely a decent account of the propositional attitudes places them in the physical world - they are not otherworldly mysteries - and given this, there seems little alternative to placing them in the brain: we know that the propositional attitudes play complex causal roles in the production of our behaviour, and only the brain has states capable of the needed complexity. Finally, it is unclear why one might suppose that functionally defined states cannot have the rationalizing properties which are alleged to vanish from the scientific picture (see Loar 1981 for further discussion).

3 Propositional attitude contents

The contents of propositional attitudes are propositions. That much is simply a matter of definition. But what are propositions? There are at least the following contenders: (1) collections of circumstances of evaluation, such as collections of possible worlds; (2) syntactic entities, such as interpreted sentences in natural language; (3) set-theoretic structures, such as Fregean thoughts (see Frege §§3-4). Only some of the strengths and weaknesses of these different approaches can be examined here (see Propositions, sentences and statements).

(1) One important virtue of the theory that propositions are collections of possible worlds - that is, collections of complete ways that things could be (see Semantics, possible worlds) - is that it directly captures the intuition that propositions involve a sorting among ways that things could be. For example, in believing that snow is white, one believes that the actual complete way that things are is among those complete ways that things could be in which snow is white; in desiring that snow falls at Christmas, one desires that the actual complete way that things are is among those in which snow falls at Christmas; and so on.

Perhaps the most important drawback of this theory is that its most plausible formulations seem to give the wrong results for attitudes with logically equivalent contents. For example, it entails that the belief that there are husbands is identical to the belief that there are wives. Because ‘There are wives’ is logically equivalent to ‘There are husbands’, the set of worlds where there are wives is one and the same as the set of worlds where there are husbands. Likewise, it makes the desire to prove that two and two are four identical to the desire to prove that arithmetic is not decidable, and the intention to draw an equiangular triangle identical to the intention to draw an equilateral triangle. As the objection is commonly put, the possible worlds theory delivers objects of the attitudes that are insufficiently fine grained - there are more distinct attitude contents than it allows.

In addition, there is a complication which needs to be addressed. Consider, for example, my belief that I am cold, or your belief that you are spilling sugar on the floor of the supermarket. These attitudes *de se* (about oneself), as they are called, need to be thought of as involving a sorting among ways that things might be centred on the subject - me, here and now, as it could be, rather than of complete ways things might be (see Demonstratives and indexicals; Content, indexical).

(2) One important virtue of the theory that propositions are syntactic entities - that is entities with sentence-like structure - is that it provides objects to discriminate between attitude contents which, as we have just seen, the possible worlds theory is unable to discriminate. For example, the sentence ‘There are husbands’ is distinct from the sentence ‘There are wives’; and the theory can exploit this fact to distinguish the belief that there are husbands from the belief that there are wives. They are attitudes to different though equivalent sentences.

However, if the sentences are thought of as being in a natural language, this advantage is purchased at a price; it provides distinctions where it is plausible that there are none. Consider, for example, our practice of translating and interpreting those who speak different languages. Clearly, 'Snow is white' and 'La neige est blanche' are distinct sentences - but surely we are not thereby obliged to say that it is impossible for a monolingual French speaker to believe that snow is white? Consider, too, the belief that p, and the belief that p and p. There is at least some temptation to think that these are not distinct beliefs.

One response is to argue that the sentences in question do not belong to natural languages but rather to a language of thought - that is, to a neural system of information or information storage which is supposed to have syntactic structure, and that this is something we all share (and share with animals like dogs that lack public language but do seem to have propositional attitudes) (see Language of thought). One difficulty with this suggestion is that it seems wrong that a mere analysis of propositional attitudes will commit us to the existence of a language of thought, a substantive view about our neural natures. It seems to be an empirical question, for example, whether mental representation is language-like or map-like (see Belief §3), but the language of thought hypothesis commits us in advance to the former alternative.

(3) One important virtue of the theory that propositions are set-theoretic structures - that is entities whose basic structure is described by set theory (see Set theory) - is that it promises to cope with the difficulties mentioned above for the syntactic and possible worlds accounts. However, before we can see why this is so, we need to give some explanation of the theory.

Consider, for example, the sentence 'London is pretty'. A natural thought is that the proposition expressed by this sentence is somehow made up of the city of London (or some way of thinking about it) and the property of being pretty (or some way of thinking about the property of being pretty). More generally, the theory holds that propositions are composed from constituents which are somehow put together in an orderly (set-theoretic) fashion. The constituents in question include: individual objects (and perhaps modes of presentation thereof), which correspond to proper names and other singular terms; properties, which correspond to predicates; functions, which correspond to sentential operators (such as 'and' and 'possibly'), and so on.

This view has the resources to discriminate between the belief that there are husbands and the belief that there are wives, one of the examples that gave the possible worlds approach trouble. Roughly, the first has in its propositional object women and the property of being married to men, whereas the second has men and the property of being married to women. The view also has the resources to identify the propositions expressed by the sentences 'Snow is white' and 'La neige est blanche' and so allow that speakers of different languages can share beliefs, one of the examples that gave the sentential theory trouble. Although the words differ, the objects and properties they stand for are the same and hence the propositions thought of as constructed from them are the same. However, the theory in its typical form has to distinguish between the belief that p, and the belief that p and p - the propositional object of the second is special in involving the operation of conjunction - and this is not clearly desirable.

Another problem with set-theoretic approaches is that they involve various kinds of 'disreputable entities' - intensional properties, modes of presentation ('Fregean senses'), perhaps even set-theoretic functions - which many philosophers would like to exclude from their theories of the attitudes (see Intensional entities). They worry about how these entities can be tied to propositional attitudes, especially if these attitudes are thought of functionally. What causal role can intensional properties or sets, for instance, play? True, this is part of a hard question that arises for any account of the attitudes. We have been discussing competing views about what the objects of the attitudes are. But there is also the question as to how a given propositional attitude state gets attached to the proposition it is attached to; the question of what makes it true, for example, that a certain state is the belief that snow is white rather than that grass is green. We will set this important question aside except to remark that the set theoretic view of the nature of the attitudes is often thought to be especially ill placed to tackle it.

Plainly, there is much more to be said about each of the options discussed here (and about other options as well). The important point is that there are many dimensions to propositional content, and that it is far from easy to provide a theory which scores well on every dimension. A good theory of propositional content should have all of

the following features:

(1) it should be naturalistically respectable - that is compatible with the outlook of natural science;
(2) it should capture the normative dimensions of propositional content;
(3) it should allow a fine-grained discrimination of the contents of propositional attitudes (logical equivalence is not sufficient for identity);
(4) it should not tie the possession of propositional attitudes too tightly to particular languages (for example, it should allow that the same attitudes can be had by creatures which do not share a language - and, indeed, that attitudes can be possessed by creatures which do not speak any language at all);
(5) it should make it clear how states which possess that kind of propositional content can play a role in the production of (rational) action.

4 Propositional attitude ‘aboutness’

Many philosophers have been puzzled by the various ways and senses in which propositional attitudes can be about particular objects and kinds of objects (see Intentionality). Think, for example, of the beliefs which one might have about Aristotle, about Einstein, about President Clinton, about Zeus, about Mickey Mouse, about Meinong’s round square, about gold, about phlogiston, about π, and about the greatest prime number. There are several different issues which can arise here.

One issue concerns the existence of the objects and kinds of objects which are the focus of the attitudes. There is a sense of ‘about’ in which one cannot have attitudes about an object (or kind of object) which does not exist - so that, for example, in this sense, one cannot have beliefs, hopes, fears, desires, and so on, about Zeus, the round square, the greatest prime number, Mickey Mouse and phlogiston, though one can have these attitudes about President Clinton and gold. (The cases of entirely past objects - Aristotle and Einstein - and numbers - π- raise other controversial questions about existence which cannot be taken up here.) This sense of ‘about’ is relational: the point parallels the impossibility of parking near the round square, shaking hands with Zeus, or adding three to the greatest prime number.

Another issue concerns the nature of the connection which obtains between subjects and those existent objects and kinds of objects which are the focus of given attitudes. If one is not appropriately connected to an object, then - in this sense - one cannot have attitudes about it. Suppose, for example, that by an extraordinary coincidence there is a planet somewhere in the universe whose history is exactly described by Tolkien’s *Lord of the Rings*. Even though there is some sense in which Tolkien’s work is true of this planet, nonetheless, the work is not about the planet (in the sense currently at issue). The reason is that the planet played no role at all in the causal history of the book. When, for example, one believes that Hawking is a genius, what makes it the case that one’s belief is about Hawking is, very likely, at least in part, that there are chains of usage of the word ‘Hawking’ which are appropriately connected to that theoretical physicist. When I believe that that tree outside my window is shedding its leaves, what makes it a belief about that tree is, at least in part, the causal processes involving electromagnetic radiation passing between the tree and my retina which are involved in the formation of my belief. In general, it seems plausible to think that some kind of causal connection is required for these kinds of attitudes about particular objects.

However, while it is clearly correct to say that there is a sense in which the aboutness of propositional attitudes is a matter of standing in a relation, an appropriate causal connection on most views, it is also clear that there is another sense of ‘about’ which cannot be analysed in this way. In some sense we can have beliefs and desires about non-existent objects - for example, some people desire to find the fountain of youth, some believe that Meinong’s round square is round, some hoped that Zeus would smite their enemies with a thunderbolt, and so on. It seems clear that those who, for example, hoped that Zeus would smite their enemies have a hope ‘about’ Zeus despite not standing in a relation, causal or otherwise, to Zeus.

How then should this final, non-relational, non-causal, sense of ‘about’ be analysed? Many alternatives present themselves. One might appeal to counterfactuals: attitudes ‘about Zeus’ are attitudes about something which does not exist, but which might have existed; attitudes ‘about phlogiston’ are attitudes about something which does not exist but which might have existed; and so on. One might hold that this sense of ‘about’ is only used in loose speech, and can be banished or paraphrased away in serious philosophical discourse: attitudes ‘about Zeus’ are

really attitudes which focus on certain myths and stories; attitudes ‘about phlogiston’ are really attitudes about a false chemical theory; and so on. One might try various other options as well. The issue is too difficult to discuss adequately here: the important point is that there is at least a loose sense in which there is ‘existentially neutral’ content, that is, content which does not commit one to the existence in the actual world of the particular individuals and kinds to which the content does give apparent commitment.

There is also an important distinction between two different kinds of content: broad content and narrow content. Roughly speaking, narrow content is purely a matter of what goes on inside the head of a subject of propositional attitude states, whereas broad content is in part a matter of how things are in the world outside the head. More exactly, if two subjects are intrinsic duplicates who live under the same laws of nature and belong to the same kinds, then their attitudes are exactly alike in narrow content but may differ in broad content (see Content: wide and narrow).

Once upon a time it was more or less taken for granted that the content of belief, for example, was narrow. How things are outside the head obviously affects whether your beliefs are true or false - your belief that London is pretty can only be true if London itself is pretty - but what you believe is a matter of how you are. But in fact the belief that London is pretty is a belief with broad content. I cannot believe that London, the capital city of England, is pretty unless I am appropriately causally connected to London, and that concerns how things are outside my head. Imagine that the universe is symmetrical, so that there is a region - call it Twin Earth - which is qualitatively identical to our own at all times. It will contain a city called ‘London’ and a twin of me. This twin will be an intrinsic duplicate of me and will produce the sentence ‘London is pretty’ as expressing what he believes. Nevertheless, his belief (and his sentence) will have a different content from mine. It will be about Twin London rather than about London, and will be true just if Twin London, rather than, London is pretty.

5 Defining propositional attitudes

The characterization of propositional attitudes in §2 was very rough, and leaves room for the thought that functionalism involves circularity of definition. We in effect identified the behaviour beliefs cause as behaviour that tends to realize desires, and this bit of inter-defining looks rather like circularity. However, the characterization can be improved in a way which makes it clear that this is not so - and, as an added bonus, also makes clear just what would be involved in the denial of the existence of propositional attitudes. The secret is to take the general account of the definition of theoretical terms provided by David Lewis and apply it to the propositional attitudes.

Roughly, the idea is to collect together a class of sentences involving the verbs of propositional attitude and treat this class as a simultaneous definition of the propositional attitudes. The same idea can be applied to other webs of theoretical terms. A very simple case is the definition of a wife as someone with a husband, and a husband as someone with a wife. We remove the air of circularity when we note that we could spell matters out as follows: *x* is a wife if there are two people, one male and one female, they are married to each other, and *x* is the female one. Consider, for a more complex example, the artworld and its denizens. Artists are producers of art works. Audiences are viewers or consumers of art works. Critics are critical assessors of art works. Galleries are places for the exhibition of art works. And art works are objects produced by artists for consumption by audiences, assessment by critics and display in galleries. Taken together, these sentences - or some more elaborate version thereof - can be taken to provide a simultaneous definition of the key terms: artist, audience, critic, gallery, artwork. And there is no (vicious) circularity involved because the other terms which feature in the various sentences - ‘production’, ‘consumption’, ‘display’, ‘object’, ‘view’, and so on - are (assumed to be) independently understood. These terms in effect specify the key interconnections, and the definition of artist, work of art, and so on that is delivered amounts to saying that an artist is anyone who is appropriately interconnected.

The sentences which are collected together are either *prima facie* analytic truths - that is, conditionals and biconditionals which encode the inferential practices constitutive of possession of the concepts under analysis - or else claims involving the concepts to be analysed which are regarded as *prima facie* common knowledge by those, or most of those, who possess the concepts. Under the former account, there is no guarantee that the sentences will be immediately available to those who possess the concepts; however, it might be hoped that reasonable and reflective people can be brought to assent to them when the sentences are drawn to their attention. Under the latter account, there is no requirement that the sentences encode inferential practices constitutive of possession of the

concepts under analysis: for all kinds of synthetic and a posteriori claims may be selected. However, the following constraint is to be observed: it should be commonly held to be a condition on attribution of possession of the concepts which are up for analysis to a subject that they do not reject, or fail to accept, too many of the chosen sentences.

On either account, one might doubt that there are sentences of the kind required by the analysis. So, for example, it might be said that it is notoriously difficult to provide exceptionless generalizations about the connections between beliefs, desires, intentions, actions, and so on. However, it is important to note that there is nothing in the account which requires that the platitudes in question must be couched as exceptionless generalisations. Indeed, a plausible thought is that most of the platitudes will instead make claims about what is normally the case - that is about what happens when all other things are equal, or when conditions are normal. So, for example, on either account, there are sentences like these: a system of beliefs and desires tends to cause behaviour that serves the subject’s desires according to that subject’s beliefs; beliefs typically change under the impact of sensory evidence; desires aim to have the world fit them; beliefs aim to fit the world, and so on.

The final point to note is that our definition tells us that the propositional attitudes are whatever it is that satisfies the chosen set of sentences and so stand in the specified relations. In the simplest case, there will be a unique natural collection of candidates for the propositional attitudes. In that case, the sentences provide an explicit definition of the propositional attitudes. But things may not be so simple: perhaps there is nothing which satisfies the sentences exactly; and perhaps the only thing which satisfies the sentences is extremely unnatural, that is, gerrymandered. In these cases, we shall look for the best near-satisfier, where the criteria which guide our search are (1) considerations of naturalness; and (2) satisfaction of as many as possible of the more central platitudes. Put another way: application of a procedure which aims at achieving reflective equilibrium may lead to the rejection of some of the sentences, and perhaps to the elevation of further sentences to membership of the favoured class. Moreover, the procedure which aims at achieving reflective equilibrium is guided by judgments about (1) the naturalness of the resulting concept; and (2) the centrality or revisability of the sentences which are incorporated or rejected.

Of course, there is nothing in the account of this procedure which guarantees that there will be a point of reflective equilibrium: and, in the case in which there is no point of reflective equilibrium, we will need to give an error theory of the propositional attitudes - that is we will hold either that there were no coherent concepts characterized by our pre-reflective usage, or else that there simply is nothing to which our concepts apply.

6 Scepticism about propositional attitudes

There are various sceptical doubts which philosophers have had about propositional attitudes. Only some of these doubts will be mentioned here.

Some philosophers - W.V.O. Quine, Paul Churchland, Stephen Stich, for example - deny that there are any propositional attitudes - they deny that there are any beliefs, desires, intentions and the like. Sometimes, the source for this denial is philosophy of language (dislike of intensional contexts, and so on); other times, the source is metaphysics (failure to find a physical structure with which the states in question can be identified, for example). Sometimes, the view is that propositional attitudes are convenient fictions - useful instruments for making predictions of behaviour, but not suited to the business of serious science. On this view, talk of propositional attitudes has a second grade, merely instrumental, status, but there is no reason why we should not continue with it for purposes of prediction and explanation - a common comparison is with talk about the average family. Other philosophers argue that talk of propositional attitudes is intended as first grade scientific theorizing, but that it just happens to be seriously astray. On this view, talk of propositional attitudes is like talk of phlogiston: talk which sought to capture an important feature of reality and to be part of serious science, but failed (see Eliminativism).

Among philosophers who accept that there are propositional attitudes - that is, among those philosophers who accept that some propositional attitude ascriptions are strictly and literally true - there are philosophers who are sceptical of the relational analysis of those attitudes. Thus, some philosophers suppose that propositional attitudes should be given an adverbial analysis according to which, for example, one’s belief that p is a matter of one’s believing p-ly (see Adverbial theory of mental states). They deny that the attitudes are relations to propositions. Sometimes, this denial is fuelled by worries about the candidate entities for those contents, the sets of worlds,

structured set-theoretic entities, and the like we discussed earlier. Theorists who deny the relational view of the attitudes typically insist that a sentence like ‘I believe that snow is white’ is best written, from a logical point of view, as ‘I snow-is-white believe’, somewhat as ‘I have a limp’ is best written ‘I limp’.

Yet other philosophers - Stephen Schiffer, for example - despair of the project of giving any philosophical account (theory, analysis) of propositional attitudes, while nonetheless supposing that our practice of making propositional attitude ascriptions is perfectly in order as it stands. The essence of their position is a denial of the need to give an account in anything like the traditional sense of the propositional attitudes.

Theorizing about propositional attitudes is a very difficult task. This brief summary only indicates a few of the controversial questions which such theorizing is bound to confront.

See also: Action; Folk psychology; Intention; Desire; Propositional attitude statements; Communication and intention

GRAHAM OPPY

References and further reading

Braddon-Mitchell, D. and Jackson, F. (1996) *The Philosophy of Mind and Cognition*, Oxford: Blackwell. (Excellent, state-of-the-art textbook for philosophy of mind.)

Churchland, P. (1981) ‘Eliminative Materialism and the Propositional Attitudes’, *Journal of Philosophy* 78: 67-90.(Well-known defence of a style of eliminativism about the propositional attitudes that sees propositional attitude talk as failed science.)

Davidson, D. (1968) ‘On Saying That’, *Synthèse* 19: 130-46.(Treatment of the contents of propositional attitude ascriptions in terms of sentences in English, but not as sentences of English.)

Fodor, J. (1975) *The Language Of Thought*, New York: Crowell(Defence of the hypothesis that there is a language of thought.)

Frege, G. (1917) ‘Thoughts’ in P. Geach (ed.) *Logical Investigations*, Oxford: Blackwell, 1977.(Famous defence of the view that the contents of propositional attitudes are abstract structured entities.)

Grice, P. (1957) ‘Meaning’, *Philosophical Review* 66: 377-88.(A classic. Includes famous analysis of meaning in terms of propositional attitudes.)

Harman, G. (1972) *Thought*, Princeton, NJ: Princeton University Press.(Early, clear, exposition of a functionalist account of the mind and, in particular, the propositional attitudes.)

Kripke, S. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell.(Classic text which deals, *inter alia*, with problems about the normativity of propositional attitude contents.)

Lewis, D. (1970) ‘How To Define Theoretical Terms’, *Journal of Philosophy* 67: 427-46.(Masterly treatment of the definition of theoretical terms. Much better - but much harder - than the outline in §5!)

Lewis, D. (1994) ‘Reduction of Mind’, in S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*, Oxford: Blackwell, 412-31.(Superb exposition of Lewis’s philosophy of mind. Assumes some familiarity with the prior literature.)

Loar, B. (1981) *Mind and Meaning*, Cambridge: Cambridge University Press.(Difficult, but very good, discussion of propositional attitudes, and their importance for philosophy of mind and philosophy of meaning.)

Pessin, A. (ed.) (1995) *The Twin Earth Chronicles*, London: Pergamon Press.(Collection of articles about Twin Earth, broad and narrow content, and so on. Good reading list.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press and John Wiley.(Famous and elegant text which includes defence of a kind of eliminativism about propositional attitudes.)

Salmon, N. (1986) *Frege’s Puzzle*, Cambridge, MA: MIT Press.(Defence of treatment of propositions as structured, set-theoretic entities.)

Salmon, N. and Soames, S. (eds) (1988) *Propositions and Attitudes*, Oxford: Oxford University Press.(Excellent collection of papers about the semantics of propositional attitude ascriptions. Good reading list.)

Schiffer, S. (1972) *Meaning*, Oxford: Clarendon Press.(Defence of the view that propositional attitudes - in particular, belief and intention - are the key to meaning. Not easy going.)

(1987) *Remnants of Meaning*, Cambridge, MA: MIT Press.(Retraction of the view that the propositional attitudes - in particular, belief and intention - are the key to meaning, and of much else besides. Still not easy going.)

Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press.(Defence of the view that the contents of propositional attitudes are sets of possible worlds. Includes best available treatment of the objection that any such treatment makes the objects of the attitudes too coarse-grained.)

Stich, S. (1983) *From Folk Psychology to Cognitive Science: The Case Against Belief*, Cambridge, MA: MIT Press.(Case for eliminativism about belief, and other propositional attitudes. Interesting, and reasonably accessible.)

Propositions, sentences and statements

A sentence is a string of words formed according to the syntactic rules of a language. But a sentence has semantic as well as syntactic properties: the words and the whole sentence have meaning. Philosophers have tended to focus on the semantic properties of indicative sentences, in particular on their being true or false. They have called the meanings of such sentences 'propositions', and have tied the notion of proposition to the truth-conditions of the associated sentence.

The term 'proposition' is sometimes assimilated to the sentence itself; sometimes to the linguistic meaning of a sentence; sometimes to 'what is said'; sometimes to the contents of beliefs or other 'propositional' attitudes. But however propositions are defined, they must have two features: the capacity to be true or false; and compositional structure (being composed of elements which determine their semantic properties).

One reason for distinguishing a sentence from 'what the sentence says' is that a sentence may be meaningless, and hence say nothing, yet still be a sentence. But perhaps the main reason is that two people, *A* and *B*, may utter the same sentence, for example, 'I am hot', and say the same thing in one sense but not in another. The sense in which they say the same thing is that they use the same words with the same linguistic 'meaning'. The sense in which they say something different is that they put the same words to different 'uses': *A* uses 'I' to refer to *A*, while *B* uses it to refer to *B*. Hence what *A* says may be true although what *B* says is false. If what is said can be true in one case and false in another, *A* and *B* have not made the same 'statement'. On the other hand, if *B* utters instead 'You are hot' when *A* utters 'I am hot', they put different sentences to the same use and make the same statement. On this view (see Strawson 1952), we must distinguish the 'sentence', the 'use' of the sentence, and the 'statement' made by using a sentence in a context of utterance. According to Strawson, it is not sentences but statements that are true or false.

Logicians usually abstract from the context of utterance of sentences in actual communication and talk of the propositions expressed as abstract entities. The main modern proponent of this conception is Frege. For Frege (1918), a proposition is a 'thought', which is both the cognitive meaning expressed by a sentence and the content of a propositional attitude such as belief or desire; thoughts are the 'senses' of sentences. Thoughts are distinguished according to the following principle: if it is possible rationally to believe that *p* and not to believe that *q* then the thought that *p* and the thought that *q* are distinct (see Frege, G. §§3-4; Sense and reference §2).

Under a different, but related, conception, proposed by Carnap (1947), the proposition expressed by a sentence *S* is the set of possible worlds in which *S* is true (see Semantics, possible worlds §9-10). This view violates the compositional structure requirement, since such propositions are not grasped by grasping their components (see Compositionality). It also seems unable to differentiate between some distinct propositions. For example, there would only be one necessarily false proposition, since there is only one empty set of worlds, but intuitively there are many different necessarily false propositions.

An alternative view of propositions as entities is the (neo-)Russellian view that they are collections of actual entities making up 'facts' or 'states of affairs' (see Facts §1). Thus the proposition expressed by 'Socrates is mortal' is the ordered pair $\langle \mathrm{Socrates,\ being\ mortal} \rangle x$, composed of the individual Socrates and the property of being mortal. Three questions for this view are: (1) Are the contents of false sentences negative facts, and are there such facts? (2) What is the criterion of identity for facts? (3) How can propositions as facts be the contents of propositional attitudes? The last of these can be seen as a version of the problem which led Frege to postulate senses: for the fact that $a = b$ must be the same fact as $a = a$, although someone who believes that $a = a$ does not thereby believe that $a = b$, and hence the propositions are not the same (see Sense and reference §1). Fregeans conclude that this shows that we cannot dispense with the notion of sense. If Russellians deny this, they must complicate their account of facts or their account of propositional attitudes (see Propositional attitude statements).

Whether one defines the intuitive notion of 'what is said' as a context-independent entity or by recourse to the notion of a statement or proposition expressed by an utterance in a particular context, one must say what it means for two sentences to express 'the same proposition' or make 'the same statement'. Both phrases rely on the notions of meaning and synonymy, criticized by Quine (1968; see Quine, W.V. §8). But if we dispense altogether with

propositions, statements or any other notion of the content of what is said, and choose instead to take sentences as truth bearers, we face two problems. The first is that sentences are unlikely candidates for the role of contents of propositional attitudes: if ‘I believe I am stupid’ introduces a relation between me and the English sentence ‘I am stupid’, it should be translated into German (say) as ‘Ich glaube "I am stupid"’, although the correct translation is ‘Ich glaube daß ich dumm bin’, which says nothing about any English sentence (Church 1950). The second problem is that, in spite of our ontological scruples about admitting propositions as entities, we still need such a notion to express the content of what is said or believed, and to account for the intentionality of thought in general. In that respect, however vague or ill-defined the notion of a proposition, it cannot be dispensed with.

See also: Intensional entities; Logical and mathematical terms, glossary of

PASCAL ENGEL

References and further reading

Blackburn, S. (1975) ‘The Identity of Propositions’, in S. Blackburn (ed.) *Meaning, Reference and Necessity*, Cambridge: Cambridge University Press.(A good statement of a deflationary alternative to the ontological conception of the problem.)

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press, 2nd edn, 1956.(The first modern formulation of the possible worlds account.)

Cartwright, R. (1962) ‘Propositions’, in R.J. Butler (ed.) *Analytical Philosophy, First Series*, Oxford: Blackwell, 81-103.(A classical discussion of the main issues.)

Church, A. (1950) ‘On Carnap’s Analysis of Statements of Assertion and Belief’, *Analysis* 10: 97-9.(States his translation argument against sentences as objects of belief.)

Frege, G. (1918) ‘Der Gedanke: eine logische Untersuchung’, *Beiträge zur Philosophie des deutschen Idealismus* 1: 58-77; trans. P.T. Geach and R. Stoothoff, ‘Thoughts’, in *Collected Papers*, ed. B.F. McGuiness, Oxford: Blackwell, 1984, 351-72.(The classic statement of propositions as ‘thoughts’ or abstract entities.)

Geach, P.T. (1965) ‘Assertion’, *Philosophical Review* 74(4); repr. in *Logic Matters*, Oxford: Blackwell, 1968.(A criticism of the notion of statement.)

Nuchelmans, G. (1973) *Theories of the Proposition*, Amsterdam: North Holland.(The best historical account, although somewhat doxographic.)

Quine, W.V. (1968) ‘Propositional Objects’, *Critica* 5: 3-22; repr. in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 1969.(Includes a summary of Quine’s misgivings about propositions as entities.)

Russell, B.A.W. (1918) ‘The Philosophy of Logical Atomism’, in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992.(Treats propositions as states of affairs in the world.)

Salmon, N. and Soames, S. (eds) (1988) *Propositions and Attitudes*, Oxford: Oxford University Press.(A good anthology on the semantics of propositional attitudes, focused on the Russellian view.)

Stalnaker, R. (1976) ‘Propositions’, in A.F. MacKay and D.D. Merrill (eds) *Issues in the Philosophy of Language*, New Haven, CT: Yale University Press, 79-92.(A good account of the possible worlds view.)

Strawson, P.F. (1952) ‘On Referring’, *Mind* 59: 320-44; repr. in *Logico-Linguistic Papers*, London: Methuen, 1971.(Puts the classical arguments in favour of the notion of statement.)

Protagoras (*c*.490-*c*.420 BC)

*Protagoras was the first and most eminent of the Greek Sophists. Active in Athens, he pioneered the role of professional educator, training ambitious young men for a public career and popularizing the new rationalist worldview that was introduced from Ionian natural philosophy. But unlike his contemporary Anaxagoras, Protagoras was sceptical of the dogmatic claims of the new science. His famous formula - 'Man is the measure of all things, of things that are, that they are, and of things that are not, that they are not' (fr. 1) - makes him the father of relativism and even, on some interpretations, of subjectivism. He was also considered the first theological agnostic: 'Concerning the gods, I am unable to know either that they exist or that they do not exist or what form they have' (fr.4). He was sometimes associated with the claim 'to make the weaker argument (*logos*) the stronger'.*

1 Career

Protagoras came from Abdera, the city of the atomist Democritus. As a professional educator, he visited Athens several times and was closely connected with the intellectual circle around the leading statesman Pericles. He is reported to have drawn up the laws for Thurii, a pan-Hellenic colony in southern Italy, founded on the site of Sybaris in 444 BC under Periclean leadership. His extraordinary prestige, as the wisest person and greatest teacher of his time, is brilliantly portrayed in the introductory scenes of Plato's dialogue *Protagoras* (see Plato §9). Protagoras there declares that he is the first Greek openly to offer moral and political training on a professional basis.

Protagoras' agnosticism regarding the existence of the gods must have shocked many of his contemporaries. A widespread ancient tradition reports that he escaped from Athens after being condemned for impiety, and that he was subsequently drowned at sea. One version of the story claims that his books were publicly burned at Athens. Some scholars have doubted this report, on the grounds that in Plato's *Meno* (91e) Socrates says that Protagoras remained in good repute until the end of his life. On the whole, however, the ancient tradition is likely to be correct. If so, Protagoras was (like Anaxagoras) one of the first victims of the popular reaction against the rationalist Enlightenment, whose most famous victim was Socrates (§1).

2 Teaching

In Plato's *Protagoras*, Protagoras describes the content of his teaching as 'good judgment in administering the affairs of one's household and those of the city, so that one may be most capable of acting and speaking on public matters' (319a). He there declares that if any pupil is not satisfied with the requested fee (which was certainly high), he need only go to a temple to swear how much he thinks the training was worth, and pay no more. The primary appeal of his teaching was to ambitious young men from wealthy families, who hoped to gain an advantage for their political career from studying with him. In historical perspective, the Sophistic movement initiated by Protagoras can be seen as the beginning of higher education in Western culture. The Sophists were the first professors without a permanent school (see Sophists).

Unlike some later Sophists who included the new natural philosophy in their curriculum, Protagoras seems to have focused his own training on practical concerns, and above all on eloquence and argument. He is credited with being the first to show that there can be an argument for and against every thesis (fr. 6a). Thus he is reported to have trained his students, like a modern debater, to argue both sides of every disputed question, and also to be able to praise and blame the same person (A20-1). We have in the *Dissoi logoi (Twofold Arguments)* an ancient collection of sample reasoning on opposite sides of issues concerning good and bad, just and unjust, whether virtue can be taught, and the like (see Dissoi Logoi). The extant text of *Dissoi logoi* is estimated to postdate Protagoras, but manuals of this type probably represent a tradition of *antilogia* or contradictory argumentation going back to Protagoras. It was presumably because Protagoras was so celebrated as a master of argument and refutation that Plato has selected him as Socrates' opponent and victim in the duel of wits in the dialogue *Protagoras*.

Protagoras' teaching is associated with the claim 'to make the weaker argument the stronger'. This is easily given a hostile interpretation: to make a bad case seem good, and thus to deceive a gullible audience. Enemies of the Sophists like Aristophanes certainly took this view of the kind of training that Protagoras inaugurated, and they regarded the moral influence of the Sophists as nefarious. Protagoras himself, however, seems to have been a

staunch defender of traditional morality. This is indicated not only by Plato's portrayal but also by his political role as legislator for Thurii. The contrast between *nomos* and *physis*, between conventional morality and the nature of things, was later used to undermine the authority of the moral tradition (see Antiphon; Callicles; Physis and nomos). If this contrast was used by Protagoras (as some scholars have supposed), it can only have functioned in favour of *nomos* and the customary moral tradition. For, as will be seen in §3, Protagoras denies the claims to absolute or objective knowledge associated with the notion of *physis*. For a relativist like Protagoras there is no knowledge of the nature of things except from a particular human perspective.

Protagoras was also a pioneer in the study of language and grammar. Two important grammatical observations are attributed to him: (1) the distinction of gender into masculine, feminine and neuter (A27), and (2) the distinction of discourse into prayer (or optative), question (interrogative), answer (indicative), and command (imperative) (A1, paragraph 53, where a more complex division is also ascribed to Protagoras). These grammatical distinctions were apparently developed in the context of critical comments on Homeric poetry in the name of *orthoepeia* or correctness of diction. Thus Protagoras is reported to have disapproved of the opening verses of the *Iliad* for using an imperative form ('Sing, muse, of the wrath of Achilles') rather than the optative, which would be more appropriate for a prayer (A29). And he seems also to have complained that Homer's word 'wrath' (*menis*) should be masculine rather than feminine (A28). Plato twice alludes to Protagoras as an expert on the correctness of names (*Cratylus* 391c) or the correctness of diction (*Phaedrus* 267c) and he presents Protagoras as claiming that skill in criticizing the poets constitutes a major part of education (*Protagoras* 338e). The complex concern with language that is characteristic of Sophist practice thus clearly begins with Protagoras, but our evidence is too meagre for an accurate appreciation of his personal contribution.

3 Philosophical views

Protagoras is famous for two quotations. The best known is from a book entitled *Truth*: 'Man is the measure of all things, of things that are, that they are, and of things that are not, that they are not' (fr.1). The other may be from a different work: 'Concerning the gods, I am unable to know either that they exist or that they do not exist or what form they have. For there are many obstacles to knowledge: the obscurity of the matter and the brevity of human life' (fr.4). There is at least a notional connection between these two doctrines, since the claims to dogmatic knowledge on the part of early poets and philosophers often invoke divine inspiration. Thus Parmenides, the most influential philosopher of the fifth century, presents his view as the revelation of a goddess. In rejecting any knowledge of the gods, Protagoras also rejects any claim to an absolute or god's-eye knowledge of the truth.

Protagoras' 'man-the-measure' formula was interpreted in two different ways: (1) the subjectivist interpretation takes the thesis to claim that whatever anyone believes is true; and (2) the relativist interpretation claims only that whatever anyone believes is true for the person who believes it. The subjectivist doctrine (1) is rather easy to refute. For if all beliefs are true, then the belief that the man-the-measure doctrine is false must also be true. So if anyone doubts the truth of the man-the-measure thesis, the thesis itself must be false. This is the argument known as the *peritropē*, the claim that the thesis overturns itself. Plato in the *Theaetetus* interprets Protagoras' doctrine as the relativist thesis (2): that what anyone believes is true *for that person* (see Relativism). Plato presents the *peritropē* argument as a refutation of this version also.

We do not know whether or not Protagoras applied his theory of truth to moral and political issues. Plato suggests that for Protagoras: 'whatever the city establishes as just, is just for that city as long as it judges so' (*Theaetetus* 177d). This form of cultural relativism, which combines the recognition of different moral codes with respect for established tradition, is likely to be historically correct for Protagoras, since we find a similar regard for *nomos* or custom, in the face of cultural diversity, defended by Protagoras' contemporary, the historian Herodotus. It would also be a convenient doctrine for an itinerant philosopher, who can thus endorse the established order in each society when travelling from city to city.

Other positions attributed to Protagoras may be connected with his relativist theory of truth. Thus it may have been a general reaction against dogmatic claims that led him to deny that the circle touches a straight line only at a single point, as is assumed in plane geometry (fr. 7). The paradoxical view that 'it is not possible to contradict someone', which was later defended by Antisthenes (§4), was originally ascribed to Protagoras. If all opinions are true, then two speakers who seem to disagree cannot really be contradicting one another: they must be talking about different things.

CHARLES H. KAHN

References and further reading

Burnyeat, M.F. (1976) ‘Protagoras and Self-Refutation in Later Greek Philosophy’ and ‘Protagoras and Self-Refutation in Plato’s *Theaetetus*’, *Philosophical Review* 85: 44-69, 172-95.(A careful analysis of the man-the-measure thesis according to the subjectivist and relativist interpretations.)

Classen, C.J. (ed.) (1976) *Sophistik*, Darmstadt: Wissenschaftliche Buchgesellschaft.(An anthology which includes seven scholarly articles on Protagoras, three of them in English.)

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, vol. 3, Cambridge: Cambridge University Press; part of vol. 3 repr. as *The Sophists*, Cambridge: Cambridge University Press, 1971.(Full scholarly account.)

Kerferd, G.B. (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press.(A briefer, more personal interpretation than Guthrie (1969).)

Plato (*c*.386-380 BC) *Protagoras*, trans. C.C.W. Taylor, Oxford: Oxford University Press, 2nd edn, 1991.(Dialogue showing Protagoras in debate with Socrates. A gathering of famous Sophists, where Socrates defeats Protagoras in a series of dialectical debates.)

Plato (*c*.370-367 BC) *Theaetetus*, trans. M.J. Levett, intro. M. Burnyeat, Indianapolis, IN: Hackett, 1990.(Contains a classic critique of Protagoras’ relativism. Elaborate examination and refutation of Protagoras’ thesis that man is the measure of what is.)

Plato (*c*.380-367 BC) *Cratylus*, trans. C.D.C. Reeve, in J. Cooper (ed.) *Plato, Complete Works*, Indianapolis, IN: Hackett Publishing Company, 1997.(Elaborate discussion on the correctness of names, whether conventional or natural; Protagoras is mentioned as having lectured on the topic.)

Plato (*c*.366-360 BC) *Phaedrus*, trans. A. Nehamas and P. Woodruff, in J. Cooper (ed.) *Plato, Complete Works*, Indianapolis, IN: Hackett, 1997.(In a long list of specialists on rhetoric, Protagoras is mentioned as having introduced refinements concerning correctness of diction.)

Protagoras (*c*.490-*c*.420 BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidmann, 7th edn, 1954, vol. 2, 253-71.(The standard collection of the ancient sources both fragments and testimonia, the latter designated with ’A’; includes Greek texts of the fragments with translations in German.)

Schiappa, E. (1991) *Protagoras and Logos, A Study in Greek Philosophy and Rhetoric*, Columbia, SC: University of South Carolina Press.(A rather speculative attempt at a systematic interpretation of Protagoras’ thought.)

Sprague, R.K. (ed.) (1972) *The Older Sophists*, Columbia, SC: University of South Carolina Press, 3-28.(Full English translation of the fragments and testimonia from Diels and Kranz (1954).)

Proudhon, Pierre-Joseph (1809-65)

Pierre-Joseph Proudhon was a French social theorist, political activist and journalist. Claiming to be the first person to adopt the label 'anarchist', he developed a vision of a cooperative society conducting its affairs by just exchanges and without political authority. In his lifetime he exercised considerable influence over both militants and theorists of the European left, and he is remembered today as one of the greatest exponents of libertarian socialism. His last writings, though still strongly libertarian, advocated a federal state with minimal functions.

Proudhon was born in Besançon, in the French Jura, to a poor artisan family. He worked first as a printer, and was largely self-educated. At the age of thirty he won a fellowship from the local Academy and went to Paris, where he spent most of the rest of his life. During the 1840s his writings included *Qu'est-ce que la propriété? (What is Property?)* (1840), and *Système des contradictions économiques ou Philosophie de la Misère (System of Economic Contradictions, or the Philosophy of Poverty)* (1846), which laid the groundwork for his social and economic theory: the latter drew an acerbic reply (*The Poverty of Philosophy*) from Marx, who regarded Proudhon's views as those of a petit-bourgeois. By the time of the 1848 revolution he was prominent enough to win election to the Constituent Assembly, where, however, he was too isolated a figure to be effectual. After the advent of the Second Empire in 1851, Proudhon suffered periods both of imprisonment in France and exile in Belgium. During this period he wrote some of his most important books, notably his three-volume *De la justice dans la Révolution et dans l'Église (On Justice in the Revolution and in the Church)* (1858). Five years after Proudhon's death, many leaders of the Paris Commune, the most important socialist insurrection ever to take place in Western Europe, regarded themselves as his disciples. Russian anarchists, such as Tolstoi and Kropotkin, also acknowledged his influence.

Proudhon wrote in a muscular and impetuous style, often producing startling aphorisms, which on first glance may seem to be paradoxes. The two for which he is best known are 'property is theft' and 'God is evil'. Neither of these, however, can be taken at face value. With regard to the first aphorism, Proudhon believed that possession of a share in productive resources was immensely important: it was essential to the development of the (male) individual's inventive and productive powers. What he rejected as 'theft' was the existence of a property right which confirmed some people in their possession of resources, when others possessed none. Proudhon favoured a society in which possession of productive resources would be widely shared (in contrast to the collective ownership favoured by socialists such as Marx (see Marx, K. §12)). Producers would exchange their goods, not at a market price determined by the interaction between supply and demand, but at a price fixed by the average labour time required by the goods' production. Efficient producers would thus be rewarded.

The second famous aphorism, 'God is evil', also requires a context. Proudhon equates 'God' with ideas of fatality which exclude or downplay the idea of human agency. When he says, in *System of Economic Contradictions*, that 'the first duty of the free and intelligent man is to hunt the idea of God from his mind', he means that, to be free, people must abandon the idea that human circumstances are something other than a human construct. Part of unfreedom is the belief that the evolution of human society is the work of fate or providence and thus beyond the control of humans: this belief will vanish when there are no longer massive concentrations of economic and political power which remove decisions from the hands of ordinary people. Another part of unfreedom is the belief that one must obey authorities endowed with some special status which places them beyond the control of their subjects, an irrational belief which would not survive without religious support.

Independence - material and intellectual - was thus an important value for Proudhon. But he also stressed the interdependence arising from the complementarities in social life. The complementarity of different skills creates an incentive to reach agreement, as well as multiplying the power of human labour. The non-synchronized character of humans' needs makes possible the institution of credit: Proudhon's writings inspired some important practical reforms in the credit system. Interestingly, within the family unit, the principle of interdependence triumphed wholly over that of independence: Proudhon held rigid views about the sexual division of labour, and would have allowed women no role outside the household. Traditionalist aspects of Proudhon's thought have led some right-wing theorists to claim him for their own.

Proudhon's thinking reveals many influences: there are echoes of Fourier, Saint-Simon, Comte, British political

economy, Hegel and Feuerbach. Rousseau (see Rousseau, J.-J. §3), however, has a special importance for him, both positive and negative. What Proudhon shared with Rousseau was the idea that absolute personal independence could be reconciled with the social condition - that one can belong to a social order yet 'obey only oneself' (perhaps in the limited sense of obeying one's own reason). What he hated in Rousseau was the conclusion that this reconciliation could be achieved only in the state: it could best be achieved, Proudhon thought, in civil society, and in the reciprocity among groups of producers, characterized by mutual respect and complementarity of interest. Proudhon was a critic of political democracy, which he thought Rousseau had inspired: although he believed in the capacity of small groups to manage concrete affairs which they could understand, he had no faith in the political discernment of the masses. When he came eventually to accept the need for government, he did so only on the condition that the primary units of government would be small communities constrained by immediate responsibility for their decisions: the radically decentralizing spirit of his earlier work is not abandoned.

See also: Anarchism §3; Freedom and liberty; Property; Socialism

RICHARD VERNON

List of works

Proudhon, P.J. (1840) *Qu'est-ce que la propriété? ou Recherche sur le principe du droit et du gouvernement*, Paris: Brocard; trans. B. Tucker, *What is Property?*, New York: Dover, 1970.(The most accessible statement of his basic position.)

Proudhon, P.J. (1843) *De la création de l'ordre dans l'humanité, ou Principes d'organisation politique (On the Creation of Human Order, or Principle of Political Organization)*, Paris: Prevot.(A systematic exposition of his views on religion, philosophy and political economy.)

Proudhon, P.J. (1846) *Système des contradictions économiques ou Philosophie de la misère*, Paris: Guillaumin; trans. B. Tucker, *System of Economic Contradictions, or the Philosophy of Poverty*, New York: Arno, 1972. (The fullest statement of his economic theory.)

Proudhon, P.J. (1851) *Idée générale de la Révolution au XIXe siècle*, Paris: Garnier; trans. J.B. Robinson, *General Idea of the Revolution in the Nineteenth Century*, London: Freedom Press, 1923.(Presents anarchism as the culmination of democratic principles.)

Proudhon, P.J. (1853) *Philosophie du progrès (Philosophy of progress)*, Brussels: Lebègue. (Depicts history as the progress of freedom.)

Proudhon, P.J. (1858) *De la justice dans la Révolution et dans l'Église (On justice in the Revolution and in the Church)*, Paris: Garnier, 3 vols.(The fullest statement of his ethical theory.)

Proudhon, P.J. (1861) *La guerre et la paix (War and Peace)*, Paris: Dentu.(Calls for the replacement of war by industry as the basis of social organization.)

Proudhon, P.J. (1863) *Du principe fédératif et de la nécessité de reconstituer le parti de la Révolution*, Paris: Dentu; trans. R. Vernon, *The Principle of Federation*, Toronto, Ont.: University of Toronto Press, 1979. (Explains his acceptance of minimal government.)

Proudhon, P.J. (1865) *De la capacité politique des classes ouvrières (On the Political Capacity of the Working Classes)*, Paris: Dentu.(Offers a programme for workers' self-emancipation.)

References and further reading

Hoffman, R.L. (1972) *Revolutionary Justice: The Social and Political Thought of P.-J. Proudhon*, Urbana, IL: University of Illinois Press.(Detailed exposition and commentary. Excellent bibliography.)

Marx, K. (1847) 'The Poverty of Philosophy', in *Selected Writings*, ed. D. McLellan, Oxford: Oxford University Press, 1977.(A polemic against Proudhon's economic analysis.)

Noland, A. (1967) 'Proudhon and Rousseau', *Journal of the History of Ideas* 28: 33-54.(Examines Proudhon's most important intellectual debt.)

Ritter, A. (1969) *The Political Thought of Pierre-Joseph Proudhon*, Princeton, NJ: Princeton University Press.(A careful, critical, yet sympathetic analysis of Proudhon's central concepts.)

Waldron, J. (1988) *The Right to Private Property*, Oxford: Clarendon Press.(Contains an account of Proudhon's theory of property.)

Woodcock, G. (1956) *Pierre-Joseph Proudhon: His Life and Work*, New York: Schocken.(The standard English-language biography. Contains a bibliography.)

Provability logic

Central to Gödel's second incompleteness theorem is his discovery that, in a sense, a formal system can talk about itself. Provability logic is a branch of modal logic specifically directed at exploring this phenomenon. Consider a sufficiently rich formal theory T. By Gödel's methods we can construct a predicate in the language of T representing the predicate 'is formally provable in T'. It turns out that T is able to prove statements of the form

(1) If A is provable in T, then it is provable in T that A is provable in T.

In modal logic, predicates such as 'it is unavoidable that' or 'I know that' are considered as modal operators, that is, as non-truth-functional propositional connectives. In provability logic, 'is provable in T' is similarly treated. We write $\Box A$ *for 'A is provable in T'. This enables us to rephrase (1) as follows:*

(1′) $\Box A \rightarrow \Box\Box A$.

This is a well-known modal principle amenable to study by the methods of modal logic.

Provability logic produces manageable systems of modal logic precisely describing all modal principles for $\Box A$ *that T itself can prove. The language of the modal system will be different from the language of the system T under study. Thus the provability logic of T (that is, the insights T has about its own provability predicate as far as visible in the modal language) is decidable and can be studied by finitistic methods. T, in contrast, is highly undecidable. The advantages of provability logic are: (1) it yields a very perspicuous representation of certain arguments in a formal theory T about provability in T; (2) it gives us a great deal of control of the principles for provability in so far as these can be formulated in the modal language at all; (3) it gives us a direct way to compare notions such as knowledge with the notion of formal provability; and (4) it is a fully worked-out syntactic approach to necessity in the sense of Quine.*

1 The prehistory of the subject

The history of provability logic proper starts with Kurt Gödel (1933), who, at the end of a discussion of the interpretation of intuitionistic propositional logic, briefly (in six lines!) discusses formal provability as a modal operator.

The next step was taken by Leon Henkin who, in 1952, asked a seemingly frivolous question:

> If Σ is any standard formal system adequate for recursive number theory, a formula (having a certain integer *q* as its Gödel number) can be constructed which expresses the proposition that the formula with Gödel number *q* is provable in Σ. Is this formula provable or independent in Σ?
>
> (Henkin 1952: 180)

In short, is a sentence that 'says' 'I am provable in Σ' provable in Σ? Note that $\underline{0} = \underline{0}$ is a fixed point of $\mathrm{Prov}_T(x)$: $\underline{0} = \underline{0}$ is provable in *T* and so is $\mathrm{Prov}_T(\underline{\#(\underline{0} = \underline{0})})$; hence, by classical logic, $(\underline{0} = \underline{0} \leftrightarrow \mathrm{Prov}_T(\underline{\#(\underline{0} = \underline{0})}))$ is provable in *T*. (See §2 below for an explanation of the notation.) $\underline{0} = \underline{0}$ is, however, not a *literal* Henkin sentence: it is not built explicitly to say of itself 'I am provable' in the way that Gödel's sentence is built to say 'I am not provable'. In 1953, G. Kreisel published a noteworthy paper answering Henkin's question with a resounding 'It depends… '. Kreisel constructed two extensionally correct provability predicates where for the one the literal Henkin sentence ('I am provable in Σ') is provably true, but for the other it is provably false.

It might have ended there. Kreisel's result might have discouraged everyone from thinking further about Henkin's 'frivolous' question. Happily for the subject, a second paper, by M.H. Löb, appeared in 1955. Löb shows that under certain natural assumptions (the ones codified by (L1)-(L3) below) on the provability predicate, the answer to Henkin's question is that the literal Henkin sentence is true. In fact, under these conditions every solution to the 'Henkin equation' is provably true. With the formulation of the Löb conditions the modal study of provability could begin. For a good part of the subsequent story the reader is referred to Boolos and Sambin (1991).

2 What is provability logic?

Provability logic is best described as the sustained investigation of formal provability and related notions as modal operators. Consider the language L of modal propositional logic (see Modal logic). Let A be a language of predicate logic and let *T* be a theory in A. *T* should satisfy a number of conditions, which we will specify later. For the moment let us simply pretend that A is the language of arithmetic and that *T* is Peano arithmetic (PA). An 'interpretation' $(.)\sigma$ of L in A with respect to *T* is given by a mapping σ of the propositional variables to sentences of A. This mapping is extended to a full interpretation via the following clauses:

$$\begin{aligned}
(p)\sigma &= \sigma(p),\\
(\bot)\sigma &= \bot,\ (\top)\sigma = \top,\\
(A \wedge B)\sigma &= (A)\sigma \wedge (B)\sigma,\\
(A \vee B)\sigma &= (A)\sigma \vee (B)\sigma,\\
(A \rightarrow B)\sigma &= (A)\sigma \rightarrow (B)\sigma,\\
(\Box A)\sigma &= \mathrm{Prov}_T(\underline{\#((A)\sigma)}).
\end{aligned}$$

Only the last clause deserves comment. Prov_T is the arithmetized provability predicate for *T*. $(A)\sigma$ is a sentence of A. $\#((A)\sigma)$ is the Gödel number of $(A)\sigma$. Finally, $\underline{\#((A)\sigma)}$ is the numeral of $\#((A)\sigma)$. Thus $\underline{\#((A)\sigma)}$ is a closed term of A denoting $\#((A)\sigma)$ under the standard interpretation. For example, $(\neg\Box\bot)\sigma$ translates to $\neg\mathrm{Prov}_T(\underline{\#(\bot)})$, the sentence expressing the consistency of *T*. $\underline{\#((A)\sigma)}$ is, in a sense, the obvious way available to *T* to talk about $(A)\sigma$.

The meaning of ' $(.)\sigma$' depends on the given *T*. Strictly speaking we should write ' $(.)(\sigma, T)$', or something similar, to make the dependence on *T* visible. However, to avoid too heavy a notational burden, we 'hide' *T*.

In the definition of $(A)\sigma$ a subformula occurrence of *B* in *A* can 'go to' two different things. If the occurrence is not in the scope of a ' $\Box$', then *B* translates to a formula of A. If the occurrence is in the scope of a ' $\Box$', then the translation of *B* contributes to the value of a term in $(A)\sigma$. This difference in treatment leads to the existence of modalized fixed points. Let us remind ourselves of the Gödel fixed point lemma (see Gödel's theorems). For any formula $C(x)$ of A with at most *x* free, there is a sentence *D* of A, such that $T \vdash D \leftrightarrow C(\underline{\#(D)})$. Observe that *D* 'occurs' in two different ways in the fixed point statement - as a formula and as a term. The fixed point lemma has the 'fixed point theorem' (1) below as a correlate in the modal language.

To formulate (1) we first give a definition. We say that *p* occurs 'only modalized' in *A* if all occurrences of *p* in *A* are in the scope of a box. For example, *p* occurs only modalized in $\Box(p \rightarrow \Box\Box p)$; *p* does not occur only modalized in $p \wedge \Box p$. Note that if *p* does not occur in *A* at all, then it occurs only modalized in *A*.

(1) Suppose that *p* occurs only modalized in *A*. Consider any σ from the propositional atoms to the sentences of A. We can find a *D* in A such that if τ is the function such that $\tau(q) = \sigma(q)$ for $q \neq p$ and $\tau(p) = D$, then we have $T \vdash (p \leftrightarrow A)\tau$.

For example, we can find a τ such that $T \vdash (p \leftrightarrow \neg\Box p)\tau$. Note that this is just a reformulation of the most celebrated application of the Gödel fixed point lemma: the proof of the existence of a Gödel sentence *G* such that

$T \vdash G \leftrightarrow \neg\mathrm{Prov}_T(\underline{\#(G)})$.

Note that if we were to drop the 'guardedness condition', that *p* must be only modalized, then we would immediately get the liar paradox by taking $A = \neg p$. We will see below that (1) has an unexpected improvement in (4).

Let Γ be a finite set of L-formulas. We write $(\Gamma)\sigma$ for $\{(A)\sigma | A \in \Gamma\}$. We define $\Gamma \models_T A$ (pronounced '*A* is a *T*-consequence of Γ') to mean that for all σ,

$T + (\Gamma)\sigma \vdash (A)\sigma$.

'$\models_T$' is a relation of arithmetical validity (with respect to *T*) for the modal language. Interpretations take the place of semantics in this definition. The dependence on *T* in the definition is two-fold: first, we are considering

principles derivable in *T*; second, the interpretation of '$\Box$' under the $(.)\sigma$ is the arithmetization of the predicate 'is provable in *T*'.

We have, for example, $\Box p \models_T \Box\Box p$. Translating this back we find that it means that, for all sentences *A* of the language of *T*,

$$T + \mathrm{Prov}_T(\underline{\#(A)}) \vdash \mathrm{Prov}_T(\underline{\#(\mathrm{Prov}_T(\underline{\#(A)}))}).$$

You could say that $\Box p \models_T \Box\Box p$ expresses the validity of a schematic step with respect to *T*. The unusual thing is that the schema is formulated in a different kind of language: the language of modal propositional logic, which contains a logical operator not occurring in the language of *T*.

The relation '$\models_T$' can be axiomatized by a theory called Löb's logic (L or PrL for provability logic, or GL for Gödel-Löb logic) with the following rules of proof:

(L1) $\vdash A \Rightarrow \vdash \Box A$
(L2) $\vdash \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$
(L3) $\vdash \Box A \rightarrow \Box\Box A$
(L4) $\vdash \Box(\Box A \rightarrow A) \rightarrow \Box A$.

(L4) is also known as Löb's principle. In view of the fact that in classical logic an implication follows from its consequent, Löb's principle tells us that an instance of the reflection principle $\vdash \Box A \rightarrow A$ is *only* provable if its consequent is already provable. In other words, *T* 'knows' that it can prove only trivial instances of reflection.

The soundness of L was, in effect, proved by M.H. Löb (1955). Completeness was proved by R. Solovay (1976). 'Solovay's completeness theorem' is as follows:

(2) Under appropriate conditions on the theory *T* (see below), $\mathrm{L} + \Gamma \vdash A \Leftrightarrow \Gamma \models_T A$.

L has a Kripke-type completeness theorem with respect to transitive models with an upwardly well-founded accessibility relation. (The accessibility relation is upwardly well-founded if there are no infinite ascending chains.) Since the logic has the finite model property, we have Kripke completeness also with respect to finite, transitive, irreflexive models. The logic is not compact for the Kripke semantics: one can produce an infinite set *X* of modal formulas that cannot be valid on any transitive, upwardly well-founded model. On the other hand, for every finite subset X_0 of *X* one can find such a model on which X_0 is valid.

There are two major results about what L can prove. The first (due to Bernardi, de Jongh and Sambin) is the fact that fixed points of modalized formulas are unique. The theorem says that if *p* and *q* are truly and provably fixed points of *A*, then they are truly and provably equivalent. We write $A(B/p)$ for the result of substituting *B* for *p* in *A*. Define $\Box^+ A := A \wedge \Box A$.

(3) Suppose that *p* occurs only modalized in *A* and that *q* does not occur in *A*. Then

$$\Box^+(p \leftrightarrow A), \Box^+(q \leftrightarrow A(q/p)) \vdash \Box^+(p \leftrightarrow q).$$

The second major result (due to de Jongh and Sambin) is the existence of explicit fixed points. The theorem says that the language L is rich enough to contain explicit definitions of the fixed points of formulas (with respect to propositional variables which occur only modalized in them).

(4) Suppose that *p* occurs only modalized in *A*. Then there is a *B* in L containing only variables occurring in *A*, but not containing *p*, such that $\vdash B \leftrightarrow A(B/p)$.

The proof of (4) yields an algorithm for computing *B* from *A* and *p*.

To illustrate (1), (3) and (4), consider $\neg\Box p$. (1) tells us that in the arithmetical interpretation this formula has a fixed point: precisely the celebrated Gödel sentence *G*. (3) tells us that, modulo provable equivalence in *T*, this sentence is unique. (4) tells us that we can find an explicit equivalent of *G* in the modal language. If we apply the algorithm associated to (4), we find that the explicit fixed point is $\neg\Box\bot$, the sentence expressing the consistency of *T*. Thus (4) links the unprovable sentence of Gödel's first incompleteness theorem to the unprovable sentence of

Gödel's second incompleteness theorem (see Gödel's theorems).

The difference between (1) and (4) is as follows. (1) tells us that for all interpretations of the propositional variables (except *p*) of *A* (with *p* modalized), we can assign a sentence to *p* yielding the fixed point property. *Prima facie* this sentence need not be expressible in the modal language. (4) tells us that the modal language is rich enough to define the fixed point explicitly.

We close this section with some remarks about the precise choice of the theory *T* and A, the language of *T*. To make the results above work, *T* must be rich enough to do arithmetization in. This means that a sufficiently rich theory of arithmetic must be interpretable in *T*. The most natural choice for this theory is Buss' S^1_2, a theory far weaker than such well-known theories as primitive recursive arithmetic, $\mathrm{I}\Sigma^0_1$ or PA. The theory S^1_2 is not contained in $\mathrm{I}\Delta_0$. S^1_2 has induction restricted to the Σ^b_1-formulas, a class connected to the logical study of feasible algorithms (see Complexity, computational). $\mathrm{I}\Delta_0$ has induction restricted to the Δ_0-formulas. In S^1_2 some functions are provably total which are not provably total in $\mathrm{I}\Delta_0$. It turns out that the presence of S^1_2 is sufficient to verify the validity of L. However, to prove Solovay's completeness theorem (2) we seem to need more. Inspection of Solovay's proof shows that the theory we (*prima facie*) need is $\mathrm{I}\Delta_0 + \mathrm{Exp}$. Here Exp is an axiom expressing that the exponentiation function is total. It is consistent with everything we know that S^1_2 has a provability logic that is stronger than L! We can now give the first condition that the theory *T* should satisfy to be sound and complete for Löb's logic.

Condition 1: $\mathrm{I}\Delta_0 + \mathrm{Exp}$ is interpretable in *T*.

We do not demand that $\mathrm{I}\Delta_0 + \mathrm{Exp}$ be literally a sub-theory of *T*, since we want the results of provability logic to be applicable to such theories as ZF (see Set theory) in which arithmetic is present only via a translation of the arithmetical language into the language of set theory.

A second condition on the theory is that it be sufficiently sound. For example, by the second incompleteness theorem $U := \mathrm{PA} + \mathrm{Prov}_{\mathrm{PA}}(\underline{\#(\bot)})$ is consistent. As is easily seen, $U \vdash \mathrm{Prov}_U(\underline{\#(\bot)})$. So $\models_U \Box\bot$ and, hence, Solovay's completeness theorem cannot hold for *U*. We exclude such theories as *U* by demanding that *T* is Σ_1-sound, that is, the Σ_1-sentences *T* proves are all true. (A somewhat sharper condition is possible, but space does not permit its introduction here.)

Condition 2: *T* is Σ_1-sound.

3 What does provability logic achieve?

We end with some comments on the usefulness of provability logic. First the mathematical side. Clearly the modal presentation of arithmetical arguments makes them very perspicuous. So provability logic achieves Kreisel's excellent aim of 'saying simple things simply'. Also, by Solovay's theorem, provability logic completely describes the arithmetically valid modal provability principles. Moreover, it enables us to study these principles using the tools of modal logic (for example, Kripke semantics). Thus, by switching to the modal language, we gain control.

Now for philosophical relevance. Provability logic offers a fully worked-out syntactic treatment of modality (in the sense of Quine 1953). Since its principles are quite different from what one would ordinarily expect of notions of knowledge, necessity and so on, one way of viewing what is achieved is to conclude that syntactic treatments of such notions are not possible (for a brief discussion, see Thomason 1980). Moreover, comparing intuitively plausible modal principles for notions such as knowledge, human provability and the like with the principles satisfied by formal provability gives some sense of how these notions differ from formal provability. The reflection principle $\vdash \Box A \rightarrow A$ is built into or internal to the ordinary notions of knowledge and (human) provability. In contrast, formal provability satisfies Löb's principle ($\vdash \Box(\Box A \rightarrow A) \rightarrow \Box A$). We can 'see' that the difference between human and formal provability is not simply one of *strength* (but, I would suggest, one of *kind*). (The popularity of the view that the difference is one of strength can be seen, for example, in the discussion concerning minds and machines, where it is sometimes claimed on the ground of Gödel's incompleteness theorems that humans can beat machines in a theorem-producing game.)

See also: Logical and mathematical terms, glossary of

ALBERT VISSER

References and further reading

Boolos, G. (1993) *The Logic of Provability*, New York and Cambridge: Cambridge University Press. (A very readable textbook; includes an up-to-date introduction to predicate provability logic.)

Boolos, G. and Sambin, G. (1991) 'Provability: The Emergence of a Mathematical Modality', *Studia Logica* 50 (1): 1-23.(A brief history of the subject for roughly the period 1933-80.)

Gödel, K. (1933) 'Eine Interpretation des intuitionistischen Aussagenkalküls', *Ergebnisse eines Mathematischen Kolloquiums* 4: 39-40; trans. 'An Interpretation of the Intuitionistic Propositional Calculus', in *Collected Works*, ed. S. Feferman, J.W. Dawson, W. Goldfarb, C.D. Parsons and R.M. Solovay, Oxford and New York: Oxford University Press, 1995, vol. 3, 296-302.(Includes a thorough discussion by A.S. Troelstra.)

Henkin, L. (1952) 'A Problem concerning Provability', *Journal of Symbolic Logic* 17: 160, problem 3. (The paper in which Henkin posed the crucial question 'Is a sentence that ''says'' ''I am provable in Σ'' provable in Σ?'.)

Kreisel, G. (1953) 'On a Problem of Henkin's', *Indagationes Mathematicae* 15: 405-6.(Kreisel's answer to Henkin: 'It depends… '.)

Löb, M.H. (1955) 'Solution of a Problem of Leon Henkin', *Journal of Symbolic Logic* 20: 115-18.(As the title of this paper suggests, Löb answered Henkin's problem, and also proved the soundness of Löb's logic.)

Quine, W.V. (1953) 'Three Grades of Modal Involvement', in *Proceedings of the 11th International Congress of Philosophy* 14, Amsterdam: North Holland, 65-81.(Quine shows that provability logic offers a fully worked-out syntactic treatment of modality.)

Smorynski, C. (1985) *Self-Reference and Modal Logic*, New York: Springer.(A well-written and elegant textbook; includes good discussions of multi-modal provability logic and of non-extensional self-reference.)

Solovay, R.M. (1976) 'Provability Interpretations of Modal Logic', *Israel Journal of Mathematics* 25: 287-304.(A proof of the completeness of Löb's logic.)

Thomason, R.H. (1980) 'A Note on Syntactical Treatments of Modality', *Synthèse* 44: 391-5. (Presents the view that syntactic treatments of knowledge, necessity and so on are not possible.)

Providence

Divine providence is God's care, provision, foresight and direction of the universe in such a way that the universe as a whole and individual creatures within it fulfil God's purposes. Belief in providence was affirmed by some Greek philosophers (especially the Stoics), and is a fundamental tenet of Judaism, Christianity and Islam. In modern times, questions have arisen about the possibility of divine intervention in worldly affairs; this creates a difficulty for belief in providence, because it makes it hard to see how God can shape events so as to carry out his providential purposes.

Theories of providence differ with regard to the extent to which God has direct and specific control over earthly events, as opposed to guiding the course of affairs in a general way towards his overall goals. The strongest affirmation of divine control comes from Calvinism, which accepts a compatibilist view of free will (that is, that free will is compatible with determinism) and affirms God's absolute control over everything that happens. Other views affirm libertarian free will for creatures; some place limitations on God's knowledge of the future, and process theology places stringent limits on God's power to affect worldly events. These limitations tend to give creatures a limited degree of independence over against God, and lessen God's direct and specific control over events.

1 A brief history of providence

Belief in a personal or quasi-personal deity has often been virtually synonymous with belief in some kind of divine providential governance of the world. Among the Homeric Greeks, to be sure, fate was an uncanny, impersonal destiny, taking precedence even over the will of the gods. But in Hesiod, the supreme governance of the universe is in the hands of Zeus, who awards both prosperity and punishments in accordance with the law of justice. Plato, in the *Laws*, wrote that 'It is a matter of no small consequence, in some way or other to prove that there are Gods, and that they are good, and regard justice more than men do' (Laws 10). Thus Plato rejected the prevalent mythology because it attributed moral defect to the gods, and the Demiurge of the *Timaeus* 'being without jealousy…desired that all things should come as near as possible to being like himself' (*Timaeus* 29e).

The most serious account of divine providence among the Greek philosophers came from the Stoics. According to them, the ruling power of the universe - God, or the Logos - orders all things for the best in accordance with a universal law of causation. Suffering and calamities are no exception to this. Thus, the path of wisdom is to accept whatever circumstances befall as divinely ordained; one should not complain about the part in the cosmic drama one has been assigned, but rather concern oneself with playing that part in the best way possible. One's inner response to circumstances is to some degree in one's own control; the circumstances themselves are not. It is not surprising that this first attempt to state a rigorous and thoroughgoing account of providence also generated the first major encounter with the problem of evil (see Stoicism §20).

The scriptures of all three major monotheistic religions place great emphasis on divine providence. The Hebrew Bible expresses strongly the belief in God's presence, activity, and direction of all things: 'The Lord is good to all, and his compassion is over all that he has made' (Psalms 145: 9). God's providential rule is seen especially in the history of the Hebrew people: 'For I brought you up from the land of Egypt, and redeemed you from the house of bondage; and I sent before you Moses, Aaron, and Miriam' (Micah 6: 4). But God's hand is not absent from the affairs of other nations: '"Are you not like the Ethiopians to me, O people of Israel?" says the Lord. "Did I not bring up Israel from the land of Egypt, and the Philistines from Caphtor and the Syrians from Kir?"' (Amos 9: 7). Similar themes are maintained in the Talmud: 'The Holy One sits and nourishes both the horns of the wild ox and the ova of lice' (*Shabbat* 107b). And with regard to human affairs, 'He is occupied in making ladders, casting down the one and elevating the other' (*Genesis Rabbah* 68.4).

The New Testament similarly attests God's intimate involvement in all aspects of existence. Best known, perhaps, are the words of Jesus: 'Are not two sparrows sold for a penny? And not one of them will fall to the ground without your Father's will. But even the hairs of your head are all numbered' (Matthew 10: 29-30). God's reign extends over worldly authorities: 'Let every person be subject to the governing authorities. For there is no authority except from God, and those that exist have been instituted by God' (Romans 13: 1). And Christ, as 'Lord of Lords and King of Kings' (Revelation 17: 14), will in the end subject all nations to himself.

The Qur'an, too, emphasizes God's universal providential activity:

> Allah is He Who raised up the heavens without any pillars that you can see…. He it is Who spread out the earth and made therein firmly fixed mountains and rivers, and of fruits of every kind He has made pairs. He causes the night to cover the day. In all this, verily, are Signs for a people who reflect.
>
> (13: 2-5)

The theme of providence is taken up, in varying degrees, by the early Fathers and theologians of the Christian Church. Their reflections are mainly inspired by Scriptural testimony, but also include themes and references derived from the Greek philosophers. According to Irenaeus, for example:

> God does… exercise a providence over all things…. And, for this reason, certain of the Gentiles, … being moved, though slightly, by His providence, were nevertheless convinced that they should call the Maker of this universe the Father, who exercises a providence over all things, and arranges the affairs of our world.
>
> (*Against Heresies*, 3.25.1)

Augustine, however, gives by far the fullest account of providence to be found in the ancient Church. In *The City of God*, he paints a grand picture of the cosmic drama, beginning with the creation and the Fall, first of Satan and his angels, then of Adam and Eve, and culminating in the new heavens and the new earth, as the City of God and the earthly city each attain their appointed destinies in heaven and hell respectively. While the main emphasis is on God's redemption of humanity through Israel, Jesus Christ and the Christian Church, God's providence is involved in the affairs of pagan nations as well:

> The one true God, who never permits the human race to be without the working of His wisdom and His power, granted to the Roman people an empire, when He willed it and as long as He willed it. It was the same God who gave kingdoms to the Assyrians and even to the Persians.
>
> (*The City of God*, bk 5, ch. 21)

The main themes of Augustine's treatment survived intact through the Middle Ages and into the post-Reformation period, though his doctrines of grace and predestination remained controversial (see Augustine §§6, 10). Providence was often considered under the three headings of conservation (or 'sustenance'), concurrence and governance. The first two of these represent metaphysical aspects of God's support of his creatures; he both sustains them in existence and 'concurs' (or cooperates) with all of their activities, so that nothing in the world takes place without the exercise of divine power (see Creation and conservation, religious doctrine of §5). Governance, on the other hand, has to do with God's specific direction of worldly events to serve his purposes; it is primarily on this topic that problems arose in the modern era.

2 Post-Enlightenment issues and problems

For a variety of reasons, the Enlightenment was characterized by a general aversion to the miraculous, sometimes supported by philosophical arguments purporting to show either that miracles are impossible or that it is impossible in principle for human beings to ascertain that a miracle has occurred. Given this aversion, which found expression first in deism and then in various forms of liberal theology, it became quite problematic to speak of divine intervention in worldly affairs. This in turn makes it very difficult to see divine providence as manifested in particular occurrences. God's conservation and concurrence are entirely general in nature, and cannot serve to pick out some events as providential in contrast with others. So if God cannot in any way intervene to shape the course of events in one direction rather than another, there is a genuine question about the significance of any claim to 'see God at work' in specific events. Religious faith, on the other hand, cannot well be sustained without some sense of the active working of God. But with a ban on 'realistic' talk of divine intervention, identifying 'God's work in the world' tends to become an entirely subjective matter, lacking not only in public verification but even in publicly accessible meaning.

A second development is found in the tendency during the eighteenth and nineteenth centuries for the doctrine of providence to be closely associated or even identified with the belief in historical optimism and inevitable progress. This may be in part a reflex from the first development mentioned above. If we cannot with any confidence identify specific events as manifestations of divine governance, where else can we hope to see

providence except in the shape and direction of the human story as a whole? But the tendency was also nourished by the general optimism of the Enlightenment, and later by the growth of evolutionary interpretations of history. Doctrines of historical optimism and inevitable progress, however, have nearly enough been refuted by the empirical facts of history in the twentieth century. And to the extent that belief in divine providence was associated with these doctrines, it suffered a blow from which it has not yet recovered.

Yet a third development in recent thought about providence is the increased concern, among both critics of religion and believers, with the problem of evil (see Evil, problem of). The problem of evil is certainly not a new one; it was used against the Stoics by the Epicureans and against the early Christians by the Marcionites and later by the Manicheans (see Manicheism). Nevertheless, it is undeniable that the preoccupation with this problem has increased, though the reasons for the increase are not entirely clear. No doubt the tendency noted in the previous paragraph to identify belief in providence with historical optimism, has played a part. The horrendous evils of the twentieth century - the two World Wars and the Holocaust, in particular - have also played a role, though it is not clear that human life on average has been worse for people of this century than for those of earlier epochs. It may be, also, that our perception of the real possibilities, through technology and medicine, of a better life for human beings has made us less tolerant of the deplorable conditions in which very many human beings have lived and still live today. In any case, it is evident that many appeal to the world's evil as a demonstration that belief in God and in providence - at least, as traditionally conceived - is no longer viable. Many theologians and philosophers of religion are more willing than their counterparts in earlier times to consider limitations on God's control over worldly events in order to lessen the directness of his involvement in and responsibility for evil.

3 The resources of providence: five options

A crucial question for any doctrine of providence concerns the resources that are available to God as he exercises his controlling oversight. The key variables here are the power of God, God's knowledge (especially his knowledge of the future), and the nature of human freedom. The views of providence that adhere generally to traditional Western theism, however, do not differ significantly concerning the power of God. (Process theism, as we shall see, is an exception here.) All the major views accept that God's power is limited only by what is logically possible, and perhaps by ethical constraints inherent in the divine nature (see Omnipotence §§1-3). This means that none of these views finds the notion of divine intervention in the world fundamentally problematic, though theologies will differ with respect to the sorts of divine interventions that actually occur. But there are serious disagreements concerning the nature of human freedom and divine knowledge (see Omniscience §§3-4). Here we survey five different views, beginning with the one which affords the highest degree of divine control over the world.

Theological determinism, or 'Calvinism' (so called because of its association with Calvin (§§3-4), though it was held already by Augustine and probably also by Aquinas), holds that everything that happens, with no exceptions, is efficaciously determined by God in accordance with his eternal decrees. This means that God's control over events is absolute; there is no non-logical limitation on the states of affairs God is able to bring about. Calvinists do usually affirm human freedom, but only in a compatibilist sense. And God is able to know the future comprehensively simply by knowing his own intentions.

The theory of divine middle knowledge is also known as Molinism, after Luis de Molina (§§1-3), who first propounded it (see Molinism). The key idea of Molinism is the attribution to God of the knowledge of a certain class of propositions, now usually called 'counterfactuals of freedom'. Such counterfactual propositions state, concerning a particular free creature, what that creature would do in a particular situation of (libertarian) free choice with which it might be confronted. According to Molinism, God knows all the counterfactuals of freedom concerning all actual and possible free creatures. Thus, the fact that the creature is endowed with libertarian freedom constitutes no limitation on God's ability to anticipate the creature's actions; God merely consults his knowledge of the relevant counterfactuals in order to know exactly what the creature would do in any circumstance in which it might find itself. And this is of immense benefit with regard to God's providential governance of the world: God can survey in advance all of the possible scenarios for creation (all the 'possible worlds'), and select the very one which best fulfils his creative purposes. And once the decision is made, the exact fulfilment of God's plan is guaranteed; the necessity for any risk-taking on God's part is eliminated entirely. God's control is not as absolute as with Calvinism, for God is limited by the counterfactuals of freedom, which are not

under his control. But given the counterfactuals, God is able to guarantee the actualization of the very best ‘feasible world’.

The theory of ‘simple foreknowledge’ agrees with Molinism that human beings possess libertarian free will, and also that God has comprehensive knowledge of the future. But it does not accept that God possesses middle knowledge, knowledge of the counterfactuals of freedom. (Many adherents of simple foreknowledge deny that there are any true counterfactuals of freedom.) God’s knowledge of the actual future is usually regarded as a kind of ‘direct vision’ - God ‘sees’ it as if in a telescope or a crystal ball.

The view that is coming to be known as ‘free-will theism’ differs from the simple foreknowledge theory by denying that God has comprehensive knowledge of the future. Free-will theism leans heavily on the arguments purporting to show that infallible foreknowledge of future free actions is impossible. God, however, has comprehensive knowledge of the world’s past and present, as well as knowledge of the dispositions, tendencies and so on of free agents. Furthermore, God is able to know many important things about the future on the basis of his own intentions, which in key respects are not contingent on the decisions of creatures. Nevertheless, God is not always able to guarantee in advance the results that will follow from his own actions and the situations he chooses to create; thus, for free-will theism, providence involves a genuine element of divine risk-taking.

The final option to be considered is process theism, a contemporary view which draws heavily on the philosophy of A.N. Whitehead (see Process theism; Whitehead, A.N. §4). Process theism agrees with free-will theism that divine foreknowledge is necessarily limited by creaturely freedom. But it deviates sharply from all versions of traditional theism in its conception of God’s power and activity. The process view of divine power may be summarized succinctly by saying that such power is always persuasive and never coercive. God is at all times attempting to persuade humans and other creatures to follow his way, by presenting to them his ‘initial aim’ for them, the ‘direction’ in which he wants them to go so as to realize rich and intense values in a harmonious fashion. But he does not, because he cannot, compel any created being to do his will, and so the actual course of events depends on how finite beings respond to God’s persuasion. This means that God’s ability to secure the fulfilment of his plan is actually very limited in comparison with that posited by all the other theories considered - and also in comparison with what seems to be presupposed in the Jewish, Christian and Muslim Scriptures.

4 Assessing the options

Clearly Calvinism affords God the maximal degree of control over worldly affairs. (It is noteworthy, however, that it does not in fact accord to God any greater power than he possesses according to Molinism, simple foreknowledge, or free-will theism.) The price to be paid for this absolute control is extreme difficulty in accounting for the world’s evil. If God has by his own decree ensured the existence of all the evil and sin that there is, how can human beings justly be held responsible for it? And how can God himself avoid being responsible for evil? How, indeed, is it intelligible that God has chosen to bring about sin and moral evil, and has then assumed towards what he has knowingly and deliberately chosen an attitude of utter, implacable hostility? It seems likely that there is no rationally comprehensible answer to these questions; Calvinists themselves frequently say that the relationship between God and evil is an impenetrable mystery.

Molinism, unlike Calvinism, accepts the existence of libertarian freedom and thus avoids the most severe of Calvinism’s difficulties over evil. At the same time, it manages to ascribe to God a degree of providential control that is second only to Calvinism and is probably the greatest control possible if determinism is rejected. There are, however, serious metaphysical difficulties involved in the postulation of counterfactuals of freedom and middle knowledge. And Molinism does not remove all the sting of the problem of evil, for each instance of evil that occurs is one that God has specifically ‘planned, ordered, and provided for’ (Freddoso 1988: 3; in Molina 1592). So one is led to the difficult assertion (though one that is often enough made) that each and every instance of evil that exists is necessary for the existence of a ‘greater good’ (or the prevention of a greater evil) that God could not bring about without the evil in question.

The arguments concerning the compatibility of comprehensive divine foreknowledge and libertarian free will are well known and highly controversial. Somewhat surprisingly, these arguments do not have any important bearing on divine providence once middle knowledge has been given up. The reason for this is that simple foreknowledge does not contribute anything to God’s providential control over the world beyond what he would have without it.

To see this, consider the following argument: suppose God possesses comprehensive knowledge of the state of the earth as of some future time *t*. The state of the earth at *t* presupposes the entire causally relevant past history leading up to *t*. Now suppose that God decides, on the basis of his knowledge of the earth at *t*, to intervene in the earth's history at some earlier time $t - n$. A little reflection shows this scenario to be incoherent. For what God knows is, by hypothesis, the actual history of the earth leading up to *t*; to suppose that, on the basis of that knowledge, God intervenes in such a way that the actual history consequent upon the intervention is different from the history God knows to be actual makes no sense. We may conclude from this that it is impossible that God should use foreknowledge derived from the actual occurrence of future events to determine his own prior actions in the providential governance of the world. If simple foreknowledge did exist, it would be useless. (The same argument applies to divine timeless knowledge of the future, where this is not accompanied by middle knowledge (see Eternity).)

So free-will theism, simple foreknowledge, and divine timelessness can be treated together. All these views entail what may be termed a 'risk-taking' view of providence; they entail that when God makes decisions that depend for their outcomes on the responses of free creatures, his making of those decisions is not guided by a prior knowledge of the outcomes. This does something to mitigate the problem of evil, in that it need not be assumed that God has specifically 'planned, ordered, and provided for' each instance of evil that actually occurs. It also, to be sure, means that God's control over events is somewhat less direct and pervasive than theology has often assumed. Whether these consequences are acceptable is an important question for theology.

The chief appeal of a process view of providence (over and above the possible attractiveness of a God who is wholly non-coercive) lies in the fact that it virtually disposes of the problem of evil: God cannot properly be said to permit the world's evil, because he has no power to prevent it from occurring. The price of this is a virtual abandonment of the view (central to all historic conceptions of providence) that God is in control of what happens in the world. Whether this is acceptable is, once again, an important question, of fundamental religious as well as theological import.

See also: Alexander of Aphrodisias §4; Deism; Eschatology; Grace; Miracles; Predestination; Reprobation; Revelation §1

WILLIAM HASKER

References and further reading

Augustine (413-27) *The City of God*, trans. G.G. Walsh, D.B. Zema, G. Monahan and D.J. Honan, Garden City, NY: Image Books, 1958.(Classic account of the Christian drama of salvation from creation to final judgment.)

Bartholomew, D.J. (1984) *God of Chance*, London: SCM Press.(Examination of the issues posed for theology by the supposition that the physical world is fundamentally indeterministic.)

Berkouwer, G.C. (1950) *De Voorversiening Gods*, Kampen: J.H. Kok; trans. L.B. Smedes, *The Providence of God*, Grand Rapids, MI: Eerdmans, 1952.(The doctrine of providence from the perspective of Reformed theology.)

Farley, B.W. (1988) *The Providence of God*, Grand Rapids, MI: Baker Book House. (A useful historical discussion of the doctrine of providence, informed by a Reformed theological perspective.)

Flint, T.P. (1988) 'Two Accounts of Providence', in T.V. Morris (ed.) *Divine and Human Action: Essays in the Metaphysics of Theism*, Ithaca, NY: Cornell University Press, 147-81.(A finely crafted comparison of Thomism (= 'Calvinism') and Molinism.)

Griffin, D.R. (1991) *Evil Revisited: Responses and Reconsiderations*, Albany, NY: State University of New York Press.(A defence of the claim that process theism represents the best response to the problem of evil.)

Hasker, W. (1989) *God, Time and Knowledge*, Ithaca, NY: Cornell University Press. (A sustained argument against Molinism and simple foreknowledge, and in favour of free-will theism.)

Hasker, W. (1992) 'Providence and Evil: Three Theories', *Religious Studies* 28: 91-105. (Expansion of material in §3 and §4.)

Helm, P. (1994) *The Providence of God*, Downers Grove, IL: InterVarsity Press. (Argues for theological determinism and against Molinist and risk-taking views of providence.)

Irenaeus (180s) *Against Heresies*, in A. Roberts and J. Donaldson (eds and trans) *The Ante-Nicene Fathers*, vol. 1, New York: The Christian Literature Co., 1896. (Defence of the Christian faith against the Gnostic heresies of the second century.)

Lanford, M.J. (1981) *Providence*, London: SCM Press.(Good general introduction to the topic in the light of

current issues and problems.)

Molina, L. de (1592) *On Divine Foreknowledge (Part IV of the Concordia)*, trans. with introduction by A.J. Freddoso, Ithaca, NY: Cornell University Press, 1988.(The classic source for Molinism; the introduction is an outstanding exposition and defence of Molinism against the background of contemporary analytic philosophy.)

Plato (*c*.366-360 BC) *Timaeus*, trans. F.M. Cornford, Indianapolis, IN: Library of Liberal Arts, 1959.(Plato's 'likely story' concerning the origin and structure of the cosmos.)

Prudence

The word 'prudence' is used in several ways in contemporary English, and its different philosophical senses to some extent reflect that variety. Traditionally, prudence is the ability to make morally discerning choices in general; but the term is also used to denote a habit of cautiousness in practical affairs; most recently, attempts have also been made to identify prudence with practical rationality, perhaps even with the pursuit of the agent's own interests, without any specifically moral implications.

1 Prudence as a cardinal virtue, and contrasted with rashness

In its most traditional sense, 'prudence' translated the Latin *prudentia*, which in turn was the technical translation of Aristotle's *phronēsis* (*Nicomachean Ethics*) (see Aristotle §23). Most contemporary English-speaking philosophers translate *phronēsis* as 'practical wisdom', and discussion of this latter topic will reveal much that traditionally was meant by 'prudence'.

Prudence was considered a virtue of the practical intellect (see Virtues and vices §3). That is to say, it consisted in an admirable intellectual skill in practical, as distinct from theoretical, matters. The details of this, in Aristotle and in Aquinas, are a matter of controversy, but in general it can be said that the prudent person has to have both an understanding of what human fulfilment is and an ability to perceive which action, here and now, can contribute to living a life of human fulfilment (see Eudaimonia).

The intellectual virtue of prudence presupposes other intellectual skills. A successful doctor needs good diagnostic abilities in order to see what is wrong with a patient, a practical knowledge of the treatment appropriate to this patient here and now, and perhaps also the ability to plan ahead for various stages of a long-term treatment. These skills in themselves were thought to be morally neutral; they might equally well characterize a doctor intent on making a fortune as a highly effective practitioner. Specific to prudence is the ability to see how these medical skills can fit into a fulfilled human life, and hence how their use can be morally significant, not simply technically or financially successful (see Morality and ethics §2).

Prudence was traditionally contrasted with three other virtues: justice, fortitude and temperance. The latter two were thought to be 'moral' virtues, as distinct from prudence, which is an intellectual virtue. The term 'moral virtue' is confusing, since prudence, too, is concerned with the moral significance of what one does. The phrase 'moral virtue' reflects Aristotle's term *ēthikēaretē*, which referred to admirable emotional dispositions in the realms of aggressiveness and desire (see Morality and emotions §3). Justice, on the other hand, was thought to reside not in the emotions (as the moral virtues do), nor in the intellect (as prudence does), but in the will. The prudent person would understand what is morally required in each situation; the just person is habitually willing to do what is morally required in giving each person their due.

These four virtues were called 'cardinal' virtues (from the Latin *cardo*, meaning 'hinge') because it was held that possession of all the more specific virtues, such as loyalty, kindness, courage and honesty, turned on the possession of these four. Indeed, Aquinas, like Aristotle, believed that one cannot have prudence without fortitude and temperance, because inappropriate emotional reactions to a particular situation would cloud one's intellectual grasp of what was morally required (*Summa theologiae* IIaIIae.47, IIaIIae.56).

The ability to grasp what is morally required in individual situations is sometimes explained as the ability to reason from general moral principles to particular cases. But the classical writers rather took the view that seeing which principle applies, and how it applies, is not a matter of argument, but, as Aristotle put it, a kind of perception (see Virtue ethics §6). Critics of a rationalistic bent have objected to this form of moral intuitionism.

Sometimes, rather than being considered as a cardinal virtue of great generality, prudence is conceived of as a particular virtue, contrasted with rashness. Thus, it might be prudent to insure one's house against fire, and prudent for the managers of a trust fund to invest in such a way that the participants in the fund were not exposed to speculative risks. Seen in this way, prudence is a particular virtue, and still belongs to the realm of ethics. On the other hand, 'prudent' is occasionally used in a more critical sense: to insist on prudence is to insist on playing safe on every possible occasion. If it is thought that some risks are reasonable, or even that it is morally required on occasion to take some risk, prudence in this last sense might even be a vice rather than a virtue.

2 Possible nonmoral senses

More recently, philosophers have used the term 'prudence' in ways which are quite different from those already discussed. Two separate contrasts are involved, which unfortunately are often confused. The first is that between the moral and the nonmoral; and the second is that between the agent's own interests and the interests of other people (see Egoism and altruism). Invoking the first contrast, prudential reasons for doing something are held to be different in kind from moral reasons. Invoking the second contrast, prudential reasons are said to be those involving the agent's own interests, rather than those of others. On these grounds, it is sometimes argued that egoism does not count as a theory of morality at all, since the reasons egoists would give in support of an action are reasons connected either immediately or in the longer term with the furtherance of their own interests, and hence are 'merely prudential'. Moreover, it does seem to be the case that we are inclined to criticize people more severely for failing in their duties towards others, even if these duties are duties of prudence, than we are to criticize for failure in prudence in respect of oneself. This thought leads to the suggestion that failure in prudence with respect to oneself, though it may be irrational, is not properly a moral fault at all.

There are problems with this overall view. First, it is not at all clear that prudential considerations are properly thought of as nonmoral. If to insure one's house is prudent, is that not a moral reason in favour of doing so? And in general, there might be some risks which it is simply morally wrong to take. Second, it is not clear that prudential considerations are restricted to the interests only of the agent. The trust fund manager has prudential duties to the participants in the trust, and it is surely reasonable to suppose that these are moral duties. Third, it is not obvious that one is subject to moral requirements only towards other people and not towards oneself as well; Kant, for example, held that one has moral duties to oneself just as much as to others, as indeed do most theories which lay stress on impartiality as an important feature of the moral (see Impartiality; Kantian ethics). But if this is so, then the distinction between self-regarding and other-regarding considerations does not coincide with the distinction between the nonmoral and the morally significant. At least part of the difficulty may stem from the fact that we ordinarily use the term 'prudence' in a variety of senses according to context, some closer to the older usage, others much less so.

Be that as it may, 'prudence' is often used in contemporary philosophy as equivalent to 'practical rationality', without any implications about moral significance. The prudent choice is that which, broadly speaking, furthers the agent's ends. The qualification is necessary because many writers have considered that we have to assume that agents formulate their decisions on the basis of full information, and do so without being influenced by such things as bias, malice or emotional stress. So it might appear that moral considerations have been reintroduced into the concept, thus bringing it closer to *phronēsis* and the medieval *prudentia*. Moreover, in default of an adequate characterization of the information and state of mind required for prudent choice, the practical usefulness of the concept is much lessened. For this reason other writers have been prepared to say simply that the only empirical basis for assessing persons' interests is to be found in the choices they actually make (see Rational choice theory). In this way, particular decisions may be assessed as irrational, and hence as imprudent, on formal grounds such as inconsistency, or being self-defeating, or being ill-adapted to stated ends; but criticism aimed directly at the content of such choices is excluded.

Theories of nonmoral rational choice themselves present problems, especially in relation to political philosophy, in which consideration has to be given to the different preferences of different individuals. This part of the modern discussion of prudence touches on some of the issues dealt with in older discussions of justice. The traditional definition, 'a willingness to give each person their due', of course does not answer the question how one is to discover what is due to different people, nor the further question whether it is possible to give a rational account of the 'best' way of reconciling conflicting preferences. These are just the problems which recent accounts of rational choice theory try to investigate.

GERARD J. HUGHES

References and further reading

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, ed. T. Gilby, Cambridge: Blackfriars, 1965, vol. 36, IIaIIae.47, IIaIIae.56.(Based on Aristotle, these questions on prudence are a systematic account of the Western Christian understanding of prudence.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985, book VI, esp. chaps 1, 2, 5, 13.(The text on which most classic discussions of prudence interpreted as practical wisdom are based.)

Broadie, S. (1991) *Ethics with Aristotle*, Oxford: Oxford University Press, ch. 4.(A modern reading of Aristotle's *Nicomachean Ethics*, excellent for drawing out the implications of the text.)

Falk, W.D. (1963) 'Morality, self and others', in H-N. Castañeda and G. Nakhnikian (eds) *Morality and the Language of Conduct*, Detroit, IL: Wayne State University Press.(Though difficult to find, this valuable article assesses some recent attempts to interpret prudence in a nonmoral way.)

Gauthier, D.P. (1967) 'Morality and Advantage' *Philosophical Review* 74: 460-75, repr. in J. Raz (ed.) *Practical Reasoning*, Oxford: Oxford University Press, 1978, 185-97.(A modern theory of the relationship between morality and self-interest.)

Nagel, T. (1970) 'Desires, Prudential Motives, and the Present', in *The Possibility of Altruism*, Oxford: Clarendon Press; repr. in J. Raz (ed.) *Practical Reasoning*, Oxford: Oxford University Press, 1978, 153-67.(A modern account of prudence interpreted as rational choice.)

Sen, A. and Williams, B. (eds) (1982) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press.(A valuable collection of articles on utilitarianism and other theories of rational choice. The editorial introduction gives a quick and helpful overview.)

Woods, M. (1986) 'Intuition and Perception in Aristotle's Ethics', in J. Annas (ed.) *Oxford Studies in Ancient Philosophy*, vol. IV, Oxford: Clarendon Press.(A discussion of some of the difficulties in interpreting Aristotle's account.)

Pseudo-Dionysius (*fl. c.* AD 500)

'Pseudo-Dionysius' was a Christian Neoplatonist who wrote in the late fifth or early sixth century and who presented himself as Dionysius the Areopagite, an Athenian converted by St Paul. This pretence - or literary device - was so convincing that Pseudo-Dionysius acquired something close to apostolic authority, giving his writings tremendous influence throughout the Middle Ages and into the Renaissance.

The extant four treatises and ten letters articulate a metaphysical view of the cosmos, as well as a religious path of purification and perfection, that are grounded in the Neoplatonism developed in the Platonic Academy in Athens. Although this strand of Neoplatonist thought, in contrast to that developed at the school in Alexandria, was deliberately pagan in its religious orientation, Pseudo-Dionysius used its conceptual resources (drawing especially on Proclus) to give precision and depth to the philosophical principles of a Christian world view. Cardinal points of Pseudo-Dionysius' thought are the transcendence of a first cause of the universe, the immediacy of divine causality in the world and a hierarchically ordered cosmos.

1 The works: character and origin

Pseudo-Dionysius' works clearly reflect the late Neoplatonism of the fifth and sixth centuries, creatively modified in accordance with fifth-century Christian ideas and commitments. These facts belie the author's claim to be the disciple of Saint Paul who is described in Acts 17: 16-34 as an instantly converted member of an Athenian judicial body (*areopagos*) asked, by certain Epicurean and Stoic philosophers, among others, to judge the acceptability of Paul's preaching in Athens. The scholarly case against Pseudo-Dionysius' assumed identity was first made by Lorenzo Valla in 1457 and was publicized by Erasmus in 1504. Pseudo-Dionysius' real identity, however, remains unknown, despite continued debate during the sixteenth and seventeenth centuries and renewed scholarly investigation since the end of the nineteenth century.

The range of possible dates for authorship of the corpus is bounded at its early end by the conclusion of Proclus' career in AD 485. This dating is confirmed by Pseudo-Dionysius' appeal to parts of the Christian liturgy and to certain doctrinal formulae current in the late fifth century, and by his adaptation of late fifth-century Neoplatonist religious rites. The latest possible date is set by the earliest unmistakable citation of Pseudo-Dionysius' works, by Severus of Antioch between AD 518 and 528.

A translation into Syriac by Sergius of Reshaina (d. AD 536) made possible the rapid spread of Pseudo-Dionysius' ideas throughout the Syrian Christian centres of learning. Armenian and Arabic translations have also been found, and there is some evidence for Pseudo-Dionysius' influence in the Islamic realm. Despite these successes, however, the pseudonymous philosopher's claims to authority were not uncontroversial; nor was it clear that his views about the Trinity or the nature of Christ were orthodox. Eventual acceptance on both counts required repeated defences, for example by John of Scythopolis and by Maximus Confessor.

Despite a florid or seemingly even obscurantist style, Pseudo-Dionysius shows himself in his works to be a serious philosopher, struggling against limitations of language and wrestling with the paradoxicality of metaphysical propositions about God, creation, cosmic order and so on. His systematic intent shows not only in his philosophically most important work, *Peri theiōn onomatōn (On the Divine Names; Latin, De divinis nominibus)*, but also in his grand project. As he explains, this project commences with some treatments of the nature of God and the supra-sensible order insofar as they are accessible to us only with the aid of revelation. (These works are not extant and may never have been finished.) Then, relying more on human reason and thereby descending, as it were, in the cosmic hierarchy, Pseudo-Dionysius undertakes to treat God metaphysically in *Peri theiōn onomatōn*. From this standpoint he descends still further, treating God and the cosmic order insofar as they can be understood through sensible creatures, more distant manifestations of God that allow a kind of metaphoric understanding of their divine cause. Such objects Pseudo-Dionysius thinks of as symbols, and the two extant treatises *Peri tēs ouraniās hierarchiās (On the Celestial Hierarchy; Latin, De coelesti hierarchia)* and *Peri tēs ekklēsiastikēs hierarchiās (On the Ecclesiastical Hierarchy; Latin, De ecclesiastica hierarchia)* can be considered partial presentations of what he calls 'symbolical theology'.

Having reached the lowest point in the cosmic order, physical things and human religious institutions (whose

rituals facilitate the metaphoric understanding of God), Pseudo-Dionysius now tries to recapitulate and complete the cognitive endeavour to reach God. He does this by exposing the limitations of the various sorts of positive knowledge he has outlined in the foregoing treatises. In this way he means to insure his reader's return to God after the descent to the level of creatures. He describes the method by which this perspective is achieved in his *Peri mystikēs theologiās (On Mystical Theology; Latin, De mystica theologia)*.

2 Doctrine and context

The outline of Pseudo-Dionysius' philosophical project indicates a basic Neoplatonist principle on which his system rests: all finite things 'proceed from' (that is, are caused by) a being that is transcendent, infinite and perfect, and they eventually 'return to' this being since their fulfillment lies in closeness to it and they are sustained by it for this end. This cosmic cycle of procession and return is the pattern according to which the first cause, or God, exercises its originative capacity and, so to speak, manifests itself through a world of creatures (see Neoplatonism).

According to its classic presentation by Neoplatonists from Plotinus through Proclus, the cosmic cycle actually involves three stages. Prior to the stages of procession and return, there is a first stage of 'immanence' in which the effects abide in their cause, being in a certain sense contained in the cause prior to their being caused. In line with that traditional theory, Pseudo-Dionysius speaks of things in the world as being 'contained beforehand' in their cause 'in a higher way', that is, as initially undifferentiated. However, by focusing on and developing the idea of the first stage, he makes one of his most original contributions to Neoplatonist metaphysics.

To understand this, one must first note two problems within Neoplatonism. First, there is a tension between the claim that the first cause is absolutely simple, containing its effects in some wholly undifferentiated way, and the claim that it does in some way really contain these effects. Second, if the first cause is radically different from its effects in virtue of its simplicity and perfection, and if its causality consists in making things like itself (bringing things to 'participate' in its perfection), how can it actually give rise to effects in the world, since they are quite different from it in nature? Exponents of classical Neoplatonism tend, to one degree or another, to meet these difficulties by supposing intermediate levels of reality between the first cause and its ultimate effects in the world. These 'hypostases' are causally prior to things in the world of nature but are causally dependent on the first cause, and they are supposed to mediate causally between the first cause and its ultimate effects. Pseudo-Dionysius reverses this multiplication of intermediary hypostases, which, one might argue, reaches excessive proportions in later Neoplatonists such as Iamblichus and Proclus. He does so by further developing the notion of the first cause's precontainment of its effects. In *Peri theiōn onomatōn*, he argues that the predicates that would be construed in classical Neoplatonism as signifying intermediate hypostases in fact simply designate the first cause itself, considered with respect to effects that it is capable of producing in virtue of its intrinsic perfection.

This account of how the first cause causes things in the world is connected with two other salient Dionysian doctrines. Classical Neoplatonism had generally characterized the first, completely simple and perfect cause as 'the One', and had claimed that this cause of all things in some radical sense transcended being itself. Though Pseudo-Dionysius has often been taken to adhere to this claim that the first cause is 'beyond being', he in fact contends that the first cause is beyond the being of anything finite, even beyond the general sort of being that it can bestow on finite things; but it itself is perfect, infinite being. Pseudo-Dionysius seems to think that this characterization, in contrast to the classical Neoplatonist characterization, accords the first cause a kind of richness that explains how it can indeed be transcendently simple and yet in some sense contain all its effects. Second, Pseudo-Dionysius develops the Platonist idea that the first cause is the good itself. His famous principle, 'The good is such as to diffuse itself', indicates that it is in the very nature of the first cause, on account of its intrinsic perfection, to give rise to the variety and plurality of finite things in the world.

With regard to the return of finite beings to their source, Pseudo-Dionysius complements the ontological conception of a comprehensive hierarchy of beings of various degrees of perfection with an epistemology that stresses gradual cognitive ascent. The finite mind approaches the infinite source by climbing from an understanding of one level in the hierarchy to an understanding of the next. This ascent, which both requires and furthers moral and spiritual purification, finally ends in a state of unification with God that Pseudo-Dionysius describes as an 'inactivity of all knowledge', or a 'knowing beyond the mind', since in this state all particular knowledge must be left behind. Progressive illumination thus culminates, paradoxically, in the 'mysterious

darkness of unknowing'. Achieving this state requires ultimately a kind of radical psychic break with ordinary ways of thinking and acting - a spiritual ecstasy.

3 Philosophical influence

Pseudo-Dionysius had a substantial influence on Christian spirituality as well as on medieval and Renaissance philosophy. The influence of his ideas about the spiritual life can be seen most clearly in the writings of medieval and Renaissance mystics who echo some of his pronouncements about the soul's union with God (see Mysticism, history of). His method of apprehending God by negating positive knowledge - and ultimately seeing God as beyond both affirmation and negation - has appealed to the mystical temper. This method (*via negativa*) must be distinguished, however, from a mystical approach to God, if by that is meant either some sort of actual convergence of the enraptured soul with divinity or a non-intellectual approach (see Mysticism, nature and assessment of; Negative theology).

Pseudo-Dionysius' particular philosophical influence is especially marked and important in the Latin West. Its first phase comprises his entry into Western thought in the Carolingian period, via a translation into Latin by Hilduin in 832, and his appropriation by Eriugena, who retranslated the works in 862 and extended his ideas in the direction of metaphysical monism. The second phase of influence occurs as part of the rebirth of learning in the twelfth century, and can be seen in such diverse philosophers as Hugh of St Victor, Isaac of Stella and Peter Lombard. His aesthetics (a chapter of *Peri theiōn onomatōn* is devoted to the predicate 'beauty') made its mark by inspiring, directly or indirectly, Abbot Suger's programme for a new architecture, the Gothic, that would lead the soul upward via sensible reality.

The third phase of influence lies in Pseudo-Dionysius' formative effect on some of the greatest philosophers of the thirteenth century: Robert Grosseteste, Albert the Great, Thomas Aquinas and Bonaventure. These philosophers (except Bonaventure) wrote substantial commentaries on Pseudo-Dionysius; Albert and Aquinas especially used their detailed knowledge of Dionysian texts to assimilate and then develop Platonist principles, making them foundational for their metaphysical systems. Finally, Pseudo-Dionysius' influence appears in the Renaissance as a factor in the revival of interest in Platonism in general. Pico della Mirandola and Marsilio Ficino, affirming the genuineness of Pseudo-Dionysius' claims to authority, see him as the exemplary Christian Platonist and hence as the exemplary philosopher.

In general, Pseudo-Dionysius appeals to medieval and Renaissance philosophers because he seems to them to be leading the way in regard to a common enterprise: the adaptation of Platonist philosophy to the purposes of monotheistic metaphysics. His ideas of creative goodness, creation against the backdrop of nothingness, the metaphysical primacy of being, the nature of evil as privation and the orderliness of the cosmos in terms of a hierarchy of perfection give his writings exceptional philosophical importance, an importance far beyond what they could have had merely by virtue of their presumed authority.

See also: Being; Creation and conservation, religious doctrine of; Neoplatonism; Platonism, medieval; Ulrich of Strasbourg

HANNES JARKA-SELLERS

List of works

Pseudo-Dionysius (*c.* AD 500) Editions of the Complete Works: G. Heil, A. Ritter and B. Suchla (eds) *Corpus Dionysiacum*, Berlin: de Gruyter, 1990-1, 2 vols; trans. C. Lubheid and P. Rorem, *Pseudo-Dionysius: The Complete Works*, London: Society for the Promotion of Christian Knowledge, 1987; trans. M. de Gandillac, *Pseudo-Denys l'Aréopagite. Oeuvres complètes. Traduction, préface et note*, Paris: Aubier, 1980.(Heil, Ritter and Suchla is a very thorough, collaborative critical edition in two volumes, with a detailed introduction to the whole edition and an extensive bibliography in Volume 1. Lubheid and Rorem is the English reader's best introduction, with accessible, fairly accurate translations with helpful notes. De Gandillac serves as a valuable resource complementing English scholarship.)

Pseudo-Dionysius (*c.* AD 500) *Peri theiōn onomatōn (On the Divine Names; Latin, De divinis nominibus)*, in *Corpus Dionysiacum*, vol. 1, Berlin: de Gruyter, 1990; trans. C. Lubheid and P. Rorem, *Pseudo-Dionysius: The Complete Works*, London: Society for the Promotion of Christian Knowledge, 1987; trans. M. de Gandillac, *Pseudo-Denys l'Aréopagite. Oeuvres complètes*, Paris: Aubier, 1980.(Other translations of importance are B.

Suchla, *Pseudo-Dionysius Areopagita. Die Namen Gottes*, Stuttgart: Hiersemann, 1988; and J. Jones, *Pseudo-Dionysius: The Divine Names and Mystical Theology*, Milwaukee, WI: Marquette University Press, 1980. Jones is an interesting if somewhat idiosyncratic rendering, with helpful references to relevant texts of Aquinas.)

Pseudo-Dionysius (*c.* AD 500) *Peri tēs ouraniās hierarchiās (On the Celestial Hierarchy; Latin, De coelesti hierarchia)*, in *Corpus Dionysiacum*, vol. 2, Berlin: de Gruyter, 1990; trans. C. Lubheid and P. Rorem, *Pseudo-Dionysius: The Complete Works*, London: Society for the Promotion of Christian Knowledge, 1987; trans. M. de Gandillac, *Pseudo-Denys l'Aréopagite. Oeuvres complètes*, Paris: Aubier, 1980.(A further translation is G. Heil, *Pseudo-Dionysius Areopagita. Uber die himmlische Hierarchie. Über die kirchliche Hierarchie*, Stuttgart: Hiersemann, 1986. This is an up-to-date, accurate translation.)

Pseudo-Dionysius (*c.* AD 500) *Peri tēs ekklēsiastikēs hierarchiās (On the Ecclesiastical Hierarchy; Latin, De ecclesiastica hierarchia)*, in *Corpus Dionysiacum*, vol. 2, Berlin: de Gruyter, 1990; trans. C. Lubheid and P. Rorem, *Pseudo-Dionysius: The Complete Works*, London: Society for the Promotion of Christian Knowledge, 1987; trans. M. de Gandillac, *Pseudo-Denys l'Aréopagite. Oeuvres complètes*, Paris: Aubier, 1980.(A further up-to-date translation is G. Heil, *Pseudo-Dionysius Areopagita. Über die himmlische Hierarchie. Über die kirchliche Hierarchie*, Stuttgart: Hiersemann, 1986.)

Pseudo-Dionysius (*c.* AD 500) *Peri mystikēs theologiās (On Mystical Theology; Latin, De mystica theologia)*, in *Corpus Dionysiacum*, vol. 2, Berlin: de Gruyter, 1990; trans. C. Lubheid and P. Rorem, *Pseudo-Dionysius: The Complete Works*, London: Society for the Promotion of Christian Knowledge, 1987; trans. M. de Gandillac, *Pseudo-Denys l'Aréopagite. Oeuvres complètes*, Paris: Aubier, 1980.(Further translations include A. Ritter, *Pseudo-Dionysius Areopagita. Über die mystische Theologie und Briefe*, Stuttgart: Hiersemann, 1994; and J. Jones, *Pseudo-Dionysius: The Divine Names and Mystical Theology*, Milwaukee, WI: Marquette University Press, 1980. Ritter's translation is supplemented by an informative introductory essay on Dionysian philosophical scholarship. Jones' edition contains useful links to relevant texts of Aquinas.)

References and further reading

1 Patristic and medieval commentaries

Albert the Great (*c.*1250-60) *Super Dionysium de divinis nominibus (Commentary on De divinis nominibus)*, ed. P. Simon in *Opera omnia*, vol. 37, part 1, Münster: Aschendorff, 1972.(A critical edition of Albert's commentary; supplements that were found in the 1892 edition of the *Opera omnia*, edited by A. Borgnet, Paris: Vives, 1892.)

Aquinas, Thomas (*c.*1265-8) *In librum beati Dionysii De divinis nominibus expositio (Commentary on Dionysius' De divinis nominibus)*, ed. C. Pera, Turin: Marietti, 1950. (A helpful study-edition, to be complemented by the Leonine critical edition in preparation. H. Jarka-Sellers (trans.) *Commentary on the Blessed Dionysius' On the Divine Names, Thomas Aquinas in Translation*, Washington, DC: Catholic University Press, forthcoming, is an accessible translation with introduction and notes.)

Chevallier, P. (ed.) (1937-50) *Dionysiaca: Receuil donnant l'ensemble des traditions latines des ouvrages attribués au Denys de l'Aréopagite*, Bruges: Desclée de Brouwer, 2 vols; repr. Stuttgart, 1989.(A valuable scholarly accomplishment, comprising translations by Hilduin, Eriugena, John Sarracenus, Grosseteste, Ficino and others.)

Eriugena, Johannes Scottus (*c.*860-70) *Expositiones in Ierarchiam coelestem (Commentaries on Dionysius' De coelesti hierarchia)*, ed. J. Barbet, Corpus Christianorum, Continuatio Mediaevalis 31, Turnhout: Brepols, 1975.(Commentaries by the most creative philosopher of the early Middle Ages.)

Gallus, Thomas (1242) *Thomas Gallus: Grand commentaire sur la Théologie mystique (Commentary on Dionysius' De mystica theologia)*, ed. G. Thery, Paris: Haloua, 1934.(A philosophically interesting commentary by a little-known contemporary of Aquinas and Bonaventure.)

Hugh of St Victor (*c.*1220-30) *Commentariorum in Hierarchima coelestem Sancti Dionysii Areopagitae (Commentary on the De coelesti hierarchia of Dionysius the Areopagite)* Patrologia Cursus Completus, Series Latinus 175, Paris: Migne, 1844-80.(A reflection of an important philosophical school's indebtedness to Platonic sources.)

John of Scythopolis *et al.* (*c.*530) *Scholia*, ed. P. Corderius, Patrologia Cursus Completus, Series Graeca 4, Paris: Migne, 1857-91.(Annotations from Patristic period, whose philosophical importance is still being assessed.)

2 Secondary sources

Brons, B. (1976) *Gott und die Seienden. Untersuchungen zum Verhältnis von neuplatonischer Metaphysik und christlicher Tradition bei Dionysius Areopagita (God and Beings: An Examination of the Relationship between Neoplatonist Metaphysics and Christian Tradition in Dionysius the Areopagite)*, Göttingen: Vandenhoeck & Ruprecht.(A stimulating and systematic study of Pseudo-Dionysius' thought and a careful examination of the texts.)

Gersh, S. (1978) *From Iamblichus to Eriugena: An Investigation of the Prehistory and Evolution of the Pseudo-Dionysian Tradition*, Leiden: Brill.(A learned study elucidating Pseudo-Dionysius' place in the Platonist tradition.)

Koch, J. (1956-7) 'Augustinischer und Dionysischer Neuplatonismus und das Mittelalter' (Augustinian and Dionysian Neoplatonism in the Middle Ages), *Kantstudien* 48: 117-33; repr. in W. Beierwaltes, *Platonismus in der Philosophie des Mittelalters*, Darmstadt: Wissenschaftliche Buchgesellschaft, 1969, 317-42.(A seminal analysis of different strands of Platonist thought in the Middle Ages.)

O'Daly, G. (1981) 'Dionysius Areopagita', *Theologische Realenzyklopädie* 8: 772-80.(An informative, judicious overview.)

O'Rourke, F. (1992) *Pseudo-Dionysius and the Metaphysics of Aquinas*, Leiden: Brill. (A suggestive study of an important topic.)

Roques, R. (1954) *L'univers Dionysien. Structure hierarchique du monde selon le Pseudo-Denys (The Dionysian Universe: The Hierarchical Structure of the World According to Pseudo-Dionysius)*, Paris: Aubier; repr. Paris: Éditions du Cerf, 1983.(The classic systematic study.)

Roques, R. *et al.* (1954) 'Denys l'Areopagite (le Pseudo-)', *Dictionnaire de spiritualité ascétique et mystique doctrine et histoire*, vol. 3, 244-429.(A highly informative treatment by one of the leaders of Dionysian scholarship, followed by discussions by an array of well-known scholars of various aspects of Pseudo-Dionysius' influence.)

Rorem, P. (1993) *Pseudo-Dionysius: A Commentary on the Texts and an Introduction to their Influence*, Oxford: Oxford University Press.(The most accessible and informative general introduction to Pseudo-Dionysius in English, with considerable attention to theological dimensions of Pseudo-Dionysius' thought. Extensive bibliography.)

Saffrey, H.-D. (1982) 'New Objective Links between the Pseudo-Dionysius and Proclus', in D.J. O'Meara (ed.) *Neoplatonism and Christian Thought*, Norfolk, VA: International Society for Neoplatonic Studies, 65-74.(A methodologically astute inquiry into the philosophical background of Pseudo-Dionysius.)

Sheldon-Williams, I.P. (1970) 'The Pseudo-Dionysius', in A.H. Armstrong (ed.) *The Cambridge History of Later Greek and Early Medieval Philosophy*, Cambridge: Cambridge University Press. (A good introduction by an eminent scholar.)

Pseudo-Grosseteste (*fl. c.*1265-75)

'Pseudo-Grosseteste' is the name given to the unidentified author of a philosophic encyclopedia written in England in the third quarter of the thirteenth century. Like other encyclopedias of that time, the Summa philosophiae (Summa of Philosophy) marshals an astonishing array of facts, principles and arguments in every field of philosophic interest, from accounts of truth or cosmologies through analyses of rational and animal souls to the properties of minerals. The Summa is distinguished not by its range of interests, but by its clear sense of its own purposes. The author intends to arrange the history and the particular doctrines of philosophy within a hierarchy built according to Augustinian principles.

To Robert Grosseteste, as to many other ancient or medieval authors, medieval scribes attributed a number of works that modern scholarship now judges spurious. Among works falsely attributed to Grosseteste there are a complete exposition of the Gospels, fragments of a summa of theology and two large-scale summaries of philosophy, one entitled *Summa philosophiae (Summa of Philosophy)* and the other *Compendium philosophiae (Compendium of Philosophy)*. The *Compendium* was quickly returned to its real author, the twelfth-century polymath Dominicus Gundissalinus, but the *Summa* has found no other author, and so it is still assigned to a 'Pseudo-Grosseteste'. It is the author of this work, and only of this work, who is treated here.

The *Summa philosophiae* cannot be by Grosseteste because it refers to an event, the death of Simon de Montfort, that occurred twelve years after Grosseteste's own death. There is little other evidence to place it precisely. It seems on internal evidence to have been written during the third quarter of the thirteenth century, probably between 1265 and 1275. Most scholars find traces of English authorship. There are citations to Alfred of Shareshill and perhaps to John of Salisbury, if it is he who is named 'John the Peripatetic'. More significantly, the *Summa* shows striking affinities with the *Opus majus* of Roger Bacon, to whom it defers on some disputed points of physical doctrine. Curiously, the *Summa* does not mention Grosseteste himself. It should also be noticed that the genre of the naturalist's encyclopedia was quite popular with English authors in the thirteenth century, attracting such notables as Alexander Neckham, Bartholomew the Englishman and Thomas of York.

The *Summa* also invokes a large number of other authors, ancient, patristic, and medieval. The largest number of references is to Aristotle, followed at some distance by Plato, Averroes, Avicenna and Augustine, but a simple count of citations is misleading. The principal philosophical authority here is not Aristotle, but Augustine. On any number of disputed points, the *Summa* sides with Augustine against Aristotle or the Aristotelians. For example, the proofs for the existence of God are not proofs from motion or causality, but from the necessity for an eternal first truth. The pattern of argument here is to be found in Augustine's *Soliloquia (Soliloquies)* and *De libero arbitrio (On Free Choice of the Will)*. Again, in its division of matter the *Summa* allows for spiritual matter as an individuating principle in immaterial beings. At these and other points, the author of the *Summa* shows himself on the side of contemporary Augustinians. This is confirmed in the references to medieval authors. The *Summa* cites approvingly Alexander of Hales and Albert the Great, both of whom it includes in its history of 'philosophy'. Alexander is the great authority among the Franciscans, and he relies throughout his work on Augustine. Albert is taken by the *Summa* either as a teacher of natural philosophy or as an eclectic thinker open to Neoplatonic doctrines. There are allusions to the teachings of Thomas Aquinas, but they are hardly approving. Indeed, the author of the *Summa* singles out distinctively Thomist positions on several occasions for sharp criticism.

The *Summa* is divided into nineteen Books, a number also adopted by other encyclopedias of the thirteenth century. The Books may be arranged thematically into four groups: an introduction on the nature of philosophy and its relation to theology (Book 1); the principles of being and knowledge (Books 2-6); God and the creation in general (Books 7-9); and specific creatures, from the immaterial to the most material (Books 10-19). The last group is by far the longest and most detailed. It offers, for example, careful analyses of the types of soul, the operations of light, the properties of the elements, meteorological phenomena and minerals.

See also: Augustinianism; Grosseteste, R.; Platonism, medieval

MARK D. JORDAN

List of works

Pseudo-Grosseteste (1265-75) *Summa philosophiae (Summa of Philosophy)* ed. L. Baur, Beiträge zur Geschichte des Philosophie und Theologie des Mittelalters, Münster: Aschendorff, 275-643; Books 2-3 trans. R.P. McKeon, *Selections from the Medieval Philosophers*, New York: Scribner's, 1929-30, vol. 1: 290-314.(A careful edition with helpful if limited annotation.)

References and further reading

McKeon, C.G. (1948) *A Study of the Summa Philosophiae of the Pseudo-Grosseteste*, New York: Columbia University Press.(A detailed retelling of the *Summa*'s doctrine according to the order of the original, with frequent comparisons to other thirteenth-century authors.)

Psychē

Conventionally translated 'soul', psychē *is the standard word in classical Greek for the centre of an animal's, and especially a human being's, 'life'. In its earliest usage (in Homer)* psychē *is a breath-like material persisting after death as a mere ghost. Its precise reference to the locus of thought and emotion only began under the influence of philosophy. From the beginning of the fourth century* BC *it became normal to pair and contrast* psychē *with 'body' (*soma*). The term generated sophisticated discussions. Leading questions include: Is* psychē *immortal? Is it corporeal or incorporeal? What are its parts or functions?*

Before philosophy there was no single Greek term for the mind or for the body. In Homer *psychē* refers to the life or ghost of a person rather than to the mind as such. Homer uses other terms for the location of thought and emotion, probably because *psychē* was taken to comprise *all* the powers a living person loses at death: Homer's disembodied shades retain the characters and histories of their embodied lives. By 500 BC Pythagoras had become famous for rejecting the Homeric Hades and for teaching the transmigration of the same *psychē* into a succession of bodies, animal as well as human. Around the same time Heraclitus treated *psychē* not only as the source of life but also as something to which the predicates 'wisest and best' can apply. Gradually the term gained wider currency as a way of referring to the self or personality. Socrates invokes this usage in Plato's *Apology* when he exhorts the Athenians to subordinate concern for their bodies to the care of their *psychē*.

Sustained reflection on *psychē* began with Plato. In his *Phaedo*, although the *psychē* animates the body, that is not its best or primary function. Incorporeal and immortal, the *psychē* is unavoidably tainted by incarnation. The task of philosophers is to detach themselves from the distractions of the body in preparation at death for a return of the *psychē* to a discarnate state of purity and wisdom. In this dialogue sensual desires originate only in the body but a thoroughly impure *psychē* will identify with the body instead of valuing intellectual goods (see Plato §13). Plato subsequently modified this doctrine. In his *Republic* IV, and repeatedly thereafter, he divides the *psychē* into three components - rational, spirited or competitive, and appetitive; this last component (and not the body itself) is the drive for instant sensual gratification. Plato's principal ground for this tripartition is the need to explain how a person can be simultaneously subject to conflicting desires. On this model of the *psychē*, each component has its own objectives. In a well-integrated *psychē*, the subordinate components perform their proper functions (of facilitating a person's embodied life) under the rule of reason (see Plato §14).

Plato's moral psychology had a decisive influence on all the subsequent philosophical tradition. Later philosophers, however, paid closer attention to the biological functions of *psychē*. For Epicureans and Stoics, the *psychē* is a physical part of the living being, centred in the chest and distributed from there to the organs of the body (see Epicureanism §§12-13; Stoicism §3). For Aristotle (§§17-19), it is not a physical part but the 'form' or 'active functioning' of the body's organs, except in the case of intellect (see Nous) which has no bodily organ. Plato had located the rational component of *psychē* in the head, but most philosophers, mistaking the functions of the cardiovascular system, favoured the chest or heart as its location. These errors continued even after the nervous system was identified in the third century BC. The first theorist to relate psychology to the most up-to-date physiology was Galen (§2), writing in the second century AD.

In spite of these errors post-Platonic philosophers did work of lasting value in theorizing about *psychē*. Taking the human *psychē* to be an assemblage of vital functions, Aristotle analysed these, starting with the 'nutritive' (common to all living beings including plants) and ending with the 'intellectual' (peculiar to human beings). He is at his most effective, perhaps, in identifying the causal components of action: desire is stimulated by sense perception or 'imagination' (*phantasia*), and then intellect (in its practical function) works out the means to achieve the desired end (see Aristotle §§20, 23-4).

Epicurus' theory of *psychē* is grounded in his physics. Being an aggregate of fine and highly mobile atoms, the *psychē* depends upon embodiment in order to exist, and cannot survive the body's dissolution. Although the motions of atoms underlie all mental phenomena, Epicurus saw no difficulty in ascribing causal agency to thought and desire. According to Lucretius, Epicureans held that the random swerve of a mind's atoms frees it from mechanistic determinism.

The Stoic doctrine of *psychē* is also anchored to basic physical postulates. The 'active principle' of the Stoic universe is 'breath' (see Pneuma) permeating everything, and the 'breath' permeating animals and human beings is their *psychē*. Unlike Aristotle, the Stoics did not attribute *psychē* to plants. In their philosophy *psychē* is the life principle of animals (creatures capable of perception and purposive motion) and more particularly of human beings, who have the additional faculties of reason and 'assent' (saying yes or no to their occurrent thoughts and impressions). In contrast with other philosophers, the Stoics ascribed rationality to the human *psychē* in its entirety. But a rational *psychē* can act irrationally through error, and it can experience conflict of desires by oscillating between one judgment and another (see Stoicism §19).

In the Neoplatonism of Plotinus (§4), *psychē* is a principle that mediates between the higher reality of 'intelligible being' and the physical world ('nature'). This doctrine applies both to the *psychē* of individual beings and to that of the world. The world-soul generates the physical world by its apprehension of intelligible reality. The souls of human individuals, even when embodied, do not 'descend' completely to the physical world. They live at two levels, one contemplative and immortal, the other mortal and engaged with embodied life. Apart from these ideas Plotinus treated numerous mental phenomena with great subtlety in his long treatise, *Problems of the Soul (Enneads)* IV.

Although discussions of *psychē* centre primarily on the vital powers of human beings, Plato and the Stoics as well as Plotinus attributed *psychē* to the world at large. Aristotle did not adopt this view in *On the Soul*, but he agreed with them in making the heavenly bodies ensouled. Only Epicurus restricted the scope of *psychē* to terrestrial creatures.

See also: Alexander of Aphrodisias §2; Alcmaeon; Anaximenes §2; Democritus §3; Iamblichus; Neoplatonism §3; Neo-Pythagoreanism; Orphism; Platonism, Early and Middle §5; Plutarch of Chaeronea §4; Posidonius §5; Proclus §5; Soul, nature and immortality of the; Strato §5; Thales §2

A.A. LONG

References and further reading

Aristotle (*c*. mid 4th century BC) *On the Soul*, trans. H. Lawson-Tancred, Harmondsworth: Penguin, 1986.(Book I discusses and criticizes Aristotle's predecessors; Books II-III present his own doctrines.)

Claus, D.B. (1981) *Toward the Soul*, New Haven, CN: Yale University Press.(Careful study of the meaning of *psychē* before Plato.)

Long, A.A. (1973) 'Psychological Ideas in Antiquity', in P.P. Wiener (ed.) *Dictionary of the History of Ideas*, New York: Scribner's, vol. IV, 1-9.(Encyclopaedic survey.)

Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press.(Translation with commentary of Epicurean and Stoic texts in volume 1, 65-77, 313-22, 410-22.)

Plato (*c*.386-380 BC) *Phaedo*, trans. D. Gallop, Oxford: Clarendon Press, 1975.(Includes notes; Plato's classic dialogue on the soul's immortality.)

Plato (*c*.380-367 BC) *Republic*, trans. R. Waterfield, Oxford: Oxford University Press, 1993.(On the soul; see Books IV and VIII-X.)

Plato (*c*.366-360 BC) *Phaedrus*, trans. R. Hackforth, Cambridge: Cambridge University Press, 1952.(Includes commentary; the soul's nature and destiny are mythologically described at 245-56.)

Plato (*c*.366-360 BC) *Timaeus*, trans. H.D.P. Lee, Harmondsworth: Penguin, 1965.(Explains the tripartition of the soul from a largely physical viewpoint.)

Plotinus (*c*.250-66) *Problems of the Soul (Enneads)* IV; trans. A.H. Armstrong, *Plotinus*, vol. 4, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1984.(The entire book is an exploration of problems concerning *psychē*; Greek text with facing translation.)

Psychoanalysis, methodological issues in

Philosophers have subjected psychoanalysis to an unusual degree of methodological scrutiny for several interconnected reasons. Even a cursory look at the Freudian corpus reveals a slender base of evidence: eleven 'case histories', including those published jointly with Breuer. On the other hand, the theoretical claims have broad scope: all psychopathology can be traced to repressed sexuality. Further, Freud and his followers have disdained the most widely accepted means of establishing theories - experimental confirmation - while allowing themselves to appeal to such apparently dubious sources of support as dream interpretation, literature and everyday life. Together, these factors conjure a picture of a 'science ' with a large gap between theory and evidence that has not and cannot be filled by solid data.

The central methodological question about psychoanalysis is whether there is now or ever has been any evidence supporting its truth. Popper rejected psychoanalysis as a science on the grounds that there could be no possible evidence against it which could test its truth. More recently, Grünbaum has objected that there are serious logical difficulties with appealing to cures as evidence of truth. Grünbaum and others have attacked both the theory of dreams and the use of dream interpretation as evidence. Sulloway and Kitcher have argued that several tenets of psychoanalysis were supported by their nineteenth-century scientific context, particularly certain aspects of Darwinian biology, but that those crucial supports have been eroded by later scientific developments. Eysenck and Wilson have examined experimental results that have been offered in support of various claims of psychoanalysis and rejected them as inadequate to establish any specifically psychoanalytic claims. By contrast, Glymour (and others) have explained how even single case histories could provide evidence in favour of psychoanalysis. Other philosophers have argued that psychoanalysis is continuous with 'common-sense' psychology and is supported by the continual reaffirmation of the essential correctness of common-sense psychological prediction and explanation.

1 Popper and Grünbaum

Two influential critics of psychoanalytic methodology have been Sir Karl Popper and Adolf Grünbaum. Although Popper did not publish his objections to psychoanalysis until 1962 (see Popper 1963), they were well-known in philosophical and psychoanalytic circles for many years before that. Dissatisfied with attempts to distinguish sciences as 'speculative' versus 'observational', he proposed a different approach to the question of demarcating the 'scientific' from other types of theories. In his view, what distinguished theoretical systems, dividing them into the 'scientific' and the 'pseudoscientific', was their 'refutability' or 'falsifiability'. For any scientific theory, there must be some possible, but not actual, observation that would be inconsistent with the claims of that theory; if a 'theory' is compatible with any possible observation whatsoever, then that theory is mere 'pseudoscience'.

Along with Marxism, psychoanalysis was Popper's prime example of a theory that is immune to refutation by any possible observation. Despite the seeming inaccessibility of unconscious motivations, however, in succeeding years philosophers and psychologists pointed out that psychoanalysis is vulnerable to refutation and that Freud took account of this vulnerability. So Fancher (1973), among others, has noted that Freud modified his theories in light of disconfirming evidence, and Grünbaum has observed that many of Freud's claims about the involvement of repressed homosexuality in paranoia are eminently capable of refutation.

Grünbaum (1984) defended psychoanalysis against Popper only to offer his own alternative critique. From Freud onwards Grünbaum believes that the most important evidence in favour of the claims of psychoanalysis has come from alleged cures. For example, in 1930 the Berlin Psychoanalytic Institute issued a report on 721 analyses, 363 of which had been concluded by the time of the report. Of these, 111 patients were classified as cured, 89 as much improved, 116 as improved, and 47 as uncured (cited in Bergin and Lambert 1978). In Grünbaum's view, psychoanalysis presents a causal account of psychopathology: various symptoms are caused by repressed memories or desires; psychoanalytic treatment lifts these repressions, which removes the symptoms; hence, when patients are cured or improved, that testifies to the correctness of the causal hypotheses offered by the analyst.

The problems with treating therapeutic success as evidence for the truth of psychoanalytic claims are threefold. First, until recently, there have been no attempts to measure success rates objectively, by using observers who are unaware of the history of treatment. Second, any successes achieved must be compared with spontaneous 'cure'

rates, to show that treatment makes a difference. Finally, many different conditions obtain in therapy sessions, including support from the therapist, personal factors in treatment and, most problematically, suggestions about the causes of the patient's illness. Grünbaum argues that, since any one of these factors might be responsible for patient improvement, the inference from partial cure to the correctness of any psychoanalytic hypothesis is invalid. Of particular concern is the therapist's suggestion that the cause of the illness has now been found and hence improvement is to be expected. Grünbaum argues that the therapist's words can act as a placebo and lead to substantial improvement, simply because patients believe that they will improve.

2 Statistical and experimental evidence

In recent years, a growing number of 'outcome' studies have compared psychoanalysis with other forms of treatment, with no treatment, and with placebo conditions. Although the results are sometimes inconsistent, a review by Bergin and Lambert (1978) concludes that psychoanalysis and a number of other therapies appear to have better results than no treatment, placebos and pseudotherapies. The problem is that it is not clear what such studies prove. They appear to support the claims of psychoanalysis to therapeutic success. As Grünbaum notes, however, they simultaneously raise serious questions about the *causal claims* of psychoanalysis. For, if the basic outlines of psychoanalysis are correct, and psychopathology is a result of repressed (sexual) memories and desires, then how is it possible to account for the successes of non-analytic therapies (especially behavioural therapies), which also exceed spontaneous 'cure' rates? Since the patients have not had their repressions lifted, there should be no improvement. Other therapies might ameliorate psychological conditions without truly curing them, so the argument is not conclusive. Still, as Hume noted long ago, if apparently different causes have a common effect, it is reasonable to look for unnoticed common factors (see Causation). And Bergin and Lambert suggest that what accounts for improvement is not so much the truth of the theories underlying the different therapies, but the common placebo factors they share, such as the understanding and support of the therapist.

Comparing treatment outcomes with different therapies is an attempt to apply J.S. Mill's methods to causal hypotheses in psychotherapy. According to Mill, the way to establish that *A* (a particular treatment) is the cause of *B* (a cure) is to show that *A* reliably produces *B* in a variety of circumstances that differ in other respects and that *B* will not occur in the absence of *A* despite the presence of other circumstances (see Scientific method). Although recent studies are more sophisticated, the work just canvassed suggests what Mill fully anticipated: in complex cases, it may not be possible to disentangle all the possible causal factors and so apply his methods. Freud's defenders often object that methodological critiques of psychoanalysis suffer from exactly this flaw. Critics complain that psychoanalysis is unable to meet particular tests of scientific proof, when the real problem may be that the types of proof available are inadequate to handle the complexities encountered in psychopathology.

Similar problems have hindered attempts to find experimental evidence for or against central theories of psychoanalysis about, for example, the three phases of sexuality, the origin of the super-ego in the resolution of the Oedipal complex, or the sexual basis for the repression of unconscious ideas. Psychoanalytic claims about the causation of symptoms are complex, but only relatively simple factors can be observed or manipulated in experiments. The gap between 'experimental confirmation' and the theory to be confirmed is evident in Eysenck and Wilson's review of this literature (1973). One study was designed to test Freud's hypothesis of oral sexuality. Although the results confirmed thumb-sucking as a substitute form of oral activity, that claim is a rather bland surrogate for Freud's rich views about the oral phase of sexuality. Another study attempted to confirm Freud's claims about the Oedipal complex by looking at the proportion of male and female strangers in dreams. Again, the results were positive. More strangers in dreams were males than females and this was true to a higher degree in the dreams of males. As Eysenck and Wilson note, however, these results are compatible with many hypotheses besides Freud's very specific claim about fear of the father in response to incestuous wishes about the mother. To take an obvious example, men may dream more about men, because they are more often in the company of men; all subjects may dream more about men, because men are more salient in our society. More recent attempts to provide experimental support for psychoanalysis by using subliminal or supraliminal stimuli may well be subject to the same criticism. Establishing that there is unconscious, but active ideation in human brains is a pale substitute for Freud's rich claims about the sexual aetiology of repression (see Unconscious mental states).

3 Dream and symptom interpretation

The study just mentioned appeals only to the 'manifest' or actual, reported content of dreams. One recurring

methodological debate in psychoanalysis concerns the use of interpretations of dreams and symptoms in explaining psychopathology. Probably more than any other single issue, the question of dream and symptom interpretation reveals a huge gulf between the supporters and detractors of psychoanalysis. Dream and symptom interpretation are critical to psychoanalysis, because, along with ‘free association’, these are the primary methods by which diagnosis and remedy are possible. According to standard psychoanalytic theory, the true causes of psychopathology are hidden. They are repressed memories or desires. Hence, the analyst must find some indirect means of uncovering those causes, and dream and symptom interpretation, interpretations of particular actions or remarks, and free association are the distinctively psychoanalytic tools for accomplishing this task.

Critics have raised a host of objections. Grünbaum and others have argued that the theory that dreams are wish-fulfilments is false. Indeed, as Grünbaum observes, some of Freud’s own reports of his patients’ dreams cast serious doubts on this hypothesis. For example, a young woman dreamed that she was going on a summer holiday with her mother-in-law, something that she did not wish to do. Freud’s explanation, that she had this dream to fulfil her wish to disprove his theory of dreams as wish-fulfilments, seems to be blatantly *ad hoc*. The more standard objection to dream and symptom interpretation is that this is either an enterprise with no rules at all or with one question-begging rule: find some interpretation that relates the symptom or manifest content to a hidden sexual wish. Echoing Popper, critics object that since, given sufficient imagination, some connection to sexual wishes is always possible, the ability of therapists to make these connections has no probative value.

More recently, Fancher, Sulloway (1991) and Kitcher (1992) have argued that the theory of dreams is not an autonomous part of psychoanalysis, but follows from some of Freud’s early, but persisting views about the organization of the nervous system. If this view is correct, then it fundamentally alters the evidentiary status of the theory of dreams. Contrary to Freud’s own position and that of many of his followers, the theory would be capable of confirmation outside the clinical setting. Although it provides the theory of dreams with an alternative source of evidence, this new understanding of the place of the dream theory in psychoanalysis leads quickly to disaster. For, if the real source of support for the dream theory is Freudian neurology, then that theory should be rejected along with its outmoded neurophysiological basis.

4 The foundation in nineteenth-century science

As Quine has argued, theories can be supported both ‘from below’ (by experimental results) and ‘from above’ (by more general theoretical considerations). In considering experimental testing and therapeutic success rates - the experimental and therapeutic ‘predictions’ of psychoanalysis - we have considered only how the theory might be confirmed from below. Analysts have not tried to provide support from above for the obvious reason. Even if Freud’s views were well supported by nineteenth-century science, so much has changed that that support has long ago eroded. In fact, some analysts have taken the opposite tack and attempted to detach a ‘clinical’ part of psychoanalysis from its ‘metapsychological’ foundations, especially Freud’s claims about the flows of energy through the ‘mental apparatus’ (Klein 1976). One problem with this approach is that it makes the enterprise of psychoanalysis entirely dependent on ‘confirmation from the couch’, which is highly problematic (see §§2, 5). It also leaves psychoanalysis vulnerable to the charge that not only is it not well supported now, it never was a plausible scientific theory, but only some type of pseudoscience.

Frank Sulloway (1979) has argued that many of the basic tenets of Freud’s theory derived from important work in nineteenth-century biology. So, for example, the prominence of sexual factors in psychopathology and mental life generally would be a plausible hypothesis, given Darwin’s claims for the importance of sex in evolution. The view that psychopathology was a result of excess neural energy, or excitation, or libido - energy that was released in dreams, symptoms, jokes and slips - seems more plausible if the nervous system is understood as a reflex mechanism that seeks to rid itself of energy that has intruded into the system. Sulloway also provides great detail about the work in ‘sexology’ that preceded Freud, which was incorporated into some of the best-known claims of psychoanalysis about infantile sexuality and psychosexual development. Although Sulloway’s research deflates some claims for the originality of psychoanalysis, it provides a powerful reply to critics who dismiss the theory as wild speculation made up on the basis of a handful of not very successful case histories.

Looking back to nineteenth-century work in linguistics, anthropology and psychology also provides some grounding for methodological strategies of psychoanalysis that have been subject to withering criticisms. Several of Freud’s least popular interpretive principles were suggested by then contemporary work in linguistics. For

example, the claim that the manifest verbal content of dreams often represented the exact opposite latent content (for example, 'strength' represented 'weakness'), was consonant with Abel's hypothesis that early concepts were formed by comparison and so indicated both sides of the comparison, 'strength' and 'weakness'. The equation of the infantile with the primitive was firmly supported by anthropology and by some biologists, who maintained a 'biogenetic law' (ontogeny is the rapid recapitulation of phylogeny). Most importantly, perhaps, the method of free association itself reflects the widespread psychological assumption that different thoughts or ideas are held together by bonds of association and that it is possible to uncover these associations by asking subjects to provide lists of associates. Although much has changed in biology and the social sciences, in so far as psychology still accepts associationism, and evolutionary theory still speaks to the importance of sexual factors in human life, some parts of psychoanalysis can still receive some mild support from above (see Evolutionary theory and social science; Learning).

5 Single case histories as evidence

Although Freud offered many theoretical discussions of psychoanalysis, he often presented his ideas through the medium of the case history. Are these cases simply meant to illustrate the principles of psychoanalysis or can they be regarded as evidence in favour of one or another of its doctrines? Seemingly, case histories can provide evidence for psychoanalysis in one of two ways: a psychoanalytic claim would be confirmed either if the diagnosis produced a cure or if patients acknowledged that they did harbour the unconscious memory or desire inferred by the analyst. The problems with arguments from cures have already been noted. Patient confirmations of analysts' hypotheses are subject to some of the same difficulties, and more besides. Even more than in the case of cures, it seems that analysts' suggestions that patients have certain unconscious desires might lead patients to assume that they do have these desires, either because they wish to please the analyst or because they assume that analysts are knowledgeable in these matters. Further, a patient's confirmation of a particular unconscious desire is evidence of the analyst's correctness only if there is good reason to believe that patients, or the population as a whole, are generally reliable about the psychological states that cause their beliefs and actions. Recent work in psychology suggests that this crucial assumption may be false (see Nisbett and Wilson 1977; Introspection, psychology of).

Despite the serious difficulties with using single case histories as evidence for psychoanalysis, both Erwin (1988) and Glymour (1982) have argued that this is possible at least in theory. Erwin notes that a single case or intrasubject design can provide evidence, if the change in the patient is sufficiently dramatic. If a patient suddenly recovers from a physical illness after unsuccessful treatment by other medications, that is strong evidence of the efficacy of the new medicine. In the same way, if a patient who has been in different psychotherapies for some time with no improvement were suddenly to improve upon being offered a psychoanalytic hypothesis about the cause of the illness, that would suggest some causal connection. Although spontaneous remission cannot be ruled out, such a case would still have probative value.

Glymour has developed a very different approach to establishing the possible use of case histories as evidence. He notes that it is possible to test hypotheses in other non-experimental sciences, such as astronomy, by making inferences from observations via particular claims of the theory to further possible observations. The observations might provide evidence only when supported by other theoretical claims, but so long as these are independent of the claims to be tested, testing is still possible. This type of 'pincers' movement can be illustrated by the well-known Rat Man case. From the Rat Man's symptoms and various principles of psychoanalysis, Freud inferred that he had an unconscious wish for his father's death. But such a wish is unobservable. Given other claims of psychoanalysis, however, he inferred that this unconscious wish must have arisen from an early conflict over sexual matters. Hence he expected to find that the Rat Man had been punished for early sexual behaviour. Had this 'observation' been made, it would have offered some confirmation that the Rat Man wished for his father's death, because the evidence and assumptions supporting the inference to this wish would be independent of the evidence and assumptions that originally led Freud to infer such an unconscious wish. In opposition to Popper (see §1), Glymour notes that when Freud concluded that no episode of sexual misconduct occurred, he also recognized that his original theory was false. Although Popper might disapprove of how Freud then modified his theory, by allowing sexual fantasies as well as actual sexual encounters to be causes of psychoneuroses, from a logical point of view, the case is an example of negative evidence leading to the rejection of an early psychoanalytic hypothesis (see Scientific method).

The use of the Rat Man, the Wolf Man, and other well-known cases in presenting psychoanalysis has led Sulloway and others to investigate their accuracy. In a recent study, Sulloway argues that Freud's case histories were not particularly accurate representations of his case notes, and that in the case of the Rat Man and, especially, the Wolf Man, they exaggerated the degree to which a cure had been achieved.

6 Scientifically informed common sense

Richard Wollheim (1971), Donald Davidson (1982), Adam Morton (1982) and others suggest a different way of regarding Freud's theories, which provides a different source of possible evidence. On their view, psychoanalytic hypotheses about the causation of behaviour are of a piece with ordinary, common-sense explanations of behaviour (see Folk psychology). The standard view of 'common-sense' psychology, which dates back at least to Hume, is that actions are caused by the agent's beliefs and desires. Although this has sometimes been debated, the general view today is that, in discovering the beliefs and desires that led to a particular action, we are discovering both the cause of the action and the reasons for which it was done (see Reasons and causes). It is the latter feature that is distinctive of psychological explanation, for it permits actions to be understood in terms of their rational justification (see Explanation).

According to these philosophers, the best way to understand psychoanalysis is by seeing it as extending common-sense psychological explanation to the realm of psychopathology or the irrational. At first, this seems paradoxical, because it suggests that Freud and his followers are trying to provide rational explanations of the apparently irrational. Although this is true, it becomes more plausible if we consider exactly why certain actions are regarded by psychoanalysis as pathological. According to psychoanalysis, pathology results when the cause or causes of actions or dreams are unconscious wishes or desires. If a wish is unconscious, then it is beyond the realm of rational control. Since the unconscious tolerates contradictions, no presumptively relevant evidence about the inappropriateness of a wish can be brought to bear on it (see Unconscious mental states). So, if a young man's belief that his father despises him is unconscious, then it cannot be undermined by words or gestures indicating quite a different attitude on the part of the father. Even though the belief is irrational in this sense, however, it can still rationalize the man's behaviour by explaining why he takes certain defensive measures to ward off possible aggression from his father. Hence, although unconscious wishes and desires are irrational in the sense that they cannot be touched by reasons, they are rational in that they can provide reasons for action. If this is the correct understanding of psychoanalytic explanation, as an extension of common-sense psychological explanation, then it can be argued that the basic approach of psychoanalysis is supported by the continuing successes of common-sense psychology. Common-sense psychology can provide no support for any particular psychoanalytic claims, such as the homosexual aetiology of paranoia, but it can lend credence to the general enterprise of looking for conscious or unconscious beliefs and desires that will make the agent's actions comprehensible in terms of reasons. Conversely, psychoanalysis can make common sense more scientific, by delineating the true scope of rational explanation.

Supporting psychoanalysis by appeal to common-sense psychology is highly controversial, for several reasons. At one extreme, some philosophers will deny that common-sense psychology can provide useful support, because it is itself an unscientific attempt to understand human behaviour. In particular, critics note that it is far from clear exactly what counts as an adequate rational justification of an action (see Rationality, practical). Even those willing to allow the acceptability of common-sense psychological explanations in straightforward cases would object that complex cases, especially those involving pathology, are too tangled to permit any degree of confidence in a particular belief-desire explanation. Finally, assimilating psychoanalytic hypotheses to common-sense psychological explanations is possible only if the latent content of dreams, slips, actions and so forth can be fathomed by the interpretive methods of psychoanalysis and, as we have seen, many critics find these methods objectionable.

See also: Freud, S.; Psychoanalysis, post-Freudian; Unconscious mental states

PATRICIA KITCHER

References and further reading

Bergin, A.E. and Lambert, M.J. (1978) 'The Evaluation of Therapeutic Outcomes', in S.L. Garfield and A.E. Bergin (eds) *Handbook of Psychotherapy and Behavior Change*, New York: Wiley & Sons, 2nd edn.(An

important review of ‘cure’ rates.)
Davidson, D. (1982) ‘Paradoxes of Irrationality’, in R. Wollheim and J. Hopkins (eds) *Philosophical Essays on Freud*, Cambridge: Cambridge University Press.(Discussion of an important problem by a well-known philosopher.)
Erwin, E. (1988) ‘Psychoanalysis: Clinical versus Experimental Evidence’, in P. Clark and C. Wright (eds) *Mind, Psychoanalysis and Science*, Oxford: Blackwell.(A useful collection of essays by leading Freud scholars.)
Eysenck, H.J. and Wilson, G.D. (1973) *The Experimental Study of Freudian Theories*, London: Methuen.(Still the best review of attempts to verify psychoanalysis experimentally.)
Fancher, R.E. (1973) *Psychoanalytic Psychology*, New York: Norton.(A clear survey of Freud’s views.)
Glymour, C. (1982) ‘Freud, Kepler, and the Clinical Evidence’, in R. Wollheim and J. Hopkins (eds) *Philosophical Essays on Freud*, Cambridge: Cambridge University Press.(A classic, if somewhat difficult, essay on the methodology of testing.)
Grünbaum, A. (1984) *The Foundations of Psychoanalysis*, Los Angeles, CA: University of California Press.(One of the most important studies of Freud’s methodology.)
Grünbaum, A. (1993) *Validation in the Clinical Theory of Psychoanalysis*, Madison, CT: International Universities Press.(A collection of essays that extends many of the arguments of Grünbaum (1984).)
Kitcher, P. (1992) *Freud’s Dream: A Complete Interdisciplinary Theory of Mind*, Cambridge, MA: MIT Press. (Analysis of the foundations of Freud’s theories in a variety of nineteenth-century sciences.)
Klein, G. (1976) ‘Freud’s Two Theories of Sexuality’, in M. Gill and P. Holzman (eds) *Psychological Issues*, vol. 9, no. 4, *Psychology and Metapsychology*, New York: International Universities Press.(A classic discussion of Freud’s metapsychology.)
Morton, A. (1982) ‘Freudian Commonsense’, in R. Wollheim and J. Hopkins (eds) *Philosophical Essays on Freud*, Cambridge: Cambridge University Press.(A clear essay that argues for continuity between Freud’s ways of explaining behaviour and our everyday explanations.)
Nisbett, R. and Wilson, T. (1977) ‘Telling More Than We Can Know: Verbal Reports on Mental Processes’, *Psychological Review* 84 (3): 231-59.(An important discussion of the illusion that we know what is going on in our minds.)
Popper, K.R. (1963) *Conjectures and Refutations*, New York: Basic Books.(A classic book by one of the best-known philosophers of the twentieth century.)
Sulloway, F. (1979) *Freud: Biologist of the Mind*, New York: Basic Books.(A brilliant and detailed account of the biological foundations of psychoanalysis.)
Sulloway, F. (1991) ‘Reassessing Freud’s Case Histories’, *Isis* 82: 245-75.(A critical and well-documented look at Freud’s clinical ‘successes’.)
Wollheim, R. (1971) *Freud*, London: Fontana.(A clear introduction to the development of Freud’s views.)

Psychoanalysis, post-Freudian

The basic concepts of psychoanalysis are due to Sigmund Freud. After establishing psychoanalysis Freud worked in Vienna until he and other analysts fled the Nazi occupation. Post-Freudian psychoanalysis has evolved in distinct ways in different countries, often in response to influential analysts who settled there.

Freud's patients were mainly adults who suffered from neurotic rather than psychotic disturbances. He found their psychological difficulties to be rooted in conflict between love and hate, caused by very disparate, often fantastic, images deriving from the same parental figure. These images provided the basic representations of the self and others, formed by processes of projection (representing the other via images from the self) and introjection (representing the self via images from the other). The internalized image of a parent could be used to represent the self as related to some version of the other, as in the formation of the punitive super-ego, or as like the other, as in the identification with the parent of the same sex through which the Oedipus complex was dissolved.

Later analysts, including Anna Freud and Melanie Klein, observed that the uninhibited play of children could be seen to express fantasies involving such images, often with striking clarity. This made it possible to analyse children, and to see that their representations of the self were regularly coordinated with fantastic representations of others, with both organized into systematically interacting systems of good and bad. Emotional disturbance was marked by a fantasy world in which the self and idealized good figures engaged in conflict with hateful bad objects, unmitigated by any sense that all derived from the same self and parental figures.

Such observations made it possible to confirm, revise and extend Freud's theories. Klein saw that symptoms, character and personality could be understood in terms of relations to internalized fantasy-figures, laid down in early childhood; and this extended to psychotic disturbances, such as schizophrenia and manic-depressive illness, which turned on the particular nature of the figures involved. This gave rise to the British object-relations approach to psychoanalysis. It also influenced the development of ego-psychology and self-psychology by Hartmann, Kohut and others in the United States, and Lacan's attempt to relate psychoanalysis to language, in France.

1 Melanie Klein, object-relations and the British school

The first, major post-Freudian innovations, and the most controversial, stemmed from the work of Melanie Klein. Klein realized that psychic structure and function could be understood in terms of relations to fantastic unconscious figures, which she called 'internal objects', and focused psychoanalytic investigation on them. For example, a little girl who suffered obsessional symptoms as well as depression played at being a queen who was getting married. When she

> had celebrated her marriage to the king, she lay down on the sofa and wanted me, as the king, to lie down beside her. As I refused to do this, I had to sit on a little chair by her side, and knock at the sofa with my fist. This she called 'churning'…immediately after this she announced that a child was creeping out of her, and she represented the scene in a quite realistic way, writhing about and groaning. Her imaginary child then had to share its parents' bedroom and had to be a spectator of sexual intercourse between them. If it interrupted, it was beaten…. If she, as the mother, put the child to bed, it was only in order to get rid of it and to be able to be united with the father all the sooner.
>
> (1975, vol. 2: 40)

Freud had noted that the parents are frequently symbolized in adult dreams by the figures of king and queen. The play described above could be seen to use the same symbolism, and so to represent the child's unconscious fantasies about her parents (whom in real life she treated with excessive but demanding fondness). Here the child's representation of adult sexuality is partly symbolic (for example, in terms of 'churning' or something knocking something); but the referent of the symbolism can none the less be inferred from the context (the activity takes place after the wedding, with the parents lying together in bed, and is followed by the birth of a child).

The fantasies expressed in this game were repeated in many others, and in many related psychological constellations. Hence it was possible to understand the little girl's obsessions and depression by reference to what she felt to be happening between herself and the fantastic unconscious versions of her parents represented in her

play (see Unconscious mental states).

How was it possible to understand the establishment of the fantastic *bad* internal objects central to conflict and psychopathology? Klein found *envy* to be particularly important in this process. This can be acute in children and infants, since they are particularly small, helpless and dependent. According to Klein, little children therefore seek relief from envy, as well as other emotions felt as bad, by both projection and introjection. In the former they try to locate in others whatever is felt as bad in the self, including envy and other forms of aggression; and in the latter to locate in themselves whatever is felt as good in others. This produces both *splitting of the self* and *splitting of the object*, in which the former is represented as good but lacking aggression, and the latter as bad and lacking goodness, both unrealistically.

These processes explained the structural features of children's fantasies, in which 'good' figures, identified with the self and sharing its goals, are systematically menaced by 'bad' figures which oppose them. Despite being kept out of consciousness, or represented as distant or alien, the 'bad' figures threaten to confront the self, or impinge upon or invade it, with mirroring directness. This is because such figures contain split-off aspects of the self, and so relate to those from which they are split with uncanny precision. Thus the little girl above dealt with aggressive envy of her parents' sexuality by such projection, and so unconsciously represented them as having sexual relations to make her envious. Hence she felt herself powerless and surrounded by bad figures, and so sought to imagine herself in an idealized way, as a rich, powerful and enviable queen.

Since such projection creates images of the other unconsciously identified with disowned aspects of the self, Klein called it 'projective identification'. (Anna Freud described a comparable mechanism as 'identification with the aggressor'; Klein's description applies where the image of the aggressor is itself formed by projection.) This mechanism could be seen as basic to both individual and social thinking, ordering human relations on primitive patterns of idealized 'us' versus bad 'them'.

Klein also proposed far-reaching hypotheses about infancy. She held that object-relations had an innate basis, and began almost from birth. Thus the infant related to the mother and to parts of her, especially her breast, before coming to understand her as a single enduring object. The disparate images of 'good' and 'bad' objects shown in analysis were laid down before the infant formed an integrated conception of the human body, or realized that the parental figures it felt as very bad or very good were in fact the same (see Cognition, infant; Cognitive development; Nativism).

In the first three or four months of life, Klein hypothesized, the infant made sense of its experience by building up concepts of episodic objects, which were anatomically incomplete, modelled on various parts of the body, and liable to be extremely good or bad, depending on the nature of the experience in which they were involved. During the fourth month, however, the infant began to unify its conception of the mother, and so to regard her as a single person, who was anatomically whole, complete and enduring.

Recent experiments seem partly to illustrate these ideas. For example, Campos observed that a 4-month-old infant made angry by someone's taking something away will direct expressions of anger towards the depriving hand, whereas by 7 months the infant will direct anger at the offending agent's face (see Campos *et al.* 1983). Bower (1982) used mirrors to display to infants three simultaneous images of the mother. Infants of 3 or 4 months are undisturbed by the apparent presence of three 'mothers', and interact with them in turn, preferring each to (images of) women not their mothers. Past 5 months, however, the sight of three 'mothers' becomes intensely disturbing. It seems that the infant does not specifically prefer enduring singularity to episodic multiplicity until this time.

Klein thought that the formation of the image of the mother as psychologically and physically unified was of particular importance. She called the pre-objective phase of life the 'paranoid-schizoid position', the term 'paranoid' indicating the extremity of the baby's potential for anxiety, and 'schizoid' the fragmentary way it represents itself and its objects. In acknowledging the mother as a single enduring object, the infant must begin to endow her with complex and apparently contradictory characteristics shown over a variety of episodes. In particular, it must recognize that she is the frustrating object towards which it feels anger and hatred, as well as the gratifying object towards which it feels love. This more complex object must now also be seen as distinct from the self, capable of absence, unique and irreplaceable. (This may explain why infants who have not before protested at their mother's leaving may at 6 or 7 months develop intense anxiety at separation; according to Klein, the infant is

then consolidating the realization that the mother leaving is the only mother it has.)

In response to these changes the infant begins to dread the loss of the object, and becomes concerned to care for it, protect it and repair the harm inflicted on it in fantasy. This Klein called the ‘depressive position’, so named because unification entails liability to depression about damaging the object (principally the feeding, caring mother). The dialectic of projection and introjection correlates these more realistic and complex attitudes to the object with complementary changes in the self. This unification constitutes the emotional aspect of the establishing of the reality principle (see Freud, S. §7), and so coincides with a diminution in the role of wish-fulfilling fantasy, splitting and projection.

According to Klein a range of psychic disturbances can be understood in terms of failure to achieve this Kantian synthesis. The worst case is schizophrenic illness, in which the infant is liable to particularly violent splitting and projection, and so cannot, in Hume’s phrase, ‘unite the broken appearances’ of self and other (see Hume, D; Mind, bundle theory of). At deep levels the object remains represented as a multiplicity of episodic presences housing projected aspects of the self, which is therefore radically fragmented, depleted and threatened by persecution. In the less severe cases of manic and depressive illness, the object consists of partly integrated bad (hated) and good (idealized) parts. In mania the subject fantasizes triumphing over the bad and possessing the good of such an object, and in depression destroying the good together with the bad.

Klein’s ideas were extended by Bion (1961) to groups and by Segal (1990) and Bion (1989) to the infantile origins of symbolism and thought. Many analysts, however, have tended to accept and extend Klein’s descriptions of relations to real and fantasized objects, while disagreeing with her hypotheses about infancy. Winnicott (1958) formulated further striking conceptions, such as the ‘transitional object’ and the ‘true’ and ‘false’ self, and Fairbairn (1954) attempted a full-scale restatement of psychoanalysis in terms of object-relations. Such accounts have now been elaborated by all schools, including that of ego-psychology (see Kernberg 1995). An ethological account of early object-relations by Bowlby (1969, 1973) fostered extensive empirical study of attachment (Ainsworth 1985).

2 Ego- and self-psychology in the United States

Freud introduced the ego and the super-ego as functional systems mediating between the individual’s innate instinctual drives and the external world. This theory was elaborated by Anna Freud (1936) and developed more systematically by Heinz Hartmann and his colleagues in the United States.

Hartmann (1939) distinguished clearly between a person’s conscious and unconscious representations of the self, and the ego as a system whose working could be described in impersonal terms. This system developed out of an ‘undifferentiated matrix’ present at birth and prior to both ego and id, and could be seen as having the general function of *adaptation*, that is, of relating the organism to the physical and social environment. Hence study of the ego could form a bridge between biology and the social sciences, rendering psychoanalysis a more general psychology.

Hartmann held that a main achievement of the ego was the attainment of*autonomy*, that is, the capacity to function with a degree of independence and self-sufficiency; and he linked this with normally conflict-free functions such as locomotion, which had received little attention in psychoanalytic theory. He related autonomy in object-relations to ‘object constancy’, the ability to represent self and other consistently, despite absence and changes in emotion (see Cognition, infant).

This notion was carried into empirical research by Renée Spitz (1965), who made a number of pioneering efforts to study ego-development in infancy, including film studies of infants and mothers, focusing on smiling, anxiety at separation or encountering strangers, grief at loss of contact, and the like. Spitz thought that the development of object-relations could be detected in a number of features of overt infantile behaviour: the way the infant related the mother’s breast and face, subsequently following her face with its eyes; the smiling response, followed by social smiling, which he took to indicate the inception of social relations; the onset of 6- to 8-month ‘stranger anxiety’, which indicated recognition of the mother as a specific individual; and the capacity to gesture (or verbalize) ‘No!’, which he took to show both identification with the mother as one who refuses, and a distinctive form of communication.

More detailed hypotheses about the development of the ego and object constancy were set out in work by Edith Jackobson (1964) and by Margaret Mahler (1968) and her associates. Mahler held with Spitz that in early life the mother serves as an 'external ego' for the infant; she argued that the infant first experiences the mother in a phase of *symbiosis*, during which it feels itself blissfully to merge with her, and imagines the two as a 'dual unity' with a common boundary, or what Jackobson called a 'self-object'. As the infant forms a more distinct image of the mother it develops a sense of basic trust, conveyed in smiling and eye contact; then, beginning to distinguish more clearly between itself and the mother, it enters the phase of 'separation-individuation', not completed until the third year. In this phase the infant builds up representations of 'good' and 'bad' self and objects, and finally unifies these into more stable and realistic representations, in which the characteristics of the self are fully distinguished from those of its objects.

Mahler divided this process into three sub-phases, marked by discernible patterns of infantile behaviour. The first is that of *differentiation*, lasting (roughly) from 5 to 9 months, during which the infant 'hatches' from the dual unity and focuses perception and action on the world. This is followed by a period of *practising*, in which the infant increasingly turns away from the mother to explore the environment, first by crawling and then by walking. As the child's enjoyment of its own capacities conflicts with its increasingly realistic sense of dependence on the mother there follows a phase of *rapprochement*, during which the child seeks to reconcile desires to remain in union with the mother with desires to become a separate and autonomous individual.

This conflict can show in a number of ways: new anxieties about separation; alternating 'shadowing' the mother, allowing no separation, and darting away, requiring her to prevent it; ignoring the mother's presence or departure, to deny her importance; 'becoming' the mother in her absence, and so creating a pseudo-independent false self; attempting to control or coerce the mother, to deny her autonomy; and so forth. In the ensuing 'rapprochement crisis' the child feels both longing and anger towards the absent mother, but directs the longing to an idealized 'good mother', while projecting the image of the 'bad mother' onto some resented substitute. The resolution of this phase, similar in structure and outcome to Klein's depressive position, is marked by diminution in splitting of this kind, so that the child approaches object constancy (see Cognitive development).

Much of the clinical work of ego-psychologists was concerned with feelings about the self linked to failures in individuation and autonomy. This set the stage for Heinz Kohut (1977) to argue that the development of a new psychoanalytic theory, 'self-psychology', was required to explain the pathology of the 'fragmented' and 'depleted' self. In attempting to frame such a theory Kohut re-employed Jackobson's term 'self-object', to designate another who is *experienced* as performing a psychological function essential to the self, and so as part of the self. This restored the phenomenological aspect of Freud's notion of an ego partly constituted by relations to its objects, lost in Hartmann's disambiguation.

In early life the primary self-objects are the parents - for example, the mother in her role as 'eternal ego' to the infant - and later versions inherit their significance from these. The 'primary psychological configuration' for self-psychology is thus the experience of the relation between the self and the self-object, and in particular that of the 'empathy', or degree of attuned adequacy, with which the self-object responds to the needs of the self. When the self-object responds appropriately to expressions of need, the infant's anxiety is replaced by an experience of merging with a powerful and calming self-object, whose functions, given repeated experience, can ultimately be internalized.

Kohut argued that two basic self-object functions show themselves in particular forms of transference to the analyst. There is an 'idealizing transference', in which the analyst is represented as embodying the kind of perfection with which the patient wishes to merge; and a 'mirror transference' in which the analyst is represented as having the function of responding empathically to the patient's narcissistic and grandiose displays, so as to sustain the patient's self-assertion through approval, mirroring and echoing. These transferences are best explained on the hypothesis that it is especially important for the infant's primary self-objects to provide both empathic mirroring for the infant's self-assertive, exhibitionistic and grandiose self, and embodiments of idealizable functioning with whom the infant can feel it merges. If these requirements are met, then the infant's grandiose self-assertions can develop towards realistic ambitions and goals, and its notions of ideal perfection towards realistic ideals and values.

Since the cohesion of the self depends upon these functions, the child (or the analytic patient in whom such

infantile needs are re-activated) responds to the parent's (or analyst's) failures in performing them with an overwhelming hatred and desire to hurt, which Kohut calls 'narcissistic rage'. The infant seeks total control over the self-object's responses, which it cannot achieve. If it remains without adequate empathic mirroring or opportunity for idealization and merging, it may become convinced that the environment is hostile, a situation Kohut compares to Klein's paranoid-schizoid position.

3 Psychoanalysis and language in France

A main tenet of all psychoanalytic approaches is that desires from infancy are constantly re-symbolized, so as to secure representational pacification throughout life. Analysis seeks to trace the constantly evolving modes of representation of the self and the objects of its desires; and this is done through language, a particular form of symbolic representation. In seminars from 1953 to the 1970s, Jacques Lacan attempted a re-presentation of Freudian theory centred on symbolism and the relation of psychoanalysis to language.

A main theme in Lacan's work, like that of Klein and Kohut, is a lack of unity in the self, which Lacan takes as an ineliminable consequence of the processes that provide representation for the self and the objects of desire (see Alterity and identity, postmodern theories of §2). As a first example, he considers the infant's recognition of its own image in a mirror. The infant joyfully assumes this external image, which represents it as possessing a wholeness, permanence and unity which it feels itself to lack, and which anticipates and facilitates its ability to move and relate to others. Paradoxically, however, this enabling identification is also an alienation; for the infant identifies itself with something it feels itself *not actually to be* and which it *may yet fail to become*. Identifications with others are alienating in the same way, so that the self is constituted by images that threaten to confront it as reminders of its own lack of being. The subject is thus formed in a potentially rivalrous and aggressive relation with its own self, and thence with others.

Lacan describes these alienating self-images, as well as others created in projection, introjection and fantasy generally, as constituting an order of representations which he terms 'the imaginary'. This he contrasts with the 'symbolic order', which encompasses the personal and social systems of signs whose elements are intrinsically constrained by rules of combination and substitution like those of natural language (see Syntax). Freud's discussions of dreams and symptoms, he argues, shows them to be constructed in accordance with such rules; so, the combinations and substitutions of (representations of) objects in dreams and symptoms, or again in the course of development, can be seen as instances of metaphor, metonymy and other linguistic forms. The unconscious is thus structured like a language, and representations of the subject and objects of desire are products of symbolic substitution, and shaped to fit the elements and rules of the combinatory systems through which they are articulated and satisfied.

This account extends to society, for Lacan regards structures of kinship and social exchange as further symbolic systems. In particular, the prohibition of incest, characteristic of all human societies, can be seen as a law regulating the relations (combinations) of men and women, taken as elements partly constituted by their role in exchanges among groups which compose society. Thus, according to Lacan, the resolution of the Oedipus complex is a development in which the boy forgoes an imaginary relation with the mother to occupy a place in a larger order, which is symbolic, social and constitutive of human culture. As in the prior instance of the mirror, the child secures a potentially fulfilling identity via the enabling but alienating assumption of an image - in this case that of the symbolic father, who embodies the social laws which regulate sexual desire and provide for its procreative satisfaction.

4 Comparisons

The diversity of psychoanalytic theories may conceal their underlying convergence. For example, Lacanian criticisms of ego-psychology assume a common basic account of the formation of the self, and an interest in symbol-formation is common to all schools. Again, Lacan's mirror stage illustrates a particular aspect of Klein's depressive position - namely, identification with something felt to be whole because seen as having spatial unity and permanence, and enviable on that account.

The main disagreements among the other approaches sketched above concern the first year of life. Klein hypothesized that envious projective identification was active from early infancy, whereas Kohut and the

ego-psychologists regarded symbiotic unity as prior to envy and aggression, ego-weakening splitting and fantasies of relating to good and bad objects generally. Studies of infant cognition, imitation and communication (see Stern 1985; Mehler and Dupoux 1994) suggest that even very young babies might engage in psychological exchanges comparable to projective identification, which can also be seen as the establishing of object- and self-object relations (see Cognition, infant; Mind, child's theory of). In addition, empirical studies of attachment suggest that patterns of object-relations formed in infancy can be traced throughout life, and from one generation to the next (see Bretherton and Waters 1985; Cichetti and Cummings 1990). Some of these patterns coincide with psychoanalytic accounts, and recent work in the United States (Kumin 1996) relates them to both Klein and Kohut. Whether a conceptual and empirical synthesis along these lines will prove successful remains to be seen.

See also: Feminism and psychoanalysis; Freud, S.; Psychoanalysis, methodological issues in

JAMES HOPKINS

References and further reading

Ainsworth, M.D.S. (1985) 'I. Patterns of Infant-Mother Attachment: Antecedents and Effects on Development' and 'II. Attachment across the Life Span', *Bulletin of the New York Academy of Medicine* 6: 771-812.(Use of significant research paradigm for studying relationships between parents and infants.)

Ainsworth, M.D.S., Blehar, M.C., Waters, E. and Wall, S. (1978a) *Patterns of Attachment: A Psychological Study of the Strange Situation*, Hillsdale, NJ: Erlbaum.(Exposition and use of significant research paradigm for studying relationships between parents and infants. See also Ainsworth *et al*. (1978b).)

Ainsworth, M.D.S., Blehar, M.C., Waters, E. and Wall, S. (1978b) *Patterns of Attachment: Assessed in the Strange Situation and at Home*, Hillsdale, NJ: Erlbaum.

Anderson, R. (ed.) (1992) *Clinical Lectures on Klein and Bion*, London and New York: Tavistock/Routledge. (Essays introducing and discussing Bion's work.)

Benvenuto, B. and Kennedy, R. (1986) *The Works of Jacques Lacan: An Introduction*, London: Free Associations Books.(Clear introductory discussion of Lacan.)

Bion, W.R. (1961) *Experiences in Groups*, London: Tavistock and New York: Basic Books.(Extension of Klein's concepts to the mental functioning of groups.)

Bion, W.R. (1977) *Seven Servants, Four Works by Wilfred Bion* incorporating *Learning from Experience*, *Elements of Psycho-analysis*, *Transformations* and *Attention and Interpretation*, New York: Jason Aaronson.(A collection of books by Bion, also published separately by Maresfield Reprints, London: Karnac Books, in 1984. These are less easy to approach than the articles in Bion (1989).)

Bion, W.R. (1989) *Second Thoughts: Selected Papers on Psychoanalysis*, London: Heinemann.(Advances in Kleinian thinking on symbolism and mental function, partly produced through work with schizophrenic patients. Emphasis on the mother's mental containment of the infant's projective identifications, and the infant's introjection of this function. Bion's work is discussed in Anderson (1992) and compared with Kohut's in Kumin (1996); Segal links her work on symbolism with Bion's on mental functioning in Segal (1990).)

Blank, G. and Blank, R. (1974) *Ego-Psychology: Theory and Practice*, New York: Columbia University Press. (Survey of ego-psychology.)

Bower, T.R.G. (1982) *Development in Infancy*, San Francisco, CA: W.H. Freeman.(Experiments and discussions concerned with the abilities of infants; on this topic see also Mehler and Dupoux (1994).)

Bowlby, J. (1969, 1973, 1980) *Attachment and Loss*, vol. 1, *Attachment*, London: Hogarth Press; vol. 2, *Separation*, New York: Basic Books; vol. 3, *Loss*, New York: Basic Books.(Series of books on early relation between infant and caretakers influenced by ethology and British object-relations theorists; gave rise to the empirical study of attachment by Ainsworth (1985) and others. See also Bretherton and Waters (1985); Cichetti and Cummings (1990).)

Bretherton, I. and Waters, E. (1985) *Growing Points in Attachment Theory and Research*, Monographs of the Society for Research in Child Development no. 209, Chicago, IL: University of Chicago Press.(Empirical study of early object-relations.)

Campos, J.J., Barrett, K.C., Lamb, M.E., Goldsmith, H.H. and Stenberg, C. (1983) 'Socioemotional Development', in Mussen (ed.) *Handbook of Child Psychology*, vol. 3, New York: Wiley & Sons.(Discussion of emotion and object-relations in infancy.)

Cichetti, D. and Cummings, E.M. (eds) (1990) *Attachment in the Preschool Years: Theory, Research, and Intervention*, Chicago, IL: University of Chicago Press.(Empirical study of early object-relations.)

Fairbairn, W.R.D. (1954) *An Object-Relations Theory of the Personality*, New York: Basic Books.(Attempted reformulation of Freudian theory in terms of object-relations.)

Feldstein, R., Fink, B. and Jaanus, M. (1996a) *Reading Seminars I and II: Lacan's Return to Freud*, Albany, NY: State University of New York Press.(Essays discussing concepts introduced in Lacan's seminars.)

Feldstein, R., Fink, B. and Jaanus, M. (1996b) *Reading Seminar XI, Lacan's Four Fundamental Concepts of Psychoanalysis*, Albany, NY: State University of New York Press.(Essays discussing concepts introduced in Lacan's seminars.)

Freud, A. (1936) *The Ego and the Mechanisms of Defense*, London: Hogarth Press; repr. in *The Writings of Anna Freud*, New York: International Universities Press, 1966.(Anna Freud's account of her father's late work which provided the starting point of ego-psychology.)

Freud, A. (1946) *The Psycho-Analytical Treatment of Children*, London: Imago Press; repr. in *The Writings of Anna Freud*, New York: International Universities Press, 1966.(Anna Freud's views on the analysis of children. Glenn (1992) provides a survey of work on child analysis influenced by both Klein and Freud.)

Freud, A. (1966) *The Writings of Anna Freud*, New York: International Universities Press.(Collected works of Freud's daughter, who influenced both child analysis and ego-psychology.)

Glenn, J. (ed.) (1992) *Child Analysis and Therapy*, New York: Jason Aaronson.(Survey of work in child analysis from the perspective of ego-psychology.)

Greenspan, S. and Pollock, G. (eds) (1989) *The Course of Life*, New York: International Universities Press. (Includes discussions of attachment at different stages of life.)

Hartmann, H. (1939) *Ego Psychology and the Problem of Adaptation*, New York: International Universities Press, 1958.(Influential attempt to set ego-psychology in unified biological and social context.)

Hartmann, H. (1964) *Essays in Ego Psychology*, New York: International Universities Press.(Influential essays by principal exponent of ego-psychology.)

Hinshelwood, R.D. (1990) *A Dictionary of Kleinian Thought*, London: Free Associations Books.(Explanations of terms and concepts in Kleinian theory.)

Hinshelwood, R.D. (1994) *Clinical Klein*, London: Free Associations Books.(Clinical illustrations of basic Kleinian concepts.)

Hopkins, J. (1987) 'Synthesis in the Imagination: Psychoanalysis, Infantile Experience, and the Concept of an Object', in J. Russell (ed.) *Philosophical Essays in Developmental Psychology*, Oxford: Blackwell, 1987. (Relates empirical work, particularly Bower (1982) to Klein and Piaget on the concept of the object in infancy.)

Jackobson, E. (1964) *The Self and the Object World*, New York: International Universities Press.(Influential account of the self in ego-psychological terms.)

Klein, M. (1975) *The Writings of Melanie Klein*, vol. 1, *Love, Guilt and Reparation and Other Works 1921-45*; vol. 2, *The Psychoanalysis of Children*; vol. 3, *Envy and Gratitude and Other Works 1946-63*; vol. 4, *Narrative of a Child Analysis*, London: Karnac Books and the Institute of Psychoanalysis; repr. New York: Delta Press, 1977; repr. London: Virago, 1989.(Volume 1 includes essays on the analysis of children, including initial descriptions of the depressive position, and Klein's seminal paper on symbolism, derived from the analysis of an autistic boy (discussed in Segal 1986; Lacan 1988). The first essays in this volume are both the easiest to read, and essential for understanding the rest. Volume 2 includes Klein's most comprehensive description of her technique and views. Volume 3 has later essays elaborating the paranoid-schizoid position and the role of envy and projective identification. Volume 4 gives a session-by-session account of an analysis of a 10-year-old boy.).

Kernberg, O. (1995) 'Psychoanalytic Object Relations Theories', in B. Moore and B. Fine (eds) *Psychoanalysis: The Major Concepts*, New Haven, CT and London: Yale University Press.(Clear, accurate and comprehensive overview of object-relations approaches of various schools; includes bibliography.)

Kohut, H. (1971) *The Analysis of the Self*, New York: International Universities Press.(Essays written during the initial framing of self-psychology.)

Kohut, H. (1977) *The Restoration of the Self*, New York: International Universities Press.(First statement of self-psychology, relating his new ideas clearly to the ego-psychology he is revising.)

Kohut, H. (1978) *The Search for the Self*, ed. P. Ornstein, New York: International Universities Press.

Kohut, H. (1984) *How Does Analysis Cure?*, ed. A. Goldberg and P. Stepansky, Chicago, IL: University of Chicago Press.(Kohut's last exposition of self-psychology: fuller but less explicitly related to other schools than Kohut (1977).)

Kumin, I. (1996) *Pre-Object Relatedness: Early Attachment and the Psychoanalytic Situation*, New York: Guilford Press.(Attempt to integrate Kleinian, Kohutian and ego-psychological thinking about early object-relations with infant research and attachment theory. The synthesis may be premature, but illustrates genuine connections among the basic ideas.)

Lacan, J. (1977a) *Ecrits*, London: Tavistock and New York: Norton.(Partial translation of the French edition of Lacan's collected essays. Deliberately difficult, and only partly understood by many enthusiastic commentators. Best approached together with a guide such as Benvenuto and Kennedy (1986) and via the texts of the *Seminars*, in which many points are discussed more fully.)

Lacan, J. (1977b) *The Four Fundamental Concepts of Psychoanalysis*, London: Hogarth Press and the Institute of Psycho-analysis; repr. New York and Harmondsworth: Peregrine Books, 1986. (Can be read with Feldstein, Find and Jaanus (1996b).)

Lacan, J. (1988) *The Seminars of Jacques Lacan*, book 1, *Freud's Papers on Technique, 1953-1954*, Cambridge: Cambridge University Press.(First volume of Lacan's seminars, notable for the way he begins by contrasting Anna Freud and ego-psychology with Melanie Klein, and uses Klein's work on symbolism and containment, later to be articulated by Bion, as the basis for his discussion of the opposition between the imaginary and the symbolic. This can be read with Feldstein, Fink and Jaanus (1996).)

Lacan, J. (1988b) *The Seminars of Jacques Lacan*, book 2, *The Ego in Freud's Theory and in the Technique of Psychoanalysis 1954-1955*, Cambridge: Cambridge University Press.(Contains an instructive treatment of the dream of Irma's injection, discussed in Freud §4, which includes an account of the mirror stage, Lacan's view of the ego, and the imaginary and the symbolic.)

Lacan, J. (1992) *The Seminars of Jacques Lacan*, book 7, *The Ethics of Psychoanalysis 1959-1960*, London: Routledge.(Includes discussion of the moral values of psychoanalysis.)

Lacan, J. (1993) *The Seminar of Jaques Lacan*, book 3, *The Psychoses 1955-1956*, London: Routledge. (Discussion of psychosis and the case of Schreber related to metapor and metonymy.)

Laplanche, J. and Pontalis, J.B. (1973) *The Language of Psycho-Analysis*, London: Hogarth Press.(Clear and scholarly exposition of basic psychoanalytic concepts, including some of Lacan's.)

Mahler, M.S. (1968) *On Human Symbiosis and the Vicissitudes of Individuation*, New York: International Universities Press.(Mahler's account of early development; includes interesting clinical material.)

Mahler, M.S., Pine, F. and Bergman, A. (1975) *The Psychological Birth of the Human Infant*, New York: Basic Books.(Study of children's behaviour via Mahler's approach; some patterns observed are closely related to those studied via attachment theory, as in Bretherton and Waters (1985); Cichetti and Cummings (1990).)

Mehler, J. and Dupoux, E. (1994) *What Infants Know: New Cognitive Science of Early Development*, Oxford: Blackwell.(Review of recent work on infant cognition.)

Moore, B. and Fine, B. (1990) *Psychoanalytic Terms and Concepts*, New Haven, CT and London: Yale University Press.(Short introductory discussions of main terms and concepts; briefer and clearer than Moore and Fine (1995). Includes discussions of Klein, Bion, Winnicott, Fairbairn and Kohut ('Self-Psychology'), all from the perspective of ego-psychology. Can be used together with Laplanche and Pontalis (1973) which is more rooted in Freud.)

Moore, B. and Fine, B. (1995) *Psychoanalysis: The Major Concepts*, New Haven, CT and London: Yale University Press.(Introductions by various authors to important concepts, from ego-psychological perspective.)

Mussen, P. (ed.) (1983) *Handbook of Child Psychology*, vol. 3, New York: Wiley & Sons.(Includes Campos *et al.* (1983).)

Petot, J.-M. (1991) *Melanie Klein*, vol. 1, *First Discoveries and First System 1919-32*, vol. 2, *The Ego and the Good Object*, Madison, CT: International Universities Press.(The most detailed and scholarly study of Klein's ideas.)

Russell, J. (ed.) (1987) *Philosophical Essays in Developmental Psychology*, Oxford: Blackwell, 1987. (Philosophical discussions of developmental psychology.)

Schneiderman, S. (1980) *How Lacan's Ideas Are Used in Clinical Practice*, New York: Jason Aaronson.(Clinical essays by Lacanian writers.)

Segal, H. (1986) *The Work of Hanna Segal: A Kleinian Approach to Clinical Practice*, London: Free Associations Books.(Exposition and development of Kleinian ideas, including influential discussion of concrete symbolism and symbol-formation, which has proven important in the analysis of psychotic states.)

Segal, H. (1989) *Klein*, London: Fontana Modern Masters.(Lucid introduction to Klein's work.)

Segal, H. (1990) *Dream, Phantasy, and Art*, London and New York: Tavistock/Routledge.(Discussion of fantasy and art which includes clear exposition of Freud on dreams and Segal's own work on dreams and symbolism, relating them to Bion's work on mental functioning. Includes a discussion of a recent analysis of an autistic child.)

Spillius, E.B. (1988) *Melanie Klein Today*, vol. 1, *Mainly Theory*, vol 2, *Mainly Practice*, London: Routledge. (Essays describing developments in Kleinian theory and practice.)

Spitz, R. (1965) *The First Year of Life: A Psychoanalytic Study of Normal and Deviant Development of Object*

Relations, New York: International Universities Press.(Pioneering ego-psychological discussion of early development; some claims require revision in light of more recent research, as in Mehler and Dupoux (1994).)

Stern, D. (1985) *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*, New York: Basic Books.(Discussion of the infant’s capacities to relate to others; attempts to integrate psychoanalytic theory with recent experimental work on infant cognition.)

Tolpin, M. and Kohut, H. (1980) ’The Disorders of the Self’, in S. Greenspan and G. Pollock (eds) *The Course of Life*, New York: International Universities Press, 1989.(Good introduction to Kohut’s understanding of disorder and the self.)

Winnicott, D. (1958) *Collected Papers: Through Pediatrics to Psychoanalysis*, New York: Basic Books; repr. London: Karnac Books and the Institute of Psychoanalysis, 1992.(Essays elaborating such influential notions as the transitional object, the value of illusion, and the true and false selves.)

Psychology, theories of

The object of study in psychology is the experience and behaviour of organisms, particularly human organisms. Psychology resembles the other sciences in employing methods appropriate to material phenomena but, unlike them, the mind (sometimes held to be immaterial) is among its objects of study. If the mind were immaterial it is difficult to see how psychology could proceed. Psychology differs from the other sciences in that understanding may not permit prediction and/or control of the phenomena. This is a consequence of the fact that human organisms may become aware of causal factors that would otherwise have determined their experience or behaviour. Being aware, they have a choice. (This is not, of course, to deny that the choice may itself be determined by more remote factors.)

In the seventeenth century, Descartes proposed that the mind (or soul), though immaterial, was nevertheless capable of two-way causal interaction with the material body. It is often held that what is immaterial cannot interact with what is material but Descartes' proposal was, in its context, so valuable that it has ever since exercised a profound influence on philosophy and on what most of us take for granted (for example, the unexamined belief that the mind/soul is, in some sense, immaterial but does, in some sense, interact with the body).

The possibility of scientific psychology was preserved by Leibniz's hypothesis of psycho-physical parallelism (according to which the deity keeps mind and body running in parallel, though there is no causal interaction between them). This view did not survive the passing of an era of universal (and largely unquestioned) religious belief. At the beginning of the twentieth century behaviourism tried to do without mind, it also tried to avoid surreptitious appeal to the lay person's unexamined 'mind' concept. In the 1960s, the cognitive revolution in psychology substantially broadened the range of phenomena under investigation but at the cost of allowing appeals to an unexamined 'mind of last resort'.

The headlong advance of the computer, which has taken over many functions previously reserved to members of the human species, leaves open the question whether it will, one day, take over all of them - or will be permitted to do so.

1 Origins

The belief is deep-seated that human persons are, by virtue of possessing conscious minds, beings of a different order from the bodies with which they are associated - which share material reality with the other physical objects in our universe. The belief is not unreasonable since each of us can validate it, so we think, from immediate experience. Until the seventeenth century it was generally agreed that the 'mind/soul' constituting the person was a substance resembling the body but with the matter taken out. Christian dogma held that, on judgment day, all souls would resume existence as physical bodies. In the seventeenth century, Descartes first proposed the more etiolated conception now current. The mind/soul remains immaterial (thus suffering none of the disabilities attached to material existence) but is no longer supposed to have the external form of the body. It is, however, still capable of two-way, causal interaction with the body. Descartes' proposal was, historically, of very great significance since it proved acceptable to the Church and allowed a distinction to be drawn between knowledge related to the soul (the prerogative of the Church) and knowledge related to the body (including the human body) which could freely be advanced by scientific investigation. Prior to Descartes no such distinction had been drawn and a scientist might be guilty of heresy for holding opinions that appear to us to have no religious content at all (for example, Galileo and the moons of Jupiter).

Descartes was aware of the physical existence and functions of the sense organs but, in his view, their purpose was to model objects in the real world for presentation to the mind/soul via the pineal gland. Collins (1973) draws attention to this radical change of emphasis. The old philosophy held, with Descartes, that objects in the real world cause changes in the perceiving organism. But what was perceived (according to the old philosophy) were objects in the real world. What was perceived according to Descartes were models of the objects, formed by the sensory systems of the perceiver. Descartes' substitution is rational for an avowed mind/body dualist but ceases to be rational if there is no postulated 'mind' to which the models may be presented. Why do we need models if we have the real things in our perceptual field? Surely all we need is information about the real things, enabling us to respond appropriately to them? The majority of contemporary theorists follow Descartes in rejecting this argument

(for example, Marr 1982). They seek only to specify how the organism constructs its hypothetical internal models. Gibson (1979) is the exception; insisting that what we perceive are objects in the real world, not models of them (see Dualism; Perception §1).

2 The emergence of psychology

Psychology was the last of the established sciences to break free from philosophy (the word 'psychology' is rarely used prior to the nineteenth century). But the change of emphasis to which Collins (1973) drew attention had profound effects on philosophy which have in turn influenced psychology. If Descartes' substitution is accepted, it becomes possible (perhaps obligatory) to adopt what Collins calls the 'egocentric, sceptical perspective' - to doubt the existence of the external world and even the continuity of the perceiving person. The models may, after all, have no material correlate; they may exist only in the immaterial mind that perceives them. Philosophers since Descartes have laboured mightily to prove the postulations of everyday thinking - that we exist and that a real world external to us exists - but without success.

Psychology cannot allow the possibility that an immaterial 'mind' that is not law-governed interacts causally with each material human body; the intervention of such a mind would put the body's behaviour beyond the reach of science; the behaviour would still be observable but no longer predictable, even in principle. At first, Leibniz's doctrine of pre-established harmony appeared to offer a way out of the dilemma, in the form of 'psycho-physical parallelism'. In a world in which dogmas of the Church were widely accepted it seemed legitimate to postulate that mind and body might not interact but might, by God's grace, run in parallel - like two ideal clocks, started at the same moment, that always show the same time though there is no connection between them. Psycho-physical parallelism was tacitly accepted during the greater part of the nineteenth century. It was compatible with the theories of memory and thinking then taken for granted according to which ideas/mental images could become associated by contiguity of time and place in individual experience, or as (apparent) cause and effect. The occurrence of one would then be likely to elicit others with which it had been associated in the past. A 'mind' was tacitly assumed to be available to draw appropriate inferences from the associations furnished by experience. At about the same time, mathematics moved on from concern with the development of new, useful mathematical techniques to worries (often associated with the problem of infinitesimals) about the validity of the techniques. They appeared to work, but could they be relied upon?

At the end of the nineteenth century the theoretical status of psychology was uncertain. There had been advances in neurophysiology sufficient to show that the brain was intimately involved in behaviour and one or two lone voices had been heard advocating a purely physiological psychology. Generally, however, physiologists were careful to keep their options open. Fritsch and Hitzig (1870: 96), for example, believed that 'some psychological functions and perhaps all of them, in order to enter matter or originate from it need certain circumscript centers of the cortex'. Wilhelm Wundt opened the first psychological laboratory at Leipzig in 1879 but he held strong views on what might and might not be done in the laboratory. Basic sensory functions were investigated to establish what was the least perceptible stimulus, or the least perceptible difference between stimuli, using controlled introspections in which the subject reported verbally whether the stimulus was present or whether two stimuli differed (Wundt 1894). Wundt did not, however, regard the higher mental functions as suitable topics for the laboratory, they might only be studied by historical and/or descriptive methods. He insisted on the mind/body distinction and succeeded (1907) in aborting a programme of laboratory investigation into higher mental functions begun in Würzburg a few years earlier (Humphrey 1951); not, however, before it had been established that many such functions proceed in the absence of introspectible mental content. This finding was important as a refutation of one central argument of associationism which maintained that associated ideas/images must necessarily be in consciousness, that is, they must be introspectible. Titchener studied in Wundt's laboratory but, on his return to the USA, he too conducted introspective investigations into 'higher mental functions'. He maintained that failure to find introspectible mental content showed only that the subject had not tried hard enough and once confessed himself 'not at all astonished to observe that the recognition of a (shade of) gray might consist of a quiver of the stomach' (1909: 179).

3 The establishment of a scientific psychology

At the beginning of the twentieth century, psychology was still obsessively concerned with the problem of consciousness. Animal psychology asked only whether or how far animals could be said to possess consciousness.

But philosophical thinking on the subject had moved on. It is a relatively short step from psycho-physical parallelism (where the hand of the deity keeps mind and body on parallel courses) to epiphenomenalism where mind and body keep on parallel courses because mind has no independent causal status and simply reflects what is going on in body (see Epiphenomenalism). Either way psychologists may legitimately proceed to the experimental investigation of behaviour but the latter view sheds surplus theological baggage that had become progressively less easy to accept. Consciousness does not need to be denied to man or animal and may continue to be employed as a means of access (via language) to human subjects. However, no separate account need be given of the functions of consciousness since it has none.

J.B. Watson, the founder of behaviourism, was a philosopher who had become interested in animal behaviour (see Behaviourism, methodological and scientific). Watson (1913) asserts that consciousness should rightfully play no greater role in psychology than it plays in chemistry or physics. He appears to mean that it should play only the role allotted to it in associationism, as an arena in which observed associations may be interpreted as evidence of the existence of laws governing the observed phenomena (for example, of physics or chemistry). This should not be an acceptable view in psychology, since the psychologist includes in his remit the mind which does the interpreting, though it does well enough in the other sciences. Watson later advocated the study of observable behaviour only, without any reference to internal, unobservable ‘minds’. (A point of view maintained even more strongly by B.F. Skinner (for example, 1974).) Neither behaviourism nor associationism offers any satisfactory account of how or where theories (of which behaviourism and associationism are themselves examples) are constructed.

Watson believed at this time that ‘thoughts’ might be nothing more than sub-threshold movements of the speech organs and tried in his laboratory to obtain photographic records of such movements. Koch (1964) reports Lashley as having told him that Watson hoped to present the photographic evidence in his presidential address to the American Psychological Association (1915) but being unable to obtain any, adopted the work of his student, Lashley, on conditioning (with due acknowledgement). Thus was born the association between behaviourism and conditioning theory which occupied most US academic psychologists for the next half century. Lashley had encountered Pavlovian theory in Germany (Pavlov’s work had been translated into German although no English translation appeared until 1927). Watson (and, subsequently, Clark L. Hull) developed Pavlov’s theory into ‘learning theory’. Something was lost in the transition, since the US theory was frankly mechanist whilst Russian theory, though materialist, was never mechanist. (Mechanism was politically unacceptable in the USSR; account had to be taken of consciousness since both Marx and Lenin had used the term.) US theorists avoided reference to ‘mind’ or ‘consciousness’ since the intellectual tradition to which they subscribed equated consciousness with mind and mind with inaccessibility in principle to scientific investigation. The fundamental unit of behaviour was, for them, the link between stimulus and observable response (the S-R bond). Hull (1930; 1935) tried to show how higher mental functions might be explained as sets or sequences of these fundamental units but the basic theoretical and experimental task was, in his view, to understand the unit. Hull and his followers are said to have obtained almost total hegemony over US academic departments of psychology in the period 1920-1950 but the hegemony did not survive Hull’s death in 1952. (US behaviourism is critically evaluated in Rozeboom (1970); for Russian work over the same period see Payne (1968) and Razran (1971).)

4 The contemporary orthodoxy

The demise of behaviourism and its replacement by cognitive, ‘information-processing’ psychology in the so-called cognitive revolution of the 1960s, was motivated primarily by a desire among experimental psychologists to work on more interesting topics than Hull had permitted. But it must also be remembered that a generation of psychologists who possessed a degree of philosophical sophistication was dying off. The new generation had often studied nothing but psychology. Until the 1960s, most psychologists had recognized the desirability of a central, theoretical framework, performing the function in psychology that atomic theory performs in physics and the theory of complex molecules, in biology. They had also recognized the undesirability of referring unresolved issues upwards to an unspecified, higher supervisory level; that is, in effect, to the operations of a (possibly immaterial) mind. Since the 1960s neither of these restrictions have been observed. There has been an exponential increase in the range of topics investigated and the sophistication of the techniques employed but few significant advances in our understanding of psychological issues.

Sutherland (1979) despaired of experimental psychology using human or animal subjects since there are simply too many possible contaminating factors to allow rational interpretation of the results obtained. He recommended that efforts should be devoted either to cognitive science (that is, the computer simulation of behaviour) or to neuroscience (the micro- or macro-physiology of the brain). There has been no significant reduction in the flow of experimental papers but both cognitive science and neuroscience have succeeded in establishing themselves as independent disciplines. Cognitive science takes symbols for granted and has tried to develop ways of manipulating them by computer that approximate to various human mental functions. Recently, technological advance has brought in the possibility of 'connectionism' (Rumelhart and McLelland 1986), which has the advantage of employing analytic methods resembling those of the brain more closely than do the methods of symbolic computing. Proponents of symbol processing (or 'artificial intelligence') and of connectionism, often profess to see their respective disciplines as incompatible alternatives but since the latter appears to model symbol formation and the former takes symbols for granted, it is likely that an accommodation will be reached. It remains to be seen whether this will contribute substantially to our understanding of human psychological function (see Artificial intelligence).

At the beginning it was suggested that human beings had, by the seventeenth century, evolved a monistic conception of human nature, in a direct perceptual relationship with the natural world outside itself, that was fundamentally sound. Unfortunately this conception had, over the centuries, acquired encumbrances relating to many tangential issues of religion and metaphysics that rendered the whole a positive block to further progress in human knowledge. In the seventeenth century the block was broken; physical and biological scientists were then free to proceed unencumbered - whether the mind was material or immaterial was of no consequence. For the psychological scientist this is not, of course, the case. If mind is immaterial then psychology has, as we have seen, no obvious way to proceed. The response of many psychologists has been to insist that the mind is, of course, material but to act as if there existed nevertheless an immaterial 'mind of last resort' that perceives and manipulates the models supposed by cognitive psychologists (following Descartes) to comprise the furniture of the mind. The activities of this 'mind of last resort' should constitute a significant part of the subject matter of psychology but are often passed over.

5 Marginal influences

The Gestalt movement (Koffka 1935) was a legitimate reaction against the late nineteenth-century view that a theory of perception could be derived from data on subjects' responses to punctate stimuli - points of light or brief pure tones. The Gestaltists' fundamental observation (the 'phi' phenomenon) was that if two adjacent points of light are flashed in quick succession, what the subject sees is one point of light moving from the first location to the second. The experience is unequivocal and many related phenomena have been discovered. The Gestalt emphasis on perceptual holism was salutary but they have advanced no viable alternative theory of perception (see Gestalt psychology).

The experimental work of Jean Piaget on human developmental psychology was influential in the UK and in the USA but his theory was not. He devoted a long life to the elaboration of a biologically-based theory of human development which may yet prove to be significant (Gruber and Vonèche 1977).

Phenomenological psychology, stemming from Brentano (§2) and Husserl (Thinès 1977), has sometimes limited itself to detailed analysis of subjects' states of consciousness but more recently (for example, Harré and Gillett 1994) emphasis has been placed on discourse between persons and on the social construction of personality. There is, some believe, no necessary conflict between a structural psychology that sets out to achieve scientific understanding of the human organism - the evolved 'thing' that is capable of doing what we know human beings have done and can do - and a discursive psychology concerned only with what the organism does in its interaction with other organisms, but both have necessarily to confront the question of logic. There is something inherently odd about founding logic, as at present, outside psychology - though logicians are unanimous in rejecting what they call 'psychologism'. It may be worth noting that Husserl (who is often cited as the final authority for the rejection of psychologism) thought his views on that subject had been grossly misrepresented. Husserl (1929) makes it clear that he thought he had shown only that logic could not be founded in an associationist psychology of mental contents. He believed that it could and must be founded in a cognitive, structural psychology (see Phenomenology, epistemic issues in).

6 Theories of personality

So far, this entry has been concerned with theories that try to answer the question 'what sort of thing is the human organism and how should it be studied scientifically?' As we have seen, another level of theorizing considers the person in interpersonal interaction, in the development of 'personality'. Such theorizing frequently starts out from the malfunctioning or neurotic personality, a theory of normal personality may or may not emerge. Sigmund Freud remains the outstanding theorist at this level. He constructed a theory of normal human development with an associated therapeutic method to correct developmental anomalies. The former was the more important in his view; he never had much confidence that the latter would do more than alleviate a patient's worst difficulties. By 1914 he had already acquired a number of US disciples with whom he lost contact during the First World War. They were largely responsible for the growth of psychoanalysis as a medical speciality, offering to cure neuroses as the orthopaedic surgeon cures broken limbs. Freud regarded neurosis as the inescapable accompaniment of civilized living - since the latter is entirely unnatural to man. Both are consequences of the diversion of libidinal energy whose natural object is physical survival and reproductive success. By complex evolutionary processes man has civilized himself, using energy surplus to what was required for survival or could be dissipated in the struggle for reproductive success (since civilization usually demands at least the pretence of monogamy). Energy not dissipated in sexual activity or sublimated in the achievement of economic, artistic or other non-sexual goals, must inevitably issue in neurotic behaviour. The developmental process determines, for better or worse, how libidinal energy will be directed in each individual. This is an unflattering account of human nature but we should not look to science for flattery. The point is that Freud's theory calls on no forces that are not already active in the evolution of lower animal species. Anyone who does not accept his account has either to argue for an alternative, working within the same constraints, or to postulate new forces specific to the formation of human (as opposed to animal) nature. Today, not many people take the latter course. Freud proposed answers to questions that no one else has asked and some would prefer never to have been asked at all. There is no guarantee that genuine alternative answers would be more flattering to human *amour propre* than Freud's were.

Following Freud, personality theories proliferated (Maddi 1989). Freud believed that successful therapy required recall into awareness of forgotten traumas of early childhood development, to redirect libidinal energy that had been misdirected or had failed to achieve any direction. The therapy was long, expensive and frequently unsuccessful. Carl Rogers (1961) proposed as an alternative a form of 'non-directive' therapy. In ordinary discourse between persons, an opinion expressed by one party is conventionally met with agreement or disagreement by the other party. Where the opinion concerns a personal predicament of the one (who knows more about it than anyone else), it is unlikely that either agreement or disagreement will persuade them to change it. Non-directive (or 'client-centred') therapy encourages the client to express opinions but the therapist avoids expressing either agreement or disagreement. The clients, not receiving the usual reinforcement, may then be inclined to review and perhaps modify their opinions for themselves. Unfortunately, Rogers later reverted to a form of directive therapy, conducted in small groups, on the (possibly mistaken) assumption that most human beings have it in them to function as therapists.

A second major development was the attempt to apply learning-theory principles (that is, behaviourism) in a clinical context; in which the client's behaviour is modified by manipulation of reinforcement by the therapist. Alcoholism may be treated by administering a drug that makes patients violently sick if they subsequently take alcohol. Phobic anxiety may be treated by teaching the patient to relax completely (a physical state incompatible with anxiety) and introducing the feared object gradually over a number of therapy sessions. All these treatments try to extinguish the learned connection between some unwanted piece of behaviour and its present, painful or pleasurable, outcome.

Treatment of personality disorders is now almost always eclectic, drawing on more than one therapeutic tradition. Since the object is to benefit the client, any treatment that works is, in principle, justifiable. The age of grand, overarching theories of personality has perhaps passed. True or false, they turn out to have few useful consequences - human nature is simply too complex.

See also: Jung, C.G.

N.E. WETHERICK

References and further reading

Arnold, W.J. (1976) *Nebraska Symposium on Motivation, 1975: Conceptual Foundations of Psychology*, Lincoln, NE and London: University of Nebraska Press. (Everything in this volume is worth reading.)
Collins, A. (1973) 'The Objects of Perceptual Consciousness in Philosophical Thought', *Social Research* 40: 153-76.(An invaluable brief account of the changes in world-view brought about by Descartes in the seventeenth century.)
Coulson, W.R. and Rogers, C.R. (1968) *Man and the Science of Man*, Columbus, OH: Charles E. Merrill. (Included because it contains a substantial contribution from Michael Polanyi whose work has so far had no impact on psychological theory but may do so in the future.)
Fritsch, G. and Hitzig, E. (1870) 'On the Electrical Excitability of the Cerebrum', in G. Von Bonin, *The Cerebral Cortex*, Springfield, IL: Charles C. Thomas, 1960.(An early classic of neurophysiology, originally published in German.)
Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Boston, MA: Houghton Mifflin.(A modern classic on visual perception.)
Gruber, H.E. and Vonèche, J.J. (1977) *The Essential Piaget*, London: Routledge & Kegan Paul.(A selection from Piaget with linking commentary.)
Harré, R. and Gillett, G. (1994) *The Discursive Mind*, London: Sage.(Concerned with what they term the 'discursive turn' in psychological theory.)
Hull, C.L. (1930) 'Knowledge and purpose as habit mechanisms', *Psychological Review* 37: 511-25.(Hull's attempt to account for 'knowledge' and 'purpose' in his own terms, that is, without reference to any higher 'mental function'.)
Hull, C.L. (1935) 'The mechanism of the assembly of behavior segments in novel combinations suitable for problem solution', *Psychological Review* 42: 219-45. (Hull's attempt to explain 'problem solving' in his own terms.)
Humphrey, G. (1951) *Thinking*, London: Methuen.(Described by the author as 'an introduction to its experimental psychology' which it is. It is probably also the best single work on psychological theory.)
Husserl, E. (1929) *Formal and Transcendental Logic*, trans. D. Cairns, The Hague: Martinus Nijhoff, 1969.(A late work of Husserl on logic, originally published in German.)
Koch, S. (1964) 'Psychology and emerging conceptions of knowledge as unitary', in T.W. Wann (ed.) *Behaviorism and Phenomenology*, Chicago, IL and London: University of Chicago Press.(This book also contains much other matter of interest.)
Koffka, K. (1935) *Principles of Gestalt Psychology*, New York: Harcourt Brace. (The standard account of Gestalt theory.)
Maddi, S.R. (1989) *Personality Theories: A Comparative Analysis*, Homewood, IL: The Dorsey Press, 5th edn.(A standard work on personality theories in all their infinite variety.)
Marr, D. (1982) *Vision*, San Francisco, CA: W.H. Freeman.(A standard work on the computational theory of vision.)
Payne, T.R. (1968) *S.L. Rubinsteyn and the Philosophical Foundations of Soviet Psychology*, Dordrecht: Reidel. (An account of psychological theory in the USSR.)
Razran, G. (1971) *Mind in Evolution*, Boston, MA: Houghton Mifflin.(A brilliant synthesis of US and Russian experimental work by a psychologist who was familiar with both.)
Rogers, C.R. (1961) *On Becoming a Person*, Boston, MA: Houghton Mifflin. (Roger's own account of his earlier work.)
Rozeboom, W.W. (1970) 'The Art of Metascience III', in J.R. Royce (ed.) *Towards Unification in Psychology*, Toronto, Ont.: University of Toronto Press.(A critique of behaviourist theory.)
Rumelhart, D.E. and McLelland, J.L. (1986) *Parallel Distributed Processing*, Cambridge, MA: MIT Press, vols 1 and 2.(The foundational work on connectionism.)
Skinner, B.F. (1974) *About Behaviorism*, New York: Knopf.(A statement of Skinner's position.)
Sutherland, N.S. (1979) 'Neurosciences Versus Cognitive Science', *Trends in Neuro-Sciences* 2 (8): I-II. (Sutherland's rejection of experimental cognitive psychology, in favour of either 'cognitive science' (computer simulation of mental function) or 'neuroscience' (direct investigation of the central nervous system).)
Thinès, G. (1977) *Phenomenology and the Science of Behaviour*, London: Allen & Unwin.(A valuable but little read work relating continental phenomenology to biology, neurophysiology and experimental psychology.)
Titchener, E.B. (1909) *Lectures on the Experimental Psychology of the Thought Processes*, New York: Macmillan.(The most satisfactory introduction to Titchener's ideas.)
Watson, J.B. (1913) 'Psychology as the Behaviorist Views It', *Psychological Review* 20: 158-77.(Watson's first statement of the behaviourist position.)

Wolman, B.B. (1980) *Contemporary Theories and Systems in Psychology*, New York and London: The Plenum Press, 2nd edn.(A detailed and critical account of movements in psychological theory in the first three-quarters of the twentieth century.)

Wundt, W. (1894) *Lectures on Human and Animal Psychology*, London: Swan, Sonnenscheim & Co.(Lectures on his experimental work.)

Wundt, W. (1907) 'On the Ausfrage Experiments and on the Methods of the Psychology of Thinking', *Psychologische Studien* 3: 301-60.(Published in German and not available in translation. For a detailed account and critique see Humphrey (1951: 107-16).)

Ptolemy (*c.* AD 100-70)

The astronomer Ptolemy was one of the leading scientific figures of Graeco-Roman antiquity. His contributions to philosophy lie in his reflections on scientific activity. In knowledge, he distinguishes a perceptual stage, which provides the natural link between knowledge and things, from a further, rational stage, governing the transition to science. The move towards science consists of the progressive distinction between concepts, initially acquired through experience and methodical observation. Many components of his thought are derived from earlier philosophy, but he excludes those aspects which bear on more general philosophical issues.

Ptolemy - Claudius Ptolemaeus - probably lived at Alexandria. His *Mathematical Composition*, better known under its Arabic title the *Almagest*, offered a complete treatment of celestial phenomena, based on the assumption of geocentricism and employing mathematical models founded on uniform circular motions. It is one of the greatest works in the history of science. In addition, he wrote monographs on astronomical themes and comprehensive treatments of nearly every science with a high mathematical content: optics, harmonics, geography, astrology. In all these areas his writings remained standard reference works for centuries.

His philosophical work, *On the Criterion and Commanding-Faculty*, concentrates on the act by which one attains knowledge of real objects - primarily sensible ones. The act consists of two stages: (1) by means of the sensory apparatus the soul is affected and perceives this; (2) the intellect reads the perception, translating it into the terms of reason (see Logos). This distinction between a perceptual and a rational stage does not make the translation artificial or arbitrary. The sensory apparatus is a set of *natural* tools, each with a proper, normal and natural mode of application which already contains a rational and conceptualizable principle. As a result, the perception yields a determinate reading, translatable into rational terms.

This theory is located within a general epistemology. The intellectual and perceptual faculties (the latter subdivided into five senses) both deal with the same domain - things and their properties. Intellect (see Nous) develops only in human beings and is later than the perceptual faculty. Starting from the data of individual perceptions, intellect comes to grasp both the functioning of the senses and aspects of external things. By means of memory, these aspects are detached from the original perceptual encounter, to form the stock of concepts used in the reading of percepts. Subsequently they become the object of separate reflection by the intellect, as by progressively articulating them it classifies things, through genera and species, all the way down to indivisible particulars. Underlying this task are certain concepts - same/different, equal/unequal, similar/dissimilar - which Ptolemy calls the intellect's proper objects, not derived from perceptual activity at all. Anyone who does not progress beyond the stage of focusing on objects in isolation remains at the level of 'opinion'. The scientific frame of mind is achieved by incorporating one's opinions into the rational order.

While perception is incontrovertible regarding effects on the soul, being a direct awareness of these, it can nevertheless be deceptive regarding its own external cause. It indicates this cause, without being the judge of the indication's correctness. However, that does not reduce perceptions of things to mere subjective experiences, because (1) the natural basis of cognitive activity guarantees that in normal conditions it functions correctly, and (2) it is the intellect's role, through its capacity for rational translation, to act as the judge. When in doubt, the intellect can have recourse to repetition of the cognitive act. Moreover, it can organize systematic observation, and this, combined with the refinement of concepts, constitutes the path from opinion to knowledge.

If perceptions can be judged by the intellect, they must be analysable. Here Ptolemy adopts the Aristotelian distinction between special and common sensibles (see Aristotle §18), and his *Optics* (partially extant in a Latin translation of an Arabic translation) offers an analysis of vision. 'In reality' vision perceives condensed illuminated surfaces; 'immediately' it perceives the special sensibles, that is, colours; 'derivatively' the common sensibles, that is, body, size, shape, position and motion. Structurally, vision is treated in geometrical terms as a cone with its apex at the eye. Ptolemy establishes the limits of vision's powers, and its internal interrelations in perceiving its different objects. And he produces a complete typology of optical illusions, both those due to natural external conditions and those due to inadequate understanding of vision's workings. The *Optics* perfectly illustrates how Ptolemy's overall conceptual framework can encase a scientific inquiry combining empirical observation, conceptual analysis and mathematics.

Ptolemy rarely invokes earlier philosophical discussions. Nevertheless, his thought and terminology include elements from every preceding major philosophy. His strategy is to extract from the tradition whatever he needs in order to define his scientific activity in its various aspects. But his principal philosophical hallmark is his way of ranking the perceptual and rational stages in knowledge, which enables him to build out of natural cognitive processes a path which, under the guidance of reason, leads to science. For any science to travel this path (as distinct from an individual's acquisition of a scientific disposition) takes longer than a human life span. He can thus, in the *On the Criterion and Commanding-Faculty*, deliberately ignore the classic Stoic -Academic debate (see Arcesilaus §2; Stoicism §§12-13) on the criterion (despite frequently drawing material from it): he may take it to bear only on practical philosophy, not philosophy of science.

According to P. Duhem's (1913) classic thesis, Ptolemy assigned a purely instrumental value to the mathematical models used in his scientific work: they helped him organize his observations and derive predictions from them, without any pretence of referring to physical fact (see Duhem, P.M.M. §4). In the *Almagest* Ptolemy certainly sometimes *uses* his models in that way. Nevertheless, his explicit defences of them do not sound instrumentalist, but normally rest on physical presuppositions. Moreover, in his *Planetary Hypotheses* he constructs an intricate system of nested corporeal spheres, and these match the components of the *Almagest*'s models. Finally, instrumentalism seems scarcely compatible with his theoretical pronouncements in the *On the Criterion and Commanding-Faculty* and the *Optics*. We may say that Ptolemy took a realist view, but that, given the absence in his day of a clearly formulated instrumentalist position, his implicit realism is not so rigorously defined as to prevent his adopting a more dismissive tone when it suits him.

FERRUCCIO FRANCO REPELLINI

List of works

Ptolemy (*c*. AD early 150s) *Almagest*, trans. G.J. Toomer, London: Duckworth, 1984.(English translation which includes notes.)

Ptolemy (*c*.AD 140-70) *On the Criterion and Commanding-Faculty*, trans. various authors in P. Huby and G. Neale (eds) *The Criterion of Truth*, Liverpool: Liverpool University Press, 1989, 179-230.(Parallel text and English translation, including notes.)

Ptolemy (*c*. AD 150s or 160s) *Optics*, in A. Lejeune (ed.) *L'Optique de Claude Ptolémée (The Optics of Claude Ptolemy)*, Leiden: Brill, 1989.(Annotated Latin text with French translation.)

Ptolemy (*c*. AD 150s or 160s) *Harmonics*, in A. Barker (trans. and ed.) *Greek Musical Writings*, vol. 2, Cambridge: Cambridge University Press, 1989, 275-391.(With extensive notes.)

Ptolemy (*c*. AD 150s or 160s) *Tetrabiblos*, trans. F.E. Robbins, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1980.(Greek text with facing translation.)

Ptolemy (*c*. AD 160s) *Planetary Hypotheses*, in J.L. Heiberg (ed.) *Claudii Ptolemaei Opera astronomica minora*, Leipzig: Teubner, 1907.(Greek text and German translation of the first (and the only extant in Greek) part of Book I, and German translation, by L.Nix, of an Arabic translation of Book II. There is an English translation of the Arabic translation of the second part of Book I (missing in Heiberg's edition) in B. Goldstein, 'The Arabic Version of Ptolemy's Planetary Hypotheses', in *Transaction of the American Philosophical Society*, n.s. 57: 3-55.)

References and further reading

Duhem, P. (1913) *Le Système du monde I (The World System)*, Paris: Hermann.(Proposes the instrumentalist interpretation of Ptolemy's mathematical astronomy.)

Lejeune, A. (1948) *Euclide et Ptolémée: Deux stades de l'optique géometrique grecque (Euclid and Ptolemy: Two Stages of Greek Geometrical Optics)*, Louvain: Université de Louvain.(The standard work on the *Optics*.)

Long, A.A. (1988) 'Ptolemy on the Criterion: An Epistemology for the Practising Scientist', in J.M. Dillon and A.A. Long (eds) *The Question of 'Eclecticism'*, Berkeley and Los Angeles, CA: University of California Press, 176-207.(A careful evaluation of Ptolemy's philosophical debts in *On the Criterion*.)

Lloyd, G.E.R. (1978) 'Saving the Appearances', *Classical Quarterly* 28: 202-22.(Refutation of Duhem's (1913) thesis.)

Taub, L.C. (1993) *Ptolemy's Universe*, La Salle, IL: Open Court.(Readable treatment of the philosophical motivations underlying Ptolemy's astronomy.)

Public interest

The concept of the public interest can be used in a wide variety of ways, and this has led many to say that it is devoid of meaning. However, the concept enables us to evaluate the tendency of policies and institutions to promote the interests of the members of a society considered in their broadest relations, for example in connection with policies to promote public health. In this sense it has significance. Historically, the concept of the public interest has drawn upon three main traditions of thought: the utilitarian idea of utility maximization; the tradition of civic republicanism; and Rousseau's idea of the general will. Nowadays, three main ways of meeting the public interest are distinguishable: the supply of certain indivisible goods like clean air; the preservation of identity-conferring social goods like a distinctive language; and the balancing of competing considerations in the making of public policy. Although the provision of goods in the public interest may be associated with injustice, there is no reason in general to think that justice and the public interest must conflict.

1 The concept of the public interest

The principle of the public interest is a broad one. Moreover, it is related to other ideas like 'common advantage', 'common good', 'public good', 'public benefit' and 'general will'. Given this wide range, it is tempting to say that the idea of the public interest is so indeterminate as to be virtually meaningless.

Despite these doubts, it is possible to identify a core sense to the term. The root idea seems to be that a policy, practice or institution is in the public interest if it is to the interest, advantage, benefit or good of some group or set of persons considered to be members of the same society. Thus, a well-functioning legal system together with an uncorrupt system of public administration will be in the public interest, by facilitating a wide range of civil and individual activities in a society.

It may turn out that a policy or practice that is in the public interest, so defined, will not be to the advantage of some subgroup in society. Does this show that all we mean by the concept is a collection of private interests? Not at all. If it would be more profitable for the Mafia to have a corrupt system of law and public administration, we may still quite properly say that there is a public interest to the contrary, since there is no way in which the minority interest could serve a more general good. For this reason, Bentham, with typical verve, called subgroups whose existence depended upon subverting the public interest 'sinister interests'.

2 Three traditions

An important influence in defining the concept of public interest has been the utilitarian tradition, in particular that strand stemming from the philosophic radicals at the beginning of the nineteenth century (see Utilitarianism). The philosophic radicals made the concept of the public interest central to their political philosophy and defined it in terms of the aggregate interests of the members of a society.

Within the tradition known as civic republicanism, by contrast, the principle of the public interest is associated with virtuous citizens upholding the institutions of a free republic, including the practices of participation in public affairs and the performance of military service (see Republicanism).

Rousseau's use of the term 'general will' forms a third tradition. According to Rousseau, the policies and practices of a society should be determined by a vote of a majority of citizens. Provided that all citizens are voting with an awareness that their choices will affect them in all the roles they occupy, the result of a majority vote will constitute the general will. In voting, citizens do not identify a pre-existing public interest, but instead define that interest in the process of choice. In this way, Rousseau hoped to combine a republican sense of virtue with a modern notion of moral autonomy (see General will §1; Rousseau, J.-J. §§1-3).

3 Modern conceptions of the public interest

The traditions listed above intermingle in modern thought in defining three varieties of meaning for the public interest.

The first sense owes much to utilitarianism and draws upon the economist's notion of a public good. Within economics a pure public good is defined as being non-rival, so that one person's consumption does not diminish

the consumption of others, and non-excludable, so that if one person in a society consumes the good, then others will as well. An example of a public good in this sense is clean air. Person A's consumption of clean air does not detract from anybody else's consumption, and if clean air is available to A it will also be available to anyone else in the society. Goods having these properties include: environmental amenity, public order, good government, a sense of fairness and obligation in contractual relations, tolerance and civic mindedness.

A familiar difficulty in the provision of such public goods is the so-called 'free rider' problem or problem of collective action (see Rational choice theory). The difficulty arises because, although it may be in everyone's interest that a public good is supplied, it may be in no particular person's interest to bear a share of the costs of supply. Since the good is non-rival and non-excludable, each individual will benefit from the existence of the good, whether or not they have made a contribution to its costs. Self-interested persons would therefore be tempted to free ride on the contributions of others, hoping that they could gain the benefit without incurring the costs. But free riding, when generalized, is collectively self-defeating, since if no one incurs the costs, the public good will not be supplied.

Various suggestions have been made for overcoming the free rider problem. A Hobbesian line of argument is that it provides a justification for the state to have coercive authority, so that it can enforce individual contributions to the supply of public goods in the public interest (Hobbes 1947). This line of argument also has affinities with Rousseau's idea that in the social contract people are 'forced to be free' in implementing the general will. Other suggestions for overcoming the free rider problem rely upon the consequences of repeated interaction among small numbers of people, in which norms of cooperation can emerge, or upon a moral education that will encourage people to internalize Kantian principles of fairness.

Public goods in this sense are instrumental, since they are compatible with widely divergent personal conceptions of the good. By contrast, the public interest may include the idea of a common good. Drawing its inspiration both from the tradition of civic republicanism and from Rousseauian/Hegelian notions of community and social identity, this formulation identifies certain irreducibly social goods that are constitutive of central human interests: a language and the cultural traditions that go with it; a sense of place and landscape for a people; the continuation of a people's identity; or popular participation in the shaping of a common political life.

The specification of the concept of a common good in this sense is central to debates between liberals and communitarians, and therefore to debates about the virtues of patriotism and the obligations of citizenship. Communitarians contest the idea that individuals in a society are bound together solely by their subscription to a common set of rules constraining the exercise of their freedom. They argue that a common social life exists prior to the formation of individuals and that notions of the right and justice have to be conceptualized so that they are consistent with a society's common good. In practice, such a view would presumably licence policies and practices that would inculcate certain patriotic virtues, for example a willingness to defend one's country in time of attack (see Community and communitarianism).

The third modern notion of the public interest is more formal than the previous two, although common in everyday political argument. Here the public interest is a balance between competing, but incompatible, goods. For example, freedom from police constraint is a good and so is personal security. To ensure personal security, it may be necessary to give the police general powers to stop people on suspicion of carrying concealed weapons. In this case two incompatible goods are set against one another, and it is sometimes said that the public interest is best promoted where a suitable balance is struck between them.

Can one define how this balancing might take place? Two methods have sought to give substance to the idea of balancing: Benthamite utilitarianism, in which the balance is struck at the point at which aggregate utility is maximized, and Rousseau's notion of the social contract, in which the balance is struck at the point of majority decision on the general will.

Both approaches are beset with difficulties, however. It is not clear that Benthamite utilitarianism can provide a weighting of utilities that does not beg the question about the value of different outcomes. The Rousseauian approach can provide little independent evidence about the terms of a social contract. In practice, despite these limitations, public policy is often made by some form of balancing, either by decision makers weighing the supposed advantages or disadvantages of various alternatives or by popular decision. Thus, the principle of the

public interest as a balancing of goods has some significance, even if its meaning is difficult to state precisely.

4 The public interest and reducibility

Who or what might be the bearer of a public interest? In speaking of the public interest, it may seem as though we are identifying an entity over and above the individuals who make up a society. Methodological individualists, who assert that there is no such entity, baulk at this thought.

In the first and the third modern senses of the public interest, however, it is not necessary to abandon methodological individualism, provided we are allowed to identify the interests of non-assignable individuals. This means that the ascription of interests cannot rest solely upon the expressed preferences of actual individuals, but must instead presuppose that there are certain interests sufficiently widely shared to ascribe them to a wide range of persons whoever they might be (see Needs and Interests §1).

The second sense of the public interest, involving the notion of a common good, raises more complicated questions. Languages and cultures are themselves non-reducible entities. Their preservation and advancement may therefore seem to require a non-reducible sense of the public interest. However, this appearance is misleading if we think that the reason why such entities should be preserved is to advance the wellbeing of those individuals who will be members of a certain society.

5 Justice and the public interest

The notion of the public interest is distinct from that of justice, since we can quite sensibly say that a practice is in the public interest but that the costs of maintaining it are unfairly distributed among the population. For example, a population might need to save water in a drought, but it might be unfair not to exempt market gardeners, whose livelihood depends upon a ready supply of water, from the restrictions.

In recent discussions of justice, it has been common to contrast justice and utilitarian conceptions of the public interest (see Justice §3). Thus, it has been urged that a commitment to an impartial consideration of the public interest cuts across the special moral obligations that individuals owe to one another in justice, and can give weight to illegitimate preferences based upon prejudice that would discriminate against classes of individuals.

It is certainly true that without some principle of equality an appeal to aggregate interests can disadvantage particular individuals. However, some aggregate balancing seems called for where general interests of similar importance are set against one another, for example freedom of movement and freedom from arbitrary attack. In these contexts, the concept of the public interest seems central to our ability to balance interests without being opposed to justice.

See also: Welfare

ALBERT WEALE

References and further reading

Axelrod, R. (1990) *The Evolution of Cooperation*, Harmondsworth: Penguin.(A clear account of the problem of collective action and an original attempt to show how repeated interaction over time will lead to a solution.)

Barry, B.M. (1967) ‘The Public Interest’, in A. Quinton (ed.) *Political Philosophy*, Oxford: Oxford University Press, 112-26.(A clear outline of a want-based theory of interests with a connection to the public interest via Rousseau.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. J. Burns and H.L.A. Hart, London: Athlone Press, 1970.(The best statement of the traditional utilitarian view of the public interest as an aggregate of individual interests.)

Hargreaves Heap, S., Hollis, M., Lyons, B., Sugden, R. and Weale, A. (1992) *The Theory of Choice*, Oxford: Blackwell. (Provides an account of public goods, the free rider problem and references for further reading.)

Hart, H.L.A. (1983) *Essays in Jurisprudence and Philosophy*, Oxford: Clarendon Press.(Part IV contains a brilliant series of essays on the relationship of individual rights to ideas of general welfare.)

Hobbes, T. (1651) *Leviathan*, ed. M Oakeshott, Oxford: Blackwell, 1947.(The principal source of the argument that coercion is necessary to provide the public good of security.)

Kymlicka, W. (1990) *Contemporary Political Philosophy: An Introduction*, Oxford: Clarendon Press.(Chapter 2 contains a useful discussion of utilitarianism and unfairness.)

Mill, J.S. (1992) *Utilitarianism*, in J. Gray (ed.) *John Stuart Mill On Liberty and Other Essays*, Oxford and New York: Oxford University Press.(Chapter 5 contains the classic attempt to reconcile the ideas of justice and the maximization of social utility.)

Pocock, J.G.A. (1975) *The Machiavellian Moment*, Princeton, NJ, and London: Princeton University Press. (Classic account of civic republican tradition in the history of political thought.)

Raz, J. (1986) *The Morality of Freedom*, Oxford: Clarendon Press.(Develops the idea that certain collective goods, like living in a tolerant society, are essential for individual freedom.)

Skinner, Q. (1992) 'On Justice, the Common Good and the Priority of Liberty', in C. Mouffe (ed.) *Dimensions of Radical Democracy*, London and New York: Verso.(A clear statement of the civic republican argument.)

Taylor, C. (1990) *Sources of the Self*, Cambridge: Cambridge University Press.(Advocates the view that individual wellbeing is tied to certain common social goods.)

Pufendorf, Samuel (1632-94)

Pufendorf was the first university professor of the law of nature and nations. His De iure naturae et gentium (On the Law of Nature and Nations) (1672) and De officio hominis et civis iuxta legem naturalem (On the Duty of Man and Citizen according to Natural Law) (1673) greatly influenced the handling of that subject in the eighteenth century. As a result Pufendorf has been recognized as an important figure in the development of the conception of international law as a body of norms commonly agreed to have universal validity by sovereign states. He regarded himself as an exponent of a new moral science founded by Hugo Grotius which transformed the natural law tradition by starting from identifiable traits of human nature rather than ideas about what human beings ought to be.

Pufendorf began his higher education by enrolling to study theology at the University of Leipzig in 1650. He soon found philosophy, history and law more to his taste. Moving to Jena in 1656, he discovered the brand of philosophy that interested him most in the works of Grotius, Descartes and Hobbes. In 1658 he was made tutor to the family of the Swedish ambassador in Copenhagen, only to be imprisoned when war recommenced between Sweden and Denmark. The ambassador escaped to Holland, where Pufendorf eventually joined him, taking the opportunity to further his studies at the University of Leiden. In 1661 he accepted a chair at Heidelberg, moving in 1668 to a new appointment at Lund. When the Danes besieged Lund in 1676 Pufendorf was called to Stockholm by Charles XI and made a privy councillor, secretary of state and Swedish historiographer royal. A decade later he received similar appointments in Prussia. He died after a brief visit to Sweden in 1694.

The parallels between the lives of Pufendorf and Grotius are striking: both became engaged in public service, particularly on Sweden's behalf; both took advantage of a period of imprisonment to develop their ideas; both wrote on a range of subjects - legal, historical and religious as well as philosophical. Where their paths most clearly diverged, however, was in Pufendorf's pursuit of a career as a university teacher. Though his *Elementa iurisprudentiae universalis (The Elements of Universal Jurisprudence)* (1660) was written after his connection with a Swedish embassy had resulted in his incarceration, it built on his earlier studies at Jena and secured his later appointment at Heidelberg. Both there and at Lund Pufendorf was employed as Professor of the Law of Nature and Nations to teach the doctrines adumbrated in Grotius' *De iure belli ac pacis (The Law of War and Peace)* (1625). Pufendorf produced a further exposition of those doctrines in his major work *De iure naturae et gentium (On the Law of Nature and Nations)* (1672) and its hugely influential epitome *De officio hominis et civis iuxta legem naturalem (On the Duty of Man and Citizen according to Natural Law)* (1673).

It is on these three treatises that his reputation as a philosopher rests. He has been hailed with Grotius as an early theorist of a new international order in which a society of sovereign states was formed by the common recognition of rules governing their relations with each other. More recently his own perception of himself as an exponent of a new moral science founded by Grotius has been revived. There can be little doubt that Pufendorf's widely disseminated works contributed greatly to the impact of this modern, post-sceptical, version of natural law theory on philosophers of the eighteenth century. Problematic is the extent to which he made a significant contribution to the substantial development of the theory.

Pufendorf was more conscious than Grotius of his role as an exponent of a distinctive discipline. Hitherto the study of universal jurisprudence had been the preserve of two academic elites: lawyers and theologians. Pufendorf insisted that the discipline of law was properly concerned with the civil laws regulating the conduct of citizens in specific states. Those laws turned on the legislator's will and, unlike natural laws, could not be ascertained by unaided reason. Similarly, the divine laws with which moral theologians were properly concerned had to be learned through revelation of the legislator's will. They regulated the thoughts as well as the deeds of Christians whose citizenship was celestial. By contrast, Pufendorf's discipline used reason to discover the laws that governed the deeds of people of all creeds in earthly societies (see Legal positivism; Sovereignty).

If Pufendorf was led to this clearer demarcation of the provinces of ethics, law and theology by his involvement in university teaching, the altered circumstances of public life in his time may explain a second way in which he went beyond Grotius. Writing against a background of devastating warfare, Grotius had tried to discover a basic moral consensus from which it could be shown how human beings might achieve peace and security through the

formation of societies. Writing after the Peace of Westphalia brought relative stability to Europe, Pufendorf was able to proceed beyond Grotius' minimalist account of natural law and to specify more fully the duties people owed to each other once they had formed societies. His exposition was therefore more extensive than Grotius' and more obviously relevant to readers who could take for granted the peace and security Grotius had sought. Later natural law theorists were to develop this trend further (see Natural law §5).

Both in defining the scope of his discipline and in exploring its contents, Pufendorf made significant adjustments to its conceptual apparatus. Grotius had still written in an Aristotelian way about human beings having an innate tendency towards life in society. In place of this essentialism Hobbes had depicted humans as beings entirely motivated by self-interest. Both authors had identified in each person's interest in self-preservation a natural right to do what each person saw fit for their own preservation. Pufendorf agreed with Grotius that human beings were sociable, but followed Hobbes in basing his claim on identifiable traits imposed on human nature by God rather than on any innate disposition in humans. Having done so, he was able to take human sociability as the foundation of all the duties of humanity and to talk more in terms of duties than of rights. For Pufendorf all rights were simply correlative to duties.

See also: Descartes, R.; Hobbes, T.; Hohfeld, W.N.; Law, philosophy of; Stair, J.D.

J.D. FORD

List of works

Pufendorf, S. (1660) *Elementorum iurisprudentiae universalis libri duo (The Elements of Universal Jurisprudence)*, trans. W.O. Oldfather, Oxford: Clarendon Press, 1931.(An early 'geometrical' synthesis of natural law theory.)

Pufendorf, S. (1672) *De iure naturae et gentium (On the Law of Nature and Nations)*, trans. C.H. Oldfather and W.O. Oldfather, Oxford: Clarendon Press, 1934.(Pufendorf's major work on natural law theory.)

Pufendorf, S. (1673) *De officio hominis et civis iuxta legem naturalem (On the Duty of Man and Citizen according to Natural Law)*, trans. F.G. Moore, New York: Clarendon Press, 1927.(A highly influential epitome of Pufendorf's major work.)

References and further reading

Dufour, A. (1991) 'Pufendorf', in J.H. Burns and M. Goldie (eds) *The Cambridge History of Political Thought, 1450-1700*, Cambridge: Cambridge University Press, 561.(Links Pufendorf's philosophical works with his writings on history and public law.)

Krieger, L. (1969) *The Politics of Discretion*, Chicago, IL: University of Chicago Press.(The only book-length study of Pufendorf's life and thought in English.)

Seidler, M. (ed.) (1990) *Samuel Pufendorf's 'On the Natural State of Man'*, Lewiston, NY: Edwin Mellen Press. (Contains a helpful 'Introductory Essay'.)

Tuck, R. (1987) 'The "Modern" Theory of Natural Law', in A. Pagden (ed.), *The Languages of Political Theory in Early-Modern Europe*, Cambridge: Cambridge University Press, 99.(Revives Pufendorf's own assessment of his place in the history of moral theory.)

Purgatory

According to Roman Catholic teaching, purgatory is the place or state of purification after death in which those who die in a state of grace (and hence are assured of being saved) make expiation for unforgiven venial sins or endure temporal punishment for mortal and venial sins already forgiven. The concept evolved to resolve the theological confusion about the state of souls between personal death and the general resurrection and Last Judgment, to explain what happens to those persons who repent before death but do not live long enough to do penance for their sins, and to make intelligible the widespread practice of praying for the souls of the departed. The doctrine developed in conjunction with a 'high' Eucharistic theology, according to which all the faithful departed take part in the liturgy of the Church. The idea of purgatory is therefore intimately connected with Christian ideas of sin, judgment, retributive punishment, the communion of saints and the idea that salvation occurs in history. It was rejected by the Reformers and, in the second half of the twentieth century, interest from Catholic theologians has waned. Nevertheless, some modern Protestant thinkers have defended the concept as an intermediate phase in salvation.

1 The development of the idea of purgatory

Purgatory is not explicitly mentioned in the Bible, but its supporters have thought it is implied in the apocryphal 2 Maccabees 12: 39-45, where Judas Maccabaeus makes propitiation for the sins of those in his army who died in battle upon discovering that they had worn pagan amulets. Several New Testament texts have also been interpreted as indirectly referring to purgatory; an example is Matthew 12: 31-2, which refers to the sin 'which will not be forgiven either in this world or in the world to come', perhaps implying that expiation is still possible after death.

The custom of offering prayers for the dead in the early Church was widespread and quickly became established practice in the liturgy. With the exception of Aerius, a fourth-century presbyter of Pontus mentioned by Augustine as a heretic, there is no evidence of dissent or hesitation on the matter of prayers for the baptized dead. Hints of the belief in a place for which expiation for sin can be made after death probably appear first in the writings of Tertullian. His contemporary, Clement of Alexandria, more explicitly refers to a place where those who have repented on their deathbed but have not had time to do penance will be sanctified after death by purifying fire. Basil, active in the third to fourth centuries, wrote in a homily on Psalm VII that those with 'any stains or relics of sin' after death 'should be detained'. At the end of the fourth century, Jerome wrote to his friend, the senator Pammachius, on the death of the latter's wife, praising him for venerating her by giving alms, 'knowing that it is written: as water extinguishes fire, so does alms, sin' (Letter LXVI).

The views of Augustine are important, because his subtlety and clarity of thought have carried his influence on Christian philosophy into the present era, but his position on purgatory is not unequivocal. Upon the death of his mother, St Monica, Augustine offered prayers for the repose of her soul, but he is reticent on the question of the nature of purgatory and of a purgatorial fire and treats it tentatively in the *Enchiridion* (chaps 69-70). However, he appears more confident in *De civitate Dei (The City of God)*:

> But temporary punishments are suffered by some in this life only, by others after death, by others both now and then; but all of them before that last and strictest judgment. But of those who suffer temporary punishments after death, all are not doomed to those everlasting pains which are to follow that judgment; for to some, as we have already said, what is not remitted in this world is remitted in the next, that is, they are not punished with the eternal punishment of the world to come.
>
> (XII: 13)

More importantly for the later development of the doctrine of purgatory, Augustine, in the *Enchiridion*, lays down rules for the use of suffrages, as prayers and alms offered for the dead were called. The first rule is that suffrages benefit not all the dead, but only those who are neither damned nor already in heaven; it is too late for the former, and the latter do not need it. The second rule is that a person qualifies for suffrages only in life. Bernstein (1993) says that these two rules came to guide the Church in regulating devotion of the living for the dead. As time went on they were interpreted as permitting the dedication of good works as well as prayers to expiate the sins of the deceased. Thus believers endowed charitable institutions and houses of prayer, and upon this basis the practice of indulgences arose.

2 The emergence of the doctrine and its decline

Atwell (1987) argues that important changes in the Church during the sixth century augmented the development of the idea of purgatory into a full doctrine in the work of Pope Gregory the Great. These developments included the systematic ordering of the penitential system within the Church, the differentiation of mortal and venial sins, and the emergence of a Eucharistic theology in which the faithful departed were commemorated and the saints petitioned for intercession. These liturgical practices were linked with the doctrine of the communion of saints, whereby the Church was understood as involving fellowship between the living and those among the dead who are saved and who therefore remain part of the Church.

Le Goff (1981) claims that the doctrine was really born in the twelfth century, when a great intellectual and moral revolution changed the way people thought about sin and profoundly altered penitential practices. This revolution started with Anselm of Canterbury (§8), whose investigation into the voluntariness necessary for sin in the *Cur Deus homo* (completed 1098) led to an era of examination of the notions of personal responsibility, guilt and punishment. This gave rise to the distinction between guilt (*culpa*) and the debt of punishment (*poena*), a distinction at the heart of the logic of purgatory. *Culpa* can be pardoned through contrition and confession, but *poena* is effaced only by 'satisfaction', that is, by completing a penance. This distinction appears in Aquinas, who says that if a soul has repented and been forgiven, but owes a debt to divine justice at death, this delays the soul's flight to heaven and makes a prior purgation obligatory (*Summa theologiae* IIIa, q.69, a.2, corpus). Aquinas maintained that souls in purgatory accept their suffering voluntarily, understanding that they are in the last stage of preparation for the beatific vision (see Heaven §3).

The official teaching of the Roman Catholic Church was defined at the Councils of Lyons (1274) and Florence (1439) with the intent of reconciling the Greek Christians, who objected to the idea of material fire and the distinction between *culpa* and *poena*. These councils avoid the term *purgatorium* for the sake of the Greeks, but speak of a purification after death for those souls 'who died truly penitent in the love of God before satisfying for their sins through worthy fruits of penance' (Denzinger and Schoenmetzer 1967: 464). There is no mention of fire. Even within the Latin Church there were disagreements on the nature of the pains of purgatory, with some emphasizing the pain of loss - the consciousness of being separated from God, albeit temporarily (and some even maintaining that such pain constitutes the entire pain of purgatory), while others emphasized the pain of sense - physical pain. Aquinas maintained that the pain of sense in purgatory comes from actual fire which is more painful than the greatest sufferings of this world, but later scholastics such as Suárez were inclined to doubt any comparison between the sensory pains of this world and those of purgatory. While the suffering in purgatory was often vividly described in imaginative literature, sometimes the hopefulness of purgatory was stressed instead. In the *Il Purgatorio* of Dante's *Divina Commedia*, purgatory is a mountain that reaches up to heaven. Dante describes how the desire for heaven makes souls fly towards God, causing the mountain to tremble constantly with a joyful thunderclap marking the arrival of each soul into paradise.

The existence of purgatory was denied by the Reformers, such as Luther, who usually also questioned the utility of prayers for the dead. In reply, the Council of Trent in 1562 proclaimed the propitiatory value of the sacrifice of the Mass even for the dead, but otherwise said only that 'Purgatory exists, and the souls detained there are helped by the prayers of the faithful' (Denzinger and Schoenmetzer 1967: 998, 983). No mention was made of the place or duration of purgatory, or of the nature of purgatorial punishment.

Interest in purgatory has declined during the latter half of the twentieth century, accompanied by a decline in interest in most of the concepts with which purgatory has been associated. The documents of Vatican II, the ecumenical council called by Pope John XXIII in the 1960s, do not mention purgatory at all, although prayers for the dead continue to be offered in liturgical services. Occasionally, however, Protestant theologians have offered defences of the existence of purgatory. Hick (1976) speaks approvingly of an intermediate state between death and heaven, but rejects many of the traditional elements of purgatory, believing that the person-making process can continue beyond the grave (see Evil, problem of §4). Brown (1985) has argued that there are conceptual problems with the idea that a human being makes an abrupt transition between a state of moral imperfection at death and a state of moral perfection in heaven. He argues that we know and identify ourselves only through continuity with our past, but a person who is suddenly morally perfected would have no reasonable grounds for believing themselves to be the same person whose past they 'remember'. Brown concludes that if there is a heaven, there

must also be a purgatory that provides a gradual transition to a perfected state. Walls (1992) is another Protestant theologian who accepts the idea of purgatory as part of the logic of the Christian conception of salvation.

See also: Hell; Limbo; Salvation

LINDA ZAGZEBSKI

References and further reading

Anselm (completed 1098) *Cur Deus homo (Why God Became a Man)*, in *St Anselm: Basic Writings*, trans. S.N. Deane, La Salle, IL: Open Court, 1966.(An account of the reasons for the Incarnation.)

Aquinas, T. (1266-73) *Summa theologica*, trans. Fathers of the English Dominican Province, Westminster, MD: Christian Classics, 1981, IIIa, qq.69-70.(Essential reading on the traditional doctrine of purgatory.)

Aquinas, T. (mid 13th century) *Comm. in lib. IV. sent.*, dist. xxi, q.1., Paris: Moos, 1933, 1947.(No English translation. Essential reading on the traditional doctrine of purgatory; see also *Quaestiones disputatae de malo (On Evil)*.)

Arendzen, J.P. (1951) *Purgatory and Heaven*, New York: Sheed & Ward.(A short and simple summary of traditional Roman Catholic teaching on judgment, the resurrection of the body, heaven and purgatory.)

Atwell, R.R. (1987) ‘From Augustine to Gregory the Great: An Evaluation of the Emergence of the Doctrine of Purgatory’, *Journal of Ecclesiastical History* 38 (2): 173-86.(A historical treatment of the beginnings of the idea of purgatory in the early centuries of the Church.)

Augustine (354-430) *Enchiridion theologicum*, Madrid: La editoria Catolica, 1961, chaps 67-70, 109-10.(An important source for the emergence of the idea of purgatory.)

Augustine (413-27) *De civitate Dei*, trans. M. Dods, *The City of God*, New York: Modern Library, 1950, chaps 21, 24, 26.(An important source for the emergence of the idea of purgatory.)

Bernstein, A.E. (1993) *The Formation of Hell*, Ithaca, NY: Cornell University Press. (Mostly about hell, but is up to date and contains some references to purgatory.)

Brown, D. (1985) ‘No Heaven Without Purgatory’, *Religious Studies* 21 (4): 447-56.(A contemporary defence of the idea of purgatory by a Protestant theologian.)

Denzinger, H. and Schoenmetzer, A. (1967) *Enchiridion symbolorum*, 34th edn, Barcelona, Freiburg im Breisgau, Rome, New York: Herder; 30th edn, trans. R.J. Deferrari, *The Sources of Catholic Dogma*, St Louis, MO and London: Herder, 1957.(A handbook in excerpt form of the sources and texts upon which Catholic dogmatic teachings are based. Texts are arranged chronologically.)

Garrigou-Lagrange, R. (1952) *Life Everlasting*, trans. P. Cummins, St Louis, MO and London: Herder.(A good explanation of Roman Catholic theology on death, judgment, heaven, hell and purgatory.)

Hick, J. (1976) *Death and Eternal Life*, New York: Harper & Row.(An influential book, which discusses both Western and Eastern eschatologies.)

Le Goff, J. (1981) *The Birth of Purgatory*, trans. A. Goldhammer, Chicago, IL: University of Chicago Press.(An important book on the historical development of the concept of purgatory; includes a discussion of Dante’s *Il Purgatorio*.)

Walls, J.L. (1992) *Hell: The Logic of Damnation*, Notre Dame, IN: University of Notre Dame Press.(This book is an examination of the Christian conception of hell, but includes a brief defence of the idea of purgatory on page 169.)

Putnam, Hilary (1926-)

Putnam's work spans a broad spectrum of philosophical interests, yet nonetheless reflects thematic unity in its concern over the question of realism. A critic of logical positivism, Putnam opposed verificationism and conventionalism, arguing for a realist understanding of scientific theories. He rejected the traditional conception of meaning according to which speakers' mental states determine meaning and consequently, reference, and put forward a conception of meaning on which external reality, for example, what one talks about, contributes essentially to meaning. Further, citing what he called the division of linguistic labour, Putnam saw the conferring of meaning as a social rather than an individual enterprise. In response to the relativistic challenge that the incommensurability of different theories precludes any possibility of intertheoretical dialogue, Putnam invoked a causal theory of reference construing reference as relatively insensitive to theoretical variation, so that the continuity and rationality of science and communication are upheld. The Copenhagen interpretation of quantum mechanics posed yet another difficulty for realism. Putnam saw quantum logic as an alternative which was compatible with realism, and argued that logic, like geometry, can be revised on the basis of empirical considerations. In the philosophy of mind, Putnam proposed functionalism, the view that mental states are characterized by function rather than material constitution. Putnam also made a substantial contribution to mathematics through his work on the insolvability of Hilbert's tenth problem.

In 1976, Putnam launched an attack on the coherence of the view he termed 'metaphysical realism'. Arguing that relativism and scepticism are disguised forms of metaphysical realism, and likewise incoherent, he suggested an alternative, referred to as 'internal realism'. Clarification of this position and its viability as a third way between realism and relativism is the focus of Putnam's later writings, and of much of the criticism they have incurred.

1 Life

Born in Chicago in 1926, Putnam spent his early years in France. His father, Samuel Putnam, was a well-known writer and translator, an active communist, and a columnist for the *Daily Worker*. Along with Noam Chomsky, Putnam majored in the emerging field of linguistic analysis at the University of Pennsylvania (and also in philosophy and German). His graduate studies were divided between Harvard, where he studied with Quine, Hao Wang, C.I. Lewis and Morton White, and UCLA, where he wrote his Ph.D. dissertation on the concept of probability under the supervision of Reichenbach. In 1953 he moved to Princeton, made the acquaintance of Carnap and, receiving informal instruction from Kreisel, worked intensively on mathematical logic. He has been at Harvard since 1965. Putnam actively protested against the Vietnam War, and was also active in Students for a Democratic Society (SDS) and The Progressive Labour Party, a Maoist group. Around 1972 he became disillusioned with communism. Since Putnam saw his early realism as embodying some of Engels' insights, it seems plausible that this turn affected his philosophical development. Later, he became interested in the study and practice of his Jewish heritage.

2 Realism

Putnam criticized such basic elements of logical positivism as the verificationist theory of meaning, reductionism and conventionalism, while sharing the positivists' interest in and respect for natural science (see Logical positivism §2). He maintained that realism is the only philosophy that does not render the success of science a miracle (see Realism and antirealism; Scientific realism and antirealism §3). His argument from success is presented as structurally similar to hypothetico-deductive arguments within science: realism provides the best explanation for the success of science in the same way atoms and genes provide explanations for observable phenomena. However, this analogy is problematic in that it is purely formal. Realism has no empirical import beyond that of its alternatives, a desideratum scientific hypotheses must meet.

Putnam's criticism of conventionalism is developed in *An Examination of Grünbaum's Philosophy of Geometry* and *The Refutation of Conventionalism*. On the Reichenbach-Grünbaum conception, the core of the transition from Newtonian to relativistic mechanics is a new definition of the spacetime metric. The definition chosen is a matter of convenience, not of truth (see Conventionalism §1). According to Putnam, however, meaning change is only part of the story; theoretical concepts have explanatory import, and must be anchored in a theory that meets both empirical and non-empirical constraints. There is, therefore, much less freedom than the conventionalist alleges.

Putnam raises a similar objection to Quine's celebrated indeterminacy of translation thesis (see Quine, W.V. §9). Here, as in the case of conventionalism, underdetermination is an illusion created by considering an unreasonably limited set of constraints. Once we recognize coherence, simplicity, and so on, as constraints on translation or theory-construction, conventionality and indeterminacy vanish.

3 The meaning of meaning

Putnam faults the traditional theory of meaning for being individualistic rather than social, and for neglecting the contribution of external reality to meaning. He construes the standard theory as based on the two assumptions that to know a term's intension is to be in a particular psychological state and that intension determines extension. Hence, it is argued, speakers who are in the same mental state when uttering a word share both its intension and its extension.

Putnam uses the twin earth (TE) thought experiment to attack this theory. TE resembles earth down to the smallest detail except that the liquid functioning as water, and called 'water', on TE is not $\mathrm{H_2O}$, but a different chemical compound. Since there is no reason to ascribe different mental states to a person using 'water' on earth, and their counterpart on TE, or, at least, there was no such difference prior to the emergence of chemistry as a science, this constitutes an example of people using the same word when in identical mental states, but ascribing different meanings to it: one refers to water, the other to the TE equivalent. One need not, however, travel as far as TE to find examples of similar mental states differing in extension. Putnam testifies that he is unable to tell a beech from an elm: his mental image of the two is the same, yet the extensions, and the meanings, of 'elm' and 'beech' are quite different in his idiolect. Can Putnam refer to an elm though unable to identify one? His answer introduces the concept of the division of linguistic labour: it is sufficient that experts can distinguish an elm from a beech, and unnecessary that each member of the linguistic community be able to do so.

The assumption underlying the TE example is that words like 'water' always refer to the stuff we call 'water' in the actual world. Though it is neither analytic nor even irrevisable that water is $\mathrm{H_2O}$, the extensions of 'water' on TE and on earth cannot be identical. To complete the account of meaning, one has to address the question of how extension is actually fixed. These aspects of Putnam's conception are related to Kripke's work on reference and rigid designation (see Kripke, S.A.; Proper names; Reference §2), but while Kripke was thinking mainly of proper names, Putnam's concern is the meaning of scientific terms. If the extensions of these terms are fixed by the theories in which they figure, then extensions will be liable to change with theoretical change. This argument can be taken as an encapsulation of Kuhn's relativism: different theories refer to different entities, and are therefore incommensurable (see Incommensurability). Viewing it as a *reductio ad absurdum* of the theory of meaning on which it rests, Putnam recommends replacement of this theory with a variant of Kripke's causal theory of reference that emphasizes not only the causal relation between speakers and what they refer to, but also such social and pragmatic factors as shared stereotypes, reliance on experts, discretion and charity (see Content: wide and narrow).

4 The philosophy of quantum mechanics

Heisenberg's uncertainty principle imposes a limit on the precision with which the values of certain pairs of physical parameters, such as position and momentum, or two spin components, can be measured simultaneously. On the Copenhagen interpretation, this principle implies that it is meaningless to ascribe simultaneous sharp values to such pairs of physical parameters, whether or not they are actually being measured. Since, however, when any one magnitude is measured separately, a sharp value is obtained, it appears that it is measurement itself which creates the transition, better known as the collapse, from the indeterminate to the well-defined state. If so, measurement does not reflect a state objectively existing prior to measurement, but points to a state of its own creation (see Quantum measurement problem). Both the inference from the impossibility of measurement to the meaninglessness of concepts and the non-classical understanding of measurement, offend the realist. In *The Logic of Quantum Mechanics*, Putnam proposed overcoming these difficulties by adopting a nonclassical logic first suggested in the context of quantum mechanics (QM) by Birkhoff and von Neumann in 1936, and developed by Finkelstein in the 1960s. The suggested logic is non-distributive - from $p \cdot (q_1 \vee q_2)$ we cannot, in general, conclude that $p \cdot q_1 \vee p \cdot q_2$. If *p* states that the system has a well-defined value *p*, of a physical magnitude P, and $q_1, q_2, \ldots, q_n$ describe all possible values of an incompatible quantity Q, then the uncertainty principle entails that $p \cdot q_i$ is false for any *i*. Yet, assuming that *p* obtains, (ascertained by measurement, say) we cannot conclude, as we classically would, that $q_1 \vee q_2 \vee \ldots \vee q_n$ is false. In fact, while $p \cdot q_i$ is a quantum logical contradiction,

$q_1 \vee q_2 \vee ... \vee q_n$ is a quantum logical tautology. The magnitude Q always has a well-defined value, which its measurement will reveal, and no collapse is called for (see Quantum mechanics, interpretation of; Quantum logic).

In light of the traditional gulf between factual and logical truth, the idea that logic can be revised on the basis of empirical considerations is revolutionary. Putnam saw this situation as analogous to the merging of physics and geometry into an interdependent whole in the framework of general relativity.

Quantum logic raises several questions. First, it is not clear that it is a logic, a way of reasoning, rather than a calculus that happens to fit the structure of the Hilbert space of QM. Second, the idea that one can save realism by rejecting classical logic, generally seen as constitutive of realism, seems paradoxical. Though intended to strengthen the analogy with logic, Putnam's operational definition of the quantum-logical operators obscures the connection to realism. Third, work on the foundations of QM by theorists such as Kochen and Specker, and Bell, put unbearable strain on the realist interpretation of QM (see Bell's theorem). Indeed, in *Quantum Mechanics and the Observer*, when Putnam had already moved away from his early realism, he assumed a verificationist understanding of quantum logic. The main point of that paper, however, is to argue for yet another interpretation of QM - perspectivism, attributed by Putnam to von Neumann. Like quantum logic, perspectivism is a way of avoiding the collapse of the wave-function upon measurement. Collapse, on this interpretation, is not a physical process but an epiphenomenon created by the shift from one perspective to another. Thus, when a system M performs a measurement on another system S, we can either view M as interfering with S from without, inducing a collapse of the wave-function of S, or view S and M as a unified system obeying QM, and the external observer as interfering with it and making its wave-function collapse. Ultimately, Putnam argues, different perspectives are empirically equivalent and congruent with the predictions of QM; hence, they are equally legitimate. But perspectives exclude each other in the sense that statements belonging to different perspectives cannot be combined to form a quantum state. Realism can be sustained within each perspective, but not across perspectives. Though this seemed an attractive way to retain 'internal' realism while forgoing metaphysical realism, upon realizing that, in some cases, different perspectives are not empirically equivalent Putnam became dissatisfied with perspectivism.

5 Mathematics and necessary truth

Putnam did significant work in mathematics, collaborating with Martin Davis and Julia Robinson in the late 1950s on proving the unsolvability of Hilbert's tenth problem, which sought an algorithm deciding the solvability of diophantine equations. The proof was completed by Yuri Matiyasevich in 1970.

The nature of logical and mathematical truth has been one of Putnam's ongoing concerns, yielding several different positions. Throughout, he rejects the standard alternatives, platonism and conventionalism. The former, is, he maintains, given twentieth-century physics, obsolete; the latter, empty: as Carroll, Wittgenstein and Quine pointed out, conventions cannot ground logic because logic is required for their application (see Mathematics, foundations of). In 'It Ain't Necessarily So', Putnam proposed replacing necessary truth with the more flexible, context-dependent notion of relative necessity, in line with his suggestion, raised regarding QM, that logic is empirical. Later, in 'Analyticity and Apriority', he argued that at least some logical truths are constitutive of rationality and, as such, cannot be rationally criticized or revised. This view is further elaborated in *Rethinking Mathematical Necessity*, where Putnam represents logical truths as 'formal presuppositions of thought' rather than as truths in the ordinary sense (see Analytic and synthetic).

6 Functionalism

In a series of papers beginning in 1960, Putnam proposed a fresh approach to the philosophy of mind, functionalism, seeking to secure the autonomy of mind without positing a non-physical mind-substance. ' The question of the autonomy of our mental life does not hinge on and has nothing to do with that all too popular…question about matter or soul-stuff. We could be made of Swiss cheese and it wouldn't matter' (1975b: 291). What matters, Putnam argued, is functional organization. Putnam's guiding analogy for functional organization was the computer (Turing machine). Evidently, different machines need not share the same hardware to carry out the same computation. Similarly, Putnam claimed, pain-states, or jealousy-states, can be functionally alike though physically different. In other words, each pain-token has a physico-chemical realization, but no reduction of the type, pain, to a given physico-chemical state is assumed. The computer analogy suggested that

mental states are computational states, characterized syntactically, the projected research programme being to provide the ’software’ for their interaction.

In the late 1970s, Putnam began to reconsider this proposal. First, there were considerations of meaning (§3). Thinking of something seems like a simple enough example of a mental state, but if, as Putnam argued, ‘meanings just ain’t in the head’, then meanings cannot be identified with internal computational states. The response of some theorists, notably Fodor and Block, was to use the distinction between narrow and wide content, presented by Putnam in The Meaning of Meaning, to save the computational picture. While acknowledging the contribution of physical and cultural environment to meaning in the wide sense, they held on to computationalism with respect to meaning in the narrow sense. Putnam’s concern over intentionality led him to reject this solution. As he argued in *Representation and Reality*, narrow-content computationalism is still an attempt to reduce the intentional to the non-intentional. But since even the ascription of meaning in the narrow sense involves interpretation, the attribution of beliefs, charity and reasonableness, eliminating intentionality is untenable. Functionalism had conceived the computational level as autonomous, that is, irreducible to, even if supervenient on, the physico-chemical level. Putnam’s critique of functionalism makes an analogous point with regard to the autonomy of the mental *vis-à-vis* the computational (see Computational theories of mind; Functionalism; Reductionism in the philosophy of the mind).

7 The incoherence of metaphysical realism

In 1976 Putnam’s philosophy underwent a major shift; rejecting what he referred to as ‘metaphysical’ realism, he adopted ’internal’ realism in its stead. The attack on metaphysical realism, first presented in the final chapter of *Meaning and the Moral Sciences* (1978) is elaborated on in *Models and Reality* and in *Reason, Truth and History* (1981).

Just as establishing the objectivity of reference was central to Putnam’s earlier realism, the dispersion of reference into a plurality of possible relations is at the heart of his later criticism. The argument draws on model-theoretic considerations. The Löwenheim-Skolem theorem entails that a first-order theory, rich enough to contain arithmetic, does not determine its models up to isomorphism (see Löwenheim-Skolem theorems and nonstandard models). Putnam extrapolates: even an ideal theory of the world, complying with all empirical and theoretical constraints, will not define a unique model, that is, a unique reference relation. In particular, causality, previously seen as anchoring language in reality, now becomes just another relation, and hence open to interpretation.

Putnam’s point is not to embrace scepticism. His argument is that both the metaphysical realist who purports to have a theory of everything, including the ’correct’ reference relation, and the sceptic who undertakes to refute that theory, are making the same mistake - they are assuming a non-existent vantage point external to any language or description-scheme. From the internal perspective, questions about reference cannot arise, ’chair’ refers to chairs, ’cherry’ to cherries. ‘ To speak as if *this* were my problem, "I know how to use my Language, but, now, how shall I single out an interpretation?" is to speak nonsense. Either the use *already* fixes the "interpretation" or *nothing* can’ (1983: 24).

Another argument against scepticism is found in *Reason, Truth and History* (1981), where considerations of reference and intentionality lead Putnam to conclude that the sceptic’s favourite fantasy - that we are all brains in a vat - is self-refuting (see Scepticism). The repudiation of scepticism is a recurrent theme uniting Putnam’s earlier and later work. The strategy, however, changes; Putnam’s responses to Quine, who has also invoked the Löwenheim-Skolem theorem, illustrate these. Whereas Putnam initially tried to reduce indeterminacy by increasing the number of constraints on an adequate translation, he later came to see the problem itself as a sceptical variation on a misguided metaphysics.

8 Internal realism

Internal realism is also referred to by Putnam as ’pragmatic’ or ’natural’ realism. Like the American pragmatists, Putnam holds that commonsense realism, which he respects, requires no ’deeper’ philosophical foundation. He also follows pragmatism in rejecting the fact/value dichotomy, and opposes moral relativism as firmly as he opposes cognitive relativism (see Pragmatism). A critic of reductionism and naturalism, he maintains that there are forms of nonscientific knowledge and that reason and morality cannot be naturalized. He distances himself from

’end of philosophy’ philosophies, and seeks to renew philosophy so as to reconnect it with human challenges and aspirations.

Internal realism is meant to avoid the pitfalls of both metaphysical realism and relativism, but Putnam has been accused of slipping back into these polar positions. At one point he was close to identifying internal realism with verificationism, with long-term warranted assertibility replacing the notion of truth. Later he rejected this view, as well as other attempts to eliminate truth.

While a long philosophical tradition sees realism and antirealism as exhaustive alternatives, Kant, the American pragmatists, and Wittgenstein, all strove to overcome this dichotomy. Putnam identifies with the latter camp. In his earlier writings, he represented realism as a kind of explanatory hypothesis, on a par with scientific theories. Internal realism, however, is not intended to play an explanatory role. Wittgenstein’s influence, and particularly, his suspicion of philosophical theories, is perceptible here, not only in the subtle similarity between the extended Löwenheim- Skolem theorem and the rule-following paradox, but also in the change of perspective that constitutes the response to these related problems. When we realize that internal realism is not an alternative theory but a call for a change of perspective, it emerges that much of Putnam’s earlier work, far from being undermined by his later philosophy, finds its proper place within it.

See also: Reference §§3-4

YEMIMA BEN-MENAHEM

List of works

Most of Putnam’s papers, and all of those cited in the text, are included in the collections - *Philosophical Papers, volumes I-III* and the two volumes edited by J. Conant. Where the title has been changed, the text cites the title appearing in these volumes.

Putnam, H. and Benacerraf, P. (eds) (1964) *Philosophy of Mathematics: Selected Readings*, New Jersey: Prentice Hall.(Introduction by the editors. Includes classic works in the foundations of mathematics and the philosophy of mathematics by leading mathematicians such as Hilbert, Brouwer and Gödel.)

Putnam, H. (1971) *Philosophy of Logic*, New York: Harper & Row.(Putnam argues for a realist position on the problem of the existence of mathematical objects.)

Putnam, H. (1975a) *Philosophical Papers*, vol. I, *Mathematics, Matter and Method*, Cambridge and New York: Cambridge University Press. (This volume contains papers on the philosophy of mathematics and the philosophy of science. It includes ‘It Ain’t Necessarily So’, which develops the notion of relative necessity, ‘An Examination of Grünbaum’s Philosophy of Geometry’, which argues against a conventionalist interpretation of the theory of relativity, and ‘The Logic of Quantum Mechanics’, advocating quantum logic.)

Putnam, H. (1975b) *Philosophical Papers*, vol. II, *Mind, Language and Reality*, Cambridge, and New York: Cambridge University Press. (This volume includes one of Putnam’s major contributions to the philosophy of language, ‘The Meaning of Meaning’, as well as ‘Explanation and Reference’ and ‘The Refutation of Conventionalism’, all of which expound a realist conception of meaning. It also includes papers in the philosophy of mind, in particular papers introducing functionalism, such as ‘Philosophy and Our Mental Life’ and ‘The Mental Life of Some Machines’.)

Putnam, H. (1978) *Meaning and the Moral Sciences*, London: Routledge & Kegan Paul. (Part One contains the John Locke Lectures, Oxford 1976, which explore the interconnections between explanation, theories of truth, and realism. Part Two marks the change in Putnam’s position from realism to ’internal realism’.)

Putnam, H. (1981) *Reason, Truth and History*, Cambridge, and New York: Cambridge University Press.(The book propounds ’internal realism’, arguing that it avoids the difficulties of metaphysical realism on the one hand, and scepticism on the other.)

Putnam, H. (1983) *Philosophical Papers*, vol. III, *Realism and Reason*, Cambridge, and New York: Cambridge University Press.(This volume reworks many of the themes previously dealt with by Putnam in the light of his new position. It includes ‘Models and Reality’, which pursues the implications of the Löwenheim-Skolem theorem; ‘Analyticity and Apriority: Beyond Wittgenstein and Quine’, which rejects some sceptical consequences of Quine’s and Wittgenstein’s positions; ‘Why Reason Can’t Be Naturalized’, which argues against naturalism; and ‘Quantum Mechanics and the Observer’, which argues for perspectivism, and abandons the earlier realist interpretation of quantum mechanics in terms of quantum logic.)

Putnam, H. (1987) *The Many Faces of Realism*, La Salle, IL: Open Court.(The Paul Carus Lectures, Washington 1985. Putnam’s ’internal realism’ is examined here in the context of American pragmatism, in particular

Peirce’s philosophy.)
Putnam, H. (1988) *Representation and Reality*, Cambridge, MA: MIT Press.(Putnam reconsiders functionalism, his early philosophy of mind, in the light of his views on meaning, and his critique of naturalism.)
Putnam, H. (1990) *Realism with a Human Face*, ed. J. Conant, Cambridge, MA: Harvard University Press.(A collection of essays by Putnam, selected and introduced by the editor. The essays in this volume are divided into three groups: metaphysics; ethics and aesthetics; studies in American philosophy. The rejection of the fact/value dichotomy is a central theme of this collection.)
Putnam, H. (1992) *Renewing Philosophy*, Cambridge, MA: Harvard University Press.(This book grew out of the Gifford Lectures, St Andrews 1990. It takes issue with scientism in several areas such as artificial intelligence and ethics, while maintaining the possibility of a cognitive relation to reality.)
Putnam, H. (1994) *Words and Life*, ed. J. Conant, Cambridge, MA: Harvard University Press.(A collection of essays by Putnam, selected and introduced by the editor. Some of the essays in this volume, such as ‘Rethinking Mathematical Necessity’ reflect Putnam’s growing attraction to the later Wittgenstein. Others discuss pragmatism, logical positivism, the philosophy of mind and the role of philosophy in our lives.)
Putnam, H. (1995) *Pragmatism*, Cambridge, MA, and Oxford: Blackwell.(First published in Italian as *Il Pragmatismo: Una Questione Aperta*, Rome, Bari: Gius Laterza & Figli Spa, 1992. Putnam’s version of pragmatism. Includes a bibliography.)

References and further reading

Hill, C.S. (ed.) (1992) ’The Philosophy of Hilary Putnam’, *Philosophical Topics* 20 (1). (Essays by D.A. Anderson, A. Bilgrami, N. Chomsky, J. Conant, B. Dreben, G. Ebbs, R. Healey, G.J. Massey, J. McDowell, R.W. Miller, and A. Sidelle, and replies by Putnam.)
Clark, P. and Hale, B. (eds) (1994) *Reading Putnam*, Oxford: Blackwell.(Essays by S. Blackburne, G. Boolos, M. Dummett, M. Hallett, C. Fuhl and C. Glymour, M. Redhead, T. Ricketts, D. Wiggins, and C. Wright, and replies by Putnam.)

Pyrrho (*c*.365-*c*.275 BC)

The Greek philosopher Pyrrho of Elis gave his name first to the most influential version of ancient scepticism (Pyrrhonism), and later to scepticism as such (pyrrhonism). Like Socrates, he wrote nothing, despite which - or thanks to which - he too became one of the great figures of philosophy. Although he has vanished behind his own legend, he must have helped nurture that legend: his unique personality palpably exercised an unequalled fascination on his acquaintances, and through them, on many others. We possess, thanks especially to Sextus Empiricus, extensive documentation of what can be called 'Neo-Pyrrhonian' scepticism, because from the time of Aenesidemus (first century BC) it invoked Pyrrho as its patron saint. But Pyrrho's own thought is hard to recover. The documentary evidence for him is mainly anecdotal, and the principal doxography is more or less directly dependent on his leading disciple Timon of Phlius, who managed to present himself as Pyrrho's mere 'spokesman', but who was in fact perhaps rather more than that. The main question, which is still unanswered, is whether Pyrrho was primarily or even solely a moralist, the champion of an ethical outlook based on indifference and insensibility, or whether he had already explicitly set up the weaponry of the sceptical critique of knowledge which underlies the epistemological watchword 'suspension of judgment'.

1 Life

According to Diogenes Laertius (IX 61-70), who is following a life of Pyrrho written around 225 BC by Antigonus of Carystus, Pyrrho had been an obscure painter before his studies, which were first with Bryson (probably the Megarian philosopher of that name), then with Anaxarchus. Anaxarchus was a companion to Alexander the Great on his Asian campaign, and brought Pyrrho along too. Alexander's biographers speak quite often of Anaxarchus, but never of Pyrrho. On his return to Greece, Pyrrho lived in the countryside near Elis, surrounded by a group of admirers, but hardly seeming a head of school.

The numerous anecdotes about him reveal a split between two concurrent images. On the one hand we hear of Pyrrho as an eccentric: indifferent towards himself as towards others, he leaves himself entirely unprotected, taking account neither of sensations nor of the beliefs which guide practical life (fr. 6). He behaves as a fakir (fr. 16) who recalls his encounters with the 'gymnosophists' or 'naked sages' of India (frs 1A, 10). He thus personifies a 'rustic' kind of scepticism, which subjects to 'suspension of judgment' not only learned doctrines but also beliefs found in ordinary life.

A second image, endorsed by Aenesidemus, is of a Pyrrho who suspends judgment when philosophizing but does not lack foresight in his practical life (fr.7). He lives in a manner which is modest, peaceful and relatively conformist (fr. 14), along with his sister and his farmyard animals, highly esteemed by his fellow citizens (fr. 11). He thus stands for an 'urbane' scepticism, which outlaws all doctrinal assertions but leaves intact the instinctive beliefs of everyday life.

This duality no doubt nurtured the debate as to what was the Pyrrhonists' chief good (*telos*) (Diogenes Laertius IX 108): insensibility (*apatheia*) according to some, gentleness (*praotēs*) according to others. It is possible, moreover, that both the 'rustic' and the 'urbane' image had their roots in the complexity of one and the same personality, as illustrated by some anecdotes. For example, the story goes that Pyrrho once fled from a vicious dog, and, reproached for violating his own principles of indifference, replied 'It is difficult to strip yourself completely of being human' (fr. 15A-B). We should take it that his ambition really was to escape the human condition, but that he knew himself incapable of unfailingly achieving this.

2 The legends

More by his personality than by his ideas, no doubt, Pyrrho created a sensation: he gave the impression of having found a new way of being happy. A disciple, Timon, asks him the secret of his superhuman serenity (frs 60-1). Another, Nausiphanes, reported that one of his own pupils used to question him eagerly about Pyrrho's conduct (fr. 28); the pupil's name was Epicurus (§1). When Cicero speaks of Pyrrho, he never describes him as a sceptical critic of knowledge (despite having the opportunity to do so), but always as an absurdly rigorous moralist (frs 69A-M), even more radical than the extremist Stoics Ariston of Chios (§2) and Herillus. Pyrrho must, nevertheless, have cared as little about the difference between knowledge and ignorance as about anything else;

Timon contrasts his tranquil lack of curiosity to 'the empty wisdom of the sophists' (fr. 60). Later, his condemnation of the branches of knowledge as useless could be presented as a way of condemning knowledge itself as impossible.

His chronological position inevitably favoured this metamorphosis. Aristotle's junior by twenty years, Pyrrho was still in his forties when Aristotle and his former pupil Alexander the Great both died (322 and 323 BC), that is, at the start of the Hellenistic age. He was a generation older than Epicurus and the Stoic Zeno of Citium, the founders of the two great new schools of this era (see Epicureanism; Stoicism). And he was two generations older than Arcesilaus, who restored scepticism to the Platonic Academy. He was therefore perfectly placed to appear a pivotal thinker in the eyes of posterity. Just before him, Aristotle seemed to have brought to completion an era of intellectual audacity, in which thinkers had vied in cognitive ambitions without necessarily considering their multiple disagreements a fatal drawback; when they had discussed knowledge, they had tended to suppose that its existence and possibility went without saying, and instead set themselves the task of analysing its nature, its tools and its methods. After Pyrrho, on the other hand, and throughout Hellenistic philosophy, there was a concern as to whether we really do have cognitive access to the world: the new schools placed the problem of the 'criterion of truth' at the head of their agenda, and their solutions to it marked the crucial distinction between dogmatists and sceptics (Hellenistic philosophy). It looks as if, in the intervening time, a radical new challenge to the very possibility of knowledge had been thrown down. When the author of this challenge is sought, Pyrrho emerges as the perfect candidate. One might even speak of a duel fought between Aristotle and Pyrrho. When Aristotle (*Metaphysics* IV) speaks of the manner of speech, thought and action to which those who deny the law of non-contradiction are condemned, he uses expressions which recur in the reports of Pyrrho's behaviour and in the formulas which he recommends using. It is as if Pyrrho had found a way to show that one could perfectly well live, speak and think in just the way Aristotle had said one could not (unless these coincidences were a later imposition designed to make Pyrrho look like the hero of an unprecedented philosophical rebellion.)

To correct this exaggerated portrayal, it is enough to notice that Pyrrho's contemporaries and successors, even the most enthusiastic of them, do not present him as an innovative theorist. Timon, despite doing more than anyone else to place him on a pedestal, nevertheless does not make him stand altogether alone: it is he who first assembles that gallery of ancestors (the Eleatics, Democritus, Protagoras and, above all, Xenophanes) which is thereafter associated with Pyrrho by the doxographers and authors of 'successions' (see Doxography).

Much later, from Aenesidemus to Sextus Empiricus, Pyrrho's name reappeared in the titles of sceptical works (*Pyrrhonian Discourses*, *Pyrrhonian Sketches*) as emblematic of a scepticism senior to that of Arcesilaus and altogether uncontaminated by the New Academy: after long neglect, it became a prestigious flag to wave. In this late era little was still known about Pyrrho. It was in any case sensible for a Pyrrhonist to disavow that name, as did Theodosius (second century AD), who remarked that 'given that the movement of someone else's thought is undiscoverable, we will never know what Pyrrho's mental disposition was; and not knowing that, we could not label ourselves Pyrrhonists' (fr. 41). It is notable that Sextus' only response to this objection is the mild one that scepticism can be called Pyrrhonism 'owing to the fact that Pyrrho seems to us to have devoted himself to the activity of sceptical inquiry (*skepsis*) more solidly and manifestly than his predecessors' (fr. 40) (see Pyrrhonism).

One may wonder, in the circumstances, whether Pyrrho's posthumous fate does not result from a sort of collage: the overwhelming originality of his ethical message, immediately recognized and celebrated, seemed to demand as its complement a radical critique of epistemology. This latter may have been bestowed on him retrospectively, in the way that philosophers often get saddled with the real or imagined consequences, or premises, of their ideas.

3 The key testimony

It would clearly be presumptuous, and would perhaps make little sense, to pretend to reconstitute the 'real' Pyrrho behind, or in front of, his multiple reflections. Reale (1981) has distinguished no fewer than eight defensible types of interpretation of Pyrrho's philosophy. Some separate him only narrowly, others widely, from the epistemological and phenomenalist scepticism of the neo-Pyrrhonists. For some he is above all an ascetic proponent of indifference; for others a nihilistic theorist of pure appearance and radical critic of all ontology. Such diversity is all the more surprising given the existence of a text (fr. 53) universally considered the centrepiece of Pyrrho's doxography, and respectable as much for its pedigree (it is cited by Eusebius from the late first century BC philosopher Aristocles, who himself drew its substance, if not its letter, from Timon) as for its content: an

overview of Pyrrho’s philosophical project. Clear and instructive in appearance, at a first reading this text seems to attest that Pyrrhonism does indeed have a fundamentally ethical goal, and that there are two closely linked means to its achievement: practical indifference and abstention from judgment, the latter going hand in hand with a complete disqualification of the ordinary instruments of knowledge. But the articulation of these different elements raises difficulties; and by scrutiny of the text we may be able to perform a task widely judged by interpreters to be impossible or useless: to separate Pyrrho’s original message from Timon’s own contribution.

Anyone who is on the verge of being happy, says Timon, should consider three questions: what the nature of things is, how we should be disposed towards them, and what benefit those who adopt this attitude will derive from it. The sequence of these three questions is itself puzzling: no latter-day sceptic would inquire about ‘the nature of things’ before asking himself whether we can know it. Pyrrho nevertheless does ask this; and he replies - in the one phrase of the text which Timon expressly attributes to him - that ‘things are entirely undifferentiated, undetermined and undecided’. This no doubt means that it is we who introduce the differences which appear to us to distinguish them (fr. 64). But what differences? He might mean all differences, including those of colour and so on as much as those between values. Or he might mean only these latter. Another summary (fr. 1A) has Pyrrho saying ‘that nothing is fair or foul, just or unjust, and that likewise in every case nothing is really (this or that), and that it is by custom and habit that people do everything they do; for nothing is any more this than that’. These generalizations seem restricted to the sphere of practical ethics.

Aristocles’ summary, after characterizing ‘things’ as ‘undifferentiated’, strangely adds that ‘for this reason, our sensations and beliefs are neither true nor false’. Here again, a latter-day sceptic would say otherwise, indeed the reverse: that our sensations and beliefs are untrustworthy, and that therefore ‘things’ are in their nature undifferentiated. Some have even emended the text to make it say this. However, once we notice that the grammar of the text attributes this bizarre inference to Timon, not to Pyrrho himself, we may prefer not to emend it, but to see in its very incongruity the traces of Timon’s own intervention, a somewhat clumsy epistemological graft onto the original Pyrrhonian trunk.

The second point on the Pyrrhonian agenda admittedly does not help much with testing this hypothesis. In the face of undifferentiated ‘things’, the attitude that we should take is that of respecting their non-differentiation: we must be ‘unopinionated, uninclined, undisturbed’ either positively or negatively towards anything. These expressions can be interpreted as describing a refusal to affirm, just as much as a refusal to choose. And the same equivocation between theory and practice persists in the description of the discourse that goes with this behaviour: it is appropriate to say, of each thing, that ‘it is no more (this or that) than it is not it’, or again (no doubt with the implication ‘if it is necessary at all costs to say something positive or negative’) that it is appropriate immediately to negate our assertion, saying that ‘it is (this or that) and at the same time that it is not it’, or that ‘it neither is (this or that) nor is not it’. These expressions are remarkable; but they do not make it clear whether ‘this or that’ covers all possible predicates, or only those which, like ‘desirable’ and ‘undesirable’, concern the sphere of action.

One who is ‘on the verge of being happy’ must last consider - and this is the final phase of the Pyrrhonian programme - ‘what will happen to those who are disposed thus’ (that is, ‘without opinions’ and so on, as already defined). Whoever is reformed by Pyrrhonian training, says Timon, will achieve first *aphasia* (not in its modern sense of mutism, but a sort of non-assertive, non-committal use of language), then *ataraxia* (complete lack of disturbance, perfect unworriedness). It is worth noticing that the attainment of this benefit is here described as predictable. In later Pyrrhonism (Sextus Empiricus, *Outlines of Pyrrhonism* I 28-9), and already in Timon (Diogenes Laertius IX 107), there was to be on the contrary an insistence that *ataraxia* results unexpectedly, by a happy chance, from suspension of judgment. Might this difference not be the result of the epistemological graft hypothesized above? That *ataraxia* should result from practical indifference makes sense: freed of all preferences, we will no longer undergo the feelings which upset us. But that *ataraxia* should result from suspension of judgment makes less sense: the renunciation of all certainty might seem, on the contrary, to condemn us to everlasting anxiety. The appeal to ‘happy chance’, an innovation on the rational planning offered by Pyrrho’s programme, may therefore reflect the new complexity introduced into the sceptic’s itinerary by the epistemological detour. If this analysis of the key testimony in Aristocles is correct, we are entitled to say, even more firmly than before, that Pyrrho was not the first of the Pyrrhonists.

JACQUES BRUNSCHWIG

References and further reading

Ausland, H.W. (1989) 'On the Moral Origin of the Pyrrhonian Philosophy', *Elenchos* 10: 359-434.(A strongly ethical interpretation of Pyrrho's thought.)

Bett, R. (1994) 'Aristocles on Timon on Pyrrho: the text, its logic, and its credibility', *Oxford Studies in Ancient Philosophy* 12: 137-81.(An acute reading of the key testimony, contemporary and independent from Brunschwig (1994).)

Brochard, V. (1887) *Les Sceptiques grecs*, Paris: Alcan; repr. 1923.(A classic which has aged well; includes the first systematically presented ethical interpretation of Pyrrho.)

Brunschwig, J. (1994) 'Once Again on Eusebius on Aristocles on Timon on Pyrrho', in *Papers in Hellenistic Philosophy*, Cambridge: Cambridge University Press, 190-211.(Defends the interpretation favoured in §3 of the entry.)

Diogenes Laertius (*c*. early 3rd century BC) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius, Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(See IX 61-108 for a life of Pyrrho, which includes an extensive doxography on neo-Pyrrhonism.)

Flintoff, E. (1980) 'Pyrrho and India', *Phronesis* 25: 88-108.(Investigates possible Indian philosophical influences on Pyrrho.)

Giannantoni, G. (ed.) (1981) *Lo scetticismo antico (Ancient Scepticism)*, Naples: Bibliopolis, 2 vols.(Important collection of articles, with a full bibliography on ancient scepticism, 1880-1978.)

Hankinson, R.J. (1995) *The Sceptics*, L ondon and New York: Routledge.(The best modern book-length study.)

Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press, 2 vols.(Volume 1 contains translations of the principal sources, with philosophical commentary (Early Pyrrhonism, pages 13-24). Volume 2 contains Greek and Latin texts with notes and bibliography (Early Pyrrhonism, pages 1-17). An outstanding work in all respects.)

Pyrrho (*c*.365-*c*.275 BC) Fragments, ed. F. Decleva Caizzi, *Pirrone - Testimonianze*, Naples: Bibliopolis, 1981. (Standard edition of the testimonies on Pyrrho from which his fragments are cited; includes Italian translation and valuable commentary.)

Reale, G. (1981) 'Ipotesi per una rilettura della filosofia di Pirrone di Elide' (Hypothesis on a Rereading of the Philosophy of Pyrrho of Elis), in G. Giannantoni (ed.), *Lo scetticismo antico*, Naples: Bibliopolis, 1981, vol. 1, 243-336.(Useful critical survey of interpretations, with an attempted genealogy of the paradoxes of a Janus-faced Pyrrho who is both dogmatic moralist and sceptic.)

Robin, L. (1944) *Pyrrhon et le scepticisme grec (Pyrrho and Greek Scepticism)*, Paris: Presses Universitaires de France; repr. New York: Garland, 1980.(Careful and balanced study.)

Stopper, M.R. (1983) 'Schizzi pirroniani' (Pyrrhonist Sketches), *Phronesis* 28: 265-97. (Important review of Giannantoni (1981), in English, with powerful defence of the traditional 'neo-Pyrrhonian' interpretation of Pyrrho.)

Pyrrhonism

Pyrrhonism was the name given by the Greeks to one particular brand of scepticism, that identified (albeit tenuously) with Pyrrho of Elis, who was said (by his disciple Timon of Phlius) to have declared that everything was indeterminable and accordingly to have suspended judgment about the reality of things - in particular whether they were really good or bad. After Timon's death Pyrrhonism lapsed, until revived by Aenesidemus. Aenesidemus held that it was inadmissible either to affirm or to deny that anything was really the case, and in particular to hold, with the Academic sceptics, that certain things really were inapprehensible. Instead, the Sceptic (the capital letter denotes the Pyrrhonists, who adopted the term, literally 'inquirer', as one of the designations for their school) should only allow that things were no more the case than not, or only so under certain circumstances and not under others. Aenesidemean Scepticism took the form of emphasizing the disagreement among both lay people and theoreticians as to the nature of things, and the fact that things appear differently under different circumstances (the various ways of doing this were systematized into the Ten Modes of Scepticism); the result was meant to be suspension of judgment about such matters, which would in turn lead to tranquillity of mind. Thus 'Scepticism' denotes a particular philosophical position, not simply, as in modern usage, that of any philosopher inclined towards doubt. Later Pyrrhonists, notably Agrippa, refined the Sceptical method and concentrated on undermining the dogmatic (that is, anti-Sceptical) notion of the criterion - there is no principled way to settle such disputes without resorting to mere assertion, infinite regress or circularity. We owe to Sextus Empiricus our most complete account of Pyrrhonian argument and the clearest exposition of the Pyrrhonian attitude. Faced with endemic dispute, Sceptics reserve judgment; but this does not render life impossible for them, since they will still react to the way things appear to be, although without believing in any strong sense that things really are as they seem. Furthermore, when Pyrrhonians describe their affective states, they do so undogmatically - and the Sceptical slogans ('I determine nothing', 'nothing is apprehended', and so on) are to be understood in a similar way, as merely reporting a state of mind and not expressing a commitment. Thus the slogans apply to themselves, and like cathartic drugs are themselves purged along with the noxious humour of dogmatism.

1 History

Greek scepticism is inextricably associated with the name of Pyrrho; the ancients themselves dubbed the most enduring and interesting form of scepticism of their times 'Pyrrhonism' after its eponymous origin. Yet surprisingly little is known about Pyrrho the man, and it is possible that he was not even a Sceptic (in the strict sense of the term) at all (see Pyrrho §3). He lived from *c*.365 to *c*.275 BC, and his name became a byword for philosophical detachment from the ordinary concerns of life. Legend has it that his friends had to prevent him from walking over cliffs and under passing traffic; this and other amusing if apocryphal stories are recorded in the biography by Diogenes Laertius. He wrote nothing, and most of what we know of him derives from the writings of his disciple and amanuensis Timon. According to Timon, Pyrrho held that to be happy we must confront three questions: How are things by nature? What attitude should we adopt towards them? What will be the outcome for those who have this attitude? He answered that all things were equally indifferent, unmeasurable and undecidable, and that our perceptions told us neither truths nor falsehoods about the way things really are. The upshot of this is that we should be free of all opinion and commitment. The proper attitude to anything is to suppose that it no more (*ou mallon*) is the case than not. As a result we should simply acquiesce in the way things appear without having any strong beliefs concerning the way they actually are.

These remained the characteristic attitudes of Pyrrhonism throughout its history. However, original though Pyrrho undoubtedly was (even if the tradition suggests that he learned philosophical detachment from Indian philosophers while following Alexander the Great's expedition with Anaxarchus), his scepticism was not entirely without precedent. Ever since Xenophanes (§5), Greek thinkers had puzzled over the nature and scope of human knowledge, and whether it was genuinely possible. Some fragments of Democritus in particular suggest a cautious attitude to the possibility of human understanding, and Democritus (§3) employed the formula 'no more' on occasion to indicate a sceptical refusal to commit himself one way or the other on some question. Moreover, Aristotle is clearly aware of sceptical challenges to our ability to rely upon our senses (*Metaphysics* IV 4-5), although he dismisses them as mere captiousness (it is possible that he knew of Pyrrho's scepticism).

There is no hint of a concerted and systematic attempt to question our justification for belief in the natures of

things before Pyrrho, however. Indeed Pyrrho himself may not have mounted a general assault on the reliability of the senses (although Timon evidently did), so much as a limited attack on our pretensions to ethical knowledge, or, more generally, knowledge of matters of value; in later antiquity, Pyrrho's name was certainly associated primarily with ethical (in the broad Greek sense) issues; and it was principally as a role model for a certain way of life, one in which tranquillity is to be achieved by a refusal to allow oneself strong commitments of any kind, that Pyrrho's fame was maintained in later antiquity.

After Timon's death Pyrrhonism as such died out, although the New Academy of Arcesilaus and Carneades kept alive the sceptical tradition for the next two centuries, in the course of their long and tortuous epistemological conflict with the Stoics. The latter held that empirical certainty was attainable, on the basis of 'apprehensive' impressions whose clarity and distinctness were assurances of their truth (see Stoicism §12); the Academic sceptics countered by arguing that there was no true impression that was such that there could not be a false one indistinguishable from it in all internal characteristics: consequently, there was no criterion by which to judge whether or not an impression was indeed apprehensive.

By the first century BC both sides had so modified their positions as a result of the dialectical sparring that it was becoming increasingly difficult to tell the two schools apart, so much so that when (a little after 90 BC) Philo of Larissa relaxed Academic strictures against the possibility of knowledge, Antiochus seceded from the Academy altogether and adopted a position very similar to the Stoics. Reacting against this dilution of the pure sceptical spirit, Aenesidemus re-founded Pyrrhonism. Aenesidemus' works are all lost, but a ninth-century summary of his eight-volume *Pyrrhonian Discourses* by Photius survives, which allows us to recreate something of its flavour. The target of his attack was the belief that apprehension, or secure knowledge of the nature of things, was ever attainable. Nobody really knows how things are, yet everybody but the Pyrrhonists claims to do so, wasting their time in futile, dogmatic bickering. Indeed, from this time on 'dogmatic' as applied to a philosopher or philosophical view, carries the connotation 'non-Sceptical': a dogmatist is anyone who professes belief in *dogmata*; and a *dogma* is, according to Sextus Empiricus, an 'assent to one of the non-evident objects of scientific inquiry' (*Outlines of Pyrrhonism* I 13). Such matters are 'non-evident' precisely because they go beyond what is strictly contained in the appearances, the *phainomena*, themselves.

Aenesidemus apparently held that there was as a matter of fact no apprehension of things. He sought to distinguish this position from that of the sceptical Academy on the grounds that they (according to him) were prepared actually to affirm certain things absolutely, notably that apprehension was impossible; by contrast Aenesidemus' position is provisional - he has no apprehension *now*, but apprehension might yet be possible. Even so, the appropriate Pyrrhonian attitude to things is to say that they are no more so than not so, or sometimes so and sometimes not so, or so for some people and not for others. Thus Aenesidemus appears to allow a form of relativism. Relativism and Pyrrhonism are incompatible however: relativism positively asserts that there is no genuine fact of the matter, a relativist in ethics appealing to the facts of cultural divergence (as indeed will the Pyrrhonist); but while the relativist will conclude that there is no such thing as an objective good, or right and wrong, the Pyrrhonist will simply infer that we do not know whether any things are genuinely good, and, if any, which. The Pyrrhonist makes no metaphysical claims about the ontological status of such 'objects'; and hence when Aenesidemus says that some things are one way for some people and otherwise for others, he is not asserting that things *really* are that way - this is simply a report of how things seem to be.

2 The Ten Modes of Scepticism

Fundamental to the Pyrrhonists' method was the collection of cases in which things seem different under different circumstances. They appear differently to people as opposed to animals (muddy water appeals to pigs but not to us: Sextus Empiricus, *Outlines of Pyrrhonism* I 56), to different groups of people ('Indians enjoy different things from us': I 80), to different sense-modalities ('Paintings seem bumpy to vision, smooth to touch': I 92), and to the same sense-modality on different occasions ('Honey seems sweet to the healthy, bitter to the jaundiced': I 101). These are examples of the first four of the Ten Modes of Scepticism. Many of the examples retailed in the modes are Greek commonplaces, deriving from much earlier times (some may be found in Aristotle; and one of Sextus' examples of the first mode - that sea water is poisonous to humans but nourishing to fish - is drawn from Heraclitus (§3)). What is important is the systematic way in which they are collected, and the use to which they are put. Aenesidemus may well have been the first to collect the material and organize it into ten separate modes

(although this is controversial); in any event, a generation or so later, in the time of Philo of Alexandria (one of our major sources for the modes; see Philo of Alexandria §1), they were firmly associated with Pyrrhonian Scepticism.

A mode (*tropos*) in this sense is a general pattern of argument, for which any number of specific instances may be found. The basic structure of all of the modes is simple and lucid. Sextus describes Pyrrhonism as a 'capacity for opposition' (*Outlines of Pyrrhonism* I 8): arguments are opposed to arguments, appearances to appearances, and arguments to appearances (I 8-9, 31-3). Whenever the issue goes beyond how things seem to particular individuals (that is, the way things *really* are), there is, the Pyrrhonists note, an 'undecidable disagreement' about the matter. Moreover, that dispute is undecidable because of the unavailability of any generally accepted means of resolving it, any criterion in the technical Greek sense (see §5). And in default of that, the Pyrrhonist urges, there is no reason to prefer one of the appearances to another, and hence we should not commit ourselves to any of them: we should suspend judgment about them (see §4). All the modes, then, share the following basic form:

(1) x appears F relative to a
(2) x appears F^* relative to b
(3) at most one of 'x is F' and 'x is F^*' can be (objectively) true
(4) no uncontroversial decision procedure tells decisively either for 'x is F' or for 'x is F^*' so
(5) we should suspend judgment as to what x is really like.

The range of x (for example, 'sea water' in the Heraclitus case) is broad and varies from mode to mode, as do the predicates covered by the variables F and F^*; what substitutes for a and b is determined by the particular nature of each mode (thus in the first mode 'a' will be humans, 'b' other animals, fish for example; in the second a and b will be different individuals or groups of humans; and so on for the other modes). The substituends for F and F^* must be incompatible predicates (for example, 'poisonous' and 'nourishing'), and since they are incompatible, (3) follows trivially. Then, on the supposition that (4) can be made good, (5) follows on reasonable assumptions about rationality.

Sextus Empiricus, our main source for the modes, spends most of his time collecting examples of (1) and (2) for each of his different modes. The first four he labels 'modes from the judger': that is, the various types of opposition are to do with variations in the individuals observing them. Modes 'from both judger and judged' include the fifth, which notes that the appearance of things varies according to the distance and standpoint from which they are viewed (this is the source of the hardy philosophical perennial that a straight oar looks bent in water: *Outlines of Pyrrhonism* I 119), and the ninth notes that our appreciation of things differs according to how familiar we are with them (naked bodies become unexciting after a while: I 142). The seventh and tenth modes are labelled 'from the thing judged' (although in the latter case it is hard to see why); the seventh is to do with supposed variations in the constitutions of things which are observed under various conditions (thus the filings of black horn look white: I 129), and it is very hard to see what Sceptical leverage such considerations might have; the tenth is the mode 'mainly concerned with ethics, having to do with ways of life, habits, laws, mythical beliefs and dogmatic suppositions' (I 145), in which Sextus collects with evident enthusiasm examples of cultural diversity ('Indians copulate in public': I 143; 'Persians marry their mothers': I 152). These examples are multiplied in a later, more general treatment of ethics (I 3 168-238).

The tenth mode in particular emphasizes an important issue. The proper Pyrrhonian conclusion from the facts of ethical diversity is, according to Sextus at least, that we cannot know whether or not anything is good or bad by nature (I 163, III 235); yet the ethical relativist will equally appeal to precisely the same examples in order to ground the different and incompatible conclusion that *nothing* is good or bad *by nature*: there simply are no fundamental facts about value. The Pyrrhonist avoids this move because it too is a species of dogmatism, albeit negative in form. Different people disagree; and that is all there is to it. For all we know there may be actual absolute values. The relativist, negative conclusion is just as unwarranted as its positive dogmatic opponent. Just as the Pyrrhonists will not say that everything is inapprehensible (although allowing that nothing actually seems to be apprehended: see §1), neither will they absolutely deny the existence of objective values.

Similarly, when propounding the eighth mode, the 'mode of relativity', the Pyrrhonist only affirms that everything *appears* to be relative (I 135), although (as Sextus notes elsewhere: I 39), relativity is in fact at the core of all of the modes' procedures (as (1) and (2) above make clear). Thus there is no second-order commitment to the *truth* of the claim that everything is relative; even that is a matter of appearances.

3 The Sceptical slogans

This is the core of the Pyrrhonian position. Pyrrhonists liked to sum up their philosophy in pithy slogans (such as 'I determine nothing': *Outlines of Pyrrhonism* I 197): but this merely means 'I am now so affected as neither to affirm nor deny dogmatically any of the subjects under investigation'. Equally, ' Everything is inapprehensible' only means 'All the non-evident subjects of dogmatic inquiry which I have investigated seem to me inapprehensible' (I 200), and similar caveats apply to 'to each argument an equal argument is opposed' (I 202-3), and *ou mallon*: ' the phrase "no more this than that" expresses our affection, by which we end up in equipoise because of the equal strength of opposing matters' (I 120). The end result is suspension of judgment, or *epochē* (stage (5) see §2 above); even then, the Pyrrhonist is careful to say only that matters *appear* equally balanced (I 196), while Sceptical *aphasia*, or non-assertion, is 'an affection of ours because of which we neither affirm nor deny anything' (I 192); moreover, its scope covers only 'what is said dogmatically in regard to things non-evident, since we yield to those things which move us affectively and force us necessarily to assent' (I 193).

Pyrrhonists, then, can say how things seem to be to them, and report the way they are affected; they live their lives according to those appearances, but 'undogmatically', that is, without any commitment to their underlying truth. Furthermore, the 'Sceptical expressions' such as 'I determine nothing' apply to themselves; for the Pyrrhonist, unlike some varieties of relativist, there is no privileged meta-language in which unalloyed second-order truths can be expressed: it is Scepticism all the way down. Thus Pyrrhonism manages to avoid being (at least in any damaging sense) self-refuting, precisely, if paradoxically, because the Sceptical slogans apply to themselves: Sextus compares them to purgative drugs which eliminate themselves along with the 'noxious humours' which are their object (I 206).

4 The criterion, signs and proof

The Ten Modes of Scepticism are ascribed by Sextus to 'the older Sceptics' (*Outlines of Pyrrhonism* I 36). By contrast, a set of five modes, more general in scope, is attributed by him to 'the more recent Sceptics' (I 164). Although in Sextus' presentation each of the ten modes concludes with suspension of judgment, the arguments by which that is supposed to commend itself are not spelled out in detail; the final steps of the basic schema outlined in §2 still require elucidation.

The modes of Agrippa organize the Sceptical material rather differently: two modes emphasize, in general terms, disagreement and the relativity of appearances, while the other three aver that any attempt to resolve the dispute will result either in mere assertion, infinite regress or circularity (see Agrippa). These three formal modes are in evidence throughout Sextus' presentation of Pyrrhonian argument; and they ultimately give the reasons why oppositions of the sort collected by the ten modes result in *epochē*. But, as Sextus was well aware, this is tricky, since elsewhere Pyrrhonists cast doubt on the validity of argument itself. If proof is doubtful, why imagine that (1)-(4) should entail (5)? Crucially, Sextus nowhere relies on the success of that entailment. Rather, perfectly consistently, he simply describes his own affective states. Persistent disputes of the sort gathered by the ten modes, and the persistent appearance that everything is relative, cause him to suspend judgment about things; but that too is simply an appearance, albeit an appearance of his own mental states, which he is at perfect liberty (non-dogmatically) to express.

Moreover, Pyrrhonian arguments are dialectical, directed against the various schools of dogmatists. If dogmatists believe in such things as proof, then they must, by their own lights, accept the Sceptics' arguments (if they are valid). The Sceptics, by contrast, are not obliged to accept them, and do not do so (although they do not reject them either); but since they do not suffer from dogmatic belief they have no need of them. The bulk of book II of *Outlines of Pyrrhonism* is taken up with such issues - and Sextus opens the argument with a vigorous defence of the salubrity of the Sceptic's procedure of investigating dogmatic doctrines on their own terms (II 1-12): the Sceptic can understand the content of the dogmatist's claims without being committed to it, and can thus point out what should (to the dogmatist) appear to be inconsistencies; and if there are such apparent inconsistencies, dogmatists themselves should abandon their own positions. All of the argument here concerns the criterion.

A criterion is, literally, something that judges: a touchstone. Criteria are standardly, in Hellenistic philosophy and after, divided into criteria of truth (means of judging true from false propositions) and criteria of action (ways to decide how best to behave). The Academics, in their battle with the Stoics (see §1), had denied the existence of the

former, while allowing that the latter were still available. Sextus seeks to cast doubt on all of the former, by the standard Pyrrhonian means of contrasting what differing dogmatic schools (primarily the Stoics and the Epicureans) had to say about the issue, as well as by pointing to internal difficulties in their several accounts. Thus Academic arguments are retailed against the Stoic criterion of the 'apprehensive impression', but not in order to show positively that there is no such thing (II 79).

Perhaps most characteristic is the claim that, since the existence of a criterion is itself a matter of dispute, it too requires a criterion to resolve it: but none is available, except on pain of infinite regress or circularity (II 20, 34-5). Given ubiquitous dispute, we might look for majority agreement, but that is hardly to be found, and even if it were would be no guarantee of the truth (you can fool almost all of the people some of the time, as the Stoics themselves allowed): Sextus employed similar considerations in discussion of the second of the ten modes (I 87-9); and elsewhere he refuses even to allow that universal agreement on an assertion is enough to show that it is true, since present consensus is no guarantee that future disagreements will not crop up (I 33-4).

Here an earlier issue resurfaces. What is wrong with adopting the Protagorean position that everyone is judge and jury in their own case? (See Protagoras §3.) Exactly that, Sextus thinks: criteria should be independent. But why? Once again it is vital to place Sextus' argumentation in its dialectical context. Again and again, he will claim that some dispute, or opposition among arguments, or disagreement about appearances, is sufficient to induce *epochē* 'concerning how things are absolutely and in their real nature' (for example, *Outlines of Pyrrhonism* I 135). He is here relying upon what had become a commonplace of non-Sceptical metaphysics, namely that if something possesses a property essentially, it must do so under all circumstances and in all conditions; but surely the irremediable variability of appearances casts doubt upon our ability to infer to such properties; and if that commonplace is to be given a strong construal, the mere fact that nothing appears the same under all circumstances is enough to show that none of its apparent properties is, in this sense, real. Once more, that conclusion is presented as being something the dogmatist is forced, willy-nilly, to accept; the Pyrrhonist, of course, simply suspends judgment.

All of this is powerful and pointed; and it is backed up by the Sceptical attack on signs in general and proofs in particular. In the latter case, Sextus seeks to show that, on their own account, dogmatic conceptions of proof are self-stultifying (his principal target here is Stoic logic, his method to show that, on the Stoics' own account of the truth-conditions for the conditional, every proof will contain a redundant conditional premise; and yet redundancy is, on the Stoics' own account, invalidating) (II 159). More generally, Sextus argues that the dogmatists' 'indicative sign' cannot fulfil its supposed function (Sextus is however, perfectly happy to allow the existence of commemorative signs, whereby one evident phenomenon serves as a sign of some other, temporarily non-evident, phenomenon, on the basis of past experience). An indicative sign was defined as 'an antecedent in a sound conditional which serves to reveal the consequent' (II 101), where the consequent is something by nature non-evident (as sweating, in the dogmatists' stock example, is supposed to reveal the existence of invisible pores in the skin). But, the Sceptics urge, if what is signified is such as never to be evident, how can we know that it has been so revealed?

5 Cause and explanation: Sceptical physics

Dogmatists may reply that perhaps the phenomena are such that there is only one possible set of underlying circumstances which could account for them (as is alleged in the sweating-pores case). The atomists held that it was only on the assumption that the world was fundamentally composed of atoms and the void that evident facts of experience (such as the existence of motion) could be explained (see Atomism, Ancient; Democritus §2; Epicureanism §2): perhaps such quasi-transcendental inferences really can take us indubitably to the heart of things. But the Sceptics simply respond that there is no reason to think that there are any such valid inferences - after all, most physical theorists reject the atomist inference from motion to the void.

Such considerations formed the core of a set of eight modes against the aetiologists, attributed to Aenesidemus. These modes consist of a set of general Sceptical reflections on the limitations of explanation: 'The first is that according to which…aetiology in general, being concerned with non-evident things, has no agreed confirmation'; 'The second shows that frequently when there are many ways of assigning an explanation to what is being investigated, some of them account for it one way only' (*Outlines of Pyrrhonism* I 181). That is, there is no general way in which we can determine the truth (if any) of non-evident matters, since they tend to be susceptible

of different interpretations. The fourth mode further notes that dogmatists assume, unfoundedly, that the micro-mechanisms to which they 'infer' will be, in general, much like their phenomenal counterparts; furthermore, all theorists are inclined to overlook awkward recalcitrant facts, as well as contradicting the evidence and sometimes themselves, and in general begging the question in favour of their own position.

Aenesidemus' eight modes attack dogmatic scientific methodology. Elsewhere (III 13-30) Sextus retails more general arguments designed to cast doubt on the coherence of the notions of cause and explanation (some of these arguments mirror those against signs from the previous book). Causes must either precede their effects, or be concurrent with them, or succeed them. The last option is absurd; and cause and effect are relative, so one cannot precede the other; yet a cause is supposed to be productive of something, and hence must precede what it is productive of. This is just one of a battery of arguments Sextus rehearses here and elsewhere (for example, *Against the Professors* IX 210-51): he also urges that causes should invariably bring about their effects, and yet everything dogmatists refer to as a cause is defeasible; thus not even fire can really be of a nature to burn since it will not burn everything under all circumstances). Sextus then uses the apparent results of the attack on causation to undermine the rest of dogmatic physics: body is that which can act and be acted upon (*Outlines of Pyrrhonism* III 38), so if there is no agency, there is no body either (and out with it go a host of related notions such as limit, surface and solidity); and equally doomed are the nest of concepts relating to change (motion, increase, rest, and so on), while if these go, so do the ideas of addition and subtraction, and with them arithmetic and the concepts of part and whole.

6 The Sceptical end

Most of Sextus' argument is devoted to attacking dogmatic concepts, yet he never loses sight of the overall goal of the Pyrrhonian procedure of constructing a dispute - it is just that the vast majority of people (and not just dogmatic theorists) believe in the existence of bodies, motion, and so on, and consequently that arm of the disagreement needs little support. Thus, first appearances notwithstanding, all of this is consistent with the general aim of Pyrrhonism. None the less, the question remains: why bother? Why should Pyrrhonists concern themselves at such length with refuting dogmatists, not to mention the views of ordinary people? The latter case is in fact particularly pointed, since Sextus frequently says that he has no quarrel with ordinary views and that he sides with life. Why, then, try to argue that motion is non-existent? Furthermore, while Sextus will on occasion argue that not even fire is warm by nature, elsewhere he will contrast the case of fire with that of ethical values, saying that the latter would have to be like fire if they were to be natural properties (*Outlines of Pyrrhonism* III 179); surely that is simple inconsistency?

The answers to these questions are controversial, but the following account seems at least coherent. Sextus does not conceive of argument in the ordinary philosophical way as entailing (or at the very least rationally buttressing) conclusions. As seen in §5, he is perfectly happy to argue against argument. Rather arguments affect us in just the same way as phenomenal appearances do; and we may, non-dogmatically, report those effects. Thus, when confronted with an equally balanced pile of considerations on both sides of an issue, I do not *conclude* that it is *rational* to suspend judgment: rather *epochē* just happens to me. Furthermore, Sextus characterizes Sceptics as perpetual investigators: why should they need to go on investigating once they have attained *epochē*? The answer is precisely that *epochē* is not rationally inferred by argument: rather it is a state, and one that is in constant need of replenishment; moreover the Sceptic goes on investigating in response to the way things seem.

Sceptics of all stripes were greatly concerned to evade the charge that life without belief (at least dogmatic belief that things *really* are thus and so) was impossible, since belief is a necessary condition for action. Dogmatic belief, Sceptics claim, is not necessary: they act on the basis of their impressions, but with none of the commitment to the truth of those impressions characteristic of ordinary people's behaviour. This characterization of the Sceptic's attitude implies a position on a much discussed question (see Barnes 1982; Burnyeat 1983; Frede 1987): namely, just what is the scope of the *dogmata* the Sceptic rejects, an answer to which is of capital importance to the associated question of whether life can be lived without them. Are *dogmata* merely *scientific* beliefs, beliefs involving explicit commitment to theory (as *Outlines of Pyrrhonism* I 13 suggests, the line adopted by Frede)? Or are they any and every non-epistemic appearance (Burnyeat's view)? More probably what makes an appearance into a *dogma* is not its content, but one's attitude to it: any appearance can, on this view, be a *dogma*, but only if its recipient accepts it as *really* true. And such commitments are not prerequisites of all action.

The upshot of this rejection of *dogmata* is *epochē*: the payoff, Sextus claims, is tranquillity, *ataraxia*, the state of being untroubled by things in general (*Outlines of Pyrrhonism* I 25-30). Distress is caused by strong evaluations - get rid of those, and life will be simpler. Of course, as a Sceptic he cannot claim certainty for his prescription; this is just the way things seem to him. But he is moved by that appearance and his general philanthropy to offer these arguments to others as well. Indeed, in a famous passage at the very end of *Outlines of Pyrrhonism*, Sextus compares Sceptical arguments to drugs: their business is to effect a cure for morbid dogmatism. How they do so is no part of the Pyrrhonist's concern.

R.J. HANKINSON

References and further reading

Annas, J. and Barnes, J. (1985) *The Modes of Scepticism*, Cambridge: Cambridge University Press.(Translation and detailed philosophical study of our sources for the Ten Modes of Scepticism.)

Barnes, J. (1982) 'The Beliefs of a Pyrrhonist', *Proceedings of the Cambridge Philological Society* 28: 1-28. (Important contribution to the debate on the liveability of scepticism; see §6.)

Brochard, V. (1887) *Les Sceptiques grecs*, Paris: Alcan; repr. 1923.(Old but still useful general treatment.)

Burnyeat, M.F. (ed.) (1983) *The Skeptical Tradition*, Los Angeles, CA: University of California Press.(Very useful collection of articles on the history of scepticism from the Greeks to Kant; includes Burnyeat's own seminal article 'Can the Skeptic Live his Skepticism?')

Dal Pra, M. (1975) *Lo scetticismo greco*, Rome and Bari: La Terza, 2nd edn.(Very good historical survey.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(Greek text with facing translation: IX 69-116 is devoted to Pyrrhonism.)

Frede, M. (1987) *Essays in Ancient Philosophy*, Oxford: Oxford University Press.(Contains several essays pertaining to scepticism, including 'The Skeptic's Beliefs'.)

Hankinson, R.J. (1995) *The Sceptics*, London: Routledge.(Full philosophical treatment of the history and development of Greek scepticism.)

Sextus Empiricus (*c*. AD 200) *Outlines of Pyrrhonism*, trans. J. Annas and J. Barnes, *Outlines of Scepticism*, Cambridge: Cambridge University Press, 1994.(Fine translation with introduction and notes.)

Sextus Empiricus (*c*. AD 200) *Against the Professors*, trans. R.G. Bury, *Against the Logicians, Against the Physicists, Against the Ethicists and Against the Professors*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 3 vols, 1935, 1936, 1949. (Parallel Greek text and English translation with minimal notes.)

Stough, C.L. (1969) *Greek Skepticism*, Los Angeles, CA: University of California Press. (Brief survey, somewhat dated.)

Pythagoras (*c*.570-*c*.497 BC)

Pythagoras of Samos was an early Greek sage and religious innovator. He taught the kinship of all life and the immortality and transmigration of the soul. Pythagoras founded a religious community of men and women in southern Italy that was also of considerable political influence. His followers, who became known as Pythagoreans, went beyond these essentially religious beliefs of the master to develop philosophical, mathematical, astronomical, and musical theories with which they tended to credit Pythagoras himself. The tradition established by Pythagoras weaves through much of Greek philosophy, leaving its mark particularly on the thought of Empedocles, Plato, and later Platonists.

1 Life and deeds

Pythagoras, son of Mnesarchus, was born on the island of Samos. For the first half of his life Pythagoras travelled widely, not only in Greece but supposedly also in Egypt, Phoenicia and Babylonia, where he is reputed to have acquired much of his knowledge and religious wisdom. Perhaps to escape the rule of Polycrates, the tyrant of Samos, he emigrated to Croton in southern Italy. His moral stature and eloquence gained him many adherents. With his followers, both men and women, Pythagoras practised a simple, communal life whose goal was to live in harmony with the divine. To that end he prescribed a regimen of purification that included dietary restrictions, periods of silence and contemplation, and other ascetic practices. In addition to the religious and monastic aspects of the Pythagorean society, we hear of Pythagorean political associations (*hetaireiai*) that played an important role in the public affairs of Croton and other southern Italian cities (it appears they initiated social reforms and supported aristocratic constitutions). After a time their dominance came to be resented and a ‘Pythagorean revolt’ ensued, in the course of which many Pythagoreans were killed or scattered abroad. Pythagoras himself, possibly as a result of this upheaval, moved to Metapontum where he died.

Already during his lifetime Pythagoras was regarded with near religious veneration. It is therefore not surprising that the stories told about him after his death should turn into hagiology and include many fantastic elements: that Pythagoras was the Hyperborean Apollo and had a golden thigh to prove it; that he was seen in two places at one time; that he could converse with animals and control natural phenomena. Pythagoras’ wonder-working clearly belongs to the realm of legend, although it reinforces the picture of him as a ‘shaman’. A more difficult matter is to establish what Pythagoras actually taught, since the oral and then the written traditions attribute to him not only miracles but also sophisticated mathematical and philosophical achievements. This habit of tracing all things back to the master, coupled with evidence of the quasi-religious avoidance of uttering his name, is typified in the expression common among Pythagoreans: ‘he himself said’ (*autos epha*; Latin *ipse dixit*). However, because Pythagoras wrote nothing and shrouded his lectures in secrecy, it is impossible to verify all that is ascribed to him. What remains certain is that he was a highly influential religious teacher whose main tenets dealt with the soul and the rites required for its purification and salvation. This made the Pythagorean movement especially popular in Magna Graecia, a fertile soil for mystery cults of all kinds. Pythagoras is also connected with certain ‘Orphic’ writings, since these share eschatological concerns similar to his oral teachings (see Orphism). For these reasons the following account will emphasize those doctrines that accord with Pythagoras’ reputation as an early Greek sage and religious innovator (for the philosophical and scientific theories traditionally associated with his name, see Pythagoreanism).

2 Teachings

Pythagoras believed that the world was animate and that the planets were gods. This view of the universe as living and divine was characteristic of early Greek thought (see Thales §2), but what appears unique with Pythagoras is the corollary he drew on an anthropological level: there is an element in human beings that is related to the universe and that, like the universe in which events recur in eternal cycles, is eternal. This divine, immortal element is the soul (see Psychē). With the death of the body the soul passes into another body, human or animal. An early witness to Pythagoras’ belief in transmigration (or *metempsychōsis*) is the poet-philosopher Xenophanes (§1), who satirizes him for claiming to recognize the soul of a friend when he heard the voice of a puppy that was being beaten. Pythagoras asserted of himself, as is typical of a religious figure who draws on personal experience, that he had once been the Homeric hero Euphorbus, and he exhorted his disciples to recall their own past lives.

The immortality of the soul underlies many of Pythagoras' practical teachings, for the soul, as the most important element within a person, required nurture to ensure not only equanimity in this life but also a better incarnation in the life to come. These ends could be achieved by bringing the soul into harmony with the divine, cosmic order (according to a disputed doxography, Pythagoras was the first to pronounce the world a *kosmos*, a term that in Greek combines the ideas of adornment, beauty and order). In so far, however, as the soul resided in a body, it needed to be freed from the turmoils and corrupting influences of the body. Hence Pythagoras preached a strict way of life that centred on purification (see Katharsis) and asceticism. Furthermore, he practised a form of musical therapy for both body and soul. He valued friendship highly as a means of promoting equality and concord; the love of friends was a specific instance of the universal sympathy existing in the cosmos.

The rules of Pythagoras found expression in short, pithy sayings known as *akousmata*, a term that implies oral transmission, or, more frequently, as *symbola*. The latter most likely functioned as secret passwords for Pythagorean initiates but, as the name suggests and their often esoteric and oracular nature prompted, they were also subjects for 'symbolic' interpretation. The *symbola* range from primitive religious taboos to simple moral precepts and various dietary prohibitions (abstinence from certain parts of animals and the famous ban on eating beans).

3 Legacy

Pythagoras, according to Plato (*Republic* X 600b), handed down to his followers a distinctive way of life (*bios*) 'they call Pythagorean to this day'. By Plato's day the Pythagorean life meant, besides purifications of the soul, inquiries in philosophy, mathematics, astronomy and music. Did Pythagoras bequeath these enterprises as well? Empedocles (§2), who was greatly influenced by Pythagoras, praised him as a man of surpassing knowledge, with a vast wealth of understanding, capable of all kinds of wise works. Heraclitus (§1), while agreeing that Pythagoras 'practised inquiry beyond all other men', saw the result as a peculiar wisdom consisting of polymathy and evil artifice. From both witnesses, however, Pythagoras emerges as a figure active in a wide variety of fields and therefore likely at least to have dabbled in those studies for which Pythagoreans were known in Plato's time. Still, the only sure legacy of Pythagoras is the immortality of the soul. Although this was primarily a religious belief, it carried philosophical import. By singling out the immortal soul as the essential element of life Pythagoras foreshadowed the Parmenidean/Platonic distinction between eternal being and changeable becoming and, in general, the dualism of mind and matter that informs so much of Western philosophy.

See also: Archytas; Iamblichus; Neo-Pythagoreanism; Philolaus; Presocratic Philosophy

HERMANN S. SCHIBLI

References and further reading

Burkert, W. (1972) *Lore and Science in Ancient Pythagoreanism*, trans. E.L. Minar, Jr, Cambridge, MA: Harvard University Press.(A scholarly and critical examination of the sources; Pythagoras the religious figure is sharply distinguished from Pythagorean philosophy and science which become identifiable only in the late fifth century BC.)

Diogenes Laertius (first half of the third century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius, Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925.(VIII 1-50 is a life of Pythagoras.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols.(A comprehensive and readable discussion of Pythagoras and early Pythagoreanism can be found in volume 1, pages 146-340.)

Guthrie, W.K.C. (1987) *The Pythagorean Sourcebook and Library*, Grand Rapids, MI: Phanes Press. (Conveniently collects the main ancient sources.)

Iamblichus (*c.* AD 290-300) *On the Pythagorean Way of Life*, trans. J. Dillon and J. Hershbell, Atlanta, GA: Scholars Press, 1991.(A Neoplatonic account of the Pythagorean life in which Pythagoras appears as the philosopher *par excellence*; the most exhaustive ancient source that survives. The edition contains text, translation and notes.)

Kingsley, P. (1995) *Ancient Philosophy, Mystery, and Magic: Empedocles and Pythagorean Tradition*, Oxford: Oxford University Press.(Includes *passim* critical, yet sympathetic, discussions of the figure of Pythagoras and the Pythagoras legends.)

Navia, L.E. (1990) *Pythagoras: An Annotated Bibliography*, New York and London: Garland. (Covers the

literature up to 1989 on all aspects of Pythagoras and Pythagoreanism.)
Navia, L.E. (1993) *The Presocratic Philosophers: An Annotated Bibliography*, New York and London: Garland. (At pages 591-7, gives an updated bibliography of works on Pythagoras and Pythagoreanism since 1988.)
Porphyry (*c*. AD 250-269) *The Life of Pythagoras*, trans. M. Smith, in M. Hadas and M. Smith, *Heroes and Gods: Spiritual Biographies in Antiquity*, New York: Harper & Row, 1965, 105-28.(The only surviving part of Porphyry's *Philosophic History*, it combines legendary material with historical information.)

Pythagoreanism

Pythagoreanism refers to a Greek religious-philosophical movement that originated with Pythagoras in the sixth century BC. Although Pythagoreanism in its historical development embraced a wide range of interests in politics, mysticism, music, mathematics and astronomy, the common denominator remained a general adherence among Pythagoreans to the name of the founder and his religious beliefs. Pythagoras taught the immortality and transmigration of the soul (reincarnation) and recommended a way of life that through ascetic practices, dietary rules and ethical conduct promised to purify the soul and bring it into harmony with the surrounding universe. Thereby the soul would become godlike since Pythagoras believed that the cosmos, in view of its orderly and harmonious workings and structure, was divine. Pythagoreanism thus has from its beginnings a cosmological context that saw further evolution along mathematical lines in the succeeding centuries. Pythagorean philosophers, drawing on musical theories that may go back to Pythagoras, expressed the harmony of the universe in terms of numerical relations and possibly even claimed that things are numbers. Notwithstanding a certain confusion in Pythagorean number philosophy between abstract and concrete, Pythagoreanism represents a valid attempt, outstanding in early Greek philosophy, to explain the world by formal, structural principles. Overall, the combination of religious, philosophical and mathematical speculations that characterizes Pythagoreanism exercised a significant influence on Greek thinkers, notably on Plato and his immediate successors as well as those Platonic philosophers known as Neo-Pythagoreans and Neoplatonists.

1 History

In the second half of the sixth century BC, Pythagoras founded a community in the southern Italian city of Croton whose members were united by the belief in the transmigration of the soul, an ascetic way of life that centred on the purification of the soul, and a political outlook that aimed at social reform along aristocratic lines. Pythagorean associations (*hetaireiai*), which also formed in other cities of Magna Graecia, acquired considerable political authority, but their dominance eventually met with opposition, both during the lifetime of Pythagoras and later again about 450 BC. In the wake of these anti-Pythagorean movements the followers of Pythagoras were scattered throughout the Greek world, so that by the time of Plato there is little evidence of formal Pythagorean societies. Individual Pythagoreans, however, continued to be recognized, some by their distinctive lifestyle in matters of food, dress and purificatory practices, others by the additional pursuit of various philosophical, mathematical and musical theories with which they credited Pythagoras. Later tradition refers to these two types of Pythagoreans as ‘hearers’ (*akousmatikoi*) and ‘learners’ (*mathēmatikoi*). The distinction supposedly goes back to the original society in which some members were only fit to accept the oral teachings of Pythagoras without arguments and proofs, while those with more leisure and perhaps philosophical ability were further instructed in the rational foundations of the master’s teachings. Whatever differences there may have been between groups of Pythagoreans, and even among individual *mathēmatikoi*, all professed allegiance to Pythagoras.

From the time of Pythagoras to Plato there were several famous Pythagoreans - Hippasus, Philolaus and Archytas. Aristotle, in his extant works, speaks only generally of the Pythagoreans (‘some Pythagoreans say… ’); his special treatise on Pythagorean beliefs unfortunately no longer survives. In the third and second centuries BC we do not hear of philosophers who were known as or called themselves Pythagoreans, but interest in ‘Pythagoreanism’ continued, as is evidenced by the wealth of apocryphal writings in prose and verse on Pythagorean themes that mostly date from this period. The actual practice of Pythagoreanism experienced a revival in the Roman world from the first century BC to the first century AD; Latin writers such as Cicero, Ovid and Seneca testify to its popularity. In the first two centuries AD the theoretical side of Pythagoreanism marked certain philosophers to the extent that they may be called Neo-Pythagoreans (see Neo-Pythagoreanism) and these in turn influenced later Platonic philosophers (see Neoplatonism).

2 Music, mathematics, and cosmology

Plato says of the true philosopher, whose mind is on the higher realities (that is, the Platonic Forms):

> he looks unto the fixed and eternally immutable realm where… all is orderly and according to reason, and he imitates this realm and, as much as possible, assimilates himself to it… and by association with the divine order becomes himself orderly and divine as far as a human being can…

(*Republic* VII 500c)

Plato’s philosophic ideal, as well as one of the methods he prescribed to achieve it - namely mathematics - owes much to Pythagoreanism. Pythagoras had taught that life should be in harmony with the divine cosmos (see Pythagoras §2). In Pythagoreanism harmony (*harmonia*) became a central tenet and was explained through numerical relations, possibly in connection with musical theory. For example, the Pythagoreans thought that the motions of the orbiting planets produced a sound which, given the belief that the intervals between the heavenly bodies corresponded to musical ratios, was harmonious. So Aristotle explains the ‘music of the spheres’ - one of several explanations offered for this famous Pythagorean image. Of seminal importance for illustrating the coherence of music and number was the discovery of musical ratios - that music should result when the first four integers of the numerical system, used as components in the harmonic ratios of the octave ($2 : 1$), the fifth ($3 : 2$) and the fourth ($4 : 3$), were imposed upon the continuum of sound. Whether or not Pythagoras, as tradition holds, was the ‘discoverer’ of the musical concords, the Pythagoreans fixed upon the first four numbers as the building blocks of nature. These four sufficed to give extension and shape to bodies in the sequence of point-line-surface-solid:

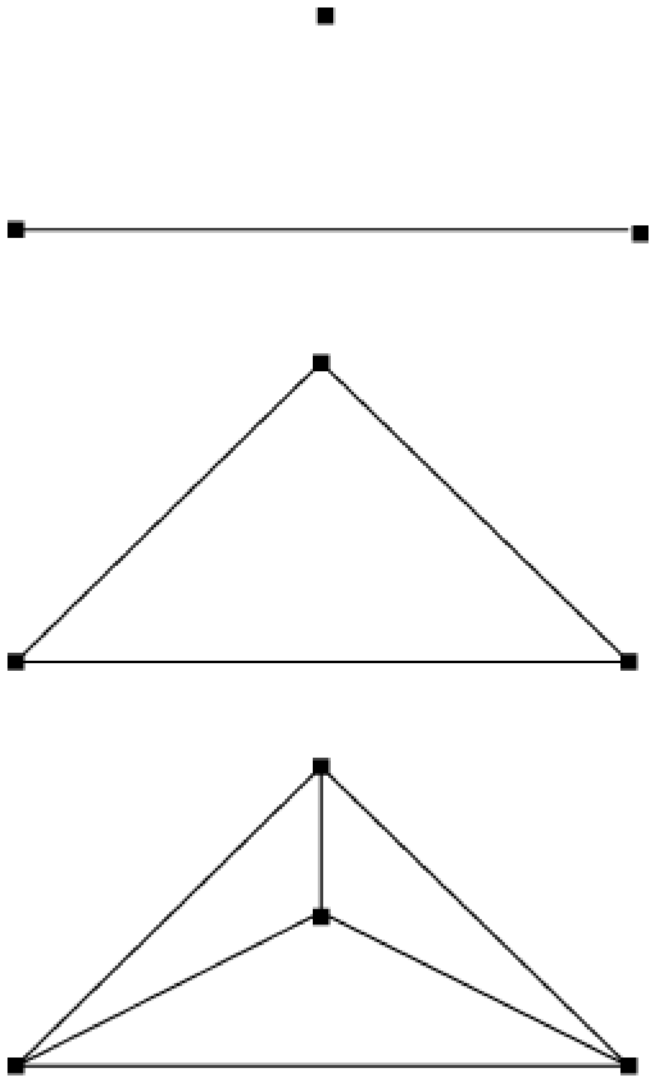

Moreover, the first four integers add up to ten, which the Pythagoreans considered a perfect number and represented in a figure called the *tetraktys*.

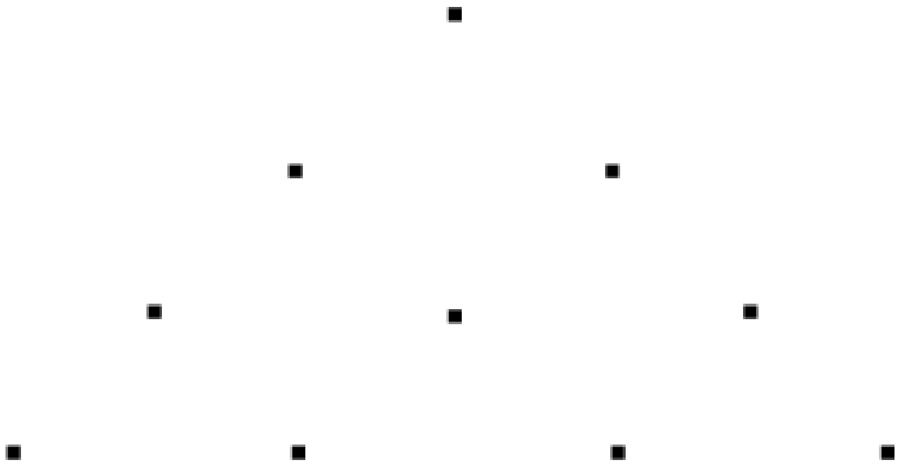

From whatever angle one approached the *tetraktys* in counting, the addition always resulted in ten, the basis of the decimal system. The *tetraktys* also produced an equilateral triangle, the simplest plane figure, and a pyramid, the simplest three-dimensional shape.

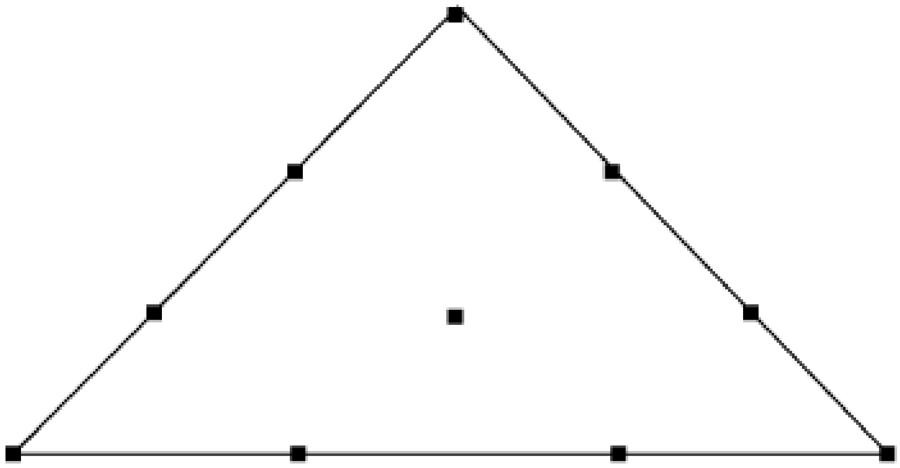

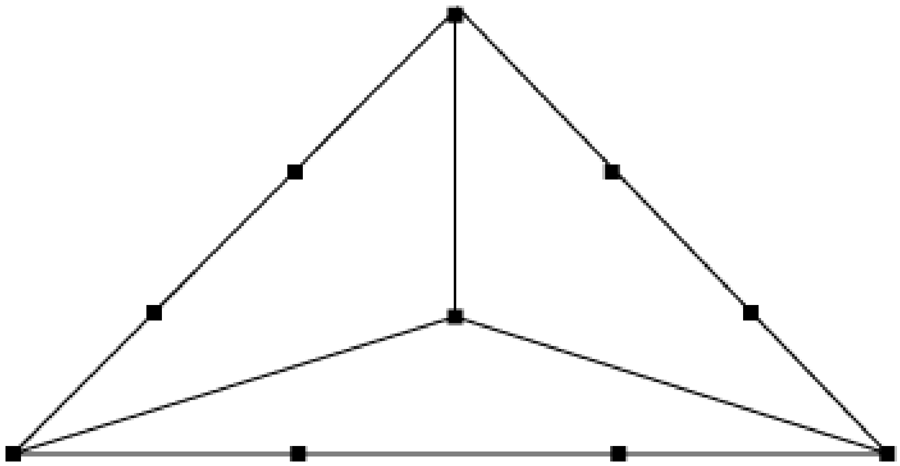

The Pythagoreans viewed the *tetraktys* as a sacred symbol and used it in the following oath: ‘By him [Pythagoras] who handed down to us the *tetraktys*, source and root of everlasting nature’. In short, the Pythagoreans supposed the nature of things could be understood numerically; indeed, some of them apparently went so far as to say that things *are* numbers: ‘…they assumed the elements of numbers to be the elements of all things that exist, and the whole universe to be *harmonia* and number’ (Aristotle, *Metaphysics* 986a2).

Since numbers functioned as the constituent elements of the cosmos, it is important to see how the Pythagoreans understood the nature and generation of number itself. Aristotle again offers the best starting-point:

> The elements of number are the even and the odd, the latter limited, the former unlimited. The One is composed of both of these (for it is both even and odd) and number comes from the one; and numbers… are the whole universe.
>
> (*Metaphysics* 986a17)

How the One can be both odd and even is best explained in the sense that the One, as the first number, is the principle of both odd and even numbers (zero was unknown in Greek mathematics). Thus ‘number comes from the One’, and the generation of number was simultaneously a cosmogonical process since ‘numbers… are the whole universe’. Although in its origin the Pythagoreans viewed the One as both odd and even, limited and unlimited, in its practical application, that is, in its interaction with other numbers, they treated it as odd and limiting. This can be seen in the schema by which the Pythagoreans illustrated the correspondence of odd/even to limited/unlimited. Gnomons (carpenter squares) were placed around an arrangement of points (or pebbles) as follows:

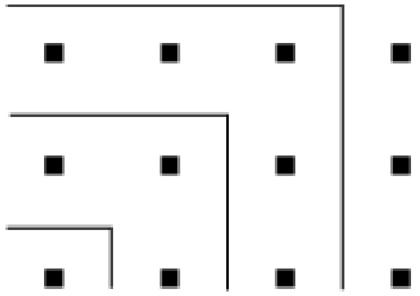

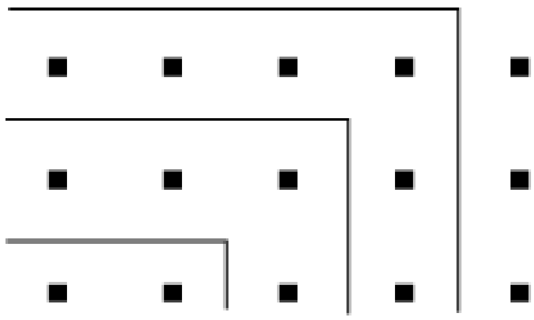

When a gnomon is placed around one point, and the process is continued in sequence, the resulting figure is always of ‘limited’ shape, that is, always a square, whereas when it is placed around two points, the result is a series of oblongs whose sides stand in an ‘unlimited’, that is, an infinite variation to each other. In this scheme the One, as a single unit or point, is equated with the odd. The One is also ranked with limited and odd in the Pythagorean table of ten opposites:

limited	unlimited
odd	even
one	many
right	left
male	female
rest	motion
straight	crooked
light	darkness
good	bad
square	oblong

The position of limited/unlimited at the beginning is not accidental, for their opposition, embodied in the original composition of the One, was considered primal and basic to the development of number and the universe, while the opposition between odd and even can be generally subsumed under this fundamental distinction. The connection of the One with the limited reappears in another Pythagorean text (Aristotle, *Metaphysics* 1091a17), where it is said that the limited (tantamount to the One) ‘breathes in’ and is thus penetrated by the unlimited. In this cosmological fragment the unlimited is equated with the void, to be thought of as infinite space and serving as a dividing principle. The One now becomes a Two (reminiscent of the mythological separation of heaven and earth), which in effect marks the beginning of number, plurality and the existence of discrete physical bodies. The cosmogonic process is sometimes put the other way round: the unlimited is limited (or penetrated) by the limit. This way of stating it brings to the fore the characteristic Greek feeling that what is unlimited, without bounds, is without order and somehow evil (hence ‘bad’ is listed under the unlimited in the above table) and therefore needs to be curtailed and bounded by limit and measure (as the limits of harmonic ratios must be imposed upon the limitless range of sound to produce music). The result, in cosmological terms, is then precisely a *kosmos*, an orderly and structured universe. In so far as the perfected and integral cosmos represents a *harmonia*, a unity stemming from a reconciliation of opposites, Pythagoreanism has a monistic aspect, yet in its source theory it remains essentially dualistic: number comes from the One, and the One is composed of limited and unlimited. From these first two principles everything derives, even though the One, when identified with limit, odd, male, and so on, was felt to be the ‘good’ element in this cosmic dualism. (There can be no question of positing the One as the sole ultimate principle without imposing Platonic and Neo-Pythagorean notions upon early Pythagoreanism.)

Limited und unlimited are the principles of number. Numbers do not merely express the substance, shape and quantitative differences of things but actually appear to constitute bodies of physical dimension. The identification of things with numbers confuses physical bodies with abstractions, or, in Aristotelian terms, the material causes of things with their formal causes. Here Pythagoreanism connects with much of Presocratic thought according to which the source and informing element of the universe was thought of as some kind of matter, be it water, air, fire, earth, or a combination of these (see Archē). At the same time the numerical theories of the Pythagoreans represent a true advance in the history of Greek philosophy: the attempt to explain the nature of things by numbers is a valid philosophical striving to understand the world by its formal or structural principles, even if the Pythagoreans then equated these with the things themselves. And while the Pythagoreans’ interest in numbers was often infused with a mystical element (their solemn veneration of the *tetraktys*) and a primitive number symbolism that blurred distinctions between abstract and concrete (a moral concept such as justice was considered the embodiment of four, a square number of ‘just’, that is, equal reciprocity), no less an authority than Aristotle acknowledged that the Pythagoreans were pioneers and made advances in the mathematical fields (arithmetic, astronomy, harmonics and some geometry). For the ‘Pythagorean theorem’, although long known to Babylonian mathematicians, the Pythagoreans are generally considered to have found the proofs and, although in the wake of this theorem they discovered that the ratios of geometrical figures to each other cannot simply be expressed by a series of rational integers - the discovery of irrational numbers and the principle of incommensurability upset the Pythagorean notion that the world is harmony and number - their very setbacks contributed to subsequent work in Greek mathematics. Plato valued mathematics as a useful discipline to train the philosopher in the perception of eternal and transcendent truth, since in his mind it dealt essentially with invisible and eternally valid realities. By divorcing number from physical substance Plato transforms Pythagorean mathematics, yet few doubt that his

interest in number, arithmetic and measure largely rests upon Pythagorean foundations (consider the mathematical model of the universe in his *Timaeus* (see Plato §16). Plato's immediate successors in the Academy were steeped in Pythagorean number theory (see Speusippus §2; Xenocrates §2).

3 Soul and ethics

Perhaps in no other ancient Greek philosophy is the human being as intimately linked to the cosmos as in Pythagoreanism. Although the correspondence between microcosm and macrocosm, the individual and the universe, is found in much of Greek thought, it received a particularly sharp outline in the Pythagorean view that by assimilation to the divine cosmos the self would come to reflect the cosmic order and harmony. The true self of every person was the soul (see Psychē), the essential element in the partnership of body and soul. Pythagoras' teaching that the soul survived the dissolution of the body and reappeared in other bodies was steadfastly adhered to throughout the history of Pythagoreanism, even if it was sometimes eclipsed by other interests. The doctrine of the immortality and transmigration of the soul translated into a practical and clearly defined way of life that combined ritual purity with high ethical standards and whose precepts were embodied in sayings known as *akousmata* and *symbola* (see Pythagoras §2). The belief in the kinship of all life, as a corollary to the belief in transmigration, imposed a great moral responsibility towards parents, children, friends and fellow citizens, and generally entailed a respectful attitude towards all forms of life in which soul may be embodied; hence the Pythagoreans were, to varying degrees, vegetarians. Their dietary laws were also intended to free them from bodily pollution. The body was seen as a prison, even as a grave (Plato, *Gorgias* 493a) from which the soul was to rise, ever achieving superior reincarnations and culminating in a state of divinity. Thus the Pythagoreans postulated three kinds of rational creatures: 'gods, men, and such as Pythagoras'. Pythagoras was thought to have achieved semi-divine status, which in Greek could be expressed by 'daemon'. It is in a Pythagorean vein that Empedocles (§2) proclaims human beings to be fallen 'daemons'. (A complex 'daemonology' was to become an ingredient of Neo-Pythagoreanism.) The recognition of the religious and moral mandates of the Pythagorean life and of its eschatological meaning constituted wisdom (*sophia*) and the lover of such a life was a *philosophos*, a term that in this sense, as certain traditions report, was first coined in Pythagorean circles.

See also: Mystical philosophy in Islam §2; Orphism; Presocratic philosophy

HERMANN S. SCHIBLI

References and further reading

Aristotle (*c*. mid 4th century BC) *Metaphysics*, trans. in J. Barnes (ed.) *The Complete Works of Aristotle*, Princeton, NJ: Princeton University Press, 1984.(A primary source on early Pythagoreanism.)

Burkert, W. (1972) *Lore and Science in Ancient Pythagoreanism*, trans. E.L. Minar, Jr, Cambridge, MA: Harvard University Press.(A scholarly examination of the sources of Pythagoreanism and a critical analysis of Pythagorean teachings; a prerequisite for all serious work on the subject.)

Diels, H and Kranz, W. (eds) (1952) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidmann, 6th edn, vol. 1, 96-113 and 375-480.(The standard collection of the ancient sources containing the main sources on Pythagoreanism.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols.(A comprehensive and readable discussion of early Pythagoreanism can be found in volume 1, pages 146-340; emphasizes the harmony between its religious and philosophical-scientific elements.)

Huffman, C.A. (1993) *Philolaus of Croton: Pythagorean and Presocratic*, Cambridge: Cambridge University Press.(Examines the fragments of a notable Pythagorean; pages 57-64 argue that 'all things are number' was not an actual claim of the Pythagoreans but rather Aristotle's summary of their number theory.)

Kahn, C.H. (1974) 'Pythagorean Philosophy before Plato', in A.P.D. Mourelatos (ed.) *The Pre-Socratics*, Garden City, NY: Doubleday; repr. 1993.(A short but useful introduction.)

Kingsley, P. (1995) *Ancient Philosophy, Mystery and Magic: Empedocles and Pythagorean Tradition*, Oxford: Oxford University Press.(Examines Pythagoreanism against the background of Greek mystery and magic; also discusses the influence of Pythagoreans on the myths in Plato.)

Navia, L.E. (1990) *Pythagoras: An Annotated Bibliography*, New York and London: Garland. (Includes the literature, up to 1989, not only on Pythagoras but also on all aspects of Pythagoreanism.)

Philip, J.A. (1966) *Pythagoras and Early Pythagoreanism*, Toronto, Ont.: Toronto University Press.(An examination, based principally on Aristotle, of the major doctrines of Pythagoreanism; in regard to cosmology, holds that the Pythagorean opposites are not the first principles of a dualistic cosmos, but rather, given a

creation in time, the constituents of one *physis*, one world-nature.)
Plato (*c*.380-367 BC) *Republic*, trans. P. Shorey, *Plato: Republic*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1930, 2 vols.(Much Pythagorean influence in books VI-VII.)

Qi

A difficult term to contextualize within Western conceptual frameworks, qi *is variously rendered as ‘hylozoistic vapours’, ‘psychophysical stuff’, ‘the activating fluids in the atmosphere and body’, and, perhaps most appropriately, ‘vital energizing field’. In the earlier texts, before the notion came to be adapted to the speculative constructions of the Han cosmologists, it had a significance not unlike the Greek* pneuma *(‘breath’ or ‘animating fluid’). In the ‘cosmological’ speculations of the Han dynasty (206* BC*-*AD *220),* qi *came to be understood as the vital stuff constitutive of all things and was characterized in terms of the active and passive dynamics of* yang *and* yin.

Most individual Western interpretations of the vital and spiritual character of things operated with a physical-spiritual dichotomy; the animating principle was largely distinguishable from the animated things (see Pneuma). In China, the animating fluid is conceptualized in terms of what we would today call an ‘energy field’. This field not only pervades all things, but in some sense is the means or process of the constitution of all things. That is to say, there are no separable ‘things’ to be animated; there is only the field and its focal manifestations. During the period of the Han cosmologists, the notions of *yin* and *yang* came to be used to characterize the dynamics of the transformations which constituted the field of *qi*.

The energizing field as the ‘reality’ of things precludes the existence of forms or ideas or categories or principles which allow for the existence of ‘natural kinds’. Thus, discriminations in the field of *qi* were made in terms of observed and conventionalized classifications associated with diurnal and seasonal changes, directions, colours, body parts and so forth. Processes associated with these correlative classifications were then charted in accordance with the rhythms of the *yang* and *yin*, or active and passive forces (see Yin-Yang).

Perhaps the most important areas in Chinese culture influenced by the notion of *qi* have to do with health, medicine and exercises leading to self-realization. Techniques such as acupuncture and the use of herbal medicines are based upon the understanding of *qi* as that fluid medium which manifests itself in the operations of the human body as well as in the bodily environs. The term ‘body’ must of course be used advisedly; everything is a continuous field of *qi* manifesting itself as both ‘body’ and ‘environs’. The purpose of acupuncture and medicine is to balance the flow of the *qi* such that environs and body are productively continuous one with the other. This productive continuity is achieved by considering the nature of the *qi* manifestation of the human body in terms of the *yin* and *yang* dynamics that are understood to be required for its health. The health of the body is, of course, dependent upon the ‘health’ of the environs. Adjustments of the *yin* and *yang* characteristics through the techniques of acupuncture or the administration of herbals serve to increase or retard the transformations of the *qi* in body and environs.

The *Huangdi neijing (Inner Classic of the Yellow Emperor)* is a pre-Han medical text which employs the sorts of correlative classifications mentioned above to promote a healthy regimen. Among many other advices, it suggests the behaviours associated with the proper response to the particular *qi* of each season. Further, each season is correlated to a part of the body. Inappropriate response to the *qi* of a particular season will harm the bodily organ associated with that season. For example, the *qi* of Spring is ‘coming to life’. In this season one must ‘sleep at night and rise early… relax the body [and] allow intent to come to life.’ Failure to respond thus to the *qi* of Spring will damage the liver and cause one to suffer chills in summer.

Techniques of self-realization have a long history in China, dating from the earliest period of Chinese written records. A fourth century BC meditative piece, ‘Inward Training’, which appears as a chapter in an eclectic work called the *Guanzi* (see Guanzi), contains techniques associated with correct posture, diet and breath control aimed at bringing individuals into harmony with the Way and thereby allowing them to achieve health and long life. Later, Mencius interpreted the field of *qi* in terms of moral energy and offered advice on the attainment of true virtue. Mencius speaks of his ability to nourish his ‘flood-like *qi*’. He describes this *qi* as that which is ‘most vast’ and ‘most firm’ (*Mencius* 2A2). His meaning seems to be, in Western philosophical parlance, that ‘flood-like *qi*’ has the greatest ‘extensive’ and ‘intensive’ magnitude. The language of (extensive) ‘field’ and (intensive) ‘focus’ allows us to understand Mencius as saying that one may nourish one’s *qi* most successfully by making of oneself the most intense focus of the most extensive field of *qi*. In this manner one gains greatest virtue (excellence,

potency) in relation to the most far-reaching elements of one's environs: 'The myriad things are here in me. There is no joy greater than to discover integrity in oneself and to seek virtue through action which persistently mirrors others (*Mencius* 7A4).

The meaning of this familiar passage from the *Mencius* is that 'all things are in me and I am with and in all things'. Recalling the meaning of *qi* as a continuous field, a better rendition might be 'The field of *qi* is focused by me, and thus all *qi* is here in me.'

See also: Chinese philosophy; Guanzi; Medicine, philosophy of; Mencius; Pneuma; Self-cultivation in Chinese philosophy; Wang Fuzhi; Yin-Yang; Zhang Zai

DAVID L. HALL
ROGER T. AMES

References and further reading

Guanzi (*c*. 4th century BC-1st century AD?), trans. A. Rickett, Princeton, NJ: Princeton University Press, 1985.(An eclectic text with important chapters on psychology.)

Huangdi neijing (The Yellow Emperor's Classic of Internal Medicine) (*c*.220-200 BC?), trans. I. Veith, Baltimore, MD: Williams & Wilkins, 1949.(A translation, dated, but all that is available.)

Major, J.S. (1993) *Heaven and Earth in Early Han Thought: Chapters Three, Four, and Five of the Huainanzi*, Albany, NY: State University of New York Press.(An authoritative translation and interpretation of the technical chapters of *Huainanzi*, a Han dynasty Daoist text.)

Mengzi (*c*.4th century BC?), trans. D.C. Lau, *Mencius*, Harmondsworth: Penguin Classics, 1970.(An authoritative translation.)

Qualia

The terms 'quale' and 'qualia' (plural) are most commonly understood to mean the qualitative, phenomenal or 'felt' properties of our mental states, such as the throbbing pain of my current headache, or the peculiar blue of the afterimage I am experiencing now. Though it seems undeniable that at least some of our mental states have qualia, their existence raises a number of philosophical problems.

The first problem regards their nature or constitution. Many theorists have noted great differences between our intuitive conceptions of qualia and those of typical physical properties such as mass or length, and have asked whether qualia could nonetheless be identical with physical properties. Another problem regards our knowledge of qualia, in particular, whether our beliefs about them can be taken to be infallible, or at least to have some kind of special authority.

1 Anti-physicalist arguments

The terms 'quale' and 'qualia' (plural) are most commonly used to characterize what may be called the qualitative, phenomenal or 'felt' properties of our mental states, such as the throbbing pain of my current headache, or the peculiar blue of the afterimage I am experiencing now. Qualitative properties can be more or less specific: the state I am in at the moment can be an example of a migraine, a headache, a pain, or, even more generally, a bodily sensation. Briefly, a mental state is considered to have qualitative properties just in case there is something it is like to be in it.

It seems undeniable that our sensations and perceptions have qualitative properties. However, many philosophers have claimed that the distinctive feels of pains, after-images and other sensations are so radically different from objective properties such as mass or length that they could not be identical with any physical or functional properties of the brain and nervous system. (Indeed, some theorists reserve the term qualia to denote those mental properties, if any, that are irreducibly qualitative: this is why disputes about the metaphysical nature of qualitative properties are sometimes characterized as disputes about whether there are qualia, rather than about what kind of properties qualia, in the metaphysically neutral sense introduced above, could be.) The view that qualitative properties are metaphysically unique has been supported by powerful argument, and may seem to be reinforced by common sense and introspection. If this view is true, however, the existence of qualia will be incompatible with a fully naturalistic account of the mind.

Anti-physicalists have argued against the identity of qualitative with physical (or functional) properties throughout the history of the philosophy of mind, but three arguments have been especially influential since the mid-1970s. The first, modelled upon Descartes' well-known argument for the 'real distinction' between mind and body in the *Sixth Meditation*, (see Descartes, R. §8; Dualism) proposes that one could, under ideal epistemic circumstances, imagine or conceive of the qualitative features of one's pains or perceptions in the absence of any specific physical or functional properties (and vice versa), and that properties that can be so conceived or imagined must be distinct. The second argument proposes that there are qualitative properties of mental states that one could not know about before actually experiencing states of that (or similar) kind, no matter how much knowledge one had about the physical and functional properties of brains and nervous systems: for example, despite having comprehensive knowledge of the relevant sciences, a person raised in an entirely black and white environment would not know what it is like to see red, and despite years of studying bat neuropsychology, we could not know what it is like to be a bat. Moreover, the argument continues, properties that do not afford this sort of knowledge could not be part of the subject matter of any of the objective sciences. The third argument states that no physical or functional characterization of sensations or perceptions could adequately explain why they feel the way they do, and that such an explanatory gap raises doubts about the identification of the properties in question. The conclusion of the first two arguments, of course, is that qualia could not be identical with physical or functional properties, and the conclusion of the third argument is that, at the very least, we have no reason to believe that these identities hold.

Although there are interesting differences between them, these arguments are linked (and can profitably be treated together) in that the first premise of each appears to depend upon the thesis (1) that there is no conceptual connection between qualitative terms or concepts and physicalistic terms or concepts. If there were such a connection, then it would be impossible, if one was fully educated and attentive, to conceive (for example) of a

pain existing apart from the relevant physical or functional property, and it would be possible, at least in principle, to know all there is to know about pain without ever having experienced pain oneself; it would also be easy to explain why it feels a certain way to have the associated physical or functional property. The second premise also appears to depend upon a common thesis, namely, (2) that given this lack of connection, our use of qualitative terms or concepts commits us to (or, at the very least, suggests) the existence of irreducibly qualitative properties.

2 Physicalist responses

Physicalists, in turn, have attempted to deny both (1) and (2). Those denying (1) have argued that, on the contrary, there is a conceptual or epistemic connection between qualitative concepts and physicalist concepts. The candidates here are usually causal, functional or otherwise topic neutral concepts that have at least some claim to be part of our common-sense understanding of these states, since it seems quite implausible to think that physical, neurophysiological or other scientific characterizations could serve as explications of our qualitative concepts (see Mind, identity theory of §3). For example, the concept throbbing headache might be characterized in terms such as 'the state typically caused by a certain range of physical syndromes, and which produces certain desires and (perhaps) behaviour directed towards the abatement of that state', and the concept appearing blue might be characterized in terms such as 'the state typically caused by certain visual stimulations and which typically results in certain similarity judgments and sorting behaviour' (see Functionalism).

Other theorists, however, have doubted that these common-sense causal or functional characterizations could be necessary or sufficient to capture qualitative concepts. The problem with sufficiency is that the attribution of such a characterization to an individual does not seem to close the question of whether they have the quale in question; this is often referred to as the 'absent qualia' objection. The problem with necessity is that some sensations or sensory features of perception seem to have no distinctive causal or functional roles that are sufficiently accessible to be part of our common sense concepts of those states: consider, for example, the strange, though mild, twinge in one's scalp that one felt for the first time yesterday. This leaves two alternatives for physicalists who accept premise (2) and thus wish to argue that there is some sort of conceptual connection between physicalistic and qualitative concepts: they can weaken the constraints on conceptual analysis, so that information from the sciences that may help to individuate sensations and perceptions can make a contribution to the explication of qualitative concepts, or they can deny that there is determinate, coherent content to our qualitative concepts over and above that which can be explicated by classical functionalist or topic neutral characterizations, and thus maintain that there are no corresponding qualitative properties at all (see Eliminativism).

The problem for the first alternative is that it is difficult to make a principled distinction between information that is relevant to conceptual explication and that which is not; defenders of this alternative respond that an important goal of explication in this case is to preserve the intuitive distinctions we make among our mental states, and that any additions to our common lore that enable us to do this (while preserving the bulk of our other commonly held beliefs about these states) will count as relevant to this project. The problem for the second alternative is that it seems counterintuitive to ignore (what seem to be) intuitively evident distinctions among qualitative states just because they cannot be captured in functional or topic-neutral terms that are part of our common knowledge. Defenders of this alternative counsel that we must be wary of judging the worth of a theory in the light of our current intuitions, since intuitions are at least partially shaped by theory and can change over time; they also point to the counterintuitive consequences of accepting a non-physicalist account of the qualitative features of our mental states, and argue that it is not clear which side has the greater burden.

On the other hand, there are many physicalists who agree that at least some qualitative concepts are ineffable, in that they cannot be captured by physical or functional descriptions, or anything other than a demonstrative such as 'feels like that' or 'seems this way to me now'. Nonetheless, they argue, one cannot draw the anti-physicalist conclusion, since the irreducibility of qualitative to physicalistic concepts does not entail the irreducibility of qualitative to physicalistic properties; that is, while accepting premise (1) of the argument against the identity of qualitative and physical or functional properties, they reject premise (2).

Physicalists using this strategy deny the Cartesian claim that our modes of conception can, even under special circumstances, be definitive guides to the nature of things. Properties exist, and are distinct or identical, by virtue of mind-independent facts about the world; the ways we conceive of properties, even clearly and distinctly, have

little to do with these facts. So stated, the argument denies that our conceptualizing has significant metaphysical import. It can be responded, however, that premise (2) is a consequence of the very possibility of reference: that non-equivalent concepts can denote the same item only by picking out different, irreducible, properties of it, and thus that even if 'pain' and 'C-fibre stimulation' did indeed denote the same property, it could only be by virtue of introducing higher-order properties that were themselves distinct. This view of how reference is determined has been considered as received wisdom since its introduction by Frege. But the argument can be countered by challenging the picture of reference it presupposes, and suggesting that the reference of qualitative concepts is determined directly, much as current causal or historical theories of reference propose for the cases of demonstratives, natural kind terms, and proper names (see Reference §4; Proper names). Anti-physicalists, in response, can mount a general challenge to these current theories of direct reference. They can also argue that qualitative concepts are a special case, one in which there is a tighter link between the concepts and the properties they pick out, but such a claim must be argued carefully, so as not to seem *ad hoc*. Alternatively, anti-physicalists can argue that, without some link between concepts (or at least clear and distinct conceptions) and the properties they denote, it would be impossible to have knowledge of metaphysical necessity and possibility at all, and thus, to preserve such knowledge, we must concede that there is at least some presumption that properties clearly and distinctly conceived to exist apart really do so.

Though all physicalists are committed to giving a physicalistic account of sensations and perceptions, those who deny that there are conceptual connections between physicalistic and qualitative concepts can concede that there is something in the realm of the qualitative that may be ineffable or irreducibly subjective, namely, the content of those qualitative concepts. In their view, however, the conceptual irreducibility of qualitative concepts implies neither the ontological irreducibility of the properties they denote nor of the concepts, as mental states, themselves; thus, this conceptual ineffability does not conflict with the physicalist requirement that all mental states and properties be fully describable in the vocabulary of the objective sciences.

Sometimes, in the post-1970s literature on this topic, anti-physicalists claim that it is only by having certain experiences that one can learn new facts about the mental. If this claim is taken to mean that experience alone provides access to a special realm of qualitative properties, then no physicalist can accept it; however, if it is taken to mean that experience affords qualitative concepts with novel content, then physicalists who accept the irreducibility of qualitative to physicalistic concepts can agree. On the other hand, physicalists who believe that qualitative concepts can be given a functional explication maintain that any knowledge we gain uniquely from experience is merely a kind of practical knowledge, or knowing how - a matter of gaining new imaginative or recognitional abilities, rather than concepts that one previously lacked.

3 Our knowledge of qualia and other issues

It is often argued that we have special epistemic access to the qualitative properties of our own mental states: it is claimed that our knowledge of them is infallible, or at least that our beliefs about them have a kind of special authority, and also that these properties are transparent or self-intimating to the minds that experience them. To have infallible knowledge of a class of properties is for it to be logically or conceptually impossible for our (sincere) beliefs about them to be false; to have special authority about a class of properties is for it to be logically or conceptually necessary for most of our (sincere) beliefs about them to be true. A property is transparent or self-intimating if an individual (with adequate conceptual sophistication) cannot help but form accurate beliefs about it whenever it is present. Insofar as they bear these special epistemic relations to those (and only those) who experience them, qualitative properties are importantly different from physical (and classically objective) properties such as shape, temperature and length, about which beliefs may be both mildly and radically fallible. The question naturally arises whether qualitative properties, if they have the above-described epistemic features, could be physical, functional or otherwise objective nonetheless.

Functionalists can accommodate infallibility or special authority and transparency, since according to (at least some versions of) this doctrine, states with qualitative properties and the beliefs they produce are interdefined. Thus, as a matter of conceptual necessity, no state will count (for example) as an instance of pain unless it produces (either always or for the most part) the belief that one is in pain, and, conversely, no state will count as a belief that one is in pain unless it is actually produced by pain.

Nonfunctionalist physicalists, on the other hand, cannot point to a definitional connection between qualitative

properties and an individual's beliefs about them. Thus, some theorists merely reject infallibility and transparency, arguing that we can and do make sincere and unambiguous errors about our experiences (and fail to notice some of their properties) when we are tired and inattentive, or when misinformation skews our expectations about the experiences we are soon to have. It is more difficult, however, to argue that one can make such mistakes in the absence of these mitigating factors. Thus, many physicalists have attempted to show that at least some sort of special epistemic access to qualia is compatible with their metaphysical views.

For example, some philosophers have argued that a certain sort of introspective accuracy is insured by the proper operation of our cognitive faculties, and thus that, as a matter of law, we cannot be mistaken about (or fail to notice certain properties of) our mental states. On a view such as this, the claim that we are infallible about our mental states (or that they are transparent to us) will be nomologically necessary, but not necessary in any stronger sense. However, on nonfunctionalist physicalist views that treat qualitative concepts as demonstrative, certain statements of infallibility can be logically or conceptually necessary, in that one cannot be wrong, on any particular occasion, to think that one's current mental state feels like that. Even so, one cannot be sure, on these views, that all states that appear to be of the same phenomenal type are in fact of that same phenomenal - that is, physical - type. Since this sort of infallibility does not guarantee that one has re-identified a single phenomenal type on different occasions, it cannot provide any substantive assurances about the nature or reality of our mental states.

There are other interesting issues about qualia which are worth mentioning, though their full discussion is beyond the scope of this entry. One involves the range or extent of their occurrence. Bodily sensations and perceptions, the most common examples of states with qualia, are often contrasted with beliefs and desires, in that the former are taken to be essentially qualitative, the latter essentially intentional (that is, purporting to be about items or events in the world) (see Bodily sensations; Perception; Intentionality; Propositional attitudes). Some theorists dispute this distinction, arguing that true intentional states must have qualitative properties as well. But whether or not intentional states must also be qualitative, it certainly seems (for occurrent beliefs and desires) that there is something it is like for us, if not all creatures, to have them. If so, then the issues raised in giving a physicalistic account of sensations and perceptions will be relevant for intentional states as well.

A related issue is whether all sensations and perceptions are indeed essentially qualitative, or whether, instead, there could be perception without qualia; that is, whether there could be bona fide perceptual states that are unconscious, such that there is nothing it is like to be in them at all.

Yet another question is whether (and if so, how) the myriad qualia we seem to experience at any given time are bound together at a given moment, and continuous with our experiences at previous and subsequent times, or whether the common sense view that we enjoy a unity of consciousness and a stream of consciousness is rather an illusion to be dispelled.

See also: Consciousness; Sense-data; Colour and qualia

JANET LEVIN

References and further reading

Block, N. (1978) 'Troubles with Functionalism', repr. in D. Rosenthal (ed.) *The Nature of Mind*, New York: Oxford University Press, 1991.(Describes and raises objections to functionalism; home of the 'absent qualia' objection.)

Dennett, D. (1991) *Consciousness Explained*, Boston, MA: Little, Brown.(Argues for eliminativism about qualia, and challenges the view that there is a unified stream of consciousness.)

Hill, C. (1991) *Sensations: A Defense of Type Materialism*, New York: Cambridge University Press.(Argues for a type-identity thesis for sensations.)

Jackson, F. (1982) 'Epiphenomenal Qualia', repr. in W.G. Lycan, *Mind and Cognition*, Cambridge, MA: Blackwell, 1990.(Develops a version of the knowledge argument against a physicalistic reduction of qualia; see §2.)

Kripke, S. (1970) from 'Naming and Necessity', repr. in D. Rosenthal (ed.) *The Nature of Minds*, New York: Oxford University Press.(Presents the neo-Cartesian argument against a physicalist reduction of qualia mentioned in §2.)

Levin, J. (1991) 'Analytic Functionalism and the Reduction of Phenomenal States', in *Philosophical Studies* 61: 211-38.(Expands on the material in §2.)

Levine, J. (1983) ‘Materialism and Qualia: The Explanatory Gap’, *Pacific Philosophical Quarterly* 64: 4. (Introduces the explanatory gap argument, discussed in §2.)

Lewis, D. (1972) ‘Psychophysical and Theoretical Identifications’, repr. in D. Rosenthal (ed.) *The Nature of Minds*, New York: Oxford University Press, 1991. (Argues for the equivalence of qualitative and functional concepts; see §2.)

Lewis, D. (1988) ‘What Experience Teaches’, repr. in W.G. Lycan, *Mind and Cognition*, Cambridge, MA: Blackwell.(Develops the view that what is gained by having new experiences are new abilities rather than knowledge of facts; see §2.)

Loar, B. (1990) ‘Phenomenal States’, in J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 4, Atascadero, CA: Ridgeview.(Introduces and develops the view discussed in §2.)

Nagel, T. (1974) ‘What is it Like to Be a Bat?’, repr. in D. Rosenthal (ed.) *The Nature of Mind*, New York: Oxford University Press, 1991.(Develops a version of the knowledge argument against physicalism, discussed in §2.)

Rosenthal, D. (ed.) (1991) *The Nature of Mind*, New York: Oxford University Press. (Anthology includes a number of articles in this list, and many others relevant to these issues.)

Searle, J. (1992) *The Rediscovery of the Mind*, Cambridge, MA: MIT.(Argues that intentional states must have qualitative character.)

Shoemaker, S. (1990) ‘First-Person Access’, in J. Tomberlin, (ed.) *Philosophical Perspectives*, vol. 4, Atascadero, CA: Ridgeview.(Develops views about self-knowledge discussed in §3.)

Quantification and inference

Quantificational reasoning in natural languages contains occurrences of expressions which, though name-like from a syntactic perspective, intuitively seem to be importantly different from ordinary names. For example, in attempting to argue that every F is G, we might consider an 'arbitrary' F, give 'it' the 'name' 'n', and go on to argue that n is G. That such an occurrence of 'n' is somehow different from occurrences of ordinary names can be seen by noting, for example, that arguing from the claim that Bertrand Russell is F to the claim that Bertrand Russell is G in general does not suffice for drawing the conclusion that every F is G. Similarly, given that some F is G, in general we do not think it legitimate to assert that Bertrand Russell is F and G, and continue reasoning from that claim. It seems that such occurrences of 'n' differ semantically from similar occurrences of ordinary names. Thus the question arises as to the semantics of occurrences of such expressions. An adequate semantic account must justify appropriate inferences to and from sentences containing these terms.

1 Condition of adequacy

Natural languages contain quantificational reasoning employing expressions which function syntactically, but not semantically, as proper names. For example, in attempting to argue that every *F* is *G*, we might consider an 'arbitrary' *F*, give 'it' the 'name' '*n*', and go on to argue that *n* is *G*. Or, having been given that some *F* is *G*, we give the 'name' '*n*' to an arbitrary *F* that is *G*, and go on to draw further conclusions concerning *n*. It will help to shed light on these terms in natural language arguments if we turn our attention to systems of natural deduction for first-order logic (see Natural deduction, tableau and sequent systems). For such systems contain rules which more or less mimic these natural language forms. Though formulations vary significantly, many systems have a 'universal generalization' (UG) rule which licenses the inference of '$(\forall x)\phi(x)$' from '$\phi(a)$' subject to certain restrictions, and an 'existential instantiation' (EI) rule which allows one to infer '$\phi(a)$' from '$(\exists x)\phi(x)$', again subject to restrictions. We shall call singular terms in systems of natural deduction which play the role played by '*a*' above 'instantial terms'. Instead of discussing terms such as '*n*' in our natural language examples, we can ask about the proper semantic treatment of instantial terms in systems of natural deduction. The regimentation and explicitness of structure of natural deduction will enable us to state conditions of adequacy precisely and to see whether those conditions are met by particular semantic theories.

As in natural language, instantial terms in natural deduction appear to behave differently from names or 'individual constants'. Both UG and EI fail to be truth-preserving if the instantial term is an individual constant. Of course, in many formulations of natural deduction, expressions from the syntactic category of individual constant (or individual variable) are pressed into service as instantial terms in applications of UG and EI. However, this probably results more from the desire not to have an additional category of singular terms than from any view about the appropriate semantics for instantial terms.

Before discussing conditions of adequacy for semantic theories of instantial terms, let us consider the view that there can be no semantic theory for instantial terms, because occurrences of formulas in applications of UG and EI which contain instantial terms should not be thought of as having any meaning. On this view, UG and EI have a certain utility in that they allow us to move from meaningful formulas to other meaningful formulas (and soundness and completeness results for systems employing these rules justify their use), but it is a mistake to think that formulas containing instantial terms in applications of UG and EI themselves have any meaning. Whatever attraction this view may have as an account of formulas containing instantial terms in formal systems, it seems quite implausible as an account of natural language sentences containing the analogues of these terms. For surely sentences containing such terms in natural language arguments are meaningful: as we read an argument containing these terms, we understand the claims made by the sentences containing them.

In an adequate semantics all allowable inferences must be truth-preserving. That is, the truth-conditions assigned to a formula must be such that, should they obtain, then so will the truth-conditions of any formula properly inferred from it. So in so far as we provide a semantics for instantial terms and the formulas containing them, it ought to account for the inferential relations such formulas manifest. Let us call this property of a semantics 'inferential adequacy'. It is important to appreciate the difference between inferential adequacy and classical completeness and soundness results. Quine's system of deduction ([1950] 1982) is illustrative. Quine defines a

'finished derivation' as a derivation whose last line neither contains an instantial term (a 'flagged variable', in his terminology) nor depends on a premise which contains an instantial term. He then proves what he calls soundness; that the premises imply the last line of a finished derivation. However, Quine provides no semantics for instantial terms that is inferentially adequate. As he himself remarks, UG and EI are not 'implicative' relative to the only semantics he gives (which essentially treats instantial terms as names).

An inferentially adequate semantics must assign distinct truth-conditions to occurrences of the same formula in different derivations. For suppose ' $(\forall x)\phi(x)$' is correctly inferred from ' $\phi(a)$' in some derivation. Inferential adequacy requires the truth-conditions of this occurrence of ' $\phi(a)$' to be such that it entails ' $(\forall x)\phi(x)$'. On the other hand, an occurrence of ' $\phi(a)$' in a derivation where it is correctly inferred from ' $(\exists x)\phi(x)$' must be assigned truth-conditions such that it is entailed by the latter. Thus ' $\phi(a)$' must have different truth-conditions in the two derivations.

There are different sorts of cases in which a given formula must be assigned different truth-conditions as it occurs in different derivations. Consider the following example.

(A)	(B)
$(1)(\forall x)(\exists y)Fxy$	$(1)(\exists y)(\forall x)Fxy$
$(2)(\exists y)Fay$	$(2)(\forall x)Fxb$
$(3)Fab$	$(3)Fab$
$(4)(\exists w)Faw$	$(4)(\forall z)Fzb$
$(5)(\forall z)(\exists w)Fzw$	$(5)(\exists w)(\forall z)Fzw$

In derivation (A), inferential adequacy requires that '*Fab*' is logically equivalent to ' $(\forall x)(\exists y)Fxy$', from which it is inferred. In (B) it must be logically equivalent to ' $(\exists y)(\forall x)Fxy$'. Note that this is unlike the previous case in that it is not simply a matter of (for example) ' $\phi(a)$' being inferred from ' $(\exists x)\phi(x)$' in one case and having ' $(\forall x)\phi(x)$' inferred from it in another. In both (A) and (B) a formula containing '*a*' has a universal generalization (on the place occupied by '*a*') inferred from it.

2 Three semantic theories of instantial terms

Three semantic accounts of instantial terms have been shown to satisfy the requirement of inferential adequacy. Though other inferentially adequate accounts can be imagined, we shall consider only these three. The first holds that in addition to individual people, numbers and so on there are 'arbitrary' people, numbers and so on. (An arbitrary number has every property common to all numbers, and so on.) Instantial terms are names of these arbitrary or indefinite objects. Prior to Frege's attack on such views (1904), they were widely held and are still sometimes suggested by informal comments accompanying the introduction of UG and EI in logic textbooks. Kit Fine (1985a) has made the view respectable by formulating a precise and powerful version that is inferentially adequate. On Fine's formulation, which arbitrary object a given instantial term names is determined by features of the derivation in which it occurs. An arbitrary object is associated with a range of individual objects (an arbitrary number is associated with a range of individual numbers, and so on), and a predicate is true of the arbitrary object just in case it is true of each individual object in the range. It is this which allows ' $\phi(a)$' to entail ' $(\forall x)\phi(x)$' under the appropriate conditions.

A second account of the semantics of instantial terms holds that they are actually quantifiers. On this view, instantial terms are expressions of generality just like ordinary quantifiers except that the quantifications they express and the scope relations they have to each other are determined by linguistic context. So, for example, in (A) and (B) above, features of the derivations determine whether '*a*' and '*b*' express universal or existential quantifications, and whether '*a*' or '*b*' has wider scope ('*a*' has wide scope in (A) and narrow scope in (B) relative to '*b*'). On this view, suggested in Wilson (1984) and formulated as inferentially adequate in King (1991), we might say that instantial terms are 'context-dependent quantifiers', thus emphasizing both that they express quantifications and that the quantifications they express and the relative scopes of those quantifications vary with the linguistic context. On such a view, ' $\phi(a)$' entails ' $(\forall x)\phi(x)$' (under the appropriate conditions) because ' $\phi(a)$' itself expresses a quantified claim.

The third view is that instantial terms are to be thought of as ε-terms. Hilbert introduced the ε-operator as a variable-binding operator used to form terms: if 'ϕ' is a formula (some presentations require 'ϕ' to contain free '*x*'), then '$\varepsilon x\phi$' is a singular term. Though Hilbert provided no explicit semantics for singular terms formed using the ε-operator, he remarked that '$\varepsilon x\phi$' should be thought of as denoting an arbitrary element in the domain satisfying 'ϕ' if anything does (if nothing satisfies 'ϕ', the term can be thought of as denoting any element in the domain). Further, Hilbert provided the following axiom to govern ε-terms: $(\exists x)\phi(x) \rightarrow \phi(\varepsilon x\phi)$. This axiom justifies the following rules of inference for ε-terms: (1) infer '$\phi(\varepsilon x\phi)$' from '$(\exists x)\phi(x)$'; and (2) infer '$(\forall x)\phi(x)$' from '$\phi(\varepsilon x{\sim}\phi)$'. Thus Hazen (1987) considers constructing a system of natural deduction in which (1) and (2) replace EI and UG respectively, but in which we abbreviate the ε-terms used in applications of (1) and (2) by single letters. Derivations using such abbreviations would be indistinguishable from those in Quine's (1950) system. A semantics for ε-terms of the sort given in Leisenring (1969) would result in an inferentially adequate theory.

Which of the three views discussed above is ultimately preferable appears to depend on philosophical and methodological considerations. In fact, it is not clear how different the views really are. One might well complain that arbitrary objects are unnecessary intermediaries, and that the ranges of individuals with which they are associated do all the real semantic work. But when arbitrary objects are eliminated from the first view and instantial terms are directly associated with ranges of individuals, the first and second views begin to look quite similar. In addition, since, following Hilbert, we can use ε-terms to 'define' the quantifiers ($(\forall x)\phi(x) \leftrightarrow \phi(\varepsilon x{\sim}\phi)$ and $(\exists x)\phi(x) \leftrightarrow \phi(\varepsilon x\phi)$), it is unclear whether the difference between the second view (instantial terms are quantifiers) and the third view (instantial terms are ε-terms) is substantive.

See also: Logical and mathematical terms, glossary of

JEFFREY C. KING

References and further reading

Fine, K. (1983) 'A Defence of Arbitrary Objects', *Proceedings of the Aristotelian Society*, supplementary vol. 57: 55-77.(An informal introduction to Fine's theory of arbitrary objects. The least technical of Fine's writings on the subject.)

Fine, K. (1985a) *Reasoning with Arbitrary Objects*, Oxford: Blackwell.(Fine's definitive and most technical work on arbitrary objects.)

Fine, K. (1985b) 'Natural Deduction and Arbitrary Objects', *Journal of Philosophical Logic* 14: 57-107.(A fairly technical work which presents some highlights of Fine (1985a).)

Frege, G. (1904) 'Was ist eine Funktion?', trans. 'What is a Function?', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 1977.(An attack on the view that instantial terms name indefinite objects.)

Hazen, A. (1987) 'Natural Deduction and Hilbert's ε-Operator', *Journal Of Philosophical Logic* 16: 411-21.(An outline of the view that instantial terms can be thought of as abbreviations of ε-terms. Informally written, but presupposes significant exposure to first-order logic and Fine (1985).)

King, J.C. (1987) 'Pronouns, Descriptions and the Semantics of Discourse', *Philosophical Studies* 51: 341-63. (Introduction to the view that instantial terms are contextually sensitive quantifier-like expressions of generality; very accessible.)

King, J.C. (1991) 'Instantial Terms, Anaphora and Arbitrary Objects', *Philosophical Studies* 61: 239-65. (Compares the account of instantial terms as context-dependent quantifiers to the account of Fine (1983, 1985). The presentation is informal; technical material in the appendix.)

Leisenring, A.C. (1969) *Mathematical Logic and Hilbert's ε-Symbol*, New York: Gordon & Breach.(The best English introduction to formal systems containing ε-terms; technical, but self-contained.)

Quine, W.V. (1950) *Methods of Logic*, New York: Holt; repr., Cambridge, MA: Harvard University Press, 4th edn, 1982.(System of natural deduction discussed in both Hazen (1987) and Fine (1985). The 1950 and 1982 editions differ significantly.)

Wilson, G. (1984) 'Pronouns and Pronominal Descriptions: A New Semantical Category', *Philosophical Studies* 45: 1-30.(The first paper to suggest that instantial terms are quantifier-like expressions of generality; very accessible.)

Quantifiers

The quantifiers 'some' and 'every' were the object of the very first logical theory, Aristotelian syllogistic. An example of a syllogism is 'Every Spartan is Greek, every Greek is European, therefore every Spartan is European'. In such inferences, no quantifier is governed by another one. Contrast this with 'Everybody loves somebody'. Modern logic is often taken to have begun when Frege systematized for the first time the logic of quantifiers, including such dependent ones. In general, much of what has passed as logic over the centuries is in effect the study of quantifiers. This is especially clear with the area of logic variously known as quantification theory, lower predicate calculus or elementary logic. Some philosophers have even sought to limit the scope of logic to such a study of quantifiers. Yet the nature of quantifiers is a delicate matter which is captured incompletely by the logic initiated by Frege and Russell.

1 Interpretation of the quantifiers

Traditionally, quantifiers have been taken to be those expressions which somehow indicate the size of the extension of a predicate (see Logic, ancient §§1-3; Predicate calculus), for example, expressions such as 'some' (the existential quantifier), 'every' (the universal quantifier), the numerical quantifiers 'one', 'two', 'three',…, 'many', 'few' and so on. The logical properties of the existential and universal quantifiers are studied in quantification theory or first-order logic. The numerical quantifiers are definable in terms of the existential and universal quantifiers, unlike quantifiers such as 'many' and 'few'. For this reason the latter are sometimes called nonstandard quantifiers. Here we shall be concerned mainly with the standard quantifiers 'some' and 'every'. When such quantifiers are applied to higher-order entities (predicates, relations, functions, predicates, relations and functions of predicates, relations and functions and so on), we are dealing with higher-order quantifiers (see Second- and higher-order logics).

The formulas of a first-order language *L* are built out of a set of nonlogical constants and individual variables by means of propositional connectives, the identity symbol and quantifiers. Typically, the formation rules for quantifiers allow us to take a formula with a free variable *x* and to prefix to it the existential quantifier '$\exists x$' or the universal quantifier '$\forall x$'. The formula is then called the 'scope' of the quantifier in question (see Scope). In the resulting formula - for example, $(\exists x)(\ldots x \ldots)$ - the variable *x* is said to be 'bound' by the quantifier '$\exists x$'. The nonlogical constants can be predicates, function symbols or individual constants.

A first-order formal language allows us to systematize the laws of quantification. Usually, a number of logically true (valid) formulas are assumed as logical axioms and a number of rules of inference are used to derive other logical truths (ideally, all of them) as theorems. A complete axiomatization of ordinary quantification theory was first given by Gödel in 1929. A theoretically more illuminating completeness proof was given by Henkin in 1949.

Quantification theory must be considered semantically, not only as a purely formal system. To specify the meaning of an interpreted first-order sentence is to specify its truth-conditions. Usually, such a specification of truth-conditions is done step by step, starting from inside, and at each step indicating how the truth or falsity of an expression depends on the truth or falsity of its component expressions. Such specifications of meaning are said to satisfy the principle of 'compositionality' (see Compositionality).

Any specification of truth-values must be relative to a model of the underlying language. The set of elements (individuals) of this model is called its domain. A quantifier is often said to range over this domain.

The truth-value of an interpreted and quantified sentence, say, $(\forall x)(\ldots x \ldots)$, cannot be specified in terms of the truth-value of the formula $(\ldots x \ldots)$, for the latter contains a free variable and hence does not have a truth-value. The truth-value of $(\forall x)(\ldots x \ldots)$ therefore has to be defined as a function of all the substitution instances of $(\ldots x \ldots)$ or else through the other semantic attributes of $(\ldots x \ldots)$. Tarski, who was the first to formulate an explicit truth-definition for quantificational languages (1933), chose the latter course and defined truth for first-order languages via the notion of satisfaction, so preserving the compositionality of his procedure (see Tarski's definition of truth).

Tarski's truth-definition is a formal specification of truth-conditions of sentences containing quantifiers, not yet a genuine interpretation of quantifiers. There are three main interpretations of quantifiers in the literature.

(A) *Quantifiers as higher-order predicates*. On this view, what, for example, the existential quantifier does in a sentence such as

(1) $(\exists x)S(x)$

is to say that the complex or simple predicate *S*(*x*) is not empty. This is the approach we described above and is *prima facie* quite natural. It goes back to Frege. Its most important systematization and further development is the theory of so-called generalized quantifiers, currently the dominant trend in the study of quantifiers among theoretical linguists (see Quantifiers, generalized).

(B) *Substitutional interpretations of quantifiers*. On such an interpretation, the import of a sentence such as (1) is explained in terms of the set of different substitution-instances of the open formula S(x) (see Quantifiers, substitutional and objectual).

(C) *Quantifiers as choice terms*. The basic idea can be explained by example. A natural interpretation of the sentence

(2) Every man owns a dog

is to say that for any man given to one, one can choose a dog such that the given man owns the chosen dog. In the formal counterpart of (2),

(3) $(\forall x)(\exists y)O(x, y)$,

we would say that, for every value of *x*, one can find a value of *y* such that *O*(*x*, *y*). We can say that on this view quantifiers codify suitable choice terms. This way of interpreting quantifiers is systematized and generalized in Hintikka's game-theoretic semantics (see Hintikka and Sandu 1996).

2 The game-theoretic approach

In this game-theoretic treatment, a sentence such as (3) above is interpreted via a semantic game played by 'myself' (the initial verifier) against 'nature' (the initial falsifier), the former trying to produce true sentences while the latter is striving for the contrary. The universal quantifier '$\forall x$' in (3) marks a move by 'nature', who must choose an individual, say *d*, from the universe of the model in which (3) is evaluated, and the existential quantifier '$\exists y$' marks a move by 'myself', who must also choose an individual, say *e*. The sentence (3) is defined to be true if 'myself' has a method (usually codified by a mathematical function *f*) which, for every individual *d* chosen by 'nature' gives him the 'right' choice *e*, that is, $O(d, e)$ holds. More precisely, $O(d, f(d))$ holds (see Semantics, game-theoretic).

The central notion involved in interpretation (C) is that of a strategy, that is, a rule which specifies which move a player is to make in any possible situation which might come up in the game. The idea of an 'independent' quantifier is captured by that of an informationally independent choice. For instance, in (3) the choice of a value of y depends on the choice of a value of *x* in the sense that the choice connected with y is made on the basis of the player's knowledge of the value that was chosen for *x*, but not vice versa. In the jargon of game-theory, a game associated with a sentence of ordinary first-order language is one of perfect information.

In the usual formal (but interpreted) first-order languages, the three interpretations (A)-(C) assign the same truth-value to each sentence. Nevertheless, several things distinguish (C) from (A) and (B).

First, only interpretation (C) does justice to the fact that the expressive power of first-order languages lies in the dependence of quantifiers on each other. To understand first-order logic is, in this sense, to understand the idea of a dependent quantifier. In contrast, in (A) and (B), quantifiers are considered in isolation from each other. Interpretations of type (C) guided the work of David Hilbert and his school in the 1920s and 1930s see (Hilbert and Bernays 1934) and similar views have enjoyed a widespread albeit inconspicuous currency in mathematical practice.

Second, (C), unlike (A), yields a non-compositional interpretation of the standard quantifiers. An existential quantifier '$\exists y$' is interpreted in (C) by a choice function whose arguments are certain universally quantified

variables. Which ones they are depends on the context.

Interpretation (C) offers an interesting possibility of generalizing quantification theory (first-order logic). If one understands the idea of a dependent quantifier as illustrated by (3), one *ipso facto* understands the idea of an independent quantifier. What often passes unnoticed is that the usual nested ordering of scopes for first-order logic rules out certain perfectly possible and understandable patterns of mutual dependence and independence between quantifiers, in that it is required that the scopes of two different quantifiers are either exclusive or nested, that is, that their scopes do not overlap only partially.

But what happens when this arbitrary restriction is given up? As an example, consider the first-order formula

(4) $(\forall x)(\exists y)(\forall z)(\exists w)S(x,y,z,w)$.

The question is what happens if in (4) we require '$\exists w$' to be within the scope of '$\forall z$' but not in that of '$\forall x$'? There is no way of expressing this complex of dependencies in the usual linear notation of first-order logic. For that purpose, a different notation has to be introduced. A special kind of informational independence has been studied by logicians under the title of 'branching quantifiers'. They were introduced by Henkin in 1959. A more general way of indicating independence is to use a slash notation:

(5) $(\forall x)(\exists y)(\forall z)(\exists w/\forall x)S(x,y,z,w)$.

The idea here is that the last existential quantifier does not depend on '$\forall x$'.

What can be said of quantifiers can also be said, *mutatis mutandis*, of disjunctions and conjunctions: the independence indicator slash can be applied to them too. The languages using the slash notation are called independence-friendly first-order languages.

Neither approach (A) nor approach (B) works for independence-friendly first-order languages. The reason is that such a language involves context dependencies which make a compositional interpretation unrealistic. For a compositional interpretation requires that the meaning of an expression be specified in terms of its component parts. This is not the case with the expression $(\exists w/\forall x)S(x,y,z,w)$ in (5), for its interpretation makes reference to an element '$\forall x$' which occurs outside it.

In contrast, interpretation (C) extends naturally to cover this new language. All we need is to drop the requirement that the semantic game associated with a sentence such as (5) is one of perfect information.

What has been said has a number of consequences not usually noted in the study of quantifiers.

(1) The meaning of quantifiers is often explained in terms of their 'ranging over' a set of entities. Such an explanation presupposes an interpretation along the lines of (A) or (B) and hence is not a general explanation of the meaning of quantifiers.
(2) The notion of quantifier scope is ambiguous. The usual formal scope notation indicates two different things. On the one hand, it indicates the relative logical priority of different quantifiers (and connectives); on the other, it indicates the segment of a formula (or a discourse) in which a variable is bound by a given quantifier. There is no reason to expect that these two always go together, and in natural languages they seem to involve entirely different mechanisms.
(3) Certain types of quantifiers, namely informationally independent ones, do not admit of a natural compositional truth-definition, but have to be treated in other ways semantically.
(4) A general theory of quantification will have to comprise more than is currently known as quantification theory or ordinary first-order logic.
(5) Since independence-friendly first-order logic is not axiomatizable, it is unrealistic to expect the logic of quantifiers to admit of a complete systematization.
(6) The logical behaviour of quantifiers in natural languages seems to differ more radically from ordinary formal first-order logic than has usually been recognized.

See also: Logical constants; Logical and mathematical terms, glossary of; Quantification and inference

JAAKKO HINTIKKA
GABRIEL SANDU

References and further reading

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. J. van Heijenoort, '*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967. (First explicit logic of quantifiers.)

Gödel, K. (1929) 'Über die Vollständigkeit des Logikkalküls' (doctoral dissertation, University of Vienna), in *Kurt Gödel: Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay and J. van Heijenoort, Oxford: Oxford University Press, 1986, 60-101.(First completeness proof for ordinary first-order logic of quantifiers.)

Goldfarb, W. (1979) 'Logic in the Twenties: The Nature of the Quantifier', *Journal of Symbolic Logic* 44: 351-68.(Changing conceptions of quantifiers from Frege to Hilbert.)

Henkin, L. (1949) 'Completeness in the Theory of Types', *Journal of Symbolic Logic* 14: 159-66.(Greatly improved completeness proof.)

Henkin, L. (1959) 'Some Remarks on Infinitely Long Formulas', in *Infinitistic Methods*, Oxford: Pergamon, 167-83.(First use of informationally independent quantifiers.)

Hilbert, D. and Bernays, P. (1934) *Grundlagen der Mathematik I-II*, Berlin: Springer.(Quantifiers and choice functions.)

Hintikka, J. and Sandu, G. (1996) 'Game-Theoretical Semantics', in J. van Benthem and A. ter Meulen (eds) *Handbook of Logic and Language*, Amsterdam: Elsevier.(A survey of game-theoretic semantics.)

Hintikka, J. and Kulas, J. (1985) *Anaphora and Definite Descriptions*, Dordrecht: Reidel.(Quantifiers in natural languages.)

Kripke, S.A. (1976) 'Is There a Problem about Substitutional Quantification?', in G. Evans and J. McDowell (eds) *Truth and Meaning*, Oxford: Oxford University Press, 325-419.(A spirited defence of the substitutional interpretation.)

Tarski, A. (1933) *Pojęcie prawdy w jezykach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956), 'On the Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(The classic work on the concept of truth.)

Westerståhl, D. (1989) 'Quantifiers in Formal and Natural Languages', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 4, 1-131.(A survey of the theory of generalized quantifiers.)

Quantifiers, generalized

Generalized quantifiers are logical tools with a wide range of uses. As the term indicates, they generalize the ordinary universal and existential quantifiers from first-order logic, ' $\forall x$*' and '* $\exists x$*', which apply to a formula A(x), binding its free occurrences of x.* $\forall x A(x)$ *says that A(x) holds for all objects in the universe and* $\exists x A(x)$ *says that A(x) holds for some objects in the universe, that is, in each case, that a certain condition on A(x) is satisfied. It is natural then to consider other conditions, such as 'for at least five', 'at most ten', 'infinitely many' and 'most'. So a quantifier Q stands for a condition on A(x), or, more precisely, for a property of the set denoted by that formula, such as the property of being non-empty, being infinite, or containing more than half of the elements of the universe. The addition of such quantifiers to a logical language may increase its expressive power.*

A further generalization allows Q to apply to more than one formula, so that, for example, Qx(A(x),B(x)) states that a relation holds between the sets denoted by A(x) and B(x), say, the relation of having the same number of elements, or of having a non-empty intersection. One also considers quantifiers binding more than one variable in a formula. Qxy,zu(R(x,y),S(z,u)) could express, for example, that the relation (denoted by) R(x,y) contains twice as many pairs as S(z,u), or that R(x,y) and S(z,u) are isomorphic graphs.

In general, then, a quantifier (the attribute 'generalized' is often dropped) is syntactically a variable-binding operator, which stands semantically for a relation between relations (on individuals), that is, a second-order relation. Quantifiers are studied in mathematical logic, and have also been applied in other areas, notably in the semantics of natural languages. This entry first presents some of the main logical facts about generalized quantifiers, and then explains their application to semantics.

1 The general concept

A (generalized) quantifier is a relation between relations on individuals (in other words, a second-order relation). Its *type*, a finite sequence of positive numbers, indicates how many relations and of what degree. For example, a quantifier of type $\langle 2,1,1,4\rangle$ is a relation between one binary relation (that is, a relation of order 2), two sets and one four-place relation. Let *M* be a non-empty universe. Then define a (local) quantifier of type $\langle 2,1,1,4\rangle$ *on M* to be any relation between a binary relation on *M*, two subsets of *M* and a four-place relation on *M*. Then a (global) quantifier *Q* of type $\langle 2,1,1,4\rangle$ is a function which associates with each universe *M* a quantifier Q_M of the same type on *M*.

The quantifiers '$\forall$' and '$\exists$' are thus of type $\langle 1\rangle$. Below are definitions of these and some other type $\langle 1\rangle$ quantifiers, with *M* any universe, *A* any subset of *M*, and $|A|$ the cardinality of *A* (Q^R is called the Rescher quantifier):

$$\begin{aligned}
\forall_M A &\Leftrightarrow A = M \\
\exists_M A &\Leftrightarrow A \\
(\exists_{\geq 5})_M A &\Leftrightarrow |A| \geq 5 \\
(\exists_{\leq 10})_M A &\Leftrightarrow |A| \leq 10 \\
(Q_0)_M A &\Leftrightarrow A \\
(Q^R)_M A &\Leftrightarrow |A| > |M - A|
\end{aligned}$$

Here are some more examples of quantifiers (types are written to the left):

$$\begin{array}{ll}
\langle 1,1\rangle & \mathrm{I}_M AB \Leftrightarrow |A| = |B| \quad \text{(the Härtig quantifier)} \\
\langle 1,1,1\rangle & \mathit{more\ldots than}_M ABC \Leftrightarrow |A \cap C| > |B \cap C| \\
\langle 2\rangle & \mathrm{W}_M R \Leftrightarrow R \text{ is a well-order of } M \\
\langle 2,2\rangle & \mathrm{IG}_M RS \Leftrightarrow R \text{ and } S \text{ are isomorphic graphs}
\end{array}$$

The notion of a type $\langle 1\rangle$ quantifier was introduced by Andrzej Mostowski (1957), and extended to arbitrary types by Per Lindström (1966). Aristotle's syllogisms (see Logic, ancient §§1-3) involve type $\langle 1,1\rangle$ quantifiers, a natural class, as will be seen in §3. The general notion of quantifiers as variable-binding operators denoting

second-order relations was in fact explicit in Frege's work, except that Frege did not relativize to an arbitrary universe (see Frege, G. §6).

The quantifiers above share an important property, closure under isomorphism (or ISOM). We use the following notation: a function *f* from a set *X* to a set *Y* is 'lifted' to subsets *A* of *X*, or to (for example) ternary relations *R* on *X*, by $f(A)=\{f(a)\colon a\in A\}$ and $f(R)=\{(f(a),f(b),f(c))\colon (a,b,c)\in R\}$. Now for a quantifier *Q* of type $\langle 2,1,1,4\rangle$, say, define *Q* to be *closed under isomorphism* if whenever *f* is a bijection from *M* to M' (a one-one function from *M* onto M'), and *R*, *A*, *B*, *S* are appropriate arguments for Q_M,

$$Q_M RABS \Leftrightarrow Q_{M'} f(R)f(A)f(B)f(S).$$

ISOM says that *Q* does not care about the individual elements of a universe, only about their 'structure' with respect to the relational arguments. This allows quantifiers to be treated as logical constants (see §2 below; Logical constants §3). ISOM has a particularly clear meaning for quantifiers which relate sets, so-called 'monadic' quantifiers (that is, of type $\langle 1,1,\ldots,1\rangle$). It then says that *Q* only cares about the cardinalities of these sets (and sets that can be formed from them by taking unions, intersections and complements), not the sets themselves. For example, for a type $\langle 1,1\rangle$ quantifier, ISOM is equivalent to the following condition:

If $A,B\subseteq M$, $A',B'\subseteq M'$, $|A\cap B|=|A'\cap B'|$, $|A-B|=|A'-B'|$, $|B-A|=|B'-A'|$, and $|M-(A\cup B)|=|M'-(A'\cup B')|$, then $Q_M AB \Leftrightarrow Q_{M'} A'B'$.

Thus, in the monadic case, ISOM quantifiers deal with *quantities*, as the name suggests they should. Indeed, such a quantifier can be redefined as a relation between numbers; the examples above provide illustrations.

2 Quantifiers in logic

Generalized quantifiers are easy to add as new logical constants to first-order logic, FO. For example, '*A* is finite' cannot be expressed in FO, but if we add the quantifier Q_0 above, the sentence $\neg Q_0 xPx$ expresses precisely that the denotation of *P* is finite. If *Q* is a quantifier of type $\langle 2,1,1\rangle$, say, the logic FO(*Q*) is formed from FO by adding a formation rule allowing formulas of the form

(1) $Qxy,z,u(A(x,y),B(z),C(u)),$

where $A(x,y)$, $B(z)$ and $C(u)$ are formulas. Free occurrences of the indicated variables become bound in (1). To interpret (1) in a model M with universe *M*, note that relative to M and the variables *x*, *y* (and a variable assignment which we suppress here), $A(x,y)$ denotes a binary relation $A(x,y)^{\mathrm{M}}$ on *M*, namely, the set of pairs (a,b) such that $A(a,b)$ holds in M. Then, (1) is defined to be true in M if and only if $Q_M A(x,y)^{\mathrm{M}} B(z)^{\mathrm{M}} C(u)^{\mathrm{M}}$. This generalizes the usual clauses for $\forall xA(x)$ and $\exists xA(x)$ in the truth-definition for FO. Logical consequence in FO(*Q*) is defined as usual (the conclusion must be true in every model of the premises - see Model theory). And just as FO can be used to formulate theories, like arithmetic or group theory (first-order theories), so can FO(*Q*). The properties of these theories, and of the consequence relation, will of course depend on *Q*.

Similarly, one defines logics of the form FO($Q_1,\ldots,Q_n$). This is a model-theoretic conception of a logic as given roughly by a class of sentences, a class of models and a truth-relation between them (in contrast, a proof-theoretic conception would emphasize a formal inference system). Some basic constraints apply; for example, only ISOM quantifiers are considered.

Abstract model theory studies the properties of model-theoretic logics in general, logics with generalized quantifiers being prime examples. Among interesting properties of a logic *L* are axiomatizability (existence of a formal inference system proving precisely the valid sentences of *L*), compactness (if every finite subset of a set Φ of *L*-sentences has a model then the whole of Φ has a model) and the Löwenheim property (if an *L*-sentence has an infinite model it has a denumerable model; see Löwenheim-Skolem theorems and nonstandard models §§2, 3). FO has these properties and several stronger properties as well, and a celebrated result of abstract model theory is that such properties in fact characterize FO. To state this one needs a way to compare the expressive power of various logics. A logic L' is said to be an extension of a logic *L*, $L\leq L'$, if everything that can be said in *L* can also be said in L', in other words if for every *L*-sentence there is a logically equivalent L'-sentence (one with the

same models). Also, $L \equiv L'$ if $L \leq L'$ and $L' \leq L$. Note that if *L* is a logic with generalized quantifiers, FO $\leq L$. Given a simple assumption about these quantifiers, one can prove the following result (Lindström's theorem):

If *L* is either compact or axiomatizable, and has the Löwenheim property, then $\mathrm{FO} \equiv L$.

Thus, increase of expressive power entails loss of some of FO's properties. For example, if '*A* is finite' is expressible in *L*, it can be seen that *L* can be neither compact nor axiomatizable. But there do exist compact and axiomatizable proper extensions of FO (hence the Löwenheim property fails for them), the simplest example being FO(Q_1), where Q_1 is the quantifier 'there exist uncountably many'. Extending FO should be done with care, to strike a balance between expressive power and model-theoretic behaviour.

How many different quantifiers (of a given type) are there? Locally, the rough answer is: many times more than the size of the universe (see Westerståhl 1989: §4.6 for more detail). For example, if *M* has *n* elements, then there are 2^{4^n} type $\langle 1,1 \rangle$ quantifiers on *M*. So there are $2^{16} = 65{,}536$ type $\langle 1,1 \rangle$ quantifiers on a universe with 2 elements! Counting only ISOM quantifiers the number is 1,024, still large compared to 2. As to global quantifiers, even if we restrict attention to finite universes there are uncountably many ISOM quantifiers of each type.

Some of these different quantifiers will have the same expressive power. For example, $\exists_{\geq 5}$ is definable in FO, so FO $(\exists_{\geq 5}) \equiv$ FO. In general, a quantifier *Q* is said to be definable in a logic *L* if it is defined by some *L*-sentence whose nonlogical symbols match the type of *Q*. So for a *Q* of type $\langle 2,1,1 \rangle$ to be definable in *L* there must exist an *L*-sentence with nonlogical symbols P_1, P_2, and P_3, with P_1 binary and P_2, P_3 unary, which is logically equivalent to $Qxy, z, u(P_1xy, P_2z, P_3u)$. We have FO $(Q_1, \ldots, Q_n) \leq L$ if and only if each Q_i is definable in *L*.

The number of non-equivalent logics of the form FO(*Q*) is still uncountable, for each type of *Q*. Moreover, the types yield a strict hierarchy of expressive power. Types can be compared with respect to the largest degree of the relation arguments. For example, $\langle 2,1,1,3,3 \rangle$ is lower than $\langle 2,4 \rangle$. A more fine-grained (partial) order also takes into account the number of arguments of each degree. Then $\langle 3,1,2,3 \rangle$ is lower than $\langle 2,1,1,3,3 \rangle$ since it has fewer unary arguments but the same number of binary and ternary arguments. It can be proved that in each type there are quantifiers not definable in any logic FO($Q_1, \ldots, Q_n$) with $Q_1, \ldots, Q_n$ of lower types (in the more fine-grained order). Again this holds even if only finite models are considered.

The study of quantifiers in mathematical logic has either been in the direction of abstract model theory or focused on logics with particular quantifiers (and theories formulated in such logics), a prominent case being FO(Q_1) and certain of its extensions. Quantifiers have also found applications outside mathematics. Restricted to finite models, they have turned up in computer science, in attempts to find extensions of FO 'capturing' certain complexity classes from the theory of computational complexity. They have been applied to modal logic, and to the study of conditionals. But perhaps the most striking application concerns the semantics of natural languages, to which we now turn.

3 Quantifiers and natural language

Consider the following sentence form, common to many languages.

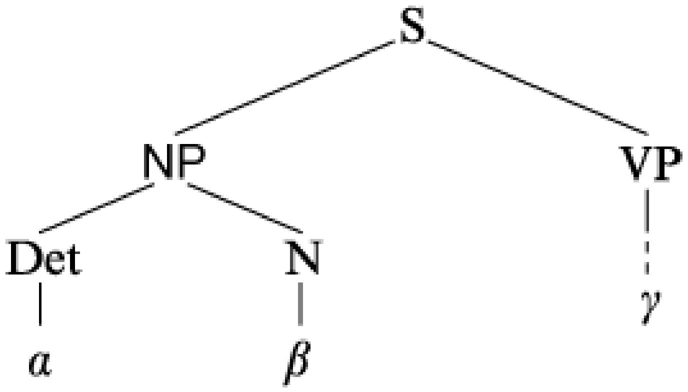

The sentence (S) consists of a noun phrase (NP) and a verb phrase (VP), and the NP in turn consists of a determiner (Det) and a noun (N). A simple English example is

(2) No student smokes

but Det, N and VP could be complex expressions. Now, given a universe *M*, N typically denotes a subset *A* of *M*, and VP another subset, *B*. Thus, Det is most straightforwardly taken to denote a binary relation between subsets of *M*, that is, a type $\langle 1,1\rangle$ quantifier on *M*. Varying *M*, Det denotes a global type $\langle 1,1\rangle$ quantifier.

In (2), this quantifier is $no_M AB \Leftrightarrow A\cap B=\emptyset$. Other Dets denote other quantifiers:

$$\begin{aligned} all_M AB &\Leftrightarrow A\subseteq B\\ at\ least\ five_M AB &\Leftrightarrow |A\cap B|\geq 5\\ most_M AB &\Leftrightarrow |A\cap B|>|A-B|\\ all\ but\ three_M AB &\Leftrightarrow |A-B|=3\\ more\ than\ two\text{-}thirds\ of\ the_M AB &\Leftrightarrow |A\subseteq B|>\tfrac{2}{3}|A|\\ all\ldots except\ John_M AB &\Leftrightarrow A-B=\{j\} \end{aligned}$$

The last example comes from a ‘discontinuous’ Det, as in, for example, ‘All students except John smoke’. A few Dets take more than one noun argument, as in

More students than teachers smoke,

which involves the type $\langle 1,1,1\rangle$ quantifier ‘more…than’ mentioned in §1.

Furthermore, NPs naturally denote type $\langle 1\rangle$ quantifiers: ‘at least five students’ denotes the set of subsets of *M* containing at least five students. In general one obtains such NP denotations from Det denotations by fixing a particular noun. Also NPs without Dets, such as ‘John’, or ‘Bill and Sue’, may be taken to denote type $\langle 1\rangle$ quantifiers:

$$\begin{aligned} &John_M B \Leftrightarrow j\in B\\ &Bill\ and\ Sue_M B \Leftrightarrow b\in B\ \&\ s\in B \end{aligned}$$

This treatment of proper names may seem complex, but has the advantage of allowing all NPs to be interpreted in the same way.

Thus, generalized quantifiers appear in a linguistic context not by means of some technical trick or invention, but rather as the most straightforward way to treat the semantics of noun phrases.

Still, is it really necessary to bring in generalized quantifiers here? After all, ‘all’ can be defined in FO by $\forall x(Ax\rightarrow Bx)$ and, similarly, quantifiers such as ‘at least five’. Nevertheless, sticking to FO is not advisable. First, the use of type $\langle 1,1\rangle$ quantifiers respects the syntactic form of sentences whereas a translation into FO does not. So generalized quantifiers facilitate a compositional semantics for natural language (see Compositionality). Second, in important cases a translation into FO is simply impossible. For example, proportional quantifiers such as ‘most’ or ‘more than two thirds of the’ are undefinable in FO. That is, there is no sentence in first-order logic, however complex, which adequately accounts for the truth-conditions of, say, ‘Most students smoke’. In fact, one can even prove that there is no such sentence in any logic FO($Q_1,\ldots,Q_n$) with $Q_1,\ldots,Q_n$ of type $\langle 1\rangle$. So type $\langle 1,1\rangle$ quantifiers are crucial also with respect to expressive power.

Most Dets have ISOM denotations, with putative exceptions including ‘every…but John’ or ‘John’s’. If ‘John’ denotes a fixed individual ISOM will fail, but there are various strategies for avoiding such Dets. The salient property of Det denotations, however, is the special role played by the first argument, coming from the noun: it restricts the domain of quantification. Formally this is reflected in two constraints, called ‘conservativity’ and ‘extension’: for all *M*, $A,B\subseteq M$ and $M\subseteq M'$,

$$\begin{aligned} &\text{CONS}\quad Q_M AB \Leftrightarrow Q_M A\ A\cap B\\ &\text{EXT}\quad Q_M AB \Leftrightarrow Q_{M'} AB \end{aligned}$$

CONS says that only that part of the verb argument matters which is common to the noun argument. Indeed, ‘All but three students smoke’ is equivalent to ‘All but three students are students which smoke’. CONS predicts, for example, that the Härtig quantifier $rmI_M AB \Leftrightarrow |A| = |B|$ (for which CONS fails) is not the denotation of any Det. EXT implies that the rest of the universe - the part lying outside the noun denotation - is irrelevant. It predicts, for example, that there can be no Det meaning ‘some’ on universes with, say, fewer than ten elements, and meaning ‘most’ on other universes.

Discounting a few marginal (and debatable) exceptions, all Det denotations appear to satisfy CONS and EXT. Other properties single out subclasses of quantifiers. Call *Q* ‘intersective’ if only the intersection of the noun and the verb argument matters; $A \cap B = A' \cap B'$ implies $Q_M AB \Leftrightarrow Q_M A'B'$. Examples are ‘some’, ‘at least five’, ‘exactly three’ and ‘only finitely many’. This turns out to be related to the acceptability of the corresponding Dets in ‘there’-constructions: ‘There are some/at least five/no cows in the barn’ is fine, but ‘There are all but five/most/all cows in the barn’, with non-intersective Dets, is not.

Monotonicity properties are also significant. *Q* is ‘right monotone increasing’ (or decreasing) if $Q_M AB$ and $B \subseteq B'$ (or $B' \subseteq B$) implies $Q_M AB'$. Similarly for monotonicity in the left argument. Many Dets denote right monotone quantifiers. For example, ‘more than two thirds of the’ is right monotone increasing, and ‘at most one third of the’ is right monotone decreasing (neither is left monotone). And many non-monotone cases, such as ‘(exactly) five’, ‘between three and five’ and ‘ten per cent of’, are conjunctions of a decreasing and an increasing quantifier. Right monotonicity has been related to various linguistic phenomena, in particular so-called ‘negative polarity items’. Left monotonicity is more scarce among Det denotations. It may be shown that, over finite models and assuming CONS and EXT, all left monotone quantifiers are first-order definable. Left monotone Det denotations are usually also right monotone. The four possible kinds of ‘double monotonicity’ are exemplified in the Aristotelian square of opposition:

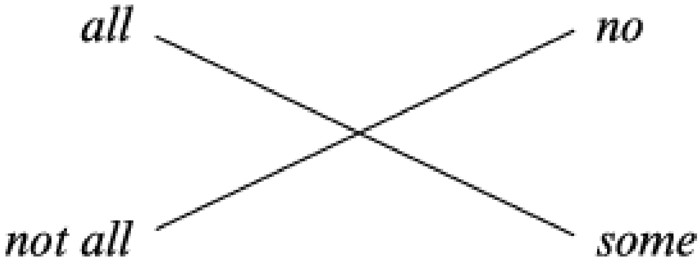

Indeed, given CONS and a certain uniformity requirement, it can be proved that these are the *only* doubly monotone quantifiers (without this requirement there are other examples: ‘at most ten’, ‘all but at most three’, ‘infinitely many’).

NP and Det denotations are normally monadic: they relate sets. But sentences can combine several NPs with a transitive verb:

(3) Most students know at least three teachers.

The truth-conditions of (3) are an *iteration* of the two type $\langle 1, 1 \rangle$ quantifiers involved: formally (with *A* the set of students, *B* the set of teachers and *R* the relation of knowing), $most_M A\{a\ at\ least\ three_M B\{b\ Rab\}\}$. However, other seemingly similar sentences combine the quantifiers differently, for example:

(4) 1,200 philosophers wrote 2,000 entries for the Encyclopedia,

which does *not* mean that each of the 1,200 philosophers wrote 2,000 entries, as an iterated reading would have it, but rather that each of them wrote *some* entry, and that each of the 2,000 entries was written by *some* one (or more) of them. This mode of combination has been called ‘cumulation’. Another suggested mode is ‘branching’:

(5) Most of the students and most of the teachers all know each other

appears to have the reading that there is a large subset *X* of the set of students (containing more than half of them) and a large subset *Y* of the set of teachers such that every student in *X* and every teacher in *Y* know each other. Again this is not an iterated reading. But note that (3), (4) and (5) can be accounted for by quantifiers of type

$\langle 1, 1, 2 \rangle$, taking two set arguments from the nouns and one relation argument from the verb. For (3) and (4) this quantifier is definable from the monadic Det denotations, but not so for (5). More precisely, it can be shown that the branching of ‘most’ as in (5) is not definable in FO/(‘most’). Examples like this indicate a possible use also for polyadic (non-monadic) quantifiers in the semantics of natural languages.

See also: Logical and mathematical terms, glossary of

DAG WESTERSTÅHL

References and further reading

Barwise, J. and Feferman, S. (eds) (1985) *Model-Theoretic Logics*, Berlin: Springer. (Classical, handbook-style reference for abstract model theory and the mathematics of quantifiers.)

Benthem, J. van (1986) *Essays in Logical Semantics*, Dordrecht: Reidel. (Includes some of the papers that initiated the subject of generalized quantifiers and natural language.)

Benthem, J. van and Westerståhl, D. (1995) ‘Directions in Generalized Quantifier Theory’, *Studia Logica* 55: 389-419. (A quick survey of work on generalized quantifiers at the time of writing.)

Keenan, E.L. and Westerståhl, D. (1997) ‘Generalized Quantifiers in Linguistics and Logic’, in J. van Benthem and A. ter Meulen (eds) *Handbook of Logic and Language*, Amsterdam: Elsevier, 837-93. (Updates the exposition in Westerståhl (1989), from a more linguistic perspective.)

Krynicki, M., Mostowski, M. and Szczerba, L.W. (eds) (1995) *Quantifiers: Logics, Models and Computation*, vol. 1, *Surveys*, Dordrecht: Kluwer. (Surveys mathematical work on quantifiers between the time of appearance of Barwise and Feferman (1985) and 1992.)

Lindström, P. (1966) ‘First-Order Predicate Logic with Generalized Quantifiers’, *Theoria* 32: 186-95. (The extension to arbitrary types of the notion of a type $\langle 1 \rangle$ quantifier.)

Mostowski, A. (1957) ‘On a Generalization of Quantifiers’, *Fundamenta Mathematicae* 44: 12-36. (The introduction of the notion of a type $\langle 1 \rangle$ quantifier.)

Westerståhl, D. (1989) ‘Quantifiers in Formal and Natural Languages’, in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, vol. 4, Dordrecht: Reidel, 1-131. (Detailed exposition of much of the material in this entry.)

Quantifiers, substitutional and objectual

Understood substitutionally, 'Something is F' is true provided one of its substitution instances (a sentence of the form 'a is F') is true. This contrasts with the objectual understanding, on which it is true provided 'is F' is true of some object in the domain of the quantifier. Substitutional quantifications have quite different truth-conditions from objectual ones. For instance, 'Something is a mythological animal' is true if understood substitutionally, since the substitution instance 'Pegasus is a mythological animal' is true. But understood objectually, the sentence is not true, since there are no mythological creatures to make up a domain for the quantifier.

Since substitutional quantifiers do not need domains over which they range, it is easy to introduce substitutional quantifiers which bind predicate or sentential variables, even variables within quotation marks. One reason for interest in substitutional quantification is the hope that it may provide a way to understand discourse which appears to be about numbers, properties, propositions and other 'troublesome' sorts of entities as being free of exceptional ontological commitments. Whether natural language quantification is sometimes plausibly construed as substitutional is not, however, clear.

1 Objectual and substitutional quantification

The existential quantification

(1) $\exists x\ x$ is a logician

is true if '*x* is a logician' is true in at least one case; the universal quantification

(2) $\forall x\ x$ is a logician

is true if '*x* is a logician' is true in every case. Two accounts may be given of what it is for '*x* is a logician' to be true for some or all cases. These yield two kinds of quantification: 'objectual' and 'substitutional'.

In an objectual account, the quantifier has a domain of objects; sentences such as '*x* is a logician' are related to the contents of the quantifier's domain by the 'true of' relation. The sentence '*x* is a logician' is true of a member of the domain if and only if that member is a logician; '*x* taught *y*' is true of a pair $\langle u, v\rangle$ drawn from the domain just in case *u* taught *v*, and so on. Understood objectually, (1) is true if '*x* is a logician' is true of at least one thing in the domain; (2) is true if '*x* is a logician' is true of everything in the domain.

An objectual quantifier is tied directly to the domain of objects over which it quantifies. A substitutional quantifier does not have a domain; it is not a device whose semantics is explained in terms of the 'true of' relation. Instead, a substitutional quantifier has a 'substitution class', a collection of expressions which may be substituted for the variables the substitutional quantifier binds. The 'substitution instances' of a quantified sentence such as (1) or (2) are the results of removing the initial quantifier and replacing what is bound with some member of the substitution class. For instance, if the substitution class consists of the names 'Frege' and 'Russell' (the names, not the people), the substitution instances of (1) and (2) are the sentences

Frege is a logician.
Russell is a logician.

In substitutional quantification the cases to be considered are substitution instances. Read substitutionally, (1) is true provided at least one of its substitution instances is true; (2) is true provided they all are.

In order to avoid confusion, it is best to use distinct notations for the two kinds of quantifiers. Hereafter '$\exists$' and '$\forall$' are the objectual quantifiers, '*x*' and '*y*' their variables; 'Σ' is the substitutional existential, 'Π' the substitutional universal quantifier, and '*X*' and '*S*' their variables.

How different are $\exists x\phi x$ and $\Sigma X\phi X$? Suppose the domain of '$\exists$' contains things not named by anything in the substitution class of 'Σ'. For example, say the domain of '$\exists$' is the set of all animals, and let the substitution class for 'Σ' be the proper names of animals. Not every horse has a name (proper or otherwise), so '$\exists x\ x$ is a horse and *x* has no name' is true. But '$\Sigma X\ X$ is a horse and *X* has no name' is false. If it were true, then

α is a horse and α has no name,

where α is a proper name of a horse, would be true. But if α names something, then ‘ α has no name’ is clearly false.

Another difference arises because there can (apparently) be true sentences containing empty (non-referring) names, such as ‘Pegasus is only a mythological animal’. If the animal name ‘Pegasus’ does not name anything, *a fortiori* it does not name anything in the domain of an objectual quantifier. So, if the domain of ‘ $\exists$’ is all animals, the sentence ‘ $\exists x$ x is only a mythological animal’ is false: there is not any animal which is only mythological. But the sentence ‘ ΣX X is only a mythological animal’ is true, as one of its substitution instances is true.

The quantifiers of (1) and (2) are ‘nominal’ quantifiers - their variables occupy only the positions of names or singular terms. To further appreciate the difference between substitutional and objectual quantifiers, let us consider how we might introduce *non*-nominal quantifiers into a language. To introduce a substitutional quantifier whose variables can occupy, say, the positions of predicates or whole sentences, all one needs to do is to specify a substitution class, add new variables and define ‘substitution instance’. For example, if we take expressions such as ‘is Greek’ to comprise the substitution class, we can introduce a substitutional quantifier which, like that of

(1′) ΣX (Socrates *X* and Plato *X*),

binds predicate variables. (Examples of substitution instances thereof would be ‘Socrates is wise and Plato is wise’ and ‘Socrates snores and Plato snores’.) If sentences make up the substitution class, we can write, for example,

(2′) ΠS (Dick said that $S \rightarrow$ it is false that *S*),

with a quantifier binding sentential variables. The truth-conditions of these sentences are exactly as before: (1′) is true provided at least one substitution instance is true; (2′) is true provided all are. Provided all the instances already had truth-values, such substitutional quantifications automatically have truth-values.

To introduce objectual quantifiers whose variables can occupy the positions of predicates or sentences requires one to assign domains to such quantifiers and to explain what it is for something such as ‘Socrates *X*’ or ‘Dick said that *X*’ to be true of something in the domain of the quantifier. One can do this only if there are objects to make up the domain. Since appropriate domains for predicate and sentential variables seem to be, respectively, sets (or, perhaps, universals) and propositions, someone with doubts about their existence ought to have doubts about the objectual analogues of (1′) and (2′) (see Nominalism §4). No such worries attach to (1′) and (2′) themselves.

Some have hoped that substitutional quantification would provide an innocuous means of interpreting non-nominal natural language quantifications, such as ‘Bill did all that John did’ and ‘There’s something Socrates is that Plato is too’. Some hope it will interpret apparently nominal quantifications (‘Whatever Dick said is false’) and numerical quantification, whose domains seem to some ontologically suspect.

Another reason for interest in substitutional quantification is worries about the coherence of using an objectual quantifier to ‘quantify into’ positions like that occupied by ‘Jane’ in

It is possible that Jane is happy.
Nancy thinks that Jane is happy.

The position of ‘Jane’ in these sentences is ‘opaque’ (it is a position for which substitution of terms which pick out the same thing can lead from truth to falsehood; see Propositional attitude statements §2; Modal logic, philosophical issues in §3). Quine has argued repeatedly that an objectual quantifier cannot bind a variable in opaque position (see Quine 1951). Now, many perfectly good English sentences, for example,

(3) Some unhappy woman could have been happy.
There’s someone [such that] Nancy thinks [that she] is happy.

seem to involve quantification into opaque position. Since there is no problem with a substitutional quantifier binding a variable in opaque position, one might conclude that sentences such as in (3) involve substitutional quantification.

In the author's opinion, it is implausible that English quantification into opaque positions as in (3) can or should be interpreted substitutionally. For one thing, Quine's arguments that an objectual quantifier cannot bind a variable in opaque position are unsound (see Kaplan 1986). Furthermore, the way we mix quantifiers and expressions such as 'believes' prohibits a substitutional reading of the quantifier. Consider

(4) Twain is Clemens; Sal believes that Twain is sad; Sal does not believe that Clemens is sad.

This does not imply

(5) Twain is something such that Sal believes that it is sad and Twain is something such that Sal does not believe that it is sad,

since (4) could be true, but (5) ascribes contradictory properties to Twain. Now (5) seems to have the form

(6) (Ex)($x =$ Twain & Sal believes that x is sad) &
(Ex)($x =$ Twain & Sal does not believe that x is sad).

If the existential quantifier 'E' in (6) is objectual, (4) does not imply (6), since no object in the domain could make both conjuncts true. But if it is substitutional, (4) does imply (6), and hence (5). But we have already said that (4) does not imply (5). So the quantifier is not substitutional. (Incidentally, this shows that when a substitutional quantifier binds variables in opaque positions, $\exists x \phi x$ and $\Sigma X \phi X$ may have different truth-values even if the substitution class for ' Σ' consists only of names of objects in the domain of ' $\exists$', and every such object is named by a member of the class.)

The above is not meant to suggest that one cannot substitutionally quantify into opaque positions, but only that English quantification into such positions is not substitutional. In fact, substitutional quantifiers are very useful when one wants to bind variables in positions, such as those inside quotation marks, where it is indeed impossible in English to bind a referential variable. A formula along the lines of

(7) ΠS: 'S' is true if and only if S

is perfectly intelligible, and may even, in certain cases, be looked upon as 'defining truth' (see Kripke 1976 and Grover 1992 for discussion).

2 Interpreting natural language quantifiers

Natural language quantifiers cannot in general be plausibly understood as substitutional. Is quantification in English (or other natural languages; the disjunct is dropped hereafter) ever plausibly taken to be substitutional? If so, what, if anything, might this say about the 'ontological commitments' of speakers?

Some philosophers (Gottlieb, for example) have suggested that English quantifiers are ambiguous, between substitutional and objectual ones. Others (Marcus 1993, for example) have suggested that substitutional quantifiers provide, in some sense, the primary interpretation of English idioms of quantification, the objectual interpretation being a limiting or special case.

Arguments for the ambiguity thesis often turn on intuitions about what (the use of) a sentence ontologically commits one to. Roughly speaking, a sentence *S* ontologically commits one to there being an *F* provided that, as a matter of logic, one cannot consistently hold that *S* is true but there is not an *F*. (More precisely, let *F* be a singular English noun and let *s* name the English sentence *S*. '*s* ontologically commits one to there being an *F*' is true if the sentence 'If *S*, then there is an *F*' is logically true, when 'there is' is understood objectually. To say '*s* commits one to *a*' is to say that '*a* exists' follows from *S*. See Ontological commitment.)

One particular argument for the ambiguity thesis runs as follows: 'Quine believes that there are no propositions' is true; and 'If Quine believes that there are no propositions, then there is a thing Quine believes' is a logical truth, and thus a truth. So 'There is a thing Quine believes' is true. But Quine is not ontologically committed to the existence of things believed (propositions) since he says that he does not think there are any. So in English 'there is' is not always an objectual quantifier. So, presumably, it is sometimes a substitutional one. Another argument observes that to explain why

(8) If there are exactly as many boys as girls, then if there are 5 boys, there are 5 girls

is logically true, one must interpret its antecedent as involving numerical quantification; but asserting its antecedent surely does not commit one to the existence of numbers (see Gottlieb 1980). The arguments depend on the neutrality thesis: a substitutional quantification ΣXFX does not commit one to there being an *F* if none of its instances do. Given the definition of ontological commitment, the neutrality thesis is trivially true, since any substitution instance of ΣXFX logically implies it.

Such arguments must be evaluated on a case-by-case basis. As for the first, one might simply note that existential generalization (the move from something whose syntactic form is '*a* is *F*' to 'Something is *F*') is not valid without restriction. The second argument seems inconclusive, at least given that what is at issue is the commitments of one who embraces the range of quantitative inferences commonly held valid in English, and accepts arithmetic practice as well as standard use of numerical talk for measurement, counting, estimating, and so on.

Even if the evidence for the ambiguity thesis is at best equivocal, the project of attempting to see how much of a stretch of a discourse (about, say, numbers, or properties, or propositions) can be interpreted substitutionally is often of interest. Even if we are most plausibly understood to be committed to existence of a sort, we might welcome the news that we need not be. In evaluating the claim that a stretch of discourse can be interpreted substitutionally, one needs to consider several questions.

First, how satisfactory is the account provided of the truth of substitution instances? If 'Some numbers are prime' is substitutional, the truth of '7 is prime' is not explained by saying that 'is prime' is true of 7. We then need an account of what makes '7 is prime' true but '8 is prime' false.

Second, how much apparently true discourse does a substitutional account render problematic or false? It is not altogether clear how much conventional mathematics can be accepted when its quantifiers are read substitutionally. (For one thing, substitutional quantification seems to require limits on expressibility akin to those in predicative logic.) It is not clear how a substitutional interpretation of 'Katya said something' can explain its truth as a report of Katya's assertive utterance of a sentence untranslatable into English. (One could say that some set theory or talk of Katya's assertions is false but useful fiction.)

Third, to what extent do the terms of a substitutional account undercut its motivation? For nominalists, substitutional quantification is of interest because it may offer a way to re-interpret a theory (without requiring significant change in the way it is used) so that it can be adopted by the ontologically scrupulous. Such nominalists need to be careful about what they themselves (objectually) quantify over, in giving a substitutional account of a theory. For example, the natural understanding of the substitution class for 'Π' for a language for arithmetic will identify it with an infinite set of expression 'types'; to give an adequate substitutional account of a language as strong as second-order arithmetic seems to require quantifying over sequences. But types and sequences are not objects acceptable to most nominalists (see Parsons 1971 and 1982).

Quine has claimed that substitutional quantification is irrelevant to questions of ontology, saying that 'the question of the ontological commitment of a theory does not properly arise' (1969) unless the theory is formulated in an objectual first-order language or translated into such. But surely this is not so. If quantifier-free expressions of English number theory do not in fact contain names of numbers, and English numerical quantification is substitutional, then English numerical discourse is free of commitment to numbers. In fact, if this is so, it is presumably untranslatable into objectual idiom, since such translation would involve wrongly ascribing such commitment.

3 Logic and semantics

There are many interesting issues in the logic and semantics of substitutional languages. The reader is directed to Kripke, Parsons and Grover for a sampling. Here we confine ourselves to mention of two fundamental facts about substitutional languages.

First, $\Sigma X\phi X$ is true provided one of its substitution instances is. Thus, if one is to define truth inductively for a language in which substitutional quantifiers occur, with the definition assigning a truth-value to every sentence, the substitution class for the quantifiers cannot contain expressions in which those very quantifiers are used.

Kripke suggests that it is best to think of substitutional quantifiers as being introduced by extending a given (interpreted) language to a new language. The extension is achieved by introducing new variables for the quantifiers, defining a substitution class *C* thereof (consisting of some expressions of the original language) and taking the atomic sentences of the new language to be all the sentences of the old language along with results of replacing members of *C* in the base language with variables. (Call the last, following Kripke, 'forms'. Strictly, forms are those results, of replacing members of *C* with variables, which always return a sentence of the base language when the new variables are replaced with members of *C*.) As Kripke (and, earlier, Marcus) observe, provided the quantifiers are introduced in this way, one can formulate 'truth-definitions' along the lines of (7) without paradox.

Second, in the case of first-order languages (without identity), choosing between objectual and substitutional quantifiers need not affect what we say about validity or satisfiability. (Again we rely on Kripke's work.) Consider the first-order substitutional and first-order objectual predicate calculi, which differ only in their quantifiers and variables. Think of an interpretation of the substitutional calculus (SC) as being given by specifying an interpreted language *L* and (non-empty) substitution class *C* therein, assigning members of *C* to SC's constants, forms (of appropriate number of argument places) of *L* to the predicate letters of SC. If validity for SC is truth in any such interpretation, and validity for the objectual predicate calculus is truth in any (objectual) interpretation, then a sentence of SC is valid if and only if its objectual analogue (the result of trading substitutional quantifiers and variables for objectual ones) is.

See also: Logical and mathematical terms, glossary of; Ontological commitment

MARK RICHARD

References and further reading

Gottlieb, D. (1980) *Ontological Economy: Substitutional Quantification and Mathematics*, Oxford: Oxford University Press.(Expounds and defends a substitutional interpretation of number theory. Includes a useful discussion of ontological commitment.)

Grover, D. (1992) *A Prosentential Theory of Truth*, Princeton, NJ: Princeton University Press.(Uses substitutional quantification in characterizing truth for various languages.)

Kaplan, D. (1986) 'Opacity', in P. Schlipp (ed.) *The Philosophy of W.V. Quine*, La Salle, IL: Open Court.(An important discussion of quantification and opacity.)

Kripke, S.A. (1976) 'Is There a Problem about Substitutional Quantification?', in G. Evans and J. McDowell (eds) *Truth and Meaning*, Oxford: Oxford University Press.(Gives what has become the paradigmatic elementary account of substitutional quantification.)

Marcus, R.B. (1993) *Modalities: Philosophical Essays*, Oxford: Oxford University Press.(Collected essays. Numbers 1, 5 and 8 defend the substitutional understanding of English quantification as well as its ontological importance.)

Parsons, C. (1971) 'A Plea for Substitutional Quantification', *Journal of Philosophy* 68: 231-7.(Argues that substitutional quantification is not ontologically neutral.)

Parsons, C. (1982) 'Substitutional Quantification and Mathematics', *British Journal for the Philosophy of Science* 33: 409-21.(A review of Gottlieb 1980.)

Quine, W.V. (1951) 'Reference and Modality', in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 1980; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971.(One source of Quine's objections to objectual quantification into opaque positions.)

Quine, W.V. (1969) 'Existence and Quantification', in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 1977.(Gives Quine's objections to the relevance of substitutional quantification to ontology.)

Quine, W.V. (1973) *The Roots of Reference*, La Salle, IL: Open Court.(Contains a subtle discussion of substitutional quantification, mathematical discourse and the conceptual priority of substitutional quantification.)

Richard, M. (1995) 'Propositional Quantification', in J. Copeland (ed.) *Logic and Reality: Essays in Pure and Applied Logic*, Oxford: Oxford University Press.(Includes a critical discussion of substitutional accounts of propositional quantification.)

Quantum logic

The topic of quantum logic was introduced by Birkhoff and von Neumann (1936), who described the formal properties of a certain algebraic system associated with quantum theory. To avoid begging questions, it is convenient to use the term 'logic' broadly enough to cover any algebraic system with formal characteristics similar to the standard sentential calculus. In that sense it is uncontroversial that there is a logic of experimental questions (for example, 'Is the particle in region R?' or 'Do the particles have opposite spins?') associated with any physical system. Having introduced this logic for quantum theory, we may ask how it differs from the standard sentential calculus, the logic for the experimental questions in classical mechanics. The most notable difference is that the distributive laws fail, being replaced by a weaker law known as orthomodularity.

All this can be discussed without deciding whether quantum logic is a genuine logic, in the sense of a system of deduction. Putnam argued that quantum logic was indeed a genuine logic, because taking it as such solved various problems, notably that of reconciling the wave-like character of a beam of, say, electrons, as it passes through two slits, with the thesis that the electrons in the beam go through one or other of the two slits. If Putnam's argument succeeds this would be a remarkable case of the empirical defeat of logical intuitions. Subsequent discussion, however, seems to have undermined his claim.

1 The logic of experimental questions

Idealizing physics somewhat, we can think of repeatable experiments as types of observation and the results as values of the corresponding observable. Given any set of possible results, *R*, for a type of observable, *O*, there is an experimental question, namely: 'Will the observation of *O* produce a value in the set *R*?'

If experiments are repeatable, then there must be some significance to the results of repeating them on systems in - idealizing again - the same state. For theories such as classical mechanics and quantum theory, therefore, we may take the statistics resulting from repetition to be manifestations of probabilities. In our idealized treatment, we then assign a precise numerical probability $u(a)$ to the 'yes' result for a given experimental question *a* when asked of a system in a state *u*. This connection between experimental questions and states imposes a logic on the experimental questions.

First we say that two experimental questions are 'equivalent' if they are assigned the same probability by every state. It is then convenient to talk as if equivalent experimental questions were identical. Next we define an entailment-like partial ordering, '$\leq$', on experimental questions: $a \leq b$ if, for all states *u*, $u(a) \leq u(b)$. The intuitive idea is that if *a* fails to 'entail' *b* then there is some state in which a 'yes' answer for *a* is more probable than a 'yes' for *b*. We define a maximum element 1 and a minimum element 0 corresponding to the trivial experimental questions which always give a 'yes' answer, or a 'no' answer, respectively. Moreover, there is a negation-like complement operator, where $a^{\perp}$, the complement of *a*, has a 'yes' answer precisely when *a* has a 'no' answer. Hence $(a^{\perp})^{\perp} = a$ for any experimental question *a*. Also for any experimental question *a* and any state *u*, $u(a) + u(a^{\perp}) = 1$. From this it follows that the logic is 'orthocomplemented', that is, $a \leq b$ implies $b^{\perp} \leq a^{\perp}$. We may then define the 'orthogonality' relation $a \perp b$ ('*a* is orthogonal to *b*') to hold if $a \leq b^{\perp}$, or, equivalently, $b \leq a^{\perp}$. In standard logic two propositions are 'inconsistent' just in case one entails the negation of the other. Orthogonality is the analogue of inconsistency.

The 'meet' $a \wedge b$, if it exists, is the greatest (with respect to '$\leq$') experimental question less than or equal to both *a* and *b*. It is analogous to the conjunction $p \,\&\, q$ of two propositions, which is the meet with respect to the ordering of entailment. Likewise the 'join' $a \vee b$, if it exists, is the least experimental question greater than or equal to both *a* and *b*. It is analogous to the disjunction, $p \text{ or } q$ of two propositions, which is the join with respect to the ordering of entailment. Because the logic is orthocomplemented, De Morgan's laws hold.

$$(a \wedge b)^{\perp} = a^{\perp} \vee b^{\perp}$$
$$(a \vee b)^{\perp} = a^{\perp} \wedge b^{\perp}$$

Thus far we have been exploring a priori the consequences of characterizing physics as having to do with idealized repeatable experiments on systems. The logic of experimental questions may be further developed by means of a

constraint which, although inspired by quantum theory and not merely the result of conceptual analysis, seems antecedently plausible. This is ‘postulate M’: given any pairwise orthogonal set of experimental questions $\{a_i\}$ there is an experimental question *b* such that, for any state *u*, $u(b)+u(a_1)+u(a_2)+\ldots=1$ (see Hughes 1989: 198). This is plausible because it amounts to there being an experimental question *b* which has the ‘yes’ answer just in case all the a_i have the ‘no’ answer. It may be inferred from postulate M by the Mackey-Maczynski theorem that the logic of experimental questions is ‘orthomodular’ (see Hughes 1989: 198), a weakened form of ‘distributive’. That is, every pairwise orthogonal countable subset has a join and the orthomodular property holds: for any experimental questions *a* and *b* such that $a \leq b$, $b = a \vee (b \wedge a^{\perp})$.

We might expect the analogue of the material conditional, $(\text{not-}p)$ or q, to be $a^{\perp} \vee b$, but since the material conditional could also be characterized as $(\text{not-}p)$ or $(p$ and $q)$, another analogue is $a^{\perp} \vee (a \wedge b)$, which I abbreviate to $a \rightarrow b$. In fact it is the latter which is appropriate, because we require the analogue of *modus ponens*, namely $a \wedge (a \rightarrow b) \leq b$ (see Gibbins 1987: 139). The orthomodular property may then be stated succinctly: for any experimental questions *a* and *b* such that $a \leq b$, $b = a^{\perp} \rightarrow b$.

2 The nonclassical character of quantum logic

It is plausible, then, but not the result of pure conceptual analysis, that the logic of experimental questions forms an orthomodular partially ordered set.

For classical physics this logic is in addition a distributive lattice and so a Boolean algebra. (To say it is ‘distributive’ is to say that the familiar distributive laws hold.

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$
$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$$

To say it is a ‘lattice’ is to say that every pair of experimental questions has a join and a meet.) The logic of experimental questions for quantum theory is not, however, a Boolean algebra. Instead it is usually taken to be isomorphic to an orthocomplemented partially ordered set which occurs naturally in an infinite-dimensional Hilbert space.

Without loss of generality we may think of a vector v in this Hilbert space as an infinite sequence of (complex) numbers $\langle v_1, v_2, v_3, \ldots\rangle$. Given numbers *s* and *t*, we can form the linear combination $\mathbf{z} = s\mathbf{v} + t\mathbf{w}$ of vectors v and w, with $z_i = sv_i + tw_i$ for all *i*. More generally, given any set *e* of vectors we can consider those vectors which are finite or infinite linear combinations of members of *e*. This is the (closed) ‘span’ of *e*, which we shall write as ‘ $[e]$’. (That it is closed with respect to the appropriate topology follows from the inclusion of infinite linear combinations.) A (closed) linear subspace is then any non-empty set *a* of vectors such that $[a] = a$.

These vectors are analogues of coordinate triples in three-dimensional Euclidean space. In this example, the linear subspaces, which are all closed, comprise (1) the coordinates of the origin, O, (2) lines through O, (3) planes through O, and (4) the whole of space. (1) and (4) correspond, respectively, to the elements 0 and 1 in the lattice, which is ordered by inclusion. So, for instance, if *a* is the set of coordinates for a line through O and *b* for a plane through O, then $a \leq b$ just in case the line is in the plane. The meet of two members of the lattice of closed linear subspaces is just their intersection. The join is not, however, the union, for the union is not in general a linear subspace (for example, the union of two lines is not a plane). Instead, the join is the *span* of the union (for instance, the set of coordinates of the plane containing the two lines).

Now some pairs of vectors in Hilbert space are orthogonal. This enables us to define the ‘orthogonal complement’ $e^{\perp}$ of a set of vectors *e* as the set of all vectors orthogonal to all vectors in *e*. This provides an alternative characterization of the span as $[e] = (e^{\perp})^{\perp}$. Hence if *a* is a closed linear subspace, then $(a^{\perp})^{\perp} = a$.

The lattice of closed linear subspaces of a Hilbert space is not a Boolean algebra, for the distributive laws fail to hold. The reason for this is easily seen in the case of the coordinate triples. Consider three different sets of coordinate triples *a*, *b* and *c* corresponding to three lines OA, OB and OC all in the one plane. Then $b \vee c$ corresponds to the plane they lie in, so $a = a \wedge (b \vee c)$. But $a \wedge b$ is the intersection of *a* and *b* and is therefore $\{\langle 0,0,0\rangle\}$. Hence $a \wedge b = 0$. Likewise $a \wedge c = 0$. So $(a \wedge b) \vee (a \wedge c) = 0 \vee 0 = 0$. The orthomodular law, a weakening of the distributive laws, does hold.

The claim that quantum logic is a lattice could be justified either by the convenience of the standard representation using closed linear subspaces of a Hilbert space, or on the basis that we should depart as little as possible from the classical case. It is not, however, easy to find any other reasons (see Hughes 1989: 200).

The nonclassical character of quantum logic is also exhibited by its valuations. Consider the following weakening of the requirements for a valuation of the standard sentential calculus: $\mathrm{Val}(1) = 1$ and if $a \perp b$ then $\mathrm{Val}(a \vee b) = \mathrm{Val}(a) + \mathrm{Val}(b)$, from which it follows that $\mathrm{Val}(0) = 0$. Any (pure or mixed) quantum state may be thought of as such a valuation with $\mathrm{Val}(a)$ being the probability assigned by the state to the experimental question *a*. It can be shown that such a valuation must take all the values between 0 and 1. (In particular, there is no two-valued or three-valued valuation. That there is no two-valued valuation is Kochen and Specker's theorem.) Following Kochen and Specker, consider a spin 1 system in which the spin component in any direction takes the values -1, 0 and $+1$ (see Kochen and Specker 1967; Hughes 1989: 164-70). So for any direction *p* we have the experimental question o_p which has the 'yes' answer just in case the spin component in that direction is 0. For any three orthogonal *p*, *q* and *r* directions, the triad o_p, o_q and o_r are pairwise orthogonal experimental questions and their join is 1. Hence we would have $\mathrm{Val}(o_p) + \mathrm{Val}(o_q) + \mathrm{Val}(o_r) = 1$. Gleason's theorem then tells us that $\mathrm{Val}(o_p)$ is a quadratic function of the coordinates of *p*. Hence it is continuous and takes all values between the minimum, 0, and the maximum, 1.

3 An alternative approach to quantum logic

The presentation so far has been in the tradition of the paper by Birkhoff and von Neumann (1936) and the work of Mackey (1963). But there is an alternative approach to quantum logic due to Kochen and Specker (1965). The idea is to think of the logic as a family $\{B_i\}$ of overlapping Boolean algebras, each, intuitively, consisting of a family of experimental questions which can be answered without interfering with each other. We require the B_i to share the same maximum and minimum elements 0 and 1 (the two trivial questions) and to be such that the operations '$\wedge_i$', '$\vee_i$' and '$\perp_i$' agree on the overlaps. The family $\{B_i\}$ is called a 'partial Boolean algebra' if a 'coherence condition' holds. Call members of the logic 'neighbours' if there is some B_i to which they all belong. Coherence then says that if *a*, *b* and *c* are pairwise neighbours then they are neighbours. The set of closed linear subspaces of a Hilbert space is a partial Boolean algebra in which the Boolean subalgebras consist of sets of subspaces which make angles of either $0°$ or $90°$ with each other. For the case of the lattice of coordinate triples, each Boolean subalgebra contains eight members and is specified by a choice of three perpendicular lines OP, OQ and OR through O. The subalgebra then consists of the coordinate sets for (1) the origin, O, (2) the lines OP, OQ and OR (3) the planes OQR, ORP and OPQ, and (4) the whole of three-dimensional space (see Hughes 1989: 192-4).

The natural way of defining a valuation for a partial Boolean algebra would be one which satisfied the usual truth-functional constraints on each Boolean subalgebra. But such a valuation would have to satisfy the constraint that $\mathrm{Val}(1) = 1$ and that if $a \perp b$ then $\mathrm{Val}(a \vee b) = \mathrm{Val}(a) + \mathrm{Val}(b)$. As we have seen, such valuations must take all values between 0 and 1.

4 The logical reformulation of quantum theory

The fact that one important difference between quantum and classical physics can be expressed as a purely algebraic difference in the corresponding logics has inspired the programme of formulating quantum theory in terms of its logic of experimental questions. There is no obstacle to this provided the logic is assumed to be that of the closed linear subspaces of a Hilbert space. Given that assumption, the first step is to classify the states, which are identified with the probability measures on the logic. Assuming postulate M (see §1 above) this amounts merely to taking the states to be valuations which assign values in the range 0 to 1. Gleason's theorem tells us that, for any Hilbert space of dimension greater than two, the states thus characterized are all the mixtures (in positive proportions adding up to 100 per cent) of the standard pure states of quantum theory, which correspond to vectors in the Hilbert space. (The state represented by vector v assigns to a closed linear subspace *a* the square of the cosine of the angle between v and *a*.)

The next step is to consider the dynamics by which states change over time. The results of Mackey and Kadison show that we may represent the vectors in the Hilbert space as suitable functions on ordinary (that is, three-dimensional Euclidean) space in such a way that the dynamics is given by the familiar Schrödinger equation (see Beltrametti and Cassinelli 1981: 252-4). The difficulty with the programme, however, is in providing simple

algebraic constraints which ensure that the logic of experimental questions is indeed represented by the set of closed subspaces of a Hilbert space (1981: ch. 21).

5 The status of quantum logic

What is the status of quantum logic? There seem to be three sorts of answer: the neoclassical, the quantum ontologists' and the quantum logicians'. The neoclassical answer is that quantum logic is more a mathematical curiosity than anything of great significance in understanding the physical world. Its non-Boolean character is the result of constraints on the questions it is physically possible for us human beings to answer, so that we might think of God as knowing the answers to two questions whose meet is the null question but which are not classically inconsistent (for example, the questions: 'Is the particle in bounded region *R*?' and 'Does it have momentum in bounded set *S*?'). Hence the meet, ' $\wedge$', would not be a genuine conjunction even though it has analogous properties. For that reason the partial Boolean algebra formulation might be judged more perspicuous because the ' $\wedge_i$' of a given Boolean subalgebra could be treated as a classical conjunction (see Garden 1984; Forrest 1988; Garola 1992).

The quantum ontologists' answer is that quantum logic tells us something important about the structure of the physical world, but that 'logic' is a misnomer, for it should not be interpreted as a theory of deductive reasoning. On this view the best way to describe the physical world is by specifying the 'logic' (better 'algebra') of experimental questions and specifying the states as probability measures on this logic (see §4). This is arguably implicit in the Copenhagen interpretation, and it would seem to be the position of van Fraassen (1991: ch. 5) and, in a more radical version, of Hughes (1989: ch. 10).

The quantum logicians' answer is that quantum logic holds because it is in fact the correct deductive logic for all thought, and we made a mistake in assuming classical logic, which we should abandon, not merely when considering quantum theory, but quite generally. This was Putnam's position and it has also been advocated by Stairs (see Putnam 1969; Stairs 1982).

How might we decide between these answers? While logic might well be empirically defeasible it does not seem to be initially derived from experience. Hence there is a burden of argument on the quantum logicians. Putnam (1969) took up this burden by arguing that abandoning the distributive law enables us to resolve various problems without abandoning a realist interpretation of quantum theory. If so, then indeed a case would have been made for the empirical defeat of classical logic, comparable to the empirical defeat of Euclidean geometry. Unfortunately, Putnam's claims are open to serious criticism. For example, Putnam noted that in the two-slit thought-experiment we might have expected the probability of a particle hitting the screen in a given region when both slits are open to be the mean of the two probabilities with one slit closed. But that result violates quantum theory. It seemed that we could, by rejecting distributivity, retain an interpretation in which particles really do pass through one slit or the other, without allowing action at a distance. The objection to this, made forcibly by Gibbins, is that even given quantum logic, realism about particles which pass through one rather than another slit, combined with the rejection of action at a distance, still gives us a result contrary to quantum theory, namely that the probability of a particle hitting the screen in a certain region should be at least as great as the mean of the probabilities with one slit closed. But in fact, for some regions that inequality does not hold (see Gibbins 1987: ch. 10).

If we reject the quantum logicians' answer then that of the quantum ontologists would be plausible enough, especially given the discovery of simple constraints on the logic which ensured it could be represented by the closed subspaces of a Hilbert space. Presumably, however, the choice between the neoclassical and the quantum ontologists' answers will depend on the overall advantages and disadvantages of their interpretations of quantum theory (see Quantum mechanics, interpretation of).

See also: Logical and mathematical terms, glossary of

PETER FORREST

References and further reading

Bell, J.L. (1986) 'A New Approach to Quantum Logic', *British Journal for Philosophy of Science*, 37: 83-99.
(Quantum logic is interpreted as a system of 'manifestations' of attributes in regions of a 'proximity space'.)
Beltrametti, E. and Cassinelli, G. (1981) *The Logic of Quantum Mechanics*, Reading, MA: Addison-Wesley.(A

comprehensive guide to the mathematical foundations, especially useful for a discussion of the 'coordinatization problem', that is, the problem of representing quantum logic using Hilbert space.)

Beltrametti, E. and Fraassen, B.C. van (eds) (1981) *Current Issues in Quantum Logic*, New York: Plenum.(A fairly up-to-date, though technical, survey, with useful introductions to the topic by Mittelstadt and van Fraassen.)

Birkhoff, G. and Neumann, J. von (1936) 'The Logic of Quantum Mechanics', *Annals of Mathematics* 37: 823-43.(Of historical interest.)

Dummett, M.E. (1976) 'Is Logic Empirical?', in H.D. Lewis (ed.) *Contemporary British Philosophy*, London: Allen & Unwin, 45-68.(A critique of Putnam's claim that logic is empirical.)

Forrest, P. (1988) *Quantum Metaphysics*, Oxford: Blackwell.(Defence of the neoclassical position on quantum logic.)

Fraassen, B.C. van (1991) *Quantum Mechanics: An Empiricist View*, Oxford: Clarendon Press. (Careful and not too technical introduction to, among other things, quantum logic, with a case against the quantum logicians' position.)

Garden, R.W. (1984) *Modern Logic and Quantum Mechanics*, Bristol: Hilger.(Argues that the peculiarities of quantum theory are features of our own description rather than of the microscopic world.)

Garola, C. (1992) 'Truth Versus Testability in Quantum Logic', *Erkenntnis* 37: 197-221.(Shows how quantum logic can be obtained as the testable fragment of the standard sentential calculus).

Gibbins, P. (1987) *Particles and Paradoxes: The Limits of Quantum Logic*, Cambridge: Cambridge University Press.(Lucid introduction to quantum logic, with an incisive critique of Putnam.)

Hallam, N. (1987) 'Logic and Indeterminacy', *Philosophical Papers* 16: 53-8.(The most recent paper in a debate between Harrison and Hallam, in which Harrison argues against the quantum logicians' position. See also their papers in *Analysis* of 1983-5.)

Holdsworth, D.G. and Hooker, C.A. (1983) 'A Critical Survey of Quantum Logic', in M.L. dalla Chiara *et al.*, *Logic in the 20th Century: A Series of Papers on the Present State and Tendencies of Studies*, Aalen: Scientia, 127-246.(Useful survey of quantum logic with critical comments.)

Hooker, C. (ed.) (1975) *The Logico-Algebraic Approach to Quantum Mechanics*, Dordrecht: Reidel. (Anthology of papers, some hard to obtain in the original.)

Hughes, R. (1989) *The Structure and Interpretation of Quantum Mechanics*, Cambridge, MA: Harvard University Press.(An excellent introduction to quantum theory and its interpretation, including a not-too-technical discussion of quantum logic generally and the partial Boolean algebra approach in particular.)

Kochen, S. and Specker, E. (1965) 'Logical Structures Arising in Quantum Theory', in J. Addison, L. Henkin and A. Tarski (eds) *The Theory of Models*, Amsterdam: North Holland; repr. in C. Hooker (ed.) *The Logico-Algebraic Approach to Quantum Mechanics*, Dordrecht: Reidel, 1975, 263-76.(The paper introducing the partial Boolean algebra approach to quantum logic.)

Kochen, S. and Specker, E. (1967) 'The Problem of Hidden Variables in Quantum Mechanics', *Journal of Mathematics and Mechanics* 17: 59-87; repr. in C. Hooker (ed.) *The Logico-Algebraic Approach to Quantum Mechanics*, Dordrecht: Reidel, 1975, 293-328.(The original statement and proof of Kochen and Specker's theorem.)

Ludwig, G. (1983) *Foundations of Quantum Mechanics*, vol. 1, Berlin: Springer.(An operational approach to quantum theory.)

Mackey, G. (1963) *The Mathematical Foundations of Quantum Mechanics*, New York: Benjamin.(A historically important, but technical, treatment.)

Mittelstadt, P. (1986) 'Empiricism and Apriorism in the Foundations of Quantum Logic', *Synthèse* 67: 497-525. (A discussion of the extent to which quantum logic has an empirical justification.)

Putnam, H. (1969) 'Is Logic Empirical?', in R. Cohen and M. Wartofsky (eds) *Boston Studies in the Philosophy of Science*, vol. 5, Dordrecht: Reidel; repr. in C. Hooker (ed.) *The Logico-Algebraic Approach to Quantum Mechanics*, Dordrecht: Reidel, 1975, 181-206.(Best-known philosophical case for the quantum logicians' position.)

Putnam, H. (1981) 'Quantum Mechanics and the Observer', *Erkenntnis* 16: 193-219.(This paper reflects some changes in Putnam's position.)

Redhead, M. (1987) *Incompleteness, Nonlocality, and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics*, Oxford: Clarendon Press.(Careful and not-too-technical introduction to, among other things, quantum logic, with a discussion of Putnam's argument.)

Stairs, A. (1982) 'Quantum Logic and the Luders' Rule', *Philosophy of Science* 49: 422-36.(Argument for the quantum logicians' position.)

Quantum measurement problem

In classical mechanics a measurement process can be represented, in principle, as an interaction between two systems, a measuring instrument M *and a measured system* S*, during which the classical states of* M *and* S *evolve dynamically, according to the equations of motion of the theory, in such a way that the 'pointer' or indicator quantity of* M *becomes correlated with the measured quantity of* S. *If a similar representation is attempted in quantum mechanics, it can be shown that, for certain initial quantum states of* M *and* S*, the interaction will result in a quantum state for the combined system in which neither the pointer quantity of* M *nor the measured quantity of* S *has a determinate value. On the orthodox interpretation of the theory, propositions assigning ranges of values to these quantities are neither true nor false. Since we require that the pointer readings of* M *are determinate after a measurement, and presumably also the values of the correlated* S*-quantities measured by* M*, it appears that the orthodox interpretation cannot accommodate the dynamical representation of measurement processes. The problem of how to do so is the quantum measurement problem.*

1 States in classical and quantum mechanics

Classical mechanics describes model universes in which the state of the universe can be specified by an assignment of values to the positions and momenta of the systems in the universe, hence to all dynamical quantities, which are defined as functions of positions and momenta. The equations of motion yield rates of change for positions and momenta in terms of the Hamiltonian, a function of positions and momenta, and generate a possible history of the universe from the position and momentum values specified at some particular time. So we can understand a classical universe as consisting of separable classical systems interacting under the influence of forces encoded in the Hamiltonian. The evolution of the universe over time - a history of the universe - is given by a particular dynamical evolution of its state (see Mechanics, classical; Models).

The state of a model classical universe plays two distinct roles: a *diachronic* role as the carrier of the dynamics over time, and a *synchronic* role as the specification of a 'possible world' at a particular time; that is, the selection of a set of determinate values for the dynamical quantities as one of the possible sets of values for these quantities, equivalently the selection of a set of properties for the systems in the universe as one of the possible sets of properties.

The dynamical quantities, represented by real-valued functions on the classical state space or 'phase space' coordinated by the positions and momenta (not to be confused with real space), form a commutative algebra with respect to addition and multiplication. Properties in classical mechanics are represented by subsets of phase space (or, equivalently, by functions that take the values 1 on the subset and 0 elsewhere). The property that the value of a dynamical variable A lies in a certain range R is represented by the subset of classical states (phase space points) for which this property holds, via the functional relationship between A and positions and momenta. The properties form a Boolean algebra, isomorphic to a set of subsets of phase space. Classical states correspond to singleton subsets, minimal non-zero elements, or atoms in this algebra. The collection of properties represented by subsets of phase space containing the state is selected as the possible world associated with the state in the synchronic sense.

Quantum mechanics is derived as a generalization of classical mechanics in which certain 'canonically conjugate' quantities (like position and momentum) fail to commute with respect to multiplication. We obtain a non-commutative algebra of dynamical quantities that is representable as an algebra of operators on a Hilbert space, a linear vector space over the complex numbers. The dynamical quantities correspond to operators that can be decomposed into linear sums or integrals of projection operators, with real coefficients (the 'eigenvalues' of the operators) that represent the possible values of the quantities. Each projection operator has eigenvalues 0 and 1, and maps the Hilbert space onto a subspace that is the range of the operator. Properties are represented by subspaces or the corresponding projection operators and form a non-Boolean algebra, isomorphic to the algebra of projection operators or subspaces of Hilbert space. The one-dimensional subspaces or rays are atoms in this algebra and represent states in the diachronic sense; that is, they represent dynamical states that evolve over time according to the linear equation of motion of the theory (see Quantum logic §2).

On the *orthodox* interpretation of quantum mechanics, the rays (or unit vectors along the rays) are also taken to represent states in the synchronic sense. The collection of properties determined by a ray in Hilbert space as

obtaining in a model quantum mechanical universe is taken as the collection of subspaces containing the ray, just as the collection of properties belonging to a classical state in the synchronic sense is the collection of properties represented by subsets of phase space containing the state as a point in phase space. Properties represented by subspaces orthogonal to the ray are taken as not obtaining in the model universe, while properties represented by subspaces that are neither orthogonal to the ray nor contain the ray are regarded as indeterminate.

2 The source of the problem

The orthodox decision to take the ray representing a quantum state as defining a state in the synchronic sense has the consequence that we can no longer apply quantum mechanics to model quantum mechanical universes in the same way that classical mechanics is applied to model classical universes. In addition to selecting properties that obtain, properties that do not obtain, and properties that are indeterminate, the ray representing the quantum state is interpreted as assigning *probabilities* to *all* the properties represented by subspaces of the Hilbert space: probability 1 to properties represented by subspaces that contain the ray; probability 0 to properties represented by subspaces orthogonal to the ray; and non-zero probabilities to all other properties. These probabilities are not representable in the usual way as measures over different possible worlds, one of which is actual, defined by the orthodox interpretation of quantum states in the synchronic sense. So a non-zero probability assigned to the property that the value of a dynamical variable *A* lies in a certain range *R* is said to be ‘the probability of finding the value of the dynamical quantity in the range *R* on measurement’. This means that the application of quantum mechanics is restricted to providing probabilities for the results of measurements on model quantum mechanical universes by some agent or device *external* to these universes. To mark this distinction between classical and quantum mechanics, dynamical quantities in quantum mechanics are referred to as ‘observables’, where an observable is understood to have no determinate value unless the ray representing the quantum state lies in one of the subspaces associated with the projection operators of the observable, a circumstance that is presumed to occur when the observable is measured.

This notion of ‘measurement’ is undefined dynamically, as is the notion of an observable having no determinate value at one time and coming to have a determinate value at some other time as the outcome of a ‘measurement’. As Bell put it (1987), one would like to have an interpretation of quantum mechanics, our most fundamental theory of motion, in terms of ‘beables’ instead of ‘observables’. What we want is an *internal* account of measurement in quantum mechanics, and this appears to be excluded by the orthodox interpretation of quantum states and the linear dynamics of the theory.

To see what goes wrong with an internal account of measurement on the orthodox interpretation, consider a model quantum mechanical universe consisting of a measuring instrument *M* and a measured system *S*. Suppose that the initial quantum state of $S+M$ is such that, according to the orthodox interpretation, an observable *A* of *S* has some determinate value and a ‘pointer’ or indicator observable *R* of *M* has a determinate zero value. Then there exists a dynamical evolution that results in a final quantum state for $S+M$ in which both *A* and *R* are determinate and *R* has some non-zero value that depends on the initial value of *A*. In this case, the dynamical evolution can be taken as representing a measurement of *A* via the pointer *R* in this model universe. But now it follows from the linearity of the dynamics of the theory that if the initial quantum state of $S+M$ is a state in which *S* has no determinate *A*-value, and *R* has the determinate zero value, then the final state of $S+M$ must be a state in which both *A* and *R* are indeterminate. So the dynamical evolution cannot be taken as representing a measurement in this case.

Schrödinger (1935) pointed out that if *M* represents a cat and *R* takes two possible values, associated with the cat being alive and dead, and the cat interacts with a microsystem *S*, such as an atom that can either decay or not decay in a certain time (where these events are associated with the two possible values of *A*), the decay event triggering a device that kills the cat, then the cat will be neither alive nor dead after the measurement interaction, according to the orthodox interpretation.

3 Proposed solutions

States represented by rays in Hilbert space are referred to as ‘pure’ states in quantum mechanics. The theory also introduces ‘mixed’ states or ‘mixtures’, which can be decomposed into probability distributions over pure states and interpreted epistemically (subject to certain qualifications introduced by the nonuniqueness of this

decomposition). In terms of this distinction, the measurement problem arises because the final state of the instrument *M* and, measured system *S*, after a dynamical interaction representing a measurement, is a pure state associated with a ray that does not lie in any of the subspaces representing pointer properties of *M* or measured properties of *S*. What we appear to require after a measurement interaction is a mixed state for the combined system $S + M$, in which pure states that are determinate for different possible measurement outcomes occur with the appropriate probabilities.

There is a series of impossibility theorems by Wigner (1963), Fine (1970), Shimony (1974), and others that show, under very general constraints on what counts as a measurement interaction, that no evolution of the combined system $S + M$ governed by the linear dynamics of the theory can result in a final state for $S + M$ that is a mixture over pure states in which the pointer properties of *M* are determinate, on the orthodox interpretation of pure states.

Standard formulations of quantum mechanics invoke two modes of dynamical evolution for the pure state of a quantum system: the usual deterministic linear evolution, described by the equation of motion of the theory, that occurs when the system does not undergo measurement; and a second stochastic evolution that occurs when the system is measured and transforms the pure state into the required mixture. This measurement evolution is characterized by a 'projection postulate' as the projection or 'collapse' of the pure state onto one of the pure states in the mixture, with the appropriate probability.

Von Neumann's justification for this duality was that a measurement divides the world into two parts: the observed system and the observer (1932). We can follow all physical processes in the observed system with arbitrary precision, in principle, but not in the observer. The linear dynamics describes systems in the observed part of the world, so long as they do not interact with the observing part. When such an interaction occurs, the projection postulate applies. Consider the measurement of *S* by an instrument *M* and observer M'. The boundary between the observed system and the observer can be drawn between *S* and $M + M'$ or between $S + M$ and M'. Von Neumann showed that the application of the projection postulate directly to the system *S* is consistent with its application to the system $S + M$, after a suitable dynamical interaction between *S* and *M* governed by the linear equation of motion of the theory.

Von Neumann's argument for two distinct dynamical processes in quantum mechanics is hardly convincing, if only because classical mechanics requires no such duality. Indeed, the measurement problem is often formulated as the problem of reconciling these two modes of evolution in some way. Proposed solutions to the problem are restricted by the impossibility theorems and either (i) accept the orthodox interpretation and modify the linear dynamics of the theory; or (ii) accept the linear dynamics and modify the orthodox interpretation; or (iii) accept both the orthodox interpretation and the dynamics, and attempt to show that the difference between the mixture required by the orthodox interpretation and the pure state obtained on the basis of the linear dynamics can be ignored for all practical purposes, if certain relevant features of real measurement interactions are taken into account.

An example of a type (i) or 'collapse' proposal is the theory of Ghirardi, Rimini and Weber (1986). This theory introduces a modified dynamics for the quantum state that results in the occasional spontaneous localization of individual microsystems - so infrequently that the time evolution of a single particle follows the usual linear dynamics over long time intervals with high probability. For macrosystems, which consist of enormously large numbers of particles, the modified dynamics results in the almost instantaneous transition of pure states to mixed states for which macroscopically distinct position properties are determinate. The underlying claim is that all measurements can be characterized as ultimately involving macroscopic changes in the position of something that can function as a macroscopic pointer.

Type (ii) or 'no collapse' proposals include Bohm's hidden variable theory (1952), Everett's (1957) 'relative state' or 'many worlds' interpretation, and 'modal' interpretations, which interpret the pure quantum state as a dynamical state specifying probabilities only, and reject the orthodox interpretation in favour of alternative specifications for what collections of properties are determinate that are intended to support a purely internal account of measurement (see van Fraassen 1991).

Finally, type (iii) proposals include 'decoherence' theories and various other attempts to exploit the large number

of degrees of freedom of the amplifying recording device, or the environment of the system, as the source of a dissipative process that effectively renders a final pure state, after a measurement process, empirically indistinguishable from the required mixture. Such proposals work only if we suppose that properties that are indeterminate in the pure state can somehow be regarded as determinate, because certain further measurements that would distinguish the pure state from the mixture are in practice impossible to perform. This can hardly be regarded as satisfactory for a fundamental theory of mechanics.

See also: Quantum mechanics, interpretation of

JEFFREY BUB

References and further reading

With the exception of the book by van Fraassen, which is addressed to a philosophical audience, all these items involve a certain amount of technicality, and in most cases a great deal.

Bell, J.S. (1987) *Speakable and Unspeakable in Quantum Mechanics*, Cambridge: Cambridge University Press.(A collection of influential papers on foundational problems of quantum mechanics. Bell's discussion of 'beables' is in chapter 19.)

Bohm, D. (1952) 'A Suggested Interpretation of the Quantum Theory in Terms of Hidden Variables, I and II', *Physical Review* 85: 166-79, 180-93; repr. in J.A. Wheeler and W.H. Zurek (eds) *Quantum Theory and Measurement*, Princeton, NJ: Princeton University Press, 1983, 369-96.(The second paper deals with measurement.)

Everett, H., III (1957) 'Relative State Formulation of Quantum Mechanics', *Reviews of Modern Physics* 29: 454-62; repr. in J.A. Wheeler and W.H. Zurek (eds) *Quantum Theory and Measurement*, Princeton, NJ: Princeton University Press, 1983, 315-23.

Fine, A. (1970) 'Insolubility of the Quantum Measurement Problem', *Physical Review D* 2: 2783-7.(Fine's original proof that the measurement problem is insoluble, under very general constraints on what counts as a measurement interaction.)

Fraassen, B. van (1991) *Quantum Mechanics: An Empiricist View*, Oxford: Clarendon Press.(A discussion of the conceptual problems of the theory and a development of the original version of the modal interpretation.)

Ghirardi, G.C., Rimini, A. and Weber, T. (1986) 'Unified Dynamics for Microscopic and Macroscopic Systems', *Physical Review D* 34: 470-91.(The original paper of the 'GRW theory' - a proposed solution to the measurement problem involving a modification of the linear dynamics of quantum mechanics.)

London, F. and Bauer, E. (1939) *La théorie de l'observation en mécanique quantique*, Paris: Hermann; trans. 'The Theory of Observation in Quantum Mechanics', in J.A. Wheeler and W.H. Zurek (eds) *Quantum Theory and Measurement*, Princeton, NJ: Princeton University Press, 1983, 217-59.(An early discussion of the measurement problem that is still an excellent introduction to the topic.)

Neumann, J. von (1932) *Mathematische Grundlagen der Quantenmechanik*, Berlin: Springer; trans. R.T. Beyer, *Mathematical Foundations of Quantum Mechanics*, Princeton, NJ: Princeton University Press, 1955.(The first mathematically rigorous formulation of quantum mechanics. Chapter VI deals with the measurement problem and is reprinted in J.A. Wheeler and W.H. Zurek (eds) *Quantum Theory and Measurement*, Princeton, NJ: Princeton University Press, 1983, 619-56.)

Schrödinger, E. (1935) 'Die Gegenwärtige Situation in der Quantenmechanik', *Naturwissenschaften* 23: 807-18, 823-8, 844-9; trans. J.D. Trimmer, in J.A. Wheeler and W.H. Zurek (eds) *Quantum Theory and Measurement*, Princeton, NJ: Princeton University Press, 1983, 152-67(Contains Schrödinger's careful discussion of the measurement problem and his 'cat paradox' referred to in §2.)

Shimony, A. (1974) 'Approximate Measurements in Quantum Mechanics', *Physical Review D* 9: 2,321-3. (Shimony's insolubility theorem for the measurement problem.)

Wheeler, J.A. and Zurek, W.H. (eds) (1983) *Quantum Theory and Measurement*, Princeton, NJ: Princeton University Press.(A reference work, containing key papers on quantum mechanics as it relates to measurement.)

Wigner, E.P. (1963) 'The Problem of Measurement', *American Journal of Physics* 31: 6-15; repr. in J.A. Wheeler and W.H. Zurek (eds) *Quantum Theory and Measurement*, Princeton, NJ: Princeton University Press, 1983, 324-41.

Quantum mechanics, interpretation of

Quantum mechanics developed in the early part of the twentieth century in response to the discovery that energy is quantized, that is, comes in discrete units. At the microscopic level this leads to odd phenomena: light displays particle-like characteristics and particles such as electrons produce wave-like interference patterns. At the level of ordinary objects such effects are usually not evident, but this generalization is subject to striking exceptions and puzzling ambiguities.

The fundamental quantum mechanical puzzle is 'superposition of states'. Quantum states can be added together in a manner that recalls the superposition of waves, but the effects of quantum superposition show up only probabilistically in the statistics of many measurements. The details suggest that the world is indefinite in odd ways; for example, that things may not always have well-defined positions or momenta or energies. However, if we accept this conclusion, we have difficulty making sense of such straightforward facts as that measurements have definite results.

Interpretations of quantum mechanics are, in one way or another, attempts to understand the superposition of quantum states. The range of interpretations stretches from the metaphysically daring to the seemingly innocuous. But, so far, no single interpretation has commanded anything like universal agreement.

1 Superposition of states

The *two-slit experiment* provides a vivid introduction to the concept of quantum superposition. Imagine a barrier with two slits, placed in front of a light source. On the other side of the barrier is a photographic plate. The intensity of the light can be made so low that photons (light quanta) hit the plate like particles, one at a time in definite though unpredictable locations. If we close the right slit, a narrow band of 'hits' will accumulate in front of the left slit. A similar result occurs if we close the left slit and keep the right one open. Now suppose both slits are left open. If photons were ordinary particles, we would predict two narrow bands of hits. What actually happens is *very* different: we get an array of bands, forming an interference pattern. If light were simply waves, this would not be mysterious: waves interfere. But individual photons are not waves. This leaves us wondering how the effect occurs and whether the individual photons even follow well-defined paths.

What does quantum theory say? Each of the one-slit experiments is associated with a particular *quantum state*. When we leave both slits open, the state becomes an odd amalgam - a *superposition* - of these. From the superposition we can derive a probability distribution for individual 'hits', a distribution that embodies the interference pattern. To say more than this, however, is to interpret quantum theory rather than simply to apply it.

Another example will make the notion of superposition more precise. Electrons have a directional property called *spin*. If we pick a direction and measure spin, we will always get one of two results: $+1$ or -1 ('up' or 'down'). Suppose we prepare a batch of electrons so that each would display spin up in measurement of spin in some specific direction x. It is usual to represent quantum states by abstract vectors in an abstract mathematical space called *Hilbert space* (see Quantum logic §2): in this instance, we will denote the quantum state of the electrons by the vector $|\,\mathrm{x}+\rangle$. Suppose now that instead of measuring x-spin we measure spin in a direction z at right-angles to x. We can use the rules of quantum theory to write $|\,\mathrm{x}+\rangle$ as a vector sum - a superposition - of z-spin states:

$$|\,\mathrm{x}+\rangle = \tfrac{1}{\sqrt{2}}\,|\,\mathrm{z}+\rangle + \tfrac{1}{\sqrt{2}}\,|\,\mathrm{z}-\rangle.$$

The numerical coefficients are used to calculate the probabilities of observing z-spin up or down in any given measurement. We get the probabilities by squaring these coefficients, in this case yielding a probability of 1/2 for each possible outcome.

Superpositions are not the same as mixtures. Suppose we prepare a batch of electrons of which half are in state $|\,\mathrm{z}+\rangle$ and half are in state $|\,\mathrm{z}-\rangle$. This is not a case of superposition. No matter what direction we choose for our spin measurement, the probabilities of spin up and spin down will be 1/2. This is very different to what we see in a superposition, for example, the state $|\,\mathrm{x}+\rangle$ discussed above. In this case, the probabilities depend on what direction we choose: for any direction in the x-z plane, the closer the direction to x the higher the probability of spin up will be, until the probability reaches certainty at x.

2 Superposition and the eigenvalue-eigenstate link

The example of spin illustrates some characteristic features of quantum mechanics. Quantum states are represented by vectors of length one in Hilbert space (an abstract mathematical shape). If we want to measure some quantity R (x-spin and z-spin being two examples of such quantities) on a system in state $\mid s\rangle$, our predictions typically involve probabilities. To find these, we write $\mid s\rangle$ as an appropriate sum of perpendicular vectors

$$\mid s\rangle = c_1 \mid r_1\rangle + c_2 \mid r_2\rangle + \ldots$$

For each possible measured value r_i of R, there is one vector $\mid r_i\rangle$. The measurable values r_i are called the *eigenvalues* of R, and the corresponding vectors r_i are the *eigenvectors* or *eigenstates* of R. If $\mid s\rangle$ happens to be one of the $\mid r_i\rangle$ - if c_i equals one and all the other c_i are zero - we predict the outcome r_i (the measurement of x-spin in the first system is an example of this: we predict that x-spin will be up, since $\mid s\rangle =\mid x\rangle$). If $\mid s\rangle$ is a superposition of the states $\mid r_i\rangle$, the probability of observing r_i in any given measurement is $\mid c_i \mid^2$. (As we saw when we measured z-spin in the first system for which $\mid s\rangle = (1/\sqrt{2}) \mid z+\rangle + (1/\sqrt{2}) \mid z-\rangle$). This rule is often referred to as the Born rule, after the physicist Max Born.

The conventional wisdom is that when $\mid s\rangle$ is a superposition of eigenstates of R, R has no definite value: so, for example, the electrons prepared with up x-spin have no definite z-spin. This principle is often called the *eigenvalue-eigenstate link*. Among other things, it would imply that the photons in the two-slit experiment do not go through either hole. To the uninitiated the principle might seem gratuitous, and in fact there is a classic argument that seems to undermine it. The modern formulation goes like this: we can produce a pair of electrons in a state called the *singlet state*:

$$\mid s\rangle = \tfrac{1}{\sqrt{2}} (\mid x_1+\rangle \otimes \mid x_2-\rangle - \mid x_1-\rangle \otimes \mid x_2+\rangle) \, .$$

Here $\mid x_1+\rangle \otimes \mid x_2-\rangle$ is a state in which electron #1 has x-spin up and electron #2 has x-spin down. $\mid x_1-\rangle \otimes \mid x_2+\rangle$ is read similarly. What this means is that, if we measure x-spin on electron #1, then an x-spin measurement on electron #2 *must* yield the opposite result no matter how well-separated #1 and #2 are. Furthermore, this holds for *all* directions: in the singlet state, parallel spin measurements *always* have opposite results. Therefore, we can predict the spin of particle #2 *in any direction we choose* by measuring particle #1. We cannot measure spin on electron #1 in two directions at once, and after the first measurement the correlation is broken: consequently, we cannot continue to accumulate information about electron #2 by repeated measurement on electron #1. But the correlation will hold for the first measurement of electron #1 that we make. Now, our measurement on electron #1 presumably does not *create* the spin of electron #2. Therefore, it apparently must have been there all along. The tempting conclusion is that electron #2 (and, by symmetry, electron #1) has a definite spin in each direction, even though *no* quantum state could tell us all those spins. If so, the eigenvalue-eigenstate link cannot hold.

This argument has its origins in the reasoning of Einstein, Podolsky and Rosen (1935), though it is not the version Einstein preferred (see Einstein, A. §5). Unfortunately, it is seriously flawed. Parallel spin measurements yield perfect anti-correlation. If the measurements are not parallel, the strength of the correlation varies with the angle. Suppose we assume that all spins are definite and that measurements on one electron do not affect the other. Can all the quantum predictions be captured consistently with the ordinary rules of probability? John Bell (1987) proved, roughly, that they cannot, and subsequent experiments have supported quantum theory's predictions. There are some loopholes here, but most physicists and philosophers agree that these results spell trouble for any wholesale rejection of the eigenvalue-eigenstate link (see Bell's theorem).

Unfortunately, retaining the eigenvalue-eigenstate link seems at least as bad. For one thing, we seem to accept that our measurement of electron #1 *does* create the spin of electron #2. Worse still, if the electrons are well separated their instantaneous 'communication' would seem to reguire faster-than-light signals, something which special relativity is usually held to forbid. There are more problems. Suppose we have an electron in the state $\mid x+\rangle$ and we decide to measure its z-spin. We would like the result to be a state in which the pointer indicates spin up or else a state in which the pointer indicates spin down. But Schrödinger's equation, the fundamental dynamical law of quantum mechanics, says that at the end of the measurement, we will get a superposition of such states, precisely because the state $\mid x+\rangle$ is not an eigenstate of the quantity (z-spin) which we are measuring. If we accept the

eigenvalue-eigenstate link, this implies, absurdly, that the measurement will fail to have a result in the oddest way possible: the pointer will not point anywhere at all (see Quantum measurement problem)!

3 Interpretations of quantum theory

Clearly quantum mechanics is puzzling. Interpretations of the theory usually focus on one or more specific puzzles, the measurement problem being the most notable. In what follows, we will look at a handful of interpretations and principles.

The most venerable is the Copenhagen interpretation, due to Niels Bohr (1935). Bohr maintains that physical concepts can only be applied in experimental contexts. This is not mere instrumentalism. Consider position and momentum. Both concepts are essential to quantum theory, but the experimental arrangements for measuring them are physically incompatible. Furthermore, the Heisenberg uncertainty relations guarantee that the more the state says about one, the less it says about the other. Bohr concludes that these concepts are *complementary*, and that each can be sensibly applied only within the appropriate experimental context.

On this view, the measurement problem is misconceived. It is a given that experiments have results. The function of quantum theory is not to tell a tale about how this comes to be, but to give an account of how the various possible experimental situations are related to one another.

A review of the literature reveals that the interpretation of the Copenhagen interpretation is no straightforward matter. Bohr may have been profound; few would claim he was clear. However, the stress on the need to ground physical concepts in experimental arrangements shows up in a different guise in a controversial principle called the projection postulate, first made explicit by von Neumann (1932). Schrödinger’s equation tells us that quantum states change smoothly and deterministically, but we have noted that this makes measurement puzzling. The projection postulate says that when we measure a quantity R, the state ’projects’ or ’collapses’ abruptly and unpredictably (but with probabitities determined by its coefficients, as above) on to one of R’s eigenstates, thereby leading to a definite measurement result.

One would like to know more. Does projection occur *only* during measurements? If so, what makes measurement special? Or does projection occur when a conscious observer enters the picture, as some interpreters claim? More recently, detailed physical theories of collapse have been offered. These amount to modifying quantum theory rather than merely interpreting it. The hope is that experiments will be able to settle the matter of whether collapses really occur.

Collapse theories deny that Schrödinger’s equation always holds. The most notorious approach that upholds Schrödinger’s equation is the many-worlds interpretation, which originated in Hugh Everett III’s relative state interpretation (1957) and was further developed by Bryce deWitt (1970). This account claims that when certain sorts of interactions occur, measurement being one of them, the universe divides into separate branches, one for each component of the superposition. This means that observers branch too, and that each observer on each branch sees one of the possible measurement results.

This approach strikes many people as metaphysically extravagant. If it fails, however, it will more likely be for technical reasons having to do with the rules governing branching (the so-called ’basis problem’) and the treatment of experimental probabilities.

Quantum logic is another scheme that seems to incorporate drastic measures. In fact, many different approaches, some quite harmless, fall under this rubric. Hilary Putnam made the most striking claims (1969). He maintained that quantum mechanics demands that we revise the laws of logic, notably the distributive law (’A and [B or C]’ is equivalent to ’[A and B] or [A and C]’). Putnam claimed that if we adopt quantum logic, all quantum paradoxes will vanish because it will be impossible to derive the apparent contradictions. He also maintained that, because statements like ’x-spin is up or down’ come out true on quantum logic, all quantities have values at all times, even though every attempt to list the values of all quantities would be inconsistent.

This is hard to understand, let alone believe, so much so that more or less no one believes it any more. However, a certain more general approach grew out of quantum logic, one that focuses on the mathematical structures underlying quantum mechanics. The hope is that by understanding the role of these structures in the theory, it will be possible to see quantum mechanics as providing explanations and not mere predictions (see Quantum logic;

Putnam, H. §4).

Finally, there are some important approaches that begin by rejecting the eigenvalue- eigenstate link, but do so in a carefully-limited way. One such is due to David Bohm (1952), but has its roots in deBroglie's 'pilot wave' interpretation. According to Bohm, particle positions are always definite. They evolve deterministically under the guidance of a 'quantum potential' operating in a space of $3n$ dimensions (*n* being the number of particles). This is the most well-worked out example of a hidden variable theory. Hidden variable theories have often been 'proved' impossible, but such proofs deal with theories that make *all* quantities definite or that deny faster-than-light causation. Bohm's theory accepts superluminal causation and takes position and functions thereof to be the *only* hidden variables. An oft-cited virtue of this approach is that it describes detailed processes underlying such cases as the two-slit experiment. The odd relationship to relativity is troublesome to many, but Bohm's theory continues to be an active area of research.

There are many other approaches to interpreting quantum mechanics and many issues that have not been raised here. For example, ensemble interpretations, which deny that the quantum state characterizes individual systems at all, are important. So-called 'modal' interpretations - a term introduced by van Fraassen - distinguish between what the state necessitates and what may actually be true in addition. Modal interpretations reject the eigenvalue-eigenstate link and have a certain kinship both with the many-worlds view (though without the extravagant metaphysics) and with hidden variables. The investigation of Bell's result led to much debate about whether there is any useful content to the idea that quantum mechanics incorporates some form of holism. And more recently, philosophers have turned their attention to foundational problems in quantum field theory (see Field theory, quantum).

The interpretation of quantum theory makes for an intriguing intersection between science and philosophy. The safest prediction is that the various controversies will not be settled soon.

ALLEN STAIRS

References and further reading

Ballentine, L.E. (1970) 'The Statistical Interpretation of Quantum Mechanics', *Reviews of Modern Physics* 42: 358-81.(Ballentine defends the view that quantum states do not describe individual quantum systems but refer instead to ensembles.)

Bell, J.S. (1987) *Speakable and Unspeakable in Quantum Mechanics*, Cambridge: Cambridge University Press.(A collection of papers by Bell on various issues in foundations of quantum mechanics. His celebrated theorem was first proved in the paper reprinted as chapter 2.)

Bohm, D. (1952) 'A Suggested Interpretation of the Quantum Theory in Terms of Hidden Variables, I and II', *Physical Review* 85: 166-79, 180-93.(These papers were the first to offer a detailed hidden variable interpretation of quantum theory and inaugurated a programme that is still vigorously pursued by some researchers.)

Bohr, N. (1935) 'Can Quantum-Mechanical Description of Reality Be Considered Complete?', *Physical Review* 48: 696-702.(Bohr's reply to Einstein, Podolsky and Rosen (1935), and as good a place as any to get a sense of Bohr's understanding of complementarity.)

Bub, J. (1974) *The Interpretation of Quantum Mechanics*, Dodrecht: Reidel.(An extended development of a quantum-logical approach to the foundations of quantum mechanics.)

Cohen, R.S. and Wartofsky, M.W. (eds) (1969) *Boston Studies in the Philosophy of Science*, vol. 5, Dodrecht: Reidel.(An important collection of papers on the foundations of quantum theory and relativity.)

Cushing, J.T. and McMullin, E. (eds) (1989) *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*, Notre Dame, IN: University of Notre Dame Press.(A watershed collection of essays on Bell's theorem, covering a broad range of interpretive issues. On the issue of holism and quantum mechanics, the papers by Fine, Howard, Hughes, Teller and van Fraassen are especially relevant.)

deWitt, B.S. (1970) 'Quantum Mechanics and Reality', *Physics Today* 23: 30-55.(A 'many worlds' version of the relative-state interpretation developed by Hugh Everett III.)

Einstein, A., Podolsky, B. and Rosen, N. (1935) 'Can Quantum Mechanical Description of Reality Be Considered Complete?', *Physical Review* 47: 777-80.(A touchstone for the foundations of quantum mechanics, though not the version of the argument that Einstein himself preferred.)

Everett, H., III (1957) 'Relative State Formulation of Quantum Mechanics', *Reviews of Modern Physics* 29: 454-62.(The paper from which the 'many-worlds' interpretation was developed. Everett's own view, as the title

indicates, relies on the notion of a ‘relative state’ - a concept that applies to coupled systems, of which the measuring device and the measured system form an example. It is arguable that the relative state interpretation avoids certain of the difficulties of the many-worlds interpretation, though it also lacks the relative intuitive clarity of the many-worlds view.)

Fine, A. (1986) *The Shaky Game: Einstein, Realism and the Quantum Theory*, Chicago, IL: University of Chicago Press.(A collection of essays. Chapter two makes a convincing case that the argument actually found in the EPR paper was not written by Einstein and was importantly different from the version that he favoured.)

Fraassen, B. van (1991) *Quantum Mechanics: An Empiricist View*, Oxford: Clarendon Press.(Includes a development of the modal interpretation, which van Fraassen first introduced in the 1970s.)

Ghirardi, G.C., Rimini, A. and Weber, T. (1986) ’Unified Dynamics for Microscopic and Macroscopic Systems’, *Physical Review* D34: 470-91.(A detailed physical proposal for augmenting Schrödinger’s equation with a process of collapse.)

Healey, R. (1989) *The Philosophy of Quantum Mechanics: An Interactive Interpretation*, Cambridge: Cambridge University Press.(An extended development of a modal interpretation of quantum theory. Healey refers to his view as an interactive approach.)

Hughes, R.I.G. (1989) *The Structure and Interpretation of Quantum Mechanics*, Cambridge, MA: Harvard University Press.(A pedagogically impressive introduction to the fundamentals of quantum mechanics in Hilbert space, as well as an interpretation of quantum mechanics in the tradition growing out of quantum logic.)

Jammer, M. (1974) *The Philosophy of Quantum Mechanics*, New York: John Wiley & Sons.(Still one of the most comprehensive introductions to the foundations of quantum mechanics.)

Neumann, J. von (1932) *Mathematische Grundlagen der Quantenmechanik*, Berlin: Springer; trans. R.T. Beyer, *Mathematical Foundations of Quantum Mechanics*, Princeton, NJ: Princeton University Press, 1955.(The first rigorous Hilbert space formulation of quantum mechanics.)

Putnam, H. (1969) ’Is Logic Empirical?’, in R.S. Cohen and M.W. Wartofsky (eds) *Boston Studies in the Philosophy of Science*, Dodrecht: Reidel, vol. 5, 181-206.(In this seminal paper, Putnam attempts to argue for a realist interpretation of quantum mechanics by an appeal to so-called ‘quantum logic’, arguing that if we give up the principle that ‘and’ distributes over ‘or’, we will be able to claim that every quantum-mechanical quantity has a definite value.)

Zurek, W.H. (1991) ‘Decoherence and the Transition from Quantum to Classical’, *Physics Today* October: 36-44. (The decoherence programme attempts to give a *quantum mechanical* explanation of why the world normally appears classical. The account begins with the fact that virtually all the systems we encounter are *open systems*, that is, are in interaction with the environment. It goes on to argue rigorously that the result is a so-called ‘reduced density matrix’ (crudely, a restriction of the total state to a subsystem) that is a stable mixture of appropriate states. In the case of a measuring device, this would be a mixture of appropriate indicator states. Zurek is one of the main proponents of this approach and his article, though technical, is relatively accessible. The ‘Letters’ portion of the April 1993 issue of the same journal includes a number of reactions to Zurek’s essay, with replies by Zurek.)

Questions

Some theorists hold that a question is an interrogative sentence; others that a question is what is meant or expressed by an interrogative sentence. Most theorists hold that each question has two or more answers, and that the point of asking a question is to have the respondent reply with one of the answers. Most hold that each question has an assertive core or presupposition that is implied by each of the answers; if it is false, then no answer is true, so we say that the question commits the fallacy of many questions and we regard the negation of the presupposition as a corrective reply to the question (it corrects the question).

For example, consider the question 'Has Adam stopped sinning?' Its answers are 'Adam has stopped sinning' and 'Adam has not stopped sinning'. It presupposes 'Adam has sinned'; thus 'Adam has not sinned' is a corrective reply. The 'safe' way to ask this question is via the conditional 'If Adam has sinned, then has Adam stopped sinning?'

We can construct formal systems for asking whether *and* which *questions in an effective way. Other types of question (for example,* who *and* why*) are still problematic. It can be proved that some questions are reducible to others, some questions raise others, and some systems for the logic of questions can never be complete in certain ways.*

1 Concepts and theories

Questions have been discussed since Aristotle, although formal, systematic study did not begin until the 1950s. Theorists disagree on basic concepts, and there are at least three approaches to theorizing about questions. In one, the essence of a question is its set of answers, in another the essence is the questioner's intentions, and in another the essence is an objective intensional entity.

The 'reduction-to-answers' view holds that every question has direct answers, these direct answers are statements, and to know what counts as a direct answer to the question is to understand the question. A question presupposes any statement that is implied by every direct answer. A complete answer is a statement that implies a direct answer; a partial answer is a statement that is implied by a direct answer. The reduction-to-answers view is exemplified in most of the systems noted in §2 below.

The 'reduction-to-*intentions*' view holds that to know the questioner's intentions is to understand the question (see §4 below). The 'reduction-to-*intensions*' view holds that questions are intensional entities that are not relative to language and might or might not be expressible by sentences in a language (see §5 below and Intensional entities).

2 Effective systems

For many purposes (for example, conversation with strangers, courtroom procedure, automated information retrieval systems), it is desirable to have systems for question and answer that are effective in the following sense: for any expression of the given language, we can effectively tell whether it is an interrogative sentence and, if it is one, tell what question it expresses. In the reduction-to-answers view this implies that, given an interrogative sentence, we can effectively tell what counts as a direct answer to the question that is expressed.

Beginning with G. Stahl and T. Kubiński in the 1950s, logicians have constructed systems that provide for *whether* and *which* questions and that satisfy the effectiveness condition. The system of Belnap and Steel (1976) is exemplary. In it, *whether* and *which* questions are expressed by interrogatives of the form *?RS*. Here the subject part *S* indicates a set of statements, which are the alternatives presented by the question. (For *whether* questions *S* is a finite list of statements; for *which* questions *S* is a formula indicating all the statements that result when names are substituted for the queriables (the free variables) in the formula.) Given *S*, the request part *R* indicates how many of the true alternatives are to be put into each direct answer and indicates whether the answer is to claim that it contains all of the true alternatives and (in the case of *which* questions) claim that the various names given in the answer denote distinct entities. The system provides for *who*, *when*, *where*, *why*, *what* questions - if these are construed as asking which person, which time, which place, which reason, which thing. It allows the questioner to call for a complete list or just some examples or an answer of the form at-least-*m*-but-at-most-*n*.

3 Some problems

Systems such as those noted above are good models for some question-asking situations but not for others. First consider ‘What is in the box?’ and ‘Some of John’s things are in the box’. If the latter is a complete answer, as some theorists say, then this *what* question cannot be construed as a *which things* question, and the system must be altered or extended to provide for it. Similar examples exist for *who* and other so-called ‘wh- questions’.

Why questions pose special problems. Consider ‘Why did this apple fall?’. We may construe this as ‘What is a good explanation of why this apple fell?’, but we will not know what counts as an answer until we know what counts as a good explanation. S. Bromberger (1992) has proposed that some types of *why* question can be adequately answered by citing (1) a general rule, (2) an abnormic law - that is, a law that specifies exceptions to the rule, and (3) some exception mentioned by the law. For example: ‘An apple remains on the tree unless it becomes ripe or a wind blows or the tree dies, and in this case the wind blew.’ It is reasonable to require that the abnormic law be empirical and true, but then (because we cannot always recognize truth) we cannot always recognize the admissible direct answers.

Effective systems have some intrinsic limitations. Suppose that the interrogatives of a system can be effectively recognized and hence listed by some algorithm and that, for each interrogative, its direct answers can be recognized and hence listed. Then it can be shown by Cantor’s diagonal method that the system is incomplete in the sense that there is a set of sentences that might be the answers to some question but cannot be the answers for any interrogative in the given list (Harrah 1984).

4 Epistemic analysis of questions

L. Åqvist (1965) proposed that questions are requests concerning epistemic states, expressible by imperatives such as ‘Let it be the case that I know that *A* or I know that *B*’ or ‘Let it be the case that, for some *x*, I know that *Px*’. In this approach there may be a variety of imperative operators, and reference made to different knowers, so that, for example, a teacher can put questions to a class in the form ‘Bring it about that I know that you know…’ J. Hintikka (1983) has developed this approach further to elaborate its pragmatic dimension. Hintikka has extended his game-theoretic semantics to provide a semantics for epistemically construed questions and rules for conducting various types of rational question-and-answer dialogue (see Semantics, game-theoretic).

In this approach every question has a presupposition (in the first example above: ‘*A* or *B*’), a desideratum (‘I know that *A* or I know that *B*’) and direct answers (*A*, *B*). If an answer meets certain conditions of relevance and conclusiveness, guaranteeing that the desideratum is satisfied, then it is a conclusive answer. The conditions vary with the type of question. In the case of a *which* question, quantified variables should range over entities with which the questioner is acquainted, and the answer should specify a noun phrase only if the questioner knows what that phrase denotes. There is no guarantee that the respondent can effectively tell from the imperative sentence what the range of the variables is supposed to be, or what noun phrases will suffice, so, in this approach, the concept of conclusive answer in general is not effective.

5 Intensional analysis of questions

P. Tichý (1978) proposed that a question is a function defined on possible worlds. Propositions, individual concepts and properties are such functions, and they are the common types of question. For any given world the value of such a function is a truth-value, an individual or a set of individuals, respectively. In general the right answer to a question is the value of the function in the actual world.

This idea can be developed on its own; it can also be combined fruitfully with the ideas of Richard Montague. In Montague’s analysis of declaratives, meanings are higher-order intensional entities. These are assigned to linguistic expressions according to Frege’s principle that the meaning of a syntactical compound is the corresponding semantic compound of the meanings of the parts of the syntactical compound. Some linguists and logicians have developed analyses of this sort that treat both declaratives and interrogatives. In most, an interrogative expresses a question and denotes its true direct answers (see Frege, G.).

6 Erotetic logic

In a general sense erotetic logic is the logic of utterance and reply (Harrah 1984). In the strict sense it is the logic

of question and answer.

For most systems, especially those noted in §2 above, an important concept is that of containment. One question contains (or covers, or obviates) another if and only if every direct answer to the first implies some direct answer to the second, and two questions are equivalent if and only if each contains the other.

For analysis of the raising of questions, the implying of questions and arguing to questions, Wiśniewski (1995) has defined some useful concepts. For example: A set of declarative sentences evokes a question if and only if the set implies that the question has a true direct answer but does not imply the truth of any particular direct answer. Also, one question implies another if and only if each direct answer to the first implies that some direct answer to the second is true and for each direct answer to the second there is a proper subset (that is, a small selection) of direct answers to the first such that the given direct answer to the second implies that a true direct answer to the first is contained in that subset (so each true answer to the second helps us find a true answer to the first).

Another of Wiśniewski's concepts is that one question is reducible to a set of other questions (call it the reducing set) if and only if each direct answer to the given question implies that all the questions in the reducing set have true direct answers; and each collection of sentences that contains exactly one direct answer to each question in the reducing set implies some direct answer to the given question; and no question in the reducing set has more direct answers than the given question does. One result is that, if a question is 'safe' (that is, in every interpretation of the language the question has a true direct answer), then the question is reducible to some set of simple yes-no questions.

DAVID HARRAH

References and further reading

All the works cited contain some technical material. All contain useful bibliographies.

Åqvist, L. (1965) *A New Approach to the Logical Theory of Interrogatives*, Uppsala: Almqvist & Wiksell. (Referred to in §4. Outlines an approach based on commands about knowledge. Technical and difficult.)

Belnap, N.D., Jr and Steel, T.B., Jr (1976) *The Logic of Questions and Answers*, New Haven, CT: Yale University Press.(Referred to in §2. Presents many useful concepts.)

Bromberger, S. (1992) *On What We Know We Don't Know: Explanation, Theory, Linguistics, and How Questions Shape Them*, Chicago, IL: University of Chicago Press.(Referred to in §3. Good discussion of *why* questions.)

Groenendijk, J. and Stokhof, M. (1984) *Studies on the Semantics of Questions and the Pragmatics of Answers*, Amsterdam: Academisch Proefschrift.(Gives an intensional analysis with a pragmatic dimension.)

Harrah, D. (1984) 'The Logic of Questions', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic, Vol. II: Extensions of Classical Logic*, Dordrecht: Reidel, 715-64.(Referred to in §§3, 6. Introduction, some history and a survey of various theories.)

Harrah, D. (1985) 'A Logic of Message and Reply', *Synthèse* 65: 275-94.(§3 presents a general erotetic logic.)

Hintikka, J. (1983) 'New Foundations for a Theory of Questions and Answers', in Kiefer, F. (ed.) *Questions and Answers*, Dordrecht: Reidel, 1983, 159-90.(Referred to in §4. Outlines an approach based on a modification of Åqvist (1965).)

Hintikka, J. (ed.) (1988) 'Knowledge-Seeking by Questioning', *Synthèse* 74: 1-262.(Papers by various authors, mainly on Hintikka's approach.)

Hintikka, J. (1992) 'The Interrogative Model of Inquiry as a General Theory of Argumentation', *Communication and Cognition* 25: 221-42.(Brief introduction to Hintikka-style systems.)

Kiefer, F. (ed.) (1983) *Questions and Answers*, Dordrecht: Reidel.(Papers presenting several approaches and discussing many aspects.)

Tichý, P. (1978) 'Questions, Answers, and Logic', *American Philosophical Quarterly* 15: 275-84.(Referred to in §5. Outlines an approach based on a theory of possible worlds.)

Wiśniewski, A. (1995) *The Posing of Questions: Logical Foundations of Erotetic Inferences*, Dordrecht: Kluwer. (Concerned mainly with raising and implying, as described in §6; technical and rigorous in some chapters but very clear throughout; has useful expositions, surveys and bibliography, and can serve as an introduction to the entire field.)

Quine, Willard Van Orman (1908-)

Quine is the foremost representative of naturalism in the second half of the twentieth century. His naturalism consists of an insistence upon a close connection or alliance between philosophical views and those of the natural sciences. Philosophy so construed is an activity within nature wherein nature examines itself. This contrasts with views which distinguish philosophy from science and place philosophy in a special transcendent position for gaining special knowledge. The methods of science are empirical; so Quine, who operates within a scientific perspective, is an empiricist, but with a difference. Traditional empiricism, as in Locke, Berkeley, Hume and some twentieth-century forms, takes impressions, ideas or sense-data as the basic units of thought. Quine's empiricism, by contrast, takes account of the theoretical as well as the observational facets of science. The unit of empirical significance is not simple impressions (ideas) or even isolated individual observation sentences, but systems of beliefs. The broad theoretical constraints for choice between theories, such as explanatory power, parsimony, precision and so on, are foremost in this empiricism. He is a fallibilist, since he holds that each individual belief in a system is in principle revisable. Quine proposes a new conception of observation sentences, a naturalized account of our knowledge of the external world, including a rejection of a priori knowledge, and he extends the same empiricist and fallibilist account to our knowledge of logic and mathematics.

Quine confines logic to first-order logic and clearly demarcates it from set theory and mathematics. These are all empirical subjects when empiricism is understood in its Quinian form. They are internal to our system of beliefs that make up the natural sciences. The language of first-order logic serves as a canonical notation in which to express our ontological commitments. The slogan 'To be is to be the value of a variable' ([1953] 1961: 15) encapsulates this project. Deciding which ontology to accept is also carried out within the naturalistic constraints of empirical science - our ontological commitments should be to those objects to which the best scientific theories commit us. On this basis Quine's own commitments are to physical objects and sets. Quine is a physicalist and a Platonist, since the best sciences require physical objects and the mathematics involved in the sciences requires abstract objects, namely, sets.

The theory of reference (which includes notions such as reference, truth and logical truth) is sharply demarcated from the theory of meaning (which includes notions such as meaning, synonymy, the analytic-synthetic distinction and necessity). Quine is the leading critic of notions from the theory of meaning, arguing that attempts to make the distinction between merely linguistic (analytic) truths and more substantive (synthetic) truths has failed. They do not meet the standards of precision which scientific and philosophical theories adhere to and which are adhered to in the theory of reference. He explores the limits of an empirical theory of language and offers a thesis of the indeterminacy of translation as further criticism of the theory of meaning.

1 Life

Willard Van Orman Quine was born on 25 June 1908. He was an undergraduate at Oberlin College and a graduate student at Harvard, where he studied with A.N. Whitehead, C.I. Lewis and Sheffer. In his dissertation 'The Logic of Sequences: A Generalization of *Principia Mathematica*' there already appears a prominent theme of Quine's philosophy - a concern with matters of ontology (Quine 1934).

In 1931, Quine had what he has described as his 'most dazzling exposure to greatness', when Bertrand Russell came to lecture at Harvard. Russell is one of the most influential figures on Quine's thought. Both share a preoccupation with questions concerning what there is. For example, Quine adopted and improved upon Russell's view of how we express ontological claims. More significantly, as the dissertation already reveals, Russell's influence was that of a rival whose theories spur Quine on to the creation of more acceptable alternatives. Wherever possible, Quine tries to get on with the fewest and most precise assumptions which will suffice to do the job at hand. Whereas *Principia Mathematica* is constructed on the basis of an ontology that comprises propositional functions, which are properties of a sort, Quine's revision tries to accomplish the same goals with concrete physical objects and sets or classes. In addition, some of Quine's most famous systems of logic and set theory are designed to achieve the same effects as *Principia Mathematica*, while avoiding Russell's theory of types (see Theory of types).

A travelling fellowship to Europe in 1932 exposed Quine to the latest developments in logic and in philosophy. In

Vienna he attended meetings of the Vienna Circle, and described the following weeks spent in Prague and Warsaw as 'the intellectually most rewarding months I have known' (see Vienna Circle). In Prague, Quine met Rudolf Carnap, one of the most careful expositors of prominent themes of analytic philosophy and especially those of the logical empiricists, such as the verifiability criterion for the empirical meaningfulness of sentences, the linguistic (analytic) character of a priori knowledge, as in mathematics and logic, and the triviality or meaninglessness of ontology as a species of metaphysics. Quine subjected each of these themes to severe criticism, resulting in some of the most important philosophical debates of the century. In Warsaw, he attended the lectures of Lesniewski, Łukasiewicz and Tarski and in the next few years was to adopt Tarski's and Gödel's 'classic' formulation of logic in formulating his most famous works. Quine was quite sympathetic to the extensionalist and the nominalist side of the Warsaw school.

At Harvard in the period prior to the Second World War, Quine worked out some of his most distinctive positions: his conception of ontological commitment (best known from his 1948 essay 'On What There Is'); his two most distinctive systems of logic and set theory 'New Foundations for Mathematical Logic' (1937) and *Mathematical Logic* (1940); and his criticisms of the position that a priori knowledge as it purportedly exists in logic and mathematics is merely linguistic. These criticisms began to appear in 1934, when Quine lectured on Carnap's work. Some of this material can be found in 'Truth by Convention' (1936) and his most famous paper, 'Two Dogmas of Empiricism' (1951).

Quine served as a naval officer in the Second World War, and after continued his work on the above topics (see *From a Logical Point of View* (1953)). Much of his most original work since then has been the formulation of a new holistic variety of empiricism and the exploration of its consequences: 'The point of holism, stressed by Pierre Duhem…, is that the observable consequence by which we test a scientific hypothesis is ordinarily not a consequence of the hypothesis taken by itself; it is a consequence only of a whole cluster of sentences… ' (1987: 141). Beginning with 'Two Dogmas', and eventually in *Word and Object* (1960), Quine employed this new holistic empiricism to criticize the concepts of meaning, synonymy and analyticity. In *Word and Object* he presented a thesis of the indeterminacy of translation as a further criticism of these notions. Later, in *Ontological Relativity* (1969), *The Roots of Reference* (1974), *Pursuit of Truth* (1992) and *From Stimulus to Science* (1995) he took a similar critical stance on concepts from the theory of reference. In essays dating from this period Quine's naturalism also comes to the fore. Though the theme of the continuity of philosophy and science is found in earlier works, he explores it more explicitly in 'Epistemology Naturalized' (1969), *Theories and Things* (1981), *Pursuit of Truth* and *From Stimulus to Science*.

2 Epistemology naturalized - nature know thyself

The problem of our knowledge of the external world is traditionally stated as one of how a self with private mental states can come to have knowledge of the external world. Quine's restatement is strikingly more naturalistic:

> I am a physical object sitting in a physical world. Some of the forces of this physical world impinge on my surface. Light rays strike my retinas; molecules bombard my eardrums and fingertips. I strike back, emanating concentric air waves. These waves take the form of a torrent of discourse about tables, people, molecules, light rays, retinas, air waves, prime numbers, infinite classes, joy and sorrow, good and evil.
>
> (Quine 1966: 215)

In its traditional statement the problem lies in how, starting with 'experience' in the form of immediately given impressions or sense-data, we justify our claims to know objects such as tables, chairs or molecules. This vantage point was that of a first philosophy, intended as providing a foundation of certainty for the sciences by standing outside of them and legitimizing their accomplishments. Quine rejects this formulation. His naturalized epistemology rephrases the problem as one of how we learn to talk about or refer to objects (ordinary as well as scientific). What are the conditions that lead to reference? How is scientific discourse possible?

The traditional accounts of the linkage between 'experience' and our knowledge vary from mentalistic conceptions, like that of Hume, in which all our ideas are copies of sense impressions, to more neutral linguistic formulations, in which cognitive claims are to be translated into observation sentences. On Quine's holistic account, one cannot deal with the empirical content of sentences, much less of terms - the linguistic correlates of ideas - one by one, either via definition, translation or some other sort of linkage. To study the relation of

knowledge and science to observation sentences is to trace the psychological and linguistic development of the knower, that is, the potential user of scientific language. Observation sentences serve as both the starting point in human language learning as well as the empirical grounds for science. The problem of knowledge now is how, starting with observation sentences, we can proceed to talk of tables, chairs, molecules, neutrinos, sets and numbers. One of the reasons for doing epistemology by studying the roots of reference is simply the failure of the traditional empiricists' programme mentioned above. Another is that it enables one to dispense with mentalistic notions such as 'experience' or 'observation'. One relies instead on two components which are already part of a naturalist's ontology: the physical happening at the nerve endings, the neural input or stimulus; and the linguistic entity, the observation sentence. These two serve as naturalistic surrogates for 'experience' and 'observation'. On Quine's empiricist and behaviourist account, observation sentences are those that can be learned independently of other language acquisition. They are the sentences that can be learned purely by ostension and as such are causally most proximate to the stimulus. This account is not vulnerable to attacks on the notion of observation as dependent on the theories one holds, since observation sentences are precisely those which are learnable without any background information. Another point of difference with empiricists concerns the alleged certainty or incorrigibility of observation. Though Quine's observation sentences are assented to with a minimum of background information and are thus included among those sentences less likely to be revised, they are not in principle immune from revision.

Unlike traditional epistemology, then, Quine's epistemology is naturalistic: we cannot stand apart from our place as part of nature and make philosophical judgments (see Naturalized epistemology). This is part of the theme that philosophy is continuous with science, science being the part of nature most suitable for knowing itself.

> The naturalistic philosopher begins his reasoning within the inherited world theory as a going concern. He tentatively believes all of it, but believes also that some unidentified portions are wrong. He tries to improve, clarify and understand the system from within.
>
> (Quine 1981: 72)

3 Dethroning the a priori

A purported stumbling block for empiricism is a priori knowledge in logic and mathematics and in such purportedly conceptual truths as 'All bachelors are unmarried' and 'Nothing is larger than itself'. Such subject matter appears to defy justification in terms of observation. J.S. Mill's empiricist programme foundered on this point, and the various forms of rationalism are unacceptable to an empiricist. A proposed solution, the dominant one favoured by analytic philosophers, involved the analytic-synthetic distinction: all a priori knowledge was said to be analytic, in the sense that the truth of sentences claimed to be known independent of experience was reducible to matters of language, for example, linguistic convention, definitions, and truth in virtue of the meaning of the expressions involved.

Quine's critique is that of an empiricist reforming empiricism, supplanting reductionist-atomistic empiricism with Duhemian holism. Experimental testing is a juncture where observation ('experience') enters as a factor in deciding whether to accept or reject a claim. An oversimplified model of how observation counts in testing is that given a hypothesis and a statement of initial conditions, we deduce by logic and/or mathematics some observation sentence as an observable consequence. If the expected observation occurs, we take this as evidence for the hypothesis. If it does not, we take this as evidence that the hypothesis is false. Such a model makes use of the dogma of reductionism by assuming that individual sense experiences-observation sentences function unequivocally for or against isolated sentences-hypotheses. Pierre Duhem showed that this model is flawed and Quine extends Duhem's point into a holistic one that embodies a critique of reductionism and the a priori.

Duhem had pointed out that where the observation fails to occur, one has leeway in dealing with the situation. All serious testing involves background assumptions, implicit in the hypothesis or in the statement of the initial conditions. A test situation underdetermines which factor should be revised and there is no way of knowing in advance where the revision should be made. Quine's insight was to take cognizance of all the assumptions that can be questioned in a test situation. The underdetermination of theory by observation does not stop with revising background assumptions relative to the hypothesis and the initial conditions. Both the purportedly disconfirming observation made and the principles involved in deriving the observable consequence can also be revised. In test

situations whole systems of beliefs go into the hopper and we have leeway as to how we make the consistency-preserving revisions. We can edit the observation. We can even question the logic and or mathematics used in deriving the observable conclusion. No sentence is in principle immune from being revised. In this spirit, all knowledge is empirical; there is no a priori knowledge. The quest for certainty is replaced by fallibilism, the associated foundationalist programmes are abandoned, as is the verifiability theory of the logical empiricists. Instead, the broad constraints on what to do in a test situation are the natural scientist's criteria for preferring one hypothesis, theory or system of beliefs over another. These criteria include explanatory power, simplicity or parsimony, conservatism, modesty and precision (1970: 54). Conservatism cautions, other things being equal, that we should accept that hypothesis-theory which clashes least with our other beliefs. This 'maxim of minimal mutilation' comes into play in explaining why logic, while in principle revisable (for example, some suggest adopting a three-valued logic), is the least likely item to be revised. Doing so would have far-reaching consequences for our other beliefs. Modesty says that all other things being equal we should hypothesize as little as is necessary for the job at hand. Precision mandates standards on introducing and explaining philosophical or scientific concepts. It requires that philosophical explications be couched in acceptable terms, the extensional or empirical notions of logic, set theory and the sciences. For example, abstract objects should not be posited without a precise account of what they are, that is to say, 'no entity without identity' (see A Priori §1; Analyticity §§1-3; Fallibilism).

4 Logic as first-order logic

Quine distinguishes the theory of reference from the theory of meaning. He is sceptical of notions associated with the theory of meaning, such as those of meaning, intension, synonymy, analyticity and necessity. By contrast, he relies on and makes contributions to the theory of reference, for example, to the understanding of logical truth, truth, reference and ontological commitment.

For Quine, the notion of logical truth falls squarely in the theory of reference. His most characteristic definition of logical truth is that a sentence is a logical truth if it is true and if it remains true when one uniformly replaces its nonlogical parts. The logical parts are the logical constants, signs for negation, disjunction, quantification and identity. 'Brutus killed Caesar or it is not the case that Brutus killed Caesar' is such a logical truth. In Quine's terminology the logical constants 'or' and 'it is not the case that' occur essentially, while the nonlogical part 'Brutus killed Caesar' can be uniformly varied and the resulting sentence will still be true. In other words a logical truth cannot be changed into a falsehood by varying the nonlogical expressions, whereas an ordinary truth can be.

The same concept of logical truth is found in Bolzano and Ajdukiewicz. One of its virtues lies in its being parsimonious, that is, in what it does not say. Logical truth and related notions are often explained in modal terms. That is, logical truths are said to be distinguished by being 'necessary' or 'true in all possible worlds', and a valid argument is defined as one in which, if the premises are true, the conclusion 'must be true'. These accounts make logic presuppose modal notions. Quine's definition leaves logic autonomous in this respect. Indeed Quine is a critic of modal logic, challenging various attempts to explain the notion of necessity. Logical truth, as defined by Quine, is a precisely explained species of truth, fitting squarely inside the theory of reference. Quine relies on the concept of truth, which he construes along the lines of Tarski's theory (see Tarski's definition of truth).

If logical truths are those in which only logical constants occur essentially, then the scope of logic is in part determined by what we take to be a logical constant. Quine lists as the logical constants the truth-functional connectives 'not', 'and', 'or', 'if… then', 'if and only if'; the quantifiers 'all' and 'some'; and the identity predicate '$a = b$'. The language of logic so construed is that of sentences formed out of truth-functional connectives, quantifiers, identity, schematic predicate letters and individual variables. Quantificational logic of this sort is also known as first-order logic. For Quine, logic *is* first-order logic with identity. Ruled out as logic on this construal are modal logic, because 'necessity' is not taken as a logical constant. Also excluded is higher-order logic, which has quantifiers for predicate positions (it is 'set theory in sheep's clothing' (1986: 66)). On other grounds other proposals such as intuitionist logic are also ineligible. Set theory, and with it mathematics, are not logic (see Logical constants §§1-2).

Quine falls in the camp of the Logicist programme in holding that mathematics is reducible to set theory. Set theory is the theory of the 'is a member of' predicate and is stated in the language of first-order logic. Given the theory of membership and logic as first-order logic plus identity, Quine introduces mathematical notions as

definitional abbreviations: for example, a number is defined as a special set, addition as a special function on these sets and so on. He argues that logic does not include set theory because membership should not be considered a logical constant for the following reasons: (1) There is a general consensus about elementary logic, which, given paradoxes such as Russell's, is lacking in the case of set theory. Alternative set theories have the status of so many tentative hypotheses. This lends credence to Quine's view that mathematics based on set theory is not very different from other sciences. (2) The incompleteness of set theory contrasts sharply with the completeness of elementary logic. (3) The ontology of set theory is not as topic neutral as that of logic. The second item in the membership relation is restricted to sets. Logic is the most general of subjects, since the variables of logic are not restricted to any one category of objects (see Logicism §1).

5 Canonical notation and ontological commitment

Ever since Frege and Russell, existential quantification has been the prevalent way in which existence assertions have been understood. The idea is that existence sentences in natural language can be paraphrased in the language of logic and 'existence' explicated by the existential quantifier of predicate logic. 'Existence is what existential quantification expresses' (1969: 97). The functions that the predicate 'exists' performs in English can be accomplished by the '$\exists x$' quantifier. Quine's version of this theme is incorporated into his account of ontological commitment. The language of first-order predicate logic in which our ontological commitments are made is his 'canonical notation'.

> Taking the canonical notation thus austerely… we have just these basic constructions: predication, quantification… and the truth functions… What thus confronts us as a scheme for systems of the world is that structure so well understood by present-day logicians, the logic of quantification or calculus of predicates.
>
> Not that the idioms thus renounced are supposed to be unneeded in the market place or in the laboratory… The doctrine is only that such a canonical idiom can be abstracted and then adhered to in the statement of one's scientific theory. The doctrine is that all traits or reality worthy of the name can be set down in an idiom of this austere form if in any idiom.
>
> It is in spirit a philosophical doctrine of categories, … philosophical in its breadth however continuous with science in its motivation.
>
> (Quine 1960b: 228-9)

A key reason for regarding this language as canonical is that in it one's use of the existential quantifier is explicit. To discover the existence assumptions or ontological commitments of a theory, we first state it in the language of truth-functional connectives and quantification, and then look at the existential quantifications we have made. The logic of '$\exists x$' is the logic of existence, and a notation that makes '$\exists x$' explicit accordingly makes our existence assumptions or ontological commitments explicit.

One of Quine's most famous remarks, 'to be is to be the value of a variable' ([1953] 1961: 15), sums up this criterion of ontological commitment. The slogan also incorporates one of Quine's elaborations on Russell's theory of definite descriptions. Russell held that in most sentences ordinary names are disguised definite descriptions, which on his theory are analysed as existential generalizations. Thus 'Socrates is human' becomes 'The one husband of Xanthippe is human' which in turn becomes the existential generalization 'There is one and only one husband of Xanthippe and he is human'. Quine's elaboration dispenses with names entirely. Wherever a name occurs in the original sentence, we can get by in the canonical notation with variables, predicates and the logical constants. If we do not have a definite description on hand to put in place of the name, we can form a predicate, such as 'Socratizes', then encapsulate the coined predicate in a definite description and define away the descriptions via Russell's theory. So names disappear and are not part of the canonical notation. Dispensing with names not only has the virtue of a more parsimonious notation: it also specifies that variables are the vehicles of reference and as such the grounds of ontological commitment. Being the value of a variable is what 'being' is all about (see Logical form; Ontological commitment).

6 Competing ontologies

Some important philosophical differences concern competing ontologies. Physicalists, for instance, have as their basic objects physical objects, while phenomenalists have sense-data. A twentieth-century version of the problem

of universals involves a dispute over the relative merits of (1) a nominalistic ontology according to which only concrete individuals exist; and (2) realist ontologies, such as those of Platonists, which involve the existence of abstract objects. The issue of nominalism versus Platonism arises for Quine in connection with the mathematics required for science and the question whether it requires hypothesizing abstract objects such as sets. Another area of ontological controversy is whether Platonist assumptions should include only extensional objects such as sets or intensional ones such as properties or propositions (see Abstract objects §3; Nominalism §4.

While we look to the existential generalizations of a theory stated in canonical notation to see what its ontological commitments are, this does not answer the question of which ontological commitments we should have. As a naturalist and scientific realist, Quine regards this as a matter of epistemology. It is the question of which theory we ought to accept. Deciding on a theory (scientific or philosophical) and its attendant ontology is once again done within a scientific perspective. Appeal is made to the same theoretical concerns mentioned earlier, that is, explanatory power, parsimony, conservatism, precision and so on. Quine's own ontological commitments are the result of just such considerations. He is a physicalist, no longer taking seriously the phenomenalist programme; partly on grounds connected with the dogma of reductionism and partly on the grounds that sense-data are not needed in a naturalized epistemology (the functions performed by sense-data are accomplished by nerve hits and observation sentences which are already part of our physicalist ontology). But he is a Platonist (a reluctant one) because of the mathematics that is required by our best scientific theories. In canonical notation, this mathematics requires quantifying over at least as many extensional abstract objects, namely sets, as there are real numbers. In an early essay co-authored with Nelson Goodman, 'Steps Towards A Constructive Nominalism' (1947), the possibilities of taking a nominalist stance were surveyed. But unlike Goodman, Quine reluctantly concluded that the nominalist programme failed. He has considered later attempts at nominalism such as by an appeal to substitutional quantification. The success of these attempts depends on whether impredicative notions are required by the mathematics embedded in the natural sciences. Others (Field 1980) try to make the case for nominalism by abiding by some and abandoning other of Quine's constraints.

However, Quine has consistently argued against theories which require abstract objects of an intensional sort. An intensional notion might provisionally be characterized as one requiring for its explanation notions which do not conform to certain basic assumptions of first-order logic and standard set theories. Intensional contexts are not truth-functional, or do not allow the substitutability of coextensive singular terms or predicates. Modal notions are a case in point. 'Necessarily P' is not a truth-functional operation; we cannot replace singular terms or predicates in P, or the whole sentence P, with coextensive expressions and be guaranteed that if the original sentence is true then the subsequent one will also be true. In non-intensional contexts, that is, extensional contexts, such replacements are truth-preserving. Intensional objects are those that require intensional notions to account for them, such as properties or propositions. Extensional objects are those that do not require intensional notions to account for them, such as sets or sentences. So sets are identical if they have the same members, for example, the set of humans is the same as the set of featherless bipeds. Properties, on the other hand, are identical only if they necessarily belong to the same objects. So the properties of being human and being a rational animal are identical, but differ from the property of being a featherless biped.

Given Quine's views on the precision of notions from first-order logic and set theory, it is not surprising that he is critical of intensional notions which are not based on standard assumptions of first-order logic and set theory. His arguments against hypothesizing intensional objects are numerous and can be sketchily listed as cases where their explanatory power is questioned, where parsimony is invoked and where precise accounts of the identity conditions of such items is lacking ('no entity without identity'). Sometimes he argues that extensional ersatz constructions can achieve the same purpose for which the intensional objects were introduced: for instance, eternal sentences rather than propositions, sets rather than properties, ordered pairs rather than relations (construed intensionally), non-modal notions rather than modal ones, and so on (see Intensional entities §1).

7 Indeterminacy of reference and global structuralism

Quine's views on ontology undergo refinements in his later writings. Most important is the recognition that empiricism does not uniquely determine which objects are required as the values of our variables. There is an indeterminacy of reference which is in keeping with empiricist strictures on deciding which ontology to accept.

The thesis of ontological relativity, also known as the inscrutability or indeterminacy of reference, is in accord

with Quine's naturalistic empiricism. It has been generalized into a view which he refers to as global structuralism. Since it is only at the observation sentences construed holophrastically as indissoluble wholes that the system is externally constrained, there are different but equally plausible ways of meeting these observational constraints and these can involve diverse ontologies, such as an ontology of rabbits or of rabbit parts. It is the structural part of the system that must be saved in order to meet the observational constraints. But this can be accomplished with quite different objects being the values of the variables. Quine endorses this global structuralist perspective by generalizing from his own cases and by noting a less global structuralist argument from the philosophy of mathematics. Quite different objects can be taken as the values of the variables for arithmetic, for example, numbers can be taken as Frege-Russell sets or as von Neumann sets without changing the truths of arithmetic. For Quine the question of whether we are really committed to rabbits as opposed to sums of rabbit parts, or to a given number as the set of all sets equinumerous to a given set (as on the Frege-Russell account) or to some different set (as on von Neumann's view) is without sense. It is without sense in that there is no natural-empirical way of raising this question. Global structuralism is a consequence of the denial of a first philosophy - a point of view in nature that transcends all natural points of view.

Quine's best-known cases which serve as evidence for his global structuralism are his rabbits case and his proxy functions, especially that of cosmic complements. Intertwined with his discussion in *Word and Object* of a linguist translating a native speaker's utterance of 'Gavagai', Quine points out that construing that expression as a referring expression leaves no empirical way of deciding whether it is used to refer to rabbits, rabbit parts, or rabbit stages and so on. As a later example he asks us to consider how sentences about concrete objects can be reinterpreted in terms of different ontologies assigned as values of the variables so that there is no empirical way of saying which is the correct one. Indeed the message of structuralism is that it is an error to speak as though there were a uniquely correct referent. The sentence 'This rabbit is furry' is true and is usually interpreted as being about individual rabbits and individual furry things. However, if we reinterpret the referring portions in terms of mereological cosmic complements the sentence remains true and there is no empirical way if we do this uniformly to say which is the correct ontology. Thus assign to 'This rabbit' the entire cosmos less this rabbit (imagine a complete jigsaw puzzle with a rabbit piece removed; the cosmic complement would be the puzzle without the rabbit piece). Assign to the predicate 'is furry' each of the cosmic complements of individual furry things. The sentence 'This rabbit is furry' is true under such an interpretation because the cosmos less this rabbit is a member of the set of cosmic complements of individual furry things (that is, that set includes the cosmic complement of that individual rabbit). One can extend this treatment of singular sentences to the remaining referential sentences.

8 The theory of meaning: its myths and dogmas

Quine's reaction in the 1930s to Carnap's work was a sceptical critique of the prevalent view among analytic philosophers that logic and mathematics are justified in some distinctively linguistic way, that is, that they are based on merely analytic-linguistic truths - truths by convention. The scepticism then and later questions proposed accounts of the distinction. Linguistic truth or truth by convention, as opposed to non-conventionally based empirical truths, appears in the end to be a purported distinction without a real difference. Of special importance is the failure to satisfy the requirement of precision in explaining the distinction. In 'Truth By Convention' (1936) and 'Carnap on Logical Truth' (1960a), Quine takes up different attempts to characterize such truths and finds that for the most part they are either too broad - not distinctive of logic or mathematics - or require non-linguistically based truth. There are as many of these characterizations as there are different senses of 'convention'. Quine considers truth by convention as based on the following: the arbitrary factor in axiomatization; formalization-disinterpretation; the arbitrary element in hypothesizing; and definition. But neither logic nor mathematics is distinguished by being axiomatized, or formalized-disinterpreted. The same can be done for other disciplines, such as physics and biology. The somewhat arbitrary choice of which sentences to take as axioms, so long as we can prove the right sentences, is also not distinctive of them. If truth by convention is taken as the somewhat arbitrary element in framing hypotheses, then this too is not distinctive of logic or mathematics. Nor are the formulas involved distinguished by being true by definition. Thus, if '$p \rightarrow p$' is defined in terms of '$-p \vee p$', then the truth of the defined formula depends on the truth of the defining formula, and that formula's truth is not a matter of definition or convention. For Quine, logic and mathematics can be precisely characterized in terms from the theory of reference. Logic is described in terms of truth, the logical constants and interchange of the extralogical elements; mathematics can be characterized in terms of set theory. But neither of these subjects is

distinct in having a different epistemological basis which results in their being in some interesting sense ‘analytic’ or mere ‘linguistic truths’.

The ‘Two Dogmas’ of Quine’s 1951 essay were the dogma of reduction (see §3 above) and the dogma of the analytic/synthetic distinction. In discussing the dogma of analyticity, Quine questions, as he did in ‘Truth by Convention’, whether the distinction can be well made. Here he rejects five ways of explaining analyticity. These involve appeals to: (1) meanings, (2) definition, (3) interchangeability, (4) semantic rules and (5) the verifiability theory of meaning.

(1) ‘Bachelors are unmarried men’ might be regarded as analytic, where that notion is explained as truth in virtue of the meanings of its words. One suggestion is that the meaning of ‘unmarried man’ is included in the meaning of ‘bachelor’. Another approach would hypothesize the existence of meanings to explain synonymy and then use synonymy in turn to show how the above sentence is a synonymous instance of a logical truth. The success of the above explanations requires assuming that there are such things as precisely characterizable meanings. Quine is sceptical of this assumption. He rejects three accounts of meanings: (a) referential theories - meanings as referents; (b) mentalism - meanings as ideas; and (c) intensionalism - meanings as intensional entities. Meaning (or sense), as Frege taught, must be distinguished from reference. The notion of meaning that is to explain synonymy and analyticity cannot simply be reference, because coreferential terms need not be synonymous and so will not distinguish analytic sentences from true non-analytic ones. Truth in virtue of meaning where meaning is simply reference is too broad as all truths would then be analytic. As to meanings as ideas, Quine brings to bear empiricist and behaviourist qualms. The account of meanings as abstract intensional entities is found in Frege, Carnap and Church. Quine maintains that such intensional objects are neither required as posits by our theories of language, nor are they precisely accounted for. The attempt to explain intensional notions is either circular or unhelpful. There is a circle of intensional notions, meaning, synonymy, analyticity, necessity, and we can define one in terms of another. Quine’s criticism is that if we do not break out of this intensional circle, then the account has failed to clarify the matter. For example, if the meaning of a predicate ‘is human’ is the property of being human, how would one go about identifying whether ‘being a rational animal’ or ‘being a featherless biped’ stood for the same property or had the same meaning? One answer is that the sentence ‘humans are rational animals’ is analytic, while the sentence ‘humans are featherless bipeds’ is not. But this relies on the notion of ‘analytic’ - which we haven’t yet defined. Another approach at giving an identity condition uses modal notions and says that the first sentence is a necessary truth while the second is not. This, however, raises the problem of giving a precise account of modal notions. Explaining modal claims in terms of analyticity, for example, ‘Necessarily humans are rational’ as explained by ‘"Humans are rational" is analytic’, will not do - since we have not defined ‘analytic’. Quine’s challenge is that one break out of this intensional circle and explain notions from the theory of meaning in more acceptable terms.

(2) He next rejects accounts of analyticity in terms of logical truth and synonymy. On this account a sentence is analytic if it is a synonymous instance of a logical truth, that is, ‘All bachelors are unmarried men’ is analytic in that it is derivable from the first-order logical truth ‘All bachelors are bachelors’ by putting a synonym ‘unmarried man’ for the second occurrence of ‘bachelor’. One account attempts to explain synonymy in terms of definition. However, the various forms of definition either presuppose synonymy or stipulate it; none explain it. Quine is sceptical of definitions or philosophical analysis when thought of as capturing or analysing some concept or meaning. Instead philosophical explication is thought of in terms of the theory of reference and scientific hypothesizing and thus again embodying the naturalistic theme of the continuity of science and philosophy. One does not capture ‘the meaning’ of an expression; one explicates or proposes a theory of the referential features one is interested in preserving.

(3) Another attempt to define synonymy asserts that two expressions are synonymous if they are interchangeable. But it is not enough to say expressions are synonymous when the interchange of the one with the other within extensional contexts does not change the truth value of the sentences involved. This has the unacceptable consequence that merely coextensive terms would be synonyms. To do better one has to require interchangeability within intensional contexts. However, this raises the problem of breaking out of the circle of intensional notions.

(4) The fourth approach is another of Carnap’s. It consists in constructing an artificial language and then defining ‘analytic’ for it. While it is possible to construct a language and specify that relative to it logic, mathematics and

such truths as ‘All bachelors are unmarried men’ and ‘Nothing is larger than itself’ are analytic, this language-relative specification of analyticity does not clarify matters. It does not help to be told that in one language, language 1 (artificial or otherwise), we have a list of sentences that are analytic 1, and that in another language, language 2, we have the list analytic 2 and so on. What we want of an explication of analyticity is an account of what analytic 1, analytic 2 and so on have in common. The appeal to artificial languages fails to provide this characterization. Moreover, the problem is precisely why ‘All bachelors are unmarried’ is on the list and ‘No bachelors are six-legged’ is not. To be told that a sentence is analytic because it is on a list (even the list of an artificial language) provides no real distinction.

(5) The last attempt to define analyticity that Quine considers appeals to the verification theory of meaning. According to this theory, ‘the meaning of a statement is the method of empirically confirming or infirming it’; ‘statements are synonymous if and only if they are alike in point of method of empirical confirmation or infirmation’ ([1953] 1961: 37). Though sympathetic towards the empiricist thrust of this theory Quine does not think it survives the holistic criticism of the dogma of reductionism.

Quine’s critique of the theory of meaning has amounted to a challenge to provide precise accounts of its notions. What counts as precise could take the form of reducing intensional notions to extensional ones. His criticisms of modal concepts has spurred a generation of responses in what is known as possible world semantics (see Modal logic, philosophical issues in §3), which in one of its variations can be seen as trying to provide a reduction of intensional modal notions via extensional metalinguistic truth conditions for modal statements. The success of this reduction is still challenged by Quinians (1981: 113-24). More in keeping with Quine’s challenge to explicate the theory of meaning is Davidson’s work on letting a Tarskian theory of truth serve as surrogate for a theory of meaning. Another way in which scepticism about the theory of meaning might be overcome would be by an empirical and behaviouristically constrained account of such notions. Carnap took up this challenge and sketched a programme for empirically identifying meanings by testing translation hypotheses, for example, a linguist’s hypotheses for translating the terms ‘Pferd’ from German to English as ‘horse’. Quine’s response was the topic of radical translation and his thesis of the indeterminacy of translation.

9 Indeterminacy of translation

How much of language is susceptible to empirical analysis? Like Carnap, Quine takes the case of linguists hypothesizing about translations as the subject matter for empirical inquiry. Both take as their data a native speaker’s response to appropriate stimuli. Quine introduces the concept of the stimulus meaning of a sentence for a person as the class of stimulations which would prompt them to assent to it. But Quine deals with the stimulus meaning of whole sentences, such as ‘This is a horse’, and not terms, such as ‘horse’. Quine’s linguist offers a hypothesis equating two sentences (one is the native’s and the other the linguist’s) and checks it against a native speaker’s assenting or dissenting to the native sentence in the presence of some nonverbal stimulus. Carnap considered translation for languages such as German and English which are known to have much in common. For Quine, the critical case is that of radical translation, that is, translations between languages that have little or nothing in common. Think of a linguist among some radically foreign tribe. The linguist observes a certain correlation between a native utterance of ‘Gavagai’ and the presence of rabbits, and proceeds to frame a hypothesis which equates ‘Gavagai’ and the one-word sentence ‘Rabbit’, short for ‘Here’s a rabbit’. The linguist could, on learning how to recognize the native’s assent and dissent, question the native by uttering ‘Gavagai’ when a rabbit appears and seeing whether the native assents. Carnap would presumably want this to count as evidence that ‘Gavagai’ and ‘Rabbit’ have the same meaning. But does the evidence really go that far? All that we have as data is the native’s expression and the rabbit stimulation. Quine claims that on these grounds one could equally well translate ‘Gavagai’ as ‘rabbit stage’ or ‘temporal part of a rabbit’ or something else. For wherever there are rabbit stimulations there are also rabbit-stage stimulations, and so on. On what basis, then, would one decide between these different translations? In the case of culturally similar languages one assumes a stock of more theoretical guides to translations, and thus one can ask the German whether all horses are *Pferde* and all *Pferde* are horses. In the case of radical translation, the linguist is not in a position to pose these more theoretical questions. At this point hypotheses less directly connected to the data - to the stimulus conditions - may be introduced by the linguist. These, which Quine calls analytical hypotheses, can be framed so as to do justice to quite different translations. Thus radical translation provides evidence for the thesis of the indeterminacy of translation (see Radical translation and radical interpretation §§2-5).

To illustrate this matter for the Gavagai case we must note that in order to ask the question 'Is this rabbit the same as that?' the linguist must have decided how to translate articles, pronouns, identity predicates and so on. To translate these sentences is to go far beyond the data provided by the stimuli and involves selecting from different sets of analytical hypotheses, that is, different possible manuals of translation. On one set of these we translate the question as 'Is this the same rabbit as that?', while on another as 'Is this rabbit stage of the same series as that?' Each of these translations is equally good, yet they are mutually incompatible. Since neither of these has any immediate connection with the Gavagai stimulation there is no way of deciding between them.

This indeterminacy provides further grounds for discrediting the notion of meaning. Philosophers have talked as if meanings are related to expressions in somewhat the same way as paintings in a museum are related to their labels. Quine dubs this 'the myth of the museum' (1969: 27). According to this view, two expressions are synonymous when they are related to a unique meaning, like two labels for the same painting. In the case of translation, one English expression is a translation of another in a different language when the two bear a relation to one and the same interlinguistic object which is their meaning. Quine is attempting to dislodge this model for thinking about language and to put in its place a more empirically based conception. According to the museum model, an expression has its meaning, pure and simple, and two synonymous expressions relate uniquely to one meaning which, as interlinguistic, is independent of the languages in which it is expressed. What Quine has shown is that it makes no sense to speak of language-independent meanings. Translation from one language to another is relative to a set of analytical hypotheses. There is no independent meaning of 'Gavagai' which the linguist can link to 'This is a rabbit' and not 'This is a rabbit stage'. The linguist is at best in a position of saying that 'Gavagai' and 'This is a rabbit' are synonymous relative to the assumption of certain analytical hypotheses. We have to study language in terms of linguistic behaviour in the face of stimulus conditions; in turn this behaviour must be interpreted in relation to more theoretical background assumptions. With the exception of observation sentences, we have at best relative notions of meaning (synonymy and analyticity), and these will not do the job philosophers have frequently assigned to them according to the myth of the museum.

Quine insists on behaviourism as the required method for studying language learning. He argues that we learn language by observing verbal behaviour and having our verbal behaviour reinforced by others. But there are naturalistic accounts of language learning which are scientifically respectable and not behaviourist (see Chomsky, N. §3). Yet although Quine insists on behaviourism as the method for studying and acquiring languages, he is not a behaviourist in psychology or the philosophy of mind. On the mind-body problem he endorses Davidson's anomalous monism - the view that our ways of speaking of the mental, for example, of perceptions and beliefs, cannot be stated in terms of the natural laws which govern the underlying physiological states, even though our mental states just are such states. Quine construes the matter so that mental ascriptions have a role in everyday life and the social sciences that cannot be precisely specified in purely physiological or physicalist terms.

See also: Analytical philosophy §4; Behaviourism, methodological and scientific

ALEX ORENSTEIN

List of works

Quine, W.V. (1934) *A System of Logistic*, Cambridge, MA: Harvard University Press.(A revised version of Quine's dissertation.)

Quine, W.V. (1936) 'Truth by Convention', in O.H. Lee (ed.) *Philosophical Essays for A.N. Whitehead*, New York: Longmans.(Critique of attempts to justify logical and mathematical truth in some distinctively non-empirical way - for example, as grounded in conventions.)

Quine, W.V. (1937) 'New Foundations for Mathematical Logic', *American Mathematical Monthly* 44: 70-80.(One of Quine's most famous alternatives to issues in the foundations of mathematics such as Russell's paradox.)

Quine, W.V. (1940) *Mathematical Logic*, Cambridge, MA: Harvard University Press.(One of Quine's most systematic logic texts, including another alternative account of the foundation of mathematics.)

Quine, W.V. and Goodman, N. (1947) 'Steps Towards a Constructive Nominalism', *Journal of Symbolic Logic* 12: 97-122.(An early examination of the prospects for a nominalist programme in mathematics and its limitations.)

Quine, W.V. (1948) 'On What There Is', *Review of Metaphysics* 2: 21-38.(The classic statement of Quine's view on ontological commitment.)

Quine, W.V. (1951) 'Two Dogmas of Empiricism', *Philosophical Review* 60: 20-43.(Quine's classic critique of

the analytic-synthetic distinction and of reductionist-atomistic empiricism.)

Quine, W.V. (1953) *From a Logical Point of View*, Cambridge, MA: Harvard University Press, revised, 1961.(An early collection of some of his most famous essays, including ‘New Foundations for Mathematical Logic’, ‘On What There Is’ and ‘Two Dogmas of Empiricism’. The 1961 edition incorporates many important revisions.)

Quine, W.V. (1960a) ‘Carnap on Logical Truth’, *Synthèse* 12: 350-74.(Continues criticism of the attempt to give a distinctly linguistic justification of logical truth.)

Quine, W.V. (1960b) *Word and Object*, New York: Wiley & Sons.(A major source for Quine’s views on meaning, translation, reference, modality and propositional attitudes.)

Quine, W.V. (1966) *The Ways of Paradox and Other Essays*, New York: Random House.(A collection of essays on various topics, including ‘Truth by Convention’ and ‘Three Grades of Modal Involvement’.)

Quine, W.V. (1969) *Ontological Relativity and Other Essays*, New York: Columbia University Press.(Collection of essays including the title essay and others such as ‘Epistemology Naturalized’.)

Quine, W.V. and Ullian, J.S. (1970) *The Web of Belief*, New York: Random House.(An introduction to the scientific method from a Quinian perspective.)

Quine, W.V. (1974) *The Roots of Reference*, La Salle, IL: Open Court.(A Quinian, methodologically behaviourist account of how we get to talk about objects.)

Quine, W.V. (1981) *Theories and Things*, Cambridge, MA: Harvard University Press.(Essays, mainly from this period, on positing objects and what it means.)

Quine, W.V. (1985) *The Time of My Life: An Autobiography*, Cambridge, MA: MIT Press.(Quine’s very readable autobiography.)

Quine, W.V. (1986) *Philosophy of Logic*, Cambridge, MA: Harvard University Press.(An introduction to the subject via a discussion of truth, modalities and propositional attitudes, logic - its nature, scope and justification.)

Quine, W.V. (1987) *Quiddities*, Cambridge, MA: Harvard University Press.(Introduction to various philosophical topics via a dictionary format.)

Quine, W.V. (1992) *Pursuit of Truth*, Cambridge, MA: Harvard University Press.(A survey and update of Quine’s views on cognitive meaning, objective reference and the grounds of knowledge.)

Quine, W.V. (1995) *From Stimulus to Science*, Cambridge, MA: Harvard University Press.(A distinctive short tour of the history of philosophy, followed by an exploration of topics such as epistemology naturalized, reference, empirical content and the philosophy of mind.)

References and further reading

Barrett, R.B. and Gibson, R. (eds) (1990) *Perspectives on Quine*, Oxford: Blackwell.(Papers given at a conference for Quine’s 80th birthday, followed by his replies.)

Davidson, D. and Hintikka, J. (eds) (1975) *Words and Objections*, Dordrecht: Reidel.(A collections of essays followed by Quine’s comments.)

Field, H. (1980) *Science Without Numbers*, Princeton, NJ: Princeton University Press.(An important attempt at a nominalistic approach to mathematics.)

Gibson, R. (1982) *The Philosophy of W.V. Quine*, Gainesville, FL: University Presses of Florida.(An extremely fine secondary source on Quine.)

Hahn, L.E. and Schilpp, P.A. (eds) (1986) *The Philosophy of W.V. Quine*, La Salle, IL: Open Court.(The Quine volume in a distinguished series, containing an intellectual autobiography by Quine, essays on his work and his replies; also contains an excellent bibliography of Quine’s works.)

Inquiry 37 (1994) *Symposium on Quine*.(A journal issue with Quine’s comments on the essays.)

Leonardi, P. and Santambroggia, M. (eds) (1995) *On Quine*, Cambridge: Cambridge University Press.(A conference volume on Quine with his comments.)

Orenstein, A. (1977) *Willard Van Orman Quine*, Boston, MA: G.K. Hall.(An introduction to Quine’s thought and its place in twentieth-century philosophy.)

Orenstein, A. and Kotatko, P. (eds) (1998) *Knowledge, Language and Logic: Questions for Quine*, Boston Studies in the Philosophy of Science, Dordrecht: Kluwer. (A collection of essays followed by Quine’s comments.)

Revue Internationale de Philosophie (January 1998), ed. P. Gochet.(A special Quine issue with papers on his work followed by his comments.)

Russell, B.A.W. and Whitehead, A.N. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(A groundbreaking classic of modern logic and of the logicist programme of reducing mathematics to higher-order logic.)

Rabelais, François (*c.*1483-1553)

Rabelais, a French humanist and comic writer of the Renaissance, is best known for his chronicles of Gargantua and Pantagruel, in which coarse popular humour, fine Lucianic irony and staggering erudition are uniquely blended, and which claim to reveal, first appearances notwithstanding, 'certain very high sacraments and dread mysteries, concerning not only our religion, but also our public and private life' ([1532-52] 1955: 38).

Rabelais has been subjected to the most contradictory interpretations and judgments. Like Erasmus, whom he admired, Rabelais was attacked in his own time by schismatic Protestants (most notably Calvin) and by reactionary Catholics (most notably the faculty of theology at Paris), as an obscene Lucianic atheist and a heretic. At the same time he was admired and supported by high-minded patrons including Francis I, the king's devout sister Marguerite de Navarre, and Cardinal Jean Du Bellay. Even today Rabelais' religion and philosophy are the subject of debate among scholars, while his work is known to non-specialists more for the 'Rabelaisian' ribaldry of a few pages than for the complex irony and profoundly humanistic design that characterize his works as a whole.

1 Intellectual orientation

Rabelais himself was a man of multiple talents and professions, prodigiously learned in all disciplines that were currently being transformed under the influence of humanism: classical languages (Latin, Greek, Hebrew), ancient and modern literatures, history, moral philosophy, medicine, law, theology. Son of a lawyer in Chinon, he served in the orders of St Francis and St Benedict before becoming a lay priest, a doctor and professor of medicine, an accomplished student of Roman law, an amateur biblical humanist, and a member of French diplomatic missions to the Holy See in the years leading up to the Gallican crisis (1534, 1535-6, 1547-9). His humanist preoccupations are clearly evident in the serious works he published in Latin and Greek in 1532. These include an annotated edition and Latin translation of medical treatises by Hippocrates and Galen, an edition of documents he took to be a will and a contract of sale dating from Roman Republican times, and an edition of a contemporary treatise on the topography of Ancient Rome. Equally revealing are several parodic almanacs Rabelais published in the vernacular, most notably the *Pantagrueline prognostication* (1533), in which astrology and prognostication generally are humorously debunked in the name of Christian scepticism.

Rabelais stated his intellectual loyalties explicitly in admiring letters to humanists like Guillaume Budé (1521) and Erasmus (1532). To the latter in particular he declared, addressing the aged humanist as his spiritual father and mother, that: 'whatever I am, whatever I am good for, it is to you alone that I owe it'. Rabelais' works in the vernacular give considerable substance to this extravagant claim, championing as they do Erasmian causes like the restoration of classical learning and the return to a non-doctrinal, eirenic, paleo-Christian church.

2 Early Works: *Pantagruel* and *Gargantua*

Rabelais' first two books, *Pantagruel* (1532) and *Gargantua* (*c.*1534), were published anonymously in Lyons under the anagrammatic pseudonym Alcofrybas Nasier. These vigorous mock epics relate the birth, education and 'horrific' exploits of two giant heroes drawn from popular tradition. Masquerading as penny chapbooks to be sold at the commercial fairs of Lyons, they adopt the format and superficial trappings of medieval popular literature in order to satirize all things medieval and represent the triumph of an enlightened Christian humanism over gothic darkness.

The principal focus of *Pantagruel* is political. Set in Utopia, the political non-place invented a mere sixteen years earlier by Thomas More (see Utopianism §1), and visibly inspired by pacificist writings like *Institutio principis christiani (Education of a Christian Prince)* (1516) by Erasmus (§3), *Pantagruel* narrates the failure of an imperialistic war of aggression. When Gargantua's realm is invaded by an usurping tyrant named Anarche (meaning 'non-rule') the eponymous hero repels the invaders, and inaugurates a new Golden Age of peace and harmony resembling Christ's kingdom of heaven (Matthew 20: 25-7 and 23: 12). He does this by abolishing the hierarchical political and religious institutions of the old order (empire and church). Pantagruel is prepared for this messianic role by a humanistic education. After a fruitless tour of the gothic universities he rejects the likes of Duns Scotus and Ockham and incorporates instead the entire curriculum of newly restored humanist disciplines,

described by Gargantua in a justly famous letter to his son ([1532-52] 1955: ch.8).

Appearing two years later, *Gargantua* rehearses the same basic plot as *Pantagruel* but in a more realistic and satirical mode. The unjust invader, Picrochole (which means ‘bitter bile’), resembles the Holy Roman Emperor Charles V in his incursions against Francis I. The scholastic old guard is represented by the ‘sophists’ of the Sorbonne, whose syllogistic illogic and wretched Latin are brilliantly parodied in a muddled speech by the unforgettable Master Janotus Bragmardo (see Major, J. §4). To these familiar themes *Gargantua* adds a sustained attack against monasticism, focusing on what Erasmus called the *constitutiones humanae* of the modern church (that is, traditions and practices unauthorized by Scripture). The most sympathetic character of the book is an anti-monk named Frère Jean des Entommeures, whose febrile activity and good-natured spontaneity are favourably contrasted with the monastic ideal of a contemplative life governed by rules and regulations. The book concludes with the founding of a utopian anti-monastery, the Abbaye de Thélème, in which the Benedictine Rule is abolished in favour of the Pauline principle of Christian freedom, expressed in the single non-rule, ‘Fay ce que vouldras’ (‘Do as you wish’).

3 Late works: *Tiers Livre* and *Quart Livre*

Many years later Rabelais wrote a sequel to *Pantagruel*, the *Tiers Livre (Third Book of Pantagruel)* (1546), followed by a sequel to the sequel, the *Quart Livre (Fourth Book of Pantagruel)* (1552). Published in Paris under the author’s own name, printed in beautiful Roman type and humanistic format, these more mature works are of a very different character. Massive humanistic learning overwhelms attenuated vestiges of popular humour and comic effects depend less on the reader’s recognition of well-known biblical and Virgilian tags, as before, than on a good first-hand knowledge of works of moral philosophy by Plutarch and Lucian, Cicero and Seneca, as well as the Hebrew Bible, the Greek New Testament, Pliny’s *Natural History* and the Justinian *Digest*. The sequels relate a pseudo-Socratic quest undertaken by Pantagruel’s foolish companion, Panurge. In both books the quest ends in failure and appears to serve chiefly as a loose structuring device for an open-ended series of episodes in which political, theological and philosophical ideologies and institutions are brilliantly satirized.

The focal character of the *Tiers Livre*, Panurge, is a sophistical egotist first seen justifying his profligacy by praising debts as the *anima mundi* in terms borrowed directly from Ficino’s *De amore (On Love)* (1469) (see Ficino, M. §4). Suspended between the desire to take a wife and the fear of being cuckolded, he tries in vain throughout the book to decide whether or not to marry. Whereas the Socratic Pantagruel would have him do this simply by knowing his own will, the gnostic Panurge insists that he must first know things above and beyond himself, believing that such knowledge will allow him to eliminate future contingents. His misguided quest leads him to consult burlesque sources of divination as well as experts in various fields of learning, none of which provides the certainty he seeks. In the lengthy discussions surrounding many of these consultations Pantagruel speaks the language of Neoplatonism in such a way as to mock the Hermetic tendencies of his wayward companion. Panurge’s thirteen consultants include a sinister astrologer, Her Trippa, in whom many have seen a caricature of Henricus Cornelius Agrippa and an exasperatingly elusive philosopher, Trouillogan, in whom Rabelais ridicules ‘ephectic’ Pyrrhonians a full decade and a half before the works of Sextus Empiricus first appeared in print.

The *Quart Livre* prolongs the quest of the *Tiers Livre* in a fantastical sea voyage to Cathay where Panurge hopes to consult the oracle of Bacbuc, the Holy Bottle. But Panurge’s personal dilemma is soon displaced by the political and religious crises of contemporary Christendom as the principal focus of the continuation. Modelled roughly on Lucian’s *True History* and profoundly influenced by Plutarch’s *On the Obsolescence of Oracles*, this bitter odyssey narrates Pantagruel’s encounters with the inhabitants of fourteen islands, which satirize the social and political institutions of modern Europe. The harshest episodes condemn the Council of Trent and the doctrines and practices it was even then legislating as dogma. The un-Christian fanaticism and intransigence of both Catholics and Protestants is a recurring theme.

The *Quart Livre* ends even more abruptly than the *Tiers Livre*, far from its ostensible goal. This deliberately non-teleological structure was so unsettling that it provoked an anonymous editor to publish an apocryphal, posthumous *Cinquiesme Livre (Fifth Book of Pantagruel)* (1564), which concludes with the Holy Bottle and the oracular word (‘Trinch’) vainly sought by Panurge in the *Quart Livre*. Although this final revelation completes Panurge’s personal quest it adds nothing to the meaning of the preceding books and seriously detracts from their

formal and ideological coherence.

See also: Erasmus, D.; Humanism, Renaissance §7; Political philosophy, history of §5-6; Scepticism, Renaissance §4

EDWIN M. DUVAL

List of works

Rabelais, F. (*c*.1483-1553) *The Complete Works of François Rabelais*, trans. D. Frame, Berkeley and Los Angeles, CA: University of California Press, 1991.(The only English translation that includes minor works and letters, all helpfully annotated by the translator and R.C. La Charité. The letters to Budé (1521) and Erasmus (1532) and the *Pantagrueline prognostication* (1533) appear in this edition, as well as the five books of Gargantua and Pantagruel.)

Rabelais, F. (*c*.1483-1553) *Oeuvres complètes*, ed. M. Huchon, Bibliothèque de la Pléiade, Paris: Gallimard, 1994.(The complete works in French and Latin, superbly edited and annotated by the leading expert on Rabelais' language and the evolution of his text. Destined to supersede the standard but incomplete edition begun by Abel Lefranc in 1913. Contains a useful chronology and a full bibliography of recent books on Rabelais.)

Rabelais, F. (1532-52) *Gargantua and Pantagruel*, trans. J.M. Cohen, London and New York: Penguin, 1955. (The most accessible and faithful of modern English versions, unfortunately virtually unannotated. Like all editions includes the posthumous *Cinquiesme Livre* (1564) along with the authentic four books, *Pantagruel* (1532), *Gargantua* (*c*.1534), *Tiers Livre* (1546) and *Quart Livre* (1552).)

References and further reading

Cooper, R. (1991) *Rabelais et l'Italie*, Études Rabelaisiennes, Geneva: Droz.(Presents most extant documents pertaining to Rabelais' personal and civic life, expertly edited. Introduction helps to fill some gaps in our sketchy knowledge of Rabelais' life and partially fills the need for an up-to-date biography.)

Defaux, G. (1973) *Pantagruel et les sophistes: Contribution à l'histoire de l'humanisme chrétien au XVIe siècle*, The Hague: Nijhoff.(An extremely erudite analysis of the conflict between scholasticism and Christian scepticism in Rabelais' works, especially *Pantagruel*. An important, rewarding study, but hard reading for those with little Latin.)

Desrosiers-Bonin, D. (1992) *Rabelais et l'humanisme civil*, Études Rabelaisiennes, Geneva: Droz.(Insightful, thorough thematic treatment of the ethical and political dimensions of Rabelais' works. Includes a complete bibliography of modern Rabelais scholarship up to 1990.)

Duval, E.M. (1991) *The Design of Rabelais's Pantagruel*, New Haven, CT and London: Yale University Press. (Demonstrates the formal and ideological coherence of Rabelais' first book as they are revealed through an overarching epic and messianic pattern.)

Duval, E.M. (1997) *The Design of Rabelais's Tiers Livre de Pantagruel*, Études Rabelaisiennes, Geneva: Droz. (Discovers the coherence and intention of Rabelais' most ambiguous book in the hero's ironic design for his companion and in the author's corresponding design for his reader, both of which turn on the oracular 'Know thyself' uttered by Panurge at the centre of a symmetrically-ordered work.)

Duval, E.M. (1998) *The Design of Rabelais's Quart Livre de Pantagruel*, Études Rabelaisiennes, Geneva: Droz. (Examines the religious basis for Rabelais' critique of the ideologies and institutions of sixteenth-century Europe, and the nature of Pantagruel's ethical stance in confronting them.)

Febvre, L. (1982) *The Problem of Unbelief in the Sixteenth Century: The Religion of Rabelais*, trans. B. Gottlieb, Cambridge, MA: Harvard University Press.(Translation of a book by the great Annales School historian, first published in 1942 to refute the then current view that Rabelais was a rationalist and an atheist. A milestone in Rabelais scholarship still useful for its analysis of Rabelais' intellectual and religious culture and for background on humanist circles in Lyons.)

Screech, M.A. (1959) *L'Évangélisme de Rabelais: Aspects de la satire religieuse au seizième siècle*, trans. by the author as *Rabelais and the Challenge of the Gospel: Evangelism, Reformation, Dissent*, Baden-Baden: Koerner, 1992.(Treats theological questions like the nature of faith, the efficacy of works, predestination and freedom of the will, and so on, according to Rabelais. Occasionally mistakes the intention of the text by reading it as a theological treatise.)

Screech, M.A. (1979) *Rabelais*, London, Duckworth and Ithaca, NY: Cornell University Press. (A useful, accessible episode-by-episode commentary on all of Rabelais' literary works, distilling the fruits of a lifetime of scholarship by the most learned of modern Rabelais scholars.)

Weinberg, F.M. (1972) *The Wine and the Will: Rabelais's Bacchic Christianity*, Detroit, MI: Wayne State University Press.(Analyses the ubiquitous wine motif in terms of a syncretistic fusing of Bacchic mysteries and Christian eucharist.)

Race, theories of

The first theories of race were attempts to explain why the peoples of Europe (or sometimes particular peoples within Europe) had developed a higher civilization than the peoples of other regions. They attributed inequality in development to different biological inheritance, undervaluing the importance of the learning process. Between the world wars social scientists demonstrated how many apparently natural differences, and attitudes towards other groups, were not inherited but learned behaviour. They asked instead why people should entertain false ideas about members of other groups. As the twentieth century comes to an end, it is claimed on the one hand that processes of racial group formation can be explained in the same terms as those used for explaining group phenomena in general. On the other hand it is maintained that the only possible theories are those explaining why, in particular societies and at particular times, racism assumes a given form.

1 Unequal development

The word race entered west European languages in the sixteenth century to designate a set of humans of common ancestry. The end of the eighteenth century saw the first attempts to utilize it as a scientific concept. These ultimately failed, but much confusion was generated by the lack of clarity in its scientific use and by the interaction between popular and scientific usage. The French anatomist Cuvier in 1817 divided *Homo sapiens* into three subspecies (or varieties) which he called races: they were Caucasians, Mongolians and Ethiopians. He maintained that differences in human physique produced differences in culture and mental quality. This explained why whites had gained dominion over the world and made the most rapid progress in the sciences. ‘Yellows’ were less advanced, and blacks degraded. So race was used as a taxon, but if it meant the same as subspecies it was redundant. It was also employed as an *explanans* of unequal development. At much the same time the idiom of race was popularized in historical novels, notably in Sir Walter Scott’s description of Normans and Saxons as races in his best-selling *Ivanhoe* (1820). Racial designations were used as a way of identifying group differences within Europe that was later extended to describe differences between Europeans and the peoples of other regions colonized by Europeans.

Any understanding of the causes of variation within the human species in the pre-Darwinian era was bedevilled by the near-impossibility of explaining how that variation could have come about if, as was generally believed, humanity had descended from Adam and Eve in the space of some six thousand years. One answer was the claim that humanity consisted of a limited number of permanent racial types (as argued by Charles Hamilton Smith in England, Robert Knox in Scotland, Arthur de Gobineau in France, J.C. Nott and G.R. Gliddon in the USA, Karl Vogt in Germany). This doctrine has often been called scientific racism. Its exponents tended to equate race with species and claimed that it constituted a scientific explanation of human history.

Darwin’s theory of natural selection was applied to human affairs by the eugenists and by writers such as Sir Arthur Keith. Their arguments in certain respects prefigured modern sociobiology. They presented races as units in the course of evolution which developed their distinctiveness by in-breeding. Racial prejudice was seen as an inherited character that helped them do this. This has also been seen as scientific racism, but whereas the typological theory looked to the past in assuming that there had once been pure races, the selectionist theory looked to the future in presenting ethnic and national groups as prospective species.

The further development of Darwinian theory put an end to the notion of race as an *explanans* of human history. If one publication marks the transition, it is *We Europeans* (1935) by Sir Julian Huxley and A.C. Haddon. This separated human physical variation, as an *explanandum* to be explained by population genetics, from the variation between national groups (now named ethnic groups) to be addressed by historians and social scientists (see Evolutionary theory and social science).

2 Early social science

Not until the 1920s was there evidence that racial prejudice was a form of learned rather than inherited behaviour. How then were the physical differences that people in western Europe and North America called ‘racial’ used to construct social categories?

The ecological theory developed by Robert E. Park in Chicago maintained that migration brought distinctive

peoples into contact; competition made them conscious of what distinguished them, and those in a superior status developed prejudice as a defensive reaction.

The Freudian theory maintained that social life built up frustration which individuals released on to scapegoats in the form of displaced aggression; this added to the direct aggression that sprang from the conflict of interests.

The first empirical studies of the relations between blacks and whites in the USA of the 1930s were cast in a structural-functionalist framework. US society was said to be founded on distinctive values that stressed the social equality of citizens. Racial discrimination conflicted with these values and generated an ‘American dilemma’.

This orthodoxy was attacked most vigorously in *Caste, Class and Race* (1948) by Oliver Cromwell Cox, a black Trinidad-born sociologist. He contended that as European capitalism expanded into territories where natural resources were abundant, demand increased for labour, especially unfree labour. So beliefs justifying black subordination were built into the structure of capitalist societies, dividing white from black workers. Cox distinguished racism and anti-Semitism as distinct social forms. The former justified exploitation. The latter was directed to the conversion, expulsion or eradication of the alien group. That the two might share a common psychological component was irrelevant.

Many subsequent writers have similarly chosen to start from postulates about power relations within and between states. While conceding that in some societies the main conflicts derived from the relation of groups to the means of production, Leo Kuper maintained that there were states (like apartheid South Africa) in which political power influenced the relation to the means of production much more than any influence in the reverse direction. These were ‘plural societies’ in which racial conflict pre-empted or preceded class conflict.

3 Recent social science

Recent theoretical writing is at one in recognizing the political significance of race as a folk concept underlying group boundaries and in denying that this concept has any value for the explanation of why humans draw racial boundaries in particular ways. To simplify what can otherwise be a confusing picture, it can be helpful to focus upon the sharp division of opinion about what are the most important *explananda* and about the kind of *explanantia* to be sought. These are expressed in two ‘problematics’ (meaning by this sets of questions of interest to scholars who share views about the best ways to answer the questions). The racial discrimination problematic (a revised version of the frequently-criticized race relations problematic) holds that the attitudes and behaviour which in popular speech are called racial are to be explained in the same terms as other forms of attitude and behaviour. The racism problematic, by contrast, holds that there is a distinctive historical phenomenon known as racism that has to be seen for itself and related to the social processes that produce it. Underlying these contentions are two contrasting epistemologies; there is no agreement on the best names for these, though the writer has called them Kant-inspired and Hegel-inspired.

The first problematic is primarily concerned with the causes of racial discrimination as a feature of individual behaviour. It holds that discrimination, like crime, is a normal feature of every kind of society. Discrimination is not necessarily unlawful or immoral, and individuals are to be held morally responsible for any social distinctions they draw. Because it starts from the individual, this approach leads to a rational choice theory that borrows from neoclassical economics and the analysis of the means-end relationship in sociology. The name has led some into assuming that it is a theory of rational choices alone, whereas the theory takes rationality (or optimization) purely as a criterion against which to assess the consequences of choices of whatever motivation or character (see Rational choice theory). The attitudes and behaviour that are popularly called racial are then explained as the product of shared tastes or values which individuals seek to satisfy in situations of scarcity and constraint. Outward differences, like those which are called racial, may be used to identify persons as belonging to particular social categories. A person’s appearance may create an expectation that they will possess or lack some social quality. Through such expectations social inequalities in one generation may be transmitted to succeeding generations. This approach attempts, within a single conceptual framework, to analyse processes of group inclusion and exclusion based upon ethnic origin in all historical periods and geographical regions.

The contrasting problematic is primarily concerned with racism as a pathological feature of society that generates structural and ideological pressures upon individuals, thereby diminishing the degree of their personal responsibility for any discriminatory action. Its starting point is with the collectivity, confining the analysis of

group inclusion and exclusion to particular societies and their spheres of influence. The utility for research of the postulate that racism is manifested in changing forms is dependent upon the reliability of the method for identifying what lies behind the form. The political dimension of the problem was illustrated in 1975 when the UN General Assembly resolved that ‘Zionism is a form of racism and racial discrimination’ (a resolution revoked in 1991). Whether Zionism is such a form depends upon the definitions of Zionism and racism.

The approach from racial discrimination follows Kant’s view of the Copernican revolution in science. The observer does not seek to comprehend objects in their entirety as things in themselves, but, by standing apart from them, to develop concepts that will help explain selected features of the observations. Definitions have to be adapted to research methods (see Kant, I. §6). The approach from racism follows Hegel’s view of the observer as part of the world they study, but able to ascertain the principles underlying the development of that world and to reach a knowledge validated by history (see Hegel, G.W.F. §8). Definitions are sought that will express the true nature of objects. There is a fundamental difference of opinion about the kinds of knowledge that are possible and the task of the social scientist.

French sociologists have recently revived and reformulated one of Cox’s arguments. They contend that whereas racial ideology in Europe used to follow a logic of inferiorization and exploitation, it now follows a logic of differentiation directed to the expulsion or destruction of groups perceived as alien. The structures of industrial society based upon stratification and the welfare state have been weakened. In the West the old nationalism has lost its appeal so that people have been turning to regional, ethnic, religious and life-style identities. These conditions stimulate racist exclusions. In the UK and North America, as in France, it is now said to be necessary to distinguish a variety of racisms that are expressed in coded language. Crucially, they can produce a racist effect while denying that this effect is the result of racism. For example, certain arguments about defending a national way of life against the pressures of immigration have been described as coded expressions of racist sentiment. ‘Race’ is said to be a political construct and the sociologist’s choice of an *explanandum* a political decision. But there remains the problem of how to delimit the number of possible racisms. For example, the Catholic Church contends that *in vitro* fertilization and genetic manipulation could give rise to a new form of racism; others might argue that whether or not such possibilities are matters of concern, their origins are not those of racial ideology. Since racism is assumed to be a hidden essence, there is no sure way of establishing any cause-and-effect relationship.

Racial differentiation has been a subject for research in other social sciences. In social psychology there is now no distinctive branch of ‘racial’ theory. Racial distinctions are explained by reference to the principles used in accounting for other forms of group behaviour, such as attitudes, attributions, identity, and so on. Geographers have studied processes of racial segregation, including the spatial dimension to the emergence of new identities based upon racial or ethnic origin. Political scientists have studied the influence of racial identification upon voting behaviour and the ways in which racial problems and policies are handled by national and international political institutions. Since the collapse of Soviet control over eastern Europe many older ethnic conflicts have reappeared, so that for some fields of research theories of racial or ethnic tensions have to entail theories of nationalism.

See also: Anti-Semitism; Multiculturalism

MICHAEL BANTON

References and further reading

Banton, M. (1983) *Racial and Ethnic Competition*, Cambridge: Cambridge University Press; repr. Aldershot: Gregg Revivals, 1992.(Pages 15-31 for the development of ideas on race in the New World, 42-50 and 78-99 for early social science theories.)

Banton, M. (1987) *Racial Theories*, Cambridge: Cambridge University Press, 2nd edn, 1998.(A history, covering the change from biological to social theories.)

Bataille, P., Couper, K., Martucelli, D. and Peralva, A. (1994) *Racisme et xénophobie en Europe: une comparaison internationale (Racism and Xenophobia in Europe: an International Comparison)*, with intro. and conclusion by M. Wieviorka, Paris: La Découverte.(Includes sections on the UK, Belgium, Germany and Italy against a background of French experience.)

Cox, O.C. (1948) *Caste, Class and Race: A Study in Social Dynamics*, New York: Monthly Review Press.(A wide-ranging and pioneering study influenced by Marxist theory.)

Huxley, J.S. and Haddon, A.C. (1935) *We Europeans: A Survey of Racial Problems*, London: Jonathan Cape. (Important to the history of thought on the subject.)

Omi, M. and Winant, H. (1986) *Racial Formation in the United States*, New York: Routledge.(An influential US work that emphasizes political and legal relations.)

Miles, R. (1989) *Racism*, London: Routledge.(Attacks the conceptual inflation resulting from over-extensive use of 'racism' and relates it more closely to class relations.)

Rex, J. and Mason, D. (1986) *Theories of Race and Ethnic Relations*, Cambridge: Cambridge University Press.(A collection of essays about the main approaches current in 1984.)

Scott, Sir W. (1820) *Ivanhoe*, Harmondsworth: Penguin, 1984.(One of the nineteenth-century's most widely-read novels.)

Solomos, J. and Back, L. (1994) 'Conceptualizing Racisms: Social Theory, Politics and Research', *Sociology* 28 (1): 143-61.(Reviews recent writing in the light of a claim that new forms of racism have emerged, requiring a revision of concepts.)

Radbruch, Gustav (1878-1949)

Gustav Radbruch is an emblematic figure in twentieth-century German legal philosophy and legal science. His particular blending of legal philosopher, dogmatist and politician, and his personal history, interwoven with the tragedy of the Weimar Republic and the rebirth of a democratic Germany after the Nazi horror, have given him special prestige and influence on both constitutional and ordinary jurisprudence in Germany. Some of Radbruch's theses, like the one in his well-known article 'Gesetzliches Unrecht und übergesetzliches Recht' (1946 - translatable as 'Statutory Non-law and Suprastatutory Law'), remain highly topical in German universities and courts.

Radbruch was born in Lübeck to a family of the mercantile bourgeoisie. He completed high school there, and entered the Munich law faculty in 1898, later studying also in Leipzig and Berlin. In Berlin he was a pupil of Franz von Liszt, probably the most influential scholar of criminal law in the new unitary German state, who upheld the line of liberal thought inaugurated by P.A. Feuerbach. Under Liszt's guidance, Radbruch took his doctorate in 1902. The following year, at Heidelberg University, Radbruch secured his *Habilitation* with a study on the relevance of the concept of action to the doctrine of criminal law. He filled various teaching posts between 1906 and 1914, took part in the First World War, and in 1919 secured a chair in criminal law at the University of Kiel.

In 1920 he was elected to the Reichstag, the German parliament, as a socialist member, and in 1921 became Minister of Justice in the Wirth cabinet, subsequently serving again as a minister in the first and second Stresemann cabinets. In November 1923, after Stresemann's withdrawal, he went back to Kiel University, where he stayed till autumn 1926 when he was called to a chair at Heidelberg University. Shortly after Hitler came to power, in May 1933, he was among the first university teachers to be removed from their posts on political grounds. In the twelve years of the Nazi regime, Radbruch retreated to the so-called *innere Emigration*, condemned to silence in his own country even though he did not leave it, devoting himself to historical and humanistic studies and publishing abroad. In September 1945, after the Nazis' defeat, he was appointed Dean of Law at Heidelberg, returning to teaching with renewed enthusiasm.

Radbruch's oeuvre is bracketed by two works, the book *Grundzüge der Rechtsphilosophie* (1914, with a definitive revised edition in 1932) and the 'Gesetzliches Unrecht und übergesetzliches Recht' ('Statutory Non-law and Suprastatutory Law') of 1946 (see Radbruch 1973). This article overturns the main thesis of the *Grundzüge*, namely a radically noncognitivist meta-ethical position, denying claims to knowledge or to objective rightness in moral values (see Moral judgment §1). For such an ethical relativism, to which all possible moralities seem acceptable, positive law is independent of moral values. The point is then to arrive at a formula or a conception of positive law not committed a priori to any ethical position, but permitting all of them. Radbruch considered this possible only if a positivistic conception of law is assumed. Here, then, as is also the case *mutatis mutandis* in Hans Kelsen's work, the endpoint of ethical relativism is legal positivism (see Kelsen, H.; Legal positivism §§1, 2).

Noncognitivism is also used by Radbruch to found or justify democracy: if no conception of any faction is capable of proof, then every conception should be fought from the viewpoint of an opposite conception; but since none of them can be refuted, each of them ought accordingly to be respected even within the viewpoint of the opposing conception (Radbruch 1934). Although he maintained that the concept of law 'is a cultural concept, that is, the concept of a reality oriented to values, which has the object of serving a value' (Radbruch 1973: 29), for Radbruch, at least until 1933, a law whose promulgation is formally and procedurally correct is valid and binding even where its content is morally reprehensible. This conclusion has, however, also a moral basis, grounded in the idea of legal certainty: 'However unjust the law may become in its content, it always, from its mere existence, meets one aim: that of legal certainty' (Radbruch 1914: 183).

Experience of the Nazi horror changed Radbruch's view radically. First and foremost, he accused legal positivism, understood more as a psychological attitude than as a legal philosophical doctrine, of having prepared, aided, or failed to obstruct the affirmation of Nazism. Jurists, accustomed to the motto '*Gesetz ist Gesetz*', that is, to regard any duly enacted statute as valid and binding law, however it be promulgated, were incapable of showing resistance to Hitler's injustice and despotism. There must therefore be a change of attitude by jurists, starting from,

or at least not neglecting, requirements of substantive justice, so that - this is the 'Radbruch formula' still applied by the German Constitutional Court - where a formally valid law proves intolerably unjust, it is no longer to be regarded as valid or binding. The 'Radbruch formula' breaks down into three parts: (1) first, it is asserted that in general positive law prevails over substantive requirements of justice, even when, while formally valid, it is unjust and irrational; (2) this primacy stops at breaches of principles of justice that are of intolerable proportions; (3) intolerable breaches of justice exist where the positive law explicitly and systematically does not have as its goal the realization or pursuit of aims of justice, and especially where the principles of formal equality (the equal treatment of equal cases) are deliberately and constantly ignored (see Justice, equity and law §1; Law and morality §1).

In German culture between the nineteenth and twentieth centuries, in particular in the stormy years of the Weimar Republic, Radbruch represents the enlightened, democratic tradition, all too rare in those times among German philosophers and jurists, often seduced by irrationalist, communitarian and authoritarian slogans. It is the Germany of Kant and of Anselm Feuerbach, the *Zivilisation* of reason and of rights, whose courageous interpreter he is. 'The expression of a spirit that has become profoundly sterile' was how he was contemptuously dismissed by the Nazi Karl Larenz. But in the end it is Radbruch's, not Larenz's, spirit that has proved fertile.

See also: Law, philosophy of; Legal positivism

MASSIMO LA TORRE

List of works

Radbruch, G. (1914) *Grundzüge der Rechtsphilosophie*, Leipzig: Duncker & Humblot.(Examines legal certainty.)

Radbruch, G. (1934) 'Le relativisme dans la philosophie du droit', *Archives de philosophie du droit*.

(1973) *Rechtsphilosophie*, Stuttgart: K.F. Koehler Verlag.(His principal work on legal philosophy in its final edition, with appendices added after the war. 'Anhang 4' is the essay 'Gesetzliches Unrecht und übergesetzliches Recht' (Statutory Non-law and Suprastatutory Law).)

(1980) *Legal Philosophy of Lask, Radbruch, and Dabin*, trans. H. Wilk, with intro. by E.W. Patterson, 20th Century Legal Philosophy, vol. 4, Cambridge, MA: Harvard University Press.(A translation of the 1936 edition of Radbruch's *Rechtsphilosophie* is to be found at pp. 45-224, capturing his thought before the later post-war revisions and adjustments.)

(1996a) *Der Mensch im Recht. Ausgewählte Vorträge und Aufsätze über Grundfragen des Rechts*, Göttingen: Vandenhoeck & Ruprecht, 3rd edn.(A selection of the author's essays and papers on fundamental questions about law.)

(1996b) *Einführung in die Rechtswissenschaft*, Stuttgart: K.F. Koehler Verlag, 12th edn.(An introductory text on jurisprudence and legal system, discussing the main branches of law.)

(1996c) *Paul Anselm Feuerbach, ein Juristenleben*, Göttingen: Vandenhoeck & Ruprecht, 3rd edn.(This life of Feuerbach gives valuable insights into Radbruch's approach and his sense of indebtedness to Kant and Feuerbach.)

References and further reading

Baratta, A. (1959) 'Relativismus und Naturrecht, im Denken Gustav Radbruchs', *Archiv für Rechts- und Sozialphilosophie* 45: 505.(A useful analysis of the tension between ethical relativism and natural law inclinations in Radbruch's work.)

Kaufmann, A. (1987) *Gustav Radbruch. Rechtsdenker, Philosoph, Sozialdemokrat*, Munich/Zurich: Pieper.(A valuable intellectual and philosophical biography of Radbruch by one of his leading successors as a critic of legal positivism in German jurisprudence.)

Stone, J. (1965) *Human Law and Human Justice*, London: Stevens and Sons.(A thoughtful account of Radbruch's work, especially on justice in relation to law, is to be found at pp. 231-57, with further useful references.)

Radhakrishnan, Sarvepalli (1888-1975)

As a modern interpreter of Indian thought to Western scholars and a major influence on later Indian thinkers, Sarvepalli Radhakrishnan's teaching, writing and worldwide lecturing introduced the West to Indian religion and philosophy as essentially an all-inclusive monism. He was actively committed to participation in Indian and international society as an educator, statesman and leader of the Indian republic and defended Hinduism against Western critics as essentially tolerant, world-affirming, progressive and socially and politically conscious. Radhakrishnan's neo-Advaita *was based on one major strand of Indian monism with modifications based on assumptions from his education in Christian institutions. While maintaining that the Absolute is identical with one's true self, he emphasized the reality of the universe. The truly religious person, he argued, does not flee the world but withdraws to attain personal realization and returns to apply the insight thereby gained to better society.*

1 Life and influences

Sarvepalli Radhakrishnan's life was that of a philosopher, definer and apologist for Hinduism and an active educator, statesman and world figure. Born in 1888 in south India, he was educated at Christian missionary schools. He received his BA and MA in philosophy from Madras Christian College. He studied under the Christian missionary and ethicist A.G. Hogg (1856-1931) who was a student at Edinburgh University of A.S. Pringle-Pattison and was influenced by A. Ritschl's school of German theology. Trained as an ethicist, Radhakrishnan emerged with the ethical concern of his teachers, but argued that this concern was found in his own tradition, which he traced back to early Indian scriptures and called Vedānta. The major influence in his definition of his tradition was the writings of the modern Indian leader Swami Vivekananda (see Ramakrishna movement), whose response to critics and definition of Hinduism resolved a number of Radhakrishnan's existential dilemmas. Radhakrishnan chose it as his own definition and devoted his thinking to explicating it as a philosophy and proclaiming it in his life and writings.

After a number of university appointments in south India, Radhakrishnan was appointed in 1921 to the most prestigious philosophical position in India, the King George V Chair of Mental and Moral Science of Calcutta University. From 1936 to 1941 he was the Spalding Professor of Eastern Religions and Ethics at Oxford University. His many posts include chief executive officer at a number of Indian universities, president of the executive board of UNESCO, a member of independent India's constituent assembly, India's second ambassador to the Soviet Union, vice president of India (1952-62) and president of India (1962-7). He lectured throughout the world, retired in 1967 and died in April 1975, having written over three dozen books. These included commentaries on four Indian religious texts and over one hundred articles on philosophy and religion.

2 Metaphysics

In his two-volume *Indian Philosophy* (1923) Radhakrishnan surveys and evaluates the major Indian philosophical traditions, but from his early writings he clearly identifies his approach and evaluation of others with the tradition of Advaita Vedānta of the eighth-century thinker Śaōkara and the early Indian scriptures called the Upaniṣads which, he believes, show evidence of that tradition. Thus, reality is ultimately a unity in which the One is the indescribable, impersonal Absolute, Brahman (see Brahman; Monism, Indian). This One is also the true self, a state of pure consciousness which is a subject rather than an object to us. In traditional terms it is *saccidānanda*, being itself, pure consciousness and unlimited bliss. Knowledge of this One is possible through rational analysis of the empirical data, but certainty comes from an intuitive, immediate awareness of Brahman. Science gradually confirms what religions have intuitively understood. This monism does not negate the world, Radhakrishnan insisted, and Śaōkara has thus been gravely misunderstood when interpreted as teaching that the world is an illusion. When Śaōkara declared all but the unqualified Brahman to be *māyā*, he meant that the world's reality is not independent, for its cause is Brahman. Only Brahman is independently real and the world is one particular manifestation of Brahman, although the actual nature of the relationship between the world and Brahman is logically inconceivable. Radhakrishnan's concern is to affirm the reality of a place in which to live the ethical life and to embrace the universe itself as a cosmic process which has meaning and value. This position he calls 'idealism' (see Idealism).

Others have understood the Absolute to be a personal, divine being. This is not a false understanding, he asserted, but a reference to the impersonal Absolute in its relationship to the world. The gods of India are personifications and symbols of the Absolute and three major deities form the *trimūrti*, the three forms of the One Brahman, at times called the Hindu Trinity. Brahma is the Divine as creator, Vishnu as preserver and Shiva as destroyer.

3 Ethics

The goal of human existence, Radhakrishnan asserts, is not a flight from this world for mere personal salvation. Hinduism has been falsely accused of teaching this. Human beings find themselves in ignorance in the rounds of karma and rebirth, but karma is not a mechanical principle; it is an ethical one. It should not be understood fatalistically, as has often been the popular understanding in India. Although actions from past lives have conditioned the present life, humanity can overcome the conditioning through choosing to live the good life. There is no mercy or grace to be sought, only human responsibility. The goal of Hinduism and true religion is *sarvamukti*, the liberation of all. One may withdraw from life in the world for a period to attain an intuitive experience of the One. However, in the spirit of the Buddhist *bodhisattva*, who refuses to enter liberation until all beings are brought in together (see Duty and virtue, Indian conceptions of §3), the realized individual returns to the world, with the insights gained thereby, to provide a religious, moral and social influence which leads all to liberation. Since all human beings are one in spirit, sacrifice and a life of service follows. Radhakrishnan chooses the *Bhagavad Gītā* (second century BC) with its call to an active lifestyle, as his major support for an ethic of unselfish action.

He does not reject the class system in India, but argues that its correct understanding is as the ordering of society, which he understands to be more like an organism. The four classes are necessary for the life of society, but these divisions are not based on birth but on ethical personality types without one being any higher, purer or more important than another. Therefore, in Radhakrishnan's opinion, this is not the corrupted system of castes found in contemporary India. The intention of the system is to solve the problem of race in India by allowing separate divisions without extermination or subordination.

4 Inclusivism

Radhakrishnan argued for the future of religion against those who proclaimed its modern irrelevance. That future faith, he said, must be consistent with the discoveries and experiential approach of science, morally active so as to solve the world's social problems and inclusive of all the world's religious viewpoints so as to promote world unity. In his later writings he advocated this 'religion of the spirit' or 'spiritual religion' which he defined in terms similar to his earlier definitions of Vedānta and Hinduism.

Radhakrishnan made religious tolerance a major theme of his work, defining it not as merely bearing with others, but as a hierarchic, subordinating inclusivism. He affirmed the truth of religious positions in terms of a hierarchy. He believed in an impersonal Absolute, identical with the self to be the highest understanding of reality as it is and that theistic positions of the highest reality as a Divine being are subordinate doctrinal understandings. These should be accepted as expressions which are relative to the historical environment of the believer (see Religious pluralism; Toleration).

Radhakrishnan quoted widely from the literature of the world's religions to support his contention that at the heart of all religions is a common religious experience which is understood differently by the religions according to their particular cumulative traditions. Hinduism, he argued, understands this best and allows for a variety of doctrinal interpretations. However, the nondual experience of the impersonal Absolute as the true Self within is, in fact, the nature of that experience.

As spokesman for the Indian state, Radhakrishnan argued that this understanding of monistic idealism provided the basis for democracy. The individual's value is rooted in its essential identity with the Absolute. Applied to economics, Radhakrishnan spoke in general terms of the value of a socialism which is 'economic democracy'. Likewise, the correct understanding of political secularism was not the indifference or rejection of religion by the state, but a nonsectarianism which affirms the essence of all religions. That essence is the 'religion of the spirit', he affirmed. Radhakrishnan's general approach to political thought and his public role as a detached teacher while president of India, positioned him as the ideal public *rājaguru*, the traditional scholar-teacher of the king, for

India's first two prime ministers, Jawaharlal Nehru and Lal Bahadur Shastri.

See also: Political philosophy, Indian; Vedānta

ROBERT N. MINOR

List of works

Radhakrishnan, S. (1920) *The Reign of Religion in Contemporary Philosophy*, London: Macmillan.(An early argument defending monistic idealism against theism.)

Radhakrishnan, S. (1923) *Indian Philosophy*, London: Allen & Unwin; 2nd edn, 1927.(A two-volume critical analysis of major traditions from his monistic perspective.)

Radhakrishnan, S. (1927) *The Hindu View of Life*, London: Allen & Unwin.(Collected lectures defending Hinduism for the modern world.)

Radhakrishnan, S. (1929) *An Idealist View of Life*, London: Allen & Unwin.(Collected lectures which present his most systematic argument.)

Radhakrishnan, S. (1939) *Eastern Religions and Western Thought*, London: Oxford University Press.(Collected lectures analysing religion and philosophy comparing East and West.)

Radhakrishnan, S. (1947) *Religion and Society*, London: Allen & Unwin.(Lectures applying religious ideals to social reconstruction.)

Radhakrishnan, S. (1955) *Recovery of Faith*, New York: Harper.(A defence of the need for a new religion for the modern world.)

References and further reading

Arapura, J.G. (1966) *Radhakrishnan and Integral Experience*, New York: Asia Publishing House.(Philosophical analysis of Radhakrishnan's thought.)

Minor, R.N. (1987) *Radhakrishnan: A Religious Biography*, Albany, NY: State University of New York Press.(Introduces his thought by placing it in the context of his life with a complete bibliography of Radhakrishnan's writings and secondary literature.)

Schilpp, P.A. (1952) *The Philosophy of Sarvepalli Radhakrishnan*, New York: Tudor. (Collection of critical, scholarly essays on a variety of topics in Radhakrishnan's thought.)

Radical translation and radical interpretation

Radical translation *is the setting of a thought experiment conceived by W.V. Quine in the late 1950s. In that setting a linguist undertakes to translate into English some hitherto unknown language - one which is neither historically nor culturally linked to any known language. It is further supposed that the linguist has no access to bilinguals versed in the two languages, English and (what Quine called) 'Jungle'. Thus, the only empirical data the linguist has to go on in constructing a 'Jungle-to-English' translation manual are instances of the native speakers' behaviour in publicly recognizable circumstances. Reflecting upon the fragmentary nature of these data, Quine draws the following conclusions:*

(1) It is very likely that the theoretical sentences of 'Jungle' can be translated as wholes into English in incompatible yet equally acceptable ways. In other words, translation of theoretical sentences is indeterminate. On the assumption that a sentence and its translation share the same meaning, the import of indeterminacy of translation is indeterminacy of meaning: the meanings of theoretical sentences of natural languages are not fixed by empirical data. The fact is, the radical translator is bound to impose about as much meaning as they discover. This result (together with the dictum 'no entity without identity') undermines the idea that propositions are meanings of sentences.
(2) Neither the question of which 'Jungle' expressions are to count as terms nor the question of what object(s), if any, a 'Jungle' term refers to can be answered by appealing merely to the empirical data. In short, the empirical data do not fix reference.

The idea of radical interpretation *was developed by Donald Davidson in the 1960s and 1970s as a modification and extension of Quine's idea of radical translation. Quine is concerned with the extent to which empirical data determine the meanings of sentences of a natural language. In the setting of radical interpretation, Davidson is concerned with a different question, the question of what a person could know that would enable them to interpret another's language. For example, what could one know that would enable the interpretation of the German sentence 'Es regnet' as meaning that it is raining? The knowledge required for interpretation differs from the knowledge required for translation, for one could know that 'Es regnet' is translated as 'Il pleut' without knowing the meaning (the interpretation) of either sentence. Beginning with the knowledge that the native speaker holds certain sentences true when in certain publicly recognizable circumstances, Davidson's radical interpreter strives to understand the meanings of those sentences. Davidson argues that this scenario reveals that interpretation centres on one's having knowledge comparable to an empirically verified, finitely based, recursive specification of the truth-conditions for an infinity of sentences - a Tarski-like truth theory. Thus, Quine's radical translation and Davidson's radical interpretation should not be regarded as competitors, for although the methodologies employed in the two contexts are similar, the two contexts are designed to answer different questions. Moreover, interpretation is broader than translation; sentences that cannot be translated can still be interpreted.*

1 Radical translation

Suppose that children, in the normal course of events, acquire their native tongue (for example, English) by observing their parents and others talking in publicly observable circumstances. A corollary of this common sense supposition is that whatever there is to linguistic meaning can be manifested in behaviour in publicly observable circumstances. However, one might ask, just how far do empirical data go toward fixing the meanings of the sentences being spoken? Quine's thought experiment of radical translation is designed to answer this question.

In the context of radical translation, the field linguist sets out to construct a manual for translating the newly discovered language of 'Jungle' into English. Of course, the linguist has a prior knowledge of English, and will exploit that knowledge in constructing a 'Jungle-to-English' translation manual. Obviously, a child learning a first language has no prior language to exploit. So, there is this fundamental difference in psychological readiness to learn 'Jungle' between the linguist and the child of native-speakers of the language. However, this difference does not affect the central fact that the ultimate empirical data available to both linguist and child are the same: facts about behaviour. The poverty of these ultimate data is what the context of radical translation is meant to bring out.

2 Indeterminacy of translation

The primary reference for indeterminacy of translation is ‘Translation and Meaning’, the second chapter of Quine’s *Word and Object* (1960). According to that account, the linguist first compiles a list of phonemes of the native-speakers’ language (which Quine dubbed ‘Jungle’), and then they settle on some expressions used by native speakers for assent and dissent.

A rabbit scurries by; the native speaker utters ‘Gavagai’. Having noticed the rabbit, and that the native speaker noticed it too, and suspecting that the rabbit probably prompted the native speaker’s utterance of ‘Gavagai’, the linguist tentatively enters into their ‘Jungle-to-English’ translation manual that the native speaker’s ‘Gavagai’ might well be translated as ‘Lo, a rabbit’. On later occasions, some rabbited, some rabbitless, the linguist will volunteer ‘Gavagai?’ with the intention of eliciting the native speaker’s assent, dissent, or neither. In this way, the linguist can assess whatever inductive support there may be for their hypothesis that ‘Gavagai’ translates as ‘Lo, a rabbit’.

The distal stimulus, in this case a rabbit, serves as the criterion for the linguist’s inductive hypothesis that ‘Gavagai’ can be translated as ‘Lo, a rabbit’. However, what a sentence (even a one word sentence) and its translation share is the same meaning. In the case at hand, is the rabbit the meaning of ‘Gavagai’? No; that would be to confuse meaning with reference. What, then, is its meaning? Quine’s answer to this question (what he calls ‘stimulus meaning’) is given in terms of the proximal stimulus, namely, patterns of a subject’s activated nerve endings. Affording the native speaker a global pattern of non-verbal stimulation for a specific duration together with the linguist’s query ‘Gavagai?’ will elicit from the native speaker either assent, dissent, or neither. Patterns of stimulation which would prompt assent and patterns which would prompt dissent belong, respectively, to the affirmative stimulus meaning and negative stimulus meaning for that sentence and native speaker at that time. Stimulus meaning *simpliciter* is the ordered pair of the two. Patterns which would prompt neither assent nor dissent are irrelevant.

Given stimulus meaning, one might say that what ‘Gavagai’ and its translation ‘Lo, a rabbit’ share is the same stimulus meaning. However, this too would be wrong, for stimulus meaning (as just defined) is private: the native speaker’s nerves are distinct from the linguist’s nerves. It would be more accurate, but still problematic, to say that the native speaker’s stimulus meaning for ‘Gavagai’ and the linguist’s for ‘Lo, a rabbit’ are approximately the same. This is Quine’s approach in *Word and Object* (although he subsequently altered this account, as we shall see).

In the context of radical translation, it is stipulated that there are no bilinguals for the linguist to consult. So, the linguist must go about compiling the translation manual by observing the behaviour of native speakers, verbal and otherwise, and querying them with sentences for assent/dissent under various publicly observable circumstances. Thus, the criteria of translation remain keyed to distal stimuli, though the definition of stimulus meaning remains keyed to proximal stimuli. After all, one cannot plausibly expect the linguist to know anything about the native speakers’ or their own global patterns of activated nerve endings.

Given the behavioural limits of the empirical data of radical translation, Quine maintains in *Word and Object* that (1) observation sentences, like ‘Gavagai’, can be determinately translated; (2) truth functions (for example, ‘and’, ‘or’, ‘not’) can be determinately translated; (3) stimulus-analytic and stimulus-contradictory sentences can be identified, but not translated; and (4) questions regarding the intrasubjective stimulus synonymy of occasion sentences can be settled, if raised. (A stimulus-analytic sentence is one that a person would assent to after every stimulation: ‘There have been black dogs’, for example. A stimulus-contradictory sentence is one that a person would dissent from after every stimulation. The occasion sentences ‘Bachelor’ and ‘Unmarried man’ could count as intrasubjectively stimulus synonymous for a person, if they have the same stimulus meaning for him.)

However, if the linguist were to become bilingual (for example, if they forgot about translation and simply acquired ‘Jungle’ as a native speaker might, or nearly so), then according to Quine (1′) they could determinately translate all occasion sentences, not just observation sentences, and (4) above becomes superfluous. (Non-observational occasion sentences are those that require a new prompting stimulus each time they are queried, but lack the community-wide agreement characteristic of observation sentences, such as ‘There goes John’s old tutor’.) This leaves the linguist of *Word and Object* with determinate translations for (1′), (2) and (3).

However, in *Roots of Reference* (1974), Quine explains that the truth functions cannot be determinately translated

after all. One barrier concerns conjunction (‘and’): if each of a pair of sentences commands neither assent nor dissent, their conjunction can sometimes command dissent and sometimes neither dissent nor assent. If (to use Quine’s examples) ‘It’s a mouse’ and ‘It’s a chipmunk’ are neither affirmed nor denied, still their conjunction will be denied. Whereas if the components are ‘It’s a mouse’ and ‘It’s in the kitchen’ and neither is affirmed or denied, their conjunction will be neither affirmed nor denied. An analogous barrier stands in the way of determinately translating alternation (‘or’).

3 Indeterminacy of translation (cont.)

In discussing the translation of observation sentences and truth functions in Word and Object, Quine makes use of Neil Wilson’s principle of charity (see Wilson 1959). The central idea of the principle is that any proffered translation that construes the native speaker as holding some patently silly belief (for example, that the law of non-contradiction is false) is less likely than that the proffered translation is a bad one (see Charity, principle of §4). Thus, there is a sound methodological reason for translating a native speaker so as to construe them as holding true beliefs (true by the linguist’s own standards). However, Quine augments this principle with another: maximize psychological plausibility. This latter principle permits the linguist to translate a native speaker’s sentence by some patently false English sentence if, given the native speaker’s ‘outlandish’ rites and taboos or whatever, doing so is more plausible than translating the sentence in question by some true English sentence.

Putting the matter overly schematically, the next steps the linguist takes in constructing a manual of translation are as follows: (a) segmenting newly heard sentences uttered by native speakers into short recurrent parts (‘words’) which these sentences share with some of the observation sentences already translated; (b) segmenting further new sentences in ways that reflect the ‘words’ of the previous step, so that a lexicon and grammar begin to emerge; (c) generating new sentences in the native speaker’s language as recursion sets in; (d) continuing to fine tune the manual. Steps (a) to (d) make use of analytical hypotheses, hypotheses that go beyond the behavioural data: though analytical hypotheses should not contravene those data, neither are they supported by them.

In ‘Indeterminacy of translation again’ (1987), Quine articulates some constraints of a pragmatic nature which help to guide the linguist’s conjectures in this area, but clearly the linguist’s latitude for conjecture remains enormous. In fact, two linguists working independently of one another might well come up with equally successful ‘Jungle-to-English’ manuals which, despite being consistent with all the behavioural data, differ from one another in assigning to countless ‘Jungle’ sentences different English sentences as translations. Of these, countless different English sentences would not be interchangeable in English contexts. But which of these rival translation manuals is the correct one? There is no fact of the matter here; they are both correct. Such is the thesis of indeterminacy of translation.

Quine sums up the philosophical import of radical translation and of indeterminacy as follows:

> The point of my thought experiment in radical translations was philosophical: a critique of the uncritical notion of meanings and, therewith, of introspective semantics. I was concerned to expose its empirical limits. A sentence has a meaning, people thought, and another sentence is its translation if it has the same meaning. This, we see, will not do.
>
> (Quine 1987: 9)

The truth is more nearly the reverse: if some sentence is the translation of another, then it is proper to say that it has the same meaning. So, the idea of propositions as objectively valid translation relations, or as meanings of sentences, cannot be maintained. Furthermore:

> The critique of meaning leveled by my thesis of indeterminacy of translation is meant to clear away misconceptions, but the result is not nihilism. Translation remains, and is indispensable. Indeterminacy means not that there is no acceptable translation, but that there are many. A good manual of translation fits all checkpoints of verbal behavior, and what does not surface at any checkpoint can do no harm.
>
> (Quine 1987: 9)

So, the radical translator imposes about as much meaning as they discover.

Before surveying some criticisms of Quine’s indeterminacy thesis, we need to clear up an ancillary matter. In

Word and Object, Quine claims that the native speaker's stimulus meaning for 'Gavagai' and the linguist's stimulus meaning for 'Lo, a rabbit' are approximately the same. But Quine realized as early as 1960 that such intersubjective comparisons of stimulus meanings are problematic, for no two people's nerve nets are isomorphic. Consequently, in *Pursuit of Truth* (1990) Quine proposes to do without intersubjective comparisons of stimulus meaning. Empathy fills the gap. A rabbit scurries by; the native speaker utters 'Gavagai'. Empathizing with the native speaker, the linguist conjectures that were they in the native speaker's place they would have uttered 'Lo, a rabbit'. Thus, the linguist forms the tentative inductive hypothesis that 'Gavagai' translates as 'Lo, a rabbit'. Appealing to empathy instead of to approximate sameness of stimulus meaning forces Quine to reformulate his definition of 'observation sentence' (see Quine [1990] 1992: 43), but he does not drop the concept of stimulus meaning from his scientific semantics. Quine leaves to the neuropsychologists the task of accounting for the mechanisms underlying humans' ability to empathize.

4 Criticism

Indeterminacy of translation has been hotly debated by philosophers ever since the publication of *Word and Object*, and it is probably fair to say that more philosophers reject what they take to be the thesis than accept it. Some of the more widespread criticisms are: (a) Quine does not prove the thesis; (b) linguistics is underdetermined, but not indeterminate; (c) the thesis is unintelligible; and (d) the evidence for actual translation is not limited to that of radical translation. Let us briefly examine each of these criticisms in turn:

(a) *In 'Word and Object' Quine does not provide a deductive proof of his thesis of indeterminacy.* This is correct, but in addition to the more or less inductive argument presented there, Quine elsewhere offers two other arguments for indeterminacy. One of these, found in 'Epistemology naturalized' (1969a), takes Peirce's verificationism and Duhem's holism as premises supporting indeterminacy:

> If we recognize with Peirce that the meaning of a sentence turns purely on what would count as evidence for its truth, and if we recognize withDuhem that theoretical sentences have their evidence not as single sentences but only as larger blocks of theory, then the indeterminacy of translation of theoretical sentences is the natural conclusion.
>
> (Quine 1969a: 80-1)

A second argument is to be found in 'On the reasons for indeterminacy of translation' (Quine 1970: 179-81). There Quine argues for indeterminacy on the grounds of underdetermination of physical theory. He argues that translation of a foreign physicist's theory which is underdetermined by observable evidence will itself be underdetermined by translation of the theory's observation sentences. Thus, the translator will have to rely on analytical hypotheses, and different systems of analytical hypotheses can result in equally good, but incompatible, translations. Thus, the extent to which a physical theory is underdetermined by observation is a measure of the extent to which its translation is indeterminate. (But see Quine 1979: 66-7 for a reservation about this way of arguing for indeterminacy.)

(b) *It is true that scientific theory is underdetermined by all possible observations, and in so far as linguistic theory is a part of scientific theory it too is underdetermined, but there is no special indeterminacy that afflicts linguistics.* Quine responds to this criticism by pointing out that once one accepts some single physical theory from among its competitors as the whole truth about what there is, still, within that theory, indeterminacy of translation can occur. The reason is that all that matters to translation are the behavioural facts, and since behavioural facts do not fix translation, no further physical facts, not even the whole truth about nature, would do so.

(c) *The thesis of indeterminacy is unintelligible, for it repudiates meanings and at the same time claims that there can be rival manuals of translation assigning to some 'Jungle' sentence English sentences which diverge in meaning. But if there are no meanings, then how can English sentences diverge in meaning? And if they can diverge in meaning, then how can there be no meanings?* The response to this criticism is to explain how two English sentences can diverge in meaning without reifying meanings. Quine responds as follows:

> A manual of Jungle-to-English translation constitutes a recursive, or inductive, definition of a *translation relation* together with a claim that it correlates sentences compatibly with the behavior of all concerned. The thesis of indeterminacy of translation is that these claims on the part of two manuals might both be true and yet the two translation relations might not be usable in alternation, from sentence to sentence, without issuing in

> incoherent sequences. Or, to put it another way, the English sentences prescribed as translation of a given Jungle sentence by two rival manuals might not be interchangeable in English contexts.
>
> (Quine [1990] 1992: 48)

Once again, Quine's solution to the difficulty is along behaviourist lines.

(d) *There may well be indeterminacy of translation in the context of radical translation, but in the context of actual translation there is more evidence available to fix the meanings of sentences. For example, it is a fact of established usage, that is, verbal dispositions, that we translate the French sentence 'Il y a peut-être des êtres intelligents sur Mars' as the English sentence 'There may be intelligent beings on Mars'.* It is true that the context of radical translation was designed to expose the ultimate empirical data for translation, where the word 'ultimate' excludes verbal behaviour linked to systems of analytical hypotheses that over time have become customary. But the point remains, even for languages as close as French and English, that some uncustomary system of analytical hypotheses could produce different translation relations between French and English which are just as effective as the customary one. Indeed, as Quine points out, indeterminacy of translation begins with the translator's home language. English, for example, could be mapped into itself by using one 'Jungle-to-English' manual to translate English into 'Jungle', then using a rival 'Jungle-to-English' manual to translate 'Jungle' back into English.

The four criticisms dealt with above are by no means the only criticisms philosophers have levelled against Quine's thesis of indeterminacy of translation, but they are some of the more interesting and widespread ones.

5 Indeterminacy of reference

As we have seen, given the empirical data, the 'Jungle' observation sentence 'Gavagai' can be determinately translated by the English observation sentence 'Lo, a rabbit'. May we conclude from this that 'gavagai' is therefore a 'Jungle' term, and if so, that it can be determinately translated as 'rabbit'? According to Quine, the answer is 'no' on both counts. Termhood and reference are matters that can be settled only by appealing to analytical hypotheses. For example, assuming the linguist decides to construe 'gavagai' as a term, is it to be construed as an abstract singular term (for example 'rabbithood'), or as a concrete general term? If the latter, then 'undetached rabbit part' could serve just as well as 'rabbit' as a translation of 'gavagai'. After all, the same portions of space-time as are occupied by undetached rabbit parts are occupied by rabbits.

The only way to settle such questions is, according to Quine, by relying upon the 'Jungle' equivalents of English plural endings, pronouns, numerals, the 'is' of identity and its adaptations 'same' and 'other'. These constitute the cluster of interrelated grammatical particles and constructions with which the individuation of terms of divided reference in English is connected. Once a linguist has fixed these equivalences, they can put to the native speaker questions like 'Is this gavagai the same as that one?', 'Is this one gavagai or two?', and so on. Once the linguist is able to ask such questions, they can begin to determine whether to translate 'gavagai' as 'rabbit' or as 'undetached rabbit part'. However, before they can ask such questions, the linguist will have to formulate a system of analytical hypotheses in connection with some other 'Jungle' expressions, namely, those playing the role of that cluster of interrelated grammatical particles and constructions of English which govern reference. But that cluster is itself susceptible to indeterminacy of translation. Thus, if one overall system of analytical hypotheses works for translating some 'Jungle' expression into 'is the same as', perhaps some other workable but systematically different system would translate the same 'Jungle' expression as 'belongs with'. Therefore, when the linguist attempts to ask 'Is this gavagai the same as that?', they could unwittingly be asking 'Does this gavagai belong with that?'. So the native speaker's assent cannot be used to settle the reference of 'gavagai' absolutely. Consequently, even though this method of translation is the best the linguist can manage, it is not sufficient for settling absolutely the indeterminacy between translating 'gavagai' as 'rabbit' or as 'undetached rabbit part', and so on. The linguist can say only that 'gavagai' refers to rabbits relative to one manual, and to undetached rabbit parts relative to a rival manual. So long as these manuals are consistent with the behaviour of all concerned, there is no fact of the matter of what is 'really' referred to by 'gavagai'.

Quine has also argued for indeterminacy of reference independent of the indeterminacy of holophrastic translation. In what he calls his 'proxy function argument', Quine explains how:

> we might reinterpret every reference to a physical object arbitrarily as a reference rather to its cosmic complement, the rest of the physical universe. The old names and predicates would be introduced by ostension as usual, but it would be deferred ostension: pointing to what was not part of the intended object. Sensory associations would carry over similarly. The word ‘rabbit’ would now denote not each rabbit but the cosmic complement of each, and the predicate ‘furry’ would now denote not each furry thing but the cosmic complement of each. Saying that rabbits are furry would thus be reinterpreted as saying that complements-of-rabbits are complements-of-furry things, with ‘complements-of-rabbits’ and ‘complements-of-furry’ seen as atomic predicates. The two sentences are obviously equivalent.
>
> (Quine 1995: 71)

Quine goes on to explain that ‘cosmic complement of’ expresses what he calls a ‘proxy’ function: a one-to-one reinterpretation of objective reference. Such proxy functions leave the truth-values of the containing sentences undisturbed.

> It is a matter of reconstruing all terms and predicates as designating or denoting the proxies of what they had designated or denoted. A term that had designated an object *x* now designates the proxy of *x*, and a predicate that had denoted *x* now denotes the proxy of *x*. No big deal; we are proxying both sides of the predication, and it cancels out. We appreciated the triviality where the proxy was the cosmic complement.
>
> (Quine 1995: 72)

Thus, unlike the conjectural status Quine accords indeterminacy of translation of theoretical sentences as wholes, indeterminacy of reference has the apodictic certainty of a proof.

6 Criticism

As with indeterminacy of translation, indeterminacy of reference (or inscrutability of reference) has been vigorously criticized. The most widespread criticism is, perhaps, that Quine’s doctrine is inconsistent: it maintains that reference is and is not determinate. For example, it asserts that the native speaker’s ‘gavagai’ is indeterminate, since it can be adequately translated by ‘rabbithood’, by ‘undetached rabbit part’, by ‘rabbit’, and so on, while taking the reference of these English terms to be determinate. A full Quinian response to this criticism would take us far afield, but perhaps the following two quotations will suffice:

> To say that ‘gavagai’ denotes rabbits is to opt for a manual of translation in which ‘gavagai’ is translated as ‘rabbit’, instead of opting for any of the alternative manuals.
>
> And does the indeterminacy or relativity extend also somehow to the home language? In ‘Ontological relativity’ I said it did, for the home language can be translated into itself by permutations that depart materially from the mere identity transformation, as proxy functions bear out. But if we choose as our manual of translation the identity transformation, thus taking the home language at face value, the relativity is resolved. Reference is then explicated in disquotational paradigms analogous to Tarski’s truth paradigm…; thus ‘rabbit’ denotes rabbits, whatever *they* are, and ‘Boston’ designates Boston.
>
> (Quine [1990] 1992: 52; for his essay ‘Ontological relativity’, see Quine 1969b)

and:

> The very freedom vouchsafed us by the indeterminacy of reference allows us to *adopt* ostension as decisive for reference to observable concrete objects. We end up as we began, then, agreeing on the denotations of ‘rabbit’ after all: rabbits for all concerned. We may then merely differ on the deeper nature of rabbits: they are spatio temporal for some… sui generis for most. Adaptation of our usage must not, however, be allowed to obscure the lesson of proxy functions. Namely, a language-wide one-to-one reassignment of values to our variables has no effect on the truth of falsity of our statements.
>
> (Quine 1995: 75)

We turn, now, to a consideration of radical interpretation.

7 Radical interpretation

Mary utters ‘It’s raining’; Tom interprets her utterance to mean that it is raining. How is this possible? What could Tom know that would enable him to interpret the sentences of a natural language? Donald Davidson argues that we would have answers to these questions - a theory of meaning - if we had a theory meeting two requirements. First, the theory must provide an interpretation for a potential infinity of utterances of a speaker or group of speakers; second, the theory must be verifiable without assuming a detailed knowledge of the speaker’s propositional attitudes. The first requirement recognizes the holistic nature of linguistic understanding; the second requirement recognizes that the theory must not beg questions by assuming too much of what it is supposed to explain.

Davidson hypothesizes that a Tarski-like truth theory, suitably modified to apply to natural languages, just might meet these two conditions. To test his hypothesis, Davidson envisages a context of radical interpretation wherein an interpreter is faced with interpreting in one idiom talk in another. (For example, Kurt utters ‘Es regnet’ and Mary interprets his utterance to mean that it is raining.)

As Davidson points out in his essay ‘Radical interpretation’:

> The term ‘radical interpretation’ is meant to suggest strong kinship with Quine’s ‘radical translation’. Kinship is not identity, however, and ‘interpretation’ in place of ‘translation’ marks one of the differences: a greater emphasis on the explicitly semantical in the former.
>
> (Davidson 1984b: 126, n.1)

Or, as Quine has put the difference:

> Translation is not the field linguist’s goal. His goal is to command the native language and perhaps to teach it, whether for reasons of ethnography and philology or simply to implement fluent dialogue and successful negotiation with the natives. His undertaking, broader than translation, is *interpretation*. An untranslatable sentence, such as the one about neutrinos [‘Neutrinos lack mass.’] can still be interpreted, and that indeed is how we have learned it ourselves. For broadly semantic purposes, as Donald Davidson appreciates, interpretation is the thing. Translation is the narrower project, pertinent specifically to my concern over the fancied concepts of proposition and sameness of meaning.
>
> (Quine 1995: 80-1)

Thus, Quine’s radical translation and Davidson’s radical interpretation are devoted to different issues. Quine wants to discover the extent to which the empirical data determine meanings of sentences of a natural language; Davidson wants to discover what one could know that would enable one to understand the sentences of a natural language. Furthermore, the knowledge required for interpretation differs from the knowledge required for translation, for one could know that ‘Es regnet’ is translated as ‘Il pleut’ without knowing the interpretation (meaning) of either sentence. Davidson is fully aware of these differences, of course:

> The idea of a translation manual with appropriate empirical constraints as a device for studying problems in the philosophy of language is, of course, Quine’s. This idea inspired much of my thinking on the present subject, and my proposal is in important respects very close to Quine’s. Since Quine did not intend to answer the questions I have set, the claim that the method of translation is not adequate as a solution to the problem of radical interpretation is not a criticism of any doctrine of Quine’s.
>
> (Davidson 1984b: 129, n.3)

What, then, is the nature of the Tarski-like theory that Davidson believes might serve as a theory of meaning?

8 Truth theory

Alfred Tarski’s theory of truth is designed to apply only to artificial languages meeting certain formal requirements: extensional languages like first-order predicate logic with relations and identity. According to Davidson:

> What characterizes a theory of truth in Tarski’s style is that it entails, for every sentence *s* of the object language, a sentence of the form:
>
> *s* is true (in the object language) if and only if *p*.

> Instances of the form (which we shall call T-sentences) are obtained by replacing '*s*' by a canonical description of *s*, and '*p*' by a translation of *s*. The important undefined semantical notion in the theory is that of *satisfaction* which relates sentences, open or closed, to infinite sequences of objects, which may be taken to belong to the range of the variables of the object language. The axioms which are finite in number, are of two kinds: some give the conditions under which a sequence satisfies a complex sentence on the basis of the conditions of satisfaction of simpler sentences, others give the conditions under which the simplest (open) sentences are satisfied. Truth is defined for closed sentences in terms of the notion of satisfaction.
>
> (Davidson 1984b: 130-1)

Following Tarski's lead, Davidson's blueprint for a theory of meaning consists of axioms, rules and theorems. A finite set of axioms state satisfaction conditions for each of the object language's *semantical primitives* (basic terms and various other basic non-sentential expressions). An example might be: '*chien* applies to a thing if and only if it is a dog'. A finite set of rules allows proving T-sentences as theorems from the axioms. The resulting theorems are a potential infinity of T-sentences stating the truth-conditions for each object-language sentence. The proof of a T-sentence amounts to an analysis of how the truth or falsity of the object-language sentence depends on how it is composed from the semantical primitives. (These axioms, rules and T-sentences are, of course, stated in the language of the theory - the metalanguage.)

This general form of Davidson's theory of meaning reflects certain features of natural languages. First, since natural languages are learnable by finite beings, they must have a finite number of semantical primitives. Second, since natural languages consist of a potential infinity of sentences, the semantic features of such sentences must ultimately depend upon those of the semantical primitives. However, this does not imply that 'word' meaning is known somehow prior to sentence meaning. In fact, Davidson argues just the opposite: sentence meaning is known prior to word meaning. Words have whatever meanings they do as a result of their functioning in meaningful sentences. Third, like the axioms, the number of rules for proving T-sentences must be finite, and for the same reason. Finally, it is noteworthy that this Tarski-like theory of meaning has several obvious ontological advantages in that it makes no use of meanings as entities, nor of properties as denotatations of predicates, nor of states of affairs as denotatations of sentences.

Tarski's *Convention T* is that a theory of truth should entail T-sentences, in which descriptions of object-language sentences are linked to translations of those sentences in the metalanguage. For his own purposes, Davidson modifies Convention T in two fundamental ways. Tarski's Convention T relies on the notion of meaning (or translation) in order to explain truth. But meaning (or translation) is just the semantic notion that Davidson wants to explain, so he cannot, without begging questions, interpret Convention T in the manner of Tarski. Rather, Davidson construes Convention T as claiming that:

> an acceptable theory of truth must entail, for every sentence *s* of the object language, a sentence of the form: *s* is true if and only if *p*, where '*p*' is replaced by any sentence that is true if and only if *s* is.
>
> (Davidson 1984b: 134)

Thus, Davidson reverses Tarski's priorities: instead of assuming meaning in order to explain truth, he assumes truth in order to explain meaning.

Furthermore, since Davidson wants to apply the theory to natural languages (something which, as Davidson reports, Tarski explicitly denies can be done), he must allow for the fact that the object language contains a wide diversity of expressions, including proper names, functional expressions, indexicals, demonstratives, and so on. Admitting indexicals and demonstratives into the object language provides for the possibility of sentences which vary in truth-value according to time and speaker. Thus, Davidson's version of a Tarskian theory must take utterances of sentences, not sentences themselves, as truth vehicles. Davidson modifies Convention T accordingly:

s is true in *L* when spoken by *x* at *t* if and only if *p*.

Recall that Davidson believes we would have an answer to the question of what a person could know that would enable them to interpret the sentences of a natural language, if we had a theory meeting two requirements. First, the theory must provide an interpretation for a potential infinity of utterances of a speaker or group of speakers - a potential infinity of T-sentences. Davidson's modified version of Tarski's theory of truth is a blueprint for

constructing such a theory. Second, since that theory is to be an empirical theory about a speaker's (or speakers') behaviour, it must be verifiable - without assuming a detailed knowledge of the speaker's (or speakers') propositional attitudes. So, how is this second requirement to be met?

This is where the notion of a radical interpreter (like Quine's radical translator) serves as a useful heuristic. By imagining the routine of a radical interpreter at work, we come to appreciate that a theory of meaning is supported by verifying its T-sentences, and we come to see what ultimate evidence there is for verifying T-sentences.

9 Truth theory (cont.)

Imagine a radical interpreter who speaks only English, does not have access to bilinguals or dictionaries, and is attempting to interpret Kurt, who speaks only German. The interpreter has no prior detailed knowledge of the meanings of Kurt's utterances nor of Kurt's beliefs. Kurt utters 'Es regnet' at a time when it is raining near him. Judging that Kurt's utterance is both linguistic and intentional, the interpreter conjectures that Kurt holds 'Es regnet' to be true. The important methodological point to note, according to Davidson, is that the interpreter can spot such utterances purely on the basis of the speaker's observable behaviour, independently of knowing anything of the utterance's meaning or the speaker's beliefs. It is those utterances that a speaker (or speakers) hold true, at a time, under publicly observable circumstances, which constitute the ultimate evidence for T-sentences.

This ultimate evidence, about what sentences a speaker holds-true, is evidence not specifically about meaning, but about a combination of meaning and belief: a speaker believes whatever it is that they mean by the sentences. However, if the interpreter knew not only that Kurt holds a certain utterance true, but also what Kurt believes at the time, then the interpreter could extract the meaning of Kurt's utterance. Similarly, if the interpreter knew both that Kurt held a certain utterance true and its meaning, then they could extract Kurt's current belief. The problem is that there is no way to discover meaning which is independent of belief, and no way to discover belief which is independent of meaning. Thus, the central problem of interpretation is simultaneously to disentangle meaning and belief. How can a radical interpreter accomplish this task?

Suppose the interpreter, to use one of Davidson's examples (1984b: 135), can formulate the following T-sentence:

(T) 'Es regnet' is true-in-German when spoken by *x* at time *t* if and only if it is raining near *x* at *t*.

The interpreter's evidence for this T-sentence is:

(E) Kurt belongs to the German speech community and Kurt holds-true 'Es regnet' on Saturday at noon and it is raining near Kurt on Saturday at noon.

The interpreter's next step is to gather generalized evidence:

(GE) (*x*)(*t*) (if *x* belongs to the German speech community then (*x* holds-true 'Es regnet' at *t* if and only if it is raining near *x* at *t*)).

The interpreter can take (E) as evidence for (T) and for (GE), if they assume Kurt believes it is raining nearby at noon on Saturday, and if they assume that particular belief played a role in the causal chain prompting Kurt's utterance, 'Es regnet'. By using this procedure of holding belief constant as far as possible while solving for meaning, the interpreter can hope to solve the problem of the interdependence of belief and meaning.

But what justifies this procedure? After all, Kurt could be in error about it raining near him, and the interpreter could be in error about what distal stimulus is prompting Kurt's utterance and, hence, about the belief that is ascribed to Kurt.

'What justifies the procedure is the fact that disagreement and agreement alike are intelligible only against a background of massive agreement' (Davidson 1984b: 137). For example, if two people agree or disagree that what they see is a rabbit in tall grass, they must be in agreement regarding an indefinite number of other beliefs about rabbits, grass, distance, and so on. Furthermore:

> The methodological advice to interpret in a way that optimizes agreement should not be conceived as resting on a charitable assumption about human intelligence that might turn out to be false. If we cannot find a way to interpret the utterances and other behavior of a creature as revealing a set of beliefs largely consistent and true

> by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything.
>
> (Davidson 1984b: 137)

This mandatory principle of charity indicates that the verification of ascriptions of meaning and of belief are holistic in character. So while the principle of charity (namely, ‘optimize agreement’) does not guarantee that Kurt’s acceptance of ‘Es regnet’, or any other sentence considered in isolation, is to be deemed correct (in that it will be interpreted to mean something we accept), Davidson contends that the principle does guarantee that most of Kurt’s acceptances are in that sense correct.

> The method is rather one of getting a best fit. We want a theory that satisfies the formal constraints on a theory of truth, and that maximizes agreement, in the sense of making Kurt (and others) right, as far as we can tell, as often as possible.
>
> (Davidson 1984b: 136)

Unlike Quine, who advocates the principle of charity at the level of translating ‘Jungle’ observation sentences like ‘Gavagai’ and the ‘Jungle’ equivalents of truth functions like ‘and’, ‘or’ and ‘not’, Davidson advocates the principle at every level of interpretation (see Davidson 1984b: 136, n.16). However, his principle of charity definitely involves more than the basic methodological precept ‘optimize agreement’:

> It is impossible to simplify the considerations that are relevant [for dealing with disagreements], for everything we know or believe about the way evidence supports belief can be put to work in deciding where the theory can best allow error, and what errors are least destructive of understanding. The methodology of interpretation is, in this respect, nothing but epistemology seen in the mirror of meaning.
>
> (Davidson 1984c: 169)

It is Davidson’s contention that for a theory that satisfies the formal constraints on a theory of truth, and which optimally fits the evidence about utterances held true, each of its T-sentences will yield an acceptable interpretation. Moreover, Davidson believes - and for Quinian reasons - that if there is one such theory for some object language, then there are many. Thus, Davidson’s theory of meaning possesses its own forms of indeterminacy and inscrutability.

10 Criticism

Davidson’s theory of radical interpretation has precipitated a tremendous amount of discussion of the topic among philosophers. Two of the more general criticisms of Davidson’s programme are: (a) the construction of a Tarski-like theory of meaning for an entire natural language is implausible; and (b) Davidson’s rendering of Convention T is too weak to yield interpretations.

(a) *The construction of a Tarski-like theory of meaning for an entire natural language is seen to be implausible when one considers the number and variety of idioms whose truth-conditions are not fully understood; idioms like ‘knows that’ and ‘believes that’, contrary to fact conditionals, and so on.* Citing some of his own work on attributions of attitudes and performatives, as well as other philosophers’ work on proper names, mass terms, ‘ought’, and so on, Davidson remains optimistic that many of these problems will be solved in ways consistent with his Tarski-like theory of meaning (see Davidson 1984b: 132-3).

(b) *Davidson’s rendering of Convention T is too weak to yield interpretations.* Recall, for Davidson, Convention T says that *s* is true if and only if *p*, where ‘*p*’ is replaced by any sentence that is true if and only if *s* is. Hence, we get the following T-sentence:

(1) ‘John is a renate’ is true if and only if John is a renate.

But, since all renates are cordates, and vice versa, we also get the following T-sentence:

(2) ‘John is a renate’ is true if and only if John is a cordate.

But even if (1) is an interpretation of the object-language sentence ‘John is a renate’, surely (2) is not. So Convention T is too weak.

Davidson's response to this problem is two-fold: he emphasizes that T-sentences are verified not individually but holistically, and he recognizes that T-sentences (and axioms) must be viewed as empirical laws. In accord with the latter point, T-sentences which yield interpretations must not only be true, they must also be capable of supporting counterfactual claims. Presumably (1) supports the counterfactual:

(1′) If 'John is a renate' were false, then John would not be a renate.

But (2) does not support the counterfactual:

(2′) If 'John is a renate' were false, then John would not be a cordate.

Davidson would explain this difference, presumably, by claiming that (1) does, but (2) does not depend ultimately on certain causal relations between speakers and the world (see Davidson 1984a: xiv).

These criticisms of Davidson's programme are by no means the only ones; many questions have been raised about his handling of indexicals, ambiguity, word-meaning versus sentence-meaning, the principle of charity, and so on.

See also: Davidson, D.; Hermeneutics; Meaning and truth; Quine, W.V.; Reference

ROGER F. GIBSON

References and further reading

Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Clarendon Press.(A collection of Davidson's essays on the philosophy of action.)

Davidson, D. (1984a) *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.(A collection containing some of Davidson's key essays on radical interpretation.)

Davidson, D. (1984b) 'Radical interpretation', in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 125-39.(One of the most important of Davidson's essays for understanding radical interpretation.)

Davidson, D. (1984c) 'Thought and talk', in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 155-70.(One of the most important of Davidson's essays for understanding his views on the principle of charity.)

Føllesdal, D. (1988) 'Indeterminacy and mental states', in R. Barrett and R. Gibson (eds) *Perspectives on Quine*, Oxford: Blackwell, 98-109.(Føllesdal gives a clear statement of Quine's indeterminacy doctrine.)

Gibson, R. (1986) 'Translation, physics, and facts of the matter', in L. Hahn and P. Schilpp (eds) *The Philosophy of W. V. Quine*, La Salle, IL: Open Court Press, 139-54.(Gives a clear statement of the difference between indeterminacy of translation and underdetermination of physical theory; excellent bibliography of Quine's writings.)

LePore, E. (ed.) (1986) *Truth and Interpretation*, Oxford: Blackwell.(A collection of twenty-eight essays on Davidson's philosophy, including three by Davidson.)

LePore, E. and B. McLaughlin (eds) (1985) *Actions and Events*, Oxford: Blackwell.(A collection of thirty essays on Davidson's philosophy of action and psychology, including three by Davidson; excellent bibliography.)

Quine, W.V. (1959) 'Meaning and translation', in R.A. Brower (ed.) *On Translation*, Cambridge, MA: Harvard University Press, 148-71.(An earlier version of Chapter 2 of *Word and Object*.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(See Chapter 2, 'Translation and meaning', for discussion of radical translation.)

Quine, W.V. (1969a) 'Epistemology naturalized', in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 69-70.(Contains the argument for indeterminacy from holism and verificationism.)

Quine, W.V. (1969b) 'Ontological relativity', in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 26-68.(Quine explains inscrutability of reference.)

Quine, W.V. (1970) 'On the reasons for indeterminacy of translation', *The Journal of Philosophy* 67 (6): 178-83. (Clarifies and expands upon earlier material.)

Quine, W.V. (1974) *Roots of Reference*, La Salle, IL: Open Court Press.(See pages 75-8 for discussion of barriers to translating truth functions.)

Quine, W.V. (1979) 'Comment on Newton-Smith', *Analysis* 39 (3): 66-7. (Quine expresses a reservation regarding the argument from underdetermination to indeterminacy.)

Quine, W.V. (1987) 'Indeterminacy of translation again', *The Journal of Philosophy* 84 (1): 5-10.(An excellent exposition of Quine's reasons for indeterminacy of translation.)

Quine, W.V. (1990) *Pursuit of Truth*, Cambridge, MA: Harvard University Press; revised edn, 1992.(Excellent

discussions of indeterminacy of translation and indeterminacy of reference.)
Quine, W.V. (1995) *From Stimulus to Science*, Cambridge, MA: Harvard University Press.(Excellent discussions of Quine's naturalism, reification and objects of thought.)
Wilson, N. (1959) 'Substances without substrata', *Review of Metaphysics* 12: 521-39.(Originator of the principle of charity.)

Rahner, Karl (1904-84)

Rahner sought to offer an account of the Christian faith that would be credible to the modern mind. His early philosophical works lay the foundation for this theological project. Using both the method and categories of the early Heidegger, Rahner placed the thought of the medieval philosopher and theologian Thomas Aquinas in conversation with modern philosophy. He asked of Aquinas' epistemology Kant's question about the conditions of human subjectivity which make knowledge possible. Rahner argued that Aquinas' description of knowledge and human freedom requires, as its necessary condition, that the subject possess an openness to a universal horizon of being, an openness to God. There is, in the structure of subjectivity, a constitutive, experiential, a priori relationship with the divine mystery. While this openness occurs within an individual's self-awareness, it is always mediated by and interpreted through the objects, people, language and ideas that make up one's historical context (the categorical). In his theology, Rahner argued that the true nature of humanity's relationship with God had been revealed by Jesus to be one of absolute nearness. Rahner rendered Christian doctrines credible by correlating them with the transcendental experience of a God who is near.

1 Life

Karl Rahner was born near Freiburg, Germany, on 5 March 1904. He entered the Society of Jesus in April 1922 and, as a Jesuit seminarian, studied the Thomistic philosophy which then dominated Roman Catholic thought. In personal study he read the works of the Belgian Jesuit Joseph Maréchal.

In 1934, two years after ordination to the priesthood, Rahner was sent to Freiburg University to take a doctorate in philosophy. While he pursued his degree under the Thomist Martin Honecker, Rahner attended Martin Heidegger's lectures and seminars. He wrote a dissertation on Aquinas' epistemology which Honecker rejected because Rahner's interpretation of Aquinas was too much influenced by Heidegger. The dissertation was published in 1939 as *Geist in Welt* (*Spirit in the World* (1968)). Its perspective is the philosophical ground for Rahner's later theology. Rahner left Freiburg in 1936. In 1937, he satisfied the doctoral and post-doctoral requirements in theology at the University of Innsbruck and lectured there until the theology faculty was abolished by the Nazis in July 1938. He spent the war years at the Pastoral Institute in Vienna and later as a pastor in Bavaria. His philosophy of religion, *Hearers of the Word*, appeared in 1941.

After the Second World War, Rahner returned to teach theology at Innsbruck. He became one of the most influential and prolific Roman Catholic theologians. During the Second Vatican Council, he served as theological advisor to Cardinal König of Vienna. He also held positions at the Universities of Munich (1963-7) and Münster (1967-71). Most of Rahner's theological publications appear as essays collected in the sixteen volumes of *Schriften zur Theologie* (1954-84) (*Theological Investigations* (1961-92)). His *Grundkurs des Glaubens* (1976) (*Foundations of Christian Faith* (1978)) offers a systematic though incomplete presentation of his thought. Rahner also edited and contributed to theological dictionaries and published a number of significant works in spiritual theology. After retiring in 1971, he continued to lecture and write until his death on 30 March 1984.

2 Context and project

Rahner is primarily a Roman Catholic theologian. The aim and content of his thought emerge from the context of the Church's intellectual life in the mid-twentieth century. In 1879, Pope Leo XIII published the encyclical *Aeterni Patris*, which granted the thought of Thomas Aquinas a privileged status within the Catholic Church. Against the scepticism of modern thought, the philosophy of Aquinas represents a realism which both offers rational proof of God's existence and establishes the divine attributes necessary to ground historical revelation. In this way it illuminates Christian revelation and serves Catholic theology (see Thomism).

Leo XIII's encyclical led to a revival of Thomistic thought within the Church. By the middle of the twentieth century, this revival had produced an unanticipated pluralism in interpretations of Aquinas. Some saw Aquinas as an alternative to the errors of modern philosophy. Others, such as the Jesuits Pierre Rousselot and Joseph Maréchal, interpreted Aquinas in the light of modern questions and insights. In *Le point de départ de la métaphysique (The Starting Point of Metaphysics)* Maréchal began a dialogue between the epistemologies of Kant and Aquinas. By Rahner's own account, the basic and decisive philosophical direction of his thought came from

Maréchal.

The Neo-Kantianism that dominated German university philosophy at the beginning of the twentieth century assumed the correctness of Kant's critique of metaphysics. In his *Critique of Pure Reason* (1781), Kant argued that human knowledge is properly directed towards the objects of sensation. Efforts to know transcendent realities such as being, human freedom and God, while inevitable, are futile. The human capacity for knowledge is properly employed only when directed towards the physical world presented in sensation. Metaphysics and natural theology do not fall within reason's competence (see Kant §§4-8).

Pursuing Maréchal's project and influenced by Heidegger's efforts to retrieve the question of being, Rahner sought a Thomistic response to Kant's critique of metaphysics. Rather than eschewing modernity's turn to the subject as the abandonment of realism, Rahner offers a transcendental interpretation of Aquinas which leads to an epistemology, an ontology, and an argument for the existence of God. His thought not only offers a Thomistic response to modern philosophical questions, it also serves as the foundation for his theology. The notion of God which Rahner draws from Aquinas serves as an experiential reference in his existential interpretation of Catholic doctrine. In sum, Rahner's thought retrieves a Thomistic ontology within the context of modern thought and, in doing so, grounds a credible account of Christian faith.

3 Fundamental ontology

Rahner observes that Aquinas, like Kant, roots all knowledge in sensation. The proper object of human knowing is the physical world perceived by the senses. Rahner's rejected doctoral dissertation *Spirit in the World* is an interpretation of *Summa theologiae* Ia, question 84, article 7, in which Aquinas asserts that there is nothing in the human intellect apart from what is attained by turning to phantasms (that is, to objects sensed and imagined). Rahner's interpretation of Aquinas is not a literal rendering of Aquinas' historical position. Rather Rahner adopts Heidegger's interpretive method of retrieval. He uncovers possibilities in Aquinas' thought (in response to Kant) which could not have occurred to a medieval thinker. If human knowledge is properly directed towards objects of sensation, as Kant and Aquinas hold, how are metaphysics and natural theology possible?

Knowledge is more than sensation (after all, animals sense the same things as humans). Knowing is the work of what Aquinas calls the agent or active intellect. Attending to sense data, one forms a concept or idea, such as the concept of 'dog'. But having concepts is not yet knowing. Ideas are the means by which the human intellect comes to the intelligibility of sensed objects. One attains truth (that is, what is) in the act of judgment that, for example, this concrete object is a dog. Thus for Aquinas, knowledge consists of sensation, conceptualization (abstraction) and judgment (see Aquinas, T. §11).

Rahner then asks of this account of knowledge Kant's transcendental question. What are the conditions that make knowing possible? What must be true of the human intellect for it to move from sensation through conceptualization to judgment? Rahner thereby follows Kant's transcendental approach and attends to the a priori conditions of subjectivity for knowing. He is a transcendental Thomist in that he asks Kant's question of Aquinas' description of knowledge.

Using a term from Heidegger, he contends that the intellect must possess a *Vorgriff*, a pre-grasp of universal being as the necessary condition for the activities of forming universal concepts and judging. The *Vorgriff* is crucial to Rahner's metaphysics and philosophy of religion. For the intellect to form universal concepts from the particular objects of sensation it must have an a priori openness to universality. Again, for the intellect to judge that this object is a dog and nothing else requires a pre-grasp of the nothing else, of a universal horizon of all other beings. The *Vorgriff* of universality is not itself a concept or an object of direct experience. It is a dynamism towards all being, manifest and experienced in the activity of the intellect as the necessary condition for that activity.

What is this universality? In judgment, one asserts what a thing is against an infinite horizon of what it is not. Judgment requires, as its necessary condition, an anticipation of the absolute range of all knowable objects. The range of the intellect is unlimited. It is ordered towards knowing everything that is, all being. If judgment attains what is (being), the absolute range of universality must be absolute being. Absolute being is what a Thomist means by God.

Furthermore, in judgment one asserts what a thing is. One attains the intelligibility of an object (for example, 'that

is a dog'). Being, what is, and intelligibility coincide. Being and knowledge must constitute an original unity. Thus absolute being is pure intelligibility, a knower knowing itself, subject and object, absolute and pure self-presence. Pure being is a pure luminosity in which knower and known are identical. There is a hierarchy of being which is a hierarchy of self-possession, of intelligibility, of knowledge. God is one simple act of self-knowledge in which God knows all that is, all being. Humans, lower in the hierarchy, come to self-possession by knowing particular beings. Every act of knowledge involves an encounter with the other in sensation, in-the-world, and the return to self in judgment ('This is a dog and not I'). One possesses the known in the return to self. Human truth is an increase in being, in that the knower possesses more of what is. Knowledge is an existential occurrence in which the knower becomes.

4 Freedom

Human nature is not static. Rahner adopts the early Heidegger's notion of human existence as freedom-in-the-world. To be human is to be handed over to oneself to become. The capacity to make specific choices is rooted in this primordial being as becoming. Human freedom is finite in that persons become through their commerce with other people and objects in-the-world, and through the possibilities offered by a particular time and place. But the performance of freedom requires, as its necessary condition, the *Vorgriff* of unlimited being. Human freedom is unavoidable transcendence, a becoming without limit. To set limits to human becoming - to assert, for example, that human beings are creatures of their environment and no more - is already to transcend those limits. To know oneself is already to transcend, to stand beyond self. Like knowing, the performance of human freedom requires as its necessary condition the subject's openness to an unlimited horizon of being. Human freedom is exercised within the ever mysterious horizon of infinite being. One determines one's being before absolute mystery.

Thus Rahner argues that an experiential openness to absolute being (God) is the necessary condition for human knowledge and freedom. But God remains a mystery. In *Hearers of the Word*, Rahner states that the divine may be merely the mysterious horizon within which human transcendence occurs. It is also possible, though, for God to establish and reveal a more intimate relationship with human creatures.

5 Theology

Rahner's fundamental ontology is the enduring context for the theology which constitutes the bulk of his work. He distinguishes two poles of human experience - object and subject, the categorical and the transcendental. The things, events, possibilities and people one encounters are the categorical. All these encounters happen to a transcendental subject. The *Vorgriff* occurs in the transcendental. Locating the experiential reference for God in the transcendental determines the direction of Rahner's theology.

Speech about God, religious and philosophical, employs categories from historical contexts (for example, horizon, depth, father, creator) to point towards and interpret humanity's transcendental relationship with the divine. Rahner concludes his philosophy of religion described in *Hearers of the Word* with the suggestion that we search history to see if God has, in fact, revealed the character of humanity's constitutive relationship with the divine mystery.

The core truth of the Christian faith is that in Jesus, God has revealed the nature of humanity's relationship with the divine to be one of absolute nearness. God is offering God's very self to each human person for eternal communion. Borrowing a Heideggerian term, Rahner asserts that God's self-offer is an existential, constitutive element of human existence. The transcendental experience is revealed to be an experience of this offer. Human freedom is the capacity to determine who one is before God, to affirm or negate the divine self-offer. Specific Christian teachings give expression to this one core truth and its implications. This truth is proclaimed by the Church within history and can be grasped by all people because it calls them to what they already are.

An example can clarify how Rahner's theology works. Belief in the resurrection of Jesus is the foundation of Christian faith. Yet many find it difficult to believe because they think of Jesus' resurrection as a categorical event, the resuscitation of a dead body within history. Setting aside historical and categorical questions about what happened to Jesus after his death, Rahner suggests that the meaning of Jesus' resurrection is attained by correlating this doctrine with the transcendental experience of God's nearness. Again following Heidegger, Rahner describes

human existence as freedom-toward-death. Human beings know they will die, and in the anticipation of death the question inevitably occurs as to whether how one lives, who one becomes in freedom, makes any difference. Does human existence have any meaning or are we simply freedom becoming towards nothing? Since we are helpless to maintain ourselves in existence at death, the question of meaning raises the possibility that what we have achieved in freedom might be granted some final, enduring validity. This possibility of meaning, which Rahner calls transcendental hope, necessarily occurs to everyone. Even the assertion that life is meaningless implies a grasp of what meaningful existence might be. This hope is the experiential manifestation of God's self-gift to every human being. Hope occurs because humanity exercises its freedom within the horizon of God's self-offer.

Transcendental hope is the horizon for hearing and believing the proclamation that Jesus is risen. God has granted final, enduring validity to Jesus' life. From this side of the grave, understanding of the resurrection is rooted in freedom's option that life is meaningful. This option occurs in a life of generosity and service. Hope is rejected in selfishness. Freedom's 'yes' to the divine self-offer may or may not take the form of explicitly religious and Christian categories. But any individual who opts for meaning implicitly affirms the truth of Jesus' resurrection and Christian revelation.

See also: Heidegger, M.

JACK A. BONSOR

List of works

Rahner, K. (1939) *Spirit in the World*, trans. W. Dych, New York: Herder & Herder, 1968.(This difficult text is Rahner's early interpretation of Aquinas, an interpretation which underlies all his later work.)

Rahner, K. (1941) *Hearers of the Word*, trans. M. Richards, New York: Herder & Herder, 1969.(This translation of *Hörer des Wortes* is frequently inaccurate. A good translation may be obtained through the journal *Philosophy and Theology* from Marquette University, Milwaukee. The anthology *A Rahner Reader* includes well-translated sections.)

Rahner, K. (1954-84) *Theological Investigations*, New York: Crossroad, 23 vols, 1961-92. (The essays in these volumes, published originally in the sixteen-volume *Schriften zur Theologie*, contain the core of Rahner's work.)

Rahner, K. (1975) *A Rahner Reader*, ed. G. McCool, New York: Seabury.(An anthology, with excellent introductions by the editor, of Rahner's philosophical and theological works.)

Rahner, K. (1976) *Foundations of Christian Faith*, trans. W. Dych, New York: Seabury, 1978.(A synthetic presentation of Rahner's theology.)

References and further reading

Aquinas, T. (1266-73) *Summa theologiae*, Blackfriars edn, New York: McGraw-Hill, 1964.(Thomas Aquinas' comprehensive treatise, which remains the most influential work in Roman Catholic theology.)

Carr, A.E. (1984) 'Karl Rahner', in D.G. Peerman and M.E. Marty (eds) *A Handbook of Christian Theologians*, Nashville, TN: Abingdon Press, 519-542.(A good introductory essay which deals with both Rahner's philosophy and theology.)

Kant, I. (1781) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1963.(Kant's epistemology and his most influential work.)

Maréchal, J. (1927-49) *Le point de départ de la métaphysique (The Starting Point of Metaphysics)*, Paris: Desclée de Brouwer, and Brussels: L'Édition Universelle, 5 vols.(Maréchal places the thought of Aquinas in conversation with Kant, seeking a Thomistic response to Kant's method and conclusions.)

Maréchal, J. (1970) *A Maréchal Reader*, ed. and trans. J. Doncell, New York: Herder & Herder.(This is an anthology of Maréchal's writings taken almost entirely from his major work, *Le point de départ de la métaphysique*.)

McCool, G. (1989) *From Unity to Pluralism: The Internal Evolution of Thomism*, New York: Fordham University Press.(A history of the Thomistic revival in the early twentieth century that formed the context for Rahner's philosophical and theological work.)

McCool, G. (1989) *Nineteeth-Century Scholasticism: The Search for a Unitary Method*, New York: Fordham University Press.(An account of the origins of the Thomistic revival within Roman Catholicism that gave rise to the intellectual context for Rahner's thought.)

Schüssler-Fiorenza, F. (1968) 'Karl Rahner and the Kantian Problematic', introduction to *Spirit in the World*, New York: Herder & Herder, xix-xlv.(A clear essay setting forth the philosophical context for Rahner's early

work.)

Sheehan, T. (1987) *Karl Rahner: The Philosophical Foundations*, Athens, OH: Ohio University Press. (An excellent though difficult study of the philosophical background, especially Heidegger, for Rahner's *Spirit in the World*.)

Tallon, A. (1989-90) 'Rahner Studies, 1939-1989: Part I, 1939-73', *Theology Digest* 36 (4): 321-46; 'Rahner Studies, 1939-1989: Part II, 1974-89', *Theology Digest* 37 (1): 17-41; 'Rahner bibliography supplement', *Theology Digest* 38 (2): 131-40.(A fairly comprehensive bibliography of works about Rahner. Tallon's bibliography is to be kept current by the journal *Philosophy and Theology* from Marquette University, Milwaukee.)

Vorgrimler, H. (1986) *Understanding Karl Rahner: An Introduction to His Life and Thought*, New York: Crossroad.(An excellent introduction to Rahner's thought.)

Ramakrishna Movement

Although the Ramakrishna Movement was born in Bengal and influenced by Christian missionary activity and Western Orientalism, its understanding of Hinduism has become the standard for modern educated Indians. Drawing on the spiritual inspiration of its guru, Sri Ramakrishna (1836-86), and the dynamic preaching of his main disciple, Swami Vivekananda (1863-1902), the Ramakrishna Order has founded centres throughout India and the West. Calling his system 'Practical Vedānta', Vivekananda laid claim to the classical Advaita Vedānta associated with Śaōkara. Unlike Śaōkara, though, Vivekananda elevated selfless social work to a spiritual path equal in value to meditation, devotion and gnosis. The swamis of the order combine traditional Hindu religious practice with the administration of educational and medical institutions on the model of Christian missions. Vivekananda's vision of Indian culture as united and renewed by his humanistic Hinduism has inspired other gurus as well as Hindu nationalists.

1 Background

The Ramakrishna Movement, a highly successful Hindu religious and missionary order, was first propagated by Swami Vivekananda (1863-1902), a disciple of Sri Ramakrishna (1836-86), in the 1890s. Vivekananda received training in Western philosophy at a college of the Scottish Free Church in Calcutta, capital of the British Raj. Before joining Sri Ramakrishna during his college years, Vivekananda belonged to the Brahmo Samaj, the first major Hindu reform movement (see Brahmo Samaj). Wounded by the scathing attacks on Hindu tradition by Christian missionaries, by the triumphalist histories of Anglocentric Orientalists, and by discriminatory policies in the colonial civil service and the courts, members of the emerging Bengali Western-educated middle class took refuge in literature, the arts and religion. A few, like Vivekananda, sought to represent their culture in a way that would integrate their conflict-torn identities within and, at the same time, stand up to Western scrutiny. By the time of Lord Curzon's partition of Bengal in 1905, a mark of Vivekananda's success was that his new Hinduism was feeding the resistance movement against British colonialism.

Those educated Bengalis who did not completely forsake their cultural tradition took heart from a long line of sympathetic Orientalists - from Sir William Jones in the late eighteenth century to Max Müller in the late nineteenth. Both the Brahmo Samaj and Vivekananda's Ramakrishna Movement felt empowered by such work to imagine a reconciliation of the conflict, as they saw it, between East and West. Both Keshub Chunder Sen, leader of the Brahmo Samaj, and Swami Vivekananda took inspiration also from Sri Ramakrishna, who seemed to offer a liberal outreach from the authentic depths of the indigenous tradition (see Brahmo Samaj §2). Keshub preached a New Dispensation based on a personal mixture of Indian and Western traditions; Vivekananda convinced his contemporaries that his synthesis was fully rooted in Hinduism, that he 'spoke nothing but the Upaniṣads'. Vivekananda followed Orientalists in presenting the classic Hindu systems in Western terms. His ultimate intention, however, was to universalize key Vedāntic concepts such as Brahman, *ātman* and *mokṣa* into a controlling framework of norms and values for the entire world.

Vivekananda assumed that each culture represents a unique national spirit which complemented that of other nations. Here Vivekananda combined European romanticism with the caste theory of *jātidharma* - the notion that each class of persons has an inherent and unique role in society. Vivekananda tried to divine this intercultural cosmology by weighing the strengths and weaknesses of each nation's religion and society:

> Which is better, the social freedom of America, or the social system of India with all its restrictions? The American method is individualistic. It gives an opportunity to the lowest. There can be no growth except in freedom, but it also has obvious dangers. Still, the individual gets experience even through mistakes. Our Indian system is based entirely upon the good of the *samāj* [society]. The individual must fit into the system at any cost. There is no freedom for the individual unless he renounces society and becomes a *sannyāsin* (monk). This system has produced towering individuals, spiritual giants. Has it been at the expense of those less spiritual than themselves? Which is better for the race? Which? The freedom of America gives opportunities to masses of people. It makes for breadth, whilst the intensity of India means depth. How to keep both, that is the problem.

(Eastern and Western Admirers 1964: 218-9)

This dialectic was typical of Vivekananda's thought process. Romantic Orientalism assumed that the great cosmic struggle of the time pitted a materialist West against a spiritually gifted but weakened East. Vivekananda assimilated these two entities to the primary *guṇas* (qualities) of Sāōkhya cosmology: *rajas* (passion, power) and *tamas* (inertia) respectively. When *rajas* and *tamas* regain balance with each other, the universe attains to *sattva* (a salvific equilibrium) (see Sāōkhya §2). Through a cultural exchange of qualities, the East would regain its vitality and teach the West its spiritual insights.

2 Principal tenets

Although Ramakrishna introduced Vivekananda to Advaita Vedānta, Vivekananda's readings in Herbert Spencer, Ernst Haeckel and Paul Deussen confirmed the viability of metaphysical nondualism as a framework for cross-cultural apologetics and assimilation. Knowledge of the Absolute equalized all competing doctrines and symbols by uniformly nullifying them, but the penultimate perception of One-and-Many equally exalted them:

> All we know about things now or may know in future are but relative truths. It is impossible for our limited mind to grasp the absolute truth. Hence, though truth be absolute, it appears variously to diverse minds and intellects. All these facets or modes of truth belong to the same class as truth itself, they being based on the same absolute truth. This is like the different photographs of the same sun taken from various distances. Each of them seems to represent a different sun. The diverse relative truths have the same kind of relation with the absolute truth. Each religion is thus true, just because it is a mode of presentation of the absolute religion.
>
> (Eastern and Western Admirers 1964: 28)

Vivekananda wanted especially to include the materialist viewpoint - 'The materialist is right! There is but One. Only he calls that One Matter, and I call it God!' (Nivedita 1910: 23). Vivekananda wanted to claim like Herbert Spencer that he had reconciled religion and science. But the styling of philosophical differences as mere terminological distinctions recalls Ramakrishna's simple dictum that the different gods of the various religions are like the different names for water in various languages.

Following the Orientalist Sir William Hunter, Vivekananda linked Vedānta cosmology with Western evolutionism. For him evolution was a soteriological process:

> From the lowest form of life to man, the soul is manifesting itself through nature. The highest manifestation of the soul is involved in the lowest form of manifest life and is working itself outward through the process called evolution. The whole process of evolution is the soul's struggle to maintain itself.
>
> (Vivekananda 1977 vol. 6: 35-6)

Vivekananda maintained that the Indian theory taught that the process did not occur through competition (Spencer's survival of the fittest) but by the 'infilling of nature' (*prakṛtyāpūrāt*). Among humans, evolution took place not by overcoming obstacles but by letting the obstacles 'give way' through 'education and culture'. This theory is more 'consonant with reason' and less 'evil'. Darwin's theory applied to plants and animals; human culture advanced not by means of human competition and cooperation with Natural Law, as Spencer argued, but rather on the basis of a struggle against Nature, as Thomas Huxley argued, and self-sacrifice for the good of society.

Vivekananda used evolutionism also to frame his philosophy of history. He endorsed the views of Spencer and Müller that religions and nations evolve and progress in history, though not without cyclical periods of decay or regression. In the case of India, he postulated a Golden Age followed by degeneration, and he looked forward to a national renewal. Vivekananda was not promoting historical determinism, however. In March 1899 he said to his main Western disciple, Sister Nivedita (the Irishwoman and Indian nationalist Margaret Noble):

> I say, Margot, I have been thinking for days about [Spencer's] line of least resistance; it is a base fallacy. It is a comparative thing.… The history of the world is the history of a few earnest men, and when one man is earnest the world must just come to his feet. I am not going to water down my ideals, I am going to dictate terms.
>
> (Eastern and Western Admirers 1964: 278)

Thomas Carlyle's Great Man theory, another favourite of Vivekananda's, here overcame Spencer's Natural Law.

Vivekananda also questioned Western progressivism from a more traditional Indian view. Ramakrishna taught that ‘The world is like a dog’s curly tail… people have been striving to straighten it out for years… when they let it go, it has curled up again’ (Vivekananda 1977 vol. 1: 79). Vivekananda preached that people should aspire to progress, to do some good for society, but with the goal in mind of spiritual progress rather than worldly perfection: ‘You cannot overcome the idea of progress. But things do not grow better. They remain as they were, and we grow better, by the changes we make in them’ (Nivedita 1910: 148).

The goal of religion in classic Advaita Vedānta was *mokṣa*, a state in which the self frees itself from consciousness of both self and world. Ramakrishna privileged the state of exalted devotion in which one retains consciousness of one’s self in order to play in the consciousness of God. Vivekananda combined Advaita Vedānta and Western humanism to preach that because every *ātman* was truly Brahman, service to humanity was service to God.

See also: Vedānta

THOMAS L. BRYSON

References and further reading

Bharati, A. [Fischer, L.] (1970) ‘The Hindu Renaissance and Its Apologetic Patterns’, *Journal of Asian Studies* 29 (February): 267-88.(Biting Western view of Vivekananda as a facile nationalist ideologue by a former member of the Ramakrishna Order.)

Burke, M.L. (1957) *Swami Vivekananda in the West: New Discoveries*, Calcutta: Advaita Ashrama, 3rd edn, 1983-7, 6 vols.(Detailed documentation by a Western devotee of Vivekananda’s tour of the West, especially the USA.)

Chatterjee, P. (1993) *The Nation and Its Fragments: Colonial and Postcolonial Histories*, Princeton, NJ: Princeton University Press.(Secular Indian view of Vivekananda as a cultural nationalist.)

Dhar, S.N. (1975) *A Comprehensive Biography of Swami Vivekananda*, Madras: Vivekananda Prakashan Kendra, 2 vols.(More historically orientated than the official ‘Life’, but underplays Western influences.)

Eastern and Western Disciples (1964) *Reminiscences of Swami Vivekananda*, Calcutta: Advaita Ashrama, 5th edn, 1979.(Uncritical but comprehensive.)

Gupta, M. [M] (1969) *The Gospel of Sri Ramakrishna*, trans. Swami Nikhilananda, New York: Ramakrishna-Vivekananda Center.(Abridged, but well-written edition of a canonical Bengali original by M, a Western-educated devotee who transcribed his conversations with Ramakrishna.)

Halbfass, W. (1981) *Indien und Europa*, trans. W. Halbfass and J. Baker, *India and Europe: An Essay in Understanding*, Albany, NY: State University of New York Press, 1988. (Sophisticated hermeneutical and philosophical analysis of the East-West encounter, with sections on Vivekananda and contemporaries.)

Jackson, C.T. (1994) *Vedānta for the West: The Ramakrishna Movement in the United States*, Bloomington, IN: Indiana University Press.(Survey of the Order’s Mission to America.)

Kripal, J.J. (1995) *Kali’s Child: The Mystical and the Erotic in the Life and Teachings of Ramakrishna*, Chicago: University of Chicago Press.(Argues for the centrality of tantra in this tradition.)

Mahadevan, T.M.P. (1961) *Outlines of Hinduism*, Bombay: Chetana.(Representative presentation of neo-Vedānta in philosophical mode.)

Müller, F.M. (1899) *Ramakrishna: His Life and Sayings*, Calcutta: Advaita Ashrama, 1984.(Reprint of Müller’s understanding and digest of the movement as it was forming in the 1890s.)

Nivedita [Margaret Noble] (1910) *The Master as I Saw Him: Being Pages from the Life of the Swami Vivekananda by his Disciple Nivedita*, Calcutta: Udbodhan Office, 3rd edn, 1923.(Record of her training by Vivekananda’s main Western disciple.)

Nivedita [Margaret Noble] (1913) *Notes of Some Wanderings with the Swami Vivekananda*, ed. Swami Saradananda, Calcutta: Udbodhan Office.(Sister Nivedita describes her travels with Vivekananda on pilgrimage in northern India.)

Rambachan, A. (1994) *The Limits of Scripture: Vivekananda’s Reinterpretation of the Vedas*, Honolulu: University of Hawaii Press.(Reveals contradictions within Vivekananda’s reinterpretations and in their claims to continuity with Śaōkara’s philosophy.)

Raychaudhuri, T. (1988) *Europe Reconsidered: Perceptions of the West in 19th Century Bengal*, Delhi: Oxford University Press.(Analysis of Vivekananda’s ‘reconsideration’ of the West.)

Rolland, R. (1931) *Life of Vivekananda and the Universal Gospel*, trans. E.M. Smith, Calcutta: Advaita Ashrama, 1965.(Classic Western idealization of the new movement.)

Vivekananda (1977) *The Complete Works of Swami Vivekananda*, Calcutta: Advaita Ashrama, 8 vols.(Collection

of letters, lectures, interviews and published writings, with little or no contextualization.)

Rāmānuja (d. *circa* 1137)

A south Indian Brahman, Rāmānuja was the theistic exegete of the Vedānta who propounded a doctrine which came to be known as viśiṣṭādvaita *or 'qualified monism'. As such, he is often said to be the founder of the most prominent of the four schools of the Vaishnava religion, the Śrīsampradāya, although in fact he considered himself to be a participant in an already ancient tradition. Rāmānuja's version of Vedānta challenges the uncompromising nondualism of Śaṇkara.*

1 Life and works

Nine works are ascribed to Rāmānuja by his biographers: three are commentaries on the *Vedāntasūtra* of which the extensive *Śrībhāṣya* is the most important; *Vedāntasāra* (The Essence of Vedānta) and *Vedāntadīpa* (Light on the Vedānta) are shorter commentaries on the same text. His other philosophical works are a commentary on the *Bhagavad Gītā* and an independent summary of his doctrine, *Vedārthasaṃgraha* (A Compendium of the Meaning of *the Veda*) (*c*.1016-*c*.1137). *Nityagrantha* (A manual of Daily Worship), as well as three shorter devotional works are also ascribed to Rāmānuja. Of these nine, only *Śrībhāṣya*, *Bhagavad Gītābhāṣya* and *Vedārthasaṃgraha* are accepted without dispute to be his own. Of the remaining six, scholarly consensus leans towards accepting the two shorter commentaries on Vedānta as genuine and the others as later ascriptions.

Rāmānuja first took instruction in the Vedānta aphorisms from Yādavaprakāśa, who himself had composed a commentary on them in the *bhedābheda* ((simultaneous) difference and identity) tradition. Following Bhāskara (*c*.900), Yādavaprakāśa gave equal ontological status to both Brahman, the ultimate reality, and the phenomenal universe. However, in Rāmānuja's opinion, his teacher was inclined towards a monistic interpretation as he considered īśvara, the personal god, to be part of the phenomenal universe and not identical with Brahman. After his break with Yādavaprakāśa, Rāmānuja transferred his loyalties to Yāmunācārya, the head priest of the Viṣṇu temple at Śrīraōgam and representative of the devotional creed of the Āḻvārs. Although he had no personal contact with Yāmuna, legend has it that Rāmānuja was given a posthumous order to undertake several tasks, including that of presenting a systematic commentary on the Vedānta which was in opposition to that of Śaōkara, through which commentary a philosophical basis would be given to the devotional religion presented in the hymns of the Āḻvārs. This Rāmānuja was able to do through combining a realistic metaphysic with a theology of revelation.

2 Epistemology

Like Śaōkara, Rāmānuja accepted three sources of knowledge: direct perception, inference and revelation (See Knowledge, Indian views of). Although immediate experience of Brahman is the ultimate proof of its truth, revelation is the only dependable source of knowledge for the aspiring seeker, direct perception being subject to various sensory defects and inference to other human failings (see Brahman). Among the sources of revelation, Rāmānuja follows Śaōkara in giving precedence to the Upaniṣads, the *Brahmasūtras* and the *Bhagavad Gītā*, as the measure of all revelatory sources. In addition, he made extensive use of the Purāṇas, primarily the *Viṣṇupurāṇa* (*c*. fourth century AD).

3 Theology

According to Śaōkara, Brahman, the supreme truth, is pure existence with no qualities; it is consciousness alone without differentiation of subject and object (see Śaōkara). Rāmānuja understood that an eternal distinction between the individual soul and Brahman is essential to monotheistic religion. The religious relationship between God and the individual which Śaōkara considered to be conventional, that is, merely an aspect of the phenomenal existence, had to be shown to be ultimate. Rāmānuja thus rejected Śaōkara's theory of two levels of revelation, claiming that all revealed statements should be given equal status. Just as certain Upaniṣadic passages illustrating the ultimacy of the monotheistic position presented a problem to Śaōkara, Rāmānuja had to accept and deal with others which showed that the individual souls and the supreme truth were not different. This he did by showing that the distinctions between the selves, the phenomenal world of matter and God were maintained even while they formed a unity. Thus, the nondual (*advaita*) ultimate reality was shown to be qualified (*viśiṣṭa*) by the existence of the finite conscious living beings and the unconscious universe. All objects of experience are characterized by generic and specific properties. Therefore, Śaōkara's characterization of Brahma as pure existence which has no

properties was seen as a figment of the imagination. Yāmunācārya had argued persuasively for the existence of the soul by showing that despite the changes in the objects of knowledge, the knower remains essentially the same. Individual consciousness remains constant and unchanging despite the changes in the body. This is further confirmed by the sense of difference between the self and the body, as evidenced by a statement such as 'this is my body'. Rāmānuja expands on this idea with his doctrine of inseparability through which he argues that no soul can be without a body, for there is no such thing as consciousness without a field of consciousness, that is, a body. Similarly, God or Brahman cannot be pure consciousness (as is held by Śaōkara), because such a thing lies outside the realm of experience. Since nothing exists outside of Brahman, that of which he is conscious is also himself. The object of God's consciousness, that is, the creation consisting of both unconscious matter and individual souls, is thus considered, by analogy with the individual consciousness, to be his body. Despite the changes in the universe, his body, he, as the soul of the universe, remains unchanged and unchanging. Through this reasoning, the universe is given the same ontological status as the ever-constant eternal ground of being, because the one is dependent on the other. Just as changes in the body of the individual do not accrue to that individual, similarly the various transformations of the universe do not constitute a change in the essence of Brahman.

4 Soteriology

The individual soul has been confounded through his association with unconscious matter and so identifies with his material body. Through surrender and devotion he can attain a body which is suitable for directly experiencing and eternally serving the personal God. This is the ultimate goal of life. After Rāmānuja's death, his followers split over whether surrender, with its implication of complete dependence on the grace of God, or devotion, interpreted as a type of doctrine of works, was the more important aspect in the path of attaining God.

See also: Brahman; God, Indian conceptions of; Monism, Indian; Vedānta

JAN K. BRZEZINSKI

List of works

Rāmānuja (*c*.1016- *c*.1137) *Vedārthasaṃgraha of ŚrīRāmānujacārya*, trans. S.S. Raghavachar, Mysore: Sri Ramakrishna Ashram, 1968.(Without dispute, this text is accepted to be the author's own.)

References and further reading

Buitenen, J.A.B. von (1953) *Rāmānuja and the Bhagavad Gītā*, The Hague: Ned Boek en Steendrukkerij.(A condensed translation of Rāmānuja's *Bhagavad GītāBhāṣya*.)

Bhagavad Gītā (200 BC-AD 200, disputed), trans. with text in Devanagari and English S. Adidevananda, Madras: Sri Ramakrishna Math, 1991.(A more recent translation.)

Carman, J.B. (1974) *The Theology of Rāmānuja: An Essay in Interreligious Understanding*, New Haven, NJ and London: Yale University Press.(Comparisons of Rāmānuja's theology with that of the Christians.)

Dasgupta, S.N. (1940) *A History of Indian Philosophy*, Cambridge: Cambridge University Press, vol. 3.(An exposition of varieties of classical Vedānta; in five volumes.)

Devamani, B.S. (1990) *The Religion of Rāmānuja: A Christian Appraisal*, Madras: Christian Literature Society. (Rāmānuja's theology is compared with Christian theology.)

Lacombe, O. (1937) *L'Absolu selon le Vedānta: Les notions de Brahman et d'Atman dans les systèmes de Çankara et Rāmānuja (The Absolute According to Vedānta: The Notions of Brahman and Atman in the Systems of Śaōkara and Rāmānuja)*, Paris: Librairie orientaliste Paul Geuthner.(Comparative study of Rāmānuja and Śaōkara.)

Lipner, J.J. (1986) *The Face of Truth: A Study of Meaning and Metaphysics in the Vedāntic Theology of Rāmānuja*, Albany, NY: State University of New York Press.(An interesting read.)

Lott, E.J. (1976) *God and the Universe in the Vedāntic Theology of Rāmānuja*, Madras: Rāmānuja Research Society.(An in-depth discussion of Rāmānuja's theology.)

Radhakrishnan, S. (1928) *The Vedānta According to Śaṇkara and Rāmānuja*, London: Allen & Unwin.(This book has not lost its value.)

Raghavachar, S.S. (1972) *ŚrīRāmānuja on the Upaniṣads*, Madras: Professor M. Rangacharya Memorial Trust. (Rāmānuja shows the Upaniṣads to be a revelatory source.)

Srinivasachari, S.M. (1956) 'Advaita and Viśiṣṭ-ādvaita', Adyar Library Series, vol. 39, Madras: Adyar Library and Research Center, 2nd edn.(A study of Rāmānuja's doctrine.)

Thibaut, G. (1904) 'The *Vedāntasūtras* with Commentary by Rāmānuja', *The Sacred Books of the East Series*, vol. 48, ed. F.M. Müller, Oxford: Clarendon Press.(Numerous summaries of Rāmānuja's thought have been written, although translations of primary sources are rare. Thibaut's English translation of the *Vedāntasūtras* (*Brahmasūtras*) commentary, the *Śrībhāṣya* is considered to be adequate to the task.)

Varadachari, K.C. (1956) *Rāmānuja's Theory of Knowledge,* Tirupati: Tirumalai-Tirupati Devasthanams Press, 2nd edn.(An examination of Rāmānuja theory of the three sources of knowledge.)

Viṣṇupurāṇa (*c*.4th century AD), trans. Nag Sharan Singh, Delhi: Nag Publishers, 1980.(Rāmānuja made extensive use of this text in his philosophy.)

Ramsey, Frank Plumpton (1903-30)

Before Ramsey died at the age of 26 he did an extraordinary amount of pioneering work, in economics and mathematics as well as in logic and philosophy. His major contributions to the latter are as follows. (1) He produced the definitive version of Bertrand Russell's attempted reduction of mathematics to logic. (2) He produced the first quantitative theory of how we make decisions, for example about going to the station to catch a train. His theory shows how such decisions depend on the strengths of our beliefs (that the train will run) and desires (to catch it), and uses this dependence to define general measures of belief and desire. This theory also underpins his claim that what makes induction reasonable is its being a reliable way of forming true beliefs, and it underpins his equation of knowledge generally with reliably formed true beliefs. (3) He used the equivalence between believing a proposition and believing that it is true to define truth in terms of beliefs. These in turn he proposed to define by how they affect our actions and whether those actions fulfil our desires. (4) He produced two theories of laws of nature. On the first of these, laws are the generalizations that would be axioms and theorems in the simplest true theory of everything. On the second, they are generalizations that lack exceptions and would if known be used to support predictions ('I'll starve if I don't eat') and hence decisions ('I'll eat'). (5) He showed how established, for example optical, phenomena can be explained by theories using previously unknown terms, like 'photon', which they introduce. (6) He showed why no grammatical distinction between subjects like 'Socrates' and predicates like 'is wise' entails any intrinsic difference between particulars and universals.

1 Mathematics

The British philosopher Frank Plumpton Ramsey graduated in mathematics from Trinity College Cambridge in 1923, became a Fellow of King's College in 1924 and a University Lecturer in Mathematics in 1926. In his short life he did work in mathematics and economics which created new branches of those subjects. But his vocation and greatest achievements were in philosophy, influenced by and influencing his Cambridge colleagues, especially Russell, Wittgenstein and Keynes. His first major work, on 'The Foundations of Mathematics' (1925) (this can be found, as can all his other works cited in this entry, in Ramsey (1990a)), makes two key improvements to the attempted reduction of mathematics to logic in Whitehead and Russell's *Principia Mathematica* (see Logicism). First, Ramsey tightens *Principia*'s definition of mathematical propositions as purely general by requiring them also to be tautologies in the sense of Wittgenstein's *Tractatus*. Second, he shows that it need deal only with purely logical paradoxes, like that of the class of all classes which are not members of themselves, not semantic ones like 'this is a lie', which depend on the meanings of words like 'lie'. This lets Ramsey simplify *Principia*'s complex hierarchy of propositions, since it no longer needs to make 'this is a lie' ill-formed by making '*p*' and '*p* is a lie' propositions of different types. This in turn lets him drop the 'axiom of reducibility' (of propositions of higher types to equivalent ones of lower types) that *Principia* needs to validate 'many important mathematical arguments' (1990a: 190). Dropping this axiom is essential to logicism, since 'there is no reason to suppose [it] true; and if it were true, this would be a happy accident and not a logical necessity, for it is not a tautology' (1990a: 191) (see Theory of types; Semantic paradoxes and theories of truth).

In 'Mathematical Logic' (1926) Ramsey defends his logicism against 'the Bolshevik menace of Brouwer and Weyl' (1990a: 219) (see Intuitionism) and the formalism of Hilbert (see Hilbert's programme and formalism). The former he attacks for denying that all propositions are either true or false - 'Brouwer would refuse to agree that either it was raining or it was not raining, unless he had looked to see' (1990a: 228) - and the latter for reducing mathematics to 'a meaningless game with marks on paper' (1990a: 233). But he remains embarrassed by logicism's need to assume that there are infinitely many things; and R.B. Braithwaite, introducing his edition of Ramsey's papers in 1931, reports that in 1929 Ramsey 'was converted to a finitist view which rejects the existence of any actual infinite aggregate'. Despite this, and logicism's general rejection by mathematicians, Ramsey's version of it remains of great interest to logicians (Chihara 1980).

2 Probability and knowledge

If Russell and Wittgenstein prompted Ramsey's work on mathematics, Keynes prompted his work on economics and probability. The latter was provoked especially by Keynes' *A Treatise on Probability* (1921), which extends the deductive logic of conclusive inference to an inductive logic of inconclusive inference by postulating a relation

of ‘partial entailment’, knowable a priori. This, when measurable, enables a probability measure of the strength of an inference from one proposition to another. But in ‘Truth and Probability’ (1926) Ramsey attacked the idea of an a priori inductive logic so effectively that Keynes abandoned it, although it was later revived (see Carnap; Probability, interpretations of).

Ramsey’s main achievement in ‘Truth and Probability’ is his probability measure of the strength (degree) of a belief. This starts from ‘the old-established way of measuring a person’s belief, [that is,] to propose a bet, and see what are the lowest odds which he will accept’ (1990a: 68). Refining this by invoking ‘the theory that we act in the way we think most likely to realise the objects of our desires’ (1990a: 69), Ramsey derives measures both of desires (subjective utilities) and of beliefs (subjective probabilities), thereby founding the now standard use of these concepts (see Decision and game theory).

Ramsey himself uses his theory to extend ‘the lesser logic … of consistency’ (1990a: 82) from full to partial beliefs. Thus to the injunction not to believe both *p* and not-*p* he adds, for example, that anyone who believes *p* to degree 1/3 must believe not-*p* to degree 2/3. Otherwise what these beliefs will make him do ‘would depend on the precise form in which the options were offered him, which would be absurd’: that is, they would make him bet at different odds on *p* and against not-*p*, so that ‘he could have a book made against him by a cunning better and would then stand to lose in any event’ (1990a: 78).

Ramsey also uses his theory to develop ‘the larger logic … of discovery’, applied (following Peirce) to belief-forming ‘habits of inference … observation and memory’ (1990a: 92). He shows why such a habit is good or bad ‘as the degree of belief it produces is near or far from the actual proportion in which the habit leads to truth’, and hence why the fact that ‘the world is so constituted that inductive arguments lead on the whole to true opinions’ makes our inductive habits reasonable (1990a: 93).

Hence also Ramsey’s claim in ‘Knowledge’ (1929) that ‘a belief [is] knowledge if it is (i) true, (ii) certain [that is, a full belief], (iii) obtained by a reliable process … [that is, one] that can be more or less relied on to give true beliefs’ (1990a: 110). This claim anticipates later accounts of knowledge (see Internalism and externalism in epistemology; Reliabilism), showing amongst other things how we can know things we do not know we know. This in particular enables Ramsey to evade several well-known objections to knowledge, conceived of as true belief that the believer could justify, which need to assume that I can only know something if I know I know it (Sahlin 1991).

In ‘Truth and Probability’ Ramsey does not apply his subjective reading of probability to physics. Unfortunately he later, in ‘Chance’ (1928), anticipates de Finetti (1937) by taking even physical chances to be only ‘in another sense objective, in that everyone agrees about them’ (1990a: 106). To this we may reply by adding to his objection to Keynes, that determining the right probabilities ‘in molecular mechanics … is a matter of physics rather than pure logic’ (1990a: 85), that it takes more than mass psychology to explain (for example) the random decay of radioactive atoms.

Fortunately Ramsey does not make all his successors’ mistakes. In particular, unlike many later decision theorists (for example, Jeffrey 1983), he never *prescribes* acting ‘in the way we think most likely to realise the objects of our desires’. He claims only that the theory that we do so is ‘a useful approximation to the truth … like Newtonian mechanics’ (1990a: 69). Here he is right: for even when I in fact do something because ‘it seemed a good idea at the time’, this fact about my action does not suffice to make it rational.

3 Belief and truth

In defining a belief’s strength by its effects on our actions, Ramsey foreshadows later theories which define mental states by their causes, effects and interactions (see Functionalism). In ‘Facts and Propositions’ (1927) he extends this idea from degrees to contents of beliefs, taking (for example) ‘the equivalence between believing Onot-*p*O and disbelieving O*p*O … to be defined [by their sharing] many of their causes and … effects’ (1990a: 44). But after failing to define the contents of beliefs as effectively as their degrees, Ramsey concludes that his view, that ‘the meaning of a sentence [expressing a belief] is to be defined by reference to the actions to which asserting it would lead’, remains ‘very vague and undeveloped’ (1990a: 51).

It is, however, developed enough to stop Ramsey’s theory of truth being, as is usually supposed, that truth is

definable by the fact that for all *p*, it is true that *p* iff *p* (see Truth, deflationary theories of). He does say that 'there is really no separate problem of truth' since (for example) 'Oit is true that Caesar was murderedO means no more than that Caesar was murdered', so that 'if we have analysed [belief] we have solved the problem of truth' (1990a: 38-9). But the solution will not be the so-called redundancy theory if our analysis of beliefs includes a substantive analysis of their truth conditions, as Ramsey's needs to do.

Ramsey starts by observing that we can 'say that a chicken believes a certain sort of caterpillar to be poisonous, and mean by that merely that it abstains from eating such caterpillars on account of unpleasant experiences connected with them'. Since this action is 'such as to be useful if, and only if, the caterpillars were actually poisonous … any set of actions for whose utility *p* is a necessary and sufficient condition might be called a belief that *p*, and so would be true if *p*, i.e. if they were useful' (1990a: 40, see Truth, pragmatic theory of).

Unfortunately Ramsey drops this idea when dealing 'with those beliefs which are expressed in words … or other symbols, consciously asserted', although it can apply to them too. In fact its only fault is to identify a belief with a set of actions, like abstaining from eating caterpillars, instead of with one of their causes. But any theory that makes beliefs entail causal functions from desires to actions can remedy this. For then the 'set of actions' of a full belief *b* will be all those that *b* would combine with some desire to cause; and *p* will be the condition in which every such action would succeed - that is, achieve the object of the desire involved, say to eat without dying. But this is obviously the condition that *b* be true, that is, *b*'s truth condition.

Ramsey can therefore let this 'success semantics' (Whyte 1990) give the truth condition of any belief definable by 'the actions to which … it would lead'. Indeed he must do so, since a belief's truth condition cannot be given just by how it makes us act, for that will be the same whether it is true or false. What does depend on a belief's truth is whether the actions it causes succeed: hence success semantics. But this, while vindicating Ramsey's claim that analysing belief solves the problem of truth, rules out the redundancy theory: for success semantics, since its contribution to the analysis of beliefs is to say what makes them true, is itself a substantive theory of truth.

4 Laws and causation

Ramsey produced two theories of laws of nature (see Laws, natural). Both are Humean in distinguishing law statements from accidentally true generalizations not by what they say but by how we use them. In 'Universals of Law and of Fact' (1928) Ramsey says they are 'consequences of those [general] propositions which we should take as axioms if we knew everything and organised it as simply as possible in a deductive system' (1990a: 150). Although Ramsey soon abandoned this 'systematic theory' (Armstrong 1983), it remains the best Humean account of laws.

Ramsey's second theory, in 'General Propositions and Causality' (1929), is that law statements like 'all men are mortal' are 'variable hypotheticals', which 'are not judgments but rules for judging OIf I meet a I shall regard it as a Čpsi;O' (1990a: 149). Thus 'a causal generalization is not … one which is simple, but one we trust' (1990a: 150), while to believe there are unknown laws is to believe there are 'such singular facts … as would lead us, did we know them, to assert a variable hypothetical [which must] be also asserted to hold within … the scope of our possible experience' (1990a: 152).

This theory explains why we invoke causation and laws in assessing action, since 'we cannot blame a man except by considering what would have happened if he had acted differently; and this … depends essentially on variable hypotheticals' (1990a: 154). But its account of why 'the deduction of effect from cause is conceived to be so radically different from that of cause from effect' (1990a: 157) will not do. For here Ramsey relies on our view that causes precede their effects, a view he identifies with the fact 'that any present volition of ours is (*for us*) irrelevant to any past event … *to us* now what we do affects only the probability of the future' (1990a: 158; emphasis added). But this implies that the only reason we can't affect the past is that we think we can't, which is absurd. Only the fact, not the view, that effects never precede their causes can explain why gluttons should believe 'I will starve if I don't eat' but not 'I will have starved if I don't eat': *pace* Ramsey, some variable hypotheticals need to be made true by facts (see Causation; Counterfactual Conditionals).

5 Theories

Scientific theories apply new predicates to unobservable entities, like photons, to explain observable, for example

optical, phenomena. How do these predicates acquire empirical meaning? Ramsey's drastic answer in 'Theories' (1929) is that there are no such predicates: we use 'is a photon', 'has frequency *n*', and so on not as predicates but as existentially bound variables. That is, a theory tacitly starts with quantifiers, 'properties exist - call them Obeing a photonO, etc. - such that …', followed by the explicit theory, in two parts. Its *axioms* link its predicate variables to each other, while its *dictionary* (see Campbell, N.R.) links them to observable predicates like 'is red' (1990a: 112). Thus if '*a*', '*b*' and '*g*' are our theoretical predicates, 'the best way to write our theory seems to be … (∃*a*,*b*,*g*):dictionary.axioms' (1990a: 131). This, which is now called the 'Ramsey sentence' of the theory, eliminates its problematic predicates while keeping its structure and observable consequences.

Although this account has been widely accepted - and explicitly applied to functionalist theories of the mind (Lewis 1972) - its explanations of other striking facts about theories are rarely noticed. It entails for example that parts of theories, since they contain variables, are not 'strictly propositions by themselves' and their meaning 'can only be given when we know to what stock of OpropositionsO … [they are] to be added' (1990a: 131). Since this makes theoretical statements in rival theories incomparable, 'the adherents of two such theories could quite well dispute, although neither affirmed anything the other denied' (1990a: 133). This both explains the phenomenon of 'incommensurability' (see Kuhn, T.S.; Incommensurability) and limits its consequences, for example for deductive accounts of theoretical explanation (see Hempel, C.G.): for as Ramsey remarks, it does not affect reasoning within the scope of a single theory's quantifiers.

Because Ramsey sentences say that certain universals (properties or relations) exist, nominalists, who deny this, must reject them (see Nominalism). Realists about universals, however, can use Ramsey sentences to determine what empirical universals exist, as follows. Since not only *un*observable properties exist, we treat *all* predicates in law statements as variables. The Ramsey sentence of all such statements then quantifies over all universals that occur in laws, which are all the empirical universals there are (Mellor 1991).

6 Universals

What lets Ramsey ignore nominalism in 'Theories' is his denial in 'Universals' (1925) that our distinction between particulars and universals shows any intrinsic difference between them. First, it cannot be based on the subject-predicate distinction, e.g. between 'Socrates' and 'is wise' in 'Socrates is wise': for the subject of the equivalent 'wisdom is a characteristic of Socrates' is wisdom, which is not a particular (1990a: 14). Also, in molecular propositions like 'Socrates is wise or Plato is foolish', the subject-predicate distinction generates complex universals, like being wise unless Plato is foolish. But Ramsey argues that if these existed, then (for example) that a universal *R* relates *a* to *b*, that *a* has the complex property *Rb* and that *b* has *aR* would 'be three different propositions because they have different sets of constituents, and yet they are … but one, namely that *a* has *R* to *b*. So the theory of complex universals is responsible for an incomprehensible trinity, as senseless as that of theology' (1990a: 14). Similarly with Socrates' apparent property of being wise-unless-Plato-is-foolish and Plato's of being foolish-unless-Socrates-is-wise. If, as Ramsey assumes, the proposition that Socrates is wise or Plato is foolish can have only one set of constituents, there can be no such complex properties.

Predicates can therefore distinguish universals from particulars only in atomic propositions, and even then the distinction will not imply an intrinsic difference unless that difference would explain our impression that (for example) 'Socrates is a real independent entity, wisdom … a quality of something else' (1990a: 19). But no such difference will do this. For our impression comes from associating 'wise' only with propositions of the atomic form '*x* is wise' while associating 'Socrates' with *all* propositions containing it, including the molecular 'Socrates is neither wise nor just'. Yet we could as easily associate 'wise' with this and all other propositions containing it, and restrict 'Socrates' to the atomic form 'Socrates is *q*', where *q* is a universal: a form which, without complex universals, can no more include 'Socrates is neither wise nor just' than '*x* is wise' can include 'neither Socrates nor Plato is wise' (1990a: 20-1). So no intrinsic difference between universals and particulars can be inferred from - since none will explain - our associating atomic forms with predicates but not subjects.

Why then do we do that, thus making universals seem less 'real and independent' than particulars? Ramsey's explanation is this. A predicate symbol 'Čphis;' can stand alone only if it names a real universal, not if it abbreviates (for example) 'has *R* to *a* or *S* to *b*'. This we must abbreviate to 'Čphis;*x*', to distinguish it from the two-place '… has *R* to *a* or … has *S* to *b*', written 'Čphis;(*x*,*y*)'. But since it is irrelevant to an extensional logic whether or not 'Čphis;' names a universal, we always write 'Čphis;' as 'Čphis;*x*', 'Čphis;(*x*,*y*)', etc., thus

associating all predicates with atomic forms (1990a: 26-8).

We cannot therefore infer from this practice that particulars differ intrinsically from universals. A logician can take 'any type of objects whatever as the subject of his reasoning, and call them individuals, meaning by that simply that he has chosen this type to reason about' (1990a: 30). We naturally choose easily discriminable objects, such as those with locations in space and time, to quantify over first; but what makes them particulars is simply *that* we choose them, not why. But then the fact that objects of certain types fail to count as particulars, just because we choose to exclude them from the range of our *first*-order quantifiers, is no reason to deny, as nominalists do, that they exist. The existence of universals - that is, of whatever we leave for our second-order quantifiers to range over - is no more problematic than that of particulars.

See also: Belief; Frege, G.; Mind, identity theory of; Probability theory and epistemology; Universals

D.H. MELLOR

List of works

Ramsey, F. P. (1931) *The Foundations of Mathematics and other Logical Essays*, ed. R.B. Braithwaite, London: Routledge & Keegan Paul.(Referred to in §1. The first published collection of Ramsey's papers.)

Ramsey, F. P. (1990a) *Philosophical Papers*, ed. D. H. Mellor, Cambridge: Cambridge University Press.(Contains all Ramsey's major philosophical papers. By permission of Cambridge University Press, this entry draws on the editor's Introduction to this book. Page references for the quotations from Ramsey's work refer to this edition; the dates given are those of first publication or, if published posthumously, of composition.)

Ramsey, F. P. (1990b) 'Weight or the Value of Knowledge, ed. N.-E. Sahlin, *British Journal for the Philosophy of Science* 41: 1-3.(Previously unpublished note calculating the value of collecting evidence for the truth or falsity of a proposition.)

Ramsey, F. P. (1990c) *Notes on Philosophy, Probability and Mathematics*, ed. M.C. Gavalotti, Naples: Bibliopolis.(Previously unpublished notes.)

Ramsey, F. P. (1991) *On Truth*, ed. N. Rescher and U. Majer (1991), Dordrecht: Kluwer. (Previously unpublished notes for an uncompleted book on truth and related matters.)

References and further reading

Armstrong, D.M. (1983) *What is a Law of Nature?*, Cambridge: Cambridge University Press, ch. 5.4.(Referred to in §4. Argues that Ramsey's first theory of laws of nature is the best Humean account of them and then attacks it.)

Chihara, C.S. (1980) 'Ramsey's Theory of Types: Suggestions for a Return to Fregean Sources', in *Prospects for Pragmatism: Essays in Memory of F. P. Ramsey*, ed. D.H. Mellor, Cambridge: Cambridge University Press, 21-47. (Referred to in §1. Defends a Fregean reading of Ramsey's theory of types.)

de Finetti, B. (1937) 'Foresight: its Logical Laws, its Subjective Sources', ch. VI, trans. H. E. Kyburg, in *Studies in Subjective Probability*, eds H. E. Kyburg Jr and H. E. Smokler (1964), New York: Wiley, 93-158.(Referred to in §2. Argues on operationalist grounds - see Operationalism - against objective interpretations of probability even in physics.)

Jeffrey, R.C. (1983) *The Logic of Decision*, 2nd edn, Chicago: University of Chicago Press, ch. 3.(Referred to in §2. Interprets Ramsey's *'Truth and Probability'* as a prescriptive subjective decision theory.)

Keynes, J.M. (1921) *A Treatise on Probability*, London: Macmillan.(Referred to in §2. Gives the interpretation of probability as a measure of a logical relation of partial entailment attacked by Ramsey and later revived by Carnap.)

Lewis, D. (1972) 'Psychophysical and Theoretical Identifications', in *Readings in Philosophy of Psychology*, vol. I, ed. N. Block, London: Methuen, 1980, 207-22.(Referred to in §5. Uses Ramsey sentences to define mental states by their causes and effects.)

Mellor, D.H. (ed.) (1980) *Prospects for Pragmatism: Essays in Memory of F. P. Ramsey*, Cambridge: Cambridge University Press.(Contains eleven original essays on different aspects of Ramsey's philosophy.)

Mellor, D.H. (1991) 'Properties and Predicates', *Matters of Metaphysics*, Cambridge: Cambridge University Press, 170-82.(Referred to in §5. Uses the Ramsey sentence of all laws to determine what empirical universals exist.)

Sahlin, N.-E. (1990) *The Philosophy of F. P. Ramsey*, Cambridge: Cambridge University Press.(A comprehensive introduction to Ramsey's philosophy and also to his mathematics and economics, not dealt with in this entry.)

Sahlin, N.-E. (1991) 'Obtained by a Reliable Process and always Leading to Success', *Theoria* 57: 132-49.

(Referred to in §2. Develops and defends Ramsey's reliabilist theory of knowledge.)

Whitehead, A.N. and Russell, B. (1927) *Principia Mathematica*, Cambridge: Cambridge University Press, 2nd edn.(Referred to in §1. The basis of Ramsey's attempt to reduce mathematics to logic.)

Whyte, J.T. (1990) 'Success Semantics', *Analysis* 50 (3): 149-57.(Referred to in §3. Gives the truth conditions of beliefs as the conditions in which the actions they combine with desires to cause achieve the objects of those desires.)

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, London: Routledge, §4.46. (Referred to in §1. The theory of tautology used by Ramsey to strengthen *Principia*'s definition of mathematical propositions.)

Ramsey, Ian Thomas (1915-72)

Ramsey's work developed within two parameters. One concerned God and language: he held that no literal statement could be true of God; all language concerning God must be metaphorical. Another concerned his epistemology: knowledge, he held, comes ultimately from experience - sensory, introspective, but also religious. Evidence that God exists comes from experience, and claims about God must be cast in nonliteral terms.

Born in 1915, Ramsey earned firsts in mathematics, moral sciences and theology at Cambridge University. He was chaplain of Christ's College, Cambridge, from 1943, and Nolloth Professor of the Philosophy of the Christian Religion at Oxford from 1951-66. In 1966, he was appointed Bishop of Durham.

Ramsey maintained that any use of religious language needs to be understood in terms of its context. He described the experiential contexts in which the word 'God' is used as a referring term. To do this, it is necessary to map the relationships between the term 'God' (and the concept it expresses) and other terms endemic to the same context. This constitutes Ramsey's 'logical empiricism'.

In endeavouring to understand and express experiences of divine activity, believers use models. For Ramsey, in no sentence of the form 'God is… ' is the predicate term that fills in the blank used either univocally, equivocally or analogically. The relation between 'God' and the predicate term will be that between the key word of a model and a term qualifying that model. Thus his analysis of religious experience is offered in the formal rather than the material mode.

As an example of how Ramsey's account works, consider the word 'wave', which is a key term in a model used to record regularities in the behaviour of light. Otherwise isolated phenomena thereby become conceptually organized, and prediction of further regularities is possible. Second-order models organize first-order models into a comprehensive theory. Without first- and second-order models, science would be impossible.

Logical positivism and its verification principle dominated Ramsey's academic setting. The plausibility and the destructive implications of the verification principle were vastly overrated, but to be heard, Ramsey had to respond to positivism's challenge. The positivists prized formal philosophy, science and sensory experience. Ramsey offered a formal mode account of models, and an analysis of science as the task of constructing appropriate first-order reportive models and second-order theoretical models. He added that theology does for religious experience what science does for sensory experience. In both cases, the relevant story is formal, experiential and model-governed. Ramsey's view was that the nature of religious experience justified his type of approach to it, and that the approach of science to observation was analogous; thus in so far as positivist concerns were legitimate, they could be met. What Ramsey called a 'disclosure model' of some thing *X* is isomorphic with *X* (shares *X*'s structure), but shares no quality with *X* and justifies no description of *X* beyond the claim that the model and *X* are structurally analogous.

Ramsey's account of self-knowledge is central to his theory. Each of us has experiences in which we know ourselves distinctively as ourselves. In such experiences, there is self-disclosure; we apply to ourselves a model in which 'I' is the key term. We may discern ourselves in circumstances in which we fall in love, fear death, or recognize an obligation. The disclosure is to the effect that the self (construed as indescribable) bears a relation to its states in a manner structurally analogous to the relation between 'I' and 'my states' in a conceptual model in which these serve as key terms. We may experience God in situations in which we are impressed by ceaseless changes in nature or note the existential dependence of what we experience, discerning a changeless being or an independent being. This amounts to the claim that God and changing or dependent things are related to one another as the 'Unchanging Independent Cause' and 'changing dependent things' in a conceptual model in which these serve as key terms.

Ramsey, then, claimed that in our recognition of change, dependence and demise we discern the presence of something unchanging, independent and unending. He held that God cannot be described in literal terms, and endeavoured to hold these ideas together by viewing the relevant experiences as properly capturable in models whose structure (and nothing else) is ascribable to dependent things and God. Disclosure experiences are for Ramsey epistemologically primitive, needing no justification. Theological reasoning is a matter of placing

collective disclosure models, which integrate particular experiences or types of experiences, under a second-order model, and theological statements 'point to' rather than describe God.

Ramsey tried to combine a stringent empiricism with an acceptance of the cognitive significance of theology. He thought that scientific theories and Christian theology were similarly dependent on the use of nonpicturing models. His work helped keep alive a stream of philosophical theology which nearly dried up during the times of positivism's dominance, but which again flourishes today.

KEITH E. YANDELL

List of works

Ramsey, I. (1960) *Freedom and Immortality*, London: SCM Press.(Presents an argument for human transcendence of empirical conditions.)

Ramsey, I. (ed.) (1961) *Prospect for Metaphysics*, New York: The Philosophical Library. (An excellent collection of essays, including one by Ramsey on the prospects of metaphysical theology.)

Ramsey, I. (1963a) *On Being Sure in Religion*, London: Athlone Press.(Develops Ramsey's religious epistemology.)

Ramsey, I. (1963b) *Religious Language*, New York: Macmillan.(Defends Ramsey's view of religious language.)

Ramsey, I. (1964a) *Religion and Science: Conflict and Synthesis*, London: SPCK.(Gives Ramsey's account of the similarities and differences between religion and science, arguing that both are cognitively significant.)

Ramsey, I. (1964b) *Models and Mystery*, London: Oxford University Press.(Presents Ramsey's views concerning what he sees as a delicate balance, mediated by models, between the incomprehensibility and intelligibility of what matters most.)

Ramsey, I. (1964c) *Christian Empiricism*, ed. J. Gill, London: Sheldon Press.(An excellent collection of Ramsey's writings, with a good bibliography.)

Ramsey, I. (1965) *Christian Discourse: Some Logical Explorations*, London, Oxford University Press.(An account of Ramsey's view of Christian religious language.)

Ramsey, I. (ed.) (1966) *Christian Ethics and Contemporary Philosophy*, London, SCM Press.(Has an introduction, an essay on moral judgment and divine commands, and an essay on natural law, by Ramsey.)

Ramsey, I. (eds) (1971) *Personality and Science*, London: Churchill Livingstone. (Introduction, epilogue, and essay on human personality, by Ramsey.)

References and further reading

Edwards, D.L. (1973) *Ian Ramsey, Bishop of Durham*, London: Oxford University Press.(A memoir.)

Lucier, P. (1976) *Empirisme logique et langage réligieux: trois approches anglo-saxones contemporaines (Logical Empiricism and Religious Language: Three Contemporary Anglo-Saxon Approaches)*, Tournai: Desclée & Cie.(Deals with R.B. Braithwaite and R.M. Hare as well as Ramsey.)

Pye, J. (1979) *A Bibliography of the Works of Ian Thomas Ramsey*, Durham: Abbey House Publications.(A useful resource.)

Ramus, Petrus (1515-72)

Petrus Ramus, for many years a professor of philosophy and eloquence at the University of Paris, wrote textbooks and controversial works in grammar, logic, rhetoric, mathematics, physics and philosophy. He was also a university reformer. His followers were prolific with commentaries, Ramist analyses of classical texts and handbooks of their own. His logical works and those of his school exercised a large influence between 1550 and 1650.

His formation was humanist, in that he attacked scholasticism and encouraged the study (and logical analysis) of classical texts, as Agricola, Sturm and Melanchthon had done. But he was far more independent-minded than them, a stern critic of the textbooks of Aristotle and Cicero, as well as an admirer of their style and intellect. His most important innovation was the method, a theory of organization which he used to simplify his textbooks. He emphasized the need for learning to be comprehensible and useful, with a particular stress on the practical aspect of mathematics. His critics would say he oversimplified. He was also a student of Gaulish pseudo-antiquities and an important proponent of the French language. His Dialectique (1555) was the first book on dialectic in French.

1 Life and educational reforms

Petrus Ramus (born Pierre de la Ramée at Cuts, near Noyon) came from a family which had lost its wealth but not its patent of nobility with the sack of Liège in 1468. He studied in Paris at the Collège de Navarre, where he took his MA in 1536, allegedly defending the thesis that ‘everything which Aristotle said was arbitrarily fabricated’. Even if this prophetic story is true, it should be pointed out that the theses chosen for MA disputations were often humorous or paradoxical. At this stage of his life, Ramus had already been influenced by Johann Sturm, who introduced the teaching of Rudolph Agricola to Paris. Ramus began teaching at the Collège du Mans, before moving to the Collège de l’Ave Maria. As a young teacher he reread the texts of the arts course and began to formulate his response to them in his earliest books *Dialecticae partitiones (Divisions of Dialectic)*, *Dialecticae institutiones (The Teaching of Dialectic)* and *Aristotelicae animadversiones (Notes on Aristotle)* (all published in 1543). These books attacking Aristotle’s approach to dialectic were condemned by the university authorities and Ramus was forbidden to teach or publish in philosophy. He turned his attention to rhetoric and mathematics, but the shortage of teachers enabled him to evade the ban, by teaching at the Collège de Presles, and he soon became head of the college. He published a new version of his dialectic under the name of his friend Omer Talon. In 1547 he was released from the ban owing to the influence of the Cardinal de Guise on the new king, Henry II. In 1551 he was named professor of philosophy and eloquence at the Collège Royal. He continued to teach and publish on rhetoric, dialectic, Latin literature, philosophy and mathematics. He became involved in many controversies with his critics, notably Jacques Charpentier, both over his attacks on Aristotle and in relation to university appointments. In 1562 he became a Calvinist. Persecution forced him to flee Paris in 1562 and 1567. From 1568 he travelled in Germany and Switzerland. He was persuaded to return to Paris in 1570, but even explicit royal protection was unable to save him from the wrath of his enemies during the St Bartholomew’s Day Massacre of 1572. That he died a Protestant martyr had considerable consequences for his later reputation. His works were massively reprinted and became very influential in Protestant parts of Germany, in Britain and in New England well into the seventeenth century.

Ramus composed several orations and prefaces on the reform of the University of Paris, in practical matters as well as in the curriculum. The main force of his curricular changes was to support and extend the growth of humanism in Paris. He wanted to purge Aristotle as well as the accretions of scholastic logic. He proposed (in 1546) that rhetoric and dialectic should be taught simultaneously, and involve composition by the pupils. Both subjects were to be illustrated through the analysis of literary texts. In 1551 he suggested a more orthodox linear syllabus, but argued for the study of mathematics and physics in the last two years (though these were also to involve literature, including Virgil’s *Georgics* and Lucretius in the final year). In 1562 he called for more practical work in physics and medicine, referring particularly to botany and dissection. Ramus’ reforms were humanist, they were methodical (in the sense of §3 below), but above all they were intended to increase the usefulness of what is learnt in the university.

2 Logic and rhetoric

For Ramus logic was the subject which organized the rest of the curriculum. Ramus published several different versions of his major works in dialectic and rhetoric, all of which were reprinted many times. Bruyère (1984) proposes five stages for the development of the dialectic; Meerhoff (1986) proposes eight for the rhetoric. Moreover, earlier versions continued to be reprinted after later ones. These facts present the student of Ramus with large bibliographical problems. It is essential to study different states of each text and to make use of the work of Bruyère, Meerhoff and Ong (1958a).

Ramus produced works of two main kinds in each subject. Where *Dialecticae institutiones* (later called *Dialectica*) is a textbook, setting out the main doctrines of the subject, *Aristotelicae animadversiones (Notes on Aristotle)* and its successors - later incorporated in the *Scholae in liberales artes (Lectures on the Liberal Arts)* - are works of controversy, attacking Aristotle’s teachings. Discoveries in the controversial works, often prompted by Ramus’ critics, are absorbed in the textbooks, and materials dropped from the textbooks often turn up later in the commentaries to the textbooks, while materials dropped from the commentaries can reappear in the controversial works. What follows is inevitably a statement of the general tendencies of Ramus’ dialectic rather than an analysis of any particular text.

Ramus aimed at clarity, simplicity and usefulness. His textbooks proceed by definition and division and they are full of examples. Ramus’ method obliged him to avoid overlaps between subjects. Accordingly a good deal of the traditional material of Aristotelian dialectic, for example the predicables and the categories, was rejected as belonging to metaphysics. For Ramus dialectic is the art of discoursing well. It consists of invention, which finds the matter for arguments through the topics (see Agricola, R. §2), and judgment, which forms arguments into propositions, syllogisms and larger structures. Within invention Ramus proceeds by a series of dichotomies to the individual topics. His list is shorter than most earlier versions; it is organized into groups and it avoids duplication. So, for example, there is no topic of definition or substance, since Ramus considers that Agricola’s new topic of subject covers the same arguments. Ramus regarded the topics as the classes of arguments (whereas his predecessors had called them the places where the arguments are found) and he left it to his examples to show how particular topical arguments function. He followed Agricola in emphasizing the usefulness of the topics and in illustrating them with passages from poetry and oratory, but he explained much less, preferring a more austere and systematic exposition.

Ramus divided judgment into axiomatic judgment, which concerns the parts, qualities, quantities and connections of the proposition (corresponding to Aristotle’s *De interpretatione*), and dianoetic judgment, which comprises the syllogism (including all three figures, and the hypothetical syllogism) and the overall shape of the piece. At first Ramus called the latter ‘disposition’, indicating a link with rhetoric and with Agricola, but later he called it ‘method’. This integration of dialectical judgment and overall organization represents an improvement on Agricola. Thus, in earlier versions Ramus sets out the imperfect method (or method of prudence) of poets and orators alongside the perfect method (from general to particular), but later the perfect method overwhelms its counterpart. In early versions of the dialectic Ramus gives a worked example of analysing a literary text into its argumentative structure, in the way Agricola had, but later this becomes part of the commentary or an exercise for the teacher. The general tendency is for the dialectic textbook to become briefer and tighter.

For Ramus rhetoric is the art of speaking well (see Rhetoric §3). It involves ornate and correct speaking and skilful delivery. Since invention and disposition belong to dialectic they are not part of rhetoric. Since correct speaking is the concern of grammar, rhetoric can begin with a discussion of the ornate, which consists of a drastically shortened account of the tropes and figures. Thus there are only four tropes (metonymy, irony, metaphor and synecdoche). Figures of diction include the rules of prose and poetic metre as well as a selection of the figures of repetition, while figures of thought are connected with the attitude of the speaker, either alone or in an imagined dialogue with the audience. The discussion of delivery is divided into suggestions on the tones of voice to be used for different effects and on suitable gestures for head and hands. Ramus carried out a thorough reduction and simplification of rhetoric. His reduced and reorganized figures are undoubtedly easier to remember, but there is also a loss of expressive force. In the rush to simplify, many valuable observations were lost along with the repetitions and the points which apply only to the Roman courtroom.

Ramus’ controversial works tend to survey selected texts from Aristotle (for dialectic), and Cicero (§2) and Quintilian (for rhetoric), pointing out what has been omitted and what is superfluous compared with a strict view

of the subject. Ramus nearly always begins by complaining that the author has not defined his subject and divided it. He often criticizes work for confused organization and for including material which is useless or which belongs to other subjects. He attacks other authors for not having written his books, rather than trying to understand what they have to say. As a result he misses Quintilian's wisdom and Aristotle's philosophical subtlety. Still, some of his remarks are well-founded: Cicero's *Orator* is badly organized; Quintilian's *Training in Oratory* does contain a considerable amount of reduplication and a good deal of material that applies mainly to Roman lawcourts; the early books of Aristotle's Organon are more metaphysical than logical, and his *Topics* is very hard to understand. It is also true that many of Ramus' criticisms are based on a consistent position outlined in his theory of method. By the end of the revisions Ramus comes to display a good knowledge of Aristotle's work, much of it owed to his opponents.

3 Method

Ramus regards method as the key to the presentation of all arts and sciences. Method first appears in Ramus' textbook in 1546 as 'the organization of different things in such a way that the whole subject may be more easily perceived and taught'. It is mainly related to the presentation of a subject and seems to derive from Ramus' former teacher Johann Sturm. Sturm had begun his 1539 commentary on Cicero's *Divisons of Rhetoric* by defining method as 'the certain, brief, correct and comprehensive way of teaching an art'. He distinguished three types of method: from general to particular, from particular to general, and by definition and division. Sturm is probably basing himself on Galen, who in turn relies on Plato's and Aristotle's remarks on method. Between them Melanchthon, Sturm, Ramus and the commentators on Aristotle's *Posterior Analytics* raised method to an issue of major significance in the mid-sixteenth century.

Ramus adopted method as the principal form of organization for an art and he developed three laws of method, which appear after 1566 and in the *Scholae*: (1) only things which are true and necessary may be included, (2) all and only things which belong to the art in question must be included, (3) general things must be dealt with in a general way, particular things in a particular way. The first two laws are mainly used to exclude material, the third to prescribe the organization of an art, from the most general principles to the most specific instances. Ramus came to see this method as the true interpretation of Aristotle's *Posterior Analytics* and himself as Aristotle's only true follower. This did not prevent him from using the laws of method to attack the content and organization of Aristotle's own writings. Equally the laws of method gave him the confidence to streamline his own textbooks, so that what began as a humanist project based on literature became (at least on the page) a structured series of definitions and divisions, almost a set of axioms, as in modern formal logic. It also enabled the whole content of a subject to be set out in the famous pictorial dichotomies which were popularized by Ramus and his followers. The austerity of his textbooks was complemented by his emphasis on usefulness and his continuing to comment on literary texts in his own teaching. Ramus' followers produced tree diagrams to express the content and organization of literary and scriptural texts. For the future Ramus increased the importance of method as a subject, emphasized the need to select material, according to what we would now call disciplinary boundaries, and encouraged the critique of Aristotle (see Logic, Renaissance §7).

4 Mathematics, physics and science

Ramus promoted mathematics. In his orations and prefaces he cited the Bible to establish the religious duty of studying mathematics and he argued that it was the oldest of the sciences. He included it in his syllabus reforms and he left a considerable part of his wealth to endow a chair of mathematics in Paris. He insisted on the practical usefulness of mathematics and argued that it should be taught in relation to practice. He proposed a new division of the subject, into the mathematics of *intelligibilia* (arithmetic, from which geometry is also derived) and the mathematics of *sensibilia* (mechanics, astronomy and optics). Although Ramus undoubtedly taught mathematics, and regarded it as an important aspect of his reforms, he did not make original contributions to the subject. He made acute criticisms of the organization of earlier mathematics texts, according to his precepts of method, but he relied on his collaborators to provide original mathematical material.

In his lectures on physics, Ramus rejected Aristotle's conception of the subject, because much of the subject-matter is not found in nature and because of Aristotle's confusions and contradictions. He wanted less on logical issues and more on problems of mechanics, astronomy and biology. Though he preferred to derive these facts from classical texts, such as Pliny's *Natural History* and Virgil's *Georgics*, he called for observation as the

basis of physics and astronomy. In particular he argued for astronomy without hypotheses, an idea which involves a rejection of traditional astronomy, and which Tycho Brahe, whom he consulted, regarded as impossible. Ramus' followers extended his methodical approach to other curriculum subjects. Theodore Zwinger produced an analysis of moral philosophy, and Johannes Freigius wrote a huge *Quaestiones physicae* (1579) which provides a systematic account of learning about the natural world (including acoustics, music, hydrography and geography), drawing on many sources besides Aristotle.

See also: Agricola, R.; Logic, Renaissance; Platonism, Renaissance §5

PETER MACK

List of works

There is no edition (modern or otherwise) of the collected works of Ramus, but the facsimiles which follow include some of the most important texts. Otherwise refer to the bibliographies of Ong (1958a) and Bruyère-Robinet (1986).

Ramus, P. [Ramée, P. la] (1543) *Dialecticae institutiones, Aristotelicae animadversiones (Teaching of Dialectic, Notes on Aristotle)*, Paris; repr. Stuttgart: Frommann, 1964.(Early versions of Ramus' dialectic textbooks and his critique of Aristotle.)

Ramus, P. [Ramée, P. la] (1549) *Rhetoricae distinctiones in Quintilianum*, trans. J. Murphy and C. Newlands, *Arguments in Rhetoric: Against Quintilian*, De Kalb, IL: Northern Illinois University Press, 1986.(Ramus' attack on classical Latin rhetoric.)

Ramus, P. [Ramée, P. la] (1569) *Scholae in liberales artes (Lectures on the Liberal Arts)*, Basle; repr. Hildesheim: Olms, 1970.(Lectures which collect Ramus' criticism of previous writers across the university syllabus.)

Ramus, P. [Ramée, P. la] (1569) *Dialectica*, Basle; English trans. *The Logike*, London, 1574; repr. Menston: Scolar Press, 1970.(First English translation of Ramus' logic.)

Ramus, P. [Ramée, P. la] (1577) *Praefationes, Epistolae, Orationes (Prefaces, Letters, Orations)*, Paris; repr. as P. la Ramée, *Oeuvres Diverses*, Geneva: Slatkine, 1971.(Useful collection of Ramus' prefaces and programmatic writings.)

References and further reading

Bruyère, N. (1984) *Méthode et dialectique dans l'oeuvre de La Ramée (Method and Dialectic in Ramus' Works)*, Paris: Vrin.(Good account of the development of Ramus' dialectic, which overstates his Platonism. Referred to in §2.)

Bruyère-Robinet, N. (1986) 'Les fonds Pierre de la Ramée des bibliothèques de France' (Holdings of Ramus' Works in French Libraries), *Nouvelles de la République des lettres* 6: 71-97.(A survey of locations for early editions of Ramus' works.)

Bruyère-Robinet, N. (ed.) (1986) *Revue des sciences philosophiques et théologiques* 70 (1), special issue.(Useful collection of recent articles about Ramus.)

Hooykaas, R. (1958) *Humanisme, Science et réforme: Pierre de la Ramée*, Leiden: Brill.(A useful overview of Ramus' approach to mathematics and science.)

Mack, P. (1993) *Renaissance Argument: Valla and Agricola in the Traditions of Rhetoric and Dialectic*, Leiden: Brill, 334-55.(Details Agricola's influence on Ramus.)

Meerhoff, K. (1986) *Rhétorique et poétique au XVIe siècle en France*, Leiden: Brill, 175-348.(Excellent account of Ramus' rhetoric, with particular attention to prose rhythm, French poetics and Latin rhetoric. Referred to in §2.)

Ong, W. (1958a) *Ramus and Talon Inventory*, Cambridge, MA: Harvard University Press.(Useful list of editions. Referred to in §2.)

Ong, W. (1958b) *Ramus, Method and the Decay of Dialogue*, Cambridge, MA: Harvard University Press.(Useful account but excessively dismissive of Ramus.)

Risse, W. (1964) *Die Logik der Neuzeit. I Band. 1500-1640 (The Logic of the Modern Period)*, Stuttgart and Bad Cannstatt: Frommann.(A survey of sixteenth-century schools of logic.)

Sharratt, P. (1986) 'Recent Work on Petrus Ramus', *Rhetorica* 5: 7-58.(An invaluable guide to recent studies, which extends his earlier survey in *Studi francesi* 47-8: 201-13.)

Sharratt, P. (ed.) (1991) *Argumentation* 5 (4), special issue.(Collection of articles on Ramus.)

Vasoli, C. (1967) *La dialettica e la retorica dell'Umanesimo (Humanist Dialectic and Rhetoric)*, Milan:

Feltrinelli.(A very good account of humanist rhetoric and dialectic, including a section on Ramus.)
Verdonk, J.J. (1966) *Petrus Ramus en de wiskunde (Petrus Ramus and the Sciences)*, Assen: Van Gorcum.(The best account of Ramus' contribution to mathematics.)
Waddington, C. (1855) *Ramus: sa vie, ses écrits et ses opinions (Ramus: his Life, his Writings and Opinions)*, Paris: Meyrveis.(The fullest Ramus biography.)

Rand, Ayn (1905-82)

Ayn Rand was a Russian-born US novelist and philosopher who exerted considerable influence in the conservative and libertarian intellectual movements in the post-war USA. Rand's ideas were expressed mainly through her novels; she set forth a view of morality as based in rational self-interest and in political philosophy defended an unrestrained form of capitalism.

Ayn Rand was born Alyssa Rosenbaum into a middle-class Jewish family in St Petersburg. Her family's expropriation by the Bolsheviks and subsequent poverty had a profound effect on her; her first novel, *We the Living* (1936), describes the tragedy of a Russian student struggling against an evil society in the 'vast prison' that was the USSR in the 1920s. Her work was marked not only by a hostility to communism but also by a strong antipathy towards any form of compromise among competing values.

Popular success came in 1943 with the publication of her philosophical novel, *The Fountainhead*, the story of an architect who refuses to compromise his independence or his integrity while good people despair in the face of evil. A deeply moral work, its theme is integrity which, for Rand, was at the root of the idea of freedom. Even greater success came with *Atlas Shrugged* (1957), Rand's final work of fiction. More explicitly political than her earlier work, it tells of the breakdown of a society of evil as the captains of capitalist industry withdraw from a world marked by political and moral corruption. As with her earlier works, the hero is uncompromising in his integrity and confidence in the value of the moral path; the bulk of the novel charts the rise of those who begin in despair. However, Rand also turns more explicitly to philosophical problems in ethics in an attempt to set morals on a more secure epistemological footing. The book contains many long philosophical speeches by characters speaking for Rand.

The popular success of her fiction brought discipleship and the 1960s and 1970s saw the growth of an 'objectivist' movement. The influence of Rand's ideas was strongest among college students in the USA but attracted little attention from academic philosophers. Her outspoken defence of capitalism in works like *Capitalism: The Unknown Ideal* (1967), and her characterization of her position as a defence of the 'virtue of selfishness' in her novel of the same title (published in 1974), also brought notoriety, but kept her out of the intellectual mainstream.

The central philosophical argument of Rand's thought is an attempt to show that the good life is itself a substantial ethical value from which may be derived important moral conclusions. In this she is self-consciously Aristotelian, although most commentators have concluded that her argument falls victim to the same difficulties, relying on a morally substantive and controversial account of human nature to generate ethical conclusions.

Rand's political theory is of little interest. Its unremitting hostility towards the state and taxation sits inconsistently with a rejection of anarchism, and her attempts to resolve the difficulty are ill-thought out and unsystematic. Of more enduring interest is her fiction, belonging to a genre she labelled 'romantic realism'. Despite her attack on altruism and insistence on the virtue of selfishness, her real concerns were the defence of the value of integrity (to the point of self-sacrifice) in the face of evil and moral despair.

CHANDRAN KUKATHAS

List of works

Rand, A. (1936) *We the Living*, New York: Macmillan.
Rand, A. (1943) *The Fountainhead*, New York: Bobbs-Merrill.
Rand, A. (1957) *Atlas Shrugged*, New York: Random House.
Rand, A. (1967) *Capitalism: The Unknown Ideal*, New York: New American Library.
Rand, A. (1974) *The Virtue of Selfishness*, New York: New American Library.
Rand, A. (1990) *Introduction to Objectivist Epistemology*, New York: New American Library.

References and further reading

Den Uyl, D.J. and Rasmussen, D. (eds) (1984) *The Philosophic Thought of Ayn Rand*, Chicago, IL: University of Illinois Press.

Merrill, R.E. (1993) *The Ideas of Ayn Rand*, Chicago and La Salle, IL: Open Court.

Randomness

The fundamental intuition underlying randomness is the absence of order or pattern. To cash out this intuition philosophers and scientists employ five approaches to randomness.

(1) Randomness as the output of a chance process. Thus an event is random *if it is the output of a chance process. Moreover, a sequence of events constitutes a* random sample *if all events in the sequence derive from a single chance process and no event in the sequence is influenced by the others.*

(2) Randomness as mimicking chance. Statisticians frequently wish to obtain a random sample (in the sense of (1)) according to some specified probability distribution. Unfortunately, a chance process corresponding to this probability distribution may be hard to come by. In this case a statistician may employ a computer simulation to mimic the desired chance process (for example, a random number generator). Randomness qua *mimicking chance is also known as* pseudo-randomness*.*

(3) Randomness via mixing. Consider the following situation: particles are concentrated in some corner of a fluid; forces act on the fluid so that eventually the particles become thoroughly mixed throughout the fluid, reaching an equilibrium state. Here randomness is identified with the equilibrium state reached via mixing.

(4) Randomness as a measure of computational complexity. Computers are ideally suited for generating bit strings. The length of the shortest program that generates a given bit string, as well as the minimum time it takes for a program to generate the string, both assign measures of complexity to the strings. The higher the complexity, the more random the string.

(5) Randomness as pattern-breaking. Given a specified collection of patterns, an object is random if it breaks all the patterns in the collection. If, on the other hand, it fits at least one of the patterns in the collection, then it fails to be random.

1 Chance

The most common conception of randomness identifies randomness with chance. Indeed, much of probability theory and statistics does not distinguish the two. Thus for a probabilist or statistician a *random event* and an *event due to chance* are typically the same thing. Moreover, processes giving rise to such events are referred to indiscriminately as *random*, *chance*, or *stochastic* processes.

Within statistics the adjective ‘random’ assumes a technical sense when it occurs in the phrase ‘random sample’. Given a chance process, one may wish to consider not just a single random event from this process, but rather an entire sequence of such events. Such a sequence is then said to constitute a random sample if: first, the same chance process is responsible for each event in the sequence; and, second, the occurrence of any event in the sequence is unaffected by the occurrence of other events in the sequence. If the first condition is satisfied one says that the events are ‘identically distributed’; if the second, that the events are ‘independent’.

Identifying randomness with chance now raises the obvious question, what is chance? While there exists a metaphysics of chance related to causation, determinism and free will, this entry approaches chance instrumentally in terms of those processes, like coin tossing and radioactive decay, for which our best understanding is irreducibly probabilistic (see Chaos theory for comparison, where the probabilities are artifacts of underlying deterministic systems).

2 Simulation

Scientific research consists increasingly of computer simulations that generate vast amounts of data. Presumably, if scientists had sufficient time and resources to examine nature directly, computer simulations that imitate nature would be unnecessary. Practical limitations, however, seem to render computer simulations indispensable to scientific research.

A dilemma now confronts the scientist. For many purposes the data a scientist wishes to obtain should properly be the output of a chance process characterized by some well-defined probability distribution. Practical limitations,

however, often prevent the scientist from actually sampling such a process and obtaining the desired data set (imagine a scientist who desires as data the sequence of heads and tails attained by flipping a coin one billion times - the scientist will expire before the billion flips can be accomplished). In this case, the scientist will want to simulate the chance process computationally.

The dilemma then is this. On the one hand computers are fully deterministic devices - specify an algorithm, and the behaviour of the machine is fixed. It follows that any probabilistic features of the data generated by a computer simulation are strictly eliminable. Yet, on the other hand, such data are to substitute for data generated by a genuine chance process, data which cannot be characterized except in probabilistic terms. As the output of a chance process, truly random data (in the sense of §1) are supposed to defy all but *post hoc* characterizations. As the output of an algorithm the (pseudo-) random data generated by a computer simulation are fully characterized in advance. How can the twain meet?

Strictly speaking they cannot. If randomness is identified with chance, then an event is random just in case it has the right sort of causal history and it was generated by a chance process. A computer is not a chance process. Ergo the data generated by a computer cannot be random. John von Neumann summed up the matter: ‘ Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin’ (Knuth 1981: 1).

Nonetheless, the incongruity of using not merely deterministic systems, but systems whose entire behaviour can be precisely specified in advance has not dampened the proliferation of random number generators (RNGs). What then justifies using the data generated in a computer simulation in place of data generated by a chance process? In practice what happens is this. Given an RNG statisticians, as it were, set up a gauntlet of statistical tests that serve to vet it. The tests specify properties which the overwhelming majority of numerical sequences should have if they were generated by the chance process that the RNG is attempting to mimic. If the numerical sequences generated by the RNG do not have these properties, the RNG is rejected. Otherwise it is considered adequate.

Although in practice RNGs do a lot of useful work, there remains a theoretical problem in justifying RNGs in this way (which unfortunately is the only way RNGs can be justified): any RNG is only as good as the last statistical test that it passed. Indeed, history is strewn with RNGs that were for a time considered adequate, and then shown to be deficient. The problem is that we can never be sure that an RNG incorporates biases which the statistical tests we have thrown at it have simply failed to detect. This is bad. Practically speaking this means that the scientific literature may contain errors which we shall be unable to root out until appropriate statistical tests are found to detect the biases. For instance, cosmologists whose computer simulations of the early universe rely on RNGs may find their models overturned if the RNGs they employ are subsequently found to be badly biased.

3 Mixing

Take a fresh deck of playing cards and begin to riffle-shuffle them. How many riffle-shuffles are required before the deck is thoroughly mixed? Persi Diaconis has shown that seven riffle-shuffles are needed. What it means for the deck to be thoroughly mixed is that any configuration of the deck is as likely as any other. The deck starts in a specified configuration. A single shuffle mixes the deck, but not enough to break all connection with the previous configuration. Only after multiple shuffles does the configuration of the deck lose its connection with the starting configuration. At this point one says that the deck has attained a random state.

Shuffling a deck of cards is an example of a group action. Group actions provide one way of mixing things up, but not the only way. Imagine a gas concentrated in one corner of a box. The particles that make up the gas are in motion. Over time the gas will reach an equilibrium state, filling the entire box uniformly. This is an example of a dynamical system from statistical mechanics. The system starts out in a low entropy state, with the particles concentrated in one corner, and eventually reaches a maximal entropy state (an equilibrium state), with the particles evenly distributed throughout the box. The system is said to be random once it reaches the maximal entropy state (see Thermodynamics).

The preceding examples illustrate several features that are common to systems which attain randomness via mixing: (1) such a system starts out from a specified configuration that is highly ordered or constrained (for example, the opposite of what we would intuitively want to call random); (2) a mixing process (for example, a group action) acts on the system, over time continually transforming the configuration of the system; (3) eventually an equilibrium state is reached after which further mixing does not affect the equilibrium. When the

equilibrium state is reached, the system is said to be random.

It is worth noting that uniform probabilities frequently characterize the equilibrium states signalling randomness. What it means for a deck of playing cards to be thoroughly shuffled is that no configuration of the deck is more likely than any other. Shuffling has therefore randomized the deck only if each possible configuration of the deck is equiprobable. Similarly, a gas within a box has reached equilibrium if temperature throughout the box is uniform and the particles are evenly distributed. Uniform probabilities are therefore intimately connected with this understanding of randomness.

4 Complexity

Consider the following two sequences of a hundred coin tosses (heads = 1, tails = 0):

(R) 1100001101011000110111111
1010001100011011001110111
0001100100001011110111011
0011111010010100101011110

and

(N) 1111111111111111111111111
1111111111111111111111111
1111111111111111111111111
1111111111111111111111111

It seems clear that any theoretical account of randomness had better make (R) more random than (N). For instance, since (R) was obtained by actually flipping a coin whereas (N) was artificially contrived, according to the causal account of randomness sketched in §1, (R) would be random, but (N) nonrandom.

If, however, we prescind from our knowledge of the causal process responsible for these sequences, could we still distinguish these sequences in terms of randomness? We could, for instance, try to find a statistical test whose rejection region includes (N) and excludes (R), and thereby justify calling (N) nonrandom and (R) random. But for every such test it is possible to find a corresponding test whose rejection region includes (R) and excludes (N). Nor do probabilities help distinguish the sequences, since both (R) and (N), and indeed all such sequences of length 100, have the same small probability of occurring by chance, namely, 2^{-100} or approximately 1 in 10^{30}.

Starting in the 1960s a group of researchers that included Gregory Chaitin and Andrei Kolmogorov proposed a way around these difficulties (Earman 1986). Instead of characterizing randomness probabilistically, they took the very different tack of characterizing randomness computationally. What they said was that a string of 0s and 1s becomes increasingly random as the shortest computer program that generates the string becomes increasingly long. In the 1980s cryptographers proposed a variant of this characterization: a string of 0s and 1s becomes increasingly random as the most efficient computer program that generates the string requires increasingly long computation times. The first approach characterizes randomness in terms of space complexity (that is, the amount of memory the program occupies); the second, in terms of time complexity (that is, the computation time the program requires). The space complexity approach to randomness is referred to in the literature as 'algorithmic information theory'.

It is now intuitively obvious why (R) is more random than (N). The shortest program that computes (N) has the form 'repeat "1" 100 times'. On the other hand, (R) seems to have no shorter description than the string itself. (N) can be drastically compressed, (R) cannot. Thus from the point of view of algorithmic information theory (R) is more random than (N) (see Information theory).

Although complexity approaches to randomness represent a genuine advance in the theoretical study of randomness, there is a limitation to these approaches that is often lost in the initial enthusiasm: all complexity approaches to randomness are relativized to a given computational environment. What this means is that even though a sequence may be random when its generating program is running in PASCAL on a standard mainframe computer, with respect to another computational device it may be nonrandom, and vice versa. In fact, since mappings between finite sets are always computable (recursion theory on finite sets is trivial), any finite string will

be random with respect to certain programming environments, nonrandom with respect to others (see Computability and information; Computability theory).

5 Pattern-breaking

Having now surveyed four distinct approaches to randomness, one is tempted to ask whether a common thread runs through these approaches? There is a common thread, but one that at first sight will seem counterintuitive. In a dictionary definition of 'randomness', the term characterizes objects or events brought about without method, plan, purpose, forethought, pattern, principle, order or design. Random objects are supposed to be higgledy-piggledy, evincing no patterns.

But what does it mean for an object to evince no patterns? Consider a spy who eavesdrops on a communication channel in which encrypted messages are being relayed. If the spy has yet to break the cryptosystem, the encrypted messages traversing the communication channel will, as far as the spy is concerned, fail to display any patterns. Yet as soon as the cryptosystem is broken, all the patterns hidden by the cryptosystem become apparent.

The point is this: what determines the patterns that must be broken for an object to be random is not some objective feature of the world - randomness is not a natural kind. Rather, what is random depends on the patterns that are specified within a given context and that must then be broken for an object to be random. What is counterintuitive about this approach is that randomness becomes thoroughly parasitic on the patterns with respect to which it is defined. Randomness on this view does not make sense until a given collection of patterns is specified.

How then does this pattern-breaking approach to randomness relate to the four preceding approaches? For the computational complexity approach to randomness, the low complexity programs specify the patterns. For the mixing approach to randomness, far-from-equilibrium-states specify the patterns. For the simulation approach to randomness, statistical tests specify the patterns. The pattern-breaking approach to randomness also makes clear why chance is so often a dependable route to randomness: in many applications the patterns specified in advance identify a set of very small probability (for example, a full complement of statistical tests used to vet an RNG will typically designate as nonrandom only a tiny proportion of possible numerical sequences). Since small probability events are rare, chance will typically deliver objects or events that break all such patterns, that is, objects that are random in the pattern-breaking sense.

WILLIAM A. DEMBSKI

References and further reading

Bayer, D. and Diaconis, P. (1989) 'Trailing the Dovetail Shuffle to its Lair', Technical Report 329, Stanford, CA: Department of Statistics, Stanford University.(See §3.)

Borel, E. (1963) *Probability and Certainty*, trans. D. Scott, New York: Walker.(Treats the problem of small probabilities. See §5.)

Dembski, W.A. (1991) 'Randomness by Design', *Nous* 25 (1): 75-106.(See §5. An exposition of the pattern-breaking approach to randomness.)

Diaconis, P. (1988) *Group Representations in Probability and Statistics*, Hayward, CA: Institute of Mathematical Statistics.(The road to randomness via group actions. See §3.)

Earman, J. (1986) *A Primer on Determinism*, Dordrecht: Reidel.(Gives a general treatment of algorithmic information theory. See §4.)

Garey, M.R. and Johnson, D.S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*, New York: Freeman.(An introduction to computational complexity in which computation time serves as the measure of complexity. See §4.)

Hacking, I. (1965) *Logic of Statistical Inference*, Cambridge: Cambridge University Press.(A good discussion of what statisticians mean by randomness in the sense of chance. See §1.)

Knuth, D.E. (1981) *Seminumerical Algorithms*, 2nd edn; in *The Art of Computer Programming*, vol. 2, Reading, MA: Addison-Wesley.(Includes Knuth's classic treatment of random number generators. See §2.)

Kranakis, E. (1986) *Primality and Cryptography*, Stuttgart: Wiley-Teubner.(Treats random number generators as well as the time complexity approach to randomness. See §§2, 4.)

Lambalgen, M. van (1989) 'Algorithmic Information Theory', *Journal of Symbolic Logic* 53 (4): 1389-400. (Technical treatment of algorithmic information theory.)

Lasota, A. and Mackey, M.C. (1985) *Probabilistic Properties of Deterministic Systems*, Cambridge: Cambridge University Press.(The road to randomness via dynamical systems. See §3.)
Mises, R. von (1957) *Probability, Statistics, and Truth*, New York: Dover, 2nd edn.(Develops many seminal ideas about randomness; although much of what von Mises did has since had to be modified or corrected.)

Rashdall, Hastings (1858-1924)

Hastings Rashdall was a utilitarian in ethics, an idealist in metaphysics and a Christian monotheist in religion. His history of medieval universities became a classic. His revisions of utilitarianism are an important part of the development of that ethical theory beyond its original version in Bentham and Mill. His religious metaphysic strongly opposed the influential nonmonotheistic idealism of F.H. Bradley.

Born in 1858, and educated at Harrow and New College, Oxford, Hastings Rashdall was a lecturer in St David's College (for clergy), a theology tutor at University College, Durham, chaplain and divinity tutor at Balliol College, and Fellow, tutor and dean at New College. He became a canon of Hereford Cathedral in 1910, and dean of Carlisle in 1917. He died in 1924.

In ethics, Rashdall rejected the hedonism of Jeremy Bentham and John Stuart Mill. He none the less accepted the general utilitarian perspective that Bentham and Mill developed. He agreed with them that an action is right only if it best serves (or, in the worst case, least harms) human wellbeing. But he refused to define the good that action should pursue in terms of pleasure alone; his utilitarianism is hence ideal rather than hedonistic. He held that rational moral judgments are possible concerning the moral worth of particular ends or goals, and it is these, rather than moral intuitions of general principles defining the nature of right action independently of consequences, that should be basic in moral reflection. A tendency to promote human wellbeing is part of the good that our actions should promote.

Rashdall held that there is a deep link between morality and monotheism:

> Belief in God… is the logical presupposition of an 'objective' or absolute Morality. Our moral ideal can only claim objective validity in so far as it can rationally be regarded as the revelation of a moral ideal eternally existing in the mind of God.
>
> (1907, vol. 2: 212)

In metaphysics, Rashdall was an idealist, maintaining that only minds and their states and characteristics exist, and that there are no spatially extended and consciousness-independent physical objects. Because, as Berkeley (§§3, 5-7) argued, neither secondary qualities nor primary qualities can exist without consciousness, what we refer to as matter is the perceptual content of mental states.

Our efficacious willing is the only causality that we experience, and thus order in nature (that is, among our perceptual states) is best explained by reference to a Mind that causes regularities among our experiences. By analogy with our minds, presumably that Mind also has thought, feeling and will. Each mind is a mental substance, and none is to be thought of as merely a mode of a divine mind.

In philosophy of religion, Rashdall was a monotheist, maintaining that God is a self-conscious, independently existing mental substance on whom all other minds depend for their existence. Our knowledge of God comes, not from religious experience, but from the arguments of natural theology, particularly the moral argument for God's existence, and from revelation in so far as it is supported by reason (see Religion, history of philosophy of §5). Our moral consciousness makes us aware that only if a preventable evil serves a greater good is its existence morally permissible.

Rashdall held that there are goods for whose existence evil is necessary: for example, he suggested that evolving 'highly organized beings without a struggle for existence' and training 'human beings in unselfishness without allowing the existence of both sin and of pain' was impossible (1909: 82).

It was Rashdall's contention, then, that there are goods that not even God can elicit without allowing evil. This, he said, follows from God's nature. He seems to have meant two things. First, he meant that regarding some actual goods *G* and evils *E*, '*G* exists and *E* does not' is self-contradictory and *G* outweighs *E*; some evils are allowed because they are logically necessary conditions for important goods. Second, he meant that in some manner the ground of logical necessity lies in God's nature. No divine defect was implied when Rashdall said that it follows from God's nature that not even God can elicit certain goods without also allowing evil. God, being good, would not have created at all if evil were going to triumph.

Rashdall was a determinist: he held that human actions occur as the result of factors over which their agents have no effective control. Hence the existence of evil cannot even in part be ascribed to free and wrong human choices. God, however, is not blameworthy for the evil in the world, for God must allow certain evils in order for there to be certain goods.

See also: Evil, problem of

KEITH E. YANDELL

List of works

Rashdall, H. (1895) *The Universities of Europe in the Middle Ages*, London: Oxford University Press. (Rashdall's classic study of medieval universities.)

Rashdall, H. (1905) *Conscience and Christ*, London: Duckworth.(Six lectures on Christian ethics.)

Rashdall, H. (1907) *The Theory of Good and Evil*, London: Oxford University Press, 2 vols. (The fullest statement of Rashdall's ideal utilitarianism.)

Rashdall, H. (1909) *Philosophy and Religion*, London: Duckworth.(Six lectures on metaphysics, ethics and revelation.)

Rashdall, H. (1912) *The Problem of Evil*, Manchester: Manchester University Press. (Rashdall's discussion of the thorniest problem for monotheists.)

Rashdall, H. (1920) *The Moral Argument for Personal Immortality*, London: Duckworth. (Rashdall's discussion of the moral argument for the existence of God.)

Rashdall, H. (1930) *God and Man*, Oxford: Blackwell.(Chapters, historical and systematic, on various issues in philosophical theology.)

References and further reading

Elliot-Binns, J.E. (1956) *English Thought 1860-1900: The Theological Aspect*, London: Longmans, Green & Company.(Discusses the historical, philosophical, theological and social setting of English theology in the indicated period.)

Langford, T. (1969) *In Search of Foundations: English Theology 1900-1920*, Nashville, TN: Abington Press. (Presentation of English theological thought in its cultural and philosophical context; good bibliography.)

Mozley, J.K. (1951) *Some Tendencies in British Theology*, London: SPCK. (Discussion of British theology from *Lux mundi* (1889) to the thought of Oman.)

Rational beliefs

To the extent that a belief is rational, it ought to be held, other things being equal; irrational beliefs should not be held. From traditional epistemological perspectives, the obligation here is narrow, concerning only good reasons for acceptance that constitute sufficient justification or warrant. Recent epistemological trends broaden the viewpoint to include also practical considerations that enter into other rational decisions, such as best use of the agent's limited resources.

A related but weaker conception of rationality appears in philosophy of mind as a necessary coherence requirement on personal identity - roughly, 'No rationality, no agent'. Such agent-constitutive rationality standards are more lenient than normative epistemic standards, since agents' belief sets can and often do fall short of epistemically uncriticizable rationality without the agents thereby ceasing to qualify as having minds.

Finally, at the widest perspective, long-standing sceptical lines of challenge to rationality of the entire structure of human belief-forming procedures conclude that we can never have the slightest good reason to accept even our most central beliefs. Recent approaches that 'naturalize' epistemology into a branch of science tend to exclude such general doubts as insignificant or meaningless; but if distinctively philosophical questions in fact do not fully reduce to regular scientific ones, sceptical-type rationality challenges may instead remain a permanent part of the human condition.

1 A priori belief

What rational belief is can be best conveyed through some major types of example. The simplest, and most extreme, case of irrationality is accepting a proposition that is inconsistent. Indeed, attribution to a person of such irrationality has sometimes been regarded as self-defeating, in that we purportedly cannot even make sense of so severe a reasoning failure. In this way charity principles of Quine and Davidson dictate that 'better translation imposes our logic' - and our rationality generally - upon the beliefs of any agent we undertake to interpret (see Charity, principle of §4). Such in-principle exclusion of the very possibility of inconsistency may seem like an overreaction, since human beings (unlike God) must often lack the insight to identify the less obvious inconsistencies in propositions they accept. To idealize away these failures as mere occasional, corrigible slips is to deny a basic fact of existence for a finite creature, human or otherwise, and to deny much of the project of the deductive sciences. We understand all too well how Gottlob Frege could genuinely accept his axioms for naïve set theory, even though they turned out to harbour Russell's Paradox; inconsistency is always with us (see Frege, G. §§8-9; Russell, B. §4).

Granting the possibility (indeed, pervasiveness) of inconsistencies, they are, as paradigmatic irrationality, conventionally viewed as disasters to be avoided or remedied at all costs. As Frege responded to Russell on the status of inconsistent foundations of mathematics, 'Arithmetic totters'. Indeed, the history of the rise of modern logic can be viewed as conditioned in large part by the discovery of, and response to, the paradoxes. However, in contrast to the traditional picture of antinomies as symptoms of disease, alternative accounts have emerged along the lines of Wittgenstein's apophthegm, 'A contradiction is not a germ which shows a general illness'. Some of these go further, to propose that inconsistency may in fact sometimes be positively healthy, as in the familiar way that inconsistent naïve set theory can be easier to use for informal purposes than later, more complicated formulations devised to avoid the paradoxes.

Psychological studies of recent decades provide some evidence that everyday human reasoning - inductive as well as deductive - might be formally incorrect to a surprising extent (see Rationality of belief; Common-sense reasoning, theories of). From the perspective of more 'realistic' theories of rationality that take account of an agent's actual cognitive processes and resources, such incorrectness need not be mysterious reasoning pathology. Such local irrationality could be made sense of as the trade-off cost of a globally rational strategy to avoid cognitive paralysis: use of 'quick but dirty', formally incorrect heuristic procedures that are more computationally efficient in many contexts than formally adequate ones can be. These accounts suggest that human beings have a multiplicity of overlapping special-purpose reasoning strategies - some formally correct, some not - that they can switch among. In this way, reasoning is evaluated in terms of, not only preservation of truth, but also wider practical issues of, for example, actual human usability of procedures given a limited cognition-budget. Paradoxes

such as Russell's then appear not necessarily as symptoms of dysfunction at the core of the human conceptual scheme, but rather as signs of fundamentally rational speed-reliability trade-offs.

2 A posteriori belief

For the empirical domain, accounts of rational belief develop along lines independent of, but still parallel to, the evolution of accounts for a priori belief. At the start of modern epistemology, Descartes proposes among rules for direction of the mind, 'I should hold back my assent from opinions which are not completely certain and indubitable just as carefully as I do from those which are patently false' (Descartes [1641] 1984: 12). He then points out that, for any belief, wide-scale sceptical doubts are always possible that would undermine any of my reasons for accepting it - for example, I might be experiencing a perfectly realistic hallucination rather than anything of a mind-independent world. Consequently, Descartes recommends a method of universal doubt, reconstructing our entire belief system from a blank slate; rational belief-management strategy must start by rejecting everything (see Descartes, R. §4; Doubt §2). However, history's majority verdict seems to be that universal doubt is not in fact advisable, that it would leave us in an epistemological tragicomedy of total, irreversible cognitive paralysis, which is rarely a rational option (see Scepticism).

Pragmatism first introduces and defines itself in terms of rejection of the Cartesian programme. Its inventor, C.S. Peirce, proposes instead 'critical commonsensism': while no belief can be immune to doubt, we can never begin with suspension of all our beliefs (see Commonsensism). From this anti-Cartesianism stems a neo-pragmatist leitmotif of much epistemology of the twentieth century. Quine (1960), for example, following Neurath, explains that Cartesian universal doubt is not only a psychological impossibility for human beings, but also self-defeating because self-undermining; to try to practise it is from the outset to sink the scientific-philosophical enterprise. The thesis that, while we must begin from whatever belief scheme with which we find ourselves, any belief can still be doubted, raises a natural next question: *Which* beliefs ought in fact to be challenged, and when?

One line of reply is that an agent is held responsible for ruling out those counter-possibilities to a belief that satisfy a 'special reasons requirement': specific, relevant ones that there is definite basis for thinking might now apply. In an example of J.L. Austin's, normally it would be humorous to criticize as hasty someone's belief that there is a goldfinch in the garden on grounds that he had not checked that the objects he glimpsed were not stuffed toys - unless, for instance, he had already been in a position to find out the neighbours' children had lately been playing with toy birds. In this way, epistemic responsibility resembles traditional doctrines of general culpable negligence, of legal responsibility for taking the due care in actions that a 'reasonable person' would. However, a thesis that the epistemic agent must take due care that falls somewhere between total irresponsibility and perfection in turn raises questions about how particular levels of required vigilance are set. Just because people in certain situations are in fact found subject to criticism if they do not maintain certain levels of care does not completely answer this question - unless actual practices are guaranteed to be optimal (see §3 below).

Thus, there has been an evolution from a simple if perfectionistic view that beliefs cannot be rational if *any* challenges to them are imaginable, to a more realistic or pragmatic type of account, where rationality of beliefs is subject to challenge only when a circumscribed set of appropriate counter-possibilities can be raised. However, if the agent is responsible for eliminating challenges whose seriousness is obviously implied by the current belief system, then formation of new rational beliefs crucially depends upon the agent's current background beliefs. Yet the background belief-set must inevitably be incomplete, particularly at cutting edges of the belief system, themselves inevitable for finite creatures like ourselves. And so breakdown scenarios can still arise where appropriateness of a counter-possibility cannot be evaluated. Even if recognition of a counter-possibility-filter in our knowledge-gathering processes may block the traditional embarrassment of sceptical paralysis, history still threatens to repeat itself in a new guise.

3 Philosophy, science and ordinary life

More generally, we seem unable to avoid posing distinctively philosophical challenges regarding adequacy of our doubt-screening procedures themselves. A question inappropriate in ordinary or scientific contexts may still not be inappropriate *simpliciter*. Rationality consists of more than just playing according to the rules of our belief-forming games; an essential part of rationality concerns the advisability of deciding to play the games in the first place. Recognition of the latter type of rationality-domain can be discerned in Kant's distinction between the

concepts of (scientific) understanding and the extra-scientific ideas of reason that guide the scientific enterprise (see Kant, I. §§6-7), in Carnap's separation of internal questions of science from external questions concerning acceptance of a whole framework for settling internal questions (see Carnap, R. §5), and in Kuhn's distinguishing 'normal' scientific activity from debate about paradigms or disciplinary matrices within which that activity proceeds (see Kuhn, T. §§2-3). *Prima facie*, we seem able to put our most basic knowledge-gathering mechanisms themselves in question and ask whether the actual is the ideal: just because a procedure is part of the system we in fact actually use, does it follow that the procedure must be perfect or otherwise uncriticizable?

One response to the intractability of such philosophical questions is to make a virtue of necessity, and to exchange impending scepticism for relativism. Basic principles cannot themselves be rationally justified without courting circularity; hence, their acceptance can be no more than a matter of epistemic taste, about which ultimately one cannot dispute or judge (see Cognitive pluralism §2; Epistemic relativism). Our own procedures may happen to be the only ones we have encountered, but they ought not to serve as a Procrustean Bed into which genuine cognitive differences must be forced. Without parochial a priori constraints on rationality, possibilities for radical cognitive diversity become wide open. However, questions in turn arise about whether radical relativism itself can even be coherently articulated. In their extreme form, these concerns tend towards a conservative conclusion that also aims to block sceptical-type challenges. One recurring argument to this effect (used, for example, by Quine) is that there is no cosmic exile from which we can take the God's-eye view required to evaluate our own basic conceptual scheme; we need a *pou sto*, a standpoint from which and by which to judge our scheme, and our actual one is the only one we have. External questions perforce become internal (see Rationality and cultural relativism).

However, one reply has been that, while we may have to employ the only system we possess, our system itself may furnish the basis for large-scale doubts about its own adequacy. For example, 'naturalized' epistemology seeks vindication of our actual procedures in a core part of the scientific worldview, evolutionary theory: natural selection has engineered cognitive structures of human beings just as it has the biological structures, and the validation of our belief-forming systems is that evolution has designed them to function well. But then in turn a natural concern is that what is efficient for information-processing tasks conducive to survival need not be epistemically sound - for instance, as mentioned in §1 above, speed and correctness are often outright antagonistic. Thus, naturalism can engender its own challenges to our scheme; such self-evaluations do not guarantee happy endings. Apparently, when sceptical challenges are eliminated by attempting to assimilate distinctively philosophical questions to those of science or ordinary life, they reintroduce themselves in new form. In this way, articulation of our conception of rationality seems a task never completed, only set aside.

See also: Justification, epistemic; Naturalized epistemology; Reasons for belief

CHRISTOPHER CHERNIAK

References and further reading

Austin, J. (1946) 'Other Minds', in *Philosophical Papers*, Oxford: Oxford University Press, 3rd edn, 1990.(The 'special reasons requirement' on challenges of rationality of empirical knowledge claims.)

Carnap, R. (1956) 'Empiricism, semantics, and ontology', in *Meaning and Necessity*, Chicago: University of Chicago Press, 2nd edn.(A positivist account of the science-philosophy division, with Kantian affinities.)

Cherniak, C. (1986) *Minimal Rationality*, Cambridge, MA: MIT Press.(Exposition of a 'resource-realistic' model of rationality.)

Descartes, R. (1641) *Meditations*, in J. Cottingham, R. Stoothoff, and D. Murdoch (eds) *Philosophical Writings of Descartes*, vol. 2, Cambridge: Cambridge University Press, 1984.(On the method of universal doubt, which pragmatism attacks. See especially Meditation I.)

Kant, I. (1783) *Prolegomena to Any Future Metaphysics*, trans. L. Beck, New York: Bobbs-Merrill, 1950.(On the relation between the domains of science and philosophy. Sections 40-4 and 56-7 are especially relevant; see §3 above.)

Peirce, C.S. (1868) 'Some consequences of four incapacities', in *Collected Papers*, vol. 5, Cambridge, MA: Harvard University Press, 1932.(The original statement of the pragmatist anti-Cartesian doctrine of 'critical commonsensism'; see §2 above.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(The starting point of much recent neopragmatist epistemology, and naturalized epistemology in particular. Chapter 1 is especially relevant.)

Simon, H. (1982) *Models of Bounded Rationality*, vol. 2, Cambridge, MA: MIT Press. (An early and important

critique of extreme rationality-idealizations of the agent in economics. See especially chapters 8.2 and 8.4.)
Stich, S. (1990) *The Fragmentation of Reason*, Cambridge, MA: MIT Press.(A relativist theory of rationality; a good overview.)
Tversky, A. and Kahneman, D. (1974) 'Judgment under uncertainty: heuristics and biases', *Science* 185 (4157): 1124-31.(Experimental studies of human use of formally incorrect inductive reasoning procedures; see §1 above.)

Rational choice theory

Rational choice theory is the descendant of earlier philosophical political economy. Its core is the effort to explain and sometimes to justify collective results of individuals acting from their own individual motivations - usually their own self interest, but sometimes far more general concerns that can be included under the rubric of preferences. The resolute application of the assumption of self-interest to social actions and institutions began with Hobbes and Machiavelli, who are sometimes therefore seen as the figures who divide modern from early political philosophy. Machiavelli commended the assumption of self interest to the prince; Hobbes applied it to everyone. Their view of human motivation went on to remake economics through the work of Mandeville and Adam Smith. And it was plausibly a major factor in the decline of virtue theory, which had previously dominated ethics for many centuries.

Game theory was invented almost whole by the mathematician von Neumann and the economist Morgenstern during the Second World War. Their theory was less a theory that made predictions or gave explanations than a framework for viewing complex social interactions. It caught on with mathematicians and defence analysts almost immediately, with social psychologists much later, and with economists and philosophers later still. But it has now become almost necessary to state some problems game theoretically in order to keep them clear and to relate them to other analyses. The game-theory framework represents ranges of payoffs that players can get from their simultaneous or sequential moves in games in which they interact. Moves are essentially choices of strategies, and outcomes are the intersections of strategy choices. If you and I are in a game, both of us typically depend on our own and on the other's choices of strategies for our payoffs.

The most striking advance in economics in the twentieth century is arguably the move from cardinal to ordinal value theory. The change had great advantages for resolving certain classes of problems but it also made many tasks more difficult. For example, the central task of aggregation from individual to collective preferences or utility could be done - at least in principle - as a matter of mere arithmetic in the cardinal system. In that system, Benthamite utilitarianism was the natural theory for welfare economics. In the ordinal system, however, there was no obvious way to aggregate from individual to collective preferences. We could do what Pareto said was all that could be done: we could optimize by making those (Pareto) improvements that made at least one person better off but no one worse off. But we could not maximize. In his impossibility theorem, Arrow (1951) showed that, under reasonable conditions, there is no general method for converting individual to collective orderings.

After game theory and the Arrow impossibility theorem, the next major contribution to rational choice theory was the economic theory of democracy of Downs (1957). Downs assumed that everyone involved in the democratic election system is primarily self interested. Candidates are interested in their own election; citizens are interested in getting policies adopted that benefit themselves. From this relatively simple assumption, however, he deduced two striking results that ran counter to standard views of democracy. In a two-party system, parties would rationally locate themselves at the centre of the voter distribution; and citizens typically have no interest in voting or in learning enough to vote in their interests even if they do vote.

The problem of the rational voter can be generalized. Suppose that I am a member of a group of many people who share an interest in having some good provided but that no one of us values its provision enough to justify paying for it all on our own. Suppose further that, if every one of us pays a proportionate share of the cost, we all benefit more than we pay. Unfortunately, however, my benefit from my contribution alone might be less than the value of my contribution. Hence, if our contributions are strictly voluntary, I may prefer not to contribute a share and merely to enjoy whatever follows from the contributions of others. I am then a free-rider. If we all rationally attempt to be free-riders, our group fails and none of us benefits.

A potentially disturbing implication of the game theoretic understanding of rationality in interactive choice contexts, of the Arrow impossibility theorem, of the economic theory of democracy and of the logic of collective action is that much of philosophical democratic theory, which is usually normative, is irrelevant to our possibilities. The things these theories often tell us we should be doing cannot be done.

1 Historical background

Of course, the role of self interest in explaining behaviour was not a discovery of the Renaissance and Reformation. From Socrates to the beneficent utilitarians, interest has been of central concern. Indeed, it has long been common to view morality as action against one's own interest in the interest of others or for otherwise good ends or good reasons. For example, Sidgwick saw beneficence and self-interest as often contrary dual motivations. Machiavelli and Hobbes were merely more insistent than many earlier philosophers on seeing the world as dominated by self-interest, and they were less ready to salvage social order in theory by supposing individuals just would or do behave well or public spiritedly.

The history of the slow invention of rational choice theory may be best told in the account of the role of fallacies of composition in political philosophy. Such fallacies in the aggregation from individual to collective choice or good are a standard background problem for rational choice theory. We are guilty of a fallacy of composition when, without argument, we assume that the attributes of an entity are the attributes of its constituents. This is a form of metaphorical or analogical reasoning that might sometimes yield correct results, but it often leads us astray. For example, we argue from the premise that every individual is rational to the conclusion that a nation or a group of individuals is rational in the same sense. It is only a slight exaggeration to say that rational-choice political theory is essentially an effort to block conclusions that are fallacies of composition. The three biggest bodies of work in rational-choice political theory - the aggregation of individual into collective preferences in the Arrow problem (§4), the rational-choice theory of voting and democracy (§5), and the logic of collective action (§6) - are deflations of fallacies of composition, as is the prisoner's dilemma game (§3), which is related to the collective action and voting problems.

Perhaps the central claim in Socrates' political philosophy is that it is in an individual's interest to act justly, to be just. This claim contradicts appearances that we, Glaucon in Plato's *Republic*, and others cannot deny. Socrates denies that there is a fallacy of composition in going from what is best for the individual to what is best for the community. It takes a great deal of devious argument and an adoring audience for him to seem convincing. Many teenage university students have since been hectored into accepting Socrates' conclusion and have therefore misconceived politics and group action from the beginnings of their supposed understanding. Perhaps it does not take more hectoring than it does merely because the fallacy is remarkably easy and seemingly natural.

After Socrates, Aristotle began his *Politics* with his own variant fallacy of composition. He supposed that, because every man acts with the intention of producing something that is a good, the polis, the most sovereign association of men, must pursue the most sovereign of goods. Aristotle's views have often been taught less with hectoring than with the weight of authority, but they too have set beginners on the path of confusion. Long after the reign of Greek philosophy, Aristotle came back into favour in a church that secured a match between individual interest and morality by inventing everlasting punishment for immorality. In effect, the church established a causal connection between individual and collective good to replace the failed conceptual connection. The causal connection it created was to make full use of self-interest.

Eventually Hobbes and others demolished the fallacy of composition at the heart of Socrates' and Aristotle's claims. In his allegory of the state of nature, Hobbes explicitly supposed that individuals would follow their own interests and that the result would be collective misery. This result gave licence to his invocation of an all-powerful governor to bring harmony and wellbeing because it can generally be in my interest to act well toward others only if there is a government prepared to coerce both me and the others to behave well. In three short chapters (*Leviathan*, chapters 13-15) Hobbes laid the foundations for rational choice theory and for political philosophy thereafter. A rational choice theorist might finally wish to say that Hobbes is the first modern political philosopher because he was the first to recognize and be persistent in avoiding fallacies of composition inherent in much of political philosophy before him (and, alas, since).

Nearly a century after Hobbes, Mandeville (§3) argued that private vices beget public virtues. In particular, universal avarice produces wide prosperity. This is a grim thought for virtue theory. All of us must wish others were avaricious because then they would serve our interests well. Avarice, the worst of all vices, has good results and its contrary virtue might have bad results. Smith, after writing his own book of near-virtue theory, *The Theory of Moral Sentiments*, then went on to write an economic theory in *The Wealth of Nations* that matched Mandeville's and severely undercut virtue theory (see Smith, A. §3). The central move, again, was to cut the supposedly conceptual connection between self interest and collective interest. There could be no conceptual, but

only a causal, connection, if any.

2 Game theory

Game theory is typically not about the source of the incentives we face but about the force of them, as is also true of rational-choice theory more generally. For most of its use, game theory is applied to actual payoff structures or to idealized structures. The payoffs are simply given. An economic theory of knowledge and preference formation can be brought to bear to explain preferences and, hence, to determine payoffs.

In social thought and in rational choice theory, games may be associated with three broad classes. There are games that involve pure conflict, games that involve pure coordination, and games that mix these two. If the only way for you to do better is for me to do worse, as when we play chess, we are in a game of *pure conflict*. If the way for each of us to do best is for both of us to do the same thing (in some sense), we are in a game of *pure coordination*, as when we drive our cars in North America and hope that every other driver stays to the right as we do (or in the UK, to the left). If I can sometimes do better only when you do worse and sometimes we can both do better by doing the same thing, we are in a *mixed-motive* game, which involves a mixture of conflict and coordination.

These three kinds of game can be represented very simply in two-person formats; each intersection of strategy choices yields a pair of payoffs. The first payoff in each pair is to Row, the second is to Column. The payoffs here are ordinal, with 1 as the first (best) choice, 4 the last (worst) choice. In the coordination and mixed-motive games, each player has two strategies from which to choose. In the pure conflict game, Column has no choice but will get whichever payoff Row chooses.

There could be any number of strategies available to both players in all categories of games. Moreover, the payoffs could be in almost anything, such as money, position, goods, or abstract utility. In most of social theory and of rational choice theory, ordinal games are adequate because there is no good sense of any quantity that could cover the range of payoffs. Even in games in which one might state what the payoffs are in each cell, one might be unable sensibly to assign them weights in anything like monetary or utility terms, so that ordinal representations still are adequate. However, ordinal representations lose important information that might alter the way people choose in a game, for example if the difference between one's first and second ranked outcome is trivial while that between the second and third is enormous.

Game theory first grew out of games of pure conflict (see game 1), perhaps especially on the model of poker. In a pure conflict game there is nothing to be gained by cooperation. When there are more than two players, there is typically the possibility of cooperation that makes the game not pure conflict. Then the analysis of the game becomes an analysis of coalitions of players that can form. Poker can be played under rules that block collusion or cooperation between players, and other games can be contrived that have more than two players but that are still games of pure conflict between every pair of players. Von Neumann's saddle-point or *maximin* theorem for two-person pure conflict games shows that one cannot do better against a symmetrically rational player than to choose a strategy that includes the so-called maximin payoff. That payoff is the best of the set of worst payoffs one could get from each strategy choice, or the maximum minimum (see Neumann, John von).

Game 1. Pure conflict

	Column	
	I	II
Row		
	II	2,1

One might think that pure conflict games are a very large and important class for social interaction. However, most social interactions that seem superficially to be pure conflict actually involve extensive opportunity for mutual benefit. For example, war might be seen as a game for victory or defeat, an analogue of chess. But actual wars, other than wars of attempted annihilation, can be fought in restricted ways that make all potentially better off. Hence, there is room for coordination as well as conflict.

Virtually the opposite of games of pure conflict are games of pure coordination, in which the incentive is to achieve harmony. Indeed, these games have been called common-interest games. Harmony can often be achieved without communication or anything even vaguely approaching agreement. The players need only recognize that their jointly best payoffs are in one or more outcomes and that there is some reason for choosing some particular strategy that includes one of those outcomes. If I am playing you in game 2 without opportunity for communication, I might choose strategy I instead of strategy II simply because the I-I outcome would come first for anyone reading English, which starts at the top left of the page - and I presume you might think the same way. Thomas Schelling (1960) has shown that people are remarkably good at coordinating in such ways without conversation.

Game 2. Pure coordination

		Column	
		I	II
	I	1,1	2,2
Row			
	II	2,2	1,1

In the pure coordination game of game 2, note that there are two outcomes with equally good payoffs for each player. Row does not care whether Column chooses strategy I or strategy II, but only cares whether both of them choose strategy I or strategy II. So long as they both choose the same strategy, they do well. Because there are two possible coordination outcomes, it is possible for the players to miscue and to finish in one of the second best payoffs. Hence, even pure coordination problems, as sanguine and simple as they may seem, can pose difficult personal and social choice problems. When such games in practical life are iterated, they are often resolved by convention (Hume 1739-40; Lewis 1969). Once we reach one of the best outcomes in an iterated class of interactions, we thereafter stick with the successful strategy, thereby establishing a *convention*.

3 Mixed-motive games

The mixed-motive game of game 3 merits extended discussion. The mixed-motive category includes many strategically distinct games. Game 3 is specifically a variant of the prisoner’s dilemma game, which is by far the most studied of all games. In the 1960s it captured the imagination of social psychologists, for whom it became a lasting focus of research. It is central in theoretical and simulation studies of altruism. And it has been associated more generally with the problem of collective action (§6). Despite all the attention it has got, however, it may have been relatively neglected in many circles because it has a misleading name.

Game 3. Mixed motive (Prisoner's dilemma or exchange)

		Column	
		Cooperate	**Defect**
	Cooperate	2,2	4,1
Row			
	Defect	1,4	3,3

As can be seen from the order of payoffs in game 3, prisoner’s dilemma has two striking features that suggest a conflict of motivations for the players. Suppose you are Column and I am Row. You can readily see that you rank your payoff from defection higher than your payoff from cooperation no matter what I do. If I cooperate, you get your best outcome in defecting. If I defect, you get your third best rather than your fourth best if you defect.

Hence, you have a *dominant strategy*: defect no matter what. My position is symmetrically equivalent to yours, so that I have the same dominant strategy. If we both follow our dominant strategy, we both defect and we reach an *equilibrium* in the game. But it is also obvious in game 3 that we both will be better off if we both cooperate than if we both defect. If we have any device outside the payoffs of game 3 to enable us to contract to insure mutual cooperation, we would benefit from doing so.

The prisoner's dilemma is unique among the seventy-eight strategically distinct two-person ordinal games in that it has an equilibrium that is Pareto inferior to some other outcome (Rapoport and Guyer 1966). (One outcome is Pareto inferior to another if moving from it to the other makes at least one person better off while hurting no one.) Defection dominates cooperation individually but mutual cooperation is better than mutual defection. The crossed incentives are maddening. But they are nothing new - we face them daily in our lives. Moreover, resolving their large-number analogue is the crux of political order (§6).

The discoverers of prisoner's dilemma were Flood and Dresher at the Rand Corporation around 1950. Flood had been trying to represent the sale of a used car as a game. Perhaps from a sense of that game's peculiarity, he and Dresher then designed the original prisoner's dilemma to test how people involved in an interaction repeatedly would behave. In essence, the first prisoner's dilemma experiment was of an iterated game. The game was sufficiently peculiar in comparison to other games, as noted, that they showed it to many people, including the Princeton mathematician, A.W. Tucker, speculating on what it might mean or represent. Tucker proposed the morality tale of two prisoners caught up in the American legal system. Kept in separate rooms, the prisoners are offered a chance at a reduced sentence if they will turn state's evidence against their partner. If both turn state's evidence, both will get lighter than the maximum sentences. If only one does, the other will get the maximum sentence, while the confessor will get a very light sentence. If both refuse to talk, both will get a mild sentence for a corollary crime (illegal possession of a firearm). The individual incentive is to turn state's evidence, but the prisoners could jointly do better if they both keep silence. The terms 'cooperate' and 'defect' correspond, for the prisoners, to keeping silence and turning state's evidence.

Unfortunately, this story of the two prisoners seems to have been more captivating than the game. Yet the real structure of that game is so common in life that, had it been properly identified in its name, it might have entered political philosophy and economics a decade or two sooner. In its payoff structure, prisoner's dilemma is fundamentally the model of ordinary *exchange*. The tale of two prisoners blurs the generality and import of the game. As is instantly obvious from their identical payoff structures, the morality tale of two prisoners and ordinary exchange are *strategically* identical. If there are morals to be drawn, they will first have to be inserted.

A particularly important variant of the prisoner's dilemma is the game in iterated play. When you and I are in an interaction beyond the immediate play of the game, we might both have incentives for action that derive from the further interaction as well as from our immediate game structure. In general, we may expect to see cooperative play in a prisoner's dilemma that is iterated many times, even into the distant future. Because much of our social life involves ongoing relationships, we will be able to map it into game theory only through iterated games. If the prisoner's dilemma is played once only between players who do not expect to meet again and without external constraints that would, for example, permit enforceable contracts, then virtually all game theorists think the players should rationally defect. What we face in real life commonly violates these pristine conditions enough to make it our interest to cooperate in manifold exchange relations (Hardin 1982; Axelrod 1984).

Game theory is a natural ground on which to investigate the plausible meaning of rationality in interactive choice. The standard statement is that it is rational to maximize (or optimize) one's own payoff against another player (or other players) who simultaneously maximizes. This is a patently circular or self-referential definition. But it sometimes works, as in pure conflict games. It also sometimes fails, as in many mixed-motive games. When it fails, we are left free to proclaim various solution concepts for games because we cannot stipulate that a particular one is rational in this natural but circular sense - and the literature is cluttered with variant solution concepts. We might conclude that rationality is ill-defined and merely wants more analysis. But we might sooner conclude that rationality is inherently indeterminate in many interactive choice contexts.

Note also that rational choice in the context of strategic interaction does not fit well with much of philosophical action theory. In philosophical action theory, I typically produce outcomes directly from my own action. My action when I choose one strategy rather than another is, of course, motivated by my concern with outcomes. But

the choice and the outcome are not directly connected. Rather, the outcome is the joint result of choices of strategy by me and by as many others as there are in the interaction. I might choose a strategy that includes some dreadful outcomes because my choice joined with strategies that I expect others to choose would produce a very good outcome. In the terminology of rational choice and standard decision theory, when I do not choose against nature but choose in combination with others with whom I interact, the meanings of my choices can be very complex.

4 The Arrow problem

The basic move from cardinal to ordinal value theory was completed in the 1930s when J.R. Hicks and R.G.D. Allen (1934) worked out the formal structure of an ordinal theory that could then be put to work in standard economic accounts. The next challenge was to move from Benthamite aggregate welfare to ordinal social welfare, still based on aggregation from individual, but now ordinal, preferences. Arrow took on this problem with the apparent expectation of resolving it positively. In a retrospective account, Arrow (1983) implicitly acknowledged that he had expected to demonstrate that it was not a fallacy of composition that we could move from individual to collective preferences of the same kind. Instead, he soon showed that it was a fallacy of composition and that Pareto was *de facto* right. He proved what is now called Arrow’s *impossibility theorem*: that there is no general way to aggregate individual ordinal preferences into collective ordinal preferences over the possible whole states of affairs.

The limited Pareto principle is often indeterminate because in many social contexts all plausible changes would make someone worse off and because, when there is more than one available Pareto improvement (a common state of affairs), the Pareto principle typically cannot say which is better. Similarly, ordinal social choice is indeterminately related to individual preference.

Arrow’s theorem is based on several assumptions that sound relatively benign and acceptable. First, he stipulates that only ordinal preference rankings be considered (the ordering condition, O). Then he stipulates that no individual’s preference ordering will be the ordering of the society (nondictatorship, D), no restrictions can be placed on any individual’s choice of orderings (universal domain, U), if everyone prefers one state of affairs to another, society prefers that state to the other (unanimity or Pareto, P), and if individuals’ rankings over the social choices are changed but no individual reverses the order of two possible choices, then the collective ordering over those two choices cannot change (irrelevance of independent alternatives, I). Finally, he implicitly assumes, reasonably enough, that our society has a finite number of individuals.

One might quarrel with each of these assumptions. But there is no point in quarrelling with O, since it is the point of Arrow’s exercise. There is also no point in quarrelling with Arrow’s unstated but implicitly invoked condition of finitude of the citizenry. And few have ever quarrelled with condition D - except to suggest it be made even stronger. Similarly, few argue against condition P, although Amartya Sen (1970) has campaigned hard against it. It is easy, however, to imagine states of affairs that everyone would rank below some other state (thus violating U). And many, perhaps most, commentators do not really know what to think of condition I. Is it naturally compelling? Or does it merely represent the lazy way we think through the problem of aggregation? In his own proof of the impossibility theorem, Arrow did not explicitly invoke the condition - although he did implicitly assume it.

As will be discussed further in §9, there is another fundamental issue that seems bothersome for Arrow’s theorem: the ranking of whole states of affairs. As Arrow says, a whole state of affairs means everything is determined. We can avoid the burden of such inclusiveness only if ordinal preferences over some things can be stated with a *ceteris paribus* clause that declares these things to be decoupled from all else. Why? Because ordinal preferences resolve two important and nearly intransigent problems of the older cardinal value theory. Under the older theory, I should apparently value two dinners tonight at twice the value of one - but clearly I do not. Indeed, I might even pay not to have to eat one of them. The two dinners would be *substitutes* for each other and therefore the second one need add no welfare to me. Also under the older theory, I should apparently value anything independently of what else there is or what else I have - but this is patently false. Some things are *complements*, and their value might be enhanced or reduced by my having something else with them, as my dinner and my glass of wine might be better enjoyed jointly than separately.

Now if the value to me of some consumption depends on whether there are substitutes or complements in my

bundle of consumptions, I cannot rank some part of the whole package. I can only rank whole packages in comparison to other whole packages. That is, as a citizen, I can only rank whole states of affairs fully determined.

Despite this relatively daunting thought, there is a reduced form of Arrow's theorem that is of great practical importance. It was recognized by the Marquis de Condorcet and by C.L. Dodgson. It is called the Condorcet paradox or the problem of cyclic majorities. If three or more people must choose from three candidates by majority rule, candidate A can defeat B, B can defeat C, and C can defeat A. There is then no outright majority winner (see Condorcet §2). Many voting systems in wide use avoid this problem by various random or controlled devices that sometimes bias outcomes. For example, legislatures commonly consider amendments to a piece of legislation one at a time. An amendment that wins a majority vote is added to the legislation; one that loses is dropped from further consideration. In the end, there is a decisive vote for or against the amended legislation, which is pitted only against the *status quo* and not against other possible variants. The *status quo* is therefore given privileged status because, if there are cyclic preferences over the *status quo*, the final piece of legislation and the same legislation with an alternative amendment, the *status quo* wins by virtue of not having to face the alternative. One might wish to say this is the best of plausible resolutions. But one cannot say any such thing with coherent meaning - that is the point of Arrow's theorem. Our presumptive notions of 'the best' are vacuous in this context.

5 Voting and democracy

Game theory and Arrow's theorem were very slow to enter debates in political philosophy and political science. Downs' economic theory of democracy, however, immediately became the starting point for most subsequent debate. Again, there are two main results of the theory. First, candidates, needing a majority of votes for election in two-party contests, would have to locate themselves near the middle of the range of popular preferences. Hence, in a catch phrase of 1950s US politics, the Democratic and Republican parties put up candidates who were close twins, Tweedledum and Tweedledee. This tendency inspired a massive literature on so-called responsible parties, a literature that argued that parties must emphasize their differences and must take strong positions. Downs' theory virtually killed that literature.

Second, the more crucial deduction was that voters have little interest in actually voting if the act of voting has real costs associated with it. If there are millions of voters and victory goes to the majority candidate, then my vote has little or no chance of affecting the outcome. Indeed, if elections were as close as a single vote, they would sometimes be voided and repeated simply because the counting of votes could not be accurate enough to yield a determinate outcome. But, in any case, genuinely close elections are extremely rare above the level of minor local offices. Hence, to weigh the value of my vote to me in affecting policies in my interest, I must discount my vote by the exceedingly low probability that it could matter. Its discounted value to me must be vanishingly small. This is the benefit side of my vote. Now consider the cost side. Suppose it takes a couple of hours out of my hard day for me to vote. This cost is not vanishingly small. On balance, it is therefore unlikely to be in my interest to vote.

Suppose voting costs could be reduced to nearly nothing. Would it then be in my interest to vote? Only if I can be reasonably sure that I know enough to vote in my interest. Karl Marx (1852) thought the French peasants of the mid nineteenth century did not know enough and that they harmed themselves with their votes. US air-traffic controllers might soon have concluded that their voting for Ronald Reagan in the presidential election of 1980 was stupid (he fired them when they went on strike). Some people know enough to judge the relation between elections and their interests accurately. Most people may have too limited knowledge and theory to judge that relation. If they are to vote rationally, therefore, they must acquire information. Typically it is costly to acquire information, and often the costs of relevant information for voting would far outweigh the discounted value to oneself of one's vote.

It is sometimes suggested that we can overcome the logic of Downs' rational voter by imposing sanctions on those who do not register and vote. But this is not a full solution - indeed, it might have perverse effects on elections because it may typically still not be rational for a voter to learn enough to make intelligent decisions. A requirement that people register and vote could be monitored relatively easily and cheaply. A requirement that they be prepared to vote intelligently could probably not be monitored without becoming massively intrusive. Moreover, the collective benefit from having everyone be well informed might be reckoned less than the collective costs of the required study. Hence, it seems reasonable to conclude from the rational-choice theory of citizen

participation in democratic politics that democracy faces severe constraints and that much of philosophical democratic theory is outside the realm of real-world politics (see Democracy §4).

Suppose the logic of the economic theory of democracy is correct and that the assumptions about individual capacities are roughly right. If ought implies can, it follows that most so-called democracies should not be the democracies of the democratic theorists. Hence, a theory that is about possibilities and that is based on simple assumptions of the weights of costs and benefits to individuals and groups yields fundamentally normative conclusions.

Unfortunately for simple, stable visions of democracy, there is a third major result in the economic theory. One of the obvious simplifications of Downs' theory was his assumption of a one-dimensional policy space along which voters are arrayed from left to right. If the world is more complex than this and there are two or more policy dimensions that do not collapse into each other, the results can be chaotic. Slight changes in location of candidates can entail radical changes in election outcomes. If candidates attempt to respond to each other, the winning policy position can be anywhere in the entire policy space. Every outcome is democratic - or none is.

6 The logic of collective action

Suppose we are members of a group who share an interest in having some policy adopted but that every one of us would rather be a free-rider on the efforts of others than contribute toward providing our common benefit. If we all attempt to be free-riders, we will none of us enjoy any collective provision. It is in the group's interest to have the good provided even at its costs; but it is in no individual member's interest to provide any part of the good. The economist Olson has called this the logic of collective action (Olson 1965; Hardin 1982).

It is commonly said that doing something is in *our* interest and therefore in the interest of *each* of us. Indeed, this is the central argument of distressingly many newspaper editorials on such problems as pollution, urban blight, war and peace, and ending poverty. The argument is wrong. Before Olson, the standard theory of groups fallaciously supposed that what was in a group's interest was in the individual group member's interest. In one very influential theory, it was supposed that, in order to predict policy, we need only do a vector sum of the force of all the relevant groups, with each group's vector length given by its number of members and its vector direction given by its policy preference. Hence, if there was a very large group with an interest, there would be policies adopted to satisfy much of the interest. This was a fallacy of composition - perhaps best called a fallacy of decomposition - but it is hard to read the error directly in traditional group theories. Anyone who stated the matter articulately would plausibly have recognized its illogic.

The full panoply of possible groups includes, in Olson's terms, privileged, latent, and intermediate groups. In a privileged group at least one person values provision (of some good that benefits a group) enough to pay fully for its provision - and therefore we might expect provision. In a latent group, every member has the preferences of the free-rider and therefore we expect no provision. In an intermediate group, every member has the preferences of a free-rider but the group is small enough and close enough for interactions that might produce greater cooperativeness and therefore, possibly, provision. In political contexts it is often seemingly true that small, well organized industrial groups can succeed in acting collectively for their political benefit, as in lobbying for privileged legislation or helping the relevant candidate win election. At the same time, the large group of much, even most, of the public is latent and fails to organize well enough to lobby for or against legislation of interest to the group or to support relevant candidates. The rational-choice theory of group behaviour therefore suggests that government outputs will be biased in favour of concentrated interests and against wider public interests.

7 Further applications

The rational choice theory of democracy suggests that we should want to understand the relation of individuals to the institutions in which they have roles and with which they have to deal in social life. If an institution is comprised of individuals in some structure of roles, then we should be able to give an individual-choice theory of institutions. This has long been a major programme, as in the work of Smith, Bentham, and Mill. It is the reputed next project for bringing Rawls' theory of justice to bear in a real world. It is a frequent worry in jurisprudence and in the resurgence of constitutional thinking after the extraordinary events of 1989 in eastern Europe. And it is a major industry in organization theory in economics, psychology, and sociology.

The modern public-choice programme in economics is often attributed to the judgment of Joseph Schumpeter that standard economics applies as well to public as to market choice because it is essentially a theory of the effects of incentives (see Schumpeter, J.A. §3). One could generalize Schumpeter's recommendation that we apply standard economic theory to government as well as the market and recommend that we apply it to all of human behaviour. This would be audacious, perhaps, and possibly misguided. But the recommendation has been taken up by many theorists, who have applied rational choice, game theory, standard price theory, equilibrium analysis, and other devices of economics to family relations, altruism, racism, crime and sex (Becker 1976), to theory of knowledge, norms, and group identification (Hardin 1995), to the rise of states (Levi 1988), to agency, religion, addiction, war, trust, and whatever ails or moves us. Rational choice theorists commonly treat motivation from anything other than interest or personal preference as an anomaly to be explained, not only for actions in the province of public choice but, increasingly, for action across the broad spectrum of life. In return, economics has benefited from feedback from social concerns in the information economics of Stigler and others, theories of cognitive limits associated with Herbert Simon and concerns with fairness and distributive justice.

8 Intentional behaviour

In rational-choice explanations there are commonly generalizations that do not turn on asking every relevant person in a decision-making group why they act as they do. Instead of asking, the rational choice theorist simply assumes a pattern of interests or of costs and benefits. Sometimes, it would be wrong to take people's claims for their actions and motives as compelling. Twentieth-century social and psychological thought has been receptive to the supposition that much of our motivation is not recognized by us. From Freudian psychology to mass movement sociology to economic and rational choice, it is commonly assumed that people act intelligibly but with poor understanding of their own actions.

Much of twentieth-century psychology and, to a lesser degree, other social sciences, was a reaction to this assumption. Since it has often been assumed that we cannot know what is going on in other people's minds many psychologists and others have insisted on considering only behavioural data, not putative mental events. In face of the Freudian vision of mental processes, one can imagine that behaviourism was appealing to many psychologists in the early twentieth century. Behaviourism has sometimes been identified with logical positivism in the philosophy of science, but the identification fails for those who are otherwise positivist but are none the less interested in mental explanations that involve such notions as rationality, instinct, or moral commitment (see Behaviourism in the social sciences; Behaviourism, methodological and scientific).

The social sciences in which behaviourism has had the least impact are economics and anthropology. There may be many reasons for the relative paucity of behaviourist theory in economics. First, and arguably the most important, is that much of economics hardly makes sense as anything but a theory of choice, and much of choice hardly makes sense as anything but mental events. Ideally, a chooser should be able to justify a choice, as must a theorist. In a similar way, rational choice explanations of behaviour cannot be merely behaviourist. They involve attributions of rationality to actors. Second, it may be true that for many problems of interest, economists can safely enough assume that behaviour is fully indicative of preference - typically a mental notion - because some preferences are virtually objective or universal. Much of what is called behavioural economics is about such quasi-objective or universal preferences. For such objects of choice, behaviour is a good proxy for mental preference. Third, economists may often find that attempting to ground theory in strictly behavioural assumptions, as in revealed preference theory, is less fruitful than relying on mental notions.

The grievous problem for explanations based on the mental notion of rationality is that, if the theory is worth doing, the explanations go beyond what the actors themselves understand of their actions. For example, when Olson presented his logic of collective action, when Flood and Dresher presented the game that is now known as the prisoner's dilemma, when Hume and, much later, David Lewis presented their theory of coordination, and when economists argue for rational expectations, they all go beyond what the subjects of their theories have recognized. If all these things had already been understood by the actors, there would be little point of doing social science, which would lead not to discoveries but merely to catalogues. Yet it seems absurd to say that people who cannot or who have never yet had occasion to understand these theories are not acting according to them. Indeed, the astonishing thing about these and many other theories is that they make sense of commonplace behaviour in *de facto* mental terms - for example, by showing that the behaviour is rational.

9 Value theory

There are several problems at the foundations of rational choice. In addition to the problem of indeterminacy discussed above (§§3-4), there are problems in the underlying value theory. I will discuss only two categories of these: variant rationality assumptions and the relations between interests, consumptions, and utility.

First consider variant rationality assumptions. When game theory first appeared it was praised as brilliant but damned as irrelevant by economists in the vanguard of the ordinal revolution. They held two counts against the von Neumann-Morgenstern utility theory. It was interpersonally comparable and it was cardinal. Decades earlier, Pareto had argued forcefully against interpersonal comparability and most economists agreed with him in principle (although they also regularly did cost-benefit analyses that violated Pareto's view). With cardinality, payoffs could be added or combined in probabilistic functions. As discussed above (§4), cardinality violated evident sense - and, perhaps worse, it was unnecessary for economic theory.

Immediately, von Neumann responded to economist critics with his proof that anyone with complete ordinal preferences, including preferences over probabilistic combinations of outcomes, has preferences that are cardinal up to a linear transformation. This claim mirrors an earlier claim by the philosopher Frank Ramsey (1931). Around this view there has risen a vast body of theory about individual decision making under uncertainty and the Bayesian theory of decision. Even a rejected value theory seems to bear fruit. Still, the ordinalists may note that Bayesian or von Neumann value theory is seemingly restricted to contexts in which there are no major effects of substitutability and complementarity among the elements of an outcome.

In the longer run, game theory and the utility theory of von Neumann and Morgenstern were seen as separable, and ordinal games are widely used. For much of game theory even in the original tradition, the use of cardinal, interpersonally additive preferences may often be little more than a mathematical crutch to ease the way to a solution.

Now turn to interests, consumptions and utility. Satisfying interests enables consumptions, which produce or have utility. Interest is therefore merely proxy for the utility of eventual consumption. If we include all consumptions now and into the future in our choice function, as in Arrow's fully determined states of affairs, interest drops out. Note that interests and consumptions trade off against each other. It is against my interest to consume an opera tonight; but if I could not do such things, I would have little interest in living. It is the very point of my interest in various resources that they enable me to consume and enjoy.

To do rational choice theory with interests instead of the full complexity of consumptions superficially might seem to simplify the account. But the costs of consumptions need not be linearly related to their utility or the enjoyment of them. Using interests as a proxy for consumptions is therefore potentially misleading, although it might often be relatively sensible. The cardinal value theory of von Neumann and Thomas Bayes might apply to simply conceived interests but not so readily to a panoply of consumptions. To put this the other way around, focusing on interests allows us to think cardinally; focusing on consumptions virtually forces us to think ordinally.

10 Preliminary conclusions

Two conclusions from the rational choice theory and one conclusion about it seem evident. The first two of these are the grievous problem of indeterminacy and the difficulties of democratic theory. Indeterminacy seems to inhere in the problems we face: aggregation from individual to social choice. It is a large part of the cause of difficulty for democratic theory. If certain choice procedures were not indeterminate, we could plausibly reach strong conclusions about the relation of democratic procedures to best outcomes. But if the category of 'best outcomes' is conceptually empty or meaningless, we may have no ground from which to judge particular procedures.

The third conclusion is that rational choice theory has been remarkably successful in the sense of systematically addressing and, in its own terms, resolving an enormous number and variety of issues. This success poses an interesting question in the philosophy of the social sciences. Why has rational choice theory been so successful? A first answer might be that it is like utilitarianism and economics - it has a programmatic approach that is broadly applicable and it has a highly refined value theory. Or rather it has several highly refined value theories, each developed over centuries by many of the best social theorists. Even in the central disagreement over the relevance of ordinal and cardinal utility theories, there is programmatic and sustained focusing of the issues.

A second answer might look instead to the standard against which the theory seems to be successful: the relative success of other approaches to social theory. Social theory often seems to be faddish - we go from one perspective to another. In many other branches of knowledge there may also be fads, but these seem to be about marginal matters while the core continues. A summary account of twentieth-century social theory would be dizzying in its rollicking from one fad to another. Many of these have gone to the core of one or more of the social sciences. A striking feature of many of the faddish movements is that they are critical, they are not constructive. Once its criticism is adequately made, each critical theory goes into decline. If rational choice theory is to go into decline, it will not be from exhaustion but from failure, which can probably happen only if a more successful approach is invented. Rational choice theory has had the advantage of centuries of coherent development that cannot soon be matched by a new approach.

See also: Decision and game theory; Social choice; Social norms

RUSSELL HARDIN

References and further reading

Aristotle (384-322 BC) *Politics*, trans. E. Barker, New York: Oxford University Press, 1958. (Sprawling normative account of politics in the polis.)

Arrow, K.J. (1951) *Social Choice and Individual Values*, New Haven, CT: Yale University Press, 2nd edn, 1963. (The master's presentation, with responses to early criticisms and extensions.)

Arrow, K.J. (1983) *Collected Papers*, vol. 1: *Social Choice and Justice*, Cambridge, MA: Harvard University Press.(Many important papers on varied topics plus a note, pp. 1-4, on Arrow's meeting with his fundamental impossibility theorem.)

Axelrod, R. (1984) *The Evolution of Cooperation*, New York: Basic Books.(On the possibility of establishing cooperative relations in dyadic interactions in large societies.)

Barry, B. (1989) *Theories of Justice*, Berkeley, CA: University of California Press. (Includes a game theoretic account of theories of distributive justice.)

Barry, B. and Hardin, R. (eds) (1982) *Rational Man and Irrational Society?*, Beverly Hills, CA: Sage Publications.(A collection of papers on the Arrow, collective action, and voting problems with commentaries and an extensive bibliography.)

Becker, G.S. (1976) *The Economic Approach to Human Behavior*, Chicago, IL: University of Chicago Press. (Collection of essays on broad range of issues. Sometimes technically demanding.)

Breton, A. and Wintrobe, R. (1982) *The Logic of Bureaucratic Conduct: An Economic Analysis of Competition, Exchange, and Efficiency in Private and Public Organizations*, Cambridge: Cambridge University Press.(A rational choice account with a strong emphasis on the sources and role of trust, esp. pp. 4-6, 61-88.)

Buchanan, J.M. and Tullock, G. (1962) *The Calculus of Consent: Logical Foundations of Constitutional Democracy*, Ann Arbor, MI: University of Michigan Press.(Rational choice applied to government institutions and to democratic theory.)

Coleman, J.S. (1990) *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.(Occasionally technically demanding sociological theory as grounded in principles of individual rational choice.)

Downs, A. (1957) *An Economic Theory of Democracy*, New York: Harper & Row. (On the problem of collective action that undercuts individual incentive to vote and on the incentives of parties and their candidates to adopt policies that produce winning votes.)

Eatwell, J., Milgate, M. and Newman, P. (1987) *The New Palgrave: A Dictionary of Economics*, 4 vols, London: Macmillan.(Includes many articles, often technical, on Arrow's problem, game theory, utility theory, and economists who have made major contributions to these issues, with extensive bibliographies.)

Elster, J. (1989) *The Cement of Society: A Study of Social Order*, Cambridge: Cambridge University Press.(On collective action, norms, and bargaining, all with some scepticism toward rational choice.)

Gauthier, D. (1986) *Morals by Agreement*, Oxford: Oxford University Press.(A rational choice theory of morality.)

Hardin, R. (1982) *Collective Action*, Baltimore, MD: Johns Hopkins University Press.(Game theoretic treatment of Prisoner's Dilemma, collective action, and convention with substantial bibliography.)

Hardin, R. (1988) *Morality within the Limits of Reason*, Chicago, IL: University of Chicago Press.(Applies game theoretic reasoning to utilitarian moral theory, esp. in chapter 2, and presents an account of institutional utilitarianism in chapters 3 and 4.)

Hardin, R. (1995) *One for All: The Logic of Group Conflict*, Princeton, NJ: Princeton University Press.(Rational choice account of individual identification with a group or community and of group conflict.)

Hicks, J.R. and Allen, R.G.D. (1934) ‘A Reconstruction of the Theory of Value, Parts I and II’, *Economica*, new series, 1: 52-76, 196-219.(The original presentation of now standard ordinal utility theory.)

Hobbes, T. (1651) *Leviathan*, C.B. Macpherson (ed.), Harmondsworth: Penguin, 1968.(Chapters 13-15 present a tight, bare-bones theory of government.)

Hume, D. (1739-40) *A Treatise of Human Nature*, L.A. Selby-Bigge and P.H. Nidditch (eds), Oxford: Oxford University Press, 3rd edn, 1978.(Book 3, part 2, §2 gives the original statement of convention as a resolution of a coordination interaction.)

Kavka, G. (1986) *Hobbesian Moral and Political Theory*, Princeton, NJ: Princeton University Press.(A game theoretic analysis.)

Levi, M. (1988) *Of Rule and Revenue*, Berkeley, CA: University of California Press. (Discusses how revenue policies drive state evolution.)

Lewis, D.K. (1969) *Convention*, Cambridge, MA: Harvard University Press. (Defines convention as the spontaneous resolution of a repeated coordination interaction.)

Luce, R.D. and Raiffa, H. (1957) *Games and Decisions*, New York: Wiley.(Dated but still valuable general survey of fundamental issues in game theory. Often quite demanding.)

Marx, K. (1852) *The 18th Brumaire of Louis Bonaparte*, New York: World Publishing, 1963.(Discusses the rational choice problems of French peasants, pp. 118-35.)

Mueller, D. (1989) *Public Choice II*, 2nd edn, Cambridge: Cambridge University Press.(An economist’s textbook survey of rational choice theory with an extensive bibliography.)

Neumann, J. von and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press, 3rd edn, 1953.(Original presentation of game theory; mathematical but generally not difficult. Later editions include, as an appendix, von Neumann’s proof that complete ordinal preferences yield cardinal preferences.)

Olson, M., Jr (1965) *The Logic of Collective Action*, Cambridge, MA: Harvard University Press.(Original presentation of the problem of collective action.)

Plato (*c*.428-*c*.348 BC) *The Republic*, trans. G.M.A. Grube, Indianapolis, IN: Hackett, 1974. (Hierarchical theory of the state. Glaucon argues for freeriding in book 2.)

Ramsey, F.P. (1931) ‘Truth and Probability’, in F.P. Ramsey, *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, London: Routledge & Kegan Paul, expanded edn 1978, 58-100.(Philosopher’s development of Baysian decision theory.)

Rapoport, A. and Guyer, M. (1966) ‘A Taxonomy of 2X2 Games’, *General Systems* 11: 203-14.(Presentation of the seventy-eight strategically non-equivalent 2X2 games.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Most important modern work on distributive justice.)

Samuelson, P. (1974) ‘Complementarity: An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory’, *Journal of Economic Literature* 12: 1255-89.(Account of development of ordinal utility theory from early beginnings with an extensive bibliography.)

Schelling, T. (1960) *The Strategy of Conflict*, Cambridge, MA: Harvard University Press, esp. pp. 83-118, 267-90. (Splendid, very accessible discussion of the real-world relevance of game theory.)

Sen, A.K. (1970) *Collective Choice and Social Welfare*, San Francisco, CA: Holden-Day.(Survey of the Arrow problem and related issues with an extensive but dated bibliography; has alternating verbal and technical chapters, the former quite accessible, the latter often difficult. The presentation in §4 uses Sen’s notation.)

Skyrms, B. (1990) *The Dynamics of Rational Decision*, Cambridge, MA: Harvard University Press.(Attempts to bring Bayesian theory and classical game theory together. Useful bibliography.)

Smith, A. (1759) *The Theory of Moral Sentiments*, D.D. Raphael and A.L. Macfie (eds), Oxford: Oxford University Press, 1976.(Presents an impartial spectator theory of morality.)

Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, R.H. Campbell and A.S. Skinner (eds), Oxford: Oxford University Press, 1976.(Economic theory that severly undercut virtue theory.)

Taylor, M. (1987) *The Possibility of Cooperation*, Cambridge: Cambridge University Press.(Rational-choice vision of anarchy, collective action, and the state with discussions of Hobbes and other classical thinkers.)

Rationalism

Rationalism is the view that reason, as opposed to, say, sense experience, divine revelation or reliance on institutional authority, plays a dominant role in our attempt to gain knowledge. Different forms of rationalism are distinguished by different conceptions of reason and its role as a source of knowledge, by different descriptions of the alternatives to which reason is opposed, by different accounts of the nature of knowledge, and by different choices of the subject matter, for example, ethics, physics, mathematics, metaphysics, relative to which reason is viewed as the major source of knowledge. The common application of the term 'rationalist' can say very little about what two philosophers have in common.

Suppose we mean by reason our intellectual abilities in general, including sense experience. To employ reason is to use our individual intellectual abilities to seek evidence for and against potential beliefs. To fail to employ reason is to form beliefs on the basis of such non-rational processes as blind faith, guessing or unthinking obedience to institutional authority. Suppose too that we conceive of knowledge as true, warranted belief, where warrant requires that a belief be beyond a reasonable doubt though not beyond the slightest doubt. Here, then, is a version of rationalism: reason is the major source of knowledge in the rational sciences. This is a weak version of rationalism which simply asserts that our individual intellectual abilities, as opposed to blind faith and so on, are the major source of knowledge in the natural sciences. It is clearly not very controversial and is widely accepted.

Suppose, however, we take reason to be a distinct faculty of knowledge distinguished from sense experience in particular. To employ reason is to grasp self-evident truths or to deduce additional conclusions from them. Suppose we conceive of knowledge as true, warranted belief, where warrant now requires that a belief be beyond even the slightest doubt. Let us also extend our attention to metaphysics and issues such as the existence of God, human free will and immortality. Here is a much stronger version of rationalism which asserts that the intellectual grasp of self-evident truths and the deduction of ones that are not self-evident is the major source of true beliefs warranted beyond even the slightest doubt in the natural sciences and metaphysics. Clearly it is highly controversial and not very widely accepted.

The term 'rationalism' has been used to cover a range of views. Scholars of the Enlightenment generally have in mind something like the first example - a general confidence in the powers of the human intellect, in opposition to faith and blind acceptance of institutional authority, as a source of knowledge - when they refer to the rationalist spirit of the period and the work of such philosophers as Voltaire. Most frequently, the term 'rationalism' is used to refer to views, like the second one above, which introduce reason as a distinct faculty of knowledge in contrast to sense experience. Rationalism is then opposed to empiricism, the view that sense experience provides the primary basis for knowledge. This entry concentrates on this still very general form of rationalism, reserving the term 'rationalism' for it alone.

1 Continental Rationalism and British Empiricism

The rationalist-empiricist division has traditionally played a major role in our understanding of the history of philosophy, particularly that of the modern period of the seventeenth and eighteenth centuries leading up to Kant. The major philosophers of the period are regularly grouped into two sets of three: Descartes, Spinoza and Leibniz are the Continental Rationalists, in opposition to Locke, Berkeley and Hume, the British Empiricists. Philosophy department curricula, textbooks, scholarly anthologies and conference programmes have all incorporated the classification for years and are likely to continue to do so.

None the less, scholars have at least five basic reservations about its accuracy. First, a close study of the three Continental Rationalists, especially their work in the natural sciences, reveals they had a great respect for the role played by experience in scientific knowledge. Second, the British Empiricists, especially Locke and Berkeley, stress, in one way or another, the importance of reason as a source of knowledge; consider Locke's account of a priori knowledge and Berkeley's acceptance of innate ideas. Third, the division encourages us to overlook important areas of agreement between philosophers on different sides of the divide, such as the views of Descartes and Locke on the nature, though not the source, of our ideas. Fourth, the division encourages us to associate irrelevant differences in language and geography - those who do not write in English against those who do; the Continent against England and Scotland - with a supposed difference in philosophical views. Fifth, the grouping of

the six philosophers in epistemological terms encourages an incorrect grouping of them in metaphysical ones. The Continental Rationalists are mistakenly seen as attempting to apply their reason-centred epistemology to pursue a common metaphysical programme, each trying to improve on the efforts of those before him; the British Empiricists are incorrectly seen as gradually rejecting those metaphysical claims, with each again consciously trying to improve on the efforts of predecessors. Defenders of the Continental Rationalists-British Empiricists distinction generally admit many, if not all, of these shortcomings but treat them as minor anomalies.

This entry does not attempt to resolve the controversy or even to lay out each side's supporting evidence from the works of the philosophers involved. A few points deserve mention, however. First, in evaluating whether it is appropriate to classify these philosophers in this way, we must consider what purpose the classification is supposed to serve. The classification might well be acceptable for pedagogical purposes at a certain level of instruction in the history of philosophy in order to initiate consideration of their views, but not at another level or for other purposes.

Second, attempts to group philosophers into such families as the rationalists and the empiricists are best understood as attempts to classify them by shared family resemblances. The question to ask, then, is not whether there is a list of interesting claims the acceptance of which is definitive of being a Continental Rationalist and the rejection of which is definitive of being a British Empiricist, so that Descartes, Spinoza and Leibniz may each be classified as a Continental Rationalist just so long as all those claims are accepted, and Locke, Berkeley and Hume may each be classified as a British Empiricist just so long as one rejects them. Critics of the classification sometimes present the issue in this way but few, if any, proponents see it in these terms. We need to consider whether there is an interesting list of claims emphasizing reason as a source of knowledge of the external world, a sufficiently large number of which are accepted by each of Descartes, Spinoza and Leibniz and rejected by each of Locke, Berkeley and Hume, though the first three may not accept all the same claims and the second three may not reject all the same ones.

Third, independently of the historical accuracy of the classification, the terms 'rationalist' and 'empiricist' are associated with some basic claims which define the family resemblance for each category. The rest of this entry focuses on the innate idea thesis and the demonstrative knowledge thesis. Minimal attention will be given to the sometimes complex scholarly questions of whether and why the claims are actually adopted by Descartes, Spinoza and Leibniz, and rejected by Locke, Berkeley and Hume, though some of these philosophers will be cited as examples where the attribution is fairly uncontroversial.

2 Innate ideas

To gain knowledge about the external world we need to think about it and thus we need the appropriate concepts. How do we gain them? One of the central theses associated with rationalism is that at least some of our concepts are not gained from experience but are instead innate. Descartes, for example, divided our ideas into three categories: adventitious ideas, such as our idea of red, are gained through sense experience; fictitious ideas, such as our idea of a hippogriff, are manufactured by us from other ideas we possess; innate ideas, such as our idea of God, of extended matter and of a perfect triangle, are placed in our minds by God at creation.

It is important to distinguish the innate-idea thesis from some other ones. The term 'innate' is often associated with things other than concepts. Some rationalists, for example Descartes and Leibniz, write of innate propositions as well as innate concepts. They seem to have in mind that we have beliefs, in particular true propositions, as part of our initial mental make-up rather than as a result of experience; the propositions are presumably constructed of innate concepts. Whether and just how such true beliefs might constitute knowledge is not clear. Some philosophers have also claimed that we have innate knowledge, though it is not always clear whether this is 'knowing that' or 'knowing how' (see Innate knowledge; Nativism). It has been maintained by contemporary philosophers, for example, that learning a natural language requires a knowledge of grammar which is 'innate' in that learners could not have inferred it from their experience. The focus in this section is solely on the thesis that some of our concepts are innate (see Concepts). The central rationalist positions on knowledge are considered in the next.

Some defences of the innate idea thesis are based in claims peculiar to particular philosophers. Leibniz's view that all substances are causally isolated monads implies a general rejection of sense experience as a source of any ideas

and so supports the claim that all ideas are innate. (‘It is a bad habit we have of thinking as though our minds receive certain messengers, as it were, or as if they had doors or windows. We have in our minds all those forms for all periods of time because the mind at every moment expresses all its future thoughts and already thinks confusedly of all that of which it will ever think distinctly,’ Leibniz 1685: Section 26.) The most common defence of the innate-idea thesis, however, takes the form of admitting experience as a source of ideas but then arguing that some concepts could not have been gained directly or indirectly from experience; that these concepts are innate is then offered as the best explanation of their existence. That the concepts could not have been gained from experience is generally defended in one or both of two ways. First, the content of the concepts is beyond what we directly gain in experience as well as anything we could gain by performing the available mental operations on what experience provides. Second, our possession of the concepts is presupposed by our ability to employ the very empirical concepts that might be thought to provide a basis for them in experience.

Consider, for example, Descartes’ defence of the claim that our concept of God, as an infinitely perfect being, is innate. This concept is not directly gained in experience, as a particular sensation of pain might be. Its content is far beyond any we could ever construct by applying available mental operations to what experience provides. Our empirical concepts include the concept of a finitely knowledgeable, powerful and good being; we can even construct the concept of a finitely but *very* knowledgeable, powerful and good being, but we cannot move on to the concept of an infinite one. We cannot, for example, gain the idea of infinite perfection by simply negating our concept of finite perfection. (‘I must not think that, just as my conceptions of rest and darkness are arrived at by negating movement and light, so my perception of the infinite is arrived at not by means of a true idea but merely by negating the finite’ (Descartes 1641: Third Meditation).)Moreover, we must possess the concept of infinite perfection in order to employ the concept of finite perfection gained from experience. (‘For how could I understand that… I was not wholly perfect, unless there were in me some idea of a more perfect being which enabled me to recognize my own defects by comparison’ (Descartes 1641: Third Meditation).)

Objections to the standard arguments for the innate-idea thesis generally take two forms. The less radical line of objection, exemplified by Locke, is to argue that the supposedly innate concepts are indeed gained by sense experience and to offer an account of the mental processes involved. The more radical critique, exemplified by Hume, is to agree that the supposedly innate concept could not be gained by experience and to argue that we do not in fact have the concept as understood by the innate-idea theorists in the first place; a failure to find an experiential source for a supposed concept should not lead to us to the innate idea thesis but to a critical examination of our concepts themselves (see Locke, J. §2; Hume, D. §2).

Besides defending their theory, innate-idea theorists face the task of explaining exactly what it is for an idea to be innate. Innate concepts are in the mind from creation but are not present in consciousness until we actually conceive them. How are these concepts ‘in the mind’ prior to being present to consciousness? How too can any of them, prior to being present to consciousness, serve as a precondition of our employing other ideas gained from experience?

Finally, innate idea theorists often assume that the ways in which we conceive of the external world using our innate concepts actually correspond to how it is, and they thus face the question of what justification, if any, there can be for this assumption. A frequent strategy is to find a non-deceptive source for our innate concepts - be it a non-deceiving God, as for Descartes, or a non-sensory experience by the soul prior to its union with the body, as in Plato’s doctrine of recollection (see Plato §§11-14).

3 Intuition and demonstration

A second thesis generally identified with rationalism is the claim that reason alone can provide us with at least some knowledge of the external world through our intuition of self-evident propositions and our subsequent deduction of additional information from those starting points. Intuition is understood as a kind of intellectual ‘grasping’ by which we comprehend a proposition in such a way as to recognize its truth; in a demonstration we reason through a series of intuited premises to a logically entailed conclusion. Propositions known by intuition are self-evident, while those known by demonstration are evidenced by the intuited premises; in either case the knowledge is a priori, independent of evidence gained from sense experience (see A priori).

It is essential to the rationalist account of knowledge by intuition and demonstration that the knowledge gained in

this way is knowledge of the external world. Many empiricists, indeed even Hume, are willing to admit that an intellectual grasping of our concepts can provide us with a priori knowledge, but they limit that knowledge to knowledge of the relations of our ideas: when we intuit that two plus three make five, we gain knowledge about how the concepts involved are related, but we do not learn anything about the world as it exists independently of our mind.

The rationalist commitment to intuitive and demonstrative knowledge of the external world is not necessarily a commitment to have such knowledge of propositions affirming the existence of particular concrete objects. It may take the form of a commitment to have such knowledge of abstract objects which none the less exist, and are as they are, independently of our thought. Descartes, for example, claimed that while he had no knowledge of the existence of any particular triangle, he none the less had an intuitive and demonstrative knowledge of propositions about the properties of a triangle and that the content of his knowledge was a 'determinate nature' independent of his mind.

The rationalist account of knowledge under consideration involves a version of foundationalism. It endorses the view that we know some propositions independently of their being evidenced for us by any other propositions; these basic propositions provide us with the evidential base for additional knowledge (see Foundationalism). Rationalists generally adopt a strong view about the strength of our intuitive and demonstrative knowledge. When we intuit a proposition we know it to be true with certainty. Our intuitive grasp places it beyond even the slightest doubt. This extreme degree of epistemic justification can only be transferred to other beliefs by deductive inference which precludes the possibility of the conclusion being false while the premise is true. Even our ability to gain additional knowledge by demonstration is limited. Once our demonstrations get so lengthy as to rely on memory, the possibility of a mistake in recall weakens our degree of justification. (On some interpretations of Descartes' epistemology, a major aim of his epistemic programme is to remove the limitation that this reliance on memory placed on the certainty of our demonstrations.)

Rationalism is sometimes identified with the view that every science has the same deductive structure in which knowledge moves from the intuition of self-evident axioms to the demonstration of theorems. It is thus seen as endorsing a single method for all sciences and as asserting the priority of the knower over the known: as we move from one subject to another - from mathematics to physics - the subject matter changes, but so long as the nature of the knower's mind remains constant the same method of intuition and demonstration may be used in each case. The Continental Rationalists were all impressed by the rigour and certainty associated with the deductive method of mathematics and they stressed the importance of meeting such standards in other areas. So too, Descartes and Leibniz, like Galileo, were quite concerned with the application of mathematics to the natural sciences. Yet, it is clearly a mistake to attribute to them individually, or to rationalism generally, the view that all science proceeds by intuition and demonstration. Even though Descartes, for example, attempted to use intuition and demonstration to establish some basic laws of nature (that whatever is in motion continues in motion in so far as it can, that all motion is rectilinear, and so on), he admitted that we need to use experiments to move beyond very basic principles.

Just as the innate idea thesis is partially motivated by the belief that sense experience can not provide some concepts we clearly have, so too the rationalist belief in knowledge of the external world by intuition and demonstration is often at least partly motivated by the conviction that sense experience falls short of providing us with knowledge in some areas. Our claims in mathematics, metaphysics and ethics, for example, seem to outstrip the content of our sense experiences. We experience lots of imperfect triangles but no perfect ones, we may experience the effects of God's creative powers but not God, and we experience how things are but not how they ought to be. So how can we have knowledge in such areas? Intuition and demonstration provide an alternative account; one that becomes increasingly attractive when the other main option seems to be scepticism (see Scepticism).

The rationalist appeal to intuition and deduction is also at times motivated by a demanding conception of knowledge that seems to require more than experience can provide. Descartes' conception of knowledge as permanent supported his requirement that it be absolutely certain which in turn placed it beyond sense experience. Whenever we form beliefs on the basis of sense experience, the possibility that our senses are deceiving us keeps us from absolute certainty (see Certainty; Doubt; Fallibilism). Here again, intuition and demonstration provide an

attractive alternative, especially given their apparently successful use in mathematics with its high degree of epistemic justification.

4 Intuition and demonstration (cont.)

Some of the difficulties for the rationalist appeal to intuition and deduction arise out of the reasons for rejecting sense experience. If knowledge requires certainty, what makes our intuitions and subsequent demonstrations certain? We have already noted the worry about the use of memory in extended demonstrations (see Memory, epistemology of). What makes our intuitions a source of certainty even while we are having them? Could not a deceitful God have us intuit false propositions, just as easily as we might deceive ourselves in a dream? Do not such epistemic possibilities as a deceitful God give us at least a slight reason to doubt our intuition that two plus three make five, preventing it from being absolutely certain?

One rationalist strategy for responding to this difficulty is to offer an argument for the veridicality of our intuitions. Once we establish that our intuitions are non-deceptive, we can gain certainty by them, and we can establish the universal veridicality of our intuitions while we cannot establish the universal veridicality of our sense experiences. Yet, the argument for the truth of our intuitions will only meet the sceptical challenge if it establishes its conclusion with certainty, and the demand for certainty thus requires the rationalists to rely on intuition to support the premises of their argument for the veridicality of intuition. They thus fall into an apparently vicious form of circular reasoning (sometimes called the ‘Cartesian Circle’ in recognition of Descartes’ struggle with the problem). A second strategy is to claim that all our intuitions are certain, even when we have not gained certain knowledge that the faculty never deceives. Whenever we intuit a proposition, we are certain of it; such hypotheses as that of a deceiving God do not give us a reason to doubt it, even if they are epistemic possibilities for us. This strategy risks being *ad hoc*: why should we accept intuition as an automatic source of certainty?

Appeals to intuition carry a related methodological problem. Let us assume that real intuitions are always true and always certain. How do we distinguish between real intuitions and only apparent ones, and how do we get ourselves in a position to have the former but not the latter? People have taken themselves to ‘just see’ the truth of all sorts of contradictory claims; they cannot all be correct (see Pyrrhonism). Descartes’ method of doubt is intended to help us gain a psychological state where we will have real intuitions, but his own extravagant list of ‘intuited’ propositions in the course of following his method illustrates its ineffectiveness.

Another difficulty for the rationalist reliance on intuition and demonstration takes us back to the view that some claims about the external world fall beyond the range of our sense experience but within that of intuition and demonstration. The rationalist must argue that the contents of intellectual intuition are independent of us in such a way that, as in Descartes’ theory of simple natures, our knowledge of how they are related is not just knowledge about the structure of our own thoughts. Rationalism, while primarily an epistemological position, thus involves its proponents in at least some metaphysical commitments (see Introspection, epistemology of).

According to many contemporary epistemologists, rationalism, like such related theories as foundationalism, is dead. It surely is beset with problems in both the innate idea thesis and the demonstrative knowledge thesis. Yet, a complete evaluation of rationalism must involve more than an examination of these two central points. It must also include an overall examination of the nature and extent of our knowledge and the nature and extent of our experience. Rationalism, in one form or another, will remain attractive so long as we find that we have knowledge of the external world which appears to go beyond what experience can provide.

See also: Empiricism; Enlightenment, continental

PETER J. MARKIE

References and further reading

Chisholm, R. (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 3rd edn, 1989.(General introduction to epistemology containing a detailed analysis of the rationalist view that reason alone can provide knowledge of the external world.)

Descartes, R. (1641) *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes*, vol. 2, trans J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1984.(Along with the *Discourse on Method* (1637) and *The Principles of Philosophy* (1644), this work provides a fine introduction to

Descartes' thought.)
Kenny, A. (1986) *Rationalism, Empiricism and Idealism*, Oxford: Oxford University Press.(Several articles in this collection question the correctness of the traditional distinction between the Continental Rationalists and the British Empiricists with regard to specific comparisons between philosophers.)
Leibniz, G.W. (1686) *Discourse on Metaphysics, Correspondence with Arnauld, Monadology*, trans. G. Montgomery, La Salle, IL: Open Court, 1973.(These three works provide a fine introduction to Leibniz.)
Loeb, L. (1981) *From Descartes to Hume: Continental Metaphysics and the Development of Modern Philosophy*, Ithaca, NY: Cornell University Press.(One of the best general challenges to the traditional distinction between the Continental Rationalists and the British Empiricists.)
Malebranche, N. (1688) *Dialogue on Metaphysics and on Religion*, trans. W. Doney, New York: Abaris, 1980. (The works of Malebranche represent an interesting development on the Cartesian tradition. This work offers the best introduction to his views.)
Quine, W.V. (1951) 'Two Dogmas of Empiricism', in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1953. (The rationalism-empiricism dispute over intuitive knowledge of the external world is sometimes conceived as involving a distinction between analytic and synthetic truths: the rationalists' commitment to such knowledge is understood as a commitment to synthetic a priori knowledge; the empiricists' rejection of such knowledge is understood as a denial of synthetic a priori knowledge; this piece contains an influential critique of the analytic-synthetic distinction.)
Rorty, R. (1979) *Philosophy and the Mirror of Nature*, Princeton, NJ: Princeton University Press.(Attack on epistemic foundationalism and more generally on epistemology as done in a tradition based in Descartes on the rationalist side, and Locke on the empiricist side.)
Spinoza, B. de (1677) *The Ethics, in The Collected Works of Spinoza*, vol. 1, ed. and trans. E. Curley, Princeton, NJ: Princeton University Press, 1985.(The best introduction to Spinoza's philosophy.)
Stich, S. (1975) *Innate Ideas*, Berkeley, CA: University of California Press.(A discussion of innate ideas and related topics.)

Rationality and cultural relativism

Under what conditions may we judge the practices or beliefs of another culture to be rationally deficient? Is it possible that cultures can differ so radically as to embody different and even incommensurable modes of reasoning? Are norms of rationality culturally relative, or are there culture-independent norms of rationality that can be used to judge the beliefs and practices of all human cultures?

In order to be in a position to make judgments about the rationality of another culture, we must first understand it. Understanding a very different culture itself raises philosophical difficulties. How do we acquire the initial translation of the language of the culture? Can we use our categories to understand the social practices of another culture, for instance, our categories of science, magic and religion? Or would the mapping of our categories on to the practices of culturally distant societies yield a distorted picture of how they construct social practices and institutions?

A lively debate has revolved around these questions. Part of the debate involves clarifying the difficult concepts of rationality and relativism. What sort of judgments of rationality are appropriate? Judgments about how agents' reasons relate to their actions? Judgments about how well agents' actions and social practices conform to the norms of their culture? Or judgments about the norms of rationality of cultures as such? Can relativism be given a coherent formulation that preserves the apparent disagreements for which it is meant to account?

Can there be incommensurable cultures, such that one culture could not understand the other? According to Donald Davidson's theory of interpretation, radical translation requires the use of a principle of charity that in effect rules out the possibility of incommensurable cultures. If this result is accepted, then a strong form of cultural relativism concerning norms of rationality is also ruled out.

Davidson's theory, some argue, does not eliminate the possibility of attributing irrational beliefs and practices to agents in other cultures, and thus still leaves some room for debate about how to understand and evaluate such beliefs and practices. Three positions frame the debate. The intellectualist position holds that judgments of rationality are in order across cultures. The symbolist and functionalist positions, here taken together, try to avoid such judgments by attributing functions or symbolic meanings to cultural practices that are generally not understood as such by the agents. The fideist position, wary of too easily being ethnocentric, assumes a more relativist stance with regard to cross-cultural judgments of rationality.

1 Background to the debate

Interest in questions of cross-cultural understanding and judgment can be found among the Greeks, in Herodotus for instance. At the beginning of the modern period, Giambattista Vico in *The New Science* (1744) in effect raised many of these issues in his attempt to develop a new science of human history (see Vico, G. §§4-6). The modern preoccupation with these questions, however, began only with the development of anthropology in the late nineteenth century. Evidence began to accumulate of apparently irrational beliefs and associated practices in diverse cultures. How was this evidence to be assessed and explained? The founders of anthropology, including Sir Edward Tylor, Sir James Frazer, Bronislaw Malinowski and Lucien Levy-Bruhl, reacting to the encounter between European and non-European civilizations and noting the diversity of cultural practices and beliefs, debated these issues largely in terms of the now suspect categories of 'civilized' versus 'savage' or 'primitive' societies, or of 'pre-logical' and 'pre-scientific' versus 'logical' and 'scientific' cultures. While largely rejecting this terminology, a later generation of philosophers and anthropologists carried on a debate about essentially the same issues, with a particularly intense period of discussion occurring in the 1960s, 1970s and into the 1980s. Some argued that the best account of the apparent irrationalities found in other cultures was in terms of a theory of cultural relativism. Others rejected relativism of any sort (see Anthropology, philosophy of §1). Rather than reaching a resolution, the debate about rationality and cultural relativism continues and draws closer to parallel debates in the history and philosophy of science and epistemology generally.

The participants in this debate use a number of difficult concepts and not always in the same way. Clarification of the concepts of rationality and relativism is required before the various positions in the debate can be presented.

2 Rationality

A wide variety of things can be said to be rational or irrational. In the context of the debate about the rationality of distant cultures, the most important are beliefs, individual actions and cultural practices. Rationality, in general, has to do with how people acquire beliefs from evidence and connect reasons to actions. At the most basic level, rationality concerns the standards for truth, consistency and deductive and inductive inference. The concept of rationality is standardly used in a normative sense, that is, one that conveys commendation or endorsement. To call an action rational is to certify that there is good reason to do the action or to judge that the action was brought about in a suitably correct way. Likewise, to describe an action as irrational is to indicate that there is not good reason to do the action or to judge that there was some error or fault in how the action came about and to criticize it (or rather, the agent performing the action) in that regard (see Rationality of belief; Practical reason and ethics; Reasons for belief).

Focusing on the rationality of action, we can usefully distinguish three levels where judgments of rationality may be made. The first level regards individual actions. According to the model of rational action suggested by Davidson, a rational action is one that stands in a certain relation to the agent's beliefs and desires, that is, their reasons for acting. For example, if a person has a desire, all things considered, to bring about goal X, and believes, all things considered, that action A is the most effective means to goal X, then the person acts rationally if they bring about action A for those reasons. This formulation simplifies things considerably. In some instances, it may be rational to choose an action that is suitably effective, but not the most effective one. Thus, there may be more than one action that would be rational, given an agent's reasons. In this case, the concept of rationality would limit but not uniquely determine the action to be chosen.

On this level, an agent would be acting irrationally if they did not choose the action that was the most effective (or at least a suitably effective) means to goal X; that is, if the agent's reasons were not good reasons for the action chosen. This failure might occur for a number of complicated reasons, most of which involve some sort of failure of critical thinking or consistency. While capturing the standard notion of instrumental rationality, this model, in this context, is meant to apply to more than what might be thought of as narrowly instrumental action. Thus, actions oriented towards successful communication or successful participation in a ritual can be understood in terms of the model with suitably adjusted notions of effective means towards those goals.

The second level where judgments of rational action may be appropriate involves questions about the agent's reasons (beliefs and desires), in particular, questions about how the agent acquired the reasons. Is the belief appropriately related to the available evidence? Does the evidence come from a reliable source? Is the evidence reliable such that the agent has good reasons for the action undertaken?

One way to understand what makes the relationships of evidence to beliefs and of reasons to action good or appropriate is to ask: (1) whether a belief was acquired by an agent in a way that conformed to the epistemic norms of the agent's culture, and (2) whether the agent's reasons (beliefs and desires) provide good reasons for the action according to the canons of reasoning of the agent's culture. Putting the questions in this way allows space for what seems to be a form of cultural relativism. If different cultures have different and conflicting epistemic norms and/or canons of reasoning, then a belief might be rationally acquired according to the norms of one culture but not according to those of another, or provide good reasons for an action in one culture but not in another. For instance, if a certain way of reading omens is considered a reliable means of evidence gathering in one culture but not in another, then reading omens to acquire evidence for beliefs and acting on those beliefs would be considered rational by the standards of the first but not the second culture. Likewise, inductive counter-evidence might be assigned different weight in different cultures, affecting what is considered a good reason for action. Thus, at this level, judgments of the rationality or irrationality of an agent's action can be regarded as intersubjective judgments relative to the norms of the agent's culture.

The move from the second level, thus construed, to the third is readily apparent. Can one make judgments of the rationality or irrationality of the norms of a culture? If we can make such judgments, then the action of an agent might on analysis turn out to be rational on the first two levels but still be irrational from the point of view of the third. For instance, if omen reading is judged not to be epistemically rational, then omen reading, and actions based on omen reading, could still be judged irrational even if they took place in a culture that endorsed the practice. (It should be noted that an action that was judged irrational on the second level because it was not properly related to the norms of the agent's culture in a way that the culture would regard as rational, might still

turn out to be rational on the third level if the agent's individual norms were judged rational.)

Given this analysis of rationality, we can restate some of the questions central to the debate. From what point of view might we make such third level judgments of the rationality or irrationality of a culture's norms? Is the only available position that of the norms of rationality of our culture? If so, should we restrict ourselves to second level judgments only? If we make third level judgments, are we really judging the other culture to be irrational, or only different? And what would we mean by such a judgment of cultural irrationality? What sort of mistake would a culture so judged be making?

3 Relativism

Relativism is a notoriously difficult concept. The form of relativism of primary concern here is cultural relativism with regard to norms of rationality. What such a relativist theory typically holds is that (1) norms of rationality differ across cultures; that (2) judgments of the rationality of a given action are relative to the governing norms of the local particular culture; such that (3) two people, depending on their cultural locations, might disagree about the rationality of the same action, one judging it to be rational and the other irrational, and both judgments would be equally correct.

One implication of this form of cultural relativism is that it denies the existence of a universal or 'objective' system of rationality that could be used to judge the rationality of actions in all cultures. In the terms used above, relativism denies that there is a culture-independent position from which to make third level judgments about the rationality of a culture's norms. The most we can say is that an action seems irrational to us, by the norms of our culture, even while recognizing that by the norms of the culture in which is occurs, the action may be rational.

There is an ongoing debate about whether relativism can be stated coherently in a way that saves the phenomenon for which it is meant to account. The relativist, at least in certain versions of the position, seems to want to endorse as correct propositions that are apparently logically incompatible. On the other hand, where claims are clearly relativized to particular theories or cultures, thus eliminating the appearance of logical inconsistency, it is not clear that genuine disagreement remains. Some maintain that if disagreement is resolved by relativizing apparently competing claims, then part of the phenomenon that relativism is meant to analyse is dissolved rather than accounted for. Also, part of the point of making normative judgments, especially critical ones, is often to offer people reasons for changing their behaviour. If judgments of rationality were thoroughly relativized to particular cultures, however, one could no longer recommend reasons in this way. The critical judgment would seemingly be simply out of place.

Despite the problems that some see with relativism, the position has enough intuitive appeal to have some purchase on us. In matters of taste, for instance, we seem ready to concede that two cultures could have very different standards such that one culture could judge an object beautiful, the other judge it ugly, and the best analysis be that each judgment is correct according to the standards of that culture. It is this sort of analysis that raises the problem of relativism concerning judgments of rationality across cultures.

4 Interpretation and translation

In order to make judgments of the rationality or irrationality of actions in other cultures, one must first understand the actions involved. If one is dealing with distant cultures where the native language is poorly understood or even not understood at all, problems of understanding are obvious, although the same problems occur much closer to home.

Could there be a culture so distant from ours, using norms of rationality so different from ours, that we could not successfully interpret it? In such a case, our culture and the other culture could be said to be incommensurable.

Davidson (1984), building on Quine's work on the radical indeterminacy of translation, argues that there could not be incommensurable human cultures (see Radical translation and radical interpretation). When confronted with the task of translating the language, hitherto unknown, of another culture, we should, counsels Davidson, proceed according to a principle of charity such that most of the beliefs of the people whose language is being translated come out true (see Charity, principle of §4). This would require, among other things, that the speakers of the translated language reason at least roughly as we do.

Some have argued that Davidson’s theory of radical translation leads to what may be called a problem of the attribution of irrationality, for it might seem that according to the principle of charity, it is improper to attribute any appreciable irrationality to those whose language is being translated. When faced with what seem provisionally to be irrational beliefs and practices, one should question the validity of the translation. If this is what Davidson’s theory implies, then it is ruling out irrationality in effect on purely conceptual grounds. This result would seem unduly strong to many, especially in light of the accumulated evidence from social anthropology. Others, however, feel that this interpretation of Davidson is itself uncharitable and that Davidson does not mean to deny the possibility of translating beliefs that may seem irrational to us.

We can assume for present purposes that the principle of charity might be amended into what is sometimes called the principle of humanity. The directive of the principle of humanity is to attempt to make the beliefs and practices of those translated intelligible rather than true. In certain circumstances at least, irrational beliefs could be intelligible, if within the context we could understand how the speakers could acquire them, but still not true. Practices informed by intelligible but false beliefs would thus be intelligible in light of the beliefs. Accordingly, Davidson’s approach thus understood would not rule out the attribution of some kinds of irrationality to those being translated. However, it would rule out translating a language of another culture as involving a radically different system of rationality, that is, different standards of truth and deductive and inductive inference. Attributing to a culture different standards on this basic level would make many of their beliefs and practices appear massively irrational to us, and such massive irrationality could no longer be made intelligible. Thus a strong form of cultural relativism would not be countenanced by a Davidsonian.

5 The debate and its errors

The debate over rationality and cultural relativism has focused on several related questions. The first asks whether we should apply the categories we (in the West) use to divide up cultural practices (for example, science, magic, or religion) to other, significantly different cultures? Do our categories enhance or distort our understanding of others? This question raises the subsidiary question of what has been called the pragmatics of belief: What are people using certain beliefs to do? Are they using the beliefs to explain part of the world, or are they using them to express participation in some aspect of their social reality or religious rituals, or something else again? The other question central to the debate asks when, if at all, and from what standpoint, can we judge the rationality of the beliefs, actions and practices in other cultures?

A variety of positions have been taken in answer to these questions. Before turning to the positions, however, we should note three especially contentious moves that often divide the field. The first is what many would consider a form of ethnocentrism; it holds in effect that our (Western) norms are objectively correct and may be used to assess the rationality of any and all cultures. Many find this position to be culturally insensitive and lacking sound epistemological grounding.

A second view that many reject as inadequate is the opposite of ethnocentrism, namely, a strong relativism. Apart from the various logical and epistemological problems of which relativism is accused, this view appears to many to paralyse judgment, preventing us from making the kinds of judgments about the rationality or irrationality of other cultures that seem intuitively in place.

The third contentious move in effect severs the link between agents’ reasons and actions by explaining the action in terms not immediately available to the agents. This is done in an effort to avoid making others seem manifestly irrational. For many, however, this move merely displaces the problem by reducing away the rationality of other cultures.

6 The intellectualist position

Horton classified the main positions in the debate in a way that will be largely adopted here. The first position is the intellectualist. Among the participants in the early phases of the debate, this position can be associated with Tylor and Frazer. Although their views differed somewhat, Tylor and Frazer sought to explain apparently irrational beliefs in terms of their purpose. For both, the general purpose of belief systems, as they understood them, is to explain, predict and control the world. On this view, systems of beliefs informing practices such as magic, witchcraft and primitive religion are seen as primitive theories, similar to Western science but for the most

part inferior to it, proto-sciences one might say.

Horton (1993) has developed the most sophisticated rendering of the neo-intellectualist position. In his analysis of traditional West African belief systems, Horton finds norms of reasoning and attempts at theory building that are similar to those at work in Western science. African systems often display considerable sophistication and explanatory prowess. Still, for the most part, their explanations are not as successful as those of our sciences. However, in certain instances, Horton notes, African systems provide arguably superior accounts of some phenomena.

By and large, the intellectualist position imports our categories such as science and magic and puts them to work in analysing other cultures. In addition, the position allows for cross-cultural judgments of rationality. Thus, in instances where the norms of a culture appear to countenance the seeming disregard of relevant evidence, or where certain inductive inferences are not drawn upon, for example, the non-efficacy of a rain dance, we can make the third level judgment that the norms are rationally deficient compared to our own.

The intellectualist takes the belief and desire statements of the members of other cultures literally as translated. The beliefs and desires are then construed as the agents’ reasons for action and a corresponding attempt is made to understand how the agents act in their cultural context. Beliefs that, when translated, appear wrong or irrational are judged false and seen as perhaps acquired in an irrational way.

7 The functionalist and symbolist positions

The second position to be examined does not approach the problem in this way. According to this position, beliefs that appear initially to be false or irrational or that appear not to offer good reasons for the associated action are not subject to rational evaluation. Rather, an alternative interpretation is offered that makes the actions rational, even if it means decoupling the actions from the reasons for it offered by the agents. Actions or practices are understood as rational in terms of their function, according to one alternative of this position, or in terms of their symbolic meaning, according to the other. Beliefs are either not considered as part of the explanation of the action or practice, or taken as expressive but not descriptive, as having a predominantly symbolic meaning. In either case, beliefs are not subject to assessment in terms of their literal meaning.

The two variants of this position, the functionalist and the symbolist, should not be entirely conflated, of course. For a functionalist, such as Radcliffe-Brown, the problem of rationality does not come up in the standard form discussed in this entry. The functionalist focuses on explaining the existence and persistence of certain social practices in terms of the effects they produce, understood against a theory of necessary social functions (see Functionalism in social science §2). Questions of the rationality of belief as informing the practices is at best of secondary interest.

In contrast, symbolists, such as Beattie, and with some qualifications, Sperber, are primarily concerned with understanding beliefs. They typically see human actions as having an expressive dimension. Beliefs that taken literally as translated seem to be indicative propositions, and problematically rational as such, symbolists interpret as having essentially symbolic meaning. They argue that this dimension of meaning can only be fully recovered by understanding the symbolic or expressive content of the ritual or practice in which the belief is embedded. Once the symbolic meanings of the belief and practice are recovered, there are no further issues of rational evaluation to be raised.

Many see a basic problem with the functionalist and symbolist approaches. In neither view is it clear how to connect the agents’ reasons to their actions. Rational agents act for reasons, expressed in terms of their beliefs and desires as they understand them. They do not typically act to satisfy the function assigned to a practice by the functionalist. Nor, for the most part, do they understand their beliefs as primarily expressive or symbolic utterances. However, reasons as understood by the agents themselves do not enter into the functionalist’s or symbolist’s explanation of the rationality of the action. Thus, agents do not appear as rational actors on the first level noted above, where actions are brought about for and in light of reasons. Practices may be understood as rational, given these approaches, but we lose our grip on the agents as rational actors.

8 The fideist position

The third position, the fideist, is most associated with Winch and influenced by Wittgenstein's remarks on Frazer's *The Golden Bough* (Wittgenstein 1979). Wittgenstein criticized Frazer in rather harsh terms for interpreting early magical and religious accounts as if they were putting forth theories to explain the world, and then judging them to be inferior, mistaken accounts at best. It is nonsense, Wittgenstein argued, to treat, as he thought Frazer did, magical accounts as a kind of false physics. Rather, we should attempt to appreciate their ritualistic and spiritual meanings as part of a culture. According to the fideist, the intellectualist imposes Western categories on other cultures too uncritically and reaches judgments of irrationality too easily. What is crucial for the fideist is attempting to understand, to as great an extent as possible, beliefs and practices as they are understood from within the other culture, or to use the Wittgensteinian term, the other form of life. Unlike the symbolist, the fideist does not want to impose a level of symbolic meaning on beliefs that is not accessible to the speakers themselves. We are to assume that agents act for reasons that they understand, but we have to be very careful about how we establish the meanings of propositions stating the agents' beliefs and desires.

Furthermore, the fideist is reluctant to make judgments of rationality or irrationality. Just as we have to understand meaning from the point of view of the agents, so we have to understand their concepts and norms of rationality. Those are the only norms that can properly be used to evaluate their beliefs and practices. Even such basic logical concepts as consistency and contradiction cannot be assumed to be held in common with us and we should not impose our concepts on others who may have very different ones. Thus, the fideist position seems to endorse a form of cultural relativism.

The fideist, then, would argue that we must stop at level two judgments of rationality, that is, judgments of the rationality of actions relative to the norms of the culture in which they occur. For the fideist, level three judgments are either incoherent, wrong, or pointless. They are incoherent if it is thought that there can be culture-independent judgments. For the fideist, all judgments are relative to a culture or form of life; we simply have to give up the idea of 'the view from nowhere', to use Nagel's phrase. Third level judgments are wrong if, thinking we are being neutral, we merely impose our norms on others. Norms and concepts cannot, according to the fideist, be exported beyond the boundaries of the form of life in which they function and have meaning. And such judgments are pointless if all we are doing is using the norms of the agents' own culture. In that case, there is no distinction between level two and level three judgments, for the norms of rationality of a culture are of course going to be judged rational from within that culture.

In so far as fideists endorse some form of relativism, they will have to contend with the various criticisms aimed at relativism in general. For instance, it is not clear that the fideists' relativism is consistent with a Davidsonian approach to interpretation. Fideists have served the important function of warning the intellectualist, and in a different way, the symbolist against uncritical use of our categories to understand others. If other cultures do have very different cultural norms and practices, we have to attempt to understand them in their own terms. It may be, for instance, that a very different culture has nothing quite analogous to Western science, or for that matter, to the practice of Western religion. What may seem initially like a rather irrational form of science might be better understood in very different terms. In cases where practices seem all but incommensurable, judgments of comparative rationality may be out of place.

Still, despite the objections of some fideists, judgments of rationality may have their place even here. What is needed is a way of comparing across cultures practices that initially seem to resist being placed in the same categories. Some have suggested that a properly developed theory of human interests might afford such common ground. Furthermore, to the degree that one is convinced by Davidson's theory of interpretation, one cannot allow for different basic norms of rationality. It would follow, then, that there are norms of rationality with which one can judge the beliefs and practices of others (and of ourselves), once the beliefs and practices are properly understood.

See also: Epistemic relativism; Moral relativism; Relativism; Social relativism

LAWRENCE H. SIMON

References and further reading

Davidson, D. (1984) *Truth and Interpretation*, Oxford: Oxford University Press.(A collection including most of Davidson's important essays on radical interpretation.)

Hollis, M. and Lukes, S. (eds) (1982) *Rationality and Relativism*, Cambridge, MA: MIT Press.(An important collection including papers by Hollis, Taylor, Sperber, Horton and Lukes. Good bibliography.)

Horton, R. (1993) *Patterns of Thought in Africa and the West*, Cambridge: Cambridge University Press.(A collection of Horton's papers stating his neo-intellectualist position. Good bibliography.)

Skorupski, J. (1976) *Symbol and Theory: A Philosophical Study of Theories of Religion in Social Anthropology*, Cambridge: Cambridge University Press.(An interesting philosophical discussion centring on the intellectualist position.)

Tambiah, S.J. (1990) *Magic, Science, Religion, and the Scope of Rationality*, Cambridge: Cambridge University Press.(An interesting survey of the debate by a leading anthropologist. Useful bibliography.)

Vico, G.B. (1744) *The New Science of Giambattista Vico*, revised trans. of 3rd edn, T.G. Bergin and M.H. Fisch, Ithica, NY and London: Cornell University Press, 1968.(Vico's attempt to develop a new science of human history.)

Wilson, B.R. (ed.) (1970) *Rationality*, Oxford: Blackwell.(An important collection of papers from the debate, mostly from the 1960s. Contributors include: Winch, MacIntyre, Horton, Beattie, Lukes and Hollis.)

Winch, P. (1958) *The Idea of a Social Science and its Relation to Philosophy*, London: Routledge & Kegan Paul.(The classic statement of Winch's views.)

Wittgenstein, L. (1979) *Remarks on Frazer's 'Golden Bough'*, ed. R. Rhees, trans. A.C. Miles and R. Rhees, Atlantic Heights, NJ: Humanities Press.(Wittgenstein's criticism of Fraser.)

Rationality of belief

Humans, claimed Aristotle, are rational animals. However, recent psychological studies purport to show that people systematically deviate from canons of logic, probability theory, decision theory and statistics. Interpretations of these studies differ about the nature of the errors: on some views, there are no real errors at all; on others, subjects are distracted from applying the proper formal rules by, for example, the influence of conversational expectations. A common suggestion is that the reasoning subjects display is nearly optimal, once we take account of our severe cognitive limits in realistic circumstances or in our evolutionary history.

These diverse views reveal two opposed tendencies in constructing a theory of rationality. One favours the formal rules which, if violated in real life, could have serious consequences. The other holds that people's actual practice provides the only standard. A popular model, incorporating both tendencies, holds that those principles or norms are justified that yield the best balance between our reasoning intuitions and the demands of theory or system.

1 Empirical studies of reasoning

Reasoning is central to the acquisition and improvement of our beliefs. If a belief is recognized as sustained through defective reasoning, it is not rational to maintain the belief. Consequently, a good deal of interest attends recent psychological studies (for example, Wason and Johnson-Laird 1972; Kahneman *et al.* 1982) that purport to show serious and systematic deviation from canons of logic, probability theory, decision theory and statistics. Although these canons do not function as theories of what *should* be believed (Harman 1986), they do provide serious *constraints* on reasoning and inference. Because these recent studies seem to show people violating these constraints, they are troubling and important.

Four of the best known of these studies are: the 'selection task', which investigates deductive reasoning with conditionals; the 'conjunction problem', the 'base rate' problem, and the 'preference reversal' problem, all of which involve features of probabilistic reasoning. (The reader is advised to try each one, before reading the results.)

The selection task. Four cards are placed in front of each subject.

$E \qquad K \qquad 4 \qquad 7$

The subjects know that on one side of each card is a letter and on the other a number. The following rule is presented: if a card has a vowel on one side, then it has an even number on the other. The task is to pick out just those cards that need to be turned over in order to determine whether the rule is true or false. Almost all subjects turn over the E card, and many the 4. More crucially, few turn over the 7 card (Wason and Johson-Laird 1972).

Conjunction. Subjects are offered the following personality sketch:

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

They are then asked to rate the probability of the following statements:

(1) Linda is a psychiatric social worker.
(2) Linda is a bank clerk.
(3) Linda is an insurance salesperson.
(4) Linda is a bank clerk and is active in the feminist movement.

The main finding in the conjunction study is that 'Linda is a bank clerk and a feminist' is judged more probable than that 'Linda is a bank clerk' (Tversky and Kahneman 1983).

Base rates. Subjects are told that a cab has been involved in a hit-and-run accident at night. They are given the following data:

(1) 85% of the cabs in the city are green and 15% are blue.
(2) A witness identified the cab as blue.
(3) The court tested the reliability of the witness under the same circumstances that existed on the night of the

accident and concluded that the witness correctly identified each one of the two colours 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was blue rather than green? Subjects largely assign the probability 0.8; that is, their judgment is based wholly on the reliability of the witness. But if the base rates (0.85 and 0.15) are used as prior probabilities, then the answer is 0.41 (Kahneman *et al.* 1982: 156-7).

Preference reversal. Subjects are offered two gambles (at different times):

(a) Choose between
gamble 1. £500,000 with probability 1; and
gamble 2. £2,500,000 with probability 0.1, £500,000 with probability 0.89.
(b) Choose between
gamble 3. £500,000 with probability 0.11; and
gamble 4. £2,500,000 with probability 0.1.

The study is cited as of 'preference reversal' because it is generally found that subjects prefer gamble 1 to gamble 2 in situation (a), while they prefer gamble 4 to gamble 3 in situation (b). Yet, one should prefer gamble 1 to 2 if and only if one prefers gamble 3 to 4, since they are equivalent. (See essays by Allais, Ellsberg, Kahneman and Tversky, and Savage in Moser (ed.) (1990).)

In each case the violation is of a simple and familiar formal principle. In the selection task, a statement of the form 'If *p*, then *q*' is incompatible with one of the form '*p* and not-*q*' (for example, a vowel on one side, and an odd number on the other). In the conjunction task, the probability that someone satisfies a *conjunction* of properties (being both a feminist and a bank clerk) cannot be greater than the probability of their satisfying one of the conjuncts *separately* (being a bank clerk alone).

The correct way to calculate the probability in the base-rate case is by using Bayes' theorem, which says that the posterior (or, conditional) probability of a hypothesis *h* (the cab was blue) given evidence *e* (the witness saw a blue cab) is determined by how likely *e* and *h* are to be true simultaneously and by the probability of *e* being true at all. But subjects ignore the likely colour of the cab and judge only on the basis of the reliability of the witness, disregarding the probability of the cab's actually being blue. In the 'preference reversal' case, one is assigning a utility to an outcome in one gamble that is inconsistent with the utility assigned to it in the other (see Rational choice theory; Rationality, practical).

2 Interpretation

Some hold that we should not treat these results as deserving a uniform account. In particular, dispute has focused on whether in the base-rate studies, subjects' answers are, in fact, incorrect. For example, in an article that generated much controversy, Cohen (1981) argued that the Bayesian analysis using the base rates as prior probabilities would be correct, if the question concerned repeated or long-run samplings. But subjects reasonably interpret their task as the determination of the colour in this particular accident. From the information given, only the reliability of the witness is relevant to that determination. The base rates would also be relevant if they were not mere frequencies, but reflected causal factors involved in an accident, for example, if the 85-15 difference was a difference in respective accident rates. In fact, when the base rates have a natural causal explanation, it was found that subjects *do* incorporate them into their answers (Kahneman *et al.* 1982: essays 8, 10).

Among further interpretational questions, four deserve mention.

(1) The task may generate conversational implicatures that lead subjects away from the target formal representation (see Grice 1989; Sperber *et al.* 1995; Implicature). For example, the statement 'Linda is a bank clerk' in the context would appear conversationally to implicate that Linda is a bank clerk *and not a feminist*. Experimenters try to control for conversational inferences. For example, instead of offering 'Linda is a bank clerk' they might offer the alternative of 'Linda is a bank clerk, who may or may not be a feminist'. Even so, it remains questionable whether habitual conversational expectations are so easily set aside (Adler 1991).

(2) Some variant presentations yield different patterns of response. In the base-rate or conjunction studies,

presentation in a *frequency* format (citing *number* of cases, rather than *ratios* or *probabilities*) leads to many more subjects responding according to a straightforward probabilistic analysis (Girgerenzer 1991).

(3) In the case of preference reversal, alternative understandings may be legitimate so that, for example, the achievement of certainty is itself valued, rather than being treated as just another degree of belief (Schick 1991).

(4) Most dramatically, at least in the selection task, when the problem is posed for a natural setting, there is enormous improvement. In one set-up there is an array of envelopes (face down and sealed; face down and unsealed; face up with a 50 lire stamp on it; face up with a 40 lire stamp on it) and the rule 'If a letter is sealed, then it has a 50 lire stamp on it'. Once subjects are presented with a realistic model, they readily recognize that the envelope with the 40 lire stamp on it (the analogue of the '7' card) must be turned over. What is crucial is the realism of the conditional, not its familiarity (Wason and Johnson-Laird 1972). In particular, striking improvement occurs when the conditional captures a connection or rule of practices involving permissions or obligations. Recent work treats such rules as modularized, products of social exchange important in our hunter-gatherer evolutionary history (Cosmides and Tooby 1996).

These interpretational questions suggest that subjects orient the problems to common expectations or habits, rather than representing their task as clearly calling on formal principles. What is it about the task or set-up that directs subjects away from representing the problem so that the familiar formal rules are seen as applicable?

The guiding thought in most answers is that these results actually reveal near optimal reasoning, given the demands on us for cognitive economy in the environments in which our reasoning capacities evolved. If subjects formed the logically equivalent contrapositive of the above rule, they would turn over the '7' card: since 'If that bird is a raven then it is black' is equivalent to 'If that bird is non-black then it is a non-raven'. But normally the contrapositive of an ordinary conditional as it involves the complement class of a natural class is not itself a natural class: black ravens form a natural class, but non-black, non-ravens do not. We do not ordinarily test whether ravens are black by examining non-black things (for example, pennies) to see whether they are non-ravens (see Confirmation theory; Hempel, C.G. §2).

Nor do we ordinarily reason with the kind of *static* information characteristic of these studies. Although other studies do reveal a tendency towards belief perseverance or conservatism (Kahneman *et al.* 1982: essays 9, 15, 25), in general dynamic or continuous reasoning is much less investigated. Where feedback from the environment is expected, the cost of initial errors may be compensated for, so long as we remain open to subsequent modification or correction (Hogarth 1986).

In reflecting on what these studies reveal about the rationality of our beliefs, there are two opposed tendencies. One says that to violate or bypass the formal rules is to commit errors, whose analogues in real life can be serious. The other says that the formal rules, and the experts in the corresponding formal systems, are not authoritative on how it is reasonable to understand the problems: we are. Suggestive of a 'naturalized epistemology', there is no authority about reason apart from the actual reasoning processes or intuitive judgments of human beings (see Naturalized epistemology).

A popular model, incorporating both tendencies, holds that those principles or norms are justified that yield the best balance between our reasoning intuitions and the demands of theory or system. The model has been especially attractive to those who hold that it is an a priori truth that canons for human reasoning must reveal its rationality, since they are derived from human judgment, at least when suitably idealized (Cohen 1981). However, the balance model supports this strong a priori rational view only if the judgments in these studies do count as telling intuitions and these intuitions are sufficiently systematic. But given the dependence on changes in content, this is disputable. Moreover, formalization allows the generation of powerful theories based on very weak axioms: the resulting gains in systematicity could quickly outweigh any appeals to the kinds of intuitions revealed by these psychological studies.

See also: Confirmation theory; Inductive inference; Rational beliefs; Rationality and cultural relativism

JONATHAN E. ADLER

References and further reading

Adler, J.E. (1991) 'An Optimist's Pessimism: Conversation and Conjunction', in E. Eells and T. Maruszewski

(eds) *Probability and Rationality*, Amsterdam: Poznan, 251-82.(The first part explores conversational influences on reasoning, especially in regard to the conjunction studies.)

Cohen, L.J. (1981) ‘Can Human Irrationality be Experimentally Demonstrated?’, *Behavioral and Brain Sciences* 4: 317-70.(With commentary in this and subsequent issues. A comprehensive critique of the experimental literature, sparking a lively debate. The nonstandard base-rate analysis is defended, via an alternative notion of probability, and an a priori balance argument is offered for the necessity of our competence in reasoning.)

Cosmides, L. and Tooby, J. (1996) ‘Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgments under Uncertainty’, *Cognition* 58: 1-73.(Building on Cosmides’ demonstration of great improvement on the Wason selection task through content corresponding to realistic problems faced by our ancestors, the two authors of this paper propose an evolutionary approach to the mind with particular application to the above studies.)

Fischhoff, B. (1982) ‘For Those Condemned to Study the Past: Heuristics and Biases in Hindsight’, repr. in D. Kahneman, P. Slovic and A. Tversky (eds) *Judgment under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press, 335-51.(One of a number of the author’s important studies of overconfidence due to knowledge after the fact.)

Girgerenzer, G. (1991) ‘How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases"’, *European Review of Social Psychology* 2: 83-115.(One of many studies by this author showing how variations emphasizing more extensional or frequentist representations facilitate judgments in line with probabilistic analyses.)

Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(Critically examines research in cognitive psychology, generally, through application of an epistemological framework. Forceful commitment to interdisciplinary research. A good deal of part 2 is devoted to the above studies.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, esp. essays 2, 3, 7, 17.(Classic presentation of influential theory of conversational reasoning and pragmatics. Methodologically fundamental for analytical philosophy and cognitive psychology.)

Harman, G. (1986) *Change in View*, Cambridge, MA: MIT Press.(An elegant approach to reasoning and change of belief, sensitive to our cognitive limits and the empirical literature, while adding novel ideas.)

Hogarth, R.M. (1986) ‘Beyond Discrete Biases: Functional and Dysfunctional Aspects of Judgmental Heuristics’, repr. in H.R. Arkes and K.R. Hammond (eds) *Judgment and Decision Making: An Interdisciplinary Reader*, Cambridge: Cambridge University Press, 680-704.(Argues for the importance of studying reasoning in continuous, not just static, environments.)

Johnson-Laird, P.N. (1983) *Mental Models*, Cambridge, MA: Harvard University Press.(An original approach to reasoning, especially deductive, turning on our interpretation of formal or syntactic presentations of inferences.)

Kahneman, D., Slovic, P. and Tversky, A. (eds) (1982) *Judgment under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.(A basic source book containing most of the research on probabilistic reasoning and many related topics, including overconfidence, risk perception and corrective procedures. Extensive references.)

Moser, P. (ed.) (1990) *Rationality in Action: Contemporary Approaches*, Cambridge: Cambridge University Press.(Includes some of the most important early and recent essays on the nature and applicability of decision theory and related problems, including the prisoner’s dilemma, Newcomb’s problem, and the Arrow-Sen results on social choice.)

Nisbett, R. (1993) *Rules for Reasoning*, Hillsdale, NJ: Erlbaum.(Studies defects in reasoning with attention to both what content in problems facilitates success and what teaching will lead to improvement.)

Nisbett, R. and Ross, L. (1980) *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, NJ: Prentice Hall.(Highly readable. A classic and comprehensive account of social judgment research, including work on heuristics, attribution and limits on our self-knowledge. Extensive references.)

Piattelli-Palmarini, M. (1994) *Inevitable Illusions*, New York: Wiley & Sons.(Highly engaging and accessible presentation of most of the problems studied in the above literature on reasoning. Argues that the results display serious, even if natural, failings of reason. Helpful guide to readings.)

Rips, L. (1994) *The Psychology of Deductive Reasoning*, Cambridge, MA: MIT Press.(Comprehensive and sophisticated studies of deductive reasoning, powerfully arguing for a large role for a syntactic, ruled-based account.)

Schick, F. (1991) *Understanding Action*, Cambridge: Cambridge University Press.(Defends different understandings of the same problem.)

Sperber, D., Cara, F. and Girotto (1995) ‘Relevance Theory Explains the Selection Task’, *Cognition* 57: 31-95.

(Thorough attempt to explain the ‘selection task’ in terms of Sperber’s influential theory of conversational reasoning.)

Stich, S. (1990) *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*, Cambridge, MA: MIT Press.(Failure of the balance model as an account of these studies, among other reasons, leads to a renouncing of any a priori epistemology in favour of a relativism and pragmatism.)

Tversky, A. and Kahneman, D. (1983) ‘Extensional Versus Intuitive Reasoning: The Conjunction Fallacy Improbability Judgment’, *Psychological Review* 90: 293-315.(The conjunction study, with many variations and extensive discussion.)

Wason, P.C. and Johnson-Laird, P.N. (1972) *Psychology of Reasoning: Structure and Content*, Cambridge, MA: Harvard University Press.(Especially good on studies of deductive reasoning, including the selection task, which one of the authors invented. Out of date now, but still a very helpful presentation of studies on reasoning from the early 1930s until the late 1960s.)

Rationality, practical

Whereas theoretical reason is that form of reason that is authoritative over belief, practical reason is that form of reason that applies, in some way, to action: by either directing it, motivating it, planning it, evaluating it or predicting it. Accounts of practical reason include theories of how we should determine means to the ends we have; how we should define the ends themselves; how we should act given that we have a multiplicity of ends; how requirements of consistency should govern our actions; and how moral considerations should be incorporated in our deliberations about how to act.

Economics has provided, in recent times, what many regard as the most compelling portrait of practical reason, called 'expected utility *theory' (hereafter 'EU theory'). On this theory, rational action is that action which yields the highest expected utility, which is calculated by measuring the utility - or the 'goodness' or 'badness' - of the possible outcomes of the action, multiplying the utility of each outcome by the probability that it will occur, and, finally, adding together the results for all the possible outcomes of each action. The action that has the highest expected utility is the rational action. Other technical representations of practical reason have been explored in the branch of social science called 'game theory', which studies 'strategic' situations in which the action that is rational for any agent depends in part on what other agents do.*

A theory of practical reason can have one or more of several different goals. If it sets out how human beings actually *reason, it functions as a* descriptive *theory of reasoning. If it sets out a conception of how our reasoning* ought *to proceed, it functions as a* normative *theory of reasoning. Theories of reason can also be about actions themselves: if a theory presents a conception of the way our actions should be intelligible or consistent or useful (regardless of the quality of the deliberation that preceded it), it functions as a (normative) theory of behavioural rationale. If it merely presents an account of consistent action that allows us to predict the behaviour of an agent whose previous actions fit this account of consistency, it functions as a descriptive theory.*

One might say that whereas theoretical reason is supposed to pursue truth, practical reason is supposed to pursue some sort of good or value in human action. Theories that take rational action to be that which achieves, furthers or maximizes (what is regarded as) good, are consequentialist *or* teleological *theories. Theories that believe rational action must sometimes be understood as action that has an intrinsic value or 'rightness' regardless of how much good it will accomplish or manifest, are non-consequentialist or non-teleological conceptions of reason. If the theory defines reason as that which serves ends defined by something other than itself, it is an* instrumental *conception. If it allows reason to have a non-instrumental role, itself capable of establishing at least some of our ends of action, it is setting out a non-instrumental conception. Theories of practical reason that recognize the existence of a special moral reasoning procedure tend to represent that procedure as non-instrumental.*

Philosophers have disagreed about whether practical reason gives us a way of reasoning prior to choice that can actually motivate us to behave in the way that it directs. Many believe it lacks motivational power, so that it can only give us authoritative directives that must be motivated by something else (for example, by our desires). Finally, the study of practical reason also considers the variety of ways in which one can fall short of being rational; and issues about the nature and possibility of irrational 'weakness of will' have been central to this discussion.

1 Representations of practical reasoning

Historical suggestions. There have been many theories of the nature of practical reason over the years, going back to the Ancient Greeks.

Aristotle, although never offering a detailed theory of practical reason, made certain remarks in various of his writings that assimilated practical reasoning to a kind of logical reasoning. One example of (what are called) Aristotelian practical syllogisms appearing in Aristotle's *Nicomachean Ethics* (1147a5) goes as follows:

Dry food suits any human.
Such-and-such food is dry.
I am human.
This is a bit of such-and-such food.

Therefore, this food suits me.

But why is this syllogism an example of practical reasoning? On the face of it, it is a logical deduction, which is a form of theoretical reasoning. Aristotle can only claim it to be a form of practical reasoning if this logical deduction eventuates in some kind of imperative with authority over action. But in the argument presented above, the conclusion is not in imperative form, and the argument makes no direct connection between the conclusion (that this food suits me) and any actions or intentions to act that I do or should have. Nor does Aristotle tell us how reasoning, operating practically, would produce a connection between the conclusion of this syllogism and my actions or intentions.

In more recent times, philosophers have argued that practical reason produces imperatives of the form ‘If you desire *x*, then since *y* is a means to *x*, do *y*’. Immanuel Kant (1785) refers to these as ‘hypothetical imperatives’ (see Kant, I. §9). So on this view, a hypothetical imperative that is authoritative over action eventuates from practical reasoning, understood as a process by which we work out effective means to the given ends of an agent. To use Aristotle’s example, if I were hungry, I would engage in practical reasoning to determine a kind of food that suits me, where my reasoning would (if I had sufficient information about the world) eventuate in a hypothetical imperative of the form ‘If I am hungry, I should eat such-and-such dry food’.

But if this is how we understand practical deliberation, it seems little different from theoretical reasoning. On this view, deliberation is primarily a matter of working out causes and effects, and if that is all reason is doing, it appears to be functioning merely theoretically, using the principles of causal reasoning. Indeed, one interpretation of the theory of reason offered by David Hume (1739-40) is that reason only has a theoretical function, and thus is a purely informational faculty, working out relations of ideas and causal connections, with no authority or motivational efficacy with respect to action (see Hume, D.). So interpreted, Hume is saying that there is no such thing as a form of reason that chooses or judges action - but only a form of reason by which we form and judge belief that is (merely) informational with respect to action. However, even among those who accept the validity of both practical and theoretical reason, there has been ongoing dispute about whether these forms of reason are genuinely distinct; even assuming that they are not, there is disagreement: some - including Plato (in the *Republic*) - have held that practical reason derives from theoretical, others (for example, Goldman 1986) have said that the theoretical derives from the practical, and still others (for example, Kant 1785) have claimed that practical reason and theoretical reason are two forms of the same thing.

Expected utility theory: the basic idea. Defenders of the reality of practical reason have attempted to define practical reasoning so as to show how it differs from theoretical reasoning. The most popular of these theories in modern times is highly mathematical, in ways that highlight how complicated and intricate idealized decision making with respect to action might be. Called ‘expected utility theory’, it was developed by John von Neumann, Oskar Morgenstern and Leonard Savage.

To get a rough idea of what the theory says, imagine yourself in a situation where you do not know for certain what the results of your various alternative courses of action might be. Suppose, for example, that you have the option to buy a lottery ticket: if it is the winning ticket, you will receive a million dollars; if it is not, you will receive nothing and you will lose the money (let us say $10) that it cost you to purchase the ticket. If you do not buy the ticket at all, you will not win, but neither will you lose. So you have a choice between a risky action (buying the ticket) and an action the results of which you know for sure, which avoids both losses and gains. What do you do? The idea behind expected utility theory is that for each possible action available to you, you develop some measure of the ‘goodness’ or ‘badness’ of that action by, first, evaluating the goodness or badness of each possible alternative outcome, second, multiplying this value by the probability of that outcome’s occurring and, finally, adding together the products for all the possible outcomes of each action.

So, in our lottery example, you proceed, first, by developing a measure of the goodness of winning the lottery prize and the badness of losing the prize and being $10 out of pocket. As we shall discuss in §3 below, there is more than one way that people who have endorsed expected utility theory have thought such a measure should be constructed; the result is standardly referred to as the ‘utility’ of the outcome. In general, positive, pleasurable or beneficial outcomes of an action are represented using a positive number, whereas negative, costly or painful results are represented with a negative number.

Second, you should multiply each of these numbers by the probability of winning the lottery or losing the lottery. The probability of a sure thing is taken to be 1; in our lottery case, this means that the probability of gaining nothing if you do not buy a ticket is 1. Note also, however, that if you do buy a ticket, it is a sure thing that you will either win the lottery or not win the lottery, meaning that the sum of the probabilities of these mutually exclusive outcomes should equal 1. Hence if *p* is the probability of winning the lottery, we can express the probability of not winning the lottery as $1 - p$ (see Probability theory and epistemology).

Third, you should add together the products of the utility and probability numbers to get the 'expected utility of the action of buying the lottery ticket'. So, if the probability of winning the lottery is *p*, and the probability of not winning the lottery is $1 - p$, and if (as economists often do) we identify utility with money, so that the utility of winning the lottery is 1,000,000 and the utility of not winning the lottery is -10, the expected utility (EU) of buying the lottery ticket is

$$(1,000,000)p + (-10)(1 - p).$$

After determining the EU of buying the ticket, you then compare it to the expected utility of not buying the ticket. Since you will not win or lose anything if you do not buy the ticket, the utility of the result of this action is 0. That number is multiplied by the probability 1 (since it is a sure thing that you will neither win nor lose if you buy the ticket), making the expected utility of not buying the ticket equal to 0. So if the expected utility of buying the ticket is greater than 0, you should buy the ticket, despite the risk of losing, and if the expected utility of buying the ticket is less than 0, you should not buy the ticket despite the chance of making great gains. In a way, one can think of expected utility theory as telling you how to determine, in each of life's risky situations, when and if a certain course of action is a 'good bet' or a 'bad bet' given the possibilities.

2 Categories of theories of practical reason

Descriptive v. normative theories of reasoning and rationale. In evaluating whether expected utility theory is a successful theory of practical reason, we should consider, first, what a theory of practical reason should be expected to do.

There have been a number of distinctive tasks that philosophers have thought their theories of practical reason must perform. For example, it is commonly thought that there is a kind of deliberation or reasoning process that does or should occur in the minds of agents prior to action, which authorizes that action and may also motivate it. Theories that describe how we *actually* deliberate prior to action are *descriptive* theories of reasoning. Theories that prescribe how we *ought* to deliberate prior to action (where this may diverge substantially from our actual reasoning process) are *normative* (or *prescriptive*) theories of reasoning.

Theories of behavioural *rationale* are a kind of normative theory that specifies one or more standards by which the behaviour of an agent is evaluated. It is a common (and understandable) presumption of many that the principle(s) used to judge behaviour should be the same as the principle(s) on which human reasoning is based prior to choice, and normative versions of theories of reasoning often function as theories of behavioural rationale. None the less, it may be that the correct conception of behavioural rationale is one that, given human constraints and limitations, is too difficult for us to use as a reasoning process, in which case, the theory used to judge the rationality of behaviour would diverge from the (less complicated) reasoning procedure human beings would be able to use prior to action.

A *predictive* theory of reason predicts the behaviour of an agent who satisfies certain consistency requirements. Such a theory may be (and often has been) linked to a theory of rationale when these consistency requirements are taken to manifest a certain normatively desirable consistency in choice. Some might speculate that if a predictive theory made successful predictions of an agent's choices, then that success would provide evidence that the agent (in some way) mentally followed the procedure described by that theory to make those choices, so that it was (in some way) a description of the reasoning processes upon which the agent operated. But the success of these predictions does not require that one draw this descriptive conclusion, and might be explained in other ways.

Two further distinctions in practical reason. There are important further issues concerning how a theory of practical reason should bear upon action. For example, should rational actions be understood as those actions that promote (what are regarded as) good objects or good states of affairs? Or as those actions that are, in some way,

intrinsically valuable or right, regardless of the good they generate? Those theorists who believe that rational actions must be defined as those actions that maximize, achieve or promote the good have a *consequentialist* or *teleological* conception of reason. In contrast, *non-consequentialist* theorists believe that a rational action has certain characteristics that make it rational other than its tendency to promote the good. They try to define rational action, at least in part, by reference to principles of what might be called inherent rightness.

Different from this distinction is that between *instrumental* and *non-instrumental* conceptions of reason itself. An instrumental conception takes reason to serve ultimate goals that are set by something *other than reason itself*. (Hume is often considered to be the source of this idea.) An instrumental theory can allow the possibility that reasoning can set goals by functioning instrumentally. That is, we can distinguish between *motivated* ends, that is, ends that are means to the attainment of other ends, and unmotivated ends, that is, ends that are not means to the attainment of other ends. Many instrumental theories accept that reason can at least fix our motivated ends; others deny even this possibility. All of them, however, agree that our unmotivated ends are fixed by non-rational forces within us, for example, our basic or innate passions or desires (see Desire).

Nor need an instrumental theory be committed to the claim that the only form of reasoning is means/end reasoning; there are other tasks that reason can perform consistent with a denial of a non-instrumental role for reason. For example, practical reasoning can include not only deliberation about instrumentally effective means to ends, but also deliberation concerning the most convenient, economical, or pleasant way of satisfying some element in an agent’s motivational set, as well as deliberations with respect to time-ordering, or weighting of conflicting desires. It is only the non-instrumental fixing of ends that the instrumental theory leaves beyond the province of rational criticism.

Non-instrumental theories of reasoning or rationale, while admitting reason has an instrumental component, also ascribe to reason a role in defining unmotivated ends of action. Such views may hold, for example, that certain things are worth having for their own sake (for example, appetitive pleasure, comfort, aesthetic appreciation, certain sorts of accomplishment) and that practical reason is capable of recognizing their appropriateness as (unmotivated) ends of human action or activity.

3 The interpretation of EU

With these distinctions between kinds of reasons in mind, we can ask whether EU theory is a successful portrayal of practical reason. It all depends upon how one interprets the theory and, surprisingly, it has been interpreted in many different ways.

EU theory can be either an instrumental or a non-instrumental theory of reason, depending upon how one understands the source of the measure of the ‘utility’ or goodness of the outcomes used to evaluate the expected utility of the actions being assessed. Suppose, for now, that we adopt the account of Jeremy Bentham (1823) that regards this measure as a way of assessing the amount of pleasure or pain that one will receive from the results of an action. In this case, the goodness or badness of any outcome is defined by the way the agent experiences that outcome, where this experience has nothing to do with the agent’s reasoning process. Pleasure and the absence of pain can therefore be understood as the agent’s good (a good that, with each action, the agent attempts to achieve or maximize), where the content of this good is not in any way determined by reason operating non-instrumentally. Expected utility theory, on this view, is merely an account of the reasoning process we use in attempting to maximize this good, and thus should be understood as an instrumental theory.

However, von Neumann and Morgenstern (1944) developed a notion of ‘utility’ that is deliberately not meant to refer to an experiential state or to be an interpersonal measure of such an experiential state. The notion of preference can be (and has been) understood either as a mental state (for example, a kind of disposition to choose), or as something not merely revealed by, but actually consisting in, a person’s publicly observable behaviour (commonly referred to by economists as the idea of ‘revealed preference’). In their development of EU theory, an agent’s utility function attempts (merely) to represent an agent’s preferences understood in either of these ways, but the theory does not offer any explanation of why the agent has these preferences. Hence the theory leaves open the possibility that what an agent prefers may be the result of reason operating non-instrumentally. Only a psychological theory of preference formation could tell us whether reason does or does not play a role in determining an agent’s preferences (interpreted in either of the ways just mentioned). This means that von

Neumann and Morgenstern's EU theory can receive either an instrumental or a non-instrumental interpretation, depending upon the kind of psychological theory of preference formation appended to it.

Interestingly, however, von Neumann and Morgenstern appended no such psychological theory to their theory. Both were influenced by the philosophical position called logical positivism, which advocates that theories, to be scientific, must concern only that which is publicly verifiable - something that behaviour and choices seem to be, unlike deliberation or inner experiences (see Behaviourism, analytic; Logical positivism). However, a behaviouristic interpretation of the theory (which would embrace the idea of revealed preference) precludes us from developing it in a way that determines whether it is instrumental or non-instrumental.

Von Neumann and Morgenstern's non-psychological understanding of their theory is connected to their interest in developing a theory that could function *predictively*, regardless of any use it might have as a theory of reasoning or behavioural rationale. On their view, we interview or observe a person in order to get them to (arbitrarily) assign utility numbers to certain outcomes which they know they will get. Call these 'riskless prizes'. Once we know a person's assignments and once we know that their preferences satisfy certain seemingly weak axioms (defined more fully below), we can predict their preferences in situations where risk is involved. That is, we do not have to ask them what their preferences are in these risky situations - this can be determined for them from our computation. Moreover, our computations do not require that they actually reason using the expected utility formula - indeed, they may be agents that do not engage in reasoning at all. As long as we know that their preferences conform to certain axioms, we know they will make their choices as if they were such a reasoner, and that is good enough for us to predict their choices, regardless of whether or not they actually use the EU reasoning process.

4 The axioms of EU

There have been two important types of axiomatization of expected utility theory: the one that was originally developed by von Neumann and Morgenstern (1944), which relies on objective probabilities, and a second kind developed by Savage (1954), using Frank Ramsey's idea (1931) that we can define probabilities subjectively from information about the agent's preferences (see Ramsey, F.P. §2). We will consider here only the following rough, non-technical statement of von Neumann and Morgenstern's axioms relying on the idea of objective probabilities.

(1) The *completeness* (or ordering) axiom specifies that all the objects over which we have preferences can be compared or ordered. To be precise: for any prizes *a* and *b*, $a > b$ or $b > a$ or they are indifferent.
(2) The *transitivity* axiom tells us that if you prefer, say, apples to oranges, and oranges to bananas, you will also prefer apples to bananas. To be precise: for any prizes *a*, *b* and *c*, if $a > b$ and $b > c$, then $a > c$.
(3) Suppose you can participate in a compound lottery (that is, a lottery that has another lottery as at least one of its prizes), where there is some chance that you will ultimately win a million dollars rather than nothing. Suppose you figure out the chance you will get the million. Call this probability *p*. The *reduction of compound lotteries* axiom says that you will be indifferent between this compound lottery and a simple lottery in which the chance that you will win the million rather than nothing is also *p*. This axiom is mathematically necessary in order for von Neumann and Morgenstern to prove their conclusion.
(4) Suppose there are three prizes *a*, *b* and *c*. The *continuity* axiom says that if *a* is preferred to *b*, and *b* is preferred to *c*, then there will be probabilities such that a person will be indifferent between *b* and a lottery with *a* and *c* as prizes. This indifference depends upon the probability of getting *a* being great enough to outweigh the risk of only getting *c*, but not so great that the person would prefer the lottery over *b*, nor so small that the person would prefer *b* over the lottery.
(5) The *independence* (or substitutability) axiom says, very roughly, that if a person is indifferent between two prizes, *a* and *b*, considered alone, they are also indifferent between two lotteries that are exactly the same, with the exception that in one of them prize *b* is substituted for prize *a*. (This axiom is roughly equivalent to what Savage calls the 'sure thing' principle in his axiomatization of EU theory. The only difference is that Savage's principle does not assume objective probabilities, allowing for the determination of a decision-maker's subjective probabilities from his utilities.)
(6) The *monotonicity* axiom says that, between any two lotteries involving only the most and least preferred alternatives, you should not prefer the lottery that renders the least preferred alternative more probable.

Many philosophers, economists and social scientists have wanted to revise von Neumann and Morgenstern's theory explicitly to make it into a theory of reasoning or a theory of behavioural rationale. These theorists have added psychological theories or normative interpretations of the axioms, so as to transform EU theory into what they regard as a plausible descriptive or normative theory. For example, Savage considers the theory a normative conception of both reasoning and rationale, whatever predictive value it might have. Other theorists (such as Kahneman and Tversky 1990) have explored the extent to which the theory might succeed as a descriptive account of human reasoning, generally reaching the conclusion that some of EU's axioms are not widely followed (especially the independence axiom, which assumes that what we do is not affected by whether, for example, we see a glass as half full or half empty), meaning that the theory, on their view, is best construed as a normative rather than a descriptive conception of reasoning (see Rationality of belief).

If this theory is construed either as a theory of reasoning or as a theory of rationale, is it an instrumental or a non-instrumental theory? As noted earlier, it can only be construed as instrumental if its supporters introduce a psychological theory that specifies that all ultimate or unmotivated ends of action are set by something other than reason. Yet the addition of such a theory may seem initially precluded by the structure of EU theory, which makes it easy to give it a non-instrumental interpretation. Consider that EU theory makes a subject's preferences the definers of the ends of their action. However, it would seem the theory does not permit just any preference to be end-defining, but rather, only those preferences that satisfy the axioms of expected utility theory explicated above. But this appears to mean that the axioms play a role in fixing the ends of action by acting as a normative 'sieve' through which our preferences must pass in order to be considered 'end-defining'. Therefore, if these principles are part of what we mean by practical reason, does this not mean that reason is playing a role in setting our ends of action? If so, EU theory, in so far as it is defined in terms of these axioms, cannot be considered an implementation of the instrumental theory of reasoning. A number of theorists who regard EU theory as a normative account of human reasoning characterize the theory in ways that strongly suggest this non-instrumental interpretation - this even includes some who elsewhere treat it as an instrumental theory of reasoning.

Some philosophers, loyal to the instrumental characterization of practical reason, have argued that EU theory's reliance on these axioms is not really enough to make it a non-instrumental theory. For example, R. Nozick (1993) argues that admitting that these axioms play a role in defining our ends is only introducing consistency standards into our preference set, which is just 'one tiny step beyond Hume'. Yet by taking this 'tiny step' we invite the possibility that there are normative standards other than those of consistency (for example, moral standards) to which our desires ought to be subject (for on what grounds can we argue for some non-instrumental standards and against others?).

How might we argue that the EU axioms can be defended within a pure instrumental theory? One obvious way to try to do so is to claim that EU axioms can be instrumentally justified. On this view, the EU axioms prescribe consistency, and agents must have consistent preferences if they are to achieve the maximal satisfaction of whatever preferences they have. But even if the principle 'To satisfy ends, one must be consistent' is true, the EU axioms are not the only representation of consistency, so it seems one could satisfy the principle by using other reasoning theories with rather different axioms. Furthermore, an instrumentalist argument must avoid circularity: to defend the axioms as instrumentally justified, the argument cannot appeal to the desirable consequences of following the axioms, if that very appeal presupposes that these consequences are desirable as the objects of preferences that satisfy the axioms to be defended. Alternatively, the axioms of EU, or some other set of axioms, might be non-instrumentally justified as intuitively compelling or as cohering well with other intuitive principles of reason.

5 Challenges to expected utility theory

Many theorists have maintained, on the basis of empirical evidence, that human beings consistently violate some of the EU axioms (see, for example, Kahneman and Tversky 1990; Allais 1979; Ellsberg 1951; Rationality of belief). That evidence appears to show that, at the very least, EU theory fails as a *descriptive* account of our practical reasoning. And the violations of some of the EU axioms seem so intuitively reasonable that many theorists believe they show EU theory fails as a *normative* account of our reasoning. While the examples are too complicated to discuss here, it is important to note that there is a pattern to many of the purported violations. Peter Hammond (1988) and others have argued that the axioms of EU theory articulate the idea that we choose actions

(solely) by assessing the consequences of those actions, but violations of these axioms occur when people's choice of action results from certain non-consequential preferences they have - for example, preferences for the situation, action or process from which the consequences will result. Economists have called these *state-dependent* preferences. Does the existence of these preferences show not only that EU theory is the wrong portrayal of practical reason but also, more fundamentally, that any successful descriptive or normative theory of practical reason will have to be non-consequentialist, relying on axioms of choice that recognize that sometimes we choose to act for reasons other than the desirability of the consequences of these actions?

Some advocates of the idea that we are consequentialist reasoners answer this last question negatively. They have proposed that purported counterexamples to EU theory can be accommodated once we reinterpret them to show that hidden in the examples are possible consequences to which people have positive or negative attitudes. If this strategy works, it would support the idea that, if we understand the consequences in any choice situation in an appropriately broad way, no human being whom we would consider rational will violate the (consequentialist) EU axioms in their choices. One worry that even the defenders of this strategy have voiced is that it might work too well - allowing us to redeem any violation of the axioms by finding some hidden consequence to which the purported violator is actually responding. Some theorists (particularly John Broome 1991) try to avoid the problem by arguing for principles that specify when a rational agent must be indifferent between two outcomes. Such principles prohibit us from postulating hidden consequentialist attitudes that, if recognized, would mean the agent was not violating the axiom after all.

Those theorists who believe that the purported counterexamples cannot be explained away have advocated rejecting one or more of EU theory's axioms, so as to develop a revised theory that attempts to be a better descriptive or normative theory of reasoning or behavioural rationale. For example, Mark Machina (1991) rejects the independence axiom, and Graham Loomes and Robert Sugden (1982) drop the transitivity axiom. These revised theories are not, however, radical departures from EU theory, and retain much of the structure of that theory.

6 Contemporary non-maximizing alternatives to expected utility theory

One central assumption of EU theory is that rational action is maximizing action: faced with a multiplicity of ends, we are supposed to try to maximize their satisfaction. But some critics of the theory insist that rational action with respect to a multiplicity of ends often permits, and sometimes even requires, that we refrain from maximizing.

For example, Herbert Simon (1982) has argued that we should understand reason in a non-maximizing way because human beings operate subject to limitations and constraints that make effective maximization impossible. These limitations and constraints come in two varieties: constraints on information-gathering relevant to action (for example, limited access, high costs of obtaining, and limited time), and constraints on computational ability (for example, limited memory, intellectual disabilities, high costs of difficult calculations, and slowness of calculation relative to time allowed). Because of these constraints, Simon argues that human beings do not (and should not) reason by straightforwardly maximizing. Instead he says they do (and should) use certain shorthand reasoning procedures that yield good results given these constraints and limitations. Simon calls reasoning using these shorthand procedures 'satisficing'.

However, Simon's satisficing can be understood as consistent with a maximizing conception of reasoning if it is defended as a form of reasoning used by agents interested in maximizing preference satisfaction while subject to the above constraints. On this view, agents maximize to the best of their ability if they use these shorthand satisficing procedures. This position therefore represents a maximizing form of reasoning as a 'meta-rational' procedure that assesses possible operational reasoning procedures for those agents who are subject to certain constraints that make maximization too costly or too difficult to operate upon directly.

A more radical anti-maximizing theory that cannot be reconciled with an overall maximizing theory of reason is proposed by Isaac Levi (1986), who argues that our decision making must sometimes contend with unresolved conflict. On Levi's view, there are times when different options available to us are good in quite different ways (reflecting different moral, aesthetic, prudential, cognitive or political values), such that we cannot conclude that one of them is, all things considered, better than the other(s). Levi argues that when such conflict between values cannot be resolved, a rational agent cannot act so as to maximize one unified conception of the good (since there is

no such thing in these circumstances), but must instead act in a way that, given their evaluation of the options, is at least admissible. Levi's non-maximizing theory is none the less still a teleological or consequentialist portrayal of practical reasoning, in so far as rational action is construed as that which serves a goal.

An even more radical anti-maximizing portrayal of practical reason has been proposed by Michael Slote (1989). Slote holds that self-regarding rationality often permits or requires maximization with respect to our own good, but he also insists that we are sometimes rationally permitted and even have positive reason, on balance, to abstain from performing actions that we know will maximize our own good. On his view, if we are moderate in our desires, we can, for example, reasonably turn down a second dessert or a luxurious house - not because we think we would not enjoy or be better off with them - but because we consider our situation to be good enough and fine just as it is. Since it says that there can sometimes be reason on balance to prefer an action whose results are *less* good for one, Slote's conception of rationality is, strictly speaking, neither consequentialist nor teleological.

Jon Elster (1979) thinks about the issue of maximization in another way. Elster distinguishes between (what he calls) a *global* maximizer conception of practical reason and a *local* maximizer conception of practical reason. Local maximizers do whatever they judge best at the time of choice, but this means they have a 'myopic eye fixed to the ground' and cannot take into account 'what happens behind the next hill'. In contrast, a global maximizer can choose the worse option now, in order to be in a position to realize a far better outcome than the presently better option would yield. Unlike theorists who think of us as reflecting in some piecemeal fashion about how to satisfy effectively a set of desires as they present themselves to us at any given time, Elster's discussion points out that human beings need, and use, a conception of means/end reasoning that stretches out over time, permitting the rationality of enduring short-term losses in order to realize long-term gains.

The importance of reasoning long-term is, in the view of many theorists, connected to the idea of prudence. It is commonly (but not universally) thought that the practically rational person can not only choose effective means to ends, but also resolve on a course of action that will advance their interests in the long run, not merely the short run. Much of the literature in this area is inspired by the work of Thomas Nagel (1970). Some of it (see, for example, Derek Parfit 1984) explores the extent to which our conception of our own identity as persons is connected to reasoning with respect to the future (see Morality and identity).

How the idea of global rationality should be developed is controversial; some, such as David Gauthier (1986) and Edward McClennen (1990), have proposed that fully rational agents should sometimes remain 'resolute' in their intention to perform an action in order to achieve some desirable end, even in situations where conventional maximizing forms of reasoning would endorse another tempting alternative. Such resolution is rational on their view, when it would make possible benefits that can only be realized if the agent maintains the intention to perform this action. Gregory Kavka's 'toxin puzzle' (1983) gives us a way to illustrate their view: imagine that there is a millionaire who will give you a million dollars if you form the intention to drink a toxin that will not kill you but will make you ill, where this millionaire will give you the money before you have to drink the toxin, as long as they are satisfied that you have formed the intention to do so. Resolute theorists argue that, to get the money, you should form the intention to drink the toxin, and then drink it. However, other theorists dispute the rationality of resoluteness, and argue that a rational agent in the toxin example would conclude that it would be irrational to drink the toxin, and thus irrational to form the intention to do so (even though this means losing a million dollars). This view presupposes a theory of rational intention that makes resoluteness irrational. Theorists on both sides of this issue believe that intentions and actions must be linked, on pain of irrationality. However, resolute theorists believe that as long as it is rational to intend to perform an action, then it is rational to perform it, whereas critics of the idea of resoluteness maintain that one can only rationally intend to perform an action when it is actually rational to perform it.

7 Contemporary non-maximizing alternatives to expected utility theory (cont.)

Elster and others have also argued that conceptions of reasoning must go beyond maximization in order to reckon with the phenomenon of 'endogenous preference formation'. This term refers to the process by which agents modify or revise what they prefer in response to constraints their environment places upon them. For example, in Aesop's fable, the fox, who cannot reach the grapes he desires, finally decides that they must be sour anyway, so that he does not want them. Thus the fox changes his preferences when he discovers that he cannot satisfy them as they are. Endogenous preference formation may not only be a fact, but a normatively desirable response to our

environment. (Consider the advice given by Crosby, Stills and Nash in a popular song: ‘If you can’t be with the one you love, love the one you’re with’.) Note that this phenomenon would seem to be neither explained nor endorsed by EU theory, since it involves a (non-instrumental) form of reasoning explicitly concerned with establishing our ends of action, in ways that have nothing to do with the axioms of consistency that define EU theory.

Theorists have also proposed non-maximizing conceptions of reasoning for situations of complete uncertainty, that is, situations in which we cannot assign probabilities to the possible outcomes. In such cases, we cannot reason as expected utility maximizers, since doing so precisely requires that we know, or be able to estimate, the probabilities of possible outcomes of alternative actions. Some theorists (such as John Harsanyi 1976) have proposed that we should use the ‘principle of insufficient reason’ to derive probabilities that enable us to use an expected utility calculation. This principle tells us to estimate as equi-probable the possible outcomes of alternative actions most relevant to our decision making, and then use these probabilities in an expected utility calculation. However, other theorists (such as John Rawls 1971) have proposed the rationality of a ‘maximin rule’ of choice in at least some situations of complete uncertainty. This rule tells us to choose that action whose worst possible outcome is better than, or at least as good as, the worst possible outcomes of all the other alternative actions. This is generally considered a very conservative principle of choice under uncertainty.

Many theorists (for example, Bratman 1987) believe progress in understanding human reasoning should be made not by concentrating on maximization, but by thinking about the way our deliberation standardly involves planning. Constructing a plan involves a complicated and interconnected series of intentions about how to act. Plans would seem to be necessary in order to achieve a wide variety of ends, and in order to ensure that the achievement of one end does not foreclose the possibility of securing ends that are as important as, or more important than, the one we are presently concerned to achieve. This suggests that part of what it means to say that a person is practically rational is that they are a good planner, so that understanding practical reasoning involves specifying what ‘good planning’ is. This requires specifying, for example, what a ‘feasible’ plan is, when sub-plans are necessary, how extensive a good plan should be, and so forth. Those who are interested in constructing artificially intelligent beings are trying to specify the nature of rational planning in a way sufficiently detailed to allow the construction of a rational planner. Such a project, involving both philosophers and researchers in artificial intelligence, attempts to produce a form of practical reasoning that we would expect any agent to perform such that we would regard their operation in the world as ‘successful’ from a rational point of view (see Pollock 1995; Artificial intelligence).

8 Game theory

Other technical representations of practical reason exist within a subject called ‘game theory’, also invented by John von Neumann and Oscar Morgenstern (see Decision and game theory). To understand game theory, consider another distinction made by Jon Elster (1979), between ‘parametric’ and ‘strategic’ forms of reasoning. Consider the parametric reasoner Harry, who treats the environment in which he will act as a ‘given’, estimating how his actions will affect it. In contrast, the strategic reasoner Harriet regards her environment as containing agents whose choice of behaviour will depend in part on what they believe she will do, even while her choice of behaviour depends in part on what she believes they will do. Game theory concerns strategic reasoning because it depicts situations in which the ultimate payoffs to each player from possible actions available to them depend upon what the other player(s) will do. These situations can involve pure conflict - in which each player will only gain if the other(s) lose(s); or pure cooperation, in which each player will only gain if the other(s) gains; or games of mixed conflict and cooperation. In general, what counts as a rational solution in game theory has been the subject of considerable controversy.

A particularly famous game is the ‘prisoner’s dilemma’ or ‘PD game’. The name for this game comes from an early example of it in which the two players are prisoners who have jointly participated in a crime, and who are separately encouraged by the police to confess, thereby informing on their partner. If neither confesses they can only be convicted of a crime that yields 1 year in jail; if both confess they will both get 5 years in jail; if one confesses and the other does not, the confessor gets no years in jail but the person who fails to confess will get 10. In a game of this sort, each player prefers most the situation where he confesses but his partner does not, he prefers second the situation where neither he nor his partner confesses, he prefers third the situation where he and his

partner both confess, and he prefers least the situation where he does not confess but his partner does. While this fanciful example may make the game seem arcane, social scientists have argued that the PD game is very common in a variety of social situations, especially in a free market. For example, consider two companies, who each have the choice between colluding on prices (the equivalent of not confessing in the original example) or competing rather than colluding (the equivalent of confessing). The payoffs to each company make this a prisoner's dilemma. The seeming ubiquity of this game has persuaded many social scientists of the importance of studying it (see Rational choice theory §§2-3).

Many theorists have proposed that rational players in a prisoner's dilemma should reason using the idea of a 'dominant strategy'. According to this form of reasoning, the only rational choice is the non-cooperative action because no matter what the other player does, each player is better off not cooperating (if the other player cooperates, he gets his favourite outcome if he fails to cooperate; if the other player fails to cooperate, he avoids his worst outcome if he fails to cooperate). In this sense, not cooperating 'dominates' over cooperating. However, whether this is the right strategy for dealing with the prisoner's dilemma has been debated, and some theorists have proposed other ways of solving it. These include using expected utility reasoning and impartial (moral) reasoning - both of which may sometimes recommend that the parties in a PD game choose the cooperative option.

The PD game has also been compared to a puzzle called 'Newcomb's paradox'. Imagine that you are given the choice between taking the contents of an opaque box, or taking the contents of that box plus the contents of another, transparent box which you can see contains $101,000. The opaque box contains either a million dollars or nothing. Which of these it contains depends upon whether a powerful being (for example, some kind of god) capable of predicting behaviour with virtually 100 per cent reliability has predicted how you will behave. If the predictor has previously determined that you will choose only the opaque box, it has put a million dollars in it. If the predictor has previously determined you will choose both boxes, it has put nothing in the opaque box. How should you rationally choose? Some theorists (for example, David Lewis 1985) maintain that the prisoner's dilemma is an instance of Newcomb's paradox. More generally, those who discuss the paradox recommend different strategies for its resolution, including dominance reasoning and expected utility reasoning, that not only generate opposing advice, but also reflect different conceptions of the structure and aim of practical reasoning.

Another game much discussed in recent years is related to the prisoner's dilemma, and the resolution proposed for it seems interestingly suggestive of a kind of moral thinking. Imagine a situation in which a prisoner's dilemma between two or more players repeatedly recurs over time, and where the last recurrence is unknown. This is called an *iterated* prisoner's dilemma. Whereas, in a single-play prisoner's dilemma, the dominant strategy is not to cooperate, in an iterated prisoner's dilemma, cooperation is rational as long as each player has reasonable assurance that the other player will also cooperate (and will not lose too much by acting on the expectation that the other cooperates and is wrong). Some theorists have argued on the basis of computer models that agents who follow this 'tit for tat' strategy (cooperating with cooperative partners, not cooperating with non-cooperative partners) do better, overall, than players who follow other, less cooperative strategies. This strategy strikes many as having moral overtones (cooperation is rewarded, non-cooperation is punished), and some have speculated that moral responses and emotions, and moral ways of thinking, may well be explained by the way in which certain cooperative or non-cooperative forms of behaviour can be beneficial or harmful to human beings in common game-theoretic situations. Such reflections on the nature of strategic reasoning encourage the hope that future theorizing in game theory may serve to link rational and moral forms of reasoning. This link is also suggested by the work of theorists (such as J.F. Nash 1951; Gauthier 1986) who have explored the extent to which the rational resolution of some of these games might be understood to mimic the results of an ideal bargaining process.

9 Practical reasoning and standards

Most of the preceding theories of reason, including both the maximizing and the non-maximizing conceptions, assume that reason involves weighing the values of various considerations (for example, outcomes, actions) pertinent to our making a choice. This idea is naturally suggested by a teleological or consequentialist conception of reason. However, there are theories that represent rational action as involving the idea of satisfying principles or standards. On this view, a rational agent is one who, at least on some occasions, chooses how to act by applying the appropriate principle to their situation. Such principle-based reasoning can be assimilated to a consequentialist conception of practical reason if the justification for applying these principles is that doing so will help an agent to

achieve or maximize the good. Non-consequentialist theorists who reject this assimilation believe that at least some of these principles of right action cannot be justified by reference to any goal or conception of the good. On this view, in addition to whatever concerns an agent might have about how much good or bad an action might produce, they must none the less deliberate or act so as to satisfy these principles in order to be considered rational.

Some non-consequentialists believe that, to be rational, an action or policy must satisfy certain moral or political principles that prescribe ‘right conduct’ (some of which we will discuss in §10 below). Moreover, although some feminists (such as Lloyd 1984) have questioned whether the whole concept of practical reason reflects the assumptions and biases of our patriarchal society, Ruddick has argued that, in order to bring up their children, mothers must engage in reasoning that determines how to act so as to satisfy different and often conflicting principles (see, for example, Ruddick 1989). Apart from presenting a principle-based view of reasoning, her argument suggests there may be distinctive forms of reasoning appropriate for different practical activities (see Feminism; Feminist ethics).

Finally, Onora O’Neill (1989) has suggested that even though the object of instrumental reasoning is a goal, *how* that reasoning ought to proceed is defined by a variety of ‘principles of rational intending’, none of which is captured by EU’s axioms. Her account clearly implies the possibility of a non-consequentialist theory of instrumental reason.

10 Moral reasoning and weakness of will

A variety of moral theorists insist that practical reason has a moral aspect. Some theorists believe moral reasoning can be accommodated within the confines of expected utility theory since, on their view, moral reasoning is simply EU reasoning with other-regarding preferences. However, others insist that moral reasoning is not merely about how we satisfy our actual other-regarding preferences, but also (and more importantly) about defining how our preference set *ought* to take others’ interests into account.

For example, Kant (1785) believed that reason structurally involves a moral reasoning procedure that indirectly defines our ends of action. Kant calls this reasoning procedure the ‘moral law’ and one of his formulations of it is: ‘Act only on that maxim through which you can at the same time will that it will be a universal law’. When I follow this procedure, I universalize my maxim of action, converting it from the form ‘I will do *x* in circumstances *c* in order to achieve *y*’ to the form ‘Everyone does *x* in circumstances *c* in order to achieve *y*’. Then I evaluate that universalized maxim by determining, first, whether I will be able to achieve *y* by doing *x* in circumstances *c* in a world where everyone does *x* in *c* to get *y*; and second, whether I can ‘will’ the world in which everyone does *x* in *c* to get *y*. If a maxim meets this test, then I am permitted to act on it - otherwise, I have an obligation not to do so. This test procedure generates what Kant calls ‘categorical imperatives’, that is, imperatives that (in contrast to hypothetical imperatives) direct us to engage in certain forms of action, or direct us to refrain from certain forms of action, regardless of the desires or interests we happen to have (see Kantian ethics).

Kant’s account of moral reasoning is non-instrumental. The moral law acts as a kind of moral ‘filtering device’ through which maxims generated by our desires must pass, and his account is also non-consequentialist, because it presents reason not as that which weighs competing goods, but as that which answers to a certain principle of right conduct.

Another representation of moral reasoning is put forward by contractarian theorists. Contractarian reasoning procedures ask us to deliberate about a course of action, plan or policy by determining whether those affected by it (understood so that they are suitably impartial and equal in power) could agree to it or, alternatively, whether it would be something none of them could reasonably reject. This type of reasoning procedure is explicitly presented by John Rawls (1971) as a non-teleological or non-consequentialist form of reasoning, and the way in which it could be used to define ends of action also makes it a non-instrumental conception of reason. Other philosophers, such as David Gauthier (1986) and Thomas Scanlon (1982), have also insisted that contractarian forms of reasoning can define and illuminate the content, and perhaps even the motivational source, of moral behaviour (see Contractarianism).

Many contemporary moral philosophers called ‘utilitarians’, following Bentham, have endorsed various forms of a moral reasoning procedure labelled utilitarian, in which actions and policies are evaluated by determining whether

they maximize the welfare (understood by different theorists in different ways, for example, as happiness or preference-satisfaction) of members of a community (see Utilitarianism). Some utilitarians (for example, Harsanyi 1976) have linked a form of a utilitarian reasoning procedure (for example, the kind called ‘average utilitarianism’) to expected utility theory. Utilitarian forms of reasoning are classic examples of consequentialist theories of reasoning, in so far as actions or policies are determined by attempting to maximize a certain conception of the good, construed as human welfare (see Consequentialism; Welfare).

Particularly when an agent’s preferences are formed by social or political processes that seem unjust (for example, practices such as slavery, or institutional discrimination), some theorists have argued that the agent’s preferences may have a content that we can evaluate as violating reason. Such an evaluation requires a non-instrumental moral reasoning procedure that determines not only the content of the preferences a rational agent ought to have, but also the ways in which such preferences are rightly formed.

Even if practical reason is the source of normatively authoritative directives (or imperatives) governing our conduct, philosophers disagree about whether these directives can (solely by virtue of that authority) motivate us to act (see Moral motivation). Hume argues that reason is motivationally inert, so that only our passions can motivate us. Reason, according to Hume, is a mere slave to our passions that never itself defines or precipitates action directly.

In contrast, Kant insists that his ‘moral law’ can move us to action, and this belief in the motivational efficacy of practical reason arises in part from his concern to show that, no matter what our interests, desires or emotions, each of us as a rational being always has the capacity not only to understand what the moral course of action is, but also to perform that action, by virtue of the fact that our reason is always in a position to comprehend and initiate appropriate moral conduct.

However, there is a related issue about how to describe cases where passion or desire seemingly leads us to act against our ‘better judgment’. There has long been controversy about how to conceive such irrational ‘weakness of will’ and even about whether the idea of such weakness is really logically coherent (see Akrasia). But it is not surprising that philosophers and others who disagree about the foundations and content of practical reason should in this way (and others) also disagree about the nature and forms of human irrationality.

See also: Economics and ethics; Economics, philosophy of; Imperative logic; Rationality and cultural relativism; Rationality of belief

JEAN HAMPTON

References and further reading

Allais, M. (1979) ‘The So-Called Allais Paradox and Rational Decision Under Uncertainty’, in M. Allais and O. Hagen (eds) *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht: Reidel, 437-681.(Presents a counterexample to EU theory’s independence axiom; discussed in §5.)

Bentham, J. (1823) *Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, London: Methuen, 1982.(Seminal work introducing utilitarian reasoning, discussed in §10; and the idea of utility as a measure of the pleasure or pain of an object or event, discussed in §3.)

Bratman, M. (1987) *Intentions, Plans and Practical Reason*, Cambridge, MA: Harvard University Press. (Expounds a conception of intention and links it to planning and practical reason; discussed in §7.)

Broome, J. (1991) *Weighing Goods*, Oxford: Blackwell.(Exploration of EU theory as a consequentialist conception of reasoning; concerned to defend EU theory from counterexamples challenging its adequacy as a normative theory, as discussed in §5.)

Campbell, R. and Sowden, L. (1985) *Paradoxes of Rationality and Cooperation: Prisoner’s Dilemma and Newcomb’s Problem*, Vancouver, BC: University of British Columbia Press.(Collection of articles, plus a useful summary by the editors, of literature on the prisoner’s dilemma and Newcomb’s paradox, and thus expands on the discussion in §8.)

Dworkin, R. (1977) *Taking Rights Seriously*, Cambridge, MA: Harvard University Press.(Classic work in philosophy of law; includes extensive exploration of legal reasoning.)

Ellsberg, D. (1951) ‘Risk, Ambiguity and the Savage Axioms’, *Quarterly Journal of Economics* 75: 653-6; repr. in Gardenfors and Sahlin (1988).(Presents counterexamples to the Savage axiomatization of EU theory; mentioned in §5.)

Elster, J. (1979) *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge: Cambridge University Press.(A work that distinguishes between global and local reasoning, discussed in §§6-7; and between parametric and strategic reasoning, discussed in §8.)

Elster, J. (1983) *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge: Cambridge University Press. (Includes an exploration of endogenous preference formation, discussed in §7.)

Frank, R. (1988) *Passions within Reason: The Strategic Role of the Emotions*, New York: Norton.(Expounds on the idea that our passions are allies in our attempts to solve certain game-theoretic problems, and may have evolved within us because of this function.)

Gardenfors, P. and Sahlin, N. (1988) *Decision, Probability and Utility: Selected Readings*, Cambridge: Cambridge University Press.(A collection of classic essays in rational choice theory, including works by Allais, Ellsberg, Savage, Ramsey and later theorists. An excellent resource.)

Gauthier, D. (1986) *Morals by Agreement*, Oxford: Oxford University Press.(Presents a kind of contractarian moral theory; also a conception of rationality that interprets EU theory as an instrumental theory as discussed in §3; and develops a bargaining-like solution to certain game-theoretic problems.)

Gibbard, A. and Harper, W.L. (1985) 'Counterfactuals and Two Kinds of Expected Utility', in R. Campbell and L. Sowden (eds) *Paradoxes of Rationality and Cooperation*, Vancouver, BC: University of British Columbia Press; repr. in Gardenfors and Sahlin (1988).(In response to Newcomb's paradox, discussed in §8, these authors develop two conceptions of decision-theoretic reasoning, 'evidential' and 'causal'; quite technical and ill-suited for beginners.)

Goldman, A. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(Explores epistemic norms, and suggests that norms of theoretical reason may be related to or derived from norms of practical reason; discussed in §1.)

Hammond, P. (1988) 'Consequentialist Foundations for Expected Utility Theory', *Theory and Decision* 25: 25-78.(Expounds the way in which EU theory expresses a consequentialist conception of reason, discussed in §5. Highly technical and ill-suited for beginners.)

Hampton, J. (1994) 'The Failure of Expected Utility Theory as a Theory of Reason', *Economics and Philosophy* 10: 195-242.(Expands on the discussion in §§3-5 of the extent to which EU theory is an instrumental conception of reason, and argues that EU theory fails as a descriptive and normative conception of reasoning, and as a theory of behavioural rationale, because it is too consequentialist.)

Harsanyi, J. (1976) *Essays on Ethics, Social Behavior and Scientific Explanation*, Dordrecht: Reidel.(A collection of essays that, among other things, expresses the author's conception of utilitarianism, influenced by EU theory, mentioned in §10, and defends the use of the principle of insufficient reason in situations of complete uncertainty, discussed in §7.)

Hart, H.L.A. (1961) *The Concept of Law*, Oxford: Clarendon Press.(A classic work in philosophy of law; considers questions about the nature of moral reasoning.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised P.H. Nidditch, Oxford: Clarendon Press, 1975.(A major work in the history of philosophy, including the thesis that reason is motivationally inert (discussed in §10) and suggesting that reasoning prior to action may be nothing but theoretical reasoning (§1); also influential in the development of the instrumental theory of practical reason (discussed in §§2-3).)

Kahneman, D. and Tversky, A. (1990) 'Rational Choice and the Framing of Decisions', in K. Cook and M. Levi (eds) *The Limits of Rationality*, Chicago, IL: Chicago University Press, 60-89.(An important article that uses empirical studies to show that people frequently violate some of the EU axioms. The collection in which it appears includes a number of classic works in decision theory.)

Kant, I. (1785) *The Groundwork of the Metaphysics of Morals*, trans. H.J. Paton, London: Harper Torchbooks, 1964.(A major work in the history of philosophy presenting a theory of moral reasoning discussed in §10. It is also the source of the distinction between categorical imperatives, discussed in §10, and hypothetical imperatives, discussed in §1.)

Kant, I. (1793) 'On the Common Saying, "This May Be True in Theory, But It Does Not Apply In Practice"', in *Kant's Political Writings*, ed. H. Reiss, Cambridge: Cambridge University Press, 1970.(An essay that proposes a contractarian reasoning procedure for political decision making; see §10.)

Kavka, G. (1983) 'The Toxin Puzzle', *Analysis* 43: 33-6.(Sets out a puzzle that some theorists have used to expound different conceptions of global reasoning; discussed in §6.)

Levi, I. (1986) *Hard Choices: Decision-Making under Unresolved Conflict*, New York: Cambridge University Press.(Proposes a non-maximizing but consequentialist account of practical reason; discussed in §6.)

Lewis, D. (1985) 'Prisoner's Dilemma is a Newcomb Problem', in R. Campbell and L. Sowden (eds) *Paradoxes of*

Rationality and Cooperation: Prisoner's Dilemma and Newcomb's Problem, Vancouver, BC: University of British Columbia Press.(Relates the prisoner's dilemma to Newcomb's problem, both discussed in §8.)

Lloyd, E. (1984) *The Man of Reason*, Minneapolis, MN: University of Minnesota Press.(Explores the extent to which conceptions of reason are generated by western patriarchal society in which men have dominated the academy; discussed in §9.)

Loomes, G. and Sugden, R. (1982) 'Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty', *Economic Journal* 92: 805-24.(Uses empirical evidence to argue that people violate EU theory's transitivity axiom; discussed in §5.)

Luce, R.D. and Raiffa, H. (1957) *Games and Decisions: Introduction and Critical Survey*, New York: Dover, 1985.(Classic statement of EU theory and a superb reference work for those wishing to study EU theory and game theory. Expands on the discussions of these fields, and includes a technical statement of the EU axioms.)

Machina, M. (1991) 'Dynamic Consistency and Non-Expected Utility Models of Choice under Uncertainty', in M. Bacharach and S. Hurley (eds) *Foundations of Decision Theory: Issues and Advances*, Oxford: Blackwell. (Uses empirical evidence to challenge EU's independence axiom; discussed in §5.)

McClennen, E. (1990) *Rationality and Dynamic Choice*, Cambridge: Cambridge University Press.(Presents a conception of rationality as resoluteness, discussed in §6; also challenges the idea that some of the EU axioms can be defended instrumentally - see §5.)

Nagel, T. (1970) *The Possibility of Altruism*, Princeton, NJ: Princeton University Press.(Important work that explores and develops conceptions of prudential and moral reasoning; mentioned in §6.)

Nash, J.F. (1951) 'Non-Cooperative Games', *Annals of Mathematics* 54: 286-95.(Develops a bargaining-like solution to some games.)

Neumann, J. von and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press, 1953.(Classic and seminal work creating both EU theory and game theory. Most parts of the book are highly technical but the first chapter is fairly accessible to the non-specialist.)

Nozick, R. (1993) *The Nature of Rationality*, Princeton, NJ: Princeton University Press.(Wide-ranging discussion of issues in the study of rationality; highly readable. Expands upon the discussion in §4 of the extent to which the EU axioms can be interpreted consistently with the idea that EU reasoning is instrumental.)

O'Neill, O. (1989) *Constructions of Reason*, Cambridge: Cambridge University Press.(An exploration of Kantian moral theory; proposes a principle-based conception of instrumental reasoning discussed in §9.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press.(A wide-ranging discussion of practical reason that explores instrumental and prudential reasoning, and personal identity; mentioned in §6.)

Pollock, J. (1995) *Cognitive Carpentry*, Cambridge, MA: MIT Press.(A presentation of practical reasoning as planning, discussed in §7, for the purpose of developing an artificial person.)

Ramsey, F. (1931) 'Truth and Probability', in D.H. Mellor (ed.) *Foundations of Mathematics*, London: Routledge & Kegan Paul, 1978; repr. in Gardenfors and Sahlin (1988).(Sets out a way of deriving subjective probabilities from an agent's utility, relevant to Savage's version of EU theory.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(This classic work in modern political theory develops an influential version of a contractarian reasoning procedure (discussed in §10) and includes an endorsement of maximin reasoning in some situations of complete uncertainty, as discussed in §7.)

Ruddick, S. (1989) *Maternal Thinking: Toward A Politics of Peace*, Boston, MA: Beacon Press.(Proposes a standard-based conception of the reasoning that mothers engage in when they bring up their children; discussed in §9.)

Savage, L. (1954) *The Foundations of Statistics*, New York: Wiley & Sons; 2nd edn, New York: Dover, 1972.(A seminal work in the development of EU theory, with an axiomatization that is an alternative to that of von Neumann and Morgenstern. Not for beginners.)

Scanlon, T. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press.(Develops a kind of contractarian moral reasoning procedure; discussed in §10.)

Simon, H. (1982) *Models of Bounded Rationality*, Cambridge, MA: MIT Press.(Expounds the idea that the reasoning upon which we operate is not concerned with maximizing but with 'satisficing'; discussed in §6.)

Slote, M. (1989) *Beyond Optimizing: A Study of Rational Choice*, Cambridge, MA: Harvard University Press. (Proposes a non-maximizing and non-consequentialist conception of practical reason; discussed in §6.)

Williams, B. (1981) 'Internal and External Reasons', in *Moral Luck*, Cambridge: Cambridge University Press, 101-13.(Influential article that develops a conception of practical reason indebted to Hume.)

Ravaisson-Mollien, Jean-Gaspard Félix Lacher (1813-1900)

Félix Ravaisson was a French philosopher, born in Namur. Apart from his three main works - Rapport sur la philosophie en France au dix-neuvième siècle (Report on Philosophy in France in the Nineteenth Century) (1863), Métaphysique d'Aristote (Metaphysics of Aristotle) (1837, 1846) and De l'habitude (On Habit) (1838) - his ideas can mainly be found in articles, fragments and short passages, such as those collected in Testament philosophique (Philosophical Testament) (1901) or others that are still unpublished. None the less, Henri-Louis Bergson, who succeeded him at the Academy of Moral and Political Sciences, according to custom gave a synthetic account of Ravaisson's work, in his Notice sur la vie et les oeuvres de M. Félix Ravaisson-Mollien (Account of the Life and Works of Mr Félix Ravaisson-Mollien) (1934). Bergson insisted on the importance of artistic research in Ravaisson's thought, and one can indeed explain his philosophy as beginning from a meditation on works of art and on beauty. The importance of this starting point can be seen in the privileged role he gives to synthesis in all explanation, for in an artistic masterpiece, it is the whole that allows for comprehension of the parts. Ravaisson gave this idea a metaphysical dimension.

In 1832 Ravaisson was awarded a prize by the Academy of Moral and Political Sciences in a competition on the study of Aristotle's metaphysics. The first volume of his *Essai sur la métaphysique d'Aristote (On Aristotle's Metaphysics)* appeared in 1837, the second in 1846. In 1838 he defended his doctoral thesis, *De l'habitude (On Habit)*. However, Ravaisson never became a professor of philosophy. In 1839 he was made Inspector of Libraries, and fifteen years later he became Inspector General of Higher Education. In 1852, the Minister of Public Education considered the issue of the teaching of drawing in schools. Ravaisson was chairman on the commission set up to report to the minister, while Delacroix and Ingres were among its members. In 1863, at the request of the government, Ravaisson wrote a *Rapport sur la philosophie en France au dix-neuvième siècle (Report on Philosophy in France in the Nineteenth Century)*, which exercised an important influence on academic philosophy. In 1893 he put forward an overview of his philosophy in an article titled 'Métaphysique et morale' (Metaphysics and Morals), which appeared as the introduction to the philosophical journal of that name. A complete exposition of his doctrine appeared in the unpublished fragments put together after his death under the title *Testament philosophique (Philosophical Testament)* (1901).

Ravaisson's most original work is *De l'habitude*, in which, according to Bergson, he broadly sets out a philosophy of nature. He refused to limit habit to its negative aspect, claiming, indeed, that it allows the mind to liberate itself more easily from nature. Nature, according to Ravaisson, is essentially spontaneity. Just as habit is voluntarily acquired, and spirit is thereby made nature, so freedom makes itself into necessity.

Ravaisson also developed a personal conception of the teaching of drawing. In his research on the Venus de Milo, he emphasized the spiritual significance of a pose that expresses love and peace. Finding the paradigm of all the stages of thought in aesthetic intuition, Ravaisson believed that the method goes back to Aristotle in his attempt to make teleological explanation the basis for all true understanding. Indeed, in his studies on the metaphysics of Aristotle, Ravaisson insisted on the fact that Aristotle saw philosophy as a bringing to completion of reality. According to Ravaisson, the essential claim in Aristotle's philosophy is that attention must be focused on the individual, and not dispersed in the general. Philosophy must not stay in the abstract, which is why Aristotle finds the highest truth in the contemplation of works of art.

Rapport sur la philosophie en France au dix-neuvième siècle gives the most complete account of Ravaisson's thought. Having distinguished the mechanical and organic modes of apprehension, he returns to the origin of things, and claims that God created the universe by a process of 'condescension' and love in a way that resembles the creation of art.

Ravaisson begins the *Rapport* by examining the ancient philosophers, and shows that Aristotle is the founder of metaphysics. After a rapid overview of pre-nineteenth-century thinkers, he summarizes the eclecticism of Victor Cousin and insists on his debt to Schelling. His examination of Comte suggests that Comte passed from physical to moral positivism, and, from the same point of view, his study of Paul Janet's works leads him to conclude that true materialism is impossible.

The *Rapport* ends with the examination of two works that seem best to express Ravaisson's fundamental ideas: those of Anthelme-Edouard Chaignet and Jean-Charles Leveque on the theory of beauty. In commenting on them, he shows that beauty, charm and grace spring from souls that are capable of goodness and love. These qualities are the essential driving forces of the world, and as a result the realm of the aesthetic blends with that of philosophy. In short, while materialism finds its origins in the mathematical and physical sciences, living things reveal the spiritual, or rather the moral and aesthetic order, being dominated more by considerations of the order and harmony of the whole than by the detail of its parts; they are more concerned with form than with content. According to Ravaisson, this leaves two main routes for philosophy: analysis and synthesis. The perspective of synthesis is essentially that of art, which consists primarily of composition and construction, and thus also of science, when it is inventive. In other words, synthesis is the main method of philosophy. We judge the whole against the model of perfection which we carry in ourselves. Synthesis strives to explain everything through the absolute perfection that nothing limits; it strives by degrees towards the infinite. Similarly, in the order of wisdom, the wise man freely chooses the good. This also leads to a reconciliation of spirit and nature, which is one way of linking the classical belief that Eros was the first of all the gods with the Christian dictum, 'God is love'.

Testament philosophique repeats the essential themes of the *Rapport*. Ravaisson once more shows the importance of the role of feeling in the search for truth and restates the claim that beauty can only be created where there is enthusiasm. He underlines the values of generosity and love as the proper foundation of education and social morality. In addition to these claims, Ravaisson adds one for a natural belief in immortality.

See also: Beauty

Translated from the French by Robert Stern
PIERRETTE BONET

List of works

Ravaisson, F. (1837) *Essai sur la Métaphysique d'Aristote (On Aristotle's Metaphysics)*, vol. 1, Paris: Imprimerie royale; vol. 2, Paris, de Joubert, 1846; vol. 1, repr. 1963, Hildesheim: Georg Olms Verlagsbuchandlung. (Ravaisson here criticizes the philosophy of Aristotle, because he sees it as analytic philosophy, while he himself sought a synthetic position.)

Ravaisson, F. (1838) *De l'habitude (On Habit)*, Paris: H. Fournier; repr. in *Revue de métaphysique et de morale*, (2), 1-35, 1894; and as *De l'habitude*, with intro. by J. Baruzi, Paris: Alcan, 1933; repr. Paris: Vrin, 1981. (Ravaisson posits that habit is not simply a negative force, but allows the mind to free itself. In this way free will can be seen as part of the spontaneity of nature.)

Ravaisson, F. (1863) *La philosophie en France au dix-neuvième siècle*, Paris: Imprimerie impériale.(Ravaisson shows that French nineteenth-century philosophy consists of two schools of thought - materialism and spiritualism - and concludes his work with the triumph of the latter, the philosophy of spirit. This work is often referred to as the *Rapport* after some later editions, entitled *Rapport sur la philosophie en France au dix-neuvième siècle*.)

Ravaisson, F. (1893) 'Métaphysique et morale' (Metaphysics and Morals), introduction to *Revue de métaphysique et de morale*.(Overview of Ravaisson's philosophy, in which he claims that metaphysical truth transcends conventional thought processes and that it is necessary 'to love in order to understand'.)

Ravaisson, F. (1901) *Testament philosophique*, ed. X. Léon, in *Revue de métaphysique et de morale*, 1-31.(This work consolidates the fragments of Ravaisson's thought and brings to light the philosophical importance of free will, emotion and love.)

Ravaisson, F. (1933) *Testament philosophique et fragments*, revised and presented by C. Devivaise, Paris: Boivin. (Revised edition of *Testament philosophique* (1901).)

Ravaisson, F. (1953) *Essai sur la métaphysique d'Aristote, fragments du tome III (Héllénisme, Judaisme, Christianisme)*, ed. C. Devivaise, Paris: Vrin.(Ravaisson's unfinished third volume on Aristotle, posthumously assembled from manuscripts.)

References and further reading

Bergson, H.-L. (1934) 'La vie et l'oeuvre de Ravaisson', in *La pensée et le mouvant*, Paris: P.U.F., ch. 9; trans. M.L. Andison in *The Creative Mind*, New York: Philosophical Library, 1946.(A clear, complete, brilliant and perfectly accessible exposition, which can perhaps only be criticized for 'bergsonifying' Ravaisson.)

Chaignet, A.E. (1882-93) *Histoire de la psychologie des Grecs*, Paris, 2 vols.(In analysing the questions of Plato

and Greek philosophy, Chaignet distinguishes the idea of beauty from that of the Good, and the idea of benevolence from that of moral perfection.)

Dopp, J. (1933) *Félix Ravaisson, la formation de sa pensée d'après des documents inédits*, Louvain: Institut supérieur de Philosophie.(Dopp shows how Ravaisson, starting from his thoughts on Aristotle, Kant and Schelling, worked towards unifying mind and nature in concrete terms.)

Janicaud, D. (1969) *Une généalogie du spiritualisme français. Aux sources du bergsonisme: Ravaisson et la métaphysique*, The Hague: Martinus Nijhoff.(Janicaud criticizes the Bergsonian interpretation of Ravaisson, while recognizing that it illuminates the real difficulties of a philosophy which explains through life what cannot be explained through thought.)

Janicaud, D. (1985) *F. Ravaisson. L'art et les mystères grecs*, Paris: L'Herme.(Texts collected and edited by Janicaud, with an interview with Alain Pasquier, head of the Department of Antiquities at the Louvre. This collection presents the original aspects of Ravaisson's personality that led him to be both a philosopher and lover of art.)

Lévêque, J.C. (1861) *La science du beau étudiée dans ses principles, dans ses applications et dans son histoire*, Paris, 2 vols.(Dissatisfied by theories in which beauty consists in proportion, Lévêque rediscovers the principles of order and strength through the works of Malebranche and Leibniz. He gives a precise definition of what beauty consists in: it is greatness tempered by order, and expresses strength and will.)

Rawls, John (1921-)

Rawls' main work, A Theory of Justice (1971), presents a liberal, egalitarian, moral conception - 'justice as fairness' - designed to explicate and justify the institutions of a constitutional democracy. The two principles of justice outlined in this text affirm the priority of equal basic liberties over other political concerns, and require fair opportunities for all citizens, directing that inequalities in wealth and social positions maximally benefit the least advantaged. Rawls develops the idea of an impartial social contract to justify these principles: Free persons, equally situated and ignorant of their historical circumstances, would rationally agree to them in order to secure their equal status and independence, and to pursue freely their conceptions of the good.

In Political Liberalism (1993), his other major text, Rawls revises his original argument for justice as fairness to make it more compatible with the pluralism of liberalism. He argues that, assuming that different philosophical, religious and ethical views are inevitable in liberal society, the most reasonable basis for social unity is a public conception of justice based in shared moral ideas, including citizens' common comception of themselves as free and equal moral persons. The stability of this public conception of justice is provided by an overlapping consensus; all the reasonable comprehensible philosophical, religious and ethical views can endorse it, each for their own specific reasons.

1 Justice as fairness

Rawls' overriding aim is 'to provide the most appropriate moral basis for a democratic society' (1971: viii). Despite its many strengths, he sees the dominant utilitarian tradition as providing deficient foundations for democracy. Rawls begins with a normative conception of persons, whom he describes as free, equal, rational and endowed with a moral capacity for a sense of justice. Because of differences in knowledge and situations, free persons inevitably will develop different conceptions of the good. To pursue their good, they make conflicting claims on scarce resources. Principles of justice regulate the division of benefits and burdens resulting from social cooperation. Rawls contends that the appropriate way to decide principles for a democratic society is by conjecturing what principles free persons would agree to among themselves to regulate basic social institutions (the political constitution, property, markets and the family). But to ensure this agreement is fair, they must abstract knowledge of their own situations - of their talents and social positions and their conceptions of the good. Since these principles will be used to assess the justice of existing institutions and the reasonableness of existing desires and claims, Rawls further envisages that contracting parties abstract not just awareness of their own, but everyone's historical circumstances, desires and conceptions of the good. They are to be placed behind a thick 'veil of ignorance'. What such free individuals do know are general social, economic, psychological, and physical theories of all kinds. They also know there are certain all-purpose means that are essential to achieving their good, whatever it might be. These 'primary social goods' are rights and liberties, powers and opportunities, income and wealth, and the basis of self-respect.

The effect of these restrictions on knowledge is to render Rawls' parties strictly equal. This enables Rawls to carry to the limit the intuitive idea of the democratic social contract tradition: that justice is what could, or would, be agreed to among free persons from a position of equality (see Contractarianism). Rawls sees his strong equality condition, along with other moral conditions on agreement (that principles be universal, general, publicly known, final, and so on), as reasonable restrictions on arguments for principles of justice for the basic structure of society. These conditions define the 'original position', the perspective from which rational agents are to unanimously agree. Parties to the original position are presented with a list of all known feasible conceptions of justice and consider them in pairwise comparisons. The parties are rational, in that all utilize effective means to secure their ends, and are motivated by their interests, and so are moved to acquire an adequate share of the primary social goods needed to pursue their interests. The parties are also assumed to be rationally prudent (with zero time-preference), mutually disinterested (of limited altruism) and without envy.

Given these conditions, Rawls argues that the parties would unanimously agree to justice as fairness over the classical and average principles of utility, perfectionist and intuitionistic conceptions, and rational egoism. Its main principles state: (1) each person has an equal right to a fully adequate scheme of equal basic liberties, compatible with a similar scheme of liberties for all; and (2) social and economic inequalities must be attached to offices and

positions open to all under conditions of fair equality of opportunity and must be to the greatest benefit of the least advantaged members of society (the ‘difference principle’). The basic liberties of the first principle are liberty of conscience, freedom of thought, equal political rights, freedom of association, freedoms specified to maintain the liberty and integrity of the person (including rights to personal property), and the rights and liberties covered by the rule of law. These liberties are basic in that they have priority over the difference principle; their equality cannot be infringed, even if inequalities would increase the opportunities or wealth of those least advantaged. Moreover, the rights implied by both principles have priority over all other social values: they cannot be infringed or traded for the sake of efficiency, others’ likes and dislikes or perfectionist values of culture.

Rawls’ argument for these principles is that, given complete ignorance of everyone’s position, it would be irrational to jeopardize one’s good to gain whatever marginal advantages might be promised by other alternatives. For included in one’s conception of the good are the religious and philosophical convictions and ethical ways of life that give one’s existence meaning. It is fundamentally irrational, Rawls contends, to gamble with these given complete ignorance of risks and probabilities. In his later work, Rawls contends that parties in the original position are also moved by ‘higher-order interests’ to develop and exercise the ‘moral powers’ that enable them to engage in social cooperation - the capacity to form, revise and rationally pursue a conception of the good, and the capacity to understand, apply and act from a sense of justice. Parties agree on the two principles underlying justice as fairness since they provide each with primary goods adequate to realize these powers; other alternatives jeopardize these conditions.

One objection to Rawls’ theory is that the parties’ ‘maximin’ strategy of choice is too conservative. Harsanyi (1982) contends that Rawls’ parties should assume an equal probability of being any member in society. Given sympathetic identification with each person’s interests, they should choose (as if they were following) the principle of average utility (see Utilitarianism). But Harsanyi’s ideal chooser, although ignorant of their own identity, still has full knowledge of everyone’s desires and situations; Harsanyi views such knowledge as necessary for sympathetic identification. But Rawls’ parties are without knowledge of anyone’s desires and circumstances, and thus are rendered incapable of sympathetic identification, as well as making interpersonal comparisons of utility. Rawls also finds that, especially under conditions of radical uncertainty, gambling freedom to practise one’s conscientious convictions against added resources betrays a failure to understand what it is to have a conception of the good. For these and other reasons Rawls contends that it is difficult to see how the argument for average utility can arise from his original position. More important, assuming publicity of basic principles, a utilitarian society will not command the willing allegiance of everyone (especially those made worse off), and so will not evince stability (see §3). That basic political principles be publicly known is required by democratic freedom; otherwise citizens are under illusions about the bases of their social relations and are manipulated by forces placed beyond their control. With Rawls’ liberal egalitarian principles, nothing is, nor need be, hidden from public view in order to maintain social stability.

2 Democratic institutions

For Rawls the role of democratic legislation is not to register citizens’ unconstrained preferences and let majority preferences rule, but to advance the interests of all citizens, so that each has the status of equal citizen, is suitably independent and can freely pursue a good consistent with justice. The two principles of justice as fairness designate a common good that provides the end of democratic legislation. Ideally it should not be individual or group interests voting, but citizens and legislators, whose judgments are based on laws that best realize the common good of justice, as defined by the two principles. These principles imply a liberal constitution that specifies basic liberties immune from majority infringement. The first principle also requires maintaining the fair value of each citizen’s political rights, thereby establishing a limit on inequalities in wealth allowable by the difference principle. The second principle, the ‘difference principle’, preserves the ‘fair value’ of the remaining basic liberties. It suggests a criterion for deciding the basic minimum of resources each citizen needs to fairly and effectively exercise the basic liberties: property and economic institutions are to be so designed that those least advantaged have resources exceeding what the worst off would acquire under any alternative economic scheme (consistent with the first principle). This implies (depending on historical conditions) either a property-owning democracy (with widespread private ownership of the means of production) or liberal socialism. In either case, Rawls assumes markets are needed for efficient *allocation* of factors of production; but use of markets for *distribution* of output is constrained by the difference principle. Whatever effect redistributions from the market

have on allocative efficiency is not a problem for Rawls, since justice has priority over efficiency. The end of justice is not to maximize productive output whatever the distributive effects, any more than it is to maximize aggregate utility (see Justice §5).

3 Stability

The argument from the original position aims to show that justice as fairness best coheres with our considered judgments of justice (in 'reflective equilibrium'). But why should we care about justice enough to allow its requirements to outweigh our other aims? Stability addresses this issue of motivation. A conception of justice is 'stable' whenever departures from it call into play forces within a just system that tend to restore the arrangement. Unstable conceptions are utopian, not realistic possibilities. Hobbes argued that stability required a nearly absolute sovereign. This is incompatible with Rawls' democratic aim. To argue that justice as fairness is stable, Rawls appeals to principles of moral psychology to show how citizens in a 'well-ordered society' can acquire a settled disposition to act on and from the principles of justice. He then argues that justice as fairness is compatible with human nature, and is even 'congruent' with citizens' good in a society well-ordered by justice as fairness.

A person's good is the plan of life they would rationally choose based on their considered interests from an informed position of 'deliberative rationality'. Rawls' congruence argument contends that it is rational, part of a person's good, to be just and reasonable for their own sake in a well-ordered society. Assuming citizens there have a sense of justice, it is instrumentally rational for them to cultivate this capacity by doing justice, in order to achieve the benefits of social cooperation. On the Kantian interpretation of justice as fairness, the capacity for justice is among the powers that define our nature as rational agents; by developing and exercising this power for its own sake, citizens realize their nature and achieve moral autonomy. The Aristotelian principle is a psychological law which implies that it is rational to want to develop the higher capacities implicit in one's nature. Since the circumstances of a well-ordered society describe optimal conditions for exercising one's sense of justice, it is rational to want to cultivate the virtue of justice for its own sake and achieve moral autonomy. Justice and moral autonomy are then intrinsic and supreme goods in a well-ordered society, so the Right and the Good are 'congruent'. If so, it is not rational to depart from justice, and a well-ordered society manifests inherent stability.

Hegel argued that Rousseau's social contract, like Hobbes', was individualistic and incompatible with the values of community. Contemporary communitarians (for example, Sandel 1982) re-state Hegel's criticism, contending that Rawls' original position presupposes abstract individualism, with a metaphysical conception of persons as essentially devoid of the final ends and commitments that constitute their identity (see Community and communitarianism). In *Political Liberalism* (1993) Rawls contends that this is mistaken. In *A Theory of Justice* he presupposes, not a metaphysical conception of persons, but a practical account of the conditions of political agency, as grounded in the moral powers. Given congruence, maintaining justice and just institutions is the shared good that underwrites the values of community (or 'social union') among free and equal moral persons.

Rawls' Kantian congruence argument addresses the classical aim of showing how justice can be compatible with the human good. It is one of Rawls' most original contributions to moral philosophy. It also bears implications that led Rawls subsequently to revise his view.

4 Political liberalism

The problem with congruence is that it conflicts with the 'reasonable pluralism' of liberal societies, which should tolerate a wide range of religious, philosophical and moral views. The 'burdens of judgment' imply certain limitations on judgment, so that under free institutions we cannot expect agreement upon a comprehensive metaphysical, religious, or moral doctrine or conception of the good. But congruence implies that widespread acceptance of the intrinsic good of moral autonomy is a condition of liberal stability. By hypothesis, most conceptions of the good in a well-ordered society can endorse Rawls' principles of justice. The problem is, some may not accept the intrinsic goodness of moral autonomy. Teleological views, such as liberal Thomism or a reasonable utilitarianism, will gain adherents in a well-ordered society, and for these views justice and autonomy are at best but instrumental to the one rational and intrinsic good (the Vision of God, and aggregate or average utility, respectively). The incompatibility of congruence with reasonable pluralism then undermines Rawls' original argument for stability.

In *Political Liberalism* Rawls reformulates the justification of justice as fairness as a 'freestanding' political conception. He aims to provide a public justification for justice as fairness acceptable to all citizens of a well-ordered democracy. This requires an argument that is not grounded in Kant's or some other comprehensive ethical doctrine, but rather in certain fundamental intuitive ideas implicit in democratic culture. Rawls argues that the features of the original position can be construed as a 'procedural representation' of the idea of social cooperation among free and equal citizens implicit in a democracy. The principles of justice can then be represented as 'constructed' from a 'model conception' of democratic citizens as free, equal and possessed of the two moral powers that enable them to participate in social cooperation. These principles are politically justified since they are presented, not as true, but as most reasonable; they fit best with the considered political convictions of justice shared by democratic citizens, at all levels of generality, in wide reflective equilibrium (see Liberalism §5).

To complete this freestanding political justification, however, Rawls needs an alternative stability argument, one that, unlike congruence, does not rely upon premises peculiar to Kant's moral philosophy. The idea of 'overlapping consensus' says that the conception of justice that is politically justified as reasonable on grounds of individuals' shared conception of themselves as democratic citizens, will also be judged most reasonable or true on independent grounds, specific to each of the reasonable comprehensive doctrines gaining adherents in a well-ordered society. For its own particular reasons, each comprehensive view (for example, Kantians, utilitarians, pluralists, and religions accepting a doctrine of free faith) can endorse justice as fairness as true or reasonable. Justice as fairness then has one public, but many non-public, justifications in a well-ordered society. Assuming an overlapping consensus of reasonable comprehensive views exists there, justice as fairness evinces willing compliance, and hence stability.

See also: Equality; Freedom and liberty; Rights

SAMUEL FREEMAN

List of works

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(Rawls' major work.)

Rawls, J. (1975a) 'The Independence of Moral Theory', *Proceedings and Addresses of the American Philosophy Association* 48 (November): 5-22.(Argues that justice as fairness relies on a moral conception of the person, not a metaphysical conception of personal identity.)

Rawls, J. (1975b) 'Fairness to Goodness', *Philosophical Review* 84 (4): 536-54.(Response to early critics.)

Rawls, J. (1980) 'Kantian Constructivism in Moral Philosophy', *Journal of Philosophy* 77 (9): 512-72. (Important transitional work, in which Rawls introduces an account of Kantian constructivism.)

Rawls, J. (1982) 'Social Unity and Primary Goods', in A. Sen and B. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press, 159-85.(Contends that primary social goods are basic needs of free and equal moral persons.)

Rawls, J. (1985) 'Justice as Fairness: Political, not Metaphysical', *Philosophy and Public Affairs* 14 (3): 223-51. (Begins the transition to political liberalism.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press. (Significant reworking of the basis of justification for 'justice as fairness'.)

References and further reading

Barry, B. (1973) *The Liberal Theory of Justice*, Oxford: Oxford University Press. (Early assessment of *A Theory of Justice*.)

Daniels, N. (ed.) (1989) *Reading Rawls*, Stanford, CA: Stanford University Press.(Collection of the major critical reactions to *A Theory of Justice* in the early 1970s by Hart, Dworkin, Daniels, Scanlon, Nagel, Hare, Sen and others.)

Harsanyi, J. (1982) 'Morality and the Theory of Rational Behaviour', in A. Sen and B. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press.(Argues that behind a veil of ignorance, rational agents would choose, not Rawls' principles of justice, but maximizing average utility.)

Hegel, G.W.F. (1821) *Philosophy of Right*, trans. H.B. Nisbet, ed. A. Wood, Cambridge: Cambridge University Press, 1991.(Critical of liberal democracy and the social contract tradition on the grounds that they undermine community.)

Martin, R. (1985) *Rawls and Rights*, Lawrence, KS: University Press of Kansas.(Focuses on the centrality of

basic rights in Rawls' view.)
Pogge, T. (1989) *Realizing Rawls*, Ithaca, NY: Cornell University Press.(Defends justice as fairness against libertarian and communitarian criticisms, and argues for applying Rawls to global justice.)
Sandel, M. (1982) *Liberalism and the Limits of Justice*, Cambridge: Cambridge University Press. (Communitarian criticism of Rawls.)
Symposium on *Political Liberalism* (1994), *Chicago-Kent Law Review* 69 (3).(Contains articles by Cohen, Freeman and Greenawalt, among others, on Rawls' *Political Liberalism*.)

Realism and antirealism

The basic idea of realism is that the kinds of thing which exist, and what they are like, are independent of us and the way in which we find out about them; antirealism denies this. Most people find it natural to be realists with respect to physical facts: how many planets there are in the solar system does not depend on how many we think there are, or would like there to be, or how we investigate them; likewise, whether electrons exist or not depends on the facts, not on which theory we favour. However, it seems natural to be antirealist about humour: something's being funny is very much a matter of whether we find it funny, and the idea that something might really be funny even though nobody ever felt any inclination to laugh at it seems barely comprehensible. The saying that 'beauty is in the eye of the beholder' is a popular expression of antirealism in aesthetics. An obviously controversial example is that of moral values; some maintain that they are real (or 'objective'), others that they have no existence apart from human feelings and attitudes.

This traditional form of the distinction between realism and its opposite underwent changes during the 1970s and 1980s, largely due to Michael Dummett's proposal that realism and antirealism (the latter term being his own coinage) were more productively understood in terms of two opposed theories of meaning. Thus, a realist is one who would have us understand the meanings of sentences in terms of their truth-conditions (the situations that must obtain if they are to be true); an antirealist holds that those meanings are to be understood by reference to assertability-conditions (the circumstances under which we would be justified in asserting them).

1 Facets of the debate

Realism became a prominent topic in medieval times, when it was opposed to nominalism in the debate concerning whether universals were independent properties of things or if classification was just a matter of how people spoke or thought (see Nominalism). The impetus for the debate in modern times comes from Kant's doctrine that the familiar world is 'empirically real' but 'transcendentally ideal', that is to say a product of our ways of experiencing things, not a collection of things as they are 'in themselves' or independently of us. Kant's 'empirical realism', confusingly, is thus a form of antirealism (see Kant, I. §5).

Closely related is 'internal realism', as represented by Hilary Putnam, according to which something may be real from the standpoint marked out by a particular theoretical framework, while the attempt to ask whether it is real *tout court* without reference to any such framework is dismissed as nonsensical (see Putnam, H. §§7-8). This re-affirms the thesis propounded earlier by Rudolf Carnap, that there are 'internal' and 'external' questions about existence or reality (see Carnap, R. §5). An internal question is asked by someone who has adopted a language of a certain structure and asks the question on that basis. Only philosophers attempt to ask external questions (are there really - independently of the way we speak - physical objects?). But this is either nonsense or a misleading way of asking whether our linguistic framework is well suited to our practical purposes. 'Internal realism', it should be noted, is certainly not a form of realism, since it admits only language- or theory-relative assertions of existence.

By the mid-1980s, largely as a result of the work of Putnam and Dummett, it had become common to formulate the distinction between realism and antirealism in a variety of what are *prima facie* quite different ways. A realist, it was said, thinks of truth in terms of correspondence with fact, whereas an antirealist defines truth 'in epistemic terms', for instance as 'what a well-conducted investigation under ideal circumstances would lead us to believe'. A realist holds that there are, or could be, 'recognition-transcendent facts', whereas an antirealist denies this. Also present was the idea that an antirealist believes that there can be a 'reductive analysis' (see §2 below) of whatever subject matter their antirealism relates to, whereas a realist holds such analysis to be impossible. Seemingly still further from the origins of the distinction, it was said to be characteristic of realism to accept, and of antirealism to deny, the general validity of the law of excluded middle. Yet another version located the basic difference in the respective theories of meaning: a realist gave the meaning of a sentence by specifying its truth-conditions, an antirealist by specifying the conditions under which it could properly be asserted.

To come to terms with this debate, the reader therefore needs an awareness of the interrelations of the many definitions of the realism-antirealism distinction, and of the inexactness of fit between some of them and others.

2 Ontological realism/antirealism

The primary form of the definition deals directly in terms of what really exists. A realist about Xs, for example, maintains that Xs (or facts or states of affairs involving them) exist independently of how anyone thinks or feels about them; whereas an antirealist holds that they are so dependent. We are not speaking here of causal (in)dependence: the fact that there would be no houses if people had not had certain thoughts should not force us into antirealism about houses. So the point of the definition is better brought out by saying that *what it is for an X to exist* does not involve any such factors (whatever their causal role in the production of Xs may be). Nor does the definition entail an antirealist stance towards the mental. Realism about mental states is a *prima facie* plausible option, holding that our mental states are what they are whatever we think they are, or whatever we would come to think they were if we investigated.

Where philosophers have argued for realism about some particular subject matter (for example, universals, ethical value, the entities of scientific theory), one particular argument is repeatedly found. For the subject matter in question, it is claimed, we find that everyone's opinion is the same, or tends to become the same if they investigate, or that (in science) theory seems to 'converge', later theories appearing to account for the partial success of their predecessors. Why should this be, unless it is the effect of a reality independent of us, our opinions and our theorizing? (See Universals; Scientific realism and antirealism.)

In consequence, there are two broad antirealist strategies, both common. One is to argue that the supposed conformity of opinion, actual or potential, does not exist - so we hear of the diversity of ethical or aesthetic judgements, for instance, or the extent to which judgements of colour depend on viewing conditions and the state of the observer. The other is to accept the conformity, but explain it as arising from a uniformity of *our* nature rather than the independent nature of things. Thus it is argued that moral 'objectivity' is really 'inter-subjectivity' - that is, a result of shared human psychological responses rather than of independent moral properties in the world - or that the similarity between different languages' schemes of classification is a product of shared basic human interests, not something forced on us by 'real' universals.

In modern times nobody has made a more radical use of this method of explaining conformity of judgement in terms of intersubjectivity than Kant. He argued that even the experience of our environment as extended in space and time was a human reaction to things that were in themselves not of a spatiotemporal nature, and to which other beings might just as legitimately react altogether differently. In the face of this it may be felt that the argument from conformity is better used to establish a very abstract realism, namely that there must be *something* independent of us, rather than that any specific property or type of thing must be so.

Two other objections have been used against certain forms of realism. One is that the realist provides no account of how the supposed real things or properties can actually have an effect on our experience. What sense do we have, it is asked, that is affected by the ethical properties of the moral realist, or by the real properties of necessity and possibility that the modal realist posits? The common realist practice of speaking of 'intuition' in these contexts is rejected as providing only a word, not an answer. The second type of objection (christened the 'argument from queerness' by John Mackie (1977), who used it in the moral context) claims that the things or properties in which the realist believes would need to be too strange to be credible (see Moral realism; Modal logic, philosophical issues in).

A closely related definition of the realism-antirealism distinction focuses not on the independence of things but on the truth of judgements about them: realism takes truth to be correspondence with fact and our knowledge of truth to be a separate matter, whereas antirealism defines truth 'in epistemic terms', that is to say as what human beings would believe after the best possible application of their cognitive faculties. This is much more a change of perspective than of substance. It is natural to think that if some object exists independently of us, then judging truly must consist in getting our judgement to match the way the object is; while if the object is determined by (perhaps a projection of) our cognitive and/or affective faculties, judging truly can only mean judging as those very faculties lead us to judge.

Harder to assess is the position of *reductive analysis* in the debate. A reductive analysis exists where what makes statements about one kind of thing, *A*, true or false are the facts about another kind *B*. (*A*s are then said to be reducible to *B*s.) Classically, phenomenalism claims that statements about physical objects are thus reducible to statements about sensory experiences; behaviourism holds that propositions about mental states are reducible to ones about dispositions to physical behaviour. Does accepting such a reduction mean accepting antirealism about

the *A*s, while rejecting reduction of *A*-statements mean accepting realism? Some philosophers speak in this way, and there is a clear point to doing so: if a reduction is possible, then a complete statement of everything there ‘really is’ would not need to mention *A*s - it could speak of *B*s instead. Besides, reductive analyses have usually been offered in opposition to a different conception of what *A*s are, and in relation to *that* (rejected) conception of an *A* the reducer is certainly saying that there are no *A*s. But it is not thereby said that *A*s and facts about them are dependent upon us - only that they are really certain sorts of fact about *B*s; our attitude to their independence is therefore a question of whatever we think about the latter (see Reduction, problems of).

3 Epistemological versions

It is common to hear realism characterized in terms of the limits of knowledge as the belief that there are, or could be, ‘recognition-transcendent facts’ (meaning thereby facts which lie beyond *our* cognitive powers - there is no intention to saddle the realist with the view that there may be facts which simply could not be recognized at all). Antirealism then becomes the view that no such facts are possible.

The motivation for this epistemic version of the realism-antirealism divide is not hard to see. If the way something is is independent of the way we are, what could rule out the possibility that there should be facts about it beyond our powers of knowledge? Conversely, if its whole nature is due to the way we ‘construct’ it through our style of experience and investigation, how could there be anything about it that our cognitive faculties cannot recover? Although understandable, this is quick and imprecise. Consider someone who holds that the nature of the physical world is utterly independent of what human beings may believe it to be, but also has such anthropocentric theological inclinations as to hold that God must have given us cognitive powers equal, in principle, to discovering every fact about it. If we call this philosopher an antirealist on these grounds, we have surely changed the original subject, not just drawn it from another perspective.

This brings out the significance of formulating the epistemic criterion in terms of mere possibility (there *could be* recognition-transcendent facts) rather than actuality, thus allowing the philosopher who thinks, for whatever reason, that our cognitive powers are in fact a match for reality, still to be a realist by virtue of accepting that our powers might have been more limited without reality being any different.

It is one thing to suggest that there may be facts beyond our powers of recognition, quite another to hold that this is true of certain specific facts; the former is just modesty about our cognitive capacities, the latter a positive scepticism. So to imply an intrinsic connection between realism and scepticism, as some do, is very different from identifying realism with a belief in the possibility of recognition-transcendent facts.

Again, there is a plausible line of thought linking realism closely to scepticism. If a certain type of fact is as it is quite independently of us, then our knowledge of it must depend on an intermediary, namely the effect that it has upon us. But then we encounter the sceptical argument of which Descartes’ fiction of a malicious demon represents the classic formulation: how are we ever to know that this intermediary effect is produced by the sort of thing we think it is produced by, and not rather by something completely different? Hence, starting with realism, we arrive at scepticism.

However, it seems undesirable to use scepticism (and the absence of it) to characterize the realism-antirealism distinction. The classic argument from realism (as independence of the subject) to scepticism may be a formidable one, but it nevertheless involves substantial assumptions which can be challenged; to adopt terminology which makes it sound as if its conclusion were true by definition invites confusion. Besides, scepticism is not itself a precise notion, and there may be forms of it which apply even under certain antirealist conceptions. For instance, one who thinks that truth is to be understood as the opinion that would be reached under ideal conditions may still be a sceptic, because they remain sceptical of our ability to recognize ideal conditions or know how closely we have approximated to them.

4 Logical and semantic versions

It is often said that realism and antirealism can be distinguished by their attitude towards the law of excluded middle (the logical principle that, given two propositions one of which is the negation of the other, one of them must be true): the realist accepts it, the antirealist does not. Again, we can understand this if we think back to the original characterization of the distinction in terms of what is independently there and what we ‘construct’, what is

the case ‘in itself’ and what is so because of our ways of experiencing (see Intuitionistic logic and antirealism).

For explanatory purposes we may consider the world of literary fiction. Most people will be happy enough with the idea that, in so far as anything can be said to be true of the world of Macbeth, just those things are true which Shakespeare wrote into it. But in that case neither ‘Lady Macbeth had two children’ nor its negation ‘Lady Macbeth did not have two children’ is true, since Shakespeare’s text (we may suppose) does not touch on that question; the law of excluded middle fails in this ‘constructed’ world.

Passing now to a genuinely disputed case, there are those who think that whether a mathematical statement is true is one thing, whether it can be proved quite another; and there are those who think that truth in mathematics can only mean provability. For the latter the law of excluded middle is unsafe. From the fact that not-*p* cannot be proved, it does not follow that *p* can be proved; perhaps neither is provable and hence, on this view of mathematical truth, perhaps neither is true. And anyone who equates truth, in whatever sphere, with verifiability-in-principle by us will be liable to the parallel conclusion: only for those propositions *p* where failure to refute *p* is *ipso facto* to verify *p* may we rely on the law of excluded middle. Where verifying *p* and verifying not-*p* are distinct procedures, excluded middle fails. (It is because they are characteristically distinct when the proposition in question makes some claim about an infinite totality that we hear so much about infinite totalities and the rejection of excluded middle.) This explains why some writers (in particular Dummett) often say that the difference between realist and antirealist lies in the difference between their conceptions of truth (see Antirealism in the philosophy of mathematics; Realism in the philosophy of mathematics).

It can also be seen why it should have become common to express the realism-antirealism opposition as an opposition between theories of meaning, and why philosophers should be found speaking of realist and antirealist semantics. Any theory which ties meaning to verification, which equates the understanding of a sentence with a knowledge of those conditions that would verify it or would justify us in asserting it, promotes the view that we have no other idea of what it is for it to be true than for these conditions to be satisfied. Hence the realism-antirealism debate often exhibits neo-verificationist features; sometimes (especially by Dummett) antirealism is presented as the outcome of Wittgensteinian ideas about meaning, sometimes (especially by Putnam) of the alleged impossibility of explaining how our language could ever come to refer to the mind-independent items that realism posits (see Meaning and verification).

EDWARD CRAIG

References and further reading

Carnap, R. (1956) ‘Empiricism, Semantics and Ontology’, in *Meaning and Necessity*, 2nd edn, Chicago, IL: University of Chicago Press, 205-21.(Carnap expounds his distinction between ‘external’ and ‘internal’ questions, as mentioned above in §1.)

Dummett, M.A.E. (1963) ‘Realism’, in *Truth and Other Enigmas*, London: Duckworth, 145-65.(Seminal - and fairly difficult - paper, shifting the debate from the perspectives of §2 above to those of §4.)

Goodman, N. (1978) *Ways of Worldmaking*, Hassocks: Harvester Press.(Chapters I, VI and VII are especially relevant. Entertainingly written presentation of an all-inclusive antirealist position. Grows a little harder and more technical in chapter VII; full understanding of some points calls for knowledge of other works by Goodman.)

James, W. (1907) ‘Pragmatism’s Conception of Truth’, Lecture VI of *Pragmatism in Pragmatism and the Meaning of Truth*, Cambridge, MA, and London: Harvard University Press, 1978.(Popular, polemical presentation of the doctrine that ‘truth is *made*’.)

Mackie, J. (1977) *Ethics - Inventing Right and Wrong*, London: Penguin.(Chapter 1.9 is especially relevant. Contains the ‘argument from queerness’ against realism over moral value, referred to in §2 above.)

Putnam, H. (1981) ‘Two Philosophical Perspectives’, in *Reason, Truth and History*, Cambridge: Cambridge University Press.(Putnam links the traditional form of the debate to the question of the rival accounts of truth and problems about linguistic reference. For the most part not difficult reading; full understanding of some points calls for acquaintance with the two preceding chapters.)

Wright, C.J.G. (1987) *Realism, Meaning and Truth*, Oxford: Blackwell, 1-43.(The Introduction provides a wide-ranging survey of the issues, with antirealist slant; at times quite intricate. Useful bibliography.)

Realism in the philosophy of mathematics

Mathematical realism is the view that the truths of mathematics are objective, which is to say that they are true independently of any human activities, beliefs or capacities. As the realist sees it, mathematics is the study of a body of necessary and unchanging facts, which it is the mathematician's task to discover, not to create. These form the subject matter of mathematical discourse: a mathematical statement is true just in case it accurately describes the mathematical facts.

An important form of mathematical realism is mathematical Platonism, the view that mathematics is about a collection of independently existing mathematical objects. Platonism is to be distinguished from the more general thesis of realism, since the objectivity of mathematical truth does not, at least not obviously, require the existence of distinctively mathematical objects.

Realism is in a fairly clear sense the 'natural' position in the philosophy of mathematics, since ordinary mathematical statements make no explicit reference to human activities, beliefs or capacities. Because of the naturalness of mathematical realism, reasons for embracing antirealism typically stem from perceived problems with realism. These potential problems concern our knowledge of mathematical truth, and the connection between mathematical truth and practice. The antirealist argues that the kinds of objective facts posited by the realist would be inaccessible to us, and would bear no clear relation to the procedures we have for determining the truth of mathematical statements. If this is right, then realism implies that mathematical knowledge is inexplicable. The challenge to the realist is to show that the objectivity of mathematical facts does not conflict with our knowledge of them, and to show in particular how our ordinary proof-procedures can inform us about these facts.

1 Platonism and intuitionism

Platonism, probably the most popular form of realism (and sometimes itself called 'realism'), is the view that mathematics is the study of a collection of distinctively mathematical objects which exist independently of us: numbers, functions, sets, lines and so on. Thus the Platonist takes mathematical discourse at face value: to say that there exists a prime number between seven and twelve is to make a straightforward existential claim; and to say that nine is greater than seven is to make a comparison between two objects. Mathematical objects, on the Platonist account, are abstract; they occupy no particular position in space or time. They are immutable and eternal, which explains the unchanging character of mathematical truth. They are related to one another in essential and also unchanging ways: the fact that nine is greater than seven, for instance, is no mere accident, but is part of the nature of the system of natural numbers.

On the Platonist account, reasoning about mathematical objects is just like reasoning about ordinary concrete objects, in that the objects, and collections of them, are simply 'out there' waiting to be investigated and described. There are of course many more mathematical than concrete objects, and the abstractness of the mathematical objects sets them apart, but they are none the less like concrete objects in their relation to our cognitive processes: in both cases the objects would exist, bearing the properties and relations they currently bear, whether or not we had ever given them any thought. And in both cases, it is reasonable to suppose that there are facts about the objects which we do not yet know, and even facts which we may be incapable of discovering.

At the other end of the realist/antirealist spectrum from Platonism is the position of 'intuitionism', according to which it makes no sense to speak of true-yet-undiscoverable mathematical facts. According to L.E.J. Brouwer, the twentieth century's paradigm intuitionist, our knowledge of mathematical objects is best explained by the view that these objects are actually *created*, not discovered, by us: mathematical objects are the products of human mental activity. As the intuitionist sees it, to say, for instance, that there is a natural number with such-and-such property is to say that we can mentally construct such a number. To say that every natural number has a given property is not to make a claim about some pre-existing infinite collection, but is rather to assert the existence of a particular kind of proof-procedure (see Intuitionism; Constructivism in mathematics).

Reasoning about infinite collections in general is a much more cautious affair under the intuitionist's strictures than under the Platonist's, since the only true universal claims, according to the intuitionist, are those which come with effective proof-procedures. And because the only states of affairs relevant to the truth of mathematical

statements are those given by our own mental constructions, a statement is counted true by the intuitionist just in case it is provable. As a result, the kinds of mathematical reasoning accepted by the intutionist differ from those accepted by the Platonist, and the intuitionist must reject some of the truths of classical mathematics.

Though the original motivation for intuitionism was a view about the creative powers of the human mind, a quite different motivation stems, in the work of Michael Dummett, from general views about the relationship between meaning, truth and verification (see Dummett, M.A.E. §3). As Dummett sees it, to grasp the meaning of a statement is to understand how it is used, and in particular to know what counts as a verification or falsification of it. Since mathematical statements are verified via mathematical proof, the view as applied to mathematics is that the meaning of a mathematical statement is given not in terms of facts which it purportedly represents, but rather in terms of the procedures which would count as a proof of it. Again, the upshot is that truth reduces to provability, and the idea of a realm of independent mathematical facts is rejected as incomprehensible.

The Platonist response to the intuitionist is that the revisionary picture recommended by the latter is under-motivated. Against Dummett, the argument is that we have a perfectly good conception of verification-transcendent truth-conditions, and hence of meaning which is given in terms of objective truth, and independently of provability. With respect to the original intuitionists' conception of mathematical objects, the Platonist's claim is that there is no reason to think of the objects of mathematics as dependent on us, and that the view of the mathematician's activity as one of creation, rather than discovery, does not do justice to this activity. An important complaint made by the Platonist in this regard is that the picture of mathematical objects as mental constructions fails to do justice to the communal nature of mathematical research, since it makes each mathematical object accessible only to its creator.

2 Platonism and epistemology

The strength of the Platonist position is that only the Platonist can hold that the ordinarily-accepted statements of classical mathematics are strictly and literally true. We all want to affirm, in some sense, that there is a prime number between three and six. But having said this, the non-Platonist must add when pressed that the claim is not to be taken quite literally, or is at least not to be taken to have objective truth-conditions. Only the Platonist can say what in weaker moments almost all of us would like to say, which is that there simply *is* a prime number between three and six.

But the literal reading of mathematical discourse does come at a price. Particularly problematic, at least on the face of it, is the explanation of mathematical knowledge. The mathematical objects posited by the Platonist are, unlike concrete objects, entirely divorced from human experience; we cannot perceive them, nor do they bring about any perceptible effects. It is therefore not easy to see how we might be in a position to know anything about them. A central challenge to the Platonist is to show that Platonism does not leave our mathematical knowledge a mystery.

There are two main lines of Platonist response to this epistemological challenge. The first is to claim that we bear a relationship to mathematical objects which is analogous to the perceptual relation we bear to concrete objects: as empirical data are to knowledge of concrete objects, so mathematical intuition is to knowledge of numbers, functions and sets. Such a position is held by Kurt Gödel (1944), who points out that even our knowledge of concrete objects is not direct, but is inferred from immediately evident sensations; so too, on this view, knowledge of mathematical objects is inferred from the immediate knowledge of simple mathematical truths. The primary challenge facing the mathematical-intuition theorist is to give a clear account of the nature of mathematical intuition, and of how, precisely, it relates us to the objects in question.

The second Platonist approach is to argue that knowledge about mathematical objects is simply knowledge of propositions, or perhaps of entire theories, in which those objects figure. The problem of explaining knowledge about the objects is then reduced to the problem of explaining the justification of those propositions or theories. Thus, for instance, Gottlob Frege (1884) holds that knowledge about numbers is simply knowledge of the propositions of arithmetic, which is explained in terms of more general capacities, and not in terms of individual contacts with the objects. An important question facing such an account is whether it is really Platonist; whether, that is, knowledge of propositions of the right form should count as, at bottom, knowledge about objects.

An explanation which attempts to reduce knowledge of mathematical objects to knowledge of entire theories is found in the work of W.V. Quine (1960) and Hilary Putnam (1971). On this account, we know about mathematical

objects in exactly the same way that we know about non-mathematical objects. Consider, for instance, the subatomic particles of modern physics. We know about these things, as Quine sees it, not in virtue of direct contact with them, but rather in virtue of the fact that our best theory includes the claim that these objects exist, and contains specific assertions about them. The evidence for the existence of these objects is just the empirical evidence we have for the truth of the theory as a whole. Similarly for ordinary concrete objects like tables and chairs, on this account: we posit these objects in order to make sense of the brute sensory data before us, and the success of the theory that posits them is the evidence we have that they exist. Given this picture of objectual knowledge, mathematical objects pose no special problem: the claim that there are such objects, and that they have the usual mathematical properties, is an indispensable part of our best theory of the world, since mathematics is an essential part of modern physics. Thus we have just as much reason to believe in the objects of mathematics as we do to believe in the existence of pieces of furniture. Whether such an account is successful will turn largely on the tenability of the holist account of knowledge in general.

An interesting response to such indispensability arguments is the ‘instrumentalist’ position, according to which the usefulness of mathematical statements in science is no evidence of their truth, and hence is no evidence of the existence of mathematical objects. As argued by Hartry Field (1980), the central claim is that the statements of mathematics offer a conservative extension of (say) physics, which is to say that, useful as mathematics is as a tool, it allows us to reach only those conclusions in physics which we would have been able to reach (albeit less efficiently) anyway. On this reading, the ordinary mathematical existence-assertions turn out to be useful but not true. The viability of this project turns on whether the conservativeness result can in fact be demonstrated and whether physics can in fact be reformulated in Field’s ‘nominalistic’ terms.

3 Anti-Platonist realism

A different response to the problems surrounding mathematical objects is to view the apparent object-directedness of mathematical discourse as just a mode of speaking, and not an indication that the truths of mathematics really are about objects. On this picture, mathematics is concerned rather with general features of collections, structures, potential measurements, or some such thing. Just as talk involving such phrases as ‘Emily’s sake’, or ‘the average family’ are easily seen to be rephrasable without the use of these terms, and to be free of commitment to any objects meeting the descriptions, so too statements purportedly about mathematical objects can be rephrased, it is argued, in such a way as to remove any *prima facie* commitment to such objects. As long as the rephrasings have objective truth-conditions which render them true or false independently of our ability to recognize whether or not they obtain, such an account is fully realist, though anti-Platonist.

One example of such a rephrasing is as follows. In recognition of the close connection between arithmetic and cardinality, one might hold that claims apparently about natural numbers can be cashed out as claims about collections and their sizes. Thus, for example, the sentence ‘Every natural number has a successor’ is taken to mean ‘For every possible finite collection, there is a possible collection with one more member’, where the phrase ‘one more’ is itself cashed out without reference to the number one. (The reference to possibility will be required in order for the claim to remain true even if the world happens to be finite.)

A more widely applicable reduction is the ‘structuralist’ account, on which mathematical statements are taken to be either about all collections of entities exhibiting the appropriate structure, or about those abstract structures themselves. Thus, the sentence ‘Every natural number has a successor’ can be taken to mean that for every (possible) collection whose members exhibit the pattern N (which we describe in terms of the successor-like relation S), for each member x of the collection there is a member y of that collection such that $S(x,y)$.

With even these simple examples, a number of important issues arise. The first is that the quantification over ‘possible collections’ may involve ontological commitments at least as strong as the commitment to numbers. Thus despite avoiding a commitment to numbers, one does not by such a strategy obviously obtain ontological parsimony, or avoid the difficulties associated with knowledge about abstract objects. Similarly, the usual ways of cashing out the necessary relations between structures or collections will involve quantification over functions; thus in the attempt to avoid one kind of mathematical object, the analysis is apparently committed immediately to another. On the other hand, such reductionist approaches have the benefit of construing arithmetical knowledge in a way which requires no epistemic relation to particular individuals (such as natural numbers), but only to whole classes of collections or patterns. This is an advantage if general properties of such collections or patterns are in

principle accessible to a priori reflection in a way in which distinguishing properties of individuals are not.

An important question about non-Platonist realism is whether it is coherent. Can one simultaneously maintain that the ordinarily accepted statements of mathematics are indeed true, but that there are no mathematical objects? The difficulty in this position is that it requires one to hold that statements such as ‘There are prime numbers between ten and twenty’ are true, while maintaining that there are no numbers. It is not enough to say that the quoted sentence can be rephrased in a way which apparently avoids mention of numbers. For if the rephrasing simply tells us what it means to say that there are prime numbers between ten and twenty, then its truth implies that there are such numbers. (See Alston (1958).)

It is tempting to try to avoid Platonism by saying that statements such as ‘There are prime numbers between ten and twenty’, and hence ‘There are numbers’ are true in some secondary sense - perhaps true ‘within mathematics’, but that they are not *really* true. But it is not clear whether this can be made sense of, since it is not clear what it is for a mathematical statement to be true in one sense and not in the other. In sum: a successful analysis of mathematical discourse which appears to eliminate reference to and quantification over particular mathematical objects does not by itself establish the claim that mathematical statements are free of commitment to mathematical objects. For all that has been said so far, such an analysis may well be regarded as an account of what it is to be committed to those objects. Whether there is more to Platonism than the acceptance of the objective truth of certain mathematical statements - and hence whether there is sense to be made of a non-Platonist realism - will turn, then, on the answers to some extra-mathematical questions concerning ontological commitment and the nature of objects in general.

See also: Antirealism in the philosophy of mathematics; Logical and mathematical terms, glossary of; Objectivity; Realism and antirealism

PATRICIA A. BLANCHETTE

References and further reading

Alston, W.P. (1958) ‘Ontological Commitments’, *Philosophical Studies* 9: 8-17.(Argues that ontological commitment cannot be avoided by ‘translating away’ explicit reference to or quantification over purported entities.)

Benacerraf, P. (1973) ‘Mathematical Truth’, *Journal of Philosophy* 70: 661-80; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 403-20.(A clear account of tensions between a Platonist account of mathematical content and acceptable theories of mathematical knowledge and reference.)

Benacerraf, P. and Putnam, H. (eds) (1964) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.(A useful collection of central papers in the field.)

Bernays, P. (1935) Sur le platonisme dans les mathématiques, *L’Enseignement mathématique*, 1st series, 34: 52-69; trans. C.D. Parsons (1964), ‘On Platonism in Mathematics’, in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 258-71.(Classic discussion of Platonism and its alternatives.)

Brouwer, L.E.J. (1912) *Intuitionisme et Formalisme*, Groningen: Noordhoff; trans. A. Dresden, ‘Intuitionism and Formalism’, *Bulletin of the American Mathematical Society* 20: 81-96, 1913; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 77-89.(Gives an account of some of the central tenets of intuitionism.)

Dedekind, R. (1888) *Was Sind und was Sollen die Zahlen?*, Brunswick: Vieweg; trans. W.W. Beman, ‘The Nature and Meaning of Numbers’, in *Essays on the Theory of Numbers*, Chicago, IL: Open Court, 1901; repr. New York: Dover, 1963.(A classic of nineteenth-century foundations of mathematics. Gives an early non-eliminative but arguably structuralist account of the natural numbers.)

Dummett, M.A.E. (1967) ‘Platonism’, in *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press, 1978, 202-14.(A criticism of the Platonist’s claim that our relation to mathematical objects is analogous to our perception of concrete objects.)

Dummett, M.A.E. (1973) ‘The Philosophical Basis of Intuitionist Logic’, in H.E. Rose and J.C. Shepherdson (eds) *Proceedings of the Logic Colloquium, Bristol, July 1973*, Amsterdam: North Holland, 1975, 5-40; repr. in *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press, 1978, 215-47; and in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 97-129.(Argues, from general principles about the theory of meaning, that classical logic should

be replaced by intuitionistic logic within mathematical reasoning.)

Field, H. (1980) *Science Without Numbers*, Oxford: Blackwell.(An argument that the usefulness of mathematics in physical science is no reason to hold that mathematical statements are true, as discussed in §2.)

Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Evanston, IL: Northwestern University Press, 2nd edn, 1980.(This classic statement of the logicist position is also an excellent introduction to many issues concerning mathematical realism. Defends the view that numbers are 'self-subsistent objects', as discussed in §2.)

Gödel, K. (1944) 'Russell's Mathematical Logic', in P.A. Schilpp (ed.) *The Philosophy of Bertrand Russell*, Evanston, IL: Northwestern University Press, 125-53; repr. in Benacerraf, P. and Putnam, H. (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 447-69.(Gödel's view that the reasons for believing in mathematical objects are analogous to the reasons for believing in concrete objects, as discussed in §2.)

Hart, W.D. (ed.) (1996) *The Philosophy of Mathematics*, Oxford: Oxford University Press.(A collection of important recent papers.)

Parsons, C. (1990) 'The Structuralist View of Mathematical Objects', *Synthese* 84: 303-46; repr. in W.D. Hart (ed.) *The Philosophy of Mathematics*, Oxford: Oxford University Press, 1996.(A survey and critique of the structuralist treatment of mathematics.)

Putnam, H. (1971) *Philosophy of Logic*, New York: Harper.(Includes a clear statement of the 'indispensability' argument for the existence of mathematical objects, as discussed in §2.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(A classic. Includes the outline of Quine's views on ontological commitment and theoretical posits, as discussed in §2.)

Resnik, M. (1981) 'Mathematics as a Science of Patterns: Ontology and Reference', *Noûs* 15 (4): 529-50.(An explanation and defence of a modern structuralist account of mathematics.)

Reasons and causes

Imagine being told that someone is doing something for a reason. Perhaps they are reading a spy novel, and we are told that their reason for doing so is that they desire to read something exciting and believe that spy novels are indeed exciting. We then have an explanation of the agent's action in terms of the person's reasons. Those who believe that reasons are causes think that such explanations have two important features. First, they enable us to make sense of what happens. Reading a spy novel is the rational thing for an agent to do if they have that particular desire and belief. Second, such explanations tell us about the causal origins of what happens. They tell us that the desires and beliefs that allow us to make sense of actions cause those actions as well.

The idea that reasons are causes has evident appeal. We ordinarily suppose that our reasons make a difference to what we do. In the case just described, for example, we ordinarily suppose that had the agent had appropriately different desires and beliefs then they would have acted differently: had the person desired to read something romantic instead of exciting, or had the person believed that spy novels are not exciting, a spy novel would not have been chosen. But if what they desire and believe makes a difference to what they do then the desires and beliefs that are those reasons must, it seems, be the cause of the person's actions.

Despite its evident appeal, however, the view that reasons are causes is not without its difficulties. These all arise because of the manifest differences between explanations in terms of reasons and causal explanations.

1 The normative character of reason explanations

If we are told that a person acts because of a desire and belief they have then we have an explanation of their action in terms of their reasons; a rationalization, as it is sometimes called. Those who think that reasons are causes claim that such rationalizations are a kind of causal explanation. This view has evident plausibility, as it allows us to make straightforward sense of the idea that our desires and beliefs make a difference to what we do. If we had had different desires and beliefs then we would have behaved differently because the cause of our behaviour would itself have been different.

Even so, many anti-causalists think that reasons are not causes. One problem, as they see it, is that if reasons were causes then there would have to be underlying causal generalizations connecting desires and beliefs on the one hand, with actions on the other. For causal explanations are derived from, grounded in, such generalizations. However, they argue, rationalizations have a normative aspect which eludes capture in the purely descriptive terms in which causal generalizations deal. No such generalizations can therefore be found. An example will help to make the problem more vivid.

Suppose someone wants to illuminate the room and believes that they can do so by flicking a particular switch. Provided nothing else is wanted, and provided there are no interfering factors that it make it pointless for the person to try to do anything, it follows that they should flick the switch. Flicking the switch would be the rational thing to do, in one perfectly ordinary sense of the term rational; it is what we would expect to happen if we are to make sense of what is going on. However, say the anti-causalists, we cannot go on to conclude that the individual will flick the switch. For that would be to move illicitly from an 'ought' to an 'is'. In this respect, they insist, causal generalizations are therefore quite different from rationalizations. Causal generalizations purport to tell us what will happen without regard to what should. If one billiard ball moves towards another at a certain speed, and if nothing else interferes, then the other ball will move off in a certain direction at a certain speed. Whether or not it should is neither here nor there.

According to these anti-causalists, the fact that rationalizations have a normative aspect means that they serve a quite different function from that of causal explanation. Of course they agree that if, in the case just described, the agent does flick the switch then it will be useful to think of her as having done so for a reason; but they insist that it is a mistake to suppose that we thereby give the cause of what happened, because the function of such a rationalization is to give a useful interpretation of what happened, not a causal explanation. We interpret each other as creatures with desires and beliefs who act in various ways, and so instantiate the various rational norms, because doing so helps us to understand what we have done, are doing, and will do; it helps us to make sense of each other. However these interpretations no more presuppose that our desires and beliefs are states that play a

causal role than do the similar interpretations we give of the behaviour of desktop computers, electronic calculators and telephone exchanges in terms of the desires and beliefs that they possess - or so say these anti-causalists. In each case we simply work with a useful fiction or manner of speaking.

Note that two crucial claims have been made. First, talk of an agent's desires and beliefs is not unlike talk of similar states in electronic calculators and the like, a fiction or manner of speaking. Second, this manner of speaking is useful because it helps us understand what has happened, does happen and will happen. The question is: how are we to put these two claims together? If desire and belief talk is indeed useful then we need to know why this is so, and the most straightforward explanation available is surely that our desires and beliefs causally explain what we do. For then our understanding of what has happened, does happen and will happen is the familiar sort of understanding that comes from knowing about the causal origins of the events in question. On this way of looking at things there is therefore a crucial difference between human beings on the one hand, and electronic calculators, telephone exchanges and the rest on the other. Desire and belief talk really is a mere manner of speaking in the latter case, but not in the former. Electronic calculators and telephone exchanges are like human beings, but not the same as them, for they do not really have desires and beliefs, whereas human beings do. We must therefore ask why anti-causalists resist giving this most straightforward explanation of the usefulness of rationalizations, at least in the case of human beings. Why do they think that all desire and belief talk is a mere manner of speaking?

The answer is that they think the normative character of rationalizations makes it impossible to state rationalizations in the form of causal generalizations. It is, however, difficult to see why they think that this is so. For whenever we explain what someone does by appeal to their reasons we in fact make an additional, descriptive assumption, one which is perfectly suited to turn the normative claim anti-causalists accept into a descriptive causal generalization: namely, the assumption that the agent in question is a rational agent, in the relevant sense of rational. Thus, even though the claim 'If an agent wants to illuminate the room, and believes that they can do so by flicking the switch, and wants nothing else more, and there are no interfering factors present, then the agent should flick the switch' does not look at all like a causal generalization, the claim 'If an agent desires to illuminate the room, and believes that they can do so by flicking the switch, and wants nothing else more, and there are no interfering factors present, and the agent is rational, then they will flick the switch' most certainly does. Seen in this light, being rational is itself a substantive disposition agents possess, a disposition which makes its own difference to what they do.

In conclusion, it seems that anti-causalists are right that rationalizations have a normative aspect, but wrong that we are therefore unable to state them in the form of causal generalizations. We have seen no reason to disagree with the causalist that the usefulness of rationalizations lies in the fact that they offer us causal explanations.

2 The logical connection argument

Anti-causalists have another objection. Hume taught us that cause and effect are logically distinct. Causal generalizations like 'If one billiard ball moves towards another at a certain speed, and if nothing else interferes, then the other ball will move off in a certain direction at a certain speed' thus report substantive facts about the world. This is because, logically speaking, we can imagine the cause with an entirely different effect, or without any effect at all. The second billiard ball might have gone in any of a number of different directions, or disappeared, or stayed precisely where it was (see Causation).

However, anti-causalists point out, in the generalizations causalists suppose ground our rationalizations the so-called cause and effect are not logically distinct. The claim 'If someone wants to illuminate the room, and believes that they can do so by flicking the switch, and wants nothing else more, and nothing else interferes, and the person is rational, then they will flick the switch' does not report a substantive fact about the world because, if the agent is indeed rational, we cannot imagine these desires and beliefs having different effects. Imagining them with different effects - imagining the agent doing something different - is to imagine an agent with different desires and beliefs. The generalization linking desires and beliefs on the one hand, and doings on the other, unlike the claim linking the separate movements of billiard balls, thus reports a logical truth, not a substantive causal generalization.

The key idea behind this logical connection argument, as it is known, is plausible enough. We can indeed define

desires and beliefs as states that occupy certain functional roles: that is, desires and beliefs are states that are caused by various perceptual inputs, and which in turn cause various sorts of behavioural and other mental outputs (see Functionalism). Those who propose the logical connection argument are thus right that the cause and effect in a rationalization are not logically distinct in at least this sense: in describing the cause as a desire - that is, as a state that occupies a certain functional role - we thereby mention events of the effect kind, for we must mention behaviour in characterizing the relevant functional role.

However, if this is right, then it becomes clear that the anti-causal thrust of the logical connection argument is undermined. For the states of an agent that occupy the functional roles characteristic of desire and belief - presumably complex states of an agent's brain - are logically distinct from the behaviour they help to produce. It is our descriptions of these states as states that are occupiers of certain functional roles - that is, as states that, inter alia, cause certain sorts of behaviour - that are logically connected with facts about behaviour, not the states themselves. Even though the generalization connecting desires and beliefs with behaviour is indeed a logical truth, it thus follows that the desires and beliefs themselves - that is, the states that occupy the functional roles characteristic of desire and belief - are still logically distinct from the behaviour that they cause, just as Hume insisted cause and effect should be.

Moreover, despite the fact that there is this sort of logical connection between desires and beliefs and behaviour at the level of description, note that it is still appropriate to cite an agent's desires and beliefs in giving a causal explanation of what they do. In order to see why, consider an analogy. To say that a glass is fragile is similarly to ascribe to it a state with a certain functional role. A fragile glass is one which is in a state that causes it to break when struck in the appropriate circumstances. The generalization 'If a glass is fragile, and it is struck in the appropriate circumstances, and nothing else interferes, then it will break' is thus a logical truth, just like the generalization connecting desires and beliefs with behaviour. However we can still cite the fact that a glass is fragile in giving a causal explanation of its breaking. For we get information about how a glass broke when we learn that it broke because it was fragile. A fragile glass that breaks need not break because it is fragile, after all. (Just imagine that it breaks when we explode a nuclear bomb next to it.) To say that a glass broke because it was fragile thus gives us causal information because it narrows down the possible ways in which the breakage was caused to those that implicate the state of the glass that occupies the functional role characteristic of fragility.

Likewise, even though desires and beliefs are themselves just states that occupy certain functional roles, including the role of producing behaviour, we can still cite the fact that an agent has certain desires and beliefs in giving a causal explanation of what they do. For when someone who, say, desires to illuminate a room and believes that a room can be illuminated by flicking a switch does indeed flick it, they need not flick it because they have those desires and beliefs. (Just imagine that they flick the switch as the result of being hit by a falling chandelier.) To say that the agent flicked the switch because they had these desires and beliefs thus gives us causal information because it narrows down the possible ways in which the flicking of the switch was caused to those that implicate the states of the agent that occupy the functional roles characteristic of desire and belief.

Those who propose the logical connection argument are right that desires and beliefs are logically connected with behaviour, but wrong to draw the conclusion that they are therefore ineligible to be causes of behaviour. Desires and beliefs may not only cause behaviour; they may also be cited in giving causal explanations of behaviour.

3 The problem of wayward causal chains

Suppose we accept that reasons are causes. Do we thereby commit ourselves to the view that it is necessary and sufficient for an agent's having acted for reasons that they had certain desires and beliefs that caused them to behave in the appropriate way, or are we merely committed to the weaker claim that it is a necessary condition? In order to support the fundamental idea mentioned earlier - the idea that our reasons make a difference to what we do - we need only commit ourselves to the weaker claim. The stronger claim is worth considering, however, for it would be a striking victory for the causalist if we could spell out the content of a rationalization in purely causal terms.

There is, however, a well-known objection to this idea. Imagine an actor playing a role that calls for them to shake as if extremely nervous. We can readily suppose that, despite the fact that the actor wants to play the role and believes that it can be done by making certain movements, once they get on stage they are so overwhelmed by

their desire and belief that they are overcome by nerves and rendered totally incapable of action. Instead of playing the role as required, the actor just stands there, shaking nervously. What this case suggests is that it is therefore insufficient for an agent's having acted for a reason that they have desires and beliefs that cause appropriate behaviour. For an agent may well have desires and beliefs that cause such behaviour, and yet, because these states cause that behaviour in the wrong way, as in this case, it would be wrong to say that they act for a reason. In order to give necessary and sufficient conditions for an agent's having acted for a reason in purely causal terms we therefore need to rule out the possibility of such wayward causal chains. In this particular case, we need to rule out the possibility of the agent's desires and beliefs causing them to shake via causing them to become nervous.

Some causalists think that it is plain what is needed. The crucial feature in all such cases, they say, is that the match between what the agent does and the content of her desires and beliefs is entirely accidental. In the case just described, for example, it is entirely an accident that the actor wanted to make just the movements that their nervousness subsequently caused. These causalists therefore suggest that, in order to state a sufficient condition for an agent's having acted for a reason, we need simply to ensure that their behaviour is especially sensitive to the content of their desires and beliefs, as opposed to being sensitive to the operation of wayward factors like nerves. An agent acts for a reason, they suggest, only if, in addition to the other conditions, over a range of desires and beliefs that differ ever so slightly in their content - suppose the actor had desired to act nervously and believed that they could do so making their teeth chatter, or desired to act nervously and believed that they could do so by walking around wringing their hands, they would still have behaved in an appropriate way. This condition is clearly violated in the case described because, no matter what the content of the actor's desires and beliefs, nervousness would still have caused them to shake.

See also: Action; Intention; Akrasia; Mental causation

MICHAEL SMITH

References and further reading

Bishop, J. (1989) *Natural Agency*, Cambridge: Cambridge University Press.(States and attempts a solution to the problem of wayward causal chains.)

Davidson, D. (1963) 'Actions, Reasons and Causes', repr. in D. Davidson *Essays on Actions and Events*, Oxford: Oxford University Press, 1981.(Classic statement and defence of the claim that reasons are causes.)

Dennett, D.C. (1987) *The Intentional Stance*, Cambridge, MA: Bradford Books. (Argues that we rationalize behaviour in order to make sense of what happens, but not in order to find out about the causal origins of what we do.)

Hempel, C.G. (1962) 'Rational Action', *Proceedings and Addresses of the American Philosophical Association* XXXV.(Explains how we can state a rationalization in the form of a causal generalization.)

Jackson, F. and Pettit, P. (1990) 'Program Explanation: A General Perspective', *Analysis* 50: 107-17.(Explains how states that are functionally characterized may still figure in causal explanations.)

Melden, A.I. (1961) *Free Action*, London: Routledge & Kegan Paul.(Classic statement and defence of the logical connection argument.)

Reasons for belief

Reasons for believing something are one or another kind of ground for believing it. Some grounds provide evidence for a belief; others explain it; some are consciously known, others not. Philosophers are concerned with these and other aspects of reason, including the questions of whether reasons are also causes, whether they yield beliefs only by inference, and whether, by suitable reflection, believers can always become aware of the reasons for a belief they hold.

1 Varieties

To see the several kinds of reasons for believing, consider these examples:

(1) There *is reason* for believing (to believe) that Israeli-Palestinian negotiations will affect the United States' standing in the Middle East.
(2) You seem to be suggesting there is (some) *reason for me* to believe he will not represent us well at the meeting - ah, you are thinking of his previous blunder - I had forgotten.
(3) She *has reason* to believe the referee was unfair.
(4) The *reason why* he believes (the reason for his believing) this myth is not evidential; it is an emotional commitment to his father's views.
(5) Despite the reliable indications that her husband will survive, the only *reason for which* she actually believes this is his saying so, not the favourable statistics.

Case (1) illustrates a normative reason, a reason there *is* to believe the proposition in question (*p*). Such a reason can also be called impersonal, since its status as a reason does not depend on anyone's being aware of it or believing the evidential proposition (the reason): I might have, in the form of propositions I know or believe, the kind of political evidence in question, but the reason could also be merely implicit in what is known by experts. Case (2) illustrates a person-relative normative reason, one there is *for S* (the person in question) to believe. Case (3) exhibits a possessed reason, one *S has*, for believing *p*. In (4) we have an explanatory reason, a reason *why S* believes *p*. And case (5) illustrates a grounding reason, one *for which S* believes *p*. Here, *S* believes *p* on the *basis* of the reason. The belief that *p* thus stands in the (epistemic) basis (or basing) relation to the reason, which is a ground *on account of which S* believes *p*. Such reasons are at least psychological grounds sustaining the belief that *p*, what might be called 'pillars of conviction'. They may also be justifying grounds for *p* itself.

2 Reasons and reason states

Reasons as just described are propositions (conceived as meeting certain conditions, for example being true or believed by *S*). But whereas a proposition *q*, an obvious truth implying *p*, may be a reason there *is* to believe *p* whether or not anyone believes *q* - and thus *q* is in that sense an *external reason* for believing *p* - *q* must be believed by *S* in order to be a reason *for* which *S* believes *p* (case 5). That an accurate thermometer says a child has fever is a reason *to* think so, but this proposition is not a reason *for* which I believe this, until I come to believe the proposition and thereby come to believe the child has fever.

Given that we do not believe *for* reasons until they enter our minds, and given the obvious truth that one belief of *S*'s can be a reason *why S* holds another belief, some philosophers speak of (certain) beliefs themselves as reasons for believing. We can preserve these points, however, and achieve greater clarity, by construing reasons for believing as propositions and using 'reason state' for a belief *expressing* such a reason, in the sense that its content constitutes the reason. Reasons proper, as abstract entities, are presumably not causes. In so far as denials that reasons are causes make this point, they are plausible. But reason states, as concrete psychological elements, are good candidates to be causes (of a certain kind). Granted, a proposition can be cited as a reason why *S* believes *p* (case 4) and this *appears* to attribute causal power to it; but the proposition can *be* a reason why *S* believes *p* only if either its truth, or at least *S*'s believing it, actually explains why *S* believes *p*: these elements, *not* the proposition itself, play the relevant causal role. That the thermometer says the child has fever is the reason *why* I believe so, because it is my *believing* that the thermometer says this that *explains* my belief that the child has fever. A reason why I believe it, such as brain manipulation that implants the belief, could, however, explain why I believe the child has fever, without being a reason *for* anyone's believing this. Explaining reasons need not be reason states.

We can illustrate all five cases of a reason progressively and in relation to the same proposition: that the thermometer says the child has fever is (1) a reason (for anyone) to believe so; (2) a reason for me - but not the child, who is too young to understand thermometers - to believe so; (3) a reason I have to believe so - once I see the mercury above normal; (4) a reason why I believe so - provided the mercury level, as opposed to, say, my feeling the burning forehead, explains my believing so; and (5) a reason for which I believe so, when my belief is *based on* my reading of the thermometer.

3 Pragmatic reasons

Suppose I desperately need a loan and that, to be persuasive, I must believe I deserve it. Some philosophers would say I thereby have a reason for believing I deserve the loan. In one tradition, deriving from Blaise Pascal and William James, this might be considered a reason for believing that I deserve the loan, even though it provides no evidence of desert. On this view, a reason for believing *p* can be a consideration - not necessarily an evidential one - that, in a certain way, makes believing *p* useful. Such non-evidential considerations are perhaps better regarded as reasons to *cause* oneself to believe, and hence reasons for *acting* rather than for believing. These pragmatic reasons regarding belief are not among the kinds discussed below.

4 Dimensions of appraisal

Among the important questions concerning the relation between believing *p* and reasons for believing it are these: Are such reasons *evidential* (providing some degree of justification of *p*), *explanatory* (at least when combined with information implicit in the context of their ascription to *S*), *causal* (in sustaining or producing the belief that *p*), *internal* (in the sense that *S* can, by suitable reflection, become aware of them, as with one's sensations), *conscious* (in being present to awareness), *inferential* (in providing a kind of premise for the belief that *p*), and psychologically *connected* (linked in *S*'s mind to *p*, for example, as evidence for it)? Consider these questions for our five kinds of reasons.

(1) Impersonal normative reasons are evidential: if there is a reason to believe the child has fever, then that reason (the thermometer's indicating fever) provides some evidence for the proposition that the child has fever. Impersonal normative reasons need not be internal. Perhaps, however, they must be *accessible*, in the sense that someone could, by suitable investigation, become aware of them: normally, in speaking of reasons there are for believing *p*, one is referring to what is known or justifiedly believed, or can be discovered or justifiedly believed, for example, by reconsidering the data or searching the literature.
(2) Person-relative normative reasons are evidential at least *for S* (for example, for me as opposed to people lacking the relevant information). Neither these nor impersonal normative reasons need be explanatory, causal, conscious, inferential or psychologically connected; for they require no belief that *p*: I may, for example, never see the thermometer, and thus may have no belief to be explained, conscious or inferential, nor any occasion to connect the normative reason with *p* as, say, good evidence for *p*. (I might be disposed to make this connection if the reason is highlighted in association with *p*; but there can be a reason for me to believe *p* even if no such psychological connection exists.) Person-relative reasons for believing are usually internal and apparently must be accessible: the thermometer's indicating fever is not a reason *for me* if I cannot become aware of that reading.
(3) Possessed reasons may be normative, as where there is an evidential reason for believing *p, and* one acquires it; they can also be explanatory and causal; and they apparently must be internal. If they are internal, then although they need not be *in* consciousness and in that sense conscious, they cannot be unconscious in the sense that only external help, such as testimony from others, can bring them to *S*'s awareness. But reasons one has can be unconscious in the common sense (associated, for example, with self-deception and buried prejudices) that they are accessible to one only by systematic self-reflection. As to psychological connectedness, perhaps if *S has* a reason for believing *p* and is not merely in possession of information that *is* a reason for believing it, then *S* somehow takes the reason to support *p*. This would not imply a specific belief, say that the reason justifies *p*; one might just see the two in a familiar pattern, like 'If *q* then *p*; and *q*. Hence, *p*'. However, the view that having a reason for believing implies psychological connectedness makes it difficult to explain such self-reproach as 'I had reasons all along to believe it; why did I not see the connections sooner?'
(4) Explanatory reasons for believing are usually possessed and are often, but not necessarily, normative. Arguably, they must be causal. How, without playing some causal role in *S*'s believing *p*, would such reasons explain *why S* believes it? Explanatory reasons can certainly be unconscious; they need not imply any inference -

since manipulation of *S*'s brain could explain why *S* believes *p*; and, as that case also shows, they need not be psychologically connected.

(5) Grounding reasons must be possessed, explanatory and, arguably, causal, internal and psychologically connected. There is better warrant for thinking them connected than for merely possessed reasons, since here *S* believes *p* on the *basis* of the ground. This basis relation is not merely a causal one that might occur as, say, an accidental effect of rays hitting the brain; believing on the basis of a reason implies a response to the reason *as* some kind of support for p. If grounding reasons are connected, however, they still need not be evidential, at least if one can believe for bad reasons *and* such reasons can constitute not just inadequate evidence but none. Grounding reasons are unlike all the others in that they must be inferential, in the broad sense that they exist only if *S* actually believes p on the basis of another belief - a premise belief for *p*. This does not require *drawing* an inference; one can have a grounding reason for *p* either without going through an inferential process or after doing so, and thereby continue to hold the belief that *p* on the basis of the inferential ground in question.

5 Reasons versus grounds

Not all grounds for belief are themselves beliefs. Perceptions, for example seeing blue, are grounds for colour beliefs. Some philosophers call such grounds 'reasons for believing'. These non-inferential reasons can certainly be reasons *why S* believes; but in more perspicuous (and perhaps prevailing) usage, non-inferential grounds for beliefs are not called reasons *for* believing. It may be significant that such grounds are naturally expressible without a 'that'-clause: the reason *why* I believe - what explains my believing - that there is something blue before me is my seeing it, whereas at best I can say that my reason *for* believing this is that I see something blue. This reason is not the perception we started with but a self-ascription of that perception; and offering the reason suggests, misleadingly, that I believe there is something blue before me on the basis of *believing* I see something blue. A further advantage of the stricter terminology is that it allows us to leave open whether some grounds of our beliefs, including justifying grounds, may be external and inaccessible by reflection, as well as to hold that people's reasons for believing (as opposed to reasons why they believe) are always internally accessible to them - which best accords with the plausible idea that, on careful reflection, we can give the reasons for which we actually believe.

See also: Inference to the best explanation; Internalism and externalism in epistemology; Justification, epistemic; Rationality of Belief; Rational beliefs; Reliabilism

ROBERT AUDI

References and further reading

Alston, W.P. (1993) *The Reliability of Sense Perception*, Ithaca, NY: Cornell University Press.(Defends a social practice theory of justification on which justifying reasons must be, in the terminology above, both internal and evidentially adequate.)

Audi, R. (1993) 'Belief, Reason, and Inference', in *The Structure of Justification*, Cambridge: Cambridge University Press.(A detailed account of what it is to believe for a reason and of how so believing is connected with inference, explanation and justification.)

Chisholm, R.M. (1966) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice Hall, 3rd edn, 1989.(Systematic statement of an internalist epistemology that construes reasons for believing as ultimately deriving from foundations in self-presenting psychological states.)

Foley, R.A. (1987) *The Theory of Epistemic Rationality*, Cambridge, MA: Harvard University Press.(Detailed defence of a subjectivist conception of justification and, thereby, of reasons for believing.)

Goldman, A.I. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(A well-developed reliabilist theory of justification, which argues that grounds of beliefs need not be internal and that reasons for believing can justify only if they are objectively adequate.)

Moser, P.K. (1989) *Knowledge and Evidence*, Cambridge: Cambridge University Press.(A theory of the nature of reasons and evidence, how they justify, and how justification is connected with explanation.)

Plantinga, A. (1993) *Warrant and Proper Function*, New York: Oxford University Press.(A theory of warrant that construes reasons, especially those actually sufficient for knowledge, as meeting conditions of proper cognitive function for human beings.)

Reciprocity

To reciprocate is to return good in proportion to the good one has received, or to retaliate proportionately for harms. The central, contested philosophical issues surrounding reciprocity are whether reciprocity is a fundamental moral principle or a subsidiary one; how we are to measure fittingness and proportionality; and whether the norm of reciprocity requires that we reciprocate for all the goods we receive, or only for the ones we invite. While most philosophers believe that reciprocity is a subsidiary principle which is unproblematic only in the context of fully voluntary transactions, there are significant minority views on this matter.

To some political philosophers, reciprocity appears to be the linchpin of a theory of justice that stands mid-way between an altruistic concern for the welfare of everyone (justice as impartiality) and an egoistic concern for reaping the benefits of cooperation (justice as mutual advantage). On this intermediate view justice involves a commitment to impartial principles that further the welfare of everyone, but only when others have reciprocal commitments. It thus defines a unique conception of justice, distinct from standard utilitarian and contractualist ones (see Justice). Other philosophers, however, see reciprocity as derived from more fundamental moral principles.

Interest in reciprocity as a separate, fundamental principle of social life and morality arose in the twentieth century. Early anthropological studies of gift-giving rituals were developed into comprehensive theories of social exchange, and integrated into structuralist accounts of social life. Stated abstractly, reciprocity appears to social scientists to be a universal social norm, implicated in a wide array of social practices from the most trivial aspects of etiquette to the most lethal aspects of retribution. To evolutionary theorists it seemed plausible to suppose that norms of reciprocity were genetically encoded. Moreover, the development of rational choice theory, and the mathematical manipulation of collective action problems, gave support to the notion that reciprocity (tit-for-tat, beginning with a beneficent opening move) was the best strategy for dealing with (rational) non-cooperative behaviour (see Rational choice theory).

Philosophical efforts to find a rational foundation for morality and justice have often alluded to reciprocity - or at least to related notions such as mutuality and fair play. There are, however, some thorny philosophical issues buried in the concept and practice of reciprocity. One is the question of measurement. In order to reciprocate we need to rank goods (and bads), so as to determine the fitting and proportional response in each case. On the good-for-good side of things, for example, ‘fittingness’ might be determined by what the potential recipients will accept as similar goods. Thus if I make you a gift of a delicacy (say a double chocolate cake) just to delight you, though I do not myself like it or other sweet desserts, you will have to find some other sort of thing - something I do like - if you want to make a fitting return to me. Further, the quantity that will count as a proportional return still needs to be decided, and it is clear that equal money value is not an adequate principle. (I may like emeralds, but giving me the market value of a chocolate cake in emerald chips is a joke, not a reciprocal response.) Besides, the declining marginal utility of money means that an equal market value rule would be oppressive for the poor. Equal marginal sacrifice is a more plausible principle. We could set proportionality to how much the gift ‘costs’ the giver on the margin. Then the reciprocator, if poor, will not ‘suffer’ more from making the return than the giver did, and if rich, will not be able to reciprocate painlessly (unless the gift was painless).

Unfortunately, this rather tidy metric does not look very plausible on the negative side of things (that is, in implementing the policy of returning bad for bad). Victims who want compensation demand plausibly to be made whole; victims who want to exact punishment demand plausibly that the malefactor’s punishment be proportional to the wrongful intent, not to the actual harm done. (People die from negligence as well as from premeditated murder. The harm done is the same; the suffering might even be greater for some victims of negligence.) So the doubts about a coherent account of fittingness and proportionality remain.

Perhaps the deepest philosophical issue raised by reciprocity, however, concerns the scope of the norm. A great deal of the good we receive from others comes to us uninvited and unavoidable. Some of it (like the care of loving parents for their young children) is given to us directly, intentionally. Some of it (like a rich cultural environment) is merely a public good from which we cannot be excluded. The ‘wide’ view of reciprocity holds that the norm covers all of one’s interactions with others; it holds that one owes a fitting and proportional return for all the good

one receives, and not merely for the good actually accepted or invited. The ‘narrow’ view argues that at most we owe a return only for the good we have accepted or invited by voluntary participation in the relevant social practices. On the wide view, reciprocity has important contributions to make to solving general cooperation problems, to giving accounts of obligations to parents, to the state, and to future generations (see Future generations, obligations to; Obligation, political). But even in its narrow form, reciprocity speaks to fundamental issues of justice and fair play.

See also: Desert and merit; Egoism and altruism; Rectification and remainders

LAWRENCE C. BECKER

References and further reading

Axelrod, R. (1984) *The Evolution of Cooperation*, New York: Basic Books.(Covers rational choice theory, sociobiology and tit-for-tat as a cooperative strategy.)

Barry, B. (1995) *Justice as Impartiality*, Oxford: Oxford University Press, 46-51. (Criticizes justice as reciprocity.)

Becker, L. (1986) *Reciprocity*, Boston, MA and London: Routledge; repr. Chicago, IL: University of Chicago Press, 1990.(Argues for the ‘wide’ view of reciprocity, namely that one owes a fitting and proportional return for all the good one receives, and not merely for the good actually accepted or invited.)

Buchanan, A. (1990) ‘Justice as Reciprocity versus Subject-Centered Justice’, *Philosophy and Public Affairs* 19: 227-52.(Addresses the question of whether reciprocity generates a distinct theory of justice.)

Gibbard, A. (1991) ‘Constructing Justice’, *Philosophy and Public Affairs* 20: 264-79. (Defends justice as reciprocity as a distinct theory.)

Gouldner, A. (1960) ‘The Norm of Reciprocity’, *American Sociological Review* 25: 161-78.(Classic sociological thesis about the universality of reciprocity.)

Mauss, M. (1923/4) ‘Essai sur le don, forme et raison de l’échange dans les sociétés archaïques’, *L’année sociologique* 2 (1): 30-186; trans. W.D. Halls, foreword by M. Douglas, *The Gift: Forms and Functions of Exchange in Archaic Societies*, New York: W.W. Norton, 1990.(The anthropology of primitive exchange.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(Views reciprocity as an element of fairness in the norms of cooperation and justice.)

Sahlins, M. (1981) *Stone Age Economics*, New York: Aldine.(The sociology of primitive exchange; includes extensive bibliography on anthropology.)

Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan; 7th edn, 1907; 7th edn repr. Chicago, IL: University of Chicago Press, 1962, books III, IV.(Material on gratitude.)

Simmons, J. (1979) *Moral Principles and Political Obligations*, Princeton, NJ: Princeton University Press.(Raises objections to using the principle of fair play as a basis for political obligation.)

Recognition

The concept of recognition has played an important role in philosophy since ancient times, when the good life was thought to depend partly on being held in regard by others. Only Hegel, however, made recognition fundamental to his practical philosophy. He claimed that human self-consciousness depends on recognition, and that there are different levels of recognition: legal or moral recognition, and the forms of recognition constituted by love and the state. A similar tripartite distinction can be used to ground a plausible modern account of ethics.

1 The history of recognition

In one way or another the concept of recognition has always played a major role in practical philosophy: in ancient ethics the conviction was predominant that only someone whose conduct met with high regard in the *polis* could lead a good life; Scottish moral philosophy was led by the idea of public recognition or disapproval as a social mechanism through which the individual was motivated to attain the desirable virtues; with Kant the term *Achtung* (respect) eventually takes over the function of a highest moral principle in the sense that it contains the nucleus of the categorical imperative: to treat any other human being only as an 'end in themselves'. None of these authors, however, made the principle of recognition itself the foundation of an ethic; the term was eclipsed by others, regarded as more fundamental. Only Hegel, in this respect a solitary pioneer, used the concept of recognition as a foundation of his practical philosophy.

2 Hegel and recognition

When Hegel undertook the project of reconstructing the development of human morality with the aid of the term recognition, he based it on Fichte's explanation of the possibility of human consciousness of freedom in terms of a theory of recognition: by examining Kant and Schelling on the foundations of natural law, Fichte (1796/7) became convinced that subjects can only achieve consciousness of their freedom if they challenge each other to exercise their autonomy and accordingly recognize each other (and themselves) as free persons (see Autonomy, ethical; Freedom and liberty; Fichte, J.; Hegel, G.; Kant, I.; Schelling, F.; Natural law).

If the idea so outlined was bolstered with elements of Hobbes' and Rousseau's political anthropology, Hegel could use the result as the foundation of his theory of recognition: human self-consciousness depends on the interpersonal experience of being recognized by other human beings (see Hobbes, T.; Rousseau. J.-J.)

For Hegel's purposes, however, the bare assertion of a necessary connection between self-consciousness and interpersonal recognition was not enough. To explain how the experience of recognition could bring about progress in morality, it was necessary in addition to explain the dynamic interrelationship that had to exist between the attainment of self-consciousness and the moral development of whole societies. Hegel's answers to these complex questions form the nucleus of his model of a 'struggle for recognition', according to which moral progress takes place in three different stages of recognition, each of them reached through an interpersonal struggle. Subjects engage in this struggle to confirm their claims to identity.

Peculiar to this approach is a thesis reaching far beyond Fichte, namely the classification into three different types of recognition once the interpersonal basis of self-consciousness had been understood. The mechanism that concedes reciprocally a sphere of individual freedom (which Fichte had in mind in his account of natural law) in fact explains the formation of a subjective consciousness of rights, but by no means does it comprehend the full positive self-understanding of a free person (see Rights).

Hegel therefore adds to legal recognition, which was to include something like what Kant understood as moral respect, two more types of reciprocal recognition which have also to correspond with particular stages of the individual's self-awareness: in love, the subjects recognize each other reciprocally in the unique nature of their desires in order to attain emotional security in the articulation of their desires (see Love). Finally, in the moral environment of the state, a type of recognition unfolds which is to allow the subjects to further their esteem for each other through the very qualities which contribute to the maintenance of social order (see State, the).

In his early works Hegel seems to have been convinced that the transition between these different areas of recognition is characterized by a struggle, conducted among persons in order to be respected in their gradually

growing self-esteem: the claim to be recognized in more and more dimensions of one's own personality leads to an inter-subjective conflict, the solution of which can only be the establishment of yet another sphere of recognition (Honneth 1992: ch. 1).

Hegel is not sociologist enough to think of this process as actually constituting modern societies. Still very much limited by the horizons of German Idealism, he sees it as the totality of the intellectual work which the subjects together have to perform in order to build a common world of the 'objective spirit' (see Idealism).

None the less, Hegel's early model of a struggle for recognition has proved complex enough still to be able to stimulate much present thought in moral philosophy and social theory (Siep 1979; Wildt 1982; Honneth 1992). Yet as early as the *Phänomenologie des Geistes (Phenomenology of Spirit)* (1807) Hegel replaced his initial programme with a concept in which the assumptions of the later system came more and more to fruition; from now on the constitution of social reality is no longer explained through an interpersonal process of conflict-formation, but seen as a result of a dialectic advance of the mind.

However, Hegel's divisions into family, civic society and state in *Grundlinien der Philosophie des Rechts (Philosophy of Right)* (1821) reflect once more the earlier distinction of three types of recognition; and it is this division into three which today allows the development of Hegel's mature system into a practical philosophy (Hardimon 1994).

3 An ethic of recognition

Somewhat as it was for Hegel, today the first step towards a contemporary ethic of recognition is to show that the potential for moral injury arises from the inter-subjectivity of the human form of life: human beings are vulnerable in that specific way we call 'moral', because their identity is constructed out of practical self-awareness which is dependent on the assistance and approval of other human beings from the very beginning (Habermas 1983: 53-67).

If we are to draw a positive idea of morals from this anthropological premise, then the 'moral standpoint' is to be defined by the network of attitudes we have to adopt in order to save other human beings from injuries resulting from the communicative nature of their self-awareness. Positively defined, morality consists of those types of recognition that we are obliged reciprocally to take up if we are to secure together the preconditions of our personal freedom (see Morality and ethics).

Hegel assumed that the differentiation of types of recognition is a matter of which levels of personal self-awareness, or of the human consciousness of freedom, can sensibly be distinguished: for each stage of positive self-awareness on which we have to rely in constructing our autonomy, another type of interpersonal recognition was required, so that the differentiation into family, civil/civic society and state defined three necessary conditions of personal freedom.

If we follow this proposition and detach it from any restriction to particular institutions, a division into three emerges, which appears plausible even today; the three types of recognition then define interpersonal conditions of personal identity as well as moral attitudes, which form together the 'moral point of view': (1) the individual is recognized as someone whose wishes and desires are of unique value to another person; for this type of recognition, which has the character of unquestioning care, the moral philosophical tradition offers terms like 'love' or 'care'; (2) the individual is recognized as a person deserving the same moral responsibility as all others. For this type of recognition, which is characterized by universally equal treatment, the Kantian term 'moral respect' has become established; (3) the individual is recognized as a person whose abilities are of constitutive value to a community; for this type of recognition, which is characterized by special esteem, philosophical tradition lacks adequate moral terms; it may however be appropriate to have recourse here to categories like 'solidarity' or 'loyalty' (see Solidarity).

To specify these three different patterns of recognition in their moral content, and if possible even to interpret them as moral duties, is the task of any theory of morality which aspires to be heir to the Hegelian theory of recognition.

See also: Finitude and vulnerability; Friendship; Moral agents; Morality and identity; Rectification and remainders

Translated from the German by Ferdinand Knapp

AXEL HONNETH

References and further reading

Bernstein, J.M. (1996) 'Confession and Forgiveness: Hegel's Poetics of Action', in R. Eldridge (ed.) *Beyond Representation: Philosophy and Poetic Imagination*, Cambridge: Cambridge University Press.(Argues for transgressive actions and misrecognition - strife - as constitutive of the recognitive relations of modern subjects.)

Fichte, J.G. (1796/7) *Grundlage des Naturrechts nach Principien der Wissenschaftslehre*, trans. A.E. Kroeger, *The Science of Rights*, Philadelphia, PA: Lippincott, 1869.(A less than reliable translation. Elegantly argues that recognizing the rights of another free self is a necessary condition for self-consciousness. Probably the first fully worked out account of a recognitive analysis of self-consciousness.)

Habermas, J. (1983) *Moralbewußtsein und kommunikatives Handeln*, Frankfurt: Suhrkamp; trans. C. Lenhardt and S. Nicholsen, *Moral Consciousness and Communicative Action*, Cambridge, MA: MIT Press, 1990.(See 'Discourse Ethics: Notes on a Program of Philosophical Justification' in Lenhardt and Nicholsen's translation (43-50). This portion of Habermas' argument builds towards an account of normative recognitive relations between selves from the example of emotional responses to actions, such as feeling resentment, only available in a performative attitude.)

Hardimon, M.O. (1994) *Hegel's Social Philosophy: The Project of Reconciliation*, Cambridge: Cambridge University Press.(Reconstructs the argument of Hegel's *Grundlinien der Philosophie des Rechts* as arguing that reconciliation between self and society is only possible if strong conditions of individuality and strong conditions of social membership are satisfied simultaneously.)

Hegel, G.W.F. (1807) *Phänomenologie des Geistes*, trans. A.V. Miller, *Phenomenology of Spirit*, Oxford: Oxford University Press, 1977.(Hegel's most systematic attempt to ground self-consciousness in structures of recognition, beginning with the dialectic of master and slave.)

Hegel, G.W.F. (1821) *Grundlinien der Philosophie des Rechts*, trans. H.B. Nisbet, *Elements of the Philosophy of Right*, Cambridge: Cambridge University Press.(Hegel's analysis of right in terms of the recognitive structures of family, civil society and the state.)

Honneth, A. (1992) *Kampf um Anerkennung*, Frankfurt: Suhrkamp; trans. J. Anderson *The Struggle for Recognition: the Moral Grammar of Social Conflicts*, Oxford: Polity Press, 1995.(An attempt to provide a formal reconstruction of Hegel's conception of self-consciousness and recognition through a series of stages: love, rights and solidarity.)

Siep, L. (1974) 'Der Kampf um Anerkennung: Zu Hegels Auseinandersetzung mit Hobbes in den Jenaer Schriften', *Hegel-Studien* 9: 155-207; trans. C. Dudas 'The Struggle for Recognition: Hegel's Dispute with Hobbes in the Jena Writings', repr. in J. O' Neill (ed.) *Hegel's Dialectic of Desire and Recognition: Texts and Commentary*, Albany, NY: State University of New York Press, 1996.(Tracks the transformation of Hobbes' notion of honour into recognition in Hegel's early writings.)

Siep, L. (1979) *Anerkennung als Prinzip der praktischen Philosophie. Untersuchungen zu Hegels Jenaer Philosophie des Geistes (Recognition as the Principle of Practical Philosophy: Investigations of Hegel's Jena Philosophy of Spirit)*, Freiburg: Alber Verlag.(Provides clarification on the distinction between intersubjective and social forms of recognition in Hegel, and argues that recognition is a synthesis of love and strife.)

Wildt, A. (1982) *Autonomie und Anerkennung. Hegels Moralitätskritik im Lichte seiner Fichte-Rezeption (Autonomy and Recognition: Hegel's Critique of Morality in Light of his Reception of Fichte)*, Stuttgart: Klett-Cotta.(Influential analysis of Hegel's critique of Kantian moral theory as grounded in appropriation of Fichte's doctrine of recognition.)

Williams, R. (1992) *Recognition: Fichte and Hegel on the Other*, Albany, NY: Sate University of New York Press.(An analysis of the concept of recognition in Hegel's *Phänomenologie des Geistes* as it emerges out of Fichte's writings.)

Rectification and remainders

Forgiveness, mercy and gratitude are rectifications, attempts to correct imbalances or set things right between us. Guilt is a moral remainder, a residue acknowledging an unexpiated wrong. Remainders offer us a limited redemption in revealing our appreciation that not everything has been made right.

Forgiveness manifests compassion for wrongdoers, who may or may not deserve it. Questions arise about when, what and whom we can forgive, what forgiving achieves, and when we ought or ought not to forgive. Forgiveness has special value in personal relationships, enabling their renewal. Connections among punishment, repentance, forgiveness and regret (another remainder) are complex, sometimes paradoxical.

Mercy often manifests forgiveness, as in pardons and amnesties. Yet we can also show mercy in administering rules where there has been no wrong. Is mercy unjust? Answers vary according to whether the case resembles a criminal offence or a civil suit. Grounded in others' sufferings rather than their deeds, mercy has us see ourselves in them, but the value of doing so can be qualified by considerations of justice and of self-respect.

Mercy and forgiveness sometimes evoke gratitude, appreciative acknowledgement of another's goodwill. Gratitude can be deserved or misplaced. Debts of gratitude are paradoxical, giving rise to ethical questions. When do we owe more than emotional response? What does reciprocity require between unequals? Some paradoxes may be solved by understanding obligations of gratitude as like those of a trustee rather than those of a debtor.

Guilt is emotional self-punishment (often relievable by forgiveness) which continues even after compensation, restitution or punishment by others. Questions arise about when it is rational and how it is related to shame, remorse, regret and repentance, which are also remainders.

1 Rectification and remainders

Rectification is setting matters right, redressing wrongs, correcting imbalances, settling scores. Remainders are rectificatory feelings that go some way towards morally redeeming us for what we cannot rectify by our actions and sometimes for actions that not even others can rectify.

Apart from punishment, praise and blame, rectificatory actions most explored in philosophy are forgiveness, mercy and gratitude (see Praise and blame; Crime and punishment). Other rectificatory actions - amnesty, apology, compensation, pardon, repentance, restitution and reward - are considered here in terms of their relationships to these three. Because we cannot claim forgiveness, mercy or gratitude as a right, these actions resemble what Kant (1797) called 'imperfect duties', which allow the agent some latitude regarding when, where, towards whom and how much (as in our duty to help others) instead of specifying exactly what must be done (see Kant, I. §10). And yet, as with 'perfect duties', such as the duty to keep a promise, these rectifications can be due specific individuals, who deserve them from us (see Promising). Omissions of gratitude, mercy or forgiveness can make us as vulnerable to moral criticism as failures of honesty, fidelity or integrity. As character traits, ingratitude, being unforgiving and being unmerciful are grave faults, while gratitude, forgiveness and mercy are cherished virtues of the moral life (see Virtues and vices §§2-3).

Ever since Bernard Williams observed that 'moral conflicts are neither systematically avoidable, nor all soluble without remainder' (1965: 179), certain emotional residues have been referred to as 'moral remainders' (see Williams, B.A.O. §5). These include guilt, remorse, regret and sometimes shame, rectificatory responses of feeling rather than action. They reveal relevant positive values of an agent who has acted wrongly or is identified with a bad action or state of affairs. Aristotle said of shame, explaining its status as a 'quasi-virtue', that ideally the occasion for it would not arise, but if it does, we do better to have shame than to be shameless. Remainders can survive rectificatory action as well as hard choices in complex situations where inevitably someone will be wronged and the best we can do is seek the least of evils. Guilt is a paradigm remainder. Other remainders - shame, remorse, regret - are discussed here in terms of their relationships to guilt, and guilt takes us back to rectificatory action.

2 Forgiveness

Forgiveness is a liberal response to those who have wronged us. Like mercy it characteristically (but not

necessarily) manifests compassion for others. It can be offered as a gift, which another may accept or reject. The questions arise when, what and whom we can forgive, what forgiveness does and implies, when we ought to forgive, when it is deserved, when it is a free choice, even when if ever it would be wrong.

We can forgive what we can resent: persons and deeds, wrongs done us and the agents of those wrongs. Hannah Arendt (1958) said we cannot forgive ourselves, that only others can forgive us. Joseph Butler (1726) presented forgiveness as a renunciation of resentment, a view that explains Arendt's observation, as we cannot resent ourselves. Although, like Butler, Arendt emphasizes the liberality of forgiveness, presenting it as letting go of the deed, she also presents it as charitable, claiming that we forgive the offence for the sake of the offender (1958: 241) (see Charity). To forgive is not simply to forget (even if we forget) nor simply to cease resenting with time. It is a choice producing a change of heart from defensive hostility at feeling disrespected or neglected to a friendlier, more pacific attitude of no longer emotionally holding against the offender what we had once resented. It is a natural response to another's sincere apology, expression of contrition, or offer to make amends or atone for an offence.

Forgiveness is of special value in personal relationships (friendships, family, work relationships), where we are most vulnerable to offence, and needs for forgiveness are probably greatest. Arendt called the power to forgive a 'redemption from the predicament of irreversibility' (1958: 237), namely past deeds. Forgiveness enables renewal of relationships but frees us to move on even if we do not renew relations.

Does forgiveness imply at least a willingness to renew relations? In the jointly authored *Mercy and Forgiveness* (1988), Jean Hampton suggests that normally it does, at least with repentant offenders, except when such renewal would be dangerous (as for a battered intimate who rightly believes the offender unregenerate), while Jeffrie Murphy regards renewal as an open question depending on many factors.

If punishment is in question, does forgiveness imply mercy? Arendt called forgiveness and punishment alternative ways of trying to put an end to what is past but also said they are not opposites (1958: 241). Hastings Rashdall (1907) found the two in tension (see Rashdall, H.). Others, including Murphy, find no incompatibility, arguing that punishment need not manifest resentment but may be required out of justice to other offenders and for deterrence to protect the innocent. Mabbott (1939) argued that duties of forgiveness and punishment fall upon different parties. If they are right, forgiveness need not imply total absolution. If it frees one only from the forgiver's resentment and not from the blame of others or the guilt that justifies punishment, it is but a limited redemption of the past's irreversibility.

If then we ought to support punishment of some forgiven offenders, we may also rightly regret that necessity. Such regret is a 'moral remainder'. This example shows, incidentally, that regret does not presuppose that the regretter acted wrongly or was even responsible for what is regretted. Such regret is primarily a sense of loss - here, loss of a permissible opportunity to engage in such otherwise natural manifestations of a forgiving attitude as remitting or not supporting punishment.

Ought we to forgive the unrepentant? Ordinarily, repentant offenders deserve forgiveness. Paradigmatically, a repentant offender apologizes and asks forgiveness, which is then granted, or a victim offers forgiveness, which is accepted and followed by repentance. Yet many find they can forgive the dead who never asked it nor even admitted wrongdoing. We can cease to resent the dead and choose to remember them for their good qualities. Forgiving the dead may be easiest. With the living, problems arise. How can we distinguish forgiving the unrepentant from condoning their offences, tacitly approving them, thus tacitly encouraging their continuation? And if an offender does repent and regenerates, what remains to forgive? Aurel Kolnai (1987) presented these questions as a paradox of forgiveness. Together they suggest that forgiveness is either inappropriate or redundant.

One way to avoid condonation is to withhold forgiveness from the unrepentant. Another may be to combine forgiveness with punishment. What remains to be forgiven even a transformed offender is the offence, for which the offender, however changed, remains responsible.

Are some deeds unforgivably heinous, however deeply repented? Atrocities of slavery and genocide are prime candidates. Survivors and descendants of survivors find the issues complex. Arendt observed that we are unable to forgive what we cannot punish nor can we punish what turns out to be unforgivable (1958: 241). Howard McGary (1992) argues that rational self-interest offers African-Americans powerful reason to move beyond resentment

regarding histories of slavery, advocating such forgiveness (in the liberal sense of letting go) for the sake of the forgiver. Simon Wiesenthal, in *Die Sonnenblume (The Sunflower)* (1969), portrays his dilemma in a concentration camp when a dying, horribly wounded German soldier Karl confessed to having participated in a mass murder of Jews and then begged forgiveness from Wiesenthal as a Jew. Wiesenthal's solution was to listen, appreciate the repentance, even not withdraw his hand when Karl took it but also finally to leave without speaking. Later, upon visiting Karl's mother (under the pretence of having been one of her son's friends), and realizing from what she told him that Karl had been truthful in his account of his early years, Wiesenthal decided not to tell her about Karl's murderous deed. By refraining from this, Wiesenthal granted Karl's wish that his mother be spared disillusionment about her son's honour. Wiesenthal circulated his meditations on these events to respected contemporaries who wrote their reflections, and he published them as a symposium with his essay. Together, they cover most ethical questions that have been asked about forgiveness.

Can forgiveness be wrong? Too ready to forgive, we may be suspected of deficiencies of the self-respect that resentment emerges to defend (see Self-respect). Forgiveness can also be misplaced, as in arrogantly or mistakenly judging that there was anything to forgive. Many of Wiesenthal's symposiasts argued that it would be presumptuous to forgive in the name of others or to forgive an offence done an entire group to which one belongs.

3 Mercy

A natural manifestation of forgiveness in some cases is to mercifully reduce or remit punishment. Mercy is compassionately withholding or mitigating a hardship or penalty that one has authority or power to inflict on another. Leniency and clemency suggest indulgence, softness, mildness. Mercy, however, can issue from strength of character.

Pardons and amnesties can exemplify mercy, although they need not. Pardons remit punishment, or the portion remaining. Ordinarily they are issued after a guilt finding, whereas amnesties preclude such findings. Mercy need not presuppose an offence, however. We can be merciful in administering requirements that can impose hardships and when we can release others from commitments that would impose unforeseen hardships.

The most difficult cases concern mercy for offenders who retributively deserve punishment (what Twambley (1975/6) calls 'the criminal court model') or who deserve to be held liable for compensation or restitution for harm to others (Twambley's 'civil court model') (see Desert and merit).

The criminal court model raises the question whether administrators of the law (or other public systems) can be merciful without being unjust. If one offender is shown mercy, every relevantly similar offender should likewise receive it. But then the reduced penalty has become an offender's right and is no longer merciful. Showing mercy to some and not others on the basis of personal inclinations is unjust favouritism. Reducing penalties for all beyond what justice requires may be unjust to the innocent who depend on deterrence for protection. Thus some philosophers hold that criminals should not be shown mercy by agents of the law (although heads of state may occasionally pardon), and others conclude that mercy for criminals is permissible only when contrary obligations (such as support of a criminal's family) require mitigation of the penalty anyhow.

A satisfactory principle setting out conditions under which offenders deserve mercy would enable administrators to discriminate among offenders without arbitrariness or favouritism. If repentant offenders deserve forgiveness, some may also (or instead) deserve mercy. The desert basis for mercy, however, is not the same. We deserve mercy when we deserve compassion even from those we have wronged or harmed, because of extraordinarily severe undeserved misfortunes in our lives, relative to the lives of our victims and others who depend for protection on enforcement of rules we violated. Such a principle helps distinguish the bases of mercy from the bases of excuse and justification, which rest on facts about the commission of the offence. Excuses remove or reduce responsibility; justifications remove or soften judgments of wrongness (see Responsibility). By contrast, a case for mercy rests primarily on what offenders have suffered, rather than simply on what they have done. It need not cite their offences, although it can, as when the crime recoils upon the perpetrator causing worse suffering than justifiable punishment would have done. The offender's undeserved misfortunes might include such things as having been victimized by yet worse crimes of others or having suffered severe natural disasters. Deserved mercy, so understood, moves towards cosmic moral compensation for inequalities in undeserved misfortune. Such deserved mercy is also a kind of equity and thus a point at which justice and charity come together.

There appears no rightful place for undeserved mercy in criminal court, no way to be merciful to those who do not deserve it without being unjust to other offenders and to the innocent who deserve protection - at least, if the laws are just. Where laws are deeply unjust, judges and juries may be lenient and heads of state issue pardons. Such compassion can respond to both injustice and undeserved misfortune.

On the civil court model (as in Shakespeare's *The Merchant of Venice* (1596/7)), mercy may be permissible or even morally required without being deserved through special misfortune. Just as one had a choice whether to bring a suit, the choices remain whether to call it off or settle. The questions are when such actions are consistent with self-respect and would not condone abuse. As with forgiveness, the offender's evident repentance and apology may obviate the problem of condonation and vindicate one's self-respect. Here mercy is a natural consequence of deserved forgiveness. Yet on the civil as on the criminal court model, offenders who have suffered extraordinary undeserved misfortune may still deserve mercy even if, as unrepentant, they do not deserve forgiveness.

Everyday situations can resemble either criminal or civil court models. Administrative positions allow us discretion in enforcing requirements on others, thus suggesting the criminal court model. Informal relations with friends, neighbours and family suggest the civil court model. In neither is mercy totally arbitrary. It is grounded in the suffering and moral responses of others, calling upon us to see ourselves in them, and it can be restricted by formal justice and by demands of self-respect.

4 Gratitude

Forgiveness and mercy, commonly motivated by compassion for wrongdoers, may evoke gratitude. Gratitude is an appreciative acknowledgement of another's goodwill (or more), often manifesting joy and accompanied normally by dispositions to reciprocate. Fred Berger (1975) argued that gratitude demonstrates that we do not value the beneficiary simply as useful for our own ends. Paradigmatically, gratitude is deserved by a benefactor from a beneficiary who was the intended recipient of the benefactor's freely bestowed (and freely accepted) goodwill. We may be grateful for gifts, favours, support and encouragement, recognition, sympathy, many things others do for us or give us beyond what they owed us or were constrained to offer (see Supererogation).

In an extended (nonmoral) sense, we can be grateful that something occurred, such as good weather, without being grateful to anyone for it. Henry Sidgwick even suggested that gratitude universalized yields the principle that good deeds ought to be rewarded (1874: 279).

Undeserved forgiveness and undeserved mercy can be virtuous, but undeserved gratitude to others is misplaced and suggests deficiencies of respect or self-respect. Gratitude is misplaced when the ostensible benefactor was only refraining from abuse or was compelled to provide the benefit or when the recipient was forced by the donor to accept it. Gratitude is misplaced for 'offers one cannot refuse', as the phrase is used regarding organized crime and likewise for a spouse's refraining from spouse battering. Gratitude for basic decencies is ordinarily misplaced. It risks offending others as well as betraying deficient self-respect.

Deserved gratitude can involve a special obligation, a 'debt' that goes beyond acknowledging appreciation and giving thanks. Richard Brandt (1974) noted that we incur obligations of two kinds, one in undertaking commitments (contracts, promises), the other in accepting benefactions. Ethical questions regarding gratitude tend to be connected with this special incurred obligation: how great the benefit must be to impose a debt, what its measure of value is, how much one should do in return, whether one can ever be done with it, even whether knowingly creating such a debt undermines desert of gratitude.

Is gratitude due one who tries but does not succeed in benefiting us? Some gratitude may be due for good intentions, but what if the would-be benefactor went to some expense or trouble? Do we measure the worth of the deed by what it cost the doer? Or by what it was worth to the recipient? Sidgwick found 'no clear accepted principle' here but noted that if either effort or benefit were great, we tend to feel strong gratitude. Aristotle said that if our relationship to the doer is based on utility, the value to the recipient is what counts, but that in true friendship, the purpose of the doer is what counts (*Nicomachean Ethics* 1163a10-24). What if the value of the effort is high but the benefit meagre, or the donor well-off and the recipient poor? Sidgwick thought that for such discrepancies reciprocating 'something between' is what 'seems to suit our moral taste' (1874: 261).

Do we owe gratitude where need compels us to accept a benefit? Such cases can create mixed feelings, gratitude undermined by humiliation. It is more considerate, as some religions teach, to give alms anonymously. Reciprocity for anonymous giving may be replaced by recipients' dispositions to do likewise for still others.

The idea of a debt of gratitude is paradoxical. Ideally, gratitude expresses gratification. Yet debts are burdensome, and paying them off a relief. Further, how can we owe for a benefit bestowed freely? How can we pay without so transforming the transaction that there is nothing for which to be grateful, and without needing to disguise the relief to avoid offending the benefactor?

These paradoxes arise from the borrower or debtor paradigm of duty or obligation. Borrowers who are granted credit occupy 'inferior' positions in relation to their creditors. Such debtors prove their reliability by paying off the debt, discharging the obligation to the creditor, terminating it. Unpayable debts of gratitude, the fear of which was expressed by Kant in his lecture 'Duties to Oneself' (1780-81), are depressing on this model, as they suggest that one can never adequately prove one's reliability.

The debtor paradigm does not cohere well with obligations of gratitude. A trustee or guardianship paradigm works better. Trustees who accept a deposit owe but are not in debt; they incur obligations but are not 'inferior' and do not have to prove their honour. It can already be an honour to be entrusted. Living up to obligations of a trustee does not terminate them. Terminating trusteeships is what others do when trustees fail to come through.

Accepting a beneficiary's goodwill is more like accepting a deposit (another's good will) than like borrowing. It can be an honour to be so 'indebted', which we express by saying we are happy to be obliged. As the anthropologist Mauss (1967) has shown, entire cultures have been built on the gift relationship, rather than on the economic relationship of buyer and seller.

Thomas Hobbes captured the basic obligation of gratitude in his 'natural law', that '[you are to] suffer him not to be the worse for you, who out of the confidence he had in you, first did you a good turn' and that one should 'endeavour, that the giver shall have no just occasion to repent him of his gift' (1642: 47). Ingratitude, a major moral failing, consists in failing to live up to this basic obligation. More than simple ungratefulness (lack of appreciation, sometimes justified), it gives the benefactor cause for regret.

Gratitude is also among the things for which gratitude is felt, which fits with the observation of Max Scheler that 'love (once it is somehow perceived) evokes a loving response' (1923: 164) (see Scheler, M.F. §7).

5 Guilt

Failure to be forgiving, merciful or even grateful can produce guilt. Unlike guilt for many offences, this guilt cannot be expiated by punishment, because forgiveness, mercy and gratitude cannot be rightfully compelled. Until we are forgiven this guilt may stay as a moral remainder.

'Guilt' is ambiguous between emotional self-punishment for having wronged others (internal guilt; guilt feelings) and the fact or finding of a transgression (objective guilt; a verdict). Internal guilt is associated with the voice of conscience, what Nietzsche (1887) called 'bad conscience' (see Nietzsche, F. §9). The debtor paradigm makes sense of guilt in both senses. Borrowers receive loans, the guilty steal; both are in debt. Borrowers need to demonstrate reliability; the guilty need to reclaim honour. Nietzsche noted that in German '*Schuld*' means both 'guilt' and 'debt'. He speculated that guilt was originally a substitute payment extracted from debtors who defaulted and that this idea was carried into punishment when legal offenders were seen as having defaulted on what they owed society (obedience).

Internal guilt is a 'moral remainder' when it continues after punishment, compensation or restitution, when no more remains to be done to set things right. Removing guilt feelings is a major point of forgiveness. Whether internal guilt is understood as the offender's introjections of victim resentment or, as Nietzsche suggested, as the offender's emotional enactments of introjected victim desires for the compensating pleasure of making the offender suffer, forgiveness naturally cancels the emotional 'debt' by communicating the victim's change of heart. It does not cancel objective verdicts of 'guilty'. Still, if accepted, it removes barriers (internal guilt and victim resentment) to renewals of relationship, barriers that might otherwise remain after punishment.

Are our guilt feelings irrational when they are not based on the belief that we culpably committed an offence?

Herbert Morris (1987) has argued that three kinds of such 'blameless guilt' are not irrational: survivor guilt (or guilt over unjust enrichment, involuntarily benefiting from injustice to others); guilt over acts of one's country or other groups to which one belongs (even though one had no effective say over those acts); and guilt for feelings and desires that one did not act upon. Guilt over unjust enrichment is what we may feel instead of gratitude for benefits made possible by injustice. Here the debtor model of obligation makes sense and suggests a way to alleviate the guilt in some cases. One may feel obliged to use the benefits to help or enrich those treated unjustly, or if that is impossible, to help or enrich their descendants.

Is internal guilt necessary? Neither guilt feelings nor the voice of conscience occur in Aristotle's ethics, although he discussed the fault of *akrasia* (acting contrary to one's better judgment) (see Akrasia). The disposition to feel guilt, as something distinct from shame, may not be learned in all cultures. Ruth Benedict (1946) characterized Japanese culture as a shame culture by contrast with guilt cultures of Europe and North America. Shame is a response to falling short of aims or standards we set ourselves, whereas guilt responds to boundary transgressions (see Piers and Singer 1971; Rawls 1971). Shame is moral when the aim or standard we fail to reach is moral; such shame injures our self-respect (see Moral sentiments §3).

Guilt differs also from remorse, regret and repentance. Repentance is that change of heart whereby a transgressor becomes contrite, remorseful. It is appropriate for major offences. Remorse is intense, often lasting, moral regret regarding our own conduct. Regret, more generally, is sorrow over a fault, offence, mistake or loss, ranging in intensity from mild to deep. It does not presuppose responsibility for what is regretted, although remorse does. More painful than guilt, remorse is emotional gnawing at oneself over one's wrongdoing. Etymologically, it means 'to bite again', suggesting continual reopening of wounds as one rehearses again and again a vivid appreciation of one's wrongdoing. By contrast, internal guilt, as emotional self-flagellation, can actually dull one's appreciation of the wrong. Remorse is often accompanied by shame, and forgiveness removes neither.

Remainders remind us that not everything can be made right, rectified. Yet they offer us a limited redemption in revealing our moral appreciation of that very fact.

CLAUDIA FALCONER CARD

References and further reading

The following works are accessible, with the exception of those marked as demanding a higher level of comprehension.

Arendt, H. (1958) *The Human Condition*, Chicago, IL: University of Chicago Press, 236-43.(Discusses forgiveness as a remedy for the irreversibility of past deeds.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. W.D. Ross, London: Oxford University Press, 1925. (Books 1108a30-35 and 1128b10-15 focus on shame; 1145a15-1152a35 deal with acting against one's better judgment; regarding gratitude, see 1123a34-1125a35 on pride and 1162a34-1163a24 on requiting services.)

Benedict, R. (1934) *The Chrysanthemum and the Sword: Patterns of Japanese Culture*, Boston, MA, and New York: Houghton Mifflin; repr. New York: New American Library, 1946.(On Japanese culture as a shame culture.)

Berger, F. (1975) 'Gratitude', *Ethics* 85: 298-309.(Classic article, presenting a nonutilitarian analysis of gratitude.)

Brandt, R. (1974) 'The Concepts of Duty and Obligation', *Mind* 73: 374-93.(Notes two kinds of incurred obligations, those from commitments, such as promises, and those from accepting benefactions.)

Butler, J. (1726) 'Upon Forgiveness of Injuries', in *Fifteen Sermons Preached at the Rolls Chapel*, London: G. Bell & Sons, 1967, Sermon IX.(Classic analysis of forgiveness as a remedy for the excesses of resentment.)

Card, C. (1972) 'On Mercy', *Philosophical Review* 81 (2): 182-207.(Offers a principle for distinguishing criminals who deserve mercy from those who do not; presents mercy as a virtue of persons rather than of institutions; argues for compatibility of charity and justice.)

Card, C. (1988) 'Gratitude and Obligation', *American Philosophical Quarterly* 25 (2): 115-27.(On debts of gratitude as paradoxical; debtor and trustee paradigms of obligation; misplaced gratitude; gratitude in relation to friendship; and gratitude between unequals; many references.)

Emerson, R.W. (1883) 'Gifts', in *Essays*, Boston, MA: Houghton Mifflin; repr. in *The Essays of Ralph Waldo Emerson*, Harvard, MA: Belknap, 1987, 311-4.(Short classic discussing ideal gifts and pitfalls of giving; helpful for thinking about gratitude.)

Feinberg, J. (1970) 'Justice and Personal Deserts' in *Doing and Deserving: Essays in the Theory of*

Responsibility, Princeton, NJ: Princeton University Press, 55-94.(Distinguishes between deserts and rights; moderately difficult but rewarding.)

Hobbes, T. (1642) 'Of the Other Laws of Nature', *De cive (The Citizen)*, New York: Appleton, Century, Crofts, 1949, 43-59.(Includes law of gratitude as third law of nature; often cited as a basis of political obligation; moderately difficult.)

Kant, I. (1780-81) 'Duties to Oneself', 'Friendship' in *Lectures on Ethics*, trans. L. Infield, New York: Harper & Row, 1963, 116-261, 200-9.(The first essay argues that accepting unnecessary favours violates a duty to oneself, the second, that ideal friends do not do material favours.)

Kant, I. (1797) 'On Duties of Love to Other Men', in *Metaphysische Anfangsgründe der Tugendlehre*, trans. M.J. Gregor, *The Doctrine of Virtue: Part II of The Metaphysic of Morals*, New York: Harper & Row, 1964, 115-38.(Presents gratitude as a sacred duty that can never be completely discharged; difficult.)

Kolnai, A. (1973/4) 'Forgiveness', *Proceedings of the Aristotelian Society*, new series, 74: 91-106; repr. in *Ethics, Value, and Reality: Selected Papers*, Indianapolis, IN: Hackett Publishing Company, 1987, 211-24.(Discusses paradoxes of forgiveness; difficult.)

Mabbott, J.D. (1939) 'Punishment', *Mind* 48: 152-67.(Classic defence of retributive theory of punishment; finds forgiveness compatible with punishment.)

Mauss, M. (1923/4) 'Essai sur le don, forme et raison de l'échange dans les sociétés archaïques', *L'année sociologique* 2 (1): 30-186; trans. I. Cunnison, *The Gift: Forms and Functions of Exchange in Archaic Societies*, New York: W.W. Norton, 1967.(Classic anthropological study of societies structured around giving rather than around buying and selling; moderately difficult.)

McConnell, T. (1993) *Gratitude*, Philadelphia, PA: Temple.(Comprehensive examination and survey of literature; chapters on political obligation and filial obligations.)

McGary, H. (1992) 'Forgiveness and Slavery', in H. McGary and B. Lawson, *Between Slavery and Freedom: Philosophy and American Slavery*, Bloomington, IN: Indiana University Press, 90-112.(Develops a case for forgiveness for the forgiver's own sake and independently of repentance by perpetrators.)

Moore, K.D. (1989) *Pardons: Justice, Mercy, and the Public Interest*, New York: Oxford University Press.(Comprehensive and illuminating discussion, rich with examples.)

Morris, H. (ed.) (1971) *Guilt and Shame*, Belmont, CA: Wadsworth.(Selections by Dostoevskii, Jaspers, Freud, Buber, Fingarette, Nietzsche, Rawls, Piers and Singer, Erikson, and Lynd; excellent place to begin.)

Morris, H. (1987) 'Nonmoral Guilt', in F. Schoeman (ed.) *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, Cambridge: Cambridge University Press.(Discusses survivor guilt, guilt feelings for what other members of one's country or other group have done, and guilt feelings for emotions not acted on.)

Murphy, J.G. and Hampton, J. (1988) *Mercy and Forgiveness*, Cambridge: Cambridge University Press.(Alternate essays by each author; topics include retribution, forgiveness, Christianity, resentment, hatred, mercy, and legal justice.)

Nietzsche, F. (1887) '"Guilt", "Bad Conscience", and the Like', in *Zur Genealogie der Moral*, trans. W. Kaufmann and R.J. Hollingdale, *On the Genealogy of Morals*, New York: Vintage Books, 1969.(Discusses punishment, responsibility, justice, law and Christian guilt; difficult but rewarding.)

Piers, G. and Singer, M.B. (1953) *Shame and Guilt: A Psychoanalytic and a Cultural Study*, Springfield, IL: Charles C. Thomas; repr. New York: W.W. Norton, 1971.(A classic by a psychologist and an anthropologist contrasting shame as a response to failure of achievement with guilt as a response to transgression, and contrasting shame cultures with guilt cultures. Very accessible.)

Rashdall, H. (1907) *The Theory of Good and Evil: A Treatise on Moral Philosophy*, Oxford: Clarendon Press; 2nd edn, London: Humphrey Milford, 1924, vol. 1, 306-12.(On tensions between forgiveness and punishment.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, ch. 8.(Discusses guilt and shame; moderately difficult.)

Scheler, M.F. (1923) *Wesen und Formen der Sympathie*, Bonn: Friedrich Cohen; trans. P. Heath, *The Nature of Sympathy*, London, Routledge & Kegan Paul, 1954; repr. Hamden, CT: Archon Books, 1970.(A classic phenomenological examination of sympathy and related emotions; helpful for thinking about forgiveness, mercy and gratitude; difficult.)

Shakespeare, W. (1596/7) *The Merchant of Venice*, New York: Bantam, 1988, act IV, scene 1.(See Portia's speech, beginning 'The quality of mercy is not strained', on the elective nature of showing mercy.)

Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan; 7th edn, 1907; 7th edn repr. Chicago, IL: University of Chicago Press, 259-61, 279.(Discusses common sense views on the ethics of gratitude; presents reward as gratitude universalized.)

Smart, A. (1968) ‘Mercy’, *Philosophy* 43 (166): 345-59.(Classic discussion of mercy on the criminal court model; interesting examples; good place to begin.)

Twambley, P. (1975/6) ‘Mercy and Forgiveness’, *Analysis* 36 (2): 84-90.(Argues against mercy in the criminal court model and presents the civil court model as an appropriate forum; moderately difficult.)

Wiesenthal, S. (1969) *Die Sonnenblume*, Paris: Opera Mundi; trans. H.A. Pichler and C.P. Pinto, *The Sunflower*, New York: Schocken, 1976.(Written with a symposium of many authors, this work comprises memoirs and meditation on the lived dilemma of forgiveness in a death camp; excellent place to begin.)

Williams, B. (1965) ‘Ethical Consistency’, *Proceedings of the Aristotelian Society*, supplementary vol. 39; repr. in *Problems of the Self: Philosophical Papers 1956-72*, Cambridge: Cambridge University Press, 1973. (Introduces the concept of ‘moral remainders’; difficult.)

Recursion-theoretic hierarchies

In mathematics, a hierarchy is a 'bottom up' system classifying entities of some particular sort, a system defined inductively, starting with a 'basic' class of such entities, with further ('higher') classes of such entities defined in terms of previously defined ('lower') classes. Such a classification reflects complexity in some respect, one entity being less complex than another if it appears 'earlier' ('lower') then that other. Many of the hierarchies studied by logicians construe complexity as complexity of definition, placing such hierarchies within the purview of model theory; but even such notions of complexity are closely tied to species of computational complexity, placing them also in the purview of recursion theory.

1 Hierarchies of formulas

Consider a formal language *L* with quantifiers $\forall$ and $\exists$ (see Formal languages and systems). Let *B* be a class of formulas of *L* such that by adding a $\forall$- or $\exists$-prefix to any member of *B* we get a non-member. We define a hierarchy of formulas of *L* that treats members of *B* as 'basic'. Set $\prod_0^B = \sum_0^B = B$. Suppose that classes $\prod_n^B$ and $\sum_n^B$ have been defined. Let $\prod_{n+1}^B$ = the class of all formulas of *L* of the form $\forall v_1 \ldots \forall v_m \varphi$ for $m \geq 1$, distinct variables $v_1, \ldots, v_m$ and $\varphi \in \sum_n^B$; let $\sum_{n+1}^B$ = the class of all formulas of *L* of the form $\exists v_1 \ldots \exists v_m \varphi$ for $m \geq 1$, distinct variables $v_1, \ldots, v_m$, and $\varphi \in \prod_n^B$. This classifies all formulas of *L* that consist of a string of quantifier-prefixes attached to a member of *B*.

Motivation The complexity of a minimally complex formula (complexity understood as above) defining a given relation may be identified with the complexity of that relation. We can now 'transfer' this hierarchy to a hierarchy of the relations defined by such formulas over an appropriate structure. This was first done by S. Kleene (1943) with first-order formulas over the standard model of arithmetic to get the arithmetical hierarchy.

2 The arithmetical hierarchy

Let L_0 = the first-order language based on function-constants for successor, addition and multiplication, an individual-constant of 0, and a two-place predicate for $<$; variables range over ω. For $\varphi(x_0, \ldots, x_m)$ a formula of L_0 with free-variables among those indicated, φ defines $\{< a_0, \ldots, a_m > \in \omega^m : \varphi(a_0, \ldots, a_m)\}$. A relation *R* is arithmetic iff *R* is defined by some formula of L_0.

Let a formula φ of L_0 be bounded iff each occurrence of $\forall$ in φ is in a context of the form $\forall v$ (if $v < t$ then …) and each occurrence of $\exists$ in φ is in a context of the form $\exists v$ ($v < t$ and …), for *t* a term and *v* a variable; take *B* = the set of bounded formulae, $\prod_n^0 = \prod_n^B, \sum_n^0 = \sum_n^B$. For any *m*, $n \in \omega, m \geq 1$, and any $R \subseteq \omega^m$, we will also let: $R \in \sum_n^0$ iff *R* is defined by some $\sum_n^0$ formula of L_0; $R \in \prod_n^0$ iff *R* is defined by some $\prod_n^0$ formula; $R \in \Delta_n^0$ iff $R \in \sum_n^0 \cap \prod_n^0$. Context should make it clear whether $\sum_n^0$ and $\prod_n^0$ is a set of formulas or a set of relations. This is the arithmetical hierarchy of relations; since any defining formula has an equivalent in prenex form, it classifies all arithmetical relations. It looks like this (with rightward = up):

$$\begin{array}{llllll} & \subset \Sigma_1^0 \subset & \subset \Sigma_2^0 \subset & \subset \Sigma_3^0 \subset & \\ \Sigma_0^0 = \Pi_0^0 = \Delta_0^0 \subset \Delta_1^0 & & \Delta_2^0 & \Delta_3^0 & \Delta_4^0 \cdots \\ & \subset \Pi_1^0 \subset & \subset \Pi_2^0 \subset & \subset \Pi_3^0 \subset & \end{array}$$

Any tuple of natural numbers can be effectively coded by a single natural number; so we lose nothing by restricting the arithmetical hierarchy to reals. The 'arithmetical hierarchy theorem' states that each class contains infinitely many 'new' reals (that is, for each new $n \in \omega \sum_{n+1}^0 - \prod_{n+1}^0$, $\prod_{n+1}^0 - \sum_{n+1}^0$ and $\Delta_{n+1}^0 - (\sum_n^0 \cup \prod_n^0)$ are all non-empty). The definitional complexity of an arithmetic relation is indicated by the class in which it 'first' appears.

This hierarchy is related to classification by recursion-theoretic properties. Most significantly: Δ_1^0 = the class of recursive relations; $\sum_1^0$ = the class of r.e. relations: Δ_2^0 = the class of relations recursive in Post's set *K* (i.e. those of T-degree $\leq 0'$). H. Putnam and Y. Ershov independently proved that Δ_2^0 is also the class of 'trial-and-error' relations: *R* is trial-and-error iff there is a recursive $m+1$- place function *f* such that given any

$\vec{a} =< a_0, \ldots, a_m > f$ guesses forever as to whether $\vec{a} \in R(f(\vec{a}, k) = \text{the kth guess})$, and eventually *f* guesses correctly for ever after. In fact, they found and studied a fine-grained hierarchy within Δ^0_2 (see Putnam 1965).

These facts generalize 'up' the hierarchy. For a real *X*, set $X^{(n+1)}$ = the jump of $X^{(n)}$, with $X^{(0)} = X$. We then have: $X \in \Delta^0_{n+1}$ iff $X \leq_T \{\}^{(n)}$ for $\{\}$ = the empty set; furthermore, $X \in \sum^0_{n+1}$ iff *X* is r.e. relative to $\{\}^{(n)}$. (H. Rogers credits these basic results to Kleene and Post; see Rogers 1967.)

For $R \subseteq \omega^m$ and real *Y*: $R \leq_1 Y$ iff there is a $1-1$ recursive function $f{:}\,\omega^m \rightarrow \omega$ so that for any $\vec{a} \in \omega^m{:}\,\vec{a} \in R$ iff $f(\vec{a}) \in Y$. 1-reducibility (that is, $\leq_1$) is the strongest of the recursion-theoretic reducibilities. For $\mathcal{C}$ a class of relations on ω: *Y* is $\mathcal{C}$-complete iff $Y \in \mathcal{C}$ and for any $R \in \mathcal{C}, \mathcal{R} \leq_\infty \mathcal{Y}$. Fact: $\{\}^{(n)}$ is $\sum^0_n$-complete and $\omega - \{\}^{(n)}$ is $\prod^0_n$-complete for each $n \leq 1$. Let $T_{\sum_n}$ and $T_{\prod_n}$ be the sets of all and only the true $\sum^0_n$ and $\prod^0_n$ sentences respectively. These sets have $\sum^0_n$ and $\prod^0_n$ definitions respectively. (There is no paradox here: 'I am not a true $\sum^0_n$ sentence' requires negation of the truth predicate for $\sum^0_n$, making it $\prod^0_n$ rather that $\sum^0_n$, and thus trivially true!) In fact, these two sets are also $\sum^0_n$-complete and $\prod^0_n$-complete respectively, and thus (by a theorem of J. Myhill) are recursively isomorphic to $\{\}^{(n)}$ and to $\omega - \{\}^{(n)}$ respectively; this expresses the sense in which the jump is the recursion-theoretic analog of $\exists$ in arithmetic, and connects recursion-theory to truth-theory (see Tarski's definition of truth; Semantic paradoxes and theories of truth).

Let us contrast the arithmetical hierarchy with classification of reals by T-degree. Each real has a T-degree; the arithmetical hierarchy only classifies arithmetic reals. In one respect T-degrees are finer than the 'compartments' of the arithmetical hierarchy (that is, $\sum^0_n - \Delta^0_n$ and $\prod^0_n - \Delta^0_n$ for $n > 0$, $\Delta^0_{n+1} - (\sum^0_n \cup \prod^0_n)$ for $n \geq 0$): for example, there are countably many T-degrees with members from $\sum^0_1 - \Delta^0_1$ (see Post's problem in Turing reducibility and Turing degrees). On the other hand, for a real $X \in \sum^0_1$ with $\deg(X) = a, \omega - X \in a$; but if $X \notin \Delta^0_1$ then $\omega - X \notin \sum^0_1$; so except for the T-degree 0 no T-degree is a subset of $\sum^0_1$; similarly for the other compartments of the arithmetical hierarchy. So in another respect these classifications are orthogonal.

3 The analytical hierarchy

Expand L_0 to L_1 by introducing infinitely many predicate- (that is, second-order) variables, allowing $\forall$ and $\exists$ to bind them; without loss of generality let us require them all to be one-place. A relation on ω is analytical iff it is defined by a formula (containing no free predicate-variables), taking each predicate-variable to range over the class of all reals, with individual-variables ranging over ω. ('Analytical' alludes to quantification over reals, as in 'Analysis' in its mathematical sense.)

Now take *B* = the class of formula of L_1 containing no bound predicate-variables and no free individual-variables (though by coding tricks we could restrict *B* further). Set $\sum^1_0 = \sum^B_0, \prod^1_n = \prod^B_0$. For any *m*, $n \in \omega, m \geq 1$, and any $R \subseteq \omega^m$, let: $R \in \sum^1_n$ iff *R* is defined by some $\sum^1_n$ formula of L_1 (with predicate variables ranging as above); $R \in \prod^1_n$ iff *R* is defined by some $\prod^1_n$ formula: $R \in \Delta^1_n$ iff $R \in \sum^1_n \cap \prod^1_n$; again, context disambiguates this double-usage. This is the analytical hierarchy. Using prenexing, it classifies all analytical relations. A picture of this hierarchy with regard to subsethood would look just like that of the arithmetic hierarchy, with superscripts changed from '0' to '1'. Again, each class contains infinitely many new reals, there are $\sum^1_n$-complete and $\prod^1_n$-complete reals, and among those are the truth-set for $\sum^1_n$ and $\prod^1_n$ respectively.

4 Two more hierarchies of relation on ω

Set RA_0 = the set of arithmetic relations on ω. Suppose RA_α has been defined, for α an ordinal. A formula $\varphi(x_0, \ldots, x_m)$ (all free variables first-order and among those shown) defines *R* over RA_α iff for any $\vec{a} \in \omega^m{:}\,\vec{a} \in R$ iff $\varphi(a_0, \ldots, a_m)$, with predicate-variable ranging over RA_α; note this change from §3. Let $RA_{\alpha+1}$ = the set of relations defined by such formulas over RA_α. For a limit ordinal λ: RA_λ = the union of all RA_α for $\alpha < \lambda$. This sequence is the ramified-analytical (hereafter *RA*) hierarchy: 'ramified' because the definitions used allow predicate-variables to range only over entities that were 'previously' defined. Note: if $\alpha \leq \beta$ then $RA_\alpha \subseteq RA_\beta$. This hierarchy is to reals what Gödel's constructible hierarchy is to arbitrary sets (see Constructible universe); indeed work by R. Jensen shows that the latter can be construed as a recursion-theoretic

hierarchy (see Devlin 1973). Note one difference between these hierarchies: the *RA* hierarchy 'peters out' (at the countable ordinal β_0 = least ordinal β such that $RA_{\beta+1} = RA_\beta$). In a natural sense the constructible relations on ω appearing the β_0-stage of the constructible hierarchy extend the *RA* hierarchy; see Odifreddi (1989) for details.

The *RA* hierarchy can be further refined by breaking up each stage $RA_{\alpha+1} - RA_\alpha$ in terms of the quantifier-prefix of the defining formulae (in prenex-form, of course): $\sum^0_{\omega\cdot\alpha+n}$ = the set of relations defined over RA_α by $\sum^1_n$ formulas of L_1; now $\prod^0_{\omega\cdot\alpha+n}$ = the set of relations defined over RA_α by $\prod^1_n$ formulas; $\Delta^0_{\omega\cdot\alpha+n} = \sum^0_{\omega\cdot\alpha+n} \cap \prod^0_{\omega\cdot\alpha+n}$. For $n \geq 1$ we again have a hierarchy theorem: each such class contains infinitely many new reals. There are connections between the *RA* hierarchy, the jump operation, and truth in arithmetic.

An ordinal is recursive if it is order-isomorphic to some recursive well-ordering of natural numbers. There is a countable ordinal ω_1^{CK} = the least non-recursive ordinal ('*CK*' for Church and Kleene, who first studied it). $\omega_1^{CK} < \beta_0$, and if we cut off the *RA* hierarchy at ω_1^{CK} we get the hyperarithmetical (hereafter HA) hierarchy, introduced independently by M. Davis, A. Mostowski and S. Kleene (in Davis 1950, Mostowski 1951, Kleene 1955). For $\alpha = \omega_1^{CK}$, RA_α = the class of hyperarithmetical relations.

The so-called Souslin/Kleene theorem connects the HA hierarchy to the analytical hierarchy: Δ^1_1 = the class of the hyperarithmetic relations; so the HA hierarchy is a classification of the Δ^1_1 relations. Δ^1_1 can be thought of as the class of relations on ω that would be computable by a super-mind capable of completing an effectively specified ω-sequence of tasks, in which case Δ^1_1 is to Δ^0_1 as $\prod^1_1$ is to $\sum^0_1$; this analogy was one of the starting points for generalized recursion theory. The class of ramified-analytical sets (= RA_α for $\alpha = \beta_0$) is a proper subset of Δ^1_2.

5 Hierarchies of relations between reals

So far we have only considered relations on ω defined by formulae of L_1; but a formula of L_1 whose only free-variables are one-place predicate-variables defines a relation between reals, that is, on Power(ω) = the class of reals. Such a relation is analytical iff it is defined in this way, allowing bound predicate-variables to range over Power(ω); it is arithmetical iff it is defined by such a formula with no bound predicate-variables (which formula may be classified like one of L_0). Thus we have an arithmetical and analytical hierarchy of such relations; for $i \in \{0,1\}, R \subseteq \mathrm{Power}(\omega)^{\mathrm{m}}\colon \mathrm{R} \in \sum^{\mathrm{i}}_{\mathrm{n}}$ iff *R* is so defined by a $\sum^i_n$ formula of L_1; $R \in \prod^i_n$ iff *R* is so defined by a $\prod^i_n$ formula.

Similarly, if we use 'ramified' definitions we obtain the *HA* hierarchy, and, going further, the *RA* hierarchy, of relations between reals. In the 1950s J. Addison (in his 1954 doctoral dissertation and in Addison (1959)) pointed out that the *HA* and analytical hierarchies of such relations where the 'effective analogues' of the so-called Borel and projective hierarchies, studied by French and Russian analysts (for example, E. Borel, H. Lesbesque, M. Suslin and N. Lusin) in the early twentieth century. Their work, usually labelled 'descriptive set-theory', was thought of as a part of topology; apparently the logicians (except for Mostowski) had worked in complete ignorance of it. Contemporary descriptive set-theory emerged from this cross-pollination as a hybrid of recursion theory and topology.

6 Other hierarchies

There is at least one interesting hierarchy within the class of recursive relations, the so-called Grzegorczyk hierarchy. Even further 'low-down' (within the class of primitive-recursive relations) hierarchies have been proposed by computer-scientists, though here as this writing (1993) ignorance outstrips knowledge (see Complexity, computational). The study of such animals is usually called 'subrecursion theory'.

All hierarchies discussed here classify relation on ω or on Power (ω). Practitioners of generalized recursion theory have generalized some of these hierarchies to classify relations on other classes of mathematical objects.

See also: Logical and mathematical terms, glossary of

HAROLD HODES

References and further reading

Addison, J. (1959) 'Separation Principles in the Hierarchies of Classical and Effective Descriptive Set Theory',

Fundamenta Mathematica 46: 123-135.(Apparently the first bridge between recursion theory and descriptive set theory.)

Davis, M. (1950) *On the Theory of Recursive Unsolvability*, Ph.D. thesis, Princeton University.(The first definition of the *HA* hierarchy.)

Devlin, K. (1973) 'Aspects of Constructibility', *Lecture Notes in Mathematics* no. 354, Berlin: Springer Verlag. (For the Jensen hierarchy, in effect Gödel's Constructible hierarchy of sets viewed recursion-theoretically.)

Hinman, P. (1978) *Recursion-Theoretic Hierarchies*, Berlin: Springer-Verlag.(A good text introducing a lot of this subject to the novice.)

Kleene, S. (1943) 'Recursive Predicates and Quantifiers', *Translations of the American Mathematical Society* 53: 41-73; repr. in M. Davis (ed.) *The Undecidable*, Raven Press, 1965, 255-287.(The first work on the arithmetical hierarchy.)

Kleene, S. (1955) 'Hierarchies of Number-theoretic Predicates', *Bulletin of the American Mathematical Society* 61: 193-213.(Kleene's first work on the HA hierarchy.)

Mosschovakis, Y. (1980) *Descriptive Set Theory*, Amsterdam: North-Holland Publishing Company.(An excellent text, connecting descriptive set theory and recursion theory.)

Mostowski, A. (1951) ' A Classification of Logical Systems, *Studia Philosophia* 4: 237- 274.(Another first paper on the HA hierarchy.)

Odifreddi, P. (1989) *Classical Recursion Theory*, Amsterdam: North-Holland Publishing. (An excellent introductory text on classical recursion theory as of 1989, with a lot on the arithmetical and analytical hierarchies, and some on the others.)

Putnam, H. (1965) 'Trial and Error Predicates and the Solution to a Problem of Mostowski', *Journal of Symbolic Logic* 30: 49-57.(A hierarchy within Δ^0_2.)

Rogers, H. (1967) *The Theory of Recursive Functions and Effective Computability*, New York: McGraw Hill Inc. (The original textbook on recursion theory, a good complement to Odifreddi (1989), with excellent exercises.)

Reduction, problems of

Reduction is a procedure whereby a given domain of items (for example, objects, properties, concepts, laws, facts, theories, languages, and so on) is shown to be either absorbable into, or dispensable in favour of, another domain. When this happens, the one domain is said to be 'reduced' to the other. For example, it has been claimed that numbers can be reduced to sets (and hence number theory to set theory), that chemical properties like solubility in water or valence have been reduced to properties of molecules and atoms, and that laws of optics are reducible to principles of electromagnetic theory. When one speaks of 'reductionism', one has in mind a specific claim to the effect that a particular domain (for example, the mental) is reducible to another (for example, the biological, the computational). The expression is sometimes used to refer to a global thesis to the effect that all the special sciences, for example chemistry, biology, psychology, are reducible ultimately to fundamental physics. Such a view is also known as the doctrine of the 'unity of science'.

1 Reduction by derivation and definition

Unity and simplicity are often touted as important virtues achieved through reduction (see Unity of science). By reducing one domain to another, we show the reduced domain to be either part of the second or eliminable in favour of it. Depending on the nature of the entities reduced, reduction will, therefore, promote ontological or conceptual economy and unity. When numbers have been reduced to sets, numbers no longer need be countenanced over and above sets; numerical concepts can be explained in terms of notions involving only sets; and laws about numbers follow from principles about sets. There will often be an explanatory gain as well: when gas laws are reduced to principles of statistical mechanics, via the kinetic theory of gases, we have an explanation of why these gas laws hold to the extent that they do - why the pressure, temperature and volume of gases behave (roughly) in accordance with the gas laws. Furthermore, reduction is sometimes thought of as a way of vindicating or grounding a possibly suspect domain of entities; if number theory is successfully reduced to logic, as Frege claimed, that would show number theory to be as firm and well-grounded as laws of logic. If the mental realm is reducible to the physical-biological domain, this would remove whatever problems and doubts might becloud the scientific status of the mental. Such are the thoughts and aspirations that inspire and sustain reductive projects and reductionisms.

There are apparently simple cases of reduction in which logical-mathematical derivation alone suffices for reduction. When Galileo's law of free fall or Kepler's laws of planetary motion are derived from the principles of Newtonian mechanics in conjunction with applicable force laws (see Mechanics, classical §2), they are reductively absorbed into a more general and comprehensive theoretical framework.

According to logical behaviourism, mental expressions are definable - that is, given synonymous translations - in terms of expressions referring to actual or possible behaviour ('behaviour dispositions'). Logical behaviourism, therefore, exemplifies an attempt to reduce the mental to the physical-behavioural through definition (see Behaviourism, analytic). Thus, one way to reduce a domain of expressions, or concepts, is to provide each expression with a definition couched solely in the expressions of the base domain. On the assumption that synonymous expressions refer to the same entities, a definitional reduction would also accomplish an ontological reduction: entities referred to by the first group of expressions have been shown to be among those referred to by the second group.

Reduction becomes more complex for domains that do not yield to direct derivational or definitional relationships, as in the case of reductions involving theories each with its own distinctive vocabulary, for example the reduction of thermodynamics to statistical mechanics.

2 The Nagel model of reduction

Ernest Nagel (1961) articulated a model of reduction for scientific theories that has served as the principal reference point in discussions of reduction (see Theories, scientific). The guiding idea of this model remains derivation, but Nagel saw that to derive laws of a theory from another which does not share the same vocabulary, certain 'bridge principles' must be assumed as additional premises. Consider two theories, T_1 and T_2, each with its distinctive vocabulary, V_1 and V_2, where T_2 is the candidate for reduction and T_1 the reduction base.

Theories in this context are construed as sets of laws. Nagel's model requires that there be connecting laws, standardly called 'bridge laws', correlating terms of V_2 with terms of V_1. What form do such laws take and how many are needed for reduction? The simple answer is that they must be available in sufficient numbers and be sufficiently powerful to enable the derivation of T_2-laws from T_1-laws. But that depends on the strengths of the specific theories involved. The only general requirement worth considering, therefore, is the following:

The condition of connectibility: For each primitive *n*-place predicate *F* of V_2, there is a predicate *G* of V_1, such that for any $x_1, \ldots, x_n$ (in the domain of entities covered by the two theories), it is a law that $F(x_1, \ldots, x_n)$ if and only if $G(x_1, \ldots, x_n)$.

Thus, each predicate of the theory to be reduced must be connected, via a *biconditional law*, with a nomologically coextensive predicate of the reducer. These bridge laws in effect licence the rewriting of every T_2-law as a T_1-statement; in general, they enable the translation of T_2-statements into the language of T_1. Thus, the only difference between Nagel reduction and definitional reduction is that in the former the translation is underwritten by empirical laws whereas in the latter it is based on meaning equivalences.

There is a sense in which the condition of connectibility alone guarantees reduction: if any T_1-rewrite of a T_2-law is not derivable from the laws of T_1 as they stand, simply add it to T_1 as a new law (assuming it, as well as the original T_1-laws, to be true). This would extend the base theory, but not its vocabulary; and the extension would be warranted since the base theory was incomplete in failing to capture a true law statable in its vocabulary: namely, the T_1-rewrite of the T_2-law.

It is evident that when a theory has been derivationally reduced, it is conserved as part of the base theory; its laws are shown to be 'derivative laws' of a more comprehensive theory. Moreover, the condition of connectibility ensures that the properties posited by the reduced theory find their nomic equivalents in the reduction base, with which they may ultimately be identified when a successful reduction becomes entrenched. Thus, Nagel reduction is a species of 'conservative' (or 'retentive') reduction, which conserves, and can legitimize, that which is reduced; it contrasts with 'eliminative' reduction which dispenses with what has been reduced.

3 Emendations to the Nagel model: the Kemeny-Oppenheim model

Nagel's model has been criticized on various grounds, of which the following three are perhaps the most important. First, biconditional bridge laws are too weak, and must be strengthened into identities if they are to yield genuine reductions (Sklar 1967; Causey 1977). We need, it is argued, the identity 'temperature = mean molecular kinetic energy', not just a law that merely affirms covariance of the two magnitudes. As long as the reduction falls short of identifying them, there would be temperatures as properties of physical systems 'over and above' their microstructural properties. Moreover such correlations ('nomological danglers') cry out for an explanation: why does temperature covary in just this way with mean molecular kinetic energy? By identifying them we provide a short and conclusive answer: because they are in fact one and the same.

Second, the reduced theory is often only approximately true, and its laws are derivable only under special simplifying assumptions (which, strictly speaking, are false). For example, in deriving the gas laws in kinetic theory of gases, one has to make various assumptions, such as that collisions between molecules are perfectly elastic, that molecules are point masses, and so on. And the laws hold only approximately and within a fairly narrow range of conditions. Third, it is often pointed out that the condition of connectibility, which requires biconditional bridge laws, is unrealistic and can seldom be satisfied. Thus, temperature cannot, it is claimed, be uniformly correlated, or identified, with a single micro-based property; it may be mean kinetic energy of molecules for gases, but something else for solids or in vacuums.

As a response to the second criticism, it has been suggested (Schaffner 1967; Churchland 1986) that the reduction of T_2 to T_1 should require only that a *corrected version* of T_2 (or its 'image' in T_1), rather than T_2 itself, be derivable from T_1. We may also explicitly allow special limiting assumptions of the sort mentioned earlier. But to what extent can we correct or revise a theory without turning it into another theory? Obviously, the required corrections must be such as to yield the necessary bridge laws or identities, but there is no guarantee that any reasonable amount of tinkering will suffice to accomplish this. This takes us back to the last of the objections mentioned above.

Kemeny and Oppenheim (1956) proposed a model of reduction that departs extensively from the Nagel model, giving us a broader conception of reduction that does not require connecting principles between the theories involved. The gist of the new model can be stated thus: T_2 is *Kemeny-Oppenheim reduced* to T_1 just in case all observational data explainable by T_2 are explainable by T_1. This definition can, and should, be understood in a way that does not presuppose a context-independent theory-observation distinction or any special conception of explanation. The core idea here is that anything, be it an observation or a low-level empirical generalization or law, that the reduced theory can explain, or predict, should be explainable and predictable by the base theory. The base theory, then, is at least as powerful, as an explanatory and predictive instrument, as the theory being reduced, making the latter at best otiose. This means that the reduced theory is ripe for elimination - unless, that is, it is also Nagel-reducible, and hence conserved as a subtheory of the reducer. Notice that any case of Nagel reduction is also a case of Kemeny-Oppenheim reduction (at least, on the standard deductive-nomological account of explanation; see Explanation §2). Unlike the Nagel model, the Kemeny-Oppenheim model requires no direct relationship between the reduced theory and its reducer; in particular, it involves no bridge laws of any form. In fact, the reduced theory may be false and the reducer true. However, it assumes that the two theories concern the same domain of phenomena (at least, that the domain of the reduced theory is included in that of the reducer). In any case, the Kemeny-Oppenheim model allows the replacement, or elimination, of the reduced theory, and hence can serve as a model of 'eliminative' reduction.

4 Microreduction and microdeterminism

A pervasive trend in modern science has been to explain macrophenomena in terms of their microstructures, and reduce theories about the former to theories about the latter. Examples abound in the physical and biological sciences. The reduction of thermodynamics to statistical mechanics (see Thermodynamics), the reduction of optics to electromagnetic theory (see Optics §1), the successes in solid-state physics and molecular biology (see Genetics; Molecular biology; Sociobiology), and a host of other examples appear to attest to the fruitfulness of microreduction as a research strategy (Oppenheim and Putnam 1958).

Both the Nagel and the Kemeny-Oppenheim models can be applied to microreduction. What is needed is the idea that one theory is a *microtheory* in relation to another. The rough idea is that the microtheory deals with objects that are*proper parts* of the objects in the domain of the macro-theory. More specifically, the domain of the microtheory will include objects that are parts of the objects in the domain of the macrotheory; in addition it will include aggregates of these micro-objects, and aggregates of aggregates, and so on. And the objects of the macrotheory are identified with certain complex aggregates in this domain. Moreover, the microtheory has a set of properties and relations characterizing its basic micro-objects, and will generate complex properties for aggregates of these objects from the basic properties and relations. These micro-based properties (for example, mean molecular kinetic energy of a gas) of aggregative structures of micro-objects are important for reduction: in Nagel reductions, it is these micro-based properties that will be correlated (or identified) by bridge laws with the properties of the macrotheory; in Kemeny-Oppenheim reductions, they will be needed in providing explanations of the data explained by the macrotheory.

The metaphysical underpinning of microreduction is the principle of *microdeterminism* or *mereological supervenience*: properties of a whole are wholly determined by, or supervenient on, the properties and relations characterizing its proper parts. No matter how complex an object (say, a human being, a planetary system) may be, if you put together qualitatively indistinguishable parts in the same structural relationship, you will get an exact duplicate of that object (barring basic physical indeterminacies). It is the task of science, then, to identify a physically significant decomposition of a whole into parts and develop a microtheory about these parts that will explain why the whole has the properties it has.

5 Multiple (or variable) realization and reduction

One influential objection often deployed against the claim that a given theory is reducible to another is 'the multiple realization argument'. This argument was initially developed as an objection against mind-body reduction, but has been used to argue for the impossibility of reduction almost everywhere (Fodor 1974). Consider a higher level property, such as pain: pain as a psychological state is 'realized' by diverse physical-biological processes in widely divergent organisms and structures. The neural realizer of pain in humans probably has little in common with its realizer in octopuses; nor can we a priori exclude the nomological possibility of pain in inorganic

electro-mechanical systems. This means that there is no single physical-biological state which could be correlated, or identified, with pain in a biconditional bridge law, and this defeats Nagel reduction of any theory of pain to an underlying physical-biological theory, as well as reductive identification of pain with a physical-biological state. The same argument is often applied to biological concepts, such as ‘heart’ and ‘digestive system’, functionally defined physical concepts like ‘carburettor’ and ‘thermometer’, and dispositional concepts like ‘transparency’ and ‘water solubility’.

Notice that the multiple realization argument does not touch Kemeny-Oppenheim reduction in general, but works only against Nagel reduction which, as we saw, aims to conserve the reduced higher level properties. In any case, one reply that is usually rejected is this: why not reduce pain to a *disjunction* of its diverse physical realizers? But given the kind of extreme diversity involved, such disjunctive states are unlikely to be projectible nomic properties and there cannot, it seems, be a unitary theory dealing with them (Kim 1992). Another, more plausible, way of dealing with the phenomenon of multiple realization is to lower our reductive aspirations: the argument perhaps shows that *uniform* or *global reduction* of pain is not feasible, but not that human pain, octopus pain and Martian pain cannot each be *locally reduced* to human physiology, octopus physiology and Martian electrochemistry, respectively (Kim 1992). Thus, multiple realization is consistent with (in fact, it arguably entails) the local reducibility of higher level sciences: for example, human psychology to human neurobiology, octopus psychology to octopus neurobiology, and so on. One possible difficulty with this approach is that pain, when considered multiply reduced to a set of diverse physical-biological bases, seems to lose its integrity as a single mental state; pain as such appears either eliminated or else remains unreduced outside the ontology of the lower level theories.

6 The primacy of physical theory

Physics is generally thought to be our basic science, the only science that aspires to ‘full coverage’ of all of the natural world. But what does this mean? Does it mean that all the special sciences - that is, laws and generalizations of these sciences and the properties posited by them - are reducible to the basic laws of physics and fundamental physical properties? It used to be thought that the primacy of physical theory was equivalent to global physical reductionism. This view will be rejected by most philosophers. Many now doubt the possibility of reduction in almost all areas of science, and downplay the scientific and philosophical significance of reductions (although of course no one has seriously suggested that we would ever actually achieve a single unified science in the vocabulary of basic microphysics). Some have argued that even where reductions are possible, we lose important explanatory information about phenomena in a given domain when we focus only on their microstructures and neglect the larger ‘patterns’ that emerge at macrolevels. These macropatterns are claimed to cut across microstructures, and be capturable only by higher level laws (Fodor 1974). These remain controversial issues, however.

Views of this kind resemble the position of emergentism (Morgan 1923) on higher level ‘emergent’ properties. Those who hold these or related views often try to explain the primacy of physical theory in a thesis of supervenience or determination, the claim that physical facts (including physical laws) determine all the facts, or that worlds that are indiscernible in respect of all physical features are one and the same world (Hellman and Thompson 1975; Kim 1984). However, the relationship between reductionism and various forms of the supervenience thesis remains controversial and continues to be debated (Kim 1984).

See also: Laws, natural; Logical positivism; Reductionism in the philosophy of mind; Simplicity (in scientific theories); Supervenience; Supervenience of the mental

JAEGWON KIM

References and further reading

Benacerraf, P. (1965) ‘What Numbers Could Not Be’, *Philosophical Review* 74: 47- 73.(Raises important ontological issues concerning the set-theoretic reduction of numbers.)

Carnap, R (1938-55) ‘Logical Foundations of the Unity of Science’, in O. Neurath, R. Carnap and C. Morris (eds) *International Encyclopedia of Unified Science*, Chicago, IL: University of Chicago Press, vol. 1, 42-62.(A classic positivist statement on reductionism.)

Causey, R.L. (1977) *Unity of Science*, Dordrecht: Reidel.(A useful and comprehensive account of reduction in science.)

Churchland, P.S. (1986) *Neurophilosophy*, Cambridge, MA: MIT Press.(Contains discussion of reductive

strategies in cognitive neuroscience.)
Fodor, J.A. (1974) ‘Special Sciences - or the Disunity of Science as a Working Hypothesis’, *Synthèse* 28: 97-115. (Presents the multiple realization argument against reduction and reductionism discussed in §5, and defends the autonomy of the special sciences.)
Hellman, G. and Thompson, F. (1975) ‘Physicalism: Ontology, Determination, and Reduction’, *Journal of Philosophy* 72: 551-64.(Presents a version of physicalism that rejects reductionism in favour of a supervenience thesis.)
Hooker, C. (1981) ‘Toward a General Theory of Reduction’, *Dialogue* 20: 38-60, 201-35, 496-529.(A detailed survey and critical examination of the literature on reduction up to the late 1970s.)
Kemeny, J. and Oppenheim, P. (1956) ‘On Reduction’, *Philosophical Studies* 7: 6-18.(Presents the classic replacement model of reduction as discussed in §3.)
Kim, J. (1984) ‘Concepts of Supervenience’, *Philosophy and Phenomenological Research* 45: 153-76. (Distinguishes various forms of supervenience and discusses the relationship between supervenience and reduction.)
Kim, J. (1992) ‘Multiple Realization and the Metaphysics of Reduction’, *Philosophy and Phenomenological Research* 52: 1-26.(An assessment of the multiple realization argument against reduction.)
Morgan, C.L. (1923) *Emergent Evolution*, London: Williams & Norgate.(A classic statement of emergentism.)
Nagel, E. (1961) *The Structure of Science*, New York: Harcourt Brace.(The classic source of Nagel’s derivational model of theory reduction as discussed in §2, and which also includes discussion of emergentism.)
Oppenheim, P. and Putnam, H. (1958) ‘Unity of Science as a Working Hypothesis’, *Minnesota Studies in the Philosophy of Science*, vol. 2.(A useful presentation of the doctrine of physical reductionism through microreduction.)
Quine, W.V. (1964) ‘Ontological Reduction and the World of Numbers’, *Journal of Philosophy* 61: 209-16. (Reviews some ontological issues in the reduction of number theory, but the discussion has wider implications.)
Schaffner, K.F. (1967) ‘Approaches to Reduction’, *Philosophy of Science* 34: 137-47.(Stresses the point that the reduced theory must be appropriately corrected.)
Sklar, L. (1967) ‘Types of Intertheoretic Reduction’, *British Journal for the Philosophy of Science* 18: 109-24. (Argues that ‘bridge laws’ should be construed as identities.)

Reductionism in the philosophy of mind

Reductionism in the philosophy of mind is one of the options available to those who think that humans and the human mind are part of the natural physical world. Reductionists seek to integrate the mind and mental phenomena - fear, pain, anger and the like - with the natural world by showing them to be natural phenomena. Their inspirations are the famous reductions of science: of the heat of gases to molecular motion, of lightning to electric discharge, of the gene to the DNA molecule and the like. Reductionists hope to show a similar relationship between mental kinds and neurophysiological kinds.

1 The road to functionalism

Reductionism in the philosophy of mind is generated by an apparent conflict between two conceptions of persons. The first we inherit from our general culture, and is often called folk psychology. Folk psychology portrays humans as having experiences, emotions, purposes, thoughts about and expectations of the world: humans are conscious intentional agents (see Folk psychology). According to the picture we derive from the natural sciences, humans are evolved animals, and animals are nothing more than complex physical systems. Reductionist strategies in the philosophy of mind all start from a firm commitment to this second conception: the idea that humans are part of the natural order. That natural order is not essentially mysterious or occult. So reductionists accept some version of physicalism (see Materialism in the philosophy of mind).

Reductionist programmes in the philosophy of mind face three problems. (1) Are humans part of the natural order - is the physicalist conception true? (2) If we are part of the natural order, does folk psychology fit, or fail to fit, with that conception? (3) What do we need to do in order to integrate folk psychology within the naturalistic conception?

It has often been thought that no physicalist theory of consciousness or subjective experience is possible; these phenomena show the outright falsity of physicalism. For example, there seems to be a fundamental contrast between my knowledge of my own mind and my knowledge of the physical world. The latter is error-prone: it is conceivable that we could be wrong in any belief about the physical world. But it is not even conceivable, the idea runs, that I am wrong in thinking that I feel sleepy. I could be wrong about the causes of the feeling, but not that I have that feeling (see Introspection; Qualia).

There are three options available for those who think humans are part of an unmysterious natural order. One idea (the ‘No real problem’ option) is that the dualists’ puzzles of how thought and consciousness fit into the natural order evaporate on a proper understanding of the nature of folk psychology. Properly understood, the folk conception of the mental is unproblematic. The strategy seeks to show that mental concepts are equivalent to, or translate into, concepts for quite innocent phenomena. The most famous variant of this line is behaviourism. Behaviourists think that mental kinds have behavioural signatures; the essence of anger, for instance, is not a particular kind of internal mental event, but a way of behaving. A claim that a person is conscious, or intelligent, or is purposive is not about their inner workings or organization. It is not a claim about the causal structure that produces behaviour; but rather about the pattern of that behaviour. It is a claim that therefore cannot, of its very nature, traffic in weird causes or weird stuff (see Behaviourism, analytic).

The eliminativist option takes the conflict between folk psychology and the idea that humans are part of the natural order to be irresolvable. They think we cannot accept the view that we are conscious intentional agents and the view that we are part of the natural order. They furthermore suggest that the folk conception must be rejected, forced as we are to choose between them. So the eliminativist proposes to abandon the folk conception of the mind (see Eliminativism).

That leaves the reductionists. They accept (1) that folk conceptions of the mental do make claims about the causal structures that produce behaviour, and (2) that folk psychologists’ inventory of mental events, processes and kinds must be shown to be compatible with the view that the mind is a wholly natural phenomenon. If folk psychology offers an incomplete but more or less correct account of some of the causal structure underlying human behaviour, and if humans are just complex physical systems, then the events, processes and properties named in folk psychology must be physical events, processes and properties.

Identity theorists developed for philosophy of mind ideas then current in the philosophy of science about the relationship between less and more fundamental sciences (see Mind, identity theory of). The idea was that a less fundamental domain could be reduced to a more fundamental domain via bridge laws that linked the domains. Thus the chemical property of valency might be linked to atomic physics via bridge laws that identified different elements' valencies with properties of their electron structure. In the philosophy of mind, this became the identity theory, holding that

Mental states = states of the central nervous system.

In J.J.C. Smart's most famous example (1959), perhaps pain just is C-fibre stimulation. If identity hypotheses can be defended for all mental states, events and processes, then facts about the mental would be reduced to facts about neural structures, events and processes.

When the identity theory was formulated in the 1950s, it had to fight against the idea that it was conceptually confused. How could physicalism be true without being obviously true? The defenders of the theory had to show that one could master and understand the language of folk psychology without any mastery of the language of the neurosciences. They had to show how it is possible to know that you are in pain without knowing, say, that your C-fibres are being stimulated.

Identity theorists met these challenges through example and analysis. Though lightning is nothing but the discharge of static electricity we can understand lightning and know plenty about lightning without knowing its physical nature. Until the discovery of the structure of DNA, population geneticists talked of genes without knowing their chemical nature. Genes were identified by their role in inheritance and development. We can understand concepts for kinds, and know something about these kinds without knowing about their intrinsic characteristics. Smart re-reinforced these examples with a 'topic neutral analysis' of mental concepts. To say that you are experiencing blue visual sensations, for example, is just to say that the experiences you are now having are similar to those you have when you see a blue wall in good light. It is obviously possible both to learn and correctly deploy topic neutral sensation concepts without knowing anything about the intrinsic nature of those sensations.

The identity theory however proved to be chauvinist in Ned Block's terminology (1978). If pain is C-fibre stimulation, then if you have no C-fibres, you cannot feel pain. Yet it is surely possible, and perhaps even actual, for organisms with different neurophysiologies to feel pain. It seemed to be a mistake to tie our account of what cognition is too closely to the details of how our brain works. So the identity theory does not quite show how it is possible to be both a folk psychologist and a reductionist.

2 Functionalism

So most current reductionists accept some version of functionalism. The many forms of that doctrine are united in holding that the essential feature of any mental state is its causal role (see Functionalism). For example, all instances of fear have a distinctive causal role: fear has a characteristic profile of causes and effects. But functionalists also claim that fear is realized by a physiological state: most probably a complex of hormonal activity and arousal of the nervous system. The functional role of fear is occupied by a physiological state. If functionalists are right, our mental life can be described in complementary ways. A functional description tells us the causal roles of the full range of human psychological states. A physical description specifies the physical states which have those causal roles. The attraction of functionalism as a theory of mind is its attempt to reconcile physicalism with folk psychology.

The relationship between functional kinds and their physical realizations is complex. Two different people can be in the same mental state, hold the same belief, for instance. A functional description specifying a certain causal role is true of them both. Psychological kinds can be multiply realized. The neural organization of our memory is flexible and shows a good deal of individual variation, so the same mental state can have different physical realizations in different people, or in the same person at different times. These complications are magnified when we consider psychological states - fear, for example - that we share with other animals. Furthermore, if there are intelligent aliens or robots, then there are systems whose intelligence has a very different physical basis from ours. So the one mental state might have wildly varied physical realizations.

The functionalist pictures psychological kinds as realized by neural kinds in complex and varied ways. So the natural kinds of psychology are not identical to the natural kinds of the physical basis of the mind. Hence, despite the fact that we are nothing but complex physical systems, functionalists often claim that psychology is relatively independent of the sciences of the brain. The functionalist programme needs to show (1) that a reductionist strategy of some kind is needed, that against the no real problem option it really is necessary to demonstrate the connection between the folk psychological and the natural science picture; (2) that folk psychology can be properly built into the physicalist world view through the idea of functional kinds as merely realized by physical kinds, and (3) that it really is plausible that all psychological kinds are some species of functional kind.

3 ‘No problem’ theories

Behaviourism is no longer widely accepted, but it is still argued that there is no potential for conflict between folk psychology and the natural sciences because folk psychology is an instrument: its use is instrumental or practical rather than theoretical. The contrast here is with the so-called ‘theory-theory’: folk psychology as an implicit theory of the causes of human behaviour. Like other theories, such a theory would need to be integrated with the rest of the natural sciences; like other theories, it could be wrong and shown to be wrong by its conflict with, or isolation from, the rest of our picture of nature.

The theory-theory may well over-intellectualize folk psychology. Folk psychology is not part of a conscious research programme into the structure of the human mind. Our use of it may be as much knowing how as knowing that. Nevertheless, folk psychology is full of claims about human minds, and about the causes of human behaviour: about, for instance, beliefs and desires causing behaviour (‘They went because they believed that they would enjoy themselves’), and about perceptions causing beliefs (‘Seeing is believing’). Paul Churchland (1981) has given one set of arguments for the theory-theory: folk psychology can be assembled into a package of generalizations connecting perceptions, thought and behaviour. Folk psychology can be organized in such a way that there are structural similarities between it and paradigm scientific theories. But a potential for conflict between the natural sciences and folk psychology does not depend on accepting a full-blooded version of the theory-theory. An analogy with other folk crafts should make that clear. The folk agriculture of, say, a New Zealand Maori tribe is not research science. It includes as much knowledge how as knowledge that. Its uses are practical. Yet none of that made it invulnerable to advances in scientific knowledge. Its central categories might turn out to be empty, to be misconceived, if their failure to correspond to the causes of growth are radical enough. Potential for conflict depends only on the idea that practices, modes of practical knowledge and skill, can depend on or presuppose conceptions of the causal structure of a domain. Even if folk psychology has different goals or purposes from the scientific behavioural sciences the demand that folk psychology be shown to be compatible with a naturalistic conception of human nature is still reasonable. An analogy with folk medicine is compelling. What would we think of a claim for the therapeutic value of some practice where we could find no physiological mechanism that would explain that value? Or, still worse, where our physiological sciences seem to show that the practice could not have therapeutic value?

4 What does physicalism require?

Though the matter remains very controversial, in my view the ‘no real problem’ option is not attractive. Folk psychology seems to make claims about the working of the human mind, about its causal structure, whose relation to the natural sciences requires explanation. But it has proved very difficult to formulate precisely this idea of demonstrating compatibility between two conceptions of human nature. At the very least, folk psychological states must supervene on the physical world. There can be no change in my folk psychological state - say, from fear to calm - without some change in my intrinsic or relational physical properties (see Supervenience of the mental). But the eliminativists argue for a much stronger condition, something close to the original idea of the identity theory’s bridge laws between the domains of psychology and neurophysiology. They argue that without a strong condition, spurious compatibility is too easy to construct, and the debris of intellectual history will escape appropriate rejection.

Despite the eliminativists’ claims, the idea that folk psychology requires bridge laws for its vindication is clearly much too strong. The double descriptions characteristic of the functionalist idea is not a fancy trick dreamed up just for the philosophy of mind. The distinction between a role and its occupant is not restricted to psychology. We can make the same distinction in politics between, say, the chief justice and the particular person who occupies

that role. We can ask questions about that role; about the legal and constitutional powers of the chief justice. We can ask questions about the occupier: about their political background and skills; about size, sex or place of birth. We can make the same distinction in biology. Zoologists investigate the causal role echo-location plays in the life of the bat. They measure the discriminatory capacities and range of the system, test its role in hunting, courtship and mating. Neuroethologists investigate the physical systems that occupy that causal role in the bat.

The idea of multiple realization is not restricted to the relationship between mental state types and the neural states realizing those types. There are no simple bridge laws between classical genetics and molecular genetics, the science which is rightly seen as vindicating classical genetics. The bodily effects of DNA sequences depend very much on the cellular environment of those DNA sequences plus many other factors. So the geneticists' fruit-fly gene for red eyes is not identical to a particular DNA sequence: there is no DNA sequence that always and only produces red-eyed fruit-flies (see Genetics). So the complexity of the relationship between functional state and physical realization is not a feature just of the relationship between folk psychology and neuroscience; it is common to many of these other domains as well.

Nonetheless if folk psychological kinds are parts of causal explanations within human psychology, an explanation of their capacity to play that causal role does seem needed. Causal powers stand in need of explanation. There is a tree of explanatory dependence that links together causal mechanisms. That tree is rooted in fundamental physical kinds and processes. Through various different branchings, all scientific kinds depend on that taproot. The kinds (and the processes in which they take part) further out in the branches need to be explained by kinds and laws closer to the root.

The importance of this idea is seen in its regulative role in scientific debate. For example, the biogeographic and geological evidence for continental drift was quite impressive even before the Second World War. But continental drift remained marginalized in the profession, in part because the proposed mechanisms to shift continents were impossible. The drifters of the 1930s conceived of continents as ploughing through somewhat less dense ocean floor rather as a concrete slab might be pushed half through, half over the top, of a layer of earth. The proposed forces were too weak, and the stresses on continental crusts would be far too great for them to survive the passage. Until drift could be plausibly placed in the explanatory tree, driftist explanations of movement were hard to accept.

There have been similar problems with mechanism in recent debates within evolutionary theory. From time to time there are announcements of counterexamples to Crick's central dogma of molecular biology, nucleic acids code proteins, but not vice versa; changes in the protein envelope of the nuclear material do not result in reverse transcription, changing RNA and thence DNA (though reverse transcription form RNA to DNA is possible). There is no plausible mechanism through which three dimensional form can be transcribed into two dimensional form, hence no mechanism through which acquired characteristics can be inherited through changes in DNA. A discoverer of the inheritance of acquired characteristics would need to give some account of a mechanism of inheritance.

The reductionist debate has not been concluded. A precise and plausible formulation of the physicalist constraint remains elusive. Many psychological kinds continue to resist functionalist theory: if they have a characteristic causal role, that causal role has been difficult to characterize. The appropriate formulation of functionalism remains a matter of intense debate.

See also: Reduction, problems of

KIM STERELNY

References and further reading

Armstrong, D.M. (1968) *A Materialist Theory of Mind*, London: Routledge & Kegan Paul.(An early functionalist classic, but written before the distinction between functionalism and the identity theory was clearly drawn. So Armstrong describes the book as a defence of the identity theory.)

Block, N. (1978) 'Troubles with Functionalism' in C. Wade Savage (ed.) *Minnesota Studies in the Philosophy of Science*, vol. 9, *Perception and Cognition*, Minneapolis. MN: University of Minnesota Press, 261-325.(Block describes the problems for the identity theory, but spends most of this paper showing how hard it is to specify the functional profile of most mental kinds.)

Churchland, P.M. (1981) 'Eliminative Materialism and the Propositional Attitudes', *Journal of Philosophy* 78:

67-90.(A clear and classic exposition of eliminativism, reprinted in W.G Lycan, 1990.)

Dennett, D.C. (1987) *The Intentional Stance*, Cambridge, MA.: MIT Press.(A very sophisticated version of the 'No problem' approach to the cognitive aspects of the mind.)

Dennett, D.C. (1991) *Consciousness Explained*, Boston, MA: Little, Brown.(A very sophisticated version of the 'No problem' strategy applied to the problems of consciousness. The first two chapters give a good account of the attractions of dualism.)

Fodor, J.A. (1968) *Psychological Explanation*, New York: Random House.(An early and very clear version of functionalism.)

Fodor, J.A. (1975) *The Language of Thought*, New York: Thomas Crowell.(A classic and very accessible paradigm of functionalism focused on thought and thinking rather than consciousness.)

Lycan, W.G. (1987) *Consciousness*, Cambridge, MA: MIT Press.(A fine and very clearly written example of contemporary functionalism.)

Lycan, W.G. (1990) *Mind and Cognition: a Reader*, Oxford: Blackwell.(A well-organized and well-chosen anthology with good introductions.)

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.(The first clear statement of the behaviourist programme.)

Smart, J.J.C. (1959) 'Sensations and Brain Processes', *Philosophical Review* 68: 141-56.(An early and very important defence of the identity theory; available in many collections.)

Reference

It is usual to think that referential relations hold between language and thoughts on one hand, and the world on the other. The most striking example of such a relation is the naming relation, which holds between the name 'Socrates' and the famous philosopher Socrates. Indeed, some philosophers in effect restrict the vague word 'reference' to the naming relation, or something similar. Others use 'reference' broadly (as it is used in this entry) to cover a range of semantically significant relations that hold between various sorts of terms and the world: between 'philosopher' and all philosophers, for example. Other words used for one or other of these relations include 'designation', 'denotation', 'signification', 'application' and 'satisfaction'.

Philosophers often are interested in reference because they take it to be the core of meaning. Thus, the fact that 'Socrates' refers to that famous philosopher is the core of the name's meaning and hence of its contribution to the meaning of any sentence - for example, 'Socrates is wise' - that contains the name. The name's referent contributes to the sentence's meaning by contributing to its truth-condition: 'Socrates is wise' is true if and only if the object referred to by 'Socrates' is wise.

The first question that arises about the reference of a term is: what does the term refer to? Sometimes the answer seems obvious - for example, 'Socrates' refers to the famous philosopher - although even the obvious answer has been denied on occasions. On other occasions, the answer is not obvious. Does 'wise' refer to the property wisdom, the set of wise things, or each and every wise thing? Clearly, answers to this should be influenced by one's ontology, or general view of what exists. Thus, a nominalist who thinks that properties do not really exist, and that talk of them is a mere manner of speaking, would not take 'wise' to refer to the property wisdom.

The central question about reference is: in virtue of what does a term have its reference? Answering this requires a theory that explains the term's relation to its referent. There has been a great surge of interest in theories of reference in this century.

What used to be the most popular theory about the reference of proper names arose from the views of Gottlob Frege and Bertrand Russell and became known as 'the description theory'. According to this theory, the meaning of a name is given by a definite description - an expression of the form 'the F' - that competent speakers associate with the name; thus, the meaning of 'Aristotle' might be given by 'the last great philosopher of antiquity'. So the answer to our central question would be that a name refers to a certain object because that object is picked out by the name's associated description.

Around 1970, several criticisms were made of the description theory by Saul Kripke and Keith Donnellan; in particular, they argued that a competent speaker usually does not have sufficient knowledge of the referent to associate a reference-determining description. Under their influence, many adopted 'the historical-causal theory' of names. According to this theory, a name refers to its bearer in virtue of standing in an appropriate causal relation to the bearer.

Description theories are popular also for words other than names. Similar responses were made to many of these theories in the 1970s. Thus, Kripke and Hilary Putnam rejected description theories of natural-kind terms like 'gold' and proposed historical-causal replacements.

Many other words (for example, adjectives, adverbs and verbs) seem to be referential. However we need not assume that all other words are. It seems preferable to see some words as syncategorematic, contributing structural elements rather than referents to the truth-conditions and meanings of sentences. Perhaps this is the right way to view words like 'not' and the quantifiers (like 'all', 'most' and 'few').

The referential roles of anaphoric (cross-referential) terms are intricate. These terms depend for their reference on other expressions in their verbal context. Sometimes they are what Peter Geach calls 'pronouns of laziness', going proxy for other expressions in the context; at other times they function like bound variables in logic. Geach's argument that every anaphoric term can be treated in one of these two ways was challenged by Gareth Evans.

Finally, there has been an interest in 'naturalizing' reference, explaining it in scientifically acceptable terms.

Attempted explanations have appealed to one or more of three causal relations between words and the world: historical, reliable and teleological.

1 Millian and description theories of proper names

The description theory of proper names stands in sharp contrast to the age-old and attractive theory that there is no more to a name's meaning than its role of designating something. Thus, John Stuart Mill claimed that 'proper names are not connotative: they denote the individuals who are called by them; but they do not indicate or imply any attributes as belonging to those individuals' ([1843] 1867: 20) (see Mill, J.S. §2). One problem for this theory is that it makes answering our central question seem hard: if a name does not 'imply any attributes' of its bearer, as the description theory claims it does, what determines which object is the bearer?

It was not primarily concern with this question, however, that led most philosophers to abandon the Millian theory and adopt the description theory. They did this largely as a result of the criticisms and counter-proposals of Frege (1893) and Russell (1911).

The most famous criticism of the Millian theory concerns identity statements. The ancient Greeks observed what they took to be a star rising in the evening and called it 'Hesperus', and what they took to be another star rising in the morning and called it 'Phosphorus'. In fact these 'two stars' were the planet Venus. Consider the following two statements:

(1) Hesperus is Phosphorus.
(2) Hesperus is Hesperus.

Sentence (1) is true. A comparison of (1) and (2) seems to indicate that they differ sharply in meaning. Various reasons have been adduced in favour of this view. Some philosophers have argued that whereas (1) is synthetic, (2) is analytic; others that whereas (1) is known empirically, (2) is known a priori; still others that whereas (1) is contingent, (2) is necessary. However, probably the most influential reason for thinking that the statements differ in meaning has been Frege's claim that they differ in 'cognitive value': (1) is highly informative, revealing an important astronomical discovery, whereas (2) is uninformative, a trivial piece of logical knowledge (see Frege, G. §3). In any case, it was generally agreed that the two statements do differ in meaning. If so, the only way to explain this seemed to be to attribute different meanings to 'Hesperus' and 'Phosphorus'. Yet, according to the Millian theory, they must have the same meaning as they both have the role of designating Venus. So the Millian theory must be wrong: a name's role of designating its bearer does not exhaust its meaning.

Another important criticism concerns existence statements. 'Vulcan does not exist' is true. Because it is true, 'Vulcan' does not designate anything. So, according to the Millian theory, 'Vulcan' should be meaningless. So, 'Vulcan does not exist' should be partly meaningless. Yet, it is perfectly meaningful. Indeed, if it were not, it could not be true. So, once again, the Millian theory must be wrong (see Fiction, semantics of).

The description theory provides neat solutions to what are problems for the Millian theory. According to the description theory, a name in effect abbreviates the definite description that competent speakers associate with the name: thus 'Aristotle' might abbreviate 'the last great philosopher of antiquity'. So it is easy to see why (1) and (2) differ in meaning: 'Hesperus' might be associated with the description 'the star that rises in the evening', and 'Phosphorus' with 'the star that rises in the morning'. And 'Vulcan does not exist' is fully meaningful because 'Vulcan' abbreviates the meaningful description 'the planet in orbit between Mercury and the Sun' (see Proper names §§1, 5).

More importantly for our purposes, the description theory provides an answer to our central question. A name designates a certain object because that object is denoted by the definite description associated with the name; 'Aristotle' designates Aristotle because 'the last great philosopher of antiquity' denotes him. Of course, this answer raises another question (see §6): in virtue of what does the description denote that object? Still, progress has clearly been made because we had that problem anyway.

Some obvious problems with this 'classical' description theory - including the problem that speakers differ in the descriptions that they associate with a name - led some philosophers, notably John Searle (1958) and Peter Strawson (1959), to modify the theory. A name is not tied tightly to one description but loosely to many. It can

designate its bearer despite the failure of some in its 'cluster' of associated descriptions to denote that object: it designates whatever object most of the descriptions in the cluster denote.

2 Three arguments against description theories of proper names

Description theories dominated for half a century until challenged by three arguments around 1970: the unwanted necessity and rigidity arguments, both due largely to Kripke (1980), and the argument from ignorance and error, due to Kripke and to Donnellan (1972).

Unwanted necessities were one of the obvious problems for the classical description theory. If 'the last great philosopher of antiquity' is synonymous with 'Aristotle', then 'Aristotle is a philosopher' should be necessarily true (provided Aristotle exists). Yet it is not: Aristotle might have died young, long before his philosophical fulfilment. The cluster theory avoided this version of the problem. The description 'the last great philosopher of antiquity' is just one among many in the cluster that expresses the meaning of 'Aristotle'. Aristotle need not have any particular one of the many properties specified by the cluster. The cluster theory does require, however, that Aristotle have *most* of the properties specified by the cluster. Kripke points out how implausible this is. Aristotle might not have had any of the properties commonly associated with him: he might not have been a pupil of Plato, taught Alexander the Great, and so on. So 'Aristotle' cannot be synonymous with the cluster of associated descriptions (see Kripke, S.A. §2).

The rigidity argument deploys the notion of 'rigid designation'. This is explained as follows: for a term *a* to be a rigid designator is for it to designate the same object in every possible world (in which it designates at all), or, less picturesquely, for it to be such that '*a* is *F*' would truly characterize some non-actual situation if and only if the object that the term *actually* designates were *F* in that situation. Kripke argues that names are rigid designators whereas the descriptions alleged to be synonymous with them are not. So description theories are wrong. Consider:

(3) Aristotle was fond of dogs.
(4) The last great philosopher of antiquity was fond of dogs.

Suppose that Aristotle had indeed died young. Then Plato, not Aristotle, would have been the last great philosopher of antiquity. In those circumstances the truth of (4) would depend on whether Plato was fond of dogs. But the truth of (3) would still depend, just as it does depend in the actual world, on whether Aristotle was fond of dogs. The name 'Aristotle' designates Aristotle in a non-actual situation just as it does in the actual situation, whereas the description 'the last great philosopher of antiquity' designates whoever is the last great philosopher of antiquity in that situation, whether Aristotle or not. So the name is not synonymous with the description. Similarly, it is not synonymous with any other description, or cluster of descriptions, that is a candidate to give its meaning. (Note that in assessing rigidity we are evaluating the truth and reference of our expressions with the meanings that they actually have as a result of our usage, but we consider them as used to characterize hypothetical situations. Of course, any expression could have a different meaning as a result of different usage in a non-actual situation - language is 'arbitrary' - but that is beside the point) (see Proper names §§2-3).

Statements (3) and (4) are not modal statements (although we have been evaluating them as characterizations of non-actual situations). Other versions of the rigidity argument concern modal statements. For example, whereas 'Hesperus is necessarily Hesperus' is true, 'Hesperus is necessarily the star that rises in the evening' is not: had the solar system been differently arranged, Hesperus might not have been visible in the evening but it still would have been Hesperus. This sort of difference between descriptions and names in modal statements had been emphasized earlier by other philosophers, particularly Ruth Barcan Marcus (1961).

Some philosophers, notably Michael Dummett (1973), resisted Kripke's two arguments by focusing on modal statements. These philosophers exploited the well-known ambiguities of scope in these statements to undermine the apparent difference between names and descriptions. Whatever the truth of this matter, the apparent difference in nonmodal statements remains.

These two arguments challenge the description theory as a theory of the meaning of a name, a meaning that determines the name's reference. This is how the theory is naturally understood. However, as Kripke points out, the theory could be understood as simply a theory of reference: the reference of a name is fixed by a description, but the name is not synonymous with that description. This weaker theory is impervious to the two arguments. Of

course, the weak theory has a defect: because it is no longer a theory of meaning, it no longer solves the problems that troubled the Millian theory. Indeed, the relation between meaning and reference becomes a pressing issue on the basis of this theory.

There is another way of saving the description theory from the two arguments while avoiding this defect. Instead of weakening the original theory into a mere theory of reference, we revise it along the following lines: a name is synonymous with a ‘rigidified’ description. Our language already seems to have descriptions that contain ‘rigidity operators’; for example, the italicized part of ‘the person who, *in the actual world*, is the last great philosopher of antiquity’ seems to make this description designate Aristotle in every possible world. If descriptions of this sort are indeed rigid, the revised theory claims that a name is synonymous with such a description. If such descriptions are not rigid, the revised theory can claim that the name itself supplies the rigidity operator and so would be synonymous with an ordinary nonrigid description governed by that operator.

All of these description theories - original, weak and revised - have the consequence that the users of a name associate with it a description that identifies its bearer. The third argument against description theories, the argument from ignorance and error, challenges this. So, if the argument is good, it counts against all description theories.

The argument shows that people who seem perfectly able to designate with a name are very often too ignorant to supply an identifying description. Thus some may fail with the name ‘Cicero’ because they associate with it only the description ‘a famous Roman orator’, which applies to many people. Others may fail because they associate ‘the man who denounced Catiline’ with ‘Cicero’ and are unable to supply an appropriate description for ‘Catiline’: the description that they associate with ‘Catiline’ is ‘the man denounced by Cicero’, which takes us in a circle and leaves both names without reference, according to the description theory.

The argument shows also that people often associate with a name a description that identifies something other than the name’s bearer: they are simply wrong about the bearer. Thus some associate ‘the inventor of the atomic bomb’ with ‘Einstein’ and some associate ‘the first person to realize that the world was round’ with ‘Columbus’. Almost everyone who has heard of Peano associates ‘the discoverer of Peano’s axioms’ with ‘Peano’, but the axioms were actually discovered by Dedekind. Despite such errors, people succeed in designating Einstein, Columbus and Peano by their names.

The description theory can be improved by allowing people to ‘borrow’ their reference from others. So the description Martha associates with ‘Einstein’ might be ‘the person Joe referred to yesterday as "Einstein"’. Provided Joe can supply an appropriate description - either one that describes Einstein directly or one that borrows reference from someone who can supply an appropriate description - Martha will succeed in designating Einstein. There is a danger of a circle, of course. Apart from that, there are problems of ignorance and error once more. Perhaps Martha cannot remember the reference lender; or she can remember the lender by his name, ‘Joe’, but cannot supply the identifying description that the theory requires; or the lender is identified but he cannot identify Einstein, perhaps identifying something else instead. The description theory still seems to place too great an epistemic burden on speakers.

An argument from ignorance and error can also be brought against another, more general, theory that some - for example, Dummett - have taken from Frege. This is the theory that to understand a name a person must be able to ‘identify’ its bearer. This ability is usually evidenced by providing a description, but it may be evidenced by ‘recognizing’ the bearer. The epistemic burden that this more general theory places on speakers still seems too great.

Various moves have been made to save the description theory in the face of these difficulties. Most popular, perhaps, have been theories that the reference of, for example, ‘Einstein’ is determined by a description along the lines of ‘the person referred to by (called, named and so on) "Einstein"’, for this description does identify Einstein and speakers surely associate it with the name. However, such theories still risk circularity.

3 General terms and mass terms

Just as there are description theories of names, so also there are description theories of general terms like ‘tiger’, ‘hammer’ and ‘bachelor’, and of mass terms like ‘gold’ and ‘paper’ (see Mass terms). Speakers of the language

associate various descriptions with a term. One of these descriptions, or most of a cluster of them, expresses the meaning of the term and determines what it applies to. If only one description does the job, the view is analogous to the classical description theory of names. If a cluster of descriptions does, the view is analogous to the cluster theory of names.

Kripke (1980) and Putnam (1975) argued that description theories are false of general and mass terms that apply to natural kinds. So they are false of ‘tiger’ and ‘gold’. The arguments are like the three against description theories of names (see §2). First, the theories yield unwanted necessities. The description we associate with ‘tiger’ is along the lines of ‘large carnivorous quadrupedal feline, tawny yellow in colour with blackish transverse stripes and white belly’. Yet it is not necessary that a tiger has four legs and is striped: a tiger might lose a leg, or in a different environment tigers might not be striped. Second, the term ‘gold’ is a rigid designator, applying to the same kind of stuff in every possible world. In contrast, an associated description like ‘dense yellow metal’ is nonrigid. Third, people who seem perfectly able to use a term are often too ignorant or misguided about the things to which it applies to supply an appropriate identifying description. Thus, some who use ‘elm’ and ‘beech’ cannot supply descriptions that distinguish elms from beeches; many who use ‘gold’ cannot distinguish gold from fool’s gold; it was once common to associate ‘fish’ with ‘whale’.

Putnam added a further argument built around the following fantasy. Imagine that somewhere in the galaxy there is a planet, Twin Earth. Twin Earth, as its name suggests, is very like Earth. In particular, each Earthling has a *doppelgänger* on Twin Earth who is a cell for cell duplicate of the Earthling. Twin Earth differs from Earth in one respect, however: the stuff that the Twin Earthians who appear to speak English call ‘water’, stuff that is superficially indistinguishable from what we call ‘water’, is not H_2O but a very different compound XYZ. So Oscar on Earth and Twin Oscar on Twin Earth are referring to different stuff by ‘water’. Yet Oscar and Twin Oscar are *doppelgängers*, associating exactly the same descriptions with ‘water’ (which is more plausible if we place Oscar and Twin Oscar in 1750 before the chemical composition of water was known). So those associations are not sufficient to determine reference and the description theory is wrong. Indeed, nothing happening in the head is sufficient to determine reference. As Putnam put it, ‘meanings just ain’t in the head’ (see Content: wide and narrow; Putnam, H. §3).

We have considered criticisms of description theories of proper names and natural-kind words. Do these criticisms extend to description theories of other words? Putnam took the arguments to apply to almost all words, including ‘pencil’ and ‘paediatrician’. Tyler Burge (1979) took a similar line, arguing that the meanings (contents) and references of a wide range of a person’s words and accompanying thoughts are not ‘individualistic’ in that they are not determined simply by that person’s intrinsic states. To a large extent they are determined by the person’s social context. Burge’s examples include ‘arthritis’, ‘sofa’, ‘brisket’, ‘clavichord’ and ‘contract’ (see Methodological individualism).

The Twin Earth fantasy brings out an important feature of description theories in general: even if a description theory gives the right answer to our central question for some word, its answer is incomplete. Thus, consider a description theory of ‘tiger’. According to the theory, the reference of ‘tiger’ is determined by the reference of such words as ‘carnivorous’ and ‘striped’. Suppose, contrary to the arguments above, that this were so. We then need to explain the reference of those words to complete the explanation of the reference of ‘tiger’. Description theories might be offered again. But then the explanation will still be incomplete. At some point we must offer a theory of reference that does not make the reference of one word parasitic on that of other words. We need an ‘ultimate’ explanation of reference that relates some words directly to the world, if there is to be any reference at all.

4 Historical-causal theories

Kripke and Donnellan followed their criticism of description theories of names with an alternative view. This became known as the ‘causal’ or ‘historical’ theory, although Kripke and Donnellan regarded their view as more of a ‘picture’ than a theory (see Proper names §4).

The basic idea of this theory is that a name designates whatever is causally linked to it in an appropriate way, a way that does not require speakers to associate an identifying description of the bearer with the name. Reference is initially fixed at a dubbing, either by perception or description of the referent. The name is then passed on from

person to person in communicative exchanges. People succeed in designating an object with a name because underlying their uses of the name are causal chains stretching back to the dubbing of the object with the name. People borrow their reference from people earlier in the chain but borrowers do not have to remember lenders; it is enough that borrowers are, as a matter of historical fact, appropriately linked to their lenders in communication. So people can designate Cicero despite their ignorance of him, or designate Einstein despite their errors about him.

Similarly, Kripke and Putnam proposed an historical-causal theory of natural-kind words. Reference is initially fixed at a dubbing, either by description or perception of samples of the kind. The reference is then to all those objects, or all that stuff, having an internal structure of the same sort as the samples: for example, in the case of gold having the atomic number 79. People at a dubbing lend their reference to others, who can then lend it to still others. People who are ignorant about the kind can use the word to refer to the kind's members because underlying the use of the word are causal chains stretching back to a dubbing.

In thus removing the epistemic burden on speakers, historical-causal theories are a radical departure from the Frege-Russell tradition. That tradition assumes that those who understand a name must know about its meaning and reference, so that if its reference is determined in a certain way, they must know that it is. The historical-causal theory must reject the assumption that speakers have this privileged 'Cartesian' access to semantic facts: the reference of a name is determined by causal chains that are likely to be beyond the ken of the ordinary speaker. Once again, we must conclude with Putnam that 'meanings just ain't in the head'. This very feature of the theory has led many to reject it and to work hard to preserve the description theory (or the more general 'identification theory' favoured by Dummett; see §3). From the traditional Cartesian perspective, the causal theory's failure to impose an epistemic burden rules it out as a candidate to explain reference.

The historical-causal theory nicely captures the rigidity of names: the reference of a name is determined by its *actual* causal relations, something that cannot change when we consider other possible worlds. Less pleasingly, by rejecting any descriptive element to the meaning of a name, the theory may seem to leave no alternative but to resurrect the Millian view, identifying a name's meaning with its role of designating its bearer. Many philosophers, influenced by the 'direct reference' approach to indexicals (see §5), have taken this route, despite the problems for the Millian view (see §1). To avoid this it seems that we must explain a name's meaning in terms of the particular sort of causal chain that determines the name's reference.

The theory faces problems arising from various confusions and mistakes that can play a role in forming the causal network underlying a name. And it must explain how the reference of a name can change even though the historical fact of the dubbing cannot change. In developing the theory to deal with these problems, Michael Devitt (1981) has emphasized that a name is typically 'grounded' in its bearer in many perceptual confrontations after the initial dubbing: it is 'multiply' grounded in its bearer.

The historical-causal theory is returned to in §8.

5 Indexicals

To answer our central question for indexicals - terms like 'I', 'now', 'here', 'you', 'she', 'that' and 'this table' - we need to consider how the context of their utterance determines their reference.

Both Russell (1918) and Hans Reichenbach (1947) explain the reference of all indexicals (called 'egocentric particulars' by Russell and 'token-reflexive words' by Reichenbach) in terms of the reference of 'this'. Thus, Reichenbach claims that '"I" means the same as "the person who utters this token"; "now" means the same as "the time at which this token is uttered"; "this table" means the same as "the table pointed to by a gesture accompanying this token"' (1947: 284). This amounts to a description theory of reference for all indexicals except 'this token' in terms of the reference of 'this token'.

Any description theory of an indexical may face arguments of the usual three sorts: unwanted necessities, rigidity, and ignorance and error. The theory is particularly vulnerable to the argument from rigidity, as David Kaplan (1989) showed. (Kaplan prefers to talk of the closely related notion of 'direct reference' rather than 'rigidity'.) Thus, compare:

(5) This table is green

and Reichenbach's interpretation of it:

(6) The table pointed to by a gesture accompanying this token is green.

Suppose that the table referred to is in fact green, so that (5) and (6) both assert true propositions. Consider now a situation in which that table was still green but the furniture had been moved around so that a different table, a brown one, would be the subject of the gesture. Would those propositions asserted by (5) and (6) still be true? (Note that this question concerns the propositions actually asserted by (5) and (6), not the propositions that would have been asserted by the sentences in that non-actual situation.) Kaplan argues that whereas what (6) asserts would be false, what (5) asserts would still be true. The indexical 'this table' is rigid, referring to the same table in each possible world, whereas the description 'the table pointed to by a gesture accompanying this token' is nonrigid, referring to whatever table fits that description in the possible world. So the demonstrative is not synonymous with the description.

Apart from this, a general theoretical consideration counts against a description theory of indexicals. We have noted the essential incompleteness of description theories (see §3): even if a description theory is right for some word, the theory's explanation must rest ultimately on the reference of some other words which must be explained nondescriptively. Indexicals seem to be the most plausible candidates for nondescriptive explanation, more so even than proper names or natural-kind words: indexicals seem to be the place where language stands in its most direct relationship to the world.

In seeking a nondescriptive theory, it helps to follow Kaplan in dividing indexicals into two groups: 'pure indexicals' like 'I', 'here' and 'now', and 'demonstratives' like 'she', 'that' and 'this table'. The nondescriptive explanations of pure indexicals are fairly simple: 'I' designates the speaker of the utterance, 'here' the place of the utterance, 'now' the time of the utterance, and so on. (These explanations may seem to be description theories once again, but they are crucially different. For example, the last explanation is not that 'now' designates the time of the utterance because it is synonymous with an associated description 'the time of this utterance', but rather, because it is governed by the rule that it designates that time.) Demonstratives are more difficult to explain.

There are three basic ideas for a nondescriptive explanation of demonstratives. According to the first, a demonstrative designates the object demonstrated by the speaker. One problem with this idea is that a demonstration is often so vague that it alone would not distinguish one object from many others in the environment. A more serious problem is that demonstratives are not always accompanied by a demonstration. Thus, where only one table is prominent in the environment, the speaker may use 'this table' without a demonstration. And reference is often to an object that is not around to be demonstrated: for example, 'That drunk at the party last night was offensive'.

According to the second idea for a nondescriptive explanation, a demonstrative designates the object that the speaker intends to refer to. Even if this is so, it does not take us far because it raises the question: in virtue of what does the speaker intend to refer to that object? This is very similar to the original problem.

According to the third idea - urged, for example, by Edmund Husserl (1900-1) - a demonstrative designates the object in which it is based perceptually (compare the perceptually based grounding of a name according to the historical-causal theory). So 'this table' designates a certain table in virtue of the fact that it was perception of that table that led to the utterance; similarly 'that drunk at the party' designates the person that caused the remark (see Demonstratives and indexicals).

6 Descriptions

Definite descriptions have the form, 'the *F*', and indefinite descriptions the form, 'a/an *F*'. In his theory of descriptions, Russell (1905) claimed that 'the *F* is *G*' is equivalent to 'there is something that is alone in being an *F* and it is *G*'; and 'an *F* is *G*' is equivalent to 'there is something that is an *F* and it is *G*'. So the descriptions are to be understood in terms of the general term, '*F*' and the existential quantifier, 'there is something'.

Under the influence particularly of Donnellan (1966), many now think that a description is 'ambiguous', having not only this 'attributive' meaning captured by Russell but also a 'referential' meaning like that of a name or demonstrative.

It has been generally agreed that descriptions have a referential use as well as an attributive use. Used attributively, ‘the *F*’ conveys a thought about whatever is alone in being *F*; ‘an *F*’ conveys a thought about some *F* or other. Used referentially, each description conveys a thought about a particular *F* that the speaker has in mind, a thought about a certain *F*. Thus, consider:

(7) The murderer of Smith is insane

used in the following two contexts. (a) We come upon Smith foully murdered. We have no idea who is responsible but the brutal manner of the killing leads us to utter (7). Its description is used attributively. (b) We observe Jones on trial for Smith’s murder. The oddity of his behaviour leads us to utter (7). Its description is used referentially. Next, consider:

(8) A man in a red cap stole Anne’s computer

used in the following two contexts. (c) Anne’s computer is discovered missing in the morning. We find signs that the burglar made a hasty escape dropping a red cap in the alley. This leads us to utter (8). Its description is used attributively. (d) After discovering that Anne’s computer is missing we remember noticing a man in a red cap behaving suspiciously earlier in the day. We utter (8) to report our suspicions to the boss. Its description is used referentially.

Despite agreement that there are these two uses, there is no agreement that descriptions are ambiguous. Appealing to ideas prominent in the work of H.P. Grice (1989), many have defended Russell. They argue that a speaker can use a description referentially, thus making the object in mind the ‘speaker referent’, even though that object is not the semantic referent. Whether a speaker has an object in mind or not, the truth-conditions of the sentence are as specified by Russell. The referential use is pragmatically, but not semantically, different from the attributive use (see Pragmatics §§3-4, 8). Thus, in context (b), although Jones is the speaker referent, the truth of (7) will depend on the sanity of whoever murdered Smith, whether Jones or not. And in context (d), although the man behaving suspiciously is the speaker referent, the truth of (8) will depend on whether some man or other in a red cap stole the computer.

Against this, many have found reasons for thinking that in contexts like (b) and (d) the speaker referent is also the semantic referent and hence that descriptions are semantically ambiguous after all. Some of these reasons - for example, those arising from failures to describe correctly the object in mind and from the behaviour of descriptions in opaque contexts - have not stood up well. Others seem more promising. First, we are not just able to use descriptions referentially, it seems that we regularly do so. This regularity suggests that there is a convention of so using descriptions. If there is, then it is hard to see why the convention is not semantic. Second, in their referential uses descriptions seem to have roles just like demonstratives: ‘the *F*’ and ‘an *F*’ function like ‘that *F*’, and are similarly based on perception of a particular object. To try to treat these demonstratives like Russellian descriptions would be to give a description theory of them, and we have already noted problems for this (see §5). Third, definite (but not indefinite) descriptions seem to have the same range of anaphoric roles as a pronoun like ‘she’ (see §7). We might then expect them also to share the pronoun’s role as a demonstrative which, as has already been pointed out, seems not to be Russellian. Fourth, consider the utterance, ‘The book is on the table’. In the right circumstances, this will seem true and yet, according to the Russellian view, it must be false: since the world is full of books and tables the two definite descriptions fail to describe unique objects. The obvious modification to save the Russellian view is to treat these ‘incomplete’ descriptions as elliptical. But this modification has problems. A speaker may have many ways to complete the description and there may be no basis for saying that any one is the correct way. Alternatively, trying to complete the description may lead to the familiar problems of ignorance and error.

Another argument against the ambiguity thesis appeals to rigidity. If referential uses of descriptions were semantically significant then, it is claimed, they should be rigid like names and demonstratives. Yet they do not seem to be. Consider the use of ‘Smith’s murderer’ in context (b) above. ‘Smith’s murderer is insane’ does not seem to be true in a world where Smith is alive and well, even if Jones is insane. However, this argument has a problem: the referential use of the demonstrative ‘that murderer’ would equally seem to fail this rigidity test in these circumstances (hence suggesting a need to revise Kaplan’s claims about demonstratives; see §5). Yet a demonstrative surely has a semantically significant referential use. If so, then a description may have one too,

despite not being rigid (see Descriptions).

7 Other terms

Many terms that we have not discussed - like adjectives, adverbs and verbs - are naturally taken to refer. It is certainly no easier to explain reference for these terms than for the terms discussed, but we may hope that doing so will not pose sharply different problems (see Adverbs; Predication).

We must also consider sentential operators like ‘and’ and ‘not’, and quantifiers like ‘every pen’, ‘some stones’, ‘most dogs’ and ‘few bachelors’. Perhaps these should be seen as largely syncategorematic (see Logical constants; Quantifiers). If we are prepared to accept the existence of certain abstract entities, however, we can take these expressions as referential also. Thus we can take ‘and’ as denoting a ‘truth function’ *conjunction* which is such that the sentence ‘*p* and *q*’ is true if and only if ‘*p*’ is true and ‘*q*’ is true. The quantifiers involve a ‘determiner’ and a general term, and can be taken as applying to sets. Thus ‘most dogs’ involves the determiner ‘most’ and the general term ‘dogs’ and can be taken as applying to any set that contains more than half the dogs; and the sentence ‘most dogs bark’ is true if and only if there is such a set and ‘bark’ applies to all its members.

In virtue of what do these expressions have these referents? The most promising answer for the sentential operators has two stages. We start by describing the ‘conceptual role’ of the operator in deductive, inductive and practical inferences. For a token to denote ‘conjunction’ it must have the appropriate conceptual role. But in virtue of what should we assign to a token with that role the denotation ‘conjunction’ rather than, say, ‘disjunction’? Because, under that assignment, deductive inferences are truth-preserving, inductive inferences are reliable, and so on. A similar line is presumably part of the answer for the quantifiers: ‘most dogs’ applies to any set containing most dogs (rather than, say, to any set containing a few cats) partly because of the reference of ‘dogs’ and partly because of the conceptual role of the determiner ‘most’ and the reliability of inferences. A worrying feature of both these answers is that they seem to make widespread irrationality impossible (see Semantics, conceptual role).

Finally, we must consider anaphoric terms. Pronouns, and even definite descriptions, often depend for their reference on other expressions in their verbal context. Thus ‘one’ in ‘John owns a car and Alice owns one too’ is (using Geach’s term) a ‘pronoun of laziness’, going proxy for the noun phrase ‘a car’ in the preceding conjunct. And consider ‘he’ in ‘John is happiest when he is alone’ and in ‘Every man knows a woman that he admires’. In the former sentence ‘he’ is naturally seen as coreferential with ‘John’, in the latter as ‘bound by’ the quantifier ‘every man’ and so functioning like a bound variable in logic.

Geach (1962) has argued that all anaphoric pronouns are either pronouns of laziness or bound by quantifier antecedents. Against this Evans (1985) has argued that some pronouns with quantifier antecedents are unbound. He calls these ‘E-type’. (a) Consider:

(9) Few congressmen admire the president, and they are very junior.

If ‘they’ were bound by ‘few congressmen’, (9) should mean that few congressmen both admire the president and are very junior. But it does not. The first clause of (9) entails that few congressmen admire the president; the second that all of those who do admire him are very junior. (b) If ‘they’ were bound in (9), then we should be able to substitute any quantifier for ‘few congressmen’ and still make sense. But ‘No congressmen admire the president, and they are very junior’ does not make sense. (c) Last, consider:

(10) If many men come to the ball, Mary will dance with them.

The quantifier ‘many men’ could bind ‘them’ only if (10) meant ‘Many men are such that if they come to the ball Mary will dance with them’, which is not the natural reading. As a result of these considerations and others - particularly pronouns in one person’s sentence that are anaphoric on quantifiers in another person’s - it is generally agreed that Evans has identified a distinct type of pronoun.

However, Evans’ view of this type has been challenged. He thinks that the reference of such a pronoun is determined in a Russellian way by a definite description that can be derived from its quantified antecedent; thus, the reference of ‘they’ in (9) is determined by ‘the few congressmen who admire the president’, and that of ‘them’ in (10) by ‘the many men who come to the ball’. Because these are definite descriptions, for (9) to be true *all* the Kennedy admirers must be junior, and for (10) to be true Mary must dance with *all* the men who come. This

consequence does not seem to generalize. ‘Some congressmen admire the president, and they are very junior’ seems to be compatible with some other congressmen admiring him and not being very junior. The problem is more acute in singular cases: ‘Socrates owned a dog and it bit him’ seems to be compatible with Socrates owning another dog which did not bite him. Finally, there are the formidably difficult ‘donkey sentences’: ‘Every man that owns a donkey beats it’ and ‘If John owns a donkey, he beats it’. On one reading, these sentences concern not simply the unique donkey of each donkey owner but *all* the owner’s donkeys (see Anaphora).

8 Naturalizing reference

From a naturalistic perspective, reference must ultimately be explained in scientifically acceptable terms. Attempted explanations have appealed to one or more of three causal relations between words and the world: historical, reliable and teleological.

Historical-causal theory of reference. Kripke, Donnellan and Putnam (see §4) did not claim to be naturalizing reference, but their theories - together with the role that perception of an object may play in determining the reference of demonstratives and referential descriptions (see §§5-6) - suggest the idea that reference might be explained naturalistically in historical-causal terms: a token refers to the object that played the appropriate role in causing it. But this idea, developed by Devitt, faces the *qua*-problem. In virtue of what is ‘Aristotle’, say, perceptually grounded in a ‘whole object’ and not a time-slice or undetached part of the object, each of which is equally present and causally efficacious? The problem is more pressing for natural-kind words. ‘Horse’ is grounded in a few horses. But those objects are not only horses, they are mammals, vertebrates and so on; they are members of very many natural kinds. Indeed, any horse is a member of indefinitely many *non*-natural kinds: it may be a pet, an investment and so on. In virtue of what is ‘horse’ grounded in an object *qua* horse, rather than *qua* mammal, pet, or whatever? So in virtue of what does it refer, as a result of such groundings, to all and only horses rather than all and only mammals, pets, or whatever?

Reliabilist theory of reference. Under the influence particularly of Fred Dretske (1981) and Jerry Fodor (1990), ‘reliabilist’, ‘informational’, or ‘indicator’ theories have been popular. The basic idea is that a token refers to objects of a certain sort because tokens of that type are reliably correlated with the presence of those objects; the tokens are ‘caused by’ those objects. The token ‘carries the information’ that a certain situation holds in much the same way that tree rings carry information about the age of a tree. There is a problem. How can the theory allow for error? Occasionally we see a muddy zebra and wrongly think ‘horse’. So, some zebras are among the things that would cause tokens of ‘horse’. What ‘horse’ is reliably correlated with is really the presence of horses, muddy zebras, the odd cow in bad light and so forth. So according to reliabilism, it should refer to horses, muddy zebras, the odd cow and so on (with the result that it was not wrong to think ‘horse’ after all). The problem is that many things that a token of a certain type does not refer to, including some denizens of Twin Earth, would cause a token of that type (see Semantics, informational).

Teleological theory of reference. Most fully developed by Ruth Millikan (1984), teleological theories explain the reference of a token in terms of its function, where that function is explained causally along Darwinian lines: a token’s function is what tokens of that type do that explains why they exist. This theory deals neatly with the problem of error because something - for example, sperm - can have a function which it does not reliably perform. An immediate consequence of the theory is that a token of a type that has not evolved will lack a referent. So the ‘thoughts’ and ‘utterances’ of an exact replica of Russell created by some cosmic accident would have no reference. This strikes many as implausible but is accepted by the theory’s proponents. To complete the theory it must be shown that tokens - even a belief like ‘computers make writing easier’ which could not plausibly be taken as innate - have a function in the required biological sense and that this function does indeed relate the token to its referent. Millikan has attempted this formidable task (see Semantics, teleological).

9 Further issues

Terms in opaque or intensional contexts cannot be seen as having their usual referential roles. For, in these contexts, particularly those of propositional attitude ascriptions (see Propositional attitude statements), the replacement of a term by a coreferential term may not preserve truth.

There are a range of what might be called ‘negative’ views of reference. (a) Some philosophers have a

‘deflationary’ view according to which there is nothing more to referential notions than is captured by all instances of a schema like ‘*e* designates *a*’, where what is substituted for ‘*a*’ ‘translates’ what is named by the term substituted for ‘*e*’; ‘"Socrates" designates Socrates’ is a typical instance. This view accompanies a similarly deflationary view of truth (see Truth, deflationary theories of). (b) W.V. Quine (1960) argues that even once the translation of a sentence has been fixed the reference of any part of the sentence is inscrutable; thus there is no fact of the matter whether an alien’s ‘Gavagai’ in response to an environment of rabbits refers to rabbits, undetached rabbit parts, time-slices of rabbits, and so on (see Radical translation and radical interpretation §§1-4). Related to this, Donald Davidson (1984) takes an instrumentalist attitude to reference, denying both the need for, and the possibility of, a theory of reference. Putnam (1983) gives a model-theoretic argument that reference is indeterminate because any theory has unintended models. (c) Kripke (1982) presents an argument (which he finds in Ludwig Wittgenstein’s discussion of rule-following) that the meanings and references of terms are not determinate (see Meaning and rule-following). (d) Less sweepingly, Hartry Field (1973) has argued that in some cases there is no determinate matter of fact whether a term refers to one thing or another and we should see it as ‘partially referring’ to both: for example, ‘mass’ as used by Newtonians does not determinately refer to either proper mass or relativistic mass but partially refers to both. (e) Finally, those in the ‘structuralist’ tradition reject reference, and hence its role in meaning, altogether. They apparently think that the only possible theory of reference is one according to which a word resembles what it refers to. But this theory is refuted by the fact that language is arbitrary, by the fact that anything could be used to mean anything. Reference is thus left as simply ‘God-given’, which is unacceptable (see Structuralism in linguistics).

Finally, views on reference can bear on realist notions about the external world (see Realism and antirealism). Putnam, for example, draws antirealist conclusions from his model-theoretic argument. And consider the consequences of a holistic description theory for scientific terms. When, in time, we come to replace one scientific theory with another, it is natural to think that part of the reason we do so is that the theory does not accurately describe reality. Combine this thought with the holistic view that the reference of each term in the theory is determined by its associations with all other terms in the theory, and we get the consequence that all terms in the theory fail to refer. So, it was a mistake to believe in the entities apparently referred to by that theory. Worse, it is probably a mistake to believe in the entities of our present theory, for that theory will surely be replaced in time too. So we should not be scientific realists. Indeed, these considerations lead Thomas S. Kuhn (1970) and others to constructivism, a radically relativistic antirealism: rather than saying that the replaced theory does not describe reality, they say that it describes *its* reality, a reality that only exists relative to that theory (see Constructivism). Each theory has its own reality and no sense can be made of scientific entities existing ‘absolutely’. This line of thought can be resisted by rejecting the holistic description theory of reference in favour of a localist theory, perhaps one explaining reference in terms of causal relations to reality (see Holism: mental and semantic).

See also: Semantics

MICHAEL DEVITT

References and further reading

Burge, T. (1979) ‘Individualism and the Mental’, in P.A. French, T.E. Uehling, Jr. and H.K. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 10, *Studies in the Philosophy of Mind*, Minneapolis, MN: University of Minnesota Press.(Argues that reference for many terms is determined by the social context; referred to in §3.)

Chastain, C. (1975) ‘Reference and Context’, in K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science*, vol. 7, *Language, Mind and Knowledge*, Minneapolis, MN: University of Minnesota Press.(A long paper on singular reference and anaphora. A source for the view that indefinite descriptions have two uses.)

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 215-25.(Argues against the need for, and possibility of, theories of reference; referred to in §9.)

Devitt, M. (1981) *Designation*, New York: Columbia University Press.(Develops a causal-historical theory of proper names, demonstratives and referential descriptions along naturalistic lines; referred to in §§4 and 8.)

Devitt, M. and Sterelny, K. (1998) *Language and Reality: An Introduction to the Philosophy of Language*, Oxford: Blackwell, 2nd edn.(Provides an introductory discussion of reference and of its significance for issues such as realism. Also includes an annotated guide to further reading.)

Donnellan, K.S. (1966) ‘Reference and Definite Descriptions’, *Philosophical Review* 75: 281-304.(Influential work proposing a distinction between referential and attributive descriptions and criticizing Russell and Strawson; referred to in §6.)

Donnellan, K.S. (1972) ‘Proper Names and Identifying Descriptions’, in D. Davidson and G. Harman (eds) *The Semantics of Natural Language*, Dordrecht: Reidel.(Criticizes the description theory of proper names and proposes a historical theory; referred to in §§2 and 4.)

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(Most influential proposal of a reliabilist theory of reference; referred to in §8.)

Dummett, M.A.E. (1973) ‘Appendix: Note on an Attempted Refutation of Frege’, in *Frege: Philosophy of Language*, London: Duckworth.(A defence of a Fregean view of proper names from Kripke’s criticisms; referred to in §2.)

Evans, G. (1982) *The Varieties of Reference*, ed. J. McDowell, Oxford: Clarendon Press.(Defends ‘Russell’s Principle’ that to think about an object a person must know which object is in question.)

Evans, G. (1985) *Collected Papers*, Oxford: Clarendon Press.(Chapters 4, 5 and 8 argue for the view that some anaphoric pronouns are ‘E-type’ rather than bound; referred to in §7.)

Field, H. (1973) ‘Theory Change and the Indeterminacy of Reference’, *Journal of Philosophy* 70: 462-81.(Argues that the reference of scientific terms is sometimes indeterminate and that they should be seen as ‘partially referring’; referred to in §9.)

Fodor, J.A. (1990) *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press.(Chapters 3 and 4 propose a reliabilist theory of reference appealing to ‘asymmetric counterfactuals’; referred to in §8.)

Frege, G. (1893) ‘On Sense and Reference’, in P. Geach and M. Black (eds) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 1952.(Classic work raising the problem of identity statements for a Millian theory of proper names and proposing the theory that names have ‘senses’; referred to in §1.)

Geach, P. (1962) *Reference and Generality*, Ithaca, NY: Cornell University Press.(Argues that all anaphoric pronouns are either like bound variables or are ‘pronouns of laziness’; referred to in §7.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.(Proposes the distinction between speaker meaning and semantic meaning that has been influential in discussions of definite descriptions; referred to in §6.)

Husserl, E. (1900-1) *Logical Investigations*, trans. J.N. Findley, London: Routledge & Kegan Paul, 1970, part 1, §26; part 6, §§1-5.(Proposes a theory of demonstratives based on perception; referred to in §5.)

Kaplan, D. (1989) ‘Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals’, in J. Almog, J. Perry and H. Wettstein (eds) *Themes from Kaplan*, New York: Oxford University Press.(Influential work arguing that the reference of a demonstrative is ‘direct’ not via a Fregean sense; referred to in §5.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.(Very influential work criticizing the description theory of proper names and natural-kind terms and proposing a causal ‘picture’; first presented in lectures in the late 1960s; referred to in §§2-4.)

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language: An Elementary Exposition*, Cambridge, MA: Harvard University Press.(Argues from a consideration of rule-following that the meanings and references of terms are not determinate; referred to in §9.)

Kuhn, T.S. (1970) *The Structure of Scientific Revolutions*, Chicago, IL: Chicago University Press, 2nd edn.(In this brilliant work on scientific change, a description theory of reference leads to antirealist constructivism; referred to in §9.)

Marcus, R.B. (1961) ‘Modalities and Intensional Language’, *Synthèse* 13: 303-22. (Argues that proper names are Millian and that true identity statements involving them are always necessary; referred to in §2.)

Millikan, R. (1984) *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.(The most detailed proposal of a teleological theory of reference; referred to in §8.)

Mill, J.S. (1843) *A System of Logic*, London: Longmans, 1867.(Book 1, Chapter 2 proposes his theory of both ‘general and singular’ names - referred to in §1.)

Neale, S. (1990) *Descriptions*, Cambridge, MA: MIT Press.(A thorough defence of the Russellian approach to definite descriptions. Includes an excellent bibliography.)

Papineau, D. (1987) *Reality and Representation*, Oxford: Blackwell.(Proposes a teleological theory of reference.)

Putnam, H. (1975) *Mind, Language and Reality: Philosophical Papers*, Cambridge: Cambridge University Press, vol. 2, 196-290.(Very influential argument against description theories and for causal theories of natural-kind and other terms; referred to in §§3 and 4.)

Putnam, H. (1983) *Realism and Reason: Philosophical Papers*, Cambridge: Cambridge University Press, vol. 3, 1-25.(Presents his model-theoretic argument that reference is indeterminate; referred to in §9.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Section 12 argues for the inscrutability of reference in the context of arguing for the indeterminacy of translation; referred to in §9.)

Reichenbach, H. (1947) *Elements of Logic*, London: Macmillan.(Section 50 proposes a theory of 'token-reflexive' words; referred to in §5.)
Russell, B. (1905) 'On Denoting', *Mind* 14: 479-93.(A classic paper proposing his famous theory of descriptions; referred to in §6)
Russell, B. (1911) 'Knowledge by Acquaintance and Knowledge by Description', in *Mysticism and Logic*, London: Allen & Unwin, 1917.(A classic source for the description theory of proper names; referred to in §1.)
Russell, B. (1912) *The Problems of Philosophy*, London: Oxford University Press, 1959. (Chapter 5 gives a brief and accessible account of Russell's theory of proper names and descriptions.)
Russell, B. (1918) 'The Philosophy of Logical Atomism', in R.C. Marsh (ed.) *Logic and Knowledge*, London: Allen & Unwin, 1956.(Discusses the referential properties of many terms, including 'egocentric particulars'; referred to in §§1, 5 and 6.)
Searle, J. (1958) 'Proper Names', *Mind* 67: 166-73.(A very clear presentation of a cluster description theory of names; referred to in §1.)
Strawson, P.F. (1950) 'On Referring', *Mind* 59: 320-44.(A critique of Russell's theory of descriptions.)
Strawson, P.F. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen, 180-3, 190-4. (Proposes a cluster description theory of names; referred to in §1.)

Reichenbach, Hans (1891-1953)

Philosophy of science flourished in the twentieth century, partly as a result of extraordinary progress in the sciences themselves, but mainly because of the efforts of philosophers who were scientifically knowledgeable and who remained abreast of new scientific achievements. Hans Reichenbach was a pioneer in this philosophical development; he studied physics and mathematics in several of the great German scientific centres and later spent a number of years as a colleague of Einstein in Berlin. Early in his career he followed Kant, but later reacted against his philosophy, arguing that it was inconsistent with twentieth-century physics.

Reichenbach was not only a philosopher of science, but also a scientific philosopher. He insisted that philosophy should adhere to the same standards of precision and rigour as the natural sciences. He unconditionally rejected speculative metaphysics and theology because their claims could not be substantiated either a priori, on the basis of logic and mathematics, or a posteriori, on the basis of sense-experience. In this respect he agreed with the logical positivists of the Vienna Circle, but because of other profound disagreements he was never actually a positivist. He was, instead, the leading member of the group of logical empiricists centred in Berlin.

Although his writings span many subjects Reichenbach is best known for his work in two main areas: induction and probability, and the philosophy of space and time. In the former he developed a theory of probability and induction that contained his answer to Hume's problem of the justification of induction. Because of his view that all our knowledge of the world is probabilistic, this work had fundamental epistemological significance. In philosophy of physics he offered epoch-making contributions to the foundations of the theory of relativity, undermining space and time as Kantian synthetic a priori categories.

1 Life

Hans Reichenbach was born in Hamburg, Germany, on 26 September 1891, where he grew up and received his primary and secondary education. In 1910 he went to Stuttgart to study engineering, but found the subject intellectually unsatisfying. He soon shifted to mathematics, physics and philosophy, which he pursued at the Universities of Berlin, Göttingen and Munich under such teachers as Max Born, Ernst Cassirer, David Hilbert and Max Planck. During the First World War he served in the German Signal Corps at the Russian front, but returned to his academic work after suffering a severe illness.

Reichenbach wrote his doctoral dissertation (1916) on the applicability of mathematical probability to the physical world; he received his degree at Erlangen in 1915. In the following year he passed a state examination in physics and mathematics at Göttingen. From then until 1920 he held a regular job in the radio industry, while pursuing academic work on a part-time basis. During that time Einstein offered his first seminar in the theory of relativity in Berlin; Reichenbach attended and established a lasting relationship with him. In 1920 he returned to Stuttgart to teach at the Technische Hochschule. In 1926, due mainly to Einstein's influence, he became Professor of Philosophy of Physics in Berlin, where he remained until Hitler's rise to power in 1933. At that time he fled to Turkey, where he taught philosophy at the University of Istanbul. In 1938 he moved to the University of California at Los Angeles, where he remained until his death from a heart attack on 9 April 1953.

2 Epistemology

Early in the twentieth century, many scientifically minded philosophers, impressed by new developments in logic and other epistemological considerations, maintained that the entire world could be viewed as a logical construction from sensations (see Phenomenalism §1; Neutral monism). This view appealed strongly to some of the logical positivists, especially Rudolf Carnap, who attempted to carry out the construction in detail (1928); it represented the epitome of logical positivism (see Carnap, R. §1). In his review of this book, Reichenbach (1933) praised its precision and scope, but expressed one misgiving. He could see no place for probability in the positivistic approach (see Logical positivism §2).

Reichenbach (1938) developed the issue of probability into a full-scale attack on logical positivism. He elaborated three major points. First, whereas the positivists had insisted that a statement must be fully verifiable in principle to be cognitively meaningful, Reichenbach enunciated a principle that required only probabilistic support (see Demarcation problem §3). Second, he denied that our factual knowledge must be based on incorrigible reports of

sensory experience; he maintained that it is founded on fallible perceptions of middle-sized physical objects and is probabilistic from the ground up. Third, he maintained that we can have probabilistic knowledge of unobservable physical entities. The atom is not a construct out of sense impressions; it is a physical object whose reality we can infer probabilistically (seeScientific realism and antirealism §4). Atoms would exist even if there were no sentient beings in the universe. This was his refutation of logical positivism.

3 Induction and probability

The concept of probability was one of Reichenbach's major concerns throughout his career; his main treatise on the subject is *The Theory of Probability* (1949). He argued that probability should be understood as the limit of a sequence of relative frequencies. Suppose, for example, that we have a coin that can be flipped in the standard way. We begin flipping the coin, and keeping track of the result (heads or tails) of each toss: H T T H T H H H T H…. This generates a sequence of fractions m/n representing the number of times *m* that heads has shown in *n* tosses of the coin: $\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{4}, \frac{2}{5}, \frac{3}{6}, \frac{4}{7}, \frac{5}{8}, \frac{5}{9}, \frac{6}{10}$, …. If this sequence of fractions has a limit, that is, if the members of this sequence become and remain arbitrarily close to some value (say $\frac{1}{2}$) as the sequence is extended indefinitely, then that limit *constitutes* the probability of heads on flips of that coin.

This concept of probability leads to many problems. First, we have no guarantee that such a limit exists; these fractions can continue fluctuating forever without approaching any particular value. Aware of this problem, Reichenbach noted its relationship to Hume's problem of justifying induction. Hume argued correctly that we have no guarantee that nature is uniform - that the future will be like the past. Similarly, since we have no guarantee that our particular sequence of coin tosses will exhibit the statistical regularity of converging to a limit, we have no guarantee that the probability in question exists. Nevertheless, we are justified in *positing* that the observed frequency in an initial section of the sequence approximates the limiting frequency, even though that posit may be incorrect. The justification is pragmatic. First, if no limit exists, we will be wrong in our posits, but no other method would succeed, for there is no limit to be discovered. Second, there may be a limit, but it may be quite different from the frequency in the observed initial portion of the sequence. It follows from the definition of 'limit', however, that persistent use of this method, 'the rule of induction', will eventually lead to posits that are accurate to any desired degree of approximation. In ascertaining probabilities (that is, limiting frequencies) we have everything to gain and nothing to lose by using the inductive method (see Induction, epistemic issues in §2; Inductive inference §§1-2; Probability theory and epistemology §1).

Why should we want to ascertain probabilities? According to Reichenbach, all of our knowledge, scientific or common-sensical, is probabilistic; moreover the frequency interpretation is the only legitimate interpretation of probability. Thus, the ascertainment of probabilities (limiting frequencies) is fundamental to all our knowledge. Unfortunately, as Reichenbach acknowledged, his justification applies to an infinite class of other rules - asymptotic rules - that direct us to posit $m/n + c_n$ where c_n is a 'corrective term' that approaches zero as *n* approaches infinity. He had no satisfactory way of showing that his rule is superior to any other asymptotic rule. The best we can say, perhaps, is that the addition of an arbitrary c_n to the observed frequency would be epistemologically unsupportable.

Reichenbach's claim that the frequency interpretation is the only legitimate one poses two other major problems. First, in many practical situations we must deal with single events rather than long sequences - for example, buying an insurance policy. Reichenbach argued that such cases can be handled by assigning the single case to an appropriate sequence and transferring the probability in the sequence, as a weight, to the single case. This will be a successful strategy if one deals with large numbers of diverse single cases. The second major problem - a special case of the first - is the probabilities of hypotheses. A given hypothesis is either true or false, just as a given toss of a coin comes up heads or tails. How can probabilities of hypotheses be construed in terms of relative frequencies? Reichenbach stated that Bayes' theorem in the probability calculus enables us to handle this problem, but he never clearly explained how.

4 Space and time

Reichenbach wrote four major treatises on space and time (1920, 1924, 1928, 1956); in the first he attempted a compromise between Kant and Einstein, but in the later works he severed his ties with Kant completely. In his

classic work, *Philosophie der Raum-Zeit-Lehre (The Philosophy of Space and Time)* (1928), he argued that the question of the geometrical structure of physical space is a matter to be decided, not a priori, but by empirical science. The situation is, however, rather complicated. The discovery of non-Euclidean geometries early in the nineteenth century did not refute Kant, even after the proof of relative consistency (if the non-Euclidean geometries are self-contradictory then so is Euclidean geometry). Kant had claimed that Euclidean geometry is synthetic; if its denial were logically inconsistent, then it would be analytic. Kant's view had been that Euclidean geometry is epistemologically, not logically, privileged; it is a necessary form of visualization of the physical world. Reichenbach argued in detail that the non-Euclidean geometries are possible forms of visualization. (Recall that Einstein had chosen non-Euclidean geometry to formulate his General Theory of Relativity before 1920.) The question is which geometry provides the most suitable one (see Space; Spacetime).

Certain facts about two-dimensional geometry were well-known in the nineteenth century. For example, in Euclidean geometry the sum of the angles of a triangle is exactly $180°$; in non-Euclidean geometries this sum is either greater or less than $180°$, and the value of the sum differs for triangles of different areas. It might seem that we could ascertain the geometry of physical space by measuring angular sums of triangles with various orientations in three-dimensional space. However, at the turn of the century Henri Poincaré had realized that to determine geometrical features of physical space we must use physical measuring instruments, and the results will depend upon the behaviour of these instruments (Poincaré 1902). According to Reichenbach, Poincaré held that the geometry of space is a matter of pure convention; although we could use any geometry we wished to describe physical space, we actually choose Euclidean geometry because it is simpler than non-Euclidean geometries.

Although Reichenbach has often been called a 'conventionalist', he was actually an anti-conventionalist regarding physical geometry (see Conventionalism §1). To interpret mathematical geometry for application to physical space we must of course adopt certain conventions, such as deciding that a certain type of measuring instrument retains its length unchanged as it is transported from place to place. By calling such decisions 'coordinative definitions' he emphasized their status as conventions. Once such a decision is made, however, the structure of physical space becomes a matter of empirical fact. A different choice of a coordinative definition, employed in the same space, may yield a different geometry, but the descriptions based on these different coordinative definitions are physically equivalent. According to Reichenbach's 'theory of equivalent descriptions', then, one set of equivalent descriptions (rather than a single description) correctly describes physical space, but other logically possible sets of equivalent descriptions give false descriptions of physical space (see Geometry, philosophical issues in).

On one aspect of time, Reichenbach took a more strongly conventionalistic position. According to special relativity, no signal can travel faster than light *in vacuo*. The fact that this speed is finite places limits on our ability to synchronize clocks located at different places in any inertial frame of reference. Given two spatially separated clocks *A* and *B*, if we send a light signal at t_1 from *A* to *B* where it is reflected back to *A*, arriving at t_3, then we can choose any time t_2 between t_1 and t_3 as its time of arrival at *B*. This amounts to saying that, while the round-trip speed of light is a matter of physical fact, the one-way speed of light is (within limits) a matter of convention. An important result by David Malament (1977) indicates that causal constraints imply that the two one-way speeds are equal (see Relativity theory, philosophical significance of §3).

One of Reichenbach's fundamental purposes in developing his theories of space and time was to implement a causal theory. He believed that causal processes, such as light pulses and moving material objects, as well as causal interactions, such as reflections of light pulses by mirrors or collisions of material particles, constitute verifiable features of the world, whereas the abstract structure of spacetime by itself does not. In his posthumously published work, *The Direction of Time* (1956), Reichenbach seeks to ground the direction of time - the difference between past and future - in causal aspects of the universe that are also probabilistic. This book contains pioneering work on the concept of probabilistic causality.

5 Other topics

Reichenbach also made significant contributions to other subjects. A major scientific revolution of the twentieth century was the advent of quantum mechanics. Reichenbach's book on this development, *Philosophic Foundations of Quantum Mechanics* (1944), emphasizes causal anomalies, and introduces a three-valued logic (in which a statement can have the value indeterminate as well as true or false) to deal with these problems (see Many-valued logics; Quantum logic; Quantum mechanics, interpretation of §3). Additional discussions of quantum mechanics

are included in *The Direction of Time*.

Reichenbach's *Elements of Symbolic Logic* (1947), his major work on symbolic logic, contains an extended analysis of conversational language, as well as an analysis of laws of nature, physical modalities (necessity, possibility and impossibility) and counterfactual conditionals (see Laws, natural §2; Counterfactual conditionals). These three topics receive a fuller treatment in his monograph, *Nomological Statements and Admissible Operations* (1954).

Although he never wrote a book on the subject, discussions of causality can be found in his writings throughout his career. He wrote very little explicitly on the nature of scientific explanation, but his work on causality contains many valuable hints (see Explanation §4).

Finally, on the basis of his verifiability criterion of meaning, Reichenbach adopted a non-cognitivist ethical theory in *The Rise of Scientific Philosophy* (1951). For him, ethical utterances were expressions of 'volitional decisions'. Fundamental value judgments cannot be considered true or false, but scientific knowledge and logic are extremely useful in dealing with 'entailed decisions', that is, hypothetical value judgments about the compatibility or incompatibility of various basic value judgments and about the selection of appropriate means to reach desired ends.

See also: Scientific method

WESLEY C. SALMON

List of works

Reichenbach, H. (1977-) *Hans Reichenbach: Gesammelte Werke in 9 Bänden (Hans Reichenbach: Collected Works in Nine Volumes)*, ed. A. Kamlah and M. Reichenbach, Wiesbaden: Friedr. Vieweg and Sohn.(All works appear in German, including those originally written or published in other languages.)

Reichenbach, H. (1978) *Hans Reichenbach: Selected Writings, 1909-1953*, ed. R. Cohen and M. Reichenbach, Dordrecht: Reidel.(All works appear in English; contains an excellent bibliography of Reichenbach's writings.)

Reichenbach, H. (1916) 'Der Begriff der Wahrscheinlichkeit für die mathematische Darstellung der Wirklichkeit' (The Concept of Probability for the Mathematical Representation of Reality), *Zeitschrift für Philosophie und philosophische Kritik* 161: 210-39; 162: 98-112, 223-53.(A reprint of Reichenbach's doctoral dissertation.)

Reichenbach, H. (1920) *Relativitätstheorie und Erkenntnis Apriori*, Berlin: Springer; trans. M. Reichenbach, *The Theory of Relativity and A Priori Knowledge*, Berkeley and Los Angeles, CA: University of California Press, 1965.(Reichenbach's first major challenge to Kant's theory of space and time. Translation contains a comprehensive introductory essay on Reichenbach's philosophy of space and time by M. Reichenbach.)

Reichenbach, H. (1924) *Axiomatik der Relativistischen Raum-Zeit-Lehre*, Braunschweig: Vieweg; trans. M. Reichenbach, *Axiomatization of the Theory of Relativity*, Berkeley and Los Angeles, CA: University of California Press, 1969.(A rather technical work in which Reichenbach attempts to clarify logical features of Einstein's theories by means of an axiomatic approach.)

Reichenbach, H. (1928) *Philosophie der Raum-Zeit-Lehre*, Berlin and Leipzig: de Gruyter; trans. M. Reichenbach, *The Philosophy of Space and Time*, New York: Dover, 1957.(*The* central classic in twentieth-century philosophy of space and time.)

Reichenbach, H. (1933) 'Rudolf Carnap, Der logische Aufbau der Welt' (Rudolf Carnap's The Logical Structure of the World), *Kantstudien* 38: 199-201.(A review of Carnap's *Aufbau* (1928), and Reichenbach's first challenge to logical positivism; it is fully developed in Reichenbach (1938).)

Reichenbach, H. (1935) *Wahrscheinlichkeitslehre (Theory of Probability)*, Leiden: A.W. Sijthoff's Uitgeversmattschappij; 2nd enlarged edn, trans. E.H. Hutton and M. Reichenbach, *The Theory of Probability*, Berkeley and Los Angeles, CA: University of California Press, 1949.(Reichenbach's major treatise on probability and induction.)

Reichenbach, H. (1938) *Experience and Prediction*, Chicago, IL: University of Chicago Press. (Reichenbach's treatise on epistemology, containing his refutation of logical positivism.)

Reichenbach, H. (1944) *Philosophic Foundations of Quantum Mechanics*, Berkeley and Los Angeles, CA: University of California Press.(Reichenbach's attempt to resolve problems of quantum mechanics using the nonstandard (3-valued) logic.)

Reichenbach, H. (1947) *Elements of Symbolic Logic*, New York: Macmillan.(Reichenbach's analysis of conversational language (including verb tenses) in chapter 7 is especially interesting.)

Reichenbach, H. (1951) *The Rise of Scientific Philosophy*, Berkeley and Los Angeles, CA: University of California Press.(Reichenbach's popularized account of his philosophy.)

Reichenbach, H. (1954) *Nomological Statements and Admissible Operations*, Amsterdam: North-Holland. (Reissued in 1976 (see below) and with a new foreword; the title later describes the topics covered in this work.)

Reichenbach, H. (1956) *The Direction of Time*, Berkeley and Los Angeles: University of California Press. (Contains many valuable ideas concerning causality and explanation, along with important approaches to problems of temporal symmetry.)

Reichenbach, H. (1976) *Laws, Modalities, and Counterfactuals*, Berkeley and Los Angeles, CA: University of California Press.(The new foreword attempts to make Reichenbach's convoluted theory accessible to a wide audience of philosophers.)

References and further reading

Carnap, R. (1928) *Der logische Aufbau der Welt (The Logical Structure of the World)*, Berlin-Schlachtenesee: Weltkreis-Verlag; trans. R. George, *The Logical Structure of the World*, Berkeley and Los Angeles, CA: University of California Press, 1967.(Carnap's monumental attempt to implement the programme of logical positivism.)

Grünbaum, A. (1963) *Philosophical Problems of Space and Time*, New York: Alfred A. Knopf; 2nd enlarged edn, Dordrecht: Reidel, 1973.(Detailed and often technical discussions of Reichenbach's views.)

Malament, D. (1977) 'Causal Theories of Time and the Conventionality of Simultaneity', *Nous* 11: 293-300.(A serious challenge to Reichenbach's doctrine of conventionality of simultaneity.)

Poincaré, H. (1902) *La science et l'hypothèse*, Paris: E. Flammarion; trans. *Science and Hypothesis*, New York: Dover, 1952.(Part 2, composed of chapters 3-5, provides the point of departure for twentieth-century discussions of the conventionality of geometry.)

Salmon, W. (1967) *The Foundations of Scientific Inference*, Pittsburgh, PA: University of Pittsburgh Press.(An elementary survey of problems of induction and probability from a Reichenbachian perspective.)

Salmon, W. (ed.) (1979) *Hans Reichenbach: Logical Empiricist*, Dordrecht: Reidel.(A collection of critical essays on Reichenbach, with an extended introductory essay, 'The Philosophy of Hans Reichenbach', by the editor.)

Salmon, W. (1975) *Space, Time, and Motion: A Philosophical Introduction*, Encino, CA: Dickenson; 2nd edn, Minneapolis, MN: University of Minnesota Press, 1980.(An elementary exposition of some of Reichenbach's theories of space and time.)

Reid, Thomas (1710-1796)

Thomas Reid, born at Strachan, Aberdeen, was the founder of the Scottish school of Common Sense philosophy. Educated at Marishal College, Aberdeen, he taught at King's College, Aberdeen until appointed professor of moral philosophy at Glasgow. He was the co-founder of the Aberdeen Philosophical Society or 'Wise Club', which counted among its members George Campbell, John Stewart, Alexander Gerard and James Beattie. His most noteworthy early work, An Inquiry into the Human Mind: Or the Principles of Common Sense attracted the attention of David Hume and secured him his professorship. Other important works are Essays on the Intellectual Powers of Man (1785) and Essays on the Active Powers of the Human Mind (1788).

Reid is not the first philosopher to appeal to common sense; Berkeley and Butler are notable British predecessors in this respect, in the discussions of perception and of free will respectively. It fell to Reid, however, to collect and systematize the deliverances of common sense - the first principles, upon the acceptance of which all justification depends - and to provide adequate criteria for that status. Reid insists we rightly rely on our admittedly fallible faculties of judgment, including the five senses, as well as memory, reason, the moral sense and taste, without need of justification. After all, we have no other resources for making judgments, to call upon in justification of this reliance. We cannot dispense with our belief that we are continually existing and sometimes fully responsible agents, influenced by motives rather than overwhelmed by passions or appetites. In Reid's view major sceptical errors in philosophy arise from downgrading the five senses to mere inlets for mental images - ideas - of external objects, and from downgrading other faculties to mere capacities for having such images or for experiencing feelings. This variety of scepticism ultimately reduces everything to a swirl of mental images and feelings. However we no more conceive such images than perceive or remember them; and our discourse, even in the case of fiction, is not about them either. Names signify individuals or fictional characters rather than images of them; when I envisage a centaur it is an animal I envisage rather than the image of an animal. In particular the information our five senses provide in a direct or non-inferential manner is, certainly in the case of touch, about bodies in space.

Reid thus seems to be committed to the position that our individual perceptual judgments are first principles in spite of his admission that our perceptual faculties are fallible. Moreover, moral and aesthetic judgments cannot be mere expressions of feeling if they are to serve their purposes; a moral assessor is not a 'feeler'. Reid is therefore sure that there are first principles of morals, a view that scarcely fits the extent and degree of actual moral disagreement.

Reid offers alternative direct accounts of perception, conception, memory and moral and aesthetic judgment. He stoutly defends our status as continuing responsible agents, claiming that the only genuine causality is agency and that although natural regularities are held to be causes they cannot be full-blooded causes. Continuing persons are not reducible to material entities subject to laws of nature, (pace Priestley); nor does the proper study of responsible agents belong within natural philosophy. Morals may be adequately systematized on a human rights basis according to which private property is not sacrosanct, once moral judgment is recognised to be based on first principles of morals. Judgments of beauty likewise rest on a body of first principles, even though Reid readily allows that there are no properties that all beautiful objects must have in common.

1 Life

Thomas Reid, the founder of the Scottish school of Common Sense Philosophy, was born in Strachan near Aberdeen, Scotland. His mother, Margaret Gregory, was a member of a very scientifically gifted Scottish family. Her brother, David Gregory, was Savilian professor of Astronomy at Oxford and close friend of Sir Isaac Newton. After a formative period in the grammar school and Marischal College, Aberdeen, where George Turnbull was his regent, marrying, and a period as minister at New Machar, Reid was appointed a regent at King's College, Aberdeen, in 1751. While there he co-founded the Aberdeen Philosophical Society or 'Wise Club' whose members included George Campbell, John Stewart, Alexander Gerard and James Beattie. He also produced his Latin 'Philosophical Orations' and wrote his first major work 'An Inquiry into the Human mind on the Principles of Common Sense', published in 1764. Hume found it challenging and it secured Reid the professorship of Moral Philosophy in Glasgow College that year, as successor to Adam Smith. Here his activities culminated in two major

works after his retirement from teaching in 1780, *'Essays on the Intellectual Powers of Man'*, published in 1785, and *'Essays on the Active Powers of the Human Mind'*, published in 1788, a controversy with Joseph Priestly and a fascinating correspondence, particularly with Lord Kames and James Gregory. He died in 1796, survived by one daughter from nine children.

2 Perception and sensation

Reid attacks the view of David Hume that nothing can ever be present to the mind but an image or perception and that the senses are only inlets through which these images are conveyed. According to Hume, the table which we see, and which appears to diminish as we move further from it, is not the real table that exists independently of us and undergoes no alteration as we move away from it; it is the image of the table that is immediately present to the mind, or seen. Reid claims that we have no good grounds for accepting the existence of such images or ideas, arguing that in perceiving something the mind has an immediate acquaintance with it. Perception of an external object involves some conception or notion of the object and an irresistible conviction of its present existence; the conviction is immediate and not inferential (1785: II.v).

But consider smelling a rose and the agreeable sensation raised when the rose is smelled. Reid allows that he is led, by his nature, to conclude some quality to be in the rose, which is the cause of this sensation. However in sensation there is no object distinct from the act of mind as there is in the cases of perception and memory; in 'I feel pain' the distinction between act and object is merely grammatical. Moreover the firm cohesion of the parts of a body is quite unlike that sensation by which I perceive it to be hard. And so sensations are not suitable as bases for inferences to qualities in objects. Given this firm rejection of the doctrine that we can have no conception of any material thing which is not like some sensation, it is hard to see how sensations can be more than occasional causes of conceptions. Reid even raises doubts about the existence of sensations in certain cases of perception (1764: VI). Thus there is no sensation for what he calls visible figure. The different positions on a visually perceived body facing an observer make a two-dimensional figure with length and breadth only. This visible figure represents the three-dimensional form of the perceived body which has length, breadth and thickness.

Sensation is an important ingredient of Reid's distinction between primary and secondary qualities. He states that our senses give us a direct and distinct notion of the primary qualities of objects, in spite of the obstacles mentioned above, but only a relative and obscure notion of secondary qualities: they are unknown qualities that produce certain sensations in us. Yet Reid freely admits (1785: II.xviii) that in seeing a coloured body the sensation is indifferent, drawing no attention, and that the (unknown) quality which we call its colour is the only object of our attention. Presumably we should follow Reid in distinguishing between the colour of a body, conceived to be a fixed and permanent quality in the body, and the appearance of colour to the eye which may be varied in a thousand ways, while not following him in his view that it is the former that we attend to when we look at a coloured object.

Reid thinks there is no visual perception of spherical objects without perception of their visible figure and colour; these are all that is originally perceived by sight even though we learn to perceive by the eye almost anything, including sphericality, that we can originally perceive by touch. It should be clear that this Berkelian view of vision sits uncomfortably with Reid's position that perception of something includes an immediate conviction of its present existence, even though it enables him to dismiss many cases of fallacy of vision as rash inferences.

Reid does not deny that there is no perception unless some impression is made upon the organ of sense either directly or through a medium which is then communicated to the nerves and by them to the brain (1785: II). However, he insists upon the following points. First, there is nothing more ridiculous than to imagine that any motion or modification of matter should produce thought. Second, when I look upon the wall of my room the wall does not act at all and is incapable of action; rather the perceiving of it is an act or operation in me, albeit an involuntary one.

3 Causality

Reid maintains that only beings which, by their active power, produce some change in themselves or in some other being are properly called causes and agents (1788: I). Moreover, whatever is the effect of active power must be something that is contingent, dependent upon the power and will of its cause. Power to produce an effect supposes

power not to produce it, otherwise it is necessity, which is incompatible with power taken in the strict sense. Reid concedes that there is another sense of the term cause, so well authorized by custom that we cannot always avoid using it. He proposes to James Gregory (see Reid 1895) that we call it the ‘physical’ sense, as when we say that heat is the cause that turns water into vapour. A cause, in this sense, means only something which, by the laws of nature, the effect always follows. There is yet another sense of the term ‘cause’ that Reid allows to have currency and which he distinguishes from the sense just explained. Specifically, laws of nature may be called ‘physical’ causes too; but note that they do not bring about anything unless put to work either by some agent or by some physical cause.

Why does Reid maintain that physical causes are not true causes? One reason is that they cannot be ultimate. Thus, that a body put in motion continues to move until stopped might be viewed as an effect of an inherent property of matter, and so as not physically ultimate. If so, we may say that that property of matter is the physical cause of the continuing motion; but surely the ultimate cause of this is the being who gave the inherent property to matter. Alternatively the continuing motion may be viewed as an arbitrary appointment of the Deity, and may then be called a law of nature and a physical cause. But such a law requires a being to enact it.

Must that being be an agent? Reid answers (1788: IV.ii) by invoking as a first principle the thesis ‘that neither existence, nor any mode of existence, can begin without an efficient cause’, from which it follows ‘that everything which undergoes any change must either be the efficient cause of that change in itself, or it must be changed by some other efficient’. Reid adds that we can only conceive kinds of active power similar or analogous to the kind we attribute to ourselves, which is exerted by will and with understanding.

Again, consider the notion of a physical cause as something which, by the laws of nature, the effect always follows. Now given this account it would seem that, on Reid’s view, the existence of such causes is, at best, contingent. Thus he says that the conjunction between a physical cause and its effect must be constant, unless in the case of a miracle, or suspension of the laws of nature. Moreover, processes such as evaporation are frequently interrupted in quite ordinary ways. Finally, it is clear that this account of physical causes will not, of itself, suffice to deal with the problem that Reid left for regularity accounts of physical causation, namely that, according to them, day is the cause of night and night of day.

Let us now turn to consider Reid’s account of how physical causes are to be discovered. Reid (1785: I) adopts a Newtonian view on this question (see Newton, I.), maintaining that if anyone claims to show us the cause of any natural effect, whether pertaining to matter or to mind, we must first consider whether there is sufficient evidence that the cause really exists. If there is not, it should be rejected as a fiction. If it does really exist, we are to consider next whether the effect necessarily follows from it. And, Reid adds, conjectures about causes should be treated as ‘reveries of the vain’. However, writing to Kames (see Home, H.), he allows that such conjectures can have definite value in the following way. Attending to a given phenomenon, I conjecture that it may be owing to a certain cause. This may lead me to make the experiments or observations appropriate for discovering whether that is really the cause or not; and if I discover that it is or that it is not, my knowledge is improved and it is clear then that my conjecture was a means to that improvement.

The establishing of efficient causes is not within the sphere of natural philosophy, the method of which is experimental and leads to the establishing of physical causes at best. Indeed, Reid writes to Kames (see Reid 1965), apart from the fact that our nature leads us to believe ourselves to be the efficient causes of our own actions, and that from analogy we judge the same of other intelligent beings (with the result that attributions of human action are outside the sphere of natural philosophy) we are left, for the establishing of efficient causes, with nothing but resort to such first principles as that every beginning of existence has an efficient cause, and that an effect which has the most manifest marks of intelligence, wisdom and goodness, must have an intelligent, wise and good efficient cause. And with final causes, apparently so pervasive, it is similar (see Causation).

4 Motivation and action

Faced with a deterministic universe the particles of which are initially set in motion by a Deity who prescribes Newton’s laws of motion for them, it is hard to see how anything that happens to particles can be an action of mine. What I call my actions would seem to become mere movements of masses of particles, given that I am such a mass. But even if we do not accept this simple picture, problems remain. To say that people act is, Reid thinks, to

say that we are efficient causes. How might that status be established for us? Reid responds by claiming (1788: I.vi) that it is a first principle that I am the cause that has power to produce certain motions of my body and directions of my thought.

Reid admits the fact of an established harmony between our willing certain motions of our bodies and the operations of the appropriate nerves and muscles. Now the willing is an act of the mind; but whether it has any physical effect upon the nerves and muscles, or whether it is only an occasion of their being acted upon by some other efficient cause according to established laws of nature, is not known to us. ‘This may leave some doubt whether we be, in the strictest sense, the efficient cause of the voluntary motions of our own body’ (1788: I.vii). He continues: ‘The man who knows that such an event depends upon his will, and who deliberately wills to produce it is, in the strictest moral sense the *cause* of the event; and the event is justly imputed to him, no matter what physical causes may have concurred in its production.’ However Reid has just shown that it is far from clear that the motion depends upon his will. For he has correctly pointed out that, for all we can know, the volition is at best an occasional cause of the operations of the nerves and muscles. Moreover, he holds that occasional causation, as in the case of night following day, does not of itself amount to physical causation. So unless Reid is willing to allow that dependency can be instanced by occasional causation, there is no clear sense in which the bodily movement depends upon the will.

Reid maintains (1788: IV) that for a bodily movement to be rated as an action of a free agent it is not sufficient that there arise a determination of the will that the movement be produced; it is also required that the agent deliberately willed to produce it. Indeed if they cannot discern one determination to be preferable to another, and thereby exercise their power to choose, their determinations can be neither right nor wrong, wise nor foolish. It is when an agent exercises their active power to determine, or decide upon, their course of action that they act with the liberty of a moral agent. Now the difficulty about whether we are in the strictest sense the cause of so-called voluntary bodily movements emerges once again at the level of voluntary determinations of the will.

The determination of the will the fulfilment of which is my bodily movement must have a cause which had power to produce it. Now the cause is either myself or some other being. But we must ask in what circumstances a determination of will to move my right arm up rates as caused by me. Clearly, first, I must have the power to bring about the determination. Second, the only way in which the exercise of power is conceivable is via a determination of the will. So the only way in which I can conceive of a determination of the will being caused by me is via a prior determination of my will. But what if that prior determination of the will is not brought about by me? Then it would appear that the prior determination - the determination via which my decision to raise my arm up was formed - was either the product of another agent or of myself. In the former case the decision to raise my arm up was not an exercise of moral liberty. Let us remember, however, that it does not follow from the contention that an exercise of active power without will is not clearly conceivable by us that it is impossible.

What is the alternative to the belief that we are efficient causes with moral liberty? We might think that every event, including thoughts and actions, has a physical cause, arguing that if thinking and behaviour are not governed by universal exceptionless laws, they must be random, and consequently that rewards and punishments are without effect. Arguably every deliberate action must have a motive. When there is no conflicting motive this motive must cause us to act. When there are contrary motives the strongest must prevail.

Reid confronts such positions vigorously in (1788: IV.iv). Whether every deliberate action must have a motive depends upon how we understand deliberation. If deliberation means the weighing of motives, then there must be motives - and conflicting ones at that. But if a deliberate action means only an action done through a cool and calm determination, it is not clear that such determinations must be reached by weighing. Consider a case where an end of some importance may be achieved equally well by various means, none of which has any special appeal to the agent. Again the position that when there is a motive on only one side that motive must determine the action ignores such factors as caprice, quite apart from its presumption that motives are the sole causes of actions. But many motives, such as money, or success, are often not even things that exist, but *entia rationis* (mental constructions), which may best be compared to advice or exhortation, leaving the agent still at liberty to choose. However, we can find entities to serve as occasional causes of determinations such as the desire for money.

When it is said of conflicting motives that the strongest always prevails, Reid rightly points out that this cannot be affirmed or denied with understanding until we know what is meant by the strongest motive. Reid allows that

when the contrary motives are of the same kind and differ only in quantity it may be easy to say which is the strongest. Thus a bribe of a thousand pounds is a stronger inducement than one of a hundred pounds. But when the motives are of different kinds, as are money and integrity, we have no rule by which to judge which is the stronger. If we go for prevalence, then the strongest motive would simply be the one that prevailed and the maxim would be trivially true.

These remarks would seem to apply to motives such as hunger and lust, which we have in common with ‘the brutes’. Reid maintains that in their case the strongest motive always prevails, since brutes lack self-command. But this must be so on the trivial interpretation of the maxim. Reid claims that in contrast humankind judges the strength of its appetites by the conscious effort which is needed to resist them. And since it is possible for a person to resist with effort a strong appetite, the strongest appetite does not always prevail (see Desire).

What of rational motives? An interesting example is our good ‘upon the whole’ - that which, taken with all its discoverable connections and consequences, brings more good than ill. Another is our duty. No sooner do we have the former conception than we are ‘led by our constitution’ to seek the good and avoid the ill. So perhaps rational principles of action motivate by themselves. Reid characterises some rational motives as the conviction of what we ought to do in order to achieve some end we have judged fit to be pursued. Such motives may very properly be compared to advocates pleading at the bar. To say that an advocate was the more powerful pleader because judgment was given on their side would be unsatisfactory; once again strength does not guarantee prevalence (see Action).

5 Language and conception

Reid distinguishes proper names and general words in (1785:V). A general word signifies either an attribute or a combination of attributes; a proper name signifies an individual. By affirming a singular substantive (for example, ‘horse’) of an individual we claim it to have all the attributes each of which must be possessed by any member of the class named by its plural (‘horses’). From the infinite number of combinations of attributes that might be formed, we choose for purposes of classification only those that are useful for arranging our thoughts in discourse and in reasoning.

Concerning the conception of individual things that really exist, Reid says (1785: IV), that the things conceived are the *actual* things, which need not act upon him, nor him on them, in order to be conceived. Such conceptions are called ‘true’ when they agree with the thing conceived in respect of its attributes and relations. In ‘general conception’ what is conceived is the meaning of the general term, attribute or combination of attributes - the conception affixed to it by those who best understand the language.

There are conceptions which Reid calls ‘fancy pictures’. Such are the conceptions which Swift formed of Laputa, and Cervantes of Don Quixote and his squire. We can give names to such creatures of the imagination, conceive them distinctly, and reason consequentially concerning them (1785: IV.i), although there are no bearers for these names outside the story. Reid insists that he can conceive a centaur (1785: IV.ii). Conceiving is an operation of the mind. The sole object in this case is a centaur, an animal which has never existed. Reid rightly insists that what he conceives is not the image of an animal but an animal. It is a body of a certain figure and colour, having life and spontaneous motion. It would seem to be an individual nonexistent. Thus, for Reid the question of how conceiving of nonexistents is possible is, arguably, equivalent to that of how the conceiving of nonexistent individuals is possible.

‘Suppose I conceive a triangle, that is, a plain figure terminated by three right lines. He that understands this definition distinctly has a distinct conception of a triangle. But a triangle is not an individual; it is a species. … the thing conceived is general, and cannot exist without other attributes’ (1785: V.ii). A triangle that really exists must have a certain length of sides and measure of angles, a fixed time and place; the definition of a triangle includes neither existence nor any of these other attributes. Such passages suggest that Reid thinks that an *ens rationis* is a combination of attributes lacking certain members. But does Reid really think that Don Quixote is a universal?

Reid readily allows that one may conceive of a plan of government which is never put into practice (1785: V.iv). Such compositions are things conceived in the mind of the author, not individuals that really exist; the same general conception which the author had may be communicated to others by language. This seems to support the view that Don Quixote is general. And a mere combination of attributes lacks real existence, although Reid allows

that we may ascribe existence to a universal in the sense of its being actually instantiated.

Reid insists that we have clear and distinct conceptions of attributes. He certainly maintains that the attributes of individuals are all we can distinctly conceive about them; but he also holds (1785: V.ii) that the whiteness of this sheet is one thing and whiteness another thing, and that, unlike whiteness, it is not a universal. Attributes are expressed by general words. The other class of general terms is of those terms that signify the genera and species into which we divide and subdivide things. This division neglects certain substantive-like expressions such as ‘entity’; if these are genuine substantives they can scarcely be said to signify a finite collection of attributes, where things that possess them all belong to one kind. Reid himself notices that the adjective ‘beautiful’ is applied to things of such different kinds that he is unable to conceive any quality that is in all the different things to which it applies (see Concepts).

6 The beautiful and the sublime

For Reid, taste is a power of the mind by which we are capable of discerning and enjoying both the beauties of nature and whatever is excellent in the fine arts (1785:VIII) (see Artistic taste). In the end, we always judge that there is some real excellence in beautiful objects. In some cases that excellence is distinctly conceived and can be pointed out (for example, a primary quality), while in other cases objects appear beautiful at first encounter without our being able to specify any excellence (for example, a secondary or occult quality).

When confronted by a beautiful object we should distinguish the pleasure produced in us from the quality in the object that causes it. Nevertheless, Reid emphasizes that since beauty is found in things so various and so very different in kind, it is at least difficult to say in what beauty consists. ‘What can it be that is common to the thought of a mind, and the form of a piece of matter, to an abstract theorem, and a stroke of wit?’ If the beauty of a stroke of wit were its excellence, and the beauty of a proof its excellence, there would be a negative answer to this question. However, since ‘excellent’ does not express an attribute or combination of attributes save when conjoined with a genuine substantive, this answer poses no problem for Reid’s account of significance for full-blooded general terms.

Reid holds that the excellence of an object justifies the judgment of beauty. Our taste ought to be accounted most just and perfect when we are pleased with things most excellent of their kind, and displeased with the contrary. Every excellence has a real beauty and charm that makes it an agreeable object to those of good taste. Of course, Reid thinks that since judgments of taste are either true or false, there must be first principles of taste. But does every judgment of beauty depend for its truth on some excellence the object has (pace Kant)?

Reid does indeed allow that some objects appear beautiful at first sight without our being able to specify any excellence that could justify our judgment, citing the plumage of birds and butterflies as examples. But he claims that by a careful examination of such objects we may perhaps discover some real excellence in them. This claim is implausible. For I may rightly exclaim, ‘What a lovely butterfly!’, confronted by a less than perfect specimen. Moreover in some cases, such as the starry heavens, any excellence to be found in the object upon deeper investigation may be in a respect quite unconnected with the features of it that first impressed me. But then if an ugly object shares that excellence, why may it not count as beautiful too?

Might a basis for my exclamation be found in the novelty of seeing such a specimen? We are so constituted, claims Reid, that what is new to us commonly gives pleasure upon that account, provided it be not in itself disagreeable. However, if ‘beautiful’ can be properly used in a case of novelty, Reid is in difficulties. For some uses of ‘beautiful’ would not then serve to represent anything in the object.

Reid avoids such difficulties by accepting the common division of the objects of taste into novelty, grandeur and beauty. In order to understand his views on the nature of beauty we should consider his thoughts on grandeur. True grandeur is such a degree of excellence as is fit to raise enthusiastic admiration; it is found originally and properly in certain qualities of mind, and only derivatively in objects of sense. (One immediately wants to ask why it is not a department of beauty.) There is no grandeur in mere matter. There is, according to Reid, a real intrinsic excellence in such qualities of mind as power, knowledge, wisdom, virtue and magnanimity. These in every degree deserve esteem, but in an outstanding degree merit admiration, and lead to imitation of what is admired. When we contemplate such vast objects as the earth, the sea or the planetary system, it seems absurd to deny there is grandeur in them; yet it deserves to be considered whether all the grandeur we ascribe to such objects is not

derived from some qualities of mind of which they are the effects or signs. Thus a great work such as the *Iliad*, is a work of great power, wisdom and goodness, well contrived for some important end; but power wisdom and goodness belong to the work figuratively, being really in its author.

It cannot be without reason that things as different as a flower and a song are called 'beautiful'. The grand is the proper object of admiration. The beautiful is the proper object of love and kind affection. A distinction can analogously be made between original and derived beauty. Original beauty is to be found in certain qualities of mind, such as innocence, gentleness and humanity, which are amiable from their very nature. Knowledge, good sense, wit and cheerfulness draw our love as do many other useful qualities. Clearly Reid needs to establish the claim that what we call beautiful among humankind, inferior animals and the inanimate, are either signs or expressions of such amiable qualities of mind or else the effects of design, art and wise contrivance (see Beauty; Sublime, The).

7 Moral judgment, rights and right bearers

As via the external senses we have both the original conceptions of the various qualities of bodies and the original judgment that this body is hard, so via our moral faculty we have the conceptions of right and wrong and the original judgment that this conduct is right and that conduct wrong. If moral approbation were simply expression of feeling and did not involve moral judgment, then quite ordinary sentences would have either no meaning or one inconsistent with all rules of grammar and rhetoric in all languages throughout all the ages (1788: V.vii).

Reid offers examples of first principles of morals (1788: V. iv and III.iii.6). The latter examples include the maxim that we ought not to do to another what we should think wrong if done to us in like circumstances. Clearly application of the maxim presupposes some basis for distinguishing right from wrong. Happily other examples such as 'no human being is born for themselves only' contain the basis of many more particular familiar moral directives.

Reid does not claim that all persons must perceive the truth of their first principles of morals. The immature may not reach the moral starting points of the mature. Just as we may rely upon the clear and distinct testimony of our eyes concerning the colours and shapes of the bodies around us, so we may rely upon the clear and unbiased testimony of our moral faculty with regard to what we ought to do. But testimony is not invariably satisfactory. Reid (1788: III.iii.6) claims that someone who in a cool moment does not accept the maxim about not doing to another what is unacceptable to us does not possess a moral outlook. He also claims that one can be misled by prejudice and bad example (III.iii.8). Consider someone brought up to pursue injury with unrelenting malice. If they have the virtues of clemency, generosity and forgiveness laid before them, and are in a suitably calm and fair frame of mind, they can come to see, Reid believes, that it is more noble to subdue vengeance than to destroy an enemy. Finally, since the exercise of moral judgment depends on how actions are classified, we cannot expect moral judgments to coincide among people who classify them differently. Reid recognizes the first point when he remarks that our first moral conceptions are probably acquired by attending coolly to the conduct of others, and by observing what moves our approval or our indignation.

The need for moral instruction justifies systematic presentations of morals. Such presentations are not accounts of the structure of the powers of our mind by which we have our moral conceptions and make moral assessments. Reid prefers a presentation in terms of rights - a system of Natural Jurisprudence (1788: V.iii). Whereas a 'right action' is an action conformable to our duty, a 'right' is a technical term from law and covers all that a person may lawfully do, all that they may lawfully possess and use, and all that they may lawfully claim of any other person.

A system of the rights of persons can be as adequate a presentation of morals as a system of their duties. What I have a liberty right to do, it is everyone's duty not to hinder me from doing. What is my property right, no one ought to take from me; or to molest me in the use and enjoyment of it. And what I have a right to demand of any one it is their duty to perform.

However it may be my duty to do a kindness to a person who has no claim of right to it. Here Reid resorts to the notion of an imperfect right. They are not owed the kindness, but it would be a wrong act on the part of the agent not to do them the kindness. Again, if my neighbour possesses a horse which he has stolen, as long as I am ignorant of the theft it is my duty to treat the horse as I would any lawful possession of theirs. Reid seems inclined to allow that the thief has a claim to the stolen goods while the theft is unknown, calling the right 'external'. But

while I indisputably have the duty not to take the horse until the theft is known, the thief's claim is groundless all along.

Reid is clear that the right of liberty extends to what I am obliged to do, to refraining from all unlawful actions and to all actions that are indifferent. The rights of liberty and property are perfect - their violation is an injury. Some claim, citing competitive sports, that Reid should not say (see 1788: V.iii) that it is the duty of all persons not to hinder me from doing what I have the liberty right to do. Clearly, if I parry and thereby thwart your thrust, I have not violated your right to thrust.

Reid holds that, although perfect, the right of property is subject to limitations and restrictions. Thus the right of an innocent individual to the necessities of life is superior to that which the rich individual has to their honestly acquired riches. Reid argues that the Utopian order of society, in which there is no private property, offers the least temptation to bad conduct (see Duty; Rights).

If someone has a limb cut off they are still the same person (1785: III.iv). The limb is no part of that person with a right to a part of their estate. My personal identity implies the continued existence of that indivisible thing which I call myself. Is it immaterial? Reid says that the changes which in common language are consistent with bodily identity differ merely in number and degree from those that are thought to undermine it; whereas personal identity does not admit of degree. Yet he allows that the testimony of witnesses to the identity of a person is commonly grounded on the same factors as our judgments of bodily identity. Nor does he think that the sole means of establishing that there is a permanent self which has a claim to all its thoughts and actions is its memory. Persons are the immediate objects of love or resentment. They are, in that respect, as far from being immediately present to the mind as other external objects.

However, Reid also argues that when I say, I see, I hear, I feel, I remember, this implies that it is one and the same self that performs all these operations. If the faculty of seeing were in the eye, that of hearing in the ear, and so on, the thinking principle, which I call myself, would be not one but many. And Kames is asked in a letter of 1775 (see Reid 1895), whether, given that we say with Joseph Priestley that mental powers are the result of brain structure, several intelligent beings formed out of Reid's brain and having its structure will all be Reid (see Persons)?

8 Common sense and first principles

Reid is not the first philosopher to have argued that since a view is contrary to common sense it should be rejected. But he is among the earliest post-medievals to reflect systematically on that argument (1785:VI). Common sense is that degree of judgment which is common to those members of humankind with whom we can converse and transact business. The same degree of understanding makes one capable of discovering what is true and what is false in matters that are self-evident. What is contrary to common sense is contrary to what is self-evident.

Sometimes Reid's characterizations of self-evidence are fairly neutral. Thus whenever we examine the evidence of any proposition either we find it self-evident, or it rests upon one or more propositions that support it, which, in turn, rest upon propositions that support them. But we cannot go back in this way to infinity. Clearly this can only come to a halt when we reach propositions which support all that are built upon them but are not themselves supported. Sometimes his characterizations serve as marks of self-evidence. One mark is that the principle be expressed in a proposition which is no sooner understood than believed. This runs the risk of fielding only trifling propositions. Another mark is that of an opinion's appearing so early in the minds of people that it cannot be the effect of education or reasoning. But this applies to a great number of beliefs that are simply false, such as the belief that parents know everything that the child wants to know, as well as to others that are true but need justification. Another mark is universality of an opinion, revealed by the tenor of human conduct. This, Reid thinks, will apply to such opinions as the existence of a material world and that there is a right and a wrong in human conduct, and no doubt many others. Another is the indispensability of the belief to our daily conduct of affairs. Yet another is the ridiculous as applied to the contraries of first principles. And, arguably, by the neutral criterion all the deliverances of the senses and of memory are self-evident, although Reid allows that the faculties in question are fallible. Clearly it is far from obvious that just any person with common sense is competent effectively to employ all of these criteria, although that is not to express doubts about their capacity to acquire the competence.

Reid is inclined to withhold the title of 'axiom' from the individual deliverances of the senses since philosophers

call only necessary truths axioms. He divides his candidates for first principles into first principles of necessary truths and first principles of contingent truths. Such principles appear to serve as starting points from which other truths can be derived - necessary truths from the former kind and contingent truths from the latter. But it is not Reid’s view that necessarily true first principles are intrinsically more certain or more surely known than contingently true first principles. Indeed Reid argues (1785: VII) that probable evidence of the highest degree - consisting of an accumulation of different arguments based on first principles of contingent truths - often leads to a degree of assurance as great as fits a proposition of Euclid.

The first principles of contingent truths concerning the reliability of the senses and of memory, of which Reid offers as direct an account as that (1785: III) for the senses, are almost immediate consequences of the principle that our faculties - senses, memory, consciousness and reasoning - are not fallacious.

In reply to a sceptic who objects that we are prone to error, Reid contends that to claim that our faculties are not fallacious is not to say that they are infallible; only by using them can we establish anything at all, including their proneness to error. But that our faculties are not fallacious is not something that can be established by using them; it is rather a presupposition of their use.

Another sceptic might resolve to withhold assent to anything until it is established that our faculties are not fallacious. Reid remarks that it would be impossible to move them out of this position by argument. Indeed, it is not possible to enter into a dispute with someone who does not share our first principles. For the procedure by which the truth of a proposition is discovered, or its falsehood detected, is to show its necessary connection with first principles or its incompatibility with them; hence one cannot reason with someone who denies one’s first principles. Indeed, if many key terms are implicitly defined by acceptance of first and derived principles it is hard to see how we can even communicate with someone who does not share our first principles.

It might be thought that Reid’s view that first principles are the voice of God should be taken as a possible justification for adhering to them, or for thinking them true. But if it is not self-evident that they are from God (even if it be evident that their acceptance is, in some instances, a part of human nature), it requires use of our faculties to establish the point that they are from God. Only then can that point be used on behalf of their acceptability.

See also: Aberdeen philosophical society; Buffier, C.; Certainty; Common sense school; Commonsensism; Enlightenment, Scottish; Moral sense theories; Primary-secondary distinction

ROGER GALLIE

List of works

Reid, T. (1764) *An Inquiry into the Human Mind, On the Principles of Common Sense*, Edinburgh; repr. as ‘An Abstract of the Inquiry into the Human Mind on the Principles of Common Sense’, ed. D.F. Norton, in S.F. Barker and T. Beauchamp (eds) *Thomas Reid: Critical Interpretations*, Philadelphia, PA: Philosophical Monographs.

Reid, T. (1780) *Lectures on Natural Theology*, repr. in *Thomas Reid’s Lectures on Natural Theology*, ed. E.H. Duncan, Washington, DC: University Press of America, 1981.

Reid, T. (1785) *Essays on the Intellectual Powers of Man*, Edinburgh.

Reid, T. (1788) *Essays on the Active Powers of the Human Mind*, Edinburgh.

Reid, T. (1895) *The Correspondence of Dr. Reid in The Works of Thomas Reid, D.D.*, ed. W. Hamilton, Edinburgh: Thin, 8th edn.

Reid, T. (1937) *Philosophical Orations of Thomas Reid*, ed. W.R. Humphries, Aberdeen: Aberdeen University Press.

Reid, T. (1965) *‘Unpublished Letters of Thomas Reid to Lord Kames’*, ed. I.S. Ross, *Texas Studies in Literature and Language* 7: 17-65.

Reid, T. (1976) *‘An Abstract of the Inquiry into the Human Mind on the principles of Common Sense’*, ed. D.F. Norton, in S.F. Barker and T. Beauchamp (eds) *Thomas Reid: Critical Interpretations*, Philadelphia, PA: Philosophical Monographs.

Reid, T. (1982) *Cura Prima*, ed. D.F. Norton, appended to L. Marcil-Lacoste, *Claude Buffier and Thomas Reid: Two Common Sense Philosophers*, Kingston and Montreal, Que.: McGill-Queens University Press.

Reid, T. (1988) *The Philosophical Orations of Thomas Reid*, ed. D.D. Todd, trans. S.D. Sulivan, Carbondale, IL:

Southern Illinois University Press.
Reid, T. (1990) *Thomas Reid: Practical Ethics*, ed. K. Haakonssen, Princeton, NJ: Princeton University Press.

References and further reading

Dalgarno, M. (1984) 'Reid's Natural Jurisprudence: The Language of Rights and Duties', in V. Hope (ed.) *Philosophers of the Scottish Enlightenment*, Edinburgh, 13-31.(A clear and helpful introductory paper, discussed in 6.)

Dalgarno, M. and Matthews, E. (eds) (1989) *The Philosophy of Thomas Reid*, Dordrecht, Boston, MA and London: Kluwer.(Contains useful papers on perception, sensation, common sense, mind and action, aesthetics, moral and political obligation, as well as material on the historical context and a helpful bibliography. Invaluable for the serious intending student.)

Daniels, N. (1974) *Thomas Reid's Inquiry: The Geometry of Visibles and the Case for Realism*, New York: Franklin.(A simple exposition of Reid's non-euclidean geometry of visibles.)

Ferreira, M.J. (1986) *Scepticism and Reasonable Doubt*, Oxford: Clarendon Press.(Ably expounds and discusses Reid on first principles and probable reasoning.)

Gallie, R.D. (1989) *Thomas Reid and 'The Way of Ideas'*, Dordrecht, Boston, MA and London: Kluwer.(Contains useful discussions of perception, conception, signification, active power, continuity and the self, and first principles.)

Gracyk, T. (1987) 'The Failure of Thomas Reid's Aesthetics', *Monist* 70 (4): 465-82.(A stimulating paper on this subject discussed in 4 and 5. There are other interesting papers on Reid in the same number.)

Haldane, J. (1993) 'Whose Theory? Which Representations?', *Pacific Philosophical Quarterly* 74 (3): 247-57. (Further valuable discussion on Reid and modern representationalism in response to Stecker (1992).)

Kivy, P. (1973) *Thomas Reid's Lectures on the Fine Arts*, The Hague: Martinus Nijhoff.(Although the lecture notes on which the text is based are not in Reid's hand the text is an invaluable source for the serious student of Reid's aesthetics.)

Lehrer, K. (1989) *Thomas Reid*, London and New York: Routledge.(Ingeniously applies a computational mind model approach to Reid in order to illuminate important areas of Reid's thought-perception, memory and conception, first principles and morality. Contains useful bibliographical information. A stimulating analytical introduction to much of Reid's philosophy suitable for advanced undergraduate students and beyond.)

Stecker, R. (1992) 'Thomas Reid's Philosophy of Action', *Philosophical Studies* 66 (2): 197-208.(Able discussion of whether Reid's insights into decision and action require a libertarian framework.)

Stecker, R. (1992) 'Does Reid Reject/Refute the Representational Theory of Mind?', *Pacific Philosophical Quarterly* 73 (2): 174-84.(Argues clearly that Lehrer's contention that there are sufficient materials in Reid to refute modern representationalism is not well founded.)

Rowe, W.L. (1991) *Thomas Reid on Freedom and Morality*, Ithaca, NY and London: Cornell University Press. (Clear and thorough exposition and discussion of Reid, as well as of Locke, Collins and Clarke, in the areas of decision, free action, causation and motivation.)

Schulthess, D. (1983) *Philosophie et Sens Commun chez Thomas Reid*, Berne: Lang.(A clear and readable exposition of nearly all aspects of Reid's philosophy.)

Reinach, Adolf (1883-1917)

Adolf Reinach, a German philosopher of Jewish extraction, was born in Mainz and died on the battlefield in Flanders. He is of principal note as the inventor, in 1913, of a theory of speech acts (or 'social acts' to use Reinach's own terminology) which in some respects surpasses the later work of thinkers such as J.L. Austin and Searle. Reinach was a leading member of the so-called Munich-Göttingen school of phenomenologists who were inspired by Husserl's early realism and who rejected Husserl's subsequent turn to 'transcendental idealism', drawing their inspiration rather from the more analytic orientation of thinkers such as Franz Brentano and other Austrian philosophers.

1 The legal a priori

Reinach and his fellows conceived phenomenology as a method for describing the 'essential structures' manifested in different realms of objects of experience, in terms of the dependent and independent parts and moments which are the constituent elements of such structures (see the third of Husserl's *Logical Investigations* (1900-01)). This method led Reinach to embrace an extreme version of 'apriorism' in legal theory. All creations of positive law, according to Reinach, presuppose certain basic legal or institutional categories prior to any human convention and subject to certain 'material a priori laws', laws which would obtain even though never recognized by any human subject. Examples of such laws are: a promise gives rise to a mutually correlated claim and obligation; a promise gives rise to a defeasible tendency on the part of the promiser to fulfil the content of the promise. Such laws are analogous to the a priori laws which obtain in other spheres - for example, laws to the effect that nothing can be red and green all over, that red is darker than orange, and so on. Only against the background of this fabric of basic legal structures and associated laws, Reinach argues, can we understand the peculiar phenomena of convention and of positive law-making - phenomena whereby uses of language give rise to certain real changes in the legal and institutional world.

2 Promises and other acts

That uses of language not only can, but even normally do, have the character of actions was a fact largely unrealized by those engaged in the study of language before the present century, not least in virtue of the influence of Aristotle, who in *De Interpretatione* (17a1-5) relegates the study of non-statement-making sentences to the disciplines of rhetoric and poetry. The first philosopher to have fought against the Aristotelian conception seems to have been Thomas Reid, who saw that in addition to judgments there are also other types of sentence permitting of a theoretical treatment (see Schuhmann and Smith 1990). Reid's remarks on social acts were, however, uninfluential, and it was not until the end of the nineteenth century that the idea of linguistic action began to rear its head once more, in the tradition initiated by Brentano and Husserl. Brentano defended the thesis that all mental acts are intentional - that is, that they are directed towards some precisely tailored object. Husserl exploited this thesis in the *Logical Investigations*, above all in the principle: every intentional experience is either an objectifying act or is dependent upon such an act for its existence. Non-objectifying acts such as feelings and emotions, Husserl held, need objectifying acts to supply them with their objects.

Husserl's theory of linguistic meaning is based on this theory of objectifying acts. For Husserl all uses of language approximate to referential uses. More precisely: all expressions are associated either with nominal acts or with acts of judgment. How, on this basis, can Husserl account for the meaningfulness of uses of language involved in asking questions, issuing commands, expressing requests, and so on? Answer: by conceiving the corresponding utterances as disguised statements about certain associated experiences on the part of the language-using subject. The command 'Sit down on the chair!', for example, is a disguised statement to the effect that 'your sitting down on the chair is my current request'. It was in response to the perceived inadequacies of this position that Reinach formulated his theory of social acts.

3 The theory of social acts

The rejection by the Munich phenomenologists of Husserl's theory of linguistic meaning was a cumulative process, instigated by Johannes Daubert (1877-1947) and culminating in the theory of speech acts put forward by Reinach in his monograph on *The A Priori Foundations of the Civil Law* (1913). On the traditional account the

action of promising is seen as the expression of an act of will or as the declaration of an intention to act in the interests of the promisee. Such an account, however, throws no light on how an utterance of the given sort can give rise to a mutually correlated obligation and claim. The bare intention to do something has, after all, no quasi-legal consequences of this sort, and it is difficult to see why things should be different just because such an intention is brought to expression in language.

Both promising and communicating one’s intention to do something, according to Reinach, belong to the category of what he calls ‘spontaneous’ acts - that is, acts which involve a subject’s bringing something about within his own psychic sphere, as contrasted with passive experiences of, say, feeling a pain or hearing an explosion (Reinach [1913: 706] 1983: 18). Certain specific types of spontaneous acts require as a matter of necessity a linguistic utterance or some other overt performance of a non-natural (rule-governed) sort. This does not hold of judging or deciding, nor even of forgiving, but it does hold of apologizing, commanding, accusing. We may accordingly divide spontaneous acts into two classes, which we might call ‘internal’ and ‘external’, according to whether the act’s being brought to overt expression is a separable or inseparable moment of the relevant complex whole.

Acts are divided further into self-directable and non-self-directable. Self-directable acts are such that the subject towards whom they are directed may be identical with the subject of the act (as in cases of self-pity, self-hatred and so on). Non-self-directable acts, on the other hand, such as forgiving or praying, demand an alien subject. A peculiarity of certain external and non-self-directable acts is that the relevant utterance must of necessity not only be directed towards a certain subject but also registered or grasped by this subject in a further act: a command must be received and understood by those to whom it is addressed (something which does not apply, for example, to an act of blessing or cursing). What has been said of commands holds also for requests, admonishments, questionings, informings, answerings, and many other types of act (Reinach [1913: 707] 1983: 19). Each such ‘social’ act constitutes a unity of deliberate execution and utterance; the utterance is not something added as an incidental extra. Certainly there exist also incidental statements relating to social acts: ‘I have just issued the command’. But such statements then relate to the whole social act, with its external aspect, as an independent unity (Reinach [1913: 708] 1983: 20).

The closeness to J.L. Austin and later speech act theorists is unmistakable. For Reinach, as for Austin, a promise is not just the expression of an act of will; it is a social act with an essential external dimension embracing utterance and execution (see Austin, J.L. §3). Reinach’s treatment of social acts includes also an account of conditional acts, of sham and incomplete and otherwise defective acts, of acts performed jointly and severally, and of that sort of impersonality of social acts that we find in the case of legally issued norms and in official declarations such as are involved in marriage and baptismal ceremonies (see Speech acts).

Reinach’s theory is embedded within a larger theory of legal formations and includes a subtle treatment of the ways in which basic legal categories may become modified in their instantiations as a result of the contingent and pragmatically motivated issuances of the positive law. It is embedded also within a larger (non-Kantian, ontological) theory of the a priori, on the one hand, and of the relation between is and ought, as well as between linguistic phenomena and legal and ethical and institutional phenomena, on the other. His brilliance and originality were greatly admired by early phenomenologists, including Husserl himself, but his influence was cut short by an early death and by a shift of direction in the phenomenological movement away from the logico-linguistic, ontological and psychological concerns manifested by Husserl himself and by his earliest followers in Munich and Göttingen.

See also: Phenomenological movement; Searle, J. §1

BARRY SMITH

List of works

Reinach, A. (1911a) ’Zur Theorie des negativen Urteils’, in A. Pfänder (ed.) *Münchener Philosophische Abhandlungen*, *Festschrift* for Theodor Lipps, Leipzig: Barth, 196-254; trans. B. Smith, ‘On the Theory of the Negative Judgment’, in *Parts and Moments*, Munich: Philosophia, 1982, 315-77.(Account of a theory of truth and meaning based on an ontology of states of affairs.)

Reinach, A. (1911b) ‘Kants Auffassung des Humeschen Problems’, *Zeitschrift für Philosophie und philosophische Kritik* 141: 176-209; trans. J.N. Mohanty, ‘Kant’s Interpretation of Hume’s Problem’, *Southwestern Journal of*

Philosophy 7 (1976): 161-88.(Criticism of Kant's misunderstanding of Hume's theory of relations of ideas and defence of an ontological theory of the a priori of which traces are found in Hume.)

Reinach, A. (1913) 'Die apriorischen Grundlagen des bürgerlichen Rechtes', *Jahrbuch für Philosophie und phänomenologische Forschung* I/2: 685-847; trans. as 'The Apriori Foundations of the Civil Law', *Aletheia* 3 (1983): 1-142.(First systematic theory of speech acts, with special attention to the structure of the promise, presented as part of a general theory of the foundations of law.)

Reinach, A. (1989) *Sämtliche Werke: Kritische Ausgabe mit Kommentar (Diverse Works: Critical Edition With Commentary)*, ed. K. Schuhmann and B. Smith, Munich and Vienna: Philosophia, 2 vols.(Critical edition of Reinach's works.)

References and further reading

Husserl, E. (1900-1) *Logische Untersuchungen*, Halle: Niemeyer; repr. in *Husserliana*, vol. 17, ed. E. Holenstein, 1975, and vols 19 (1) and 19 (2), ed. U. Panzer, 1984, The Hague: Nijhoff; trans. J.N. Findlay, *Logical Investigations*, London: Routledge & Kegan Paul, 1970.(Husserl's *magnum opus*.)

Mulligan, K. (ed.) (1987) *Speech Act and Sachverhalt: Reinach and the Foundations of Realist Phenomenology*, The Hague: Nijhoff.(Essays on Reinach's work, particularly on his philosophy of language and law, with extensive bibliography.)

Pfänder, A. (ed.) (1911) *Münchener Philosophische Abhandlungen*, *Festschrift* for Theodor Lipps, Leipzig: Barth. (Early collection of papers by leading Munich phenomenologists.)

Schuhmann, K. and Smith, B. (1990) 'Elements of Speech Act Theory in the work of Thomas Reid', *History of Philosophy Quarterly* 7: 47-66.

Smith, B. (1993) 'An Essay on Material Necessity', *Canadian Journal of Philosophy*, supplementary vol. 18: 301-22.(Comparison of the views of Hume, Reinach and Searle on speech acts and the is-ought problem.)

Spiegelberg, H. (1960) *The Phenomenological Movement: A Historical Introduction*, The Hague: Nijhoff, 2 vols. (Extensive historical survey.)

Reincarnation

The doctrine of reincarnation teaches that each human being has been born and died, and again been born and died, over and over again in a beginningless process that will never end unless they become enlightened. The doctrine of karma asserts that right and wrong actions bring, respectively, positive and negative consequences. For monotheistic religious traditions that accept reincarnation and karma, each person beginninglessly depends on God, and karmic consequences are under God's providential control; repentance and faith may lead to God graciously cancelling negative consequences. Nonmonotheistic religious traditions that embrace reincarnation and karma doctrine see karma as operating in terms of what is, in effect, a moral version of natural or causal law. Both sorts of religious tradition view escape from the reincarnation cycle - 'the wheel' - as the ultimate goal of one's existence and as possible only if one can escape from having karmic consequences still coming at one of one's deaths. Monotheistic traditions see escape as continuance of personal identity, and living in the presence of God (as in monotheistic Hinduism). Nonmonotheistic traditions range from seeing escape as continuance of personal identity in a disembodied condition of omniscience (Jainism, an atheistic religion), loss of all personal identity in entering a changeless nirvāṇa, *or annihilation of all undesirable states but continued existence composed of only desirable states (as in different Buddhist traditions), or simply the realization of identity with a qualityless ultimate reality, so that there only apparently are either persons or reincarnations (as in Mahāyāna voidism and Advaita Vedānta).*

1 Karma and Brahman

Karmic consequences accrue over lifetimes. A right or wrong action performed by an agent *A* in one lifetime may only result in consequences to *A* in *A*'s next lifetime or in *A*'s next lifetime plus one, or later still. One's circumstances of birth (one's sex, caste, economic status, place in society, health, psychological structure, and so on) are functions of one's karmic history. While popularly humans are thought sometimes to be reborn as animals, scholarly Hindu tradition tends to take the idea of a person being reborn as something other than a person as underlining the importance of acting rightly, and accounts of human persons being reborn as animals as stories warning against wrong actions, not literal accounts of the future fate of wicked people.

Within Hindu monotheistic traditions, Brahman or God is held to superintend the reincarnation cycle and its karmic distributions; karma is thought of as God's way of acting justly, parcelling out the appropriate rewards for right action and punishments for misdeeds. Further, repentance of one's sin and trust in God's forgiving mercy is held to provide a way of escape from one's bearing the full brunt of one's negative karma, as well as from the cycle of reincarnation (see Brahman; Karma and rebirth, Indian conceptions of).

Nonmonotheistic South Asian (Indian) traditions - Jainism and, in so far as it takes reincarnation and karma literally, Buddhism - of course have no notions of divine forgiveness or of God providing release from the wheel of rebirth. Their account of what makes escape possible is that if one becomes enlightened in a given lifetime, this signals that at the moment of enlightenment one is free both from karmic credit and karmic debit. Taking one's life would be wrong, so following enlightenment by suicide is not regarded as a solution. One can keep oneself from deserving negative consequences in a future life by not acting wrongly. It is possible, we are told, to perform right actions in a way that does not bring positive consequences, namely by performing them disinterestedly, without emotional investment in their outcome. Thus by maintaining a detached attitude regarding right actions and avoiding wrong actions, an enlightened person can avoid piling up further karmic credits and debits. 'Becoming enlightened' for these traditions is a matter of having an esoteric religious experience; such an experience is a pre-*nirv(ṇa* state - an event in one's current lifetime that promises final *nirv(ṇa* at this lifetime's end (see Nirvāṇa).

2 The importance of escape

One might consider the doctrines of reincarnation and karma as exceedingly good news, since if they are true then apparently one could improve one's overall lifestyle and status by being a morally decent person until one reached a lifetime in which one enjoyed highly satisfactory circumstances. Then, one might think, one could maintain that level of happy life while being born into various healthy bodies and living in sundry pleasant surroundings. Yet any such entrepreneurial alternative is universally rejected by the reincarnation and karma traditions.

There are two reasons why the entrepreneurial perspective is nowhere taken in Indian thought. The more superficial reason is that the nature of reincarnation is not such as to allow one, in one lifetime, to make an efficacious decision regarding one's moral efforts in all future lifetimes. Thus one's decision now to be decent ever after, even if it holds for one's current lifetime, does not guarantee that one will make a similar choice the next time around or guarantee the effectiveness of that new choice should it be made. The more important reason is that human existence within the reincarnation cycle is itself held not to be worthy of one's continued attachment. A Hindu text reads as follows:

> In this body, which is afflicted with desire, anger, covetousness, delusion, fear, despondency, envy, separation from the desirable, union with the undesirable, hunger, thirst, senility, disease, sorrow, and the like, what is the good of the enjoyment of desires…we see that this whole world is decaying…. In this sort of cycle of existence, what is the good of the enjoyment of desires, when after a man has fed on them there is seen repeatedly his return here to earth?… [I]n this cycle of existence I am a frog in a waterless well.
>
> (Maitri Upaniṣad, I, 3-4)

It is basic Theravāda Buddhist doctrine that this world is a place of suffering and 'unsatisfactoriness'. No unenlightened state that a human can be in is satisfactory or worthy of being clung to. The basic religious problem is that of escaping the round of birth and rebirth, and the way to do so is through enlightenment. Theravāda, and Indian Buddhism generally, is nonmonotheistic; there is no doctrine of a creator and providence, and what gods and goddesses there may be are themselves involved in the reincarnation cycle and seeking their own way out.

3 Persons and enlightenment

The Theravāda Buddhist account of what a person is contends that *at* a single time a person is nothing more or other than a bundle of nonconscious and conscious, or physical and mental, states; *over* time a person is nothing more or other than a series of such bundles. An individual person is a set of elements, each momentary and transitory, and everything else is made up of momentary, transitory states. Much of the intellectual context within which Buddhism developed held that each person (*ātman*) is a mental substance that endures throughout its reincarnation history and (should that person become enlightened) on into the enlightenment state. It also held that Brahman is a supreme person who exists everlastingly, is not subject to reincarnation or karma, and gives aid in escaping the reincarnation cycle to those who seek it. For the Buddhist (save for a heretical personalist school) there is no enduring self (*ātman*), nor is there an unchanging ultimate reality (Brahman). It is basic Buddhist doctrine that nothing noncomposite endures even two moments; everything is impermanent. The person, or self, of everyday experience - the 'empirical self' - is held to be unreal or illusory, since we ordinarily or commonsensically take ourselves to be enduring self-conscious things - mental substances - and Buddhist doctrine is adamantly nonsubstantialist. For the Theravādin, there is only a stream of successive, causally linked bundles of momentary physical and mental states. Later Indian Buddhism, for example Mahāyāna, retains the structure of this analysis, but, as an idealist tradition, rejects the view that there are physical states. Causal and (potential) memory connections link one series of bundles in one lifetime to another series in another to constitute reincarnation. Reincarnation is simply a matter of this stream continuing.

If one becomes enlightened, one has a pre-*nirv(ṇa* experience in which one learns the truth concerning 'impermanence' and 'selflessness'; final, postmortem *nirv(ṇa* is the cessation of even this transitory self with consequent release from all desire and unsatisfactoriness. For the Theravāda tradition, the goal is simply to achieve *nirv(ṇa*, and no one who has the sort of pre-*nirv(ṇa* experience that makes final *nirv(ṇa* an option could rationally turn down reaching the final goal upon death. The Mahāyāna tradition, however, holds that one who becomes enlightened in the sense of attaining the relevant pre-*nirv(ṇa* state will put off achieving *nirv(ṇa* upon death and opt to return to the reincarnation cycle in order to help others become enlightened. Final *nirv(ṇa* will be accepted by such a person only when everyone has become enlightened, a condition that for some schools is far off and for others will never arise. Other Mahāyāna schools hold 'voidist' doctrines and contend that even the doctrines of reincarnation and karma presuppose the existence of distinct elements, whereas all reality is seamless; since no distinction that can be made corresponds to anything real, there is no reincarnation cycle and no karma, and coming to see that all is void is *nirv(ṇa* (see Buddhist concept of emptiness).

The Jaina tradition holds that 'modifications cannot exist without an abiding or eternal something - a permanent

substance' (Radhakrishnan and Moore 1957: 269). A 'substance' in this sense is something that has qualities, is not itself a quality or a bundle of qualities, remains the same through change, and endures over time; indeed, on the Jaina view, a substance is indestructible. On this account, then, persons are mental substances, self-conscious souls whose existence neither begins nor ends. The participation of a soul in the reincarnation cycle involves its embodiment, with the undesirable consequence that its inherent omniscience is not realized and it appears to itself to be destructible. By achieving an esoteric religious experience, one can come to a true notion of one's personhood and, upon one's next death, attain *nirv(ṇa*. Both embodiment and karmic consequences, positive and negative, become things of the past, and, with full retention of personal identity (remaining numerically the same self-conscious substance that one has always been), one is freed from all errors and need never return to the wheel. For the Jaina, then, there is an infinite number of real (nonillusory) things which are eternally distinct from one another. Most important among these distinct things are persons (souls, *jīvas*), who can attain enlightenment in which their inherent omniscience is no longer dimmed by embodiment and their individuality is not lost but enhanced. Though the Jainas believe in individual immortality, they do not believe in a supreme deity.

Advaita Vedānta is a nonmonotheistic, monistic variety of Hinduism, according to which only Brahman without qualities exists (see Vedānta). Each person (*ātman*) is identical to the qualityless Brahman. Freedom from the reincarnation cycle involves coming to see that this identity holds. For the Advaitin, ultimate reality (that which does not depend on anything else for its existence) is apersonal and eternal, unchanging and permanent. The real self is identical to this ultimate reality and in a this-life esoteric religious experience, one learns this truth, thereby escaping the cycle of rebirth, losing all individuality, and being 'absorbed' into Brahman. There are obvious similarities between Buddhist 'voidism' and Advaita Vedānta, as was noted by those Hindus who accused Śaōkara, perhaps the greatest of Advaita Vedānta thinkers, of being a crypto-Buddhist. Critics of Advaita Vedānta have argued that if reincarnation and karma are illusions, then there must be someone whose illusions they are, and since by nature Brahman cannot be subject to illusion there must be persons who are not identical to Brahman.

Reincarnation and karma doctrines, then, vary as notions of what a person is and what is ultimately real vary. Agreement on the desirability of escape from the cycle of reincarnation coexists with quite different notions of what sort of thing is in the cycle, what exactly the cycle is and whether the cycle is itself guided by an intelligent and gracious hand.

4 Reincarnation and evil

It is sometimes said that the doctrines of reincarnation and karma make it clear why there is evil in the world and why apparently innocent people suffer. A baby born with disease or defect cannot deserve its fate due to anything it has done in its lifetime. Very bad things happen to very good people and very good things happen to very bad people in ways that seem quite unjust in the light of their current conduct and character. The 'luck of the draw' regarding one's family, intelligence, social status, economic position and the like can powerfully affect the degree to which one is healthy and happy. Wellbeing seems poorly correlated with personal merit or character. In sum, the world plainly seems not to be characterized by a just distribution of goods or of (what might be seen as) rewards and punishments.

If the doctrines of reincarnation and karma are true, however, each person now reaps what they have sown in the past. One deserves whatever one gets, though perhaps not because of anything one has done in one's current lifetime. If a person exists before each of their embodied lifetimes, then the conditions of embodiment and the pleasures and sufferings they undergo may reflect their character in previous lives; so the apparently unjust distribution of goods, or of rewards and punishments, may in fact be exquisitely matched to their deserts.

The claim is that whatever circumstances a person is born into, and whatever befalls that person, are what is morally appropriate in the sense that they are the sort of circumstances and consequences that a morally perfect being, fully informed and capable, would provide them with. For the monotheistic reincarnation traditions, things are this way precisely because such a being has so acted; for the nonmonotheistic traditions, the universe is so structured that this happens by law.

On this account, a person's life-conditions in a given lifetime are a function of their moral conduct and character in a previous lifetime, and so on back forever. The resulting situation is perfectly analogous to that in which the existence of one dependent thing is explained by reference to another, and that to another, without ever coming to

anything that exists independently. In this latter case, one always has, by hypothesis, an answer to ‘What caused this particular thing to exist?’ but never, in principle, an answer to the question ‘Why does anything at all exist?’ In the former (reincarnation and karma) case, there is always, by hypothesis, an answer to the question ‘Why does this particular evil exist?’ but never, in principle, an answer to the question ‘Why is there any evil at all?’ The reincarnation and karma explanation simply assumes the existence of evil, and then goes on to present its blueprint for explaining the fact that evils are distributed as they are. Of course, it involves the general claim that there are such explanations without providing specific explanations in particular cases. Hence no explanation of the fact of evil is forthcoming from a reincarnation and karma perspective.

See also: Salvation; Soul, nature and immortality of the

KEITH E. YANDELL

References and further reading

O’Flaherty, W.D. (1985) *Karma and Rebirth in Classical Indian Traditions*, Berkeley, CA: University of California Press.(Collection on the historical, social, religious and philosophical settings of reincarnation and karma doctrines; includes a long bibliography.)

Radhakrishnan, S. and Moore, C.A. (eds) (1957) *A Sourcebook in Indian Philosophy*, Princeton, NJ: Princeton University Press.(Still the standard anthology of Indian philosophy.)

Smart, N. (1964) *Doctrine and Argument in Indian Philosophy*, London: Allen & Unwin.(Clear general discussion of Indian philosophy.)

Reichenbach, B. (1990) *The Law of Karma*, London: Macmillan.(A philosophical discussion of karma.)

Warren, H.C. (1962) *Buddhism in Translations*, New York: Atheneum Publications. (Standard collection of Buddhist texts.)

Reinhold, Karl Leonhard (1757-1823)

A catalyst in the rise of post-Kantian idealism, Reinhold popularized Kant's critical philosophy by systematizing it in the form of a theory of consciousness. Reinhold shifted from one position to another, however, each time declaring his latest philosophical creed as ultimate. For this he was ridiculed by his more famous contemporaries, including Fichte, Schelling and Hegel, and his historical reputation suffered accordingly. Recent re-evaluations, however, suggest that there was considerable coherence to his philosophical wanderings.

A sometime priest who converted to Protestantism, active freemason and popular teacher, Reinhold advocated political intervention in the promotion of enlightened practices. He steadfastly defended the French Revolution.

A secret promoter of the Austrian Enlightenment while still a priest and monk in Vienna, Reinhold eventually left the Church and fled to Protestant Germany in 1783. There he continued his campaign against 'monkish' obscurantism, mostly in C.M. Wieland's journal, *Der Teutsche Merkur*. Appointed to the chair of Kantian philosophy at Jena in 1787, he later took up a professorship in Kiel (1794), where he remained - popular teacher and voluminous writer - to the end.

In 1784 Reinhold defined the Enlightenment as the effort to accelerate through political and educational means the otherwise slow natural advance of the masses towards rational life. This advance, though retarded by priestly and political reaction, is unstoppable. It is also subject to an inner logic discernible in the history of metaphysics - itself only a reflective, specialized expression of common sense. But metaphysics, Reinhold later repeatedly declared, had run the full gamut of logically possible one-sided views concerning truth. The time was therefore ripe for the revelation of the one true position that would reconcile all past ones.

In 1786-7, this belief led Reinhold, in the midst of the Jacobi-Mendelssohn dispute (see Jacobi, F.H. §2), to portray Kant as the philosopher who, in response to the needs of the time, had cut across old dividing lines with a few new and bold distinctions. Kant was therefore capable of reconciling such extreme opposites as Mendelssohn's rationalism and Jacobi's faith, thereby satisfying the full aspirations of common sense (see Kant, I. §4; Mendelssohn, M. §1).

Reinhold drew this portrait of Kant in *Der Teutsche Merkur*, in eight 'Kantian Letters' which won immediate popularity for critical philosophy and notoriety for himself. However, since Reinhold held that truth cannot but be recognized and accepted as such by all immediately upon display, he had to explain the failure of critique to win universal following. The reason was that Kant - a pioneer - had neither declared his principles explicitly nor systematized his philosophy accordingly. To remedy this failure, in 1789 Reinhold developed a theory of the faculty of representation in which he expounded Kantianism in 88 propositions, all based on a statement of supposedly the most simple fact of consciousness, to wit: 'Representation is distinguished in consciousness by the subject from both subject and object, and is referred to both' (*Beyträg* 1790: 167). Reinhold's new theory, though presented as a systematization of Kant's position, differed from it substantially. Behind its veneer of deductions, it was a description of presumed facts of consciousness - a phenomenology rather than critique. Moreover, Reinhold introduced a form-content distinction at every representational level, and explained each level as the product of a reflection upon a more elementary one - the 'form' in one providing the 'content' of the next.

In two volumes of essays (1790-4) Reinhold sought to refine his *Elementarphilosophie* (Philosophy of the elements), as his theory came to be known. There, and in the second of two other volumes of new 'Kantian Letters' (1790-2) the first incorporating the original eight), he also reacted to recent attacks on critical moral theory. He rejected Kant's identification of 'will' with 'reason', and sought to establish in the faculty of choice the basis of 'personality'. However, Reinhold's efforts in the promotion of Kantianism proved counterproductive, precipitating G.E. Schulze's successful sceptical attack on critical philosophy in 1792.

But Reinhold began to study Fichte's new *Wissenschaftslehre* (which his own theory had influenced) and, for a brief period, even declared himself for it (1797) (see Fichte, J.G. §3). He then tried, in 1799, to position himself between Jacobi and Fichte by defining philosophy as the artificial (religiously neutral) conceptual reconstruction of truths already possessed through faith, only to turn his attention to the 'rational realism' of C.G. Bardili. Bardili held that it is possible by means of a reflection upon pure thought to establish the first predicates of 'being as

such', and the necessity of a First Being. But the abstract categories thereby obtained only yield a science of nature when applied to a matter that must be presupposed independently of all thought. Reinhold refined and defended this dualism in six volumes of essays (1801-3). After 1800, however, he gravitated ever closer to Jacobi, criticizing idealism and expressing sympathy for the synthesis of empirical and critical elements which younger philosophers such as F. Bouterwek, in part inspired by Jacobi, were formulating. Reinhold accepted Jacobi's claim that the first access to truth is through faith, but staunchly defended reason's independent ability to grasp and express this truth scientifically.

Reinhold sided with Jacobi in the latter's final attack on Schelling (see Jacobi, F.H. §3). In 1812, inspired by Jacobi and indirectly by the later J.G. Hamann (§2), he tried his hand at a meta-critique of reason by means of an analysis of language. He attributed all past philosophical errors to confusions (to which reason is held hostage) which are inherent in accepted usages of language. Critique must dispel these confusions. Accordingly Reinhold identified eight 'families of words', in an effort to define the relations of meaning that obtain within each and the irreversible (linear and non-dialectical) order connecting one group to the other. True to his Enlightenment heritage, Reinhold assumed that underlying language is a natural system of universal meanings which now lie hidden under the sedimentation of historically determined usages, but which, once brought to consciousness as they inevitably shall, cannot but be universally recognized and accepted.

See also: Enlightenment, Continental; German idealism

GEORGE DI GIOVANNI

List of works

Reinhold, K.L. (1977) *Karl Leonhard Reinhold: Schriften zur Religionskritik und Aufklärung, 1782-1784*, ed. Z. Batscha, Bremen: Jacobi.(Includes Reinhold's anonymous polemical reviews of theological literature written in 1782-3 while he was still a monk in Vienna, the 1784 essay on Enlightenment and an informative introductory study.)

Reinhold, K.L. (1786-7) 'Briefe über die Kantische Philosophie' (Letters Concerning Kant's Philosophy), *Der Teutscher Merkur* 3 (1786): 99-127, 127-41; 1 (1787): 3-39, 117-42; 2 (1787): 167-85; 3 (1787): 67-88, 142-65, 247-78.(The original eight 'Kantian Letters' that helped popularize Kant's *Critique*.)

Reinhold, K.L. (1789a) 'Über das bisherige Schicksal der Kantischen Philosophie' (On Kantian Philosophy's Present Fortune), *Der Teutscher Merkur* 2: 3-37, 113-35.(An attempt to place Kant's *Critique* in a historical context; republished by Reinhold in various places.)

Reinhold, K.L. (1789b) *Versuch einer neuen Theorie des menschlichen Vorstellungsvermögen (Attempt at a New Theory of the Human Faculty of Representation)*, Prague and Jena: Widtmann & Mauke.(Reinhold's attempt at systematizing Kant's philosophy. Reprinted 1963, Darmstadt: wissenschaftliche Buchgesellschaft.)

Reinhold, K.L. (1790, 1792) *Briefe über die Kantische Philosophie (Letters Concerning Kantian Philosophy)*, Leipzig: Göschen, 2 vols.(Two volumes of twelve 'Kantian Letters' each. The first volume, which incorporates the original eight Letters, deals with theoretical issues; the second, with moral issues. Both volumes republished 1923, ed. R. Schmidt, Leipzig: Reclam.)

Reinhold, K.L. (1790, 1794) *Beyträge zur Berichtigung bisheriger Mißverständnisse der Philosophen (Contributions to the Correction of Present Misunderstandings by the Philosophers)*, Jena: Mauke, 2 vols.(Two volumes of essays - the first containing some refinements of the Theory of Representation and Reinhold's reaction to his critics; the second, reflections on a variety of topics including common sense, scepticism, morality, religion and taste.)

Reinhold, K.L. (1791) *Über das Fundament des philosophischen Wissens (Concerning the Foundation of Philosophical Knowledge)*, Jena: Mauke; republished 1978, ed. W.H. Schrader, Hamburg, Meiner.(The translation of a substantial excerpt is included in G. di Giovanni and H.S. Harris (1985) *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism*, Albany, NY: State University of New York.)

Reinhold, K.L. (1797) *Auswahl vermischter Schriften (Miscellaneous Writings: A Selection)*, Jena: Mauke, 2 vols.(Reinhold declared himself for Fichte's *Wissenschaftslehre* in the Preface of vol. 2.)

Reinhold, K.L. (1799) *Über die Paradoxien der neuesten Philosophie (Concerning the Paradoxes of the Most Recent Philosophy)*, Hamburg: Perthes.(Reflections on post-critical idealism.)

Reinhold, K.L. (1799) *Sendschreiben an J.C. Lavater und J.G. Fichte über den Glauben an Gott (Letter to Lavater and Fichte Concerning Faith in God)*, Hamburg: Perthes. (Reflections on the relation of faith to reason occasioned by the then ongoing dispute regarding atheism. Critical edition in *Fichte-Gesamtausgabe*, ed. R. Lauth and H. Gliwitzky, Stuttgart/Bad Cannstat: Frommann, 1972, 3: 3.)

Reinhold, K.L. (1801-3) *Beyträge zur leichtern Übersicht der Philosophie beym Anfange des 19. Jahrhunderts (Contributions to an Easier Overall View of Philosophy at the Beginning of the Nineteenth Century)*, Hamburg: Perthes, 6 vols.(These volumes include essays on a variety of topical subjects, some polemical tracts against idealism and Reinhold's exposition of Bardili's philosophy.)

Reinhold, K.L. (1812) *Grundlegung einer Synonymik für den allgemeinen Sprachgebrauch in den philosophischen Wissenschaften (Groundwork of a Synonymy for Universal Language Usage in the Philosophical Sciences)*, Kiel: Schmidt.(Reinhold's linguistic turn. Reprint in preparation, Frankfurt/Main: Minerva.)

Reinhold, K.L. (1816) *Das menschliche Erkenntnißvermögen, aus dem Gesichtspunkte des durch die Wortsprache vermittelten Zusammenhangs zwischen der Sinnlichkeit und dem Denkvermögen (The Human Cognitive Faculty, from the Standpoint of the Connection between Sensibility and the Faculty of Thought as Mediated through Language)*, Kiel: Academische Buchhandlung.(Reprint in preparation, Frankfurt am Main: Minerva.)

Lauth, R., Heller, E. and Hiller, K. (eds) (1983) *Karl Leonhard Reinhold: Korrespondenz 1773-1788*, Stuttgart/Bad Cannstatt: Frommann, Holzbook, Österreichische Akademie der Wissenschaften.(The first of a planned six-volume edition of Reinhold's correspondence.)

Schonborn, A. von (1991) *Karl Leonhard Reinhold: Eine annotierte Bibliographie*, Stuttgart/Bad Cannstatt.(A critical bibliography of authenticated works with a very informative introductory study.)

References and further reading

Bardili, C.G. (1800) *Grundriß der ersten Logik gereiniget von den Irrhümmern bisherigen Logiken überhaupt, der kantischen insbesondere; keine Kritik sondern eine Medicina mentis, brauchbar hauptsächlich für Deutschlands kritische Philosophie (Outline of First Logic, Purified of the Mistakes of All Prior Logics, the Kantian in Particular; Not a Critique But a Medicina mentis, of use Chiefly for Germany's Critical Philosophy)*, Stuttgart.(Reinhold studied this text even before its official publication date of 1800. It was instrumental to his conversion to 'rational realism'. Reprinted 1970, Brussels: Culture et Civilisation.)

Beiser, F.C. (1987) *The Fate of Reason*, Cambridge: Harvard.(Includes a very readable chapter on the *Elementarphilosophie*.)

Bondeli, M. (1995) *Das Anfangsproblem bei Karl Leonhard Reinhold (The problem of a beginning in K.L. Reinhold)*, Frankfurt/Main: Vittorio Klostermann.(Scholarly, detailed presentation of the development of Reinhold's philosophy from 1789 to 1803.)

Bouterwek, F. (1799) *Idee einer Apodiktic. Ein Beytrag zur menschlichen Selbverständigung und zur Entscheidung des Streits über Metaphysik, kritische Philosophie und Skepticismus (Idea of an Apodictic. A Contribution to Human Self-Understanding and to the Resolution of the Conflict over Metaphysics, Critical Philosophy and Scepticism)*, Halle: Renger, 2 vols.(An example of traditional empiricism modified by the importation of Kantian critical elements.)

Breazeale, D. (1981-2) 'Between Kant and Fichte: Karl Leonhard Reinhold's "Elementary Philosophy"', *The Review of Metaphysics* 35 (4): 785-821.(Very readable and informative.)

Cloeren, J.-J. (1972) 'Philosophie als Sprachkritik bei K.L. Reinhold. Interpretive Bemerkungen zu seiner Spätphilosophie' (Philosophy as Critique of Language in K.L. Reinhold. Interpretive Comments on His Late Philosophy), *Kant-Studien* 63 (2): 225-36.(An interpretation of Reinhold's late linguistic turn.)

Giovanni, G. di (1985) 'The Facts of Consciousness', in G. di Giovanni and H.S. Harris, *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism*, Albany, NY: State University of New York, 1985, 2-50.(Reinhold in the context of the early reception of Kantian Critique.)

Klemmt, A. (1958) *Karl Leonhard Reinholds Elementarphilosophie (Karl Leonhard Reinhold's Philosophy of the Elements)*, Hamburg: Meiner.(A detailed and readable study of the 'Philosophy of the Elements' as transition point from Kantian Critique to subsequent Idealism.)

Klemmt, A. (1961) 'Die philosophische Entwicklung Karl Leonhard Reinholds nach 1800' (Karl Leonhard Reinhold's Philosophical Development After 1800), *Zeitschrift fur philosophische Forschung* 15 (1-2): 79-101, 250-77.(Reinhold's development after 1800.)

König, A. (1980) *Denkformen in der Erkenntnis. Die Urteilstafel Immanuel Kants in der Kritik der reinen Vernunft und in Karl Leonhard Reinhold Versuch einer neuen Theorie des menschlichen Vorstellungsvermögen (Forms of Thought in Cognition. Immanuel Kant's Table of Judgments in the Critique of Pure Reason and in Karl Leonhard Reinhold's Attempt at a New Theory of the Human Faculty of Representation)*, Bonn: Bouvier.(A technical examination of Reinhold's difference with respect to Kant.)

Lauth, R. (ed.) (1974) *Philosophie aus einem Prinzip Karl Leonhard Reinhold (Philosophy on One Principle: Karl Leonhard Reinhold)*, Bonn: Bouvier.(Seven important specialized essays and a catalogue of letters.)

Pupi, A. (1966) *La formazione della filosofia di K.L. Reinhold, 1784-1794 (The Fashioning of K.L. Reinhold's Philosophy, 1784-1794)*, Milan: Vita & Pensiero.(A careful and readable historical study of Reinhold's Kantian period, with special attention to the debates and the journal literature of the day.)

Reichner, W. (1976) *Rekonstrucktion oder Reproduktion des Grundes. Die Begründung der Philosophie als Wissenschaft durch Kant und Reinhold (Ground Reconstruction v. Ground Reproduction. The Grounding of Philosophy as Science by Kant and Reinhold)*, Bonn: Bouvier.(Kant's and Reinhold's contrasting strategies in the founding of science.)

Schulze, G.E. (1792) *Aenesidemus oder über die Fundamente der von dem Herrn Prof. Reinhold in Jena gelieferte Elementar-Philosophie. Nebst einer Verthedigung des Skepticismus gegen die Anmaßungen der Vernunftkritik (Aenesidemus or Concerning the Foundations of The Philosophy of the Elements Delivered in Jena by Prof. Reinhold, Together with a Defence of Scepticism Against the Pretensions of the Critique of Reason)*, n.p.p.; substantial excerpts trans. G. di Giovanni and H.S. Harris in *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism*, Albany, NY: State University of New York, 1985.(The attack on Critical Philosophy precipitated by Reinhold's attempt at reforming it.)

Valenza, P. (1994) *Reinhold e Hegel. Ragione storica e inizio assoluto della filosofia (Reinhold and Hegel. Historical Reason and Philosophy's Absolute Beginning)*, Padua: CEDAM.(An instructive study of possible points of contact between the philosophy of history of Reinhold and of Hegel.)

Relativism

Someone who holds that nothing is simply good, but only good for someone *or from a certain point of view, holds a relativist view of goodness. Protagoras, with his dictum that 'man is the measure of all things', is often taken to be an early relativist. Quite common are relativism about aesthetic value, about truth in particular areas such as religious truth, and (arising from anthropological theory) about rationality. There are also a number of ways of answering the question 'relative to what?' Thus something might be said to be relative to the attitudes or faculties of each individual, or to a cultural group, or to a species. Relativism therefore has many varieties; some are very plausible, others verge on incoherence.*

1 Grades of relativism

What people believe is affected by their circumstances. A twentieth-century business executive will probably not hold the same views on morality as a medieval peasant, nor a medieval monk the same views on the nature of the physical world as a twentieth-century physicist. An Indian is unlikely to have the same religious opinions as an Italian. This rather commonplace point is hardly exciting enough to deserve the name of 'relativism'. Admittedly, some have concluded from it that there is no one truth about these things, only what seems true from certain perspectives. This argument has proved especially appealing in the case of morality, but it is quite certainly invalid. From the fact that two people hold (even predictably hold) different opinions on a question, it does not follow that there is no truth of the matter - only that, if there is, at least one of them has got it wrong.

But at least the idea that there is no one truth about things sounds closer to something which we might pointfully call 'relativism'. And perhaps there are in some cases (morality looks, *prima facie*, a more promising candidate than physics) better reasons for believing it than this crude and manifestly invalid reasoning. For suppose two people each set themselves to think of a moral code, conformity to which will, in their differing cultural and economic environments, result in as much good as possible - may we not expect that the codes will differ? In widely differing circumstances, the same practices may have widely differing results.

The (moral) relativism which this argument yields is still a rather dilute one. It is fully compatible with there being utterly objective moral principles on which both our protagonists agree. The difference arises only when each asks what these principles dictate when applied to their own situation. Similarly, we can imagine circumstances in which certain procedures were irrational (in the sense of being unlikely to lead to true beliefs) which are rational in ours. In a society whose members were much more wary and suspicious of strangers than they are in our own, it might be irrational to believe information from anybody one does not know. Again, the (epistemic) relativism here is superficial.

A fully-blown relativism denies that there is any such deeper unity beneath the diversity. There is simply what *they* consider right (or rational), and what *we* consider right, what passes for true among them, and what so passes among us.

2 The credibility of relativism

Relativism in its strong form is a version of anti-realism, and how credible it is will depend very much on what we are asked to be relativists about (see Realism and antirealism). Few will have difficulty about 'gastronomic relativism': whether peaches taste nice or not is just a matter of how individuals respond to eating peaches. Many will not object to relativism about colour, according to which the colour of something is a matter of how an observer's visual system responds to it (see Secondary qualities). Even if this does involve accepting that in some sense grass is not 'really' green, we can still react to colours, enjoy certain combinations of colour, attach emotive associations to colours, stop at red traffic lights, and in short go on as we always have.

With moral relativism things are more difficult. To hold that something is wrong for you, or in your society, but perfectly permissible for someone else, or in their society, and that *that is all there is to be said about it*, comes very close to giving up one's own moral view. Are you really to think that there is nothing morally objectionable going on in that other society, just because it is that society? How are you to judge its members if their conduct comes to affect you? An uninvolved Olympian spectator (this may be why relativism sometimes seems easier while we are philosophizing!) might be able to see that there is no more to it than the different reactions and

feelings of members of the two societies, but whether that can be believed from within the melée - even if it is actually true - is quite a different question.

Trickier still is the relativist doctrine that ‘true’ always really means ‘true-by-the-standards-of-X’, where X is some individual or group. An example might be a theory which understands truth in terms of what it is satisfactory to believe (see Truth, pragmatic theory of). Such a theory allows that truth may differ from group to group or from person to person, since social conditions, and individual psychology, may affect the satisfaction to be had from a given belief. Proponents of this kind of theory face various problems: have they anything at all to say to someone who questions whether the standards adopted by their chosen X are good ones? And are they offering their theory of truth as being itself true, or merely ‘true-by-the-standards-of-X’?

One response, favoured by many postmodernists, is to flit ironically from perspective to perspective, espousing none (see Postmodernism). Another response would be to say nothing. How many postmodernists have adopted this latter alternative is, from the nature of the case, hard to determine.

See also: Anthropology, philosophy of; Epistemic relativism; Moral relativism; Pluralism; Protagoras; Rationality and cultural relativism; Social relativism

EDWARD CRAIG

References and further reading

Barnes, J. (1979) *The Presocratic Philosophers*, London and New York: Routledge & Kegan Paul, 1982. (Chapters 23 and 24 discuss some ancient forms of relativism, moral and epistemic, including that ascribed to Protagoras.)

Hollis, M. (1994) *The Philosophy of Social Science*, Cambridge: Cambridge University Press.(Chapter 11 provides an introduction to relativism about rationality. Quite accessible, given that this is not an easy topic.)

Morton, A. (1996) *Philosophy in Practice: An Introduction to the Main Questions*, Oxford: Blackwell.(Written in quite an unusual style for a philosophy textbook, with ‘reader-participation’. See pages 105-23 for an introduction to moral relativism.)

Williams, B.A.O. (1972) *Morality: An Introduction to Ethics*, Cambridge: Cambridge University Press.(Pages 34-9 offer a brisk and hostile introduction to moral relativism.)

Relativity theory, philosophical significance of

There are two parts to Albert Einstein's relativity theory, the special theory published in 1905 and the general theory published in its final mathematical form in 1915. The special theory is a direct development of the Galilean relativity principle in classical Newtonian mechanics. This principle affirms that Newton's laws of motion hold not just when the motion is described relative to a reference frame at rest in absolute space, but also relative to any reference frame in uniform translational motion relative to absolute space. The class of frames relative to which Newton's law of motion are valid are referred to as inertial frames. It follows that no mechanical experiment can tell us which frame is at absolute rest, only the relative *motion of inertial frames is observable. The Galilean relativity principle does not hold for accelerated motion, and also it does not hold for electromagnetic phenomena, in particular the propagation of light waves as governed by Maxwell's equations. Einstein's special theory of relativity reformulated the mathematical transformations for space and time coordinates between inertial reference frames, replacing the Galilean transformations by the so-called Lorentz transformations (they had previously been discovered in an essentially different way by H.A. Lorentz in 1904) in such a way that electromagnetism satisfied the relativity principle. But the classical laws of mechanics no longer did so. Einstein next reformulated the laws of mechanics so as to make them conform to his new relativity principle. With Galilean relativity, spatial intervals, the simultaneity of events and temporal durations, did not depend on the inertial frame, although, of course, velocities were frame-dependent. In Einstein's relativity the first three now become frame-dependent, or 'relativized' as we may express it, while for the fourth, namely velocity, there exists a unique velocity, that of the propagation of light* in vacuo, *whose magnitude c is invariant, that is, the same for all inertial frames. It can be argued that c also represents the maximum speed with which any causal process can be propagated. Moreover in Einstein's new mechanics inertial mass* m *becomes a relative notion and is associated via the equation* $m = E/\mathrm{c}^2$ *with any form of energy* E. *Reciprocally inertial mass can be understood as equivalent to a corresponding energy* $m\mathrm{c}^2$.

In the general theory Einstein ostensibly sought to extend the relativity principle to accelerated motions of the reference frame by employing an equivalence principle which claimed that it was impossible to distinguish observationally between the presence of a gravitational field and the acceleration of a reference frame. Einstein here elevated into a fundamental principle the known but apparently accidental numerical equality of the inertial and the gravitational mass of a body (which accounts for the fact that bodies move with the same acceleration in a gravitational field, independent of their inertial mass). By extending the discussion to gravitational fields which could be locally, but not globally, transformed away by a change of reference frame, Einstein was led to a new theory of gravitation, modifying Newton's theory of gravitation, which could explain a number of observed phenomena for which the Newtonian theory was inadequate. This involved a law (Einstein's field equations) relating the distribution of matter in spacetime to geometrical features of spacetime associated with its curvature, considered as a four-dimensional manifold. The path of an (uncharged spinless) particle moving freely in the curved spacetime was a geodesic (the generalized analogue in a curved manifold of a straight line in a flat manifold).

Einstein's theories have important repercussions for philosophical views on the nature of space and time, and their relation to issues of causality and cosmology, which are still the subject of debate.

1 The Lorentz transformations

Consider two inertial reference frames *K* and K', where the spatial origin of the K'-frame moves along the positive *X*-axis with velocity *v* as viewed from the *K*-frame. If we introduce Cartesian spatial coordinates *x*, *y*, *z* and a time coordinate *t* to locate some arbitrary event relative to the *K*-frame, and similarly use x', y', z' and t' to locate the event relative to the K'-frame, where we suppose the spatial axes are parallel in the two frames, and the origins of the frames coincide at time $t = t' = 0$, then the classical Galilean transformations (see Mechanics, classical §1) read:

$$to\mathrm{x}' = x - vty' = yz' = zt' = t(1)$$

The corresponding Lorentz transformations of special relativity are:

$to x' = \frac{1}{\sqrt{1-v^2/c^2}}(x-vt) y' = y z' = z t' = \frac{1}{\sqrt{1-v^2/c^2}}\left(t - \frac{vx}{c^2}\right) (2)$

where c is the magnitude of the velocity of light *in vacuo*.

For sufficiently small values of v/c the Lorentz transformations are approximately the same as the Galilean transformations.

Einstein argued for (2) by considering the propagation of a spherical light wave centred on the origin of the *K* and K' frames at the time $t = t' = 0$. He employed two principles:

(i) the ‘light constancy principle’, that the velocity of light was a constant independent of the motion of the source; and
(ii) the ‘special principle of relativity’, that the laws of nature are invariant (take the same form) in all inertial reference frames.

From (i) and (ii) Einstein inferred:

(iii) the ‘light invariance principle’, that the velocity of light was the same in any direction and had the same numerical magnitude as viewed from the *K* and K'-frames.

By invoking various other ‘homogeneity’ and ‘symmetry’ properties of space and time, Einstein proceeded to derive the Lorentz transformation equations. Since Einstein’s 1905 paper there have been many other derivations of the Lorentz transformations from various combinations of plausible principles. In particular it has been argued that the principle of relativity only licences a weak form of the light invariance principle, committed just to invariance of the shape of the light wave, not also the magnitude of the light velocity, but it turns out that the Lorentz transformations can still be derived without invoking the strong form of the light invariance principle. Returning to Einstein’s own derivation, the light constancy principle already follows from assuming that light propagates in accordance with Maxwell’s equations (see Maxwell, J.C.; Optics §1; Electrodynamics), and the light invariance principle would follow from assuming the invariance of Maxwell’s equations. But Einstein did not want to assume the full validity of Maxwell’s equations, but rather to base his relativity theory on specific principles that he thought would survive possible alterations to Maxwell’s theory that would be engendered by the quantum theory.

It is easily checked that the light invariance principle is satisfied by the Lorentz transformation equations, whereas the Galilean transformations clearly violate it. It also turns out that the full set of Maxwell’s equations *are* invariant under the Lorentz transformations although, as we have just mentioned, this was not built into Einstein’s derivation.

H. Minkowski (1909) gave a geometrical interpretation of the Lorentz transformations in terms of a four-dimensional geometry equipped with an indefinite metric, which defined the four-dimensional interval between two punctiform events located at x_1, y_1, z_1, t_1 and x_2, y_2, z_2, t_2 relative to the *K*-frame as:

$$s = (t_2 - t_1)^2 - \frac{1}{c^2}\left[(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2\right]$$

The events are said to be time-like separated, light-like separated, or space-like separated according as *s* is greater than zero, equal to zero or less than zero.

The metric corresponding to this interval is said to have Lorentzian signature, which distinguishes it from a four-dimensional *Euclidean* space, where the interval would be positive definite.

The interval between two events is then invariant under the Lorentz transformations. That $s = 0$ corresponds to the fact that the two events are connectable by a light signal. Time-like separated events are connectable by a signal travelling slower than light, while for space-like separated events they can only be connected by signals travelling faster than light. If we adopt a limiting velocity principle and take the velocity of light *in vacuo* to be the maximum velocity with which any signal can be propagated, then space-like separated events cannot be connected by a signal. The status of the limiting velocity principle *vis-à-vis* the light invariance principle will be discussed in §3.

If connectibility by a physically propagating signal is identified with causal connectibility, then the causal structure imposed on events in Minkowski spacetime comprises a double light-cone constructed at any point *P* in spacetime which divides the whole of spacetime into the locations of three classes of event: those that can be causally influenced by an event at *P*, bounded by the future-directed light-cone; the events which can causally effect an event at *P*, bounded by the past-directed light-cone, points on the light-cone being connectable with *P* by light signals; while the third region comprises the locations of events external to the light-cone, which are causally isolated from an event at *P* (see Spacetime §2). These facts about Minkowski spacetime have been employed by philosophers such as Hans Reichenbach and Adolf Grünbaum to provide a causal underpinning for Einstein's light invariance principle (see §3).

2 The relativity of simultaneity, duration and length

From the point of view of the Galilean transformations all inertial reference frames agree on which events are simultaneous with one another. The simultaneity relationship is an equivalence relation which partitions the set of all actual and possible events in spacetime into equivalence classes, which are indexed by the absolute time at which they occur. Absolute time can be understood as the linearly ordered quotient set associated with this equivalence relation, corresponding to Leibniz's order of succession, while the equivalence classes themselves, the spatial slices at a given time, correspond to Leibniz's order of coexistence (see Leibniz, G.W. §11).

In the case of the Lorentz transformations, however, simultaneity becomes a three-place relation, the third place being filled by the inertial reference frame relative to which the simultaneity of events is being assessed. Thus, referring to the transformation equations, (2), the locus of events simultaneous with the origin according to the K'-frame, that is, events satisfying the condition $t' = 0$, correspond according to the *K*-frame to the locus $t = vx/c^2$, which is a straight line inclined at a slope v/c^2 to the *X*-axis, which is the locus $t = 0$ of events judged to be simultaneous with the origin by the *K*-frame (for simplicity we ignore the *y* and *z* spatial dimensions for the time being). This has immediate repercussions for reconciling the *A*-theory of time (see Time §1) with the special theory of relativity. For the *A*-theorists the present is a distinguished simultaneity class of events which alone are co-real. The future comprises unrealized possibilities, the past, events that were actual when their occurrence coincided with the then present, but are no longer actual now. According to special relativity theory, if we assume that the relation of co-reality is *not* frame-dependent, then it is easy to prove that all events in spacetime are co-real. This arises simply because if we take any two events *A* and *B* in spacetime there always exists a possible event *C* which is simultaneous with *A* according to one frame of reference *K* and simultaneous with *B* according to another frame of reference K'. If *A* occurs now, then *C* is co-real with *A* according to *K*; so *C* is now real, while *B* is co-real with *C*, according to K'; so *B* is also real now, but since *B* is an arbitrary event it follows that all events in spacetime are real now.

Such an argument has also been used to claim that special relativity is inconsistent with indeterminism, since the occurrence of all events has a determinate truth-value now. The latter argument can be resisted by distinguishing determinateness from determinism (see Determinism and indeterminism). The former argument shows, however, that special relativity is inconsistent with a unique global frame-independent notion of becoming and hence is fundamentally opposed to the underlying intuition of the *A*-theorists, which requires a uniquely distinguished present. The best that can be done in rescuing a notion of becoming is to relativize it to individual time-like worldlines, but the local 'nows' cannot be assembled into a global 'now'.

We turn now to the relativity of temporal duration. Reverting to the transformation equations, (2), it follows immediately that a time interval T' assessed by a clock at rest in the frame K' is assigned an interval *T* by the frame *K*, where *T* and T' are related by

$$T = \frac{1}{\sqrt{1-v^2/c^2}} T' (3)$$

So the moving clock suffers a dilatation of its time-scale in accordance with equation (3).

But it also follows from (2) that clocks stationary in the *K*-frame suffer the *same* dilatation in time-scale as assessed relative to the K'-frame. This reciprocity of clock dilatation I shall refer to as the *clock paradox* of special relativity. How can two clocks each run slow as assessed by each other? It has even been claimed that the clock paradox shows special relativity to be a logically inconsistent theory. But the resolution is simple when

proper account is taken of the relativity of simultaneity. Each clock is not being assessed by the other clock in isolation, but only in conjunction with the relevant standard of simultaneity, which changes according to which reference frame is being used to assess the rate of the clock.

The clock paradox is closely related to another famous paradox in special relativity, the twin paradox. Here one considers two twins (biological clocks), one of which remains at rest in the *K*-frame, the other moves away and then returns to meet up again with the stationary twin. On comparing clock readings (biological ageing), the stay-at-home twin shows a greater lapse of time than the travelling twin. This follows from the time dilatation formula of equation (3). But from the reciprocity of time dilatation would it not follow that, describing the situation from a reference frame in which the traveller is at rest, the original stay-at-home twin would be the younger when they meet up again? Once more a logical contradiction seems to threaten.

The resolution of the twin paradox arises from noting that there is no single *inertial* reference frame relative to which the travelling twin is always at rest. And the reciprocity result only holds between inertial frames. Does this mean that the resolution of the twin paradox requires consideration of accelerated (that is, non-inertial) reference frames for a complete elucidation, as has been widely claimed in the literature? We shall return to this matter in §3.

Just as temporal duration becomes a relative notion in special relativity, the same is also true of spatial lengths. In place of the dilatation formula, (3), it is easy to derive a corresponding contraction formula for length:

$$L = \sqrt{1 - v^2/\mathrm{c}^2}\,L'$$

where L' is the length of a moving rod as assessed by K', and *L* the length assessed by *K*.

But the relativization of simultaneity, temporal duration and spatial length does not mean that *all* spatial-temporal quantities are relativized to reference frames. The central idea in special relativity is that there is a unique velocity, whose magnitude is the *same* as assessed by all inertial referenced frames. In its geometrical interpretation by Minkowski, the invariant notion in the geometry is the interval *s*, and relativity theory should more accurately be called the invariance theory of the interval, that is, the theory of transformations under which the interval is invariant. The attempt to use Einstein's theory to motivate a general thesis of the relativity of local particulars to their global context, as in the philosophy of Alfred Whitehead, is thus seen to be misguided (see Whitehead, A.N. §3).

The question is often asked, do clocks really run slower, or rods really contract, when they move, or do they only *appear* to do so? The answer according to the presentation of the theory we are adopting in this section (but for a quite different answer see §3), is that if the behaviour of the clocks and rods is governed by the laws of *relativistic* mechanics, then it is a consequence of those laws that clocks will run slower, and rods contract, as a result of the detailed mechanics of what happens when they are suitably set in motion. There is thus a harmonious consistency between relativistic kinematics and relativistic mechanics. But at a deeper level we may ask, are the moving rods and clocks *measuring* relativized features of space and time, so is special relativity a theory about space and time *per se*, or is it a theory about how measuring instruments behave when in motion, and not a revision of our concepts of space and time at all? In particular, does the invariance of the speed of light, as measured by different observers in relative motion to each other, arise as a remarkable conspiracy of compensating effects in the instruments which prevent us from ever being able to know our state of absolute motion? The thoroughgoing relativist rejects such conspiracies, and moves directly to new metaphysical theses about the nature of space and time. The compensating theories were espoused by H.A. Lorentz and H. Poincaré and in that sense the real conceptual revolution must be ascribed to Einstein. But was Einstein genuinely a relativist in the sense we have been describing in this section? In the next section a quite different outlook on the theory will be described, the origins of which can also be traced to Einstein's 1905 paper.

3 Causality and the conventionality of simultaneity

Einstein (1915) recognized that the light invariance principle (see §1) could not be directly verified by measurement. With a single clock it would be possible to measure the average two-way speed of light by emitting a light signal at spatial location *A* (marked as event *a* in the figure) reflecting it at location *B* (event *b*) and noting the time of the event a'' when the signal returns to *A*. If *d* is the spatial separation of *A* and *B*, then the average

two-way speed of light would be equal to $2d/t_{a''} - t_a$, where the time interval $t_{a''} - t_a$ between the events *a* and a'' is measured by a single clock stationary at *A* throughout the experiment. But how could we measure the one-way speed of light from *A* to *B* or from *B* to *A*? This would require recording the time of event *b* by a clock located at *B*. But this means we would first of all have to synchronize the clocks at *A* and *B*. Einstein assumed that this was done by defining *b* to be simultaneous with the event a', midway in time, according to the clock at *A*, between *a* and a''. This is equivalent to assuming that light propagates with the same speed from *A* to *B* as from *B* to *A*. But Einstein states that this is a matter of *definition*, the one-way speed cannot be measured independently of a decision as to how to synchronize spatially separated clocks.

The line of argument was further developed by Hans Reichenbach (1928) who claimed that one could set

$$t_b = t_a + \varepsilon(t_{a''} - t_a)(5)$$

where the parameter εhas any value between 0 and 1, the choice being a matter of convention, the Einstein convention corresponding to the choice $\varepsilon = 1/2$. What Reichenbach is claiming then is that *b* can be identified as simultaneous with any event in the closed interval [*a*, a'']. The conventionality thesis is based, not just on the invariance of the light speed, but on the limiting velocity principle, that light is the fastest possible signal, thus preventing us from using a faster signal to narrow down the conventional latitude [*a*, a''] of events at *A* ascribed simultaneity with the event *b* at *B*.

Adolf Grünbaum expressed the matter somewhat differently by defining a concept of topological simultaneity which relates any two events not connectable by a causal signal. Thus any two events at space-like separation are topologically simultaneous. Grünbaum claimed that any more precise identification of distant simultaneity could have no ontological status, since it could not be specified in purely causal terms.

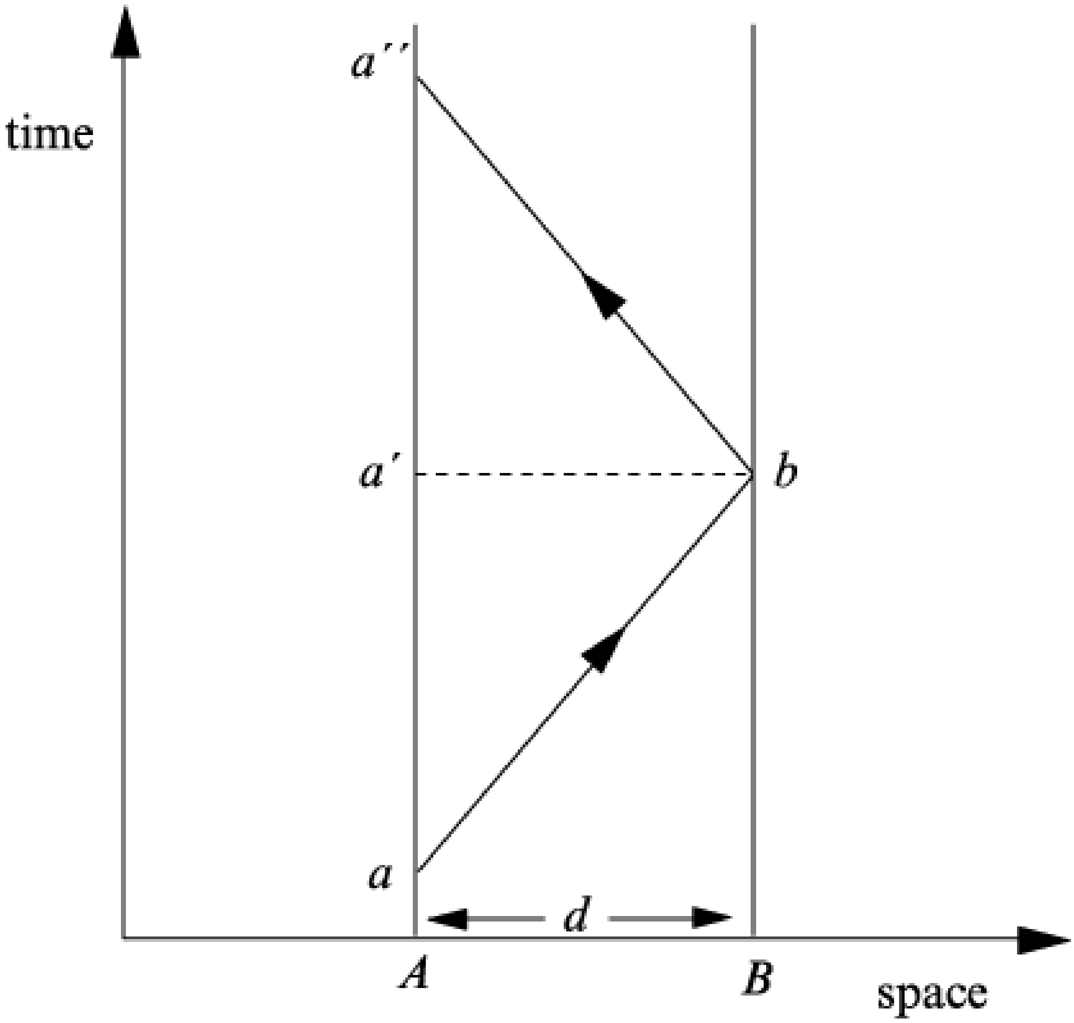

Measuring the speed of light by sending a signal from A to B and back again

It is easy to show that any two events which are topologically simultaneous can be made metrically simultaneous relative to the Einstein convention applied in a suitably moving inertial reference frame. In other words the range of events, such as [a, a''] in the figure which are topologically simultaneous with the distant event (such as b in the figure) in a *given* frame of reference is exactly the same as the range of events that are declared metrically simultaneous with the distant event in *different* frames of reference, according to the relativity of simultaneity (which, of course, follows from the Einstein convention).

By exploiting the conventionality of simultaneity in a given frame, we can actually eliminate the relativity of simultaneity, temporal duration and length as between different frames! So, from this point of view, these effects are not ‘real’, as opposed to effects such as the invariance of the average two-way speed of light which do not depend on any simultaneity convention.

The conventionality thesis has attracted considerable discussion in the literature. First it has been argued that Einstein synchrony *is* causally definable. To explain this, we can note that Einstein synchrony in a given reference frame is equivalent to the locus of events which are orthogonal in Minkowski geometry to the world-line of the origin of the reference frame. There has been a series of results in the literature showing that Minkowski-orthogonality is causally definable. So given the world-line of the origin of a reference frame, then Einstein synchrony in that frame can be defined causally. But the conventionalist response is just to note that relative to a given frame we are free to use the causally defined Einstein synchrony relative to a moving frame, but

as we have seen this is the same as using nonstandard synchrony in the given frame. The same argument can be used to resist the admitted fact that, in the limit of a sufficiently slow transport of a clock, such transport can be used to pick out Einstein synchrony. But again the conventionalist retorts that in a given frame we can use clock transport which is sufficiently slow relative to a moving frame to establish Einstein synchrony relative to that frame, and hence nonstandard synchrony relative to the original frame.

It is a remarkable fact that in four-dimensional Minkowski spacetime the causal structure determines the complete metric structure (*modulo* dilatations). From a point of view which emphasizes the geometrical structure of spacetime the issue of how we coordinatize the spacetime can appear trivial, but from the causal point of view the conventional latitude in simultaneity in a given frame can be regarded as licensing the discordant ascriptions of simultaneity relative to moving frames in the orthodox interpretation.

The limiting velocity principle is a logically stronger principle than the light invariance principle. Indeed the latter is consistent with the possibility of tachyons, particles which *always* travel faster than light. However it has been argued by Redhead that there are inherent limitations in the use of tachyons as a signalling device that arise out of their relativistic properties, and that a suitable emendation of Grünbaum's causal criterion for simultaneity can still yield his same notion of topological simultaneity even in the presence of tachyons.

Finally we may note that the conventionality approach can be usefully applied to the twin paradox, avoiding the use of accelerated reference frames (Redhead 1993).

4 General covariance

We turn now to the development of the general theory of relativity (GR). Historically Einstein believed this to be an extension of the relativity principle from inertial frames to arbitrary reference frames. However subsequent analysis has shown that GR is a quite different sort of theory from special relativity (SR) as we shall now attempt to explain. GR considers spacetime as a semi-Riemannian manifold. By a manifold one means that spacetime can be coordinatized by quadruples of real numbers and this can be carried out in many different ways, different possible coordinates being related by smooth mathematical transformations. This generality of description replaces the coordinates supplied by inertial reference frames and their relationship via the linear transformations of the Lorentz group. Just as in Minkowski spacetime there is defined an invariant interval between events specified by a metric with a Lorentzian signature, so in GR we assume spacetime to be equipped with a variable metric, but still with the property that by a suitable smooth transformation the metric can be reduced to the Lorentzian form at a given spacetime point, but in general this cannot be achieved over a region of spacetime, only at a single point. As first shown by Bernhard Riemann this corresponds to the fact that the manifold is curved as opposed to flat over that region (see Space §4; Spacetime §3). Moreover local inertial reference frames can be introduced relative to which gravitational effects can be (locally) eliminated, and Einstein assumed a generalized equivalence principle according to which *all* physical phenomena reduced to their SR description relative to such frames.

Einstein's field equations can be written symbolically in the form

$$G = \kappa T (6)$$

where *G* is a quantity related to the curvature of spacetime and *T* represents the distribution of matter and electromagnetic energy throughout spacetime; κis a fundamental constant representing the coupling of the geometry to the distribution of mass/energy.

The fact that the field equations can be written in so-called component form using *any* system of coordinates in spacetime was called by Einstein the principle of general covariance. But the fact that we can redescribe spacetime points in many different ways does not of itself have any physical content, as was pointed out by E. Kretschmann in 1917. In no sense is the principle of general covariance a generalization of the principle of Lorentz invariance that arises in SR.

Important clarification of this issue was introduced by J. Anderson in 1967, who expressed covariance and invariance principles in terms of active point transformations on the spacetime manifold *M* instead of the passive coordinate transformations. Smooth point transformations are known to mathematicians as diffeomorphisms. Coordinate transformations induce diffeomorphisms according to the rule that corresponding points have the *same* numerical coordinates in the two coordinate systems. A geometric object field 0 is transformed to a new object

field, the ‘drag-along’ of 0, written d_*0, under the diffeomorphism *d*, where d_*0 evaluated at the transformed point $d(p)$ has the same numerical components with respect to the new coordinate system as does 0 evaluated at *p*, with respect to the old coordinate system.

Anderson partitions the geometric objects into absolute objects A_i which characterize the fixed geometrical spacetime background and dynamical objects P_j which characterize the physical contents of spacetime. The active version of general covariance then states that if $\langle M, A_i, P_j \rangle$ is a model of a spacetime theory *T*, then so also is $\langle M, d_*A_i, d_*P_j \rangle$. The diffeomorphisms are said to comprise the covariance group of *T*. A spacetime symmetry is a diffeomorphism for which $d_*A_i = A_i$, for all values of *i*. SR then is a spacetime theory in which the Minkowski metric is the absolute object, and the spacetime symmetries are the Lorentz transformations that maintain the invariance of the metric. But in GR there are no absolute objects: the geometrical structure characterized by the metric is itself a dynamical object. In a degenerate sense all diffeomorphisms become spacetime symmetries, so the spacetime symmetry group and the covariance group are formally identical, but the symmetry no longer has physical content just because it provides no *restriction* on the covariance group.

Anderson failed to provide a precise mathematical definition of absolute object, although the intuitive meaning seems clear enough. This technical problem has still not been given a satisfactory solution, despite a notable attempt by Friedman (1983) to clarify Anderson’s ideas.

General covariance is not the distinctive feature of GR. Indeed Newtonian gravitational theory can also be expressed in generally covariant form. The distinctive feature of GR is the dynamic interplay between the geometrical structure of the spacetime and the distribution of mass and energy in the spacetime, and the fact that the geometric structures (curvature, affine connection, and so on) are all specifiable in terms of a metric.

Solutions of the Einstein field equations tell us how the geometric structure (the metric and associated curvature) vary throughout spacetime. Einstein originally augmented the field equations with a principle stating that an uncharged test-particle introduced into the spacetime would follow a geodesic path defined by the metric field. So, in a crisp slogan, a test-particle according to Einstein moves in a straight line in a curved spacetime, whereas in Newtonian gravitational theory it would move in a curved line, in a flat spacetime. In GR the metric does ‘double duty’, defining both the local spacetime geometry, and also specifying the effective gravitational field acting on a test-particle. It was later shown that the geodesic equation of motion is not an independent principle, but actually is a consequence of the field equations themselves.

5 Determinism and substantivalism in general relativity

We shall now examine a number of arguments concerned with the question of whether GR is consistent with a substantival interpretation of spacetime or with a relationist interpretation (see Space §3).

We begin with an ingenious argument that if we assume determinism then GR is inconsistent with a substantivalist conception of spacetime points. Suppose $\langle M, g, T \rangle$ is a model of GR with metric *g* and mass/energy distribution *T*. *M* is the spacetime manifold whose points are treated by the substantivalist as ‘individuals’ to which properties, the values of the *g*- and *T*-fields, can be attributed. Then, according to general covariance, for an arbitrary diffeomorphism *d*, $\langle M, d_*g, d_*T \rangle$ is also a model. Now consider an arbitrary proper neighbourhood in spacetime, which we refer to as a ‘‘hole’, and take for *d* a diffeomorphism such that it is the identity map everywhere outside the hole, but unequal to the identity inside the hole. The original model and the new model produced by dragging inside the hole with such a so-called hole diffeomorphism are distinct spacetimes, in the sense that inside the hole the same manifold points are associated with different values of the metric field and mass/energy field. But the two models agree everywhere outside the hole, so what has been shown is that there can be no unique physics *inside* the hole that is controlled by the state of affairs *outside* the hole, so determinism in the weakest possible sense is violated.

Notice that the argument fails in SR, where the symmetry group is sufficiently restrictive that hole diffeomorphisms, which are also symmetries of the Minkowski metric, cannot be constructed.

In standard modern text books of GR, spacetimes are specified only up to diffeomorphism on the grounds that diffeomorphically related spacetimes are empirically equivalent. This of course is true, but does not help the thoroughgoing substantivalist, who does not believe that empirical equivalence implies ontological equivalence.

The worrying point is that the failure of determinism demonstrated is a pervasive feature of GR spacetimes, and does not depend on the detailed physics of field equations, equations of motion, and so on.

There has been much discussion of the ‘hole’ argument in the philosophical literature. Possible responses include just accepting the failure of determinism as a metaphysical discovery about GR spacetimes, or giving up a version of substantivalism in which the identity of spacetime points is independent of properties instantiated at those points, and allowing the identity of spacetime points to follow the dragged metric field (metrical essentialism). Another approach is to resort to counterpart theory and refuse to allow that the dragged model refers to the *same* points of spacetime as the original model, but just to counterparts of those points, the counterparthood relation being formulated so as to rescue a definition of determinism that does not fall foul of the hole argument.

No completely satisfactory solution of the problem has been proposed by the substantivalists, but as we shall argue in the next section, this does not mean that GR should be interpreted relationally; that is, that spacetime just expresses relational properties of the contents of spacetime as encoded in the *T*-field.

6 Mach’s principle

In GR we can raise the question: what determines the local inertial reference frames relative to which the laws of SR hold, according to the generalized equivalence principle? The local inertial frames are themselves determined by the *g*-field, so we can rephrase the question as: what determines the *g*-field? Looking at Einstein’s field equations the possible answer suggests itself that it is the *T*-field, that is, that matter (or other forms of energy) is the source of the *g*-field. This would suggest that mass/energy is the ultimate origin of inertia, rather than spacetime itself. If correct this would demonstrate the implementation in GR of the principle enunciated by Ernest Mach in his critique of classical mechanics at the end of the nineteenth century (1893), that inertial effects in a terrestrial laboratory arise from motion relative to the average matter in the whole universe. If the matter were removed, inertia would disappear, since there would be no chimerical absolute Newtonian space surviving the removal of the contents of the space. Einstein certainly anticipated that GR would support a relationist conception of spacetime, and that Mach’s principle would be a crucial feature of his theory. This turned out not to be the case, for a number of reasons we shall now try to explain.

First we should note that in the field equations the specification of the density of the *T*-field on the right-hand side of the equation already involves the *g*-field, so that the *T*-field cannot be regarded as an independent input to the equations, independent, that is, of the spacetime geometry. Second, the solutions of the field equations involve in general boundary conditions, which again have to be specified in terms of the behaviour of the *g*-field (Einstein tried to obviate this difficulty by requiring the universe to be spatially closed, thus eliminating the need to specify boundary conditions at spatial infinity). Third, it is possible to define the rotation of the matter field with respect to the local inertial frames specified by the *g*-field. So it makes sense in GR to talk of matter rotating locally relative to spacetime, rather than with respect to the rest of the material contents of spacetime. A nice example is the rotating universe discovered by Kurt Gödel (see Spacetime §4). Fourth, we can ask what happens if we set $T = 0$ throughout spacetime. In the spirit of Mach’s principle, the vacuum equations $G = 0$ should admit of no solutions. However, many non-flat matter-free solutions of the field equations exist, in which, informally speaking, gravitational energy provides the source of the curvature (note that in the Einstein field equations gravitational energy is *not* included in the *T*-field).

So one concludes that GR is quite inimical to a relationist conception of spacetime. Far from reducing spacetime to matter, one may indeed envisage spacetime as the primary ontological category, and regard matter (and electromagnetism) as reducible to certain geometrical features of the spacetime.

7 Supersubstantivalism

Supersubstantivalism is the view that spacetime is not just *a* substance but the only substance. What we regard as matter is just a region endowed with certain sorts of geometrical structure, and other candidates for physical reality, such as electromagnetic fields, are also to be given a geometrical interpretation. Precursors of this view include Isaac Newton, who interpreted matter in terms of a property of impenetrability impressed on a region of space, and Rudjer Boscovich and Michael Faraday who regarded material atoms as centres of active force, rather than substantial bodies. William Kingdon Clifford applied Riemann’s theory of curved manifolds to suggest a

geometrical interpretation of matter in terms of local regions of curvature impressed on physical space. Einstein in his own later work sought geometrical explanation of electromagnetic forces in terms of additional geometric structure such as asymmetric connections or higher dimensions of the spacetime manifold. But in terms of orthodox GR, a theory known as geometrodynamics was especially espoused by John Wheeler (1962) and his associates. This exploited an identification of the electromagnetic field with aspects of the four-dimensional geometry, due originally to G. Rainich, electric charges being associated with novel small-scale topology in the spacetime structure, while mass was identified with local concentrations of curvature in the style of Clifford (1876). Technically geometrodynamics foundered in the 1960s due to the pervasive existence of singularities in the solutions of the GR field equations, and problems associated with formulating the initial value problem for the time-evolution of the metric. Nevertheless the spirit of supersubstantivalism has continued to exercise a strong hold on further developments in theoretical physics associated with the programme of quantizing gravitation and unifying it with the other fundamental interactions of electromagnetism and nuclear physics in the general framework of a geometrically interpreted quantum field theory.

See also: Cosmology; General relativity, philosophical responses to

MICHAEL REDHEAD

References and further reading

Anderson, J.L. (1967) *Principles of Relativity Physics*, New York: Academic Press. (Stresses important distinction between covariance principles and symmetry principles - see §4.)

Clifford, W.K. (1876) 'On the Space-Theory of Matter', *Proceedings of the Cambridge Philosophical Society* 2: 157-8.(Interprets matter as the curvature of space.)

Earman, J. (1989) *World Enough and Space-Time: Absolute versus Relational Theories of Space and Time*, Cambridge, MA: MIT Press.(Comprehensive discussion of absolute versus relational theories of space and time.)

Earman, J. (1995) *Bangs, Crunches, Whimpers and Shrieks: Singularities and Acausality in Relativistic Spacetimes*, Oxford: Oxford University Press.(Discussion of spacetime singularities and their significance.)

Earman, J., Glymour, C. and Stachel, J. (eds) (1977) *Foundations of Space-Time Theories*, Minneapolis, MN: University of Minnesota Press.(Papers on philosophy of relativity expressing the state-of-the-art in the late 1970s.)

Einstein, A. (1905) 'Zur Electrodynamik bewegter Körper', *Annalen der Physik* 17: 891-921; trans. W. Perrett and G.B. Jeffery, *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity*, London: Methuen, 1923.(Classic paper introducing the special theory of relativity.)

Einstein, A (1915) 'Die Feldgleichungen der Gravitation', *Preussische Akademie der Wissenschaften*, Berlin: Sitzungsberichte, 844-7.(Statement of the final form of the Einstein field equations without the so-called cosmological constant, introduced by Einstein in 1917 to secure a static bounded universe - see §6.)

Einstein, A (1916) 'Die Grundlagen der Allgemeinen Relativitätstheorie', *Annalen der Physik* 49: 769-822; trans. W. Perrett and G.B. Jeffery, *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity*, London: Methuen, 1923.(Einstein's own views on the foundations of general relativity.)

Friedman, M. (1983) *Foundations of Space-Time Theories*, Princeton, NJ: Princeton University Press.(A modern classic - a good deal of mathematical technicality.)

Grünbaum, A. (1973) *Philosophical Problems of Space and Time*, Dordrecht: Reidel, 2nd enlarged edn.(Classic philosophical discussion of Einstein's theories.)

Kretschmann, E. (1917) 'Ueber den Physikalischen Sinn der Relativitätspostulaten', *Annalen der Physik* 53: 575-614.(Important critique of the principle of general covariance.)

Lorentz, H.A. (1904) 'Electromagnetic Phenomena in a System Moving with any Velocity Less than that of Light', *Proceedings of the Royal Academy of Amsterdam* 6: 809-31.(Pre-Einstein derivation of the Lorentz transformations.)

Lucas, J.R. and Hodgson, P.E. (1990) *Spacetime and Electromagnetism: An Essay on the Philosophy of the Special Theory of Relativity*, Oxford: Clarendon Press. (Comprehensive discussion of different approaches to deriving the Lorentz transformations.)

Mach, E. (1893) *Die Mechanik in ihrer Entwickelung, historischkritisch dargestellt*, Leipzig: Brockhaus; trans. T.J. McCormack, *The Science of Mechanics: A Critical and Historical Account of its Development*, La Salle, IL: Open Court, 1960.(Mach's formulation of his principle - see §6.)

Malament, D. (1977) 'Causal Theories of Time and the Conventionality of Simultaneity', *Nous* 11: 293-300. (Influential attack on the conventionality of simultaneity.)

Minkowski, H. (1909) 'Raum und Zeit', *Physikalische Zeitschrift* 10: 104-11; trans. W. Perrett and G.B. Jeffery, *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity*, London: Methuen, 1923. (Introduces the four-dimensional geometrical interpretation of special relativity.)

Nagel, E. (1980) 'Relativity and Twentieth-Century Intellectual Life', in H. Woolf (ed.) *Some Strangeness in the Proportion: A Centennial Symposium to Celebrate the Achievements of Albert Einstein*, Reading, MA: Addison-Wesley, 38-45. (Nontechnical introduction to the general philosophical import of relativity theory.)

Perrett, W. and Jeffery, G.B. (1923) *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity*, London: Methuen; repr. New York: Dover, 1952.(Useful collection of English translations of the classic papers on relativity theory.)

Poincaré, H. (1905) 'Sur la Dynamique de l'Electron', *Comptes Rendus de l'Academie des Sciences* 140: 1,504-8. (Summarizes Poincaré's version of relativity theory - see §2.)

Redhead, M.L.G. (1993) 'The Conventionality of Simultaneity', in J. Earman, A.I. Janis, G.J. Massey and N. Rescher (eds) *Philosophical Problems of the Internal and External Worlds: Essays on the Philosophy of Adolf Grünbaum*, Pittsburg, PA: University of Pittsburg Press, 103-28.(Critical assessment of the conventionality thesis in special relativity and its application to the twin paradox - also discusses the significance of tachyons in special relativity.)

Reichenbach, H. (1928) *Philosophie der Raum-Zeit-Lehre*, Berlin: de Gruyter; trans. M. Reichenbach and J. Freund, *The Philosophy of Space and Time*, New York: Dover, 1957.(Important philosophical commentary on Einstein's theories.)

Salmon, W.C. (1975) *Space, Time and Motion: A Philosophical Introduction*, Encino, CA: Dickenson.(Clear nontechnical discussion of special relativity in the Reichenbach-Grünbaum tradition.)

Sklar, L. (1976) *Space, Time, and Space-Time*, Berkeley, CA: University of California Press.(General introduction to the philosophy of relativity theory.)

Sklar, L. (1985) *Philosophy of Spacetime Physics*, Berkeley, CA: University of California Press.(Collection of papers by Sklar updating his 1976 book.)

Stein, H. (1991) 'On Relativity Theory and the Openness of the Future', *Philosophy of Science* 58: 147-67. (Defends a notion of becoming in special relativity, but relativized locally to the here and now, rather than the global 'now'.)

Torretti, R. (1983) *Relativity and Geometry*, Oxford: Pergamon Press.(Philosophical discussion of both the special and general theory, with much technical detail.)

Wheeler, J.A. (1962) *Geometrodynamics*, New York: Academic Press. (Comprehensive collection of papers on geometrodynamics - see §7.)

Relevance logic and entailment

'Relevance logic' came into being in the late 1950s, inspired by Wilhelm Ackermann, who rejected certain formulas of the form $A \rightarrow B$ *on the grounds that 'the truth of A has nothing to do with the question whether there is a logical connection between B and A'.*

The central idea of relevance logic is to give an account of logical consequence, or entailment, for which a connection of relevance between premises and conclusion is a necessary condition. In both classical and intuitionistic logic, this condition is missing, as is highlighted by the validity in those logics of the 'spread law', $A \;\&\; {\sim}A \rightarrow B$*; a contradiction 'spreads' to every proposition, and simple inconsistency is equivalent to absolute inconsistency. In relevance logic the spread law fails, and the simple inconsistency of a theory (that a set of formulas entails a contradiction) is distinguished from absolute inconsistency (or triviality: that a set of formulas entails every proposition). The programme of relevance logic is to characterize a logic, or a range of logics, satisfying the relevance condition, and to study theories based on such logics, such as relevant arithmetic and relevant set theory.*

1 The calculus of entailment

'Relevance logic' (also known as 'relevant logic') was born in the late 1950s, when Alan Ross Anderson and Nuel D. Belnap were inspired by reading a paper by Wilhelm Ackermann (1956). In that paper, 'Begründung einer strengen Implikation' (A Foundation for a Rigorous Implication), Ackermann set out to characterize an implication relation, or connective, $A \rightarrow B$ in which there was always a connection between *A* and *B*. 'Rigorous implication', he wrote, 'is intended to express the idea that a logical connection holds between *A* and *B*, that the content of *B* is part of the content of *A*' (1956: 113). Formulas which, though valid in classical logic, were rejected by Ackermann on the grounds that 'the truth of *A* has nothing to do with the question whether there is a logical connection between *B* and *A*' included

$A \rightarrow (B \rightarrow A)$ $\qquad A \rightarrow (B \rightarrow A \;\&\; B)$

$A \rightarrow ({\sim}A \rightarrow B)$ $\qquad A \rightarrow ((A \rightarrow B) \rightarrow B)$

$B \rightarrow (A \rightarrow A)$ $\qquad A \rightarrow (A \rightarrow A)$

$A \;\&\; {\sim}A \rightarrow B.$

Ackermann began his paper by recalling C.I. Lewis' reasons for developing his calculus of strict implication as a calculus expressing a more reasonable notion of 'implication' or 'entailment' than was current at the time (see Lewis 1914). (That the quest for a relevant notion of entailment goes back well before Lewis has been demonstrated more recently by C.J. Martin (1986) in his investigations into twelfth-century logic.) But Ackermann believed Lewis had not gone far enough. Although the first four formulas above fail when '$\rightarrow$' is interpreted as strict implication, the last three still hold. Moreover, taking '$\Box$' to represent 'is necessary', $\Box A \rightarrow (B \rightarrow A)$ and $\Box{\sim}A \rightarrow (A \rightarrow B)$ were theses of Lewis' system of strict implication. Ackermann added '$\Box$' to his calculus of rigorous implication (via an absurdity constant '$\bot$' and a definition of '*A* is impossible' as $A \rightarrow \bot$) and showed that the above so-called 'paradoxes of strict implication' fail when '$\rightarrow$' is taken as rigorous implication.

Belnap reports returning to Yale University from a year in Europe working with Feys on composing a Gentzen system for Ackermann's system, to write up his Ph.D. dissertation, and ingenuously asking Anderson if he knew anyone in the United States who was interested in Ackermann's work. Anderson, feeling equally isolated in his fascination with Ackermann's ideas, having reviewed them in the *Journal of Symbolic Logic*, was delighted to find, as Belnap put it, a 'fellow Strengenite'. Together they started the research programme which led to more than thirty years' creative work on every aspect of a logic of relevance.

How does their work relate to Ackermann's? The initial motivation was the same, and the same goal which had inspired Lewis: to find a satisfactory explication of the concept of entailment, or (formal) logical consequence. Therefore, in the first decade of work, the focus was on the calculus of entailment (E) incorporating aspects of both relevance and necessity. E validates the same theses as Ackermann's system, and differs from it in several apparently minor ways, which in fact are quite significant. First, '$\Box A$' can be defined directly in the basic system

as $(A \to A) \to A$, yielding the same theses, so '$\bot$' and the associated apparatus can be removed. '$\to$' is itself a modal connective (the '$\to$'-fragment of E is a subsystem of the arrow fragment of S4). Second, two of Ackermann's four rules of inference, namely (γ) and (δ), though unproblematic in the pure logic, are unacceptable when the logic is extended to a substantive theory. Anderson and Belnap showed that they could be dispensed with. Ackermann's rule of inference (γ) says that a proof of *B* results immediately from proofs of $A \vee B$ and $\sim A$. The point is this: whenever $A \vee B$ and $\sim A$ are provable in the pure logic of E, so too is *B*, that is, (γ) is an admissible rule of E, and so is redundant. (Conjectured in 1958, the first proof was not found until ten years later.) But in a theory based on E, there might be proofs of $A \vee B$ and $\sim A$ (that is, these formulas might belong to the theory, extending E by non-logical axioms) without there necessarily being a proof of *B* (or of $\sim B$ - the theory need not be negation-complete). One major task in the research programme of relevance logic became to show the admissibility of (γ) for various E-theories, once E was shown to admit (γ) itself. We will note later the major disappointment in this search, the failure of (γ) for relevant arithmetic.

The problem with the rule (δ) is similar, but less deep and seminal. (δ) says that from $A \to (B \to C)$ and a proof of *B* one can infer $A \to C$. Again, the proof of *B* must be logical. Even if $A \to (B \to C)$ and *B* belong to a theory, $A \to C$ might not. Indeed, what (δ) seems to promise is a permutation, from $A \to (B \to C)$ to $B \to (A \to C)$, and that is unacceptable in the modal logic E, being a fallacy of modality. (Let *A* be $B \to C$ itself, and consider how the permutation would fail in S4 with '$\to$' as strict implication.) Such a move is legitimate only if *B* is appropriately modal, which we can capture with the connective '$\Box$'. (δ) can therefore be replaced by an axiom,

$$(A \to (B \to C)) \to (\Box B \to (A \to C)),$$

or, equivalently, as Anderson and Belnap chose to do in E,

$$(\Box A \mathbin{\&} \Box B) \to \Box(A \mathbin{\&} B),$$

with '$\Box$' defined as above.

E, the calculus of entailment, had thus been born; its study was to reveal new worlds in logic, and, before long, to shift attention to a range of related logics. (See Anderson and Belnap 1975, 1992.)

2 The calculus of relevant implication

Anderson and Belnap called E the logic of relevance and necessity. However, it is important to understand that the programme they embarked on did not try to give any general characterization of relevance, or even of relevance as a logical notion. What they identified were what they called fallacies of modality and of relevance. Fallacies of modality were first characterized by them for the '$\to$'-fragment as taking as valid any formula of the form $B \to (A \to C)$, where *B* was a propositional variable. Ackermann had suggested the criterion for his full calculus that neither '$\to$' nor '$\bot$' should belong to *B*.

Arguably, neither characterization really identifies what a modal fallacy is. For one thing, neither condition easily generalizes beyond the propositional language given. Adding further connectives or quantifiers and the language of predicates, one must generalize the condition, yet that means one needs some understanding of what motivates it. Pottinger tried to capture a notion suitable for E (see Anderson and Belnap 1975: 348), but again, how it would generalize further is unclear. These are technical criteria, not proper philosophical analyses. Moreover, while separating the modal from the relevance fallacies is a promising idea, a natural consequence is that the modal fallacies should equally be rejected in S4. Yet $B \to (A \to C)$ is a thesis of S4 whenever $A \to C$ is, regardless of the nature of *B*, even if it is a propositional variable. The answer is that a modal fallacy is essentially a permutation move, promoting *B* from its secondary position to a primary one: from $A \to (B \to C)$ to $B \to (A \to C)$. Such a permutation will permit, *inter alia*, validation of formulas $B \to (A \to C)$ where *B* is a propositional variable (or contains neither '$\to$' nor '$\bot$'). Checking that such formulas are not theses is a useful test that a system does not commit any modal fallacies. But it is not an adequate explication of the notion.

Fallacies of relevance were also characterized by Anderson and Belnap only in technical terms, without any wider elaboration of what was really wrong with them. Their formal tests of relevance were variable-sharing and derivational utility. The former again only applies to the propositional fragment. The idea was that $A \to B$ is valid only if *A* and *B* share a propositional variable. Belnap presented matrices which test for this condition. The

other criterion arose from an attempt by Anderson and Belnap to prove a deduction theorem for Ackermann's system, and led to the creation of a (Fitch-style) natural deduction formulation of E. They developed sets of indices which kept track of relevance through the proof. Conditional proof (the analogue in the natural deduction formulation of the deduction theorem as a result about axiomatic formulations) was permissible only when in the proof of *B* from *A*, *B*'s index contained that assigned to the assumption of *A*. The idea has been fecund in the development of relevance logic, especially when one realizes that the indices play the role of the assumption lists in a Gentzen-style formulation of the logic as a natural deduction calculus of sequents. But it is still a restricted and technical idea: how will it generalize when we extend the calculus to, say, second-order, and outstrip proof theory? In brief, what is really meant here by relevance?

But separating ideas of relevance from those of modality permitted the creation of the system R of relevant implication. This results from E by relaxing the restriction on permutation. The '$\rightarrow$' of R is a non-modal relevant conditional. It was seen as relating to the '$\rightarrow$' of E as material implication relates to strict implication. Yet in this development lay the effective death-knell of E. If one now starts with R and adds an S4-necessity operator '$\Box$', there are theses containing '$\Box$' only in the position $\Box(A \rightarrow B)$ which do not result from theses of E in which $A \rightarrow B$ has been replaced by $\Box(A \rightarrow B)$. Since this discovery in 1973 (see Anderson and Belnap 1975: 351-2), work on R and its extensions and subsystems has put E in the shade.

3 The systems of relevance logic

With the creation of R, and more so with its eclipse of E, the stage was set for a study of a full range of systems of relevance logic. Immediate neighbours of E and R were the system T of 'ticket entailment' and the 'mingle' systems EM and RM. Ticket entailment picked up an idea of Ryle's; that conditionals are a kind of inference-ticket, licensing the inference from one factual (non-conditional) statement to another. T is a subsystem of E which effectively restricts even the permutation of conditional antecedents which is allowed in E.

The mingle systems extend E and R by the addition of the axiom $A \rightarrow (A \rightarrow A)$. This looks harmless, lacking the obvious irrelevance of $A \rightarrow (B \rightarrow A)$. But adding the mingle axiom to R leads to the validity of $(A \rightarrow B) \vee (B \rightarrow A)$, and we find one of the classic paradoxes of material implication has returned. RM is an interesting source of study; but it is, in Meyer's phrase, a 'semi-relevant logic'.

With hindsight, one can see the study of E and R as the start of research into a range of systems which in the early 1990s came to be called 'substructural logics'. Two insights led to this development. One was recognition of the Curry-Howard correspondence, the treatment of formulas as types (see Lambda calculus). This reveals that the '$\rightarrow$'-fragment of R was being studied even in the 1930s by Church, in his λ-I calculus. In the λ-calculus one studies functions, developing a framework for the general treatment of functional abstraction and application. The reductions and equivalences of functional expressions are governed by combinators, so that, for example, the combinator *B* asserts that the composition of functions is associative: $B(fg)h$ reduces to $f(gh)$ (see Combinatory logic §1). Correlating functional abstraction with conditional proof, functional application with *modus ponens*, there is a natural correlation of the definability of λ-terms, or combinators, with the provability of '$\rightarrow$'-formulas (which extends to Cartesian product and conjunctions, and so on). R (or at least, its '$\rightarrow$'-fragment) is then seen as BCW-logic, the logic of associativity, permutation and contraction, placed in a family with BCK-, BCI- and other logics. BCI-logic, otherwise known as the '$\rightarrow$'-fragment of RW (R without contraction) corresponds to Church's λ-I calculus. The relevance logics can then be classed as those representing functions that 'really depend on their arguments', rejecting the combinators such as *K*, corresponding to the λ-term $\lambda x, y.x$ (see Combinatory logic §1).

The other development was 'correspondence theory', the study of the correlation between logical principles and constraints on corresponding relations in the model theory. It started in the study of semantics for modal logics, whereby different conditions on accessibility in possible worlds semantics are seen to match various axioms for necessity, '$\Box$', in the logic modelled. The semantics for relevance logic began to be worked out in the late 1960s and early 1970s by Dunn, the Routleys, Meyer, Maximova and Fine, working largely independently. There were two aspects needing special attention. First, a way of invalidating the spread law had to be found, and that seemed to require modelling *A* and $\sim A$ simultaneously. One solution was the negation-shift operation, pairing each world *w* with w^*, whereby $\sim A$ is true at *w* if and only if *A* is true at w^*. Suitable constraints on '*' give the right involute properties to '$\sim$'. The other aspect was the relevance '$\rightarrow$' itself, whether for E or R (or related logics).

The approach which has come to dominate uses a three-place relation on worlds, *R*, whereby $A \rightarrow B$ is true at *x* if *B* is true at *b* whenever *A* is true at *a* and *Rxab*. Again, suitable constraints on *R* give the desired properties of ‘ $\rightarrow$’. Fine’s version used a pre-ordering of worlds (as in S4) and an operation of fusion, ‘ $\circ$’, such that $A \rightarrow B$ is true at *x* if *B* is true at $x \circ a$ whenever *A* is true at *a*. This operational semantics makes more transparent the correspondence with the combinators. The correspondences thus set up identify a range of logics, the substructural logics (logics with restricted structural rules), which differ essentially only in their structural rules.

This assumes, however, the invariance of the Boolean apparatus of ‘Čamp;’ and ‘ $\vee$’ erected on the underlying calculus of ‘ $\sim$’, ‘ $\rightarrow$’, and perhaps fusion ‘ $\circ$’, with $A \circ B$ true at $a \circ b$ whenever *A* is true at *a* and *B* at *b*. ‘ $\circ$’ is a ‘relevant conjunction’, $A \circ B \rightarrow C$ being equivalent to $A \rightarrow (B \rightarrow C)$. ‘Čamp;’ and ‘ $\vee$’ need not be so invariant, however. Ackermann noted that distribution of ‘Čamp;’ over ‘ $\vee$’ no longer followed from the other axioms when restricted to yield rigorous implication, and so added it as a separate axiom. Anderson and Belnap followed him, supporting the case by yet another criterion of relevance, applied to first-degree entailments ($A \rightarrow B$ where *A* and *B* contain only ‘ $\sim$’, ‘Čamp;’ and ‘ $\vee$’); that they be tautological: putting *A* in disjunctive normal form (DNF) and *B* in conjunctive (CNF), every disjunct on the left shares an atom with every conjunct on the right, roughly, *B* is ‘contained in’ *A*. But some have argued that dropping the combinator *K* (‘weakening’) as the source of irrelevance in full classical logic, should yield a non-distributive calculus. Interestingly, every classical tautology (in ‘ $\sim$’, ‘Čamp;’ and ‘ $\vee$’) is provable in R. But this fact fails for the non-distributive calculus. For example, the axiom ‘sum’ of *Principia Mathematica*,

$$(Q \supset R) \supset ((P \vee Q) \supset (P \vee R)),$$

with ‘ $A \supset B$’ for ‘ $\sim A \vee B$’, is unprovable in LR (R without distribution). LR was the focus of much research in automated theorem proving in the 1980s.

Dropping *W* (‘contraction’) from LR yields a fragment of Girard’s linear logic (1987), as does dropping distribution from RW (see Linear logic). Girard’s rich theory, encompassing additive connectives (‘Čamp;’ and ‘ $\vee$’ - there written ‘Čamp;’ and ‘ $\oplus$’), multiplicatives (‘ $\rightarrow$’ and ‘ $\circ$’ - there ‘ $\multimap$’ and ‘ $\otimes$’) and exponentials (the modal operators ‘!’ and ‘?’), tracking the essential and profligate use of resources, picks up Church’s original interest in BCI and has played a central role recently in theoretical computer science.

Those relevance logics (Ackermann’s system is one exception) which reject the spread law in the form $A \mathbin{\&} {\sim}A \vdash B$, are paraconsistent logics (see Paraconsistent logic). For some, they even share a motivation, namely, to allow logical treatment of inconsistent theories without making them trivial (their deductive closure does not embrace everything, as in classical logic). But for others, their motivation is different: many paraconsistent logics arise from an interest in dialetheism, the belief that some inconsistent theories are true. Wider still run the ‘sociative logics’ (a term of Sylvan’s), encompassing relevance, connexive, non-transitive and a whole range of, as he puts it, ‘broadly relevant’ logics.

4 Relevant predication and arithmetic

This survey of relevance logics suggests an undue concentration on the zero-order (propositional) fragment, and that reflects the facts. But important results have been established for first-order relevance logic, relevant arithmetic and so on. For too long, perhaps, it was thought that the extension to first-order was straightforward. Fine’s discovery that plausible ways of extending the proof theory and semantics to first-order quantifiers led to incompleteness was a surprise, not really softened by his later elaboration of a semantics which did not validate the unprovable formulas (see Anderson and Belnap 1992: §§52-3). Urquhart’s demonstration of undecidability for propositional T, E and R settled an open problem of twenty years’ standing, again in a surprising way (1992: §65). (The non-distributive logics are decidable.) The failure of (γ) for the formulation $\mathrm{R}^{\#}$ of relevant arithmetic was another surprise, blocking a natural line of research.

The exposure of these negative results in the 1980s was not unlike the exposure of the limitative results for classical mathematics (incommensurability, uncountability, incompleteness). At first, relevance logic lost momentum as they were settled, and settled in such surprising and negative ways. Nonetheless, limitative results notoriously give rise, once assimilated, to new avenues of research. They are appearing in relevance logic, two notable ones being the attempt to characterize a notion of relevant predication by Dunn and others (Anderson and Belnap 1992: §74) - recognizing that the extension to first-order is not straightforward; and the search for a

relevant arithmetic which admits (γ) (1992: §72) - perhaps by strengthening the rule of induction.

See also: Consequence, conceptions of; Logical and mathematical terms, glossary of; Logical laws; Modal logic

STEPHEN READ

References and further reading

Ackermann, W. (1956) 'Begründung einer strengen Implikation' (A Foundation for a Rigorous Implication), *Journal of Symbolic Logic* 21: 113-28.(The original paper which started it all.)

Anderson, A.R. and Belnap, N.D. (eds) (1975, 1992) *Entailment: The Logic of Relevance and Necessity*, Princeton, NJ: Princeton University Press, 2 vols.(The classic source for their work and others'; includes an extensive bibliography.)

Došen, K. and Schroeder-Heister, P. (eds) (1993) *Substructural Logics*, Oxford: Clarendon Press. (Places relevance logic in the framework of wider study of logics with restricted structural rules.)

Dunn, J.M. (1986) 'Relevance Logic and Entailment', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, vol. 3, *Alternatives to Classical Logic*, Dordrecht: Reidel, 117-229.(A very useful technical survey.)

Friedman, H. and Meyer, R.K. (1992) 'Whither Relevant Arithmetic?', *Journal of Symbolic Logic* 57: 824-31. (Shows the failure of (γ) for relevant arithmetic.)

Girard, J.-Y. (1987) 'Linear Logic', *Theoretical Computer Science* 50: 1-102.(The first and definitive presentation of linear logic.)

Lewis, C.I. (1914) 'The Calculus of Strict Implication', *Mind* 23: 240-7.(One of Lewis' early papers in which he described his project of formalizing a logic of entailment.)

Martin, C.J. (1986) 'William's Machine', *Journal of Philosophy* 83: 564-72.(The debate over relevance in twelfth century logic.)

Read, S. (1988) *Relevant Logic*, Oxford: Blackwell.(An elementary presentation of the philosophical basis of relevance logic and a formal system for a range of relevance logics.)

Sylvan, R. [néé Routley] with Plumwood, V., Meyer, R.K. and Brady, R. (1982) *Relevant Logics and Their Rivals*, vol. 1, *The Basic Philosophical and Semantical Theory*, Atascadero, CA: Ridgeview. (Places zero-order relevance logic in a setting of non-classical intensional logics.)

Sylvan, R. and Norman, J. (eds) (1989) *Directions in Relevant Logic*, Dordrecht: Kluwer.(A collection of papers including those given at a memorial conference for Alan Anderson in 1974, some updated and others additional.)

Thistlewaite, P.B., McRobbie, M.A. and Meyer, R.K. (1988) *Automated Theorem-Proving in Non-Classical Logics*, London: Pitman.(The computer program KRIPKE, using a search tree based on models for non-distributive relevance logic.)

Reliabilism

Reliabilism is an approach to the nature of knowledge and of justified belief. Reliabilism about justification, in its simplest form, says that a belief is justified if and only if it is produced by a reliable psychological process, meaning a process that produces a high proportion of true beliefs. A justified belief may itself be false, but its mode of acquisition (or the way it is subsequently sustained) must be of a kind that typically yields truths. Since random guessing, for example, does not systematically yield truths, beliefs acquired by guesswork are not justified. By contrast, identifying middle-sized physical objects by visual observation is presumably pretty reliable, so beliefs produced in this manner are justified. Reliabilism does not require that the possessor of a justified belief should know that it was reliably produced. Knowledge of reliability is necessary for knowing *that a belief is justified, but the belief can* be *justified without the agent knowing that it is.*

A similar reliabilist account is offered for knowledge, except that two further conditions are added. First, the target belief must be true and, second, its mode of acquisition must rule out all serious or 'relevant' alternatives in which the belief would be false. Even an accurate visual identification of Judy does not constitute knowledge unless it is acute enough to exclude the possibility that it is her twin sister Trudy instead.

One major virtue of reliabilism is its ability to secure knowledge against threats of scepticism. In place of excessive requirements often proposed by sceptics, reliabilism substitutes more moderate conditions. People do not need infallible or certainty-producing processes to have justified beliefs, according to reliabilism, only fairly reliable ones. Processes need not exclude radical alternatives like Descartes' evil demon in order to generate knowledge; they need only exclude realistic possibilities like the presence of an identical twin.

1 Reliabilism and naturalistic epistemology

Reliabilism is often regarded as a species of naturalistic epistemology (see Naturalized epistemology). There are various properties of an epistemology that might invite the label 'naturalistic', and reliabilism has most if not all of them. First, an epistemological theory may be called naturalistic if it holds that normative epistemic properties are reducible to natural, non-epistemic properties and relations. The reliabilist theory of justified belief sketched above fits this description. It tries to reduce the normative property of 'justifiedness' to three types of 'natural' properties: psychological properties (for example, being a psychological process), causal relations and truth. It is generally agreed that psychological properties are non-normative (certainly not epistemically normative, at any rate), and similarly for causation. Although some philosophers maintain that truth is an epistemic concept, which should be defined in terms of justification or rationality, reliabilism typically rejects epistemic analyses of truth in favour of a realist, non-epistemic account, such as a correspondence theory (see Truth, correspondence theory of). If the latter approach is correct, then only natural properties are invoked by reliabilism.

Reliabilism is also naturalistic in portraying the cognitive agent as a natural - that is, physical or biological - system and interpreting epistemic accomplishments as the products of natural processes or mechanisms. Psychological operations, processes or mechanisms may be conceptualized as input-output devices, where the inputs are doxastic or nondoxastic states and the outputs are beliefs. One variant of reliabilism, the information-theoretic approach, views the cognitive system as an 'information-processing' system. Knowledge that *s* is *F* is acquired when the system receives the 'information' that *s* is *F*, that is, when it receives a signal that eliminates all (relevant) counter-possibilities. Since the concept of information transmission has been applied to a variety of physical systems, this approach has a strong naturalistic flavour (see Information theory and epistemology).

Epistemologies are called naturalistic in a third sense when they appeal to empirical science for the execution of certain epistemological tasks. Under one brand of reliabilism, for example, it is a task for cognitive science to identify the specific psychological operations available to human cognizers and to determine their reliability. Of course, ordinary people must have rough and intuitive notions of our psychological processes; otherwise they could not segregate beliefs into the categories of justified and unjustified. But cognitive science may be expected to produce more revealing and refined accounts of belief-forming processes.

At what age, for example, are children able to form reliable beliefs about the number of objects in a collection, and

what mental operations enable them to do so? Surprisingly, it appears that infants as young as five months are sensitive to numbers and can calculate the results of small additions and subtractions. Judging by the amount of time infants stare at a display, psychologists have determined that they expect one item added to one item to yield two items, and expect the subtraction of one item from two items to leave one item. The plausibility of early - indeed, innate - counting abilities in humans is strengthened by the discovery of impressive feats of counting in non-human animals. Rats, for example, can identify the number of times they have pressed a lever, up to at least twenty-four presses, when they are trained to press a specified number of times on a particular lever before pressing once on a second lever for a reward (see Innate knowledge).

What operation or mechanism might subserve reliable counting? One possibility proposed by researchers is an accumulator device. A hypothesized pacemaker puts out pulses at a constant rate, which can be passed into an accumulator by the closing of a mode switch. Every time an entity is experienced that is to be counted, the mode switch closes for a fixed interval, passing energy into the accumulator. The accumulator fills up in equal increments, one for each entity counted. A device of this sort would be a reliable counter. If animals have some such mechanism, their numerical beliefs generated by it would qualify as knowledge according to reliabilism.

Reliabilism does not claim, of course, that knowledge depends exclusively on specialized innate mechanisms. Culturally acquired methods, such as a method for deriving square roots, can also breed knowledge if they are reliable. Although native resources must be employed to mentally represent and execute learned methods, the methods themselves need not be ‘hard-wired’.

According to reliabilism the epistemic status of a belief depends on its mode of causation (see Knowledge, causal theory of). This contrasts with purely evidentialist theories, which consider a belief to be justified as long as it is logically or probabilistically supported by a corpus of evidence that the agent possesses (see Probability theory and epistemology). Reliabilism contends that mere possession of strong evidence is not enough for justified belief. A detective, for example, may possess assorted strands of evidence that jointly incriminate a certain suspect, but the slow-witted detective may not notice that these disparate strands of evidence can be woven into an airtight case. If the detective is prompted to accept the suspect’s guilt for entirely different and illegitimate reasons - for example, the suspect’s unsavoury appearance - then the belief is unwarranted.

Not only is mere possession of adequate evidence insufficient for justification, but it is also unnecessary. At least the *current* possession of adequate evidence is not required for justification. Suppose you once had excellent evidence for *p* (having heard it from a trustworthy source) but you no longer recall this evidence: your belief in *p* may still be justified. A person’s failure to keep track of his original evidence does not destroy the justifiedness of a belief, as long as the belief in fact had a suitable provenance, namely a history of reliable belief-forming and belief-preserving steps.

2 Reliabilism as a species of externalism

Reliabilism is a species of epistemological externalism rather than internalism (see Internalism and externalism in epistemology). Although there is no unanimity in defining these terms, the rough idea is that a theory is internalist if and only if all of the factors it requires for a belief to be justified are directly accessible to the cognizer, whereas a theory is externalist if and only if some of the justificatory factors are beyond the cognizer’s (direct) ken. Since neither the truth-propensity of a belief-forming process nor the history of a belief’s acquisition is directly accessible to the cognizer, reliabilism is a form of externalism. Externalism is often criticized for departing from traditional epistemology, which is allegedly internalist. It is debatable, however, whether the dominant tradition in epistemology is really internalist. In fact, many historical epistemologies can be interpreted as having strong elements of reliabilism, including the epistemologies of Plato, Descartes, Hume, Reid and Peirce - see, for example, Schmitt (1992).

Historical epistemology aside, it is clear that the concept of knowledge (as contrasted with justification) has externalist elements. First, knowledge requires the truth of what is known, and truth is the quintessential externalist element. Second, a satisfactory solution to the Gettier problem about knowledge seems to require a further externalist component (see Gettier problems). For example, a person might have a justified belief that the object they see in the nearby field is a barn, and this may be true. But if, entirely unknown to them, there are papier-mâché barn facsimiles in the neighbourhood which they would mistake for a barn if one of them were there

instead, then intuitively they do not know what they see to be a barn. This seems to show that the external situation of the cognizer, and not just their internal mental condition, is relevant to the possession or non-possession of knowledge. The reliabilist theory of knowledge covers this by saying that knowledge is only acquired when all relevant alternatives are excluded, and the external situation - such as the presence of papier-mâché facsimiles in the neighbourhood - helps to determine which alternatives are relevant.

Although knowledge clearly has externalist components, the concept of justification may yet be thoroughly internalist, as internalists contend. But what, more precisely, is required by internalism? The typical internalist requirement of cognitive accessibility can be interpreted in at least two ways: as requiring that the believer actually should be aware of the justifying factors (at the time of the belief), or as requiring only the capacity to become aware of these factors by appropriate focusing of attention, without any change of position, new information, and so on. Either of these criteria, however, is probably unsatisfiable by many traditional theories of justification. One brand of coherentism, for example, says that a belief is justified only if it coheres with the totality of the cognizer's other current beliefs (see Knowledge and justification, coherence theory of). The justifying factors in this theory cannot meet the direct-accessibility requirement (in either version) because it is psychologically impossible to retrieve and examine the totality of one's beliefs. The beliefs of a normal adult number in the thousands or even millions, are mostly stored in long-term memory and are often difficult to recall without appropriate retrieval cues. Furthermore, many brands of coherentism, as well as other forms of evidentialism, impose complex demands in terms of the probabilistic relations that a justified belief must bear to the background corpus of beliefs. Whether a target belief satisfies these demands is not readily accessible to ordinary people who lack training in probability theory. So coherentism, no less than reliabilism, will turn out to be a brand of externalism.

Is it so bad for a theory of justification to be externalist? A common objection is that externalism, at least in its reliabilist incarnation, renders mistaken classifications. Imagine a person with a reliable clairvoyance faculty, who occasionally forms beliefs 'out of the blue' (that is, unguided by any sense-like appearances) about properties and locations of far-off objects; and these are usually correct. This person never checks up on these beliefs, and therefore has no independent evidence that the faculty is reliable; or we may suppose that the person has evidence to the contrary. Would a clairvoyantly-formed belief of that person's be justified? Reliabilism in its simplest form says yes, because the belief would be caused by a reliable process; but critics reject this assessment as intuitively mistaken.

Reliabilists may respond by complicating their theory. A standard gambit is to add a 'non-undermining' requirement, namely, a provision that the cognizer must neither believe nor possess evidence supporting the proposition that the target belief is *un*reliably caused (see Knowledge, defeasibility theory of). In at least one version of the clairvoyance case, the agent has evidence against his possession of a (reliable) clairvoyance faculty. This might give him evidence that his belief is unreliably caused (since it has no other familiar mode of causation), which would violate the new requirement. Even in the variant of the example in which the cognizer merely lacks positive evidence in support of his possessing a clairvoyance faculty, it might be argued that background scientific information makes the existence of such a faculty highly improbable; so if one lacks specific evidence of this power, the overall evidence cuts against it. The non-undermining condition would again be violated, and hence the revised version of reliabilism would declare the belief to be unjustified. Addition of the non-undermining condition does not transform reliabilism into a form of internalism, for it retains the reliability element that would violate the direct-accessibility requirement.

Whereas the clairvoyance counterexample tries to show that reliability is not sufficient for justification, a second type of objection claims that it is not necessary. Imagine a person regularly deceived by a Cartesian demon. They have the same patterns of sensory experiences as you or I, but in this case the resulting perceptual beliefs are systematically mistaken. Despite the unreliability of their perceptual belief-forming processes, surely their beliefs are as justified as yours or mine, since subjectively the grounds for belief are exactly the same. This challenges the very core of externalism by suggesting that only subjective, internal factors matter to justification, not external ones.

Reliabilists may respond by first pointing out that reliabilism also makes internal factors such as sensory experiences and background beliefs relevant to justification. These internal states are the inputs to mental

processes, and only in combination with them can mental processes yield accurate output-beliefs. Furthermore, mental processes themselves are internal events, so reliabilism cannot justly be accused of ignoring internal matters completely. But these points do not blunt the force of the objection. After all, reliabilism does appeal to 'external' truth-ratios to distinguish justified from unjustified beliefs, and yet the demon case seems to show that this appeal is misplaced.

The objection assumes, however, that the reliability of a process is assessed by its performance in the world of the example. If this assumption is correct - if reliability is determined by a process's performance in the example world - then reliabilism does imply the unjustifiedness of the demon-world beliefs. But some alternate interpretations of reliability might be proposed under which the epistemic status of the beliefs (as judged by reliabilism) may be different. For example, a reliabilist might propose that what counts is reliability in the *actual* world. Since perceptual belief-formation is reliable in the actual world, use of the same process or processes would yield justified beliefs even in a demon world. This particular reformulation of reliabilism may not quite succeed, but it shows how reliabilism has greater flexibility than might initially be supposed.

A third kind of problem for reliabilism - one not directly related to the internalism-externalism issue - is the 'generality' problem. Reliabilism says that a belief's justificational status depends on the truth-conduciveness of the psychological processes that cause it. But processes can be individuated in many ways; indeed, each token belief can be viewed as the outcome of numerous process types of differing grain or generality, with widely varying truth-ratios. What principle selects the appropriate process type, or set of process types, for each case?

3 A variant form of reliabilism

In light of these problems, different variants of reliabilism have been proposed. Let us examine one such variant and see how it addresses two of the foregoing problems. As a preliminary point, we should be clearer about what is expected of a theory of knowledge or justified belief. On one approach, such a theory should reveal the 'nature' or 'essence' of knowledge or justification, where these are construed as something like natural kinds, independent of our human conceptualization of them. On another approach, the aim of a theory is to reveal the concepts or mental representations associated with these epistemic predicates, and to reveal the ways that these representations are deployed by epistemic evaluators when they make epistemic judgments. This latter approach is the one that is pursued here.

The form of reliabilism considered in this section starts with a familiar idea in cognitive science, namely that a great deal of cognitive activity consists of 'pattern matching'. A pupil doing algebra problems, for instance, will try to match a new problem with some familiar pattern or structure of problems. It is much easier to solve a problem with a familiar solution pattern than an entirely novel one. The view considered here is that epistemic evaluation similarly proceeds by pattern matching. Evaluators are assumed to represent, in their heads, various patterns (or prototypes) of belief formation, some categorized as good patterns and some as bad. Examples of good patterns might include the use of various types of perception and the use of the 'straight rule' in inductive reasoning (see Theoretical (epistemic) virtues §1). Examples of bad patterns might include forming beliefs by guesswork and ignoring relevant evidence. Good patterns may be called epistemic 'virtues' and bad patterns epistemic 'vices'. The present hypothesis is that when an evaluator is presented with a real or imagined case of belief, they consult their mentally stored list of virtues and vices, and judge the belief to be justified or unjustified by seeing whether its own mode of production matches virtues or vices. If its mode of production matches virtues only, it is categorized as justified; otherwise it is categorized as unjustified. If a mode of production is unfamiliar or non-standard, the epistemic judgment will depend on the judged similarity of the process to the antecedently stored patterns of virtue or vice.

How would this approach apply to our previously discussed examples? In the demon-world case, beliefs are produced by standard perceptual belief-forming processes. So these are matched to patterns that are clearly virtues. In the clairvoyance cases, matters are somewhat less clear. In one version of this case, the epistemic agent is described as having evidence against possessing a reliable clairvoyance power; but they apparently ignore this evidence, since they still rely on the clairvoyance faculty. If ignoring relevant evidence is on an evaluator's list of epistemic vices, it is predictable that the evaluator would judge the belief to be unjustified. Furthermore, the process of clairvoyance might seem very similar to a process of guesswork, since there are no appearances or sensory presentations associated with clairvoyance as there are with sight, hearing, smell and so forth. Thus, a

clairvoyantly formed belief might be assimilated to the 'vice' of guesswork, and judged to be unjustified on that basis.

Thus far, nothing in the present approach invokes reliability. The reliability element is now introduced to explain how 'patterns' come to be classified as good or bad. The theory suggests that evaluators form opinions about the truth-ratios associated with sundry belief-forming processes. Those thought to have high truth-ratios are treated as good patterns and those with low truth-ratios as bad patterns. Thus, reliability is the criterion to which evaluators appeal in establishing patterns as good or bad. Evaluators might not do this purely individually. They might inherit some evaluative prototypes from others in the linguistic community. Whether individually or socially, however, the good and bad prototypes are selected by considerations of (judged) reliability.

The theory that emerges, then, is a two-level theory, somewhat analogous to rule-utilitarianism (see Utilitarianism §3). Judgments of justifiedness in particular cases do not appeal directly to the reliability of the belief's generating processes. Rather, such judgments involve pattern-matching to stored prototypes of good and bad belief-generating processes. Reliability enters the theory as the basis for classifying these processes as epistemically good or bad.

Recalling the demon-world example, it may be asked whether the basis for process evaluation is performance in the actual world or performance in other possible worlds as well. Presumably, evaluators appeal to real-world track records in making their evaluations, but they may also tend to assume that these observed track records can be extrapolated to non-observed and non-actual cases. In general, it is doubtful that ordinary evaluators, who lack ways of thinking systematically about 'possible worlds', use any sharply defined, world-relativized basis for reliability assessments. The crucial point, however, is that evaluators do not directly apply reliability considerations to novel cases. According to the present theory, they do not say (or think): 'Since perception is unreliable in the demon world, therefore perception-based beliefs in that world are unjustified'. Instead, they use pattern-matching to previously entrenched prototypes to arrive at an epistemic assessment of the target belief. With this understood, we have a form of reliabilism that has notable success in explaining evaluators' judgments in 'hard' cases, that is, cases that are hard for simple reliabilism.

See also: Justification, epistemic; Knowledge, concept of; Normative epistemology; Virtue ethics

ALVIN I. GOLDMAN

References and further reading

All of the following works involve intricate argument but little or no technicality.

Alston, W.P. (1989) *Epistemic Justification: Essays in the Theory of Knowledge*, Ithaca, NY: Cornell University Press.(Contains several essays defending a form of reliabilism and discussing background meta-epistemological issues.)

Alston, W.P. (1996) 'How to Think about Reliability', *Philosophical Topics* 23 (1): 1-29. (Attempts to resolve the generality problem for reliabilism.)

Dretske, F.I. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(Expounds the information-theoretic version of reliabilism mentioned in §1 above.)

Feldman, R. (1985) 'Reliability and Justification', *Monist* 68 (2): 159-74.(Presses the generality problem for reliabilism.)

Goldman, A.I. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(Systematic exposition of reliabilism plus an examination of psychological processes and their reliability.)

Goldman, A.I. (1992) *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge, MA: MIT Press.(The essay 'What Is Justified Belief?' is a standard formulation of reliabilism; a second, 'Epistemic Folkways and Scientific Epistemology', presents the two-tiered version of reliabilism sketched in §3 above.)

Goldman, A.I. (1994) 'Naturalistic Epistemology and Reliabilism', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Midwest Studies in Philosophy, XIX, Naturalism*, Notre Dame, IN: University of Notre Dame Press.(Expansion of some of the material in §§1-2 of this entry.)

Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press.(Chapter 3 presents a 'tracking' theory of knowledge.)

Schmitt, F.F. (1992) *Knowledge and Belief*, London: Routledge.(The historical section examines reliabilist themes in the epistemologies of Plato, Descartes and Hume.)

Sosa, E. (1991) *Knowledge in Perspective: Selected Essays in Epistemology*, Cambridge: Cambridge University Press.(Several essays defend a form of reliabilism, including a virtues/vices approach similar to the one

sketched in §3 above.)
Wynn, K. (1992) ‘Evidence against empiricist accounts of the origins of numerical knowledge’, *Mind and Language* 7 (4): 315-32.(Expansion of the material on counting in §1 above.)

Religion and epistemology

Epistemology is theory of knowledge; one would therefore expect epistemological discussions of religion to concentrate on the question as to whether one could have knowledge of religious beliefs. However, discussions of religious belief have tended to focus on arguments for and against the existence of God: the traditional theistic arguments on the one hand and, on the other, such arguments against the existence of God as the argument from evil.

To see why, we must think about evidentialism with respect to religious belief ('evidentialism' for short), the doctrine that a religious believer must have evidence for their beliefs if they are to be rationally justified. In particular, they must have propositional evidence: evidence from other things they believe, evidence that can be put forward in the form of argument. And going with evidentialism is the evidentialist objection to religious belief: the objection that religious belief is unjustified because there is not enough evidence for it. Evidentialism begins with the classical foundationalists René Descartes and (especially) John Locke. According to Descartes and Locke, some beliefs are certain for us. There are two kinds of certain belief: first, self-evident beliefs, such as ' $2 + 1 = 3$*', and second, beliefs about one's own mental life, such as 'it now seems to me that I see a hand'. According to Locke, I am, of course, clearly justified in accepting those beliefs that are certain; indeed, it is not within my power to reject them. For any belief that is not certain, however, I am justified in accepting it only if I can see that it is probable or likely with respect to beliefs that* are *certain for me.*

What is this 'justification' and why does it matter whether or not my beliefs have it? Locke believed that human beings are rational creatures: creatures capable of forming, holding and criticizing beliefs. And rational creatures, he thought, have an intellectual duty to believe only those propositions they can see to be probable with respect to beliefs that are certain for them. This is our duty as rational agents. And justification, as Locke thinks of it, is simply the condition of being within your rights, of not having gone against your duties. You are justified in doing a given thing if it is not contrary to duty for you to do it.

Locke's view of this matter has been extremely influential among epistemologists in general and among those who think about the epistemology of religious belief in particular. Furthermore, given his views it is easy to see why there should be so much concern with proofs or arguments for the existence of God. It is not self-evident that God exists - otherwise there would be no atheists and agnostics - and of course the belief that God exists is not about one's own mental life. But then, according to this Lockean way of thinking, anyone who accepts this belief must see that it is probable with respect to what is certain for them, else they will be going contrary to their duty and deserve blame and disapprobation. And proofs or arguments are just the vehicles by means of which one sees (and shows) that a given belief is *probable with respect to what is certain.*

Evidentialism has come to seem less compelling. First, the whole history of Western philosophy from Descartes to Hume shows that there is little one can really see to be probable with respect to what is certain. If we may only believe propositions that meet that condition, then most of what we believe - that there is an external world, that there are other people, that there has been a past - will not be (or will not clearly be) justified. And second, on sober reflection it just does not seem that there really is a duty to restrict belief to what is probable with respect to what is certain.

If we step back for a broader look, we can distinguish two different kinds of question about religious belief, and two corresponding kinds of criticism or objection. First, there are the claims that religious belief - Christianity, say - is not true: it simply is not true, for example, that there is such a person as God, or that Jesus Christ is the divine son of God. We may call such an objection a de facto *objection: the claim is that the religious belief in question is false, is not factual. But there is another kind of criticism or objection as well. Here the claim is not that religious belief is false, but that whether or not it is false, it is in some way improper - unjustified, irrational or in some way not worthy of belief. The evidentialist objection to religious belief is one version of such a* de jure *criticism of religious belief, and it has been the most prominent objection. But there is another* de jure *objection that has been increasingly important. This is the objection, raised by Freud, Marx and Nietzsche, that religious belief is irrational. What does that mean? According to Freud, religious belief arises out of illusion or wish-fulfilment: we find ourselves confronted by a cruel and heartless nature that delivers pain, fear and hurt, and in the end demands our death. As a response, we (subconsciously) invent a father in heaven who loves us and is*

really in charge of nature; otherwise we would sink into depression, stupor and death.

According to Freud, therefore, religious belief arises out of illusion. And the reason this constitutes a criticism of such belief is that this mechanism of illusion is not aimed at the production of true *belief, but rather at belief that has some other, non-truth-related property - in this case, the property of enabling us to carry on in this otherwise discouraging world. It is for this reason that religious belief, on this account, is irrational.*

This is an intriguing criticism of religious belief. But perhaps the most important thing to see about it is that while it is an allegedly de jure *objection to religious belief, it is not really independent of the* de facto *question. For if Christianity is true, then Christian belief pretty clearly is not irrational in Freud's sense at all. If it is true, then indeed there is such a person as God, who intends that we should have knowledge of him; and the cognitive processes that produce belief in God and in the other truths of the Christian religion very likely have as their function the production of true belief in us. On the other hand, if Christianity is false, then it is very likely that Christian belief is not produced by cognitive processes whose purpose it is to produce true belief. This* de jure *criticism of Christian belief, therefore, presupposes that Christian belief is not true; it is viable only if Christian belief is false. If it is intended as a reason for rejecting Christian belief, it is question-begging.*

More generally, it seems that there is no sensible de jure *epistemic criticism of religious belief that is independent of the* de facto *question as to the truth of the belief in question. (The evidentialist criticism is a failure and Freud's complaint is not independent of the* de facto *question whether the religious belief in question is true.) One fairly common critical attitude towards religious belief can be expressed as follows: 'As to whether religious belief is true I am completely agnostic - but I do know this: religious belief is irrational.' The above considerations show that this attitude is at best problematic.*

1 Evidentialism

Epistemology is theory of knowledge: an inquiry into whatever it is that distinguishes knowledge from mere true belief. (For example: due to pathological optimism, you are convinced that you will win the lottery; if by some fluke it turns out you *do* win, your belief will be an example of true belief that is not knowledge.) We need a name for that quality or quantity, whatever precisely it is, that makes the difference between knowledge and mere true belief: call it 'warrant'. Then one would expect the epistemology of religious belief to centre on whether religious belief has warrant, and if so, how much and how it gets it. As a matter of fact, however, epistemological discussion of religious belief, at least since the Enlightenment (and in the Western world), has mostly focused not on the question as to whether religious belief has warrant, but on arguments - in particular, on arguments for and against theistic belief. This is the belief that there exists a person like God as conceived in traditional Christianity, Judaism and Islam: an almighty, all-knowing, wholly benevolent and loving immaterial person who created the world, created human beings in his own image, and continues to act in the world by way of providential care for his creatures.

The most popular theistic proofs or arguments have been the traditional big three - the ontological, cosmological and teleological arguments, to use Kant's terms for them - together with the moral argument (see God, arguments for the existence of). Of these, the teleological argument, the argument from design (as in Swinburne 1979), is perhaps both the most popular and the most convincing. On the other side, the anti-theistic side, the principal argument has traditionally been the deductive argument from evil: the argument that the existence of an omnipotent, omniscient, wholly good God is logically inconsistent with the existence of evil, or with all the pain, suffering and human wickedness actually found in the world (see Evil, problem of). The deductive argument has fallen out of favour over the last quarter-century as philosophers have come to think that there is no inconsistency here; it has been replaced by the probabilistic argument, according to which it is *unlikely* that there is such a person as God, given all the evil the world displays. The argument from evil is flanked by subsidiary arguments, such as the claim that the very concept of God is incoherent (because, for example, it is thought to be impossible that there be an omnipotent person, or an omniscient person, or a transcendent person who can act in the world, or a person without a body). (As we shall see below, in the nineteenth century Freud and Marx introduced a different style of argument against theistic belief, one according to which such belief is irrational in that it arises out of cognitive malfunction or wish-fulfilment.)

Which of these groups of arguments is the stronger? According to *evidentialism*, belief in God is justifiable only if

the former is stronger than the latter.

2 Evidentialism characterized

Why has discussion centred on these arguments, and why does it matter which group is stronger? Indeed, suppose there are no good arguments for religious or theistic belief at all: why should that be thought to create a problem for the believer? To see why we must understand evidentialism with respect to religious belief ('evidentialism' for short). Evidentialism has been the dominant (though not the sole) way of thinking about these matters from the Enlightenment to the present (again, in the Western world). The evidentialist thinks a person who accepts a religious belief must have evidence for that belief. In particular, they must have propositional evidence for it - evidence from other things they believe. And this is why arguments are crucially important: an argument is simply a way of organizing and presenting your propositional evidence for some belief or other. Just to simplify matters, let us follow current custom and think for the most part about theistic belief, belief in God; later we can think also about beliefs that go beyond theism, such as beliefs that distinguish Christianity from Islam and Judaism.

The evidentialist, therefore, thinks a believer in God must have evidence for that belief - but must for what? What will be the matter with them if they do not? The answer, according to the evidentialist, is twofold. In the first place, if they have no (or insufficient) evidence, then the belief will not constitute knowledge. The believer will know that there is such a person as God (that belief will have warrant for them) only if they have propositional evidence for it - only if they have a good argument for it. But the stakes here are considerably higher than that. For if the believer does not have evidence, then, according to the evidentialist, not only will they lack knowledge, they will also be unjustified in holding this belief; they will not be rationally justified.

But what *is* this justification, and why do we need it? Why is lacking it a problem? And why should the justification of theistic belief be so closely linked to the discussion of theistic and anti-theistic arguments? To answer these questions, we must go back to the beginnings of modern and Enlightenment thought on these questions, to the genesis of Locke's *Essay Concerning Human Understanding* (1689), one of the most influential sources of thought on these topics (see Locke, J. §7).

Locke lived through one of the most turbulent periods of British intellectual and spiritual history; it was in particular the religious ferment and diversity, the enormous variety of divergent religious opinion, that caught his attention. There was the Catholic-Protestant debate, and within Protestantism there were countless disagreements and controversies and many warring factions. Locke was deeply concerned about this blooming, buzzing confusion of religious opinion and the civil unrest that went with it. His *Essay* is at least in part designed to ameliorate this problem.

The source of this confusing welter of inconsistent religious opinion, Locke thought, was the propensity of people to indulge themselves in unjustified belief. But what is it to be justified (or unjustified) in holding a belief? According to Locke (as well as René Descartes (§7), the other of the twin towers of modern Western epistemology), there are epistemic or intellectual duties, or obligations, or requirements:

> Faith is nothing but a firm assent of the mind: which if it be regulated, as is our duty, cannot be afforded to anything, but upon good reason; and so cannot be opposite to it. He that believes, without having any reason for believing, may be in love with his own fancies; but neither seeks truth as he ought, nor pays the obedience due his maker, who would have him use those discerning faculties he has given him, to keep him out of mistake and error. He that does not this to the best of his power, however he sometimes lights on truth, is in the right but by chance; and I know not whether the luckiness of the accident will excuse the irregularity of his proceeding. This at least is certain, that he must be accountable for whatever mistakes he runs into: whereas he that makes use of the light and faculties God has given him, and seeks sincerely to discover truth, by those helps and abilities he has, may have this satisfaction in doing his duty as a rational creature, that though he should miss truth, he will not miss the reward of it. For he governs his assent right, and places it as he should, who in any case or matter whatsoever, believes or disbelieves, according as reason directs him. He that does otherwise, transgresses against his own light, and misuses those faculties, which were given him.
>
> (1689: IV.xvii.24)

We have a duty to regulate our belief a certain way. If we do not, we are accountable. One who regulates belief in

this way has the satisfaction of 'doing his duty as a rational creature'; such a person 'governs his assent right and places it as he should'. Rational creatures, creatures capable of holding and withholding belief, have duties and obligations with respect to the regulation of their belief or assent. And the central core of the notion of justification (as the etymology of the term indicates) is this: a person is justified in an action or belief if, in taking that action or holding that belief, they violate no duties or obligations, conform to the relevant requirements, are within their rights. To be justified in believing something is to be *responsible* in forming and holding those beliefs; it is to be flouting no duty in holding them. This way of thinking of justified belief (together with analogical extensions of one kind or another) has remained dominant from Locke's day to this.

Now Locke believed that if only everyone could be persuaded to restrict assent to those beliefs in which they were justified, we would no longer be confronted with this confusing welter of religious belief. His reasoning is as follows. A belief is justified for me, he says, just if I am within my epistemic rights, am flouting no duties or obligations in holding that belief. But what are my duties or obligations in this area? According to Locke, the central epistemic duty is this: to believe a proposition only to the degree that it is probable with respect to what is *certain* for you. Which beliefs are certain for you? There are two kinds. First, according to Locke (here again he concurs with Descartes), there are propositions about your own immediate experience that are certain for you: that you have a mild headache, for instance, or that you are thinking about dinosaurs. And second, there are propositions that are self-evident for you: necessarily true propositions so obvious that you cannot so much as entertain them without seeing that they must be true. (Examples would be simple arithmetical and logical propositions, together with such propositions as that red is a colour, and that whatever exists has properties.) Propositions of these two sorts are certain for you, and you are automatically justified in believing them; as for other propositions, you are justified in believing any one of them, says Locke, only to the degree to which that proposition is probable with respect to what is certain for you. (Here he differs from Descartes, who seems to say that you are justified in believing a proposition that is not certain for you only if that proposition follows deductively from propositions that are certain for you.) According to the whole modern foundationalist tradition initiated by Locke and Descartes, therefore, you have a duty not to accept a proposition unless either it is certain for you, or it is (at least) probable with respect to what is certain for you. If a proposition is not certain for you, you have a duty to refrain from believing it unless you have propositional evidence for it. But it was Locke's belief that if we all did our epistemic duty and believed only propositions we could see to be probable with respect to what is certain for us, we would not find ourselves in disagreement, or at least would not disagree nearly as often.

Returning to theistic belief, we can see how the above thought about justification applies. First, this belief is not certain for us: it is neither self-evident (since there are atheists and agnostics, it is not such that grasping or understanding it guarantees accepting it); nor is it about one's own mental states. Therefore theistic belief, on Locke's way of thinking, is justified only if there is propositional evidence for it - more exactly, a person is justified in accepting theistic belief only if they believe it on the basis of a good argument for it, an argument whose premises are certain for them. And this is the linchpin of the way of thinking about the justification of theistic belief that has prevailed from Locke's time to ours. Theistic belief is justified only if it is probable with respect to what is certain; it is probable with respect to what is certain only if there are good arguments for it; therefore theistic belief is justified only if there are good arguments for it. But then of course it is very easy to see why discussion of theistic belief has tended to focus on the arguments for and against theism.

Now Locke does not *argue* for this position; he simply *announces* it. Subsequent thinkers who have discussed the justification of theistic belief have for the most part followed him here, both in announcing the position and in failing to argue for it. Thus W.K. Clifford trumpets that 'it is wrong, always, everywhere, and for anyone to believe anything upon insufficient evidence' (1879: 183); his is perhaps the most prominent in a considerable chorus of voices insisting that there is an intellectual duty not to believe in God unless you have a good argument for that belief. (A few others in the choir are Brand Blanshard (1974: 400-), H.H. Price, Bertrand Russell and Michael Scriven.) According to all these thinkers, if you accept theistic belief without having a good argument for it, you are going contrary to epistemic duty, are therefore unjustified in accepting it, and are living in epistemic sin.

3 Evidentialism criticized

It is widely thought that there are several important problems with this approach to the epistemology of theistic

belief. First, the standards for theistic arguments have traditionally been set absurdly high (and perhaps part of the responsibility for this must be laid at the door of some who have offered these arguments and claimed that they constitute wholly demonstrative proofs). The idea seems to be that a good theistic argument must start from what is self-evident or utterly obvious and proceed majestically by way of self-evidently valid argument forms to its conclusion. It is no wonder that few if any theistic arguments meet that lofty standard - after all, almost no philosophical arguments of any sort meet it.

Second, attention has been mostly confined to three theistic arguments: the traditional ontological, cosmological and teleological arguments. But in fact there are many more theistic arguments. For example, there are arguments from the nature of proper function, and from the nature of propositions, numbers and sets. There are arguments from intentionality, from counterfactuals, from the confluence of epistemic reliability with epistemic justification, from reference, simplicity, intuition and love. There are arguments from colours and flavours, from miracles, play and enjoyment, from morality, from beauty and from the meaning of life. There is even a theistic argument from the existence of horrifying evil.

But there is a third and much deeper problem. The basic assumption underlying traditional thought about the justification of theistic belief is that such belief is justified only if there is a good argument for it from other propositions you believe. But why believe that? Perhaps *some* beliefs are like that. Scientific hypotheses - special relativity or the theory of evolution, for example - are such that if you believe them without evidence you are irresponsible. For such theories have been devised to explain certain phenomena, and they get all their warrant from their success in so doing. (Even so, would you really be irresponsible in believing such a proposition without evidence, or just foolish or irrational?) But other beliefs - such as memory beliefs, or belief in other minds - are not like that at all; they are not hypotheses, and are not accepted because of their explanatory powers. Now why assume that theistic belief, belief in God, is in this regard more like a scientific hypothesis than like, say, a memory belief? Why think that the justification of theistic belief depends upon the evidential relation of theistic belief to other things one believes?

According to Locke, it is because there is a duty not to assent to a proposition unless you can see that it is probable with respect to what is certain to you; but is there really any such duty? No one has succeeded in showing that, say, belief in other minds or the belief that there has been a past is probable with respect to what is certain for us. Suppose it is not: does it follow that you are living in epistemic sin if you believe that there are other minds? Or a past? Nearly everyone recognizes such duties as that of shunning gratuitous cruelty, of taking care of one’s children and one’s aged parents, and the like; but do we also find ourselves recognizing that there is a duty not to believe what is not probable (or what we cannot see to be probable) with respect to what is certain for us? Hardly. But then it is hard to see why being justified in believing in God requires that the existence of God be probable with respect to some such body of evidence as the set of propositions certain for you. Perhaps theistic belief is properly basic - that is, such that one can be justified in accepting it without accepting it on the evidential basis of other propositions one believes.

For consider a typical Christian believer: they have been brought up as a Christian, and for the most part Christian belief has always seemed to them clearly true. While they have never looked carefully into the alleged objections to Christian belief, what they have heard of them does not seem promising; those whom they respect on these matters tell them the objections are without foundation and they accept this. Such a person, surely, is not to be censured; they are not a proper subject of moral disapprobation. They may be mistaken, deluded, even foolish; they may be insufficiently critical; but there is no reason to think them unjustified or derelict in their epistemic duties.

On the other hand, consider someone sophisticated in these matters and very well aware of the critics. This person does not believe on the basis of propositional evidence; they therefore believe in the basic way. Can they be justified? They read the critics, but on careful reflection do not find them compelling; likewise, although they are aware of theistic arguments and find some of them not without value, they do not believe on the basis of them. Rather, this person has a rich inner spiritual life; it seems to them that they sometimes catch a glimpse of the overwhelming beauty and loveliness of God; they are often aware, as it strongly seems to them, of the work of the Holy Spirit in their heart, comforting, encouraging, teaching, and leading them to accept the ‘great things of the gospel’ as Jonathan Edwards (1746) calls them. After long, hard, conscientious reflection, they find all this

enormously more convincing than the complaints of the critics. Are they then going contrary to duty in believing as they do? Are they being irresponsible? Clearly not. There could be something defective about them, some malfunction not apparent on the surface. They could be mistaken, a victim of an illusion; they could be a victim of wishful thinking, despite their best efforts. They could be wrong, desperately wrong, pitiably wrong, in thinking these things. But they are not flouting any discernible duty; they are doing their level best to fulfil their epistemic responsibilities. They are certainly justified.

Taking justification in that original etymological sense, therefore, there is every reason to doubt that one is justified in holding theistic belief only if one has evidence for it. Of course, the term ‘justification’ has undergone various analogical extensions in the work of various philosophers. Thus it is sometimes used just to mean propositional evidence; then to say that a belief is justified for someone is to say that they have (sufficient) propositional evidence for it. In that sense of ‘justified’, of course, a person will be justified only if they have evidence. But you cannot settle a substantive question just by giving a definition. If you accept that definition, then the real question is whether there is anything amiss with holding beliefs that are unjustified. Perhaps one does not have propositional evidence for one’s memory beliefs; if so, those beliefs would be unjustified in that sense, but none the worse for that; it would not suggest that there is something wrong with holding them. And the same goes for theistic belief.

4 *De facto* and *de jure*

Suppose we try to set evidentialism in a broader perspective. Ever since the Enlightenment, there have been two kinds of critical question about religious belief. On the one hand, there are those who argue that religious beliefs are false or at any rate improbable: it is at best unlikely that (say) there is such a person as God, or that, if there is, Jesus Christ is his divine Son. (Here typical arguments would include the anti-theistic arguments mentioned above.) Since this question is about the truth or factual character of Christian belief, we may call it the *de facto* question. On the other hand, there is the question of the propriety, or reasonability, or justification, or rationality, or to combine those last two, the rational justification of Christian belief. Christian belief may be true and it may be false; but even if it happens to be true, so these critics say, there are serious questions as to whether it is rational or rationally justifiable to accept it. Call this the *de jure* question. The claim that belief in God is unjustified (that is, irresponsible) because there is insufficient evidence for it, is an example of a *de jure* criticism; as we have seen, however, this claim has very little to be said for it. This claim has been the dominant *de jure* criticism of religious belief; but the nineteenth century saw the rise of another kind of *de jure* criticism, one associated with those three great ‘masters of suspicion’, Nietzsche, Freud and Marx.

Freud and Marx insisted that Christian belief is irrational. (Of course it was not only Christian belief that drew their fire). Nietzsche’s complaint could also be examined here: that religion originates in slave morality, in the *ressentiment* of the oppressed (see Nietzsche, F. §§8-9). As Nietzsche puts it, Christianity both arises from and fosters a sort of weak, snivelling, cowardly, servile and generally disgusting sort of character, which is at the same time envious, self-righteous and full of hate disguised as loving kindness. In what follows, we shall ignore Nietzsche, considering the objection Freud and Marx bring against religious belief, with the emphasis upon Freud.

5 Freud and Marx

Some of Freud’s treatment of religious belief is perhaps rather fanciful - for example, his suggestion that religion originated in a remarkable transaction among the ‘primal horde’. (The sons of the dominant male, jealous of their father because he had seized all the women for himself, killed and ate him; religion somehow emerged from the resulting guilt and remorse.) But he also makes a much more sober claim about the ‘psychical origin of religious ideas’:

> These, which are given out as teachings, are not precipitates of experience or end-results of thinking: they are illusions, fulfilments of the oldest, strongest and most urgent wishes of mankind. The secret of their strength lies in the strength of those wishes. As we already know, the terrifying impression of helplessness in childhood aroused the need for protection - for protection through love - which was provided by the father; and the recognition that this helplessness lasts throughout life made it necessary to cling to the existence of a father, but this time a more powerful one. Thus the benevolent rule of a divine Providence allays our fear of the dangers of life; the establishment of a moral world-order ensures the fulfilment of the demands of justice, which have so

> often remained unfulfilled in human civilization; and the prolongation of earthly existence in a future life provides the local and temporal framework in which these wish-fulfilments shall take place.
>
> (1927: 30)

The idea is that theistic belief arises from a psychological mechanism Freud (§4) calls ‘wish-fulfilment’; in this case, the wish is father, not to the deed, but to the belief. Nature rises up against us, cold, pitiless, implacable, and blind to our needs and desires. It delivers hurt, fear and pain, and in the end demands our death. Paralysed and appalled, we invent (unconsciously, of course) a Father in Heaven who exceeds our earthly fathers as much in power and knowledge as in goodness and benevolence; we believe he loves and cares for us. The alternative would be to sink into depression, stupor and finally death. According to Freud, belief in God is an illusion, in a semi-technical use of the term: a belief that arises from the mechanism of wish-fulfilment. But if Freud intends this as a criticism of religious belief, he must be thinking that what he says in some way discredits it, casts doubt upon it, and, in a word, shows that there is something wrong with it. So what, precisely, is the problem? We can put it by saying that religious belief (specifically theistic belief) is irrational; but what does that mean? Irrational in what way? In order to understand what is really involved in that complaint, we must make a brief excursion into the assumptions underlying the sort of criticism it represents.

It is natural to think that there are intellectual, or cognitive, or rational powers or faculties - for example, perception and memory. These powers or processes produce in us the myriad beliefs we hold. They are something like instruments; and like instruments, they have a function or purpose. If we thought of ourselves as created and designed either by a Master Craftsman or by Evolution, these cognitive faculties would be the parts of our total cognitive establishment whose purpose it is to produce beliefs in us. Their purpose, furthermore, is presumably to produce true beliefs in us; to put it a bit less passively, they are designed in such a way that by using them properly we can come to true belief. Our cognitive faculties work over an immensely large area to deliver beliefs on many different topics: beliefs about our immediate environment, about the external world at large, about the past, about numbers, propositions and other abstract objects and the relations between them, about other people and what they are thinking and feeling, about what the future will be like, about right and wrong, about what is necessary and impossible, and about God himself.

These faculties are aimed at the truth in the sense that their purpose or function is to furnish us with true belief. And like any other instruments or organs, they can work either well or ill; they can function properly or malfunction. A wart or a tumour does not function properly, but nor does it malfunction (although it might be by virtue of malfunction that the tumour is present). That is because it has no function or purpose. But an organ - such as your heart, liver or pancreas - does have a function, and does either work properly or malfunction. And the same goes for cognitive faculties or capacities: they too can function well or badly. Further, we ordinarily take it for granted that when our cognitive faculties are functioning properly, when they are not subject to dysfunction or malfunction, then for the most part the beliefs they produce are at least close to the truth. There is, we might say, a presumption of reliability for properly functioning faculties; we are inclined (rightly or wrongly) to take it that properly functioning cognitive faculties for the most part deliver true belief. No doubt there will be mistakes and disagreements, and we may be inclined to scepticism about certain special areas of belief - political beliefs, for example, as well as beliefs formed at the very limits of our ability, as in cosmology and subatomic physics - but the bulk of the everyday beliefs delivered by our rational faculties, so we think, are true.

But when our cognitive faculties malfunction, they do not fulfil their purpose of furnishing us with true belief - or if they do, it is by accident. Insanity is an extreme case of malfunction of the rational faculties. But there are more subtle ways in which irrational or non-rational beliefs can be formed in us. First, there are belief-forming mechanisms that are not aimed at the formation of true belief, but at the formation of belief with some other property - contributing to survival, perhaps, or peace of mind or psychological comfort. Someone with a lethal disease may believe their chances of recovery much higher than the statistics in their possession warrant; their so believing itself improves their chances of recovery. The function of the process producing this belief is not that of furnishing true beliefs, but beliefs that make it more likely that the believer will recover. A person may be blinded (as we say) by ambition, failing to see that a certain course of action is wrong or stupid, even though it is obvious to everyone else. Our idea is that the inordinately ambitious person fails to recognize something they would otherwise recognize; the normal functioning of some aspect of their cognitive powers is inhibited or overridden or impeded by that excessive ambition. You may be blinded also by fear, lust, anger, pride, grief, social pressure, and

even loyalty, continuing to believe in the honesty of your friend long after an objective look at the evidence would have dictated a reluctant change of mind.

So there are at least three ways in which a belief can fail to be a proper deliverance of our rational faculties: it may be produced by malfunctioning faculties, or it may be produced by cognitive processes aimed at something other than the truth, or the proper function of rational faculties can be impeded and overridden by lust, mother love, ambition, greed, grief, fear, low self-esteem and other emotional conditions. And here we come to the heart of Freud’s objection: when Freud says that theistic belief is irrational, his basic idea is that belief of this sort is not produced by the unimpeded proper function of belief-producing processes whose purpose it is to furnish us with true belief. This means that the presumption of reliability attaching to properly functioning cognitive faculties does not apply to the processes that yield belief in God. The idea is that theistic belief has a source distinct from those of our faculties that are aimed at the truth - or, if such belief does somehow issue from those truth-aimed faculties, their operation, in producing such belief, is impeded by other factors. It is therefore irrational; and it is also irrational in the further sense that it is inconsistent with rational belief. For this reason the presumption of reliability does not attach to religious belief.

Marx’s views differ from Freud’s here in an interesting way. Marx (§3) thinks that religion arises from cognitive malfunction: ‘Man is *the world of man*, the state, society. This state, this society, produce religion, *a perverted world consciousness*, because they are a *perverted* world’ (Marx and Engels [1844] 1964: 41-2; original italics). So Marx’s idea, fundamentally, is that religious belief is produced by *malfunctioning* cognitive faculties - malfunctioning in response to social and political malfunction. Freud, on the other hand, thinks theistic belief is an *illusion* in his special sense of a cognitive mechanism that is aimed, not at the truth, but at psychological wellbeing. But illusions have their functions; they may serve important ends (such as the end Freud thinks religious belief does serve). When our cognitive faculties produce illusions, therefore, they are not necessarily malfunctioning. So Marx thinks theistic belief is the product of malfunctioning cognitive faculties, whereas Freud thinks it is the product of cognitive processes not aimed at the truth. They concur in thinking that theistic belief is irrational in the double sense that it is not produced by unimpeded properly functioning cognitive faculties aimed at the truth, and it runs counter to the deliverances of our rational powers. As Freud puts it, religious belief is ‘patently infantile’, and ‘foreign to reality’.

6 The religious riposte

Of course, Christians and other theists will not agree that theistic belief is irrational in this sense; they will have their own views as to how it is that belief in God is formed, their own candidates for the sources of belief in God. Thomas Aquinas (§11), for example, speaks of a ‘natural (and confused)’ knowledge of God. John Calvin (§2) develops this idea, suggesting that we human beings have a *sensus divinitatis* (a sense of divinity) whereby under a wide variety of circumstances - danger, perception of the beauties and wonders of nature, perception of our own sinful condition - we form true beliefs about God. The *sensus divinitatis*, according to Calvin, is a natural, inborn faculty with which human beings have been created; in this respect it resembles perception, memory, reason and other cognitive faculties. What is important in all these views is that our beliefs about God are produced by cognitive faculties aimed at the truth and functioning properly; hence on these views such belief is not at all irrational in the sense intended by Freud and Marx.

There is a complication here: according to Christianity, our natural knowledge of God has been compromised by sin, which has cognitive or noetic as well as moral and spiritual results. As a result, God, as Anselm says, is obscured by the smoke of our wrongdoing. But (again, according to Christianity) God has made a gracious response to our human sinful condition. There is the offer of salvation and eternal life, available through the sacrificial death and resurrection of Jesus Christ, the divine Son of God. But there is also a cognitive component to the divine response. This is God’s providing a way by which we human beings can come to know and appropriate the gracious offer of salvation. Here Christians will think of the Scripture, the Church, and perhaps above all, the inward testimony of the Holy Spirit. The doctrine of the work of the Holy Spirit was developed perhaps most fully in the Reformed tradition and in particular by the great Puritan divines. According to John Calvin, the principal work of the Holy Spirit in the life of a Christian is the production of faith, which includes at the least a deep acceptance of the gracious offer of salvation, and a deep conviction of the truth of the essentials of Christian teaching. In so far as Christian belief (including belief in God) is a result of the work of the Holy Spirit, it is not a

product just of natural faculties; the activity of the Holy Spirit is supernatural. But of course it is not at all irrational in the sense of the objection from Freud and Marx: it is not produced by cognitive faculties that are malfunctioning or aimed at something other than the truth. Instead, these beliefs are produced in us (with our concurrence) by the work of God himself, as part of his gracious response to the human sinful condition. From this point of view, then, Christian faith is in no way irrational, and, indeed, is a form of knowledge (given most accounts of knowledge), so that it is an error to *contrast* faith with knowledge; for faith is one kind of knowledge.

7 *De jure* reduced to *de facto*

Freud, Marx, Nietzsche and their confrères claim that theistic belief is irrational (in the sense explained); Christians and other believers in God deny this. Who is right? Here it is important to see that this question cannot really be settled apart from the question as to whether Christian or theistic belief is true. First, it is clear that if Christian belief is true, then very likely it is rational; it is produced by unimpeded cognitive faculties functioning properly and aimed at the production of true belief. For if Christian belief is true, then we have been created by God and created in his image, one aspect of which involves our being able to resemble him with respect to knowledge. Further, God has instituted a way of salvation for human beings, and has also made available to us the means to know of and apprehend that salvation. When Christians form these beliefs, therefore, they do so by way of mechanisms that are working properly and successfully aimed at the production of true belief. So if Christian belief is true, it is rational. On the other hand, if Christian or theistic belief is false (if, for example, naturalism is true), then these beliefs constitute massive error, and it is hard to see how cognitive faculties functioning properly and aimed at the truth could produce them. Whatever mechanisms do produce them then, those mechanisms must be either malfunctioning or, like wishful thinking, be aimed at something other than the truth.

Once we see this, however, we see an important point about the version of the *de jure* criticism offered by Freud and Marx. What we see is that this question as to the rationality (or lack thereof) of Christian belief is not really just an epistemological question at all; it is at bottom a metaphysical, or theological, or religious question. For it is to be answered in terms of the answer to another question: what sort of beings are human persons, and what sorts of belief do their noetic faculties produce when they are functioning properly? Your view as to what sort of creature a human being is will determine or at any rate heavily influence your views as to what it is rational or irrational for human beings to believe. But the answer to that question depends on whether or not Christian theism is true. And so the dispute as to whether theistic belief is rational, in the present sense, cannot be settled just by attending to epistemological considerations; it is at bottom not merely an epistemological dispute, but a metaphysical or theological dispute.

You may think humankind is created by God in the image of God - and created both with a natural tendency to see God’s hand in the world about us, and with a natural tendency to recognize that we have indeed been created and are beholden to our creator, owing him worship, obedience and allegiance. You may add that the source of distinctively Christian belief lies in the work of God himself. Then of course you will not think of belief in God or Christian belief as (in the typical case) a manifestation of cognitive dysfunction or any other kind of intellectual defect; nor is it a product of some mechanism not aimed at the truth. (It is then more like a deliverance of sense perception, or memory, or sympathy - or perhaps the faculty responsible for a priori knowledge.) On the other hand, you may think we human beings are the product of blind evolutionary forces; you may think that we are part of a Godless universe. Then you will no doubt be inclined to go along with Freud and Marx in seeing belief in God as either a product of cognitive dysfunction or of a mechanism whose function is not that of the production of true belief. If you adopt the former view, you will of course think Christian belief eminently rational; if you adopt the latter you will think it irrational. But the important thing to see is that this dispute cannot be settled by attending only to epistemology: at bottom it is a dispute about the truth of Christian belief. To determine whether Christian theism is rational, therefore, one must first determine whether it is true. But then this *de jure* question - the one associated with Freud and Marx - is not after all independent of the *de facto* question. To answer the former, we must already know the answer to the latter.

What we have seen so far is that there are fundamentally two *de jure* criticisms of Christian theism: the evidentialist objection and that from Freud and Marx. The first is easily seen to be mistaken and the second is not independent of the *de facto* question of the truth of Christian belief. But then it seems that there are not any sensible *de jure* questions or criticisms that are independent of the *de facto* question of the truth of Christianity.

And this means that the only possibly successful objections to Christian or theistic belief will have to be to the *truth* of such belief, not to its rationality, or justification, or intellectual respectability, or rational justification, or whatever. The only possibly successful objections are *de facto* objections; the *de jure* objections drop away.

See also: Agnosticism; Atheism; Deism; Epistemology, history of; Faith; Natural theology; Religion, history of philosophy of §8

ALVIN PLANTINGA

References and further reading

Alston, W. (1991) *Perceiving God*, Cornell, NY: Cornell University Press.(Powerful and probing inquiry into the justification of Christian belief; it concludes that religious belief can receive justification by way of *perceiving* God. Requires a little preliminary work in epistemology.)

Blanshard, B. (1974) *Reason and Belief*, London: Allen & Unwin, 400-.(Elegant statement of a rationalist position with respect to the justification of religious belief.)

Clifford, W.K. (1879) ‘The Ethics of Belief’, *Lectures and Essays*, London: Macmillan.(Classic statement of the position that belief without evidence is unjustifiable and contrary to duty.)

Edwards, J. (1746) *Religious Affections*, ed. J.E. Smith, New Haven, CT: Yale University Press, 1959.(Not explicitly on the topic of the epistemology of religious belief, but contains a wealth of suggestions about the work of the Holy Spirit in producing Christian belief.)

Freud, S. (1927) *Die Zukunft einer Illusion*, Leipzig and Zurich: Internationaler Psychoanalytischer Verlag; trans. and ed. J. Strachey, *The Future of an Illusion*, New York and London: W.W. Norton.(Freud’s account of the nature and prospects of religious belief.)

Gale, R. (1991) *On the Nature and Existence of God*, Cambridge: Cambridge University Press.(An inquiry into the question as to whether there are any good arguments for or against the existence of God. Technical in places.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, with Locke’s ‘Prolegomena’, ed. A. Fraser, Oxford: Oxford University Press, 1894; New York: Dover, 1959, bk IV.(An influential source of the dominant contemporary way of thinking about the justification of religious belief.)

Mackie, J. (1982) *The Miracle of Theism*, Oxford: Clarendon Press.(Perhaps the best contemporary argument for the position that theistic belief is unjustified.)

Marx, K. and Engels, F. (1844) *Contribution to the Critique of Hegel’s Philosophy of Right*, in R. Niebuhr (ed.) *On Religion*, Chicago, IL: Scholars Press, 1964. (Contains Marx’s and Engels’ account and criticism of religious belief.)

Plantinga, A. (1974) *God, Freedom and Evil*, Grand Rapids, MI: Eerdmans.(Argues that there is no inconsistency in the propositions that God exists and that there is evil.)

Plantinga, A. (1998) *Warranted Christian Belief*, New York: Oxford University Press. (Argues that the epistemological objections to Christian belief fail, that Christian belief is warranted if true, and that belief in naturalism is irrational.)

Swinburne, R. (1979) *The Existence of God*, Oxford: Clarendon Press; revised edn, 1991.(Detailed and powerful development of the theistic argument from design.)

Wolterstorff, N. and Plantinga, A. (eds) (1983) *Faith and Rationality*, Notre Dame, IN: University of Notre Dame Press.(An influential collection of essays by the editors and others on the rationality or justification of Christian and theistic belief; for the most part the essays reject classical foundationalism with respect to theistic belief.)

Wykstra, S. (1989) ‘Towards a Sensible Evidentialism: On the Notion of "Needing Evidence"’, in W. Rowe and W. Wainwright (eds) *Philosophy of Religion: Selected Readings*, New York: Harcourt Brace Jovanovich. (Argues that the sort of evidentialism associated with classical foundationalism is extravagant and clearly mistaken, but that there is a more restrained variety of evidentialism that makes much better sense.)

Religion and morality

The relationship between religion and morality has been of special and long-standing concern to philosophers. Not only is there much overlap between the two areas, but how to understand their proper relationship is a question that has stimulated much debate. Of special interest in philosophical discussions has been the question of divine authority and the moral life. If there is a God, how are we to understand the moral status of his commands? Are there moral standards that even God must acknowledge? Or does God's commanding something make *it morally binding? Secular thinkers have insisted that these questions pose a serious dilemma for any religiously based ethic: either the moral standards are independent of God's will, with the result that God's authority is not supreme, or God's will is arbitrary, which means that what appears to be a morality is really a worship of brute power. Many religious ethicists have refused to acknowledge the dilemma, arguing for an understanding of divine moral directives as expressions of the complexities and excellences of God's abiding attributes.*

The impact of religion on moral selfhood has also been much disputed. Secularists of various stripes have insisted that religion is not conducive to moral maturity. Religious thinkers have responded by exploring the ways in which one's notion of moral maturity is shaped by one's larger worldview. If we believe that there is a God who has provided us with important moral information, then this will influence the ways we understand what is to count as a 'mature' and 'rational' approach to moral decision making.

Religious ethicists have had a special interest in the ways in which worldviews shape our understandings of moral questions. This interest has been necessitated by the fact of diversity within religious communities. Different moral traditions coexist in Christianity, for example, corresponding to the rich diversity of theological perspectives and the plurality of cultural settings in which Christian beliefs have taken shape. This complexity has provided some resources for dealing with the 'postmodern' fascination with moral relativism and moral scepticism.

The relationship between religion and morality is also important for questions of practical moral decision. Religious ethical systems have often been developed with an eye to their 'preachability', which means that religious ethicists have a long record of attempting to relate theory to practice in moral discussion. The ability of a moral system to provide practical guidance is especially important during times of extensive moral confusion.

1 God and moral authority

It is not surprising that philosophers have devoted considerable attention to the relationship between religion and morality. Both areas of discourse address the issue of how people are to order their lives so as to promote human flourishing in the light of 'ultimate concerns'. The relationship is also one that easily generates controversies. The notion that morality has a close link to religion has been deeply offensive to those philosophers who have wanted to protect the 'autonomy' of ethics. Their passion in addressing the topic has often been matched by those philosophers who have been convinced that the moral life requires a commitment to 'absolutes' that can only be provided by a religious perspective.

Religious activity extends, of course, well beyond the range of specifically moral concerns. Believers have typically insisted, however, that religious teachings provide the larger context in which the claims of morality find their proper place. Most religious systems portray the good life, including the morally good life, as grounded in obedience to the will of God. (Some religious perspectives are polytheistic, and a few are even properly characterized as atheistic; here, however, we will focus on monotheistic religion (see Atheism §1; Monotheism §§1-2).) It is often argued from a religious perspective that the moral convictions of all human beings, whether or not they consciously acknowledge God's moral authority, presuppose, or are in some other sense dependent upon, a divinely instituted moral ordering of reality. Philosophers who have been critical of religious belief as such have often focused on this alleged connection in arguing for the independence of morality from religious teachings.

The terms of the debate over these issues were established early on in the Western philosophical tradition, when Plato asked in the *Euthyphro* whether the gods themselves must honour the standards of 'the pious or holy'. The issues at stake here have loomed large in philosophical discussions of religion and morality. Are there standards of goodness and obligation that even God must acknowledge, since they are independent of the divine will? Or are we to view divine directives as morally binding simply because they are willed by God? For many religious

thinkers, neither option has been especially attractive as stated. If moral standards have their own reality, independent of the divine will, then God's supreme moral authority is called into question. If, on the other hand, moral standards are themselves created by God, then the divine will would seem to be arbitrary: God could have created a moral order that conflicts in very basic ways with the convictions acknowledged by existing religious systems (see Voluntarism).

Philosophical critics of religious belief have persistently pressed these points. On occasion, they have done so in the hope of demonstrating the basic incoherence of any attempt to provide a religious grounding for morality. If God's will is viewed as arbitrary, then it is difficult to see, the argument goes, how we can rescue the *moral* enterprise. In such a universe, human obedience to divine directives is essentially a subjection to brute power; here, it would seem, duty and moral deliberation have no obvious role to play. Some believers have been willing to concede this point. Peter Geach once gave a blunt answer to the charge that obedience to the will of God is 'plain power worship': indeed it is, he responded, 'but it is worship of Supreme power, and as such is wholly different from, and does not carry with it, a cringing attitude toward earthly powers' (1969: 127).

Many religious ethicists have stopped short of that kind of unnuanced understanding of the situation. In formulating their alternatives, they have generally opted for one of two strategies. The first is to defend a *compatibilist* perspective. Here it is argued that the acceptance of religiously based ethical directives is compatible with adherence to a more generally accepted philosophical perspective. A number of major philosophical ethicists have encouraged such an approach. For example, Jeremy Bentham (§2) defended the ultimacy of the utilitarian principle by attempting to show that the principle of utility was simply an explication of that which is accepted by all human beings, 'wheresoever they have a clear view of their own interest'. He immediately applied this to religious believers: those who submit to the commands of God are aiming at that kind of personal happiness that can best be attained 'by God's appointment, either in this life, or in a life to come' ([1823] 1961: 40). Similarly, John Stuart Mill (§§8-11) attacked the notion that utilitarians advocate 'a *godless* doctrine' by suggesting that to submit to the will of a God who wants his creatures to be happy is to 'fulfil the requirement of utility in a supreme degree' ([1861] 1961: 423). Adopting a very different approach to moral justification, Immanuel Kant (§§9-11) held that many of the moral directives found in the Bible can be viewed as conforming to the categorical imperative. Ethicists working within specific religious traditions have sometimes seized on the opportunities offered by such accounts, arguing that engaging in religious moral practices is quite consistent with the acceptance of one or another of the perspectives in the philosophical tradition. One of the more sophisticated attempts along these lines can be found in William Frankena's 'Ideal Observer' theory of moral justification (Frankena 1976). Frankena contends that in the making of moral judgments we are implicitly claiming a consensus for what we are advocating: to say that something is right or good is to insist, in effect, that we would be sustained in our verdict by everyone who takes the moral point of view, and who thinks about the subject in a clear-headed and rational way in the light of all relevant facts. In short, as Frankena puts it, it is to insist that our position is the one that would hold up on 'the Day of Judgment'. This is not to prove, of course, that the making of moral judgments entails the existence of a divine Judge. But it does suggest that a wise, omniscient and benevolent deity would be in the best position to determine questions of moral value.

A second strategy is to take a *distinctivist* approach, portraying religious ethics as a unique alternative to the standard options in philosophical understandings of moral justification. On such a view, God is not simply a Platonic contemplator of rationally discernible ideals, nor is he a utilitarian calculator, a Kantian universalizer or an Ideal Observer. None of these philosophical perspectives, it is insisted, can do justice to God's moral authority as we experience it from within the religious life. A common line of argument here is that God's will is not arbitrary, but is an expression of his attributes. For example, Eleonore Stump and Norman Kretzmann (1991) have defended such a view by appealing to Thomas Aquinas' thesis that God is essential goodness itself; on this understanding of God's nature, the relationship between God and moral principles is neither one of subjective divine arbitrariness nor of God subscribing to standards that are independent of the divine nature (see Simplicity, divine).

A 'third way' approach of this sort points to the need for a somewhat fluid understanding of the process of moral formation. To the obvious question, for example, of how we know that God's attributes are morally good unless we employ some independent standard of moral value, the answer can be given that the ways in which we learn, refine and employ moral standards and beliefs are complicated. The adult convert to a religious perspective surely

brings certain moral notions to the initial encounter with God; but these notions can themselves be transformed as the person internalizes a new and complex belief system. In such a process, it is not always clear where we apply ‘independent’ standards of evaluation in assessing divine directives and where we submit our previous moral commitments to divine scrutiny for correction and even transformation. Basil Mitchell makes this case by observing that while submission to God’s moral authority does require that ‘we can recognize what is excellent’, it surely is not the case ‘that the only sufficient reason for believing God to be good is our own ability, independently, to think up the divine decrees for ourselves’ (1980: 150).

2 Moral maturity

Critics of religiously based moral perspectives have not only focused on divine moral authority. They have also objected strenuously to the notion that submission to divine authority is a way of promoting human flourishing. The objection takes many forms. Some have argued that a morality based on obedience to a divine will is ‘infantile’ (Patrick Nowell-Smith 1966); others see it as ‘prehuman’ (Erich Fromm) or ‘bad faith’ (Simone de Beauvoir), or as promoting a ‘loss of self’ (Karl Marx (§3)). These various formulations embody a common core objection: that the posture of obedience to a will that is ‘external’ to, or ‘higher’ than, one’s own is an impediment to human flourishing. Moral maturity requires the development of a healthy sense of freedom, as autonomous selves chart their own moral courses without coercion. To accept God’s moral directives, however ‘benevolent’ they might be, is destructive in a very fundamental way to the moral enterprise.

One prominent religious response to this analysis is to challenge the ways in which concepts such as autonomy and maturity are being used. Autonomous selfhood does not require the creation of moral principles and values *ex nihilo*. It is enough to take responsibility for the standards one accepts and the decisions one makes. And such a sense of responsible freedom is not alien to a life that centres on obedience to the will of God. For example, the biblical notion of the ‘covenantal’ framework for the believer’s relationship to God highlights the importance of responsible selfhood in the religious life.

Obedience to the divine will can be viewed as an important manifestation of moral maturity. To be a mature person requires a willingness to take the facts seriously. If one happens to believe that God in fact exists, and is infinitely greater than human beings in his knowledge, goodness and power, then it would be immature, even self-deceptive, to ignore God’s moral greatness. If God is in fact a loving Creator, who desires that human beings experience the divine peace, then to serve such a God is to realize new dimensions of mature human selfhood. Someone might insist, of course, that the factual claims that are basic to a religious perspective - that there is a God, that God has revealed important information about moral reality, that human beings are sinners who are very much in need of moral guidance - are so outlandish that their endorsement is itself a sign of immaturity. While these are important issues to raise, to do so is to shift the discussion away from specifically moral themes to more general philosophical topics having to do with the ways in which we form and justify our beliefs.

A religious understanding of morality, then, cannot be criticized without also challenging the larger complex of beliefs with which it is intimately associated. To criticize such a perspective in the name of autonomy or moral maturity only raises broader questions about the ‘thick’ theoretical frameworks in which those terms can be understood. Ultimately, one is led to issues that have to do not only with the existence and nature of the deity, but with the contours of the human condition, the character and extent of divine revelation, the proper locus of moral authority, and the ways in which religious teachings actually bear on the rich diversity of human practices.

3 Morality and worldview

Religion and morality interact in many different ways. For one thing, the moral universe of the believer is typically more densely populated than it is for other perspectives: even if the believer is, for example, a fairly straightforward utilitarian in their understanding of moral justification, they would still have to take the concerns of saints and angels, to say nothing of God, into account in ascertaining what will promote ‘the greatest good for the greatest number’. Believers also regularly operate with a lengthier sense of the temporal context for thinking about the implications of moral activity, since in most religious traditions moral decisions have eternal significance. Religious teachings are also narratively rich: the believer is provided with abundant examples of spiritual saints, heroes, heroines and villains, stories of people who live out their moral lives in response to a variety of challenges and temptations. These same narratives provide the believer with an expanded sense of what

is morally possible: the belief in miracles and a Final Judgment, and a sense of access to divine sources of strength and blessing, can have an important impact on moral motivation.

If these 'thick' relationships are to be explored properly, the philosophical discussion of religion and morality cannot be confined to formulating philosophical responses to the critics of religious ethics. What is also required is the more positive, constructive task described by Marilyn McCord Adams (1988): she has challenged Christian philosophers to engage in a 'mapping of… alternative positions' that will inevitably lead to the 'mounting of theological theories' about the nature of 'moral value and the human good'. For some religious ethicists, this more constructive dimension of their work has been partially fulfilled by demonstrating the ways in which a religious understanding of morality can be seen as more comprehensive, and/or more satisfactory, than the secular ethical alternatives. One common strategy is to portray non-religious schools of thought as emphasizing, in a reductionistic manner, some aspect of morality that is an integral component of a larger religious scheme. Thus deontologists are attempting to account for what religionists view as God's 'intrinsic' ordering of moral reality; utilitarians are sensitive, albeit in a distorted manner, to the moral universe's divinely ordained *telos*; secularist treatments of virtue and moral character are probing the contours of the Creator's designs for human flourishing; and so on. Some religious ethicists, especially in the Christian tradition, have also delineated the ways in which the biblical call to a life of self-giving love offers a richer understanding of the good life than any non-religious alternative theory.

The mandate to engage in more constructive efforts has also motivated many religious ethicists to explore a rather broad interdisciplinary territory, looking at the ways in which the general patterns of 'the good life' relate to even larger philosophical and theological topics. One way of viewing this broader constructive endeavour in moral thought is to see it as an exercise in a more expanded understanding of the scope of the work of philosophical theologians. Philosophical theology has often been construed too narrowly, as a discipline that focuses exclusively on epistemological questions regarding faith and reason, and metaphysical questions about the nature of God. Philosophical theology can be viewed more broadly, as encompassing the whole theological curriculum, so that it addresses philosophical topics relating to biblical studies, human nature, evangelism and the like. In this sense, moral philosophical theology can be treated as a sub-discipline of philosophical theology, just as moral theology is seen as a sub-discipline of theology proper in standard theological curricula.

Needless to say, these topics are much debated among theologians. The constructive task of religious moral thought, then, must pay considerable attention to the different ways in which specific theological perspectives shape understandings of moral topics. The ethics of Reform Judaism, for example, differ significantly from the moral teachings of the Hasidim. Similar differences exist within, say, Buddhism and Islam. Comparative moral thought has, in fact, been a major preoccupation of Christian thinkers. The Eastern and Western Churches developed divergent understandings of the metaphysics of 'spirit' and 'matter', which has led to alternative prescriptions regarding many significant moral issues (see Sanctification §§1-2). Within Protestantism, to cite just one contrast, thinkers in the Anabaptist tradition have developed an ethical perspective that emphasizes 'simple living' and pacifism, whereas those in the Calvinist tradition have often emphasized 'work ethic' themes and a 'just war' perspective on the use of violence. In Roman Catholicism, where ecclesiastical authorities officially function as a moral magisterium, a significant body of practical moral teachings has been developed, although the fostering of a variety of special 'religious communities' (Benedictine, Dominican, Franciscan, Jesuit, Carmelite, and the like) has also resulted in a diversity of spiritualities (contemplative, activist, eremitic, liberationist, and so forth) that correspond, in turn, to differing patterns of living.

The comparative study of diverse moral perspectives within religious traditions has also focused heavily on cross-cultural issues. Again, this has been a special emphasis within the Christian community, where the phenomenon of 'contextualization' has been emphasized by thinkers who want to draw sympathetic attention to the different ways in which the Christian message is received, appropriated and interpreted in a variety of cultural contexts. Moral matters have loomed large in these investigations, especially when missionary efforts have forced Christians to think about how cultural factors contribute to the shaping of moral differences, even when there is a common core of religious convictions.

The question, though, of how a traditional religious moral outlook can shed light on 'postmodern' discussions of moral relativism and moral pluralism raises new challenges for philosophical theology. The widespread conviction

that ‘the Enlightenment project’, with its strong assumption of the sufficiency of autonomous human reason to address basic issues of human concern, has failed to deliver on its promises, has generated a quest for alternatives. For some, both in academia and popular culture, this has stimulated a renewed interest in ‘pre-modern’ metaphysical schemes; for others, it has meant some sort of fundamental accommodation to radical pluralism, whether in the form of a thoroughgoing, ‘deconstructing’ epistemological and moral scepticism, or a more modest advocacy of a ‘perspectivalist’ approach.

One significant attempt to address this complex of issues has been the emergence in the late twentieth century of a strong interest in the role of ‘narrative’ in the moral life. This interest has been especially strong among religious ethicists. In his elaboration of a Christian narrativism, Stanley Hauerwas (Hauerwas and MacIntyre 1983) has been very explicit about the anti-Enlightenment motives at work. We are not ‘noumenal selves’ who best actualize our natures by formulating moral universals that strip us of our particularities; rather we are ‘historically formed’ communal beings whose selfhood is realized only in the telling of our particularized narratives. The good life, as Hauerwas views it, is most effectively attained when we immerse ourselves in communities whose narratives embody social practices that enable us to be ‘faithful to the one true God of the universe’.

An important question for religious narrativism is, of course, how we know which communities are the ones whose practices are ‘true’ to the will of God. In addressing this question, religious narrativists refuse to join those who express great scepticism about the plausibility of ‘meta-narratives’. They insist on the importance of formulating a ‘larger story’ that makes sense of more particular stories - where ‘making sense of’ includes the ability to expose weaknesses and inadequacies. Here divine revelation is viewed as a resource that provides not primarily a set of commands or moral principles, but a larger worldview, embedded in an elaborate narrative of God’s dealings with humankind.

Discussions of the larger context for understanding the relation between religion and morality have also been strongly influenced by developments in feminist thought. Unlike nineteenth-century feminist reformers, who often took traditional theism for granted in their campaigns for equality, the emergence of feminism in the twentieth century featured a critical re-examination of widely accepted metaphysical and epistemological assumptions. Many issues raised in these explorations focused on topics fundamental to a religious understanding of reality. Some feminists argue, for example, that any notion of a ‘hierarchy’, including the idea that God has transcendent being and power ‘over’ humankind, is designed to reinforce gender oppression.

To be sure, not all feminist thought is critical of divine authority or religious transcendence as such. But even where feminist philosophers and theologians have worked within the general contours of a traditional religious perspective, they have probed deeply entrenched patterns of construing some basic moral themes. In doing so, they have called attention to long-neglected thinkers, such as medieval women mystics, and highlighted less ‘masculine’ motifs (the ‘ethics of care’, relationality) in religiously based moral reflection (see Feminist ethics; Feminist theology).

4 Practical ethics

The relationship between religion and morality has also been highlighted in the late twentieth century by an intense concern with practical moral questions. Alasdair MacIntyre has suggested that there are ‘periods in the history of morality when the deepest moral divisions exist in areas of central human concern: social justice, war, sexuality, the uses of technology’ (Hauerwas and MacIntyre 1983: 1). When these times come, many people who are not themselves professional philosophers see a need to discuss practical moral concerns ‘with an unusual degree of articulateness and theoretical depth’, and they want very much to know what academic moral philosophy ‘has to contribute at the level of practice’. Of course, the capacity to generate practical guidance is not by itself a test of the truth of a theoretical perspective on morality. But practical relevance is certainly one factor that is to be taken into account in the overall assessment of an ethical system.

Both theologians and philosophers are notable for the degree of abstraction that characterizes their investigations. But the theological enterprise has typically been closely connected with efforts to ‘preach’ theories in popular settings dominated by practical concerns. These religious efforts have often been characterized by an unusual degree of ‘thickness’. In discussions of medical ethics, for example, religious ethicists have regularly insisted that the ‘ought’ topics addressed by medical ethics are seldom simply questions of moral rightness and wrongness.

Since they also have to do with our larger worldview, we must pay explicit attention to a broad range of philosophical and theological topics: metaphysical, epistemological and anthropological questions concerning the genesis and cessation of human life; dualistic and monistic accounts of human composition, as they touch on basic issues concerning the significance of medical treatments in the lives of human persons; and many others. What is true for medical ethics will also hold for other areas where practical moral guidance is offered, although it is not only religious believers who will insist on calling attention to these complexities. Proponents of a variety of non-religious points of view, such as Marxists, Freudians, deconstructionists and secular feminists, have their own reasons for insisting on the recognition of theoretical 'thickness'. Here, as in other discussions of moral topics, though, the religious believer has a special role in pointing to the unique and important ways in which religious questions enter the picture.

See also: Halakhah; Morality and ethics

RICHARD J. MOUW

References and further reading

Adams, M.M. (1988) 'Problems of Evil: More Advice to Christian Philosophers', *Faith and Philosophy* 5 (April): 121-43.(Cited above for the case it makes for placing philosophical analyses of specific issues in the larger context of theologically informed theories of reality.)

Adams, R.M. (1987) *The Virtue of Faith and Other Essays in Philosophical Theology*, New York: Oxford University Press.(Includes several essays that have been highly influential in discussions of religion and morality.)

Alcoff, L. and Potter, E. (eds) (1993) *Feminist Epistemologies*, New York: Routledge.(An anthology of essays setting forth feminist perspectives on basic philosophical questions, including religious and ethical ones.)

Benhabib, S. (1992) *Situating the Self: Gender, Community and Postmodernism in Contemporary Ethics*, New York: Routledge.(Essays covering a broad range of themes in postmodern and feminist moral thought.)

Bentham, J. (1823) *An Introduction to the Principles of Morals and Legislation*, in *The Utilitarians*, Garden City, NJ: Doubleday, 1961.(Cited above for its argument for utilitarianism's compatibility with a religious belief in the afterlife.)

Frankena, W.K. (1976) *Perspectives on Morality: Essays by William K. Frankena*, ed. K. Goodpaster, Notre Dame, IN: University of Notre Dame Press.(Includes several essays in which Frankena elaborates on his 'Ideal Observer' theory of moral justification.)

Geach, P. (1969) *God and the Soul*, New York: Schocken Books.(Argues in favour of obeying divine commands that are backed by 'supreme power'.)

Hauerwas, S. and MacIntyre, A. (eds) (1983) *Revisions: Changing Perspectives in Moral Philosophy*, Notre Dame, IN: University of Notre Dame Press.(The collected essays explore various links between religion and morality; the editors' essays provide helpful overviews.)

Helm, P. (ed.) (1981) *Divine Commands and Morality*, New York: Oxford University Press.(A helpful anthology setting forth various perspectives on the relevance of God's directives for the moral life.)

Mill, J.S. (1861) *Utilitarianism*, in *The Utilitarians*, Garden City, NJ: Doubleday, 1961.(Cited above for its argument that utilitarianism is not 'a godless doctrine'.)

Mitchell, B. (1980) *Morality: Religious and Secular*, Oxford: Clarendon Press.(A broad-ranging discussion of philosophical and theological themes in ethics.)

Nowell-Smith, P.H. (1966) 'Morality: Religious and Secular', in I.T. Ramsey (ed.) *Christian Ethics and Contemporary Philosophy*, New York: Macmillan.(Cited above for its critique of religious ethics as 'infantile'.)

Quinn, P.L. (1978) *Divine Commands and Moral Requirements*, Oxford: Clarendon Press.(A careful study of issues at stake in formulating a divine-command ethics.)

Stump, E. and Kretzmann, N. (1991) 'Being and Goodness' in S. MacDonald (ed.) *Being and Goodness: The Concept of the Good in Metaphysics and Philosophical Theology*, Ithaca, NY: Cornell University Press. (Develops Aquinas' thesis regarding the relationship between God's being and goodness; the collection includes several essays on morality and the divine nature.)

Religion and political philosophy

Political philosophy began in Athens, but the large-scale impact of religion upon it had to await Christianity. Biblical Christianity portrays human beings as subjects of a kingdom of God, destined for a supernatural end and bound to love one another. This view is potentially in tension with the demands of the various political societies to which Christians belong. The requirement of devotion to God might conflict with the allegiance that temporal government demands; human beings' attempts to attain their supernatural end can bring them into conflict with civil laws. The power and structure of the Church in the Middle Ages opened the possibility of tensions between the authority of the institutional Church and of various national states. These tensions, potential and actual, set much of political philosophy's agenda from the fourth to the fourteenth century.

The tension between membership of the kingdom of God and of an earthly polity was forcefully described by Augustine. He likened faithful Christians to pilgrims journeying through the world, who avail themselves of the peace temporal authority provides. Political thinkers of the early Middle Ages examined the conditions under which war, regicide and disobedience were permissible, and queried whether the Pope had authority over temporal rulers. Thomas Aquinas elaborated a theory of natural law according to which valid human law cannot conflict with the dictates of morality. Temporal rulers, he argued, are responsible for promoting their subjects' common good and eternal salvation.

Since the sixteenth century, political philosophy has been concerned with problems set by the religious developments that ushered in the modern period. The Reformation brought religious diversity to European nations on a large scale. It thereby raised questions about how policy could be set and unity maintained without a shared religion to provide common goals and social bonds. The seventeenth-century philosopher Thomas Hobbes opposed the toleration of religious diversity and argued that states could remain unified only if their religious unity were maintained by an absolute sovereign. John Locke, on the other hand, argued for the right to religious liberty. Locke and other liberals associated with the movement of thought known as the Enlightenment were opposed by classical conservatives such as Edmund Burke. Burke argued that human society depended upon willing adherence to traditional customs and social institutions, including an established national Church. More recently, liberalism has also been opposed by Marxism. Marxists argue that religion helps to maintain social stability under modern conditions by masking the exploitation of the working class.

Contemporary political philosophy in the English-speaking world is descended from the Enlightenment liberalism of Locke. John Rawls argues that social cooperation must be based only upon what citizens of liberal democracies can reasonably affirm under ideal conditions. Religious critics of contemporary liberalism argue that it unduly restricts religiously inspired political argument and activism.

1 Political philosophy in the Middle Ages

Classical Athens was the birthplace of systematic philosophical reflection on politics, but Greek religion made little impact upon it. There have been bodies of political thought associated with Judaism, Islam and Byzantine Christianity, but these religions too exercised little lasting influence on the development of political philosophy. By contrast, the development of the subject has been profoundly influenced by the doctrinal and institutional developments of Latin Christianity and the religions descended from it.

The Christian Scriptures portray all human beings as equal members of God's kingdom, bound to love God with their whole hearts and their neighbours as themselves. The Scriptures are also standardly interpreted as promising righteous people eternal life with God. The demands of the Christian life portrayed in the Scriptures are potentially in tension with the demands of membership of political society. Conformity with God's commands might require disobedience to human law. The demands of Christian charity may conflict with the requirements of armed military service. National unity may seem to depend less upon mutual love than upon patriotism or economic reciprocity. Laws, it might seem, are more often used to promote national purposes than eternal salvation. During the Middle Ages, the increasing power, wealth and organization of the Church paved the way for jurisdictional controversies between political authorities and the institutional Church. These controversies centred on the relative primacy of the pope and various secular rulers, the government of Church property, the power to appoint bishops, and the laws governing the clergy. Political philosophy between the fourth and the fourteenth centuries was

concerned in large part with addressing these tensions, both potential and actual.

Augustine, the fourth-century North African bishop, was the first great Christian social thinker. His political thought is unsystematic, but he frequently touched on political subjects in his greatest work, *The City of God* (413-27). There Augustine argued that political authority is inherently coercive and would be unnecessary had the Fall not occurred. Pride, Augustine thought, is the legacy of this original sin and the source of actual sin. God ordained that humans be subject to political authority in part to teach them humility. Augustine enjoins Christians not to take pride in the glory of the state to which they belong. Christians are, he said, pilgrims journeying through a sinful world as through an alien land. They are to make use of the peace earthly authority makes possible, but are not to love the things of this world. They may hold positions of authority, but ought not covet them. Political authority, Augustine argued, should sometimes be used to suppress heresy and to force schismatics to return to the Church.

The adoption of Christianity as the official religion of the Roman Empire by the emperor Constantine (*c*.275-337) and the Empire's later collapse brought the Church a power and wealth which required increasingly complex administration. This increasing organizational complexity set the stage for both conflict and cooperation between ecclesiastical and political institutions. Potential conflicts of jurisdiction between ecclesiastical and civil authorities, the power of the Pope to depose temporal rulers, and the permissibility of killing tyrants were among the subjects that engaged political thinkers between Augustine and the thirteenth century. The political thought of this period was largely occasional in character. It was more a product of immediate political circumstance than an element of larger philosophical enterprises.

Political philosophy was reinvigorated in western Europe in the thirteenth century. This occurred when the works of Aristotle were rediscovered by the Latin-speaking world and introduced into the medieval universities. Thomas Aquinas, the greatest of Aristotle's followers in the Middle Ages, was a practitioner of the method of inquiry known as scholasticism. It was characteristic of scholastics to integrate the truths they found in diverse sources into systematic wholes. Thus Aquinas sought to combine the truths he found in the Scriptures and in patristic sources with the insights of Aristotle's writings.

Aquinas touched on a number of political questions while constructing his own philosophical and theological system. In his treatment of politics, as of other subjects, he attempted to combine what he learned from Aristotle and from the Christian tradition. Thus he argued, following Aristotle (§27), that the purpose of political society is to make the good life available to its members. Political authorities, he concluded, are responsible for promoting the common good and eternal salvation of their subjects. Because he thought that political authority had these positive functions, he disagreed with Augustine's view that subjection to political authority is merely a consequence of sin.

Aquinas' most famous contribution to political philosophy is his theory of law (*c*.1270) (see Aquinas, T. §13). He argued that the precepts of human morality are found in what he called 'the natural law'. The natural law is ordained by God to lead human beings to their good and is accessible to human reason. It consists of injunctions such as 'Parents are to be honoured' and prohibitions such as 'Innocents should not be killed'. Human law, Aquinas argued, must be consistent with natural law. A law enjoining the ritual killing of the innocent, for example, would be inconsistent with the natural law. Aquinas would therefore argue that such a law is not, properly speaking, a law at all, and that no-one is obliged to obey it. Aquinas also thought that some societies - various Christian societies, for example, and the Israelites of the Hebrew Scriptures - enjoy the benefit of God's revelation. Through revelation, these societies have access to what Aquinas called 'divine law'. The divine law includes precepts about how God is to be worshipped. Authorities in societies to which God has revealed such precepts should enact laws which promote proper worship, where such laws are feasible. They also have a duty to suppress heresy where possible.

Another of Aquinas' important contributions to political philosophy is his account of the conditions under which a war is just. He argued that wars must be declared by proper authority rather than be the result of private initiative. He also agreed with Augustine that combatants must have the right motives. They ought not be moved by belligerence or desire for revenge. The war must be fought for a just cause, such as self-defence or the defence of the innocent. Finally, the damage done and casualties inflicted must be proportionate to the offence done by the enemy and must be necessary to achieve the just aims of the war.

In the fourteenth century, questions about the proper extent of Papal authority provoked bitter political controversy and philosophical debate. William of Ockham (§11) and Marsilius of Padua, a follower of Aristotle, argued for severe limitations on papal authority over political rulers and political matters.

During much of the Middle Ages, political philosophers thought of Europe as a single entity unified by religion. The common religion was Christianity, headed by the pope, whose proper relationship to various temporal rulers remained a matter of debate. The ascendancy of Holy Roman Emperors, such as Charlemagne in the ninth century and Frederick II in the thirteenth, kindled hopes for a Europe that would be politically as well as religiously unified. In such a world, it was natural for political theorists to argue that religion grounded the legitimacy of political authorities. Rulers, it was sometimes thought, derived their moral authority from the fact that they were doing God's will. Thus Charlemagne was crowned by the pope on Christmas Day. He was often portrayed holding a copy of Augustine's *The City of God* and thought of his empire as an attempt to realize the city of which Augustine wrote. Much later, Frederick II employed highly religious imagery to legitimate his authority. He portrayed himself as a Christlike figure who pronounced law on God's behalf.

The Reformation early in the sixteenth century brought large-scale religious diversity to Europe and effectively ended the hope of a unified Christendom. It thereby posed new questions for political philosophy, questions about how social unity could be maintained in the face of religious diversity, about what the proper ends of politics should be, and about what gives rulers their political legitimacy.

2 Political philosophy in the modern period

One challenge to the claim that human beings naturally form a spiritual unity came with the discovery of the Americas. The native peoples of the Americas had had no previous exposure to Christianity and lived very differently from Europeans. European political thinkers of the sixteenth century therefore asked what the status of the native peoples was and how Europeans were obliged to treat them. Some argued that the native peoples were either animals or natural slaves who could be deprived of their property. Most religious political philosophers, notably the Spaniards Francisco de Vitoria, Luis de Molina (§4) and Domingo de Soto (§4), argued that the native peoples of the Americas were fully human, that they should not be compelled to accept Christianity, that their property rights should be respected and that the *conquistadores* should return what they had taken. These arguments, sadly, had little practical impact.

The last great defender of a unified Christendom was Robert Bellarmine, a cardinal of the Roman Catholic Church and opponent of the Reformation. In his *De summo pontifice*, Bellarmine argued that human beings have a common human nature and common supernatural end, and are thus spiritually united. He concluded that all human beings are subject to a single ultimate spiritual authority, whom he identified as the pope. Because the pope has ultimate authority over spiritual matters, he also has authority over those political matters that affect people's spiritual welfare. Bellarmine concluded that the pope has the power to declare laws invalid that endanger the one true faith and to depose rulers who fall into heresy or who leave the one true Church.

Like Bellarmine, the English philosopher Thomas Hobbes (§7) recognized that religious diversity was a source of political conflict. Religious diversity, Hobbes recognized, carries with it a diversity of religious authorities. Members of political societies who owe their allegiance to different religious authorities can find themselves under obligations which bring them into conflict. Some, for example, may be obliged to worship in ways that others find intolerable. Rulers who follow different religions can think themselves obliged to spread their religion to other nations. Like Bellarmine, Hobbes therefore located the source of religiously based conflict in the potentially conflicting demands of diverse authorities. Where Hobbes and Bellarmine disagreed was on how these problems were to be addressed.

In *Leviathan* (1651), Hobbes denied Bellarmine's claim that all human beings naturally form a spiritual unity which has but one spiritual leader. Instead, he argued, groups of human beings unite themselves by subjecting themselves to a common sovereign. Hobbes thought that the toleration of religious diversity within a state was an invitation to civil war. He also thought that recognizing a spiritual authority (such as the pope) not subject to the national sovereign would breed political division. A sovereign, Hobbes argued, can maintain the unity and peace of the subjects only if the sovereign has absolute authority in both political and spiritual matters. Therefore each national state is to have a Church of its own. The doctrine and practices of a national Church should, in Hobbes'

view, be dictated by a political ruler with the power to punish dissenters.

According to Hobbes' analysis of moral obligation, individuals can have obligations only when there is someone with absolute power to ensure compliance. A contract between two people is morally binding, for example, only if there is someone with absolute power to enforce its terms. It follows that members of one society, with religious and political obligations to its absolute sovereign, cannot have any conflicting obligations to a foreign religious authority such as the Pope. Moreover, no one exercises absolute sovereignty over the world. It follows, therefore, that no national sovereign can have an obligation to spread religion to other nations, since there is no one to compel the sovereign to fulfil that obligation.

John Locke (§7) endorsed a different solution: religious toleration. This is the doctrine that state power should not be used to enforce religious conformity. Locke's defence of this doctrine, as found in his *A Letter Concerning Toleration* (1689), rests upon two arguments. First, Locke argued that religious belief is a mental state into which people cannot be forced. It is therefore irrational to employ coercive force in an attempt to secure religious conformity. Second, Locke held that religious diversity is not itself a source of social and political conflict. What leads to such conflict is religious persecution. Cease persecuting dissent, and religiously based political conflict will largely disappear.

The toleration Locke defended was limited in scope. He was suspicious of Catholics because of their allegiance to the pope. He was also suspicious of atheists who, he thought, lacked sufficient motivation to perform their civic duty because they lacked a belief in eternal punishment. Locke was therefore reluctant to extend religious toleration to members of either of these groups.

Locke said that public officials may prohibit religious conduct if it interferes with valid political purposes, but denied that they may promote or discourage religion as such. His view, as he lays it out in his *Letter*, therefore contains the germ of an important idea: the separation of Church and state. However, Locke did not explicitly develop and defend this view himself. Church-state separation and religious toleration were, however, very important ideas to the founders of the American republic a century later. Some, such as Thomas Jefferson, defended religious freedom because of the importance they attached to individual liberty. James Madison (1785) argued against establishing the Church of the majority. He thought that establishment and the governmental support that came with it conferred too privileged a position on the Church in question, which in turn led to complacency and even corruption among the clergy. Thus disestablishment, Madison concluded, benefited the Church of the majority as well as minority denominations.

Historians of philosophy commonly associate Locke and Jefferson with the movement of thought known as the Enlightenment. Its most prominent philosophers were Locke, Montesquieu, Rousseau and Kant. These Enlightenment liberals reposed great confidence in the power of human reason to understand the natural world and to solve social and ethical problems. This confidence in human reason was central to the political philosophies of Enlightenment philosophers. Political thinkers of the Enlightenment typically argued that political authority is legitimate only if society is organized in ways that each individual could understand and approve, at least under proper circumstances.

The Enlightenment account of legitimacy was not only a reaction to those such as Hobbes who thought that rulers should have absolute power. It was also a reaction against earlier thinkers in the modern period who invoked religion to legitimate political authority. According to the doctrine of the divine right of kings, for example, monarchs were entitled to rule by the grace of God and were responsible only to God for their rule. Divine-right theory was at its height in England in the century prior to the Civil Wars of 1642-51 and in France under Louis XIV.

The insistence that the legitimacy of a political order rests upon the rational consent of individuals who live under it has provoked charges that Enlightenment liberalism is unduly individualistic and rationalistic. Enlightenment liberalism has also been criticized as overly hostile to religion, in part because it purposely excludes religion from its account of the foundations of government. Whether or not the charge of hostility can be sustained, the fact remains that the theological doctrines of Catholic and Protestant Christianity had little impact on the substantive views of Enlightenment political philosophers. These thinkers were, however, concerned with the power of ecclesiastical institutions. They generally advocated religious toleration and favoured restricting the political

influence of the clergy and of institutions such as the Catholic Church.

Among the critics of Enlightenment liberalism was the British thinker and political figure Edmund Burke (§3). Burke is considered a classical conservative. While contemporary conservatives are usually identified by the value they attach to the operation of free markets and to minimalist government, classical conservatives are identified by the value they accord to tradition. Thus Burke argued that emphasis on individual rational consent neglects the importance of tradition in maintaining political unity and in setting social aims. He claimed that social unity and peace depend upon the willing adherence to traditional mores, social arrangements and institutions. These institutions should, in Burke's view, include an established national Church such as the Church of England.

More recently, liberalism has also been criticized by Karl Marx (§3) and his followers. Marxists argue that religion is an ideology. It helps to maintain social stability under conditions of industrial capitalism by obscuring the operations of an economic system that exploits the working classes. It does so by systematically encouraging false social analysis. Its emphasis on individual charity as an ideal human motivation masks the fundamental character of economic forces in human history. More importantly, its other-worldly emphasis distracts members of the working class from their condition and makes them more content with it than they would otherwise be. A fully communist society would abolish exploitation of the working class and vest political power in its hands. In such a society, religion would wither away because the felt need for it would disappear. Twentieth-century countries that have purported to be Marxist - such as China, the former Soviet Union, the former East Germany, and Albania - have discouraged or prohibited organized religion, ostensibly to help realize full communist society.

3 Contemporary political philosophy

The first half of the twentieth century saw a revival of interest in the thought of Thomas Aquinas among Catholic scholars. With this Thomistic revival came a renewed interest in natural law political philosophy (see Natural law; Thomism). Natural law theory has traditionally emphasized the common good as the goal of government. Pursuit of the common good, it is sometimes thought, must come at the expense of individuals whose opinions and ways of life are at odds with it. Natural law theory has therefore been criticized for an inability to accommodate the rights of individuals. This is a criticism that twentieth-century natural law thinkers have tried to meet. Jacques Maritain (§3), for example, argued that natural law theory could lend philosophical support to the UN's Declaration of Human Rights. John Courtney Murray, an American Jesuit, used natural law reasoning to argue that the Catholic Church should accept freedom of conscience (1960). The Catholic Church endorsed Murray's position in 1965 at the Second Vatican Council. John Finnis' *Natural Law and Natural Rights* (1980) is an attempt to ground natural rights in an original, and controversial, account of natural law.

Most contemporary political philosophy in the English-speaking world is descended from the Enlightenment liberalism of Locke and Kant. Contemporary liberalism finds its most sophisticated expression in *A Theory of Justice* (1971) by the American political philosopher John Rawls. Rawls claims that liberal democracies must conform with moral principles that would be acceptable to every individual. This claim raises questions about the circumstances under which hypothetical consent must be registered. Rawls argues that it should be rendered under circumstances of freedom and equality.

Consent to moral principles cannot, Rawls says, depend upon the religious or philosophical beliefs of the consenting parties. Therefore the moral foundations of liberal democracy are, on Rawls' view, independent of religion. He also believes that religion should play only a very restricted role in debates about legislation and public policy. He argues, for example, that using *only* religious considerations to justify sharp legal restrictions on abortion compromises the reasonableness and civility of public debate. Other contemporary liberals, such as Robert Audi, have defended even more severe moral (not legal) restrictions than Rawls endorses. Audi argues that it is wrong to participate in political debate from religious motives, even if the arguments introduced into debate are thoroughly secular (Audi 1989). Such religiously motivated participation, Audi says, compromises reasonableness and civility.

Some religious thinkers (Wolterstorff 1997, for example) have criticized Rawls and other liberals for these restrictions. They argue that religion is so important to its adherents that they should be able to express their convictions in public debate. They also point out that religious themes and stories have a political appeal that transcends denominational lines. Use of religious imagery and argumentation, they maintain, is therefore not as

unreasonable as they think Rawls, Audi and other liberals take it to be.

See also: Liberation theology; Political philosophy, nature of: Religion and morality

PAUL J. WEITHMAN

References and further reading

Aquinas, T. (*c*.1270) *The Treatise on Law*, trans. R.J. Henle, Notre Dame, IN: University of Notre Dame Press, 1993.(A recent translation of Aquinas' discussion of law in the *Summa theologiae*.)

Audi, R. (1989) 'The Separation of Church and State and the Obligations of Citizenship', *Philosophy and Public Affairs* 18 (3): 259-96.(Defends normative constraints on the role of religion in liberal democracies committed to separating Church and state.)

Augustine (413-27) *The City of God*, trans. H. Bettenson, New York: Penguin, 1972.(Augustine's most important work of social theory; see especially Book XIX.)

Bellarmine, R. (1581-92) *De summo pontifice*, V, in *Opera omnia*, vols 1-2, ed. J. Fèvre, Paris: Louis Vives, 1870-4.(This is the third of Bellarmine's *General Controversies*.)

Black, A. (1992) *Political Thought in Europe 1250-1450*, Cambridge: Cambridge University Press.(A brief and historically sensitive discussion of the political thought of the late Middle Ages.)

Bossuet, J. (1709) *Politics Drawn from the Very Words of Holy Scripture*, trans. and ed. P. Riley, Cambridge: Cambridge University Press, 1990.(The *locus classicus* of divine-right theory.)

Carlyle, R.W. (1950) *A History of Medieval Political Theory in the West*, New York: Barnes & Noble.(Accessible and thorough history of medieval political philosophy.)

Filmer, Sir R. (1680) *Patriarcha*, ed. J.P. Sommerville, Cambridge: Cambridge University Press, 1991.(This work, the classical source of English divine-right theory, was composed sometime between 1636-40 and enjoyed wide circulation in manuscript form until its publication in 1680. Locke wrote his *First Treatise of Government* in the late 1670s in reply.)

Finnis, J. (1980) *Natural Law and Natural Rights*, Oxford: Oxford University Press.(Contemporary statement of natural law political philosophy.)

Hobbes, T. (1651) *Leviathan*, ed. C.B. MacPherson, New York: Penguin, 1985.(A summary and definitive statement of Hobbes' political philosophy.)

Locke, J. (1689) *A Letter Concerning Toleration*, ed. J. Tully, Indianapolis, IN: Hackett, 1983.(Locke's defence of religious toleration.)

Madison, J. (1785) 'A Memorial and Remonstrance', in S. Padover (ed.) *The Complete Madison*, New York: Harper, 1953.(Madison's argument against establishment.)

Murray, J.C. (1960) *We Hold These Truths*, New York: Sheed & Ward.(An influential example of a natural law argument.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(The major contemporary statement of the liberal position.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(A collection of Rawls' papers; essay 6 addresses the place of religious considerations in political debate.)

Skinner, Q. (1978) *The Foundations of Modern Political Thought*, Cambridge: Cambridge University Press.(A history of political philosophy from the late middle ages to the seventeenth century. Contains an excellent bibliography.)

Wallace Hadrill, J.M. (1965) 'The *Via Regia* of the Carolingian Age', in B. Smalley (ed.) *Trends in Medieval Political Thought*, Oxford: Blackwell, 22-41.(A useful and accessible essay on the political thought of Charlegmagne's time.)

Walzer, M. (1965) *The Revolution of the Saints*, Cambridge, MA: Harvard University Press.(A study of the Puritan origins of radical politics.)

Wolterstorff, N. (1997) 'Why We Should Reject What Liberalism Tells Us About Thinking and Acting in Public for Religious Reasons', in P.J. Weithman (ed.) *Religion and Contemporary Liberalism*, Notre Dame, IN: University of Notre Dame Press.(An accessible criticism of the restrictions on religion in political debate desired by some liberals.)

Religion and science

Philosophical discussion of the relation between modern science and religion has tended to focus on Christianity, because of its dominance in the West. The relations between science and Christianity have been too complex to be described by the 'warfare' model popularized by A.D. White (1896) and J.W. Draper (1874). An adequate account of the past two centuries requires a distinction between conservative and liberal positions. Conservative Christians tend to see theology and science as partially intersecting bodies of knowledge. God is revealed in 'two books': the Bible and nature. Ideally, science and theology ought to present a single, consistent account of reality; but in fact there have been instances where the results of science have (apparently) contradicted Scripture, in particular with regard to the age of the universe and the origin of the human species.

Liberals tend to see science and religion as complementary but non-interacting, as having concerns so different as to make conflict impossible. This approach can be traced to Immanuel Kant, who distinguished sharply between pure reason (science) and practical reason (morality). More recent versions contrast science, which deals with the what and how of the natural world, and religion, which deals with meaning, or contrast science and religion as employing distinct languages. However, since the 1960s a growing number of scholars with liberal theological leanings have taken an interest in science and have denied that the two disciplines can be isolated from one another. Topics within science that offer fruitful points for dialogue with theology include Big-Bang cosmology and its possible implications for the doctrine of creation, the 'fine-tuning' of the cosmological constants and the possible implications of this for design arguments, and evolution and genetics, with their implications for a new understanding of the human individual.

Perhaps of greater import are the indirect relations between science and theology. Newtonian physics fostered an understanding of the natural world as strictly determined by natural laws; this in turn had serious consequences for understanding divine action and human freedom. Twentieth-century developments such as quantum physics and chaos theory call for a revised view of causation. Advances in the philosophy of science in the second half of the twentieth century provide a much more sophisticated account of knowledge than was available earlier, and this has important implications for methods of argument in theology.

1 Religion and Western predecessors of science

Western interest in a systematic account of the natural world is an inheritance from the ancient Greeks rather than from the Hebrew tradition, which tended to focus on the human world. The Greek concept of nature was not set over against a concept of supernature, as it has been in more recent centuries, so it is possible to say that Greek philosophy of nature was inherently theological. Early Christian scholars were divided in their approach to Greek natural philosophy, some making great use of it for apologetic purposes (see Origen §3; Augustine §§1, 4, 8), others rejecting it (see Tertullian).

After the fall of Rome, the centre of scholarship shifted eastward. Islamic scholars in the Middle Ages were largely responsible for preserving the learning of the Greeks, as well as for significant scientific developments of their own in the fields of optics, medicine, astronomy and mathematics. It was through Muslims in Spain that important scientific works by Aristotle were introduced to western Europe in the twelfth century. The influence of these works on Christian thought culminated in Thomas Aquinas' two Summas (see Aquinas, T. §§5-6; Aristotelianism, medieval §§3-4).

2 Early modern science and worldview

At the end of the nineteenth century, White (1896) and Draper (1874) promoted the view of science and religion as traditional enemies. However, revisionist history at the end of the twentieth century presents a much more complex picture. It is true that the Catholic Church silenced Galileo (§§1, 4) in 1633, that René Descartes' mechanicist conception of matter was condemned, and that fear of censorship had a generally chilling effect on scientific theorizing throughout the seventeenth century (see Descartes, R. §11). However, it must be noted that not all of the Catholic officials were opposed to Galileo. In addition, a number of the century's greatest scientists were Catholic: Pierre Gassendi (§1), Marin Mersenne (§§1, 2), Blaise Pascal (§§3, 6) and Nicolas Steno, as well as Galileo and Descartes. The Jesuit order was home to a number of scientists who were not outstanding theorists but contributed

significantly to experimental science.

In the early modern period, it is difficult to distinguish conflicts between science and religion on the one hand from intra-theological conflicts and conflicts between the new science and the Aristotelian scholastic synthesis on the other. The Galileo affair needs to be interpreted in the light of both these complications, since it is not possible to understand the resistance to Galileo's astronomy without recognizing the fact that it called into question an entire socio-political order founded on a picture of the cosmos and of the place of humans in it. The affair was also an internal church struggle concerning the proper interpretation of Scripture. Galileo followed Augustine's rule that an interpretation of Scripture should be revised when it is found to conflict with other knowledge. This put him in conflict with conservative church officials who adopted a more literalist interpretive strategy. A further complication is the fact that the new science was often liberally mixed with magic and astrology, which the Catholic Church condemned both because they dabbled with the demonic and because of suspicion that they confirmed Calvinist views of determinism against the Catholic view of free will.

Robert Merton (1938) argued that Puritanism promoted the scientific revolution, a thesis still debated over half a century later. While Merton's thesis was overstated, it is likely that a particular Reformed doctrine of the sovereignty of God - that God's sovereignty excludes all active contributions of lesser beings to his work - made the modern scientific and philosophical conception of matter as inert or passive more acceptable to Isaac Newton, Robert Boyle (§4) and other Protestants than it would otherwise have been. Here again it is important to recognize the interplay of Aristotelianism and intra-theological disputes. The mechanicist conception of matter was a direct rejection both of Neoplatonic magical conceptions and of the Aristotelian teleological and organicist view, that 'forms' inherent in substances provided built-in powers and goals. At the same time, it furthered theological convictions first expressed by late medieval nominalist theologians (see Nominalism §§1-2). It is one of the great ironies of history, then, that Newton's mechanicist conception of the material universe so quickly evolved into Pierre Simon de Laplace's purely materialist and determinist view, the latter being absolutely incompatible with religion.

3 Indirect relations

If direct conflicts between Christian theology and the various theories of modern science have often been overemphasized, the deleterious effects on theology of indirect conflicts between religion and science have received too little attention. These indirect interactions can be considered under the headings of metaphysics and epistemology.

Metaphysics. Descartes' mechanicist view of matter as pure extension, accompanied by a view of mind as 'thinking substance', inaugurated a metaphysical dualism that has replaced older and more nuanced views of Christian anthropology. In so far as this dualism has been shown to be philosophically untenable, Christianity, with its view of the soul and afterlife, has appeared untenable as well (see Dualism).

The clockwork image of the universe as a closed system of particles in motion, strictly governed by the laws of physics (the image epitomized in the nineteenth century by the work of Laplace), created insuperable problems in accounting for divine action. A popular variety of deism offered the most reasonable account: God was the creator of the universe, and responsible for the laws of nature, but has no ongoing interaction with the natural world or with human history (see Deism §1). The alternatives for theists were accounts of miraculous interventions or an account of God as an immanent sustainer of natural processes (see Miracles §§1-2). The former seemed to make God irrational (contradicting God's own decrees) or inept (needing to readjust the system). The latter view made it difficult to maintain any more sense of God's personal involvement in human life than was possible for the deists. Much of the difference between liberal and conservative Christianity can be traced to theories of divine action: conservatives tend to take an interventionist, liberals an immanentist, view.

Epistemology. Medieval theologians had two sets of epistemological categories at their disposal, those relating to *scientia* (demonstrative or scientific knowledge) and those relating to *opinio* ('probable' beliefs, including those based on authority). So those theological conclusions that could not be deduced from first principles could, happily, be based on unimpeachable authority, the very word of God. However, in the modern period, the range of *scientia* contracted to the spheres of mathematics and formal logic; Hume (§6) and Kant (§8) both provided powerful critiques of deductive arguments for the existence of God and of natural theology generally. Furthermore,

when probable knowledge took on its contemporary sense of knowledge based on the weight of empirical evidence, appeals to authority became irrelevant, and most judged it impossible to provide empirical evidence for theological claims. Thus the central question for modern liberal theologians has been how, if at all, theology is possible.

Liberal theology diverged from more traditional accounts as a result of its strategies for meeting the problems raised directly or indirectly by science. Following Friedrich Schleiermacher (§7), many liberal theologians have understood religion to constitute its own sphere of experience, unrelated to that of scientific knowledge. Theological doctrines are expressions of religious awareness, not accounts of a supernatural realm. God works immanently, not by interventions in either the natural world or human history. Thus liberal theology has avoided direct conflict with modern science, at the cost (or with the beneficial consequence) of a radical revision of the very concepts of religion and theology. However, Ian Barbour's *Issues in Science and Religion* (1966) presented an encyclopedic overview of the points at which scientific claims *are* relevant to religious thought, and in *Myths, Models, and Paradigms* (1974) he argued for significant epistemological similarities between science and religion. Since then, a growing number of scholars from the liberal wing of Christianity have begun to call the modern division of territory into question.

4 Geology, evolution and the age of the earth

Physics and astronomy were the main scientific foci for theologians in the seventeenth and eighteenth centuries; geology and biology held an analogous place in the nineteenth and twentieth. For centuries, the biblical narrative from creation to Christ and the projected Last Judgment provided the skeletal outline for accounts of natural as well as human history. For instance, the story of Noah and the Flood served as a useful explanation for marine fossils found high above sea level. However, by the seventeenth century, the short span of history calculated from the Bible was being challenged from a number of directions. (James Ussher, a seventeenth-century Irish archbishop, has been credited with the calculation that creation took place a mere 4004 years before Christ.) Although sporadic attempts to reconcile geological history with Genesis continue up to the present, in the eighteenth century a large number of geologists already recognized that the Flood hypothesis could not explain the growing body of knowledge regarding rock stratification and the placement of fossils. A much longer history of the Earth, prior to human history, had to be presumed. At the same time, Egyptian and Chinese records were calling into question the short span of human history calculated from the Bible.

While some contemporary opposition to evolutionary theory involved 'young earth' chronology, negative reactions in the nineteenth century to Charles Darwin's *The Origin of Species* (1859) were more often objections to social Darwinism and to the claim that humans were kin to the 'lower animals'; other negative reactions focused on the fact that natural selection provided an alternative to divine design for explaining the fit of organisms to their environments, thus undermining an important apologetic argument. Nonetheless, many theologians and other believers readily accepted the theory and judged the changes it required in theology to be salutary rather than mere accommodation (see Darwin, C. §§2-4; Evolution, theory of).

5 Biological sciences

The theory of evolution is a surprisingly hot issue again at the end of the twentieth century. A Gallup poll published in the magazine *US News and World Report* (December 23, 1991) reported that a majority of North American Christians are sceptical of the macroevolutionary paradigm. The best explanation for this resistance is probably the fact that the issue has come to be framed in terms of creation *versus* chance as an account of the origin of the human species. That the issues can be formulated in these terms is due in part to a (defective) theory of divine action that contrasts God's creative acts with natural processes, rather than allowing that God may work through natural processes, including those that involve random events. The controversy is exacerbated by the use made of evolutionary biology by proselytizing atheists.

Genetics provides a new area for dialogue between religion and the biological sciences. Studies showing a genetic basis for human characteristics and behaviour raise questions about the status of the human person - for example, questions about free will and determinism - that have been the province of philosophy and religion. Of particular interest are studies of twins suggesting a genetic factor in religious behaviour (for example, Eaves *et al*. 1990).

Genetic research in general and genetic engineering in particular have raised a number of ethical questions that relate to theological ethics. For example, while most people favour genetic treatment for illnesses, many are opposed to germ-line intervention, which would affect all succeeding generations. Some objections are based on quasi-religious positions: scientists should not ‘play God’. This line of thinking calls for theological scrutiny: are not human beings themselves created in order to participate in God’s ongoing creative process? It is noteworthy that in 1991, the US National Institutes of Health awarded its first grant ever to a theological institution to the Center for Theology and the Natural Sciences (Berkeley, California) to study the theological and ethical implications of the Human Genome Initiative, the project to map human DNA (see Genetics and ethics; Religion and morality §4).

6 Cosmology

Physical cosmology is the branch of science that studies the universe as a whole. Beginning in the 1920s, developments in this field have sparked lively debate at the interface between theology and science. The Big-Bang theory, based on the expansion of the universe and a variety of other data, postulates that the universe originated in an extremely dense, extremely hot ‘singularity’ some 15 to 20 billion years ago (see Cosmology §3). Many Christians, including Pope Pius XII, greeted this theory as a confirmation of the biblical doctrine of creation. It was not only religious people who saw it as such; Frederick Hoyle defended a steady-state model of the universe, in which hydrogen atoms come into being throughout an infinite time span, partly because he saw it as more compatible with his atheism.

The discussion among theologians on the relevance of Big-Bang cosmology to the doctrine of creation involves controversy over the very nature of theology. As mentioned above, it has been common among liberal theologians since Schleiermacher to claim that religious meaning is entirely independent of scientific fact. Theologians who hold this position claim that the doctrine of creation, having to do only with the relation of all that exists to God, says nothing about the temporal origin of the universe, and is therefore equally compatible with any cosmological model.

A more recent area of research that has occasioned theological speculation can be referred to as the issue of the anthropic principle. A number of factors in the early universe had to be adjusted in a remarkably precise way to produce the universe we have. These factors include the mass of the universe, the strengths of the four basic forces (electromagnetism, gravitation, and the strong and weak nuclear forces), and others. Calculations show that if any of these numbers had deviated even slightly from its actual value, the universe would have evolved in a radically different manner, making life as we know it - and probably life of any sort - impossible. An example of the ‘fine-tuning’ required is that if the ratio of the strength of electromagnetism to gravity had varied by as much as one part in 10^{40}, there would be no stars like our sun.

Many claim that this apparent fine-tuning of the universe for life calls for explanation. To some, it appears to provide grounds for a new design argument (see God, arguments for the existence of §5). Others believe that it can be explained in scientific terms - for example, by suggesting that there are vastly many universes, either contemporaneous with our own or in succession, each of which instantiates a different set of fundamental constants. One or more of these universes would be expected to support life, and it is only in such a universe that observers would be present to raise the question of fine-tuning. Whether or not the fine-tuning is taken as evidence for the existence of God, it has important consequences for theology in that some philosophers believe that it argues against an interventionist account of continuing creation and divine action, since the prerequisites for human existence were built into the universe from the very beginning.

7 Physics and metaphysics

A variety of developments in physics since the end of the nineteenth century have called into question the determinist worldview. Quantum physics has introduced indeterminacy into the worldview of physics. Quantum theory generally allows only for probabilistic predictions regarding classes of events, not for prediction of individual events. It is unclear whether this limitation represents only a limit of human knowledge, or whether it signifies genuine indeterminacy in nature (see Quantum mechanics, interpretation of §3). However, scholarly opinion tends towards the latter view. Thus, most physicists reject the determinism of the Newtonian worldview, at least at this level. ‘Quantum non-locality’ refers to the peculiar fact that electrons and other sub-atomic entities that

have once interacted continue to behave in coordinated ways, even when they are too far apart for any known causal interaction in the time available. This phenomenon calls radically into question the Newtonian picture of the universe as discrete particles in motion, interacting by means of familiar physical forces. If Newtonian determinism had strong implications for theories of divine action, it is surely the case that these developments in quantum physics must have theological implications as well. What these implications are is still very much an open question.

A more recent development, which cuts across physics and the other natural sciences, is chaos theory (see Chaos theory). This is the study of systems whose behaviour is highly sensitive to changes in initial conditions. What this means can be illustrated with an example from classical dynamics: the movements of a billiard ball are governed in a straightforward way by Newton’s laws, but very slight differences in the angle of impact of the cue stick have greatly magnified effects after several collisions; moreover, initial differences that make for large differences in later behaviour are too small to measure, so the system is intrinsically unpredictable. Chaotic systems are found throughout nature - in thermodynamic systems far from equilibrium, in weather patterns and even in animal populations. Chaos theory is relevant to discussions of divine action not because chaotic systems are indeterminate (that is, not causally determined) and thus open to divine action without violation of laws of nature. Rather, the recognition of the ubiquity of chaotic systems shows the intrinsic limitations of human knowledge, and leads to the negative but important conclusion that one is rarely (or never) in a position to know that God is not acting in natural processes.

Another development throughout science with important implications for the issue of determinism and divine action is the recognition of ‘top-down causation’. The sciences can be conceived as a hierarchy in which higher sciences study progressively more complex systems: physics studies the smallest, simplest components of the universe; chemistry studies complex organizations of physical particles (atoms and molecules); biochemistry studies the extremely complex chemical compounds making up living organisms, and so on. The dream of the logical positivists was to provide an account of the sciences wherein the laws of the higher-level sciences could all be reduced to the laws of physics (see Logical positivism). This concept of explanatory reductionism followed naturally from the ontological reductionism that has become an important tenet of the modern scientific worldview: if all entities and systems are ultimately made up of the entities studied by physics, their behaviour ought to be understandable in terms of the laws of physics. So ontological and explanatory reductionism entail causal reductionism, or ‘bottom-up causation’. If the laws of physics are deterministic, we have a deterministic account of the whole of nature. However, it has become apparent that the behaviour of entities at various levels of the hierarchy of complexity cannot always be understood entirely in terms of the behaviour of their parts; attention to their interaction with non-reducible features of their environments is also required. Thus, the state or behaviour of a higher-level system exercises top-down causal influence on its components.

Arthur Peacocke (1990) has used this development in scientific thought to propose new directions for understanding divine action. In his ‘panentheist’ view, the universe is ‘in’ God, and God’s influence on the cosmos can then be understood by analogy with top-down causation throughout the hierarchy of natural levels (see God, concepts of §8). While this proposal does not answer questions about how God affects specific events within the cosmos, it does dissolve the long-standing problem of causal determinism.

8 Epistemology and language

The shift from medieval epistemology to modern empiricism required radical revision of religious epistemology. Various strategies were employed during the modern period to show theology to be epistemologically respectable. However, the increasing prevalence of atheism in scholarly circles suggests that these strategies have not been successful. At a point in intellectual history that some would call the end of the modern period, theories of knowledge have changed enough that the question of the epistemic status of theology needs to be examined afresh. Our concern here will be only with changes relating directly to science.

Theologians’ statements have sometimes been dismissed on the grounds that they describe states of affairs that are unimaginable or non-picturable. However, quantum theory and other recent scientific developments describe a physical reality that is equally unimaginable and, some would say, calls into question traditional two-valued logic. This line of argument is intended to point out that a view of knowledge more humble than that of the modern period is called for; reality is more complex and mysterious than anything our language and concepts allow us to

capture.

It has often been said (especially by theologians) that theology differs radically from science in that science is objective while all religious knowledge is self-involving, the product of an interaction between God and the human subject. Another way in which science has tempered older views of knowledge, and narrowed the difference between science and theology, is in its recognition that scientific knowledge itself is interactive. Measurements are interactions with the phenomena being measured, especially at the subatomic level.

Most modern thinkers have judged it impossible to provide empirical support for theology. However, beginning with the work of Ian Barbour (1974), there has been an investigation of the ways in which theological reasoning resembles that of science, including accounts of suitable data for theology. This development was made possible by advances in philosophy of science that show science itself to be a more complicated, and more human, enterprise than the positivists assumed (see Kuhn, T.; Lakatos, I. §3).

9 Religion's implications for science

Most of this entry has focused on the implications of science for religion. However, it is also the case that religion has implications for science. It has been argued that Christian doctrine was an important contributor to the rise of modern science: God's freedom entailed that features of the natural world could not be deduced a priori from rational principles, yet God's goodness and faithfulness suggested that the world would not be so chaotic as to be unintelligible. The very existence of religion is a valuable reminder that there are boundaries beyond which scientific explanation cannot go, and its doctrines help to answer questions that lie beyond those boundaries. The Newtonian era saw the separation of natural philosophy (science) from natural theology, and since then it has been a methodological presupposition of science that it should provide purely natural explanations. Science has thereby set boundaries on its own competence, but this does not mean that what is beyond its competence is therefore unimportant (or non-existent). Cosmology and physics raise questions they cannot answer: Why is the behaviour of natural processes law-like? What caused the Big Bang? Why is there a universe at all? While theology and science may interact in minor ways within each of their proper domains, it is here that theological explanation comes into its own.

See also: Creation and conservation, religious doctrine of §6; Personalism §2; Religion and epistemology; Science in Islamic philosophy; Wittgenstein, L.J. §15

NANCEY MURPHY

References and further reading

Barbour, I. (1966) *Issues in Science and Religion*, New York: Harper & Row.(Very accessible survey of relations between science and Christianity, beginning in the modern period.)

Barbour, I. (1974) *Myths, Models and Paradigms*, New York: Harper & Row. (Methodological comparison of science and religion, making use especially of Thomas Kuhn's philosophy of science. Very accessible.)

Barbour, I. (1990) *Religion in an Age of Science: The Gifford Lectures, 1989-1991*, San Francisco, CA: HarperCollins, vol. 1.(Combines and updates material from previous books. A particularly good introduction.)

Darwin, C. (1859) *The Origin of Species*, London: John Murray.(Classic statement of Darwin's thesis.)

Draper, J.W. (1874) *History of the Conflict between Religion and Science*, New York: D. Appleton.(A once-popular denunciation of Catholicism for its interference with scientific development.)

Eaves, L.J., Martin, N.G. and Heath, A.C. (1990) 'Religious Affiliation in Twins and Their Parents: Testing a Model of Cultural Inheritance', *Behaviour Genetics* 20 (1): 1-21.(Provides some evidence for a genetic component in religious behaviour.)

Hefner, P. (1993) *The Human Factor: Evolution, Culture, and Religion*, Minneapolis, MN: Fortress Press. (Example of theological use of the theory of evolution.)

Leslie, J. (1989) *Universes*, London: Routledge.(An accessible account of the anthropic or fine-tuning issue.)

Lindberg, D.C. (1992) *The Beginnings of Western Science: The European Scientific Tradition in Philosophical, Religious, and Institutional Context, 600 BC to AD 1450*, Chicago, IL: University of Chicago Press. (Accessible.)

Lindberg, D.C. and Numbers, R.L. (eds) (1986) *God and Nature: Historical Essays on the Encounter between Christianity and Science*, Berkeley, CA: University of California Press.(Criticizes accounts of the history of science and religion that presuppose the warfare model.)

Merton, R. (1938) *Science, Technology and Society in Seventeenth-Century England*, New York: Harper & Row, repr. 1970.(Presents the thesis that the development of science was encouraged by Puritanism.)

Murphy, N. (1990) *Theology in the Age of Scientific Reasoning*, Ithaca, NY: Cornell University Press.(On the relation between theological method and philosophy of science. Presupposes some knowledge of philosophy.)

Murphy, N. and Ellis, G.F.R. (1996) *On the Moral Nature of the Universe: Theology, Cosmology, and Ethics*, Minneapolis, MN: Fortress Press.(Comprehensive model for relating natural and social sciences to theology and ethics. Moderate technicality.)

Peacocke, A.R. (1979) *Creation and the World of Science: The Bampton Lectures, 1978*, Oxford: Clarendon Press.(Thorough survey of issues in the relation of contemporary science to Christian theology. Expands on issues of evolution and creation, and the hierarchical ordering of the sciences. Less accessible than Barbour.)

Peacocke, A.R. (1990) *Theology for a Scientific Age*, Oxford: Blackwell; enlarged edn, Minneapolis, MN: Fortress Press, 1993.(Relates top-down causation to divine action.)

Rolston, H. (1987) *Science and Religion: A Critical Survey*, New York: Random House.(A good introduction to the field; less readable than Barbour, but aesthetically pleasing.)

Russell, R.J., Murphy, N. and Peacocke, A. (eds) (1994) *Chaos and Complexity: Scientific Perspectives on Divine Action*, Vatican City State: Vatican Observatory; distributed by University of Notre Dame Press.(A series of articles on divine action, of various levels of technicality.)

White, A.D. (1896) *A History of the Warfare of Science with Theology in Christendom*, New York: D. Appleton, 2 vols.(Referred to in introduction and §2.)

Religion, critique of

During the Enlightenment a new philosophy of religion arose, one which was not connected with metaphysics or philosophical theology. It asked to what extent religion could be legitimated philosophically, to what extent it could be shown to be reasonable.

*The reasonableness of religion was taken to be significant for the political as well as the confessional clash between Christian denominations in Europe, all of which tried to justify their conflicting religious doctrines by reference to a supernatural revelation. The philosophical debate that began in the Enlightenment with regard to the criteria and arguments for a religion connected either to human nature or to public reason can be called a 'critique of religion' (*Religionskritik*), although the expression is not common before the critical philosophy of Kant and his school. Hegel followed the programme of Kant's philosophy, maintaining a philosophical concept of religion as falling 'within the limits of reason alone'. The radical left-wing school of Hegelianism transformed Hegel's approach, which was a critical legitimation of religion, into its destruction. Presupposing materialism in ontology and atheism, Feuerbach held that religion should be interpreted as a kind of anthropology. Marx claimed that religion is an expression of a certain sort of ideology and a necessary illusion within a class-structured society.*

In twentieth-century philosophy, the critique of religion can be found in two positions. The first is a rational reconstruction of the practical intentions or semantic content of religious belief; the second is a continuation of the interpretation of religion as ideological or illusory. In addition, we can identify certain other varieties of the critique of religion, including the theological critique of religion (found, for example, in the work of Barth and Bonhoeffer) and the philosophical critique of particular religious traditions (found, for example, in the romantic and postmodern rejections of Christian monotheism by Hölderlin, Nietzsche, Klages and Heidegger).

1 The concept of the critique of religion

In the history of Western philosophy, the specific contents of mythologies and religions as well as the forms of religious representations and practices have always been the subject of philosophical reflection. But only at the time of the European Enlightenment do we find the term 'critique of religion' (*Religionskritik*). This may indicate that the philosophy of the seventeenth and eighteenth centuries discovered the problem of a general philosophical theory of religion as such beyond the earlier metaphysical theories of a first cause or highest good. The immediate cause of the Enlightenment reflection was concern regarding the conflicting claims of reason and faith, of rational human insight and of divine revelation. While Pierre Bayle maintained the fundamental incompatibility of reason and religion, the rationalist Leibniz (§3) claimed a conformity of faith with reason. But in so far as the religious confessions had to refer to historical events or traditions, Leibniz's belief in such a conformity lost its plausibility. The search for a natural religion as a common property of all human beings sought to overcome the uncertainties of the particular confessions, their conflicting claims and their recourse to revelation. Belief in a 'natural religion' by thinkers such as Reimarus, Lessing and Wolff (§6) was clearly a criticism of the unrestrained claims of the historical religions based on revelation. None the less, it was the philosophy of Kant that first introduced the concept of a critique of religion into the philosophical debate. His special critique of religious belief is part of the general critical task he defined as essential for philosophy. In the preface to the first edition of his *Critique of Pure Reason*, Kant describes the duty of philosophy as 'a call to reason to undertake anew the most difficult of all its tasks, namely that of self-knowledge, and to institute a tribunal which will assure to reason its lawful claims, and dismiss all groundless pretensions' (1781: A11).

According to that account of the task of philosophy, all historical forms of religion have to be examined in the light of a general and public reason. Kant presupposes a concept of reason which is able to develop the criteriology to judge by its own competence. This critical approach is clearly distinct from the earlier positions. In defining the essence of religion as the recognition of the duties of human beings discovered by pure practical reason 'as divine commands', Kant is able to escape from the conceptual choice offered by deism, between a natural religion and a historical and revealed one. For Kant, it is not the origin of a religion but its 'qualification to be universal' which defines religious legitimacy. As he explains in *Religion within the Limits of Reason Alone*:

> Such a religion can be natural, and at the same time revealed, when it is so constituted that men could and

ought to have discovered it of themselves merely through the use of their reason, although they would not have come upon it so nearly, or over so wide an area, as is required.

(1793/1794: A219/B233)

2 Critique of religion after Hegel

In Hegel's idealistic philosophical system, religion and philosophy have the same subject, namely God or the Absolute Idea. One difference between religion and philosophy has to do with how each is able to express or to comprehend that subject. While Hegel defines religion as a system of symbolization or a 'form of imagination', it lacks a specific stage of conceptual reflection and self-awareness. It is the task of philosophy to reflect the imaginative form of the religious representation of the Absolute and, in doing so, to overcome its limits. Therefore, Hegel concludes that philosophy can claim the higher insight (see Hegel, G.W.F. §8).

This idealistic critique of religion was the immediate cause for a series of severe controversies after Hegel's death (1831). Karl Marx's famous opening statement in his *Critique of Hegel's Philosophy of Law* (1844), that 'the critique of religion is the precondition for all further criticism', has to be seen as a reaction to Hegel's idealistic view of religion's function and limits and the identical content of philosophy and religion (see Marx, K. §§2-3). On the other hand, Kierkegaard's existentialist protest against Hegel must be understood as criticism of the devaluation of religion and its interpretation within the idealistic system of philosophy (see Kierkegaard, S.A. §2). Feuerbach's theory of the anthropological content of religion and his critique of religion as an illusory projection of human attributes onto God participates in the general rejection of Hegel's philosophy by the critical movement of the 'Hegelian left' (see Feuerbach, L.A.). One of the consequences of this programme to reduce the idealistic philosophy to its true - that is, social-historical - origin and its practical human content is found in Marx's work. Religion for Marx is no longer the symbolic expression of the Absolute Idea, but the 'sigh' of the oppressed creature, a necessary illusion in a class society and a protest against injustice and exploitation.

3 Critique of religion in the twentieth century

The philosophical critique of religion in the twentieth century reflects the broader interest and engagement of philosophers in the field of philosophy of religion. In addition to a continuation of a rational critique of religion in the tradition of the Enlightenment, Kant and Hegel (examples are the Neo-Kantian philosophers Paul Natorp, W. Windelband, Hermann Cohen (§3) and Ernst Cassirer, and the philosophers of religion Paul Tillich and R. Schaeffler), and a tradition of a more or less Marxist critique (examples are R. Garaudy and Ernst Bloch), we can identify various new types of philosophy of religion. These often include new categories of a critique of religion, with both affirmative and disapproving statements of religion. In the tradition of psychoanalysis, there are the works of Freud and Mitscherlich. In the context of historicism we find Ernst Troeltsch. In the philosophical school of phenomenology there are Rudolf Otto and Max Scheler, and in the analytical tradition there are Russell, Carnap, Wittgenstein, Ayer and Hare. In addition to these, we can recognize two further contributions to the critique of religion, especially in Germany. First, there is the dialectical theology of Karl Barth, in which religion is supposed to be a work of human beings which is opposed to the 'word of God', and the position of Dietrich Bonhoeffer, who favours a Christian belief without religious qualities (*religionsloses Christentum*); second, there is the (neo-)romantic proposal to resuscitate elements of ancient Greek mythology and a polytheistic view of the world in the tradition of Hölderlin, Nietzsche, Klages and Heidegger.

See also: Atheism; Enlightenment, Continental; Deism; Existentialist theology; Natural theology; Religion and epistemology

MATTHIAS LUTZ-BACHMANN

References and further reading

Bayle, P. (1695-7) *Dictionnaire historique et critique (Historical and Critical Dictionary)*, ed. J.C. Gottsched, Hildesheim: Olms, 1973-.(A very influential critique of prejudice and ignorance in the early European Enlightenment.)

Bloch, E. (1968) *Atheismus im Christentum (Atheism in Christianity)*, Frankfurt: Suhrkamp Verlag.(A Neo-Marxist critique of Christianity.)

Feuerbach, L. (1841) *Das Wesen des Christentums (The Essence of Christianity)*, Stuttgart: Reclam Verlag,

1971.(The classical critique of religion in the post-Hegelian era of German philosophy in the nineteenth century.)

Hegel, G.W.F. (1832) *Vorlesungen zur Philosophie der Religion (Lectures on the Philosophy of Religion)*, in *Werke in 20 Bänden*, vols 16-17, Frankfurt: Suhrkamp Verlag, 1971.(Contains Hegel's doctrine concerning the relation between reason and religious belief.)

Kant, I. (1781) *Critique of Pure Reason*, trans. N. Kemp Smith, New York: St Martin's Press, 1965.(The most important work of modern German philosophy, in which Kant intended to construct a critical metaphysics on the basis of a critique of human understanding.)

Kant, I. (1793) *Religion within the Limits of Reason Alone*, trans. T.M. Greene and H.H. Hudson, New York: Harper & Row, 1960.(This text contains the basic assumptions of Kant's doctrine of religion.)

Leibniz, G.W. (1710) 'Discours préliminaire de la conformité de la foi avec la raison' (Introductory Discourse on the Conformity of Faith with Reason), in C.J. Gerhardt (ed.) *Essais de Théodicée*, Hildesheim: Olms, 1961.(An important early-Enlightenment reflection on the possibility of a fundamental harmony between religious belief and philosophical reason.)

Lutz-Bachmann, M. (1977) *Kritische Theorie und Religion (Critical Theory and Religion)*, Würzburg: Echter Verlag.(A collection of articles on the critique of religion in Critical Theory.)

Marx, K. (1844) *Critique of Hegel's Philosophy of Law*, in *Collected Works of Marx and Engels*, vol. 3, London: Lawrence & Wishart, 1975-.(The introduction to this book contains Marx's critique of religion.)

Post, W. (1969) *Kritik der Religion bei K. Marx (Marx's Critique of Religion),* Munich: Kösel Verlag.(A brilliant dissertation on Marx's critique of religion.)

Reimarus, H.S. (1754) *Die vornehmsten Wahrheiten der natürlichen Religion (The Most Exalted Truths of Natural Religion)*, Hamburg.(An important document of the German enlightenment.)

Thielicke, H. (1957) *Offenbarung, Vernunft und Existenz, Studien zur Religionsphilosophie Lessings (Revelation, Reason and Existence, Studies in Lessing's Philosophy of Religion)*, Gütersloh: Gütersloher Verlagshaus Gerd Mohn.(A qualified summary of the philosophical thought of Lessing, the most influential German poet of the Enlightenment.)

Religion, history of philosophy of

The philosophy of religion comprises any philosophical discussion of questions arising from religion. This has primarily consisted in the clarification and critical evaluation of fundamental beliefs and concepts from one or another religious tradition. Major issues of concern in the philosophy of religion include arguments for and against the existence of God, problems about the attributes of God, the problem of evil, and the epistemology of religious belief.

Of arguments for the existence of God, the most prominent ones can be assigned to four types. First, cosmological arguments, which go back to Plato and Aristotle, explain the existence of the universe by reference to a being on whom all else depends for its existence. Second, teleological arguments seek to explain adaptation in the world, for example, the way organisms have structures adapted to their needs, by positing an intelligent designer of the world. Third, ontological arguments, first introduced by Anselm, focus on the concept of a perfect being and argue that it is incoherent to deny that such a being exists. Finally, moral arguments maintain that objective moral statuses, distinctions or principles presuppose a divine being as the locus of their objectivity.

Discussions of the attributes of God have focused on omniscience and omnipotence. These raise various problems, for example, whether complete divine foreknowledge of human actions is compatible with human free will. Moreover, these attributes, together with God's perfect goodness give rise to the problem of evil. If God is all-powerful, all-knowing and perfectly good, how can there be wickedness, suffering and other undesirable states of affairs in the world? This problem has been repeatedly discussed from ancient times to the present.

The epistemology of religious belief has to do with the questions of what is the proper approach to the assessment of religious belief (for rationality, justification, or whatever) and with the carrying out of such assessments. Much of the discussion has turned on the contrast between the roles of human reason and God's revelation to us. A variety of views have been held on this. Many, such as Aquinas, have tried to forge a synthesis of the two; Kant and his followers have sought to ground religion solely on reason; others, most notably Kierkegaard, have held that the subjecting of religious belief to rational scrutiny is subversive of true religious faith. Recently, a group of 'Reformed epistemologists' (so-called because of the heavy influence of the Reformed theology of Calvin and his followers on their thinking) has attacked 'evidentialism' and has argued that religious beliefs can be rationally justified even if one has no reasons or evidence for them.

1 What is philosophy of religion?

The term 'philosophy of religion' is a relative newcomer, dating only from the late eighteenth century. It became widespread under the influence of Hegel, whose system of philosophy featured various 'philosophy of' components - history, mind, art, as well as religion. But philosophical reflection on religious issues is as old as philosophy itself. One of the earliest spurs to philosophy in ancient Greece was the emergence of doubts concerning the religious tradition. The present article construes 'philosophy of religion' as the philosophical discussion of questions growing out of religion, whether conducted under that title or not.

These discussions are highly diverse. Different religions give rise to different issues, depending, for example, on whether the religion is centred around faith in a personal deity. The philosophizing that is done will reflect the thinker's conception of the philosophical task. The philosophy of religion of a speculative metaphysician such as Hegel will differ widely from that produced by an analytic philosopher who thinks of philosophy as concerned primarily with the clarification of concepts. And even when these two variables are assigned specific values, there are different views as to how philosophy can best contribute to the understanding of religion. The main concern has been the clarification and critical evaluation of fundamental beliefs and concepts from some particular religious tradition. But some philosophers in the last two centuries, influenced by Hegel and by phenomenology, have set out to produce a synoptic view of religion in all its aspects, not only the doctrinal and conceptual ones, and to survey and interrelate the different forms in which it is realized (see van der Leeuw 1963).

Faced with this bewildering diversity and long history, an article of this length can provide only a brief sample. It will therefore concentrate on the attempts of Western philosophy to deal with philosophical questions arising from the Judaeo-Christian tradition, and, among these attempts, on what can more specifically be called *philosophical*

theology. This involves the attempt to deal with fundamental religious concepts and beliefs in a philosophical way, where ‘philosophical’ implies two restrictions. First, when our aim is philosophical, we are concerned with *understanding* (clarifying, explicating) the material, and critically evaluating it for truth, coherence and rationality, rather than with *describing* it or discovering laws that govern it. (This distinguishes philosophy of religion from the history, sociology or psychology of religion.) Second, the enterprise is not conducted from the standpoint of any religious commitments. It appeals only to what is (claimed to be) available to any rational person who reflects carefully on the matter. (This distinguishes philosophical theology from dogmatic theology.)

Even with this restriction there is an enormous choice of topics on which to concentrate. Here is a list of those that most frequently appear in current treatises and textbooks on the subject:

(1) arguments for and against the existence of God;
(2) problems concerning the various attributes of God;
(3) the problem of evil;
(4) religious issues concerning human nature, particularly the possibility of human life after death, and free will;
(5) the bearing of religion on ethics;
(6) how to understand basic religious concepts, such as faith, salvation, creation and spirituality;
(7) questions about divine action, particularly the notion that God ‘intervenes’ in the world, bringing about events that would not have occurred had only purely natural causes been involved (miracles);
(8) the epistemology of religious belief.

The present entry deals with the most significant elements in the histories of topics (1), (2), (3) and (8); this choice is prompted by the main areas of interest among contemporary philosophers of religion.

2 Arguments for the existence of God: the cosmological argument

Philosophical theology goes back to the ancient Greeks, to a time at which intellectual dissatisfaction with the polytheistic religious tradition led to attempts by philosophers to develop a more rational religious orientation. Plato, in the *Laws*, may be said to have inaugurated the project of giving philosophical arguments for the existence of God. Interestingly enough, it occurs in a political context. Arguing that the denial of certain theological truths is a grave offence against the social order, he proceeds to demonstrate those truths, beginning with the existence of God. He presents a form of the cosmological argument that turns on the cause of motion. Briefly put, the argument is that motion produced by something else presupposes a self-mover, a being that spontaneously moves. This is identified by Plato with soul (*psychē*). He does not seek to show that all change can be traced to a single self-changing source. On the contrary, he holds that there must be at least two, though he takes the ‘best soul’ to be supreme.

A more elaborate argument from motion is found in Book XII of Aristotle’s *Metaphysics* (see Aristotle §16). Time is necessarily everlasting, for if it came into being, that would presuppose a time before which there was time; and if it were to cease there would be a time after that, a time with no time. Aristotle also thinks that there cannot be time without motion. Therefore motion is everlasting. And since motion presupposes substances that move, it is necessary that there are always substances in motion. But if all substances were destructible (had a potentiality for nonexistence), it would be possible that at some time there were no substances. Hence there must exist at least one substance that is pure actuality, with no unrealized potentialities, and it is this that guarantees the eternality of moving substances. Since it is pure actuality, it does not move, and since all matter involves potentiality it is immaterial. As the source of all motion, it is the ‘first mover unmoved’. Since the first mover is immune to any kind of change, it can be the source of motion only as what Aristotle calls a ‘final cause’, the end or goal towards which a process tends. It moves the heavens (which in turn move all else) as an object of desire or love. As unmoved, it does not enter into causal interaction with the world, as the Judaeo-Christian God does, but confines itself to pure thought. It is ‘thought thinking itself’.

Neither Plato nor Aristotle thought of God as creating the physical universe, but Plato’s God is somewhat more active than Aristotle’s. In the *Timaeus*, Plato depicts God as a ‘Demiurge’ who introduces order into the chaos with which he is initially confronted. But it was left to Christian theology to develop the radical idea of the divine creation of everything other than God *ex nihilo*, from nothing.

Neoplatonism, as developed most fully by Plotinus in the *Enneads*, involved a more mystical view, in which the supreme reality is an absolute One, devoid of any distinctions whatever, and accessible only to a unitive mystical experience in which even the distinction between subject and object is obliterated. Neoplatonism profoundly influenced Christian theology, primarily through the thinker who wrote mystical treatises around AD 500 under the name of 'Dionysius' (see Pseudo-Dionysius). His influence is evident even in thinkers who were not primarily mystical theologians, including Anselm and Aquinas, both of whom held the doctrine of the absolute simplicity of God.

Neoplatonism was the main philosophical basis of Christian theology from Augustine to the twelfth century, when Aristotle's works were rediscovered in western Europe. By the latter half of the thirteenth century the philosophical thought of the West was thoroughly Aristotelian. As for arguments for the existence of God, the 'five ways' of Thomas Aquinas, in his *Summa theologiae* Ia, q.2, came to dominate Roman Catholic philosophical theology up to our own day. Three of the five ways can be viewed as cosmological. There is an argument, reminiscent of Aristotle, to a first mover unmoved, a parallel argument to a first cause (of existence) uncaused, and an argument, again reminiscent of Aristotle, that is designed to show that the universe cannot consist merely of contingent beings (beings that may or may not exist), but must contain a necessary being, one that exists by the necessity of its own nature. But though these arguments are heavily, and admittedly, indebted to Aristotle, Aquinas draws from them conclusions much more congenial to Christian theology. Like Aristotle, he argues that the first cause uncaused must be pure actuality, since it is prior to all other beings and actuality is prior to potentiality. But he also maintains that the first cause must contain all the perfections of those things that depend on it for their existence, since causality consists in actualizing potentialities, and an agent can bestow only those actualities it already possesses. That is the thin end of the wedge Aquinas needs to show that the first cause is absolutely perfect, and from that he proceeds to derive the basic attributes of the Christian God - omniscience, omnipotence, perfect goodness, love, and so on.

Aquinas' first mover and first cause arguments involve the thesis that an infinite series of causes or movers is impossible; hence any such series must stem from a first member, on which all succeeding members depend for their efficacy. It is natural to read Aquinas as denying that a *temporally* infinite series of causes is possible, in which case his conclusion would be that there is a temporally first cause of the universe, and hence that the universe came into existence some finite period of time ago. But since he elsewhere denies that it can be shown by reason that the world had a beginning, this cannot be his intention. His claim is rather that there cannot be an infinite hierarchy of levels of dependence. Such a series would begin, for example, with the motion of a pencil being dependent on the simultaneous motion of my hand, which in turn is dependent on my will, which in turn…. However, Aquinas' development of this way of thinking was in terms of medieval physics, which is far removed from anything taken seriously today. Hence there have been repeated attempts in the modern period to free the argument from entanglement with outmoded science. A notable example is Samuel Clarke (§1) (1738). A brilliant contemporary critical discussion of this argument is put forward by Rowe (1975). Avoiding the sometimes labyrinthine details of Clarke's presentation, the basic idea can be presented as follows:

(1) The physical universe is temporally either finite or infinite.
(2) If finite, the beginning of its existence must have been due to some cause.
(3) If infinite, there must be some cause, the activity of which is responsible for the fact that this universe exists rather than some other or none at all.
(4) Even if the cause of the existence of our universe owes its existence to some other cause, an infinite regress of such causes is impossible.
(5) Therefore, the entire system ultimately owes its existence to an uncaused first cause, one that exists by the necessity of its own nature (a necessary being).

In this version, the series of causes that is claimed to be necessarily non-infinite is wholly outside the spatiotemporal universe. And since the argument begins with the universe as a whole, it does not depend on any particular scientific view of the internal economy of that universe.

In this version, the argument's dependence on what has come to be called the Principle of Sufficient Reason is apparent. According to that principle, for any fact (state of affairs) there must be a sufficient reason for its obtaining, a reason either in itself or in some external cause. Much discussion of the argument has centred around

that principle, which received a historically important defence by Leibniz (§§3, 11).

3 Arguments for the existence of God: the teleological argument

Here the crucial claim is that a cosmic designer is required to explain the design or adaptation we find in the world, the way in which things are fitted to achieve certain ends. Aquinas' fifth way is an early example. He thought in terms of a teleological physics in which all physical processes were explained by an orientation to goal states. For example, each type of matter was thought of as seeking its proper place in the universe - solids the centre (identified with the centre of the earth), fire the periphery, and so on. He then argued that such goal direction is possible for inanimate beings only if they are directed to these goals by an intelligent being. But with the demise of teleological physics in the modern period the main examples of design are found in the organic realm, in the ways in which the structure and function of organisms are nicely adapted to the satisfaction of their needs. For example, the outer ears that land animals, but not fish, possess are fitted to focus sound waves that are weaker in air than in water. And the ears of predators are fitted to receive sounds from in front, while their intended victims have ears that pick up sounds better from the rear. A famous example of this argument is found in Paley's *Natural Theology* (1802), in which he imagines coming across a watch on the ground and becoming convinced, from the way in which its design is conducive to keeping the time, that it was produced by an intelligent designer. He then argues that we must say the same of the human eye (see Paley, W. §2).

The argument was subjected to a variety of criticisms in the *Dialogues Concerning Natural Religion* (1779) of David Hume (§6), many of which involve doubts that the universe as a whole is enough like a watch or other artefact to sustain the argument. In the last century and a half, the chief criticism of the argument has been that the theory of evolution provides a superior, naturalistic explanation for the cases of organic adaptation on which the argument is based. Organisms are well adapted to survive in the environments in which they find themselves because otherwise they would die out. It is the fittest that survive in the competition for scarce resources. In the face of such criticism, the argument has shifted in a way analogous to the shift noted above in the cosmological argument. Instead of starting with particular cases of adaptation, for which scientific explanations might be advanced, we start with the universe as a whole and ask why it is so constituted as to be fitted for life, however particular forms of life are to be explained in terms of factors within the universe. To that question no science could possibly provide an answer, since science is concerned with the internal economy of the universe, how one part or aspect is related to another. With recent developments in the study of the origin and development of the physical universe, this more cosmic form of the argument has received a shot in the arm from considerations concerning the 'fine tuning' that is required of basic physical constants if the universe is to be supportive of life (see Leslie 1989) (see Religion and science §6). The argument has also been extended to include a larger variety of features said to require a designer, such as mathematical order and beauty (see Tennant 1930).

4 Arguments for the existence of God: the ontological argument

This is the final member of the classic trio of arguments. It was first stated clearly by Anselm of Canterbury (§4), the most important thinker of the eleventh century, in his *Proslogion* (1077-8). It differs from the other two in not seeking an explanation of empirical facts. It is purely a priori, consisting in the analysis of concepts. In Anselm's version, it is a *reductio ad absurdum*. Suppose, he says, that the being than which no greater can be conceived ('the perfect being', for short) exists only in the mind and not in reality. Then we could conceive a greater being, one just like the former except that it also exists in reality. But it is impossible that we should conceive a being greater than a being than which none greater can be conceived. Hence the supposition with which we began - that the perfect being does not exist in reality - must be rejected.

A monk named Gaunilo sent Anselm a criticism of the argument, in which one of the points he made was that there must be something wrong with it, since one could use an exactly parallel argument to show that a perfect island, or a perfect anything else, exists. Aquinas rejected the argument on the grounds that it wrongly assumes that we know the essence of God. However, the most persistent criticism, classically expressed by Immanuel Kant in his *Critique of Pure Reason* (1781), is that the argument falsely assumes that existence is a property, whereas we do not add anything to our concept of a being (to the list of its properties) when we point out that it exists.

The argument has had a number of supporters through the centuries, including Descartes (§6) and Leibniz (§3), as well as a number of critics. Leibniz realized that questions could be raised about the *possibility* of a perfect being,

and he undertook to establish that. A major step forward occurred in this century when Charles Hartshorne (1962) and Norman Malcolm (1963) independently discovered that there are two versions of the argument in the *Proslogion*. The first, the one presented above, is designed to show that a perfect being exists, the second that a perfect being necessarily exists. Both Hartshorne and Malcolm point out that the second version is immune to the charge of treating existence as a predicate. There are not the same objections to treating necessary existence as a predicate. More recently, Alvin Plantinga (1974) has made use of contemporary modal logic to develop the most sophisticated formulation of the second form of the argument.

5 Arguments for the existence of God: the moral argument

This argument takes its start from premises concerning moral obligation, moral principles, moral goodness, and the like. It is a more recent arrival than the three just discussed. It was first given prominence by Kant (§11) in his *Critique of Practical Reason* (1788), where we find the following argument: Moral endeavour presupposes that the virtuous will eventually attain happiness, for only in the union of virtue and happiness do we have the supreme good, and it is a fundamental moral obligation to strive to realize the good. But this can be an obligation only if it is possible to attain the supreme good. And this is possible only if the universe is under the control of a morally good personal being. Therefore the existence of God is a necessary presupposition of morality. Kant did not claim that this reasoning enables us to *know* that God exists; it only shows the existence of God to be a presupposition of the moral life.

Other thinkers in the nineteenth and twentieth centuries have presented more straightforward moral arguments. The most common form starts from the claim that moral distinctions - between right and wrong, between what we are and are not obliged to do - are objective. It is an objective fact that it is wrong wantonly to inflict suffering on another human being, and that one is obliged to care for one's children. But this can be so only if there is a God to 'anchor' these facts. According to the argument, there is no other possible objective locus for morality. In some versions it is the divine will - what God commands us to do and not to do - that renders moral facts objective. In others it is the goodness of God that plays this role. God *is* the good, and human morality constitutes the approximation to the divine goodness that is most appropriate for human beings (see Adams 1987).

There have been some interesting recent reformulations of traditional arguments for the existence of God. Chief among them is the restatement of cosmological and teleological arguments in terms of probability and confirmation theory by Swinburne (1979).

6 Divine attributes

As Christian theology developed, the nature of God was thought of in terms of various attributes, such as omnipotence, omniscience and perfect goodness. He was also thought of as a personal agent, one who has purposes and who acts to carry them out in the light of his knowledge. The attempt to think all this through gives rise to various difficulties with which philosophical theologians have been concerned since the early centuries of our era.

Omnipotence. It is not difficult to see that this attribute must be limited in some way if we are to avoid paradoxes and contradictions. For one thing, we have to avoid the idea that God could bring about contradictory states of affairs. If he could, then he could both save Jones and damn Jones, an inconceivable result. Aquinas put this by saying that God can do only what is possible absolutely. Since a contradiction is not a possible state of affairs, it is no limitation on God's power that he cannot bring it about. From the middle ages the 'paradox of the stone' has exercised philosophers. Can God create a stone so heavy he cannot lift it? If we say 'Yes', then there is something he cannot do, namely, lift such a stone. If we say 'No', there is still something he cannot do, namely, create such a stone (see Rosenkrantz and Hoffman 1980).

A more obviously religiously important problem has to do with God's ability to do wrong. If God is essentially perfectly good, it is impossible that he should act wrongly. But then this is something he cannot do, so how can he be omnipotent? Both Anselm and Aquinas answer this by maintaining that since to do wrong is to fall short of a perfect action, omnipotence implies the inability to do wrong rather than the ability to do it. But this looks like fudging. A better answer would be that since God's doing wrong is an inherently impossible state of affairs, there would be no lapse from omnipotence in not being able to realize it, any more than in not being able to realize any

other impossible state of affairs (see Pike 1969) (see Omnipotence).

Omniscience. The most widely discussed problem in this area concerns divine foreknowledge and its relation to human free will. If God is omniscient, he knows everything there is to know, including future human actions. But then how can any of those actions be free? If God knows now that I will accept a certain invitation tomorrow, how can it be that I have any choice in the matter? Surely I can't do anything contrary to what God knows; and so it is not within my power to do anything on this point except accept the invitation. I have no choice; I am not free. But if I never have a choice about what I do, how can I be responsible for my actions?

The problem is given classic expression in Book V of *The Consolation of Philosophy* (525-6) by Boethius (§5). His solution, which was accepted by Anselm, Aquinas, and many other medieval philosophers, was that God's being is not temporal. God does not live through a succession of moments but enjoys, in Boethius' memorable phrase, 'the complete possession all at once of illimitable life' (*The Consolation of Philosophy* V, pr. 6). Hence God does not know in advance what choices I will make, for God never occupies any temporal standpoint. As eternal rather than temporal, he is all at once simultaneous with every moment of time. Hence he knows what I do because, so to say, he 'sees' me doing it when I do it. Therefore there is no threat to human free will (see Eternity).

Though the doctrine of divine atemporality was widely held in the Middle Ages, as well as by many modern theologians, including Schleiermacher, the father of liberal Protestant theology, it has been severely criticized in this century as being unbiblical and as rendering incoherent any notion of divine action in the world (see Pike 1970). It has been vigorously defended recently by Stump and Kretzmann (1981), and by Leftow (1991). However, various attempts have been made to reconcile divine omniscience and human free will without denying divine temporality. One of the most influential is that of William of Ockham, whose seminal idea ran something like this: Even though God's knowing today that I will wash the windows tomorrow entails that I will wash the windows tomorrow, that does not mean that his foreknowledge renders it *necessary* that I will wash the windows tomorrow, and hence does not imply that I have no choice. It is all a question of what depends on what. The past is necessary in the sense that nothing we can do now can alter it. But God's knowledge of future contingent facts is not a 'pure' past fact. It depends for what it is on the future. God knows that I will wash the windows tomorrow because I will wash the windows tomorrow. That is, God's knowledge of what will happen tomorrow is partly a function of what will happen tomorrow. So that leaves me free to have a choice. It is because of the choice that I will make at that time that God's past state is one of knowledge (see Ockham's *Predestination, God's Foreknowledge, and Future Contingents*, q.1, M). A number of contemporary philosophers - for example Plantinga (1986) and Freddoso (1983) - have developed these Ockhamist ideas in new and fruitful ways. Various other solutions to the problem have been developed recently. Some have suggested a limitation of divine omniscience to exclude foreknowledge of free actions, on analogy with limitations on omniscience mentioned above (Swinburne 1977; Hasker 1989). Freddoso has revived Luis de Molina's doctrine of 'middle knowledge', knowledge of what free creatures would freely do if placed in a certain situation (see Molina, L. de §3; Omniscience). The problem of divine foreknowledge and human free will has been widely debated in contemporary philosophy of religion; a useful selection of articles can be found in the anthology edited by Fischer (1989).

7 The problem of evil

'Evil' as understood here ranges over anything undesirable, anything we would be better off without. It is not restricted, as it is in ordinary usage, to wrong actions, or wickedness. It also includes the suffering of sentient creatures, which is termed 'natural evil', in contrast to 'moral evil'. Evil is a serious, perhaps the most serious, problem for theistic religion just because it seems obvious at first sight that an omnipotent and perfectly good God would not allow any evil in his creation (see Evil, problem of). Being omnipotent, he could prevent it, and being perfectly good, he would do so. How then can there be evil if the world is governed by such a deity?

This problem has been with us since the beginning of philosophy in ancient Greece, and it has been extensively discussed in Christianity and other theistic traditions (see Evil). The most common line in Christian theology has been that moral evil is the result of wrong choices (sins) by free created agents (angels and humans), and that natural evil is either a divine punishment for sin or is involved in the corruption of nature that is consequent on angelic and human sin. Here, however, we are concerned with treatments of the problem that are philosophical rather than theological. The most prominent of these can be traced back at least as far as St Augustine (§9), and

they were taken up by many of his successors. Their exposition in Aquinas' *Summa theologiae* Ia, qq.48-9, will be utilized here. First, all evil is a *privation* of something a being would have naturally, like blindness, which is a lack of normal sight. (Only a naturally sighted being can be called 'blind'; a stone is not blind.) Thus evil is not a positive 'nature' or attribute. Everything positive is good, and since there can only be a privation where there is something with a positive (good) nature to be deprived, evil always presupposes good and is parasitic on it. This is not to say that evil is unreal or an illusion; it is only to say what kind of metaphysical status it has. And since evil is real, we are still faced with the question of why God brings it about, or at least permits it. Here Aquinas treats moral and natural evil separately. Wrongdoing is in no sense the work of God, though he permits it, because he has chosen to create free agents who have the power of choice between right and wrong, and sometimes choose the wrong. As for natural evil, that is the work of God, and his (allegedly sufficient) reason for including it is the Principle of Plenitude:

> The perfection of the universe requires that there should be inequality in this, so that every grade of goodness may be realized. Now, one grade of goodness is that of the good which cannot fail. Another grade of goodness is that of the good which can fail in goodness…. Now it is in this that evil consists, namely, in the fact a thing fails in goodness.
>
> (*Summa theologiae* Ia, q.48, a.2)

It is for the perfection of the universe as a whole that God brings it about that there are creatures subject to privation as well those that are not. This has felicitously been termed an 'aesthetic' approach to the problem, since it depicts God as a supreme artist, aiming at the most perfect composition of the whole universe.

These Augustinian themes have persisted in discussions of the problem up to the present, though they fail to convince many contemporary thinkers. One of the most famous post-medieval treatments is that of Leibniz (§3) in his *Theodicy* (1710). This is most famous for Leibniz's insistence that God created the best of all possible worlds, since otherwise he would not be perfectly good. (It is noteworthy that Aquinas (*Summa theologiae* Ia, q.25, a.6) denies that God's goodness has this implication. He argues that since whatever the state of creation, God could make it better by adding goods, there can be no best possible creation.) By holding there to be no alternative to regarding this as the best world possible, Leibniz raised the stakes even higher, making it more difficult to understand why there is evil in the world. He makes many suggestions as to how things that are undesirable in themselves can lead to goods that far outweigh them, but he makes no major advances in this beyond the medieval tradition.

A contemporary thinker, John Hick (1978), has unearthed and developed an important alternative to the Augustinian approach, one which he traces to the second-century Christian theologian Irenaeus. The point of natural evil in the divine plan is its usefulness in 'soul making', giving people the opportunity to develop morally and spiritually in dealing with the suffering and misery in their lives (see also Stump 1985).

Since evil arises in its acutest form only if God is regarded as omnipotent, it is not surprising that it has occurred to some to challenge that assumption. A relatively early example is Mill (1874). In this century, Charles Hartshorne (1941, 1984), following the lead of Alfred North Whitehead (§4), has forged a highly influential 'process theology', in which God is far and away the most powerful being, but is not omnipotent. Instead, God and the world are coexistent and equally necessary. What God does is to work for order, harmony and fulfilment in the world, but without an absolute ability to bring any of this about by a simple act of will. For anything undesirable in the world, we can say that God is doing the best he can to bring good from it (see Process theism).

8 The epistemology of religious belief

Though it is only recently that this topic has been a major preoccupation in philosophy of religion, it is not without historical roots. The question of the nature and status of religious faith has concerned theologians and philosophers at least since the early Christian centuries. Faith is a multifaceted attitude, involving affective and behavioural as well as cognitive components. Concentrating here on the belief component, what are the appropriate standards by which to assess religious belief for rationality, justification, knowledge, and the like? A major issue from the beginning has been the place of reason in all this. Must one have adequate reasons for a religious belief in order to hold it justifiably? Or is that the wrong way to look at the matter? Views on this issue range all the way from that of Tertullian (160-225), who is famous for the dictum concerning the Incarnation, 'I believe it because it is

absurd', thus hurling his opponents' reproach into their teeth, to Kant's rationalism, which expresses itself in the title of his major work in the field, *Religion Within the Bounds of Reason Alone* (1793). Most treatments fall somewhere between these extremes. On the one hand, when we look at the full range of the beliefs of an actual religion, there is general agreement that not all those beliefs can be established by unaided human reason. If we are to have any basis for them they must be disclosed to us by God, by divine revelation. In Christianity this is true of the Incarnation, as well as beliefs about divine purposes for, requirements on, and interaction with human beings. Though it is often held that God accompanied his major revelations with miracles to indicate their status, it is also widely held that these indications do not constitute a decisive proof of that status. Hence it has been thought that divine assistance is required for a person to have firm faith in divine revelation, that faith is, at least in part, a gift of God. On the other hand, there are basic beliefs that are often regarded as susceptible of rational proof - the existence and nature of God, and the immortality of the soul, for example. During the first Christian millennium no sharp distinction was made between these two groups of beliefs. With Augustine, for example, it is difficult to say whether he meant to give rational arguments for the doctrine of the Trinity, or only to explain its meaning. By the thirteenth century, however, we find Aquinas (§14) making a clear distinction between the preambles of faith that could be established by philosophical argument and the remainder that rested on divine testimony.

The Thomistic synthesis fell apart in the modern period under the impact of increasing doubts about both of Aquinas' divisions. David Hume (§2), in the chapter on miracles in his *Enquiry Concerning Human Understanding* (1748) argued that since in order to count as a miracle an event must be contrary to our general experience, and since that experience makes it extremely unlikely that such a thing should happen, a report of a miracle would need an enormous weight of testimony to outweigh that intrinsic improbability, a weight that no report of the miraculous ever displays. Thus, said Hume, we can never be justified in accepting a report of a miracle. As for faith as a gift of God, believing in the absence of sufficient reasons has often been deemed unworthy of a rational agent. This attitude was given classic expression in the late nineteenth century in W.K. Clifford's 'The Ethics of Belief', where he says that 'It is wrong always, everywhere, and for every one, to believe anything upon insufficient evidence' (1879: 183). Finally, with the rejection of the Aristotelian philosophy on which Aquinas' arguments for the existence and nature of God rest, those arguments lost their appeal. The enterprise of providing other arguments for the existence and nature of God was continued by such thinkers as Descartes, Leibniz, Locke (§§6-7) and Berkeley (§§3, 6-7), but by the latter part of the eighteenth century Kant was arguing that it is impossible that any such arguments could be productive of knowledge, though, as pointed out above, he held that we could exhibit the existence of God as a necessary presupposition of morality. Since Kant, although many thinkers have continued to propound arguments for the existence of God, the enterprise has not regained its former prestige.

Not everyone took the failings of natural theology to reflect badly on religion. The nineteenth-century Danish religious thinker Søren Kierkegaard (§§4-5) was a modern Tertullian who passionately argued that rational argument, so far from being required by faith, is incompatible with it. Faith requires a leap, a risk. Seeking the security of rational proof is a snare and a deception; it is a way of avoiding the trust that God requires of us. Though Kierkegaard has been enormously influential, especially in this century, other voices have also been heard. John Henry Newman, in his *An Essay in Aid of a Grammar of Assent* (1870), stressed the continuity of religious belief with other areas in which trained sensitive judgment is needed to assess the bearing of a complex set of diverse considerations. This approach has been continued in this century in the work of Basil Mitchell (1981).

The most prominent position on the current scene is 'Reformed epistemology', so called because its advocates are self-consciously working in the Reformed tradition stemming from John Calvin. A notable example is Alvin Plantinga (1983), who attacks 'evidentialism', the view that religious belief is irrational unless supported by adequate evidence (see Religion and epistemology §§1-3). In this he sounds like Kierkegaard, but the alternatives they suggest are quite different. Plantinga argues that religious beliefs can be 'properly basic', that is, rationally justified even though based on no reasons whatever. Thus he is no more irrationalist than the evidentialists he opposes. He simply has a different theory as to the conditions for rational religious belief.

See also: God, arguments for the existence of; God, concepts of

WILLIAM P. ALSTON

References and further reading

Adams, R.M. (1987) 'Moral Arguments for Theistic Belief', in *The Virtue of Faith*, New York: Oxford University Press.(A useful survey of moral arguments for the existence of God.)

Anselm of Canterbury (1077-8) *Proslogion*, trans. S.N. Deane, Chicago, IL: Open Court, 1939.(The classic source for the ontological argument.)

Aquinas, T. (1266-73) *Summa theologiae*, London: Eyre & Spottiswoode, 60 vols, 1964.(Question 2 of Part I contains the celebrated five ways of proving the existence of God, and questions 3-26 draw out the consequences of that for the divine nature.)

Aristotle (mid-4th century BC) *Metaphysics*, trans. W.D. Ross, in R. McKeon (ed.) *The Basic Works of Aristotle*, New York: Random House, 1941.(Book XII contains Aristotle's philosophical theology.)

Boethius, A.M.S. (525-6) *The Consolation of Philosophy*, Indianapolis, IN: Bobbs-Merrill, 1962.(A classic late ancient work, containing an influential discussion of divine foreknowledge.)

Clarke, S. (1738) *A Discourse Concerning the Being and Attributes of God, The Obligations of Natural Religion and the Truth and Certainty of the Christian Revelation*, London: John & Paul Knapton.(Classic presentation of the eighteenth-century version of the cosmological argument. Not easy going.)

Clifford, W.K. (1879) 'The Ethics of Belief', in *Lectures and Essays*, London: Macmillan.(Trenchant statement of 'evidentialism', the view that it is irrational to believe without adequate evidence.)

Fischer, J.M. (ed.) (1989) *God, Foreknowledge, and Freedom*, Stanford, CA: Stanford University Press.(A good selection of essays on the subject.)

Freddoso, A.J. (1983) 'Accidental Necessity and Logical Determinism', *Journal of Philosophy* 80 (5): 257-78. (Contemporary version of Ockhamism.)

Hartshorne, C. (1941) *Man's Vision of God*, Hamden, CT: Archon Books.(A primary source for 'process theology'.)

Hartshorne, C. (1962) *Anselm's Discovery: A Re-examination of the Ontological Proof for God's Existence*, La Salle, IL: Open Court.(Launched the concentration on a modal version of the ontological argument.)

Hartshorne, C. (1984) *Omnipotence and Other Theological Mistakes*, Albany, NY: State University of New York Press.(The case for a God of limited power.)

Hasker, W. (1989) *God, Time, and Knowledge*, Ithaca, NY: Cornell University Press.(Fine survey of issues concerning God's knowledge of future contingent events.)

Hegel, G.W.F. (1984-7) *Lectures on the Philosophy of Religion*, trans. C.P. Hodgson and R.F. Brown, Berkeley, CA: University of California Press, 3 vols. (Religion as a pictorial version of Hegelian metaphysics.)

Hick, J. (1978) *Evil and the God of Love*, New York: Harper & Row, revised edn. (The classic presentation of natural evil as a contribution to 'soul making'.)

Hume, D. (1748) *An Enquiry Concerning Human Understanding*, La Salle, IL: Open Court, 1907.(Section 10 contains Hume's famous attack on the credibility of miracles.)

Hume, D. (1779) *Dialogues Concerning Natural Religion*, Harmondsworth: Penguin, 1990.(The classic critique of the teleological argument.)

Kant, I. (1781) *Critique of Pure Reason*, trans. N.K. Smith, New York: St Martin's Press, 1929.(One of the major landmarks in the history of philosophy. Difficult.)

Kant, I. (1788) *Critique of Practical Reason*, trans. T.K. Abbott, London: Longmans, Green & Co., 1898.(Kant's major work on ethics, and the source of his moral argument for the existence of God.)

Kant, I. (1793) *Religion Within the Limits of Reason Alone*, trans. T.H. Greene and H.H. Hudson, New York: Harper & Row, 1960.(Kant's major philosophical treatment of religion.)

Kierkegaard, S. (1846) *Concluding Unscientific Postscript*, trans. D.F. Swenson, Princeton, NJ: Princeton University Press, 1944.(A major source of Kierkegaard's radical views on religious faith.)

Leeuw, G. van der (1963) *Religion in Essence and Manifestation*, New York: Harper & Row.(Classic work on the phenomenology of religion.)

Leftow, B. (1991) *Time and Eternity*, Ithaca, NY: Cornell University Press.(The latest major work on divine eternity and a rich treatment of the subject. Difficult.)

Leibniz, G.W. (1710) *Theodicy*, trans. E.M. Huggard, La Salle, IL: Open Court, 1985.(Leibniz's extended treatment of the problem of evil.)

Leslie, J. (1989) *Universes*, London: Routledge.(The teleological argument in the light of contemporary science.)

Malcolm, N. (1963) 'Anselm's Ontological Arguments', in *Knowledge and Certainty*, Englewood Cliffs, NJ: Prentice-Hall.(A discovery of a modal ontological argument in Anselm.)

Mill, J.S. (1874) *Three Essays on Religion*, New York: Liberal Arts Press, 1958. (A powerfully argued case for a God of limited power.)

Mitchell, B. (1981) *The Justification of Religious Belief*, New York: Oxford University Press.(A 'cumulative

case' defence of belief in God.)
Molina, L. de (1588) *On Divine Foreknowledge*, Part IV of the *Concordia*, trans. A.J. Freddoso, Ithaca, NY: Cornell University Press, 1988.(An important sixteenth-century view on divine foreknowledge and human free will.)
Newman, J.H. (1870) *An Essay in Aid of a Grammar of Assent*, Notre Dame, IN: University of Notre Dame Press, 1979.(A major nineteenth-century work on the epistemology of religious belief. Not easy going.)
Paley, W. (1802) *Natural Theology, or Evidences of the Existence and Attributes of the Deity Collected from the Appearances of Nature*, London: R. Faulder & Son; ed. F. Ferré, Indiana, IN: Bobbs-Merrill, 1964.(Important statement of the teleological argument.)
Pike, N. (1969) 'Omnipotence and God's Ability to Sin', *American Philosophical Quarterly* 6 (3): 208-16. (Important discussion of this problem.)
Pike, N. (1970) *God and Timelessness*, London: Routledge & Kegan Paul. (Powerful case for the temporality of God.)
Plantinga, A. (1967) *God and Other Minds*, Ithaca, NY: Cornell University Press. (A masterful discussion of various arguments for the existence of God.)
Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Clarendon Press.(Very technical, but worth the effort for the latest word on the ontological argument.)
Plantinga, A. (1983) 'Reason and Belief in God', in A. Plantinga and N. Wolterstorff (eds.) *Faith and Rationality*, Notre Dame, IN: University of Notre Dame Press. (Important statements of new trends in the epistemology of religious belief.)
Plantinga, A. (1986) 'On Ockham's Way Out', *Faith and Philosophy* 3 (3): 235-69.(A sophisticated 'Ockhamist' way of reconciling divine foreknowledge and human freedom. Difficult.)
Plato (366-360 BC; 360-347 BC) *Laws* and *Timaeus*, in E.Hamilton and H. Cairns (eds.) *The Collected Dialogues of Plato*, Princeton, NJ: Princeton University Press, 1961.(The main sources for Plato's conception of God and arguments for the existence of God.)
Plotinus (*c*.250-66) *Enneads*, Cambridge, MA: Harvard University Press, 1966-88.(The fountainhead of Neoplatonism.)
Rosenkrantz, G. and Hoffman, J. (1980) 'The Omnipotence Paradox, Modality, and Time', *The Southern Journal of Philosophy* 18 (4): 473-9.(A brilliant solution of the paradox of the stone.)
Rowe, W.L. (1975) *The Cosmological Argument*, Princeton, NJ: Princeton University Press.(The best survey of the topic. Rather difficult.)
Stump, E. (1985) 'The Problem of Evil', *Faith and Philosophy* 2 (4): 392-423. (Natural evil as a device by God to 'fix' human wills. Very readable.)
Stump, E. and Kretzmann, N. (1981) 'Eternity', *Journal of Philosophy* 78: 429-57.(An essay that sparked the contemporary revival of interest in divine eternity.)
Swinburne, R. (1977) *The Coherence of Theism*, Oxford: Clarendon Press. (Impressive discussion of divine attributes.)
Swinburne, R. (1979) *The Existence of God*, Oxford: Clarendon Press.(Reformulation of traditional arguments for the existence of God in terms of modern probability theory.)
Tennant, F.R. (1930) *Philosophical Theology*, Cambridge: Cambridge University Press.(Important twentieth-century formulation of the arguments for the existence of God.)
Whitehead, A.N. (1929) *Process and Reality*, New York: Macmillan.(A major metaphysical work, and the source of 'process philosophy' and 'process theology'.)
William of Ockham (after 1317) *Predestination, God's Foreknowledge, and Future Contingents*, trans. M. Adams and N. Kretzmann, New York: Appleton-Century-Crofts, 1969.(One of the most important approaches to this problem.)

Religion, philosophy of

Philosophy of religion is philosophical reflection on religion. It is as old as philosophy itself and has been a standard part of Western philosophy in every period (see Religion, history of philosophy of). In the last half of the twentieth century, there has been a great growth of interest in it, and the range of topics philosophers of religion have considered has also expanded considerably.

Philosophy of religion is sometimes divided into philosophy of religion proper and philosophical theology. This distinction reflects the unease of an earlier period in analytic philosophy, during which philosophers felt that reflection on religion was philosophically respectable only if it confined itself to mere theism and abstracted from all particular religions; anything else was taken to be theology, not philosophy. But most philosophers now feel free to examine philosophically any aspect of religion, including doctrines or practices peculiar to individual religions. Not only are these doctrines and practices generally philosophically interesting in their own right, but often they also raise questions that are helpful for issues in other areas of philosophy. Reflection on the Christian notion of sanctification, for example, sheds light on certain contemporary debates over the nature of freedom of the will (see Sanctification).

1 Philosophy and belief in God

As an examination of mere theism, the core of beliefs common to Western monotheisms, philosophy of religion raises and considers a number of questions. What would anything have to be like to count as God? Is it even possible for human beings to know God's attributes (see God, concepts of; Negative theology)? And if so, what are they? Traditionally, God has been taken to be a necessary being, who is characterized by omniscience, omnipotence, perfect goodness, immutability and eternity (see Necessary being; Omniscience; Omnipotence; Goodness, perfect; Immutability; Eternity), who has freely created the world (see Creation and conservation, religious doctrine of; Freedom, divine), and who is somehow specially related to morality (see Religion and morality).

This conception of God takes God to be unique (see Monotheism), unlike anything else in the world. Consequently, the question arises whether our language is capable of representing God. Some thinkers, such as Moses Maimonides, have argued that it is not and that terms applied to God and creatures are equivocal. Others have argued that our language can be made to apply to God, either because some terms can be used univocally of God and creatures, or because some terms used of creatures can be applied to God in an analogical sense (see Religious language).

Not everyone accepts the traditional characterization of God, of course. Pantheists, for example, reject the distinction between God and creation (see Pantheism). Certain philosophers have objected to the traditional conception on the grounds that it leaves certain philosophical problems, such as the problem of evil, insoluble (see Process theism). And many feminists reject it as patriarchal (see Feminist theology).

Given the traditional conception of God, can we know by reason that such a God exists? There are certain arguments that have been proposed to demonstrate the existence of God so understood (see God, arguments for the existence of; Natural theology). The ontological argument tries to show that a perfect being must exist (see Anselm of Canterbury). The cosmological argument argues that the existence of the world demonstrates the existence of a transcendent cause of the world. And the teleological argument argues from design in nature to the existence of a designer. Some philosophers have maintained that the widespread phenomenon of religious experience also constitutes an argument for the existence of a supernatural object of such experience (see Religious experience; Mysticism, history of; Mysticism, nature of). Most contemporary philosophers regard these arguments as unsuccessful (see Atheism; Agnosticism).

But what exactly is the relation between reason and religious belief? Do we need arguments? Or is faith without argument rational? What is faith? Is it opposed to reason? Some philosophers have argued that any belief not based on evidence is defective or even culpable. This position is not much in favour any more. On the other hand, some contemporary philosophers have suggested that evidence of any sort is unnecessary for religious belief. This position is also controversial (see Faith; Religion and epistemology).

Some philosophers have supposed that these questions are obviated by the problem of evil (see Evil, problem of), which constitutes an argument against God's existence. In their view, God and evil cannot coexist, or at any rate the existence of evil in this world is evidence which disconfirms the existence of God. In response to this challenge to religious belief, some philosophers have held that religious belief can be defended only by a theodicy, an attempt to give a morally sufficient reason for God's allowing evil to exist. Others have thought that religious belief can be defended without a theodicy, by showing the weaknesses in the versions of the argument from evil against God's existence. Finally, some thinkers have argued that only a practical and political approach is the right response to evil in the world (see Liberation theology).

Those who use the existence of evil to argue against the existence of God assume that God, if he existed, could and should intervene in the natural order of the world. Not everyone accepts this view (see Deism). But supposing it is right, how should we understand God's intervention? Does he providentially intervene to guide the world to certain ends (see Providence)? Would an act of divine intervention count as a miracle? What is a miracle, and is it ever rational to believe that a miracle has occurred (see Miracles)? Some people have supposed that a belief that miracles occur is incompatible with or undermined by a recognition of the success of science. Many people also think that certain widely accepted scientific views cast doubt on particular religious beliefs (see Religion and science).

2 Philosophy and religious doctrines and practices

In addition to the issues raised by the traditional conception of God, there are others raised by doctrines common to the Western monotheisms. These include the view that the existence of a human being does not end with the death of the body but continues in an afterlife (see Soul, nature and immortality of the; Reincarnation; Resurrection). Although there is wide variation in beliefs about the nature of the afterlife, typically the afterlife is taken to include heaven and hell. For some groups of Christians, it also includes limbo and purgatory. All of these doctrines raise an array of philosophical questions (see Heaven; Hell; Limbo; Purgatory).

There is equally great variation in views on what it takes for a human being to be accepted into heaven (see Salvation). Christians generally suppose that faith is a necessary, if not a sufficient, requirement (see Justification, religious). But they also suppose that faith is efficacious in this way because of the suffering and death of Jesus Christ (see Incarnation and Christology; Trinity). Christians take sin to be an obstacle to union with God and life in heaven, and they suppose that Christ's atonement is the solution to this problem (see Sin; Atonement). Because of Christ's atonement, divine forgiveness and mercy are available to human beings who are willing to accept it (see Forgiveness and mercy). Most Christians have supposed that this willingness is itself a gift of God (see Grace), but some have supposed that human beings unassisted by grace are able to will or even to do what is good (see Pelagianism). How to interpret these doctrines, or whether they can even be given a consistent interpretation, has been the subject of philosophical discussion.

The religious life is characterized not only by religious belief and experience but by many other things as well (see Phenomenology of religion). For many believers, ritual and prayer structure religious life (see Ritual; Prayer). Christians also suppose that sacraments are important, although Protestants and Catholics differ on the nature and number of the sacraments (see Sacraments). For Christians, the heart of the religious life, made possible by the atonement and the believer's acceptance of grace, consists in the theological virtues - faith, hope, and charity (see Theological virtues).

Many religious believers suppose they know that these and other things are essential to the religious life because God has revealed them (see Revelation). This revelation includes or is incorporated in a book, the Qur'an for Muslims, the Hebrew Bible for Jews, and the Old and New Testaments for Christians. How the texts in this book are to be understood and the way in which religious texts are to be interpreted raise a host of philosophical issues (see Hermeneutics, biblical).

Certain thinkers who are not themselves philosophers are none the less important for the philosophy of religion and so are also included in this encyclopedia. These include, for example, John Calvin and Martin Luther, whose views on such issues as justification and atonement significantly influenced the understanding of these notions, and Jacques Maritain and Pierre Teilhard de Chardin, whose influence on contemporary philosophical theology has been significant.

See also: African traditional religions; Aquinas, T.; Augustine; Barth, K.; Boehme, J.; Bowne, B.P.; Brahman; Buddhist philosophy, Chinese; Buddhist philosophy, Indian; Buddhist philosophy, Japanese; Buddhist philosophy, Korean; East Asian philosophy; Edwards, J.; Eschatology; Existentialist theology; Gnosticism; God, Indian conceptions of; Illumination; Indian and Tibetan philosophy; Islamic philosophy; Islamic theology; Jaina philosophy; Jewish philosophy; Manicheism; Niebuhr, R.; Occasionalism; Omnipresence; Otto, R.; Personalism; Pietism; Postmodern theology; Predestination; Prophecy; Rahner, K.; Religion and political philosophy; Religion, critique of; Religious pluralism; Reprobation; Shintō; Simplicity, divine; Theology, Rabbinic; Theosophy; Tillich, P.; Tindal, M.; Tonghak; Voluntarism; Zoroastrianism

ELEONORE STUMP

References and further reading

Murray, M. and Stump, E. (eds) (1999) *Philosophy of Religion: The Big Questions*, Oxford: Blackwell.(A broad and inclusive anthology of readings in philosophy of religion.)

Quinn, P. and Taliaferro, C. (1997) *A Companion to Philosophy of Religion*, Oxford: Blackwell.(A helpful and comprehensive reference work for philosophy of religion.)

Religious experience

Philosophy is interested in religious experience as a possible source of knowledge of the existence, nature and doings of God. The experiences in question seem to their possessors to be direct, perceptual awarenesses of God. But they may be wrong about this, and many philosophers think they are. Many philosophers think that such experiences are never what they seem, and that no one has a veridical experience of the presence and/or activity of God. The main philosophical reason for supposing that such experiences are in fact sometimes veridical is a principle according to which any apparent experience of something is to be regarded as veridical unless we have sufficient reasons to the contrary. Experiences are innocent until proven guilty. If we do not accept that principle, we will never have sufficient grounds for taking any experience to be veridical - religious, sensory or whatever.

There are critics who think that we do have sufficient reasons to the contrary in the case of religious experience. For one thing, we do not have the same capacity for intersubjective checks of religious experiences that we have with sense perceptions. But to this it can be replied that we should not suppose that sense perception represents the only way in which we can achieve genuine cognitive contact with objective reality. For another thing, it is widely supposed that religious experience can be adequately explained by psychological and social factors, without bringing God into the picture. But even if this-worldly factors are the only immediate causes of the experience, God could figure as a cause farther back in the causal chain. Finally, the disagreements between alleged experiences of God, especially across different religions, provide a reason for doubting the deliverances of religious experience. But it is possible for a number of people to be genuinely experiencing the same thing, even though they disagree as to what it is like. This is a common occurrence in sense perception.

1 The experience of God

The term ‘religious experience’ is properly used for any experiences one has in connection with one’s religious life, including a sense of guilt or release, joys, fears, longings, a sense of gratitude, and so on. But the usual philosophical concern with religious experience has a much narrower focus. It is concerned with experiences taken by the subject to be an experiential awareness of God. To cast the net as widely as possible, and to avoid restricting the discussion to one kind of religion, such as ‘theistic’ religions that think of God in personal terms, let us understand ‘God’ here to range over any supreme reality, however construed.

Here is an anonymous report of such an experience:

> All at once I… felt the presence of God - I tell of the thing just as I was conscious of it - as if his goodness and his power were penetrating me altogether… Then, slowly, the ecstasy left my heart; that is, I felt that God had withdrawn the communion which he had granted… I asked myself if it were possible that Moses on Sinai could have had a more intimate communication with God. I think it well to add that in this ecstasy of mine God had neither form, color, odor, nor taste; moreover, that the feeling of his presence was accompanied by no determinate localization… But the more I seek words to express this intimate intercourse, the more I feel the impossibility of describing the thing by any of our usual images. At bottom the expression most apt to render what I felt is this: God was present, though invisible; he fell under no one of my senses, yet my consciousness perceived him.
>
> (James [1902] 1982: 68)

This report is typical in several respects. First, the awareness of God is experiential, as contrasted with abstract thought (thinking of God, reasoning about God, or asking questions about God). Like sense experience it seems to involve a presentation of the object. The subject takes God to have been present, to have been given to the subject’s awareness, in something like the way in which a tree is presented to one’s awareness when one sees it.

Second, the experience is direct. The subject feels immediately aware of God, rather than being aware of God through being aware of something else. That is, it seems to be analogous to directly seeing another human being in front of you, rather than seeing that person on television, where one is aware of the person through being aware of something else, in this case the television screen. People also report indirect experiences of God:

> There was a mysterious presence in nature and sometimes met within the communion and in praying by oneself, which was my greatest delight, especially when as happened from time to time, *nature became lit up*

> *from inside* with something that came from beyond itself (or seemed to do so to me).
>
> (Beardsworth 1977: 19; original italics)

Third, the experience is completely lacking in sensory content. It is a wholly non-sensory presentation of God. But there are also experiences of God that involve seeing or hearing something:

> I awoke and looking out of my window saw what I took to be a luminous star which gradually came nearer, and appeared as a soft slightly blurred white light. I was seized with violent trembling, but had no fear. I knew that what I felt was great awe. This was followed by a sense of overwhelming love coming to me, and going out from me, then of great compassion from this Outer Presence.
>
> (Beardsworth 1977: 30)

Many find it incredible that a non-sensory experience should involve a presentation of something, but this seems a baseless prejudice. Why should we suppose that our modes of sensory receptivity constitute the only possible vehicles of an experiential awareness of external reality?

Finally, it is a focal experience, one in which the awareness of God attracts one's attention so strongly as to blot out all else for the moment. But there are also lower-intensity experiences that persist over long periods of time as a background to everyday experiences, as in this report from James: 'God surrounds me like the physical atmosphere. He is closer to me than my own breath. In him literally I live and move and have my being' ([1902] 1982: 71-2).

The present discussion will be limited to direct, non-sensory, focal experiences, since they constitute the most distinctive and striking claims to be experientially aware of God.

A great deal of the literature on this subject concentrates on mystical experience, understood as a state in which all distinctions are transcended, even the distinction between subject and object. The person is aware of a seamless unity. Such experience falls under our general category, for it is typically taken by the mystic to be a direct awareness of supreme reality. But since experiences like this pose special problems of their own, the present entry will focus primarily on more moderate cases like the ones cited, in which subjects do not seem to lose their own identity. In spite of this, the term 'mystical experience' will be used as a convenient way of designating what is taken by the subject to be a direct experience of God. Since these subjects suppose themselves to be aware of God in a way analogous to that in which one is aware of things in the physical environment by sense perception, the term 'mystical perception' will be used.

2 Mystical experience as a basis for beliefs about God

The chief philosophical interest in mystical experience concerns the possibility that it serves as a source of knowledge, or justified belief, about God. Those who have such experiences typically take themselves to have learned something from them, as well as receiving additional confirmation of beliefs already held. Usually they suppose only a limited set of beliefs to be justified in this way. These include the belief that God exists, certain beliefs about his nature (for example, that he is loving or powerful), and beliefs about what God is doing *vis-à-vis* the subject at the moment - comforting, condemning, forgiving, inspiring, communicating a certain message. It is rare, at best, for someone to think it possible to come to know that God delivered the Israelites from slavery in Egypt or that he is three persons in one substance just from an experience of God (unless that is something that the subject takes God to be saying). But the beliefs that are derived from such experiences typically play a central role in one's religious life. In another anonymous report, the writer, after speaking of an experience in which he 'felt the perfect unison of my spirit with His' goes on to say: 'My highest faith in God and truest idea of him were then born in me.... My most assuring evidence of his existence is deeply rooted in that hour of vision, in the memory of that supreme experience' (James [1902] 1982: 66-7).

The fact that subjects take themselves to acquire knowledge of God from mystical perception does not guarantee that they do. One can be misled even by sense experience. One can suppose that one saw, at dusk, that there was a car in the distance when it was actually a cow. And in extreme cases one can be subject to complete hallucinations, like Macbeth falsely taking there to be a dagger in front of him. With both sense experience and mystical experience, contradictions between reports prevent us from taking all of them to be veridical. As for mystical

experience, we need look no further than the many cases in which someone supposes that God commanded the murder of as many people of a certain sort as possible, in contrast to those instances in which one is aware of God as supremely loving. In both areas of experience there are two ways in which one can acquire false beliefs. One may not be experiencing the thing, or sort of thing, one supposes one is experiencing. The two cases of sensory illusion just mentioned illustrate this, as do the cases of people supposing themselves to be aware of God telling them to do things that God would not tell anyone to do. But second, even if one is genuinely aware of the sort of thing in question, the beliefs one forms about it may be mistaken. I may genuinely see a car in the distance, but mistakenly think it is moving. I may genuinely be aware of the presence of God, but mistakenly suppose that he refuses to forgive me. A comprehensive discussion would go into both these ways in which experiences can be misleading. But since sceptics about religious experience usually take it that no one is ever genuinely aware of a really existing God, we will concentrate on the former, more radical criticism. The main issue will be whether mystical experiences are significantly often genuine experiences of God, are veridical experiences.

Reasons for a positive answer can be divided into theological and philosophical ones. The former comprise any components of the belief system of a given religion that give us reason for thinking either that God is in principle accessible to human experience, or that particular persons have experienced God's presence on one or another occasion. Theological considerations are out of bounds here. As for philosophical reasons, the most important one, perhaps the only important one, goes as follows.

Any supposition that one perceives something to be the case - that there is an elephant in front of one or that God is strengthening one - is *prima facie* justified. That is, one is justified in supposing this unless there are strong enough reasons to the contrary, strong enough overriders. Overriders can be divided into *rebutters*, strong enough reasons for the falsity of the belief, and *underminers*, strong enough reasons for supposing that the experience, in this instance, does not sufficiently support the belief. In the elephant case, strong reasons for thinking that there is no elephant in the vicinity would be a rebutter. A reason for supposing oneself to be subject to hallucinations, perhaps because of some drug, would be an underminer. According to this position, beliefs formed on the basis of experience possess an initial credibility by virtue of their origin. They are innocent until proven guilty.

This position has been widely advocated for sense perception (see Chisholm 1977, ch. 4). Swinburne (1979, ch. 13) applies it to mystical experience, terming it 'The Principle of Credulity'. Alston (1991) gives it a more social twist; the claim is that any socially established 'doxastic (belief-forming) practice' is to be accepted as a source of (generally) true beliefs unless there are sufficient reasons against its reliability. The position is usually supported by the claim that unless we accord a *prima facie* credibility to all experiential reports, we can have no sufficient reason to trust any experiential source of beliefs. This is the only alternative to a thoroughgoing scepticism.

As applied to mystical experience, this '*prima facie* credibility' position implies that whenever anyone reports having perceived God as being or doing so-and-so, the report is to be accepted as true unless we have sufficient reasons for refusing to do so. Lest one fear that this lets in any crackpot report, remember that we wouldn't regard reports as 'crackpot' unless we had reasons, from other reports of mystical perception or from other theological sources, for denying that what the 'crackpot' believes about God is correct.

3 Objections to regarding experiences of God as veridical

A large number of contemporary philosophers deny that anyone ever genuinely perceives the presence or activity of God. Most of the reasons for this are based on differences, real or alleged, between sense experience and mystical experience. Since we are all, in practice, completely confident of the by-and-large reliability of sense perception, positive analogies between the two will be taken to support a positive epistemic assessment of mystical perception, whereas differences between them will be taken to support a negative assessment (see Alston 1991, chaps 5-7, for further discussion; also Wainwright 1981, ch. 3).

The first and most obvious reason for a negative assessment concerns certain striking differences between sense experience and mystical experience. Sense experience is a common possession of all human beings, whereas, so far as we can tell, mystical experience is not. Even if we are careful to include dim, background experiences in the reckoning, it is foreign to a considerable proportion of the population. Moreover, even for those who are privileged with both, they play vastly different roles in one's life. Sense experience is continuously, insistently and unavoidably present during all our waking hours, while mystical experience, except for a few choice souls, is a

rare phenomenon. Sense experience, especially vision, is vivid and richly detailed, bursting with information, more than we can possibly encode; while mystical experience, by comparison, is meagre and obscure. These differences can lead one to deny that mystical experience is a conduit of information about objective reality.

The differences certainly render mystical experience much less useful as a source of information, but that is quite different from saying that, where it exists, it is never or seldom a veridical experience of God, or that it provides no information about God. Regarding the partial distribution in the population, the defence can point out that there is no a priori reason to suppose that a doxastic practice engaged in by only a part of the population is less likely to be a source of truth than one we all engage in. There are many reliable practices engaged in by only a small minority - for example, research in the physical sciences and wine tasting. We have to learn from experience which features of the world are open to all and which are open only to some people who satisfy special requirements. And it would beg the question to think that experience tells us that no aspect of reality is disclosed in mystical experience. Similar points can be made about the lesser frequency and lesser detail of mystical experience. That shows that it provides less information than sense experience, but not that it provides none.

A second common charge is that the mystic is simply reading prior religious beliefs into a cognitively indifferent (perhaps purely affective) experience, rather than being directly aware of supreme reality. By and large, only Christians report experiences of Christ, only Hindus of Vishnu, and only Buddhists of *nirvāṇa*. Nevertheless, the charge is a serious oversimplification. There are many conversions on the basis of an alleged experience of God. In these cases, subjects do not suppose themselves to be aware of what they were antecedently expecting; quite the contrary. To be sure, one cannot report what one was aware of except by using concepts one has (what else?). But this is as true of sense perception as of mystical perception. When I report what I saw from the train window, I use my conceptual repertoire; I say that I saw houses, trees, a river, and so on. Sense perception and mystical perception would not seem to differ in this respect.

A third reason for partiality is the supposition that mystical experience can be adequately explained in purely naturalistic terms, and that this fact shows it not to constitute an experience of a supernatural reality (see Mackie 1982). It is a basic principle of perception that we cannot be genuinely perceiving an entity that does not make a significant causal contribution to the experience involved. If a tree on the other side of a high stone wall plays no role in eliciting my present visual experience, then I cannot be *seeing* that tree, whatever my experience is like. But if the causes of a religious experience are purely this-worldly, then God plays no role in producing it. Hence I am not really perceiving God in that experience, however it might seem.

There are a number of points the defence might make. First, the claim to an adequate naturalistic causal explanation can be challenged. We are certainly not in possession of any such explanation. At most there are programmatic suggestions - from psychoanalysis, social psychology, and other quarters - as to the general form such an explanation might take. Second, the causal requirement for genuine perception might be challenged, though not with much plausibility. The most impressive defence would be the following. A consideration of sense perception shows that even there the object perceived need not be among the *proximate* causes of the experience. The immediate causes of a sense experience are all within the subject’s brain, which is not perceived. When the causal requirement is satisfied in sense perception, it is by the perceived object’s figuring further back along the causal route leading to the experience. Hence, even if the immediate causes of a mystical experience are all natural, that leaves open the possibility that God figures among the more remote causes of the experience, and in such a way as to qualify as what is perceived. Hence a proximate naturalistic explanation of mystical experience poses no threat to the veridicality of (some) mystical perception.

A fourth important difference between sensory and mystical perception is that there are effective tests for accuracy in the case of the former but not the latter. When someone claims to have seen something - for example, an aeroplane overhead - there are procedures that can yield a conclusive verdict on that claim. We can look into whether suitably qualified observers saw an aeroplane, and check appropriate sources to find whether there are any records of an aeroplane being there at that time. But nothing like this is available for mystical perception. There are checks that are commonly applied in established mystical communities, such as conformity with the background system of religious doctrine, and conducivity to spiritual development and purity of life. But they are far from yielding comparable results. Moreover, there is nothing like the check of other observers we have for sense perceptual reports. If I claim to have been aware of God’s sustaining me in being, there are no conditions

such that if someone else who satisfies those conditions is not (at that time or any time) aware of being sustained in being by God, I will take that as showing that I was mistaken. And the argument of the critic is that this discredits the claim of the mystic to be aware of an objective reality. If my claims to perceive something objective cannot be validated by intersubjective agreement, they have no standing (Martin 1959; Rowe 1982).

The best response to this criticism is to charge the critic with what we might call ‘epistemic imperialism’, unwarrantedly subjecting the outputs of one doxastic practice to the requirements of another. The complaint is that a mystical perception cannot lay claim to putting its subject into effective touch with objective reality because the subject cannot validate this status in the same way as with sense perceptions. But there are various sources of belief that work quite differently from sense experience in this respect. Consider introspection, one’s awareness of one’s own conscious states. My report that I feel relieved cannot be validated by considering whether someone else, who satisfies certain conditions, feels my relief. But it would be absurd to reject introspection as a source of knowledge because of the unavailability of such tests. Unless critics can give a convincing reason for supposing that the criteria available for sense perception constitute a necessary condition for *any* experiential access to objective reality, they are guilty of epistemic chauvinism (to change the metaphor) in rejecting mystical perception for this reason. Epistemic chauvinism is also exhibited by the first criticism in this section, to which the response was to ask why one should suppose that a mode of experience different from sense experience in the ways specified should be less likely to be a source of knowledge.

Another disability of some of the above criticisms is the use of a double standard, whereby mystical perception is taken to be discredited by some feature it shares with sense perception. This is exhibited by the second and third criticisms above. We saw that both modes of perception use antecedent conceptual schemes to report what is perceived, and that both locate the object perceived among remote rather than proximate causes of the experience.

The final criticism concerns the fact of religious pluralism, probably the most epistemologically significant difference between mystical perception and sense perception. Human beings at all times and places and in all cultures use pretty much the same conceptual scheme to specify what they perceive by the senses. The extent to which there is less than complete agreement is a matter of controversy among anthropologists and historians, but the commonality is clearly much greater than it is with mystical experience. The full range of religions, past and present, differ widely in their conceptions of ultimate reality, and their beliefs about it differ even more widely. This point still holds if we confine ourselves to contemporary major world religions. Since the conceptions and beliefs of the various religions are the ones used by people to articulate what they take themselves to be aware of in mystical experience, those articulations are correspondingly different and often contradict each other. ‘Theistic’ religions such as Judaism, Christianity and Islam think of supreme reality as a personal creator who enters into personal relationships with us. Buddhism and certain forms of Hinduism think of ultimate reality in more impersonal terms, as an undifferentiated unity of which nothing can literally be said, or as a void or ‘Nothingness’. The theistic religions also differ in what they believe about what God has done in history, what his plans are and what he requires of us. In the face of this unresolved diversity, how can we think that mystical experience is a conduit of objective truth?

It is important to be clear as to what is and is not involved in this problem. It was pointed out earlier that it would not be sensible to think that mystical perception can generate a complete system of religious belief. Hence the experiential reports of Christians and Muslims might not contradict each other, even though there are contradictions between the two complete bodies of belief. But even if the experiential reports are contradictory to a certain extent, mystical perception could still constitute a genuine awareness of a supreme reality. Two people can both genuinely perceive something even if they disagree as to what it is like. Two witnesses to a car accident might both have really seen the accident, though their accounts of what happened are importantly different. By the same token, mystics from widely different religions might genuinely perceive the same ultimate reality, even if what they perceive it as is quite different.

Still, it must be admitted that the unresolved incompatibilities in mystical perceptual reports count as negative evidence for the claim that mystical perception is, often, a genuine awareness of supreme reality, though for the reasons just given it can hardly be regarded as conclusive evidence against that claim. On this point there would seem to be a standoff between the mystic and the critic, a standoff that would be resolved by strong enough reasons for regarding one of the competing religious traditions to be closest to the truth or, contrariwise, by strong

enough reasons for regarding no religions to have any truth about supernatural dimensions of reality.

See also: Mysticism, nature of; Religion and epistemology; Revelation

WILLIAM P. ALSTON

References and further reading

Alston, W.P. (1991) *Perceiving God: The Epistemology of Religious Experience*, Ithaca, NY: Cornell University Press.(A defence of the thesis that mystical experience is a source of justification for certain kinds of religious beliefs. Rather advanced.)

Alston, W.P. (1993) *The Reliability of Sense Perception*, Ithaca, NY: Cornell University Press.(A critical analysis of attempts to show that sense perception is a reliable source of belief.)

Beardsworth, T. (1977) *A Sense of Presence*, Oxford: Religious Experience Research Unit.(A fine collection of reports of experiences of God.)

Chisholm, R.M. (1977) *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice-Hall, 2nd edn.(A concise presentation of the epistemology of one of America's leading philosophers. Difficult.)

James, W. (1902) *The Varieties of Religious Experience*, New York: Penguin, 1982.(The classic work on the subject. Very readable, as always with James.)

Mackie, J. (1982) *The Miracle of Theism*, Oxford: Clarendon Press.(Critical of theistic religion. Fairly accessible.)

Martin, C.B. (1959) *Religious Belief*, Ithaca, NY: Cornell University Press. (An unsympathetic treatment of many aspects of religious belief, including the claims of mystical experience. Not very difficult.)

Poulain, A. (1950) *The Graces of Interior Prayer*, trans. L.Y. Smith and J.V. Bainvel, London: Routledge & Kegan Paul.(A classic of Catholic mystical theology which presents a strong case for the veridicality of mystical perception. Very readable.)

Rowe, W.L. (1982) 'Religious Experience and the Principle of Credulity', *International Journal for Philosophy of Religion* 13 (2): 85-92.(Fairly technical.)

Swinburne, R. (1979) *The Existence of God*, Oxford: Clarendon Press.(A contemporary classic. Moderately difficult.)

Wainwright, W. (1981) *Mysticism*, Brighton: Harvester Press.(Clearly written.)

Yandell, K.E. (1993) *The Epistemology of Religious Experience*, Cambridge: Cambridge University Press.(A defence of the claim that religious experience provides evidence for religious belief. Difficult.)

Religious language

The main philosophical interest in religious language is in the understanding of what purport to be statements about God. Can they really be what they seem to be - claims to say something true about a divine reality? There are several reasons for denying this. The most prominent of these stems from the verifiability criterion of meaning, according to which an utterance can be a statement that is objectively true or false only if it is possible to verify or falsify it empirically. It is claimed that this is not possible for talk about God. However, the verifiability criterion itself has been severely criticized. Moreover, many religious beliefs do have implications that are, in principle, empirically testable, though not conclusively.

If one is moved to reject the idea that statements about God are what they seem to be, they can be taken as expressions of feelings and attitudes, and/or as guides to a life orientation. To be sure, religious utterances can have these functions even if they are also genuine statements of fact.

If one believes there to be genuine true-or-false statements about God, there are still problems as to how to understand them. We can focus on the construal of the predicates of such statements - for example, 'made the heavens and the earth' and 'commissioned Moses to lead the Israelites out of Egypt'. There is a serious problem here because of two basic features of the situation. First, the terms we apply to God got their meaning from their application to creatures, particularly human beings. Second, God is so radically different from us that it seems that these terms cannot have the same meaning in the two uses. One possibility here is that all these terms are used metaphorically when applied to God, which obviously often happens ('The Lord is my shepherd'). But are there some terms that can be literally true of God? This may be the case if some abstract aspect of the creaturely meaning of a term can be literally applied to God. For example, if one aspect of the meaning of 'makes' when applied to one of us is 'brings about some state of affairs by an act of will', the term 'makes' with that particular meaning might be truly applied to God.

1 The topic

The title of this entry is a misnomer. There is no language peculiar to religion ('Do you speak English, French or religious?'), nor is religion restricted to any particular type of language. A more accurate term for the topic would be 'religious uses of language'. Here we find an enormous diversity. Worshippers engage in praise, thanksgiving, petition, confession, instruction and exhortation. Sacred writings contain cosmological speculations, fictional narratives, historical records, predictions, commandments, theological pronouncements and legal codes. In devotional literature there are biographical reminiscences, theologizing, rules for the spiritual life and descriptions of religious experience. Philosophers of religion have concentrated on a very restricted portion of this plenitude, namely on statements about God or, more generally, about the objects of religious devotion and worship.

There is more than one reason for this selectivity. First, it is the belief aspect of religion with which philosophy is most concerned. This is partly an occupational bias. Philosophy is largely taken up with a critical examination of beliefs, assumptions and presuppositions in all areas. When the inquiry takes a linguistic turn, it is deflected to the linguistic formulations of beliefs. Since the most central statements in a religion are statements about God, they get most of the philosophical attention. Second, the belief system of a religion underlies everything else. Religious believers pray as they do, worship as they do, lead, or try to lead, their lives as they do, take up the attitudes they do, because of what they believe about God. If we are concerned to evaluate a religion, we are well advised to evaluate the belief system of that religion, since its pluses and minuses will have implications for the evaluative status of the whole.

Philosophers raise questions of various degrees of generality about religious statements. Some pose difficulties of understanding. The Christian doctrines of the Trinity (God is three persons in one substance) and the dual nature of Christ (the divine and human natures of one person), the Buddhist doctrine of *nirvāṇa* as the ultimate human fulfilment (is it pure nothingness, or does it have a positive aspect?), and the Hindu doctrine of Brahman (the absolute undifferentiated unity that constitutes all reality) are famous cases. But philosophers have also been concerned with more general questions as to how religious statements are to be understood. Traditionally this investigation has centred on the question of how predicates are to be understood in their application to God, predicates such as 'made the heavens and the earth' and 'knows the inmost secrets of our hearts'. For our grasp of

the subject, ‘God’, presumably comes from some predicates or other. Even if reference to God is at least partly on the basis of experience of God (Alston 1989: ch. 5), still it is also at least partly on the basis of what predicates we apply to him. Thus the question of how to understand predicates as applied to God is the most fundamental issue here.

It has generally been assumed that most apparent statements about God really do have that status, that they are used to make truth claims about a reality that is what it is independently of us, our beliefs, attitudes and conceptual schemes. But in the twentieth century that assumption has frequently been questioned. Hence, before tackling issues about the understanding of theological predicates, the reasons that have been given for denying the credentials of religious statements will be considered, along with the alternative construals of such utterances that have been proposed.

2 Reasons for denying that there are genuine statements about God

These reasons can be ranked under three headings: *metaphysical* - if treated as statements that are true or false, they are all false; *epistemological* - we have no effective way of determining their truth value; and *semantic* - because of the previous epistemological criticism, sentences predicating properties of God do not satisfy necessary conditions for having the kind of meaning (factual meaning) that would fit them for being used to make statements that are true or false.

The *metaphysical* claim is based on a naturalistic or materialistic metaphysics that takes reality to be confined to the ‘natural’ order, that is, to the physical universe in space and time (see Naturalized philosophy of science §1). Since that leaves no room for God, as usually conceived in religion, all statements that purport to refer to such a being are false. To be sure, this is quite compatible with taking what appear to be religious statements to have that status; they just all happen to be false. But if, while embracing naturalism, one is still motivated to hang on to something like traditional religion, the only option is to give religious utterances some nonstandard interpretation.

This line of thought can be no better than the naturalistic metaphysics on which it is based. Though materialists can claim some support from the developments of modern science for the thesis that everything in the spatiotemporal universe is purely physical in nature, it is not clear that they have any significant reason for denying that there are realities of a different order altogether, such as God is typically taken as being.

The *epistemological* reason is that we lack sufficient grounds for supposing that religious statements are true. Whether this is so is an extremely complicated issue that is treated elsewhere (see God, arguments for the existence of; Religion and epistemology; Religious experience). But even if it is so, that only implies that they must be accepted on faith. If one feels uncomfortable with that, it would be another motive for holding that apparent religious statements are not what they seem.

The *semantic* approach draws a stronger conclusion from an epistemological claim like the foregoing. The principle on which this reasoning is based is the verifiability criterion of meaningfulness. According to this principle, a sentence has factual meaning (the kind of meaning that renders it usable to make a statement with a truth value) only if it is in principle possible to verify it or falsify it empirically, on the basis of observations (see Meaning and verification §§2-3). The argument is that alleged statements about God fail this test and hence are not genuine statements of fact. In a famous passage, Antony Flew (1955) posed the rhetorical question, ‘What would have to occur or to have occurred to constitute for you a disproof of the love of, or of the existence of, God?’. The implied answer is that nothing would fit this bill, and, by the same token, nothing would amount to a proof either. Alleged statements about God are only pretend-statements because they are not empirically testable.

It is this argument against the genuineness of statements about God that has received the most press. Again, it can be no stronger than the principle on which it is based, and the verifiability criterion has repeatedly been severely criticized (see Plantinga 1967). But in any case, do statements about God really fall foul of the requirement? That depends on a number of things. First, it depends on what counts as an ‘observation’ or ‘empirical datum’. It makes a big difference whether mystical experience is allowed to count as ‘observation’. Second, it depends on the viability of arguments for the existence of God, such as the teleological argument, that are based on observable features of the universe. And, on the other side, many thinkers take it that massive and apparently undeserved suffering constitutes empirical evidence against the existence of God. Third, it depends on details of the particular religious belief system in question. Many such systems involve fairly straightforward predictions as to what the

gods will do under certain conditions, where these divine actions manifest themselves in sensorily observable changes. In many primitive religions it has been believed that the gods will bring rain or military victory if they are approached through certain rites. In the Judaeo-Christian tradition it is believed that God will see to it that the Church or the chosen people will be finally victorious on earth and that prayers, made in the right spirit and in the right conditions, will be effective. It is true that such predictions come with severe qualifications. We cannot say *when* the Church will be victorious or when the second coming of Christ will inaugurate a new era. And it is impossible to be sure that one has prayed in the right spirit. Thus it is rare to find a decisive empirical test for religious beliefs. (But it is often held that this is true of science as well.) In any event, the issue of whether belief about God is empirically testable to some extent is by no means easily answered.

3 Nonassertive construals of theological statements

If one is convinced that talk of God can be meaningful only if it does not involve statements that are assessable as true or false, there are several options. Here are two:

Expressivism-instrumentalism. This is an analogue of ‘noncognitivism’ or ‘emotivism’ in ethics, a view that also springs from attachment to verificationism and which takes ethical utterances to be expressions of attitudes and emotions, or as recommendations of a policy of action, rather than as statements of fact. A classic version of the application to religion is found in Santayana (1905). On his view, there are two components to a religious doctrine, or ‘myth’. There is (a) an evaluation of some sort, which is (b) expressed in the form of a picture or story. Thus the Christian myth of God’s incarnation in Jesus Christ, and his death on the cross to atone for our sins can be regarded as an expression of the moral value of self-sacrifice. That is the expressive side of the position, but Santayana also thinks of religious myths as *guiding* our lives, our responses to the world. This side of the matter is indicated by the term ‘instrumentalism’, taken from the philosophy of science. The function for which religious beliefs are ‘instrumental’ is not predictive, as in science, but rather ‘life-orienting’ (see Braithwaite 1955).

Symbolisticism. Here a leading figure is Paul Tillich (§§3-5), according to whom it is misguided to ask whether religious doctrines are true or false of God. Since God (the true God, Being-Itself) is beyond any conceptualization (1953: 264-5), it is hopeless to seek any correspondence between what we say about God, and God. Instead, our ‘God-talk’ is made up of symbols of God, which ‘point to’ his reality by ‘participating’ in his power and being. Not only what are commonly recognized as symbols - the lamb, water, the shepherd, and so on - are to be so construed; anything concrete or conceptualizable is a symbol, including Christ, God the Father, and the Holy Spirit. Speaking of God the Father is an appropriate way of symbolizing Being-Itself because fatherhood is one of the ‘places’ in the world where we are ‘grasped’ by the power of Being. But any literal correspondence of our religious utterances with the divine is out of the question.

It must be noted that even if what seem to be factual religious statements do genuinely have that status, they can also function in an expressive-instrumental and a symbolic role.

4 Theological predicates

In discussing the question of how to understand predicates in their application to God we will assume that the arguments of §2 against the genuineness of statements about God are not cogent, and that such statements are to be understood as what they appear to be. Each such statement can be thought of as applying a certain predicate to God. Here is a sample:

(1) made the heavens and the earth;
(2) became incarnate in Jesus of Nazareth;
(3) is omnipotent (perfectly good, loving, wise, omniscient);
(4) forgives the sins of those that repent and turn to him;
(5) told me not to worry so much about trifles;
(6) wills that all people should be saved.

We can roughly divide this list into attributes (3), actions (1, 2, 4, 5) and intentional psychological states (6). The attributes can be seen as derivative from the other types, being properties of God that fit him for actions (loving, omnipotent) or for psychological states (omniscient, wise). Discussions of the topic have tended to concentrate on attributes, perhaps because they can be more concisely formulated, but this tends to distort the subject.

Why should there be a special problem here? We know what it is to tell someone something, to know something, to comfort someone, and so on. Is that not what we are saying of God in these cases?

There is a problem because, first, as the last paragraph suggests, we get our terms (concepts) for talking (thinking) about God from our talk and thought about creatures, particularly human beings, and, second, God is so different from human beings as to make the use of these terms problematic.

Regarding the first point, does it just happen that all the terms we use to specify God's attributes, actions and psychological states are terms we also use for human beings, or is there some deep reason for this? The latter alternative is supported by the following considerations. We have the kind of cognitive access to each other that makes it possible to establish a common language for talking about us. But we do not have the same resources for setting up an independent language for talking of God. To put it roughly, the parent can know when the child is perceiving another person talking or hugging someone or criticizing someone, and this makes it possible to introduce the child to the publicly shared meanings of those terms in application to human beings. But we cannot do anything analogous in the case of God. Even if the child can be aware of being spoken to or forgiven or comforted by God, the parent cannot tell when the child is aware of this unless the child tells the parent, which presupposes that the child already has learned how to apply the terms to God.

The second point is that creaturely terms cannot, in general, be used of God in exactly the same sense because of the ways in which God is different from creatures. There is no universal agreement as to just what these ways are, but some of those most commonly cited are as follows. God is infinite in power, knowledge and goodness; each of us is very limited in these respects. God is purely spiritual; we are embodied. God is omnipresent; we are severely restricted in spatial location. According to some theologies, God enjoys an atemporal mode of being - the simultaneous and complete possession of illimitable life; we live our lives successively, one moment at a time.

These differences have implications for what it is for God to be or do the various things we believe of him. God's knowledge is a very different thing from human knowledge. It is not built up by inference from the deliverances of perception. God has no sense organs, and no need to infer some truths from others. He knows everything directly. Since God is immaterial, divine overt action is a fundamentally different thing from human overt action. The only things we can effect directly are changes in our bodies; we bring about changes in the external world only by moving our bodies in certain ways. But God, having no body, will directly bring about changes in the world. Finally, if God's mode of being is atemporal, the picture will be still more different. Virtually every aspect of our activity is deeply imbued with temporality. We acquire information, think, deliberate and act by a temporal succession of stages.

This makes it clear that predicates cannot be truly applied to God in just the sense in which they apply to creatures. They cannot be used univocally of God and creatures. But that leaves several alternatives. They can be used metaphorically of God. That is very common in religious discourse. When we say 'His *hands* prepared the dry land' or 'The Lord is my *rock* and my *fortress*', the italicized terms are obviously used metaphorically. We do not think that God literally has hands or is a rock. And some people take all, or virtually all, theology to be metaphorical (see McFague 1982). But most theologians have supposed that we can make some literally true statements about God, and that it is important for the foundations of faith that we are able to do so. This has led them to explore other alternatives. A popular tack is that classically taken by Aquinas and developed in various ways by his successors. On this view, (some) creaturely terms are used analogically of God. There is an important analogy between their creaturely and theological import. According to Aquinas (§9), there is not even a partial univocity between 'know', for example, as applied to humans and God, even though there is enough analogy between their senses to enable us to understand what is said of God when he is said to know something. Aquinas' denial of partial univocity stems from his belief in a particularly radical divine-human difference, namely that God is absolutely simple. There is no difference of any sort between 'aspects' of the divine being. Since God is not different from any of his attributes or actions, no term as applied to God can mean even partly the same as to it does when applied to creatures (see Simplicity, divine §§1-2).

Contrary to Aquinas, there is a way in which, for some terms, there can be a partial overlap of meaning in the divine and human applications. The crucial point here is that the meaning of a term often includes more abstract, generic aspects and more concrete, specific ones. The term might carry over its more abstract meaning, though not

its more concrete meaning, from one application to another. This would make it possible for the more abstract component of the concept of knowledge to be common to God and humans, even though the way this is realized (and so the more concrete meaning of the term) would be different in the two cases. Consider an action term, such as ‘makes’. It seems reasonable to suppose that the full concrete meaning of this term, as it is used of human beings in saying ‘John made a bookcase’, includes ‘transforming some pre-existing material’ and ‘effecting this by moving his body in certain ways’. If so, that full meaning does not carry over to ‘God made the universe’, assuming that we mean this to involve creation *ex nihilo* and take God to be immaterial. But there could be a more abstract component of the meaning that is carried over, for example, ‘brings it about by one or more acts of will that a certain state of affairs obtains’. In this way, some parts of the creaturely meanings of some terms can be used literally to make statements about God that have a chance of being true.

See also: Negative theology

WILLIAM P. ALSTON

References and further reading

Alston, W.P. (1989) *Divine Nature and Human Language: Essays in Philosophical Theology*, Ithaca, NY: Cornell University Press.(Philosophical essays on the nature of God and our talk of God.)

Aquinas, T. (1259-65) *Summa contra gentiles*, trans. V.J. Bourke, Garden City, NY: Doubleday, 1956, I, chaps 29-36.(See next item.)

Aquinas, T. (1266-73) *Summa theologiae*, London: Eyre & Spottiswoode, 60 vols, 1964, Ia, q.13.(This and the preceding item provide the classic statement of the view that terms applied to God have an analogical mode of meaning.)

Barbour, I.G. (1974) *Myths, Models, and Paradigms*, New York: Harper & Row. (Religious belief as based on models. Very readable.)

Braithwaite, R.B. (1955) *An Empiricist’s View of the Nature of Religious Belief*, Cambridge: Cambridge University Press.(An interpretation of religious beliefs in terms of stories told for their moral point. Influenced by the verifiability criterion of meaningfulness.)

Crombie, I.M. (1957) ‘The Possibility of Theological Statements’, in B. Mitchell (ed.) *Faith and Logic*, London: Allen & Unwin.(A rich and complex view of the understanding of talk about God.)

Flew, A. (1955) ‘Theology and Falsification’, in A. Flew and A. MacIntyre (eds) *New Essays in Philosophical Theology*, London: SCM Press.(Classic application of the verifiability criterion to religious discourse.)

McFague, S. (1982) *Metaphorical Theology*, Philadelphia, PA: Fortress Press.(A stimulating presentation.)

Plantinga, A. (1967) *God and Other Minds*, Ithaca, NY: Cornell University Press.(A classic in philosophy of religion. Advanced.)

Santayana, G. (1905) *Reason in Religion*, New York: Charles Scribner’s Sons. (An early account of religious discourse as non-factual.)

Tillich, P. (1953) *Systematic Theology*, vol. 1, London: Nisbet & Co.(Sets forth his view of theological discourse as symbolic; see also his ‘Theology and Symbolism’, in F.E. Johnson (ed.) *Religious Symbolism*, New York: Harper & Bros, and ‘Religious Symbols and Our Knowledge of God’, *Christian Scholar* 38 (3): 189-97, both published in 1955.)

Religious pluralism

Religion displays a luxuriant diversity of beliefs and practices. Crusades and colonialism, preaching and proselytizing, argument and apologetics have failed to produce worldwide agreement. In order to understand this situation, four possibilities are worth considering. The first is reductive naturalism. On this view, religious beliefs about a supernatural or transcendent dimension of existence are all false. They are to be explained as products of a merely human projection mechanism. The writings of such naturalistic philosophers and scientists as Feuerbach, Marx, Freud, and Durkheim suggest ways in which such projections might occur. A second possibility is exclusivism. Doctrinal exclusivism is the view that the doctrines of one religion are completely true; the doctrines of all others are false whenever there is conflict. Soteriological exclusivism is the view that only one religion offers an effective path to salvation or liberation. Though the two kinds of exclusivism are logically independent, they are usually held together. A third option, which has found increasing favour in the second half of the twentieth century, is inclusivism: one religion contains the final truth and others contain only approaches to or approximations of it; the privileged religion offers the most effective path to salvation, but those outside it can somehow be saved or liberated. The final option, pluralism, is a relative newcomer. According to pluralism, a single ultimate religious reality is being differently experienced and understood in all the major religious traditions; they all, as far as we can tell, offer equally effective paths to salvation or liberation.

These options raise interesting questions. What accounts for the growing popularity of inclusivism and pluralism? How are we to articulate pluralism? Does exclusivism remain a rational option in spite of what is known about the whole range of religious traditions? Is pluralism, once clearly stated, a rational option?

1 The emergence of inclusivism and pluralism

Many Westerners will best understand the emergence of inclusivism and pluralism in terms of the history of Christianity. For most of its history Christianity has been resolutely exclusivist. In late antiquity it was a new religion, struggling to establish itself in the face of criticism and persecution. It is not surprising to find it making exclusive claims on behalf of its charismatic founder, Jesus of Nazareth. Support for exclusivism was derived from sayings attributed to Jesus in scripture, such as ‘No one comes to the Father except through me’ (John 14: 6). By the third century many Christians held that there was no salvation outside their church (*extra ecclesiam, nulla salus*). During the Middle Ages, Christian Europe had little direct contact with the non-theistic religions of Asia. Islamic civilization was at the height of its powers and so posed a cultural and sometimes a military threat to Christian civilization. Jews were often despised and persecuted. To be sure, thinkers such as Thomas Aquinas managed to learn from Maimonides and Ibn Sina (Avicenna), but there were few opportunities for Christians, Muslims and Jews to interact on a footing of equality and develop mutual religious respect. And when in the modern era Christian Europe became the dominant world power, colonial expansion began and Christian missionaries, whose chief aim was conversion, spread over the face of the earth.

Of course, Christianity is not the only religion to have fostered exclusivist attitudes. In their more militant moments, Muslims have done the same. Some Jews cherish an ethnically exclusive identity as God’s chosen people, and some Hindus revere the Vedas as a source of absolute truth. Buddhists often see in the teachings of Guatama the only *dharma* that can liberate humans from illusion and suffering. Still, the exclusivist strain in Christianity has been unusually strong and persistent.

The turn away from exclusivism on the part of many but not all Christians in the latter half of the twentieth century is therefore a dramatic development. No doubt the full explanation of it will be complex, but it seems clear that at least the following three factors will figure in it. First, many Christians have learned tolerance for other religions. Christian denominations learned tolerance for one another in the aftermath of the Wars of Religion; Christianity is learning tolerance of other religions in the aftermath of colonialism and the Holocaust. In retrospect, many Christians have come to see in exclusivist attitudes both a source of complicity in the evils of colonialism and a stimulus for the anti-Semitism that contributed causally to the Holocaust. Second, many Christians have also come to understand (at least to some extent) other religions, a benefit often conferred on them by the modern academic study of religion. When one can read the important texts of other religious traditions in good translations with helpful commentaries, one can experience for oneself their power, nobility and allure. When one learns that the

beliefs of other religious traditions are supported by experiences and arguments similar to those that support the beliefs of one's own tradition, one is apt to acknowledge the epistemic rationality of the participants in other traditions. And one is likely to acknowledge their practical rationality once one comes to see that their traditions, judged by the fruits they yield, are comparable to one's own in helping people to live well and in producing saintly people.

The third and most important factor is the increasing frequency of intense and cooperative personal interaction among participants in diverse religious traditions. Exclusivist attitudes are under pressure in such contexts. Such interactions have played a major role in the careers of two of the leading pluralist thinkers, Wilfred Cantwell Smith and John Hick. Smith has written about the influence on his thought of having taught for some years at a small Christian college in Lahore in which the majority of the teaching staff and students were Hindus, Muslims or Sikhs (Smith 1963). He also organized and worked for several years in the Institute for Islamic Studies at McGill University, which is structured by the proviso that half the faculty and half the graduate students should be Muslims. Hick has written of the impression made on him by cooperating with Muslims, Jews, Hindus and Sikhs in campaigns for tolerance in Birmingham, England, during the 1960s (Hick 1993). As religious diversity increases within our societies, more and more Christians are involved in and being influenced by such personal interactions, and the influence in many cases has operated to undermine exclusivist attitudes.

If one moves away from exclusivism, the step to pluralism is larger than the step to inclusivism. Since the Second Vatican Council in the 1960s, inclusivists among Roman Catholic theologians have proposed that human salvation depends entirely on the sacrificial death of Christ, but that all humans can somehow be united with Christ and so share in his redemption. To pluralists, this view seems to be an unsatisfactory halfway house. If one is going to allow that people in religious traditions other than Christianity can be saved, it appears to be simpler and more natural to claim that their salvation is mediated by their traditions than to insist that it is mediated by Christ. Thinking that something more radical than inclusivism is called for, Smith, himself a Christian theologian, argues that it is an urgent task for us to work out a world theology or a theology of comparative religion. He suggests that such a theology would involve discourse about the transcendent dimension of human life and of the universe, to which the history of religion bears witness and which it elucidates (Smith 1981). However, this suggestion raises philosophical problems. According to Muslims, the transcendent dimension of the universe is a personal deity, Allah; according to Advaitic Hindus, it is an impersonal absolute, Brahman. These claims seem to be logical contraries. Hence it appears that any theology that takes a stand on the specific nature of the transcendent dimension of the universe will be unable to escape doctrinal exclusivism.

2 The pluralistic hypothesis

Appearances, however, can deceive. Hick has set forth a philosophically sophisticated pluralistic hypothesis that may avoid problems of this sort (1987, 1989). As he sees it, each of the major religious traditions offers a path to salvation or liberation that involves a transformation of human existence from self-centredness to reality-centredness. Judged by their results, all of these traditions are, as far as we can tell, of roughly equal effectiveness in producing this transformation. This suggests that a single ultimate reality is being differently conceived, experienced and responded to from within different traditions. Following Kant (§3), Hick develops the suggestion using the distinction between the phenomenal and the noumenal. His hypothesis is that the noumenal Real is experienced and conceived within different religious traditions as the range of gods and absolutes reported by the phenomenology of religion. Being phenomenal, such gods and absolutes are not illusory but are empirically real manifestations of the noumenal Real.

There are two ways of interpreting these claims. According to the first, the pluralistic hypothesis postulates a single noumenal Real and diverse ways in which it appears and is experienced within different religious traditions. On this interpretation, apparently conflicting claims about ultimate reality are to be understood as truths about how the noumenal Real appears to various groups, but as falsehoods about how it is in itself. It is not contradictory to suppose that the noumenal Real appears as and is experienced as personal by Muslims and impersonal by Advaitic Hindus. Moreover, if we assume that being personal and being impersonal are contraries rather than contradictories, it is consistent to suppose that the noumenal Real is in itself neither personal nor impersonal. On this view, the major religions will be on a par in two senses: they are all equally correct if their apparently conflicting claims are taken to be about the ways in which the noumenal Real appears to them, and they are all

equally mistaken if those claims are taken to be about the ways in which it is in itself.

According to the second interpretation, the pluralistic hypothesis postulates not only a single noumenal Real but also many phenomenal Reals that are joint products of the interaction of the noumenal Real and various human religious traditions. On this interpretation, apparently conflicting claims about ultimate reality are to be understood as truths about the attributes of the diverse phenomenal Reals of the various religious traditions, but falsehoods about the attributes of the noumenal Real. It is not contradictory to suppose that the phenomenal Real of Islam is personal, the distinct phenomenal Real of Advaitic Hinduism is impersonal, and the noumenal Real, which is yet a third thing, has neither of the contrary attributes of being personal and being impersonal. On this view, the major religions will also be on a par in two senses: they are all equally correct if their apparently conflicting claims are taken to be about the attributes of their diverse phenomenal Reals, and they are all equally mistaken if those claims are taken to be about the attributes of the noumenal Real.

It should be noted that, on either interpretation, Hick's pluralistic hypothesis purchases such parity at a high price. It is a rival to the main lines of self-understanding within the major religious traditions. Most members of such traditions would reject the claim that their beliefs are true only of ways in which ultimate reality appears to them, or of phenomenal objects it contributes to producing, and are not true of that reality as it is in itself. If they employed the distinction between the phenomenal and the noumenal, Muslims would be likely to insist that the noumenal Real is personal, and Advaitic Hindus would be likely to insist that it is impersonal. Hick can attribute only mythological truth to such claims, and he takes mythological truth to be nothing more than literal falsity plus a tendency to evoke appropriate dispositional attitudes. So Hick's version of pluralism attributes massive literal error to both Muslims and Advaitic Hindus. This is not surprising, because their traditions remain, for the most part, stoutly pre-Kantian in their self-understandings. Of course, Hick's pluralism attributes equally large literal error to all the other major religious traditions. Hence it will be unacceptable to most people who at this time participate in any of the major religions; such people will prefer to hang on to their doctrinal exclusivism. Is it rational for them to do so in the face of what is known about religious diversity?

3 The rationality of exclusivism

A positive answer to this question is contained in the religious epistemology of William P. Alston. He operates within a doxastic practice approach to epistemology. A doxastic practice is a way of forming beliefs and evaluating them epistemically in terms of a background system of beliefs that furnishes potential overriders. For example, forming beliefs on the basis of sense perception is a doxastic practice. Alston (1991) argues at length that it is practically rational to engage in socially established doxastic practices that are not demonstrably unreliable or otherwise unqualified for rational acceptance.

In the religious sphere, mystical perception is defined by Alston as religious experience in which a presentation or appearance to the subject of something the subject identifies as ultimate reality occurs (see Religious experience). As he sees it, there are different socially established mystical doxastic practices in the major religious traditions because there are substantial differences in their overrider systems of background beliefs. One such practice is Christian mystical practice. Alston argues that it is not demonstrably unreliable, and parallel arguments can be made for the mystical practices of other major religions. However, both the outputs of Christian mystical practice and its overrider system are massively inconsistent with their counterparts in the mystical practices of other religious traditions, and those mystical practices are in turn massively inconsistent with one another. Hence at most one of them can be a reliable way of forming beliefs about ultimate reality. But why should one suppose that Christian mystical practice is the one that is reliable if any is?

Of course, Christian mystical practice can come up with internal reasons for supposing that it is more reliable than its rivals. But each of them can do the same. Moreover, as Alston points out, Christian mystical practice enjoys significant self-support from the way in which promises it represents God as making are fulfilled in the spiritual lives of its practitioners. But each of its rivals enjoys similar self-support. Thus it seems that Christian mystical practice is disqualified for rational acceptance unless it can come up with sufficient independent reasons for supposing that it is reliable or, at least, more reliable than its rivals. By parity of reasoning, the same thing seems to be true of the mystical practices of other religious traditions.

Alston does not try to provide independent reasons for the reliability of Christian mystical practice, and he

acknowledges that its epistemic efficacy is weakened in the absence of such reasons. However, he denies that it is reduced to the point at which it is no longer qualified for rational acceptance, and supports this denial with an analogical argument. It proceeds by comparing the actual diversity of mystical perceptual practices and a hypothetical diversity of sensory perceptual practices. Suppose that in certain cultures there were a socially established Cartesian practice of seeing what is visually perceived as an indefinitely extended medium that is more or less concentrated at various points, rather than (as in our actual Aristotelian practice) as made up of more or less discrete objects scattered about in space. Further suppose that both these practices are equal in terms of the fruits they produce in terms of enabling their practitioners to deal successfully with their physical environment. Imagine also that in this scenario we are as firmly wedded to our Aristotelian practice as we are in fact, yet can find no independent reasons for supposing that it is more reliable than the rival Cartesian practice. According to Alston, in the absence of independent reasons for thinking that the Cartesian practice is more reliable than our own, the rational thing for us to do is to stick with our Aristotelian practice, of which we are masters and which serves us well in guiding our activity in the world. But the imagined situation is precisely parallel to our actual situation with respect to Christian mystical practice. Therefore, by analogy, the rational thing for those engaged in Christian mystical practice to do is to stick with it and continue to accept the Christian system of belief. Since that system of belief as currently constituted is exclusivist, it follows that it is rational for its adherents to remain Christian exclusivists. And, of course, by parity of reasoning, similar conclusions hold for those engaged in the mystical practices of other religious traditions. Leaving aside independent reasons for thinking that alternatives are more reliable than their own practices, the rational thing for them to do is also to stick with those practices and to accept their exclusivist systems of belief.

4 The rationality of pluralism

Even if it is granted that Alston has convincingly argued for the practical rationality of sticking with the exclusivist beliefs of one’s own mystical practice, the claim that this is the only rational thing to do can be challenged. To return to his analogy, sticking with our own Aristotelian practice and switching to the rival Cartesian practice are not the only options. We can also revise our Aristotelian practice from within and try to get the revised practice socially established. Revisions might proceed in a Kantian direction. Suppose it occurs to us that the success of both Aristotelian and Cartesian sensory perceptual practices in the imagined situation can be explained by the hypothesis that each is reliable with respect to the appearances things present to its practitioners, but neither is reliable with respect to how things are in themselves. Moved by this consideration, we decide to modify our perceptual practice so that it comes to map sensory inputs onto doxastic outputs about the appearances things present to us, but not about how things really are independent of us. Precedent for such modifications is found in the way people respond to learning that such things as phenomenal colours are not mind-independent. It seems that it would be rational to proceed in accordance with this scenario. In the imagined situation, then, it would be rational to stick with our Aristotelian practice, but it would also be rational to transform it into a Kantian practice.

The imagined situation is, of course, relevantly parallel to the actual situation with regard to competing mystical practices. Hence, by analogy, though it is rational for those engaging in Christian mystical practice to continue to do so, it is not the only rational thing to do, there being more than one thing it is rational to do in the face of competing mystical practices. Another such thing would be to modify Christian mystical practice from within to make it more Kantian. Each of these courses of action is rationally permissible in the light of religious diversity. Neither is irrational, but neither is rationally required. And setting aside any relevant dissimilarities, the lesson applies not only for those engaged in Christian mystical practice, but also for those engaged in other socially established mystical practices with significant self-support.

See also: Mysticism, history of; Mysticism, nature of

PHILIP L. QUINN

References and further reading

Alston, W.P. (1991) *Perceiving God: The Epistemology of Religious Experience*, Ithaca, NY and London: Cornell University Press.(Chapter 7 provides full details of the argument summarized in §3.)

Hick, J. (1987) ‘Religious Pluralism’, in M. Eliade (ed.) *The Encyclopedia of Religion*, New York: Macmillan.(A brief and informal summary of the position discussed in §2.)

Hick, J. (1989) *An Interpretation of Religion: Human Responses to the Transcendent*, New Haven, CT and

London: Yale University Press.(Detailed presentation of the position discussed in §2.)
Hick, J. (1993) *Disputed Questions in Theology and the Philosophy of Religion*, New Haven, CT: Yale University Press.(Essay 8, 'A Personal Note', contains the autobiographical information referred to in §1.)
Plantinga, A. (1995) 'Pluralism: A Defense of Religious Exclusivism', in T. Senor (ed.) *The Rationality of Belief and the Plurality of Faith*, Ithaca, NY and London: Cornell University Press.(Presents a critique of Hick's religious pluralism. The volume also contains interesting essays on religious pluralism by George Mavrodes, Peter van Inwagen and William Wainwright.)
Quinn, P.L. (1995) 'Towards Thinner Theologies: Hick and Alston on Religious Diversity', in E.T. Long (ed.) *God, Reason and Religions*, Dordrecht, Boston, MA and London: Kluwer Academic Publishers.(Expansion of the material of §§2-4 of this entry.)
Smith, H. (1991) *The World's Religions*, San Francisco, CA: Harper.(A readable and informative overview of the major religious traditions.)
Smith, W.C. (1963) *The Faith of Other Men*, New York: New American Library. (The introduction contains the autobiographical information referred to in §1.)
Smith, W.C. (1976) *Religious Diversity*, New York: Harper & Row.(Essays on various issues connected with religious diversity.)
Smith, W.C. (1981) *Towards a World Theology*, Philadelphia, PA: The Westminster Press. (Contains details of the proposal for a theology of comparative religion mentioned in §1.)

Renaissance philosophy

The term ‘Renaissance’ means rebirth, and was originally used to designate a rebirth of the arts and literature that began in mid-fourteenth century Italy (see Humanism, Renaissance). Here the term is simply used to refer to the period from 1400 to 1600, but there are ways in which Renaissance philosophy can be seen as a rebirth, for it encompasses the rediscovery of Plato and Neoplatonism (see Platonism, Renaissance), the revival of such ancient systems as Stoicism and scepticism (see Scepticism, Renaissance; Stoicism), and a renewed interest in magic and the occult. Continuity with the Middle Ages is equally important. Despite the attacks of humanists and Platonists, Aristotelianism predominated throughout the Renaissance, and many philosophers continued to work within the scholastic tradition (see Aristotelianism, Renaissance).

1 Historical and social factors

Three historical events were of particular importance. First is the Turkish advance, culminating in the capture of Constantinople in 1453. This advance produced a migration of Greek scholars (like George of Trebizond) and Greek texts into the Latin-speaking West (see Humanism, Renaissance §4; Platonism, Renaissance §3). It also led to a search for new trade routes. The European discovery of the Americas and the first voyages to China and Japan widened intellectual horizons through an awareness of new languages, religions and cultures (see Scepticism, Renaissance §2). New issues of colonialism, slavery and the rights of non-Christian peoples had an impact on legal and political philosophy (see Latin America, colonial thought in §1; Vitoria, F. de; Soto, D. de; Suárez, F.). The study of mathematics and science (especially astronomy) was also affected by developments in navigation, trade and banking, by new technology such as the telescope and other instruments (see Kepler, J.; Galilei, Galileo), as well as by the recovery of Greek mathematics and the favourable attitude of Plato towards mathematical studies (see Platonism, Renaissance §6).

Second is the development of printing in the mid-fifteenth century (see Humanism, Renaissance §6). This allowed for the publication of scholarly text editions, for the expansion of learning beyond the universities, and for the increased use of vernacular languages for written material (see Humanism, Renaissance §4). These changes particularly affected women, who were most often literate only in the vernacular. Christine de Pizan, Paracelsus, Ramus, Montaigne, Bruno and Charron are among those who used vernacular languages in at least some of their works.

Third is the Protestant reformation in the first part of the sixteenth century (see Luther, M.; Calvin, J.). Protestant insistence on Bible reading in the vernacular strengthened both the use of the vernacular and the spread of literacy (see Melanchthon, P. §1). The Catholic Counter-Reformation also affected education, particularly through the work of the Jesuit Order (founded 1540), which set up educational institutions throughout Europe, including the Collegio Romano in Rome (founded 1553) and the secondary school at La Flèche, where Descartes was educated (for further examples, see Collegium Conimbricense; Aristotelianism in the 17th century §2). Political philosophy took new directions (see Hooker, R., for example) and theological studies changed. As the Protestants abandoned the *Sentences* of Peter Lombard and emphasized the church fathers, so the Catholics replaced the *Sentences* with the *Summa theologiae* of Thomas Aquinas. In turn, these changes affected the undergraduate curriculum, which (for other reasons as well) became less technically demanding, especially in relation to logic studies (see Logic, Renaissance). Personal liberties, too, were affected. Both Catholics and Protestants censored undesirable views, and the first Roman Catholic Index of Prohibited Books was drawn up in 1559. Bruno was burnt for heresy, and Campanella was imprisoned. Calls for tolerance by such men as Montaigne and Lipsius were not always favourably received. The books of all these men, and others such as Erasmus, Machiavelli and Rabelais, were placed on the Index or required to be revised. At the same time, Calvinist Geneva prohibited the printing of Thomas Aquinas and Rabelais.

Social factors also affected philosophy which, as an academic discipline, was tied to the universities. These continued to accept only male students, and to teach in Latin, the universal language of learning and of the Roman Catholic Church, but more students came from higher social classes than during the Middle Ages. They expected a curriculum with less emphasis on technical logic and natural science and more on rhetoric, modern languages, history and other practical disciplines. Such curricular changes owed much to humanism, as did the spread of new

secondary schools (see Humanism, Renaissance §7; Montaigne, M. de §1).

The Renaissance was also notable for the spread of learning outside the university. Some men largely relied on the patronage of nobles, princes and popes (among them Valla, Ficino, Pico della Mirandola and Erasmus), some were medical practitioners (including Paracelsus and Cardano), some had private resources (like Montaigne). Nor was it only men that were involved: Christine de Pizan, for example, was a court poet.

2 Humanism and the recovery of ancient texts

Humanism was primarily a cultural and educational programme (see Humanism, Renaissance; Petrarca, F.; Agricola, R.; Erasmus, D.; Vives, J.L.; Rabelais, F.). Humanists were very much concerned with classical scholarship, especially the study of Greek, and with the imitation of classical models. Despite their frequent criticisms of scholastic jargon and techniques, they were not direct rivals of scholastic philosophers, except in so far as changes to the university curriculum brought about by the influence of humanist ideals diluted or squeezed out scholastic subjects. It was humanism that led to the rediscovery of classical texts, and their dissemination in printed form, in Greek and in Latin translation. Plato is the most notable example, but he was rediscovered with the Neoplatonists, and was often read through Neoplatonic eyes (see Ficino, M.; Platonism, Renaissance §1). The so-called ancient wisdom of Hermeticism (also known as Hermetism) was also recaptured within a Neoplatonic framework (see Ficino, M. §2; Patrizi da Cherso, F.), and, along with the Kabbalah (see Kabbalah), led to a revived interest in magic and the occult (see Alchemy §5; Agrippa von Nettesheim, H.C.; Bruno, G.; Paracelsus). These streams also fed into the new vitalistic philosophy of nature (in such thinkers as Paracelsus, Bruno, Campanella, Cardano and Telesio). Other ancient schools of thought that were revived include Epicureanism (see Valla, L. §2), scepticism (see Agrippa von Nettesheim, H.C.; Erasmus, D. §4; Sanches, F.; Montaigne, M. de; Charron, P.) and Stoicism (see Lipsius, J.).

Some humanists wrote important works on education, including the education of women (see Erasmus, D. §2; Vives, J.L. §§1, 4). The Lutheran Aristotelian Melanchthon was also an educational reformer; and the Jesuits drew up the *Ratio Studiorum* (Plan of Studies) which prescribed texts for all Jesuit institutions (see Collegium Conimbricense; Fonseca, P. da §1). Humanism also affected Bible studies (see Erasmus, D. §1; Luther, M. §3; Humanism, Renaissance §6) and Aristotelianism itself (see Aristotelianism, Renaissance §1).

3 Scholasticism and Aristotle

Scholastic philosophy was the philosophy of the schools, the philosophy which was taught in institutions of higher learning, whether the secular universities or the institutions of religious orders. The association of late scholastic philosophy with institutions of higher learning carried with it a certain method of presentation, one which is both highly organized and argumentative, with a clear account of views for and against a given thesis. It also carried with it a focus on Aristotle, for it was Aristotle who provided most of the basic textbooks in the sixteenth- and even the seventeenth-century university. Nor was the study of Aristotle necessarily carried on in a rigidly traditional manner, for many different Aristotelianisms were developed (see Aristotelianism, Renaissance). Moreover, particularly within the Jesuit order, there was a strong inclination to include new developments in mathematics and astronomy within the framework of Aristotelian natural philosophy (see Aristotelianism in the 17th century §1).

Aristotelians include Paul of Venice, George of Trebizond, Vernia, Nifo, Pomponazzi, Melanchthon, Zabarella and the Thomists (see below). Anti-Aristotelians include Petrarch (see Petrarca, F.), Blasius of Parma, Valla, Ramus, Sanches, Telesio, Patrizi da Cherso and Campanella. Some philosophers sought to reconcile Platonism and Aristotelianism (see Pico della Mirandola, G. §4; Platonism, Renaissance §§3, 6).

A very important characteristic of late scholastic philosophy is its use of medieval terminology, along with its continued, explicit, concern both with problems stemming from medieval philosophy and with medieval philosophers themselves. There are fashions here as elsewhere. Albertism (the philosophy of Albert the Great) was important in the fifteenth century (see Aristotelianism, medieval §5); nominalism more or less disappeared after a final flowering in the early sixteenth-century (see Biel, G.; Major, J.). Scotism declined significantly, but was still present in the seventeenth century (see Aristotelianism in the 17th century §§2-4; Latin America, colonial thought in §2). Thomism underwent a strong revival especially through the work of the Dominicans (Capreolus, Cajetan,

Silvestri, Vitoria, Soto, Báñez and John of St Thomas) and the Jesuits (Fonseca, Toletus, Suárez and Rubio: see Latin America, colonial thought in §§2, 5).

4 Philosophical themes

It is difficult to map the interests of Renaissance philosophers on to the interests of contemporary philosophers, especially as the main form of writing remained the commentary, whether on Aristotle or Aquinas. Suárez is the first well-known author to write a major systematic work of metaphysics that is not a commentary, though earlier authors (such as Nifo and Pomponazzi) had written shorter works on particular themes. Nonetheless, certain general themes can be isolated:

4.1 Logic and language. Logic was basic to the curriculum of all educational institutions, and many Renaissance philosophers wrote on logic (see Logic, Renaissance; Language, Renaissance philosophy of; Aristotelianism, Renaissance §2; Logic in the 17th and 18th centuries §§1-2). Individual humanists who worked in this field include Valla (§4), Agricola, Vives; Melanchthon (§2) and Ramus (§2); individual scholastics include Soto (§2), Toletus and Fonseca (§2). Theories of logic and language were often closely related to metaphysics and philosophy of mind, as well as to science.

4.2 Metaphysics and philosophy of mind. Among the themes that overlapped with theories of logic and language were: (i) mental language (see Language, Renaissance philosophy of §2; Language, medieval philosophy of §2; see also Language of thought); (ii) analogy (see Language, Renaissance philosophy of §4; Capreolus, J.; Cajetan §2; Silvestri, F.; Fonseca, P. da §3); (iii) objective and formal concepts (see Capreolus, J.; Fonseca, P. da §3; Suárez, F. §2; Language, Renaissance philosophy of §4); (iv) beings of reason (see Language, Renaissance philosophy of §3; Fonseca, P. da §3; John of St Thomas §4). A specifically Thomistic theme in metaphysics was the relation between essence and existence (see Aquinas, T. §9; Cajetan §5; Fonseca, P. da §3; Báñez, D.; Suárez, F. §2). Other metaphysical issues include: (i) universals (see Paul of Venice; Biel, G. §4; Suárez, F. §2); (ii) individuation (see Capreolus, J.; Cajetan §3; Suárez, F. §2); and (iii) the Great Chain of Being (see Paul of Venice; Ficino, M. §3; Pomponazzi, P. §4; Bruno, G. §5; Aristotelianism, Renaissance §6). Issues in the philosophy of mind included the existence of an agent sense (see Blasius of Parma §3; Nifo, A. §3; Aristotelianism, Renaissance §4) and of intelligible species (Cajetan §4; Nifo, A. §3; Toletus, F. §5; Aristotelianism, Renaissance §4).

4.3 Immortality. The biggest single issue was the nature of the intellectual soul, whether it was immortal, and if so, whether its immortality could be proved (see Paul of Venice; Blasius of Parma §2; Ficino, M. §3; Cajetan §4; Silvestri, F.; Pomponazzi, P. §2; Nifo, A. §3; Toletus, F. §5; Suárez, F. §3; John of St Thomas §3; Aristotelianism, Renaissance §5; see also Soul, nature and immortality of the).

4.4 Free will. Free will was a topic closely connected with the religious issues of grace, predestination and God's foreknowledge (see Valla, L. §3; Biel, G. §3; Pomponazzi, P. §4; Luther, M. §3; Erasmus, D. §1; Calvin, J. §4; Molina, L. de; Molinism; Báñez, D.; Providence §3).

4.5 Science and philosophy of nature. The discussion of scientific method also overlaps with logic (see Logic, Renaissance §7; Aristotelianism, Renaissance §2; Vernia, N. §2; Zabarella, J. §§4-5; Latin America, colonial thought in §5). Themes include (i) traditional Aristotelian discussions about the object of natural philosophy (see Aristotelianism, Renaissance §3; Vernia, N. §3; Zabarella, J. §6, John of St Thomas §3); (ii) Anti-Aristotelian materialism (see Blasius of Parma); (iii) the new philosophies of nature which saw the universe as full of life (see Paracelsus; Bruno, G.; Campanella, T.; Cardano, G.; Telesio, B.) or as explicable in terms of light-metaphysics (see Patrizi da Cherso, F.); (iv) tentative approaches to empiricism (see Vives, J.L. §4; Ramus, P. §4; Sanches, F. §3). Finally, there are the thinkers who set science on a new path by using a combination of mathematical description and experiment (such as Copernicus, Kepler and Galileo).

4.6 Moral and political philosophy. Humanists were deeply concerned with moral and political philosophy (see Humanism, Renaissance §2; Erasmus, D. §3; Vives, J.L. §2), as were Protestant reformers (see Melanchthon, P. §3; Calvin, J. §5). Although the central focus remained on Aristotle (see Aristotelianism, Renaissance §7), Epicurean moral philosophy was taken up by Valla and Stoic moral philosophy was also influential (see Lipsius, J.; Charron, P. §3). Major political thinkers included Machiavelli, Vitoria and Bodin. Many discussions of forms of government, the status of law, and the notion of a just war grew out of the Aristotelian-Thomistic tradition - prominent contributors to this tradition include Christine de Pizan, Vitoria, Soto (§4), Toletus (§2), Suárez (§4),

Molina (§4) and Hooker (see also Latin America, colonial thought in §1). Other significant types of Renaissance political philosophy include: (i) conciliarism (see Ailly, P. d'; Gerson, J.; Nicholas of Cusa; Major, J. §3); (ii) utopianism (see Utopianism; Rabelais, F. §2; Campanella, T. §3); (iii) Neostoicism (see Lipsius, J.). (See also Political philosophy, history of; Religion and political philosophy; Natural law.)

4.7 The human being. Themes related to the human being that were prominent in the Renaissance include: (i) the distinction between microcosm and macrocosm (Nicholas of Cusa; Pico della Mirandola, G. §2; Paracelsus §2; Campanella, T. §2); (ii) love (Ficino, M. §4; Pico della Mirandola, G. §2); (iii) the ability to shape one's own nature (Pico della Mirandola, G. §3; Pomponazzi, P. §4).

E.J. ASHWORTH

Renner, Karl (1870-1950)

Karl Renner was a leading contributor to democratic-socialist legal theory within the Neo-Kantian 'Austro-Marxist' interpretation of socialism that developed in late nineteenth-century Vienna. For Renner and his associates, law is a fundamental and universal institution in any ordered human society. Given the universal necessity of law, the development of socialism may well proceed in an evolutionary rather than a revolutionary manner - and certainly, within the context of Viennese social democracy, it would proceed better by evolution.

A view of Soviet Marxism is expressed by such as E.B. Pashukanis, for whom law in its full bourgeois development is simply the commodity structure reflected, while the juridical subject is an ideological representation of *homo economicus*. Law can neither contribute to nor survive the revolution. By contrast, Renner analyses human society in terms of three universal 'orders': the orders of power, of labour and of goods. Without power-relations, he holds, there can be no order in society; without social labour, there can be no material means for survival; goods for use and consumption are the product of labour and a necessity for social existence. So Marx's analysis of capitalist society can be recast in terms of the subordination of power and labour to property: the order of goods dominates the order of power and the order of labour. Any transformation in fundamental social forms will be expressed through a reshaping of the three orders (see Social theory and law §3).

Though it may seem paradoxical, the institutions of private law, at the level of their formal constitutive and regulative rules, persist in a superficially unchanged way through economic and societal change. The slave-owner, the subsistence farmer, and the joint-stock company of different epochs and modes of production enjoy property rights expressible in substantially unchanging juridical terms. But the institution's function in society will be transformed by shifts in the three orders. This in turn is reflected in development of 'complementary institutions', earlier of private law, later of labour law, welfare law, and planning law. The development through public law of institutions that subordinate the order of goods to those of labour and power is the evolutionary route to socialism.

As well as contributing significantly to legal theory, Renner played a large and important part in Austrian public life, with the obvious exception of the years 1934-45. One of the founding fathers of the Austrian republican constitution in 1919, he was first Chancellor of the Republic in 1919, and remained active in law and legal politics until the troubled thirties. After 1945, he became briefly Chancellor of the restored Republic, and was subsequently its President till his death in 1950. At the end of the twentieth century, his work is in something of an eclipse; but it remains worthy of attention as an attempt to state a democratic-socialist conception of law outside the sphere of the socio-legal ideas discredited in the revolutions of 1989.

See also: Law, philosophy of

RICHARD KINSEY NEIL MacCORMICK

List of works

Renner, K. (1949) *The Institutions of Private Law and their Social Functions*, trans. A. Schwarzschild, ed. with intro. O. Kahn-Freund, London: Routledge & Kegan Paul. (An excellent and clear translation of Renner's *magnum opus*, his only substantial work available in English. Kahn-Freund's introduction, though somewhat dated, remains of real value for the contemporary reader with an English-speaking or common-law background and approach.)

References and further reading

Robson, P. (1977) 'Renner Revisited', in E. Attwooll (ed.), *Perspectives in Jurisprudence*, Glasgow: Glasgow University Press, 221-35.(A critical rereading of Renner's work which looks at it in terms of a socialist perspective in legal theory of the 1970s in the UK.)

Kinsey, R. (1983) 'Karl Renner on Socialist Legality', in D. Sugarman (ed.), *Legality, Ideology and the State*, London: Academic Press, 11-42.(A very clear account of Renner's theory, setting it properly in the context of the debate between Austro-Marxism and Soviet Marxism, and explaining the 'three orders' briefly sketched in the present article.)

Renouvier, Charles Bernard (1815-1903)

Charles Renouvier is the main representative of French Neo-Kantianism in the nineteenth century. Following Kant, he delimited the conditions for the legitimate exercise of the faculty of knowledge, and denounced the illusions of past metaphysics. Wishing to go further than Kant in this direction, he criticized the notions of substance and of actual infinity. According to him, relation is the basis of all our representations, reality is finite, and certainty rests on liberty. In ethics, he took into consideration, beyond the ideal of duty, the existence of the desires and interests to which history testifies.

Renouvier became known shortly after leaving the École polytechnique through an academic treatise on Cartesianism. His first philosophy, known as 'the philosophy of manuals', is marked by the attempt to reconcile contradictories in being and by republicanism in political thought.

The discovery of contradictions of the actual infinite led to the real system of 'neo-criticism', a doctrine combining phenomenalism, finitism and apriorism. Neo-criticism rests on the principle of relativity: our knowledge is restricted to things as they appear to us, as they enter into our representations, which are subject to the laws that govern our experience a priori. Thus things of which we have no representation do not exist, or are as nothing to us. Agreeing with Auguste Comte, Renouvier substitutes the investigation into phenomena and laws for the investigation into essences. He formulates the 'principle of number', according to which everything given is numbered, and shows that space, time, matter and movement could not be posited in themselves without implying the contradictory notion of an actual infinity of parts. He undertakes to destabilize metaphysics, which he calls an 'idolology', by refuting its illusions or 'idols' (the Absolute, the Substance, the Infinite, pure Oneness, pure Being) and to refine the Kantian system, in which he perceives a remnant of the substantialism of pre-critical philosophy. His criticisms are many: Kant makes the mistake of identifying the real with the unknowable thing in itself, and of relegating liberty to the noumenal sphere; sensibility is not a separate faculty from the understanding; relation is the principle law of representation, which underlies all the other categories (number, position, succession, quality, becoming, causality, finality, personality); the enumeration of these categories is not justifiable, but they must be simply submitted to the verification of each person; finally, the system of antinomies juxtaposes theses of unequal value, as the antitheses are contradictory in themselves, while the theses are simply incomprehensible.

The affirmation of liberty marks the transition from theoretical to practical reason. Following his friend Jules Lequier, Renouvier maintains that liberty is the basis of certainty. The scope of evidence is limited to the perception of actual phenomena. Certainty rests on belief, and belief on liberty. Our affirmations are governed, at bottom, by free acts. Thus the existence of external things, pushed back beyond the limits of knowledge, is rehabilitated in the field of belief. Opposed to the doctrine of 'inevitable progress', Renouvier proposes an 'analytic' philosophy of history, proceeding from complex data to elementary facts, without claiming that these facts could not be combined in another way, and explaining rise and decline in history through the free actions of individuals and societies. He proposes on the other hand a 'science of ethics', comprising a pure morality, with the notions of justice and obligation or the categorical imperative, and an applied morality, understood as a collection of principles inspired by pure morality, but taking into account the interests and desires revealed by experience and history. Renouvier considered that morality, like mathematics, must be based on pure concepts, but by confronting the ideal and the real, the state of peace and the state of war, and in formulating a right distinct from pure rationality and adapted to real humanity, he escapes from formalism, and constructs a morality that deals with concrete questions, such as marriage, property, taxation, universal suffrage and the death penalty.

Renouvier's final philosophy of 'personalism' is characterized by a metaphysics inspired by Leibniz and by new postulates concerning the divine and the origin of evil. Being, that is not in itself, *is* nevertheless, as it is representation for itself; it is composed of monads, defined by the functions of spontaneous activity, perception and appetition. Man, in so far as he is disposed towards a reflective consciousness and freedom, is a person. God is a person also, but the first and creator of the universe, understood in reference to the world and to our consciousness, which constitute the sole data by which we can formulate the idea of him, through acts of induction and generalizations. By admitting the fact of a first cause, Renouvier avoided the thesis, which to him was contradictory, of the infinite regress of phenomena; he recognized that it was a question of a limit for

understanding, but notes that the beginning of a series of phenomena due to our will is no less incomprehensible. In this period, Renouvier applied himself to summarizing the conflict of philosophical doctrines in the form of dilemmas, and opposed, on each subject, the 'doctrine of consciousness' to the 'doctrine of things'. The postulate required by practical reason to account for the existence of evil leads to an astonishing cosmogony: primitive humanity, a happy society instituted in a perfect natural order, was taken to collapse due to the effects of self-love and injustice; once the harmony of cosmic forces and human ends was corrupted, the whole was doomed to break up, forming the nebula, which is the probable origin of the present world.

See also: Neo-Kantianism §1; Personalism

Translated from the French by Robert Stern
LAURENT FEDI

List of works

Renouvier, C.B. (1842) *Manuel de philosophie moderne (Manual of Modern Philosophy)*, Paris: Paulin.(A manual of philosophy, including a close study of Cartesianism.)

Renouvier, C.B. (1844) *Manuel de philosophie ancienne (Manual of Ancient Philosophy)*, Paris: Paulin.(A study of ancient philosophy.)

Renouvier, C.B. (1848) *Manuel républicain de l'Homme et du Citoyen (Republican Manual of Man and Citizen)*, Paris: Pagnerre; ed. J. Thomas, Paris: Colin, 1904.(A work on politics.)

Renouvier, C.B. (1854) *Essais de critique générale. Premier Essai (Essays of General Critique. First Essay)*, Paris: Ladrange; 2nd edn, expanded, *Traité de logique générale et de logique formelle (Treatise on General and Formal Logic)*, Paris: Bureau de la Critique philosophique, 1875; reprinted, Paris: Colin, 1912. (Renouvier's theory of knowledge.)

Renouvier, C.B. (1859) *Essais de critique générale. Deuxième Essai (Essays of General Critique. Second Essay)*, Paris: Ladrange; 2nd edition, expanded, *Traité de psychologie rationnelle (Treatise on Rational Psychology)*, Paris: Bureau de la Critique philosophique, 1875; reprinted, Paris: Colin, 1912.(Renouvier's theory of faculties of mind, certainty and human freedom.)

Renouvier, C.B. (1864a) *Essais de critique générale. Troisième Essai (Essays of General Critique. Third Essay)*, Paris: Ladrange; 2nd edn, modified, *Les principes de la nature (The Principles of Nature)*, Paris: Alcan, 1892; reprinted, Paris: Colin, 1912. (Renouvier's philosophy of nature.)

Renouvier, C.B. (1864b) *Essais de critique générale. Quatrième Essai (Essays of General Critique. Fourth Essay)*, Paris: Ladrange; 2nd edn, modified, *Introduction à la philosophie analytique de l'histoire (Introduction to the Analytic Philosophy of History)*, Paris: E. Leroux, 1896.(Renouvier's philosophy of history.)

Renouvier, C.B. (1869) *Science de la morale (Science of Ethics)*, Paris; Ladrange; 2nd edn, Paris: Alcan, 1908. (Renouvier's theory of ethics, a response to fundamental problems in political theory, concerning justice and democracy.)

Renouvier, C.B. (1876) *Uchronie*, Paris: Bureau de la Critique philosophique; recent edn, Paris: Fayard, 1988. (A fictitious reconstruction of European history.)

Renouvier, C.B. (1885-6) *Esquisse d'une classification systématique des doctrines philosophiques* (Sketch of a Systematic Classification of Philosophical Doctrines), Paris: Bureau de la Critique philosophique.(A history of philosophy based on logical alternatives, including Renouvier's intellectual autobiography.)

Renouvier, C.B. (1896-7) *Philosophie analytique de l'histoire* (Analytic Philosophy of History), Paris: E. Leroux.(A history of ideas, religions and philosophical systems.)

Renouvier, C.B. and Prat, L. (1899) *La Nouvelle Monadologie* (The New Monadology), Paris: Colin.(A theory of being along Leibnizian lines, but free of infinitism.)

Renouvier, C.B. (1901a) *Les dilemmes de la métaphysique pure* (The Dilemmas of Pure Metaphysics), Paris: Alcan.(A further history of philosophy based on dilemmas.)

Renouvier, C.B. (1901b) *Histoire et solution des problèmes métaphysiques* (History and the Solution of Metaphysical Problems), Paris: Alcan.(A history of major philosophical systems.)

Renouvier, C.B. (1903) *Le Personnalisme* (Personalism), Paris: Alcan.(A theory of creation, of divinity and of the existence of evil.)

Prat, L. (1904) *Les Derniers Entretiens de C. Renouvier* (The Last Conversations of C. Renouvier), Paris: Colin. (Renouvier's last conversation about his theories.)

Renouvier, C.B. (1906) *Critique de la doctrine de Kant* (Critique of Kant's Philosophy), Paris: Alcan.(A critical analysis of the *Critique of Pure Reason*.)

References and further reading

Foucher, L. (1927) *La jeunesse de Renouvier et sa première philosophie (1815-54)* (Renouvier's Youth and His First Philosophy (1815-54)), Paris: Vrin.(A study of the 'philosophie des Manuels'.)

Hamelin, O. (1927) *Le système de Renouvier* (Renouvier's System), ed. P. Mouy, Paris: Vrin.(A series of lectures given at the Sorbonne.)

Séailles, G. (1905) *La philosophie de Renouvier* (Renouvier's Philosophy), Paris: Alcan.(A critical study that has become a reference work.)

Verneaux, R. (1945) *L'idéalisme de Renouvier* (Renouvier's Idealism), Paris: Vrin.(A study of the problem of knowledge in Renouvier's philosophy.)

Representation, political

Political representation - the designation of a small group of politically active citizens to serve as representatives of the political community as a whole - is a central feature of contemporary states, especially of those that claim to be democratic. But what does it mean to say that one person or one group of people represents a larger group? Representatives are sometimes understood as agents of those they represent, sometimes symbolizing them, sometimes typifying their distinctive qualities or attitudes. Although political representation has something in common with each of these, it has its own special character. The missing idea here may be that the group represented authorizes *the representative to make decisions on its behalf. This still leaves open one crucial question, however: how far should political representatives remain answerable to those they represent, and how far should they have the freedom to act on their own judgment?*

1 The concept of representation

In what sense or senses (if any) can one person represent another? Apart from the philosophical problems involved, there is also a problem about the political relevance of any of the senses identified. This may be brought out by considering the approach of Anthony Birch. In three books, Birch has analysed the notion of representation and its political application. He identifies three senses of ‘representation’, in relation to the general issue of how one person may be thought to represent another. The first is manifest in the relationship between a principal and an agent. For example, a person may employ others, such as lawyers and accountants, to act on their behalf with respect to a particular facet of their affairs. The second is when one person (or sometimes an institution or even an item like a flag) may be thought of as representing persons in a symbolic sense. A flag may be seen as a symbol of a nation; or a sports team may be regarded as representing the country for which it competes. Third, a person may represent others by being in some way typical of them. A jury may represent a wider community because it reflects in some way the composition of the general public. Birch, however, holds that none of these three logically distinct senses of what it is to be a ‘representative’ is a good guide to the meaning of the term applied in a political setting. This consideration reverts to the question of ‘political relevance’.

On the face of it, it is not difficult to provide political analogies to Birch’s three senses of representation. A person who votes for a candidate in an election might be thought to empower that person to act on their behalf, reproducing the relationship between principal and agent. Just as persons might arrange for their affairs to be dealt with by an accountant or lawyer, so they might pass the expression of their political concerns to someone for whom they vote. Again, symbolic representation seems to have a clear application with respect to heads of state: a monarch or a president may be thought of as representing the nation (or state) of which they are the titular head. Finally, a political representative may have characteristics, such as class, domicile, attitude and policy preference, which are shared with those persons represented.

Nevertheless, Birch maintains that these three senses of representation do not capture the meaning of political representation. He claims that there is a fourth sense of representation, applied to members of a representative assembly, ‘because they have been appointed by a particular process of election to occupy that role’, which authorizes them to ‘exercise certain powers’ (1993: 74).

2 Representation, election and authorization

This perspective introduces two important considerations: the relation between representation and election, and the relation between authorization and representation.

Despite the historical importance of arguments about the role of elections in a system of representation, there is no straightforward logical connection between election and representation. This is clear from a review of the three central understandings of ‘representation’ identified above. In the relation between a principal and an agent, in the usual case, an agent is selected rather than elected. This is because one person is choosing another to pursue their interests or to look after their affairs. The powers that each has as a result of their agreement are primarily a matter of the contract between them. For example, someone might allow a lawyer to pursue a case for them as the lawyer saw fit; but another individual might reserve judgment on the suggestions made by the lawyer and thus make any ‘final’ decisions. The amount of discretion enjoyed by an agent varies according to the agreement with the

principal.

In political terms, there is a distinction between a representative conceived of as a delegate and a representative conceived less narrowly. Delegates are bound by a prior decision of those for whom they speak; a delegate is to carry out the wishes, already determined, of the principals. The distinction was explored by Edmund Burke in his famous speech in Bristol in 1774. He contrasted the role of the representative, who uses his best judgment about the issue in hand, with the role of the delegate, who was expected to carry out the instructions of those who had chosen him, whatever the 'delegate's' own opinion. The 'political relevance' of this notion of representation depends on the space for negotiation between principal and agent. While it is true that a representative might be thought of as an agent, nothing much follows about the relative powers of the parties until the relation between them is understood. On the one side, the representative could be seen as having *carte blanche*; on the other, as having no role other than to register the decisions of those represented. Clearly, the understanding of this relation will impinge on how the representative is selected. If the representative is a delegate, then the delegate might be chosen at random. If the representative is expected to use judgment, then a procedure enabling those represented to register their acceptance of that potential for judgment, such as election, might seem more appropriate. Some specifications of the principal-agent relation will support the appropriateness of election; others will not.

Symbolic representation is connected with elections in an even more contingent manner. Many political systems have an elected head of state, but these systems vary widely in the amount of executive power which that head enjoys. On the other hand, many political systems have non-elected heads of state: for example, a hereditary monarch. Whatever powers are enjoyed, and irrespective of election, the head of state's claim to represent the nation as a whole, symbolically, can be advanced.

The last notion of representation - as related to typicality - is obviously the one which allows the greatest space for political argument; chiefly, disputes about the political relevance of the representative's various characteristics, on the one side, and the political relevance of various distinctions between those represented on the other. For example, recent feminist arguments have advocated the importance of quotas to ensure that allegedly relevant distinctions are properly reproduced in the representative body. The contingency of the connection between election and representation is also clear: in any election, the individual voter may aim not to support a candidate who possesses similar characteristics to the elector.

3 The limits of representation

The question arises, therefore, as to whether Birch is correct in claiming that none of the three understandings of representation helps in giving meaning to political representation. As we have seen, it is possible to produce political examples of the three general interpretations; on the other hand, there are limits as to how far such examples illuminate the special circumstances of politics. Birch's preferred solution is to rely on the idea that electors authorize their representatives to exercise particular powers. This is to tie the theory of representation to a particular theory of representative democracy, which has been historically contested (see Democracy).

The concept of authorization can indeed illuminate some difficulties. For example, the election of a pope may be seen as authorizing his role in the apostolic succession, thereby legitimizing his claim to the papacy. There is nevertheless considerable scope for political argument about the nature of authorization: when, and in what circumstances, may it be withdrawn? Which voting system is best equipped to sustain claims that those who are successful have been authorized (or have a mandate)? Representative government has been promoted as a political system solving many problems: for James Mill, that of harnessing the interests of the governed to those of the decision makers; for J.S. Mill (1861), that of balancing expertise with participation; for the Founding Fathers of the American Republic, that of balancing virtue with improving popular sentiment, or perhaps, again, restraining power. Rousseau (§§2-3) objected to political representation because the general will of the people (which was the proper source of legislation) could not be represented. Whatever the (no doubt considerable) merits of representative government, there are likely to be tensions between any government's claim to representativeness and the policy preferences of those they claim to represent.

ANDREW REEVE

References and further reading

Birch, A.H. (1972) *Representation*, London: Macmillan.(A conceptual exploration rooted in the writings of

specific authors.)
Birch, A.H. (1993) *Concepts and Theories of Modern Democracy*, London: Routledge. (Relates representative government to the development of political science; especially useful for those without a background in political science.)
Burke, E. (1774) 'Speech at the Conclusion of the Poll, Bristol, 1774', vol. III, *The Works of Edmund Burke*, ed. F. Lawrence and W. King, London, 16 vols, 1815-27.(Eloquent argument defending the view that elected representatives should use their own judgment, rather than merely transmit the opinions of the represented.)
Dunn, J. (ed.) (1993) *Democracy - The Unfinished Journey 508 BC to AD 1993*, corrected edn, Cambridge: Cambridge University Press.(An excellent collection of essays with enormous historical sweep)
Lively, J. and Lively, A. (eds) (1994) *Democracy in Britain: A Reader*, Oxford: Blackwell and The British Council.(An anthology of debate, from both literary and political sources, about the development of representative institutions in the UK.)
Mill, J.S. (1861) *Considerations of Representative Government*, London: Dent, 1910. (Classic statement of the benefits of representative government when it combines expertise with participation.)
Phillips, A. (1995) *The Politics of Presence*, Oxford: Clarendon Press.(Most recent writing of an author who has been concerned with the relationship between democracy and feminism.)
Phillips Griffiths, A. (1960) 'How Can One Person Represent Another?', *The Aristotelian Society*, supplementary volume 34: 187-208.(A philosophical enquiry into the question posed.)
Pitkin, H. (1967) *The Concept of Representation*, Berkeley and Los Angeles, CA: University of California Press. (Another exploration of ideas about representation drawn from a criticism of theorizing presented by important writers in the historical canon.)

Reprobation

Reprobation is an eternal decision by God that results in everlasting death and punishment for some persons. The doctrine of reprobation typically takes one of three forms: (1) that God from eternity decreed to elect some without regard to faith or works and to reprobate others without regard to sin or unbelief, both to display his glory and for reasons we do not know (sometimes called double predestination); (2) that God from eternity decreed to elect some, despite their sin, and to abandon the rest, with the cause of their reprobation being sin and unbelief; or (3) that God from eternity elected those he foreknew would believe in Christ and reprobated those he foreknew would persist in sin and unbelief.

Reprobation doctrine was developed by Augustine and appears in the theology of Thomas Aquinas, Martin Luther and John Calvin, who were deeply indebted to Augustine's thought. Although some Lutheran and Roman Catholic theologians have defended reprobation doctrine since the sixteenth century, Reformed theologians have stressed it and made it the occasion of controversy.

1 Predestination, election and reprobation

Reprobation is closely linked to election and predestination. Election is an eternal divine decision to choose some people to be the recipients of special saving grace. Sometimes the term predestination is synonymous with election, and sometimes it encompasses both election and reprobation.

Since the Middle Ages, many theologians have identified two aspects of reprobation. Negative reprobation, or preterition, is God's decision to 'pass by', or withhold saving grace from, certain individuals, thereby permitting them to sin, to reject the gospel, and to undergo eternal punishment for their sin and unbelief. Positive reprobation is the divine decision or will to punish eternally those who sin and do not receive saving grace.

The doctrine of reprobation is rooted in the Bible's statements that God loved Jacob and hated Esau 'even before they had been born or had done anything good or bad' (Malachi 1: 2-3; Romans 9: 10-13), that God hardened Pharaoh's heart (Exodus 4: 21, 7: 3, 9: 12, 10: 27, 11: 10; Romans 9: 14-18), and that God, who is like a potter working with clay, has the right to make some individuals for destruction in order to demonstrate his wrath, power and glory (Romans 9: 19-24). The most troublesome biblical text for reprobation doctrine is the statement that God 'desires everyone to be saved and to come to knowledge of the truth' (1 Timothy 2: 4).

2 Augustine and the Middle Ages

In opposition to Pelagian teachings that God gives grace to some and punishes others according to their merits, Augustine of Hippo affirmed the existence of an eternal divine decision, not based on merit, to give faith and eternal life to some but not to others (see Augustine §§6, 10, 13). Predestination is God's preparation of kindnesses that certainly deliver the elect, while the rest are denied the means of believing and left to God's righteous judgment upon their sin.

Recognizing that speaking of some who are foreordained to damnation might make God seem unjust, Augustine argued that reprobation is deserved, while election in Christ occurs by grace, not by merit. Even if no human being had been redeemed, God's justice would have been untarnished. Moreover, by showing what the whole race deserved, the everlasting punishment of the reprobate impresses upon the elect the benefits they have received by grace. Basing his argument on the parable of the labourers in the vineyard (Matthew 20: 1-16), Augustine said that God is not unjust if he is just towards some and benevolent towards others. Still, God's ways are past finding out, and Augustine cannot say why God mercifully chooses one person and abandons another in his justice even though both could have been saved if God had so willed it. Nevertheless, the will of the omnipotent God is never evil, and is never defeated, 'because even when it inflicts evil it is just, and what is just is certainly not evil' (*Enchiridion*, 102). By interpreting the biblical statement that God desires all people to be saved to mean that God desires every *sort* of person to be saved, Augustine avoided saying that 'the omnipotent God has willed anything to be done which was not done' (*Enchiridion*, 103).

In the ninth century, Gottschalk of Orbais (*c*.804-69), who defended double predestination (the doctrine that God from eternity decreed to elect some without regard to faith or works and to reprobate others without regard to sin

or unbelief), was condemned by the council of Mainz (848) and by the first and second councils of Quiercy (849 and 853), which held that God predestines certain individuals to life only. The decisions of these councils were repudiated by the council of Valence (855), which upheld the notion of double predestination.

Thomas Aquinas said that God predestines some to eternal life through his providence, and reprobates others by *permitting* them to fall into sin and suffer its consequence. Reprobation is more than simply knowing that a person will sin and not reach eternal life; it also includes God’s *will* ‘to permit a person to fall into sin, and to impose the punishment of damnation on account of that sin’ (*Summa theologiae* Ia, q.23, a.3). So predestination (or election) causes both saving grace and eternal salvation, whereas reprobation causes abandonment by God and future eternal punishment, but not sin in the present life. Still, God’s will alone determines which individuals are elect and which are reprobate.

Aquinas responded to the objection that God is unjust because he treats equals unequally, electing some and reprobating others. First, God can justly deny something to someone to whom it is not owed. Since all are equally undeserving of salvation, God is free to give grace to one person but not to another. Second, God elects some and reprobates others, not on the basis of foreknowing their good and evil actions, but to display his goodness, which is one and undivided in himself, but must be manifested in many ways in creatures. God manifests his goodness by mercifully sparing the elect and justly punishing the reprobate.

Rejecting Aquinas’ position, Henry of Ghent argued that God elects those he foreknows will accept grace freely and reprobates those he foreknows will not. John Duns Scotus, drawing his doctrine of election from Aquinas and his doctrine of reprobation from Henry of Ghent, held that God eternally wills salvation for the elect apart from divine foreknowledge, but eternally wills damnation for the reprobate only on the basis of foreseeing their sin. William of Ockham (§10) came close to Henry of Ghent’s view of reprobation, although his view that there are no standards of justice independent of God’s policies led him to hold that God’s free choice determined that obedience leads to eternal life while ultimate and unrepentant sinning results in eternal punishment.

Thomas Bradwardine (§3) revived the notion of double predestination, saying that reprobation is a sovereign act of God’s will and is not based on foreknowledge of sin. In response to the objection that it would be unjust for God to predestine anyone to damnation apart from foreknown guilt, Bradwardine argued both that the reprobate serve the purpose of leading the elect to flee from sin and choose the good and that, since God’s will alone is the righteous law for all creation, no creaturely lump of clay can dispute the divine potter’s actions.

3 The Reformation and post-Reformation period

Martin Luther taught that God predestines some individuals to salvation and others to damnation. In his treatise The Bondage of the Will, Luther based his view on statements in Romans 9 that God loved Jacob and hated Esau prior to their birth and apart from any merits or demerits of their free actions, and also that the clay has no right to question the actions of the potter. He attacked human reason for holding that God, in order to qualify as good, ‘may damn none but those who in our judgment have deserved it’. Even if God should send all people to perdition, says Luther, he is nevertheless good. Still, God’s hardening or working evil in people does not involve ‘creating evil in us from scratch’ ([1525] 1972: 178). Rather, God allows a person’s evil will to proceed according to its own bent and confronts the person with words and actions with which this evil bent will clash. Departing from Luther’s view of unconditional reprobation, the Lutheran Formula of Concord (1577) says that predestination causes the salvation of the elect, whereas the non-elect perish because of their own sinfulness, not because of God’s will or foreordination.

John Calvin (§4) held that God, on the basis of his will alone and not on the basis of foreknown holiness or sinfulness, foreordains eternal life for some and eternal damnation for others. God’s sovereign election of some and reprobation of others reveals his glory by showing his justice as well as by impressing upon the elect God’s infinite mercy to them.

The Roman Catholic Council of Trent (1545-63) responded to Protestants’ renewed emphasis on election and reprobation by anathematizing anyone who says that the grace of justification extends only to those who are predestined to life and that others who are called do not receive grace because they are by divine power predestined to evil.

In the seventeenth century, a dispute arose in the Reformed churches, sparked by Jacob Arminius (1559-1609) and his followers, the Remonstrants. Arguing against the view of Calvin and others that God reprobates without giving any consideration to sin and unbelief, the Remonstrants held that God elects those he foreknows will respond to grace and reprobates those he foreknows to be unrepentant unbelievers. In response, the Reformed Synod of Dort (1618-19) declared that God's eternal decision is the reason he softens the hearts of some, giving them the gift of faith, and leaves others 'in their wickedness and hardness of heart' (*Canons*, I, art. 6). Election to salvation is based not on foreseen faith, but exclusively in God's good pleasure in adopting as his own certain persons 'from among the common mass of sinners' (*Canons*, I, arts 9-10). Reprobation, according to Dort, is God's free and irreproachable decision not to grant 'saving faith and the grace of conversion' to some, but rather to leave them in the misery into which, 'by their own fault, they have plunged themselves'; and finally to condemn and eternally punish them for their sin and unbelief 'in order to display his justice' (*Canons*, I, art. 15). By reflecting the medieval distinction between negative and positive reprobation, Dort followed a more moderate line than Calvin on the doctrine of reprobation. Dort defended God's justice by arguing that God would have done no one an injustice if he had condemned the entire human race on account of their sin and also by pointing out, on the basis of Romans 9: 20, that we human beings have no place finding fault with God.

4 The twentieth century

Karl Barth emphasized and recast the doctrines of election and reprobation. For Barth, Jesus Christ is both the elect one in whom all humanity is elect and the reprobate or rejected one. Arguing that morality disallows ascribing to God a justice that would lead to the reprobation or rejection of individuals, Barth said that God in love from eternity transfers to Jesus Christ his rejection of and wrath against all humanity. So in Jesus Christ, the rejected one who suffers the wrath of God, we see that rejection means being cast out from the presence of God and suffering eternal death.

The Reformed theologian Louis Berkhof (1874-1957) and the Thomist Réginald Garrigou-Lagrange held kindred views regarding reprobation. Both distinguished preterition, or negative reprobation, and condemnation, or positive reprobation, arguing that positive reprobation is based on foreseen sin, while negative reprobation is not. Although both held that there is no reason other than God's will for God's decision to predestine some to salvation and to pass by others, Garrigou-Lagrange added that the motive for negative reprobation is God's will to manifest his goodness by his justice as well as by his mercy, while Berkhof said that God alone knows his reason for preterition.

See also: Eternity; Grace; Heaven; Hell; Justification, religious; Omniscience §§3-4; Predestination; Providence; Salvation

RONALD J. FEENSTRA

References and further reading

Adams, M.M. (1987) *William Ockham*, Notre Dame, IN: University of Notre Dame Press, 2 vols.(Contains a very good chapter on election and reprobation in thirteenth- and fourteenth-century theology.)

Aquinas, T. (1256-9) *The Disputed Questions on Truth*, trans. R.W. Mulligan, Chicago, IL: Henry Regnery, 1952. (An early treatise by Aquinas that discusses reprobation.)

Aquinas, T. (1266-73) *Summa theologica*, trans. Fathers of the English Dominican Province, New York: Benziger, 1948.(Includes a classic statement of Aquinas' position on predestination and reprobation.)

Augustine (*c*.421) *Enchiridion*, in P. Schaff (ed.) *Nicene and Post-Nicene Fathers of the Christian Church*, series 1, 1886-9; reprint, Grand Rapids, MI: Eerdmans, 1978. (Includes Augustine's anti-Pelagian works.)

Barth, K. (1956-77) *Church Dogmatics*, ed. G.W. Bromiley and T.F. Torrance, Edinburgh: T. & T. Clark. (Election and reprobation are dealt with in volume 2, part 2.)

Berkhof, L. (1939) *Systematic Theology*, Grand Rapids, MI: Eerdmans, 4th edn. (Includes a classic Reformed definition and defence of the doctrine of reprobation.)

The Book of Concord: The Confessions of the Evangelical Lutheran Church (1580), trans. and ed. T.G. Tappert, Philadelphia, PA: Fortress Press, 1959.(Includes the Formula of Concord, discussed in §3.)

Bradwardine, T. (*c*.1344) *The Cause of God Against the Pelagians*, in H. Oberman (trans. and ed.) *Forerunners of the Reformation: The Shape of Late Medieval Thought Illustrated by Key Documents*, 1966; reprint, Philadelphia, PA: Fortress, 1981.(An important document by Bradwardine on predestination and reprobation.)

Calvin, J. (1552) *Concerning the Eternal Predestination of God*, trans. and ed. J.K.S. Reid, London: James Clarke

& Co, 1961.(An important treatise by Calvin on election and reprobation.)
Calvin, J. (1559) *Institutes of the Christian Religion*, The Library of Christian Classics, ed. J.T. McNeill, trans. F.L. Battles, Philadelphia, PA: Westminster, 1960.(A summary of Calvin’s theology, including his views on reprobation.)
Garrigou-Lagrange, R. (1939) *Predestination*, trans. B. Rose, St Louis, MO: Herder.(Includes a summary of Roman Catholic teaching on, and a restatement of, the doctrine of reprobation.)
Klooster, F.H. (1984) *Calvin’s Doctrine of Predestination*, Grand Rapids, MI: Baker, 2nd edn.(Includes a good summary of Calvin’s views on reprobation.)
Luther, M. (1525) *The Bondage of the Will*, in P.S. Watson (ed.), *Luther’s Works*, vol. 33, general eds J. Pelikan and H.T. Lehmann, St Louis, MO: Concordia and Philadelphia, PA: Fortress, 1972.(Contains Luther’s most extensive comments on reprobation.)
Muller, R.A. (1986) *Christ and the Decree: Christology and Predestination in Reformed Theology from Calvin to Perkins*, Durham, NC: Labyrinth Press.(An important study of predestination and reprobation in sixteenth-century Reformed theology.)
Schaff, P. (ed.) (1931) *The Creeds of Christendom*, revised D. Schaff, 6th edn; reprint, Grand Rapids, MI: Baker, 1983.(Contains the Canons and Decrees of the Council of Trent, the summary Epitome of the Lutheran Formula of Concord, and the Canons of the Synod of Dort.)
Sinnema, D.W. (1985) *The Issue of Reprobation at the Synod of Dort (1618-1619) in Light of the History of This Doctrine*, PhD dissertation, Toronto, Ont.: University of St Michael’s College.(A thorough study of the history of the doctrine of reprobation, with an emphasis on Reformed theology.)
William of Ockham (*c*.1320-3) *Predestination, God’s Foreknowledge, and Future Contingents*, trans. and ed. M.M. Adams and N. Kretzmann, Indianapolis, IN: Hackett.(Both the introduction and the translated material deal with reprobation.)

Reproduction and ethics

The first reproductive issue debated extensively by philosophers was abortion. Debates about its morality were, and still are, dominated by the issue of the moral status of the foetus, on which a wide variety of views has been defended. The most 'conservative' view is usually associated with very restrictive abortion policies, inconsistent with 'a woman's right to choose' (though the connection has been challenged by Judith Jarvis Thomson). However, all but the most conservative find it hard to ground prevailing moral intuitions concerning the newer issue of using human embryos for research purposes. Embryos, and even gametes, also assume importance in the context of methods for overcoming infertility (artificial insemination by donor (AID), egg and embryo donation involving in vitro fertilization (IVF), surrogacy) where issues about rights and ownership may arise. Considerations of 'the welfare of the child', often used to settle surrogacy disputes, also bear on questions of what should, or may, be done to avoid bringing a child with a genetic abnormality into the world. Current philosophical literature on reproductive issues is largely limited to a vocabulary of rights and little attention is paid to the social and familial contexts in which reproductive decisions are usually made

1 l The status of the foetus

A wide range of views on the moral status of the foetus, and whether it is the sort of thing that may, or may not, be killed has been defended. (1) According to the 'conservative' position, the foetus has, from the moment of conception, the same moral status as an adult human being. (2) At the other, 'liberal', extreme, the foetus is claimed to be nothing but a collection of cells, part of the pregnant woman's body, like her appendix, until the moment of birth. Oddly enough, these diametrically opposed views share two assumptions; both assume that the moral status of the foetus remains unchanged from conception to birth and both assume that the foetus is, morally speaking, like something else - an adult or an appendix. Three other views reject at least one of these assumptions. (3) A 'moderate' view claims that the moral status changes at some determinate 'cut off point' such as motility or viability. (4) According to the 'gradualist' view, the moral status changes gradually, increasing as the foetus develops. (5) According to the 'potentiality' view, the foetus has a unique moral status, being quite unlike anything else, that of a potential human being, from the moment of conception. Minor variations on these views exist, beyond the scope of this entry; however, a well-known sixth must be mentioned. (6) According to Michael Tooley and his followers, whatever the foetus is, it is not a person, that is, does not have a right to life; and the same is true of infants (see Rights; Moral standing §3).

Each of these views has been not only defended, but contested. None has an argument to establish what the moral status of the foetus is that its opponents regard as conclusive and each has difficulties concerning its *prima facie* consequences, which may be seized upon by opponents as unacceptable. So, for example, the potentiality view's 'difficulty' is that it hardly counts as yielding any consequences at all. The conservative view appears to yield the consequence that if you can save only a baby or a two-day-old embryo in vitro from certain death, you are faced with just the same moral dilemma as when you can save only one of two babies in cradles. Tooley's view, notoriously, licenses infanticide, and so on. The difficulties become even more acute when we come to consider the treatment of foetuses not *in utero*.

2 Embryonic and foetal research

The debate about the status of the foetus assumed prominence in the days when abortion came to be seen as an issue of women's rights, and this was well before we had acquired the technique of fertilizing extracted ova in vitro, or discovered that foetal brain tissue might help adult human beings suffering from Parkinson's disease, or started on the 'genome project' (see Genetics and ethics §1). The gradualist or moderate positions were, perhaps, gaining ground, as the views that underpinned the legislation governing abortion in many Western countries, wherein increasingly 'serious' reasons for abortion are required as the pregnancy develops but abortion 'on demand' is allowed in the first trimester, until the new questions about the treatment of first-trimester-age foetuses (or embryos) hit the headlines, and people started manifesting qualms. In fact, research on quite well-developed - even viable - foetuses had been going on, but few people knew about it, and, in many countries, there was no legislation that covered it. Now there is, but the substantial restrictions laid down seem to fit badly with the policies which are fairly 'liberal' about first trimester abortions. If the moral status of a ten-week-old foetus is so

minimal that abortion 'on demand' is morally permissible, why do we insist on laws restricting the use of even two- or three-day-old embryos for research? In particular, why is it always assumed (as it is) that embryos may only be used as a last resort (when, that is, no other animals can be used to further the research), a restriction that those concerned about our exploitation of other animals rightly point out calls for some justification (see Animals and ethics).

In fact, the debates about these issues tend to dodge questions about morality and centre instead around legislation where considerations of the general consequences of allowing or forbidding certain practices become obviously relevant. Hence *prima facie* inconsistent positions which combine liberal abortion legislation with very restrictive legislation on the use of embryos and foetuses may be defended on the grounds that, as things are at the moment, liberal abortion legislation is a necessary evil, the only available way of avoiding desperate women resorting to backstreet abortionists or suffering the emotional and economic hardship of having babies they did not want and could not afford. However, this is hardly a defence that conservatives about the moral status of the foetus can employ; nor does it tend to recommend itself to those who defend 'a woman's right to choose'.

3 The right over one's own body

The prevailing emphasis on the status of the foetus has an odd effect; one could, if ignorant of the facts of human reproduction, read hundreds of articles written on abortion and be left wondering what they were. Those who maintain that, at least in the early stages, the foetus is just a growth in the woman's body, fail to mention the fact that it is a growth like no other, namely a growth that, uniquely, results from the cells of two human beings, and is a growth that will usually become a baby, someone's child, if allowed to develop. Those who hold the 'conservative' or 'potential' view emphasize the fact that a fertilized ovum naturally develops into a baby; some mention the fact that an ovum is fertilized by a male cell and that this usually happens as a result of sexual intercourse, but remarkably few mention the fact that the nine-month development into a baby standardly - and arduously - takes place in a woman's body. If one did not know better, one might reasonably infer that parthenogenesis was common and that many of the results of sexual intercourse were raised in incubators as in Aldous Huxley's *Brave New World*.

Conservatives about not only the status of the foetus, but also abortion do not deny that we all have some sort of right over our own bodies; they merely claim that it is restricted, or outweighed, by the foetus' right to life. This initially plausible move has been challenged, however, in a deservedly famous article by Judith Jarvis Thomson (1971), which, despite its age, still stands almost alone in its attempt to take account of what is special about abortion, namely that it is the termination of a (human) pregnancy. A human pregnancy, regardless of one's views about the status of the foetus, is a condition of a human body, and usually results from sexual intercourse, voluntary or involuntary, in the hope, or not, of conceiving. Thomson daringly allows the conservatives their premise about the moral status of the foetus and argues that, even if this is granted, the impermissibility of abortion does not follow in many cases. Her argument depends on the claim that the right to life does not, as such, include the right to the use of another person's body to survive. If you can survive only by being connected to my circulatory system for nine months, then my right to decide what happens to my body allows me not only to refuse and let you die, but, moreover, to disconnect us and thereby kill you if I have not granted you the right to use my body but have been kidnapped and connected up to you while unconscious.

The wild unlikelihood of the latter scenario reflects Thomson's heroic attempt to describe something that is not a pregnancy but is like at least unintended pregnancies in the (assumed) relevant respects, namely that one person (the foetus, granting the conservative view) needs the use of another person's body to survive, while the second person has not done anything that can be construed as giving them the right to use it. The strained nature of Thomson's analogies has attracted much criticism, but few of her critics have paused to reflect that any analogies to pregnancy are bound to be strained because there simply is not any other condition of the human body remotely like it.

Her article manages to take cognizance of remarkably many of the unique features of pregnancy but still leaves several out. When we turn from the issue of abortion to those of surrogacy, artificial insemination by donor (AID) and in vitro fertilization (IVF), the fact that a fertilized ovum has resulted from the cells of two human beings, a man and a woman, and the fact that it would become a baby, someone's child, if enabled and allowed to develop *in utero*, assume unavoidable prominence. Moreover, questions about what is involved in 'the right to decide what

happens to my body' and, more generally, what counts as 'mine', become increasingly problematic.

4 Rights and ownership

AID, IVF and surrogacy, as moral and political issues, revolve around those people who very much want to have a child; hence the question of whether the foetus has the moral status of something that may be killed tends to fade into the background, and passionate feelings about parenthood, families and 'my (our) child' come to the foreground. Those who espouse the conservative view on the status of the foetus do indeed object to the current practice of IVF on the ground that it tends to involve producing 'spare' embryos, which must either be allowed to die or be frozen and stored; but the various methods of alleviating infertility involve many further problems.

Feminists have found themselves divided over what to say about surrogacy, inclined, on the one hand, to defend the view that a woman's right to decide what happens to her body surely extends to deciding whether or not to act as a surrogate mother but inclined, on the other, to liken surrogacy to female prostitution, regarding both as practices in which the woman's body is *used* in an exploitative and degrading manner, even if she has, in some sense, freely consented.

Anyone who is, for whatever reason, inclined to defend at least some forms of surrogacy, has to consider what should be done when the agreement between the parties concerned breaks down - when, say, the surrogate mother changes her mind, and wants to keep the child, or the commissioning couple change theirs and do not want to take it. Some aim to settle these unhappy questions by arguing that surrogacy is a form of 'pre-natal adoption', whereby the commissioning couple acquire parental rights (and duties) to the embryo as soon as fertilization takes place; the surrogate mother cannot abort it, nor claim it, nor can the commissioning couple refuse to take it without this counting as their immorally abandoning it. Others aim to settle them by considering who owned the sperm and ovum involved initially, who can be said to have 'donated' or given up their rights to these gametes and who has retained them, whether the surrogate mother acquires a right to the baby by (in a new sense of the phrase) 'mixing her labour with it', and hence who 'owns' the baby, as though gametes, foetuses and children were like any other possessions. But, like the two most extreme positions on the moral status of the foetus, each of these approaches takes cognizance of only some of the facts relevant to human reproduction, ignoring those emphasized by the other side and leaving some out entirely.

5 Welfare considerations

What about, in particular, 'the welfare of the child', the consideration which, in practice, has pre-eminently been appealed to in resolving disputes when surrogacy agreements have broken down? This undoubtedly goes beyond the debates about the various adults' rights but it is not, thereby, a consideration independent of the facts that are appealed to in those debates. For, it may be said, it is in the best interests of any child not only to be wanted and loved by two adults, but also for its mother's love to spring from the natural bond that exists between the child and the mother who carries it and gives birth to it and, even further, for it to have the opportunity to know, and, we hope, love, its genetic parents. But these considerations resolve a surrogacy-agreement breakdown adequately (if that) only in the particular case in which the surrogate child is the genetic offspring of the surrogate mother and her partner and they decide mutually they want to keep the child. Otherwise, the decision has to be over the circumstances in which the child will be least disadvantaged.

That surrogacy and, indeed, AID, egg and embryo donation are all methods of overcoming infertility that, arguably, lead to the production of a disadvantaged child probably forms the strongest basis for those who are morally opposed to them all. Common claims that they are all unnatural, or introduce a third party into what should be 'the exclusive relationship between wife and husband', thus undermining the family, tend to fall foul of the obvious counter that few things are more 'natural', or more affirmative of the value of family life, than a couple's desire to have a child. But no-one thinks that this natural, proper, and in some cases, quite consuming, desire can, morally, be satisfied by any means. You cannot steal a child in order to have one, and, it may be said, you cannot set about bringing a disadvantaged child into the world in order to have one either.

The extent to which different methods of overcoming infertility produce, or would produce, a disadvantaged child is usually thought to vary. As I write, many people in Britain have said that it would be a terrible thing for a child to know that its 'mother' was an aborted foetus and, on those grounds, supported legislation designed to forbid any

future use of the ova already present in female foetuses. It is important to remember in cases such as these that the very existence of the child whose welfare would be at stake depends on the decision taken. The choice here, for example, is not between existence with a foetus as mother and existence without, but between existence with a foetus as mother and nonexistence. However, egg donation by mature women, embryo donation and AID, when uncomplicated by surrogacy, often pass unquestioned - though the recent discovery that an unscrupulous doctor at an infertility clinic in America was the genetic father of hundreds of children, having used his own sperm to fertilize all his patients, gave some people pause for thought.

Considerations of what sort of life a child produced in certain circumstances will have may also form the basis of adverse moral judgments of people’s selfishness and irresponsibility. Many condemned a fifty-nine-year-old woman who chose to have a child by IVF (though the same judgment is rarely passed on even older men who father children), and some insist that people carrying certain genes should get themselves sterilized and resign themselves to childlessness or adoption. It is sometimes even said that it is selfish and irresponsible of pregnant women to reject screening for genetic abnormality or to reject abortion when it is identified; but whether this is so surely depends on their reasons for the rejections. If they think, for instance, that the genetic abnormality does not prevent one’s life being a good one (perhaps because they or their partner have it themselves), or if they think that abortion, at least in their circumstances, is wrong, then they have a good reason for not trying to ‘maximize happiness’ on this occasion (see Utilitarianism).

6 Morality, legislation and rights

At this point we return to the abortion debate, but now under a new aspect and in a way that brings questions about infanticide more clearly to the fore. The most fundamental objection to the conservative position on the status of the foetus is that it puts abortion, at any stage, on a par with infanticide, and, in the usual context of the abortion debate, infanticide equals murder; on the conservative view, abortion in the case of rape is not one whit more justifiable than murdering a child conceived through rape, and arguably less justifiable or excusable (given the innocence of the child) than the woman’s murdering the rapist. But in the context of euthanasia, killing an infant, or allowing it to die, without its consent, may well not be counted as murder, even on a conservative view; not because, quite generally, infants fail to be ‘persons’ with a right to life, who can thereby be killed for any old reason (as Tooley suggests), but because, when a human being is by virtue of extreme youth (or perhaps extreme impairment) incapable of autonomy, ‘paternalistic’ considerations of their welfare form, quite properly, the determining factor (see Life and death).

So even on the conservative view on the status of the foetus, abortion in the form of ‘foetal euthanasia’ may be justified, as infant euthanasia may be; and on any view that ranks the foetus as somehow not quite the same as a born baby, the possibilities of justification increase. But a new twist has been added; ‘euthanasia’ has its motivation built in, since it is done for the sake of the one who dies. Morally speaking, there is all the difference in the world between seeking an abortion because one wants a ‘designer baby’, and seeking it because one thinks it is wrong to bring a disadvantaged child into the world, just as there is a difference between my instructing the doctors to take my mother off life support for her sake, and my doing it to save myself the medical bills. But, given that people can and will lie about their motives, there is no way in which legislation can effectively permit the well-motivated cases but prohibit the callous ones.

Much of the literature devoted to reproductive ethics in fact vacillates between discussing morality and legislation, frequently leaving it unclear which is at issue. This is, no doubt, in part the result of the current tendency to talk almost exclusively in terms of rights, for we tend to think of (moral) rights as things that should be protected by good legislation. But we too readily forget that, particularly within families, it may be morally quite wrong for me to exercise a right that I certainly have, and, more generally, that there is much opportunity within families for acting morally well or ill where questions of rights, and even duties, do not arise. Most reproductive decisions are made by couples who love each other, who discuss what ‘their’ decision will be; the discussion is often extended to other members of the family who will say ‘we’ decided; most of the couples want to be good parents, and morally good people. This is, indeed, how things should mostly be, but none of it can be brought about by legislation, and all of it is almost universally ignored in the current literature (see Family, ethics and the §5).

See also: Applied ethics; Bioethics; Nursing ethics

ROSALIND HURSTHOUSE

References and further reading

Alpern, K.D. (1992) *The Ethics of Reproductive Technology*, Oxford: Oxford University Press.(A wide range of readings, not exclusively modern, with useful case studies.)

Dyson, A. and Harris, J. (1991) *Experiments on Embryos*, London: Routledge. (Articles by lawyers, scientists and theologians as well as philosophers; more appropriate for a British audience than a North American one.)

Feinberg, J. (1984) *The Problem of Abortion*, Belmont, CA: Wadsworth, 2nd edn. (Still the most comprehensive collection of articles on the topic, including Judith Jarvis Thomson's and Michael Tooley's, with a helpful introduction.)

O'Neill, O. and Ruddick, W. (1979) *Having Children, Philosophical and Legal Reflections on Parenthood*, New York: Oxford University Press.(Too early to contain any articles specifically concerned with surrogacy, but still the best collection of readings on parenthood.)

Thomson, J.J. (1971) 'Defense of Abortion', *Philosophy and Public Affairs* 1. (Despite its age, still the most original article on abortion available.)

Republicanism

Significant divisions exist in all societies and communities of any size. The expression of these divisions in politics takes many forms, one of them republican. The hallmark of republican politics is the subordination of different interests to the common weal, or what is in the interest of all citizens. To ensure this outcome, government in a republic can never be the exclusive preserve of one interest or social order; it must always be controlled jointly by representatives of all major groups in a society. The degree of control exercised by representatives of different social elements may not be equal, and different styles of government are compatible with republican objectives. However, all republican governments involve power-sharing in some way. Even in a democratic republic political majorities must share power with minorities for the common good to be realized.

Maintaining an appropriate balance of political power is the chief problem of republicans. One or another faction may obtain control of government and use it to further its own interests, instead of the common weal. To prevent this republicans have developed a variety of strategies. Some rely on constitutional 'checks and balances' to cure the mischief of factionalism. Others seek to minimize factionalization itself by regulating the causes of faction - for example, the distribution of land and other forms of property. Still others promote civic religions in order to bind diverse people together. All these methods accept the inevitability of conflicting interests, and see the need to accommodate them politically. Hence, civic life is at the heart of republicanism.

1 Political diversity of republicans

Republicanism is a remarkably protean sentiment. It embraces the most famous republic of all, Plato's *politeia*, wherein philosophers rule. It includes Cicero's Rome under Numa and Servius Tullius, statesmen of the highest rank. It encompasses Machiavelli's Florence under the financial dominance of the Medici family. Republicanism includes Harrington's Oceana, and the representative government of Thomas Paine, wherein the common sense of ordinary people holds sway. It even includes Jean-Jacques Rousseau's vision of a polity ruled by the *volonté générale*, or general will, not to mention most contemporary regimes.

In terms of their politics, then, republicans are a diverse lot: they stand to the left of absolutism, to the right of majoritarianism, and at every point between these extremes. The political diversity of republicans arises from different conceptions of the public good, the object of republicans in all times and places. Plato's *politeia* was meant to be good for all residents of the city, not just for those who ruled (see Plato §14). For that reason Cicero called it *res publica*, a Latin term referring to 'the public affair'. Cicero's own version of the just regime, *De re publica, de legibus* (54-51 BC), or *On the Commonwealth* as the work is known in English, likewise stressed the ways in which the different peoples of a vast empire might be served by a common legal order. It highlighted the sense in which persons differentiated by rank, privilege and duty nevertheless might be unified in spirit, law and political allegiance (see Cicero).

Republicans of all stripes are committed to a public order in which every citizen has a definite standing or place. Associated with this place in the community are specific responsibilities, for example, the defence of the homeland, as well as certain rights or legal protections for citizens and their property. As long as all citizens assume their proper place and perform their appointed role in civic life, the public good is served. Justice prevails, in the sense that each person receives their due, whatever that may be (and it need not be the same for all). Conversely, if citizens fail to act as they should, or when they step out of place and exercise prerogatives rightly belonging to others, injustice occurs. The public order is disturbed, and if it cannot be restored the sense of commonality will unravel and the republic will expire.

Although republicans are committed to justice, they are divided over its meaning. They also disagree about the most telling signs of injustice and the best means of preventing harm to the common weal. Such differences are reflected in preferences for various forms of government and strategies for insulating government from factional strife.

2 Mixed forms of government

Classical republicans had a strong affinity for mixed government, an ideal that dominated Western political thought after Plato. According to this way of thinking, all pure forms of government tend to degenerate. Monarchy

generally gives way to tyranny, aristocracy becomes oligarchy, and democracy declines into ochlocracy, or mob rule. In each case, the interests of a single order or faction prevail at the expense of others' interests, and at the expense of the common weal. Mixed government is a solution to the problem, in so far as it combines elements of all three forms of rule: the one, the few and the many, each influences mixed government. Since all interests must be taken into account in policy making; justice is the necessary outcome of this balance of power.

In a mixed government, the power of one, few and many need not be equal. Historically, monarchic republicans favoured a mixture that gave kings the upper hand, on the presumption that only a strong monarch was capable of constraining ambitious nobles, on the one hand, and intemperate masses bent on furthering their own interests, on the other. Thus, Cicero preferred ‘'that there should be a dominant and royal element in the commonwealth' ([54-51 bc] 1976: 151). The sixteenth-century historian Guicciardini was of the same opinion; his Bernardo mourned the passing reign of the Medici in Florence. And in England, the Reverend Philip Hunton, Provost of Durham University, defended what he called 'mixed monarchy' against those who would oppose Charles I in the Civil War.

Aristocratic republicans preferred mixed governments dominated by the few, as long as they exhibited proper republican virtues: courage in military affairs, skill in diplomacy and a commitment to order and justice at home. Where such aristocrats held the balance of power, they resisted tyrannically inclined monarchs - and unruly peoples. This was the version of republicanism favoured by Machiavelli (§§5-6), who saw it as the most stable form of polity (Machiavelli 1531). In England, aristocratic republicanism was the choice of Algernon Sidney, who was executed for his role in a plan to assassinate Charles II and James II, although his democratic inclinations were exaggerated for political purposes by opponents and proponents of monarchy alike.

For their part, democratic republicans sought a larger role for the many in defining the public good and pursuing it by political means. Publius Valerius (friend of the plebians in Rome), Savonarola (religious and political reformer in Florence) and James Harrington (utopian writer in England), all favoured republics in which the people or their representatives exercised control over policy makers. In fact, it was during Harrington's life that 'republican' and 'republicanism' first entered the English language as an appellation for ‘Commonwealth men' and anyone else who wanted to reduce the power of hereditary rulers in England. Few of these republicans were democrats in a modern sense, although most remained dedicated to mixed government with a strong popular element. Not until the late eighteenth and early nineteenth century did republicanism come to be associated exclusively with the cause of popular sovereignty and majoritarianism, and applied to large, commercially-oriented nations (as distinct from the urban polities of the Renaissance or ancient Rome and Greece).

3 Republicanism and majoritarianism

Opponents of popular sovereignty always assumed that democratic republicanism led straight to majoritarianism, a form of rule that was suspect on two grounds. First, majoritarianism was a form of class rule, and therefore unjust on its face; and second, it was rule by the class least suited for political rule - ordinary folk, who were apt to place the interest of commoners above the common interest of all. Radicals such as Thomas Paine (§1)seemed to confirm this suspicion when they rejected the very idea of mixed rule in favour of a government composed solely of representatives elected by 'the people'. Paine's vision of simple government, infused with common sense and beholden to the majority, became the rallying cry for opponents of monarchy and aristocracy in Europe. In the subsequent political struggle against hereditary rule a modern conception of republicanism took root and flourished.

Interestingly, democratic republicanism stopped short of outright majoritarianism. Even Paine admitted that the first principle of the republic is justice, which rules out 'any species of despotism over each other, or doing a thing, not right in itself, because a majority of them may have the strength of number sufficient to accomplish it' (1989: 233). In other words, in a republic there must exist a balance between the will of a majority and the interests of political minorities. Majorities should defer to the legitimate rights of minorities, who must nevertheless accept the right of majorities to determine the direction of communal affairs. Thus, majorities and minorities each have a recognized place in the political order of democratic republics, which are not only more egalitarian, but also much larger, than the city-states of an earlier era (see Democracy).

Whereas classical republicans sought a balance of more or less immutable social orders, Paine and other modern

republicans, such as the US statesman James Madison, wanted a balance of political interests that might change places, so to speak. The composition of political majorities is not fixed; they may decline into minorities, just as minorities may ascend to majority status over time. To preserve this fluidity it was necessary to prevent majorities from abusing minorities, and to do so without undermining popular sovereignty. In the USA the Federalists' plan to 'extend the republic' was meant to encompass a broader range of interests, inhibiting the formation of majorities of the moment, the most likely source of abuse in a popular regime. Should this kind of majority ever form, the constitutional system of 'checks and balances' was supposed to frustrate any tyrannical impulses that might appear (see Constitutionalism).

On the other hand, power in the US system could be exercised by extraordinary majorities, which were much less likely to behave tyrannically than majorities of the moment (since the formation of large majorities depends on moderation and a regard for a broad spectrum of opinion, without which the coalition will not grow to the necessary size). In other nations the same effect has been achieved under political arrangements that require the concurrence of important linguistic, ethnic or religious minorities for policies to be adopted and implemented by numerical majorities. The presumption is that larger, more comprehensive majorities have a better view of the common good than bare majorities, who in turn see the common good more clearly than minorities of any sort.

4 Maintaining the republic

The preservation of a balance of interests is the leitmotif of republicanism. The constitutional plans of the Federalists were preceded by Harrington's reliance on the Agrarian Law as a means of stabilizing relations between diverse interests in the republic. Before that, republicans emphasized the importance of civic virtue as the foundation of order and political stability. Civic virtue implied a willingness to place one's talents in service to the republic. For most citizens this meant temporary service in militias raised to repel invaders. For a select few it meant a longer stint of public service, perhaps as political leaders, or in some other important capacity. The performance of various kinds of public service was deemed patriotic, and in that sense it was self-sacrificing. But in another sense it was self-serving; by contributing to the maintenance of the republic, citizens defended their place in the republic and whatever liberty it conveyed upon them (see Citizenship §2).

For many classical republicans instilling a love of ordered liberty among citizens required something like a civil religion. Certainly this is what Plato had in mind when he endorsed the propagation of a 'noble lie' concerning the origin of human differences and their embodiment in the division of political labour outlined in *Republic* (*c*.380-367 BC). Other republics had comparable myths of origin or founding, the purpose of which was to reconcile people to their place in a social order. The idea of a 'social contract' performs the same function in modern republicanism, which rests on a 'civic culture' that acknowledges different interests even as it stresses the importance of balancing them by political means. An even stronger version of this sentiment may be found today in communitarian thought, which stresses the bonds that unite different members of a polity (see Community and communitarianism §2).

Perhaps that is why Marx and other radicals of the nineteenth century favoured democracy over republicanism. For them the primary division in a society is between classes, defined in terms of their relationship to the means of production. With the abolition of private property, classes were expected to disappear and the state would no longer be a site of class struggle. In that sense radical democrats looked forward to the elimination of divisions among citizens, and hence an end to politics. Their solution promised to remove the entire *raison d'être* of republicanism: the reconciliation or accommodation of powerful and persistent divisions of interest in human communities.

RUSSELL L. HANSON

References and further reading

Cicero, M.T. (54-51 BC) *De re publica, de legibus*, trans. and intro. G.H. Sabine and S.B. Smith, *On the Commonwealth*, Indianapolis, IN: Bobbs-Merrill, 1976.(Translation of Cicero's dialogue on republicanism in Rome.)

Fontana, B. (ed.) (1994) *The Invention of the Modern Republic*, Cambridge: Cambridge University Press. (Intellectual historians discuss the adaptation of republicanism in revolutionary America and France, and its subsequent emergence in large, commercial societies in western European nations; complements Rahe (1994).)

Harrington, J. (1656) *The Commonwealth of Oceana and A System of Politics*, ed. J.G.A. Pocock, New York: Cambridge University Press, 1992.(Description of a democratic republic based on the widespread ownership of land and frequent rotation of crops.)

Machiavelli, N. (1531) *The Discourses*, ed. B. Crick, Harmondsworth: Penguin, 1981.(The single best explication of classical republicanism in the context of mixed government.)

Paine, T. (1989) *Political Writings*, ed. B. Kuklick, Cambridge: Cambridge University Press.(A stirring statement of modern republicanism and the need for simple government.)

Plato (*c*.380-367 BC) *Republic*, trans., with notes and an interpretive essay, by A. Bloom, New York: Basic Books, 1968.(Socrates explains why the just city is ruled by philosopher kings, protected by a class of guardians, and economically sustained by farmers and craftsmen..)

Pocock, J.G.A. (1975) *The Machiavellian Moment: Florentine Political Thought and the Atlantic Tradition*, Princeton, NJ: Princeton University Press.(The best single-volume discussion of classical republican thought and its modern variants.)

Rahe, P. (1994) *Republics Ancient and Modern*, Chapel Hill, NC: University of North Carolina Press.(A compendious, and at times tendentious, history of republics and republicanism in the West; anticipates Fontana (1994).)

Skinner, Q. (1978) *The Foundations of Modern Political Thought*, Cambridge: Cambridge University Press, 2 vols.(A treatise on political thinking in Europe during the Renaissance and Reformation, with particular stress on the evolution of republicanism and notions of the state.)

Respect for persons

The idea that one should treat persons with due *respect is an important part of common sense morality, but opinions differ about when respect is called for, what it requires, and why. Respect for persons is also a central concept in many ethical theories. Some theories even hold respect for persons to be the foundation of all other moral duties and obligations.*

Respect *is distinguished commonly, on one side, from fear and submission, and on another, from admiration, liking and affection.* Respect for all persons as such *is distinguished normally from* esteem or special regard for persons of unusual merit. *Some philosophers identify respect with* agapē, *a special kind of love, but respect is perhaps most often regarded as a distinct attitude that should constrain and complement the promptings of love. Kant, for example, held that the requirements of respect and love are different, though compatible, and that both are dependent upon the more general and fundamental idea that humanity in every person is an end in itself.*

Other key issues in discussions of respect for persons include: what moral requirement, if any, there is to respect all persons; what the grounds, scope, and theoretical status are of that requirement; whether one can forfeit all claim to respect as a person; what 'respect for persons' demands with regard to specific problems, such as conflicts rooted in race and gender differences; and whether there is the same ground and obligation to respect oneself as to respect others.

1 Aspects and kinds of respect

The root idea of the word 'respect' is 'to look back' or 'to look again'. People speak of respecting a variety of things in addition to persons, for example, talents, achievements, character, laws, authorities, social positions, opinions, powerful forces, and even nature. A common thread seems to be the idea of 'paying heed' or 'giving proper attention' to the object of respect. Respect is generally an acknowledgement of the value or importance of something (or someone) from some perspective (presupposed in the context). It is often associated with awe, reverence, uncoerced willingness to obey or conform, or at least symbolic recognition of status, excellence or power. Depending on the situation, we can show respect *for persons* in a variety of ways: for example, by praising, giving tokens of honour, and accepting orders or advice, or merely by maintaining an appropriate social distance and refraining from expressions of contempt and arrogance. We may respect enemies without necessarily liking them, agreeing with their opinions, approving of their projects, or obeying them; but respecting them is incompatible with regarding them as utterly 'worthless' or 'insignificant'.

Is respect a feeling, a belief, a behavioural pattern, a disposition to act, or an attitude? Respect is often felt, for example, towards persons, standards, or institutions, but a mere feeling or sentiment would not count as respect without some corresponding disposition to *treat* the object in appropriate ways ('respectfully') and a *belief* that the object is worth such treatment. Respect is also more than merely respectful behaviour, for one can show respect deceptively when one does not feel or have it. In discussions of the morality of respect for persons, respect seems most often treated as an *attitude* that might be analysed as a combination of elements: beliefs, evaluative judgments, and policy commitments, as well as dispositions of behaviour and feeling, towards the person who is respected.

Respect for persons *as persons* is to be distinguished from respect for persons as professionals, officials, or members of more specific groups. We might, for example, respect an individual as a musician though not as person, if we thought that the person, though dishonest, is an extraordinarily talented and creative pianist. Even if we regard someone as both professionally inept and morally corrupt, we might, in a sense, respect the person as an official (such as a judge, a priest, or a prime minister) within an institution the authority of which we acknowledge. Here what is respected primarily is the office itself; but in other cases when we say we respect someone as, say, a judge, a priest, or a prime minister, we mean to imply that the person is a good official who fulfils the specified role well. Thus 'respecting a person as a -', where the blank is filled with a reference to a role or office, is ambiguous between, on the one hand, an attitude of due regard to the rights and privileges of the person's position and, on the other hand, an attitude appreciative of the excellence of the person's ability and performance in the position.

Although being a person is not the same as having a specific profession or public office, there is an ambiguity, similar to the one above, in the expression ‘respect for a person as a person’. In one sense, this indicates an appreciation that the person has good traits of character or personality, that is, traits that are generally considered desirable in persons whatever their particular roles and circumstances. For example, one might say of persons of known integrity, ‘They may be ineffective lawyers, but I respect them highly as persons’. In another sense, however, we might respect individuals as persons, even though we believe them to be immoral and lacking in most other human excellences. Here, ‘being a person’ is treated as if it were a widely inclusive, nonconventional status, analogous to holding a public office, with at least some minimum rights (or consideration due) independently of individual merit. To illustrate with an extreme and controversial example, some might say even of unrepentant Nazi war criminals that although their outrageous crimes deserve nothing but condemnation and they have forfeited their right to liberty, perhaps even life, we cannot simply discard them as worthless rubbish but must continue to respect them as persons (see Good, theories of the §3).

Darwall (1977) introduced the now common term, ‘appraisal respect’, for respect based on a positive assessment of the merits of individuals, whether as persons or as professionals. This is contrasted with ‘recognition respect’, which is a disposition to give appropriate weight in one’s deliberations to the fact that someone is a person (whether meritorious or not). It is obvious that we need not, and indeed cannot, respect every person as a person in the first sense, for in that sense respect implies a belief that the person in question has individual merits that deserve respect. What is controversial is whether we should respect all persons in the second sense - regardless of their individual merits - and if so, why.

When we say that persons should be respected *as persons*, the qualification ‘as persons’ may reflect, ambiguously and indefinitely, our view of the scope of the requirement, its grounds, how it is to be fulfilled, or several of these. That is, first, the point could be merely to say that *all* persons should be respected, without saying how or why. Alternatively, the suggestion could be, instead or in addition, that persons should be respected *because* they are persons. That is, the fact that people have the various features constitutive of persons might be seen as the pivotal premise in some argument, presupposed but not yet spelled out, to justify the claim that they should be respected. Finally, the point might be to give clues as to how respect is to be shown. Assuming a background of shared understanding about how various other things are treated appropriately, the injunction to respect persons as persons offers at least negative counsel: for example, persons should not be treated as mere expendable commodities, as instruments or playthings existing solely for our convenience or pleasure, as dogs, vermin, or dirt, or, more metaphorically, as mere statistics (as opposed to individuals with rights, feelings and particular interests). Again, assuming a common agreement on positive requirements for the appropriate treatment of persons, the injunction to treat persons as persons can serve rhetorically as a reminder to apply those standards to a particular case at hand. Here it says, in effect, ‘Treat these individuals in the manner that, as you know, is befitting their status as persons.’

Since commonly accepted standards for the appropriate treatment of persons and other things may be challenged, there is an obvious need for further thinking. Some ethical theories raise and respond to the following natural questions. Why must I respect all persons? What argument can justify the inference from ‘They are persons’ to ‘They should be treated in such-and-such a manner’? What, specifically, are reasonable, and not merely widely shared, standards for how persons should be treated (see Moral standing §2)?

2 Respect for persons in Kant’s ethics

The historical roots of most contemporary discussion of respect for persons are in the moral philosophy of Immanuel Kant (see Kant, I. §§9-11; Kantian ethics §1). Previous practice and theory often called for proper respect for individuals according to their social rank and individual merit, but Kant (1785, 1788), influenced by Rousseau, argued that all human beings have a dignity that is independent of rank and merit. All moral agents, by virtue of their rationality and autonomy of will, are jointly ‘authors’ of moral law, bearers of fundamental rights, and pursuers of ends that others may not ignore (see Autonomy, ethical; Moral agents). Moral duties are categorical imperatives, that is, rational requirements to which we are subject independently of the various personal ends that we choose to set for ourselves. According to Kant, there could be categorical imperatives only if there is something of absolute value; and only persons, as the source of all other values, could have this status. Persons exist as ‘ends in themselves’, of unconditional and incomparable worth, in contrast to things valuable

merely as means or as objects of affection. All rational persons, Kant thought, attribute this special value to themselves. Furthermore, he argued, they must acknowledge, on due reflection, that the ground for this self-attribution is not something unique to them but something they have in common with all other persons, namely, 'humanity' or the special human capacities identified with reason and freedom. Accordingly, a version of the categorical imperative, which is supposed to be a basic, comprehensive, and universal requirement of reason, is, 'Act so that you treat humanity, whether in your own person or in that of another, always as an end, and never as a means only.'

In *Die Metaphysik der Sitten (The Metaphysics of Morals)* (1797), Kant used this idea as a basis for both legally enforceable rights and ethical duties to respect others and oneself. In his theory of justice he implied that the status of persons as ends in themselves underlies all rights and, in particular, imposes limits on what legal punishments are permissible. In his theory of virtue Kant maintained that proper self-respect requires one to avoid drunkenness, gluttony, and servility and that respect for others is incompatible with arrogance, defamation, and ridicule. He also held that respect, together with love, is an essential element in friendship.

What is required morally, according to Kant, is to act with due respect, not to have or cultivate a mere 'feeling' of respect. This is partly because we cannot in general control our feelings by will in the way we can and must control our behaviour. Another reason is that respect for persons is derivative from respect for the moral law, and respect for the moral law, in Kant's view, is not something we choose to have or not but rather is something that, as human moral agents, we cannot help but feel. It is the humbling feeling that moral requirements reasonably impose limits on our attempts to satisfy our desires and pursue our personal ends. In so far as it is a form of respect for morality, then, the imperative to respect persons does not ask us to try to conjure up immediately, or even to cultivate over time, an 'affect' or sentiment that we might lack. Rather, what is required is that we choose to act so that, in practice, we live up to our own rational assessment of the worth (dignity) of humanity, a worth that all human moral agents (to some degree) recognize and feel.

3 Some later developments and applications

The idea of respect for persons has been interpreted in various ways, and has served different functions, in philosophical discussions before and after Kant. Contemporary moral philosophers often appeal to principles of respect in support of their judgments regarding practical issues, and some have attempted to develop ethical theories with respect for persons as the basic moral requirement. Downie and Telfer (1969), for example, treat the attitude of respect as morally basic. This attitude is characterized as 'valuing and cherishing persons for what they are'. It is identified as a kind of love, *agapē*, which requires 'active sympathy'. The concept of 'persons' itself is viewed as an evaluative concept, though not in a way that makes the injunction to respect persons trivial. What is morally salient about persons is that they have 'rational wills'; that is, they are 'self-determining and rule-following'. This is understood in a sense, taken to be different from Kant's, that includes affective and emotional capacities. Respect for persons requires that we make the ends of others our own and be ready at least to consider how others' rules apply to them and ourselves. This requirement is supposed to be consistent with utilitarianism, but not derived from it; to the contrary, Downie and Telfer (1969) have argued that utilitarianism is in fact derivable from a principle of respect for persons (see Utilitarianism).

Alan Donagan (1977) also presented respect for persons as the core of morality. He interpreted Kant's imperative to treat humanity as an end in itself as the principle, 'It is impermissible not to respect every human being, oneself or any other, as a rational creature.' This, he claimed, is the basic principle of those aspects of traditional Western morality not based on theological beliefs. Respect for persons 'as rational creatures', in Donagan's view, is a descriptive term, difficult to define, but familiar in social science and everyday life. Thus he thought it possible to spell out the meaning of the phrase in 'specificatory premises' which enable one to derive from the basic principle a system of intermediate moral principles regarding truth-telling, promise-keeping, non-injury, charity, family, property, obedience to law, and cultivation of one's own health and mental powers. Guided by prior work of Jewish and Christian casuists, he argued that the resulting principles would incorporate more flexibility for exceptional circumstances than Kant admitted but far less latitude than advocated by contemporary consequentialists (see Casuistry; Consequentialism). The system of principles, he claimed, is grounded in reason, not faith. Although he presented morality as a system of principles derived deductively from a first principle, he disclaimed the 'foundationalist' idea that the first principle is 'self-evident' or more certainly rational than the

intermediate principles (see Foundationalism).

Attempts to make respect for persons a comprehensive standard invite various sceptical responses (see, for example, Cranor 1975; Frankena 1986; Hill 1993). It is questionable, for example, whether the idea of respect is sufficiently substantive and descriptive to serve the role it has in Donagan's theory. Further, common usage suggests, and many philosophers agree, that respect is not at issue in all moral contexts, even though it is an important consideration in many. Rather than an all-encompassing first principle, it is one moral consideration along with many others (such as beneficence and gratitude) (see Moral pluralism). Treating respect for persons as the comprehensive source of all moral duties also raises serious questions about the grounds for decent treatment of nonrational animals and human beings with severe mental incapacities (see Animals and ethics). Also, most utilitarians would deny the primacy of respect for persons, and, to be consistent, all of them must object to principles of respect if they demand acts, rules, and motives that promote less than maximum utility.

Not all contemporary discussions of respect for persons, however, are focused on this controversy about the primacy and scope of respect principles. Many concentrate instead on the implications of respect for persons with regard to particular moral issues, such as racism and sexism, violent protest, pornography, punishment, privacy, legal enforcement of community values, and paternalism in medicine. Self-respect has also been a common theme. Most notably, Rawls (1971) has argued that his theory of justice is morally preferable to utilitarian theories because it better affirms and supports the self-respect of citizens. Sachs (1981) and others have noted significant distinctions between self-respect and self-esteem, and the idea that one should respect oneself is a common premise in recent moral arguments, for example, that no one should passively tolerate humiliation and oppression.

See also: Self-respect

THOMAS E. HILL, JR

References and further reading

Cranor, C. (1975) 'Toward a Theory of Respect for Persons', *American Philosophical Quarterly* 12: 309-19. (Reviews previous accounts, offers a complex analysis of respect, and argues that respect for persons alone cannot be an adequate basis for moral theory.)

Darwall, S. (1977) 'Two Kinds of Respect', *Ethics* 88: 36-49.(A discussion of the distinction between 'appraisal respect' and 'recognition respect'.)

Dillon, R. (1992) 'Respect and Care: Toward Moral Integration', *Canadian Journal of Philosophy* 22: 105-31.(A discussion of how moral considerations of respect for persons might be combined with a feminist ethics of care.)

Donagan, A. (1977) *The Theory of Morality*, Chicago, IL and London: University of Chicago Press.(A systematic attempt to derive moral duties from a basic and comprehensive principle of respect for persons.)

Downie, R.S. and Telfer, E. (1969) *Respect for Persons*, London: Allen & Unwin. (Develops the idea of the supreme worth of the individual person as the basis for a comprehensive requirement of respect.)

Frankena, W.E. (1986) 'The Ethics of Respect for Persons', *Philosophical Topics* 14: 149-67.(A clear critique of the project of grounding all of morality on respect for persons.)

Fried, C. (1978) *Right and Wrong*, Cambridge, MA: Harvard University Press. (Employs the idea of respect for persons in discussions of many particular moral problems.)

Green, O.H. (ed.) (1982) *Respect for Persons*, Tulane Studies in Philosophy, vol. 31, New Orleans, LA: Tulane University.(A collection of twelve short essays on respect for persons in relation to moral theory, self-respect, and various moral problems.)

Hill, T.E., Jr (1993) 'Donagan's Kant', *Ethics* 104: 22-52.(A comparative and critical discussion of Donagan and Kant on respect for persons and humanity as an end in itself.)

Hudson, S.D. (1980) 'The Nature of Respect', *Social Theory and Practice* 6: 69-90.(A detailed analysis of four kinds of respect.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(Kant's classic, but not easy, work on the foundations of ethics, important here especially for its discussion of humanity as an end in itself.)

Kant, I. (1788) *Critik der practischen Vernunft*, trans. L.W. Beck, *Critique of Practical Reason*, New York: Macmillan, 1965; 3rd edn, 1993.(Another important but difficult work on the foundations of morals, with an extended treatment of respect for moral law, which underlies respect for persons.)

Kant, I. (1797) *Die Metaphysik der Sitten*, trans M.J. Gregor, *The Metaphysics of Morals*, Cambridge: Cambridge University Press, 1991.(Kant's late work on the intermediate principles of morals, including quite readable sections on respect for others, respect in friendship, and, under 'duties to oneself', self-respect.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Now a classic of moral and political theory, this work argues that self-respect is a primary good better secured by Rawls' two principles of justice than by utilitarian principles.)

Sachs, D. (1981) 'How to Distinguish Self-Respect from Self-Esteem', *Philosophy and Public Affairs* 10 (4): 346-60.(A subtle account of self-respect and ways in which it differs from self-esteem.)

Responsibilities of scientists and intellectuals

Do scientists and intellectuals bear responsibilities peculiar to them? If an 'intellectual' is whoever has a committed interest in the truth or validity of ideas for their own sake and a 'scientist' anyone possessing a special competence in the natural or social sciences, they may indeed be more likely to find themselves in certain characteristic positions of responsibility. In the case of intellectuals, the importance of providing checkable justification of claims made in their pursuit of truth brings certain responsibilities. Scientists may be said to have responsibilities for pursuing truth in their own areas of competence, for wielding their social power appropriately, for making their results generally accessible and for using resources properly. But these apparently special responsibilities are nevertheless to be understood as rooted ultimately in those which any human being may, in the relevant circumstances, be thought to bear to their fellows.

1 The terms of debate

In what sense might scientists and intellectuals have responsibilities different from those of any other member of society (see Responsibility)? Clearly the question is not one of their strictly contractual responsibilities. Anyone employed to teach or to engage in research on the basis of a contract of service will legally have whatever responsibilities are specified in their contract. Of course, teachers, doctors, priests and plumbers among many others will be widely assumed to have responsibilities stemming from the roles that they occupy going far beyond any that may be set out in their contracts of employment. In that sense one might say that all members of society have an obligation to fulfil whatever responsibilities are generally taken to be incumbent upon them in virtue of the roles that they may occupy, whether specified in their contracts or not. But there are wide differences between different types of role in different types of society. In most contemporary societies, for example, nobody can qualify as a doctor or lawyer without passing certain well-defined tests. One can be expelled from these roles by the appropriate authorities acting in accordance with recognized procedures, or one may cease to occupy them of one's own accord. Moreover, lawyers, doctors and many others have produced for themselves fairly elaborate codes of ethics designed to make explicit many of their most recognizably important responsibilities, which does not mean that it always puts their proper line of conduct beyond all possible doubt (see Professional ethics). (Among lawyers' chief responsibilities are those of seeking always to do their best for their clients and of never knowingly misleading the court. What to do when these responsibilities clash is, very sensibly perhaps, left open to potentially divergent judgment.) Other social roles, however, may differ in a whole variety of ways. The institution of marriage, for instance, differs notoriously from one society to another (and even within some overall societies) not least in the degrees of freedom, if any, with which those concerned may or may not enter upon or renounce that state - as also in the diverse responsibilities which may be considered to go with it. The same is true of many other socially recognizable roles.

Problematic as the search for common agreement as to the responsibilities attendant upon such roles and institutions may be, there are at least generally clear criteria for determining whether or not someone actually is a doctor or lawyer, husband or wife. This is somewhat less clear in the case of so-called common-law husbands and wives, much less clear in the case of scientists and very much less so in that of intellectuals. Is the term 'scientist' to be reserved for those working as teachers or researchers in the area of the natural sciences? What exactly is the extent of that area? What level must they have achieved to be entitled to this appellation? What of those who once worked in that area, but who have moved to other fields or have simply retired? Nevertheless, if anyone asked me, as a university teacher, which of my colleagues were scientists, I should not expect many of them to disagree with my response. If asked how many were intellectuals, however, I should first need to settle on what exactly we might both understand by that expression.

For present purposes, then, we need to stipulate some guiding definitions. The term 'intellectual' I take to refer to anyone who takes a committed interest in the validity and truth of ideas for their own sake independently of their causal relationships to whatever other ends. The term 'scientist' I take to refer to anyone possessing a specialized competence in the natural or social sciences, leaving open questions as to which of the social studies at what stage of their development should be treated as sciences (in whatever sense of that term) and what level of competence should here be counted as sufficiently specialized. Thus defined, the classes of intellectuals and of scientists will, of course, show a substantial, if indeterminate, overlap. They are nevertheless by no means identical, for not only

will there be many nonscientific intellectuals, but many scientists will be interested in the truth and validity of their ideas not so much for their own sake as for that of what can be done with them.

However, in so far as scientists and intellectuals (thus defined) occupy social roles, they do not do so in such explicitly determinable manner as do doctors or barristers. Moreover, there are no constituted bodies who might properly take it upon themselves to draw up a code of ethics for them. If, then, we attribute to them certain special responsibilities, it must be in virtue of certain generally recognizable characteristics of their roles rather than of any formally defining constitution.

2 Intellectuals

Let us take first the case of intellectuals, whether they be scientists or not. Our definition is, of course, cast in terms that reflect the relative indeterminacy of the concept of an intellectual itself. There is much room for debate and disagreement about what might count as validity and truth in the assessment not just of statements, propositions or sentences, which is complicated enough, but even more so of something as indefinite as ‘ideas’. Still, such debates are themselves typical of intellectual concerns. And whatever its degree of indeterminacy on the margins, the overall difference is clear enough between those whose interest in ideas is primarily from the point of view, say, of their history or that of imaginative play, but for whom consideration of their validity or truth remains firmly in brackets, and those for whom that consideration is their first preoccupation. In other words, intellectuals, according to our definition, are interested not only in ideas, but also in questions concerning their possible justification.

Justification, however, is a fundamentally serious matter. If one genuinely wants to assure oneself whether a given set of ideas is justified or not, one cannot knowingly rest content with only half-thought-out approximations. One is, so to speak, responsible to oneself, to others and before the tribunal of whatever ideas are under consideration for the seriousness of one’s endeavour to probe the matter of their justification. There is nothing essentially secret about this responsibility to oneself. On the contrary, the first responsibilities, which in a sense underpin all others, are those reciprocally reflexive responsibilities generally incumbent on every speaker of a language to respect those overall norms of truth and reference on which the stability of its meanings depends. To that not insignificant extent all participants in meaningful discourse have an implicit share of responsibility for the validity, even indirectly the truth, of the meanings and concepts on which their discourse is based; and to that same extent, however surprising it may appear to them, all participants in discourse have at least an implicit foot within the terrain of intellectuals. Those who have come to occupy positions which may justify their being picked out as intellectuals, however, can justly be said to have a more special and explicit responsibility; they may rightly be called upon to answer to their communities of common communication for the truth or validity of what they may have to say. This is no light responsibility; for, as Vaclav Havel has shown so well in his essay on *The Power of the Powerless* (1985), a society which has lost the most fundamental capacities for ‘living in the truth’ is one which has exposed itself to the most fundamental forms of corruption and demoralization.

It should be noted, incidentally, that it is not only intellectuals as such who bear a special responsibility for truth and validity. It falls also upon teachers to a very special degree. (Of course, all participants in human discourse may, at times and often willy-nilly, also find themselves fulfilling some of the functions of a teacher; but again the more explicit responsibilities fall upon those who openly occupy such a role.) Naturally, many teachers will also be intellectuals, just as many intellectuals will be teachers; but again the two classes are not coextensive, and neither is wholly included in the other.

3 Scientists

To those (very many) scientists who are also intellectuals the same responsibilities may be ascribed as to any other intellectual as such. What other questions of responsibility arise? They may seem to fall in the main under three heads, none concerning scientists alone, but all perhaps presenting themselves to scientists with an often peculiar urgency.

First, scientists may be held responsible, by others but more particularly by the scientific community itself, not simply for respecting the truth about whatever their domain of research may be, but, more positively, for remaining uncompromisingly active in their pursuit of as yet undiscovered truth. Many would say that this respect for truth

must include a commitment to make the results of their research publicly available. Some scientists, indeed, would maintain that this is, strictly speaking, their only responsibility *qua* scientists.

Second, however, in the world as it now is, scientists, or at any rate certain among them, are in a position to furnish it with its most powerfully necessary and perhaps sometimes sufficient means of life and death - not to mention many of its other most powerful instruments for the ongoing transformation of our environment and of our lives together. (One may think of nuclear research and of its applications to nuclear weapons, or of research into the human genome; but there are many other examples (see Technology and ethics; Genetics and ethics §1).) Scientists may thus possess very real social power. Should they be held peculiarly responsible - as scientists - for the ways in which that power may be exercised?

Third, certain scientific research programmes may cost sums of money that are already very large in themselves, but, more to the point, quite enormous as a proportion of the money socially available. Of course, the results of such research may bring enormous more or less long-term returns - to society at large maybe, but also to more restricted sectional interests, perhaps even to the scientists concerned themselves. At all events, those seeking such investment, whether from public or from private sources, will be expected to justify their requests, and may naturally be held responsible for their proper seriousness as well as for the honest and non-wasteful use of the resources allocated to them. They may also find that their sponsors, whether governmental or commercial, will make it a condition of their support for research that its results should not be made public.

In principle, however, none of these issues and dilemmas really concerns scientists alone. Power is distributed widely in society in all sorts of complex and interconnected ways. In particular, other nonscientific intellectuals, teachers or researchers may also occupy positions of diverse forms and degrees of power. Indeed, as Michel Foucault persistently pointed out, knowledge and culture themselves constitute formidable resources for the exercise of power. Again, many nonscientific researchers also may seek and receive substantial financial support - not to mention all those other forms of cultural activity which can flourish only on the basis of such subsidy. Nor is it scientists alone who may be held responsible for continuing unremittingly in the search for as yet undiscovered truths. Historians furnish just one among a whole number of other possible examples.

There are also important connections, including potential trade-offs, between these three groups of issues. For example, the (presumed) importance of whatever one is seeking to discover as the result of any given research has to be measured against its likely cost. Again, the more money one has at one’s disposal in mounting a research programme, the greater one’s potential power for all sorts of purposes. Conversely, the stronger one’s initial position in the field, the greater one’s chances of obtaining substantial financial and other institutional support.

None of these responsibilities, of course, exists as pure indisputable fact. The questions of their exact nature and weight, of whom they should fall on and even of their very existence, are all open to controversy. In the end, however, it is fair to say that they all derive ultimately from the reciprocal moral responsibilities of human beings as such - of human beings who may find themselves in one role or another, endowed with one set of capacities or powers or another, but whose responsibilities in the exercise of those roles or powers derive fundamentally from the general responsibilities which human beings may be thought to have to one another in so far as they are able to think of themselves as moral beings at all (see Moral agents; Universalism in ethics).

See also: Animals and ethics; Risk

ALAN MONTEFIORE

References and further reading

Gordon, C. (ed.) (1980) *Power/Knowledge: selected interviews and other writings, 1972-1977*, New York: Pantheon.(Selection of interviews with, and excerpts from the major works of, Michel Foucault; illustrates his views on the relation of knowledge to power.)

Havel, V. (1985) *The Power of the Powerless*, London: Hutchinson.(A remarkable analysis of the pervasively demoralizing effects on society of a ruthless totalitarian insistence on lip service to the ‘truth’ of what all concerned know to be false.)

Kieffer, G.H. (1979) *Bioethics: A Textbook of Issues*, Reading, MA: Addison-Wesley. (A clear discussion of ethical issues arising in the biomedical sciences, with a final chapter on science and society and the moral responsibility of scientists.)

MacLean, I., Montefiore, A. and Winch, P. (eds) (1990) *The Political Responsibility of Intellectuals*, Cambridge: Cambridge University Press.(A collection of essays by writers from a number of different countries.)

Singer, P. (ed.) (1991) *A Companion to Ethics*, Oxford: Blackwell.(A good collection of papers covering abstract issues of ethical theory as well as a number of practical examples.)

Various, *Science and Engineering Ethics*.(A relatively new journal, which first appeared in 1995 and whose nature is well indicated by its title.)

Responsibility

To be responsible for something is to be answerable for it. We have prospective *responsibilities, things it is up to us to attend to: these may attach to particular roles (the responsibilities of, for instance, parents or doctors), or be responsibilities we have as moral agents, or as human beings. We have* retrospective *responsibilities, for what we have done or failed to do, for the effects of our actions or omissions. Such responsibilities are often (but not always) moral or legal responsibilities.*

The scope of our retrospective moral responsibilities is controversial. We are responsible for the intended results of our actions, but how far we are responsible for their foreseen effects, or for harms that we do not prevent when we could, depends on how we should define our prospective responsibilities, that is, on how far we should regard such foreseen effects, or such preventable harms, as our business. To say that I am responsible for some foreseen effect, or for a harm which I did not prevent, is to say that I should have attended to that effect or to that harm in deciding how to act; our retrospective responsibilities are partly determined by our prospective responsibilities.

I am responsible for something only if it is within my control. It is sometimes argued that I am therefore not responsible for that whose occurrence is a matter of luck; but it is not clear that we can or should try to make responsibility wholly independent of matters of luck.

We have responsibilities not merely as individuals, but also as members of organizations (organizations themselves have responsibilities in so far as they can be seen as agents). This raises the question of how far we are responsible for the actions of groups or organizations to which we belong.

1 Prospective and retrospective responsibility

Our concern is with responsibility-ascriptions of the form '*A* is [was] responsible for *x*'. Such locutions may ascribe merely causal responsibility ('the earthquake was responsible for the damage'); but our concern is with normative ascriptions which hold an agent answerable for something.

My *prospective* responsibilities are those I have before the event, those matters that it is up to me to attend to or take care of. These are often tied to specific roles: I may have responsibilities as a teacher, or parent, or doctor (see Professional ethics). But we also have responsibilities as friends, as citizens, perhaps even as human beings (to pay some attention to the interests of other human beings), or as inhabitants of this planet (to have some practical concern for its future) (see Help and beneficence; Future generations, obligations to; Environmental ethics). My prospective responsibilities are those I have in virtue of satisfying some description - 'teacher', 'parent', 'human being'; and we may disagree about just what responsibilities, if any, belong with particular descriptions.

My *retrospective* responsibilities are those I have after the event, for events or outcomes which can be ascribed to me as an agent. I am retrospectively responsible for what I do, or fail to do, in discharging my prospective responsibilities. A doctor who cures a patient is responsible for that cure; one whose treatment (or lack of it) causes death is responsible for that death. More generally, though, we are responsible for at least some of the results of our actions, most obviously for those we bring about intentionally (see Intention). If I kill *V* intentionally, I am responsible for *V*'s death (see §§2-3).

To hold *A* responsible for an event is not yet to say that *A* should be blamed for it, partly because praise, rather than blame, may be due; and partly because I can avoid blame for an untoward event by justifying my action (I killed *V* intentionally, but did so in self-defence) (see Praise and blame). To be responsible for *x* is to be answerable for it. I can be called to account for it: to explain, justify or admit my culpability for bringing it about or failing to prevent it. Retrospective responsibilities may also bring liabilities. If I am responsible for the harm that someone suffered, I may be liable to censure or punishment for causing that harm, to make compensation for it, or to apologize for it.

Such responsibilities, prospective and retrospective, flow from a conception of our moral duties and relationships, or are ascribed by a legal system. But they may be neither moral nor legal: a sports referee, for instance, has prospective and retrospective responsibilities, but they are not all a matter either of morality or of law.

Only responsible agents, those with the capacities necessary for accepting and discharging responsibilities, can be

held prospectively or retrospectively responsible. Just what those capacities are depends on the kind of responsibility at stake. Only moral agents can have moral responsibilities (see Moral agents; Free will); but the law specifies its own criteria of responsible agency, which might differ between criminal and civil law. However, in so far as criminal responsibility entails liability to punishment, which itself involves censure, we can hold that the criteria of criminal responsibility should match those of moral responsibility (see Crime and punishment §1).

A responsible agent is one who can be held prospectively or retrospectively responsible. A different use of 'responsible' concerns the agent's attitude to their responsibilities. A responsible person is someone who can be trusted to discharge their prospective responsibilities, and to accept their retrospective responsibilities; an irresponsible person is not one who has no responsibilities, but one who does not take their responsibilities seriously.

2 The scope of responsibility

One general question about retrospective moral responsibility concerns its scope: what are we responsible *for*? An answer to this question depends partly on an account of the scope of our prospective responsibilities.

I am most obviously responsible for the directly intended results of my actions. I make myself responsible for them by directing my action towards bringing them about. I am also responsible for at least many events which I am certain will ensue from my actions. An arsonist who realizes that the fire they start will kill someone is responsible for that death, whether or not they directly intend it (but see §3).

Those who deny any intrinsic moral difference between direct intention and foresight argue that I am as responsible for any certainly foreseen consequence of my action as I am for its directly intended results (see Double effect, principle of). Those who deny any intrinsic moral difference between 'acts' and 'omissions' also argue that I am as responsible for events which I could but do not prevent as I am for effects which I actively bring about. For whether I act with the intention of bringing *x* about, or simply knowing that x will ensue if I act thus; whether I actively bring *x* about, or do not do what I know would prevent *x*; in each case I know that I can determine, by my action or inaction, whether *x* occurs or not.

To deny any intrinsic moral difference between direct intention and foresight, or between acts and omissions, is to hold that I am retrospectively responsible for whatever was within my control; and *x* was within my control if I knew that it was within my power to determine, by my action or inaction, whether *x* occurred or not. More precisely, control as thus strictly defined is held to be a *sufficient* condition of responsibility. I am responsible at least for what I thus control. It is not a necessary condition of responsibility. I am responsible for harm I cause recklessly or negligently; but a reckless agent realizes only that their action might cause harm, while a negligent agent is unaware even of a risk that harm will ensue. A weaker 'control' requirement still holds here, however. I act recklessly or negligently only if I know, or could and should have realized, that my conduct might cause harm and that I could avoid that risk by acting differently. (Such retrospective responsibility depends on prospective responsibility. I act recklessly or negligently as to some harm only if I should have taken care with regard to it.)

But is control sufficient for responsibility? To argue that it is, that I am retrospectively responsible for all that lay within my control, is to take a particular moral view of our prospective responsibilities as moral agents. If we believe that agents should attend to all the foreseen effects of their projected actions (in particular that they should see the fact that some harm will ensue from an action as a reason against acting thus), we will hold them responsible for all such foreseen effects; if we believe that agents have a general, stringent duty of care towards all others, to save them from harm, we will hold them responsible for any harms which they knew they could have prevented. But such beliefs are morally controversial. If I believe in the free market, I might deny responsibility for my competitor's ruin, which I knew would ensue from my commercial activities. Such an outcome of my legitimate commercial activity, I may say, is not my concern; it is not something I should attend to as a reason against acting thus. But a critic of the free market might insist that I am responsible for my competitor's ruin; that I can avoid blame for it only if I can justify my action of ruining them (see Market, ethics of the). Likewise, while I may recognize some general responsibility towards others, simply in virtue of our common humanity, to help them if they are in grave distress, I might insist that that responsibility is limited. If I fail to save the lives of distant strangers, my responsibility for their deaths is less than it would be if I made myself fully responsible for their deaths by deliberately killing them; if I do not prevent some trivial harm befalling another, I may deny any

(retrospective) responsibility for that harm, by denying any (prospective) responsibility to take such care for their interests. So to determine the scope of our retrospective responsibilities, we must determine the scope of our prospective responsibilities; where moral responsibility is at stake, that must be a moral determination.

Our particular roles, which help determine our prospective responsibilities, thus also help determine our retrospective responsibilities, most obviously by extending them. I could, but do not, stop a child committing a minor act of vandalism. If the child is a stranger, I might deny responsibility for the damage - it was not my business to control the child; but if the child is in my care, I cannot thus deny responsibility. Our roles may also, however, limit our responsibilities. I know that if I fail this thesis, the candidate will lose their academic job. But I might insist that my responsibility as an examiner is to judge the thesis on its merits; the further effects of my decision are not my concern (the candidate's prospective dismissal is not something I should attend to as a reason against failing the thesis), and I am not responsible for them.

Conceptions of prospective and retrospective responsibility help to determine our very descriptions of people's actions or omissions. As a free marketeer I might deny, but others might insist, that I 'ruined my competitor', since whether that is an appropriate description depends on whether I am responsible for my competitor's ruin. Likewise, we can say that someone 'let *x* happen', or 'failed to do *y*', only if there was some expectation that they would or should prevent *x* or do *y*.

3 Responsibility and luck

The connection noted above between responsibility and control underpins the idea that moral responsibility is incompatible with luck. How can I be morally responsible for some harm if its occurrence was a matter of luck, and thus not within my control?

But this idea threatens to undermine my responsibility for the actual (intended or foreseen) effects of my actions. For the occurrence of such effects depends on factors outside my control, and is thus arguably a matter of luck. The success of an attempt depends not merely on the agent, but on external factors which they do not control; if it fails because of one of those factors, that failure is a matter of luck. Similarly, it is a matter of luck whether a reckless or negligent action causes harm. How then can I be held morally responsible for the actual effects of my actions, if their occurrence is a matter of luck?

Those who argue in this way do not typically argue that the intrusion of luck precludes every kind of responsibility: I am responsible for the actual harm that I cause in that I may be liable to pay compensation for it. But luck does, they insist, preclude responsibility as *blameworthiness*. I am blameworthy only for what is strictly within my control, and not for that whose occurrence depends on luck. One who attempts to do harm is blameworthy for the attempt, but their culpability is not affected by the chance fact of success or failure; those who act recklessly or negligently are culpable for their reckless or negligent acts, but are no more culpable if they actually cause harm than if they do not.

To assess this argument, we must attend to three matters. First, it seems to misuse the contrastive concepts of control and luck. We do not normally think that we lack control over all the effects of our actions, or that those effects always depend on chance. One confusion here is to think that if the occurrence of x depends on any factors outside my control, I do not (really) control it; another is to move from the true claim that my control over anything is not invulnerable (I could lose it) to the false conclusion that I therefore do not (really) have control.

Second, it might be argued that our ordinary talk of control and luck is loose or metaphysically shallow. There is a deeper sense in which all that we do is vulnerable to chance, which captures a deeper truth about our lack of (real) control over the external world. But this metaphysical thought (if it is coherent) threatens any idea of control. If for this reason I do not (really) control the effects of my actions, I do not control my actions or intentions either. For these too are vulnerable to chance, and depend on factors outside my control.

Third, we must none the less admit that the actual outcome of an action is sometimes a matter of luck. It might be a matter of luck that an attempt fails, or succeeds; that a reckless action causes harm, or does not. Should we then, in ascribing moral blame, ignore such matters of luck; draw no distinction between the successful and the failed assailant, or between lucky and unlucky reckless agents? This suggestion, however, requires us to draw a sharp distinction between moral blame, or judgments of moral culpability, and those other aspects of our understanding

of and responses to a person's action which are structured by its actual effects; it is not clear whether that distinction can be drawn (see Moral luck).

4 Collective responsibility

We must also attend to collective, or shared, responsibilities: the responsibilities of organizations or groups; the responsibilities individuals have as members of organizations; and whether such membership can make us responsible for the actions of others.

We can talk of the responsibilities of organizations such as corporations, universities, governments and nations in so far as we can see such collectives as agents - as capable of acting purposively, as being potentially answerable for what they do or fail to do. The actions of such organizations involve, of course, the actions of individual agents within them; but that is not to say that descriptions of corporate actions can be reductively analysed without loss into descriptions of individual actions of individual agents (see Holism and individualism in history and social science).

Our prospective and retrospective responsibilities, as individuals, are often defined partly by our roles within organizations: I may have responsibilities, for instance, as a teacher, as an employee, as a director, as a minister. Such responsibilities are defined initially by the organization itself, in the light of its purposes and values. But there is obviously room for controversy (often moral controversy) about what our role-responsibilities ought to be, and about when and whether they may be overridden by moral responsibilities which transcend our particular roles.

In so far as I have responsibilities within an organization, I share, with other members of the organization, responsibility for what the organization does. But is that responsibility limited to the part which I play directly, or could and should play, in the organization's actions? Or can I properly be held responsible for the actions of others in the organization even though I did not myself have authority or control over them? Can I be held responsible for actions of the organization to which I did not contribute directly, and which I could not control? If the government of my country behaves despicably, can I disavow responsibility by pointing out that I voted against that party or that measure; or must I still see that action as being in some sense 'mine', as a citizen of this country? Indeed, can or should I accept responsibility for my country's actions before I was born?

Answers to such questions depend on getting clear about what it is to be a member of a group, and about what it is to be responsible. It would, for instance, be absurd to blame myself for governmental actions over which I had no control, or which were done before I was born; but it is not obviously absurd to feel shame for them, or think that it is up to me (though not of course up only to me) to apologize for them, or to try to find some way of making up for them.

See also: Action; Confucian Philosophy, Chinese §5; Legal concepts

R.A. DUFF

References and further reading

Ashworth, A.J. (1987) 'Belief, Intent and Criminal Liability', in J. Eekelaar and J. Bell (eds) *Oxford Essays in Jurisprudence*, 3rd series, Oxford: Oxford University Press, 1-31.(An account of retrospective criminal responsibility which tries to eliminate the role of luck.)

Casey, J. (1971) 'Actions and Consequences', in J. Casey (ed.) *Morality and Moral Reasoning*, London: Methuen, 155-205.(A sophisticated discussion of the ways in which descriptions of actions and omissions depend on ascriptions of prospective responsibility.)

Duff, R.A. (1990) *Intention, Agency and Criminal Liability*, Oxford: Blackwell, chaps 4-5.(Discusses competing conceptions of our retrospective responsibility for the foreseen effects of our actions.)

Hart, H.L.A. (1968) *Punishment and Responsibility*, Oxford: Oxford University Press.(See especially 'Intention and Punishment', on attempted crimes and on intention and foresight; and 'Postscript: Responsibility and Retribution', on different aspects of responsibility.)

Lucas, J.R. (1993) *Responsibility*, Oxford: Oxford University Press.(Analyses responsibility as answerability; discusses omissions, the responsibilities attaching to roles, and collective responsibility.)

Nagel, T. (1976) 'Moral Luck', *Proceedings of the Aristotelian Society*, supplementary vol. 50: 137-51; repr. in *Mortal Questions*, Cambridge: Cambridge University Press, 1979, 24-38.(A seminal discussion of the ways in

which our actions are vulnerable to luck, and their implications for responsibility.)

Wells, C. (1993) *Corporations and Criminal Responsibility*, Oxford: Oxford University Press.(Argues that corporations can and should be seen as criminally responsible agents.)

Williams, B. (1976) 'Moral Luck', *Proceedings of the Aristotelian Society*, supplementary vol. 50: 115-36; repr. in *Moral Luck: Philosophical Papers 1973-80*, Cambridge: Cambridge University Press, 1981, 20-39.(Another key discussion of the significance of luck for responsibility.)

Williams, B. (1993) *Shame and Necessity*, Berkeley, CA: University of California Press, ch. 3.(An account of the central elements of retrospective responsibility, comparing our conception with those of the ancient Greeks.)

Zimmerman, M.J. (1988) *An Essay on Moral Responsibility*, Totowa, NJ: Rowman & Littlefield.(An exhaustive analysis of retrospective moral responsibility, covering most of the topics discussed in this entry; has a useful bibliography.)

Resurrection

*The Judaeo-Christian belief in a future general resurrection of the dead arose in late second-temple Judaism (see, for example, Daniel 12: 2 and John 11: 24). (Whether there would be a resurrection of the dead was one of the main points that divided the Pharisees and the Sadducees.) When the new Christian movement appeared - before it was clearly something other than a party or sect within Judaism - it centred on the belief that the crucified Jesus of Nazareth had been, in a literal, bodily sense, raised from the dead (*resurrectus*) and that his resurrection was, in some way, the means by which the expected general resurrection of the dead would be accomplished. Indeed, resurrection was so pervasive a theme in early Christian preaching that it was apparently sometimes thought that Christians worshipped two gods called 'Jesus' and 'Resurrection' (*Anastasis*). The early Christians generally said that 'God raised Jesus from the dead'. In post-New Testament times, it became more common for Christians to say that 'Jesus rose from the dead'. Belief in the resurrection of Jesus and a future general resurrection continue to be central to Christianity. Christians have always insisted that resurrection is not a mere restoration of what the resurrected person had before death (as in the story in the fourth Gospel of the raising of Lazarus) but is rather a doorway into a new kind of life. The status of a belief in the general resurrection in rabbinic Judaism is difficult to summarize. It should be noted, however, that a belief in the resurrection of the dead is one of Maimonides' 'thirteen principles', which some Jews regard as a summary of the essential doctrines of Judaism. A belief in a general resurrection of the dead is one of many Judaeo-Christian elements that have been incorporated into Islam.*

1 The concept of resurrection

The concept of the resurrection of the body (or of the dead) is most easily explained by laying out the ways in which it differs from the most important competing picture of the survival of death, the Platonic picture. According to Plato (§13), when one dies (that is, when one's body dies), one will continue to be what one has been all along, a soul: an immaterial centre of consciousness, reason and action. One's death is, therefore, an extrinsic change in one: being dead means simply no longer having a body to animate. Since one's death is an extrinsic change in one, one's survival of death is something that happens in the natural course of events: one continues to exist after death by the continued exercise of the same powers or capacities that enabled one to exist when one still animated a body. (This inference is natural and plausible, but, as Descartes would later point out, it is not logically valid: for all logic can tell us, animating a body might be essential to the existence of a soul.) Death is, moreover, not a bad thing, as the vulgar believe, but a liberation, for the body is a prison of the soul - or it might be likened to a millstone that drags the soul down into the world of flux and impermanence. The liberation of the soul by death will not, unfortunately, be permanent, for the soul is destined repeatedly to suffer the misfortune of embodiment.

Christians, Jews and Muslims who believe in the resurrection of the dead will accept two of Plato's theses about death: that the person does survive death, and that dead persons will not be forever disembodied. But everyone who believes in resurrection will dispute the following elements of Plato's metaphysic of body, soul and death: that the body is a prison; that the soul must by its very nature survive the death of the body; and that the embodied soul has been disembodied in the past and will experience a large, perhaps infinite, number of 'reincarnations' in the future. Christians, moreover, will insist that the new bodily life that awaits the soul (the saved soul, at least; perhaps this is not true of the damned) will not be of the same sort as its earlier life. The doctrine of resurrection, however, is no more than a doctrine. It is not a worked-out metaphysic of body, soul and death. (The primary biblical data concerning the metaphysics of resurrection are found in 1 Corinthians 15: 35-55. This passage, however, is open to a variety of interpretations.)

There are several competing philosophical theories of the metaphysics of resurrection. Some who accept the doctrine of resurrection deny the existence of a separable, immaterial soul. Examples include Tertullian, who argued in his *De anima* (*c*.210-13) that the soul is corporeal, and, in the twentieth century, the Scottish computer scientist D.M. MacKay (1987) and the English physicist J.C. Polkinghorne (1994). Others accept the existence of an immaterial soul, but differ on the question whether the person, the 'I', *is* the immaterial soul. Aquinas (§10), for example, sees the human person as essentially a composite of a human soul and a human body. According to the 'composite' theory, a person cannot exist without a body: to exist is for one's soul (always numerically the same) to animate some human body or other. (In the interval between one's death and one's receiving a new body at the time of the general resurrection, one's soul exists and thinks and has experiences, but *one* does not, strictly

speaking, exist.) However, others who believe in a separable soul - most of the Fathers of the Church, and, probably, most Christians who have not given the matter much thought - accept a metaphysic of soul and body that is deceptively similar to Plato's: one *is* an immaterial soul, and one will exist and think and have experiences throughout the interval during which one is without a body. But even the members of this party - the theologically well instructed among them, at any rate - would accept the following anti-Platonic theses: that the death of one's 'first' body is not a natural consequence of the impermanence of material things, but is rather a result of the Fall; that the soul's survival of death is not a natural consequence of its immateriality or simplicity, but is rather a miracle, a special gift from God; that existing without a body is not a good thing for the soul, an essential part of the *telos* of which is to animate a body; and that the life of the 'spiritual' or 'glorified' body that the saved soul will be given at the general resurrection will be qualitatively different from (and superior to) the life of the soul's first or 'natural' body. (It must be emphasized that, whatever 'spiritual body' may mean, it does not mean 'immaterial body'.)

2 Philosophical difficulties

Each of these metaphysical theories of resurrection faces philosophical problems. Believers in resurrection who are dualists face the problems any dualist faces (see Dualism). Since these problems are the same whether or not the dualist believes in resurrection, they will not be discussed here. Believers in resurrection who are materialists (as regards human beings) face the problems any materialist faces (see Materialism in the philosophy of mind). Since these problems are the same whether or not the materialist believes in resurrection, they will not be discussed here. In addition, however, believers in resurrection who are materialists face a special philosophical problem about personal identity. The remainder of this entry will discuss this special problem.

It can be plausibly argued that the doctrine of the resurrection of the dead presupposes some form of dualism. For if human persons are not immaterial souls, if they are living animals, then it would seem that death must be the end of them. A living animal is a material object. A material object is composed, at any given moment, of certain atoms. But if one is composed of certain atoms today, it is clear from what we know about the metabolisms of living things that one was not composed of those same atoms a year ago: one must then have been composed of a set of atoms that hardly overlaps the set of atoms that composes one today - and so for any living organism. This fact, the fact that the atoms of which a living organism is composed are in continuous flux, is a stumbling block for the materialist who believes in resurrection.

Suppose, then, that God proposes to raise Socrates from the dead. How shall he accomplish this? How shall even omnipotence bring back a particular person who lived long ago and has returned to the dust? - whose former atoms have been, for millennia, spread pretty evenly throughout the biosphere? This question does not confront the dualist, who will say either that there is no need to bring Socrates back (because, so to speak, Socrates has never left), or else that Socrates can be brought back simply by providing his soul (which still exists) with a newly created human body. But what will the materialist say? From the point of view of the materialist, it looks as if asking God to bring Socrates back is like asking him to bring back the snows of yesteryear or the light of other days. For what can even omnipotence do but reassemble? What else is there to do? And reassembly is not enough, for Socrates was composed of different atoms at different times. If someone says, 'If God now reassembles the atoms that composed Socrates at the moment of his death, those reassembled atoms will once more compose Socrates', there is an obvious objection to the thesis. If God can reassemble the atoms that composed Socrates at the moment of his death in 399 BC - and no doubt he can - he can also reassemble the atoms that composed Socrates at some particular instant in 409 BC. In fact, if there is no overlap between the two sets of atoms, God could do both of these things, and set the two resulting men side by side. And which would be Socrates? Neither or both, it would seem, and, since not both, neither.

It might be objected that God would not do such a frivolous thing, and this may indeed be so. Nevertheless, if God were to reassemble either set of atoms, the resulting man would be who he was, and it is absurd, it is utterly incoherent, to suppose that his identity could depend on what might happen to some atoms other than the atoms that composed him (for this is what a materialist who holds that the reassembled '399 BC' atoms compose Socrates just so long as the '409 BC' atoms are not also reassembled is committed to). In the end, there would seem to be no way round the following requirement: if Socrates was a material thing, a living organism, then, if a man who lives at some time after Socrates' death and physical dissolution is to be *Socrates*, there will have to be some sort of

material and causal continuity between the matter that composed Socrates at the moment of his death and the matter that at any time composes that man. (St Paul seems to suggest, in the passage from 1 Corinthians cited above, that this will indeed be the case.) But 'physical dissolution' and 'material and causal continuity' are hard to reconcile. To show how the continuity requirement can be satisfied, despite appearances - or else to show that the continuity requirement is illusory - is a problem that must be solved if a philosophically satisfactory 'materialist' theory of resurrection is to be devised.

See also: Eschatology; Personal identity; Reincarnation §3; Soul, nature and immortality of the

PETER VAN INWAGEN

References and further reading

Cullmann, O. (1958) *Immortality of the Soul or Resurrection of the Dead?*, London: Epworth; also in K. Stendahl (ed.) *Immortality and Resurrection*, New York: Macmillan, 1965.(A classic and much discussed essay that contrasts the two concepts of its title and argues that the concept of the immortality of the soul is foreign to the New Testament.)

Edwards, P. (ed.) (1992) *Immortality*, New York: Macmillan.(A very useful collection, which contains an extensive bibliographical essay. Includes the relevant passages from Plato's *Phaedo* and Thomas Aquinas' *Summa theologiae*.)

Geach, P. (1969) *God and the Soul*, London: Routledge.(A modern defence of the 'composite' thesis; see especially chapter 2, 'Immortality', which is reprinted in Edwards.)

MacKay, D. (1987) 'Computer Software and Life after Death', in R. Gregory (ed.) *Oxford Companion to the Mind*, Oxford: Oxford University Press.(Argues that the distinction between computer hardware and software may be used to defend a 'materialist' account of resurrection. Reprinted in Edwards.)

Polkinghorne, J. (1994) *The Faith of a Physicist: Reflections of a Bottom-up Thinker*, Princeton, NJ: Princeton University Press.(Recent defence of a broadly 'materialist' account of resurrection; see especially chapter 6, 'Crucifixion and Resurrection', and chapter 9, 'Eschatology'.)

Tertullian, Q.S.F. (*c*.210-13) *De anima (On the Soul)*, in R. Arbesman, E.J. Daly and E.A. Quain (eds) *Tertullian: Apologetical Works*, New York, 1950.(Contains Tertullian's account of the soul as a corporeal entity.)

Van Inwagen, P. (1978) 'The Possibility of Resurrection', *International Journal for Philosophy of Religion* 9: 114-21.(This essay (which is reprinted in Edwards) and the following essay attempt to devise a metaphysic of resurrection that presupposes that human beings are living organisms.)

Van Inwagen, P. (1995) 'Dualism and Materialism: Athens and Jerusalem?', *Faith and Philosophy* 12: 475-88. (See previous item.)

Wolfson, H.A. (1961) 'Immortality and Resurrection in the Philosophy of the Church Fathers', in *Religious Philosophy: A Group of Essays*, Cambridge, MA: Belnap; also in K. Stendahl (ed.) *Immortality and Resurrection*, New York: Macmillan, 1965.(A very learned account of the positions of the Church Fathers on the metaphysics of resurrection.)

Revelation

All major theistic religions have claimed that God has revealed himself in some way, both by showing something of himself in events and also by providing some true, important and otherwise unknowable propositions. Event-revelation may include both general revelation (God revealing himself in very general events, observable by all, such as the existence of the universe and its conformity to natural laws), and special revelation (God revealing himself in certain particular historical events). The events are a revelation in the sense that God has brought them about and they show something of his character. Thus Judaism teaches that God manifested his nature and his love for Israel when he brought his people out of Egypt and led them to the promised land through the agency of Moses. Christianity traditionally affirms that God has revealed himself in a much fuller sense in Jesus Christ - because Jesus did not merely show us something of the character of God but was God himself. God reveals propositions by some chosen prophet or society telling us truths orally or in writing which we would not have adequate grounds for believing unless they had been announced to us by persons who showed some mark of God-given authority. Thus Islam teaches that God inspired Muhammad to write the Qur'an *in the seventh century* AD*, and that its success (its proclamation throughout a large part of the civilized world), content and style (deep thoughts expressed in a beautiful way, not to be expected of an uneducated person) show its divine origin.*

1 Event revelation

Any event brought about by an agent reveals something of that agent's character. If God made the world, one would expect its general character to show something about its Creator. The prophet Jeremiah represented God as speaking of his 'covenant of day and night', the regular succession of day after night and night after day, as showing the reliability of God and providing a reason why his solemn undertakings should be believed (Jeremiah 33: 20-6). Some recent writers have seen the focus of God's general revelation as the general progress of humanity in its understanding of the world and moral truth and in its working together for the advancement of all the human race. Most theologians have believed in special, as well as general revelation. But some recent 'neo-orthodox' Christian theologians, inspired by Karl Barth (§2), have held that God does not reveal himself in any comprehensible way in the natural world or secular history - he is too 'other' and too different from such things. By contrast, the eighteenth-century deist movement denied that God had manifested his presence in some historical events more than others; and this spirit has been in evidence in much recent liberal Christian theology (see Deism). It has been characterized by a denial that Christianity is in any unique way the true religion and so has held that if some events manifest God's nature more than others, they include events within the histories of all the great religious traditions and secular cultures. This is contrary to unanimous Christian tradition from the first to the eighteenth century, that God has manifested his presence pre-eminently in Christian history and above all in Jesus Christ (see Incarnation and Christology §1).

If one has any reason to believe that there is a God, one would expect the natural world to reveal to some extent what he is like. Only if one has other beliefs about what God is like would one deny this; and those other beliefs would require justification. Why believe that certain historical events reveal God better than others? Perhaps reason, and in particular our understanding of moral goodness, might lead us to see that if God is at work in the world, he is at work in some places more than others. Our private religious experience or the miraculous nature of the special events might also show us this. But if one allows general as well as special revelation, one can learn something about what God is like from the natural world; one can then use that to recognize in which particular events God was at work and, by reflecting on them, gain more knowledge of him.

Theistic religions have generally maintained that God manifests his presence not merely publicly in the natural order or in events within human history, but privately to particular individuals. But such events, except when they involve the conveying of information, are more naturally called religious experiences than instances of revelation.

If God reveals himself in particular events of human history and these are events of significance for the human race, then if later generations are to benefit from them, they will presumably need to know about them. Some Scripture or Church proclamations will presumably be needed. Many twentieth-century theologians have been unwilling to regard such written records - for example, the Bible - as themselves items of revelation. Barth preferred to call them 'witnesses' to revelation. Such theologians have regarded the scriptural texts as not always

true but as useful and inspiring signposts to what is true. Earlier Christian centuries generally regarded the Bible as containing only truths, some reporting significant historical events, others conveying God's commandments and truths about his nature.

2 Propositional revelation: its grounds

The major theistic religions have all held that God has revealed certain propositions, which are incorporated in written works. For Christians, they are in the Bible; for Jews, in the Hebrew Bible (known to Christians as 'The Old Testament' of their Bible); for Muslims, in the Qur'an. Why should anyone believe this?

As most theologians have supposed, there must be two kinds of relevant evidence. The first is the consonance of the propositions purportedly revealed with what is known by other routes of the nature of God (including the evidence of other propositions purportedly revealed). God is supposed, by definition, to be morally good. If the purportedly revealed propositions commend what is not morally good, they cannot be the commands of God (see Religion and morality §1). Again, God is supposed to be non-embodied and omnipotent (and natural theology claims that arguments from the natural world show this). If the propositions purportedly revealed speak of God having arms and legs, and being unable to do various actions, they cannot be revealed.

But if the propositions are to reveal anything to us, they must go beyond our prior knowledge while remaining consonant with it. Why believe one rather than another claim consonant with prior knowledge? The normal answer is that the purported revelation is accompanied by miracles, events contrary to natural laws, which, since God keeps natural laws operative, God alone could bring about (see Miracles). Miracles are thus the second kind of relevant evidence. If the miracles produce or forward the teaching of a purportedly revealed book, that is reasonably understood as an act of God (like a nod or a signature) authenticating that teaching. The major theistic religions have all claimed evidence of this sort for their revealed documents. The Qur'an was to be recognized as of divine origin by its 'inimitability' - it was a document of a kind that an uneducated person could not have written unless God had intervened to give him knowledge and literary ability beyond his education. The Hebrew Bible was to be recognized as of divine origin because it was the record of miraculous events (such as the parting of the Red Sea) and of the people founded as a result of those events. The Christian Bible was to be recognized as of divine origin because it was founded upon the resurrection of Jesus from the dead and other miracles which accompanied the foundation and growth of Christianity. That miracles are vital to authenticate claims to revelation was the normal teaching of Christian theologians, at least until the last century or two, when many have tried to dispense with any appeal to the miraculous.

Analogy with human cases explains why we look for these two kinds of evidence. If I receive a letter purportedly from John, my grounds for believing it to come really from John are first, that everything written in it is the sort of thing John could be expected to write; it is in character and shows knowledge of our previous meetings. And, second, it is written in a way or comes by a route that letters from John, and very few others, alone would do. Thus it may be that the handwriting is almost exactly like previous specimens of John's handwriting, and very few others can write in that way. Or it may be that the letter is given to me by a friend who says that John gave it to him, a friend whom in the past I have found to be completely trustworthy. The two kinds of evidence are sometimes called respectively internal and external evidence of authorship. We rightly look for both kinds of evidence of the divine authorship of purportedly revealed propositions.

A private religious experience had by the prophet writing the teaching down could perhaps provide him with the requisite authentication; and quite a few Christian writers, especially Protestants, have stressed the importance of the inner conviction that God is speaking to us through the text as grounds for believing its truth. But in view of the conflicting revelations purportedly inspired by God to which 'inner conviction' has seemed to testify, and the danger of self-deception in such matters, it would be dangerous for those other than the prophet to rely too much on inner conviction to authenticate such detailed claims about God as a revelation purports to give. There is need for public miracles.

How are you to know that a purported miracle took place and was indeed a miracle? We need to have evidence that if the event in question took place, it would have been contrary to natural laws; for example, that if the Red Sea parted in the way described in Exodus, that could not have been produced by natural processes of wind and tide operating normally. We also need evidence that the event in question took place. Evidence that some event

took place will always include both historical evidence of its purported effects - testimony of witnesses and any physical traces of its occurrence - and also evidence from any well-authenticated general theory of how likely such an event is to occur. The more evidence we have from overall theory that a certain kind of event is likely to happen some time or other, the less we need in the way of detailed historical evidence to show that it has happened. If your theory of astronomy leads you to suppose that sometimes stars will explode, then it is reasonable to interpret what you see through your telescope as the debris of a stellar explosion. The better justified your overall theory, the better justified your interpretation of the data. But if a well-justified theory tells you that these things cannot happen, then you need a great deal of detailed historical evidence to outweigh the countervailing theory. Similarly, the better established is theism, and the better established is the view that God has reason to intervene in history, the less you need by way of historical evidence to show that he has done so (see Atheism; God, arguments for the existence of). We must consider here whether if there is a God, he has reason to give us a revelation and so to authenticate it by miracle. If so, we need less in the way of detailed historical evidence to show that he has done so.

Some have supposed revelation to be unnecessary, since reason or natural theology can show us what God is like and which acts are morally good or bad, and that is all we need to know. The theistic religions have, however, generally claimed that while natural theology can show us quite a bit, there is also a lot that it cannot show us and that we need to know in order to have the right relation to God and to live the right sort of life on earth. St Thomas Aquinas (§14) in particular claimed that while natural theology can show us God’s existence and the most general moral truths, we need revelation to show us in much more detail what God is like (*Summa contra gentiles* I, chaps 2-13). Revelation also fleshes out our general duty of conformity to the will of God, by showing us, for example, that God wills us to worship him in particular ways on particular occasions. Even with respect to the former truths which, Aquinas claims, reason can discover, many humans may simply not have the time and ability to discover them; and revelation could also inform us about them. It is plausible to suppose that if there is a God, he wants us to know a lot more about him than pure reason can discover in order to enter into a personal relationship with him. And if he has provided some sort of community and institutions to help us to do this, he will need to inform us. The Catholic tradition has generally followed Aquinas, and seen Christian faith as involving the acceptance of the Christian revelation. The First Vatican Council (1870) declared that faith is a supernatural virtue whereby ‘we believe that what has been revealed by God is true’.

But if God dictates a book, it may get destroyed or misinterpreted, and with the passage of centuries any book can get badly misinterpreted. So it is plausible to suppose that a propositional revelation would need to be backed up by the provision of an institution or tradition of interpretation itself based on a miraculous foundation, which authenticated the book and its subsequent interpretation. Both Islam and Judaism have claimed that there are authorized interpreters of revelation. Catholic and Eastern Orthodox Christianity claim the Christian Church to be such a divinely authorized interpreter; and even many Protestants have held that the Christian tradition in a wide sense is a divinely guided tradition for interpreting the Bible.

3 Propositional revelation: its content

How is the Qur’an or the Hebrew Scriptures or the Bible to be interpreted? The Qur’an was written down at the dictation of a single human author, Muhammad, allegedly inspired by God. The Hebrew Scriptures were a collection of books compiled over many centuries, having different human authors; and the Bible was put together over even more centuries with even more human authors. How these Scriptures are to be interpreted depends on the genres to which we suppose their various parts to belong.

There are many kinds of literary genre: history, historical novel, narrative poetry, philosophical dialogue, science fiction, moral fable, and so on; and there are many genres known to the ancient world which are unfamiliar today, such as wisdom literature, apocalyptic, allegory, and Greek tragedy. They differ in various stylistic respects, and above all in the respect of whether they have a truth-value. Furthermore, in varying ways, for some but not all genres the truth-value of the whole is a function of the truth-value of the component sentences. Thus a newspaper story is true to the extent to which its component sentences are true. The whole is more or less true according to how many of the component sentences are true, and even the truth of individual sentences may be a matter of degree; some sentences may be only approximately true. By contrast, a simple modern novel is neither true nor false. But literary works may also have a true message, even though that is not a matter of having most of their

sentences true. Poems or fables, such as Aesop's fables, often tell us things deep and important about human nature or moral worth by means of fictional stories (about historical or non-existent persons), each of whose sentences taken on its own is false or lacks a truth-value.

So in interpreting an ancient work, purportedly revealed, we need to inquire about the genre of the whole and the way in which it affects the genre of the constituent books. Is, for example, the Book of Jonah, a constituent of the Hebrew Scriptures and the Bible, to be taken as straight history or as a moral fable? Our judgments about the genre of a work (and so how its sentences are to be interpreted) must depend on our judgments about who is the author and who are the intended audience. For the authorship of the Qur'an, there are only two possible candidates - Muhammad (if he was not inspired by God) and God. For the Bible, there are very many candidates, for the Bible is a patchwork; small passages written by one human author or group were put together into larger strands and then into 'Books', and those eventually into the whole Bible. Passages may have one meaning when written by one author on their own, a different meaning when used as part of a larger work by another author, and so on. Thus the central part of the Book of Ecclesiastes is an atheistic work, but this was sandwiched by another author within chapters having a theistic tone, which - if you think of the work as a whole - leads to reinterpretation of the central parts; the later author having taken over the central chapters, they must be interpreted as having a meaning consistent with the other chapters.

The document of revelation is, for Jews, the Hebrew Scriptures; for Christians, their whole Bible, Old and New Testaments. The Christian Church promulgated that work as having a central core, condensed into the Christian creeds, in the light of which the whole must be interpreted. It declared that Scripture was totally true and that God was the author of the Bible. It was meant for one Church of many centuries. What is the genre of the Bible? There is little with which to compare it. But there is one obvious principle of interpretation: that, like any other work, it must be interpreted in the light of other things we know about the author's beliefs. Hence there grew up in the Christian Church a tradition of biblical interpretation, first stated explicitly by Origen (§2), and continued in many writers of the first 1500 years AD: that passages are to be taken literally (as well, perhaps, as in other senses) where that is consonant with other things known about God, both through natural reason and as part of the central core of Christian revelation, and otherwise they are to be taken only in figurative senses (see Origen's *On First Principles*, 4.3). This led to books and passages of the Bible being interpreted in analogical, allegorical or metaphorical ways. For example, any passages at variance with the best science of the day or suggesting that God had a body were not to be taken literally; much in the early chapters of the book of Genesis needed to be taken in figurative senses.

What then is at stake with respect to any purported propositional revelation is whether the evidence of its content and any accompanying miracle suggests that God is its author and so that it is a genuine revelation. Competing claims to provide a revelation must be weighed against each other. But in judging what the content of a purported revelation is, the inquirer must do so on the assumption that God is its author.

See also: Hermeneutics, biblical; Jewish philosophy in the early 19th century §4; Natural theology; Prophecy; Religious experience

RICHARD SWINBURNE

References and further reading

Aquinas, T. (1259-65) *Summa contra gentiles* I, trans. A.C. Pegis, *On the Truth of the Catholic Faith*, bk I, Garden City, NY: Doubleday., 1955.(Statement of Aquinas' distinctions between natural and revealed theology; see chapters 2-13.)

Arberry, A.J. (1957) *Revelation and Reason in Islam*, London: Allen & Unwin.(Useful guide to the history of Islamic thought on the issue.)

Butler, J. (1736) *The Analogy of Religion*, London: George Bell & Sons, 1902, pt 2, 'Of Revealed Religion'.(A classic statement of the need for propositional revelation and for miracles as evidence for it.)

Dulles, A. (1983) *Models of Revelation*, Dublin: Gill & Macmillan.(A survey of twentieth-century theological accounts of revelation.)

Duns Scotus, J. (*c*.1305) *Ordinatio*, Prologus, 2, 100-19, vol. 1 of *Omnia opera*, Vatican City: Scotistic Commission, 1950-.(A very clear fourteenth-century account of the grounds for believing what is revealed in the Bible. There is no available English translation.)

Newman, J.H. (1845) *An Essay on the Development of Christian Doctrine*, London: Longmans, Green & Co. (Newman's classic statement of the view that Christianity involves a progressive development in understanding

the doctrines originally revealed.)
Origen (*c*.225) *On First Principles*, trans. G.W. Butterworth, New York: Harper & Row, 1966, 4.3.(First and classic Christian statement of the principles of biblical interpretation.)
Pannenberg, W. (1991) *Systematic Theology*, trans. G.W. Bromley, Edinburgh: T. & T. Clark, vol. 1, ch. 4.(A history of Christian understanding of revelation, including an exposition of revelation as God acting through historical events.)
Swinburne, R. (1992) *Revelation: From Metaphor to Analogy*, Oxford: Clarendon Press.(Develops an account of the meaning and justification of Christian propositional revelation.)

Revolution

There have been revolutions in politics, science, philosophy and most other spheres of human life. This entry discusses revolution mainly through concepts pertaining especially to the political realm. Attempts to define political revolution have been controversial; as a consequence there is dispute about whether specific occurrences were revolutions, rebellions, coups d'état *or reformations.*

If we define revolution as the illegal introduction of a radically new situation and order for the sake of obtaining or increasing individual or communal freedom, we may list those characteristics most often ascribed to it. These characteristics distinguish it from its earlier use where revolution referred to the return of an original state of affairs, as in astronomy; they also allow its distinction from related concepts such as reformation. At least at a superficial level this definition can do justice to early modern (seventeenth and eighteenth) as well as late modern (nineteenth and twentieth century) revolutions. Through these periods there has, however, been sufficient change in concepts closely related to revolution to require the definition's openness to nuances for it to apply to both periods. It is unclear whether even such a nuanced definition can apply in postmodern thought.

1 Related concepts

Revolution should be distinguished from *reformation*, *rebellion* and *coup d'état*. *Coup d'état* refers to attempts to change the rulers of a system, often with the intent of making more accessible or changing the benefits derivable from that same system; and rebellion indicates attempts to redress specific grievances within a particular system without necessarily a change of that system's leadership. Reformation denotes events potentially more extensive than either of these; it aims at general overhaul by insisting that pre-existing fundamental norms judge both rulers and systems. If a system is not structured by these norms and its rulers refuse their implementation, reformers take it as their task to change the system to conformity, if necessary against the rulers' wishes. Not infrequently, one result is a split in the system, with part remaining unreformed under the old leadership and part claiming successful confirmation of old norms under new leadership. Revolution designates events more radical than any of these; its intent is not modification but uprooting and transformation.

To clarify the difference between reformation and revolution, consider two events whose traditional labels correctly mark it: the early sixteenth-century Lutheran Reformation and the late eighteenth-century French Revolution. Although they share their stress on the individual and its concomitant rejection of certain kinds of authority, there are profound dissimilarities reflecting an essential distinction.

Notwithstanding Luther's stress on individuals in his rejection of the church's mediating role between individual and God, he insists primarily on extrapersonal unchanging norms - revealed by God and codified in the Scriptures - which are to judge both church and individual. Since he believed these norms no longer characterized institutional and individual practice, their practice needed re-formation in order to restore institutional legitimacy and secure the individuals' justification. Underlying Luther's slogan 'through faith alone' is the requirement that we should return to intended and traditional ways. Equality before God includes equal individual responsibility for implementing God's unchanging edicts (see Luther).

Of the French Revolution's slogans 'no God, no master' and 'freedom, equality, brotherhood', the first demands extirpation of all external religious and secular hierarchical power, while the second indicates the outcome of heeding that call. Liberation from divine and royal edicts enfranchises a common humanity expressed through each free and equal individual. The norms which are to structure life now find their source in each individual's reason; and the assumed commonality of human nature, including reason's universal sameness, guarantees the uniformity and authorizes the universally compelling force of these rules. The Revolution meant to effect a total reversal: institutions sanctioned in the past through extra-individual powers are found wanting when judged by new norms demanding a new order derived from its subjects.

This comparison distinguishes reformation from revolution through their different radicalism. Whereas reformers return to roots, insisting that institutions and actions conform to root situations, revolutionaries pluck up and discard roots. Reformers reinstitute old norms, revolutionaries establish new ones. When they aim to re-establish a quite probably mythical past golden age, reformers are appropriately called 'utopian reactionaries'. In contrast to

reformers, revolutionaries are anti-tradition in their suspicion and rejection of whatever their cultural present or past would have them accept on trust.

Both rebellion and *coup d'état* can function in reforming and revolutionary movements. Rebellions may occur when a reforming movement's leadership is deemed to effect change too quickly or too slowly, or to aim at reinstitution of the wrong norms. If the leadership resist the rebels' demands a coup may follow - although often not, strictly speaking, a *coup d'état*, since the leaders of a reformation need not be leaders of a state. The success of a *coup* allows for continuing reformation at a different pace or in a different direction. Similarly for revolutions: rebellions or coups are then not themselves revolutionary to the extent that the revolution is already under way; rebels then demand a change of pace or of direction (that is, new ideals different from those pursued by the current leadership). If the current revolutionary leadership rejects such demands a *coup* may follow. Regimes that revolutionaries want to depose seldom leave quietly; revolutions tend to be violent events. Similarly for *coups* within revolutionary movements: of the French Revolution it was said that it devoured its children.

2 Defining characteristics

The definition introduced above contains three characteristics of revolution: (1) radical novelty, (2) illegality, and (3) freedom. These three are both essential and most frequently present in writings on revolution, and will remain focal.

Radical novelty is generally seen as fundamental to revolution; both illegality and freedom presuppose it. Writers speak of revolution as involving 'drastic substitution', 'fundamental change', 'the speaking of a new, unheard of language, another logic, a revaluation of all values', a change 'affecting the fundamental laws' and expressing itself in 'radical transformation' (Brinton 1952; Friedrich 1966). This radical novelty as fundamental change makes revolutions typically modern phenomena. Pre-modern times accepted a transcendent reality (as in the Greeks' Platonic Good or the Medievals' Christian God) which characteristically imposed eternally and universally binding norms for human institutions and actions. The appearance of revolutions coincides with the irrelevance (at least in principle) of the transcendent in some of the most important events of human history.

Not all alteration of fundamental laws or constitutions is illegal; constitutions often include directives for their legal amendments. Revolutionary change is illegal change of fundamental laws, unconstitutional change of constitutions. Because revolutionary change is meant to be sweeping and fundamental, it intends to remove whatever opposes it which is, primarily, the current order. Since the order to be instituted is radically different from the order it means to replace, the latter contains no provision for sanctioning the former.

Revolutions occur for the sake of promoting human freedom (Arendt 1963). This freedom is both positive and negative, freedom from a current order creating room for freedom to institute a new order. The pertaining order is always seen as constricting, its revolutionary abolition as enfranchising, the new order as guaranteeing greater freedom. Like illegality so also is freedom intrinsically connected with radical novelty: what is completely swept away has lost all power to prevent the revolutionaries' establishment of their new order. Freedom is connected, as well, with passion, for revolutionaries are persons chafing under constraints which they want removed; they are driven by anger about the present, confidence in the possibility of progress and hence hope for a better future. Passion precipitates revolutions, while reason serves as an instrument to direct both liberating activity and the establishment of new freedom.

3 Revolution and reason

An intimate relationship between revolution and reason in the early modern period raises the question whether a single definition of revolution is possible given changing views of the nature and role of reason as we move from this period through historicism into postmodernism (see Historicism; Postmodernism).

Influential early modern thinkers took reason to dictate the revolutionary stance of plucking up and discarding roots. Descartes was the first to present this stance as integral to all systematic thinking. In his *Discourse on the Method* (1637) reductionistic reason dictates non-acceptance of whatever experience initially provides as true or good unless the individual is able to free such items from their given contexts and then fit them into a new rational scheme. His *Meditations* (1641) provide the new foundation for this rational scheme in the activity of the individual's autonomous thought. Descartes's project displays all the major characteristics of revolutionary

activity: it opposes predominant systems of his day which upheld the absolute necessity of tradition and so could not possibly legitimate its uprooting; it introduces the radical novelty of locating the criteria of truth in each individual thinker; its initial rejection of all givens aims to free thinkers from historical contexts to enable their progress in knowledge, power and freedom. In the following century figures important to the French Revolution spoke of Descartes as their precursor in breaking the tyranny of tradition.

With Hegel (§8), Marx (§8) and Engels, historicism, for which anything must be understood in its past-dependent changing context, becomes part of the Western outlook. Marx and Engels therefore locate initiating revolutionary force in material and economic conditions rather than in persons no longer characterized by reductionistic reason. *The Communist Manifesto* (1848) states as conditions for revolution: exploitation of the suppressed class to the degree that it understands and becomes ready to shake off institutional shackles; and incapacity of the oppressors to continue exploitation competently. Both conditions appear through the historically inexorable development of class antagonism under changing forms of economic production. Reason still has a place - revolution is a rational attempt by the oppressed to achieve change - but economic forces expressed as macro-historical laws are primary in determining the possibility of revolutionary activity. Economic forces, argues *The German Ideology* (1846), cause changes in material production and alter the oppressed's thinking and products of their thinking so that, generally, it is not thought which determines life but life which determines thought. Nevertheless, chief characteristics of the earlier definition remain. The socialist revolution is seen as introducing novelty greater than that of the French Revolution, for it will abolish classes as well as the old division of labour; it will certainly be illegal in the eyes of both oppressors and oppressed, for it will annul the oppressing legal system and thereby progress towards new forms of freedom.

Postmodernism considers itself free from all forms of modern rationality; in its rejection of both reductionistic and holistic past narratives it embraces epistemic, moral and cultural relativism. Consequently, postmodernism neither advocates a particular political stance nor identifies an overarching cause for revolution (see Postmodernism in political philosophy §2). It rejects 'the *foundational* character of the revolutionary act, the institution of a point of concentration of power from which society could be "rationally" reorganized' (Laclau and Mouffe 1985: 177; original emphasis). Nevertheless, the term revolution is ubiquitous in its political, sociological and artistic writings and 'revolutionary' now conveys no less praise than during late-modernity's struggles for liberation. But can revolution retain any of its modern content given a radical diversity of values? Are the dislocated selves of postmodernity still capable of deliberate revolutionary activity in a relativistic culture? Or is it only possible to rebel against contingent structures, or perpetrate *coups* at best resulting in transfer of control from one group to another without a fuller realization of individual or communal freedom?

Modernity's revolution became possible given its concepts of rationality, freedom, radical novelty and progress, where at least some of the meaning of these terms carried over from its earlier to its later phase. Postmodernity's questioning the meaning of each of these terms makes the univocal use of revolution more problematic when we move from modernity to postmodernity than from early to late modernity.

PETER A. SCHOULS

References and further reading

Arendt, H. (1963) *On Revolution*, Westport, CT: Greenwood Press, 140-1, 235-6.(Referred to in §2. A classic modern introduction to the topic, illustrated with reference to major European events.)

Brinton, C. (1952) *Anatomy of Revolution*, New York: Vintage Books, esp. its opening pages and 277-8.(Referred to in §2. Presents a comparative historical study of major Western revolutions with the French Revolution as an archetypical phenomenon.)

Calvert, P. (1990) *Revolution and Counter-Revolution*, Milton Keynes: Open University Press.(Useful general introduction to theoretical accounts of the concept.)

Descartes, R. (1637) *Discourses on the Method*, in J. Cottingham, R. Stoothoff and D. Murdoch (eds) *The Philosophical Writings of Descartes*, vol. 1, Cambridge: Cambridge University Press, 1985.(Introduces the modern stance for which reductionistic human reason dictates that, fundamentally, truth is known non-contextually.)

Descartes, R. (1641) *Meditations*, in J. Cottingham, R. Stoothoff and D. Murdoch (eds) *The Philosophical Writings of Descartes*, vol. 2, Cambridge: Cambridge University Press, 1985.(Presents the influential early-modern argument for the autonomy of the individual's thought which places the criteria for truth in each

individual thinker.)

Dunn, J. (1989) *Modern Revolutions: An Introduction to the Analysis of a Political Phenomenon*, Cambridge: Cambridge University Press, 2nd edn.(An important theoretical introduction followed by an analysis of eight twentieth-century revolutions; provides an extensive bibliography.)

Friedrich, C.J. (1966) *Revolution*, New York: Atherton Press, esp. 4, 37, 124. (Referred to in §2. Contains various useful articles on theories of revolution.)

Kristeva, J. (1984) *Revolution in Poetic Language*, New York: Columbia University Press.(Explores the meaning of revolution in postmodernist semiotic and aesthetic terms.)

Kuhn, T.S. (1970) *The Structure of Scientific Revolutions*, Enlarged, Chicago, IL: University of Chicago Press, 2nd edn.(Classic treatment of revolutions in science, which argues that the characteristics of radical novelty and illegality hold for scientific as well as political revolutions.)

Laclau, E. (1991) *New Reflections on the Revolution of Our Time*, London: Verso. (Argues that, free from modern notions of rationality, certain kinds of revolution are still possible in a postmodern world.)

Laclau, E. and Mouffe, C. (1985) *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics*, London: Verso.(Referred to in §3. Defines 'radical democracy' in the postmodern condition of social and cultural relativism, a definition which displaces the old leftist understanding of and commitment to revolution.)

Marx, K. and Engels F. (1846) *The German Ideology*, ed. C.J. Arthur, London: Lawrence & Wishart, 1970. (Referred to in §3. Explores relationships among material and economic conditions, modes of thought and sociopolitical activity.)

Marx, K. and Engels F. (1848) *The Communist Manifesto*, London: Lawrence & Wishart, 1983.(Referred to in §3. A popular tract which states the conditions for, and urges the working class to, revolution.)

rGyal tshab dar ma rin chen (1364- 1432)

rGyal tshab dar ma rin chen (Gyeltsap darma rinchen) was a disciple of the great Tsong kha pa. Like much Tibetan philosophy, his work is commentarial in style. In his commentaries on the work of Dharmakīrti he developed a moderate realist position with regard to abstract entities, claiming that this was what Dharmakīrti (apparently an antirealist) really intended. Properties, rGyal tshab maintained, do exist, but not independently of things; universals are separable from particulars in thought, but never in perception. Perception straightforwardly presents us with real objects; inference, on the other hand, tends to present reality in a distorted way.

rGyal tshab dar ma rin chen (Gyeltsap darma rinchen) was one of the foremost disciples of Tsong kha pa Blo bzang grags pa (Dzongkaba, 1357-1419), the founder of the dGa'-ldan (Ganden) school, later known as dGe-lugs-pa (Geluk). rGyal tshab was born in Western Tibet at Ri-nang and moved to Sa-skya, where he acquired a solid scholarly reputation prior to meeting Tsong kha pa. He became one of Tsong kha pa's closest disciples. After his teacher passed away, rGyal tshab succeeded him to the throne of dGa'-ldan monastery.

Even more than other Tibetan thinkers, rGyal tshab discusses classical problems both on the basis of their philosophical merits and in relation to, and under the form of, commentaries on basic texts formative of the tradition. He does not examine philosophical propositions independently of the traditional framework, but always examines other systems from within his own and defends or develops his own from within. This he sees as responsible philosophizing, as distinguished from irresponsible sophistry. rGyal tshab offers a number of clear expositions of important Buddhist commentaries omitted by his master, and among his commentaries those elucidating the texts of Maitreya and Dharmakīrti are his greatest contributions.

A peculiarity of the way rGyal tshab interprets Dharmakīrti's epistemology is that he presents a realist version of a system that is fundamentally antirealist. Although Dharmakīrti (§2) describes universals as unreal conceptual constructs, rGyal tshab does not accept this description as representing Dharmakīrti's true thought. Following a realist tradition of interpretation that goes back to at least Phya pa chos kyi seng ge (Chaba chögyi sengge, 1109-69), rGyal tshab holds that Dharmakīrti's many passages that seem to reject universals must be reinterpreted, for they lead to unacceptable philosophical consequences.

rGyal tshab's position is a form of moderate realism. Like its Western counterparts, it is born from the attempt to find a middle ground between the extreme realists and the antirealists. Moderate realists are sympathetic to the antirealist's rejection of the extreme realist assertion that properties exist independently of the particulars that instantiate them. They are unwilling, however, to take the next antirealist step, namely the complete rejection of abstract entities. If abstract entities were totally dependent on our conceptual schemata, as argued by the antirealists, our concepts would be arbitrary. To avoid this unwanted consequence, moderate realists hold that we must admit that properties exist independently of the mind. This does not mean, argue the moderate realists, that they exist independently of their instances. How, they ask, can properties exist without any support in empirical reality? Consequently, moderate realists hold that universals exist in things (see Abstract objects).

The problem with this position is that it makes it difficult to spell out the relation between universals and their instances. Since they do not exist apart, they cannot be different. And yet they cannot be identical either, since the former are not reducible to the latter. rGyal tshab and his tradition respond to this predicament, which seems inherent in moderate realism, by positing a new type of relation: particulars and universals are described as being distinct and yet the same entity (*ngo bo gcig la ldog pa tha dad*). Suffice it to say here that, for rGyal tshab, particulars and universals are distinct because they can be separated by thought. Hence they can be thought of as apart, but, since they cannot be taken apart by perceptual experience, they are the same entity.

Epistemologically, the consequence of rGyal tshab's realism is that he presents a picture of perception that differs markedly from that of Sa skya Paṇḍita (Sagya paṇḍita, 1182-1251). For rGyal tshab, perception presents us with a fully fledged common-sense object. When we perceive, for example, a jar, we do not merely sense a presence that we interpret as a jar. Rather, we directly see the object as being a jar. This common-sense account of perception is possible for rGyal tshab because of his acceptance of real universals. Because they exist in reality, perception, which accurately reflects reality, can perceive them and deliver a cognitively articulated vision of reality.

A similar realist streak is present in rGyal tshab's account of inference. He accepts that conceptual cognition in general, and inference in particular, are distorted. They do not apprehend the object as it is. This does not mean, argues rGyal tshab, that they are totally mistaken, as antirealist interpreters wrongly assume. Conceptuality should not be thought of as being cut off from reality, but rather as putting us in touch with reality, albeit in an indirect and distorted way.

See also: Nominalism, Buddhist doctrine of; Tibetan philosophy

GEORGES B.J. DREYFUS

List of works

rGyal tshab dar ma rin chen (1364-1432) *tshad ma rnam 'grel gyi tshig le'ur byas pa'i rnam bshad thar lam phyin ci ma log par gsal bar byed pa (The Faultless Revealing of the Path to Liberation, a Complete Explanation of the Stanzas of [Dharmakīrti's] Pramāṇavārttika)*, Varanasi: Ge-luk-ba Press, 1974-5; second chapter trans. R. Jackson, *Is Enlightenment Possible? Dharmakīrti and rGyal tshab on Knowledge, Rebirth, No-Self and Liberation*, Ithaca, NY: Snow Lion Publications, 1993.(rGyal tshab's major commentary on Dharmakīrti's main work on epistemology and logic, the 'Commentary on Valid Cognition'.)

rGyal tshab dar ma rin chen (1364-1432) *tshad ma rigs pa'i gter gyi rnam bshad legs par tshad pa'i snying po (Essence of Good Sayings, an Explanation of [Sa skya Paṇḍita's] tshad ma rigs gter)*, in G. Dreyfus and S. Onoda, *A Recent Rediscovery: rGyal tshab's Rigs gter rnam bshad*, Kyoto: Biblia Tibetica, 1994.(This commentary on Sa skya Paṇḍita's work on logic and epistemology - the 'Treasure on the Science of Valid Cognition' - was not included in rGyal tshab's collected works, but then resurfaced in Tibet.)

rGyal tshab dar ma rin chen (1364-1432) *phar phyin rnam bshad snying po rgyan (The Essential Ornament, a Complete Explanation of the Perfection [of Wisdom Sūtra])*, Varanasi: Ge-luk-ba Press, 1980.(rGyal tshab's main commentary on Maitreya's *Abhismayā laṃkara*.)

rGyal tshab dar ma rin chen (1364-1432) *Commentary on Āryadeva's Catuḥśataka*, trans. R. Sonam and Geshe Sonam Rinchen, *The Yogic Deeds of Bodhisattvas; Gyel-tsap on Āryadeva's Four Hundred*, Ithaca, NY: Snow Lion Publications, 1994.(Being an explanation of Buddhist practice according to the Madhayamaka school by Nāgārjuna's main disciple.)

References and further reading

Dreyfus, G. (1996) *Recognizing Reality*, New York: State University of New York Press.(A philosophical discussion of Sa skya Paṇḍita's epistemology in the context of the Tibetan reception of Dharmakīrti.)

Rhetoric

Rhetoric is the power to persuade, especially about political or public affairs. Sometimes philosophy has defined itself in opposition to rhetoric - Plato invented the term 'rhetoric' so that philosophy could define itself by contrast, and distinctions like that between persuasion and knowledge have been popular ever since. Sometimes philosophy has used rhetorical techniques or materials to advance its own projects. Some of its techniques, especially topics of invention, the classification of issues, and tropes or figures of speech, are occasionally employed by philosophers. The philosophical question is whether these techniques have any interest beyond efficacy. What is the relation between techniques effective in persuading others and methods for making up one's own mind? Is there any connection between the most persuasive case and the best decision? Is there a relation between the judgments of appropriateness and decorum exercised by the rhetorician, and the judgments of appropriateness exercised by the person of practical wisdom? Do judgments about probability, ambiguity and uncertainty, and judgments under constraints of time or the need for decision, aspire to the ideal of perfect rationality, to which they are doomed to fall short, or do these kinds of judgment have an integrity of their own?Apart from supplying useful techniques, an art of persuasion also raises philosophic questions concerning the relation between rhetoric and logic, rhetoric and ethics, and rhetoric and poetics.

1 Philosophy and rhetoric at their beginning

Plato invented the term 'rhetoric' as a contrast term against which philosophy could define itself. Philosophy and rhetoric both proposed new truths and apparently powerful methods that threatened existing moral codes and authorities. As modes of empowerment, self-making and self-consciousness they could as easily become enemies as allies. Thus while philosophy attempted to achieve self-consciousness and power through an awareness of thought and being, rhetoric focused on an awareness of language and the circumstances of speaking and acting. Cicero interpreted the Platonic separation of philosophy from rhetoric as that of Socrates dividing wisdom from eloquence, and saw that the task of philosophical rhetoric was to overcome this separation. Prior to Socrates, he says in *de Oratore*, 'the same teaching seems to have imparted education both in right conduct and in good speech…the same masters gave instruction in both ethics and rhetoric' (see Cicero §§1, 2). From the beginning, then, rhetoric already had the pejorative connotation it continues to have, in spite of periodic attempts to speak in its defence.

Rhetoric's importance for philosophy seems greatest in periods of political and cultural confrontation and pluralism, for example, during Roman expansion, the Renaissance and today. These are all times of 'linguistic turns' in philosophy and, in addition, of fruitful interactions between philosophic and legal argumentation and interpretation. Linguistic, legal and pragmatic turns are rhetorical phenomena that direct attention to effective reasoned communication. In such times, the revival of rhetoric is motivated by a sense that philosophy itself has become too professionalized and too remote from human concerns. Rhetoric is then invoked, as in Cicero, to bring philosophy back down to earth. That neither Rome nor the Renaissance is considered a golden age of philosophy is probably a reflection of their rhetorical character.

Whether or not the history of philosophy is a series of footnotes to Plato, the history of the philosophical issues of rhetoric certainly can be read that way. The *Gorgias* begins by showing the difficulty in defining rhetoric. Gorgias finds it hard not only to define rhetoric logically, but to offer a definition in the more serious sense of keeping it within its proper sphere. In Socrates' hands the universality of rhetoric becomes a sign of emptiness, not power, and its flexibility turns from resourcefulness into slavishness. The conversation between Socrates and Polus make this philosophy of rhetoric the first example of 'applied ethics', in which one searches for ethical constraints on an activity. The conversation between Socrates and Callicles originates the tradition of 'professional ethics' which relates the internal norms of a practice to the sort of person the practitioner is.

Other Platonic dialogues amplify these considerations and raise further questions. The same rhetoric that appears in the *Gorgias* as a power of domination appears in the *Protagoras* in a context of mutual agreement. The most obvious use of rhetoric - arguing both sides of a question - can be either to exercise power unilaterally or to provide the basis of community, just as competitive Homeric virtues either underlie or are transformed into the cooperative virtues of the polis. In the twentieth century, Kenneth Burke (1950) highlights this shift by changing

the central occupation of rhetoric from persuasion to 'identification'. The relation between the competitive and cooperative facets of rhetorical argument sets another philosophic problem for rhetoric: Under what conditions can the power to argue both sides of a question become the power to uncover truth through the free competition or marketplace of ideas? (In *On Liberty*, Mill develops his vision of truth emerging through controversy by reference to Cicero, 'the second greatest orator in antiquity'.) The introductory conversation in the *Protagoras* between Hippocrates and Socrates raises questions about the relation between learning from the sophists and being a sophist oneself. The dialogue then presents Socrates in conversation with the sophists Protagoras, Hippias and Prodicus. Protagoras limits the domain of rhetoric to political questions, justice and civic virtue, while Hippias would expand it to cover the natural sciences, and Prodicus would extend it to literary interpretation. Whether rhetoric is restricted to political questions, whether there is a 'rhetoric of science' or further uses of rhetoric outside politics all become standard questions for rhetoric from then onwards.

The *Phaedrus* turns to a third possible alliance which defines rhetoric in relation to poetics. In addition to rhetoric and logic (what does the rhetorician know?) and rhetoric and ethics (what are the uses and abuses of this power, and what sort of person does one become in practising rhetoric?) there is the relation between rhetoric and poetry, another activity that either employs charms not available to more rational approaches or operates at a further remove from reality. The *Phaedrus* raises further questions concerning the relation between written and spoken language, and so between the production and reception of discourse, a line of investigation that eventually leads from rhetoric to hermeneutics. Questions from the *Gorgias* concerning rhetoric and knowledge are formulated in the *Phaedrus* in a pointed manner as the relation between rhetoric and sincerity and deception, or between persuasion, *eros*, beauty, and seduction. Burke's investigations of the relations between serious and playful uses of language are a modern rediscovery of this set of questions, as are the recent revivals of rhetoric in de Man and Derrida.

The *Phaedrus* ends with Socrates asking whether there is a rhetoric for philosophy - that is, whether there are modes of communication that are especially suited for philosophy. But Socrates' questions about written versus spoken language, and long speeches versus questions and answers, at the same time pose the parallel problem of whether rhetoric has a philosophy, whether it has ethical or metaphysical commitments. The answer is parallel to that for the analogous question about mathematics: does the mathematical enterprise inherently have a 'philosophy'. There seems to be a natural affinity between mathematics and one type of philosophy, namely Platonism. On the other hand there are as many philosophies of mathematics as there are types of philosophy. Similarly, there are affinities of rhetoric with scepticism and relativism, and there are as many philosophies of rhetoric as there are kinds of philosophy (see Plato §§4, 5).

2 Philosophical problems posed by rhetoric

Rhetoric is persuasive communication, especially about practice and politics. Whether rhetoric is a subject for philosophy depends on whether persuasion raises issues of its own apart from those exhausted by logic, ethics and psychology. If the way people were persuaded had nothing at all to do with logical cogency or ethical responsibility, or if it was simply reducible to one of them, there would be nothing philosophic about rhetoric. It is only because there seems to be some connection between logical cogency and rhetorical persuasiveness, between the trust accorded a persuasive speaker and the confidence we have in good ethical agents, and between figurative language and something more than mere ornament, that rhetoric is a subject for philosophy. While rhetorical theory has had negligible influence on the history of philosophy, the problems raised by rhetorical practice are significant philosophical problems.

All the issues raised by Plato's dialogues could have become philosophical problems in a variety of ways, but they were disciplined by Aristotle's organization of knowledge. The details of Aristotle's *Rhetoric* have had little impact on the history of either philosophy or rhetoric, but his placement of rhetoric relative to other arts and sciences has organized the trajectory of the relation of philosophy and rhetoric. Rhetoric, he says, is the offshoot of dialectic and politics. There are three ends and kinds of rhetoric: political or deliberative rhetoric directed towards utility; judicial or forensic rhetoric which determines justice and injustice; and epideictic or demonstrative rhetoric which concerns worthiness and blameworthiness. There are three sources of persuasion or belief (*pistis*) in rhetoric - the character (*êthos*) of the speaker, the passions of the audience, and the 'speech (*logos*) itself, in so far as it proves or seems to prove' - generating the trio implicit in Plato: rhetoric and logic, rhetoric and ethics, rhetoric and

poetics.

The meaning and scope of all these terms of opposition (logic, ethics and poetics) shifts as the domain of rhetoric itself expands and contracts. Rhetoric is potentially about everything that can be communicated, and so about all thought and language. But the use of speech in deliberation about the concrete and practical particular is always at its core. As ethics after Aristotle became less political, so too did rhetoric. In Plato's dialogue, Protagoras defines his subject politically as the abilities without which men could not be citizens. Others softened the definition to the ability to speak on those matters on which people would be ashamed not to have an opinion, that is, the field of common sense and common understanding. Austin's (1962) inquiry captures part of the idea of persuasive communication apart from a political context, and shows the alliance between the linguistic and the rhetorical turns of philosophy (see Austin, J.L.). Some argue that rhetoric is as broad as human speech in general, with extensions to general rhetorics like the 'rhetoric of inquiry' or 'rhetoric of the human sciences' or, as with Burke, to all strategies for encompassing situations. Others would restrict it to public or political affairs. The broadening of rhetoric is sometimes a reflection of the lack of political, or even practical, functioning for rhetoric, and sometimes a corrective to the impractical specialization and professionalization of rhetoric itself. The logic, ethics and poetics against which rhetoric defines itself are usually much broader than the three sources of proof in Aristotle's account.

3 Rhetorical methods and philosophy

As rhetoric developed, it elaborated specific techniques for persuasion, and even if rhetorical activity had no philosophic dimensions, the history of rhetoric would still offer a usable history of persuasion. The first rhetorical method with philosophic import is that of topical invention, which brings into focus the relation between rhetoric and logic. *Topos* is the Greek word for place, translated into Latin as *locus*. *Koinoi topoi* became *loci communi* and then, in English, 'commonplaces,' places to find arguments, pigeon-holes for classifying appeals, or major premises from which to derive particular conclusions. Topics are the means for finding something persuasive to say in a practical situation. Aristotle defines rhetoric as an art of finding in any case the available means of persuasion. Subsequent rhetoricians make invention (*heuresis*, *inventio*) the first of five parts of rhetoric, alongside judgment (sometimes called disposition or arrangement), style, memory and delivery. Any of these can be a merely verbal technique or an art with philosophical significance. The topics can range from the logical ('similar effects have similar causes') to more substantive considerations ('follow the money'; '*cherchez la femme*'; 'we all know that Athenians are acquisitive'). Often there is a connection between methods for finding arguments and systems of artificial memory for recalling them. When the rhetorical invention of arguments is replaced by a scientific discovery of things, only style remains in rhetoric's domain. Ramus' (1574) division of labour between a topical dialectic or logic and a rhetoric confined to style and arrangement in the early modern period is perhaps the most historically influential of such moves.

Scientific method might supplant rhetorical invention, but the methods of scientific discovery which replace rhetorical invention were themselves derived from topical invention. Much of the meaning of necessity in seventeenth-century science, for example, derives from pleas of necessity as an excuse or justification in legal proceedings. Stephen Toulmin's (1964) 'inference warrants', which supply the connection between grounds and claims, is one recent reappearance of the topics; Chaim Perelman's (1969) *loci* for establishing connections is another.

Hume finds the meaning of 'personal merit' through epideictic rhetoric by 'displaying the praises of any humane, beneficent man'. He refers to the considerations used as the 'topics of praise'. The inter-relations among the three kinds of rhetoric, and the kinds of argument appropriate to each, generate significant philosophical problems. Why, as Hume asks, does utility please? What is the relation between the impartial spectator whose judgment is the focus of epideictic rhetoric, the judge whose verdicts are the end of forensic rhetoric, and the ideal deliberative agent? What is the relation between the right and the good, or between what is best and what should be done? What is the connection between action-guiding and agent-evaluating reasons?

The other two philosophically interesting rhetorical methods besides topical invention can trace their ancestry back to Plato and Aristotle, but were really developed in later rhetorics which emphasized judicial oratory and the performance values associated primarily with demonstrative rhetoric. Issues - *staseis* in Greek and *constitutiones* in Latin - were rhetoric's central technique for most of its history. The classification of issues evolved out of

practices of judicial rhetoric, which is always the most systematic of the three kinds of rhetoric, and they draw attention to the relation between rhetoric and ethics. There are, usually, four *staseis*: (1) the question of fact or conjectural issue; (2) the definitive stasis; (3) the qualitative issue concerning mitigating or aggravating circumstances; and (4) the translative issue of whether a given court is the appropriate forum for judgment. Kant, not long after Hume, explains the 'principles of any transcendental deduction' by referring to stasis theory: 'Jurists, when speaking of rights and claims, distinguish in a legal action the question of right (*quid juris*) from the question of fact (*quid facti*)' (1787: B116-7). His question of right is the translative issue, by what right knowledge can be judged to have a 'legal title', so the foundations of knowledge cannot be secured by pointing to facts about the success of science. Austin's 'plea for excuses' and contemporary arguments about responsibility invite reconsideration of the rich problems that revolve around the qualitative issue.

If the topics emphasize the relations between rhetoric and logic, and the theory of issues or *staseis* develops the relation between rhetoric and ethics, the third rhetorical method, that of figures or tropes, concerns the relation of rhetoric to poetics. Augustine, in the *de Doctrina Christiana*, used rhetoric to teach Christians to read their Bible figuratively and to adjust contradictory passages by considering the circumstances to which speakers and authors adapted their statements and meanings. For Vico (1744), a near contemporary of Hume and Kant, the 'tropes are corollaries of poetic logic'. Borrowing from a formulation that seems to originate with Ramus, Vico names four primary tropes: metaphor, metonymy, synecdoche and irony. These tropes are not merely elegant variations on things that could be expressed more directly, but reveal original truths hidden by more sophisticated and professionalized philosophy. In the hands of our own contemporaries such as Kenneth Burke (1950) and Hayden White (1973), the figures move from methods of description or redescription into methods of seeing and encompassing situations, as rhetoric is expanded from a verbal to a universal art. Perelman, similarly, constructs a relation between argumentative forms and figures of speech and thought. Vico has become a hero to those who think that progress will come in philosophy from being anti-Cartesian, and championing conversation and communication over more limited, anti-rhetorical, forms of rationality. These examples show how even the most language-centred or poetic dimension of rhetoric has an ethical agenda. The rhetorical emphasis on saying what is appropriate provides the connection between ethical and poetic demands for timeliness and decorum. The current revival and popularity of rhetoric in philosophy often takes the form of a programme to replace science by conversation as the model for community and rationality and the hope for democracy.

See also: Language, medieval theories of; Language, Renaissance philosophy of; Legal reasoning and interpretation

EUGENE GARVER

References and further reading

Aristotle (*c.* mid 4th century BC) *On Rhetoric: A Theory of Civic Discourse*, trans. with introduction, notes and appendices by G.A. Kennedy, New York and Oxford: Oxford University Press, 1991.(The most readable and reliable English translation available.)

Augustine (*c.*396-427) *De Doctrina Christiana* (*On Christian Doctrine*), trans. D.W. Robertson, Jr, Indianapolis, IN: Bobbs-Merrill, 1958.(The most influential work explicitly adapting rhetoric to Christian religious and theological purposes.)

Austin, J.L. (1962) *How to do things with words*, Cambridge, MA: Harvard University Press.(A recovery of the rhetorical dimension of language under the title 'performative utterance'.)

Barthes, R. (1970) 'L'Ancienne Rhetorique: Aide-mémoire', *Communications* 16 (1970): 172-229; 'The Old Rhetoric: An Aide-Mémoire', in *The Semiotic Challenge*, trans. R. Howard, New York: Hill & Wang, 1988, 11-94.(Designed as a summary of the history of rhetoric for students of contemporary literary criticism.)

Burke, K. (1950) *A Rhetoric of Motives*, New York: Prentice Hall.(A cataloguing of the formal techniques for persuasion.)

Burke, K (1965) *Permanence and Change: An Anatomy of Purpose*, Indianapolis, IN: Bobbs-Merill.(An analysis of uses of poetry to solve ethical conflicts through rhetoric as 'an art of appeal'.)

Cicero (55 BC) *de Oratore*, trans. E.W. Sutton and H. Rackham, London: Loeb Classical Library, 1959.(Cicero's most philosophical treatment of rhetoric. Not a practical handbook but a dialogue on the place of rhetoric in ethical and political life.)

Conley, T.M. (1990) *Rhetoric in the European Tradition*, New York and London: Longman; Chicago, IL: University of Chicago Press, 1994.(A textbook, full of summaries and bibliographies; the best one-volume

history of rhetoric.)

Garver, E. (1994) *Aristotle's Rhetoric: An Art of Character*, Chicago, IL: University of Chicago Press.(A philosophical analysis of Aristotle's Rhetoric.)

Hume, D. (1748, 1751) *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1966.(A moral theory based on the inference patterns of epideicic rhetoric.)

Kant, I. (1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1929.(A critical philosophy rooted in justificatory rhetoric.)

Kennedy, G. (1980) *Classical rhetoric and its Christian and secular tradition from ancient to modern times*, Chapel Hill, NC: University of North Carolina Press.(One of a series of informative histories of rhetoric by this author.)

McKeon, R. (1987) *Rhetoric: Essays in Invention and Discovery*, Woodbridge, CT: Oxbow Press.(Contains the influential 'Rhetoric in the Middle Ages' and other articles by an important but difficult historian and philosopher of rhetoric.)

Mill, J.S. (1859) *On Liberty*, New York: Bobbs-Merill, 1956.(Chapter 2, in particular, concerns the power of controversy and discussion to reveal truth.)

Perelman, C. and Olbrecht-Tyteca, L. (1969) *The New Rhetoric: A Treatise on Argumentation*, Notre Dame, IN: University of Notre Dame Press.(A treatment of the argumentative force of figures of speech and the topics or *loci communes*.)

Plato (*c*.390-347BC) *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1971.(All Platos's dialogues collected in one volume, with an introduction and prefatory notes.)

Ramus, P. (1574) *Logikē*, trans. M. Roll, London: Scholar Press, 1966.(The only edition available in English of this highly influential work.)

Richards, I.A. (1936) *The Philosophy of Rhetoric*, New York and London: Oxford University Press.(Proposes a new rhetoric that studies misunderstandings and their remedies.)

Ricoeur, P. (1977) *The Rule of Metaphor*, Toronto, Ont.: University of Toronto Press. (A historical and philosophical investigation of the interrelations between rhetorical and poetic approaches to language and literature.)

Trimpi, W. (1983) *Muses of One Mind: The Literary Analysis of Experience and Its Continuity*, Princeton, NJ: Princeton University Press.(Traces the competition and joint contributions of philosophy and rhetoric to literary theory, from the Greeks to the Renaissance.)

Toulmin, S. (1964) *The Uses of Argument*, Cambridge: Cambridge University Press. (A re-discovery, under the name 'inference warrants', of the topics of classical rhetoric.)

Vico, G. (1744) *The New Science*, trans. T. Bergin and M. Fish, Ithaca, NY: Cornell University Press, 1984.(A defence of rhetoric against philosophy, and in particular the epistemic and ethical value of figurative over literal language.)

White, H. (1973) *Metahistory: The Historical Imagination in Nineteenth-Century Europe*, Baltimore, MD and London: Johns Hopkins University Press.(Uses of Kenneth Burke's theory of tropes to analyse historical narratives and philosophies of history.)

Richard of Middleton (*c.*1249-1302)

Richard was a Franciscan philosopher and theologian. In general he followed the tradition flowing from Bonaventure, although on some questions he sided with Thomas Aquinas. However, there is also a strong anti-Thomist reaction in his work. Many of the questions raised in the condemnations of 1277 at Paris and Oxford are central in Richard's works. His answers often echo Bonaventure, William of Ware and Matthew of Aquasparta; yet his argumentation carries his personal stamp and shows him deeply engaged with the definitions and arguments of the authors of his own era.

Richard studied in Paris during the intellectually exciting years 1276-87, probably under William of Ware and Matthew of Aquasparta. In general, the *Doctor solidus* (Solid Doctor) followed the tradition flowing from Bonaventure, although on some questions he sided with Thomas Aquinas, at least in his sober commentary on Peter Lombard's *Sentences* (probably begun in 1281 and completed in 1284). However, while regent master of the Franciscan house of studies in Paris from 1284-87, his *Quodlibeta (Quodlibetal Questions)* manifest a strong anti-Thomist reaction, as do his earlier *Quaestiones disputatae (Disputed Questions)* (disputed after 1277 but before the *Sentences* commentary). In 1295 he was elected Provincial Minister of France, and he died at Reims in 1302.

When Aquinas, for instance, argued that philosophical reason cannot show the impossibility of an eternal creation, Richard was quick to point out the ambiguity hidden in Aquinas' portrait of creation. When the latter speaks of *creatio ex nihilo* (creation out of nothing), he correctly stresses the absolute dependence of creatures on God for existence. Richard charged, however, that Aquinas' definition had excluded any consideration of the issue of duration. By neglecting a treatment of the temporal or eternal character of creation, Aquinas, according to Richard, had built a prejudice in favour of the possibility of an eternal world into his very definition of the term 'creation' (see Eternity of the world, medieval views of).

Richard likewise rejects Aquinas' theory of individuation by matter. For him, the singular is singular by its proper essence, since this concrete essence is of its very nature indivisible. He thus follows in the footsteps of Bonaventure, for whom the ultimate cause of individuation is the very 'act of existence'. Richard agrees with the opinion that 'by the very fact that a thing actually exists, it is numerically one… Because a being in actual existence is composed of essence and actual existence, its being one adds nothing positive to such an actually existing being but only a negation of division' (*Sententiarum* I, d. 24, a. 1, q. 2).

By holding that in each of the elements there is only one substantial form, Richard disagreed with Bonaventure. However, as he moved on to consider the substantial forms in more complex beings, and especially in man, he reaffirmed the arguments of Bonaventure and John Pecham for the plurality of substantial forms. Richard could not see how, for instance, those who argued for the unicity of substantial form or the dominance of the intellective or rational soul in man could explain why the accidents of the body remain the same in number after the recession of the intellective soul at death. The unicity-of-form theory, defended by Aquinas, Giles of Lessines, Giles of Rome and many others, is also incapable of explaining human generation. Richard taught that if there is to be any meaning to the phrase 'man generates man', then the vegetative and sensitive forms of the human being must exist before, and must continue to exist after, the arrival of the rational soul created directly by God. There must, therefore, be a plurality of substantial forms in composite beings.

In his theory of universals and in his theory of knowledge, Richard comes close to Aquinas: 'Whatever God has created is singular. A universal exists only through an act of the intellect' (*Sententiarum* I, d. 36, a. 1, q. 1). Intellectual knowledge arises from the singulars of experience by means of the abstraction performed by the agent intellect in human beings. Richard rejects the need for special divine illumination, defended by Bonaventure, to account for our intellectual knowledge. Human beings know, and they know by means of their own proper intellectual efforts.

This Aristotelian bent continues in Richard's discussion of man's knowledge of God. There is no innate or immediate knowledge of God. He can be known only in starting with creatures and by our reasoning about their origins or purposes: 'We only come to a knowledge of spiritual beings by means of sensible things… and to a

knowledge of God only through sensible and intelligible effects…' (*Sententiarum* II, d. 25, a. 5, q. 1).

Richard's teaching on the primacy of the will returns him to the Augustinian camp. 'The will', he declares, 'is unqualifiedly more noble than the intellect' (*Sententiarum* II, d. 24, a. 1, q. 5). Richard argues that it is more noble to love God than to understand him, since such understanding, if not linked with love, keeps us aloof from God' (*Quaestiones disputatae* q. 10, a. 3). The nobility of our will, furthermore, is found in its freedom. Unlike the intellect, which can be forced by evidence, the will remains free except in the case of its end, the good. This object the will must pursue by the radical tendency built into it by its creator. However, the will is free in regard to all the means leading to that predetermined goal. There is no parallel between the case of the intellect, in connection with which we are forced to accept a logically drawn conclusion once we have accepted the premises, and the will in regard to those things that lead to our end. Even when our intellect shows us the best means to our end, we do not have to adopt them: 'For although the intellect, like a servant with a lamp, points out the way, the will, like the master, makes the decisions and can go in any direction it pleases' (*Sententiarum* II, d. 38, a. 2, q. 4).

Richard, with a questioning spirit, most often follows the tradition of Bonaventure. His commentary on the *Sentences* in particular represents a solid, searching mind in an age of great philosophical challenge and confusion.

See also: Aquinas, T.; Augustinianism; Bonaventure; Giles of Rome

STEPHEN F. BROWN

List of works

Richard of Middleton (1284-7) *Quodlibeta quaestiones octuaginta (Eighty Quodlibetal Questions)*, Brescia, 1590; repr. Frankfurt: Minerva, 1963.(The reprinted edition also includes the *Sententiarum*.)

Richard of Middleton (1281-4) *Super IV libros Sententiarum Petri Lombardi quaestiones subtilissimae (Questions on the Four Books of Peter Lombard's Sentences)*, Brescia, 1591; repr. Frankfurt: Minerva, 1963, 4 vols.(The reprinted edition also includes the *Quodlibeta*.)

Richard of Middleton (1277-84) *Quaestio de gradu formarum (A Question on the Hierarchy of Forms)*, ed. R. Zavalloni, *Richard de Mediavilla et la controverse sur la pluralité des formes*, Louvain: Éditions de l'Institut Supérieur de Philosophie, 1951, 35-169.(On the hierarchy and plurality of forms.)

Richard of Middleton (1277-84) *Quaestiones disputatae (Disputed Questions)*.(The following few questions have been edited: *Quaestio de unitate formae* (Question on the Unity of Form), ed. R. Zavalloni, *Richard de Mediavilla et la controverse sur la pluralité des formes*, Louvain: Éditions de l'Institut Supérieur de Philosophie, 1951, 173-80; *Quaestio de immortalitate animae* (Question on the Immortality of the Soul), ed. S. Vanni Rovighi, *L'immortalità dell'anima nel maestri francescani del secolo XIII*, Milan: Pubblicazioni dell'Università Cattolica del Sacro Cuore, 1936, 349-70; *Quaestio disputata: Utrum voluntas angeli possit movere se ipsam (A Disputed Question: Can an Angel's Will Move Itself?)*, ed. O. Lottin, *Psychologie et morale aux XIIe et XIIIe siécles*, Louvain: Abbaye du Mont César, 1942, vol. I, 293-9; *Quaestio: Utrum angelus vel homo naturaliter intelligat verum creatum in veritate aeterna (A Question: Can an Angel or a Human Being Naturally Understand Created Truth in Eternal Truth?)*, ed. Patres Franciscani, *De humanae cognitionis ratione anecdota*, Quaracchi: Ad Claras Aquas, 1883, 221-45.)

References and further reading

Cova, L. (1984) *Originale peccatum e concupiscentia in Riccardo di Mediavilla: vizio ereditario e sessualità nell'antropologia teologia del XIII secolo*, Rome: Edizioni dell'Ateneo.(Studies man's hereditary corruption and sensuality according to Richard.)

Cunningham, F.A. (1970) 'Richard of Middleton, O.F.M. on *esse* and Essence', *Franciscan Studies* 30: 49-76. (Richard's views on essence and existence.)

Henninger, M.G. (1994) 'Hervaeus Natalis (b. 1250/60; d. 1323) and Richard of Mediavilla (b. 1245/49; d. 1302/07)', in J.J.E. Gracia (ed.) *Individuation in Scholasticism: The Later Middle Ages and the Counter-Reformation, 1150-1650*, Albany, NY: State University of New York Press, 299-318.(Compares Richard's views with those of Hervaeus.)

Hocedez, E. (1916) 'Les "Quaestiones disputatae" de Richard de Middleton', *Recherches de science religieuse* 6: 493-513.(Provides a clear portrait of the content of Richard's disputed questions.)

Hocedez, E. (1925) *Richard de Middleton: sa vie, ses oeuvres, sa doctrine (Richard of Middleton: His Life, His Works and His Teaching)*, Louvain: Spicilegium Sacrum Lovaniense, and Paris: Librairie Ancienne Honoré Champion, Edouard Champion.(Provides a solid overview of Richard's life, works and teaching.)

Lampen, W. (1930) ‘Richard de Mediavilla’, *La France Franciscaine* 13: 388-90. (Updates two earlier articles concerning nationality of Richard.)

Lippens, H. (1944) ‘De modo celebrandi Capitulum deque observantia Regulae in Provincia Franciae post annum 1517 (iuxta documenta inexplorata)’, *Archivum Franciscanum Historicum* 37: 3-47.(Concerns the election of Richard as Provincial Minister of France: notes that this election does not determine his nationality.)

Sharp, D.E. (1930) ‘Richard of Middleton’, *Franciscan Philosophy at Oxford in the Thirteenth Century*, London: Oxford University Press; repr. Farnborough: Gregg Press, 1966, 211-76.(Summarizes the main doctrinal positions.)

Van Veldhuijsen, P. (1990) ‘Richard of Middleton contra Thomas Aquinas on the Question Whether the Created World Could Have Been Eternally Produced by God’, in J.B.M. Wissink (ed.) *The Eternity of the World in the Thought of Thomas Aquinas and his Contemporaries*, Studien und Texte zur Geistesgeschichte des Mittelalters 24, Leiden: Brill, 69-81.(Richard’s views on eternity and creation.)

Zavalloni, R. (1951) *Richard de Mediavilla et la controverse sur la pluralité des formes (Richard of Middleton on the Plurality of Forms)*, Philosophes médiévaux 2, Louvain: Éditions de l’Institut Supérieur de Philosophie. (Details Richard’s role in the controversy over the unicity and plurality of forms in man.)

Richard of St Victor (d. 1173)

Richard is most famous for his contemplative doctrine, which is based on a biblical anthropology that involves a philosophical psychology and noetic theory. Richard's writings should be understood in the context of Hugh of St Victor's programme for a complete theological pedagogy, organized according to the threefold sense of Scripture (literal, allegorical, tropological). Richard's specifically exegetical works include an encyclopedic introduction to the methods of interpreting Scripture, the Liber exceptionum, and important commentaries on the Apocalypse and Ezekiel. Like Hugh, he stresses that the literal sense of Scripture is the foundation of its spiritual senses.

Richard of St Victor was a native of Scotland or Ireland. He became an Augustinian Canon Regular at the Abbey of St Victor in Paris, where he was a student of Hugh of St Victor. He later served as sub-prior and prior of the monastery and was the master of its famous school. He died in 1173.

Richard's most important speculative work is *De trinitate (On the Trinity)*. In the manner of Anselm of Canterbury, through a 'faith seeking understanding' he discovers 'necessary reasons' for God's existence, the divine attributes and the Trinity. Richard does not 'demonstrate' the Trinity, but through a conceptual logic of perfection considers the divine being as it is revealed in Scriptures. The reasons for the Trinity are intrinsic to God's 'necessary' being and attributes. God is love or Goodness itself, which by nature is wholly self-giving. God is also Glory itself and Felicity itself, which likewise communicate themselves wholly. This threefold communicative fullness requires that there be one who gives himself exhaustively, one who receives and gives back himself exhaustively and one who from both receives exhaustively. A further existential category of 'one who gives nothing and receives nothing' is inconceivable. Therefore the circle of self-giving and receiving is consummated in three 'persons', each of whom possesses divine existence incommunicably and independently. Richard's account of the Trinity had considerable influence on Alexander of Hales, Bonaventure and John Duns Scotus (see Trinity).

According to Richard's contemplative teaching, union with God is possible because the soul's rational 'force' is created in the 'image' of God, and its affective 'force' is made in the 'likeness' of God. Originally the human body also reflected the divine nature. By the Fall, human reason was corrupted by ignorance, the affections by concupiscence and the body by infirmity and death. The light of divine wisdom heals the soul's ignorance, divine charity heals its concupiscence and the virtues alleviate the body's infirmity, although they cannot prevent its death. Within this dynamic of sin and grace, Richard develops a philosophical psychology of contemplation and spiritual perfection (see Grace).

In *De duodecim patriarchis (The Twelve Patriarchs)*, or *Benjamin minor*, Richard interprets the affiliations among the wives and sons of Jacob tropologically (that is, morally and mystically), as signifying relations among powers and virtues of the soul. In *De arca mystica (The Mystical Ark)*, or *Benjamin major*, he tropologically interprets the features of the Ark revealed to Moses, as signifying cognitive powers of the soul. The soul's imaginative, reasonable and 'intelligent' powers apprehend, respectively, sensible, intelligible and transcendent 'intellectible' realities. These modes of apprehension diversify into six genera of contemplation. The first four are 'in the imagination according to imagination', or wonder at the order and form of sensible realities; 'in the imagination but according to reason', whereby the spirit discerns the principles underlying sensible realities; 'in reason according to imagination', whereby one discovers the similitudes of invisible realities in visible symbols; and 'in reason according to reason', whereby one contemplates purely spiritual created beings (for example, the human soul and its operations, angelic beings). These kinds of contemplation are attainable by human 'industry' aided by God, but the last two depend solely on the light of grace shining in the 'fine point of the intelligence'. The fifth kind, which contemplates the divine unity and attributes, is 'above reason but not without it'. The sixth 'is above and beyond reason' and seemingly against it, inasmuch as it regards mysteries (the Trinity, the Eucharist) that elude the priniciples of human thought.

Richard's philosophical psychology derives ultimately from Neoplatonic sources. The human soul is able to know itself immediately. Its rational and affective 'forces' are reciprocal and not really distinct, but are simply different modes of its operation. Hence love and desire motivate the search for wisdom, and understanding elicits love and desire. The soul's highest cognitive power, the 'intelligence', transcends discursive reasoning and penetrates

spiritual realities intuitively (see Platonism, medieval).

Richard of St Victor's contemplative writings profoundly influenced Bonaventure's *Journey of the Mind into God* and Dante's *Comedy* (see Alighieri, Dante). Bonaventure designates Richard (with Pseudo-Dionysius) as the exemplar of 'anagogic' or mystical theology. Accordingly, in the later Middle Ages Richard was a standard authority among mystical writers.

See also: Hugh of St Victor; Platonism, medieval

KENT EMERY, JR

List of works

Richard of St Victor (before 1173) *Opera omnia*, ed.F. Hugonin and J.-P. Migne, Patrologia Latina, Paris: 1855; repr. 1880, vol. 196.(A re-presentation of the seventeenth century edition prepared under the care of Victorine fathers. The volume contains some spurious works and omits some authentic ones. The text is often faulty.)

Richard of St Victor (before 1173) *De quatuor gradibus violentae caritatis (The Four Degrees of Fervent Charity)*, ed. and trans. G. Dumeige, *Épître à Séverin sur la charité; Le quatre degrés de la violente charité (De quatuor gradibus violentae caritatis)*, Textes philosophiques du moyen âge 3, Paris: Vrin, 1955.(An edition of the Latin texts, with French translations, of a small treatise by a 'Brother Ivo', falsely attributed to Richard, and of *De quatuor gradibus*.)

Richard of St Victor (before 1173) *Liber exceptionum*, ed. J. Châtillon, *Liber exceptionum: Texte critique avec introduction, notes et tables*, Paris: Vrin, 1958.(A critical edition of the Latin text of Richard's methodological introduction to the reading of Scripture.)

Richard of St Victor (before 1173) *De trinitate (On the Trinity)*, ed. J. Ribaillier, Textes philosophiques du moyen âge 6, Paris: Vrin, 1958.(A critical edition of the Latin text of Richard's most important speculative work. A French translation can be found in *La Trinité: Texte latin, introduction et notes*, ed. and trans. G. Salet, Sources Chrétiennes 63, Paris: Éditions du Cerf, 1959, which includes a Latin text with the French translation on facing pages. A reliable English translation of Book III only can be found in *The Twelve Patriarchs, The Mystical Ark, Book Three of The Trinity*, trans. G.A. Zinn, New York: Paulist Press, 1979.)

Richard of St Victor (before 1273) *De duodecim patriarchis (The Twelve Patriarchs)* or *Benjamin minor*, trans. G.A. Zinn, *The Twelve Patriarchs, The Mystical Ark, Book Three of The Trinity*, New York: Paulist Press, 1979.(A reliable English translation, with a good introduction to Richard's spiritual teaching.)

Richard of St Victor (before 1273) *De arca mystica (The Mystical Ark)* or *Benjamin major*, trans. G.A. Zinn, *The Twelve Patriarchs, The Mystical Ark, Book Three of The Trinity*, New York: Paulist Press, 1979.(A reliable English translation, with a good introduction to Richard's spiritual teaching.)

Richard of St Victor (before 1273) Theological Works, in J. Châtillon and W.J. Tulloch (eds), J. Barthelemy (trans.) *Sermons et opuscules spirituels inédits 1: Texte latin, introduction et notes*, Paris: Desclée de Brouwer, 1951.(This volume presents the first edition of Richard's *Exiit edictum* or *De tribus processionibus*, with a French translation. The work gives spiritual interpretations of the processions celebrated on the feasts of the Purification, Palm Sunday and the Ascension.)

Richard of St Victor (before 1273) Theological Works, in J. Ribaillier (ed.) *Opuscules theologiques: Texte critique avec notes et tables*, Textes philosophiques du moyen âge 15, Paris: Vrin, 1967.(A critical edition of the Latin texts of speculative treatises by Richard concerning the power of absolving sins, the spirit of blasphemy, the final judgment, the Holy Spirit, the appropriations of the Trinity, the incarnate Word, the difference between venial and mortal sin and the explanation of difficult passages in Scripture.)

Richard of St Victor (before 1273) Three Treatises, in J. Châtillon (ed.) *Trois opuscules spirituels de Richard de Saint-Victor: Textes inedits accompagnes d'etudes critiques et de notes*, Paris: Études augustiniennes, 1986.(A critical edition of the Latin texts of three previously unedited treatises concerning the inner life of the soul.)

References and further reading

Châtillon, J. (1987) 'Richard de Saint-Victor', *Dictionnaire de spiritualité ascétique et mystique, histoire et doctrine*, Paris: Beauchesne, 1987, 13: 593-654.(An excellent, ample and detailed introduction to Richard's life, works, doctrine and influence, with a full bibliography of publications since 1952, by the leading student of Richard's texts and thought.)

DiLorenzo, R.D. (1982) 'Imagination as the First Way to Contemplation in Richard of St. Victor's *Benjamin minor*', *Medievalia et Humanistica* 11: 77-96.(An interesting study of the high value Richard places on the imagination and on its power to envision suprasensible realities. The essay lays the groundwork for assessing

Richard’s influence on Dante.)

Dumeige, G. (1952) *Richard de Saint-Victor et l’idée chrétienne de l’amour (Richard of St Victor and the Christian Idea of Love)*, Paris: Presses universitaires de France, 1952.(A study of the central importance of the idea of love in every aspect of Richard’s thought, especially important for a critical bibliography of publications on Richard until 1952.)

Lubac, H. de (1959-64) *Exégèse médiévale: Les quatre sens de l’Écriture (Medieval Exegesis: The Four Senses of Scripture)*, Paris: Aubier, 4 vols.(In vol. 3, 361-435, de Lubac treats Richard’s method of biblical interpretation and its place in the history of exegesis.)

Robilliard, J.A. (1939) ‘Les six genres de contemplation chez Richard de Saint-Victor et leur origene platonicienne’ (Richard of St Victor’s Six Kinds of Contemplation and their Platonic Origin), *Revue des sciences philosophiques et théologiques* 28: 229-33.(A brief but important article tracing the Neoplatonic sources of Richard’s genera of contemplation.)

Zinn, G.A. (1977) ‘Personification, Allegory and Visions of Light in Richard of St. Victor’s Teaching on Contemplation’, *University of Toronto Quarterly* 46: 190-214. (A study of Richard’s ‘historical’ personification allegory, distinct from rhetorical techniques of personification and allegory.)

Richard Rufus of Cornwall (d. after 1259)

A thirteenth-century philosopher and theologian, Rufus was among the first Western medieval authors to study Aristotelian metaphysics, physics and epistemology; his lectures on Aristotle's Physics are the earliest known surviving Western medieval commentary. In 1238, after writing treatises against Averroes and lecturing on Aristotle - at greatest length on the Metaphysics - he joined the Franciscan Order, left Paris and became a theologian.

Rufus' lectures on Peter Lombard's Sentences were the first presented by an Oxford bachelor of theology. Greatly influenced by Robert Grosseteste, Rufus' Oxford lectures were devoted in part to a refutation of Richard Fishacre, the Dominican master who first lectured on the Sentences at Oxford. Though much more sophisticated philosophically than Fishacre, Rufus defended the more exclusively biblical theology recommended by Grosseteste against Fishacre's more modern scholasticism.

Rufus' Oxford lectures were employed as a source by Bonaventure, whose lectures on the Sentences were vastly influential. Returning to Paris shortly after Bonaventure lectured there, Rufus took Bonaventure's lectures as a model for his own Parisian Sentences commentary. Rufus' Paris lectures made him famous. According to his enemy Roger Bacon, when he returned to Oxford after 1256 as the Franciscan regent master, his influence increased steadily. It was at its height forty years later in the 1290s, when John Duns Scotus was a bachelor of theology. Early versions of many important positions developed by Duns Scotus can be found in Rufus' works.

1 Natural philosophy

As far as we know, the earliest work we have by Rufus is a set of lectures on Aristotle's *Physics*, presented at Paris shortly before the ban against teaching the *libri naturales* was formally lifted. A series of brief questions on issues raised by the Aristotelian text, they have never been published. The published works they most closely resemble are Roger Bacon's post-1237 question-commentaries.

Rufus' *Physics* commentary is important in part because of his unorthodox approach, often departing from Averroes' interpretation of Aristotle. Those departures Bacon consciously rejects. A case in point is the explanation of projectile motion. Rufus is the first Western scholastic to declare Aristotle's account of projectile motion inadequate. No explanation involving only the medium could be adequate because there can be projectile motion in opposite directions in a uniform medium, and because different projectiles move through the same medium at different speeds. To explain projectile motion, Rufus posits an imprint of violent motion in the projectile, but he does not completely reject Aristotle's account. Arguing that the projector affects the medium as well as the projectile, he explains projectile motion in terms of the effect of violent motion both on the projectile and on the medium. Postulating an imprint of violent motion on the projectile, he also accepts *antiperistasis* as an account of the effect of violent motion on the medium. By contrast, Bacon rejects imprint theory, arguing that there is constant substantial contact between mover and moved object; 'everything that moves is moved by another'. Other contemporaries rejected imprint theory on the grounds that Rufus could not precisely explain what was imprinted. However, Rufus' views had continued currency and were cited, accepted and given a fuller exposition in the fourteenth century by Franciscus de Marchia (see Natural philosophy, medieval).

Among the first medieval Western philosophers to encounter Aristotle's arguments for the eternity of the world, Rufus also presented some of the most cogent counter-arguments. One argument is based on a contradiction between the definitions of 'past' and 'infinity'. It is impossible to traverse an infinity, but it pertains to the nature of the past to have been traversed; therefore, past time cannot be infinite. Rufus presents the argument with characteristic brevity: 'Having been traversed' is incompatible with the definition of 'infinity', but 'having been traversed' belongs to the definition of the past. Therefore, being past is incompatible with the definition of infinity.

This argument, first presented in late antiquity by John Philoponus, is now associated with Immanuel Kant; in medieval philosophy, it is ordinarily ascribed to Bonaventure, who advanced it in 1250 or 1251. It occurs in different versions, some more persuasive than others. Grosseteste, for example, mistakenly seeks to apply it to the future as well as the past, claiming that the argument can be used to show that time could not be infinite *a parte post*. Rufus sees even more clearly than Philoponus that the direction of time is an important part of this argument.

He notices that the argument needs to be based on the fact that the *whole* of past time has been traversed, rather than on the claim that the whole of the past and the future will have been traversed. In his later work he seeks to force his opponents to see that they are committed to the claim that some past days are not now and never were present. By contrast, Philoponus sees this as an argument about the impossibility of completely counting an infinite series, with no particular focus on the direction of time.

Rufus' version of another of Philoponus' arguments is based on our concept of priority. If the number of days before today is infinite, and the number of days before tomorrow is infinite, then the number of days before today is not less than the number of days before tomorrow. Consequently, today does not arrive sooner than tomorrow, which is absurd. Rufus assumes here that unequal infinities are impossible. Following Georg Cantor, modern mathematicians reject this assumption. However, Rufus needs only the uncontroversial claim that mappable infinities are equal: if we postulate beginningless time, the number of days before today and the number of days before tomorrow are mappable infinite series. Rufus might still argue that if the world has no beginning, then we must give up the belief that less time elapses before earlier events than before later events.

Philoponus' original version of this argument is not based on the claim that more time transpires before later events than before earlier events. The absurdities he asks us to reject are mathematical: that it is possible to add to an infinity, or that one infinity can be multiplied by another, so that one infinity would be greater than another by a determinate proportion. By contrast, the absurd conclusion Rufus asks us to reject is that 'today does not come sooner than tomorrow'; he emphasizes the unique properties of time (see Eternity of the world, medieval views of).

2 Epistemology

The fullest statement of Rufus' epistemological views now known is a treatise entitled *Speculum animae (A Mirror of the Soul)*, probably written to explain problems in Aristotelian philosophy to his Franciscan confrères. In this ambitious treatise, Rufus attempts to explain what Aristotle means when he says that 'in some manner the soul is every thing'. Rufus asks in what sense the soul becomes an object when it understands or senses that object. He rejects the view his predecessors based on patristic authorities: the soul is everything because it shares being with rocks, life with animals and understanding with angels. He also rejects a literal interpretation of the dictum. This leaves Rufus in a difficult position, since he holds that species in the soul are really identical with the common natures exemplified in external objects. If the material in the sensory soul combines with the sensible species, why does not the soul become green when it perceives something green? There are two elements in Rufus' reply: one is to postulate a different kind of being for sensible and intelligible species. Sensible species are not natural beings, in the sense that they are not included in the Aristotelian categories; they are neither substances nor accidents. Because they are different kinds of entities, when they combine with the soul, what is produced is not the object itself but cognition. The second element of Rufus' reply is to argue that what is formally distinct may be really identical. Species have the same real nature, but are not formally identical with, or predicable of, the objects they represent in the soul. This safeguards the claim that what we perceive is really the same as external objects; in some sense the soul really is all things.

3 Metaphysics

Postulating a kind of identity which permits real but not formal predication is a conceptual tool which Rufus employs when discussing a variety of philosophical topics, for example, the problem of individuation. Like Duns Scotus in his *Metaphysics* commentary, Rufus postulates individual forms to explain individuation. Individual forms are really, but not formally, the same as specific forms. Specific forms are principles of shared identity; they pertain to common natures capable of instantiation (*multiplicabilis*). By contrast, individual forms pertain to the same individual natures as are actually instantiated (*actu multiplicata*).

Rufus' arguments against alternative theories were initially more influential than his own views. He holds that the cause of individuation cannot be an accident or an aggregation of accidents, since individual primary substances are ontologically prior to accidents. Though he allows a role for matter as an occasional cause of individuation, Rufus argues that even determinate matter could not by itself be the principle of individuation. Being an individual means being distinct and united, both of which are functions of form, the active principle of substance, not matter, the passive principle.

Holding that individual forms added to an aggregate of matter and specific form must be the principle of individuation, Rufus denies that the ultimate constituents of individuals are knowable. He is not sure whether what is added to the common nature can be located within an Aristotelian category. He suggests that, strictly speaking, the cause of individuation may be neither a substance nor an accident. Identifying individual forms as perfections of the specific form, he suggests that they may be substantial without being substances. Specific and individual forms provide different degrees of unity: specific unity is less than individual unity and greater than generic unity.

Like Rufus' views on individuation, his argument for the existence of God was accepted and modified by John Duns Scotus. Rufus rejected Anselm's famous ontological argument as sophistical (though subtle) (see Anselm of Canterbury). In its place he advanced a modal argument based on the concept of God as an independent being (*a se et non ab alio*). The existence of independent beings is either necessary or impossible. Therefore, if an independent being can exist, it does exist. Rufus employs logically sophisticated arguments to show that an independent being can exist.

4 Logic

Rufus' works are all characterized by rigour and logical subtlety, but whether he wrote any independent treatise in logic is not yet clear. One such treatise, called *Abstractiones*, has been plausibly attributed to him. It is a guide to the proper exposition of ambiguous propositions, often propositions including syncategorematic words such as 'only' and 'every' or modal operators such as 'possible' and 'necessary'. If this attribution is correct, Rufus was known to fourteenth century Franciscans such as William of Ockham as the 'Magister Abstractionum'.

5 Influence

Rufus' importance has long gone unrecognized, in part because he preferred not to take credit for his own work and in part because, unlike his contemporaries, he provided lengthy quotations of the positions he treated seriously. Since his own views were often stated briefly, historians who overlooked his critical bent saw him as a derivative figure. Now that Bonaventure's borrowing from Rufus has been discovered, and we are beginning to appreciate the significance of citations by Robert Grosseteste (Magister Richardus), John Duns Scotus (*Doctor antiquus*) and Franciscus de Marchia (Richardus), the question of Rufus' influence will have to be reconsidered. The first prerequisite is a critical edition of his works.

See also: Aristotelianism, medieval; Bacon, R.; Eternity of the world, medieval views of; Grosseteste, R.

REGA WOOD

List of works

As noted, there is no critical edition of Richard Rufus' works, most of which survive only in a single manuscript. Extant works include commentaries on the *Physics* and *Metaphysics* of Aristotle; two treatises against Averroes (*De ideis (On Ideas)* and *De causa individuationis (On the Basis of Individuation)*; disputed questions; two commentaries on Peter Lombard's *Sentences*, lectures given at Oxford and Paris; and an epistemological treatise (*Speculum animae (A Mirror of the Soul)*). Lost or unidentified is a commentary on Aristotle's *Meteorology*, cited by Bartolomeus Anglicus.

References and further reading

Callus, D. (1939) 'Two Early Oxford Masters on the Problem of Plurality of Forms: Adam of Buckfield - Richard Rufus of Cornwall', *Revue Néoscolastique de Philosophie* 42: 425-9.(Notes that Rufus does not support the doctrine of plurality of forms.)

Dales, R.C. (1971) 'The Influence of Grosseteste's *Hexaemeron* on the *Sentences* Commentaries of Richard Fishacre, O.P. and Richard Rufus of Cornwall, O.F.M.', *Viator* 2: 270-300.(Considers Rufus a derivative figure.)

de Libera, A. (1983) 'La littérature des *Abstractiones* et la tradition logique d'Oxford' ('The Abstractiones Literature and the Oxford Logical Tradition'), in O. Lewry (ed.) *The Rise of British Logic*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 63-114.(Discusses the evidence for identifying Richard Rufus as Richardus Sophista, the Magister Abstractionum.)

Doucet, V. (1948) 'Prolegomena', in *Summa Halesiana* IV part 1, Quaracchi: Typographia Collegii S. Bonaventurae, 243-5.(A discussion of Rufus' life and works, not rendered obsolete by the works of Raedts or

Wood.)
Gál, G. (1950) ‘Commentarius in Metaphysicam Aristotelis cod. Vat. lat. 4538, fons doctrinae Richardi Rufi’(‘Cod. Vat. lat. 4538, a Commentary on Aristotle’s *Metaphysics* and a Richard Rufus Source’), *Archivum Franciscanum Historicum* 43: 209-42. (Concludes prematurely on the basis of Rufus’ misleading references to this as the work of a secular author that Rufus is not the author of the *Metaphysics* commentary. Provides an outline of the contents of the work and publishes excerpts. Best introduction to the *Metaphysics* commentary.)
Gál, G. (1956) ‘Viae ad exsistentiam Dei probandum in doctrina Richari Rufi’ (’Proofs for the Existence of God in Richard Rufus’), *Franziskanische Studien* 38: 177-202.(Publishes substantial excerpts from Rufus’ Oxford commentary and briefer excerpts from Assisi 138. Excellent study of Rufus’ Oxford theology lectures, which does not include Book IV of the *Sentences*.)
Gál, G. (1975) ‘Opiniones Richardi Rufi Cornubiensis a Censore Reprobatae’ (’Richard Rufus of Cornwall’s Censored Opinions’), *Franciscan Studies* 35: 136-93.(Correctly attributes the *Metaphysics* commentary to Rufus and documents his influence on Scotus. As Bacon notes, Rufus’ Paris theology lectures brought him censure as well as fame. Gál prints substantial excerpts from the Paris commentary as well as marginal notes criticizing Rufus, and points outs that many of the censured opinions were not asserted but recited. Of special interest are Rufus’ views on relations and on the human soul. Best introduction to Rufus’ Paris theology lectures.)
Noone, T. (1987) *An Edition and Study of the Scriptum super Metaphysicam, Book 12, Dist. 2: A Work Attributed to Richard Rufus of Cornwall*, Ottawa, Ont.: National Library of Canada, 1989. (Correctly attributes the *Metaphysics* commentary to Rufus and publishes substantial portions of the text. A good exploration of the relation to Robert Grosseteste.)
Noone, T. (1989) ‘Richard Rufus of Cornwall and the Authorship of the *Scriptum super Metaphysicam*’, *Franciscan Studies* 49: 79-80.(Notes Adam Buckfield’s debt to Richard Rufus; despite statements in early Franciscan sources that Rufus was a Parisian master when he joined the Order, suggests that his pre-Franciscan *Metaphysics* commentary originated in Oxford.)
Pelster, F. (1926) ‘Der älteste Sentenzenkommentar aus der Oxforder Franziskanerschule’, *Scholastik* 1: 50-80. (Establishes that Rufus was the author of the first surviving Oxford *Sentences* commentary by a bachelor of theology, and also the first Franciscan commentary from Oxford.)
Pelster, F. (1950) ‘Richardus Rufus Anglicus O.F.M. (c. 1250), ein Vorläufer des Duns Scotus in der Lehre von Wirkung der priestlichen Lossprechung’, *Scholastik* 25: 550. (Notes Rufus’ rejection of priestly character and his influence on Scotus.)
Plevano, R. (1993) ‘Richard Rufus of Cornwall and Geoffrey of Aspall: Two Questions on the Instant of Change’, *Medioevo* 19: 167-232.(The only recent discussion of Rufus’ disputed questions, includes an edition.)
Raedts, P. (1987) *Richard Rufus of Cornwall and the Tradition of Oxford Theology*, Oxford: Clarendon Press.(A survey of the literature on Rufus’ life and works published before 1987, this book deals chiefly with Rufus’ Oxford lectures. Rufus is compared with Richard Fishacre and Robert Grosseteste; no notice is taken of another major source, Alexander of Hales.)
Wood, R. (1992a) ‘Richard Rufus of Cornwall and Aristotle’s Physics’, *Franciscan Studies* 53: 247-81. (Presents the evidence for attributing the *Physics* commentary preserved in Erfurt, Quarto 312 to Rufus; publishes his account of projectile motion.)
Wood, R. (1992b) ‘Richard Rufus on Creation: The Reception of Aristotle’s Physics in the West’, *Medieval Philosophy and Theology* 2: 7-30.(Discusses the development of Rufus’ view on the eternity of the world from 1235-55.)
Wood, R. (1994) ‘Richard Rufus: Physics at Paris before 1240’, *Documenti e Studi sulla Tradizione Filosofica Medievale* 5: 87-127.(Summarizes Wood’s earlier work on the *Physics* commentary. Concentrates on Rufus’ discussion of place. Trying to answer the question how the heavens can be in a place, Rufus argues that in its most general sense place indicates a fixed relation of a surface to the center of the earth.)
Wood, R. (1995) ‘Richard Rufus’ *Speculum animae*: Epistemology and the Introduction of Aristotle in the West’, in A. Speer (ed.) *Die Bibliotheca Amploniana ihre Bedeutung im Spannungsfeld von Aristotelismus, Nominalismus und Humanismus*, Miscellanea Mediaevalia 23, Leiden: Brill, 86-109.(Identifies the manuscript sources of the *Speculum* and summarizes its contents.)
Wood, R. (1996a) ‘Richard Rufus and English Scholastic Discussion of Individuation’, in J. Marenbon (ed.), *Aristotle in Britain in the Middle Ages*, Rencontres de Philosophie Médieévale 5, Turnhout: Brepols, 117-143. (Shows the influence of Rufus’ theories of individuation prior to Scotus; argues against accounts of English scholasticism which minimize continental influences.)
Wood, R. (1996b) ‘Angelic Individuation: According to Richard Rufus, St. Bonaventure and St. Thomas

Aquinas', in *Individuum und Individualität in Mittelalter*, ed. A. Speer, *Miscellanea Mediaevalia* 24: 209-29. (Shows Bonaventure's borrowing from Rufus, and presents Aquinas' views in the context of Rufus' Aristotelianism.)

Zimmermann, A. (1992) 'Some Aspects of the Reception of Aristotle's *Physics* and *Metaphysics* in the Thirteenth Century', in M. Jordan and K. Emery (eds) *Ad Litteram*, Notre Dame, IN: University of Notre Dame Press.(Notes Bacon's use of Rufus' *Metaphysics* commentary. Includes a discussion of Rufus on the subject of metaphysics.)

Ricoeur, Paul (1913-)

Paul Ricoeur is one of the leading French philosophers of the second half of the twentieth century. Along with the German philosopher Hans-Georg Gadamer, Ricoeur is one of the main contemporary exponents of philosophical hermeneutics: that is, of a philosophical orientation which places particular emphasis on the nature and role of interpretation. While his early work was strongly influenced by Husserl's phenomenology, he became increasingly concerned with problems of interpretation and developed - partly through detailed inquiries into psychoanalysis and structuralism - a distinctive hermeneutical theory. In his later writings Ricoeur explores the nature of metaphor and narrative, which are viewed as ways of creating new meaning in language.

1 Early work

Born in Valence in 1913, Ricoeur began his philosophical career at a time when phenomenology and existentialism were influential in French intellectual circles. As a student in Paris in the late 1930s, and subsequently as a prisoner of war in Germany during the Second World War, Ricoeur read the work of Husserl, Heidegger, Jaspers and Marcel. Following the war, Ricoeur and Mikel Dufrenne published a lengthy study of Jaspers, *Karl Jaspers et la philosophie de l'existence* (1947); in the same year, Ricoeur published his own comparative study of Jaspers and Marcel. Ricoeur also translated Husserl's *Ideen I* into French, thereby establishing himself as a leading authority on phenomenology (see Phenomenological movement).

In 1948 Ricoeur was elected to a chair in the history of philosophy at the University of Strasbourg. During the following decade Ricoeur sought to develop an ambitious project on the 'philosophy of the will': that is, a systematic reflection on the affective and volitional aspects of human existence. In the first volume of this project, *Freedom and Nature* (1950), Ricoeur employed a phenomenological method to explore the interplay between the voluntary and the involuntary features of life, such as the way in which action is mediated by the organs of the body and by a range of factors which lie beyond the individual's control. The second volume of the philosophy of the will was published in 1960 as two separate books: *Fallible Man* and *The Symbolism of Evil*. In these books Ricoeur moved away from a strict phenomenological method and began to develop a more interpretative approach to symbols and myths, as a way of gaining access to problems of human fallibility and fault. Ricoeur sketched the plan for a third volume which would be concerned with the 'poetics of the will', but postponed this project while he turned his attention to other matters.

2 Psychoanalysis and structuralism

In 1957 Ricoeur was elected Professor of Philosophy at the Sorbonne where he remained until 1966, when he was appointed Dean at the University Paris X, Nanterre - a position he currently holds in conjunction with Professorships in the Department of Philosophy and the Divinity School at the University of Chicago. In the late 1950s and early 1960s, the intellectual climate in Paris was changing in significant ways: there was a growing interest in psychoanalysis, and structuralist ideas were beginning to have a significant impact. Ricoeur responded to this change by undertaking a detailed inquiry into the work of Freud and by engaging critically with some of the key texts underpinning the structuralist approach (see Psychoanalysis, methodological issues in; Structuralism).

Freud and Philosophy appeared in 1965 and is one of Ricoeur's best-known works. Ricoeur provides a thorough analysis of the development of Freud's work, with the aim of bringing out its philosophical relevance. Unlike philosophical orientations (like phenomenology) which take the contents of consciousness as a starting point, Freud begins by questioning the primacy of consciousness and looking elsewhere, at the level of the unconscious, for the key processes that shape psychic life. But, as Ricoeur shows, Freud's postulation of unconscious processes is inseparable from the activity of interpretation. Psychoanalysis is, on Ricoeur's account, a form of hermeneutics in which the analyst employs a specialized set of rules, assumptions and models in order to make sense of actions and utterances that are initially difficult to understand. Hence psychoanalysis should be regarded, not as a natural science like physics or chemistry, but rather as an interpretative discipline similar in character to history and literary criticism; it is a discipline concerned with deciphering symbols and signs in relation to the expression of desire.

Ricoeur's critical discussions of structuralism appeared in a series of essays which were brought together in a

volume entitled *The Conflict of Interpretations* (1969). There were many thinkers in France and elsewhere who argued that the principles of structural linguistics - especially as elaborated by the Swiss linguist Ferdinand de Saussure - provided a generalizable model that could be used for the analysis of social and cultural phenomena (see Structuralism in linguistics). Ricoeur had serious reservations about this line of argument. He wanted to show that, while structuralist methods of analysis could be useful for certain purposes, they were based on a number of assumptions which limited their validity. Hence structuralist methods could not provide a comprehensive and self-sufficient approach; they had to be integrated into a broader hermeneutical theory (see Hermeneutics).

Ricoeur developed this argument through a rigorous analysis of the work of some of the authors who were influential in the rise of structuralism. This included the work of structural linguists, such as Saussure, Jakobson and Hjelmslev, as well as the work of the anthropologist Claude Lévi-strauss - one of the few thinkers who explicitly and consistently described himself as a structuralist. Through a brilliant analysis of Lévi-Strauss's writings on myth, Ricoeur showed that Lévi-Strauss was constantly overstepping the limits of validity of the structuralist approach, ending up in what Ricoeur described as 'a Kantianism without a transcendental subject'. Ricoeur also argued that the structuralist analysis of myth was at best a partial account, since it neglected, among other things, the ways in which myths were embedded in historical traditions that were handed down from one generation to the next.

3 Text and interpretation

Ricoeur's critique of Lévi-Strauss and others remains one of the best accounts available of the strengths and weaknesses of the structuralist approach. It also formed the backcloth against which Ricoeur developed his own hermeneutical theory. In a short text called *Interpretation Theory* (1976) and in a series of essays reprinted in *Hermeneutics and the Human Sciences* (1981), Ricoeur worked out a systematic theory of interpretation based on a distinctive notion of the text. A text, according to Ricoeur, is an instance of written discourse; it involves four forms of 'distanciation' which differentiate it from the conditions of speech. First, in spoken discourse there is an interplay between the event of saying (the utterance) and the meaning of what is said; but in the case of written discourse, the event of saying is eclipsed by the meaning of what is said: it is the meaning which is inscribed in the text. Second, whereas in spoken discourse there is an overlap between the intention of the speaking subject and the meaning of what is said, in the case of written discourse these two dimensions of meaning drift apart: the meaning of the text does not coincide with what the author meant. Third, whereas spoken discourse is addressed to a specific recipient, written discourse is addressed to an unknown audience and potentially to anyone who can read. Fourth, whereas the shared circumstances of the speech situation provide some degree of referential specificity for spoken discourse, in the case of written discourse these shared circumstances no longer exist.

These various characteristics of the text provide the basis upon which Ricoeur elaborates his theory of interpretation. Like earlier thinkers within the tradition of hermeneutics, Ricoeur treats the text as a privileged object of interpretation. But unlike earlier thinkers, Ricoeur argues that the process of interpretation can be facilitated and enriched by the use of explanatory methods of analysis. Since a text is a structured totality of meaning which has been distanced from its conditions of production, it can be analysed fruitfully and legitimately by means of structuralist methods. But this type of analysis can never be an end in itself. It can only be a step along the path of interpretation - that is, a partial contribution to the broader hermeneutical task of unfolding the 'world' of the text and the significance it has for its readers.

While Ricoeur's theory of interpretation is formulated with regard to texts, he argues that it can be extended to non-textual phenomena like action (see *Hermeneutics and the Human Sciences* (1981) and *From Text to Action* (1986)). Ricoeur's argument is that action involves the four forms of distanciation characteristic of written discourse; and hence, for the purposes of analysis, action can be treated as a text. One advantage of this approach is that it enables Ricoeur to propose a novel solution to the problem of the relation between explanation and understanding in the social sciences. Just as the interpretation of texts can be facilitated by the structuralist analysis of their contents, so too the understanding of action can be enriched by an explanatory account. Hence explanation and understanding are not necessarily opposed to one another, as some philosophers of social science have suggested (see Social science, contemporary philosophy of). Rather, the explanation of action can be treated as an integral part of a process of understanding.

4 Metaphor and narrative

Throughout his writings Ricoeur has been interested in the phenomena of creativity and imagination, and in the 1970s and 1980s he pursued this interest through his extensive studies of metaphor and narrative. *The Rule of Metaphor* was published in 1975, and the three volumes of *Time and Narrative* appeared in 1983, 1984 and 1985. Metaphor and narrative are particularly interesting for Ricoeur because they are sites of creativity in language: they are works of displacement and synthesis in which the productive imagination is expressed.

On Ricoeur's account, the creative character of metaphor stems from the fact that a metaphorical utterance sets up a tension between two terms in a sentence through the violation of a linguistic code. The metaphor emerges through a creative semantic pertinence which reduces the tension established by the incongruous attribution. Metaphor thus operates not simply at the level of words, but rather at the level of the sentence. The emergent meaning can be grasped only through a constructive interpretation, an imaginative restructuring of semantic fields, which makes sense of the sentence as a whole. The creative character of metaphor does not destroy the referential dimension of language but rather endows metaphor with a new kind of referential power: the power to redescribe reality (see Metaphor).

In the case of narrative, the semantic innovation consists in the invention of a plot. By bringing together characters, goals, causes and chance, the plot of a narrative creates something new: a new pattern in the organization of events. To understand this pattern, one must use the imagination to grasp the diverse elements of the plot, integrating them into a complete and intelligible whole. Moreover, just as metaphor has the referential power to redescribe reality, so too narrative has what Ricoeur calls, following Aristotle, a 'mimetic function'. The invention of plots - or, more generally, the telling of stories - is the means by which we refashion the field of human action and refigure the temporality of human existence. In this respect, suggests Ricoeur, the difference between fictional narratives (such as novels) and historical narratives (that is, works of historical scholarship) is not as great as it might at first appear to be. Of course, historians generally assume that their narratives bear some relation to a past that is presumed to be 'real'. But, at a more general level, both historical narratives and works of fiction are concerned with the refiguration of the temporality of human existence.

As his writings on metaphor and narrative illustrate, Ricoeur addresses problems which are of interest not only to philosophers but also to individuals in other disciplines. His work has had a significant impact on debates in literary criticism, theology, sociology, anthropology and history. There is a consistent vision that animates Ricoeur's work, a vision that emphasizes creativity, interpretation and the social-historical embeddedness of human action and existence. But Ricoeur's work is also distinguished by its generous assessment of other thinkers and its openness to other disciplines and traditions of thought.

See also: Hermeneutics

JOHN B. THOMPSON

List of works

Ricoeur, P. (1947) *Gabriel Marcel et Karl Jaspers: Philosophie du mystère et philosophie du paradoxe*. Paris: Temps Présent.

Ricoeur, P. and Dufrenne, M. (1947) *Karl Jaspers et la philosophie de l'existence*, Paris: Éditions du Seuil.

Ricoeur, P. (1950) *Philosophie de la volont, I: Le Volontaire et l'involontaire*, Paris: Aubier; trans. E.V. Kohk, *Freedom and Nature: The Voluntary and the Involuntary*, Evanston, IL: Northwestern University Press, 1966.

Ricoeur, P. *Histoire et vérité*. (1955) Paris: Éditions du Seuil; trans. C.A. Kelbley, *History and Truth*, Evanston, IL: Northwestern University Press, 1965.

Ricoeur, P. (1960a) *Philosophie de la volonté. Finitude et culpabilité, I: L'homme fallible*, Paris: Aubier; trans. C.A. Kelbley, *Fallible Man*, Chicago: Henry Regnery, 1965.

Ricoeur, P. (1960b) *Philosophie de la volonté: Finitude et culpabilité, II: La symbolique du mal*, Paris: Aubier; trans. E. Buchanan, *The Symbolism of Evil*, New York: Harper & Row, 1967.

Ricoeur, P. (1965) *De l'interprétation: Essai sur Freud*, Paris: Éditions du Seuil; trans. D. Savage, *Freud and Philosophy: An Essay on Interpretation*, New Haven, CT: Yale University Press, 1970.

Ricoeur, P. (1967) *Husserl: An Analysis of his Phenomenology*, trans. E.G. Ballard and L.E. Embree, Evanston, IL: Northwestern University Press.

Ricoeur, P. (1969) *Le conflit des interprétations: Essais d'hermneutique*, Paris: Éditions du Seuil; trans. W.

Domingo *et al*., *The Conflict of Interpretations: Essays in Hermeneutics*, ed. D. Ihde, Evanston, IL: Northwestern University Press, 1974.

Ricoeur, P. (1974) *Political and Social Essays*, ed. D. Stewart and J. Bien, Athens, OH: Ohio University Press.

Ricoeur, P. (1975) *La métaphore vive*, Paris: Éditions du Seuil; trans. R. Czerny, *The Rule of Metaphor: Multi-Disciplinary Studies of the Creation of Meaning in Language*, London: Routledge, 1978.

Ricoeur, P. (1976) *Interpretation Theory: Discourse and the Surplus of Meaning*, Fort Worth, TX: Texas Christian University Press.

Ricoeur, P. (1981) *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*, ed. and trans. J.B. Thompson, Cambridge: Cambridge University Press.

Ricoeur, P. (1983) *Temps et récit, I*, Paris: Éditions du Seuil, 1983; trans. K. McLaughlin and D. Pellauer, *Time and Narrative, Volume I*, Chicago: University of Chicago Press, 1984.

Ricoeur, P. (1984) *Temps et récit, II: La configuration du temps dans le récit de fiction*, Paris: Éditions du Seuil; trans. K. McLaughlin and D. Pellauer, *Time and Narrative, Volume 2*, Chicago: University of Chicago Press, 1985.

Ricoeur, P. (1985) *Temps et récit, III: Le Temps raconté*, Paris: Éditions du Seuil; trans. K. Blamey and D. Pellauer, *Time and Narrative, Volume 3*, Chicago: University of Chicago Press, 1988.

Ricoeur, P. (1986) *Du texte à l'action: Essais d'hermeneutique, II*, Paris: Éditions du Seuil; trans. K. Blamey and J.B. Thompson, *From Text to Action: Essays in Hermeneutics, II*, Evanston, IL: Northwestern University Press, 1991.

Ricoeur, P. (1986) *Lectures on Ideology and Utopia*, ed.George H. Taylor, New York: Columbia University Press.

Ricoeur, P. (1990) *Soi-Même comme un autre*, Paris: Éditions du Seuil; trans. K. Blamey, *Oneself as Another*, Chicago: University of Chicago Press, 1992.

References and further reading

Bourgeois, P.L. (1973) *Extension of Ricoeur's Hermeneutic*, The Hague: Martinus Nijhoff.(A study of Ricoeur's earlier work for the advanced student.)

Clarke, S.H. (1990) *Paul Ricoeur*, London: Routledge.(A general and accessible introduction to Ricoeur's work, placing particular emphasis on its relevance to debates in literary theory.)

Ihde, D. (1971) *Hermeneutic Phenomenology: The Philosophy of Paul Ricoeur*, Evanston, IL: Northwestern University Press.(An authoritative work on Ricoeur, which deals with his writings up to 1970.)

Kemp, T.P. and Rasmussen, D. (eds) (1989) *The Narrative Path: The Later Works of Paul Ricoeur*, Cambridge, MA: MIT Press.(A collection of essays on Ricoeur's later work, for the advanced student. This volume contains an excellent bibliography of Ricoeur's publications in English up to 1987.)

Klemm, D.E. (1983) *The Hermeneutical Theory of Paul Ricoeur: A Constructive Analysis*, East Brunswick, NJ: Associated University Presses, 1983.(An introduction to Ricoeur's work for students with an interest in religious language and theology.)

Madison, G.B. (ed.) (1975) *Sens et existence: En hommage Paul Ricoeur*, Paris: Éditions du Seuil.(A collection of essays for the advanced student.)

Palmer, R.E. (1969) *Hermeneutics: Interpretation Theory in Schleiermacher, Dilthey, Heidegger, and Gadamer*, Evanston, IL: Northwestern University Press.(An excellent introduction to the tradition of hermeneutics.)

Philibert, M. (1971) *Paul Ricoeur ou la liberté selon l'espérance*, Paris: Seghers.(A study of Ricoeur's earlier work for the advanced student.)

Rasmussen, D. (1971) *Mythic-Symbolic Language and Philosophical Anthropology*, The Hague: Martinus Nijhoff.(A study of Ricoeur's earlier work, for the advanced student.)

Reagan, C.E. (ed.) (1979) *Studies in the Philosophy of Paul Ricoeur*. Athens, OH: Ohio University Press.(A collection of essays for the advanced student. This volume includes an excellent bibliography of Ricoeur's writings up to 1976.)

Thompson, J.B. (1981) *Critical Hermeneutics: A Study in the Thought of Paul Ricoeur and Jürgen Habermas*, Cambridge: Cambridge University Press.(A comparative study of the work of Ricoeur and Habermas, aimed at the advanced student.)

Valdz, M.J. (ed.) (1991) *A Ricoeur Reader: Reflection and Imagination*, Toronto: University of Toronto Press.(A selection of essays by, and interviews with, Ricoeur. The volume has a good introduction which highlights the relevance of Ricoeur's work to literary theory and criticism.)

Vanhoozer, K.J. (1990) *Biblical Narrative in the Philosophy of Paul Ricoeur: A Study in Hermeneutics and Theology*, Cambridge: Cambridge University Press.(A study of Ricoeur's philosophy of religion, for the

advanced student.)
Wood, D. (ed.) (1991) *On Paul Ricoeur: Narrative and Interpretation*, London: Routledge.(A collection of essays on Ricoeur's later work, for the advanced student.)

Right and good

'Right' and 'good' are the two basic terms of moral evaluation. In general, something is 'right' if it is morally obligatory, whereas it is morally 'good' if it is worth having or doing and enhances the life of those who possess it.

Acts are often held to be morally right or wrong in respect of the action performed, but morally good or bad in virtue of their motive: it is right to help a person in distress, but good to do so from a sense of duty or sympathy, since no one can supposedly be obliged to do something (such as acting with a certain motive) which cannot be done at will.

Henry Sidgwick distinguished between two basic conceptions of morality. The 'attractive' conception, favoured by the ancient Greeks, views the good as fundamental, and grounds the claims of morality in the self-perfection to which we naturally aspire. The 'imperative' conception, preferred in the modern era, views the right as fundamental, and holds that we are subject to certain obligations whatever our wants or desires.

1 Meaning of the terms

There are two different ideas of moral value, which the everyday usage of 'right' and 'good' comes close to matching (see Morality and ethics). The idea of right refers to what is obligatory, to a prescription to which we ought to conform (see Duty). The idea of good, by contrast, refers to what is desirable; it applies to whatever is worth having or doing and enhances the life of which it is a part. Clearly, the obligatory and the desirable are not the same. Pleasure at the sight of another's success may well be morally admirable without being obligatory; so, too, some duties are ours, it seems, whatever their performance may mean for the quality of our lives. Each of these fundamental notions plays an essential role in our moral thinking, and morality as a whole turns on the relations assumed to exist between them.

W.D. Ross' classic study, *The Right and the Good* (1930), explains the difference between moral right and good along these lines. But Ross wrongly supposed that it is as easy to determine which things are morally right and which things morally good as it is to distinguish conceptually between the obligatory and the desirable. What counts as obligatory and what as desirable are often controversial. Like Kant before him, Ross held that acts are morally right or wrong in respect of the action performed, but morally good or bad in virtue of their motive (see Kantian ethics §2). In Ross' view, it cannot be obligatory or morally right that I act with a certain motive - that I keep my promise, not out of fear for my reputation, but because I believe it the proper thing to do. For 'I ought' implies 'I can', whereas I cannot by choice produce a certain motive.

Though near-canonical, this analysis of moral action rests on shaky ground. True, it is often said that we cannot have an obligation to do what it is not in our power to do, and the feelings or convictions that may move us to act cannot be brought about at will. Yet there is also the common thought that it is right, not just to acknowledge our benefactors, but to do so with a sense of gratitude. In addition, the principle that 'ought' implies 'can' is not really so evident as sometimes believed. It does not even seem to fit every case of obligatory action. As Aristotle observed, people may pursue a life of injustice for so long that they become unable to do what justice still requires. No doubt, these conflicting commitments can be qualified so as to hang together with some coherence. But we should not suppose that reasonable people will agree upon the way to do it.

2 Ancient and modern ethics

Another question has to do with how the right and the good are themselves to be related at the foundations of ethics. Henry Sidgwick gave this problem a probing analysis in *The Methods of Ethics* (1874). The nature of moral value, he suggested, assumes two different forms, depending on whether the right or the good is considered the more basic notion. In the one case, he wrote, morality has an *imperative* form, in the other an *attractive* one. In addition, these two views were in his eyes historically distinct: the priority of the good was central to Greek ethics, whereas modern ethics has largely embraced the priority of the right. Sidgwick's remarks contain important insights, both philosophical and historical.

Sidgwick distinguished the imperative and attractive conceptions of morality in the following way. To suppose that the right is fundamental is to hold that there are principles of conduct which are binding or obligatory,

whatever our wants or desires may be. By contrast, to assume that the good is basic is to believe that proper conduct is what we would display if we were sufficiently informed about what at bottom we really desire. Of course, each conception makes use of the other moral notion as well, though in a subordinate position. If the right is fundamental, then, to be good, what the agent wants must conform to the demands of obligation; the good is the object of right desire. If the good is fundamental, then the right is what one ought to do in order to attain what one would want if adequately informed. Thus, the imperative view of morality makes the right prior to the good, whereas the attractive view makes the good prior to the right.

These two theoretical conceptions become clearer if we look at the evidence which led Sidgwick to correlate them with the different outlooks of ancient and modern ethics. The ethical thought of Plato and Aristotle drew upon an attractive notion of good, for they regarded the exercise of moral virtue as an intrinsic part of the 'happy' or fulfilled life that by nature each human being desires as the ultimate end (see Eudaimonia). For them, morality is the agent's interest well-understood. This does not mean that their ethical outlook was a species of egoism, at least as that position is usually defined (see Egoism and altruism). For they believed that the agent's self-fulfilment consists in the exercise of virtue (instead of being a condition instrumentally advanced by it), so that we achieve the fulfilled life we want only by being generous, courageous, and so on, for their own sakes. But the crucial point is that, though both philosophers indeed referred to what we 'ought' to do, they appealed nowhere to the idea that there are obligations binding on all agents unconditionally (or 'categorically', as Kant would say). Theirs was an ethics in which the good is prior to the right.

The priority of the right over the good is an expression that comes from Kant. But Sidgwick rightly maintained that the imperative conception is in reality a broad one, encompassing much of modern ethics. It seems to have first appeared among late medieval Franciscans such as Scotus and Ockham. Rejecting the idea of a perspicuous natural order in the name of God's inscrutable omnipotence, they shifted the source of moral principles from what human beings naturally desire to what God commands (see Duns Scotus, J. §14; William of Ockham §10). In fact, Scotus anticipated developments we associate with Kant, arguing that the Christian rule of loving others for their own sake and thus the idea of justice in its fullest sense cannot rest on the natural desire for self-perfection that underlies Aristotelian and Thomistic ethics, but only on a freedom of will that can suspend that desire (see Perfectionism; Autonomy, ethical §2). Christian theology, with its image of God as moral legislator and its ideal of disinterested love, played a decisive role in the rise of an ethics of the right (see Love §2).

Freed from a theological idiom, the imperative conception of morality has dominated modern moral philosophy. The utilitarian tradition, at least since Sidgwick himself, is no exception (see Utilitarianism). In *Ethics* (1963), William Frankena wrongly, but influentially, described the difference between deontological and utilitarian theories as lying in whether the right or the good is made the fundamental moral concept (see Deontological ethics). In reality, both sorts of theory turn on the priority of the right over the good. The essential principle of any 'consequentialist' theory (utilitarianism being the version that conceives of the good as subjective happiness) is that right action consists in bringing about the most good overall, for all those affected by the action, each 'counting for one and only one' (see Consequentialism). This principle does not subordinate the idea of right action to an independent notion of the good. For it defines the good to be maximized by reference to a categorical principle of right: the maximizand is determined by considering everyone's good impartially, as it is claimed we ought to do, so that the pursuit of the greatest good overall is a duty binding on us unconditionally, whatever the direction in which our own interests would lead us (see Impartiality).

Deontological theories generally hold that we are bound by certain obligations even if an alternative course of action is known to bring about more good overall. But it is not this principle that expresses the priority of the right characteristic of modern ethics. It is rather the view such theories share with consequentialism, namely, the view that moral duty is independent of the agent's own good. This common assumption is foreign to the spirit of ancient ethics, which is why the unending debate between deontological and consequentialist outlooks, so central to modern ethics, never engaged the Greek moralists and their successors.

3 Contemporary problems

Two important qualifications must be made to Sidgwick's account of the difference between ancient and modern ethics. First, even if we agree that we are subject to categorical duties and thus continue to understand ourselves as 'modern', we need not suppose that morality as a whole must fit the imperative conception. A more plausible view

would have a place, not just for a core morality unconditionally binding on all, but also for duties arising from friendship, the ties of place and culture, and membership of particular associations (see Friendship; Family, ethics and the; Solidarity). Such duties are ours only so long as we remain interested in maintaining these social relations. It would be a hollow idea of friendship that implied that we owe our friends the special concern they deserve whatever we may feel about them.

Second, we should note that some important modern philosophers have denied that ‘’right’ or ‘ought’ should play the fundamental role in moral theory. In *An Enquiry Concerning the Principles of Morals* (1751), Hume suggests that the idea of morality as a system of laws is an invention of Christian theology, which, focusing moral approval on what can be performed at will, misses the central importance of the virtues. Schopenhauer (1840) took Kant to task for assuming that ethics must have an imperative form (see Schopenauer, A. §6). That assumption makes sense, he insisted, only within a theological framework, so that the practical postulate of belief in God which Kant drew from his ethics was really its guiding premise all along. And in the twentieth century, G.E.M. Anscombe argued that the modern emphatic sense of ‘ought’ has its home only within the conception of a divine legislator (see Anscombe, G.E.M. §2). While we may feel unconditionally bound by certain obligations, without thinking of any theological framework, these convictions are in her view only ‘survivals’. Having lost all intelligible meaning, they urge themselves upon us with the authority of conscience only because we have been shaped by a tradition in which, however, we no longer believe.

These dissenting voices do not refute Sidgwick’s thesis about the difference between ancient and modern ethics. For all three philosophers understood themselves as protesting against what had become the dominant conception of ethics in modern times. Indeed, Hume and Anscombe urged a return to ancient models. Anscombe’s essay, ‘Modern Moral Philosophy’ (1958), is a charter document of recent neo-Aristotelianism or ’virtue ethics’, which seeks to conceive of the virtues as fundamentally forms of human flourishing (see Virtue ethics). Whether the good should thus be regarded as prior to the right, or whether the modern priority of the right should be maintained, is a cardinal debate of contemporary ethics.

See also: Confucian Philosophy, Chinese §6; Good, theories of the

CHARLES LARMORE

References and further reading

Anscombe, G.E.M. (1958) ‘Modern Moral Philosophy’, *Philosophy* 33 (1): 1-19; repr. in *Collected Philosophical Papers*, Minneapolis, MN: University of Minnesota Press, 1981, vol. 3, 26-42.(An influential critique of the modern emphasis on the notion of right.)

Frankena, W. (1963) *Ethics*, Englewood Cliffs, NJ: Prentice Hall, 2nd edn, 1973. (Contains an influential, if erroneous, account of the distinction between ‘teleological’ and ’deontological’ theories.)

Hume, D. (1751) *An Enquiry Concerning the Principles of Morals*, in *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 3rd edn, 1975, appendix IV.(Discusses the difference between ancient and modern ethics.)

Larmore, C. (1996) ‘The Right and the Good’, in *The Morals of Modernity*, Cambridge: Cambridge University Press.(Explores the roles of the right and the good at the foundations of ethics.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Oxford University Press.(A classic work containing an acute analysis of the moral notions of right and good.)

Schopenhauer, A. (1840) *Über die Grundlage der Moral*, trans. E.F.J. Payne, *On the Basis of Morality*, Indianapolis, IN: Bobbs-Merrill, 1965, sect. 4.(Contains his critique of Kant’s imperative conception of morality.)

Sidgwick, H. (1874) *The Methods of Ethics*, Macmillan: London; 7th edn, 1907; repr. Indianapolis, IN: Hackett Publishing Company, 1981, bk I, ch. 9.(Contains his insightful contrast between ancient and modern ethics.)

Wolter, A.B. (1990) *The Philosophical Theology of John Duns Scotus*, Ithaca, NY: Cornell University Press. (Excellent treatment of Scotus’ ethics, bringing out those elements which anticipate modern conceptions.)

Rights

There is widespread consensus that rights are ways of acting or of being treated that are beneficial to the rightholder. Controversy begins, however, when one attempts to specify the notion of rights further.

(1) It is sometimes said, perhaps too casually, that all rights carry with them correlated obligations - things that other persons are supposed to do or refrain from doing when some given person is said to have a right to something. The question is: how is it best to state this relationship between rights and correlated obligations?

(2) Most people think that rights are, in some sense, justified. But there is considerable controversy as to what, precisely, is the proper focus of justification. Some say that rights are practices (certain ways of acting or of being treated) that are established, typically socially established. Thus, the issue for them is whether the fact of social recognition and enforcement is justified (or could be). Others say that rights themselves are claims; hence a right is a justified claim or principle of some sort (whether the practice identified in that claim exists or not). This dispute, between rights as justified practices and rights as justified claims, needs to be explored and, if possible, resolved.

Other topics need addressing beyond the question of the initial characterization of rights. One of them is the question of the function of rights: what good are they anyway? what can one do with rights? Another is the question of how best to justify particular kinds of rights, such as human rights and basic constitutional rights. Is there a substantive theory of critical morality that can do the job? Many people are concerned, especially, with whether utilitarianism (one of the dominant ethical theories in the West today) is up to this task. Finally, mention should be made of one other issue much talked about of late: what kinds of beings can have rights, and under what conditions of possession and dispossession?

1 Initial characterization and some points of consensus

Rights are an important issue in contemporary social and political philosophy. For it is widely held that rights, by providing a significant protection of important interests of individuals against the state and against other persons (even a majority), give a person something to stand on. One may not want to go so far as to say that rights are ‘trumps’ (as some have), but it is none the less clear that rights are valuable things. So it seems natural to ask: ’what, then, is a right?’

Rights are socially established ways of acting or ways of being treated (or, alternatively, such ways as ought to be so established). More specifically, a right so understood is a right *to* something that is (1) fairly determinate and that (2) can be similarly distributed on an individual basis to each and all of those who are said to be rightholders. A right is always regarded as (3) a beneficial way of acting or of being treated both for the rightholder and, more generally, for society. Thus, (4) it is or should be something socially accepted - recognized and protected in given societies. Such acceptance would be (5) deemed reasonable, even by outsiders, in that it made explanatory sense. For the way of acting or of being treated in question could be exhibited, plausibly, as a means to or as a part of accomplishing some interest or perceived benefit or other good (or desirable) thing. Accordingly, (6) directives could be issued to others, to those who are not rightholders. And (7) further initiatives could be taken as a feature of any such successful claim to rights status.

This initial characterization constitutes common ground in the arguments people make about rights. Indeed, several of its features are not particularly controversial at all. Thus, there would today be widespread consensus on the idea that rights are ways of acting or ways of being treated that are (1) appropriately determinate, (2) equitably distributable on an individual basis and (3) beneficial. Even the central characterization, concerning social acceptability in (4), is not unduly contestable as stated; but dispute would break out as soon as we tried to determine what to emphasize - whether rights are *socially established* or merely *ought* to be. Finally, the idea (6) that rights always involve some sort of normative direction of the behaviour of others might also appear to be universally agreed upon; but there are problems with alleging consensus on this particular point.

2 Normative direction

The view in question in (6) is often put by saying that rights correlate with duties - meaning thereby that a right

always implies or has attached some distinctive and closely related duty of others. But serious difficulties arise for the thesis in this precise form.

The most interesting arguments against such correlations derive from Wesley Hohfeld's highly influential classification of rights (Hohfeld 1964) (see Hohfeld, W.N.). On his view a legal right could be constituted by any one of four elements: a claim; a liberty; a power; or an immunity. And each type of right has a unique second-party correlative. Thus, for a legal *claim right* to some thing the correlative element is a legal duty of some second party. Analogously, a person's *immunity right* from some thing is necessarily correlated with a lack of power - with a legal disability - on the part of others to do that thing (for example, the constitutional inability of the US Congress to 'abridge' free political speech). Hohfeld's point is simply that a legal duty and a no-power (a legal *in*ability) are significantly different. Accordingly, the existence of immunity rights tells against the view that the correlative of every right is always going to be a closely related second-party duty (see Duty).

Thus, the thesis that rights logically correlate with specific duties is not sound. A weaker but more defensible view is that any genuine right must involve some normative direction of the behaviour of persons other than the holder. Even this weaker thesis, however, seems to run up against the authority of Hobbes (in his account of rights in the state of nature) and of Hohfeld. We can see this most clearly by looking at the Hohfeldian liberty right.

Here the legal *liberty right* to do some thing - which consists in the absence of any duty on the agent's part to refrain from doing that thing - is matched with other people's lack of a claim that such a thing not be done by the agent. The point is, this is the *only* directive incumbent on the conduct of second parties in the case of a liberty. They can make no claim on the duties of a liberty-rightholder (to refrain); beyond that their own action is relatively unencumbered.

A liberty right, so conceived, is indeed an odd one. For it fails to capture the common-sense notion that when one has a liberty right to do a thing someone else is directed not to interfere with that doing. The problem is that literally *no* normative direction at all is involved for second parties (in Hobbes' case) and no *significant* normative direction against interference is involved (in Hohfeld's).

If the common-sense notion of a liberty right is correct, there ought always to be some sort of strong mandate for non-interference, either explicitly stated in our formulation of a given liberty or present at least in the context in which that liberty normally occurs. Thus, to take the latter case, certain standing duties of second parties (such as the duty not to assault or batter others or to trespass on their property), even though these are relatively independent of a given liberty (for example, the liberty to paint one's barn a shocking purple), would none the less afford the exercise of that liberty a considerable degree of protection. Without some such fairly robust mandates against interference (either closely connected to the liberty in question or permanently and independently in place on the 'perimeter' of its usual exercise), we would probably be inclined to call that liberty, not a right, but a *mere* liberty or a privilege.

The upshot, then, is that we should state our main contention so as to emphasize *significant* normative direction (on the conduct of second parties). This is the focal point of what appears to be an emerging consensus on the matter at issue.

3 Accreditation

No real consensus has emerged, however, on the point we now turn to: whether rights, in order to be rights, require social recognition (and beyond that, social maintenance). In considering this issue one school of thought - embracing both classical natural rights theorists and contemporary advocates of human rights - has tended to emphasize that individuals can have rights independently of organized society, of social institutions, and hence of social recognition and maintenance in any form. The rather common characterization that rights are essentially claims can be taken as a way of emphasizing that rights hold irrespective of whether they have been acknowledged, either in the society or, more specifically, by that person against whom the claim is made.

Against the view that rights are essentially claims are ranged a number of philosophers. Bentham (§5) comes most readily to mind, and his polemic on this very point against natural rights as 'nonsense' still adds relish to philosophical discussions. T.H. Green, in his insistence that rights require social recognition and that without it they are something less than rights, would be another. And, oddly enough, Lenin would be a third.

The problem we are examining arises, in part, because the procedure for deciding whether something is a right is not wholly settled. We find that the vocabulary of rights, in particular, of human rights, may actually be used at any of several steps: that of mere claim, that of entitlement (where only the claim-to element is really settled), that of fully validated claim (where we have the idea both of a justified claim *to* something and of a justified claim *against* someone for it) and, finally, that of satisfied or enforced claim (where the appropriate measures required to support or to fulfil the claim have been given effective embodiment as well). The presence of these possible stages has introduced a degree of ambiguity into assertions that a right exists.

Accordingly, we find a significant variety of contemporary opinion as to the point at which such assertions can most plausibly be thought to take hold. While some have said simply that rights are *claims* (Mayo 1965), others say they are *entitlements* (McCloskey 1965), and yet others (most notably, Feinberg 1973, 1980) say they are *valid claims*. Ranged against them have been those (such as Sumner 1987; Martin 1993) who emphasize that rights, even human rights, are basically established ways of acting or being treated. And, last of all, some (for example, Rawls 1971; Melden 1988) have treated rights as legitimate expectations and, hence, have landed more or less in the middle (see Rawls, J.).

The main backdrop to the view that rights are (valid) claims is, I think, the common opinion (emphasized by Dworkin (1978), Raz (1986), MacCormick (1977) and Held (1984) among others) that to have a right is to have a justification for acting in a certain way, or a justification for being treated in a certain way (see Dworkin, R.). Now, suppose that a candidate for rights status had all the rights-making features (mentioned in §1) but one. Although accredited (in the sense of justified), it was not established; it lacked the *social* recognition which it ought to have.

Why should the lack of such recognition deprive it of rights status? For, clearly, if we modelled the rights-making features on what was justified (what was accredited in *that* sense), the thing was already a right even before it was recognized, even before it became a practice. And when it was recognized it would be recognized as a *right* (as something that was fully justified) and would not simply *become* a right in being recognized.

The opposing view, that rights are socially recognized practices, rests on three main contentions. The first of these is the contention that the notions of authoritative recognition (if not explicit, then at least implicit, as evidenced by conduct) and of governmental promotion and maintenance (usually on a wide variety of occasions) are themselves part of the standard notion of a *legal* right, that is, when we are concerned with rights that are more than merely nominal ones.

Thus, on the social recognition view, the fatal flaw in the theory of rights as valid claims (in any of its formulations) is the suggestion that practices of governmental recognition and enforcement in law can be dispensed with in the case of legal rights. Indeed, this is the very point at which both Dworkin and Raz, who might otherwise be taken to be supporters of some form of the valid claims thesis, desert that thesis for one that emphasizes the necessity of institutionally establishing ways of acting/being treated, if these are to count as *legal* rights (see Dworkin 1978; Raz 1984).

The second point put forward by the social recognition view is that it is desirable to have, if possible, a single, unequivocal sense of ’rights’: one that is capable of capturing both legal rights and human (and other moral) rights under a single generic heading. Now, if the argument just sketched is to be credited, then the view of rights as valid claims does not provide an adequate generalized notion of rights, one that can comfortably include both legal and human rights. For we have already seen that legal rights cannot be satisfactorily accounted for under the heading of valid claims.

This brings us to the third point urged by the social recognition view. Here the argument is that *all* moral rights can, indeed must, be construed as involving established practices of recognition and maintenance. Since human rights (as a special case of moral rights) are thought to be addressed to governments in particular, we must regard practices of governmental recognition and promotion as being the form for such recognition and maintenance to take for these rights. Here we have, in brief compass, then, the social recognition view that opposes the contention that rights are essentially justified or valid claims.

4 Functions of rights

Rights have many functions. Two in particular are emphasized in the contemporary literature: the conferring of liberty or autonomy (on rightholders) and the protection of their interests, especially their basic interests.

Rights in the seventeenth and eighteenth centuries were largely discussed as if they were simply liberties and, hence, ways of acting on the part of the rightholder. Indeed, this tendency is deeply rooted in the tradition of rights discourse. It is hard to say when 'a right' was first spoken of in a way continuous with current usage, but many careful expositors locate that first recognizable use with William of Ockham (§11), when he talked of a right (*ius* or *jus*) as a power or capacity (*potestas*) to act in accordance with 'right reason' or, in the special case of a legal right (*ius fori*), with an agreement. Such a usage was well established by the seventeenth century (with Hobbes and, some would say, with Locke) and has been widespread ever since.

It constitutes, none the less, a drastic oversimplification - even if the rights referred to are, as they often are, the classic rights of the eighteenth-century declarations. For these rights include important rights to ways of being treated and such rights are *not* things the rightholder does or can do. Even so, the oversimplification continues to prevail in philosophical literature (for example, Hart 1973; Rawls 1971, 1993). Thus, Rawls' 'equal basic liberties' (enshrined in his first principle of justice) include both liberties of action and ways of being treated, typically ways of not being injured by the actions of others (see Freedom and Liberty §3).

It is clearly possible to have both important functions (the conferring of liberty on rightholders and the protection of their interests) as functions of rights, often of a single right. Thus it seems arbitrary, where both functions are normally served by almost all rights, to single out just one of these functions (typically the function of conferring autonomy) and to give it *definitional* weight (see Sumner 1987).

In fact, in line with the contemporary understanding of rights (as expressed, for example, in the UN's *Universal Declaration of Human Rights* of 1948), it might be best to stress three main functions of rights. Thus, the central content of some rights will be a way of acting (for example, a liberty of conduct of some sort). But at the core of other rights will be a way of being treated: a non-injury of some sort or, alternatively, the provision of a service.

Corresponding to each main heading or class of rights (as determined by these central cores), there is an appropriate or characteristic normative response enjoined for the conduct of others. But the essential character of this normative direction of the conduct of second parties shifts from main case to main case. Allowing or even encouraging a piece of conduct is what these parties are normatively directed to do in the case of a liberty; prohibiting their doing of an injury to the rightholder or requiring of them a service, again to the rightholder, is the incumbent directive in the other two cases.

5 Critical justification

Rights are eminently plausible candidates for justification, an idea that I tried to capture in §1 with the notion that rights are *accredited* ways of acting or of being treated. This section will consider some of the main full-blown theories offered to justify rights. One proviso is that the most important rights are universal rights - in particular, human rights and constitutional rights, which are fundamental or basic civil rights of all persons (or all citizens) within a given politically organized society. My account is limited to theories that attempt to justify such universal rights.

All civil rights are important rights and all reflect a high level of social commitment. But not all can be justified as representing individuated and practicable and universal moral claims which serve as proper conclusions to sound arguments from objective principles of critical morality (or at least from principles widely regarded as reasonable).

Some can be, however. Indeed, in the social recognition view (described in §3), human rights would be, simply, constitutional rights that embodied precisely such morally valid claims. These claims, then, when on their own, could be described (relatively noncontroversially and giving due weight to both the opposing views canvassed in §3) as human rights *norms*.

We can ask: what might be involved, then, in the justification of human rights - or human rights norms - and of those constitutional rights susceptible of the same sort of justification? One thing seems clear: the norms which constitute or back up human rights are moral norms. Thus human rights can exist only if substantive moral norms in some sense exist (or, at least, can be objectively described and argued for). Now, it is possible for moral, and hence human, rights to exist even if moral norms are conventional or are relative to culture. But if human rights -

or human rights norms - are to serve their role as international standards of political criticism then such a conventional morality would have to include some norms that are accepted worldwide. More important, if such norms are to have weight and bearing for future human beings in societies not yet existing (and this much would seem to be involved if we are to call these norms *universal* in any significant sense), then these norms cannot be merely conventional (see Universalism, ethical).

Thus, in classifying human rights as moral rights one may wish to distinguish between actual and critical moralities. What seems especially crucial to human rights, then, is the belief that there are objectively correct, or objectively reasonable, critical moral principles. Often, human rights - or human rights norms - are traced back to such foundational ideas as human dignity or moral personality or moral agency or moral community (for example, Melden 1988). But the exploration of such possibilities has failed to gain widespread support, perhaps because such notions as moral agency do not themselves seem sufficiently distinct from the very norms or rights they are being called upon to justify. Or perhaps because such notions seem, in the end, to stand in need of a more basic sort of justification themselves. Thus, we might do well to consider other grounding principles, principles that could be regarded as rock bottom and, arguably, as objectively reasonable.

One appropriate way to narrow the field among these is to consider first those substantive theories of critical justification that have grown up in proximity to serious talk about human and constitutional rights. Three important contemporary theories fit this description: utilitarianism (in particular, the theory developed by J.S. Mill (§10) and advocated recently under the name of ’indirect’ utilitarianism), the theory of John Rawls, and rational-choice ethical theory, especially that of David Gauthier (1986) (see Contractarianism §3). I will confine the discussion to one example.

Rawls’ theory (like Dworkin’s) emphasizes the standing priority of basic liberties and other constitutional rights over such things as the common good or perfectionist values (for example, the value of holiness, as religiously conceived, or the values of Nietzschean elitism). In his 1993 book Rawls sketches a complex theory of justification; it has two main parts. He starts with what he calls a ’freestanding’ justification of the political conception of justice, drawing here on certain fundamental ideas which he finds ’implicit’ in the contemporary democratic tradition. Next he claims that this political conception will also be endorsed and supported as the focus of an ’overlapping consensus’ among the proponents of various comprehensive religious and moral doctrines that exist in the Western world today. Historic utilitarianism is prominently mentioned as one of these doctrines.

But it is doubtful that the utilitarian principle of general happiness could support the assignment of basic rights - constitutionally guaranteed benefits - to individuals if such rights prevented the utilitarian politician from allowing policies favourable to corporate or aggregate interests to override or supersede constitutional rights when those interests could be seen to conduce to greater benefit. In that sense, then, philosophical utilitarianism is incompatible with the notion of basic rights developed by Rawls, Dworkin and others. For utilitarianism cannot possibly accept a critical justification of a scheme of basic institutions in which constitutional civil rights have a standing priority over policies favouring corporate goods or aggregate welfare.

Much hinges, it would seem, then, on how discussion of the critical justification of rights is set up and conducted. And if questions regarding the distribution of rights are best taken up *after* successful or at least plausible attempts at justification, then such issues as what kinds of beings can have rights are seen to hang in the balance as well.

Currently, we find highly agitated discussions about whether foetuses have rights or whether animals can have them or about limits to the right to life (in cases of mercy killing, for example, or in requests for assisted suicide). But a serious attempt to give answers to questions such as these, questions of distribution and of scope and of defeasibility, cannot be clearly addressed until they can be considered in the light of adequate accounts of the function of rights and with one or more substantive theories of critical justification in hand (see Animals and ethics; Reproduction and ethics).

See also: Law, limits of

REX MARTIN

References and further reading

Dworkin, R. (1978) *Taking Rights Seriously*, Cambridge, MA: Harvard University Press, ch. 4.(Referred to in

§§3, 5. Originally published 1977; important appendix added in 1978. Essays vary in difficulty; some are intended for a popular audience, others are law review articles. Essays 2-4, 6, 7, 12 and 13 are especially relevant to rights.)

Dworkin, R. (1985) *A Matter of Principle*, Cambridge, MA: Harvard University Press.(Essays, with a similar range of difficulty to Dworkin 1978. Essays 2, 5 and 17 are especially relevant to rights.)

Feinberg, J. (1973) *Social Philosophy*, Foundations of Philosophy series, Englewood Cliffs, NJ: Prentice Hall. (Referred to in §3. Feinberg is probably the leading US writer on the concept of rights; his argument - also found in Feinberg 1980 - that rights are valid claims has been especially influential. This book is intended for introductory reading by students in political studies. Chapters 4-6 are relevant to rights.)

Feinberg, J. (1980) *Rights, Justice, and the Bounds of Liberty*, Princeton, NJ: Princeton University Press. (Referred to in §3. A collection of Feinberg's writings, originally published in scholarly journals. Feinberg is an unusually clear writer. Essays 6-8 and 11 are relevant to rights.)

Gauthier, D. (1986) *Morals by Agreement*, Oxford: Oxford University Press.(Referred to in §5. An impressive attempt to ground morality in rational choice theory. It is elegantly written; some parts are quite technical. Chapters 7 and 9 are especially relevant to rights.)

Hart, H.L.A. (1973) 'Bentham on Legal Rights', in A. W. Simpson (ed.) *Oxford Essays in Jurisprudence*, Oxford: Clarendon Press, 2nd series, 171-201; repr. unchanged, in H.L.A. Hart, *Essays on Bentham: Studies in Jurisprudence and Political Theory*, Oxford: Oxford University Press, 1982, ch. 7, 162-93.(Referred to in §4. A very influential paper.)

Held, V. (1984) *Rights and Goods: Justifying Social Action*, New York and London: Free Press.(Referred to in §3. Well-written book, with many interesting examples.)

Hohfeld, W.N. (1964) *Fundamental Legal Conceptions*, New Haven, CT: Yale University Press.(Referred to in §2. The two papers printed under the above title first appeared as articles in the *Yale Law Journal*, one in 1913, the other in 1917. Earlier editions of this book appeared in 1919 and 1923. Hohfeld's idea of four distinct rights patterns has left an indelible impression on Anglo-American theory.)

MacCormick, D.N. (1977) 'Rights in Legislation', in P.M.S. Hacker and J. Raz (eds) *Law, Morality and Society: Essays in Honour of H.L.A. Hart*, Oxford: Oxford University Press, ch. 11, 189-209.(Referred to in §3. Criticizes Hart's idea of rights as protected choices, discussed in §4.)

McCloskey, H.J. (1965) 'Rights', *Philosophical Quarterly* 15 (59): 115-27.(Referred to in §3; this paper is one of the foremost interpretations of rights as entitlements. See also McCloskey's 1976 paper 'Rights - Some Conceptual Issues', *Australasian Journal of Philosophy* 54 (2): 99-115.)

Martin, R. (1993) *A System of Rights*, Oxford: Clarendon Press.(Expansion of the material of this entry, esp. of §§1, 5; referred to in §3. See chapter 13 for a discussion and criticism, with extensive citation, of utilitarian attempts at a critical justification of rights and those of Rawls', especially his 1982 article and his writings from 1985 onwards. This book contains a considerable bibliography relevant to rights.)

Martin, R. and Nickel, J.W. (1980) 'Recent Work on the Concept of Rights', *American Philosophical Quarterly* 17 (3): 165-80.(Expansion of the material of §§2-4 of this entry. The 1980 paper contains lengthy discussions of human rights and of the defeasibility of rights. It contains a useful bibliography. For an even more extensive bibliography, with some annotations, by the same authors, see their 1978 piece 'A Bibliography on the Nature and Foundations of Rights, 1947-77', *Political Theory* 6 (3): 395-413.)

Mayo, B. (1965) 'Symposium on ''Human Rights'', II', *Proceedings of the Aristotelian Society Supplementary Volume* 39: 219-36.(Referred to in §3. Well-written exposition of the idea that rights are simply claims.)

Melden, A.I. (1988) *Rights in Moral Lives: A Historical-Philosophical Essay*, Berkeley, CA: University of California Press.(Referred to in §§3 and 5. Some parts are easy going, others not. Interesting attempt to justify our current set of rights by moving along two main lines: through considering relations of persons in a moral community and through comparing rights, or their absence, in moral communities at different times in history.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(Referred to in §§3, 4. A magisterial treatise. Not easy going - Rawls' prose is reminiscent of the high Victorian philosophical style of Green and Sidgwick. Rawls' characterization of rights is vague, but the book contains one of the most impressive attempts at a critical justification of rights.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(Although Rawls here attempts a more conversational style, this book is ultimately no easier than his 1971 book. In his 1993 book Rawls attempts to rework his theory of justice as a political theory. It also contains, unrevised, his important 1982 essay on 'The Basic Liberties and Their Priority'. Rawls' 1993 theory of critical justification is described and criticized in §5.)

Raz, J. (1984) 'Legal Rights', *Oxford Journal of Legal Studies* 4 (1): 1-21.(Cited in §3. Raz is a principal

representative of the jurisprudential tradition of Hart. The essay, although difficult, repays close reading.)
Raz, J. (1986) *The Morality of Freedom*, Oxford: Oxford University Press.(Referred to in §3. Chapter 7 is on rights.)
Sumner, L.W. (1987) *The Moral Foundation of Rights*, Oxford: Oxford University Press.(Referred to in §§3 and 4. This well-written book also represents one of the most important attempts to provide a critical justification of rights on a utilitarian basis.)
Tuck, R. (1979) *Natural Rights Theories: Their Origins and Development*, Cambridge: Cambridge University Press.(A major study - provides a brief early history of rights and an indispensable background to seventeenth-century thought, in particular Hobbes.)
Waldron, J. (ed.) (1987) *Nonsense Upon Stilts: Bentham, Burke and Marx on the Rights of Man*, New York and London: Methuen.(This book consists mainly of excerpts from the three theorists named plus introductions to each by Waldron and, also, a lengthy general introduction and a lengthy concluding essay by him. Valuable for background to discussions on rights in the eighteenth and nineteenth centuries - although weak on Rousseau and on the US contribution. Includes a very useful bibliographical essay.)

Risk

Every day persons face threats from natural disasters such as hurricanes and from technological hazards such as exposure to more than 60,000 different chemicals. The increase of pervasive, human-caused hazards raises a number of philosophical issues, most notably in the areas of ethics and epistemology. There are three main classes of ethical issue associated with risk. (1) Who should define risk, and how should it be defined? (2) Who should evaluate risk, and according to which rules? (3) What are the conditions under which it is ethically acceptable to impose societal risk? Societal risks (such as those from liquefied natural gas facilities) tend to be involuntarily imposed, whereas individual risks (such as those from dietary consumption of saturated fats) are more voluntarily chosen. This discussion addresses societal, rather than individual, risks because they involve less individual choice and hence more ethical controversy.

1 Ethics and risk definition

Moral philosophers use the term ‘risk’ in at least five distinct ways. In the context of ordinary language analysis and normative ethics, a ‘risk’ is the possibility that some harm will occur. In the (second) context of Bayesian decision theory, a ‘risk’ is the probability of an undesirable outcome (see Decision and game theory). Risks are thus distinguished from certain outcomes (having probability 1) and from uncertain outcomes (to which no probabilities can be assigned). In the (third) context of quantitative risk assessment (QRA), a ‘risk’ is the probability that some consequence will occur. Typically a risk is expressed in QRA as the average annual probability of fatality that a particular situation imposes on an individual, such as a coal miner (see Risk assessment). In the (fourth) context of risk-benefit analysis (RBA), a ‘risk’ is the monetary value assigned to some probable negative outcome such as loss of life. A variant of cost-benefit analysis, RBA is used commonly to evaluate whether some risk (such as the use of a pesticide) is worth the benefits. In the (fifth) context of insurance, a ‘risk’ is the chance of loss, often financial loss.

One of the most basic ethical issues associated with risk is how and by whom it is to be defined. Scientists and engineers who study societal risks - such as those associated with toxic waste dumps - typically define ‘risk’ in the third way just mentioned, as the probability that some harm, like death, will occur (National Research Council 1983). Social scientists and philosophers often argue that risk cannot be reduced to a mathematical expression, as in the third definition, because it also is a function of qualitative components (captured by the first definition) such as citizen consent and trust in government risk managers (MacLean 1986). Some moral philosophers also argue that typical definitions of risk exhibit a ‘naturalistic fallacy’ in the sense that they attempt to reduce ethical to purely scientific concepts (Shrader-Frechette 1985).

Much ethical debate focuses on whether societal risk ought to be defined (as in the second, third and fourth senses) by members of the scientific community, or by laypersons (as in the first sense) who are most likely to be its victims. Scientists tend to treat risk definition as the paternalistic prerogative of experts, in part because they claim that the definitions of irrational, ignorant, or risk-averse laypersons could impede social progress (Douglas and Wildavsky 1982) (see Paternalism). Many moral philosophers argue, in return, that rationality is a matter not merely of scientifically defensible outcomes but - because risk affects public welfare - also of just procedures for defining and evaluating risk (Cranor 1992; Shrader-Frechette 1991). Ethicists also claim that because there are no established hazard frequency records for new risks and because scientists have well-known heuristic biases in defining and estimating risks - such as the overconfidence biases in estimates of nuclear risk (Cooke 1993; Kahneman *et al*. (eds) 1982) - it is important to shape risk definitions through participatory democracy as well as by scientific fiat.

2 Ethics and risk evaluation

Still other ethical controversies concern how one ought to evaluate risks. A major issue is whether one ought to judge risk acceptability according to Bayesian or maximin rules. Should one maximize average expected utility (where ‘utility’ is the ability to satisfy some human want and where ‘expected utility’ is the subjective probability of some state multiplied by its ability to satisfy some human want)? Or should one minimize the likelihood of the worst outcome (consequences)? Utilitarians like Harsanyi (1975) argue for the Bayesian position on the grounds that ‘worst cases’ of technological risk occur rarely. They say that maximin decisions are overly conservative,

impede social progress, and overemphasize small probabilities of harm. Egalitarians, many influenced by Rawls (1971), argue for maximin on the ground that the subjective risk probabilities are both uncertain and dwarfed by potentially catastrophic consequences, such as global warming, toxic leaks, or nuclear core melts. They also point out that a small (close to zero) probability of catastrophe does not outweigh infinitely serious risk consequences (see rationality, practical §§6-7).

Other ethical controversies regarding risk evaluation concern whether, in cases of uncertainty where both cannot be avoided, one ought to minimize false positives (false assertions of harm) or false negatives (false assertions of no harm). Following traditional norms, many scientists argue for minimizing false positives on grounds that this stance is conservative, avoids positing an effect where there may be none, and places the burden of proof on those attempting to confirm some harm. Many moral philosophers, however, claim that traditional, pure-science norms for dealing with uncertainty are inapplicable to societal risk decisions because they affect human welfare. They argue for minimizing false negatives because doing so gives greater protection to public health and places the burden of proof on risk imposers rather than risk victims (Shrader-Frechette 1991). More generally, ethicists have challenged the traditional legal dictum that a potentially hazardous substance (such as a carcinogen) is 'innocent until proved guilty'. In cases involving 'toxic torts', they argue that fairness and equal treatment require risk evaluators and decision-makers to reverse the burden of proof, in part because causal chains of harm are difficult to prove and in part because risk victims are less able than risk imposers to bear the costs of faulty risk evaluations (Cranor 1992).

3 Ethics and risk imposition

Apart from the quantitative and scientific rules (such as expected utility) for evaluating risk, the most significant set of ethical issues regarding risk concerns the qualitative criteria under which it is acceptable to impose some hazard (such as chemical effluents) on workers or the public. One important criterion is the equity of distribution of the risks and benefits associated with some activity. Parfit (1983), for example, argues that temporal differences among persons/generations are not a relevant basis for discriminating against them with respect to risk. He and others maintain that a risk imposition is less acceptable to the degree that it imposes costs on future persons but awards benefits to present persons. Commercial nuclear fission, for example, benefits mainly present generations, whereas its risks and costs will be borne by members of future generations (see Future generations, obligations to). Many economists, however, question notions of distributive equity and continue to discount future costs such as deaths caused by hazardous technologies. They also question whether 'geographical equity' (Shrader-Frechette 1993) or 'environmental justice' (Bullard 1993) requires risks to be distributed equally across generations, regions and nations. Proponents of siting hazardous facilities - for example, toxic dumps - in economically and socially disenfranchised areas argue that such risks have been accepted voluntarily and that they provide employment as well as tax benefits. They say that a bloody loaf of bread is better than no loaf at all. Opponents of such risk impositions argue that life-threatening risks rarely bring substantial benefits. They maintain that economically and socially disenfranchised persons typically bear larger burdens of societal and workplace risk.

Economically, educationally, or socially disfranchised persons also are less likely than others to be able to give genuine free informed consent to public and workplace risk (MacLean 1986; Rescher 1983) (see Consent). Some utilitarian ethicists have argued, however, that no instances of consent are perfect, that workers and members of the public have the moral right to trade bodily security for higher wages or economic benefits, and that the greater good is achieved by risk-for-money trade-offs (see Utilitarianism). Ethical debate over such trade-offs focuses on opposed views about rights, paternalism, human dignity, equal treatment, and adequate compensation for risk. On the one hand, many utilitarians and traditional economists believe that the 'compensating wage differential', for example, gives workers in riskier occupations fair or equitable treatment. They also argue that overall economic prosperity justifies the imposition of industry-generated risks and that allowing companies to buy and sell legal rights to pollute leads to the greater good of all. Moreover, they claim that the contemporary federalist economy requires a 'politics of sacrifice' within which guaranteeing free informed consent to all societal and workplace risks is unrealistic and unattainable. On the other hand, MacLean (1986) and others claim that some things (like bodily health and environmental security) ought not be traded for compensation. Gewirth (1982), for example, argues that persons have a moral and legal right not to be caused to have cancer. Philosophers who argue for the ethical justifiability of the economics-safety trade-off tend to be utilitarians, to have more lenient conceptions of free informed consent, to underemphasize sociological differences among those who accept risky jobs and living

conditions, and to believe that pursuing neoclassical economics leads overall to social and ethical benefits. Philosophers who argue against the economics-safety trade-off tend to be egalitarians or social-contract theorists, to have more stringent conceptions of free informed consent, to overemphasize the sociological differences among persons allegedly choosing different levels of societal risks, and to be more critical of the ethical assumptions underlying neoclassical economics.

Another aspect of the consent and compensation debate over risk imposition concerns liability for societal risks. Current US laws, for example, excuse nuclear power plant licensees from full liability for accidents on grounds of economic efficiency and the greater good. Many ethicists argue that such exclusions violate rights such as due process (Shrader-Frechette 1993), legal rights to full administration of the law and to redress under the law.

Apart from the question whether risk victims deserve due process, a related ethical issue is the magnitude of societal risks that ought to be subject to regulation. How far do legal or moral rights to bodily security extend? Ought government or industry be allowed to impose even a minimal risk on persons without their consent (MacLean 1986)? The Delaney Clause to the US food additive amendments, for example, demands a zero-risk level and prohibits the direct addition of any amount of carcinogens to food. For US hazards not related to food, the maximum level of societal risk that can be imposed on the public, without any government regulation, is 10^{-6} (the risk imposes on each person an average annual chance of death of 1 in 1 million). The comparable level for workplace risks is approximately 10^{-3}. Proponents of uniform standards argue that zero risk is unachievable and unrealistic and that economic efficiency and fairness require such standards. Opponents of uniform standards argue that because the public is more averse to some risks, like cancer, different standards (including zero risk) are necessary for different risks. They also argue that all societal and workplace risks should be kept as low as possible, perhaps below the 10^{-6} level, and that uniform standards are too lenient in protecting only ‘average’ persons rather than especially sensitive individuals such as children (Shrader-Frechette 1991).

Apart from the particular norms one accepts for societal risk definition, evaluation and imposition, there is consensus on several general conclusions. Not all societal risks can be reduced by a technological ‘fix’. Not all can be resolved by a legislative ‘fix’ that imposes a hazardous facility on a group without its consent. Not all societal risks can be ameliorated by a public-relations ‘fix’. Most risk problems can be solved only by ethical analysis and democratic process: the most important aspect of risk is not scientific but ethical.

See also: Public Interest

KRISTIN SHRADER-FRECHETTE

References and further reading

Bullard, R.D. (1993) *Confronting Environmental Racism*, Boston, MA: South End Press.(Criticism of distributive inequities regarding risk.)

Cooke, R. (1993) *Experts in Uncertainty: Opinion and Subjective Probability in Science*, New York: Oxford University Press.(Provides suggestions for more reliable estimates of probabilistic risk.)

Cranor, C. (1992) *Regulating Toxic Substances: A Philosophy of Science and the Law*, New York: Oxford University Press.(An argument for changing the burden of proof to the plaintiff in toxic-tort cases.)

Douglas, M. and Wildavsky, A. (1982) *Risk and Culture*, Berkeley, CA: University of California.(Provides arguments that risk is a social construct.)

Gewirth, A. (1982) *Human Rights*, Chicago, IL: University of Chicago Press. (Explanation and defence of a theory of human rights.)

Harsanyi, J. (1975) ‘Can the Maximin Principle Serve as a Basis for Morality?’, *American Political Science Review* 69: 594-605.(Arguments in favour of utilitarian risk evaluations.)

Kahneman, D., Slovic, P. and Tversky, A. (eds) (1982) *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press. (Criticism of traditional scientific accounts of risk estimates and probabilistic reasoning.)

Kneese, A.V., Ben-David, S. and Schulze, W.D. (1982) ‘The Ethical Foundations of Benefit-Cost Analysis Techniques’, in D. MacLean and P. Brown (eds) *Energy and the Future*, Totowa, NJ: Rowman and Littlefield. (Argues that economic methods ought to take account of ethical values.)

MacLean, D. (ed.) (1986) *Values at Risk*, Totowa, NJ: Rowman and Allanheld. (Argues that risk decisions ought to take account of social and ethical values.)

National Research Council (1983) *Risk Assessment in the Federal Government: Managing the Process*,

Washington, DC: National Academy Press.(Classic survey of how to manage/control societal risks.)

Parfit, D. (1983) ‘The Further Future: The Discount Rate’, in D. MacLean and P. Brown (eds) *Energy and the Future*, Totowa, NJ: Rowman and Littlefield, 31-7. (Argues against economists’ use of discount rates.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Classic contemporary defence of social-contract ethics and egalitarian views of risk.)

Rescher, N. (1983) *Risk: A Philosophical Introduction*, Washington, DC: University Press of America.(Readable philosophical survey of problems related to risk.)

Sagoff, M. (1988) *The Economy of the Earth*, Cambridge: Cambridge University Press.(Criticism of traditional scientific accounts of economic methods for dealing with risk.)

Shrader-Frechette, K. (1985) *Science Policy, Ethics, and Economic Methodology*, Boston, MA: Kluwer. (Analysis of the values issues underlying risk-benefits analysis.)

Shrader-Frechette, K. (1991) *Risk and Rationality: Philosophical Foundations for Populist Reforms*, Berkeley, CA: University of California Press.(Presents arguments for lay, as well as expert, evaluations of risk.)

Shrader-Frechette, K. (1993) *Burying Uncertainty: Risk and the Case Against Geological Disposal of Nuclear Waste*, Berkeley, CA: University of California Press.(Criticism of risk evaluation methods that ignore ethical and social values.)

Risk assessment

Probabilistic or quantitative risk assessment (QRA) aims to identify, estimate and evaluate a variety of threats to human health and safety. These threats arise primarily from particular technologies (such as commercial nuclear fission) or from environmental impacts (such as deforestation). Defined in terms of the probability that some consequence will occur, 'risk' typically is expressed as the average annual probability of fatality that a particular activity imposes on one individual. For example, because of normal lifetime exposure to dichloromethane (DCM), a multipurpose solvent, the average member of the public has an annual probability of dying from cancer of 0.0000041 *or (* 4.1×10^{-6} *). Or, for every million persons exposed to DCM throughout their lifetimes, on average the chemical will cause four cancer deaths each year.*

Although risks may be individual *(such as those from consuming saturated fats) or* societal *(such as those from liquified natural gas facilities), government typically regulates only societal risks. By definition, they are largely involuntarily imposed, whereas individual risks affect only the persons voluntarily choosing them. Most QRAs address societal risks, either because a government seeks a scientific basis for particular risk regulations, because some industry wishes to determine possible liability for its processes or products, or because actual or potential victims want to protect themselves or to allocate risks by means other than market mechanisms.*

Philosophical contributions to QRA are of three main types: assessments of particular risks, criticisms of existing assessments, and clarifications of important QRA concepts, methods or theories. Such contributions usually focus on either epistemology (including philosophy of science) or ethics. Epistemological analyses address, for example, the adequacy and appropriateness of some scientific, probabilistic or policy technique used in QRA; the status of a specific causal hypothesis about risk; or the rationality of alternative decision rules for evaluating risks. Ethical analyses investigate, for instance, the equity of the risk distributions presupposed in a specific QRA or by general QRA methodology; the degree to which a particular method of risk evaluation accounts for crucial social values, such as free informed consent and due process; and the extent to which a given QRA technique, such as discounting the future (see Parfit 1983), begs important ethical questions such as rights of future generations.

1 The need for risk assessment

Agricultural, military and industrial technologies typically cause occupational and public fatalities. The World Health Organization estimates, for instance, that normal use of pesticides kills $40,000$ persons every year in developing countries. Even in developed nations like the USA, approximately $100,000$ workers die annually from accidents or exposure to compounds such as toluene or chromium. Members of the public also are at risk, as shown by catastrophes in Bhopal (India), Seveso (Italy) and Chernobyl (Ukraine). For example, experts predict that the 1986 Chernobyl nuclear-reactor accident will cause between $25,000$ and $475,000$ premature deaths from cancer worldwide.

Recognizing the enormity of societal hazards, and spurred by environmental legislation, in the 1960s many governments intensified their efforts in quantitative risk assessment (QRA). By the early 1980s, most regulatory agencies, industries and environmental groups - at least in developed nations - employed decision theorists, economists, engineers, epidemiologists, mathematicians, philosophers, various physical and social scientists, and toxicologists to assess risks associated with many technological activities and environmental impacts. In the USA during the 1980s committees of the National Academy of Sciences and the National Science Foundation developed both methodological standards and research priorities for QRA. Their goals were assisting society in determining which societal risks, relative to others, are likely to be most dangerous or most in need of regulation; protecting public health and safety; distributing societal risks and benefits equitably; and providing a framework for efficient and wise risk management.

2 Three tasks of assessment

Most QRAs include three tasks: risk identification, risk estimation and risk evaluation (National Research Council 1983, 1993; National Academy of Engineering 1986). Assessors perform *risk identification* by looking for clusters of harm associated with some environmental impact or technological activity. In the 1970s, for example, risk assessors identified vinyl chloride as hazardous by noticing the liver cancers occurring among rubber workers

exposed to the chemical. To identify vinyl chloride or other substances as hazardous, assessors need to be able to make a number of causal inferences and to show that particular clusters of disease or death are neither random occurrences nor the result of hidden variables (see Causation).

At the step of *risk estimation* scientists employ sensitive toxicological, biostatistical and epidemiological methods to estimate the population at risk (from a particular hazard), its level of exposure and the associated dose-response curve. By gathering data among affected persons, risk assessors were able to determine, for example, the average vinyl-chloride exposure for rubber-industry workers. Using exposure data, assessors develop a dose-response curve by determining which of many possible curves best explains and predicts the numbers of cases of disease, injury or death as a function of a particular dose of (or exposure to) a substance like vinyl chloride. Because data on humans frequently are not available, assessors usually interpolate and extrapolate from animal data to arrive at a dose-response curve. Particular curves are controversial because data are almost always incomplete and because different curves have different consequences for human health, for government regulations and for industry expenditures to control hazards.

Performing *risk evaluation*, assessors analyse whether a given societal activity - for example, use of commercial nuclear fission - is socially and ethically acceptable, relative to other risks associated with similar benefits. To determine risk acceptability, assessors generally employ economic methods (such as risk-cost-benefit analysis), psychometric techniques (such as expressed preferences or revealed preferences), and ethical analyses (such as weighting risk parameters on the basis of Paretian or Rawlsian rules (Kneese, Ben-David and Schulze 1982)) (see Applied ethics). A major problem with risk-evaluation conclusions, however, is their sensitivity to specific assumptions about measuring preferences, determining social choices, or quantifying risks, costs and benefits. For example, when risk assessors include the value of government subsidies for radioactive waste storage/disposal in their calculations of the costs of nuclear-generated electricity, it is virtually impossible to show a favourable benefit-cost ratio for this technology. Studies that assert the cost-effectiveness of commercial nuclear energy, however, typically ignore the effects of government subsidies - on the grounds that they are external to normal market processes. As this nuclear example illustrates, reliably assessing a particular risk-evaluation conclusion requires one to know both the methodological assumptions and the factual parameters to which the conclusion is most sensitive.

3 Epistemological issues in risk assessment

Because of its emphasis on methodological assumptions, QRA provides a new and important context for investigating traditional questions in epistemology and philosophy of science. Some of these questions, for example, address the realist-versus-antirealist debate over the status of scientific knowledge (see Scientific realism and antirealism). A central controversy in QRA concerns the degree to which both general and particular risk estimates are realistic. Natural scientists, more traditionally oriented philosophers of science, and proponents of normative epistemology tend to be more realist. They argue that although all risk estimates are assumption-laden, they are based on actual characteristics of phenomena that can, in principle, be analysed by mathematicians and scientists. Risk assessors in more realist camps also defend their claims by arguing that obtaining new data (such as accident frequencies) forces analysts to modify their risk estimates and illustrates that their assessments are not purely subjective (National Research Council 1983). Social scientists, more sociologically oriented philosophers of science, and proponents of naturalized epistemology tend to be more conventionalist. Arguing that all risk estimates are socially constructed and that no risk knowledge is unbiased, they claim that all risks are assessed on the basis of social, cultural and methodological values. More conventionalist assessors also maintain that, even if risk estimation were objective, risk evaluation would nevertheless be determined by our perceptions and constructs (Johnson and Covello 1987; Douglas and Wildavsky 1982).

Other epistemological debates over risk assessment mirror continuing metascientific controversies concerning the nature of rationality and the reliability of inductive inferences (see Induction, epistemic issues; Inductive inference). Does Bayesian rationality adequately characterize successful risk evaluation? Are given data adequate to estimate a particular risk? Do expert or lay opinions provide better evaluations of risk acceptability (as opposed to risk magnitude)? Some assessors argue that they, not members of the public, have more reliable knowledge of various risk probabilities and therefore are able to make more rational decisions about risk acceptability (Douglas and Wildavsky 1982). More populist assessors respond that, although laypersons may have inaccurate knowledge

of risk probabilities, lay ignorance in this regard is beside the point. They claim that members of the public have ‘the right to be wrong’ because rational decisions about risk acceptability are less a function of *knowledge* of risk *probabilities* than a function of the *value* they attach to avoiding potentially catastrophic *consequences* and to obtaining benefits from taking some risk (Shrader-Frechette 1985, 1991; MacLean 1986).

Philosophical debates over subjective probabilities and the role of expert judgment are central to QRA because risk controversies typically arise as a result of new technologies and activities. For example, roughly $60,000$ different chemicals are used annually in developed nations. For the $10,000$ new chemicals introduced each year, there are no established frequency records regarding potential hazards. Without such records, there are numerous debates over which estimates and models accurately represent particular risks. It is often not easy to settle these conflicts because of the many inferences, extrapolations and interpolations associated with QRA methods; because ethical sanctions against human experimentation limit available data; because of the expenses associated with long-term epidemiological testing and the large sample sizes or time periods needed to detect small effects, for example, of low-level exposures to toxic substances. In response to conflicts over the estimation of risk probabilities, philosophers of science have proposed new techniques for calibrating the subjective estimates of different assessors (Cooke 1992).

In addition to problems with risk estimation, empirical underdetermination in QRA also causes many metascientific and philosophical controversies over risk evaluation. One conflict, for example, concerns which decision-theoretic rules, are more appropriate to risk evaluation - those based on average expected utility, on maximin, or on some other rule (Giere 1991)? Proponents of expected utility would maximize average welfare in making risk decisions because any other procedure assigns too much weight to small risk probabilities. Proponents of maximin rules argue, in response, that in cases involving probabilistic uncertainty and potentially catastrophic risks, it is more rational to avoid the worst possible outcomes. Defending maximin rules, they claim that rational persons would accept neither potentially catastrophic consequences, even if their probability were low, nor risks with uncertain probabilities (Slovic 1987). Other risk debates focus on whether, in a situation of uncertainty, one ought to make the ‘equiprobability assumption’ that all unknown probabilities have the same value. What is appropriate risk/scientific behaviour in a situation of uncertainty?

Some of the most basic questions raised by philosophers of science interested in risk assessment concern the adequacy of particular causal inferences or predictions. What causal accounts explain a given number of fatalities? When should a scientist claim that the evidence of harm is adequate and therefore abandon the search for confounders (Cranor 1992)? When are analogies between animal models and human models close enough to infer that similar causal processes create similar risks? Or, which scientific accounts (such as hydrogeological models for groundwater flow at hazardous waste sites) adequately explain and predict phenomena relevant to risk estimation and risk evaluation (Shrader-Frechette 1993)? When is a particular scientific model too idealized to be useful in the practical work of risk estimation and evaluation? Why does one risk-estimation curve fit the dose-response data better than another? Is predictive power a necessary condition for explaining risk phenomena? Are assessors justified, especially in cases of potential catastrophe, in making an inference to the best explanation (see Idealizations; Inference to the best explanation; Statistics and social science)?

As the disputes over causal inferences reveal, epistemological controversies regarding QRA frequently concern the methodological adequacy of various techniques of risk assessment. These controversies are likely to persist, so long as we continue to purchase societal progress at the price of increasing particular risks and costs.

See also: Confirmation theory; Environmental ethics; Observation; Probability, theory and epistemology; Risk; Scientific method

KRISTIN SHRADER-FRECHETTE

References and further reading

Cooke, R. (1992) *Experts in Uncertainty: Opinion and Subjective Probability in Science*, New York: Oxford University Press.(Mathematical and philosophical analysis of how to take account of differing subjective probabilities used in risk assessment by one of the prominent European risk assessors.)

Cranor, C. (1992) *Regulating Toxic Substances: A Philosophy of Science and the Law*, New York: Oxford University Press.(Philosophical and ethical arguments for putting the burden of proof, in legal controversies over toxic substances, on the alleged perpetrators of the risk.)

Douglas, M. and Wildavsky, A. (1982) *Risk and Culture*, Berkeley, CA: University of California.(Analysis of risk behaviour in terms of three main constructivist or relativist accounts of risk, and an attack on environmentalist views of risk.)

Freudenburg, W. (1988) 'Perceived Risk, Real Risk: Social Science and the Art of Probabilistic Risk Assessment', *Science* 242 (7 October): 44-9.(Classical analysis and defence of the position that all risks are perceived, and that experts do not really know real risk, independent of their perceptions.)

Giere, R. (1991) 'Knowledge, Values, and Technological Decisions: A Decision-Theoretic Approach', in D. Mayo and R. Hollander (eds) *Acceptable Evidence: Science and Values in Risk Management*, New York: Oxford University Press.(Example of how to use decision-theoretic rules in risk assessment.)

Hornstein, D.T. (1992) 'Reclaiming Environmental Law: A Normative Critique of Comparative Risk Analysis', *Columbia Law Review* 92: 501-71.(Classic legal and ethical criticism of risk assessment.)

Humber, J. and Almeder, R. (eds) (1987) *Quantitative Risk Assessment*, Clifton, NJ: Humana Press.(An early philosophical analysis of some key issues in risk assessment.)

Johnson, B. and Covello, V. (eds) (1987) *The Social and Cultural Construction of Risk*, Boston, MA: Kluwer.(An early sociological analysis of risk assessment.)

Kahneman, D., Slovic, P. and Tvesky, A. (eds) (1982) *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.(Classic analysis of the ways in which subjective judgements of probabilities err, even among experts.)

Kneese, A.V., Ben-David, S. and Schulze, W.D. (1982) 'The Ethical Foundations of Benefit-Cost Analysis Techniques', in D. MacLean and P. Brown (eds) *Energy and the Future*, Totowa, NJ: Rowman & Littlefield. (Economists' analysis of how results of benefit-cost analysis and risk assessment can change if one employs ethical weighting criteria.)

Krimsky, S. and Golding, D. (1992) *Social Theories of Risk*, London: Praeger.(Well-known social and philosophical views of risk.)

MacLean, D. (ed.) (1986) *Values at Risk*, Totowa, NJ: Rowman & Allanheld.(Analysis of how risk assessment challenges some important values, especially ethical values.)

National Academy of Engineering (1986) *Hazards: Technology and Fairness*, Washington, DC: National Academy Press.(Important survey of risk-assessment problems that arise in connection with engineering risk assessment.)

National Research Council (1983) *Risk Assessment in the Federal Government: Managing the Process*, Washington, DC: National Academy Press.(Known as the 'Redbook', this is the 'how to' manual for all risk assessors.)

National Research Council (1993) *Issues in Risk Assessment*, Washington, DC: National Academy Press. (Work on several important problems in ecological risk assessment.)

Parfit, D. (1983) 'The Further Future: The Discount Rate', in D. MacLean and P. Brown (eds) *Energy and the Future*, Totowa, NJ: Rowman & Littlefield, 31-7.(Well known argument against the use of discount rates in benefit-cost analysis and risk assessment.)

Rescher, N. (1983) *Risk: A Philosophical Introduction*, Washington, DC: University Press of America.(Good mathematical and philosophical survey of risk assessment: key methods and issues.)

Schultze, W. and Kneese, A. (1981) 'Risk and Benefit-Cost Analysis', *Risk Analysis* 1: 81-8.(Overview of how one applies benefit-cost analysis to risk assessment.)

Shrader-Frechette, K. (1985) *Risk Analysis and Scientific Method*, Boston, MA: Kluwer.(Analysis of some of the main epistemological, scientific and ethical problems in quantative risk assessment.)

Shrader-Frechette, K (1991) *Risk and Rationality: Philosophical Foundations for Populist Reforms*, Berkeley, CA: University of California Press.(Argument against letting the expert and scientific community make decisions about acceptable risk, and argument in favour of more populist, procedural and ethical accounts of acceptable risk.)

Shrader-Frechette, K (1993) *Burying Uncertainty: Risk and the Case Against Geological Disposal of Nuclear Waste*, Berkeley, CA: University of California Press.(Scientific, ethical and logical analysis of arguments for the world's first permanent geological repository for nuclear waste.)

Slovic, P. (1987) 'Perception of Risk', *Science* 236: 280-5.(Explanation of how and why people perceive risks in particular ways.)

Ritual

Ritual, present throughout human affairs and central to many religious and cultural traditions, presents perplexities. One important question concerns the worth of such repetition and fixety - for example, in prayer, in human interaction, sometimes even in eating and drinking. To consider prayer, why not encourage the direct expression of religious thought and affect - from the heart, as it were, rather than in prescribed ways? It is sometimes suggested, and tempting to suppose, that to regularize such expression is to constrict it, ultimately to demean it. It is difficult to locate value in such apparently unnecessary regulation of human affairs.

In the context of philosophy the question becomes striking. None of the prevailing approaches to ethics makes it easy to see how ritual might possess ethical value or figure crucially in the ethical life. Yet this is precisely how ritualized ways are often seen within communities of practitioners.

1 The ethical import of ritual and traditional Judaism

Ethical theory, as we have it, seems unfriendly to ritual. For the Kantian, what distinguishes moral norms is their applicability to all rational agents. Rituals, however, are owned by communities, in whose highly particular and idiosyncratic idioms they speak. Indeed, rituals are sometimes optional even within the community. Nor does ritual come to mind when utilitarians reflect on behaviour that contributes to human betterment. Finally, engagement with ritual does not suggest itself to us, nor did it to Aristotle, as a trait that figures crucially in flourishing.

It will be helpful to examine a tradition or cultural setting in which ritual is taken seriously, in which it has a weighty, even a central, role. The present entry will focus upon traditional Judaism, a highly ritualized, comprehensive system of practice, indeed an inclusive legal system in which ritual permeates areas such as torts, contracts and divorce law. Ritual is pervasive, and since communal customs attain something of the status of law, the domain of ritual increases over time. How are we to think about the ethical status of ritual in the Jewish context?

A striking and suggestive feature of Jewish ritual is its essential role in the tradition's distinctive approach to human flourishing. Flourishing, of course, is the pivotal ethical notion, at least according to Aristotle and his followers.

Aristotle himself did not see ritual as essential to flourishing. Different cultural settings and traditions, however, may yield different ideas about human flourishing. And while not just anything should count as flourishing, a broadly Aristotelian view will want to allow for some latitude. The Aristotelian project as understood here - the empirical study of flourishing, and of the traits of character that contribute to and partly constitute human flourishing - thus becomes applicable to a wide range of cultural settings. Carried out in connection with traditional Judaism, the project yields the conclusion that ritual observance is ethically of the first importance. It is noteworthy that Mencius, in the Confucian tradition, counted *li* - often translated as 'propriety', a trait that centrally involved engagement with ritual - among the virtues.

What is the distinctive approach of traditional Judaism to human flourishing, and how is it that ritual is essential? At the heart of that approach is the conception of developed religious character. Ritual plays an essential role in the growth and sustenance of religious character.

2 Developed religious character: the role of awe

We are apt to think of a deeply religious person as a 'true believer'. However, there is no expression in biblical Hebrew that corresponds to our term, 'believer'. The phrase that comes closest is *Y'rei Adonai*, which means 'one who stands in awe of the Lord'. (The Hebrew expression *yirah*, translated here as 'awe', also means 'fear'. As A.J. Heschel notes, the term and its cognates are used in the Bible primarily in the sense of awe.) This suggests an emphasis on affect, orientation and responsiveness, rather than on the doxastic. Specifically, this Hebrew expression suggests that awe plays a central role. 'Awe rather than faith is the cardinal attitude of the religious Jew', writes A.J. Heschel (*1959*: 77).

One can see why awe might be of special interest by an examination even of quite ordinary - not religiously charged - experiences. Particularly important is a curious duality. In the grip of awe, one feels humbled; in the

extreme case, overwhelmed. At the same time, remarkably, one does not feel crushed or diminished, but rather elevated, exhilarated. This duality - humbled yet elevated - is of great importance for the religious orientation at which Judaism aims. (Needless to say, the features of awe experiences highlighted here and below vary in intensity and relative prominence from experience to experience.)

Awe experiences, perhaps as a consequence of the elevation-cum-humility, characteristically engender generosity of spirit, lack of pettiness, and increased ability to forgive and to contain anger and disappointment. The tradition notably associates just such affect and behaviour with God. Turning to the cognitive, awe experiences engender a Godlike perspective, the ability to see things, or almost see things, under the aspect of eternity, as Spinoza put it. One often also feels a powerful sense of gratitude.

Reflection on awe provides an entry point into the concept of holiness, a concept that is as important to Judaeo-Christian (as opposed to Greek) thinking about flourishing as it is difficult. In the grip of powerful awe, one often feels oneself to be in the presence of something sacred. To destroy the object of such an experience - the Grand Canyon, say - or to allow it to be destroyed, would be sacrilege. If ‘holy’ and ‘sacred’ have any natural application for non-theists, it is in connection with such moments. It is sometimes suggested that such reactions on the part of non-theists reflect the lingering presence of a religious upbringing, or an earlier time in our history. Instead, such reactions might be taken at face value, and as shedding light on the concept of holiness by providing a beginning point for reflection: awe seems to engender a sense of the holy.

These considerations provide a sense of what we might call the religious content of awe, and thus provide motivation for Heschel’s making awe pivotal. At the same time, powerful awe experiences are relatively rare and transitory. How can such uncommon and fleeting experiences bear so great a weight?

It is here that the concept of the *y’re Adonai*, one who stands in awe of God, comes to the fore. What is distinctive about such a religiously developed person is not only the object of awe. Perhaps even more important is the habitual quality, the steadiness, of awe. The *y’re Adonai* is one who has made awe a regular, albeit not a constant, companion. And with awe comes its concomitants: the sense of being humbled and yet elevated, the Godlike tendencies of thought, feeling and behaviour, the perspective *sub specie aeternitatis*, the gratitude, the sense of being confronted by the holy in all sorts of unlikely places. ‘Ordinary’ experiences, interactions with other human beings, for example, or with nature, become encounters with the holy. It has been said that close to the core of the Jewish religious attitude - and this perhaps represents an important contribution of Judaism to culture - is the idea of the sanctification of the ordinary.

To say that awe goes to the heart of the religious attitude is not to say that there are no other important aspects. A more complete treatment than is possible here would need to explore other such features, for example, the love of God, an aspect of the religious attitude that seems related, but not reducible to, awe.

3 Developed religious character: the role of ritual

To attain the heightened responsiveness of the *y’re Adonai* is quite an achievement. What is called for is a substantial change of orientation, a heightened responsiveness, and a deepening of wonder, of appreciation and of one’s character. Effecting such change is intrinsically difficult, and external factors often make it more so. The frailties and limitations of one’s fellows often inhibit their support for such development. And the distractions, discouragements, frustrations and sufferings of the human situation only increase the difficulty.

How then is such character development and sustenance possible? What are the tools by which the tradition means to effect its exceedingly ambitious plan? One often said to be the most fundamental is study of the tradition. Indeed, ‘study’ is not adequate, for it fails to convey the intensity of intellectual engagement. Jewish liturgy highlights an adaptation of Deuteronomy 30: 20, ‘For they [the words and teachings of the Torah] are our lives and the length of our days, and with them we are engaged [or we meditate] day and night’. Saintly personalities throughout post-biblical Jewish history are paradigmatically giants of scholarship, those who quite literally are engaged day and night.

Ritual plays a major role in effecting and sustaining the transformation. Consider the practice of saying blessings: on eating and drinking, on smelling fragrant spices, herbs, plants, on seeing lightning, shooting stars, vast deserts, high mountains, a sunrise, the ocean, on seeing trees blossoming for the first time in the year, on seeing natural

objects (including creatures) of striking beauty, on meeting a religious scholar, on meeting a secular scholar, on seeing a head of state, on hearing good news, on hearing bad news.

Heschel suggests that the practice of saying blessings is training in awe. One develops the habit, before so much as sipping water, to reflect and appreciate. In addition to the blessings' training function, blessings also function as reminders. Ordinary experience is distracting, and the tradition has assembled reminders to assist in maintaining focus.

In addition to occasion-related blessings, Jewish practice includes thrice-daily fixed prayer. Whatever one thinks of the practice of saying occasion-related blessings, the thrice-daily fixed prayer is likely to appear highly constricting. The blessings, at least, are appropriate to one's current experience.

We should remember, however, the magnitude and ambition of the tradition's project. While fixed prayer can degenerate into mechanical, unthinking, unfeeling performance, it offers great opportunities. Some of the usual translations notwithstanding, Jewish liturgy is a compilation of passages of literary magnificence. That such literature, the Psalms, for example, has survived the ages is a tribute to its expressive power, its ability to articulate and illuminate religious experience. To engage regularly with such literature - not merely to read the words but to declare them, to wrestle with them - is to occupy oneself with the tradition's project. The encounter with literature of such power encourages regularization of the attitudes to which it so ably gives voice. Indeed, ritualization turns out to be a great virtue: the agent need not wait until the appropriate experiences present themselves.

Ritualized prayer has another distinct advantage over spontaneous prayer. Spontaneous expressions - for example, of awe - are limited by the expressive capacities of the agent. How many among us are capable of summoning words adequate to powerful experiences and their concomitant thoughts and emotions?

Ritualized prayer does indeed present challenges of its own. The challenge is presented not by the repetition, but rather by the difficulty, the sheer hard work, involved in summoning up the thoughts and feelings appropriate to such literary magnificence. The founder of Hasidism, the Ba'al Shem Tov, is reputed to have said that it would be easier to deliver two advanced Talmudic lectures than to offer a single *amidah*, a fixed prayer of a few pages.

4 Rituals: linguistic and nonlinguistic

The examples of ritual we have been exploring all involve ritualized speech. Rituals need not involve linguistic performance, however, even in the highly word-oriented context of Judaism. Some may involve nonlinguistic behaviour that accompanies speech, as in bowing at appropriate times during prayer. Some may involve pure behaviour, that is behaviour not accompanying a linguistic performance - for example, the donning of a prayer shawl, or of phylacteries, or the burning of the leaven before Passover. These performances are accompanied by blessings (expressive, *inter alia*, of appreciation for the opportunity to perform the ritual), but the blessing accompanies them (and not vice versa). These are distinct from examples in which the action is a mere adjunct to the linguistic performance, for example, in bowing during prayer. Still other nonlinguistic rituals may involve no overt behaviour, as in the practice of hearing the searing call of the ram's horn every day during a month of the year devoted to reflection on the direction of one's life.

Ritualized prayer is ritualized engagement with dramatic and literary representation. It might seem that nonlinguistic ritual is distinctive in that it involves the actual performance of a dramatic representation, as opposed to merely engaging with one, as in prayer. There is, no doubt, something to this. Dramatic representations of ideas, for example, have a power beyond mere discursive articulations of the ideas themselves. It would be difficult to find a verbal equivalent of, say, wrapping oneself in a prayer shawl (where part of what this may mean to the agent is the wrapping of oneself in the tradition). Similarly, the call of the ram's horn is simply not to be equalled by any discursive substitute. Still, we should not forget that to engage in the prayer of Jewish tradition is not simply to read the text. It is rather to declare it, to make it one's own in speech and thought.

For the stranger to ritual, ritual will seem like rote performance, constricting rather than expansive, mechanical rather than expressive. One of the things lacking in such an attitude, according to the advocate of ritual, is a healthy respect for repetition. One who regularly engages with prayer, for example, can come not only to take comfort in the familiar, but also to see, in the repeated sentences, new depths. Since we are speaking of great literature, it will inevitably be the case that different sentences, or different turns of phrase, leap off the page, as

will different emphases, and different meanings and levels of meaning. And what is true of prayer is also true for the nonlinguistic rituals.

For the stranger to ritual, multiplication of ritual forms will seem like a burden. But for one's day to be sprinkled with such dramatic representations is for one's day to be sprinkled with powerful expressions of the attitudes and ethical motifs in question. It is also to be in possession of a multiplicity of reminders, mechanisms by means of which one may retain one's focus.

5 Conclusion

Our look at traditional Judaism suggests that in the right cultural setting, ritual can figure importantly in human flourishing. Having come this far, we can extend the point: whether ritual is ethically relevant for the utilitarian depends upon what counts as well-being. A different sense of what counts as flourishing may indeed yield a different sense of well-being. Even for Kantians, community-based ritual might well come to be seen as ethically relevant, as conducing, say, to an increased respect for persons.

The foregoing suggests that ethical relevance is not a purely formal matter, but depends upon the particulars of the culture or tradition. One might conclude that although ritual may indeed be ethically relevant, it is not so in our cultural setting. One may wonder, however, whether ritualized ways of acquiring, sustaining and celebrating values might play a more useful role than we have allowed.

See also: Prayer; Religious experience; Sacraments; Shintō

HOWARD WETTSTEIN

References and further reading

Douglas, M. (1970) *Natural Symbols: Explorations in Cosmology*, New York: Pantheon Books.(An argument for the indispensability of ritual.)

Durkheim, É. (1912) *Elementary Forms of the Religious Life*, trans. J.W. Swain, London: Allen & Unwin, 1915. (Classic discussion of the ritual nature of society and religion.)

Frank, D.H. (ed.) (1993) *A People Apart: Chosenness and Ritual in Jewish Philosophical Thought*, Albany, NY: State University of New York Press. (Contemporary discussion, including some discussion of historical figures. See especially Part 2, 'Ritual', and chapter 4, 'Rational Law/Ritual Law' by L.E. Goodman.)

Heschel, A.J. (1959) *God in Search of Man*, New York: Meridian Books, and Philadelphia, PA: Jewish Publication Society.(See especially chapter 4 (43-53), which is a discussion of wonder and its role in the religious life, of wonder and blessings, and of the ordinary as wondrous; chapter 7 (73-9, especially note 6), which examines awe and its relation to and primacy over faith; and chapter 15 (152-7), which characterizes faith as steady and heightened responsiveness.)

Rohault, Jacques (1617-72)

Rohault belongs (with Régis and de Cordemoy) to a generation which did much to consolidate the position of Cartesian physics in France. He is particularly famous for his experimental attitude. He contributed to the debate over the physical interpretation of the Eucharist.

Jacques Rohault was born in France at Amiens. Although he obtained a degree at Paris University, he seems to have been an autodidact in philosophy. He was a friend of Cyrano de Bergerac, of Molière and Clerselier (the best-known Cartesian philosopher of his time often cited in conjunction with the editor of Descartes' works), whose daughter Geneviève he married in 1664. Rohault was most famous for his public lectures on Cartesian philosophy, which from 1665 he held in his own house every Wednesday. In 1670 Rohault published *Traité de physique (Treatise on Physics)*, probably based on these lectures, followed by *Entretiens sur la philosophie (Conversations on Philosophy)* (1671). This book is essentially a defence of Descartes, dealing with controversial topics such as the relation between Descartes and Aristotle, Descartes' interpretation of the Eucharist and Descartes' ideas on the animal soul. Both of these works remained influential until the second quarter of the eighteenth century. Other works by Rohault, mostly on mathematics, were published posthumously by Clerselier in 1682.

According to Rohault, traditional natural philosophy is too 'metaphysical'. His own physics, on the other hand, is experimental. Indeed, the experiments and demonstrations were the great attraction of his Wednesday lectures. His concepts, however, are unmistakably Cartesian, although he does believe that Aristotelian concepts like 'matter' and 'form' can be made more precise: matter is extended matter, and forms are the various ways in which matter is modified into particular things. Apart from the fact that Rohault's physics is more complete (at least in terms of contemporary notions) than Descartes' (the *Treatise on Physics* also contains a part on animated bodies), it is also more overtly experimental (he even has a theory of experimentation) and therefore more attractive for students. Moreover, Rohault never claims more than probability for his theories. Physical explanations have the status of scientific hypotheses. They are true not because they are metaphysically warranted but because they provide the best account of experience.

One of the classical objections to Cartesian philosophy in Catholic countries was that Descartes' rejection of 'real accidents' makes it impossible to account for the fact that in the Eucharist the accidents belonging to the bread and the wine continue to exist independently from the original substance (which was changed into the body and blood of Christ). Also the definition of matter as extension (which implies that wherever there is any extension there is matter) would be contrary to the idea that Christ is really present under the dimensions of bread and wine. Rohault's general answer, to which he devoted his first 'Entretien' is that his theory is not contrary to that of Aristotle, so that, given that the traditional theory was based on Aristotle, there is nothing to worry about. Further, that theology and philosophy are mutually independent, so that no philosophical theory can ever be dangerous for theology. His specific reply is twofold. First he maintains that, since 'real accidents' cannot exist independently from a mind that perceives them, there is nothing strange in the fact that after the Consecration, God continues to give us the sensations connected with the bread and wine, although their substance is no longer there. Next he repeats what Descartes himself had said in the Fourth Set of Replies, namely, that the body of Christ affects the senses in the same way as the bread and the wine did because, in replacing them, its physical limits are formed by the same surface. This surface however is not part of the substance as such. Accordingly, the substance can change (the bread can change into the body of Christ), while the impression its surface makes on the senses remains the same.

See also: Sacraments §§2-3

THEO VERBEEK

List of works

Rohault, J. (1670) *Traité de physique*, Paris; Latin trans. *Tractus physicus*, 1674; English trans. J. Clarke as *Rohault's System of physics*, London, 1728-9, 2 vols; repr. with intro. L. Laudan, New York and London, 1969.
Rohault, J. (1671) *Entretiens sur la philosophie*, Paris; repr. ed. P. Clair in *Recherches sur le XVIIe siècle*, vol. 3, Paris, 1978, 101-64.

Rohault, J. (1682) *Oeuvres posthumes*, Paris; new edn, The Hague, 1690, 2 vols.(Of the items published in this collection, Rohault's treatise on mechanics was also published separately, first in Latin (*De arte mechanica tractatus mathematicus*, London, 1692), then in English (*A treatise of mechanicks*, London, 1716) and finally in French (*Traité de mécanique*, 1723).)

References and further reading

Armogathe, J-R. (1977) *Theologia cartesiana*, The Hague: Nijhoff, 108-10.(A good - and the only - work on the physical interpretation of transubstantiation.)

Balz, A.G.A. (1951) *Cartesian Studies*, New York, 28-41.

Bouillier, F. (1868) *Histoire de la philosophie cartésienne*, 3rd edn, Paris: Delagrave, vol. 1, 508-11.(Fundamental on the history of Cartesian philosophy in general; still useful, but on the whole now outdated.)

Clair, O. (1978) *Jacques Rohault (1618-1672): Bio-bibliographie avec l'édition critique des 'Entretiens sur la philosophie'*, *Recherches sur le XVIIe siècle*, vol. 3, Paris: CNRS.(Critical edition of an important text of Rohault; definitive on biography.)

Clarke, D. (1983) *Occult Powers and Hypotheses: Cartesian Natural Philosophy under Louis XIV*, Oxford: Clarendon Press.(Fundamental on Cartesian natural philosophy in France at the end of the seventeenth century.)

Gouhier, H. (1978) *Augustinisme et cartésianisme*, Paris: Vrin, 71-88.

Martinet, M. 'Jacques Rohault', in *Grundriß der Geschichte der Philosophie* (Überweg): Die Philosophie des 17. Jahrhunderts, Basel: Schwabe, Bd.2/2, 660-70. (Reliable, but short.)

Mouy, P. (1937) *Le développement de la physique cartésienne*, Paris: Vrin, 113-45.

Roman law

Law was Rome's greatest gift to the intellect of modern Europe. Even today the Roman law library, and the achievements of the jurists who built it up, live on in the law of the Continental jurisdictions and of other countries farther afield. It is true that over the past two centuries codification has largely interrupted the long tradition of direct recourse to the Roman materials, but the concepts applied in civilian jurisdictions and the categories of legal thought which they use are still in large measure those of the Roman jurists. In England, perhaps for no better reason than that from the late thirteenth century the judges of the King's Bench and Common Pleas happened to come from a background which cut them off from the clerical education which had given their predecessors access to the Roman library, there was no reception of Roman law. Post-Norman England thus became the second Western society to set about building up a mature law library from scratch. The common law (being the law common to the whole realm of England) and the civil law (being the ius civile*, the law pertaining to the* civis*, the citizen, initially of course the Roman citizen) thus became the two principal families within the Western legal tradition. It is wrong, however, to suppose that the development of the common law was constantly isolated. There have on the contrary been important points of contact at almost all periods. One result is that the categories of English legal thought are not in fact dissimilar to those of the jurisdictions of continental Europe. The study of Roman law has contributed immeasurably to the idea of a rational normative order, an idea fundamental to legal philosophy as indeed to all practical philosophy.*

1 Markers in time

The first life of Roman law, when it was the law of the Romans themselves, is conventionally allotted to the millennium which is almost exactly divided by the change of era, five hundred years before and five hundred years after Christ. There are convenient markers. Emerging from its prehistory, the tough agricultural society which had recently expelled the last of its kings enacted a crude code, the Twelve Tables, in 451 BC. Nearly a thousand years later, and no longer in Rome but in the predominantly Greek capital of the eastern Roman empire, the Emperor Justinian and his chief minister, Tribonian, carried through a task which to lesser men would have seemed intellectually and administratively impossible. They set up Commissions to read the whole law library and to compact it. The result was two huge volumes and one small one, the *Digest*, the *Codex* and the *Institutes*. These came out in the early 530s AD. Private collections of the emperor's subsequent pronouncements later added a fourth volume, the *Novels*. The four make up what we call the whole body of the civil law, the *Corpus iuris civilis*.

Within this thousand years, the focus of most modern scholarship has been the age of the greatest names, from the reign of Hadrian to that of Alexander Severus, roughly the century from AD 125 to 225. It is sometimes said that the strongest court that ever sat in England sat in York in AD 211, the year that the Emperor Septimius Severus died there, on campaign against the Scots. Papinian was certainly there, possibly also both Paul and Ulpian. One cannot but wish that the three greatest Roman jurists were there. The end of the golden age, the classical period of Roman law, is usually marked by the murder of Ulpian in the summer of AD 223. At the time of the assassination he was the first man in the empire below only the Emperor Alexander Severus himself.

Hardly less interest attaches to the preceding periods, since it is of them that the question has to be asked why Rome, and no other ancient society, managed to develop a mature law library. The end of the beginning was the work of Servius Sulpicius Rufus, who died in AD 43, in the tense months which followed the murder of Julius Caesar, on a mission to Marcus Antonius in Mutina. Pomponius' account of the history of Roman jurisprudence shows that Servius, both a prolific writer and an innovator in methods, more than doubled the number of law books which had been written before his time. He and his distinguished pupils prepared the way for the exponential improvement and expansion of the law library in the pre-classical or, as some would prefer, early classical first century AD, from Augustus to Hadrian.

After the assassination of Ulpian, the great names soon dry up, and in terms of military and political history the empire runs into deeply troubled times, from which it was saved, albeit at the cost of sowing the seeds of later division, by the Diocletianic settlement at the end of the third century. Whether the post-classical period was a time of steep and absolute decline in legal science is now doubtful. The form certainly changed. The responsibility for the interpretive development of the law shifted decisively to the emperor and the imperial bureaucracy, whose

characteristic utterance was the rescript, the written reply to a request for a ruling on the problems arising from a given fact-situation. The jurists who in earlier times would have been the great names achieving personal immortality through their books were now civil servants working in the anonymity of the imperial chancery. Decline and vulgarization cannot be denied, especially in the collapsing West, but the phenomena are more complex than was previously supposed. Although the *Codex Theodosianus*, a collection of imperial pronouncements put together and published in the early fifth century, does evidence a failing grip on the precise use of language, it is broadly true that the centre held. In the sixth century Justinian's Commissions are themselves evidence of complete recovery, based largely on the great law schools of the East at Beirut and Constantinople. Professors of law from those schools were leading members of the Commissions and had without doubt been vigorous advocates of the seemingly superhuman task of renewing the law library.

The western empire is taken to have ended with the abdication of an emperor whose name seems with specious elegance to close a great cycle of history, Romulus Augustulus (475-6). The second life of the Roman law, when its library was taken back into the foundations of western Europe, began about five hundred years after that collapse. From the embers of the western empire there ultimately sprang the European unity of the middle ages, that pre-Reformation Catholic civilization which is now most obviously attested by the great cathedrals and universities. The secular learning of that civilization was law, the law of Justinian's *Corpus iuris civilis*. The University of Oxford owes its existence in large measure to the fact that the Church happened to make Oxford a centre of jurisdiction and administration. By 1150 Roman law was being studied there, if not from the great Bolognese jurist Vacarius then at least from his writings. Oxford was not unique. In the twelfth century law was the study of the age. The appetite of both spiritual and secular powers for competent administrators meant that there were careers to be made by those who knew their law, and this in turn drew jurists out across Europe to spread the learning which was so much in demand. The centre from which this learning radiated was northern Italy. In the universities there a legal renaissance had begun. The study of the *Corpus iuris*, in particular of the *Digest*, had not merely been restarted but had quickly been raised to a remarkably high scholarly level. This legal renaissance is usually dated from the work of Irnerius in the University of Bologna. He died about 1125.

By 1100 the authors whose works contributed to the *Digest* had mostly been dead for upwards of eight hundred years. The generations of medieval professors of law, first the glossators and then the commentators, did not expound their texts as historians, revealing the law of a lost society which had itself changed as it developed through the centuries of its first life. On the contrary, they used the texts for their own time, as though they had all been written yesterday or outside time, and with scriptural authority. With brilliant ingenuity they explained away contradictions which we would recognize as due to juristic disagreement or to historical development. In some areas, though in remarkably few, no intellectual somersaults would suffice, and the Roman materials then had to be supplemented from more recent materials. Feudal land-tenure is one such case. The Romans knew nothing of feudalism.

In short the medieval expositors treated the Roman texts as a law library ready made for themselves and their own conditions of life, supplementing where absolutely necessary. In the sixteenth century humanist scholars reacted both against the scriptural authority of the Roman texts, with its tendency to stifle thought and undervalue contemporary experience, and against the intellectual dishonesty implicit in an unhistorical approach to the texts. Scholars such as Cujas (1522-90) and Hotman (1524-90) advocated and pursued a more critical and historical approach, more in the style of modern Roman lawyers who, relieved of the burden of adapting and applying the Roman law library to contemporary conditions, aim chiefly to rediscover the Roman law as it was and as it differed between, say, Sabinus in the first century, Julian in the second and Ulpian in the early third. Since the dissident humanists were mostly French, their critical historicism became known as 'the French way' (*mos Gallicus*) as opposed to 'the Italian way' (*mos Italicus*).

2 The *Corpus iuris civilis*

The work of Justinian's Commissions is the fulcrum upon which the story turns. The *Corpus iuris* sums up the first life of Roman law, and its rebirth in the West begins the second. Certain aspects of the work come as a surprise. The *Digest*, twice the size of the Bible, has 432 titles. One, for example, is 'Marriage', another 'Formation of the Contract of Sale' and another 'Theft'. Under those and the other headings the law was stated, not in the words of the Commissioners, but by their choice of extracts from the works written by the classical and pre-classical

authors. The book is thus a classified anthology of excerpts, each one carefully attributed.

The excerpts from the jurists as we see them have to a certain extent been managed. Many have been abbreviated, or doctored so as to be able to stand alone. Some routine modernizations have been introduced. Some imperfections and impurities in the manuscript tradition have survived. But, while these species of interpolation or disfigurement are not to be denied, few scholars now believe in the wholesale rewriting by the Commissioners that was suspected by many Romanists earlier in this century.

The *Codex* is somewhat less huge than the *Digest*. It is also a classified collection of material, but the matter collected, and once again managed by abbreviation and so on, is different. The *Codex* collects the pronouncements of the emperors, chiefly rescripts. The relationship between the two great books is chronological. The *Digest* compacts the library of the classical and pre-classical periods, while the *Codex* is the book of the post-classical period, when the imperial chancery monopolized all forms of authoritative utterance. Because in reality there was no abrupt transition from classical to post-classical or from private to nationalized utterance, there is some chronological overlap between the two volumes. The word ‘codex’ has nothing to do with codes and codification. It means ‘book’ and, in particular, the kind of book with pages and a spine. The transition from scrolls to that kind of book, the greatest revolution in information storage before printing and the computer, was complex but is generally located in the third century. The codex was thus the medium of the age which is represented in Justinian’s ‘Book’.

The *Institutes* was by comparison tiny. In size it is to the *Digest* much as St Matthew to the Bible. The genre can be traced back to Gaius in the second half of the second century. Indeed Justinian’s *Institutes* is essentially a new edition of the work of Gaius and actually re-uses substantial parts of the original. Gaius himself was long denied the full credit, because through the whole second life of Roman law no copy of his *Institutes* was known to have survived until in Verona in 1816 one was miraculously discovered, overwritten, by Niebuhr. Of Gaius himself we know only what can be wrung from his writings. It seems likely that he was a teacher of law. He certainly had a strong interest in legal history and believed that full understanding of an area of law could only be achieved by trying to go back to the beginning.

The function of the Institutes was to provide a coherent map of the whole law. It had a fixed place in Justinian’s scheme of legal education. Most of the first year was to be spent studying this overview of the whole law. Even deep in the jungle of the *Digest* and *Codex*, the mind thus formed would never be lost. This is an educational lesson that the contemporary study of law, throughout the civilized word, had taken to neglecting. However, in that the mind cannot dispense with categories, least of all the legal mind, jurisprudence cannot safely continue to neglect the fundamental categories of legal thought and their relation one with another (see Legal concepts).

3 Intellectual achievements

There are many negatives. The jurists were not interested in jurisprudence in the narrow modern sense, meaning legal philosophy. Cicero (§2) engaged in some discussion of the nature of law and its relation with the state and political theory, but Cicero, a contemporary of Servius Sulpicius, lived in the period before the law library had put on any weight; and he was, anyhow, not a jurist. Roman law produced no Austins, Kelsens, Harts or Dworkins. Next, from the standpoint of social justice there was little to admire. Slavery was an unchallenged feature of the Roman legal landscape. Again, access to justice was not obviously of much concern. It was a rich man’s system. Politically, the picture is equally grim. Though lawyers could fairly claim to have been a moderating and a stablizing influence, the constitutional structure within which they worked had no democratic aspirations. Indeed as time went on its tendency was to ever more absolute autocracy.

What can there be left? The question needs to be answered against an awareness of two facts. First, the requirement of justice that like cases be treated alike requires that legal science achieve stable analyses and solutions of a multitude of problems of responsibility, problems over which moral philosophers have the luxury of continual debate and uncertainty. The requirement of stability, not, be it noted, morbid rigidity, adds a dimension of intellectual difficulty to law which is absent from ethics and goes some way to explain why Blackstone presented law as the highest branch of ethics. Second, there is the not unrelated fact, perhaps naïve-seeming but none the less true, that there is a vast amount of law which is to a large extent politically neutral. For example, unless the very concept of private property is rejected, it is necessary to know what interests in land and material

goods can be held by individuals and how they are acquired, transferred and inherited. Again, raising the level of magnification, a mature legal system must be able to say exactly how a sale works, how the contract is formed, how it differs from neighbouring transactions such as hire and exchange, what obligations as to quality and title the seller undertakes, when the risk of damage passes to the buyer, at what moment and by virtue of what acts the buyer becomes owner. These and many other questions need stable answers applicable or adaptable to all imaginable versions of the facts. The analyses and their implications for litigation are likely to be more or less immune to politico-moral upheavals in the larger structures of society. If this were not so, it would be difficult to explain how it could be possible for medieval and early modern Europe to borrow the law library of an ancient and largely alien society.

The intellectual achievement of Roman law has to be seen at two levels. One is represented by the *Digest* and the *Codex*, the other by the *Institutes*. In large matters these great volumes are thoroughly disordered, rather as the English common law of the nineteenth century was, and today largely remains, in structural disarray. Nevertheless, when focused on matters of detail - when is legacy invalid, what if a neighbour's activities threaten damage or if he refuses you access to the aqueduct which brings your water across his land, what if a buyer is mistaken as to the thing bought, can an owner ever steal his own property, is it possible for a person to own a wild animal? - the jurists achieved a level of stable analysis unsurpassed by modern legal science. The acid test is whether, and how often, the intellect of the law is defeated by novel or unusual facts and driven back to abstract appeals to justice and good sense, alias gut reaction or mere personal prejudice. Although there were of course disagreements and uncertainties, some of them long unresolved, the ordered intellect of jurists such as Papinian, Paul and Ulpian was rarely defeated in matters of detail. Their concepts and methods set a standard of legal rationality which both challenges modern law and legal philosophy and explains the continuing importance of these practical and untheoretical jurists.

The second great intellectual achievement is represented by the *Institutes*. When he wrote his book, or delivered his lectures, Gaius developed a way of looking at the whole law in such a way that all the parts related intelligibly to each other. Mediated through Justinian's *Institutes* with some minor modifications, Gaius' scheme has dominated legal thinking ever since, and, though there have been bold attempts to escape it or improve it, it has by and large survived. Like many great intellectual advances it has about it a simplicity which almost defies attempts to recreate the pre-existing situation in which people had to do without it. Common lawyers, no strangers to serious disorder, can probably do it better than civilians (see Common law).

One kind of disorder, of which English law is overfond, is the alphabetical list of topics, but that was not quite how Roman law was before Gaius. It should rather be conceived as a historically determined list of actions (that is, of claims claimable), each having attracted its own packet of juristic interpretation. Thus the action of debt, called the *condictio*, was, only slightly simplified, 'I say you ought to give me 1000'. Generations of juristic tradition then built up a detailed commentary on that claim, showing who could make it, against whom, and, especially, what facts would as a matter of law substantiate it. Models of all the claims were set out and publicly displayed in the edict of the minister responsible for litigation, called the Urban Praetor. And if you knew all that the law library had to say about each one you knew virtually the whole law, albeit thrown together in a heap.

The institutional scheme put that heap in order. In Justinian's version, only minutely different from that of Gaius, it affirmed that the law was either public or private. Public law had the barest toe-hold in the *Institutes*, represented by the pages on the constitutional sources of law at the very beginning and the single chapter on crime at the end. Private law, the meat of the book, was first divided into three: all private law was about persons, things or actions - persons who make and defend claims, things claimed, and claiming. The law of things, essentially the law relating to assets of all kinds, was the largest of the three divisions of private law: assets were either corporeal, like land or a ship, or incorporeal. Incorporeal things were different kinds of right, rights over other people's property, such as a right of way or water, the right of inheritance, and rights to compel another person to make a performance, called (from the other end of the relationship) obligations. Obligations in turn arose from contracts or from wrongs or as though from a contract or as though from a wrong. Of course the *Institutes* is not just a diagrammatic representation of the law, like a family tree. At every level of classification it introduces the reader to the substance, so that it both reveals how the different areas of the law fit together and introduces the reader to the general nature of each one.

In the early modern period jurists who considered themselves emancipated from dependence on the Roman texts and wanted to write in their own language and of the contemporary law of their own country still used the scheme of the *Institutes* to provide their structure. In the seventeenth century Grotius did so in his *Introduction to the Jurisprudence of Holland*; and in Scotland Mackenzie did the same, as in the next century did Erskine. Scotland also provides a particularly fine example of a jurist who wrestled with the Roman scheme and hoped to improve it. Viscount Stair, a contemporary of Mackenzie, argued that the law of persons should be eliminated and, since the different conditions of persons were reflected in the different incidence of obligations, absorbed into the law of obligations.

Just as the *Institutes* either provided the structure of law books or the structure against which the author reacted, so also, when from the late eighteenth century the movement for codification gathered momentum on the continent of Europe, the codifiers either followed the institutional scheme, as in France, or sought to improve upon it, as in Germany. In fact, however, even the improvers, whether codifiers or authors, always left a good deal of the Roman scheme intact. It thus remains true that one who knows the *Institutes* is instantly at home in any of the civil codes, as also in the legal bookshops of the Continent. More remarkably, but rarely noticed, the categories of English legal thought and the curricula of English law schools reflect the same enduring influence. It is not by fortuitous parallelism that English law students read books and attend courses on contract and tort (or, as is becoming fashionable, on obligations), on property and on succession, and, on the public law side, on criminal law and constitutional law.

Classification lies at the foundation of legal rationality. The best evidence of the Roman genius for law is that the lawyers of Europe still use their scheme. It is a tribute both to Gaius and to the power of the teacher of law that if today we want to understand the categories of modern legal thought, or to improve them, we still have to follow generations of earlier lawyers in making ourselves thoroughly conversant with the *Institutes*.

See also: Bartolus of Sassoferrato; Law, philosophy of; Legal hermeneutics; Pothier, R.J.

P.B.H. BIRKS

References and further reading

Cairns, J.W. (1984) ‘Blackstone, an English Institutist: Legal Literature and the Rise of the Nation State’, *Oxford Journal of Legal Studies* 4: 318.(Studies the influence of the *Institutes* on early modern law. Cairns shows how in Blackstone English law flirted with the same systematizing tradition.)

Gaius (*c.* AD 161) *The Institutes of Gaius*, ed. and trans W.M. Gordon and O.F. Robinson, London: Duckworth. (Gives the Latin text of the *Institutes*, with a parallel English translation.)

Gaius (1981) *Emperors and Lawyers*, 2nd edn, Oxford: Oxford University Press, 1994. (This is a detailed study of the transition from the juristic tradition of classical Roman law to the bureaucratic law-making of the post-classical period, when jurists became anonymous civil servants in the ministry of justice.)

(1982) *Ulpian*, Oxford: Oxford University Press.(The first chapter tells us all we know of the life and times of the early third-century jurist whose work contributed more than any other to the *Corpus iuris civilis*. The rest of the book is an advanced study of the jurist’s style and pattern of work.)

Justinian (533) *Justinian’s Institutes*, trans. with intro. P. Birks and G. McLeod, London: Duckworth, 1987.(Gives the Latin text of the *Institutes*, with a parallel English translation. Also gives a substantial introduction explaining the importance of the *Institutes* in the later development of the Western legal tradition.)

Luig, K. (1972) ‘The Institutes of National Law in the Seventeenth and Eighteenth Centuries’, *Juridical Review* 17: 193.(Studies the influence of the *Institutes* on early modern law.)

Nicholas, B. (1962) *Introduction to Roman Law*, 3rd edn, Oxford: Oxford University Press, 1987.(This is a marvellous introduction not only to Roman law but to law in general. Read with the *Institutes* it provides the best foundation for further study of Roman law and legal thought and, incidentally, the best point from which any aspiring lawyer can get to know the concepts which law school takes for granted.)

Watson, A. (1981) *The Making of the Civil Law*, Cambridge, MA, and London: Harvard University Press. (Provides a way into the story of the second life of Roman law, its reception into modern private law.)

Wieacker, F. (1967) *Privatrechtsgeschichte der Neuzeit*, 2nd edn, Göttingen:Vandenhoeck and Ruprecht; trans. T. Weir, *The History of Private Law in Europe*, Oxford: Oxford University Press, 1995.(Provides a way into the story of the second life of Roman law, its reception into modern private law. A classic, to some extent requiring modification as a result of more recent research but still the most powerful and subtle account of the formation of modern European law.)

Zimmermann, R. (1990) *The Law of Obligations: Roman Foundations of the Civilian Tradition*, Cape Town: Juta.(This is a great work of comparative legal history which, confining itself to the law of obligations, shows how the Roman law was developed into the *ius commune* of Europe and then fixed in different forms by Continental fashion for codification. The author makes constant reference to parallels in the development of the common law. Though long and detailed, the book is extremely accessible even to the reader from the Anglo-American tradition.)

Romanticism, German

Because Romanticism has many meanings which vary according to time and place, it is best to examine the movement in a specific culture and period. Of all the phases of Romanticism, early German Romanticism is of special importance in the history of Western philosophy. The early German Romantics - Friedrich Schlegel, Friedrich von Hardenberg, Schleiermacher and Schelling - developed influential ideas in the fields of metaphysics, ethics, aesthetics and politics. The aim of their movement was essentially social and political: to overcome the alienation and disenchantment created by modernity, and to restore unity with oneself, others and nature. In accord with this aim, the Romantics advocated an ethics of love and self-realization, in opposition to hedonism and the Kantian ethic of duty. They championed an ideal of community against the competitive egoism of modern society; and, finally, they developed an organic concept of nature against the mechanistic worldview of Cartesian physics. Romantic ethics, politics and aesthetics should all be seen in the light of their essential cultural goal: to cure humanity of homesickness and to make people feel at home in the world again.

1 Intellectual geography

Romanticism was a European-wide literary and artistic movement, appearing chiefly in Germany, England, Italy and France, but also in virtually every other country. It began in the late eighteenth century, fully blossomed by the 1830s and began to dissipate by mid-century; some currents continued until the late nineteenth century. Although Romanticism is usually understood as an artistic or literary movement, it also has an important place in the history of philosophy. While rarely systematic, the Romantics developed original and influential ideas in epistemology, metaphysics, ethics and politics.

Considered in all its breadth, it is impossible to generalize about Romanticism. The movement differed from country to country and underwent many changes, one stage often contradicting another. To make any generalization about Romanticism, therefore, it is necessary to limit oneself to a specific time and place. It is best to begin with German Romanticism, partly because its ideas were deeply influential elsewhere, and partly because, in many respects, it marks the beginning of the movement as a whole.

German Romanticism is often divided into three phases: *Frühromantik* (early Romanticism) from 1797 to 1802; *Hochromantik* (high Romanticism) from 1803 to 1815; and *Spätromantik* (late Romanticism) from 1816 to 1830. From a philosophical viewpoint, the most important phase was *Frühromantik*, which flourished chiefly in the literary salons of Berlin and Jena. The leading figures of the early Romantic circle were August Wilhelm Schlegel (1767-1845), his brother Friedrich Schlegel (1772-1829), Ludwig Tieck (1773-1853), Wilhelm Heinrich Wackenroder (1773-1801), Friedrich Schleiermacher (1768-1834), and Friedrich von Hardenberg (1772-1801), chiefly known by his pseudonym ‘Novalis’. Other important thinkers on the fringes of this circle and connected with it in various ways and from time to time include Wilhelm von Humboldt (1767-1835), Friedrich Hölderlin (1770-1843) and August Ludwig Hülsen (1765-1810). The most important philosophers of the circle were Friedrich Schlegel, Schelling, Schleiermacher and Novalis. It is chiefly their views that are summarized below.

2 The early Romantic agenda

Early German Romanticism grew out of a disillusionment with some of the fundamental tendencies of modernity, more specifically the growth of science and technology, the division of labour, and a competitive market economy. For the Romantics, these tendencies had created the fundamental malaise of modern culture: alienation or disenchantment. There were three facets to this alienation: that people had become divided from themselves, from others and from nature. Each facet was the result of one of the tendencies of modernity. People became divided from themselves because of the increasing division of labour, which forced them to ignore their inner selves for the demands of productive labour, and which made them neglect the diverse sides of their humanity to specialize in one activity. They became divided from others because they had to compete in the market place rather than cooperate with them in a community. And they became divided from nature because the sciences treated it as an object to be dominated and controlled instead of a realm of beauty, mystery and magic (see Alienation).

The Romantics’ reaction to modernity has to be placed within the context of their general philosophy of history. They see history, along the lines of the Biblical myth, as a drama of innocence, fall and redemption. Kant and

Schiller had shown how this myth could be rationalized and secularized in explaining moral development; the Romantics duly followed this lead, applying the myth to all history. Innocence consists in unity with oneself, others and nature; the fall starts with division and alienation; and redemption consists in re-achieving unity after division. The present modern epoch is the depth of the fall. The task of the modern human is to achieve redemption by recovering unity with the self, others and nature.

The Romantics saw the rise of modernity as tragic, as the inevitable product not just of capitalism but of civilization. They looked back with nostalgia to earlier cultures, such as those of ancient Greece and the High Middle Ages, which enjoyed much greater unity. But their pessimism was also tempered by the optimism of their utopian belief that it was still possible for modern humans to approach redemption, even if they could not achieve it. If people directed their energies in the right direction, they would be able to recreate through reason that unity with the self, nature and society given to earlier humans through nature.

Hence the Romantic agenda sought to heal the wounds of modernity: to restore unity with the self, with others and with nature. They wanted everybody to become a whole person again, so that each realizes their unique individuality and all their distinctively human characteristics. They championed a community based on love and cooperation, rather than a society torn by self-interest and competition. And, finally, they hoped to restore the beauty, magic and mystery of nature in the aftermath of the ravages of science and technology.

3 Aestheticism

Early German Romanticism has often been understood as an aesthetic or literary movement. The Romantics indeed saw art as the highest form of human experience and self-expression, making it their instrument for the education and redemption of humanity. Romantic art had a definite objective, set by the Romantics' social and political agenda to restore the unity with the self, society and nature that had been destroyed by modernity. In this regard, the Romantics were deeply influenced by Schiller's programme for the aesthetic education of humankind, which gives primacy to art as the means to restore humanity to wholeness.

The early Romantic aesthetic, especially as defined by Friedrich Schlegel, has its origins in Schiller's reflections on naïve and sentimental poetry. According to Schiller, the poetry of the ancients was naïve because it simply and directly imitated nature; naïve poetry thus reflected the ancients' harmony with the world around them. The poetry of the moderns, however, is sentimental because it expresses a sentiment or feeling: the *longing* to return to that unity with nature that had once been given to the ancients.

Schlegel believed that Romantic poetry should adopt some of the characteristics which Schiller had attributed to sentimental poetry. It should be free of all the constraints of classicism, allowing the poet to express the freedom characteristic of modernity; but it should above all express the feeling and longing of the poet, the striving to return to unity with the self, society and nature.

Novalis' definition of Romantic art should also be understood in the same context. The aim of the artist, Novalis wrote, is *to romanticize* the world, to see the infinite in the finite, the extraordinary in the common place, the wonderful in the banal. Here again the mission of Romantic art was to restore unity: its purpose was to restore the magic, mystery and beauty of nature that had been lost with the growth of science and technology.

Romantic aestheticism has often been dismissed as quixotic, because it seems to exaggerate the power of art to motivate human action and to reform morals. But the Romantic faith in art should not be understood simply as a claim about the influence of the arts, and still less as a claim about their autonomy. Rather, Romantic aestheticism is based on the classical equation of the good with the beautiful. The Romantics believe that beauty should have a central role in culture because they think that beauty is involved in all forms of perfection, whether individual, social or political. The Romantic ideal of a self-realized person, of the ideal society and of the perfect state is that they should be aesthetic wholes. Hence the aim of the Romantic was not only to produce good poems, plays and novels, but first and foremost to make the individual, society and the state into works of art.

4 Ethics

The Romantic ethic is best understood in terms of a response to the classical question of the highest good, to the problem of the supreme value of life. The Romantics saw the highest good as *Bildung*, loosely translatable as

education or personal development. Theirs was an ethic of self-realization, which stressed the development not only of one's individuality but also of all one's characteristic human traits. This view of the highest good should be seen as a reaction against the two prevalent views in the late eighteenth century: hedonism and the Kantian-Fichtean ethic. According to the hedonists, the highest good is pleasure, while according to Kant and Fichte it is happiness in accord with virtue.

The Romantics rejected pleasure as the highest good because it would not develop those powers characteristic of our humanity and individuality. Their critique of hedonism is most apparent in their indictment of the philistines, those who devote themselves to a life of comfort and security at the expense of the real self.

The Romantics saw two fundamental difficulties for the Kantian ethic. First, Kant and Fichte had stressed reason at the expense of sensibility, ignoring how our senses are just as much a part of our humanity and just as in need of cultivation and development. It is not simply a purely rational being who acts morally but the whole individual who does their duty not contrary to but from their inclination. Second, by emphasizing action according to universal laws, Kant and Fichte had failed to see the importance of individuality.

The fundamental ethical value for the Romantics was love, which they believed should be the basis of social life and the heart of personal development. They pitted their ethic of love against the egoism of modern society, and the abstract and artificial commands of the Kantian-Fichtean ethic. They saw no conflict between the demands of love and individualism: the individual could become a unique whole only in virtue of love, which is the source of all personal development. Love rather than law was the chief social bond, holding that the state could become secure only through ties of affection and brotherhood.

5 Politics

The early Romantics came of age in the 1790s, a decade dominated by the French Revolution. Not surprisingly, their social and political ideals were formed in the crucible of that epochal event. At first they celebrated the Revolution as the dawn of a new age, embracing the ideals of liberty, equality and fraternity. As moderate republicans they believed that an ideal constitution should be a mixture of democracy, aristocracy and monarchy. They became progressively disillusioned with the Revolution, however, when they fully realized its cultural implications, and were horrified by the egoism, materialism and anomie of modern French civil society, which seemed to leave no place for community and spiritual values.

Because of its reaction against modernity, it would seem that Romanticism was, or at least became, an essentially conservative movement. Some of the Romantics were indeed eager to preserve some traditional aspects of European society, such as guilds and monarchy, because they saw these as a source of community and as a bulwark against materialism. Some looked back with nostalgia upon the society of the Middle Ages, which was more devoted to communal and spiritual values.

It is necessary, however, to place the Romantics' traditionalism in the context of their general social and political values. What they especially wished to maintain in traditional society was its pluralistic structure, its autonomous guilds and local corporations. They valued such a pluralistic structure chiefly because they saw it as a safeguard of liberty and a bulwark against tyranny. They were critical of all forms of political centralization and absolutism, whether it came from a revolutionary dictator or an absolute prince. In this regard they were moderates, as critical of the *ancien régime* as of the Revolution in France.

For all their disillusionment with modernity, the early Romantics endorsed one of its fundamental values: freedom of the individual. They never ceased to embrace this ideal, seeking to liberate the individual from all forms of oppression, whether political, social or cultural, so that people could fully realize and develop their individuality. Friedrich Schlegel, Humboldt and Schleiermacher were early champions of sexual liberation, of the right to divorce and to love without marriage. The main political problem for the young Romantics was how to square the freedom of individuality with the need for community. Their solution was a pluralistic society, whose autonomous groups would provide for community while serving as a defence of liberty.

6 *Naturphilosophie*

One of the characteristic products of early Romanticism was *Naturphilosophie*, which flourished in the first decade

of the nineteenth century (see Naturphilosophie). Its chief representative was Schelling, but it is important to see that he was only one figure in a much wider movement. Other important *Naturphilosophen* include Carl August Eschenmeyer (1771-1852), Lorenz Oken (1779-1851), Franz von Baader (1765-1841), Johann Wilhelm Ritter (1776-1810) and Gotthilf Schubert (1780-1860). It would be wrong to see all these thinkers as forming a single school, for they usually worked independently and were often critical of one another. Nevertheless, they did have some common characteristics: the vision of humans and nature as a unity, the attempt to see all the various phenomena of nature (magnetism, gravity and electricity) as manifestations of a single fundamental force, and a rejection of the old mechanical physics. The impetus for *Naturphilosophie* came from several sources: Kant's dynamic concept of nature, Herder's vitalistic pantheism, Goethe's natural investigations, and the vital materialist tradition of the English freethinkers and French *philosophes* (see Herder, J.G.; Goethe, J.W.).

Naturphilosophie developed from the Romantic concern to overcome the alienation between humans and nature. It saw the fundamental expression for this alienation in the Cartesian mental-physical dualism. The ultimate source of this dualism lay with the concept of matter and the mechanical model of explanation of Cartesian physics. According to Descartes, the essence of matter consists in inert extension; and we explain matter only if we show how the motion of one body is caused by the motion of another body upon it. This concept of matter and paradigm of explanation seemed to make the explanation of the mind impossible according to natural laws. Consciousness is not within space, and hence we cannot explain its changes by the impact of other bodies on it. If we follow Descartes' paradigm of explanation, the only options then seem to be dualism or materialism (see Descartes, R. §8).

Responding to this dilemma, Schelling developed a dynamic concept of matter, according to which the essence of matter consists in living force or power. The advantage of this concept of nature is that it seemed to surmount the dualism between the mental and the physical: they are now simply differing degrees of organization and development of living force. Mind is the highest degree of organization and development of the living forces of matter, while matter is simply the lowest degree of organization and development of the living forces of the mind (see Panpsychism).

For Schelling, the fundamental category of *Naturphilosophie* is that of an organism. He extends this metaphor to all nature, so that we should regard all nature as one vast organism, and mechanism itself as only an appearance of it. In his 1798 *Von der Weltseele (Of the world-soul)* Schelling explicitly revived the ancient Stoic doctrine of a single soul pervading all of nature. His general vision of nature could be described as a vitalistic monism or as a monistic vitalism.

FREDERICK BEISER

References and further reading

Collected works, critical editions and English translations of philosophers mentioned in the text can be found in the individual biographical entries throughout the *Encyclopedia*.

Ayrault, R. (1961) *La Genèse du romantisme allemand (The genesis of German Romanticism)*, Paris: Aubier, 3 vols.(A thorough historical account of the origins of Romanticism.)

Beiser, F. (1992) *Enlightenment, Revolution and Romanticism: the Genesis of Modern German Political Thought, 1790-1800*, Cambridge, MA: Harvard University Press.(Chapters 8 to 11 place early German Romanticism in its social and political context).

Behler, E. (1992) *Frühromantik (Early Romanticism)*, Berlin: Walter de Gruyter.(A useful survey that treats Romanticism as a literary movement.)

Bowie, A. (1997) *From Romanticism to Critical Theory*, London: Routledge.(Explores the origin of German Romanticism and the development of Romantic literary theory, tracing its continuation into the work of Heidegger, Benjamin and Adorno.)

Haym, R. (1870) *Die romantische Schule (The Romantic School)*, Berlin: Gaertner. (A brilliant classic, still the most substantial work on the early history of Romanticism.)

Heine, H. (1833) *Die romantische Schule (The Romantic School)*, Stuttgart: Reclam, 1976.(Witty and provocative critique of the Romantic movement.)

Huch, R. (1924) *Die Romantik (The Romantic)*, Leipzig: Haessel.(Perceptive study of central themes of Romanticism from its birth to its decline.)

Izenberg, G. (1992) *Impossible Individuality: Romanticism, Revolution, and the Origins of Modern Selfhood*,

Princeton, NJ: Princeton University Press.(A study of the early Romantic concept of the self in Germany, France and England.)

Kluckhohn, P. (1941) *Das Ideengut der deutschen Romantik (The ideas of German Romanticism)*, Tübingen: Niemeyer.(Introductory thematic treatment of central philosophical themes of early Romanticism.)

Pikulik, L. (1992) *Frühromantik (Early Romanticism)*, Munich: Beck.(Very useful survey of some of the basic works and themes of early Romanticism.)

Prang, H. (ed.) (1968) *Begriffsbestimmung der Romantik (Conceptual determination of the Romantic)*, Darmstadt: Wissenschaftliche Buchgesellschaft.(Useful anthology of essays dealing with the definition and historiography of Romanticism.)

Prawer, S. (ed.) (1970) *The Romantic Period in Germany*, New York: Schocken.(A helpful anthology on the literary and aesthetic dimension of Romanticism.)

Schelling, F.W.J. (1798) *Von der Weltseele, eine Hypothese der höheren Physik zur Erklärung des allgemeinen Organismus (On the world-soul: a hypothesis of higher physics in order to explain the general organism)*, Hamburg: Perthes; repr. in *Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart 1856-61, vol. 2, 345-583. (Explanation of organic nature in terms of a general dualism of opposing forces.)

Schmitt, C. (1925) *Politische Romantik (The Political Romantic)*, Munich: Duncker & Humblot.(Influential and controversial interpretation of Romanticism arguing that the movement was essentially a-political.)

Seyhan, A. (1992) *Representation and its Discontents: The Critical Legacy of German Romanticism*, Berkeley, CA: University of California Press.(An incisive treatment of the aesthetic and critical dimension of Romanticism.)

Silz, W. (1929) *Early German Romanticism*, Cambridge, MA: Harvard University Press.(A clear and straightforward account of early Romanticism that exposes many anachronistic interpretations.)

Walzel, O. (1918) *Deutsche Romantik*, Leipzig: Teubner; trans. A.E. Lussky as *German Romanticism*, New York: Putnam.(Incisive treatment of the philosophical aspects of early Romanticism.)

Rorty, Richard McKay (1931-)

Richard Rorty is a leading US philosopher and public intellectual, and the best-known contemporary advocate of pragmatism. Trained in both analytic and traditional philosophy, he has followed Dewey in attacking the views of knowledge, mind, language and culture that have made both approaches attractive, drawing on arguments and views of the history of philosophy from sources ranging from Heidegger and Derrida to Quine and Wilfrid Sellars. He takes pragmatism to have moved beyond Dewey by learning from analytical philosophy to make 'the linguistic turn', and from Thomas Kuhn that there is no such thing as 'scientific method'. Language and thought are tools for coping, not representations mirroring reality. Rorty's characteristic philosophical positions are what might be called 'anti-isms', positions defined primarily by what they deny. In epistemology he endorses anti-foundationalism, in philosophy of language anti-representationalism, in metaphysics anti-essentialism and anti- both realism and *antirealism, in meta-ethics ironism. He extols pragmatism as the philosophy that can best clear the road for new ways of thinking which can be used to diminish suffering and to help us find out what we want and how to get it. In the public arena, he is a leading exponent of liberalism and critic of both left and right.*

1 Life

Rorty's philosophical training at the University of Chicago and Yale University grounded him in the history of philosophy and the main currents of pragmatism and traditional philosophy dominant in the USA in the first half of the twentieth century, as well as the techniques and goals of the analytical philosophy then winning the allegiance of younger philosophers. Starting in the 1960s, he published articles and reviews addressing a wide range of philosophical subjects, the philosophies of thinkers as disparate as Whitehead, Dewey, Royce, Austin and Wilfrid Sellars, and, always, the metaphilosophical issues that arise from the multiplicity of conceptions of philosophy and its methods. He surveyed these metaphilosophical issues from a pragmatist point of view in the introduction to his anthology *The Linguistic Turn: Recent Essays in Philosophical Method* (1967). During the later 1960s and the 1970s he won wide recognition as a leading contributor to debates in the philosophy of mind, metaphysics, epistemology and the philosophy of language, staking out radical positions in each of these fields and drawing on the attacks on the philosophical tradition in Heidegger, Derrida, the later Wittgenstein and especially Dewey. He reached the first rank of US philosophers and won international attention outside philosophical circles with the publication of his major work *Philosophy and the Mirror of Nature* (1979). This book brought together all his lines of thought in an attempt to get philosophers, and those who look to philosophy to lend authority to their own cultural activity, to abandon the views of knowledge, mind, language and culture which give appeal to movements like analytic philosophy that claim to put philosophy on the true path of a science. Since then he has, in many articles, reviews and lectures, extended his attack on that philosophical tradition. He has defended his version of pragmatism as the best way of thinking about those things philosophers should still be thinking about; he has sought to place his thought in relation to an ever-widening range of other twentieth-century writers both within and outside philosophy (for example, Castoriadis, Davidson, Dennett, Foucault, Freud, Habermas, Lyotard, Nabokov, Orwell, Putnam and Roberto Unger) and to make pragmatism useful to workers in other areas (for example, feminism, education, jurisprudence and literary criticism). And he has taken up the role, well-established in Europe but rare among philosophers in the United States after Dewey's death, of a public intellectual, commenting in articles and interviews in journals of general circulation on affairs of common interest in democratic societies. Rorty's wide command of modern philosophy is remarkable, as is his dialogue with a broad spectrum of philosophical movements, and he is unusual also in giving credit to a number of other thinkers (Dewey, Heidegger, Davidson, Freud, Kuhn, Dennett, Quine, Sellars, Derrida, Nietzsche) for most of the key ideas in his writings. (Dennett, however, has enunciated the *Rorty Factor*: 'Take whatever Rorty says about anyone's views and multiply it by .742' to derive what they actually said.) Many of his most important papers and lectures have been collected in four volumes (listed below). Having urged a break with the traditional project of philosophy, he left the philosophy department at Princeton University to become a professor of humanities at the University of Virginia in 1982.

2 Metaphilosophy

Rorty takes the three models of philosophy most attractive to contemporary philosophers to be the scientistic (Carnap), the poetic (Heidegger) and the political (Dewey); he endorses the last as most congenial to pragmatism.

Philosophers should emulate the moral virtues which scientific communities have exemplified, but give up belief in ‘scientific method’ as well as philosophical method. They should foster the invention of new metaphors that is the work of poets and prophets, creating new vocabularies and social hopes; but philosophers have no special mission to be either poets or prophets. Philosophy should be hermeneutic, helping different areas of culture to make contact with each other and to sort out conflicts, just as liberal politics tries to find a working relationship among the different and conflicting desires and hopes at play in the world (see Hermeneutics). Its most important task now is to clear away past bad philosophy and bad ideas that are getting in the way of thinking usefully about, and doing, what can be done to make people happier.

As his metaphilosophy suggests, Rorty’s characteristic philosophical positions are what might be called ‘anti-isms’, positions defined primarily by what they deny. In epistemology he endorses anti-foundationalism, in philosophy of language anti-representationalism, in metaphysics anti-essentialism and anti- both realism *and* antirealism, in meta-ethics ironism (see Foundationalism; Essentialism; Realism and antirealism). None of these is a theory but, rather, a collection of considerations for the rejection of theorizing in these areas.

3 Knowledge and truth

Rorty’s anti-foundationalism, drawing on arguments of Quine and Wilfrid Sellars, denies that the justification of our knowledge claims must or can terminate in beliefs or statements that provide a foundation for knowledge. Nor does Rorty think that knowledge has any other overall structure. On his view (epistemological behaviourism), one justifies a belief or statement by adducing other beliefs or statements that do not require justification in that context, so as to satisfy the standards implicit in our social practice of justification (see Contextualism, epistemological). He consequently denies both that scepticism is a problem for philosophy and the theories whose appeal is that they purport to solve the problem of scepticism (see Scepticism). Rorty’s anti-foundationalism also denies that there is some particular discipline or part of culture that has the job of providing justification for, or making sense of, all the rest. In particular, he insists that it is a mistake to try to justify practices and institutions like liberal democracy, academic freedom and scientific research by appeal to philosophical theories that show them to correspond to or be in touch with the ultimate nature of reality or the human self or the will of God or something else bigger or deeper than our actual practices. This is not the view that such practices do not need any justification, which some critics have accused Rorty of holding, but rather that appropriate justifications will always be piecemeal and local (see Justification, epistemic §7).

Epistemological behaviourism is an account of justification but not a theory of truth. Rorty denies that there is any interesting theory of truth, that is, any general account of what makes beliefs and sentences true. He endorses James’ dictum that the true is just the good in the realm of belief; there is no general account of why beliefs are true any more than there is a general account of why things are good. He rejects correspondence theories and coherence theories of truth alike. His anti-representationalism denies that the essence of language is to represent or picture reality in such a way that bits of language match up with bits of reality; languages - the noises, gestures and marks we humans make - and thoughts - the brain states we get into - are part of the repertory of devices we have accumulated for coping with the world (including ourselves). While some stretches of language or thought might work by such matching-up techniques, these stretches have no special privilege or philosophical importance. Rorty rejects both realism and antirealism (or idealism) as products of a misguided representational view of language. No class of truths has foundational status with respect to the rest of the truths. While we usually call beliefs true when they are better justified than their competitors, ‘true’ does not *mean* ‘justified’ or ‘warranted’; it is indefinable and ineliminable.

Since there is no foundation for truth or knowledge, there is nothing outside our social practices to ground them, according to Rorty. Hence, he has sometimes rejected objectivity as a goal of inquiry (and objective-subjective as a relevant dimension of appraisal in the cognitive realm) in favour of solidarity with our community of inquirers; less provocatively, he urges that objectivity be understood as intersubjectivity or taken as shorthand for practices, like not taking bribes, that we have found very helpful in most kinds of inquiry.

Rorty is often accused of relativism and as often rejects the accusation. He claims that relativism about truth is easily refutable. The alternative to relativism about justification that he espouses is ‘ethnocentrism’, the view that justification is relative to *our* practices. The defence of our beliefs against challenges by other communities (Nazis, religious fundamentalists, the Nuer) must always be question-begging, but this does not vitiate the defence, since

no other kind of defence is better or even as good, and appraisal must always be against relevant alternatives (see Cognitive pluralism).

Opponents with epistemological and political concerns have regarded Rorty's ethnocentrism and epistemological behaviourism as viciously circular and conservative, making existing practices and institutions self-justifying and impervious to rational criticism, an objection also brought against other epistemological behaviourists such as Wittgenstein. Rorty's response has been to appeal to Kuhn and Davidson. According to Kuhn's account of scientific revolutions, relatively large-scale changes in the standards and practices of rhetorical communities are not justifiable by criteria available to those in the community before the change, but one result of the change is that new criteria become available to the reconstituted community after the revolution. Rational justification of the new beliefs and vocabulary, and rational criticism of the old beliefs and old vocabulary are available, but only when the new vocabulary and its attendant standards are established. According to Davidson's account of metaphor, metaphorical uses of old words to make new judgments that are false and even irrational by existing practices of justification, can change people's practices so that, as dead metaphors, the sentences are true and justified. Our accounts of the development of thought to its present (or future) achievements are always 'Whiggish'. The recognition that our most important values and practices are without foundation or non-question-begging justification is *ironism*.

4 Mind and society

As a social philosopher, Rorty advances two different lines of thought: on the one hand, he is a bourgeois liberal democrat celebrating 'the end of ideology' and advocating an incremental meliorism; on the other hand, he tries to keep the field open for, and even to cheer on, radically imaginative utopianism (for example, by feminists), however uncongenial to the stodginess of his first line of thought, provided it does not actually clash with either his liberalism or his anti-foundationalism. This has made him a target for political thinkers from all points on the political spectrum, who accuse him, even when he agrees with them, of giving comfort to the enemy.

In philosophy of mind, Rorty, following Ryle and Dennett, takes our talk of beliefs, desires and so on to be a way of talking which we find useful for predicting, controlling, imagining and making sense of some of the things that we and some other things do, just as physics and neurology give us ways of talking which are useful for predicting, and so on, other things that we and some other things do. No one of these is primary or needs to be reduced to the others, though we need not suppose that when we mention someone's beliefs and desires we are talking about something other than what we are talking about when we mention their neural processes (non-reductive physicalism). The mind as a philosophical subject Rorty takes to be a monstrous fiction cobbled together by Descartes out of thoughts and sensations; philosophers should drop talk of the mind, as it is confusing and unhelpful (eliminative physicalism).

Many strands in Rorty's thought have been neglected in the preceding sketch - his continuing skirmishes with philosophy as a profession, his ideas on the historiography of philosophy and his attacks on socialism, for example. Rorty has exhibited the virtues of modesty and willingness to learn from sympathetic criticism, and he has modified and extended his thought considerably throughout his career. He will undoubtedly have added new twists by the time you read this.

See also: Liberalism; Pragmatism

MICHAEL DAVID ROHR

List of works

Rorty's output runs to hundreds of articles and grows by a dozen or more items each year, often published in out-of-the-way places. A nearly complete list up to 1994 may be found in Saatkamp's 1995 work. I have listed his books and a few of the most influential or representative articles that have not, at the time of publication, yet appeared in his books.

Rorty, R. (1965) 'Mind-Body Identity, Privacy and Categories', *Review of Metaphysics* 19 (1): 24-54.(Much anthologized early defence of eliminative materialism with metaphilosophical corollaries.)

Rorty, R. (ed.) (1967) *The Linguistic Turn: Recent Essays in Philosophical Method*, Chicago, IL, and London: The University of Chicago Press, 2nd edn, 1992. (Anthology of metaphilosophical essays by ordinary language and ideal language analytic philosophers; long introduction by Rorty.)

Rorty, R. (1979) *Philosophy and the Mirror of Nature*, Princeton, NJ: Princeton University Press.(His major work; see §1 above.)

Rorty, R. (1982a) *Consequences of Pragmatism (Essays: 1972-1980)*, Minneapolis, MN: University of Minnesota Press.('Essays which might have some interest for readers outside of philosophy'.)

Rorty, R. (1982b) 'Contemporary Philosophy of Mind', *Synthese* 53 (2): 323-48. (Places his views in the context of views of contemporary analytic philosophers.)

Rorty, R. (1984) 'Diskussion/Discussion: A Reply to Six Critics', *Analyse & Kritik* 6: 78-98.(A good example of Rorty's responses to criticisms.)

Rorty, R. (1987) 'Thugs and theorists: A Reply to Bernstein', *Political Theory* 15 (4): 564-80.(Defence of Rorty's anti-Marxist social democratic liberalism.)

Rorty, R. (1989) *Contingency, Irony, and Solidarity*, Cambridge: Cambridge University Press.(Lectures on language, selfhood and politics plus interpretations of other writers.)

Rorty, R. (1991a) *Objectivity, Relativism, and Truth: Philosophical Papers*, vol. 1, Cambridge: Cambridge University Press. (Papers, 1980-89, dealing with issues and figures within analytic philosophy, especially Davidson.)

Rorty, R. (1991b) *Essays on Heidegger and Others: Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press.(Papers, 1980-89, dealing with Heidegger, Derrida and Foucault.)

Rorty, R. (1993a) 'Feminism, ideology, and deconstruction: a pragmatist view', *Hypatia* 8 (Spring): 96-103.(What Rorty thinks pragmatism can contribute to feminism.)

Rorty, R. (1993b) 'Putnam and the relativist menace', *Journal of Philosophy* 90 (9): 443-61.(An attempt to settle some differences with Putnam.)

Rorty, R. (1994) 'Does Academic Freedom Have Philosophical Presuppositions?', *Academe* 80 (6): 52-63. (Consequences of anti-foundationalism for academic freedom.)

Rorty, R. (1996) 'Who are we? Moral Universalism and Economic Triage', *Diogenes* 44 (1): 5-15.(Consequences of ethnocentrism for the politics of rich and poor.)

References and further reading

Geras, N. (1995) *Solidarity in the Conversation of Humankind: The Ungroundable Liberalism of Richard Rorty*, London: Verso.(Criticism of Rorty's ethnocentrism.)

Hall, D.L. (1994) *Richard Rorty: Prophet and Poet of the New Pragmatism*, Albany NY: State University of New York Press.(Good survey of Rorty's philosophy.)

Kolenda, K. (1990) *Rorty's Humanistic Pragmatism: Philosophy Democratized*, Tampa, FL: University of South Florida Press.(Sympathetic exposition of Rorty's philosophy.)

Malachowski, A. (ed.) (1990) *Reading Rorty: Critical Responses to Philosophy and the Mirror of Nature (and Beyond)*, Oxford: Blackwell.(Articles, mostly attacking Rorty's views; good bibliography.)

Saatkamp, H.J., Jr (ed.) (1995) *Rorty & Pragmatism: The Philosopher Responds to His Critics*, Nashville, TN, and London: Vanderbilt University Press.(Articles, mostly attacking Rorty's views, with Rorty's responses; good bibliography.)

Roscelin of Compiègne (c.1050-after 1120)

Roscelin of Compiègne was one of a group of logicians in late eleventh and early twelfth-century Europe who, in defiance of most of their predecessors in the field, treated logic as dealing with the concrete physical things that serve as verbal signs of realities rather than the realities those signs signified. This meant that although the things logic talks about are part of the physical world, it talks about them not as things referred to by language but as parts of language itself. All the technical notions of Aristotelian logic, for example 'universal', 'individual', 'category', 'genus' and 'species', apply only to those linguistic signs themselves qua *signs.*

Very little is known of Roscelin's life beyond the facts that he was born in Brittany around 1050, taught in France, was forced in 1092 to renounce certain positions he had taken on the doctrine of the Holy Trinity, and died after 1120. Of his work, only a letter to Peter Abelard is known to have survived. Consequently, what knowledge we have of his views comes to us second-hand, and in this case from the witness of two thinkers, Abelard and Anselm of Canterbury, known to be unsympathetic to his positions. Anselm tells us that Roscelin thought of universals, such as genera and species, as mere 'puffs of air', by which was meant the air expelled in uttering a word such as 'animal' or 'donkey' (see Universals). Also according to Anselm, Roscelin denied that qualities such as colours are anything over and above the subjects that possess them.

Roscelin's view of universals as concrete physical things used as verbal signs goes along with his apparent refusal to allow that anything could be common to many things independently of being a signifier of many. Abelard reports that Roscelin was 'insane' enough to draw from this view that wholes should not be thought of as divided into parts. This opposition to commonness is probably what led him to a form of tritheism in which there is no room for a divine essence that is genuinely common to all three divine Persons. He characterized the doctrine that the three Persons have but one substance as 'a manner of speaking', and such remarks were sufficient to bring down on him the charge of heresy.

What we know of Roscelin's views in logic entirely coincides with some positions adopted in a work entitled *Dialectica* dating from the mid or late eleventh century and ascribed to a certain 'Garland'. Professor L.M. de Rijk (1959) in his edition of this work has made a case for the work being by Garland 'the Computist', the author of an astronomical work called *Compotus*, probably a native of Lorraine and eventually a master at Besançon in 1084; but the authorship of the *Dialectica* is still very much in doubt. On the assumption that 'Garland' is the author's name, we can note that his overall approach to logic is probably very close to what Roscelin accepted. For example, when discussing the Aristotelian categories (see Categories), Garland equates a given category, such as substance, with the physical thing that serves as the category-word, for example, 'substance', plus all the words which are its species and individuals, such as 'body', 'human' and 'Socrates'. Any word 'signifies' the items it is true of, that is, those we would now say are in its extension. 'Animal', for example, signifies each and every animal. As for the intensional side of meaning, Garland speaks of 'modes of signifying' rather than anything signified. Thus both 'human' and 'wise' signify Socrates, but they do so in different modes. On this basis Garland draws the conclusion that of the category-words, that is, the ten highest genera serving to name the ten Aristotelian categories, each signifies everything, but in different modes. 'Substance', 'quantity', 'quality' and so on all signify the individual things named by the proper names in the category of substance; in other words, their extensions entirely coincide. It is very likely that something close to this was Roscelin's position as well. This relegation of the intensional side of signification to mere 'modes' removes from logic the temptation to posit real entities common to many individuals prior to any signification, the real universals we know Roscelin abhorred. Although Roscelin agreed that universals were things - that is, uttered puffs of air - their universality resulted from the use of those things as signs, and in that sense he was an anti-realist on the question of universals (see Realism and anti-realism).

This anti-realism doubtless had an effect on Roscelin's pupil, Peter Abelard, who in the early twelfth century conducted a sustained polemic against the various forms of realism which were current in his day. However, Abelard never credits his former teacher with having been right about anything, perhaps because he wished to distance himself as much as possible from a suspected heretic. Nevertheless, Roscelin's position that universals were only words and that commonness is a result, not a presupposition, of signification remained an important

option on this question throughout the medieval period, indeed throughout all subsequent philosophy in the West.

See also: Abelard, P.; Realism and anti-realism; Universals

MARTIN M. TWEEDALE

List of works

Of Roscelin's works, only a letter to Abelard is still extant; the text of this can be found in the appendix of F. Picavet *Roscelin philosophe et theologien d'après la légende et d'après l'histoire; sa place dans l'histoire generale et comparée des philosophies medievales*, Paris: Alcan, 1911. The further readings suggested for this entry offer further sources for Roscelin's thought.

References and further reading

de Rijk, L.M. (1959) *Garlandus Compotista: Dialectica*, Assen: Van Gorcum.(Contains the Latin text of a logical work which sustains positions very like those ascribed to Roscelin, as well as de Rijk's defence of the work's being ascribed to Garland 'the Computist'.)

Kluge, E-H. (1976) 'Roscelin and the Medieval Problem of Universals', *Journal of the History of Philosophy* 14 (4): 405-14.(The best attempt so far to reconstruct Roscelin's philosophical views.)

Meier, H.C. (1974) *Macht und Wahnwitz der Begriffe: der Ketzer Roscellinus (The Power and Madness of Concepts: Roscelin the Heretic)*, Aalen: Spieth Verlag.(An imaginative presentation of Roscelin's conflict with the church authorities.)

Mews, C.J. (1992) 'Nominalism and Theology Before Abaelard: New Light on Roscelin of Compiègne', *Vivarium* 30 (1): 4-33.(Claims Roscelin was not anti-realist about universals after all. Relates him to the theological and grammatical discussions of his day.)

Picavet, F. (1911) *Roscelin philosophe et theologien d'après la légende et d'après l'histoire; sa place dans l'histoire generale et comparée des philosophies medievales (The Philosopher and Theologian Roscelin, in Legend and History: His Place in History Generally as well as in the Comparative History of Medieval Philosophies)*, Paris: Alcan.(The appendix contains Roscelin's surviving letter and other important texts.)

Tweedale, M.M. (1988) 'Logic(i): From the Late Eleventh Century to the Time of Abelard', in P. Dronke (ed.) *A History of Twelfth Century Western Philosophy*, Cambridge: Cambridge University Press, 196-226.(Gives an overview of the trends in logic and grammar at the time Roscelin taught.)

Rosenzweig, Franz (1886-1929)

*An outstanding Hegel scholar - his Hegel und der Staat (Hegel and the State) (1920) remains a standard work on Hegel's political philosophy - Franz Rosenzweig elaborated, in Der Stern der Erlösung (The Star of Redemption) (1921) and several articles - notably, his 1925 article 'Das neue Denken' ('The New Thinking') - a philosophy of revelation that breaks with the systematic and rationalistic premises of German Idealism. The nisus of Rosenzweig's New Thinking was formulated as early as 1917, in a letter containing the germ (*Urzelle*) of Der Stern: 'after reason, ''philosophical reason'', has absorbed everything in itself', Rosenzweig writes, 'after it has proclaimed its sole existence, man suddenly discovers that he is still here, although he was digested long ago…. I am still here, I - plain, private subject, with first and last name, I - dust and ashes….* Individuum ineffabile triumphans*'. How can this be? The human being, Rosenzweig explains, can acquire personal identity as an individual only through the call, that is, the revelation of the Other: God - but also, some other human being. Dialogue, communication in language, comes to the fore in this philosophy, developed around the same time as Martin Buber's* Ich und Du (I and Thou) *(1923). Often treated as a common ground and basis for understanding between Jews and Christians, Der Stern der Erlösung is a major source of inspiration for such contemporary philosophers as Lévinas and is widely regarded as a masterpiece of Jewish philosophy.*

1 Life

Born in Cassel, Germany, Rosenzweig studied history (under Friedrich Meinecke) and philosophy (under Heinrich Rickert) at Freiburg and Berlin. His dissertation on Hegel's political philosophy became the first part of *Hegel und der Staat (Hegel and the State)* (1920). While in Berlin Rosenzweig discovered a manuscript in Hegel's hand, *Das älteste Systemprogramm des deutschen Idealismus (The Earliest Programmatic System of German Idealism)*, which he thought to be Hegel's copy of an original sent to Hegel by Schelling. The manuscript and Rosenzweig's analysis of it are still at the core of many discussions of German Idealism. But Rosenzweig himself turned in another direction. In July 1913, after intense discussions with friends and relatives who had converted from Judaism to Christianity, Rosenzweig resolved to become a convert himself. Determined to enter Christianity as a Jew and not 'a pagan', he attended the worship services of Rosh ha-Shanah and Yom Kippur and found there a spirituality he had not expected to find in his ancestral faith. Abandoning his planned conversion, he committed himself to the Jewish religious tradition. Drawing nearer to the Jewish Neo-Kantian philosopher Hermann Cohen, Rosenzweig wrote *Der Stern der Erlösung (The Star of Redemption)* (1921), most of it in the trenches during his years of military service in the First World War, completing the book in 1918. He settled in Frankfurt in 1920, where he founded the *Freies Jüdisches Lehrhaus* (Independent House of Judaic Studies). From 1922 until his death he battled with a severe illness which progressively paralysed him, but he continued to write on the New Thinking, Jewish thought, Jewish life in Germany, and the problems of translating the Hebrew Bible into German, a project which he undertook in collaboration with Martin Buber.

2 German Idealism

Rosenzweig's *Hegel und der Staat* shows the influence of his teacher, Meinecke. But it also breaks with Meinecke's widely influential reading of Hegel and evinces a pioneering independence of mind. In *Weltbürgertum und Nationalstaat (Cosmopolitanism and the National State)* (1907), Meinecke had made Hegel an early advocate of the Machtstaat, or power-state, a Machiavellian entity grounded in *Realpolitik* and exempt from all moral obligations towards the individual. In *Hegel und der Staat*, Rosenzweig breaks with this reading of Hegel's political philosophy. He offers a meticulously documented and still authoritative analysis of Hegel's intellectual development. Adopting a stance that has been enlarged and deepened but broadly confirmed by later research, Volume II interprets Hegel's mature idea of the state as a counter to both revolutionist and royalist/legitimist claims, owing much more to British parliamentary models than to prevailing Prussian traditions (see Hegel, G.W.F. §8).

Rosenzweig's reading of Hegel still holds to many of Meinecke's ideas (see Bienenstock 1992a). For example, it takes over Meinecke's definition of *Macht*, power, in terms of *Gewalt*, violence. Hegel in fact identified *Macht* with *Geist* (spirit). Rosenzweig notes that fact but never elucidates the true sense that Hegel gives the idea of Geist. Rather, he interprets Hegel's assertion of a link between the *Volksgeist* and the state in terms of a distinction

borrowed, once again, from Meinecke: between *Staatsnationen* and *Kulturnationen*; that is, nations, founded on the unifying force of a common political history and constitution, and those founded on some common cultural experience - linguistic, literary, or religious. To Rosenzweig it seemed clear that Hegel, who claimed that there can be no universal history without the state - that is, without a common *political* history - would have been able to account for the nation state but not for the *Kulturnation*, in which a common language or religion grounds a common history.

The idea of culture that Rosenzweig used here is again heavily indebted to Meinecke and tinged with a non-Hegelian, romantic conception of culture as the irrational source of a people's life. The manuscript that Rosenzweig had found in Berlin contains sharp criticism of the very idea of the state, this 'mechanical thing', and ends with a call for the foundation of a 'new mythology' among the people. It was because he saw these ideas as more in keeping with Schelling's project than with Hegel's that Rosenzweig ascribed the work to Schelling, but he felt deep affinities with the Programme. In a letter to his mother (3 July 1918) he remarks with approbation that the Jewish people is now 'the only people with a national myth' - and, for this reason, the only people beyond the state and beyond history. The remark sheds light on Rosenzweig's attitude toward German idealism in the *Der Stern der Erlösung*. Reacting there against Hegel's logocentrism, Rosenzweig draws upon Schelling's late, 'positive' philosophy. It was Schelling's *Philosophie der Offenbarung (Philosophy of Revelation)* (1856-8) and *Die Weltalter (The Ages of the World)* (1811-15) (read by Rosenzweig in the 1913 version) that helped him find his way back to a Jewish reading of the Bible.

3 The New Thinking

Rosenzweig's New Thinking originates in his rejection of Hegel's philosophy, which he sees as the climax of the philosophical enterprise at large, 'from Ionia to Jena'. This tradition, which endeavours to grasp the All, Rosenzweig argues, always reduces reality, in the end, to a single element, whether world (in ancient Greek philosophy), God (in the medieval period), or human spirit (in the modern). The New Thinking sets in with a broken totality. It speaks for the irreducibility of each element: God, world, the human being. Drawing on Kierkegaard, but also on Schopenhauer and Nietzsche, Rosenzweig writes at the very beginning of *Der Stern* that when I, as a human being, am confronted with death, my own personal death, I experience the impossibility of reducing it to nought (*Nichts*), and the corresponding impossibility of reducing the world to some rationally comprehensible 'being' (see Kierkegaard, S.A.).

Being, Rosenzweig insists, is not identical with thought or, for that matter, with God conceived as *logos*. Setting himself against the opening verse of John's Gospel, 'In the beginning was the Word', he contends that the Word (*logos*) comes second: 'The beginning is: *God Created*'. Here Rosenzweig returns not just to Genesis but to one of the great issues tackled in Jena around 1800 by Hölderlin, the young Schelling and Hegel: the problem of the nexus of myth and language. To ascribe the creation of all things to the Word, he argues, confines us to myth, to a world produced by a 'creative' activity like that of the poet or artist. Such an activity, he insists, involves no real language. For, instead of using words as symbols denoting things and thus different from those things, it presumes the identity of words with things; that is, it relies upon myth. To account for language, one must break with myth and so with the logos-philosophy grounded on the identification of word and thing, thought and being. For Rosenzweig this requires us to acknowledge the reality of Creation, and the main part of *Der Stern* - its 'heart' in Rosenzweig's words (Book II) - thus begins with Creation. Here Rosenzweig describes Creation - like Schelling in the *Philosophie der Offenbarung* - as a 'birth from the ground' (*Geburt aus dem Grunde*), manifesting itself through the 'alternation' (*Wechsel*) of two 'primal Words' (*Urworte*), irreducible to one another: Yes and No, 'Being' (*Sein*) and 'Nought' (*Nichts*), one principle through which God's 'essence' will assert itself, and another through which it will negate or restrain itself.

These Words, Rosenzweig insists, constitute no real language. The close affinity some have seen in Rosenzweig's conception of a *creatio ex nihilo* (*Schöpfung aus Nichts*) to Kabbalistic ideas is open to question (see Scholem 1930; Idel 1988). But its conceptual roots in Schelling should not be overlooked. Like Schelling, Rosenzweig aims at exposing the failure of any kind of idealistic dialectic that would arrogate to itself the power of explaining Creation - the emanationist doctrines elaborated in the Middle Ages and their culmination in the more recent, idealistic attempts of philosophers to produce the Many from the One. Against Hegel's 'pantheism', Rosenzweig builds upon Schelling's understanding of the relationship between myth and language and argues in

Der Stern that myth, which is typical of paganism, remains alive in us all. To a mythical consciousness, he writes, the world appears endowed with life, animated by forces perceived as both natural and divine. Such a consciousness resists distinctions between nature and culture, but also between past and present. To it, past events are still here, still living, permanently present in consciousness (see Schelling, F.W.J.).

For consciousness to distinguish between past, present and future, it must break loose from the mythical world in which it first lives. This, Rosenzweig claims, is what Revelation achieves. It replaces the mythic past, in which neither past nor present exist, with a past comprehended as such, thus making possible both a present and a future. Here Rosenzweig draws on the opening sentences of Schelling's *Weltalter*: 'the past is known (*gewußt*), the present is understood (*erkannt*), the future is foreseen (*geahndet*). That which is known is narrated (*erzählt*), that which is understood is described (*dargestellt*), that which is foreseen is prophesied (*geweissagt*)'. Building on these distinctions, Rosenzweig makes the past the age of Creation, the present the age of Revelation, and the future the age of Redemption (the three parts of the 'heart' of *Der Stern*). Creation is *narrated*, in the Bible; the present is *described*, and Redemption is *prophesied*.

In *'Das neue Denken' ('The New Thinking')* (1925), Rosenzweig urges that the human experience of past, present and future, is existential; it grows out of the inner life of consciousness; and its temporality, therefore, cannot be erased by thought. Thus the error of all philosophical systems in which thought is made timeless. Rosenzweig's insistence on temporality (*Zeitlichkeit*) anticipates Heidegger's *Sein und Zeit (Being and Time)* (published six years after *Der Stern*; see Löwith 1942). But where Heidegger acknowledges only 'existential' (temporal) truths, Rosenzweig discovers a truth beyond time, eternal, revealed truth. Revelation, he argues in *Der Stern*, enables us to apprehend Creation as an event belonging to the past, but is not itself an event in the past. Rather, revelation takes the form of a command issued to us in the present: the command to love God and, through God, our fellow human beings. To explain why this command dawns upon us as a revelation, Rosenzweig goes back beyond the Revelation at Mount Sinai, to God's call to Adam in Genesis (1: 9): 'Where are you?'. God's original Revelation occurs through this call. For Revelation is made to all human beings, not only to Israelites; and it dawns upon us from the outside. It is the revelation of a 'Thou', of an Other issuing a command. There is at first no 'I' to answer. Adam does not answer God's call. He tries to escape his responsibility - to evade the summons made to him personally, to him as an individual, who has a 'private name' - by hiding behind a pronoun: 'it', 'she' or 'he' (did it). An answer to God's call becomes audible in Abraham's 'Here am I' (Genesis 22: 1) - then, louder, in Israel's acceptance, renewed each day, of God's command. Only so does the revelation of a 'Thou' issue in a true dialogue with an 'I'.

Rosenzweig elaborates here, almost simultaneously with the Catholic teacher Ferdinand Ebner in *Das Wort und die geistigen Realitäten (The Word and the Spiritual Realities)* (1921) and with Martin Buber in *Ich und Du* (1923), what may be considered the first systematic presentation of the philosophy of dialogue (see Buber, M.). Rosenzweig, who knew both works, also criticized them. According to him, both Ebner and Buber so narrowed the 'I-Thou' relation that they lost sight of the 'We' and of nature.

Rosenzweig himself drew on the work of his friend Eugen Rosenstock-Huessy, a historian and sociologist of Jewish origin who had converted to Christianity and who was then teaching at the University of Leipzig. In his *Angewandte Seelenkunde (Practical Knowledge of the Soul)* (1924), a text Rosenstock-Huessy sent to Rosenzweig as early as 1916, the former had argued that the Cartesian *cogito ergo sum* should be replaced with a so-called 'grammatical' proposition: 'God called me, therefore I am. I am given a private name, therefore I exist'. He had made the human being 'a being that is spoken to' (*ein angeredtes Wesen*). He had sought to establish on this basis a true 'science of the soul' (*Seelenkunde*), a discipline that would do justice to the soul, instead of just giving up the word, as William James had done in his *Psychology*. According to Rosenstock, those who reduce the science of the soul to a 'science of the mind' (*Geisteswissenschaft*), turn language to logic and cannot understand what a 'community' (*Gemeinschaft*) is. They reduce it either to a mere collection of individuals moved by private interests, or to an entity understandable only as an 'it'. Such reductions are politically dangerous.

Rosenzweig fully agreed with his friend's diagnosis. He saw in the Jewish people a true community in Rosenstock's sense. Making much use of the Song of Songs, where 'strong as death is love' (8: 6), he contends in *Der Stern* that the dialogue of the 'lover' (God) with the 'beloved' (Israel) announces a Redemption already lived by the Jewish people in the present, although it is pregnant with the future. The eternal people are 'beyond

history' - always and already 'with God'. In this sense, Israel is the true incarnation of Eternity in time. Other peoples, particularly Christian ones, are 'on the way to God', that is, invested with the responsibility to realize the divine mission in history.

In keeping with this model, Rosenzweig opposed Zionism, thinking that Judaism can be lived only in Exile, at least as long as the history of the peoples of the earth has not ended. This conviction of his, much criticized for its seeming acceptance of the notion and the myth of the Wandering Jew, may also be seen as a variant of Hegel's idea that eternity itself is in time. Despite all his criticism of German Idealism, Rosenzweig remained enmeshed in it.

See also: Hegelianism; Jewish philosophy in the early 19th century

MYRIAM BIENENSTOCK

List of works

Rosenzweig, F. (1976-) *Franz Rosenzweig: Der Mensch und sein Werk: Gesammelte Schriften (Franz Rosenzweig: The Man and his Work: Collected Writings)*, The Hague, Dordrecht, Boston, MA and Lancaster: Martinus Nijhoff.(Collected works; four volumes published so far, under various editors.)

Rosenzweig, F. (1917) *Das älteste Systemprogramm des deutschen Idealismus: Ein handschriftlicher Fund (The Earliest Programmatic System of German Idealism: A Manuscript Find)*, Heidelberg: Winter; repr. in *Franz Rosenzweig: Der Mensch und sein Werk: Gesammelte Schriften*, vol. III, 3-44.(Rosenzweig's first, scholarly publication. Also sheds light upon his later thought.)

Rosenzweig, F. (1917) '"Urzelle" des *Stern der Erlösung*' ('"Germ" of The Star of Redemption'), in *Franz Rosenzweig: Der Mensch und sein Werk: Gesammelte Schrift*en, vol. III, 125-38.(First formulation of Rosenzweig's main ideas.)

Rosenzweig, F. (1920) *Hegel und der Staat (Hegel and the State)*, Munich and Berlin: Oldenbourg; repr. Aalen: Scientia, 1962.(Rosenzweig's scholarly study of Hegel's political thought.)

Rosenzweig, F. (1921a) *Der Stern der Erlösung (The Star of Redemption)*, Frankfurt: Kauffmann; trans. W.W. Hallo, New York: Holt, Rinehart & Winston, 1970; paperback edn, Boston: Beacon Press, 1972. (Rosenzweig's main work, in which he sets out a 'new thinking' to outline a history of culture and proposes a philosophical theology of Judaism and Christianity.)

Rosenzweig, F. (1921b) *Das Büchlein vom gesunden und kranken Menschenverstand*, D üsseldorf: Melzer; trans. and ed. N.N. Glatzer, *Understanding the Sick and the Healthy: A View of World, Man and God*, New York: Noonday Press, 1954.(Rosenzweig's world view.)

Rosenzweig, F. (1925) 'Das neue Denken: Einige nachträgliche Bemerkungen zum *Stern der Erlösung*' (The New Thinking: Some additional remarks to The Star of Redemption), *Der Morgen* 1 (4); repr. in *Franz Rosenzweig: Der Mensch und sein Werk: Gesammelte Schriften*, vol. III, 139-61.(Rosenzweig's own synopsis of his masterpiece.)

Rosenzweig, F. (1937) *Kleinere Schriften (Smaller Writings)*, Berlin: Schocken.(Contains many of Rosenzweig's 'smaller' but important writings, such as the 'Urzelle', 'Atheistiche Theologie' and 'Das neue Denken'.)

References and further reading

Anckaert, L. and Casper B. (1990) *An Exhaustive Rosenzweig Bibliography: Primary and Secondary Writings*, Leuven: Library of the Faculty of Theology; 2nd edn, 1995.(Comprehensive bibliography.)

Bergman, S.H. (1991) *Dialogical Philosophy from Kierkegaard to Rosenzweig*, New York: State University of New York Press, 173-214.(Vivid introduction to Rosenzweig.)

Bienenstock, M. (1992a) 'Rosenzweig's Hegel', *The Owl of Minerva* 23 (2): 177-82.(Detailed discussion of Rosenzweig's relationship to Hegel.)

Bienenstock, M. (1992b) 'Mythe et révélation dans *l'Étoile de la rédemption*: Contemporanéité de Franz Rosenzweig' ('Myth and Revelation in The Star of Redemption: Rosenzweig, Our Contemporary'), *Archives de Philosophie* 55: 17-34.(Analysis of the philosophical and historical context of Rosenzweig's thought.)

Buber, M. (1923) *Ich und Du*, Leipzig: Insel; trans. R.G. Smith, *I and Thou*, New York: Charles Scribner's Sons, 1958.(Buber's most famous presentation of his philosophy of dialogue.)

Cohen, R.A. (1994) *Elevations: The Height of the Good in Rosenzweig and Levinas*, Chicago, IL: University of Chicago Press.(Comparative study.)

Ebner, F. (1921) *Das Wort und die geistigen Realitäten: Pneumatologische Fragmente (The Word and the Spiritual Realities: Pneumatological Fragments)*, Innsbruck: Brenner; trans. H.J. Green, *The Word and the*

Spiritual Realities, Evanston, IL: Northwestern University Press, 1980.(The philosophy of dialogue, by one of Rosenzweig's contemporaries.)

Freund, E. (1933) *Die Existenzphilosophie Franz Rosenzweigs: Ein Betrag zur Analyse seines Werkes Der Stern der Erlösung (Franz Rosenzweig's Philosophy of Existence: An analysis of The Star of Redemption)*, Berlin and Leipzig: Felix Meiner; trans. S.L. Weinstein and R. Israel, The Hague, Boston, MA and London: Martinus Nijhoff.(General introduction.)

Glatzer, N.N. (1953) *Franz Rosenzweig: His Life and Thought*, New York: Schocken; 2nd edn, 1961.(The classic English introduction to Rosenzweig. Contains extracts from Rosenzweig's works in English translation.)

Idel, M. (1988) 'Franz Rosenzweig and the Kabbalah', in P. Mendes-Flohr (ed.) *The Philosophy of Franz Rosenzweig*, Hanover, NH and London: University Press of New England, 162-71.(The classic English introduction to Rosenzweig. Contains extracts from his works in English translation.)

Levinas, E. (1961) *Totalité et infini: Essai sur l'extériorité (Totality and Infinity: An Essay on Exteriority)*, The Hague: Nijhoff; trans. A. Lingis, Pittsburgh, PA: Duquesne University Press, 1969.(The work in which Levinas explicitly acknowledges his debt to Rosenzweig.)

Löwith, K. (1942) 'M. Heidegger and F. Rosenzweig, or Temporality and Eternity', *Philosophy and Phenomenological Research* 3 (1): 53-77.(Comparison between Heidegger's and Rosenzweig's philosophical points of departure.)

Meinecke, F. (1907) *Weltbürgertum und Nationalstaat (Cosmopolitanism and the Nation State)*, Munich and Berlin: Oldenbourg.(The work of the great German historian who was Rosenzweig's teacher.)

Rosenstock-Huessy, E. (1924) *Angewandte Seelenkunde (Practical Knowledge of the Soul)*, Darmstadt: Roether; trans. Jericho, VT: Argo Books, 1988.(A text written for Rosenzweig and sent to him as early as 1916, in which it is argued that the Cartesian *cogito ergo sum* should be replaced with a so-called 'grammatical' proposition: 'God called me, therefore I am. I am given a private name, therefore I exist'.)

Rosenstock-Huessy, E. (ed.) (1969) *Judaism despite Christianity: The letters on Christianity and Judaism between Eugen Rosenstock-Huessy and Franz Rosenzweig*, Tuscaloosa, AL: University of Alabama Press. (Contains a good introductory article by A. Altmann, and several other contributions.)

Schelling, F.W.J. (1811-15) *Die Weltalter (The Ages of the World)*, ed. L. Kuhlenbeck, Leipzig: Reclam, 1913; trans. with introduction and notes by F. de W. Bolman, Jr, *The Ages of the World*, New York: Columbia University Press, 1967.(Rosenzweig probably used the Kuhlenbeck edition. Schelling wrote several drafts of the text from 1810 onwards. The major source of Schelling's later philosophy, in which he outlines God's progressive realization of the divine self through self-consciousness of ideas within himself.)

Schelling, F.W.J. (1856-8) *Philosophie der Offenbarung (Philosophy of Revelation)*, Stuttgart: Cotta. (Posthumously published on the basis of numerous lecture notes. In this work, he outlines his belief that revelation, the incursion of the divine into history, is the point around which persons organize their world and experiences.)

Schmied-Kowarzik, W. (ed.) (1988) *Der Philosoph Franz Rosenzweig (1886-1926)*, Freiburg and Munich: Alber. (The Proceedings, in two volumes, of the International 1986 Conference on Rosenzweig in Kassel. An extensive collection, containing articles on most aspects of Rosenzweig's thought, with useful bibliographical indications.)

Scholem, G. (1930) 'Franz Rosenzweig and His Book *The Star of Redemption*', in P. Mendes-Flohr (ed.) *The Philosophy of Franz Rosenzweig*, Hanover, NH and London: University Press of New England, 1988, 20-41. (Scholem's speech at the Hebrew University of Jerusalem, thirty days after Rosenzweig's death.)

Rosmini-Serbati, Antonio (1797-1855)

In the reactionary, anti-Enlightenment, spiritualistic climate of Italy and Europe in the first decades of the nineteenth century, the Italian philosopher Rosmini set out to elaborate a Christian, Catholic system of philosophy which drew elements from Platonic, Augustinian and Thomist thought, while also taking account of recent philosophical developments, especially Kantian ones, as well as of the new liberal political trends in the culture of the time. His aim was to restore the principle of objectivity in the field of gnoseology, as well as in ethics, law and political thought.

Rosmini studied at Padua and was ordained in 1821. In 1826 he settled in Milan and two years later founded a religious congregation, the Istituto della Carità (Institute of Charity). In 1830 he published his first and most important philosophical work, the *Nuovo saggio sull'origine delle idee (New Essay on the Origins of Ideas)*, in which he distanced himself from empiricism and sensism, on the one hand, and idealism and rationalism, on the other. Empiricism and sensism, for Rosmini, offer less than what is needed, in that they attempt to derive knowledge and spirituality exclusively from experience. Idealism and rationalism, by contrast, offer more than is necessary, claiming, as Kant did, the innateness of all ideas as formal elements of knowledge. Both positions are characterized by subjectivism and thus lead to scepticism.

To get beyond subjectivism and scepticism, Rosmini argues, it is necessary to proceed from the proposition that the human intellect contains an original, essential idea, which does not derive from experience but is innate and thus imprinted by God, and which can be sensed through intuition. This idea is the idea of being in general: of indeterminate, possible being, in the sense that it can take on any determination and that it is the presupposition of all ideas and all judgments. It is this idea of being, this light of reason, that guarantees the objectivity of knowledge. This owes a great deal to the doctrines of Plato, Augustine and Aquinas, especially where the elements of 'innatism' and the illumination of reason are concerned. In the part of his system that deals with the processes of knowledge, however,Rosmini accepts certain fundamental aspects ofKant's critical philosophy.

One notion he derives from Kant is the distinction between material and formal elements and the need for a synthesis between them in the act of knowing (which, forRosmini, as for Kant, is essentially the act of judging) (see Kant, I. §6). He departs sharply from Kantian tradition, however, regarding Kant's a priori synthesis - what Rosmini calls '*percezione intellettiva*' ('intellective perception'). For Rosmini, the formal element is represented not by a mere transcendental category, but by the idea of being. Although this idea guarantees the objective validity of knowledge, it is not sufficient in itself to form the basis for determinate knowledge, to allow us to apprehend the reality of things. A direct encounter between the intellect and the sensible world is needed. This 'intellective perception' is the product of the synthesis between form, given by the idea of being and intuited by the intellect, and matter, which derives from sensations and from the '*sentimento fondamentale*': the underlying sense that all human beings have of their own life in its entirety and of which sensations are merely modifications.

The above notion of being or ideal being was particularly central to Rosmini's thought, supplying a philosophical foundation for his reconstruction of a spiritualistic metaphysics in his posthumously published *Teosofia (Theosophy)* (1859-74). In particular, it provided a grounding for the essential principle of that metaphysics, the existence of God. ForRosmini, ideal being is in fact merely one divine element of a personal Being, God. Ideal being thus postulates the real and absolute Being of God. In addition, Rosmini sees general ideas, such as identity, non-contradiction, and substance and cause, as deriving from the same notion of being through a process of abstraction and universalization. In this way, he constructs a comprehensive philosophical system, spanning the 'ideological' sciences, whose object is ideal being; the metaphysical sciences; and the deontological sciences, which concern themselves with ethics, law and politics.

In ethics, as in metaphysics, with such works as *Antropologia in servizio della scienza morale (Anthropology at the Service of Ethics)* (1838), Rosmini aims to restore the principle of objectivity, which he regards as having been abandoned in favour of subjectivism by both utilitarians and hedonists and by Kant. From the notion of being, which for him was the premise of gnoseological truth, there also stemmed the notion of the Good, which is the end of morality and the object towards which our love should be directed, according to a scale of values that has God at its summit, but which also has a place for persons, who have the status of ends in themselves, as they do in Kant.

The Good, which translates into moral law, not only can be known by the intellect, but should also be pursued by the will. The achievement of this objective can be impeded by the senses and by our instincts, as well as by individuals' free choice.

Rosmini's contributions to jurisprudence and political philosophy are closely bound up with his moral thought. It is moral law, in fact, which justifies personal rights - the ends to which society and state are the means. Civil society should under all circumstances adhere to the model provided by ecclesiastical society, which aims at a happiness lying beyond this world. In other words, it should look to the Church, which Rosmini considered was in need of reform, as he argues in his *Delle cinque piaghe della santa Chiesa (The Five Wounds of the Holy Church)* (1848a). From 1836, Rosmini lived in Piedmont, where, under the influence of liberal trends in political thought, he underwent an intellectual evolution that took him as far as a moderate constitutionalism and support for the notion of a federation of Italian states against Austria, expressed in his *Costituzione secondo la giustizia sociale (The Constitution According to Principles of Social Justice)* (1848b).

Rosmini was accused of psychologism (that is, subjectivism) by Vincenzo Gioberti, and he came under attack from Catholic critics (in 1888, the *Congregazione del Saint'Uffizio* condemned forty propositions drawn from his works). Idealists, however, considered him an Italian and Catholic version of Kant, and a point of transition to Neapolitan neo-Hegelianism. His attempt at a modern synthesis of Catholic philosophy has influenced modern Christian spiritualism.

See also: Italy, philosophy in §2

Translated from the Italian by Virginia Cox
GUIDO VERUCCI

List of works

Rosmini-Serbati, A. (1934-) *Opere edite e inedite di Antonio Rosmini (Complete Works)*, ed. E. Castelli and M.F. Sciacca.(The National Edition of Rosmini's works is published with the collaboration of the Istituto di studi filosofici, Rome and the Centro di studi rosminiani, Stresa.)

Rosmini-Serbati, A. (1830) *Nuovo saggio sull'origine delle idee (New Essay On The Origin Of Ideas)*, ed. F. Orestano, Rome: Anonima Romana Editoriale, 1934.(Rosmini's central work, this forms the basis for his entire theoretic, metaphysical, moral and legal-political system of philosophy. Vols 3-4 of the *Complete Works*.)

Rosmini-Serbati, A. (1831, 1837) *Principi della scienza morale e storia comparativa e critica dei sistemi intorno al principio della morale (Principles of Ethics, with a Comparative Critical History of Systems Concerned with the Principle of Morality)*, ed. D. Morando, Milan: Fratelli Bocca, 1941; trans. D. Cleary and T. Watson, *Principles of Ethics*, Durham, Rosmini House, 2nd edn, 1989.(The first work in which Rosmini uses the foundations laid by the philosopher di Rovereto to re-establish, on an objective basis, a universally valid moral science. Vol. 21 of the *Complete Works*.)

Rosmini-Serbati, A. (1838) *Antropologia in servizio della scienza morale (Anthropology at the Service of Ethics)*, ed. C. Riva, Roma and Milan: Fratelli Bocca, 1954.(Complements and enlarges on his principles of moral science. Vol. 25 of the *Complete Works*.)

Rosmini-Serbati, A. (1841-5) *Filosofia del diritto*, ed. R. Orecchia, Rome: Città nuova, 1967-8, 2 vols; vol. 1, trans. D. Cleary and T. Watson, *The Philosophy of Right*, Durham, Rosmini House, 1993.(In this work Rosmini illustrates the essential human rights - rights to truth and justice, rights regarding ones body and life, and property rights - which cannot be postponed, even for the sake of public welfare. Vols 35 and 36 of the *Complete Works*.)

Rosmini-Serbati, A. (1848a) *Delle cinque piaghe della Santa Chiesa (The Five Wounds of the Holy Church)*, ed. A. Valle, Rome: Città nuova, 1981.(In perhaps his best-known work, Rosmini illustrates what he perceived as the worst evils of the Church at the time and proposes remedies, from a liberal-catholic viewpoint.)

Rosmini-Serbati, A. (1848b) *Progetti di Costituzione. La Costituzione secondo la giustizia sociale (Constitutional Projects. The Constitution According to Principles of Social Justice)*, ed. C. Gray, Milan: Fratelli Bocca, 1952. (These reflect the constitutional phase of Rosmini's political thought, in which he forces himself to accept as fully as possible, with limitations of a confessional catholic nature, the aspirations of the moderate liberal movement. Vol. 24 of the *Complete Works*.)

Rosmini-Serbati, A. (1859-74) *Teosofia (Theosophy)*, ed. C. Gray, Rome: Edizioni Rome and Milan, Fratelli Bocca, 1938-9, 1940-1.(A posthumous work in which Rosmini brings out the first and most important consequence of the construction of his philosophical system, namely the restoration of a spiritualistic

metaphysics. Vols 7-14 of the *Complete Works*.)

References and further reading

Bergamaschi, C. (1967-82) *Bibliografia rosminiana (Bibliography of Works by and on Rosmini)*, Milan: Marzorati, 6 vols.(Very ample and helpful.)

Bulferetti, L. (1942) *A. Rosmini nella Restaurazione (Rosmini during the Restoration)*, Florence: Le Monnier.(On the evolution of Rosmini's philosophical and political thought.)

De Giorgi, F. (1995) *La scienza del cuore. Spiritualità e cultura religiosa in Antonio Rosmini (The science of the heart. Spirituality and religious culture in Antonio Rosmini)*, Bologna: Il Mulino.(In depth study of the spirituality and religious culture of Rosmini, set against the larger picture of contemporary agnosticism; with an extensive bibliography.)

Gentile, G. (1898) *Rosmini e Gioberti (Rosmini and Gioberti)*, Florence: Sansoni, 3rd edn, 1958.(Presents the idealist intepretation of Rosmini's thought.)

Piovani, P. (1957) *La teodicea sociale di Rosmini (Rosmini's Social Theodicy)*, Padua: Milani.(On the relation between theology and society in Rosmini's thought.)

Solari, G. (1957) *Studi rosminiani (Studies on Rosmini)*, Milan: Giuffrè.(A rigorous and scholarly study relating the development of Rosmini's thought to the cultural climate in which he lived.)

Traniello, F. (1966) *Società religiosa e società civile in Rosmini (Religious society and civil society in Rosmini)*, Bologna: Il Mulino.(An acute study of Rosmini's life and thought, from a fundamentally ecclesiastical and religious perspective.)

Ross, Alf (1899-1979)

Famous for his contribution to Scandinavian legal realism, Alf Niels Christian Ross was among the major philosophers of the latter half of the twentieth century. He was Kelsen's ideal successor and on a par with H.L.A. Hart in terms of notoriety and international influence.

Ross opposed equally natural law theories and formalistic legal positivism and, following Hans Kelsen whose Vienna seminars he attended in 1924-5, he insisted that the central questions of law were epistemological (what is law? what is legal knowledge?) yet rejected the 'two-world' dichotomy of *Sein* and *Sollen* ('is' and 'ought') with their separate methods of investigation. The rejection of dualism and dilemmas on how to characterize the interplay of law's normative and factual character condition his entire life's work (see Legal positivism §5; Natural law).

Under the influence of Axel Hägerström (§4) Ross found the philosophical basis on which to ground his need for realism. Ross embraced the theory that a concept of reality is not logically conceivable outside the context of space and time, the emotivistic concept of values and the noncognitive notion of value judgments, and the rejection of voluntarism as a theory of positive law. Ross claimed that 'legal phenomena' rather than 'positive law' must be investigated. These comprise mental impulses and attitudes of impartial behaviour, rationalized in the use of normative and evaluative linguistic expressions (validity), as well as attitudes of biased behaviour, construed as the subjectivization of coercion. Although formally distinguishing them, Ross equates legal and moral validity, both understood as psychological phenomena of rationalization. Naturally he now abandoned his previous view that specific normative knowledge is possible. Legal science, including legal dogmatics, was to be subsumed in the wider sphere of psychological and sociological sciences which study human behaviour. Yet the concern did not abandon the need to explain at least the feeling of obligation, being bound by law.

In the final and most famous phase of his work, Ross turned to logical positivism. In the preface to the English edition of *On Law and Justice*, Ross coupled Anglo-American jurisprudence from John Austin to Oliver Wendell Holmes with the Scandinavian tradition from Anders Sandoe Orsted to Hägerström. Ross declared that his aim was to trace empiricist principles to their ultimate conclusions in the field of law. The principle of verifiability provided the means to do this: any statement has a meaning in so far as it implies that, following certain procedures, certain results can be obtained, the set of results representing the 'real content' of the statement. Hence legal statements of the type 'Directive D is a valid law' signifies that 'by applying the system of norms as a scheme of interpretation, we are enabled to comprehend the actions of the courts as meaningful responses to given conditions' (Ross 1958: 39). The procedure of verifying such propositions will be the behaviour of the courts which settle legal disputes (see Meaning and verification).

In *Directives and Norms* (1968) Ross adopts and modifies R.M. Hare's distinction between phrastic and neustic in order to ascribe to norms a semantic meaning denied earlier. Legal dogmatics none the less remains an empirical science, no longer a branch of sociology since this discipline is concerned with law in action while legal dogmatics studies the abstract ideal content of legal norms. Now too the biased and impartial attitudes are separated out, no longer construed as part of the reality of law, but rather, part of the phenomenon of reasons that induce people to obey the law.

See also: Law, philosophy of; Legal realism §1; Voluntarism

ENRICO PATTARO

List of works

Ross, A. (1933) *Kritik der sogenannten praktischen Erkenntnis (Critique of So-called Practical Knowledge)*, Copenhagen and Leipzig: Munksgaard.(A critique of theories about 'so-called practical cognition', presenting a strongly anti-cognitivist moral and legal theory.)

Ross, A. (1946) *Towards a Realistic Jurisprudence: A Criticism of Dualism in Law*, Copenhagen: Munksgaard. (An attack on the dualism of norms and facts, arguing that in a scientific perspective norms are merely psychological contents of thought which have a causal impact on behaviour.)

Ross, A. (1957) 'Tû-Tû', *Harvard Law Review* 70: 812-25.(A witty and influential essay, using imaginary

anthropology to suggest that legal terms like ‘right’ and ‘ownership’ have no semantic meaning, serving in legal discourse only as a link between conditioning facts and conditioned consequences.)

Ross, A. (1958) *On Law and Justice*, London: Stevens and Sons.(Readable statement of ethical noncognitivism coupled with legal realism, reducing legal statements in the last resort to predictions of behaviour and emotive utterances tending to affect behaviour.)

Ross, A. (1968) *Directives and Norms*, London: Routledge & Kegan Paul.(Final and most mature statement of Ross’ position, modifying it to take into account the work of R.M. Hare and other analytical moral philosophers, developing a theory of norms that makes it possible to ascribe meaning to them.)

References and further reading

Harris, J.W. (1978) *Law and Legal Science*, London: Butterworth.(Advanced discussion of leading positions in modern jurisprudence, containing a sympathetic account of Ross’ psychological approach to the explanation of law, but preferring a more Kelsenian thesis about the nature of legal norms and legal science.)

MacCormack, G.D. (1971) ‘Scandinavian Realism’, *Juridical Review*, new series, 14: 259- 81.(Discussion of realist accounts of law in terms of ‘facts’, criticizing the psychologism of Ross and Karl Olivecrona.)

Marshall, G. (1957) ‘Law in a Cold Climate: the Scandinavian Realism’, *Juridical Review*, new series, 1: 259-81. (Discussion in dialogue form of some main points of Scandinavian realism, with prominent attention to Ross.)

Simpson, A.W.B (1964) ‘The Analysis of Legal Concepts’, *Law Quarterly Review* 80: 535-58.(Critical account of certain leading discussions of legal concepts, notably those of H.L.A. Hart and of Ross.)

Ross, William David (1877-1971)

W.D. Ross was a British ancient and moral philosopher. In terms of his moral thinking, he was a pluralist, who held that there are several distinct moral considerations which bear on the rightness of an action. Among the things we need to take into account are promises we have made, the need to avoid harming others, gratitude to benefactors, and the amount of good our action will produce. That these considerations are morally relevant is something we can know, but which action is the right one is a matter of fallible judgment, because that will depend upon how these considerations are to be weighed against each other in the particular case. Ross' contributions to the study of ancient philosophy mainly concerned Aristotle. He is now best known, however, for his moral philosophy.

W.D. Ross was a Fellow of Merton College, Oxford, before moving to Oriel College, where he became Provost. He was knighted in 1938. Ross made contributions to the study of ancient philosophy and in particular Aristotle, though he also published a book on the Platonic Forms (1951). He edited the Oxford Translation of Aristotle, wrote an influential book on Aristotle, and published several important editions of Aristotelian texts, with commentaries. Now, though, he is best known for his moral philosophy.

In his two major ethical works, *The Right and the Good* (1930) and *The Foundations of Ethics* (1939), Ross offers the clearest and most ably defended account of a pluralist deontology in the twentieth century (see Deontological ethics; Moral pluralism). One of Ross' main targets is the consequentialism of G.E. Moore (1903), which held that, in determining whether an action is right, there is just one consideration that is relevant, namely the amount of good that will result from doing that action (see Consequentialism; Moore, G.E.). Ross held, by contrast, that a number of features of actions are morally significant, and that which action is right will depend on which of these features carries the most weight in a particular case.

Deontology thus distinguishes itself from consequentialism in claiming that the right is independent of the good (see Right and good); the right action is not necessarily the one that produces the most good. Ross defends this claim by arguing that being productive of the most good is neither what it *is* for an action to be right, nor what *makes* an action right. The right action is also distinct from the morally good action; the latter depends on the agent's motives, whereas the former does not. His method throughout is to appeal to our ordinary reflective consciousness, which for Ross is the final arbiter on all ethical issues. He is confident that reflection on attempts to define 'right' will reveal that it is an irreducible notion. It cannot be analysed, but it can be elucidated: Ross suggests that rightness is a species of the relational property of fittingness or suitability. Just as a telling word or phrase may be called for at a particular juncture in a poem, so a particular moral response may be called for by various features of a situation, and the right action will be the act which constitutes the most suitable response to all the factors in that situation, taken as a whole.

Which features of an action are relevant to its rightness? Ross offers a list of what he terms '*prima facie* duties', each of which may have a bearing on the rightness of an action. These are duties of fidelity (promise-keeping), reparation, gratitude, justice, beneficence, self-improvement and nonmaleficence (not injuring others). Ross thus agrees with the consequentialist that we have a duty to do as much good as possible (beneficence) but he maintains that this is merely one duty among several, and one or more of the others might outweigh the duty of beneficence on some particular occasion (see Help and beneficence).

Despite having passed into common philosophical parlance, the phrase '*prima facie* duty' is, as Ross is quick to point out, doubly misleading. First, this is not strictly a list of duties, of things I ought to do, since, for example, it may precisely be my duty not to act in the way which will do the most good if I have promised to do something else. In here allowing considerations of fidelity to override those of beneficence I am not failing to carry out any sort of duty. Second, the term '*prima facie*' might suggest that we are speaking only of an appearance which a moral situation presents at first sight and which might turn out to be illusory. But even where some morally significant consideration is overridden by others, it does not cease to be relevant. (For this reason, some writers now prefer to use the term '*pro tanto*' duty.) It is perhaps best to think of Ross' list of *prima facie* duties as a list of fundamental moral considerations, or right-making characteristics, which always count in favour of an action (see Duty §2).

Ross' list is not an arbitrary one. He offers it as a list of the most basic morally relevant features. He does not claim that it is final - he is perfectly prepared to consider the possibility, for example, that one of the general features on the list might turn out not to be basic, or to be genuinely distinct from the others - but he does hold that what should be on the list is a subject for reasoned debate, drawing again on careful reflection on our moral beliefs.

That certain features of actions are right-making features is self-evident, according to Ross, in much the way that a mathematical axiom or the validity of a form of inference is self-evident. It is not that they are immediately obvious, but that, since they are basic, we cannot appeal to anything more fundamental to justify them. What we ought to do in any particular case is, by contrast, something of which we can never be certain, according to Ross, but only have a better or worse founded opinion. This pessimism about the possibility of moral knowledge in the particular case is perhaps over-stated, but it rests on a structural difference between knowledge of a general moral principle and appreciation of the particular case. Ross holds that, in all but the most trivial cases, whether an action is right will depend on the way that conflicting moral considerations bear on the particular case. There is no mechanical method, no algorithm, for calculating which of these considerations is the weightiest in some specific case. In most cases, there will be just one action that is the right one, but deciding which that is calls for judgment and practical wisdom (see Moral judgment §4).

See also: Axiology; Intuitionism in ethics; Values

DAVID McNAUGHTON

List of works

Ross, W.D. (1923) *Aristotle*, London: Methuen.(A general introduction to Aristotle.)

Ross, W.D. (ed.) (1924) *Aristotle: Metaphysics*, Oxford: Clarendon Press, 2 vols.(Contains Greek text and introduction by Ross.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Clarendon Press.(The finest modern systematic exposition of a moderate deontology.)

Ross, W.D. (ed.) (1936) *Aristotle: Physics*, Oxford: Clarendon Press.(Contains Greek text and introduction by Ross.)

Ross, W.D. (1939) *The Foundations of Ethics*, Oxford: Clarendon Press.(Ross' further thoughts on these matters, being the Gifford Lectures 1935-6.)

Ross, W.D. (1951) *Plato's Theory of Ideas*, Oxford: Clarendon Press.(A discussion of Platonic metaphysics.)

Ross, W.D. (1954) *Kant's Ethical Theory*, Oxford: Clarendon Press.(A commentary on Kant's *Groundwork of the Metaphysics of Morals* (1785).)

Ross, W.D. (ed.) (1955) *Aristotle: Parva Naturalia*, Oxford: Clarendon Press.(Contains Greek text and introduction by Ross.)

Ross, W.D. (ed.) (1957) *Aristotle: Prior and Posterior Analytics*, Oxford: Clarendon Press.(Contains Greek text and introduction by Ross.)

Ross, W.D. (ed.) (1961) *Aristotle: De Anima*, Oxford: Clarendon Press.(Contains Greek text and introduction by Ross.)

References and further reading

Audi, R. (1996) 'Intuitionism, Pluralism, and the Foundations of Ethics', in W. Sinnott-Armstrong and M. Timmons (eds) *Moral Knowledge?: New Readings in Moral Epistemology*, Oxford: Oxford University Press, 101-36.(A full-scale examination and defence of an intuitionist epistemology.)

Dancy, J. (1993) 'An Ethic of *Prima Facie* Duties', in P. Singer (ed.) *A Companion to Ethics*, Oxford: Blackwell, 230-40.(Offers a number of serious challenges to Ross' account of a *prima facie* duty.)

Gaut, B. (1993) 'Moral Pluralism', *Philosophical Papers* 22: 17-40.(A forceful defence of moral pluralism.)

McNaughton, D. (1996) 'An Unconnected Heap of Duties?', *Philosophical Quarterly* 46: 433-47.(Argues that deontology, in the form advocated by Ross, can offer a systematic account of our distinct duties.)

Moore, G.E. (1903) *Principia Ethica*, ed. T. Baldwin, Cambridge: Cambridge University Press, revised edn, 1993. (Moore here makes consequentialism true by definition by defining the right action as the action which produces the most good. This claim is vigorously rebutted by Ross.)

Moore, G.E. (1912) *Ethics*, London: Oxford University Press.(By the time he wrote this book, Moore had abandoned the view that the right could be *defined* in terms of the good, but he still held that what made an

action right was solely that it produced the greatest amount of good. Ross also vigorously rebutted this claim.)

Urmson, J.O. (1975) 'A Defence of Intuitionism', *Proceedings of the Aristotelian Society* 75: 111-19.(A lively defence of moral pluralism of a Rossian kind.)

Rousseau, Jean-Jacques (1712-78)

Rousseau was born in Geneva, the second son of Isaac Rousseau, watchmaker. His mother died a few days after his birth. From this obscure beginning he rose to become one of the best known intellectual figures of the eighteenth-century French Enlightenment, taking his place alongside Diderot, Voltaire and others as one of the emblematic figures of this period, for all that he came to differ violently in view from them. He died in 1778 and in 1794 his body was transferred to the Panthéon *in Paris.*

Rousseau always maintained that he regretted taking up a career of letters. His first love was music and he composed a number of operas in the 1740s with some success. The turning point in his life occurred in July 1749. He was on his way to see his then friend Diderot who was imprisoned at Vincennes. He read in the newspaper a prize essay question, asking whether advances in the sciences and arts had improved morals. So overcome was he by the flood of ideas that this question aroused in him the realization that he had to break his journey. The rest of his life's work was, he claimed, determined for him at that moment. Rousseau's primary claim to fame depends on his ideas about morals, politics and society. Perhaps his best-known remark is 'Man is born free; and everywhere he is in chains'; this reveals his preoccupation with issues of freedom in the state.

In answer to the prize essay question Rousseau argued that men and morals were corrupted and debilitated by advances in higher learning. The goal of prestigious distinction is substituted for that of doing useful work for the good of all. This theme, of people seeking invidious ascendancy by doing others down - the effect of exacerbated amour-propre *- pervades Rousseau's social theorizing generally. His essay,* Discourse on the Sciences and Arts *(1750), won the prize; related concerns shape the more profound* Discourse on the Origin of Inequality *of 1755. In his most famous work of political theory,* The Social Contract *(1762), Rousseau presents an alternative approach to how we might achieve a just and legitimate civil order. All members of society should take an equal place as members of the sovereign authority and societal laws should come from the general will by which a people gives rules to itself. Only under such a system, Rousseau argues, will humankind live on equal terms bound by fraternal ties, enjoying as much freedom and rights of self-determination as is possible in a stable community. Speaking up in this way for the equal political standing of all, regardless of birth or wealth, Rousseau points the way towards the dissolution of the* ancien régime *and the emergence of more democratically based polities. Precisely what influence his ideas had on the French Revolution is impossible to determine, although his name was often invoked.*

Rousseau also wrote extensively on education. In his Émile *(subtitled* On Education*, 1762) he tries to show how a child could be brought up free of the aggressive desire to dominate others. Instead that child can be caused to want to cooperate with others on a footing of mutual respect. He hopes by this to show that his social proposals are not an unrealizable dream. In this work there are also criticisms of religious dogma and church practices which brought severe condemnation onto Rousseau. He had to flee Paris in 1762 to avoid imprisonment. This, and other related experiences, plunged him into a protracted period of mental distress in which he feared he was the object of the plotting of others. These others came to include David Hume, with whom Rousseau had hoped to find refuge in England in 1766.*

Still troubled in mind, Rousseau returned to France the next year, and during the last decade of his life he wrote several works of self-explanation and self-justification. The greatest of these is his autobiography, Confessions *(written between 1764 and 1775, published posthumously), but there are other more prolix writings. After an accident in 1776, the worst of Rousseau's mental disturbance seems to have cleared and his last substantive work, an album of miscellaneous reflections on his life, ideas and experiences (*Reveries of the Solitary Walker*, written 1776-8), has a clarity and balance which had been absent for so long.*

1 Life and writings

Brought up by his father for the first ten years of his life after the death of his mother, Rousseau traced his love of republican Rome to the reading of Plutarch that he and his father used to do. This love, along with the idolization of his native Geneva, provided the inspiration for many of his political ideas. After being involved in a fight, Rousseau's father fled Geneva in 1722 and Rousseau was sent to live with his cousin not far from Geneva for a couple of years. This period in his youth is exquisitely evoked by Rousseau in Book One of *Confessions* (1764-75). When he returned to Geneva his more lowly social station became apparent and he was indentured to

an engraver, Abel Ducommun, a brutal and ill-educated man.

Restless and dissatisfied, Rousseau was more than glad to take advantage of the mischance of being locked out of the city on a Sunday in 1728. He walked away from that life, seeking the help of a Catholic priest who sent him to see Françoise-Louise de la Tour, Baronne de Warens, who was in receipt of money to secure more Catholic converts. She sent him, in turn, to Turin for instruction and Rousseau was admitted to the Church in April 1728. It is doubtful that Rousseau had any deep spiritual involvement in this process; he was more anxious to retain others' interest in him. He had a number of shortlived jobs in Turin. In one of these, he lied about stealing a ribbon and put the blame on a servant girl. This wicked deed preyed on his mind for the rest of his life.

The next year, he made his way back to Madame de Warens. He learned the rudiments of music and his passion for music was a dominant force in his life at this time. By the Autumn of 1731 he had moved in permanently with Madame de Warens. They lived a life of innocent delight for some years, she calling him *petit* and he calling her *maman*. He became her lover in 1733, although he appears never to have enjoyed this almost incestuous relationship. He read avidly during this time, laying a foundation for many of his later writings.

This idyll did not endure, however. Rousseau was displaced in Madame de Warens's affections in 1738. Considerably aggrieved, he took up the post of tutor to the two sons of Jean Bonnot de Mably in Lyons in 1740. Not an adept teacher, he gave up the post after a year determined to make his way in the larger world of Paris where he moved in 1742 (two short essays on education date from this time).

Once there, he presented a paper on musical notation to the Academy of Sciences; this was published in 1743 as *Dissertation on Modern Music*. In that year, Rousseau went to Venice as secretary to the French Ambassador. They quarrelled and Rousseau returned to Paris to resume his musical compositions. About this time he set up home with his mistress, Thérèse Levasseur, who was to be his lifelong companion. He had a number of children by her, whom he abandoned to his later shame. Rousseau had also begun to keep the company of the rising Parisian intelligensia. Diderot was a personal friend and it was while on the way to visit him during one of his periodic bouts of imprisonment that Rousseau had the experience that fixed the course of the rest of his life. The Academy at Dijon had advertised a prize essay question asking whether the advancements in the sciences and arts had improved morals. Rousseau saw this and was so overwhelmed by a flood of insights evoked by it that (he said) he spent the rest of his life trying to put into words what he had seen in one hour. Rousseau, answering the question with a firm 'No', won the prize and his essay was published in 1750 under the title *Discourse on the Sciences and Arts*. He was poised to begin a new career as social critic, moralist and philosopher, but his last triumphs as a composer and musical theoretician also occur about this time. His opera *Le Devin du Village* was performed before the King at Fontainebleau in 1752 and his *Letter on French Music* (1753) created an enormous stir as part of a large-scale argument over the relative merits of the French and Italian styles.

Rousseau was soon to turn his back on Parisian society. He wrote a further, very original, essay on social questions, *Discourse on the Origin of Inequality* (1755) but then withdrew to the countryside the better to meditate and write about his new concerns, attracting the scorn of many of his erstwhile friends. Around this time he returned to the protestant faith of his childhood, and reclaimed his citizenship of Geneva. The *Discourse on the Origin of Inequality* has a passionate dedication to Geneva.

During the next six years, Rousseau wrote the bulk of his greatest work: his masterpiece of educational theory, *Émile* (1762); of political theory, *The Social Contract* (1762); but also a best-selling novel, *La Nouvelle Héloïse* (1761) and a host of smaller pieces: the *Letter to M. d'Alembert on the Theatre* (1758); the *Letter to Voltaire on Providence* (1756), the *Moral Letters* (1757-8), written to Sophie d'Houdetot with whom Rousseau was then desperately in love.

Catastrophe befell Rousseau in 1762 after the publication of *Émile*. A section of it, the so-called *Creed of a Savoyard Vicar*, was judged unacceptable by the religious authorities and out of fear of being imprisoned Rousseau fled Paris in June 1762. Unsettled years followed, mostly spent in different parts of Switzerland. Rousseau wrote extensively in defence of himself and his work during this time, including his *Letter to Christophe de Beaumont, Archbishop of Paris* (1763) written in reply to the condemnation of the *Creed of a Savoyard Vicar*; and his *Lettres Écrites de la Montagne (Letters Written from the Mountain)* (1764), a response to criticism of him made by Geneva's attorney-general. From January 1766 Rousseau spent just over a year in England at the

invitation and in the company of David Hume. Rousseau, almost always a touchy and suspicious person, was at that time in the grip of a severe paranoiac breakdown and he became convinced Hume was plotting to humiliate him. An account of this sorry episode was given by Hume (*A Concise Account*, 1766). Exhausted and ill, Rousseau returned to France in early 1767 and sought refuge well away from the public gaze near Grenoble where he married Thérèse.

The tide of public opinion was slowly turning, and in 1770 he returned to Paris very much a celebrated figure and object of curiosity, even though he was banned from writing and speaking on controversial matters. Despite his grave mental distress, Rousseau had, from around 1764, been working on his great autobiography, *Confessions*. He completed part one by 1770. He gave some private readings of parts of the text; these also were banned. Other personal works occupied the bulk of the last decade of his life. There is an extensive essay in self-justification and defence, *Rousseau Judge of Jean-Jacques: Dialogues* (1772-6). Its completion was marked by another episode of desperate mental anguish as Rousseau attempted to place the manuscript on the altar at Nôtre Dame. Later in 1776, returning home from a walk, Rousseau was knocked down by a dog. This accident seems, miraculously, to have cleared his mind and his last work, *Reveries of the Solitary Walker* (1776-8) has a simplicity and clarity of manner missing from the writings of the preceding years.

Not all his work was in self-vindication however. He wrote at length on the political problems of Poland (*Considerations on the Government of Poland* (1769-70); he prepared a *Dictionary of Music* (1767) and he botanized extensively, also writing some short works on botany (*Elementary Letters on Botany and Dictionary of Botanical Terms*, uncompleted). He died at Ermenonville in June 1778, outlived by Thérèse for twenty-two years.

It is useful to give more information about Rousseau's life than is usual for most philosophers or political theorists, since so much of his work arises from events in his life or is directly about himself. This is, however, not so true of his principal works of social and political theory, just because they are works more purely of theory. They provide the most solid basis for Rousseau's reputation, and an account of these follows.

2 Works leading up to *The Social Contract*

From 1750 onwards, Rousseau developed increasingly deeper and more sophisticated ideas about the origin and nature of the condition of man in society and about what could and should be done to ameliorate that condition. His discussion of these themes in his first serious work, *Discourse on the Sciences and Arts*, is fairly shallow. He argues that increasing scientific knowledge and refinement of arts and letters does not at all produce an improvement of morals either in individuals or in society at large. On the contrary, such sophistication is the offshoot of luxury and idleness and it has developed principally to feed people's vanity and desire for ostentatious and aggressive self-display. All these features work against the moral virtues of loyalty to one's country, courage in its defence and dedication to useful callings. Rousseau allows for the fact that there are a few people of genius who genuinely enrich humanity by their ideas. But the majority of us are not improved, but harmed, by exposure to the 'higher learning'.

This essay attracted considerable notice and a number of replies, to which Rousseau responded with care. But he did not continue immediately with his works of social criticism. His musical interests intervened, although with some of these his social and moral ideas became entwined. In his *Letter on French Music* (1753), Rousseau criticizes French music as monotonous, thin and without colour because the spoken language (in which all music is rooted) is thus also. This is because, as Rousseau explains in his *Essay on the Origin of Languages* (1755-60, but never completed), the French language has been shaped by the imperatives of calling for help and controlling other people, which require harshness and clarity above all else. In warmer southern climes it is the sweet accents of love and passion which colour the language and hence the supremacy of Italian opera. Thus social and political demands shape even the nature of music, according to Rousseau. Effective government also requires sharp, impressive utterance, he maintains, and it is to the origin and function of government that Rousseau turns in his so-called second *Discourse*, the *Discourse on the Origin and Foundations of Inequality*, to give it its full title. This is a very substantial essay and one of Rousseau's most important works.

In *Discourse on the Origin and Foundations of Inequality* Rousseau gives an account of the 'fall' of natural humankind, its degeneration and corruption as it joins together with others to make up tribes, societies and eventually states. Natural man (the 'noble savage') left alone in his natural environment is self-sufficient, largely

absorbed in present feeling without foresight or recollection, solitary, peaceable and, in fact, most often asleep. (Rousseau may well have had the orang-utan in view here.) Inclement circumstances, increase in numbers (arising from hasty couplings in the forest devoid of all the artificial trappings of romantic love), force people to live together. Sexual jealousy, the desire for domination, vindictive resentments grow up as men come to demand esteem and deference. *Amour-propre*, an anxious concern for tribute to be paid to one's status, replaces *amour de soi*, a simple healthy concern for one's own natural wellbeing. Men begin to compete for precedence and life is tainted by aggression and spite. Those who have acquired dominance then conspire together to consolidate their position. They argue that everyone needs a more peaceable and stable society, which can only be achieved through the apparatus of government, law, punishments. Thus it is that they consolidate the *status quo*, but without right or justice and acting only to perpetuate unfair privilege and the oppression of the weak.

This extraordinarily subversive essay seems to have attracted no official censure; that came later in connection with other works. Rousseau's other significant essay on political themes from this period is the so-called *Discourse on Political Economy* (1755, first published separately 1758), which began life as an entry for Diderot and D'Alembert's *Encyclopaedia* (1751-72; see D'Alembert). A very eloquently written piece, it shows clear signs of being a preliminary study for *The Social Contract*. Much play is made with the idea of the sovereignty of all the people over themselves, expressing their legislative intent through the general will. Emphasis is laid on the need to cultivate patriotic republican loyalties in citizens if a just society of equals united by common care and respect is ever to arise and to survive. The essay ends with a discussion of taxation and fiscal issues, but the principal force of the argument lies in the discussion of the source of legitimate law in 'the people'. It is this same issue which Rousseau places at the centre of his now most famous work, *The Social Contract*.

3 *The Social Contract*

This work is generally regarded as an essential entry in the canon of classic works in political theory and as Rousseau's masterpiece. Many people read nothing else of his. This is a pity, for many of the themes in it are rendered unnecessarily hard to understand by being taken in isolation. Also, the work is in some ways poorly constructed and uses idea drawn from different times in Rousseau's development. As he says in a prefatory note to the work, it is the only residue remaining of a project begun many years before. However, we must take the work as we find it. Its present reputation would perhaps have surprised contemporary readers. *Émile* was considered a more seditious work; and *La Nouvelle Héloïse* regarded as the most perfect exhibition of Rousseau's genius. Certainly Hume regarded Rousseau's own good opinion of *The Social Contract* as quite absurd.

The Social Contract is divided into four parts. Roughly speaking, Book One concerns the proper basis for the foundation of a legitimate political order; Book Two the origin and functions of the sovereign body within that order; Book Three considers the role of government, which Rousseau treats as a subsidiary body in the state deriving its powers from the sovereign; and Book Four considers more issues regarding a just society, treating of the Roman republic at some length and of the functions of civil religion. It is important always to remember that the book is subtitled: *The Principles of Political Right*. Rousseau's paramount concerns are normative, with the nature and basis of legitimacy, justice and right and not simply with *de facto* political structures. A useful brief summary of the principal themes of the work is given in Book Five of *Émile*, as part of Émile's political education.

Rousseau argues that it is our lack of individual self-sufficiency that requires us to associate together into society. But, when we do so, we do not want to have to accept a condition of enslavement as the price of our survival. Freedom is an essential human need and the mark of humanity; mere survival without that does not constitute a human life. Rousseau holds that freedom and association can only be combined if all the persons of the association make up the sovereign body for that association, that is, the final authoritative body which declares the law by which the people wish to bind themselves. This law is a declaration of the 'general will'.

The notion of the general will is wholly central to Rousseau's theory of political legitimacy (see General will). It is, however, an unfortunately obscure and controversial notion. Some commentators see it as no more than the dictatorship of the poletariat or the tyranny of the urban poor (such as may perhaps be seen in the French Revolution). Such was not Rousseau's meaning. This is clear from the *Discourse on Political Economy* where Rousseau emphasizes that the general will exists to protect individuals against the mass, not to require them to be sacrificed to it. He is, of course, sharply aware that men have selfish and sectional interests which will lead them to try to oppress others. It is for this reason that loyalty to the good of all alike must be a supreme (although not

exclusive) commitment by everyone, not only if a truly general will is to be heeded, but also if it is to be formulated successfully in the first place.

This theme is taken up in Book Two. Here Rousseau appeals to the charisma of a quasi-divine legislator to inspire people to put the good of their whole community above their own narrow selfish interest and thereby gain a greater good for themselves. In the course of this Book, Rousseau alludes to Corsica as having a people who have the sentiments and capacities to establish just laws and a good state (Book Two, ch. 10). His passing remark that ‘I have the feeling that some day that little island will astonish Europe’ has caused some fancifully to suppose that he foresaw the emergence of Napoléon.

Book Three of *The Social Contract* concerns the role of government. Rousseau knows that governors often rule in their own interest, not in the interests of their community. For this reason he argues that governmental functions must be thoroughly subordinate to the sovereign judgment of the people and that it is essential to adjust the form and powers of government to suit the different circumstances (size, dispersion and so on) of different states. It still surprises some readers that Rousseau has no particular enthusiasm for democratic government. Of course, the constitution and functions of the sovereign body are a different matter.

Book Four has something of a disjointed character. Rousseau discusses the Roman republic at considerable length, principally to hold it up as a model from which, in his opinion, there has been a terrible falling away. But he also discusses civil religion, arguing that divine sanctions should be joined to civil laws the better to procure obedience to them and people’s loyalty to the common good of all in their nation.

Rousseau made wholly central to his vision of political right the union of free and equal men devising for themselves the laws under which they shall then proceed to live their lives as citizens one with another of their own state. In doing so he depicted a form of political community which exerts a very great appeal and influence on the modern imagination. We are still learning to live with the consequences of that appeal.

Rousseau’s political concerns were not confined to theory alone. On two occasions he was approached for help with the political affairs and constitutional problems of countries. In 1764 he wrote an unfinished fragment, A *Project for a Constitution for Corsica*, in response to a plea for help and guidance from the Corsican rebels. Then again, in 1769-70, Rousseau wrote extensively on the constitutional and legislative problems facing Poland (*Considerations on the Government of Poland*) in response to a request from persons opposed to Russian domination. This work (not properly published in Rousseau’s lifetime) is a substantial essay which throws a lot of light on how Rousseau envisaged his theoretical notions working out in historically specific situations. He reveals many shrewd and hard-headed practical insights.

4 *Émile* (or *On Education*)

There is some evidence that Rousseau regarded *Émile* as his most mature and well-achieved work. In his self-evaluating *Rousseau Judge of Jean-Jacques: Dialogues* (1772-6) he specifies it as the book in which someone who is truly concerned to understand him will find his ideas most deeply and comprehensively expressed. Posterity has, perhaps unfortunately, not generally endorsed this evaluation.

Precisely when he began work on *Émile* is unclear, but it must have been around 1759 when Rousseau was at the peak of his creative powers. Its immediate occasion seems to have been a request from certain of his distinguished women friends to give them his advice about the upbringing of their children; and indeed, the subtitle of the book is *On Education*. However, within this framework Rousseau gives us his deepest ideas about the origins of human evil and wickedness and about the prospects for a whole and happy life.

Émile is structured as the narrative of the upbringing of a young man (Émile himself) who is to be spared the pain and loss of human corruption but made whole and entire by following the teachings of ‘nature’. The work also includes in Book Four, a long more-or-less self-contained essay on the basis and nature of religious belief, called the *Creed of a Savoyard Vicar*. Rousseau puts his religious ideas into the mouth of a fictitious priest, although one modelled on priests he had previously known. It was this section which attracted the condemnation of the Catholic authorities and led to the burning of the book in June 1762. Rousseau then fled Paris, condemned to almost ten years of distress and displacement. He wrote at some length in defence of these religious ideas, in his *Letter to Christophe de Beaumont, Archbishop of Paris* of 1763 (see §5).

Rousseau held that most men and women in contemporary society were corrupted and their lives deformed because of the nature and basis of their social relationships and of the civil order (see §2). That man is good by nature, but is perverted by society is perhaps the dominant theme of *Émile*. Thus, if a person is to live a whole and rewarding life they must first be protected from such damaging influences and then be given the personal resources and emotional and moral dispositions to enable them to develop in a creative, harmonious and happy way once they do enter society. The principal discussions in *Émile* are devoted to studying the deepest causes of health or sickness in human development, both those internal to the individual and those coming from external influences, at each stage in the growth and maturation of a person from infancy to adulthood.

In early life it is the tendency to imperious rage and the petulant demand for others' immediate compliance to one's wishes that must be checked. Children must certainly not be tormented, but neither must they be indulged since that gives rise to both misplaced expectations and to even less capacity to cope with setbacks. Children need to be treated in a steady, predictable and methodical way, as if not in contact with other humans at all at first. Thus they learn to manage in a practical and efficacious way with concrete issues and not to engage in a battle of wills and in contention for dominance.

This motif of living according to nature - that is, according to the actualities of our powers and real circumstances - continues as Émile matures. As and when he needs to find a place for himself in society he will not try to control all that is around him and be aggrieved if he cannot as if he were a despot. Rather, he will seek to establish relations grounded in friendship, mutual respect and cooperation proper to finite and needy beings. Our capacity to feel compassion for each other and our acceptance of compassion with gratitude forms, in Rousseau's view, the fundamental basis for human union and the true explanation of the Golden Rule. Real moral demands are not imposed on us from outside, nor are they precepts discovered by reason. Rather, they express the requirements by which a bond of creative respect can be sustained between equals. This same issue of maintaining self-possession and mutual respect shapes Rousseau's treatment of marriage and sexual relations in Book Five. Such intimate union holds out the greatest hopes of human happiness, but can also lead to enslavement to the whims of the beloved. Feminist critics have found Rousseau's depiction of the character and role of Émile's intended, Sophie, objectionable, in that she appears to be stereotyped as largely passive and destined for traditional domestic occupations.

In the controversial material on religious belief Rousseau argues that we know God not by reason, but through simple feelings and convictions much deeper and more permanent than any theorems of reason. Such feelings teach us that the world is animated by a loving and powerful intelligence, who is God. Rousseau spends some time denouncing religious factionalism and intolerance which he sees as wholly incompatible with Christ's message of love and forgiveness. There can be no serious doubt that these are Rousseau's own thoughts. The rhetorical distance provided by the figure of the vicar is very slight.

In his deep and subtle psychological insights into the damage aggression does, not just simply to the victims of aggression, but in a complex and concealed way to the aggressors themselves, Rousseau shows the greatness of his mind in *Émile*.

5 Controversial works

Rousseau did not take the condemnation of the *Creed of a Savoyard Vicar* lightly. Almost as soon as he had settled again after his flight from Paris, he wrote a lengthy reply to the criticism of his work made by Christophe de Beaumont, Archbishop of Paris. Written in the form of a letter, Rousseau defends the fundamental tenets of his work. He has always held, he says, that man is naturally good, but corrupted by society. It is therefore a mystery why his work should only now be singled out for condemnation. He then mounts a point-by-point reply to the Archbishop's criticisms, arguing that the religion of the priests and the dogmas of the church must never be confused with the true gospel of charity and love taught by Christ. This *Letter to Christophe de Beaumont, Archbishop of Paris* was published in 1763. It ends on a note of self-aggrandizement. So far from being reviled, Rousseau says, statues of himself should be erected throughout Europe.

Just a year later, Rousseau published his extensive *Lettres Écrites de la Montagne (Letters Written from the Mountain)*. In 1763, the Genevan attorney-general, Jean-Robert Tronchin, had written in defence of the authorities in their condemnation of Rousseau's works in his *Letters Written from the Country* (hence Rousseau's oppositional

title). Rousseau again replies at length, arguing that the Geneval political system had become very corrupt, but also defending once more the basic principles of his thought.

These two works date from the period just after the condemnation of Rousseau's most famous works. But prior to that he had also been engaged in some controversial exchanges. In 1758 he wrote the *Letter to M. d'Alembert on the Theatre* in which he argues that to establish a theatre in Geneva would corrupt the honest morals and civil integrity of that city-state. This was in opposition to D'Alembert's argument, presented in an article on Geneva for the *Encyclopaedia*, that a theatre would improve the cultural life of that city. However, such sophistication is not a benefit, in Rousseau's opinion; it goes along with deceit and the abandonment of morally commendable activities. Rousseau writes with great verve in this essay, harking back to some of his themes in his first discourse, *Discourse on the Sciences and Arts*.

A similar clash between urbane civilization and (what Rousseau liked to see as his own) plainness and simplicity of heart occurs in the exchange with Voltaire on the providence of God (written in 1756; Voltaire may also have been an influence on D'Alembert). Voltaire, in his poem on the Lisbon earthquake, had written scornfully of Leibnizian optimism that all is for the best in the best of all possible worlds. Rousseau retorts that one must rest one's certainty of God's providence in feeling, not on the subtleties of philosophical reasoning. More personally, he says that it is surprising to find the wealthy and successful Voltaire complaining against God when he, Rousseau, who lives in poverty and obscurity, sees only the blessings of existence. This *Letter to Voltaire on Providence* also makes some very sharp points against religious intolerance, prefiguring the ideas of the *Creed of a Savoyard Vicar*. Rousseau's controversial writings are among his most eloquent, even though they do not generally add much to our appreciation of his overall intellectual achievement.

6 *La Nouvelle Héloïse* and other literary works

Outside the narrow circle of the intelligensia and his aristocratic patrons, Rousseau was probably best known during his lifetime for his novel of illicit passion, reconciliation and self-transcendence, *Julie, ou La Nouvelle Héloïse*. At a time when works in different genres were not so sharply compartmentalized, David Hume for one saw in this novel the most perfect expression of Rousseau's genius and could not understand why Rousseau seemed to value *The Social Contract* more.

Rousseau professed himself surprised and dismayed to be writing the work. In 1756 he had turned his back on the hot-house world of fashionable Parisian society, wishing to dedicate the rest of his life to working for the good of humankind. However, as he took solitary walks in the forest of Montmorency, be became absorbed in an imaginary world of passion and illicit love. In a fever of erotic ecstasy he wrote the first of the letters between Julie, her tutor-lover Saint-Preux and her friend and cousin Claire. Saint-Preux confesses his love for Julie; a love which she tries to fend off through intimate conversations instructing him to be virtuous and pure. Of course, this does not work and she finally gives herself over to him in a passion. But her father has other plans for her. She is betrothed to the Baron de Wolmar; Saint-Preux leaves and not until much later does he return to become tutor once more, but now to Julie's two young sons. Wolmar (who has come to know of their earlier intimacy), leaves Julie and Saint-Preux alone on his model estate at Clarens. Saint-Preux confesses that he had never ceased to love Julie, but the novel ends tragically with the death of Julie, who has contracted pneumonia after having saved one of her children from drowning.

As Rousseau wrote this work nature seemed to imitate art. Sophie d'Houdetot, the sister-in-law of Madame d'Épinay whose house Rousseau was then living in, visited him and Rousseau fell in love with her. He saw in her the incarnation of his imaginary Julie. The relationship did not endure. Rousseau became morbidly suspicious that his middle-aged love was being mocked behind his back by his erstwhile friends.

La Nouvelle Héloïse was published in 1761 and was a bestseller. It is seldom read these days, except to be mined for ideas which might illuminate Rousseau's social and political philosophy. Wolmar's model estate is sometimes argued to be Rousseau's own vision of an ideal community, with rigid paternalistic control and substantial manipulation of the inhabitants by the all-seeing, all-knowing Wolmar. It is scarcely clear that this was Rousseau's intention. The fact that Julie dies, despite living at Clarens, may be taken to imply that it provides no adequate human habitation.

The rest of Rousseau's literary output is slight. It includes a number of mostly short poems, dating from the early

1740s, some plays (also mostly early), one of which *Narcissus* (Self-Lover), received performance in 1752 and for which Rousseau wrote a substantial preface explaining how his theatrical writing could be squared with his then political and social polemic against civilized letters.

7 Autobiography and other personal works

The last ten or so years of Rousseau's life were primarily given over to the writing of works of autobiography and other substantial essays of self-explanation and justification. These were presaged in his four *Letters to Malesherbes* written in 1762, just before the catastrophe of the banning of *Émile* and *The Social Contract*. Malesherbes, although official censor and likely to be suspicious of Rousseau's subversive ideas, in fact took a highly intelligent and sympathetic interest in his work. Rousseau became fearful that the printing of *Émile* was being held up by Jesuit plotting. Through Madame de Luxembourg, Malesherbes was contacted and able to put Rousseau's mind to rest. Rousseau expressed his gratitude by writing to Malesherbes four semi-confessional letters, setting out the principal events of his life and trying to make his motives and character plain to Malesherbes. Rousseau writes that he is not a misanthrope; he seeks the country only because he can live there more freely and fully as himself. He wants, in fact, nothing more than to serve humankind, but he can best do this by keeping himself apart and not getting embroiled in quarrels and back-stabbing.

Around this time Rousseau began to assemble materials towards writing an autobiography. He worked on and off at this until 1767, by which time part one of what we now know as *Confessions* was completed. This still extraordinary work of self-disclosure and candour is one of the most remarkable books ever written. It includes some beautiful writing about childhood and about his travels, but also revelations of a most intimate and shameful kind. Part one covers the period up to 1741-2, when Rousseau left Madame de Warens to make his way in Paris. Part two (1769-70) is less successful. Rousseau's morbid fears sometimes surface here as he describes the events of the years 1742-65, including his foolish passion for Sophie d'Houdetot, the writing and publication of *Émile*, and so on. The book breaks off in 1765, just as Rousseau is about to leave for England in the company of Hume, there (he believed) further to be ensnared. As the revelation of the quality of being of another human soul, *Confessions* is almost without equal.

Another lengthy work of self-explanation is *Rousseau Judge of Jean-Jacques: Dialogues* (published, like *Confessions*, posthumously). This is in three parts, cast in the form of dialogues between 'Rousseau' and 'a Frenchman', who together try to delve beneath the surface to find out the true nature of 'Jean-Jacques', the real Rousseau. The innumerable lies put about regarding Jean-Jacques are considered and exposed in the first part; in the second, a visit is paid to him and his true character is revealed; in the third part, a careful reading of his works is made and their true meaning explained. Although sometimes obsessively detailed and very repetitive, this work has considerable interest for the light it throws on Rousseau's own estimate of his achievement. The overall tone is, it is thought, marred by lengthy self-justification.

Rousseau's last work of self-accounting is *Reveries of the Solitary Walker* (left unfinished at his death). Cast in the form of 'Walks', it comprises a series of reflections, ideas and meditations which supposedly occurred to Rousseau as he went on his perambulations in and around Paris. Rousseau returns to special moments in his life - his love for Mme de Warens; the episode of the stolen ribbon; his 'illumination' on the way to see Diderot. But he also reflects for one last time on some of his major intellectual preoccupations - on the depth of *amour-propre*; on the sources of malice and on the nature of happiness. There is a clear steadiness of vision which pervades this work which contrasts markedly with the often distressed and distressing writing of the preceding five years.

After his death, Rousseau's grave on the Île des Peupliers at Ermenonville became a place of pilgrimage for Parisians and Rousseau was embraced as one of the great sons of France. His influence remains very great, not only because of his political writings which have become part of the permanent canon of works in political theory, but also because of his more imponderable effect on sensibility and attitudes. His love of nature and stress on the value of the simple life, as well as his far-reaching explorations of his own character and feelings, make him a central figure in the development of romanticism. The emphasis in his educational writings on discouraging the coercion of the child into tasks which are apparently pointless, undoubtedly influenced the work of Montessori and A.S. Neill (see Education, history of philosophy of). Even Rousseau's musical writings and compositions, seldom studied these days, made a marked impact on the history of opera in particular. His place as one of the major figures of Western civilization is secure, even though he can still attract violent differences of opinion.

See also: Contractarianism; Enlightenment, continental; Language, early modern philosophy of; Political philosophy, history of; Society, concept of

NICHOLAS DENT

List of works

There is no complete standard English translation of Rousseau's works. The most ambitious undertaking (edited by Masters and Kelly) is still in progress. It is listed here with two French editions of Rousseau's complete works, followed by selected works in French and English and individual works in English.

1 Collected works

Cole, G.D.H. (ed. and trans.) (1973) *The Social Contract and Discourses*, London: Dent.(Revised and augmented by J.H. Brumfitt and J.C. Hall. Contains principal writings on political theory; an excellent selection in English.)

Gagnebin, B. and Raymond, M. (eds) (1959-95) *Jean-Jacques Rousseau: Oeuvres Complètes*, Paris: Gallimard, 5 vols.(The standard modern edition.)

Grimsley, R. (ed.) (1970) *Rousseau: Religious Writings*, Oxford: Clarendon Press. (Selected works in French.)

Launay, M. (ed.) (1967-71) *Rousseau: Oeuvres Complètes*, Paris: Éditions du Seuil, 3 vols.(Also valuable.)

Leigh, R.A. (ed.) (1965-89) *Correspondance Complète de Jean-Jacques Rousseau*, Geneva and Banbury: Voltaire Foundation, 50 vols.(One of the greatest scholarly achievements of this century and an indispensable source for Rousseau studies.)

Mason, J.H. (ed.) (1979) *The Indispensable Rousseau*, London: Quartet.(Offers selections in English from across the total range of Rousseau's writings.)

Masters, R.D. and Kelly, C. (eds) (1990-) *The Collected Writings of Rousseau*, Hanover, NH and London: University Press of New England, 5 vols to date.(Vol. 1: *Rousseau Judge of Jean-Jacques* (1990); vol. 2: *Discourse on the Sciences and Arts and Polemics* (1992); vol. 3: *Discourse on the Origin of Inequality, Polemics and Political Economy* (1992); vol. 4: *The Social Contract and short pieces* (1994); vol. 5: *Confessions* (1995) - the most up-to-date English translation, with excellent notes.)

Vaughan, C.E. (ed.) (1915) *The Political Writings of Jean-Jacques Rousseau*, Cambridge: Cambridge University Press, 2 vols.(Selected works in French.)

2 Individual works

Rousseau, J.-J. (1743) *Dissertation on Modern Music*, trans. L. Hewitt, in *Project Concerning New Symbols for Music*, Kilkenny, 1982.

Rousseau, J.-J. (1750) *Discourse on the Sciences and Arts*, in G.D.H. Cole (ed. and trans.) *The Social Contract and Discourses*, London: Dent, 1973; also in V. Gourevitch (ed.) *The First and Second Discourses together with the Replies to Critics*, New York: Harper & Row, 1986.

Rousseau, J.-J. (1753) *Letter on French Music*, trans. in part in *Source Readings in Music History*, O. Strunk (ed.) London: Faber & Faber, 1950.

Rousseau, J.-J. (1755) *Discourse on the Origin of Inequality*, in G.D.H. Cole (ed. and trans.) *The Social Contract and Discourses*, London: Dent, 1973; also in V. Gourevitch (ed.) *The First and Second Discourses together with the Replies to Critics*, New York: Harper & Row, 1986.

Rousseau, J.-J. (1755) *A Discourse on Political Economy*, in G.D.H. Cole (ed. and trans.) *The Social Contract and Discourses*, London: Dent, 1973; also in R.D. Masters (ed.) *On The Social Contract with Geneva Manuscript and Political Economy*, New York: Harper & Row, 1978.

Rousseau, J.-J. (1755-60) *Essay on the Origin of Languages*, in J.H. Moran and A. Gode (eds) *On the Origin of Languages*, New York: University of Chicago Press, 1966; also in V. Gourevitch (ed.) *The First and Second Discourses together with the Replies to Critics*, New York: Harper & Row, 1986.(Written around 1755-60, but not completed.)

Rousseau, J.-J. (1756) *Letter to Voltaire on Providence*, trans. in part in *The Indispensible Rousseau*, J.H. Mason (ed.) London: Quartet, 1979.

Rousseau, J.-J. (1757-8) *Moral Letters*.(No English translation currently available.)

Rousseau, J.-J. (1758) *Letter to M. d'Alembert on the Theatre*, in A. Bloom (ed. and trans.) *Politics and the Arts: Rousseau's Letter to M. d'Alembert*, Ithaca, NY: Cornell University Press, 1960.

Rousseau, J.-J. (1761) *La Nouvelle Héloïse*, abridged and trans. J.H. McDowell, Philadelphia, PA: Pennsylvania State University Press, 1968; also trans. W. Kenrick, *Eloisa, or a Series of Original Letters*, Oxford:

Woodstock Books, 1989.(Facsimile reprint of an 1803 translation).

Rousseau, J.-J. (1762) *Letters to Malesherbes*, in C.W. Hendel (ed. and trans.) *Citizen of Geneva*, Oxford: Oxford University Press, 1937.

Rousseau, J.-J. (1762) *Émile*, or *On Education*, trans. A. Bloom, New York: Basic Books, 1979; repr. London: Penguin, 1991.

Rousseau, J.-J. (1762) *The Social Contract*, in G.D.H. Cole (ed. and trans.) *The Social Contract and Discourses*, London: Dent, 1973; also in R.D. Masters (ed.) *On the Social Contract with Geneva Manuscript and Political Economy*, New York: St Martin's Press, 1978; also in M. Cranston (ed.) *The Social Contract*, Harmondsworth: Penguin, 1968.

Rousseau, J.-J. (1763) *Lettre à Christophe de Beaumont (Letter to Christophe de Beaumont)*, in J.H. Mason (ed.) *The Indispensable Rousseau*, London: Quartet, 1979.(Parts only; this work does not exist in whole in English translation.)

Rousseau, J.-J. (1764) *Lettres Écrites de la Montagne (Letters Written from the Mountain)*, in J.H. Mason (ed.) *The Indispensable Rousseau*, London: Quartet, 1979.(Parts only; this work does not exist in whole in English translation.)

Rousseau, J.-J. (1764-5) *Project for a Constitution for Corsica*, in F. Watkins (ed.) *Rousseau: Political Writings*, Edinburgh: Thomas Nelson, 1953.

Rousseau, J.-J. (1764-75) *Confessions*, trans. J.M. Cohen, Harmondsworth: Penguin, 1954.

Rousseau, J.-J. (1769-70) *Considerations on the Government of Poland*, in F. Watkins (ed.) *Rousseau: Political Writings*, Edinburgh: Thomas Nelson, 1953; also trans. W. Kendall, *Considerations: The Government of Poland*, Indianapolis, IN: Hackett, 1985.

Rousseau, J.-J. (1767) *Dictionary of Music*.(Published posthumously; no English translation currently available.)

Rousseau, J.-J. (1771-3) *Elementary Letters on Botany and Dictionary of Botanical Terms*, in *Botany: A Study of Pure Curiosity*, trans. P.J. Redouté, London: Michael Joseph, 1979.(The *Dictionary* is unfinished.)

Rousseau, J.-J. (1772-6) *Rousseau Judge of Jean-Jacques: Dialogues*, trans. J.R. Bush, C. Kelly and R.D. Masters, ed. R.D. Masters and C. Kelly, Hanover, NH and London: University Press of New England, 1990.

Rousseau, J.-J. (1776-8) *Reveries of the Solitary Walker*, trans. P. France, Harmondsworth: Penguin, 1979; trans. C.E. Butterworth, New York: Hackett, 1979.

References and further reading

Bloom, A. (1979) 'Introduction', *Émile*, or *On Education*, New York: Basic Books; repr. Harmondsworth: Penguin, 1991.(Revised edition of 'The education of democratic man', *Daedalus*, 107 (1978). An outstanding essay on *Émile*.)

Charvet, J. (1974) *The Social Problem in the Philosophy of Rousseau*, Cambridge: Cambridge University Press. (A lively analysis of Rousseau's social philosophy.)

Cranston, M. (1983-97) *Jean-Jacques*, vol. 1, London: Allen Lane, 1983; *The Noble Savage*, vol. 2, London: Allen Lane, 1991; *The Solitary Self*, vol. 3, London: Viking and Penguin, 1997.(A beautifully written and very insightful work.)

Crocker, L.G. (1968) *Jean-Jacques Rousseau: The Quest (1712-58)*; *The Prophetic Voice (1758-78)*, New York: Macmillan; 2nd edn 1973.(A very thorough treatment by a fairly hostile critic.)

Dent, N.J.H. (1988) *Rousseau*, Oxford: Blackwell.(Covers most of the psychological and political works.)

Gay, P. (1964) 'Reading about Rousseau', in *The Party of Humanity: Studies in the French Enlightenment*, London: Weidenfeld & Nicolson.(An exceptional essay on the reception and assessment of Rousseau's work.)

Grimsley, R. (1973) *The Philosophy of Rousseau*, Oxford: Oxford University Press.(The best short introduction to Rousseau's work overall.)

Guehenno, J. (1966) *Jean-Jacques Rousseau*, London: Routledge, 2 vols.(An engaging biography, making good use of the correspondence.)

Lovejoy, A.O. (1948) 'Rousseau's supposed primitivism', in *Essays on the History of Ideas*, Baltimore, MD: John Hopkins Press, 1948.(A remarkable study of the *Discourse on the Origin of Inequality*.)

Masters, R.D. (1968) *The Political Philosophy of Jean-Jacques Rousseau*, Princeton: NJ, Princeton University Press.(A ground-breaking study of Rousseau's political ideas.)

Plamenatz, J. (1992) *Man and Society*, vol. 2, London: Longman.(A judicious and elegant short study of the political ideas.)

Shklar, J. (1969) *Men and Citizens*, Cambridge: Cambridge University Press.(A subtle and insightful study of the psychology and politics.)

Starobinski, J. (1971) *Jean-Jacques Rousseau: La Transparence et l'Obstacle*, trans. A. Goldhammer,

Transparency and Obstruction, Chicago, IL: University of Chicago Press, 1988.(A very influential study by one of France’s foremost critics.)

Wokler, R. (1987) *Rousseau on Society, Politics, Music and Language*, New York: Garland.(An exemplary scholarly work, particularly on the *Essay on the Origin of Languages*.)

Royce, Josiah (1855-1916)

Josiah Royce rose from a humble background in the California of the Gold Rush period to become Professor of the History of Philosophy at Harvard University and one of the most influential American philosophers of the so-called 'period of classical American philosophy' from the late nineteenth to the early twentieth century. He was also (along with F.H. Bradley) one of the two most important English-speaking philosophers of the period who defended philosophical idealism: the doctrine that in some sense or other all things either are minds or else are the contents of minds. Royce remained loyal to his own idealist commitments throughout his life, despite the fact that his friend and Harvard colleague William James was extremely hostile to idealism, and that his intellectual environment was increasingly dominated by the 'pragmatism' of which James was an outspoken champion. In later years, however, under the influence of another pragmatist, Charles S. Peirce, Royce gave the themes of his idealist thought a naturalistic social foundation rather than the abstract metaphysical foundation of his earliest writings. Royce's entire corpus is perhaps best seen as representing a bridge from the German world of Neo-Kantianism and various varieties of philosophical idealism to the American world of pragmatism and of philosophical naturalism.

1 Life

Josiah Royce was born on 20 November 1855 in Grass Valley, California. In the spring of 1866 Royce's family moved to San Francisco, where he attended grammar school and high school. From 1870 to 1875 Royce was a student at the University of California. From 1875 to 1876 he studied in Germany, principally among thinkers who were revitalizing Kant's thought or defending and modifying Hegel's thought. From 1876 to 1878 Royce did graduate work at The Johns Hopkins University, where he gave a series of lectures entitled *On the Return to Kant*, in which he both expounded and defended Kant. Royce received the Ph.D. degree from Johns Hopkins in 1878.

From 1878 to 1882 Royce taught English at the University of California at Berkeley, and during this period he married. In 1882, taking his wife and their new child with him, Royce went to Harvard University to replace William James, who was on temporary leave. This appointment developed into an assistant professorship in 1885; eventually, in 1892, Royce obtained a full professorship in the history of philosophy. In 1914 Royce was appointed Alford Professor at Harvard. Except for brief excursions, Royce remained at Harvard from his initial appointment in 1882 until his death on 14 September 1916.

In American philosophy prior to the First World War, Royce was an even greater force than James. He had many distinguished students, among them George Santayana, W.E.B. Du Bois, T.S. Eliot, and C.I. Lewis.

2 Ethical theory

Besides his work in mathematical logic, Royce's philosophical thought concentrated on ethics and on metaphysics, which latter topic included philosophy of religion. In both ethics and metaphysics Royce's thought was heavily influenced by Kant, much of whose philosophy Royce absorbed from Neo-Kantians during his year of study in Germany.

In his ethical theory, espoused for example in his book *The Philosophy of Loyalty* (1908a), Royce's highest ethical principle appealed to the idea of acting from 'loyalty to loyalty', a formula in which Royce captured an ethical notion closely akin to Kant's notion of acting 'from duty'. Just as Kant held that acting from duty meant acting in such a way that one's motive for acting is to be in accord with the ideal of duty itself, so Royce held that acting out of loyalty to loyalty meant acting in such a way that one's motive for acting is to be consistent with the ideal of loyalty in general (see Kantian ethics). With the beginning of the First World War the ethical dimensions of Royce's thinking took on a pronounced social and geopolitical air. Royce even wrote a treatise in 1914 on geopolitical morality, published as *War and Insurance*. Shocked by the German sinking of the *Lusitania* in 1915, Royce became a vehement supporter of the Allies, and 'loyalty' acquired in his usage a certain patriotic flavour. In general, however, the meaning of 'loyalty' in Royce's ethical thought is akin to that of 'commitment', and not to any especially patriotic or chauvinistic notion.

In his ethical thinking during the last four or five years before his death, Royce was going beyond a mere ethics of duty and freedom and was pressing towards a more general ethics of responsibility and the 'fitting'. The very last

of Royce's ethical thoughts were devoted not so much to an attempt to specify the nature of right and wrong as to an effort to describe and justify the proper approach to life and life's challenges that a wise person takes.

In this regard Royce distinguished three stages of ethical maturity. In the first and least mature stage, the moral agent acts more or less naively, aiming their efforts at achieving goals. As a moral being, the agent attempts to be loyal to loyalty, and so to fulfil their duties and in general live a responsible life. But the agent has not yet confronted the difficulties life has to offer, its pains, frustrations, anxious cares, crushing burdens, demoralizing defeats. With such features of life the ethical agent still must learn to contend.

At the second, somewhat more mature, ethical stage the agent responds to the pains of life by achieving a kind of quiet resignation, turning inwards onto the self, that is to say towards meditation, self-control and self-alteration. But in this resignation to suffering, the agent ceases to attempt to alter the external world, or indeed anything in their relation to the external world other than the self. Indeed, this approach is broadly the stance of the mystic, the introvert, the meditator. Royce associates the second stage of ethical existence with some of the disciplines of oriental philosophy and religion.

At the third and highest stage of ethical maturity, the agent has transcended both the naive stance of the first stage and the quietistic resignation of the second. Once again the agent is an active agent attempting to change the external world and not merely the self. But now the agent is fully aware of life's sufferings and of the fact that very probably most of their efforts, perhaps even all of them, will be in vain. Nevertheless, the agent acts, not expecting that the efforts will be rewarded with success. At this third stage, the object of acting is moral action in itself.

Royce's late ethics has a great deal in common with certain strains of twentieth-century thought in what has been called 'existentialism'. In particular it has a great deal in common with that variety of existentialism labelled 'absurdism' by the French philosopher Albert Camus. For Royce, just as for Camus, the mature life of bliss is to be found not in achievements but rather in strivings, not in what is accomplished but rather in the manner and quality of what is attempted (see Existentialism). Only in the ethical life of striving while remaining free of all expectations regarding results is an agent immune from the frustrations of the first ethical stage and from the introspective isolation of the second.

3 Metaphysics

Royce's position in metaphysics was his own particular brand of idealism, which, despite its unique features, naturally had affinities with other forms of idealism. It had, for example, affiliations with the so-called 'subjective idealism' of Berkeley, whose principal doctrine was that all things are either minds or the contents of minds. Royce's idealism also in many key respects resembled the so-called 'transcendental idealism' of Kant, whose principal doctrine is that the objects of experience have no intelligibly conceivable existence outside of experience. In other respects Royce's position resembled the so-called 'metaphysical idealism' of Hegel, whose principal doctrine is that the world as we understand it is the evolutionary unfolding of a mental principle that we can call 'Spirit' (see Berkeley, G. §3; Hegel, G.W.F. §4; Idealism; Kant, I. §5).

The chief feature of Royce's early idealism that set it apart from other varieties of idealism was his doctrine of the existence of what he called a 'Universal Consciousness' or 'Absolute Mind'. He conceived this as a kind of overarching, all-embracing mind or consciousness in which all particular thoughts, as well as all the objects of these thoughts, along with all finite minds, are in some fashion 'contained'. This doctrine of the 'Universal Consciousness' was the burden of Royce's first major book *The Religious Aspect of Philosophy* of 1885, the year of his appointment at Harvard to an assistant professorship. In his doctrine of the Universal Consciousness, Royce aimed to capture in philosophical terms a sense of the radically immediate presence of the Divine Being. The spiritual thrust of Royce's philosophizing remained a steady feature of it throughout his career.

Despite the religious import of the doctrine of the Universal Consciousness, however, Royce tried in his work of 1885 to base this doctrine entirely on a semantical and logical analysis of the concepts of knowledge and truth. He began with two initially plausible assumptions. The first was that knowledge and truth presuppose reference. The second was that reference is to be analysed in such a way that objects of reference are completely determined purely by the referential intentions of referring minds. We may call this latter assumption the assumption of pure referential intentionalism.

Pure referential intentionalism has the consequence that all objects of reference are internal to minds. In particular, objects of reference are internal to the minds that achieve reference to such objects simply by means of having the referential intentions that they do. This consequence obviously involves an idealist commitment of some sort. But in order to see the particular form of idealism that Royce elicited from it, let us consider that the consequence raises a number of problems. How, for example, given this consequence, can two different minds refer to (and thus know) one and the same thing? Indeed, how even can one mind refer to (and thus know) one and the same thing on two different occasions? How can anyone ever make an error in judgment (since even in making a mistake in judging one must, it would seem, surely know that to which one refers)? Royce argued that the doctrine of the Universal Consciousness was required in order to solve these and other such problems, in particular that of the possibility of error.

With only minor changes this doctrine of the Universal Consciousness persisted in Royce's writings through his turn-of-the-century Gifford Lectures that were published as the two-volume work *The World and the Individual* (1899-1901).

The theory of knowledge and truth defended by Royce was soon to be attacked by William James in his works on pragmatism. Indeed, James's own pragmatic theory of truth was a direct response to the Roycean idealist challenge (see James, W. §5). James's main task was that of accounting for reference and truth without committing himself to an idealism that he considered absurd. Thus his task was to account for reference and truth on some basis other than pure referential intentionalism. Ultimately his pragmatism took a position according to which reference is determined not purely by referential intentions but also and more importantly by unbroken chains of causal connection stretching from referential intentions to objects of reference. Such objects, then, need not have a purely mental status but may consistently be regarded as being outside minds and in the external world. Not surprisingly, Royce and James enjoyed almost daily debates with one another on the topics of reference and truth until James's death in 1910. One of the most important of the historical consequences of Royce's thought is this influence on the development of James's version of pragmatism.

Not only was Royce an original philosopher - in particular an original idealist philosopher - in his own right: he was also a serious student of the history of philosophy. Indeed, although the bulk of his work was devoted to his own philosophical creation, still great portions of it were almost entirely devoted to the history of philosophy, for example almost all of *The Spirit of Modern Philosophy* and *Lectures on Modern Idealism*. His interest in the history of philosophy was simply a part of his interest in history more generally, and he devoted several years to writing on the history of his native state, California. Royce's interest in the history of thought led him to become accomplished in a wide variety of subjects, among them even such specialized subjects as the Sanskrit language.

4 Mathematical logic

A little-known side of Royce's intellectual life is his deep interest in mathematical logic and his extensive work in this area. Indeed, Royce introduced the topic of mathematical logic into the curriculum at Harvard, and by means of his own efforts and later those of his student C.I. Lewis, logic remained one of the mainstays of the Harvard philosophical experience. Royce's interest in logic was present as early as 1881, when he wrote for the benefit of his students at Berkeley his *Primer of Logical Analysis*. At this time he evidenced in his private diaries his concern to develop a logic that would be relevant to the theory of space. In this project Royce was aligning himself with the Kantian idea of a 'transcendental logic', that is to say a logic in which the structure of space and time plays a major role.

Royce seems to have left his logical concerns to lie dormant, however, until they were revived by Charles S. Peirce in 1898, when Royce attended Peirce's Cambridge Conference Lectures entitled *Reasoning and the Logic of Things* (see Peirce, C.S. §8). In these lectures Peirce showed that the forms of logic were in fact intimately connected with the forms of spatial structure; indeed, Peirce presented in the lectures a geometric form of logic that he had invented, called 'Existential Graphs'. Thus Peirce showed that Royce's earlier ideals with regard to logic were capable of detailed, technical fulfilment.

Peirce also fired Royce's logical and mathematical imagination in other ways. For example, he introduced Royce to the thought of the English mathematician A.B. Kempe, who had been a student of Augustus de Morgan, and who in the 1890s was an officer and leading figure in the London Mathematical Society. Kempe was considered

by Peirce to be the world's greatest logician. As early as 1890, Kempe had published a number of works in which he had attempted to reduce all mathematical form to certain special forms, and in which he had attempted to show deep similarities between the most basic structures of logic and the fundamental geometrical structures. Of particular interest to Royce was Kempe's paper 'The Relation between the Logical Theory of Classes and the Geometrical Theory of Points'.

From the time of Peirce's Cambridge Conference Lectures until at least as late as 1910, Royce worked assiduously on mathematical logic, indeed with a fierce enthusiasm. His goal was in accord with his earlier, ultimately Kantian ideals, namely to amalgamate mathematical logic with the theory of space. By the second volume of *The World and the Individual*, composed around 1900, Royce was already utilizing the fruits of his own original mathematical and logical researches, for example in describing space as a certain kind of logical order. By 1905 he had developed what he called his 'System Sigma'. This mathematical system was similar to Kempe's system for analogizing logic and geometry, but it attempted to remove various imperfections that Royce perceived in Kempe's theory, the most important of which concerned the notion of an individual.

Royce had a subsidiary and ancillary mathematical interest in what were called Boolean functions. In connection with this subject Royce studied A.N. Whitehead's researches into Boolean algebra. Royce was one of the first mathematical thinkers to discuss what are now called 'Boolean rings'. Another of his interests was the mathematical and logical paradoxes, including 'Russell's Paradox', and the foundations of set theory (see Russell, B.A.W. §§4-7). By 1905 Royce had proposed a set theory that would be 'non-foundational', that is to say, a set theory in which it is allowable to have infinitely descending (ordered by the relation of set membership) sequences of sets.

5 Ultimate position

During the last years of his life, Royce was beset by various health and family problems, but his philosophical production was undiminished. Indeed, *The Problem of Christianity*, which is widely regarded as his greatest work, was published only in 1913. In this work, under the influence of Peirce, Royce modified his earlier idealism into a powerful philosophical position, not unlike that of the later Wittgenstein in that its major themes are language and in particular the social functions of language. Central to the development in this work of Royce's own theory of the community is his employment of Peirce's ideas concerning the 'triadic' character of the notion of interpretation. For Royce the community is in essence a community of interpretation. It is an organic whole whose parts are individual persons, in which each of the parts has its place in a crisscrossing series of interpretations: each member interprets (things) to each other member. The resultant network of interpretation is the community. This idea is Royce's most mature expression of the idealistic thesis that was expressed in his earlier work by employing the metaphysical notion of the Absolute Consciousness. For the philosophical role played earlier by the Universal Consciousness comes to be assumed by the 'Community of Interpretation'. In this down-to-earth, as it were 'naturalized', version of idealism, the spiritual concerns that had driven Royce's earliest idealistic thought were unattenuated: indeed, as the title of his master work indicates, these concerns were made even more conspicuous in that work than ever before.

See also: Absolute, the; Hegelianism §5; Logic in the 19th century; Pragmatism

ROBERT W. BURCH

List of works

Royce, J. (1885) *The Religious Aspect of Philosophy*, Boston, MA: Houghton Mifflin. (Royce's first major work on ethics and metaphysics.)

Royce, J. (1886) *California from the Conquest in 1846 to the Second Vigilance Committee in San Francisco [1856]: A Study of the American Character*, Boston, MA: Houghton Mifflin.(Royce's history of California in the state's early days.)

Royce, J. (1887) *The Feud of Oakfield Creek: A Novel of California Life*, Boston, MA: Houghton Mifflin. (Royce's only novel.)

Royce, J. (1892) *The Spirit of Modern Philosophy: An Essay in the Form of Lectures*, Boston, MA: Houghton Mifflin.(An exposition of the major idealist themes presented in *The Religious Aspect of Philosophy*.)

Royce, J. (1897) *The Conception of God*, New York: Macmillan.(An exposition of Royce's idealism with a special emphasis on the nature of god and the finite individual.)

Royce, J. (1898) *Studies of Good and Evil*, New York: D. Appleton.(An analysis of fundamental ethical ideas.)
Royce, J. (1899-1901) *The World and the Individual*, New York: Macmillan, 2 vols.(One of Royce's greatest works: his two-volume set of Gifford Lectures.)
Royce, J. (1908a) *The Philosophy of Loyalty*, New York: Macmillan.(Royce's systematic presentation of his ethical theory based on loyalty.)
Royce, J. (1908b) *Race Questions, Provincialism, and Other American Problems*, New York: Macmillan. (Miscellaneous ethical and social-philosophical essays.)
Royce, J. (1911) *William James and Other Essays on the Philosophy of Life*, New York: Macmillan.(A collection of essays on various topics.)
Royce, J. (1912) *The Sources of Religious Insight*, New York: Charles Scribner's Sons. (Royce's treatment of religious epistemology.)
Royce, J. (1913) *The Problem of Christianity*, Chicago, IL: University of Chicago Press, 1968.(One of Royce's greatest works: his transformation of his early idealism into a theory of thought that is socially and communally based.)
Royce, J. (1914) *War and Insurance*, New York: Macmillan.(Royce's effort to promote world peace.)
Royce, J. (1916) *The Hope of the Great Community*, New York: Macmillan.(A further work on the topic of community.)
Royce, J. (1919) *Lectures on Modern Idealism*, ed. J. Loewenberg, New Haven, CT: Yale University Press.(Text of Royce's lectures of 1906.)
Royce, J. (1920) *Fugitive Essays*, ed. J. Loewenberg, Cambridge, MA: Harvard University Press.(A posthumously published set of miscellaneous essays.)
McDermott, J.J. (ed.) (1969) *The Basic Writings of Josiah Royce*, Chicago, IL: University of Chicago Press, 2 vols.(An extensive anthology of Royce's writings.)
Clendenning, J. (ed.) (1970) *The Letters of Josiah Royce*, Chicago, IL: University of Chicago Press.(An excellent edition of Royce's letters.)

References and further reading

Clendenning, J. (1985) *The Life and Thought of Josiah Royce*, Madison, WI: University of Wisconsin Press.(The definitive biography of Royce, very carefully and elegantly presented. It is a natural starting place for anyone interested in Royce's life, times and thinking.)
Kempe, A.B. (1889-90) 'The Relation between the Logical Theory of Classes and the Geometrical Theory of Points', *Proceedings of the London Mathematical Society* 21: 147-82.(An exposition of fundamental similarities among the topics of set theory, logic and geometry.)
Kuklick, B. (1985) *Josiah Royce: An Intellectual Biography*, Indianapolis, IN: Hackett Publishing Company.(A general account of Royce's life and thought, written by a historian rather than a philosopher.)
Oppenheim, F.M. (1993) *Royce's Mature Ethics*, Notre Dame, IN: University of Notre Dame Press.(An account of Royce's ultimate ethical position.)
Oppenheim, F.M. (1987) *Royce's Mature Philosophy of Religion*, Notre Dame, IN: University of Notre Dame Press.(An interpretation of Royce's *The Problem of Christianity*.)
Oppenheim, F.M. (1980) *Royce's Voyage Down Under: A Journey of the Mind*, Lexington, KY: The University Press of Kentucky.(An account of Royce's recovery from a nervous breakdown by means of travel to Australia.)
Peirce, C.S. (1992) *Reasoning and the Logic of Things*, ed. K.L. Ketner, Cambridge, MA: Harvard University Press.(This is a recent edition, indeed the only adequate edition, of Peirce's Cambridge Conference Lectures of 1898, the lectures that inspired Royce to return to mathematical logic with vigour. This edition includes an extremely valuable introduction by Kenneth Laine Ketner and Hilary Putnam.)

Rozanov, Vasilii Vasil’evich (1856-1919)

Vasilii Rozanov, a prominent spokesman of the Russian Religious-Philosophical Renaissance, is known for his writings on sex, marriage and the family, his attacks on Christian asceticism, and his love-hate attitude to Judaism. He termed Judaism a religion of life because it sanctified sex and the family, and Christianity a religion of death because it exalted celibacy. But Rozanov also charged that Judaism mandated ritual murder and that Jews were feeding off Russians. In Apokalypsis nashego vremeni (The Apocalypse of Our Time) (1917-18), he apologized for his anti-Semitic pronouncements and blamed the Bolshevik Revolution on Christian otherworldliness and sexlessness. In private life, Rozanov was a pillar of the Church, which he regarded as a haven of beauty, warmth, and spiritual succour.

Rozanov’s writings contain brilliant insights, contradictions, distortions and outright lies. ‘Falsehood never tormented me’, he wrote, ‘and for a strange reason. What business is it of yours precisely what I think?*’ He championed conservative policies in articles published under his own name and radical policies under a pseudonym. It is clear, however, that Rozanov was deeply religious and that he associated sex and the family with God. ’Dirty diapers and a naked wife: this is the truth of Bethlehem around you.’*

Rozanov’s interpretations of Dostoevskii, Gogol’ and other Russian authors were original and perceptive. His collections of aphorisms Solitaria (1912) and Opavshie list’ia (Fallen Leaves, 2 vols) (1913b, 1915), are considered masterpieces of Russian style.

1 Biography and early works

Rozanov was born in Vetluga, Kostroma province. The death of his father, a minor forestry official, when Rozanov was 5 plunged the family into poverty. His mother was too overworked to provide the love and attention that he craved. Rozanov was 13 when she died; his first reaction was that now he could smoke openly. He attended gymnasia in Kostroma, Simbirsk and Novgorod, and the Historical and Philological faculty at Moscow University, graduating from the latter in 1881. In his third year he married Apollinaria Suslova, Dostoevskii’s former mistress and twenty years Rozanov’s senior, in a bizarre attempt to be close to the great writer. Suslova left him in 1888, but cruelly refused consent to a divorce. In 1889, Rozanov met Varvara Rudneva, a young widow. Two years later they were married in an unofficial church ceremony which was conducted secretly lest he be prosecuted for bigamy. He had to adopt their five children formally.

Between 1880 and 1893 Rozanov taught history and geography in provincial gymnasia, hating every moment and writing in his spare time. His articles appeared in a variety of newspapers including the ultra-conservative *Novoe vremia (New Times)*. His first book, *O ponimanii (On Understanding)* (1886), an attack on positivism, was ignored by the critics and sold very few copies. But *Legenda o velikom inkvizitore F.M. Dostoevskogo (Dostoevsky and the Legend of the Grand Inquisitor)* (1891) attracted the attention of the influential literary critic Nikolai Strakhov (1828-96) who found Rozanov a minor bureaucratic job in St Petersburg, thereby enabling him to leave teaching. In 1899, Rozanov became a staff writer for *Novoe vremia*. His next books - *Literaturnye ocherki (Literary Sketches)* (1899), *Sumerki prosveshcheniia (The Twilight of Enlightenment)* (1899), *Religiia i kul’tura (Religion and Culture)* (1899) and *Priroda i istoriia (Nature and History)* (1900) - were collections of previously published articles, most of which had a conservative Slavophile bent. He was then seeking a Christian metaphysic of culture. Later on, he publicly endorsed the Holy Synod’s excommunication of Tolstoi, but he criticized its impersonal and bureaucratic handling of the matter. Rozanov would have preferred that Tolstoi be ’driven out of the Church’ by a band of outraged peasants whose religious sensibilities he had offended.

Around the turn of the century, Rozanov began to associate with Symbolists and Decadents even though he disapproved of their ‘unnatural’ writings. He published in *Mir iskusstva (The World of Art)*, had a regular column, ’In My Own Corner’, in *Novyi put’ (New Path)*, and was a prominent member of the Religious-Philosophical Society of St Petersburg. Instrumental in obtaining permission for the meetings, he was the only founding member who was a personal friend of some of the clerics.

In 1905-6, Rozanov wrote pro-revolutionary articles under a pseudonym (Varvarin) for *Novoe slovo (New Word)* in which he associated revolution with youth and described the autocracy as a ’weakening fetish’. He reprinted

these articles under his own name in *Kogda nachal'stvo ushlo (When the Authorities Went Away)* (1910). In *Fallen Leaves II*, he wrote, 'Politics must be destroyed, apoliticism established… by confusing all political ideas. …By making ''red'' ''yellow'' and ''white'' ''green'''. Laura Engelstein maintains that Rozanov was consciously trying to 'disorganize' public discourse by 'wreak[ing] havoc with the conventions of serious prose and responsible argument' (1992: 314).

Rozanov welcomed the First World War as the beginning of Russia's renaissance. He hailed the February Revolution and was going to 'compose an ideology' for it, but the Bolshevik Revolution was a disaster for him. When the Bolsheviks closed *Novoe vremia* he moved, with his family, to Sergeev Posad, near the Holy Trinity Monastery. He died of starvation and exhaustion, after having the priests administer the last rites nine times. At his request he was buried next to Konstantin Leont'ev (1831-91) in the monastery cemetery.

Leont'ev, Dostoevskii and Nietzsche were the major intellectual influences on Rozanov (see Leont'ev, K.N.; Dostoevskii, F.M.; Nietzsche: impact on Russian thought). He and Leont'ev corresponded in 1890-1, drawn together by mutual hostility to positivism, liberalism and democracy, and by concern for the future of Russia and of Christianity. But Leont'ev ridiculed the 'rosy-coloured' Christianity of Dostoevskii and Tolstoi and regarded Old Testament Judaism as a religion of fear. Rozanov pictured Jehovah as a father, a giver of life and a founder of families. Both writers were called the 'Russian Nietzsche' - Leont'ev because of his elitist aestheticism and Rozanov because of his diatribes against Christianity and his unabashed amoralism. 'Morality', Rozanov wrote, 'I don't even know how to spell it'. In sharp contrast to Nietzsche, Rozanov idealized domesticity and found supreme beauty in the face of Jesus and in church rituals. Trotsky called Rozanov the 'poet of the cosy corner'.

2 Sex, marriage and the family

Rozanov's first writings about sex, marriage and the family were motivated by his personal situation. *Semeinyi vopros v Rossii (The Family Question in Russia)* (1903), treated the legal and social status of the family, interspersing letters, newspaper reports and statistics with his own commentary. Rozanov considered the increase in illegitimacy, abortion, and prostitution symptoms of the decay of marriage, which he blamed on the Church. He advocated easy divorce as a way to revitalize the family and urged that the very concept of illegitimacy be abolished. An earlier work, *V mire neiasnogo i nereshennogo (In the World of the Obscure and Unresolved)* (1901), was banned as pornographic because it included a letter from a correspondent who claimed that circumcision had improved his sexual performance. In articles published in *Mir iskusstva*, Rozanov praised the sun-worshipping (life-affirming), phallic religions of the Ancient Near East.

Throughout his life, Rozanov proclaimed the sanctity of family life and the metaphysical and religious significance of the sexual act. He believed that during intercourse the couple's souls come into contact with other worlds and create a new soul, and that pregnancy was a sacred state. Contrasting the two poles of Christianity, Bethlehem (childbirth) and Golgotha (death), Rozanov maintained that by concentrating on Golgotha, the Church has turned Christianity into a religion of suffering, sorrow and death. He wanted Christians to resurrect the spirit of Bethlehem and for Christianity itself to become phallic, 'at least in part'. Independently of Freud, he extolled a good sex life as essential to physical and mental health.

3 Christianity and Judaism

Around 1905-6 Rozanov went through a religious crisis. Soon after, he began to attack Christianity *per se* as a religion more suited to funerals than to weddings, a religion that teaches man *'to live for death and to die for life'*. No Christian conception of the family exists, he argued. Monasticism was the Christian ideal - denial of life and love, rejection of the world and the family, renunciation of pleasure of any kind for the sake of heaven. One cannot even imagine any of the apostles or their young wives in love. Only love of Jesus is allowed. Through love of 'the sweetest Jesus', the fruits of this world tasted bitter. The Gospels were a religiously cold, even religiously indifferent book, in which the stamp of sadness is ever present. Jesus never laughed, never sang, never danced, and did not marry. Christian monks and nuns were akin to the castrated priests of the Phoenician gods Moloch and Astarte. 'And this hard crystal [monasticism] is indissoluble in Christian civilization. Moreover, it is slowly leading the Christian peoples…and Christianity itself to disunion, destruction and to leaving behind on earth [only] the "chosen few"'. (see Roberts 1978: 194).

Rozanov expressed these ideas in 'On Sweetest Jesus and the Bitter Fruits of the World' (1907), published in *Tëmnyi lik. Metafizika khristianstva (Dark Face: The Metaphysics of Christianity)* (1911a), *Liudi lunnogo sveta. Metafizika khristianstva II (People of the Moonlight: The Metaphysics of Christianity II)* (1911b) and *Apokalypsis nashego vremeni (The Apocalypse of Our Times)* (1917-18).

Rozanov termed ascetics 'lunar types' who live outside the family and do not procreate. Some ascetics had a weak or non-existent sex drive, but most were repressed homosexuals. As examples, Rozanov gave Jesus, Plato, Solov'ëv, monks and nuns, career women, and secular radicals from 'that half Urning' (homosexual), Chernyshevskii to the present. Rozanov associated atheism and socialism with 'seedlessness', non-procreation. Homosexuals and lesbians were a 'third sex'. They have contributed greatly to society and culture and should not be persecuted. But Rozanov also claimed that because they have no children, they are indifferent to the future, including their country's future.

Rozanov admired the sex-positive teachings of Judaism, the strong family life of contemporary Jews, and Jewish survival in the face of centuries of persecution. But he found Jewish energy menacing, and attributed it to their sex-positive religion and to circumcision. He portrayed certain Jewish rituals in eerie, other-worldly terms and described Jews as spiders feeding off Russian blood and as destroyers of Russian culture. During the Beiliss Case (1911-13) - the government's frame-up of a Jew, Mendel Beiliss, for ritual murder - Rozanov falsely claimed that ritual sacrifice was and remains the most essential aspect of Judaism. Denying that he was exacerbating anti-Semitism in a land rife with pogroms, Rozanov declared that he wanted to move the case from politics to a higher 'religious' plane. Some of his writings were so scurrilous that even *Novoe vremia* refused to print them. Because of them Rozanov was expelled from the Religious-Philosophical Society.

In *Apokalypsis nashego vremeni*, Rozanov's response to the Bolshevik Revolution, he wrote that the Gospels have led humanity to a 'blind alley of castration and death' and called *Revelations* an anti-Christian book, because it was based on birth ('The Woman Clothed in the Sun'). He claimed that the masses have abandoned Christianity because it is impotent. 'Despite all the beauty of Christianity', it cannot provide bread. Christianity's dissolution has left 'huge voids' into which 'all humanity is falling'. Rozanov also criticized Russian literature, especially nihilism, Gogol' and himself, for causing Russians to lose sight of reality. And he apologized to the Jews for defaming them, lauded their positive role in history and in Russian society, and condemned pogroms.

Rozanov's diatribes against Christianity and Christ were motivated as much by love as by hate. His writings also contain innumerable positive statements about them. For example: 'only in the Church is it warm'; 'of course I will *nevertheless die in the Church*, of course *I need the Church incomparably more than literature*'; and 'I need only consolation, and therefore I need only Christ… Paganism and Judaism do not even come to mind' (see Roberts 1978: 304-10; original emphases). His writings constitute an extreme example of the revaluation of Christian dogma that was a feature of the Religious-Philosophical Renaissance.

See also: Russian Religious-Philosophical Renaissance

BERNICE GLATZER ROSENTHAL

List of works

Rozanov's writings, consisting of over twenty books and 2,500 articles, have never been collected.

Rozanov, V. (1886) *O ponimanii (On Understanding)*, Moscow: E. Lissiev & Iu. Roman.

Rozanov, V. (1891) *Legenda o velikom inkvizitore F.M. Dostoevskogo* serialized in *Russkii vestnik*, Moscow: Nikolaev, 1894; repr. of St Petersburg 1906 edn, Munich: W. Fink, 1970; trans. S. Roberts, *Dostoevsky and the Legend of the Grand Inquisitor*, Ithaca, NY: Cornell University Press, 1972.

Rozanov, V. (1899a) *Literaturnye ocherki (Literary Sketches)*, St Petersburg: Merkushev.

Rozanov, V. (1899b) *Religiia i kul'tura (Religion and Culture)*, St Petersburg: Merkushev.

Rozanov, V. (1899c) *Sumerki prosveshcheniia. Sbornik statei po voprosam obrazovaniia (Twilight of Enlightenment. A collection of articles about education)*, St Petersburg: Merkushev.

Rozanov, V. (1900) *Priroda i istoriia (Nature and History)*, St Petersburg: Pertsov.

Rozanov, V. (1901) *V mire neiasnogo i nereshennogo (In the World of the Obscure and Unresolved)*, St Petersburg: Merkushev; 2nd edn, 1904; repr. Moscow: Respublika, 1995.

Rozanov, V. (1903) *Semeinyi vopros v Rossii (The Family Question in Russia)*, St Petersburg: Merkushev, 2 vols.

Rozanov, V. (1906) *Okolo tserkovnykh sten (Near the Church Walls)*, St Petersburg: Vaisberg i P. Gershunin, 2

vols.
Rozanov, V. (1910) *Kogda nachal'stvo ushlo (When the Authorities Went Away)*, St Petersburg: Suvorin.
Rozanov, V. (1911a) *Tëmnyi lik. Metafizika khristianstva (Dark Face: The Metaphysics of Christianity)*, St Petersburg: Vaisberg i Gershunin.
Rozanov, V. (1911b) *Liudi lunnogo sveta. Metafizika khristianstva II (People of the Moonlight: The Metaphysics of Christianity II)*, St Petersburg: Merkushev; 2nd edn 1913; repr. of 2nd edn, Wurzburg, 1970, with intro. by G. Ivask; repr. of 2nd edn Moscow: Druzhba narodov, 1990.
Rozanov, V. (1912) *Uedinënnoe*, St Petersburg: Suvorin; trans. S.S. Koteliansky, *Solitaria*, London: Wishart, 1927.
Rozanov, V. (1913a) *Lituraturnye izgnanniki (Literary Exiles)*, St Petersburg: Suvorin.
Rozanov, V. (1913b, 1915) *Opavshie list'ia (Korob pervyi, Korob vtoroi)*, St Petersburg: Suvorin, 2 vols; trans. of 1st vol., S.S. Koteliansky, *Fallen Leaves, Bundle One* London: Mandrake, 1929.
Rozanov, V. (1914a) *Sredi khudozhnikov (Among the Artists)*, St Petersburg: Suvorin.
Rozanov, V. (1914b) *Oboniatel'noe i osiazatel'noe otnoshenie evreev k krovi (The Attitude of the Jews towards the Smell and Touch of Blood)*, Petrograd: Suvorin; repr. Stockholm: n.p., 1932.
Rozanov, V. (1915) *Voina 1914 g. i russkoie vozrozhdenie (The War of 1914 and the Russian Renaissance)*, Petrograd: Suvorin.
Rozanov, V. (1917-18) *Apokalypsis nashego vremeni (The Apocalypse of Our Time)*, Sergiev Posad: Ivanov.
Rozanov, V. (1990) *V.V. Rozanov*, Moscow: Nauka, ed. E.V. Barabanov, 2 vols.(Volume 1 contains *Religiia i kul'tura*, *Russkaia tserkov'*, *L.N. Tolstoi i Russkaia tserkov'*, and *Tëmnyi lik. Metafizika khristianstva*. Volume 2 contains *Liudi lunnogo sveta*, *Uedinënnoe* and *Opavshie list'ia*, both bundles.)

References and further reading

Banerjee, M. (1971) 'Rozanov and Dostoevsky', Slavic and East European Journal 15: 411-24. (Focus on Rozanov's The Legend of the Grand Inquisitor.)
Crone, A.L. (1978) *Rozanov and the End of Literature*, Würzburg: Arnult Liebing.(Part I - Rozanov's aphorisms as a polyphony of eight voices; Part II - Rozanov on literature and his use of literary forms.)
Crone, A.L. (1986) 'Nietzschean, All Too Nietzschean? Rozanov's Anti-Christian Critique', in *Nietzsche in Russia*, ed. B.G. Rosenthal, Princeton, NJ: Princeton University Press, 95-112.(Argues that Nietzsche influenced Rozanov.)
Engelstein, L. (1992) 'Sex and the Anti-Semite: Vasilii Rozanov's Patriarchal Eroticism', in *The Keys to Happiness*, Ithaca, NY: Cornell University Press, 299-333.(Emphasizes gender issues.)
Fateev, V.A. (ed.) (1995) *V.V. Rozanov: Pro et contra*, St Petersburg: Izdatel'stvo russkogo Khristianskogo gumanitarnogo instituta, 2 vols.(Anthology of writings by Russian thinkers and scholars about Rozanov's life and work.)
Glouberman, E. (1976) 'Vasilii Rozanov: The Anti-Semitism of a Russian Judaephile', in Jewish Social Studies 38 (2): 117-44. (Argues that Rozanov's anti-Semitism is endemic to Christianity.)
Gollerbakh, E.V. (1922) *Rozanov: Lichnost' i tvorchestvo*, Petrograd: Poliarnaia zvezda.(Reminiscences and bibliography of works about Rozanov to 1922.)
Hutchings, S. (1993) 'Breaking the Circle of the Self: Domesticiation, Alienation, and the Question of Discourse in Rozanov's Late Writings', in Slavic Review 52 (1): 67-86.(Emphasizes genre issues.)
Kline, G.L. (1968) 'Religious Neo-Conservatives: Leontyev and Rozanov', 35-72.(Synopsis and comparision of their views of Christianity.)
Kornblatt, J.D. (1997) 'Russian Religious Thought and the Jewish Kabbala', in *The Occult in Modern Russian and Soviet Thought*, ed. B.G. Rosenthal, Ithaca, NY: Cornell University Press, 75-95. (Includes section on Rozanov's misuse of the Kabbala to 'prove' ritual murder in the Beiliss Case.)
Nikoliukin, A.N. (ed.) (1995) *V.V. Rozanov. O pisatel'stve i pisateliakh (V.V. Rosanov. On Writing and Writers)*, Moscow: Respublik. (Anthology of Rozanov's essays on the subject.)
Putnam, G. (1971) 'Vasilii Rozanov: Sex, Marriage, Christianity', in Canadian Slavic Studies 3: 301-26.(A pioneering study of these subjects.)
Roberts, S. (1978) *Four Faces of Rozanov: Christianity, Sex, Jews, and the Russian Revolution*, New York: Philosophical Library. (Introductory and concluding essays by the editor; translation of 'On Sweetest Jesus', excerpts from *Liudi lunnogo sveta* and *Apokalypsis nashego vremeni*.)
Romanoff, J. (1974) 'Vasilij Rozanov, the Jurodivij of Russian Literature', Ph.D. dissertation, Stanford University.(Detailed study of Rozanov's life and thought.)
Stammler, H. (1984) *Vasilij Vasil'evic Rozanov als Philosoph (Vasilii Vasil'evich Rozanov as Philosopher)*,

Giessen: W. Schmitz. (Contains extensive bibliography of works by and about Rozanov.)

Ruge, Arnold (1802-80)

Arnold Ruge was the most influential liberal writer and activist of the radical wing of Young Hegelianism. For him philosophy was a challenge to translate the humanist ideals of emancipation and self-determination into the realities of moral, cultural and political practice. As editor of powerful intellectual journals such as Hallesche und Deutsche Jahrbuecher (1838-43) with Theodor Echtermeyer, 'Anekdota zur neuesten deutschen Philosophie und Publizistik' (1843), 'Deutsch-Franzoesische Jahrbuecher' (1844) with Karl Marx, and 'Die Akademie' (1850), he became the leading promotor of liberal philosophy and civic emancipation in Germany. Ruge represented the citizens of Breslau in the Frankfurt Paulskirche parliament in 1848-9 and worked briefly with Alexandre Ledru-Rollin and Guizeppe Mazzini in establishing a short-lived 'European Democratic Committee' in London in 1849.

Ruge understood his critical educational, cultural and political activities as a direct calling from the heritage of European enlightenment and German idealism, thus transforming idealistic theory and vision into the realities of political practice and agitation. In this manner he promoted such radical figures as Bruno Bauer, Max Stirner, David Friedrich Strauss and Ludwig Feuerbach.

1 Philosophy and revolution

Arnold Ruge, born 9 September 1802 in Bergen on the German island of Rügen, died in 1880 in Brighton, England after thirty-one years of exile. One of the most influential and radical German liberal intellectuals and democrats, he had to take refuge first in Paris and London in 1843 where he worked with Karl Marx, and again after 1849 in London and Brighton. Ruge was a successor to the tradition of academic philosophy which runs from the Enlightenment through Kant, Fichte and Hegel. But Kant's concept of autonomy and Hegel's model of emancipation still were only theoretical and academic: 'Today we have to resolve not only the theoretical, but also the practical conflicts. Only the unity of free thought and free will, only the dissolution of the conflict between free thought and un-free mankind will result in the complete realisation of liberty' (Ruge 1852: 18). Times have changed: 'Our times are political, and politics work for the liberty of this world', he wrote in 1842 (1985 vol. 2: 254) in his influential critique of Kant's and Hegel's reluctance to address real practical matters of political change.

It is by free and reasoning people that the unconscious self-determination of humankind will be elevated to self-conscious self-determination. Only the free person is the true person; the realization of liberty lies in emancipation and humanization - that is, in humanism. Ruge held that the humanism of the future will translate traditional principles of existing and historic religions into the 'humanistic religion', determined to realize the 'essence of man through knowledge, beauty, and liberty'. Means to achieve these goals are 'a constitution of society rooted in self-determined development, free communities, schools and academies of the sciences and arts' resulting in a humanism which is characterized by 'true religious practice in ethics and the education of all humans to their true essence in equal, political, and economic community' (Ruge 1852: 26, 45).

2 Protestantism and romanticism

Contrary to Hegel's complex and consensus-oriented hermeneutical interpretation of dialectical processes in personal development and in the history of ideas and in world history, Ruge in 1842 identified only two conflicting forces in history as well as in personal development: Protestantism and romanticism. Protestantism is the progressive and evolutionary progress of political, cultural and personal development, a self-reforming and self-emancipating principle, 'free and autonomous as science which is an offspring of protestant reasoning'. Romanticism, the contradicting principle, dwells on emotions and sentiments, on 'inwardness' ('*Innerlichkeit*') which defines itself in antagonistic and dualistic terms against the outside world and its challenges (Ruge 1985 vol. 2: 128). Modern 'romantic inwardness' itself must be understood as a product of Protestant emancipation of individual conscience from heteronomous religious, metaphysical or political powers. In political terms, romanticism can easily be identified as the anti-emancipatory and reactionary force, and is politically represented by the suppressive and reactionary political powers on the European continent. While German Protestantism, in Ruge's assessment, does not live up to its calling and heritage any more as it has become a part of the reactionary powers of romanticism, the principle of Protestantism has already emigrated into liberal philosophy and radical

political theory. Ruge's conflict theory - influenced by Bruno Bauer's Manichean model of antithetical progress in philosophy and politics (see Bauer, B.) - becomes the blueprint for Marx's concept of ideational antithetics between materialism and idealism, and of political and economic antithetics between capitalism and proletarism. It marks the departure of radical and confrontational antithetical Hegelianism from the non-confrontational dialectical method of Hegel.

3 Religion of humanism

Ruge regards his political vision of a socialist and democratic republic of responsible and freely contracting individuals as the manifestation of free religion, the religion of humanism. In *Die Loge des Humanismus (The Lodge of Humanism)* (1852) and *Unser System (Our System)* (1850) he outlines his cultural, social and political programme of the religion of humanism as the logical consequence of the history of visions, ethics and economics. He holds that religious principles such as work-ethics and virtues, love and solidarity, altruism and mutual understanding have to be released from the bondage of established churches and religious sects in the same way as individuals have to be freed from the dominating powers of heteronomous tutelage by reactionary and elitist models of government. Not only individuals, but also the institutions of society must be driven by self-determination and self-emancipation.

Ruge calls for free education, free religious societies, academies for the arts and sciences, for free speech and free public opinion, and a free economic market within a socialist framework guided by a culture of work ethics, justice, solidarity and equality. But such a vision, in his understanding, does not exclude the 'organisation' of free institutions of learning, arts, and crafts, of a system of labour organisational and distribution by state authorities. Heine called Ruge the grim doorkeeper of Hegelian philosophy; to be a philosopher for Ruge meant not only to analyse the good and to argue for it but also to will it and to fight for it.

See also: Hegelianism §3; Humanism; Stirner, M.; Strauss, D.F.

HANS-MARTIN SASS

List of works

Ruge, A. (1985-) *Werke und Briefe*, ed. H.-M. Sass, Aalen: Scientia, 12 vols.(To date the following volumes have been published: volume 2 (1988) *Philosophische Kritiken 1838-1846*; volume 3 (1988) *Literarische Kritiken (1838-1846)*; volume 4 (1988) *Politische Kritiken 1838-1846*; volume 10 (1985) *Briefwechsel und Tagebuchblaetter 1825-1847*; volume 11 (1985) *Briefwechsel und Tagebuchblaetter 1848-1880*.)

Ruge, A. (1846-8) *Saemtliche Werke*, 10 vols, Mannheim: J.P. Grohe.

Ruge, A. (1842) 'Hegels Rechtsphilosophie und die Politik unserer Zeit', *Deutsche Jahrbücher für Wissenschaft und Kunst* 189, 190; trans. J.A. Massey as 'Hegel's Philosophy of Law and the Politics of Our Times', in L. Stepelevich (ed.) *The Young Hegelians*, New York: Cambridge University Press, 1983, 211-37.(Offers a critique of Hegel from a Young Hegelian standpoint.)

Ruge, A. (1843) 'Eine Selbstkritik des Liberalismus', *Deutsche Jahrbücher für Wissenschaft und Kunst* 1843 (1); trans. J.A. Massey as 'A Self-Critique of Liberalism', in L. Stepelevich (ed.) *The Young Hegelians*, New York: Cambridge University Press, 1983, 237-60.

Ruge, A. (1852) *Die Loge des Humanismus (The Lodge of Humanism)*, Bremen: Arnold Ruge.(Ruge's thoughts on liberalism and humanism.)

Ruge, A. (1850) *Unser System (Our System)*, Leipzig: Verlagsbureau; repr. Frankfurt: Neuer Verlag, 1903. (Detailed presentation of Ruge's philosophical ideas.)

Ruge, A. (1854) *New Germany, its Modern History, Literature, Philosophy, Religion and Art*, London: Holyoake & Co.(Identical to *Die Loge des Humanismus*, 1852.)

Ruge, A. (1868) *Eight Lectures on Religion*, St Louis, MO (German edition was published in Berlin in 1869.)

References and further readings

Brazill, W.J. (1970) *The Young Hegelians*, New Haven, CT, and London: Yale University Press.(Concentrates on the interpretation of the 'young Hegelian metaphysic' (27-70) and Ruge's role in promoting Young Hegelian politics (227-60); includes a bibliographical essay (283-96).)

Eßbach, W. (1988) *Die Junghegelianer. Soziologie einer Intellektuellengruppe (The young Hegelians: Sociology of an Intellectual Movement)*, Munich: Fink.(An intellectual history of the ideas of Young Hegelian thought and their mode of argumentation.)

Loewith, K. (1964) *From Hegel to Nietzsche. The Revolution in Nineteenth Century Thought*, trans. D. Green, New York: Holt, Rinehart & Winston.(Classic contribution on Young Hegelian theory-formation and the process of radicalization within the movement.)

Neher, W. (1933) *Arnold Ruge als Politiker und politischer Schriftsteller (Arnold Ruge as politician and political writer),* Heidelberg: Winter.(Describes Ruge's political theory in the context of his political activities.)

Sass, H.-M. (1960) *Untersuchungen zur Religionsphilosophie der Hegelschule (Studies on the religion philosophy of the Hegelian School)*, Phil. Diss. Münster: Universität Münster.(Dissertation, discussing Ruge's humanist and liberal theory within the context of the Young Hegelian movement, also Ruge's political role as editor of influential critical journals.)

Walter, S. (1995) *Demokratisches Denken zwischen Hegel und Marx. Die politische Philosophie Arnold Ruges (Democratic thinking between Hegel and Marx: the political philosophy of Arnold Ruge)*, Düsseldorf: Droste. (Reconstructs and analyses Ruge's political philosophy, and analyses his liberal ideas of transnationalism, international peace, European Unity and democracy.)

Zanardo, A. (1970-1) *Arnold Ruge giovane hegeliano 1842-1849 (The young Hegelian Arnold Ruge 1842-1849)*, Contributo bibiografico Annali Istituto Giangiacomo Feltrinelli, 12, Milano: Istituto Feltrinelli. (Comprehensive annotated bibliograph, 1842-9.)

Rule of law (Rechtsstaat)

The 'rule of law' most simply expresses the idea that everyone is subject to the law, and should therefore obey it. Governments in particular are to obey law - to govern under, or in accordance with, law. The rule of law thus requires constitutional government, and constitutes a shield against tyranny or arbitrary rule: political rulers and their agents (police and so on) must exercise power under legal constraints, respecting accepted constitutional limits. The British and US conceptions of this ideal find a parallel in the Germanic concept of the Rechtsstaat, *or 'state*-under-*law', where the state as an organized entity is conceived to be limited by laws and by fundamental principles of legality, rather than being a purely political organization that can dispense with law in the interests of policy. Such concepts play an essential part in the political philosophy of liberalism; yet, characteristically, their more detailed exposition and indeed their nature and meaning are contested and controversial.*

In a wider sense, the rule of law articulates values of procedural fairness or due process which affect the form *of legal rules and govern the manner of their application. Those values both enhance the utility of legal regulation and also acknowledge underlying ideas of human dignity and autonomy. In a further sense, the rule of law refers to the faithful application of those rules and principles which constitute the law of a particular legal system. It expresses the idea that legal obligation should always be determined in particular cases by analysis of existing law - as opposed to* ad hoc *legislation by judges - even where disagreement may exist about the true meaning or content of the law.*

The connection between the rule of law and justice is complex. The rule of law cannot itself guarantee justice, but it forms an essential precondition. In so far as it imposes formal constraints on the laws enacted or enforced, which ensure that they are capable of being obeyed and that they are fairly administered, the rule of law assumes a conception of moral personality - of how individuals should be treated, as responsible human beings, capable of a sense of justice - which links the idea with the values of freedom and autonomy, and the ideal of equality.

1 Constitutionalism and the separation of powers

Constitutional government entails some form of separation of powers. Although the law may be changed by the exercise of political power, it may be altered only in compliance with established constitutional procedures; and all acts of government and public officials must respect the requirements of the law which apply at the time of acting. The legislative, executive and judicial powers of the state must not be simultaneously exercised by the same persons or the same institutions, without appropriate restraints. The independence of the judiciary is fundamental. Although various arrangements for the division of legislative and executive powers may suffice - the rule of law does not necessarily require complete institutional independence - the judicial power must be wholly distinct. Both government and legislature must act in accordance with the existing law - as ultimately interpreted by the judiciary - until it is altered by lawful (constitutional) means.

F.A. Hayek, however, defended a more rigorous interpretation. The law constituted a shield against arbitrary rule by virtue of its generality. If the political rulers were obliged to govern within the constraints of general laws, they could not single out particular persons for special treatment. John Locke also emphasized the importance of rule by general laws, as opposed to 'arbitrary decrees' tailored to the circumstances of particular cases.

Hayek objected that the rule of law had been undermined by the modern blurring of the legislative and executive functions. Every coercive act of government should be authorized by law in the sense of a 'universal rule of just conduct' - a general rule which was also abstract in the sense that it was not directed to the attainment of any specific policy objective or desirable set of circumstances. Hayek considered that the executive government should always respect the ordinary rules of private law, applicable to the actions of private citizens. The authorization by the legislature of particular acts of government which would otherwise be unlawful permitted the executive to escape the constraints of general laws. Acting as a representative assembly, whose majority is committed to the support of an administration or of schemes for the use of public resources, the modern legislature fulfils a primarily executive role in breach of the theory of separation of powers.

A.V. Dicey attributed similar importance to the subservience of government to the ordinary private law. In his analysis of the rule of law as a principle of British constitutional law, Dicey stressed the idea of equality, in the

sense of the equal subjection of everyone - public official as well as private citizen - to the civil and criminal law. The rule of law meant not only the absence of arbitrary or even discretionary governmental power, but also that constitutional law was the product of the courts' decisions in ordinary litigation involving public officials.

It is now widely considered that Dicey gave exaggerated emphasis to private law, at the expense of public law. It is not possible for government to accomplish necessary tasks of economic control and social administration without special powers conferred by, and subject to independent principles of, public law. None the less, Dicey's famous repudiation of 'administrative law' or *droit administratif* reflected the underlying rationale of the rule of law. Hayek distinguished between rules regulating the activities of governmental agencies, or regulations made by such agencies, on the one hand, and 'administrative powers over persons and property', on the other. The latter entailed the exercise of discretion and permitted discrimination between persons. The pursuit of 'social justice', if that involved treating the citizen and his property as objects of administration, was not compatible with a free society under the rule of law.

Constitutionalism means more than adherence to a purely formal concept of legality which satisfies the principle of separation of powers. If there were no limits on the nature and scope of government powers - because the laws permitted the exercise of unfettered discretion - there would be no safeguard against arbitrariness, and individual freedom would be constantly in peril. Hans Kelsen's 'pure' theory of law seems deficient in this respect (see Kelsen, H.). Kelsen identified the 'state' with the legal order, in the sense that the state was a 'personification' of the 'unity' of the legal order. Acts could be imputed to the state where they were performed by persons or institutions occupying certain superior positions within the hierarchy constituted by the legal order. The law of a state consisted of those norms whose validity could be logically derived from acts of will by the ultimate authority, which was itself empowered by the 'basic norm' necessarily presupposed by lawyer and 'legal scientist'. It was not necessary, then, that 'law' should have any particular content, provided only that it was enacted or laid down by the appropriate authority. In this minimal sense, every state is a *Rechtsstaat* - a state defined and regulated by law. It is apparent, however, that in this sense the rule of law has no value as a principle of constitutionalism, or safeguard of liberty. The contemporary German conception of *Rechtsstaat* postulates observance of fundamental principles of legality and proportionality, and respect for guaranteed constitutional rights expressing the inviolability of human dignity. Kelsen's formalism is largely rejected.

2 Justice and liberty

The special process-values which are closely associated with the rule of law, in its wider conception, were articulated by Lon Fuller. Fuller argued that the rule of law was the special virtue of law, in the sense that the existence of a legal system - as opposed to the exercise of unbridled power - involves compliance with the 'inner morality of law'. In addition to the requirement of generality - the exercise of power must be constrained by rules - the laws must be published so that the citizen can know their content. Obedience is possible, moreover, only if the laws are clear, and not contradictory. Legality is equally frustrated by laws requiring the impossible. Laws imposing strict liability for civil wrongs pose special problems; and strict criminal liability, whereby someone can be punished despite acting innocently and with due care, undermines the rule of law.

Laws should be prospective for similar reasons, although a retroactive statute might sometimes be needed as a curative measure to correct lapses in other requirements of the law's internal morality. The precept *nulla poena sine lege* expresses the widely shared condemnation of *ex post facto* criminal laws. There must also be reasonable constancy of law through time; and the actions of officials must conform to the rules declared. The rules of 'natural justice', or 'procedural due process', assist in achieving such congruence by ensuring that the law is correctly applied in particular instances.

Although Fuller asserted a connection between procedural legality and the moral quality of the law enforced, he failed to establish a necessary connection. Other writers have observed that legality is compatible with great iniquity because it places no limitations on the substantive content of law (see Legal positivism §1). It excludes certain sorts of arbitrary power, but fails to guarantee respect for human rights, civil liberties or social equality. The rule of law has been likened to a sharp knife - an efficient instrument for governance, but morally neutral between different purposes. In this respect, attempts to associate the rule of law with a wider vision of political liberty and social justice have been criticized for dissolving its discrete contribution to political theory in more diffuse and contentious philosophical claims.

A connection between the rule of law and the moral quality of law may, however, be made by paying closer attention to the underlying rationale of legality. Adherence to the rule of law acknowledges human dignity and autonomy by treating people as capable of using the law as a guide in determining their actions. Where the rule of law exists, citizens may take account of legal requirements in formulating their plans; and their powers of self-determination are not frustrated by the unpredictable exercise of political authority. John Finnis has observed that the rule of law recognizes the intrinsic importance of a certain quality of interaction between ruler and ruled, expressing values of reciprocity and procedural fairness whose preservation may often justify sacrifices in efficiency or other political ends (see Natural law §4).

Moreover, Fuller emphasized the likelihood of frequent conflict between the various canons of legality - there could be no utopia in which all the desiderata were always fully satisfied. The rule of law may therefore be thought to embrace an approach to the resolution of such conflicts, in any particular case, which itself respects the values of dignity and autonomy constituting its essential point. In this respect, the coherence of the 'inner morality' of law ultimately depends on considerations which cannot be regarded as wholly instrumental.

It is also mistaken to regard the rule of law merely as a fetter on the exercise of political power. The law does not serve merely to execute the will of political rulers, but to provide a stable framework within which each person may pursue their own ends while respecting a similar freedom in others. In this way, adherence to the rule of law serves to protect legitimate expectations, formed in reliance on private as well as public law and affording secure grounds for individual planning and independent action.

John Rawls has made a similar connection between justice and the rule of law. The regular and impartial administration of public rules - or 'justice as regularity' - is a necessary condition of a just scheme of social cooperation. The more perfectly a legal order fulfils the precepts of the rule of law, the more securely is liberty protected and the more effective is the organization of social conduct. The boundaries of our liberty are uncertain if laws are vague and imprecise, or the rules of natural justice are neglected. The principle of responsibility, whereby criminal liability should not exist in the absence of *mens rea*, also serves to protect individual liberty (Justice, equity and law; Law and morality).

3 Judicial obligation and legal decisions

The separation of powers assumes that judges are bound by existing law even when its application to particular cases presents questions of interpretation, whose proper resolution may be controversial. The ideal of the rule of law would seem to be seriously compromised if, in doubtful cases, judges were free to usurp the legislative function and create new law according to their own perceptions of justice or public policy.

Ronald Dworkin has offered a theory which seeks to explain the sense in which every judicial decision is determined by existing law, even in cases where good lawyers may reasonably disagree about what the law requires. The principle of separation of powers is reflected in Dworkin's distinction between principle and policy. Questions of policy are those which concern the public good or the general welfare, and they are primarily the province of government or legislature. Judicial decisions should be grounded mainly on principle, which concerns the scope and content of individual rights. When accepted rules do not clearly determine questions about legal rights, the judge must resort to analysis of underlying principle. He is bound by law in the sense that he must decide each case in accordance with the scheme of principle which best explains and justifies existing legal rules and practices. Since Dworkin supposes that, in any highly developed legal system, a sufficiently rigorous analysis would display good reasons of morality or consistency for judicial decision, even when the outcome may be controversial, he claims that there is in principle a correct answer to every question of law.

Argument about the requirements of the rule of law, in this further sense, focuses on the content of existing law, properly understood, rather than on more formal criteria which are independent of any particular legal system. Dworkin's account has the merit of showing the importance of the individual conscience in making judgments about the content of law. It is part of the idea of the rule of law that legal obligation is ultimately a matter of reason, and is therefore susceptible of understanding and intelligent obedience by every citizen. The identification of legal analysis with moral reasoning and political principle, however, makes the content of law closely dependent on controversial criteria of justice and fairness. It may be argued, none the less, that the values of autonomy and individual responsibility, which underlie formal accounts of the rule of law, assume precisely the

kind of independent moral judgment which Dworkin's theory makes central to legal analysis.

The concept of 'integrity', which expresses Dworkin's model of judicial reasoning, represents an accommodation between conflicting values of justice, concerning individual rights, and fairness, in the sense of the appropriate distribution of political power. Integrity requires that the law should reflect a consistent and coherent body of principles, in order that everyone is treated alike. The analysis therefore reveals the importance of equality as an ingredient of the rule of law. Dworkin shows that the formal equality which is ensured by the uniform and consistent application of legal rules - the least controversial feature of the rule of law - must be supplemented by a more thoroughgoing consistency of legal principle. The ideal of equality before the law ultimately assumes a substantive equality of legal protection for the rights and interests of every citizen.

See also: Law, philosophy of; Social theory and law

T.R.S. ALLAN

References and further reading

Altman, A. (1990) *Critical Legal Studies: A Liberal Critique*, Princeton, NJ: Princeton University Press.(A sophisticated but lucid defence of the liberal ideal of the rule of law against contemporary critics.)

Dworkin, R. (1977) *Taking Rights Seriously*, London: Duckworth.(Series of essays which challenge positivist conceptions of law; some assume a background knowledge of other works.)

(1985) *A Matter of Principle*, Oxford: Oxford University Press.(Essays on various aspects of legal philosophy which, though closely argued, do not generally assume specialized knowledge.)

(1986) *Law's Empire*, London: Fontana.(More difficult, but elaborates the author's theory of law as integrity.)

Finnis, J. (1980) *Natural Law and Natural Rights*, Oxford: Oxford University Press, 270-76.(Short, readable discussion, but proper appreciation requires wider study of the work as a whole.)

Fuller, L. (1969) *The Morality of Law*, revised edn, New Haven, CT: Yale University Press.(Very readable and easily understood explanation of procedural legality.)

Hayek, F. (1960) *The Constitution of Liberty*, London: Routledge & Kegan Paul; Chicago, IL: University of Chicago Press.(Easily readable; no specialized knowledge required.)

(1973-9) *Law Legislation and Liberty*, vol 1: *Rules and Order* (1973), vol 2: *The Mirage of Social Justice* (1976), vol 3: *The Political Order of a Free People* (1979), London: Routledge & Kegan Paul; Chicago, IL: University of Chicago Press.(Perhaps more difficult, but no previous study of philosophy necessary.)

Kelsen, H. (1945) *General Theory of Law and the State*, trans. A. Wedberg, Cambridge, MA: Harvard University Press, 1949.(For the specialist reader: full statement of the author's conception of legal positivism.)

Locke, J. (1690) *Two Treatises of Government*, P. Laslett (ed.), Cambridge: Cambridge University Press, vol. 2, chaps 9-12, 1960.(Early and readable treatment of separation of powers.)

Loughlin, M. (1992) *Public Law and Political Theory*, Oxford: Clarendon Press.(An analysis of the evolution of modern British public law in response to conflicting traditions of thought about political theory and the nature of law.)

Lucas, J. (1966) *The Principles of Politics*, Oxford: Oxford University Press, 78-143. (Accessible to the general reader; the rule of law is set in the context of other associated ideas and principles.)

Rawls, J. (1972) *A Theory of Justice*, Oxford: Oxford University Press, 235-43. (Straightforward account, but should be read in the light of the author's wider theory of law and justice.)

Raz, J. (1977) 'The Rule of Law and its Virtue', *Law Quarterly Review* 93, 195-211; repr. in J. Raz, *The Authority of Law*, Oxford: Clarendon Press, ch. 11, 1979. (Good general discussion which does not assume specialized knowledge.)

Russell, Bertrand Arthur William (1872-1970)

Bertrand Russell divided his efforts between philosophy and political advocacy on behalf of a variety of radical causes. He did his most important philosophical work in logic and the philosophy of mathematics between 1900 and 1913, though later he also did important work in epistemology, metaphysics and the philosophy of mind, and continued to contribute to philosophy until the late 1950s. He wrote relatively little on ethics. His political work went on until his death.

In the philosophy of mathematics his position was logicism, the view that all of mathematics can be derived from logical premises, which he attempted to establish in detail by actual derivations, creating in the process what is essentially now the standard formulation of classical logic. Early in this work he discovered the self-referential paradoxes which posed the main difficulty for logicism and which he eventually overcame by the ramified theory of types.

Logic was central to Russell's philosophy from 1900 onwards, and much of his fertility and importance as a philosopher came from his application of the new logic to old problems. Among his most important logical innovations were the modern theory of relations and the theory of descriptions. The latter enabled him to reparse sentences containing the phrase 'the so-and-so' into a form in which the phrase did not appear. The importance of this theory for subsequent philosophy was that it enabled one to recast sentences which apparently committed one to the existence of the so-and-so into sentences in which no such commitment was suggested. This laid the basis for a new method in metaphysics (widely pursued by Russell and others in the first half of the century) in which theories about items of a given kind are reformulated so as to avoid reference to items of that kind.

Logicism itself offers just such a treatment of mathematics and in his later work Russell used the method repeatedly, though the reformulations he suggested were rarely so explicit as the ones he had offered in mathematics. In 1914 he proposed a solution to the problem of the external world by constructing matter out of sensibilia. After 1918 he proposed to construct both mind and matter out of events. After 1940 he treated all particulars as bundles of qualities. In each case his motivation was to avoid postulating anything that could be constructed, thereby eliminating ontological commitments which had no independent evidential support. Outside mathematics, his starting-point was the empirically given and he attempted to make his constructions depend as little as possible upon items not given in experience. He was not, however, a strict empiricist, since he did not think that empirical evidence alone would be sufficient for the constructions and he was always prepared to supplement it in order to obtain them. He wanted to construct, not those items which were empirically warranted, but those which were required by the relevant scientific theories, for he regarded science as the best available, though by no means an infallible, source of truth. The task, in each case, was therefore to reveal the least amount of apparatus that would have to be assumed in addition to the empirical data in order for the constructions required by science to be possible. This methodology, which he pursued throughout his career, gives an underlying unity to what, more superficially, appears as a series of abrupt changes of position.

1 Life

Russell was born into the Whig aristocracy and inherited many of the values of its most radical wing. His grandfather, a prominent Whig reformer of the 1830s, had twice been prime minister. Russell was orphaned before he was 4 and was brought up by his grandmother who had him educated at home. In 1890 he went to Trinity College, Cambridge, to study mathematics. He gained a first in mathematics in 1893 and then turned to philosophy for his fourth year.

In 1894 he married Alys Pearsall Smith and spent part of his honeymoon in Berlin working on his first book, *German Social Democracy* (1896). The following year he published *An Essay on the Foundations of Geometry*, a revised version of his dissertation for which he had been awarded a six-year fellowship at Trinity. This was the first of a projected series of books on the sciences. After some aborted work on physics, he turned to pure mathematics and logic, producing many of his most important contributions in the period to 1913. While doing so, he lived mainly off unearned income, but in 1910, his capital depleted, he took up a lectureship at Trinity where he taught until 1916.

Though Russell had been born with an interest in politics, it had not hitherto occupied much of his time (despite his study of the German Marxists and vigorous interventions on behalf of free trade and women’s suffrage). With the outbreak of war in 1914, however, his interest in philosophy lessened and he threw himself into writing, speaking and organizing on behalf of the pacifists. A conviction under the Defence of the Realm Act led to his dismissal from Trinity in 1916; a further conviction led to his being jailed in 1918. After 1916 he had only relatively short periods of academic employment and was dependent upon writing to make his living - a fact which only partially explains his huge subsequent output.

After the war he visited Russia to see the work of the Bolshevik government, but was disillusioned with its authoritarianism. In 1920-1 he spent a year at the University of Peking (Beijing). Returning to England he married Dora Black in 1921, just in time to legitimize his son; a daughter was born in 1923. Parenthood led him to take an interest in education and with his wife he started an experimental school. It was not, in Russell’s eyes, a success, and it was very expensive to run, requiring that he undertake regular lecture tours of the USA to raise money. His involvement with it ended, along with his marriage, in 1932, though Dora continued to run it on her own until 1943.

Through the 1920s and early 1930s Russell wrote prolifically on an astonishing range of topics, producing (among much else) books on Russia, China, relativity, history, education, sexual morality, international relations, religion and the future of society. Though much of this activity was necessary to make ends meet, Russell was a tireless advocate of progressive causes. By the mid-1930s, however, tiring of the precariousness of this way of life and needing now to support two families (he had remarried in 1936 and a third child was born in 1937), he hoped to return to academic life. This was not easy: positions were scarce and Russell was a controversial figure. In 1938 he lectured at Oxford, followed by visiting appointments at Chicago and UCLA and finally an offer of a permanent position at City College, New York. This last, however, provoked opposition from New York’s Catholic community and the appointment was overturned in a celebrated court case.

In 1940, therefore, Russell found himself unemployed and marooned in America by the war. At this point, the eccentric millionaire Albert Barnes came to his rescue with a five-year appointment to lecture on the history of philosophy at the Barnes Foundation in Philadelphia. Although Barnes fired Russell at the end of 1942, he permanently solved Russell’s financial problems, for not only did Russell collect a sizeable sum for breach of contract but the lectures he gave for Barnes became the basis for his hugely successful *History of Western Philosophy* (1945).

Russell returned to England in 1944 to take up a fellowship at Trinity College, where he completed his last great philosophical work, *Human Knowledge* (1948). His return marked not only a mending of his relations with Trinity but also with the British establishment. His continued condemnation of communism in general and the Soviet government in particular was well suited to the beginnings of the cold war and Russell enjoyed a period of unaccustomed respectability. Throughout the 1950s he continued to write prolifically, including a philosophical autobiography, *My Philosophical Development* (1959). He married for a fourth time in 1952.

His respectability was short-lived. After the death of Stalin his hatred of the Soviet government moderated, and the threat of nuclear war came to dominate his thoughts. He wrote extensively on the danger of war during the 1950s but increasingly felt the need for action and in 1958 helped found the Campaign for Nuclear Disarmament, and then the more militant Committee of 100. In 1961 he was jailed once more for inciting demonstrators to civil disobedience. The 1960s were a time of hectic political work for Russell. He lent support to many causes and was especially active in opposing the Vietnam War. His last political statement, on the Middle East, was written two days before his death.

2 Early work

As an undergraduate Russell came under the influence of the neo-Hegelianism which then dominated British philosophy, especially the work of J.M.E. McTaggart, James Ward and F.H. Bradley. His earliest logical views were influenced most by Bradley, especially Bradley’s rejection of psychologism. But, like Ward and McTaggart, he rejected Bradley’s metaphysical monism in favour of monadism. Even as an idealist, he held that scientific knowledge was the best available and that philosophy should be built around it. Through many subsequent changes, this belief about science and his pluralism remained constant.

In 1895 he conceived the Hegelian plan of writing an encyclopedia of the sciences, together with another series of books on social and political issues, the two to be united in a synthesis of theory and practice. He followed this plan long after its Hegelian inspiration had passed, and in old age he noted that he had written as much of both series as could have been expected - except for the final synthesis, which still eluded him. As a system, the encyclopedia was to be dialectical, with supercessions between the individual sciences. Within each science, however, transcendental arguments were used to establish its a priori elements. In *An Essay on the Foundations of Geometry*, for example, he attempted to show that projective geometry was a necessary condition for the possibility of external experience. At the same time, he argued that geometry gave rise to contradictions which could only be resolved by a supercession to kinematics. From geometry he turned to arithmetic (which preceded it in the dialectic) and to physics (which followed it), though what he wrote on these subjects remained largely unfinished (see 1983-: vol. 2).

Two main problems caused Russell to abandon this work. First, there was (as G.E. Moore pointed out) a lingering psychologism associated with his use of transcendental arguments. Second, the contradictions which supposedly arose in the special sciences generally were not eliminated by the transition to a new science, but merely reappeared in a new form. They resulted from the use of asymmetrical relations, but only if these relations were treated as internal. Russell resolved both problems in 1898 by abandoning idealism (including internal relations and his Kantian methodology). He called this the one real revolution in his philosophy (see Hegelianism §5; Transcendental arguments).

3 Logicism

The new philosophy was little help with Russell's immediate task - an analysis of the concept of number - until, in 1900, he discovered Peano's symbolic logic. Significantly, Russell's first contribution to Peano's logic was a formal theory of relations on the basis of which he was able to define the cardinal number of a class u as the class of classes coextensive with u and produce a theory of progressions which formed the basis for ordinal arithmetic (1901).

Using the new logic Russell propounded logicism, the view that the whole of pure mathematics can be derived deductively from logical principles, a position he arrived at independently of Frege, who held a similar but more restricted view (and who had anticipated Russell's definition of cardinals in 1884) but whose work Russell discovered only later (see Frege, G. §8). Russell's first full statement of logicism was published in *The Principles of Mathematics* (1903). The interest of this work is mainly philosophical: it contains none of the actual derivations of mathematical results from logical principles. The latter was to be carried out in a projected second volume to be written in collaboration with A.N. Whitehead (see A.N. Whitehead §2). Their work expanded over the next ten years and resulted in the three-volume *Principia Mathematica* (1910-13), in which detailed derivations were given for Cantor's set theory, finite and transfinite arithmetic (including an elegant generalization of ordinal arithmetic called 'relation-arithmetic') and elementary parts of measure theory (see Cantor, G.). A fourth volume, by Whitehead alone, on geometry (including complex number theory) was never completed. As a demonstration of logicism, therefore, *Principia* depends upon much prior arithmetization of mathematics, for example, of analysis, which is not explicitly treated. Even with these allowances much is still left out, such as abstract algebra and statistics (see Numbers §6).

4 Absolute realism

The philosophy which Russell had adopted after abandoning idealism in 1898 and which underlay his early logicism was an extreme realism (to which Moore also subscribed). It was fully stated in *The Principles of Mathematics* and had been anticipated in *A Critical Exposition of the Philosophy of Leibniz* (1900) and in unpublished writings (see 1983-: vol. 2).

In Russell's absolute realism, everything which can be referred to or made the subject of a proposition is a term which has being (though not necessarily existence). Terms are either things, which can occur in a proposition only as its subject(s), or concepts, which may occur as subjects but which may also occur as relations. A proposition is a complex of terms related to each other. All complexes are propositions; all propositions are terms. Terms are neither linguistic nor psychological, but objective constituents of the world. The first task of philosophy is the theoretical analysis of propositions into their constituents. The propositions of logic are unique in that they remain

true when any of their terms (apart from logical constants) is replaced by *any* other term.

In 1901 Russell discovered that this position fell prey to self-referential paradoxes. For example, if the combination of any number of terms is a new term, the combination of *all* terms is a term distinct from any term. This paradox is related to the greatest cardinal paradox which arose in Cantor's transfinite arithmetic. Cantor had shown that the power-set, $\wp(S)$, of any set *S* has a higher cardinality than *S*. But if *S* is the set of all cardinals, the cardinality of $\wp(S)$ will then be a cardinal greater than any in *S*. While thinking about this problem Russell came upon an even more troubling contradiction, now known as Russell's paradox, which affected the foundations of set theory itself. It seems natural to suppose that some sets are, but most sets are not, members of themselves. If we consider, however, the set of all sets that are not members of themselves and ask whether this set is a member of itself we arrive at Russell's paradox. For supposing that it is a member of itself, then it lacks the defining property of the set and so is not a member of itself. But if it is not a member of itself then it has the defining property of the set and so is a member of itself. The paradox is simply proven given the unrestricted comprehension axiom of naive set theory, that for any property (or propositional function) ϕ there is a set $\{x\colon \phi x\}$ of just those things that satisfy ϕ; that is to say, $a \in \{x\colon \phi x\} \equiv \phi a$. Let ϕx be $\sim(x \in x)$ and let *a* be $\{x\colon \sim(x \in x)\}$, then substituting in the comprehension axiom gives:

$$\{x\colon \sim(x \in x)\} \in \{x\colon \sim(x \in x)\} \equiv \sim\big(\{x\colon \sim(x \in x)\} \in \{x\colon \sim(x \in x)\}\big).$$

These paradoxes in set theory are related to other self-referential paradoxes. The best-known is the liar paradox, generated by the person who says 'I am lying': if what they say is true, it is a lie; if it is a lie then they speak the truth. The long delay in the publication of *Principia* was largely due to the difficulty of devising a logic which was both paradox-free and strong enough to support pure mathematics (see Paradoxes of set and property §4; Cantor's theorem; Continuum hypothesis).

5 Solutions to the paradoxes: the zig-zag theory

In *The Principles of Mathematics* Russell offered no firm solution to the paradoxes, but in the next few years he tried a variety of approaches. The first was what he called the 'zig-zag theory'. The intuitive idea behind it was simple: only 'fairly simple' propositional functions determine classes; 'complicated and recondite' functions do not (1906a: 145-6). The problem was to specify the degree of complexity which marked the divide. Russell's only published remarks on the theory were in a survey article written after he had abandoned the theory and are vague on this point (1906a). Unpublished writings reveal that he tried a number of formulations (1983-: vol. 4). The chief difficulty is to find an appropriately restricted comprehension axiom: versions with weak constraints permit new contradictions to be derived, but if the constraints are too strong, logicism (for example, Russell's definition of cardinal number) is undermined. Russell found no way of showing that a comprehension axiom strong enough to support logicism would be paradox-free.

The zig-zag approach has some affinity to Quine's 'New Foundations' theory (Quine 1937) and some more recent type-free theories (Feferman 1984), but the former imposes quite different constraints on comprehension, while the latter is inconsistent with logicism.

6 Solutions to the paradoxes: the substitutional theory

Russell's next attempt went considerably further. It occupied him from 1905 to 1906 and started from the theory of incomplete symbols he had devised in 1905 (see §9 below). Under the substitutional theory Russell eliminated classes, relations and predicates as terms over which variables ranged. Variables now range over entities, which are either simple individuals or complexes (including propositions). The fundamental notion of the theory is that of substituting one entity for another in a complex to yield a new complex. This is written '$p\frac{b}{a}!q$' and read '*p* with *b* substituted for *a* yields *q*', where *p* and *q* need not be propositions, though in the interesting cases they will be. '$p\frac{b}{a}!q$' will always be true or false whatever *p*, *a* and *b* are. If *a* does not occur in *p* then $p\frac{b}{a}!p$. Propositional functions and classes of individuals are expressed by matrices, for example, 'p/a', in which *p* is called the prototype and *a* the argument. *b* is a member of the class represented by 'p/a', if $p\frac{b}{a}!q$ and *q* is true. Propositional functions of propositional functions (or classes of classes) are expressed by the matrix '$q/(p/a)$'. Russell hoped to avoid the paradoxes because the application of a matrix to itself, which generates many of them, is impossible in the system. For example,

$$p\frac{p/a}{a}$$

is ill-formed, because ‘ p/a’ is an incomplete expression, not the name of an entity, and thus there is nothing which is substituted for *a*. Thus one can express with matrices everything that could previously be expressed by means of classes or functions except, as Russell puts it, for ‘certain limiting cases…[which] are precisely those that lead to contradictions’ (1906b: 171-2).

The ability to capture an iterative concept of classes by substitutional means preserved Russell’s definition of cardinal number and thus permitted the logicist programme to proceed. Russell spent much of 1906 developing logicism on a substitutional basis (see 1983-: vol. 5). The above account illustrates the lengths he was prepared to go to in order to preserve the unrestricted variables of the *Principles*. However, the fact that the theory allowed propositions as substituends permitted a propositional paradox to be derived. After several attempts to eliminate it and finding that, each time, it returned in more complicated forms, Russell reluctantly abandoned the substitutional theory.

7 Solutions to the paradoxes: the theory of types

In 1906 Russell adopted the solution which appears in *Principia*, namely the theory of types which banned self-reference by stratifying terms and expressions into complex hierarchies of disjoint sub-classes. The expression ‘all terms’, for example, is then meaningless unless restricted to terms of specified type(s), and a combination of terms of a given type is a term of different type; functions are of a different type to their arguments; classes of a different type to their members.

Ironically, a simple version of the theory had appeared in *The Principles of Mathematics* (appendix B). There, propositional functions are assigned a range of significance, ‘a range within which *x* must lie if $\phi(x)$ is to be a proposition at all’ (1903: 525). Significance ranges form types in that ‘if *x* belongs to the range of significance of $\phi(x)$, there is a class of objects, the *type* of *x*, all of which must belong to the range of significance of $\phi(x)$’ (1903: 525). Every significance range either is a type or is the union of several types. Types are hierarchically arranged: individuals constitute the lowest type, classes of individuals the next, classes of classes of individuals the next after that, and so on. Two-place relations (in extension) form a different hierarchy, three-place relations yet another, and so on. The theory bears resemblance to Chwistek’s simple type theory (Chwistek 1921), though there are important differences, notably that in Russell’s theory all ranges form a type, and propositions and numbers form distinct types outside the hierarchy. Since propositions are not stratified into types, Russell’s theory yields a propositional version of the contradiction (1903: 527).

Russell anticipated that the type theory of the *Principles* would have to be ‘transformed into some subtler shape before it can answer all difficulties’ (1903: 523). One reason he did not pursue this immediately was due to his reluctance to abandon the unrestricted variables of the *Principles* upon which his account of logic depended (see Urquhart 1989). Only when other avenues had been explored did he return to type theory. In the type theory of the *Principles* the type of a function had been determined exclusively by the type(s) of its argument(s). The ‘subtler shape’ of the theory Russell developed after 1906 - the ‘ramified theory of types’ - was obtained by adding order distinctions within each type based upon the form of the function. In a controversy with Poincaré, Russell had acknowledged that all the paradoxes involved violations of the ‘vicious circle principle’: No totality can contain members defined in terms of itself; or, in the formulation Russell preferred, ‘whatever involves an apparent [bound] variable must not be among the possible values of that variable’ (1906c: 198). The ramified theory of types was designed to build order-restrictions into the bound variables themselves in such a way that violations of the vicious circle principle would be prevented by the symbolism itself.

Functions are significant only when their arguments for each argument-place are drawn from a specified range of items appropriate to that place. Russell expresses this by saying that a function presupposes a totality for each of its argument-places; such a totality is the significance range for that argument-place. As before, individuals constitute the lowest such totality. First-order functions take only individuals as arguments; second-order functions take either first-order functions or individuals as arguments; third-order functions take either second-order functions, first-order functions or individuals as arguments; and so on. All the arguments of any function must be of a lower order than that of the function. If an *n*th-order function has an argument of order $n-1$, the function is

said to be *predicative*. Non-predicative functions all involve quantification over functions of lower order. (For example, if Russell is an individual, ‘Russell is clever’ employs a first-order predicative function of individuals, whereas ‘Russell has all the (first-order) qualities of a great philosopher’ employs a non-predicative function of second-order involving the quantification $(\phi)(x \text{ is a great philosopher } \supset \phi x)$, where the variable ϕ ranges over the totality of first-order functions. All bound variables range over only one such totality.)

Propositions form a similar hierarchy: *elementary* propositions have only individuals as terms; first-order propositions are either elementary or involve quantification over individuals; second-order propositions involve quantification over first-order propositions; and so on. First-order propositions presuppose no totality except that of individuals and, in general, propositions of *n*th-order presuppose totalities only of propositions of order $n-1$. The hierarchy of propositions can be derived from the hierarchy of functions since a proposition involving a totality of *n*th-order propositions can be reduced to one involving a totality of *n*th-order functions. Observing the restrictions imposed by these hierarchies ensures that no violations of the vicious circle principle can occur. The liar paradox is avoided, for instance, because what the liar asserts is ‘(*p*)(I affirm *p* and *p* is false)’. For this to be significant the bound variable ‘*p*’ must range over propositions of some specified order, while the proposition the liar asserts will be of a higher order and thus not among the propositions the liar asserts to be false.

In 1908, Russell sketched the development of logic and set theory up to the definition of cardinal and ordinal numbers on the basis of the ramified theory, a process that is completed in *Principia*. There are some differences between the 1908 presentation followed above and that in *Principia*: for instance, in *Principia* elementary propositions are not included among first-order propositions and propositional quantification is avoided. This is related to a more important change, namely Russell’s treatment of propositions in *Principia* as incomplete symbols (see §10) which requires a radical distinction between elementary propositions and those involving first-order quantification and is the basis for his claim that the propositional hierarchy is derived from the functional one. Some elements of the substitutional theory survive into ramified type theory, indeed Landini (1993) goes so far as to call the 1908 version a ‘modification’ of the substitutional theory. As in the substitutional theory, classes remain incomplete symbols, and Russell presents the hierarchy of functions in explicitly substitutional terms (see Theory of types).

8 Problems with ramified types

Russell’s reluctance to embrace the ramified theory of types was well-motivated, for it raised serious obstacles to logicism. In the first place, Russell’s definition of a cardinal number as the class of all similar classes is no longer admissible. There is no longer a class of all similar classes, but only a class of all similar first-order classes, a class of all similar second-order classes, and so on. (Strictly, of course, the definition will proceed in terms of the hierarchy of functions, but the class terminology is more familiar and admissible as a *façon de parler*.) There then ceases to be a unique cardinal 2, for example, but a different 2 at each level of the hierarchy; ordinary talk about 2 is typically ambiguous between these different cardinals. Similar problems affect the statement of logical principles which, strictly, would need to be stated for each order. In *Principia* this is avoided by the use of free variables which are typically ambiguous. Thus $\vdash . \ \phi x \vee \sim\!\phi x$ asserts the law of excluded middle for monadic propositions of undetermined order; whereas $\vdash .(\phi).\phi x \vee \sim\!\phi x$ asserts it for those of some particular order. These results are untoward but not intolerable.

A similar, but more troubling difficulty threatened Russell’s treatment of the continuum, for on the ramified theory the least upper bound of a bounded set of real numbers is a number of a higher level. The most serious difficulty, however, concerns the principle of mathematical induction: namely, that any function satisfied by 0, and satisfied by $x+1$ if satisfied by *x*, is satisfied by all natural numbers. These two difficulties threaten the logicist programme for transfinite and finite arithmetic respectively. Russell surmounted them by making the assumption that for every nonpredicative function there is an equivalent predicative function in the same variables. This is his ‘axiom of reducibility’: $(\exists f)(\phi x \equiv_x f!x)$ (1910-13: *12.1).

It has often been objected that the axiom of reducibility undoes all the good that ramifying type theory achieved, on the grounds that it renders the ramified theory equivalent to the simple theory, thereby readmitting all the paradoxes ramification was designed to prevent (Ramsey 1925; Copi 1950). Myhill (1979) shows that this objection is ill-founded.

Frank Ramsey (1925) proposed that the paradoxes should be divided into those which belong to logic and set theory and those which are semantical or linguistic and that logic need solve only the former. Since the simple theory of types solves the logical paradoxes, the ramified theory is not needed in logic. This suggestion has been very influential. However, it still leaves the semantic paradoxes to be dealt with and Ramsey's demarcation between the two kinds is hardly adequate. His suggestion that the semantic paradoxes are in some way empirical is neither clear nor plausible. Moreover, the fact that both types of paradox arise from self-reference suggests a uniform treatment. Above all, Ramsey's distinction does little justice to Russell's appealing intuition that the same underlying logic should be applied universally. It is clear now that semantics is capable of just as rigorous, formal treatment as set theory and the suggestion that the proper formal treatment of each requires different logical principles lacks *prima facie* plausibility (see Semantic paradoxes and theories of truth §2).

Under the influence of Wittgenstein, Russell came to accept the view that logical propositions were all tautologies. From this point of view the axiom of reducibility is clearly unsatisfactory (Wittgenstein 1922). From much the same point of view, Ramsey argued that it was not only contingent but a contingent falsehood - though it is not clear that his counterexample is a good one (Ramsey 1925). In the second edition of *Principia*, therefore, Russell tried to dispense with the axiom. He conceded that much of transfinite arithmetic would be lost as a result, but hoped none the less to save the theory of natural numbers by proving a version of the principle of mathematical induction within the constraints of the ramified theory and without the axiom of reducibility. Gödel (1944) noted the flaw in the proof and Myhill (1974) has since shown it to be irremediable. It follows, therefore, that even finite arithmetic cannot be logicized within the constraints of the ramified theory without the axiom of reducibility.

9 The theory of incomplete symbols

Both the substitutional theory and ramified type theory were closely linked to one of Russell's most important contributions to logic: his theory of definite descriptions and the more general theory of incomplete symbols it presaged. The theory of descriptions arose out of problems in Russell's theory of denoting in *The Principles of Mathematics*. It is plain that 'I met Quine' and 'I met the author of *Quiddities*' are different propositions, even though Quine is the author of *Quiddities*. In the *Principles*, the first proposition contains Quine, while the second contains a 'denoting concept' (expressed by the definite description 'the author of *Quiddities*') which denotes Quine. When Quine occurs in a proposition the proposition is about Quine, but when a denoting concept occurs the proposition is not about the concept but what the concept denotes. This theory, if it is to be statable, requires that there be some way in which a denoting concept, rather than its denotation, can be denoted. After much effort, Russell in 'On Denoting' (1905) concluded that this was impossible and eliminated denoting concepts as intermediaries between denoting phrases and their denotations by means of his theory of descriptions.

Russell showed in a broad (though not comprehensive) range of cases how denoting phrases could be eliminated in favour of predicates and quantified variables by providing an analysis of the sentence in which they occurred. This advance was made possible by Russell's study of Frege, whose theory of quantification replaced the intricate but unavailing theory of the *Principles*. Russell analysed sentences containing definite descriptions, such as, 'The successor of 1 is even', thus:

(1) $(\exists x)(x \text{ succeeds } 1 \;\&\; (y)(y \text{ succeeds } 1 \;\supset y = x) \;\&\; x \text{ is even})$

The three conjuncts in this analysis ensure, respectively, that a successor of 1 exists, that there is at most one successor of 1 and that whatever succeeds 1 is even.

In cases where the definite description does not denote, the sentence gets the value *false* on account of the falsity of the first conjunct of its analysis. This enabled Russell to treat as meaningful sentences containing non-denoting definite descriptions without having to suppose either that nonexistent objects were denoted (as in Meinong's theory of objects) or that there were such things as denoting concepts (as in Russell's 1903 theory) or Fregean senses. The subsequent literature has emphasized the first point, but it was the second which motivated Russell.

Some counterintuitive consequences of Russell's analysis were removed by his introduction of scope distinctions among the occurrences of descriptions within sentences. Not all problems are thus removed, however, and there have been many subsequent attempts to do better either through improvements to Russell's theory, or through 'free logics', or by means of Meinongian theories which permit quantification over nonexistent objects.

Obviously the theory will not permit the use of definite descriptions as substitution-values for the variables in (1) and this leads to a sharp distinction in Russell's philosophy of language between definite descriptions and names. Since Russell thought that most ordinary names were in fact disguised definite descriptions, he distinguished between descriptions and 'logically proper names' which alone could be substituted for variables. (The latter were names of objects of acquaintance - represented in ordinary language by 'this' and 'that'.) By treating ordinary names as descriptions he could use the theory of descriptions to account for the informative nature of identity statements such as 'Hesperus = Phosphorus'.

The principle of contextual definition employed in the theory of descriptions, which allowed sentences containing a description to be reparsed to eliminate the description, led Russell, after 1905, to make a sharp distinction between the grammatical and the logical form of a proposition. He applied similar techniques to other kinds of expression (for instance, to class names and to propositions). Russell held that the expressions thus eliminated, which he called 'incomplete symbols', had meaning only in the context of a sentence and were meaningless in isolation. The various ways in which they were eliminated constituted his 'theory of incomplete symbols' (see Descriptions; Existence §2; Free logics §§1, 3; Proper names §4; Sense and reference; Strawson, P.F. §2).

10 Theories of truth and judgment

The theory of incomplete symbols made possible Russell's substitutional theory, in which class names and relations were treated as incomplete symbols. The appearance of a propositional paradox in the substitutional theory convinced Russell that propositions should also be treated as incomplete symbols. This was accomplished by his multiple relation theory of judgment which became part of the philosophical underpinnings of ramified type theory. In Russell's early realism, true and false propositions alike were treated as subsistent complexes and belief as a relation between a mind and a proposition. This leaves it obscure why we prefer to believe true propositions. In the multiple relation theory, belief and other 'propositional' attitudes were treated as many-place ('multiple') relations between a mind and the individual constituents of the erstwhile proposition. Thus '*m* believes that aRb' has the form '$B(m, a, R, b)$', not '$B(m, p)$'. Apparent references to propositions are eliminated by subsuming their constituents within an actual psychological complex including a mind and related by some 'propositional' attitude. Propositions thus become fabrications of the mind. The belief represented by '$B(m, a, R, b)$' will be a true belief just in case there is a complex a-R-b.

Russell had considered such a theory as early as 1906, but put it aside while he worked on the substitutional theory in which propositions were needed as entities. The theory was taken up again and developed in writings from 1910 to 1913. The final development of the theory, in *Theory of Knowledge* (1913), was left unpublished by Russell because of criticisms from Wittgenstein, then his student at Cambridge. Wittgenstein's criticisms are perhaps most simply expressed as a dilemma. Either the constituents of a belief (*a*, *R* and *b* in the example above) are assigned to types or they are not. If they are not, then the 'propositions' fabricated by thought will not be subject to the ramified type hierarchy (it will be possible to judge nonsense, as Wittgenstein puts it) and the paradoxes will reappear. If they are, then they must be assigned to types by some prior judgment, to which the same considerations apply, and an infinite regress results (Sommerville 1981; Griffin 1985).

11 Logical atomism

Russell used the name 'logical atomism' for all his philosophical work after 1898, though now it is used mainly to describe his position from 1905 to 1919. The philosophy was atomistic because it took items in certain categories to be basic; it was logical because it sought to construct items in other categories by rigorous logical means. This involved a two-fold process: an analysis of concepts in ordinary or scientific use to discover which items were basic, followed by a logical reconstruction of the science from the basic items up. Although Russell's early logicist constructions were clearly along these lines, this analytic methodology came fully into its own only with the advent of the theory of incomplete symbols in 1905.

One important consequence of the theory of incomplete symbols was that the ontological commitments of a theory could be reduced by reformulating the theory to remove expressions which apparently denoted problematic entities. It was unnecessary to posit the existence of items which could be logically constructed. This formed the basis for what Russell called the 'supreme maxim in scientific philosophizing', his version of Ockham's razor: 'Wherever possible, logical constructions are to be substituted for inferred entities' (1983-: vol. 8, 11).

Parallel results held in epistemology. The theory of incomplete symbols showed how knowledge of a wide range of items could be achieved by knowledge (by acquaintance) of a much narrower range. Acquaintance was a direct, two-place cognitive relation, and objects of acquaintance were epistemically basic. Knowledge of other items was obtained by description, which always depended upon knowledge by acquaintance (1911). The distinction between knowledge by acquaintance and knowledge by description first appeared in 'On Denoting', where Russell first formulated his 'principle of acquaintance': 'in every proposition that we can apprehend…all the constituents are really entities with which we have immediate acquaintance' (1983-: vol. 4, 427). All other apparent constituents were to be eliminated by analysis. These innovations in 1905 marked the end of Russell's extreme realism, though he remained a Platonist in that he included universals among the objects of acquaintance (see Knowledge by acquaintance and description; Logical atomism §2; Ontological commitment).

12 Our knowledge of the external world

Russell has traditionally been regarded as the direct heir of the British empiricists, yet his work prior to 1910 belongs more naturally to the continental rationalist tradition. None of the thinkers who influenced him during this period - Kant, Bradley, Cantor, Peano, Frege - were empiricists, and it was not until 1912, when he was commissioned to write a short introduction to philosophy, *The Problems of Philosophy*, that he undertook a serious study of the British empiricists, especially Berkeley and Hume. Hume's influence appears directly in the chapter on induction (and in a paper of 1912, 'On the Notion of Cause'); Berkeley's was more indirect (since Russell rejected his idealism) and appears in Russell's treatment of the external world.

In *Problems* Russell held that empirical knowledge is based on direct acquaintance with sense-data and that matter itself, of which we have only knowledge by description, is postulated as the best explanation of sense-data. He soon became dissatisfied with this idea and, inspired by his logicist constructions of mathematical concepts, proposed instead that matter be logically constructed out of sense-data, thereby obviating dubious inferences to material objects as the causes of sensations. The actual data of sense, however, are too fragmentary for the construction of items with the expected properties of matter (such as permanence). To solve this problem Russell was led to postulate unsensed sensibilia in addition to sense-data. It is important to realize that for Russell, unlike most sense-datum theorists, sense-data were physical, located in physical space at varying distances from the place at which common sense located the material object. It is thus logically, though not practically, possible for more than one mind to be acquainted with the same sense-datum at the same time. Sense-data are merely sensibilia with which a mind happens to be acquainted, so properly speaking matter is constructed out of sensibilia rather than sense-data. In addition to matter, Russell also sought to construct space and time. All the sense-data with which a single mind is acquainted are located within a perspective, a private space peculiar to that mind. Similar spaces exist at every point at which a mind could be located. These private spaces Russell calls perspectives. Physical space is the set of all such perspectives. The arrangement of perspectives in physical space is achieved by the variation in the qualities of the sensibilia (such as their size and shape) as they appear in the perspectives: for instance, if the sensibilia in perspectives P_1 and P_2 are more similar than those in perspectives P_3 and P_4, P_1 and P_2 will be located closer together in physical space than P_3 and P_4.

This theory received its fullest expression in *Our Knowledge of the External World* (1914), though it remained in programmatic form; detailed constructions along similar lines were undertaken by Carnap (1928) (see Carnap, R. §1). Although Russell hoped that eventually a correlation between sense-data and the differential equations of physics would be forthcoming, it is difficult to see how this would ever be possible. Many of the standard objections to sense-data do not apply to Russell's theory, but it is still difficult to see how the gap between experiential evidence and physical theory could be bridged in the deductively rigorous way Russell wanted. For example, it is difficult to see how perspectives could be uniquely arranged in physical space when no sensibile occurs in more than one perspective. Nor is it clear how sensibilia could be grouped together to form the material objects required by either common sense or physics. Moreover, since sensibilia are physical and objective it would seem that they are still subject to misapprehension by the mind, thereby reintroducing the difficulties with illusion which had allegedly made material objects unviable as objects of acquaintance (see Perception, epistemic issues in).

13 Neutral monism

Before 1919 Russell wrote little about the philosophy of mind, though his dualism was evident in his analyses of judgment and acquaintance. By 1919, however, he came to think that mind, like matter, should be constructed out of more primitive material, a view akin to James's neutral monism (see James, W.). This change is somewhat puzzling, since Russell had criticized James's theory in some detail in *Theory of Knowledge* (1913) and, more tentatively, in 'The Philosophy of Logical Atomism' (1918). He was influenced partly by Hume's arguments for the unintrospectability of the self, but most importantly by Wittgenstein's treatment of belief, which 'shows…that there is no such thing as the soul' (Wittgenstein 1922: 5.5421).

Russell's neutral monism began with a new theory of belief in 'On Propositions' (1919a). He identified the contents of beliefs with images ('image-propositions') and words ('word-propositions'), understood as certain sorts of events - now the basic elements of his construction - which are neutral in the sense of being neither physical nor mental. Truth was analysed in terms of resemblance and causal relations. Wider issues are treated in *The Analysis of Mind* (1921), where the influence of contemporary trends in psychology, especially behaviourism, is evident. He never fully embraced behaviourism (for example, he did not attempt a behaviourist elimination of images), but he was prepared to push it as far as he thought it would go. He offered, for example, a behaviourist account of desire, about which he subsequently had doubts.

On Russell's neutral monism, minds and matter are constructed out of events and differ only in that they are organized by different causal laws: material objects by those of physics, minds by 'mnemic' causal laws, in which a past state of the mind is among the proximate causes of its current state. The theory is a precursor of later mind-brain identity theories.

In 1927, Russell returned to the problem of matter - this time, paying more attention to contemporary physics. Part I of *The Analysis of Matter* is devoted to relativity and quantum theory, and Part III deals with the causal and spatiotemporal order of events. In Part II, starting from a causal theory of perception, Russell asks what can be inferred from percepts (considered as events occurring within one's head) about the nature of their supposed external causes. Using the dictum that like causes have like effects, Russell answers that we can infer only the structure of the external world from the structure of our percepts - a doctrine known as structural realism.

Russell quickly abandoned structural realism when M.H.A. Newman (1928) showed that any set with the right cardinality could be arranged so as to have the same structure as the world - a result analogous to that claimed in Putnam's model-theoretic argument against realist theories of reference (Demopoulos and Friedman 1989). Russell, however, never abandoned neutral monism, though he left the theory incomplete. Constructions were suggested rather than provided (the fullest sketches offered were the treatment of belief in the second edition of *Principia* (1925) and a 1936 paper on temporal order). Moreover, while Russell had much to say about consciousness, he said very little about intentionality and (more surprisingly) personal identity. On the physics side, the breakdown of causality in quantum theory was a serious problem, given the importance Russell attached to causal relations in his constructions, but in 1927 it was perhaps not yet so well-established a fact as to demand explanation (see Neutral monism).

14 Later epistemology

Russell's penchant for reductive analysis was taken further in *An Inquiry into Meaning and Truth* (1940) where the terms of atomic facts, hitherto sense-data or events, are construed as bundles of compresent qualities. These qualities are the meanings of object-words, and bundles of them are denoted by logically proper names. This position was retained in *Human Knowledge* (1948); it marks the culmination of Russell's long battle with substance/attribute metaphysics. Despite its title, the *Inquiry* was more concerned with metaphysics and epistemology than with semantics. The meaning theory proposed there was causal and (in inspiration) behaviourist, though once again Russell did not expect a purely behaviourist analysis to be adequate.

Russell's allegiance to empiricism, like behaviourism, was never very strong and by 1936 he was writing on 'The Limits of Empiricism'. Even in *The Analysis of Matter* there are clear indications that he thought pure empiricism would render science impossible. In the *Analysis* he thought that Berkeleyan scepticism about matter could be avoided by construction, but that Hume's scepticism about induction was so far untouched. He tackled induction most thoroughly in *Human Knowledge* (1948), which contains his only extended treatment of probability. The problem for this book was set in 'The Limits of Empiricism', where he asked what non-empirical principles must

be assumed in order for science to be possible. Originally, he thought that induction itself would be sufficient. In *Human Knowledge*, however, he showed that inductive inference was not always truth-preserving (for reasons later popularized by Nelson Goodman as ‘the new riddle of induction’) (see Induction, epistemic issues in §1).

In place of induction, he proposed five ‘postulates of non-demonstrative inference’ as the minimum extra-logical, non-empirical apparatus required for science. His methodology here returned to that which had guided his work on *Principia* where he had sought the minimum logical apparatus necessary to support mathematics. It even had some resemblance to his earlier Kantian methodology, though Russell pointed out that he proposed his postulates as scientific hypotheses, not as synthetic a priori truths. The postulates, unfortunately, were vaguely stated in *Human Knowledge*, and Russell left more detailed work on them unpublished, probably on account of the frosty reception the book received.

By the time *Human Knowledge* was published linguistic philosophy was at its height and Russell’s work had become very unfashionable. Russell had no time for Wittgenstein’s later philosophy, nor for the ordinary language philosophy it inspired. He continued to think that understanding the world, rather than language, was the chief task of philosophy and his last philosophical writings, in the 1950s, were unavailing attempts to stem the tide of linguistic philosophy (see Ordinary language philosophy, school of §§1-3).

See also Analytical philosophy §§1-3; Idealism §§5, 7; Logicism; Monism; Moore, G.E. §4

NICHOLAS GRIFFIN

List of works

Russell, B.A.W. (1983-) *The Collected Papers of Bertrand Russell*, The McMaster University Edition, various editors, 30 vols anticipated, London: Routledge.(The definitive edition of Russell’s shorter writings and unpublished works. His philosophical papers are to be found in volumes 1-11.)

Russell, B.A.W. (1896) *German Social Democracy*, London: Allen & Unwin, 1965.(Russell’s first book. A study of German Marxism.)

Russell, B.A.W. (1897) *An Essay on the Foundations of Geometry*, New York: Dover, 1956. (The first instalment of the encyclopedia of the sciences. An attempt to rehabilitate a Kantian philosophy of geometry in the light of non-Euclidean geometry.)

Russell, B.A.W. (1900) *A Critical Exposition of the Philosophy of Leibniz*, London: Routledge, 1992.(For long an influential work on Leibniz, it also contains early intimations of Russell’s own realist philosophy.)

Russell, B.A.W. (1901) ‘Sur la logique des relations avec des applications à la théorie des séries’ (The Logic of Relations with Some Applications to the Theory of Series), *Revue de mathématiques* 7: 115-48; trans. R.C. Marsh, in Russell (1983-), vol. 3: 314-49. (The formalization of Russell’s theory of relations in Peano’s logical notation; an early statement of logicism which includes the Frege-Russell definition of cardinal number.)

Russell, B.A.W. (1903) *The Principles of Mathematics*; 2nd edn, 1937; London: Routledge, 1992.(A philosophical prolegomena to logicism.)

Russell, B.A.W. (1905) ‘On Denoting’, *Mind* 14: 479-93; repr. in Russell (1983-), vol. 4: 415-27.(The first published statement of Russell’s theory of definite descriptions.)

Russell, B.A.W. (1906a) ‘On Some Difficulties in the Theory of Transfinite Numbers and Order Types’, *Proceedings of the London Mathematical Society* series 2, 4: 29-53; repr. in Russell (1974), 135-64.(A survey of different approaches to solving the paradoxes.)

Russell, B.A.W. (1906b) ‘On the Substitutional Theory of Classes and Relations’ in Russell (1974), 165-89.(A survey of Russell’s substitutional theory, prepared for publication but withdrawn in the face of Russell’s discovery of new paradoxes.)

Russell, B.A.W. (1906c) ‘Les Paradoxes de la Logique’, *Revue de métaphysique et de morale* 14: 627-50; trans. ‘On ‘‘Insolubilia’’ and their Solution by Symbolic Logic’, in Russell (1974), 190-214.(A defence of logicism against criticisms by Poincaré. In it Russell recognizes violations of the vicious circle principle as the source of the paradoxes.)

Russell, B.A.W. (1908) ‘Mathematical Logic as based on the Theory of Types’, *American Journal of Mathematics* 30: 222-62; repr. in Russell (1956), 59-102.(The first version of ramified type theory with which Russell finally hoped to eliminate the paradoxes.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(The definitive statement of Russell’s logicism, containing detailed derivations of cardinal and ordinal arithmetic including Cantor.)

Russell, B.A.W. (1911) ‘Knowledge by Acquaintance and Knowledge by Description’, *Proceedings of the Aristotelian Society* 11: 108-28; repr. in Russell (1983-), vol. 6: 148-61.(An early epistemological paper by Russell, dominated by insights gained by his theory of descriptions.)

Russell, B.A.W. (1912a) *The Problems of Philosophy*, Oxford: Oxford University Press, 1974.(A popular introduction to philosophy, with the emphasis on epistemology.)

Russell, B.A.W. (1912b) ‘On the Notion of Cause’, *Proceedings of the Aristotelian Society* 13: 1-26; repr. in Russell (1983-), vol. 6: 193-210.(A sceptical, neo-Humean account of causation, and a denial of its importance for science.)

Russell, B.A.W. (1913) *Theory of Knowledge*, in Russell (1983-), vol. 8.(A major unfinished work on epistemology. Extant parts give detailed discussion of acquaintance and belief.)

Russell, B.A.W. (1914) *Our Knowledge of the External World*, London and Chicago, IL: Open Court; 2nd edn, London: Allen & Unwin, 1926; London: Routledge, 1993.(A deductivist treatment of the external world, programmatically proposing to construct material objects out of sensibilia.)

Russell, B.A.W. (1918) ‘The Philosophy of Logical Atomism’, *The Monist* 28: 495-527; 29: 32-63, 190-222, 345-80; repr. in Russell (1983-), vol. 8: 160-244.(A summary of Russell’s philosophical development to date with a survey of outstanding problems.)

Russell, B.A.W. (1919a) ‘On Propositions: What They Are and How They Mean’, *Proceedings of the Aristotelian Society* supplementary volume 2: 1-43; repr. in Russell (1983-), vol. 8: 278-306.(The first statement of Russell’s neutral monist theory of belief.)

Russell, B.A.W. (1919b) *An Introduction to Mathematical Philosophy*, London: Routledge, 1993. (A popular introduction to the logic of *Principia*.)

Russell, B.A.W. (1921) *The Analysis of Mind*, London: Routledge, 1992.(The fullest statement of Russell’s neutral monism.)

Russell, B.A.W. (1927) *The Analysis of Matter*, New York: Routledge, 1992.(The definitive statement of Russell’s structural realism about the external world.)

Russell, B.A.W. (1936a) ‘The Limits of Empiricism’, *Proceedings of the Aristotelian Society* 36: 131-50.(An admission that pure empiricism is an insufficient basis on which to construct science.)

Russell, B.A.W. (1936b) ‘On Order in Time’, *Proceedings of the Cambridge Philosophical Society* 32: 216-28; repr. in Russell (1956), 347-63.(A formal construction of temporal order based on Russell’s event ontology.)

Russell, B.A.W. (1940) *An Inquiry into Meaning and Truth*, London: Routledge, 1992.(As much metaphysics and epistemology as semantics. An important statement of Russell’s later position.)

Russell, B.A.W. (1945) *A History of Western Philosophy*, London: Routledge, 1993.(A popular, entertaining and brilliantly witty history. Scholars may sneer, but it’s well worth reading.)

Russell, B.A.W. (1948) *Human Knowledge: Its Scope and Limits*, London: Routledge, 1992. (Russell’s last major statement in epistemology.)

Russell, B.A.W. (1956) *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992.(A collection of important essays, edited by Marsh in collaboration with Russell.)

Russell, B.A.W. (1959) *My Philosophical Development*, London: Routledge, 1993.(Russell’s philosophical autobiography.)

Russell, B.A.W. (1967-9) *The Autobiography of Bertrand Russell*, London: Allen & Unwin, 3 vols.(The full autobiography of Russell’s life and times, including his political activism.)

Russell, B.A.W. (1974) *Essays in Analysis*, ed. D. Lackey, London: Allen & Unwin.(Further essays, with more of a bias towards logic. Contains a very useful three-part bibliography.)

References and further reading

Carnap, R. (1928) *Der Logische Aufbau der Welt*, trans. R.A. George, *The Logical Structure of the World*, London: Routledge, 1967.(The only serious attempt actually to furnish the logical constructions proposed in Russell (1914). Technical.)

Chwistek, L. (1921) ‘Antynomje logiki formalnej’, *Przeglad Filozoficzny* 24: 164-71; trans. ‘Antinomies of Formal Logic’ in S. McCall (ed.) *Polish Logic*, Oxford: Oxford University Press, 1967, 338-45.(Proposes a simple type theory reminiscent of that of Russell’s *Principles of Mathematics*. Technical.)

Clark, R.W. (1975) *The Life of Bertrand Russell*, London: Cape and Weidenfeld & Nicolson.(The standard biography.)

Copi, I. (1950) ‘The Inconsistency or Redundancy of *Principia Mathematica*’, *Philosophy and Phenomenological Research* 11: 190-9.(Claims that ramified type theory with the reducibility axiom is equivalent to simple type theory.)

Copi, I. (1971) *The Theory of Logical Types*, London: Allen & Unwin.(A simple introduction to simple type theory.)

Demopoulos, W. and Friedman, M. (1989) 'The Concept of Structure in *The Analysis of Matter*', in C.W. Savage and C.A. Anderson (eds) *Rereading Russell. Essays on Bertrand Russell's Metaphysics and Epistemology*, Minneapolis, MN: University of Minnesota Press, 183-99.(An excellent account of Newman's criticism of the *Analysis of Matter* and its ramifications.)

Feferman, S. (1984) 'Towards useful Type-free Theories', *Journal of Symbolic Logic* 49: 75-111.(Theories with some resemblance to Russell's zig-zag theory are proposed - though the Frege-Russell definition of cardinal number fails in them.)

Gödel, K. (1944) 'Russell's Mathematical Logic', in P.A. Schilpp (ed.) *The Philosophy of Bertrand Russell*, La Salle, IL: Open Court, 1989, 123-54.(Still one of the best things written on Russell's logic.)

Grattan-Guinness, I. (1977) *Dear Russell - Dear Jourdain. A commentary on Russell's logic based on his correspondence with Philip Jourdain*, London: Duckworth.(Difficult and technical, but an important and fascinating source on Russell's logic.)

Grayling, A.C. (1996) *Russell*, Past Masters series, Oxford: Oxford University Press. (A very concise popular introduction to Russell's thought, including his politics.)

Griffin, N. (1985) 'Russell's Multiple Relation Theory of Judgment', *Philosophical Studies* 47: 213-47.(An account of the development of the multiple relation theory and its collapse under Wittgenstein's criticisms.)

Griffin, N. (1991) *Russell's Idealist Apprenticeship*, Oxford: Clarendon Press.(A detailed survey of Russell's neo-Hegelian work and the reasons that led him to abandon it.)

Jager, R. (1972) *The Development of Bertrand Russell's Philosophy*, London: Allen & Unwin.(The most comprehensive book on Russell - including his social and political thought. Sometimes superficial, but a useful survey for the beginner.)

Hylton, P. (1990) *Russell, Idealism, and the Emergence of Analytic Philosophy*, Oxford: Clarendon Press.(One of the best books on some of the central themes of Russell's philosophy to 1913. Not introductory, but not technical.)

Landini, G. (1987) 'Russell's Substitutional Theory of Classes and Relations', *History and Philosophy of Logic* 8: 171-200.

Landini, G. (1993) 'Reconciling *PM*'s Ramified Type Theory with the Doctrine of the Unrestricted Variable of the *Principles*', in A.D. Irvine and G.A. Wedeking (eds) *Russell and Analytic Philosophy*, Toronto, Ont.: University of Toronto Press, 361-94. (Landini's two articles offer the best and most detailed account of Russell's substitutional theory. Both are technical, but those new to the subject should begin with the first.)

Monk, R. (1996-) *Bertrand Russell*, London: Jonathan Cape, 2 vols.(An important, insightful and controversial biography.)

Myhill, J.R. (1974) 'The Undefinability of the Set of Natural Numbers in the Ramified *Principia*', in G. Nakhnikian (ed.) *Bertrand Russell's Philosophy*, London: Duckworth, 19-27.(Proves the impossibility of Russell's attempt in the second edition of *Principia* to derive a principle of mathematical induction within the constraints of ramified type theory without the axiom of reducibility. Technical.)

Myhill, J.R. (1979) 'A Refutation of an Unjustified Attack on the Axiom of Reducibility', in G. Roberts (ed.) *Bertrand Russell Memorial Volume*, London: Allen & Unwin, 81-90.(A refutation of the claim in Copi (1950). Technical.)

Newman, M.H.A. (1928) 'Mr. Russell's Causal Theory of Perception', *Mind* 37: 137-48.(The definitive refutation of the structural realism of Russell (1927).)

Pears, D.F. (1967) *Bertrand Russell and the British Tradition in Philosophy*, London: Fontana.(The classic statement of the traditional view of Russell as the heir of the British empiricist tradition. Intricate in places, but not technical.)

Pears, D.F. (1972) *Bertrand Russell: A Collection of Critical Essays*, Garden City, NY: Doubleday.(Though dated, the essays are generally of high quality. Contains an exceptionally useful, though now somewhat dated, bibliography.)

Quine, W.V. (1937) 'New Foundations for Mathematical Logic', *American Mathematical Monthly* 44: 70-80; repr. in W.V. Quine, *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1953, 80-101. (Includes a proposal, like that in Russell's zig-zag theory, to avoid the paradoxes by restricting comprehension - though Quine's restriction differs from any proposed by Russell.)

Ramsey, F.P. (1925) 'The Foundations of Mathematics', *Proceedings of the London Mathematical Society* series 2, 25: 338-84; repr. in F.P. Ramsey, *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, ed. D.H. Mellor, London: Routledge, 1978, 152-212.(Contains the influential proposal to distinguish logical

from semantic paradoxes and to solve only the former within logic, thereby obviating the need for ramified, as distinct from simple, type theory. An important original contribution to logic in the Russellian tradition.)

Ryan, A. (1988) *Bertrand Russell: A Political Life*, London: Allen Lane.(A survey of Russell's political engagements; more narrative than analysis.)

Savage, C.W. and Anderson, C.A. (eds) (1989) *Rereading Russell: Essays in Bertrand Russell's Metaphysics and Epistemology*, Minneapolis, MN: University of Minnesota Press.(Several useful papers on Russell's early logic and his later epistemology.)

Schilpp, P.A. (ed.) (1944) *The Philosophy of Bertrand Russell*, La Salle, IL: Open Court; 5th edn, 1989.(An old collection of essays of variable quality - Gödel's is superb and Weitz's is still worth reading. The collection is of special value because Russell replies to the papers - but not, unfortunately, to Gödel's.)

Sommerville, S.T. (1981) 'Wittgenstein to Russell (July, 1913): ''I am very sorry to hear…my objection paralyses you''', in *Language, Logic and Philosophy: Proceedings of the 4th International Wittgenstein Symposium*, Vienna: Holder-Pichler-Tempsky, 182-8.(The first paper to identify Wittgenstein's objections to Russell's multiple relation theory.)

Urquhart, A. (1989) 'Russell's Zig-Zag Path to the Ramified Theory of Types', in I. Winchester and K. Blackwell (eds) *Antinomies and Paradoxes. Studies in Russell's Early Philosophy*, Hamilton, Ont.: McMaster University Library Press, 82-91.(An excellent, concise account of Russell's attempts to solve the paradoxes between 1903 and 1908 and why he was so reluctant to embrace ramified type theory. Not difficult.)

Watling, J. (1970) *Bertrand Russell*, Edinburgh: Oliver & Boyd.(An admirably concise and clear book concentrating on Russell's less technical work to 1914. The best 100-page introduction to the topic.)

Wittgenstein, L.J.J. (1922) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London: Routledge & Kegan Paul, 1966.(An astonishing and exceedingly difficult work by the other logical atomist, cited here for its criticisms of the logical foundations of *Principia*.)

Russell: The Journal of the Bertrand Russell Archives (1971-) publishes articles on all aspects of Russell's life and work.

Russian Empiriocriticism

Russian empiriocriticism was an ephemeral movement within Russian Marxism of the early twentieth century. Its brief existence and deep involvement in politics invite the judgment, 'of historic interest only'. But that pat phrase dodges the problem of comprehending history and acting appropriately, which obsessed such thinkers as A.A. Bogdanov (1873-1928) and A.V. Lunacharskii (1875-1933), the best-remembered of the Russians who looked to Mach and Avenarius for philosophic support of Marx. In its German origins empiriocriticism was an academic effort to avoid metaphysics while analysing experience as the source of knowledge. In Russia the focus moved outside of academic cloisters. How is one to relate action to social understanding, if one knows that action and understanding shape each other within an overwhelming process of socioeconomic transformations? Analysis of 'experience' or 'practice' in the Russian context - a 'backward' society under a tyrannical state in an age of total war - nullified the academic calm of Mach and Avenarius. Their Russian admirers wanted to justify Marx's claims of social knowledge that would be both scientific and revolutionary; they rejected philosophies that merely interpret the world in different ways, while the task, described in Marx's final 'Thesis on Feuerbach', is to change it.

1 Philosophy and politics

Insistence on a revolutionary fusion of philosophy, social science and political action was a common feature of Russian radicals, offered in justification of diverse philosophical and political positions. 'Orthodox' believers in 'dialectical materialism', a name that Plekhanov devised in 1891, were Menshevik as well as Bolshevik (the two wings of the Russian Marxist party, which split into antagonistic parties by 1912). Countercurrents of 'revisionist' philosophy emerged on both sides of the split, though mostly among the Bolsheviks, perhaps because the greater wilfulness of Bolshevik politics invited some drastic revision of 'materialist fatalism', some turn to schools of philosophy which privileged 'experience' and 'will'. Lenin denounced them all as 'subjective idealism', in *Materializm i ėmpiriokrititsizm (Materialism and Empiriocriticism)* (1909), but his polemic did not end their linkage with Bolshevism. Even after the 1917 Revolution Bogdanov defended his mixture of Marxism and empiriocriticism within the Communist Academy. Lunacharskii, who was the Soviet Commissar of Education until 1929, Pokrovskii, the chief of higher education, and Bukharin, the major communist ideologist of the 1920s, showed their partiality more discretely, for confession of orthodoxy was becoming obligatory within the communist state (see Marxist Philosophy, Russian and Soviet §§1-3).

'Revisionism' in philosophy should not be equated with 'reformism' in politics just because a few of Marx's German disciples proposed that equation at the turn of the century. It was formally rejected even in Germany, where the Marxist party was becoming reformist in political action while clinging to a professed faith in orthodox revolutionary theory. In Russia virtually all Marxists disdained political reformism both in theory and in practice; they took for granted an inevitable lower-class upheaval, and split on political strategies and theoretical justifications for guiding the working classes through revolution to democracy and socialism.

Such tangles of political and philosophical issues have preoccupied scholarly interest in Russian empiriocriticism, with a chronic temptation to impose simplistic correlations - such as the linking of Bolshevik wilfulness with radical positivism or Nietzschean voluntarism. The Russian Marxists themselves, whether orthodox or revisionist, were sufficiently open-minded to restrain that temptation during their first four decades, from the 1890s until Stalin's 'revolution from above' in 1929-32 (see Marxist philosophy, Russian and Soviet §4). His 'great break' initiated a strict party line in philosophy; before that violent establishment of uniformity Russian Marxists usually took for granted some variety of debatable correlations between philosophical views and political positions.

Nevertheless a difference in mentalities was emerging in those first forty years. Whether Menshevik, Bolshevik or in between, the Marxists who were not embarrassed to call themselves 'orthodox', though they rejected traditional religions and were closing their minds to new developments in philosophy. Their critics within the Marxist movement were trying to adapt it to the new positivism, which was retreating from the visionary claim that science yields a guiding plan of human history. Some of the critical Marxists were also succumbing to the new Romanticism, as we may call the insistence by Nietzsche and James that beliefs precede the reasoned inquiry which serves them, that action is the ground of human being and reason its ancillary. Goethe's blasphemous

inversion of Genesis - 'In the beginning was the deed' - was a favourite quote within all Marxist trends; they divided on ways to confess such irrationality and still to justify their claims of rationality. The orthodox way was to barricade the mind within a confessional creed, favouring defensive apologetics rather than critical inquiry, claiming to be rational critics of their own beliefs while actually shutting off the possibility. The revisionist way was to resist that closing of the Marxist mind; revisionists would neither quit the movement nor quit criticizing it.

The movement - parties committed to Marxism - was swiftly growing in Germany and in Russia, while grand philosophical systems were disintegrating, obliging thoughtful militants to decide whether they were committed to a grand system, to a cluster of particular theories claiming scientific justification, to a mode of inquiry for the guidance of action, or to some combination of all three - with the possible admixture of an irrational will to power. Orthodox Marxists, such as Plekhanov and Lenin, tended to beat down such questions in polemics against those who asked them, in vehement defence of vague generalities. For example, they denounced positivism in name and praised dialectical materialism, but tended to favour the notion that a queenly discipline of philosophy withers away as each branch of knowledge becomes a genuine science. The Russian Marxists who scoffed at dialectical materialism and studied empiriocriticism were less content with generalities, and more receptive to the corrosive inquiry that disintegrates grand schemes. Yet they wound up, paradoxically, doing more system-building than their orthodox opponents.

2 Bogdanov

The paradox was most evident in Bogdanov, known as the chief 'Machist' among Russian Marxists, though he repudiated the label. He created 'empiriomonism', which was supposed to overcome, unlike Mach's empiriocriticism (see Mach, E. §2), the fragmentation of modern knowledge and the surrender to history as blind process or 'elementalism'. (The Russian word, *stikhiinost'*, is commonly mistranslated as 'spontaneity', though the root is Greek *stoicheion*, element, and the Russian meaning is not free agency but mindless process.) Seeking a unified science of all sciences, which would bring organized consciousness to the 'elemental' upheaval of the lower classes, Bogdanov drew first on Ostwald's 'energeticism'. He liked its vision of science as the monistic inquiry that studies energy in its transformations, culminating in human consciousness with its capacity for rational organization. But he criticized Ostwald for confusing the unity of scientific method with a notion of energy as metaphysical substance, and found in Mach's analysis of experience complete escape from metaphysics into methodology. Yet Bogdanov's Mach was ancillary to his Marx, methodology to worldview, for he insisted that human experience is collective, historically developing, organized by evolving 'ideology', Bogdanov's umbrella category for all forms of knowledge and art at all stages of human development, which he brought to order in empiriomonism.

Bogdanov claimed, in effect, to be creating a system that was not a system. 'Empiriomonism' indicated both 'the final philosophical goal, the ideal of cognition' and 'the way to it, the effort to give as coherent a world picture as possible for our time and for that social class to which I have devoted myself'. To support that version of Marxism he relied not only on contemporary positivists. Zarathustra's declaration, 'Man is the bridge to the superman', served as epigraph to an early article, gesturing towards a future transcendence of present values. But a subsequent explanation pointed backward in time:

> The ideal of a complete and powerful, harmonious and all-victorious life, which is expressed in the sculptures of the ancient world, and in the statues of their gods, is clearer and, I would suggest, deeper than Nietzsche's 'superman' in philosophy.
>
> (Bogdanov 1911: 17)

He did not examine in depth the implicit clash between Romanticism and classicism, that is, between a vision of human values in open-ended creation and a vision of endless striving towards a changeless ideal.

Bogdanov's eclectic mixture was still Marxism, he insisted, for various philosophies were 'materials' for his theorizing, but 'the social philosophy of Marx…has been the regulator and the method of my work'. He disagreed with revisionists who argued that Marx's social theorizing lacked philosophical justification. Bogdanov found it in Marx's vision of the historical process, which, in his rendition, resembled Comte's three-stage scheme: first an authoritarian social system justified by religious thought; then an individualist stage with an exchange economy and speculative systems of thought; finally the collective stage and scientific thought, which were coming together

in the proletarian struggle for emancipation within industrial society (see Comte, A.). Out of that struggle Marx had distilled a guiding consciousness, which needed further development as ‘empiriomonism’ or ‘universal organizational science’ or ‘tectology’. Bogdanov moved from one term to the next in the course of his prolific writing from the 1890s to the 1920s. (He borrowed the word ‘tectology’ from another scientific monist of the period, Ernst Haeckel, who grounded his grand system in biological rather than social science.)

Some scholars have interpreted Bogdanov as a thinker like Sorel, who declared social theorizing to be myth-making that justifies the wilful behaviour of social groups. Others have narrowed Bogdanov’s vision of organizational science to mathematical modelling of ‘systems’, and picture him as a forerunner of cybernetics. Still others, the present author included, take due note of Bogdanov’s strenuous *resistance* to the separation of mythopoeic from scientific modes of thought, of subjective from objective aspects of experience. That resistance sustained his old-fashioned positivist faith in science as unified knowledge that reveals the whole pattern of history, while he ostensibly confronted the downfall of such pretensions, the rising critical insistence that knowledge is irremediably partial, in both senses of the word: limited and biased. Bogdanov tended to evade the version of this antinomy that was most distinctive of Marx, the striving for a history of socioeconomic formations that would be simultaneously particular and universal, true to the variegated past of human societies and predictive of an authentically human future for all. A dreamy abstraction marked the ‘general task’ that Bogdanov set for ‘empiriomonism’: ‘to find the way by which it would be possible to systematically reduce all the discontinuities [*pereryvy*] of our experience to a principle of continuity [*nepreryvnost’*]’ (*1905-7: vol. 1, 173*).

Bogdanov’s disagreements with Lenin have sometimes been interpreted as a clash between freedom and authoritarianism within emergent communism, though both agreed, while founding the Party, that it must bring disciplined collective ‘consciousness’ to the ‘elemental’ struggle of the lower classes. Bogdanov distinguished between authentic revolutionary consciousness, a free collectivism that he attempted to encourage in the movement for ‘proletarian culture’, and the false consciousness of a social group in bondage to ‘verbal fetishism’ or ‘vampirism’. Those were his favourite cursewords for blind worship of dead formulas, idolatrous cults that feed on the revolutionary movement. Since he published such criticism of the Leninists as they took power, he has been praised as an original Bolshevik who tried to keep alive the yearning for ‘utopia’ as present consciousness, endlessly disturbing to those whose chants of a distant ‘utopia’ soothe the slaves of ‘verbal fetishism’. That praise may not be compatible with tributes to him as a prophet of ‘systems theory’ and computerized society - or with the view of him as ‘the red Hamlet’, tragically baffled by the conflicting demands of thinking and doing. It remains to be seen which view of Bogdanov, or indeed whether any significant interest in him, will survive the collapse of the state that turned Russian Marxism into a petrified orthodoxy.

3 Lunacharskii and others

The same question hangs over the other theorists who tried to keep Russian Marxism alive by infusion of new philosophies. In the Soviet period Lunacharskii seemed to have a lasting hold on posterity, but the reasons may have been more political than intellectual. He had been expelled from Lenin’s party along with Bogdanov in 1909, but he rejoined in 1917, as Bogdanov did not. He was Commissar of Education for the first twelve years of the Soviet state, died before the wild destruction of ‘old Bolshevik’ survivals in the late 1930s, and enjoyed a revival in the post-Stalin efforts to revitalize the communist vision. His flood of publications was more varied than Bogdanov’s, with artistic flair to draw readers in and less abstractness to tax their patience. Literary critic and playwright as well as philosophical essayist and revolutionary boss, Lunacharskii was, like Trotsky, a spectacular analogue to the men of letters who played major roles in the ‘bourgeois’ revolutions of an earlier age. Yet these Russian stand-ins for Jefferson and Condorcet, however energetic and creative, were at odds with their era, which splinters the intellect into specialized disciplines and puts the label ‘superficial’ or ‘dilettante’ on those who cling to the tradition of the universal mind.

Lunacharskii studied with Avenarius at the University of Zürich, but grinding at the self-doubts of academic positivism could not confine a spirit captivated very early, as he confessed, by ‘religious and artistic expressions of man’s life of the heart’. They became his ‘authentic *life’s problem*’ after an ‘ecstatic "conversion to Marxism"’. Most Marxists recoiled against such talk, insisting that their doctrine was not religion but science. Lunacharskii argued that it was both. Marx had endowed the socialist faith with knowledge of the historical process, but that ‘cognition of the world’ still needed an appropriate ‘evaluation of the world’, which projected human preferences

into an otherwise indifferent universe. Thus Lunacharskii became a leader of ‘God-building’, as a sympathetic Maksim Gorkii named the effort to create a faith out of human aspirations, to connect each ephemeral life with a progressive reality larger than itself.

> To live the life of a human being means to see and to feel the past as one’s own, even before one’s day of birth, and the future as one’s own, *really one’s own*, even beyond the grave; this means to be conscious of oneself as a mortal part of an immortal element [*stikhiia*] - of the collective life.
>
> (1909: 87)

Lunacharskii dismissed the similarity with Comte’s religion of humanity, which he considered a ridiculous imitation of Catholicism. He derided ’Plekhanovist’ orthodoxy as ’materialist fatalism’, an unacknowledged religion that invoked cosmic processes for assurance of human progress, culminating in socialism. Not mindless atoms and indifferent galaxies, he insisted, but evolving social practice was the ground of Marx’s theorizing, and therefore Bogdanov was ’the sole Marxist philosopher continuing the pure philosophical tradition of Marx’. Lunacharskii ridiculed ’the guards at the Marxist museum…who have raised cries of heresy’, while they reviled the demand for clear distinctions between the true faith and taboo beliefs. Such mockery was part of his serious argument that knowledge of humanity’s progress towards socialism must be charged with intense feeling if the movement is to go forward as our knowledge tells us it must.

Lunacharskii’s aesthetic theorizing also rested on insistence that knowledge and feeling are necessarily intermixed. He joined Nietzschean versions of that theme with biologistic talk, taken from Avenarius and Spencer, of nerve energies raised towards a destructive level, and released in some life-enhancing way. He distinguished two types of that aesthetic discharge: art as play and art as mobilization of emotion for work or struggle. Thus, as early as 1904, he came to a formula for a kind of art that would be called ’socialist realism’ in the Soviet period:

> To beautify the people’s life as much as possible, to draw pictures of the future gleaming with happiness and perfection, and along with that [to picture] the repulsive evil of the present, to develop the sense of tragedy, the joy of struggle and of triumph, of promethean striving, of stubborn pride, of unyielding courage, to join hearts in a general feeling of the urge toward the superman [see Nietzsche: impact on Russian thought §3].
>
> (1967: vol. 7, 99)

Such formulaic prescriptions were at odds with Lunacharskii’s enthusiasm for modernist revolts against prescriptive formulas, for art as experience reshaped in unpredictable forms. Praise of modernist writers flowed continuously from his pen, whether in theatre reviews for a pre-revolutionary newspaper or in lectures on West European literature at a communist university in the 1920s. Indeed, an enthusiastic preface to a Soviet translation of Proust was the last thing he wrote, in 1933, just before his heart stopped. Yet his aesthetic theorizing anticipated the formulaic rigidity of ’socialist realism’, and so did the aesthetics of other Russian Marxists, both revisionist and orthodox. (See especially essays by Lunacharskii, Trotsky, Bazarov, Iushkevich and so on, attacking ’literary decadence’ in *Literaturnyi raspad*, 1908-9). The anarchic creativity of the artistic avant-garde and the mobilization of political commitment by the revolutionary vanguard were at odds within Lunacharskii’s mind, as in Russia at large. He tried to reconcile the contrary currents, ultimately without success. He was committed to action on behalf of a revolution that was drifting into cultural petrifaction, a dreadful position for one who held social practice to be the criterion of truth.

The same can be said of nearly all the Russian thinkers who tried to enrich Marxism with new forms of positivism or Nietzscheanism. After the Soviet revolution most of them tried to work for the new regime. N. Valentinov - pen-name of N.V. Vol’skii (1879-1964) - who went West and turned against the whole revolutionary project, was exceptional, and even he tried to serve the Soviet cause for twelve years. V.A. Bazarov - pen-name of Rudnev (1874-1939) - hung on, tried to bend socialist economic theorizing towards realism, and was condemned for ’wrecking’ in 1931. Ia.A. Berman (1868-1933) seems to have suffered a similar fate. P.S. Iushkevich (1873-1945), creator of ’empirio-symbolism’, singled out by a recent historian as ’the most eloquent and outspoken protestor against dogmatism’ before 1917, survived afterward as a translator.

The experience of these thinkers must have some lesson for philosophies that claim ’experience’ or ’practice’ as the criterion of knowledge, but it may be impossible to specify the lesson without falling into vicious circularity - claiming to know how ongoing experience will work out - or leaping beyond historical experience, to romantic

claims of transcendent knowledge. If the social process to be comprehended is both a determinant and a product of the action to be taken, claims of comprehension may be masks of interests to be served, or worse yet, of interests poorly served, by a rash will to power or a cowardly fatalism, or by some confused fluctuation between the two extremes. Such issues, which threaten vicious circularity for all believers in the primacy of experience or practice, become painful dilemmas when would-be activists confront economic backwardness, tyranny, catastrophic wars, lower-class revolts and national conflicts, such as the Russian Empire and the USSR experienced in the twentieth century.

DAVID JORAVSKY

List of works

Bogdanov, A.A. (1905-7) *Ėmpiriomonizm; stat'i po filosofii (Empiriomonism: Articles on Philosophy)*, 3 vols, Moscow: Dorovatovskii & Charushnikov; 2nd and 3rd edns, 1906-8. (See Bogdanov for detailed bibliography.)

Bogdanov, A.A. (1911) *Kul'turnye zadachi nashego vremeni (The Cultural Tasks of Our Time)*, Moscow: Dorovatovskii & Charushnikov.(A pamphlet arguing for 'cultural revolution', and against the interpretation of Marxism by Plekhanov and Lenin.)

Lunacharskii, A.V. (1908-11) *Religiia i sotsializm (Religion and Socialism)*, vols 1 and 2, St Petersburg.(His major work on religion.)

Lunacharskii, A.V. (1909) *'Dvadtsat' tretii sbornik "Znaniia"' (The Twenty-third Collection of 'Knowledge')*, in *Literaturnyi raspad; Kriticheskii sbornik (Literary Decadence; a Critical Collection)*, vol. 2, St Petersburg: EOS, 1908-9.(Lunacharskii's contribution to a major attack on modernistic literature.)

Lunacharskii, A.V. (1964) *A.V. Lunacharskii o literature i iskusstve: bibliograficheskii ukazatel', 1902-1963 (A.V. Lunacharskii on Literature and Art: Bibliographical Index)*, ed. K.D. Muratova, Leningrad.(Bibliography of Lunacharskii's works on literature and art.)

Lunacharskii, A.V. (1967) *Sobranie sochinenii (Collected Works)*, vol. 7.(Includes the 1904 article 'Osnovy pozitivnoi estetiki' (Foundations of Positive Aesthetics).)*Sborniki* (collections) with articles by empiriocriticists:

Ocherki realisticheskogo mirovozzreniia; sbornik statei po filosofii obshchestvennoi nauke i zhizni (Essays on the Realistic Worldview: A Collection of Articles on Philosophy, Social Science and Life) (1904) St Petersburg: Dorovatovskii & Charushnikov.

Ocherki po filosofii marksizma (Essays on the Philosophy of Marxism) (1908) St Petersburg.

Literaturnyi raspad: kriticheskii sbornik (Literary Decadence: A Critical Collection) (1908-9) St Petersburg: Izdatel'skoe biuro.

Ocherki filosofii kollektivizma; sbornik pervyi (Essays on the Philosophy of Collectivism; First Collection) (1909) St Petersburg: Znanie.

References and further reading

Bailes, K.E. (1967) 'Sur la "théorie des valeurs" de A.V. Lunacharskij' (On A.V. Lunacharskii's Theory of Values), in *Cahiers du monde russe et soviétique* 8: 223-43.(A revealing analysis, using archival sources.)

Kelly, A. (1981) 'Empiriocriticism: A Bolshevik Philosophy?' in Cahiers du monde russe et soviétique 22 (1): 89-117.(A richly documented argument for an affirmative answer.)

Kline, G. (1968) '''âNietzschean Marxism" in Russia', in *Boston College Studies in Philosophy II: Demythologizing Marxism*, The Hague: Nijhoff.(The Nietzschean elements in Bogdanov *et al*.)

Kolakowski, L. (1978) *Main Currents of Marxism II: The Golden Age*, Oxford: Clarendon Press.(Ch. 17 analyses the conflict over empiriocriticism.)

Lebedev, A.A. (1970) *Ėsteticheskie vzgliady A.V. Lunacharskogo (A.V. Lunacharskii's Aesthetic Views)*, Moscow: Iskusstvo.(A Soviet effort to rescue Lunacharskii's aesthetics from orthodox hostility.)

Lenin, V.I. (1909) *Materializm i ėmpiriokrititsizm: kriticheskie zametki ob odnoi reaktsionnoi filosofii (Materialism and Empiriocriticism: Critical Notes on a Reactionary Philosophy)*, Moscow: Zveno.(Many republications and translations in the Soviet era.)

Plekhanov, G.V. (1957) *Izbrannye filosofskie proizvedeniia (Selected Philosophical Works)*, vol. 3, Moscow: Gosizdat.(Polemics against Bogdanov *et al*.)

Read, C. (1979) *Religion, Revolution and the Russian Intelligentsia, 1900-1912*, London: Macmillan.(Includes Marxist issues in larger debates about religion.)

Rosenthal, B.G. (ed.) (1986) *Nietzsche in Russia*, Princeton, NJ: Princeton University Press.(Part 3: 'Nietzsche's Influence on Russian Marxism'.)

Scherrer, J. (1978) 'La crise de l'intelligentsia marxiste avant 1914: A.V. Lunacharskij et le *bogostroitel'stvo*' (The Crisis of the Marxist Intelligentsia before 1914: A.V. Lunacharskii and God-building), in *Revue des études slaves* 51: 207-15.(Lunacharskii's role in God-building.)
Tait, A.L. (1984) *Lunacharskii, Poet of the Revolution, 1875-1907*, Birmingham: University of Birmingham Press. (Richly informed analysis of Lunacharskii's early development.)
Utechin, S.V. (1958) 'Bolsheviks and Their Allies after 1917: The Ideological Pattern', in *Soviet Studies* 10: 113-35.
Williams, R.C. (1986) *The Other Bolsheviks: Lenin and His Critics, 1904-1914*, Bloomington, IN: Indiana University Press.(A history that emphasizes political issues.)

Russian literary Formalism

Russian literary Formalism, an active movement in Russian literary criticism from about 1915 to 1929, approached the literary work as a self-referential, formed artefact rather than as an expression of reality or experience outside the work. It asked the question, 'How is the work made?' rather than 'What does the work say?' Its founding assumption, that poetic language differs from the language of ordinary communication, spawned numerous investigations of what the Formalists called 'literariness' - the qualities that make a work artistic. This distinction between practical and poetic language also allowed the Formalists to argue that literature was an autonomous branch of human activity, evolving according to its own immanent laws rather than as a consequence or reflection of historical events. Proceeding from this theoretical model, the Formalists viewed literary works as responses to previous literature rather than to the outside world.

In their literary theory and their interpretations of particular literary works, the Formalists were reacting to the predominant tendency of Russian literary criticism to draw direct correspondences between lived experience and the literary work. Boris Eikhenbaum, Roman Jakobson, Viktor Shklovskii, Boris Tomashevskii, Iurii Tynianov and other Formalists questioned accepted correspondences between life and art, casting doubt upon realist interpretations of Russian authors such as Gogol' and Tolstoi, and examining the narrative structure of non-Russian works such as Tristram Shandy and O. Henry's short stories. Their analyses showed how intonation, word order, rhythm and referential meaning interact within a literary work, and they argued that literary works are less a reflection of life than an attempt to refresh conventional perceptions. The influence of Russian literary Formalism is felt in more recent theoretical schools such as semiotics, structuralism, deconstruction, feminist criticism and new historicism, in so far as all of these take account of the particular use of language in any literary work.

1 Practical versus poetic language

Russian literary Formalism can be roughly divided into three periods. From 1915 to 1919, it sought to establish the distinction between practical and poetic language; from 1919 to 1921 it investigated the use of poetic language in particular literary works; and from 1921 until 1929 it examined literary works as responses to previous literary history. The movement began as part of the avant-garde experimentation in the arts in the years surrounding the Bolshevik revolution of 1917. Its practitioners formed two groups: the Moscow Linguistic Circle, which included Roman Jakobson, Pëtr Bogatyrev, N. Trubetskoi and Grigorii Vinokur; and its sister group in St Petersburg, the Society for the Study of Poetic Language (*Opoiaz*), whose members included Osip Brik, Boris Eikhenbaum, Lev Iakubinskii, Evgenii Polivanov, Viktor Shklovskii, Boris Tomashevskii and Iurii Tynianov. This collaboration between linguists and literary scholars addressed language as the focal point of literature. Verbal texture was examined as the artistic medium of literature, which contributed to message. In seeking to build an objective theory of literature, Russian literary Formalism ignored authorial intention, biography and social and historical conditions, the better to focus on the work itself.

Taking its cue from the sound experiments of Russian Futurist poetry, early Formalism drew attention to the non-referential aspects of language, particularly to the role of sound. In 'Voskreshenie slova' (The Resurrection of the Word) (1914), which may be considered the inaugural work of Russian literary Formalism, Viktor Shklovskii declared that the expressive impact of words was dulled by habitual usage and that the purpose of literature was to restore to words a sense of newness and to stimulate new perceptions. In 'Iskusstvo kak priëm' (Art as Device) (1916), Shklovskii introduced the concept of *ostranenie* ('defamiliarization') to refer to the ways in which literary works may overturn conventional perceptions.

The whole edifice of Russian Formalist theory rests on the differentiation between practical and poetic language, formulated by Iakubinskii in his 1916 article, 'O zvukakh stikhotvornogo iazyka' (On the Sounds of Verse Language). Iakubinskii argued that the goals of ordinary speech and poetic language were fundamentally opposed. Everyday utterances were aimed at rapid, efficient communication, and to that end they employed readily recognizable formulations. This 'automatized' language, as the Formalists termed it, grown familiar through repeated, habitual use, had attained a narrow functionality at the expense of its vitality and richness. The goal of poetic speech was to restore the palpability of language that is lost with quick recognition. Poetic language, said

Iakubinskii, retards the process of comprehension while making it multilayered, bringing out a wealth of accumulated meanings. Where practical language is highly referential, pointing unambiguously to objects and ideas, poetic language is largely self-referential, drawing attention to the verbal pattern of the work.

In their preoccupation with the renewal of automatized perceptions and with the properties of language, the Formalists were influenced by Andrei Belyi and Henri-Louis Bergson. Belyi's 1910 book, *Simvolizm (Symbolism)*, had attempted to establish an 'anatomy of style' by counting syllables, accents and parts of speech in thousands of lines of verse. Following Belyi, the Formalists took an almost scientific approach to literary analysis.

Until about 1919, Russian literary Formalism was occupied with attempts to characterize poetic language, investigating the question of how the renewal of automatized language was achieved. Studies by Brik, Polivanov and Shklovskii showed that in poetry and other short genres words were selected for their auditory properties no less than for their referential meaning. They explored sound patterns in literary texts, showing that meaning resides in sound as well as in semantics. Shklovskii in particular identified literary 'devices', such as retardation (the intentional slowing down of plot development to create suspense), parallel story lines and repetitions. Plot development itself was shown to be a device: the chronology of a narrative could be rearranged to achieve particular effects. The presumed chronological order of events in a narrative the Formalists termed *fabula* (frequently translated as 'story'); the artistic rearrangement of events for purposes of narration they called *siuzhet* (usually rendered as 'plot'). As with the distinction between practical and poetic language, here too the Formalists cast matters in terms of an opposition between the non-literary (*fabula*) and the literary (*siuzhet*).

2 Literariness

The Formalists' emphasis on the device and their attention to arrangement and selection of verbal elements were an attempt to account for what Jakobson termed the *literariness* of the work. Message was necessarily affected by presentation, so that a work's literariness was an integral and inseparable part of its message or content. Instead of seeing form as a covering imposed on a pre-existing content, the Formalists rejected the dichotomy of content and form, seeing words, syntax and intonation as simultaneously both content and form. The literary work consisted exclusively of formed content.

Formalist theory and practice were part of a strident literary debate with Russian Marxist critics, and many Formalist statements were deliberately polemical and one-sided. Formalism shared its polemical stance and emphasis on form with Russian Futurist poetry, which emphasized sounds in isolation from meaning; the 'self-valuable' word; neologisms; and shock value. Shklovskii's widely publicized statement, 'art is the sum of its devices', was an example of the polemical nature of Formalist utterances, as were his assessment of Laurence Sterne's *Tristram Shandy* as the most typical novel of world literature, and his categorical statement that a literary work is nothing but form.

In spite of such catchwords and slogans, the Formalist emphasis on form in no way dismissed meaning. Eikhenbaum's study of verse intonation in his book *Melodika russkogo liricheskogo stikha (Melody in Russian Lyric Verse)* (1922) examined the interaction of lexical meaning and syntax in dozens of poems, showing that literary interpretation rested on the blend of the semantic and the formal. Formalism sought to displace the usual explanation of content as a reflection of the non-literary world, replacing it with an approach to content as a literary component of a work. What the Formalists objected to (like the American New Critics after them) was the attempt to paraphrase a work's content, extracting it from the verbal blend. The problem with the content-form dichotomy was that it implicitly equated content with meaning.

With the distinction between poetic and practical language in place, the Formalists began to examine the workings of poetic language in specific literary texts. This middle stage of Formalism, lasting from about 1919 to 1924, treated the literary work as a 'dynamic system' in which contrasting and often conflicting features competed for primacy. Instead of matching each other and forming a congruous whole, elements such as rhythm, syntax and intonation were shown often to be at odds, forcing concessions from one another in a struggle for dominance. At the heart of poetic language the Formalists saw not harmony but dissonance, incongruity and struggle. This followed logically from their premise of automatization: the need for poetic language constantly to renew and defamiliarize meant the creation of new, unexpected combinations of linguistic and semantic material. Borrowing from the German aesthetician Broder Christiansen the concept of the *dominanta*, Tynianov in *Problemy*

stikhotvornogo iazyka (Problems of Verse Language) (1924) theorized that the prevailing element in this struggle deformed the others. Tynianov introduced the term ‘constructive principle’ to refer to the relationship between the *dominanta* and the other elements in a work. Shklovskii’s catch-phrase ‘art is the sum of its devices’ was now shown to be an oversimplification: art was far more than the mere *sum* of its parts, because the meaning of each part was dependent on the whole context created by the work. Each element served to place the others in relief.

To view the literary work as a ‘system’ was to see it as a self-created world rather than an imitation of outside reality or referent. In this the formalist concept of the literary work paralleled the work of Ferdinand De Saussure (§2) on the sign. Saussure’s sign, which consisted of mental concept (signified) and word (signifier), took no account of the referent, which was completely outside the sign. Similarly, for the Formalists ‘content’, like referent, was completely outside the literary work, which consisted of aestheticized meaning - the interaction of all components of the literary work.

Since the goal of poetic language was to renew familiar, automatized language, and since what was familiar changed over time, it followed that what constituted poetic language in one epoch would cease to be poetic in the next. Poetic language itself was susceptible to automatization through repeated usage in different works, thereby losing its poeticity. Writers trying to overcome the automatized conventions of their literary ‘fathers’ would turn to works by their ‘grandfathers’ or ‘uncles’ for devices that might revitalize exhausted genres. In ‘Noveishaia russkaia poéziia’ (Modern Russian Poetry) (1919), Jakobson stated that new literary forms arise to replace previous literary forms that have become exhausted. This position, for which the Formalists found precedent in the work of the nineteenth-century Russian ethnographer Aleksandr Veselovskii, contrasted with the widely held view of the nineteenth-century civic critics and the Bolsheviks that new forms in literature are needed to express a new content, such as a change in the structure of society.

3 Literary evolution

In building change into its model of poetic language, Russian literary Formalism differed from theories that applied a formulaic description to all works of a given genre. Poetic language was by definition *new* language. In ‘O literaturnoi ėvoliutsii’ (On Literary Evolution) (1927), Tynianov described this process of renewal and change in poetic language as a four-step cycle of literary evolution: (1) a new constructive principle arises to replace the previously dominant one that has finally become automatized; (2) the new constructive principle gains currency in new literary works; (3) it becomes more widespread; (4) it becomes automatized and evokes opposing constructive principles.

The concept of a constructive principle within a dynamic system whose elements are constantly colliding allowed the Formalists to challenge accepted interpretations of major literary works, especially those widely characterized as examples of realism. Eikhenbaum’s seminal article on Gogol’’s short story ‘The Overcoat’ called into question the consensual interpretation of that story as a portrayal of the humble man disdained by his colleagues and superiors in an inhumane society. Eikhenbaum argued that the ‘realist’ passages, including the famous words ‘I am your brother’, universally interpreted as a plea for recognition of human dignity, constitute only one strand of Gogol’’s narrative, and that they are undercut by another, comic strand woven out of sheer linguistic play. When a realist narrative is placed in the context of a linguistic game, Eikhenbaum argued, its meaning is altered.

Eikhenbaum’s article reflected the Formalist bias that realistic mimesis was not the main business of literature. Having established that each work constituted an aesthetic system, the Formalists now had an approach to literature that could be offered as an alternative to realist readings. A system of interacting elements meant that the function of a device could change depending on its context. Furthermore, the workings of this dynamic system, the demands of artistic presentation, worked against an accurate reflection of the world. Aesthetic requirements conflicted with factual accuracy; literary works on historical themes might need to sacrifice historical accuracy in the interests of aesthetics.

4 Autonomy of literature

The differentiation between practical and poetic language, as well as the notion of system, allowed the Formalists to claim that literature was an autonomous activity not chained to economic, social or political reality. The Formalists thus liberated literary studies from the mimetic orientation that sees the work as a reflection of the

world. Such harnessing of the literary work to the outside world had been the practice of mainstream Russian literary criticism. It characterized the radical civic critics of the mid-nineteenth century, who pressed literature into the service of social justice; the academic critics of the cultural-historical and cultural-psychological schools; and also the Marxist critics who gained legitimacy after the Bolshevik Revolution and with whom the Formalists engaged in bitter polemics. Russian literary Formalism set out to make literary criticism into a scholarly discipline by constructing an 'objective' theory of literature, centred on the work itself, that would not subordinate the study of literature to other disciplines. Most of the Formalists' theoretical concepts, such as their early emphasis on sound and the palpability of the word, pointed to the work itself rather than to the cultural context that produced it.

The case for literary autonomy was a major achievement of Formalism while simultaneously containing the germ of its demise. The notion on which autonomy rested - that of a poetic language whose definition lay in the revitalization of automatized forms - accounted merely for the fact of literary change, not for the actual direction of change at any given time. By the mid-1920s, it became clear that the automatization/revitalization dynamic, while at least partly valid, left too many questions unanswered, and that the answers were to be found in the non-literary contexts that Formalism had dismissed as irrelevant to literariness and aesthetics. Within the confines of the purely work-centred poetics, no further insights were possible.

In an attempt to incorporate some of these contexts without sacrificing the literary autonomy for which Formalism had fought so hard, Tynianov proposed a modification of Formalist theory. His article of 1927, 'On Literary Evolution', put forth a model of culture presenting its various aspects - economic, social, political, religious, linguistic, literary - as parallel, autonomously developing lines, or 'series'. Tynianov's model preserved the Formalist premise of literary autonomy, since each series had its own immanent development that was not directly affected or 'caused' by any other series. Indirect influences from other series, however, could occur, passing vertically from one series to the next as though through a porous membrane. The most significant influence, Tynianov suggested, would come from the neighbouring series; influences from more distant series, which would have to pass through the intervening series, would be modified along the way and more difficult, if not impossible, to trace. Initially Tynianov maintained that the literary series had as its closest neighbour the linguistic series; subsequently, in a collaborative article of 1928, 'Problems in the Study of Literature and Language', Tynianov and Jakobson proposed that the series change their positions, so that in different eras the literary series might border on the economic, the social or the political.

At the same time that the model of the parallel series was intended to reinvigorate the increasingly repetitive practice of Formalist criticism by admitting a select amount of extra-literary information into literary interpretation, it also challenged the Marxist model of economic base and cultural superstructure (see Marx, K §8). Tynianov's model, in providing for immanent development within each series, conferred on literature and all other activities an autonomy that was incompatible with the notion of a base reflected in a superstructure.

With the rigid controls imposed on literature, philosophy and theory at the close of the 1920s when Stalin consolidated his power, the Formalists had to abandon the attempt to open their literary theory to include extra-literary factors. Those Formalists who remained in Russia by and large omitted theoretical issues from their subsequent writings. The tendency of post-structuralist literary theories to place literary works in their broader cultural context addresses an omission of Formalism; but these same critical schools have internalized the founding Formalist assumption that the construction of the linguistic/literary utterance is central to meaning.

See also: Deconstruction §1; Semiotics; Structuralism; Structuralism in literary theory

CAROL ANY

References and further reading

Any, C. (1994) *Boris Eikhenbaum: Voices of a Russian Formalist*, Stanford, CA: Stanford University Press.(See ch. 3 for analysis of Formalist theory.)

Belyi, A. (1910) *Simvolizm: Kniga statei (Symbolism: A Book of Articles)*, Moscow: Musaget.(Referred to in §1.)

Bennett, T. (1979) *Formalism and Marxism*, London: Methuen.(Examines Formalism with respect to Saussure, Bakhtin, Althusser and post-Althusserians.)

Eikhenbaum, B. (1922) *Melodika russkogo liricheskogo stikha (Melody in Russian Lyric Verse)*, St Petersburg: Opoiaz.(Contrasts the melodic intonational structure of poems by Zhukovskii and Lermontov with unmelodic

poems of Pushkin, Tiutchev and Fet.)

Eikhenbaum, B. (1919) ‘Kak sdelana "Shinel’" Gogolia’; trans. R. Maguire, ‘How Gogol’s "Overcoat" Is Made’, in R. Maguire (ed.), *Gogol from the Twentieth Century: Eleven Essays*, Princeton, NJ: Princeton University Press, 1974, 269-92.(The classic example of Russian Formalism, showing how syntax and word play can undermine conventional semantics.)

Erlich, V. (1981) *Russian Formalism: History - Doctrine*, New Haven, CT: Yale University Press.(Fourth edn of Erlich’s definitive 1955 work on Russian Formalism, with full bibliography.)

Gorman, D. (1992) ‘A Bibliography of Russian Formalism in English’, in *Style* 26 (4): 554-76.(Thorough listing of anthologies, works by individual Formalists, and works about Formalism.)

Hansen-Love, A. (1978) *Der russische Formalismus*, Vienna: Akademie der Wissenschaften.(Written after Erlich’s study and assumes greater knowledge of Formalism.)

Iakubinskii, L. (1916) ‘O zvukakh stikhotvornogo iazyka’ (On the Sounds of Verse Language), in *Sborniki po teorii poėticheskogo iazyka (Collected Articles on the Theory of Poetic Language)*, Petrograd: Opoiaz, vol. 1. (Referred to in §1.)

Jakobson, R. (1919) ‘Noveishaia russkaia poėziia’; trans. E. Brown, ‘Modern Russian Poetry: Velemir Khlebnikov’, in E.J. Brown (ed.) *Major Soviet Writers: Essays in Criticism*, New York, NY: Oxford University Press, 1973, 58-82, 413-14.(A Formalist study of the Futurist poetry of Velemir Khlebnikov.)

Jakobson, R. and Tynianov, Iu. (1928) ‘Problemy izucheniia literatury i iazyka’; trans. H. Eagle, ‘Problems in the Study of Literature and Language’, in L. Matejka and K. Pomorska, eds, *Readings in Russian Poetics: Formalist and Structuralist Views*, Cambridge, MA: MIT Press, 1971, 79-81.(Written when Russian Formalism had exhausted its possibilities, this brief outline suggests new directions for literary theory.)

Jameson, F. (1972) *The Prison-House of Language: A Critical Account of Structuralism and Russian Formalism*, Princeton, NJ: Princeton University Press.(Theoretical analysis for knowledgeable readers.)

Shklovskii, V. [Shklovsky] (1914) ‘Voskreshenie slova’; trans. R. Sherwood, ‘The Resurrection of the Word’, in S. Bann and J. Bowlt, (eds), *Russian Formalism: A Collection of Articles and Texts in Translation*, Edinburgh: Scottish Academic Press, 1973, 41-7.(The first work of Russian Formalism argues that a poet’s task is to make us perceive familiar words in a new way.)

Shklovskii, V. [Shklovsky] (1916) ‘Iskusstvo kak priëm’; trans. B. Sher, ‘Art as Device’, in V. Shlovskii [Shklovsky], *Theory of Prose*, Elmwood Park: Dalkey Archive, 1990.(How literary works renew conventional perceptions.)

Shukman, A. (1977) ‘A Bibliography of Translations of Formalist Writings’, in *Russian Poetics in Translation*, vol. 4, *Formalist Theory*, eds L. O’Toole and A. Shukman, 100-8.(123 titles in English, French, German and Italian. Includes a category on ’Bakhtin and his School’.)

Steiner, P. (1984) *Russian Formalism: A Metapoetics*, Ithaca, NY: Cornell University Press.(Argues for metaphorical basis of Formalist theory, which variously approaches literature as machine, organism or system.)

Striedter, J. (1989) *Literary Structure, Evolution, and Value: Russian Formalism and Czech Structuralism Reconsidered*, Cambridge, MA: Harvard University Press.(Ch. 1 explicates Formalist theory of prose and evolution; ch. 2 examines relationship of Russian Formalism with the Prague School.)

Tynianov, Iu. (1924) *Problemy stikhotvornogo iazyka*; trans. M. Sosa and B. Harvey, *Problems of Verse Language*, Ann Arbor, MI: Ardis, 1981.(Distinguishes literary criteria for poetry versus prose.)

Tynianov, Iu. (1927) ‘O literaturnoi ėvoliutsii’; trans. C. Luplow, ‘On Literary Evolution’, in L. Matejka and K. Pomorska, (eds), *Readings in Russian Poetics: Formalist and Structuralist Views*, Cambridge, MA: MIT Press, 66-78.(Theoretical discussion of how new literary genres eventually become automatized.)

Russian Materialism: ‘the 1860s’

No tradition of philosophical materialism existed in Russia until the years conventionally called ‘the 1860s’ - roughly, the period from the death of Tsar Nicholas I in 1855 to the attempted assassination of Tsar Alexander II in 1866. During that time philosophical freethinking, under the delayed influence of the French Enlightenment and the contemporaneous influence of post-Hegelian German materialism, came together with political radicalism to create a major social and intellectual movement with a broadly materialist philosophical foundation.

The theoretical underpinnings of the movement were elaborated in Russia (as far as tsarist censorship would permit) by Nikolai Chernyshevskii, Dmitrii Pisarev, Nikolai Dobroliubov, Ivan Sechenov and others, and more freely in emigration by Mikhail Bakunin. Their ‘materialism’ was less a precisely articulated ontological position than a grand, science-worshipping worldview that sought to undermine both religion and the state; its elements included naturalism and universal causal determinism in metaphysics, empiricism in epistemology, reductionism in the philosophy of mind, ‘rational egoism’ in ethics, revolutionary socialism in political philosophy and realism in aesthetics. Because of their extreme opposition to established authority and traditional values, the representatives of this movement came to be called ‘nihilists’, and under that name they were portrayed in best-selling novels of the day by Ivan Turgenev and Fëdor Dostoevskii.

Government repression after 1866 put an end to the open development of this materialist movement, but the writings of its leaders proved to be an inspiration to Georgii Plekhanov and Vladimir Lenin, the founders of Russian Marxism. Under Communism, the materialists of ‘the 1860s’ were honoured in Russia as great philosophers and important precursors of Marx.

1 Intellectual lineage

Although Soviet Marxist historians of philosophy laboured for decades to find a native materialist tradition in Russian philosophy extending back to the eighteenth century, no such tradition is discernible until the mid-nineteenth. The influence of the French Enlightenment, encouraged in Russia for a time by Catherine the Great, can be seen in the thought of Aleksandr Radishchev (1749-1802) and the ‘Russian Voltairians’, but none of them actually advanced a materialist ontology or significantly influenced the later Russian thinkers who did, and the same is true of other alleged early ‘materialists’ such as Mikhail Lomonosov (1711-65) and the theoreticians of the abortive Decembrist uprising of 1825. The eighteenth-century Enlightenment affected only a small stratum of Russian society and did little to change the overall religious, idealistic direction of Russian philosophical thought - a direction that was reinforced with a vengeance by the repressive measures of Nicholas I, the ‘Tsar-disciplinarian’, upon his accession to the throne in 1825 (see Enlightenment, Russian).

Materialism came to Russia in the nineteenth century as it had come to Germany - as a reaction against German Idealism; and in both countries the trend was initiated by Ludwig Feuerbach. Among the liberally minded, Western-oriented Russian intelligentsia, brief but intense infatuations with Schelling, Hegel and Fichte (see Schellingianism; Hegelianism, Russian; Fichte, J.G.) were followed by enthusiasm for Feuerbach, whose views became known in Russia shortly after the publication of his *Das Wesen des Christentums (Essence of Christianity)* in 1841, despite official prohibition of his writings. But the materialist trend in philosophy that Feuerbach inspired did not become a broad movement in Russia until the death of Nicholas I in 1855 and the end of the Crimean War a year later. When it did, its tone was set more by the German ‘scientific materialists’ who had succeeded Feuerbach - Ludwig Büchner, Jacob Moleschott and Karl Vogt (see Materialism §3).

The bitter experiences of the Crimean War persuaded many Russians of the need for radical social reform, and the accession of Alexander II offered hope of movement in that direction. Censorship was eased somewhat, and as a consequence new books were available and new journals appeared. A vigorous progressive spirit arose, leading in 1861 to the long-sought emancipation of the serfs. Philosophically this new spirit was manifested in a secular, empiricist attitude that put its trust in reason and science to sweep away the accumulated evils of the past and set society on the path of progress. It was the Enlightenment revisited. Indeed, since Russia’s eighteenth-century Enlightenment was so shallow and limited that Nicholas could easily obliterate it, this was in effect the true Russian Enlightenment. Leon Trotsky called ‘the 1860s’ ‘our short-lived eighteenth century’ (see Lampert 1965: 6).

If the German materialists were the Diderot and Holbach of the delayed Russian Enlightenment, Büchner’s notorious *Kraft und Stoff (Force and Matter)* (first published in German in 1855) was its *Encyclopédie*. Called ‘the Bible of the German materialism of the present day’ by the historian of philosophy Friedrich Ueberweg, the book was no less a sensation in Russia than in Germany. Like the works of Feuerbach, it was never published legally in pre-Soviet Russia; but beginning about 1859, Russian university students became *samizdat* pioneers in the translation and reproduction of prohibited books. It was *Kraft und Stoff* that the ‘nihilist’ younger generation in Ivan Turgenev’s novel *Otsy i deti (Fathers and Sons)* (1862) recommended to the ‘fathers’.

These Russian materialists made little distinction between the views of Feuerbach and the ‘scientific materialists’, whose impact on most of them was concurrent. Like Feuerbach they often referred to themselves as ‘realists’ rather than ‘materialists’, but this was at least in part owing to the fact that the typical Russian censor found ‘realism’ less alarming than ‘materialism’. They appear to have regarded the two terms as synonymous, and they looked upon the intellectual movement from Feuerbach to Büchner, Moleschott and Vogt not (as some critics did) as a degeneration of philosophical materialism but as the productive elaboration of a common philosophical core. Chernyshevskii singled out Feuerbach for special praise, but he never explicitly criticized the later thinkers or pointed to differences between their outlook and Feuerbach’s. Pisarev enthusiastically endorsed the views of Büchner and company, but without rejecting Feuerbach. Among Russians who developed materialist views in ‘the 1860s’, only Bakunin (writing abroad, apart from the active domestic scene) expressed reservations about the scientistic extremes of materialist thought.

The philosophical manifesto of ‘1860s’ materialism - Chernyshevskii’s essay ‘Antropologicheskii printsip v filosofii’ (‘The Anthropological Principle in Philosophy’) (1860) - is sometimes considered a Russian popularization of the ideas of Feuerbach. Much of its content is indeed reminiscent of Feuerbach, but it is also reminiscent of Büchner, and in fact Chernyshevskii’s central idea of the organic connection between all human activity and an integral, physical human nature was already expounded in his diary in 1848, before he had read Feuerbach, not to mention Büchner. Others have pointed to La Mettrie and Holbach (especially the latter’s *Système de la nature*) as Chernyshevskii’s sources. In the end, specific influences on particular works of Chernyshevskii or the other Russian materialists cannot be pinpointed. All of them were stimulated by Feuerbach, impressed by the further advances of the German ‘scientific materialists’, influenced directly or indirectly by various thinkers of the French Enlightenment (Chernyshevskii, for one, was fascinated by the figure and the thought of Rousseau); but they were also guided by their own reactions both to modern European philosophy and to the social and intellectual realities of their immediate Russian environment.

2 Philosophical content

The actual content of the ‘materialism’ preached by the radicals of ‘the 1860s’ is not always clear. As indicated, they often avoided the term itself for reasons of censorship, and when they used it they often did so without delineation of its scope. At the same time, they did subscribe to a loose constellation of ideas that can be identified as a materialist philosophical programme. Its fullest legal expression in Russia came in the published writings of Chernyshevskii, Sechenov and Pisarev (especially Chernyshevskii’s ‘Antropologicheskii printsip v filosofii’); more explicit presentation of these ideas may be found in some unpublished writings such as a series of letters Chernyshevskii wrote from his Siberian exile and a number of essays by Bakunin published abroad.

The clearest and most succinct expression of the materialists’ ontology was provided by Chernyshevskii in a letter from Siberia: ‘That which exists is matter. Matter has properties. The manifestations of properties are forces. What we call laws of nature are the modes of action of forces’ (see Scanlan 1970: 78). With this basic outlook and with its most immediate implications, all the Russian materialists of ‘the 1860s’ would agree. Those implications include, first, the ontological thesis that nothing immaterial, such as a supernatural supreme being or an immortal spiritual soul, exists; second, the epistemological thesis that only empirical investigation, not faith or pure reason, can provide knowledge; and third, the reductionist thesis that all human functions or ‘forces’ (including of course what we call mental operations) are properties of matter.

Because outright denial of God’s existence or rejection of faith as a source of knowledge could not readily be camouflaged to avoid censorship, the attention of Chernyshevskii, Sechenov and others in their works published in Russia in ‘the 1860s’ was directed primarily towards establishing the reductionist thesis - that is, towards

demonstrating that all human functions, from the most 'animal' to the most refined, are materially based and can be exhaustively comprehended by the natural sciences. Chernyshevskii's apparent rambling in 'Antropologicheskii printsip v filosofii' is actually a sustained defence of this thesis, following two complementary lines of argument: that other animals possess all the traits sometimes thought to be restricted to man, and that human behaviour is explicable solely and completely in the same physiological and ultimately chemical terms used to explain processes in the 'lower' orders of nature (Chernyshevskii implies, for example, that because thinking can be reduced to processes of sensation, which is a form of assimilation comparable to digestion, thinking is a chemical process like digestion). The physiologist Sechenov's contribution to this argument is evident from the alleged subtitle (struck out by the censors) of his work *Refleksy golovnogo mazga (Cerebral Reflexes)* (1863): 'An Attempt to Reduce the Mode of Genesis of Psychical Phenomena to Physiological Bases'. Pisarev contributed some of the more imaginative applications of the reductionist thesis, such as his contention that the intellectual vitality of eighteenth-century Europe was attributable to the introduction of the stimulants tea and coffee.

A corollary of the reductionist approach is that the laws of nature apply to acts of the human will just as they do to all other natural 'forces', and this determinist thesis is vigorously championed by Chernyshevskii, Sechenov and Bakunin. Pisarev, on the other hand, offers an inconsistent mixture of materialistic reductionism and 'subjectivity': he attributes an undefined but supremely important 'individual autonomy' to human beings. Bakunin, too, is concerned to retain a place for human freedom in a deterministic universe, but he does so by locating 'freedom' in the conscious awareness that one is determined *only* by natural laws and not by the will of another; thus 'materialism denies free will and ends in the establishment of liberty' (Bakunin [1870-1] 1970: 48).

Their determinism did not prevent most of the materialists of 'the 1860s' from espousing an ethical theory, despite their expressed abhorrence of 'morality' and 'ideals'. Chernyshevskii and Pisarev, in particular, advanced a form of ethical egoism, holding that people ought to act in accordance with their own best interests or true needs. This theory was said to be 'scientific' because it was based on the elementary truth of *psychological* egoism - that is, on the supposed 'natural law' that people invariably *do* act in accordance with what they *think* is their own best interest, with their *perception* of their needs. The task, then, is simply to bring perceptions of needs into line with genuine needs and provide social arrangements for their satisfaction; once people are shown what their true needs really are, and are allowed to pursue them, psychological egoism guarantees that they will do so. The materialists were not worried that a society of perfect egoists would be chaotic or conflictual, because they were convinced that the true, 'natural' needs of all people are harmonious and mutually reinforcing. Hence 'rational egoism', as they called it, would coincide in practice with altruism.

In its most general terms, the social ideal of all these thinkers, including Bakunin, was a society consisting of institutions that allowed the expression of 'natural' human needs rather than generating 'false' or 'artificial' needs such as acquisitiveness and hostility. Because existing institutions were so distant from this ideal, so 'unnatural', an initial posture of revolutionary destruction was called for, expressed most vehemently by Pisarev when he wrote: 'What can be smashed should be smashed. What withstands the blow is fit to survive. What flies into pieces is rubbish. In any case, strike out right and left; no harm can come of it' (see Edie *et al.* 1965: vol. 2, 65). The exact contours of what was expected to survive were never fully elaborated, though it was widely assumed that the society attuned to people's true needs would be secular, socialist and non-authoritarian (to the point of anarchism, in the case of Bakunin).

An important place in the materialists' worldview was occupied by their philosophy of art. This was the first element of their outlook to receive broad public attention after the publication of Chernyshevskii's essay 'The Aesthetic Relation of Art to Reality' in 1855. In opposition to the prevailing idealist theories, Chernyshevskii argued that the purpose of art is simply to reproduce reality critically - that is, to portray what exists in such a way as to explain it, identify its shortcomings and project a more desirable future, all in the service of true human needs. The notion that art is subordinate to life and is justified only to the extent that it contributes to social progress was carried to an extreme by Pisarev, who under the banner of 'the destruction of aesthetics' rejected almost all art of the past on the ground that it catered to false, artificial needs.

3 Sociocultural significance

The Russian materialism of 'the 1860s' is less important for the philosophical content of its ideas than for their influence on Russian society and culture. Materialism in Russia, as it was in France in the eighteenth century and

in Germany around 1848, was closely allied with the revolutionary movement. It opposed, in the words of T.G. Masaryk, both 'theology and theocracy' - both God and the autocratic state ruled in God's name (Masaryk [1913] 1919 vol. 2: 48). It offered not merely an alternative philosophical perspective but a system of values and purposes that were radically different from traditional attitudes and that required the sweeping rejection of existing institutions. The challenge to construct a new life both personally and socially was taken up by throngs of restive Russians, most of them young and many with little interest in philosophy. Called 'nihilists' because of their seemingly total denial of standards of thought and behaviour, they were in fact dogmatically committed to their own antireligious, antitraditionalist and anti-authoritarian standards.

By 1861 the radicals were disappointed by the slow pace of reform, and especially by the illiberal terms of the emancipation of the serfs in that year. Disturbances were answered with intensified repression; both Chernyshevskii and Pisarev were arrested and imprisoned in 1862. Earlier hopes of change from above grew into a determination to bring change about by force, and a 'nihilist' student, Dmitrii Karakozov, effectively ended the relatively open decade of 'the 1860s' by his unsuccessful attempt on the life of Alexander II in 1866. Thereafter, Russian materialism evolved into a philosophy of revolutionary activism, soon with Marxist overtones. A disciple of Chernyshevskii named Pëtr Nikitich Tkachëv (1844-86 NS), one of the first Russian writers to come under the influence of Marx, brought elements of Marxist historical materialism into his literary criticism in the mid-1860s; subsequently, writing abroad, Tkachëv became an extreme advocate of revolutionary terrorism and the violent seizure of power.

Among the enduring cultural influences of the Russian materialists' worldview was its impact in the field of literature. The Russian tradition of 'civic criticism', inaugurated by Vissarion Belinskii, was developed further by Chernyshevskii, Pisarev, Dobroliubov and others, in part because the discussion of literature offered them a relatively protected forum for the social critique they could not publish directly. Creative literature, too, felt the powerful impact of 'nihilism'. The very term was popularized by Turgenev in *Otsy i deti (Fathers and Sons)*, in which the character Bazarov is the quintessential '1860s' materialist, bent on saving the world through studying the physiology of frogs. A more sympathetic, if aesthetically less successful, portrayal of the 'new people' of the day was presented in Chernyshevskii's own widely read novel, *Chto Delat'? (What Is to Be Done?)* (1863). The deepest literary impact of the movement, albeit a negative one, came in the works of Fëdor Dostoevskii. His *Zapiski iz podpol'ia (Notes From Underground)* (1864) is a frontal assault on the determinism and 'rational egoism' of Chernyshevskii, and all of Dostoevskii's great novels deal in one way or another with the materialist outlook and what Dostoevskii took to be its pernicious implications; in *Besy (The Possessed)* (1871-2) he paints a highly unflattering portrait of 'nihilist' revolutionaries.

Although the writings of the materialists of 'the 1860s' inspired one of the liveliest intellectual debates in nineteenth-century Russia, the debate did not result in the philosophical refinement of their materialist outlook (except to some degree in Chernyshevskii's unpublished and publicly unknown letters from Siberia of 1876-8). In 1861 and 1862, the religious philosopher Pamfil Danilovich Iurkevich (1826-74) published a series of essays containing careful and philosophically sophisticated criticism of materialism and especially of Chernyshevskii's 'Antropologicheskii printsip v filosofii'. Because of censorship, however, Chernyshevskii could not reply substantively to the charges against him; his side of the philosophical argument was effectively proscribed, and he and his supporters resorted to invective and crude personal attacks on Iurkevich. The only strictly philosophical legacy of the materialists came in the form of their influence on Russian Marxism. Georgii Plekhanov and Vladimir Lenin, the two thinkers most responsible for the development of Marxism in Russia, credited Chernyshevskii with having, respectively, 'massive' and 'overwhelming' influence on them (see Scanlan 1970: 65). During the communist period of Russian history, the principal 'nihilist' theoreticians were officially lionized under the designation 'Russian revolutionary democrats' and were called the most important materialist thinkers in the history of philosophy before Marx.

See also: Nihilism, Russian

JAMES P. SCANLAN

References and further reading

Bakunin, M. (1870-1) *God and the State*, ed. P. Avrich, New York: Dover, 1970.(Referred to in §2. The most popular of Bakunin's writings, a lively statement of his worldview.)

Büchner, L. (1855) *Kraft und Stoff (Force and Matter)*, Frankfurt.(Referred to in §1. A presentation, highly influential in Russia, of extreme reductive materialism.)

Chernyshevskii, N. (1855) *Esteticheskie otnosheniia iskusstva k deistvitel’nosti (The Aesthetic Relation of Art to Reality)*, St Petersburg.(Referred to in §2. Chernyshevskii’s statement of a realist philosophy of art.)

Chernyshevskii, N. (1860) ‘Antropolgicheskii printsip v filosofii’ (‘The Anthropological Principle in Philosophy’), in *Sovremennik*, vol. 4, 329-36, vol. 5, 1-46.(Referred to in §§1-3. Chernyshevskii’s principal philosophical essay.)

Edie, J.M. *et al*. (eds) (1965) *Russian Philosophy*, Chicago, IL: Quadrangle Books, 3 vols.(Referred to in §2. A comprehensive anthology of Russian philosophy containing extensive selections from the writings of the thinkers discussed in this entry, as well as bibliographies of their works.)

Feuerbach, L. (1841) *Das Wesen des Christentums*, trans. M. Evans (George Eliot) as *Essence of Christianity*, 1854; new edn, intro. K. Barth, foreword H.R. Niebuhr, New York: Harper & Row, 1957.(Feuerbach’s most influential work, referred to in §1.)

Lampert, E. (1965) *Sons against Fathers: Studies in Russian Radicalism and Revolution*, Oxford: Clarendon Press.(Referred to in §1. A highly readable and thorough study of the Nihilist movement, with major chapters on Chernyshevskii, Pisarev and Dobroliubov.)

Masaryk, T.G. (1913) *The Spirit of Russia: Studies in History, Literature and Philosophy*, trans. E. and C. Paul, 2 vols, London: Allen & Unwin, 1919.(Referred to in §3. An older but still useful intellectual history of Russia with a lengthy chapter on ‘Realism and Nihilism’ (vol. 2, 1-114).)

Scanlan, J.P. (1970) ‘Nicholas Chernyshevsky and Philosophical Materialism in Russia’, in Journal of the History of Ideas 8 (1): 65-86.(Referred to in §§2-3. Expansion of many of the points in this entry.)

Sechenov, I. (1863) *Refleksy golovnogo mazga (Cerebral Reflexes)*, Moscow.(Referred to in §2. A Russian physiologist’s influential statement of reductive materialism.)

Russian philosophy

Russian thought is best approached without fixed preconceptions about the nature and proper boundaries of philosophy. Conditions of extreme political oppression and economic backwardness are not conducive to the flowering of philosophy as a purely theoretical discipline; academic philosophy was hence a latecomer on the Russian scene, and those (such as the Neo-Kantians of the end of the nineteenth century: see Neo-Kantianism, Russian) who devoted themselves to questions of ontology and epistemology were widely condemned for their failure to address the country's pressing social problems. Since Peter the Great's project of Westernization, Russian philosophy has been primarily the creation of writers and critics who derived their ideals and values from European sources and focused on ethics, social theory and the philosophy of history, in the belief that (as Marx put it in the first 'Thesis on Feuerbach') philosophers had hitherto merely interpreted the world: the task was now to change it. This passionate social commitment generated much doctrinaire fanaticism, but it also inspired the iconoclastic tendency made philosophically respectable by Nietzsche: the revaluation of values from an ironic outsider's perspective. The principal contribution of Russian thinkers to world culture has so far consisted not in systems, but in experiments in the theory and practice of human emancipation. Some of these led to the Russian Revolution, while others furnished remarkably accurate predictions of the nature of utopia in power. Like Dostoevskii's character Shigalëv who, starting from the ideal of absolute freedom, arrived by a strict logical progression at the necessity of absolute despotism, Russian philosophers have specialized in thinking through (and sometimes acting out) the practical implications of the most seductive visions of liberty that Europe has produced over the last 200 hundred years.

1 The development of Russian philosophy

What Berdiaev called the 'Russian Idea' - the eschatological quest that is the most distinctive feature of Russian philosophy - can be explained in terms of Russian history. The Mongol yoke from the twelfth to the fourteenth century cut Russia off from Byzantium (from which it had received Christianity) and from Europe: it had no part in the ferment of the Renaissance. Its rise as a unified state under the Moscow Tsardom followed closely on the fall of the Orthodox Byzantine Empire, and the emerging sense of Russian national identity incorporated a messianic element in the form of the monk Philotheus' theory of Moscow as the 'Third Rome', successor to Rome and Constantinople as guardian of Christ's truth in its purity (see Medieval philosophy, Russian). 'There will not be a fourth', ran the prophecy: the Russian Empire would last until the end of the world. Russian thought remained dominated by the Greek patristic tradition until the eighteenth century, when the Kievan thinker Skovoroda (sometimes described as Russia's first philosopher) developed a religious vision based on a synthesis of ancient and patristic thought. He had no following; by the mid-century Russia's intellectual centre was St Petersburg, where Catherine the Great, building on the achievements of her predecessor Peter, sought to promote a Western secular culture among the educated elite with the aid of French Enlightenment ideas. But representatives of the 'Russian Enlightenment' were severely punished when they dared to cite the *philosophes'* concepts of rationality and justice in criticism of the political status quo (see Enlightenment, Russian). The persecution of advanced ideas (which served to strengthen the nascent intelligentsia's self-image as the cultural and moral leaders of their society) reached its height under Nicolas I (1825-55), when philosophy departments were closed in the universities, and thought went underground. Western ideas were the subject of intense debate in small informal circles of students, writers and critics, the most famous of which in Moscow and St Petersburg furnished the philosophical education of such intellectual leaders as the future socialists Herzen and Bakunin, the novelist and liberal Ivan Turgenev, the literary critic Belinskii (from whose 'social criticism' Soviet Socialist Realism claimed descent), and the future Slavophile religious philosophers Kireevskii and Khomiakov (see Slavophilism). As a critic has noted: 'In the West there is theology and there is philosophy; Russian thought, however, is a third concept'; one which (in the tsarist intellectual underground as in its Soviet successor) embraced novelists, poets, critics, religious and political thinkers - all bound together by their commitment to the goals of freedom and justice.

In the 1830s these beleaguered individuals encountered German Idealism: an event of decisive significance for the future development of Russian thought. The teleological structures of idealist thought provided Russian intellectuals with a redemptive interpretation of their conflicts and struggles as a necessary stage in the dialectical movement of history towards a transcendent state of harmony. Idealism (notably in its Hegelian forms: see

Hegelianism, Russian) left its mark on the vocabulary of subsequent Russian philosophy, but its principal legacy was the belief, shared by the vast majority of Russian thinkers, that an ‘integral worldview’, a coherent and unified vision of the historical process and its goal, was the essential framework both for personal moral development and social theorizing. The question of history’s goal became a matter for intense debate among the intelligentsia with the publication in 1836 of Chaadaev’s ‘Philosophical Letter’, which posed Russia’s relationship to the West as a central philosophical problem, maintaining that Russia’s historical separation from the culture of Western Christianity precluded its participation in the movement of history towards the establishment of a universal Christian society. Chaadaev’s version of the march of progress was much indebted to French Catholic conservatism, while the nationalist riposte to his ideas drew heavily on the Romantics’ critique of the Age of Reason and Schelling’s organic conception of nationhood: the Slavophiles held that Western culture was in a state of terminal moral and social decline, suffering from an excess of rationalism, which had led to social atomization and the fragmentation of the individual psyche (see Chaadaev, P.I.; Schellingianism). These divisions could be healed only by religious faith in its purest form, Russian Orthodoxy, whose spirit of organic ‘togetherness’, uncontaminated by Western rationalism, they presented as a model for Russian society and a beacon for mankind. They thereby laid the foundations of a distinctively Russian tradition of cultural and religious messianism which includes Dostoevskii’s political writings, the Pan-Slavist and Eurasian movements (see Dostoevskii, F.M.; Pan-Slavism and Eurasian movement), and the apocalyptic vision of Berdiaev, whose philosophy was highly popular among the Soviet underground.

Secular and Westernist thinkers tended to be scarcely less messianic in their response to Chaadaev’s pessimism. The first philosophers of Russian liberalism (see Liberalism, Russian) interpreted their country’s past and future development in the light of Hegel’s doctrine of the necessary movement of all human societies towards the incarnation of Reason in the modern constitutional state, while the Russian radical tradition was shaped successively by the eschatological visions of the French utopian socialists, the Young Hegelians and Karl Marx. Herzen defined the distinctive characteristic of Russian radical thought as the ‘implacable spirit of negation’ with which, unrestrained by the European’s deference to the past, it applied itself to the task of freeing mankind from the transcendent authorities invented by religion and philosophy; and the radical populist tradition that he founded argued that the ‘privilege of backwardness’, by permitting Russia to learn both from the achievements and the mistakes of the West, had placed it in the vanguard of mankind’s movement towards liberty.

Russian religious philosophers tended to see themselves as prophets, pointing the way to the regeneration of human societies through the spiritual transformation of individuals. Vladimir Solov’ëv (regarded by many Russians as their greatest philosopher) believed that his country’s mission was to bring into being the Kingdom of God on Earth in the form of a liberal theocracy, which would integrate knowledge and social practice and unite the human race under the spiritual rule of the Pope and the secular rule of the Russian tsar. His metaphysics of ‘All-Unity’ was a dominant force in the revival of religious and idealist philosophy in Russia in the early twentieth century, inspiring an entire generation of thinkers who sought to reinterpret Christian dogma in ways that emphasized the links of spiritual culture and religious faith with institutional and social reform, and progress in all other aspects of human endeavour. Among them were leading Russian émigré philosophers after 1917, such as Semën Frank, Bulgakov (who sought to create a new culture in which Orthodox Christianity would infuse every area of Russian life), Berdiaev (who was strongly influenced by the messianic motifs in Solov’ëv), and Hessen, who offered a Neo-Kantian and Westernist interpretation of the notion of ‘All-Unity’. A number of émigré philosophers (notably Il’in and Vysheslavtsev) interpreted Bolshevism as the expression of a spiritual crisis in modern industrialized cultures. Many blamed the Russian Revolution on infection from a culturally bankrupt West which (echoing the Slavophiles, Dostoevskii and Leont’ev) they presented as corrupted by rationalism, positivism, atheism and self-centred individualism (although few have gone as far as the fiercely polemical Losev who, up until his death in the Soviet Union in 1988, maintained that electric light expressed the spiritual emptiness of ‘Americanism and machine-production’). Most maintained a historiosophical optimism throughout the catastrophes of the first half of the twentieth century, which Berdiaev saw as a precondition for messianic regeneration, while Hessen believed that religious and cultural values would emerge triumphant from the carnage in a dialectical *Aufhebung*.

2 Major themes in Russian philosophy

The main impetus of Russian philosophy has always been towards the future, as its representatives strained to

discern the features of the 'new man' (the term favoured by the left from the 1860s, with the addition of the adjective 'Soviet' after 1917), or the 'integral personality', as Slavophiles and neo-idealists preferred to describe the individual who would one day be free from the cognitive and moral defects that had hitherto prevented mankind from realizing its potential. The nature of these flaws and the specifications of the regenerated human being were the subject of bitter disputes between rival movements. Even on the left, models of the 'new man' varied widely, from the narrow rationalist who was the ideal of the 'nihilists' of the 1860s (see Nihilism, Russian; Russian Materialism: 'the 1860s') and subsequently of Lenin and Plekhanov, to Bakunin's eternal rebel, who would embody the spontaneous spirit of freedom in defiance of all established authorities and orders. At the end of the nineteenth century, in the cultural ferment produced by new movements in philosophy and the arts emanating from the West, radical thinkers began *en masse* to renounce their predominantly rationalist models of the individual and society (see Russian Religious-Philosophical Renaissance). Nietzsche's Superman had a pervasive influence on the ensuing 'revaluation of values', undertaken with the aim of formulating moral and social ideals that would embrace the manysidedness of human creativity (see Nietzsche: impact on Russian thought). Some radical philosophers (such as Berdiaev and Frank), in the process of moving from Marxism to neo-idealism, sought to reconcile Nietzsche's aesthetic immoralism with Christian ethics, while the 'Empiriocriticist' group of Bolsheviks attempted to inject Russian Marxist philosophy with an element of heroic voluntarism by synthesizing it with Nietzschean self-affirmation and the pragmatism of Ernst Mach (see Russian Empiriocriticism). Nietzschean influences combined with the mechanistic scientism of Soviet Marxism in the Soviet model of the 'new man' (whose qualities Lysenko's genetics suggested could be inherited by successive generations). In the post-Stalin 'thaw' some Soviet philosophers, including Il'enkov and Mamardashvili, began a critical rereading of Marx's texts from an anthropocentric standpoint which emphasized the unpredictable and limitless potential of human consciousness (see Marxist philosophy, Russian and Soviet).

This open-ended view of progress (officially encouraged in the Gorbachev period) is uncommon in Russian philosophy, where epistemological scepticism is more often to be encountered in uneasy combinations with eschatological faith. Like other rootless groups, Russian intellectuals were drawn to compensating certainties that seemed capable of resisting their corrosive critique. The radical humanism of much Russian thought placed it at the forefront of the developing critical insistence on the context-dependent nature of truth; but many thinkers who attacked the claims of systems and dogmas to encompass and explain the experience and creative needs of living individuals in specific historical contexts, nevertheless retained a belief in a final, ideal state of being in which the fragmentation of knowledge would be overcome and all human purposes would coincide: a condition for whose principles some looked to science, others to religious revelation. The nihilists, who rejected metaphysics and all that could not be proven by rational and empirical methods, fervently believed that progress would inevitably lead to the restoration of a natural state of harmony between the individual and society. The empiriocriticist movement within Russian Marxism opposed the idolatry of formulas with the claim that experience and practice were the sole criteria of truth, but the group's leading philosopher, Bogdanov, looked forward to a metascience that would unify the fragmented world of knowledge by reducing 'all the discontinuities of our experience to a principle of continuity', predicting that under communism, when all would share the same modes of organizing experience, the phenomenon of individuals with separate mental worlds would cease to exist. Solov'ëv's pervasive influence on subsequent Russian religious idealism owed much to the charms of his vision of 'integral knowledge' and 'integral life' in an 'integral society'. Religious and socialist motifs were combined in some visions of an earthly paradise, such as Bulgakov's 'Christian Socialism', or Gorkii's and Lunarcharskii's creed of 'God-building', which called for worship of the collective humanity of the socialist future. In the revolutionary ferment of the first two decades of the twentieth century many religious and radical philosophers, together with Symbolist writers and poets, envisaged the leap to the harmonious future in apocalyptic terms: the novelist and critic Merezhkovskii prophesied the coming of a 'New Christianity' which would unite Christian faith with pagan self-affirmation in a morality beyond good and evil (see Nietzsche: impact on Russian thought §1; Russian Religious-Philosophical Renaissance §§3-4). In the aftermath of 1917 some thinkers (notably Berdiaev and members of the Eurasian movement) found consolation in apocalyptic fantasies of a new light from the East shining on the ruins of European culture.

Herzen memorably ascribed such doctrinaire utopianism to the Russian tendency to march 'in fearless ranks to the very limit and beyond it, in step with the dialectic, but out of step with the truth'. The most original and subversive Russian thinker, he was the first of a significant minority who directed the iconoclastic thrust of Russian

philosophy against all forms, without exception, of messianic faith. Contending that there was no basis in experience for the belief in a purposeful universe on which the great optimistic systems of the nineteenth century were built, he urged his contemporaries to adapt their categories to the flow of life, to accept (and even welcome) the dominant role of contingency in human existence, on the grounds that individual freedom and responsibility were possible only in an unprogrammed world. Herzen's critique of the claims of metaphysical systems to predict or regulate the course of history was echoed by the 'subjective sociology' developed by Mikhailovskii and Lavrov in opposition to the deterministic scientism of the dominant Russian radical tradition. Tolstoi pointed to the chanciness of life and history in order to demonstrate the inadequacy of all attempts to formulate general rules for human societies; Dostoevskii confronted the systematizers with the lived experience of human freedom as the ability to be unpredictable; in their symposium of 1909 (frequently cited in the West as a pioneering analysis of the psychology of political utopianism) the neo-idealists of the *Signposts* movement explored the ways in which obsession with an ideal future impoverishes and distorts perception of the historical present (see Signposts movement).

Under the Soviet system a few representatives of this anti-utopian tradition ingeniously evaded the pressure on philosophers (backed up by the doctrine of the 'partyness' of truth - see Partiinost') to endorse the official myths of utopia in power. The history of the novel form was the vehicle for Bakhtin's reflections on the 'unfinalizability' of human existence (see Bakhtin, M.M.); similar insights were expressed by the cultural-historical school of psychology established by Vygotskii, who drew on Marx to counter the mechanistic determinism of Soviet Marxist philosophy with a view of consciousness as a cultural artefact capable of self-transcendence and self-renewal. In the 1960s Soviet psychologists and philosophers such as Il'enkov helped to revive an interest in ethics with their emphasis on the individual as the centre of moral agency, while in its historical studies of culture as a system of semiotic signs, the Moscow-Tartu school brought a richly documented and undoctrinaire approach to important moral and political topics.

The insights of some of these individuals and movements into the attractions and delusions of utopian thought are lent added conviction by their own often spectacularly unsuccessful efforts to overcome what Nietzsche called 'the craving for metaphysical comfort'. Tolstoi was torn all his life beween his pluralist vision and his need for dogmatic moral certainties, while Dostoevskii in his last years preached an astonishingly crude variety of religio-political messianism. The humanism of some later religious philosophers (including the *Signposts* authors Berdiaev and Bulgakov) is hard to reconcile with their eschatological impatience.

See also: Asmus, V.F.; Chernyshevskii, N.G.; Fëdorov, N.F.; Florenskii, P.A.; Kropotkin, P.A.; Lossky, N.O.; Positivism, Russian; Rozanov, V.V.; Russian literary Formalism; Shestov, L.; Shpet, G.G.; Trotsky, L.

AILEEN KELLY

References and further reading

Berlin, I. (1978) *Russian Thinkers*, London: The Hogarth Press.(Essays in nineteenth-century Russian thinkers, including Tolstoi, Herzen, Bakunin and Belinskii.)

Edie, J.M., Scanlan, J.P. and Zeldin, M.B. (eds) (1966) *Russian Philosophy*, Chigaco, IL: Quadrangle Books, 3 vols.(A selection of texts, well annotated and introduced, from the beginnings of Russian philosophy until the Soviet period.)

Masaryk, T.G. (1955) *The Spirit of Russia. Studies in History, Literature and Philosophy*, trans. E. and C. Paul, with additional chapters and bibliographies by J. Slavik, London: Allen & Unwin, 3 vols.(First published in German in 1913, and still an excellent introduction to Russian philosophy.)

Walicki, A. (1980) *Russian Thought From the Enlightenment to Marxism*, Oxford: Clarendon Press.(Covers the main movements in Russian thought from the eighteenth to the early twentieth century.)

Zenkovsky, V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, 2 vols; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, London and New York: Routledge & Kegan Paul and Columbia University Press, 1953.(A general history.)

Russian Religious-Philosophical Renaissance

The Russian Religious-Philosophical Renaissance was created by lay intellectuals who found rationalism, positivism and Marxism inadequate as explanations of the world or guides to life. They were deeply engaged in finding solutions to the problems of their time, which they saw as moral or spiritual/cultural in nature. Some were already devout Christians; others became so later on. Collectively known as the God-seekers, they propounded their ideas in numerous publications and in the Religious-Philosophical Societies of St Petersburg and Moscow. The meetings of these societies attracted capacity audiences and helped disassociate religion from reaction. Branches were founded in Kiev and Vladimir. The founding members were mainly Symbolist writers and idealist philosophers. Both groups sought a new understanding of Christianity, but the Symbolists emphasized psychological and literary/aesthetic issues and the idealists focused on ethics, epistemology and political and social reform. The Revolution of 1905 was a watershed for all of them. The hitherto apolitical Symbolists perceived it as the start of the apocalypse and championed anarchistic political doctrines. The idealists continued to champion reform. After the revolution, some of them called for a new religious intelligentsia that respected culture and the creation of wealth, spiritual/cultural and material. Both groups began to talk about national identity and destiny. The Bolshevik Revolution signalled the end of the Religious-Philosophical Renaissance. In 1922-3, over 160 non-Marxist intellectuals were forced into exile, where they continued their work. Inside Russia private religious-philosophic study circles carried on illegally.

The Religious-Philosophical Renaissance had a profound impact on Russian thought and culture. It inspired attempts to ground metaphysics and political doctrines in Christianity, demands for church reform, visions of a new culture, sophiology, religious existentialism and new interpretations of Orthodox ritual and dogma. Its proponents made people aware of the needs of the 'inner man', the soul or the psyche, and the importance of art and myth. Symbolism became the dominant aesthetic, shaping literature, poetry, painting and theatre. Theorists of Symbolism tried to make it the basis of a new cosmological worldview. The Religious-Philosophical Renaissance was rediscovered by Soviet intellectuals in the 1960s, nourished the dissident movement from then on, and is extensively discussed in Russia today.

1 The leading figures

The leading figures of the Religious-Philosophical Renaissance opted for spirituality over materialism, art over utility, and mysticism or idealism over positivism. Focusing on spiritual, psychological and cultural transformation rather than social engineering, they debated issues such as the relevance of Christianity to the political and social order, Christian attitudes to sex, philosophy and the law, art and cognition, the intelligentsia and the people, and the essence of 'Russianness'. Although they argued about the same problems and were occupied in a common search for new values, their approaches were somewhat different so that in essence there were two groups. The first group included the Symbolist writers Dmitrii Merezhkovskii (1865-1941), his wife Zinaida Gippius (1869-1945), Dmitrii Filosofov (1872-1940), Nikolai Minskii (real name Vilenkin, 1855-1937), Andrei Belyi (1880-1934), Aleksandr Blok (1880-1921), Vyacheslav Ivanov (1866-1949) and Aleksei Remizov (1877-1957). Associated with them were Fëdor Sologub (Teternikov, 1863-1927), Vasilii Rozanov (1856-1919) and Aleksandr Benois (1870-1960), of the journal *Mir iskusstva (The World of Art)* (1899-1904). The Symbolist writer Valerii Briusov (1873-1924) was not a God-seeker, but he used their terminology and published their works in his journal *Vesy (The Balance)* (1904-9). Influenced by Nietzsche and Wagner, and in some cases by Solov'ëv, this group was primarily concerned with art, self-expression, and personal freedom (see Nietzsche: impact on Russian thought; Solov'ëv, V.S.). Once they became politicized most of them favoured anarchism but, except for the mandate to feed the hungry, they ignored economic issues and were indifferent or hostile to industrialization. The second group included the contributors to the symposium *Problemy idealizma (Problems of Idealism)* (1902), Pavel Novgorodtsev (1866-1924), Sergei Trubetskoi (1861-1905), Evgenii Trubetskoi (1863-1920), Nikolai Berdiaev (1847-1948), Sergei Bulgakov (1871-1944), Semën Frank (1877-1950), and Pëtr Struve (1870-1944). Influenced by Neo-Kantian idealism and by theories of natural law, they advocated economic growth if it be guided by ethical considerations, self-discipline and a sense of responsibility for one's actions (see Natural law; Neo-Kantianism, Russian §5). The idealists were also influenced by Solov'ëv, but they highlighted different writings. Neither group was homogeneous. The Trubetskoi brothers objected to the Nietzscheanism of Berdiaev's and Frank's essays in *Problemy idealizma*. Belyi and Rozanov defended the symposium *Vekhi (Signposts)* while Merezhkovskii attacked

it (see Signposts movement). Berdiaev and Bulgakov ultimately rejected Kant. Ivanov, a proponent of Mystical Anarchism during the Revolution of 1905, contributed to the idealists' anti-Bolshevik symposium *Iz glubiny (Out of the Depths)* (1919). Another leading figure, the philosopher Lev Shestov (1866-1938) was neither a Symbolist nor an idealist. The philosopher-priest Pavel Florenskii (1882-1937) was a God-seeker in the early twentieth century.

2 The Religious-Philosophical Society meetings

The Religious-Philosophical Society of St Petersburg was founded by Merezhkovskii, Gippius and Filosofov. After obtaining permission from the Overprocurator of the Holy Synod Konstantin Pobedonostsev, they held the first meeting in November 1901. Bishop Sergei (Stragorodskii), Rector of the Theological Academy of St Petersburg (university level), was the President and a different Bishop Sergei, Rector of the Theological Seminary (secondary-school level) was Vice-President. Also presiding over the meetings were V.M. Skvortsov, editor of *Missionerskoe obozrenie (Missionary Survey)*, Merezhkovskii, Gippius, Filosofov, Minskii, Benois, Valentin Ternavtsev, the secretary of the Holy Synod, and two professors from the Theological Academy, Anton Kartashev and Vasilii Uspenskii.

At the meetings a paper was read and discussed; the discussion sometimes continued for several sessions. The minutes were published serially in the Merezhkovskii's revue *Novyi put' (New Path)* (1902-4). Ternavtsev's keynote address, 'The Clergy and the Intelligentsia', was a call for their reconciliation. To combat atheistic socialism, he argued, the Church must find Christian solutions to the problems of everyday life, not ignore them. Several sessions were devoted to Merezhkovskii's studies of the life and work of Tolstoi, Dostoevskii and Gogol' and their contemporary relevance. Sessions devoted to the relation of church and state, to Christian attitudes to sex, and to whether or not Orthodox dogma is complete were particularly stormy. Outraged at the excommunication of Tolstoi by the Holy Synod in 1901, Merezhkovskii called for a re-opening of the issue of the Synod's authority and charged that by excommunicating Tolstoi it had alienated everyone who believes in freedom. Father Mikhail Semёnov's paper on Christian marriage, which maintained that sexual desire must be expressed only in marriage and for the purpose of begetting children, inaugurated a particularly stormy series of sessions on 'the mystery of the flesh'. Rozanov argued that the sexual act is itself holy and denied the right of the Church to regulate marriage and divorce (see Rozanov §2). Merezhkovskii talked about 'holy flesh', the permeation of flesh by spirit. Uspenskii's paper on Christian dogma provoked challenges to the authority and position of the clergy as its sole interpreter. Some members denied the need for any dogma at all. Others maintained that new dogma was necessary and that laymen must share in its creation. In April 1903 Pobedonostsev ordered that the society be closed down. From his point of view the 'mission to the intelligentsia' had failed. But he could not surpress the rethinking of Orthodoxy that the society had stimulated. The meetings, twenty-two in all, provided a platform for new religious views and forced clergymen to defend their positions. This hitherto unprecedented spectacle encouraged more challenges to clerical authority and, during the Revolution of 1905, demands for a Church council (*Sobor*). The St Petersburg society was revived in 1907 but without the clergy.

The Solov'ёv Moscow Religious-Philosophical Society (1906-17) was founded by Berdiaev, Bulgakov, Florenskii, Vladimir Ern (1881-1917) and others. Social and political issues dominated the agenda, as they did that of the revived St Petersburg society (sometimes the same paper was presented to both), but issues such as Christian asceticism and church reform were also on the agenda. In 1906, members or former members of these societies formed two short-lived groups: The Group of Thirty-Two, priests who demanded various reforms, and the Christian Brotherhood of Struggle, which preached a kind of liberation theology (see Liberation theology). Bulgakov attempted, unsuccessfully, to form a Union of Christian Politics.

3 Merezhkovskii's 'new religious consciousness'

Merezhkovskii was the dominant figure in the original St Petersburg society, which he founded to proselytize his 'new religious consciousness' and his quest for a new Christianity. In the 1890s, he had initiated the Symbolist movement and popularized Nietzsche's thought. In the early twentieth century, he shaped the framework of the Religious-Philosophical Renaissance and, as he had done for Symbolism, defined the issues which others treated in a more sophisticated manner. He propounded his views in a trilogy of historical novels (about Julian the Apostate, Leonardo da Vinci and Peter the Great), in studies of Tolstoi, Dostoevskii and Gogol', and in numerous articles about current social, political, cultural and religious issues.

Merezhkovskii taught that ‘historical Christianity’ (Christianity as taught by the churches) was becoming obsolete; he expected a New Revelation, a Third Testament, to be granted by Christ himself, whose Second Coming was imminent. At that point, the joyousness of Jesus’ Resurrection will prevail over the suffering of Golgotha and the hitherto proscribed pagan values of sensuality, worldly beauty, and self-affirmation will be sanctified. Russians were to prepare for His Advent by studying all religions and by joining Merezhkovskii in his religious quest. At first Merezhkovskii envisaged a new church developing out of the Orthodox Church. Later on, he advocated a complete break with it.

When Merezhkovskii began to preach his ‘new religious consciousness’, religion was so unpalatable to the intelligentsia that he found it necessary to expend a great deal of energy arguing that people need religion and that surrogate religions - whether scientism, aestheticism or socialism - were inherently unsatisfactory. Religion, he said, begins when man realizes that he is mortal. Therefore, personal immortality, the resurrection of both the body and the soul, is the heart of religion. All other aspects, including ethics, are mere by-products. Faith nourishes the soul just as food nourishes the body. Individuals are part of a culture shaped by religious faith; when the faith breaks down the culture loses its unity and individuals are left isolated and alone. The effects of spiritual deprivation are nameless fear and loneliness, and the sufferers do not even know what ails them. Western civilization was nourished by the sap of Christianity, but rationalism has destroyed that sap. Only a ‘Christian renaissance’ would enable contemporary humankind to overcome despair and decadence. Merezhkovskii also argued that Nietzsche’s objections to Christianity applied only to ‘historical Christianity’. True Christianity is not life-denying; it is the supreme affirmation of life. Christianity does not teach a slave morality but a new morality beyond conventional notions of good and evil. And Jesus Christ is the Superman that Nietzsche sought in vain.

Previously, in his essay ‘Pushkin’ (1896), Merezhkovskii had argued that Christianity and paganism (really Nietzscheanism) represented the ‘truth of heaven’ (personal immortality and love) and the ‘truth of the earth’ (worldly pleasures) respectively. Each constituted half of a yet unknown greater truth. Merezhkovskii’s search for this truth led him to postulate that the Old Testament was the Revelation of God the Father and the New Testament of God the Son. A Third Revelation will be of the Holy Spirit. In emigration, Merezhkovskii maintained that the Holy Spirit was feminine.

In *Tolstoi i Dostoevskii* (1900-1), written in the first flush of his ‘new religious consciousness’, Merezhkovskii reformulated the ‘two truths’ as polarities *within* Christianity. Comparing the Russian giants with each other and with Nietzsche, Merezhkovskii argued that Dostoevskii represented ‘the spirit’ and Tolstoi represented ‘the flesh’ which ‘historical Christianity’ denied. Merezhkovskii attributed Tolstoi’s asceticism to an unconscious preoccupation with the flesh, accused him of hypocrisy and of a ‘flight from culture’, and claimed that Tolstoi’s real religion was Buddhism. But Merezhkovskii also saw an unconscious ‘pagan holiness’ in Tolstoi’s *The Cossacks* and he praised Tolstoi’s courage in defying the Orthodox Church. Even so, he preferred Dostoevskii, seeing him as the prophet of the spiritualization of the flesh and of a new humanity united by ‘evangelic love’. Even though Dostoevskii was not always faithful to his own insights, even though he had the Grand Inquisitor advocate the slavery of ‘miracle, mystery, and authority’ in *Brat’ia Karamazovy (The Brothers Karamazov)*, Elder Zosima’s message in the same book showed that Dostoevskii anticipated the coming Christ. Merezhkovskii believed that the Grand Inquisitor and Elder Zosima voiced contradictory aspects of Dostoevskii’s incompletely integrated worldview, but that Elder Zosima and his disciple Alësha Karamazov represented Dostoevskii’s truly prophetic insights. A decade later, Merezhkovskii described his attempt to reconcile Christianity and paganism as dangerous heresy. As he distanced himself from Nietzsche, his admiration for Tolstoi grew correspondingly (see Dostoevskii, F.; Tolstoi, L.N.).

To Merezhkovskii, beauty outranked sex as the greatest worldly pleasure and beauty was incarnated in the statues of Venus or Aphrodite that the early Christians destroyed. Hence his objection to Tolstoi’s *Chto takoe iskusstvo? (What is Art?)* (1898), which subjugated art to moralism. In a study of Gogol’ (1903) Merezhkovskii portrayed his confessor, Father Matvei (Konstantinovskii), as the personification of Christian asceticism, and blamed him for Gogol’s ruin. Merezhkovskii’s interpretation of Tolstoi, Dostoevskii and Gogol’ as religious thinkers rather than social critics inspired other ‘religious’ interpretations.

Novyi put’ provided an outlet for other God-seekers. Blok made his debut there. Rozanov had a regular column, as did Bulgakov who came on board with Berdiaev in mid-1904. Ivanov’s influential study, ‘Ėllinskaia religiia

stradaiiushchego Boga' (The Hellenic Religion of the Suffering God) (1904), which treated Dionysus as a precursor of Christ, first appeared in *Novyi put'*. Florenskii published some early articles in it. Berdiaev and Bulgakov became co-editors of its successor journal *Voprosy zhizni (Questions of Life)* (1906), which featured essays by Symbolists and idealists.

During and after the Revolution of 1905, political and other disputes fractured the Symbolist group and created a breach between them and the idealists. Merezhkovskii advocated a 'religious revolution' that would institute the Kingdom of God on Earth and proclaimed that the atheistic intelligentsia, not the Orthodox Church, was the carrier of Christ's message: 'They are not yet with Christ but Christ is with them'. Ivanov supported Mystical Anarchism, a doctrine that advocated the abolition of government, law and morality. Love (*eros*), ritual and myth would be the social cement. The idealists continued to champion reform. Members of both groups predicted that 'Socialism as Religion' would lead to despotism.

4 Symbolism as a worldview

Russian Symbolism was not just an artistic school but a worldview, based on a mystical epistemology which held that the empirical world is but a reflection of a higher reality, and that art, imagination and intuition, rather than reason and science, are the means to new psychological and religious truths. The Symbolists saw themselves as creators of a new culture of freedom, beauty and love, which was opposed in all respects to contemporary materialist and positivist-utilitarian culture, including such 'bourgeois' institutions as constitutions and parliaments.

Basic to Russian Symbolism were the Orthodox concepts of transfiguration and incarnation, and the Orthodox belief that man is called upon to fulfil the divine Plan - to be co-creator with God. Also basic, especially to Ivanov, Belyi and Blok, were Solov'ëv's apocalypticism, and his concepts of all-unity (*vseedinstvo*), Godmanhood (*Bogochelovechestvo*) and Sophia (divine wisdom, the 'eternal feminine', or the World Soul) (see Solov'ëv, V.S. §§1-2). Melding these concepts with ideas derived from Nietzsche, Wagner, Dostoevskii, Fëdorov and occult doctrines, the Symbolists regarded art as a theurgical activity that would transfigure the world and human beings. The poetic word was itself a symbol, bearing the transparent reflection of the eternal in the temporal. The Symbolist poet had a divine mission, the articulation of a new saving word which would generate a new world. Within this broad framework, each Symbolist had special themes. Blok wrote poems to 'the Beautiful Lady' and to 'the Stranger', a fallen Sophia. Ivanov advocated self-transcendence in the mystical ecstasy of the Dionysian rites. Maintaining that poetry requires a special hieratic language, he was interested in the roots of language and the origins of myth and cult. Merezhkovskii opposed all attempts to subvert individual integrity and used language that ordinary people could understand. Belyi tried to make Symbolism the basis of a new religious worldview that encompassed art, philosophy and science. Briusov, however, maintained 'Symbolism was and always wanted to be only art'. He made this statement in 1910, during the 'crisis of Symbolism'. After the crisis, individual Symbolists remained creative but went their separate ways. *Vesy* and *Zolotoe runo (Golden Fleece)* (1906-9) had ceased publication, but they had access to mainstream journals such as *Russkaia mysl' (Russian Thought)* and to a new religious journal *Trudy i dni (Works and Days)* (1912-16).

The Symbolists refused to defer to 'necessity', including the 'laws of nature'. Ivanov titled one of his essays 'On the Non-Acceptance of the World' (after Ivan Karamazov's refusal to accept the world that God has created) and praised 'struggling with God' (*Bogoborchestvo*), by which he meant rejection of that which exists in the name of a higher ideal. Some Symbolists, and Berdiaev, rejected the traditional family (husband, wife, children). They counterposed resurrection to procreation and separated sex from procreation in the hope of breaking the endless cycle of birth and death. Their ideal human being was an androgyne, which implies the end of the sexual act.

The Symbolists' sociopolitical ideal, which they borrowed from the Slavophiles and Solov'ëv, was *sobornost'*, originally an ecclesiastical concept that pertained to the unity of all believers in the mystical body of Christ (see Slavophilism §3). To the Symbolists and to most idealists as well, *sobornost'* connoted a society united by love and common ideals whose members retain their individuality, as opposed to individualism which they rejected as self-affirmation against or apart from society. They also rejected impersonal collectivism.

As the eschatological expectations aroused by the Revolution of 1905 died down, the Symbolists and some idealists condemned the 'pagan' Renaissance and romanticized the medieval West. Ivanov made Dante Alighieri

into a Symbolist and equated Beatrice with Sophia. Paradoxically, the very same people pontificated about Russia's cultural/national identity and destiny. Ivanov tried to articulate the 'Russian Idea' - that which distinguishes Russia from other nations - insisting that it was a Christian idea and that only Christianity could provide the spiritual integration that Russia needed. He urged the intelligentsia to 'descend' to the people in the spirit of Christian humility and self-sacrifice. Merezhkovskii objected to Ivanov's view on various grounds, but he shared Ivanov's belief that Russia stood at a crossroads and must choose between 'the beast' and Christ. Blok began to write about an anthropomorphized Russia, at once mother and wife, and to prophesy the destruction of the Westernized intelligentsia. The latter was also the theme of Belyi's novel *Serebrianyi golub' (The Silver Dove)* (1909), the first of a projected trilogy on the collision of East and West. *Petersburg* (1911) was the second. The third novel, which Belyi never wrote, was to be about the legendary city of Kitezh which sank to the bottom of Lake Svetloiar at the time of the Mongol invasion and would rise at the end of time. Reports at Religious-Philosophical Society meetings included attempts to distinguish between 'nationality', love of one's own land and culture, and aggressive, self-aggrandizing 'nationalism'. Ivanov maintained that the 'sin' of Slavophilism was its inclusion of spiritually extraneous factors such as the state in its metaphysics. He viewed the nations of humanity as members of a cosmic *sobornost'*.

Nevertheless, Ivanov welcomed the outbreak of the First World War as the start of a new constructive era, as did most other Religious-Philosophical Society members. Merezhkovskii was a notable exception. During the war, Ern gave a lecture 'From Kant to Krupp' in which he argued that Kant's philosophy leads inevitably to a drive for world hegemony. In another lecture, Ern proclaimed that 'the time is slavophilizing'. In his major philosophical work *Struggle for the Logos* (1911), Ern distinguished between the Logos, a real entity, and abstract rationalism, associating the latter with the West.

5 The idealists

The symposium *Problemy idealizma* was a foundational work of the Religious-Philosophical Renaissance. All the contributors emphasized the importance of the individual person, grounded this tenet in a metaphysical worldview, and attacked positivism, materialism and determinism. They came from different milieus. Novgorodtsev, who was also the editor, and the Trubetskoi brothers were devout Christians and long-standing members of the Moscow Psychological Society (founded 1885). Berdiaev, Bulgakov, Frank, Struve and S.A. Askoldov had been 'legal Marxists' in the 1890s. All the idealists were indebted to Kant, and most of them to Solov'ëv as well. Berdiaev and Frank were also captivated by Nietzsche. Novgorodtsev, a liberal and a founding member of the Kadet (Constitutional Democratic) Party, did not write about religion *per se*. Rather, he argued for moral renewal based on 'the principle of the unconditional significance of the person'. Novgorodtsev based his respect for the individual person on Christian and Kantian ethics, which he viewed as compatible, even though he recognized a certain tension between Christian personalism and Kantian rationalism (see Kantian ethics). He regarded the rule of law as a safeguard against arbitrary political power and as a precondition for the development of individuality, and argued that for law to be respected it must accord with morality and justice. Novgorodtsev never gave a paper or participated in the Religious-Philosophical discussions, but he knew about them. During the Revolution of 1905 he argued that political and social reform could not, by themselves, bring about a good society; the prerequisites were moral regeneration and self-perfection. Convinced that there would always be tension between the personal and the political, he wanted to keep them separate. Novgorodtsev did not contribute to *Signposts*, but he sympathized with its critique of revolutionism and gave it his qualified support. In *On the Social Ideal*, serialized 1911-16, he warned that utopianism could lead to despotism. Only in exile did Novgorodtsev write about subjects such as 'The Essence of the Russian Orthodox Consciousness' (1923).

Sergei Trubetskoi held that philosophy must be based on a metaphysics that is deeply rooted in Christianity. He claimed that rationalism, empiricism and mysticism each omit essential aspects of reality, and hoped to unite their insights in 'absolute all-unity'. He called his philosophy 'concrete idealism'. One of its tenets was the communal (*sobornyi*) nature of human consciousness. His most important books are two studies of the history of philosophy, *Metaphysics in Ancient Greece* and *The Doctrine of the Logos*, and three works in which his own views are stated: *On the Nature of Human Consciousness*, *The Foundations of Idealism* and *The Belief in Immortality*. One of Trubetskoi's concerns was to overcome Solov'ëv's pantheism. Another was to resist the 'mystical alogism' of Berdiaev, Bulgakov, and Florenskii by arguing that 'Revelation itself is inseparable from the Logos and is subject to its valuation'.

Evgenii Trubetskoi is best known for his two volume study of Solov'ëv's worldview (Mirosozertsanie Solov'ëva) (1912) which describes the collapse of Solov'ëv's utopianism, but he also wrote about early Christianity, epistemology, and Russian icons. In his newspaper, *Moskovskii ezhenedel'nik (Moscow Weekly)* he argued that Christianity did not mandate either socialism or anarchism, but was compatible with several political systems. He was a co-director of the publishing house Put', which specialized in books and articles about Orthodox philosophy and its contemporary applications.

Signposts was the high-water mark of the former legal Marxists' liberalism and Kantianism. Soon after, their thought took a different direction. Berdiaev and Bulgakov rejected Kant. Frank continued to grapple with 'Kantian' ideas, but was no longer a Neo-Kantian. Struve advocated a mystique of the state, nationalism and imperial expansion.

6 Responses to the Bolshevik Revolution

Merezhkovskii considered Bolshevism the reign of the Anti-Christ and worked against it in Russia and in emigration. Rozanov bewailed 'the apocalypse of our time' and attributed it to the Church's neglect of everyday life. By contrast, Belyi and Blok perceived the Bolshevik Revolution as part of a spiritual revolution that would culminate in a non-materialistic, artistically creative new man. Ivanov regarded it as a religious trial, a Golgotha which would eventually result in Resurrection. In 'Our Language', his contribution to *Iz glubiny*, Ivanov attacked the Bolshevik orthographic reform as a secularizing measure. In their contributions, Berdiaev, Bulgakov, Novgorodtsev, Frank and Struve blamed the Revolution on the secular intelligentsia. Struve emigrated soon after. Berdiaev, Bulgakov, Frank and Novgorodtsev were expelled in 1922. Ivanov left voluntarily in 1924.

Soviet intellectuals discovered the God-seekers in the 1960s. By 1980 there was so much interest that the government began publishing books that debunked them. In 1988-9, however, Gorbachev permitted official publications of their works. Interest grew rapidly and continues to the present day.

BERNICE GLATZER ROSENTHAL

References and further reading

Edie, J.M.; Scanlan, J.P. and Zeldin M.B. (eds) (1965) *Russian Philosophy*, vol. 3, Chicago, IL: Quadrangle Books.(This anthology includes critical commentary on twentieth-century Russian philosophers, edited with the collaboration of G.L. Kline.)

Hughes, R.P. and Paperno, I. (eds) (1994) *Christianity and the Eastern Slavs, vol. 2, Russian Culture in Modern Times*, Berkeley, CA: University of California Press.(Based on papers presented to two conferences commemorating the centennial of the Christianization of Rus.)

Khoruzhii, S. (1994) *Posle pereryva. Puti russkoi filosofii (After the Break. Paths of Russian Philosophy)*, Moscow: Aleteiia. (Interpretation of the distinctive aspects of the Russian religious-philosophical tradition by a contemporary philosopher.)

Kornbatt, J.D. and Gustafson, R. (eds) (1996) *Russian Religious Thought: Contexts and New Perspectives*, Madison, WI: University of Wisconsin Press.(Articles on Bulgakov, Florenskii, Frank and Solov'ëv.)

Kuvakin, V. (ed.) (1994) *A History of Russian Philosophy*, Buffalo, NY: Prometheus Books, vol. 2, part 9, 'Idealism and New Systems'.(Articles on Neo-Kantianism, Rozanov, Berdiaev, Shestov, Bulgakov, Florenskii, Lossky and Frank, by academics at Moscow State University and other Russian institutions.)

Lossky, N.O. (1951) *History of Russian Philosophy*, New York: International Universities Press.(Chapters on the leading religious philosophers, except for Rozanov.)

Pachmuss, T. (1971) *Zinaida Hippius: An Intellectual Profile*, Carbondale, IL: Southern Illinois University Press. (Emphasizes the Christian aspects of Gippius' life and work.)

Poltoratzskii, N. (ed.) (1975) *Russkaia religiozno-filosofskaia mysl' xx veka (Russian Religious-Philosophical Thought of the Twentieth Century)*, Pittsburgh, PA: University of Pittsburgh Press. (Articles about the major figures and themes of the Religious-Philosophical Renaissance by American scholars.)

Putnam, G. (1977) *Russian Alternatives to Marxism: Christian Socialism and Idealistic Liberation in Twentieth Century Russia*, Knoxville, TN: University of Tennessee Press.(This work takes the sociology of knowledge approach, viewing Bulgakov and Novgorodtsev as examples of the claims of the Church and the academy to the moral leadership of Russia.)

Pyman, A. (1994) *A History of Russian Symbolism*, Cambridge: Cambridge University Press.(Analyses the works of the major Symbolists from a literary perspective; includes some discussion of underlying philosophical and

religious issues.)
Read, C. (1979) *Religion, Revolution, and the Russian Intelligentsia, 1900-1912: The Vekhi Debate and its Intellectual Background*, London: Macmillan.(Describes the intelligentsia's turn to religion and its debate about the revolutionary tradition.)
Rosenthal, B.G. (1975) *D.S. Merezhkovsky and the Silver Age: The Development of a Revolutionary Mentality*, The Hague: Martinus Nijhoff.(Merezhkovskii's evolution from Symbolist, to God-seeker, to revolutionary.)
Rosenthal, B.G. and Chomiak, M.B. (eds) (1990) *A Revolution of the Spirit: Crisis of Value in Russia, 1890-1924*, Bronx, NY: Fordham University Press. (Translations of Merezhkovskii's 'Revolution and Religion', Belyi's 'Revolution and Culture', Ivanov's 'The Crisis of Individualism', Berdiaev's 'Socialism as Religion', Bulgakov's 'An Urgent Task', Novgorodtsev's 'The Essence of the Russian Orthodox Consciousness' and other important essays.)
Scanlan, J.P. (ed.) (1994) *Russian Thought After Communism: The Rediscovery of a Philosophical Heritage*, Armonk, NY: M.E. Sharpe.(Articles on the Religious-Philosophical Renaissance by Western and Russian scholars.)
Scherrer, J. (1973) *Die Religios-Philosophischen Vereinigungen (The Religious-Philosophic Unions)*, Wiesbaden: Otto Harrassowitz.(The first and still the most complete study of the Religious Philosophical Societies.)
Walicki, A. (1992) 'Pavel Novgorodtsev: Neo-Idealism and the Revival of Natural Law', in *Legal Philosophies of Russian Liberalism*, Notre Dame, IN: Notre Dame University Press, 291-341.(Shows that Novgorodstev was indebted to Solov'ëv as well as to Kant.)
Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, 2 vols; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, vol. 2, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, 760-80.(Chapters on the leading religious philosophers.)
Zernov, N.N. (1963) *The Russian Religious Renaissance of the Twentieth Century*, London: Darton, Longman & Todd.(The pioneering work on the subject.)

Ryle, Gilbert (1900-76)

Alongside Wittgenstein and Austin, Ryle was one of the dominant figures in that middle period of twentieth-century English language philosophy which became known as 'Linguistic Analysis'. His views in philosophy of mind led to his being described as a 'logical behaviourist' and his major work in that area, The Concept of Mind (1949), both by reason of its style and content, has become one of the modern classics of philosophy. In it Ryle attacked what he calls 'Cartesian dualism' or the myth of 'the Ghost in the Machine', arguing that philosophical troubles over the nature of mind and its relation with the body arose from a 'category mistake' which led erroneously to treating statements about mental phenomena in the same way as those about physical phenomena. For Ryle, to do something was not to perform two separate actions - one mental, one physical - but to behave in a certain way.

Much of Ryle's work had a similar theme: philosophical confusion arose through the assimilation or misapplication of categorically different terms, and could only be cleared up by a careful analysis of the logic and use of language. He later became preoccupied with the nature of reflective thinking, since this stood as an example of an activity which seemed to evade the behaviouristic analysis that he recommended. Ryle was also a considerable Plato scholar, though his work in this area has been less influential.

1 Early influences

Gilbert Ryle was born in 1900 in Brighton, England, and went to the Queen's College, Oxford, as a Scholar in Classics. He read Literae Humaniores (Classics and Philosophy) and then Modern Greats (Philosophy, Politics and Economics). In 1924 he became a Lecturer in Philosophy at Christ Church, Oxford, and was to remain at Oxford University for the whole of his academic life.

In a deliberate attempt to avoid sinking into what he felt was a rather parochial philosophical atmosphere at Oxford, in the first few decades of the twentieth century Ryle deliberately directed his academic gaze outwards. He travelled to Cambridge to hear Moore, Russell and Wittgenstein, and set himself the task of reading Wittgenstein's *Tractatus Logico-Philosophicus* as well as dipping into Russell's *Principia Mathematica*. Ryle also set out to gain some knowledge of contemporary continental European philosophy. He taught himself German and read Husserl's *Logische Untersuchungen*, as well as some Croce, Gentile, Meinong, Brentano, Bolzano and Frege. Against the prevailing fashion he offered at Oxford 'an unwanted course of lectures, entitled "Logical Objectivism: Bolzano, Brentano, Husserl and Meinong"', which became known as 'Ryle's three Austrian railway stations and one Chinese game of chance'. Indeed Ryle's first published pieces of philosophical writing were book reviews of works by the Polish philosopher Ingarden (a follower of Husserl) and Heidegger.

As with the Cambridge school, so with the Continentals, Ryle was principally interested in their philosophical logic, and this is reflected in the titles of his earliest papers, 'Negation' (1929), 'Are There Propositions?' (1930) and following shortly after, 'Imaginary Objects' (1933a), 'About' (1933b) and 'Internal Relations' (1935).

In the introduction to his *Collected Papers*, Ryle remarks that 'to elucidate the thoughts of a philosopher we need to find the answer not only to the question "What were his intellectual worries?" but, before that question and after that question, the answer to the question "What was his overriding worry?"' (1971: ix). Ryle himself declared that his overriding worry concerned the nature and function of philosophy itself. This question has arisen from time to time in philosophy but, through the influence both of the scientistic doctrines of logical positivism and Wittgenstein's *Tractatus*, and also because of the remarkable inroads which natural science seemed to be making into the subject matter of philosophy, the question became particularly insistent in the 1920s.

Ryle gave his answer in the seminal paper, 'Systematically Misleading Expressions' (1932). His uncompromising and restricted view of philosophy, as 'the exercise of systematic restatement', became one of the major influences on the style and content of mid-century English-speaking philosophy. It led to those who felt sympathetic to Ryle's (and to some extent Wittgenstein's) view of philosophy being dubbed 'Ordinary Language Philosophers' and to their movement, if it could be called that, being called 'Linguistic Analysis' (seeOrdinary language philosophy, school of §§2-3).

When Ryle came to the task of working out his programme of 'systematic restatement' in more detail, it turned out to consist mainly of two subsidiary tasks, a positive and a negative one. The negative task was concerned with

exposing and correcting conceptual mistakes perpetrated by philosophers' mishandling of ordinary language in the course of propounding and defending philosophical theories. The positive task consisted in what he later described as 'conceptual cartography' or the job of getting clear about the basic categories that were or should be operative in some area of knowledge.

These views were among the influences that shaped the style of English-speaking philosophy in the twentieth century. He believed that the 'growing passion for ratiocinative rigour' was to be satisfied by ever more careful, step-by-step conceptual analysis. Great emphasis was to be placed on getting clear about the exact meaning of terms and, in consequence, on the exact usage of them. Ryle's views about the nature of philosophy also influenced the subject matter of philosophy, because he felt, with Wittgenstein, that philosophers should eschew theorizing and be content with the logico-linguistic analysis of concepts. In turn, this meant that for several decades philosophy in much of the English-speaking world became heavily weighted towards the analysis of language.

For most of the Second World War Ryle was engaged in intelligence work. In 1945, he was elected to the Waynflete Chair of Metaphysical Philosophy at Oxford, which had lain vacant since R.G. Collingwood's death in 1943. In his Inaugural Lecture, entitled 'Philosophical Arguments', Ryle expanded on his views about the nature of philosophy. 'Philosophical reasoning,' he proclaimed, ' separates the genuine from the erroneously assumed logical powers of abstract ideas by using the *reductio ad absurdum* argument as its flail and winnowing fan,' and philosophers of genius are those who have the insight to see, then flail and winnow, new abstract ideas which then become the core of some new area of enquiry or some new way of making sense of an old area of enquiry.

2 *The Concept of Mind*

In 1947, Ryle became editor of *Mind*, then the most influential philosophy journal in the English-speaking world, and set about turning its gaze outwards as well. A few years before, in 1945, he was beginning to crystallize his ideas about a number of issues in the philosophy of mind. The outcome of this, in 1949, was *The Concept of Mind*. As he himself readily admitted, *The Concept of Mind* was 'a sustained piece of analytical hatchet-work' on Cartesian dualism, whose view of mind he gleefully lampooned as an account of 'the life of a ghostly Robinson Crusoe' and as 'the dogma of the Ghost in the Machine'. Strictly speaking, his target was not so much Descartes himself as Cartesian doctrines which he believed had lain more or less unmolested in the philosophy of mind up to that time.

Ryle believed that Cartesian dualism was one large category mistake - an incorrect assignment of the terms of our common-sense psychological vocabulary to one logico-linguistic category or type when they should be assigned to another. After careful philosophical analysis it would be discovered that our mental terms, such as 'mind', 'thought' or 'belief', are not words which refer to or describe an inner private mental world of faculties with their proprietary activities but are mostly to be analysed as dispositional terms whose attribution depends on the ordinary observation of ordinary human behaviour. For dispositions are nothing but an ability, propensity, liability or capacity to do things of a certain type in certain specifiable circumstances. To be intelligent is to be disposed to accomplish successfully such tasks as doing mathematical problems or intelligence tests; it is not to provoke an inner Cartesian faculty, called 'the intellect', into producing its private and proprietary mental acts (see Dualism §1).

Even attributions about what seem to be the most 'inner' and 'private' of mental activities are still to be cashed out in a behaviouristic way. Thus imagining is paradigmatically an outward performance, such as when a boxer imagines, in an afternoon's shadow-boxing performance, how the fight will go that evening. So imagining in one's head should also be analysed in relation to ordinary 'outer' behaviour. To imagine my childhood nursery is not to engage in inner Cartesian performances, but simply to realize, anticipate or expect what I would be seeing if I were once again in my nursery. It is to be disposed to see something *if*…

To take another example, to introspect is not to observe inwardly, in some non-sensuous manner with our 'mind's eyes', some item in our streams of consciousness. Rather, it is to 'retrospect' in memory some behaviour of our own, covert or public, which we have perceived in the past with our ordinary 'outer' senses, and then to make some overall dispositional claims about what we have been able to recall. Indeed, Ryle notoriously asserted in *The Concept of Mind* that 'the sorts of things that I can find out about myself are the same as the sorts of things that I

can find out about other people, and the methods of finding them out are much the same' (1949: 155)

It was conclusions such as these which led many to describe Ryle as a logical behaviourist. In *The Concept of Mind* at least, he seemed to arrive at much the same conclusions as did the psychological behaviourists, but his reasons were logico-linguistic, not methodological. On the other hand, it is probably true to say that both those who assent that Ryle was a logical behaviourist and those who deny it find support for their arguments in *The Concept of Mind* (see Behaviourism, analytic).

3 *Dilemmas* and 'Le Penseur'

In 1953, Ryle gave the Tarner Lectures at Cambridge; they were published the following year as *Dilemmas*. The paradigmatic type of dilemma which Ryle had in mind, and which he believed would yield to the right sort of philosophical analysis, was that which arose when answers to different questions were mistakenly taken to be answers to one and the same questions and so to be in competition with one another. The solution to such dilemmas was to demonstrate that the purported answers cannot in fact be answers to one and the same question because they are of a different logical type or category. Thus, for example, the neurophysiologist's account of perception, the philosopher's account and the ordinary person's account are not in competition with each other, though they might at first appear to be so. When examined closely, each account has been generated in response to quite different questions. The neurophysiologist's question is 'What are the mechanisms of perception?', the philosopher's 'What is perception?', and the ordinary person's 'What is perceived?'

Ryle dealt with other dilemmas by analysing them as arising through the misinterpretation of some unproblematic and innocuous proposition, such that the misinterpretation then emerges as a rival to the common-sense view of the matter. For example, a dilemma occurs when the dictum, 'That whatever is, was to be' is misinterpreted by the Fatalist as 'That for whatever takes place, it was antecedently true that it was going to take place', instead of being given the quite innocuous interpretation, 'That for everything that happens, if anyone had at any previous time made the guess that it would happen, the guess would have turned out correct'. The misinterpretation occurs, Ryle suggested, when the term 'true' is incorrectly assigned to the category of 'property' when it should be assigned to the category of 'verdict'.

Though less spectacular and on the whole less successful than in *The Concept of Mind*, the Tarner Lectures were a further display of his view that many philosophical problems can be solved by 'systematic restatement' through 'the replacement of category-habits by category-disciplines'. This in turn meant that category mistakes are to be corrected by having misapplied terms and the misunderstood concepts underlying them reassigned to their correct categories or types. It was not an idle comment of Ryle's when he said that Russell's Theory of Types had been an important influence on his work (see Categories §2).

Between 1949 and 1954 Ryle wrote a further series of influential papers on philosophical logic, including '"If", "So" and "Because"' (1950a), 'Heterologicality' (1950b), 'Thinking and Language' (1951), and 'Ordinary Language' (1953). At about the same time Ryle also embarked on the first in the series of papers on the theme which was to occupy him on and off for the rest of his life: the nature of thinking, in particular reflective or contemplative thinking. Indeed, he believed that an inadequate account of this latter aspect was one of the major omissions in *The Concept of Mind*.

Ryle acknowledged that mental activities such as doing mental arithmetic or composing a tune in one's head or, in general, doing whatever it was that 'Le Penseur' (Rodin's 'Thinker') was doing posed a particular problem for his approach to the analysis of mind. Such activities were circumstance-disengaged and behaviour-free, and thus there was nothing of which a dispositional analysis could take hold. Eventually Ryle toyed with the idea of giving 'an adverbial account'. Such an account worked well enough in the context of practical thinking. Thinking, Ryle argued, was often doing something, such as playing chess or driving a car, 'thinkingly' or 'wittingly', that is, 'with initiative, care, patience, pertinacity and interest'. However, to give an adverbial account of what 'Le Penseur' is doing, one must first nominate some inner activity of which thinking is the modification. This, Ryle admitted, he was never able satisfactorily to do, and he was still wrestling with the problem at the time of his death in 1976. The record of his struggles with this problem - in the case of true contemplative thinking, of 'finding the peg' on which to hang the adverbial modifications - was published posthumously as *On Thinking* (1979) (see Mental states, adverbial theory of).

4 The return to Greek philosophy

Besides his collected papers, the other major work published by Ryle after *Dilemmas* was a series of articles on Plato which culminated in his book *Plato's Progress* (1966). This was a reappraisal of the chronology of Plato's dialogues, and included the provocative claim that the dialogues had been written for dramatic performance. In the following year his magisterial entry on Plato appeared in Paul Edwards' *Encyclopedia of Philosophy* (1967). These essays in Greek philosophy were not the sign of a change of direction in the last part of his life, but a reversion to one of his earliest philosophical interests. His interest in Greek philosophy, which encompassed both Plato and Aristotle, had never waned. In the 1930s he had immersed himself in Plato's later dialogues; this concentration culminated in his paper on the Parmenides (1939) and in a review of Cornford's book on the same dialogue. These works initiated a revolution in the philosophical interpretation of Plato's later work. In the 1950s Ryle was still enthralling undergraduates with a detailed series of lectures on the Theaetetus; he was a powerful lecturer - tall, firm-voiced, dryly humorous and incisive. When Ryle died in 1976, among his 'work in progress' there was a paper on Plato's *Meno*.

WILLIAM LYONS

List of works

Ryle, G. (1971) *Collected Papers*, London: Hutchinson, 2 vols.(These two volumes comprise all the papers, except for a few very minor pieces, such as reviews and obituary notices, which Ryle published before 1971. Of the papers mentioned in the body of this entry (with their original dates of publication in brackets, but given here with the pagination of their most accessible form, in the *Collected Papers*) - namely 'Negation' (1929: 1-11), 'Are There Propositions?' (1930: 12-38), 'Systematically Misleading Expressions' (1932: 39-62), 'Imaginary Objects' (1933a: 63-81), 'About' (1933b: 82-4), 'Internal Relations' (1935: 85-100), 'Plato's *Parmenides*' (1939: 1-44), 'Review of F.M. Cornford: *Plato and Parmenides*' (1939: 45-53), 'Philosophical Arguments' (1945: 194-211), '"If", "So" and "Because"' (1950a: 234-49), 'Heterologicality' (1950b: 250-7), 'Thinking and Language' (1951: 258-71), and 'Ordinary Language' (1953: 301-18) - all appear in volume 2, with the exception of the two papers on the *Parmenides*, which are reprinted in volume 1.)

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.(In this, his *magnum opus*, Ryle sets out to dispel 'The Cartesian Myth' on account of which both philosophy and psychology since Descartes have been haunted by the distinction between the inner and outer world. Ryle makes explicit the traces of Cartesianism in traditional theories of will, feeling, imagination, perception and thought, and tries to obliterate those traces. In place of the myth, he proposes an alternative behaviouristic account of our mental vocabulary.)

Ryle, G. (1954) *Dilemmas: The Tarner Lectures 1953*, Cambridge: Cambridge University Press.(Ryle sets out to resolve certain false dilemmas wherein the ordinary person's account of the world or of their knowledge of it, in terms of, for example, the concepts of pleasure or motion or perception, is made to look as if it is opposed to and so inferior to some technical or scientific view of the world. These resolutions are a practical demonstration of Ryle's view that many philosophical problems are just cases of conceptual confusion which can be dissolved by 'systematic restatement' through 'the replacement of category habits by category disciplines'.)

Ryle, G. (1966) *Plato's Progress*, Cambridge: Cambridge University Press.(A provocative intellectual history of Plato, in which Ryle questions the accepted chronology of Plato's dialogues and suggests that by and large they were composed for public recitation at the great Hellenic festivals. He also disputes whether Aristotle was in any meaningful sense ever a disciple of Plato.)

Ryle, G. (1979) *On Thinking*, ed. K. Kolenda, Oxford: Blackwell.(This collection of eight papers on topics of thinking encompasses most of Ryle's papers which were published after his *Collected Papers* went to press. It also includes an interesting preface by Geoffrey Warnock as well as a very brief appendix, entitled 'On Bouwsma's Wittgenstein'.)

Ryle, G. (1993) *Aspects of Mind - Gilbert Ryle*, ed. R. Meyer, Oxford: Blackwell. (Includes some of Ryle's previously unpublished 'publishable' papers, most of which were written between the late 1950s and the late 1960s. Among them is a paper on Plato's *Meno*, a short piece tracing the development of Oxford Philosophy in the mid-century, a set of notes of Ryle's lectures and a seminar taken by Meyer, and two tributes to Ryle. In his 'Annotations', Meyer mentions that 'when he retired as Waynflete Professor of Magdalen College in 1967, Ryle donated his books and some papers to Linacre [College, Oxford]. The collection consists of about 1,100 books, some papers, rough notes and letters'.)

References and further reading

Lyons, W. (1980) *Gilbert Ryle: An Introduction to His Philosophy*, Brighton: Harvester Press, and Atlantic Highlands, NJ: Humanities Press.(This is a critical introduction and guide, written for students, to the central themes in Ryle's philosophical work. It contains also a short biographical chapter plus a bibliography of works which discuss topics in Ryle's philosophy.)

Magee, B. (1971) 'Conversation with Gilbert Ryle', in *Modern British Philosophy*, London: Secker & Warburg. (This transcript of a conversation with Ryle about his own philosophical work is part of a book that originated in a series of conversations with contemporary British philosophers about their own or others' work which were first broadcast on BBC radio during the winter of 1970-1.)

Passmore, J. (1957) *A Hundred Years of Philosophy*, London: Duckworth, ch. 18.(This chapter, entitled 'Wittgenstein and Ordinary Language Philosophy', rightly places considerable emphasis on the work of Ryle and in a very scholarly way places his work in context.)

Quinton, A.M. (1964) 'Contemporary British Philosophy', in *A Critical History of Western Philosophy*, ed. D.J. O'Connor, New York: The Free Press.(This extended essay gives a synoptic account of the period leading up to Ryle and of that in which Ryle flourished.)

Rée, J. (1993) 'English Philosophy in the Fifties', *Radical Philosophy* 63.(This essay covers much the same ground as Quinton (1964) but puts an interestingly different and more sceptical slant on things.)

Wood, O.P. and Pitcher, G. (eds) (1970) Ryle, *Modern Studies in Philosophy series*, ed. A. Rorty, London: Macmillan.(This is a large collection of articles and reviews on Ryle's work and remains the best secondary source.)

Régis, Pierre-Sylvain (1632-1707)

Régis helped to define and disseminate Cartesianism. He proselytized on its behalf, defended it against its critics and innovators, and wrote the systematic textbook for which Descartes had hoped. Although primarily an expositor of Descartes' views, he sometimes developed them in creative ways that tended towards empiricism. He seems to have been led in this direction by Robert Desgabets who, while adhering to Descartes' principles, consciously departed from what Descartes actually said in order to be 'more Cartesian than Descartes himself'. The same may often be said of Régis.

'One of the ornaments and pillars of the Cartesian sect', as Pierre Bayle called him, Régis was born near Agen, France. After his education by the Jesuits at Cahors, he studied theology in Paris, where he became interested in the philosophy of Descartes. A clearly important influence was the weekly lectures given by the leading proponent of Cartesianism, Jacques Rohault. A more important influence came from Robert Desgabets, who may have introduced Régis to Descartes' philosophy. Desgabets would have provided him with an understanding of those of its principles, particularly the doctrine of created eternal truths, that gave it a quasi-empiricist complexion.

Around 1665, Rohault sent Régis to Toulouse to spread the gospel of Cartesianism. There, at Montpellier (where he met Locke), at Aigues-Mortes and finally back in Paris in 1680 after the death of Rohault, Régis' lectures enjoyed enormous success. However, given the official opposition to Cartesianism because of the theological controversies it generated, the lectures were suppressed by the Archbishop of Paris, and the publication of Régis' first, longest and most important work, the *Système de philosophie* (1690), was delayed for ten years.

Régis was less a creative thinker than an expositor of what he took to be Cartesianism (even if he sometimes developed it in surprising ways), which he defended in a number of polemical works. His *Système* criticized Malebranche on several topics, on three of which Malebranche and he then exchanged 'Replies': the nature of sensual pleasure, the problem of the horizontal versus meridional moon (on which Régis was demonstrably mistaken), and, most importantly, the nature of ideas (see Malebranche, N. §§1, 2).

As had been the case in Malebranche's much longer debate with Arnauld over the nature of ideas, the central issue was whether the soul's modifications are essentially representative (as Arnauld and Régis held), or whether ideas different from those modifications are required for knowledge of material things (as Malebranche held). Many of the same arguments from the Arnauld debate reappear - for example, Malebranche's contention that a modification of the soul, which is a particular, cannot represent a circle in general (see Arnauld, A. §3). But novel points also emerge, one of the most important being Régis' contention that the exemplary cause of an idea is the thing it represents, which he deduces from the principle that a representation cannot be of nothing. Régis and Malebranche, in fact, share this principle but draw different conclusions from it: Malebranche sees perception mediated by God, while Régis has it directly of objects. Malebranche's problem is to explain how we know material things; Régis' problem is to explain how we can fail to do so.

Malebranche's cause in general was taken up by his disciple Henri Lelevel, who found Régis committed both to Spinozism and to scepticism (see Spinoza, B. de). Régis tried to respond to the charge of Spinozism by condemning in no uncertain terms the view that there is only one substance, which is God, but he seems not to have rejected or modified his earlier view that individual minds are modes of thought, just as individual material things are modes of extension. Thus, while holding with Desgabets that substances are indefectible, the basis for personal immortality seems none the less upset.

In other debates, Régis defended Cartesianism against the sceptical attacks of P.-D. Huet, against the objections of Jean Du Hamel on a range of issues beginning with the theory of ideas, and finally against the misgivings of Leibniz, with whom he also exchanged 'Replies' concerning the consequences of Cartesianism for religion and piety. In his last work, *L'usage de la raison et de la foi* (1704), Régis sought to establish the compatibility of faith and reason, essentially by establishing their independence. He was one of the few Cartesians to subscribe to Descartes' doctrine of the created eternal truths, which he thought followed from God's omnipotence (see Descartes, R. §6). He held, moreover, that anything dependent on God's will can be known only through experience or revelation, and thus the apriorism usually associated with Cartesianism is contrary to the central

thrust of Régis' philosophy. This thrust is clear, too, in his conception of the soul, which thinks through the body.

Régis disavowed the occasionalism that he might have been expected to hold (see Occasionalism). His emphasis on divine omnipotence led him to the principle on which Malebranche had based his occasionalism: 'there is such a necessary connection between His will and the existence of the thing He wills to produce that it is incomprehensible that God should will a thing to be produced and that it not be' (1690: 110). But the created things between which this necessary connection fails to occur are none the less real secondary causes, and not just occasional causes of God's operation. As occasional causes determining God to act, they would violate his immutability. It is open to question, however, whether Régis really understood the doctrine of occasionalism, and whether the occasional causes of Malebranche really differ from Régis' secondary causes, which he describes as having 'no proper causality' and as 'contributing to the production of things only by serving as instruments by which God modifies His action' (1690: 125).

In accordance with the arboreal metaphor in the introduction to his *Principles*, Descartes thought that all philosophy was connected and based on metaphysics. Aside from the earlier sketchy effort of François Bayle, however, Régis' *Système* was the first attempt to produce a systematic and comprehensive account of the Cartesian view of things that Descartes had only imperfectly realized. In addition to logic and metaphysics, an extended physics, biology and so on, the *Système* also offers ethical, political, legal and theological theories. The title of its second edition (1691) seems fully justified: *Cours entier de philosophie ou système général selon les principes de Mr. Descartes*.

THOMAS M. LENNON

List of works

Régis, P.-S. (1690) *Système de philosophie contenant la logique, la métaphysique, la physique et la morale (System of philosophy containing logic, metaphysics, physics and morality)*, Paris; 2nd edn, Amsterdam, 1691; repr. with introduction by R.A. Watson, New York and London: Johnson, 1970.(A textbook account of the Cartesian system, but one that departs from the views of Descartes in a number of crucial respects, for example in holding that individual minds are modes of a single substance.)

Régis, P.-S. (1691) *Réponse au livre qui a pour titre P. Danielis Huetii censura philosophiae cartesianae (Response to book which has as its title P. Daniel Huet's critique of philosophy)*, Paris.(A defence of Cartesianism against the sceptical attack of Pierre-Daniel Huet.)

Régis, P.-S. (1692) *Réponse au réflexions critiques de Du Hamel (Response to the critical reflections of Du Hamel)*, Paris.(Important for Regis' views on, among other things, the creation of the eternal truths.)

Régis, P.-S. (1694) *Première réplique á…Malebranche [Seconde… Troisième…] (First reply to Malebranche [second… third…])*, Paris.(Polemical works directed against Malebranche on three topics: the nature of ideas, the problem of the horizontal versus meridional sun, and the nature of sensual pleasure.)

Régis, P.-S. (1697) 'Réflexions sur… Leibniz… ' (Reflections on Leibniz), and 'Réflexions pour… réplique…' (Reflections for… response…), *Journal des savants* 17 June and 16 November.(A defence of Cartesianism against Leibniz's criticisms with respect to religion and piety. Compare Leibniz's *Die philosophischen Schriften*, ed. C.I. Gerhardt, Berlin: Weidmannsche Buchhandlung, vol. 4, pages 333-42.)

Régis, P.-S. (1704) *L'usage de la raison et de la foy ou l'accord de la foy et de la raison (The use of reason and faith or the relation of faith and reason)*, Paris.(Important for the author's views on causation, immortality, the mind-body connection, as well as the relation between faith and reason; a refutation of Spinoza is appended.)

References and further reading

Rodis-Lewis, G. (1993) 'Régis', in R.W. Meyer (ed.) *Friedrich Ueberwegs Gundriss der Geschichte der Philosophie (Friedrich Ueberweg's outline of the history of philosophy)*, Basel: Verlag Schwabe, 431-9.(The most comprehensive account of both the primary and the secondary literature.)

Watson, R.A. (1988) *The Breakdown of Cartesian Metaphysics*, Atlantic Highlands, NJ: Humanities Press, 89-93. (Sees Régis as sharing the Cartesian failure to reconcile dualism with causal and epistemological likeness principles.)

Sa skya Paṇḍita (1182-1251)

The philosophical importance of Sa skya Paṇḍita (Sagya Paṇḍita) lies in his clarification of the tradition of logic and epistemology established by Dharmakīrti. He actively promoted the study of Dharmakīrti's thought in Tibet as a propaedeutic to the study of other systems of Buddhist philosophy as well as to a Buddhist account of knowledge; knowledge is a crucial element in the Buddhist tradition, for ignorance is considered the main obstacle to liberation, the summum bonum *of the tradition. Like Dharmakīrti, Sa skya Paṇḍita held that the only two types of knowledge are perception and inference. Perception presents us with real individual objects, while inference enables us to consider these individuals in a conceptual way, in terms of universals; however, it is a mistake to regard these universals as real.*

Sa skya Paṇḍita (Sagya Paṇḍita), whose name is often shortened to Sa-paṇ, is a dominant Tibetan figure whose accomplishments go well beyond the domain of philosophy. Politically, he was the first ruler of a unified Tibet after the collapse of the Tibetan empire in 842. He was chosen by the Mongolian Prince Godan to govern Tibet under Mongolian supervision, and was summoned to the Mongolian court, where he spent his last years. Religiously, Sa skya Paṇḍita was the fourth leader of the school named after the monastery of Sa-skya in southwestern Tibet. He played an important role in bringing the monastic and Tantric aspects of Tibetan Buddhism together, furthering Atīśa's attempt to harmonize these two traditions by describing the practice of Tantra from a monastic perspective in his *ldom gsum rab byed (Differentiation of the Three Vows)*.

Sa skya Paṇḍita is important philosophically for his contribution to the tradition established by Dharmakīrti, which was introduced into Tibet by rNgog lōtsa ba Blo ldan shes rab (Ngok lodzawa loden shayrap, 1059-1109) and Phya pa Chos kyi seng ge (Chaba Chögyi sengge, 1109-69). Through these two thinkers, an important indigenous tradition of logico-epistemological study developed, based on realist principles. Its orientation was actually quite different from Dharmakīrti's antirealist philosophy, as Sa skya Paṇḍita discovered while studying with the recently exiled Indian pandit, Śākya Śrībhadra (1127-1225). To expose and correct the revisionism of rNgog, Phya pa and their followers (among which the later dGe-lugs tradition is included), Sa skya Paṇḍita wrote his famous *Tshad ma rigs gter (Treasure on the Science of Valid Cognition)*.

Sa skya Paṇḍita interprets Dharmakīrti as espousing a conceptualist view of universals, in opposition to Phya pa's realist interpretation, which allows for the existence of real universals. In the Buddhist tradition, reality is composed of momentary and causally effective objects set apart from the realm of unreal conceptual entities. For Sa skya Paṇḍita, only individuals can be real, since an entity can be considered real if, and only if, it has its own distinctive essence (*rang bzhin*; *svabhāva*). The term 'specifically characterized phenomenon' (*rang mtshan*; *svalakṣaṇa*) used by Dharmakīrti indicates quite well the individual nature of real entities. For Sa skya Paṇḍita, the distinctive essence corresponds to clear identity conditions, spelled out in terms of definite spatiotemporal location. Only things clearly located in space and time are real. Tested this way, individuals pass, while universals do not.

Abstract entities such as properties or universals are described by Dharmakīrti as 'generally characterized phenomena' (*sāmānyalakṣaṇa*; *spyi mtshan*). They are conceptually superimposed onto reality and hence are without the power to produce an effect. This does not mean, however, that they are completely nonexistent. Although less than real, universals have a minimal ontological status as objects of concepts. They are quasi-entities, whose existence does not rest on any extra-mental reality and is presupposed by thought.

Sa skya Paṇḍita's epistemology reflects his parsimonious ontology. Following Dharmakīrti, he asserts the existence of only two types of knowledge, perception (*pratyakṣa*; *mngon sum*) and inference (*anumāna*; *rjes dpag*). These two types of cognition are distinguished on the basis of their objects and their modes of operation: whereas perception is nonconceptual and relates to real objects through experience, inference apprehends unreal conceptual constructs on the basis of reasoning.

It would be a mistake, however, to think that Sa skya Paṇḍita's view of perception is that of a straightforward empiricist who would hold perception to be cognitive by itself. For Sa skya Paṇḍita, we do not know things only by sensing them, for we need concepts to identify them. Since perception is nonconceptual, it cannot deliver fully

articulated objects, but only initiates a contact that becomes cognitively relevant through categorization. Perception passively holds an object and induces the appropriate conceptual judgement, thereby creating a fully cognitive situation. Whether or not the perceiver adequately ascertains the situation depends less on the perception than on the perceiver’s conceptual state.

The other form of knowledge is inference, which is conceptual. This is the more active form of knowledge, which categorizes the objects perceived and synthesizes them on the basis of a reason. In this way it brings signification to the bare data apprehended by perception, and extends knowledge beyond the purview of the evident. Unlike perception, conception does not mirror the reality of individuals, but relates to universals, which it mistakenly projects onto the former. In this way, conceptuality constructs the universe of meaning in which humans live. Sa skya Paṇḍita’s conceptualist and constructivist bent also concerns language, which like thought is based on conceptual activity.

The process of conceptual construction, which characterizes thinking and language, is negative in nature: conceptual constructs are formed by eliminating their contradictories. For example, the concept of tree is constructed by eliminating the non-tree. This process requires thought to create a dichotomy between a conceptually constructed property (being a tree) and its contradictory (not being a tree). By eliminating this contradictory, a property is constructed. Such a quasi-entity is just an elimination of others (*anyāpoha*; *gzhan sel*); it does not exist in reality and is merely superimposed by thought. In this way, the strict distinction between the real and the conceptual is maintained.

See also: Momentariness, Buddhist doctrine of §§1, 5; Nominalism, Buddhist doctrine of; rGyal tshab dar ma rin chen; Tibetan philosophy; Universals, Indian theories of §3

GEORGES B.J. DREYFUS

List of works

Sa skya Paṇḍita (1182-1251) *ldom gsum rab byed (Differentiation of the Three Vows)*, Complete Works of the Great Masters of the Sa sKya Sect, Tokyo: Toyo Bunko, 1968, V.297.1.1-323.2.6.(This work, which has been central to the Tibetan tradition, seeks to conciliate exoteric and esoteric aspects of Buddhism. It also contains a fierce polemic against the views of some of Sa skya Paṇḍita’s contemporaries.)

Sa skya Paṇḍita (1182-1251) *mkhas pa rnams ’jug pa’i sgo (The Door of the Wise People)*, Complete Works of the Great Masters of the Sa sKya Sect, Tokyo: Toyo Bunko, 1968, V.183.3.4- 154.4.6.(Sa skya Paṇḍita’s statement concerning Buddhist scholarship and its philosophy.)

Sa skya Paṇḍita (1182-1251) *thub pa’i dgongs pa’i rab gsal ba (Revealing the Thought of the Muni [the Buddha])*, Complete Works of the Great Masters of the Sa sKya Sect, Tokyo: Toyo Bunko, 1968, V.1.1.1-50.1.6.(A presentation of the main practices of the Buddhist path, an example of Gradual Path (*lam rin*) literature.)

Sa skya Paṇḍita (1182-1251) *tshad ma rigs gter (Treasure on the Science of Valid Cognition)*, Complete Works of the Great Masters of the Sa sKya Sect, Tokyo: Toyo Bunko, 1968, V.155.1.1-167.1.6.(A masterful summary in mnemonic verses of Dharmakīrti’s logico-epistemological system.)

Sa skya Paṇḍita (1182-1251) *tshad ma rigs gter rang ’grel (Auto-commentary to the Treasure on the Science of Valid Cognition)*, Complete Works of the Great Masters of the Sa sKya sect, Tokyo: Toyo Bunko, 1968, V.167.3.1-264.2.6.(Sa skya Paṇḍita’s own commentary on the previous text.)

References and further reading

Dreyfus, G. (1996) *Recognizing Reality*, New York: State University of New York Press.(A philosophical discussion of Sa skya Paṇḍita’s epistemology in the context of the Tibetan reception of Dharmakīrti.)

Jackson, D. (1987) *The Entrance Gate for the Wise*, Vienna: Arbeitkreis für Tibetische und Buddhistische Studien.(No complete study of Sa skya Paṇḍita’s work is available, but this is the best historical overview. Also contains a translation of part of Sa skya Paṇḍita’s work on scholarship.)

Jackson, D. (1990) ‘Sa-skya Paṇḍita the "Polemicist": Ancient Debates and Modern Interpretations’, *Journal of the International Association of Buddhist Studies* 13 (2):17-116.(A sympathetic view of Sa skya Paṇḍita’s polemical activities.)

Jackson, D. (1994) ‘The Status of Pramāṇa Doctrine according to Sa-skya Paṇḍita and Other Tibetan Masters: Theoretical Discipline or Doctrine of Liberation?’, in T. Skorupski and U. Pagels (eds), *The Buddhist Forum*, London: School of Oriental and African Studies, vol. 3: 85-129.(A discussion of Sa skya Paṇḍita’s view of the

religious implications of Dharmakīrti’s works.)

Jackson, R. (1982) ‘Sa-skya Paṇḍita’s Account of the bSam yas Debate: History as Polemic’, *Journal of the International Association of Buddhist Studies* 5: 89-99.(A particular example of Sa skya Paṇḍita’s polemical activity.)

Kuijp, L. van der (1983) *Contributions to the Development of Tibetan Epistemology*, Wiesbaden: Franz Steiner. (An exploration of some of the philosophical and historical details of Sa skya Paṇḍita’s epistemology.)

Saadiah Gaon (*fl.* early 10th century)

Saadiah Gaon al-Fayyumi was the first systematic philosopher of Judaism and a pioneering exegete, grammarian, lexicographer, liturgist and chronologist. His Kitab al-mukhtar fi ’l-amanat wa-’l-’i‘tiqadat (Book of Critically Chosen Beliefs and Convictions) uses reason, experience and tradition to elaborate a monotheistic theology and pluralistic ethics.

Organized in ten thematic treatises, the work, familiarly known by the title of its Hebrew translation, Sefer Emunot ve-De‘ot (The Book of Beliefs and Convictions), opens with a striking epistemological prelude laying out the sources of knowledge in sense experience, reason and (for the recipients of Scripture) tradition. Saadiah defends sense experience on the grounds that scepticism is self-undermining. He defends reason as the basis of critical and scientific knowledge; and he defends the Jewish sources of traditional learning on the grounds of the continuity and trustworthiness of their transmission. He treats tradition not as an independent source of knowledge but as a means of preserving primary knowledge acquired in the past.

The ten treatises of the work defend creation against alternative cosmological theories, argue for God’s unity and incorporeality, explain the human situation as a trial designed to test and reward human goodness, defend God’s justice and the substantiality and immortality of the soul, affirm the national restoration promised by the prophets of Israel, and lay out the constituents of the good life, which Saadiah argues is undermined by excessive attention to any one of the varied goods available to us. Committed to reason, science, free will and God’s ultimate justice, Saadiah champions the veracity of Scripture using his formidable philological skills and learning to find appropriable interpretations of biblical language whenever the apparent textual sense to be ruled out by reason, science, another text, or a sound tradition. In keeping with the Rabbinical and Biblical outlook, he upholds the value of this life on the grounds that only here are authentic choices possible. Saadiah’s philosophy profoundly influenced Maimonides and later Jewish thinkers; his biblical commentaries are still consulted for their philosophical and philological insights.

1 Life and works

Born in Egypt well and educated in Scripture and Rabbinic law, Saadiah published his first Hebrew-Arabic lexicon in 913, expanding it gradually to over one thousand entries. He sought a philosophical correspondence with the Jewish Neoplatonist Isaac Israeli. His polemics against the Karaites evidently led to his removal from Egypt. Living in Palestine, Iraq and Syria, he deepened his knowledge of history, philosophy and Scripture, studied with Abu Kathir Yahya al-Katib of Tiberias, and absorbed the ideas of the Jewish philosopher/*mutakallim* (dialectical theologian) Daud al-Muqammas and the Masoretes of Tiberias. His writings show a thoughtful acquaintance with the Arabic texts of Plato and Aristotle.

Gaining repute for his successful defense of the traditional Hebrew calendar against proposed alterations, Saadiah was appointed an *alluf* (master) of the ancient Talmudic academy of Pumpedita, now, like its sister academy of Sura, relocated in Baghdad. In 928 the Exilarch David ben Zakkai made Saadiah head of the Sura academy, with the traditional title Gaon (eminence). Ben Zakkai was not put off but only piqued when warned that Saadiah seemed to fear no one. By 930 the two men were at loggerheads over Saadiah’s refusal to sign a testamentary judgment in which ben Zakkai had improperly awarded himself a fee from the proceeds. Bans, depositions, factions and even riots ensued.

Refusing a bribe from a jealous rival, the Khalif assigned the case to ‘Ali ibn ‘Isa, ‘the good wazir’, as ‘Ali was known, restored Saadiah to office and was seeking to reconcile the rival leaders when the Khalif was killed in a coup d’etat and ‘Ali’s government was prorogued. When the new and impecunious Khalif al-Qahir ascended the throne, Saadiah was deposed. At length a reconciliation was achieved, consummated in a moving ceremony in 937. The anathemas were withdrawn and Saadiah restored to office with approval of the new Khalif al-Radi and the reinstated ‘Ali ibn ‘Isa.

Deprived of judicial authority for seven years, Saadiah pursued scholarship and philosophy. His commentary in 931 on the Kabbalistic *Sefer Yetzirah (Book of Creation)* adopts a scientific cosmology. Like the Muslim savants of his day, he knows that the earth is round; he has no use for a flat-earth cosmology and refuses to prop it up

against Scripture. Philosophy and history, he urges, are the proper work of man, in which God will aid us, ‘disclosing deep things’ (Job 12: 22). By 933 Saadiah completed his chief philosophic work, opening with a critical epistemology, a constructive, rationalistic empiricism, buttressed by a subdued Platonism and enriched by a chastened traditionalism that relies on trustworthy ancestors for their histories and hermeneutics but does not treat tradition as a source of knowledge independent of reason and experience.

Saadiah finds little use for the Neoplatonic ‘active intellect’ (see Neoplatonism). Yet, his naturalism and rationalism do not exclude mysticism. Prophets and the blessed derive comfort and inspiration from a ‘created light’ that is no rival to God’s absoluteness but allows the immanence of divine action. Anchoring what would later become a central Kabbalistic tenet, reciprocity between the human and the divine, Saadiah systematically glosses ascriptions to God of emotions like yearning, satisfaction and joy, as transferred epithets whose logical subject is a human being. The approach finds consummate expression in the liturgical phrase applied to Jacob, and thus to all Israel: ‘whom Thou didst love with Thine own love and rejoice with Thine own joy’. Spinoza later uses the same idea in explaining how the Infinite can care for finite individuals.

Saadiah’s books on language, surviving only in fragmentary form, include the earliest known Hebrew grammar. He composed the first scholarly *Siddur* or Hebrew prayerbook. Most of his liturgical poetry is lost, but his didactic poem on the Ten Commandments and his penitential and petitionary prayers, found in the Cairo Geniza, convey the flavour: highly allusive writing of the philological type favoured in his day. His prose prayers are more straightforward, and his Arabic prayers and translations of Hebrew prayers reveal his expressive powers.

Saadiah’s Arabic Bible translations and commentaries assign each book a thematic title and include an introduction explaining the book’s problematic and complementing the linear commentary to reveal the higher order argument. The commentaries are openly philosophical and typically fight shy of midrashic embroideries.

Most of Saadiah’s *halakhic* contributions are lost, but in the fragmentary corpus that survives, thematic introductions and conceptual organization reveal the role of philosophy in structuring Saadiah’s thought. Saadiah’s lost *Kitab al-ta’rikh (Chronology)*, apparently was intended as a summary of history from the creation. As Franz Rosenthal (1968: 139) showed, the work represented the movement towards linear historiography that was ongoing in Saadiah’s time. Saadiah’s goal was to situate Biblical and other events chronologically. His exegetical writings show a similar interest in geographical orientation.

2 *The Book of Critically Chosen Beliefs and Convictions*

Saadiah’s philosophical *chef d’oeuvre* is commonly known by the Hebrew title, *Sefer Emunot ve-De‘ot (The Book of Beliefs and Opinions)*. But in what seems to be his final revision it bears the title *Kitab al-mukhtar fi ’l-amanat wa-’l-’i‘tiqadat (The Book of Critically Chosen Beliefs and Convictions)*, ‘And rightly so’, Joseph Kafih explains. ‘For our teacher did not set out simply to gather a compendium of beliefs and convictions, but to demonstrate which beliefs were worthy of choice and which convictions were true in his estimation’ (Kafih 1970: 1). Like Aristotle and Plato, Saadiah surveys and criticizes rival views, choosing one to be accepted. As in *kalam* (dialectical theology), he answers objections and arrays arguments, scriptural and rational, against the rejected views and in support of the view he deems orthodox; but the enterprise is never purely apologetic, since what passes muster cannot remain unaffected by the demands of critical scrutiny.

The work comprises ten ‘treatises’. The introduction lays out Saadiah’s epistemology. Doubt is a natural concomitant of our finitude but can be overcome by subduing its causes, ignorance and impatience. Subjectivism is absurd, since opinions do not determine reality; nor does disbelief exempt us from our obligations. Saadiah defends perceptual knowledge, arguing that systematic scepticism about perception itself is self-destructive. He goes on to urge that the methods needed to render perceptions trustworthy lead inevitably to general theories and thus to the sciences. Only the superstitious forbid speculation, fearing for the faith. Such fears are as irrational as the fantasy of the ignorant that whoever travels to India will grow rich.

Treatise 1 defends creation as the bulwark of theism, warning against attempts to explain the ultimate Cause found by reason. ‘Explaining’ the ultimate in terms, for example, of sensory phenomena would reduce God to the very facts that creation was invoked to explain. The revolution of the heavens proves the cosmos finite; its finitude and compositeness, the temporality of all accidents in nature, the inability of finite particulars to cause their own existence, all prove the world to have been created. True, nothing comes from nothing, but for that very reason,

creation provides the best explanation for nature’s existence.

Neoplatonic attempts to derive the physical from the ideal explain the obscure by the more obscure and bring matter and spirit no closer together than does the sheer Biblical creation. Materialism, by contrast, hides a tendency to confuse cause with effect. Our aim is not to deny causality; but proximate causes are just part of the cosmic system. They are insufficient without the ultimate cause. Thus, when Aristotelians ascribe the cosmic order to the celestial motions, Saadiah emphasizes that circular motion is natural to the heavenly bodies, and is thus part of what we are seeking to explain.

Against the idea that chance is the ultimate cause, Saadiah argues that chance could not produce a stable system. Besides, chance can be defined only relative to a natural order; so it is incoherent to treat it as the ultimate cosmic principle. The idea that nature has always been as we observe it, by contrast, extrapolates our experience to the wholly unknown.

Treatise 2 derives God’s unity from his incorporeality as the absolute creator, from his oppositeness to the world’s multiplicity, and from the economy of explanations: one cause is sufficient, more would be redundant and would require proof beyond the sole proof we have, the act of creation. Dualists and polytheists have no way of limiting divinities once they begin; in the end they cannot help making a god of every element or principle. If God needed help or cooperation to make or rule the world, he would be powerless; and if some other god were not his aide but free to contradict him, the two would either limit one another’s power, making neither divine, or overrule each other - so that the same object, for example, could have contradictory characteristics.

As for attributes, God’s life, power, knowledge and transcendent goodness are known from the act of creation. However, God’s attributes represent a single reality; they differ only in our understanding. Thus they are figures, like the Biblical anthropomorphisms, not to be taken literally. Paradigmatically, the ascription of loves and hates to God expresses normative intentions. God’s speech is a created sound, and God’s ‘back’, as seen by Moses, is his created glory.

God created the world, Treatise 3 explains, to allow humans to earn blessedness. Earned desert is far more precious than mere bliss, but the effort to earn blessedness entails real risks: trials of our mettle, accountability for our choices, sufferings that may be warnings or chastisements, or ‘sufferings of love’, whose purpose, which we cannot know when we undergo them, is enhancement of our reward in the hereafter through recompense for preserving our integrity. The chief vehicle of our test, for which the world was created, is the system of our obligations. The most basic of these are well known to reason, as when we recognize the wrongfulness of causing bloodshed or pain, fornicating, stealing or lying. But many of the commandments are known to us only because they are revealed. Their performance is our obligation not because of any intrinsic value in the acts they involve but because they function symbolically, as expressions of gratitude, a fitting means of honouring our Creator.

The rational commandments are not derivable from hedonism or from any merely empiric naturalism. Hedonism will make the same act, such as a theft, both good and evil, since it brings pain to the victim and pleasure to the thief. But if we recognize the need to differentiate ourselves from animals (and so do not fornicate), understand that misrepresentation is a grotesque perversion of creation (and so do not lie), and see that bloodshed thwarts fulfilment of God’s plan (by violating the potentials God imparted to be realized), we discover the values underlying some of the precepts of the Law. We can even find the rational basis for the ritual commandments, those that would have had no strict standing as obligations had they not been commanded. Reason demands a response to generosity, and all the ritual commandments, although they may yield benefits (Sabbath rest, Levitical purity), are ultimately expressions of gratitude. Such laws, then, have a purpose, but that purpose alone does not define, but as we might put it today, underdetermines their content. Prophets are needed to spell out God’s expectations. Miracles corroborate the claims of prophets, and tradition preserves their message, vouching for its authenticity, but also interpreting it. Just as reason is prior to revelation, tradition is posterior to it; none of the three can stand alone.

Treatise 4 presents man as the object of creation, at the centre of the cosmos, endowed with moral freedom. Our bodies are small, but are the best that could be given mortal beings; and the human soul is vaster than the cosmos, which its knowledge embraces. That cosmos was created in our interest, not for our pleasure. Life is short, but the choices made while choice is possible are of absolute significance. As with any growth, dead wood must be

cleared away. Thus there is capital punishment in this world and hellfire in the next, where the very light that comforts the blessed torments the damned. God gives us the capacity to act and choose, but some choices may diminish our degrees of freedom. What God forms is our underlying nature; our character is our own work.

We are judged, Treatise 5 teaches, by the preponderance of our good or evil actions, whose inner worth God knows irrefragably. Reconciling virtue ethics with command ethics, Saadiah finds each act significant, but insists that one act is not our character. Thus penitence is possible, fulfilling regret, just as action fulfills intent. The weight of our choices, however, precludes redemption of every act: once choices have sealed our character they have sealed our destiny, and even penitence becomes impossible. Prayer goes nowhere when it is insincere or intransigent; it is bootless in one who is actively neglectful of the Torah or the poor, mired in embezzlement or impurity. Three sins will not be expiated: slander, misleading others and retention of ill-gotten gains. Three merits are rewarded in this world, even for those who reject God’s service: filial piety (Exodus 20: 12), kindness to animals (Deuteronomy 22: 7) and honest dealing (Deuteronomy 25: 15). Like Maimonides after him, Saadiah derives not merely the commandment (and its reward) but the generalized theme of each precept: not simply releasing the mother bird, but kindness to animals; not simply fair weights and measures, but honest dealing.

The soul, Treatise 6 argues, is created on completion of the body with which it is united. Neither soul nor body is impure; sin results only from our own wrong choices. Like the heavenly spheres, the soul draws luminosity from God, gaining life and consciousness, which allow it to animate a body that would otherwise be passive and inert. Once its destiny is complete, it returns to God, who allowed it to act through the body’s intermediacy. When the tally of souls God destined for existence is complete, all are reunited with their bodies and judged. Those whose lives were cut short or who suffered undeservedly are recompensed, including the slain infants of Israel’s ancient conquest, and even any animals that suffered unduly in the Temple sacrifices.

The soul is not an accident; it is not a function of the body or adjunct of the blood. Nor is it a self-moving number, an entelechy or an epiphenomenon. The world was not created for the sake of an accident. The soul is not made of fire or air, for it lacks their qualities. It is not of two parts locatable in the head and heart, for the soul would be what enables these to interact. Nor are there three separate souls, as suggested by Plato. Appetite, ire and reason are faculties of one soul, called alive in virtue of the immortality to which it is heir. The demands of theodicy may give colour to the theory of metempsychosis, but understanding that God’s grace and justice assure us of recompense for all unrequited sufferings and retribution for all unpunished wrongs deflates the appeal of the notion of transmigration.

Treatises 7-9 differentiate resurrection, redemption and requital. Resurrection reunites body and soul here in the world. All monotheists will share in it, but Israel will have the leading role because of her long sufferings. God did not include resurrection in the first redemption, the exodus from Egypt, but promised it for the future, because Israel’s present bondage is heavier than the Egyptian enslavement.

Redemption is the historical vindication of God’s promises to Israel: the ingathering of her exiles, the return of prophecy and the restoration of the Davidic monarchy. But the ultimate requital must be otherworldly, to respond to transcendent human goodness and suffering, sin and cruelty. In this world, Saadiah argues, following Al-Razi, pains outweigh pleasures, the wicked often triumph, and suffering innocents are not requited. God’s mercy must remedy life’s defects. Were it not for future requital, fire and brimstone would surely have fallen on the earth long ago, as they did on Sodom and Gomorrah.

But the hereafter is not an earthly place. Only metaphorically is it called Tofet or Eden. Time itself will be transmuted in the new heaven of which Isaiah spoke. The most striking transcendence will be moral: our trials over, there will be no more need, or chance, for choices, but the infinite consequence of our decisions in this world will be played out to eternity.

3 The good life

Treatise 10 concerns the good life, defined in moral terms, for we do not know the reward of the ritual commandments even in this life, still less in the hereafter. Saadiah’s eudaimonism is pluralistic, based on the complexity of human nature. Like Plato, he sees the good life as a balancing of interests (see Good, theories of the §5). He does not break these down into intellectual, appetitive and spirited, but elicits his own list from Scripture and from observation of human nature: abstinence, eating and drinking, sex, passionate love, wealth, progeny,

agrarian and urban development, longevity, power, vengeance, knowledge, worship and rest. Each of these (even vengeance) is in some sense a good, but none, as devotees might imagine, offers a fulfilling life. To make one of them our sole object is to cheat ourselves of the rest.

Self-denial is a valuable discipline, but the pure ascetic is misanthropic and embittered; his isolation feeds his envy and deprives him of the piety he may have sought. Food and drink sustain body and mind and foster reproduction, but the gourmand is bloated, unhealthy, selfish, foggy-headed and licentious. Sex is a unique delight, countering melancholy, cementing human relations and well accepted by the prophets, but the lascivious are unhealthy and typically adulterous. Passionate love has its place, sustaining the marital relationship; but, as a way of life, it is an absurd obsession, often a source of regret or hatred. Progeny perpetuate the world and give solace and joy, but offspring are also a hardship and source of anxiety; they are not sufficient to give meaning to our lives. Development is useful and satisfying; but taken beyond human need and made our overriding object, it distracts us from intellectual and spiritual goals and becomes a source of anxiety and compulsion. Longevity too is a means to an end, allowing us to attain our spiritual and worldly goals, but the valetudinarian must know that even the vigorous often die young, and there are higher goals than maintenance of the body.

Power or authority is necessary to the ordering of the world; but, if made all-sufficient, it promotes arrogance and injustice and becomes self-destructive, transforming a ruler’s ebullience from overconfidence to the tyrant’s suspicions and hatred of humanity. Vengeance is the most specious of *prima facie* goods. It gives momentary satisfaction, but the scheming it engenders fosters anxiety and ruthlessness. Hatred cannot be made a way of life. Vengeance may spur us in pursuit of justice, but vengeance is not justice. Like all other *prima facie* goods, it becomes an actual good only when mitigated by the rest.

Knowledge is a good, but even the quest for knowledge can be excessive; pursued to the exclusion of all else, our appetite for knowledge would ruin our health, even dull our mind. Worship is a fitting expression of gratitude; but, as an exclusive goal, it is self-undermining. How can we fulfill the commandment to keep just weights and measures if we eschew all social engagement and economic activity? Pietists may aspire to devote themselves wholly to God, but their neglect of their bodies and their offspring cannot be condoned. Finally, rest is needed, and is prescribed for sabbaths and holy days. But rest is possible and valuable only through work; laziness is destructive. Rest imparts a taste of the world to come, but our obligations are for this world.

Saadiah underlines his integrative pluralism with a brief discussion of aesthetics. Blending, he argues, is the key to beauty. Tastes, colours, sounds, even smells are beautiful when mingled; rough, even injurious, when left simple. All goods - and Saadiah acknowledges more *prima facie* goods than he has listed - have their proper place and context, and this can be found, if one is not simply rationalizing one’s appetites and desires but inquiring after the truth with humility and sincerity. The Torah lays out a way of life that enables us to balance the goods pertinent to our nature, allowing none to usurp the place of reason.

See also: Bible, Hebrew; Good, theories of the

L.E. GOODMAN

List of works

Saadiah Gaon (early 10th century) *Kitab ’usul al-shi‘r al-‘ibrani (Ha-Egron, The Book of the Roots of Hebrew Poesy)*, ed. N. Allony, Jerusalem: The Academy of the Hebrew Language, 1969.(This is Saadiah’s dictionary, presented here in a modern, critical edition.)

Saadiah Gaon (early 10th century) *Kitab al-mukhtar fi ’l-amanat wa ’l-’i‘tiqadat (The Book of Critically Chosen Beliefs and Opinions)*, trans. J. Kafih, *Sefer ha-Nivar ba-Emunot uva-De‘ot*, Jerusalem: Sura, 1970; trans. S. Rosenblatt, *The Book of Beliefs and Opinions*, New Haven, CN: Yale University Press, 1948.(Kafih provides Arabic text with a modern Hebrew translation. Rosenblatt’s English translation is in need of reworking.)

Saadiah Gaon (early 10th century) *Kitab al-ta‘dil (The Book of Theodicy)*, trans. L.E. Goodman, New Haven, CN: Yale University Press, 1988.(Saadiah’s commentary and translation of the Book of Job.)

References and further reading

Cohen, B. (ed.) (1943) *Saadiah Anniversary Volume*, New York: American Academy for Jewish Research. (Includes papers by Salo Baron, Solomon Gandz, Harry Wolfson, A.S. Halkin and others.)

Efros, I. (1974) *Studies in Medieval Jewish Philosophy*, New York: Columbia University Press, 1-137.(These

pages represent a monograph on Saadiah’s philosophy.)
Finkelstein, L. (ed.) (1944) *Rab Saadia Gaon*, New York: Arno Press; repr. 1988. (Symposium volume, includes papers by A. S. Halkin, Ben Zion Bokser, Robert Gordis and others.)
Goodman, L.E. (1976) ‘Saadya Gaon on the Human Condition’, *Jewish Quarterly Review*, new series 67: 23-9. (Explains Saadiah’s views about the prevalence of suffering in this life and the warrant he gives for deriving an afterlife from the recognition of God’s justice.)
Goodman, L.E. (1990) ‘Saadiah Gaon’s Interpretive Technique in Translating the Book of Job’, in *Translating Scripture*, *Jewish Quarterly Review* supplement, Philadelphia, PA: Annenberg Research Institute, 47-76. (Analyses Saadiah’s exegetical method with concrete examples of their application and explication of their philosophical and hermeneutical import.)
Goodman, L.E. (1996) *God of Abraham*, New York: Oxford University Press.(Chapter 6, ‘Monotheism and Ritual’, contains a detailed examination and defence of Saadiah’s treatment of the ritual commandments, arguing against the widely held view that Saadiah differentiated these, or any class of *mitzvot*, as non-rational.)
Heschel, A.J. (1944) *The Quest for Certainty in Saadia’s Philosophy*, New York: Feldheim.(Early study of Saadiah’s epistemology.)
Katz, S. (ed.) (1980) *Saadiah Gaon*, New York: Arno Press.(A collection of classic papers in English, French and German, by David Neumark, Georges Vajda, Alexander Altmann and others.)
Malter, H. (1969) *Saadiah Gaon: His Life and Works*, New York: Hermon Press.(Reprint of the edition of 1926.)
Neuman, A.A. and Zeitlin, S. (eds) (1943) *Saadiah Studies*, Philadelphia, PA: Dropsie College.(Includes the epistemological studies of Heschel and Efros, along with papers by Solomon Skoss, Harry Wolfson, Louis Ginzburg (on Saadiah’s *Siddur*) and the two editors.)
Rosenthal, E.I.J. (1943) *Saadya Studies*, repr. New York: Arno Press, 1980.(Includes papers by Alexander Altmann, Chaim Rabin, Simon Rawidowicz and the editor, among others.)
Rosenthal, F. (1968) *A History of Muslim Historiography*, Leiden: Brill, 2nd revised edn.(This general work on the development of Islamic writing about history contains a brief discussion of Saadiah’s lost work on history, situating it against the background of Islamic works of world history and chronology.)
Skoss, S. (1955) *Saadia Gaon, The Earliest Hebrew Grammarian*, Philadelphia, PA: Dropsie College Press.(A detailed analysis of Saadiah’s pioneering work in the grammar of the Hebrew language.)
Vajda, G. (1967) ‘Autour de la Théorie de la Connaissance chez Saadia’ (On Saadiah’s Theory of Knowledge), *Revue des Études Juives* 126: 375-97.(Classic investigation of Saadiah’s epistemology.)
Ventura, M. (1934) *La Philosophie de Saadia Gaon*, Paris: Vrin.(A detailed commentary on the theses and arguments of Saadiah’s *Book of Critically Chosen Beliefs and Convictions*.)

Sacraments

The Christian theory of 'sacraments' underlies ideas of a general 'sacramentality' in the universe whereby ordinary things have religious significance by their own nature or by virtue of some hidden power within them. The pre-Christian Latin word sacramentum *meant a non-returnable gift marking the taking on of some binding obligation; more informally it meant an oath, and later a secret or mystery. Latin theology turned it to Christian use, initially in rough translation of the Greek* mysterion, *applied to the Church, to the Scriptures and to Old as well as New Testament rites. The word then became the predominant medieval and modern term specifically designating those rites in permanent use in the Church which human authority was conceived not to be free to abolish, add to or change in their essentials.*

Each such rite presupposes that the creaturely things used have some aptitude which allows or invites the particular ritual use concerned, that is, which presupposes some more general sacramental potential in natural things. The conceptual tools developed in Catholic theology - 'effective sign', 'matter and form', 'sacrifice', 'authority', 'power' and 'institution' - sharpen enquiry into the phenomenology of rituals within many different religious traditions.

1 Explaining early conceptions: the 'seven sacraments'

The so-called seven sacraments existed in the Church long before being classified together as 'sacraments' to distinguish them from other Church rituals, and long before any discussion of their precise number or what they had in common - what it is to be a 'sacrament'.

The Fourth Lateran Council (1215) set forth the Western Catholic conception of seven sacraments: two of initiation - Baptism and Confirmation; the Eucharist, the central action supposed constitutive of the Church, around which the others are grouped (commonly referred to as the Mass or the Divine Liturgy, or, among Protestants, Holy Communion or the Lord's Supper); two of healing - Reconciliation (absolution from sin, commonly referred to as Confession or Penance) and the Anointing of the Sick (Extreme Unction); Ordination, conveying authority to celebrate the Eucharist and to teach; and Marriage, which is unique as the context of cooperation with God in his creation of new persons, and gives grace enabling marriage to imitate the mystery of the love between Christ and the Church. The Eastern Orthodox Churches generally agree in the same list. Protestants tend to similar conceptions of which rites are integral to the life of the Church, but regard only baptism and the Eucharist as instituted by Christ himself, and so only these as sacraments in the strict sense (although Luther initially ranked confession as a third). All agree on the higher status of baptism and the Eucharist.

Catholic theological tradition saw a sacrament as essentially an action that is effective in communicating grace, the action being such as to signify what is being done. The philosophical concept of 'effective sign' defined what the seven rites had in common.

How was effectiveness viewed? God alone (acting immediately, not deputing his power) gives grace inwardly in the sacraments, with Christ himself acting as principal agent through the Holy Spirit; the ministers of the sacraments are essentially instruments. In this conception, it is primarily Christ who baptizes, absolves, celebrates the Eucharist, heals, joins in marriage, ordains and so forth - a view echoed in the Anglican archbishop William Temple's insistence that in ordination the minister acts as the representative of Christ, not of the Church. The effectiveness of the sacraments seems an immediate corollary of this conception, so that, if there is no obstacle on the side of the recipient, then the appropriately intended performance of the rite by ministers with appropriate authority must be effective for God to give grace, in the traditional phrase *ex opere operato* (that is, effective simply 'out of the work's having been worked'), although the fruitfulness of the grace is affected by the dispositions and prayer of the persons concerned.

This belief in effectiveness was what first made the notion of 'institution by Christ' pivotal, since evidently no merely human authority could arrogate to a human ritual (even granted appropriate intentions and ministers) certainty of being thus instrumental to an action of God. This consideration made Thomas Aquinas reject earlier ideas that some sacraments might have been instituted by the apostles or by the Church. He argued that such lesser authorities could no more do this than institute or alter the essential constitution of the Church. The list of

sacraments was supposedly compatible with many different explanations of how and when they were instituted. Aquinas, for example, distinguished the institution of a sacrament (the establishment of its power), its being published (perhaps by Christ instructing the apostles before his Ascension) and its being promulgated (made known to the whole world); the promise of the Holy Spirit without measure implied the completion of baptism in confirmation and the elevation of marriage, instituted at the beginning of mankind, to a sacramental means of grace. Which rites were actions of Christ and therefore effective was not held in doubt, only their mode of establishment.

The notion of ‘sign’ seems less problematic, and customarily two aspects were distinguished: matter (water, bread, wine, oil, the laying on of hands - that is, certain materials and actions with them) and form (what indicates the meaning of the actions, including the kind of minister required). But how the matter/form distinction applies in the cases of confession and marriage is obscure; where both the laying on of hands and the application of oil are used, which is essential has been disputed; and the Eucharist presents another difficulty, described in §2 below. Plainly, the form is paramount; but many obscurities arise of the kinds philosophy often finds in the areas of law and action (for example, how to distinguish the publicly recognizable context of words or actions, which determines ‘the intention of the rite’ - that is, the character of the action - from the private intentions of ministers or recipients). Problems are resolvable only by regarding the essence as lying in some action, an action by the doing of which grace is given, not in the applicability of this matter/form distinction. These distinctions, problems and approaches to resolution arise in considering the rituals of almost every religion.

The list of seven has a rationale in its relation to the universal structure of human life. However, its significance has always lain partly in what it excludes, for example, kings being anointed or crowned, monks and suchlike taking vows, places and objects being dedicated (except in the Eucharist). It also excludes all other public or private ritual postures, acts, ceremonies and processions, including those centred on sacred icons, statues or images. To distinguish them from the sacraments, they all are commonly grouped as ‘sacramentals’, along with various uses of candles, incense, ashes or water (except in baptism) - the last sometimes associated with natural springs and wells.

2 The Eucharist: presence and sacrifice

The Eucharist is described by Aquinas as the greatest of the sacraments because ‘it contains Christ substantially’. According to ancient Catholic and Orthodox tradition, what is sacramentally present and received under the forms of bread and wine is not bread and wine as such, but Christ himself, mysteriously present in his complete humanity. Receiving Christ’s body and blood is not itself the grace received, but only the means of grace. The aim is not cannibalism: Christ’s body is said to be present but ‘not in the way a body is in a place in its own dimensions’ (*Summa theologiae* IIIa, q.76, aa.3-6); therefore it is not available so as to be altered and consumed in physical nutrition. Rather the grace given is that of being ‘nourished with the divine life’, ‘divinized’, ‘made partakers of the divine nature’ (2 Peter 1: 4). Anscombe (1981) argues that it is a secondary question whether Christ, as wholly present in the sacrament, is thus present really or only symbolically or only in the act of reception; the primary datum is that in the Eucharist it is Christ himself - mysteriously present in body, blood, soul and divinity - who is the matter or means of grace, whereas in other sacraments it is only water, oil, the laying-on of hands, and suchlike. Describing the matter of the sacrament as ‘bread and wine’ is oversimple.

The philosophical difficulties in Catholic and Orthodox teaching are signalled by the absence of any positive ways of describing Christ’s mode of presence (the word ‘sacramentally’ tells us no more than the word ‘mysteriously’) or of describing the conversion of the bread and wine into Christ’s body and blood (denied to be an annihilation), beyond saying that they constitute a change in ‘reality’ or ‘substance’ (a ‘transubstantiation’). The use of the term ‘transubstantiation’ (Greek, *metaousiosis*) was ancient, recurring in the Fourth Lateran Council (1215), and not tied to Aristotelian doctrine, let alone to the contrast between essential and accidental predicates. The Council of Trent (1547-63) used the term ‘appearance’ (*species*), not the term ‘accident’, thereby implying no more than that the change concerned is knowable by faith alone.

However, such philosophical difficulties have been exacerbated by considering the aspect of presence out of its original context, namely that of a memorial (Greek, *anamnesis*). The Jewish conception of memorial is not of individuals gathered in an assembly each, in their individual minds, remembering the thing being remembered, but rather of laying this out in front of other people, the world or God - in the case of an action, it is more like

re-enacting it than doing anything merely psychological (etymologically, ‘re-member-ing’ is putting parts together again). Memorial is also the context for conceiving the Eucharist as a ‘sacrifice’, not in the modern sense of killing or mutilating something, but still in the sense of an offering, not just of ‘ourselves’, but of Christ - it involves pleading again in an unbloody manner the one bloody sacrifice of Christ on the cross. Dummett (1987) argues that the reality of the presence of the risen Christ is a condition of the objectivity of the memorial made, so we cannot make sense of the ancient understanding that the primary action of the Eucharist is the making of this memorial without registering that it was understood to require a real presence. The generic meaning of the word ‘sacrifice’ as ‘making holy’ or ‘separating from human use for divine use’ adds nothing, except in being one of a whole family of features anticipated in Jewish as well as pagan rites. In the Eucharistic memorial, Christ as having been sacrificed is presented, now risen, before the Father under the forms of food and drink; these forms mark this action as to be completed by Christ’s being received by the participants as food and drink, with all the symbolism of a shared meal and of being nourished, the participants being thereby also set apart for divine use.

Roman Catholic and Orthodox Christians reckon the Eucharist to contain, not just the pledge, but the reality of the incarnate Christ now risen, and so to represent and embody the change in the world brought about by the Incarnation - in Athanasius’ words, ‘God’s being made man in order that man might be made divine’. It was thus that the Eucharist with its many aspects first became envisaged as the Mystery constitutive of the Church.

3 Post-medieval developments

The primary feature of Protestant accounts was, while retaining the notion of outward signs of inward grace, to repudiate any *ex opere operato* effectiveness, and to make the receiving of grace wholly dependent on the faith of the recipient, the minister being the representative of the congregation rather than of Christ. Institution by Christ was now important only because obedience to explicit Scripture required it, not because the sacraments were actions of Christ in the Church. In the Protestant conception, a sacrament was a human work, of no intrinsic merit, so that efficacy lay only in faith in God, who instituted the sacrament; in the Catholic conception, by contrast, a sacrament was in itself a divine work, so that reliance on it constituted direct reliance on God rather than on one’s own psychological state.

Philosophical conceptions varied: Luther’s consubstantiation might seem only a slight variation on transubstantiation (the body of Christ co-present with the bread, as fire in molten iron, instead of taking its place), except it was conceived more as a temporary indwelling than a real change, the change not conceived as persisting after the ceremony; Calvin’s doctrine involves only a power (*virtus*) in the bread and wine, a power realized only in their being received in faith; Zwingli began by regarding the bread and wine as symbols of Christ’s body and blood, but later considered only the ceremony as a whole as significant, emphasizing its aspect as a community meal. Recent Catholic theology has added other non-realistic theories, such as transignification (criticized by Anscombe (1981) and Dummett (1987)), as well as attempts to explain how the non-empirical change of bread into Christ’s body can be ontological and real without being physical (for example, Schillebeekx 1968; Dummett 1987).

Twentieth-century Catholicism, transformed by Pius X’s drive for ‘frequent and even daily communion’ and Pius XII’s systematic emphasis on the people’s participating in more ways in every part of the liturgy, is now experiencing the resultant vernacularization initiated by Vatican II. This has dramatized the character of the sacraments as bringing people into relation to one another as well as to God. Each sacrament relates people to the Church in the same act as relating them to Christ, so that (for example) being joined or reconciled to Christ is being joined or reconciled to the Church. The Eucharist presents itself as constitutive of the Church. Vatican II situates both sacraments and sacramentals within the context of the Church as ‘Sacrament or instrumental sign of intimate union with God and of the unity of all humanity.’

4 General sacramentality of the universe

Their realistic standpoint with regard to the Eucharist, despite the absence of any satisfactory account of it, constitutes the high point in Catholic and Orthodox resistance to the desacralization of the world. In the lesser sphere of ‘sacramentals’, the numinous significance of icons for the Orthodox presents a dramatic case of the same resistance.

In the theistic perspective, God, as cause of being, immediately present in and upholding everything, is never an absentee. Therefore, within the world thus loaded with his presence and activity, there is always the possibility, through the Holy Spirit, of God's presence and activity taking forms beyond mere causation, forms richer in personal significance and power, and instanced not only in the sacraments but more widely. This perception appears in Donald Baillie's emphasis on the sacramentality of the universe (1957), and is given vivid expression in Gerard Manley Hopkins' poem 'God's Grandeur' (1877). The way God's immanence in all things makes all potentially open to religious significance has also been important for Christian understanding of very differently expressed ideas - for example, in Hinduism. It is the background for the permanent potential that nature and particular places have to arouse religious awe, a sense of the 'numinous'.

In the Catholic conception, relationship to God in any such rich mode of presence in creation is constitutive of relationship to the Church. God's giving grace is not limited to the sacraments, but the symbolism they embody signifies or explains the direction of the grace which he may give without them. What happens when a person receives grace by a sacrament is representative of what happens when the same grace is given without sacrament.

Within a dualist framework of thought, identifying the spiritual with the immaterial, the sacraments have been thought of as an accommodation of God's mode of action to the lowly body-and-imagination-bound character of the human soul, as if human worship and its expression were ideally purely verbal or intellectual. However, the sacraments can be thought of as reflecting God's express purpose in creating the physical universe; the human being's unitary psychophysical nature is not a weakness or a privation, but a source of enrichment, modes of worship include a physical aspect, and voluntary physical action often provides the most definitive expression of the human heart. Light, heat, the air, the sea, the heavens, food and water provide images pregnant with religious meaning, not to be disparaged in a false body/mind dualism, but integral to a rounded communication between God and the human race, the background to sacrament and sacramental alike.

See also: Grace; Ritual

DAVID BRAINE

References and further reading

Anscombe, G.E.M. (1981) 'On Transubstantiation', in *Collected Papers, III: Ethics, Religion and Politics*, Oxford: Blackwell.(A modern philosopher discusses a key mislocation of the problem of the nature of the Eucharist.)

Aquinas, T. (1266-73) *Summa theologiae* IIIa, qq.60-90, London: Blackfriars, vols 56-60, 1961-70; also the Supplement, qq.1-68, in *Summa theologiae*, London: Burns & Oates, 2nd edn, vols 18-20, 1941.(The Supplement was compiled after Aquinas' death from earlier writings to complete his plan for the whole *Summa*. See especially qq.64-6, 72, 75-7 and Suppl. qq.29, 41-2. Of special philosophical interest are q.75, on 'transubstantiation', and q.76 and q.77, a.7, on the peculiar mode of Christ's spatial presence.)

Baillie, D. (1957) *The Theology of the Sacraments*, London: Faber & Faber.(Part II gives an eirenic Protestant view - in Lecture 1, of the general sacramentality of the universe, in Lecture 2, of the rooting of Christian sacraments in the Incarnation and in Lecture 4, of the meaning of the Eucharist.)

Calvin, J. (1536) *Institutes of the Christian Religion*, ed. J.T. McNeill, trans. F.L. Battles, London: SCM, 1960, bk IV, chaps 14, 17-19.(This systematic Protestant masterwork treats the sacraments in general in chapter 14, the Eucharist in chapters 17-18 and lesser rites in chapter 19.)

Dummett, M. (1987) 'The Intelligibility of Eucharistic Doctrine', in W.J. Abraham and S.W. Holtzer (eds) *The Rationality of Religious Belief: Essays in honour of Basil Mitchell*, Oxford: Oxford University Press.(A modern analytic philosopher examines the difficulties of Eucharistic doctrine.)

Hopkins, G.M. (1877) 'God's Grandeur', in *Poems and Prose*, ed. W.H. Gardner, London: Penguin, 1953. (Displays the principle of Hopkins' poetry, a world whose every feature is pregnant with meaning.)

Luther, M. (1520) 'The Babylonian Captivity of the Church', *Luther's Works*, vol. 36, *Word and Sacrament II*, Philadelphia, PA: Fortress Press, 1959.(A pioneer elaboration of a Protestant view; see especially pages 18, 28-41, 51-7, 62-8.)

Schillebeekx, E. (1968) *The Eucharist*, London: Sheed & Ward.(An attempt to give a phenomenological interpretation of a realist view of the Eucharist.)

Saint-Simon, Claude-Henri de Rouvroy, Comte de (1760-1825)

An influential French social theorist, Saint-Simon propounded a philosophy of history and an account of the future organization of industrial society. He predicted a 'golden age', where harmony between individual capacities and social structures, reflected in a reordering of 'temporal' and 'spiritual' power, would overcome disorder and banish idleness. He has been variously portrayed as a utopian socialist, the founder of sociology and a prescient madman.

Saint-Simon's life was eventful, even if the more apocryphal details of the hagiography transmitted by his followers - including his descent from Charlemagne, being tutored by d'Alembert, childhood imprisonment, and marriage proposal to Madame de Staël - are omitted. As a commissioned French officer he participated in the American Wars of Independence, was wounded and imprisoned by British forces, and commended for bravery by the American government. Following his release, Saint-Simon pursued an unsuccessful canalization project in Panama before returning to France. During the Revolution, he (prudently) renounced his title, underwent a 'republican baptism', and made (and subsequently lost in litigation) a fortune speculating on the property market - his brief incarceration during the Terror was the result of mistaken identity. From 1802 Saint-Simon devoted his energies to developing and promoting his distinctive views on social organization in a torrent of books, pamphlets, prospectuses, articles and letters. The interpretative difficulties created by the diversity of these projects are compounded by the frequency of joint and disputed authorship, his opportunistic willingness to adapt his emphasis to imagined audiences as varied as the European proletariat and the French king, and the indifference to consistency that accompanied his predilection for descriptive detail. His productivity was punctuated by an unconsummated and hastily terminated marriage, a period in a private mental hospital, trial and acquittal on charges of subversion, and a suicide attempt (involving seven shots to his head) which resulted in the loss of his right eye.

In advocating a 'science of social organization', Saint-Simon depicted himself as emulating physiology in applying the methodology of the natural sciences - basing arguments on 'observed and examined facts', rather than revelation or deductions from reason - to the study of humankind. History, he claimed, was driven by the progression of the human mind, in which successive advances in understanding are embodied in a (twelve-stage) chronological narrative. In its most schematic form - as promulgated bySaint-Amand Bazard - history is structured into 'organic' epochs (in which ideas and institutions are in harmony) divided by 'critical' interludes (where social arrangements lag behind the progress of intellect). Development proceeded from the ancient and Christian worlds (grounded in polytheism and theism respectively), through the contemporary period of transition, to a final 'golden age' of industrialism (founded on positive science). The role of human agency was limited to hastening the arrival, rather than determining the content, of 'the perfection of the social order' that lay ahead.

'Social physiology' revealed inequalities in human nature, both between 'Africans' and 'Asians' (the 'descendants of Cain') on the one hand and 'Europeans' (the 'children of Abel') on the other, and within the latter category, where one of three capacities - emotive, rational and motive - would typically predominate within any particular individual. This psychology motivates the class structure of industrialism in which social differentiation according to capacity - into artists, scientists and 'industrials' (the latter group including wage labourers and owners of capital) - is functional to both the individual aim of fulfilment (since 'every citizen must naturally tend to confine himself to the role for which he is most suited') and the social goal of production (which requires the invention, examination and execution of useful projects). Saint-Simon portrays the shared ends of industrial society in moral terms, but reduces morality to the general happiness which is in turn identified with productivity - 'work is the essence of all virtues'. The same reductive strategy defines society as 'the ensemble and union of men engaged in useful work', describes individuals as free when they are 'unrestricted in productive work', and reveals 'true equality' as consisting in an individual's right to benefits 'exactly proportionate' to the contribution of that person (and their capital) to productivity.

The universal human drive to seek power would in future be sublimated, deflected away from immoral and destructive expression in conflict between individuals and into a cooperative victory over nature. Politics would be transformed into a positive science, developing 'laws of hygiene' to eradicate idleness ('a state of sickness in

man'), whilst 'government' would transmogrify into a much reduced 'administrative' structure facilitating production (divided accordingly into Chambers of Invention, Examination and Execution). His scientism and belief in natural inequality led Saint-Simon to assign 'temporal power' to those most competent to exercise it (the most successful 'industrials'). 'Spiritual power' would be the responsibility of a clerisy of scientists and artists, providing the moral and educational leadership necessary for social stability. Initially Saint-Simon advocated a 'Religion of Newton', with its own forms of worship and dogma, as an alternative to Christianity (optimistically suggesting that existing clergy might retrain for the new order). Increasingly, however, he portrayed his 'holy enterprise' as embodying the 'true spirit' of a Christianity uncontaminated by the heresies of the Church. This 'New Christianity' would utilize rhetoric, music and imagery to 'fill the souls of the faithful with feelings of terror or joy' designed 'to direct their ardour…towards works of general utility'. Institutionalized cooperation between the 'temporal' and 'spiritual' powers of Europe would lead to a confederation 'without wars, catastrophes or political revolution', in which a 'new family sentiment of Europeanism' would supersede competing patriotisms.

Saint-Simon's legacy resists summary. Any adequate account would include and assess his impact on his distinguished and estranged secretaries (Auguste Comte, Augustin Thierry) and claims of influence upon a succession of renowned social and political theorists (John Stuart Mill, Karl Marx), writers (Heinrich Heine, George Sand) and musicians (Franz Liszt, Hector Berlioz). The Saint-Simonian 'movement' did not survive intact in France, but fractured into the worldly and messianic currents that Saint-Simon had sought to incorporate. The achievements of the former, a distinguished group of engineers, mathematicians, economists and bankers, include the Suez Canal Company, the Paris-Lyon-Méditerranée railway and the Crédit mobilier. The latter, a more wayward group under Père Enfantin, scandalized Paris with the elaborate public rituals of their community at Ménilmontant and talk of an androgynous (and seemingly lascivious) God, before embarking on a crusade to North Africa in search of the 'Female Messiah'.

See also: History, philosophy of; Positivism in the social sciences §1; Socialism; Utopianism

DAVID LEOPOLD

List of works

Saint-Simon, C.-H. de R., Comte de (1966) *Oeuvres de Claude-Henri de Saint-Simon*, Paris: Éditions Anthropos, 6 vols.(The most complete modern edition of Saint-Simon's writings.)

Saint-Simon, C.-H. de R., Comte de (1975) *Selected Writings on Science, Industry and Social Organisation*, trans. and ed. K. Taylor, London: Croom Helm.(An extensive and useful collection of excerpts in translation from Saint-Simon's writings.)

Saint-Simon, C.-H. de R., Comte de (1976) *The Political Thought of Saint-Simon* ed. G. Ionescu, Oxford: Oxford University Press.(A selection of excerpts in translation from Saint-Simon's writings focusing on his political ideas.)

References and further reading

Carlisle, R.B. (1987) *The Proffered Crown. Saint-Simonianism and the Doctrine of Hope*, Baltimore, MD: Johns Hopkins University Press.(A sympathetic historical account of the Saint-Simonian movement.)

Durkheim, É. (1958) *Socialism and Saint-Simon*, ed. A.W. Gouldner, Yellow Springs, OH: Antioch Press.(An interpretation of Saint-Simon as the founder of a positivist science of society by the renowned French sociologist.)

Manuel, F. (1956) *The New World of Henri Saint-Simon*, Cambridge, MA: Harvard University Press.(An erudite and interesting general work on Saint-Simon.)

Salvation

For there to be such a thing as salvation, there must be someone to be saved, something from which they need to be saved, and some way in which they can be saved from it. 'Salvation' is primarily a religious term, and religious traditions typically assume that there is some basic religious problem that all people face. Monotheistic religions (for example, Judaism, Christianity, Islam, Viśiṣṭādvaita and Dvaita Hinduism) whose central doctrine concerns God conceived as Creator and Providence take this basic problem to lie in the fact of sin. Human persons have sinned (knowingly acted against the will of God) and sinning has become habitual. Thus there is need for forgiveness and reformation, which are available only in God's gracious pardon and restorative power. People can receive forgiveness and reformation through repentance and faith. Salvation by sheer self-effort is impossible. Nonmonotheistic traditions (for example, Buddhism, Jainism, Advaita Vedānta Hinduism) take a particular sort of ignorance to be the basic problem. The ignorance in question involves having false beliefs about the nature of persons and their cosmic environment. The proper treatment and cure is the achievement of an esoteric religious experience in which calm and bliss are accompanied by an understanding of the true nature of reality. The different traditions give very different accounts of what this nature is. Thus religious traditions differ greatly in the ways in which they conceive persons, their basic religious problem, and the proper treatment and cure. Secular notions of salvation, as in classical Marxism, tend to be secularizations of one or another religious conception - in the Marxist case, of the notion of the Kingdom of God.

1 Indian Buddhism

According to Indian Buddhism, there is a cycle of birth and rebirth in which every person participates; the cycle is beginningless and, unless one becomes enlightened, endless. There is a law of karma: one's misdeeds yield bad consequences for oneself, and one's good deeds yield good consequences for oneself. These consequences frequently come in a later life than that in which one performed the original action. Life in this cycle is inherently unsatisfactory, so the basic religious problem is that of escaping the round of birth and rebirth. The way to do so is through having a pre-*nirvāṇa* enlightenment experience in one's present life, which prepares one for release from the cycle and entry into *nirvāṇa* upon one's (next) death.

To prize anything in the reincarnation cycle more than escape from that cycle is, for reincarnation traditions, the height of folly. A person who enjoys the best of earthly conditions and fails to see that they are momentary and meaningless compared to the value of escape and enlightenment is desperately ignorant.

The natural world (roughly, the set of objects of sensory experience) does not consist of mind-independent, enduring physical objects. Rather, it consists of either physical (nonconscious) states or mental (conscious) states with perceptual content; in Theravāda Buddhism - the 'tradition of the elders' - it is held to consist of both, whereas in Mahāyāna Buddhism - the 'greater vehicle' (as contrasted to Hīnayāna, the 'lesser vehicle', Mahāyāna's term for Theravāda) - it consists only of mental states. Perhaps more importantly, persons themselves are nothing more *at* a given time than a bundle of states (conscious and nonconscious for the Theravādin, only conscious for the Mahāyānin), and *over* time they are nothing more than a series of such bundles. The idea seems to be that were persons more than this, enlightenment and escape from the reincarnation and karma cycle would be impossible. This tells us, then, the nature of what requires salvation according to Indian Buddhism, and what it needs salvation from (see Buddhist philosophy, Indian §§1-2).

The nature of salvation is another matter. What is it to be in the condition of having been saved? Buddhist salvation is a matter of becoming enlightened in this lifetime, which involves having an esoteric religious experience in which one comes to recognize the truth of the Buddhist account of reality, particularly concerning persons, and experiences calm and bliss. This experience in turn allows one release from the reincarnation cycle at death. For the Theravāda tradition, as for Buddhism generally, there is nothing permanent - nothing endures. Everything noncomposite that exists at a given moment exists only at that moment and everything composite either exists only at one moment or else is composed of elements each of which exists at only one moment. Further, every state or condition that can enter into the composition of a person is inherently unsatisfactory. Given these doctrinal claims, it follows that continued existence as a person inevitably involves continuing to have the basic problem that attaining *nirvāṇa* is intended to solve. Hence *nirvāṇa* will be extinction as a person; it will involve

extinction or annihilation of any state or series of states that could enter into the composition of a person. This is what Theravāda doctrine entails and what various texts seem to teach. This interpretation, however, is controversial, and *nirvāṇa* is sometimes said to be ineffable. For part of the Mahāyāna tradition, achieving a pre-*nirvāṇa* enlightenment state involves achieving a condition in which one is still composed of conscious states, but states which are not unsatisfactory. Hence the earlier claim that every state or condition that can enter into the composition of a person is inherently unsatisfactory is rejected. On this view, one who becomes enlightened will indefinitely postpone entering final *nirvāṇa* until the time at which everyone is able to do so, and it is sometimes held that this time will never come. Thus, on this account, whatever personal identity holds regarding a person within one lifetime can also hold over lifetimes after one has become enlightened, and over enlightened lifetimes in which none of one's composing bundles contains anything unsatisfactory. By contrast, Mahāyāna voidist tradition denies that even such persons as Mahāyāna idealism allows are anything other than illusory, and so takes the solution to one's religious problem to be that the problem, and those who have it, are themselves only illusory (see Buddhist concept of emptiness).

2 Jainism

Jainism holds that there exist a great number both of real (nonillusory) self-conscious substances and non-self-conscious substances. To be a substance is to be a possessor of properties, not to be a property oneself, to endure over time and to remain numerically the same through qualitative change. Indeed, Jaina mental substances are permanent and indestructible, and their goal is to escape being embodied because embodiment causes ignorance, namely the false beliefs that we depend on other things for our existence and could be destroyed, and that our knowledge is inherently limited. Each person or soul is eternally distinct from every other and from any physical (nonconscious) substance. Though the Jainas believe in individual immortality, they do not believe in a supreme deity (see Jaina philosophy §1).

The depth and importance of the difference between Buddhism and Jainism concerning the nature of persons is illustrated by one of their more important disputes. The Jaina objection to the Buddhist account of persons and karma essentially runs as follows. Justice requires that the same person who performs an action must reap the positive or negative karmic consequences that follow from it. According to the Buddhist tradition, a person at a time is but a momentary bundle composed of momentary states. Call this a single-time person; all there is to any person at any time is a single-time person. According to the Buddhist tradition, a person over time is a series of momentary bundles - a series of single-time persons - and so is what we might call a multiple-time person. Every action is performed at some time, and so is done by some single-time person. Every consequence is received at some time, and so by a single-time person. But since consequences typically take time to come along, the same single-time person does not both perform the action and receive the consequences, and justice is not served.

The obvious Buddhist answer is that it is enough to serve justice that the same multiple-time person include both the doer of the deed and the receiver of the consequences thereof. The Jaina response is that all there is to the multiple-time person at the time of the action is the single-time person who then exists, and all there is to the multiple-time person at the time of the consequences is the single-time person who then exists; multiple-time persons are agents or patients only in so far as their component single-time person parts act or receive. Single-time persons must *be* persons, or else agents are not persons. But then multiple-time persons are not persons in the sense in which agents are persons, but are classes of persons; single-time persons are persons in one sense (the basic) and multiple-time persons are persons in another (derivative) sense. If the person-in-the-basic-sense who acts is different from the person-in-the-basic-sense who receives the consequences of the action, then justice is not served.

Jainas, like Buddhists, hold that salvation requires the having of an enlightenment experience; this involves feelings of calm and bliss, and essentially includes a recognition of the correctness of the Jaina view of reality, particularly concerning persons. One who has such a this-life experience is released from embodiment and karmic obligations upon death. Personal identity is retained.

3 Advaita Vedānta

For Advaita Vedānta, a nonmonotheistic variety of Hinduism, ultimate reality, which does not depend on anything else for its existence, is nonpersonal and eternal, unchanging and permanent. The real self (*ātman*) of each person

is identical to this ultimate reality (Brahman-without-qualities). Having an enlightenment experience (*mokṣa*) involves realizing this truth, thereby escaping the reincarnation cycle, losing all individuality, and being ‘absorbed’ into Brahman. Here, in contrast to Jainism, personal identity is not retained. What was you, or seemed to be, no longer exists as you when final enlightenment occurs. Critics of this view point out that it means that there can be neither persons who have a basic religious problem, nor a problem for them to have - nor, of course, on this view can persons exist problemless (see Vedānta §1). Viśiṣṭādvaita Vedānta (whose major figure is Rāmānuja) and Dvaita Vedānta (whose greatest representative is Madhva) are monotheistic varieties of Hinduism that offer a very different account of the sacred texts to which Śaōkara, the most influential Advaita Vedānta philosopher, appeals.

4 Monotheistic traditions

Monotheistic traditions typically hold that God, an omnipotent, omniscient and morally perfect being, depends for existence on nothing else, and everything else that exists depends on God for its existence. Whether beginningless, or created *at* a time, or *with* time, the world exists by divine courtesy. Persons resemble God in ways that nonpersons do not, having some degree of knowledge, power and (potential) goodness. God is unique and holy, and, typically, in some sense the source of morality. For at least Semitic (Jewish, Christian and Islamic) as well as Hindu monotheism, God is characterized by unselfish love (see God, concepts of §§1-6).

According to these traditions, the basic religious problem all human beings face is that they have sinned - that is, they have knowingly acted wrongly against God’s will, and overrated their own significance while undervaluing the worth of others. Sin is not merely manifested in individual actions, but comes from a disposition towards wrong action that comes to lie deep in our nature. Our individual and collective existence is plagued by it; sin has social and institutional as well as individual manifestations. Thus there is a need for human repentance, for sorrow for one’s sins that includes the resolve not to repeat them, and for prayer for divine forgiveness. This prayer God graciously answers, enabling forgiven sinners to love others as they love themselves. Love in this sense is primarily volitional, not emotional, and involves concern for the ultimate good of the beloved, for which one is prepared to act sacrificially (see Sin §§1-2). Human individuality, then, is viewed as real, not illusory, and good. Finite persons typically are viewed, as in Jainism, as mental substances, but also, as in Buddhism, as not indestructible.

In Christianity, it is the incarnation, death and resurrection of Jesus Christ, God incarnate, that brings salvation. Christ lived a perfect human life as our example, died for our sins to make a just forgiveness possible and rose from the dead; the power to forgive and to progressively free us from the domination of sin is exercised through Christ’s life, death and resurrection (see Incarnation and Christology §1; Atonement).

Sin is not seen as the violation of arbitrary laws, but as a disposition (realizing itself in action) that thwarts one’s potential to realize one’s capacity to become as like God as is possible for a creature, and hence to fulfil one’s nature. It is in worship, obedience and service that one both shows love to God and becomes fully oneself. Creation in God’s image involves having an intrinsic worth that God sustains in existence. Hence, in the monotheistic traditions, personal identity is typically fully retained in the afterlife. Union with God is a matter of oneness of purpose, not of being. Salvation is seen as resulting in a personal relationship with God that occurs in the context of a redeemed community.

5 Agreements and disagreements

Semitic and Hindu monotheism, generally speaking, agree on these matters. For both sorts of monotheism, God is providential, though in Hindu monotheism this involves God governing the course of reincarnation and distributing karma with a justice tempered by mercy, whereas for Semitic monotheism, God’s being providential involves such actions in history as calling Abraham to be father of the nation of Israel, giving the law to Moses, sending the Old Testament prophets, becoming incarnate in Jesus Christ (for Christianity) and revealing the Qur’an to Muhammad (for Islam). (There are traditions associated with particular Hindu temples that Krishna, for example, appeared to someone and commanded that a temple be built at a particular place, but there is not the robust notion of God acting in history that characterizes the Semitic traditions, and no notion of God moving the course of history towards the Kingdom of God). For each sort of monotheism, worship is not a preliminary religious experience to be later transcended; its appropriateness is built into the nature of the distinction between

Creator and creature, which is not a dissolvable distinction. In Semitic, but not Hindu, monotheism central religious doctrines make essential reference to certain persons and events. For Christianity, God's becoming incarnate in Jesus Christ to die for our sins and be raised from the dead for our justification is an essential objective basis for divine forgiveness of the repentant.

There are similarities between Indian monotheistic religious traditions and Indian nonmonotheistic traditions. They agree that there is no movement of history towards an end or fulfilment. For Advaita Vedānta and Buddhist voidism, temporal and historical processes are illusory. For Buddhist idealism, historical events are all states within the minds of persons - parts of transitory bundles. For earlier Indian Buddhism, historical events include physical or nonconscious states, many of which do not belong to bundle-persons, but, like Jainism and Hindu monotheism, it denies that a providence guides those events to any historical conclusion or goal. All the Indian perspectives initially agree that the doctrines of reincarnation and karma are true, though some of them so radically reinterpret these doctrines as in the end to reject them. For no Indian tradition does basic religious doctrine essentially concern certain persons or contain essential references to places.

There are also differences between Indian monotheistic and nonmonotheistic perspectives. For the nonmonotheistic traditions, the basic religious problem is ignorance, not sin (there is no God to sin against), and the basic religious need is for enlightenment, not for forgiveness (there is no God to forgive). For Indian monotheism, it is sin that is our problem, and God's active forgiveness that solves it. For the nonmonotheistic traditions, religious knowledge is gained through meditation and through sacred texts that have no nonhuman personal source, not through revelation (there is no God to offer revelation). For Indian monotheism, Brahman is the source of all religious knowledge, scriptural or experiential. For the nonmonotheistic traditions, salvation comes through one's own efforts (there is no God to help). These are do-it-yourself religions: however necessary the aid of a teacher, guide or guru prior to one's enlightenment, it is one's own efforts that ultimately succeed or fail. For all of these traditions, it is in an enlightenment experience that one is said to learn the ultimate religious truth, and by having the experience one learns it in such a manner that one is transformed. But the truths that one is alleged to learn cannot all be true; at most one is. The claim is that enlightenment experience yields saving religious knowledge, not merely that it provides psychological certainty or great comfort. For monotheistic traditions, salvation requires that one repent of one's sins and seek God's forgiveness, and is constituted by one's flourishing in a community that lives in God's presence.

See also: Heaven; Justification, religious; Reincarnation

KEITH E. YANDELL

References and further reading

Brandon, S.F.G. (1962) *Man and his Destiny in the Great Religions*, Manchester: Manchester University Press. (Brandon's Wilde Lectures, delivered 1954-7 at Oxford, dealing widely with notions of salvation.)
Brandon, S.F.G. (1963) *The Saviour God*, Manchester: Manchester University Press.(Collection of essays covering many notions of religious salvation.)
Carman, J.B. (1974) *The Theology of Ramanuja*, New Haven, CT: Yale University Press.(The classic discussion in English of Rāmānuja's theology.)
Carman, J.B. (1994) *Majesty and Meekness*, Grand Rapids, MI: Eerdmans.(Comparative study of concepts of God.)
Denney, J. (1912) *The Death of Christ*, London: Hodder & Stoughton.(A classic statement of the Christian understanding of salvation through Christ.)
Lott, E. (1980) *Vedantic Approaches to God*, London: Macmillan.(Excellent discussion of Indian monotheism.)
McGrath, A.E. (1986) *Iustitia Dei*, Cambridge: Cambridge University Press, 2 vols. (A history of the Christian doctrine of justification.)
McGrath, A.E. (1994) *Christian Theology: An Introduction*, Oxford: Blackwell. (Comprehensive, accessible introduction to Christian theology.)
Sharma, N.K. (1902) *Philosophy of Madhacarya*, Delhi: Motilal Barnasidas.(Major study of the philosophy and theology of Madhva.)
Smart, N. (1964) *Doctrine and Argument in Indian Philosophy*, London: Allen & Unwin.(Clear general discussion of Indian philosophy.)
Smart, N. (1968) *The Yogi and the Devotee*, London: Allen & Unwin.(Comparison of Indian and Catholic

theology.)
Wood, T.E. (1990) *The Mandukya Upanishad and the Agama Sastra*, Honolulu, HI: University of Hawaii Press. (Discussion of Advaita Vedānta.)
Wood, T.E. (1991) *Mind-only: A Philosophical and Doctrinal Analysis of the Vijnavada*, Honolulu, HI: University of Hawaii Press.(A lucid and superb discussion of early Buddhism with some reference to later Buddhism.)

Sanches, Francisco (1551-1623)

Francisco Sanches was a sceptical philosopher and a professor of medicine at the University of Toulouse in southern France in the late sixteenth and early seventeenth century. He was born in Spain to a family of Jewish ancestry that had been forcibly converted to Catholicism, but he was brought up in France. Though he was a distant cousin of the sceptic Michel de Montaigne, he independently advanced what was perhaps the strongest sceptical critique of Aristotelianism and Platonism. In addition he developed a scepticism about mathematical knowledge claims. At the same time, he offered the first form of constructive scepticism, a way of solving intellectual problems without antecedently overcoming the sceptical challenge to traditional kinds of knowledge. He thus presented science as a way of dealing with experience, rather than as a way of gaining knowledge, and in this his views anticipate some twentieth-century philosophies. Sanches was also an important empirical medical practitioner, who presented the newest medical findings in his courses at Toulouse. His sceptical-critical views were influential in the first half of the seventeenth century, and were still being studied in Leibniz's time.

1 Life and works

Francisco Sanches (or Sánchez, though he is not to be confused with Francisco Sánchez de la Brozas (1523-1600), also known as Sanctius) was born in 1551 on the Spanish-Portuguese border to a family which had been very influential in Spain before the Expulsion of the Jews in 1492, and were now Marranos or New Christians (Jews who had been forcibly converted to Christianity). They first moved to Portugal and then, in 1564, to southern France in order to escape religious and political persecution by the Inquisition. The young Sanches studied at the Collège de Guyenne in Bordeaux, the same school that Montaigne had attended. After these studies he travelled in Italy for a while, visiting various universities there, and then settled in Rome. After two years of study at Rome with leading medical innovators who brought him into contact with new medical discoveries and theories, he went to the University of Montpellier, where he received a degree in medicine in 1574. After several years as a successful medical practitioner he was appointed first as professor of philosophy at the University of Toulouse in 1585, and then in 1612 as professor of medicine at the same university. He continued to work there until his death in 1623.

Sanches wrote extensively. The first of his philosophical writings that has survived is a letter to the Jesuit mathematician, Father Christopher Clavius, who had just edited Euclid's works, and whom Sanches had met in Rome. In the letter, Sanches offered a sceptical attack on the possibility of attaining genuine truth in mathematics. This was followed by his most famous writing, *Quod nihil scitur (That Nothing is Known)*, composed between 1574 and its first publication in 1581. Soon thereafter he wrote *Carmen de Cometa*, published in 1578, a critical examination in poetic form of the astrological interpretations of the comet of 1577. He also wrote some commentaries on portions of Aristotle's writings, as well as many medical works. Sanches criticized various Renaissance naturalistic views, such as those of Cardano (§§3, 6), and he may have actually debated with Giordano Bruno in person.

Limbrick's introduction to *That Nothing is Known* stresses how much of Sanches' limited sceptical view grew out of his lifelong medical work: 'it is the combination of two approaches to the theory of knowledge, the philosophical and the medical, which distinguishes Sanches' contribution to the history of ideas in the sixteenth and seventeenth centuries' ([1581] 1989: 25-6).

2 Attacks on Aristotle and Plato

The most important work of Sanches, *That Nothing is Known*, is perhaps the best technical exposition of philosophical scepticism produced during the sixteenth century (see Scepticism, Renaissance §5). Unlike other philosophers who were diffuse and digressive in their presentations of scepticism, appealing to the history of human stupidity and the variety of theories that have been accepted, Sanches instead systematically developed a radical sceptical critique of the Aristotelian theory of knowledge (see Aristotle §6), pointing out its many errors and inconsistencies. All sciences, he declared, begin with definitions. But these definitions are arbitrarily assigned to things without having any definite relation to them, and so they do not state the nature of things. Moreover, names keep changing. Hence when we think we are saying something about the nature of things by means of combining words and definitions, we are unfortunately just fooling ourselves.

This nominalistic thesis, when applied to Aristotle's view that science consists of knowledge that is certain, acquired by syllogistic demonstrations from true definitions, leads quickly to a scepticism about whether such knowledge can be achieved. As already seen, we do not possess true definitions, but only arbitrary ones. In fact, the particulars about which we are trying to demonstrate are better known than the general terms in the premises. Further, the syllogistic method of reasoning does not lead to any new knowledge. There is a vicious circularity in any knowledge claims based on demonstrative syllogisms, for the conclusion that is being proved actually constitutes part of the evidence for the premises. In order to demonstrate that Socrates is mortal, it is argued that all men are mortal and that Socrates is a man. However, the premises involve the conclusion, since the particular, Socrates, is needed in order to have a conception of man and of mortality. This criticism was later given by both Gassendi (§3) and J.S. Mill (§2) in their discussions of Aristotelian logic. Moreover, one could not have a true definition of the terms involved in the syllogism without already knowing their meaning, in which case one would not need the syllogism. On the basis of all this, Sanches said of Aristotle's notion of science, 'Do you call this science? I call it ignorance!'

Next Sanches contended that in a formal sense anything can be proved syllogistically if one just begins with the proper premises. Thus the fact that one can construct a syllogistic demonstration constitutes no indication that the conclusion of such a demonstration is true, for if the initial premises are carefully chosen, both true and false propositions can be proved in this manner.

Finally he attacked the Aristotelian claim that true knowledge is knowledge of things in terms of their causes. In order to know something in this manner one would have to know the cause of the cause as well, and the cause of the cause of the cause, and so on *ad infinitum*.

In his letter to Clavius, Sanches attacked a form of the Platonic theory of knowledge (see Plato §10). We cannot gain knowledge of things through mathematical study, since the objects studied by mathematics are not the natural real ones encountered in human life. Rather these objects are ideal, or perhaps even impossible ones, like points and lines. The mathematical relations that are demonstrated about such objects do not help explain anything in nature or experience, unless we happen to know independently that the experienced objects have mathematical properties, and also know that the principles of mathematics are in fact true. We cannot learn this from mathematics itself. So far as we can tell, mathematics is just conjectural or hypothetical until we can independently determine the nature of things.

3 Constructive scepticism

To replace the approaches that Sanches regarded as false conceptions of science, he insisted that true science is perfect knowledge of a thing ('Scientia est rei perfecta cognitio'). This conception, Sanches contended, was perfectly clear. Genuine knowledge is the immediate, intuitive apprehension of all of the real qualities of an object. Therefore true science will deal with particulars, each of which will somehow have to be understood individually. Any generalizations that go beyond this level of scientific certainty will introduce abstractions, fictions and so on. Scientific knowledge in its ideal form for Sanches would then consist of the experiential apprehension of each particular thing in and by itself.

None the less Sanches casts doubt on the possibility of such an experiential apprehension. At the beginning of *That Nothing is Known*, Sanches had said that he did not know if he actually knew anything at all. After casting doubt on whether anything can be known by using Aristotle's theory of science, Sanches then analysed his own view, and showed that in a strict sense humans are incapable of reaching any certainty at all. A full apprehension of objects which are known only one by one cannot be reached, in part because of the nature of objects, and in part because of the human being's nature. Everything is related to everything else, and therefore cannot be known individually. And there are an unlimited number of things, all of which are different, so they could never all be known. And, to make matters worse, things change. Thus they cannot be truly known since they never reach a final or complete state.

Turning to the human side of the question, Sanches spent much time discussing the difficulties that prevent human beings from obtaining true knowledge. Our ideas are dependent on our senses which are only capable of perceiving the surface aspects of things, the accidents, and never the substances. Appealing to his medical information Sanches was able to indicate how unreliable sense information is, how it changes as our state of health

changes, and so on. The many imperfections and limitations that God has seen fit to leave us with do, in fact, prevent our senses and our other faculties and powers from ever reaching true knowledge. All that we can attain is a quite limited, imperfect knowledge of some things which are present in our experience through observation and judgment. And, as Sanches seems to have found out through his medical studies and work, few scientists in fact make use of experience, and few persons actually know how to judge.

Sanches' sweeping negative conclusion that 'nothing is known' (*nihil scitur*), goes far beyond the Pyrrhonian balancing of opposing views and suspension of judgment offered by his distant cousin Montaigne. Yet Sanches' conclusion is the result of a most careful epistemological examination of the nature of our knowledge. Unlike the scepticism presented by Agrippa von Nettesheim (§3), which is destructive in its tendencies, Sanches attempted to state precisely what genuine knowledge would be, and then to show why it is not possible to attain this knowledge. He differs from both Montaigne (§§2-3) and Agrippa in that there is also a positive side to his work. He put forward a procedure which was not aimed at gaining knowledge, but at developing a way of dealing constructively with human experience. This procedure, which Sanches explicitly called scientific method ('método universal de las ciencias'), consists in patient, careful empirical research and cautious judgment and evaluation of the data observed. This would not provide, as his contemporary Francis Bacon (§§3, 6) thought, a key to knowledge of the world, but it would allow us to obtain the best information available. In advancing this limited or constructive view of science, Sanches was the first Renaissance sceptic to conceive of science in its modern form, as the useful activity concerning the study of nature that remained after one had given up the search for absolutely certain knowledge of the nature of things (see Scientific method).

Unlike other Renaissance sceptics, Sanches only briefly raises the fideistic solution to sceptical difficulties, the appeal to knowledge gained by religious faith alone. This lack of a serious presentation of fideism has led José Faur (1992) to suggest that Sanches was really secretly putting forth a non-Christian message, an assertion of his ancestral Jewish faith rather than the Christianity of the other fideists. Faur shows that Sanches' use of biblical material, together with his knowledge of early Jewish non-biblical literature could be taken to suggest that he was indicating his real beliefs. However, the fact that two of his sons became priests may be more indicative of Sanches' actual views.

4 Influence

Sanches had a good deal of influence in his own day and throughout the seventeenth century. *That Nothing is Known* was reissued several times up to 1665, and he is mentioned by quite a few authors, including Leibniz, who refers directly to some of Sanches' views from both *That Nothing is Known* and the letter to Clavius. People have seen possible influences on Descartes (§4), Gassendi (§3), Mersenne (§2) and Spinoza (§8), among others. However, in general it is hard to determine his exact influence as distinct from that of Montaigne, Sextus Empiricus, Cicero, Charron, and other available sceptical sources that were read by most intellectuals of the time.

See also: Montaigne, M. de §§2-3; Scepticism §1; Scepticism, Renaissance

RICHARD H. POPKIN

List of works

Sanches, F. (*c.*1570-1623) *Opera philosophica (Philosophical Works)*, ed. J. de Carvalho, Separata da Revista de Universidade de Coimbra 18, Coimbra, 1955; repr. 1957.(The only complete edition of Sanches' philosophical works.)

Sanches, F. (1578) *Carmen de Cometa*, in J. de Carvalho (ed.) *Opera philosophica*, Coimbra, 1955.(Poetic analysis of different interpretations of the 1577 comet.)

Sanches, F. (1581) *Quod nihil scitur*, ed. and trans. E. Limbrick and D.F.S. Thomson, *Franciscus Sanches: That Nothing is Known*, Cambridge: Cambridge University Press, 1989.(A systematic sceptical critique of Aristotelian epistemology; Sanches' most important work. Includes Limbrick's excellent introduction (referred to in §1).)

References and further reading

Faur, J. (1992) *In the Shadow of History, Jews and Conversos at the Dawn of Modernity*, Albany, NY: State University of New York Press.(Referred to in §3. Chapter 6, 'Francisco Sanchez and the Quest for a New Rationality' is an attempt to see Sanches' contribution in terms of Jewish intellectual developments resulting

from forced conversions.)

Limbrick, E. and Thomson, D.F.S. (eds) (1989) *Franciscus Sanches: That Nothing is Known*, Cambridge: Cambridge University Press, 1989.(Translation of Sanches' most important work, *Quod nihil scitur*, with an extensive bibliography that includes secondary literature in Portuguese. Limbrick's introduction, referred to in §1, gives an excellent overall presentation of Sanches' life, thought and influence.)

Popkin, R.H. (1979) *The History of Scepticism from Erasmus to Spinoza*, Berkeley, CA: University of California Press.(Places Sanches in the context of the development of scepticism in the Renaissance.)

Sanctification

Sanctification, the process of becoming holy, is closely connected to justification, although Roman Catholic and Protestant theologians differ on how closely. According to the Roman Catholic Church, sanctification takes place in justification. In justification, sins are forgiven and there is an infusion of sanctifying grace, whereby one is made just and holy; this is what it is to be sanctified. In this state of grace, one merits heaven. Nonetheless, there is room for spiritual growth in the Christian life (though this is not called sanctification, as it would be in Protestantism) because concupiscence remains and appetites are still not fully under control. Justification is the beginning of a new life (the life of grace) in which we may grow towards integrity, the proper use of appetites; moreover, the gifts of faith, hope and charity, which enable one to perform meritorious works, can also increase. The state of grace, which is the result of justification, can be lost by mortal sin, but can be fully restored through the sacrament of penance.

Protestants teach that in justification one's sins are forgiven and one is fully reconciled to God, but that one is not wholly sanctified (that is, renewed or made holy). Luther's formula that we are simul iustus et peccator *(both just and sinner) is widely accepted by Protestants. Justification is only the beginning of sanctification, which, most Protestants believe, is never completed in this life. The exceptions are Methodists and members of some holiness churches (groups that either broke away from Methodism or were influenced by it).*

1 The influence of Eastern Orthodoxy on the Western Church

The Orthodox Fathers never engaged in controversies over justification and sanctification comparable to those of the Western churches. Their debates concerning humanity's salvation focused on the nature and person of Christ, that is, on what Christ must be in order to effect salvation. At the Council of Nicaea (325) it was affirmed that Jesus is fully God and fully man. At the Council of Chalcedon (451) it was made clear that Jesus is the mediator between God and humanity because Jesus has two natures, a divine one and a human one. God saves his followers through the work of Christ, which is brought to fullness by the Holy Spirit. The teachings of these two councils were accepted as definitive by both the Roman Catholic and the Protestant churches. The outlook of the Eastern Church was summarized by Saint Athanasius in his *Incarnation*: 'He was made man that we might be made divine' (ch. 54: 3). That is, human beings, while remaining human, are able through union with the crucified and risen Christ to participate in the eternal (uncreated) life of God.

The Christian life, whether Roman Catholic, Protestant or Orthodox, begins with baptism, which remits sin and enables one to participate in the death and life of the risen Christ. What distinguishes the Orthodox Church is its focus on the Christian life as the increase of one's participation in the uncreated life of God; this process is referred to as *theiopoiesis* (divinization). The liturgy of the Eucharist (Holy Communion) is an eschatological event at which the present world and the world to come intersect. The congregation, in the presence of the risen Christ, the prophets, the apostles and the saints, has a foretaste of heaven. The Eucharist is often referred to as the 'medicine of immortality', whereby one is increasingly healed of the effects of original sin, which introduced death into human life, and is increasingly enabled to participate in the eternal life of God.

Eastern Orthodoxy deeply influenced the spirituality of the Western churches, largely through the ascetic practices of the Desert Fathers and Mothers (from the third century) and the mystical theology of Pseudo-Dionysius, who was active in the sixth century. Evagrius of Pontus summarized their ascetic practices in his *Praktikos (Practice)*. He distinguished eight deadly thoughts (gluttony, lust, avarice, anger, sadness, sloth, vainglory and pride), which are said to assail every Christian and which must be mastered so that one may obey Christ's commandment to love one's neighbour. Reading, vigils, prayer, almsgiving and fasting are particularly recommended as effective remedies for the passions they arouse. Once sufficiently freed of the distractions of the eight deadly thoughts by the development of *apatheia* (understood as purity of heart, and not as lack of feeling, as in Stoicism), Christians are able to practise the contemplative life. Contemplative prayer in particular increases knowledge and love of God, culminating, as Paul taught, in knowledge of God face to face (1 Corinthians 13: 12). The sixth-century Pope Gregory the Great, through his knowledge of John Cassian, a disciple of Evagrius and other Desert monastics, recast Eastern ascetic teaching for the Western Church; the eight deadly thoughts, for instance, became the seven deadly sins (lust, gluttony, avarice, sloth, wrath, envy and pride).

The teachings of Pseudo-Dionysius, first through Maximus the Confessor and then, from the tenth century, through a direct knowledge of some of his writings, brought into the Western Church the practice of an elaborate spiritual ascent into God, said to culminate for a very few Christians in mystical ecstasy. Some of the most important Western spiritual teachers that Pseudo-Dionysius influenced are Bernard of Clairvaux, Bonaventure (§§2-3) and Saint John of the Cross. John particularly stressed Pseudo-Dionysius' *via negativa,* in which all images and concepts of God are set aside as inadequate, and the 'dark nights' of purification, which enable one to ascend into mystical union with God. The most influential alternative to this speculative spirituality ('speculative' meaning knowledge of God as reflected in God's works and culminating in direct mystical knowledge of God) is the practical spirituality of Ignatius Loyola's masterpiece, *Spiritual Exercises*. Its practical nature lies in the stress on the discernment and pursuit of God's will for one's individual life.

All Christian spirituality uses the distinction between the active or ascetic life, and the contemplative life (see Asceticism). Thanks mostly to the influence of Pseudo-Dionysius, the contemplative life is usually subdivided, yielding the famous threefold way: the purgative way of ascetic practices, and the illuminative and unitive ways of contemplation. The way of illumination is an indirect knowledge of God through contemplation of God's two books: the book of nature, which reflects God's power, wisdom and goodness, and the book of Scripture, usually interpreted allegorically for hidden spiritual wisdom. The unitive way is a direct knowledge of God through such practices as the *via negativa* and prayer. The thirteenth-century writer Hugh of Balma's widely used terms, beginner, proficient and perfect, more or less correspond to the threefold way (to asceticism, the illuminative way and the unitive way respectively).

2 The Christian life in Roman Catholicism

The Roman Catholic Church clearly affirms that Christians grow in their participation in the divine life. Nonetheless, it placed its emphasis on the other aspect of the Christian life, namely the forgiveness of sin, which, by renewing one, puts one in a state of grace that merits heaven. The tradition of spirituality, with its stress on ascent to God, is primarily for those who have a special vocation. Even so, the central concern for all Catholics is to obey the teachings of Christ, which is possible through the infused grace of the sacraments. It is by participation in the sacraments that one grows in faith, hope and charity, and achieves greater integrity. But since one is not granted the gift of stability in the restoration from the Fall, there is a continuing concern to preserve the state of grace and to restore it whenever it is lost by mortal sin.

There is a strong emphasis on the objective character of the sanctifying grace that justifies. In justification, one is actually renewed and is just or holy regardless of one's psychological state. That is, one does not have to experience grace in order for sanctifying grace actually to have changed one's state from that of sin to that of grace. This affects pastoral practice, since there is a marked distance between the ontological condition of holiness and the actual experience of Christians. Although there is no theoretical incompatibility between doctrinal teaching on justification, spirituality and pastoral practice, the strong interpersonal nature of the Christian life with God (made possible by the indwelling of Christ) is in practice in danger of being neglected by the stress on sanctifying grace as objective, since the personal appropriation, however desirable, is not necessary.

In justification, one is just or holy by the standard of justice, because one has been actually renewed and made innocent. In the state of grace, one is able to perform actions which are meritorious (worthy of reward) according to the standard of justice; this is called condign merit. By good deeds, one may also merit an increase of grace in this life and a greater glory (degree of beatitude) in the next life (see Heaven §3); this is called congruous merit and, in contrast to condign merit, stems from divine liberality rather than justice. One may congruously merit special graces for oneself and for others, including the sanctifying grace of justification. These views on merit underlie the propriety of prayer for others.

3 Protestant views: Luther, Calvin and Wesley

Martin Luther has been widely misunderstood by both Roman Catholics and Protestants (including Lutherans) as teaching a view of justification that has no place whatsoever for sanctification. For Luther, in justification we are forgiven because Christ's righteousness is imputed to us. But unlike Roman Catholic teaching, we are not made holy or sanctified. Luther formulated his view as *simul iustus et peccator* (both just and sinner). But in theological polemics, it is little noticed that Luther held that justification is the beginning of sanctification, the beginning of

the reception of God's promise in Christ to establish his kingdom or rule in his followers. The coming of this eschatological fact into their lives is how Luther conceived of sanctification. Although they are forever inadequate according to the law (that is, they remain sinners because they are incapable of making themselves acceptable to God by their efforts), God in Christ makes life with God and in God possible. Whenever and wherever that life manifests itself, sanctification occurs. One has regeneration of one's life, with the eschatological reality of God's kingdom breaking into one's life.

Luther's conceptions of ordinary occupations as genuine spiritual vocations (with the rejection of the monastic or priestly life as spiritually superior) and justification by faith are hallmarks of Protestantism, marking it as sharply different from ancient and medieval conceptions of the Christian life. Furthermore, in order to avoid any suggestion that righteousness (justice, holiness) could be achieved by one's own efforts (by works), Luther also rejected any scheme of stages of growth for achieving sanctification. For Luther, there is no path or course that describes the Christian journey, a position which would include a rejection of the pre-eminently influential Protestant account, Bunyan's *The Pilgrim's Progress*. Yet Luther did not repudiate spirituality entirely, despite the rejection of stages of growth and of monasticism. On the contrary, he deeply admired *Theologica germanica*, an anonymous classic of spirituality written in the fourteenth century, and wrote that, next to the Bible and Augustine, he had read nothing more helpful and true.

John Calvin (§4) was in essential agreement with Luther, but unlike Luther he wrote extensively on sanctification. For Calvin, the issue of one's salvation is firmly and completely settled with justification. Although one's primary concern is to discern and do the will of God, one is still under the power of sin, with a will that resists God's will. Sanctification consists of the conflict between sin and grace, flesh and spirit, which lasts until the will becomes wholly subject to Christ's will. One then becomes free of the tyranny of sin and finds perfect freedom. In this struggle to become holy, one realizes how much one is utterly dependent on God's mercy. So although one can do good deeds, they are acceptable to God only because of God's liberality (this corresponds to 'congruous merit' rather than 'condign merit' in Roman Catholic theology). Calvin makes it clear that sanctification (actual regeneration and holiness), no less than justification (the gift of Christ's righteousness), comes to the believer from Christ. The two are distinct, but are not to be disjoined, since they both result from one's union with Christ.

Both Luther and Calvin agreed that, even though one is justified apart from the law, there are three valid uses of the law: enforced by the civil government, it permits civil society; it convicts one of sin; and it guides the life of believers in discernment of God's will. Both were opposed to an antinomianism in which it was inferred that since one is saved by faith alone, the law is not binding; however, both believed that through justification, one is motivated by love and not by compulsion to obey the law. The extensive debate between Lutherans and Calvinists over the third use of the law appears in retrospect to have been based on misunderstanding.

Unlike Luther and Calvin, John Wesley agreed with the Roman Catholic view that the gift of salvation can subsequently be lost. However, his main concern was the completion of salvation by achieving freedom from all effects of the Fall ('entire sanctification'). For him, the sanctification described by Roman Catholic teaching is not entire because one is not freed from all effects of sin (concupiscence remains), nor does one achieve perfection. It is difficult to tell what Wesley meant by the perfection he believed attainable in this life, but this belief divides him from Luther and Calvin, and Roman Catholicism. He clearly thought that there is an experience of sanctification in which there is a total death to sin and a complete renewal of the image of God. His various qualifications concerning the nature of perfection did not, however, weaken the Methodist stress that one must press on towards perfection in this life. Much of the social activism of Methodism sprang from this stress.

4 Pastoral psychology and sanctification

The nineteenth-century camp meeting revivals, with their stress on a 'born again' experience led Horace Bushnell to champion Christian nurture as the normal path to godliness. Through his reliance on modern pedagogical principles, including elementary psychology, he became the father of the Christian education movement. In Protestantism, Christian education continues to be the primary means for promoting spiritual growth.

In post-war years, the psychology of Sigmund Freud and, to a lesser extent, that of Carl Jung was adapted for the purposes of understanding and promoting spiritual development, especially in US Protestantism. It is generally held that the pastoral psychology movement has become increasingly in danger of complete secularization, losing

a transcendent dimension. Since about 1970, this has provoked a marked resurgence of interest in spiritual classics, because of their explicit stress on the work of the Holy Spirit in renewal, as well as new attempts to integrate spirituality and psychology.

See also: Grace; Justification, religious

DIOGENES ALLEN

References and further reading

Anon. (14th century) *Theologica germanica*, New York: Paulist Press, 1980.(A spiritual classic that emerged from a movement known as the Friends of God.)

Aquinas, T. (1266-73) *Summa theologiae*, New York: McGraw-Hill, 1964.(A compact statement of Roman Catholic teaching on justification and sanctification; book III, section 4 is particularly useful.)

Athanasius (*c*.335) *Incarnation*, New York: Macmillan, 1946.(A classic statement of our redemption by Christ.)

Berkouwer, G.C. (1952) *Faith and Sanctification*, Grand Rapids, MI: Eerdmans.(A major Reformed study from Calvin to Barth, with comparisons to Lutheranism.)

Bunyan, J. (1678) *The Pilgrim's Progress from This World to That Which is to Come*, ed. J.B. Wharey, Oxford: Clarendon Press, 1960.(The most influential Protestant treatment in the English-speaking world of the Christian life.)

Calvin, J. (1957) *Institutes of the Christian Religion*, trans. H. Beveridge, London: James Clarke, bk III.(The most important and influential work in Reformed theology.)

Evagrius of Pontus (*c*.400) *Praktikos and Chapters on Prayer*, trans. J.E. Bamberger, Spencer, MA: Cistercian Publications, 1972.(The most influential work on ascetic theology.)

Flew, R. (1934) *The Idea of Perfection in Christian Theology*, London: Oxford University Press.(Especially valuable for the account of Wesley.)

Forde, G.O. (1984) 'The Christian Life', in C.E. Braaten and R.W. Jensen (eds) *Christian Dogmatics*, Philadelphia, PA: Fortress, vol. 2.(A thorough discussion of Luther and Lutheranism on justification and sanctification.)

Imbelli, R. (1994) 'Catholic Identity after Vatican II', *Commonweal* (11 March): 12-16. (A progress report on Jozef van Beeck's theology, which seeks a new spiritual-pastoral-theological synthesis in response to Vatican II.)

Loyola, I. (1548) *Spiritual Exercises*, New York: Paulist Press, 1991.(The major statement of a practical - in contrast to a speculative - spirituality.)

Luther, M. (1516) *Commentary on the Epistle to the Romans*, Grand Rapids, MI: Zondervan, 1954.(A major account of Luther's development of what became the Protestant doctrine of justification, with indications of his views on sanctification.)

Pseudo-Dionysius (*c*.500) *Theologia mystica*, West Park, NY: Holy Cross Press, 1944. (The classic account of the spiritual ascent into divine ecstasy.)

Stump, E. (1993) 'Sanctification, Hardening of the Heart, and Frankfurt's Concept of Free Will', in J.M. Fischer and M. RaVizza (eds) *Perspectives on Moral Responsibility*, Ithaca, NY: Cornell University Press.(An account of how a good God could sanctify some people and not others, and how sanctification is compatible with free will.)

Veith, G.E., Jr (1985) *Reformation Spirituality*, Lewisburg, PA: Bucknell University Press.(A valuable study of sanctification, through the poetry of George Herbert.)

Wingren, G. (1957) *Luther on Vocation*, Philadelphia, PA: Muhlenberg.(A presentation of the Protestant view of the spiritual value of ordinary occupations.)

Santayana, George (1863-1952)

George Santayana was a philosopher, essayist, novelist and poet. Born in Spain, he moved to America as a child and attended Harvard, studying under William James and Josiah Royce. The philosophical world first took note of Santayana for his work in aesthetics. The Sense of Beauty (1896), his attempt to give a naturalistic account of the beautiful, remains influential. He wrote exquisitely crafted essays on literature and religion, viewing both as articulating important symbolic truths about the human condition. His mature philosophical system is a classical edifice constructed out of positions adopted from Plato and Aristotle, which he modified in light of the naturalistic insights of his beloved Lucretius and Spinoza and steeped in pessimism reminiscent of Schopenhauer. Although in close touch with the philosophical developments of his day, he always viewed human life and its problems in a calming cosmic perspective.

1 Life and works

George Santayana was born Jorge Augustin Nicolas Ruiz de Santayana in Madrid, Spain, on 16 December 1863. At nine years of age, he accompanied his mother on her move to Boston. He attended Harvard University and studied with William James and Josiah Royce. Upon receiving his doctorate, he joined his former teachers on the faculty and rose to professor of philosophy. He never married, took early retirement from his teaching post and spent the last forty years of his life travelling and writing in Europe. He died in Rome on 26 September 1952.

Santayana's philosophical work is best understood in the context of the rejection of idealism at the beginning of the twentieth century. He was a realist in every significant sense of the word, believing that the world exists independently of human cognitive efforts and that not even the themes of thought we entertain depend for their reality on the human mind. Although his early, five-volume *The Life of Reason* (1905-6) has been compared to Hegel's *Phenomenology of Spirit*, the two works are fundamentally different. Santayana's philosophical edifice lacks Hegel's optimistic teleology, his devotion to totality and his structuring commitment to the power of consciousness (see Hegel, G.W.F. §5).

Beginning around 1920, Santayana developed an ontology designed to identify fundamentally different types of being. In *Scepticism and Animal Faith* (1923) and the much-misunderstood four-volume *Realms of Being* (1927-40), he attempted to show the ways in which essence, matter, truth and spirit differ from one another and are jointly adequate to account for everything in human experience and in the world. The terms in which Santayana cast his system derive from the great Western philosophical tradition. References to essence and substance may even make his writing appear scholastic. His message, however, is thoroughly modern, even postmodern, in its rejection of essentialism, of the possibility of certainty, and of a universal hierarchy of values.

2 Santayana's realism

Scepticism is irrefutable, Santayana argues, as long as knowledge is supposed to involve certain possession of its object. Such a stringent demand as this, however, commits us to a wrong-headed standard for cognition. In the affairs of life, we neither seek nor need assurance of the highest order. Accordingly, scepticism is defeated not by its logical weaknesses but by the irrelevance of the criterion it imposes on knowledge.

Santayana issues a call for honesty in philosophy, meaning by this the refusal to believe in our reflections what we would never countenance in daily life. This opens the door to a new method of doing philosophy and to a resultant system of ideas we might call 'the philosophy of animal faith'. The method consists of taking animal action as a starting point and attempting to disentangle the beliefs implicated in it. When animals eat, for example, they take the independent existence of their food for granted. Consciously or otherwise, they believe that what they eat is not a part of themselves, that it continues to exist even after it slips from sight, that they have the power to engulf it and that doing so is beneficial. The existence of a dynamic and enduring environment consisting of differentially powerful and desirable items is thus one of the beliefs involved in and justified by activities in which we all engage. The job of philosophy is to identify these beliefs and to present them as a coherent system of common sense.

What Santayana calls 'the discovery of essence', however, is often thought to be anything but common-sensical. But criticisms of the infinity of the forms of which the realm of essence consists are based largely on

misunderstandings. The essences he distinguishes do not *exist*: they are the qualities and relations, the structures and the event-types that may gain embodiment in the world. Santayana focuses on the ways in which these forms are different from the flux of physical events. As objects of thought, features of existence and terminal points of change, they remain free of the ravages of time. Their eternal self-identity renders them beacons in the turning world: they make temporal process possible without sharing its fate.

Essences are what the tradition has called ‘universals’. But since every distinguishable characteristic and the object of every possible thought is an essence, generic forms and the forms of natural kinds enjoy no privilege. This disintegrates the metaphysical, epistemic and axiological prerogatives of essences: if forms set the standards for being, knowing and the good, the infinity of essences means that we must choose which of a very large number of possible norms to embrace. Santayana accepts Aristotle’s idea that nature determines perfection. He thinks, however, that human nature is not defined by a single generic form, consisting instead of a vast continuum of the variously resembling forms of individuals. This leaves ample room for diversity and calls for choice in deciding which essence to enact.

Embodiment is the work of matter - a restless, mindless, impersonal force. Existence consists of a jostle of miscellaneous events: its external relations result from instantiating a shifting and arbitrary collection of essences. Santayana refers to the force of embodiment as a ‘whirlwind’ that creates ‘an insane emphasis’. Why there is something rather than nothing and why this world exists rather than some other are, therefore, questions that admit of no intelligible answer. Existence is a surd whose ultimacy cannot be eliminated by imagining a God who confers it as an expression of benevolence.

The realm of truth consists of all the essences that have been, are being and will be actualized in the history of the world. Truth is, in this way, a complete record of the agency of matter. Even if this realm contains an infinite number of forms, there remains an infinity never actualized. For if there is a monstrously complex essence that contains the form of everything that ever happened in the history of the world, there is another, unembodied essence differing from it in a tiny particular. In addition, there is an infinity of other forms that differ from it in indefinitely many ways.

Although this view of truth makes it eternal and unchanging, Santayana thinks that its tie to time cannot be severed. All truths relate to the changing world: only those essences that matter endows with external relations are members of the realm of truth. Santayana calls the truth about a fact its ‘standard, comprehensive description’. He does not mean by this, however, that it involves some actual description by a living being and thus that it belongs in the sphere of language or discourse. On the contrary, though it may be difficult to know, it is altogether objective and independent of what anyone says or thinks. It is the standard against which such thoughts are measured and is a ‘description’ only in the sense of being the total embodied form that may be described.

3 Values and spirituality

‘Spirit’ is Santayana’s word for consciousness, which he thinks is a product of biological causes. The immediate objects of consciousness are the nontemporal qualities and relations that belong to the realm of essence. Animal eagerness takes the presence of such themes in consciousness as signs of surrounding things. Perception and the knowledge of the world it makes possible are, therefore, always symbolic: they involve the use of present essences in deciphering the nature and causal powers of the substances on which our weal and woe depend. Such knowledge is never certain. Continued experience can correct it and successful action serves as its final criterion. The growth of knowledge in this way expresses the interests of the animal and not the intrinsic nature of awareness.

Consciousness in its purity pays no heed to animal life and knowledge. It finds no essences more worthy of attention than any others; it enjoys the presence of every form with equal readiness. By the claim that spirit is intrinsically spiritual Santayana means that it naturally transcends the concerns of daily life to engage in the carefree contemplation of eternal forms. This is ‘pure intuition’, a moment of aesthetic vision in which, as he puts it, the ultimate becomes immediate.

The spiritual life is not an existence that rivals successful rational life in this world. It consists of pure intuitions that can be had anywhere because their objects may be any of the infinity of essences. The only requirement of this achievement is that we liberate ourselves from animal concern and attend to the intrinsic features of what is

present, rather than to its causal or cognitive properties. This stress on the enjoyment of immediacies connects Santayana to the long history of spirituality and to such modern advocates of aesthetic immediacy as Schopenhauer (see Schopenhauer, A. §5).

What generates and justifies this account of ultimate human satisfaction is Santayana's view of the impotence of consciousness. He thinks that awareness is epiphenomenal, a 'lyric cry in the midst of business'. Although a product of the human nervous system, it is ontologically different from anything material and can wield no formative power. Its only function is to light up the realm of forms and, in the process, to enable us to enjoy things under the form of eternity. The transcendence it offers has a material ground and natural limits: spirituality is a perfection of this life, not the promise of another (see Epiphenomenalism).

Intuition of essence does not yield knowledge. The immediate objects of consciousness are universals and it takes animal faith to convert the colours and sounds that appear to us into signs of a surrounding world. Perception is veracious when the direct object of mind is an apt symbol of the features and movements of matter. This critical realism sees the identity of the essence we intuit with the essence embodied as a limiting case of knowledge; it is not necessary for cognition, we can probably not achieve it, and even if we did, we could have no valid evidence of it. There is little need of literal knowledge so long as the symbols we use help us live successfully.

Santayana's conviction that one's good is determined by one's nature is not enough to lead him back to Greek virtue ethics. He rejects the tidy universality that imposes a single perfection on all humans or even on all the members of a community. Working on the assumption that the role of the moral philosopher is to understand, not to condemn, he insists on the internal justification of every coherent life and unified set of values. The personal good is relative to the nature of the individual and the social good to the natures with which a community endows its members. The great problems of human life derive not from wickedness but from the clash of incompatible commitments, from internal confusion about the good and from the militancy that attempts to impose alien values.

By the time of his death, the loss of interest in systematic philosophy fostered by logical empiricism sent Santayana's popularity into decline. But the last two decades of the twentieth century witnessed the rediscovery of his thought. He is valued not only for his sharp insights, but also for his tolerant understanding of the varieties of human perfection. His naturalistic ontology is strikingly contemporary and his realism can serve as a welcome counterweight to prevailing historicist and constructivist positions. His categories offer a sensible and unified picture of the world, along with an account of human life that is free of illusions. His greatest contribution may be the way in which he combined an uncompromising naturalism with the highest demands of spirituality.

See also: Spain, philosophy in §8

JOHN LACHS

List of works

Santayana, G. (1896) *The Sense of Beauty*, New York: Charles Scribner's Sons; New Critical Edition, ed. W. Holzberger and H. Saatkamp, Cambridge, MA: MIT Press, 1988.(Naturalistic aesthetics, arguing that beauty is objectified pleasure.)

Santayana, G. (1900) *Interpretations of Poetry and Religion*, New York: Charles Scribner's Sons; New Critical Edition, ed. W. Holzberger and H. Saatkamp, Cambridge, MA: MIT Press, 1990.(Literature and religion as symbolic of the struggles of the human spirit.)

Santayana, G. (1905-6) *The Life of Reason*, New York: Charles Scribner's Sons, 5 vols. (Elegant account of the development of human ideas and institutions.)

Santayana, G. (1913) *Winds of Doctrine*, New York: Charles Scribner's Sons.(Contains important essays on Bergson, Russell and others.)

Santayana, G. (1923) *Scepticism and Animal Faith*, New York: Charles Scribner's Sons. (Central text for understanding Santayana's relation to scepticism and his philosophy of animal faith.)

Santayana, G. (1927-40) *Realms of Being*, New York: Charles Scribner's Sons, 4 vols.(Full development of Santayana's mature ontology.)

Santayana, G. (1936) *Obiter Scripta: Lectures, Essays, and Reviews*, New York: Charles Scribner's Sons. (Philosophical and literary essays.)

Santayana, G. (1946) *The Idea of Christ in the Gospels*, New York: Charles Scribner's Sons. (Naturalistic interpretation of the Gospels.)

Santayana, G. (1951) *Dominations and Powers*, New York: Charles Scribner's Sons; New Critical Edition, ed. J. McCormick, New Brunswick, NJ: Transaction, 1995. (Santayana's ideas on social and political life.)

Santayana, G. (1969) *Physical Order and Moral Liberty*, ed. J. Lachs and S. Lachs, Nashville, TN: Vanderbilt University Press.(Selections from Santayana's unpublished manuscripts.)

References and further reading

Arnett, W.E. (1984) *Santayana and the Sense of Beauty*, Magnolia, MA: Peter Smith. (Presents Santayana as a philosopher of disillusionment.)

Jones, J. and Saatkamp H. (1982) *George Santayana: A Bibliographical Checklist, 1880-1980*, Bowling Green, OH: Philosophy Documentation Center.(Exhaustive listing of works on Santayana.)

Kerr-Lawson, A. and Saatkamp, H. (eds) *Overheard in Seville*, College Station, TX: Texas A&M University. (Journal devoted to Santayana studies.)

Lachs, J. (1988) *George Santayana*, New York; Macmillan.(Introduction to Santayana's system of philosophy.)

Leitz, R.C. (ed.) (1991) *Critical Essays on George Santayana*, New York: Macmillan. (Essays on Santayana as literary artist.)

Levinson, H. (1992) *Santayana, Pragmatism, and the Spiritual Life*, Chapel Hill, NC: University of North Carolina Press.(Lively discussion of Santayana's moral and political views.)

McCormack, J. (1988) *George Santayana: A Biography*, New York: Paragon House.(Detailed but controversial biography.)

Singer, B. (1970) *A Rational Society*, Cleveland, OH: The Press of Case Western Reserve University.(Perceptive account of Santayana's social and political views.)

Sprigge, T.L.S. (1995) *Santayana: An Examination of His Philosophy*, New York: Routledge, 2nd edn.(Fine assessment of Santayana's philosophical contributions.)

Woodward, A. (1988) *Living in the Eternal*, Nashville, TN: Vanderbilt University Press.(Appreciation of Santayana's spirituality.)

Śaōkara (early 8th century)

Śaōkara has been a highly influential figure in Hindu philosophy and religion from his lifetime (early eighth century; traditionally 788-820) to the present day. He is the most renowned teacher of nondualist (Advaita) Vedānta, which emphasizes realizing the nondual reality, Brahman, through hearing and contemplating the Upaniṣads, sacred knowledge which reveals the nature of human existence and the cosmos. Unlike many Western thinkers, who consider themselves forward-looking individuals putting forth new insights, Śaōkara is self-consciously part of an ongoing tradition committed to scriptural exegesis. He honours prior teachers (such as Gauḍapāda), and his own writings are primarily explanatory commentaries on sacred Vedānta texts. Śaōkara also requires certain purifying qualifications to pursue liberation, and vigorously contests other views prevailing in his time.

1 Life and works

The traditional Advaitic account of the life of Śaōkara (Śaōkarācārya) is as follows: born a Brahman in Kaladi, South India, he was a worshipper, and reputedly an incarnation, of Śiva. Becoming a renunciant when still a youth despite his mother's wishes, he went to Banaras to study and teach. During extensive travels, he set up monasteries in the four corners of India and founded an ascetic order with ten divisions. He debated with (and often converted) scholars of other religious sects and philosophical schools, both Hindu and Buddhist. Although he is said to have died at the age of 32, his Advaita Vedānta philosophical school and the renunciant Brahmanical tradition continues, from immediate followers such as Sureśvara and Padmapāda to contemporary teacher-leaders called Śaōkarācārya ruling most prominently at the Śṛōgeri and Kanchipuram monasteries in South India.

Śaōkara's primary writings include commentaries on a number of Upaniṣads (such as the Bṛhadāraṇyaka, Chāndogya and Taittirīya), Bādarāyaṇa's *Brahmasūtra* (concise aphorisms on Brahman which attempt to explain and show the coherence of Upaniṣadic teachings while refuting other views), and the *Bhagavad Gītā*, the immensely influential poetic discourse found in the *Mahābhārata*. He also wrote an independent philosophical work, the *Upadeśasāhasrī*, and is said to have composed many devotional hymns (*stotras*) to Śiva, though his authorship of these writings is contested by many scholars.

2 Śaōkara on Brahman

Throughout his writings, most notably in comments on *Brahmasūtra* 1.1.1-4, Śaōkara holds that Vedāntic texts teach that the basis of reality is nondual Brahman, which is eternal, qualityless (*nirguṇa*) and pervades everything without limitation. It is the single, efficient, substantial and conscious cause of existence. Brahman is identical with our true unqualified and unchanging self (*ātman*), which is immediate, self-luminous and distinctionless consciousness. The *ātman* is 'intuited' as that pure subject prior to all particular contents of consciousness (and free from the limitations which condition waking and dreaming); it is sometimes called *sākṣin*, the self-luminous and unconditioned but passive witness. The self is undeniable and self-authenticating; one who doubts it in fact presupposes it.

Śaōkara's Advaita teaches that human beings normally suffer bondage to transmigratory existence (*saṃsāra*), which requires experiencing the fruits of all one's actions (karma) and thus repeatedly entering the cycle of rebirth. The delusive manifestation of diversity and change (*māyā*) appears due to beginningless ignorance (*avidyā*), which is the superimposition (*adhyāsa*) of illusory phenomenal diversity on *nirguṇa* Brahman. Brahman is, however, unaffected by the superimposition of objects and conditions upon it. The knowledge that *ātman* and Brahman are identical removes ignorance, and this removal brings liberation (*mokṣa*) from *saṃsāra*. Liberation is knowing *and* being Brahman, our original nature. Liberation has no beginning, end or degrees; it is not a place or an attainment. It is solely a change in understanding through the removal of ignorance, not a product or result of any action.

Śaōkara's well-known introduction to the *Brahmasūtra* commentary states that ignorant beings such as ourselves have a general sense of the pure subject, nondual Brahman, but then, from the memory of a thing previously experienced elsewhere, we erroneously superimpose objective limiting conditions on Brahman, and properties of Brahman on limited, unconscious objects. We begin to identify all things as 'I/mine' or 'other', oblivious to the

ultimate nondifference of subject/object and knower/known. Through this natural and eternal *adhyāsa*, we falsely suppose that we are finite and embodied agents and enjoyers (called *jīvas*), but neither *jīvas* nor the world are truly real. Śaōkara calls the *jīva* a reflection of the self, like an image of the sun reflected in a pool of water; changes or imperfections in the reflection do not affect the source. Inquiry into Brahman is necessary and proper because Brahman is now known imperfectly, and there are many mistaken opinions about it (such as regarding it as a body, agent or momentary entity). Śaōkara concludes that Vedānta study should be begun to remove the error of superimposition and to realize the unity of the self.

An obvious problem here is how to explain the common-sense perception of duality. Śaōkara relies on the idea of two levels of truth, one apparent, known as practical or 'everyday' worldly activity (*vyavahāra*), and the other the higher nondual reality indicated in the Upaniṣads. Śaōkara does not merely dismiss the everyday activity of *vyavahāra*, sphere of reasoning and experience. He claims that despite lacking ultimate reality, objects appear - thus, they are neither existent nor nonexistent. *Vyavahāra* is real as far as it goes; those in it should follow the ritual and social order according to *dharma* (though without attachment to results) to prepare properly for liberating knowledge. He often describes delusive everyday appearance as being like a rope mistakenly seen as a snake due to superimposition - the snake is never there and never different from the rope (see Error and illusion, Indian conceptions of §1). Still, unreal imaginings (snake, *vyavahāra*) are based on a real substratum (rope, Brahman). Śaōkara's 'realism' also asserts that waking experience is normally true on its own terms, and dreams are mental creations less real than, and largely derived from, waking. On the other hand, both these states are shown to be false by the highest truth, and he praises deep sleep as suggestive of Brahman: luminous, serene and objectless bliss.

For Śaōkara, then, we are always Brahman, but do not know it - just as the 'snake' is always the rope, but we realize that only after being told. Still, according to Śaōkara, some things in the realm of ignorance, such as the Upaniṣads and wise teachers, can move us towards higher knowledge. One can in fact be liberated while living, but the knower remains in a body as long as needed to experience currently manifesting fruits of actions, as one temporarily continues trembling even after realizing the 'snake' is but a rope.

Perhaps one of Śaōkara's strengths is that he does not feel obliged to resolve all the difficult philosophical problems his outlook raises. This is but one manifestation of his reservations about reasoning and the need for reliance on scripture. He is not deeply concerned with how diversity arises from unity, and he acknowledges that why superimposition takes place, or exactly how Brahman and ignorance are related, is ultimately inexplicable. Neither does he resolve whence ineluctable karma arises, or whether karma limits Brahman. He asserts that such issues are only relevant to those caught in *vyavahāra* and cannot be understood until ignorance ends.

3 Qualifications for pursuing liberation

Śaōkara holds that one must be qualified to attain liberation. In commenting on *Brahmasūtra* 1.1.1, he describes four personal or psychological requirements for pursuing knowledge of Brahman: discriminating between non-eternal and eternal things, detachment from enjoyments here or in the afterlife, serenity and control of mind and senses, and desire for liberation. He implicitly endorses the later formula which states that one must hear, reflect on and meditate on Vedic texts.

In addition, and contradicting modern claims of Advaitic inclusivism, Śaōkara held that access to the Veda (the source of liberating knowledge) is externally or socially restricted. That is, in everyday reality Śaōkara remains committed to Brahmanical orthodoxy, which stresses the caste system and the importance of Vedic duties and right action according to *dharma* as a preliminary leading to renunciation and knowledge. Since only higher-caste males can hear the Veda, no women, low-caste or outcaste men may get the proper training which brings liberation, a state that Śaōkara generally seems to hold can only be gained as a male Brahman renunciate.

4 Śaōkara on other Indian schools

While Brahman is one, it has many manifestations. Śaōkara recognizes a lower Brahman 'with qualities' (*saguṇa*), often called *īśvara* ('lord'). *Īśvara* is omnipotent and omniscient, and the efficient and material cause of world appearance, creating due to motiveless play. Despite *īśvara*'s relative inferiority, Śaōkara holds that belief in and devotion to *īśvara* is acceptable as a preliminary to, and helpful auxiliary on, the path to liberation. Worshipping a

god may lead one to the path of the gods, but not to liberation; that is, it may lead to the heavens and a better rebirth as a high-caste renunciate, well-placed to pursue *mokṣa*, but it does not lead directly to the ultimate end.

Śaōkara's Vedānta is sometimes called the later (*uttara*) Mīmāṃsā, or Exegesis (of Brahman), to differentiate it, but show its indebtedness to, earlier (*pūrva*) Mīmāṃsā (of *dharma*), which explains duties described in the ritual-oriented portions of the Veda (see Mīmāṃsā). Śaōkara repeatedly rejects the Pūrva Mīmāṃsāview that actions are necessary or sufficient to bring liberation; he argues that the works portion of the Veda, which contains injunctions about rituals and social duties, is inferior to the knowledge portion, which makes statements about Brahman. As stated earlier, release comes from knowledge of nondual Brahman alone (and this knowledge is not an action); actions cannot affect or purify the self. Performing Vedic rituals and caste duties may, like worship, lead to prosperity, increased purity or better rebirth (which make one eligible for knowledge). Still, only Brahman knowledge, eternally present and not dependent on human activity, can bring final liberation. Even yogic meditation, which stills the mind, cannot bring liberation; it is a humanly dependent action and presupposes duality.

Despite major philosophical differences with Pūrva Mīmāṃsā, Śaōkara largely shares its commitment to textual exegesis, its modes of argumentation, and its understanding of the means of knowledge (*pramāṇas*). For the highest truth, Śaōkara holds that the most important means is Vedic text (*śruti* or *āgama*), which is eternal and transcends human senses and reason. While *pramāṇas* like sense perception and reasoned inference are generally valid and legitimate in their own (duality-based) sphere, they are relevant only to the ignorant *vyavahāra*-bound being, and cannot reveal nondual Brahman on their own. Although Brahman is not directly communicable in language, even by scripture, the Vedāntic texts are vital since they inform us about Brahman and evoke an immediate intuition of it through identity statements like *tat tvam asi*, 'you (*ātman*) are that (Brahman).'

Śaōkara's writings reveal him to be a superb debater and polemicist. He generally puts forth the opponent's view (*pūrvapakṣa*) and then counters with his response, which brings the correct conclusion (*siddhānta*). His commentaries often include critiques of other philosophical schools of his time, attempting to show Advaita's coherence and the incoherence of his opponents, particularly Sāōkhya and the Sarvāstivāda and Vijñānavāda schools of Buddhism. He focuses on issues concerning causality. For example, Śaōkara generally agrees with the Sāōkhya view (called *satkāryavāda*) that an effect pre-exists in a cause (like a pot is 'in' a lump of clay); otherwise, events could happen randomly and anything could arise anywhere. However, unlike Sāōkhya, Śaōkara holds that change in form is not real transformation, but only appearance; that is, the unchanging cause (Brahman, clay) merely appears as an effect (*māyā*, pot), and is unaffected by the latter's impurities (see Causation, Indian theories of §§3-4; Sāōkhya §5).

Later schools, within and outside Vedānta, engage extensively with Śaōkara's ideas (see Rāmānuja; Madhva). There are also many modern interpreters of Śaōkara's Advaita; some, such as Swami Vivekananda and Sarvepalli Radhakrishnan, have been influenced by Western thought, and claim Śaōkara as an ecumenicist, reformer and 'philosopher' (see Ramakrishna movement). While it should be acknowledged that great thinkers allow for many readings, readers should be wary of how some modernize Śaōkara's views to fit the temper of other times.

See also: Brahman; Monism, Indian; Vedānta

ANDREW O. FORT

List of works

Śaōkara (early 8th century) *Brahmasūtrabhāṣya*, in G.F.W. Thibaut (trans.) *The Vedānta Sūtras of Bādarāyaṇa with the Commentary of Śaṃkara*, Sacred Books of the East 34, 38, New York:EDover, repr. 1962; trans. Swami Gambhirananda, *Brahmasūtrabhāṣya of ŚrīŚaṃkarācārya*, Calcutta: Advaita Ashrama, 1965. (Thibaut's is the standard translation, with a good introduction comparing Śaōkara's and Rāmānuja's commentaries; however, Gambhirananda's largely accurate translation is more lucid than Thibaut's.)

Śaōkara (early 8th century) *Upadeśasāhasrī*, ed. and trans. Sengaku Mayeda, *A Thousand Teachings (The Upadeśasāhasrī of Śaṃkara)*, Tokyo: University of Tokyo Press, 1979; Albany, NY: State University of New York Press, 1992.(A translation of Śaōkara's major independent work with an introduction to his thought.)

References and further reading

Clooney, F.X. (1993) *Theology After Vedanta*, Albany, NY: State University of New York Press.(Includes an

exceptionally informative close reading of Śaōkara’s *Brahmasūtrabhāṣya*.)

Pande, G.C. (1994) *Life and Thought of Saṃkarācārya*, Delhi: Motilal Banarsidass. (Includes much historical context and uses a wide variety of sources, but is sometimes insufficiently analytical.)

Potter, K.H. (ed.) (1981) *Encyclopedia of Indian Philosophies*, vol. 3, *Advaita Vedānta up to Śaōkara and his Pupils*, Princeton, NJ: Princeton University Press. (Includes summaries of Śaōkara’s works and a thorough introduction to Advaita philosophy.)

Sāōkhya

Considered one of the oldest classical Hindu schools by Indian tradition, Sāōkhya is most famous in Indian philosophy for its atheism, its dualist model of puruṣa *(passive, individual consciousness) and* prakṛti *(nonconscious, cognitive-sentient body) and its theory that effects pre-exist in their cause. In its classical formulation the* puruṣa-prakṛti *model is analysed into twenty-five components (*tattva*) intended to encompass entire metaphysical, cognitive, psychological, ethical and physical worlds in terms of their embodiment as individual constituents and the creative and interpretive projection of those worlds as experience by and for individuals. Both the world and the individual, in other words, are considered a phenomenological refraction and projection of the underlying and constitutive components of the conscious body.*

Falsely identifying with the cognitive and sensory components of prakṛti *(which according to orthodox Sāōkhya performs cognitive and sentient operations, but is bereft of consciousness;* puruṣa *alone is conscious), Sāōkhyans believe themselves to be the agents of their actions, rather than recognizing that actions are processes lacking any selfhood. Sāōkhyans claim that liberation from the suffering of repeated rebirths can only be achieved through a profound understanding of the distinction between* puruṣa *and* prakṛti. *The latter is not abandoned after liberation, but continues to operate, observed with detachment by* puruṣa. *However, according to some versions of Sāōkhya,* prakṛti *eventually becomes dormant.* Puru ṣa *and* prakṛti *both are considered to be eternal and to have no beginning. Since liberation is achieved through knowledge, Sāōkhya stresses the importance and efficacy of knowledge over ritual and other religious endeavours.*

Sāōkhya is cognate to saōkhyā, *meaning 'to count' or 'enumerate'. Thus Sāōkhya seeks to enumerate the basic facts of reality so that people will understand them and find liberation. Basic Sāōkhyan models and terms appear in some Upaniṣads and underlie important portions of the epic Mahābhārata, especially the Bhagavad Gītā and Mokṣadharma. No distinct Sāōkhyan text prior to Īśvarakṛṣṇa's Sāōkhyakārikā (*c.*350-*c.*450) is extant. It enumerates and explains the twenty-five components and a subsidiary list of sixty topics (*ṣaṣtitantra*), which are then subdivided into further enumerative lists. Most of the subsequent Sāōkhyan literature consists of commentaries and expositions of the Sāṃkhyakārikā and its ideas, which continued to be refined without major alterations well into the eighteenth century. Sāōkhyan models strongly influenced numerous other Indian schools, including Yoga, Vedānta, Kashmir Shaivism and Buddhism.*

1 Historical overview

Sāōkhya claims origins in remote antiquity, identifying its founders with names found in the Vedas - the earliest Indian texts (*c*.2000-*c*.800 BC) - such as Kapila and Āsuri, but nothing in the early material points to these people as having particularly Sāōkhyan views. Some Sāōkhya texts assert that its teachings predate the Vedas. Kauṭilya's *Arthaśāstra* (*c*.300 BC) mentions Sāōkhya as one of the leading teachings of the day. The epic *Mahābhārata* (*c*.400-200 BC) contains sections revealing a sophisticated, elaborate Sāōkhyan system. In one section, the *Mokṣadharma*, the basic system of twenty-five components is already found, along with a twenty-six component system, adding a universal single *puruṣa* as the twenty-sixth item. This universal, cosmic *puruṣa*, or 'ultimate self', was retained in the Yoga school, which in many other respects developed parallel to Sāōkhya. The *Mahābhārata*'s most famous section, the *Bhagavad Gītā*, not only presents a detailed discussion of early Sāōkhya theories, but structures its characters and plot to reflect those models. Both sections, emphasizing the indispensability of knowledge, use the terms 'field' and 'knower of the field' as synonyms for *prakṛti* and *puruṣa* respectively. The Buddhist poet Aśvaghoṣa, in his *Buddhacaritam (Life of the Buddha)* (first to second century), treats Sāōkhya thought as one of the formative teachings studied by the Buddha on his way to enlightenment, suggesting that by Aśvaghoṣa's time there already existed a developed Sāōkhyan tradition considered ancient and influential even by its opponents.

Orthodox Sāōkhya begins with Īśvarakṛṣṇa's *Sāōkhyakārikā*, which synthesized centuries of conflicting Sāōkhyan speculation. Its seventy verses (although the number varies in different commentaries) concisely presents the models, terminology, arguments and systematic configurations that were to be definitive for Sāōkhya from that time on. Virtually every subsequent Sāōkhya text is a commentary on either the *Sāōkhyakārikā* or the *Sāōkhyasūtra* (about fifteenth century). The latter itself is considered to be an expanded, reorganized version of the

Sāōkhyakārikā. The many dozens of commentaries written over the centuries reflect the changing concerns and growing sophistication of Indian thought. While some of the commentaries display clever strategies and great erudition, and minor points are continually being redefined and reinterpreted, the basic parameters set by Īśvarakṛṣṇa are scrupulously followed. The same stock arguments and examples invariably appear century after century in every commentary (Mainkar 1964; Raja 1963) - sometimes with embellishments - but there are no truly creative innovations to the system itself.

The earlier commentaries endeavour to deploy the latest developments in Indian epistemology and argumentative discourse to defend the statements of the *Sāōkhyakārikā* from actual and possible objections. The later commentaries make increasing concessions to non-Sāōkhyan ideologies, often to the point of subverting or reversing the point of distinctive Sāōkhyan teaching while attempting to retain the terminology and basic structure established by Īśvarakṛṣṇa. While orthodox Sāōkhya claimed individuals possessed their own distinct *puruṣa*, later commentators sought to ground this multiplicity of selves in a universal single self. Orthodox Sāōkhya denied the existence of God, but some later commentators reinstated God, or encouraged their readers to reject Sāōkhya's atheistic claims. These later concessions, whose most important advocates were Vacāspati Miśra (ninth to tenth century), Aniruddha (fifteenth century) and Vijñānabhikṣu (sixteenth century), were not so much efforts to change or reform the Sāōkhya system as attempts to make Sāōkhya palatable to contemporary audiences or promote reforms in the non-Sāōkhyan ideologies of the day to which these commentators owed their true allegiance. They may have been emboldened by the Yoga School (whose classical text is Patañjali's *Yogasūtra*), which unlike Sāōkhya, accepted the existence of God who was identified as the best of the *puruṣas*. Yoga also rejected the Sāōkhyan duality of *puruṣa* and *prakṛti*, claiming instead that ultimately the latter dissolves into the former.

2 *Puruṣa* and *prakṛti*

The term *puruṣa* originally meant 'person' and is used in the *Ṛg Veda* to signify the primordial, cosmic person from whom the universe is created (*Puru ṣasūkta* 10.90). As *Ṛg Veda* states, 'Two birds, inseparable companions, have found refuge in the same sheltering tree. One incessantly eats from the peepal tree; the other, not eating, just looks on' (I.24.7). This image of an inseparable dyad, one part actively engaging its appetites and appropriational desires and the other passively observing the activity of the first part, prefigures the notion of *puruṣa* and *prakṛti*.

In Sāōkhya *puruṣa* signifies the observer, the 'witness'. *Prakṛti* includes all the cognitive, moral, psychological, emotional, sensorial and physical aspects of reality. It is often mistranslated as 'matter' or 'nature' - in non-Sāōkhyan usage it does mean 'essential nature' - but that detracts from the heavy Sāōkhyan stress on *prakṛti*'s cognitive, mental, psychological and sensorial activities. Moreover, subtle and gross matter are its most derivative by-products, not its core. Only *prakṛti* acts. *Puruṣa* and *prakṛti* are radically different from each other, although both are considered to be eternal, without a beginning and ultimately inseparable. Every person is constituted of conjoined *puruṣa-prakṛti*. Misunderstanding how these two aspects function and interrelate is the root cause of all problems. Liberation arises from properly distinguishing between them such that one ceases falsely to identify with the components of *prakṛti* as one's self and instead correctly understands that *puruṣa* is a dissociated watcher of the activities of one's *prakṛti*.

Primordially, *prakṛti* is composed of three primary strands or qualities (*guṇa*): *sattva*, which is light, clear, tranquil, joyous, kind; *rajas*, which is passionate, moving, dynamic, agitated, angry; and *tamas*, which is dull, inert, dark, depressed, stupid. Everything in the universe (except *puruṣa*) is composed of varying proportions of these three qualities. Physical objects, mental and emotional states, moral qualities, types of food and personality, for example, are all definable according to which quality predominates. *Rajas* and *tamas* signify the inertia of motion and rest, respectively. *Sattva* signifies the clarity and tranquillity that comes from 'rising above' the negative properties of the other two qualities. For example, by shaking up or moving (*rajas*) something sedimented, depressed, stuck (*tamas*), it may either return to *tamas* (resediment) or remain dynamic (*rajas*), or it may purify and become sattvic. A clear and peaceful condition, if disturbed by passion or violent activity, can transform into reciprocal anger and violence, or produce stupidity, depression and stubbornness. Such dynamics can be applied to psychology, politics, physics, soteriology, or any other field. The three qualities undergo perpetual transformation, those transformations being the empirical universe. The essence of the three qualities remains unmanifest, acting as the cause of the empirical world (manifest *prakṛti*) which, as effect, shares the same nature as its cause. *Sāōkhyakārikā* characterizes their function thus: *sattva* is illuminating, *rajas* is activating or

unfolding and *tamas* imposes limitations and restrictions. They are inseparable, like the wick, oil and flame of a lamp (*tamas*, *rajas* and *sattva*, respectively). *Sattva* and *tamas* are comparable to the Chinese principles of *yang* and *yin*, with *rajas* as the catalyst and dynamic agent that keeps the other two active.

Orthodox Sāōkhya analyses *puruṣa-prakṛti* into twenty-five fundamental components. One is *puruṣa*; the remaining twenty-four are aspects of *prakṛti*, each a further development or transformation of the three qualities. These twenty-five components account for the totality of the universe, as well as the components of each individual. The second component consists of the three qualities taken together, also called 'fundamental *prakṛti*' and the 'unmanifest'. The remaining components, lumped together as the 'manifest' are all formed from combinations of these three. The third of these components is known as *buddhi* which represents reflective discernment and discrimination. *Puruṣa* is most directly connected to *buddhi* through which it is aware of the activities of *prakṛti*. It is *buddhi*'s task to distinguish effectively between *puruṣa* and *prakṛti*. *Ahaṃkāra*, the fourth component, is literally the 'I-maker' or 'the constructor of the "I am"'. The 'I-maker' refracts the three qualities to generate the sensorium, which it then appropriates for itself, thus constructing a sense of subjective selfhood. Thus, it interprets the activities of the three qualities in such a way that it sees itself as the agent or origin of the experience or interpretive theory. It experiences the results as '*my* experience'. The intensity of its sense of self-concern disrupts and obscures *buddhi*'s understanding of its relation with *puruṣa*. The fifth component is the empirical mind (*manas*), which interprets the sensorium and coordinates the discrete sense fields, such as audition and vision, into coherent experience. The next five components are the sense capacities, namely, hearing, touching, seeing, tasting and smelling. The eleventh to fifteenth components are the five activity capacities, such as speaking, grasping, mobility, excreting and procreating. The next five are the subtle elements, such as sound, tactility, colour, form, taste and smell. The five gross elements, ether, air, fire, water and earth, represent the final five components.

3 *Buddhi*

Buddhi alone among the twenty-four components of *prakṛti* directly interacts with *puruṣa*. Orthodox Sāōkhya wishes to maintain a radical distinction between *puruṣa* and *prakṛti*, as well as a *prakṛti* that is fully cognitive, intellective, sensorially aware and active. At the same time it must also be nonconscious (*acetana*). The question of how *buddhi* can be rational, discerning, reflective, discriminative and cognitive and yet lacking consciousness is an interesting one. *Buddhi* must provide both the linkage or communication between *puruṣa* and *prakṛti* and yet generate the discernment that realizes their ultimate separation. Since Sāōkhyan soteriology relies on *buddhi* to discern that separation, and the relation between the radically incommensurate *puruṣa* and *prakṛti* resides in the function of *buddhi*, the coherence of the entire system rests on the coherence of the notion of *buddhi*.

Prakṛti is not 'insentient', since the senses and their functions operate entirely within *prakṛti*. The five sense capacities are said to possess 'bare awareness' of their objects (see Sense perception, Indian views of). The empirical mind is described as both a sense organ and a conceptualizer. The 'I-maker' generates the sense of subjectivity. *Buddhi* discriminates, decides, reflects and provides certainty.

Reserving conscious knowing exclusively for *puruṣa*, while assigning all cognitive, psychological and sensory activities to *prakṛti* only seems confusing, argues *Sāōkhyakārika*. This is because we transfer mistakenly the properties of one to the other due to their proximity: *prakṛti* seems to be conscious and *puruṣa* seems to be active, but in fact they are not. Iron becomes hot when in proximity with heat, but heat is not in its nature, they argue. *Puruṣa* and *prakṛti* support each other and are compared to the symbiosis of a blind man and a lame man. In order to travel, the blind man (that is, *prakṛti*), who cannot see where he is going can carry the lame man (that is, *puruṣa*), who cannot walk but can see and direct the blind man. Since directing a blind man may be construed as an activity and the blind man must be sentient to cooperate with and be useful to the lame man, this example confuses rather than clarifies the problem.

In the later Yoga system, which also adopted a version of the twenty-five components, *puruṣa* shines a light into *buddhi* which *buddhi* then disperses throughout the rest of *prakṛti*. In other words, *puru ṣa* infuses *prakṛti*. Thus the 'light' of reason, sensation and cognition in *prakṛti* is the refracted vision of *puruṣa*. Enlightenment and liberation are produced by realigning *buddhi* such that *puruṣa*'s light is reflected back into *puruṣa*, producing 'self-illumination'. This structure, however, is not entertained by Sāōkhya, according to which *puruṣa* is passive and thus can illuminate nothing. *Sāōkhyakārika* 33-7 explains its relation to *buddhi* in terms of a distinction

between the internal organs (*buddhi*, the 'I-maker' and empirical mind) and the external organs (the ten capacities, that is, five sense capacities and five activity capacities). The external organs only perceive objects in the present; the internal organs can also perceive objects in the past and future. All objects perceived by the external organs are also perceived by the internal organs. Although each organ is different as each is derived from distinct differentiations of the three qualities, together 'like a lamp', they illumine *buddhi* for the sake of *puruṣa*. This is so *buddhi* can provide *puruṣa* with every experience and differentiate the subtle otherness between *puruṣa* and *prakṛti*. Thus, *Buddhi*, both causes *puruṣa*'s suffering and effects its liberation (see Awareness in Indian thought).

4 Orthodox epistemology

Sāōkhyakārikā, like Buddhism and many other Indian schools, claims that the root problem as experienced is a profound sense of dis-ease and dissatisfaction (*duḥkha*) that can only be cured by knowledge. Some commentaries say this knowledge concerns the difference between *puruṣa* and *buddhi*; some say it is knowing the twenty-five components or the list of sixty topics. Sacrifice, ritual and scriptural remedies, the usual methods for an orthodox Hindu, are declared ineffective for anything except temporary rebirth in heaven. Moreover, those methods cannot guarantee success (for example, even authentic prayer and religious observance may not cure barrenness) and they often involve 'impurity' (the killing during sacrifice, the caste requirements to have intercourse with one's mate). Heavenly births and impurities both perpetuate the cycle of life and death rather than resolving it. Like Buddhists, Sāōkhyans consider existence in any realm into which one is reborn, whether human, or heavenly, to be transitory.

Sāōkhya accepts three means of acquiring knowledge: perception, inference and reliable testimony. Descriptions of all three vary. Among Īśvarakṛṣṇa's predecessors, Vārṣagaṇya (*c*.300-*c*.100 BC) reportedly defined perception as the function of sense organs; Vindhyavāsin (*c*.400-*c*.300 BC) added that valid perception was devoid of mental or linguistic construction. Īśvarakṛṣṇa added that perception ascertains specific and definite objects. The *Yuktidīpikā*, one of the fullest and most sophisticated of the commentaries on *Sāōkhyakārika*, further asserts that the instrumentality of perception (that is, the *pramāṇa*) is located in *buddhi*, but its results occur in *puruṣa*.

The discussions of inference also vary across the commentaries, reflecting developments in Indian logic (see Inference, Indian theories of). *Sāōkhyakārikā* cryptically states that inference is threefold, based on a mark and its antecedent mark. The *Sāōkhyavṛtti* (author and date unknown) describes the three types of inference as inferring from what precedes, for example, imminent rain from storm clouds; inferring the whole from a part, for example, inferring all sea water is salty by tasting one drop; and generalization, for example, inferring that because one tree is in bloom others must be likewise. The *Suvarṇasaptati*, a *Sāōkhyakārikā* commentary that only survives in a Chinese translation made by Paramārtha in the sixth century, states that inference is dependent on perception. It redefines the three types in the following terms: inference from what precedes means to infer an effect from the perception of a cause, that is, rain from a black cloud; inferring the whole from the part means to infer a cause from the perception of an effect, such as previous rain from a flood; and examples for generalization are self-evident. Gau ḍapāda (sixth century) in his commentary *Sāōkhya kārikābhāṣya*, for 'generalization' or 'general correlation' gives as an example seeing someone first in one place and then another, which denotes movement although the moving itself may not have been perceived.

This last example is important as orthodox Sāōkhya argues that merely because something is not perceived, this does not mean that it does not exist, as when one knows the Himalayas have a top without actually seeing it. Imperceptibles may be proven to exist through inference. Since Sāōkhyans are atheists, it is not God whose existence they are trying to prove but rather *puruṣa* and unmanifest *prakṛti*. *Sāōkhyakārikā* 8 states that *prakṛti* is too subtle to be perceived, but it is 'perceived through its effects'. Sāōkhyan arguments for the knowability of imperceptibles by means of inference are disappointing since they fail to distinguish between what is not perceived but in principle perceptible, and what is genuinely imperceptible. All their examples address the former variety. Even when the *Yuktidīpikā* records this objection from a Buddhist, its response continues to ignore the distinction.

The most repeated 'proof' for the existence of *puruṣa* is a questionable argument from design (for example, *Sāōkhyakārikā* 17). It states that the body, or *prakṛti*, like a bed, is composite in nature (implying it was put together), therefore it must be for something other than itself. The only existent other than *prakṛti* is *puruṣa*, therefore *prakṛti* and its operations must be for the benefit of *puruṣa*. There is a usual argument offered to prove that the imperceptible unmanifested three qualities are the cause of the manifest world. It is that an effect must be of the same nature as its cause and the effects of the three qualities are observable everywhere in the manifest

world, as everything can be seen as varying proportions of the three qualities. This is illustrated by an analogy of cloth and thread. Thread when woven becomes cloth, although in nature there is no difference between thread and cloth. Since for the Sāōkhyans the qualities function as material causes as well as efficient causes, they were satisfied with this argument, although the thread does not transform the manner posited of *prakṛti*. The *Yuktidīpikā* states that fundamental *prakṛti*, because it is too subtle lacks the 'function of movement', but instead has the 'function of transformation'. Transformation, it says, is when an object obtains new qualities without deviating from its essence. An example might be found in the way a piece of black *palāśa* wood becomes yellow because of exposure to heat and yet never stops being *palāśa* wood. However, the change from threads to cloth is certainly not a transformation in this sense.

The *Yuktidīpikā* describes the ten factors that constitute logical proof. These include five psychological 'conditions for an inference': desire to know, doubt, purpose, examining alternative possibilities, removal of doubts; and the five parts of an actual syllogism: thesis, reason, example, application and conclusion. The *Mātharavṛtti*, another *Sāōkhyakārikā* commentary, discusses the later Indian three- and five-part syllogisms in a manner consistent with their explication in Nyāya and Buddhism.

As for testimony or reliable authority, like most Hindu schools Sāōkhya accepts scripture and the testimony of a reliable person as valid means to knowledge. The *Yuktidīpikā*, for example, asserts that the legendary ancient preceptors of Sāōkhya, such as Pañcaśikha (mentioned in the *Mahābhārata*), actually perceived the effect in the cause so that the Sāōkhyan causal theory rests on their authority (see Testimony in Indian philosophy).

5 Causal theory

Sāōkhyakārikā holds that the effect pre-exists in the cause (*satkāryavāda*) in a latent or potential state, arguing that since something cannot arise from nothing, the effect must pre-exist. It further claims that all effects rely on a material cause. Things do not arise indiscriminately from just anything: certain types of causes produce certain types of effect, for example, cows do not give birth to puppies. Something can only produce what it is capable of producing and a producer can only produce what is capable of being produced; the nature of the cause is in the effect. The same text says elsewhere that all manifest things must have a single ultimate cause to avoid an infinite regress of causes and effects. This ultimate cause is *prakṛti*.

Other causal notions are found in the commentaries. Gauḍapāda's commentary states that each of the three qualities produces conditions conducive to the other two, as well as to itself: 'thus *sattva*, like a beautiful woman, is a joy to her husband, a trial to her co-wives and arouses passion in other men' (*sattva*, *tamas* and *rajas*, respectively). The *Sāōkhyasaptativṛtti* discusses two types of cause: productive causes, which entail fundamental *prakṛti*, *buddhi*, the 'I-maker' and the subtle elements and cognitive causes, which includes five types of misconception, twenty-eight types of dysfunction (of the sense and activity capacities, mind and *buddhi*), nine contentments, eight types of attainment and the cognitive constructions of the eight predispositions.

There are also various discussions on material, efficient and instrumental causes, but these are merely adopted from other Indian schools. Sāōkhya seemed more interested in specific causal sequences and correlations between its various enumerated lists than in an elaborate theory of causality itself. The *Yuktidīpikā* criticizes Buddhist 'momentariness' and the causal theories of Vaiśeṣika and Nyāya, but the arguments shed little light on causal theory in general (see Causation, Indian theories of).

6 Soteriology

Sāōkhyakārikā states that the purpose for which the experiential world is created by the conjoining of *puruṣa* and *prakṛti* is so that the former can 'see' the latter, and the latter can liberate or isolate the former. Everything *prakṛti* does, she does for the liberation of *puruṣa*: 'just as unaware milk operates as an efficient cause to nourish a calf, so does *prakṛti* operate as an efficient cause for the liberation of *puruṣa*' (*Sāōkhyakārikā* 57). As a dancer stops dancing once the audience has seen the performance, so does *prakṛti* stop after having illumined *puruṣa*. She, like a self-sacrificing mother, does everything for the sake of *puruṣa*, while *puruṣa* lends no assistance whatsoever.

The *Sāōkhyasaptativṛtti* declares that *buddhi* enlightens and attains liberation for *puruṣa*. Critics of Sāōkhya have been quick to point out that *puruṣa* is defined in *Sāōkhyakārikā* 19 as inherently liberated and indifferent to pain and pleasure, thus it should be in no need of liberation. Sāōkhya's response to this criticism again hinges on

buddhi, but with an additional notion known as the ‘predispositions’ (*bhāva*). The predispositions that bind and liberate are classified into three types and into eight types: the three are innate, natural and acquired; the eight are *dharma* (meritorious action), which leads to rebirth in a higher life, adharma (demerit), which leads to lower births, knowledge, which leads to liberation, ignorance leading to bondage, detachment leading to dissociation from the activities of *prakṛti*, attachment producing the cycle of birth and death, power conducive to controlling circumstances and impotence leading to loss of control. Innateness, acquisition, detachment leading to dissociation from activities of *prakṛti* and power conducive to controlling forces are considered sattvic; their opposites are considered tamasic.

The *Sāōkhyasaptativṛtti* defines the innate predispositions as *dharma*, knowledge, detachment and power; the natural predispositions are latencies that emerge at certain times in one’s life and the acquired dispositions are derived from learning the truth from a teacher. Gauḍapāda defines the innate as the eight predispositions constituting one’s inherent nature; nature is the eight dispositions resulting from previous lives and the acquired is the eight predispositions obtained by hearing the truth from a teacher. He adds that the predispositions reside in *buddhi* and shape the (gross) embryo. Of the eight, knowledge alone leads to liberation. To the extent that the other seven contribute to knowledge they are conducive to liberation, but only knowledge can effect release.

The predispositions and the subtle body cannot operate without each other. The subtle body is composed of the internal organs, the ten capacities and the five subtle elements. It does not experience anything, but is the repository that holds the predispositions. Impelled by these predispositions, the subtle body is reborn from life to life until the predispositions are eliminated by knowledge. The subtle body is neither a self nor invariant since it is perpetually being modified by the predispositions influenced by the fluctuations of *prakṛtic* experience.

Puruṣa, prior to liberation, watches *prakṛti*’s transformations and suffers the pain of old age and death. But those transformations are merely an unconscious movement, a dance designed to show *puruṣa* that in its own nature it is never bound or liberated. At the next stage, while ‘standing aside like a spectator, *puruṣa* views *prakṛti* who, having fulfilled her purpose, stops, turning her back on the seven forms’ (v.65), that is, all the predispositions except knowledge. Although *puruṣa* and *prakṛti* are still conjoined, no new creation is generated and no new predispositions are created. By the attainment of correct knowledge, the seven predispositions cease to cause further embodiment and yet, ‘like a potter’s wheel that continues to spin even after the potter has stopped applying force’, embodiment continues for a while. Having fulfilled its purpose, *prakṛti* ceases functioning, sometimes understood to mean that the qualities return to an equilibrium from which no further transformations emerge. Attaining separation from the body, *puruṣa* attains everlasting ‘isolation’ or ‘freedom’ (*kaivalya*).

See also: Dualism; Karma and rebirth, Indian conceptions of; Self, Indian theories of

DAN LUSTHAUS

References and further reading

Āraṇya, Svāmī Hariharānanda (1977) *The Sāṃkhya-Sūtras of Pañcaśikha and the Sāṃkhyatattvāloka*, Delhi: Motilal Banarsidass.(Includes the Sanskrit text in Devanagri with an English translation, although the English leaves much to be desired.)

Chakravarti, P. (1951) *Origin and Development of the Sāṃkhya System of Thought*, Calcutta: Metropolitan Printing and Publishing House Ltd.(A substantial treatment of the early sources.)

Chapple, C. (1986) *Karma and Creativity*, Albany, NY: State University of New York Press.(Deals with Sāōkhyan models found in several texts, including the *Mahābhārata* and *Yogavāsiṣṭha* in an insightful and lucid manner.)

De Nicolas, A. (1976) *Avatāra: The Humanization of Philosophy Through the Bhagavad Gītā*, New York: Nicolas Hays.(Penetrating, original interpretation of the *Gītā* that includes a superior English translation.)

Johnston, E.H. (1937) *Early Sāṃkhya*, London: Royal Asiatic Society.(An important examination of the early Sāōkhya tradition.)

Larson, G. (1979) *Classical Sāṃkhya*, Delhi: Motilal Banarsidass.(Includes review of earlier scholarship, an analysis of *Sāōkhyakārikā* with romanized Sanskrit text and translation in an appendix.)

Larson, G. and Ram Shankar, B. (eds) (1987) *Sāṃkhya: A Dualist Tradition*, Princeton, NJ: Princeton University Press.(A volume in the *Encyclopedia of Indian Philosophies*, this is the most comprehensive English work to date on the subject. Includes historical and philosophical overviews, summaries of every major text and thinker in the tradition and a careful delineation of all the major and minor terms and lists.)

Mainkar, T.G. (trans.) (1964) *The Sāṃkhyakārikā of Īśvarakṛṣṇa with the Commentary of Gauḍapāda*, Poona: Oriental Book Agency.(An interesting translation.)

Raja, C.K. (1963) *The Sāṃkhyakārikā of Īśvarakṛṣṇa*, India: V.V. Research Institute.(Translation with romanized Sanskrit text, includes the translator's commentary drawn from Gauḍapāda and Vacāspati Miśra.)

Takakusu, M. (1904) 'La *Sāṃkhyakārikā* étudiée à la lumière de sa version chinoise (II)' (The *Sāōkhyakārikā* Studied in the Light of the Chinese Version), *Bulletin de l'École Française d'Extrême Orient* 4: 978-1064.(French translation of the commentary translated into Chinese by Paramārtha.)

Sapir-Whorf hypothesis

The Sapir-Whorf Hypothesis is a widely used label for the linguistic relativity hypothesis, that is, the proposal that the particular language we speak shapes the way we think about the world. The label derives from the names of American anthropological linguists Edward Sapir and Benjamin Lee Whorf, who persuasively argued for this idea during the 1930s and 1940s - although they never actually characterized their ideas as an 'hypothesis'. In contrast to earlier European scholarship concerned with linguistic relativity, their approach was distinguished by first-hand experience with native American languages and rejection of claims for the superiority of European languages.

Early in the twentieth century, American anthropologist Franz Boas (1858-1942) inaugurated an important expansion of scientific investigation of the languages of native North America. As part of a broad critique of nineteenth-century evolutionary arguments he stressed the equal value of each language type and their independence from race and cultural level. He argued that each language necessarily represents an implicit classification of experience, that these classifications vary across languages, but that such variation probably has little effect on thought or culture.

His student Edward Sapir (1884-1939) accepted the main thrust of Boas' position but came to feel that the closely knit system of categories in a language could represent incommensurable analyses of experience with effects on speakers' conceptual view points and aesthetic interpretations. Gestalt and psychoanalytic psychology and Sapir's own literary efforts also played a role in his thinking on this issue. Sapir's concern was not with linguistic form as such (for example, whether a language uses inflections or not), nor with linguistic content or meaning as such (for example, whether a language could refer to a particular referent), but rather with the formal organization of meaning characteristic of a language, the regular ways meanings are constructed (for example, grammatical categories and patterns of semantic composition). Despite the suggestiveness of his formulation, Sapir provided few specific illustrations of the sorts of influences he had in mind.

Benjamin Lee Whorf (1897-1941), a gifted amateur linguist independently interested in these issues as they related to the nature of science, came into contact with Sapir in 1930 and began developing these views in a more systematic way. He analysed particular linguistic constructions, proposed mechanisms of influence, and provided empirical demonstrations of such influences on belief and behaviour. However, his views on this issue are known to us largely through letters, unpublished manuscripts and popular pieces, which has led to considerable debate about his actual position. In this context, the one article on this issue prepared for a professional audience must be given special weight (see Whorf 1956).

Whorf argued that each language refers to an infinite variety of experiences with a finite array of formal categories (both lexical and grammatical) by grouping experiences together as analogically 'the same' for the purposes of speech. These categories also interrelate in a coherent way, reinforcing and complementing one another, so as to constitute an overall interpretation of experience. Languages vary considerably not only in the basic distinctions they recognize, but also in the assemblage of these categories into a coherent system of reference. Thus the system of categories which each language provides to its speakers is not a common, universal system, but one peculiar to the individual language, and one which makes possible a particular 'fashion of speaking'.

But speakers tend to assume that the categories and distinctions of their language are natural, given by external reality. Further, speakers make the tacit error of assuming that elements of experience which are classed together on one or another criterion for the purposes of speech are similar in other respects as well. The crux of Whorf's argument is that these linguistic categories are used as guides in habitual thought. When speakers attempt to interpret an experience in terms of a category available in their language they automatically involve the other meanings implicit in that particular category (analogy) and in the overall configuration of categories in which it is embedded. And speakers regard these other meanings as being intrinsic to the original experience rather than a product of linguistic analogy. Thus, language does not so much blind speakers to some obvious reality, but rather it suggests associations which are not necessarily entailed by experience. Ultimately, these shaping forces affect not only everyday habitual thought but also more sophisticated philosophical and scientific activity. In the absence of another language (natural or artificial) with which to talk about experience, speakers will be unlikely to

recognize the conventional nature of their linguistically-based understandings.

The ideas of Sapir and Whorf have attracted widespread attention in the humanities and social sciences. Their views have been important in leading many students into comparative linguistics and played a crucial role in giving rise to the field of psycholinguistics in the 1950s. Despite this wide currency, their views have not been subjected to much empirical research. In large part, acceptance or rejection of their proposals has had more to do with the personal and professional outlook of the investigator and the prevailing temper of the times than with any solid evidence. In particular, philosophical responses, whether sympathetic or derisive, have rarely engaged with real linguistic phenomena (that is, how languages actually differ) or with practical cognition (that is, what kinds of effects one might expect). Instead debate centres on the logical (im)plausibility of linguistic incommensurability and determinism.

Existing empirical research on Whorf's claims has consisted primarily of tests for a relationship between lexical or grammatical categories (for example, words for colour, number marking) and experimental assessments of patterns of memory and classification. There have also been some attacks on Whorf's particular evidence, especially his claims about the Hopi language. This empirical research remains controversial, but some results clearly support Whorf's proposals and none decisively contradicts them. Another important strand of thinking has considered whether diverse uses of language (in particular the specialized discursive forms associated with language standardization, literacy and formal education) might have effects on thinking either in their own right or by mediating the structural effects proposed by Sapir and Whorf. This research often focuses on the cognitive importance of decontextualized speech and on the nature of speakers' conscious awareness and control of language - this latter an issue of concern to Whorf himself. Direct empirical research on such possible discursive effects remains quite limited.

There is also a growing body of research exploring the historical development of Sapir and Whorf's views, including intellectual biographies, more complete collections of their works, and accounts of the relation of their views to earlier European thinkers, especially in Germany. It seems certain that they were aware of this earlier work: Boas trained in Germany, Sapir wrote a Masters paper on Herder, and both Sapir and Whorf explicitly reject Humboldt's main substantive thesis, though not by name (see Herder, J.G.; Humboldt, W. von). However, their work is better understood, as with the rest of twentieth-century American anthropology, as an innovative departure from Europeanist views stimulated by direct first-hand contact with native American languages and cultures in a context of an emerging non-hierarchical view of culture.

See also: Austin, J.L.; Cassirer, E.; Condillac, E.B. de; Determinism and indeterminism; Diderot, D.; Hamann, J.G.; Language, philosophy of; Putnam, H.; Quine, W.V.; Radical translation and radical interpretation; Relativism; Searle, J.R.; Wittgenstein, L.

JOHN A. LUCY

List of works

Sapir, E. (1921) *Language: An Introduction to the Study of Speech*, New York: Harcourt, Brace & Co.(Classic treatment of language from an anthropological perspective; important for his early views; chapters 1, 4, 5, 10 and 11 deal with matters relevant to language and thought.)

Sapir, E. (1924) 'The Grammarian and His Language', repr. in D.G. Mandelbaum (ed.) *The Selected Writings of Edward Sapir in Language, Culture, and Personality*, Berkeley, CA: University of California Press, 150-9. (States his views on the diversity among languages as systems.)

Sapir, E. (1929) 'The Status of Linguistics as a Science', repr. in D.G. Mandelbaum (ed.) *The Selected Writings of Edward Sapir in Language, Culture, and Personality*, Berkeley, CA: University of California Press, 160-6. (States his view, especially on page 162, that language differences shape a speaker's view of reality.)

Sapir, E. (1931) 'Conceptual Categories in Primitive Languages', repr. in D.H. Hymes (ed.) *Language in Culture and Society: A Reader in Linguistics and Anthropology*, New York: Harper & Row, 1964, 128.(Dense, one-page summary of the shaping power of language at several levels.)

Sapir, E. (1933) 'Language', repr. in D.G. Mandelbaum (ed.) *The Selected Writings of Edward Sapir in Language, Culture, and Personality*, Berkeley, CA: University of California Press, 7-32.(Important for the explicit dismissal of nineteenth-century views of the Humboldtian sort regarding the correlation of language type, such as inflectional or isolating, with cultural type or level.)

Whorf, B.L. (1956) *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed. and intro. J.

Carroll, Cambridge, MA: MIT Press.(The basic source for Whorf's writings on language and thought. Reprints 'The relation of habitual thought and behavior to language' his most important work for a professional audience (pages 134-59), several articles written for a popular audience (pages 207-70), and important unpublished manuscripts (pages 57-125). Carroll's introduction provides a brief intellectual biography and a bibliography of published and unpublished works.)

References and further reading

Aarsleff, H. (1988) 'Introduction', in W. von Humboldt, *On Language: The Diversity of Human Language-Structure and its Influence on the Mental Development of Mankind*, trans. P. Heath, Cambridge: Cambridge University Press, vii-lxv.(Useful summary of eighteenth- and nineteenth-century European views on linguistic relativity making clear en route their differences from contemporary views.)

Bertefflany, L. von (1955) 'An Essay on the Relativity of Categories', *Philosophy of Science* 22: 243-63. (Sympathetic treatment of the Sapir-Whorf hypothesis as analogous to other, commonly accepted types of incommensurability.)

Black, M. (1959) 'Linguistic Relativity: The Views of Benjamin Lee Whorf', *The Philosophical Review* 68: 228-38.(Influential article attacking Whorf's views as logically flawed.)

Boas, F. (1911) 'Introduction', *Handbook of American Indian Languages, Bureau of American Ethnology Bulletin* 40 (1): 1-83.(Classic statement of the Boasian view of language; see especially §2 on diversity among languages; various reprints available.)

Hymes, D. (1964) 'Two Types of Linguistic Relativity', in W. Bright (ed.) *Sociolinguistics, Proceedings of the UCLA Sociolinguistics Conference*, The Hague: Mouton, 114-65.(Early effort to consider the implications of work on language use or pragmatic function for Whorf's ideas.)

Gipper, H. (1972) *Gibt es ein sprachliches Relativitätsprinzip? Untersuchungen zur Sapir-Whorf-Hypothese*, Frankfurt am Main: S. Fischer; trans. R. Pinxton, 'Is There a Linguistic Relativity Principle?', in R. Pinxton (ed.) *Universalism versus Relativism in Language and Thought*, The Hague: Mouton, 1977, 217-28. (Contemporary German view on the Sapir-Whorf hypothesis, partly reproduced in translation.)

Gumperz, J.J. and Levinson, S.C. (eds) (1996) *Rethinking Linguistic Relativity*, Cambridge: Cambridge University Press.(Collection of contemporary anthropological research on issues related to renewed interest in the Sapir-Whorf hypothesis. Includes a review by Lucy of approaches suggesting relativity due to differences in language use.)

Koerner, E.F.K. (1992) 'The Sapir-Whorf Hypothesis: A Preliminary History and a Bibliographic Essay', *Journal of Linguistic Anthropology* 2: 173-8.(Useful brief introduction to the various historical studies of the Sapir-Whorf tradition and its various precursors. Good bibliography.)

Lee, P. (1996) *The Whorf Theory Complex: A Critical Reconstruction*, Amsterdam: John Benjamins.(Close intellectual history of Whorf and those who have worked on his ideas; makes heavy use of archival materials; extensive bibliography.)

Lucy, J. (1992) *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*, Cambridge: Cambridge University Press.(Close textual analysis of the views of Boas, Sapir and Whorf and a critical analytic review of subsequent empirical research. Extensive bibliography of the social science literature.)

Lucy, J. (1992) *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis*, Cambridge: Cambridge University Press.(Outlines a method for investigating the Sapir-Whorf hypothesis and provides an extensive empirical investigation.)

Penn, J. (1972) *Linguistic Relativity versus Innate Ideas: The Origins of the Sapir-Whorf Hypothesis in German Thought*, The Hague: Mouton.(Traces important historical connections with earlier work.)

Rollins, P. (1972) *Benjamin Lee Whorf: Transcendental Linguist*, unpublished Ph.D dissertation, Harvard University, Cambridge, MA.(Biographical information on Whorf's intellectual interests. Volume 3 reprints some of Whorf's unpublished materials. The main arguments are reworked in *Benjamin Lee Whorf: Lost Generation Theories of Mind, Language, and Religion*, Ann Arbor, MI: University Microfilms, 1980.)

Silverstein, M. (1979) 'Language Structure and Linguistic Ideology', in P. Clyne, W. Hanks and C. Hofbauer (eds) *The Elements: A Parasession on Linguistic Units and Levels*, Chicago, IL: Chicago Linguistic Society, 193-247.(Influential extension of Whorf's arguments to questions of language use and native-speaker ideology.)

Sartre, Jean-Paul (1905-80)

Sartre was a philosopher of paradox: an existentialist who attempted a reconciliation with Marxism, a theorist of freedom who explored the notion of predestination. From the mid-1930s to the late-1940s, Sartre was in his 'classical' period. He explored the history of theories of imagination leading up to that of Husserl, and developed his own phenomenological account of imagination as the key to the freedom of consciousness. He analysed human emotions, arguing that emotion is a freely chosen mode of relationship to the outside world. In his major philosophical work, L'Être et le Néant (Being and Nothingness) (1943a), Sartre distinguished between consciousness and all other beings: consciousness is always at least tacitly conscious of itself, hence it is essentially 'for itself' (pour-soi) *- free, mobile and spontaneous. Everything else, lacking this self-consciousness, is just what it is 'in-itself'* (en-soi)*; it is 'solid' and lacks freedom. Consciousness is always engaged in the world of which it is conscious, and in relationships with other consciousnesses. These relationships are conflictual: they involve a battle to maintain the position of subject and to make the other into an object. This battle is inescapable.*

Although Sartre was indeed a philosopher of freedom, his conception of freedom is often misunderstood. Already in Being and Nothingness *human freedom operates against a background of facticity and situation. My facticity is all the facts about myself which cannot be changed - my age, sex, class of origin, race and so on; my situation may be modified, but it still constitutes the starting point for change and roots consciousness firmly in the world. Freedom is not idealized by Sartre; it is always within a given set of circumstances, after a particular past, and against the expectations of both myself and others that I make my free choices. My personal history conditions the range of my options.*

From the 1950s onwards Sartre became increasingly politicized and was drawn to attempt a reconciliation between existentialism and Marxism. This was the aim of the Critique de la raison dialectique (Critique of Dialectical Reason) *(1960) which recognized more fully than before the effect of historical and material conditions on individual and collective choice. An attempt to explore this interplay in action underlies both his biography of Flaubert and his own autobiography.*

1 Background

Sartre's prestige as a philosopher was at its peak, in France at least, in the late 1940s, in the aftermath of the Second World War, when a philosophy of freedom and self-determination fitted the mood of a country recently liberated from the Occupation. It was at its nadir in the late 1960s and 1970s when structuralism had discredited, temporarily, both humanism and existentialism, and proclaimed 'man' to be no more than a locus of forces traversed and indeed produced by social and linguistic structures (see Structuralism). The British analytic tradition has never had much time for the literary and dramatic aspects of existentialism and phenomenology, though some recent critics, such as Phyllis Morris (1974) and Gregory McCulloch (1994), have attempted to take Sartre seriously as a philosopher and to assess his contribution in terms more accessible to analytically trained minds.

The emotive responses tend in their different ways to distort Sartre's arguments and to focus, for example, on one of the poles of the many paradoxes which his philosophy implies. For Sartre is indeed a philosopher of paradox - deliberately facing his readers with logically 'impossible' or self-contradictory statements in order to force them to think beyond the confines of the binary oppositions to which common sense and analytic reason have accustomed them. For example, 'Man is what he is not and is not what he is' (1943a: 97), provocatively compels the reader who perseveres to confront the difficult issues of the relationship between essence, existence and negation. 'Man is what he is not', that is to say, man is a being without an essential nature, a being who operates through negation, who cannot be identified with his past, or indeed his present self, and so 'who is not what he is'.

1940s Paris overestimated Sartre's faith in human freedom and lauded him for it; 1960s Paris made the same mistake and discarded him along with all other relics of mid-century humanism. Neither period read Sartre carefully enough to recognize the constraints and limits within which freedom was, from the outset, deemed to operate.

2 Early philosophy

Sartre's first published philosophical works were *L'Imagination* (1936a), a history of theories of imagination up to

the theory of Edmund Husserl, and 'La Transcendance de l'ego' (*The Transcendence of the Ego*) (1936b). *The Transcendence of the Ego* shows hostility to any kind of essentialism of the self. In it Sartre argues (against Husserl) that the ego is not transcendental but transcendent, that is, it is not an inner core of being, a source of my actions, emotions and character, but rather a construct, a product of my self-image and my image in the eyes of others, of my past behaviour and feelings. Sartre maintains that consciousness is not essentially first-person but is impersonal, or at most pre-personal, and that it is characterized by intentionality, that is to say it is always directed at something other than itself. In this context Sartre positions himself in relation to the Kantian 'unity of apperception', arguing that although the 'I think' must be able to accompany all my representations, it does not always do so, at least explicitly. I may turn my attention at any moment away from what I am doing and direct it towards myself as agent, but this reflexivity is not a permanent, thetic feature of consciousness. Later, in *L'Être et le Néant (Being and Nothingness)* (1943a), Sartre claims that it is precisely this very reflexivity - the self-consciousness of consciousness - that personalizes consciousness and constitutes the human subject, but in *The Transcendence of the Ego* such a notion is absent and he is more concerned to argue against the identification of consciousness with selfhood than to explore the ways in which consciousness relates to the notion of subject.

In his *Esquisse d'une théorie des émotions (Sketch for a Theory of Emotions)* (1939) Sartre turns his attention to another area of human experience in order to show that this, in its turn, cannot be described in essentialist terms. Emotions, in Sartre's account, are chosen rather than caused: emotion involves a 'magical' attempt to transform reality by changing what can be changed (my own feelings) rather than what is less easily malleable, that is, the outside world. In the face of extreme danger I may faint from fear: the danger has not disappeared but I am no longer conscious of it. Sartre here takes a radical position which he maintained but modified in later years, as his recognition of the degree to which we are formed by external conditions gradually increased. He is careful to distinguish between various areas related to emotion - passion, feeling and so on. Emotion is not sustainable continuously through time, but is subject to fluctuations of intensity, and may at times be replaced by alternative feelings. In this sense too Sartre rejects essentialism: like Proust he believes in the 'intermittances of the heart': love, for example, is not a continuous emotional state, but an amalgam of affection, desire, passion, as well as, perhaps, jealousy, resentment and even occasionally hatred. Love is not the permanent compelling state we may like to imagine: it is the product of a decision and a commitment (see Emotions, nature of §4; Emotions, philosophy of §4).

These two works form the grounding for Sartre's early theory of human freedom along with a second work on the imagination. In *L'Imaginaire, psychologie phénoménologique de l'imagination (The Psychology of the Imagination)* (1940) Sartre picks up the threads of Husserl's theory of imagination and develops it further by showing how phenomenological psychology works in practice. Unlike traditional empirical psychology it is not based in a positivist methodology in which evidence depends on an accumulation of examples. The phenomenological method operates through a particular type of introspection or intuition in which the phenomenologist examines a single example, or a series of examples, of the phenomenon to be analysed (here imagination) and deduces from the example the general principles and features of the phenomenon. In this way Sartre describes what he calls the 'poverty' of the image - the fact, that is, that I can never find in it any more than I have already put there. If, say, I do not know the number of columns in the Parthenon, I can count them if I look at the temple in reality; if I merely imagine the temple the number of pillars will depend not on the real building but merely on my own implicit estimate. I cannot learn anything from imagination as I can from perception. But the reverse of this 'poverty' of the imagination is its *freedom* - in imagination I am not constrained as I am in perception by the material world around me. Indeed, imagination is not merely image formation - in Sartre's account it is itself constitutive of the freedom of consciousness. Without imagination we would be 'stuck in the real', unable to escape from the present moment of time and our immediate surroundings. It is imagination that allows us to step back from our material environment and take up an (imaginary) distance from it, in Sartre's terms to 'totalize' it, to see it as a 'world' with order and pattern. In *Being and Nothingness* Sartre will also maintain that the imagination is the source of the purpose and finality we see in the world, but *The Psychology of the Imagination* concentrates rather on the different functions of imagination and image formation in the narrower sense (see Imagery; Imagination §2).

3 *Being and Nothingness*

Being and Nothingness sets out the main philosophical tenets of the 'classical' Sartre. Being is subdivided, as it

were, into two major regions - being for-itself (*l'être pour-soi*) or consciousness, and being-in-itself (*l'être en-soi*) which is everything other than consciousness, including the material world, the past, the body as organism and so on. To being-in-itself Sartre devotes no more than six of his 660 pages; there is little to be said about it other than it is, it is what it is, and it is 'in itself'. Only through the 'for-itself' of consciousness does the 'in-itself' become a world to speak of. Indeed, Sartre argues, we cannot know anything about being as it is, only about being as it appears to us. It is through consciousness that the world is endowed with temporality, spatiality and other qualities such as usefulness. This is where the imagination in its broadest sense may be seen as primary: 'imagination is the whole of consciousness as it realizes its freedom' (1940: 236). Imagination makes a world of the 'in-itself', it totalizes and 'nihilates' it. Nihilation (*néantir*) is a term that is specific to Sartre, and means not annihilation but rather the special type of negation that consciousness operates when it 'intends' an object: it differentiates the object from its surroundings and knows itself *not* to be that object. But consciousness is not alone in the world it has created from the brute 'in-itself', indeed it has not created the world individually, but rather as part of an intersubjective community. And other people, or their consciousnesses, are not an afterthought for Sartre. Like Heidegger he sees man as always already engaged in relationships with others; unlike Heidegger he sees these not in terms of *Mitsein* (Being-with), but in terms of conflict in a manner reminiscent of the account given by Hegel of the relationship between masters and slaves. The other is in permanent competition with me. I wish to be a subject and make of the other an object, while he or she attempts to make me an object in my turn. In Sartre's account, this battle is the key to all human relationships, and not merely those which might appear conflictual, but also those of sexual desire and even love. Consciousness is engaged in a permanent struggle to maintain its freedom in the face of onslaughts from all sides.

These aspects of Sartre's early philosophy are probably the best known. Less familiar but no less significant are his accounts of the limits within which human freedom operates. The battle of consciousnesses is not disembodied, and my own body constitutes not only the condition of possibility but also one of the major constraints on my freedom. Consciousness and imagination are free, but they are free against a background of facticity and situation. Facticity in particular is rarely given due weight by exegetes of Sartre's philosophy. My facticity is all the facts about myself which cannot be changed - my age, sex, height, class of origin, race, nationality, for example. (Later Sartre comes to include in facticity more psychological elements of genetic or environmental origin.) One's situation may be modified, but it still constitutes the starting point for any change, and roots consciousness firmly in the world about it. All this means that the Sartrean philosophy of freedom is less idealized than it might at first appear. I am not free to change a whole multiplicity of aspects of my condition, and those I am free to change may not prove easy. As I live I create a self which does not bind me but which certainly makes some courses of action easier and more attractive than others. My own self-image and the image others hold of me also condition the range of possibilities open to me. I make a character for myself over the years, and though it is always open to me to act 'out of character' - after all it is a self I have constituted, not an essence I was born with - such a decision is not usually easy. Sartre describes this self-constitution in terms not so much of character as of 'project', each person having a fundamental project of being, which is not necessarily the result of a conscious decision, and possibly elaborated gradually over time. This project forms the core of a whole nexus of choices and behavioural decisions which form the totality that constitutes my self. My actions form a meaningful whole, each act relates to others before and since, and so the decision to make significant changes always comes up against resistance from already existent patterns and structures. Discussing, for example, an episode when a man gives up on a long hike declaring he is 'too tired' to continue, Sartre discusses the abandonment of the walk in terms of a project which does not put persistence in the face of setbacks at much of a premium. He 'could have acted differently, of course,' Sartre comments, 'but at what cost?' (1943a: 531). Our personal history does not eradicate our freedom, but in practice it is often easier to deny our freedom than to employ it. We hide behind the selves we have constructed, fearing change and convincing ourselves that our choices are limited. Freedom is threatening to us, it opens up a range of possibilities which we find daunting, and we flee from it in what Sartre calls 'bad faith'. Ideally we would like the positive aspect of liberty - free choice, a lack of constraints - together with the security and comfort of a fixed character or nature. The two are incompatible, and our desire to combine them is termed by Sartre a 'useless passion' (see Self-deception, ethics of §2).

In 1943, then, Sartre already sets freedom firmly against a background of constraint - constraints which arise from the features of the material world, from other people whose projects may not coincide with mine, from bodily existence, from facticity and from fear of freedom itself. Freedom is always within and starting from situation, and

it is on the determinants and conditioning power of situation that Sartre increasingly focuses in his later writings.

4 Literary works

The 1940s were the period of Sartre's most prolific literary production. From *La Nausée (Nausea)* (1938) which explores the relationship of contingency and necessity in life and art through the experiences of Roquentin, Sartre moves on in the war years to a contemporary trilogy *Les Chemins de la liberté (The Roads to Freedom)* (1945-9). The trilogy (or unfinished quadrology?) portrays the lives of a varied group of Parisian intellectuals at the outbreak of war, and in particular the ways in which they hide their freedom from themselves while convincing themselves that it is their ultimate goal. Mathieu, a university academic, is the main focus for such ambivalence as he tries to find money for an abortion for his long-term mistress Marcelle.

Sartre also wrote several very successful plays in this period - *Les Mouches (The Flies)* (1943b), a wartime allegory of resistance to German occupation, which uses the Orestean myth to explore the power of human liberty in the face of oppression. *Huis Clos (In Camera)* (1944) shows the deadly consequences of conflictual human relations and self-deception in a hell comprising three characters doomed to remain together for ever in a Second Empire drawing room. *Les Mains Sales (Dirty Hands, or Crime Passionel)* (1948) debates the issues of realism and idealism, means and ends, truth, lies and political commitment in Illyria, an imaginary Communist country in Eastern Europe. This finely balanced and complex play received an unexpectedly positive response from the bourgeois press who interpreted it, against Sartre's intentions, as predominantly anti-communist. In consequence Sartre felt obliged to ban its production for about ten years.

5 Later philosophy

The increasing politicization of Sartre's postwar writing meant that he left both literature and philosophy to one side in the 1950s as he became increasingly engaged as a writer, lecturer and public figure in concrete political issues and endeavours. His next major philosophical work, the *Critique de la raison dialectique (Critique of Dialectical Reason)*, did not appear until 1960 and is clearly marked by his increasing intellectual engagement with Marx. The *Critique* is an attempt to do the impossible: to reconcile existentialism and Marxism; to revivify Marxism, which Sartre believed was becoming sclerotic, by reawakening its awareness of individual and collective subjectivity; and to bring existentialism into closer contact with the material conditions of historical existence. Sartre examines social and political issues such as group action, historical change, revolution and behaviour in the face of material scarcity of resources. He modifies his radical position on the extent of human freedom by recognizing more fully than before the effect of historical and material conditions on individual and collective choice. He takes as his own the famous slogan of Engels: 'Men make history on the basis of what history has made them.' We are not pawns or cogs in a machine, nor do we simply participate in processes of internalization and externalization: we are free agents, but agents who are profoundly and inescapably situated in specific social and material conditions. Indeed Sartre later uses the (Jansenist) term 'predestination' to explain how his views differ from positivist theories of human determinism. Material conditions set up the environment in which we operate. They do not causally determine our behaviour, but they do prescribe the (limited) range of options open to us. A white bourgeois male in a prosperous suburb has a vastly wider range of choices on which to exercise his freedom than an elderly black women living in the poverty of an inner city ghetto. Both are free in the ontological sense, but their possibilities for making use of that freedom are not comparable. And in 1960 Sartre is as concerned with the restrictions imposed on freedom by the material world as with human liberty itself.

It is this preoccupation with the absolute and yet circumscribed nature of human freedom that underpins Sartre's two last major works: his autobiography, *Les Mots (Words)* (1963), a brief and finely wrought literary masterpiece, and *L'Idiot de la famille (The Idiot of the Family)* (1971-2), a 3,000-page biography of Flaubert which draws on a vast range of different disciplines. 'What can one know of a man, today?' was the question Sartre set out to answer in his account of Flaubert, and in it he synthesizes not only existentialism, phenomenology and Marxist theory and method, but also psychoanalysis, sociology, history of literature, aesthetics and anthropology. What did Flaubert make of what was made of him? Educated in a family embodying the historical conflicts of its age, second son of a doctor and expected to become a lawyer, the young Gustave Flaubert constructed a very different career for himself. Resistant to adult pressures to perform, he learned to read late (hence the *Idiot* of the title), lived in his elder brother's shadow and opted out of law school through a hysterico-epileptic crisis ('intentional' but not 'deliberate', in Sartre's terms) which made him an invalid - the 'hermit of Croisset' - and thus permitted him to

live in the family home and become a writer. Sartre's account of his own choice of the same career is more succinct and more ironic: the Sartre and Schweitzer (maternal grandfather) families are not spared in the biting and witty descriptions of the 'family comedy' which made of young Jean-Paul a precocious charlatan, writing to please adults, writing for future fame - a superman author - and finally writing as a professional. The gap between choice and destiny is shown to be very small, but it has not closed. Even when analysing with cruel perspicacity his own formation, Sartre maintains the framework he set up thirty years earlier: freedom within situation, even when the situation may leave little room for manoeuvre. Subjectivity is now defined as the *décalage* or difference between the processes of internalization and externalization; liberty may be no more than the 'play' in the mechanism, but the permanent dialectic between the poles of freedom and conditioning remains untotalized.

See also: Beauvoir, S. de; Camus, A.; Existentialism; Existentialist ethics; Merleau-Ponty, M.; Phenomenological movement

CHRISTINA HOWELLS

List of works

Sartre, J.-P. (1936a) *L'Imagination*, Paris: Alcan; trans. F. Williams, *Imagination, a Psychological Critique*, Ann Arbor, MI: University of Michigan Press, 1962.(A history of theories of the imagination leading up to Husserl.)

Sartre, J.-P. (1936b) 'La Transcendance de l'ego, Esquisse d'une description phénoménologique', *Recherches Philosophiques* 6; repr. in *La Transcendance de l'ego, Esquisse d'une description phénoménologique*, ed. S. le Bon, Paris: Vrin, 1965; trans. F. Williams and R. Kirkpatrick, *The Transcendence of the Ego. An Existentialist Theory of Consciousness*, New York: Noonday, 1962.(A phenomenological account of the ego.)

Sartre, J.-P. (1938) *La Nausée*, Paris: Gallimard; trans. L. Alexander, *Nausea, or The Diary of Antoine Roquentin*, New York: New Directions, 1949; trans. R. Baldick, *Nausea, or The Diary of Antoine Roquentin*, Middlesex: Penguin, 1965.(Novel in diary form about the discovery by Antoine Roquentin of the contingency of existence.)

Sartre, J.-P. (1939) *Esquisse d'une théorie des émotions*, Paris: Hermann; trans. B. Frechtman, *The Emotions: Outline of a Theory*, New York: Philosophical Library, 1948; trans. P. Mairet, *Sketch for a Theory of the Emotions*, London: Methuen, 1962. (Study of the psychology of the emotions.)

Sartre, J.-P. (1940) *L'Imaginaire, psychologie phénoménologique de l'imagination*, Paris: Gallimard; trans. B. Frechtman, *The Psychology of the Imagination*, New York: Philosophical Library, 1948.(A phenomenological study of imagination.)

Sartre, J.-P. (1943a) *L'Être et le Néant. Essai d'ontologie phénoménologique*, Paris: Gallimard; trans. H.E. Barnes, *Being and Nothingness: An Essay of Phenomenological Ontology*, New York: Philosophical Library, 1956; London: Methuen, 1957.(Sartre's major philosophical work: a study of the relationship between consciousness and the world, and between consciousness and other consciousnesses.)

Sartre, J.-P. (1943b) *Les Mouches*, Paris: Gallimard; repr. Paris: Livres de Poche, 1971; trans. S. Gilbert, *The Flies*, in *No Exit and Three Other Plays*, New York: Vintage Books, 1949.(Resistance play based on the Greek myth of Orestes.)

Sartre, J.-P. (1945) *Huis Clos*, Paris: Gallimard; repr. Paris: Livres de Poche, 1971; trans. S. Gilbert, *In Camera*, in *No Exit and Three Other Plays*, New York: Vintage Books, 1949.(Drama of existence in which three people are trapped together for eternity.)

Sartre, J.-P. (1945-9) *Les Chemins de la liberté (The Roads to Freedom)*, vol. 1, *L'Âge de raison*, Paris: Gallimard, 1945; trans. E. Sutton, *The Age of Reason*, New York: Knopf, 1947; vol. 2, *Le Sursis*, Paris: Gallimard, 1945; trans. E. Sutton, *The Reprieve*, New York: Knopf, 1947; vol. 3, *La Mort dans l'âme*, Paris: Gallimard, 1949; trans. G. Hopkins, *Troubled Sleep*, New York: Vintage Books, 1951.(Trilogy of novels set in Paris of the early 1940s.)

Sartre, J.-P. (1946) *L'Existentialism est un humanisme*, Paris: Nagel; trans. B. Frechtman, *Existentialism*, New York: Philosophical Library, 1947, and Citadel, 1957.(A lecture purporting to present existentialist philosophy as a humanism - later repudiated by Sartre as over-simple.)

Sartre, J.-P. (1948) *Les Mains Sales*, Paris: Gallimard; trans. L. Abel, *Dirty Hands*, or *Crime Passionel*, in *No Exit and Three Other Plays*, New York: Vintage Books, 1949. (Political play opposing realism and idealism.)

Sartre, J.-P. (1960) *Critique de la raison dialectique, précedé de Questions de methode, I, Théorie des ensembles pratiques*, Paris: Gallimard; repr. in new annotated edn, 1985; first essay trans. H.E. Barnes, *Search for a Method*, New York: Knopf, 1963; main text trans. A. Sheridan-Smith and ed. J. Rée, *Critique of Dialectical Reason*, London: New Left Books, 1976, and Atlantic Highlands, NJ: Humanities Press, 1976.(A lengthy

attempt to reconcile existentialism and Marxism within a philosophy of history.)
Sartre, J.-P. (1963) *Les Mots*, Paris: Gallimard; trans. B. Frechtman, *The Words*, New York: Braziller, 1964; trans. I. Clephane, *Words*, London: Hamish Hamilton, 1964.(Sartre's (ironic) account of his childhood.)
Sartre, J.-P. (1971-2) *L'Idiot de la famille, G. Flaubert de 1821 à 1857*, Paris: Gallimard; trans. C. Cosman, *The Idiot of the Family*, Chicago, IL: University of Chicago Press, 3 vols, 1981, 1987, 1989.(A three-volume existential biography of Flaubert, intended to answer the question, 'What can we know of a man today?')
Sartre, J.-P. (1983a) *Les Carnets de la drôle de guerre*, Paris: Gallimard; trans. Q. Hoare, *The War Diaries of Jean-Paul Sartre*, New York: Pantheon Books, 1984.(Sartre's diaries at the onset of the Second World War.)
Sartre, J.-P. (1983b) *Cahiers pour une morale (Notebooks for an Ethic)*, Paris: Gallimard. (Notebooks attempting a sketch for an ethics with which Sartre was never fully satisfied.)
Sartre, J.-P. (1985) *Critique de la raison dialectique, tome II (inachevé), L'Intelligibilité de l'Histoire (Critique of Dialectical Reason, vol. 2 (incomplete), The Intelligibility of History)*, ed. A. Elkaim-Sartre, Paris: Gallimard. (Volume 2 of the *Critique*, focusing in particular on the question of the intelligibility of history.)
Contat, M. and Rybalka, M. (1970) *Les Ecrits de Sartre*, Paris: Gallimard; trans. *The Writings of Jean-Paul Sartre*, Evanston, IL: Northwestern University Press, 1973.(A full bibliography of Sartre's works up to 1969. Later supplements are given in the English translation, and in *Obliques* 18-19 (1979), ed. M. Sicard.)

References and further reading

Aronson, R. (1980) *Jean-Paul Sartre: Philosophy in the World*, London: New Left Books.(A lively critical account.)
Caws, P. (1979) *Sartre*, London: Routledge & Kegan Paul.(A clear and comprehensive study.)
Chiodi, P. (1976) *Sartre and Marxism*, trans. K. Soper, Hassocks, Sussex: Harvester, and New York: Humanities Press.(A lucid critical discussion of Sartre's *Critique*, from a Marxist perspective.)
Cohen-Solal, A. (1987) *Sartre: A Life*, New York: Pantheon, and London: Heinemann.(Much the best biography of Sartre so far; well-informed and not uncritical.)
Danto, A. (1975) *Sartre*, Modern Masters, London: Fontana.(Short and entertaining account.)
Fell, J. (1965) *Emotion in the Thought of Sartre*, New York: Columbia University Press.(The best study of Sartre's early phenomenological psychology.)
Fell, J. (1979) *Heidegger and Sartre*, New York: Columbia University Press.(A well-informed comparison of Sartre and Heidegger.)
Howells, C. (1988) *Sartre: The Necessity of Freedom*, Cambridge: Cambridge University Press.(A study of both literature and philosophy.)
Howells, C. (ed.) (1992) *The Cambridge Companion to Sartre*, Cambridge: Cambridge University Press.(A stimulating collection of essays by American and European specialists.)
Jeanson, F. (1947) *Le Problème moral et la pensée de Sartre*, Paris: Seuil; trans. R. Stone, *Sartre and the Problem of Morality*, Bloomington, IN: Indiana University Press, 1980.(An early discussion by a close friend of Sartre, especially good on Sartre's early ethical theory.)
McCulloch, G. (1994) *Using Sartre*, London: Routledge.(An attempt to connect Sartre's early philosophy with themes from analytical philosophy.)
Merleau-Ponty, M. (1945) *Phénoménologie de la Perception*, Paris: Gallimard; trans. C. Smith, *Phenomenology of Perception*, London: Routledge, 1962.(A major work of philosophy in its own right, part 3 contains penetrating criticisms of *Being and Nothingness*.)
Morris, P. (1976) *Sartre's Concept of a Person*, Amherst, MA: University of Massachusetts Press.(An account of Sartre from an analytic perspective.)
Olafson, F. (1967) *Principles and Persons*, Baltimore, MD: Johns Hopkins University Press.(A well-informed critical discussion of Sartre's ethical theory.)
Olafson, F. (1965) *The Philosophy of Sartre*, London: Hutchinson, and New York: Hillary House.(A classic, if now out-of-date, account.)
Warnock, M. (ed.) (1971) *Sartre: A Collection of Critical Essays*, New York: Anchor.(An early collection of essays, many of them still well worth reading.)
Wilcocks, R. (ed.) (1988) *Critical Essays on Jean-Paul Sartre*, Boston: G.K. Hall. (An excellent collection of essays.)

Saussure, Ferdinand de (1857-1913)

*Though he made a major contribution to the comparative and historical studies which dominated nineteenth-century linguistics, Saussure is best known today for the development of a radically different conception of language and of the methodology of linguistics which became central to twentieth-century structural linguistics. According to this conception a language is a system of signs which are radically arbitrary, so that their significations are determined only by the historically constituted systems of conventions to which they belong - such a system Saussure called '*la langue*'. It follows, therefore, that a linguistic study is first and foremost one of* la langue*, that is, of the conventional relations obtaining at a given time between signs belonging to the same system, rather than one of the development of linguistic forms over time, as the comparativists had maintained.*

1 The primary object of linguistic study

In the nineteenth century linguistics was dominated by comparative and historical studies the primary aims of which were to compare forms in different languages and to trace their development. The Swiss linguist Ferdinand de Saussure made a major contribution to these studies at the age of twenty-one in his *Memoires sur le système primitif des voyelles dans les langues indo-européenes (Memoir on the Primitive System of Vowels in Indo-European Languages)* (1878). However, he became increasingly dissatisfied with the assumptions made by these studies. In particular, they failed to ask fundamental questions: 'What is the primary object of a linguistic study?' and 'What is the nature of the object studied?' Saussure proposed novel answers to these questions in three lecture courses given in Geneva between 1906 and 1911. Unfortunately, he died in 1913 before he was able to publish his conclusions; the text by which he is best known, *Cours de linguistique générale (Course in General Linguistics)*, was reconstructed from his lecture notes by two of his pupils and published in 1916.

Saussure's argument about the nature of the object of linguistics is a complex one, in the course of which he develops a conception of a language as a social institution which is, at the same time, radically arbitrary. The starting point of the argument is a claim that for each language one can make a distinction between *la langue* (the language itself, for example, French or German) and *le parole* (its uses, the primary one for Saussure being speech). *La langue* is a system of conventions the knowledge of which enables speakers to communicate. It both logically precedes and is more abstract than *le parole*, since speech sounds, while being the primary means of its articulation, are nevertheless only one of its instruments. The existence of other instruments, such as writing, or the signs of a language for the deaf, show that what is natural to man is not 'oral speech but the faculty of constructing a language' ([1916] 1977: 26). Thus, to concentrate attention on speech sounds, or indeed on any other instrument of articulation, would be to neglect the underlying system (*la langue*) which makes articulation possible and is, therefore, both more fundamental and more abstract than any of its instruments of expression; in short, it is 'a form not a substance' (1977: 113). Since *la langue* is a system of conventions, it is inherently social, but it has no architect and so is not an intentional product. On the other hand, aspects of *le parole* - the uses of that system - such as the production of speech sounds, written texts or speech acts and so on, are individual acts or products which are typically intentional. Nevertheless, they are logically parasitic on the existence of an underlying system; for instance, without a conventional way of naming days of the week (a fact about *la langue*), one could not assert that today is Tuesday by uttering the words 'It is Tuesday'. So, Saussure argues, the fundamental study is one of *la langue*: '*from the very outset we must put both feet on the ground of language (*la langue*) and use language as the norm of all other manifestations of speech*' (1977: 9; original emphasis).

To appreciate the importance of Saussure's claim, and the distance which he had moved from the position of historical linguistics, it is necessary to consider at this point the other major distinction he introduced; that between 'synchronic' and 'diachronic' linguistics. One way in which he presents this distinction seems relatively unproblematic, involving no more than a difference, created by the inquirer's point of view, between studying something at a moment in time (a synchronic study) and studying it as it changes over time (a diachronic study). But his account of the distinction is misleading if it suggests that what is studied from the two perspectives is one and the same thing. On the contrary, Saussure maintains that one can study *la langue* only from the first perspective. By contrast, diachronic linguistics is the study of changing relations between individual items which, though they may impact on *la langue*, are not designed or intended by anyone to do so. They do not, therefore, form a system, so that 'there is no such thing as ''historical grammar''' (1977: 134). Moreover, Saussure argues,

before one can compare items belonging to different stages of a language one first needs to establish what the respective systems of conventions are or were, so that a synchronic study logically precedes a diachronic one. For instance, it is a synchronic fact that a number of English nouns mark the plural with a vowel change: ‘tooth’/‘teeth’, ‘goose’/‘geese’ and so on. This is a fact about the conventions of contemporary English, which is psychologically real for native speakers. On the other hand, it is a diachronic fact that this way of marking the plural was made possible by a series of purely phonetic changes which occurred previously; changes which were in no way designed or intended to make possible the modern way of forming a plural in the cases in question. So if we want to understand how plural forms are generated within a language, we can do so only by studying the language synchronically: ‘the linguist can neither describe it nor draw up standards of usage except by concentrating on one state’ (1977: 82). Hence Saussure’s conclusion that the fundamental linguistic study is a synchronic one of *la langue* - a root-and-branch rejection of the assumptions of nineteenth-century historical linguistics.

2 Language as a system of signs

Central to the argument underpinning this conclusion is Saussure’s theory of signs. For him each sign has two aspects, one acoustic, the other conceptual, and he designates them by the terms ‘signifier’ and ‘signified’ respectively. For example, an acoustic impression of an utterance of ‘cat’ is the word’s signifier, and, *as a first approximation*, its signified is the concept *cat* (see comments on values below). It is a fundamental principle that signs are arbitrary: ‘the signified ‘‘ox’’ has its signifier *b-ö-f* [*boeuf*] on one side of the border and *o-k-s* (*Ochs*) on the other’ (1977: 67). Saussure’s principle implies, of course, that the connection between a signifier and its signified is contingent; but it asserts more than that, namely that there are no extra-linguistic facts which determine either the existence of, or the relation between, a signifier and a signified. For instance, it is not because of facts about the structure of our eyes that we have the colour vocabulary that we have, nor because of facts about the structure of thought that we employ the linguistic categories we do. Thought without language is amorphous, so that ‘nothing is distinct before the appearance of language’ (1977: 112). The development of language does not involve, as some think, a naming ceremony at which signifiers are invented to correspond to pre-existing signifieds. Rather, the differentiation of distinct signifiers and signifieds has to proceed together.

However, Saussure argues, the differentiation has to proceed systematically: ‘the principle of the arbitrariness of the sign would lead to the worst sort of complication if applied without restriction’ (1977: 133). Numerical signs illustrate this point: ‘twenty-one’, ‘thirty-one’ and so on are constructed in a systematic way, and if this were not so it would not be possible to learn them all. Moreover, Saussure maintains that it is because a sign has a unique position in a system that we can identify occurrences of it. What makes two spoken occurrences of ‘Gentlemen!’ the same is not identity of sound or content, but the fact that each differs in the same way from instances of other English words - ‘in language there are only differences *without positive terms*’ (1977: 120; original emphasis). These differences depend on two kinds of relations that signs can enter into, called ‘syntagmatic’ and ‘associative’ respectively. The first kind of relation holds between a sign and the elements which can precede or succeed it; for example, there is a syntagmatic relation between ‘talk’ and ‘ing’ in ‘John is talking.’ The second relation holds between items which by virtue of a similarity in form or meaning can be substituted for each other in a range of contexts. For instance, ‘talking’ and ‘walking’ are associatively related because of similarities in form, while ‘talking’ and ‘speaking’ are related because of similarities in meaning. And since the *only* way to identify a sign is by identifying the syntagmatic and associative relations it has to other signs - so we know what it is not - signs present themselves as elements of a system and cannot be identified independently of it.

Hence, Saussure argues, signifieds cannot be identified, as we did as a first approximation, with concepts, for that would suggest that they have a content independent of the linguistic system to which they belong. Instead, they should be thought of as values which are wholly dependent on the system to which they belong. Colour words provide an illustration; for example, the English word ‘brown’ has no exact equivalent in French, since it can be translated as ‘brun’, ‘marron’ or even ‘jaune’ depending on context (Lyons 1968: 56). This makes it clear why a synchronic study of *la langue* is the primary study, since a study of values is a study of the system to which they belong.

3 Semiology and structuralism

The contention that linguistic units can only be identified as terms of a system is the central tenet of European

structural linguistics (see Structuralism in linguistics), a movement for which Saussure's work provided the foundations. However, not least because they were work in progress, his theories leave much scope for interpretation and for further development; for instance, there is no serious theory of *le parole* in *Course in General Linguistics*, though one was promised, and there is no *formal* account of the relations which signs can enter into, something which is particularly serious in the case of associative relations given the vagueness of the idea of a similarity of form or meaning. So there was much work still to be done by linguists both to develop and to explore the limitations of Saussure's ideas.

As well as its importance for the development of European structural linguistics, Saussure's work also had a much wider influence. The central importance which he gave to the theory of the linguistic sign led him to speculate whether that theory were not in fact an instance of a more general study of signs of any kind that express ideas: '*a science that studies the life of signs within society* is conceivable; it would be a part of social psychology and consequently of general psychology; I shall call it *semiology* (from the Greek *sēmeîon* ''sign'')' (1977: 16; original emphasis). This claim was programmatic, but, albeit belatedly, it undoubtedly affected and enriched the thought of many who tried to adapt the methodology of structural linguistics as a model for that of the study of cultural signs in general, including such thinkers as Merleau-Ponty, Lévi-Strauss, Barthes and Lacan. So the impact of Saussure's work has to be measured not only in terms of its effects on the development of linguistics, especially European structural linguistics, but, more widely, in terms of its impact on the development of a methodology for the study of the social sciences in general.

See also: Semiotics; Structuralism in social science; Structuralism in literary theory

DAVID HOLDCROFT

List of works

Saussure, F. de (1878) *Memoires sur le système primitif des voyelles dans les langues indo-européenes* (Memoir on the Primitive System of Vowels in Indo-European Languages), Leipzig: Teubner.(Saussure's contribution to comparative and historical linguistic studies; not available in English.)

Saussure, F. de (1916) *Cours de linguistique générale*, ed. C. Bally and A. Sechehaye, with the collaboration of A. Riedlinger, Lausanne and Paris: Payot; trans. W. Baskin, *Course in General Linguistics*, Glasgow: Fontana/Collins, 1977.(The key source of Saussure's ideas.)

Saussure, F. de (1993) *Saussure's Third Course of Lectures in General Linguistics (1910-1911): From the Notebooks of Emile Constantin*, Language and Communication series, vol. 12, trans. and ed. E. Komatsu and R. Harris, Oxford: Pergamon.(This account of Saussure's final lecture course was compiled from notes not available to the editors of the original 1916 publication. It contains new material and gives a more advanced picture of Saussure's final ideas.)

References and further reading

Culler, J. (1976) *Saussure*, Glasgow: Fontana/Collins.(A good general introduction, with an excellent account of Saussure's impact on structuralist thought.)

Ducrot, O. and Todorov, T. (1981) *Encyclopedic Dictionary of the Sciences of Language*, trans. C. Porter, Oxford: Blackwell.(As well as a sympathetic essay on Saussure, there are very helpful separate essays on all the main Saussurean themes.)

Harris, R. (1987) *Reading Saussure*, London: Duckworth.(A good recent detailed critique by a linguist.)

Holdcroft, D. (1991) *Saussure: Signs, System, and Arbitrariness*, Cambridge: Cambridge University Press.(A lengthier exposition of the themes of this essay.)

Lyons, J. (1968) *An Introduction to Theoretical Linguistics*, Cambridge: Cambridge University Press.(Includes a sympathetic discussion of the main Saussurean theses.)

Savigny, Friedrich Karl von (1779-1861)

*Friedrich Karl von Savigny was a powerfully influential student of Roman law both in its medieval manifestations and in the contemporary ‘*Pandektenrecht*’ (law based on Justinian’s Pandects, or Digest) of nineteenth-century Germany. His contributions to the philosophy of law are in the spirit of the Romantic movement, and lay stress on the organic character of the legal experience of a people, hence favouring customary law over statute law, and opposing the contemporary movement towards codification. A founder of what is sometimes called the ‘historical school’ in the philosophy of law, he argues that law is to be understood always in its historical setting, the result of a process of historical development, not simply as the arbitrary command of a - perhaps transitory - sovereign power.*

Savigny was a nobleman whose family roots were in Lorraine; his decision to become an academic was unusual for one of his class. His doctoral thesis of 1803, on ‘Possession’, made his reputation; analysing the civilian texts, he advanced the thesis that possession in law requires both physical control and an appropriate intention, namely the intention to hold the object in question for oneself. In 1810, he was called to his chair in Berlin, and taught there until 1842. Appointed then to be Prussian Minister for Legislation, he served until 1848. His two major scholarly works, the first on the *The History of the Roman Law during the Middle Ages*, the second on the *System of Modern Roman Law*, were published in the years 1815-21 and 1840-49 respectively. Especially in the analysis of personality in law, and in relation to contractual obligations, he had influence beyond the German-speaking world; at a time of revival in academic legal studies in England and America, his works had considerable influence in the common-law world.

His historicist view of law was most famously expressed in a pamphlet of 1814, *Zum Beruf unsrer Zeit für Gesetzgebung und Rechtswissenschaft (On the Vocation of our Age for Legislation and Jurisprudence)*. This was a polemic against a proposal by Thibaut for codification of law in the Germanic states along the lines of the already-celebrated Code Napoléon; as the German states emerged from their conquest and domination by Napoleonic France, Savigny issued his critique of the positivistic view implicit in the codification project. Law, he argued, emerges from the ‘spirit of the people’, the *Volksgeist*, and develops organically like language. As society becomes more advanced, there has to develop a special profession of lawyers, and it is their customs, including their practices of interpretation of received texts such as the *Pandects*, that express the people’s spirit so far as concerns technical law. As a protest against what F.A. Hayek was later to dub ‘constructivist rationalism’ in law, Savigny’s theory makes an important point; in awakening legal studies to the significance of history (and in founding a legal-historical journal that survives to this day), Savigny made an enduring contribution. His rather mysterious and holistic *Volksgeist* is, however, less convincing, and coloured by an unacceptable ethnic nationalism.

See also: Common law; Jurisprudence, historical §1; Justinian; Law, philosophy of; Roman law

NEIL MacCORMICK

List of works

(1814) *Zum Beruf unsrer Zeit für Gesetzgebung und Rechtswissenschaft*, trans. A. Hayward, *On the Vocation of our Age for Legislation and Jurisprudence*, Birmingham, AL: Legal Classics Library, 1986.(Savigny’s most celebrated single contribution to legal theory, where he developed his theory of the customary and popular or national character of law, in opposition to Thibaut’s proposals for codifying German law.)

(1815-21) *Geschichte des römischen Rechts in Mittelalter*, trans. E. Cathcart, *The History of the Roman Law during the Middle Ages*, repr. West Port, CT: Hyperion Press, 1979.(Reprint provides an accessible modern edition.)

(1840-49) *System des heutigen römischen Rechts (System of Modern Roman Law)*, Berlin: O.L. Heuser.(The full text of Savigny’s most enduringly influential work, in which the theory of the *Volksgeist* is stated.)

References and further reading

Cohen, M.R. and Cohen, F.S. (1952) *Readings in Jurisprudence and Legal Philosophy*, New York: Prentice Hall. (A student’s reader, containing at pp. 386-93 some well-chosen excerpts from translations of Savigny, set in a

helpful context of related readings.)
Guthrie, W. (1880) *Private International Law, and the Retrospective Operation of Statutes*, 2 vols, Edinburgh: T. & T. Clark.(A treatise on the conflict of laws and the limits on their operation in respect of space and time, translating and working from one of Savigny's most influential contributions.)
Paton, G.W. and Derham, D.P. (1972) *Jurisprudence*, 4th edn, Oxford: Clarendon Press.(A useful textbook of jurisprudence, containing at pp. 19-21 a discussion of Savigny's stance and influence in legal theory.)
Perry, Sir E. (1848) *Treatise on Possession, or the Jus Possessionis of the Civil Law*, 6th edn, London: Sweet. (Savigny's elaborate discussion of the physical and mental elements of possession considered as a right set the tone for analytical jurisprudence in this domain for a century to follow.)
Rattigan, W.H. (1884) *Jural Relations, or the Roman Law of Persons as Subjects of Jural Relations*, London: Sweet.(Translation of book 2 of Savigny's *System of Modern Roman Law*; this gives Savigny's view on the nature of personhood in law.)
Stone, J. (1966) *Social Dimensions of Law and Justice*, London: Stevens and Sons.(See pp. 86-101, for a sympathetic and revealing discussion of Savigny's historicist legal theory, set in the context of contemporary debate and also against the background of a discussion of the development of sociological approaches to law.)

Scandinavia, philosophy in

The three countries of Scandinavia - Sweden, Denmark and Norway - share much of their history and culture with Finland and Iceland, and it is natural to treat all five Nordic countries together in any philosophical survey. The first universities in this region were founded more than 500 years ago, in Sweden at Uppsala in 1477, and in Copenhagen, the Danish capital, two years later. Over the years, the main trends of philosophical thought, from Descartes and Locke to Hegelianism, existentialism and logical positivism have all impinged upon philosophy in these countries. A unique feature of philosophy in Norway and Iceland, and until 1971 also in Denmark, is that all university students, including students in law, medicine and dentistry, spend all or most of their first semester preparing for a compulsory exam in philosophy which comprises some philosophy of science and philosophy of language, and some history of philosophy and history of science. This requirement has meant much for recruiting and employment opportunities for philosophers. Thus, for example, the University of Oslo has sixty-five tenured philosophers, while Denmark has suffered a dramatic reduction in the number of philosophical positions since the requirement was abolished. In Sweden philosophy is a compulsory subject in some branches of study in secondary schools.

That philosophers from the Nordic countries have gained a reputation for broad interests and familiarity with several philosophical traditions may largely be due to two factors: small countries *increase the likelihood that they will get involved in popularization and public affairs, and* small language communities *induce the learning of other languages, notably English, German and French, which makes developments in other countries more accessible.*

1 From the Middle Ages to *c.*1870

Philosophy in the Nordic countries began in the eleventh and twelfth centuries with the introduction of Christianity and the establishment of monasteries. Many went abroad to study at the leading universities, and some made notable contributions, such as the Swedish Averroist Boëthius of Dacia who, together with Siger of Brabant, taught at the University of Paris until they were both condemned in 1277. Boëthius ended his days in the 1290s as canon in Linköping.

When the first universities were set up in the late fifteenthth century, Aristotelianism continued to be the main philosophical pursuit. The University of Copenhagen, founded in 1479, almost perished during the Reformation and was re-established in 1537 as a Lutheran institution. The University of Uppsala was suppressed from 1515 to 1593. After the Reformation, a short period of interest in Ramism was followed by a rise in Cartesianism, due in no small way to the fact that in 1649 Descartes came to Stockholm as Queen Christina’s personal tutor. A widespread interest in natural law culminated with Samuel Pufendorf leaving his chair in Heidelberg in order to serve as Professor of Natural Law at the University of Lund during its first nine years. He thereafter went to Stockholm where he spent ten years writing a thirty-three-volume history of Sweden.

The mid-1700s were marked by the influence of Wolff, especially in Sweden and Finland. Subsequently, interest in Locke became prevalent in Finland and Sweden due to the work of Henrik Hassel (1700-76) and Henrik Gabriel Porthan (1739-1804) in Turku and Pehr Niclas Christiernin (1725-99) in Uppsala. Kant was introduced into Danish discussion by Børge Riisbrigh (1731-1809), and Kant’s ethics and philosophy of law had a strong influence on the lawyer and politician Anders Sandøe Ørsted (1778-1860) and through him on Danish law. In Sweden, Kant was introduced by Daniel Boëthius (1751-1810), while in Finland the poet and professor Franz Mikael Franzén (1772-1847) and others developed various versions of Kantianism. Kant had little impact in Norway, mainly as a consequence of the criticism of Niels Treschow (1751-1833), the first professor of philosophy at the new university of Oslo (founded 1813). Treschow, who developed a Spinozistic conception of evolution, was born in Norway and held a chair in Copenhagen from 1803 until taking up the appointment in Oslo. He left the university after just one year to go into government.

Schelling had some influence in Denmark, mostly due to Henrik Steffens (1773-1845). Born in Norway, Steffens studied in Copenhagen and went on to hold professorships in Halle, Breslau and Berlin. The dominant influence on philosophy in the Nordic countries in the late 1800s was, however, Hegel. In Uppsala, Benjamin Höijer (1767-1812) anticipated in interesting ways the post-Kantian developments of Fichte and Schelling. He is for this

reason often called ‘Sweden’s Fichte’. However, his later development was influenced by Hegel. The Uppsala professor of history Erik Gustaf Geijer (1783-1847), in opposition to Hegel, developed a personalistic philosophy, arguing that self-consciousness is possible only in interaction with other individuals: ‘No you - no I’. In Turku the philosopher and statesman Johan Vilhelm Snellmann (1806-81), who has been honoured as the ‘national philosopher’ of Finland, transformed Hegel’s idea of absolute spirit into a doctrine of national spirit expressed through a nation’s language and literature. Snellman gave philosophy a central public role in Finland, that it has retained since. In Denmark and Norway the influence of Hegel went so far as to become extremely oppressive, and Kierkegaard reacted strongly to Hegelianism and every other ‘system philosophy’. In Norway Hegelianism was introduced by the Danish philosopher and poet Poul Martin Møller (1794-1838), who taught philosophy at the University of Oslo from 1827 to 1831. However, it was through Marcus Jacob Monrad (1816-97), who became professor in Oslo at age 29 and retained the position for more than fifty years, that Hegelianism became more protracted in Norway than in any of the other Nordic countries. Following this period of widespread Hegelianism, philosophy developed in a different directions in each country.

2 Sweden

Christopher Jacob Boström (1797-1866), professor of practical philosophy in Uppsala from 1842 to 1863, developed a kind of personalism which combined elements of Plato’s theory of ideas, Leibniz’s monadology and Berkeley’s idealism (see Plato; Leibniz, G.W.; Berkeley, G.). Boström (who has been called ‘Sweden’s Plato’) exercized a strong influence on Swedish intellectual life in the latter part of the nineteenth century, acquiring a dominance similar to that of Hegel in the other Nordic countries. After Boström’s death, three of the four philosophy chairs in Sweden were held by his students: the chair of theoretical philosophy in Lund was held by a Hegelian.

The Uppsala school. In the first decade of the twentieth century the dominance of Boströminianism was broken by what was to become the most remarkable philosophical movement to come out of Sweden: the Uppsala school of philosophy, which flourished until the Second World War. This was the result of the work of two gifted philosophers, Axel Hägerström and his student and later colleague Adolf Phalén (1884-1931). In 1905 Hägerström started to develop the idea that was to become basic for the Uppsala school, that moral sentences express emotions and contain an imperative element. They may in addition have a descriptive ingredient, but their emotive component is the proper topic of ethics, which is an investigation into the noncognitive nature of moral sentences.

Hägerström expanded this view into philosophy of religion and philosophy of law. He rejected metaphysics and studied the ways in which language deceives us into thinking that there are values and that moral statements are true or false. While there are many similarities between the Uppsala school and later logical empiricism, there are also important differences: the Uppsala school combined moral subjectivism with epistemological objectivism, it rejected the view that the subject has immediate knowledge solely of its own experiences, and placed a strong emphasis on the history of philosophy. These historical studies were particularly important for Phalén. Whereas Hägerström maintained and supported his theses with great intensity, Phalén explored with calm objectivity the various possible ways in which the major philosophers have tried to deal with the contradictions in common-sense notions of time, motion, reality, knowledge and so on. This work started in 1910 with a criticism of subjectivism and continued in his large dissertation on Hegel two years later. Historical analysis is also evident in his works on space and time and other systematic issues, as well as in his lectures (unpublished until 1973-8). Phalén’s main interest, however, was epistemology, where his position resembles that of Husserl in some respects, and in 1911 he was the first philosopher in Scandinavia to write on Husserl’s phenomenology. After the Second World War, philosophy in Sweden became much more varied both in theme and perspective.

History of philosophy. The two most notable Swedish philosophers in the generation that succeeded the Uppsala School were Konrad Marc-Wogau (1902-91), professor of theoretical philosophy in Uppsala from 1946 to 1968, and Anders Wedberg (1913-78), professor of theoretical philosophy in Stockholm from 1949 to 1976. Marc-Wogau published a number of books and essays on Plato, Descartes, Marx, Hägerström and other figures in the history of philosophy and on contemporary Soviet thought. He also wrote a logic text and several studies of systematic issues, including a book on psychoanalysis. Together with his two prewar great books on Kant and his 1945 work *Die Theorie der Sinnesdaten* they establish him as one of Scandinavia’s most notable

philosopher-scholars. Wedberg's few, but high-quality publications, on a broad spectrum of topics, are mostly historical. He had a strong interest in Bolzano, which has been followed up by Jan Berg (1928-), since 1969 in Munich. Wedberg was above all a great systematic philosopher, and his works set a high standard for his students, the foremost of whom were Kanger, Berg and Prawitz.

In Lund, all the major philosophers during that period worked in the history of philosophy. Particularly important was Hans Larsson (1862-1944), professor of theoretical philosophy from 1901 to 1927, who wrote on Kant and Spinoza, and whose work earned him one of the eighteen places in the Swedish Academy. Åke Petzäll (1901-57) introduced logical empiricism in Sweden and founded the journal *Theoria*. His successor in the chair of practical philosophy, Manfred Moritz (1909-90) was Sweden's foremost expert on Kant's ethics after Hägerström. He also further developed Wesley Hohfeld's system of basic legal concepts, as did Kanger later.

Philosophical logic. The most prominent Swedish contributions to philosophy today are within the field of philosophical logic. Sören Halldén (1923-), who held the chair in theoretical philosophy in Lund from 1964 to 1988, produced the first major work in this field in Sweden in his Uppsala dissertation of 1950. His later contributions include one of the first studies of preference logic, several books on decision logic and its bearing on epistemology, and a number of other books and articles within several areas of philosophy. Together with the legal scholar Per Olof Ekelöf (1906-90) in Uppsala and the philosopher of science Martin Edman (1945-) in Umeå, Halldén has been working on the evaluation of evidence, particularly in courts, and developed the important notion of evidentiary mechanisms and the evidentiary value model.

A most significant Swedish contribution to philosophical logic was the dissertation which Stig Kanger (1924-88) wrote in Stockholm for Wedberg in 1957. Kanger here presented the key ideas of what has later come to be called 'Kripke semantics' for modal logic, named after Saul Kripke who published them in 1959 and 1963. Although Hintikka and Kripke discovered this type of semantics independently, it would be appropriate to call it Kanger-Kripke semantics. Kanger also made several further important and original contributions to logic and its applications and to the theory of measurement. He also developed and refined Hohfeld's system of basic legal terms. In 1968 he succeeded Marc-Wogau in the chair in theoretical philosophy in Uppsala.

Dag Prawitz (1936-) also wrote his dissertation with Wedberg and in 1976 succeded Wedberg in the chair of theoretical philosophy in Stockholm, after five years as a professor in Oslo. In his 1965 dissertation he was the first to prove Gentzen's *Hauptsatz* directly for certain systems of natural deduction. Prawitz has extended this result and has also written on causality, utilitarianism and a number of other topics, and has developed a Dummett-inspired theory of meaning. Similar work on meaning has been done by Per Martin-Löf (1942-), research professor in logic in Stockholm. Martin-Löf has also developed an intuitionistic-type theory and has done important work on the foundations of probability theory.

A main contributor to modal logic is Krister Segerberg (1936-), who in 1990 succeeded Kanger in Uppsala. Among his many results are a generalization of Lemmon and Scott's 'filtration method', a way of replacing complicated models with simple ones. Bengt Hansson (1943-), who succeeded Halldén in Lund in 1989, has worked in modal logic, preference logic, philosophy of science and philosophy of language. Among the many other philosophers in Sweden who have worked on logic and its applications, Per Lindström (1936-) in Gothenburg has done particularly important work in model theory. In 1969 he proved what has become known as Lindström's theorem.

Ethics. Among Swedish postwar philosophers, the best known to the general public has been Ingemar Hedenius (1908-82) who, from 1947 to 1974, held the chair once held by Hägerström. Hedenius' 1941 *Om rätt och moral (On law and morality)* exerted a strong influence on legal philosophy and ethics in Sweden, while his 1949 work *Tro och vetande (Faith and knowledge)* played an extraordinary role in the debate on religion in Scandinavia. Hedenius also developed a version of utilitarianism and wrote several articles and books on ethics and the history of ethics (Berkeley, Hume and Plato) as well as on other topics. His successor in Uppsala was Lars Bergström (1935-), who produced his dissertation under the direction of Ofstad and Wedberg in 1966. In 1987 he left Uppsala for Stockholm. Most of Bergström's work has been in ethics, much of it connected with his very precise statement of utilitarianism in his dissertation, but he has also written articles on other areas. In Lund, Göran Hermerén (1938-) has written on *Influence in Art and Literature* (1975) and other topics in aesthetics, is now professor of medical ethics.

Decision theory. Among Hedenius' and Bergström's many gifted students some have turned to decision theory and related fields. Sven Danielsson (1939-), who succeeded Bergström in Uppsala, and Wlodek Rabinowicz (1947-), professor of practical philosophy in Lund, have together with Peter Gärdenfors (1949-), professor of cognitive science in Lund, and Nils-Eric Sahlin (1954-) done important work on decision theory, group preferences and distributive justice. Gärdenfors and Sahlin have also worked on many other topics, including dynamic models of belief and empirical evidence.

Other activities. The fields that have been mentioned are particularly vigorous in Sweden, but there is also activity in many other fields. Gothenburg, in particular, has had two philosophers with highly distinctive profiles. Ivar Segelberg (1914-87), professor of philosophy from 1951 to 1979, wrote several phenomenological studies inspired by Phalén, and has in turn stimulated a third generation of Swedish phenomenolologists. Segelberg's successor, Mats Furberg (1933-) has written extensively on speech-act theory, and has in recent work attempted a reconciliation between speech-act theory and hermeneutics. However, his most distinctive contribution is a number of exceptionally engaging books in Swedish, on topics such as death, the meaning of life and the riddle of the world.

3 Denmark

Harald Höffding (1843-1931) was the dominant figure in Danish philosophy from the 1870s until he retired in 1915 after thirty-two years as professor in Copenhagen. His humanistic naturalism was inspired by, among others, the positivism of Comte and the empiricism of John Stuart Mill and Herbert Spencer. He defended a double-aspect theory of mind and body, and in a somewhat eclectic way, via utilitarianism and religion, he transposed ideas from the Danish idealist tradition into the framework of naturalism. Jørgen Jørgensen (1894-1969), professor in Copenhagen from 1926 to 1964, favoured Russell's logic and the logical positivist movement, and criticized 'metaphysical' assumptions in the sciences. In his later works Jørgensen developed a psychologistic understanding of epistemology and logic, and maintained that nerve processes, behaviour and ideas are different aspects of the same thing. In meta-ethics he defended an emotivist version of noncognitivism and in consequence a radical distinction between values and facts. A similar positivistic view on values was introduced in jurisprudence by Alf Ross, professor of law in Copenhagen from 1938 to 1969. Having studied under Hans Kelsen in Vienna, Ross adopted his tutor's interpretation of propositions in law as predictions of the behaviour of a judge in court.

Partly under Jørgensen's influence, several Danish philosophers have taken up philosophy of science, notably Johannes Witt-Hansen (1908-86), professor of philosophy in Copenhagen between 1959 and 1978, who wrote on Eddington, on matter, on generalization and on historical materialism. David Favrholdt (1931-), who in 1966 became the first professor of philosophy at the University of Odense, wrote his dissertation on Wittgenstein's *Tractatus*, but has been working on Niels Bohr. So has Jan Faye (1947-), who received his doctorate in Odense in 1981 with a dissertation in which he argued that backward causation is possible in principle. A major issue between Favrholdt and Faye is to what extent Harald Höffding had a decisive influence on Bohr's interpretation of quantum mechanics.

In ethics, a most influential Danish contribution has come from Knud Ejler Løgstrup (1905-81), professor of theology at the University of Aarhus. Inspired by the phenomenological philosophy of Hans Lipps, in the 1950s he developed an ethical theory which is widely studied in several countries, notably among theologians. Mogens Blegvad (1917-), professor of philosophy in Copenhagen from 1964 to 1987, has written on the naturalistic fallacy, action explanation and a variety of other topics.

The later Wittgenstein has influenced philosophy in Denmark mainly through Justus Hartnack (1912-), who dominated the philosophical scene in Aarhus during his time there as professor (1954-72). Peter Zinkernagel (1921-) argued that all language-use presupposes the truth of certain language rules embedded in ordinary language. These rules include substantial philosophical claims, for example that there is an external physical world. Thus phenomenalism and other traditional philosophical views which run counter to common sense turn out to be self-refuting.

The abolition of the compulsory philosophy exam in 1971, together with a negative attitude to universities among politicians, has been disastrous for Danish philosophy. In Copenhagen the number of chairs in philosophy was reduced from four to one. However, several young Danish philosophers went abroad to study and many of them

have returned, bringing with them a conception of philosophy as a discipline with fundamental systematic ambitions and introducing the works of Donald Davidson, Michael Dummett and Saul Kripke into Danish philosophy. Among these is Finn Collin (1949-), associate professor in Copenhagen who, after a dissertation on sensations in Berkeley, has written books on interpretive social science and on the construction of social fact. He has also written on causality and explanation, on phenomenology and hermeneutics, and on philosophy of language. His Copenhagen colleague Peter Sandøe (1955-), who studied in Frankfurt and Oxford, has written about moral realism.

History of philosophy. The most notable Danish contributions to philosophy in recent decades have come within the history of philosophy, most of them from Copenhagen, which is a centre for the study of ancient and medieval philosophy. Karsten Friis-Johansen (1930-) has followed his early important Plato studies with work on Democritus and on Kierkegaard, and has also written a monumental first volume of a four-volume history of philosophy. In the Classics Department, Johnny Christensen has written *An Essay on the Unity of Stoic Philosophy* (1962) and has produced several other works. In the theological faculty Troels Engberg-Pedersen (1948-) has written *Aristotle's Theory of Moral Insight* (1983) and *The Stoic Theory of Oikeiosis* (1990).

In medieval philosophy, Jan Pinborg (1937-1982) published a number of important studies on semantics and logic in the Middle Ages. Several of his articles were co-authored with Sten Ebbesen (1946-), whose main work is his three-volume *Commentators and Commentaries on Aristotle's Sophistici Elenchi* (1981). Niels Jørgen Green-Pedersen (1942-) has also written *The Tradition of the Topics in the Middle Ages* (1984).

In Aarhus, at the Institute for the History of Science, Olaf Pedersen (1920-) and his colleagues have produced a number of renowned studies on ancient and medieval history of science, mathematics and technology, among them Pedersen's *Early Physics and Astronomy* (1974) and his *Galileo and the Council of Trent* (1983).

Modern philosophy is studied in Copenhagen by Carl Henrik Koch (1938-), who has written the large third volume (from the Reformation to the Enlightenment) of a four-volume history of philosophy. Koch has also written a survey of eighteenth century philosophy for the *Cambridge History of 18th Century Philosophy*. Knud Haakonssen (1947-), who took his doctorate in Edinburgh on what later became the book *The Science of a Legislator. The Natural Jurisprudence of David Hume and Adam Smith* (1981), is now teaching at Boston University.

4 Finland

Like other countries, Finland experienced a reaction against Hegelianism towards the end of the nineteenth century, inspired in part by Höffding in Denmark. Particularly important was Edward Westermarck (1862-1939), whose work on ethical relativism reflects his combined interest in ethics and sociology. He divided his time between a chair of sociology in London and a professorship of philosophy in Helsinki, and later in Turku (Åbo). Westermarck and the aesthetician Yrjö Hirn (1870-1952) were the first Finns to influence international philosophical discussion.

In the 1930s, Eino Kaila (1890-1958) had an important influence on Finnish philosophy. He participated in the meetings of the Vienna Circle and wrote on the philosophy of psychology and the philosophy of physics (see Vienna Circle). He introduced symbolic logic and logical empiricism as well as Gestalt psychology and empirical psychology to Finland. Ultimately, however, he went more in the direction of philosophy of nature than philosophy of science. His students included Erik Stenius (1911-90), Oiva Ketonen (1913-) and Georg Henrik Von Wright, all of whom did some of their work in logic. Ketonen, who succeeded Kaila at the University of Helsinki in 1948 when Kaila was appointed to the Academy of Finland, has written influential works in Finnish on science policy and university affairs, as well as on general cultural issues. Stenius wrote his dissertation on paradoxes in logic and set theory. He later worked on the Presocratics, epistemology, logic and philosophy of language, but is best known for his incisive interpretation and defence of Wittgenstein's picture theory of language (1960).

Jaakko Hintikka (1929-) is, together with his teacher von Wright, among the leading figures on the international philosophical scene. Like so many Finnish philosophers, Hintikka came to philosophy from mathematics and did his first work in logic, developing further von Wright's idea of distributive normal forms, which continues to play a role in his later work. In 1957 Hintikka - independently of the slightly earlier proposal by Kanger and of the later one by Kripke - presented the basic idea of possible worlds semantics for modal logic. There is not room here to enumerate Hintikka's many important and seminal contributions to modal, epistemic and deontic logic, to model

theory, the interpretation of quantifiers, game theoretic semantics, to philosophy of science, epistemology and philosophy of language and to the history of philosophy from Plato and Aristotle, through the medievals, Descartes, Leibniz, Kant, and Husserl to Wittgenstein.

Hintikka has attracted a large number of talented students who, in the United States and in Finland, work in all the different areas of philosophy in which Hintikka has been working. His Finnish students are main contributors within all the three most active fields of philosophy in Finland today: philosophy of science, philosophy of language and the history of philosophy.

Philosophy of science. Two of Hintikka's earliest students teach at the University of Turku: Risto Hilpinen (1943-) and Juhani Pietarinen (1938-). Both wrote their dissertations on inductive logic. Hilpinen has also written on deontic logic, norms and imperatives and their connection with action theory, and on C.S. Peirce, in particular his theory of 'indeterminate reference'. Several of Hilpinen's latest publications explore a question-theoretic approach to scientific inquiry.

Raimo Tuomela (1940-) in 1970 was appointed to the newly established chair in the methodology of the social sciences in Helsinki, having received his Ph.D. at Stanford the previous year. He has worked mainly on action theory, particularly social action. Together with Ilkka Niiniluoto (1946-), he has written on theoretical concepts and hypothetico-deductive inference. Niiniluoto's later works, on scientific progress and especially truthlikeness, establish him as a leading philosopher of science. Niiniluoto has also contributed to inductive logic. He has published several studies on research policy and the evaluation of research, and has also branched into ethics and law, as in. In 1975 he became the first to hold a chair in the foundations of mathematics in Helsinki.

Jan von Plato (1951-) is working on the foundations of probability, particularly as applied in statistical physics and in dynamical systems. He has also written outstanding historical studies on the foundations of probability.

Philosophy of language. Hintikka's many ideas in the philosophy of language have been followed up by several of his students, in particular game-theoretic semantics, which has been applied to the study of quantifiers, partially ordered connectives, and anaphora. Hintikka-Åqvist's logic of questions has been applied to throw light on explanations.

History of philosophy. A characteristic of Finnish philosophy has been to combine historical and systematic work. Finnish systematic works are rich in historical references, and all the main Finnish historians of philosophy have made contributions to contemporary systematic discussion. Kaila had historical interests, but the combination of high-quality historical scholarship and systematic discussion starts with von Wright, Stenius and Hintikka. Greek philosophy has been studied by all three, especially Stenius and Hintikka. Important contributions have also come from Mårten Ringbom (1934-) and the classicists Rolf Westman (1927-) and Holger Thesleff (1924-). Simo Knuuttila (1946-) in the theological faculty in Helsinki has written on Greek philosophy, but his main work has been on medieval philosophy, to which he is one of the main contributors in Scandinavia. His 1976 dissertation showed that the assumptions concerning modality that Hintikka had attributed to Aristotle, were commonplace throughout most of the Middle Ages. Knuuttila has written on modality in *The Cambridge History of Later Medieval Philosophy* and has also produced numerous articles on subjects ranging from Plato to Descartes.

Finland also has a leading Descartes scholar, Lilli Alanen (1941-). As well as her books and articles on Descartes she is taking up systematic issues in the philosophy of mind and is directing a research group on intentionality. In 1997 she was appointed to a new chair in the history of philosophy at the University of Uppsala in Sweden. Another prominent woman philosopher in Finland is Leila Haaparanta (1954-), who has written a number of books and articles on Frege, Husserl, Peirce and other nineteenth-century thinkers.

Wittgenstein studies are particularly strong in Finland. In addition to the publications of Stenius, von Wright and Hintikka, important contributions have come from some of von Wright's and Stenius' students, notably Lars Hertzberg (1943-) and Heikko Kannisto (1945-). Two other philosophers who have been influenced by von Wright and Stenius are Ingmar Pörn (1935-) and Ghita Holmström-Hintikka (1936-). Pörn has written on action theory and social science, on deontic logic and also on power, Kierkegaard, the emotions and on health, health care and medical ethics, while Holmström-Hintikka has developed a theory of action and will, and applied it to the philosophy of law and in the analysis of medieval discussions, notably in Augustine and Ockham. She has also written on the philosophy of law and on medical ethics.

5 Norway

Monrad's long Hegelian reign was followed by a combination of philosophy and psychology, typical of the time. This combination culminated with Harald Schjelderup (1896-1974) who became professor of philosophy at age 26, but had his chair changed to psychology in 1928. From then until 1954 there remained only one professorship in philosophy in Norway. This was held from 1908 to 1937 by Anathon Aall (1867-1943), who was opposed to Hegelianism and combined empirical psychology, Greek philosophy and the history of Christian thought. Aall was succeeded in 1939 by Arne Næss (1912-), an appointment opposed by many who felt that empirical tendencies were a danger to philosophy. However, fortunately for the openness and the international orientation of Norwegian philosophy, Næss was appointed.

Although it is more than hundred years since philosophy became compulsory for all university students, there have been few teaching positions. In Oslo the situation has improved greatly during the last fifteen years. The philosophy department in Oslo now has sixty-five full-time permanent faculty members spread over most main areas of philosophy. The other universities are following suit. Areas of particular strength at present include logic, philosophy of science and the history of philosophy, notably Greek philosophy, Kant, Kierkegaard and Husserl. There is also an effort to build up pure and applied ethics and to attract more women to philosophy.

Logic and philosophy of language. The earliest international contributions to philosophy from Norway came in logic, starting with Axel Thue (1863-1922) and culminating with Thoralf Skolem (1887-1963). Oslo now has six professors of logic, three in the mathematics department and three in linguistics. Jens Erik Fenstad (1935-) has taught most of them.

In the Philosophy Department in Oslo, Dagfinn Føllesdal (1932-) started out as a student of Skolem, but went to Harvard to take a Ph.D. in philosophy with Quine, and now concentrates mainly on philosophy of language and phenomenology. Also in the Philosophy Department in Oslo is Andrew Jones (1947-), who is working in deontic logic and also in philosophy of mind, philosophy of language, artificial intelligence and particularly the philosophy of law. Jones has also been advisor to an impressive number of doctoral students. Olav Gjelsvik (1956-), also in Oslo, received his D.Phil. in Oxford 1986 and is working in the philosophy of mind, philosophy of language and action theory. In Trondheim Ingemund Gullvåg (1925-) has written on definiteness of intention, and in his later work has explored a wide variety of ideas in epistemology, metaphysics and philosophy of logic.

Philosophy of science. Philosophy of science, including philosophy of the humanities and the social sciences, is a main theme in the compulsory philosophy exam at the Norwegian universities. This has led to considerable activity within this field.

Jon Elster (1940-), based at Columbia University, has written an impressive number of widely read and highly regarded books on Marx and Leibniz, but mostly on issues relating to rationality and deviations from rationality, including addiction. He has also published empirical studies of justice. Nils Roll-Hansen (1938-), who combines philosophy of science with history of biology, has written on Pasteur, Wilhelm Johannsen and Lysenko, arguing from detailed analysis of historical case studies that there is a basis for objectivity and rationality in empirical science.

In Bergen, Gunnar Skirbekk (1937-) and Ragnar Fjelland (1947-) have established a centre for the study of the sciences and humanities. Fjelland, who came from engineering to philosophy, is working on philosophical issues connected with science and technology, while Skirbekk has written several books in French, English and German, on political philosophy, rationality and other topics. Together with Nils Gilje (1947-) he has written a history of philosophy, which has been translated into several languages. Erik Brown (1943-) in Bergen and Magne Dybvig (1940-) and Sverre Sløgedal (1930-) in Trondheim are working in epistemology, while Audun Øfsti (1938-) in Trondheim and several of the philosophers in Troms, the world's northernmost university, are working in transcendental pragmatics.

Greek philosophy. The study of Greek philosophy in Norway is concentrated in Oslo, which has a special chair. Egil Wyller (1925-), who has written on Plato's Parmenides and worked out a Plato-inspired philosophy of unity called 'henology', has been succeeded in this chair by Eyjólfur Kjalar Emilsson (1943-), a native of Iceland. Emilsson, whose main work has been on Plotinus, now has two younger colleagues who, like himself, each

received their Ph.D. in Greek philosophy from Princeton.

Kant. Several philosophers are working on Kant and German transcendental philosophy, among them two women in Oslo, Hjørdis Nerheim (1940-) - Norway's only woman professor in philosophy - and Camilla Serck-Hanssen (1960-), who has her Ph.D. from the University of California, San Diego. She worked there with Henry Allison (1937-), who since 1995 has been Professor at the University of Oslo.

Kierkegaard studies. Alastair Hannay (1932-) came to Oslo from London in 1961. He served as professor in Trondheim from 1975 to 1985 before returning to Oslo, where he was appointed professor in 1990. He has written extensively on Kierkegaard and has translated several of Kierkegaard's works into English. He has also written on the philosophy of mind, and since 1971 has been editor of the journal *Inquiry*, which was founded by Arne Næss in 1958.

Husserl's phenomenology. In Norway, the study of phenomenology started in Oslo in the 1950s and then spread to the other universities. In 1956 Dagfinn Føllesdal wrote his M.A. thesis on Husserl and Frege, and the following year Hans Skjervheim (1926-) wrote his on 'Objectivism and the study of man'. Skjervheim, who was professor in Bergen from 1982 until he retired, criticized various tendencies to 'objectivize' mental and social phenomena and not recognize their intentionality. Skjervheim has been important for an anti-positivist movement in Norway. There are now several young philosophers in Norway, chiefly in Oslo, working in phenomenology and related fields.

Wittgenstein. Wittgenstein had his hut in Skjolden, one of the inner branches of the Sogne Fjord north of Bergen. Appropriately, the University of Bergen has taken the responsibility for The Norwegian Wittgenstein Project, whose aim is to register Wittgenstein's *Nachlaß* in electronic form.

Ethics. Ethics is an area of rising activity, and the central figure here has been Knut Erik Tranøy (1918-), professor in Bergen from 1959 to 1978, and in Oslo from 1978 to 1986. Tranøy was the first professor of medical ethics in Oslo (1986-8). In 1989 a centre for medical ethics was founded in Oslo on Tranøy's initiative. Tranøy has written on ethics, on science as a norm-guided activity, on Aquinas and on Wittgenstein. Also Tore Nordenstam (1934-) in Bergen has written extensively on ethics and also on aesthetics, as has Kjell S. Johannesen (1936-). In Oslo, Jon Wetlesen (1940-) has been a main contributor to ethics.

To build up more competence in ethics in Norway, the Norwegian Research Council started a ten-year ethics programme in 1991, which has thirty people working towards doctorates in ethics. Most concentrate on applied ethics and already have a doctorate or similar background in the field whose ethics they are studying.

There is also an effort to bring more women into philosophy. In addition to those mentioned above, there is Vigdis Songe-Møller (1949-) in Bergen, who has written on Parmenides and on women's issues. Also Else Viestad (1940-) is working on women's issues, as well as Else Barth (1928-), professor emeritus of analytic philosophy in Groningen.

6 Iceland

Iceland, with a population of merely 266,000, has two universities: the University of Iceland in the capital, Reykjavik, founded in 1911, and a small university in Akureyri in the north, established in 1987. At both universities the compulsory 'philosophicum' is taught. This has led to there now being five professorships in philosophy in Reykjavik and several lectureships there and in Akureyri. Since its beginning, Icelandic philosophy has been internationally oriented, and Icelandic philosophers have usually gone abroad for their doctorates. However, after their return to Iceland much of their effort has gone into making classical and contemporary philosophy accessible to Icelanders through translation and popular publications, and they have published relatively little for an international public. They have been working in all the central areas of philosophy, and also in the history of philosophy, on Plotinus, Descartes, Hume, Nietzsche, Husserl, Heidegger, Ingarden and Ricoeur.

DAGFINN FØLLESDAL

References and further reading

1 General surveys

Bostad, I. and Svenneby, E. (1994) *Gender - an Issue for Philosophy?*, Oslo: Nordic network for women in philosophy.(Articles by women philosophers from the Nordic countries, with a roster of Nordic women philosophers and their work.)

Burr, J.R. (ed.) (1980) *Handbook of World Philosophy: Contemporary Developments Since 1945*, Westport, CT: Greenwood Press.(Articles on philosophy in the Nordic countries by Wedberg, Hintikka, Hartnack, Næss and Hellesnes.)

Embree, L. (ed.) (1997) *Encyclopedia of Phenomenology*, Dordrecht: Kluwer.(The article on Scandinavia surveys work in phenomenology.)

Klibansky, R. and Pears, D. (eds) (1993) *La philosophie en Europe*, Paris: Gallimard. (The article 'Pays nordiques' (Nordic countries) is a guide to young philosophers and their work.)

Olson, R.E. and Paul, A.M. (1972) *Contemporary Philosophy in Scandinavia*, Baltimore, MD: Johns Hopkins University Press.(Articles by Scandinavian philosophers, with an excellent survey of philosophy in Scandinavia by Georg Henrik von Wright.)

2 Sweden

Bergström, L. (1966) *The Alternatives and Consequences of Action*, Stockholm Studies in Philosophy, Stockholm: Almqvist & Wiksell.(A very precise, critical discussion of some of the basic notions in utilitarianism.)

Dahlquist, T. (1976) 'Theoretical Philosophy, Practical Philosophy. Some pages from the History of Philosophy in Uppsala', *Uppsala University 500 years*, 5: Faculty of Arts at Uppsala University: History, Art and Philosophy, 129-46.(A brief survey of the history of philosophy in Uppsala.)

Halldén, S. (1986) *The Strategy of Ignorance: From Decision Logic to Evolutionary Epistemology*, Library of Theoria 17, Stockholm: Thales.(How decision logic helps us to understand the wisdom of living beings.)

Henschen-Dahlquist, A.-M. (1993) *En Ingemar Hedenius bibliografi*, Stockholm: Thales, 1993.(A bibliography of one of Sweden's publicly most visible philosophers.)

Kanger, S. (1957) *Provability in Logic*, Stockholm Studies in Philosophy, Stockholm: Almqvist & Wiksell. (Dissertation, supervised by Wedberg, which anticipated Kripke.)

Marc-Wogau, K. (1967) *Philosophical Essays*, Library of Theoria 11, Lund: Gleerup.(Contains a bibliography of the writings of Konrad Marc-Wogau.)

Ofstad, H. (1990) *Our Concept of Weakness*, Stockholm: Almqvist & Wiksell.(Major study of Nazism by a former Professor of Practical philosophy in Stockholm.)

Phalén, A. (1914) *Zur Bestimmung des Begriffs des Psychischen*, Skrifter utgifna af K. Humanistiska Vetenskaps-Samfundet i Uppsala, Leipzig: Harrassowitz.(A critical discussion of various attempts to define consciousness, settles for a Husserl-inspired conception of consciousness as directedness.)

Prawitz, D. (1965) *Natural Deduction. A Proof Theoretical Study*, Stockholm Studies in Philosophy, Stockholm: Almqvist & Wiksell.(Dissertation, supervised by Wedberg, which was the first to prove Gentzen's Hauptsatz directly for certain systems of natural deduction.)

Sahlin, N.-E., and Segerberg, K. (eds) (1993) *The Philosophy of Sören Halldén*, (*Theoria* vol. 59), Stockholm: Thales.(Contains a bibliography of the writings of Sören Halldén.)

Wedberg, A. (1955) *Plato's Philosophy of Mathematics*, Stockholm: Almqvist & Wiksell.(One of Wedberg's few, but high-quality publications.)

Wedberg, A. (1982-4) *History of Philosophy*, vols 1-3, Oxford: Oxford University Press.(A history of philosophy where the arguments stand in focus.)

3 Denmark

Blegvad, M. (1972) 'The Philosophy of Value of Harald Höffding', in R.E. Olson and A.M. Paul, *Contemporary Philosophy in Scandinavia*, Baltimore, MD: Johns Hopkins University Press, 1972.(The author's major works in ethics are all written in Danish.)

Collin, F. (1985) *Social Reality*, London: Routledge & Kegan Paul.(Clarification of key issues in contemporary social science by a young Copenhagen philosopher.)

Favrholdt, D. (1965) *An Interpretation and Critique of Wittgenstein's 'Tractatus'*, Copenhagen: Munksgaard. (Important dissertation which earned its author the first professorship in Odense.)

Hartnack, J. (1995) *Hegels Logik*, Frankfurt: Peter Lang.(The latest book by the prominent Aarhus philosopher.)

Jørgensen, J. (1964) 'Bibliography', *Danish Yearbook of Philosophy*, vol. 1, Copenhagen: Munksgaard, 183-96. (The bibliography is the last section of a Festschrift for Jørgensen.)

Løgstrup, K.E. (1997) *The Ethical Demand*, Notre Dame, IN: University of Notre Dame Press.(The main work of Denmark's most influential theologian/philosopher.)

Sandelin, K. (ed.) (1932) *Harald Höffding in Memoriam*, Copenhagen: Munksgaard.(Comprehensive bibliography of works by and on Höffding.)

4 Finland

Aarnio, A. (1983) *Philosophical Perspectives in Jurisprudence*, Helsinki: Acta Philosophica Fennica 36.(Starts with a forty-page survey of the development of legal theory and philosophy of law in Finland; extensive bibliography.)

Alanen, L. (1982) *Studies in Cartesian Epistemology and Philosophy of Mind*, Helsinki: Acta Philosophica Fennica 33.(Dissertation of Scandinavia's foremost Descartes-scholar.)

Bogdan, R. (ed.) (1987) *Jaakko Hintikka*, Profiles 8, Dordrecht: Reidel.(Essays on Hintikka's philosophy, with responses by Hintikka and a bibliography of his writings.)

Hintikka, J. (1996-) *Selected Papers*, Dordrecht: Kluwer.(Volume 1, *Ludwig Wittgenstein* and volume 2, *Lingua Universalis vs. Calculus Ratiocinator*, are the first of a multivolume edition of Hintikka's papers.)

Kaila, E. (1979) *Reality and Experience: Four Philosophical Essays*, Vienna Circle Collection, Dordrecht: Reidel. (One of the most characteristic works of the philosopher who inspired von Wright, Stenius and Ketonen.)

Niiniluoto, I. (1993) 'From logic to love: The Finnish tradition in philosophy', *Proceedings of the Estonian Academy of Sciences, Humanities and Social Sciences* 42 (4): 369-77.(A brief history of philosophy in Finland.)

Plato, J. von (1994) *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*, Cambridge: Cambridge University Press.(A major contribution to the history of probability theory by a young Finnish philosopher.)

Pörn, I. (1984) *The Philosophy of Erik Stenius*, (*Theoria* vol. 50), Stockholm: Thales. (With a bibliography of the writings of Erik Stenius.)

Stenius, E. (1960) *Wittgenstein's Tractatus*, Oxford: Oxford University Press. (Incisive interpretation and defence of Wittgenstein's picture theory of language.)

5 Norway

Elster, J. (1979) *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge: Cambridge University Press.(One of the author's many influential books on rationality and deviations from rationality.)

Hannay, A. (1990) *Human Consciousness*, London: Routledge.(The author's most recent book on the philosophy of mind.)

Roll-Hansen, N. (1979) 'Experimental Method and Spontaneous Generation: The Controversy between Pasteur and Pouchet 1859-64', *Journal of the History of Medicine and Allied Sciences* 34: 273-92.(Combines the history and the philosophy of biology in a way characteristic of the author.)

Skirbekk, G. (1993) *Rationality and Modernity: Essays in Philosophical Pragmatics*, Oslo: Universitetsforlaget. (Essays by a productive Bergen philosopher.)

Tranøy, K.E. (1988) *The Moral Import of Science*, ed. A.J.I. Jones, Bergen: Sigma. (With a bibliography of the writings of Knut Erik Trany.)

6 Iceland

Emilsson, E.K. (1988) *Plotinus on Sense Perception*, Cambridge: Cambridge University Press.(A main work by an Icelandic philosopher who is now professor in Oslo.)

Karlsson, M.M. (1978) 'Doubt, reason and Cartesian therapy', in M. Hooker (ed.) *Descartes: Critical and Interpretive Essays*, Baltimore, MD: Johns Hopkins University Press.(One of the most influential articles from Iceland.)

Scepticism

Simply put, scepticism is the view that we fail to know anything. More generally, the term 'scepticism' refers to a family of views, each of which denies that some term of positive epistemic appraisal applies to our beliefs. Thus, sceptical doctrines might hold that none of our beliefs is certain, that none of our beliefs is justified, that none of our beliefs is reasonable, that none of our beliefs is more reasonable than its denial, and so on. Sceptical doctrines can also vary with respect to the kind of belief they target. Scepticism can be restricted to beliefs produced in certain ways: for example, scepticism concerning beliefs based on memory, on inductive reasoning or even on any reasoning whatsoever. And sceptical views can be restricted to beliefs about certain subjects: for example, scepticism concerning beliefs about the external world, beliefs about other minds, beliefs about value and so on. Solipsism - the view that all that exists is the self and its states - can be seen as a form of scepticism based on the claim that there are no convincing arguments for the existence of anything beyond the self.

The philosophical problem of scepticism derives from what appear to be very strong arguments for sceptical conclusions. Since most philosophers are unwilling to accept those conclusions, there is a problem concerning how to respond to the arguments. For example, one kind of sceptical argument attempts to show that we have no knowledge of the world around us. The argument hinges on the claim that we are not in a position to rule out the possibility that we are brains-in-a-vat being artificially stimulated to have just the sensory experience we are actually having. We have no basis for ruling out this possibility since if it were actual, our experience would not change in any way. The sceptic then claims that if we cannot rule out the possibility that we are brains-in-a-vat, then we cannot know anything about the world around us.

Responses to this argument often fall into one of two categories. Some philosophers argue that we can rule out the possibility that we are brains-in-a-vat. Others argue that we do not need to be able to rule out this possibility in order to have knowledge of the world around us.

1 The philosophical problem of scepticism

Most contemporary discussions of scepticism have focused on scepticism concerning the external world. We can use this type of scepticism to illustrate the broader philosophical problem, as many of the arguments we consider can be applied *mutatis mutandis* to other types of scepticism.

One type of scepticism denies that we know anything about the external world. The view is not simply that, for example, by gathering more evidence we could come to know. Rather, it is that we are unable to attain knowledge. On the plausible assumption that knowledge entails justified belief, scepticism concerning knowledge follows from scepticism concerning justified belief - the view that justified belief about the external world is unattainable.

Scepticism is of philosophical interest because there appear to be very strong arguments that support it. This presents us with the problem of how to respond to these arguments. One way would be to accept their conclusion. Of course, very few philosophers are willing to do this. There are very few actual sceptics. So the problem of scepticism is how to refute or in some way neutralize or deflate the force of these arguments.

In the history of philosophy, some sceptical arguments have been based on the unreliability or relativity of our senses (see Pyrrhonism), or upon the inability of reason to produce non-question begging arguments for our beliefs (see Hume, D. §2). Nearly all sceptical arguments exploit sceptical hypotheses or alternatives. Sceptical alternatives suppose that the world is very different from what we would normally believe on the basis of our sensory evidence. This entails that our sensory evidence is radically misleading. More precisely, suppose we claim to know a proposition q on the basis of evidence e. Let (proposition) h be an *alternative* to q just in case h is incompatible with q (q and h cannot both be true). Then h is a *sceptical alternative* to q provided h is an alternative to q compatible with e. An alternative of this kind has sceptical force precisely because it is compatible with the evidence we claim gives us knowledge of q. For example, ordinarily, I would claim to know on the basis of my visual evidence that I am currently looking at my computer monitor. One sceptical alternative, introduced by Descartes (1641), is that the world of familiar objects does not exist and that I am being deceived into thinking it does by a powerful demon. The demon causes me to have just the sensory experiences I would have if the world of familiar objects existed (see Descartes, R. §4). According to a modern version of this alternative, I am a

brain-in-a-vat being artificially stimulated to have all the experiences I would have if I had a body and interacted, in the normal way, with the world of familiar objects. These alternatives are incompatible with what I claim to know about the familiar world around me since according to those alternatives, that world does not exist. Moreover, since these alternatives entail that it appears to me as if that world exists, they are compatible with my evidence.

Sceptical alternatives provide the basis for very powerful sceptical arguments. Exactly how they do this is a matter of some controversy. The quickest route to scepticism is through what I will call the entailment principle:

S knows *q* on the basis of (evidence) *e* only if *e* entails *q*

Since a sceptical alternative is, by definition, a proposition incompatible with *q* but compatible with *e*, it follows from the mere existence of sceptical alternatives of the kind we have been considering that we do not know those empirical propositions we ordinarily claim to know. But, this argument is only as good as the entailment principle. Should we accept this principle? In effect, the principle says I can know *p* only if my evidence precludes the possibility of error. Though many philosophers concede that this principle has considerable intuitive force, most have thought, in the end, that it should be rejected. This position is sometimes called fallibilism (see Commonsensism §§1-2; Fallibilism). Of course, few philosophers believe that scepticism should be avoided at all costs. But when given a choice between scepticism and fallibilism, most philosophers opt for fallibilism (at the expense of the entailment principle).

Does fallibilism beg the question against scepticism? After all, precisely what the sceptic claims is that the existence of alternatives consistent with our evidence undermines our claims to know. Fallibilists merely respond that the alternatives the sceptic has invoked do not undermine our knowledge claims: that is, we can know even when there are such alternatives. Since this is the point at issue, fallibilists seem to need an argument in support of this crucial claim.

Here, fallibilists can appeal to our strong intuition that in many cases we do know things, despite the existence of sceptical alternatives. And it is not clear that the sceptic can undermine those intuitions except by appealing to the entailment principle - which is itself undermined by those very intuitions. Thus neither side of the debate may be able to defend its position without begging the question.

Unfortunately scepticism is not so easily dispatched. The sceptic can turn the appeal to our ordinary intuitions against fallibilism. For some of those intuitions can provide the basis for a new sceptical argument. This argument begins by claiming, quite plausibly, that whatever else we may say about the significance of sceptical alternatives, we cannot claim, plausibly, to know they are false. For example, we cannot claim, plausibly, to know that we are not brains-in-a-vat being artificially stimulated to have exactly the same experience we would have as normal human beings. None of our evidence counts against this hypothesis since if it were true, we would have precisely that evidence.

But how, exactly, does this permit the sceptic to conclude we do not know the propositions we ordinarily claim to know? At this point, the sceptic appeals to a very intuitive principle that is weaker than the entailment principle. This principle says that the set of known (by *S*) propositions is closed under known (by *S*) entailment:

If *S* knows *q*, and *S* knows that *q* entails not-*h*, then *S* knows not-*h*

While one could quibble with some details about this principle, it (or something very much like it) seems compelling (see Deductive closure principle). From this principle and the claim that we fail to know sceptical alternatives are false, it follows that we fail to know the propositions we ordinarily claim to know (since we know those propositions entail the falsity of sceptical alternatives).

2 Responses to scepticism

This argument presents problems for fallibilism, as I have characterized it, since the argument at no point presupposes the entailment principle. The sceptical argument we are now considering merely exploits the fallibilist position that permits the existence of alternatives to known propositions.

Fallibilist responses come in two forms, each of which corresponds to the denial of one of the two premises of the

sceptical argument. One response denies the closure principle. For example, Dretske (1970) has argued that the fact that we do not know the falsity of sceptical alternatives shows that the closure principle is false, since we do know the truth of many empirical propositions that (we know) entail the falsity of sceptical alternatives. According to this view, certain alternatives are not relevant to whether one knows a proposition: one does not have to know such an alternative to *q* is false in order to know *q*. So, for example, one can know that one sees a zebra without knowing that the alternative - that one sees a cleverly disguised mule - is false, because that alternative is not relevant. This version of fallibilism is sometimes called the 'relevant alternatives' view.

The other fallibilist response to the sceptical argument agrees with the sceptic that the closure principal is true. But, against the sceptic, these fallibilists deny the claim that we fail to know the falsity of sceptical alternatives. One version of this fallibilist response uses the closure principle along with the claim that we do have knowledge, to reject the claim that we do not know that sceptical alternatives are false. They argue from the premise that we know some ordinary proposition *q* and the premise that if we know *q* then we know any proposition that we know is entailed by *q* (the closure principle), to the conclusion that we know that we are not seeing a cleverly disguised mule. We can call this view '*modus ponens* fallibilism'.

3 Relevant alternatives fallibilism

As we have noted, the sceptic attempts to undermine our claims to know by calling attention to sceptical alternatives. The relevant alternatives response to this sceptical manoeuvre is to deny that these alternatives are relevant. An alternative, *h*, to *q*, is relevant just in case we need to know *h* is false in order to know *q*. So if *h* is not a relevant alternative, we can still know *q* even if we fail to know *h* is false. This view entails that the deductive closure principle is false.

There are two ways to argue for this view. The direct way is to cite alleged counterexamples to the deductive closure principle. Some philosophers have done this by appealing both to our intuition that we know many propositions about the external world and to our intuition that we fail to know the falsity of sceptical alternatives. So my strong intuition that I know I am looking at my computer monitor and my strong intuition that I fail to know I am not a brain-in-a-vat constitute the basis for such a counterexample.

A more indirect way to argue for this view is to construct a theory of knowledge that has as a consequence, the failure of the closure principle, as in Nozick (1981). The basic idea of these kinds of theories is that knowing requires the truth of certain subjunctive conditionals. On one (simplified) version, my knowing *q* requires that:

(S) If *q* were false, I would not believe *q*

This requirement for knowledge precludes my knowing I am not a brain-in-a-vat. For I would still believe I am not a brain-in-a-vat, even if I were a brain-in-a-vat. But, this requirement allows me to know I see a computer monitor. For it seems plausible to claim that I would not believe I see the computer monitor if I were not seeing it.

A significant difficulty for the direct way of arguing for the relevant alternatives view - the appeal to counterexamples to the closure principle - is that the intuitions that support the counterexamples seem no more compelling than the intuitions in favour of the closure principle. Many think that the closure principle expresses a fundamental truth about our concept of knowledge. So much so that if a certain theory of knowledge entails the falsity of the closure principle, some philosophers are inclined to take the fact as a *reductio ad absurdum* of that theory.

But this presents problems for the indirect way of arguing for the relevant alternatives view: some philosophers reject theories that endorse condition (S), for the very reason that it entails the falsity of the closure principle. Moreover, there are other difficulties for theories that endorse conditions like (S). One problem for these theories is that they seem to preclude our knowing much of what we take ourselves to know inductively. Consider an example where you leave a glass containing some ice cubes outside on an extremely hot day (Vogel 1987). Several hours later, while you are still inside escaping the heat, you remember the glass you left outside. You infer that the ice must have melted by now. Here we have an ordinary case of knowledge by inductive inference. According to the theories we are now considering, my knowing that the ice cubes have melted requires the truth of this subjunctive conditional:

(S′) If the ice cubes had not melted, I would not believe that they had

But (S′) looks false. It seems plausible to claim that had the ice cubes not melted, it might have been for some reason (for example, someone putting them in a styrofoam cooler) that would still leave me believing they had melted. Thus, it looks as if theories which endorse this condition are too strong. If this is correct, then the anti-sceptical results afforded by condition (S) come at the cost of scepticism about certain kinds of inductive knowledge.

We should note, however, that there is some controversy over the evaluation of subjunctive conditionals like (S′). But I think it is fair to say that standard semantics for subjunctive conditionals would render (S′) as false (see Deductive closure principle §§2-3).

4 *Modus ponens* fallibilism

Modus ponens fallibilists accept, along with the sceptic, the deductive closure principle. But they attempt to turn that principle against the sceptic. Like relevant alternatives fallibilists, they take as a starting point the strongly intuitive claim that we do know many things about the world. They then note that, given the closure principle, it follows that we know the falsity of sceptical alternatives. For example, I now know that I am looking at my computer monitor. I also know that my looking at a computer monitor precludes my being a brain-in-a-vat. It follows by the closure principle that I know I am not a brain-in-a-vat.

Is this piece of reasoning legitimate? One might challenge those who reason in this way to explain how we know sceptical alternatives are false. How, for example, do I know I am not a brain-in-a-vat? After all, the sceptical problem arises because we seem to lack any reason for believing sceptical alternatives are false. These alternatives are constructed so as to make it impossible for our evidence to count against them. Presumably, our recognition of this explains, at least in part, our intuition that we fail to know sceptical alternatives are false.

One way for the *modus ponens* fallibilist to try to meet this challenge is to claim that I can know:

not-*h*: I am not a brain-in-a-vat

by inferring it from:

q: I am looking at my computer monitor

According to this way of proceeding, even though none of my evidence for *q* counts in favour of not-*h*, it does not follow that I have no reason to believe not-*h*. For that reason can be *q* itself. Since I know *q* (on the basis of my visual evidence) and I know that *q* entails not-*h*, I can infer not-*h* from q and thereby come to know not-*h*.

Is this reasoning legitimate? Let’s compare it with another case. Suppose I park my car in front of the market and go inside. Although I am not currently looking out the window I can still know:

p: My car is parked in front of the store

Can I then come to know:

r: My car has not been towed away

simply by inferring it from *p*? Notice that *p* entails *r*. It seems, none the less, that I would already need to have sufficient evidence to know *r* before I could infer *p*. And if my initial evidence is insufficient for me to know *r*, I cannot infer *p* and so I cannot infer *r* from *p*.

The *modus ponens* fallibilist reasoning concerning sceptical alternatives looks suspicious because it seems like the reasoning in the parked car case. Intuitively, I need to have reason to believe not-*h* before I can infer (and thereby come to know) *q*. Thus I cannot first infer *q* and then go on to infer (and thereby come to know) non-*h*.

Another version of fallibilism argues for the claim that we know sceptical alternatives are false by appealing to principles of inductive inference. One version of this view argues that the hypothesis that the familiar world of objects exists is the best explanation of our sensory evidence (and so a better explanation than sceptical

alternatives). This licenses an inference from our sensory evidence to the familiar-world hypothesis (see Inference to the best explanation). We can thereby come to know that this familiar world exists. And since we know that the familiar-world hypothesis rules out the sceptical alternatives, it follows by the closure principle that we know sceptical alternatives are false.

The burden for this view is to say why the familiar-world hypothesis is a better explanation of our sensory evidence than any sceptical alternative. This is not easy to do since sceptical alternatives are designed to explain our sensory evidence. Proponents of the view that sceptical alternatives provide inferior explanations often appeal to pragmatic considerations like simplicity and conservatism. But there are several problems with this approach. Even if we could establish that the familiar-world hypothesis is, for example, simpler than any sceptical alternative, why should we think that this supports the claim that the hypothesis is true? Unless this crucial link can be made, it is not clear how this response to the sceptic can succeed (see Theoretical (epistemic) virtues).

Moreover, often arguments that the familiar-world hypothesis is the best explanation of our sensory data are quite sophisticated and complex. This raises the worry that only those who are philosophically sophisticated enough to follow such an argument can have knowledge of the external world.

5 The role of intuitions

Many fallibilist responses to the sceptic take as their starting point our ordinary intuitions about knowledge or our everyday pattern of knowledge attributions. But how exactly can our everyday pattern of knowledge attributions have force against sceptical arguments, since the sceptic is calling into question precisely these attributions?

The reason our ordinary intuitions about knowledge have force against the sceptic is that these intuitions persist even in the face of sceptical arguments. When we confront a sceptical argument, even though we may not be able to say where the argument goes wrong, we are reluctant to withdraw our everyday knowledge attributions. This is the basis of G.E. Moore’s famous response to sceptical arguments. Moore claimed to be more sure that he knew some things, for example, that he has a hand, than he is that the sceptical argument is sound. So even though he could not say where the sceptical argument goes wrong, he thought it more rational to suppose that there is a mistake in the sceptical argument than to suppose that the conclusion of the argument - we fail to know anything - is true (see Moore, G.E. §3; Commonsensism).

The sceptic could try to dismiss the significance of our reluctance to withdraw our everyday knowledge attributions as nothing more than the persistence of old habits. This persistence of our habitual ways of thinking about knowledge even after we have been confronted with sceptical arguments was noticed by Descartes and by Hume.

But in response, we can note that, often, we find our everyday pattern of knowledge attributions compelling even while we are in the midst of sincere philosophical reflection. The fact is that when we think about sceptical arguments, we often find ourselves pulled in two directions. We feel the pull of the sceptical argument and yet we remain reluctant to give up our claims to know. This phenomenon cannot be dismissed as nothing more than an unreflective habit. So the fallibilist can maintain that our everyday knowledge attributions reflect deep-seated intuitions about our concept of knowledge. Since our intuitions are a kind of data that any theory of knowledge must explain, they present a formidable challenge to the sceptical position.

Nevertheless, there is something unsatisfying about rejecting scepticism just because it conflicts with our intuitions about knowledge. For, again, it is hard to deny the force of the sceptical argument. And just as our intuitions about our everyday knowledge attributions present a problem for scepticism, so our sceptical intuitions present a challenge to our everyday knowledge attributions. If scepticism is a strongly counterintuitive view, then why do sceptical arguments have any grip on us at all? Why do we not immediately respond to sceptical arguments by objecting, for example, that sceptical hypotheses are too remote and fanciful to undermine our knowledge claims? (Either we can know that sceptical alternatives are false or we need not know they are false in order to know things about the external world.) Sometimes we are inclined to do just that. But the sceptical problem arises precisely because we cannot always sustain that attitude. Sometimes, when we consider sceptical arguments, we begin to worry that sceptical alternatives really do threaten our knowledge claims.

What we are confronting here is a paradox - a set of inconsistent propositions, each of which has considerable

independent plausibility:

(1) We know some ordinary empirical propositions.
(2) We do not know that sceptical alternatives are false.
(3) If *S* knows *q*, and *S* knows that *q* entails not-*h*, then *S* knows not-*h*.

One of these propositions must be false (on the assumption that we know *q* entails not-*h*). Yet each of them is very difficult to deny. This is what explains our vacillation over scepticism. The arguments for scepticism and for fallibilism attempt to exploit the intuitions favourable to them. The sceptic appeals to (2) and (3), and concludes that (1) is false. Relevant alternatives fallibilism appeals to (1) and (2), and concludes that (3) is false. *Modus ponens* fallibilism appeals to (1) and (3), and concludes that (2) is false. Because each member of the set has independent plausibility, it seems arbitrary and unsatisfying to appeal to any two members of this triad as an argument against the third. Such a strategy does not provide what any successful resolution of a paradox should provide, namely an explanation of how the paradox arises in the first place. Any satisfying resolution of the paradox that defends our claims to know against the sceptic must explain the appeal of sceptical arguments. For it is that very appeal that gives rise to the paradox.

This is where Moore's response to the sceptic goes wrong. Many philosophers think that Moore begged the question against scepticism. In a way he did, but no more so than the sceptic begs the question against him. Still, there is something quite unsatisfying, philosophically, about Moore's treatment of the sceptical argument. But the problem with it is not that it begs the question against the sceptic. Rather the problem is that it fails to explain the dialectic force of sceptical arguments. Though it is possible that the apparent cogency of sceptical arguments is explained by some very subtle error in our reasoning, the simplicity of these arguments suggests that their appeal reveals something deep and important about our concept of knowledge. That is why we can learn much about the nature of knowledge by grappling with the problem of scepticism.

See also: Causation, Indian theories of §2; Induction, epistemic issues in; Internalism and Externalism in epistemology; Justification, epistemic §6; Knowledge, concept of §8; Knowledge, Indian views of §4; Perception, epistemic issues in; Phenomenology, epistemic issues in; Rational beliefs §2; Reliabilism; Scepticism, Renaissance; Solipsism

STEWART COHEN

References and further reading

Cohen, S. (1988) 'How to be a fallibilist' in J. Tomberlin (ed.) *Philosophical Perspectives*, 2, Atascadero, CA: Ridgeview, 91-123.(This contributor's treatment of scepticism.)
Cornman, J. (1980) *Scepticism, Justification, and Explanation*, Dordrecht: Reidel.(Defends inference to the best explanation response - see §4.)
Descartes, R. (1641) *Meditations on First Philosophy*, in E. Haldane and G.R.T. Ross (eds) *The Philosophical Works of Descartes*, Mineola, NY: Dover Publications, 1955. (Meditation I contains a classic statement of the sceptical problem.)
Dretske, F. (1970) 'Epistemic Operators', *Journal of Philosophy* 67: 1007-23. (Argues against the deductive closure principle for knowledge.)
Klein, P. (1981) *Certainty*, Minneapolis, MN: University of Minnesota Press. (Detailed defence of *modus ponens* fallibilism.)
Moore, G.E. (1959) 'Certainty', in W. Doney (ed.) *Descartes: A Collection of Critical Essays*, Garden City, NY: Doubleday, 1967.(Classic statement of *modus ponens* fallibilism.)
Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press.(Detailed defence of a theory that endorses condition *S* - see chapter 3.)
Unger, P. (1975) *Ignorance*, New York: Oxford University Press.(Influential defence of scepticism.)
Vogel, J. (1987) 'Tracking, Closure and Inductive Knowledge', in S. Luper-Foy (ed.) *The Possibility of Knowledge: Nozick and His Critics*, Totowa, NJ: Rowman & Littlefield.(Critical discussion of condition *S*.)
Wittgenstein, L. (1921) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London: Routledge & Kegan Paul, 1961.(Interesting treatment of solipsism.)

Scepticism, Renaissance

Ancient Greek scepticism was revived during the Renaissance, and played an important role in the religious and philosophical controversies of the time. There is little evidence that ancient scepticism was known directly during the Middle Ages, or that its perplexing questions played any significant role in medieval thought. It was indirectly known from the writings of Augustine; some manuscripts of the texts of Cicero and Sextus Empiricus were available; and occasionally vague reference to some sceptical details appears in medieval discussions. However, the interests of scholastic philosophers were, by and large, far removed from the questions about the sources, reliability and certainty of knowledge claims that concerned the ancient sceptics. With the humanistic revival of interest in ancient literature there came a rediscovery of scepticism as presented in the writings of Cicero, Sextus Empiricus and Diogenes Laertius. Cicero's Academics (Academica) was read from the fourteenth century on; Life of Pyrrho by Diogenes Laertius was rediscovered in the early fifteenth century; and Greek manuscripts of the writings of Sextus were brought from Constantinople into Italy in the mid-fifteenth century. These treasuries of sceptical argumentation were used in many ways in the Renaissance. At first they were seen largely as sources of information about the ancient world, but gradually more attention was paid to the actual arguments they contained. Some saw these arguments as a basis for rejecting Aristotelian philosophy, as well as other ancient dogmatic claims about nature and humanity. Others used them as ammunition in the great religious controversies between Catholics and Protestants. A full-fledged scepticism about knowledge claims was developed in the second part of the sixteenth century, through the work of Sanches and Montaigne. Montaigne was particularly inspired by the first published Latin translations of the writings of Sextus Empiricus. Scepticism in all its forms was closely associated with fideism. If it is impossible to acquire knowledge of anything through the senses and reason, then it is impossible to acquire knowledge of God in these ways, and one can argue that religious truth must be accepted on the basis of faith in divine revelation. The weak recommendation of ancient sceptics to suspend belief while accepting local customs as the guide for conduct was thus turned into a strong recommendation to adopt Christian beliefs. Only later did epistemological scepticism become associated with scepticism about religious beliefs themselves. Renaissance scepticism in its various guises was a major intellectual force in the transition from scholasticism to modern thought.

1 Recovery of ancient texts

The most important direct sources of ancient scepticism are Cicero's philosophical dialogue, *Academics (Academica)*, written in Latin (see Cicero §3); *Outlines of Pyrrhonism* and *Against the Professors (Adversus Mathematicos)* by Sextus Empiricus, both written in Greek; and the *Life of Pyrrho* which Diogenes Laertius, included in his *Lives of the Philosophers*, also written in Greek. Through these works Renaissance authors became acquainted with two types of scepticism, both of which employed arguments and dialectical puzzles to cast doubt on sense perception and reason as sources of knowledge and on the possibility of formulating any reliable criterion for judging knowledge claims. Academic sceptics (whose views were formulated from the third to the first century BC), at least as interpreted during the Renaissance, were those who denied the possibility of any knowledge, thus adopting negative dogmatism. They were also said to have allowed probable judgments, a stance criticized by Montaigne (§3) who said 'Either we can judge absolutely, or we absolutely cannot.' Pyrrhonian sceptics, who were followers of Pyrrho, preferred to say that judgment should be suspended altogether. This epistemological stance had moral implications, for the sceptic was then exhorted to live peacefully in accordance with the laws and customs of society.

During the Middle Ages, scepticism was known only indirectly through Augustine's attack on Academic scepticism in his *Contra Academicos* and *De Trinitate* (see Augustine §2). The *Life of Pyrrho* was not available at all, and while there are some medieval manuscripts of Cicero's *Academics* and some early fourteenth-century manuscripts of a partial Latin translation of *Outlines of Pyrrhonism* by Sextus Empiricus, there is no evidence that anyone made use of this material. Only in fourteenth-century Byzantium was there any interest in ancient scepticism. Nor was there any independent development of scepticism. Neoplatonic doubts about sense perception as a basis for knowledge, various types of mysticism, and fourteenth-century discussions both of intuitive cognition and of God's absolute power certainly allowed opportunity for such development, but the opportunity was not taken. The main factor in the revival of scepticism was the recovery of the ancient texts.

These texts first became known in Italy through the work of the humanists, though after the 1520s the focus of attention shifted to northern Europe. Cicero’s *Academics* was the first to be read, though it never became very popular. While it had been available before the fourteenth century, it became better known after Petrarch (see Petrarca, F.) referred to it. It was printed with the other works of Cicero in 1471, but the first separate printed text only dates from 1535 and the first commentary from 1536. Diogenes Laertius’ *Lives of the Philosophers* (including the *Life of Pyrrho*) was widely known and studied from the early fifteenth century in Latin and Italian translations, and the first Greek text was printed in 1533. It was used to supplement the information available from Latin sources.

The most important ancient source is Sextus Empiricus. His work was almost unknown until the mid-fifteenth century, when manuscripts from Constantinople began to circulate in Italy. By the end of the century Greek manuscripts were available in Rome, Venice and Florence, especially in the library of the convent of San Marco where Girolamo Savonarola was the prior. At first, Sextus Empiricus seems to have been regarded as a literary, philological and historical source, rather than as a source of epistemological problems. Serious philosophical attention was paid to him in northern Europe after the publication of two Latin translations. His *Outlines of Pyrrhonism* was published by the Protestant Henri Estienne in 1562, and his *Against the Professors* was published (along with Estienne’s translation of *Outlines of Pyrrhonism*) in 1569 by the Catholic Counter-Reformer Gentian Hervet. The Greek texts were not printed until 1621.

The term ‘sceptic’ itself comes from the recovery of the ancient texts, for *scepticus* and its Latin cognates became an accepted part of the intellectual lexicon only during the fifteenth century.

2 Informal sceptical currents

Many developments in the Renaissance raised problems and questions about accepted beliefs, and gave rise to an informal scepticism about previous knowledge claims in many areas. One important factor was the European discovery of America and the first voyages round Africa to India, China and Japan which led to the discovery of new cultures, civilizations and religions all over the planet. Information brought back by the explorers and later by Jesuit missionaries, raised doubts about all previous claims about geography, and even about the nature of the human being. Could one still accept the biblical view about human origins? Where did all these different people come from? Were they truly human? Attempts were made to fit everybody into the biblical framework by referring to possible migrations after the Flood and after the Tower of Babel episode. But could this really account for how people crossed wide oceans, and had such different ways of living? Some hardy souls suggested that the American Indians were, perhaps, pre-Adamites, peoples who had an independent origin. If this were the case then the Bible could no longer be considered the complete history of humanity. Thomas Harriot, who accompanied Sir Walter Ralegh to Virginia, is supposed to have presented an ‘atheist’ lecture on his return in which he said that American Indian history was 16,000 years old, while biblical history was supposed to have started only in 4004 BC. Giordano Bruno suggested that there were three original protoplasts from which three different kinds of human being developed. In the seventeenth century, Isaac La Peyrère (1596?-1676) in his book *Prae-Adamitae* (1655) offered the hypothesis that there were humans before Adam who were not created by God, but were part of an eternal universe. Only Adam was divinely created as a way of starting divine history, the history of the Jews, but not the rest of humanity.

Another source of informal scepticism was the influence of humanism on Bible scholarship. The increased use of Greek and Hebrew raised serious sceptical questions as to whether the biblical text is correct, complete and accurate. Erasmus questioned whether the doctrine of the Trinity was really in the original version of the New Testament. Others, especially those using Jewish materials, questioned the accepted Christian readings of the text. Some, like Michael Servetus, who was burned at the stake in 1553 by John Calvin and his followers, even suggested that the Bible had to be understood contextually, in terms of what was going on in ancient pre-Christian times, rather than in terms of foreshadowings of the New Testament. Scholars began questioning the authenticity of some of the canonic writings and the status of the Apocrypha (see Humanism, Renaissance §6).

Along with the growing knowledge of Hebrew during the Renaissance came the awareness of a wholly different way of understanding Scripture and the cosmos, namely that presented in the Jewish Kabbalistic writings. The Kabbalah was a series of medieval Jewish treatises containing Neoplatonic and Gnostic elements (see Kabbalah). The treatises purported to contain an ancient wisdom and to offer non-rational ways of understanding the world.

Many thinkers combined the study of the Kabbalah with the study of the *Corpus Hermeticum*, a body of writings from the second and third centuries AD (see Hermetism). These too contained a mixture of Neoplatonic, Gnostic, Jewish and other elements, and they purported to record a wisdom which predated Greek philosophy. These sources influenced such authors as Giovanni Pico della Mirandola, Leone Ebreo and Agrippa von Nettesheim. The mix of Kabbalistic learning with Neoplatonism gave rise to a wide variety of new religious interpretations, some of them like those of Guillaume Postel and Giordano Bruno, definitely heretical, and sceptical of accepted Christianity.

Another, originally less intellectual, critique of accepted Christianity was offered by Spanish mystics (usually Jewish forced converts) who claimed to be directly illuminated by God. At the beginning of the sixteenth century, some Spanish women with no theological training began offering a version of what they considered pure Christianity, gained through special non-rational practices. Despite persecution by the Spanish Inquisition, the mysticism of the *alumbrados* became a most important spiritual force, especially as expressed by Teresa of Avila and Juan de la Cruz (John of the Cross), and led many to scepticism about accepted Christian theology and the efficacy of Church practices (see Mysticism, history of §6).

Developments in science challenged accepted views in medicine, astronomy and physics. Paracelsus (§3) presented a new view of the role of medicine in relation to human nature and the cosmos; Copernicus raised questions about the accepted geocentric theory of the universe; and Bruno (§§4-5) and others cast doubt on Aristotelian science.

Some people (for example, Jardine 1983) have suggested that humanist dialectic, in placing an emphasis on techniques of persuasion and probable argumentation, was influenced by Academic scepticism as presented by Cicero. However, Lorenzo Valla (§4) rejected such nonstandard argument forms as *sorites*, and later humanist logicians failed to present a genuine informal logic of probabilities (see Logic, Renaissance §4).

All these currents none the less created an informal kind of scepticism that was pervasive in the late Renaissance world. As John Donne put it, ‘new philosophy calls all in doubt’. This informal general scepticism was often combined with an informal use of the ancient sceptical texts. References to items that appear in Sextus and Cicero are in many authors of the period. Rabelais (§2), for instance, introduced a Pyrrhonian professor as a comic foil to Pantagruel in a discussion of whether Panurge should marry. The Pyrrhonian professor uses the terms of ancient scepticism without really developing the view. Other thinkers contented themselves with a general attack on the possibility of knowledge. For instance, Agrippa von Nettesheim (§3) in his *De incertitudine et vanitate scientiarum et artium (Of the Vanity and Uncertainty of Arts and Sciences)* (1526) - a most popular work of the time, translated into many languages - attacked all kinds of intellectual activity and knowledge claims in a way that is anecdotal rather than systematic, though he did draw materials from both Cicero and Sextus.

3 Early scepticism and the Reformation

The first use of Greek scepticism in the religious controversies of the time appears in the views of Girolamo Savonarola and his leading disciple, Gianfrancesco Pico della Mirandola (1469-1533), the nephew of the great fifteenth-century humanist Giovanni Pico. Savonarola was a Dominican monk and philosophy teacher in Florence who pressed for reform of the Catholic Church. Although he campaigned against pagan philosophy, he apparently suggested the reading of Sextus as an introduction to Christian faith, and he wanted his monks to prepare a Latin translation of Sextus’ works. This did not happen, as he was condemned as a heretic and burned at the stake in 1498. His disciple, Gianfrancesco Pico, presented Sextus’ arguments for the European audience in his anti-Aristotelian work, *Examen vanitatis doctrinae gentium (An Examination of the Emptiness of Pagan Learning)* (1520). Like Savonarola, he saw Sextus’ attacks on all forms of intellectual pursuits as a way of making people accept religious revelation as the sole basis for knowledge. He made serious use of Sextus’ arguments against various ancient schools of philosophy, and he attacked Aristotelian sense-based epistemology. Pico’s work had some influence on various Renaissance thinkers, and was known to Gassendi (§3) and Leibniz (§10). It was used along with Sextus’ arguments by a Venetian rabbi in the early seventeenth century as a means of getting Jews to accept the Torah as the sole source of truth.

The early use of scepticism in the religious controversies of the Reformation is illustrated by the debate between Desiderius Erasmus (§§4, 6) and Martin Luther. Erasmus was interested in Academic scepticism as presented by

Cicero. In his *In Praise of Folly* (1511), where he criticized all kinds of philosophical views, Erasmus said that he found the Academic sceptics ‘the least surly’ of all the philosophers because they had the good sense not to dogmatically assert anything. Erasmus’ own *philosophia Christi* was an acceptance of the spirit of Christ’s teachings without buttressing them with any metaphysical principles or systems.

When Luther challenged the Church of Rome’s unique claim to determine Christian truth, Erasmus was initially supportive of this challenge to an apparently overbearing authority, but as Luther himself became more and more dogmatic in his pronouncements, Erasmus slowly withdrew his support and remained within the Catholic Church. He was urged by Catholic leaders to speak out against the Reformation and finally did so in his *Discourse on Free Will* (1524), which challenged Luther. Erasmus contended that the subject was too complicated for human beings to understand, and the Scriptures were too difficult to interpret on the matter. Because of this Erasmus recommended that one adopt the sceptical attitude of suspending judgment while accepting the traditional view, that of the Church, undogmatically.

Luther found Erasmus’ gentle scepticism, offered as a reason for remaining a Catholic, totally unacceptable. Christians, Luther insisted, cannot be sceptics, since they must be certain rather than dubious of their beliefs, given that salvation is at stake. Faith and scepticism are incompatible, for a Christian must ‘delight in assertions’. Luther declared that his erstwhile friend, Erasmus, could remain a sceptic if he wished, but he should be aware that Judgment Day was coming and that *Spiritus Sanctus non est scepticus* (‘The Holy Ghost is not a sceptic’).

The Erasmus-Luther debate raised a fundamental sceptical problem: how does one establish what criterion to employ in determining religious truth? Luther had challenged the Church’s traditional criteria (the Pope, Church councils, Church tradition), and he defiantly appealed to Scripture alone. But people’s readings of Scripture varied, and Luther then appealed to Scripture as interpreted by the Holy Spirit. This introduced what was in effect a subjective private criterion - the dictates of the Holy Spirit to each person’s conscience. In the on-going battle between Catholics and Protestants, the problem of the criterion raised by the ancient sceptics thus became a living issue, and the texts of Cicero and Sextus became a vital part of the argument.

Both Estienne and Hervet in their prefaces to their translations of Sextus said that this ancient treasury of sceptical arguments would be useful for those engaged in the religious debates of the time. In fact Hervet, who was the secretary of the Cardinal of Lorraine and a participant at the Council of Trent, said that ancient scepticism would provide the perfect answer to Calvinism, then making great inroads in France. If nothing can be known, then Calvin’s knowledge claims cannot be known; the Reformers made the same claim about Catholicism. Each side then proposed a sceptical fideism, or a Christian Pyrrhonism. If sceptical difficulties about knowledge cannot be resolved, then religious views must be accepted on the basis of faith alone.

4 Scepticism and anti-scholasticism

The sceptical texts were also used in the anti-scholastic controversies. Cicero’s *Academics* and commentaries upon it were appealed to in support of Petrus Ramus’ drive for a new non-scholastic method of understanding and teaching various subjects, and in support of Francesco Patrizi’s 1581 blast against all forms of Aristotelianism (see Ramus, P.; Patrizi da Cherso, F. §2). Some of these writers were accused of being, or were labelled by themselves, *nouveaux académiciens* or *nouveaux pyrrhoniens*, and some traditionalists decried the menace of these new sceptical groups.

An important figure in this context was Omer Talon, a close associate of Petrus Ramus in Paris. In 1547 (with a reissue in 1550) he published Cicero’s *Academics* with a lengthy introduction and a commentary. Whether or not Talon was a *nouveau académicien*, as he was accused of being, he used ancient scepticism to show the need for less dogmatism and for free inquiry in philosophizing. This would make the entrenched scholasticism of the universities less powerful, and so clear a path for Ramism.

Talon’s work was attacked by Pierre Galland, Professor at the Collège Royal, who published a polemical oration in 1551 in favour of the schools of Paris against ‘the new Academy of Petrus Ramus’. The aim of Ramus and Talon, according to their critic, was to attack the Gospel after having destroyed all of philosophy! Galland’s long oration was apparently taken sufficiently seriously that Rabelais saw fit to ridicule it. Another criticism of the *nouveaux académiciens* was written by Guy de Brués, a friend of Ramus, and a member of the group of writers, La Pléiade, connected with the great poet, Pierre Ronsard. In 1557, he published a set of dialogues in which four

actual writers of the time discuss scepticism, and the sceptic finally is led to abandon his doubts, saying ‘O miserable Pyrrho, who has made all into opinion and indifference!’ In his preface De Brués claimed he was writing in order to prevent young people from giving up religion because of sceptical doubts. The dialogues do not really bring out either the strong force of Greek sceptical arguments, or the struggle to answer them, but they indicate the importance intellectuals of the time gave to the presence of scepticism.

Another work which should be mentioned is the *Academica* of Pedro de Valencia (1596) which gives a favourable historical analysis of the development of Academic scepticism, and which ends on a fideist note, saying that God is the only source of truth.

5 Philosophical scepticism: Sanches, Montaigne, Charron

Modernized versions of ancient philosophical scepticism were advanced by two distant cousins, Francisco Sanches and Michel de Montaigne. Both were educated, though not at the same time, at the Collège de Guyenne in Bordeaux, and both also studied at the University of Toulouse, then a centre of the ferment of new philosophical and religious ideas. Sanches then studied medicine at Montpellier and Rome and became professor of philosophy and later of medicine at the University of Toulouse. Montaigne studied law and became a political figure in Bordeaux, serving as mayor of the city and a councillor to Henri de Navarre, the Protestant leader who later turned Catholic and became King Henri IV. There is no evidence that they ever knew each other or read each other’s writings, but each produced his version of modern scepticism at about the same time. Montaigne wrote the major part of his longest and most sceptical essay, ‘The Apology for Raymond Sebond’, in 1575-6, when Sanches was already at work on his book *Quod nihil scitur (That Nothing is Known)*, published in 1581.

As the title of his book suggests, Sanches (§2) was most influenced by Academic scepticism, and indeed, he makes no explicit reference to Sextus Empiricus. The main thrust of his work was an attack on Aristotelian science, viewed as a demonstrative process that employs definitions to give necessary reasons or causes for natural events. He used sceptical arguments to show that neither definitions nor syllogistic reasoning could produce knowledge, and that the appeal to knowledge of an object’s causes leads to an infinite regress. His own view was that knowledge would have to deal with particulars, each understood individually, but, he argued, human beings are incapable of reaching any certainty about individuals, both because of the unlimited and changing nature of reality, and because of our own imperfections. As a result, nothing can be known. However, again following the spirit of Academic scepticism, Sanches did allow some constructive results. Patient, careful empirical research combined with cautious judgment and evaluation of observed data will not give us certain knowledge, but it will allow us to obtain the best information available, and give us a useful way of dealing with human experience.

Montaigne explicitly rejected Academic scepticism. He regarded the claim that nothing could be known as just the kind of dogmatic assertion the sceptic should reject, and he did not approve of any appeal to probable judgment, as that too involves assertion. He preferred to follow the road of Pyrrhonian scepticism as presented by Sextus Empiricus. In ‘The Apology for Raymond Sebond’ he developed in a gradual manner the many kinds of problem that make people doubt the reliability of human reason. He cast doubt on human intellectual pretensions by comparing human abilities to those of animals; and he drew on the reports of ancient philosophers and modern explorers concerning the great variety of customs and behaviour in different parts of the planet to suggest that there was no way of determining right or true standards. He argued that even scientists were merely offering personal opinions that were likely to be replaced by other people’s opinions. Finally, he considered in more detail the unreliability of the information gained by the senses or by reason, and the inability of human beings to find a satisfactory criterion of knowledge. He concluded that people should suspend judgment on all matters, and that they should follow customs, traditions and social rules undogmatically, while being tolerant of other people’s views. He combined these conclusions with fideism. The state of complete doubt and suspension of judgment leaves the human mind ‘a blank tablet prepared to take from the finger of God such forms as he shall be pleased to engrave on it’. Religious beliefs should be based solely on faith rather than on dubious evidence (see Montaigne, M. de).

Montaigne’s Pyrrhonian scepticism was presented and popularized in didactic form by Pierre Charron (§3). In his *De la Sagesse (On Wisdom)* (1601), a work often republished during the seventeenth century, he advanced a fideistic defence of religious thought which was based on accepting complete scepticism while appealing to faith alone as the source of religious knowledge.

6 Impact of Renaissance scepticism

Renaissance scepticism, in both its informal manifestations and especially as put forth by Sanches, Montaigne and Charron, played a most important role in the development of modern thought. Seventeenth-century thinkers either accepted scepticism, offered ways of living with it, or tried to refute it. Scepticism was used by Mersenne (§2) and Gassendi (§§3-5) to attack Renaissance naturalism, astrology and alchemy, and as the basis for proposing a constructive or mitigated scepticism, a way in which the emerging modern mechanical science could be accepted hypothetically rather than as a necessarily true picture of the world. Others, such as René Descartes (§4), demanded some certainty that no sceptics could challenge, and so developed the first modern philosophical systems.

See also: Cicero §3; Diogenes Laertius; Humanism, Renaissance; Montaigne, M. de; Pyrrhonism; Sanches, F.; Scepticism; Sextus Empiricus

RICHARD H. POPKIN

References and further reading

Floridi, L. (1995) ‘The Diffusion of Sextus Empiricus’s Works in the Renaissance’, *Journal of the History of Ideas* 56: 63-85.(Floridi surveys the manuscripts of Sextus that were available in the fifteenth and sixteenth centuries, and argues that the failure of the humanists to use Sextus’ sceptical arguments was due more to lack of interest than to ignorance.)

Jardine, L. (1983) ‘Lorenzo Valla: Academic Skepticism and the New Humanist Dialectic’, in M. Burnyeat (ed.) *The Skeptical Tradition*, Berkeley, CA, and London: University of California Press, 253-86.(Referred to in §2. A controversial discussion of links between scepticism, humanist logic and probable argumentation.)

Popkin, R.H. (1979) *The History of Scepticism from Erasmus to Spinoza*, Berkeley, CA: University of California Press.(Basic study of the development of scepticism from the early sixteenth century on; substantial bibliography.)

Popkin, R.H. (1993) ‘The Role of Scepticism in Modern Philosophy Reconsidered’, *Journal of the History of Philosophy* 31: 501-17.(An examination of how the work of Popkin and Schmitt has stood up, and directions for future investigations.)

Popkin, R.H. (1993) ‘Scepticism and Modernity’, in T. Sorell (ed.) *The Rise of Modern Philosophy*, Oxford: Clarendon Press, 15-32.(Recent evaluation of how sceptical thought helped bring about ‘modernity’.)

Popkin, R.H. (1996) ‘Scepticism and Prophesy’, *British Journal for the History of Philosophy* 4: 1-20.(On Savonarola and Gianfrancesco Pico and the use of scepticism to advance prophetic religion.)

Popkin, R.H. and Schmitt, C.B. (eds) (1987) *Scepticism from the Renaissance to the Enlightenment*, Wolfenbütteler Forschungen Band 35, Wiesbaden: Harrassowitz. (Several relevant articles, including Schmitt on the development of the historiography of scepticism.)

Schmitt, C.B. (1969) *Gianfrancesco Pico della Mirandola (1469-1533) and His Critique of Aristotle*, The Hague: Martinus Nijhoff.(The only study of Gianfrancesco Pico’s philosophy, especially his sceptical work.)

Schmitt, C.B. (1972) *Cicero Scepticus: A Study of the Influence of the Academica in the Renaissance*, The Hague: Martinus Nijhoff.(A study of the role played in the Renaissance by Cicero’s *Academics (Academica)*.)

Schmitt, C.B. (1983) ‘The Rediscovery of Ancient Skepticism in Modern Times’, in M. Burnyeat (ed.) *The Skeptical Tradition*, Berkeley, CA, and London: University of California Press, 225-51.(The best available information on how ancient scepticism entered Renaissance thought; good bibliography.)

Scheler, Max Ferdinand (1874-1928)

Max Scheler, usually called a phenomenologist, was probably the best known German philosopher of the 1920s. Always an eclectic thinker, he was a pupil of the neo-idealist Rudolph Eucken, but was also strongly influenced by the life-philosophies of Dilthey and Bergson. While teaching at Jena he regularly met Husserl, the founder of the phenomenological movement, and his mature writings have a strongly phenomenological, as well as a Catholic, stamp. Later he turned towards metaphysics and the philosophical problems raised by modern science.

Scheler's interests were very wide. He tried to do justice to all aspects of experience - ethical, religious, personal, social, scientific, historical - without doing away with the specific nature of each. Above all, he took the emotional foundations of thought seriously. Many of his insights are striking and profound, and sometimes his arguments are very telling, but his power to organize his material consistently and to attend conscientiously to the business of justification is poorly developed.

Scheler is best known for his anti-Kantian ethics, based on an a priori emotional grasp of a hierarchy of objective values, which precedes all choice of goods and purposes. He himself describes his ethics as 'personalist', and makes personal values supreme, sharply distinguishing the 'person' from the 'ego', and linking this with his analysis of different types of social interaction. In epistemology he defends a pragmatist approach to science and perception; thus philosophy, as the intuition of essences, requires a preparatory ascetic discipline. His philosophy of religion is an attempt to marry the Augustinian approach through love with the Thomist approach through reason. In his later work, to which his important work on sympathy provides the transition, he defends a dualist philosophical anthropology and metaphysics, interpreting the latter in activist terms as a resolution of the tensions between spiritual love and vital impulse.

1 Life

Scheler was a Bavarian, born at Munich on 22 August 1874 to a Protestant father and a Jewish mother. He was three times married and twice divorced. He died at Frankfurt am Main, 19 May 1928.

Between 1894 and 1899 he studied at the universities of Munich, Berlin (under Dilthey), Jena (under Rudolph Eucken), and Heidelberg (to hear Max Weber). Both his doctoral promotion and habilitation were achieved underEucken. He was baptised a Catholic on 20 September 1899.

While lecturing at Jena, he had several meetings with Husserl, who helped him transfer to Munich in 1906. He quickly won a leading role among the Munich phenomenologists (see Husserl, E.; Phenomenological movement). However, irregularities in his private life led the university authorities to withdraw his license to teach. There followed a period of unofficial lecturing at Göttingen in 1910-11, thanks to the phenomenologists there, and then also in Berlin from 1912, where his first mature works began to appear: *Zur Phänomenologie und Theorie der Sympathiegefühle und von Liebe und Haß (On the Phenomenology and Theory of Sympathy and of Love and Hate)* (1913), *Der Formalismus in der Ethik und die materiale Wertethik* (*Formalism in Ethics and Non-Formal Ethics of Values*) (1913, 1916) and *Vom Umsturz der Werte (The Subversion of Values)* (1915a). Poor eyesight disqualified him from active service in the war, but he was employed by the state to give propaganda lectures.

In 1918 he was called to the chair of Philosophy and Sociology and the Directorship of the Institute of Social Sciences at Cologne, where he published *Vom Ewigen im Menschen (On the Eternal in Man)* (1921), *Die Wissensformen und die Gesellschaft (Society and the Forms of Knowledge)* (1926) and *Die Stellung des Menschen im Kosmos (The Place of Man in the Cosmos)* (1927). In 1928 he was called to Frankfurt, but died of a heart attack before he could take up office. *Philosophische Weltanschauung (Philosophical Perspectives)* was published in 1929.

2 General characteristics of his philosophy

Scheler's work is usually divided into three periods. His few early writings show the influence of Eucken's neo-idealism, and are separated by about eight years from a productive period beginning around 1912, with his second marriage, and lasting till about 1922.Scheler is here at his most phenomenological and Catholic. But in the early 1920s, at Cologne, he begins to pay less attention to what is timeless in human experience, and more to what

is changing and evolving. Theism gives place to pantheism. This change is first apparent in his sociologically inclined writings, but then dominates his metaphysics. This late work only reached the stage of partial sketches before his death.

Scheler's philosophy is hardly ever academic or dry. Like Eucken, he thought that philosophy should change people's lives, and this gives his phenomenology a very different feel from that of Husserl. Even when not primarily engaged in some critique of the modern world, as in many of his minor works, he often seems to be trying to create a change of heart in the reader. Even his metaphysics is conceived in terms of 'engagement' for the full unfolding of the deity, in which he sees the meaning of the developing universe. He never preaches, but his writings have a vital, passionate and challenging quality.

One formal key to his thought is his use of the categories of higher and lower. Like Greek and medieval thinkers, he is always revealing hierarchies: of values, forms of social relating, levels of consciousness and embodied existence, of types of being. These often furnish an effective instrument in the critique of other philosophical positions. They culminate in his late philosophical anthropology, where the hierarchical structure of human being, now focused on the duality of spirit and drive (a dualism nearer to Kant than Descartes), becomes the key to the metaphysics.

Scheler's thought is highly eclectic. His receptive mind could not help learning even from those he most strongly criticized, above all Kant. But he also took much fromSpinoza, Hegel, Schopenhauer, Nietzsche and Dilthey. Even his conscious borrowings (from Augustinianism, Malebranche,Pascal, Maine de Biran, for example) are always thoroughly assimilated. He took nothing that he could not experience for himself. It is this that gives his work its ultimate unity.

Scheler made an extremely powerful impression on his contemporaries, both academic and lay, especially as a lecturer and teacher, but also as a writer, clearing the ground in many spheres and making new beginnings or revivals possible. Had it not been for the twelve-year Nazi ban on his work, his influence on later generations would probably have been much greater. In fact, many related thinkers, including Heidegger and the early Sartre, owe more to him than is often acknowledged, and his influence in the social sciences was considerable, and is growing once more. Both the sociology of knowledge and philosophical anthropology received their initial impetus from his later work. But perhaps his greatest achievement was to re-open people's eyes to a world where vital, spiritual and religious values and phenomena (essences and necessary relations between them) could all be acknowledged in their own right, free of the monistic reductions of psychologism or positivism, or the anthropocentric distortions of Neo-Kantianism.

3 Philosophy and other forms of knowledge

Although Scheler also called himself an ethical personalist, he implies that he was already thinking along phenomenological lines when he met Husserl, by whom he was clearly influenced. However, he differs from him in important respects. Apart from a lack of interest in method as such and a refusal to follow him in his 'idealist' turn, Scheler has a different approach to the intuition of essence, the question 'what exactly is given in experience?'. Husserl saw 'phenomenological reduction', which enabled us to investigate essences without the distractions of real objects and states of affairs, as a 'bracketing' of existence, a purely intellectual operation. But Scheler interpreted existence as 'resistance to striving'. It is 'given' to us (though never known) at the lowest level of psychic functioning (analogous to Aristotle's 'vegetable soul'), where our vital impulses blindly encounter something that confines them. Thus phenomenological reduction must be an inhibition of impulse. This removes the barriers to spiritual cognition of all kinds. Philosophy thus requires a moral culture of humility, self-knowledge and self-control.

This account of philosophy occurs in *On the Eternal in Man*. Later he calls it *Bildungswissen*, or knowledge for 'formation', since spiritual persons are 'formed' by the essences and values that become part of them. All knowing is, in Greek and Scholastic fashion, 'having a share in' an object without changing it. The object known, *qua* essence, is henceforward 'in' the mind of the knower. Philosophy is also a 'desymbolization of the world'. Our natural, self- or group-centred 'vital' approach to things (later writers talk of the 'life-world') rests on a classification according to vital importance. Perception itself thus selects from the totality of the given. The partial aspects of an object's nature that we take in (the redness of a cherry, for example) function as symbols of the

object both as it essentially is and as it relates to our purposes. But the philosophical inhibition of impulse enables us to get beyond the screen of concepts to the entire essential nature, and see what ‘can only be seen’.

The ‘natural worldview’ given in culture and language is the basis for control over our environment. Knowledge from this viewpoint can, however, be as adequate as that of philosophy or, at the other extreme, science. But its objects are relative to the concerns of the human species (with cultural modifications). Scientific objects, on the other hand, are relative to those of living beings as such, since science transcends the limits of human sensory equipment, and thus the range of human concerns. The world as reduced by science, largely unintelligible in terms of our sensory imaginations, is an almost pure (and largely probabilistic) schema of the world, from which the essential content of things has almost disappeared. The schema is structured as a device for control, the world of pure pragmatism. Scientific knowledge is thus *Leistungswissen*, or knowledge for achievement.

The third form of knowledge is *Heilswissen*, or knowledge for salvation. In Scheler’s Catholic period, salvation concerns moral and religious knowledge and practice, and the salvation of the spiritual individual as member of the ‘community of love’. But in the last period *Heilswissen* is metaphysics, and salvation itself concerns the resolution of all tensions and oppositions in the absolute being, whose spirit is the sphere of emergent essences and whose body is the material universe. Human beings provide the arena where this reconciliation of the spiritual and the material takes place. Thus metaphysics is engagement for the deity, an active spiritual love of essences and values together with an ecstatic participatory ‘yes’ to the blind processes of material and vital energy. The *Nachlaß* volumes which supplement the last lectures and papers show us something more of this grand conception, but very little is worked out in any detail.

4 Ethics

Scheler’s ethics depends on his theory of value, which did not change in his last years. His first phenomenological and Catholic works are, *inter alia*, concerned to defend the objectivity of values and a less materialistic approach to the world. Among his particular targets are Nietzsche, with his attack on Christian morality, and the utilitarian or calculating attitude to morals, based on an all-devouring resentment (‘Das Ressentiment im Aufbau der Moralen’ (Resentment in the Structure of Morals) (1915)) and the elevation of hedonic values above those of vitality. His great ethical treatise, *Formalism in Ethics and Non-Formal Ethics of Values*, rich in subtle analyses and distinctions, is primarily intended as a refutation of Kantian ethics, together with the philosophical assumptionsKant uses to justify his position (see Kantian ethics). The huge, rambling, but immensely impressive work has six parts, each criticizing some specific Kantian theme.

In the course of it Scheler defends his thesis that values are the intentional objects of feeling. They are qualities felt in or on objects, and are hence inaccessible to the understanding. They constitute an ‘emotional a priori’, since the phenomena of goods and purposes, and even the attractions of pleasure, presuppose them. Far from being formal, they are given to us, in positive and negative forms, in distinct kinds of material and in a hierarchical order, with the hedonic or sensory values at the bottom, then those of vitality, then spiritual values of beauty, justice and assent to pure (non-pragmatic) truth, and lastly the values of holiness. A separate category of welfare values is sometimes inserted between the hedonic and the vital. There are also other hierarchies of ‘relatively formal’ values, depending on the nature of the value-bearers concerned. But moral values are personal and of a logically second order. A person’s act has moral value when it intends the realization of a first-order value intrinsically higher than the given alternatives. Since choice of value depends on the structure of one’s loving, the main foundation of personal being (see *Bildungswissen* above), Scheler’s ethics has something in common with virtue ethics, and a corresponding distaste for moral obligation, though this is accepted as a regrettable necessity. It also contains illuminating discussions of ethical relativism (where he defends the absoluteness of values as such against various relativities of ethos, ethic, moral code and moral conduct), levels of the emotional life, and retribution and punishment.

5 The person and society

The last section of *Formalism in Ethics and the Non-Formal Ethics of Values* is a long treatise on the person as individual and social being. The influence of Kant is strong in the first part, though Scheler insists that the spiritual person is not just a rational or intellectual being, but also performs emotional acts, acts of will and, above all, acts of love. These depend on the otherwise ‘automatic’ mental functioning of the ‘ego’, ‘through’ which the

'functional unity' of the person acts, thereby raising ego-functioning to a higher level, where questions of meaning and validity are inescapable.

The analyses of forms of social relating begin from the distinction between community and association, but are illuminatingly extended, and integrated with the value-philosophy and analysis of the individual person. Scheler's interesting treatment of the problem of 'other minds' comes in *Wesen und Formen der Sympathie* (*The Nature of Sympathy*) (see Other minds). He here argues that the acquisition of knowledge is, *inter alia*, a matter of filling with particular content and in a definite order the initially empty spheres of being which are given to human consciousness as such. The social sphere (a possibly unfulfilled experience of 'we') is, in fact, filled before the sphere of individuality. Hence the main assumption of the 'problem' of other minds, that self-awareness must come first, is the false insinuation of 'individualism'. Our actual knowledge of the ego-functioning of other egos depends on the 'universal grammar of expression' to which we have access as living beings; in so far as this behaviour is raised to the dignity of spiritual activity, we can know it by performing the acts 'with' or 'after' the persons concerned.

6 Religion

The theory of spheres of being goes with a theory of specific spiritual acts relating to the typical contents of each sphere. In *On the Eternal in Man* Scheler argues that at some time or other we all perform religious acts, which include repentance, petitionary prayer, thanks, praise and worship, in relation to the absolute sphere. It is only through performing such acts that we fully grasp the meaning of holiness, the highest value. This gives us a key to answering the most fundamental question: whether our own religious acts are directed to an adequate object, or only to an idol, such as knowledge, fame or comfort. He also held that the absolute sphere of being is the first to be given some determinate content. Since religious acts presuppose the absoluteness and holiness of their object, which for Scheler denotes the supreme Good, our first and most deeply bedded experiences are of perfect objects and absolute qualities.

This religious, or 'Augustinian', road to God, open to all who try to work out their salvation in practical life, has its counterpart in a philosophical road, which starts from wonder and a concern to grasp the nature of ultimate reality. Scheler briefly expounds the 'three fundamental philosophical truths' in 'Probleme einer Soziologie des Wissens' (*Problems of a Sociology of Knowledge*) (1926). They are, first, that there is 'something and not nothing', second, that existing beings must either be absolute or dependent on something else for their own being, and third, that all being has both essence and existence. Scheler insists that the first two have an emotional content, which can only be grasped in wonder, humility and reverence, and a shuddering glance into the abyss of absolute nothingness. The absolute being which can thus be directly cognized is in complete 'conformity' with the perfect being encountered in religious acts. The two approaches to God are complementary.

7 Sympathy

We may finally return to *The Nature of Sympathy*, which many critics have considered his best work. The general importance of the emotional inScheler's thought is clear, but this work also contains some excellent phenomenological analysis of the different types of sympathy itself, ignored by most writers of 'sympathy ethics'. He goes on to discuss metaphysical theories of sympathy, especially Schopenhauer's, and devotes much attention to *Einsfühlung* ('feeling one with'), which is needed for the 'engagement' of the later metaphysics. The second part of the work contains an analysis of love and hate. Love is the key to the spiritual and personal, since only love opens up the world of essences and values. Humanity is, for Scheler, the *Ens amans*, the being that loves. His own analysis of love as a creative movement of the heart which brings out the higher values of its objects (hatred is destructive, working in a contrary direction) is difficult but profound. The metaphysical writings introduce us to *Eros* or vital love, which plays an important role in his account of evolution, and in the metaphysical reconciliation between spirit and drive.

FRANCIS DUNLOP

List of works

Scheler, M.F. (1954) *Gesammelte Werke (Collected Works)*, Bonn: Bouvier Verlag. (Vols 10-14 contain papers and fragments from the *Nachlaß*.)

Scheler, M.F. (1899) *Beiträge zur Feststellung der Beziehungen zwischen den logischen und ethischen Prinzipien(On the Relations between Logical and Ethical Principles)*, Jena: Vopelius; in *Gesammelte Werke*, vol. 1.(Aspects of the opposition between thought and action, the true and the good.)

Scheler, M.F. (1900) *Die transcendentale und die psychologische Methode (The Transcendental and Psychological Methods)*, Jena: Dürr; in *Gesammelte Werke*, vol. 1.(Dialectical critique of the transcendental and psychological methods in philosophy and establishment of the 'noological'.)

Scheler, M.F. (1913) *Zur Phänomenologie und Theorie der Sympathiegefühle und von Liebe und Haß (On the Phenomenology and Theory of Sympathy, and of Love and Hate)*, Halle: Max Niemeyer.(Phenomenology and theories of sympathy and love; essay on 'other minds'.)

Scheler, M.F. (1913, 1916) *Der Formalismus in der Ethik und die materiale Wertethik*, Halle: Max Niemeyer; in *Gesammelte Werke*, vol. 2; trans. M.S. Frings and R. Funk, *Formalism in Ethics and Non-Formal Ethics of Values*, Evanston, IL: Northwestern University Press, 1973.(Scheler's own position on values and goals, the a priori, consequentialism, imperatives, eudaimonism and the person discussed in opposition to Kant.)

Scheler, M.F. (1915a) *Vom Umsturz der Werte (The Subversion of Values)*, Leipzig: Verlag der weißen Bücher, 2 vols; in *Gesammelte Werke*, vol. 3.(A collection of ethical and phenomenological papers on virtue, ressentiment, human nature, self-knowledge and other topics. Includes 'Das Ressentiment im Aufbau der Moralen', trans. W. Holdheim, ed. Lewis A. Coser, *Ressentiment*, New York, The Free Press of Glencoe, 1961.)

Scheler, M.F. (1915b) *Der Genius des Krieges und der deutsche Krieg (The Spirit of War and the German War)*, Leipzig: Verlag der weißen Bücher; in *Gesammelte Werke*, vol. 4. (Philosophical discussion of war, with special application to the First World War.)

Scheler, M.F. (1917) *Die Ursachen des Deutschenhasses (Why the Germans are Hated)*, Leipzig: Kurt Wolff; in *Gesammelte Werke*, vol. 4.(A moral-pedagogical call to the nation on the basis of psychological analysis.)

Scheler, M.F. (1921) *Vom Ewigen im Menschen*, Leipzig: Neue Geist; in *Gesammelte Werke*, vol. 5; trans. B. Noble, *On the Eternal in Man*, London: SCM Press, 1960.(The nature of philosophy, religion and Christian love.)

Scheler, M.F. (1923) *Wesen und Formen der Sympathie*, Bonn: Friedrich Cohen; in *Gesammelte Werke*, vol. 7; trans. P. Heath, *The Nature of Sympathy*, London, Routledge & Kegan Paul, 1954.(As (1913) with much additional material on the metaphysics of sympathy.)

Scheler, M.F. (1923-4) *Schriften zur Soziologie und Weltanschauungslehre (Papers on Sociology and the Theory of Worldviews)*, Leipzig: Neue Geist; in *Gesammelte Werke*, vol. 6.(Subjects include suffering, love and knowledge, national characteristics in thinking, and Christian social problems.)

Scheler, M.F. (1926) *Die Wissensformen und die Gesellschaft (Society and the Forms of Knowledge)*, Leipzig: Neue Geist; in *Gesammelte Werke*, vol. 8.(Contains 'Probleme einer Soziologie des Wissens', trans. M.S. Frings, ed. K.W. Stikkers, *Problems of a Sociology of Knowledge*, London, Routledge & Kegan Paul, 1980. Also contains the important 'Knowledge and Work: the limits of the pragmatic approach in our world-knowledge'.)

Scheler, M.F. (1927) 'Idealismus-Realismus', in *Philosophischer Anzeiger II*, Bonn: Friedrich Cohen; in *Gesammelte Werke*, vol. 9.(Unfinished; includes material on the metaphysical foundations of knowledge.)

Scheler, M.F. (1928) *Die Stellung des Menschen im Kosmos*, Darmstadt: Otto Reichl; in *Gesammelte Werke*, vol. 9; trans. H. Meyerhoff, *Man's Place in Nature*, New York, Noonday Press, 1961.(Important summary of Scheler's work in philosophical anthropology.)

Scheler, M.F. (1929) *Philosophische Weltanschauung*, Bonn: Friedrich Cohen; in *Gesammelte Werke*, vol. 9; trans. O. Haac, *Philosophical Perspectives*, Boston, MA, Beacon Press, 1958.(Papers on forms of knowledge, history, the future of man, Spinoza.) >

References and further reading

Dunlop, F. (1991) *Scheler*, in *Thinkers of our Time* series, London: The Claridge Press.(A short, clear and untechnical introduction toScheler's life and thought.)

Dunlop, F. (1978) 'Scheler's Theory of Punishment', *The Journal of the British Society for Phenomenology* 9 (3): 167-74.(Gives a short, easy introduction to the social philosophy. The same issue also contains a bibliography of English translations of works by Scheler and five other articles on aspects of his thought.)

Dupuy, M. (1959) *La Philosophie de Max Scheler: son évolution et son unité*, Paris: Presses Universitaires de France, 2 vols.(An excellent introduction to Scheler's thought, including the early works.)

Dupuy, M. (1959) *La Philosophie de Religion chez Max Scheler*, Paris: Presses Universitaires de France.(A full-length study of Scheler's philosophy of religion.)

Frings, M.S. (1965) *Max Scheler: a Concise Introduction into the World of a Great Thinker*, Pittsburgh, PA: Duquesne University Press.(Good coverage of the later work.)
Frings, M.S. (1974) *Max Scheler: Centennial Essays*, The Hague: Nijhoff. (Contains continuation of Hartmann's bibliography.)
Gabel, M. (1991) *Intentionalität des Geistes (What Does Scheler Mean by Intentionality?)*, Leipzig: St Benno Verlag.(A good recent work.)
Good, P. (ed.) (1975) *Max Scheler im Gegenwartsgeschehen der Philosophie (Reactions to Scheler by Contemporary Thinkers)*, Bern: Francke Verlag.(A useful collection of papers, especially on Scheler's phenomenology and anthropology. Good bibliography.)
Hartmann, W. (1963) *Max Scheler: Bibliographie*, Stuttgart-Bad Canstatt: Friedrich Frommann Verlag. (Thorough list of works by and about Scheler, including translations, up to 1963.)
Mader, W. (1980) *Max Scheler: in Selbstzeugnissen und Bilddokumenten (Max Scheler's Life, on the Basis of Autobiographical and Pictorial Sources)*, Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.(Illustrated. Especially good on the relation between the life and the work.)
Spiegelberg, H. (1971) *The Phenomenological Movement: A Historical Introduction*, 2nd edn, The Hague: Nijhoff.(Contains a substantial section onScheler's place in the phenomenological movement.)
Stegmüller, W. (1970) *Main Currents in Contemporary German, British and American Philosophy*, Bloomington, IN: Indiana University Press.(Contains a good critical chapter on Scheler, 101-32.)

Schelling, Friedrich Wilhelm Joseph von (1775-1854)

Like the other German Idealists, Schelling began his philosophical career by acknowledging the fundamental importance of Kant's grounding of knowledge in the synthesizing activity of the subject, while questioning his establishment of a dualism between appearances and things in themselves. The other main influences on Schelling's early work are Leibniz, Spinoza, J.G. Fichte and F.H. Jacobi. While adopting both Spinoza's conception of an absolute ground, of which the finite world is the consequent, and Fichte's emphasis on the role of the I in the constitution of the world, Schelling seeks both to overcome the fatalism entailed by Spinoza's monism, and to avoid the sense in Fichte that nature only exists in order to be subordinated to the I. After adopting a position close to that of Fichte between 1794 and 1796, Schelling tried in his various versions of Naturphilosophie *from 1797 onwards to find new ways of explicating the identity between thinking and the processes of nature, claiming that in this philosophy 'Nature is to be invisible mind, mind invisible nature'. In his System des transcendentalen Idealismus (System of Transcendental Idealism)1800) he advanced the idea that art, as the 'organ of philosophy', shows the identity of what he terms 'conscious' productivity (mind) and 'unconscious' productivity (nature) because it reveals more than can be understood via the conscious intentions that lead to its production. Schelling's 'identity philosophy', which is another version of his* Naturphilosophie*, begins in 1801, and is summarized in the assertion that 'Existence is the link of a being as One, with itself as a multiplicity'. Material nature and the mind that knows it are different aspects of the same 'Absolute' or 'absolute identity' in which they are both grounded. In 1804 Schelling becomes concerned with the transition between the Absolute and the manifest world in which necessity and freedom are in conflict. If freedom is not to become inexplicable, he maintains, Spinoza's assumption of a logically necessary transition from God to the world cannot be accepted. Philosophische Untersuchungen über das Wesen der menschlichen Freiheit und die damit zusammenhängenden Gegenstände (Of Human Freedom) (1809) tries to explain how God could create a world involving evil, suggesting that nature relates to God somewhat as the later Freud's 'id' relates to the developed autonomous 'ego' which transcends the drives which motivate it.*

The philosophy of Die Weltalter (The Ages of the World), on which Schelling worked during the 1810s and 1820s, interprets the intelligible world, including ourselves, as the result of an ongoing conflict between expansive and contractive forces. He becomes convinced that philosophy cannot finally give a reason for the existence of the manifest world that is the product of this conflict. This leads to his opposition, beginning in the 1820s, to Hegel's philosophical system, and to an increasing concern with theology. Hegel's system claims to be without presuppositions, and thus to be self-grounding. While Schelling accepts that the relations of dependence between differing aspects of knowledge can be articulated in a dynamic system, he thinks that this only provides a 'negative' philosophy, in which the fact of being is to be enclosed within thought. What he terms 'positive' philosophy tries to come to terms with the facticity of 'being which is absolutely independent of all thinking' (2 (3): 164). Schelling endeavours in his Philosophie der Mythologie (Philosophy of Mythology) and Philosophie der Offenbarung (Philosophy of Revelation) of the 1830s and 1840s to establish a complete philosophical system by beginning with 'that which just exists…in order to see if I can get from it to the divinity' (2 (3): 158), which leads to a historical account of mythology and Judeo-Christian revelation. This system does not, though, overcome the problem of the 'alterity' of being, its irreducibility to a philosophical system, which his critique of Hegel reveals. The direct and indirect influence of this critique on Kierkegaard, Nietzsche, Heidegger, Rosenzweig, Levinas, Derrida and others is evident, and Schelling must be considered as the key transitional figure between Hegel and approaches to 'post-metaphysical' thinking.

1 Transcendental philosophy and *Naturphilosophie* (1795-1800)

Schelling was born in Leonberg, near Stuttgart, on 27 January 1775. He attended a Protestant seminary in Tübingen from 1790 to 1795, where he was close friends with both Hegel and Friedrich Hölderlin. He moved to Leipzig in 1797, then to Jena, where, via Goethe's influence, he took up his first professorship from 1798 to 1803. From 1803 to 1806 he lived in Würzburg, whence he left for Munich, where he mainly lived from 1806 onwards, with an interruption from 1820 to 1827, when he lived in Erlangen. He moved to Berlin in 1841 to take up what had been Hegel's chair of philosophy. He died on 20 August 1854 in Ragaz, Switzerland.

Schelling's early philosophy was inspired by the French Revolution and by the revolution in philosophy

inaugurated by Kant, particularly as interpreted in the work of J.G. Fichte. The tensions in Schelling's philosophy of this period, which set the agenda for most of his subsequent work, derive from a series of related sources. In the view of the early Schelling, Kant failed to explain the nature of the subject's knowledge of itself: in Kantian terms knowledge could only result from judgments, the synthesis by the subject of intuitions which were given to it from the external world. Although the subject was the condition of possibility or ground of *knowledge*, it seemed unable to ground itself. Kant regards the condition of possibility of the syntheses of knowledge as a 'spontaneity', as cause of itself rather than as the result of other natural causes, but does not succeed in explicating this spontaneity. Along with Kant's approach to the question of grounding knowledge, the most significant other approaches to the issue for Schelling were those of F.H. Jacobi and Fichte.

In 1783 Jacobi became involved in the 'Pantheism controversy', an influential dispute with the Berlin Enlightenment philosopher Moses Mendelssohn over the claim that G.E. Lessing had admitted to being a Spinozist, an admission which at that time was regarded as tantamount to an admission of atheism. In his *Über die Lehre von Spinoza in Briefen an den Herrn Moses Mendelssohn (On the Doctrine of Spinoza in Letters to Herr Moses Mendelssohn)* (1785; 2nd, revised edition 1789), which was influenced by his reading of Kant's first *Critique*, Jacobi revealed a problem which recurs in differing ways throughout Schelling's work. Jacobi's interpretation of Spinozism was concerned with the relationship between what he termed the 'unconditioned' and the 'conditioned', between God as the ground of which the laws of nature are the consequent, and the chain of the deterministic laws of nature. Cognitive explanation relies, as Kant suggested, upon finding a thing's 'condition'. Jacobi's question is how this can ultimately ground the explanation, in that the explanation leads to a regress in which each condition depends upon another condition *ad infinitum*. Any philosophical system thus 'necessarily ends by having to discover *conditions* of the *unconditioned*'. For Jacobi this led to the need for a theological leap of faith if philosophy were to be grounded. In the 1787 Introduction to the first *Critique* Kant maintains that this problem can be overcome by acknowledging that, while reason must postulate the 'unconditioned…in all things in themselves for everything conditioned, so that the series of conditions should thus become complete', by restricting knowledge to appearances, rather than 'things in themselves', the contradiction of seeking conditions of the unconditioned can be avoided.

The condition of the knowledge of appearances for Kant was the 'transcendental subject', but what sort of 'condition' was the transcendental subject? This problem initially united Schelling and Fichte. Fichte insisted in *Wissenschaftslehre* that the establishing of the unconditioned status of the I was required for Kant's system to legitimate itself. He asserts that 'It is…the ground of explanation of all facts of empirical consciousness that before all positing in the I, the I itself must previously be posited', thereby giving the I the founding role which he thought Kant had failed adequately to explicate. Fichte does so by suggesting that the cognitive activity of the I, via which it can reflect upon *itself*, cannot therefore be understood as part of the causal world of appearance and must therefore be part of the noumenal realm, where Kant had located the 'unconditioned'.

Schelling takes up the problems posed by Jacobi and Fichte in two texts of 1795: *Vom Ich als Prinzip der Philosophie oder über das Unbedingte im menschlichen Wissen (Of the I as Principle of Philosophy or on the Unconditional in Human Knowledge)*, and *Philosophische Briefe über Dogmatismus und Kriticismus (Philosophical Letters on Dogmatism and Criticism)*. He reinterprets Kant's question as to the condition of possibility of synthetic judgments a priori as a question about why there is a realm of judgments, a manifest world requiring syntheses by the subject, at all. In *Of the I* Schelling puts Kant's question in Fichtean terms: 'How is it that the absolute I goes out of itself and opposes a Not-I to itself?'. He maintains that the condition of knowledge, the 'positing' by the I of that which is opposed to it, must have a different status from what it posits: 'nothing can be posited by itself as a thing, that is, an absolute/unconditioned thing (*unbedingtes Ding*) is a contradiction'. However, his worry about Fichte's position already becomes apparent in the *Philosophical Letters*, where he drops the Fichtean terminology: 'How is it that I step at all out of the Absolute and move towards something opposed [*auf ein Entgegengesetztes*]?'. The problem Schelling confronted was identified by his friend J.C.F. Hölderlin, in the light of Jacobi's formulation of the problem of the 'unconditioned'. Fichte wished to understand the Absolute as an I. For something to be an I, though, it must be conscious of an other, and thus in a relationship to that other. The overall structure of the relationship could not, therefore, be described from only one side of that relationship. Hölderlin argued that one has to understand the structure of the relationship of subject to object in consciousness as grounded in 'a whole of which subject and object are the parts', which he termed 'being'.

Schelling sought a philosophical way to come to terms with the 'ground' of the subject's relationship to the object world, which avoided the fatalist consequences of Spinoza's system by taking on key aspects of Kant's and Fichte's transcendental philosophy and yet which did not fall into the trap Hölderlin identified in Fichte's conception of an absolute I. In his *Naturphilosophie* (Philosophy of Nature), which emerges in 1797 and develops in the succeeding years, and in the *System des transcendentalen Idealismus* (*System of Transcendental Idealism*) (1800) Schelling wavers between a Spinozist and a Fichtean approach to the problem of the 'unconditioned' (see Naturphilosophie; Spinoza, B. §§2-4). In the *Naturphilosophie* the Kantian division between the appearing world of nature and nature in itself results from the fact that the nature theorized in cognitive judgments is wholly objectified in opposition to the knowing subject. This fails to account for the living dynamic forces in nature, including those in our own organism, with which Kant himself became concerned in the third *Critique* and other late work, and which had played a role in Leibniz's account of nature. Schelling thinks of nature in itself as a 'productivity': 'As the object [*qua* conditioned condition] is never absolute/unconditioned (*unbedingt*) then something *per se* non-objective must be posited in nature; this absolutely non-objective postulate is precisely the original productivity of nature'. The Kantian dualism between things in themselves and appearances is a result of the fact that the productivity can never appear as itself and can only appear in the form of 'products', which are the productivity 'inhibiting' itself. The products are never complete in themselves: they are like the eddies in a stream, which temporarily keep their shape, despite the changing material flowing through them.

Schelling then tries to use the insights of transcendental philosophy, while still avoiding Kant's dualism, to explain our knowledge of nature. Given the fact of knowledge, things in themselves and 'representations' cannot be absolutely different:

> One can push as many transitory materials as one wants, which become finer and finer, between mind and matter, but some time the point must come where mind and matter are One, or where the great leap that we so long wished to avoid becomes inevitable.
>
> (1797, 1 (2): 53)

The *Naturphilosophie* includes ourselves within nature, as part of a necessarily interrelated whole, which is structured in an ascending series of 'potentials' that entail a polar opposition within themselves. The model is a magnet, whose opposing poles are inseparable from each other, even though they are opposites. As productivity, nature cannot be conceived of as an object, since it is the subject of all possible real 'predicates', but its 'inhibiting' itself means that the 'principle of all explanation of nature' is 'universal duality', an inherent difference of subject and object which prevents nature from ever reaching stasis. The sense of nature as an absolute subject links it to the spontaneity of the thinking subject, which is the condition of the syntheses required for the constitution of objectivity. The problem for Schelling lies in explicating how these two subjects relate to each other.

In the *System of Transcendental Idealism* Schelling returns to Fichtean terminology, though he soon finally abandons it. He endeavours to explain the emergence of the thinking subject from nature. This emergence is thought of in terms of an absolute I coming retrospectively to know itself in a 'history of self-consciousness'. The *System* recounts the history of which the transcendental subject is the result. A version of the model Schelling establishes was to be adopted by Hegel in the *Phenomenology of Mind*. Schelling conceives of the whole process in terms of the initially undivided I splitting itself in order to articulate itself in the syntheses, the 'products', which constitute the world of knowable nature. The founding stages of this process, which bring the world of material nature into being, are 'unconscious'. These stages then lead to organic nature, and thence to consciousness and self-consciousness. Schelling claims that the resistance of the noumenal realm to theoretical knowledge results from the fact that 'the [practical] act [of the absolute I] via which all limitation is posited, as condition of all consciousness, does not itself come to consciousness'. He prophetically attempts to articulate a theory which comes to terms with the awareness that thought is driven by forces which are not finally transparent to it, of the kind later to become familiar in psychoanalysis. How, though, does one gain access by thought to what cannot be an object of consciousness?

Schelling adopts the idea from the early Romantic thinkers Friedrich Schlegel and Novalis, whom he knew in Jena at this time, that art is central to understanding what cannot appear as an object of knowledge. Philosophy cannot represent nature in itself because access to the sphere of the unconscious must be via what appears to

consciousness in the realm of theoretical knowledge. The work of art is an empirical object, but if it is not more than what it is *qua* determinable object it cannot be a work of art, which requires the free judgment of the subject. Although the *System* depends upon the transition from theoretical to practical philosophy, which involves breaking Jacobi's chain of 'conditions', Schelling is concerned to understand how the highest insight must yet be into reality as a product of the interrelation of both the 'conscious' and the 'unconscious'. It is not, therefore, a re-presentation of the latter by the former. Whereas in the *System* nature begins unconsciously and ends in consciousness, in the work of art: 'the I is conscious according to the production, unconscious with regard to the product'. The product cannot be understood via the intentions of its producer, as this would mean that it became a 'conditioned' object, which would lack that which makes mere craft into art. Art is 'the only true and eternal organ and document of philosophy, which always and continuously documents what philosophy cannot represent externally'. The particular sciences can only follow the chain of conditions, via the principle of sufficient reason, and must determine the object via its place in that infinite chain. The art object, on the other hand, manifests what cannot be understood in terms of its knowable conditions, in that an account of the materials of which it is made does not constitute it as art. It shows what cannot be said. Philosophy, therefore, cannot positively represent the Absolute, because 'conscious' thinking operates from the position where 'absolute identity' has always already been lost in the emergence of consciousness.

2 Identity philosophy (1801-*c.*1808)

Although the period of Schelling's 'identity philosophy' is usually dated from the 1801 *Darstellung meines Systems der Philosophie* (Presentation of My System of Philosophy) until some time before the 1809 *Philosophische Untersuchungen über das Wesen der menschlichen Freiheit und die damit zusammenhängenden Gegenstände* (*Of Human Freedom*), the project of that philosophy is carried on in differing ways throughout his work. The identity philosophy derives from Schelling's conviction that the self-conscious I must be seen as a result, rather than as the originating act as it is in Fichte, and thus that the I cannot be seen as the generative matrix of the whole system. Again, the problem is to articulate the relationship between the I and the world of material nature, without either reverting to Kantian dualism or falling into the traps of idealism and materialism.

Schelling's mature identity philosophy, which is contained in *System der gesammten Philosophie und der Naturphilosophie insbesondere (System of the Whole of Philosophy and of the Philosophy of Nature in Particular)*, written in Würzburg in 1804, and in other texts between 1804 and 1807, breaks with the model of truth as correspondence:

> It is clear that in every explanation of the truth as a correspondence [*Übereinstimmung*] of subjectivity and objectivity in knowledge, both, subject and object, are already presupposed as separate, for only what is different can agree, what is not different is in itself one.
>
> (1804a, 1 (6): 138)

The crucial problem is explaining the link of the subject and object world, which is what makes judgments possible. For there to be synthetic judgments at all, what is split must, Schelling contends, in some way already be the same (see Truth, correspondence theory of). This has often been understood as leading Schelling to a philosophy in which, as Hegel puts it in the *Phenomenology*, the Absolute is the 'night in which all cows are black', because it swallows all differentiated knowledge in the assertion that everything is ultimately the same. This is not a valid interpretation of Schelling's argument.

In order to get over the problem in monism, of how the One is also the many, Schelling introduces the notion of 'transitive' being, which links mind and matter as predicates of itself. Schelling explains this 'transitivity' via the metaphor of the earth:

> You recognize its [the earth's] true essence only in the link by which it eternally posits its unity as the multiplicity of its things and again posits this multiplicity as its unity. You also do not imagine that, apart from this infinity of things which are in it, there is another earth which is the unity of these things, rather *the same* which is the multiplicity is also unity, and *what* the unity is, is also the multiplicity, and this necessary and indissoluble One of unity and multiplicity in it is what you call its existence… Existence is the link of a being [*Wesen*] as One, with itself as a multiplicity.
>
> (1806a, 1 (7): 56)

‘Absolute identity’ is the link of the two aspects of being, which, on the one hand, is the universe, and, on the other, is the changing multiplicity which the knowable universe also is. Schelling insists now that ‘The *I* think, *I* am, is, since Descartes, the basic mistake of all knowledge; thinking is not my thinking, and being is not my being, for everything is only of God or the totality’: the I is ‘affirmed’ as a predicate of the being by which it is preceded.

Schelling is led to this view by his understanding of the changing and relative status of theoretical knowledge. It is the inherent incompleteness of all finite determinations which reveals the nature of the Absolute, as is evident in his description of time: ‘time is itself nothing but *the totality appearing in opposition to the particular life of things*’, so that the totality ‘posits or intuits itself, by not positing, not intuiting the particular’. The particular is determined in judgments, but the truth of claims about the totality cannot be proved because judgments are necessarily conditioned, whereas the totality is not. Given the relative status of the particular, though, there must be a ground which enables us to be aware of that relativity: this ground must have a different status from the knowable world of finite particulars. At the same time, if the ground were wholly different from the world of relative particulars the old problems of dualism would recur. As such the Absolute *is* the finite, but we do not *know* this in the manner in which we know the finite. Without the presupposition of ‘absolute identity’, therefore, the evident relativity of particular knowledge becomes inexplicable, since there would be no reason to claim that a revised judgment is predicated of the same as the preceding - now false - judgment.

Schelling summarizes the theory of identity as follows:

> For being, actual, real being is precisely self-disclosure/revelation (*Selbstoffenbarung*). If it is to be as One then it must disclose/reveal itself in itself; but it does not disclose/reveal itself in itself if it is not an other in itself, and is *in* this other the One for itself, thus if it is not absolutely the living link of itself and an other.
>
> (1806a, 1 (7): 54)

The link between the ‘real’ and the ‘ideal’, the physical and the mental, cannot, Schelling maintains, be seen as a causal link. Although there cannot be mental events without physical events, the former cannot be causally reduced to the latter: ‘For real and ideal are only different views of one and the same substance’. Schelling wavers at this time between a position of the kind which Hegel soon tried to articulate, in which, in Schelling’s terms, ‘the sameness of the subjective and the objective is made the same as itself, knows itself, and is the subject and object of itself’, in the ‘identity of identity and difference’, and the sense that this position cannot finally circumscribe the structure of the Absolute. The structure of reflection, where each aspect mirrors itself and then is mirrored in the other, upon which this account of the identity of subject and object relies, must be grounded in a being which carries it:

> Reflection…only knows the universal and the particular as two relative negations, the universal as relative negation of the particular, which is, as such, without reality, the particular, on the other hand, as a relative negation of the universal… something independent of the concept must be added to posit the substance as such.
>
> (1804a, 1 (6): 185)

Without this independent basis, subject and object would merely be, as Schelling thinks they are in Fichte, relative negations of each other, leading to a circle ‘inside which a nothing gains reality by the relation to another nothing’. Schelling prophetically distinguishes between the cognitive - reflexive - ground of finite knowledge, and the real - non-reflexive - ground that sustains the movement of negation from one finite determination to another. As a two-sided relationship, reflection alone always entails the problem that the subject and the object in a case of reflection can only be *known* to be the same via that which cannot appear in the reflection: if I am to recognize *myself* in a mirror, rather than a random object in the world, I must already be familiar with myself before the reflection. This means a complete system based on reflection is impossible, because, in order to ground the system, it must presuppose as external to the system what it claims is part of it. From the 1820s onwards, Schelling raises this objection against Hegel’s system of ‘absolute reflection’.

Schelling’s own dissatisfaction with his early versions of identity theory derives from his rejection of Spinozism. Spinoza saw the move from God to the world of ‘conditions’ as a logical consequence of the nature of God. Schelling becomes convinced that such a theory gives no reason why the Absolute or the ‘unconditioned’ should

manifest itself in a world of negative ‘conditions’ at all. Schelling is confronted with the task of explaining the transition from the Absolute to the finite world. In *Philosophie und Religion (Philosophy and Religion)* (1804), he claims, like Jacobi, that there is no way of mediating between conditioned and unconditioned, and already makes the distinction between ‘negative’ and ‘positive’ philosophy, which will form the heart of his late work. Explicating the structure of the finite world leads to ‘negative philosophy, but much has already been gained by the fact that the negative, the realm of nothingness, has been separated by a sharp limit from the realm of reality and of what alone is positive’. The next stage of his philosophy will become concerned with the transition between infinite and finite.

3 The *Ages of the World* (1809-*c.*1827)

Schelling’s work from his middle period is usually referred to as the philosophy of *Die Weltalter* (*Ages of the World*). It begins with *Philosophische Untersuchungen über das Wesen der menschlichen Freiheit und die damit zusammenhängenden Gegenstände* (*Of Human Freedom*) (1809), written in Stuttgart, and lasts until the late 1820s. The *Weltalter* philosophy is an attempt to explain the emergence of an intelligible world at the same time as coming to terms with the inextricable relation of mind to matter. The initial concern is to avoid Spinoza’s fatalism, which renders the human freedom to do good *and* evil incomprehensible. Schelling’s crucial objection is to the idea that evil should be understood as merely another form of negativity, which can therefore be understood by insight into a necessitated totality, rather than as a fact relating to the nature of human freedom. He now sees the fundamental contradictions of the *Naturphilosophie* in terms of the relationship of the intelligibility of nature and ourselves to a ground without which there could be no intelligibility, but to which intelligibility cannot be reduced. In *Of Human Freedom* he introduces, against both Spinoza and Fichte, a conception of ‘willing’, which was later influential for Schopenhauer’s conception of the ‘Will’: ‘In the last and highest instance there is no other being but willing. Willing is primal being, and all the predicates of primal being only fit willing: groundlessness, eternity, being independent of time, self-affirmation’ (1809, 1 (7): 350). Schelling establishes a more antagonistic version of the structure of the identity philosophy. The ‘ground’ is now in one sense ‘groundless’, that is, uncaused: it must be understood in terms of freedom, if Spinozism is to be avoided. At the same time there must be that against which freedom can be manifest for it to be freedom at all. The theory is based on the antagonisms between opposing forces which constitute the ‘ages of the world’. He argues, though, that the world whose origins the *Weltalter* wishes to understand must entail the *same* conflicting forces which still act, though not necessarily in the same form, in this world, of which the mind is an aspect: ‘Poured from the source of things and the same as the source, the human soul has a co-knowledge/con-science (*Mitwissenschaft*) of creation’. Schelling suggests that there are two principles in us: ‘an unconscious, dark principle and a conscious principle’, which must yet in some way be identical. The same structure applies to what Schelling means by ‘God’. As that which makes the world intelligible, God relates to the ground so that the ‘real’, which takes the form of material nature, is ‘in God’ but ‘is not God seen absolutely, that is, in so far as He exists; for it is only the ground of His existence, it is "nature" in God; an essence which is inseparable from God, but different from Him’. The point is that God would be meaningless if there were not that which God transcends: without opposition there is no life and no sense of development.

Wolfram Hogrebe has convincingly claimed that the *Weltalter* philosophy is a theory of predication (1989). In it, being is initially One, is not manifest and has no reason to be manifest: Hogrebe terms this ‘pronominal being’. The same being, given that there is now a manifest world, must also be ‘predicative being’ (*ibid.*), which ‘flows out, spreads, gives itself’. The contradiction is only apparent. Schelling maintains in line with the identity philosophy that the ‘properly understood law of contradiction really only says that the same cannot be *as the same* something and also the opposite thereof, but this does not prevent the same, which is A, being able, as an other, to be not A’. One aspect of being, the dark force, which he sometimes terms ‘gravity’, is contractive, the other expansive, which he terms ‘light’. Dynamic processes are the result of the interchange between these ultimately identical forces. If something is to be *as* something, it must both be, in the positive sense in which everything else is, which makes it indeterminately positive, and it must have a relationship to what it is not, in order to be determinate. In the *Weltalter* the One comes into contradiction with itself and the two forces constantly vie with each other. Differences must be grounded in unity, however, as otherwise they could not be manifest at all as differences. The ground is, though, increasingly regarded as the source of the transitory nature of everything particular, and less and less as the source of tranquil insight into how we can be reconciled to finite existence.

The abandonment of his residual Spinozism leads Schelling to a growing concern with the tensions which result from contradictions which we also embody. The ages of the world are constituted by the development of forms and structures in the material and the mental world. The development depends upon the expanding force's interaction with the contracting force's slowing of any expansion, which allows transient but determinate forms to develop. This process gives rise most notably to language, which Schelling sees as the model for the development of the whole world:

> It seems universal that every creature which cannot contain itself or draw itself together in its own fullness, draws itself together outside itself, whence, for example, the elevated miracle of the formation of the word in the mouth belongs, which is a true creation of the full inside when it can no longer remain in itself.
>
> (1946: 56-7)

Language as 'contracted' material signifier, and 'expanding' ideal meaning repeats the basic structure of the *Weltalter* philosophy. This interaction between what is contained in itself and what draws something beyond itself is also what gives rise to consciousness, and thus to an inherent tension within consciousness, which can only be itself by its relation to an other. Hegel uses a related model of subjectivity, particularly in the *Phenomenology*, but Schelling later rejects this model. Schelling's later philosophy will present a subject whose origin in nature prevents it from ever achieving the 'self-presence' Hegel thinks he can explicate via the completed structure of 'self-reflection' in the other. Schelling's *Weltalter* philosophy is never completed: its Idealist aim of systematically unifying subject and object by comprehending the real development of history from the very origins of being founders on problems concerning the relationship between philosophical system and historical contingency which do not admit of solutions.

4 Positive and negative philosophy, and the critique of Hegel (*c*.1827-54)

Schelling has usually been understood to provide the transitional 'objective idealist' link between Fichte and Hegel. By regarding Hegel's system as the culmination of German Idealism this interpretation fails to do justice to Schelling's real philosophical insights. Many of these insights, particularly in the later philosophy, directly and indirectly influenced the ideas of thinkers, such as Feuerbach, Kierkegaard, Nietzsche and Heidegger, who were critical of Hegel's attempt at a complete philosophical system (see Hegelianism §2).

The differences between Hegel and Schelling derive from their respective approaches to understanding the Absolute. For Hegel the Absolute is the *result* of the self-cancellation of the finite. It can therefore be presented in the form of the successive overcoming of finite determinations, the 'negation of the negation', in a system whose end comprehends its beginning. For Hegel the result becomes known when the beginning moves from being 'in itself' to being 'for itself' at the end of the system. Schelling became publicly critical of Hegel while working on a later version of the *Weltalter* philosophy in Erlangen in the 1820s, but made his criticisms fully public in lectures given in Munich in the 1830s, and in the 1840s and 1850s as professor in Berlin. The aim of the Idealist systems was for thought to reflect what it is not - being - as really itself, even as it appears not to be itself, thereby avoiding Kant's dualism. The issue between Schelling and Hegel is whether the grounding of reason by itself is not in fact a sort of philosophical narcissism, in which reason admires its reflection in being without being able to articulate its relationship to that reflection. Schelling's essential point is that it is not the particular manifestation of knowledge which tells me the truth about the world, but rather the necessity of movement from one piece of knowledge to the next. This much can be construed in Hegelian terms. However, a logical reconstruction of the process of knowledge can, for Schelling, only be a reflection of thought by itself: the real process cannot be described in philosophy, because the cognitive ground of knowledge and the real ground, though inseparable from each other, cannot be shown to *reflect* each other.

Dieter Henrich characterizes Hegel's Absolute as follows: 'The Absolute is the finite to the extent to which the finite is nothing at all but negative relation to itself' (1982). Hegel's system depends upon showing how each limited way of conceiving of the world cannot grasp the whole, because it has an internal contradiction. This necessarily leads thought to more comprehensive ways of grasping the world, until the point is reached where there can be no more comprehensive way, because there is no longer any contradiction to give rise to it. The very fact of the limitations of empirical thought therefore becomes what gives rise to the infinite, which, in Hegel's terms, is thought that is bounded by itself and by nothing else.

Schelling accepts such a conception, to which he substantially contributed in his early philosophy, as the way to construct a 'negative' system of philosophy: it explains the logic of change, once there is a world to be explained. It does not, though, explain why there is a developing world at all, but merely reconstructs in thought the necessary structure of development. Schelling's own attempt at explaining the world's facticity led him to a 'philosophical theology' which traces the development of mythology and then of Christian revelation in his *Philosophie der Mythologie (Philosophy of Mythology)* and *Philosophie der Offenbarung (Philosophy of Revelation)*, which like all his substantial works after 1811, were not published in his lifetime. The failure of his philosophical theology does not, though, invalidate his philosophical arguments against Hegel. The alternative to the 'common mistake of every philosophy that has existed up to now' - the 'merely logical relationship of God to the world' - Schelling terms 'positive philosophy'. The 'merely logical relationship' entails reflexivity, in which the world necessarily follows from the nature of God, and God and the world are therefore the 'other of themselves'. Hegel's system removes the facticity of the world by understanding reason as the world's immanent self-articulation. Schelling insists that human reason cannot explain its *own* existence, and therefore cannot encompass itself and its other within a system of philosophy. We cannot, he maintains, make sense of the manifest world by beginning with reason, but must begin with the contingency of being and try to make sense of it with the reason which is only one aspect of it.

Schelling contends that the identity of thought and being cannot be articulated *within* thought, because this must presuppose that they are identical in a way which thought, as one side of a relation, cannot comprehend. By redefining the 'concept' such that it is always already both subject and object, Hegel's aim is to avoid any presuppositions on either the subject or the object side, allowing the system to complete itself as the 'self-determination of the concept'. Schelling presents the basic alternative as follows:

> For either the concept would have to go first, and being would have to be the consequence of the concept, which would mean it was no longer absolute being; or the concept is the consequence of being, then we must begin with being without the concept.
>
> (1842-3, 2 (3): 164)

Hegel attempts to merge concept and being by making being part of a structure of self-reflection, rather than the ground of the interrelation of subject and object. He invalidly assumes that 'essence', which is one side of the relationship between being and essence, can articulate its identity with the other side in the 'concept', because the other side is revealed as being nothing until it has entered into a relation which makes it determinate as a moment of the whole process.

The problem that Hegel does not overcome is that this identity cannot be known, because, as Schelling argues of his concept of being, 'existing is not here the consequence of the concept or of essence, but rather existence is here itself the concept and itself the essence'. The problem of reflection cannot be overcome in Hegel's manner: identifying one's reflection in a mirror as oneself (understood now as a metaphor for essence) entails, as we saw above, a prior non-reflexive moment if one is to know that the reflection *is* oneself, rather than a random reflected object. How far Schelling moves from any reflexive version of identity philosophy is evident in the following from the *Einleitung in die Philosophie der Offenbarung oder Begründung der positiven Philosophie (Introduction to the Philosophy of Revelation or Foundation of the Positive Philosophy)*:

> Our self-consciousness is not at all the consciousness of that nature which has passed through everything, it is precisely just *our* consciousness…for the consciousness of man is not = the consciousness of nature… Far from man and his activity making the world comprehensible, man himself is that which is most incomprehensible.
>
> (1842-3, 2 (3): 6-7)

Schelling refuses to allow that reason can confirm itself via its reflection in being:

> what we call the world, which is *so completely contingent* both as a whole and in its parts, cannot possibly be the impression of something which has arisen by the *necessity of reason*…it contains a *preponderant* mass of *unreason*.
>
> (1832-3: 99)

Schelling is, then, one of the first philosophers seriously to begin the destruction of the model of metaphysics

based on the idea of representation, a destruction which can be seen as one of the key aspects of modern philosophy from Heidegger to the later Wittgenstein and beyond. At the same time, he is committed, unlike some of his successors, to an account of human reason which does not assume that reason's incapacity to ground itself should lead to the abandonment of the question of truth.

See also: German idealism; Romanticism, German

ANDREW BOWIE

List of works

Schelling, F.W.J. (1856-61) *Friedrich Wilhelm Joseph Schelling's Sämmtliche Werke*, ed. K.F.A. Schelling, Division 1 (10 vols), Division 2 (4 vols), Stuttgart: Cotta.(An easily accessible, substantial selection of the complete works has been published as *Friedrich Wilhelm Joseph von Schelling, Ausgewählte Schriften*, ed. M. Frank, Frankfurt: Suhrkamp, 1985, 6 vols.)

Schelling, F.W.J. (1946) *Die Weltalter*, ed. M. Schröter, Munich: Biederstein.(This has other versions than the version from 1813 printed in the *Sämmtliche Werke*.)

Schelling, F.W.J. (1976-) *Historisch-kritische Ausgabe, im Auftrag der Schelling-Kommission der Bayerischen Akademie der Wissenschaften*, ed. H.M. Baumgartner, W.G. Jacobs and H. Krings, Stuttgart.(This is still a long way from completion, but will become the new standard edition.)

Schelling, F.W.J. (1794) *Über die Möglichkeit einer Form der Philosophie überhaupt*, trans. and with commentary by F. Marti, On the Possibility of an Absolute Form of Philosophy, in *The Unconditional in Human Knowledge: Four early essays 1794-6*, Lewisburg, PA: Bucknell University Press, 1980.(Argues the need for an 'absolute founding proposition' in philosophy in the light of Reinhold's attempt to establish the foundations Kant had failed ultimately to provide.)

Schelling, F.W.J. (1795a) *Vom Ich als Prinzip der Philosophie oder über das Unbedingte im menschlichen Wissen*, trans. and with commentary by F. Marti, Of the I as the Principle of Philosophy or on the Unconditional in Human Knowledge, in *The Unconditional in Human Knowledge: Four early essays 1794-6*, Lewisburg, PA: Bucknell University Press, 1980.(Attempts in a mainly Fichtean manner to establish the 'unconditioned' for philosophy in the wake of Kant.)

Schelling, F.W.J. (1795b) *Philosophische Briefe über Dogmatismus und Kriticismus*, trans. and with commentary by F. Marti, Philosophical Letters on Dogmatism and Criticism, in *The Unconditional in Human Knowledge: Four early essays 1794-6*, Lewisburg, PA: Bucknell University Press, 1980.(Schelling's attempt to reconcile a Spinozist and a Fichtean account of the Absolute.)

Schelling, F.W.J. (1796-7) *Abhandlungen zur Erläuterung des Idealismus der Wissenschaftslehre (Essays in Explanation of the Idealism of the Doctrine of Science)*, in *Friedrich Wilhelm Joseph Schelling's Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (1) 345-452.(Further exploration of Kantian and Fichtean questions concerning the foundation of philosophy.)

Schelling, F.W.J. (1797) *Ideen zu einer Philosophie der Natur als Einleitung in das Studium dieser Wissenschaft*, trans. E.E. Harris and P. Heath, with intro. by R. Stern, *Ideas for a Philosophy of Nature: Introduction to the Study of This Science*, Cambridge: Cambridge University Press, 1988.(First version of a system of *Naturphilosophie*, which begins to move away from Fichte.)

Schelling, F.W.J. (1799) *Erster Entwurf eines Systems der Naturphilosophie (First Plan of a System of the Philosophy of Nature)*, in *Friedrich Wilhelm Joseph Schelling's Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (3) 5-268.(Further attempt to develop a systematic *Naturphilosophie*.)

Schelling, F.W.J. (1800) *System des transcendentalen Idealismus*, trans. P. Heath, with intro. by M. Vater, *System of Transcendental Idealism* Charlottesville, VA: University Press of Virginia, 1978.(Remarkably internally consistent account of the history of self-consciousness, using a model which Hegel will adopt in the *Phenomenology*. Sees art as the 'organ of philosophy'.)

Schelling, F.W.J. (1801a) *Darstellung meines Systems der Philosophie (Presentation of My System of Philosophy)*, in *Friedrich Wilhelm Joseph Schelling's Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (4) 107-212.(First, clearly flawed, presentation of the 'identity philosophy'.)

Schelling, F.W.J. (1801b) *Über den wahren Begriff der Naturphilosophie und die richtige Art, ihre Probleme zu lösen (On the True Concept of the Philosophy of Nature and the Right Way to Solve its Problems)*, in *Friedrich Wilhelm Joseph Schelling's Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (4) 81-103.(The text in which Schelling clearly distances his philosophy from that of Fichte for the first time.)

Schelling, F.W.J. (1802) *Bruno oder über das göttliche und natürliche Prinzip der Dinge*, trans. and with intro. by M. Vater, *Bruno, or On the Natural and the Divine Principle of Things*, Albany, NY: State University of New

York Press, 1984.(Further explorations of questions of *Naturphilosophie* and ‘identity philosophy’.)
Schelling, F.W.J. (1802-3) *Philosophie der Kunst*, trans. *The Philosophy of Art*, Minneapolis, MN: Minnesota University Press, 1989.(The first work ever to carry this title: a systematic presentation of the philosophical significance of the differing forms of art. Particularly impressive in relation to music.)
Schelling, F.W.J. (1803) *Vorlesungen über die Methode des akademischen Studiums*, trans. E.S. Morgan, ed. N. Guterman, *On University Studies*, Athens, OH: Ohio University Press, 1966.(Texts concerned with the nature of university study which relate to the foundation of new German universities at this time.)
Schelling, F.W.J. (1804a) *System der gesammten Philosophie und der Naturphilosophie insbesondere (System of the Whole of Philosophy and the Philosophy of Nature in Particular)*, in *Friedrich Wilhelm Joseph Schelling’s Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (6) 133-577.(The most coherent and important version of the ‘identity philosophy’. The philosophical reflections are among Schelling’s most important: the account of *Naturphilosophie* is at times merely bizarre.)
Schelling, F.W.J. (1804b) *Philosophie und Religion (Philosophy and Religion)*, in *Friedrich Wilhelm Joseph Schelling’s Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (6) 13-70.(Text in which Schelling first makes the distinction between ‘negative’ and ‘positive’ philosophy. Also signals the beginning of the move away from the Spinozist aspects of the ‘identity philosophy’.)
Schelling, F.W.J. (1806a) *Darlegung des wahren Verhältnisses der Naturphilosophie zu der verbesserten Fichteschen Lehre (Explanation of the True Relationship of the Naturphilosophie to the Improved Fichtean Doctrine)*, in *Friedrich Wilhelm Joseph Schelling’s Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (7) 3-126. (Contains important formulations of the principles of identity philosophy.)
Schelling, F.W.J. (1806b) *Aphorismen zur Einleitung in die Naturphilosophie (Aphorisms as an Introduction to the Philosophy of Nature)*, in *Friedrich Wilhelm Joseph Schelling’s Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (7) 140-197.(Further important development of the mature identity philosophy.)
Schelling, F.W.J. (1807) *Über das Verhältnis der bildenden Künste zur Natur (On the Relationship of the Fine Arts to Nature)*, in *Friedrich Wilhelm Joseph Schelling’s Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (7) 291-329.(Further development of Schelling’s conception of art.)
Schelling, F.W.J. (1809) *Philosophische Untersuchungen über das Wesen der menschlichen Freiheit und die damit zusammenhängenden Gegenstände*, trans. and with critical notes by J. Gutmann, *Of Human Freedom*, Chicago, IL: Open Court, 1936.(The text which inaugurates a major change in Schelling’s thought, which is concerned with the question of evil and its relation to human freedom, and with God’s relationship to nature.)
Schelling, F.W.J. (1810) *Stuttgarter Privatvorlesungen (Stuttgart Private Lectures)*, in *Friedrich Wilhelm Joseph Schelling’s Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (7) 419-84.(Takes up some of the themes from *Of Human Freedom*, developing them in a more systematic manner.)
Schelling, F.W.J. (1811-15) . *Die Weltalter*, trans. and with intro. and notes by F. de W. Bolman, Jr, *The Ages of the World*, New York: Columbia University Press, 1967.(Schelling’s attempt to give a metaphysical account of why the world becomes intelligible at all, and how the conflicts in reality can be grasped within philosophy. Contains vital remarks on the question of predication and identity.)
Schelling, F.W.J. (1820-1) *Initia Philosophiae Universae*, ed. H. Fuhrmans, Bonn: Bouvier, 1969. (Important text which begins Schelling’s move away from Idealist philosophy. Contains key account of the problem of ‘reflection’, which paves the way for the critique of Hegel. Contains *Ueber die Nature der Philosophie als Wissenschaft*.)
Schelling, F.W.J. (1821) *Ueber die Nature der Philosophie als Wissenschaft (On the Nature of Philosophy as a Science)*, in *Friedrich Wilhelm Joseph Schelling’s Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 1 (9) 209-46.(Also contained in *Initia Philosophiae Universae*.)
Schelling, F.W.J. (1827-8) *System der Weltalter (System of the Ages of the World)*, ed. S. Peetz, Frankfurt: Klostermann, 1990.(Development of *Die Weltalter*. Contains one of the earlier attacks on Hegel’s system.)
Schelling, F.W.J. (1830) *Einleitung in die Philosophie (Introduction to Philosophy)*, ed. W.E. Ehrhardt (Schellingiana, vol. 11), Stuttgart: Frommann-Holzboog, 1989.(Further evidence of Schelling’s move away from the Idealist paradigm towards the ‘positive philosophy’.)
Schelling, F.W.J. (1832-3) *Grundlegung der positiven Philosophie (Foundations of the Positive Philosophy)*, ed. H. Fuhrmans, Turin: Bottega d’Erasmo, 1972.(First extant systematic account of the ‘positive philosophy’, contains vital remarks against Hegel’s account of reason.)
Schelling, F.W.J. (*c*.1833-4) *Zur Geschichte der neueren Philosophie*, trans. and with intro. by A. Bowie, *On the History of Modern Philosophy*, Cambridge: Cambridge University Press, 1994.(Uniquely valuable account of the history of philosophy from Descartes to Schelling himself, covering Spinoza, Leibniz, Wolff, Kant, Fichte, Jacobi, and containing the most extensive extant critique of Hegel. A text which anticipates many of the major

themes in European philosophy after Schelling.)

Schelling, F.W.J. (1841-2) *Philosophie der Offenbarung (Philosophy of Revelation)*, ed. M. Frank, Frankfurt: Suhrkamp, 1977.(Illegally transcribed, but clearly accurate outline of the whole of the *Philosophy of Revelation*. Contains vital remarks on Hegel and positive philosophy. Excellent introduction.)

Schelling, F.W.J. (1842) *Philosophie der Mythologie (Philosophy of Mythology)*, in *Friedrich Wilhelm Joseph Schelling's Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 2 (2) 3-674.(Remarkable historical account of the emergence and development of mythology: contains important methodological reflections on the study of mythology that influenced many subsequent thinkers.)

Schelling, F.W.J. (1842-3) *Philosophie der Offenbarung (Philosophy of Revelation)*, in *Friedrich Wilhelm Joseph Schelling's Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 2 (3) 3-530.(The official version of *Philosophy of Revelation*, compiled by Schelling's son. The first ten lectures contain some of Schelling's most brilliant critical remarks on modern philosophy.)

Schelling, F.W.J. (between 1847 and 1852) *Philosophische Einleitung in die Philosophie der Mythologie oder Darstellung der reinrationalen Philosophie (Philosophical Introduction to the Philosophy of Mythology or Presentation of the Purely Rational Philosophy)*, in *Friedrich Wilhelm Joseph Schelling's Sämmtliche Werke*, ed. K.F.A. Schelling, Stuttgart: Cotta, 2 (1) 255-572.(Important further reflections on negative and positive philosophy. Contains some of Schelling's most significant remarks on ontology, particularly in relation to Aristotle and the question of being and non-being.)

References and further reading

Bowie, A. (1990) *Aesthetics and Subjectivity: from Kant to Nietzsche*, Manchester: Manchester University Press. (Contains a chapter on Schelling which characterizes him in relation to Hölderlin and to Romantic and post-Romantic theories of aesthetics, and as a theorist of subjectivity who does not rely on the idea of self-presence.)

Bowie, A. (1993) *Schelling and Modern European Philosophy: An Introduction*, London: Routledge.(The first full-length account of Schelling in English to consider him as a major philosopher in his own right, rather than as a pendant to Hegel. Connects Schelling to issues in contemporary analytical and European philosophy.)

Fichte, J.G. (1971) *Werke (Works)*, Berlin: de Gruyter, 1, 1.(The standard edition of Fichte's writings; referred to in §1.)

Frank, M. (1975) *Der unendliche Mangel an Sein (The Unending Lack of Being)*, Frankfurt: Suhrkamp.(The classic modern account of Schelling's critique of Hegel: a dense, very difficult, but indispensable work.)

Frank, M. (1985) *Eine Einführung in Schellings Philosophie (An Introduction to Schelling's Philosophy)*, Frankfurt: Suhrkamp.(A detailed account of Schelling's early work until the end of the identity philosophy: see §2.)

Frank, M. (1991) *Selbstbewußtsein und Selbsterkenntnis (Self-Consciousness and Self-Knowledge)*, Stuttgart: Reclam.(Contains a vital essay on Schelling's identity theory, 'Identität und Subjektivität' (Identity and Subjectivity), which sees the theory as a major event in Western philosophy.)

Frank, M. and Kurz, G. (eds) (1975) *Materialien zu Schellings philosophischen Anfängen (Material on Schelling's Philosophical Beginnings)*, Frankfurt: Suhrkamp. (Essays on various aspects of Schelling's philosophy between 1795 and 1804, with accompanying historical material.)

Heidegger, M. (1971) *Schellings Abhandlung über das Wesen der menschlichen Freiheit (Schelling's Essay on the Essence of Human Freedom)*, Tübingen: Niemeyer. (Dense and difficult, but essential commentary on Schelling's *On the Essence of Human Freedom*, with material from later lectures by Heidegger. Referred to in §3.)

Henrich, D. (1982) *Selbstverhältnisse (Self-Relationships)*, Stuttgart: Reclam. (Important essays on Schelling, Hegel and modern philosophy.)

Heuser-Kessler, M.-L. (1986) *'Die Produktivität der Natur' Schellings Naturphilosophie und das neue Paradigma der Selbstorganisation in den Naturwissenschaften (The 'Productivity of Nature', Schelling's Naturphilosophie and the New Paradigm of Self-Organization in the Natural Sciences)*, Berlin: de Gruyter. (Claims that Schelling's philosophy of nature can be linked to developments in non-linear dynamics and to the theory of self-organizing systems.)

Hogrebe, W. (1989) *Prädikation und Genesis. Metaphysik als Fundamentalheuristik im Ausgang von Schellings 'Die Weltalter'(Predication and Genesis. Metaphysics as Fundamental Heuristics Beginning with Schelling's 'The Ages of the World')*, Frankfurt: Suhrkamp.(A brilliant, but demanding account of the *Weltalter* as a theory of predication, which uses the tools of analytical philosophy to show how consistent much of Schelling's position is.)

Jähnig, D. (1966, 1969) *Schelling, Die Kunst in der Philosophie (Schelling: Art in Philosophy)*, Pfullingen: Neske, 2 vols.(Detailed and impressive account of the importance of art for Schelling's philosophy as a whole.)

Jaspers, K. (1955) *Schelling: Größe und Verhängnis (Schelling: Greatness and Disaster)*, Munich: Piper.(An interesting, if outdated, account of Schelling's life and work, which sees Schelling as failing to achieve his philosophical goals.)

Marx, W. (1984) *The Philosophy of F.W.J. Schelling: History, System, Freedom*, Bloomington, IN: Indiana University Press.(General and fairly accessible account, mainly of earlier work by Schelling, as far as *On the Essence of Human Freedom*.)

Sandkaulen-Bock, B. (1990) *Ausgang vom Unbedingten. Über den Anfang in der Philosophie Schellings (Going Out from the Absolute: On the Beginning in the Philosophy of Schelling)*, Göttingen: Vandenhoeck & Ruprecht.(Excellent account of Schelling's response to questions posed in particular by Jacobi concerning the grounding of philosophy in the Absolute: historically detailed and very thorough on the early work.)

Sandkühler, H.J. (1970) *Friedrich Wilhelm Joseph Schelling*, Stuttgart: Metzler. (Contains bibliography, which complements that of Schneeberger (1954).)

Sandkühler, H.J. (ed.) (1984) *Natur und geschichtlicher Prozeß (Nature and Historical Process)*, Frankfurt: Suhrkamp.(Selection of essays on the philosophy of nature with useful bibliography of writings on that philosophy.)

Schneeberger, G. (1954) *Friedrich Wilhelm Joseph von Schelling. Eine Bibliographie*, Bern: Franke.(The standard bibliography, to be complemented by those cited above.)

Scholz, H. (ed.) (1916) *Die Hauptschriften zum Pantheismusstreit zwischen Jacobi und Mendelssohn (The Main Texts in the Pantheism Controversy Between Jacobi and Mendelssohn)*, Berlin: Reuther & Reichard.(Contains most of the key texts by Jacobi in the Pantheism controversy.)

Schulz, W. (1975) *Die Vollendung des deutschen Idealismus in der Spätphilosophie Schellings (The Completion of German Idealism in the Late Philosophy of Schelling)*, Pfullingen: Neske.(The book which reoriented the study of Schelling after the Second World War towards the study of the later work, particularly the critique of Hegel, and linked Schelling to Kierkegaard and Heidegger. Difficult but thought-provoking.)

Snow, D.E. (1996) *Schelling and the End of Idealism*, Albany, NY: State University of New York Press.(Lucid historical and philosophical introduction to Schelling's work from §§1-3. Excellent on the relationship to Jacobi, but does not deal with the Schelling of 'positive philosophy' and the critique of Hegel, §4.)

Tilliette, X. (1970) *Schelling une philosophie en devenir (Schelling: A Philosophy in Development)*, Paris: Vrin, 2 vols.(Encyclopedic historical account of the development of Schelling's work: stronger on general exposition and on theology than on Schelling's philosophical arguments.)

Tilliette, X. (1983) *Schelling: Introduction to the System of Freedom*, New Haven, CT, and London: Yale University Press.(Good introduction to Schelling's work as a whole, which tends to focus, though, on its undoubted weaknesses, at the expense of its strengths.)

Zizek, S. (1996) *The Indivisible Remainder. An Essay on Schelling and Related Matters*, London and New York: Verso.(Very wide-ranging, fascinating, but dense and often difficult reflections on the implications of Schelling's relationship to Jacques Lacan and Hegel. Tends to underestimate the radical differences between Schelling and Hegel, but gives a very good sense of why Schelling matters to contemporary philosophy.)

Schellingianism

Schelling's philosophy, spread by German professors teaching at Russian universities and by Russians who had studied in Germany, some with Schelling himself, had an early and lasting influence in Russia. It was greatest in aesthetic theory and in the philosophy of history, but it was noticeable even in the natural sciences. Schelling appealed especially to Russians who were inclined to reconcile a modern scientific worldview with religious faith or with an exalted view of art. Particularly prominent figures were Odoevskii, Belinskii, Chaadaev, Venevitinov and Solov'ëv.

1 Schellingianism at Russian universities

Danilo Vellanskii (1774-1842), who had studied with Schelling in Germany, was a professor of the natural sciences at St Petersburg University from 1805 and published several works on medicine, biology and physics in which he faithfully followed Schelling's *Naturphilosophie* (see Naturphilosophie §2). Mikhail Pavlov (1793-1840), another student of Schelling's, was appointed professor of physics and agronomy at Moscow University in 1820. His *Outline of Physics* (1825-36) is clearly under the sway of Schelling's conception of nature. Aleksandr Galich (1783-1848), who had studied in Germany, was made a professor of philosophy at St Petersburg University in 1819. His 'Opty nauki iziashchnogo' (Essay on a Science of the Beautiful) (1825) accurately presents Schelling's aesthetics. Nikolai Nadezhdin (1804-56) won an appointment as professor of fine arts and archaeology at Moscow University in 1831, on the strength of a dissertation (in Latin) on Romantic poetry. Nadezhdin knew Schelling well and used his ideas in developing his own aesthetic theory, which features an organic conception of the social, national and historical aspects of art. The list of professors known to have taught Schellingian ideas may be readily expanded.

The student circle 'Lovers of Wisdom' in Moscow (1823-5), a secret discussion group, was led by Prince Vladimir Odoevskii (1804-69), president, and Dmitrii Venevitinov (1805-27), secretary. It had perhaps a dozen members and sympathizers, among them the Kireevskii brothers, Ivan (1806-56) and Pëtr (1808-56), both students of Schelling, Aleksei Khomiakov (1804-60), Mikhail Pogodin (1800-75) and Stepan Shevyryov (1806-64), all of whom would later become prominent Slavophiles (see Slavophilism). The circle was a hotbed of Schellingian ideas, which were launched in the publications of its members.

Odoevskii's *Russkie nochi (Russian Nights)* (published 1844, but written largely in the 1820s) has a group of young Russians gather to read their stories and discuss their ideas. Schelling's philosophy is explicity identified as the focus of these discussions. Faust, the discussion leader, presents it as a synthesis of all sciences, an inspired vision of the universe as a living organism in goal-directed flux. The power of a great artist (Bach is the case in point) is seen as in tune with the very rhythm of the cosmos. The idea that the ultimate truth, though understood intuitively, eludes rational expression, also appears as a leitmotif in *Russian Nights*. Venevitinov's essays 'Sculpture, Painting, and Music' (1827) and 'Anaxagoras: A Platonic Dialogue' (1830) are poeticized versions of Schellingian conceptions, the former being an allegoric exposition of Schelling's hierarchy of art forms, the second a poetically stylized statement of Schelling's idea that the human spirit passes from a Golden Age of instinct into freedom, as it enters history, and will return to a second Golden Age when all self-will is eliminated and pure reason reigns.

The poet Fëdor Tiutchev (1803-73) might have been among the Lovers of Wisdom, had he not joined the Russian foreign service in 1822. Like them, he later became a Slavophile. While stationed in Munich, he met Schelling and became intimately acquainted with his philosophy. A basic theme of Schelling's thought is dominant in Tiutchev's poetry: nature is an organism possessed of spirit - a cosmos; but when intuition penetrates the surface of this cosmos, it faces chaos and senses man's affinity to it, perceiving the chaos of night as the mother of all things.

A circle led by Nikolai Stankevich (1813-40) at Moscow University in the early 1830s became yet another hotbed of Schellingianism. Among its members and sympathizers were Vissarion Belinskii (1811-48), Mikhail Bakunin (1814-76), Timofei Granovskii (1813-55) and Konstantin Aksakov (1817-60), to mention the most prominent. Stankevich and several members of his circle abandoned Schelling for Hegel around 1838 (see Hegelianism, Russian).

2 Schellingian aesthetics

Schelling's organic aesthetics dominated Russian literary criticism mainly as a result of Belinskii's pervasive influence. Belinskii, a student of Nadezhdin's, turned Hegelian after 1838, but the principles of Schelling's romantic aesthetics were retained by some of his followers, particularly Apollon Grigor'ev (1822-64), who in turn influenced Dostoevskii and even the Symbolists at the turn of the century. These principles were: true art is not *mimesis* but a creative extension of nature; it is inherently symbolic and has cognitive power, expressing the universal through the particular, the ideal through the real, and the eternal through the temporal; it is free, yet it necessarily expresses the truth of life; it is in phase with history, creating the mythology by which nations and societies live. Grigor'ev, who called his method 'organic criticism', defined historical sense as 'an understanding of the organic nature of history, a recognition of eternal, absolute ideals, the presence of which gives meaning to history' ([1858] 1876: 223). He denounced the Hegelian 'historical school' for seeing only the progress of history, rather than the absolute and lasting values created by each stage of the historical process.

3 Schellingian philosophies of history

While the Left of Russian thought followed Hegel and the Left Hegelians once they had become known in Russia, the Right stayed with a Schellingian philosophy of history, essentially in an effort to retain religion as a foundation of national life. This is true of Pëtr Chaadaev (1794-1856), the first prominent Westernizer. Chaadaev met Schelling in Germany and corresponded with him. He was the first Russian to believe that humankind would enter the millennium through the gates of history, seen as a movement generated by a striving for unity, perfection and universality. Unlike Schelling, he saw Christian religion not as a mere stage of this process, but as its sole generating power.

Slavophiles, specifically Khomiakov and Ivan Kireevskii, and the *pochvenniki* (from *pochva*, 'soil', Schelling's *Boden*) Grigor'ev and Dostoevskii shared Schelling's organic conception of nationhood and saw the historical process in terms of revealed absolute ethical values undergoing temporal metamorphoses through a dialectic of resolved contradictions. The Slavophiles, in particular, stressed the primacy of revealed religion: in a dialectic of material and spiritual values, the burden was on science to raise itself to the level of religion, not vice versa.

Khomiakov perceived history as a dialectic process whose opposite poles are 'Iranian' humanity, guided by ideal (spiritual and aesthetic) values and its 'Kushite' antipodes, guided by material aspirations of power and possessions. Christianity, the culmination of Iranian humanity, amounts to the realization of the idea that man is so godlike - in a spiritual sense - that God could become man: we have a capacity for infinite perfection, which gives us infinite freedom.

Lev Shestov said of Vladimir Solov'ëv (1853-1900), Russia's leading academic philosopher, that 'Schelling so permeated him that he apparently lost his ability to distinguish himself from him' (1964: 38). In particular, it was Schelling's *Philosophy of Revelation* that appealed to Solov'ëv, who also followed Schelling in incorporating Jakob Boehme's mystic doctrines into his system. Solov'ëv's first work was pure Schelling: *'The Mythological Process in the Paganism of Antiquity'* (1873) sees history as a mythmaking and theogonic process. Solov'ëv's conception of history is that of a complex, living, personal God, realizing himself through a process of overcoming what is alien to him: chaos, disorder, evil. In his *Opravdanie dobra (Justification of the Good)* (1897), Solov'ëv followed Schelling in dealing with the problem of evil: God denies evil as final or everlasting, but allows it to exist as a passing condition of freedom. The nature of man is one of divine potential. Once the process of divine revelation is completed, evil will become unreal. Solov'ëv also followed Schelling in presenting his historical insights in terms of dialectic potencies. In an essay of 1877, 'Three Powers', the inhuman God of the East is countered by godless man of Western capitalist society, with a synthesis of godly man produced by a union of East and West.

Schellingian ideas appear massively in the writings of Solov'ëv's followers, such as Sergei Trubetskoi (1862-1905), Viacheslav Ivanov (1866-1949), and Nikolai Berdiaev (1874-1948).

VICTOR TERRAS

References and further reading

Grigor'ev, A. (1858) 'Kriticheskii vzgliad na osnovy, znachenie i priemy sovremennoi kritiki iskusstva' (A

Critical View of the Foundations, Importance and Devices of Contemporary Art Criticism), in his *Sochineniia (Works)*, 1876.(Referred to in §2.)

[Grigoryev] (1862-4) *Moi literaturnye i nravstvennye skital'chestva*; trans. R.E. Matlaw, *My Literary and Moral Wanderings and Other Autobiographical Material*, New York: E.P. Dutton, 1962.

Lehmann, J. (1975) *Der Einfluss der Philosophie des deutschen Idealismus in der russischen Literaturkritik des 19. Jahrunderts: Die 'organische Kritik' Apollon A. Grigor'evs (The Influence of the Philosophy of German Idealism in Russian Literary Criticism of the Nineteenth Century: the Organic Criticism of Apollon A. Grigor'ev)*, Heidelberg: Carl Winter. (Covers Schelling in great detail.)

Leighton, L.G. (1975) 'D.V. Venevitinov: Schellingism and Philosophy; the Poet's Own Feelings', in *Russian Romanticism: Two Essays*, The Hague and Paris: Mouton, 79-84.

Odoevskii, Prince V. [Odoevsky] (1844) *Russkie nochi*; trans. O. Koshansky-Olienikov and R.E. Matlaw, *Russian Nights*, New York: E.P. Dutton, 1965.(Written in the 1820s. Referred to in §1.)

Setschkareff, W. (1939) *Schellings Einfluss in der russischen Literatur der 20er und 30er Jahre des XIX. Jahrhunderts (The Influence of Schelling in Russian Literature of the 1820s and 1830s)*, Leipzig: Harrassowitz; repr. Nendeln/Liechtenstein: Kraus, 1968.(Covers Schelling's influence in all areas (literature, philosophy, the sciences) up to the end of the 1830s.)

Shestov, L. (1964) 'Umozrenie i apokalipsis: Religioznaya filosofiya Vl. Solov'eva', in *Umozrenie i otkrovenie (Speculation and Revelation)*, Paris: YMCA-Press, 23-91.(Written in 1927, this examines Solov'ëv's efforts, futile in Shestov's opinion, to prop up revelation with ratiocination. Points out Solov'ëv's dependence on Schelling.)

Solov'ëv, V. [Solovyov] (1952) *A Solovyov Anthology*, arranged by S. Frank, trans. N. Duddington, New York: Scribner.(Compiled from a broad range of Solov'ëv's works.)

Venevitinov, D.V. (1826) 'O sostoianii prosveshcheniia v Rossii'; trans. and ed. L.G. Leighton, 'On the State of Enlightenment in Russia', in *Russian Romantic Criticism: An Anthology*, New York and London: Greenwood Press, 1987, 111-18.

Schiller, Ferdinand Canning Scott (1864-1937)

F.C.S. Schiller was the outstanding exponent of pragmatism in Britain. His views, which he referred to at various times as humanism, voluntarism and personalism, as well as pragmatism, were strongly influenced by William James, to whom he paid great tribute, although he claimed to have arrived at his opinions independently. Schiller pursued the subjective and personal aspects of James's psychology, whereas Dewey built on its objective and social elements. In taking the process of knowing as central to reality, Schiller was also influenced by Hegel. Schiller's philosophy may be best approached in terms of his opposition to the absolute idealism of the then-dominant British Hegelians (particularly F.H. Bradley, his bête noire*); Schiller thought their monism, rationalism, authoritarianism and intellectualism denied the basic insight of Protagoras that it is man who is the measure of all things.*

Schiller was born in Schleswig-Holstein and studied at Oxford, where he gained an MA. In 1893 he went to Cornell as a graduate student and instructor. He left there without a doctorate, however, to go to Corpus Christi College, Oxford, where he became senior tutor and fellow. He received the D.Sc. degree in 1906, as well as other academic honours. In 1926 he began to spend part of each year at the University of Southern California, and moved there permanently in 1935 as professor.

Schiller argued that all acts and thoughts are irreducibly the products of human beings, and are therefore inescapably associated with the needs, desires and purposes of the individual actor or thinker. Such concepts as 'reality' and 'truth' do not denote independent and absolute entities; they are permeated by human intentions and activities. The absolute idealists maintained that 'reality' is a seamless logical unity in which all separateness vanishes, and that nothing finite, transitory or changeable is quite real. But, Schiller pointed out, all that exists for us are the bits of matter we encounter, the individual acts we perform and the private thoughts we think. Reality for us is indeed piecemeal, incomplete and plastic. The reality revealed by our active inquiries is not rigid but malleable, not complete but evolving, responsive to our probing and manipulating, and therefore to our needs and purposes. What we call 'real' is what we evaluate as important. It is the result of the kind of activity by which we reduce the chaos about us to order. To be sure, there are patent limits to human powers, and the world obviously preceded our existence; Schiller later reluctantly accepted the distinction between 'making' the real and 'finding' it, but he reiterated the meaninglessness of the 'real-as-it-is-in-itself'. He used the Greek term '*hyle*' to refer to the indeterminate formless chaos beyond our ability to perceive or manipulate. His metaphysics can thus accommodate Darwinian evolution and the emergence of genuine novelty (always a problem for the absolute idealist). It is a firm foundation for human freedom, and legitimizes human progress. Metaphysical systems reflect personality and temperament, and are thus quasi-ethical, or even aesthetic, in character.

Schiller's doctrine of truth likewise raised British philosophical hackles. The truth of a proposition can be determined only by what follows from it in the court of experience. Truth is a valuation applied as the result of a procedure called verifying, or making true. Since no verifying can establish the absolute truth of a statement, Schiller held that truth is particular, personal and progressive. That which furthers the purpose of our inquiries we call true. 'Truth is that manipulation of [objects] which turns out upon trial to be useful, primarily for any human end, but ultimately for that perfect harmony of our whole life which forms our final aspiration'. Of course it is fallacious to infer 'the useful is true' from 'the true is useful' - all sorts of statements (political propaganda, advertising, outright lies, for example) can be useful, though clearly false. Schiller distinguished seven levels of truth claims. Thus a postulate is a statement which becomes an axiom if it is fully verified, 'serves as principle for a fully established science' and 'rests securely on the solid mass of scientific fact it has been instrumental in eliciting'. There are also methodological assumptions (for instance, determinism) and methodological fictions (for instance, the use of plane geometry in cartography).

Schiller accused traditional formal logic of having 'etherealized' and 'depersonalized' truth in its search for validity. It had become a word game, with no concern for actual thinking processes, or for meaning and context. In two books, *Formal Logic* (1912) and *Logic for Use* (1929), Schiller showed that formal logic, even on its own terms, was not free from ambiguity - how can there be anything in the conclusion of a syllogism not contained in its premises? What is the precise import of the copula in a proposition? How can logic consistently appeal to such

psychological notions as the 'necessity' of implication or the 'self-evidence' of axioms? Schiller urged that logic should become a systematic evaluation of actual knowing. His resolute experimentalism led him to claim in 'Axioms as Postulates' (1902) that even the basic Aristotelian principles of identity, contradiction and excluded middle were postulates. He also argued that the 'facts' of the scientist are relative to our senses, memory, language, instruments, aims and hypotheses.

Schiller carried his pragmatism into ethics, social policy (he was a great advocate of eugenics) and religion. He shared with William James and Henri Bergson an interest in psychical research, and translated James's 'will to believe' into the 'right to postulate'. Schiller was a prolific writer, a sprightly stylist and a spirited polemicist; he rather enjoyed being an *enfant terrible*. He produced a parody of that stately establishment journal *Mind* (which he called *Mind!*), an example of that scarce commodity, philosophical humour.

See also: Personalism; Pragmatism; Truth, pragmatic theory of §2

REUBEN ABEL

List of works

Schiller, F.C.S. (1891) *Riddles of the Sphinx: A Study in the Philosophy of Evolution*, London: Swan Sonnenschein and New York: Macmillan; new and revised edn, subtitled *A Study in the Philosophy of Humanism*, London: Macmillan, 1910.(A traditional discussion of traditional philosophical problems, written before Schiller saw the light of pragmatism; not very useful now. First published under the pseudonym 'A. Troglodyte'.)

Schiller, F.C.S. (1902) 'Axioms as Postulates' in H. Sturt (ed.) *Personal Idealism*, London: Macmillan, 47-133.(A pragmatic and fertile approach to the theory of knowledge.)

Schiller, F.C.S. (1903) *Humanism: Philosophical Essays*, London and New York: Macmillan, 2nd edn, 1912. (Fifteen essays on various subjects, mostly based on or reprinted from lectures or papers in learned or philosophical journals.)

Schiller, F.C.S. (1907) *Studies in Humanism*, London and New York: Macmillan.(Twenty essays on various subjects. Schiller preferred this form to writing a systematic single expository or controversial volume.)

Schiller, F.C.S. (1908) *Plato or Protagoras?*, Oxford: Blackwell and London: Simpkins, Marshall. (Subtitled, 'A Critical Examination of the Protagoras Speech in the *Theaetetus*, with some remarks upon Error'.)

Schiller, F.C.S. (1912) *Formal Logic: A Scientific and Social Problem*, London: Macmillan. (Schiller's basic contention: that you cannot consider the abstract 'forms of thought' apart from their content without losing sight of the fundamental issues of meaning and truth.)

Schiller, F.C.S. (1917) 'Scientific Discovery and Logical Proof', in C.J. Singer (ed.) *Studies in the History and Methods of Science*, Oxford: Clarendon Press, vol. 1, 235-89. (Schiller argues that logic has treated scientific discoveries as illustrations of a preconceived ideal of proof, and thus misunderstands the problem of science.)

Schiller, F.C.S. (1921) 'Hypothesis', in C.J. Singer (ed.) *Studies in the History and Methods of Science*, Oxford: Clarendon Press, vol. 2, 414-46.(Schiller argues for the inter-relationship of 'fact' and 'hypothesis': a good hypothesis does not conform to abstract prior rules, but is made good and knocked into shape by the procedure of verification.)

Schiller, F.C.S. (1924a) *Problems of Belief*, London: Hodder & Stoughton.(Schiller was interested in the problem of immortality and in the methods of science at its periphery, for instance, in psychical research. Here he argues that non-anthropomorphic thought is absurd.)

Schiller, F.C.S. (1924b) *Tantalus, or the Future of Man*, London: Kegan Paul and New York: Dutton.(Part of the 'Today and Tomorrow' series, this is written in a popular style. The survival of the human race is threatened by the progress of science, which helps the weak and least fit to survive.)

Schiller, F.C.S. (1926) *Eugenics and Politics*, London: Constable and Boston, MA: Houghton Mifflin.(Schiller was an advocate of eugenic improvement, before Hitler put it into practice.)

Schiller, F.C.S. (1929) *Logic for Use: an Introduction to the Voluntarist Theory of Knowledge*, London: G. Bell. (A constructive sequel to the destructive *Formal Logic*; here he argues that logic can and should be a theory of actual reasoning, not of empty and abstract forms.)

Schiller, F.C.S. (1934) *Must Philosophers Disagree?*, London: Macmillan.(Collection of essays in popular philosophy. In the title essay Schiller argues that philosophy should be concerned with human actions, not with 'word games'.)

Schiller, F.C.S. (1939) *Our Human Truths*, New York: Columbia University Press.(Twenty-seven essays on various subjects, selected and edited by his widow, Louise S. Schiller.)

References and further reading

Abel, R. (1955) *The Pragmatic Humanism of F.C.S. Schiller*, New York: King's Crown Press of Columbia University.(A critical study of Schiller's philosophy, with an exhaustive bibliography.)

Schiller, Johann Christoph Friedrich (1759-1805)

Schiller was an artist first - a major poet and the leading dramatist of eighteenth-century Germany - and an aesthetician second. At the height of his involvement in aesthetics, he calls the philosopher 'a caricature' beside 'the poet, the only true human being'. But reflection had deep roots in his nature, to the point where he felt it inhibited his creativity, yet would also have to be the means to restore it. He eventually came to terms with this paradox by devising a typology of 'naïve' and 'reflective' artists that explained his problem - and incidentally the evolution of modern European literature (On Naïve and Reflective Poetry, 1796). Schiller was also driven by a passionate belief in the humanizing and social function of art. His early speech The Effect of Theatre on the People (1784; later title The Stage considered as a Moral Institution) celebrated the one meeting-place where our full humanity could be restored. In the mature essays of the 1790s, an immensely more complex argument cannot hide the ultimate simplicity of his faith in art, even and especially in the midst of historical crisis: his culminating statement on beauty, On the Aesthetic Education of Man (1795) is at the same time a considered response to events in France, where a 'rational' Revolution had turned into a Reign of Terror. Schiller proposes an education for humane balance as the only sufficiently radical answer to the violent excesses of impulse, and argues that art is its only possible agent. Schiller's ideas are imaginative, generous and intuitively appealing as an account of what art is and might do. With the authority of his poetic standing and the high eloquence of his prose, they are powerful cultural criticism. Arguably they could have been more effective still and less vulnerable if he had not tried to make them something else by giving them a systematic quasi-Kantian form, as a result of which philosophical commentators have often patronized him while the Common Reader has been scared off.

1 Life

Schiller was born at Marbach in Württemberg. At fourteen he was placed in the Duke of Württemberg's military academy against the wishes of his parents, and made to study medicine against his own. Apart from an occasional autopsy and some observations of melancholia in a fellow-student, 'medicine' largely meant philosophical speculation on the workings of mind and body.Schiller was fascinated by the problem of mind-body interaction which was to remain central to his aesthetics. His 'medical' dissertations emphasize the effect and value of the physical component, true to the pattern of German eighteenth-century aesthetics which upgrades sense experience and alters the priorities of rationalism. Frustrated by a lowly post as regimental doctor and a ducal prohibition on writing any more drama after his sensational début *The Robbers* (1782),Schiller fled from Württemberg into the dubious freedom of a hand-to-mouth existence. There followed years of editing journals and writing pot-boiler fiction and popular history, necessary means of survival but obstacles to his higher ambition. In 1789 he was appointed Professor at the University of Jena, lecturing on history and aesthetics: a post that brought prestige but little pay. He only became relatively free of financial pressures in the very last years before his premature death. By then at least his fame was secure as poet, dramatist and critic and (from 1794 on) the literary partner and acknowledged equal of Johann Wolfgang Goethe in what came to be known as Weimar Classicism.

2 First essays: the sublime and tragedy

Schiller's first work in aesthetics proper is avowedly derivative from Kant's account of the sublime, except that Schiller takes over this standard eighteenth-century concept as the basis of his definition and practice of tragedy (see Kant, I.). Burke in 1757 had already suggested that sublime objects - towering crags, desert wastes and the like - 'exercise the finer parts of the system' in a virtually physiological way. By 1790 Kant is explaining the pleasures of sublime phenomena by an intricate psychological-cum-moral sequence in which their threat to overwhelm us forces us back on our distinctively human rational nature. By now, too, human actions such as principled self-sacrifice have crept in among the sublimities. The proximity to tragedy is obvious. Schiller locates tragedy in the necessary assertion of 'moral being' (*Sittlichkeit*) in the face of threats from or to the agent's 'sensuous being' (*Sinnlichkeit*). This may seem a reaffirmation of rationalist moral values, but for Schiller 'sensuous being' is not so readily abandoned as in earlier heroic and Christian drama. The pleasure of tragedy is not the simple one of moral triumph at the rejection of something worthless, but the poignant mixed feeling at the necessary sacrifice of something intensely valuable. In his last essay on the sublime (probably late 1790s), the tragic vision has been intensified by European wars, and the primacy of the stoic moral will is even more in the foreground. Yet tragedy still only treats the harsher aspect of life, is only one literary genre, and not for Schiller

the highest. This rank he gives to the genre he never worked in, comedy, because of its final harmony and balance.Schiller was the last writer to make significant use of the sublime in its traditional sense before it expired with the century, to be reassumed into a single larger conception of beauty (although the concept recurs with a different function in Hegel's *Aesthetics* and is resurrected in twentieth-century theories of 'Post-Modernism').

3 Beauty and aesthetic experience

On this central theme,Schiller once again starts from Kant, and after following some false trails goes well beyond him. In particular he does fuller justice to the complexity of both the aesthetic object and its reception. Both aspects bear out Schiller's aperçu that it was crucial for a theory of art that the thinker should be a practising artist. Kant's principle that aesthetic experience consisted in 'pleasure without interest' had set the object apart from practical purposes like use or desire, to be dwelt on for its own sake; while the effect of the pure form-percept lay in its power to seem nevertheless mysteriously purposive. This is conventionally seen as a philosophical breakthrough, but it only codified what contemporary writers were already perceiving and demanding in their practice, namely that the work of art and its constituent elements must be free of the constraints, moral, political or religious, placed on utterances in the practical world, and that it had formal laws of its own. Art was a distinct realm of contemplation and representation where the writ of authority did not, or at least should not, run - a recognizable off-shoot of the central Enlightenment principle of independent thought and free public communication. This is not of course to explain away in social terms the new insight that art is a means to hold external demands in suspense while life is inspected with a fresh eye and reshaped by the imagination. To all this, Schiller the writer was deeply committed.

For Kant, however, aesthetic judgment and the very attribution of aesthetic status were subjective, resting on an adjustment of the observer's mode of seeing. This leftSchiller dissatisfied. Was there nothing in the constitution of the object that positively invited such a mode of seeing - more compellingly than Kant's example of the arabesque, which inspired no 'interest' the observer would need to disregard and could thus, by a too simple logic, actually rank higher in beauty than the human form? An equally dubious logic had established Kant's next step. If observers all abstracted from their personal 'interests', then what was left must be common ground, for they had categorically lost the wherewithal to disagree. Ergo, aesthetic judgments were universal. But this was to reduce the beholder to a pure Nobody, emptying both subject and object of the reality that must surely be present - transformed, yes, but not purified out of existence - if art is to matter to anybody.

4 The *Kallias* Letters

Kant's limitations set Schiller's programme. In the never-completed project *Kallias*, sketched in 1793, he tried to establish an objective (but not empirically dependent) definition of beauty. He begins in Kantian spirit by calling it the 'form of a form', that is, a pure percept which is, as it were, permitted by the functional form of objects. (In line with much eighteenth-century aesthetics, he has natural objects in mind at least as much as artefacts.) But where for Kant it was the observer who disregarded purpose, for Schiller the object has itself 'overcome' its purpose, to appear free. Hence beauty is 'freedom in the phenomenon' or 'in appearance'. Though the German *Erscheinung* does not mean 'mere' or 'illusory appearance', Schiller cannot get round the fact that his freedom *is* merely in the eye of the beholder, or even just a metaphor for beauty's effect. Nor is there, incidentally, any clear reason why aesthetic pleasure should stem from the sight of freedom, other than that it offers 'an analogy with the form of the pure will', and that Schiller is avowedly one of those 'for whom freedom is the highest principle'. His fall-back position - that beauty comes about when nothing in the phenomenon too obviously *belies* the appearance of freedom - remains just as open to his friend Körner's objection that the whole theory rested on 'the autonomy which the observer mentally adds to the phenomenon', that is, it was irretrievably subjective.

5 *Grace and Dignity* (1793)

But instead of pursuing this central problem with a changed approach, Schiller keeps his flawed approach and changes the problem, to a secondary one which promised to fit it better. He defines grace as 'beauty in motion': motion at least was objective. Yet, lacking a satisfactory definition of the beauty to be observed in motion, he was (as he admitted) trying to fly before he could walk. 'Grace' moreover, when further unpacked, turned out to depend on the same criterion that had given trouble in *Kallias*: it was 'the beauty of the [human] form [*Gestalt*] under the influence of freedom'. Admittedly freedom in human beings, even if problematic in other ways, is not in

doubt as an illusion of the beholder. And it soon becomes clear thatSchiller is attempting an aesthetics not just of physical movement *per se*, but of total modes of behaviour in real situations. This might seem to cross the border from aesthetics into ethics altogether.Schiller distinguishes two basic responses to two equally basic types of situation. When human impulses are in harmony with each other and not under pressure from the surrounding world, the agent can freely choose how to act, and as an unconscious effect (Schiller constructs careful provisos against its conscious pursuit) grace may result. In contrast, when the impulses are at war with each other or under attack from the world, so that only a stoic defence is left, dignity may result. The first case shows freedom *in*, the second freedom *from* nature, where 'nature' is both the system of inner purposes that constitute our organic form, and the nexus of outer constraints that shape our lives.Schiller also posits a character type - the 'beautiful soul' or 'fine mind' (*schöne Seele*) - whose serene inner nature entails grace as a constant attribute. Yet even this ideal type cannot be immune to life's tribulations. In adversity, the harmony of mind and body has to be replaced by the freedom of mind *from* body, the 'sublime' quality of dignity. The links with Schiller's theory of tragedy are plain, as is the implication that the *schöne Seele* would be in its literary element as the hero or heroine of comedy or idyll.

The direct claims of morality that had been expelled from eighteenth-century aesthetics seem at first sight to have got in again. Schiller however is asserting a value higher than mere ethical outcomes. This emerges clearly in his discreet rebellion against the moral authority of Kant. If, as Kant had argued, an action can only be accounted moral when duty is kept sharply distinct from inclination, and indeed has normally had to overcome inclination, then there can be no such thing as an inherently moral being. Surely, Schiller argues, the ideal should be an *inclination to* duty? Only then would virtue be a value we embrace with our whole self, and no longer merely an external requirement that we obey in an endlessly repeated struggle with ourselves. Schiller concedes that Kant's rigorism may have been a necessary prescription for their age. But the highest moral norm should surely not be set by emergency; it must lie in the kind of primal spontaneity of action that defines the *schöne Seele*. Kant'sfootnote to the second edition of *Religion within the Bounds of Mere Reason*, acknowledging Schiller's 'masterly treatise', is a response but hardly an answer. Schiller, diplomatically, hastened to declare himself satisfied. Yet his conception remains a significant challenge toKant's axiom of pessimistic dualism. It is an appeal to balance spirit with sense, and moral with aesthetic judgment.

6 *Letters on the Aesthetic Education of Man* (1795)

Balancing opposed elements is the central idea of Schiller's major essay. It is the key both to understanding the nature of beauty and to creating a stable free society in some future age. These disparate aims make an unlikely pair, but to Schiller art seemed to offer precisely what the politics of the 1790s cried out for. Imbalance was the common condition in a world already suffering from division of labour. HereSchiller prepares the ground for Hegel and Marx on alienation, drawing a familiar eighteenth-century contrast with the wholeness of the human individual that was possible in Ancient Greece (L6). The added stress of revolutionary upheaval in France had now produced internecine violence. The French had acted with the ruthless rationality of 'barbarians' untouched by feeling, or the crude sensuousness of 'savages' unchecked by spirit (L4), an analysis that incidentally echoes the comments of observers in Paris. The project of an ideal society had failed not through the use of reason to guide reform, as conservatives like Burke held, but through its unmediated use by the unreformed. A force was needed that would smooth and stabilize the transition from sense to reason, exercising the human agent as both sensuous and rational being and strengthening or moderating each element as needed. That, according to Schiller's philosophical poem 'The Artists', was the role beauty had played in human evolution, and it could be repeated now. Art was the only force potentially free from the vicious circle of unregenerate humanity and the society that resulted. Art's highest exemplars fused substance and form and thus could balance and integrate sense and spirit in the beholder. The structural correspondence is clear. Less clear is how art might get a purchase on political reality. Schiller however is not offering a plaster for present ills, but a means to stop them recurring in the long term and so to create a new civic starting-point. There is no question of art's relapsing into didacticism to achieve this. Its effect for any specific purpose is avowedly 'nil'. Yet precisely by nullifying all previous determinations, its effect is 'infinite' because that restores human beings to themselves and sets them free once more to be - 'what they are meant to be' (L21). This is the still undaunted Enlightenment or Rousseauian faith that a new start will necessarily be a start in the right direction. The more pessimistic view, that our only realistic course is to practise a stoic rationality since life will always be crisis, is left for that tailpiece on the sublime (see §2 above) which in a sense

completes the programme of aesthetic education.

We are again at frontiers, with ethics, politics, social anthropology. To get there has required new solutions in aesthetics, and it was the pressure of an urgent problem in those areas that droveSchiller to find them. After sketching the crisis and arguing that art alone can help (L1-9), he has analysed human makeup into ‘person’ and ‘condition’, the core identity and its changing determinations (L10); and identified two matching drives, a sensuous impulse (*Stofftrieb*) towards realization and involvement in the material world, and a formal impulse (*Formtrieb*) towards moral and intellectual control (L12). These elements are only separable in the abstract; we actually experience them in their various interactions and imbalances. The ideal fulfilment of our dual nature would be to assimilate a maximum of reality but with a maximum of formal coherence. Such a state would bring into being a third drive, the play impulse (*Spieltrieb*), whose concrete object would be a fusion: the material world, or in its broadest sense ‘life’ (*Leben*), would join with formal structure or ‘shape’ (*Gestalt*) to produce ‘living shape’ (*lebende Gestalt*) (L15). More important surely for art’s role as a sociopolitical remedy, thoughSchiller does not say this, would be the reverse sequence: once aesthetic play was in some measure achieved, the two original impulses would be brought into some kind of balance. ‘Living shape’ meantime becomes Schiller’s definition of beauty, and ‘play’ his account of aesthetic experience. Although he proposes the aesthetic state as the long-term answer to a political problem, in the still longer term he intends it to be self-sufficient. It can be borrowed to meet a need, but is ultimately an ideal (L27). For ‘human beings only play when they are human in the fullest sense of the word, and they are only fully human when they play’ (L15). The language of the *Letters*, incidentally, enacts the complexities of opposition and reconciliation in a play of its own with a rich and at first confusing array of concepts, all of them however variations on an underlying pattern. Readers will find the game is better played fast for its form than slowly for its detailed content.

Leaving the political application aside, the *Letters* are a great advance on earlier aesthetics, bothSchiller’s own and Kant’s. No longer is an ‘objective’ theory insisted on, though there are vestiges of it in the original draft of Letter 1. Instead, both the constitution and the effect of beauty are grounded in the psycho-physical constitution and needs of human beings. Whereas Kantian ‘disinterestedness’ was a passive state that denied individuality, ‘play’ is an activity that subtly redeploys it. People bring to aesthetic encounters not an ‘empty indeterminateness’ but their whole past experience in a state of ‘active determinability’ (L20ff). This ensures, not a depersonalized universality of aesthetic judgment as in Kant, but a deeply personal response to both the forms and the substance of art. It restores reality to the abstractions of aesthetics: a landscape painting works with our accumulated experience of landscapes, a statue of Venus provokes more than an appreciation of assorted arabesques. The notion of ‘play’ means we take things seriously but not solemnly, in and for themselves, in an act of contemplation and savouring. In Brecht’s deceptively simple words, ‘In art, people enjoy life’. Such enjoyment involves separating ‘appearance’ - *Schein* this time, the standard word for illusion - from reality (L26). But it clearly no longer bothersSchiller that this makes his theory a subjective one. It has strong enough roots in the individual experience art springs from and appeals to, while ‘play’ is universal not as a logical abstraction but as an anthropological phenomenon.

7 *On Naïve and Reflective Poetry* (1796)

Schiller’s final large-scale essay is not so much aesthetic theory as literary and cultural history, with sections of brilliant practical criticism. In the wake of long-running European debates on the relative merits of Ancients and Moderns, it traces a movement from the Greeks’ oneness with the natural world to the hyperconsciousness of post-Christian Europe where, except for a few outstanding cases like Shakespeare, Molière and now Goethe, reflection necessarily intrudes between the poet and the object. Modern writing is constituted by reflection’s many forms, but their underlying constant, and hence the deepest shaping influence on European literature, is an elegiac sense of lost harmony. To restore that harmony and the perfection of artistic form that went with it in antiquity, while managing somehow still to retain the riches of the modern sensibility - a postmodernism of serious substance - is the near-impossible millennial goal. But Schiller was never short on aspiration.

See also: Aesthetics and ethics; Burke, E.; Cassirer, E.; Goethe, J.W. von; Sublime, the §§2, 3

T.J. REED

List of works

Schiller, J.C.F. (1943-) *Nationalausgabe*, ed. B. von Wiese, L. Blumenthal and N. Oellers, Weimar. (The standard edition of Schiller's works and letters. The philosophical writings are in vols 20-22. There is also a reliable complete edition in 5 vols, ed. H. Göpfert and G. Fricke, Munich, 1958 and reprints.)

Schiller, J.C.F. (1795) *On the Aesthetic Education of Man in a Series of Letters*, trans and with intro. by E.M. Wilkinson and L.A. Willoughby, Oxford, 1967; repr. in paperback 1982.(German text with facing English translation. Editorial apparatus includes a glossary of concepts, diagrammatic representations ofSchiller's thought, and an extensive bibliography.)

Schiller, J.C.F. (1796a) *On Naïve and Sentimental Poetry*, trans. H. Watanabe, Manchester: Carcanet Press, 1981. ('Sentimental' in the normal English sense is not Schiller's meaning, which is why I translate his word 'sentimentalisch' as 'reflective' in this entry.)

(1796b) *On Naïve and Sentimental Poetry*, with On the Sublime, trans. J.A. Elias, New York: Ungar, 1966.

Schiller, J.C.F. (1985) *German Aesthetic and Literary Criticism*, ed. H.B. Nisbet, Cambridge.(Reprints the Elias version in a collection of the most important eighteenth-century German aesthetic essays.)

Schiller, J.C.F. (1978) *Medicine, Psychology and Literature*, ed. K. Dewhurst and N. Reeves, Oxford. (Presents in full the young Schiller's medical dissertations, which contain the seeds of his mature thought, along with much editorial material on medicine in the period.)No reliable translation is available of Schiller's other essays.

References and further reading

Burke, E. (1757) *A Philosophical Enquiry into the Origin of our Ideas of the Sublime and Beautiful*, Oxford: Oxford University Press, 1990.(See especially Part 4 for Burke's views on the sublime.)

Cassirer, E. (1916) *Freiheit und Form*, Berlin.(A classic account of the intellectual tradition in which Schiller's thought has its roots.)

Ellis, J.M. (1969) *Schiller's Kalliasbriefe and the Study of his Aesthetic Theory*, The Hague: Mouton.(Acute arguments whose import extends well beyond the named text.)

Mann, T. (1932) 'Goethe and Tolstoy', in *Three Essays*, New York: Knopf.(Elaborates on Schiller's naïve/reflective typology, and applies it to the pair Tolstoy/Dostoevsky as well as to Goethe/Schiller.)

Podro, M. (1972) *The Manifold in Perception. Theories of Art from Kant to Hildebrand*, Oxford: Oxford University Press.(Schiller's theory forms the core of the book, which proceeds from a perceived affinity between painting since Impressionism and classical German aesthetics in which the author finds 'valuable critical ideas quite absent from contemporary philosophical writing'.)

Reed, T.J. (1991) *Schiller*, Past Masters series, Oxford: Oxford University Press.(A concise introduction with emphasis on the historical context and connections in the history of ideas.)

Reiner, H. (1953) *Pflicht und Neigung*, Meisenheim: Anton Hain.(A full and cogent treatment ofSchiller's differences withKant over duty and inclination.)

Schaper, E. (1979) *Studies in Kant's Aesthetics*, Edinburgh: Edinburgh University Press.(The chapter 'Schiller's Kant: a chapter in the history of creative misunderstanding' sees confusions in Schiller's reading of Kantian terminology, but is not dismissive: his early essays point up areas of difficulty in Kant's aesthetics.)

Sharpe, L. (1991) *Friedrich Schiller. Drama, Thought and Politics*, Cambridge: Cambridge University Press.(A substantial and coherent survey of all areas ofSchiller's work.)

Sharpe, L. (1995) *Schiller's Aesthetic Essays: Two Centuries of Criticism*, Columbia, SC: Camden House.(A wide-ranging survey of the literature in both English and German.)

Simmel, G. (1904) *Kant. Sechzehn Vorlesungen*, Leipzig: Duncker & Humblot.(Lecture 15 is a classic statement of the essentials ofKant's and Schiller's aesthetic thought.)

Wilkinson, E.M. (1960) 'Reflections on translating Schiller's *Letters on the Aesthetic Education of Man*', in *Schiller Bicentenary Lectures*, ed. F. Norman, London: Institute of Germanic Studies, 45-82.(Illuminates the relation of Schiller's discrete concepts to his underlying binary structures and attempted resolutions.)

Schlegel, Friedrich von (1772-1829)

Schlegel was the major aesthetician of the Romantic movement in Germany during its first formative period (1797-1802). In these years he developed his influential concepts of Romantic poetry and irony, created an original approach to literary criticism and edited the journal of the early Romantic circle, Athenäum. Along withF. von Hardenberg (Novalis),F.W.J. Schelling andF.D.E. Schleiermacher, he was also a guiding spirit in the development of a Romantic metaphysics, ethics and politics. His metaphysics attempted to synthesizeFichte's idealism and Spinoza's naturalism. His ethics preached radical individualism and love against the abstract formalism ofKant's ethics. In his early politics Schlegel was very radical, defending the right of revolution and democracy against Kant. In his later years, however, he became much more conservative. His final works are a defence of his neo-Catholic mysticism.

Schlegel began his intellectual career as a classical scholar. His *Über das Studium der Griechischen Poesie (On the Study of Greek Poetry)* (1797) attempted to write a history of Greek poetry and formulate a neoclassical aesthetics. Schlegel wanted to do for the history of Greek poetry what J.J. Winckelmann had done for the history of Greek sculpture. Like Winckelmann, Schlegel admired the ancient Greeks for attaining a purely 'objective' norm of beauty, which was independent of national taste, personal caprice or even moral and political ends. He criticized modern literature because, rather than observing such a norm, it pandered to national taste with novel and striking gimmicks. Schlegel believed that the artist should attempt to imitate the purity and simplicity of classical models; in the recent works of Goethe he saw promising signs of a new classicism in Europe.

However, as early as 1797, largely under the influence of Friedrich Schiller, Schlegel began to have doubts about his neoclassicism. He now started to appreciate some of the distinctive values of modern Christian culture, especially its ideal of the infinite, its ethic of love and its emphasis upon personal freedom. In his *Athenäums Fragmente* (1798) Schlegel then gave a condensed account of his new Romantic aesthetic. The aim of the Romantic artist is to express the characteristic feature of modern culture: the striving for the infinite, the longing for the kingdom of God on earth, the struggle to realize Kant's ideal of the highest good (see Kant, I. §11). What is distinctive of Romantic art in contrast to classical, Schlegel wrote in his *Gespräch über Poesie (Conversation on Poetry)* (1800a), is that it expresses sentiment, especially the feeling of love, which is the longing to realize the infinite. In attempting to express such a grand ideal, the Romantic artist should attempt to cultivate irony, a critical detachment towards his own productions, for any of them are limited and therefore inadequate to express his unlimited ideal.

In several essays of the late 1790s, 'Über Lessing', 'Georg Forster' and 'Jacobis Woldemar', Schlegel developed a new method of literary criticism, which he described as 'characterization' (*Charackteristik*). The aim of this method was to understand a work as a unique whole, to reconstruct an author's characteristic style. Rather than criticize a work according to some norm of objective beauty,Schlegel insisted upon evaluating it according to its own aims and upon exposing its inconsistencies. This internal approach to a work was, according to Wilhelm Dilthey, an important step in the development of hermeneutics (see Hermeneutics).

It was also in the late 1790s that Schlegel made his most important contributions to the ethics and politics of Romanticism. His 'Versuch über den Begriff des Republikanismus' ('Essay on the Concept of Republicanism') (1796 (1996)), one of the most radical writings of the 1790s in Germany, was a defence of the right of revolution and direct democracy against Kant. His *Athenäums Fragmente* (1798-1800) and *Lucinde* (1799) were also far ahead of their time in championing such progressive causes as sexual liberation and the emancipation of women. Along with Schleiermacher, Schlegel developed an ethic of love and individuality in reaction to the abstract formalism of Kant's ethics (see Kantian ethics). He maintained that love cannot be understood as a legal obligation, and that it should involve a sensualization of the spirit as much as a spiritualization of the senses.

Schlegel's contribution to the metaphysics of Romanticism is his *Vorlesungen über die Transzendentalphilosophie (Lectures on Transcendental Philosophy)*, delivered as lectures in Jena in 1800. This work is an attempt, several years before Hegel, to develop a synthesis of naturalism and idealism, of Spinoza and Fichte. While Schlegel argued that Kant and Fichte had wrongly separated the self from nature and history, he also criticized Spinoza for his static and ahistorical conception of the divine substance.

In the early 1800s the flame of Schlegel's early radicalism dimmed and his thought moved steadily in a conservative direction. He became disillusioned with the French Revolution, which seemed to end in anarchy, commercialism and military dictatorship. Increasingly, he saw the defence of the Catholic Church and the old social hierarchy as the only safeguards against these disturbing trends, and as the only pillars of spiritual and communal values. His growing conservatism culminated in his conversion to the Roman Catholic Church in 1808 and in his diplomatic and literary activity on behalf of Metternich between 1809 and 1818. In his later political writings, especially his *Signatur des Zeitalters (Sign of the Age)* (1820), Schlegel defended a virtually reactionary position: that the basis of all right is tradition, that society should be organized according to estates, and that social order depends upon the restoration of the Church.

His later works are mainly an apology for his neo-Catholic mysticism. His Philosophie des Lebens (Philosophy of Life) (1828) taught that the aim of philosophy should be to develop the spiritual life of a person, their receptivity for a divine revelation. His *Philosophie der Geschichte (Philosophy of History)* (1829) held that world history is not a progression towards greater rationality, but an attempt to return to spiritual grace and harmony with the divine. Although this work does give great importance to non-Western cultures, Schlegel still sees European Christianity as the turning point of history and the culmination of civilization.

See also: Romanticism, German

FREDERICK BEISER

List of works

Schlegel, F. von (1958-) *Kritische Friedrich Schlegel Ausgabe (Critical Friedrich Schlegel Edition)*, ed. E. Behler, J.J. Anstett and H. Eichner, Paderborn: Schöningh. (The standard edition ofSchlegel's works.)

Schlegel, F. von (1797) *Über das Studium der Griechischen Poesie (On the Study of Greek Poetry)*, Berlin: Unger. (The chief work of Schlegel's early classical phase.)

Schlegel, F. von (1798-1800) *Athenäum. Eine Zeitschrift von August Wilhelm und Friedrich Schlegel* (Athenaeum: A Journal by August Wilhelm and Friedrich Schlegel), Berlin: Vieweg, 3 vols.(The main journal of the Romantic movement.)

Schlegel, F. von (1799) *Lucinde*, Berlin: Fröhlich.(Schlegel's first and only novel.)

Schlegel, F. von (1800a) Gespräch über Poesie (Conversation on Poetry), in *Athenäum*, vol. 3, 169-87. (Schlegel's reflections on the nature of art and the aims of the Romantic artist.)

Schlegel, F. von (1800b) *Vorlesungen über die Transzendentalphilosophie* (Lectures on Transcendental Philosophy), in *Neue philosophische Schriften*, ed. J. Körner, Frankfurt: Schulte Verlag, 1935.(Written as lectures in 1800, these were published only in 1935, and are the main source for Schlegel's early philosophy.)

Schlegel, F. von (1808) *Über Sprache und Weisheit der Indier* (On the Language and Wisdom of India), Heidelberg: Mohr.(Schlegel's pioneering work on Indian philosophy.)

Schlegel, F. von (1820) *Signatur des Zeitalters (Sign of the Age)*, in *Concordia*, Vienna: J.B. Wallishausser, vols 1-4: 3-70, 164-90, 343-98.

Schlegel, F. von (1828) *Philosophie des Lebens (Philosophy of Life)*, Vienna: Schaumberg.(The main exposition of Schlegel's mature philosophy.)

Schlegel, F. von (1829) *Philosophie der Geschichte (Philosophy of History)*, Vienna: Schaumberg.(The main work for Schlegel's mature philosophy of history.)

Schlegel, F. von (1968) *Dialogue on Poetry and Literary Aphorisms*, trans. and ed. E. Behler, University Park, PA: Penn State Press.

Schlegel, F. von (1957) *Literary Notebooks 1797-1801*, ed. H. Eichner, Toronto, Ont.: Toronto University Press.

Schlegel, F. von (1971) *Lucinde and the Fragments*, trans. P. Firchow, Minneapolis, MN: University of Minnesota Press. (Translations of Schlegel (1798-1800) and (1799).)

Schlegel, F. von (1996) *Early Political Writings of the German Romantics*, trans. and ed. F. Beiser, Cambridge: Cambridge University Press.(Contains 'Essay on the Concept of Republicanism' (1796), *Athenäums Fragments* and parts of *Lectures on Transcendental Philosophy*.)

References and further reading

Ayrault, R. (1961) *La Genèse du romantisme allemand (The Genesis of German Romanticism)*, Paris: Aubier. (The most thorough account of the early years of German Romanticism; vol. 4, 279-311 is devoted toSchlegel's Romantic aesthetic.)

Eichner, H. (1970) *Friedrich Schlegel*, Twayne World Author Series 98, New York: Twayne.(A useful introduction to all aspects of Schlegel's thought.)
Haym, R. (1890) *Die romantische Schule (The Romantic School)*, Berlin: Gaertner. (The classic account of the young Schlegel and the origins of the Romantic school.)
Silz, W. (1929) *Early German Romanticism*, Cambridge, MA: Harvard University Press.(A useful introductory survey.)
Walzel, O. (1932) *German Romanticism*, New York: Putnam.(A useful and careful survey of early Schlegel and the Romantic movement.)

Schleiermacher, Friedrich Daniel Ernst (1768-1834)

Friedrich Daniel Ernst Schleiermacher was the most notable German-speaking protestant theologian of the nineteenth century. He gave significant impetus to the re-orientation of theology after the Age of Enlightenment (see his speeches Über die Religion (On Religion) (1799), and also Kurze Darstellung des theologischen Studiums (Brief Outline of Theology as a Field of Study) (1811a)) and he enjoyed a wide audience in Berlin both as preacher and Professor of Theology and Philosophy. Throughout his life he was a fervent advocate of the union between the Lutheran and the Reformed Church established in the so-called Old Prussian Union, and his compendium Der christliche Glaube (The Christian Faith) (1821, 1822) is held to be the first dogmatics transcending the denominational boundaries between the Reformation Churches. His translation ofPlato attained the status of a classic. In his university lectures and academic speeches on philosophy he made a profound and lasting impression on his audience, both in his historical and systematic thought. He also had an important hand in the reform of the German Universities. In theology and philosophy he strove to find an independent and intermediate position between the Enlightenment, German Idealism and Romanticism.

1 Life

Schleiermacher was born on 21 November 1768 in Breslau, the second child and eldest son of the Reformed Church chaplain Johann Gottlieb Adolph Schleyermacher and his first wife Elisabeth Maria Katharina Schleyermacher (born Stubenrauch). After a childhood spent in Breslau, Pless and Anhalt (Upper Silesia), Schleiermacher was educated among the Moravians (Herrnhuter Brüdergemeine), a small Lutheran community established by Zinzendorf in 1727. He was a pupil at the *Pädagogium* (grammar school) in Niesky and then went to the *Seminarium* (college) in Barby. After leaving the Moravians, he studied theology and philosophy at the Friedrichs-Universität in Halle an der Saale from 1787 to 1789. From Johann August Eberhard he gained familiarity with ancient and contemporary philosophy. Schleiermacher sought to find his own position within the intellectual dialogue that was being conducted between the Leibniz-Wolff school and Kant's critical philosophy, which had begun to gain ascendency in Germany. He passed his first Church examination in 1790 and took a post in Schlobitten, East Prussia, as tutor to the family of Count Dohna. After successfully negotiating his second Church examination in 1794, he became a preacher in the Reformed Church, initially at Landsberg an der Warthe; then, from 1796 to 1802, at the Charité hospital in Berlin and eventually at Stolp in Pomerania. In 1804 he was appointed Professor of Reformed Theology in Halle an der Saale, but after the Prussian defeat at the hands of the French and the truce at Tilsit in 1807, he moved back to Berlin. He participated in the literary discussion concerning the reform of the university, the subject of his *Gelegentliche Gedanken über Universitäten in deutschem Sinn (Occasional Thoughts on Universities in the German Sense)* (1808a). In 1809 he married the widow of a friend, Henriette von Willich (born von Mühlenfels), and took up the pastorate of the Reformed Church at the Dreifaltigkeitskirche in Berlin. In the following year Schleiermacher also became Professor of Theology at the newly founded Friedrich-Wilhelms-Universität, and a member of the philosophical section in the Royal Academy of Sciences in Berlin. In this capacity he gave lectures both in theology and philosophy at the university. As secretary of the philosophical section from 1814 and, from 1826, also secretary of the section in history and philology, he undertook the reorganization of the Academy, becoming secretary of the historical-philosophical section in 1827. While Schleiermacher worked actively in the Prussian reform movement after 1807, participating officially in educational reform from 1810-14, he took a stance in opposition to the government during the Restoration era. He continued his official ecclesiastic and academic duties in Berlin until his death on 12 February 1834.

2 Philosophical work

Schleiermacher's main philosophical interest was in the field of ethics. As a member of the Berlin circle of Early Romantics he anonymously published several small essays on ethical themes in the *Athenaeum* journal of the Schlegel brothers (1798-1800) (see Schlegel, F.). His poetical early writings were in part dedicated to the exposition of his ethical views: *Monologen. Eine Neujahrsgabe (Soliloquies)* (1800a) as the manifesto of an individualist ethics, and his *Vertraute Briefe über Friedrich Schlegel's Lucinde (Letters concerning Friedrich Schlegel's Lucinde)* (1800b) as the Romantic interpretation of erotic love. His first systematic work was *Grundlinien einer Kritik der bisherigen Sittenlehre (Principles of a Critique of all Doctrines of Ethics Hitherto)*

(1803), which was a critical examination of ethical systems from Antiquity to Kant (§§9-11) and Fichte (§6). Schleiermacher laid the plans for his own exposition of ethics in his lectures, but never published them. Instead he published detailed ethical investigations in the course of his academic treatises (1819-30). Areas to which he directed much philosophical attention were his research into the history of philosophy (for instance, *Herakleitos* (1808b)) and to his translation of Plato's works (1804-28). However,Schleiermacher also undertook comprehensive systematic studies in the course of his lecturing career and it is here that he made his seminal contribution to hermeneutics, pedagogics and psychology. He never managed to work out his rich and diverse lectures on different philosophical disciplines for publication; his philosophical system has to be reconstructed from the relevant posthumously published manuscripts and his students' notes of his lectures.

3 The system of sciences

Schleiermacher's system follows the ancient division of philosophy into dialectics, physics and ethics. His idea of philosophy is the perfection of knowledge in the mutual interpenetration of six basic sciences (*Grundwissenschaften*), four with regard to the material aspects of knowledge and two to the formal. He structures the system of material sciences with the help of the dual opposition between reason and nature on the one hand and essence and existence on the other. The sciences of nature and reason, which differ with respect to their objects, fall under the modal distinction of the speculative and the empirical. The science of nature and the science of reason are thus subdivided into a speculative science of the essence of nature and reason respectively and into an experiential science of the appearance of nature and reason respectively. Schleiermacher calls these four material sciences 'physics' and 'the study of nature' (*Naturkunde*), 'ethics' and 'the study of history' (*Geschichtskunde*). The formal sciences are dialectics and mathematics. Dialectics is the supreme science, which replicates the supreme knowledge formally, just as mathematics is concerned with the form of the particular as such.

Schleiermacher did not explore the natural sciences to any great degree, and in this domain he relied in particular on the work of Henrik Steffens (*Grundzüge der philosophischen Naturwissenschaft*, Berlin, 1806). What he did attempt to reconstruct philosophically was the self-realization of reason in the process of history. The study of history illustrates the doctrine of ethics and the doctrine of ethics structures the study of history. Since there is no transition, but rather a hiatus between ethics as a science of essence and the study of history as a science of appearances, both ethics and history have to be related to each other in theory and in practice through critical and technical procedures. Critical procedure is investigative, whereas technical procedure is regulative. While critique (*Kritik*) judges the particular appearances as representations of the ideas, technology (*Technik*) gives instructions as to how, under different natural conditions, the production of particular appearances might take place. Psychology, aesthetics, political science and the philosophy of religion belong to the critical disciplines, while hermeneutics, pedagogics and political wisdom belong to the technical disciplines.

4 Dialectics

Schleiermacher's elementary philosophy, which sets out the procedural rules for the production of knowledge, he calls 'dialectics' (first lectures 1811). In his dialectics he expounds the principles according to which speech (both dialogue and monologue) which is aimed at attaining knowledge must proceed, if it is to fulfil its task of developing contested claims into completely accepted, universally valid knowledge. Dialectics is an artificial doctrine, designed to avoid conflict and doubt in the field of pure thought and reach a permanent and harmonious accord. Thus it stands opposed to scepticism. But since dialectics presupposes a conflict about knowledge, it cannot be the beginning of all knowledge. It thus gets its meaning from its intermediary position between the already extant will to know and the perfection of knowledge. The historical and social limitations of dialectics result from the fact that all thought is bound to language.

Schleiermacher's dialectics joins metaphysics (transcendental philosophy) and logic (formal philosophy). In the transcendental part he deals with the fundamental principles, conditions and structures of knowledge, and in the technical part he gives directions as to how knowledge is to be produced. In the first part dialectics is *prima philosophia*, in the second it is a logical apparatus ('organon') for the construction of the totality of knowledge and for the evaluation of individual items of knowledge.

The mutual play of reason and sensuality lies at the root of all knowledge. Reason as an intellectual function gives overall unity and determines the form of thinking. Sensuality as an organic function furnishes the manifold

material contents. All concrete perceptions are conceptually structured, and all concepts are saturated with experience. Logic and ontology are parallel. Forms of being and forms of knowledge correspond to one another.

Two transcendental ideas which transgress the limits of real knowledge - the idea of the world and the idea of God - are none the less constitutive of all knowledge. The idea of the world signifies the totality of the oppositions which beset finite beings, and to which real thinking approximates. The idea of God means the unity without oppositions, which is presupposed by all knowledge, and to which knowledge cannot approximate. Schleiermacher pairs the ideas of the world and God, because God without the world would be an empty representation, and the world without God something purely contingent.

5 Ethics

Ethics, or the philosophical doctrine of morals (first lectures 1804-5 (1805-6)), next to which Schleiermacher places the Christian doctrine of morals (first lectures 1806 (1834-64: 1, 12)), is the science of the principles of history. In the ethical process of history, reason appropriates nature in such a way that their opposition is eventually overcome. History is complete when nature has become the organ and symbol of reason, and when what is individual and what is universal have been wholly evened out in a comprehensive process of interaction.

Philosophical ethics is a theory of culture and formulates the essential content of history, in so far as it understands the present as a phase on the way to perfection. Ethics is more descriptive than prescriptive. It tries to avoid the Kantian dualisms of duty and inclination, universal reason and individual will, ought and is (see Kantian ethics). Its central concern is the productive moral power of the individual, which manifests itself in different social situations. The morality of an individual is embedded in sociality. The moral law becomes analogous to a law of nature and understood as its improvement.

Schleiermacher divides ethics into the doctrine of the good, the doctrine of virtue and the doctrine of duty, and gives precedence to the doctrine of the good, as in ancient philosophy. The doctrine of virtue asks the question, ‘Who is good and who does good?’ and looks at the moral power of the individual. Following the ancient canon of cardinal virtues Schleiermacher recognizes four main virtues: wisdom, love, prudence and perseverance. The doctrine of duty asks the question, ‘What is good?’ and looks at the moral deed. It teaches a system of actions and modes of behaviour which shows how everyone can best promote the universal moral goal of the highest good. In the process Schleiermacher looks at the compatibility of the different fields of social action, so that law, profession, love and conscience might fit together. The doctrine of the good asks after the very concept of moral goods, which human beings produce through their moral powers and which constitute the world of ethical action. The doctrine of the good constructs the concept of the highest good through the interrelation of two pairs of opposites: the action of reason on nature is organizing or symbolizing, and it is individual or universal (identical). Accordingly Schleiermacher divides up the manifold of concrete actions using the following schema: identical (universal) organization is the concern of work, business and economics; individual organization is the concern of property, friendship, sociality and hospitality. Identical (universal) symbolization, which is articulate in speech and in thought, is the concern of science; individual symbolization is the concern of art and religion. These four fields of ethical or moral life become relatively independent through the institutions of the state, the household, the scientific academy and the church. Moral demands on the individual are embedded in the forms of community, within which morality is always already present. Imperatives aim at the development of the highest good and to this extent they can always be based upon the fact that reason has already become in some measure objective.

6 Psychology

Schleiermacher’s psychology (first lectures 1818 (1834-64: 3, 6)) complements and concretizes his ethics. The basic structures of ethics are traced to the characteristics of individuals and peoples. Schleiermacher argues that the soul, which must not be isolated from the body, is only present in the context of one’s life as a whole. The ‘I’, the basic structure of the soul, is not thought of as the aggregate of human faculties, but is disclosed through its opposite activities. The receptive activity processes the impressions which are received from outside, while the spontaneous activity is creatively directed towards the outside world. Both these directions of activity are always woven together and increase in determinacy only in the course of their development. The task of psychology is to comprehend both basic activities in terms of their genesis and their different characteristics.

Receptivity is characterized by the activity of the senses. This activity leads to a doubling of all perceptive consciousness into an objective consciousness of things and a subjective self-consciousness. Within objective and subjective consciousness there is a tendency towards ever greater generalization. By dint of its realization in concepts and language, objective consciousness strives towards knowledge in the sense of a rational species-consciousness (*Gattungsbewußtsein*). In subjective consciousness the tendency towards generalization is made manifest in the feeling of sociality and in the formation of an individual consciousness which presages the infinite at the limits of the finite. Spontaneity is likewise analysed in terms of its subjective and objective aspects. The acquisitive appropriation of the external world of nature is juxtaposed to the manifestation of the self in art and science.

7 The philosophy of religion

Schleiermacher conceives the philosophy of religion as a critical discipline, which makes a structural comparison of the various forms of religion from the starting point of the ethical basis of religion. However, he never devoted any of his lectures specifically to the philosophy of religion. In his theological writings, *Kurze Darstellung des theologischen Studiums (Brief Outline of Theology as a Field of Study)* (1811a), *Der christliche Glaube (The Christian Faith)* (1821, 1822) and *über die Religion (On Religion)* (1799), one can find numerous references to the philosophy of religion. The essence of Christianity can be ascertained from a comparison of the different forms of religion.

Religion is a sense for the infinite and is based on the feeling of absolute dependency in the immediacy of self-consciousness. It is a universal element of life which is based in the universal essence of mankind and it does not conflict with the knowledge through which it is vouchsafed in reason and experience. Christianity, philosophy and the causal sciences are all compatible.

Piety and understanding are the poles between which the life of the spirit unfolds. Philosophy, which remains a kind of negative theology, points towards its opposite pole, which it is unable to attain. Philosophy and Christian piety gravitate together in a continual process of approximation. Philosophy does not offer the foundation for piety, but provides religion with concepts for its exposition.

Theology has no space of its own in the system of sciences. As a positive science, it is functionally constituted through its external task of offering its knowledge and skills for the governing and management of the church. Dogmatic propositions describe the self-consciousness which is germane to Christian piety; their truth is grounded in this self-consciousness.

8 Aesthetics

Schleiermacher understands aesthetics as a critical discipline which speculatively comprehends art in the context of all human activities and encyclopedically grasps the historical occurrence of the different forms of art (first lectures 1819 (1819, 1825)). According to ethics, art is the symbolization of something individual. It is generated from the individual's coming to self-awareness, from the individual's sensitivity for the world and from the expressive power of the individual's imagination. Art is the organ of the representation of what is individual. All human beings are therefore artists in this sense.

Schleiermacher's aesthetics is an aesthetics of production. It begins from the standpoint of the activity of the artist, not from the content of the art work. It looks at art's process of becoming and marks the stages in which feelings become objective in a work. The original datum is the becoming of the original image in the joining together of enthusiasm and temperance. The imagination fashions an internal image into the form of a work. This presentation in the material can then be experienced by others.

Schleiermacher's aesthetics is an aesthetics of expression. It examines the expression of subjective feeling, which at the same time brings to appearance something that is universally human. What is only inchoate in nature is made complete and explicit in artistic activity. In art the productive spirit shows itself to be that which interprets and advances creation.

Schleiermacher's division of the arts follows the fundamental distinction of his psychology between self-consciousness and object-consciousness. The visual arts of painting and sculpture proceed from

object-consciousness, by forming and physically realizing representations in free imaginative productions in external media. Music and mimic art proceed from self-consciousness by creatively shaping the articulation of feeling through gesture and sound. Poetry unites both poles, although it is more strongly attracted to the pole of object-consciousness.

9 Hermeneutics

Schleiermacher removes hermeneutics from its traditional contexts of theology, philology and jurisprudence, and develops it into a general theory of understanding (first lectures 1805 (1819)). Hermeneutics allows for thoughts which are articulated in language to be partially understood and partially not, and thus it operates neither on the level of complete knowledge nor on the level of complete ignorance. It claims to reconstruct individual and concrete linguistic utterances methodologically, in order to determine the thought which they articulate. Diametrically opposed to dialectics, hermeneutics considers the transition between linguistic communication and universal reason. The process of understanding is open-ended and interminable, like the process of history. Every interpretative act must have a provisional grasp of the whole in order to comprehend the parts, which in their turn reveal the contours of the whole more precisely. Thus hermeneutics does have an ethical orientation, since it participates theoretically and practically in the realization of the highest good, in so far as it furthers successful communication and the formation of the scientific community.

What characterizes Schleiermacher's hermeneutics, which counts as one of the technical disciplines, is that it prescribes a method. The method specifies four procedures with which given speeches or texts have to be reconstructed: grammatical, psychological, comparative and divinatory interpretation. Since every linguistic utterance is both the product of an author and a component of a linguistic system, grammatical (objective) interpretation examines the utterance in the context of language as a whole, while psychological (subjective) interpretation focuses on the context of the individual's production of thoughts. The comparative procedure clarifies obscurities with the help of what is already understood, and to this end it requires particular historical and philological investigations. The divinatory procedure elucidates meaning and context intuitively, and to this end it is based on the autonomous productivity of the interpreter. The four basic procedures are normally put to use in combinations of two. The grammatical method is predominantly comparative, while psychological interpretation is mainly divinatory. Nevertheless, attention should be paid to all four procedures in the course of every interpretation.

The hermeneutic formula of method is complemented by its indication of the goal of inquiry - to understand a given utterance or text as well as its author, and then better than its author. So first of all interpreters have to attempt to put themselves in the position of the author, and then, distancing themselves from this proximity with the author, to make new sense of the text or utterance in the context of the linguistic system.

10 Pedagogics

As a technical discipline, pedagogics inherits its guiding concepts and basic structure from the field of ethics (first lectures 1813 (1834-64: 3, 9)). The ethical relation between individual and community, between nature and reason, provides education with its two guiding perspectives: individual education - the drawing out of the student's individual nature - and social education - the integration of the student into ethical life.

Schleiermacher distinguishes three periods of education: the family upbringing and acquisition of language; school education; and vocational training. The activity of education which is directed towards the formation of character and capacities has to promote good and suppress evil. The intellectual capacities which involve the student's *Weltanschauung* (worldview) are honed on the scientific study of history and nature; the ethical capacities, which concern their *Weltbildung* (their way of making the world) are formed through the undertaking of ethical and human tasks.

Since pedagogics as a technical discipline presupposes that a certain level of culture has already been attained, it imposes the twofold task of preserving the ethical *status quo* and ameliorating what is still ethically imperfect. Schleiermacher puts his pedagogics in the historical situation of Europe after the French Revolution, namely in the conflict between a traditionalism hostile to all reform and a revolutionism hostile to all tradition. His great aim is the overcoming of social inequality, of the division between aristocracy and bourgeoisie. The pedagogics of ethical

perfection is supposed to smooth the path of progress without revolutionary force and thus shows itself to be in accord with the Prussian policy of reform.

11 Politics

Schleiermacher conceives politics as a combined critical and technical discipline (first lectures 1808-9 (1980-: 2, 8)). Politics indicates the general perspectives from which existing states are to be appraised. Thus political activity can be harmonized with the level of culture, which for Schleiermacher is characterized above all by the Prussian reform movement and the national uprising against France.

The state is constituted by the distinction between government and people, and by the presence of an accepted law. It is a social institution which regulates identical organization (the communal mastery of nature) and thus guarantees the subsistence of the citizens. The state therefore has a particular orientation towards administration and economics. In this way Schleiermacher downgrades the significance of the military. The progress of culture is furthered by labour. Schleiermacher rejects capital punishment, offensive wars, violent revolutions and enforced colonization.

The state has to overcome the basic tension between the individual and the universal will; it must take upon itself the task of restraining private interest where it threatens the public good, but must also restrict itself and leave private interest to its own devices where such interest serves the ends of humanity. The state must take appropriate measures to intervene in the labour process, in order to ensure that the universal interest is served. It must not become an economic subject in the process, however, but must guarantee social welfare in the way in which it would uphold a legal duty.

Schleiermacher holds constitutional monarchy to be the most appropriate form of constitution for large, modern nation-states. In a constitutional monarchy the legislative power extends from the people to the king, and the executive power extends from the king to his subjects. Standing above all private interests, the king is in an excellent position to guarantee the universality of freedoms and justice. Aristocracy has the disadvantage that it fixes differences in class and status. Democracy has the disadvantage that public and private interest can come into intractable conflict with each other.

Schleiermacher accords the state no primacy over other forms of community. He rather places the state beside church, household and academy of science, so that religion, free sociality and science are thereby removed from its jurisdiction.

See also: Hermeneutics

Translated from the German by J.G. Finlayson
GÜNTER MECKENSTOCK

List of works

Schleiermacher, F.D.E. (1980-) *Kritische Gesamtausgabe (KGA)*, ed. H.-J. Birkner, G. Ebeling, H. Fischer, H. Kimmerle, G. Meckenstock, K.-V. Selge, Berlin and New York: de Gruyter.(New Critical edition of all works, lectures, sermons, translations and letters.)

Schleiermacher, F.D.E. (1834-64) *Sämmtliche Werke* (Collected Works), Berlin: Reimer.(Early thirty-volume edition of the works, sermons and lectures.)

Schleiermacher, F.D.E. (1799) *Über die Religion. Reden an die Gebildeten unter ihren Verächtern*, Berlin: Unger; 2nd edn, revised, 1806; 3rd edn, revised and extended, 1821; 4th edn, 1831; 1st edn trans. R. Crouter, *On Religion. Speeches to its Cultured Despisers*, Cambridge: Cambridge University Press, 1988; new edn, 1996; 4th edn trans. J. Oman, London: Paul, Trench, Trübner, 1893; New York, Harper, 1959.(Classical document of the early Romantic understanding of religion; critical edition in *KGA* 1, 2; 2nd-4th edns in *KGA* 1, 12.)

Schleiermacher, F.D.E. (1800a) *Monologen. Eine Neujahrsgabe (Soliloquies: A New Years Gift)*, Berlin: Spener; 2nd edn, revised, 1810; 3rd edn, revised, 1822; 4th edn, revised, 1829; 1st edn trans. H.L. Friess, *Soliloquies*, Chicago, IL: Open Court, 1926.(Manifesto of an ethics of individuality; critical edition in *KGA* 1, 3; 2nd-4th edns in *KGA* 1, 12.)

Schleiermacher, F.D.E. (1800b) *Vertraute Briefe über Friedrich Schlegel's Lucinde (Confidential Letters on Friedrich Schlegel's Lucinde)*, Lübeck and Leipzig: Bohn.(Apology for Friedrich Schlegel's erotic novel,

which caused a great scandal; critical edition in *KGA* 1, 3.)
Schleiermacher, F.D.E. (1803) *Grundlinien einer Kritik der bisherigen Sittenlehre (Principles of a Critique of All Hitherto Doctrines of Ethics)*, Berlin: Realschulbuchhandlung. (Critical analysis of ancient and modern systems of ethics.)
Schleiermacher, F.D.E. (ed. and trans.) (1804-28) *Platon's Werke (Plato's Works)*, Berlin: Realschulbuchhandlung and Reimer, 6 vols; introductions trans. W. Dobson, *Introduction to the Dialogues of Plato*, Cambridge: Deighton, 1836.(Classic translation of Plato's dialogues into German with critical introductions.)
Schleiermacher, F.D.E. (1805-6) *Brouillon zur Ethik (Outline of Ethics)*, ed. H.-J. Birkner, Hamburg: Meiner, 1981.(Lectures of 1805-6. The introduction contains a bibliography.)
Schleiermacher, F.D.E. (1808a) *Gelegentliche Gedanken über Universitäten in deutschem Sinn. Nebst einem Anhang über eine neu zu errichtende*, Berlin: Realschulbuchhandlung; trans. T.N. Tice and E. Lawler, *Occasional Thoughts on Universities in the German Sense*, San Francisco, CA: Edwin Mellen Press, 1991. (Programmatic work of educational and political significance on the founding of the University of Berlin.)
Schleiermacher, F.D.E. (1808b) *Herakleitos der dunkle, von Ephesos, dargestellt aus den Trümmern seines Werkes und den Zeugnissen der Alten (Heraclitus the Obscure of Ephesus, reconstructed from the Remnants of his Work and the Testimonies of the Ancients)*, in *Museum der Alterthums-Wissenschaft* 1, 313-533.(Historical reconstruction of the doctrine of Heraclitus.)
Schleiermacher, F.D.E. (1811a) *Kurze Darstellung des theologischen Studiums zum Behuf einleitender Vorlesungen*, Berlin: Realschulbuchhandlung; 2nd edn, 1830; trans. T.N. Tice and E. Lawler, *Brief Outline of Theology as a Field of Study*, San Francisco, CA: Edwin Mellen Press, 1988.(Functionally structured encyclopedia of theology. 2nd edition heavily reworked and extended.)
Schleiermacher, F.D.E. (1811b) *Dialektik (Dialectics)*, ed. A. Arndt, Hamburg: Meiner, 1986.(Lectures of 1811. The introduction contains a bibliography.)
Schleiermacher, F.D.E. (1812-13) *Ethik* (Ethics), ed. H.-J. Birkner, Hamburg: Meiner, 1981.(Lectures of 1812-13. The introduction contains a bibliography.)
Schleiermacher, F.D.E. (1814-15) *Dialektik (Dialectics)*, ed. A. Arndt, Hamburg: Meiner, 1984. (Lectures of 1814-15 and a fragment from 1833. The introduction contains a bibliography.)
Schleiermacher, F.D.E. (1819) *Hermeneutik (Hermeneutics)*, ed. H. Kimmerle, Heidelberg: Winter, 1974. (Lectures of 1819 and other source texts on hermeneutics.)
Schleiermacher, F.D.E. (1819, 1825) *Ästhetik (Aesthetics)*, ed. T. Lehnerer, Hamburg: Meiner, 1984. (Lectures from 1819 and 1825. The introduction contains a bibliography.)
Schleiermacher, F.D.E. (1821, 1822) *Der christliche Glaube nach den Grundsäzen der evangelischen Kirche im Zusammenhange dargestellt (The Christian Faith according to the Principles of the Protestant Church Expounded in Context)*, Berlin: Reimer, 2 vols; 2nd edn revised, 1830, 1831; trans. H.R. Mackintosh and J.S. Stewart, *The Christian Faith*, Edinburgh: T. & T. Clark, 1928, 1948; New York, Harper, 1963.(The most significant protestant dogmatics of the nineteenth century.)
Schleiermacher, F.D.E. (1992) *On Freedom*, trans. A.C. Blackwell, Lewiston, NY: Edwin Mellen Press. (Translation of *KGA* 1, 1: 217-356. Detailed analysis of the concept of freedom.)
Schleiermacher, F.D.E. (1992) *On the Highest Good*, trans. H.V. Froese, Lewiston, NY: Edwin Mellen Press. (Translation of *KGA* 1, 1: 81-125. Critical discussion of Kant's *Critique of Practical Reason*.)
Schleiermacher, F.D.E. (1995) *On What Gives Value to Life*, trans. E. Lawler and T.N. Tice, Lewiston, NY: Edwin Mellen Press.(Translation of *KGA* 1,1: 391-471. Phenomenological analysis and moral rating of human life and fate.)

References and further reading

Lehnerer, T. (1987) *Die Kunsttheorie Friedrich Schleiermachers (Schleiermacher's Theory of Art)*, Stuttgart: Klett-Cotta.(Systematic exposition of Schleiermacher's theory of art.)
Meckenstock, G. (1988) *Deterministische Ethik und kritische Theologie. Die Auseinandersetzung des frühen Schleiermacher mit Kant und Spinoza 1789-1794 (Deterministic Ethics and Critical Theology: The Early Schleiermacher's Encounter with Kant and Spinoza 1789-1794)*, Berlin and New York: de Gruyter.(Detailed investigation of the early writings.)
Meckenstock, G. and J. Ringleben (eds) (1991) *Schleiermacher und die wissenschaftliche Kultur des Christenthums* (Schleiermacher and the Scientific Culture of Christianity), Berlin and New York: de Gruyter. (Contains essays on historical, theological and philosophical themes.)
Meding, W. von (1992) *Bibliographie der Schriften Schleiermachers nebst einer Zusammenstellung und Datierung seiner gedruckten Predigten (Bibliography of Schleiermacher's Writings with a Collection and*

Chronology of His Published Sermons), Berlin and New York: de Gruyter.(Bibliography of Schleiermacher's works.)

Moxter, M. (1992) *Güterbegriff und Handlungstheorie (The Concept of the Good and Theory of Action)*, Kampen: Kok Pharos.(Historical and systematic reconstruction of Schleiermacher's ethics of the good.)

New Athenaeum - Neues Athenaeum (1989-) A Scholarly Journal Specializing in Schleiermacher Research and Nineteenth-Century Studies- Zeistschrift für Schleiermacher-Forschung und für Studien zum 19. Jahrhundert, Lewiston, NY: Edwin Mellen Press.(Contains articles on and translations of Schleiermacher's work.)

Pleger, W. (1988) *Schleiermachers Philosophie (Schleiermacher's Philosophy)*, Berlin and New York: de Gruyter.(Introductory survey of the many subjects thematized by Schleiermacher's philosophy.)

Reymond, B. (1984) 'Schleiermacher en Français' (Schleiermacher in French), in *Archivio di Filosofia* 52: 489-97.(Bibliographical list of the French translations and the French Schleiermacher literature.)

Rieger. R. (1988) *Interpretation und Wissen (Interpretation and Knowledge)*, Berlin and New York: de Gruyter. (Historical account of Schleiermacher's hermeneutics.)

Scholtz, G. (1984) *Die Philosophie Schleiermachers (Schleiermacher's Philosophy)*, Darmstadt: Wissenschaftliche Buchgesellschaft.(Offers a very good introduction to the system and the history of Schleiermacher's philosophy, with all the important earlier literature.)

Selge, K.-V. (ed.) (1985) *Internationaler Schleiermacher-Kongreß Berlin 1984 (International Schleiermacher Congress, Berlin 1984)*, Berlin and New York: de Gruyter, 2 vols.(Contains essays on philosophical, theological and historical themes.)

Sorrentino, S. (ed.) (1992) *Schleiermacher's Philosophy and the Philosophical Tradition*, Lewiston, NY: Edwin Mellen Press.(Contains essays on various philosophical themes.)

Tice, T.N. (1966) *Schleiermacher Bibliography*, Princeton, NJ: Theological Seminary. (Bibliographical list of Schleiermacher's writings and the Schleiermacher literature, 1800-1964.)

Tice, T.N. (1985) *Schleiermacher Bibliography: Corrections, New Information and Comments*, Princeton, NJ: Theological Seminary.(Bibliographical corrections to the 1966 list.)

Tice, T.N. (1985) *Schleiermacher Bibliography (1784-1984), Updating and Commentary*, Princeton, NJ: Theological Seminary.(Supplements and additions to the 1966 list.)

Tice, T.N. (1989) 'Schleiermacher Bibliography: Update 1987', in *New Athenaeum - Neues Athenaeum* 1, 280-350.(Supplements and additions to the 1966 list.)

Tice, T.N. (1991) 'Schleiermacher Bibliography: Update 1990', in *New Athenaeum - Neues Athenaeum* 2: 131-63. (Supplements and additions to the 1966 list.)

Tice, T.N. (1995) 'Schleiermacher Bibliography: Update 1994', in *New Athenaeum - Neues Athenaeum* 4, 139-94. (Supplements and additions to the 1966 list.)

Zeitschrift für Neuere Theologiegeschichte - Journal for the History of Modern Theology (1994-), Berlin and New York: de Gruyter.(Articles on and editions of Schleiermacher's work.)

Schlick, Friedrich Albert Moritz (1882-1936)

Moritz Schlick is usually remembered as the leader of the Vienna Circle, a group that flourished from the late 1920s to the mid-1930s, and made an important contribution to the philosophical movement known as 'logical empiricism'. Yet many of Schlick's most original contributions to philosophy antedated the hey-day of the Circle, providing the foundations for much of its subsequent development. He started his academic career as a physicist, and his early contributions to philosophy include an influential conventionalist interpretation of general relativity and a new account of the definitions of the basic terms of theoretical science. In the debates that flourished within the Vienna Circle he is famous for his commitment to the Principle of Verifiability and his defence of a correspondence theory of truth. In addition, his works during the final years of the Vienna Circle represent some of the most sober reflections on the problems that vexed the early logical empiricists. Although few of the views identified with logical empiricism currently find favour among philosophers, their approach to philosophy, especially their identification of its central perplexities, still wields enormous influence among contemporary thinkers. Since Schlick contributed significantly to the form logical empiricism assumed during its period of dominance, there can be little doubt that his thought continues to inspire much philosophical thinking today.

1 Spacetime and conventionalism

Schlick originally planned a career in physics, and finished his doctorate under the direction of the Nobel prize-winner Max Planck, in 1904. Within a few short years his interests turned to philosophy and he soon he produced a youthfully enthusiastic ethical tract (*Lebensweisheit*) as well as a long essay on the concept of truth. Schlick's work on truth was an impressive achievement, for it not only provided an extensive discussion of the major treatments of truth then prevailing in German academic philosophy, it also contributed an original analysis of the nature of truth. Schlick's basic idea was that truth consists in what he called the 'univocal designation' by a judgment of a situation in the world. Judgments consist of constituents that must be coordinated with elements of reality, in order for the judgment to possess any external significance at all. If the constituents of a judgment are combined in such a way that the judgment as a whole univocally designates a situation in the world, then it is true; if not, the judgment is false. In the prevailing context of German Neo-Kantianism, Schlick's analysis of truth was strikingly innovative. Yet its greatest significance derives from the role it subsequently played in his interpretation of the new physics of relativity theory.

It was Schlick's philosophical interpretation of relativity theory that first brought him to eminence, not only among scientifically-minded philosophers but also in the eyes of prominent physicists. Schlick pioneered the conventionalist understanding of relativity that prevailed in the 1920s and eventually became a cornerstone of logical empiricist philosophy of science. After treating the Special Theory in an article in 1915, he applied his insights to the General Theory in a highly reputed monograph, *Space and Time in Contemporary Physics*. There Schlick's interpretation results from combining his insights concerning the nature of truth with a geometric conventionalism drawn from Jules Henri Poincaré.

Schlick recognized that the system of geometry and physics is not just a series of purely symbolic formulas unrelated to experience; the symbols occurring in geometric and physical expressions must be coordinated with elements of reality. Not only must the geometry and physical principles be chosen, they must be interpreted in experience by coordinating their symbols with identifiable empirical situations, or elements thereof. But once the system of geometry, physics, and symbolic coordinations is in place, there are still choices to be made. For, according to Schlick's work on the concept of truth, the goal is to provide a univocal designation of reality and, as the history of physics has taught, there are often distinct systems, each of which provides an unambiguous representation of reality. The choice among equivalent systems of representation can only be founded on considerations of simplicity, for no other consideration can adjudicate between equivalent systems that univocally designate reality.

A similar line of reasoning had led Poincaré to his own geometric conventionalism. Poincaré reasoned that, since the choice of a geometry can only be guided by considerations of simplicity, physicists will always opt for the simplest geometric system. Moreover, since Euclidean geometry is the simplest, physicists will always choose it over any alternative. Departing from Poincaré's conclusion, Schlick insisted that it is not just geometry that is

conventionally selected, but the whole system - including physical principles and empirical coordinations as well. Since, as Kant had already noted, it is never space itself, but only the motion of bodies in space that is the object of perception and measurement, it makes no sense to speak of the geometry of space apart from a specific understanding of the physical principles governing motion. There simply is no fact of the matter concerning which geometry best describes space, independently of a specific understanding of the behaviour of bodies. Consequently, the choice of a particular geometry can never be made apart from the construction of a specific theory (see Kant, I. §5).

Schlick's conventionalist interpretation would eventually prevail as the dominant understanding of the new physics; it was immediately acclaimed by Albert Einstein himself, who aided Schlick's efforts to obtain a position at the University of Kiel, which he assumed in 1921 (see Conventionalism).

2 Epistemology

Schlick then applied his understanding of physical knowledge to other domains, in his *General Theory of Knowledge* of 1918. There Schlick furthered his claim that his conventionalist epistemology distinguished itself not only from the metaphysics of Kant but from the phenomenalism of Ernst Mach as well. Philosophers like Ernst Cassirer and Hans Reichenbach rejected Schlick's conventionalism, claiming instead that if Kant were 're-worked', his philosophy could comprehend Relativity. But Schlick argued that, once Cassirer had 're-worked' the synthetic a priori (see Kant, I. §4) to fit relativity theory, little was left of what was ingenious about Kant's doctrine. Reichenbach's objection was different: although he explicitly subscribed to many of Schlick's own premises, he could not see how concepts could be formed without the Kantian a priori.

It was at this point that Schlick made a contribution that helped wrest twentieth-century philosophy of science from the grip of the past. Schlick had, in *General Theory of Knowledge*, provided an account of concept-formation by means of the doctrine of implicit definition, and Reichenbach himself had relied on it without fully appreciating its significance. Schlick borrowed the idea of implicit definition from the mathematician David Hilbert, in order to demonstrate how scientific concepts are formed according to a conventionalist philosophy. Hilbert utilized the method for application to geometry, arguing that the geometric primitives were defined by the axioms of the system, whether the system was Euclidean or one of the non-Euclidean geometries. But the philosophical significance of the method was not recognized until Schlick promoted it as a method of concept-formation generally applicable throughout the sciences. Schlick's idea was simply that one function of the most fundamental postulates of any science is to give sense to the key terms occurring in it. Thus generalized, the doctrine of implicit definition not only obviates the need for a Kantian a priori to constitute concepts, it reveals a method of concept-formation distinctly at odds with the 'abstraction and generalization' pattern characteristic of empiricist philosophies, like the phenomenalism of Ernst Mach. Thus Schlick's new epistemology blazed an original trail that departed from time-worn paths.

3 The Vienna Circle

By the time the second edition of *General Theory of Knowledge* appeared in 1925, Schlick had already been exposed to new ideas that would deeply affect his thinking. In 1922, he assumed the prestigious chair of *Naturphilosophie* at the University of Vienna, which had been occupied by such luminaries as Ludwig Boltzmann and Ernst Mach. There he became involved with other scientifically-minded thinkers, including mathematicians and social scientists as well as philosophers. This group, the precursor of the Vienna Circle (see Vienna Circle), was interested in the latest developments in logic and the foundations of mathematics; indeed, one of their earliest projects was to read and discuss Ludwig Wittgenstein's *Tractatus Logico-Philosophicus*. Soon Rudolf Carnap would join the faculty at Vienna and provide Schlick with yet another source of innovative ideas.

By the late 1920s, Schlick began meeting with Wittgenstein in Vienna, discussing all sorts of philosophical topics, especially ones that arose from scientific contexts. The result was one of the most thoughtful interchanges on issues in geometry, physics, and philosophy of science between two of the major thinkers of the twentieth century. These discussions, as well as Schlick's interaction with his University colleagues, served as a prelude to the period of greatest activity in his philosophical career. Indeed, it was from the heady intellectual atmosphere created by these exchanges that the Vienna Circle would emerge, and Schlick would become its natural leader. More significant, perhaps, is that it was in this period that Schlick first articulated his own variety of logical empiricism.

And though many have cited this period as one in which Schlick abandoned the views of his earlier years, there is a natural continuity that ties his entire career together.

These new influences first became apparent in Schlick's attempts to address the challenge to traditional conceptions of causality raised by microphysics. He envisaged a stratification of principles, ranging from statements describing singular statements of fact, through empirical hypotheses, to philosophical principles. Statements at the lowest level, asserting the existence of individual states of affairs, embody all the evidential content of science. Wittgenstein had suggested that hypotheses are introduced to systematize these singular statements, by providing universalized characterizations of a limitless number of cases. Thus the utility of hypotheses lies in their function in the framing of predictions. From the hypothesis 'All *F*s are *G*s', along with the knowledge that some particular *x* is an *F*, we could predict that *x* will be a *G*. Hypotheses may thus be conceived as rules for forming expectations, as Wittgenstein suggested. But, as Wittgenstein further noted, to accept an hypothesis is not just to recognize that it yields accurate predictions, it is also a commitment to the existence of the ontology it posits, to the range of entities described by the terms of the hypothesis. Just employing the term 'electron' and its ilk in an hypothesis commits one to a microphysical ontology. Schlick took Wittgenstein's ideas a step further, and argued that, just as hypotheses may be understood as rules for forming predictions, the causality principle is a rule for framing hypotheses. Analogously to the relation of hypotheses to singular statements, certain philosophical principles control our formation of scientific claims. One such principle is the causality principle. Although the principle of causality is usually taken to express a claim about the nature of the world, it is really a prescription for framing hypotheses. Taken as such, the principle recommends forming hypotheses that yield absolutely precise predictions. And the new developments in quantum physics show that the utility of the causal principle is limited by discoveries in the submicroscopic domain.

4 Empiricism

Schlick's reasoning was guided by a precept that would play a larger and larger role in his thinking, the Principle of Verifiability. Schlick regarded this principle as the core of his empiricism. And though it has been interpreted in nearly as many ways as there were early logical empiricists, Schlick's understanding of the Verifiability Principle was one of the most liberal. Roughly speaking, what this principle required was that every meaningful statement asserts only what can be verified in experience. Schlick hastened to add that this does not mean that only experience is real, for such a claim would itself be meaningless. Nor did Schlick employ the principle to exclude from significant discourse physical statements which assert more than can be verified in a finite number of immediate experiences. Rather, Schlick used the Verifiability Principle only to dismiss as empty typical philosophical theses that assert the reality of one domain of entities to the exclusion of others, such as the thesis of positivism, construed as the assertion that only immediate experience is real and implying that physical objects are mere constructions out of immediate experience that summarize regularities within the given. Since whether this thesis were true or false would make no difference in our experience, the thesis is meaningless. And the same holds for the realist thesis that only physical objects are real. Like positivism, this thesis purports to assert the reality of one domain while denying others, but does not make any claim that can be verified in experience.

Schlick also deployed the Verifiability Principle to revive, in a somewhat altered form, a distinction already present in his earlier thought. In *General Theory of Knowledge*, he had created a deep division between concepts, formally constituted by means of implicit definitions, and intuitions, the experiences to which concepts must be linked to provide us with knowledge of the world. Later, Schlick identified the conceptual structure of our knowledge with its form, and intuitions as its material content, citing the concurrence of Carnap and Wittgenstein. The structure of our knowledge is embodied in the logic of the language in which we express it, but the intuitions to which each of us connects concepts are purely subjective. Thus, although private experience is the ultimate touchstone of the truth of our beliefs about the world, it can never be expressed, although its structure may be. For what, Schlick asked, would it mean to compare the subjective experiences of distinct individuals, other than to relate the ways in which the individuals have structured those contents? Since there is no means of verifying sameness or difference of the contents of distinct individuals, the question of their comparison cannot be answered. The consequence, which Schlick called 'The Thesis of the Incommunicability of Content', is that content, the ineffable material of experience, can never be expressed.

Although other members of the Circle, especially Carnap, had developed versions of the form-content dichotomy,

most had abandoned it by the time Schlick was working out his own ideas on the matter. By then the other Circle members were embracing physicalism: the idea that all scientific discourse could be expressed in the language of physics. In particular, they thought that statements expressing the results of observation, so-called ‘protocol sentences’, could be expressed in the everyday language of material things. Recognizing that such statements are, by their very nature, fallible, they admitted there was no absolutely secure basis for science. Some members of the Circle even flirted with the coherence theory of truth, the idea that truth consists in the consistency of the statements accepted by the community of scientific practitioners (see Truth, coherence theory of). Schlick was adamantly opposed to this last development, for it contradicted his long-standing view that truth is a property of statements that correspond to the facts (see Truth, correspondence theory of). In his classic essay, ‘On the Foundation of Knowledge’ (1934), he acknowledged that coherence was necessary for truth, but denied that it was sufficient. His argument involves little more than pointing out that, if coherence sufficed for truth, a well-crafted fairy tale would have to be accorded the same status as the finest textbooks in physics, chemistry, etc. Schlick then proceeded to sketch his own positive account, in which the correspondence doctrine was substantiated by the idea that the body of knowledge rests on a foundation of beliefs which directly correspond to what is observed. The foundation consists of physicalistic protocol sentences which, as Schlick and the other Circle members recognized, are fallible. But these observation statements are, in turn, grounded in fleeting momentary experiences he called ‘affirmations’. Affirmations themselves are absolutely certain, though this epistemic property is not transmitted to the fallible, physicalistic protocols to which they give rise. Naturally, Schlick’s claim that affirmations were ‘certain’ provoked immediate opposition within the Circle, mostly from members who did not understand that, on Schlick’s view, affirmations were little more than occasions for the assertion of proper observation reports and, as such, lie outside the system of science.

But the criticism and opposition to Schlick’s ‘Foundation’ essay gave him the opportunity to clarify his views on fundamental philosophical issues. He continued to maintain his earlier conventionalism, though expressed in terms of the choices to be made between alternative languages. He especially emphasized the limits of his conventionalism, denying that it allowed the possibility of adopting languages in which observational statements did not play a privileged epistemic role. This latter qualification was founded on his empiricism, expressed in the form of the Verifiability Principle. Thus, in his last essays, Schlick defended a form of empiricism that acknowledged the limited role of conventions, as well as the fallibility of scientific beliefs, while emphasizing the foundationalist view that observation is the final touchstone - indeed, the only touchstone - of the truth of our claims about the world.

On 6 June 1936, Schlick was on the steps of the University of Vienna when he was approached by one Johan Nelböck. Nelböck had received an advanced degree in philosophy from the University and, for some reason, had been threatening Schlick’s life for several years. Indeed, Nelböck had menaced Schlick to the extent that the authorities had committed the former to an asylum, where he was diagnosed as a paranoid schizophrenic, and released after observation. Although the simplest explanation of Nelböck’s behaviour is that he was deranged, the only evidence of his *dementia* was his threat against Schlick. But on that June day in 1936, he was in earnest, shooting Schlick four times in the abdomen and leg with an automatic pistol. Schlick died within a few hours.

See also: Logical postivism; Meaning and verification

THOMAS OBERDAN

List of works

Schlick, M. (1979) *Philosophical Papers*, vols I and II, eds B. v. de Velde-Schlick and H. Mulder, Dordrecht: Reidel.(The best current source of Schlick’s writings.)

Schlick, M. (1908) *Leibensweisheit (Life wisdom)*, Munich.(Schlick’s youthful ethical treatise.)

Schlick, M. (1910) ‘Das Wesen der Wahrheit nach der modernen Logik’, *Vierteljahrschrift für wissenschaftliche Philosophie und Soziologie* 34: 386-477; trans. P. Heath, ‘The Nature of Truth in Modern Logic’, in *Philosophical Papers*, vol. I, 41-103.(This highly original essay lays the groundwork for Schlick’s subsequent conventionalist interpretation of relativity theory.)

Schlick, M. (1915) ‘Die philosophische Bedeutung des Relativitätsprinzips’, *Zeitschrifft für Philosophie und philosophische Kritik* 159: 129-75; trans. P. Heath as ‘The Philosophical Significance of the Principle of Relativity’, *Philosophical Papers*, vol. I, 153-89.(Schlick’s treatment of the special theory of relativity.)

Schlick, M. (1917) ‘Raum und Zeit in der gegenwärtigen Physik: Zur Einführung in das Verständnis der

allgemeinen Relativitätstheorie', *Die Naturwissenschaften* 5: 161-7, 177-86; enlarged edn *Raum und Zeit in der gegenwärtigen Physik: Zur Einführung in das Verständnis der Relativität- und Gravitätionstheorie*, Berlin: Springer, 1919; 3rd edn trans. H. Brose as *Space and Time in Contemporary Physics*, New York: Oxford University Press, 1920; enlarged, revised 4th edn, Berlin: Springer, 1922, trans. P. Heath as 'Space and Time in Contemporary Physics', *Philosophical Papers*, vol. I, 153-89.(This work set the standard for the understanding of relativity theory that prevailed in the 1920s among such luminaries as Albert Einstein, Arthur Eddington and Hans Reichenbach.)

Schlick, M. (1918) *Allgemeine Erkenntnislehre*, Berlin: Springer. A revised, second edition appeared in 1925 and was translated into English by A. Blumberg and H. Feigl as *General Theory of Knowledge*, New York and Vienna: Springer, 1974.(This is a classic text - truly a 'must read' for anyone interested in the origins of analytic philosophy.)

Schlick, M. (1926) 'Erleben, Erkennen, Metaphysik', *Kant Studien* 31: 146-58; trans. P. Heath, 'Experience, Cognition, and Metaphysics', in *Philosophical Papers*, vol. II, 99-111.(Here Schlick introduced his form-content duality, citing - for the first time - the work of Wittgenstein and Carnap.)

Schlick, M. (1930) 'Die Wende der Philosophie', *Erkenntnis* 1: 4-11; trans. P. Heath, 'The Turning Point in Philosophy', in *Philosophical Papers*, vol. II, 154-60.(This is the 'lead' article in the first issue of the journal which served as the 'house organ' of the early logical empiricists, edited by Rudolf Carnap and Hans Reichenbach.)

Schlick, M. (1931) 'Die Kausalität in der gegenwärtigen Physik', *Die Naturwissenschaften* 19: 145-62; trans. P. Heath, 'Causality in Contemporary Physics', in *Philosophical Papers*, vol. II, 176-209.(Schlick presents the idea that results in the investigation of the quantum domain present an insuperable limit to the applicability of the principle of causality. His view was immediately attacked by Albert Einstein in correspondence.)

Schlick, M. (1932) 'Positivismus und Realismus', *Erkenntnis* 3: 1-31; trans. P. Heath, 'Positivism and Realism', in *Philosophical Papers*, vol. II, 259-84.(This essay evoked a sharp response from Schlick's mentor, Max Planck, who seems to have misunderstood its central thesis that neither of the positions mentioned in the title was significant.)

Schlick, M. (1934) 'Über das Fundament der Erkenntnis', *Erkenntnis* 4: 79-99; trans. P. Heath, 'On the Foundation of Knowledge', in *Philosophical Papers*, vol. II, 370-87. (This is the sharp response to developments by Carnap and Neurath that provoked the protocol sentence controversy.)

Schlick, M. (1936) 'Meaning and Verification', *The Philosophical Review* 45: 339-69; reprinted in *Philosophical Papers*, vol. II.(A full development of Schlick's view of language, which departs significantly from Wittgenstein's at that time.)

References and further reading

Friedman, M. (1983) 'Critical Notice: Moritz Schlick, Philosophical Papers', *Philosophy of Science* 50: 498-514. (This very helpful survey of Schlick's thought throughout his entire philosophical career focuses critically on his later work.)

Friedman, M. (1993) 'Geometry, Convention, and the Relativized *Apriori*: Reichenbach, Schlick and Carnap', in W. Salmon and G. Wolters (eds) *Logic, Language, and the Structure of Scientific Theories*, Pittsburgh: University of Pittsburgh Press, 21-34.

Lewis, J. (1988) 'Schlick's Critique of Positivism', in *PSA* 1988, East Lansing, MI: The Philosophy of Science Association, vol. 1, 110-17.(Like her subsequent work, this essay addresses problematic aspects of Schlick's epistemology in the context of his commitment to scientific realism.)

Lewis, J. (1990) 'Hidden Agendas: Knowledge and Verification', in *PSA* 1990, East Lansing, MI: The Philosophy of Science Association, vol. 2, 159-68.(Further development of the author's work on Schlick's epistemology in the context of scientific realism.)

Lewis, J. (1996) 'Conceptual Knowledge and Intuitive Experience: Schlick's Dilemma', in R. Giere and A. Richardson (eds) *The Origins of Logical Empiricism*, Minnesota Studies in the Philosophy of Science, Minneapolis, MN: University of Minnesota Press, 292-308.(Further addresses problematic aspects of Schlick's epistemology in the context of his commitment to scientific realism.)

McGuinness, B. (ed.) (1967) *Wittgenstein und der Weiner Kreis*, Berlin: Suhrkamp. (A record of the conversations between Wittgenstein, Waismann and Schlick in the late 1920s and early 1930s.)

McGuinness, B. (ed.) (1988) 'Moritz Schlick', special issue of *Synthèse* 64 (3).(A collection of essays exploring Schlick's philosophy from a variety of viewpoints.)

Oberdan, T. (1993) *Protocols, Truth and Convention*, Studien zur Oesterreichischen Philosophie, vol. 19, Amsterdam and Atlanta, GA: Rodopi.(An extended attempt to integrate Schlick's epistemology into the

background of his (largely unpublished) views on the philosophy of language.)

Oberdan, T. (1996) ‘Postscript to Protocols: Reflections on Empiricism’, in R. Giere and A. Richardson (eds) *The Origins of Logical Empiricism*, Minnesota Studies in the Philosophy of Science, Minneapolis, MN: University of Minnesota Press, 269-91. (Condenses many of the arguments of Oberdan (1993) in order to explore the nature of Schlick’s commitment to empiricism.)

Uebel, T. (1992) *Overcoming Logical Positivism from Within: The Emergence of Neurath’s Naturalism in the Vienna Circle’s Protocol Sentence Debate*, Studien zur Oesterreichischen Philosophie, vol. 17, Amsterdam and Atlanta, GA: Rodopi.(A detailed study of the philosophies of the leading protagonists in the protocol sentence controversy, based on primary sources and unpublished materials.)

Uebel, T. (1996) ‘Anti-Foundationalism and the Vienna Circle’s Revolution in Philosophy’, *British Journal for the Philosophy of Science* 47: 415-40.

Schmitt, Carl (1888-1985)

Carl Schmitt was a conservative critic of the Weimar Republic's liberal-democratic constitution. After Hitler's rise to power, he allied himself briefly to Nazism, and despite having fallen from favour and having revised his position even before the war, was never able to rehabilitate himself from the Nazi taint. Interned at Nuremberg in 1945, he was never brought to trial, but was banned from teaching thereafter. His critique of liberalism lay in liberalism's alleged inability to deal with the nature of politics. Schmitt continues to exert a vast influence on German public law, legal theory and political philosophy, as well as on European right-wing thought. His work remains important for liberals and opponents of liberalism for the challenges it poses to the neutrality of the liberal state and its legal order.

Schmitt claims that there is no rational way of deciding political conflict since the characteristic of politics is the distinction between irreconcilable friends and enemies in a pluriverse of ideologies. Liberalism attempts to construct a legal order which gets rid of political conflict by subjecting all individuals to the rule of law, the order of positive law of parliamentary democracy. But liberalism finds itself compelled to adopt a stance of neutrality between individual conceptions of the good, which allows groups of individuals to capture the legal order and thus the liberal state. The only way to end this conflict between private interest groups is for a (dictatorial) leader to reconstruct political and legal order on the basis of the vision he articulates of the substantive homogeneity of the people (*Volk*). Liberalism finds itself unable to preclude this event, since its commitment to legalism disables it from dealing with the 'state of exception', the moment when a sovereign decision is required to resolve a fundamental challenge to the *status quo* (see Rule of law (Rechtstaat); Law and morality §§1-3).

Schmitt's position is perhaps best explained in a work published in 1938, after he had fallen from Nazi favour. There he meditates on the passage from the publication of Hobbes' *Leviathan* (1651) to his own quite precarious situation.

According to Schmitt, Hobbes saw that the political order and stability of the modern state turned on eradicating political conflict within the state and displacing it to a matter of external affairs; he also saw that such order had to be maintained by a system of positive law founded on the myth of the great monster Leviathan. However, that myth cannot survive Hobbes' attempts to provide in addition a rational justification for political and legal order. Whenever Hobbes appeals to or makes room for individual reason, he subverts his own project (see Hobbes, T.). Hence Schmitt can conclude that what is required is a deeply anti-individualist or anti-liberal myth, which has the consequence that he seems to accept that those who do not fit a particular myth's otherwise arbitrary criteria for inclusion should expect no protection from the state.

Liberal thinkers today are deeply divided on such issues as state neutrality, the nature of legal order, and the proper response to the fact of pluralism and hence to potentially fundamental challenges to the legitimacy of liberal democracy. It may prove fruitful for the friends as well as the enemies of liberalism to take Schmitt seriously.

See also: Law, philosophy of; Radbruch, G.

DAVID LUDOVIC DYZENHAUS

List of works

Schmitt, C. (1922) *Politische Theologie: vier Kapitel zur Lehre von der Souveränität*, trans. and intro. G. Schwab, *Political Theology: Four Chapters on the Concept of Sovereignty*, Cambridge, MA: MIT Press, 1988.(This work claims that all secular political ideologies are disguised theologies.)

Schmitt, C. (1923) *Die geistesgeschichtliche Lage des Heutigen Parlamentarismus*, trans. and intro. E. Kennedy, *The Crisis of Parliamentary Democracy*, Cambridge, MA: MIT Press, 1998.(Here Schmitt explores an alleged contradiction between liberalism and democracy.)

Schmitt, C. (1928) *Verfassungslehre (Constitutional Theory)*, Berlin: Duncker & Humblot, 1989; trans. into French L. Deroche, intro. O. Beaud, *Théorie de la Constitution*, Paris: Presses Universitaires de France, 1993. (Schmitt's most elaborate work on legal theory.)

Schmitt, C. (1932) *Der Begriff des Politischen*, trans. and intro. G. Schwab, *The Concept of the Political*, New Brunswick, NJ: Rutgers University Press, 1976.(Here Schmitt claims that the fundamental distinction of

politics is that between friend and enemy.)

Schmitt, C. (1938) *Der Leviathan in der Staatslehre des Thomas Hobbes: Sinn und Fehlschlag eines politischen Symbols*, ed. G. Maschke, Cologne: Hohenheim; *The Leviathan in the State Theory of Thomas Hobbes: Meaning and Failure of a Political Symbol*, trans. G. Schwab and Erna Hilfstein, Westport, CT: Greenwood, 1996.(A translation of Schmitt's commentary on Hobbes' *Leviathan*.)

References and further reading

Note that any evaluation of Schmitt is complicated by the writer's understanding of the relationship between that work and his involvement with Nazism.

Bendersky, J. (1983) *Carl Schmitt: Theorist for the Reich*, Princeton, NJ: Princeton University Press.(A political biography of Schmitt.)

Caldwell, P. (1997) *The Theory and Practice of Weimar Constitutionalism*, Durham, NC: Duke University Press. (This book contains an important discussion of Schmitt's theory in the context of Weimar debates about law.)

Dyzenhaus, D. (ed.) (1997) 'Carl Schmitt', *Canadian Journal of Law and Jurisprudence* 10.(Special issue containing essays on Schmitt's political and legal theory.)

Dyzenhaus, D. (1997) *Legality and Legitimacy: Carl Schmitt, Hans Kelsen, and Hermann Heller in Weimar*, Oxford: Clarendon Press.(A critique of Schmitt's legal and political philosophy.)

McCormick, J. (1997) *Against Politics as Technology: Carl Schmitt's Critique of Liberalism*, New York: Cambridge University Press.(A discussion and critique of the main themes of Schmitt's work.)

Scheuerman, W. (1994) *Between the Norm and the Exception: The Frankfurt School and the Rule of Law*, Cambridge, MA: MIT Press.(This book contains an important discussion and critique of Schmitt's legal theory.)

Schwab, G. (1989) *The Challenge of the Exception: An Introduction to the Political Ideas of Carl Schmitt between 1921 and 1936*, New York: Greenwood Press.(First monograph in English on Schmitt's political and legal theory.)

Schopenhauer, Arthur (1788-1860)

Schopenhauer, one of the great prose-writers among German philosophers, worked outside the mainstream of academic philosophy. He wrote chiefly in the first half of the nineteenth century, publishing Die Welt als Wille und Vorstellung (The World as Will and Representation), Volume 1 in 1818 and Volume 2 in 1844, but his ideas became widely known only in the half-century from 1850 onwards. The impact of Schopenhauer's philosophy may be seen in the work of many artists of this period, most prominently Wagner, and in some of the themes of psychoanalysis. The philosopher most influenced by him was Nietzsche, who originally accepted but later opposed many of his ideas.

Schopenhauer considered himself a follower of Kant, and this influence shows in Schopenhauer's defence of idealism and in many of his central concepts. However, he also departs radically from Kant. His dominant idea is that of the will: he claims that the whole world is will, a striving and mostly unconscious force with a multiplicity of manifestations. Schopenhauer advances this as a metaphysical account of the world as it is in itself, but believes it is also supported by empirical evidence. Humans, as part of the world, are fundamentally willing beings, their behaviour shaped by an unchosen will to life which manifests itself in all organisms. His account of the interplay between the will and the intellect has been seen as a prototype for later theories of the unconscious.

Schopenhauer is a pessimist: he believes that our nature as willing beings inevitably leads to suffering, and that a life containing suffering is worse than nonexistence. These doctrines, conveyed in a literary style which is often profound and moving, are among his most influential. Equally important are his views on 'salvation' from the human predicament, which he finds in the denial of the will, or the will's turning against itself. Although his philosophy is atheist, Schopenhauer looks to several of the world religions for examples of asceticism and self-renunciation. His thought was partially influenced by Hinduism at an early stage, and he later found Buddhism sympathetic.

Aesthetic experience assumes great importance in Schopenhauer's work. He suggests that it is a kind of will-less perception in which one suspends one's attachments to objects in the world, attaining release from the torment of willing (desire and suffering), and understanding the nature of things more objectively. The artistic genius is the person abnormally gifted with the capacity for objective, will-free perception, who enables similar experiences in others. Here Schopenhauer adopts the Platonic notion of Ideas, which he conceives as eternally existing aspects of reality: the genius discerns these Ideas, and aesthetic experience in general may bring us to comprehend them. Music is given a special treatment: it directly manifests the nature of the will that underlies the whole world.

In ethics Schopenhauer makes thorough criticisms of Kant's theory. He bases his own ethical views on the notion of compassion or sympathy, which he considers a relatively rare quality, since human beings, as organic, willing beings, are egoistic by nature. Nevertheless, compassion, whose worldview minimizes the distinctness of what are considered separate individuals, is the only true moral impulse for Schopenhauer.

1 Life

Arthur Schopenhauer was born in Danzig in 1788 into a wealthy and enlightened business family. Following a childhood of sound school education and wide travel in Europe he attended the universities of Göttingen and Berlin, gravitating to philosophy after studying a number of subjects. He was most impressed by the writings of Kant and Plato, but also came across the Hindu Upaniṣads. These three he later claimed as his greatest influences. However, Schopenhauer's early impressions of the German university system were not favourable. He found the lectures of Fichte in particular to be pretentious and vacuous. While his inheritance from his father allowedSchopenhauer financial security, he developed a contempt for professional university philosophy which eventually became focused on Hegel, its leading figure.

The most creative period of Schopenhauer's life was the decade 1809-18, when he was in his twenties. He gained his doctorate in 1813 with a dissertation entitled *Über die vierfache Wurzel des Satzes vom zureichenden Grunde (On the Fourfold Root of the Principle of Sufficient Reason)*, a work which he always regarded as integral to his philosophy and which he revised substantially for republication in 1847. In 1816 he published a short work *Über das Sehn und die Farben (On Vision and Colours)*, which had its origin in a collaboration with Goethe.

The work which outshines all Schopenhauer's others, *Die Welt als Wille und Vorstellung (The World as Will and Representation)*, was published in 1818. It aimed to present a complete philosophical system, starting from a modified Kantian idealism and a metaphysics of the will embracing both the self (the microcosm) and the world (the macrocosm), and moving on to present original doctrines concerning aesthetics, ethics and the nature of human existence. Schopenhauer adhered to the philosophy of this work for the remainder of his life, and revised it for publication in 1844 with the addition of a second volume of elucidatory essays that more than doubled its length. There is little genuine intellectual development in Schopenhauer beyond 1818. His later writings give more reflective expression to the same set of doctrines.

Schopenhauer made an abortive attempt to begin a career at the University of Berlin in 1820, but scheduled his lecture at the same hour as Hegel's and had no audience. His hatred for Hegel's philosophy and the university system became intense. His own work, which he considered a great contribution to philosophy, went almost unnoticed. After a period of instability Schopenhauer settled in Frankfurt am Main in 1833, and remained there, leading a largely solitary life, until his death in 1860. His publications during this period were *Über den Willen in der Natur (On the Will in Nature)* (1836), *Die beiden Grundprobleme der Ethik (The Two Fundamental Problems of Ethics)* (1841), the second, two-volume edition of *The World as Will and Representation* (1844), the second edition of *On the Fourfold Root* (1847) and *Parerga und Paralipomena (Parerga and Paralipomena)* (1851). In the final decade of his life he prepared second editions of *On the Will in Nature* and *The Two Fundamental Problems of Ethics*, and in 1859 a third edition of *The World as Will and Representation*.

All Schopenhauer's publications from the 1830s onwards were designed to defend and amplify the ideas set out in *The World as Will and Representation*. *On the Will in Nature* attempted to support the metaphysical doctrine of the will with empirical evidence taken from the various sciences. *The Two Fundamental Problems of Ethics* is a compilation of the separately written essays *Über die Freiheit des menschlichen Willens (On the Freedom of the Human Will)* (1839) and *Über die Grundlage der Moral (On the Basis of Morality)* (1840). These were composed as entries to essay competitions set in Norway and Denmark. Because of the anonymity of the competitions, Schopenhauer could not rely on a full exposition of his philosophical system. The result is a pair of well-argued, self-contained essays which make an interesting contribution to ethics.

After the publication of *Parerga and Paralipomena*, a wide-ranging collection containing substantial philosophical essays, polemical pieces and popular aphorisms, Schopenhauer's philosophy was more widely recognized, and in old age he began to enjoy something of the popularity which continued for half a century after his death. During this time he became through his writings one of the greatest intellectual figures in European culture, and his influence may be traced in different ways in, among others, Wagner, Nietzsche, Tolstoy, Hardy, Freud,Jung, Proust,Thomas Mann and Wittgenstein.

2 Early work

At university Schopenhauer took greatest interest in the writings of Kant and Plato, and in certain doctrines of Hinduism. Volume 1 of his surviving *Manuscript Remains* (1966-75) shows that from around 1813 he was trying to produce a critical synthesis of these sources. Schopenhauer worked initially with a dichotomy between 'empirical consciousness' and what he called 'better consciousness'. He associated empirical consciousness with appearance, individuality and suffering, and saw the better consciousness as the experience of a higher reality in which the mind could penetrate beyond appearances, lose its sense of individuality, and enter a state free of suffering. The idea had both religious and aesthetic associations. Kant's influence is present in Schopenhauer's use of the appearance/thing-in-itself distinction and his attempt to characterize the empirical world in terms of the a priori forms of space, time and causality imposed by the subject of experience. The Platonic influence shows in the notion of a higher, pain-free cognition of a timeless reality lying beyond the empirical. From the Hindu writings he adopted the doctrine of the veil of *māyā*, the view that ordinary consciousness is enmeshed in illusion, and the idea that at a fundamental level the distinctness of individuals is illusory (see Monism, Indian). At this stage Schopenhauer was prone to conflate his disparate sources. For example, he assumed that Kant's thing-in-itself andPlato's Ideas were the same - a mistake which he later corrected, but whose effects his theory never entirely lost (see the opening of *The World as Will and Representation*, Third Book).

During the same period Schopenhauer wrote his dissertation, *On the Fourfold Root*, taking as his theme the principle of sufficient reason which states '*nihil est sine ratione cur potius sit quam non sit*' (nothing is without a

ground or reason why it is rather than is not). This principle may concern distinct species of ‘reason’ or ‘ground’, and hence different species of explanation. Schopenhauer seeks to clarify matters by mapping out four distinct kinds of explanatory principle, which he calls the sufficient reason of becoming, the sufficient reason of knowing, the sufficient reason of being and the sufficient reason of acting. The framework of *On the Fourfold Root* is Kantian. Schopenhauer uses the dichotomy of subject and object, in which objects are the known and the subject is the knower that can never itself be an object of knowledge. What can be known as objects are representations (*Vorstellungen*) which the subject has. The four kinds of explanation in the dissertation concern different classes of representations and the connections between them.

Schopenhauer followsKant in describing empirical consciousness as consisting of representations organized by the a priori forms of space and time. The empirical content that fills these forms is matter, appearing to the subject as distinct spatiotemporal objects. Schopenhauer’s first kind of connection among representations is the ‘principle of the sufficient reason of becoming’, which asserts that every state that appears must have resulted from a change that preceded it. This version of the principle is thus the ‘law of causality’. Schopenhauer’s discussion of causality builds on but also criticizesKant’s, and is the longest and most successful section of *On the Fourfold Root*. Space and time also yield the distinct explanatory principle which Schopenhauer calls the ‘principle of the sufficient reason of being’. This is supposed to cover reason-giving in mathematics.Schopenhauer here relies onKant’s idea that mathematics involves non-empirical, a priori intuition of spatial position and temporal succession. Space and time themselves are described as objects or representations for the subject, and the connections in space and time cognized a priori are termed relations of ‘mathematical necessity’. Concepts are held to be another distinct class of representations. Schopenhauer calls them ‘representations of representations’, regarding them as derivative from perception by a process of abstraction. Concepts enable the subject to make judgments, and the ‘principle of the sufficient reason of knowing’ states that if a judgment is to express knowledge, it must be related to a ground - which may lie in perception, in inference from another judgment, or in the possibility of experience or thought as such. Finally,Schopenhauer states the ‘principle of the sufficient reason of acting’, or law of motivation, which says that every act of will is related to a motive which causes it. The subject’s own will is the unique object of experience which this form of the principle concerns. Schopenhauer later extended his account of the will, but in general regarded his early analysis of the forms of the principle of sufficient reason as indispensable for understanding the rest of his thought.

3 *The World as Will and Representation*, First Book

Schopenhauer divides his main work into four books, both in the original Volume 1 and in the parallel, elucidatory Volume 2. Volume 1 also contains a long appendix entitled ‘Critique of the Kantian Philosophy’. Schopenhauer regards himself as a follower of Kant, but he has many criticisms to make. For instance, he is scathing in his attack on Kant’s architectonic ambitions, his style of writing and his account of the Ideas of soul, world and God - Ideas of the ‘unconditioned’ supposedly produced by ‘pure reason’ itself, but which Schopenhauer finds suspiciously amenable to the parochial Christian tradition. This ‘Critique of the Kantian Philosophy’ is closely linked to Schopenhauer’s concerns in the rest of the work.

In the First Book Schopenhauer is primarily concerned with epistemology. He defends transcendental idealism, a doctrine whose outline and terminology he takes from Kant. Transcendental idealism states that the world of empirical things is a world of objects existing for the subject’s experience, not existing in itself; and that the reality of the empirical world (our representations) consists in its being organized in space, time and causality, the necessary principles of connection among representations, whose origin is in the subject. Of the Kantian categories Schopenhauer retains only causality. The empirical world consists of causally efficacious matter filling the forms of space and time.

A feature of Schopenhauer’s account which would have been anathema to Kant is his appeal to Berkeley’s philosophy. Schopenhauer argues that Kant’s transcendental idealism shares with Berkeley the claim that the world of empirical things is mind-dependent, the difference being that Berkeley’s contribution is exhausted by this insight, whileKant adds the account of the a priori conditions of experience to explain objectivity. Schopenhauer had researched Kant’s *Critique of Pure Reason* thoroughly, making an extensive comparison of the second edition with the first, which had fallen into neglect. He alleges that Kant’s commitment to idealism is more clearly apparent in the first edition but that he wavers in the second, in fear of being classified as a Berkeleyan.

Schopenhauer uses a number of arguments for idealism which are variants on arguments found in Berkeley (§3). He suggests that idealism is the only viable alternative to scepticism about the external world, that to imagine a world existing without the subject is impossible, and that realism commits its proponent to a world existing in the subject's representations and to a parallel, redundant mind-independent world. But he chiefly relies on his concepts of 'subject' and 'object'. All experience requires that there be both - 'No subject without object' and 'No object without subject' are, he says, self-evident truths. Schopenhauer appears confident that this is sufficient to show that material objects would not exist without a subject.

Schopenhauer's account of the subject of experience is significant. This subject, he claims, can never be an object of experience. It is not identical with the person (since persons are, at least in part, bodily, and bodies are objects of experience). It is not identical with any part of the spatiotemporal, empirical world - not any individual thing within the world. Rather, he says, 'each of us finds himself as this subject' ([1818, 1844] 1969: 1, 5).Schopenhauer uses a number of images in describing the subject: it is like an eye that cannot see itself, or the focal point at which light rays are concentrated by a concave mirror.

At the same time Schopenhauer is keen to stress that each person, while 'finding' themselves as this pure subject, is a bodily thing within the world of objects. There is a deliberate tension in his account here. Throughout Schopenhauer's philosophy idealism is set in contrast with a blunt form of materialism. Materialism, he says, cannot be the whole truth because it cannot account for the subject's experience: 'materialism is the philosophy of the subject which forgets to take account of itself'. Yet in the same passage he maintains: 'It is just as true that the knower is a product of matter as that matter is a mere representation of the knower' ([1818, 1844] 1969: 2, 13). From an objective standpoint there is ultimately only matter in space and time - and this is true also of oneself considered from an objective standpoint. But from a subjective standpoint, he thinks, we must embrace transcendental idealism and its conception of the pure, non-objective subject.

Schopenhauer makes a sharp distinction between perceptual and conceptual representations. He claims (heretically, as far as Kant is concerned) that causality, one of the basic organizing forms of experience, is not conceptual. Like space and time, it is for Schopenhauer a form of intuition (*Anschauung*). Intuition or perception is the awareness of particular, causally connected, spatiotemporal objects through the senses. Concepts, for Schopenhauer, are quite different from this. Their role is in discursive thought or judgment which may have linguistic expression, and thus in reasoning. (He accusesKant, perhaps with some justification, of not clearly separating the discursive role of concepts from their alleged role in organizing perception.) Schopenhauer has a clear view about the distinction between humans and other animals: it is simply that humans alone have concepts, language and reason. Animals have understanding (*Verstand*), however, which is the ability to perceive a world of objects, and is different not in kind but only in degree from human understanding. An interesting sub-theme in Schopenhauer is the kinship of humans and animals, rooted in his conviction that the possession merely of reason provides no grounds for regarding a species as superior. Schopenhauer regards reason (*Vernunft*) as a secondary capacity, whose 'abstract, discursive concepts… have their whole content only from the knowledge of perception, and in relation to it' ([1818, 1844] 1969: 1, 35).

4 *The World as Will and Representation,* Second Book

The Second Book moves away fromKant and brings into play Schopenhauer's conception of the will. He argues that all processes in nature are fundamentally a kind of striving or end-seeking (usually unconscious) for which the term 'will' is the most appropriate. The governing aim here is metaphysical - the will provides Schopenhauer with an account of the nature of the world-in-itself, including the underlying nature of the individual human being. At the same time he believes that evidence from animal behaviour, psychology, the natural sciences and ordinary human experience gives confirmation of his view.

The argument begins with the question: How is one aware of one's own body? Schopenhauer's previous account of the subject in the First Book severed the knower from the known: the possessor of empirical knowledge surveyed the totality of objects comprising the spatiotemporal world, but was at no place within that world. But if I am such a subject, the body I call mine will be for me simply 'an object among objects', and I will not understand its movements except by the kind of observation and inference I apply to all empirical processes. As Schopenhauer says, this is not how things are: I do not generally relate to my body's movements in this way. I understand them 'from inside', and I understand them as my 'will'.

Schopenhauer's view of action is firmly anti-dualist. 'Act of will' is not, for him, the description of any purely mental event:

> The act of the will and the action of the body are not two different states objectively known, connected by the bond of causality; they do not stand in the relation of cause and effect, but are one and the same thing, though given in two entirely different ways, first quite directly, and then in perception for the understanding.
>
> ([1818, 1844] 1969: 1, 100)

This means that the account Schopenhauer gave earlier of the subject and its relation to a world of objects is now seen as inadequate. As subjects of action, we are bodily: when someone acts, the manifestation of will occurs directly in their body.

Schopenhauer goes further and states that 'the whole body is nothing but objectified will'. Here we must be careful. He calls the process of digestion, for example, one in which will manifests itself. But he does not mean that the digestive system develops and functions in conscious or rational pursuit of a goal. The term 'will' applies equally to 'blind' processes, and Schopenhauer wishes to regard the whole body as an expression of will only in the sense that its processes, such as digestion, can be explained by the end they serve for the organism. This part of his philosophy centres around his conception of the will to life (*Wille zum Leben*). Life is not an end which is consciously or rationally chosen by living things - primarily, this is not even the case with living things that are conscious and rational - but their morphology, behaviour and psychology convince Schopenhauer that life is the end for which they are organized. He paints a vivid picture of the whole of animate nature as forever striving, struggling and competing to live and to further life by producing offspring. Conscious, rationally caused willing in humans is merely the highest sophistication of this will to life that permeates nature.

Since human beings are as much organic expressions of the will to life as any other living thing, Schopenhauer thinks we should not overestimate the fact that we are subjects of knowledge, applying the classifications of space, time and causality, understanding the empirical world and making rational judgments. We are organisms whose brains and other physiological processes enable us to perform these functions, but the innermost core of the human being, as of every organism, is the will. Our mental processes are almost always at a deeper level subservient to the 'blind' will to life. Schopenhauer sees a complex interplay between this will and the conscious intellect. His idea that the intellect is often forced to follow the 'secret purposes' of an underlying will which it cannot control has been seen as a precursor of Freud's view of the unconscious. Freud also acknowledged that Schopenhauer had prefigured him in his treatment of sexuality. Schopenhauer links sexuality with the drive to reproduce, one principal way in which the will to life manifests itself throughout nature. He is not surprised, therefore, to find that sex is constantly, if 'secretly', present in human behaviour: 'It is the ultimate goal of almost all human effort; it has an unfavourable influence on the most important affairs, interrupts every hour the most serious occupations' ([1818, 1844] 1969: 2, 533).

The doctrine of the will is, however, supposed to extend well beyond animate nature. All natural processes, including those such as magnetism and gravity, are to be seen as manifestations of will: as if, with or without consciousness, with or without life, every bit of the world must be striving for some end or other. It is essential to Schopenhauer's thought that there is no supreme end, no grand design, purpose or meaning. There is no answer to the question why the will wills as it does. Nevertheless Schopenhauer discerns a kind of internal order within the world as will. There are a determinate number of natural kinds: nature is not haphazard but falls into distinct species and repeatable law-like processes.Schopenhauer says that in addition to the individual things and events of the empirical world, there are eternal Ideas, those forms, such as 'lion' or 'oak tree', which may be shared in by many individual lions or oak trees. He calls these Ideas 'grades of the will's objectification'.

Schopenhauer's 'world as will' is an exercise in metaphysics, an attempt to say how the world is in itself. Thus it becomes clear that his idealism has a different purpose fromKant's: he wants to separate the empirical world, existing only in the subject's representations, from the world as it is in itself, in order to give a positive account of the latter. Schopenhauer's 'key' to the will as thing-in-itself is provided by action. The 'inner' awareness I have of my own will manifesting itself in the body supposedly points me towards what exists beyond the realm of representations altogether. My 'inner' awareness shows me that what I *am* in myself is will. Rather than maintain a kind of theoretical egoism (or solipsism) in which I alone have this essence - a view which he regards as

irrefutable but mad - Schopenhauer advocates extending the same insight to the world as a whole.

Schopenhauer uses the expression ‘the will’, implying that the whole world of objects is the expression (or ‘objectification’) of one thing-in-itself. There cannot be a plurality of things-in-themselves: space and time are the principle of individuation (*principium individuationis*), but space and time do not apply to the thing-in-itself. Also the relation between the ‘in itself’ and the empirical cannot be causal, because causality has legitimate application only within the realm of representations. Schopenhauer says instead that the thing-in-itself (the will) ‘objectifies itself’ as a multiplicity of empirical things. This means simply that the world’s experienceable aspect (the world as representation) consists of many spatiotemporal things, whilst considered as it is in itself (the world as will) it is not composed of distinct individuals (see Monism).

This metaphysical system is beset by problems. Even if it makes sense to say that the same world is, under one aspect, divisible into many empirical things, and under another aspect one single thing, there remain at least two further difficulties. First, it is unclear how Schopenhauer is entitled to any knowledge at all of the thing-in-itself. And second, there is some mystery as to why the thing-in-itself is best called ‘will’. Schopenhauer says that our own willing is the nearest we ever come to knowing the ‘in itself’ of anything, so ‘will’ is the best term available to describe *the* thing-in-itself. Yet clearly neither the world-in-itself nor the majority of its phenomenal manifestations exhibit will in the way a human agent does. ‘Will’ threatens to become just a proper name for the world - but that in turn robs Schopenhauer’s theory of any power to understand or interpret the world. He sets out to ‘solve the riddle’ of the thing-in-itself. It remains unclear how that is to be achieved.

5 *The World as Will and Representation*, Third Book

Schopenhauer associates the will with misery. The will to life drives us on through an ever-ramifying set of desires and goals, but we reach no ultimate point or final satisfaction. To have desires unsatisfied is to suffer, to have needs is to be vulnerable to deprivation, and - the final irony - to be without needs usually brings only a state of empty boredom waiting to be filled by a further cycle of desires.

Yet there is alleviation of this condition, in the form of aesthetic experience, which is the topic of Schopenhauer’s Third Book. The unifying thought in Schopenhauer’s aesthetic theory is that one may have perceptual experience while the will is suspended. Such an occurrence is comparatively rare because the intellect is by nature a tool of the will and not prone to contemplating reality with the objectivity and freedom from desires that aesthetic experience demands. In aesthetic experience one is sunk in contemplation of some object and ceases to impose upon it the usual spatial, temporal and causal connections: ‘we no longer consider the where, the when, the why and the whither of things, but simply and solely the *what*’ ([1818, 1844] 1969: 1, 178).

Aesthetic experience, whether of nature or of art, has value for Schopenhauer because it is a temporary state of calm will-lessness, from which desire and suffering alike are excluded. But he also sees in aesthetic contemplation a cognitive change and an alteration in one’s sense of self. The subject in aesthetic experience becomes unaware of its separateness from that which it experiences: ‘the person who is involved in this perception is no longer an individual, for in such perception the individual has lost himself; he is *pure* will-less, painless, timeless *subject of knowledge*’ ([1818, 1844] 1969: 1, 179). At the same time, the object of the experience is not merely the individual spatiotemporal thing, but one of the eternal Ideas fixed in nature. Schopenhauer’s thinking here is as follows: individual empirical things are experienced when and only when the subject applies to its representations the a priori forms of space, time, and causality. But ordinary empirical knowledge is driven by the will: it consists in brain processes whose occurrence subserves the ends of the organism. Thus, if the intellect breaks away from its service to the will, it must leave behind the forms of space, time and causality. And since we experience individual spatiotemporal things only because we impose these forms, a timeless and spaceless experience must have as its object something beyond individual spatiotemporal things. This doctrine is the descendant of Schopenhauer’s earlier thinking about the ‘better consciousness’. He believes that, by freeing one’s intellect temporarily from the will, one gains a higher form of knowledge, and becomes a pure subject, objectively mirroring reality, and leaving behind one’s identification with any individual part of the empirical world.

Although he recognizes that nature provides many opportunities for this kind of elevated contemplation, Schopenhauer’s main interest is in art, where the spectator’s experience is mediated by the activity of the artist. For Schopenhauer the true artist is a genius, by which he means someone whose intellect - the capacity for

perception, not concept-use or reasoning - is abnormally powerful and able to function in greater isolation from the will. The genius can discern the universal in the particular with greater objectivity, as it were on behalf of the rest of us, and convey this insight in perceptible form. Schopenhauer says emphatically that conceptual thinking, by contrast, is unfruitful in the arts.

Schopenhauer's writing is informed by a wide knowledge and appreciation of the various arts. He discusses architecture, painting of different genres, sculpture, poetry and drama. Each art form has Ideas (or 'grades of the will's objectification') which it is especially able to reveal. This enables Schopenhauer to place the arts in a hierarchy. At the lower end, architecture, for example, enables us to know the fundamental Ideas of gravity, cohesion and rigidity. Landscape painting displays Ideas of inanimate nature and the plant world. Other paintings reveal the Ideas of different species of animals, and finally of human beings - but the highest art form, whose speciality is the Idea of humanity in all its complexity, is poetry.

At the very pinnacle stands tragedy, which has a special significance for Schopenhauer, since the eternal Idea of humanity which it makes known contains the most profound picture of the misery of our condition. Yet tragedy's value for Schopenhauer does not lie solely in this knowledge of what he calls 'the conflict of the will with itself'. He contends that the best tragedies present the hero's will turning away from life and adopting a sublime resignation in the face of suffering - thus exemplifying the attitude which he later argues is the only genuine 'salvation' (see Tragedy §5).

Schopenhauer's account of music is especially noteworthy. He suggests that while the other arts all attempt to stimulate knowledge of Ideas by depicting individual things, music 'is as *immediate* an objectification and copy of the whole *will* as the world itself is' ([1818, 1844] 1969: 1, 257). The will expresses itself as the phenomenal world; the same will expresses itself again in music, which bypasses the level of Ideas altogether. Thus Schopenhauer maintains that the appeal of music lies in its copying the patterns of striving and resolution of the will, to which we respond because they resonate with our lives as willing beings. But no personal strivings or sufferings enter into music:

> Music does not express this or that particular and definite pleasure, this or that affliction, pain, sorrow, horror, gaiety, merriment, or peace of mind, but joy, pain, sorrow, horror, gaiety, merriment, peace of mind *themselves*, to a certain extent in their abstract nature.
>
> ([1818, 1844] 1969: 1, 261)

Hence the value of music to the listener may also be that of a will-less calm (see Music, aesthetics of §6).

6 *The World as Will and Representation*, Fourth Book

The Fourth Book of *The World as Will and Representation* concerns ethics, taken broadly to include questions about the value of human existence, and what kind of happiness or salvation we may hope for. Schopenhauer gives his views here on the nature of morality and the question of free will and responsibility. These views are equally well stated (or perhaps better stated) in *On the Freedom of the Will* and *On the Basis of Morality*. It is only in *The World as Will and Representation*, however, that his conclusions about the human condition and his advocacy of denial of the will are presented with their full power.

Schopenhauer gives an elegant case for determinism, making a distinction between freedom to act, which one has when there are no impediments to one's doing what one wills, and freedom to will. The latter raises the important question: given that one willed to do such and such, could one have willed a different course of action? Schopenhauer suggests that acts of will are caused by a combination of one's permanent unchanging character and motives, which are representations of states of affairs in the world. No act of will could have failed to occur if the same motives and the same character had been present. In this sense there is no free will. Yet the feeling that we are responsible for our actions remains.Schopenhauer tries to account for this by saying that we feel responsible for what we are - for our unchanging, intelligible character, a character which is supposedly what we are beyond the realm of the empirical (see Free will).

This view has a Kantian ancestry. But in his moral theory Schopenhauer is generally critical of Kant. He is unimpressed by the notion of a categorical imperative, seeing it as a relic of the idea that commands are issued by the absolute authority of a divine being. He also questions Kant's linking of morality to rationality; other animals

for Schopenhauer should be accorded moral status, and the fact that they lack rationality is irrelevant (see Kant, I.). His own account of morality is simple. The one genuine moral impulse is compassion (or sympathy, *Mitleid*), which he says is present in each human being in some degree. Individuals governed by compassion apprehend the world and their place in it in a superior way, and they and their actions are good.

Each individual's character - which for Schopenhauer is inborn and unchanging - has some combination of the ingredients of egoism, compassion and malice. Malice is the impulse to seek another's harm, egoism the impulse towards one's own well-being and the avoidance of harm to oneself. Egoism is the greater part of most natures, according to Schopenhauer, since as manifestations of the will to life we must strive continually to survive and further ourselves. Pure malice is as much an exception as pure compassion, but both impulses must be accepted as facts of human nature. Compassion is the impulse to seek another's well-being and to prevent their suffering, and is grounded in a vision of the world which sets less store than usual on divisions between individuals: the good man, says Schopenhauer, sees everywhere 'I once more'. The metaphysical foundation for this is the claim that individuation is not an ultimate truth in the universe, since whatever appears as distinct at the empirical level is, at the level of the 'in itself', one and the same will. If individuality is thus illusory, compassion is more profoundly justified than egoism.

The core of Schopenhauer's pessimistic assessment of the value of human life lies once again in the opposition between the individual and the world as a whole. One's existence as a bodily, striving individual emerges not only as illusory (when viewed from the highest metaphysical vantage point), but as pernicious. To be an individual in which will manifests itself is inevitably to be open to suffering. But neither the striving of which the individual's life is full, nor the suffering which accompanies it, nor the temporary achievement of satisfaction which so soon induces boredom, has any higher point or value. Schopenhauer concludes that existence as an individual human being is always a worse alternative than nonexistence. 'In fact,' he proclaims, 'nothing else can be stated as the aim of our existence except the knowledge that it would be better for us not to exist' ([1818, 1844] 1969: 2, 605). He also argues that we inhabit the worst of all possible worlds.

Death is not something to fear, in Schopenhauer's view, since the world-will which expresses itself in this one fleeting individual continues undisturbed; death is 'the great opportunity no longer to be I…the moment of that liberation from the one-sidedness of an individuality that does not constitute the innermost kernel of our true being' ([1818, 1844] 1969: 1, 507-8). Despite this, Schopenhauer does not approve of suicide, which he regards as a failure to accept life on proper terms. The suicide affirms life, but revolts against the particular sufferings life contains. The contrasting attitude whichSchopenhauer advocates is denial of the will to life, an attitude which accepts the state of being alive but acquiesces in the suffering and the non-fulfilment of desires which it brings.

The question arises whether denial of the will to life is, paradoxically, something one can bring about at will. Schopenhauer appears to think not, since he talks of 'those in whom the will has turned and denied itself' and says denial of willing 'is not to be forcibly arrived at by intention or design…it comes suddenly, as if flying in from without' ([1818, 1844] 1969: 1, 404). It is as though the will to life is a distinct agency from the individual in whom it dwells. There are two routes to this turning of the will. One is the life of what Schopenhauer calls a saint. Such a person has knowledge of the illusoriness of individuation, and their individual will is 'quieted' thereby. The attitude of saints is one of such overwhelming compassion that they do not seek to further their own ends in distinction from those of others, nor to avoid harm to themselves. The other path leading to the turning of the will is to undergo suffering so great that one's will to life gives out spontaneously, while yet one is still alive. Those in whom the will has turned attain a state which Schopenhauer describes as 'resignation, true composure, and complete will-lessness'. He asks us to consider the blissful state of aesthetic contemplation, and then to imagine it prolonged: such, he claims, is the state of will-less self-denial that is the only genuine 'salvation' for humanity. Since our existence as bodily, striving individuals is one we would have been better without, the only remedy lies in achieving a vision of the world which attaches the lowest possible importance to one's individuality.

See also: Art, value of; Sexuality, philosophy of

CHRISTOPHER JANAWAY

List of works

Schopenhauer, A. (1946-50) *Schopenhauers sämtliche Werke*, ed. A. Hübscher, 7 vols, Wiesbaden: Brockhaus.

(The definitive edition of Schopenhauer's published works.)
Schopenhauer, A. (1977) *Arthur Schopenhauer: Zürcher Ausgabe*, ed. A. Hübscher, 10 vols, Zurich: Diogenes. (An accessible paperback version of the above.)
Schopenhauer, A. (1813, 1847) *Über die vierfache Wurzel des Satzes vom zureichenden Grunde*, trans. E.F.J. Payne, *On The Fourfold Root of the Principle of Sufficient Reason*, La Salle, IL: Open Court, 1974; trans. F.J. White as *Schopenhauer's Early Fourfold Root*, Aldershot: Avebury, 1997.(Schopenhauer's doctoral dissertation, which he later substantially revised, and regarded as integral to his philosophical system. The White translation also provides a commentary on the shorter 1813 version.)
Schopenhauer, A. (1816) *Über das Sehn und die Farben*; trans. E.F.J. Payne, *On Vision and Colors*, Oxford and Providence, RI, Berg, 1994.(A short work instigated byGoethe's theory of colours, but diverging from it. Not regarded as important to Schopenhauer's philosophy.)
Schopenhauer, A. (1818, 1844) *Die Welt als Wille und Vorstellung*, trans. E.F.J. Payne, *The World as Will and Representation*, New York: Dover, 2 vols, 1969; ed. D. Berman, trans. J. Berman as *The World as Will and Idea: abridged in one volume*, London: Everyman, 1995.(The main work of Schopenhauer's life, from whose central doctrines he never deviated. The only work containing his whole system of thought. The abridged version is greatly condensed and loses the grand sweep of the original, but makes the remaining argument more accessible to the general reader.)
Schopenhauer, A. (1836) *Über den Willen in der Natur*, trans. E.F.J. Payne, *On the Will in Nature*, New York and Oxford: Berg, 1992.(Schopenhauer's attempt to corroborate the doctrine of will stated in his main work by finding confirmations in science.)
Schopenhauer, A. (1841) *Über die Freiheit des menschlichen Willens*, trans. K. Kolenda, *On the Freedom of the Will*, Oxford: Blackwell, 1985.(A clear, self-contained essay on the problem of free will.)
Schopenhauer, A. (1841) *Über die Grundlage der Moral*, trans. E.F.J. Payne, *On the Basis of Morality*, Providence, RI and Oxford: Berhahn Books, 1995.(Schopenhauer's account of morality as based on compassion, preceded by a critique of Kant's ethics.)
Schopenhauer, A. (1851) *Parerga und Paralipomena*, trans. E.F.J. Payne, *Parerga and Paralipomena*, Oxford: Clarendon Press, 1974, 2 vols.(A late work containing a mixture of philosophical essays expanding on themes dealt with in earlier works, and various aphorisms and ruminations.)
Schopenhauer, A. (1966-75) *Schopenhauer: Der handschriftliche Nachlaß*, ed. A. Hübscher, Frankfurt am Main: Waldemar Kramer, 5 vols; trans. E.F.J. Payne, *Manuscript Remains*, Oxford, New York and Hamburg: Berg, 1988-90, vols 1-4. (Schopenhauer's unpublished notes, ranging throughout his career.)

References and further reading

Atwell, J.E. (1990) *Schopenhauer: The Human Character*, Philadelphia, PA: Temple University Press. (Concentrates on action, the will and ethical theory.)
Fox, M. (ed.) (1980) *Schopenhauer: His Philosophical Achievement*, Brighton: Harvester.(A collection of articles written from various points of view.)
Gardiner, P. (1967) *Schopenhauer*, Harmondsworth: Penguin; repr. Bristol: Thoemmes Press, 1997.(A comprehensive account ofSchopenhauer's philosophy.)
Hamlyn, D.W. (1980) *Schopenhauer*, London: Routledge & Kegan Paul.(A treatment of all major aspects of Schopenhauer's philosophy.)
Higgins, K.M. (1993) 'Arthur Schopenhauer', inR.C. Solomon and K.M. Higgins (eds) *Routledge History of German Philosophy, vol. 6: The Age of German Idealism*, 330-62.(A short introductory essay.)
Jacquette, D. (ed.) (1996) *Schopenhauer, Philosophy and the Arts*, Cambridge: Cambridge University Press. (Essays by various authors on Schopenhauer's aesthetics and its influence.)
Janaway, C. (1989) *Self and World in Schopenhauer's Philosophy*, Oxford: Clarendon Press.(Discusses the self, idealism, subjective and objective standpoints and Kant.)
Janaway, C. (1994) *Schopenhauer*, Oxford: Oxford University Press.(Written at an introductory level for non-specialists.)
Luft, E. von der (ed.) (1988) *Schopenhauer: New Essays in Honor of his 200th Birthday*, Lewiston, NY: Edwin Mellen.(A varied collection of articles with a thorough bibliography covering works by and on Schopenhauer.)
Simmel, G. (1907) *Schopenhauer and Nietzsche*, trans. H. Loiskandl, D. Weinstein and M. Weinstein, Amherst, MA: University of Massachusetts Press, 1986.(A classic comparative study of the two philosophers.)
White, F.C. (1992) *On Schopenhauer's Fourfold Root of the Principle of Sufficient Reason*, Leiden: E.J. Brill.(A commentary on *On the Fourfold Root*.)
Young, J. (1987) *Willing and Unwilling: A Study in the Philosophy of Arthur Schopenhauer*, Dordrecht: Martinus

Nijhoff.(A treatment of all major aspects of Schopenhauer’s philosophy.)

Schütz, Alfred (1899-1959)

Alfred Schütz, Austrian-American philosopher and social scientist. Combining ideas from Weber, Bergson and Husserl, Schütz developed a methodology for social science that integrated subjectivist, phenomenological elements with the causal-explanatory aspects of traditional objectivist approaches.

Schütz was born in Vienna, where he studied law, economics and sociology, and subsequently worked as a lawyer for a banking firm while pursuing his scholarly interests in his spare time. He emigrated to the USA in 1939, and was appointed professor at the New School for Social Research in New York in 1956.

Schütz's work stands in the German *geisteswissenschaften* tradition of Wilhelm Dilthey and Max Weber (§3), which insists that social science must methodologically reflect the specific nature of its human subject matter. Human and social reality - in particular, human action - is distinctive, in that it springs from the free flow of human subjectivity; hence, a method of interpretation (*verstehen*) of this subjective basis of action is required. Following Weber, Schütz endeavoured to combine *verstehen* with an objectivistic, causal approach. He held, however, that Weber's methodology lacked an adequate philosophical analysis of human subjectivity, a gap which Schütz sought to fill, first with elements from Bergson's doctrine of the flow of consciousness, later with tenets from Husserl's transcendental phenomenology. In turning to Husserl, Schütz broadened the investigation of the subjective springs of action into a phenomenological analysis of the acts of consciousness through which social reality is constituted. Schütz, however, did not conduct a full investigation of this domain within phenomenological brackets in his main work, *Der sinnhafte Aufbau der sozialen Welt (The Phenomenology of the Social World)* (1932), but rather examined the way social reality is grasped by subjects in the everyday, 'mundane' attitude. Indeed, Schütz eventually came to doubt the power of transcendental phenomenology fully to resolve the problems of *verstehen*; instead, in his US period, he strengthened the empirical side of his work by incorporating ideas from Mead and James, among others.

Schütz's phenomenological investigation brought to light certain characteristic structural properties of social reality as grasped by the agent (the agent's social *lebenswelt*). Social reality is organized around the experiencing subject in spheres of increasing concreteness and vividness the closer they are to the subject. Closeness here refers both to an item's spatial proximity to the subject and its relevance with respect to the subject's projects; Schütz stressed the pragmatic nature of our conceptualization of social reality, a nature manifest also in the way that this conceptualization is taken for granted and never exposed to theoretical doubts. The innermost circle of social reality is inhabited by persons with whom the subject is in direct face-to-face contact. This permits a full and detailed grasp of the other's unique individuality, a privileged mode of social cognition of which other modes are derivatives. Outside this zone lies the world of our contemporaries, people who exist simultaneously with ourselves but beyond our sensory horizon; then the world of our predecessors; and finally the world of our successors, the people who shall live after we are dead. Denizens of these spheres cannot be grasped in their full individuality, but must be understood in terms of general typifying concepts of increasing anonymity the further away they are from their source in the subject's experienced here-and-now. Schütz's analysis of the social *lebenswelt* shows his debt to Bergson in the stark contrast drawn between immediate intuitive experience and abstract conceptual understanding, and is tinged with a distinctly existentialist pathos when the unique significance of the face-to-face encounter is described.

The phenomenological account of the agent's subjective life-world is an indispensable part of social inquiry in its own right and, in addition, provides tools for the construction of an objective and genuinely theoretical stratum of social science. The social scientist must push further the process of typification that is present in inchoate form in the everyday agents' understanding of their life-world, analysing the social world in terms of idealized types that are derived from the agents' own abstract classifying concepts. In expounding a methodology of ideal types, Schütz borrowed heavily from Weber. Weberian ideal types are derived from experience by the one-sided accentuation of certain selected features, which are extrapolated to full logical perfection. Ideal types permit the formulation of law-like generalizations about human action, a paradigm case of which was, for Schütz as well as for Weber, the principles of rational action as found in particular in economics. Schütz also embraced Weber's methodological individualism and his insistence that science be 'value-free'.

The suggested approach is meant to solve what Schütz viewed as the crucial methodological problem for social science, namely how to construct an objective science of the agents' subjective meaning structures. The subjectivist desideratum is satisfied because the concepts used in the proposed science retain a link to the agents' naïve, unreconstructed notions. This link is formally guaranteed in terms of a general principle of methodology, which requires that the ideal concepts be understandable to the agents themselves in terms of everyday interpretations of action. On the other hand, the scientific concepts are refined and idealized, and hence satisfy the usual scientific standards with respect to clarity, consistency and precision.

Schütz conceived of his social theory as a 'sociology of everyday knowledge', but it differs from sociology of knowledge in the Marxist tradition in its lack of critical intent. It simply describes the structure of everyday social knowledge and the way it is utilized in the agents' understanding of their social life-world. Schütz's dual methodology thus offers no directions for the social scientist in the event of a clash between its objective and subjective components, such as a failure of an ideal type to be instantiated in the conduct of the agents, despite their endorsement of the same. This might occur if the agents were the victims of an ideological misconstrual of their own motivations. Indeed, the notion of 'ideology' and related concepts are signally absent from Schütz's methodology. Thus, the compatibility of the subjectivist and objectivist aspects of Schütz's methodology of social science is tacitly assumed, rather than effectively achieved.

See also: Explanation in history and social science; Phenomenological movement §3

FINN COLLIN

List of works

Schütz, A. (1932) *Der sinnhafte Aufbau der sozialen Welt*, Vienna: Julius Springer; Eng. trans. *The Phenomenology of the Social World*, Evanston, IL: Northwestern University Press, 1967.(Schütz's earliest and most penetrating work, in which he endeavours to ground Weber's interpretive sociological method on Husserlian phenomenology; rather dense.)

Schütz, A. (1962-66) *Collected Papers*, 3 vols; vol. 1, *The Problem of Social Reality*, 1962; vol. 2, *Studies in Social Theory*, 1964; vol. 3, *Studies in Phenomenological Philosophy*, 1966, The Hague: Martinus Nijhoff. (These volumes chiefly contain material from Schütz's US period. Volume 1 shows the development of Schütz's philosophy of social reality under the influence of US philosophers and social scientists. Volume 2 contains Schütz's essays within empirical social thought. Volume 3 contains Schütz's reflections, at times very critical, on the chief figures of phenomenology.)

Schütz, A. (1970) *Reflections on the Problem of Relevance*, R.M. Zaner (ed.), New Haven, CT: Yale University Press.(An edited version of a highly fragmentary manuscript found in Schütz's literary remains. Partially overlaps the *Collected Papers* and the 1973 book, but contains some new material on the phenomenology of social action.)

Schütz, A. and Luckmann, T. (1973) *The Structures of the Life-World*, Evanston, IL: Northwestern University Press.(Thomas Luckmann's edition of a manuscript left unfinished at the time of Schütz's death. Offers a more empirical and more accessible, but often less concise, version of some of the topics covered in the 1932 book.)

References and further reading

Collin, F. (1985) *Theory and Understanding: A Critique of Interpretive Social Science*, Oxford: Blackwell. (Contains a critique of Schütz's methodology.)

List, E. and Srubar, I. (eds) (1988) *Alfred Schütz: Neue Beiträge zur Rezeption seines Werkes (Alfred Schütz: New Contributions Regarding the Influence of His Work)*, Amsterdam: Editions Rodopi.(A collection of essays in German and English on the impact of Schütz's work.)

Natanson, M. (ed.) (1963) *Philosophy of the Social Sciences*, New York: Random House.(A selection of texts contrasting interpretive and positivist approaches to social reality.)

Thomason, B. (1982) *Making Sense of Reification*, Atlantic Highlands, NJ: Humanities Press.(Discusses the idealist and realist aspects of Schütz's methodology.)

Wagner, H.R. (1983) *Alfred Schütz: An Intellectual Biography*, Chicago, IL: University of Chicago Press. (Combines biography with a comprehensive treatment of the intellectual influences upon Schütz. Includes a good bibliography.)

Schumpeter, Joseph Alois (1883-1950)

*Schumpeter is best known for his seminal work in economics, but he also made important contributions to the fields of political science and sociology. He aimed to create a broad economic science that he called 'social economics' (*Sozialökonomik*), which was to include not only economic theory but also economic history, statistics and economic sociology. Inspiration for this project came in particular from his colleague Max Weber.*

As an economist Schumpeter is primarily remembered for his theory of the entrepreneur and for his emphasis on the dynamic aspects of economic reality: capitalism, as he saw it, meant first and foremost change. But Schumpeter also made a number of interesting observations about theorizing in economics and the role that vision plays in the work of the economist. His trenchant critique of the conventional theory of democracy and advocacy of a more realistic theory is generally recognized as a major contribution to political theory. Many of Schumpeter's most important ideas on economics and politics can be found in his bookCapitalism, Socialism and Democracy (1942), which has become something of a classic in the social sciences.

1 Life

Joseph Alois Schumpeter was born in the small town of Triesch, part of the Austro-Hungarian Empire. His family was Catholic and belonged to the local elite. At the age of ten Schumpeter moved with his mother and stepfather to Vienna, where he was educated at an exclusive imperial school. As a student at the University of Vienna, he soon focused on economics and received his doctorate in 1906 at a record early age. After a sojourn in England and Egypt (where he worked as a lawyer), in 1908 Schumpeter returned to his native country to pursue an academic career as an economist. Initially very successful, he soon earned a full professorship at the University of Graz - again at a record early age. During the years before the First World War Schumpeter produced three books, which established him as one of the country's most promising young economists. One of these books - *Theorie der wirtschaftlichen Entwicklung (The Theory of Economic Development)* (1911) - is considered to be his most successful work in economic theory.

After the First World War Schumpeter tried his luck first in politics - he was finance minister for a few months in 1919 - and then as a banker and investor. Both of these enterprises ended badly, and by the mid-1920s Schumpeter was happy to return to academia. In 1932 he emigrated for personal reasons to the USA, where he was to teach at Harvard University until his death. While in the USA, Schumpeter produced three major works: the enormous two-volume set *Business Cycles* (1939), the popular *Capitalism, Socialism and Democracy* (1942) and the extensive *History of Economic Analysis*, published posthumously in 1954.

2 Early works

As a young man Schumpeter produced three books: a short history of economics, a study of economics using a conventional static approach and a study of economics using an innovative dynamic approach. The most important and original of these is the third work, which appeared in 1911 and which was translated as *The Theory of Economic Development* in 1934. It is in this work that Schumpeter for the first time presented his famous theory of the entrepreneur. In a dynamic approach to the economy, he argued, one has to assume that there exists a force *within* the economy itself that can account for change and development. This force is embodied in the entrepreneur; and entrepreneurship is defined as 'the carrying out of new combinations' (Schumpeter [1911] 1934: 66). An entrepreneur does not invent, Schumpeter says, but innovates. The original innovation produces huge profits and is soon imitated by others, thereby leading to a wave of economic change that affects the whole economy.

According to Schumpeter, there exist five basic types of entrepreneurship: (1) the introduction of a new good; (2) the introduction of a new method of production; (3) the opening up of a new market; (4) the conquest of a new source of raw materials or half-manufactured goods; and (5) a new way of organizing an industry (Schumpeter [1911] 1934: 66). Schumpeter also emphasized that the entrepreneur is not primarily motivated by the prospect of gain - the textbook image of *homo economicus*. What makes the entrepreneur act is something else: 'the dream and the will to found a private kingdom', 'the will to conquer' and 'the joy of creating' ([1911] 1934: 93).

Finally, it should be noted that the first edition of *The Theory of Economic Development* included a last chapter in

which Schumpeter suggested that the basic principles of dynamic economics could also be applied to all other social sciences. Schumpeter even supplied an outline of what such an analysis would look like. The fear of being misunderstood made Schumpeter eliminate this evocative chapter in later editions of his work however.

3 *Capitalism, Socialism and Democracy*

After not having published a book for more than twenty years, Schumpeter's *Business Cycles* finally appeared in 1939. Its reception was a disappointment to Schumpeter however, and while planning his next work in economics he undertook an interim project: the book that was to become *Capitalism, Socialism and Democracy* (1942). This work contains a provocative analysis of Marx (§12), a novel approach to the theory of democracy and the famous prophecy that capitalism will soon be replaced by socialism: 'Can capitalism survive? No. I do not think it can.' 'Can socialism work? Of course it can' (Schumpeter [1942] 1994: 61, 167).

According to Schumpeter, there existed a number of reasons why capitalism was doomed, including the fact that the entrepreneur was everywhere being replaced by managers, the intellectuals were getting out of hand and the old sense of property was vanishing. Schumpeter was very unhappy about these developments since he personally detested socialism. He did not want his own personal feelings to interfere with his scientific judgement, however, and agreed that it was, in principle, possible to have a democratic form of socialism (see Socialism §§1, 4).

While Schumpeter's approach to Marx and his thesis of the demise of capitalism are still occasionally referred to, it is however his theory of democracy that has survived best. What Schumpeter called 'the classic doctrine of democracy' had, in his mind, a simplistic and naïve approach to politics. He characterized its basic tenets in the following way: 'the democratic method is that institutional arrangement for arriving at political decisions which realizes the common good by making the people itself decide issues through the election of individuals who are to assemble in order to carry out its will' (Schumpeter [1942] 1994: 250). According to Schumpeter, this approach to democracy implied a series of errors. For one thing, there is no such thing as a 'common good' that all people can agree on. Second, if such a common good none the less existed, it would provide no guidance whatsoever in individual cases. And third, the important role that politicians play in the political process tends to disappear if they are seen simply as those persons who carry out the will of the people. A much more realistic attitude to democracy, Schumpeter continued, would be to conceptualize it as a way of selecting those people who are to govern the masses. Schumpeter's 'alternative theory of democracy' is defined in the following way: 'the democratic method is that institutional arrangement for arriving at political decisions in which individuals acquire the power to decide by means of a competitive struggle for the people's vote' (Schumpeter [1942] 1994: 269) (see Democracy §1).

Some commentators have felt that Schumpeter's own theory of democracy is cynical and elitist since it views democracy as an active enterprise only for those who govern. This view has some truth to it and Schumpeter's book contains many contemptuous remarks about the political capacity of the average person. Other commentators, however, have felt that Schumpeter's approach to democracy is refreshingly realistic and non-ideological. Finally, it should be mentioned that Schumpeter's conception of the politician as a kind of political entrepreneur (a view that has its origins in the work of Max Weber) has inspired the public choice school (see Rational choice theory §7).

4 Schumpeter's last work

When Schumpeter died in January 1950 he had almost completed the extensive history of economics which he had begun work on in the early 1940s. In *History of Economic Analysis*, published posthumously in 1954, Schumpeter tried to summarize his experience of a life as an economist, and his book covers a vast array of topics. It is here, for example, that we find Schumpeter's discussion of what he called an economist's 'vision'. Behind all the great works in economics, Schumpeter said, there is always a specific vision of how the economy works and what its building blocks are. Schumpeter defines vision in the following way: 'a preanalytic act that supplies the raw material for the analytic effort' ([1954] 1994: 41).

It is also in *History of Economic Analysis* that we find Schumpeter's most extensive presentation of his own vision of economics. Inspired very early in his career by Max Weber, Schumpeter envisioned economics as a very broad kind of science, which would encompass not only economic theory but also economic history, economic sociology

and statistics. Schumpeter emphasized that ‘the all-round economist’ must in principle be knowledgable in all of these fields. His statement that if he were to relive his life and could choose only one field, he would chose economic history, shocked his colleagues. A careful reading of Schumpeter’s last work shows, however, that its author held economic theory in very high regard. Indeed, *History of Economic Analysis* contains a number of interesting observations about the role that theory plays in economics. Theory, for example, must be closely linked to empirical reality if it is to be useful. Schumpeter also emphasized that creative theorizing has little to do with the testing of hypotheses (this comes at a later stage of the scientific process), but is rather a way of initially conceptualizing economic reality and thereby adding to the economist’s ‘box of tools’. Thus, despite his many contributions to a variety of fields in the social sciences, Schumpeter’s primary interest as an economist was, and always remained, economic theory.

See also: Economics, philosophy of

RICHARD SWEDBERG

List of works

Schumpeter, J.A. (1911) *Theorie der wirtschaftlichen Entwicklung*, Leipzig: Duncker & Humblot; trans. of 2nd edn from 1926 by R. Opie, *The Theory of Economic Development*, Cambridge, MA: Harvard University Press, 1934.(Schumpter’s most famous work in economics, containing his well-known theory of entrepreneurship. Opie’s translation is the standard text, although considerably shorter.)

Schumpeter, J.A. (1939) *Business Cycles*, New York and London: McGraw-Hill, 2 vols.(Schumpeter’s giant work on business cycles is simultaneously a history of capitalism from 1787 to 1930s.)

Schumpeter, J.A. (1942) *Capitalism, Socialism and Democracy*, London: Routledge, 1994.(Probably Schumpeter’s best-known work, with brilliant passages on contemporary capitalism as well as on the problems democracy.)

Schumpeter, J.A. (1954). *History of Economic Analysis*, London: Routledge, 1994.(An enormous history of economic thought which is very useful. Published posthumously.)

References and further reading

Allen, R.L. (1991) *Opening Doors: The Life and Work of Joseph Schumpeter*, New Brunswick, NJ: Transaction Publishers.(An important but somewhat unwieldy biography.)

Harris, S.E. (ed.) (1951) *Schumpeter, Social Scientist*, Cambridge, MA: Harvard University Press.(An excellent anthology of articles by some of Schumpeter’s colleagues and pupils.)

Swedberg, R. (1991) *Joseph A. Schumpeter: His Life and Work*, Cambridge: Polity Press.(An intellectual biography of Schumpeter which emphasizes his vision of economics as a very broad kind of science.)

Schurman, Anna Maria van (1607-78)

The first woman to attend a Dutch university, Schurman studied ancient languages and theology. Her Latin treatise on the expedience of scholarship for women made this 'Star of Utrecht' the most famous female intellectual in seventeenth-century Europe. She was among the few women to publish views on Counter-Reformation controversies concerning predestination and transubstantiation. Her autobiography served as an apology for the Pietist sect, Labadism.

1 Philosophical and theological influences

Anna Maria van Schurman studied theology, Hebrew, Arabic, Chaldanean, Aramaic and Syrian under the mentorship of Gisbertus Voetius, Calvinist theologian of the University of Utrecht. Her first publication, *Amica dissertatio…* (1638), was a scholastic defence of women's fitness for scholarship, which included a correspondence with the Calvinist theologian André Rivet; part of it was translated as *The Learned Maid…* (1659). Her correspondence on the issue of predestination with the physician Johan van Beverwijck was published as *De vitae humanae termino* (1639). This treatise, the *Dissertatio* and selected letters were published in her collected works, *Opuscula…* (1648). Throughout the 1640s, Schurman communicated both with noted philosophers, such as Gassendi, Elisabeth of Bohemia, Marie le Jars de Gournay and, via Constantijn Huygens, Mersenne, and with such theologians as Jacob Lydius, Frederick Spanheim, Gisbertus Voetius and Daniel Heinsius. In response to Claude de Saumaise's treatise on transubstantiation, she wrote an epistolary essay commenting on the views of the Church Fathers. Though she had contact with Descartes and Queen Christina of Sweden, she came to reject both for religious reasons. During the 1650s, her Pietist leanings led her to abandon all scholarship except the study of Revelation, opting to pursue a life of faith in service to others (see Pietism). After joining Jean de Labadie's Pietist community in Amsterdam in 1669, she obtained a haven for her persecuted 'household', first via Abbess Elisabeth of Bohemia at Herford in Germany, later at Altona and Wieuwerd. Her autobiography and defence of Labadism appeared in 1673 as *EYKΛHRIA: seu Melioris partis electio…* In response to charges by Huygens, Drechssler and others against Labadism, she published *Korte Onderrichting…* and wrote *EYKΛHRIA II*, which, like her biblical criticism, was published posthumously.

2 *The Learned Maid…*

Schurman's scholastic treatise begins by defining and conservatively limiting terms: 'Maid' refers to a Christian woman of 'indifferent good wit', enough wealth to pay a teacher, sufficient free time to devote to study and the following goals for her scholarship: God's glory, personal salvation, virtue, happiness, the instruction of the family, and aid to other women. 'Scholar' refers to one 'given to the study of… all kinds of Learning' and Schurman is among the first to urge that no restrictions be placed on learning for women. By 'may be' she means 'expedient, fit, decent' - a normative notion weaker than 'ought'. She advances fifteen syllogistic arguments, with refutations of adversaries' arguments, to show that: (1) women have the appropriate nature, intellectual powers and desire for all the arts and sciences; (2) study will provide them with 'a similitude of Divine gladness', or pleasure; (3) it is fitting for women to study sciences instrumental to theology, since these lead to greater love of God; (4) scholarship is especially appropriate for those with leisure time to spend at home; and (5) learning will aid women in acquiring virtue and escaping vice and heresy.

The arguments for (1) rely upon the Aristotelian view that the Form of a human being is the rational soul and 'that which is in the whole Species or kind, is in every Individual… in Maids also' (1659: 8). Schurman's assumption that there is no deformation of the rational soul in woman rests, in part, on the argument that since some maids actually learn arts and sciences (together with the Aristotelian assumption that there are no acts without corresponding powers or principles), they must be naturally endowed with the powers or principles for the arts and sciences. The arguments for (2) turn on Aristotle's views about pleasure in the *Nicomachean Ethics*. The arguments for (4) and (5) constitute one of the earliest philosophical responses by a woman to the *querelle des femmes* tradition in moralistic writings of the period (see Feminism §2). These writings, derived from the Church Fathers, assume women's inconstancy, vanity, imprudence and small-mindedness; humanist writers such as Erasmus saw these vices as augmented by women's increased leisure time. Because Schurman turns diverse moralist assumptions to her own advantage in the course of her arguments, there are occasional tensions between

syllogisms. For example, Argument IV assumes that ‘leisure… is… the Mother of wickedness’ to conclude that women must have no free time, rather ‘continual employment’. Argument V, however, positively depicts the leisurely life of maids as one with the ‘Tranquillity and Liberty’ conducive to study. But these tensions are inherent in the conflicting positions of the moralists themselves. Schurman’s purpose is to show that all roads lead to women’s fitness for learning. Indeed, Schurman’s very production of this treatise stands as an argument for her conclusion. This work, together with Marie le Jars de Gournay’s *Egalité des hommes et des femmes* (1622), spawned numerous defences of women’s intellectual and moral capacities, many by women.

3 *De Vitae Termino*

This epistolary treatise (‘Concerning the Limit of Life’) responds to Beverwijck’s questions concerning predestination: (1) How can we reconcile God’s decree concerning the ‘limit of human life’ as an event ‘fixed and immovable’, when, as part of the contingent events of this world, this limit appears ‘changeable’? and (2) Of what use are doctors or medicines ‘if they have no power at all to delay the death which fate threatens?’ ([1648] 1650: 8). In answer to (1), Schurman argues that while a limit for each mortal creature’s life has been ‘unalterably established’ by a decree of God, nevertheless, because this limit is produced ‘freely and contingently’ by the power of the subordinate causes, ‘it can most lawfully be called changeable and movable’ ([1648]1650: 6). The decree of God, then, which necessarily determines all future events, also determines the manner in which events are to take place by secondary causes, either contingently or necessarily. So, from the standpoint of a creature, and in relation to secondary causes that could have been different, the limit of life may be called ‘contingent’. Like Calvin and Descartes, Schurman none the less admits that no one can understand how ‘divine providence bends the wills of men so sweetly and strongly that they freely and willingly choose and pursue that which God has decreed’ ([1648] 1650: 23). In reply to (2), Schurman argues that while ‘the art of medicine falls short of its own end, and the limit of the dying man is ever the same’, still we should take physicians to be ‘the instruments or rather "partners" of the very Author of life and health’ ([1648]1650: 26). For, the duty of a physician is not to ’cheat the hidden plans of God, but to break the hold of disease, to soothe pains, to restore natural strength, especially when they see that death is surely imminent’ ([1648]1650: 25). Versions of Schurman’s responses can be found in earlier Reformed authors such as Calvin, Voetius, Rivet and Johannes Wollebius, and in later ones such as Francis Turretin. The treatise’s uniqueness lies in the erudition displayed in support of her Calvinist position - ancient Stoic, Sceptic, Ciceronian, Platonic, Aristotelian, Judaic, early Christian, Muslim, Thomist and Italian Renaissance sources - coupled with its authorship by a woman, given its discussion of such controversial points of Calvinist doctrine (see Predestination).

EILEEN O’NEILL

List of works

Schurman, A.M. van (1638) *Amica dissertatio inter noblissimam virginem Annam Mariam a Schurman et Andream Rivetum de ingenii muliebris ad scientias et meliores literas capacitate (A friendly discourse between the renowned Miss Anna Maria van Schurman and André Rivet concerning the suitability of woman’s mode of thinking for the arts and sciences)*, Paris; repr. as *Dissertatio logica… (A logical discourse…)*, Leiden, 1641; French trans. as *Question célèbre… (The celebrated question…)*, Paris, 1646; English trans. C. B[arksdale] as *The Learned Maid…*, London, 1659.(One of the most important early modern defences of woman’s nature and fitness for scholarship; the fifteen syllogisms, with refutations of adversaries’ arguments, reproduced in the English translation, are joined with Schurman’s correspondence with Rivet on this topic in the Latin and French versions.)

Schurman, A.M. van (1639) *De vitae humanae termino (Concerning the limit of human life)*, Leiden; repr. as *De vitae termino in Opuscula*; Dutch trans. as *Paelsteen van den Tijt onzes Levens*, Dordrecht, 1639.(In response to the physician Beverwijck, this epistolary treatise defends the Calvinist position on predestination, drawing support from ancient pagan, Christian, Judaic, Islamic and Italian Renaissance sources.)

Schurman, A.M. van (1648) *Nobiliss. Virginis Annae Mariae à Schurman. Opuscula, hebraea, graeca, latina, gallica, Prosaica et metrica (The works of the renowned Miss Anna Maria van Schurman: Prose and verse in Hebrew, Greek, Latin and French)*, Leiden, 1650; repr. London, 1649; repr. Utrecht, 1652; new edn, D. Loeber, Leipzig, 1794.(Schurman’s collected works, including her *Dissertatio* and *De vitae termino*, a substantial correspondence with noted philosophers, theologians and other intellectuals, some of her poems, and elegies dedicated to her.)

Schurman, A.M. van (1673) ΕΥΧΛΗΡΙΑ*: seu melioris partis electio brevem religionis ac vitae ejus*

delineationem exhibens (Eukleria: or the choice of the better part, as presenting a brief sketch of her religion and life), Altona; repr. Dessau, 1782; Dutch trans. as *Eucleria of Uitkiezing van het Beste Deel*, Amsterdam, 1684, repr. Leeuwarden, 1978; English trans L. Richards as *Choosing the Better Part: Anna maria van Schurman (1607-1678)*, ed. M. de Baar, Dordrecht and London: Kluwer, 1996.(In the tradition of Augustine's *Confessiones*, this work combines Schurman's autobiography and defence of her faith.)

Schurman, A.M. van (1675) *Korte Onderrichtinge Rakende de Staat en de manier van leven der Personen die god t'samen vergadert en tot sijnen dienst vereenigt heeft door de bedieninge sijnes getrouwen dienstknechts Johannes de Labadie en sijner Broeders en Mede-Arbeiders Petrus Yvon en Petrus Dulignon (Short instruction concerning the state and the way of life of those persons whom god gathers and has united in his service through the actions of his faithful servant Jean de Labadie and his brothers and fellow-workers Pierre Yvon and Pierre Dulignon)*, Amsterdam.(In the year following Labadie's death, Schurman published this defence of the Labadist community, which had finally settled in Wieuwerd, Friesland, after being expelled from Herford and Altona in Germany.)

Schurman, A.M. van (1684) ΕΥΧΛΗΡΙΑ *II*, Amsterdam; Latin edn, Amsterdam,1685.(This posthumously published work, completed in the last days of Schurman's life, resumes the Labadist defence begun in the original ΕΥΧΛΗΡΙΑ, responding, in part, to Drechssler's criticisms of that earlier work.)

Schurman, A.M. van (1699) *Mysterium magnum oder Grosses Geheimnis (Mysterium magnum or the great mystery)*, Wesel.(The posthumous publication of Schurman's speculations about the future kingdom of God.)

Schurman, A.M. van (1732) *Uitbreiding over de drie eerste Capittels van Genesis (Elaboration on the first three chapters of Genesis)*, Groningen.(The posthumous publication of an example of Schurman's most cherished intellectual pursuit: Biblical criticism.)

References and further reading

Birch (Pope-Henessy), U. (1909) *A nna van Schurman: Artist, Scholar, Saint*, London: Longmans, Green & Company.(The most comprehensive discussion in English. As Moore notes, 'Schotel, Birch and Douma still provide the foundation of Schurman scholarship'.)

Bourignon, A. (1686) *Toutes les Oeuvres de Mlle. Antoinette Bourignon (All the works of Miss Antonette Bourignon)*, Amsterdam.

Douma, A.M.H. (1924) *Anna Maria van Schurman en de Studie der Vrouw (Anna Maria van Schurman and the study of woman)*, Amsterdam: H.J. Paris.(Treats Schurman on women's education; reviews secondary literature up to 1915.)

Drechssler, J.G. (1673) *Eukleria Eukeatos seu Melioris Partis Electio rescissa (Eukleria Eukeatos or the choice of the better part torn apart)*, Leipzig.(This Lutheran educator accused ΕΥΧΛΗΡΙΑ of claiming that internal illumination provides revelation independently of the Scriptures and that the pursuit of the sciences is morally useless.)

Gournay, M. le J. de (1622) *Egalité des hommes et des femmes (The equality of men and women)*, n.p.; English trans. E. O'Neill in J. Sterba (ed.) *Social and Political Philosophy in perspective: Classical Western Texts in a Feminist and Multicultural Perspective*, Belmont, CA: Wadsworth, 1994.(Like Schurman's *Dissertatio*, this is one of the most important early modern philosophical works on the issue of woman's moral and intellectual capacities.)

Huygens, C. (1670a) *Heusche Vermaaning, aan de doorlugtige en wel-edele juffr. Anna Maria van Schurman, om haar selven te kennen, en af te wijcken van Jean de Labadie (Serious admonition to the serene and noble Miss Anna Maria van Schurman, to know herself, and to break with Jean de Labadie)*, Amsterdam.(An anti-Labadist tract by a previous supporter of Schurman.)

Huygens, C. (1670b) *Sedige en sielveroerende aenspracke aen Juffr. Anna Maria van Schurman, om haer af te trecken van Jan de Labadie (Modest and soul-stirring appeal to Miss Anna Maria van Schurman, to pull her away from Jean de Labadie)*, Amsterdam.(Another important anti-Labadist tract by this Dutch poet, musician and correspondent of Descartes.)

Irwin, J.L. (1977) 'Anna Maria Van Schurman: From Feminism to Pietism', *Church History* 46: 48-62.(An examination of Schurman's changing religious views and their relation to their position on what role scholarship should play in a pious life.)

Irwin, J.L. (1989) 'Anna Maria van Schurman: Learned Woman of Utrecht', in K. Wilson and F. Warnke (eds) *Women writers of the Seventeenth Century*, Athens, GA: University of Georgia Press.(A survey of Schurman's life and works; excerpts from *The Learned Maid*, translated letters and and translated selection from ΕΥΧΛΗΡΙΑ are appended.)

Moore, C.N. (1994) 'Anna Maria van Schurman', in K. Aercke (ed.) *Women Writing in Dutch*, New York and London: Garland.(A survey of Schurman's life and works, a slightly corrupt text of *The Learned Maid*, a translation of a letter to Dorothea Moor and translated selections from ΕΥΧΛΗΡΙΑ are appended.)
Schotel, G.D.J. (1853) *Anna Maria van Schurman*, 's-Hertogenbosch: Gebroeders Muller.(Among the first book-length treatments of Schurman, it remains a central secondary source.)

Science in Islamic philosophy

Islam attempts to synthesize reason and revelation, knowledge and values, in its approach to the study of nature. Knowledge acquired through rational human efforts and through the Qur'an are seen as complementary: both are 'signs of God' that enable humanity to study and understand nature. Between the second and eighth centuries AH *(eighth and fifteenth centuries* AD*), when Muslim civilization was at its zenith, metaphysics, epistemology and empirical studies of nature fused to produce an explosion of 'scientific spirit'. Scientists and scholars such as Ibn al-Haytham, al-Razi, Ibn Tufayl, Ibn Sina and al-Biruni superimposed Plato's and Aristotle's ideas of reason and objectivity on their own Muslim faith, thus producing a unique synthesis of religion and philosophy. They also placed great emphasis on scientific methodology, giving importance to systematic observation, experimentation and theory building.*

Initially, scientific inquiry was directed by everyday practices of Islam. For example, developments in astronomy were influenced by the fact that the times of Muslim prayer were defined astronomically and its direction was defined geographically. In the later stage, the quest for truth for its own sake became the norm, leading to numerous new discoveries and innovations. Muslim scientists did not recognize disciplinary boundaries between the 'two cultures' of science and humanities, and individual scholars tended as a general rule to be polymaths. Recently, Muslim scholars have started to develop a contemporary Islamic philosophy of science by combining such basic Islamic concepts as 'ilm *(knowledge),* khilafa *(trusteeship of nature) and* istisla *(public interest) in an integrated science policy framework.*

1 Science and metaphysics

The Muslim inspiration for the study of nature comes straight from the Qur'an. The Qur'an specifically and repeatedly asks Muslims to investigate systematically natural phenomena, not simply as a vehicle for understanding nature but also as a means for getting close to God. In Surah 10, for example, we read:

> He it is who has made the sun a [source of] radiant light and the moon a light [reflected], and has determined for it phases so that you might know how to compute years and to measure [time]…in the alternative of night and day, and in all that God has created in the heavens and on earth, there are messages indeed for people who are conscious of Him.
>
> (Surah 10: 5-6)

The Qur'an also devotes about one-third of its verses to describing the virtues of reason. Scientific inquiry, based on reason, is thus seen in Islam as a form of worship. Reason and revelation are complementary and integrated methods for the pursuit of truth.

The philosophy of science in classical Islam is a product of the fusion of this metaphysics with Greek philosophy. Nowhere is this more apparent than in Ibn Sina's theory of human knowledge (see Ibn Sina §3) which, following al-Farabi (§3), transfers the Qur'anic scheme of revelation to Greek philosophy. In the Qur'an, the Creator addresses one man - the Prophet - through the agency of the archangel Gabriel; in Ibn Sina's Neoplatonic scheme, the divine word is transmitted through reason and understanding to any, and every, person who cares to listen. The result is an amalgam of rationalism and ethics. For Muslim scholars and scientists, values are objective and good and evil are descriptive characteristics of reality which are no less 'there' in things than are their other qualities, such as shape and size. In this framework, all knowledge, including the knowledge of God, can be acquired by reason alone. Humanity has power to know as well as to act and is thus responsible for its just and unjust actions. What this philosophy entailed both in terms of the study of nature and shaping human behaviour was illustrated by Ibn Tufayl in his intellectual novel, *Hayy ibn Yaqzan*. Hayy is a spontaneously generated human who is isolated on an island. Through his power of observations and the use of his intellect, Hayy discovers general and particular facts about the structure of the material and spiritual universe, deduces the existence of God and arrives at a theological and political system (see Epistemology in Islamic philosophy; Ethics in Islamic philosophy).

While Mu'tazilite scholars had serious philosophic differences with their main opponents, the Ash'arite theologians, both schools agreed on the rational study of nature. In his *al-Tamhid*, Abu Bakr al-Baqillani defines science as 'the knowledge of the object, as it really is'. While reacting to the Mu'tazilite infringement on the

domains of faith, the Ash‘arites conceded the need for objective and systematic study of nature. Indeed, some of the greatest scientists in Islam, such as Ibn al-Haytham (d. 1039), who discovered the basic laws of optics, and al-Biruni (d. 1048), who measured the circumference of the earth and discussed the rotation of the earth on its axis, were supporters of Ash‘arite theology (see Ash‘ariyya and Mu‘tazila).

The overall concern of Muslim scientists was the delineation of truth. As Ibn al-Haytham declared, ‘truth is sought for its own sake’, and al-Biruni confirmed in the introduction to his *al-Qanun al-mas‘udi*: ‘I do not shun the truth from whatever source it comes.’ However, there were disputes about the best way to rational truth. For Ibn Sina, general and universal questions came first and led to experimental work. He begins his *al-Qanun fi'l-tibb (Canons of Medicine)*, which was a standard text in the West up to the eighteenth century, with a general discussion on the theory of drugs. For al-Biruni, however, universals came out of practical, experimental work; theories are formulated after discoveries. But either way, criticism was the key to progress towards truth. As Ibn al-Haytham wrote, ‘it is natural to everyone to regard scientists favourably…. God, however, has not preserved the scientist from error and has not safeguarded science from shortcomings and faults’ (see Sabra 1972). This is why scientists so often disagree amongst themselves. Those concerned with science and truth, Ibn al-Haytham continued, ‘should turn themselves into hostile critics’ and should criticize ‘from every point of view and in all aspects’. In particular, the flaws in the work of one's predecessors should be ruthlessly exposed. The ideas of Ibn al-Haytham, al-Biruni and Ibn Sina, along with numerous other Muslim scientists, laid the foundations of the ‘scientific spirit’ as we have come to know it.

2 Methodology

The ‘scientific method’ (see Scientific method), as it is understood today, was first developed by the Muslim scientists. Supporters of both Mu‘tazilism and Ash‘arism placed a great deal of emphasis on systematic observation and experimentation. The insistence on accurate observation is amply demonstrated in the *zij*, the literature of astronomical handbooks and tables. These were constantly updated, with scientists checking and correcting the work of previous scholars. In medicine, Abu Bakr Muhammad al-Razi's detailed and highly accurate clinical observations in the early third century AH (ninth century AD) provide us with a universal model. Al-Razi was the first to observe accurately the symptoms of smallpox and described many ‘new’ syndromes. However, it was not just accurate observation that was important; equally significant was the clarity and precision by which the observations are described, as was demonstrated by Ibn Sina in his writings.

The emphasis on model construction and theory building can be seen in the category of Islamic astronomical literature known as *‘ilm al-haya*, or ‘science of the structure (of the universe)’, which consists of general exposition of principles underlying astronomical theory. It was on the strength of both accurate observation and model construction that Islamic astronomy launched a rigorous attack on what was perceived to be a set of imperfections in Ptolemaic astronomy (see Ptolemy). Ibn al-Haytham was the first to declare categorically that the arrangements proposed for planetary motions in the *Almagest* were ‘false’. Ibn Shatir (d. 1375) and the astronomers at the famous observatory in Maragha, Adharbayjan, built in the thirteenth century by Nasir al-Din al-Tusi, developed the Tusi couple and a theorem for the transformation of eccentric models into epicyclic ones. It was this mathematical model that Copernicus used to develop his notion of heliocentricity, which played an important part in the European ‘scientific revolution’.

Apart from the exact sciences, the most appropriate and interesting area in which theoretical work played an essential role was medicine. Muslim physicians attempted to improve the quality of *materia medica* and their therapeutic uses through continued theoretical development. Emphasis was also placed on developing a precise terminology and ensuring the purity of drugs, a concern that led to a number of early chemical and physical procedures. Since Muslim writers were excellent organizers of knowledge, their purely pharmacological texts were themselves a source for the development of theories. Evolution of theories and discovery of new drugs linked the growth of Islamic medicine to chemistry, botany, zoology, geology and law, and led to extensive elaborations of Greek classifications. Pharmacological knowledge thus became more diversified, and produced new types of pharmacological literature. As this literature considered its subject from a number of different disciplinary perspectives and a great variety of new directions, there developed new ways of looking at pharmacology; new areas were opened up for further exploration and more detailed investigation. Paper-making made publication more extensive and cheaper than use of parchment and papyrus, and this in turn made scientific knowledge much

more accessible to students.

While Muslim scientists placed considerable faith in scientific method, they were also aware of its limitations. Even a strong believer in mathematical realism such as al-Biruni argued that the method of inquiry was a function of the nature of investigation: different methods, all equally valid, were required to answer different types of questions. Al-Biruni himself had recourse to a number of methods. In his treatise on mineralogy, *Kitab al-jamahir (Book of Precious Stones)*, he is the most exact of experimental scientists. However, in the introduction to his ground-breaking study *India* he declares that ‘to execute our project, it has not been possible to follow the geometric method’; he therefore resorts to comparative sociology.

The work of a scholar of the calibre and prolificity of al-Biruni inevitably defies simple classification. He wrote on mineralogy, geography, medicine, astrology and a whole range of topics which dealt with the dating of Islamic festivals. Al-Biruni is a specific product of a philosophy of science that integrates metaphysics with physics, does not attribute to either a superior or inferior position, and insists that both are worthy of study and equally valid. Moreover, the methods of studying the vast creation of God - from the movement of the stars and planets to the nature of diseases, the sting of an ant, the character of madness, the beauty of justice, the spiritual yearning of humanity, the ecstasy of a mystic - are all equally valid and shape understanding in their respective areas of inquiry. In both its philosophy and methodology, Islam has sought a complete synthesis of science and religion.

Polymaths such as al-Biruni, al-Jahiz, al-Kindi, Abu Bakr Muhammad al-Razi, Ibn Sina, al-Idrisi, Ibn Bajja, Omar Khayyam, Ibn Zuhr, Ibn Tufayl, Ibn Rushd, al-Suyuti and thousands of other scholars are not an exception but the general rule in Muslim civilization. The Islamic civilization of the classical period was remarkable for the number of polymaths it produced. This is seen as a testimony to the homogeneity of Islamic philosophy of science and its emphasis on synthesis, interdisciplinary investigations and multiplicity of methods.

3 Revival attempts

At the end of the twentieth century, scholars, scientists and philosophers throughout the Muslim world are trying to formulate a contemporary version of the Islamic philosophy of science. Two dominant movements have emerged. The first draws its inspiration from Sufi mysticism (see Mystical philosophy in Islam) and argues that the notions of ‘tradition’ and the ‘sacred’ should constitute the core of Islamic approach to science. The second argues that issues of science and values in Islam must be treated within a framework of concepts that shape the goals of a Muslim society. Ten fundamental Islamic concepts are identified as constituting the framework within which scientific inquiry should be carried out, four standing alone and three opposing pairs: *tawhid* (unity), *khilafa* (trusteeship), *‘ibada* (worship), *‘ilm* (knowledge), *halal* (praiseworthy) and *haram* (blameworthy), *‘adl* (justice) and *zulm* (tyranny), and *istisla* (public interest) and *dhiya* (waste). It is argued that, when translated into values, this system of Islamic concepts embraces the nature of scientific inquiry in its totality; it integrates facts and values and institutionalizes a system of knowing that is based on accountability and social responsibility. It is too early to say whether either of these movements will bear any real fruit.

See also: Aristotelianism in Islamic philosophy; al-Farabi; Greek philosophy: impact on Islamic philosophy; Ibn Sina; Islamic philosophy: transmission into Western Europe; Neoplatonism in Islamic philosophy; Religion and science; Scientific method

ZIAUDDIN SARDAR

References and further reading

Bakar, O. (1996) ‘Science’, in S.H. Nasr and O. Leaman, *History of Islamic Philosophy*, London: Routledge, ch. 53, 926-46.(Discussion of some of the main thinkers and principles of science in Islam.)

Dani, A.H. (1973) *Al-Biruni’s India*, Islamabad: University of Islamabad Press.(Al-Biruni’s research on the people and country of India.)

Fakhry, M. (1983) *A History of Islamic Philosophy*, London: Longman, 2nd edn.(A general introduction to the role of reason in Islamic thought.)

Hill, D. (1993) *Islamic Science and Engineering*, Edinburgh: Edinburgh University Press.(The classic work on the practical aspects of Islamic science.)

Hourani, G. (1975) *Essays on Islamic Philosophy and Science*, Albany, NY: State University of New York Press. (An important collection of articles on particular theoretical issues in the philosophy of science.)

Hourani, G. (1985) *Reason and Tradition in Islamic Ethics*, Cambridge: Cambridge University Press.(A discussion of the clash between reason and tradition in Islamic culture as a whole, especially in ethics.)

Ibn Tufayl (before 1185) *Hayy ibn Yaqzan (The Living Son of the Vigilant)*, trans. S. Oakley, *The Improvement of Human Reason Exhibited in the Life of Hai Ebn Yokhdan*, Zurich: Georg Olms Verlag, 1983.(This translation of *Hayy ibn Yaqzan* was first published in 1708.)

Kirmani, Z. (1992) 'An Outline of Islamic Framework for a Contemporary Science', *Journal of Islamic Science* 8 (2): 55-76.(An attempt at conceptualizing modern science from an Islamic point of view.)

Leaman, O. (1985) *An Introduction to Medieval Islamic Philosophy*, Cambridge: Cambridge University Press.(A general approach to the role of philosophy in Islam.)

Nasr, S.H. (1993) *The Need for a Sacred Science*, Richmond: Curzon Press.(An argument for the significance of religion in any understanding of science.)

Pines, S. (1964) 'Ibn al-Haytham's Critique of Ptolemy', in *Actes du Xe Congrès internationale d'histoire des sciences*, Paris: Ithaca.(One of the most important works in Islamic astronomy.)

Sabra, A.I. (1972) 'Ibn al-Haytham', in C.C. Gillispie (ed.) *Dictionary of Scientific Biography*, New York: Charles Scribner's Sons, 6th edn. (An excellent introduction to the thought and work of Ibn al-Haytham.)

Said, H.M. (ed.) (1979) *Al-Biruni Commemorative Volume: Proceedings of the International Congress held in Pakistan, November 26-December 12, 1973*, Karachi: Hamdard Academy.(Contains numerous papers discussing all the major works of al-Biruni.)

Saliba, G. (1991) 'The Astronomical Tradition of Maragha: A Historical Survey and Prospects for Future Research', *Arabic Sciences and Philosophy* 1 (1): 67-100.(A study of a particularly well-developed period of astronomical research in the Islamic world.)

Sardar, Z. (1989) *Explorations in Islamic Science*, London: Mansell.(Some contemporary debates on the nature of Islamic science.)

Young, M.J.L., Latham, J.D. and Serjeant, R.B. (1990) *Religion, Learning and Sciences in the Abbasid Period*, Cambridge: Cambridge University Press.(The leading work on the most important period for science in the Islamic world.)

Science, 19th century philosophy of

In the nineteenth century, science was organized, it tested and confirmed positive knowledge of the natural world and achieved remarkable theoretical development and hitherto unimagined practical application. Science drove industry and free enterprise, and became a powerful catalyst in the battle between defenders of knowledge as power and advocates of knowledge as love.

Fruitful scientific theories and observations were plentiful. Darwin, Wallace and Spencer caused a revolution in biology. Faraday, Maxwell and Hertz contributed seminal ideas in electromagnetic theory. Hermann von Helmholtz studied the physiology of tones and discovered a principle of the conservation of force. Lyell's efforts established geology as a science. Ernst Mach argued for the elimination of absolute space in favour of a space and time consisting of observable relations between things, thus providing incentive for Einstein's theory of relativity. Sir John Herschel added many observed double stars to the growing catalogue of celestial bodies.

These and other observational, theoretical and applied achievements in nineteenth-century science were replete with philosophical consequences. Until the nineteenth century natural philosophy and science coexisted as a single discipline. Now science and traditional philosophy drew apart. Some held that henceforth science would deal with the world revealed in experience, and philosophy with the world existing (if any does) beyond what we experience. Others (including prominent scientists) were unwilling to yield to philosophy licence to speculate beyond the limits of what could be ascertained by means of observation and experimentation: even if science and philosophy were no longer one unified intellectual enterprise, philosophy had a substantial role to play in philosophizing about science.

To satisfy changing expectations, a new intellectual discipline was created in the nineteenth century: the philosophy of science. Unlike previous philosophy, whose subject matter was everything that is (or is not), the philosophy of science had a distinct and determinate subject matter: theoretical texts and experimental and observational reports of scientists (the word 'scientist' having been invented by William Whewell).

Theoretical scientific systems and their logical structure were one focus of attention. Science was also said to discover laws. Were such laws timeless and exceptionless truths about nature, or simply convenient, economical ways of cataloguing information? These laws were discovered (or invented) generalizations that provided tested information about nature. This discovery and confirmation relied upon the method of induction - thought by most nineteenth-century philosophers of science to have a logic - to involve decisions concerning the validity or invalidity of inferences based on knowledge from experience. Was this alleged logic trustworthy? These questions exemplify the complex problems concerning the epistemic reliability of scientific explanation.

1 'Man is the interpreter of nature, science, the right interpretation'

William Whewell provides the aphoristic watchword expressive of the new nineteenth-century preoccupation with the success of science. For him, as for many others, certain fundamental scientific theories had achieved the status of complete and unrevisable truths, among them, the physical synthesis of Newton and the wave theory of light. These theories stood as exemplars of what could be achieved in the inductive natural sciences. Study of the historical development of these, and other, sciences reveals the methodological strategies involved in the establishment of theories now seen to be established truth. Above all, the history of the inductive sciences confirms that science is progressive, moving always towards theories that are simpler, more comprehensive and more accurate as instruments of prediction. New sciences may emerge, but the established ones now require, not additional evidential support, but only refinement.

Others, like Ernst Mach, pointed out that the major theoretical triumphs of science should be construed as invitations to strive for completion, for the end of a science is an ideal, not an achievable result. Nevertheless, Mach shared the general acceptance of the fact that science is progressive, that what we know of nature now is somehow more reliably known than heretofore. A completed science may never be attained but this admission need not entail that we have not progressed in our positive knowledge of how the world works.

John Stuart Mill argued for the epistemic ascendancy of science by claiming that because science has discovered some invariable laws of nature, we can justly conclude that all phenomena are governed by such laws. The

progress of science is characterized by the discovery of lawfulness.

Progress as completion of theory, as epistemic ideal, as discovery of laws - three complex philosophical versions of the Victorian aphorism: nature as text receives its best reading from the precise and comprehensive testimony of science.

2 ‘The general property of the age’

John Stuart Mill used this phrase to describe what Auguste Comte conceived as the concluding phase of human cultural history: entry into a positivistic age in which both religion and metaphysics no longer function as reliable guides to human action. That which transcends human experience is to be replaced forever by a cooperative social reliance upon the facts disclosed by the sciences.

Many of the thinkers of the century can be classed as positivists, not in the sense of those in the Vienna Circle (see Logical positivism §4; Vienna Circle) but in the sense of subscribing to the main tenets of a new, although not always explicitly articulated, philosophical programme involving five commitments. First, Victorian positivists accepted Whewell’s aphorism: only science provides reliable knowledge of nature. Second, these philosophers were convinced that metaphysics could no longer be regarded as a trustworthy source of knowledge. Third, for the majority of nineteenth-century philosophers it was almost a truism to hold that, whatever is to count as reliable knowledge must make contact with the world of human experience. Whewell’s ‘facts’, Mach’s ‘elements’, Mill’s ‘particulars’ and Helmholtz’s ‘perceptions’ (see Helmholtz, H. von §2) form a family of conceptualizations of meaningful human experiences. For these thinkers (Mill may be the exception) the basic units of experience are laden with theory. Whewell’s facts come about by means of the controlled imposition of an idea upon given data. Mach’s elements are creatures of reflection upon common human episodes. Helmholtz’s perceptions are produced by unconscious inferences based upon rememberings. Nevertheless, the implications of an idea can be tested, an abstraction can be decomposed into its participating components and an inference can be found to be valid or invalid. The point is that, if illusions and imperfect conceptual constructions can somehow be accounted for, it will have to be agreed that no legitimate knowledge can proceed from a foundation other than what is directly available in human contact with the environment.

Fourth, mathematics was elevated to a high position both in education and in the work of scientists. The arguments on behalf of the primacy of mathematics came from Whewell and his contemporaries at Cambridge, later from Henri Poincaré, and from the new formal logicians, for example Stanley Jevons, George Boole and Augustus DeMorgan. Fifth, an element of pragmatism became prominent in the thinking of philosophers of science. Part of what is typical of the application of scientific inductive method involves the making of decisions based on considerations of utility, workability and other values not necessarily associated with truth.

Nineteenth-century philosophers of science did not engage the five points as planks in a philosophical platform. There were important exceptions. Not everyone developed an ambiguity-free positivism of the sort described. What is involved is not so much intellectual acceptance of a set of propositions; rather, our philosophers of science in the main assumed the five points as philosophical obligations, matters of trust to be taken to heart if one is to understand science. From a historical perspective the five points ought to be construed as sign posts directing the investigation of how philosophers of the nineteenth century understood science.

3 The status of scientific laws

In an age of sociopolitical empires, when half the known world was being cut up piecemeal by piratical European governments, it was not difficult for philosophers to call attention to the universal tendency of human beings to generalize, to project universal claims on the basis of small samples. Ever since Aristotle, such generalization had been identified as induction. The formula, though vacuous as a classification, persisted: deduction moves from the general to the particular; induction, from the particular to the general (see Inductive inference; Logic in the nineteenth century). Whewell and Sir John Herschel believed they had discovered an inductive ‘propensity’. Herschel characterized this tendency as an

> irresistible impulse of the mind to generalize *ad infinitum*, when nothing in the nature of limitation or opposition offers itself to the imagination, [and as an] involuntary application of the law of continuity to fill up, by the same ideal substance of truth, every interval which uncontradicted experience may have left blank in our

inductive conclusions.

(1857a: 198)

Whewell was also convinced that ‘ the mind has a perpetual propensity to consider… individual propositions as cases of more general ones and to frame and contemplate these latter’ (1830).

In one of his notebooks, Whewell illustrates this propensity by means of a humorous, if logically badly flawed, example:

[We start with the particular observation that:]

(1) That apple tree is full of fruit;

[and then move from this observation through the following generalizing steps:]

(2) All apple trees of this district are full of fruit;
(3) All the fruit trees of the kingdom are this season full of fruit;
(4) This season is one of great fruitfulness;
(5) All things are at present going on very well;

[to this grand generalized conclusion:]

(6) All things always go on well[!]

(1830)

This propensity provides the nerve centre of inductive logic. The conclusions of generalizing, which in common affairs are maxims or rules of behaviour, in the sophisticated reaches of scientific inference, are laws. For most philosophers of science a law represents a proposition expressing universally invariant regularity of some sort, and possessing what is now called ‘counterfactual force’ (see Counterfactual conditionals; Laws, natural §2). Even if no piece of lead is ever again submerged in *aqua regia*, it is still true that ‘lead dissolves in *aqua regia*’.

Mill and Whewell, believing as they did that the logic of induction possesses resources for distinguishing valid from invalid inductive inferences, held also that the results of valid inductions, laws, are necessary truths. Herschel provided the conceptual flag carried by the growing regiment of scientific realists who regarded the observed regularities as fixed, ‘ the laws of nature are not only permanent, but consistent, intelligible, and discoverable with such a modest degree of research, as is calculated rather to stimulate than to weary curiosity’ (1830) (see Scientific realism and antirealism §1). Whewell argued that the acceptability of necessary truths was demonstrated by the self-contradictory character of their logical opposites. Mill held a somewhat more complex view of the epistemic status of natural laws.

Pause to reflect. The world could, for all we know, be very different. There is a clear case for concluding that any generalization over items of experience is contingent; it might, again for all we know, be false. More seriously, all known regularities could be replaced, in accordance with a plan or a successful wager on random possibilities. Such logical possibilities led some to propose that natural laws are endowed with a somewhat attenuated epistemic status.

Mill insisted that natural laws are inductively adequate descriptions of observed empirical regularities. Karl Pearson refined this idea: scientific laws, unlike civil laws which are prescriptive, are descriptive. They are valid for all normal human beings, not just for special communities at special times, and are ‘ unchangeable so long as their perceptive faculties remain at the same stage of development’ (1911). A natural law is not descriptive of repeated sequences of events existing outside us, but is what Pearson calls a ‘mental formula’. This initially perplexing idea had been given fuller treatment by several earlier nineteenth-century philosophers of science; surprisingly, Mill is one of them.

Although Mill thought that a scientific law can be understood as part of our knowledge of nature, he also thought that a law can be taken to be a ‘memorandum for our guidance’. As knowledge of nature, a law ‘registers’ facts; as a memorandum, it can be used as a guide to inference. Mill wrote: ‘ General propositions are merely registers of such [inductive] inferences already made, and short formulae for making more’ (1843). We do not draw inferences

from generalizations, but *in accordance with* generalizations taken as licensing certain inferences. We arrive at generalizations by collecting resembling particulars. No one of these particulars is specially privileged, and when we conclude that 'all particulars of sort *x* are particulars of sort *y*', and then go on to infer that some specific *x* is a *y*, we are, says Mill, 'merely deciphering our own notes'.

In commenting upon the generally accepted characterization of a law as a description, Mach prefers the term 'restriction on expectations' to point to the biological significance of scientific laws: ' in origin, the "laws of nature" are restrictions that under the guidance of our experience we prescribe to our expectations' ([1905] 1976). More fully:

> A law always consists in a restriction of possibilities, whether as a bar on action, as an invariable course of natural events, or as a road sign for our thoughts and ideas that anticipate events by running ahead of them in a complementary manner. Galileo and Kepler imagine the various possibilities of free fall and planetary motion, trying to guess those that correspond to observation and making them more precise. The law of inertia which assigns uniform rectilinear motion to a body once all forces disappear, selects one of infinitely many possible thoughts as the decisive one for our ideas.
>
> (Mach 1976: 352)

Mach is not denying that laws pick out observed invariable regularities. But human knowers are biological beings whose expectations are motivated by the need for preservation; laws of nature are instruments of orientation, rules of preservation. It is true that the refinement of laws requires abstraction from given experiences, simplification, idealization of facts and mental decomposition of facts into more cognitively manageable units, which allow us to deal with facts more efficiently (see Idealizations). Uniform and uniformly accelerated motions of masses, stationary (steady) thermal and electric currents, are examples of such idealized factual simples; as creatures of mind, they never occur in reality. 'Natural science may be viewed as a kind of collection of instruments for the intellectual completion of any partially given facts or for the restriction, as far as may be, of expectations in future cases'. This is one formulation of Mach's famous thesis of the economy of thought.

The instrumentalism of laws argued for by Mill, Pearson and Mach appears entirely consistent with the pragmatic positivism of the time. Instrumentalism is anti-metaphysical - in Mach and Mill it is even anti-theoretical. In stressing the specific human purposes that are to be fulfilled by seeking and employing laws, our three philosophers of science join forces with other nineteenth-century thinkers, who, like Herbert Spencer and Thomas Huxley, provide related arguments on behalf of the thesis that man is a progressive being.

It can be argued that instrumentalism is incoherent because it really cannot state its own case. Were there no ways in which nature invariably worked before the idealizing techniques of science came into play? Yes and no, answers the instrumentalist. Yes (perhaps), but who, as human knower, cares? No, because laws can only be useful for human knowers; laws restrict what we can know and do. Instrumentalism becomes a matter of perspective; it is a manner of viewing the efficacy of known natural law. But the perspective, though it provides orientation and patterns of adaptation, also displays limitations on what we can know if what we can know is truly only what we can successfully use. The laws, in addition, must be construed as mind-dependent, possible only because of the imposition and manipulation of ideas.

4 The logic of induction

The characteristics of inductive inference (as learning from experience) were much discussed by nineteenth-century philosophers of science. Whewell and Mill debated the question of just what constituted proper induction; Herschel provided the first rules of inductive procedure since Francis Bacon; Mach argued that conclusions of inductive arguments are as vacuous as those of deductive arguments; Mach, Pearson and Ernst Friedrich Apelt stressed the role of abstraction in scientific enquiry; James Clerk Maxwell championed a complex anti-theoretical methodology based on reasoning from analogies.

The core consideration here is the unwavering confidence of some philosophers of science in what they regarded as the logic of induction, a confidence emphatically proclaimed in Mill's statement that 'there are… certain and universal inductions; and it is because there are such, that a Logic of Induction is possible'. Herschel's nine 'general rules' were intended to provide guidance in the search for natural laws (or common causes of similar

phenomena), by permitting deliberate experimental manipulation of data in order to disclose the causes or laws. The suggestion is that, properly applied, the rules will generate valid inductive inferences.

Mill appropriates Herschel's rules in his simplified and generalized 'canons of induction', now better known as 'Mill's methods'. Induction, which for Mill is always inference from particular to particular, is based on consideration of sets of data constructed out of empirically noted regularities and similarities. Factors in a set are varied so that agreements and differences are detectable. These features may refer to invariable antecedence or consequence of certain factors, increase or diminution of certain factors, discovered proportionality of certain factors, and the like. In cases of valid inference, the causal analysis of varying factors will lead to discovery of a true cause, or a law of nature.

Mill thus follows one lead from the insightful writings of Herschel on scientific method. Whewell follows another. Herschel's rules and Mill's methods deal with what we might call 'local induction', generalization over known cases, and inference to others of the same kind not now known. Here, the only evidence that an induction is valid is that it covers the known cases, that it is empirically adequate. It had been recognized since the work of David Hume that any inferences beyond the known cases to all cases of the same (and sometimes different) kind (of global inductions), are hazardous at best, logically unwarranted at worst. Herschel suggests a test of such inferences that Whewell makes the cornerstone of his theory of method. Herschel writes:

> The surest and best characteristic of a well-founded and extensive induction… is when verifications of it spring up, as it were, spontaneously, into notice, from quarters where they might be least expected, or even among instances of that very kind which were at first considered hostile to them. Evidence of this kind is irresistible, and compels assent with a weight which scarcely any other possesses.
>
> (1830: 170)

Whewell calls this epistemic phenomenon a 'consilience of inductions', and regards the test as one capable of establishing the validity of some inductive inferences:

> The instances in which [consilience] has occurred, indeed, impress us with a conviction that the truth of our hypothesis is certain. No accident could give rise to such an extraordinary coincidence. No false supposition could, after being adjusted to one class of phenomena, exactly represent a different class, where the agreement was unforeseen and uncontemplated. That rules springing from remote and unconnected quarters should thus leap to the same point, can only arise from *that* being the point where truth resides.
>
> (1858a: 88)

Whewell regarded hypotheses (he called them 'colligations') as initiated by a mental act: the imposition of an idea on some body of data.Mach complimented Whewell on what Mach mistakenly thought was Whewell's concession to the role of freely fantasized and intuited hypotheses. But Whewell was more cautious. There are at least two kinds of restraint imposed on the selection of colligating ideas: success in the past employment of similar ideas, and fruitful discussions in the peer group of relevant scientists. Thus constrained, the idea used is to 'save the appearances', to provide a useful conduit to predictions of phenomena of the same kind as are in the original body of data. Successful prediction is thus one evidential test of a hypothesis, but it still operates on the local level, serving only to enlarge the items in the original set of data. Mill argued that successful prediction is of no evidential value; it only provides more items falling under the same description.

Whewell, the arch scientific realist, finally staked everything on global inductions, those successfully colligating *and* consiliating a hypothesis, adequately covering initial data, predictions of data of the same kind, and unexpected data originally thought to be different in kind, but now seen to be data of the same kind as items in the initial set. Recall Newton's hypothesis of universal gravitation, which covers motions of earthly bodies and of heavenly bodies, and the motions of the tides. Mill thought all induction was local induction, induction by enumeration. Although he agreed with Whewell that induction arrives at necessary truths, his comment on the value of consilience was deliberate silence.

A well-recognized feature of deduction is that deductive conclusions from premises convey no new information that is not contained in the premises but merely*analyse* the premises. The test of a valid deductive inference is that one should not be able to deny the conclusion, given the premises, without logical contradiction. It had also been a

long-standing view that inductive inferences yield conclusions that convey new information. It was generally thought that induction moves from the known to the unknown. Valid induction is discovery. The problem, underscored by the devastating analysis of causality by Hume, is to understand the respect in which induction can, like deduction, have a logic, for the conclusions of inductive inferences (as generalizations about the world) seem to involve contingent claims. Alternative conclusions are always *logically* possible.

Mach argued that inductive inference, like deductive inference, yields only vacuous conclusions. If it is simply a matter of piling up more and more instances, as in Mill's favoured induction by simple enumeration, then nothing really new is ever learned that could count as a law. All that one learns is that some data sets are larger than others. Mach's objection may apply to local induction of Mill's sort, but what about global induction? Here also Mach thinks no new information is obtained, for although in consilience data of an apparently new kind are involved, the consilience itself shows that they are not really different in kind from the initial data. If consilience really happens, it converts new data to old, and only extends the magnitude of the enumeration.

William Stanley Jevons delivers what looks like a fatal blow to any claim to have a logic of induction. Granting that science yields only approximations of the given in experience, involving as it does those factors stressed by Mach and Whewell - abstraction, simplification, idealization - any perfect fit between theory and observation must be grounds for suspicion that we are wrong, not for confidence that we are right. In the consideration of rival hypotheses, which are always possible in science, Jevons introduced subjectivist probability theory.

How, then, did believers in the possibility of a logic of induction justify induction?

5 The justification of induction

Herschel thought that the risk of error involved in Jevons's challenge is infinitesimal when we are dealing with ' the unbroken experience of all observers, in innumerable instances'. A conclusion from induction can 'enjoy [no] more than a provisional security'. Such a small risk should produce at least a practical, as distinct from a mathematical, certainty, 'in all physical inquiry, and in all transactions of life' (1857b). This pragmatic appeal to limited certainty would satisfy Mach, whose views on science as providing instruments for orientation have been discussed above. It would also satisfy Karl Pearson, and the biologically oriented Herbert Spencer. It would not, however, satisfy Mill or Whewell.

Whewell, the scientific realist, believed that science does establish laws that are universal truths, that contingent claims become necessary ones. The warrant for this belief is partly pragmatic: Whewell does appeal to the success of certain kinds of methodological technique as disclosed by the history of the inductive sciences. He also seeks a metaphysical justification of his scientific realism.

The historical record of attempts to justify induction reveals that for the most part philosophers rested content with an appeal to the principle of the uniformity of nature. Why trust induction? Inductive conclusions are based on invariable sequences of events. Why take them to be invariable? Nature is uniform. But the argument is viciously circular. We can only know that nature is uniform because we observe lots of sequences taken to be invariable. Or the argument appeals to metaphysical knowledge: nature is uniform because God made it that way. Or, as Herschel thought, the principle of uniformity is 'an axiom drawn from the inward consciousness of our nature, by involuntary generalization. We acknowledge it expressly or impliedly in every instant of life'.

Mill accepts a form of this argument as his justification of induction. We discover what we take to be laws of nature, invariable repeated sequences of events. What is the simplest assumption that can guarantee that we are right in our assumption that we have in fact discovered laws of nature? The assumption that there are laws of nature (that nature is uniform). This argument is not circular. It asks for the simplest assumption from which all of the known laws of nature can be deduced. The assumption is not inductively established, it is posited. Perhaps it is the best we can hope for (see Induction, epistemic issues in).

6 'The ultimate problem of all philosophy'

Whewell admitted with characteristic candour that what he called 'the ultimate problem of all philosophy' is profound, and probably unsolvable. This did not stop him from offering a solution. The problem is as old as scepticism: by means of human mental resources, we arrive at what we take to be knowledge of causal

connections, of natural necessity, how can we be sure that what we frame in mind exists in reality outside us? Whewell's committed Christianity provided him with the motivation that ensured a solution. The orders of human knowing correspond to real orders because God makes it so. Scientists share God's thoughts when they hit upon natural laws. That those laws involve independent realities is thus assured, since what God thinks, in these cases at least, he enacts.

Whewell's realism thus shifted its troubled burden on to a metaphysics as old as Leibniz's. He carries that burden alone. For Helmholtz, Hertz, Emil DuBois-Reymond and others, mechanism, the view that all natural phenomena conform to the fundamental principles of mechanics, provides the answer. Although there is no qualitative similarity between our sensations and ideas and their causes in the external world, there are the activities of nerve impulses, which mediate between the two. Physiology, as a part of physics, provides the argument for this reductionist position. Our ideas do not directly correspond to external realities, but both these realities and our ideas are, ultimately, to be viewed as the motions of atomic parts over distances.

We know that Mach was strongly opposed to the mechanical philosophy. He was also strongly opposed to considering Whewell's ultimate problem as a genuine one. Taking his cue from biology, he argued for the thesis that science provides human orientation in a complex world, not metaphysical guidance. Science aids us in our efforts to get along in this world, helps us to realize purposes. But there can be no question of identifying the economies of our successful ideas with an established pattern existing independently of us and holding for all time.

Herbert Spencer's similar biological interests led him to hold a view not unlike that of Mach. It may be true that the order of human knowing reflects accurately and without obdurate distortions the order of reality independent of the human mind, but this is because human beings evolve in the direction of better intellectual adaptation; any correlation between experience and law is not fixed. In the biological domain all change is towards equilibrium; this includes changes in knowledge as well as changes in social contexts.

The ultimate problem is basically the problem of scepticism: how can I trust that I know anything at all? Whewell leaves us in perplexity. There are those who will believe that the avenue to God's mind is forever closed for repairs. The same persons will likely also accept that those who coupled pragmatism and biology created, not an avenue, but a naturalistic horizon whose philosophical enticements constitute one of the benefits of later philosophy of science.

See also: Darwin, C.R.; Geology, philosophy of §1; Hertz, H.; Wallace, A.R.

ROBERT E. BUTTS

References and further reading

Butts, R.E. (ed.) (1989) *William Whewell: Theory of Scientific Method*, Indianapolis, IN, and Cambridge: Hackett Publishing Company.(An anthology of Whewell's basic writings in philosophy of science, with an introduction and bibliography.)

Butts, R.E (1993) *Historical Pragmatics. Philosophical Essays*, Boston Studies in the Philosophy of Science, vol. 155, Dordrecht: Kluwer.(Essays 8-13 supply details of nineteenth-century philosophy of science, especially that of Herschel, Mill and Whewell. The interpretive analyses are sometimes complicated.)

Cannon, W.F. (1961) 'John Herschel and the Idea of Science', *Journal of the History of Ideas* 22: 215-39.(Good background. Not too technical.)

Giere, R. and Westfall, R.S. (eds) (1973) *Foundations of Scientific Method: The Nineteenth Century*, Bloomington, IN: Indiana University Press.(A conference volume, including papers on Whewell, Peirce, Maxwell, Darwin and Bernard. Nontechnical, good bibliographies.)

Helmholtz, H.L.F. von (1921) *Schriften zur Erkenntnistheorie*, P. Hertz and M. Schlick (eds), Berlin: Springer; trans. M.F. Lowe, *Epistemological Writings*, Boston Studies in the Philosophy of Science, vol. 37, Dordrecht: Reidel, 1977.(Major works of Helmholtz on epistemology.)

Herschel, Sir J.F.W. (1830) *A Preliminary Discourse on the Study of Natural Philosophy*, London: Longman; repr. *The Sources of Science*, no. 17, New York and London: Johnson Reprint Corp., 1966.(Introductory work on philosophy of science.)

Herschel, Sir J.F.W. (1857a) 'Whewell on the Inductive Sciences', *Essays from the Edinburgh and Quarterly Reviews*, London: Longman, 142-256.(Critical review of Whewell's *Philosophy of the Inductive Sciences*.)

Herschel, Sir J.F.W. (1857b) ‘Quetelet on Probabilities’, *Essays from the Edinburgh and Quarterly Reviews*, London: Longman, 365-465.(Excellent statement of Herschel’s pragmatic acceptance of natural laws.)

Jevons, W.S. (1874) *The Principles of Science: A Treatise on Logic and Scientific Method*, London: Macmillan; repr. New York: Dover, 1958.(Difficult sections, but a good introduction to his thought.)

Mach, E. (1883) *Die Mechanik in ihrer Entwicklung. Historisch-kritisch dargestellt*, Leipzig; trans. *The Science of Mechanics. A Critical and Historical Exposition of its Principles*, Chicago, IL, and London, 1893.(Difficult, presupposes knowledge of nineteenth-century physics.)

Mach, E. (1886) *Die Analyse der Empfindungen und das Verhältnis des Physischen zum Psychischen*, Jena; trans. *The Analysis of Sensations and the Relation of the Physical to the Psychical*, Chicago, IL, 1897.(Technical and involved argument, but essential for understanding Mach.)

Mach, E. (1905) *Erkenntnis und Irrtum*, Leipzig: J.A. Barth; trans. T.J. McCormack and P. Foulkes, *Knowledge and Error*, Dordrecht: Reidel, 1976.(Straightforward and easy introduction to Mach’s philosophy of science.)

Mandelbaum, M. (1971) *History, Man, and Reason. A Study in Nineteenth-century Thought*, Baltimore, MD, and London: Johns Hopkins University Press.(An indispensable guide to the history of nineteenth-century ideas. Presupposes some knowledge of the history of philosophy.)

Mill, J.S. (1843) *A System of Logic: Ratiocinative and Inductive; Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, London, Longman; 8th edn, 1872; repr. in J.M. Robson (ed.) *Collected Works of John Stuart Mill*, London: Routledge and Toronto, Ont.: University of Toronto Press, vols 7 and 8, 1991.(Mill’s major work on scientific method, including his criticisms of Whewell. The argument is often involved and difficult. The pagination of volumes 7 and 8 is consecutive; volume 8 begins at page 639.)

Mill, J.S. (1865) *Auguste Comte and Positivism*, London; repr. in J.M. Robson (ed.) *Collected Works of John Stuart Mill*, London: Routledge and Toronto, Ont.: University of Toronto Press, vol. 10, 261-368, 1991. (Invaluable source of Mill’s thoughts about positivism. Also contains his critique of Herbert Spencer. Nontechnical.)

Pearson, K. (1911) *The Grammar of Science*, London; repr. New York: Meridian Books, 3rd edn, 1957. (Presupposes some knowledge of nineteenth-century science, but not a difficult book.)

Santillana, G. de (1950) ‘Aspects of Scientific Rationalism in the Nineteenth Century’, *International Encyclopedia of Unified Science*, vol. II, no. 8, Chicago, IL: University of Chicago Press.(Valuable nontechnical account of the scientific background.)

Spencer, H. (1855) *The Principles of Psychology*, London: Williams & Norgate; 2nd edn, New York: Appleton, 1871-3, 2 vols.(The two editions comprise Spencer’s basic ideas in psychology. Few technicalities.)

Spencer, H. (1864, 1867) *Principles of Biology*, London: Williams & Norgate, 2 vols; 2nd edn, 1898, 1899. (Spencer’s basic ideas on evolution of species predated those of Darwin.)

Whewell, W. (1830) ‘Induction III (I) genl. views/geology/etc’, unpublished notebook in black metal chest marked INDUCTION, Wren Library, Trinity College, Cambridge.(A historical treasure with views predating published ones.)

Whewell, W. (1840) *Philosophy of the Inductive Sciences*, London, 2 vols. (Whewell’s *magnum opus* in philosophy of science; relatively nontechnical.)

Whewell, W. (1858a) *Novum organon renovatum*, London: J.W. Parker & Sons.

Whewell, W. (1858b) *History of Scientific Ideas*, London: J.W. Parker & Sons.

Whewell, W. (1860) *On the Philosophy of Discovery*, London: J.W. Parker & Sons.(1858a and b and 1860 comprise the 3rd edition of *The Philosophy of the Inductive Sciences*. New material is added. 1860 contains Whewell’s ‘solution’ to the ultimate problem of philosophy.)

Wolters, G. (1987) *Mach I, Mach II, Einstein und die Relativitätstheorie. Eine Fälschung und ihre Folgen (Mach I, Mach II. Einstein and the Theory of Relativity. A Forgery and its Consequences)*, Berlin and New York: de Gruyter.(Unfortunately not translated into English. Enlightened understanding of Mach’s philosophy of science. Difficult sections.)

Wolters, G. (1989) ‘Phenomenalism, Relativity and Atoms: Rehabilitating Ernst Mach’s Philosophy of Science’, in J.E. Fenstad, I.T. Frolov and R. Hilpinen (eds) *Logic, Methodology and Philosophy of Science VIII*, Proceedings of the Eighth International Congress of Logic, Methodology and Philosophy of Science, Moscow, 1987, Amsterdam: North Holland, 641-60.(A clear, concise statement of Wolter’s reading of Mach’s philosophy of science.)

Science, philosophy of

1 Historical background and introduction

Science grew out of philosophy; and, even after recognizable, if flexible, interdisciplinary boundaries developed, the most fruitful philosophical investigations have often been made in close connection with science and scientific advance. The major modern innovators - Bacon, Descartes, Leibniz and Locke among them - were all centrally influenced by, and in some cases significantly contributed to, the science of their day. Kant's fundamental epistemological problem was generated by the success of science: we have obtained certain knowledge, both in mathematics and - principally due to Newton - in science, how was this possible? Unsurprisingly, many thinkers who are principally regarded as great scientists, had exciting and insightful views on the aims of science and the methods of obtaining scientific knowledge. One can only wonder why the epistemological views of Galileo and of Newton, for example, are not taught along with those of Bacon and Locke, say, in courses on the history of modern philosophy. Certainly it can be argued very convincingly that the former two had at least as much insight into the aims and methods of science, and into how scientific knowledge is gained and accredited as the latter two (see Galilei, G. §3; Newton, I. §§2-3; also see Boyle, R.; Copernicus, N.; Kepler, J.).

In the nineteenth century, Maxwell, Hertz and Helmholz all had interesting views about explanation and the foundations of science, while Poincaré who was undoubtedly one of the greatest mathematicians and mathematical physicists, was arguably also one of the greatest philosophers of science - developing important and influential views about, amongst other things, the nature of theories and hypotheses, explanation, and the role of probability theory both within science and as an account of scientific reasoning (also see Duhem, P.M.M.; French philosophy of science; Le Roy, É.; Meyerson, É.; Science, 19th century philosophy of).

The period from the 1920s to 1950s is sometimes seen as involving a movement towards more formal issues to the exclusion of detailed concern with the scientific process itself (see Logical positivism). While this has been over-exaggerated - Carnap, Hempel, Popper and especially Reichenbach for example all show sophisticated awareness of a range of issues from contemporary science (also see Bridgman, P.W.; Operationalism) - there is no doubt that general attention in philosophy of science has been redirected back to the details of science, and in particular of its *historical* development, by 'post-positivist' philosophers such as Hanson, Feyerabend, Kuhn, Lakatos and others.

Current philosophy of science has developed this great tradition, addressing many of the now standard philosophical issues - about knowledge, the nature of reality, determinism and indeterminism and so on - but by paying very close attention to science both as an exemplar of knowledge and as a source of (likely) information about the world. This means that there is inevitably much overlap with other areas of philosophy - notably epistemology (the theory of scientific knowledge is of course a central concern of philosophy of science) and metaphysics (which philosophers of science often shun as an attempted a priori discipline but welcome when it is approached as an investigation of what current scientific theories and practices seem to be telling us about the likely structure of the universe). Indeed one way of usefully dividing up the subject would see scientific epistemology and what might be called scientific metaphysics as two of the main branches of the subject (these two together in turn forming what might be called general philosophy of science), with the third branch consisting of more detailed, specific investigations into foundational issues concerned with particular scientific fields or particular scientific theories (especial, though by no means exclusive, attention having been paid of late to foundational and interpretative issues in quantum theory and the Darwinian theory of evolution). Again not surprisingly, important contributions have been made in this third sub-field by scientists themselves who have reflected carefully and challengingly on their own work and its foundations (see Bohr, N.; Darwin, C.R.; Einstein, A.; Heisenberg, W.; Planck, M.), as well as by those who are more usually considered philosophers.

2 Contemporary philosophy of science: the theory of scientific knowledge

Scientists propose theories and assess those theories in the light of observational and experimental evidence; what distinguishes science is the careful and systematic way in which its claims are based on evidence (see Scientific method). These simple claims, which I suppose would win fairly universal agreement, hide any number of

complex issues.

First, concerning theories: how exactly are these best represented? Is Newton's theory of gravitation, or the neo-Darwinian theory of evolution, or the general theory of relativity, best represented - as logical empiricists such as Carnap supposed - as sets of (at least potentially) formally axiomatized sentences, linked to their observational bases by some sort of correspondence rules? Or are they best represented, as various recent 'semantic theorists' have argued, as sets of models (see Models; Theories, scientific)? Is this simply a representational matter or does the difference between the two sorts of approach matter scientifically and philosophically? This issue ties in with the increasingly recognized role of idealizations in science and of the role of models as intermediates between fundamental theory and empirical laws (see Campbell, N.R.; Idealizations). It also relates to an important issue about how best to think of the state of a scientific field at a given time: is a scientist best thought of as accepting (in some sense or other) a single theory or set of such theories or rather as accepting some sort of more general and hierarchically-organized set of assumptions and techniques in the manner of Kuhnian paradigms or Lakatosian research programmes? It seems likely that arriving at the correct account of scientific development and in particular of theory-change in science will depend on identifying the 'right' account of theories.

Next concerning the *evidence*: it has long been recognized that many of the statements that scientists are happy to regard as 'observation sentences' in fact presuppose a certain amount of theory, and that *all* observation sentences, short perhaps of purely subjective reports of current introspection, depend on some sort of minimal theory (even 'the needle points to around 5 on the scale' presupposes that the needle and the scale exist independently of the observer and that the observer's perception of them is not systematically deluded by a Cartesian demon). Does this mean that there is no real epistemic distinction between observational and theoretical claims? Does it mean that there is no secure basis or foundation for science in the form of observational and experimental results (see Observation)? If so, what becomes of the whole empiricist idea of basing scientific theories on the evidence? It can be argued that those who have drawn dire consequences from these considerations have confused fallibility with (serious) corrigibility: that there are observation statements, such as reports of meter readings and the like, of a sufficiently low level as to be, once independently and intersubjectively verified, not seriously corrigible despite being trivially strictly fallible (see Measurement, theory of). Aside from this issue, experiment was for a long time regarded as raising barely any independent, philosophical or methodological concern - experiments being thought of as very largely simply means for testing theories (see Experiment). More recently, there has been better appreciation of the extent to which experimental science has a life of its own, independent of fundamental theory, and of the extent to which philosophical issues concerning testing, realism, underdetermination and so on can be illuminated by studying experiments.

Suppose that we have characterized scientific theories and drawn a line between theoretical and observational statements, what exactly is involved in 'basing' theoretical claims 'systematically and carefully' on the evidence? This question has of course been perhaps *the* central question of general philosophy of science in this century. We have known at least since David Hume that the answer cannot be that the correct theories are *deducible* from observation results. Indeed not only do our theories universally generalize the (inevitably finite) data as Hume pointed out, they also generally 'transcend' the data by explaining that data in terms of underlying, but non-observable, theoretical entities. This means that there must always in principle be (indefinitely) many theories that clash with one another at the theoretical level but yet entail all the same observational results (see Underdetermination). What extra factors then are involved over and beyond simply having the right observational consequences? What roles do such factors as simplicity (see Simplicity (in scientific theories)), and explanatory power (see Explanation), play in accrediting theories on the basis of evidence? Moreover what status do these factors have - are they purely pragmatic (the sorts of features *we* like theories to have) or are they truth-indicating, and if so why? Some have argued that the whole process can be codified in probabilistic terms - the theories that we see as accredited by the evidence being the ones that are at any rate *more probable* in the light of that evidence than any of their rivals (see Confirmation theory; Inductive inference; Probability theory and epistemology).

Finally, suppose we have characterized the correct scientific way of reasoning to theories from evidence, what exactly does this tell us about the theories that have been thus 'accredited' by the evidence? And what does it tell us about the entities - such as electrons, quarks, and the rest - apparently postulated by such theories? Is it reasonable to believe that these accredited theories are *true* descriptions of an underlying reality, that their theoretical terms refer to real, though unobservable entities? (Or at least to believe that they are *probably* true? or

approximately true? or perhaps probably approximately true?) More strongly still, is any one of these beliefs the *uniquely* rational one? Or is it instead more, or at least equally, reasonable - at least equally explanatory of the way that science operates - to hold that these ‘accredited’ theories are no more than empirically adequate, even that they are simply instruments for prediction, the theoretical ‘entities’ they involve being no more than convenient fictions (see Conventionalism; Fictionalism; Incommensurability; Putnam, H.; Scientific realism and antirealism)? One major problem faced by realists is to develop a plausible response to once accepted theories that are now rejected either by arguing that they were in some sense immature - not ‘fully scientific’ - or that, despite having been rejected, they nonetheless somehow live on as ‘limiting cases’ of current theories (see Alchemy; Chemistry, philosophical aspects of §2; Field theory, classical; Mechanics, Aristotelian; Mechanics, classical; Optics; Vitalism).

Clearly an antirealist view of theories would be indicated *if* it could convincingly be argued that the accreditation of theories in science is not simply a function of evidential and other truth-related factors or even of epistemic pragmatic factors, but *also* of broader cultural and social matters. Although such arguments are heard increasingly often, many remain unconvinced - seeing those arguments as based *either* on confusion of discovery with validational issues *or* on fairly naïve views of evidential support (see Constructivism; Discovery, logic of; Gender and science; Marxist philosophy of science).

3 Contemporary philosophy of science: ‘scientific metaphysics’

Suppose that we take a vaguely realist view of current science, what does it tell us about the general structure of reality? Does a sensible interpretation of science require the postulation, for example, of natural kinds (see Natural kinds) or universals? Does it require the postulation of a notion of physical necessity to distinguish natural laws from ‘mere’ regularities (see Laws, natural)? What is the nature of probability (see Probability, interpretations of) - is a probabilistic claim invariably an expression of (partial) ignorance or are there real, irreducible ‘objective chances’ in the world? What exactly is involved in the claim that a particular theory (or a particular system described by such a theory) is deterministic (see Determinism and indeterminism), and what would it mean for the world as a whole to be deterministic? Does even ‘deterministic’ science eschew the notion of *cause* (as Russell argued)? Does this notion come into its own in more ‘mundane’ contexts, involving what might be called ‘causal factors’ and probabilistic causation? What exactly is the relationship between causal claims - such as ‘smoking causes heart disease’ - and statistical data (see Causation)? How should spacetime be interpreted (see Spacetime): as substantive or as ‘merely’ relational? Does current science plus whatever ideas of causality are associated with it unambiguously rule out the possibility of time travel (see Time travel), or does this remain at least logically possible given current science?

Finally, and most generally, what is science (or, perhaps more significantly, the *direction* of scientific development) telling us about the overall structure of the universe - that it is one simple system governed at the fundamental level by one unified set of general laws, or rather that it is a ‘patchwork’ of interconnected but separate, mutually irreducible principles (see Unity of science; Reduction, problems of)? Although it is of course true - despite some exaggerated claims on behalf of ‘theories of everything’ - that science is very far from reducing everything to a common fundamental basis, and although it is of course true that, even in cases where reduction is generally agreed to have been achieved, such as that of chemistry to physics, the reduction is *ontological* (that is, chemistry has been shown to need no essential, non-physical primitive notions) rather than *epistemological* (no one would dream of trying actually to *derive* a full description of any chemical reaction from the principles of quantum mechanics), some would nonetheless still argue that the overall *tendency* of science is in the reductionist direction (see Chemistry, philosophical aspects of §5).

These are examples of the more or less general, and impressively varied, ‘metaphysical’ issues informed by science that have attracted recent philosophical attention.

4 Contemporary philosophy of science: foundational issues from current science

Many of the most interesting issues in current philosophy of science are closely tied to foundational or methodological concerns about current scientific theory. One fertile source of such concerns is quantum theory. How much of a revolutionary change in our general metaphysical view of the world does it require? Is the theory irreducibly indeterministic or do ‘hidden variable’ interpretations of some sort remain possible despite the negative

results? What does quantum mechanics tell us about the notion of cause? Does quantum mechanics imply a drastic breakdown of ‘locality’, telling us that the properties of even vastly spatially separated systems are fundamentally interconnected - so that we can no longer think of, for example ‘two’ spatially separated electrons as separate, independent ‘particles’? More directly, is there, in view of the ‘measurement problem’ a coherent interpretation of quantum mechanics at all? (It has been argued that when the theory is interpreted universally so that *all* systems, including ‘macroscopic’ ones, such as measuring apparatuses, are assigned a quantum state then the two fundamental principles of quantum theory - the Schrödinger equation and the projection postulate - come into direct contradiction (see Bell’s theorem; Field theory, quantum; Quantum measurement problem; Quantum mechanics, interpretation of; also see Randomness; Statistics).)

Although perhaps attracting relatively less attention than quantum theory, the other two great theories that form the triumvirate at the heart of contemporary physics - relativity (both special and general) and thermodynamics - pose similarly fascinating problems. In the case of relativity theory, philosophers have raised both ontological issues (for example, concerning the nature of spacetime) and epistemological issues (concerning for example the real role played in Einstein’s development of the theory by Machian empiricism, the role of allegedly crucial experiments such as that of Michelson and Morley (see Crucial experiments), and the evidential impact on the general theory of the Eddington star-shift experiment). There are also important issues about the consistency of relativity and quantum theory - issues that in turn feed into the more general questions concerning the unity of science and realism (see General relativity, philosophical responses to; Relativity theory, philosophical significance of).

Thermodynamics raises issues about, amongst other things, probability and the testing of probabilistic theories, about determinism and indeterminism, and about the direction of time (see Thermodynamics; Determinism and indeterminism; Duhem, P.M.M. §2; Time). Other current areas of physics, too, raise significant foundational issues (see Chaos theory, Cosmology).

For a long time, philosophy of science meant in effect philosophy of *physics*. A welcome broadening-out has occurred recently - especially in the direction of philosophy of biology. The central concern here has been with foundational issues in the Darwinian theory of evolution (or more accurately the neo-Darwinian synthesis of natural selection and genetics). Questions have been raised about the testability and, more generally, the empirical credentials of that theory, about the scope of the theory (in particular what it can tell us about humans and human societies), about the appropriate ‘unit of selection’ (individual, gene, group), about what exactly are genes and what exactly are species, and about whether evolutionary biology involves distinctive - perhaps even in *some* sense ‘teleological’ - modes of explanation (see Darwin, C.R.; Ecology; Evolution, theory of; Functional explanation; Genetics; Huxley, T.H.; Life, origin of; Linnaeus, C. von; Sociobiology; Species; Taxonomy; Wallace, A.R.). More recently philosophy of biology has started to widen its own scope by considering issues outside of evolutionary theory (see Molecular biology; Medicine, philosophy of), where, however, issues of reductionism and of the possibility of distinctive modes of explanation still loom large.

See also: Archaeology, philosophy of; Atomism, ancient; Colour, theories of; Computer science; Conservation principles; Decision and game theory; Demarcation problem; Democritus; Electrodynamics; Facts; Feminism and social science; Geology, philosophy of; Heideggerian philosophy of science; Information theory; Jung, C.G.; Keynes, J.M.; Koyré, A.; Matter; Mead, G.H.; Naturalized philosophy of science; Paranormal phenomena; Postcolonial philosophy of science; Proclus §10; Ptolemy; Religion and science; Responsibilities of scientists and intellectuals; Risk assessment; Science in Islamic philosophy; Scientific realism and social science; Space; Technology, philosophy of; Thought experiments

JOHN WORRALL

Further reading

Kitcher, P. (1993) *The Advancement of Science: Science without Legend, Objectivity without Illusions*, New York, and Oxford: Oxford University Press.(Thorough and illuminating account of the general issues surrounding theory-change in science; also useful as an introduction to the methodological issues raised by Darwinian theory.)

Maudlin, T. (1994) *Quantum Non-Locality and Relativity*, Oxford, and Cambridge, MA: Blackwell.(Given its subject matter, an exceptionally clear, accessible account of some of the foundational issues in quantum theory, especially concerning its reconcilability with relativity theory.)

Papineau, D. (ed.) (1996) *The Philosophy of Science*, Oxford Readings in Philosophy, Oxford: Oxford University Press.(Recent collection of articles, especially on the realism/antirealism issue, but also on issues of empirical support.)

Salmon, M.H. *et al.* (1992) *Introduction to the Philosophy of Science*, Englewood Cliffs, NJ: Prentice Hall.(A text written by members of the internationally celebrated History and Philosophy of Science Department at the University of Pittsburgh and covering general philosophy of science, as well as philosophy of physics, of biology, and of the behavioural and social sciences.)

Scientific method

Procedures for attaining scientific knowledge are known as scientific methods. These methods include formulating theories and testing them against observation or experiment. Ancient and medieval thinkers called any systematic body of knowledge a 'science', and their methods were aimed at knowledge in general. According to the most common model for scientific knowledge, formulated by Aristotle, induction yields universal propositions from which all knowledge in a field can be deduced. This model was refined by medieval and early modern thinkers, and further developed in the nineteenth century by Whewell and Mill.

As Kuhn observed, idealized accounts of scientific method must be distinguished from descriptions of what scientists actually do. The methods of careful observation and experiment have been in use from antiquity, but became more widespread after the seventeenth century. Developments in instrument making, in mathematics and statistics, in terminology, and in communication technology have altered the methods and the results of science.

1 'Method' and 'science'

'Method' comes from the Greek *meta* (after) plus *hodos* (path or way). A method is a way to achieve an end; a scientific method is a way to achieve the ends of science. What those ends are depends on what science is or is taken to be. The word 'science' now means primarily natural science, examples of which are physics, astronomy, biology, chemistry, geology and psychology, and it applies secondarily to social sciences such as economics and sociology. Discussions of method focus on the cognitive aims of science, which may include knowledge, understanding, explanation, or predictive success, with respect to all or part of nature or to some domain of natural or social phenomena. Abstractly described, scientific method is the means for attaining these aims, especially by forming models, theories, or other cognitive structures and testing them through observation and experiment (see Experiment; Models; Observation; Theories, scientific). Investigations of scientific method may describe the methods actually employed by scientists, or they may formulate proposals about the procedures that should be followed to achieve scientific knowledge (see Naturalized philosophy of science).

The main features traditionally ascribed to 'the scientific method' - including clear statement of a problem, careful confrontation of theory with fact, open-mindedness, and (potential) public availability or replicability of evidence - are common to many cognitive endeavours, including much work in the humanities. Although there is no single method that distinguishes science from other intellectual practices, the following features are characteristic of the natural and social sciences: the use of quantitative data and of theories formulated mathematically; the use of artificially created experimental situations; and an interest in universal generalizations or laws (see Demarcation problem; Laws, natural §1; Unity of science). (But note that biological taxonomies are not intrinsically quantitative, and that neither astronomy nor economics is based primarily on experiment.) In both the natural and social sciences, the formation of models or theories may involve skills in mathematical computation and derivation, in evaluating consistency, in imagining new theoretical possibilities, in assessing the structure of a taxonomy, or in relating one area of investigation to other areas. The means for testing theories or generating new empirical knowledge vary widely, and include systematic observation, unsystematic observation, checks against background theories or knowledge, and various experimental procedures, including sophisticated statistical techniques, the construction of special instruments or apparatus and the use of specially bred laboratory animals.

2 Ideas of method from the Greeks to Thomas Kuhn

Because 'science' originally meant any systematic body of knowledge, ranging from mathematics to theology, the method of science was the method for obtaining, or perhaps merely for presenting and teaching, knowledge in general. Methods varied in relation to beliefs about what there is to be known.

The writings of Plato and Aristotle embody contrasting conceptions of both the objects of knowledge and the method for knowing them. In *Republic* VI, Plato divided the objects of knowledge into two realms: the visible and the intelligible (see Plato §14). He considered the former to include the objects of the senses, about which only opinion but not genuine knowledge is possible, and the latter to include geometry and astronomy, in which investigators assume the existence of their objects (such as geometrical objects) and reason from them as from hypotheses. In the highest reaches of the intelligible realm, reason attempts to reach 'the first principle of all that

exists', from which it then 'comes down to a conclusion…proceeding by means of Forms and through Forms to its conclusions which are Forms', without any reference to the visible world. Plato conceived the sensible world as a dim reflection of the intelligible Forms, and he held that the Forms themselves are best known through direct intellectual contemplation, independent of sensory experience. The notion of an intelligible world behind the sensible world, and especially of a world described by mathematics, has played an important role in physical science since Plato's time.

Aristotle rejected Plato's intelligible realm because it removed the objects of mathematical sciences such as astronomy from the sensible world, where he believed the forms of things are to be found. He carried out extensive observations (including dissections) in biology and developed a preliminary taxonomic scheme. Aristotle's principal discussion of method is the *Posterior Analytics*, a founding work in the philosophy of science. He accepted the Platonic distinction between a direction of cognition that is going 'to the forms' and the direction of cognition (as in syllogistic demonstration) that proceeds 'from the forms', but he conceived these processes as starting from sensible objects and arriving at knowledge of the common natures or essences of things as existing in those objects. Such knowledge (for instance, of the essence of a specific kind of mineral, or kind of living thing) yields a set of core propositions in each science, from which other knowledge is to be deductively derived.

Medieval philosophers, including Roger Bacon, John Duns Scotus and William of Ockham, commented extensively on Aristotle's methodological writings, which were later discussed at the University of Padua together with those of Galen. A central topic was the distinction between analysis and synthesis, or, as it was sometimes called, between resolution and composition. In the analytic phase of inquiry, one resolves the object of investigation into its basic constituents or least elements so as to determine its first principles. In the synthetic phase, one explains a subject matter from its first principles, or presents a body of knowledge by deriving it from such principles. In a common example, the analytic phase would include the search for the axioms, postulates and definitions of geometry; the synthetic phase the demonstration of theorems from those axioms and postulates. Bacon suggested that the first principles achieved in the analytic or inductive phase can be tested by deducing and checking new consequences (a feature of 'hypothetico-deductive' tests of theories). Scotus outlined a method of agreement, in which a possible cause for an effect is found by listing the circumstances that co-occur with the effect and looking for one that is present every time. Ockham suggested a method of difference, in which a circumstance that is present when the effect is present and absent when it is absent is considered as a possible cause for the effect.

The seventeenth century, a time of fundamental change in physics and astronomy, saw continuing attention to method within the inductive-deductive framework established by Aristotle. Francis Bacon outlined inductive procedures in detail, calling for extensive collections of data (named 'histories') which are to be culled systematically for general principles or laws. Galileo urged that mathematical descriptions be fitted to natural phenomena through observation and experiment (see Galilei, G.). Descartes wrote in *Discourse on the Method* (1637) that the derivation of an effect from a cause may serve as an explanation of the effect, and also as an empirical 'proof' of the posited cause (see Descartes, R. §6). Newton, in *Mathematical Principles of Natural Philosophy* (1729), laid down several 'hypotheses' or 'rules' for reasoning in natural philosophy. He advised investigators to avoid multiplying causes in relation to effects, to generalize from properties found in bodies that have been observed to all bodies in the universe, and to accept inductively supported propositions as 'accurately or very nearly true' until new observations improve upon their accuracy or limit their scope (see Newton, I. §4).

During the nineteenth century, the 'philosophy of science' or the 'logic of science' became, in the writings of William Whewell, John Stuart Mill and others, a main staple of philosophy (see Philosophy of science in the nineteenth century). Whewell's *Philosophy of the Inductive Sciences* (1840) analyses scientific knowledge of 'external' nature (excluding the mind itself). Whewell held that scientific knowledge is based upon sensations and ideas, the former being the 'objective' element (caused by objects), the latter a 'subjective' element (provided by the knowing subject). Consciously entertained facts and theories correspond to sensations and ideas, but not completely, because all facts implicitly include ideas (and so, possibly, theory). Whewell divided the methods of science into methods of observation, of obtaining clear ideas, and of induction. The methods of observation include quantitative observation (as in chemistry or astronomy) and the perception of similarities (as in natural history); observation includes both the collection and classification of facts. Clear ideas result from intellectual education (including both the mathematical sciences and natural history), and from discussion, including

(sometimes metaphysical) discussions of definitions, such as whether uniform force acting in free fall should be defined relative to space or to time. Science proceeds by ‘induction’, including the use of quantitative techniques to smooth out the irregularities of observation (that is, the ‘method of curves’ by which a curve is fitted to data points, the ‘method of means’ and the ‘method of least squares’) and the formation and empirical testing of tentative hypotheses. Laws of phenomena are usually formed first, but theories of true causes are desired, such as (Whewell explained) have been found in physical astronomy, physical optics, and geology, and might be found with respect to heat, magnetism, electricity, chemical compounds and living organisms.

In *A System of Logic* (1843), Mill analysed the methods of science even more fully than had Whewell, now including psychology and the social sciences (also known as the ‘mental’ and ‘moral’ sciences). In his analysis of experimental method, Mill included the methods of agreement and difference already mentioned, and added the method of residues, which directs the investigator to look for the (as yet unknown) causes of those effects that remain after all other effects have been assigned to known causes, and the method of concomitant variations, according to which those phenomena that vary regularly in quantitative degree with one another are assumed to be causally related. Like Whewell, Mill emphasized the role of new or pre-existing concepts and names in scientific observation, and the role that classification plays in induction. He proposed that psychology and the social sciences should adopt the explanatory structures of the natural sciences, a proposal frequently criticized since his time, notably by the Neo-Kantians Wilhelm Dilthey, Heinrich Rickert and Ernst Cassirer, who discussed methods pertaining to the ‘social’ or ‘human’ sciences, and who included the humanities as a form of ‘science’ (in German, *Wissenschaft*) (see Neo-Kantianism; Positivism in the social sciences).

Philosophers in the first half of the twentieth century, especially the Vienna Circle and Karl Popper, sought to analyse science and to reconstruct scientific reasoning using the new symbolic logic or the new theory of probability (see Vienna Circle). They continued the investigation of theory confirmation, focusing on recent theories in physics (see Confirmation theory). Rudolph Carnap attempted unsuccessfully to develop a quantitative theory of inductive support. Carl Hempel, who favoured a hypothetico-deductive account of scientific confirmation (involving the testing of theories by their deductive consequences), revealed certain paradoxes that result when the relation between scientific generalizations and confirming instances are expressed in predicate logic and certain (plausible) assumptions are made. Popper concluded that the defining feature of the empirical methods of science is that statements are always subjected to falsification by new data.

In the second half of the twentieth century philosophical analyses of scientific method broadened to again include sciences such as biology and geology, and paid greater attention to the history of science. N.R. Hanson recalled the often implicit role of theory in observation, questioning the notion of a theory-neutral observation language (see Observation §4). Thomas S. Kuhn emphasized the social nature of scientific communities and the common training that produces a shared vocabulary and set of experimental procedures. Kuhn and Paul Feyerabend stressed the need to distinguish the idealized accounts of scientific method given by some scientists and philosophers from the actual methodological practices of scientists. In studying the latter, historical and sociological investigation supplements the participatory acquaintance with scientific research possessed by scientists themselves and by some philosophers.

3 Scientific method in scientific practice

The methods of careful observation and description (including quantitative description) and of controlled experiment arose in antiquity and were practised by Greek, Hellenistic and Islamic investigators (see Philosophy of science in Islam §2). Examples include Aristotle’s biological observations, the many Greek and Arabic (and earlier Babylonian) tables of astronomical data, Ptolemy and Alhazen’s careful studies of binocular vision and distance perception, and Galen’s use of the ligature and other experimental techniques in physiology.

From antiquity, instruments aided the precision of observation in astronomy and optics. In the early seventeenth century Johannes Kepler used Tycho Brahe’s precise astronomical data (obtained with improved instruments) to establish the elliptical orbits of the planets and to determine the relations among the sizes and periods of those orbits. In 1609 Galileo used the newly invented telescope to observe previously unseen heavenly bodies (including the moons of Jupiter). Later in the century the microscope opened new fields of observation. The nineteenth and twentieth centuries have seen the development of refined and often complex instrumentation in all branches of natural science, including biology, chemistry and psychology, ranging from improved balances to the electron

microsope and the space telescope. Photography has been used to record data in nearly every field. The computer permits collection and manipulation of larger bodies of data than was previously practical. In psychology, the computer allows generation of precisely controlled stimuli and the recording of data with highly accurate temporal measurement.

The development of mathematics, including probability and statistics, has yielded new forms of theory statement and new descriptions of observational or experimental data. Mathematical sciences from antiquity to the seventeenth century used geometry almost exclusively. The development of algebraic geometry and of the calculus permitted new statements of Newtonian mechanics in the eighteenth century, and opened up new possibilities (theoretical as well as experimental) for describing and investigating functional relations among quantities. In the nineteenth and twentieth centuries, new mathematics has been demanded by or has facilitated mathematical physics. Thus, the discovery of non-Euclidean geometries opened up hitherto unforeseen theoretical possibilities in physical cosmology. The development of probability and statistics permitted the formulation of statistical laws, as in mathematical genetics, quantum physics and sociology (see Probability, interpretations of; Statistics; Statistics and social science). Inferential statistics is widely used in the analysis of quantitative data in psychology and other sciences.

Clear and precise terminology is an important feature of scientific methodology. Astronomy, optics (as the science of vision), natural history and medicine developed technical vocabularies in antiquity. Newton profoundly altered the terminology of physics, which continues to change as theories change. Carl von Linnaeus invented important taxonomies in botany and zoology; after Darwin’s theory of natural selection gained acceptance, evolutionary history influenced biological taxonomy (see Evolution, theory of; Taxonomy). Molecular biology has produced another new terminology (see Molecular biology). In psychology, long-standing mentalistic terminology was purged by twentieth-century behaviourists (see Behaviourism, methodological and scientific), and has since been reintroduced, partly under the influence of the computer metaphor for mental processes (see Mind, computational theories of). Likewise, economics, anthropology and sociology use refined technical vocabularies.

A sense of the various instruments and techniques of data collection and analysis now used can be gleaned from the materials, methods and results sections of scientific journals. Journals and other means of communication are themselves of methodological significance. The available methods for presenting observational data were radically altered by the development of printing (for both text and images), and again through computer-generated images and electronic communication. The mass production of standardized illustrations and printed data permits worldwide dissemination, utilization and hence testing of scientific findings.

The structure of scientific research groups and their interaction with scientific institutions, including the processes for deciding whether to fund research or to publish results, are also part of the method of science (broadly conceived). The methodological effectiveness of science can be evaluated at various scales, including the individual experiment, the individual investigator, the laboratory group, or the institutional structures by which collective instruments such as particle accelerators are administered. One might further examine the normative consequences of having relative homogeneity of methodological and theoretical belief across an active science, as opposed to the hedged bet of methodological and theoretical diversity. The student of scientific methods may investigate any aspect of the linguistic, conceptual, psychological, instrumental, social and institutional features of the sciences that affects their cognitive products.

See also: Crucial experiments; Discovery, logic of; Explanation; Inductive inference; Feminism and social science; Gender and science; Heideggerian philosophy of science; Objectivity; Philosophy of science in Islam; Postcolonial philosophy of science

GARY HATFIELD

References and further reading

Aristotle (*c*.mid-4th century BC) *Posterior Analytics* (Analytica Posteriora), in J. Barnes (ed.) *Complete Works of Aristotle*, vol. 1, Princeton, NJ: Princeton University Press, 2 vols, 1984.(A founding work in philosophy of science.)

Bacon, F. (1620) *Novum organum*, London: Longmans; repr. in T. Fowler (ed.) *Bacon’s Novum organum*, Oxford: Clarendon Press, 1888; trans. P. Urbach and J. Gibson, Chicago, IL: Open Court, 1994.(Seminal statement of an empirical methodology for investigating nature.)

Descartes, R. (1637) *Discours de la méthode pour bien conduir sa raison et chercher la vérité dans les sciences plus la dioptrique, les meteores, et la geometrie, qui sont des essais de cete methode (Discourse on the Method for Properly Conducting Reason and Searching for Truth in the Sciences, as well as the Dioptrics, the Meteors, and the Geometry, which are essays in this method)*, in J. Cottingham, R. Stoothoff, D. Murdoch and A. Kenny (eds and trans.) *The Philosophical Writings of Descartes*, vol. 1, Cambridge: Cambridge University Press, 1984-91, 2 vols.(Seminal statement of a programme for investigating all of nature.)

Hanson, N.R. (1958) *Patterns of Discovery*, Cambridge: Cambridge University Press.(Important study of fundamental concepts in science, with reference to actual historical cases.)

Jardine, N. (1984) *Birth of History and Philosophy of Science: Kepler's 'Defence of Tycho against Ursus', with Essays on Its Provenance and Significance*, Cambridge: Cambridge University Press.(Detailed analysis of an early modern work in philosophy of science.)

Kitcher, P. (1993) *The Advancement of Science: Science without Legend, Objectivity without Illusions*, New York: Oxford University Press.(Advanced work in philosophy of science, relevant to method.)

Kuhn, T. (1962) *Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press; 2nd edn, 1970. (Widely read analysis of the historical development scientific knowledge.)

Losee, J. (1993) *Historical Introduction to the Philosophy of Science*, Oxford: Oxford University Press, 3rd edn. (An introduction, with references, to many topics in the philosophy of science.)

Mill, J.S. (1843) *A System of Logic*; repr. in J.M. Robson (ed.) *Collected Works of John Stuart Mill*, London: Routledge and Toronto, Ont.: University of Toronto Press, 1991.(Lengthy but readable and important nineteenth-century examination of the structure and empirical basis of scientific knowledge.)

Newton, I. (1687) *Philosophiae naturalis principia mathematica*, London: Joseph Streater (for the Royal Society); 2nd edn, Cambridge, 1713; 3rd edn, London: Guil. & Joh. Innys (for the Royal Society), 1726; repr. in A. Koyré and I.B. Cohen (eds) with assistance of A. Whitman *Isaac Newton's Philosophiae naturalis principia mathematica* (3rd edn 1726, with variant readings), Cambridge, MA: Harvard University Press, 1972, 2 vols. (Newton's monumental contribution, forever transforming science. Contains the 'Rules of Reasoning in Philosophy'.)

Plato (*c*.380-367 BC) *Republic*, trans. G.M. Grube, revised by C. Reeve, Indianapolis, IN: Hackett Publishing Company, 1992.(Contains a discussion of knowledge and methods of knowing.)

Whewell, W. (1840) *The Philosophy of the Inductive Sciences Founded upon their History*, London: J.W. Parker, 2 vols.(Difficult but seminal work in philosophy of science.)

Scientific realism and antirealism

Traditionally, scientific realism asserts that the objects of scientific knowledge exist independently of the minds or acts of scientists and that scientific theories are true of that objective (mind-independent) world. The reference to knowledge points to the dual character of scientific realism. On the one hand it is a metaphysical (specifically, an ontological) doctrine, claiming the independent existence of certain entities. On the other hand it is an epistemological doctrine asserting that we can know what individuals exist and that we can find out the truth of the theories or laws that govern them.

Opposed to scientific realism (hereafter just 'realism') are a variety of antirealisms, including phenomenalism and empiricism. Recently two others, instrumentalism and constructivism, have posed special challenges to realism. Instrumentalism regards the objects of knowledge pragmatically, as tools for various human purposes, and so takes reliability (or empirical adequacy) rather than truth as scientifically central. A version of this, fictionalism, contests the existence of many of the objects favoured by the realist and regards them as merely expedient means to useful ends. Constructivism maintains that scientific knowledge is socially constituted, that 'facts' are made by us. Thus it challenges the objectivity of knowledge, as the realist understands objectivity, and the independent existence that realism is after. Conventionalism, holding that the truths of science ultimately rest on man-made conventions, is allied to constructivism.

Realism and antirealism propose competing interpretations of science as a whole. They even differ over what requires explanation, with realism demanding that more be explained and antirealism less.

1 Arguing for realism

Late nineteenth- and early twentieth-century debates over the reality of molecules and atoms polarized the scientific community on the realism question. Antirealists like Mach, Duhem and Poincaré - representing (roughly) phenomenalist, instrumentalist and conventionalist positions - at first carried the day with a sceptical attitude towards the truth of scientific theories and the reality of the 'theoretical entities' employed by those theories (see Phenomenalism; Conventionalism). Led by the successes of statistical mechanics (see Thermodynamics) and relativity (see Relativity theory, philosophical significance of), however, Planck and Einstein helped turn the tide towards realism. That movement was checked by two developments. In physics the quantum theory of 1925-6 quickly ran into difficulties over the possibility of a realist interpretation (see Quantum mechanics, interpretation of) and the community settled on the instrumentalist programme promoted by Bohr and Heisenberg. This was a formative lesson for logical empiricism whose respect for developments in physics and whose positivistic orientation led it to brand the realism question as metaphysical, a pseudo-question (see Logical positivism, philosophy of). Thus for a while empiricist and instrumentalist trends in science and philosophy eclipsed scientific realism.

The situation changed again in the 1960s, by which time science and its technological applications had become a ubiquitous and dominant feature of Western culture. In this setting philosophers like Smart (1963) and Putnam (1975) proposed what came to be known as the 'miracles' argument for scientific realism (see Smart, J.J.C. §3; Putnam, H. §2). They argued that unless the theoretical entities employed by scientific theories actually existed and the theories themselves were at least approximately true of the world at large, the evident success of science (in terms of its applications and predictions) would surely be a miracle. It is easy to see, at least with hindsight, that the most one could conclude from scientific success, however impressive, is that science is on the right track. That could mean, as the argument concludes, on the track to truth or it could just mean on the track to empirical success, perhaps with deeply flawed representations of reality. The 'miracles' argument is inconclusive. Nevertheless, during the next two decades it was compelling for many philosophers. Indeed, during this period realism became so identified with science that questioning realism was quickly put down as anti-science.

Realist orthodoxy found support in Popper's attack on instrumentalism, which he criticized as unable to account for his own falsificationist methodology (Popper 1956) (see Popper, K.R. §2). Broadening this line, Boyd developed an explanationist version of the 'miracles' argument that focused on the methods of science and tried as well to give proper due to the human-centred (constructivist and conventionalist) aspects of science emphasized by Kuhn and Feyerabend. Boyd asks why methods crafted by us and reflecting our interests and limitations lead to

instrumentally successful science. Contrasting realism with empiricism and constructivism, he finds that realism offers the best (indeed, the only) explanation. That is because, he argues, if we begin with truths or near-truths the methods we have crafted for science produce even more of the same. Since it is only realism that demands the truth of our scientific theories, then realism wins as giving the best explanation for the instrumental success of science. Hence, like a scientific hypothesis, realism is most likely to be true and we should believe in it.

The explanationist argument is carefully framed so that we ask only about the instrumental success of science; that is, success at the observational level. To take science as successful (for example, truth-producing) at the theoretical level would beg the question against empiricism and instrumentalism. Once this is recognized, however, we can see a significant gap in the reasoning. The argument is driven by a picture of science as generating new truths from old truths, but the explanatory issue raised is only about truths at the level of observation, not about truths in general. Antirealists might well reject this as an illegitimate request for explanation. If they accept it, there is an obvious empiricist or instrumentalist response: namely, that our scientific methods are made by us to winnow out instrumentally reliable information. If we begin with fairly reliable statements, the methods we have crafted for science will produce even more. Thus the explanation for scientific success at the instrumental level need not involve the literal truth of our scientific principles or theories, just their instrumental reliability. This move nicely converts the argument for realism as the best explanation of scientific success into an argument for instrumentalism.

There is a second problem with the explanationist tactic, perhaps even more serious. The conclusion in support of realism depends on an inference to the best explanation (see Inference to the best explanation). That principle, to regard as true that which explains best, is a principle that antirealisms (especially instrumentalism and empiricism) deny. Van Fraassen, for example, regards being the best explanation as a virtue, but one separate from truth. (He reminds us that the best may well be the best of a bad lot.) Although not required, there could perhaps be an instrumentalist principle of inference to the best explanation. It would not infer to the truth of the explanation but to its instrumental reliability (or empirical adequacy) - precisely the strategy pursued above where we infer instrumentalism from the instrumental success of science. Thus the explanationist argument uses a specifically realist principle of inference to the best explanation and, in so doing, begs the question of truth versus reliability, one of the central questions at issue between realism and antirealism.

2 Piecemeal realisms

Inference to the best explanation promised the most cogent version of the ‘miracles’ argument. Its inadequacy hastened a retreat from realism’s original undertaking as a global interpretation of science. Retreat was fostered by two other antirealist developments. One was the pessimistic meta-induction to the instability of current science, a conclusion based on the repeated overthrow of scientific theories historically and the consequent dramatic alterations in ontology. The other was a sharpening of the underdetermination thesis associated with Poincaré and Duhem, suggesting that there may be empirically equivalent theories between which no evidence can decide (see Underdetermination). Both developments tended to undermine claims for the reality of the objects of scientific investigation and the truth of scientific theories.

Pursuing a salvage operation, several philosophers suggested that realism could confine itself to being a doctrine about the independent existence of theoretical entities (‘entity realism’) without commitment to the truth of the theories employing them. Hacking (1983) proposed an ‘experimental argument’ for this entity realism; roughly, that if you can deploy entities experimentally to discover new features of nature (for example, use an electron gun to learn about quarks), then the entities must be real whether or not the covering theories are true (see Experiment §3). Cartwright (1983) suggested that the strategy of inference to the best explanation be confined to an inference to the causes of phenomena, since causes are unquestionably real. To the antirealist, however, these related strategies seem far from compelling. For one thing, it is not clear that one can so neatly disengage theoretical entities from their covering theories. Moreover, in both cases, we can see that the basis on which one is asked to draw a realist conclusion need support no more than a conclusion about utility or reliability. In Hacking’s case one need conclude only that electrons are a useful theoretical construct (perhaps a useful fiction?) and in Cartwright’s that certain causal hypotheses are reliable in certain domains.

Faced with these difficulties realism has fragmented even further. Sometimes it takes an historicist turn, countering the pessimistic meta-induction by endorsing as real only those fruitful entities that survive scientific revolutions.

Sometimes realism becomes highly selective in other ways; for example, looking only at what seems essential in specific cases of explanatory or predictive success, or at entities that stand out as supported by only the very best scientific evidence. Although each of these principles locates matters of scientific significance, it is not clear that such criteria overcome the general strategies that have undone global realist arguments. In particular they do not seem to discriminate effectively between what is real and what is merely useful (and so between realism and instrumentalism).

3 Alternatives to realism

Several alternatives to realism have developed in the course of these debates. Principal among them are Putnam's 'internal realism' (1981) (see Putnam, H. §8), van Fraassen's 'constructive empiricism' (van Fraassen 1980) and what Fine (1986) calls 'the natural ontological attitude', or NOA. In a chameleon-like move, Putnam switched from being realism's champion to its critic. Rejecting what he called 'metaphysical realism' (associated with a 'God's eye view'), Putnam proposed a perspectival position in which truth is relative to language (or conceptual scheme). He could then allow scientific claims to be true in their proper domain but deny that they tell the whole story, or even that there is a whole story to tell. His picture was that there could be other truths - different stories about the world - each of which it may be proper to believe. Van Fraassen's constructive empiricism eschews belief in favour of what he calls commitment. He takes the distinguishing features of realism as twofold: realism seeks truth as a goal, and when a realist accepts a theory it is accepted as true. Constructive empiricism, by contrast, takes empirical adequacy (not truth) as the goal of science, and when it accepts a theory it accepts it as empirically adequate. This involves commitment to working within the framework of the theory but not to believing in its literal truth. Unlike these others, Fine's NOA is not a general interpretive scheme but simply an attitude that one can take to science. The attitude is minimal, deflationary and expressly local. It is critically positive, looking carefully at particular scientific claims and procedures, and cautions us not to attach any general interpretive agenda to science. Thus NOA rejects positing goals for science as a whole, as realists and constructive empiricists do. NOA accepts 'truth' as a semantic primitive, but rejects any general theories or interpretations of scientific truth, including the perspectivalism built into internal realism and the external-world correspondence built into realism itself. NOA is perhaps better classified as a nonrealism than as an antirealism.

It is interesting to contrast how these positions respond to good science. Realism accepts good science as true of an observer-independent world; internal realism accepts it as true relative to our scheme of things; constructive empiricism accepts it only as empirically adequate. NOA simply accepts it. This brings out two significant features of the recent debates. One is that they are more about the reach of evidence (what kind of acceptance is warranted) than about the metaphysical character of the objects of belief. The contrast also shows that major contenders, whether realist or not, share a basically positive attitude towards science. This has not always been acknowledged and a contrary suspicion still attaches to constructivism, which is frequently regarded as anti-science.

4 The constructivist challenge

Contemporary developments in the history and sociology of science have revived constructivist approaches (see Constructivism). Sharing with instrumentalism and other forms of pragmatism an emphasis on science as an activity, constructivism borrows the Marxist vocabulary of the 'production' of ideas (see Marxist philosophy of science) to place science among the manufacturing institutions. Specifically, what science makes is knowledge, which includes concepts and theories, along with things and even facts. Constructivism also emphasizes that science is open-ended. It highlights the role of unforced judgment in scientific practice, challenging the picture of a strict scientific method and of decision-making forced by rationality at every turn. The upshot is to see science as a form of human engagement like others; just people doing their own thing as best they can. Many regard this placement of science as a displacement, demoting science from its privileged position as the paradigm of rational and objective inquiry.

The emphasis on human constructions may challenge the mind-independence that is the hallmark of realist metaphysics. The respective roles of the social order and of nature in shaping these constructions, however, differ among constructivists, making for strong idealism at one pole (see Idealism) and pragmatic realism at the other. Despite these differences, constructivism challenges the unique position that realism marks out for itself with respect to ongoing science. If we look beyond the relatively sophisticated arguments for realism rehearsed above, perhaps realism's major hold on our attention is its claim to offer the only viable setting for understanding

scientific practice. We are told that unless we take scientists to be engaged in finding out about a world not of their own making we cannot begin to understand how science works. The major constructivist challenge is right here. The heart of constructivism consists in richly detailed studies of science in action. These studies set out to understand how science actually proceeds while bracketing the truth-claims of the area of science under investigation. Instead, constructivists typically employ little more than everyday psychology and an everyday pragmatism with respect to the common objects of experience. To the extent to which these studies succeed they paint a picture of science quite different from realism's, a constructivist picture that may undermine not only the arguments but also the intuitions on which scientific realism rests.

See also: Dewey, J.; Empiricism; Fictionalism; Pragmatism; Realism and antirealism; Theories, scientific

ARTHUR FINE

References and further reading

Cartwright, N. (1983) *How the Laws of Physics Lie*, New York: Clarendon Press.(Advances the causal argument mentioned in §2.)

Devitt, M. (1983) 'Realism and the Renegade Putnam', *Nous* 17: 291-301.(Critical discussion of Putnam's 'internal realism' of §3.)

Fine, A. (1986) *The Shaky Game: Einstein, Realism and the Quantum Theory*, Chicago, IL: University of Chicago Press.(Elaborates on NOA of §3. The Afterword of the 2nd edition, 1996, brings the discussion up to date.)

Fraassen, B.C. van (1980) *The Scientific Image*, Oxford: Clarendon Press.(The 'constructive empiricism' of §3.)

Galison, P. and Stump, D. (eds) (1996) *The Disunity of Science: Boundaries, Contexts, and Power*, Stanford, CA: Stanford University Press.(Essays by leading figures touching the constructivist themes of §4.)

Hacking, I. (1983) *Representing and Intervening*, Cambridge: Cambridge University Press.(Advances the 'experimental argument' mentioned in §2.)

Leplin, J. (ed.) (1984) *Scientific Realism*, Berkeley, CA: University of California Press.(Significant essays that signalled a turning away from realism. The essay by Boyd expands on the material of §1, that of Laudan responds to the 'miracles' argument of §1, and purses the pessimistic meta-induction referred to in §2, while the essay by McMullin develops the fruitfulness line mentioned in §2.)

Miller, R.W. (1987) *Fact and Method*, Princeton, NJ: Princeton University Press. (Book-length defence of a piecemeal realism, as in §2.)

Papineau, D. (ed.) (1996) *The Philosophy of Science*, Oxford: Oxford University Press.(Essays by leading figures touching on many of the themes related to §2 and §3.)

Popper, K. (1956) 'Three Views Concerning Human Knowledge', in H.D. Lewis (ed.) *Contemporary British Philosophy*, London: Allen & Unwin; repr. in *Conjectures and Refutations*, London: Routledge & Kegan Paul, 1972, 97-119.(Contains the attack on instrumentalism mentioned in §1.)

Putnam, H. (1975) *Mathematics, Matter and Method*, vol. 1, Cambridge: Cambridge University Press.(The 'miracles' argument of §1 occurs on page 73.)

Putnam, H. (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press.(The 'internal realism' of §3.)

Rosen, G. (1994) 'What is Constructive Empiricism?', *Philosophical Studies* 74: 143- 78.(Critical discussion of van Fraassen's 'constructive empiricism' of §3.)

Rouse, J. (1996) *Engaging Science: How to Understand its Practices Philosophically*, Ithaca, NY: Cornell University Press.(Critical discussion of Fine's NOA of §3, and the constructivism of §4.)

Smart, J.J.C. (1963) *Philosophy and Scientific Realism*, London: Routledge & Kegan Paul.(Defence of realism by means of the 'miracles' argument of §1.)

Scientific realism and social science

A central issue in the philosophy of the social sciences is the possibility of naturalism: whether disciplines such as sociology, anthropology, economics and psychology can be 'scientific' in broadly the same sense in which this term is applied to physics, chemistry, biology and so on. In the long history of debates about this issue, both naturalists and anti-naturalists have tended to accept a particular view of the natural sciences - the 'positivist' conception of science. But the challenges to this previously dominant position in the philosophy of science from around the 1960s made this shared assumption increasingly problematic. It was no longer clear what would be implied by the naturalist requirement that the social sciences should be modelled on the natural sciences. It also became necessary to reconsider the arguments previously employed by anti-naturalists, to see whether these held only on the assumption of a positivist conception of science. If so, a non-positivist naturalism might be defended: a methodological unity of the social and natural sciences based on some alternative to positivism. That this is possible has been argued by scientific realists in the social sciences, drawing on a particular alternative to positivism: the realist conception of science developed in the 1970s by Harré and others.

1 Realist v. positivist views of science

Within the philosophy of science, scientific realists criticize positivism both on philosophical grounds, and for failing adequately to represent the nature of scientific theories and their historical development. Rejecting the positivists' deductive-nomological model of explanation as unable to distinguish predictive from explanatory power, realists argue that scientific explanation consists in describing the processes through which observable phenomena, and whatever regularities obtain between them, are generated by the operation of (typically unobservable) underlying structures and mechanisms. In doing so, they likewise reject the regularity view of causation and defend a concept of natural necessity together with a Lockean view of causal powers.

This view of explanation brings with it a far less restrictive attitude towards the ontological commitments of scientific theories than positivism permits. For although (most) positivists accept what may be termed the 'minimal realist' claim that science aims at knowledge of an external world whose existence and nature are independent of the knowing subject (see Realism and antirealism), scientific realists criticize the positivists' tendency to limit what can be said to exist to what is observable, directly measurable, and so on. They therefore reject both operationalist and instrumentalist interpretations of the cognitive status of theoretical concepts, instead regarding them as having, at least potentially, a straightforwardly referential function (see Theories, scientific). Further, whereas for positivists, scientific progress is largely a matter of the increasing scope and hence predictive-explanatory power of universal laws, for scientific realists it consists primarily in increasing theoretical depth, with each level of causally operative structures and mechanisms being successively explained by reference to other, deeper levels. In such developments, existential hypotheses concerning previously unimagined entities - by contrast with the universal hypotheses emphasized by positivists - often play a central role.

Adopting this alternative to the positivist conception of science, realist naturalists have attempted both to articulate the character of a realist social science (§2), and to disarm a number of influential anti-naturalist arguments (§3).

2 The nature of a realist social science

Some realists, notably Bhaskar in *The Possibility of Naturalism* (1979), have seen it as their task to identify the ontological presuppositions of a realist social science, especially the nature of the agency-structure relationship (see Critical realism). But others, such as Keat and Urry in *Social Theory as Science* (1975), have pursued more modest aims: to free social science from the unnecessary and impoverishing limitations of positivism, and to indicate what kinds of theoretical work in the social sciences would be consistent with, or suggested by, their being modelled on a realist view of natural science.

Obvious examples of such positivistically inspired limitations are provided by the various forms of behaviourism in psychology and other social sciences (see Behaviourism, methodological and scientific). For realists, mentalistic concepts can be seen as performing a similar function to theoretical concepts in the natural sciences: potentially referring to unobservable entities, structures and processes, and employed in theories that can be tested for their explanatory power like any others. Hence there is no a priori reason either to eliminate them altogether, to provide

them with operational definitions in behavioural terms, or to regard them instrumentally, as no more than convenient fictions or predictive devices.

Nor, for the realist, need the realm of the mental be restricted to that of consciousness. Hence, whereas positivists have often regarded psychoanalytic theory as ‘unscientific’ or even ‘metaphysical’, realists may view Freud’s arguments for the postulation of ‘the unconscious’ as on a par with those typically adduced in postulating theoretical entities in the natural sciences, and his attempts to specify the nature of its structures and mechanisms as, in principle, a perfectly legitimate exercise in scientific theorizing.

However, none of this should be taken to imply that realists do or must endorse the substantive truth or explanatory power of mentalistic or psychoanalytic theories. A realist philosophy of the social sciences provides no more guidance about the truth or falsity of specific theoretical claims than does realism in the natural sciences. This applies equally to what has been, in the work of many realist naturalists, the main example of a substantive social theory open to a realist interpretation, namely Marx’s materialist theory of history and his attempts to identify the underlying structures of the various ‘modes of production’, such as capitalism and feudalism, and the nature of their respective generative mechanisms, such as their specific means of extracting surplus value.

Since Marxism, psychoanalytic theory and mentalistic psychology all seem to provide bona fide examples of the kinds of theoretical enterprise permitted by realism, its adoption should not be seen as implying commitment to a materialist ontology: not only material but also ideational items (such as beliefs, values and conceptual schemes) may legitimately be referred to by the theoretical concepts of a realist social science. Indeed realism would seem to be consistent even with ‘idealism’ in the social sciences, that is, with the claim that the social world is exclusively ideational in character: provided, of course, that these ideational items meet the minimal realist requirement of existing independently of the social scientist’s supposed knowledge or theoretical representation of them.

There is, however, a quite different sense of ‘idealism’, according to which the ‘objects of knowledge’ are themselves constructed or constituted by the knowing subject, and/or by the conceptual frameworks within which specific scientific theories are articulated (see Idealism). Idealism in this latter sense is clearly incompatible not just with scientific realism in the social sciences, but with any minimally realist view of either the social or natural sciences. These two senses of ‘idealism’ are often confused in debates about the social sciences: in particular, by some post-structuralist social theorists who adopt both forms of idealism without noticing their inconsistency in doing so.

3 Scientific realism and anti-naturalism

In its most general form, the anti-naturalist claim is that there is something about the character of the social world that makes it an impossible or inappropriate object of inquiry modelled upon the natural sciences. What realist naturalists try to show, in effect, is that this ‘something’, while excluded by a positivist social science, can find a suitable home in a realist one.

The most influential versions of anti-naturalism focus on the distinctively ‘meaningful’ character of social phenomena, and argue that these require, in place of scientific explanation, some form of ‘understanding’: the so-called ‘understanding *v.* explanation’ thesis. According to one version of this thesis, such understanding consists in identifying the reasons that people have for their actions, and/or in relating these to social rules of various kinds. To this the realist may respond by claiming that, once positivistic restrictions on theoretical entities are removed, there are no grounds for excluding such ‘meaningful’ items from the social sciences, nor for rejecting the possibility of causal relations obtaining between them. Thus such ‘understanding’ of actions can be seen as a form of causal explanation, with agents’ reasons as the mental causes (beliefs, desires and so on) of their behaviour. Correspondingly, the distinction between (physical) behaviour and (meaningful) action, much cited by anti-naturalists, may be interpreted by realists as one between observable phenomena and a theoretical redescription of these referring to their presumed causal determinants, something that is commonplace in the natural sciences.

But there are other versions of the understanding *v.* explanation thesis, especially those associated with the hermeneutic tradition, that are less obviously open to this naturalist rejoinder from a realist standpoint (see Hermeneutics). Here the concern is not with the nature of the relationship between reasons/rules and actions, but with the epistemological and methodological character of the ‘interpretive’ process through which the

identification of these reasons and rules itself takes place.

Consider, for example, the claim that agent A’s reason for action X is their belief that P, where P is some proposition. Even if, as the realist maintains, A’s believing that P can in principle be seen as (part of) the cause of their doing X, there remains the question of what exactly is involved in the identification of P itself. There is surely some process of ‘interpretation’ here, through which the meaning or content of this proposition is grasped or understood; and similarly for social rules and the like. Nor need this interpretive process stop at ‘surface’ meanings: one may also go further, through various ‘depth-hermeneutic’ procedures, to arrive at more fundamental levels of belief, meaning and so on.

Admittedly, there is also a necessary role for such interpretive processes in the conduct of the natural sciences, namely in the communicative interactions that take place between members of a scientific community. But there, it would seem, such interpretive procedures play no part in identifying or understanding the objects of scientific inquiry themselves. For these, unlike their counterparts in the social sciences, are essentially non-meaningful in character.

The scientific realist may respond to such claims by arguing that provided the structures of meaning arrived at in this way meet the minimal realist requirement of existing independently of the interpretive procedures employed - something that is accepted by some, though by no means all hermeneutic theorists - a qualified form of naturalism remains defensible. For the interpretation of meanings may then be seen, not as replacing the explanatory aims of a realist social theory, but rather as serving to identify the nature of the entities to which reference must be made in pursuing that explanatory project. Thus interpretation, while necessary given the distinctive nature of social phenomena, is not by itself enough: it must be combined with causal explanation of a recognizably ‘scientific’ kind.

However, this partial concession by the realist is unlikely to satisfy most anti-naturalists. Many would deny, for example, the possibility of objective hermeneutic interpretation, or would view with suspicion the realist’s attempt to assimilate the ‘unobservability’ of structures and mechanisms in nature and that of meanings in the social world. At this point, perhaps, what is required is a more nuanced and discriminating account of naturalism and anti-naturalism themselves, to make clear just what it is that the natural and social sciences are being asserted or denied to have in common.

See also: Naturalism in social science; Positivism in the social sciences; Scientific realism and antirealism

RUSSELL KEAT

References and further reading

Benton, T. (1977) *Philosophical Foundations of the Three Sociologies*, London: Routledge & Kegan Paul. (Includes a realist interpretation of Marxist theory.)

Bhaskar, R. (1979) *The Possibility of Naturalism*, Brighton: Harvester Wheatsheaf, 1989.(Defends a qualified naturalism based on a philosophically distinctive form of scientific realism.)

Harré, R. (1970) *The Principles of Scientific Thinking*, London: Macmillan.(A key statement of the realist alternative to positivism in the natural sciences.)

Harré, R. (1986) *Varieties of Realism*, Oxford: Blackwell.(A useful guide to different forms of realism in science.)

Keat, R. (1981) *The Politics of Social Theory*, Oxford: Blackwell.(A realist response to Habermas’ critique of positivism, including discussion of the scientific status of psychoanalytic theory.)

Keat, R. and Urry, J. (1975) *Social Theory as Science*, London: Routledge, 2nd edn, 1982.(A statement of the case for realism in the social sciences.)

Outhwaite, W. (1987) *New Philosophies of Social Science*, London: Macmillan.(A succinct and informative comparison of realism, hermeneutics and critical theory.)

Outhwaite, W. (1990) ‘Realism, Naturalism and Social Behaviour’, *Journal for the Theory of Social Behaviour* 20: 365-77.(Reviews much of the literature on realism in the social sciences over the preceding twenty years.)

Sayer, A. (1984) *Method in Social Science*, London: Hutchinson.(Develops the implications of Bhaskar’s realism for research methods in the social sciences.)

Stockman, N. (1983) *Antipositivist Theories of the Sciences*, Dordrecht: D. Reidel.(Criticizes scientific realism from a standpoint sympathetic to Habermas’ critical theory; extensive bibliography.)

Scope

Scope is a notion used by logicians and linguists in describing artificial and natural languages. It is best introduced in terms of the languages of formal logic. Consider a particular occurrence of an operator in a sentence - say, that of ' $\rightarrow$ ' in (1) below, or that of the universal quantifier ' $\forall$ ' in (2) below.

(1) $A \rightarrow (B \,\&\, C)$
(2) $\forall x(Bxy \rightarrow \exists yAxy)$

Speaking intuitively, the scope of the operator is that part of the sentence which it governs. The scope of ' $\rightarrow$ ' in (1) is the whole sentence; this renders the whole sentence a conditional. The scope of 'Čamp;', on the other hand, is just '(B & C)'. In (2), the scope of the quantifier ' $\forall$ ' is the whole sentence, which allows it to bind every occurrence of x. The scope of ' $\exists$ ' is only ' $\exists yAxy$ '. Since 'Bxy' is outside its scope, the 'y' in 'Bxy' is left unbound.

Although the importance of scope is semantic, it is usually identified with a syntactic relation between an occurrence of an expression in a sentence and a part of that sentence. In sentence and quantifier logic, the scope of an occurrence of an operator in a sentence *S* is usually defined as the smallest sub-sentence of *S* containing the occurrence. This ensures that for any two (occurrences of) operators with overlapping scope, exactly one is in the other's scope. In '$A \rightarrow (B \,\&\, C)$', for example, 'Čamp;' is in the scope of ' $\rightarrow$ ', and not vice versa. When the scope of one operator includes that of a second, the first is said to have wider scope than the second; the second's scope is narrower than that of the first.

Many semantically significant properties are characterized in terms of scope. The main logical operator of a sentence is the operator with the whole sentence as its scope. Thus conjunction, universal quantification, necessitation and so on are defined in terms of scope. A variable is bound whenever it occurs in the scope of a quantifier on it. A *de re* sentence is a sentence, such as ' $\exists x$ I believe that *x* is a spy', in which a variable in the scope of a non-extensional operator (like 'believes' or 'necessarily') is bound by a quantifier outside that operator's scope. *De dicto* sentences are ones which do not involve such 'quantifying in' (see De re/de dicto).

Differences in relative scope can affect a sentence's truth-conditions. 'Many hate a few' has an interpretation, in which 'many' gets wide scope, on which it says that many have the property 'hating a few'. It has a quite different reading in which 'a few' gets wide scope, on which it says that a few have the property 'being hated by many'. If the quantifier 'someone' in 'I believe that someone is a spy' has a narrower scope than 'believes', the sentence does not imply that there is some individual whom I think is a spy; if the quantifier's scope is wider, it does imply this.

In a well-designed formal language, the relative scope of operators is indicated by surface syntax, but this is not inevitably true in natural language. A natural language sentence with *n* operators has potentially *n*! (*n* factorial: $n(n-1)\ldots 3 \times 2 \times 1$) readings, since there are in principle that many ways of ordering the scopes of the operators. (Not all such orderings need be linguistically possible, of course.) Such ambiguities need to be resolved by an interpretive process, or a user's intentions, before the sentence can be evaluated for truth. Some syntactic theories posit a stage in sentence generation ('logical form') at which quantifier scope ambiguities are resolved by moving quantifiers so that their relative positions encode their scope (see Logical form).

A quantifier's scope should be the domain in which pronouns 'anaphoric on' (referring back to) the quantifier behave as variables bound by it. For example, the occurrence of 'her' in

(1) Every woman knew a man who loved her,

if anaphoric on 'every woman', behaves as a bound variable, since (1) is then understood as equivalent to

(1′) $\forall x(x$ is a woman $\rightarrow x$ knew a man who loved *x*),

and the scope of 'every woman' is therefore the whole sentence.

Observe that only some uses of pronouns anaphoric on quantifiers behave like variables bound by those

quantifiers. For example, consider the sentence

(2) Garth bought some dogs and Wayne vaccinated them.

Even when anaphoric on ‘some dogs’, ‘them’ in (2) cannot be treated as a bound variable. If it were a bound variable, then the sentence as a whole would be equivalent to

(3) There are some dogs Garth bought which Wayne vaccinated,

but (3) does not imply that Wayne vaccinated *all* the relevant dogs, as (2) does. (Gareth Evans (1985) seems to be the first person to have noticed this.) So ‘them’ is to be excluded from the scope of the quantifier ‘some dogs’. In accounts of natural language grammar which identify sentences with tree structures, quantifier scope is often identified with a structural relation between tree parts called c-command. While an expression in one sentence may be anaphoric on one in another, it is not possible for an expression in one sentence to c-command an expression in another. In particular, ‘some dogs’ in (2) does not c-command ‘them’. It follows, on such accounts, that not all pronouns anaphoric on a quantifier are bound thereby.

Russell (1910-13) speaks of the scope of a description in giving his ‘definition in use’ for (that is, method for eliminating) definite descriptions. When the scope of a description ‘the *F* in a sentence *S* is a sub-sentence ‘*T*(the *F*)’ of *S*, Russell says *S* is equivalent to the result of substituting (a formalization of) ‘there is exactly one *F*, and *T*(it)’ for ‘*T*(the *F*)’ in *S*. For example, if, in ‘Possibly, the mayor is mad’, the scope of the description is the whole sentence, the sentence is equivalent to ‘There is exactly one mayor and possibly it is mad’. If the scope of the description is just ‘The mayor is mad’, the sentence is equivalent to ‘Possibly: there is exactly one mayor and it is mad’. Russell’s account makes the definite article semantically analogous to a binary quantifier.

Expressions other than quantifiers and connectives also have scope. Variable binders (such as function abstractors) and verbs which take sentential complements (for example, ‘to believe’) are examples. Standard definitions extend straightforwardly here.

One hears it said that proper names ‘always take wide scope’. This seems intended as the claim that what is said by a use of a sentence *S*(*n*), with *n* a proper name, is what would have been said by an appropriately related use of a sentence of the form ‘*n* is such that *S*(it)’, with ‘it’ anaphoric on *n*.

See also: Anaphora; Descriptions; Quantifiers, substitutional and objectual

MARK RICHARD

References and further reading

Evans, G. (1985) *Collected Papers*, Oxford: Oxford University Press.(Essays 4, 5 and 8 discuss the interpretation of pronouns as bound variables. Somewhat difficult.)

Higginbotham, J. (1980) ‘Pronouns and Bound Variables’, *Linguistic Inquiry* 14: 679-708.(Proposes identifying Russell’s notion of scope with c-command. Difficult.)

Neale, S. (1990) *Descriptions*, Cambridge, MA: MIT Press.(Discusses and defends Russell’s account of descriptions. Also discusses contemporary accounts of binding and anaphoric relations. A good introduction to these topics.)

Russell, B. (1905) ‘On Denoting’, *Mind* 14: 479-93.(Russell’s first presentation of his account of definite descriptions.)

Russell, B. and Whitehead, A.N. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(Volume 1 includes a formal treatment of descriptions and the scopes thereof.)

Searle, John (1932-)

John Searle was a pupil of J.L. Austin at Oxford in the 1950s. He is the Mills Professor of Mind and Language at the University of California, Berkeley, where he has taught philosophy since 1959. According to Searle, the primary objects of analysis in the philosophy of language are not expressions but the production of expressions, speech acts, in accordance with rules. Learning a language involves (often unconsciously) internalizing rules that govern the performance of speech acts in that language. Speech-act theory aims to discover these rules and is itself a part of action theory, which concerns intentional states directed at or about something. It follows that speech-act theory is part of a more comprehensive theory of intentionality.

1 Speech acts and intentionality

Searle's career has primarily been devoted to developing a general theory of intentionality, beginning with *Speech Acts* (1969), moving deeper into his analysis in *Intentionality* (1983). Pursuit of questions about how words relate to the world crystallized for Searle questions about how the mind relates to the world. The conditions of satisfaction of speech acts derive from the conditions of satisfaction of corresponding psychological states. In *The Rediscovery of the Mind* (1992), Searle reaffirms his general picture. A theory of language is incomplete without an account of the relation between mind and language and of how meaning - the derived intentionality of linguistic elements - is grounded in the more biologically basic intrinsic intentionality of the mind/brain.

Searle's view that intentional states are related to a network of other intentional states, though not indisputable, is nevertheless in the mainstream. Commitment to the network is a form of holism; only relative to a vast range of other intentional states does any single intentional state have conditions of satisfaction. His view that intentional states function only against a background of skills is less widespread. These skills cannot be reduced to rules, nor can they be explicated by appeal to intentional states themselves, for they do not function as representations. But understanding literal meaning, and in particular, metaphor, requires this background. An oddity of the background is that, though the network fades off into it, the background itself is not intentional, but still involves some kind of stance towards the world (see Holism: mental and semantic).

Underlying Searle's general approach is the view that any divide between the mental and the physical is logical, not ontological, since in his view intentional states are caused by, and realized only in, physical structures. However, unlike mainstream philosophers of mind who worry how physical phenomena can have intentional properties, Searle since the late 1970s has been saying that the mind-body problem has a simple solution. Instead of this problem, Searle takes consciousness to be the central issue in the philosophy of mind. He has been successful in recruiting lots of others to refocus philosophical attention on consciousness (see Consciousness).

In *The Construction of Social Reality* (1995) Searle promotes his view about collective intentional behaviour. There is a difference between a group of people merely running around a field and a group playing football. According to Searle, this difference cannot be analysed as the sum of the behaviour of individuals. Since bodily movements can be (type-)identical in both cases, Searle infers that the difference lies in their minds. For those playing football, each has a collective intention of the form 'we intend to play football', distinct from, and irreducible to, individual intentions of the form 'I intend to run downfield twenty yards' or even 'I intend to play football'. Rejecting any suggestion that collective intentions are the product of a mysterious Hegelian communal mind, Searle incorporates collective intentions into his individualistic theory of intentionality. Since all mental phenomena are caused by operations of the brain and realized in its structure, there too must collective intentions lie.

2 Defending common sense

Though §1 sketches some of Searle's contributions to philosophy to date, it fails to reveal his greatest impact. He has enraged the philosophical community simply by defending the unexceptional. Here is a sampling of views he fiercely and proudly defends in the face of so-called academic achievement in fields as diverse as literary criticism, artificial intelligence, cognitive science, animal and child psychology, and of course, contemporary continental and analytical philosophy: the physical world is independent of its inhabitants and would continue to exist even if they did not (see Realism and antirealism); the world comprises facts which make statements and beliefs about it true or

false (see Truth, correspondence theory of); some statements are true solely because of what they mean, others because of the way the world is (see Analytic and synthetic); some translations are correct, others are incorrect; because and only because linguistic expressions have meaning, they can determinately apply to things in the world (see Reference); thoughts are in the head (internalism); because thoughts are (individuated by factors) in the head, it is (logically) possible that we are radically deceived about (the nature of) the external world (see Scepticism); I am (or can be) consciously aware of my thoughts and feelings in a way that others cannot (see Introspection, epistemology of); small children and some animals think and feel; one couldn't have a mind in virtue of running a computer program (see Chinese room argument); mental phenomena are just biological phenomena - that is, higher-level features of the brain - and so, studying the mind is studying the brain (biological naturalism); but since Searle maintains a logical distinction between mental and neurophysiological predicates, neurophysiology wouldn't give one any philosophical or conceptual insight into mental phenomena.

That it has become fashionable to contradict the views Searle defends might amaze a novice to philosophy. Even more amazing is that, though speech act theory has become academically fashionable (more in linguistics than in philosophy), it is Searle's defence of the unexceptional that has won him the reputation of philosophical maverick. Some have branded Searle an iconoclast sans argumentation, as philosophically unsubtle and obtuse. But with unflinching commitment, Searle defiantly rejects any philosophy that controverts the commonplace. Searle, not unlike J.L. Austin, Wittgenstein and Moore before him, looks to philosophy for analytic clarification, not metaphysical reduction or elimination.

See also: Intentionality; Speech acts; Analytical philosophy; Ordinary language philosophy

ERNIE LEPORE

List of works

Searle, J.R. (1969) *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.(Develops a theory of how language is an expression of thought.)

Searle, J.R. (1979) *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge: Cambridge University Press.(Further development of his theory of speech acts.)

Searle, J.R. (1983) *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press. (Develops a theory of intentionality.)

Searle, J.R. (1984) *Minds, Brains and Science*, Cambridge, MA: Harvard University Press.(Criticizes the very idea of artificial intelligence; includes famous 'Chinese room argument'.)

Searle, J.R. (1987) 'Indeterminacy, Empiricism, and the First Person', *The Journal of Philosophy* LXXXIV: 123-46.(An important critisism of Quine's indeterminancy thesis.)

Searle, J.R. (1992) *The Rediscovery of the Mind*, Cambridge, MA: MIT Press.(Criticizes the dominant traditions in the philosophy of mind- in particular, materialism and dualism.)

Searle, J.R. (1995) *The Construction of Social Reality*, New York: Free Press.(Examines the structures of the world that are facts only by human agreement.)

References and further reading

Lepore, E. and van Gulick, R. (eds) (1991) *John Searle and his Critics*, Oxford: Blackwell.(A useful collection of essays with a reply by Searle).

Second- and higher-order logics

In first-order predicate logic there are symbols for fixed individuals, relations and functions on a given universe of individuals and there are variables ranging over the individuals, with associated quantifiers. Second-order logic adds variables ranging over relations and functions on the universe of individuals, and associated quantifiers, which are called second-order variables and quantifiers. Sometimes one also adds symbols for fixed higher-order relations and functions among and on the relations, functions and individuals of the original universe. One can add third-order variables ranging over relations and functions among and on the relations, functions and individuals on the universe, with associated quantifiers, and so on, to yield logics of even higher order. It is usual to use proof systems for higher-order logics (that is, logics beyond first-order) that include analogues of the first-order quantifier rules for all quantifiers.

An extensional n-*ary relation variable in effect ranges over arbitrary sets of* n-*tuples of members of the universe. (Functions are omitted here for simplicity: remarks about them parallel those for relations.) If the set of sets of* n-*tuples of members of a universe is fully determined once the universe itself is given, then the truth-values of sentences involving second-order quantifiers are determined in a structure like the ones used for first-order logic. However, if the notion of the set of all sets of* n-*tuples of members of a universe is specified in terms of some theory about sets or relations, then the universe of a structure must be supplemented by specifications of the domains of the various higher-order variables. No matter what theory one adopts, there are infinitely many choices for such domains compatible with the theory over any infinite universe. This casts doubt on the apparent clarity of the notion of 'all* n-*ary relations on a domain': since the notion cannot be defined categorically in terms of the domain using any theory whatsoever, how could it be well-determined?*

1 Standard second-order logic

Only second-order quantification over relations will be discussed here. Higher-order quantification and quantification over functions can be treated in an entirely parallel way. Symbols for higher-order relations and functions will be omitted for simplicity. Supplement the language of first-order predicate logic (see Predicate calculus) with variables X_n^m for *m*-ary relations, for all natural numbers *m* and *n*. Add to the definition of a formula of first-order predicate logic that, for all natural numbers *m* and *n*, $X_n^m t_1 \ldots t_m$ is an atomic formula, where each t_i is a constant symbol or first-order variable; and that if ϕ is a formula, then so are $(\forall X_n^m)\phi$ and $(\exists X_n^m)\phi$. Free and bound second-order variables are defined in exact parallel to the definitions for first-order variables, and a sentence is a formula with no free variables of any order.

It will be convenient always to require of any given language that there be infinitely many relation symbols of each possible number of places and infinitely many constant symbols not in it. The definition of truth in a structure for first-order predicate logic is extended to the *standard semantics* for second-order logic by supplying additional clauses for second-order quantification as follows. Consider a structure A with language $\mathcal{L}$. Let *P* be some definite *m*-placed relation symbol not in $\mathcal{L}$, and define $\psi(P)$ to be the sentence which results from replacing every free occurrence of X_n^m in ψ by *P*.

(1) If the sentence ϕ is of the form $(\forall X_n^m)\psi$ for some *m* and *n*, then the truth-value of ϕ is T if the truth-value of $\psi(P)$ is T in every expansion of A to a structure the language that includes every member of $\mathcal{L}$ and also *P*. It is F otherwise.
(2) If the sentence ϕ is of the form $(\exists X_n^m)\psi$ for some *m* and *n*, then the truth-value of ϕ is T if the truth-value of $\psi(P)$ is T in at least one expansion of A to a structure for the language that includes every member of $\mathcal{L}$ and also *P*. It is F otherwise.

Once given the standard semantics a logical truth is defined to be a sentence true in every structure for the language of the sentence and a valid argument to be one such that there is no structure in which all the premises are true and the conclusion false, as for first-order logic. Moreover, since the notion of structure is unchanged from first-order logic, standard model-theoretic notions such as isomorphism and categoricity can be taken over unchanged from the model theory of first-order logic. (Two structures are said to be isomorphic if, roughly, there is a one-to-one correspondence between their universes that takes the interpretations of constant and relation

symbols in one to the corresponding interpretations in the other. A theory is said to be categorical if all its models are isomorphic.)

It is natural to turn now to a deductive system for second-order logic to characterize logical truths and valid arguments. This does not work out nearly so well for second-order logic as for first-order.

The core of any deduction system for second-order logic will be a first-order system supplemented with quantifier rules for the second-order quantifiers that exactly parallel those for the first-order ones, with *n*-placed relation symbols playing the role constant symbols did for first-order quantification in the case of second-order quantification over *n*-ary relations.

In standard second-order logic one assumes that for any formula ϕ with all of its free variables from the list $v_1, \ldots, v_m, v_{m+1}, \ldots, v_{m+n}$ and for any *n* individuals $b_1, \ldots, b_n$, there exists an *m*-ary relation comprehending exactly those *m*-tuples of individuals $a_1, \ldots, a_m$ which are such that

$$\phi[a_1, \ldots, a_m, b_1, \ldots, b_n]$$

holds, where the square brackets indicate that v_1 is replaced everywhere it occurs free in ϕ by a new constant symbol interpreted by a_1 in an expanded structure, and so on. The relation is said to be defined by ϕ with parameters $b_1, \ldots, b_n$. More precisely, one assumes that the following axiom schema of comprehension holds:

$$(\forall v_{m+1}) \ldots (\forall v_{m+n})(\exists X_1^m)$$
$$(\forall v_1) \ldots (\forall v_m)(X_1^m v_1 \ldots v_m \leftrightarrow \phi),$$

where *m* and *n* are natural numbers and ϕ is a formula with all of its free variables among $v_1, \ldots, v_{m+n}$. If the variables $v_1, \ldots, v_m$ do not occur in the formula ϕ, then the comprehension axiom entails that the empty *m*-ary relation is in the domain of the second-order variables when ϕ does not hold of $b_1, \ldots, b_n$, and that the universal *m*-ary relation is in the domain of the second-order variables when it does hold.

It is enough for many purposes to consider only the predicative axiom schema of comprehension, which requires additionally that ϕ be a formula that does not contain any second-order quantifiers.

One might add other axioms to a deduction system for second-order logic. However, there is no usable system of axioms that yields all the logical truths as theorems: if a deduction system for second-order logic is *sound* (that is, if it cannot be used to derive sentences that are not logical truths) and the theorems are recursively enumerable (that is, there are effective procedures for listing the axioms and checking deductions - see Computability theory), then it is incomplete in the sense that there is a logical truth that cannot be deduced in it.

The impossibility of providing a complete deduction system for second-order logic follows from Gödel's incompleteness theorems (see Gödel's theorems §§3-4) and the fact that there is a second-order sentence that is a categorical axiomatization of arithmetic (that is, an axiomatization such that any model is isomorphic to a model with universe the set of natural numbers with, say, for definiteness, the relation 'strictly less than' and the addition and multiplication operations - in less technical terms, this is an axiomatization that characterizes the natural numbers completely for all mathematical purposes). (A sentence that is a categorical axiomatization of arithmetic is described and discussed in §3 below.) Suppose the sentence Φ is a categorical axiomatization of arithmetic. A truth of arithmetic is just a sentence true of the natural numbers. Thus, a sentence ϕ is a truth of arithmetic just if it is true in every model of Φ, that is, just if the argument from Φ to ϕ is valid. But that means that a complete deduction system for second-order logic would yield a decision procedure for the truths of arithmetic: ϕ is true if deducible from Φ and not true if $\neg\phi$ is deducible from Φ, and a systematic enumeration of all possible deductions from Φ will eventually decide which of these is the case. Gödel's incompleteness theorem entails that no such decision procedure is possible, and so no such deduction system is possible.

2 Alternative semantics

Instead of requiring, as in the standard semantics, that second-order variables range over all relations on a domain, one can be more explicit about their range by specifying as part of a structure the domains of the second-order variables in addition to that for the first-order variables. A Henkin structure for second-order logic is, by definition, an ordinary structure (with a universe and an interpretation function) supplemented by a third

component: a domain function $\mathcal{D}$ on the natural numbers that takes each natural number *n* to a set of *n*-ary relations on the universe of the structure. For each natural number *n*, the set $\mathcal{D}(n)$ must contain every *n*-ary relation that is the interpretation of an *n*-placed relation symbol in the language of the structure. (This condition is frequently omitted. But no use is ever made of the Henkin structures that do not satisfy it.)

Let A be a Henkin structure with language $\mathcal{L}$, and let *P* and $\psi(P)$ be as above. In Henkin semantics, the clauses of the definition of truth for the quantifiers will be

(1′) If the sentence ϕ is of the form $(\forall X_n^m)\psi$ for some *m* and *n*, then the truth-value of ϕ is T if the truth-value of $\psi(P)$ is T in every expansion of A to a Henkin structure for the language that includes every member of $\mathcal{L}$ and also *P* and interprets *P* by a member of $\mathcal{D}(n)$. It is F otherwise.
(2′) If the sentence ϕ is of the form $(\exists X_n^m)\psi$ for some *m* and *n*, then the truth-value of ϕ is T if the truth-value of $\psi(P)$ is T in at least one expansion of A to a Henkin structure for the language that includes every member of $\mathcal{L}$ and also *P* and interprets *P* by a member of $\mathcal{D}(n)$. It is F otherwise.

One can define a Henkin logical truth to be a sentence true in every Henkin structure for the language of the sentence and a Henkin valid argument to be one such that there is no Henkin structure in which all the premises are true and the conclusion false. If one were to restrict attention to *full* Henkin models - in which for each *n* the value of $\mathcal{D}(n)$ is the set of all *n*-ary relations on the domain - one would recover the logical truths and valid arguments of standard second-order logic. There are possibilities intermediate between considering all Henkin structures and considering only full ones: for example, that of restricting attention to Henkin structures that are models of the comprehension schema.

The second-order quantification of Henkin semantics resembles first-order quantification in that the universe of quantification is specified. That suggests going even further - to a first-order semantics for (syntactically) second-order logic. The domains of *n*-ary relations of Henkin structures are here replaced by perfectly arbitrary domains of objects - no longer required to be relations on the universe of individuals - and the atomic formulas of the form $X_n^m t_1 \ldots t_m$, which are now just of the semantic form of a list of $m+1$ first-order constants and variables, are interpreted using what are called ‘predication relations’ (which should really be called ‘relation relations’ but are not, for obvious reasons). For example, $X_n^m v_1 \ldots v_m$ becomes $P_m X_n^m v_1 \ldots v_m$, where P_m is the predication relation for *m*-ary relations, which may be read ‘the relation X_n^m holds of $v_1, \ldots, v_m$’.

Once there are predication relations, there is no need to introduce any other relations, except as objects in the first-order domains - all the relations can be taken to be members of the universes of the second-order variables. It is therefore possible to assume that the only relation symbols in a first-order structure for second-order logic are those for the predication relations and equality. This is not absolutely essential, but it does simplify matters considerably. A *first-order structure for second-order logic* will be constituted, as usual, of a non-empty universe of individuals, a domain function $\mathcal{D}$ on the natural numbers that takes each number to a non-empty set, and an interpretation function. The interpretation function interprets constant symbols as members of the various universes, and it interprets each predication relation symbol P_n as a relation that can hold only between a member of $\mathcal{D}(n)$ and *n* members of the universe of individuals, that is, the interpretation of each P_n is a set of $n+1$-tuples with first members from $\mathcal{D}(n)$ and all other members from the universe of individuals.

A Henkin structure can be converted into a first-order structure for second-order logic as follows. The domain function $\mathcal{D}$ is simply the domain function of the Henkin structure. The predication relations are based on membership relations: the *n*th predication relation holds between $R \in \mathcal{D}(n)$ and *n* objects just if the *n*-tuple formed by the *n* objects is a member of *R*. Each *n*-placed relation symbol must be replaced by a constant symbol that has the same interpretation, now viewed as a member of $\mathcal{D}(n)$. The semantics for second-order logic with first-order structures is just a standard first-order semantics for a multi-sorted structure, that is, one with many universes. The variables associated with a given universe (in our example, variables X_m^n are associated with $\mathcal{D}(n)$) are only interpreted by (or replaced by constant symbols interpreted by) members of that universe. The device of using many universes instead of one involves less than meets the eye: it is possible using entirely elementary methods to replace structures with many universes with ordinary single-universe structures that are equivalent in a suitably strong sense.

One can make use of intensional first-order structures for second-order logic, but here we restrict our attention to

extensional ones, which are, by definition, those that are models of the following schema:

$$(\forall X_1^n)(\forall X_2^n)((\forall v_1)\dots(\forall v_n)$$
$$(X_1^n v_1 \dots v_n \leftrightarrow X_2^n v_1 \dots v_n) \rightarrow X_1^n = X_2^n).$$

The schema says that relations that hold of the same *n*-tuples are identical.

In extensional first-order structures for second-order logic, the set of *n*-tuples of which a relation holds fully determines the relation, and so extensional first-order structures for second-order logic might as well just be Henkin structures: every such structure A gives rise to a Henkin structure by dropping the predication relations and, for each *n*, replacing each member *R* of $\mathcal{D}(n)$ by the set of all *n*-tuples $(a_1, \dots, a_n)$ such that $(R, a_1, \dots, a_n)$ is in the *n*th predication relation of the first-order structure. Constant symbols that stand for objects in $\mathcal{D}(n)$ must be replaced by *n*-placed relation symbols that are interpreted as the relations corresponding to the given objects. The resulting Henkin structure, converted to a first-order structure by the procedure described above, yields a first-order structure suitably similar to the original structure A.

3 Metamathematics

Dedekind observed that a set is finite if and only if there is no one-to-one correspondence between the set and any of its proper subsets. Thus, the following formula, $\Psi(X_1^1)$, on the standard semantics, expresses that the predicate X_1^1 holds of finitely many things:

$$(\forall X_1^2)(((\forall v_1)(X_1^1 v_1 \rightarrow (\exists v_2)(X_1^1 v_2 \wedge X_1^2 v_1 v_2))$$
$$\wedge (\forall v_1)(\forall v_2)(\forall v_3)(X_1^2 v_1 v_2 \wedge X_1^2 v_1 v_3 \rightarrow v_2 = v_3)$$
$$\wedge (\forall v_1)(\forall v_2)(\forall v_3)(X_1^2 v_1 v_3 \wedge X_1^2 v_2 v_3 \rightarrow v_1 = v_2))$$
$$\rightarrow (\forall v_2)(X_1^1 v_2 \rightarrow (\exists v_1)(X_1^1 v_1 \wedge X_1^2 v_1 v_2))).$$

Using this formula one can categorically axiomatize the natural numbers as a discrete linear order with a least element and no greatest element (‘discrete’ means that every element that has a predecessor (successor) has an immediate predecessor (successor)) in which the set of elements less than any element is finite. One can then supplement this with standard first-order axiomatizations of, for example, addition and multiplication. It is straightforward to give a first-order axiomatization of the discrete linear orders with a least element and no greatest element. The only second-order part of the categorical axiomatization is the use of the notion of finitude. That is our main concern here.

Note that on the Henkin interpretation of second-order logic $\Psi(X_1^1)$ does not define finitude: in order for the definition to work, it is crucial that the domain of the variable X_1^2 include relations that yield *all* one-to-one correspondences from X_1^1 to itself. To give a trivial example, $\Psi(X_1^1)$ holds vacuously of any set - including an infinite one - in a Henkin structure in which there simply are not any binary relations that relate objects in the set to other objects in the set. It is clear that this example may in some sense be taken to be artificial. Indeed, the constraints we have placed on which Henkin structures are to be allowed exclude ones in which there are no binary relations whatever on some part of the universe. But if constraints are to be imposed axiomatically - that is, by restricting attention to Henkin models of some set of axioms - then no constraints help to ensure that Ψ, or indeed any other formula, characterizes finitude. Any suitable axiomatic constraints would, as was in effect shown in §2, carry over into first-order axiomatic constraints such that finitude could be characterized over the first-order structures for second-order logic that satisfy them.

But the notion of finitude cannot be characterized axiomatically in first-order logic, since first-order logic is *compact*. The compactness theorem for first-order logic states that if every finite subset of a set of sentences of first-order logic has a model, then so does the whole set. Suppose a set *S* of sentences of first-order logic has models in which the interpretation of a predicate symbol *P* is of arbitrarily large finite size. Let S' be the set of sentences including those in *S* plus sentences of first-order logic formalizing ‘There is at least one thing *v* such that $P(v)$’, ‘There are at least two things *v* such that $P(v)$’, and so on. Then every finite subset of S' has a model. But then by the compactness theorem the whole set has a model, in which the extension of *P* must be infinite. So if a set *S* of sentences of first-order logic has models in which the interpretation of *P* is of every finite size then it has a model in which the interpretation of *P* is infinite. It fails to characterize finitude. (See Model theory.) Thus, the

notion of finitude cannot be axiomatically characterized in second-order logic interpreted using either first-order or Henkin structures.

By a contraposition of this argument, since finitude can be characterized in standard second-order logic, standard second-order logic is not compact. Thus, a sentence of standard second-order logic can be a logical consequence of infinitely many sentences of standard second-order logic without being a consequence of any finite subset of them. This is a very strong failure of the very idea of having a deduction system for standard second-order logic.

In §1, it was shown that no deduction system captured the logical truths of standard second-order logic. It now follows that standard second-order logic is also difficult to characterize in a second sense: the Henkin structures that are associated with standard second-order logic - the full Henkin structures - cannot be characterized axiomatically, in the sense that there is no set of axioms, not even the set of all the logical truths of standard second-order logic, that hold of all and only the full Henkin structures. That follows because if there were such a set of axioms (since finitude can be characterized in standard second-order logic) they could be used to characterize axiomatically the notion of finitude.

The situation is entirely similar with many other familiar mathematical notions: like finitude, many notions that mathematicians seem to feel they understand have adequate definitions in standard second-order logic but not in first-order logic.

See also: Logical and mathematical terms, glossary of; Second-order logic, philosophical issues in

SHAUGHAN LAVINE

References and further reading

Church, A. (1956) *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press.(An influential introduction to second- (and first-) order logic.)

Ebbinghaus, H.-D., Flum, J. and Thomas, W. (1984) *Mathematical Logic*, Berlin and New York: Springer, 2nd edn, 1994.(The mathematical results about second-order logic found in this article are proved in detail. Results characterizing first-order logic - and hence distinguishing it from second-order and other logics - are presented in an unusually accessible manner.)

Field, H. (1994) ‘Are our Logical and Mathematical Concepts Highly Indeterminate?’, in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 19, *Philosophical Naturalism*, Notre Dame, IN: University of Notre Dame Press, 391-429.(The ways in which second-order logic does and does not help in the determination of our logical and mathematical concepts are analysed in some detail, and a number of arguments and positions are carefully laid out.)

Hilbert, D. and Ackermann, W. (1928) *Grundzüge der theoretischen Logik*, Berlin: Springer, 2nd edn, 1938; trans. L.M. Hammond, G.G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, ed. R.E. Luce, New York: Chelsea, 1950.(A classic introduction to logic, including second-order logic; quite clear and detailed.)

Lavine, S. (1994) *Understanding the Infinite*, Cambridge, MA: Harvard University Press.(Though the main focus of this work is the theory of sets, it includes a detailed discussion of how a part of second-order logic plays a role in fixing mathematical concepts.)

Monk, J.D. (1976) *Mathematical Logic*, Graduate Texts in Mathematics 37, New York: Springer.(This text includes a concise but readable introduction to the basics of the syntax and semantics of second- and higher-order logic.)

Moore, G.H. (1988) ‘The Emergence of First-Order Logic’, in W. Aspray and P. Kitcher (eds) *History and Philosophy of Modern Mathematics*, Minneapolis, MN: University of Minnesota Press, 95-135.(A stellar synoptic treatment of the emergence of first-order logic - and hence, as a sort of residue, second-order logic - from earlier systems of logic that classified things rather differently. This article provides an excellent entry point into the modern history of the subject, with a very useful bibliography.)

Shapiro, S. (1991) *Foundations without Foundationalism*, Oxford: Clarendon Press.(A comprehensive work on basics of second-order logic and the ways in which it is relied on in much of ordinary mathematical practice. Extensive bibliography.)

Secondary qualities

Primary qualities are generally defined as those properties that objects have independently of being perceived. Standard examples would include properties of shape, weight, position, electric charge, atomic structure. These properties characterize the way the world is in itself, separately from mind. Secondary qualities, by contrast, are defined as those properties that incorporate sensory responses in their conditions of application, so that the idea of a perceiver is built into their nature. It is more controversial which properties, if any, belong to this category, since not all philosophers agree that the standard alleged examples of secondary qualities - colours, sounds, tastes, smells, feels - are really correctly so classified. Some thinkers hold that objects have only primary qualities. Let us note the significance of the question, concentrating on the case of colour, which is the one most frequently discussed.

Objects appear to have both shape and colour in equal measure, but is this really how things are? Depending upon how we answer this question, we get very different pictures of the relation between appearance and reality. If both sorts of property are equally out there, equally objective, then what appears to us in perception is reality itself. When we see a material object we see something that exists independently of our seeing it, and we see the object as it is whether or not there are (or even could be) any perceivers. But if the colour of the object is inherently dependent upon our sensory responses, then the question arises as to whether what we see is really in some way itself mental. If colour is a secondary quality, in other words, do we see things as they really are? What is it that bears colour if colours are in some way mentally constituted? Do we indeed see anything at all, as distinct from introspecting the features of our own subjective states?

1 Some theories

To the question, ‘What is colour?’, a number of answers might be given; we shall mention five, and then offer some evaluation of them.

(1) An answer that appeals to the disciple of science is that to be red is to instantiate some physical property of surfaces and media - as it might be, reflecting or transmitting light of a particular wavelength. That is, whatever science says is the primary quality basis of colour in the object is literally identical with that colour. The case is thus analogous to other theoretical identifications with which we have become familiar, say heat with molecular motion or light with volleys of photons. Science has, on this view, discovered what redness really is, deep down in its hidden nature. Possibly, there is no one physical property to which redness reduces, since different types of red object might not be unified by any single physical property; in which case the proponent of the current theory will opine that it is the disjunction of each of the many physical properties that underlie redness in different objects with which redness is identical. When we see an object as red, then, we are seeing the relevant physical property as presenting a certain sensory appearance; but the redness itself has a nature that is not given to us in the sensory presentations we record with the word ’red’. Redness thus has an invisible nature, not revealed to ordinary perception. The seen is but the appearance of the unseen.

(2) Discomfort with a reducibility thesis of the previous kind is apt to prompt a claim of irreducibility. Just as some philosophers have felt that there is no alternative to accepting the ultimate irreducibility of mental states, so it might be thought that colour properties simply are what they are and not anything else: to be red is to be, precisely, red - period. Nothing more can be said, nor need it be. There is no doubt an operative causal basis of colour in objects, where this basis explains how light of certain wavelengths reaches the human eye; but it is not true that colour is simply identical with that basis. There are no informative identities of the kind envisaged by the previous theory. As with so many other concepts, colour concepts pick out their corresponding properties in the only way they can be picked out. Nothing is hidden in the nature of colour itself: how we see redness is how it is, *in toto*. Colours are not physical properties at all, save in the trivial sense that physical objects have them; they form an irreducible family of properties, not assimilable to properties of other kinds. In an important sense, they are inexplicable, primitive aspects of the way things are.

(3) Sensing a subjective ingredient in colour, some thinkers have gone to the extreme of denying outright that external objects are coloured; what bear colour properties are rather perceptual experiences themselves, episodes in the mind. Just as it is mental states to which the predicate ’pain’ properly applies (see Bodily sensations), so it is

states of mind that are rightly said to be red or green. Seeings themselves are what is coloured; or, put less bluntly, we must introduce items called 'sense-data' to serve as the bearers of colour, where sense-data are construed as inherently mind-dependent (see Sense-data). When we say that a tomato is red we make an error of projection, spreading the mind upon the world. This type of view will then naturally go all the way and accept that the shapes of things we see are also not properties of external objects, since the shaped object we see is surely identical with the coloured object we see - it is not like a case of one object lying behind another. It will turn out then that we are not perceptually acquainted with objects in the external world at all but only with items that dwell in our minds. If there are any external objects, we do not see them; indeed, given that we do not see them, we must face the question of what point there is in assuming their existence. Hence idealism - the denial of anything aside from minds and their states. Because of colour the ordinary world gets extinguished (see Idealism).

(4) A radical move now might be to deny that anything is ever coloured: not any physical object, because colour is too sense-dependent; and not any mental object, because then the objective world becomes invisible to us, at best a lurking presence shrouded in deep shadow. We are, to be sure, acquainted with colours in experience, but it is a mistake to think that anything has them, in the way that primary qualities are had. To experience a red tomato is like having a red after-image: there is colour in the offing in the latter case, but no physical thing is thereby red and nor is any mental thing. Colour might be compared with a medium through which you see things; it is not a property of what is seen or of the seeing of it. Colour is like perspective, a mode of seeing rather than a thing seen. When we say that the tomato is red we can only mean that it is presented though the medium of redness - from that chromatic perspective, as it were - not that it literally instantiates a property. Colour terms are more like adverbs of seeing than adjectives of objects: just as we see things clearly, so we see things redly (see Mental states, adverbial theory of).

(5) The final theory has been historically the most popular: it attempts to register the subjectivity inherent in colour while not denying that external objects are the proper bearers of colour. By situating colours as deftly as possible between the external object of perception and the inner experience of seeing, the theory aims to preserve as much as possible of our common-sense view of the world. Colours, on this theory, are dispositions that objects have to cause sensory responses in perceivers under appropriate conditions of lighting and perceptual receptivity. Objects have physical properties that interact with light and thence with our nervous systems to produce colour experiences in us: for an object to have a given colour is then for it to have a certain complex relational property - that of having the physically grounded capacity to produce specific modes of experience in suitably placed perceivers. To be red is to have the disposition or ability to appear red to perceivers in virtue of properties that do not themselves include the property of being red - such as reflecting certain wavelengths of light. Colours are thus like such properties as fragility, digestibility, potency - powers to produce specific effects. They are indeed in objects, since it is the object that has the disposition, but they are defined by reference to the sensory effects that objects produce, so that an element of subjectivity is introduced. In effect, they are higher-order psychophysical properties - logical constructions from a (possibly variable) physical basis and a specific sensory type. The mistake of the other theories was to regard colours as monolithic non-relational properties; once we recognize their complex psychophysical relational character colours can be everything we want them to be. Out there, yes; but also rooted in what is in here.

2 Assessment

We shall mention here just a few of the salient problems with each approach. Theory (1) suffers from two main problems: first, it makes our knowledge of the nature of colour overly conjectural and empirically vulnerable; second, it imposes an absoluteness on colour concepts that they cannot plausibly bear. Thus, to take the first, if red is constituted by some underlying micro-property, then we do not know the true nature of red until we know the identity of that micro-property; and we do not know how red relates logically to other colours until we know how the underlying micro-properties relate. But both of these consequences are highly counterintuitive. And, secondly, when we consider a hypothetical case in which the same micro-property systematically interacts with various nervous systems so as to produce different types of experience from case to case, we find that our intuitions favour the idea that in such circumstances the perceivers in question would speak truly in predicating of the objects a colour property that matches their experience, not one that matches the experience of some other privileged group of perceivers. Less cumbrously put, if things that look red to us look green to Martians, in virtue of our different sensory systems, then Martians speak truly when they describe as green what we describe (truly) as red. Truths

about colour are relative, while micro-properties are absolute, so the former cannot be identical to the latter. To be more exact, given that a particular colour is essentially the colour it is - which seems hard to deny - we cannot allow that different micro-properties might constitute the same colour or the same micro-property different colours.

Theory (2) ventures little and so risks little: its chief failing, viewed as offering the whole truth, is that it turns its face against a plain conceptual connection, namely that between the colour of an object and how it appears to perceivers. As we have just noted, colour truths are sensitive to the facts of sensory response, and simply declaring them irreducible does nothing to accommodate that constitutive connection. The dispositional theory, by contrast, explicitly builds the link to experience into its definition of colour, thus acknowledging what the simple irreducibility theory ignores.

The trouble with theory (3) has already been indicated: it cuts us off perceptually from the real world, by making the proper objects of sight purely mental. And there is this correlative problem: if mental items are what bear colour, and if what bears colour is the same as what bears shape, then we are committed to holding that mental items instantiate primary qualities too. Not only are bits of my mind red - they are also spherical and hollow and heavy. But that is surely false, even a category-mistake: no experience of a red sphere can itself be a red sphere, on pain of cramming the mind with spatially extended objects, with all that that implies. The theory roundly conflates what an experience is of with the experience itself, with the absurd result that the mind comes to be credited with all the properties of the physical world.

Theory (4) is an obscure and desperate attempt to wriggle out of the kinds of problem just rehearsed. No matter where we put the colour property we get into trouble, so put it nowhere! Here the difficulty is immediate: surely it is excessively revisionary to maintain that colour words are incapable of predicating properties of anything. Grammatically they are just like other predicates, and we have no hesitation in saying that what we are seeing is red, just as it is spherical. When you perceive a multi-coloured expanse you surely experience something as having various colours; so your experience itself would have to be grossly delusory if this were, in truth, not so. Your experience is certainly not as of there being colourless things viewed through a coloured filter (whatever that might mean). The phenomenology of colour experience, as well as the language of colour, is powerfully object-ascriptive - as much so as for primary qualities.

The dispositional theory aims to preserve this feature of colour by identifying colours with relational properties of objects, while not severing the link with sensory response. It also allows us to accord the right kind of epistemological status to our knowledge of colour, since we know our own experiences with an immediacy that is not vulnerable to empirical upset. Relativity to perceivers is explicitly registered, as is the need for a categorical physical base in the object. Theory (5) thus appears to paint colour in the right hues, locating it at the appropriate juncture between subjective and objective.

But there is a fly in the ointment, which concerns the phenomenology of colour experience. For ordinary visual experience does not, on the face of it, present colours to us in the way they intrinsically are according to that theory: a red object does not look as if it instantiates a complex relational disposition in respect of appropriate perceivers; it looks, rather, as if it instantiates a simple monadic property of the object’s local surface. Indeed, dispositions do not in general look any way at all (consider solubility), while colour is paradigmatically perceptual, seeming to inhere firmly in the seen object and nowhere else. This would not be such a problem if we did not hold that you can tell the nature of colours just by seeing them - that they are transparent to perception. But that axiom of common sense is put under pressure if their nature is declared to be intrinsically relational, since this is not how they appear. Should we then give up the assumption of the perceptual transparency of colour, or should we conclude that the dispositional theory gets the colours wrong? Neither alternative delights the spirit. What we need, ideally, is a way to retain the benefits of the dispositional theory while avoiding its revisionary consequences. How can colours be dispositions and yet not look like them, given that they have to look the way they are?

This may seem a difficult feat to accomplish; the authority of colour vision seems to set an impossibly high standard for the dispositional theory to meet. But there does, happily, appear to be a way through the problem, which requires but a little delicacy in formulating the purport of the dispositional theory. Typically, the theory is propounded in a strong form, as maintaining an identity thesis as between the property of being red and the

property of having a categorically based disposition to appear red in certain circumstances to perceivers. The problem, then, is that this putative identity is not borne out by colour experience itself, which sets a condition of adequacy for any account of the kind of property redness is. Might we weaken the theory so as to relinquish any claim of identity while preserving the conceptual link we want to respect? We might indeed: let us say that being red is *dependent* upon the corresponding disposition but not identical with it. Or, to bring out the analogy with a similar strategy in the philosophy of mind, let us say that colour properties are supervenient upon the corresponding dispositions but they are not reducible to those dispositions - in somewhat the way that pain is supervenient on brain states yet is not reducible to them (see Supervenience; Supervenience of the mental). Moreover, let us add that the dependence in question is conceptual, so that we can recognize it to hold a priori. Thus nothing else has to be true of an object for it to be red than that it instantiates the right disposition, but it does not follow from this - and is not true - that the colour is the disposition. In other words, we combine an irreducibility thesis about the colour property itself, after the fashion of theory (2), with a dependency claim about what disposition that property necessarily covaries with. Accordingly, the colour essentially depends upon the relational disposition but it does not thereby inherit all the logical features of what it depends upon. And this allows us to say that the colour is itself a simple monadic property of surfaces, which is precisely how it appears, while its instantiation is wholly controlled by a complex disposition that has no perceptual counterpart (is not seen). Note that the disposition does not constitute the hidden nature of the colour, since it is not identical to it; rather, it provides an extrinsic condition that nevertheless completely fixes the distribution of colours across the world. It is indeed a conceptual truth that colours map on to dispositions to appear, but we need not infer that we have any collapse of the former into the latter. Insisting on this logical gap enables us to respect colour experience to the hilt while recognizing the important vein of truth in the dispositional theory. Here, as elsewhere, holding philosophical explication to the strict standard of reduction - with the identity relation prized as the only respectable relation to invoke - leads to implausibility in our account of the concept at issue. Beguiling to the philosophical imagination though identity is, it is not the only relation that can serve our philosophical purposes. We need to acknowledge a hierarchy of distinct properties here - from categorical physical basis in the object to perceiver-relative disposition to simple colour property - instead of just so many descriptions of a single property.

3 Consequences

Having canvassed the options and hit upon the most lissom one, let us satisfy ourselves that matters have worked out as we would wish. Does the non-reductive dispositional theory leave us where we are content to rest? For it does have one large metaphysical implication that some may find disturbing, namely that the world cannot be divided exhaustively into a mental part and a physical part. On the pure dispositional theory, colour properties turn out to be constructions from physical bases in the object and sensory responses in perceivers, so that a fundamental metaphysical dualism is observed. But if colour properties are strictly irreducible to such dispositions, then there are properties in the world that are neither mental nor physical nor a combination of both. These properties belong to a third great category, not reducible to the other two. True, colours depend upon properties from the other two categories, but that is by no means the same as belonging to them; they are, as it were, emergent with respect to mental and physical properties. Thus, if we are asked whether colours are inherently subjective on the theory in question, the only correct answer is that strictly speaking they are not; though they may be rated as subjective in the weak sense that their conditions of instantiation include subjective facts. Metaphysically, then, colours force a pluralism of properties into our ontology, exceeding the limits of the mental and the physical.

This result has consequences for a general materialism (see Materialism). It is commonly assumed, by those who favour a dispositional theory of colour, that colours would come within a materialist ontology if experiences themselves could be brought within that ontology. But that is not quite right if colours are construed according to the weaker dispositional theory, since even if experiences themselves were wholly physical it would not follow that the colours they fix are themselves physical, on account of the non-identity of colours with their corresponding dispositions. Materialism would fail even for material objects, even if it could be made to work for the mind.

Colour logic comes out satisfactorily under the weak dispositional theory, perhaps even more so than under the strong version. For the weak version allows us to hold on fully to the idea that we know, ahead of doing any science, what the intrinsic nature of each colour is, and hence how the various colours relate to each other in terms of exclusion, similarity, and the like. No knowledge of the underlying micro-properties is needed; nor indeed are

we required to translate colour talk into sensory disposition talk. What we know of colour logic is thus derived directly from our acquaintance with colours as they are presented to us in all their simplicity; there is no call to reformulate such necessities into other terms (though there is a sense in which they have their origin in experiential necessities).

See also: Colour, theories of; Primary-secondary distinction

COLIN McGINN

References and further reading

Hardin, C.L. (1988) *Colour for Philosophers*, Cambridge, MA: Hackett Publishing Company.(Relates the philosophical issues to the empirical facts about colour vision.)

Johnston, M. (1992) ‘How to Speak about the Colours’, *Philosophical Studies* 68: 221-63.(Fine recent discussion.)

Locke, J. (1690) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1975, esp. bk II, ch. viii.(Classic source for the distinction between primary and secondary qualities.)

McGinn, C. (1983) *The Subjective View*, Oxford: Oxford University Press.(Sets the discussion in the context of the relationship between the world as it is independently of the mind and the world as it is presented to the mind.)

McGinn, C. (1996) ’Another look at colour’, *The Journal of Philosophy* 93: 537-53.(Discusses the phenomenology of colour experience and the dispositional theory of colour.)

Smith, A.D. (1990) ‘Of Primary and Secondary Qualities’, *Philosophical Review* 99: 221-55.(Fine recent discussion.)

Second-order logic, philosophical issues in

Typically, a formal language has variables that range over a collection of objects, or domain of discourse. A language is 'second-order' if it has, in addition, variables that range over sets, functions, properties or relations on the domain of discourse. A language is third-order if it has variables ranging over sets of sets, or functions on relations, and so on. A language is higher-order if it is at least second-order.

Second-order languages enjoy a greater expressive power than first-order languages. For example, a set S of sentences is said to be categorical if any two models satisfying S are isomorphic, that is, have the same structure. There are second-order, categorical characterizations of important mathematical structures, including the natural numbers, the real numbers and Euclidean space. It is a consequence of the Löwenheim-Skolem theorems that there is no first-order categorical characterization of any infinite structure. There are also a number of central mathematical notions, such as finitude, countability, minimal closure and well-foundedness, which can be characterized with formulas of second-order languages, but cannot be characterized in first-order languages.

Some philosophers argue that second-order logic is not logic. Properties and relations are too obscure for rigorous foundational study, while sets and functions are in the purview of mathematics, not logic; logic should not have an ontology of its own. Other writers disqualify second-order logic because its consequence relation is not effective - there is no recursively enumerable, sound and complete deductive system for second-order logic.

The deeper issues underlying the dispute concern the goals and purposes of logical theory. If a logic is to be a calculus, an effective canon of inference, then second-order logic is beyond the pale. If, on the other hand, one aims to codify a standard to which correct reasoning must adhere, and to characterize the descriptive and communicative abilities of informal mathematical practice, then perhaps there is room for second-order logic.

1 First-order and second-order logic compared

The downward Löwenheim-Skolem theorem states that if *S* is a set of first-order sentences that has an infinite model, then *S* has a model whose domain is at most countably infinite (or the cardinality of *S*, whichever is larger). The upward Löwenheim-Skolem theorem states that if *S* is a set of first-order sentences such that, for each natural number *n*, *S* has a model whose domain has at least *n* elements, then for every infinite cardinal κ, *S* has a model whose domain has cardinality at least κ. It follows that no set of first-order sentences that has an infinite model is categorical. (A set of sentences is said to be categorical if any two models satisfying it are isomorphic.)

There is a complete, sound and effective deductive system *D* for first-order logic (see Predicate calculus). That is, if *S* is a set of sentences and ϕ is a single sentence, then ϕ is deducible from *S* in *D* if and only if ϕ holds in every structure that satisfies every member of *S*. Finally, first-order logic is compact, in that for every set *S* of first-order sentences, if every finite subset of *S* has a model, then *S* itself has a model.

(In standard semantics for second-order languages, the set or predicate variables range over the entire power set of the domain. It follows that the domain of discourse determines the range of the other variables. There is an alternative semantics, due to Leon Henkin (1950), in which this does not hold. Each model consists of not only a domain of discourse, but also separate ranges for the predicate, relation and function variables. With Henkin semantics, second-order logic is complete and compact, and the Löwenheim-Skolem theorems hold. In effect, with Henkin semantics, second-order languages are treated as multi-sorted first-order languages. The foregoing discussion presupposes standard semantics.)

Second-order logic has none of the above properties. The upward Löwenheim-Skolem theorem fails because there is a categorical characterization *AR* of the natural numbers. Every model of *AR* is countably infinite. The downward Löwenheim-Skolem theorem fails because there is a second-order categorical characterization *AN* of the real numbers. Every model of *AN* has the cardinality of the continuum, and so is uncountable (see Cantor's theorem; Continuum hypothesis). A corollary of these features, and Gödel's incompleteness theorem, is that second-order logic is not complete (see Gödel's theorems §1). There is no effective, sound and complete deductive system for second-order logic. Likewise, second-order logic is not compact.

The sentence *AR* characterizing the natural numbers is a straightforward formalization of common informal

axiomatizations of arithmetic. The ‘second-order’ item is the induction principle, stating that if a property (set) *P* of numbers holds for (contains) zero and is closed under the successor function, then *P* holds for (contains) all natural numbers. In symbols,

$$\forall P[(P0 \mathbin{\&} \forall x(Px \rightarrow Psx)) \rightarrow \forall xPx].$$

In first-order arithmetic, this principle is replaced by a scheme. Every instance of

$$[\phi(0) \mathbin{\&} \forall x(\phi(x) \rightarrow \phi(sx))] \rightarrow \forall x\phi(x)$$

in which ϕ is a formula in the language of first-order arithmetic, is an axiom. Thus, first-order arithmetic has infinitely many axioms.

In the characterization *AN* of real analysis, the second-order item is the principle of completeness, stating that every non-empty set of real numbers with an upper bound has a least upper bound:

$$\begin{aligned}&(\forall P)\{(\exists xPx \mathbin{\&} \exists x\forall y(Py \rightarrow y \leq x))\\ &\qquad \rightarrow \exists x[\forall y(Py \rightarrow y \leq x) \mathbin{\&}\\ &\qquad\qquad \forall z(\forall y(Py \rightarrow y \leq z) \rightarrow x \leq z)]\}\end{aligned}$$

First-order real analysis is obtained by replacing the completeness axiom with the completeness scheme,

$$\begin{aligned}&(\exists x\phi(x) \mathbin{\&} \exists x\forall y(\phi(y) \rightarrow y \leq x))\\ &\quad \rightarrow \exists x[\forall y(\phi(y) \rightarrow y \leq x) \mathbin{\&}\\ &\qquad\quad \forall z(\forall y(\phi(y) \rightarrow y \leq z) \rightarrow x \leq z)]\end{aligned}$$

one instance for each formula ϕ that contains neither *x* nor *z* free.

The difference between, say, second-order real analysis and its first-order counterpart is that in the former one cannot directly state that every non-empty bounded set has a least upper bound. Instead, there is a separate principle for each such set which is definable by a formula in the language of first-order analysis. Since there are sets which are not definable, the first-order theory has models which are not isomorphic to the real numbers. Such structures are called ‘nonstandard models’ (see Löwenheim-Skolem theorems and nonstandard models).

One common characterization of finitude, due to Richard Dedekind, is that a set is finite if there is no one-to-one function from the set to a proper subset of itself. This can be rendered as a second-order sentence with no non-logical terminology:

$$\begin{aligned}&\mathrm{FIN}(X)\text{:}\ \sim\exists f[\forall x(Xx \rightarrow Xfx) \mathbin{\&}\\ &\qquad \forall x\forall y((Xx \mathbin{\&} Xy \mathbin{\&} fx = fy) \rightarrow x = y)\\ &\qquad\quad \mathbin{\&} \exists z(Xz \mathbin{\&} \forall x(Xx \rightarrow fx \neq z))]\end{aligned}$$

It follows from the compactness property that there is no first-order characterization of finitude.

2 Defence

It is widely held, often implicitly, that mathematicians succeed in describing various structures, up to isomorphism, and in communicating information about them. When various mathematicians refer to ‘the natural numbers’, ‘the real numbers’, ‘Euclidean space’ and so on, they are talking about the same structures. Similarly, there is little doubt that such mathematical notions as finitude and well-foundedness are clear, definite and unequivocal. In short, it is agreed that the *informal* language of mathematics is sufficient for ordinary description and communication of the basic structures and notions.

One purpose of logic, model theory in particular, is to ‘model’ the semantics of the informal languages of mathematics. Thus, one desideratum is that logic should somehow register the successful description and communication of mathematical structures and notions. As we have seen, first-order logic falls short of this. As Barwise and Feferman put it, ‘As logicians, we do our subject a disservice by convincing others that logic is first-order and then convincing them that almost none of the concepts of modern mathematics can really be captured in first-order logic’ (1985: 5).

A further claim is that it is unnatural to rely on infinitely many axioms, when a single second-order axiom is available. For example, if someone is asked why they believe that the second-order completeness axiom holds for the real numbers, the reply might be that the principle characterizes the numbers, perhaps citing the categoricity result. However, if asked why they believe every instance of the first-order completeness scheme, the answer is not that simple. One cannot claim that the scheme characterizes the real numbers since, as we have seen, the scheme does not characterize the numbers. Georg Kreisel suggests that the reason that mathematicians believe the instances of the various schemes is that they are consequences of the respective second-order axiom (1967). Indeed, in common deductive systems for second-order logic, each instance of each scheme can be deduced from the corresponding axiom. An opponent might retort that one can see from the scheme itself that all of its instances are true. This might be called a 'metalinguistic' move, since it invokes the scheme itself at the level of justification. Notice that this manoeuvre involves a generalization, since it refers to *all* instances of the scheme. It remains to be seen whether this generalization is any more problematic than the generalization over predicates, properties or sets. Notice also that the scheme itself is an infinite structure, and cannot be characterized in a first-order metalanguage.

Another problem with the schemes is that each one is tied to the ingredients of the particular language in use at the time. For example, if the induction scheme is formulated in a language whose only non-logical term is the successor symbol, then a very weak, decidable theory results. To obtain a richer system, closer to classical arithmetic, symbols for addition and multiplication must be added, and the induction scheme must itself be expanded to include instances with the new terminology. On the other hand, the second-order induction axiom, which contains only the successor symbol, is sufficient to characterize the natural numbers. Moreover, addition and multiplication can be defined in second-order arithmetic, as can any recursive function. Beyond that, new terminology can be added as needed, without having to modify the very definition of the natural numbers. This fits the intuition that one cannot tell in advance what resources, and thus what terminology, are needed to shed light on the natural numbers, even after the structure has been characterized.

In short, 'the natural numbers' has an informal interpretation that outruns what is captured in first-order arithmetic, and this informal interpretation plays a central role in actual mathematical practice. Alonzo Church agrees:

> our definition of the consequences of a system of postulates…can be seen to be not essentially different from [that] required for the…treatment of classical mathematics…. It is true that the non-effective notion of consequence, as we have introduced it…presupposes a certain absolute notion of ALL propositional functions of individuals. But this is presupposed also in classical mathematics, especially classical analysis
>
> (1956: 326)

3 Prosecution

A number of authors have either rejected second-order logic outright, or else claimed that it is not logic, but rather an obscure form of mathematics. One broad attack is aimed at 'traditional' versions of second-order and higher-order systems that have variables ranging over intensional items such as properties, attributes and propositional functions (see Intensional entities). There is no consensus on which properties, attributes and so on exist, and there is no consensus on when two properties are identical or distinct. These sticky metaphysical matters should not interfere with serious work in the foundations of mathematics, the argument goes. Mathematics itself is reasonably clear, and we should not invoke obscurity in order to develop foundations for it.

This argument does not affect contemporary systems. The items in the range of the second-order variables in standard semantics are extensional - they are sets or classes. For any collection *C* of elements of the domain of discourse, there is an item in the range of the 'property' variables whose elements are the members of *C*. The development of modern set theory shows that the extensional notion of class is at least relatively clear (see Set theory).

The problem now, however, is that by invoking set-theoretic notions in the model theory, we may have crossed a border from logic into mathematics. W.V. Quine calls second-order logic 'set theory in disguise' and 'set theory in sheep's clothing'. The metaphor suggests that, unlike grizzly mathematics, logic is, or ought to be, a mere 'sheep'.

Quine and others have brought a number of arguments for this judgment. One involves ontological commitment

(see Ontological commitment). A first-order theory, with an intended interpretation, has variables ranging over a given domain of discourse. A second-order theory, with standard semantics, has variables ranging over the entire power set of the domain, a larger collection. Moreover, the usual metatheory for second-order languages, where we turn for information about the power set operation, uses Zermelo-Fraenkel set theory (see Set theory, different systems of §3), which does have a vast ontology. As Quine puts it, 'Set theory's staggering existential assumptions are…hidden…in the tacit shift from schematic predicate letter to quantifiable set variable' ([1970] 1986: 68).

Because of the expressive resources of second-order languages, including the terminology for sets, there are a number of rather substantial mathematical statements that can be formulated in second-order languages. There is, for example, a sentence that is a logical truth if and only if the generalized continuum hypothesis is true, and there is a sentence that is a logical truth if and only if the axiom of choice holds. But it is on the face of it counterintuitive to hold that these statements are logically true or logically false.

It follows from Gödel's completeness theorem that the consequence relation of first-order logic is effective. The set of first-order logical truths, for example, is recursively enumerable. On the other hand, the set of second-order logical truths is not. The set is not even definable in (second-order) arithmetic, although it is definable in set theory. Again, we seem to be beyond the purview of logic (see Non-constructive rules of inference).

In summary, many of those who hold that second-order logic is not logic seem to agree that the *informal* language of mathematics is somehow adequate to describe and communicate the various structures and notions, but they add that logic must fail where informal mathematics succeeds. There is another line of argument, less popular today, that rejects the presupposition that the mathematical structures and notions in question are unequivocal. Accordingly, the Löwenheim-Skolem theorems indicate that there is no unambiguous notion of 'finite', 'countable', 'natural number' and so on. All of these notions are relative to a background (first-order) set theory. This sceptical view is sometimes called 'Skolemite relativism'. Advocates of it adopt first-order logic because they claim that its model theory accurately reflects the ontological/epistemic/semantic situation. There is nothing unequivocal to capture or describe with either formal or informal languages of mathematics.

4 Settlement?

By itself, a border dispute concerning mathematics may not be all that interesting. The deeper matters here concern the role and purposes of logical theory. Why do we develop logical systems in the first place? Historically, a number of goals have been articulated. One of these is to present a calculus, a deductive system to codify or serve as a canon of correct reasoning. In light of incompleteness, this goal is not served by the consequence relation of second-order logic (as long as a system must be effective, or recursive, in order to qualify as a calculus; see Church's thesis). There are, to be sure, good deductive systems for second-order languages, but they are all incomplete for standard semantics.

On the other hand, it is widely held that deductive systems must themselves adhere to a prior notion of logical consequence. This notion is often understood in modal terms. Near the beginning of the *Prior Analytics*, Aristotle defines a 'syllogism' to be 'a discourse in which, certain things having been supposed, something different from the things supposed results *of necessity*' (see Logic, ancient §§1-3). In advance of reflection, there is no reason to expect that the pre-theoretic, modal notion of logical consequence is effective, and that there is a complete deductive system for it. Advocates of second-order logic claim that the second-order consequence relation captures an underlying notion of logical consequence, one to which deductive systems must answer.

There is a trend in philosophy to understand at least some modality in terms of language. In particular, it is held by many that necessities are grounded in the meanings of terms. In this spirit, logical consequence is sometimes defined in terms of the meanings of a certain collection of terms, the so-called 'logical terminology'. This is consonant with the slogan that logical consequence is a matter of 'form'. The logical terminology of a discourse is what marks its form, while the non-logical items determine content. From this perspective, the issue of second-order logic is whether the membership (or predication) relation and bound variables ranging over classes are logical. This, of course, is another border dispute. It should be settled by examining the various systems, to determine which are the most useful and insightful accounts of logical consequence. Perhaps one can be eclectic (see Logical constants; Logical form).

Second-order, and even higher-order, languages have found applications outside of logic and the foundations of mathematics. Two noteworthy examples are Montague grammar, for natural language (1974; see Montague, R.M.), and Aldo Bressan's higher-order modal logic for the language of science (1972). Although, as noted, many logicians have come to see first-order logic as too weak, it is still widely held that the consequence relation of second-order logic is too intractable for useful mathematical study. A growing research programme in mathematical logic is to develop and study alternatives which are, in a sense, intermediate between first-order and second-order.

See also: Consequence, conceptions of; Logical and mathematical terms, glossary of; Second- and higher-order logics; Theory of types

STEWART SHAPIRO

References and further reading

Barwise, J. and Feferman, S. (eds) (1985) *Model-Theoretic Logics*, New York: Springer.(Extensive treatment of many logics, some of which are intermediate between first-order and second-order. Quite technical, for the most part. Extensive bibliography.)

Bressan, A. (1972) *A General Interpreted Modal Calculus*, New Haven, CT: Yale University Press.(An extensive, higher-order modal logic, put to a number of philosophical uses.)

Church, A. (1956) *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press.(Still an excellent introduction to both first-order and second-order logic.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel*, Cambridge, MA: Harvard University Press.(Collection of many original papers in logic, translated into English. Includes papers by Frege, Peano, Dedekind, Cantor, Hilbert, Russell, Zermelo, Löwenheim, Skolem, Weyl, Ackermann, von Neumann, Bernays and Gödel.)

Henkin, L. (1950) 'Completeness in the Theory of Types', *Journal of Symbolic Logic* 15: 81-91.(Henkin's alternative semantics, which treats second-order languages as multi-sorted first-order languages, with separate ranges for the predicate, relation and function variables.)

Hilbert, D. and Ackermann, W. (1928) *Grundzüge der theoritischen Logik*, Berlin: Springer; trans. L. Hammond, G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, New York: Chelsea, 1950. (Historical source; quite readable.)

Kreisel, G. (1967) 'Informal Rigour and Completeness Proofs', in I. Lakatos (ed.) *Problems in the Philosophy of Mathematics*, Amsterdam: North Holland, 138-86. (Among other things, a multi-faceted philosophical defence of the use of higher-order languages in foundational studies. Moderate difficulty.)

Montague, R. (1974) *Formal Philosophy*, ed. R. Thomason, New Haven, CT: Yale University Press.(Higher-order logic, put to linguistic use. The start of an extensive research programme.)

Quine, W.V. (1970) *Philosophy of Logic*, Englewood Cliffs, NJ: Prentice Hall; 2nd edn, Cambridge, MA: Harvard University Press, 1986, esp. ch. 5.(Argument against second-order logic. Lucid.)

Shapiro, S. (1991) *Foundations Without Foundationalism: A Case for Second-Order Logic*, Oxford: Oxford University Press.(Extensive development and justification for higher-order logic. Philosophical, mathematical and historical treatment. Most of the book requires some background in logic. Extensive bibliography.)

Selden, John (1584-1654)

Antiquarian, philologist, parliamentarian, legal historian and practising lawyer, John Selden was a major figure in the renaissance and systematization of common law. In jurisprudence, his importance lies in his attempt to develop certain elements of an epistemology of common law. He made use of history to criticize current legal doctrines, and developed a philosophical methodology in relation to the interpretation of precedent.

Educated at Oxford, Selden joined Clifford's Inn (1602) and Inner Temple (1604) and was called to the Bar in 1612. A friend of the historian William Camden and the antiquary Sir Robert Cotton, his historical writings were in large measure responses to the Reformation and to the vernacularization of English law.

Selden's work was in essence a nascent form of common law hermeneutics, a critical interpretation of legal texts and institutions. His use of philology to reconstruct the development of common law indicated not only that it was a member of the European tradition of law but also that, conceived as a tradition of judgment, English law was far more complex and diverse a cultural form than the insular profession of his day would allow. In his extra-legal writings, most notably in his posthumously published *Table Talk*, Selden did much to indicate the cultural significance of legal institutions.

In common with the historical and classicist tendencies of the Elizabethan era, Selden was concerned primarily to collect for posterity what he termed the neglected and fragmentary evidence of antique common law. In an early and in many respects exemplary work on the ancient constitution, *Jani Anglorum facies altera (The Other Face of the English Janus)* (1610a), he thus traced the inhabitants and laws of England through an imaginative and remarkably wide-ranging study of sources as diverse as mythology, literature, etymology and statuary inscriptions. The work explicitly placed the common law under the ambivalent sign of Janus and indicated in detail the diversity of nationalities and of languages that had played a part in the development of common law. The constitution was in his view plural and in many respects uncertain. The histories of English law which Selden subsequently wrote concentrated upon detailing the foreign influences and jurisdictions which subsisted within the common law. In his history of *The Duello or Single Combat* (1610b), he showed the importance of French law to English custom. In his *Titles of Honour* (1614) and in his editions of the early treatises *Fleta* and *De Laudibus Legum Angliae (In Praise of the Laws of England)* he indicated the pervasiveness of Roman influence. In *Mare Clausum (The Enclosed Sea)* (1636), his polemical tract directed against the famous treatise by Hugo Grotius on the freedom of the seas, Selden urged the use of natural law, while in *De Diis Syris (Of Syrian Gods)* (1617b) and in the *History of Tythes* (1618), his last major work, he argued for the importance of divine law (*ius divinum*). In each case the importance of historical research lay not simply in rectifying the neglect of evidence and precedent but also and more remarkably in providing a tool for the criticism of England's 'dulling custom' and of some of its substantive faults. He argued thus in *Jani Anglorum* for the recognition of the right of women to inherit public office, while elsewhere he urged reform of the law of tithes and more broadly supported an expansion of the historical method and sources of common law judgment.

It was Selden's lifelong belief that what he coined, in a work on *Titles of Honour*, as 'the Lady Common Law' could not lie alone but was rather to be understood as a member of a family of legal traditions whose ancestry could be traced across Europe and whose reason was both divine and human.

See also: Common law; Law, philosophy of; Legal discourse; Legal hermeneutics

PETER GOODRICH

List of works

Selden, J. (1610a) *Jani Anglorum facies altera (The Other Face of the English Janus)*, trans. R. Westcot, London: T. Bassett, 1683.(Study of the other face or 'backface' of English law, concerned particularly with fragments and evidence of female rule and female lawgivers.)

Selden, J. (1610b) *The Duello or Single Combat*, London: I. Helme.(History of the law of duels, with particular attention to the comparative and civilian sources from which English practice developed.)

Selden, J. (1614) *Titles of Honour*, London: W. Stansby.(Extended analysis of the civilian *ius imaginum*, the law of nobility and lineage, as applied within the common law.)

Selden, J. (1617a) *Table Talk*, London: E. Smith, 1689.(Selden's most popular work, an aphoristic collection of moral conversations concerned with Christian manners.)
Selden, J. (1617b) *De Diis Syris (Of Syrian Gods)*, Lipsiae: L.S. Corneri, 1672.(Treatise in two volumes addressing the history of pagan gods and the laws against idolatry.)
Selden, J. (1618) *History of Tythes*, BL 517.b.4, London: Private Circulation.(History of the canon, civil and common law relating to tithes, the tenth part of the annual increase of the profits of land, originally due to the Church. Published privately, available from the British Library.)
Selden, J. (1636) *Mare Clausum (The Enclosed Sea)*, London: published by Royal Command, repr. and trans. M. Needham, 1652; a further edn, 1663.(Polemical tract attacking the international lawyer Grotius, and the argument that the seas be kept free.)

References and further reading

Christianson, P. (1985) 'John Selden, the Five Knights Case, and Discretionary Imprisonment', *Criminal Justice History* 6: 65-87.(Interesting account of Selden in his capacity as lawyer and politician at a dangerous time.)
Kelley, D.R. (1990) 'English Developments: The Common Law', in *The Human Measure: Social Thought in the Western Legal Tradition*, Cambridge MA: Harvard University Press.(This contains a discussion of Selden's part in giving relatively early definition to common law and the character of common law thinking.)
Ogg, D. (1925) 'Introduction', in *Ioannis Seldeni Ad Fletam Dissertatio*, Cambridge: Cambridge University Press. (This gives a very helpful scholarly account of Selden's work as a legal historian and antiquary.)
Tuck, R. (1979) *Natural Rights Theories*, Cambridge: Cambridge University Press. (Selden's work is placed in its context in a historical discussion of ideas of natural law and natural right.)

Self, Indian theories of

Hindu thought traces its different conceptions of the self to the earliest extant Vedic sources composed in the Sanskrit language. The words commonly used in Hindu thought and religion for the self are jīva *(life),* ātman *(breath),* jīvātman *(life-breath),* puruṣa *(the essence that lies in the body), and* kṣetrajña *(one who knows the body). Each of these words was the culmination of a process of inquiry with the purpose of discovering the ultimate nature of the self. By the end of the ancient period, the personal self was regarded as something eternal which becomes connected to a body in order to exhaust the good and bad karma it has accumulated in its many lives. This self was supposed to be able to regain its purity by following different spiritual paths by means of which it can escape from the circle of births and deaths forever.*

There is one more important development in the ancient and classical period. The conception of Brahman as both immanent and transcendent led to Brahman being identified with the personal self. The habit of thought that tried to relate every aspect of the individual with its counterpart in the universe (Ṛg Veda X. 16) had already prepared the background for this identification process. When the ultimate principle in the subjective and objective spheres had arrived at their respective ends in the discovery of the ātman *and Brahman, it was easy to equate the two as being the same spiritual 'energy' that informs both the outer world and the inner self. This equation had important implications for later philosophical growth.*

The above conceptions of the self-identity question find expression in the six systems of Hindu thought. These are known as āstikadarśanas *or ways of seeing the self without rejecting the authority of the Vedas. Often, one system or the other may not explicitly state their allegiance to the Vedas, but unlike Buddhism or Jainism, they did not openly repudiate Vedic authority. Thus they were* āstikadarśanas *as opposed to the others who were* nāstikadarśanas*. The word* darśana *for philosophy is also significant if one realizes that philosophy does not end with only an intellectual knowing of one's self-identity but also culminates in realizing it and truly becoming it.*

1 Vedic literature

Hindu thought traces most of its philosophical ideas to the time of the Vedas (*c*.2000-1000 BC) composed in Sanskrit. While the corpus called Vedic literature cannot be accurately classified, it is by convention considered to comprise the four Vedas (*Ṛg Veda*, *Atharva Veda*, *Yajur Veda*, *Sāma Veda*) along with the *Brāhmaṇas* (liturgical works), *Āraṇyakas* ('books composed in the forest') and Upaniṣads (philosophical literature) attached to each of the Vedas. The *Ṛg Veda*, which is uniformly considered to be the most ancient contains statements that formed, in some way, the basis of important notions of the self that were developed in the philosophical schools later (*c*.200-600 AD). By the time of the early Upaniṣads (*c*.800-400 BC) the different words used for the self were entrenched in Hindu thought. All the different genres of Vedic literature collectively enjoy a sacrality which is denoted by designating them as 'not the work of humans' (*apauruṣeya*) and by the fact that they were never written down at their time of origin. They were 'heard' (*śruti*) and are considered to be the result of the intuitive insight of the Vedic wise sages (*ṛṣis*); they were handed down orally from generation to generation and many methods were devised to preserve the accuracy of the texts.

There are many terms used to denote the individual self in Hindu thought; some of them are *jīva* (from the root *jīv* - to live, be alive), *ātman* (from the root *an* - to breathe), *jīvātman* (combining the first two), *puruṣa* (from the root *puri śete* - lying in the body, embodied), *kṣetrajña* (from *kṣetram jānāti* - one who knows the body, embodied self). Thus, it can be seen that the self in all these meanings differs from the usual notion of ego: what a person is, their identity, etc., with which one associates the self in ordinary understanding. While we cannot believe that Vedic literature, as it exists today, contains early thought pertaining to the question of personal identity, what we do have is fairly representative of the directions this inquiry took. More importantly, these meanings persisted and were standardized in the course of time, so that in the later schools of philosophy (*c*.200-600 AD), as well as in popular Hinduism, these were accepted as words that convey the idea of the self.

The inquiry in the Upaniṣads was mainly speculative. It had the dual purpose of finding the ultimate principle that was the basis of the universe and the true nature of the individual self. The first inquiry led to the postulation of Brahman (from root *bṛh* - to expand, burst forth) as the source behind the universe. This was the culmination of the trend set in hymns like *Ṛg Veda* 1.164.46, 10.81, 10.90, 10.129, where different ideas are expressed regarding the

ultimate source of the universe. This concept finally settled down with the notion of Brahman being described both as immanent in the cosmos (Chāndogya Upaniṣad 3.14.1, 7.25.2) and as transcending the same (Bṛhadāraṇyaka Upaniṣad 3.8).

The other inquiry was directed towards discovering the identity of the self by analysing the inner nature of the human being. Having arrived at the conclusion that the individual self cannot be identified with any one of the senses (both motor and cognitive), through a process of elimination (Kena Upaniṣad 3.4) the mind as being the self was also ruled out as it was obvious that all mental states had something other than the mind as their referent which was also constant. While this form of inquiry was a process from the gross to the subtle elements, another feature was the observation that the 'sense-of-I' (personal identity) was invariably associated with consciousness (self-awareness) and knowledge. This, combined with the persistence of the 'sense-of-I' in all the three states of waking, dreaming and deep sleep led to the postulation of the *ātman* being the inner self, which is the ultimate principle (Bṛhadāraṇyaka Upaniṣad 4.3 and 4).

There were other trends which were growing in the same period which would contribute to the enrichment of the concept of *ātman*. By the time of the Upaniṣads the concepts of *dharma* (from root *dhṛ* - to sustain, support, behave in a human way) and that of karma (from root *kṛ* - to act (ritually)) had evolved into metaphysical principles. The goals of life called *puruṣārthas* had also been formulated by this time as the pursuit of *dharma*, *artha* (economic goal constituting *dharma*), *kāma* (sensual fulfilment within *dharma*) and finally *mokṣa* (liberation from bondage, or to a state of one's true nature).

All these ideas working on the self meant that good karma will result in a good life on earth, while its opposite will lead to a bad life. Since one single life could not adequately explain the inequalities present in life, the same self came to be considered eternal, taking birth repeatedly after the fall of the body (death) until such time as it realizes the futility of this worldly existence (see Karma and rebirth, Indian conceptions of). The way to realize that was to overcome the ignorance of identifying the self with the body through experiencing the true nature of the self, either by following the path of knowledge, of disinterested action (action for its own sake without thinking of the result or reward), or the path of devotion. In the course of time these came to be the famous threefold paths of *jñāna mārga* (knowledge path), *karma mārga* (action path) and *bhakti mārga* (devotion path). Yoga also came to be recognized as a path by the time of the *Bhagavad Gītā* (*c*.200 BC). The state of liberation was called variously *mokṣa*, *apavarga*, *nirvāṇa*, *kaivalya*, *sāyujya* and so on.

2 The Upaniṣadic view

By pursuing a methodology that mainly combined perception, inference, experience and speculation the Upaniṣads reach a startling and bold conclusion. This was the identification of the ultimate truth underlying the universe called Brahman with the inner self of all beings called the *ātman*. It was the same *ātman* that was present in all living beings, the differences seen being attributed to the nature of the respective body it occupies (see Brahman). The journey that ended in this equation can be described in a twofold manner: one that approached the question in an objective, external direction; the other which turned its attention to the inner nature of oneself, or adopted a subjective approach.

The objective search went through the exercise of finding the ultimate principle that underlies the universe (*Ṛg Veda* 10.31.7). But significant to the *ātman-Brahman* equation is the tendency to find a correspondence between the individual and the universe in a 'micro-macro' paradigm that manifests itself very early in the period, for example, in the *Ṛg Veda* 10.90, 10.16, *Atharva Veda* 11.4, Aitareya Upaniṣad 1, Chāndogya Upaniṣad 3.16.1, 5.11-18. This came to be known as the affinity (*bandhutā*) theory. Side by side there was the tendency to assign the world an inferior status compared to the transcendent nature of Brahman (*Ṛg Veda* 10.129).

The subjective approach, on the other hand, analysed the inner person in a twofold manner. One was the approach from the gross to the subtle and the other was from the point of view of knowledge and the principle of consciousness. The first approach led to the elimination of all motor senses and the senses of knowledge as being the self (Kena Upaniṣad 3, 4). Even the mind was ruled out as the enduring self for it was obvious that all mental states referred to something other than themselves as a basis for those states and also this referent was a constant factor underlying all mental states, as in the Kaṭha Upaniṣad 2.2.1, 3.5, 3.10. and the Kena Upaniṣad 1.4.8. It was also observed that the self was always associated with any event of cognition.

Another interesting discovery was the fact of the persistence of the ‘sense-of-I’ in the three states of waking, dreaming and deep sleep. The deep sleep identity was explained by the fact that the person who gets up from such a slumber recognizes their own identity as the one who had that experience of sleep. As these three states were still within the realm of worldly existence, *ātman* was viewed as transcending these three states and reaching its pure form in a fourth stage which is identified with the state of freedom (Chāndogya Upaniṣad 8.6, Māṇḍūkya Upaniṣad 1).

The affinity phenomenon, combined with the immanent and transcendent nature of both *ātman* and Brahman, made it easy to equate the one with the other. Thus by the time of the Upaniṣads the *ātman* was established as the self in any living being, it was of the nature of pure consciousness, it was identified with Brahman very often denoted by the word *paramātman* (highest self) and was of the nature of both immanence and transcendence (Bṛhadāraṇyaka Upaniṣad 2.4.14, Taittirīya Upaniṣad 2.1-5). As for the question of whether this *ātman* can be known like any other event in knowledge, the Upaniṣads seem to suggest only an intuitive realization of the self which is different from knowledge understood in everyday life (Kena Upaniṣad 2.2, Kaṭha Upaniṣad 2.23).

3 *Dharma*, karma, human goals and transmigration

There were many other ideas which reached maturity during the time of the Upaniṣads and which led to the theory of the permanent nature of the *ātman*. The Brahman- *ātman* equation and their status as the final truth in an interchangeable manner posed a problem for the philosopher; if something is existent in an absolute sense it cannot be negated at any time or in any sense. Thus, in order to preserve the immutable nature of *ātman* it was conceived as being a permanent entity, existing even after the fall of the body (Māṇḍūkya Upaniṣad 4.40, Kaṭha Upaniṣad 2.18, *Bhagavad Gītā* 2.16).

The concepts of *dharma* and karma had evolved into metaphysical principles that held sway over the lives of individual selves on this earth. It is not possible to trace all the nuances of meaning that *dharma* went through before it settled down to mean virtue, merit, good conduct and also to be raised to the status of the first of the goals of human beings. These goals were four in number, called *dharma*, *artha* (pursuit of economic ends within *dharma*), *kāma* (legitimate sensual pleasure within *dharma*) and *mokṣa* (liberation from repeated births to deaths).

Karma meanwhile had also grown into a fully fledged theory out of the necessity to explain inequalities seen in the world. It was a theory of ethics relating one’s life on earth to the kind of deeds one had done in past lives (Kaṭha Upaniṣad 2.2.7). In this way, since the inexplicable details of a single life could not be explained, the self came to be accepted as a permanent entity that transmigrates from one birth to another until it realizes its true nature. This realization could come through knowledge, disinterested action or devotion. Later Yoga also came to be recognized as a means to the final goal.

We now get a concept of the self which is eternal, connected with a body due to its own good or bad deeds, capable of realizing its true nature by following the spiritual path best suited to itself, which finally escapes the circle of births and deaths attaining *mokṣa*. The nature of *mokṣa* that the self attains will depend on the kind of ontological status assigned to the final truth in each philosophical system. Some retain a theistic, cosmic view of the ultimate while others favour a monistic, acosmic perception. While the goal of *mokṣa* may be defined variously depending on the nature of Brahman, the self is viewed in the same way in practically all schools of Hindu thought.

4 Six schools of Hindu thought

The six schools of Hindu thought are collectively called the *āstikadarśanas*. *Āstika* stands for that which does not reject the authority of Vedic literature and *darśana* literally means vision. Thus, *darśana* is the equivalent of philosophy in the Hindu context. It also indicates that philosophy does not end with just an intellectual understanding of what the ‘self’ is, but it culminates in ‘seeing it’ or ‘becoming it’.

The six *darśanas* are Nyāya, Vaiśeṣika, Sāōkhya, Yoga, Mīmāṃsāand Vedānta. While each of these systems had an independent beginning, some came to be designated jointly as twin systems because of certain affinities in their metaphysical theories. Thus, after the grouping, the six systems are referred to as Nyāya-Vaiśeṣika, Sāōkhya-Yoga and Mīmāṃsā-Vedānta in the philosophical treatment of the schools. Each of these developed its own extensive theories regarding both ontological and epistemological categories, but all were in agreement regarding the eternal

nature of the individual self and in its ultimate liberation from births and deaths, however they may differ in detail. These schools can also be viewed as representing different philosophical approaches to the understanding of the ultimate categories. They range from the most realistic (Nyāya-Vaiśeṣika) to the most idealistic (Advaita Vedānta) in the way they approach the theory of the *ātman*.

5 Nyāya-Vaiśeṣika

As the name signifies Nyāya (reasoning) lays emphasis on the method of critical reasoning. Vaiśeṣika, on the other hand, is derived from the word for particularity (*viśeṣa*) which is one of its ultimate categories. It conveys a sense of differentiation even between things which cannot be otherwise differentiated. Among the six schools Nyāya-Vaiśeṣika is the most realistic as it believes in the independent existence of objects in the world. All objects of experience are divided into seven classes in the Vaiśeṣika: one of them is known as substance (*dravya*). We will pay attention only to the substances, of which there are nine, one of which is *ātman*; the other eight are earth, water, light, air, ether, space, time and the mind. The Nyāya has a scheme of sixteen ultimate philosophical topics and the seven of Vaiśeṣika come under its second topic called knowable things (*prameya*) (see Ontology in Indian philosophy).

Since the treatment of *ātman* is similar in both Nyāya and Vaiśeṣika, henceforth it will be treated as a single school for purposes of the *ātman* problem. The *ātman* is unique, eternal, all-pervading and there are as many selves as there are bodies associated with them. Consciousness is not intrinsic to *ātman*, but it is only one of the many qualities that the self has. It is not only an essential quality but also an adventitious attribute along with desire, volition, pleasure, pain, love, aversion, virtue and vice.

The system is in agreement with the general theory of the bondage of the *ātman* in the round of births and deaths. The reason for this state is lack of true knowledge of the self as distinct from all other things, which is the reason for pain. When correct knowledge dawns through study of the sacred books and by the practice of Yoga, there is the destruction of ignorance and the achievement of the goal of liberation. The fact that the self does not also have the attribute of consciousness in liberation means that the self is not conscious of its own freedom. Although a Supreme self known as the Lord (*Īśvara*) is admitted in the school, it is restricted to being the efficient cause of the world and does not extend to the sphere of liberation of the individual self. Nyāya and Vaiśeṣika do not believe in liberation while still in an embodied state as some of the other systems do (see Nyāya-Vaiśeṣika).

6 Sāōkhya-Yoga

The school of Sāōkhya-Yoga recognizes two ultimate ontological categories called *prakṛti* (material nature) and *puruṣa* (personal self). The selves are also many as in Nyāya and Vaiśeṣika. While Sāōkhya and Yoga agree in this classification, Yoga also believes in God, who is described as a unique person different from both material nature and the personal selves in bondage; God has never been subject to the bondage of births and deaths (see God, Indian conceptions of §4). The evolution of the universe takes place from material nature, which is composed of three material qualities. Material nature is perpetually in motion and replicates itself when there is an equilibrium of the constituent qualities. When this balance is upset evolution is set in motion and leads to the unfolding of the different principles. In Sāōkhya the basis for the evolution to start is the presence of the *puruṣa*: it is not clear whether one self or many selves are intended here, but some kind of connection with karma seems to be suggested. It is this difficulty that Yoga tries to overcome by postulating a God who effects the association of material nature with the unseen karma of selves and starts the process of evolution. However, the concept of a Supreme self over and above the other selves seems more a practical necessity and does not serve any philosophical purpose.

In Sāōkhya-Yoga the self is characterized by pure consciousness, unlike in Nyāya and Vaiśeṣika, where consciousness is only one of the attributes of the soul. It is the false identity of the *puruṣa* with material nature (which in its psychological and bodily effects are represented by intellect, ego-sense, mind, five motor senses, five senses of knowing, five subtle elements, the five gross elements of earth, water, fire, wind and ether), which is called ignorance that causes bondage to the self. This false identification comes about due to the capacity of the intellect to reflect *puruṣa* onto itself as a mirror, which then imagines itself to be the experiencer of all that happens. This false knowledge has to be rectified by true knowledge. This is the stage when discriminate discernment as to the true nature of *puruṣa* and material nature dawns. There is a difference in the way this stage is described in Sāōkhya and Yoga, but both generally agree to it in essence.

It is perhaps in the means that one adopts for liberation that the two still retain their distinctiveness. The Sāōkhya does not elaborate on it and mentions only meditation on the difference between material nature and *puruṣa* as the means (*Sāōkhyakārikā* 64). This resembles the Nyāya and Vaiśeṣika approach through study, reflection and applied meditation. On the other hand, Yoga lays down an elaborate plan of psychological preparation which will result in isolating *puruṣa* from its entanglement with material nature through an eightfold path of Yoga.

While Sāōkhya describes this state of liberation as *apavarga* (escape from pain), Yoga calls it *kaivalya* (the self abiding alone in its own nature). Both Sāōkhya and Yoga believe that this liberation can be attained while still living in the world. The second liberation is that which comes on the fall of the body and is termed *videhamukti* (see Sāōkhya).

7 Mīmāṃsā and Vedānta

These two systems can be discussed together as they have some common understanding of the Vedas, although they differ in their basic philosophy. Of all the six systems, the Mīmāṃsā assigns a paramount position to the ritualistic portion of the Vedas (*brāhmaṇas*). Ritual is the pivot around which its ontology and epistemology are built. As it accords an inferior status to the Upaniṣads (that come at the end of the Veda and are known as Vedānta), it is called Pūrva Mīmāṃsā (reflection on the earlier portion of the Veda). Opposed to this is the Uttara Mīmāṃsā, for whom the Upaniṣads are the quintessence of philosophy.

The selves in Mīmāṃsā are eternal, all-pervasive, have knowledge as their attribute, are many and are completely under the control of ritualistic activity (karma), which is also at the back of the formation of the world of objects. As karma is self-sufficient to explain both the world and the selves there is no belief in the existence of God. It is thus a realistic school which postulates the continuation of the self after death in another world until it returns to the earth after reaping the fruit of its ritual acts in the previous life.

Virtue (*dharma*) and vice (*adharma*) are called together *apūrva* (something never known before) and are the result of ritual action. *Dharma* is defined in the school as an injunction to perform a deed (sacrifice), and *adharma* is its opposite. One learns of what is *dharma* and *adharma* from the Vedas. Karma in this scheme is described mainly as threefold: first, done to obtain some desired result, like the gaining of heaven after death; second, prohibited, as its indulgence will lead to undesirable results; third, mandatory, the performance of which is for its own sake (see Duty and virtue, Indian conceptions of).

If one then acts as instructed by the Veda, the highest result of liberation (which in its earlier stages was equivalent to attaining heaven through *dharma*, or even equivalent *dharma* per se) would be accomplished. This is now understood, like all the other systems, as an end to the round of births and deaths for the individual self and a cessation of its relation to the world. In liberation all attributes of the self, including that of consciousness, also disappear. Thus it resembles Nyāya-Vaiśeṣika in the state of liberation, except that it retains the potential to manifest these qualities. However, once liberation is achieved, according to the Mīmāṃsāview, there will be no need for the self to manifest such attributes as consciousness. Again as in Nyāya-Vaiśeṣika and Sāōkhya-Yoga, the state of liberation is described negatively as *apavarga*, an escape from worldly existence (*saṃsāra*). As in Nyāya-Vaiśeṣika, liberation is also only obtained on the fall of the body and there is no belief in being liberated while still alive (see Mīmāṃsā).

Vedānta is based on the three works the *Brahmasūtra*, the Upaniṣads and the *Bhagavad Gītā*. These three together are known as the Three Texts and at least six schools of Vedānta, along with many variations, have sprung up, divided according to the way the Three Texts have been interpreted. While these schools differ in the ultimate ontological categories and in the relationship of the ultimate being to the individual self, they do not differ substantially in their understanding of the nature of the self.

Since Advaita is the earliest of the Vedānta schools, we can start with the nature of the self given in it. This system pursues relentlessly the logic of monism in the Upaniṣads and maintains the nondual nature of the ultimate truth called Brahman, which is identified with *ātman* (individual self) and is the only reality, all others including the world being explained as a superimposition on Brahman due to ignorance. Brahman is intrinsically without attributes, but it is of the essence of existence, consciousness, and bliss, and it is infinite (Taittirīya Upaniṣad 2.1). Even God is only a superimposition and is brought in to make sense of the world of objects. The individual self, which in all other systems is an ultimate entity, is here a complex of Brahman characterized as an eternal witness

and the objective internal organ (see Monism, Indian).

This complex called the living soul (*jīva*) serves to explain among other things the notion of ego that is superimposed on the pure witness. The living soul is caught up in the world of births and deaths due to beginningless karma and identifies itself with the body and the senses. Since the self is really Brahman, only in association with the internal organ, liberation is not some state to be achieved but is to be realized in experience. The obstacle of ignorance has to be removed and this makes Advaita assign paramount importance to knowledge as the only means to liberation.

Such knowledge, which results in an identity experience with Brahman, can only be of a rare kind. Although the direct means is knowledge, there is a preliminary training which strengthens detachment from the world and cultivates a strong desire for liberation. Then comes the rigorous study under a *guru*, someone who has realized the truth of the Upaniṣads. Next there is the following of the instructions introspectively through reasoning and finally the aspirant must engage in constant meditation on the truth reasoned out. Such a course will result in an immediate experience of the identity with Brahman in the form *aham brahma asmi* ('I am Brahman' (Bṛhadāraṇyaka Upaniṣad 1.4.10)). This is possible while still in the body. When the body falls, the self, being disassociated from ignorance, stays as Brahman.

The other Vedānta schools consider the self not in the complex way that Advaita does; they essentially differ in recognizing a personal god as the ultimate entity; the self is eternal, characterized by knowledge and bliss in all these schools; in addition the Dvaita school also believes that each living soul is different from every other. All the schools believe in the bondage of the selves due to karma and ignorance; they are all in agreement about devotion to God being the highest means to liberation. But the liberation is one in which duality persists; they also do not believe in liberation during life, as the ontology assumes attainment of liberation in another world, in which one's God resides (see Vedānta).

See also: Buddhist concept of emptiness; Buddhist philosophy, Indian; Heaven, Indian conceptions of; Hindu philosophy; Mind, Indian philosophy of; Rāmānuja; Śaōkara; Soul, nature and immortality of the

T.S. RUKMANI

References and further reading

Belvakar, S.K. and Ranade, R.D. (1927) *History of Indian Philosophy*, Poona: Bilvakunja Publishing House.(An excellent background to Indian philosophy.)

Bhagavad Gītā(Song of God) (*c.*200 BC), trans. F. Edgerton, *The Bhagavad Gītā*, Harvard Oriental Series 38-9, Cambridge, MA: Harvard University Press, 1944.(A fairly accurate translation of the most important text of Hinduism.)

Chatterjee S. and Datta, D. (1960) *An Introduction to Indian Philosophy*, Calcutta: University of Calcutta.(A good introduction to the six schools.)

Das Gupta, S.N. (1975) *A History of Indian Philosophy*, vols 1, 3 and 5, Delhi: Motilal Banarsidass.(Has detailed information on the six schools.)

Hiriyanna, M. (1952) *Outlines of Indian Philosophy*, London: George Allen & Unwin.(One of the best books in Indian philosophy both for undergraduate and graduate students.)

R.E. Hume (trans.) (1931) *The Thirteen Principal Upaniṣads*, 2nd edn trans. R.E. Hume, London: Oxford University Press, 2nd edn.(Gives a good introduction to Upaniṣadic thought.)

Radhakrishnan, S. (1966) *Indian Philosophy*, vols 1 and 2, London: George Allen & Unwin.(Detailed approach to philosophical schools along with comparisons to Western thought where relevant. As a general survey of Indian thought these two volumes are still the best.)

Griffith, R.T.H. (trans.) (1987) *Hymns of the Ṛg Veda*, revised 3rd edn with popular commentary, New Delhi: Munshiram Manoharlal, 2 vols.(Provides a good background to Vedic poetry.)

Sāōkhyakārikāof Īśvara-kṛṣna, trans. T.G. Mainkar, Poona: Oriental Book Agency, 1964.(Deals with the Sāōkhya school and has a good translation of the kārikās.)

Upaniṣads (800-300 BC), trans. P. Olivelle, *Upaniṣads*, Oxford: Oxford University Press, 1996.(A new translation that may well be the best; readable and accurate.)

Self-control

A human person or self possesses powers that can come into conflict. Reason may have to struggle to overcome contrary desire. Self-control may be characterized as the ability to regulate or resolve such conflict correctly. A well-ordered self has self-control; a disordered self lacks it. When an agent lacks self-control, inner conflict often results in the victory of evil over good. Philosophers such as Plato, Aristotle, and Kant have painted portraits of the disordered self. They are alike in portraying its actions as stemming from appetites or desires that are not properly ruled by some 'higher part' of the self. These philosophers have also proposed accounts of the well-ordered self. They are alike in depicting it as a self in which a part of the self connected with reason or wisdom governs a 'lower part' associated with appetites or desires.

1 The disordered self

According to Plato, a disordered self is a soul whose parts or components are in conflict. The *Phaedrus* likens the soul to the union of powers in a pair of horses and their charioteer. One of the horses struggles against control by the charioteer. If the rebellion were to succeed, the soul would do something evil the charioteer does not want to do. In Plato's *Republic*, the soul is divided into rational, spirited and appetitive parts. A struggle for control between the rational part and the spirited part or the appetitive part occurs in a disordered self.

For Aristotle, a disordered self is an incontinent or akratic self (*Nicomachean Ethics*). It suffers from weakness of will or *akrasia* (see Akrasia). Suppose some people on a diet ought not to eat sweets. Knowing this, when they perceive that this piece of cake is sweet, their practical reasoning should move them to reject it. But if they are incontinent, an appetite for pleasure may make another line of reasoning more salient. In the presence of an appetite for pleasures of taste, if they believe that all sweet things are pleasant, their perception that the piece of cake is sweet may move them to eat it. Though they are aware that they ought not to eat sweets, they eat the cake because their reasoning is unduly influenced by an appetite.

Aristotle recognized two kinds of incontinence: the weak deliberate about what to do but fail, on account of their passions, to stand by their conclusions; the impetuous fail to deliberate, in some cases because they are too quick to make up their minds and in others because their appetites are too strong. Both the weak and the impetuous are apt to be aware of the general practical principle that should determine their actions. Hence, as Aristotle observes, the incontinent are likely to repent, acknowledging in retrospect that they have done something they should not have done.

According to Kant (1793), the disordered self suffers from a radical evil in human nature. Influenced by the Christian doctrine of original sin, Kant believed that there is a propensity to evil deep in all of us. We bring it upon ourselves by freely adopting a policy for acting that does not respect the proper order of the incentives for the will. The incentives provided by our inclinations or sensuous desires always ought to be subordinated to the incentive of respect for the moral law in the maxims on which we act. A policy that does not respect this order manifests itself in particular cases in the frailty, impurity, or wickedness of the human heart. The frail are those in whom the incentive of respect for moral law proves in practice to be weaker than their inclinations; they want to act morally but do not do so. The impure are those by whom dutiful actions are done not purely for duty's sake; they act in accord with the moral law but from mixed motives because respect for the law is not a sufficient incentive for their wills. The wicked neglect the incentive of respect for moral law altogether; even when they act in accord with the moral law, they do the right thing for the wrong reasons and so are corrupt in their cast of mind (see Sin §3).

These three pictures of the disordered self have something in common. They portray its actions as flowing from appetites or desires that have not been subordinated properly to the rule of some higher part or aspect of the self. Bad or evil action results when the self allows unruly appetites or desires to get out of hand.

2 The well-ordered self

For Plato, the remedy for lack of self-control consists of establishing the mastery of the rational part of the soul and rendering docile the other two parts. The rational part of the soul is described in the *Republic* as the ruling principle: the self is well-ordered when reason's two subjects are agreed that reason ought to rule and so do not rebel against it.

As Aristotle sees it, practical wisdom determines the right rule for action. Since an incontinent person is one who allows bodily appetites to move them to act contrary to the right rule, a continent person will be one who possesses such appetites but does not allow them to have such motivating force. Aristotle tells us that the continent person does nothing contrary to the rule for the sake of the bodily pleasures. A person who is merely continent will have bad appetites and will feel pleasure contrary to the right rule but not be led by it to act contrary to the rule. Hence, Aristotle's initial remedy for lack of self-control is continence. A continent self is well-ordered because it has self-control with respect to action. For Aristotle, however, merely being continent is not the ideal state of the self; it is better to be temperate. Like the continent agent, the temperate agent does nothing contrary to the rule for the sake of bodily pleasures. But the temperate agent does not have bad appetites and does not feel pleasure contrary to the right rule. In such a self not only actions but also appetites and feelings are in accord with the rules specified by practical wisdom. Thus, Aristotle's ultimate remedy for lack of self-control is the virtue of temperance. Yet it is doubtful that temperate people should be thought of as having self-control. Lacking bad appetites, they seem to have nothing unruly within them that needs to be kept under control. It appears, then, that a person who has self-control is one who achieves continence but falls short of temperance (see Virtues and vices §§2-3).

According to Kant, the remedy for radical evil in human nature that manifests itself in frailty, impurity, and wickedness is a moral revolution in character in which one freely adopts a policy for acting that respects the proper order of incentives for the will. There is in all of us a capacity for respect for the moral law to serve by itself as a sufficient incentive for the will, and this capacity is not destroyed by our propensity to evil. Kant thought it can be restored to its original purity if we reverse, in a single unchangeable decision, the evil policy of not always subordinating the incentives of inclination or sensuous desire to the incentive of respect for the law. Kant did not profess to know how such radical character change is possible. He was confident, however, that we can bring it about because he was persuaded both that it is our duty to do so and that whatever we ought to do we can do. And such a moral revolution will, he thinks, set us on a path of continual progress from bad to better in our actions. So the Kantian remedy for lack of self-control is a revolution in character that results in respect for the moral law, which is the legislation of our own practical reason, being restored to its dominant place in the hierarchy of our incentives for action.

These three accounts of self-control resemble one another to a considerable extent. In all of them, self-control consists of a 'higher part' of the self, associated with reason or wisdom, governing a 'lower part' of the self, associated with bodily appetite, sensuous desires, passions or feelings. Though such rationalistic accounts of self-control have been dominant in philosophy, they have not escaped criticism. Few philosophers would consider unbridled wantonness or self-indulgence virtuous. But many contemporary feminists, for example, are apt to suspect gender bias in the thought of the self as a monarchy ruled by reason. Such critics are motivated to construct more democratic models of the well-ordered self in which parts of the self cooperate, reason need not always be in control, and the cultivation and expression of feeling are valued (see Feminist ethics).

See also: Emotions, philosophy of; Morality and emotions

PHILIP L. QUINN

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985, VII.(Discusses incontinence and continence.)

Kant, I. (1793) *Die Religion innerhalb der Grenzen der blossen Vernunft*, trans. T.M. Greene and H.H. Hudson, *Religion Within the Limits of Reason Alone*, New York: Harper & Row, 1960, I.(Presents the doctrines of radical evil and moral revolution.)

Plato (*c*.380s-370s BC) *Phaedrus*, trans. R. Hackforth, Cambridge: Cambridge University Press, 1952; repr. in *The Collected Dialogues including Letters*, ed. E. Hamilton and H. Cairns, Oxford: Oxford University Press, 1975. (Compares the human psyche to a charioteer and two horses.)

Plato (*c*.380-367 BC) *Republic*, trans. P. Shorey, Cambridge, MA: Harvard University Press, 1930; repr. in *The Collected Dialogues including Letters*, ed. E. Hamilton and H. Cairns, Oxford: Oxford University Press, 1975, IV.(Presents the theory of the tripartite psyche.)

Rorty, A.O. (1980) '*Akrasia* and pleasure: *Nicomachean Ethics* Book 7', in A.O. Rorty (ed.) *Essays on Aristotle's Ethics*, Berkeley and Los Angeles, CA: University of California Press.(Useful discussion of *akrasia* or weakness of will.)

Wood, A.W. (1970) *Kant’s Moral Religion*, Ithaca, NY: Cornell University Press, ch. 6.(Discusses radical evil and its remedy.)

Self-cultivation in Chinese philosophy

Chinese philosophy may be viewed as disciplined reflections on the insights of self-cultivation. Etienne Balazs asserted that all Chinese philosophy is social philosophy and that, even if Chinese thinkers dwell upon metaphysical speculation, they will sooner or later return to the practical issues of the world here and now. This concern for the concreteness of the life-world gives the impression that the social dimension of the human condition features so prominently in the Chinese world of thought that the idea of the group takes precedence over conceptions of the individual self. The anthropological studies that contrast the Chinese sense of shame with the Western sense of guilt further enhance the impression that external social approval, rather than internal psychological sanction, defines the moral fabric of Chinese society. The prevalent sociological literature on the mechanism of 'saving face' as a key to understanding Chinese interpersonal relationships also stresses the centrality of external conditioning in Chinese ethics.

If we follow this line of thinking, it is easy to assume that Chinese philosophers are preoccupied with neither the transcendent referent nor the inner psyche. They are not particularly interested in questions of ultimate reality such as the creator, the origin of the cosmos or the existence of God. Nor are they engrossed in problems of the mind such as consciousness, self-identity or moral choice. Indeed, Chinese philosophy as social philosophy seems exclusively immersed in issues of correct behaviour, familial harmony, political order and world peace. Even strands of thought that emphasize the aesthetic experience of the self are all intimately bound up with the highly ritualized world of human-relatedness. Actually the spirit of spontaneity, as a liberation from social constraints, should be appreciated in terms of a conscious reflection on and critique of society and thus inherently sociological.

However, this widely held opinion of Chinese philosophy is seriously flawed. While it offers a common-sense picture of where the strength of Chinese thought lies, it does not address the underlying reasons or the actual processes that define the main trajectory of the Chinese modes of thinking. Wing-tsit Chan suggests a more comprehensive characterization of Chinese philosophy as humanism: 'not the humanism that denies or slights a Supreme Power, but one that professes the unity of man and Heaven' (Chan 1963: 3).

It is crucial to note that 'humanism' so conceived is diametrically opposed to secular humanism as a distinctive feature of the Enlightenment mentality of the modern West. Western humanism emerged as a thorough critique of spiritualism and a radical departure from naturalism, or a sense of affinity with nature; it was the result of secularization. Chinese humanism, on the other hand, tends to incorporate the spiritual and naturalist dimensions in a comprehensive and integrated vision of the nature and function of humanity in the cosmos.

The advantage of characterizing Chinese philosophy as humanistic rather than sociological is to open the possibility of allowing aesthetic, religious and metaphysical as well as ethical, historical and political perspectives to shape the contours of the Chinese reflective mind. This synthetic approach better captures the spirit of Chinese thought because it was historical and social change, rather than speculation, which was instrumental in the outgrowth of humanism as a defining characteristic of Chinese philosophy.

1 Chinese philosophy: the beginnings

The advent of Chinese philosophy was not marked by rupture from a mythic, religious world of shamanistic panpsychism to 'rational' reflection, but by a moral discourse on the interplay between the Mandate of Heaven and human virtue. As the pre-Confucian *Shijing (Book of Odes)* notes, the establishment of the Zhou civilization in twelfth century BC was the result of virtuous deeds:

> Don't you mind your ancestors!
> Cultivate your virtue.
> Always strive to be in harmony with Heaven's Mandate.
> Seek for yourselves the many blessings.

(*Shijing*, in Chan 1963: 7)

The idea of cultivating virtue as a human endeavour to harmonize with Heaven's Mandate and to seek blessings

provides a root metaphor for self-cultivation.

2 Philosophy as a spiritual quest: Confucius

This lack of differentiation between religious commitment and intellectual reflection, as implied in this idea of self-cultivation, has profound implications for our understanding of the nature of Chinese philosophy, which is a spiritual discipline directed toward human flourishing. The realms of abstract theory, devoid of significance for practical daily living, are rejected as the proper domain for the life of the mind. Indeed, the insistence that thinking must be grounded in the lived concreteness of the human condition is based on the 'anthropocosmic' vision that human beings, through self-effort, can fully realize their nature as co-creators of the cosmic process.

Confucius' pithy account of his life history is a paradigmatic example that self-cultivation is pivotal in each stage of his learning to be human:

> At fifteen, I set my heart on learning.
> At thirty, I took my stand.
> At forty, I came to be free from doubts.
> At fifty, I knew the Decree [Mandate] of Heaven.
> At sixty, my ear was attuned.
> At seventy, I followed my heart's desire, without overstepping the line.

(*Analects* 2.4)

The very fact that Confucius located the beginning of his spiritual journey at fifteen indicates the conscious choice to shape the direction of his life through learning as awakening, and marks the definitive boundary between givenness and wilful transformation. Meaningful life begins when one has made a conscious choice to learn to become human through self-effort. The ability to 'plant one's feet firmly on the ground' and find a proper niche in society implies that the inherently personal undertaking of self-cultivation necessarily involves a social dimension. However, despite the commitment that learning as awakening entails the internalization of social norms in every facet of ordinary daily living, the inner strength of self-knowledge does not depend upon standards superimposed from the outside. Indeed, when one is free from 'perplexities', there is a depth of self-awareness which engenders a profound sense of direction.

The transcendent dimension at the age of fifty, instead of a departure from the 'humanist' quest for self-realization, offers a new insight on the human condition. The significance of the evocation of the Mandate of Heaven is twofold: on the one hand, it indicates a critical, if not tragic, recognition of the structural limitations inevitably imposed by aging; at the same time, it also symbolizes a transcendental breakthrough in self-awareness. The former shows that since we are merely human; we cannot but come to terms with our mortality. As Confucius said, to know what we do not know is wisdom; we may analogize that to accept fully the responsibility for what we cannot do is a sign of courage. This 'transcendental breakthrough', then, implies that despite our conditionality, what we do and what we become as humans has cosmological significance.

The deliberate choice of aural rather than visual perception as a way of depicting Confucius' stage of self-cultivation at sixty is profoundly meaningful. The art of listening requires the calmness of the heart and the patience to receive all the messages without prejudging any of them. As an accomplished listener, the dynamic temporality broadens and deepens one's awareness of the outside world without affecting inner tranquillity. The harmony between what one does out of natural impulse and what one ought to do as dictated by moral obligations is, therefore, a further refinement on training the ear to be docile. Within this context, the spirit of spontaneity and freedom that Confucius experienced at seventy, as the result of decades of continuous self-cultivation, has become a standard of inspiration in Chinese moral philosophy.

While Confucius' life history is itself a lesson of self-cultivation through exemplary teaching, it is not meant to be a rigid model for human flourishing. Confucius' attainments depended so much on a confluence of conditions involving economic circumstances, social class, political situation and cultural milieu that it is unique and unrepeatable (see Confucius). Confucian followers rejected the idea of imitating Confucius but insisted upon the relevance of his example for emulation (see Confucian philosophy, Chinese). The best way to emulate Confucius is to practise his way of learning to be human through our own self-cultivation. A commitment to internalize the

central values of his teachings and integrate them into our daily life is required. However, this does not mean that we have to suspend our rational thinking and critical judgment in order to embrace the Confucian project. The fusion of intellectual inquiry and spiritual transformation suggests that the philosophy of self-cultivation is predicated on an experiential understanding of how it actually works; acceptance of any unexamined dogmas or implicit doctrines is not a precondition.

In response to the query, ‘is there a single saying that one can act upon all day and every day?’ Confucius said, ‘Perhaps the saying about reciprocity: "Never do to others what you would not want them to do to you".’ What is the status of this ‘golden rule’ in Confucian moral philosophy? On the surface, it seems to be a dogma and a doctrine, yet, as presented, it does not have the force of ‘thou shall not do otherwise’. It is an observation, a suggestion, a recommendation. The implicit message is: put this ‘rule’ into practice in your ordinary interpersonal relationships, and see how it actually works. Assuming that this emphasis on considerateness is sometimes unworkable because the anticipated sympathetic response is absent, should the practice be continued? As concrete practice rather than abstract principle, a realistic assessment of the situation is critical in the application of the Confucian golden rule. When Confucius was asked about the merit of repaying malice with kindness, instead of encouraging the student to cultivate what may be called a higher morality of considerateness, Confucius retorted, how are we to repay kindness? If our reciprocal response to malice is kindness, we may have difficulty responding appropriately to kindness. The recommendation that we repay malice with justice and kindness with kindness is to encourage attention to the specificity of the situation in formulating a measured response.

However, it is mistaken to characterize Confucian self-cultivation as a form of situational ethics. As a virtue-centred moral philosophy, an overriding concern is the cultivation of a personality through character-building that will be able to respond to all situations with dignity and openness:

> The profound persons have nine areas for which they are mindful. In seeing they are mindful of seeing clearly; in hearing they are mindful of hearing distinctly; in looks they are mindful of being good-natured; in their manners of being reverential; in their words of being trustworthy; in their works of being diligent. When in doubt they are mindful of asking for information; when angry they are mindful of the consequences, and when they see a chance for gain, they are mindful of whether the pursuit of it would be consonant with rightness.
>
> (*Analects* 16.10)

Concrete situations provide the opportunity to put general principles, such as the golden rule, into practice, but self-cultivation as a way of learning to be human takes as its primary purpose the fostering of balanced personalities:

> When there is a preponderance of native substance over acquired refinement, the result will be churlishness. When there is a preponderance of acquired refinement over native substance, the result will be pedantry. Only a well-balanced admixture of these two will result in gentlemanliness.
>
> (*Analects* 6.18)

It should be noted that the self in Confucian self-cultivation is not an isolated or even isolatable individual, but a centre of relationships. As a centre, there is a unique configuration of social roles evolving around the autonomous and independent self. The self as an end, rather than a means, can never be reduced to these roles. Confucius’ injunction that the profound person is not a vehicle clearly shows that the dignity of the person cannot be fully appreciated by referring merely to their role and function in society. The paradox is of course that each social role (that is, parent, child, teacher, student, patron, client, employer, employee, farmer, worker, merchant, business executive, lawyer, doctor, journalist, artist, writer or computer programmer) provides an opportunity for self-realization. As a centre of relationships, the self is enriched and empowered by a continuous interaction and interchange with other selves. Nevertheless, balanced personalities are not simply well-adjusted to the vicissitudes of the external environment; an inner core of self-knowledge serves as a spring of rich spiritual resources.

3 Ultimate justification: Mencius

The great contribution that Mencius made to the Confucian project is precisely locating this inner core in human nature:

> A profound person [gentleman] steeps himself in the Way because he wishes to find it in himself. When he

> finds it in himself, he will be at ease in it; when he is at ease in it, he can draw deeply upon it; when he can draw deeply upon it, he finds its source wherever he turns. That is why a profound person wishes to find the Way in himself.
>
> (*Mengzi* 4B14)

There is a paradox in the assertion that profound persons, intent on finding the Way in themselves, steep themselves in the Way. If the Way is already in us, all we need to do is to find it through introspective self-examination. Indeed, Mencius seems to believe that is the case:

> All the myriad things are present in me. There is no greater joy for me than to find, on self-examination, that I am true to myself. Try your best to treat others as you would wish to be treated yourself, and you will find that this is the shortest way to humanity.
>
> (*Mengzi* 7A4)

However, by steeping ourselves in the Way, we are enjoined to make a conscious effort to realize that, if we follow our original nature, we are intrinsically congenial to the Way. By implication, since we are not necessarily what we ought to be, we may easily mistake that part of us which is contrary to the Way as the true self. The art of self-cultivation, then, is to know the difference and to act accordingly. Mencius straightforwardly articulates what he observes as the central *Problematik* in self-cultivation:

> When our senses of sight and hearing are used without thought and thus become obscured by material things, the material things act on the material sense and lead them astray. That is all. The function of the mind-and-heart is to think. If we think, we will get the principles of things. If we do not think, we will not get them. This is what Heaven has given us. If we first build upon the ‘great body’ of our nature, then the ‘small body’ cannot overcome it. It is this which makes human beings great.
>
> (*Mengzi* 6A15)

Although we are not what we ought to be, the process by which we learn to become so can be initiated from the very structure of what we are here and now. The assumptive reason is that the ultimate justification for self-cultivation lies in the feeling, willing, knowing, and sensing capacities of the heart-and-mind. As long as our heart-and-mind can feel the suffering of others, there is the authentic possibility of self-cultivation. Human beings so defined are primarily sentient beings. Our sensitivity, which includes rationality, enables us to experience directly the sympathetic and reciprocal resonances with other human beings and the world at large.

Yet, it is necessary that we know the difference between the ‘great body’ and the ‘small body’ in our nature. We must not allow our desires, especially appetite and sex, to dominate, or we will be obsessed with the demands of the small body. As a result, our behaviour can be easily defined in terms of the rules of the game common to all animals: survival and procreation. On the other hand, if we cultivate our great body in order to realize that ‘all the myriad things are present in me’, the potential for our interconnectedness is immense. Mencius’ art of nourishing the flood-like vital energy addresses this in seemingly mythical terms:

> [The flood-like vital energy] is the sort of vital energy which is utmost in vastness and utmost in strength. If through uprightness you nourish it and do not interfere with it, it fills the space between heaven and earth. It is the sort of vital energy which matches rightness with the Way. Without these it starves. It is generated by the accumulation of right deeds, and is not the sporadic right deeds that one might make a grab at. If anything in conduct is dissatisfying of the heart, it starves…. You must work at it and never let it out of your mind. At the same time, while you must not let it out of your mind, you must not forcibly help it either.
>
> (*Mengzi* 2A2)

Viewed from this perspective, Mencius’ thesis that human nature is good provides an ultimate justification for self-cultivation. The reason that we can cultivate ourselves is because the feeling, willing, knowing and sensing capacities of the heart-and-mind are inherent in our nature (see Xin (heart-and-mind)). Virtues of humanity, rightness, wisdom and propriety can be cultivated by appealing directly to the basic manifestations of human sensitivity: sympathy, empathy, the feeling of shame and dislike, the ability to know the difference between right and wrong and a sense of appropriateness.

4 Cultural transformation: Xunzi

While Xunzi, the great Confucian thinker of the third century BC, shares Mencius' vision that the human condition is improvable through self-effort and that moral education is absolutely necessary for the maintenance of social order, he doubts that human nature will automatically generate the creative process for human flourishing. He insists that the coercive power of the law, the influence of ritual and music, the authority of the ruler, the guidance of the teacher, and the enduring presence of good customs are all necessary for the harmonious functioning of society. His justification for self-cultivation is twofold: it is sociologically necessary on the one hand, and psychologically and epistemologically possible on the other.

Xunzi subscribes to an evolutionary theory of the advent of the human species with particular emphasis on its sociality. He observes that whereas all modalities of being have vital energy, plants have life and animals have consciousness, human beings alone possess rightness. A distinctive feature of the human sense of rightness is the ability to know the need and the intelligence to design the structure to organize human communities for survival and flourishing. Self-cultivation is central to the stability and solidarity of any human community. Human desires are numerous, but the material goods to satisfy them are limited. If desires are not properly channelled, social order cannot be maintained. Although coercive measures from the outside are necessary, in the long run the surest way is still moral self-cultivation. To be sure, human nature suffused with the passions of appetite, sex and aggression cannot in itself generate the mechanism for the development of pro-social sentiments, but the cognitive function of the mind has the capacity to make distinctions and intelligent choices.

Xunzi compares the mind to a pan of water: 'If you guide it with reason, nourish it with clarity, and do not allow external objects to unbalance it, then it will be capable of determining right and wrong and of resolving doubts' (*Xunzi*, in Liang 1996: 222). Understandably, an important concern of Xunzi's moral education is to free the mind from obsessions:

> What are the sources of obsession? One may be obsessed by desires or hates, by the beginning of an affair or the end, by the far away or close by, by breadth of knowledge or by shallowness, by the past or by the present. When one makes distinctions among myriad beings of creation, these distinctions all become potential sources of obsession. This is the danger in the use of the mind which is common to all human beings.
>
> (*Xunzi*, in Liang 1996: 222)

However, despite the mind's vulnerability, three distinctive qualities it possesses enable it to save humanity from self-destruction: emptiness, oneness and tranquillity.

The mind's receptivity is inexhaustible; no matter how much information it receives, there is always room for new. The emptiness of the mind offers a great capacity for learning. Xunzi believes that although human beings, as compared with other animals, are weak in physical strength, through learning they can be socialized to become the most intelligent. The mind also has the synthetic power to integrate fragmented ideas into a coherent pattern, and can do so without losing sight of its own inner identity. Similarly, while the mind responds dynamically to external stimuli, it can always maintain its internal tranquillity. These qualities enable the mind to exercise its cognitive function to control the passions of human nature.

Consequently, what Xunzi recommends as the central task of self-cultivation is to use the mind to govern human nature. Specifically, he recognizes the necessity of artfully constructed cultural forms to help the mind do this work. Law, ritual, music, norms and customs are all critically important in helping the mind to transform human passions so that their destructive influence on society can be minimized. Both Mencius and Xunzi see the need to control desires. They do not advocate asceticism, but they are fully aware that the best way to nourish the mind is to make desires few.

5 The root for human flourishing

The centrality of self-cultivation in classical Confucian humanism is captured in a pithy statement in the *Daxue (Great Learning)*:

> From the Son of Heaven on down to the common people, all must regard cultivation of the personal life as the root. There is never a case where the root is in disorder and yet the branches are in order. There has never been a case that what is treated with great importance becomes a matter of slight importance or what is treated with

> slight importance becomes a matter of great importance.
>
> (*Daxue* 1)

It is profoundly significant that the root and importance of self-cultivation is extended to the family, state and the world. It is not only a subject in moral education of the young but also the major concern of the body politic as a whole:

> The ancients who wished to manifest their clear character to the world would first bring order to the states. Those who wished to bring order to their states would first regulate their families. Those who wished to regulate their families would first cultivate their personal lives. Those who wished to cultivate their personal lives would first rectify their minds. Those who wished to rectify their minds would first make their wills sincere. Those who wished to make their wills sincere would first extend their knowledge. The extension of knowledge consists in the investigations of things.
>
> (*Daxue* 1)

Self-cultivation so conceived occupies a pivotal position in the link between the self and the community mediated by a variety of political, social and cultural institutions. At the personal level, it involves a complex process of empirical studies and mental disciplines. In terms of the whole project of human flourishing, it serves as the foundation of stable families, ordered states and a peaceful world.

The Confucians may have been particularly concerned about the social and political implications of self-cultivation, but self-cultivation features so prominently in all Chinese philosophy that it is central to Chinese modes of thinking in ethics, aesthetics, metaphysics and epistemology. In the Chinese life of the mind, the primary concern in ethics is how to establish the standards of inspiration for the young through exemplary teaching, in aesthetics how to acquire the real taste of beauty through rigorous discipline, in metaphysics how to understand ultimate reality from within, and in epistemology how to know in the spirit of impartiality and mutuality. Indeed, the centrality of self-cultivation compels Chinese thinkers to ground ethics in practice, aesthetics in experience, metaphysics in wisdom and epistemology in communication. The Chinese equivalent of the Greek dictum that an unexamined life is not worth living is that an uncultivated person is not yet human. A defining characteristic of learning to be human is to engage in a ceaseless process of self-cultivation. Philosophy so conceived is neither abstract theorizing nor speculative thinking but a spiritual discipline. As a spiritual discipline, the love of wisdom is a transformative act.

6 The embodiment of the Way: the Daoist approach

Self-cultivation for the Daoists requires the art of unlearning (see Daoist philosophy). While in the pursuit of knowledge one gains by day, in the quest for the *dao* one must let go, lose and be silent:

> To rarely speak - such is the way of Nature.
> Fierce winds don’t last the whole morning.
> Torrential rains don’t last the whole day.
> Who makes these things?
> If Heaven and Earth can’t make these last long-
> How much more is this true for man!
>
> (*Daodejing* 23, in Liang 1996: 223)

The constant and enduring way of nature is therefore undramatic, nonviolent, peaceful, tranquil and colourless. The precondition to emulate this is to acknowledge our fatal flaw of not knowing that we do not know:

> To know you don’t know is best.
> Not to know you don’t know is flaw.
> Therefore, the Sage’s not being flawed
> Stems from his recognizing a flaw as a flaw.
> Therefore, he is flawless.
>
> (*Daodejing* 71, in Liang 1996: 222)

Having acknowledged our ignorance, we begin to appreciate that, in the spirit of the *dao*, self-cultivation does not take the form of action:

> The softest, most pliable thing in the world runs roughshod over the firmest thing in the world.
> That which has no substance gets into that which has no spaces or cracks.
> I therefore know that there is benefit of taking no action.
> The wordless teaching, the benefit of taking no action -
> Few in the world can realize these!

(*Daodejing* 43, in Liang 1996: 222)

Even though the *dao* is ineffable, it can be experienced. The tacit 'knowledge' of non-action can help us to embody it. To experience the *dao* or to acquire a taste of it is the focus of Daoist self-cultivation:

> To know the constant is to be all-embracing;
> To be all-embracing is to be impartial;
> To be impartial is to be kingly;
> To be kingly is to be like Heaven;
> To be like Heaven is to be one with the Dao;
> If you're one with the Dao, to the end of your days you'll suffer no harm.

(*Daodejing* 16, in Liang 1996: 223)

Zhuangzi describes vividly how, for the Daoists, an experiential understanding of the *dao* is the most efficacious way of realizing the *dao* in our ordinary daily living. The relationship between the perception of what the dao is, and the spiritual journey through which the *dao* is embodied, is articulated by the language of nature symbolizing spontaneity, constant flux and incessant transformation:

> Tao [Dao] has reality and evidence but no action or physical form. It may be transmitted but cannot be received. It may be obtained but cannot be seen. It is based in itself. Before heaven and earth came into being, Tao existed by itself from all time. It gave spirits and rulers their spiritual powers. It created heaven and earth. It is above the zenith but it is not high. It is beneath the nadir but it is not low. It is prior to heaven and earth but it is not old. It is more ancient than the highest antiquity but is not regarded as long ago.

(*Zhuangzi*, in Chan 1963: 194)

The spiritual implications for this perception of the *dao* are equalizing all things and opinions, abandoning selfishness of all descriptions and attaining enlightenment by moving into the realms of 'great knowledge' and 'profound virtue'. In Zhuangzi's own style of expression, the person who has learned *dao* is able to transcend this world, all material things and all life:

> Having transcended all life, he became as clear and bright as the morning. Having been as clear and bright as the morning, he was able to see the One. Having seen the One, he was able to abolish the distinction of past and present. Having abolished the past and present, he was then able to enter the realm of neither life nor death. Then, to him, the destruction of life did not mean death and the production of life did not mean life. In dealing with things, he would not lean forward or backward to accommodate them. To him everything was in the process of destruction, everything was in the process of perfection. This is called 'tranquillity in disturbance'. Tranquillity in disturbance means that it is especially in the midst of disturbance that [tranquillity] becomes perfect.

(*Zhuangzi*, in Chan 1963: 195)

This remarkable convergence of an ontological reading of the *dao* and an existential interpretation of the true person suggests a fruitful interplay between the aloofness and elusiveness of the *dao*, and the concreteness and practicality of the Daoist experience. The *dao* as an idea is conceptually forever inaccessible, and as a way of life is experientially always present.

7 An anthropocosmic vision

It is easy to assume that the Confucians, being politically concerned, socially engaged and culturally sensitive, take self-cultivation as a this-worldly value-orientation, whereas the Daoists, being disinterested in politics, society and culture, define self-cultivation in other-worldly terms. In a deeper sense, however, Confucians and Daoists are both immersed in the same *Problematik*: how can we respond to the vicissitudes of a constantly changing world in the spirit of spontaneity without losing the inner tranquillity of the mind? In the words of a syncretic text of the Han dynasty, how can we transform together with external things but not lose our true nature? A classical formulation of this central issue is found in the *Zhongyong (Doctrine of the Mean)*:

> When emotions, such as pleasure and anger and sorrow and joy, have not been aroused, the state is called that of centrality. When these emotions are aroused and each and all attain due measure and degree, it is called the state of harmony. The state of centrality is the great root and the state of harmony is the far-reaching way of all existence in the world. Once centrality and harmony are realized, Heaven and Earth take their proper place and all things receive full nourishment.
>
> (*Zhongyong* 1)

The ‘anthropocosmic’ vision embedded in this statement is both Confucian and Daoist. Indeed, it is predicated on an ontological commitment in the ‘Appended Remarks’ of the *Yijing (Book of Changes)*: the great transformation of the cosmos is the highest standard of inspiration for human self-cultivation, and human beings can assist in the transforming and nourishing processes of Heaven and Earth and thus form a trinity with Heaven and Earth through personal self-realization. By implication, what we do in the privacy of our homes and in the practical living of our everyday existence is profoundly meaningful in a cosmological as well as in a psychological sense.

8 Discipline of the mind: the Buddhist insight

This central and enduring motif of self-cultivation philosophy has been greatly enriched by the introduction of Buddhism from India (see Buddhist philosophy, Chinese). The most obvious Buddhist contribution to Chinese self-cultivation philosophy is the discipline of the mind. An early Buddhist master, Huisi (514-77), remarked that, through concentration, we can realize the pure mind; insight derived from the realization of the pure mind enables us to perceive that all sentient beings are ‘harmoniously identified to form one body of a single character’ (Liang 1996: 225). The power of concentration helps us not only to attain calmness, stillness and purity of the mind, but also to function without the form of functioning, and to act without the form of acting. This resembles the Daoist idea of acting in the spirit of non-action.

The introduction of Buddhism was accompanied by a host of spiritual disciplines hitherto unavailable to the Daoist practitioner. Furthermore, its new perceptions of the human condition fundamentally reshaped the basic tenets of Chinese self-cultivation philosophy. The issue of ‘attachment’ loomed large in the new configuration. As Jizang (549-623) notes, ‘if there is an attachment (in the mind), there is a fetter, and one cannot obtain release from birth and old age, sickness and death, care and sorrow, pain and suffering.’ To rise above attachment and perceive things as they are, a combined effort of meditation and wisdom is required. Yet, as Huineng noted, meditation and wisdom are a unity, rather than two separate disciplines:

> Calmness [stillness] itself is the substance of wisdom; wisdom itself is the function of meditation. At the very moment when there is wisdom, then meditation exists in wisdom; at the moment when there is meditation, then wisdom exists in meditation. Good friends, this means that wisdom and meditation are alike.
>
> (Chan 1963: 433)

The use of substance and function to define the relationship between meditation and wisdom is instructive. The penetrating insight which emerges from the meditative practice is itself an illumination of the mind. Since our original nature is the source of wisdom, the meditative practice as a way of clarifying the mind enables our original nature to present itself. As long as that happens, wisdom is there. Meditation is therefore a return and discovery of the original nature whereas wisdom is the character of the purified mind. Huineng insists that the internal and sudden approach is the authentic method of true awakening (see Platform Sutra). If we do not place our trust in external practices but cultivate right views in regard to our original nature in our own minds, self-awakening is always possible. Unfortunately, our deluded minds compel us to seek wisdom by external practice and thus prevent us from awakening to our true nature.

It seems on the surface that Huineng is oblivious to the 'weakness of the will' problem. Where is the source of strength for this sudden awakening? His rival, Shenhui (670-762), offered a reasonable gradualist solution with emphasis on the power of good habits:

> After a master founder has smelted and refined the material, gold and the mineral will presently be differentiated. The more refined, the purer the gold will become, and with further smelting, the residual mineral will become dust. The gold is analogous to Buddha-nature, whereas minerals are analogous to afflictions resulting from passions. Afflictions and Buddha-nature exist simultaneously.
>
> (Liang 1996: 226)

This is the reason that, in his famous verse of self-cultivation, Shenhui perceived the body as the 'tree of perfect wisdom' (*bodhi*) and the mind as 'the stand of a bright mirror' (Chan 1963: 431). The method of self-cultivation is analogous to diligently wiping the mirror at all times so that it will never become dusty.

This common-sense approach, however, was emphatically rejected by Huineng as inadequate. Instead, he advocated sudden enlightenment as not only a more efficacious but also the only true way to understand how the mind actually works in spiritual self-cultivation. Sudden awakening is possible even for those sentient beings filled with passions and sentiments:

> It is like the great sea which gathers all flowing streams, and merges together the small and large waters into one. This is seeing into your own nature. Such a person does not abide either inside or outside; he is free to come or go. Readily he casts aside the mind that clings to things, and there is no obstruction to his passage.
>
> (Liang 1996: 226)

The freedom one experiences as the result of sudden awakening, presumably occasioned by the combined effort of meditation and wisdom, is neither cumulative nor sequential; it happens all at once. Understandably, Huineng's verse, as a response to Shenhui's formulation, is to deny the authenticity of a gradualist approach to self-cultivation:

> The mind is the tree of perfect wisdom.
> The body is the stand of a bright mirror.
> The bright mirror is original clear and pure.
> Where has it been defiled by any dust?
>
> (Chan 1963: 432)

However, in a deeper sense, the strength of Chinese Buddhist self-cultivation philosophy lies in its integrated thinking as the basis for guiding practical living in ordinary daily existence. The fruitful interaction between the gradualist and suddenist approaches enabled the Buddhist teachers, especially Chan masters, to subscribe to suddenist insight as ontological wisdom without losing sight of the gradualist effort as an existential discipline. Fazang (643-712), for example, offered the merging of the one and many as an ontological insight:

> When the feelings have been eliminated and true substance revealed, all becomes an undifferentiated mass. Great function then arises in abundance, and whenever it does, there is surely Perfect Reality. All phenomena are in great profusion, and are interfused but not mixed. The all is the one, for the relation between cause and effect is perfectly clear. As the power of the one and the function of the many embrace each other, their expansion and contraction are free and at ease.
>
> (Chan 1963: 410)

What Fazang propounded here is both an attained state of the mind and the true nature of ultimate reality such a mind perceives. With regard to self-cultivation, Fazang's ontological insight is not at all in conflict with the recommendation of Huangbo (d. 850) of the concrete steps for purifying the mind:

> First, learn to be entirely unreceptive to sensations arising from external forms, thereby purging your bodies of receptivity to externals.
> Second, learn not to make any distinctions between this and that arising from your sensations, thereby purging your bodies of useless discernments between one phenomenon and another.

> Third, take great care to avoid discriminating in terms of pleasant and unpleasant sensations, thereby purging your bodies of vain discriminations.
> Fourth, avoid pondering things in your mind, thereby purging your bodies of discriminatory cognition.
>
> (Liang 1996: 227)

Surely, the Buddhist thinkers, by their rigour of ascetic practices, persistent meditation and profound wisdom, fundamentally reshaped the spiritual landscape of self-cultivation in China, but the distinctive Buddhist traditions that emerged in the Zhongtu (the Middle Country) assumed new forms significantly different from their Indian origins. Chan Buddhism, which exemplifies the creative transformation of Buddhist teaching in China, is a paradigmatic case. The recognition that *buddha*-hood is inherent in our human nature, that our physical bodies are capable of fully realizing our nature and that we can attain *buddha*-hood in this life extends the Buddhist community to embrace the world at large and instills a sense of urgency, a heightened temporality, among the Chan practitioners. The simple and direct method Master Yixuan (d. 867) employed to inspire his disciples is a case in point:

> Time is precious. Don't make the mistake of following others in desperately studying meditation or the Path, learning words or phrases, seeking after the Buddha or patriarchs or good friends. Followers of the Path, you have only one father and one mother. What else do you want? Look into yourselves. An ancient sage said that Yanjnadatta thought that he had lost his head and sought after it, but when his seeking mind was stopped he realized he had never lost it.
>
> (Liang 1996: 228)

Huineng pointedly remarked that our self-nature, which entails the Three Bodies of Buddha (the Dharma-body, the Reward-body, and the Transformation-body), is in our physical bodies. In the classical theory of the Three Bodies, the Dharma-body is the Buddha-body in its true original self-nature that is empty (no bodily existence at all). The Reward-body is the enlightened body with real insight into the 'thusness' of things in its eternal bliss of self-enjoyment, and the Transformation-body is the compassionate manifestations of buddha-hood appearing in various forms to save people. Although the three bodies are in one, are present in all *buddhas*, and are potentially realizable by all human beings, the conditions for their realization are so complex and the time required so long that it is beyond imagination that any physical body has a chance to attain them in its own life time. Reflecting on the implications of Huineng's insistence that 'the Three Bodies are inherent in one's physical nature', Wing-tsit Chan commented:

> The doctrine of 'becoming Buddha in this very body' is a far cry from the original Indian idea that the body is a hindrance to freedom. One cannot help recalling that the Confucianists have always regarded the body as a gift from parents and as such it is a sacred trust and therefore to be well taken care of, and that for centuries the Taoist [Daoist] religion had tried in many ways, including medicine, diets, exercise, sex technique, and breath control to make the body suitable for everlasting life on earth. These are some of the roots that make Zen [Chan] essentially Chinese.
>
> (Chan 1963: 437)

9 The cultivation of the body

The Chinese heritage as a whole takes our physical nature, the body, absolutely seriously. Self-cultivation, as a form of mental and physical development and rejuvenation is an ancient Chinese art involving such exercises as rhythmic bodily movements and breathing techniques. Understandably, the classical Chinese conception of medicine focuses on healing rather than merely curing illness because its primary function is not only preventing sickness but also restoring vital energy essential for the wholeness of the body. Since the level of vital energy required for health varies according to the specific conditions of a particular person, the wholeness of body is situationally defined as a dynamic process rather than a static structure. The maintenance of health, accordingly, involves a delicate balance among a wide range of environmental, dietary, physiological and psychological factors. The rhythm of the living body is seasonal, for the body is temporal as well as spatial. The temporalization of the body as budding (spring), growing (summer), withdrawing (autumn) and preserving (winter), symbolizes longevity as both an attainment and a natural process.

Our bodies are not merely physiological entities; they are energy fields rather than conglomerates of discrete organs. Indeed, their functions are analogous to 'power stations' and 'communication centres'. The body is, therefore, a conduit through which we communicate with all modalities of vital energy in order to realize the ultimate meaning of life in ordinary human existence. Religious Daoism, especially the cult of immortality, has developed one of the richest traditions of cultivating the body as a spiritual discipline.

10 Learning to become a sage: the neo-Confucian Way

To the neo-Confucians, the central problem in self-cultivation philosophy is how to learn to become a sage (see Neo-Confucian philosophy). This is founded on the faith in the perfectibility of human nature. Zhou Dunyi (1017-73), having responded positively to the question whether or not sagehood can be learned, succinctly outlined a course of action for the spiritual journey:

> The essential is to concentrate on oneness. By concentrating on oneness is meant having no desire. Having no desire, one is vacuous while tranquil, and straightforward while in action. Being vacuous while tranquil, one becomes intelligent and hence penetrating; being straightforward while active, one becomes impartial and hence all embracing. Being intelligent, penetrating, impartial, and all-embracing, one is almost a sage.
>
> (Chan 1963: 473)

Sagehood, as the most authentic manifestation of humanity, is the result of self-cultivation. Indeed, the state of being desireless is existentially impossible for any human being to attain. This is because we observe the human condition from the perspective of the senses. If we focus our attention on the heart-and-mind rather than the senses, we can experientially understand that it is possible to rise above all desires, if only for a moment. If we can experience this even once, we recognize that sagehood, far from being an unattainable ideal, can be a standard of inspiration for daily practical living. As Zhang Zai (1020-77) notes, if we are not encumbered by the perceptions of the senses and engage ourselves in realizing fully the inner abilities of the heart-and-mind, we will appreciate sagelike qualities in our own ordinary experience.

This distinction between what we ontologically are, and what we existentially can be and must strive to become, is critical for self-cultivation philosophy. All three teachings - Confucianism, Daoism and Buddhism - hare a profound faith in human nature as a necessary and sufficient condition for self-realization (see Xing). Since ontologically sagehood, *dao* or *buddha*, is inherent in our nature, we are potentially sages, true persons and *buddha*s. Existentially, however, no matter how hard we try we can never fully realize sagehood, *dao*, or *buddha*. There is always room for improvement. Yet, although we are not what we ought to be, the ideals of what we must strive to become are inherent in our nature, indeed our bodies.

This non-dualistic thinking offers a radically different approach to the mind/body problem. The deliberate and arbitrary split of the self into *res cogitans* and *res extensa* is absent (see Mind, philosophy of). The body is never reduced to a reified corporeality, nor are the thinking and reflecting functions of the living person totally subsumed under the category of the mind. The sharp contrast between the mental and the physical totally undermines the affective dimension, feeling, as an essential feature of being human. Surely human beings are capable of willing, sensing and knowing, but feeling underlies all mental and physical activities. No expression of willpower, sense of appropriateness or cognition can be devoid of feeling. It requires a stretch of our imagination to believe that we can will, sense or know independent of how we feel. It is human to feel. Natural feelings that arise in response to the suffering of others defines who we are.

The cultivation of moral values such as humanity, rightness, propriety, wisdom, and faithfulness depends upon the nourishment of the appropriate feelings: commiseration for humanity, shame and dislike for rightness, deference for propriety and a sense of right and wrong for wisdom. Cheng Hao (1032-85) addresses this issue in his commentary on Mencius:

> Humanity implies impartiality, that is, to make (the moral principle) human. Rightness means what is proper, the standard for weighing what is of greater or smaller importance. Propriety means to distinguish (to determine ranks and functions). Wisdom is to know. And faithfulness (belief) means 'We have it'. All things have nature. These Five Constant Virtues are nature. As to commiseration and so forth (the sense of shame, the sense of deference and compliance, and the sense of right and wrong), they are all feelings. Whatever is aroused is called feeling.

(Chan 1963: 537)

This integrated vision of virtues and feelings in the context of human nature and the heart-and-mind enabled the neo-Confucians to practice self-cultivation on a daily basis. For example, Cheng Hao remarked: 'When I practice calligraphy, I am very reverential. My purpose is not that the calligraphy must be good. Rather, my practice is the way of moral training.'

Cheng Yi (1033-1107) made it clear that self-cultivation can be conceived as the site of a contest between nature and emotions, but since nature also reveals itself in feelings, it is the contest between feelings such as commiseration, shame, deference and a sense of right and wrong, and emotions such as pleasure, anger and sorrow (see Xin (heart-and-mind)):

> As the physical form appears, it comes into contact with external things and is aroused from within. As it is aroused from within, the seven emotions - called pleasure, anger, sorrow, joy, love, hate, and desire - ensue. As emotions become strong and increasingly reckless, his nature becomes damaged. For this reason the enlightened person controls his emotions so that they will be in accord with the Mean. He rectifies his mind and nourishes his nature. This is called turning emotions into original nature. The uncultivated person does not know how to control them. He lets them loose until they are depraved, fetter his nature, and destroy it. This is therefore called turning one's nature into emotions.

(Chan 1963: 548)

Following Cheng Yi's teaching, Zhu Xi (1130-1200) singled out reverence as the psychological state most congenial to self-cultivation. As a gradualist, Zhu Xi strongly advocated the necessity of a cumulative process for nurturing the right beliefs, attitude and behaviour. He urged his students to cultivate their hearts-and-minds as a daily routine:

> The mind is that with which man rules his body. It is one and not a duality, is subject and not object, and controls the external world instead of being controlled by it. Therefore, if we examine external objects with the mind their principles will be apprehended.

(Liang 1996: 230)

While he believed that when it is in a state of reverence, the mind will naturally nourish the proper dispositions for human flourishing, he insisted that the advancement of learning lies in the extension of knowledge. By implication, the investigation of things, a kind of empirical study, is relevant to and essential for moral self-cultivation.

Wang Yangming (1472-1529), however, criticized the Cheng-Zhu line of thinking for its failure to fully comprehend the true nature of moral knowledge. His perception that, in the actual practice of self-cultivation, knowing entails a transformative act serves as the basis for this doctrine of the unity of knowledge and action:

> In their learning people of today separate knowledge and action into two different things. Thereupon when a thought is aroused, although it is evil, they do not stop it because it has not been translated into action. I advocate the unity of knowledge and action precisely because I want people to understand that when a thought is aroused that it is already action. If there is anything evil when the thought is aroused, one must overcome the evil thought. One must go to the root and go to the bottom and not allow that evil to lie latent in the mind. This is the basic purpose of my doctrine.

(Chan 1963: 670)

On the surface, Wang's theory is basically therapeutic; its primary purpose is to bring about concrete results in moral behaviour. However, despite its remedial function, the inseparability of knowing and acting is predicated on the centrality of 'primordial awareness' in self-cultivation.

'Primordial awareness' (*liangzhi*), variously referred to as 'innate knowledge', 'the knowledge of the good', 'conscientious consciousness' or simply 'conscience', is what Wang takes to be the true function of human nature and the ultimate reality of all beings. This form of knowing is inherently creative; such knowledge in itself engenders new realities. Strictly speaking, 'primordial awareness' is not a reflection or meditation on external objects. Rather, it is an awareness which arises through embodying the inner patterns of things in an ever-expanding sensitivity of the heart-and-mind. Self-cultivation so conceived involves the extension of one's

‘primordial awareness’ through analogical thinking. The quality of the heart-and-mind is vitally important. Regardless how much empirical knowledge one accumulates, one’s ethical intelligence will only be enhanced when the mind is properly cultivated. Therefore, in Wang’s self-cultivation philosophy, the sincerity of the will, as an effort to purify one’s motivation, takes precedence over the investigation of things.

In the thought of Liu Zhongzhou (1578-1645), Zhu Xi’s painstaking effort to ritualize the body through reverence and Wang Yangming’s ‘direct and immediate’ approach to the self-illumination of the heart-and-mind reaches a new synthesis. Liu recommends on the one hand that we dwell in vigilant solitude in order to strengthen the roots of our intentions; on the other hand, he also recommends that this emphasis on the interiority of self-cultivation be supplemented by focused attention to the details of our daily behaviour. The authentic way of learning to be human is through examining and correcting mistakes. Self-cultivation combines the purification of an ever-deepening subjectivity with the rectification of all the mistakes made in ordinary practical living. This interplay between alertness of the subtle, incipient activation of the heart-and-mind and rigour in the ritualized discipline of the body offers a balanced programme for self-cultivation.

An elaborate project defining self-cultivation as a nine-stage effort to purify the mind was designed by the seventeenth century syncretist Lin Chao’en as an attempt to integrate the Three Teachings (Confucianism, Buddhism and Daoism) into one coherent idea of human flourishing. Lin enunciated his doctrine in a language richly textured by the symbolism of the *Yijing*. This claim that the microcosmic process in the physical body directed by the heart-and-mind actually emulates the macrocosmic process of the Heaven and Earth is characteristic of virtually all forms of self-cultivation in popular religion.

11 Modern significance

The primacy of experience, the centrality of exemplary teaching, and the necessity of personal commitment beyond reason may imply that, while such an approach is religiously meaningful, it is not philosophically significant. However, the pervasiveness of self-cultivation in all strata of Chinese society and all periods of Chinese history, clearly indicates that it is reasonable to characterize Chinese philosophy as disciplined reflections on the insights of self-cultivation. Even in Chinese Marxist literature, self-cultivation is assumed to be an integral part of the political ideology (for example, Liu Shaoqi’s essay on ‘How to be a Good Communist’) (see Marxism, Chinese). Feng Qi’s philosophy of wisdom, arguably one of the most original Chinese contributions to Marxism, focuses on the transformation of praxis into virtue-centred ethics. Self-cultivation provides the necessary background for Feng’s thought. Self-cultivation also features prominently in New Confucian Humanism. Indeed, a defining characteristic of New Confucians in the twentieth century is the prominence of the moral discourse on the learning of the heart-and-mind in their styles of philosophy.

Feng Youlan [Fung Yu-lan] (1895-1990), a major thinker under the influence of New Realism, characterized philosophy as a reflection on the human spirit, self-reflexivity and thinking about thinking. In a highly original philosophical move, he constructed his New Learning of the Principle around four cardinal concepts: principle, vital energy, substance of the *dao*, and the Great Whole. He also described Wang Yangming’s ‘primordial awareness’ (*liangzhi*) as a basic philosophical postulate. Feng’s use of propositional language in defining ‘primordial awareness’ as a postulate was severely criticized by Xiong Shili (1885-1968), one of the founders of New Confucian Humanism. Xiong insisted that ‘primordial awareness’ as an experience is a manifestation, indeed a realization of the human spirit. If it is defined merely as a postulate, the advantage of philosophical impartiality will be offset by its spiritual irrelevance. How could a mere postulate generate a profound sense of commitment to its full realization? On the surface, the difference between Feng and Xiong lies in the choice between an analytical reflection and an ethico-religious assertion. The former belongs to philosophy, whereas the latter smacks of religious or theological advocacy. At stake, however, is the role and function of self-cultivation in our definition of philosophy.

If philosophy is itself a spiritual quest and its purpose is self-knowledge, the disciplined reflection on insights ought to lead to wisdom necessary for human flourishing. ‘Primordial awareness’, the innate knowledge and innate ability that enables us to cultivate ourselves through our own effort, should not be affirmed merely as a postulate. Rather, it should be demonstrated as a presence and experienced as a reality. The danger of the propositional language is not only the misleading impression that the whole exercise is noncommittal but also the mistake in confusing the really real with an imagined possibility. We may detect a dogmatic streak in presenting

the ‘primordial awareness’ as a true presence and an experienced reality. However, the purpose is to show that philosophers must share their insights the way they experientially understand them to be. If ‘primordial awareness’ is what philosophers believe to be a genuine insight on the human condition (both Feng and Xiong, as Confucian thinkers, certainly did), then to them it must be more than a postulate.

On the surface, this may appear to be a conflict between different styles of philosophizing. Feng seems justified in having made ‘primordial awareness’ a mere postulate before he fully demonstrated its validity. After all, he could follow a series of assumptive reasons until he finally arrived at the conclusion that without ‘primordial awareness’ his whole philosophical enterprise would fall. Nevertheless, Xiong would contend that this style of philosophizing cheapens the seriousness of the spiritual exercise; it can be reduced to an inconsequential intellectual wordplay.

An underlying conviction of the New Confucian Humanists is that philosophy rooted in self-cultivation, far from being an academic discipline confined to a coterie of professionals in the ivory tower, is available to and necessary for all people. To paraphrase a statement in the *Daxue*, from the ruling minority and cultural elite to people in the street, all should regard self-cultivation as the root. If the root is not established, the Way cannot prevail; all the moral, social and political institutions instrumental for human flourishing (family, school, community, state and the world) depend upon self-cultivation for their stability, regularity, governance and peace. This personalist approach to morality, society and politics is based on the simple notion that the health of the whole is determined by the vitality of its constitutive parts. Ultimate human flourishing means that each and every member of the family, school, community, nation and the world is well-cultivated. Yet, to understand the significance of self-cultivation simply as a means to serve the ends of community well-being is one-sided. It is also an end in itself. Indeed, the justification for regulated families, functioning schools, ordered communities, governed states and a peaceful world is that they provide a wholesome environment for self-cultivation.

What is the relevance of self-cultivation philosophy to a complex modern society? A smooth functioning of the market, a vibrant democracy and the societal respect for the dignity of the person all depend upon the quality of the people involved. As we become increasingly self-interested, profit-oriented, litigious, competitive and individualistic, the need to develop a sense of public-mindedness, rightness, civility, trust and group spirit is compelling. The philosophy that enjoins us to elevate our common sense to good sense and transform the private ego into an open, dynamic and creatively transformative selfhood is vitally important for our life of the mind.

Perhaps the most enduring contribution of self-cultivation to the current style of philosophizing with emphasis on the analytical technique is to broaden the scope of rigorous thinking to include reflection on things at hand, meditation on ultimate concerns and examination of practical matters of daily living. Philosophy so conceived is no longer the privilege of a small coterie of trained professionals. Rather, it is a cultivated art accessible to all reflective, meditative and examining minds in the human community. While highly complex abstract reasoning may still characterize what teachers and researchers of philosophy do, the philosophical enterprise of self-knowledge is then an open invitation to all who cherish the value of the life of the mind. Surely, the degree of sophistication and refinement in self-cultivation varies from individual to individual and, like art, literature or music, the difference in proficiency and accomplishment between an initiating student and a virtuoso can be immense; but, to the extent that philosophy is viewed as self-cultivation, it is accessible to all members of the human community. While there is no need to delink philosophy from logic, epistemology and natural sciences, philosophy as self-cultivation or self-cultivation philosophy (they are of course distinguishable) must involve scholars in the humanities and social sciences. Indeed, for philosophy to regain the meaning and vitality as the ‘love of wisdom’ it once possessed, it must also transcend the boundaries of the ivory tower. Attention to self-cultivation is essential for such a task.

See also: Buddhist philosophy, Chinese; Chinese Classics;; Chinese philosophy Confucian philosophy, Chinese; Confucius; Dao; Daodejing; Daoist philosophy; Daxue; Mencius; Neo-Confucian philosophy; Xin (heart-and-mind); Xing; Yijing; Zhongyong

TU WEI-MING

References and further reading

Balazs, E. (1964) ‘Political Philosophy and Social Crisis’, in H.M. Wright and A.F. Wright (ed. and trans.)

Chinese Civilization and Bureaucracy, New Haven, CT: Yale University Press.(A collection of essays on Chinese institutions, history and thought.)

Chan Wing-tsit (ed. and trans.) (1963) *A Source Book of Chinese Philosophy*, Princeton, NJ: Princeton University Press.(A survey of original sources in translation.)

Confucius (551-479 BC) *Analects*, trans. D.C. Lau, *Confucius: The Analects*, Harmondsworth: Penguin, 1979.(An authoritative translation.)

Cua, A.S. (1982) *The Unity of Knowledge and Action: A Study in Wang Yang-ming's Moral Psychology*, Honolulu, HI: Unversity of Hawaii Press.(A study of Wang's doctrine in light of present inquiry concerning the relation between moral thought and action.)

Ivanhoe, P.J. (1993) *Confucian Moral Self Cultivation*, New York: Peter Lang.(A study of major Confucian thinkers that aims to provide a philosophically sensitive account of Confucian moral self-cultivation.)

Liang Congjie (1996) *The Great Thoughts of China*, New York: Wiley.(A source book on Chinese philosophy.)

Mencius (4th century BC) *Mengzi*, trans. D.C. Lau, *Mencius*, Harmondsworth: Penguin, 1970.(An authoritative translation.)

Tu Wei-ming (1979) *Humanity and Self-Cultivation*, Berkeley, CA: Asian Humanities Press.(A collection of articles on Confucian self-cultivation.)

Tu Wei-ming (1985) *Confucian Thought: Selfhood as Creative Transformation*, Albany, NY: State University of New York Press.(A collection of articles on Confucian thought.)

Tu Wei-ming (1989) *Centrality and Commonality: An Essay on Confucian Religiousness*, Albany, NY: State University of New York Press.(Although focused primarily on the *Zhongyong*, this monograph provides understanding of moral self-cultivation in light of Confucian religious thought.)

Xunzi (*c.*298-238 BC) *Xunzi*, trans. B. Watson, *Hsün Tzu: Basic Writings*, New York: Columbia University Press, 1966.(An authoritative translation of selected chapters.)

Self-deception

Philosophical work on self-deception revolves around a trio of questions. What is self-deception? Is self-deception possible? How are cases of self-deception to be explained? The extent to which self-deception is analogous to interpersonal deception is controversial, partly because certain analogies threaten to render the possibility of self-deception deeply problematic. The problems concern both the mental state of self-deceived individuals at a particular time (static problems) and the dynamics of self-deception (dynamic problems). For example, in normal interpersonal deception the deceivers know something, p, or at least believe truly that p, while getting their victims to believe the opposite, ~p. So if self-deception is strictly analogous to (normal) interpersonal deception, self-deceivers know or believe truly that p while getting themselves to believe that ~p. If this entails simultaneously believing that p and believing that ~p, self-deception may seem impossible. (For example, how can I believe that someone will read this entry while also believing that no one will read it?) Moreover, even if this state is possible, the suggestion that people can get themselves into it by deceiving themselves is problematic. It may seem, for instance, that any project describable as 'getting myself to believe what I now know to be false' is bound to be self-defeating. Self-deception may be dynamically impossible.

1 Intentional self-deception and partitioning strategies

Philosophers who deem self-deception conceptually or psychologically impossible typically contend that deceiving, by definition, is an intentional activity and that putative self-deceivers, therefore, must intentionally deceive themselves. In various time-lag scenarios, intentionally deceiving oneself clearly is possible. Paul, a pudgy prankster who has intentionally deceived others about their weight by secretly adjusting their bathroom scales, now intends to deceive himself into believing that he is lighter than he in fact is. Cognizant of his forgetfulness, the prankster adjusts his own scales for this purpose and counts on eventually forgetting that it was done. In this way, Paul may intentionally bring it about that he believes himself to be lighter than he is. In problematic cases, there is no such time lag. Rather, at some point during agents' intentional efforts to deceive themselves, they succeed in so doing. Some contend that such cases are impossible and that self-deception, therefore, is nonexistent. (These theorists suppose that self-deception is essentially intentional and involves no time lag.) Others defend the possibility of such cases. Yet others argue that intentional self-deception is remote from garden-variety self-deception and that even if the hard cases are impossible, that leaves the reality of self-deception intact.

Some philosophers propose to account for the possibility of self-deception by postulating a partitioned mind. Partitioning strategies range from postulating full-blown sub-agents with their own motives and intentions to employing modest distinctions among kinds or levels of awareness (see Modularity of mind; Cognitive architecture). In general, partitioners accept a strict interpersonal model of self-deception and seek to locate the requisite divisions within a single mind. A seriously divided mind may contain both a deceiving agent and a deceived subject. The deceiver (D) may believe that p while intentionally causing the subject (S) to believe that ~p. To the extent to which a divided mind is analogous to a pair (or group) of persons, static problems seem manageable. As you may believe that p while I believe that ~p, D may believe that p while S believes that ~p. Dynamic problems would be equally manageable. The basic idea of a representative dynamic problem has two major components. First, given that self-deceivers intentionally deceive themselves, they must intentionally follow a plan in so doing. Second, setting aside time-lag scenarios, any attempt at such plan-following would be self-defeating: just as I will not deceive you during a temporal span t, if you know what I am up to throughout t, you will not deceive yourself during t, if you know what you are up to throughout t. In a radically divided mind, S may have no idea what D is up to, and the problem vanishes.

Radical partitioners are often charged with substituting a species of 'other-deception' for self-deception. The criticism is that no human beings both deceive and are deceived by themselves on the radical view; rather, a sub-agent is the deceiver and something else is deceived. Further, the mental condition of radically partitioned persons strikes some theorists as disturbingly similar to multiple personality disorder. Surely, ordinary self-deceivers need not be so fragmented? Arguably, more modest proposals can account for self-deception.

Less radical partitioners eschew appealing to internal agents. Normally, however, their partitioning hypotheses are designed specifically to handle static problems, and dynamic counterparts remain. It has been proposed, for

example, that a person may simultaneously believe that p and believe that ~p, because the true belief may be possessed only unconsciously and therefore need not pre-empt the simultaneous holding of a contrary belief (see Unconscious Mental States). However, this does not answer the charge that intentional processes logically required for self-deception are necessarily self-defeating.

2 Departing from interpersonal models and related issues

Another approach to defending the possibility of self-deception is to motivate a departure from strict interpersonal models of the phenomenon. The assumption that deceiving is, by definition, an intentional activity drives many familiar puzzles about self-deception; and it is false. Using deceived in the passive voice, we properly say such things as ‘Unless I am deceived, I left my keys in my car’. Here ‘deceived’ means ‘mistaken’. In a corresponding use of ‘deceive’ in the active voice, to deceive is ‘to cause to believe what is false’; and not all instances of this are intentional. Your father accidentally misreads an encyclopedia article, taking it to report that p. He then tells you that p, intending to report something true but actually uttering something false. If you believe him, he has caused you to believe a falsehood: he has deceived you, in one legitimate sense of the word, but not intentionally.

This point has been exploited. When we know that people accidentally caused themselves to believe a falsehood (by innocently misreading an encyclopedia article, for example), we do not charge them with self-deception. However, when motivation enters the causal story in certain ways, a charge of self-deception seems warranted. Ordinarily, people who allegedly deceive themselves into believing that p want p to be true. For example, people who deceive themselves into believing that their spouses are not having extramarital affairs normally want it to be true that they are not so engaged, and parents who deceive themselves into believing that their children were erroneously convicted of a crime normally want it to be true that they are innocent. Owing significantly to the influence of relevant desires, Edna falsely believes that her husband, Ed, is not having an affair, in the face of strong evidence to the contrary. Is it likely that Edna first comes to believe that Ed is having an affair and then forms the intention to get herself to believe that he is not so engaged - which intention, perhaps after a bit of means/end reasoning, she successfully executes? Or is it more likely that, owing partly to her desire that Ed not be having an affair, Edna acquires or retains a false belief that he is innocent of infidelity without having come to believe that he is guilty and without intending to deceive herself?

A proper understanding of the dynamics of self-deception requires an appreciation of the ways in which motivation may bias reasoning and the formation, perseverance, and revision of beliefs. How might it happen that Edna believes that Ed is innocent of infidelity even though an impartial person, presented with the same evidence, reasonably concludes that he is guilty? Owing to her desire for Ed’s innocence, Edna may misinterpret data, reasoning, for instance, that if Ed were having an affair he would hide it and that his seemingly intimate meetings with his apparent lover in local taverns and cinemas consequently indicate that he is not sexually involved with her. Edna may even recruit Ed (in effect) in her motivated interpretive activities, by asking him for an explanation of the data, or by suggesting for his approval some acceptable hypothesis about his conduct. Further, again owing to her desire, Edna’s attention may be selectively attracted by evidence of Ed’s innocence. There is considerable empirical evidence that what we believe tends to be affected by the salience of our data; and Edna’s desire for Ed’s innocence may render corresponding data more salient than they would otherwise be. Edna may even set out to conduct an impartial investigation, but, owing to her desire, locate and focus on less accessible support for Ed’s innocence while overlooking more readily attainable support for his guilt.

All this can happen without Edna’s intending to deceive herself. Still, given the motivated way in which Edna’s pertinent belief is produced or sustained, even in the face of stronger evidence to the contrary, Edna is plausibly regarded as self-deceived. She is deceived as a consequence of her motivationally biased cognitive conduct.

The alleged requirement that self-deception be intentional is separable from another putative requirement - namely, that the person who is self-deceived in believing that p also knows or believes, at some level, that ~p. Some allege that self-deceived individuals characteristically behave in ways that clash with their sincere assertions and that this is explained by their knowing the truth at some level, which knowledge manifests itself in their overt behaviour. This is disputable. Edna, for instance, arguably is self-deceived even if she does not know or believe, at any level, that Ed is having an affair - even if her self-deception consists, in part, in a motivated failure to acquire the true belief that Ed is guilty of infidelity. Further, if Edna’s behaviour towards Ed has become cool, and if she has begun to enquire regularly of his whereabouts, this may be explained on the less demanding hypothesis that she suspects

that Ed may be having an affair. Edna may believe that Ed is faithful while also believing that there is a significant chance that she is wrong about this: ‘I believe that p, but I may be mistaken’ is utterly intelligible. In a thoroughly self-deceived person, beliefs of the latter sort presumably are absent. A self-deceived person - like a person deceived by someone else into believing that p - may be fully confident that p.

In garden-variety instances of self-deception, on this view, we do not deceive ourselves with the intention of so doing, or with the intention of causing ourselves to believe (or to cease believing) something. The absence of such an intention does not, however, preclude personal responsibility for garden-variety self-deception. Often, we are properly held responsible for unintended deceiving: for example, when we should have taken more care to get the facts straight, or to express them clearly. Sometimes, when we should have been more vigilant, we are justifiably held accountable for being deceived. This is so even when the deceiving and deceived persons are one and the same.

A thorough understanding of self-deception would clarify such related issues as the nature of belief, the influence of motivation on belief, and the structure of human minds. Certain controversial views about self-deception contravene the popular idea that belief is essentially exclusive, in the sense that a person’s believing that p is incompatible with the person’s simultaneously believing that ~p. Various discussions of the dynamics of self-deception illuminate the influence of motivation on belief. The literature also features a lively debate about the extent to which the human mind is unified and about the range of entities (sub-agents, for example) to which such mental items as beliefs, desires and intentions are plausibly attributed.

See also: Action; Self-deception, ethics of

ALFRED R. MELE

References and further reading

All items cited are comprehensible by advanced undergraduates.

Davidson, D. (1985) ‘Deception and Division’, in E. Lepore and B. McLaughlin (eds) *Actions and Events*, New York: Blackwell.(Advances a much-discussed partitioning account.)

Demos, R. (1960) ‘Lying to Oneself’, *Journal of Philosophy* 57: 588-95.(Seminal paper advancing modest partitioning.)

Fingarette, H. (1969) *Self-Deception*, London: Routledge & Kegan Paul.(Emphasizes the dynamic nature of self-deception and develops an action-oriented response to familiar worries about the possibility of self-deception.)

Haight, M. (1980) *A Study of Self-Deception*, Sussex: Harvester Wheatsheaf.(Defends the impossibility of self-deception.)

Lockard, J. and Paulhus, D. (eds) (1988) *Self-Deception: An Adaptive Mechanism?*, Englewood Cliffs, NJ: Prentice-Hall.(Useful collection of literature in psychology and related fields.)

McLaughlin, B. and Rorty, A. (eds) (1988) *Perspectives on Self-Deception*, Berkeley, CA: University of California Press.(Useful collection; extensive bibliography.)

Martin, M. (ed.) (1985) *Self-Deception and Self-Understanding*, Lawrence, KS: University of Kansas Press. (Useful collection of philosophical and psychological literature; extensive bibliography.)

Mele, A. (1987) *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*, New York: Oxford University Press.(Criticizes partitioning and advances an alternative resolution of the static and dynamic problems.)

Pears, D. (1984) *Motivated Irrationality*, Oxford: Oxford University Press.(Advocates radical partitioning.)

Sartre, J-P. (1943) *L’être et le néant*, Paris: Gallimard; trans. H. Barnes, *Being and Nothingness*, New York: Washington Square Press, 1956.(Includes an influential discussion of bad faith, a phenomenon sometimes identified with self-deception.)

Self-deception, ethics of

Self-deception is complicated and perplexing because it concerns all major aspects of human nature, including consciousness, rationality, motivation, freedom, happiness, and value commitments. In a wide sense, 'self-deception' refers to intentional activities and motivated processes of avoiding unpleasant truths or topics and the resulting mental states of ignorance, false belief, unwarranted attitudes, and inappropriate emotions. Deceiving oneself, like deceiving other people, raises a host of questions about immorality. These include whether self-deception is always immoral or only when it conceals and supports wrongdoing; whether self-deception about wrongdoing and character faults compounds or mitigates guilt for causing harm; how important the value of authenticity is, and whether it can be sacrificed in an attempt to cope with reality; what the relation is between self-deception and responsibility; and whether groups can be self-deceived. Ultimately, the moral status of any instance of self-deception depends on the particular facts of the case.

1 Explaining wrongdoing

The activities and the mental states involved in self-deception - especially the suggestion that we can wilfully get ourselves to believe the opposite of what we know is true - seem paradoxical in ways that have generated a large literature in epistemology and philosophy of mind (see Self-deception). That literature is germane to the moral issues, especially as it serves to clarify the extent of voluntariness in deceiving ourselves.

Self-deception is objectionable when it camouflages and supports greed, cruelty and other forms of wrongdoing. Moreover, often the first step towards self-reform is to remove self-deception about our character faults. And we would make fairer appraisals of other people if we were less self-deceiving about condemning them and excusing ourselves. Or so traditional ethicists have contended.

In his discussion of this in two early lectures (1722), Joseph Butler focused on two contexts. First, when moral rules are vague we tend to interpret them in ways convenient to ourselves. For example, we employ legalistic ruses in narrowly interpreting duties to help others and prohibitions against corruption. Second, we excuse our overt violations of clear duties so as to make them 'sit a little easy' in conscience. In both contexts, our motives might be general selfishness or specific passions. Our tactics include rationalization, wilful ignorance, and systematic ignoring. Having (and using) suspicions and partial knowledge of our faults, we avoid further inquiry, closing 'the eyes of the mind' as easily as those of the body. Because the avoidance is intentional, Butler concludes that self-deception compounds rather than lessens guilt for harming others and for corrupting our own consciences. It makes us hypocrites before others and also 'inner hypocrites' who culpably pretend to ourselves that we are better persons than we are.

Kant (1797) went further in suggesting that self-deception about virtually any topic - not just immorality - can indirectly cause violations of duties to others, either by making us ignorant of morally-relevant information or by encouraging habits of self-deception. Kant's view is implausible regarding many examples but it has some plausibility regarding general habits of engaging in self-deceit.

David Pears (1984) and Alfred R. Mele (1987), among others, have explored parallels and interactions between self-deception and moral weakness (see Akrasia). Self-deception explains many instances of acting against our better judgment by revealing how we blur that judgment or ignore morally relevant information. Even when moral weakness is fully conscious, self-deception causes a loss of moral perspective that might have prevented it. Conversely, some self-deception may itself qualify as moral weakness in so far as we voluntarily violate our standards of morality and rationality.

2 Authentic values

No philosopher was harsher in condemning self-deception ('bad faith') than Jean-Paul Sartre (§3). His condemnations were made despite and partly because of his renunciation of traditional beliefs in objectively justifiable values (see Existentialist ethics §3).

Sartre (1943) defined authenticity as fully acknowledging our complete freedom to choose values and to interpret the world in light of them (see Existentialist ethics §§1-2). By definition, authenticity is incompatible with

self-deceptive excuses and evasions. The worst self-deception is the direct denial of freedom of consciousness because it leads to violating other persons' freedom of conduct, for example through anti-Semitism or political oppression. Yet self-deception about all topics, not just about freedom in general, implies a cowardly refusal to acknowledge our freely chosen mental states and actions. Tactics of self-deception include quick oscillations of attention towards pleasing aspects of our lives and away from anxiety-producing ones. For example, we affirm our freedom ('transcendence') in so far as it promises happiness, but we appeal to hard facts ('facticity') whenever it assuages the anxieties in admitting responsibility for 'authoring' our choices. The fundamental strategy ('original project') in self-deception is to distort the standards of rationality for belief by exaggerating favourable evidence for what we want to believe, disregarding contrary evidence, and resting content with minimal evidence for pleasing beliefs. Sartre's discussion is often insightful, but he conflates key distinctions such as those between free consciousness and free conduct and between causal responsibility and moral accountability.

Nietzsche (1888) offered a more nuanced critique of self-deception than Sartre (see Nietzsche, F. §9). Certainly he could be equally harsh in condemning traditional beliefs in the objectivity of moral values, with the possible exceptions of honesty and courage. Perhaps inconsistently, however, he praised the need for some self-deception in asserting our 'will to power' in developing and living according to authentic value-perspectives.

A more traditional approach to authentic values and personal identity retains the traditional faith in objective moral values by reference to which authenticity is understood. This approach does not entail dogmatism and leaves room for a pluralistic conception of diverse forms of goodness from which we choose in creating meaningful lives (see Moral pluralism). On this view, the authentic self is honestly responsive to general moral requirements, attuned to the particular goods we care most deeply about, and courageous in dealing with ineluctable realities such as death.

3 Healthy coping

Despite claims of neutrality, psychologists and social scientists who discuss self-deception frequently presuppose evaluative perspectives. Those who focus on mental illness emphasize the harmful aspects of self-deceiving 'defence mechanisms'; those who focus on healthy behaviour stress the value of self-deception as an adjustment mechanism.

Although Sigmund Freud (1920) rarely used the term 'self-deception', many of his followers interpret defence mechanisms (such as repression, projection, and sublimation) as strategies of self-deception. Using these strategies, the ego (the coping part of the mind) defends itself against onslaughts from the id (unconscious sexual and aggressive desires), the superego (socially instilled values), and external reality. In his early work Freud portrayed the ego as entirely conscious, but later he emphasized that an unconscious part of the ego engages in defence.

Cognitive psychologists have explored the positive roles of self-deception in promoting mental wellbeing, especially in maintaining healthy self-esteem by exaggerating our virtues, supporting optimism by playing down risks and creating illusions of control, and coping with traumatic events and threatening uncertainties. Similarly, sociobiologists hypothesize that we evolved as creatures who compete more successfully when we can deceive ourselves into believing the falsehoods we perpetrate on others (in order not to betray the deception with behavioural signals). These scientists provide needed correctives to one-sided condemnations of self-deception by Kant and Sartre.

4 Ambiguous responsibility

Responsibility for self-deception is obscure and ambiguous because the notion suggests a self divided against itself, rather than a unified agent, accountable for its voluntary choices.

In *Self-Deception* (1969), Herbert Fingarette linked self-deception to wider issues about personal identity and divided selves. In his view, self-deceivers pursue 'engagements' (ways of participating in the world) that they disavow. Disavowal implies a refusal to become explicitly conscious of engagements and also of the activities involved in concealing them. More fundamentally, disavowal constitutes a refusal to integrate disavowed engagements into the network of other engagements that comprise the dominant aspects of one's personality. Isolated and without rational supervision, disavowed engagements can wreak havoc on oneself and others. We are responsible for initiating disavowal, but responsibility becomes ambiguous in so far as we lose control over

disavowed engagements.

Fingarette has been criticized for being too sympathetic to psychoanalysis and insufficiently rooted in more empirical psychology (although it should be noted that his book inspired many subsequent empirical studies of self-deception). It has also been noted that he unhelpfully plays down cognitive terms like ‘belief’ and ‘knowledge’ as he shifts to action-oriented terms like ‘avow’ and ‘disavow’. Finally, he has been accused of overgeneralizing about the moral ambiguities of (all) self-deception, rather than attending to the moral variety of particular cases.

5 Collective self-deception

Collective self-deception is self-deception by entire groups who share false beliefs and unwarranted attitudes. Charges of collective self-deception are both provocative and contentious. Thus, Marxists portray workers who accept capitalism as suffering from ‘false consciousness’, that is, illusions about the economic origins and political implications of economic ‘ideologies’ (see Marx, K. §10; Ideology §1). Some feminists portray satisfied homemakers as having ‘internalized’ patriarchal ideologies. Religious and nonreligious thinkers criticize each other’s perspectives as self-deceiving. And in *1984* (1949), George Orwell suggested that political oppression is sustained by fostering widespread ‘doublethink’, a process he described in terms reminiscent of Sartre’s description of bad faith.

Some charges of collective self-deception are illuminating, although troublesome in thinking about personal and collective responsibility (see Responsibility). In so far as my self-deception is pressured or encouraged by strong social forces, am I less blameworthy because my autonomy is lessened, or is my responsibility compounded because I allowed myself to be a passive conspirator? (In *Invisible Man* (1947), Ralph Ellison suggests that both may be true, in different respects.) Nevertheless, allegations of collective self-deception are easily abused and tend to degrade the concept of self-deception into a rhetorical weapon that undermines tolerance. Ascriptions of self-deception about morality are rooted most firmly where values are clear and well-justified; they are best avoided in contexts where morally reasonable persons interpret and balance values in alternative ways.

6 Conclusion

Self-deception is morally complex for two main reasons. First, in evaluating particular instances and in ascribing responsibility for their consequences, attention must be paid to an array of contextual factors: what the self-deception is about, what motivates it, which obligations it threatens, the degree of harm it causes to oneself and to others, the vital needs it may serve, how much control an individual has in initiating and sustaining it, what help is available to overcome it, and whether it is episodic or habitual. When motives are good and consequences are desirable, self-deception may be excusable, forgivable, desirable, or simply comic, depending on the circumstances. When motives and consequences are evil, as they were in the self-deception that Albert Speer ascribes to himself in his memoirs, *Erinnerungen (Inside the Third Reich)* (1969), self-deception becomes a form of cruelty and callous moral negligence.

Second, different moral emphases can be brought to bear in understanding self-deception. We may agree that honesty (with oneself and others) is an important virtue, but exactly how important relative to other virtues is open to dispute and reasonable disagreement (see Truthfulness). The insistence that honesty is always paramount and forbids all self-deception can result in inhumane neglect of other important values such as love and self-respect. Indeed, idealistic demands for perfect honesty can themselves be permeated with self-deceptive self-righteousness, as playwrights remind us: see Molière’s *Le Misanthrope (The Misanthrope)* (1666), Henrik Ibsen’s *Vildanden (The Wild Duck)* (1884), and Eugene O’Neill’s *The Iceman Cometh* (1957). It is equally clear, however, that love and self-respect are themselves undermined when habitual self-deception distorts caring for others and for oneself.

MIKE W. MARTIN

References and further reading

Ames, R.T. and Dissanayake, W. (eds) (1996) *Self and Deception*, Albany, NY: State University of New York. (Essays contrasting Western and non-Western views of self-deception.)

Butler, J. (1722) ‘Upon Self-Deceit’ and ‘Upon the Character of Balaam’, in *The Works of Bishop Butler*, Oxford: Oxford University Press, 1896.(Insightful sermons that focus on the corruption of conscience.)

Dilman, Ï. and Phillips, D.Z. (1971) *Sense and Delusion*, New York: Humanities Press.(Argues that self-deception undermines meaningful life.)
Ellison, R. (1947) *Invisible Man*, New York: Vintage Books.(A novel about a young man's search for authenticity in a racist society.)
Fingarette, H. (1969) *Self-Deception*, London: Humanities Press.(Focuses on personal identity, integrating ideas of Sartre, Kierkegaard and Freud.)
Freud, S. (1920) *A General Introduction to Psychoanalysis*, New York: Liveright.(Readable and lively overview of the role of unconscious motives.)
Haight, M.R. (1980) *A Study of Self-Deception*, London: Humanities Press.(Explores paradoxes of freedom, determinism, and the divided self.)
Ibsen, H. (1884) *Vildanden*, trans. D.B. Christiani, *The Wild Duck*, New York: W.W. Norton, 1968.(A play unmasking dangerous illusions about human pretensions.)
Kant, I. (1797) *Metaphysische Anfangsgründe der Tugendlehre*, trans. M.J. Gregor, *The Doctrine of Virtue: Part II of The Metaphysics of Morals*, Philadelphia, PA: University of Pennsylvania Press, 1964.(Early reflections on the divided self.)
Lockard, J.S. and Paulhus, D.L. (eds) (1988) *Self-Deception: An Adaptive Mechanism?*, Englewood Cliffs, NJ: Prentice Hall.(Psychological studies of self-deception as a coping device.)
Martin, M.W. (1986) *Self-Deception and Morality*, Lawrence, KS: University Press of Kansas.(Discussion of four moral perspectives: inner hypocrisy, authenticity, moral ambiguity, and vital lies.)
Martin, M.W. (ed.) (1985) *Self-Deception and Self-Understanding*, Lawrence, KS: University Press of Kansas. (Philosophical and psychological essays on ethics, paradoxes, and mental structures of self-deception.)
Mele, A.R. (1987) *Irrationality: An Essay on Akrasia, Self-Deception and Self-Control*, Oxford: Oxford University Press.(Explores interconnections between weakness of will and self-deception.)
Molière, J-B.P. (1666) *Le Misanthrope*, trans. R. Wilbur, *The Misanthrope*, New York: Harcourt Brace, 1954. (Comic exploration of the idea that honesty threatens personal relationships unless constrained by common decency.)
Nietzsche, F. (1888) *Ecce Homo*, trans. W. Kaufmann and R. Hollingdale in *Basic Writings of Nietzsche*, New York: Modern Library, 1968.(Wide-ranging epigrams dealing with knowledge and values.)
O'Neill, E. (1957) *The Iceman Cometh*, New York: Vintage Books.(Drama that exposes the violence in personal relationships camouflaged by 'pipe dreams'.)
Orwell, G. (1949) *1984*, New York: New American Library, 1961.(Nightmarish vision of a totalitarian government which uses mind control centred around 'doublethink'.)
Pears, D. (1984) *Motivated Irrationality*, Oxford: Oxford University Press.(Studies the parallels between self-deception and other forms of irrationality.)
Pines, C.L. (1993) *Ideology and False Consciousness*, Albany, NY: State University of New York Press. (Introduction to Marxist conceptions of collective self-deception.)
Sartre, J-P. (1943) *L'Être et le Néant. Essai d'ontologie phénoménologique*, Paris: Gallimard; trans. H.E. Barnes, *Being and Nothingness: An Essay of Phenomenological Ontology*, New York: Washington Square, 1966, 86-116.(Famous chapter on 'bad faith' that sparked extensive philosophical interest in self-deceit.)
Speer, A. (1969) *Erinnerungen*, trans. R. Winston and C. Winston, *Inside the Third Reich*, New York: Avon, 1970.(Memoirs of Hitler's leading bureaucrat, who explains his cruelty by ascribing to himself self-deception.)

Self-realization

'Self-realization' is the development and expression of characteristic attributes and potentials in a fashion which comprehensively discloses their subject's real nature. Usually, the 'self' in question is the individual person, but the concept has also been applied to corporate bodies held to possess a unitary identity.

What constitutes the self's 'real nature' is the key variable generating the many conceptions of self-realization. These can be grouped broadly into two types: (1) the 'collectivist', in which the self-realizing lifestyle, being either the same for all or specific to a person or subgroup of people, is ultimately definable only in the context, and perhaps with reference to the common purposes, of a collective social body; (2) the 'individualist', in which a person's self-realization has no necessary connection with the ends of a particular community.

As an ethic, self-realization can be proposed as the means to achieve a life identified as good by some criterion independent of the self-realizing process, or held to be that which actually defines the good. Its critics typically argue that human nature is such that any equation of 'self-realization' and 'goodness' is implausible or undesirable.

1 Collectivist self-realization

Aristotle's moral and political work, especially *Nicomachean Ethics*, is an enormously influential example of a self-realizationist theory, which holds that the development and expression of people's potentials is the goal for the best moral - and, for Aristotle, political - life (see Aristotle §§20-6). Using 'self-realization' as much to describe the good society as the good person, he proceeds from the belief that humans naturally possess a unique, essential state of being - the life of reason-guided activity - to argue that achievement of this state is the goal (*telos*) towards which normal human development aims (see Human nature §1). Subsequently, he identifies those personal dispositions and social institutions necessary for securing this end.

Aristotelian self-realization is collectivist not only in the sense of its essence being a shared characteristic but also in its requirement that it be achieved within a very particular form of social life: the polis or city-state. Theoretically, it permits some diversity across the range of self-realizing lives. Other versions of collectivist self-realization can be much stricter in detailing exactly what form of life is required, tying the 'real nature' of the self to specific social roles or the performance of onerous moral duties. Often, these doctrines adopt a version of positive liberty by characterizing self-realization as 'true freedom', which facilitates their descriptions of non-self-realizing lifestyles as not 'really' wanted by people despite any 'mistaken' preferences to the contrary (see Freedom and liberty).

The sustainability of collectivist conceptions in Western thought was weakened significantly (particularly from the seventeenth century onwards) with the emergence of more atomized, individualistic notions of the self, generated not least by a growing appreciation of human diversity. This distinctively modern form of identity self-consciously sought individuation, conceptualizing the self and its ends apart from any particular, metaphysically independent moral order. Increasingly, people were seen as empowered to direct their own lives according to their own desires. The postulation of substantive shared natures and comprehensive collective purposes, and the equation of their realization with human freedom and morality, became much less credible. Indeed, for some, such conceptions were downright pernicious in their attempt to attribute an artificial uniformity to human nature.

That some versions of collectivist self-realization remained vibrant in the nineteenth and twentieth centuries was due largely to Hegel and his intellectual heirs, but the transformation in the concept of the individual left its mark on them (see Hegelianism). Reflecting high Enlightenment optimism over the goodness of human potential and the continuing reconcilability of freedom, morality and social order, their theories held that self-realization involved the expression of full individuality and a unification of every self with each other in a qualitatively new, yet genuinely communal, form of social life (see Enlightenment, Continental). Marxism became the most influential exponent of this approach, viewing self-realization as based upon the capacity for free creative labour supposedly shared by all and insisting that, under full communism, the free self-realization of each would ultimately be dependent upon the free self-realization of all (see Marx. K.).

2 Individualist self-realization

Romanticism was the prime influence in bequeathing conceptions of self-realization to the individualized notions of selfhood, prompting a view of the self as invested with a capacity to develop - freely and creatively - a naturally given potential to lead a morally worthy, aesthetically stimulating life (see Romanticism, German). This was distinguishable from the collectivist variant in that the self-realizing lifestyle was held to be unique to each individual and no longer required reference to the purposes of a larger whole for its full meaning. Many individualistic conceptions nevertheless stressed self-realization's dependence upon specific forms of community as necessary means for its achievement, retaining the concept's potential as critic of degenerate societies and inspiration for social and political reform. The emphasis had nevertheless switched to the individual's uniqueness as the end in itself and in some other versions, notably that of the later Rousseau, modern society is seen as so irredeemably corrupt as to require virtual abandonment by the self seeking realization (see Rousseau, J.-J.).

The 'individualistic' label should not obscure the fact that these conceptions generally reaffirm the idea that there are certain features common to everyone's self-realization: similar kinds of potential, for example. This preserves the viability of 'self-realization' as something identifiably separate from other ways of living. Yet the same sensitivity to human diversity which undermined the collectivists' credibility poses serious challenges here, too. One may doubt that even the relatively minimal shared features specified by these conceptions really exhaust all cases of what self-realization might legitimately involve. One may believe there is no warrant for characterizing self-realization as anything other than however individuals would like to live, in which case the need for - perhaps even the very intelligibility of - the concept as distinct from, say, mere 'autonomy' must be questioned (see Autonomy, ethical).

Alternatively, if self-realization is still defined in sufficiently substantive terms to avoid a content-draining equation of it with 'whatever one chooses to be or do', but one also desires to respect individuals' autonomy (as individualistic conceptions usually profess to do), self-realization very readily reduces simply to one lifestyle option among many which might be chosen by free people. Though this need not force a denial that it is the best form of life, preference for self-realization over respect for autonomy as a social and political ethic consequently seems to lose justification.

3 Further problems

The concept's detractors may also be critical of its presentation as an ethical standard, in which self-realization is defined by some independent specification of goodness or is deemed to be what actually yields the criterion of goodness itself (see Perfectionism). Arguing that humanity's behaviour reveals nothing good in people's potentials, they may contend that self-realization is far from the best route to goodness and patently unsatisfactory as a defining characteristic of the good. Alternatively they may urge that, even if it is not intrinsically bad, self-realization creates a 'culture of narcissism', encouraging those forms of selfhood which, while perhaps harmless enough in themselves, nevertheless fail to show respect for others and threaten to enervate social life itself.

Critics also claim that the self-realization of people with morally positive natures might still entail struggle, frustration and disillusionment, which could be worse (in hedonistic or any other terms) than the outcome of a non-self-realizing life. Equating 'good' directly with 'self-realization' theoretically dissolves such worries ('whatever pain it may cause, self-realization remains the best by definition'). Yet this move seems eminently resistible if we consider the gamut of human potentials to be profoundly ambivalent in character. There are both good and bad potentials, with some people possessed of a surfeit of one and most, perhaps, laden with a mixture of both. To yield this plausible discrimination, we clearly need to conceptualize 'good' independently of 'self-realization'. However, the same ambivalence militates against the thought that self-realization is the best way to achieve an independently defined good: the latter might require the restraint of at least part of one's nature.

Furthermore, the very concept of the self as possessed of a given identity capable of realization has been challenged in postmodernist thinking (see Subject, postmodern critique of the). While many conceptions of self-realization readily accept that the 'givenness' of the self's nature does not preclude its fluidity and revisability, postmodernism wants to resist any such fixity and, accordingly, prefers putatively distinct notions of self-*creation*.

Yet 'self-realization' can be revised to insist that the highest realizable values must spring from part of human nature, urging that we learn to identify with that part which promises the most fulfilling and sustainable form of

personal and social life, and fashion our lifestyles accordingly. Green moral and political thought has utilized the concept thus, interestingly de-emphasizing its usual ‘activist’ insistence upon personal striving in a reconciliation of the self with nature which is more reminiscent of the spiritual, contemplative self-realizationist ethics in Eastern philosophy (see Environmental ethics §2; Buddhist philosophy, Indian §1).

If it can, then, be rendered a distinctive and attractive concept, ‘self-realization’ might plausibly continue to be a central concept in critiques of those ways of life and forms of society which leave people unfulfilled and alienated.

See also: Daodejing; Happiness; Japanese philosophy §§8-9;Life, meaning of; Nirvāṇa; Self-cultivation in Chinese philosophy; Self-deception, ethics of

MARK EVANS

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985.(Arguably the single most influential text in Western self-realizationist thinking.)

Berlin, I. (1969) ‘Two Concepts of Liberty’, in *Four Essays on Liberty*, Oxford: Oxford University Press. (Describes self-realization as a version of positive freedom, which it attacks for its totalitarian potential.)

Elster, J. (1986) ‘Self-Realization in Work and Politics: The Marxist Conception of the Good Life’, in E.F. Paul, F.D. Miller, Jr, J. Paul and J. Ahrens (eds) *Marxism and Liberalism*, Oxford: Blackwell.(A clear exposition and defence of a self-realizationist ethic loosely inspired by Marx.)

Lasch, C. (1978) *The Culture of Narcissism*, New York: W.W. Norton.(Attacks what it considers to be a pervasive and overly indulgent, narrow-minded obsession with oneself which self-realization has been accused of promoting.)

Naess, A. (1995) ‘Self-Realization: An Ecological Approach to Being in the World’, in G. Sessions (ed.) *Deep Ecology for the 21st Century*, Boston, MA: Shambala Publications.(Representative of the contemporary Green interest in conceptions of self-realization.)

Taylor, C. (1979) *Hegel and Modern Society*, Cambridge: Cambridge University Press. (A compact discussion not only of the Hegelian but also the Romantic and Marxist versions of self-realizationist theory.)

Taylor, C. (1993) *The Ethics of Authenticity*, Cambridge, MA: Harvard University Press. (Using ‘authenticity’ synonymously with ‘self-realization’, an ideal introduction to a contemporary moral and political theory based upon the concept.)

Self-respect

Persons are said to respect themselves when they have an appropriate sense of their own worth either as persons generally or as individuals occupying particular roles. In respecting themselves as persons, people may recognize their worth as persons ('recognition' self-respect) or value the positive aspects of their character ('evaluative' or 'appraisal' self-respect). On the 'subjective' view of self-respect, persons have self-respect so long as they value themselves according to their own standards of worthiness. On the 'objective' view, there are certain attitudes or actions that show self-respect, regardless of whether the agents exhibiting them are conforming to their own standards. Self-respect plays a central role in the ethical philosophy of Kant and the political philosophy of Rawls. Kant maintains that persons have a duty to respect themselves, which consists in regarding themselves as equal in moral status to other persons. Rawls holds that a just society must provide the social bases of self-respect for all citizens, for without self-respect, their lives are severely diminished. In this same spirit, some critiques of oppression emphasize the injustice of robbing people of their self-respect, which is often a consequence, it is claimed, of being oppressed.

1 The concept of self-respect

Self-respect is a multifaceted notion involving a constellation of attitudes, beliefs, desires, dispositions, commitments, expectations, actions and emotions that express or constitute one's sense of one's worth. It includes a recognition and understanding of one's worth, as well as a desire and disposition to protect and preserve it. One can respect oneself as a person, as when one stands up for one's rights, or in a specific role or capacity, as when one adopts the code of ethics associated with one's work or profession. One may also merit one's own respect, as a person or as the occupier of a role, by conducting oneself honourably or by living up to one's standards. Since philosophers tend to be concerned with respect for oneself as a person, 'self-respect' will henceforth be assumed to have persons as such as its object.

Most writers on self-respect agree that it can be divided into two kinds, according to its appropriate grounds. Robin Dillon (1992a) calls these two kinds 'recognition' self-respect and 'evaluative' self-respect. The latter kind is also widely referred to as 'appraisal' self-respect (see Respect for persons §1). Recognition self-respect is a type of regard towards oneself that all persons are obliged or entitled to acquire. Unworthy attitudes or conduct, such as servility or needless self-deprecation, are viewed as evidence of, or as constitutive of, its absence. The label 'recognition self-respect' derives from the fact that the attitude to which it refers consists primarily in recognizing that one is a person and weighing appropriately this fact in deliberation and action. Such recognition is not to be understood as mere acknowledgement of the fact that one is a person. It also includes appreciating the special status one has as a person and that this status may require certain kinds of responses, conduct, or restraint. One is obliged or entitled to respect oneself in virtue of the fact that one is a person.

The concept of recognition self-respect relies heavily on a normative conception of the person. Moreover, one must have some sense of what that conception entails to understand what would count as respecting or showing disrespect to persons. It follows that we can infer much about philosophers' understanding of the nature and status of persons from their accounts of the content of recognition self-respect. The notion of the person that is presupposed by most contemporary discussions of recognition self-respect is either strictly Kantian, or Kantian in spirit (see Respect for persons §2). On this view, persons are essentially autonomous rational agents who possess a special worth grounded in their capacity for moral agency and who enjoy a status of moral equality with other persons on the basis of this special worth (see Autonomy, ethical). Recognition respect for oneself as a rational autonomous agent involves acknowledging and appreciating one's autonomy, one's rationality, the value one has in virtue of having these capacities, and the moral status of equality that is grounded in this shared value (see Kantian ethics).

In contrast with recognition self-respect, which all persons are required or entitled to secure, evaluative self-respect is a kind of self-respect that individuals may or may not merit and may merit in varying degrees. Where the presence of recognition self-respect prompts us to engage in worthy conduct or refrain from engaging in unworthy conduct, the presence of evaluative self-respect results from an assessment of conduct already undertaken. Because persons must earn evaluative self-respect by conforming their actions, emotional responses,

attitudes and so on to certain standards of worthiness, not everyone is entitled to evaluative self-respect, and some people's evaluative self-respect is unwarranted. Likewise, some people's lack of evaluative self-respect is also unwarranted. The normative ground of evaluative respect is the quality of one's moral character and those actions that reflect it. One merits evaluative self-respect, in other words, not in virtue of *being* a person, but because of the virtue *of* one's person.

Self-respect is distinct from self-esteem, though the latter notion bears some resemblance to evaluative self-respect. Both evaluative self-respect and self-esteem involve having a favourable opinion of oneself that arises from an assessment of one's traits or actions. Moreover, both attitudes can be undeserved as well as deserved but absent. None the less these two notions differ in their respective normative grounds. A favourable regard for oneself based upon features that do not bear directly upon one's moral character, such as one's sense of humour or one's having won a chess match, is self-esteem. Evaluative self-respect involves the positive regard one has for oneself based upon one's moral character. Although this method of distinguishing self-respect from self-esteem is by no means uncontroversial, it has the advantage of allowing us to distinguish clearly the positive attitudes towards oneself that are directly relevant to morality and those that are less so. Evaluative self-respect, since it involves one's moral character, is certainly an appropriate subject for moral theorizing. Self-esteem, in contrast, if it is relevant to morality, is so indirectly.

On the preceding account, recognition self-respect was described as an objective notion. That is, it was assumed that there are certain attitudes, dispositions and responses that a person *must* have in order to be adequately self-respecting. Those who propound an objective view would probably claim, for instance, that persons who habitually ingratiate themselves with others for personal gain are lacking in recognition self-respect. According to the subjective view, on the other hand, so long as such persons do not *themselves* regard ingratiation for personal gain as objectionable or 'beneath them', their self-respect is intact. It is compromised only if they behave in ways that *they* judge to be inappropriate. When moral philosophers claim that all persons ought to have recognition respect for themselves, they are generally treating self-respect as an objective concept. Their claim expresses the idea that all persons, regardless of those persons' own views of what is fitting, ought to value their rights, assert their needs when appropriate, formulate and commit themselves to ideals and goals worthy of them as persons, abstain from self-destructive behaviour, avoid subordinating themselves to others unnecessarily and take responsibility for their conduct.

Evaluative self-respect can also be viewed either as an objective or as a subjective notion. Regarded as an objective concept, evaluative self-respect is judged to be warranted only if the standards by which one has appraised oneself positively are morally worthy. As a subjective concept, evaluative self-respect is seen to be warranted so long as one's positive self-assessment is based upon standards that one deems worthy of oneself.

2 Self-respect in moral and political philosophy

Though many philosophers have remarked upon self-respect, or related notions such as pride, integrity and dignity, self-respect figures most prominently in the work of Kant and Rawls. In *Die Metaphysik der Sitten (Metaphysics of Morals)* (1797), Kant claims that persons have a duty to respect themselves. Self-respect, on his view, is a valuing stance that one wilfully adopts or neglects to adopt towards oneself. It consists in an understanding and appreciation of the equal moral status one shares with other persons and a disposition to act in ways that express this understanding and appreciation. The requirement that persons regard and treat themselves as the moral equals of others is grounded in Kant's contention that all rational beings have a special moral value, which he calls 'dignity', that is itself rooted in our capacity for autonomous rational agency. Since every person as such has dignity, each person as such is equal in basic moral status to every other person and ought to appreciate this fact and act accordingly.

Our special moral worth, Kant claims in *Grundlegung zur Metaphysik der Sitten (Groundwork of the Metaphysics of Morals)* (1785), requires that all persons be treated and treat themselves as ends in themselves and never merely as a means to another's ends. In other words, the capacity of rational agents to set ends freely for themselves should be treasured and respected by all (see Kant, I. §§9-11). This duty, which is one formulation of the categorical imperative, has been interpreted by some philosophers as a duty to respect others and to respect oneself. For Kant, the requirement that we respect ourselves consists not only in a duty to regard ourselves as having the same moral standing as others, but also in an array of duties requiring us to value our autonomy and

rational agency and to treat ourselves in accordance with our special moral status.

In contrast to Kant, John Rawls views self-respect as an entitlement rather than a duty. In *A Theory of Justice* (1971), he maintains that justice requires that self-respect, or more precisely its social bases, be provided for all members of a society. Moreover, he sees self-respect not as an attitude that individuals may wilfully adopt, but as a social good that can be secured by individuals only when certain social and political conditions prevail. He defines self-respect as a conviction that one's plan of life is worth pursuing accompanied by the belief that one is well-suited to pursue it. The value of self-respect, Rawls claims, lies in the fact that no one can carry out or achieve their aims adequately without it.

For Rawls, civil equality is the key to citizens' acquiring and maintaining adequate self-respect. In the absence of civil equality, those with fewer rights and privileges will regard their life plans as less worthwhile than the life plans of citizens with more rights and privileges, for, by depriving certain citizens of goods essential to the effective pursuit of their ends, a society implies that the ends of those so deprived have a lesser value. Since self-respect is fundamental to the basic wellbeing of all persons, it follows that principles prescribing the design of basic social structures are just only if they foster or do not hinder civil equality. The two principles of justice proposed by Rawls, the equal liberty principle and the 'difference principle', are designed to ensure civil equality and hence help guarantee self-respect for all citizens.

Influenced by Rawls' realization that self-respect is largely a product of one's social and political circumstances, some political theorists have employed the notion of self-respect in their critiques of the oppression of white women, people of colour, and gays and lesbians. These writers point to the damage done to the self-respect of members of groups that are marginalized, stigmatized or exploited by the dominant culture. Oppressed people may internalize the disparaging images and views of them produced by the dominant culture and hence come to see themselves as inferior or as undeserving of equal treatment. Some writers discuss the ways in which both individual and collective resistance to oppressive institutions, actions and images can empower those who suffer from oppression, thereby augmenting and reinforcing their sense of worth. Boxill (1976), for example, claims that oppressed people can have faith in their worth only if they protest against injustices done to them. In the course of discussing the relation between oppression and self-respect, some theorists have attempted to reconceive traditional conceptions of self-respect so as to expunge from them features that are judged to reflect or champion objectionable aspects of the dominant culture. Dillon (1992b), for example, has argued that the Kantian notion of the person as an independent, autonomous, self-legislating will that underlies most conceptions of self-respect is androcentric (see Feminist ethics §4).

Recent philosophical explorations of self-respect have thrown into relief its value and significance. They have shown that self-respect is a central component of human wellbeing, since its absence is painful and debilitating, and illustrated its connection with many other important moral and social values such as rights, autonomy, dignity, freedom and equality.

See also: Autonomy, ethical; Equality; Freedom and Liberty; Rights

CYNTHIA A. STARK

References and further reading

Boxill, B. (1976) 'Self-Respect and Protest',*Philosophy & Public Affairs* 6: 58-69.(Argues that those whose self-respect is at risk must protest against injustices done to them in order to be confident that they are self-respecting.)

Dillon, R. (1992a) 'How to Lose Your Self-Respect', *American Philosophical Quarterly* 29: 125-39.(Explains the distinction between 'evaluative' and 'recognition' self-respect and identifies a variety of failures of self-respect.)

Dillon, R. (1992b) 'Toward a Feminist Theory of Self-Respect', *Hypatia* 7: 52-69.(A feminist evaluation of traditional conceptions of self-respect.)

Hill, T.E., Jr. (1973) 'Servility and Self-Respect', *Monist* 57: 12-27.(Extensively anthologized discussion of the concept of self-respect from a Kantian perspective.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. with notes by H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(Develops an ethical theory centred around a principle of morality that emphasizes the absolute

worth of rational agents.)
Hill, T.E., Jr. (1797) *Die Metaphysik der Sitten*, trans. M.J. Gregor, *The Metaphysics of Morals*, Cambridge: Cambridge University Press, 1991.(Presents the various duties, including the duty of self-respect, which are derived from the supreme principle of morality established in *Groundwork of the Metaphysic of Morals*.)
Massey, S. (1983) 'Is Self-Respect a Moral or a Psychological Concept?', *Ethics* 93: 246-61.(Offers a distinction between 'objective' and 'subjective' conceptions of self-respect and endorses the latter.)
Mohr, R. (1988) 'Dignity vs. Politics: Strategy When Justice Fails', in *Gays/Justice: A Study of Ethics Society and Law*, New York: Columbia University Press: 315-7.(Argues that the gay and lesbian rights movement ought to adopt strategies that promote the dignity and self-respect of gays and lesbians.)
Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(Presents a theory of distributive justice according to which self-respect is a prominent social good.)
Thomas, L. (1978) 'Rawlsian Self-Respect and the Black Consciousness Movement', *Philosophical Forum* 9: 303-14.(Argues that the black consciousness movement was in part a call for African-Americans to respect themselves as persons.)

Sellars, Wilfrid Stalker (1912-89)

Wilfrid Sellars was among the most systematic and innovative of post-war American philosophers. His critical destruction of the 'Myth of the Given' *established him as a leading voice in the Anglo-American critique of 'the Cartesian concept of mind' and in the corresponding shift of attention from the categories of thought to public language. His own positive views were naturalistic, combining a robust scientific realism with a thoroughgoing nominalism which rejected both traditional abstract entities and ontologically primitive meanings. In their place, Sellars elucidated linguistic meaning and the content of thought in terms of a sophisticated theory of conceptual roles, instantiated in the linguistic conduct of speakers and transmitted by modes of cultural inheritance. He combined this theory with a form of 'verbal behaviourism' to produce the first version of functionalism in the contemporary philosophy of mind. Besides his profoundly original philosophical contributions, his long career as a distinguished teacher and influential editor earned him justified acclaim as one of the definitive figures of the post-war period.*

1 Life and works

As the son of the eminent Critical Realist philosopher Roy Wood Sellars, Wilfrid Sellars' own philosophical calling was almost preordained. Educated at the University of Michigan and as a Rhodes Scholar at Oxford, he held positions at the University of Iowa, the University of Minnesota and Yale University before becoming University Professor of Philosophy and Research Professor of the Philosophy of Science at the University of Pittsburgh, where he remained from 1963 until his death. In 1950, with Herbert Feigl, he founded the first scholarly journal explicitly devoted to analytic philosophy, *Philosophical Studies*, edited jointly until 1971 and by Sellars alone for three more years.

Sellars saw post-war philosophers as confronted with two 'images', each of which laid claim to being a (potentially) complete picture of man-in-the-world - the 'manifest image', which had been the focal concern of 'perennial philosophy' from Plato and Aristotle to Strawson and Austin, and the 'scientific image', a complex understanding of man-in-the-world still in the process of emerging from the fruits of theoretical reasoning. Philosophy was challenged to explore the possibility of a 'stereoscopic understanding', in which the two images would become 'fused' into a single synoptic vision. Sellars' writings are consequently dialectical and synthesizing, typically undercutting or evading accepted dichotomies and attempting to mediate conflicting intuitions. They are shaped by the standing conviction that it is a philosopher's duty not merely to show that a received position has gone wrong, but also to explain why and how it could ever have appeared to be right.

2 Ontological perspectives

Along with Wittgenstein, Sellars figured among the pre-eminent contributors to the twentieth century's thoroughgoing critique of Descartes' picture of first-person privileged access to the states of a thinking substance (see Descartes, R.; Privacy). Unlike Wittgenstein, however, Sellars proceeded to develop a positive reconception of the problem space traditionally centred on a dualism of bodies and minds as a transposition of a more fundamental dualism of facts and norms (see Dualism; Fact/value distinction). Kant's analysis of the distinct contributions of the sensibility and understanding in perceptual experience became the guiding thread for Sellars' further sharp separation of problems regarding 'raw feels' from questions regarding intentional (object-directed) thought.

Along with Quine, but without his global rejection of semantic, mental and modal contexts, Sellars helped consummate the shift of semantic attention from the categories of thought to those of public language. Sellars differs from Quine, too, in that his own rejection of Cartesian models of the mind never issued in a correlative (radically behaviourist) abandonment of inner episodes *per se*. Instead, Sellars' critique of the positivist picture of scientific theories enabled him to develop an alternative, thoroughly realistic, understanding of sophisticated scientific inquiry in terms of which the status of concepts pertaining to inner episodes could be constructively understood on the model of theoretical discourse (see Scientific realism and antirealism; Theories, scientific).

Sellars' fundamental metaphysical conviction that the distinguishing mark of the real is the power to act or be acted upon was reflected in his thoroughgoing naturalism. This, in turn, imposed strong nominalistic constraints,

not only on the overall synoptic project, which consequently needed to find a place for mind that did not require independent ontological status for intentional entities, but also on traditional categorial ontology, which could legitimately maintain the reality of abstract entities only if it were possible simultaneously to supply an adequate account of their place within the causal order, broadly conceived. Sellars' response to both of these naturalistic challenges was to develop a sophisticated theory of conceptual roles at the heart of which was the idea that one and the same item could be at once both a causally evoked response to the environment and a normatively significant item in a rule-governed network of reasons. On Sellars' view, a person's grasp of a concept, for example, the concept 'red', consists precisely in their differential disposition to produce such inferentially articulated responses to red things (see Semantics, conceptual role).

Sellars consequently proposed to interpret both categorial ontological idioms and mentalistic intentional contexts in terms of a semantic discourse fundamentally conceived in terms of the inferential roles of public 'natural linguistic objects'. Like Rudolf Carnap, Sellars embraced a form of 'linguistic nominalism', which treated traditional categorial discourse as the classificatory discourse of a functional metalanguage transposed into the 'material mode of speech'. Unlike Carnap, however, he refused to identify the extensional constructs of 'pure' formal syntax or semantics with their corresponding 'descriptive' pre-philosophical counterparts, arguing that to do so is to fail to acknowledge properly the irreducibly normative aspects of syntactical and semantical words functioning as such.

In the philosophy of mind, Sellars adopted 'psychological nominalism', the denial that any sort of commerce with abstract entities is an essential ingredient of mental acts, and espoused instead what he came to call 'verbal behaviourism', according to which the primary sense of (occurrently) thinking that-*p* is a 'thinking-out-loud that-*p*', that is, in first approximation, an event of candidly and spontaneously saying *p* (see Behaviourism, methodological and scientific; Language of thought). The conceptual framework of inner episodes was to be understood as built upon the semantical characterization of overt verbal episodes in a manner analogous to the way in which, for instance, the framework of molecules is built on the observable behaviour of gases. On Sellars' account, the concept of a thought is fundamentally the concept of a causally mediating logico-semantic (inferential) role player, the determinate intrinsic ontological (empirical) character of which is left open. The identification of such intentionally-characterized inner episodes with, for instance, occurrences in an organism's central nervous system thus generates no ontological tension, and the manifest image's conception of persons as thinkers can fuse smoothly with the scientific image's story of complex material organisms. These proposals constituted the original version of functionalism in the contemporary philosophy of mind, the subsequent manifold varieties of which were both facilitated and directly or indirectly inspired by the theoretical space of philosophical options they had opened (see Functionalism).

As in the case of thoughts, Sellars argued that talk about sensations ('raw feels') can also usefully be interpreted on the model of a species of theoretical discourse. Since, however, with Kant and contrary to the Cartesian tradition, Sellars viewed sensations as *non*-intentional items, his account both of their place within the manifest image and their ultimate relocation within the scientific image differs essentially from his ontologically noncommittal functionalist picture of occurrent thoughts. Sellars proposed that it is the job of analogical thinking to construct new forms of concepts pertaining to theoretical entities. Within the philosophically refined manifest image, sensation concepts pick out non-intentional states of individual perceivers, 'sensings', whose intrinsic characters are analogically indicated by adverbial transpositions of qualitative predicates of objects. In a series of essays stretching over a period of forty years and widely regarded as among his most difficult and challenging works, however, Sellars argued further that, because the manifest image's unitary perceiving subjects have ontological pluralities as their scientific image counterparts, the fusion of the two images at the point of sensations will in fact require the postulation of further (theoretically) basic entities which actually ontologically instantiate qualitative sensory contents conceived in yet another analogically constructed categorial guise distinct from both 'properties of objects' and 'manners of sensing'. In particular, he concluded, sensory contents can be non-epiphenomenally integrated into the scientific image only after it has as a whole been transposed from its classical 'thing-like' (particulate) form into a categorially monistic framework all of whose fundamental entities are 'event-like' absolute processes (see Mental states, adverbial theory of; Epiphenomenalism; Qualia).

3 Epistemological perspectives

Sellars' key nominalistic strategies could succeed only if the notion of a linguistic item's having a semantic or inferential function could itself be explicated without recourse to irreducibly platonistic or mentalistic conceptions. One of the greatest strengths of his systematic philosophy is the exquisite care with which he proceeded to locate the normative conceptual order within the causal order and to interpret the modes of causality exercised by linguistic rules. For Sellars, inference itself is always a normative affair, a matter of the judgments one ought to or is entitled to make. He defuses the circularity which threatens such an account by arguing that our knowledge of what implies and follows from various claims is, in the first instance, a *practical* ability to discriminate, that is, to respond differentially to, good and bad inferences. Rule-governed linguistic behaviour develops out of multiple repertoires of 'pattern-governed behaviour', behaviour which exhibits a pattern because the propensity to produce behaviour belonging to that pattern has been selectively reinforced and contrary propensities selectively extinguished (see language, social nature of; Meaning and rule-following). The pattern-governed behaviour characteristic of language includes 'language-entry transitions', propensities to respond to non-linguistic states of affairs (such as sensory stimulations) with appropriate linguistic activity; 'language-departure transitions', propensities to respond to a subset of linguistic representings (for example, such first-person future-tensed conduct-ascriptions as 'I shall now raise my hand') with appropriate corresponding behaviour; and 'intra-linguistic moves', propensities to respond to linguistic representings with further linguistic episodes (only) in patterns corresponding to valid theoretical and practical inferences. Linguistic roles or functions, Sellars suggested, are ultimately individuated in terms of the structures of positive and negative uniformities generated in the natural order by such pattern-governed activities.

In the Kantian tradition Sellars insisted that, in contrast to the mere capacity to be sensorily affected by external objects, *perception* of how things are requires not only systematic differential response dispositions but also the ability to respond to sensory stimulation with a *judgment*, that is, the endorsement of a claim (see Perception, epistemic issues in). Sellars went on, however, to propose that reports of how things look or seem, rather than employing 'more primitive' concepts, result instead from withholding these characteristic endorsements. This account enabled him to explain the incorrigibility of 'seems' judgments that Cartesianism takes as its fundamental datum. Their incorrigibility is simply a matter of their tentativeness; a 'seems' judgment expresses a perceptual ascription without endorsing it. It follows that 'seems' judgments do not express a special class of immediate cognitions. Applying the concept 'looks red' requires the same mastery of inferential articulations, the same inferential 'know how', as does applying the concept 'is red'.

Sellars' analysis of the Cartesian incorrigibility of perceptual 'seemings' is one strand of the philosophical dialectic most frequently associated with his name, his comprehensive critique of the 'Myth of the Given'. Basic to this critique is his insistence on the irreducibly normative character of epistemic discourse. Characterizing an episode or state in epistemic terms is not giving an empirical description of it but rather placing it within a social framework of justifications, of having and being able to give reasons for what one says. All knowledge that something is the case - all 'subsumption of particulars under universals' - presupposes intersubjective learning and concept formation. It follows that the ability to be (epistemically) aware of a sort of thing rests upon a *prior* command of the concept of that sort of thing and cannot account for it - and this holds equally true for concepts pertaining to 'inner episodes'. The first-person reporting role of such concepts, a use Cartesians interpret as evidencing the 'privacy' of the mental and one's 'privileged access' to one's own mental states, is necessarily built upon and presupposes their intersubjective status.

The idea that a language necessarily contains a stratum of claims any of which, on various occasions, can occur either as a report (an unmediated causally evoked response) or as the conclusion of an inference (mediated by other linguistic-conceptual representations) suggests the possibility of a further stratum of purely inferential claims. That is how Sellars proposes that we understand theoretical claims, that is, as claims that one can *de facto* become entitled to endorse only as the conclusion of an inference. On this view, the difference between theoretical objects and observables is not ontological but methodological. It is not a matter of the kind of thing our claims are about; it is a matter of how we come to be entitled to make them - and, as our technological sophistication increases, this can change. Once it is recognized that there is no epistemically privileged stratum of judgments with which to respond to one's sensory environment, and that such a 'language-entry' response is what 'observation' is, it follows that nothing is in principle unobservable. Noninferential and purely inferential claims can both be about the same kind of object (see Observation).

Sellars' holistic view of justification implied that the reasonableness of accepting even first principles is a matter of the availability of good arguments warranting their acceptance. Since it is definitive of first principles that they cannot be derived by sound theoretical reasonings from still more basic premises, Sellars located the requisite arguments in a stratum of 'vindications', practical reasonings in which the adopting of specific principles (an epistemic conduct) is demonstrated to contribute to the realization of non-arbitrary epistemic ends. From the ontological point of view, however, such practical cognitions - intentions and volitions - are simply species of occurrent thinkings whose unique conduct-structuring functional role is to be understood in terms of their special ('language-exit') relationships to non-linguistic behaviour (see Justification, epistemic).

For Sellars, it is finally practical thinking which lies at the centre of the framework of persons as such (see Persons). To think of an entity as a person is to think of it as actually or potentially a member of a community, and it is the most general common intentions of a community - its 'we-intentions' - that fundamentally define the structure of norms and values, crucially including those which make meaningful discourse and rationality itself possible, in terms of which the conducts of its members come to be appraised. A genuinely synoptic vision of man-in-the-world, Sellars concluded, can therefore be achieved only by enriching the scientific image with the language of individual and shared intentions.

See also: Categories; Concepts; Foundationalism; Intentionality; Nominalism; Ontology

JAY F. ROSENBERG

List of works

Sellars, W.S. (1963) *Science, Perception and Reality*, London and New York: Routledge & Kegan Paul; repr. Atascadero, CA: Ridgeview, 1991.(Contains Sellars' best-known essays, including his classical critique of the 'Myth of the Given', 'Empiricism and the Philosophy of Mind', and his metaphilosophical manifesto, 'Philosophy and the Scientific Image of Man'.)

Sellars, W.S. (1967) *Philosophical Perspectives*, Springfield, IL: Charles C. Thomas; repr. Atascadero, CA: Ridgeview, 1977, 2 vols.(Work from the 1950s and early 1960s. Part 1 consists of historical essays on Plato, Aristotle and the Rationalists. Part 2 includes important original essays in metaphysics and the philosophy of mind.)

Sellars, W.S. (1968) *Science and Metaphysics: Variations on Kantian Themes*, The John Locke Lectures for 1965-6, London and New York: Routledge & Kegan Paul; repr. Atascadero, CA: Ridgeview, 1992(Based on his John Locke Lectures for 1965-6. The most systematic presentation of Sellars' views on perception, intentionality, representation, truth and the metaphysics and ethics of persons, developed in relation to the work of Kant, Peirce and the early Wittgenstein.)

Sellars, W.S. (1974) *Essays in Philosophy and Its History*, Dordrecht: Reidel.(A rich collection of historical interpretive and systematic essays from Sellars' most productive years. Includes his presidential address to the American Philosophical Association on the Kantian text 'this I or he or it (the thing) which thinks…', as well as significant contributions to the philosophy of language, metaphysics and the philosophy of science.)

Sellars, W.S. (1980) *Naturalism and Ontology*, Atascadero, CA: Ridgeview.(Sellars' John Dewey Lectures for 1973-4 present his mature views on the interrelationships of language, logic and ontology. The most systematic defence of his 'linguistic nominalism'.)

Sellars, W.S. (1981a) *Pure Pragmatics and Possible Worlds*, ed. and with intro. by J.F. Sicha, Atascadero, CA: Ridgeview.(A comprehensive collection of Sellars' earliest works (1947-53). These richly textured and dialectically complex early essays are difficult to access, but, besides offering a developmental perspective on Sellars' overall corpus, they include a number of interesting ideas and theses not explicitly developed in his later work.)

Sellars, W.S. (1981b) 'Foundations for a Metaphysics of Pure Process', *Monist* 64: 3-90.(A special issue of the *Monist*, containing Sellars' Carus Lectures from 1977-8. This is the most developed and systematic presentation of Sellars' unique views in the philosophy of mind. The title reflects his conviction that sensible qualities can be satisfactorily accommodated within a naturalistic worldview only in the context of a (speculatively-envisioned) ontology of 'pure processes'.)

Sellars, W.S. (1989) *The Metaphysics of Epistemology: Lectures by Wilfrid Sellars*, ed. P.V. Amaral, Atascadero, CA: Ridgeview.(This posthumously published work reproduces a series of university lectures given by Sellars in 1975 on central topics in epistemology and metaphysics. In sharp contrast to his densely written essays, Sellars' classroom presentations tended to be discursively relaxed and informal. This charmingly illustrated book is consequently a good first introduction to his complex views regarding perception, knowledge, minds,

meaning and representation.)

References and further reading.

Castañeda, H.-N. (ed.) (1975) *Action, Knowledge, and Reality: Critical Studies in Honor of Wilfrid Sellars*, Indianapolis, IN: Bobbs-Merrill.(A collection of critical studies of Sellars' work by his students and colleagues. The book also contains two important pieces of original Sellarsiana, his intellectual autobiography and 'The Structure of Knowledge', a systematic treatment of epistemological themes based on Sellars' 1971 Machette Foundation Lectures at the University of Texas.)

Delaney, C.F. *et al*. (1977) *The Synoptic Vision: Essays on the Philosophy of Wilfrid Sellars*, Notre Dame, IN: University of Notre Dame Press.(Interpretive and critical essays on Sellars' views and contributions in epistemology, ontology, philosophy of science, philosophy of mind and ethics by members of the philosophical faculty of Notre Dame University. The title echoes Sellars' thesis that contemporary philosophy is challenged with fusing two comprehensive worldviews, the 'manifest image' and the 'scientific image' into a single 'synoptic' picture.)

Pitt, J.C. (ed.) (1978) *The Philosophy of Wilfrid Sellars: Queries and Extensions*, Dordrecht: Reidel.(Fourteen interpretive and critical essays deriving from a Workshop on the Philosophy of Wilfrid Sellars held at the Virginia Polytechnic Institute and State University in 1976.)

Seibt, J. (1990) *Properties as Processes: A Synoptic Study of Wilfrid Sellars' Nominalism*, Atascadero, CA: Ridgeview.(The first full-scale, systematic, book-length study of Sellars' philosophy, written by one of his last doctoral students and described by Sellars himself as 'one of the best essays on my work that I have ever seen'. This presentation of Sellars as 'a unique example of radical and systematic nominalism' is indispensable for any serious student of his views.)

Semantic paradoxes and theories of truth

The Cretan philosopher Epimenides said that Cretans always lie. Assuming, for the sake of argument, the mendacity of all other statements by Cretans, we get a paradox: if what Epimenides said was true, it must have been a lie, whereas if what he said was a lie, it would have made his statement true. The citizens of Crete have long since forgiven the insult, but semantics has never recovered.

Alfred Tarski perceived the consequences of Epimenides' paradox with particular clarity. Our common-sense intuitions about truth follow the paradigm: 'Snow is white' is true if and only if snow is white. As Tarski rigorously shows, if the language we are describing (the object language) is the same as the language in which we are formulating our theory (the metalanguage), this paradigm will be inconsistent with the rudimentary laws of syntax. The conclusion Tarski drew was that, if we are to develop a satisfactory theory of truth, our metalanguage must be essentially richer in expressive power than the object language. Since there is no human language essentially richer than English (or any other natural language), there can be no satisfactory theory of truth for English.

One earnestly hopes that this is not the end of the matter. Tarski's analysis leaves open the prospect that we can develop a fully satisfactory theory of truth for a substantial fragment of English; also the prospect that we can develop a theory of truth for English as a whole which, while not fully satisfying our intuitions, is none the less useful and illuminating.

Both prospects have been substantially advanced by Saul Kripke's 'Outline of a Theory of Truth', *which exploits the idea that there are truth-value gaps, statements which are neither true nor false, and that Epimenides' insult was one of them.*

Invocation of truth-value gaps does not resolve the paradox in any straightforward way. If we let the phrase 'the simple liar sentence' refer to the sentence 'The simple liar sentence is false', we see that we can readily account for the paradoxical features of the sentence by declaring the sentence neither true nor false; but if we let the strengthened liar sentence be 'The strengthened liar sentence is not true', we get a sentence we cannot dispose of so tidily. If the strengthened liar is neither true nor false, then it is not true; but that it is not true is precisely what the sentence says.

Truth-value gaps have not vanquished the liar paradox. Nor have any of the alternatives, the most prominent of which are a contextualist account, which sees the English word 'true' as radically ambiguous, and so-called 'revision theory', which investigates the cyclic reasoning that occurs when we try to evaluate the simple liar sentence: if the sentence is true, then it must be false; but if, then, it is false, it must be true; and so on. While these approaches have not eliminated the paradox, they have opened new approaches that have greatly improved our prospects for finding a comfortable way to live with it.

1 Tarski's requirement of a richer metalanguage

Kurt Gödel (1931) proved the incompleteness of a version of Peano Arithmetic by associating a code number, denoted $\ulcorner\phi\urcorner$, with each sentence ϕ of the language of arithmetic, then constructing a sentence χ which denied that $\ulcorner\chi\urcorner$ was the code of a theorem (see Gödel's theorems). Tarski (1935) used Gödel's construction to give a formalized version of the liar paradox. If $Tr(x)$ were a formula of the language of arithmetic which represented the set of code numbers of true sentences, we would expect to have

$$Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi(\mathrm{T})$$

for each sentence ϕ. But Gödel's methods can be used to construct a sentence λ such that $\lambda \leftrightarrow \neg Tr(\ulcorner\lambda\urcorner)$ is a theorem. We conclude that there is no formula of the language of arithmetic which represents the set of code numbers of true sentences of the language of arithmetic. Indeed, this result holds not only for the language of arithmetic but for any language which includes the language of arithmetic.

Arithmetic plays an inessential role here. Quine (1946) shows that we can encode the language of arithmetic into any language which can provide a rudimentary theory of syntax. We are forced to conclude that, if $\mathcal{L}$ is a language that is able to describe its own syntax, then there is no open sentence $Tr(x)$ of $\mathcal{L}$ such that the schema (T) is consistent with the laws of syntax; here $\ulcorner\phi\urcorner$ is a standard name of ϕ which describes the sentence in terms of its

syntactic structure. On the other hand, any semantic theory which adequately represents our intuitive understanding of truth must surely imply all the (T)-sentences. We conclude that there is no possibility, consistent with the laws of syntax, of finding an open sentence of $\mathcal{L}$ which adequately represents the set of truths of $\mathcal{L}$. The conclusion Tarski drew, and which nearly everyone has accepted, is that, to develop a satisfactory theory of truth for $\mathcal{L}$ we must work within a metalanguage richer than $\mathcal{L}$ in expressive power.

2 *Principia Mathematica* and its successors

Whitehead and Russell responded to Epimenides' paradox by proposing a layered multiplicity of notions of truth. Simple statements of fact enjoy the first order of truth; attributions of truth at order one will be true or false at order two; and so on. An attribution of *n*th order truth must always have order higher than *n*. In this way (though they did not give details), Whitehead and Russell were able to develop the semantics of *Principia Mathematica* without contradictions. *Principia Mathematica* was designed to resolve a variety of paradoxes in addition to the liar. Here we shall follow contemporary practice in separating the semantic from the set-theoretic paradoxes (as recommended by Ramsey 1925), and in focusing on the liar as our paradigmatic representative of the former.

Tarski's approach to the liar was similar to Whitehead and Russell's, but simplified from the intricacies of the theory of types. Tarski spoke of a hierarchy of languages. The ground language has no semantic predicates, while the $(n+1)$th language is derived from the *n*th language by adding a predicate true_n whose extension is the true sentences of the *n*th language.

Whitehead and Russell proposed to apply their analysis to natural languages by treating the English word 'true' as ambiguous; contradictions arise from failing to recognize the ambiguity.

Their proposal has been vigorously pursued, by Tyler Burge (1979), Charles Parsons (1974), Jon Barwise and John Etchemendy (1987) and others, who have treated 'true' as an indexical term, like 'here' and 'now', whose extension is determined in part by the context of utterance. If everything Plato says is either true_{12} or false_{12}, the use of 'true' in 'Everything Plato says is true' will be contextually determined to mean 'true_{13}'. One quickly requires transfinite indices, going beyond the finite levels of *Principia Mathematica*.

Paradoxes arise when the customary rules for placing subscripts falter. Ordinarily, the subscript taken by 'false' in the sentence 'Everything a Cretan says is false' will be greater than the subscripts supplied to any occurrence of 'true' or 'false' in any utterance by a Cretan. But it is not possible to observe this rule when the sentence is uttered by a Cretan. If we take the word 'false' in Epimenides' statement 'Everything a Cretan says is false' to mean 'false_α', we find that his statement is neither true_α nor false_α, since the predicates 'true_α' and 'false_α' only apply to sentences with subscripts less than α.

From a higher level of reflection of which Epimenides himself was in principle incapable, we can see that, since Epimenides' statement that everything a Cretan says is false_α is not false_α, we can make out a sense in which the sentence 'Everything a Cretan says is false' is false, namely, 'Everything a Cretan says is false_α' is $\mathrm{false}_{\alpha+1}$. We should not say, however, that what Epimenides said was false after all, rather, the sentence Epimenides uttered expresses something false when we say it, though in Epimenides' mouth it was neither true nor false.

Talk of levels of reflection is something the theorist engages in when describing the language, but not something speakers of the language engage in. If this were not so, Epimenides could say 'Nothing a Cretan says is true at any level of reflection' and reinstate the paradox. Thus we still submit to Tarski's requirement that the metalanguage have expressive powers not available in the object language. Again, the theorist can make the subscripts under 'true' and 'false' explicit, and use variables as subscripts. The speakers of the language cannot do this, otherwise they could get a contradiction by formulating the explicit liar ('The explicit liar is not true_α, for any α'). Contextualists have not overturned Tarski's requirement that the semantics of a language be developed in a richer metalanguage, nor, for the most part, have they attempted to do so. Instead they have tried to show that, contrary to first impressions, Tarski's requirement is relaxed enough to make room for virtually everything speakers of English want to say.

3 Kripke's theory of truth

We have been pretending that, apart from one notorious remark by Epimenides, everything ever said by a Cretan is

outright false. This pretence is, of course, contrary to fact, and, once we drop the pretence, Epimenides' statement is no longer problematic. If we hear a Cretan remark that snow is white, we can conclude at once that not all Cretans always lie, hence Epimenides' statement is false. To reach this judgment, it is not necessary to examine everything a Cretan has ever said; it is enough to find one Cretan truth.

The assignment of levels we get from *Principia Mathematica* is unnecessarily extravagant. It is based upon the idea that, to determine the truth of 'Everything a Cretan says is false', we need to take account of everything any Cretan ever says, so it tries, unsuccessfully, to assign to the sentence a level higher than the level assigned to any Cretan statement. But to determine that Epimenides' statement is false, we do not need to take account of everything a Cretan ever says, we only need to find one truth uttered by a Cretan. So, if some Cretan statement is true at the first level, we can declare Epimenides' statement false at the second level, without waiting to examine a raft of irrelevant further statements.

This adjustment in the assignment of levels effects a remarkable simplification and clarification of the hierarchical structure. In describing the construction, we shall utilize a first-order language with primitive connectives 'and', 'not' and 'all', other connectives defined as usual. All terms other than '*Tr*' are assumed already to have a fixed meaning. We assume that our universe of discourse is a set, and we suppose that every member of the universe is named by some constant, expanding the language if necessary to ensure this. (Alternatively, we could define truth in terms of satisfaction, as Tarski does; but this would complicate matters.) We assume that the language can describe its own syntax, either directly or via a Gödel coding.

A non-semantic atomic sentence will be either true_0 or false_0.

A sentence of the form $Tr(c)$ is $\mathrm{true}_{\alpha+1}$ if *c* denotes a sentence which is true_α, and $\mathrm{false}_{\alpha+1}$ if *c* denotes a sentence which is false_α. If the individual denoted by *c* is not a sentence, $Tr(c)$ is false_0.

If ϕ is true_α and ψ is true_β, the conjunction $(\phi \wedge \psi)$ will be $\mathrm{true}_{\max(\alpha,\beta)}$. If either conjunct is false_γ for some γ, the conjunction is false_γ for the least such γ.

A negation is true_α if the negatum is false_α and false_α if the negatum is true_α.

A universally quantified sentence is true_α if α is the least ordinal such that every substitution instance of the sentence is true_β for some $\beta \leq \alpha$. If some substitution instance of $\psi(x)$ is false_γ, then $(\forall x)\psi(x)$ is false_γ for the least such γ.

Kripke's proposal (1975) is that a sentence be regarded as true if it is true_α for some α, false if it is false_α for some α, and undetermined otherwise. Applying his construction to a suitably regimented fragment of English, we find that it manages narrowly to skirt the edge of paradox without falling into contradiction, and the truth-values it supplies, when it supplies truth-values, are invariably the intuitively correct ones. For example, suppose that Socrates says only

Everything Plato says is true.

while Plato says

Not everything Socrates says is true.
Everything Epimenides says is true.

and Epimenides says

Nothing a Cretan says is true.
Snow is white.

Epimenides' second statement is true_0, so his first statement is false_1. Hence Plato's second statement is false_2, Socrates' statement is false_3, and Plato's first statement is true_4.

Not only does Kripke's theory get the right truth-values, but, more importantly, it explains them. Sentences declared true are true because they are grounded in the non-semantic facts. Thus a sentence is true_α for some α just in case it is derivable from true atomic and negated atomic non-semantic sentences by the following rules,

while a sentence is false_α, for some α, just in case its negation is derivable:

'*Tr*'-introduction: If *c* denotes a sentence ϕ, infer $Tr(c)$ from ϕ.
'$\neg Tr$'-introduction: If *c* denotes a sentence ϕ, infer $\neg Tr(c)$ from $\neg\phi$. Also, if *c* denotes a non-sentence, you may write $\neg Tr(c)$.
Conjunction introduction: Infer a conjunction from its conjuncts.
Negated conjunction introduction: Infer the negation of a conjunction from the negation of either conjunct.
Double negation introduction: From a sentence, infer the negation of its negation.
Universal quantification introduction: Infer a universal quantified statement from all of its instances. (This is an infinitary rule.)
Negated universal quantification introduction: Infer the negation of a universally quantified statement from the negation of any of its instances.

4 The fixed-point theorem

We have been speaking of 'true' as an ambiguous predicate. It is perhaps simpler, and certainly closer to Kripke's original way of thinking, instead to regard 'true' as a univocal predicate in a language with a nonclassical logic. We employ Kleene's 'strong three-valued logic' (see Kleene 1952: §54), according to which a predicate has an 'extension', consisting of those things to which it is determined that the predicate applies, and an 'anti-extension', consisting of those things to which it is determined that the predicate does not apply. The two cannot overlap, but there may be things to which it remains unsettled whether the predicate applies. An atomic sentence $Rc_1 \ldots c_n$ is true if the *n*-tuple consisting of the things denoted by c_1 to c_n lies in the extension of *R*, false if it lies in the anti-extension, and truth-valueless otherwise. A conjunction is true if both conjuncts are true and false if either conjunct is false. A negation is true if the negatum is false and false if the negatum is true. A universally quantified statement is true if every instance is true and false if at least one instance is false.

In the application we are interested in, all the terms other than '*Tr*' already have a fixed interpretation in which the anti-extension is the complement of the extension, so all we need to determine is the extension and the anti-extension of '*Tr*'. Where *A* and *B* are non-overlapping sets, let $\Gamma(A, B)$ be the pair $\langle C, D\rangle$, where *C* is the set of sentences which are declared true if *A* is assigned as the extension of '*Tr*' and *B* as its anti-extension; and *D* is the union of the set of non-sentences and the set of sentences declared false when *A* is assigned as the extension of '*Tr*' and *B* as its anti-extension. $\langle A, B\rangle$ is a 'fixed point' if and only if $\langle A, B\rangle = \Gamma(A, B)$.

Kripke's construction can now be succinctly characterized. Let *C* be the set of sentences which are true_α for some α, and let *D* be the set consisting of the non-sentences and the sentences which are false_α for some α. Then $\langle C, D\rangle$ is the smallest fixed point; that is, it is a fixed point included in every other fixed point. (We say $\langle A, B\rangle$ is included in $\langle C, D\rangle$ if $A \subseteq C$ and $B \subseteq D$.)

The existence of fixed points is startling. The liar paradox shows that, within the context of classical two-valued logic, it is not possible to assign an extension to '*Tr*' in such a way that, for any sentence ϕ, ϕ is true or false according as $Tr(\ulcorner\phi\urcorner)$ is true or false. We can, however, solve the analogous problem for three-valued logic. At a fixed point, we assign an extension and an anti-extension to '*Tr*' in such a way that, for any sentence ϕ, ϕ is true, false or unsettled according as $Tr(\ulcorner\phi\urcorner)$ is true, false or undecided.

There are fixed points other than the smallest. If *A* and *B* are non-overlapping sets with $\langle A, B\rangle \subseteq \Gamma(A, B)$, then there is a smallest fixed point which contains $\langle A, B\rangle$. The extension of '*Tr*' in this smallest fixed point consists of the sentences which are derivable from the true atomic and negated atomic non-semantic sentences, together with the sentences in *A* and the negations of the sentences in *B*.

Fixed points other than the smallest are useful in describing the different sorts of pathological behaviour we find among self-referring sentences. Tarski's sentence λ, which asserts its own untruth, is viciously paradoxical; whether we declare it true or false, we fall into contradiction (see Gödel's theorems). By contrast, if we formalize the construction of the truth-teller sentence ('The truth-teller sentence is true'), getting a sentence τ which is provably equivalent to $Tr(\ulcorner\tau\urcorner)$, we again get a sentence which is paradoxical, but in a more compliant fashion; whether we declare the sentence true or false, the sentence is amiably willing to go along, but no such classification has any basis in extra-semantic reality.

The difference between these two sorts of pathologies is reflected in the structure of the fixed points. There are no fixed points in which λ is assigned a truth-value. By contrast, there are some fixed points in which τ is true, others in which it is false, and still others in which it lacks a truth-value.

There are a great many fixed points; if the language has κ sentences, there will be 2^κ of them. The fixed points constitute a complete lower semi-lattice, that is, a partially ordered set in which every non-empty subset has a greatest lower bound. If $\mathcal{K}$ is a non-empty collection of fixed points, the union of all the fixed points which are included in every member of $\mathcal{K}$ will be a fixed point, so it will be the greatest lower bound of $\mathcal{K}$.

5 Classical-logic versions of Kripke's theory

Where Kripke's construction assigns a truth-value, the value it assigns invariably agrees with our intuitive understanding of the everyday application of 'true'. However, there are many sentences to which the construction fails to assign a truth-value but about which ordinary intuition fails to detect even a whiff of paradox. For example, 'A conjunction is true if both its conjuncts are true' is, according to intuition, harmlessly and obviously true, yet the construction fails to give it a truth-value. Indeed, the construction fails to provide even the simplest semantic generalizations, withholding its assent from such harmless platitudes as 'Every true sentence is a true sentence'; since $Tr(\ulcorner\lambda\urcorner)$ is truth-valueless, so is $(Tr(\ulcorner\lambda\urcorner) \rightarrow Tr(\ulcorner\lambda\urcorner))$, hence $(\forall x)(Tr(x) \rightarrow Tr(x))$ falls into the gap between truth and falsity.

This limitation is easily remedied by extending the rules of inference to encompass classical logic and taking as axiomatic the principle that, if a conditional and its antecedent are both true, the consequent is also true. Thus the variant construction treats a sentence as true just in case it is derivable from true atomic and negated atomic non-semantic sentences, together with instances of the schema

$$(Tr(\ulcorner\phi \rightarrow \psi\urcorner) \rightarrow (Tr(\ulcorner\phi\urcorner) \rightarrow Tr(\ulcorner\psi\urcorner))),$$

by '*Tr*'-introduction, '$\neg Tr$'-introduction, universal quantification introduction, and the laws of classical logic.

According to this variant construction, all classically valid sentences are true, in particular, $(\forall x)(Tr(x) \rightarrow Tr(x))$. For each ϕ and ψ, the conditional

$$((Tr(\ulcorner\phi\urcorner) \wedge Tr(\ulcorner\psi\urcorner)) \rightarrow Tr(\ulcorner\phi \wedge \psi\urcorner))$$

is derivable from the fact that '*Tr*' distributes over conditionals, and the generalization that a conjunction is true if both its conjuncts are true follows by universal quantification introduction. Indeed, we can show that an inference schema is valid in classical logic if and only if all substitution instances of the schema are truth-preserving.

There is a price to pay. In the original construction, truth is unmistakably grounded in the non-semantic facts. In the alternative construction, this is not so evident, since we want an explanation of the truth of the laws of classical logic. In the original construction, the truth of a disjunction was based upon the truth of one of its disjuncts, but, in the alternative construction, this is no longer so.

Another construction with much the same effect employs Bas van Fraassen's method of 'supervaluations' (1966) in place of Kleene's three-valued logic; this is the method suggested by Kripke himself. The sentences declared true in $\Gamma(A, B)$ will be those sentences true in all the classical models we get from the original model by taking the extension of '*Tr*' to be a consistent set of sentences, closed under first-order consequence, which contains *A* and is disjoint from *B*; the sentences declared false will be those whose negations are declared true. A sentence is regarded as true if it is declared true in the smallest fixed point.

The two constructions give similar results. Indeed, if the language is countable (which implies, since everything has a name, that the universe of discourse is countable), they give precisely the same results.

The classical-logic versions of Kripke's theory capture a great many of our common-sense intuitions about truth and paradox, but there remain fundamental intuitions that they leave aside. Notably, the intuition driving the theory is that the paradoxical sentences are neither true nor false. But the statement that the strengthened liar sentence is neither true nor false, while formalizable in the language, is never accepted as true. That the strengthened liar sentence is neither true nor false is something we can only see from the perspective of an essentially richer metalanguage. Within the object language, we can only recognize a sentence as untrue if at some

stage it is declared false; only from the external vantage point of the metalanguage can we regard a sentence as untrue because it is never declared true.

6 Revision theory

There are other proposals for obtaining an unambiguous truth predicate within a classical framework, notably the ‘revision theory’ of Hans Herzberger (1982), Anil Gupta (1982) and Nuel Belnap (1982), which sees the paradoxes as arising because the processes by which we ordinarily introduce new terms into the language, as seen in ordinary explicit definitions and in inductive definitions, partially break down when the (T)-sentences are taken to define ‘*Tr*’. The processes fail to stabilize upon a single choice for the extension of ‘*Tr*’. Instead, if V_0 is a candidate for the extension of ‘*Tr*’ we get another candidate, at least as good, by taking the extension of ‘*Tr*’ to be the set V_1 of sentences true in the classical model obtained by setting the extension of ‘*Tr*’ equal to V_0. We get another good candidate by taking the extension of ‘*Tr*’ to be the set V_2 of sentences true in the model in which the extension of ‘*Tr*’ is V_1. The process never settles down. Whereas Plato’s first statement above is in V_n for all $n > 5$, if the simple liar sentence is in V_n, it is outside V_{n+1}, but it is back inside V_{n+2}. The process continues into the transfinite, and, as it proceeds, we get better and better candidates for the extension of ‘*Tr*’. The process must eventually repeat itself, and, after it starts to do so, the candidates we get are all optimal candidates for the extension of ‘*Tr*’. Because the starting point V_0 was arbitrary, there are many different revision sequences. Count a sentence as ‘stably true’ if there is an ordinal level after which it always appears at every stage in every revision sequence.

The details of extending the process into the transfinite - what exactly to do at limit stages - is a matter of some disagreement among the three authors mentioned above, but all agree that a sentence is in V_λ, for λ a limit, if it is in all V_α, for α a sufficiently large ordinal less then λ, and a sentence is outside V_λ if it is outside V_α for sufficiently large $\alpha < \lambda$.

Revision theory still requires a richer metalanguage. We can only construct the revision sequences when the universe of discourse of the original language is a set, so we cannot carry out the construction for the language of set theory or for English, languages whose universe of discourse is not a set. (In those languages, we can talk about all sets, and there is no set that includes all sets.) The hope is that, in spite of this limitation, the fragments of English amenable to revision-theoretic treatment will be rich enough so that their study will cast a useful light upon our ordinary conception of truth in English.

7 English as an object language?

Most of the recent work on the paradoxes has been aimed at explicating the intuitions that govern ordinary speakers’ usage of the word ‘true’. For such purposes, a theory which only takes account of a fragment of the natural language may well suffice. For some theoretical purposes, however, one would like a global semantic theory, which takes full account of the linguistic practices of English speakers. Should it turn out, on the contrary, that a comprehensive theory of language is inherently impossible, so that part of our language lies irretrievably beyond the reach of linguistic inquiry, this would severely curtail the hopes of science, and it would also gravely undermine the prevalent mood of metaphysical naturalism, according to which human language is a part of the natural order, not intrinsically more mysterious or less comprehensible than the planetary orbits.

Hopes for a comprehensive theory are greatly discouraged by the liar paradox, but, if our optimism persists, we have only two alternatives: either to restrict classical logic or to restrict schema (T) (see §1 above). Restricting the logic does not seem to help. If we replace classical logic with Kleene’s three-valued logic, we still do not get the (T)-sentences. Indeed, Feferman’s investigations (1984) would seem to show that we do not get the (T)-sentences within any reasonable logic.

We might hope to preserve at least one direction of schema (T), but even here our hopes are thwarted by Montague’s theorem (1963) that any logical system that includes the left-to-right direction of (T) together with the ‘*Tr*’-introduction rule will be inconsistent with the laws of syntax. A dual result shows that the right-to-left direction is incompatible with ‘$\neg Tr$’-introduction and also that it is incompatible with ‘Tr’-elimination, which is the converse of ‘*Tr*’-introduction.

Another possibility, recommended by Vann McGee (1991), is to replace schema (T) with a system of rules of

inference: 'Tr'-introduction and -elimination and ' $\neg Tr$'-introduction and -elimination. The paradoxes are diagnosed as arising from the misapplication of these rules of inference within conditional proofs. The classical-logic versions of Kripke's theory can be employed to show that this system of rules is consistent.

As this consistency proof shows, Kripke's construction is a highly versatile tool, which can be usefully applied to a wide variety of philosophical purposes.

See also: Inductive definitions and proofs; Logical and mathematical terms, glossary of; Paradoxes of set and property; Tarski's definition of truth; Theory of types

VANN McGEE

References and further reading

Barwise, J. and Etchemendy, J. (1987) *The Liar*, New York: Oxford University Press.(A mathematically and philosophically sophisticated development of the thesis that the referent of 'true' depends on context, developed within situation semantics, as found in Barwise and Perry (1983).)

Barwise, J. and Perry, J. (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press.(Situation semantics, utilized in Barwise and Etchemendy (1987).)

Belnap, N. (1982) 'Gupta's Rule of Revision Theory of Truth', *Journal of Philosophical Logic* 11: 103-16. (Proposed modifying Gupta's revision theory by changing the procedure for limit stages, a proposal accepted by Gupta.)

Burge, T. (1979) 'Semantical Paradox', *Journal of Philosophy* 76: 169-98; repr. in Martin (1984), 83-117.(Lucid and highly influential development of a context-dependent conception of truth.)

Feferman, S. (1984) 'Toward Useful Type-Free Theories, I', in Martin (1984), 237-87.(Careful investigation of a variety of unified approaches to semantic and set-theoretic paradoxes.)

Fraassen, B.C. van (1966) 'Singular Terms, Truth Value Gaps, and Free Logic', *Journal of Philosophy* 64: 464-95.(Invented supervaluations, used in §5.)

Gödel, K. (1931) 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38: 173-98; trans. 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 592-617.(In proving the incompleteness of Peano arithmetic, Gödel developed the fundamental tools of modern logic.)

Gupta, A. (1982) 'Truth and Paradox', *Journal of Philosophical Logic* 11: 1-60; repr. in Martin (1984), 175-235. (Gupta and Herzberger independently discovered revision theory.)

Gupta, A. and Belnap, N. (1993) *The Revision Theory of Truth*, Cambridge, MA: MIT Press.(Extends the revision theory of truth, according to which 'true' is defined by the (T)-sentences, to an illuminating general theory of circular definitions.)

Herzberger, H. (1982) 'Notes on Naïve Semantics', *Journal of Philosophical Logic* 11: 61-102; repr. in Martin (1984), 133-74.(Herzberger and Gupta independently discovered revision theory.)

Kleene, S.C. (1952) *Introduction to Metamathematics*, New York: Elsevier, 1971.(An invaluable textbook. Three-valued logic is in §54.)

Kripke, S.A. (1975) 'Outline of a Theory of Truth', *Journal of Philosophy* 72: 690-716; repr. in Martin (1984), 53-81.(The centre of all contemporary discussion of the liar paradox.)

Martin, R. (ed.) (1984) *Recent Essays on Truth and the Liar Paradox*, New York: Oxford University Press.(A collection of the most important papers.)

Martin, R. and Woodruff, P. (1985) 'On Representing ''True-in-*L*'' in *L*', *Philosophia* 5: 213-17; repr. in Martin (1984), 47-51.(First proved the existence of fixed points, independently of Kripke.)

McGee, V. (1991) *Truth, Vagueness, and Paradox*, Indianapolis, IN: Hackett Publishing Company. (Explores prospects for a theory of truth developed without recourse to a richer metalanguage.)

Montague, R. (1963) 'Syntactical Treatments of Modality, With Corollaries on Reflexion Principles and Finite Axiomatizability', *Acta Philosophica Fennica* 16: 153-67.(A version of the liar paradox with 'necessary' in place of 'true'.)

Parsons, C. (1974) 'The Liar Paradox', *Journal of Philosophical Logic* 3: 381-412; repr. in Martin (1984), 9-45. (Insightfully develops a unified approach to the semantic and set-theoretic paradoxes, based upon the context dependence of both the extension of 'true' and the domain of quantification.)

Quine, W.V. (1946) 'Concatenation as a Basis for Arithmetic', *Journal of Symbolic Logic* 11: 105-14. (Codes arithmetic into the theory of syntax.)

Ramsey, F.P. (1925) 'The Foundations of Mathematics', *Proceedings of the London Mathematical Society* 25: 338-84; repr. in *Philosophical Papers*, Cambridge: Cambridge University Press, 1990, 164-224.(Distinguishing semantic from set-theoretic paradoxes, Ramsey effects a great simplification in the theory of types by neglecting the former.)

Tarski, A. (1935) 'Der Warheitsbegriff in den formalisierten Sprachen', *Studia Logica* 1: 261-405; trans. J.H. Woodger (1956), 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(Virtually all modern work on the logic of truth takes this fundamental paper as its starting point.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7; repr. 1994.(The development of all of mathematics on a single axiomatic basis. Discussion of the semantic paradoxes is in the introduction, 41-47 and 60-65.)

Semantics

Semantics is the systematic study of meaning. Current work in this field builds on the work of logicians and linguists as well as of philosophers. Philosophers are interested in foundational issues in semantics because these speak to the nature of meaning, as it embeds in our thinking and in our relations to each other and to the world. Of special interest are questions about how a semantic theory should respect the connections of meaning to truth and to understanding. In addition, numerous semantic problems concerning particular linguistic constructions bear philosophical interest, sometimes because the problems are important to resolving foundational semantical issues, sometimes because philosophical problems of independent interest are expressed using the constructions, and sometimes because clarity about the semantic function of the constructions enables clarity in the development of philosophical theories and analyses.

1 Historical sources and goals

Systematic theorizing about meaning in language always has been of philosophical interest. In addition to being a science with interesting foundational problems, semantics interests philosophers because of their special interests in the nature of meaning and in the meaning of particular sorts of language, notably the language that serves as our primary means of expressing philosophically interesting concepts and thoughts (see Language, philosophy of).

Contemporary semantics owes a great deal to work in formal logic by Gottlob Frege, Bertrand Russell and Alfred Tarski, among others (see Logic, philosophy of). This work has provided techniques for constructing *formal systems* - abstract systems of symbols and sets with rigorously defined logical and structural relationships (see Formal languages and systems). Ludwig Wittgenstein and logical positivists such as Rudolf Carnap found these techniques so compelling that they took certain formal, logical systems to reveal the kind of structure that any coherent language - in particular, any language suited to serious science - must possess at its roots (see Logical positivism). A less extreme but more enduring view is that many aspects of natural languages are aptly *modelled* by formal systems - whether formal systems devised independently by logicians or designed expressly to reveal the systematic structure of the languages in question.

Recently, semantics has profited from the development, under the aegis of Noam Chomsky, of systematic accounts of natural language syntax (see Syntax). Chomsky regards the grammar of a person's language as a partly innate theory that is implicitly mastered in acquiring the language (see Language, innateness of). The linguist's task is to make this theory explicit, guided by the person's spontaneous judgments about which constructions are grammatically acceptable, by experimental evidence about their cognitive faculties, and by general evidence about innate constraints on human grammars. Armed with this conception of the linguist's task, Chomsky and others have developed powerfully sophisticated theories of natural language grammar. Contemporary semanticists find this work useful for a number of reasons. Many semanticists now view their task in a Chomskian light: to uncover partly innate, cognitively real theories. Also, most contemporary semantical theories begin with something very like a grammar: a systematic account of the way sentences of the language in question are constructed from their meaningful parts - and it is quite reasonable to expect that the semantically relevant kind of sentence-structure bears a close relation to the correct grammar for the language.

2 Contemporary semantical traditions

The meaning of a statement is a matter of how things are, according to the statement; and the meaning of any expression is what users attach to it in their competent understanding of it. Semantics, then, ought to respect the outward-looking aspect of meaning: meaning fixes the representational and descriptive powers of language. And semantics must also respect the cognitive and epistemological aspect of meaning: meaning is cognized - it is grasped, known, understood.

A language (whether it be a shared social entity such as English, a narrower dialect, or even an individual's own 'idiolect') contains reusable words and expressions such as words and sentences. These are called expression-types because different tokenings of the expressions (different utterances or inscriptions of them) are importantly similar - they are of the same type (see Type/token distinction). Expression-types in a language have 'timeless' meaning that is drawn on in the various tokenings of the expressions. However, meanings can be expressed in the tokenings

that complete, or go beyond, the timeless meanings of the expressions used. For instance, a particular utterance of the sentence ‘That was worse’ would express a meaning to do with a particular subject matter (‘that’) and would involve (because of ‘was worse’) a gesture to another subject matter and to a dimension of evaluation. An attractive principle is that the timeless meaning of an expression simply is a matter of the expression’s distinctive potential for expressing meanings in particular tokenings. The discipline of pragmatics (whose border with semantics is notoriously porous) seeks to understand the relations between the timeless meaning of expression-types and the meanings achieved by tokenings, as well as the variety of ways tokenings (or speech acts) can achieve communicative and other purposes (see Pragmatics; Presupposition; Propositions, sentences and statements; Speech acts).

A fact about natural languages that is no less impressive for being obvious is that every natural language issues in infinitely many sentences with different meanings. In English, ‘Ann ate a pear’ differs in meaning from ‘Ann ate a pear, and then Ann ate a pear’, which differs too from ‘Ann ate pear, and then Ann ate a pear, and then Ann ate a pear’, and so on. Any acceptable semantic account of a natural language will reveal its meanings as depending systematically on its repeatable, recombinant features (features of its expression-types and of their tokenings). One way to accomplish this is to show how meanings of complex expressions are generated from the meanings of simpler expressions, according to how the simpler expressions are combined. Such a strategy might employ a notion of *semantic values* of the meanings that are assigned to the simpler expressions (or tokenings) in the generation of the meanings of the more complex. That structure of a linguistic expression which, on these accounts, provides the key to the compositional derivation of its meaning sometimes is called the expression’s *logical form*, because of assumptions about the nature of meaning to which we now turn (see Compositionality; Logical form).

The representational side of meaning has led many semanticists to take as their primary goal systematic accounts of the *truth-conditions* of sentences, as well as accounts of such related phenomena as the conditions of true application of predicates (such as ‘is a horse’) and the conditions of the denotation or reference of various noun phrases (such as proper names, definite descriptions like ‘the last Tsar’, and demonstratives like ‘this’). It is immediately plausible that conditions of truth, applicability and reference are proper subjects for a semantic theory. For instance, the meaning of ‘is a felucca’, very plausibly, is a matter of what it is for a thing to be a felucca, which, again plausibly, is given by revealing what it takes for a thing to be truly described as a ‘felucca’. Similarly, the meaning of ‘In Syracuse there are many feluccas’ is elucidated by characterizing how things would have to be to be truly so described (see Meaning and truth).

Because of the widespread view that a complete semantic account of a language must include systematic explanations of what it takes for its expressions to be true, to refer, or to apply truly, the central foundational issues in semantics include the nature of truth and reference, as well as the proper explanation and use within a semantic theory of the notions of truth- and applicability-conditions. Since truth- and application-conditions stand to each other in logical relations such as consistency and entailment, accounts of these conditions can serve what is regarded as another important goal in semantics: to explain the logical relations among linguistic expressions (and among their tokenings) (see Truth, coherence theory of; Truth, correspondence theory of; Truth, deflationary theories of; Truth, pragmatic theory of; Semantic paradoxes and theories of truth; Reference; Sense and reference).

A style of contemporary semantics that can be traced to Frege and Russell employs as semantic values abstract ‘universals’ such as *propositions* and *properties* - entities that embody truth- or application-conditions, in that (unlike linguistic expressions) they are not intrinsically tied to any particular language and have their truth- or application-conditions timelessly and essentially. Thus, some semantical frameworks start with conceptions of the nature of propositions and properties; within these frameworks the principal task in giving a semantical description of a given language is to discover the principles governing which expression-tokenings have as their semantic values which propositions and properties (see Propositions, sentences and statements; Universals; Intensional entities).

Some of these frameworks employ devices from the logical theory of models, in which the meanings of expressions are represented by set-theoretical structures (see Model theory). A particularly influential tradition in semantics deriving from the work of Carnap, Saul Kripke and Richard Montague, uses or imitates the notion of a ‘possible world’: the meaning of a statement is given by the sets of possibilities it excludes and allows, and the

meanings of smaller expressions are their contributions to this semantic function of the sentences in which they figure (see Semantics, possible worlds; Carnap, R.; Montague, R.M.; Kripke, S.A.; Semantics, situation). Among the philosophical insights claimed for this work is the idea that in addition to revealing a statement's simple truth-condition (the condition that the utterance must meet if it is to be true), a semantical theory ought further to reveal its *modal content*, which determines what sorts of possible situation the utterance correctly describes. Focus on this distinction has promoted intense debate about metaphysical necessity and possibility, in particular, on whether certain devices of natural language (such as proper names) presuppose substantive views about necessity and essence, and on whether these views are seriously tenable (see Proper names; Essentialism; Kripke, S.A.).

A somewhat different tradition in semantics has been pioneered by Donald Davidson, employing techniques due to Tarski (see Tarski's definition of truth). The goal in offering a Davidsonian semantics for a given language is to find an appropriate 'truth theory', thought of as a deductive system whose axioms and rules issue in 'T-sentences' such as 'The sentence "Schnee ist weiss" is true if and only if snow is white'. An appropriate truth-theory will not only be correct and complete, but also will in some sense explain the language-user's competence by revealing what it is that the speaker knows in knowing the language (see Meaning and truth; Meaning and understanding §2).

Recently, too, philosophers, linguists and cognitive scientists have developed systematic semantic accounts based on algorithmic or procedural conceptions of meaning. These traditions borrow from the logical theory of proof (see Semantics, game-theoretic; Proof theory).

All semantical traditions, even those that focus heavily on truth-conditions, must aim as well at least for compatibility with satisfactory explanations of the cognitive/epistemological aspects of meaning, including understanding (see Meaning and understanding; Meaning and verification; Sense and reference). One strategy is to keep issues about understanding at a remove from detailed semantic theorizing, by viewing understanding as involved not so much in the nature of meaning as in the uses to which meaning is put in thought and communication. David Lewis, for instance, proposes that a semantic account of a language is simply a correct description of the abstract function that assigns to statements in the language sets of possible worlds as meanings; understanding enters the picture only when we consider what constitutes the language's actually being used. For Lewis, the competent use of a particular language amounts to participation in a certain sort of conventional practice (see Language, conventionality of; Lewis, D.K.). Other strategies take semantic theorizing to be more directly focused on what the competent understanding of a language consists in. Some accounts in the Davidsonian tradition, for instance, submit truth-theories to constraints of cognitive plausibility: a truth-theory must not employ conceptual resources beyond those plausibly required of a competent language-user. Other accounts allow alien concepts in a semantic description of a language and employ a notion of *tacit knowledge* to explain how the semantic description represents what the language-user knows (see Knowledge, tacit). Also heavily cognitive in orientation are accounts within the Chomskian paradigm; these take the primary goal of semantics to be the discovery of the cognitive capacities and structures underlying linguistic competence (see Chomsky, N.). Some semantic accounts take as their task the explanation of how linguistic meaning is fixed by the meaning of mental items such as concepts and thoughts; for the ultimate source of meaning, these accounts defer to theories of cognitive semantics (see Mind, philosophy of; Grice, H.P.; Communication and intention; Concepts; Language of thought; Semantics, conceptual role; Semantics, informational; Semantics, teleological).

3 Issues and problems in semantics

Some debates about the semantic features of particular sorts of expression arise only within particular semantic traditions. Questions of *reference*, in contrast, are of very broad interest, since reference is widely taken to be a semantically central notion. Philosophers have hotly pursued questions about the nature of reference, about the proper role of reference in semantic theories and about how the reference of particular kinds of linguistic item is determined (see Reference; Sense and reference). Receiving special attention have been proper names, natural kind terms (like 'tiger' and 'gold'), context-dependent terms (such as the demonstrative 'that' and the indexical 'today') and theoretical terms of science (such as 'mass' and 'oxygen') (see Proper names; Natural kinds; Demonstratives and indexicals; Scientific realism and antirealism). Among the questions under heavy discussion are the extent to which the determinants of reference are settled by intrinsic features of a speaker's cognition, rather than being settled by features of the speaker's society and environment (see Language, social nature of; Content: wide and

narrow; Methodological individualism).

Issues about the semantical or logical structure of various locutions, in contrast, are often discussed in the context of particular traditions of semantic theory (see Logical form). Still, a number of such issues have interest that is not confined to any one semantic tradition. Among these are the proper treatment of definite descriptions such as ‘the short spy’. Definite descriptions resist easy assimilation either to the paradigm of proper names or to that of quantifying phrases such as ‘all circles’ and ‘three dogs’ (see Descriptions; Quantifiers). Also of persisting interest is the behaviour of ‘anaphoric’ pronouns - ones which in one way or another ‘refer back’ (or even ahead) - such as occur in ‘The dog is happy; he is wagging his tail’ and ‘Probably some dogs can climb trees; it is likely that they are small’. Difficulties in providing systematic explanations of the roles of anaphoric pronouns has led to a number of theories according to which semantic accounts of sentences must view their semantically relevant structure as inextricable from the larger discourse in which the sentences figure (see Discourse semantics; Anaphora). Even the logical form of apparently simple subject-predicate sentences has been debated. For instance, Davidson proposes that the logical form of ‘Bill is running’ involves not the simple predication of an attribute to Bill, but rather an existence claim about an event (perhaps: an event exists, which is a running, which is in progress, and of which Bill is the agent). This account is meant to cohere with a straightforward account of adverbs, in which, for instance, it is clear why ‘Bill is running’ is entailed by ‘Bill is running slowly’ (see Predication; Events; Adverbs).

So-called propositional-attitude ascribing statements (such as ‘Sue believes that oranges are expensive’) also have prompted extensive discussion. An ancient puzzle still has received no widely accepted solution: how is it that Hesperus and Phosphorus are one and the same entity, and yet ‘Hammurabi believed that Hesperus is bright’ might be true while ‘Hammurabi believed that Phosphorus is bright’ is false (see Propositional attitude statements; Indirect discourse; Intensionality)?

Another source of fascinating, but inconclusive, debate among semanticists is the treatment of conditionals (if/then sentences), including indicative conditionals like ‘If Oswald didn’t shoot Kennedy, then someone else did’ and subjunctive, counterfactual conditionals like ‘If Oswald hadn’t shot Kennedy, then someone else would have’. Concerning indicative conditionals, one current view is that their truth-conditions are aptly modelled on those of the material conditional of formal logic, while another is that indicative conditionals have no truth-conditions at all but only conditions of proper assertion (see Indicative conditionals). Counterfactual conditionals have been linked to metaphysical issues about necessity, causation and natural law, and debates about their interpretation are enmeshed with metaphysical debates (see Counterfactual conditionals).

Also among the issues that bear on semantic theories are questions about necessity and possibility, the nature and ontological acceptability of intensional entities, tense, metaphor, vagueness and language apparently about nonexistent and fictional entities (see Modal operators; Modal logic, philosophical issues in; Intensional logics; Intensional entities; Tense and temporal logic; Metaphor; Vagueness; Existence; Fiction, semantics of; Fictional entities; Necessary truth and convention).

See also: Ambiguity; Analyticity; Emotive meaning; Imperative logic; Intuitionistic logic and antirealism; Logical constants; Mass terms; Meaning, Indian theories of; Meaning in Islamic philosophy; Ontological commitment; Questions; Semiotics; Structuralism in linguistics

MARK CRIMMINS

References and further reading

Larson, R. and Segal, G. (1995) *Knowledge of Meaning*, Cambridge, MA: MIT Press. (A lucid and philosophically sophisticated introduction to semantics developing a broadly Davidsonian approach.)

Chierchia, G. and McConnell-Ginet, S. (1995) *Meaning and Grammar*, Cambridge, MA: MIT Press.(An excellent introduction to semantics in the tradition of Montague grammar.)

Semantics, conceptual role

According to conceptual role semantics (CRS), the meaning of a representation is the role of that representation in the cognitive life of the agent, for example, in perception, thought and decision-making. It is an extension of the well-known 'use' theory of meaning, according to which the meaning of a word is its use in communication and, more generally, in social interaction. CRS supplements external use by including the role of a symbol inside *a computer or a brain. The uses appealed to are not just actual, but also counterfactual: not only what effects a thought* does *have, but what effects it* would *have had if stimuli or other states had differed. Of course, so defined, the functional role of a thought includes all sorts of causes and effects that are non-semantic, for example, perhaps happy thoughts can bolster one's immunity, promoting good health. Conceptual roles are functional roles minus such non-semantic causes and effects.*

The view has arisen separately in philosophy (where it is sometimes called 'inferential' or 'functional' role semantics) and in cognitive science (where it is sometimes called 'procedural semantics').

1 Motivations for CRS

There are two quite different projects that go by the name 'semantics'. One, which we might call *linguistic* semantics, deals with the meanings of particular expressions in particular languages and how they fit together to make up meanings of larger expressions. The second project, *metaphysical* semantics, is one of investigating the fundamental nature of meaning, especially what it is about a person that gives their words or thoughts whatever meanings they have in the first place. Conceptual role semantics (CRS) is in the domain of metaphysical semantics: it says that the nature of meaning is functional. It does not have anything very informative to say about linguistic issues, about particular languages or about how a language user works out the meanings of sentences on the basis of the meanings of their component words. But if correct, it can contribute to these enterprises by discouraging false and confused foundational views (see Semantics).

One major motivation for CRS is a functionalist approach to the mind generally (see Functionalism). Functionalism says that what makes a state a mental state and what gives a mental state the specific content that it has is the role it plays in interacting with other mental states in a creature's psychology.

This idea motivates a reply to theories that insist that a mind requires something more. For example, Searle (1980) has argued that computers cannot understand language in virtue of their programs or, more generally, by manipulating symbols in a certain way. He rests his case on a thought experiment, the Chinese room, in which a non-Chinese speaker manipulates Chinese symbols by following rules that do not require him to understand the meanings of the symbols he is manipulating. The rules are so devised that he produces sensible responses in Chinese to any Chinese inputs. Searle says that none the less he does not understand Chinese: he is just mindlessly manipulating symbols. CRS motivates the 'systems reply': if we can programme a computer to be intelligent, it will not be the central processing unit (CPU) all by itself that is intelligent or that understand the symbols, but rather all the complex relations between the CPU and other subsystems of the mind, for example, for perception, reasoning and decision making. So the whole system understands Chinese even if the person who is simulating the CPU does not (see Chinese room argument).

Approaching the matter from the point of view of language rather than thought, what makes CRS plausible is the fact that many terms seem definable only in conjunction with one another. For example, in learning the theoretical terms of Newtonian mechanics - 'force', 'mass', 'kinetic energy', 'momentum' and so on - we do not learn definitions outside the circle. There are no such definitions. We learn the terms by learning how to use them in our thought processes, especially in solving problems. Indeed, CRS explains the fact, noted by Thomas Kuhn (1962), that modern scientists cannot understand the phlogiston theory without learning elements of an old language that express the old concepts. The functional role of, for example, 'principle' as used by phlogiston theorists is very different from the functional role of any term or complex of terms of modern physics, and hence we must acquire some approximation of the eighteenth-century functional roles if we want to understand their ideas (see Definition; Scientific method).

Moreover, CRS does seem to give a plausible account of the meanings of the logical connectives. For example, we

could specify the meaning of ‘and’ by noting that certain inferences - for example, the inferences from ‘*p*’ and ‘*q*’ to ‘*p* and *q*’, and the inference from ‘*p* and *q*’ to ‘*p*’ - have a special status (they are ‘primitively compelling’, in the terminology of Peacocke 1992).

A further motivation for CRS is that it explains a reasonable version of a principle of charity according to which we cannot rationally attribute irrationality to a person without limit (see Charity, principle of §4). Attributing unexplainable irrationality leads to a poor match of roles. If the best translation yields poor enough matches, then the alien conceptual system is not intelligible in ours.

2 Two-factor CRS

Putnam (1975) raised what might seem to be a powerful objection to any CRS. He pointed out that many natural kind concepts, such as ‘water’ and ‘gold’, depend in part for their meaning upon something other than the role of a representation in a person’s head, namely upon what happens to be in their external environment (see Content: wide and narrow; Methodological individualism).

Some proponents of CRS have responded by favouring a ‘two-factor’ version of CRS. On this view, meaning consists of an internal, ‘narrow’ aspect of meaning - which might be handled by functional roles that are within the body - and an external referential/truth-theoretic aspect of meaning, which might be handled by some other metaphysical theories of meaning (for example, a causal one). According to the external factor, ‘Superman flies’ and ‘Clark Kent flies’ are semantically the same since Superman = Clark Kent; it is the internal factor that distinguishes them. But the internal factor counts ‘Water is more greenish than bluish’ as semantically the same in my mouth as in the mouth of my twin on twin earth (see Content: wide and narrow §2); in this case, it is the external factor that distinguishes them.

Two-factor theories gain some independent plausibility from the need for them to account for *indexical* thought and assertions, assertions whose truth depends upon facts about when and where they were made and by whom (see Content, indexical). For example, suppose that you and I say ‘I am ill’. One aspect of the meaning of ‘I’ is common to us, another aspect is different. What is the same is that our terms are both used according to the rule that they refer to the speaker; what is different is that the speakers are different. White (1982) generalized this distinction to apply to the internal and external factors for all referring expressions, not just indexicals.

In a two-factor account, the conceptual roles stop at the skin in sense and effector organs; they are ‘short-arm’ roles. But CRS can also be held in a one-factor version in which the conceptual roles reach out into the world - these roles are ‘long-arm’. Harman (1987) has advocated a one-factor account which includes in the long-arm roles much of the machinery that a two-factor theorist includes in the referential factor, but without any commitment to a separable narrow aspect of meaning.

3 Criticisms of CRS

Error. Actual conceptual roles involve errors, even dispositions to err. For instance, in applying the word ‘dog’ to candidate dogs, one will make errors, for example, in mistaking coyotes for dogs (see Fodor 1987). This problem arises in one form or another for all naturalistic theories of truth and reference, but in the case of CRS it applies to erroneous inferences as well as to erroneous applications of words to things. Among all the conceptual connections of a symbol with other symbols, or (in the case of long-arm roles) with the world, which ones are correct and which ones are errors? Saul Kripke (1982), for example, wonders what distinguishes someone who mistakenly says ‘ $57 + 65 = 5$’ from someone who says it correctly, meaning by ‘Čplus;’ a function that agrees with addition except in yielding a value of 5 with 57 and 65 as arguments. The answer a person gives in the two cases could be the same, correct in one and erroneous in the other.

Some think we can solve the problem by appealing to dispositions to ‘correct’ previous answers, or to ‘correct’ those corrections. But others wonder why all these dispositions could not be the same for two persons who use ‘Čplus;’ to designate different functions. (The problem of error is sometimes said to be the problem of specifying semantic ‘norms’, although norms in this sense should not be confused with norms in the sense of how one ought to apply a word; see Horwich 1994.) Another line of reply is to attempt to specify some sort of naturalistic idealization which specifies roles that abstract away from error, in the way that laws of free fall abstract away from friction.

Words/world. Fodor criticizes a computer-oriented form of CRS for confusing what words denote with the words themselves. The functional roles in the target version of CRS stress searching data banks and manipulating representations, and this Fodor says is like claiming that the meaning of ‘Napoleon won at Waterloo’ is a set of instructions for finding that sentence in a book in the New York Public Library. All such a search yields is more words: we never get the semantic values of those words, namely Napoleon or Waterloo. But, the CRS theorist says in response, long-arm roles include causal chains outside the machine. And the two-factor version of CRS relies on a second factor, the referential factor, to explain the relation between the word ‘Napoleon’ and Napoleon.

CRS is often criticized from the point of view of truth-conditional theories of meaning (see Meaning and truth). If the meaning of a sentence is its truth-conditions, then the meaning cannot be its conceptual role. But with the two-factor theory, proponents of CRS have the option of counting meanings as the same or different in accordance with whether the external factor specifies truth-conditions that are the same or different. Further, there is reason to suppose that meaning is more fine-grained than truth-conditions. For example, the truth-conditions of ‘I am happy’ and ‘Ned is happy’ are the same (since I am Ned), but the meanings of those sentences differ. The further machinery involved in the internal factor can capture the differences among sentences with the same truth-conditions.

Sensory properties. Fodor also criticizes CRS for giving the wrong account of how I and Helen Keller (who was blind and deaf from an early age) can mean the same thing by, for example, ‘Water tastes great’. After all, none of her thoughts bears the same relation to the evidence of sight and sound that mine do. But here Fodor assumes that CRS only has the resource of appealing to similarity in *inferential* role, which is entirely internal. He disparages such an account in favour of a referential view: we mean the same because our concepts of water are concepts of the same *thing*. But a two-factor CRS, relying in part on a referential component, has the option of giving exactly the same account as can a long-arm one-factor account.

What glues the two factors together. Fodor and Lepore (1992) object to the two-factor account, wondering what glues the two factors together. Why can there not be a sentence that has the inferential role of ‘Water is greenish’ but is true if and only if 3 is a prime number? But there is nothing in the CRS approach that dictates that there is any restriction at all on what roles can go with what truth-conditions. This is an independent question that both proponents and opponents of CRS can ask. Everyone who accepts the existence of inferential roles and truth-conditions should find the question meaningful, whether or not they think these are two factors of meaning.

4 Criticisms of CRS (cont.)

Holism. CRS is often viewed as essentially holistic, but the CRS theorist does have the option of regarding some proper subset of the functional roles in which an expression participates as the ones that constitute its meaning. Thus the subset could be taken to be those that are analytic (or ‘true by virtue of meaning’); or as the primitively compelling inferences (Peacocke 1992) plus those generated by them; or the explanatorily basic regularities (Horwich 1994).

One natural and common view of what distinguishes the meaning-constitutive roles is that they are analytic, or played by an expression by virtue of its meaning, as in the case of an inference from ‘bachelor’ to ‘male’. Proponents of CRS are thus viewed as having to choose between accepting holism and accepting that the distinction between the analytic and synthetic is scientifically respectable, a claim that has been seriously challenged by Quine (1954) (see Analyticity). Indeed, Fodor and Lepore (1992) argue that, lacking an analytic/synthetic distinction, CRS is committed to semantic holism, regarding the meaning of any expression as depending on its inferential relations to *every* other expression in the language (see Holism: mental and semantic). This, they argue, amounts to the denial of a psychologically viable account of meaning.

Proponents of CRS can counter as follows. First, there is a question of whether a meaning-constitutive inference is thereby analytic. *If* what is meaning-constitutive is analytic, then holistic versions of CRS need analyticity too, since they regard all inferences as meaning-constitutive. But if what is meaning-constitutive is *not* thereby analytic, then neither holistic nor non-holistic versions of CRS need analyticity. So analyticity is not the issue between holistic and non-holistic versions of CRS.

Second, proponents of CRS can reply that the view is not committed to regarding what is meaning-constitutive as

analytic. In terms of our earlier two-factor account, they can, for example, regard the meaning-constitutive roles as those that are explanatorily basic in a *narrow* psychology: they are the rules that explain other rules of use and determine *narrow* content (Horwich 1994). Narrow content does not involve *truth*-values; these arise only with regard to *wide* content, and so *a fortiori* it does not involve any commitment to *truth by virtue of meaning alone*.

A third approach to accommodating holism with a psychologically viable account of meaning is to substitute close enough similarity of meaning for strict identity of meaning. That may be all we need for making sense of psychological generalizations, interpersonal comparisons and the processes of reasoning and changing one's mind.

Compositionality. Fodor and Lepore (1992) raise a further worry that links the metaphysical semantic issue with a linguistic one: a CRS would seem to risk violating 'compositionality', that is, the requirement that the meaning of a complex expression be a function (in the mathematical sense) of the meanings of its parts (see Compositionality). It is widely thought that such a property of both language and thought is required to explain how human beings seem to able to grasp indefinitely many ever more complicated thoughts, and how they can learn to understand complex sentences on the basis of simple ones. CRS threatens this principle, since, Fodor and Lepore say, the conceptual role of a complex non-idiomatic representation is not always a function of the conceptual roles of its parts. Someone who thinks that rattling snakes, especially, are dangerous is disposed to infer 'This is dangerous' from 'This is a rattling snake' for reasons that may not depend at all on any inferences they are disposed to make from 'This is rattling' or 'This is a snake' separately.

Advocates of non-holistic versions of CRS should regard the argument's assumption that all inferences are to be included in inferential roles as question-begging. Non-holistic versions of CRS can deal with compositionality by counting only a subset of inferences as meaning-constitutive. As mentioned above, these inferences could be identified as the analytic ones, the explanatorily basic ones, or as those that are primitively compelling or generated by them. The threat to compositionality can be avoided by not counting the inference from 'This is a rattling snake' to 'This is dangerous' as part of the meaning-constitutive roles of either sentence.

Advocates of holistic versions of CRS may wish to go along with Fodor and Lepore in assuming that all inferences are part of inferential roles. They should point out that the inferential role of 'rattling' and 'snake' is a matter not just of their roles in *isolation* from one another, but also their roles in contexts involving 'rattling' and 'snake' *together*. The 'rules of use' of these terms are context-sensitive, not context-free.

Once we allow context-sensitive rules of use, compositionality can be trivially satisfied. For example, we can characterize the meaning of a word as an ordered pair, $\langle X, Y \rangle$, where *X* is the set of inferences *to* sentences containing the word and *Y* is the set of inferences *from* sentences containing the word. This is a holistic version of the view, for it includes the inference from 'rattling snake' to 'dangerous' in the meaning of 'rattling' and 'snake', and this example stands proxy for the inclusion of *every* inference in the meaning of every word involved in those inferences. Now the roles just mentioned satisfy the requirements of compositionality from a metaphysical point of view without being a psycholinguistic or a linguistic theory of the representations on the basis of which language is learned or sentences are understood.

5 Framework, not theory

CRS is more of a framework for a theory than an actual theory. There is no agreement among proponents of this framework about how the roles are constituted. By actual causal interactions among thoughts? All? Some? If some, which ones? And what about systematically mistaken inferences (for example, the 'gambler's fallacy')? Do widespread cognitive illusions contribute to the determination of meaning? Or are the roles normative? If the roles are idealized to avoid mistakes, how is the idealization supposed to be understood? Inference can be understood in intentional terms or in purely causal terms, and the latter would be preferable from the point of view of avoiding circularity in specifying roles. And is there any way to distinguish correcting an old practice from changing to a new one (Kripke 1982)? Many successful philosophical theories are quite sketchy. Some say that CRS is no worse than many of them, but others say that the problems in filling in these details involve difficulties that are fatal to the whole project.

See also: Concepts

NED BLOCK

References and further reading

Block, N. (1987) 'Functional Role and Truth Conditions', *Proceedings of the Aristotelian Society* 61: 157-81. (Argues that two-factor and one-factor theories do not differ substantively.)

Field, H. (1977) 'Logic, Meaning and Conceptual Role', *Journal of Philosophy* 69: 379-408.(Two-factor version of CRS based in conditional probability.)

Fodor, J. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.(Raises the 'error' or 'disjunction' problem for theories of meaning.)

Fodor, J. and Lepore, E. (1992) *Holism: A Shoppers' Guide*, Oxford: Blackwell.(A critique of arguments for holism. Chapter 3 is concerned with CRS.)

Harman, G. (1987) '(Non-Solipsistic) Conceptual Role Semantics', in E. Lepore (ed.) *New Directions in Semantics*, London: Academic Press.(A defence of a one-factor CRS.)

Horwich, P. (1994) 'What it is Like to be a Deflationary Theory of Meaning', in E. Villanueva (ed.) *Philosophical Issues*, vol. 5, *Truth and Rationality*, Atascadero, CA: Ridgeview.(Shows how use theories can accommodate the truth-conditional, normative and compositional aspects of meaning.)

Kripke, S. (1982) *Wittgenstein: On Rules and Private Language*, Oxford: Blackwell.(Influential discussion of the 'error' problem, which Kripke claims to find in Wittgenstein.)

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(A famous argument that the history of science is a series of routine periods punctuated by revolutions in which the scientific community changes its conception of the problems and of what the criteria are for a solution.)

Miller, G. and Johnson-Laird, P. (1976) *Language and Perception*, Cambridge, MA: MIT Press.(A CRS proposal in psychology.)

Peacocke, C. (1992) *A Theory of Concepts*, Cambridge, MA: MIT Press.(A CRS-oriented account of the nature of concepts.)

Putnam, H. (1975) 'The Meaning of "Meaning"', in K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science*, vol. 7, *Language, Mind and Knowledge*, Minneapolis, MN: University of Minnesota Press.(The source of the famous twin earth thought experiment.)

Quine, W.V. (1954) 'Carnap and Logical Truth', in *Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press; 2nd edn, 1976, 107-32.(Challenges the claim that the distinction between the analytic and synthetic is scientifically respectable.)

Searle, J. (1980) 'Minds, Brains and Programs', *Behavioral and Brain Sciences* 13 (4): 610-42.(Presentation of the 'Chinese room argument' against computational theories of mind and meaning.)

White, S. (1982) 'Partial Character and the Language of Thought', *Pacific Philosophical Quarterly* 63: 347-65. (The source of the view of narrow meaning based on Kaplan's theory of demonstratives.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(Source of the view that 'the meaning of a word is its use', an important inspiration for CRS.)

Woods, W. (1981) 'Procedural Semantics as a Theory of Meaning', in A. Joshi, B. Webber and I. Sag (eds) *Elements of Discourse Understanding*, Cambridge, MA: MIT Press.(A CRS proposal in artificial intelligence.)

Semantics, game-theoretic

Game-theoretic semantics (GTS) uses concepts from game theory to study how the truth and falsity of the sentences of a language depend upon the truth and falsity of the language's atomic sentences (or upon its sub-sentential expressions). Unlike the Tarskian method (which uses recursion clauses to determine satisfaction conditions for nonatomic sentences in terms of the satisfaction conditions of their component sentences, then defines truth in terms of satisfaction), GTS associates with each sentence its own semantic game played on sentences of the language. This game defines truth in terms of the existence of a winning strategy for one of the players involved. The structure of the game is determined by the sentence's structure, and thus the semantic properties of the sentence in question can be studied by attending to the properties of its game.

For each sentence *S* of the language, there is a semantic game G(*S*) between two players, the defender of *S* and the attacker. *S* is true when the defender has a way of winning the game, no matter what the attacker does - that is, when the defender has a winning strategy in G(*S*). *S* is false when the attacker has a winning strategy. The defending player defends a disjunction by selecting a disjunct to defend, since the truth of a disjunction requires only a single disjunct to be true. Since the truth of a conjunction requires both its conjuncts to be true, the attacker attacks a conjunction by selecting a conjunct for the defender to defend. This means that the defender must be able to defend both conjuncts, if the conjunction is to be true. To defend a negation $-S$, it suffices to show *S* false, so the defender of $-S$ is the attacker of *S*, and vice-versa.

A play of G(*S*) eventually comes down to a single atomic sentence *A*. Since the truth-values of nonatomic sentences depend on those of atomic ones, G(*S*) is played against the background of a specified assignment of truth-values to the language's atoms. If *A* is true according to the given assignment, the defender of *A* wins this play of G(*S*), but if it is false, the attacker wins. Truth of *A* is not necessary in order for the initial sentence *S* to be true; for instance, the defender may defend a disjunction unwisely by picking a false disjunct, even if the disjunction is true. *S* is true when the defender of *S* has a winning strategy, which means that there is a way to force any play of G(*S*) to end in victory for that player.

To show an existential quantification to be true, it suffices to find a single true instance. To show a universal quantification false, it suffices to find a single false instance. Thus the defender of an existential sentence selects an individual as an instance of the existential claim, and the attacker of a universal selects an instance of that universal. Again, the universal is true only if the defender can defend every instance that the attacker might select.

When applied to an ordinary first-order language (with no restrictions imposed on strategies available to the players), the assignment of truth-values to sentences produced by game-theoretic semantics (GTS) - on the basis of an assignment of values to the atoms - coincides with the results of the more familiar semantic frameworks. The usefulness of GTS appears when attention turns to natural language, and also when restrictions are imposed on available strategies in games for formal languages.

Unlike most formal languages, English is structurally ambiguous in ways that a semantics must treat (see Ambiguity). For instance, 'Everyone loves someone' is ambiguous between two readings, $(\forall x)(\exists y)(x \text{ loves } y)$ and $(\exists y)(\forall x)(x \text{ loves } y)$, with the former reading strongly preferred. This ambiguity arises in GTS because there are distinct games for the sentence. In one game, the universal quantifier 'everyone' is processed first, thus receiving a wide-scope interpretation so that the first reading is produced. In the second, the existential quantifier 'someone' is processed first. The first reading is preferred because 'everyone' commands 'someone' in the sentence. ('Command' refers to the structural relation of the two quantifiers: the first is 'higher' in the sentence than the second). Thus GTS for English involves a principle that commanding operators must be processed first. This is an 'ordering principle' which governs the structures of semantic games so as to produce preferred readings of ambiguous sentences. Some ordering principles can be absolute, others (like the example above) can be defeasible. Another defeasible ordering principle is that operators should be processed following the left-to-right order of their occurrence in the sentence in question. This principle also applies to 'Everyone loves someone'.

Unlike many approaches to the semantics of English, GTS is not compositional. The interpretation of a sentence need not be determined in a 'cumulative' way based on the predetermined meanings of its components. This

allows GTS to treat various semantic phenomena that do not respond to compositional semantic analysis (see Compositionality).

GTS applied to English yields interesting and novel treatments of such well-known semantic problems as anaphora (determining the conditions for co-reference of pronouns and related singular terms and quantifiers), conditionals, intentional identity, negation and nonstandard informational relationships among operators in a sentence. This can be seen in so-called 'branching quantifiers' and in the interpretation of relational questions. For example, the branching sentence:

$$\left.\begin{array}{l}\forall x \exists y \\ \\ \forall z \exists w\end{array}\right\rangle Fxyzw$$

has *y* within the scope of *x* and *w* within the scope of *z*, but *y* is not within the scope of *z*, nor *w* within the scope of *x*. This pattern of quantifier relationships cannot be represented in a linear first-order quantifier prefix. The semantic game for this sentence is like the game for $\forall x \exists y \forall z \exists w Fxyzw$, except that the game for the branching sentence is a game of imperfect information. The defender of the sentence is ignorant of certain previous moves in the game, whereas the defender of the linear sentence knows what has transpired in the game for that sentence (see Quantifiers §2).

Branching quantification is a well-understood example of informational independence of operators. Informational independence is currently the object of intense study in GTS, which has led to the formulation of 'independence-friendly' logics and the discovery of conditions governing informational variability. Other areas currently of great interest include negation in English and the role of subgames in interpretation. The subgame of a semantic game is a self-contained game serving as a component of the larger game, so that the larger game involves results of the subgame in its play. For example, interpretation of a conditional requires information concerning the semantic game for its antecedent, in order to interpret pronominal connections between the antecedent and consequent. Recent developments have generalized the subgame idea, using 'tangent games' to provide a semantics of relative clauses and to link them to homophonic embedded questions.

See also: Decision and game theory; Semantics

MICHAEL HAND

References and further reading

Hand, M. (ed.) (1994) *Synthèse* 99 (3), special issue.(Special GTS issue containing important new work on negation, informational variability, games with chance moves and other topics.)

Hintikka, J. (1976) *The Semantics of Questions and the Questions of Semantics*, Amsterdam: Acta Philosophica Fennica.(An early, extensive semantic investigation of questions.)

Hintikka, J. (1983) *The Game of Language: Studies in game-theoretical semantics and its applications*, Netherlands: Kluwer.(Papers by Hintikka - some in collaboration with Kulas - on negation, tense and other topics.)

Hintikka, J. and Kulas, J. (1985) *Anaphora and Definite Descriptions*, Dordrecht: Reidel.(A GTS account of pronominal relations.)

Hintikka, J. and Sandu, G. (1990) *On the Methodology of Linguistics*, Dordrecht: Reidel.(Focuses largely on a GTS account of pronominal relations.)

Saarinen, E. (1979) *Game-Theoretical Semantics*, Dordrecht: Reidel.(A collection of early papers in GTS, including Hintikka's influential discussions of branching quantification, Hintikka and Carlson's original discussion of subgames in the semantics of conditionals, Saarinen on backward-looking operators and intentional identity, and Rantala's urn models.)

Semantics, informational

Information-theoretic semantics (ITS) attempts to provide a naturalistic account of the conditions under which a psychological state such as a belief or desire has a particular mental content: what it is by virtue of which, say, a psychological state is a belief 'that it is raining' or a desire 'that it stop raining'. Because of the complexities of an entirely general account, ITS typically attempts to provide merely a sufficient *naturalistic condition for a belief content of the sort normally acquired by perception (for example, that it is raining). It is expected that other sorts of mental contents may require that ITS be supplemented in various ways.*

ITS was inspired by Claude Shannon's theory of 'information' (1948), which provided a mathematical measure of the amount of information carried by a signal. Employing a notion of 'natural meaning' discussed by Peirce (1931) and Grice (1957), Dretske (1981) supplemented Shannon's work with an account of what *information a signal carries. The intuitive idea is that a signal carries the information 'that* p*' if and only if it naturally means (that is, indicates) that* p*, as when smoke 'means' there is fire.*

Natural indication is a key ingredient in ITS accounts of mental content. In their accounts, Stampe (1977) and Stalnaker (1984) appeal to the notion of what a state indicates under 'optimal' conditions. Fodor (1987) appeals to 'asymmetric dependencies' between the meaning-forming and the non-meaning-forming indication conditions in the causation of psychological states. Dretske (1988) appeals to the idea that, via operant conditioning, a state can acquire a functional role vis-à-vis *behaviour because it naturally indicates 'that* p*' and thereby can acquire the natural function of indicating 'that* p*'.*

1 Content and indication

In an early proposal, Dretske (1981) maintained that a state, *S*, carries the information 'that *p*' (that is, has the informational content 'that *p*') if and only if the probability of *S*, given *p*, is equal to 1 (see Information theory). One issue this proposal raises is whether the probability is 'objective' or 'subjective' (a matter of degrees of belief) (see Probability, interpretations of). If the notion of information content is to be suitable for a naturalistic account of mental content, the notion of probability invoked must be objective. But, as Dretske himself pointed out, the relevant notion of probability is not relative frequency. Moreover, it seems not to be propensity either: the direction of propensity is the direction of causation; but the conditional probability Dretske would have us consider is the opposite direction: the probability of there being 'fire', given that there is 'smoke'. In a later proposal, Dretske (1988) employed a counterfactual: the occurrence of a state, *S*, carries the information that *p* if and only if *S* would not have occurred unless *p*. He claimed that such counterfactuals express objective, mind-independent facts.

2 Two problems

Error. A state carries the information 'that *p*' if and only if it indicates (or naturally means) 'that *p*'. But a state indicates 'that *p*' only if *p* is indeed the case. So information, in the sense of natural indication, does not admit of misinformation, or error. However, belief does: one can believe 'that *p*' even when *p* is false. How can information-theoretic semantics (ITS) accommodate this fact?

Fine-grained individuation. Moreover, even when a belief 'that *p*' indicates 'that *p*', the indication relation will not uniquely pair the belief with *p*. The belief 'that *p*' can differ from the belief 'that *q*', even when *p* and *q* are logically equivalent. For example, the belief that 'there is a chair in the room' has a different content from the belief that 'there is a chair in the room and it is either an antique or not an antique', even though these contents are logically equivalent. However, if a state indicates *p*, then it indicates anything that is logically equivalent to *p*. Indeed, if a state indicates 'that *p*', then it indicates anything that is merely nomologically equivalent to *p*: for example, anything that indicates the presence of a renate (creature with kidneys) indicates the presence of a cordate (creature with a heart). Moreover, if something indicates 'that *p*', then it indicates anything that logically or even nomologically follows from *p*: anything that indicates that something is a cow indicates that it is a herbivore. Yet the belief that something is a renate is different from the belief that it is a cordate, a belief that something is a cow is different from the belief that it is a herbivore.

3 Three proposals

Optimal conditions. Stampe (1977), Fodor (1990b) and Stalnaker (1984) try to accommodate the possibility of error or misrepresentation by maintaining that the content of a psychological state is what the state would indicate under 'optimal' conditions. Stalnaker's proposal is that a belief state has the content 'that *p*' if and only if, in optimal conditions, the state indicates *p,* and the subject is in the state because *p* or because of something that entails *p*. The idea is that in cases of false belief, conditions are not optimal (see Idealizations).

One problem is that it is by no means clear that the notion of ideal conditions is suitably naturalistic. Belief fixation is holistic: what someone would come to believe in some situation depends on what they already believe about other things. Even in a good light, someone presented with a horse might not think it is one if they think appearances are misleading. If optimal conditions must include the subject's being in belief states with certain contents, then, as an account of belief content, ITS is circular.

Asymmetric dependencies. Fodor (1987) formulates his ITS theory as one for mental symbols - expressions in a 'language of thought' (see Language of thought). He calls tokens of mental symbols meaning 'cow' that are caused in the absence of cows 'wild', and calls the property whereby symbols can mean things that on occasion are not causes of their tokening 'robustness'. He points out that in solving problems of robustness and error, ITS needs to solve the 'disjunction' problem: given that among the causes of a symbol's tokenings, there are both meaning-forming and wild causes, what distinguishes them? In particular, what makes it true that some symbol *F* means [horse] and not [horse or cow on a dark night] or [horse or cow on a dark night or w_2 or w_3 or…] (where each w_i is a property the exemplification of which could cause the tokening of *F*)?

Fodor speculates that a state's being caused by wild conditions depends upon its being caused by meaning-constitutive ones, but not vice versa: ill-lit cows causing horse thoughts depends upon well-lit horses causing them, but well-lit horses causing them does not similarly depend upon the ill-lit cows doing so (getting things wrong depends upon getting things right in a way that getting things right does not depend upon getting things wrong). Fodor draws from this speculation the proposal that mention of ideal conditions is inessential: the structure of this asymmetric causal dependency alone, abstracted from any specific conditions or causal chains, can do all the work.

One problem with this suggestion is that there are many asymmetric dependencies even within meaning-forming cases: for example, small horses causing horse thoughts might asymmetrically depend upon all horses doing so, but not vice versa. If this were so, Fodor would be committed to claiming that 'horse' meant 'non-small horse'.

Another problem is that, in avoiding any mention of mentality, the account risks gratuitous meanings that are brought about by the physics of the world but have no cognitive significance for the agent: electrical stimulation by a poking neurosurgeon, or by cosmic rays, causing tokenings of a mental state in one set of circumstances (for example, when potassium levels were high), might depend upon their lawfully causing those tokenings under other circumstances (when sodium levels were low, for example), but not vice versa. Fodor's view would seem to be committed to treating these further phenomena as meanings, albeit of no cognitive relevance to the agent.

Fodor (1987, 1990a) has replied to these objections with considerable ingenuity. However, what is needed is not only to rule out cases one-by-one, but some general reason to believe that the relevant notion of asymmetric dependency can be explicated without appeal to any mentalistic notions.

Indicator functions. Dretske's most developed ITS proposal is that a mental state has the content 'that *p*' if and only if it is has the *function* of indicating *p*. This distinguishes a proper subset of natural indications; moreover, errors arise when the state fails to indicate what it has the function of indicating. The problem of consequences is handled by observing that from the fact that a state has the function of indicating 'that *p*', and *p* implies *q*, it does not follow that the state has the function of indicating that *q* (see Semantics, teleological; Functional explanation).

The main burden of this approach is to say what makes it the case that a state has a certain indicator function. Dretske offers an account of how a state can acquire such a function, and uses it to offer an account of what he calls 'proto-beliefs' and 'proto-desires'. Suppose that a state *B* of an organism indicates that some property or kind *F* (for example, water) is present in the organism's vicinity, and that a state *D* of the organism renders it receptive to a reward *R*. Then, when in *D*, the organism's behaviour can be reinforced by *R*. Suppose further that when *B* and

D jointly contribute to producing a movement *M* in circumstances in which *F* is present, this results in the organism's receiving *R*. Then, through a process of operant conditioning, *B* and *D* can come to have the 'control' duty of producing *M*. State *B* can thus acquire the control duty because it indicates *F*, and *D* can thus acquire this duty because it is a state of receptivity for *R*. According to Dretske, if *B* and *D* were recruited in this way for the control duty of producing some movement, *B* has the natural function of indicating that *F* is present, and *D* counts as a proto-desire for *R*.

A problem for this account is whether it can be naturalistically supplemented to yield an account of beliefs and desires themselves, since pairs of beliefs and desires do not generally have a specific control duty *vis-à-vis* movements: desiring water and believing there is some in the brook, does not necessarily give rise to any specific behaviour. And, of course, this account is inapplicable for beliefs and desires not acquired by operant conditioning (see Behaviourism, analytic; Behaviourism, methodological and scientific).

Moreover, appeals to function arguably fail to solve the problem of fine-grained individuation. If *p* and *q* are logically equivalent or even nomologically equivalent, then the mechanisms of operant conditioning will be insensitive to their difference (see Intentionality; Propositional attitude statements). Consequently, a state will have the natural function of indicating 'that *p*' if and only if it has the natural function of indicating 'that *q*'. Indeed, whenever *p* and *q* are such that the mechanisms of operant conditioning are insensitive to their difference, whatever has the natural function of indicating *p* has the natural function of indicating *q*. Thus, if, for instance, such mechanisms are insensitive to the difference between the presence of a 'rabbit' and the presence of 'undetached rabbit parts' (to borrow a famous example from Quine 1960), it seems that a state will have the function of indicating that a rabbit is present if and only if it has the function of indicating that undetached rabbit parts are present.

4 Two-factor theories

Since Brentano, it has seemed to many philosophers undeniable that the mind makes more distinctions than even all possible worlds provide: the mind can conceive the impossible, and distinguish among even necessarily co-instantiated properties (like 'being an equiangular triangle' and 'being an equilateral triangle'). 'Two-factor' theories claim that there is something about a mental state in addition to what is captured by ITS that supplies the additional distinctions, notably, its structure or its conceptual role. Fodor (1990a), doubting that conceptual roles can be specified in a way immune to Quine's criticisms of the analytic/synthetic distinction, holds that what distinctions are needed can be captured by structures specified in a language of thought. Others (Block 1986, for example) think that once conceptual roles are specified purely internally, independently of issues of truth, Quine's criticisms are no longer a worry. Most defenders of ITS agree that capturing the contents of logically complex, non-perceptual beliefs (for example, that 'not every particle retains charge if divided') requires reference to some internal factor (see Semantics, conceptual role §3; Concepts §10).

See also: Content, non-conceptual; Information theory; Information theory and epistemology; Semantics; Semantics, teleological

BRIAN P. McLAUGHLIN
GEORGES REY

References and further reading

Block, N. (1986) 'Advertisement for a Semantics for Psychology', in P. French, T. Uehling, Jr., and H. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 10, *Studies in the Philosophy of Mind*, Minneapolis, MN: University of Minnesota Press.(Defence of a 'two-factor' theory of content, allowing for information-theoretic semantics and conceptual role semantics.)

Boghossian, P. (1989) 'The Rule-Following Considerations', *Mind* 98: 507-50. (Raises the holism objection to information-theoretic semantics accounts that appeal to optimal conditions, and argues that ITS accounts cannot capture the normativity of mental content.)

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press/Bradford Books. (Dretske's original proposal of information-theoretic semantics.)

Dretske, F. (1988) *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: MIT Press/Bradford Books.(Dretske's later development of information-theoretic semantics.)

Fodor, J. (1984) 'Semantics, Wisconsin Style', *Synthese* 59: 231-50.(Criticism of Stampe- and Dretske-style

theories (see Stampe 1977; Dretske 1981) from a fellow endorser of the information-theoretic semantics strategy.)

Fodor, J. (1987) *Psychosemantics*, Cambridge, MA: MIT Press/Bradford Books. (Source of the 'asymmetric dependency' proposal referred to in §3.)

Fodor, J. (1990a) *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press. (Criticisms of conceptual role theories of content and a defence of his asymmetric dependency proposal.)

Fodor, J. (1990b) 'Psychosemantics, or Where Do Truth Conditions Come From?', in W. Lycan (ed.) *Mind and Cognition: A Reader*, Oxford: Blackwell, 312-38.(A vigorous defence of an 'optimal conditions' version of information-theoretic semantics; widely circulated as a manuscript in 1980 and published in 1990 despite Fodor's repudiation of the view (see Fodor 1987).)

Gates, G. (1996) 'The Price of Information', *Synthese* 107: 325-47.(Forceful criticism of information-theoretic semantics on the basis of the argument of Quine 1960.)

Grice, H. (1957) 'Meaning', *Philosophical Review* 80: 377-88.(Famous effort to distinguish 'natural' from 'non-natural' meaning, aimed against ITS-style theories.)

Loewer, B. (1986) 'From Information to Intentionality', *Synthese* 2: 287-317.(Makes the point about objective probability and natural meaning made in §1; raises the holism objection to accounts using information-theoretic semantics that appeal to optimal conditions.)

Loewer, B. and Rey, G. (eds) (1991) *Meaning in Mind: Fodor and His Critics*, Oxford: Blackwell.(Criticisms by a number of philosophers of Fodor's theory of mind generally, but especially of his 'asymmetric dependency' theory of content; with replies by Fodor and a useful introduction to his theory by the editors.)

McLaughlin, B. (1986) 'What is Wrong With Correlational Psychosemantics', *Synthese* 2: 271-86.(Raises the holism objection to accounts using information-theoretic semantics that appeal to optimal conditions; points out that such accounts cannot cover mental contents that are satisfied by conditions that cannot obtain in optimal conditions, and thus cannot account for beliefs such as that one is not in optimal conditions.)

McLaughlin, B. (ed.) (1991) *Dretske and His Critics*, Oxford: Blackwell.(Contains several essays on Dretske's information semantics and his replies to these essays.)

Peirce, C.S. (1931) *Collected Papers*, ed. C. Hartshorne and P. Weiss, vol. 2, Cambridge, MA: Harvard University Press.(Perhaps the earliest proposal of an information-theoretic semantics.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(A highly influential discussion of meaning that contains a limited, behaviouristic version of information-theoretic semantics, raising (in chapter 2) the famous 'gavagai' problem for any such theory: how could, for example, 'rabbit' be distinguished from 'undetached rabbit parts'.)

Shannon, C.E. (1948) 'The Mathematical Theory of Communication', repr. in C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, Chicago, IL: University of Illinois Press, 1949.(Presents a mathematical theory of *quantity* of information.)

Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press/Bradford Books.(Defends a version of information-theoretic semantics that appeals to optimal conditions.)

Stampe, D. (1977) 'Toward a Causal Theory of Linguistic Representation', in P. French, T. Uehling, Jr, and H. Wettstein (eds) *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press, 81-102. (One of the original proposals of information-theoretic semantics.)

Semantics, possible worlds

Possible worlds semantics (PWS) is a family of ideas and methods that have been used to analyse concepts of philosophical interest. PWS was originally focused on the important concepts of necessity and possibility. Consider:

(a) Necessarily, $2+2=4$.
(b) Necessarily, Socrates had a snub nose.

Intuitively, (a) is true but (b) is false. There is simply no way that 2 and 2 can add up to anything but 4, so (a) is true. But although Socrates did in fact have a snub nose, it was not necessary that he did; he might have had a nose of some other shape. So (b) is false.

Sentences (a) and (b) exhibit a characteristic known as intensionality*: sentences with the same truth-value are constituent parts of otherwise similar sentences, which nevertheless have different truth-values.* Extensional semantics *assumed that sentences stand for their truth-values, and that what a sentence stands for is a function of what its constituent parts stand for and how they are arranged. Given these assumptions, it is not easy to explain the difference in truth-value between (a) and (b), and hence not easy to give an account of necessity.*

PWS takes a sentence to stand for a function from worlds to truth-values. For each world, the function yields the truth-value the sentence would have if that world were actual. '$2+2=4$*' stands for a function that yields the truth-value 'true' for every world, while 'Socrates had a snub nose' stands for a different function that yields 'true' for some worlds and 'false' for others, depending on what Socrates' nose is like in the world. Since these two sentences stand for different things, sentences that have them as constituents, such as (a) and (b), can also stand for different things.*

This basic idea, borrowed from Leibniz and brought into modern logic by Carnap, Kripke and others, has proven extremely fertile. It has been applied to a number of intensional phenomena in addition to necessity and possibility, including conditionals, tense and temporal adverbs, obligation and reports of informational and cognitive content. PWS spurred the development of philosophical logic and led to new applications of logic in computer science and artificial intelligence. It revolutionized the study of the semantics of natural languages. PWS has inspired analyses of many concepts of philosophical importance, and the concept of a possible world has been at the heart of important philosophical systems.

1 Intensions demeaned

Traditionally, the 'intension' of a predicate was distinguished from its 'extension'; the former is a property, the latter is a set. The predicates 'is a featherless biped that is not a plucked chicken' and 'is human' have (one can imagine) the same extensions but different intensions. (The example is from Bertrand Russell.) Gottlob Frege's concepts of *Sinn* and *Bedeutung* extend this idea: the *Sinn* of a singular term is an identifying condition (or 'individual concept'), the *Bedeutung* the individual designated. The *Sinn* of a sentence is a proposition, the *Bedeutung* a truth-value. Frege defended his choice of truth-values as the *Bedeutung* of sentences on systematic grounds (see Frege, G.; Sense and reference).

As model theory was developed by Tarski and others, a version of Frege's choices for *Bedeutung* became the standard values in 'extensional semantics'. The extension of an *n*-place predicate is the set of *n*-tuples of objects of which the predicate is true (thus, the extension of 'gives' might be the set of those four-tuples containing two persons, an object and a time, such that the first person gives the object to the second person at that time). The extension of a singular (object-denoting) term is the object it designates. The extension of a sentence is a truth-value (see Model theory). The packaging together of the predicate calculus with an extensional semantics proved adequate for important work in mathematical logic and overshadowed older approaches to logic. In contrast, no understanding of intensions emerged that is generally agreed on. In the middle part of the century interest in intensional phenomena waned. In fact, the success of extensional logic led to somewhat uncharitable attitudes towards any non-extensional phenomena. In Quine's influential view non-extensional constructions are not suited for scientific work; they are more in need of regimentation than straightforward analysis (see Quine, W.V.). (Non-extensional constructions are those that apparently distinguish between different phrases or sentences

used to complete them, even though the phrases or sentences have the same extensions. One example is 'Elwood believes that…'. 'Elwood believes that Stanford is east of Hawaii' might be true, while 'Elwood believes that Stanford is east of Berkeley' might be false, even though 'Stanford is east of Hawaii' and 'Stanford is east of Berkeley' both have the same extension (the truth-value 'true'). 'Intensional' is sometimes used simply to mean 'non-extensional', and is sometimes given a narrower meaning.)

2 Modal logic

A number of philosophers and logicians continued to attempt to provide straightforward analyses of intensional phenomena, however. Until the 1950s, the emphasis was on syntactic approaches. A key figure was C.I. Lewis, whose dissatisfaction with the extensional treatment of 'if…, then…' as the material conditional led him first to the logic of 'strict implication', then to 'modal logic', the logic of necessity and possibility (see Lewis and Langford 1932). The modal operators (typically translated 'necessarily' and 'possibly', and usually symbolized as $\Box$ and $\Diamond$) are not truth-functional, and so require intensional analysis.

The language of propositional modal logic (ML) consists of the language of propositional logic, plus the rule that if ϕ is a well-formed formula (wff), then so are $\Box\phi$ and $\Diamond\phi$. Lewis and others worked out a number of axiom systems for ML and studied and compared them proof-theoretically.

More semantically oriented approaches to intensionality emerged later in the century, beginning with Carnap (1946, 1947). One of the most important of Carnap's many contributions to the study of intensionality was to recruit Leibniz's idea that necessary truth was truth 'in all possible worlds' to the task of building an intensional semantics. This is the guiding idea of possible worlds semantics (PWS). Carnap's version of this idea, less straightforward than those that were to follow, relies on linguistic representations of possible worlds which he called 'state-descriptions'.

The basics of the now-standard treatment came in the late 1950s and early 1960s with results obtained by Stig Kanger (1957; also Føllesdal 1994) and Saul Kripke (1959, 1963a, 1963b; also Hintikka 1957, Montague 1974a). We shall look briefly at K, S4 and S5, three among the plethora of axiom systems for modal logic that have been studied.

S5 includes:

- all propositional tautologies and *modus ponens*,
- the definition $\Diamond\phi =_{\mathrm{df}} \neg\Box\neg\phi$,
- the rule of necessitation ($\vdash\phi \;/\; \vdash\Box\phi$, that is, if ϕ is deducible from the null set of premises, then so is $\Box\phi$),
- the axioms

[*K*]: $\Box(\phi\rightarrow\psi)\rightarrow(\Box\phi\rightarrow\Box\psi)$
[*T*]: $\Box\phi\rightarrow\phi$
[*4*]: $\Box\phi\rightarrow\Box\Box\phi$
[*B*]: $\Diamond\Box\phi\rightarrow\phi$

If we drop B we have S4; if in addition we drop *4* and *T* we have $\mathbf{K}$, Kripke's minimal system.

A modal model structure is a pair $\langle K,R\rangle$. *K* is the set of worlds; *R* we will consider later. A modal model will tell us which atomic sentences of the base language *L* are true at which worlds of *K*. For the connectives of propositional logic the rules remain unchanged. To extend the system to include $\Box$ it is natural, on the Leibnizian conception, to use the rule (we symbolize 'ϕ is true in *w*' as $\phi[w]$):

$\Box_A$: $\Box\phi[w]$ iff $\forall w'\phi[w']$.

Given the definition of $\Diamond$ we have:

$\Diamond\phi[w]$ iff $\exists w'\phi[w']$.

The reader can check that on this conception, all of the axioms for S5 are valid, that is, true in every world of every model. Consider *B*, for example. Suppose the antecedent is true at *w*. Then $\exists w'\forall w''\phi[w'']$. Since the existential

quantification is vacuous, this reduces to $\forall w' \phi[w']$. Then, by Universal Instantiation (U.I.), we have $\phi[w]$.

Note that the $\Diamond$ was vacuous; $\Diamond\Box\phi$ collapsed into $\Box\phi$. This is characteristic of S5: iterated modalities collapse to the right.

S5 is a natural logic for metaphysical necessity, which was doubtless the conception Leibniz had in mind. But there are other coherent concepts of necessity, for which some of the axioms of S5 do not seem correct, and for which distinctions among iterated modalities are significant (see Modal logic).

Consider physical necessity. We have $\Box\phi[w]$ if ϕ is true in every world that obeys the laws of physics of *w*. Suppose w' is a world which has all of our laws and more. Certain events may be ruled out by the physics of w' that are not ruled out by our physics, so our world is not physically possible relative to w'. Suppose, for example, that it is a law of physics in w' that no golf ball travels more than 200 yards. Then even though w' is physically possible (it obeys our laws) and 'No golf ball travels over 200 yards' is necessary in w', it is not true that no golf ball travels over 200 yards. So Axiom *B* is incorrect for physical necessity.

T is intuitive in the case of metaphysical and physical necessity, but not for 'deontic logic', in which '$\Box\phi$' is interpreted as 'It ought to be the case that ϕ' (see Deontic logic).

In discussing these alternative conceptions of necessity and possibility, we move from an absolute to a relative conception of possibility, the idea that the set of worlds relevant to issues of necessity varies from world to world. This is the information given by the second member of the model structure above. *R* is a relation on *K*, the 'accessibility relation'. Different accessibility relations correspond to different conceptions of necessity. We replace our absolute rule $\Box_A$ with a relative rule:

$\Box_R$: $\Box\phi[w]$ iff $\forall w'$, if w' is accessible from w, then $\phi[w']$.

The axioms that characterize the various systems of modal logic correspond to the logical properties of the relation *R*. The axiom *K* places no restrictions on it; *T* requires reflexivity; *4* requires transitivity; and *B* symmetry. Thus absolute necessity, captured by S5, is the case where the accessibility relation is an equivalence relation.

3 Other applications

The semantical apparatus developed for modal logic has been used to investigate a number of other logical systems.

In 'epistemic logic', for example, a knowledge operator, indexed by knowers, is patterned after $\Box$; $K_\alpha\phi$ means 'ϕ holds in all of α's epistemic alternatives' (Hintikka 1962; see Epistemic logic).

It is important for the philosophically oriented reader to keep in mind that for the purposes of developing and applying semantical treatments of intensional languages, for example, in completeness proofs, the possible worlds of PWS need not be invested with any important metaphysical properties; they are just indices for models. The basic apparatus has been used to study a number of areas in which the metaphor of a possible world is inapplicable. In dynamic logic, for example, the apparatus of modal logic is applied to programs. The 'worlds' are states of a machine. Accessibility relations are indexed by programs. Where α is a program, $\Box[[\alpha]]\phi$ means 'ϕ holds after every terminating execution of α' (see Pratt 1976; Dynamic logics).

The interplay between semantic structures and logical systems involved in these investigations constitute a development in logic comparable to the move in geometry away from Euclidean geometry, conceived as the one true system, to geometry as the study of alternative axiom systems for spaces with diverse properties.

4 Temporal logic

The apparatus of modal model structures works nicely to provide a semantics for temporal logic - the logic of operators modelled after the tense and temporal adverb systems of natural languages. Let '*G* mean 'It will always be the case' and '*F* mean 'It will sometimes be the case'; '*F* can be defined as '$\neg G \neg$'. Thus *G* is a universal operator, analogous to $\Box$, and *F* is an existential operator, analogous to $\Diamond$. Similarly, let '*H* mean 'It has always been the case' and define '*P*' as '$\neg H \neg$'. Then instead of a set of worlds and an accessibility relation, take a model structure to be a set of moments of time and an ordering relation between them. The need for an accessibility

relation is rather more intuitive here than in the case of necessity and possibility since, unlike worlds, we usually think of times as ordered by the relation of ‘before’. As with modal logic, different logics correspond to different conceptions of the ordering relation. One minimal tense logic (Benthem 1988) contains the axioms:

- $G(\phi \rightarrow \psi) \rightarrow (G\phi \rightarrow G\psi)$
- $H(\phi \rightarrow \psi) \rightarrow (H\phi \rightarrow H\psi)$
- $\phi \rightarrow GP\phi$
- $\phi \rightarrow HF\phi$

and the rules *modus ponens* and the analogue to necessitation, sometimes called ‘eternity’:

- $\vdash \phi / \vdash G\phi$
- $\vdash \phi / \vdash H\phi$

As with modal logic, there is a precise correspondence between ordering conditions and additional axioms. For example $PP\phi \leftrightarrow P\phi$, which seems plausible enough, requires that the structure of moments be transitive (if *t* is before t' and t' is before t'', *t* is before t'') and dense (if *t* is before t', there is a t'' between them, that is, after *t* and before t') (see Prior 1967; Benthem 1988; Tense and temporal logic).

5 Conditional logic

As we noted, dissatisfaction with the material conditional as an explication of the ordinary-language conditional was an early complaint against extensional logic. There is some connection between the antecedent and the consequent that the semantics for the material conditional misses. For one thing, $\phi \rightarrow \psi$ is true whenever ϕ is false, making all counterfactual conditionals trivially true. Let us use ‘ $\Rightarrow$’ as a symbol for a better approximation. Another key way in which $\Rightarrow$ should differ from $\rightarrow$ is that it should not always permit *strengthening the antecedent*. We have:

$(\phi \rightarrow \psi)$ only if $(\phi \wedge \chi \rightarrow \psi)$

but not:

$(\phi \Rightarrow \psi)$ only if $(\phi \wedge \chi \Rightarrow \psi)$.

For example:

ϕ: I put water in my canteen when I start my hike.
ψ: I have water when I stop for a rest.
χ: There is a hole in my canteen.

Robert Stalnaker’s version of Frank Ramsey’s test for evaluating conditionals is ‘make the minimal revision of your stock of beliefs required to assume the antecedent. Then, evaluate the acceptability of the consequent on the basis of this revised body of beliefs’ (Ramsey 1931). Stalnaker (1968) and David Lewis (1973) have proposed analyses that implement this idea within the possible worlds framework:

Stalnaker: $\phi \Rightarrow \psi[w]$ iff ψ holds in the ϕ-world that is closest to *w*.
Lewis: $\phi \Rightarrow \psi[w]$ iff ψ holds in *all* ϕ-worlds which are closest to *w*.

These analyses require a relation of overall similarity or closeness among worlds. (It can be argued that the relation of overall similarity of worlds is vague and context sensitive; it is replied that this captures the vagueness and context sensitivity of the ordinary conditional.)

On either analysis, strengthening the antecedent fails for $\Rightarrow$, because when the antecedent is strengthened, different worlds might be the closest in which the antecedent is true.

The choice between Lewis’ definition and Stalnaker’s depends on such issues as whether there is always a unique closest world. One important principle that turns on this is ‘conditional excluded middle’,

$(\phi \Rightarrow \psi) \vee (\phi \Rightarrow \neg\psi)$

which Stalnaker endorses and Lewis rejects (see Indicative conditionals; Counterfactual conditionals).

6 Quantified modal logic

In 1946, Ruth Barcan Marcus and Rudolf Carnap independently published systems of quantified modal logic (QML), in which principles like the following were considered:

$$\forall x \Box \phi(x) \rightarrow \Box \forall x \phi(x)$$

This is the ‘Barcan formula’, attractive to those who would reduce *de re* to *de dicto* necessity (see De re/de dicto).

Kripke (1963b) has provided a semantics for these systems. A quantificational modal structure $\langle K, R, \Psi \rangle$ adds a function Ψ which assigns a domain of individuals to each possible world in *K*. A model assigns extensions to each predicate at each world. On the natural, ‘world-bound’ interpretation of universal quantification, $\forall x \phi(x)$ is true in a world *w* if and only if in *w* $\phi(x)$ is true of every member of the domain of *w*.

If we suppose, as is natural, that worlds have varying domains (that objects that might have been in our world, for example, actually exist in other worlds), then it seems the Barcan formula is not valid. Even if every object we find in the actual world is ϕ in all possible worlds, there could be objects in other possible worlds that are not ϕ. But the formula can be validated if we interpret quantifiers as ranging over all possible objects, or if we suppose that the domain is constant across worlds; each of these alternatives have found proponents.

‘Quantifying in’ (that is, quantification across modal operators as in the antecedent of the Barcan formula) was deemed by Quine to put us in danger of a commitment to ‘essentialism’. Suppose Quine is in the extension of ‘is a Kantian’ in some possible world *w*. This fact about the realm of possibility is not a fact about the necessary or possible truth of some sentence, so there is some commitment to *de re* modality. Note however that accepting such facts as these does not commit one to a view that Quine must have an essence that distinguishes him from all other objects, one natural understanding of ‘essentialism’ (Føllesdal 1986) (see Modal logic, philosophical issues in §3).

7 Index theory and intensional logic

Starting in the late 1960s, PWS began to find a role beyond providing semantics for particular logical systems. One development was the development of ‘intensional logics’ that combined modal, temporal and other operators (see Intensional logics).

We can think of time and possible worlds, for example, as two dimensions along which the truth of a sentence can vary; from this point of view, it is natural to provide a semantics for a system in which sentences are true in worlds at times, and that contains both temporal and modal operators. Thus, for example,

$$\Box(H\phi)$$

will be true in world *w* at time *t* if and only if in every world w' accessible to *w*, at every time t' earlier than *t*, ϕ is true at w' and t'.

But there are also other factors relative to which the truth of sentences can vary. In particular, sentences containing indexicals (such as ‘I’, ‘you’, ‘here’ and ‘now’) will vary in truth depending on who says them, to whom, where and when.

Thus we could think of:

I will walk to the shop

as being true at a world *w*, time *t* and person *a*, if in *w*, *a* walks to the shop at some time subsequent to *t*.

Following the advice of Dana Scott, Montague, Lewis and others developed and explored versions of ‘index theory’, systems in which sentences were true at an index, where indices were *n*-tuples of worlds, times, speakers and other factors (see Scott 1970; Montague 1974b; Lewis 1970).

Kamp’s ‘double-indexing’ (1971), Segerberg’s ‘two-dimensional modal logic’ (1973) and Kaplan’s ‘logic of demonstratives’ (1989) provide alternatives to index theory. These authors emphasize the difference between *reliable* sentences such as ‘I am here now’, that, even if they express contingent propositions, cannot be uttered

falsely, and sentences such as ‘either there are cats or there are not’, that are valid in the standard sense, of being true in every world in every modal model.

8 Montague semantics

Richard Montague was a leader in the development of intensional logic. In his early papers, he developed PWS as a tool for investigating a number of phenomena of philosophical interest, such as sense-data and events. In later work he developed PWS as a powerful tool for model-theoretic treatments of the semantics of English and other natural languages. Montague’s work has had a profound influence in linguistics (see Partee 1989).

As the body of Montague’s work developed, intensional phenomena were increasingly seen not as exceptional and marginal, but as at the core of the way language works. In modal logic, intensionality derives from special operators added onto a base language that works on extensional principles. In Montague semantics, intensionality is a ubiquitous phenomenon; there are not only intensional operators, but intensional verbs, adjectives, adverbs and so on. In Montague’s later work, intensionality is basically the default case, with special postulates to guarantee extensionality (1974c).

By using intensional logic and PWS to give a precise semantics for constructions of natural language, Montague developed an important new subdiscipline of linguistics, often simply called ‘Montague Grammar’.

9 Intensions triumphant

PWS provides philosophy with a toolkit of entities for the analysis of intensional phenomena:

- For individual concepts: functions from worlds to individuals.
- For properties: functions from worlds to extensions.
- For propositions: functions from worlds to truth-values (or sets of worlds).

Note that these functions are themselves extensionally understood, and so analysable within the framework of set theory, and (in that sense at least) free of obscurity.

By the 1970s, philosophers were availing themselves of these tools to talk in disciplined ways about many traditional and some new issues, issues often not directly connected with the interpretation of systems of logic. A few of these are:

- Quantified modal logic has been at the heart of a productive rethinking of issues involved in the distinction between *de dicto* and *de re* necessity, beginning with Quine’s charge that quantified modal logic commits us to essentialism (see Føllesdal 1986; Essentialism).
- PWS provides two models for the semantics of names. On the possible worlds version of the Frege-Russell-Searle descriptive account of names, the meaning of a name is an individual concept. An alternative is to model names on variables directly assigned to individuals, the same for all worlds, irrespective of their properties in the worlds. Marcus (1961) suggested the latter possibility, and in *Naming and Necessity* (1980), Saul Kripke has mounted a full-scale challenge to the descriptive account, arguing that names are ‘rigid designators’ (referring to the same object in all worlds), and providing a causal account of the link between name and thing as an alternative to the descriptive account (see Proper names).
- David Kaplan, whose lectures, seminars and unpublished writings stirred much interest in PWS throughout the 1970s, worked out an account of indexicals and demonstratives in the context of index theory. This has led to a clarification of a number of issues involving the semantics and epistemology of indexicals (see Kaplan 1989; Demonstratives and indexicals).
- David Lewis has used the apparatus of PWS to make significant contributions to our understanding of convention, the semantics of natural language, the understanding of counterfactuals (see above) and many other issues in metaphysics, epistemology and the philosophy of science (see Lewis 1970, 1973, 1979).

Lewis’ own view of possible worlds maintains that possible worlds are alternative concrete realities; they are actual for their inhabitants, as ours is for us. The inhabitants of other worlds are not identical with the inhabitants of the actual world (that is, our world), but are their *counterparts*; Lewis’ account of quantification is based on the counterpart relations rather than identity.

10 All the intensions we need?

Can PWS supply philosophy with all the intensions that are needed to understand intensional phenomena? Can all intensions be understood as functions from worlds to appropriate sets?

One of the most difficult challenges is the problem of propositional attitudes. The basic problem is that PWS supplies only one necessary proposition (the set of all worlds) and only one contradictory proposition (the null set). This seems to pose a severe problem for dealing with mathematical knowledge. Given usual principles of compositionality, we could infer from 'Elwood knows that $7+5=12$' to 'Elwood knows that *S*' for any true mathematical sentence *S* or any other necessary truth for that matter.

Robert Stalnaker (1984) has given an careful and extended defence of the use of PWS in epistemology. He argues that the concept of content needed for propositional attitudes is grounded in pragmatic relations and such informational relations as indication. The problem of mathematical knowledge, he argues, can be resolved by seeing a linguistic element in our knowledge of mathematical truths.

Advocates of situation semantics have argued, however, that the possible worlds analysis of indication is also vitiated by the problem of the single necessary proposition. Where *P* is a contingent proposition and *N* is the necessary proposition, $P = P \,\&\, N$. So if the tree rings indicate that the tree is one hundred years old, they also indicate that it is one hundred years old and $7+5=12$. But indication appears to distribute over conjunction, so we could infer that the tree rings indicate that $7+5=12$ (see Perry 1993; Semantics, situation).

While these and other problems have stimulated great interest in other approaches to intensionality in recent years, it seems fair to say that PWS has had by far the most impact on the disciplined investigation of intensional phenomena and that no alternative treatment yet devised provides as natural and comfortable a scheme for thinking about intensional matters.

See also: Intensional entities; Possible worlds

JOHN R. PERRY

References and further reading

Benthem, J. van (1983) *The Logic of Time*, Dordrecht: Reidel.(Important contributions to a wide range of issues in tense logic.)

Benthem, J. van (1988) *A Manual of Intensional Logic*, Stanford, CA: Center for the Study of Language and Information.(A survey of modal logic, tense logic and other applications of the possible worlds framework.)

Carnap, R. (1946) 'Modalities and Quantification', *Journal of Symbolic Logic* 11: 33-64.(Introduces quantifiers into modal logic.)

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press.(Carnap's classic work; explains his system of intension and extensions, and presents a version of possible worlds semantics.)

Føllesdal, D. (1986) 'Essentialism and Reference', in L.E. Hahn and P.A. Schilpp (eds) *The Philosophy of W.V. Quine*, Library of Living Philosophers, La Salle, IL: Open Court, 97-115.(Discusses Quine's charge that quantification into modal contexts involves essentialism; argues that it is valid in a relatively benign sense of 'essentialism', but not valid in other senses.)

Føllesdal, D. (1994) 'Stig Kanger In Memoriam', in D. Prawitz, B. Skyrms and D. Westerståhl (eds) *Logic, Methodology and Philosophy of Science IX*, Uppsala: Elsevier Science, 885-888.(Discusses the contributions by Kanger, Kripke and others to the semantics of modal logic.)

Hintikka, J. (1957) 'Quantifiers in Deontic Logic', *Societas Scientiarum Fennica, Commentationes humanarum litterarum* 23 (4).(An early use of the idea of relative necessity.)

Hintikka, J. (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Ithaca, NY: Cornell University Press.(Explores the logic of knowledge and belief within a version of the possible worlds framework.)

Hughes, G. and Cresswell, M. (1968) *An Introduction to Modal Logic*, London: Methuen.(Excellent introduction to modal logic and possible worlds semantics.)

Kamp, H. (1971) 'Formal Properties of "Now"', *Theoria* 37: 227-74.(Develops a two-dimensional version of index theory to treat temporal indexicals.)

Kanger, S. (1957) *Provability in Logic*, Stockholm Studies in Philosophy 1, Stockholm: Almqvist & Wiksell. (Basic completeness results in modal logic using accessibility relations.)

Kaplan, D. (1989) *Demonstratives*, in J. Almog, J. Perry and H. Wettstein (eds) *Themes from Kaplan*, 481-563, New York: Oxford University Press.(Classic treatment of indexicals within a possible worlds framework; discussions of related philosophical and semantical issues.)

Kripke, S.A. (1959) 'A Completeness Theorem in Modal Logic', *Journal of Symbolic Logic* 24: 1-15.(Uses possible worlds semantics to prove the completeness of S5.)

Kripke, S.A. (1963a) 'Semantical Analysis of Modal Logic I, Normal Propositional Calculi', *Zeitschrift für mathematische Logic und Grundlagen der Mathematik* 9: 67-96.(Uses possible worlds semantics with accessibility relations to obtain a number of completeness results.)

Kripke, S.A. (1963b) 'Semantical Considerations on Modal Logic', *Acta Philosophical Fennica* 16: 83-94. (Provides possible worlds semantics for quantified modal logic.)

Kripke, S.A. (1965) 'Semantical Analysis of Modal Logic II, Non-normal Propositional Calculi', in I.W. Addison, L. Kenkin and A. Tarski (eds) *The Theory of Models*, Amsterdam: North Holland, 206-20.(Extends semantical treatment to logics without necessitation and other special cases.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press. (Groundbreaking work on the semantics of proper names, arguing that they are 'rigid designators'.)

Lewis, C.I. and Langford, C.H. (1932) *Symbolic Logic*, New York: The Century Company.(Includes a survey of modal logics.)

Lewis, D.K. (1969) *Convention*, Cambridge, MA: Harvard University Press.(Develops a conception of convention and argues that language is conventional.)

Lewis, D.K. (1970) 'General Semantics', *Synthèse* 22: 18-67.(Possible worlds semantics in the index style applied to problems in natural language.)

Lewis, D.K. (1973) *Counterfactuals*, Cambridge, MA: Harvard University Press.(Proposes an analysis of conditionals within the possible worlds framework.)

Lewis, D.K. (1979) 'Attitudes *De Dicto* and *De Se*', *Philosophical Review* 88: 513-43. (Treats problems of indexicality and self-knowledge within a possible worlds framework.)

Lewis, D.K. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.(Defends a conception of possible worlds as alternative concrete realities.)

Marcus, R.B. (1946) 'A Functional Calculus of the First Order Based on Strict Implication', *Journal of Symbolic Logic* 11: 1-16.(Introduces quantifiers into modal logic; identifies crucial principles.)

Marcus, R.B. (1961) 'Modalities and Intensional Languages', *Synthèse* 13: 303-32. (Discussion of philosophical issues in the interpretation of quantified modal logic; argues that names are like tags rather than descriptions.)

Montague, R.M. (1974a) 'Logical Necessity, Physical Necessity, Ethics and Quantifiers', in R. Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, CT, and London: Yale University Press, 71-83.(Notes quantificational structure of various operators; uses a relative notion of necessity.)

Montague, R.M. (1974b) 'Pragmatics and Intensional Logic', in R. Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, CT, and London: Yale University Press, 119-147.(Develops an index-style treatment of indexicality within an intensional language.)

Montague, R.M. (1974c) 'English as a Formal Language' and 'The Proper Treatment of Quantification in Ordinary English', in R. Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, CT, and London: Yale University Press, 188-221, 247-70.(In these two classic papers Montague develops a formal semantics for significant fragments of a natural language.)

Montague, R.M. (1974c) 'On the Nature of Certain Philosophical Entities', in R. Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, CT, and London: Yale University Press, 148-87.(Applies possible worlds semantics and other concepts of formal philosophy to sense-data and other classic philosophical problems.)

Partee, B. (1989) 'Possible Worlds in Model-Theoretic Semantics: A Linguistic Perspective', in S. Allen (ed.) *Possible Worlds in Humanities, Arts and Sciences. Proceedings of Nobel Symposium 65*, Berlin and New York: de Gruyter, 93-123.(Survey of the contribution that possible worlds semantics has made to the field of linguistics.)

Perry, J. (1993) *The Problem of the Essential Indexical and Other Essays*, New York: Oxford University Press. (This collection includes papers on indexicality and critiques of possible worlds semantics.)

Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Oxford University Press.(Explores a number of metaphysical issues from the perspective of a version of possible worlds semantics.)

Pratt, V.R. (1976) 'Semantical Considerations on Floyd-Hoare Logic', *Proceedings of the Seventeenth Annual IEEE Symposium on Foundations of Computer Science*, 109-21.(A seminal work in the logic of programs.)

Prior, A.N. (1959) *Time and Modality*, Oxford: Oxford University Press. (Explorations in temporal and modal

logics by a pioneer in the field.)
Prior, A.N. (1967) *Past, Present and Future*, Oxford: Oxford University Press.(The classic presentation of temporal logic.)
Prior, A.N. and Fine, K. (1977) *Worlds, Times and Selves*, London: Duckworth. (Examines metaphysical issues relevant to modal logic and its interpretations.)
Ramsey, F.P. (1931) *The Foundations of Mathematics*, London: Routledge & Kegan Paul.(Collection of Ramsey's philosophical writings.)
Scott, D. (1970) 'Advice on Modal Logic', in K. Lambert (ed.) *Philosophical Problems in Logic*, Dordrecht: Reidel.(An important work in the development of index theory.)
Segerberg, K. (1973) 'Two-Dimensional Modal Logic', *Journal of Philosophical Logic* 2: 77-96.(Develops a version of index theory in which the accessibility relation holds between pairs of indices.)
Stalnaker, R. (1968) 'A Theory of Conditionals', in N. Rescher (ed.) *Studies in Logical Theory*, Oxford: Blackwell.(Proposes an analysis of conditionals within the possible worlds framework.)
Stalnaker, R. (1981) 'Indexical Belief', *Synthese* 49: 129-51.(Introduces the concept of 'diagonal' propositions to handle cases involving indexicals.)
Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press.(An analysis of basic epistemological concepts within possible worlds semantics.)

Semantics, situation

Situation semantics attempts to provide systematic and philosophically coherent accounts of the meanings of various constructions that philosophers and linguists find important. It is based on the old idea that sentences stand for facts or something like them. As such, it provides an alternative to extensional semantics, which takes sentences to stand for truth-values, and to possible worlds semantics, which takes them to stand for sets of possible worlds.

Situations are limited parts or aspects of reality, while states of affairs (or infons) are complexes of properties and objects of the sort suitable to constitute a fact. Consider the issue of whether Jackie, a dog, broke her leg at a certain time T. There are two states of affairs or possibilities, that she did or she did not. The situation at T, in the place where Jackie was then, determines which of these states of affairs (infons) is factual *(or* is the case *or* is supported*).* Situation theory*, the formal theory that underlies situation semantics, focuses on the nature of the* supports *relation.*

Situation semantics sees meaning *as a relation among* types of situations. *The meaning of 'I am sitting next to David', for example, is a relation between types of situations in which someone A utters this sentence referring with the name 'David' to a certain person B, and those in which A is sitting next to B. This relational theory of meaning makes situation semantics well-suited to treat indexicality, tense and other similar phenomena. It has also inspired relational accounts of information and action.*

1 History of situation semantics

Situation semantics was originally conceived as an alternative to extensional model theory and possible world semantics especially suited to the analysis of various problematic constructions, including naked-infinitive perception verbs (Barwise 1981) and belief-reports (Barwise and Perry 1981a) (see Model theory; Semantics, possible worlds). In its earliest forms, the central ideas were:

• *Partiality.* Situations are contrasted with worlds; a world determines the answer to every issue - the truth-value of every proposition. A situation corresponds to the limited parts of reality we in fact perceive, reason about, and live in. What goes on in these situations will determine answers to some issues, but not all. In 'Scenes and Other Situations' (1981), reporting his initial work on situation theory, Jon Barwise represents scenes, the situations we perceive, as partial first-order models.
• *Realism.* Basic properties and relations are taken to be real objects - uniformities across situations and objects - not bits of language, ideas, sets of *n*-tuples, or functions. In *Situations and Attitudes* (Barwise and Perry 1983), 'courses of events' are partial functions from sequences of locations, relations and objects to truth-values. Complex properties and relations, and various types of objects were full-fledged objects, entering into courses of events.
• *The relational theory of meaning.* The meaning of an expression ϕ is conceived as a relation between a discourse situation, a connective situation and a described situation, written:
$_{d,c}[[\phi]]_e$

The meaning of 'I am sitting next to David', for example, would obtain between courses of events *d*, *c* and *e* if there are individuals *a* and *b* such that (1) in *d*, *a* is the speaker of the sentence; (2) in *c*, *a*'s use of 'David' is used to refer to *b*; and (3) in *e*, *a* is sitting next to *b*.

A number of trenchant criticisms were made of *Situations and Attitudes* (see especially Soames 1985). In reaction to them, Barwise and Perry (1985) recognized the need to rethink the foundations of situation semantics. Two main developments bridge the early versions of situation semantics and the later ones which emerged from this rethinking:

(1) In early versions, situation semantics was developed within standard set theory; this led to foundational problems. In the mid-1980s, Barwise and others developed various versions of 'situation theory', in which all of the various entities that had become necessary were treated axiomatically (Barwise 1989; Devlin 1991; Westerståhl 1990) or within Peter Aczel's version of set theory (Barwise and Etchemendy 1987).

(2) The concept of a constraint, developed in *Situations and Attitudes* as an adjunct to the relational theory of meaning, has become central to the development of situation semantics as a general account of informational and intentional content (Barwise 1993; Israel and Perry 1990; Perry 1993).

2 Situations

The basic idea of situation semantics is that in thought and action we use complexes of objects and properties to classify 'directly' and 'indirectly' parts and aspects of reality, or 'situations'. This sort of realistic classification is more basic than linguistic classification, and it underlies linguistic classification. Consider a simple dialogue:

'What happened in the woods this afternoon?'
'Jackie broke her leg.'

The question concerns a certain situation, a bit of reality: the events in the woods this afternoon. The answer directly classifies the situation in terms of an object (the dog Jackie) and a property (acquiring a broken leg). We classify situations by what goes on in them, by which properties objects have, and by the relations they stand in to one another in virtue of the events that comprise the situation.

Consider the issue of whether Jackie broke her leg at a certain time *t*. There are two dual possibilities or 'states of affairs', corresponding to whether her leg was broken or not, which we can represent as:

σ: $\langle\langle \mathrm{breaks\ leg}, t, \mathrm{Jackie}; 1\rangle\rangle$

and

σ': $\langle\langle \mathrm{breaks\ leg}, t, \mathrm{Jackie}; 0\rangle\rangle$

Of course, what goes on in the whole world (if we assume there is such a totality) will determine whether or not Jackie broke her leg, but this will also be determined by much smaller situations. Let *s* be the situation in the woods this afternoon. Then,

$s \models \sigma$

that is, *s supports* σ, or (in more traditional philosophical terms) *s makes it the case* that σ, or *makes* σ *factual*.

In situation theory, various objects are built from the basic interplay of situations and states of affairs, permitting complex and abstract ways of classifying situations, including complex states of affairs, properties and relations. A key concept is a 'type of situation', such as the type of situation in which a dog breaks its leg (call it *S*) and the type of situation in which a dog does not run (call it S').

There are states of affairs involving these abstract objects. In particular, one type of situation may involve another: if there is a situation of the first type, there will also be one of the second type. *S* involves S': dogs with broken legs do not run. These sorts of states of affairs are 'constraints'.

3 Meaning

Constraints give rise to the possibility of 'indirect classification': classifying situations by what they mean. That is, classifying situations not by the states of affairs they support, but by the types of situations they involve, relative to some constraint.

Indirect classification is how situation semantics conceives of informational and intentional content. Classifying situations by their contents is what organisms do under the influence of what Hume calls 'custom': confronted with a situation, they form expectations, or at least contemplate possibilities, on the basis of what the situation involves relative to some constraint, factual or not, to which they have become attuned. Situation semantics interprets informational and intentional content as a system that exploits such indirect classification. Situations are indirectly classified relative not only to laws of nature and other actual constraints (informational content), but also to conventions, rules, customs, plans and other constraints, both factual and fictional, of human contrivance (intentional content).

Consider:

(1) Jackie has a broken leg.
(2) The x-ray shows that Jackie has a broken leg.
(3) The vet said that Jackie had a broken leg.

In (1), we have direct classification; in (2) and (3), indirect classification. In the latter a common pattern is discernible, involving three types of situation and a constraint:

- A local situation.
- Connections between objects in that situation and other objects, the 'subject matter'.
- A remote situation, the content, involving the subject matter.
- A constraint according to which a combination of situations of the first two types involves a situation of the third type.

In (2) the local situation is the x-ray having certain characteristics. The x-ray is connected to Jackie: it was taken of her. The complex type of situation, in which an x-ray taken of a certain dog exhibits those features, involves a situation in which the dog has a broken leg. Given the connections between the x-ray and Jackie, its having those characteristics shows that she has a broken leg. Here the constraint is factual and the content is informational.

In (3) the local situation is the utterance, in which the vet utters the words 'Jackie has a broken leg'. The vet's use of the word 'Jackie' is connected through various mental and conversational links to the dog Jackie. The rules of English provide the constraint: given the characteristics of the utterance and its connections, it is true if and only if Jackie has a broken leg. English speakers are attuned to these constraints, not in the sense that they automatically form expectations when they hear utterances, but in the sense that they grasp the type of situation meant. Here the content is intentional.

Situation semantics then conceives of meaning as a relation between types of situations. A key advantage of this conception is that it allows us to see how different information can be gleaned from the same 'signal' given different starting points.

In the case of (2), we think naturally of a case in which an experienced vet in a well-organized office studies an x-ray known to be of Jackie, and learns that she has a broken leg. In another case, an experienced vet in a poorly organized office might infer from the x-ray and the fact that Jackie is the only dog in the place with a broken leg, that it was of her. And a would-be vet might learn how to read x-rays, knowing that the x-ray is of Jackie and that she has a broken leg.

Similarly, in (3) we think of a person who knows to which dog the vet refers when he says 'Jackie' and knows English, learning that Jackie has a broken leg. But attunement to the same constraint and a different starting point might allow someone to learn which of the dogs in the office was named 'Jackie'.

4 Accomplishments

Situation semantics has been used to analyse a wide variety of linguistic phenomena (see, for example, Gawron and Peters 1990; Cooper 1992), the liar paradox (Barwise and Etchemendy 1987), heterogeneous reasoning and representation (Barwise and Etchemendy 1991), diagrammatic reasoning (Shin 1990), the nature and structure of information and action (Israel and Perry 1990; Devlin 1991; Barwise 1993), and a number of other issues involving language, representation and computation.

It is probably fair to say, however, that up to this point situation semantics has been more successful in terms of adoption of its broad themes than in terms of adoption of its specific formalism and proposals. The main themes of early situation semantics (partiality, realism and the relational nature of meaning; see §1) have been incorporated into the (generally) received wisdom of philosophy and linguistics. But situation semantics remains only one of a number of alternative semantical frameworks that exhibit these virtues in various ways, from which a theorist may choose.

JOHN R. PERRY

References and further reading

Aczel, P., Israel, D., Katagiri, Y. and Peters, S. (eds) (1993) *Situation Theory and Its Applications*, vol. 3,

Stanford, CA: Center for the Study of Language and Information. (This volume contains a number of articles about situation theory and situation semantics; it is the third in a series of volumes devoted to these topics; see Cooper, Mukai and Perry (1990) and Barwise *et al*. (1991) for the other volumes.)

Barwise, J. (1981) 'Scenes and Other Situations', *Journal of Philosophy* 77: 369-97; repr. in *The Situation In Logic*, Stanford, CA: Center for the Study of Language and Information, 1989.(In this paper Barwise describes his original idea of a scene semantics, a precursor of situation semantics in which partial first-order models are used to provide an analysis of naked-infinitive perception reports.)

Barwise, J. (1989) *The Situation In Logic*, Stanford, CA: Center for the Study of Language and Information.(The papers in the collection developed many of the basic ideas of later versions of situation theory.)

Barwise, J. (1993) 'Constraints, Channels, and the Flow of Information', in P. Aczel, D. Israel, Y. Katagiri and S. Peters (eds) *Situation Theory and Its Applications*, Stanford, CA: Center for the Study of Language and Information, vol. 3, 3-27.(In this paper Barwise develops an abstract theory of information flow based on the idea of constraints.)

Barwise, J. and Etchemendy, J. (1987) *The Liar*, New York: Oxford University Press.(Barwise and Etchemendy provide an analysis of the liar paradox in terms of situation semantics.)

Barwise, J. and Etchemendy, J. (1991) 'Visual Information and Valid Reasoning', in W. Zimmerman (ed.) *Visualization in Mathematics*, Washington, DC: Mathematical Association of America, 9-27.(Barwise and Etchemendy use a situation-based concept of information to defend the view that reasoning using diagrams and reasoning that combines diagrams and sentences can provide valid proofs and can be studied model-theoretically.)

Barwise, J., Gawron, J.M., Plotkin, G. and Tutiya, S. (eds) (1991) *Situation Theory and Its Applications*, vol. 2, Stanford, CA: Center for the Study of Language and Information.(This volume is the second in a series devoted to articles about situation theory and situation semantics; see Cooper, Mukai and Perry (1990) and Aczel *et al*. (1993) for the other volumes.)

Barwise, J. and Perry, J. (1981a) 'Semantic Innocence and Uncompromising Situations', in P. French, T. Uehling and H. Wettstein (eds) *Midwest Studies in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press, vol. 6, 387-403.(In this article, which announced situation semantics, Barwise and Perry criticize a famous argument used by Church, Davidson and others to show that facts cannot be the semantic values of sentences.)

Barwise, J. and Perry, J. (1981b) 'Situations and Attitudes', *Journal of Philosophy* 77: 668-91.(In this early article Barwise and Perry describe the application of situation semantics to belief-reports and explain the foundational issues that plagued early versions of situation semantics.)

Barwise, J. and Perry, J. (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press/Bradford.(This book developed a version of situation theory within set theory, and applied it to linguistic meaning, the propositional attitudes and the semantics of mental states.)

Barwise, J. and Perry, J. (1985) 'Shifting Situations and Shaken Attitudes', *Linguistics and Philosophy* 8: 105-61.(Responding to the criticisms of Soames and others, Barwise and Perry recognize that their theory (1983) has serious problems and explore strategies for repairing it.)

Cooper, R. (1992) 'A Working Person's Guide to Situation Theory', in S. Hansen and F. Sørensen (eds) *Semantic Representation and Interpretation*, Fredericksberg: Samfundslitteratur.(A lucid introduction to situation theory for the working cognitive scientist.)

Cooper, R., Mukai, K. and Perry, J. (eds) (1990) *Situation Theory and Its Applications*, vol. 1, Stanford, CA: Center for the Study of Language and Information.(This volume is the first in a series devoted to articles about situation theory and situation semantics; see Barwise *et al*. (1991) and Aczel *et al*. (1993) for the other volumes.)

Devlin, K. (1991) *Logic and Information*, Cambridge: Cambridge University Press. (Devlin's book explains the basics of the current version of situation theory and applies it to a number of problems in linguistics and psychology.)

Gawron, J.M. and Peters, S. (1990) *Anaphora and Quantification in Situation Semantics*, Stanford, CA: Center for the Study of Language and Information. (Situation semantics is used to explain the certain aspects of anaphoric relations.)

Israel, D. and Perry, J. (1990) 'What is Information?', in P. Hanson (ed.) *Information, Language, and Cognition*, Vancouver, BC: University of British Columbia Press, 1-19.(Israel and Perry develop a distinction between pure and incremental information within situation theory.)

Israel, D. and Perry, J. (1991) 'Information and Architecture', in J. Barwise, J.M. Gawron, G. Plotkin and S. Tutiya (eds) *Situation Theory and Its Applications*, Stanford, CA: Center for the Study of Language and

Information, vol. 2, 147-60.(Israel and Perry use the theory they developed in ‘What is Information?’ (1990) to distinguish among various types informational structures.)

Perry, J. (1993) *The Problem of the Essential Indexical and Other Essays*, New York: Oxford University Press.(A collection of Perry’s essays on situation semantics and related topics.)

Shin, S.J. (1990) *The Logical Status of Diagrams*, Cambridge: Cambridge University Press.(Situation theory is used to provide a precise analysis of reasoning with Venn diagrams.)

Soames, S. (1985) ‘Lost Innocence’, *Linguistics and Philosophy* 8: 59-71.(A trenchant critique of Barwise and Perry (1983).)

Westerståhl, D. (1990) ‘Parametric Types and Propositions in First-Order Situation Theory’, in R. Cooper, K. Mukai and J. Perry (eds) *Situation Theory and Its Applications*, Stanford, CA: Center for the Study of Language and Information, vol. 1, 193-230.(A subtle analysis of the role of parameters in situation theory.)

Semantics, teleological

Teleological/biological theories of meaning use a biological concept of function to explain how the internal states of organisms like ourselves can represent conditions in the world. These theories are controversial, as they have the consequence that an organism's history affects the content of its present thoughts. These theories have advantages over other naturalistic theories of meaning in the task of explaining the possibility of error and unreliable representation.

1 Basic ideas

Teleological/biological theories (hereafter, 'teleological theories') of meaning claim that certain internal states of organisms represent particular external conditions because of the biological *functions* of these inner states and mechanisms that interact with them. These functions derive from evolutionary history, or from some other history of selection; something's function is (roughly) the thing it does which explains why it is there (Wright 1973). For example, the function of the heart is pumping blood, as that is the thing hearts do that explains why they are there. Teleological theories of meaning, such as those theories of Millikan (1984) and Papineau (1987), apply this strategy of analysis to inner representational states (see Functional explanation).

These theories claim that inner states that help an organism coordinate its behaviour with its environment can represent environmental conditions that have played a certain causal role in the organism's history. For example, an animal may be put together in such a way that when a dark shadow appears overhead, an inner state occurs which causes the animal to hide. This inner state has the content 'Predator!' (or perhaps 'Danger!') if it is the presence of predators (or danger) on past occasions of this type that explains this pattern of inner wiring being favoured by natural selection. Similarly, though a frog might snap at any small dark object in its visual field, if the pattern of wiring responsible for this behaviour has been selected over alternatives because of past occasions in which this behaviour resulted in frogs obtaining food, then 'Food!' or perhaps 'Edible insect!' is the content of the inner state that is caused by the dark spot and which causes the behaviour. To use the language of Ruth Millikan (1984), who has developed the most sophisticated teleological theory, the inner state represents the environmental condition that must obtain for the state to help the parts of the organism that make use of (or 'consume') it to perform their functions in a historically normal way.

This type of explanation accounts for the content of belief-like states. The content of desire-like states can be explained in terms of satisfaction conditions. A desire's satisfaction condition is the condition in the world that the desire is supposed to bring about (see Belief; Desire).

Teleological theories do not hold that all beliefs or concepts have a specific evolutionary history. First, these theories can explain some novel beliefs as novel combinations of old concepts. Second, there can be teleological explanations of concepts that are not innate. Millikan explains the semantic properties of learned concepts in terms of the biological functions of parts of the cognitive system that are designed to adapt the organism to novel conditions by means of learning. An alternative approach, defended by Papineau (1987, 1993), is to view learning as a selection process itself, one akin to natural selection across generations, and an independent source of biological functions. In principle there are also other mechanisms which could bestow functions of the relevant sort, such as cultural evolution and deliberate design.

Though Millikan and Papineau give general analyses of all mental representation in these terms, it is also possible to use these ideas in less ambitious ways. A teleological theory could be used to explain only the most basic types of internal representation. Other theories of meaning could then make use of the representational abilities explained by the teleological theory (Sterelny 1990). For example, we might explain a basic stock of primitive concepts in teleological terms, and see more complex concepts as logical constructions out of these (an old-fashioned possibility). Or the teleological view might explain how it is possible for organisms to represent basic environmental features, including other organisms, and then social factors could be used to assign semantic properties to more complex thoughts and public representations. Either way, the teleological view would then be part of a package of different theories which work together in explaining meaning.

2 Related theories

One way to motivate the teleological view is by contrasting it to other naturalistic semantic theories. Indicator theories explain representation in terms of the existence of a reliable correlation, in certain conditions, between an inner state and its object (see Dretske 1981; Fodor 1990; Semantics, informational). These theories have difficulty explaining error and unreliable representation. Teleological theories are attractive to some because they allow an organism to represent something that it cannot, in principle, reliably detect. Suppose an organism's means of detecting predators is not reliable, as it is prone to 'false alarms'. No matter how unreliable it is, if this inner wiring was selected because it enabled organisms to escape predators, the inner state in question can be seen, on a teleological view, as about predators. For an indicator theory, the content of the inner state must be something weaker, which is reliably correlated with the inner state, such as dark shape. Although most teleological theories do assign these 'ecologically salient' states as contents of representations - 'Predator!' rather than 'Dark shape!' - this is not strictly necessary. Neander (1995) has defended a theory which is teleological in structure, but which converges with indicator theories in the contents it assigns in these cases.

Indicator theories are 'upstream-looking' theories; they link a representation to its object by looking to the processes involved in the bringing about of the representational state. They are based upon the organism's powers of discrimination by means of perception. Teleological theories look instead to connections 'downstream' of the representational state, connections going via behaviour and its consequences. In this respect there is a kinship between the teleological approach and some older ideas about thought associated with pragmatism (see Pragmatism). A simple way to link behaviour and belief content is to suppose that the truth condition of a belief is the condition such that actions based on the belief will be successful if and only if that condition obtains. The teleological approach is one way of developing this idea. It replaces the problematic notion of 'success' with a precise concept based upon natural selection. In some respects the debate between the indicator view and the teleological view is an expression of a more general opposition concerning the relative importance of perception and behaviour in understanding the mind. However, there have also been attempts to combine the indicator and the teleological approaches, such as the theory defended by Dretske (1988). For Dretske an inner state represents predators if it has the 'function to indicate' predators; both reliable correlation and function are required.

3 Current debates

Teleological theories of representation have been the subject of lively debate. I will discuss three types of objection.

First, these theories claim that organisms can only represent the world in virtue of facts about their history. Some find this unacceptable, as it entails that a molecule-for-molecule replica of you (such as the 'swampman' of Davidson 1987) which arose instantly by sheer chance would not have thoughts with any semantic content. Millikan's response is simply to embrace this result. Note that while an evolutionary history is required for content on Millikan's view, a more moderate line can be taken by theories such as Papineau's, which treat individual learning as an independent source of functions. Then a swampman would have no thoughts with content initially, but could acquire them within its lifetime.

Second, Jerry Fodor (1990) has argued that teleological theories cannot account for the 'opacity' of attributions of content (see Intentionality; Propositional attitude statements). Fodor claims that if all *F*s are *G*s in some environment, then anything that was selected in that environment for responding to *F*s was selected also for responding to *G*s. If so, attributions of content based on biological function cannot distinguish in such cases between representing the presence of an *F* and representing the presence of a *G*. Millikan replies that 'selection for' is a causal matter, and hence that attributions of function have as much opacity as causal explanations in general. Then teleological theories will distinguish between representing the presence of an *F* and representing the presence of a *G*, when all *F*s happen to be *G*s. The theories will only fail to distinguish between these two if the properties of being an *F* and being a *G* are indistinguishable in their causal powers. Then the question becomes whether a teleological theory ought to distinguish, for example, thoughts about equilateral triangles from thoughts about equiangular ones (see Causation; Property theory).

Lastly, suppose an organism avoids predators by avoiding some totally different environmental condition, such as sunrise, which happens to be reliably correlated with danger from predation (Pietroski 1992). Teleological theories say that the condition in the world with explanatory importance is the object of representation. So an animal that instinctively flees the sun and hence avoids predators is representing to itself 'Predator!' rather than 'Sun!'. This

can be so even if the animal has no ability to detect predators when they are right in front of it. This is a problem for most teleological views, including Millikan's and Papineau's, although not for Dretske's 'mixed' view nor Neander's theory. The problem is hard for standard teleological views because on such views there is not supposed to be a sharp distinction between the object of a representation and the object which explains the representational practice. These theories are explicitly aimed at assimilating the former to the latter.

See also: Evolution, theory of; Functional explanation; Semantics

PETER GODFREY-SMITH

References and further reading

Davidson, D. (1987) 'Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association* 60: 441-58.(Introduces 'swampman'.)

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(Classic text for the 'indicator' approach.)

Dretske, F. (1988) *Explaining Behavior*, Cambridge, MA: MIT Press.(Defends a view which combines the teleological approach and the indicator approach.)

Fodor, J.A. (1990) *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press.(Attacks the teleological approach and defends an indicator theory.)

Godfrey-Smith, P. (1994) 'A Continuum of Semantic Optimism', in S. Stich and T.A. Warfield (eds) *Mental Representation: A Reader*, Oxford: Blackwell.(Surveys teleological theories, and other theories which use a link between meaning and success.)

Millikan, R.G. (1984) *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.(The most sophisticated teleological theory.)

Millikan, R.G. (1993) *White Queen Psychology, and Other Essays for Alice*, Cambridge, MA: MIT Press.(Various papers defending and developing Millikan's theory; includes 'Biosemantics', the best introduction to her view, and her response to Fodor, 'Speaking up for Darwin'.)

Neander, K. (1995) 'Misrepresenting and Malfunctioning', *Philosophical Studies* 79: 109-41.(Presents a teleological theory and defends it against Fodor, Pietroski and others.)

Papineau, D. (1987) *Reality and Representation*, Oxford: Blackwell.(Defends a teleological theory and discusses connections to debates about realism.)

Papineau, D. (1993) *Philosophical Naturalism*, Oxford: Blackwell.(A further development of Papineau's teleological theory.)

Pietroski, P. (1992) 'Intentionality and Teleological Error', *Pacific Philosophical Quarterly* 73: 267-82.(An interesting criticism of the teleological approach.)

Sterelny, K. (1990) *The Representational Theory of Mind: An Introduction*, Oxford: Blackwell.(A more 'modest' teleological approach.)

Wright, L. (1973) 'Functions', *Philosophical Review* 82: 139-68.(Classic article on functions.)

Semiotics

As the study of signification, semiotics takes as its central task that of describing how one thing can mean another. Alternatively, since this philosophical problem is also a psychological one, its job could be said to be that of describing how one thing can bring something else to mind, how on seeing 'x' someone can be induced to think about 'y' even though 'y' is absent.

A person in whose head 'y' has been brought to mind may be responding to an 'x' someone else has transmitted with the intention of its signifying 'y'; or, mistakenly, responding to an 'x' someone has transmitted in the guileless expectation of its signifying some 'z'; or, often, responding to an 'x' that comes to his notice without anybody's apparent intention at all. Words, for example, generally signify because someone intends them to, and ideally (though not always) they signify what is intended; whereas clouds signify - a coming storm, a whale - because we so interpret them, not because they shaped themselves to convey some meaning.

Obviously the study of signification forms an integral part of the study of thinking, since no object can itself enter the brain, barring fatal mischance, and so it must be represented by some mental (that is, neural) 'x' that signifies it.

Signifiers are equally essential for creatures far lower than humans, as when a chemical signal 'x' emitted by some bacterium signifies to one of its colleagues some 'y' such as 'there's a dearth of food hereabouts'.

There are a number of ways in which an 'x' can signify some 'y', but for humans these are chiefly: by physical association; by physical resemblance; and/or by arbitrary convention.

When we take some 'x' as signifying some 'y' we are often guessing; our guess is subject to checking by interpretative (re)appraisal.

1 The signifier, the signified, the appraisal

In what might be called the standard version of semiotic theory, which in most essentials we owe to the American polymath Charles S. Peirce, our knowledge of the external world, in fact all thinking of any kind, is composed of chains and skeins of linked representations or signifiers ('signs' for short). The world enters our consciousness as highly processed sense-reports, which begin as representations (for instance, the electro-chemical signals into which our light-recepters transform the light striking the retina). So the world is understood, to the extent that it is, via signs, many of which stand for still other signs, each bound to what it signifies by interpretive (re)appraisals, which take account of such things as general knowledge and immediate context. An illustrative example of such a skein of signs is provided by the various ways in which an English word can be encoded. In Morse code the word 'card' is ' —·—· ·— ·—· —··', where, to take just the first Morse letter, ' —·—·' stands for:

(1) the relatively long and short electrical impulses we call 'dash, dot, dash, dot' (thus sequenced);
(2) the relatively long and short audible beeps we call 'dash, dot, dash, dot' (thus sequenced);
(3) the words 'dash, dot, dash, dot' (thus sequenced);
(4) the Roman letter 'C';
(5) the sound that 'C' spells in the context '-ard', namely /k/.

Each of these stands, in turn, for any or all of the others. Even this is but a bare beginning, since the entire word 'card', however expressed, stands ultimately for a concept, which interpreted or appraised in a particular context might be a playing card, a visiting card, an index card or part of a machine for carding wool.

Though he had many eminent predecessors, it was Peirce who first undertook, in plangent insights scattered over his (mostly unpublished) eighty-odd volumes of papers, to build a comprehensive discipline of semiotics within which all varieties and manners of signification could be given a unified explanation. He entered his final decline with this mighty task unachieved, but he had at least clearly delineated the framework within which almost all semiotic issues, even the semiotic investigations of antiquity, are currently best understood. In particular, it was Peirce who first definitively gave equal triadic footing to the signifier, what it signifies and its interpretative (re)appraisal.

The two greatest semiotic theorists of antiquity were Aristotle and Augustine, with other important contributions from Plato and the Greek Stoics, especially Chrysippus (see Language, ancient philosophy of). Taken as a group, these philosophers distinguished with some care, first, between '*x*' the signifier and '*y*' what it signifies and, second, between the signified '*y*' and, where at issue, the object in the putatively real world that '*y*' represents (Plato, primarily in *Cratylus*; Aristotle, most importantly in *De Interpretatione (Perihermenias)*; Chrysippus; and Augustine in *De Doctrina Christiana*). In addition Aristotle established in the *Prior Analytics* that in the process by which sign is tied to signified, the determination of what the sign signifies is often a probabilistic guess. If I interpret the apparent fact that 'That weathervane is pointing north' as signifying that 'The local wind is a northerly', I am jumping from premise to conclusion without invoking the universal premise ('A north-pointing weathervane *always* signifies that the local wind is a northerly') that would clinch this enthymeme as a syllogism. (Note that here, as so often, 'signifies' functions like 'implies'.) In fact this missing universal premise is quite properly omitted, since it is false. Weathervanes sometimes mislead. They can be rusted in place; be under the control of mischievous children; be indicating, in the absence of any more recent wind, the direction from which it blew a day ago. In short, once a sign like a weathervane's pointing north has been guessed to be signifying a current local northerly, there is much room left for an interpretative (re)appraisal of such vital facts as the presence or apparent absence of mischievous children.

Chrysippus nicely demystified the tie between the signifier and what it signifies by arguing that any such tie is made in the mind. (He so emphasized this point, in fact, as to hold that the tie itself is the *only* mental entity, since for him both signifier and signified - his *semenon* and *semenomenon* - were real things in the real world. Whereas nowadays we recognize that it is not the weathervane itself that signifies to us, but our mental image of it. If its image is lacking, the weathervane has not even been seen.)

It was well recognized in antiquity that, while words may be the archetypical signifiers, many signifiers are not words. Aristotle, for instance, mentions in the *Prior Analytics* that a woman's lactating constitutes a sign (it signifies that she is pregnant); and Augustine, to take another instance, writes in *De Doctrina Christiana* of footprints as signs (of the passage of whoever owned the feet). In this fashion the groundwork was laid for a general science of signs, that is, for semiotics.

Interest in signs and their relations continued during the medieval and early Renaissance periods, notably among the many who seem to have devoted the bulk of their waking hours to writing exegeses of Aristotle. These activities centred, for reasons partly accidental, in Iberia: in Portugal (Fonseca, P. da and John of St Thomas, also known as Jean Poinsot) and, in an earlier and more arcane outbreak, in Catalonia (Llull, R.). Although this work retains much historical interest, it added little to semiotic theory, perhaps because so much of it was devoted to weighing in on one or another of the theological or quasi-theological debates that in those days took the place of free inquiry. With the full Renaissance, however, and the rebirth of interest in everything human, interest in semiotic problems occasionally took a different and more organized turn, notably in the work of George Dalgarno of Scotland (*c*.1619-87).

2 Modern developments

The next important advances came in the late nineteenth and twentieth centuries. First, Charles S. Peirce codified Aristotle's notion of enthymemic signification into the trivium: 'sign' (signifier) + signified objective (which Peirce unfortunately shortened to 'object') + 'interpretant'. It is worth especial notice that Peirce's 'interpretant' plays an essential role in his system, since for him (as for Aristotle) many and perhaps most signs are understood to be tied to their object(ive)s by probabilistic guesswork - by the intuitive leap that Peirce called 'abduction' - requiring the amplification and partial (de)confirmation available, with luck, from an appraising interpretant. More importantly, in Peirce's codification each such interpretant is often itself a sign and so in turn is tied to some signified object, or perhaps to more than one such, each being in need of a still further interpretant, and so on, in some cases *ad infinitum*. In this fashion Aristotle's enthymemic sequence of (minor premise) + (consequent) + (probabilistic substitute for the missing major premise) is converted into a unified chain of signs, each engaged in a triadic structure of sign + object + interpretant.

The three most important theorists since Peirce have been Charles Morris (1901-79), Thomas A. Sebeok (b.1920), and Umberto Eco (b.1932).

Morris' chief contribution was to put living flesh on Peirce's 'interpretant', which for Peirce was a discarnate or at least incorporeal activity - 'pure mind' perhaps - but which for Morris was an interpreter with a brain and a body. This is a gain, since who or what produces and interprets signs has a physical shape that can influence how signs are shaped and related to other signs. (Simple example: the deaf 'speak' their signs with their hands, face and torso; and, compared to vocal signs, the signs of the deaf are predictably different in form, in their modifiability, in their mutual resemblances, and in the ways in which they can mimic what they signify.) Similarly, the primary signs that begin the process of someone's assigning to some sign some significance - the process that is nowadays called 'semiosis' - are for Morris real and observable 'sign-vehicles' of interest in their own right. Both points are important, because how signs are emitted and received by physical entities (such as bodies and brains), and how the signs of a given system (such as the English language) are structured and inter-related, are topics to which Peirce himself gave short shrift. The distinction between corporeal and discarnate semiosis can be summed up in two terms: 'pragmatics' versus 'pure rhetoric', respectively Morris' and Peirce's terms for the study of how signs are related to their interpretants.

Thomas Sebeok's effect on semiotics as a field of study has been pervasive, and is universally recognized. He was co-founder (with Margaret Mead) of the modern semiotic discipline (in 1962); he was founding editor of the defining journal *Semiotica* (from 1968); and he has been the guiding mentor to the entire field. His theoretical contributions are likely to prove just as lasting. He has steadily broadened the scope of semiotic inquiry, for instance by delving into zoösemiotics (the study of infrahumans' signifying activities, in both fact and myth, and of their 'prefigurements of art') and into the sub-area of endosemiotics (the study of signifying among microbiota, neurons included). He has sharpened some of Peirce's most basic notions (see §3) and he has opened the eyes of many to the necessity of defining the logical nature of the basic connective 'signifies' (or 'is a sign of'), as by noting that that connective must be at least potentially reflexive and symmetric.

Umberto Eco, besides having written the 'semiotic' novels that have brought him deserved fame, has been deeply involved in the development of semiotic theory since the late 1960s. He is the author of well-known basic texts, has written on the history of semiotic inquiry and has striven to refine some of Peirce's fundamental concepts.

Among the many other twentieth-century scholars who have also made noteworthy contributions to this burgeoning field are Ferdinand de Saussure, Jacob von Uexküll (1864-1944), Karl Bühler (1879-1963), Roman Jakobson (1896-1982) and Louis Hjelmslev (1899-1965) (see Structuralism in linguistics).

3 Signifying by association, by resemblance, by convention

A given sign may do its signifying in different ways: in the widely used terminology laid down by Peirce, it may signify by physical association with what it signifies (such signifiers are called 'indexes'); by mimicry, that is, by physical resemblance to it ('icons'); and/or arbitrarily and merely by convention ('symbols'). For instance the sketchy representation of a man found on certain urgently sought doors in airports indicates by physical association (location) that this and not the next one is the door to the men's toilet; it indicates by recognizable physical resemblance to a man that this door leads to the men's facility (and not the women's); and it indicates the latter fact partly by convention, since generally it is a stick-figure with a pumpkin head and bears little resemblance to an actual person.

This three-way distinction of types of sign was first drawn by Augustine (*De Doctrina Christiana*), though the distinction between imitative icon and arbitrary symbol, or between symbol and associative index, had been hinted at earlier, respectively by Plato (especially *Cratylus*) and Aristotle (*Prior Analytics*). The distinctions are clear enough in theory, but their application is sometimes aporetic, as when deciding the extent to which a word like 'ding-dong' is imitative (hence 'iconic' in Peirce's sense) or arbitrary (hence 'symbolic'). Indeed, as Plato showed, albeit inadvertently, such discussions are often inconclusive or plain silly. (He argued in *Cratylus* that no Greek noun could be wholly arbitrary, unless imported from Phrygian or some other barbarian language, since any native Greek word could be partitioned, in the crackpot philology of the time, into meaningful subsegments.)

The ternary division of sign-types has a broad acceptance, though alternatives are possible. Certainly a binary division is conceivable - for instance, into Saussure's 'arbitrary' and 'non-arbitrary' types, the latter including both the iconic and the indexical - and it is not difficult to add a fourth or even fifth sign-type. One such is the sign that signifies by modelling the physical process of producing its real object, as the characters of the Korean Hangul

writing-system do (they model the mouth in the act of articulating the characters' corresponding sounds).

On the other hand, Peirce's basic tripartite division of sign-types seems quite robust in comparison with some of the other categorizations that have been proposed. For example, in George Dalgarno's tripartite typology, the earliest fully coherent one (in *Didascalocophus* (1680)), three kinds of signification are defined, but by such a hotchpotch of criteria as to leave the categories leaking rather badly. Dalgarno reserves 'conventional' signs (Peirce's 'symbols') for the exclusive use of humans, and 'natural' signs (Peirce's 'indexes') for infrahumans; supernatural signs, such as 'dreams' and 'apparitions' (these correspond to nothing in Peirce's basic typology) he reserves for the use of 'Almighty God'. (This last category is consistent with the pervasive belief in theurgy typical of Dalgarno's day, but it may have sprung from one of the more bizarre passages in *Cratylus*, where Plato avers that the gods have their own language. Proof: they call 'Xanthus' the river that humans call 'Scamander'.) Dalgarno would thus implicitly deny to humans the ability to signify anything by indexes (such as by correctly placing the 'men's toilet' sign) or, presumably, by such humanly transmitted 'apparitions' as plays or films.

4 Systematicity

Signs typically fall into sets, each sign sharing with its fellows distinctive properties of similarity of form and/or function. Thus the words of English are similar in their partaking of the English sound-system (so that a French word, if properly pronounced, cannot be an English word). Again, any arrangement of signs properly called a sentence is only the overt aspect of one or more covert underlying structures. The sentence: 'Politicians who hold floozies to be congenial are likely to be denounced from the pulpit', can be assigned either of two rather different structures and corresponding meanings. In one, certain politicians embrace floozies in order to be congenial; in the other, it is to the floozies that congeniality is attributed, by certain politicians (see Ambiguity). Such structures have been paid far more attention by linguists than by other semioticists, partly because they are presumably more intricate in language than elsewhere; but, clearly, some other systems of signification have intricacies of their own (think of a map of the London Underground), and much remains to be done.

Another way in which signs are parts of systems is this: the connective 'signifies' has logical or functional properties, which when well defined may be uniform. While the colloquial verb 'signifies' does not map neatly into a single logical connective, any more than the colloquial verb 'implies' does, much is gained by pondering this issue. As a defined propositional connective, 'signifies' is functionally equivalent to 'implies' or even 'if and only if'. For instance, when 'signifies' is narrowly defined then '(*p* signifies *q*) & (*q* signifies *r*)' signifies '*p* signifies *r*'. (Let '*p*' = 'that weathervane points southward', '*q*' = 'the local wind is southerly', and '*r*' = 'it will soon warm up'.) However, 'signifies' as a nominal connective is rather different. From 'this drawing signifies that cloud' & 'that cloud signifies an impending storm' it certainly does not follow that therefore 'this drawing signifies an impending storm'. Two different meanings of 'signifies' are being played on here - 'represents' and 'betokens' - and mixing them must often produce nonsense.

This said, it remains to observe that the construction of a comprehensive semiotic calculus is very much a work-in-progress.

5 The field

As presently constituted, the field of semiotics exists less as a discipline than as a speciality. The process by which someone receives or emits a sign, and interprets or conceives it - the process generally called 'semiosis' - is investigated by both semioticists and psychologists, mainly in a domain such as the processing of language, whether written or spoken. (In earlier semiotic work issues of semiosis were rather slighted, apart from some inconclusive remarks by Peirce on, for example, the 'firstness' (roughly, accessibility without mediation) of sense-impressions. This is nice, but it sharpens only slightly Augustine's pellucid definition of a sign as 'a thing which, over and above the impression it makes on the senses, causes something else to come into the mind as a consequence of itself' (*De Doctrina Christiana* 2.1.1).) The functioning of signs in literature, music, architecture, painting and drama is mainly pursued by semioticists who are also students of those respective arts. Examples of work in these and other areas of semiotics can be found in the references.

See also: Language, philosophy of; Moscow-tartu school

W.C. WATT

References and further reading

Aristotle (*c.* mid-4th century BC) *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984, 2 vols.(Contains *De Interpretatione*, which discusses signification, among other topics. Also contains *Prior Analytics*)

Augustine (396-426) *De Doctrina Christiana (On Christian Instruction)*, New York: Cima, 1947.(Brilliant and clear-headed discussion of signs and their function.)

Bréhier, E. (1951) *Chrysippe et l'ancien stoïcisme*, Paris: Presses Universitaires de France.(The best introduction to Chrysippus.)

Dalgarno, G. (1680) *Didascalocophus*, Menston: The Scolar Press, 1971.(Odd but fascinating early attempt to categorize signs.)

Deely, J. (1982) *Introducing Semiotic*, Bloomington, IN: Indiana University Press.(Especially good on medieval work.)

Eco, U. (1979) *A Theory of Semiotics*, Bloomington, IN: Indiana University Press.(On semiotics as a study of communicative codes.)

Eschbach, A. and Trabant, J. (eds) (1983) *History of Semiotics*, Philadelphia, PA: John Benjamins.(The study of signs from Plato to Wittgenstein.)

Harris, W.V. (1983) 'On Being Sure of Saussure', *Journal of Aesthetics and Art Criticism* 41: 387-97.(A nice account of Saussure's views on the arbitrariness or conventionality of the signs of language.)

Plato (427-347 BC) *Works*, trans. T. Taylor, London, 1804; repr. New York: Garland, 1984.(Contains *Cratylus*, with notes, which discusses signification, among other topics.)

Posner, R. (1982) *Rational Discourse and Poetic Communication*, Berlin: Mouton.(On the semiotic study of language and literature.)

Ross, W.D. (1965) *Aristotle's Prior and Posterior Analytics*, Oxford: Clarendon Press.(Excellent introduction to Aristotle's insights into signification in relation to logic.)

Sebeok, T.A. (1989) *The Sign and Its Masters*, Lanham, MD: University Press of America.(Authoritative account of modern semiotics.)

Tarasti, E. (1994) *Musical Semiotics*, Bloomington, IN: Indiana University Press.(The standard work in the field.)

Watt, W.C. (ed.) (1994) *Writing Systems and Cognition: Perspectives from Psychology, Physiology, Linguistics, and Semiotics*, Dordrecht: Kluwer Academic Press.(Modern work within the intersection of semiotics and psychology.)

Seneca, Lucius Annaeus (4/1 BC-AD 65)

Lucius Annaeus Seneca, Roman statesman and Stoic philosopher, is the earliest Stoic of whose writings any have survived intact. Seneca wrote, in Latin, tragedies and a wide range of philosophical works. His philosophical and literary work was carried out in the intervals of an active political career. He is most important for his ethics and psychology, although natural philosophy was not neglected. Unlike many Stoics he showed little interest in logic or dialectic. His most influential work was on the psychology of the passions, the nature of the human will and techniques of moral education; he also wrote extensively on social and political issues from a distinctively Stoic perspective.

1 Life and works

Seneca was born into a wealthy family of the equestrian class at Cordoba in Spain. His father, the Elder Seneca, saw to it that his son was educated at Rome, where he rose to become a senator. Exiled and recalled by the emperor Claudius, Seneca became the teacher and advisor of the emperor Nero. His influence on Nero was considerable until AD 62; Seneca eventually withdrew from active politics, but nevertheless was compelled to commit suicide in AD 65 for his presumed support of a conspiracy against Nero.

Seneca's Stoicism was affected by his early adherence to the Sextian school of philosophy (see Neo-Pythagoreanism), which emphasized asceticism and moral training. His Stoic education was thorough, and his works reveal the influence of Panaetius, Hecaton and Posidonius as well as the early heads of the school (see Stoicism §1). He rethought many aspects of Stoic philosophy and continued the work of Cicero in developing a Latin philosophical vocabulary. His prose writings display a balance between his personal contribution and inherited school doctrine. He was a part of contemporary literary culture, famous for his distinctive rhetorical prose style and as the author of justly admired tragedies. The relationship between his philosophical convictions and the tragedies is controversial, as is the question of the impact of his philosophical convictions on his political activity.

Many of Seneca's works are lost, including a biography of his father, speeches, letters and a late treatise entitled *Moral Philosophy*. His surviving prose works include three consolatory works, *To Marcia*, *To Polybius* and *To Helvia* (his mother). Most of his treatises on ethics were dedicated to close friends or family members, although *On Mercy* was addressed to Nero. Also of political note is a viciously witty satire on the dead emperor Claudius, the *Pumpkinification*. Late in his career Seneca wrote a lengthy work on physics, the *Natural Questions*.

The other extant ethical works are: *On Anger* (in some ways a companion piece to *On Mercy*); *On the Brevity of Life*; *On the Steadfastness of the Wise Man*; *On Mental Tranquillity*; the fragmentary *On the Private Life*; *On the Happy Life*; the long treatise *On Favours*, which deals with social relations as well as personal ethics; the fragmentary *On Providence*; and his most influential work, the *Letters to Lucilius* (more correctly *Letters on Ethics*, henceforth *Letters*).

Seneca is important for the history of Stoicism because he is the earliest professed Stoic any of whose works survive in complete form. It is not his aim to report on the history of the school, and his evidence for its early period must be used with care. His treatises confirm and elaborate on what is known about early Stoicism from other sources, but divergences and changes of emphasis are not uncommon. Examples of this include the treatment of a political figure, Cato the Younger, as a sage; an emphasis on suicide as the ultimate expression of personal freedom; and a humane application of the traditional Stoic view that there are no natural slaves.

Seneca's philosophical works have been persistently influential, first on Latin Church Fathers and again in the Renaissance; Montaigne's *Essays* owe much to Seneca's *Letters* (see Patristic philosophy; Montaigne, M.E. de; Renaissance philosophy).

2 Psychology

Seneca's greatest contribution is in moral psychology. In numerous works it is apparent that personal moral decision (*voluntas* or 'will') is given greater emphasis than it received in earlier Stoicism. The treatise *On Anger* (especially book II) also suggests (although this is a controversial point) that he has rethought the earlier Stoic

theory of assent and reassessed its role in the explanation of passions. He treats assent as a consciously made decision, a moment of personal self-assertion which shapes (rather than just reveals) our moral character. Will, for Seneca, has become the focus of individual moral freedom, much as *prohairesis* was to be for Epictetus (§3). It would be extreme to say that Seneca has posited a distinct faculty of 'will'; but he certainly took some steps in that direction and inspired others, such as Augustine, to go even further.

His views on the relationship between reason and the passions are also of note. He held to the view that passions are produced by the rational mind when it makes errors about the values of things, and denied the claim of Plato (§14) and Aristotle (§§22-3) that they are rooted in a non-rational part of the soul. Seneca did, however, take an interest in the complexity of the interactions between bodily reactions to external stimuli and our mental responses to them. His psychology may sometimes seem Platonic, but if there is 'dualism' in his theory, it is the body-soul dualism of Plato's *Phaedo* rather than the psychological dualism of the *Republic* (see Plato §§13-14). Hence Seneca rejected the Peripatetic view (see Peripatetics) that passions should merely be moderated; *apatheia*, complete freedom from passions, remained the ideal.

3 Ethics, physics, logic

Seneca emphasizes that he is not himself a wise man, but merely a *prokoptōn*, a person making moral progress. Like other Stoics of the first century AD and like Panaetius two centuries earlier, he focuses on the moral needs of his audience of imperfect but serious students of philosophy, rather than on the somewhat abstract deductions of earlier theorizing. The paradoxical technicalities of Stoic ethics are not unknown to Seneca, however; in the treatise *On Favours* he applies them effectively to serious ethical questions.

Seneca's attitude to physics is typical of his time, and is important as background to ethics: one must grasp the basic outlines of a providentially organized cosmos held together by a divine plan with which mankind is in harmony, but detailed debate about physics and cosmology is not worth much effort. Specific phenomena, such as those explored in the *Natural Questions*, are dealt with out of scientific curiosity or as illustrations of the divine rationality of the world we inhabit. Logic and dialectic are almost entirely neglected by Seneca. He was aware of work in this area, but comparison with Epictetus shows Seneca's limitations in logic (see Epictetus §2).

4 Seneca and philosophy in the first century AD

Seneca's works provide us with an invaluable picture of philosophical life in Rome in the first century AD. His reaction to Epicureanism (which ranges from selective appreciation in the early books of the *Letters* to the hostile denunciations of *On Favours* 4) shows that the school was still to be taken seriously. Similarly, *Letters* 58, 65 and 89 show that Seneca was familiar with the more technical doctrines of contemporary Platonic and Aristotelian philosophy, just as *Letters* 121 and 124 display his detailed knowledge of the technical side of Stoic ethics. Cynic philosophy was important in Rome, and Seneca anticipates Epictetus in his attitude to it: a Cynic (such as Seneca's friend Demetrius) can be presented as a paradigm of the wise person, despite the Cynic rejection of physics and logic as areas of study.

5 The *Letters*

The *Letters to Lucilius* have proved the most influential of Seneca's works. The dedicatee, Lucilius, was a real person, but the letters are fictitious, philosophical open letters modelled on those of Epicurus and designed to emulate the published correspondence of Cicero. Seneca plays the role of a moral advisor to Lucilius, whose progress can be followed through the sequence of letters. The tone and style of the letters also owes much to the rhetorical form known as the diatribe (see Cynics §3), and is often similar to that of his moral treatises; but there is a greater fluidity and intimacy. Although the facade of epistolary realism is often dropped, Seneca's pretence of writing to a close personal friend draws the reader into a personal engagement with Stoicism that could not otherwise be achieved; finely judged autobiographical revelations and confessions sustain a sympathy for Seneca as a moral teacher which is not evoked by the rest of his works. The *Letters*, in fact, paved the way for the development of the personal philosophical essay as a genre.

The *Letters* as we have them culminate with an investigation of a technical problem in Stoicism: is 'the good' grasped by the senses or by the intellect? Beginning with a quotation from Virgil, Seneca guides the reader through the complexities which establish that the genuine good is grasped by the mind and can only be found in a

rational soul. That rational virtue is the only good is a view fundamental to Stoic ethics, yet detailed physical and metaphysical argument is required to establish it. The reader is soothed by the literary grace of Seneca's style; the strategy of personal address promotes personal conviction. Seneca goes beyond Lucretius, who claimed to mask the bitterness of philosophical technicality with the honey of his poetry. However, Seneca used the tools of his literary art not to disguise the philosophy we must swallow, but to draw us as willing participants into its inner workings. Like the best of Plato's dialogues, Seneca's letters exploit the illusion of personal dialogue to make us partners in his own search for wisdom.

BRAD INWOOD

List of works

The dating system used below follows Griffin (1992); *On Providence*, being fragmentary, is hard to date with accuracy.

Seneca, L.A. (late 30s AD) *Consolation to Marcia*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1932.(Latin text with facing translation; addressed to the daughter of the historian Cremutius Cordus.)

Seneca, L.A. (40s AD, during his exile) *Consolation to Polybius* trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1932.(Latin text with facing translation; in part an attempt to regain favour at the court of Claudius.)

Seneca, L.A. (40s AD, during his exile) *Consolation to Helvia*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1932.(Latin text with facing translation; addressed to his mother, concerning his own exile.)

Seneca, L.A. (before AD 52) *On Anger*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1917-25, 10 vols.(Latin text with facing translation; especially important for Stoic theory of passions.)

Seneca, L.A. (*c*. AD 48-55) *On the Brevity of Life*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1928.(Latin text with facing translation; life is long enough for one guided by philosophy.)

Seneca, L.A. (*c*.50s) *On the Steadfastness of the Wise Man (De constantia sapientis)*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1928.(The wise man is immune to insult and injury.)

Seneca, L.A. (*c*.50s) *On Mental Tranquillity*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1932.(Latin text with facing translation; Stoic rationality as the key to contentment.)

Seneca, L.A. (*c*.50s) *On the Private Life*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1932.(Latin text with facing translation; on the relation between political and philosophical life.)

Seneca, L.A. (*c*.50s) *Pumpkinification (Apocolo-cyntosis)*, trans. W.H.D. Rousé, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1917-25, 10 vols.(Latin text with facing translation; satire on the emperor Claudius.)

Seneca, L.A. (*c*.50s) *On Mercy*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1928.(Latin text with facing translation; important for political theory as well as ethics.)

Seneca, L.A. (*c*.50s) *On the Happy Life*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1932.(Latin text with facing translation; a vigorous defence of the Stoic conception of life according to nature.)

Seneca, L.A. (*c*. AD 56-62) *On Favours (De beneficiis)*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1935.(Latin text with facing translation; general theory of social relations and ethically appropriate action.)

Seneca, L.A. (by AD 64) *Natural Questions*, trans. T. Coscoran, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1971-2.(Latin text with facing translation; essays on a range of topics in natural philosophy with ethical digressions.)

Seneca, L.A. (*c*. AD early 60s) *Letters*, trans. R.M. Gummere, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1917-25, 10 vols.(Latin text with facing translation.)

Seneca, L.A. (*c*. mid 1st century BC) *On Providence*, trans. J.W. Basore, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1928.(Latin text with facing translation.)

References and further reading

Arnold, E.V. (1911) *Roman Stoicism*, Cambridge: Cambridge University Press.(Standard work, somewhat dated but still useful.)

Chaumartin, F.-R. (1989) 'Quarante ans de recherche sur les œuvres philosophiques de Sénèque (Bibliographie 1945-1985)' (Forty years of research on Seneca's Philosophical Works), in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II.36: 3, 1545-1605.(Well articulated bibliography, arranged chronologically within each topic.)

Cooper, J.M. and Procopé, J.F. (1995) *Seneca: Moral and Political Essays*, Cambridge: Cambridge University Press.(Better translations of *On Anger*, *On Mercy*, *On the Private Life*, and *On Favours* 1-4, plus excellent introduction and commentary.)

Griffin, M. (1992) *Seneca: A Philosopher in Politics*, Oxford: Clarendon Press, 2nd edn.(Definitive historical treatment; offers a likely order of composition for Seneca's works.)

Grimal, P. (1979) *Sénèque, ou la conscience de l'Empire*, Paris: Les Belles Lettres.(A general assessment of his philosophical work and the historical context.)

Grimal, P. (1989) 'Sénèque et le stoïcisme romain', in W. Haase (ed.) *Aufstieg und Niedergang der römischen Welt*, Berlin and New York: de Gruyter, II.36: 3, 1962-92.(An assessment of Seneca's relation to the early Stoics.)

Sengzhao (AD 384?-414)

Sengzhao was one of the first native Chinese thinkers to develop a distinctive version of Buddhist philosophy. He blended the dialectical logic of Indian Mādhyamika Buddhism with ideas and terms often borrowed from Chinese Daoism. His collected treatises, the Zhaolun, argued that language is inadequate for capturing reality. There is a need for an intuitive insight to penetrate reality in its nonlinguistic form as a means to understanding the process and limitations of subsequent conceptualization.

Born in Changan (present day Xian) in AD 384 (or possibly 374), Sengzhao was from a poor family and found work as a copyist, an occupation that gave him broad exposure to a variety of literatures. Originally enamoured of the neo-Daoism popular at the time, Sengzhao converted to Buddhism after reading a copy of the *Vimalakīrti Sūtra*. As a Buddhist monk in northwestern China and then in Changan, he studied under the famous Indian Buddhist translator and commentator Kumārajīva (AD 344-413). Kumārajīva introduced Sengzhao to Mādhyamika thought, known in China either as the 'Four Treatise School' or the 'Three Treatise School' (see Buddhism, Mādhyamika: India and Tibet). As a chief disciple of Kumārajīva, Sengzhao achieved great eminence and his writings have been regarded as foundational to the Sinicization of Indian Buddhism.

Sengzhao's contribution to Chinese philosophy had multiple dimensions. First, he introduced a new form of rhetoric for philosophical argument. In this regard, Sengzhao often used models of argumentation taken from the second-century founder of Indian Mādhyamika philosophy, Nāgārjuna. Nāgārjuna had developed a form of dialectical analysis to show that opposing concepts could only be understood in relation to each other. Yet, the realities to which they referred were typically considered ontologically discrete. For example, the concepts of 'past', 'present' and 'future' are understood in terms of each other, yet the realities to which they supposedly refer cannot temporally coexist and, therefore, cannot be directly related. So, Nāgārjuna argued that the way concepts interrelate cannot mirror the way their supposed referents interrelate. Sengzhao assimilated aspects of this form of argument in analysing pairs of concepts. However, he did not always analyse the same pairs of concepts as had Nāgārjuna. Often his arguments seemed to address issues more central to the indigenous concerns among the neo-Daoist philosophers, for example. In this regard, it is important to note that some of Sengzhao's dialectical style of analysing opposing pairs of concepts also resembled the form of argument in Chapter Two of *Zhuangzi*, the famous Daoist philosophical classic (see Zhuangzi). Therefore, Sengzhao was developing a form of argumentation that drew on both Indian and indigenous Chinese traditions. This may account for the almost immediate reaction of his contemporaries that he had introduced a new and important style of philosophical argument.

Second, and closely related to the first point, Sengzhao's writings heightened the Chinese awareness of the logical structure of argumentation. Sengzhao did not seem to have a logician's grasp of the principles involved in Mādhyamika argumentation, sometimes making invalid inferences in his arguments. It is likely that he did not have access to Indian logical treatises. Therefore, without the help of theoretical explanation, he probably developed his logical arguments by trying to mimic the style of Mādhyamika writings. Despite this limitation, he often formulated arguments with a logical tightness that had not been attained previously by Chinese writing in their native language. This set the stage for further Chinese developments in this area as the Chinese Buddhists learned more of the Indian forms of rhetoric and logical analysis.

Third, Sengzhao gave the Mādhyamika idea of intuitive insight (*prajñā* in Sanskrit; *banru* in Chinese) a new explication. As in the Indian tradition, this intuitive insight was considered an immediate, perhaps mystical, grasp of reality without the intervention of conceptual understanding. One important difference from the Indian tradition, however, was that Sengzhao associated this intuitive insight with the Chinese tradition of the sage. In various indigenous philosophies, the Chinese sage had been described as having extraordinary powers of insight into reality. Sengzhao identified this quality with the Buddhist ideal of intuitive insight. This had the effect of making Chinese Buddhism, like Daoism and Confucianism, a 'sage religion'. Furthermore, it reinforced the indigenous idea that an immediate experience of reality was possible. Sengzhao stressed that since this intuitive insight did not have a conceptual object, it could not be a form of knowledge. It was something more 'spontaneous' or 'natural' than conceptual modes of apprehension. This emphasis on merging with the natural gave the Buddhist idea a

distinctively Daoist flavour (see Daoist philosophy §1; Zhi).

Lastly, although Sengzhao stressed the primacy of intuitive insight, he also stressed the dual-layered character of reality. On one level, there is a natural, organic source that can be grasped immediately. On another level are the functions of that source as the realm of conceptual expression, yielding a necessary but limited grasp of truth. This gave Sengzhao's Chinese Buddhism an emergent or process form of ontology describing how the non-conceptual became conceptualized. This organic model of reality was not present in the Indian Mādhyamika tradition and was much closer to ideas in Daoism and such ancient Chinese Classics as the *Yijing* (see Yijing).

Because of these varied dimensions to Sengzhao's work, scholars disagree about the best way to classify his own philosophical perspective. Was Sengzhao essentially a product of the Chinese, fundamentally neo-Daoist worldview who managed to graft on to the tradition some new rhetorical structures imported through Indian Buddhism? Or was he a convert to the Indian philosophy of Mādhyamika Buddhism who was able to stretch his native language and indigenous ideas to accommodate the new way of thinking he had adopted? Or was he a creative, syncretistic thinker who developed his own distinctively cross-cultural philosophy? Because his writings are not extensive and because he was breaking out of the boundaries of the traditional styles of writings, some critical points are open to multiple interpretation. What is undisputed, however, is Sengzhao's impact on the ensuing Chinese philosophers of various schools.

See also: Buddhist philosophy, Chinese; Daoist philosophy; Buddhism, Mādhyamika: India and Tibet

THOMAS P. KASULIS

List of works

Zhaolun (Treatises of Zhao) (AD *c*.384-414); ed. and trans. Tsukamoto Zenryū, *Jōron kenkyū (Studies of the Zhaolun)*, Kyoto: Hōzōkan, 1955; ed. and trans. W. Liebenthal, *Chao Lun: The Treatises of Seng-chao; A Translation with Introduction, Notes and Appendices*, Hong Kong: Hong Kong University Press, 1968. (Tsukamoto provides the best critical edition of the original Chinese text as well as a Japanese translation and philological essays in Japanese. Liebenthal is a full English translation of the *Zhaolun* with commentary emphasizing the Chinese, rather than Buddhist, nature of the text.)

References and further reading

Cai Zongqi (1993) 'Derrida and Seng-Zhao: Linguistic and Philosophical Deconstructions', *Philosophy East and West* 43 (3): 389-404. (Interesting comparison of the deconstructive strategies in the language use of Jacques Derrida and Sengzhao.)

Fung Yu-lan (1953) *A History of Chinese Philosophy*, vol. 2, trans. D. Bodde, Princeton, NJ: Princeton University Press. (A classic survey of Chinese philosophy with a section (pages 258-70) devoted to a summary of the three central fascicles of the *Zhaolun*.)

Robinson, R.H. (1965) *Early Mādhyamika in India and China*, Madison, WI: University of Wisconsin Press. (Good scholarly study of the Mādhyamika tradition, including a chapter on Sengzhao and a translation of three fascicles of the *Zhaolun*.In contrast to Liebenthal, Robinson emphasizes Sengzhao as a continuation of the Indian Mādhyamika tradition.)

Sense and reference

The 'reference' of an expression is the entity the expression designates or applies to. The 'sense' of an expression is the way in which the expression presents that reference. For example, the ancients used 'the morning star' and 'the evening star' to designate what turned out to be the same heavenly body, the planet Venus. These two expressions have the same reference, but they clearly differ in that each presents that reference in a different way. So, although coreferential, each expression is associated with a different 'sense'. The distinction between sense and reference helps explain the cognitive puzzle posed by identity statements. 'The morning star is the evening star' and 'The morning star is the morning star' are both true, yet the sentences differ in cognitive significance, since the former may be informative, whereas the latter definitely is not. That difference in cognitive significance cannot be explained just by appeal to the references of the terms, for those are the same. It can, however, be naturally accounted for by appeal to a difference in sense. The terms 'the morning star' and 'the evening star' used in the first sentence, having different senses, present the referent in different ways, whereas no such difference occurs in the second sentence.

The distinction between sense and reference applies to all well-formed expressions of a language. It is part of a general theory of meaning that postulates an intermediate level of sense between linguistic terms and the entities the terms stand for. Senses give significance to expressions, which in and of themselves are just noises or marks on a surface, and connect them to the world. It is because linguistic terms have a sense that they can be used to express judgments, to transmit information and to talk about reality.

1 Sense, reference and cognitive significance

The distinction between sense and reference was originally drawn by Gottlob Frege in his 1892 article 'Über Sinn und Bedeutung' ('On Sense and Reference'; see Frege, G. §3). According to Frege, associated with each meaningful expression of a language there is a 'sense' (*Sinn*) that makes the expression significant and determines its 'reference' (*Bedeutung*), that is, the entity the expression applies to or designates. ('*Bedeutung*' is also translated as 'designation', 'designatum', 'denotation' and 'nominatum'. It is also sometimes translated as 'meaning'. This can create confusion, since some of Frege's contemporaries, including Russell, used 'meaning' as the translation of '*Sinn*', most likely on the assumption that the notion corresponding to '*Sinn*' is that of cognitive meaning or significance.) Before 1892, Frege had espoused the view that the significance or content of an expression is given exclusively by that which the expression designates or stands for (what he came to regard as the 'reference'). Observe that on such a view two coreferential terms, such as 'Cicero' and 'Tully', are supposed to have the same content or significance. But if that is so, it seems impossible to explain the obvious difference in cognitive significance between, say, 'Cicero is Cicero', a trivially true and uninformative statement, and 'Cicero is Tully', a potentially informative sentence. The problem of how to explain differences in cognitive significance is not restricted exclusively to statements of identity. 'Cicero was an orator' and 'Tully was an orator' also differ in cognitive significance, since it is possible for a competent speaker to accept one of the statements while rejecting the other. Without postulating a difference in the content or significance of the terms 'Cicero' and 'Tully', it would seem impossible to explain the differences between those two pairs of sentences. The distinction between sense and reference provides for a natural explanation: although the two names are coreferential, they are associated with different senses and those senses present the referent in different ways (see Proper names §1).

2 The sense and reference of different types of expressions

The distinction between sense and reference is intended to apply to all meaningful expressions of a language, whether simple or complex. In the case of definite descriptions it is an extremely natural distinction since it is obvious that different descriptions of one and the same object present that object in different ways. In the case of proper names, the reference is the bearer of the name and the sense is traditionally conceived as some sort of condition satisfied uniquely by the referent. There is no unanimity, however, regarding how to understand such a uniquely identifying condition. Frege suggests (1892) that the sense of a proper name is some descriptive information that applies uniquely to the bearer. Thus, some philosophers interpret Frege's suggestion as entailing the claim than each proper name is synonymous with some definite description. Others, including John Searle (1958), view the sense of a given proper name as a cluster of descriptive information that speakers associate with

the name, and the referent as whatever satisfies most of the attributes in the cluster. Other semanticists do not uphold the descriptive nature of the senses of proper names. For example, David Kaplan (1969) has argued that on Frege's view the sense of a name may be seen as a representation of the referent and so the relation between the sense of a name and its reference is much like the relation between picture and pictured object. Finally, Michael Dummett (1981) is the foremost representative of the view that regards the sense of a proper name, as the sense of any expression, as a mechanism or procedure that, under ideal circumstances, would enable a competent speaker to single out the referent.

The reference of a general term is traditionally taken to be the extension of the term in question. Thus, for example, in the case of a predicate such as 'red' the extension is the set of red things, that is, the set of things to which the predicate correctly applies. And the extension of a relational verb such as 'is taller than' is a set of ordered pairs (since 'being taller than' is a relation between two things) such that the first member of each pair is taller than the second member of the pair. The sense of a general term is a mode of presentation of the set and is responsible for the cognitive content or cognitive significance of the term.

Although it is traditional to represent the reference of a general term as a set, such a choice is strictly incorrect, from the standpoint of pure Fregean orthodoxy. Frege distinguished sharply between concepts and objects. Only the latter are, on his view, 'saturated' entities; entities that subsist on their own. People and artefacts, also numbers and sets, are saturated entities. On the other hand, concepts are 'unsaturated'; they do not have an independent existence, but can only subsist as completed by a saturated entity. On Frege's view, the reference of a general term is a concept, and thus an unsaturated entity. Thus, the reference of a general term is not a set, for the latter is a saturated entity. Still, in the posthumously published comments on 'Über Sinn und Bedeutung' (1892-5), Frege himself stresses that any two general terms refer to the same concept just in case their extension is the same. So, it seems that, even for Frege, equality of extension is a criterion of identity for concepts, so taking the reference of a general term to be a set is not a radical departure from the standard Fregean theory.

Finally, according to the theory, whole sentences express complete thoughts. The 'thought' or sense associated with a sentence is the claim or assertion, traditionally called the 'proposition', expressed by the sentence. In 'Über Sinn und Bedeutung' Frege argues that the sentence's reference is its truth-value. The relation between a sentence and its truth-value should therefore be understood in terms of the relation between a name and its bearer or a definite description and the individual it designates. A true sentence is like a name for 'True', so that all the true sentences (like all the false ones) have one and the same reference; however, since different, non-synonymous sentences express different thoughts, they present that reference in different ways. Frege's argument relies on the assumption (see §4 below) that the reference of a complex expression does not change if components with the same reference are intersubstituted. Thus, for example, if we substitute the coreferential 'the orator who denounced Catiline' for 'Cicero' in the complex expression 'the father of Cicero', the reference of the expression (whoever fathered the famous orator) does not change. In this vein Frege asks what it is that remains constant when we intersubstitute coreferential components in a sentence, for that should be the reference, if anything is. He observes that the thought cannot be the reference, for typically the thought expressed by the sentence is altered: 'Cicero was an orator' and 'The orator who denounced Catiline was an orator' clearly express different thoughts. The truth-value of the sentence is what remains constant in spite of the substitution, so that should be the reference.

3 The multiple roles of sense

The sense of an expression, as we have seen, embodies its cognitive value and is solely responsible for its significance. A term that lacks a sense is therefore not a meaningful constituent of a language; however, an expression may have a sense (that is, be meaningful) and still lack reference. For example, 'the least integer' is a meaningful expression even if there is no object to which it refers. The sense of an expression determines the referent (if the expression has a referent) but, as Frege points out, there is no road back from reference to sense. Aristotle, for example, is the referent of many different terms; there is no mechanism, no procedure that would single out a sense only on the basis of what the reference is.

The sense of a sentence - the thought or proposition it expresses - determines a truth-value as reference. But truth-value typically depends on specific features of the world. Thus, given the way the world is, the sense of the sentence 'Aristotle was a philosopher' determines (we suppose) 'True' as reference. Had Aristotle never thought about philosophical questions, the sense of that sentence would determine 'False' as reference. It is the role of the

sense of a sentence to determine truth-value on whatever conditions obtain, thus the sense accounts for the truth-conditions of a sentence. Because the thought expressed by a sentence is a function of the senses of the component expressions, the sense of an expression is that expression's contribution to the truth-conditions of the sentences in which it figures.

Senses are also what speakers grasp when they understand words. Hence, language learning and linguistic competence are a matter of grasping senses. Moreover, speakers communicate linguistically with one another by grasping the same thoughts. So senses, although mentally graspable entities, should not be confused with private images or ideas. Senses are objective, so that different speakers can grasp the same senses and transmit a given thought to other speakers.

4 Compositionality and substitutivity

The Fregean theory of sense and reference espouses two basic and closely connected semantic principles: the principle of compositionality and the law of substitutivity.

The principle of compositionality states that the reference of a complex expression is a function of the references of its components, plus the way in which those components are combined (see Compositionality). Similarly, the sense of a complex expression is a function of the senses of the component expressions. Thus the reference of the sentence 'The Eiffel Tower is taller than the Empire State Building', a truth-value, is determined exclusively by the references of the meaningful components and their combination. The mode of combination is important, since a different reference would be determined if the components were combined differently, as with the sentence 'The Empire State building is taller than the Eiffel Tower'.

Since, by compositionality, the references of the parts are solely responsible for the reference of the whole, it obviously follows that coreferential terms are intersubstitutable without altering the reference of the complex expressions in which they occur. The latter is known as the law of substitutivity, which states that an expression which figures as part of another expression can be replaced with a coreferential one, and the reference of the complex will remain the same. However, the *sense* of a complex is typically altered when coreferential components are intersubstituted, if they have different senses. Thus, for example, 'Aristotle was Greek' and 'The tutor of Alexander the Great was Greek' express different thoughts but, since the latter sentence results from the exchange of coreferential terms, the reference of the two sentences is the same. (The sense of a complex remains the same if components with the same sense are intersubstituted.)

5 Non-customary contexts

The law of substitution seems to have obvious counterexamples. For example, although 'Cicero' and 'Tully' are coreferential names, the reference of sentences such as 'Mary said that Cicero was a historian' or 'John said, "Cicero was a historian"' may be altered if 'Tully' is substituted for 'Cicero'. But, according to the standard theory of sense and reference, some contexts cause shifts in the sense and reference of the expressions that occur in them. Thus the alleged failure of substitutivity is only apparent, for in the contexts in question the terms being substituted do not have their customary reference. The Fregean theory of sense and reference is essentially contextual: what an expression refers to and what its sense is depend on the context in which the expression occurs.

Consider the sentence 'The word "Cicero" contains six characters'. The reference of a word within quotation marks is not the customary referent, but rather the word itself (see Use/mention distinction and quotation). In our example, the occurrence of 'Cicero' does not refer to Cicero but to the word 'Cicero'. Clearly, the substitution of 'Tully' for 'Cicero' in that context would alter the truth-value of the sentence in question. But since 'Tully' and 'Cicero' are different words, the terms of the attempted substitution are not coreferential in that context, and therefore the observed phenomenon does not violate the principle of substitutivity.

Something similar occurs in the case of indirect quotation, such as 'Mary said that Cicero was a historian' (see Indirect discourse), and belief ascriptions, such as 'John believes that Cicero was a historian' (see Propositional attitude statements §§1-2). Here too the intersubstitution of what seem to be coreferential terms may alter the reference of the sentence. And here again, the breakdown of substitutivity is only apparent, for, according to the theory of sense and reference, expressions in contexts such as these, which Frege called 'oblique', have a non-customary referent. The non-customary referent in these cases - the oblique referent - is the customary sense

of the expression. In oblique contexts the reference of the embedded clause is the thought which that clause customarily expresses, and the reference of each of the clause's component expressions is its customary sense. So, the reference of 'Cicero' in the two sentences considered above is not Cicero but rather the sense 'Cicero' has in customary contexts such as 'Cicero was an orator', 'Cicero was a historian' and so on. In oblique contexts expressions have in turn an oblique sense, whose role is precisely to determine the oblique reference. It is an open question whether the semantic treatment of more complex embeddings requires that expressions have a higher-order sense that determines the oblique sense as reference in the context in question.

See also: Intensional entities §1

GENOVEVA MARTÍ

References and further reading

Burge, T. (1990) 'Frege on Sense and Linguistic Meaning', in D. Bell and N. Cooper (eds) *The Analytic Tradition: Meaning, Thought and Knowledge*, Philosophical Quarterly Monographs series, vol. 1, Oxford and Cambridge, MA: Blackwell, 30-60. (Proposes a nontraditional interpretation of the notion of sense.)

Church, A. (1951) 'A Formulation of the Logic of Sense and Denotation', in P. Henle, H. Kallen and S. Langer (eds) *Structure, Method and Meaning*, New York: Liberal Arts, 3-24.(A formal development of Frege's ideas. Church argues for the need for an infinite hierarchy of senses.)

Donnellan, K.S. (1972) 'Proper Names and Identifying Descriptions', in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel, 356-79.(A criticism of the Fregean theory of sense and reference as applied to proper names.)

Dummett, M. (1981) *Frege: Philosophy of Language*, Cambridge, MA: Harvard University Press.(A comprehensive discussion of various aspects of the theory of sense and reference.)

Evans, G. (1982) *The Varieties of Reference*, Oxford: Clarendon Press.(A contemporary Fregean approach to sense and reference. Chapter 1 has a thorough exposition and discussion of the doctrine of sense and reference for proper names.)

Evans, G. (1985) 'Understanding Demonstratives', in *Collected Papers*, Oxford: Oxford University Press. (Applies the theory of sense and reference to indexicals.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildeten Formelsprachen des reinen Denkens*, Halle: Nebert; trans. 'Conceptual Notation', in *Conceptual Notation and Related Articles*, ed. T. Bynum, Oxford: Clarendon Press, 1972, 101-203.(A sample of Frege's thought before the distinction between sense and reference. §8 gives Frege's first attempted solution to the problem posed by true and informative identity sentences.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. and ed. P.T. Geach and M. Black, 'On Sense and Meaning', in *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 3rd edn, 1980, 56-78; trans. H. Feigl, 'On Sense and Nominatum', in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Ridgeview, 1982, 85-102.(This is the essay in which Frege introduces the distinction between sense and reference.)

Frege, G. (1892-5) 'Ausführungen über Sinn und Bedeutung', in *Nachgelassene Schriften*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Hamburg: Meiner, 1969-76; trans. P. Long and R. White, 'Comments on Sense and Meaning', in *Posthumous Writings*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Chicago, IL: Chicago University Press, 1979, 118-25.(A posthumously published discussion of some of the major themes of 'Über Sinn und Bedeutung'. Frege focuses specially on the application of the notions of sense and reference to general terms.)

Frege, G. (1918) 'Der Gedanke: eine logische Untersuchung', *Beiträge zur Philosophie des deutschen Idealismus* 1: 58-77; trans. A.M. and M. Quinton, 'The Thought: A Logical Inquiry', *Mind* 65: 289-311; trans. P.T. Geach and R. Stoothoff, 'Thoughts', in *Collected Papers*, ed. B. McGuinness, Oxford: Blackwell, 1984.(Frege's defence of the objectivity of the realm of sense. It also includes a few remarks on the sense of indexicals.)

Furth, M. (1968) 'Two Types of Denotation', *Studies in Logical Theory*, American Philosophical Quarterly Monograph series, vol. 2, Oxford: Blackwell, 9-45. (Discusses unsaturated entities and the reference of general terms.)

Kaplan, D. (1969) 'Quantifying In', in D. Davidson and J. Hintikka (eds) *Words and Objections*, Dordrecht: Reidel, 178-214.(A discussion of Frege's treatment of oblique contexts. Argues for the representational conception of the sense of proper names.)

Parsons, T. (1981) 'Frege's Hierarchies of Indirect Senses and the Paradox of Analysis', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 6, 37-57.(Contends, against Church,

that just one level of oblique sense and reference is sufficient to provide an adequate semantic treatment of all contexts.)

Perry, J. (1977) ‘Frege on Demonstratives’, *Philosophical Review* 86: 474-97. (Criticizes the theory of sense and reference as applied to indexicals.)

Salmon, N. (1986) *Frege’s Puzzle*, Cambridge, MA: MIT Press.(A critical discussion of the puzzle of cognitive significance.)

Searle, J.R. (1958) ‘Proper Names’, *Mind* 67: 166-73.(On the cluster conception of the sense of proper names.)

Sluga, H. (ed.) (1993) *The Philosophy of Frege*, vol. 4, *Sense and Reference in Frege’s Philosophy*, New York: Garland.(Large collection of articles about the topic.)

Sense perception, Indian views of

Sense perception is considered in classical Indian thought in the context of epistemological issues - in particular, perception as a source of knowledge - and of psychological and metaphysical issues, for example, the relations of sense experiences to objects, to language and to the perceiving self or subject. The Sanskrit word used most commonly in philosophical investigations of sense perception is pratyakṣa, *a compound of* prati, *'before', and* akṣa, *'eye' or any 'organ of sense'; thus it should be understood as 'being before the eyes' or 'experientially evident' as an adjective, and 'immediate experience' or 'sense experience' as a noun. The meaning 'sense perception' is normal within philosophical inquiries. But just how many sense modalities there are is not to be taken for granted. In addition to the five types of sense experience commonly identified, 'mental' perception (as of pleasures, pains and desires), apperception (awareness of awareness) and extraordinary or yogic perception are sometimes counted as* pratyakṣa.

Views about the psychology of perception or, more broadly, about perception considered as part of the world are developed in religious and soteriological literature (literature about enlightenment and liberation) predating classical philosophical discussions. In Upaniṣadic, Buddhist and Jaina texts over two millennia old, perception is painted in broad strokes within spiritual theories of self and world that promote ideas of the supreme value of a mystical experience. Sense perception is usually devalued comparatively. Later, the psychology of perception becomes very advanced and is treated in some quarters independently of soteriological teachings.

Classical Indian philosophy proper is marked by tight argumentation and self-conscious concern with evidence. The justificational value of perception is recognized from the outset, in so far as any justifiers, or knowledge sources, are admitted at all. Nāgārjuna and others challenge the epistemological projects of Nyāya and other positive approaches to knowledge, prompting deep probing of perception's epistemic role. Views about veridicality, fallibility and meaningful doubt become greatly elaborated.

What do we perceive? Throughout classical thought, sharp disagreements occur over the perceptibility of universals, relations, absences or negative facts (such as Devadatta's not being at home), parts versus wholes, and the self or awareness itself. Issues about perceptual media (such as light and ether, ākāśa, *the purported medium of sound), about occult or spiritual perceptibles and about the very existence of objects independently of consciousness are hotly debated. A Buddhist phenomenalism is polemically matched by a Mīmāṃsā and Nyāya realism on a range of concerns.*

Probably through the influence of mysticism, verbalization of experience, however simple and direct, becomes suspect in comparison with experience itself; this suspicion is evident in concerns over the value of each in presenting reality, as well as in other, sometimes rather indefinite, ways. The judgment is prevalent that what prevents a person from living in an enlightened or liberated state is thinking - verbalizing experience, calculating, planning, and so on - instead of having pure experience, perceptual and otherwise, and thus living with a 'silent mind'. This attitude emerges in treatments of sense experience, reinforcing what is perhaps a natural tendency among philosophers to find the relations of experience and language problematic. Even in the root text of Nyāya, where the influence of yoga and mysticism is not so strong, perception is said to be a cognition that is nonverbal, avyapadeśya, *although there is considerable dispute about precisely what this means. The relations between various modes of experience and the language used with respect to them remains an ongoing concern of the very latest and most complex classical Indian philosophy.*

1 Psychology of perception

This entry is concerned with classical Indian views of perception. The reflections of modern philosophers not writing in Sanskrit (such as Aurobindo Ghose, whose innovative theory has not been sufficiently explored) and academic philosophers fall outside its purview. In this section, four topics will be discussed: (a) views of perception as episodic, along with the role of psychological dispositions, called *saṃskāra*, in recognition and related cognitive processes; (b) internal perception and the internal instrument, the *manas* or 'mind'; (c) peculiarly Indian views about the operation of sense organs; and (d) the relation of perception to consciousness. With the first three topics, labelling according to school - Buddhist Yogācāra, Nyāya, Mīmāṃsā, and so on - or individual thinker is done incidentally, mainly to provide minimal historical context. Most of the positions are pan-Indian,

common across periods as well as the usual metaphysical divides, although not invariably. With the fourth topic, the stances of individual schools are discriminated, since there is much disagreement about consciousness.

In early Buddhist texts, discussion of sense perception occurs within the context of the doctrine of 'no-self', *an-ātman* (in Pāli, *anatta*), which is advanced by the Buddha apparently to facilitate disidentification with desires and other phenomenal presentations - such disidentification being considered crucial to the attainment of a mystical enlightenment. Varieties of consciousness are listed by the Buddha's interpreters in efforts to analyse the make-up of the false conglomerate commonly thought to be a self. Thus emerges a picture of streams of experience integrated into an ego that is erroneously thought to endure. Perception is considered episodic in character; a perceptual moment lasts only an instant and is followed by a distinct occurrence.

So, too, throughout the approximately two-thousand-year history of Nyāya-Vaiśeṣika and most other classical schools is found a view of perception as episodic. (The *vṛtti* notion used in Sāōkhya and Advaita Vedānta, roughly 'modification', is a change to be understood temporally.) Despite the realism of some of these philosophies, standing in sharp opposition to the most prominent Buddhist schools, there is agreement that sense experiences occur as brief episodes in a temporal series, or as an intersection of multiple series (a sense experience results from both physical and psychological factors, according to some views). Much classical Indian psychology of perception traces a complex of causal factors giving rise to particular types of awareness, with sense experiences included within the broader project.

Some of these analyses are intricate. For example, with the phenomenon of recognition (as in 'This is that Devadatta whom I saw yesterday') a perceptual factor and a memory factor are identified. The memory factor is analysed in terms of *saṃskāra*, dispositions or subconscious traces formed by a blend of past experiences and actions. Disputes centre on how the perceptual factor (expressed by 'this' in the verbalization 'This is that Devadatta') fuses with or relates to the memory dispositions. In later thought, it is well recognized that verbalizable perceptual awareness always, or almost always (or on almost all analyses), contains an admixture of concepts understood dispositionally. What some Western thinkers would call beliefs are talked about in these same terms, namely, with reference to *saṃskāra*, as is a whole range of cognitive phenomena, such as doubting, planning, expecting, and so on.

2 Psychology of perception (cont.)

Dispositions, or subconscious traces, are not the only internal factors talked about in analyses of perceptual awareness. Though there is much controversy about the nature of a cognizer or self, there is surprising consensus about an internal perceptual organ called the *manas*, commonly translated 'mind' (and here rendered as 'sense mind'), and sometimes also called *antaḥkaraṇa*, the 'internal organ'. Phenomenologically, perceptions of pleasures, pains and desires are similar to perceptions of colours, sounds, tastes, and so forth. In the latter cases, sense organs mediate, or deliver, the perceptions; thus it would appear that there must be an organ for 'internal' perceptions, too. Moreover, awarenesses originating through different sense media are synthesized - or channelled or bundled together - in such a way that a person at least seems to get a multidimensional view of the self and the world, a presentation informed by multiple senses, including the internal colourings of hedonic tone and appetitive hue. The sense mind is postulated to serve the dual function of grasping desires and the like and bundling the displays of the external senses. It is normally considered 'internal' with respect to the physical organs, as somehow nearer the seat of consciousness, but is 'external' to the self or consciousness itself. It is an instrument of awareness, but, as Sāōkhya philosophers stress, it is part of nature, not of consciousness, nor is it in itself conscious, except according to some Buddhist theorists (see Sāōkhya §2).

The nature of *manas* is filled out differently according to different schools. One important difference concerns the apparent synthesis of the deliverances of the distinct senses. In Nyāya-Vaiśeṣika, it is held that the appearance of a synthesis is an illusion comparable to the firebrand that appears to be a circle but is really a single torch whirled around rapidly. Similarly, the sense mind, which is considered atomic and capable of delivering to the self only one bit of sense-mediated information at a time, moves very rapidly among the organs and internal occurrences, giving the impression of a unified, multidimensional moment of consciousness, whereas the mosaic is really a rapid series of perceptions (see Nyāya-Vaiśeṣika §6). Again, this fuller view is by no means held by all schools, although a core theory or simple postulation of a sense mind is.

What relations hold between sense organs and the objects they grasp? According to some, all the organs come into direct contact with their objects, while others maintain that direct contact is the operative relation only with certain organs and certain objects. Most parties agree that touch operates by direct contact, also taste and smell, and most hold that the sense mind, or internal organ, works similarly too. A prominent view, though less universally espoused, is that with the seat of vision being the eyes, the visual organ itself is a ray that goes out of the body as swiftly as light (which carries it), comes into contact with objects at a distance and spreads over their illuminated surfaces. To obvious objections, proponents come up with ingenious rejoinders. Concerning hearing, there are advocates of a travelling organ, but a more prominent view is that sounds are transmitted in a special element called ether, *ākāśa*, alternatively through air, reaching the organ located in the hole in the ear. Some hold that the ether as delimited by the earhole is the sense organ.

There are issues about the relation to the grasping sense organ of objects not plausibly in direct contact with it, such as the colour and shape of a pot. We touch a pot, but not its shape, at least not directly. The visual organ may come into direct contact with the pot, but colour is not the sort of thing with which there could be direct contact, contact being a relation that holds between substances. Such issues are not pressing for Buddhist phenomenalists, since they hold that there is no essential difference between the means to awareness and awareness itself: talk of 'organs' is just a manner of speaking. But on almost all other views, considerable attention is paid to the nature of the operative relation for each type of object perceived, for example, substance, quality (and different types of quality - colours, sounds, and so on), motion, universal or absence.

Finally, the major schools take distinct positions on the question of the consciousness in perception. Yogācāra Buddhists (Buddhist phenomenalists) generally hold that every perception is conscious in itself. To postulate an additional cognizer would be otiose. Sāōkhya-Yoga propounds a strict dualism of consciousness and nature, with all perceptual apparatus as well as content regarded as part of nature. Consciousness is purely a witness, with no further, thicker relation to that which it beholds. Advaita Vedānta takes a similar view: consciousness has no relation to the modifications of the internal organ other than to be their witness. Mīmāṃsā and Nyāya-Vaiśeṣika philosophers uphold a more integrated position. In Mīmāṃsā (or, more precisely, the Mīmāṃsā of Kumārila and followers), awareness is considered a personal act, creating the quality of 'being-perceived' in an object perceived. Nyāya-Vaiśeṣika thinkers view awareness, perceptual and otherwise, as a quality or attribute of the self, a quality of brief duration, as explained above, and the result of complex causal processes. In philosophy, an awareness may be verbalized in abstraction from the person to whom it belongs, but there are no free-floating awarenesses. Nor are selves conscious except in possessing awarenesses (see Awareness in Indian thought).

3 Epistemology of perception

It is difficult to appreciate classical views on the epistemic value of perception in isolation from fuller epistemological projects where perception is identified as one of several sources of knowledge (see Knowledge, indian views of). The Sanskrit word *pratyakṣa* is normally used with an implication of veridicality; a nonveridical sense experience is not counted as *pratyakṣa*. Westerners are prone to react to this by suspecting classical Indian thinkers of missing crucial issues by presupposing too much in how they understand a term. But it is the translation 'perception' that is a little misleading. The English term 'knowledge' is normally used to imply truth of a belief. If what was considered a bit of knowledge is discovered to be in fact a false belief, then it is said that there was no knowledge in the first place, only a belief that was considered true. The same logic holds for *pratyakṣa*, substituting veridicality for truth and awareness for belief: if what was considered a perceptual awareness is discovered to be in fact nonveridical, then it is said that there was no perception in the first place, only an awareness that was considered veridical.

Little is to be gained in epistemology from semantic reflections. There is, however, a nice tie between the veridicality presupposition in the use of the term *pratyakṣa* and an important point about default justification, fallibility and the burden of proof that is central to much classical Indian epistemology. This is that doubt, like all awareness, has grounds, indeed causal grounds (as is brought out lucidly in later reflection), such as one's inability to discern a distant object that appears, say, to be perhaps a post, perhaps a person. Disagreement can be a cause for doubt, as can other sorts of cognitive tension. In such situations, we are called to look to reliable sources of information to resolve the doubt. But when the conditions for such meaningful doubt do not obtain - regarding sense experience, when conditions for doubting the reality of what we see or otherwise perceive do not obtain -

then we proceed, and have every right to proceed, assuming the experience's veridicality. Veridicality is the default assumption. And the defeasibility, or fallibility, of perceptual judgments, contributes to this right, to the default being an assumption of veridicality (along with an assumption of the truth of perceptual judgments). Were we to claim infallibility, or certitude, then the burden of proof would be on our shoulders; that is to say, the conditions for meaningful doubt would be a lot less stringent. Much of this position is worked out by members of the Nyāya school in response to epistemological scepticism advanced by the Buddhist Nāgārjuna and others. But it is also developed by the Yogācāra Buddhist Dharmakīrti and some of his followers.

Perception, *pratyakṣa*, is held to be the fundamental knowledge source, the 'eldest' (*jyeṣṭa*) source, by philosophers belonging to several distinct schools. The Cārvāka sceptic - who, with Nāgārjuna and his school and the later dialectical Advaita Vedānta of Śrīharṣa, is a common target of those with positive epistemological projects - accepts the epistemic value of perception; what the Cārvāka questions is the authority of inference and other putative knowledge sources. Nyāya itself is concerned with inference only in so far as it is to provide knowledge of the world, holding that any correct inference must have a perceptual premise. Moreover, inferential cognitions that are *prima facie* correct can be overridden by perception. This is clearly made out. Less clearly discerned, and often not at all, is the importance of inference to show the nonveridicality of a sense experience. The Dharmakīrti school, however, does propound a combination of coherentism and experiential foundationalism where considerations of coherence are viewed as potential overriders and weighted more heavily than any perceptual judgment (see Dharmakīrti; Epistemology, Indian schools of §1).

4 Perception and the world

A perceptual awareness is intentional; that is to say, it has an object or content, namely what it is *of*. But just what types of things is perception capable of revealing, and - because much of perception's epistemic value hangs on the following refinement - just what is it capable of *directly* revealing?

Yogācāra Buddhists, who are nominalists as well as phenomenalists, hold that what perception reveals is invariably unique and particular (see Nominalism, Buddhist doctrine of §§1-2). Dharmakīrti develops this stance into a pragmatism about concept-formation, justification and truth, finding inferential relations to hold only on the level of our mental fictions, not among things (or events) themselves. Indian realists, in contrast, paint a much richer picture of what we directly perceive.

To take up a highly contentious example, on the issue of how, according to the realists, an inference-underpinning pervasion, *vyāpti* (every locus of *F* as also a locus of *G*), is known, the view is that universals or class characters (natural kind characters) and relations among them are presented perceptually. To counter such objections as that from knowing the universal we should be acquainted with things strewn throughout all time, realists admit that internal factors are relevant to what in perception we are capable of verbalizing and acting upon. Of course, not everything *F* or *G* is known perceptually; what is known is *F*-hood and *G*-hood and the relation between them. A child may not be able to recognize cowhood in its generality, though the universal is available in the presence of Bessie the family cow; there is a difference between seeing cowhood and seeing it as cowhood (see §6). The realists stress their fallibilism, as well as the training often required to recognize something for what it is. Thus the relation between smokiness and fieriness, for example, may not be recognized; or we may think we perceive a connection, and a future counter-instance proves us wrong. It is wide experience (*bhūyo-darśana*) of correlations of things *F* and *G* that gives us confidence. Nevertheless, since a pervasion is a natural occurrence, one can be grasped on a single occasion. To use less technical terms, can we not know from a single witnessing that smoke is caused by fire? (Causal relations are important types of pervasion, on the realists' view, but not the only types.)

Other important ontological issues fought over with reference to direct perceptibility are relations (in particular that which the realists find between a property and a property-bearer - there is no such distinction, claim the Buddhists and, for different reasons, the Advaita Vedāntins), absences or negative facts, and the self or awareness itself (see Ontology in Indian philosophy §§1-2, 10-11; Negative facts in classical Indian philosophy). However, the focus here will be on the debates about the reality of wholes and the very existence of objects independently of consciousness; these stand close to the heart of the Buddhist-realist controversy, where so much of the best of classical Indian argument over perception is to be found.

The Buddhist tradition that the great Dharmakīrti defines, both by summarizing the arguments of his predecessors

and by setting the terms of the debate for his followers, finds a chasm between what experience directly presents and conceptual constructions. Vasubandhu (*c*.400) entertains the possibility of a universal illusion - the world as presented perceptually does not differ from the presentations of a dream - to land in a phenomenalism of event-instants. These presentations (atomic events of colour-shape, of sound, of feeling, and so on) are common to illusions and veridical experiences, common to dreaming and to waking. Wholes are mentally or imaginatively constructed (*kalpanā* is the key term: 'imagination' in classical aesthetics) out of these parts, as are enduring objects. And though important variations among Buddhists appear (whole schools are differentiated by Vasubandhu and others), generally speaking it is a mark of our spiritual ignorance that we take our constructions to be real. The only true difference between dreaming and waking is that between individual and mass delusion, between an imaginative construction or projection due to individual karma (dispositions built from past actions) and projections due to common karma. (Dharmakīrti says the nature of our constructions are due to our desires.)

With this view as background, Buddhists find an easy target in the realist view of wholes (such as in Nyāya), for the claim is not only that a whole, such as a tree, is in itself a reality other than the totality of its parts, but that when we perceive a tree we directly perceive the whole. (How ridiculous! Are the roots then directly perceived?) The deep issue is what counts as basic evidence. The Buddhists maintain that it is the atomic-event presentations common to illusions and veridical experiences. Nyāya philosophers counter that we see wholes, and they do not mean to use a different sense of 'see': we directly perceive wholes; perception of wholes is the basic evidence.

5 Perception and the world (cont.)

The *Nyāyasūtra* (2.1.33-5) presents two nonphenomenological lines of argument to shore up the position (phenomenological appeals, at this point, would beg the question). The first is that the view that we see a part presupposes that we see a whole - a part of what? Moreover, if we see only parts, then presuming a part is a macro object, it must have parts, and so *it* turns out not to be seen. Or if it is partless, it is so minute as to be imperceptible (some Buddhists do admit that, strictly speaking, the event-instants are not consciously perceived). Thus if we cannot see wholes, we can see nothing at all. The second line of argument is that only by assuming a whole can we explain holding and pulling. Only wholes, not parts (at least of some things), can be moved about. The realists also purport to have an answer to the Buddhist argument about a common core to veridical and illusory experiences in their theory of nonveridical awareness (see Error and illusion, Indian conceptions of §§1-2). Concerning objects' independence of consciousness, they continually invoke considerations of intersubjectivity and then, in later periods, supplement such arguments with an appeal to parsimony (*lāghavatva*, theoretic 'lightness') as the decisive advantage of their position.

The Buddhists, for their part, seem driven by an idealism that is tied to an understanding of what makes *nirvāṇa* experience possible: if objects were independent of experience, there could be no enlightenment, no consciousness of a single interdependent reality as is taught by the Buddha. On the other hand, Dharmakīrti achieves a pragmatism built on a more moderate view of a fundamental reality, indeed a sophisticated scepticism. That is, what is fundamentally real may be problematic, but as far as our everyday activities are concerned, all we need take to be real is what is capable of satisfying our desires. Our perceptions are to be understood as revealing things that are in this way causally efficient, so long as they cohere with other bits of knowledge.

Backing away from extreme idealist statements would seem an advantage polemically. But the realists get the last word in the debate with the Buddhists, for Nyāya and Mīmāṃsā as schools outlast Buddhist philosophy in India by several centuries. In the later period, many Buddhist arguments and positions are, however, taken over by Advaita Vedāntins.

Concerning what some call yoga-born perception and the spiritual matters putatively revealed by it, many philosophers addressing the topic do not seem themselves to have been yogins, and the views presented involve only minor extensions of common positions in the psychology of perception. Through yoga and the accumulation of religious merit, the organs of perception are modified, particularly the sense mind (*manas*), such that spiritual matters, or things subtle, remote or normally hidden, are directly perceived. Such speculation is supposed to show how extraordinary perception is possible; given the possibility, an epistemic parallelism between ordinary and extraordinary perception is upheld.

6 Perception and language

A range of views about the relation, or multiple relations, between perception and language-mediated thought - running alongside it, unfolding concepts implicit in it, perverting it, abstracting from it, referring to things revealed by it - occurs in classical Indian schools. We shall begin with a quick look at the panoply of positions and views in an abbreviated and general form.

Bhartṛhari, an astute grammarian and psychologist of the fifth century, appears to embrace a metaphysics where all awareness would be manifestation of a primordial Word, with several gradations or levels of ideas/language. He more than anyone else is responsible for provoking classical philosophers to suspect that there is no non-language-mediated consciousness. But Patañjali too, the author of the *Yogasūtra* (*c.*400), promotes the view that short of the highest mystic trance, the ultimate liberation and enlightenment, all mentality is shot through with seeds of concepts and speech. As mentioned, it is in Buddhist traditions that we find views of perception as utterly foreign to language: the mind-stream runs alongside perception or alters it so that it is no longer in its language- and concept-free native state. The realists are influenced on this score by their opponents, and the *Nyāyasūtra* pronounces perception to be nonverbal. Nevertheless, as a knowledge source, perception is taken not only to reveal things but to warrant judgments about the things that it reveals. Dharmakīrti himself, though in a hedged and qualified manner, accords perception a similar role. Moreover, it seems all the later philosophers come to recognize that it is only speech behaviour, *vyavahāra*, reflecting experience (presumably), that is of any use in philosophical debate. Universally, the need is felt to be able to account for, or at least take into consideration, common perceptual judgments and a wide range of speech acts.

The fourteenth-century Navya-Nyāya (‘New Logic’) philosopher Gaōgeśa (§2) accepts as a first principle of verbalizable awareness - that is, awareness reflected in speech behaviour and useful in our lives - that whatever an awareness is of, namely a qualificandum, is cognized as something or other: anything *x* is known as some predicate , a qualifier. But in order to be able to recognize and talk about a qualificandum as , we have to have a prior awareness of ; otherwise, we would not be able to see or talk about the object as . Thus it would seem that has to be cognized as something or other, and an infinite regress of qualification looms. In other words, Gaōgeśa views even the simplest verbalizable awareness as having an object that is ontologically complex. Such an awareness is *viśiṣṭa*, ‘qualified’, by which he means that a qualificandum is cognized through a qualifier. Thus in order to avoid an infinite regress of qualification (the qualifier$_1$ of qualificandum$_1$ as itself a qualificandum, qualificandum$_2$, to be known through another qualifier, qualifier$_2$, *ad infinitum*), Gaōgeśa posits indeterminate awareness, where qualifiers are directly cognized. To cognize Bessie as a cow presupposes acquaintance with cowhood, that is, awareness of the ground for application of the term ‘cow’. Thus there appears a danger of an infinite regress: *how* is the qualifier cognized? And so qualifiers such as cowhood must be directly available - in what Gaōgeśa calls, borrowing a Buddhist term, indeterminate (or non-propositional) awareness, which may be considered as a first phase of the causal process that results in a determinate, verbalizable awareness. Thus, in sum, there are two phases of sense perception according to New Logic, an indeterminate one where qualifiers are directly grasped in an immediate and nonverbalizable way, followed by determinate perception, which is propositional in content and verbalizable in terms that capture a fact perceived (that is, given that the perception is veridical).

Seeing cowhood and seeing it *as* cowhood are not the same. The ground for recognizing the universal would be ‘cowhoodness’, which is, according to Gaōgeśa’s followers, not an additional real. But here we threaten to pass into ontological and logical matters that occupy the latest of the New Logicians, and to exceed the confines of our charge.

See also: Buddhism, Yogācāra school of §§3-8; MīmāṃsāČsect;2; Mind, Indian philosophy of; Perception

STEPHEN H. PHILLIPS

References and further reading

Bhattacharya, G. (trans.) (1976) *Tarkasaṃgraha-dīpikāon Tarkasaṃgraha by Annaṃbhaṭṭa*, Calcutta: Progressive Publishers.(Lucid presentation of Nyāya views related to perception - through the vehicle of translation and explanation of a Nyāya textbook.)

Bhattacharyya, S. (1993) *Gaōgeśa’s Theory of Indeterminate Perception*, New Delhi: Indian Council of Philosophical Research.(Translates Gaōgeśa on indeterminate awareness.)

Dasgupta, S. (1922) *A History of Indian Philosophy*, vol. 1, Cambridge: Cambridge University Press, repr. 1969. (Remains probably the best overall introduction to classical Indian systems.)

Dignāga (c.500) *Pramāṇasamuccaya (Collected Writings on the Acquisition of Knowledge)*, ch. 1 trans. Masaaki Hattori, *Dignāga, On Perception*, Cambridge, MA: Harvard University Press, 1968. (Dignāga is the most important Yogācāra Buddhist philosopher except for Dharmakārti, who saw himself as Dignāga's follower.)

Matilal, B.K. (1986) *Perception: An Essay on Classical Indian Theories of Knowledge*, Oxford: Oxford University Press.(A modern classic by a premier scholar and philosopher.)

Mukhopadhyay, P.K. (1984) *Indian Realism*, Calcutta: K.P. Bagchi.(A modern defence of the spirit - and often the letter as well - of Nyāya realism.)

Nyāyasūtra (*c*.150), trans. M. Gangopadhyay, *Nyāyasūtra with Vātsyāyana's Commentary*, Calcutta: Indian Studies, 1982.(A readable and accurate translation.)

Sinha, J. (1969) *Indian Epistemology of Perception*, Calcutta: Sinha Publishing. (Probably too difficult for those unfamiliar with Sanskrit philosophical terminology, but a storehouse of information and astute commentary on classical views.)

Warder, A.K. (1970) *Indian Buddhism*, Delhi: Motilal Banarsidass, 2nd edn, 1980.(A classic historical study.)

Sense-data

A philosophical theory of perception must accommodate this obvious fact: when someone perceives, or seems to perceive something, how things appear may differ from how they are. A circular coin tilted will look elliptical. A stick partially immersed in water will look bent. Noting that appearance and reality do not always coincide, some philosophers have given the following account of the contrast between the two. Suppose someone seems to see a book with a red cover. Whether or not there is any book to be seen, the individual seeming to see the red book will be aware of something red. What they are aware of is called a sense-datum. According to a sense-datum theory, any perceptual experience involves awareness of a sense-datum whether or not it is an experience of a physical object.

Some philosophers link a sense-datum theory with certain views about knowledge. According to foundationalists all knowledge of the external world must rest on a foundation of beliefs that are beyond doubt. We can always be mistaken about what physical objects are like. On the other hand, we cannot be mistaken about what sense-data are like. So, all knowledge about the external world rests on beliefs about sense-data. In this way a sense-datum theory is supposed to do double duty in contributing towards an account of perception, and an account of knowledge based on perception.

1 Sense-datum theories

The term 'sense-datum' was coined early this century by Bertrand Russell. However, sense-datum accounts of perception go back much further. When Locke, Berkeley and Hume refer to ideas of sense, sensible qualities or impressions of sense frequently they appear to be referring to what we would now call sense-data.

In this century early advocates of sense-data intended the term 'sense-datum' to be neutral. Philosophers adopting opposed views about perception were supposed to be able to agree that any case of perceiving involves an awareness of sense-data, even if they then disagreed about what sense-data actually are and how they relate to the physical. A sense-datum may be a physical object, or it may be non-physical; it may exist only when perceived, or exist unperceived; it may be an event, or belong to some other category; it may be perceivable by only one individual, or more than one. But, however these issues are resolved, it is said to be undeniable that I am aware of a circular gold sense-datum. In this way sense-data, understood neutrally, are supposed to provide the foundation on which to erect competing theories of perception.

Sense-data are supposed to explain how perceptual experiences can fail to match reality. The book appears red even though it is some other colour. How can this be? Because the individual looking at it is aware of a red sense-datum. The coin appears elliptical even though it is circular: the individual looking at it is aware of an elliptical sense-datum. Sense-data, it is said, are what one is aware of in virtue of having perceptual experiences.

Call a perceptual experience veridical just in case in having it one seems to perceive an object with those qualities. I have a perceptual experience as of seeing a grey surface. My experience is veridical because I am seeing a grey surface. Now consider someone who is having a non-veridical experience. Consider the alcoholic who has an experience as of seeing a pink rat. According to the sense-datum theorist, even if not seeing a pink rat, the alcoholic is having an experience as of seeing something pink because they are aware of something pink. The alcoholic is aware of a pink sense-datum. Hence, sense-datum theorists are committed to:

(i) Every perceptual experience involves an awareness of a sense-datum.

G.E. Moore once toyed with the idea that a sense-datum can appear to have characteristics it lacks. However, allowing that sense-data may not be as they appear undermines a central role that sense-data are supposed to play in perception. Sense-data are appealed to, in part, to explain how the apparent characteristics of a physical object can differ from its actual characteristics. That explanation would be undermined if the apparent characteristics of a sense-datum could, likewise, differ from its actual characteristics. For that reason, sense-datum theorists are also committed to:

(ii) At least for a certain range of characteristics, it is impossible for a sense-datum to appear to have a characteristic that it lacks.

A physical object can appear red or round without being red or round. A sense-datum cannot appear red or round without being red or round.

Apart from (i) and (ii) other features have commonly been attributed to sense-data. Some say that sense-data are private: it is metaphysically impossible for more than one individual to be aware of the same sense-datum. Some say that sense-data are transient: it is metaphysically impossible for one sense-datum to be perceived discontinuously. Some say that sense-data are self intimating: for a restricted range of characteristics, a sense-datum must appear to have any characteristic that it has. These features, privacy, transience, mind dependence, and self intimation are optional in the following sense. Lacking any of the features in question will not disqualify sense-data from playing a role in the theory of perception.

A.J. Ayer proposed an unorthodox interpretation of sense-datum theories. His interpretation has become known as the alternative language view. Ayer accepts that a sense-datum theory is not open to empirical confirmation or disconfirmation. Since that is so, he is reluctant to treat a sense-datum theory as a theory about the existence of a special class of entities called sense-data. Instead, he views advocacy of a sense-datum theory as a recommendation to use language in a particular way. Consider again the coin that first appears circular to Jayne, then elliptical. She could describe the perceptual facts by saying that she was aware first of a circular sense-datum, then of an elliptical sense-datum. Alternatively, she could describe them by saying that the coin she perceived changes from being circular to being elliptical. Which description is correct? In Ayer’s view neither. Which description Jayne selects, the sense-datum description or the one that has implausible implications about the stability of objects, is simply a matter of convenience.

2 Arguments for the existence of sense-data

I will use ’sense-datum’ to refer to things that, at least, satisfy (i) and (ii). What reason is there to think that there are sense-data? Many arguments have been offered, but all depend on contestable assumptions that are rarely defended in the literature.

The perception of parts argument One cannot perceive an object unless one perceived every part of it. Whenever one looks at a physical object one does not perceive every part of it. So, whenever one looks at a physical object, one perceived something distinct from it: a sense-datum. This argument is only as plausible as its first premise, which to many people does not seem plausible at all.

The certainty argument Suppose one takes oneself to see a tomato. One cannot be certain that one is seeing a tomato. Nevertheless, one can be certain that one is seeing something that looks red, round and bulgy. So, the red-, round- and bulgy-looking thing one is certain of seeing is not the tomato, or any part of it. It is a sense-datum.

This argument depends on three assumptions. First, it is assumed that, in the case of seeing the tomato, one cannot be certain that a tomato exists. Second, it is assumed that one can be certain that a red-, round-, and bulgy-looking thing exists. Finally, it is assumed that if I am certain that a red-, round- bulgy-looking thing exists but uncertain that a tomato exists, then the red-, round- bulgy-looking thing is not the tomato. In general, however, it does not follow from someone being certain that a exists, and uncertain that b exists, that a is not identical with b.

The causal argument Perception involves a causal chain extending from a physical object to a perceiver. What is really perceived is only what is at the end point of the causal chain: a sense-datum. This argument depends upon the unwarranted assumption that only the end point of a causal process which enables one to see something can be perceived.

The time lag argument When someone perceives a very distant object, for example a star, light takes a long time to reach their perceptual apparatus. Because it does so, the star may have ceased to exist by the time the silvery speck taken to be the star is perceived. Hence, the silvery speck is not identical with the star. It is a sense-datum. This argument depends upon the assumption that nothing can be perceived unless it exists at the same time that it is perceived.

The argument from analysis Any case of veridical or non-veridical perception involves having a perceptual experience and, it is agreed, having a perceptual experience is best analysed as awareness of an object. The object in question is a sense-datum.

Why should we think that having a perceptual experience is best analysed as awareness of an object? Suppose someone sees a red after-image against a white wall. Surely that individual is aware of *something* red which is both an after-image and sense-datum. After all, there seems to be something, an after-image, obscuring part of the wall in much the way that a physical red patch would do.

Or consider the following three cases. In the first someone is looking at some yellow curtains through blue tinted spectacles. In the second someone is looking at the same curtains without anything impeding or distorting that person's vision. In the third someone is hallucinating a yellow surface without seeing any physical surface with that colour. The second and third cases share something is common which they do not share in common with the first. In the first someone in aware of some yellow curtains without being aware of their colour. In the second and third someone is aware of the colour, yellow, which is in fact the colour of the curtains. How can someone be aware of the colour of the curtains in the third case, unless they are aware of *something*, a sense-datum, with that colour?

The problem with the argument from analysis is that it relies on the sense-datum theorist's analysis of having a perceptual experience being superior to any other. However, alternative analyses have been defended in the literature. One implies that perceptual experiences invariably have objects, but treats the objects of perceptual experiences as intentional (see Intentionality). Another treats perceptual experiences as dispositions to believe (see Perception). According to a third, the adverbial analysis (see Mental states, adverbial theory of), perceptual experiences need not have objects. For example, if it is said that someone is experiencing a red after-image, the expression 'red' is used to specify how the person is experiencing rather than a quality of an experienced object.

Frank Jackson has recently argued for the superiority of a sense-datum analysis over these competitors. For example, he argues that an analysis of perception in terms of sense-data can, in contrast to an adverbial analysis, explain the validity of certain incontestably valid arguments concerning appearances.

3 The relationship between sense-data and physical objects

Direct realists (see Perception §4) hold that in perception one is often directly or immediately aware of the surfaces of physical objects. The famous argument from illusion has been used to refute direct realism and is said to show that sense-data are never identical with the surfaces of physical objects. Suppose Jayne looks at the same surface of a coin over an interval of time during which it is rotated. At first the coin looks circular. Later it looks elliptical. Since the coin does not change shape, the elliptical items that Jayne is later aware of are, presumably, not identical with the surface of the coin. As the coin rotates Jayne is aware of a series of things. The first member is circular, and the last elliptical. It is implausible to maintain that the first member of the series is a physical surface, and the later elliptical members things of a totally different kind. We should allow that, as the coin rotates, all of the differently shaped things Jayne perceives are sense-data. In general, one immediately perceives only sense-data whenever one perceives a physical object.

What is the relationship between immediately perceived sense-data and physical objects? According to representative realists, sense-data interpose between the physical world and the perceiver's awareness of it and represent physical objects to the perceiver in virtue of standing in an appropriate relation, usually though of as causal, to those objects (see Perception §2). According to one version of idealism physical objects are collections of mind dependent sense-data (see Idealism §2). According to one version of phenomenalism statements about physical objects are analyzable as statements about the sense-data one would have if one were suitably located (see Phenomenalism).

It should be noted that, of the above views about the relationship between sense-data and physical objects, only representative realism requires reference to sense-data for its formulation. There are versions of idealism and phenomenalism that make no mention of sense-data. On the other hand, any version of representative realism will have to incorporate reference to sense-data.

4 Objections to sense-data

A multitude of objections have been raised against sense-data, to which, naturally, sense-datum theorists have replies.

The individuation objection. When I see something am I aware of a single visual sense-datum, or a number of sense-data? How can I tell whether I am aware of the same sense-data at different times? These questions, it is said, are unanswerable since we lack criteria for the identity of sense-data.

This objection presupposes that lack of precise criteria for individuating sense-data constitutes an objection to their existence. However, we lack precise criteria for individuating many things we believe in such as persons or ships.

The private language objection. If sense-data exist, we can refer to them. Sense-data are essentially private objects. Hence, sense-data can only be referred to using a private language. However, it is agreed, private languages are impossible (see Private Language Argument). This objection presupposes that sense-data must be private, and, more dubiously, that one cannot refer to private objects using a public language.

The intelligibility objection. For a sense-datum theory to be intelligible certain distinctions have to be grasped such as the distinction between sense-data and material objects, or the distinction between direct and indirect perception. However, these distinctions are not really intelligible when they are applied as the sense-datum theorist wishes to apply them. For example, it only makes sense to say that something is directly perceived if it is possible to indirectly perceive it. According to the sense-datum theorist, a sense-datum can only ever be directly perceived.

This objection appears to rest on much disputed theses about the nature of language, and in particular on the thesis that an expression only meaningfully applies to something if there could be a thing of the same kind to which the expression does not apply.

The indeterminacy objection. If something exists it must be determinate in the sense that, for any property, it either has or lacks that property. But sense-data are not determinate. I see a hen which appears to have a certain number of speckles. Despite that, for any number of speckles, the hen does not appear to have that number of speckles. So, the sense-datum I am aware of when I see the hen is speckled without having any particular number of speckles. It is indeterminate.

This objection assumes that, not only must a sense-datum be as it appears, but a sense-datum must appear in every respect as it is. Unless this is so, the sense-datum of the speckled hen may, like the hen itself, have a determinate number of speckles without appearing to have that number of speckles.

See also: Foundationalism; Perception, epistemic issues in §5

ANDRÉ GALLOIS

References and further reading

Armstrong, D.M. (1961) *Perception and the Physical World*, London: Routledge & Kegan Paul.(A critique of sense datum theories combined with an alternative analysis of perception in terms of the acquisition of belief.)

Austin, J.L. (1962) *Sense and Sensibilia*, Oxford: Oxford University Press.(A sustained attack on sense datum theories.)

Ayer, A.J. (1964) *The Foundations of Empirical Knowledge*, London: Macmillan. (An exposition and defence of Ayer's alternative language view.)

Jackson, F. (1977) *Perception: A Representative Theory*, Cambridge: Cambridge University Press.(An ingenious defence of a sense datum analysis of perception in the context of defending a representative theory of perception.)

Price, H.H. (1932) *Perception*, London: Methuen.(A classic and detailed exposition of a sense datum theory.)

Sprigge, T.L.S. (1970) *Facts, Words, and Beliefs*, London: Routledge & Kegan Paul.(Chapter I contains an excellent response to the objections raised to sense data by Austin and others.)

Swartz, R.J. (ed.) (1965) *Perceiving, Sensing, and Knowing*, New York: Doubleday.(A collection on perception which includes an excellent selection of pieces on sense data in section II entitled Sensing, Sense Data, and Appearing.)

Sergeant, John (1623-1704)

John Sergeant was the last of the Blackloists - the faction of English Catholics who followed the thought of Thomas White in the middle decades of the seventeenth century. As such his major philosophical concern was with a refutation of scepticism, and to that end he adopted a synthesis of Aristotelianism and aspects of the new philosophy. Sergeant is now best remembered as the main Catholic protagonist in the theological 'rule of faith' debates, when he engaged with such eminent Anglicans as Tillotson and Stillingfleet to argue the cause of religious certainty, attainable only through Catholic tradition. Turning later to the need for certainty in philosophy and science, he formulated his own 'solid philosophy': within an essentially Aristotelian framework, he incorporated aspects of the new thought that had been earlier adopted by Thomas White and Kenelm Digby.

John Sergeant's conversion to Catholicism took place after his graduation from St John's College, Cambridge, in 1643. Having then spent some twelve years at the English Catholic College at Lisbon, he returned to England in 1655, and acted as Secretary to the English Catholic Chapter during the height of its Blackloist influence, until 1667. His reputation as a disciple of the discredited Thomas White, and his own incrimination in the 'Popish Plot', have resulted in long historical neglect. He remains, however, not only of theological importance, but also of considerable philosophical interest. In particular, his two late works, *The Method to Science* (1696) and *Solid Philosophy Asserted* (1697), are significant as purporting to provide the means for a last-ditch defence against encroaching scepticism.

Sergeant is essentially a late Aristotelian, and he specifically opposes such proponents of new thought as Descartes and Locke. The trouble with these 'ideists' is that they reduce our ideas or conceptions to mere representations or similitudes. We are then left without criteria for distinguishing which similitudes are true and which false, and we are reduced to predicating reality only of the contents of our own minds; and this is to be reduced to scepticism. The remedy is to ground one's knowledge in empiricism: simple notions are to be assumed, which leave bodies and enter through our sense-organs into our minds. Such notions are actual physical entities, actual parts of the original body; so that our apprehension of them is incontrovertible, inasmuch as they do not *represent* the external object, but actually *are* a part of that object.

With notions thus conceived as sense-data, with respect to which there is no room for misunderstanding, some solid foundation has been laid for the acquisition of certain knowledge. Admittedly, there is room for subsequent misunderstanding, as cognition consists in linking together those notions in an orderly way, such that there is 'conformity of the understanding to the thing'; but, by taking care to avoid misunderstanding and linguistic ambiguities, we may proceed to certain knowledge. Whereas 'ideists' ground their philosophical work on abstractions which have a dubious relationship with the object they purport to represent, Sergeant claims that his own empirically-based philosophy is founded on direct knowledge of 'things themselves'. From that secure starting-point, further knowledge can be attained by means of traditional Aristotelian syllogistic reasoning.

Sergeant's position, then, is derived in part from Aristotle, for whom direct apprehension of external objects gave clear though necessarily limited knowledge, and for whom syllogistic reasoning provided the subsequent path to truth. But Sergeant is indebted also to his Blackloist friends Thomas White and Kenelm Digby, who had earlier adopted a 'mechanical' theory of perception, in terms of which material particles were directly received by our sense-organs, having emanated from bodies in an imperceptible but material 'effluvium'.

Sergeant's philosophy is ultimately underpinned by his theology. In this his prime concern is with asserting the necessity of Catholic tradition to ensure certainty, but he also reveals the religious foundations of his assumed philosophical certainties. Thus, the self-evident propositions on which his logic is to be based are themselves derived from a deity who justifies the most supremely self-evident proposition that 'self-existence is self-existence'. Without God's self-existence, the whole of logic itself might be suspect; and without belief in a God who created an orderly nature, our assumption of a correlation between our logical theories and the natural world would be further undermined. 'An atheist can have no perfectly certain knowledge or evidence of anything'; and such a sceptical outcome has by all means to be resisted.

See also: Aristotelianism in the 17th century

BEVERLEY SOUTHGATE

List of works

Sergeant, J. (1665) *Sure-footing in Christianity*, London.(Sergeant's first major contribution to the 'rule of faith' debate, in which he argues in favour of Catholic tradition.)

Sergeant, J. (1667) *Faith Vindicated from Possibility of Falsehood*, Louvain.(An assertion of the 'firmness and certainty' of Catholic faith.)

Sergeant, J. (1696) *The Method to Science*, London.(A spirited attempt to refute scepticism by establishing grounds for philosophical certainty.)

Sergeant, J. (1697) *Solid Philosophy Asserted against the Fancies of the Ideists*, London; repr. New York: Garland, 1984.(An anti-sceptical treatise, which includes in particular a criticism of Locke's *Essay*.)

Sergeant, J. (1698) *Non Ultra: or a Letter to a learned Cartesian*, London.(A criticism of Cartesian claims for certainty, and a recommendation of syllogistic reasoning.)

References and further reading

Bradish, N.C. (1929) 'John Sergeant: a forgotten critic of Descartes and Locke', *Monist* 39: 571-628.(An early attempt to resuscitate Sergeant as a critic of canonical philosophers.)

Cooney, B. (1973) 'John Sergeant's criticism of Locke's theory of ideas', *Modern Schoolman* 50: 143-58. (Discusses relationship of Sergeant's anti-Lockean epistemology with his theory of mind-body interaction.)

Krook, D. (1993) *John Sergeant and his Circle: A Study of Three Seventeenth Century English Aristotelians*, ed. and intro. B.C. Southgate, Leiden: E.J. Brill. (Analyses Sergeant's philosophy as essentially Aristotelian; includes bibliography.)

Phemister, P. (1993) 'Locke, Sergeant, and Scientific Method', in T. Sorell (ed.) *The Rise of Modern Philosophy: The Tension between the New and Traditional Philosophies from Machiavelli to Leibniz*, Oxford: Clarendon Press.(Compares Sergeant with Locke in relation to scientific method.)

Southgate, B.C. (1998) '"Beating down scepticism": the solid philosophy of John Sergeant, 1623-1707', in M.A. Stewart (ed.) *English Philosophy in the Age of Locke*, Oxford: Clarendon Press.(Outlines Sergeant's anti-sceptical philosophy and argues for Blackloist influence.)

Set theory

In the late nineteenth century, Georg Cantor created mathematical theories, first of sets or aggregates of real numbers (or linear points), and later of sets or aggregates of arbitrary elements. The relationship of element a to set A is written $a \in A$*; it is to be distinguished from the relationship of subset B to set A, which holds if every element of B is also an element of A, and which is written* $B \subseteq A$*. Cantor is most famous for his theory of transfinite cardinals, or numbers of elements in infinite sets. A subset of an infinite set may have the same number of elements as the set itself, and Cantor proved that the sets of natural and rational numbers have the same number of elements, which he called* $\aleph_0$*; also that the sets of real and complex numbers have the same number of elements, which he called c. Cantor proved* $\aleph_0$ *to be less than c. He conjectured that no set has a number of elements strictly between these two.*

In the early twentieth century, in response to criticism of set theory, Ernst Zermelo undertook its axiomatization; and, with amendments by Abraham Fraenkel, his have been the accepted axioms ever since. These axioms help distinguish the notion of a set, which is too basic to admit of informative definition, from other notions of a one made up of many that have been considered in logic and philosophy. Properties having exactly the same particulars as instances need not be identical, whereas sets having exactly the same elements are identical by the axiom of extensionality. Hence for any condition Φ *there is at most one set* $\{x|\Phi(x)\}$ *whose elements are all and only those x such that* $\Phi(x)$ *holds, and* $\{x|\Phi(x)\} = \{x|\Psi(x)\}$ *if and only if conditions* Φ *and* Ψ *hold of exactly the same x. It cannot consistently be assumed that* $\{x|\Phi(x)\}$ *exists for every condition* Φ*. Inversely, the existence of a set is not assumed to depend on the possibility of defining it by some condition* Φ *as* $\{x|\Phi(x)\}$*.*

One set x_0 *may be an element of another set* x_1 *which is an element of* x_2 *and so on,* $x_0 \in x_1 \in x_2 \in \ldots$*, but the reverse situation,* $\ldots \in y_2 \in y_1 \in y_0$*, may not occur, by the axiom of foundation. It follows that no set is an element of itself and that there can be no universal set* $y = \{x|x = x\}$*. Whereas a part of a part of a whole is a part of that whole, an element of an element of a set need not be an element of that set.*

Modern mathematics has been greatly influenced by set theory, and philosophies rejecting the latter must therefore reject much of the former. Many set-theoretic notations and terminologies are encountered even outside mathematics, as in parts of philosophy:

pair	$\{a,b\}$	$\{x \mid x = a \text{ or } x = b\}$
singleton	$\{a\}$	$\{x \mid x = a\}$
empty set	$\emptyset$	$\{x \mid x \neq x\}$
union	$\cup X$	$\{a \mid a \in A \text{ for some } A \in X\}$
binary union	$A \cup B$	$\{a \mid a \in A \text{ or } a \in B\}$
intersection	$\cap X$	$\{a \mid a \in A \text{ for all } A \in X\}$
binary intersection	$A \cap B$	$\{a \mid a \in A \text{ and } a \in B\}$
difference	$A - B$	$\{a \mid a \in A \text{ and not } a \in B\}$
complement	$A - B$	
power set	$\wp(A)$	$\{B \mid B \subseteq A\}$

(In contexts where only subsets of A are being considered, $A - B$ *may be written* $-B$ *and called the complement of B.)*

While the accepted axioms suffice as a basis for the development not only of set theory itself, but of modern mathematics generally, they leave some questions about transfinite cardinals unanswered. The status of such questions remains a topic of logical research and philosophical controversy.

1 The axioms of set theory

The axiom of *extensionality* states that sets having exactly the same elements are identical. The axiom of *foundation* states that there is no infinite descending sequence of sets $\ldots \in x_2 \in x_1 \in x_0$. The basic

Zermelo-Fraenkel axioms (ZF; see Zermelo 1908) consist of the axioms of extensionality and foundation and existence axioms stated below. The accepted axioms ZFC consist of ZF plus the axiom of choice (AC): given a partition or set *x* of sets that are non-empty (each has at least one element) and disjoint (no two have a common element), there exists a choice set *y* having exactly one element in common with each element of *x*.

The intuition behind the axioms is that sets lie in a hierarchy of levels (see Set theory, different systems of §5), and when $x \in y$, then *x* lies below and *y*, above. At the lowest level there is just $\emptyset$ (though a variant of the standard notion allows at this level 'individuals' or items that have no elements and are not sets but may be elements of sets of any level). There is no highest level (though a variant allows a top level of 'classes' or items that may have sets of any level as elements but are not themselves sets nor elements of anything). If set *x* lies at one level then the 'successor' of *x*, $x^{\#} = x \cup \{x\}$ lies at the next higher level. The levels continue into the infinite, so that above $\emptyset$, $\emptyset^{\#} = \{\emptyset\}$, $\emptyset^{\#\#} = \{\emptyset, \{\emptyset\}\}, \ldots$ lies the set $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots\}$, whose existence is asserted by the axiom of *infinity*. The hierarchy is supposed to be as 'tall' and as 'wide' as possible, an intuition (only) partly expressed by the remaining existence axioms. These consist of axioms asserting the existence of the pair, union and power sets, along with infinitely many further axioms of each of two kinds. First, for any condition $\Phi(x)$ there is an axiom asserting that, for any set *u*, *separation* from *u* of those elements *x* for which $\Phi(x)$ holds forms a set:

$$\{x \in u | \Phi(x)\} = \{x | x \in u \text{ and } \Phi(x)\}$$

Second, for any binary condition Ψ there is an axiom asserting that if for every *x* there is a unique $y = \psi(x)$ such that $\Psi(x, y)$ holds, then for any set *u*, *replacement* of each element *x* of *u* by $\psi(x)$ forms a set:

$$\{\psi(x) | x \in u\} = \{y | \Psi(x, y) \text{ for some } x \in u\}$$

(In a rigorous formalization, the vague, intuitive notion of 'condition' would be explained in terms of a precise, logical notion of a 'formula' of a symbolic language.)

To develop mathematics on the basis of these axioms, for each system of numbers, points or other mathematical objects, a system of set-theoretic objects must be specified to serve as substitutes. Usually the system of set-theoretic substitutes is easy to specify intuitively, in terms of the relationship of the substitutes to the original objects, but it may be less easy to define formally, in terms taken only from the language of set theory. Set-theoretic counterparts to the basic operations on the original objects must also be defined in the language of set theory, and the counterparts of the basic laws for these operations deduced from the axioms of set theory. When this is done, the set-theoretic substitutes are thereafter called by the name of the mathematical objects in question, which are said to have been 'identified with' their set-theoretic substitutes.

No more can be done here than to describe intuitively for the natural, integral, rational, real and complex numbers the related sets with which these numbers are identified. First, each natural number $n \in \mathbb{N}$ is identified with the *set* of natural numbers less than *n*, so that $0 = \emptyset$, $1 = \{0\} = \{\emptyset\}$, $2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}$, and so on; and $\mathbb{N}$ itself with the set whose existence is asserted by the axiom of infinity (above). For other kinds of numbers, one needs the identification of the ordered pair (a, b) of a, b with the set $\{\{a\}, \{a, b\}\}$, which allows one to define

$$\begin{aligned}
A \oplus B &= \{(0, a) | a \in A\} \cup \{(1, b) | b \in B\} \\
A \otimes B &= \{(a, b) | a \in A \text{ and } b \in B\} \\
B^A &= \{F \subseteq A \otimes B | \text{ for all } a \in A \text{ there is} \\
&\qquad \text{a unique } b \in B \text{ with } (a, b) \in F\} \\
\mathrm{dom}(F) &= \{a | (a, b) \in F \text{ for some } b\} \\
\mathrm{ran}(F) &= \{b | (a, b) \in F \text{ for some } a\} \\
F \lceil C &= \{(a, b) \in F | a \in C\} \\
F[C] &= \{b | (a, b) \in F \text{ for some } a \in C\} \\
F^{-1}[C] &= \{a | (a, b) \in F \text{ for some } a \in C\}
\end{aligned}$$

and to define a relation on *A* or a function from *A* to *B* to be a subset of $A \otimes A$ or an element of B^A, respectively, so that, intuitively, a relation *R* or function *F* is identified with the set of pairs (a, b) such that *aRb* or $F(a) = b$. (Then the notions dom(*F*), ran(*F*), $F \lceil C$, $F[C]$ and $F^{-1}[C]$ agree with certain customary notions from

mathematical analysis, where they are called the ‘domain’ (or set of arguments), ‘range’ (or set of values), ‘restriction’, ‘image’ and ‘pre-image’.)

Now an integer *w* can be identified with the set of pairs of naturals whose difference is *w*; a rational *x* with the set of pairs of integers whose quotient is *x*; a real *y* with the set of rationals $x < y$ or, alternatively, with the set of functions *f* from naturals to rationals such that $f(x)$ for larger and larger *x* gives a better and better approximation to *y*; and a complex *z* with the ordered pair of its real and imaginary parts (each a real number). Ultimately any hypothesis that can be stated in conventional mathematical language can be stated in the language of set theory, and any theorem provable by conventional mathematical methods can be proved from the axioms ZF, with AC, or one of its many known implications, such as ‘dependent choice’ (DC), being needed in certain branches of higher mathematics. (DC asserts that if *R* is a relation on *A* such that for every $a \in A$ there exists at least one $b \in A$ with *aRb,* then there exists a function *F* from $\mathbb{N}$ to *A* with $F(n)RF(n+1)$ for all $n \in \mathbb{N}$.)

2 Ordinal numbers and ordinal arithmetic

One equivalent of AC that is widely used in mathematics is called ‘Zorn’s lemma’. A ‘partial order’ on a set *A* is a relation *R* such that for all $a, b, c \in A$ one has

aRa	(reflexivity)
If aRb and bRa then $a = b$	(antisymmetry)
If aRb and bRc then aRc	(transitivity).

R is an ‘order’ (also called a ‘total order’ or ‘linear order’) on a given $C \subseteq A$ if for any $a, b \in C$ either *aRb* or *bRa* (dichotomy). If *R* is an order on (all of) *A*, one often writes ‘ $\leq$’ (‘less than or equal to’) for *R*, and uses the common notations and terminology ‘ $<$’, ‘ $\geq$’, ‘ $>$’ and so on in the usual way. Even if *R* is only a partial order on *A*, one can use such notions as that of an ‘upper bound’ *d* for $C \subseteq A$ (*cRd* for all $c \in C$), but one must be careful to distinguish the notion of a ‘maximum’ element ($d \in C$ such that *cRd* for all $c \in C$) from that of a ‘maximal’ element ($d \in C$ such that, for any $c \in C$, *dRc* implies $c = d$), and similarly for ‘minimum’ and ‘minimal’.

Zorn’s lemma states that if *R* is a partial order on *A* and every $C \subseteq A$ on which *R* is a total order has an upper bound, then *A* has a maximal element.

The equivalent of AC most widely used in set theory is the ‘well-ordering’ principle, which asserts that every set can be well-ordered. A ‘well-order’ on a set *A* is a total order which also has the property that every non-empty subset of *A* has a minimum element (groundedness).

Cantor introduced two kinds of transfinite numbers: cardinal and ordinal. The cardinal (number) of a set is concerned only with the ‘size’ of the set; an ordinal also with the relationship of its elements. To define an ordinal, we must first define isomorphism. A function $F: A \rightarrow B$ is a ‘bijection’ (that is, a one-one, onto mapping) if for every $b \in B$ there is exactly one $a \in A$ with $F(a) = b$. Given (total) orders *R, S* on sets *A, B,* a function $F: A \rightarrow B$ ‘preserves order’ if for any $a, a' \in A$ one has aRa' if and only if $F(a)SF(a')$. An ‘isomorphism’ is a bijection that preserves order. The ‘order type’ of an order *R* is written $|R|$. $|R| = |S|$ if there is an isomorphism from *R* to *S*, and $|R| \leq |S|$ if there is an isomorphism from *R* either to *S* or to an initial sub-order of *S*, $S_b = S \lceil B_b$. An ‘initial sub-order’ is the order on the ‘initial subset’ $B_b = \{b' \in B | b'Sb \text{ and } b' \neq b\}$, for some $b \in B$. An ‘ordinal’ is the order type of a well-order.

If ρ is the ordinal of a well-order on a set of cardinality α, then ρ will also be said to be of cardinality α. By the well-ordering principle, every cardinal is that of some ordinal; and it follows from the definitions that if ordinals ρ, $\mathbf{\Sigma}$ are of cardinality α, β and $\rho \leq \mathbf{\Sigma}$, then $\alpha \leq \beta$. (To avoid confusion arising from the traditional use of the same notations for both cardinal and ordinal numbers, we will use the letters α, β, γ for cardinals, and ρ, $\mathbf{\Sigma}$, τ for ordinals.)

Cantor proved that the order on the ordinals (defined above) is a well-order. Hence there is a least or zero ordinal. For every ordinal ρ there exists a greater ordinal and hence, again by the definition of a well-order, a *least* greater ordinal, called its successor, $\rho^{\#}$. For every set *T* of ordinals there is an ordinal greater than or equal to each of its elements, hence a least such ordinal, called its ‘supremum’, $\sup T$. (Note this implies that there can be no set of all ordinals.) A ‘limit’ (ordinal) is one that is neither zero nor a successor.

The ordinals 0, $1 = 0^{\#}$, $2 = 1^{\#}, \ldots$ less than the least limit are called 'finite', and the least limit $\sup\{0, 1, 2, \ldots\}$, the ordinal of the natural order on the natural numbers, is called ω_0 or simply ω. It follows from the definitions that if $\mathbf{\Sigma}$ is the ordinal of a well-order *S* on a set *B*, then there is an isomorphism *F* from *S* to the order on the set $\{\rho | \rho < \mathbf{\Sigma}\}$ of ordinals less than $\mathbf{\Sigma}$ (namely, $F(b) = |S_b|$ for $b \in B$). It is now customary to identify the ordinal $\mathbf{\Sigma}$ with this set of lesser ordinals (generalizing what was done in the case of finite ordinals - natural numbers - in §1). With this identification, it is easy to describe explicitly $\rho^{\#}$ and $\sup T$: they are just $\rho \cup \{\rho\}$ and $\cup T$, respectively. Thus

$$\omega^{\#} = \{0, 1, 2, \ldots, \omega\}$$
$$\omega^{\#\#} = \{0, 1, 2, \ldots, \omega, \omega^{\#}\}$$

and so on; and the next limit is

$$\omega^{\dagger} = \{0, 1, 2, \ldots, \omega, \omega^{\#}, \omega^{\#\#}, \ldots\}.$$

Addition and multiplication for order types can be defined as follows. Given linear orders *R*, *S* on sets *A*, *B*, one defines orders $R \oplus S$, $R \otimes S$ on $A \oplus B$, $A \otimes B$ by

$$\begin{aligned} (a,b)(R \oplus S)(c,d) \quad & \text{if } (a = c = 0 \text{ and } bRd) \\ & \text{or } (a = 0 \text{ and } c = 1) \\ & \text{or } (a = c = 1 \text{ and } bSd) \\ (a,b)(R \otimes S)(c,d) \quad & \text{if } bSd \\ & \text{and } (\text{if } b = d \text{ then } aRc). \end{aligned}$$

Intuitively, $R \oplus S$ is like a copy of *A* in the *R* order, followed by a copy of *B* in the *S* order, while $R \otimes S$ is like a copy of *B* but with each element replaced by an entire copy of *A*. Then $|R| + |S| = |R \oplus S|$ and $|R| \cdot |S| = |R \otimes S|$.

For ordinals the methods of proof by transfinite induction and definition by transfinite recursion, analogous to the ordinary methods of induction and recursion in higher arithmetic or number theory, are available. To prove by transfinite induction that $\Phi(\rho)$ holds for all ordinals ρ, one must prove, as in finite induction, that $\Phi(0)$ holds and if $\Phi(\rho)$ holds then $\Phi(\rho')$ holds, and additionally that if τ is a limit and $\Phi(\mathbf{\Sigma})$ holds for all $\mathbf{\Sigma} < \tau$, then $\Phi(\tau)$ holds.

Similarly, to define by recursion $\psi(\rho)$ for all ordinals ρ, one must define (1) $\psi(0)$, (2) $\psi(\rho')$ in terms of $\psi(\rho)$, and (3) $\psi(\tau)$ in terms of $\psi(\mathbf{\Sigma})$ for $\mathbf{\Sigma} < \tau$ at limits τ. Equivalent, more convenient, definitions of addition and multiplication, plus a definition of ordinal exponentiation, can be given for ordinals by transfinite recursion:

$$\begin{aligned} \rho + 0 &= \rho & \rho + (\mathbf{\Sigma}^{\#}) &= (\rho + \mathbf{\Sigma})^{\#} \\ \rho \cdot 0 &= 0 & \rho \cdot (\mathbf{\Sigma}^{\#}) &= (\rho \cdot \mathbf{\Sigma}) + \rho \\ \rho^0 &= 1 & \rho^{(\mathbf{\Sigma}^{\#})} &= (\rho^{\mathbf{\Sigma}}) \cdot \rho \end{aligned}$$

and at the limits:

$$\begin{aligned} \rho + \tau &= \sup\{\rho + \mathbf{\Sigma} | \mathbf{\Sigma} < \tau\} \\ \rho \cdot \tau &= \sup\{\rho \cdot \mathbf{\Sigma} | \mathbf{\Sigma} < \tau\} \\ \rho^{\tau} &= \sup\{\rho^{\mathbf{\Sigma}} | \mathbf{\Sigma} < \tau\}. \end{aligned}$$

If $\rho, \mathbf{\Sigma}$ both have cardinality α, then so do $\rho + \mathbf{\Sigma}$, $\rho \cdot \mathbf{\Sigma}$ and even (in contrast to cardinal exponentiation) $\rho^{\mathbf{\Sigma}}$. Note $\rho^{\#} = \rho + 1$. Also $\omega^{\#}$, $\omega^{\#\#}$ and $\omega^{\dagger}$ are $\omega + 1$, $\omega + 2$, $\omega + \omega$.

3 Cardinal numbers and cardinal arithmetic

The cardinal number ('size') of a set is written $||A||$. The order '$\leq$' on the cardinals is defined as follows. $||A|| = ||B||$ if there is a bijection from *A* to *B*, and $||A|| \leq ||B||$ if there is a bijection from *A* to a subset of *B*. This order also has the properties of a well-order, by its connection with ordinal order. Hence there is a least or

zero cardinal. Cantor proved that for any cardinal α there is a greater cardinal and hence a least greater cardinal, its successor, α^+ (an explicit description is given below). For every set *C* of cardinals there is a cardinal greater than or equal to each of its elements, hence a least such cardinal, the supremum of *C*, $\sup C$. If $C = \{\beta | \beta = ||B||$ for some $B \in X\}$, then $\sup C = ||\cup X||$. (Note this implies there can be no set of all cardinals.) A limit (cardinal) is one that is neither zero nor a successor.

As with ordinals, the cardinals 0, $1 = 0^{\#}$, $2 = 1^{\#}, \ldots$ less than the least limit are called finite, and the least limit $\sup\{0, 1, 2, \ldots\}$, the cardinal of the natural numbers, is called $\aleph_0$ ('aleph-nought' or 'aleph-null'). Sets of cardinality $\aleph_0$ are called 'countable'. $\aleph_0^+$ is called $\aleph_1$, and the least ordinal of cardinality $\aleph_1$ is called ω_1 or Ω. $\aleph_1^+$, $\aleph_1^{++}, \ldots$ are called $\aleph_2$, $\aleph_3, \ldots$, and the next limit cardinal $\sup\{\aleph_0, \aleph_1, \aleph_2, \ldots\}$ is called $\aleph_\omega$, while the least ordinals of these cardinals are called ω_2, $\omega_3, \ldots, \omega_\omega$. There is only one ordinal of each finite cardinality (since there is an isomorphism between any two orders on a finite set). Not so for infinite cardinals, since the ordinals $\omega^{\#}$, $\omega^{\#\#}, \ldots, \omega^{\dagger}$ can also be described as the ordinals of certain unnatural well-orders on the natural numbers, namely

$1, 2, 3, \ldots, 0$

$2, 3, 4, \ldots, 0, 1$

$1, 3, 5, \ldots, 2, 4, 6, \ldots, \ldots$

and hence all have cardinality $\aleph_0$. Indeed, Cantor gave an explicit description of α^+ by proving that it is the cardinal of the set of ordinals of cardinality α. It is now customary to identify a cardinal with the least ordinal of that cardinal, so that $\aleph_0 = \omega_0$, and so on.

Cardinal addition, multiplication and exponentiation are defined as follows: $||A|| + ||B|| = ||A \oplus B||$, $||A|| \cdot ||B|| = ||A \otimes B||$ and $||A||$. These agree with the usual arithmetic notions on natural numbers, and some of the usual arithmetic laws generalize from the finite to the infinite (the associative, commutative and distributive laws of addition and multiplication, and the laws of exponents). But there are some differences. Bijections from $\mathbf{N}$ to $\mathbf{N} \oplus \mathbf{N}$ and to $\mathbf{N} \otimes \mathbf{N}$ are suggested by the following.

$(0,0)$	$(0,0)$
$(1,0)$	$(0,1)$
$(0,1)$	$(1,1)$
$(1,1)$	$(1,0)$
$(0,2)$	$(0,2)$
$(1,2)$	$(1,2)$
$(0,3)$	$(2,2)$
$(1,3)$	$(2,1)$
$(0,4)$	$(2,0)$
$(1,4)$	$(0,3)$
$\vdots$	$\vdots$

This shows that $\aleph_0 + \aleph_0 = \aleph_0 \cdot \aleph_0 = \aleph_0$. This result can be generalized in two directions to show (1) that a union of countably many countable sets is countable and (2) that

$$(\aleph_0)^n = \aleph_0 \cdot \ldots \cdot \aleph_0 = ||\mathbf{N} \otimes \ldots \otimes \mathbf{N}|| = \aleph_0$$

for all $n \in \mathbf{N}$, and hence (3) writing $\mathbf{N}^{<\omega}$ for the set of all finite ordered sequences (pairs, triples,…) of natural numbers and $\aleph_0^{<\omega}$ for $\sup\{(\aleph_0)^n | n = 0, 1, 2, \ldots\}$, one has

$$\aleph_0^{<\omega} = ||\mathbf{N}^{<\omega}|| = \aleph_0.$$

These 'abstract' results have 'concrete' counterparts for the sets $\mathbf{Z}$, $\mathbb{Q}$ and $\mathbf{A}$ of integers, rational and algebraic numbers (these last being the real numbers that are roots of polynomial equations with rational coefficients); namely,

$$||\mathbf{Z}|| = ||\mathbb{Q}|| = ||\mathbf{A}|| = \aleph_0.$$

Finally, there is the generalization that $\alpha + \alpha = \alpha \cdot \alpha = \alpha^{<\omega} = \alpha$ for all infinite α, and $\alpha + \beta = \alpha \cdot \beta = \sup\{\alpha, \beta\}$ for all infinite α, β, so cardinal addition and multiplication trivialize.

Exponentiation is not trivial. A bijection between subsets of *A* and functions from *A* to $\{0, 1\}$ is obtainable by relating a subset *B* with its 'characteristic function', χ_B.

$$\chi_B(a) = \begin{cases} 1 & \text{if } a \in B \\ 0 & \text{if } a \notin B \end{cases}$$

Thus $||\wp(A)|| = 2^{||A||}$. Cantor proved that $||A|| < 2^{||A||}$. (Indeed, no function $F : A \rightarrow \wp(A)$ can be a bijection, since $F(a) \neq \{x \in A | x \notin F(x)\}$ for any *a*.)

The cardinal c (for 'continuum') is defined to be $||\mathbb{R}||$, where $\mathbb{R}$ is the set of real numbers (or points in the line). By the general result $||A|| \cdot ||A|| = ||A||$, c is equal to $||\mathbb{R} \otimes \mathbb{R}||$ or $||\mathbb{C}||$, where $\mathbb{C}$ or $\mathbb{R} \otimes \mathbb{R}$ is the set of complex numbers (or points in the plane). It can also be shown to be equal to $||\wp(\mathbb{N})||$ or $2^{\aleph_0}$, and by the general result $||A|| < 2$ one has $\aleph_0 < \mathbf{c}$. A 'concrete' consequence of this 'abstract' result is that there exist transcendental (that is, non-algebraic real) numbers.

The 'generalized continuum hypothesis' is the conjecture that $2^\alpha = \alpha^+$ for all cardinals α. The 'continuum hypothesis' (CH) is the special case for $\alpha = \aleph_0$. Equivalently, CH asserts there is no intermediate cardinal α with $\aleph_0 < \alpha < \mathbf{c}$. The most interesting alternatives to CH include the hypotheses CH': that there is just one such intermediate; and $\mathrm{CH}^{\#}$: that there are c intermediates. Where CH asserts that $\mathbf{c} = \aleph_1$ and CH' asserts that $\mathbf{c} = \aleph_2$, $\mathrm{CH}^{\#}$ implies that $\mathbf{c} > \aleph_1, \aleph_2, \ldots, \aleph_\omega$, and more. Assuming the generalized continuum hypothesis, exponentiation simplifies.

$$\alpha^\beta = \begin{cases} \alpha & \text{if } \beta < \mathrm{cf}\alpha \\ \alpha^+ & \text{if } \mathrm{cf}\alpha \leq \beta < \alpha \\ \beta^+ & \text{if } \mathrm{cf}\alpha < \beta \end{cases}$$

Otherwise, one must distinguish between a cardinal α being a limit, meaning that $\beta^+ < \alpha$ whenever $\beta < \alpha$, and α being a 'strong limit', meaning that $2^\beta < \alpha$ whenever $\beta < \alpha$.

The 'cofinality' $\mathrm{cf}\alpha$ of a cardinal α is the least β such that there is a set *B* of sets with $\beta = ||B||$ and $||b|| < \alpha$ for all sets $b \in B$, but $|| \cup B|| = \alpha$. α is 'regular' or 'singular' according as $\mathrm{cf}\alpha$ is equal to or less than α. $\aleph_0$ is regular (otherwise $\mathrm{cf}\aleph_0$ would be finite, but a union of finitely many finite sets is finite), $\aleph_1$ is regular (otherwise $\mathrm{cf}\aleph_1$ would be $\aleph_0$, but a union of countably many countable sets is countable), and generally α^+ is regular. $\aleph_\omega$ is singular, with $\mathrm{cf}\aleph_\omega = \aleph_0$ (a union of countably many sets with cardinality $\aleph_0$, $\aleph_1$, $\aleph_2$ and so on has cardinality $\aleph_\omega$). It can at least be proved that $\mathbf{c} \neq \aleph_\omega$, and more generally that $\mathrm{cf}2^\alpha > \mathrm{cf}\alpha$.

4 Combinatorial set theory

Analogous to finite combinatorics based on the ordinary arithmetic of the natural numbers there is an extensive 'combinatorial' set theory or infinitary combinatorics based on the transfinite arithmetic of cardinals and ordinals. A typical cardinal result is the 'delta system' theorem: if *A* is an uncountable set of finite sets, then there is a fixed finite set *b* and an uncountable $B \subseteq A$ such that $b' \cap b'' = b$ for all $b', b'' \in B$.

A typical ordinal result is the 'pressing down' theorem. Let *A* be a set of ordinals. *A* is called 'unbounded' in τ if $\sup(A \cap \tau) = \tau$, or for every $\rho < \tau$ there exists $\mathbf{\Sigma} \in A$ with $\rho < \mathbf{\Sigma} < \tau$. Given an uncountable regular cardinal α, *A* is called 'closed' in α if $\tau \in A$ whenever $\tau < \alpha$ and *A* is unbounded in τ. $B \subseteq \alpha$ is called 'stationary' in α if for all sets *A* that are closed and unbounded in α, $B \cap A \neq \emptyset$. The theorem asserts that if *F* is a function from α to α and $\{\rho < \alpha | F(\rho) < \rho\}$ is stationary, then there is a $\mathbf{\Sigma} < \alpha$ such that $\{\rho < \alpha | F(\rho) = \mathbf{\Sigma}\}$ is stationary.

Also included in combinatorial set theory are results about order types other than ordinals, beginning with the theory of the order types η, θ of the usual orders on $\mathbb{Q}$ and $\mathbb{R}$. Given a linear order '$\leq$' on a set *A*, an 'open' interval is a set of the form $]a, b[= \{x | a < x < b\}$ for $a < b$ and a 'closed' interval is a set $[a, b] = \{x | a \leq x \leq b\}$, $a \leq b$. $B \subseteq A$ is 'dense' if it is disjoint from no open interval and 'bounded' if it is a subset of some closed interval. $\leq$ is 'separable' if there exists a countable dense subset *B*, and satisfies the

'countable chain condition' (or 'is CCC') if there exists no uncountable set of disjoint open intervals. (Separability implies CCC since each of a set of disjoint open intervals must contain a distinct element of the countable dense set.) '$\leq$' is 'complete' if every chain of closed intervals has a non-empty intersection. $\mathbb{Q}$, $\mathbb{R}$ are each unbounded and dense in themselves, and $\mathbb{Q}$ is dense in $\mathbb{R}$ (and countable), so $\mathbb{R}$ is separable and CCC. $\mathbb{R}$ is complete, but $\mathbb{Q}$ is not (for example, the intervals

$$\ldots [1.414, 1.415] \subseteq [1.41, 1.42] \subseteq [1.4, 1.5] \subseteq [1, 2]$$

have intersection $\{\sqrt{2}\}$). Cantor proved that any unbounded, dense linear order on a countable set has order type η. Similarly, any unbounded, dense, separable, complete linear order has order type θ. The 'Suslin hypothesis' is that the latter result holds with CCC in place of separability.

5 Point sets

The oldest part of set theory and the part most closely connected with the rest of mathematics (analysis, topology) is the theory of sets of points in the line, plane and so on (identified with $\mathbb{R}, \mathbb{R} \otimes \mathbb{R}, \ldots$). Modern analysis and topology consider more complicated functions and sets than traditional calculus and geometry. A function $F : \mathbb{R} \to \mathbb{R}$ is 'continuous' if $F^{-1}[U]$ is open for every open $U \subseteq \mathbb{R}$, where an 'open' set is a union of open intervals and a 'closed' set is the complement of an open set. The Cantor 'discontinuum', the complement in the unit interval $[0, 1]$ of the union of the open intervals $]\frac{1}{3}, \frac{2}{3}[, \ ]\frac{1}{9}, \frac{2}{9}[, \ ]\frac{7}{9}, \frac{8}{9}[, \ ]\frac{1}{27}, \frac{2}{27}[, \ ]\frac{7}{27}, \frac{8}{27}[, \ ]\frac{19}{27}, \frac{20}{27}[$ and so on, is a famous example of a closed set (see Continuum hypothesis §2) more complicated than any considered in traditional mathematics. Sets obtainable from open sets by iterated application of complementation and countable union are called 'Borel' sets. They occur ubiquitously in modern mathematics (as in the rigorous formulation of probability theory), while sets beyond Borel occur sporadically.

An 'analytic' set is the image of a Borel set under a continuous function, a 'co-analytic' set is the complement of an analytic set, a 'pro-co-analytic' (PCA) set is the image of any set under a continuous function, and a 'co-pro-co-analytic' (CPCA) set is the complement of a pro-co-analytic set. Sets obtainable from Borel sets by iterated application of continuous image and complementation are called 'projective'. Projective sets are rather special compared with arbitrary sets. It can be shown that there are only c projective sets, whereas there are $2^{\mathfrak{c}}$ subsets of $\mathbb{R}$. The study of such special classifications of sets is called 'descriptive' set theory. In one notation, open sets are denoted $\mathbf{\Sigma}^0_1$, closed sets, $\mathbf{\Pi}^0_1$, unions of countably many closed sets, $\mathbf{\Sigma}^0_2$ and their complements, $\mathbf{\Pi}^0_2, \ldots$. Analytic, co-analytic, PCA and CPCA sets are denoted, respectively, $\mathbf{\Sigma}^1_1$, $\mathbf{\Pi}^1_1$, $\mathbf{\Sigma}^1_2$ and $\mathbf{\Pi}^1_2, \ldots$ sets.

Central to attempts to define integration for complicated functions, and length, area and so on for complicated linear, planar,… sets, is the notion of a (complete probability) 'measure' on a set or 'space' *X*. Such a measure is a function μ from the set of '(μ-)measurable' subsets of *X* to the unit interval $\mathrm{J} = [0, 1]$, $\mu(A)$ being called the (μ-)measure of $A \subseteq X$. μ has the following properties: (1) 'normality' - the whole space has measure one; (2) 'non-triviality' - any one-point set has measure zero; (3) 'completeness' - any subset of a set of measure zero has measure zero; (4) 'complementarity' - the complement of a measurable set is measurable; and (5) 'additivity' - the union of a countable set of measurable sets is measurable and, if the sets are disjoint, the measure of their union is the sum of their measures. The analogue of additivity with 'countable set' replaced by 'set of cardinality less than that of the whole space' may be called ($5'$) 'superadditivity'. If we restrict ourselves to the unit interval, the 'translation' of $A \subseteq \mathrm{J}$ by $t \in \mathrm{J}$ is defined to be $\{a \# t | a \in A\}$, where $x \# y$ is defined to be whichever of $x + y$, $x + y - 1$ is an element of J, and one may consider the further property (6) 'invariance' - a translation of a measurable set is measurable, and its measure is the same. The key positive result is the existence of a unique measure on J with properties (1)-(6), called 'Lebesgue measure'. (Analogues can also be defined on $\mathbb{R}, \mathbb{R} \otimes \mathbb{R}$ and so on.) For Lebesgue measure, $\mu(A)$ intuitively represents the probability that a point randomly chosen from J will be an element of *A*. One respect in which the discontinuum is 'pathological' is that it is large in the sense of having cardinality c but small in the sense of having Lebesgue measure 0. (Like some other positive results, the existence and uniqueness of Lebesgue measure require for their proof only DC (see §1 above), not full AC.)

To mention a negative result, the hypothesis that all sets are Lebesgue measurable (AM) fails. Let *S* be a choice set for $\{y \in \mathrm{J} | (y - x) \in \mathbb{Q}\}$. If *S* were measurable, then using invariance it would follow that the disjoint sets of form $\{y \in \mathrm{J} | y \# p \in S\}$ for $p \in \mathbb{Q}$, whose union is all of J, would all have the same measure ε. But then using additivity, μ(J) would have to be $\varepsilon + \varepsilon + \varepsilon + \ldots$, which is 0 or ∞ according as $\varepsilon = 0$ or > 0. The failure of

AM leads to consideration of some alternative hypotheses, notably the hypothesis (known to imply $\sim\mathrm{CH}$ and $\sim\mathrm{CH}'$) of the existence of a measure with properties (1)-(5) for which all sets are measurable, or the hypothesis (known to imply $\mathrm{CH}^{\#}$) of the existence of a measure with properties (1)-(4) and superadditivity for which all sets are measurable. A more positive result using AC may also be mentioned. Given $Q \subseteq \mathbb{R}$ and a continuous *F* from $\mathbb{R}$ to $\mathbb{R}$ a uniformization for *Q*, *F* is a choice set *P* for $\{\{y \in Q|F(y) = F(x)\}|x \in Q\}$. AC immediately implies the existence of a uniformization for any *Q*, *F*.

The failure of AM also leads to consideration of some restricted hypotheses: for Γ one of the special classifications considered above, 'Γ-measurability' is the hypothesis that all sets in Γ are measurable. Similarly, 'Γ-uniformization' is the hypothesis that for any $Q \in \Gamma$ and any continuous *F* there is a uniformization $P \in \Gamma$. It can be shown that while $\mathbf{\Sigma}^1_p$-measurability and $\mathbf{\Pi}^1_p$-measurability imply each other for the same *p*, $\mathbf{\Sigma}^1_p$-uniformization and $\mathbf{\Pi}^1_p$-uniformization cannot both hold for the same *p*. Positive results of descriptive set theory obtained by the end of the 1930s include $\mathbf{\Sigma}^1_1$- and $\mathbf{\Pi}^1_1$-measurability, an example of a 'regularity' theorem or result of the form 'Every set in Γ has some ''nice'' property', and also $\mathbf{\Pi}^1_1$- and $\mathbf{\Sigma}^1_1$-uniformization, an example of a 'structure' theorem or result of the form 'For every set in Γ there is another set in Γ related to it in some ''nice'' way'. All attempts to prove $\mathbf{\Sigma}^1_2$- and $\mathbf{\Pi}^1_2$-measurability, and $\mathbf{\Sigma}^1_3$- or $\mathbf{\Pi}^1_3$-uniformization have failed.

6 New axioms for set theory: large cardinals and determinacy

Beginning in about 1940, many conjectures Δ (including those mentioned at the end of §5) have been proved by mathematical logicians to be 'undecidable' relative to ZFC, meaning that it has been shown that if ZFC is consistent, then so are both $\mathrm{ZFC} + \Delta$ (relative consistency of Δ) and $\mathrm{ZFC}+\sim\Delta$ (relative independence of Δ). Note that the consistency of ZFC, let alone of $\mathrm{ZFC} + \Delta$ (or $\mathrm{ZFC}+\sim\Delta$), cannot be proved within ZFC itself, by celebrated results of mathematical logic (see Gödel's theorems), and hence cannot be proved by ordinary mathematical methods, since these can be developed within ZFC. The most one can hope to prove is the *relative* result that if a contradiction could be deduced from $\mathrm{ZFC} + \Delta$ (or $\mathrm{ZFC}+\sim\Delta$), then a contradiction could be deduced from ZFC. There are two basic methods for proving relative consistency. A first method was used by Kurt Gödel to prove the consistency of CH relative to ZFC (and of AC relative to ZF) (see Constructible universe). A second method was used by Paul Cohen to prove the independence of CH relative to ZFC (and of AC relative to ZF) (see Forcing). Both methods have been widely applied since.

For some hypotheses, however, there can be no hope of a proof of relative consistency, namely for any hypothesis that itself implies the consistency of ZFC. An example is the hypothesis of the existence of an 'inaccessible' or regular strong limit cardinal α (IC). (How IC implies the existence of a model of ZFC, and hence the consistency of ZFC, will be indicated in §7.) Even in such cases there may be heuristic or inductive evidence of consistency (failure to deduce a consequence in years of work with the hypothesis). In order to decide questions undecidable in ZFC, new axioms have from time to time been advocated. The two kinds most often advocated (which unfortunately still leave CH undecided) have been 'large cardinal' and 'determinacy' axioms. The former, which pertain to sets larger than any encountered in ordinary mathematical practice, are often advocated on the intrinsic ground that they are further expressions of intuitions only partly expressed by the ZFC axioms (to the effect that the hierarchy of sets is 'tall'). The latter, which pertain to special classifications of sets of linear, plane,… points, and are connected with problems in analysis and topology, are often advocated on the extrinsic ground that their consequences are 'nice'.

Large cardinal axioms include IC and also MC, the hypothesis that there exists a measurable cardinal β. Here β is measurable if there is a superadditive measure (as defined in §5), taking only the two values 0 and 1, for which all subsets of β are measurable. It follows from non-triviality and superadditivity that any subset of cardinality $< \beta$ will have measure 0, and by superadditivity again that any union of a set of cardinality $< \beta$ of such subsets will still have measure 0, whence β is regular. It also follows that if $\alpha < \beta$ and *F* is a function from β to $\wp(\alpha)$, then for any $\rho < \alpha$ one of the sets

$$D^+(\rho) = \{\mathbf{\Sigma} < \beta|\rho \in F(\mathbf{\Sigma})\}$$
$$D^-(\rho) = \{\mathbf{\Sigma} < \beta|\rho \notin F(\mathbf{\Sigma})\}$$

will have measure 0, the other, measure 1, and by superadditivity the union *D* of the ones with measure 0 will have

measure 0. But for $\mathbf{\Sigma}, \tau \notin D$ one has

$F(\mathbf{\Sigma}) = F(\tau) = \{\rho < \alpha | \mu(D^+(\rho)) = 1\}$

so *F* is not a bijection, $2^\alpha = ||\wp(\alpha)|| < \beta$, and β is a strong limit. So a measurable cardinal β is inaccessible, and indeed it can even be shown that $\{\alpha < \beta | \alpha$ is an inaccessible $\mathrm{cardinal}\}$ has cardinal β, so MC is a much stronger hypothesis than IC. Still stronger is the hypothesis (SC) that there exists a 'supercompact' cardinal γ. For such a γ there is for any $\delta > \gamma$ an analogue of a measure as above for the subsets of δ of cardinality γ.

Determinacy axioms pertain to certain infinite games considered in mathematical game theory. Given $A \subseteq$ J there is a game for two players I, II: they alternately pick the terms in the decimal expansion of a real *x*, and I or II wins according as *x* is or is not an element of *A*. A 'strategy' for a player is, intuitively, a rule telling the player how to play as a function of the opponent's previous moves, and is winning if by so playing the player will win the game, no matter what the opponent's moves. *A* is 'determinate' if one or other player has a winning strategy. $\mathbf{\Pi}^1_1$-AD, PD and AD (to add to the notation set up in §5) are, respectively, the hypotheses that all $\mathbf{\Pi}^1_1$, projective or arbitrary sets are determinate. AD is refutable using AC, among other reasons because AD implies AM, while, by essentially the same proof, PD implies projective measurability, and various other regularity theorems. Similarly, PD implies projective uniformization and various other structural theorems. $\mathbf{\Pi}^1_1$-AD yields, correspondingly, $\mathbf{\Sigma}^1_2$- and $\mathbf{\Pi}^1_2$-measurability and $\mathbf{\Pi}^1_3$-uniformization.

Despite superficial differences, there are deep connections between large cardinals and determinacy. By the 1970s it was established that a large cardinal axiom somewhat weaker than MC implies a conclusion somewhat stronger than $\mathbf{\Pi}^1_1$-AD, with the partial converse that such a determinacy hypothesis implies the consistency of such a large cardinal hypothesis. In the 1980s an analogous relation between SC and PD was established. So the cases for large cardinal and for determinacy axioms reinforce each other. The search for further axioms to decide CH continues.

7 Models of set theory

In Gödel's method for proving relative consistency of a conjecture Δ, one seeks a formula $\Xi(x)$ and proves that for any axiom of $\mathrm{ZFC} + \Delta$ and hence for any finite conjunction Ψ of such axioms, Ψ^Ξ is provable in ZFC, where Ψ^Ξ is the result of replacing each quantifier 'for all *x*' or 'for some *x*' in Ψ by the 'restricted' quantifier 'for all *x* such that $\Xi(x)$' or 'for some *x* such that $\Xi(x)$'.

Supposing this had been done, if a contradiction could be deduced from $\mathrm{ZFC} + \Delta$, the deduction would use only finitely many of the axioms of $\mathrm{ZFC} + \Delta$, and so a contradiction $\Theta \text{ and } {\sim}\Theta$ could be deduced from some finite conjunction Ψ of axioms of $\mathrm{ZFC} + \Delta$ and, restricting quantifiers, a contradiction $\Theta^\Xi \text{ and } {\sim}\Theta^\Xi$ could be deduced from Ψ^Ξ. By exhibiting the deduction of a contradiction from Ψ^Ξ, ${\sim}\Psi^\Xi$ would be provable in ZFC, and hence a contradiction (both Ψ^Ξ and ${\sim}\Psi^\Xi$) would be deducible in ZFC. Such a Ξ (and derivatively, the 'universe' of all *x* such that $\Xi(x)$ holds) is said to constitute an 'inner model'.

In Cohen's method for proving relative consistency, one assumes the existence of a set *X* that is a model of ZFC in the sense that for any axiom of ZFC, and hence for any finite conjunction Φ of such axioms, Φ^X holds, where Φ^X is the result of replacing each quantifier in Φ by the restricted quantifier (as above). One then proves the existence of a set *Y* that is a model of $\mathrm{ZFC} + \Delta$, but in such a way that for any finite conjunction Ψ of axioms of $\mathrm{ZFC} + \Delta$, the proof that Ψ^Y holds depends only on the assumption that Φ^X holds for some finite conjunction $\Phi = \Phi_\Psi$ of axioms of ZFC. Supposing this done, one uses the fact that while the existence of a model *X* of all of ZFC cannot be proved in ZFC (since this would imply the consistency of ZFC), for any finite conjunction Φ of axioms of ZFC the existence of a model of Φ can be proved in ZFC (as will be indicated below). It follows that the existence of a model for any finite subset Ψ of $\mathrm{ZFC} + \Delta$ is provable in ZFC. Then if a contradiction could be deduced from $\mathrm{ZFC} + \Delta$, a contradiction could be deduced from some finite conjunction Ψ of axioms of $\mathrm{ZFC} + \Delta$, and by exhibiting this deduction the non-existence of any model of Ψ could be proved in ZFC, and hence a contradiction (existence and non-existence of a model of Ψ) could be proved in ZFC.

The general theories of set models *X* (as in the second method) and formula models Ξ (as in the first method) are closely parallel, and only the theory of set models will be sketched here. For applications it suffices to consider only sets *X* that are 'transitive' in the sense that whenever $u \in v \in X$, then $u \in X$. Informally, if Φ^X holds, it will be said that it 'appears' to *X* that Φ holds, or that Φ holds 'inside' *X*. Central to the theory of models is the

determination of sufficient conditions for various axioms of ZFC to hold inside *X*. Central to this determination are considerations of absoluteness, where a formula Φ is called 'absolute' for *X* if for any $u, v, \ldots \in X$, $\Phi(u, v, \ldots)$ appears to *X* to hold if and only if it really holds. This will always be so if Φ is a 'limited' formula, mentioning only $u, v,$ and their elements, elements of elements and so on, since by transitivity these will all be elements of *X*.

For instance, the condition 'u, v have different elements' or 'For some *w*, either $w \in u$ and $w \notin v$ or $w \in v$ and $w \notin u$' is limited, and so $u, v \in X$ appear to *X* to have different elements if and only if they really do have different elements, and so the axiom of extensionality, 'For every u, v, if $u \neq v$ then u, v have different elements', holds inside *X* for any transitive *X*. Similarly, it can be shown that the axiom of foundation holds inside *X* for any transitive *X*. As for pairing, the condition '$w = \{u, v\}$' or '$u \in w$ and $v \in w$ and for all $z \in w$, $z = u$ or $z = v$' is limited, and so *w* appears to *X* to be the pair $\{u, v\}$ of u, v if and only if it really is that pair, and so the axiom of *pairs*, 'For every u, v there exists a *w* such that $w = \{u, v\}$', holds inside *X* if and only if $\{u, v\} \in X$ whenever $u, v \in X$, or, as is said, if and only if *X* is closed under the pairing operation $\{\ ,\ \}$. Similarly it can be shown that the axiom of *union* holds in *X* if and only if *X* is closed under the union operation '$\cup$'; that the axiom of infinity holds in *X* if and only if $\omega \in X$; and that the axiom of choice holds in *X* if and only if every partition in *X* has a choice set in *X*.

As for power, *u* appears to *X* to be a subset of *v* if and only if it really is, but *w* appears to *X* to be the set $\wp(v)$ of all subsets *u* of *v* if and only if *w* really is the set $\wp(v) \cap X$ of all subsets *u* of *v* with $u \in X$. Thus the axiom of *power* appears to *X* to hold if and only if *X* is closed under the operation of forming $\wp(v) \cap X$ from *v*, which may be the case even if *X* does not contain all subsets of *v* and therefore is not closed under the operation of forming $\wp(v)$ from *v*. Similarly, a necessary and sufficient condition for separation to hold inside *X* is that for each condition Φ, *X* is closed under the operation that applies to any *v* forms $\{u \in v | \Phi^X(u)\}$. Similarly, a necessary and sufficient condition for replacement to hold inside *X* is that for each condition Φ such that for every $x \in X$ there is a unique $y = \psi^X(x) \in X$ such that $\Phi^X(x, y)$ holds, *X* is closed under the operation that applied to *v* forms $\{\psi^X(u) | u \in v\}$. *X* is 'supertransitive' if whenever $u \subseteq v \in X$, then $u \in X$. For supertransitive *X*, $\wp(v) \cap X = \wp(v)$ for any $v \in X$, so the necessary and sufficient condition for power to hold is just that *X* is closed under the power operation $\wp$. Supertransitivity is a more than sufficient condition for separation to hold. Similarly, a more than sufficient condition for replacement to hold is that whenever $u \subseteq X$, $||u|| = ||v||$ and $v \in X$, then $u \in X$.

One kind of model of a large part of ZFC is obtained as follows. For any *x* let the 'transitivization' $t(x)$ of *x* be $\cup\{x, \cup x, \cup\cup x, \ldots\}$, the set of all elements, elements of elements and so on of *x*. For any uncountable cardinal α, let $H(\alpha)$ be $\{x | \alpha > ||t(x)||\}$. It follows from the definition that $H(\alpha)$ is supertransitive, and it can be checked, using the necessary and sufficient conditions indicated above, that $H(\alpha)$ is a model of all axioms of ZFC except perhaps power and replacement. Moreover, if α is a strong limit, then $H(\alpha)$ will be closed under $\wp$ and hence a model of power; while if α is regular, then one will have $u \in H(\alpha)$ whenever $u \subseteq H(\alpha)$, $||u|| = ||v||$, $v \in H(\alpha)$ and $H(\alpha)$ will be a model of replacement. (Thus if α is inaccessible, $H(\alpha)$ is a model of the whole of ZFC.)

Another kind of model of a large part of ZFC is obtained as follows. The hierarchy of levels of sets, intuitions about which underlie the axioms (as in §1), can be defined by transfinite recursion.

$$
\begin{aligned}
V(0) &= \emptyset \\
V(\rho+1) &= \wp(V(\rho)) \\
\text{at limits } V(\tau) &= \cup\{V(\boldsymbol{\Sigma}) | \boldsymbol{\Sigma} < \tau\}
\end{aligned}
$$

That $V(\rho)$ is supertransitive can be proved by transfinite induction. Given the other axioms of ZFC, the principle of replacement can be proved equivalent to the following principle of 'reflection': for any formula Φ there are arbitrarily large ordinals ρ such that Φ is absolute for $V(\rho)$. In particular, for any finite conjunction Φ of axioms of ZFC there are arbitrarily large ordinals ρ such that Φ holds in $V(\rho)$.

While both the $V(\rho)$ (for $\rho > \omega$) and the $H(\alpha)$ (for $\alpha > \aleph_0$) are uncountable, it can be shown that if a ZFC or some variant set theory has a transitive model, then it has a countable transitive model, using a celebrated result of mathematical logic (see Löwenheim-Skolem theorems and nonstandard models). Since the uncountability of the

power set of the set of natural numbers is a theorem of set theory, the existence of a countable model for set theory is sometimes considered paradoxical, and called the ‘Skolem paradox’. There is, however, no genuine contradiction. (For the conditions ‘ $x = \wp(\omega)$’ and ‘ $||x|| > \aleph_0$’ are not absolute: the former holds inside a model *X* if $x = \wp(\omega) \cap X$, which may well be countable even though $\wp(\omega)$ is uncountable; the latter holds inside the model *X* if there is no surjection $F \in X$ from ω to *x*, which there may well not be even if there is such an $F \notin X$.)

See also: Cantor’s theorem; Logical and mathematical terms, glossary of

JOHN P. BURGESS

References and further reading

Barwise, J. (ed.) (1977) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 317-522.(A standard reference including surveys on a generous scale of the different branches of set theory, with attribution of results to their original authors and references to the original technical literature.)

Benacerraf, P. (1965) ‘What Numbers Could Not Be’, *Philosophical Review* 74: 47-73; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.(A philosophical puzzle about ‘identification’ as in §1.)

Cantor, G. (1895, 1897) ‘Beiträge zur Begründung der transfiniten Mengenlehre’, *Mathematische Annalen* 46: 481-512, 49: 207-46; trans. P.E.B. Jourdain (1915), *Contributions to the Founding of the Theory of Transfinite Numbers*, New York: Dover, 1955.(The original source for §§2-3, with an introduction setting the work in its historical context.)

Halmos, P. (1960) *Naive Set Theory*, New York: Van Nostrand.(A perennially popular elementary textbook, expounding for students of mathematics the development of mathematics on the basis of ZFC mentioned in §1.)

Jech, T. (1977) ‘About the Axiom of Choice’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 345-70.(Surveys all aspects of AC mentioned here, and many more besides.)

Kunen, K. (1977) ‘Combinatorics’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 370-402.(Surveys at length combinatorial set theory as mentioned in §4; and, briefly, large cardinals as in §6.)

Maddy, P. (1988) ‘Believing the Axioms’, *Journal of Symbolic Logic* 53: 481-511, 736-64.(Surveys axioms old and new as in §§1, 6; especially advances made in the 1980s.)

Martin, D.A. (1977) ‘Descriptive Set Theory’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 783-818.(Surveys theory of projective sets (as in §§5-6) through the 1970s.)

Shoenfield, J.R. (1977) ‘Axioms of Set Theory’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 321-44.(Surveys material mentioned in §§1-4, 6.)

Zermelo, E. (1908) ‘Untersuchungen über die Grundlagen der Mengenlehre, I’, *Mathematische Annalen* 65: 261-81; trans. S. Bauer-Mengelberg, ‘Investigations in the Foundations of Set Theory’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Sourcebook in Mathematical Logic*, Cambridge, MA: Harvard University Press, 1967, 200-15.(The original source for the axiomatization of set theory, with an introduction setting the work in its historical context.)

Set theory, different systems of

To begin with we shall use the word 'collection' quite broadly to mean anything the identity of which is solely a matter of what its members are (including 'sets' and 'classes'). Which collections exist? Two extreme positions are initially appealing. The first is to say that all do. Unfortunately this is inconsistent because of Russell's paradox: the collection of all collections which are not members of themselves does not exist. The second is to say that none do, but to talk as if they did whenever such talk can be shown to be eliminable and therefore harmless. This is consistent but far too weak to be of much use. We need an intermediate theory.

Various theories of collections have been proposed since the start of the twentieth century. What they share is the axiom of 'extensionality', which asserts that any two sets which have exactly the same elements must be identical. This is just a matter of definition: objects which do not satisfy extensionality are not collections. Beyond extensionality, theories differ. The most popular among mathematicians is Zermelo-Fraenkel set theory (ZF). A common alternative is von Neumann-Bernays-Gödel class theory (NBG), which allows for the same sets but also has proper classes, that is, collections whose members are sets but which are not themselves sets (such as the class of all sets or the class of all ordinals).

Two general principles have been used to motivate the axioms of ZF and its relatives. The first is the iterative conception, according to which sets occur cumulatively in layers, each containing all the members and subsets of all previous layers. The second is the doctrine of limitation of size, according to which the 'paradoxical sets' (that is, the proper classes of NBG) fail to be sets because they are in some sense too big. Neither principle is altogether satisfactory as a justification for the whole of ZF: for example, the replacement schema is motivated only by limitation of size; and 'foundation' is motivated only by the iterative conception.

Among the other systems of set theory to have been proposed, the one that has received widespread attention is Quine's NF (from the title of an article, 'New Foundations for Mathematical Logic'*), which seeks to avoid paradox by means of a syntactic restriction but which has not been provided with an intuitive justification on the basis of any conception of set. It is known that if NF is consistent then ZF is consistent, but the converse result has still not been proved.*

1 Virtual collections

Sometimes when we speak of collections it is just a manner of speaking and no more: if we say that Sarah belongs to the collection of all linguists, we might just as well say straight away that she is a linguist; nothing seems to be lost in the translation. This way of regarding collections is quite old: it is how Peano conceived of them in the 1880s. If we already have before us some formal theory which does not have collection-talk in it, then we can add it straightforwardly. If $\Phi(x)$ is a formula of the theory in question, we introduce a new term $\{x\colon \Phi(x)\}$ (called a 'collection term') into the language. Collection terms are subject to the following two rules:

(1) If $\Phi(x)$ is a formula then $x \in \{y\colon \Phi(y)\}$ is an abbreviation for $\Phi(x)$.
(2) If $\Phi(x)$ and $\Psi(x)$ are both formulas of the theory then $\{x\colon \Phi(x)\} = \{x\colon \Psi(x)\}$ is an abbreviation for $\forall x(\Phi(x) \leftrightarrow \Psi(x))$.

We permit the use of collection terms only in contexts from which these two rules allow them to be eliminated. Collections used harmlessly in this way have been called 'virtual' by Quine and we shall follow this terminology here. Let us introduce the convention that lower-case Greek letters stand schematically for collection terms. So the second re-writing rule allows us to assert, for instance, the following *extensionality schema*:

$\forall z(z \in \alpha \leftrightarrow z \in \beta) \rightarrow \alpha = \beta.$

By this schematic device we generate a considerable amount of the naïve theory of collections. We can, for instance, introduce the usual Boolean operations as follows:

$$\alpha \cap \beta = \{x\colon\ x \in \alpha \wedge x \in \beta\}$$
$$\alpha \cup \beta = \{x\colon\ x \in \alpha \vee x \in \beta\}$$
$$\emptyset = \{x\colon\ x \neq x\}$$
$$\mathbf{V} = \{x\colon\ x = x\}$$
$$\{x\} = \{y\colon\ y = x\}$$
$$\{x, y\} = \{x\} \cup \{y\}$$

and so on. The theory thus generated is elegant enough in itself, but it is difficult to be satisfied with it for long. (We have to remember, for instance, that in this theory expressions such as $\{x, y\}$ are merely notational devices and do not, despite the appearance to the contrary that their 'explicit' definitions might suggest, have any meaning on their own.) Collection terms stand in sentences in some of the places where the grammar leads us to expect proper names; schematic letters stand where we might expect variables. There is a drive to objectify at work here; to suppose that collection terms can stand in *all* the places where names can occur. It is of course this step which, if carelessly done, leads straight to contradiction. How to do it without contradiction will be the subject of the rest of this entry.

2 Real collections

In the virtual theory just discussed the relation of membership is defined contextually. In the real theory with collections regarded as objects in their own right membership is taken by almost all authors as primitive. (Lewis (1991) is a rare exception.) Formally, this means that the language in which the theory is couched contains, in addition to the usual logical constants, a binary predicate symbol '$\in$'. It will be convenient to introduce a few notational conventions straight away. Upper-case Greek letters such as Φ and Ψ will stand for formulas. The notation $\Phi(x)$ stands for a formula in which the variable x occurs free: it does not indicate that no other variable occurs free in the formula. If Φ is a formula, we write $\Phi^{(a)}$ for the 'relativization' of Φ to a, that is, the formula obtained from Φ by replacing each quantifier $\forall x$ or $\exists x$ by $(\forall x \in a)$ or $(\exists x \in a)$, as the case may be. We write $x \subseteq y$ as an abbreviation for $\forall z(z \in x \rightarrow z \in y)$. We introduce ordered pairs by means of the device

$$(x, y) = \{\{x\}, \{x, y\}\}.$$

In the virtual theory, the question of whether there are any objects which are not collections is scarcely an issue: since there are not really any collections, if there is nothing else then there is nothing at all; the theory therefore collapses into vacuity. In any non-trivial real theory things are different: there are plenty of collections to talk about, whether or not there are any non-collections. Since this issue arises uniformly for all the real theories we shall be discussing, it is best to address it now. The objects of a theory which are not collections are called 'atoms', 'individuals' or (even in some non-German texts) '*Urelemente*'. If we are to allow for their presence, we need a primitive unary predicate $\mathrm{atom}(x)$ to represent that x is an atom. We then introduce as an axiom the *axiom of extensionality* (*applied form*):

$$\forall x \forall y(\mathrm{atom}(x) \vee \mathrm{atom}(y) \vee \forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y).$$

This asserts no more than that objects which are not atoms are collections, since (as we have already observed) it is part of what we mean by a collection that it is completely determined by its members. If the atoms are to play any role in the theory at all, we shall also need the *collection axiom*, which states that the atoms form a collection:

$$\exists y \forall x(x \in y \leftrightarrow \mathrm{atom}(x)).$$

If, on the other hand, it is pure collections that interest us, then the beginnings of the theory are a little more concise. We do not need the predicate 'atom(x)' nor the collection axiom, and we can use the simpler *axiom of extensionality* (*pure form*):

$$\forall x \forall y(\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y).$$

Throughout what follows we shall state theories in their pure form and adopt the convention that if *T* is a pure theory (that is, one which includes the pure form of the axiom of extensionality) then *TU* is the corresponding applied theory (that is, the one obtained from *T* by deleting the pure form of extensionality and substituting the

applied form together with the collection axiom). In the virtual theory we had collection terms defined contextually, provided that they occurred in certain positions (on the right of the '$\in$' symbol or on either side of the '$=$' symbol). Now that we are regarding collections as real we cannot be so indiscriminate. Let us say that a formula $\Phi(x)$ is 'collectivizing' in x for a theory *T* if

$$T \vdash \exists y \forall x (x \in y \leftrightarrow \Phi(x)).$$

Suppose now that *T* includes extensionality among its axioms. If $\Phi(x)$ is collectivizing in x for *T* then, by extensionality, we have uniqueness:

$$T \vdash \exists ! y \forall x (x \in y \leftrightarrow \Phi(x)).$$

It is therefore legitimate to introduce a term $\{y\colon \Phi(y)\}$ with the property that

$$\forall x (x \in \{y\colon \Phi(y)\} \leftrightarrow \Phi(x)).$$

We thereby retrieve the virtual theory of §1 but limited to collectivizing formulas. What remains is to specify which formulas are collectivizing; it is this that is dealt with differently in the various competing theories.

3 Zermelo-Fraenkel set theory

In 1908 Ernst Zermelo published the first axiomatization of set theory. His choice of axioms seems to have been driven largely by his desire to ground a proof of the well-ordering theorem (see Axiom of choice §2; Set theory §2), which he published in the same year. Zermelo's axioms were those of ZCU^- (that is, the applied form of Z^- plus the axiom of choice) except that the axiom of infinity was replaced by Zermelo's own form (see below).

Although Zermelo's system is strong enough for many mathematical purposes, it is weaker than ZCU^- in one important respect: in it, it is impossible to develop a satisfactory theory of ordinal numbers (see Set theory §2). During the 1920s this problem was solved in two ways: ingenious methods were found, principally by Kuratowski (1921), of achieving without ordinals what is most naturally done with them; and Zermelo's system was strengthened, principally by Fraenkel and Skolem, so as to permit the theory of ordinals to be embedded in it. Zermelo-Fraenkel set theory (ZF) was the result.

We start by listing the axioms of a fragment of Zermelo-Fraenkel set theory called Z^-.

Axiom of extensionality. As before.
Axiom schema of separation. If $\Phi(x)$ is a formula in which y does not occur free, then this is an axiom:

$$\forall z \exists y \forall x (x \in y \leftrightarrow x \in z \wedge \Phi(x)).$$

$$Empty\ set\ axiom\colon \exists x \forall y\ y \notin x.$$

$$Pairing\ axiom\colon \forall x \forall y \exists z (x \in z \wedge y \in z).$$

$$Union\ set\ axiom\colon \forall a \exists x \forall y \forall z ((z \in y \wedge y \in a) \rightarrow z \in x).$$

$$Power\ set\ axiom\colon \forall x \exists y \forall z (z \subseteq x \rightarrow z \in y).$$

$$Axiom\ of\ infinity\colon \exists x (\emptyset \in x \wedge \forall y \forall z ((y \in x \wedge z \in x) \rightarrow y \cup \{z\} \in x).$$

To get Z we add:

$$Axiom\ of\ foundation\colon \forall x (x \neq \emptyset \rightarrow (\exists y \in x) x \cap y = \emptyset).$$

(See Set theory §1 for an alternative formulation.) To get ZF we add:

Axiom schema of replacement. If $\Phi(x, y)$ is a formula in which z does not occur free then this is an axiom:

$$\forall a ((\forall x \in a) \exists ! y \Phi(x, y) \rightarrow \\ \exists z (\forall x \in a)(\forall y \in z) \Phi(x, y)).$$

To get ZFC we add:

$$Axiom\ of\ choice\colon \forall a \exists b (\forall x \in a)(x \neq \emptyset \rightarrow \exists ! y (y \in x \cap b)).$$

Many other variants are possible. Below are some examples.

(1) In the presence of the replacement schema the exact form of the axiom of infinity is not critical: any axiom asserting the existence of a set which can be proved to be infinite will do. Two popular choices are the following. *Axiom of infinity (weak form)*:

$$\exists x(\emptyset \in x \wedge \forall y(y \in x \rightarrow y \cup \{y\} \in x).$$

Axiom of infinity (Zermelo's form):

$$\exists x(\emptyset \in x \wedge \forall y(y \in x \rightarrow \{y\} \in x).$$

(2) In ZF^{-} (that is, ZF without foundation) and stronger theories the pairing axiom is redundant.
(3) The separation and replacement schemata and the union set axiom can be amalgamated into the following single schema.
Axiom schema of selection and union. If $\Phi(x, y)$ is a formula in which a and b do not occur free, then this is an axiom:

$$\forall y \exists a \forall x(\Phi(x, y) \rightarrow x \in a) \rightarrow$$
$$\forall b \exists z \forall x(x \in z \leftrightarrow (\exists y \in b)\Phi(x, y)).$$

4 Von Neumann-Bernays-Gödel class theory

Zermelo-Fraenkel set theory is adequate for almost all mathematical purposes. However, its restrictions can sometimes be inconvenient. Mathematicians who use ZF and its variants are therefore accustomed to using the device of virtual sets (see §1 above), except that in order to distinguish virtual sets from the real thing they are called virtual *classes*. If we stop there, we have a good compromise: the technical simplicity of ZF plus the convenience of virtual classes when we need them. However, the urge to treat virtual objects as real is strong, and there is a popular alternative, von Neumann-Bernays-Gödel class theory (NBG), which allows us to say the same things but with classes treated as real.

Before we state the axioms of NBG we need to express what it is for a relation to be 'functional': we do this by writing $\mathrm{func}(r)$ as an abbreviation for the formula $\forall x \forall y \forall z(((x, y) \in r \wedge (x, z) \in r) \rightarrow y = z)$.

$Axiom\ of\ extensionality$: As before.

$Axiom\ of\ classes$: $\exists x \forall y((\exists z\ y \in z) \rightarrow y \in x)$.

$Empty\ set\ axiom$: $\exists x(\forall y\ y \notin x \wedge \exists z\ x \in z)$.

$Pairing\ axiom$: $(\forall x \in \mathbf{V})(\forall y \in \mathbf{V})(\exists z \in \mathbf{V})\forall u(u \in z \leftrightarrow u = x \vee u = y)$.

Axiom schema of class existence. If Φ is any formula with free variables $v_1, v_2, \ldots, v_n$ then this is an axiom:

$$(\forall v_1, v_2, \ldots, v_n \in \mathbf{V})$$
$$\exists z \forall x(x \in z \leftrightarrow (x \in \mathbf{V} \wedge \Phi^{(\mathbf{V})}).$$

$Union\ set\ axiom$: $(\forall x \in \mathbf{V})(\exists y \in \mathbf{V})\forall u(u \in y \leftrightarrow \exists v(u \in v \wedge v \in x))$.

$Power\ set\ axiom$: $(\forall x \in \mathbf{V})(\exists y \in \mathbf{V})\forall u(u \in y \leftrightarrow u \subseteq x)$.

$Axiom\ of\ infinity$: $(\exists x \in \mathbf{V})(\emptyset \in x \wedge (\forall u \in x)(\forall v \in x)(u \cup \{v\} \in x))$.

$Axiom\ of\ replacement$: $(\forall x \in \mathbf{V})\forall r(\mathrm{func}(r) \rightarrow (\exists y \in \mathbf{V})\forall u(u \in y \leftrightarrow (\exists v \in x)(v, u) \in r))$.

$Axiom\ of\ foundation$: As before.

Most of these axioms are unexciting transcriptions of the axioms of ZF. However, there is one striking difference: the axiom schema of replacement has turned into a single axiom. Even more surprisingly, the axiom schema of class existence can also be replaced by a finite list of axioms, such as the following.

$\exists a(\forall x, y \in \mathbf{V})((x, y) \in a \leftrightarrow x \in y)$
$\forall a \forall b \exists c \forall x(x \in c \leftrightarrow (x \in a \wedge x \in b))$
$\forall a \exists b(\forall x \in \mathbf{V})(x \in b \leftrightarrow x \notin a)$
$\forall a \exists b(\forall x \in \mathbf{V})(x \in b \leftrightarrow (\exists y \in \mathbf{V})((x, y) \in a))$
$\forall a \exists b(\forall x, y \in \mathbf{V})((x, y) \in b \leftrightarrow x \in a)$
$\forall a \exists b(\forall x, y, z \in \mathbf{V})((x, y, z) \in b \leftrightarrow (y, z, x) \in a)$
$\forall a \exists b(\forall x, y, z \in \mathbf{V})((x, y, z) \in b \leftrightarrow (x, z, y) \in a)$

So NBG is finitely axiomatizable. This is quite striking since, by contrast, ZF cannot be finitely axiomatized (if it is consistent). And yet NBG and ZF are of equal strength, in the sense that if Φ is any formula in the language of set theory then

$\mathrm{ZF} \vdash \Phi$ if and only if $\mathrm{NBG} \vdash \Phi^{(\mathbf{V})}$.

And moreover, NBG is consistent if and only if ZF is consistent.

Once we have taken the step of regarding classes as objects, it is puzzling why we should constrain the class existence schema by insisting that the formula defining a class can mention only sets, not proper classes. So if we are serious about the reification of proper classes, we should replace the class existence schema with the following.

Morse-Kelley schema. If $\Phi(x)$ is any formula in which y does not occur free then this is an axiom:

$\exists y \forall x(x \in y \leftrightarrow (x \in \mathbf{V} \wedge \Phi(x)))$.

The system thus obtained is known as Morse-Kelley class theory (MK; see Kelley 1955). It is (if consistent) a genuine extension of NBG, even with respect to sets. In other words, there are formulas Φ such that $\mathrm{MK} \vdash \Phi^{(\mathbf{V})}$ but $\mathrm{NBG} \nvdash \Phi^{(\mathbf{V})}$. However, the difficulty now is to know why we should stop: if classes are sufficiently real to occur legitimately in the definitions of sets, there seems no good reason why they should not be members of still more collections. As long as we refrain from calling the new collections thus countenanced ‘sets’ or ‘classes’, there is no fresh danger of inconsistency. Theories arranged in this way are not popular, though, despite their practical convenience.

5 Other axiomatizations

So far we have said little about the motivation for the axioms. Two principles have guided traditional accounts of this question: the iterative conception; and the doctrine of limitation of size. However, it is not easy to justify all the axioms of ZF on the basis of either principle on its own: for example, the replacement schema is motivated only by limitation of size; and foundation is motivated only by the iterative conception.

Other axiomatizations have been given which are closer in spirit to these principles. Here we shall describe two. The first is Scott’s way of axiomatizing the iterative conception of set, according to which sets are divided up into ‘levels’ (see Scott 1974). The set of all levels belonging to a given level is called its ‘history’. The formal details are as follows. First define the ‘accumulation’ of a set to be the set of all the elements and subsets of its members (if this set exists):

$\mathrm{acc}(a) = \{x\colon (\exists b \in a)(x \in b \vee x \subseteq b)\}$.

Next define the predicate ‘history(*a*)’ to be an abbreviation for

$(\forall b \in a) b = \mathrm{acc}(b \cap a)$.

Then define ‘level(*b*)’ to be an abbreviation for

$\exists a(\mathrm{history}(a) \wedge b = \mathrm{acc}(a))$.

Now adopt the convention that the variables V, V', V'' are restricted to range over levels, so that, for instance, $\forall V \ldots$ is an abbreviation for $\forall V(\mathrm{level}(V) \rightarrow \ldots)$. Then the axioms of Z' are as follows:

Axiom of extensionality: As before.

$Axiom\ schema\ of\ separation$: As before.

$Axiom\ of\ creation$: $\forall x \exists V\ x \in V$.

$Axiom\ of\ infinity$: $\exists V (\forall V' \in V)(\exists V'' \in V)(V' \in V'')$.

The axioms of ZF' are as follows:

$Axiom\ of\ extensionality$: As before.

$Axiom\ schema\ of\ separation$: As before.

Axiom schema of reflection. If $\Phi(x)$ is any formula then this is an axiom:

$$\exists V (\forall x \in V)(\Phi(x) \to \Phi^{(V)}(x)).$$

Then Z' is equivalent to Z and ZF' is equivalent to ZF.

The second theory we shall describe is due to Ackermann (1956). It takes $\mathbf{V}$ as a primitive constant. The axioms of the system A are as follows.

$Axiom\ of\ extensionality$: As before.

$Axiom\ schema\ of\ separation$: As before.

$Axiom\ of\ transitivity$: $\forall x \forall y (x \in y \wedge y \in \mathbf{V} \to x \in \mathbf{V})$.

$Axiom\ of\ subsets$: $\forall x \forall y (x \subseteq y \wedge y \in \mathbf{V} \to x \in \mathbf{V})$.

Ackermann's schema. If Φ is a formula in which the free variables are $v_1, v_2, \ldots, v_n$ and in which $\mathbf{V}$ does not occur, then this is an axiom:

$$(\forall v_1, v_2, \ldots, v_n \in \mathbf{V})(\forall x(\Phi(x) \to x \in \mathbf{V}) \\ \to (\exists y \in \mathbf{V})\forall x(x \in y \leftrightarrow \Phi(x))).$$

The motivation for A is not that of the iterative conception; rather, it is based on the idea that anything formed solely from sets is itself a set. The modified system A^* is something of a hybrid: it grafts onto A the following axiom, which is derived from the iterative conception and cannot be justified by the motivating idea of A:

$Axiom\ of\ foundation$: $(\forall x \in \mathbf{V})(x \neq \emptyset \to (\exists y \in x)(y \cap x = \emptyset))$.

If Φ is a formula in which the constant $\mathbf{V}$ does not occur, then

(1) if $\mathrm{Z}^- \vdash \Phi$ then $\mathrm{A} \vdash \Phi^{(\mathbf{V})}$;
(2) $\mathrm{A}^* \vdash \Phi^{(\mathbf{V})}$ if and only if $\mathrm{ZF} \vdash \Phi$.

6 Anti-foundation axioms

We have already mentioned that not all the axioms of ZF can be motivated by any one account of the notion of sethood. In particular, the axiom of foundation cannot be sustained solely on the basis of the doctrine of limitation of size. But without it where do sets get their individuation? To tell whether two sets are equal we compare their members; if necessary, we compare *their* members, and so on. In the presence of the axiom of foundation, this process cannot go on indefinitely; in its absence, it can. Now if the pattern of membership exhibited by the two sets is the same, then it seems reasonable to say that they are equal: they have done all that can be expected of them, as sets, to be regarded as identical. We are therefore led to the following principle.

Axiom of extensionality (strong form). If two sets have isomorphic $\in$-graphs, then they are equal.

(The $\in$-graph of a set is the membership relation restricted to the set, its members, their members, and so on.) As we have already noted, this assertion is provable in ZF, but in ZF^- (ZF without foundation) we need to add it as an axiom. It ensures, for instance, that there is at most one set Ω such that $\Omega = \{\Omega\}$ (since any two such sets would have to have isomorphic $\in$-graphs). However, it does not guarantee that there *is* such a set. Aczel (1988) and others have studied the consequences of adding to ZF^- not only the strong extensionality axiom just mentioned (which is essentially limitative) but also a permissive axiom generating non-well-founded sets (which

can be thought of as limit points in much the same way that irrational numbers are limits of rational numbers). The most natural of the axioms that have been studied is the following *anti-foundation axiom*:

Every graph is the $\in$-graph of some set.

7 Quine's systems

Another, more radical strand in attempts to resolve the set-theoretic paradoxes (see Paradoxes of set and property) is represented by the various forms of the theory of types (see Theory of types). We cannot, in any of the systems we have been describing so far, collect together all the objects to form another object, but we can certainly quantify over them; in the theory of types we cannot. Objects are stratified into 'types', which are not cumulative but otherwise resemble the levels of Scott's theory; variables are labelled with superscripts to indicate the type they range over. The formula $x^m \in y^n$ is taken to be well-formed only if $n = m + 1$. Now type theory is not, according to the strict definition adopted here, a form of set theory, because it does not have the axiom of extensionality. What it has instead is a schema of axioms, one for each type:

$$\forall x^{n+1} \forall y^{n+1} (\forall z^n (z^n \in x^{n+1} \leftrightarrow z \in y^{n+1}) \\ \rightarrow x^{n+1} = y^{n+1}).$$

Not to call this the axiom of extensionality may seem a pedantic cavil, particularly if the device of 'typical ambiguity' is introduced, whereby formulas are written with superscripts omitted and taken to stand schematically for any way of adding superscripts which produces a grammatical result. Thus the extensionality schema above is usually represented

$$\forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y).$$

Nevertheless, we maintain, this is not the axiom of extensionality. Typical ambiguity is not true generality. To see the difference, consider the formula $\exists y \forall x\ x \in y$: taken as typically ambiguous this stands schematically for such formulas as

$$\exists y^{n+1} \forall x^n\ x^n \in y^{n+1}$$

and is therefore true; taken as a formula in its own right it is false. The restrictions on what can be said in the theory of types have struck many writers as too severe. However, the means by which the theory avoids paradox are as secure as those of iterative set theories. Quine has proposed a system which uses type theory's method of paradox-avoidance but abandons its grammatical restrictions. Specifically, he regards formulas as genuine, not typically ambiguous, but calls a formula 'stratified' if it is possible to decorate it with superscripts so as to make it well-formed according to the theory of types. The system, which is known as NF because it was proposed in an article called 'New Foundations for Mathematical Logic' (1937), has the following axioms.

Axiom of extensionality: As before.

Axiom schema of stratified comprehension. If $\Phi(x)$ is any stratified formula in which y does not occur free then this is an axiom:

$$\exists y \forall x (x \in y \leftrightarrow \Phi(x)).$$

NF is very different from ZF. For one thing NF is finitely axiomatizable. Moreover, since the formula $x = x$ is stratified, it follows at once that there is a set of all sets: if we denote it $\mathbf{V}$ then $\mathbf{V} \in \mathbf{V}$. The axiom of choice is provably false in NF. When we come to arithmetic, it is possible in a natural way to define zero and the successor operation in NF, and then to define the set of natural numbers to be the intersection of all sets containing zero and closed under successors. This works well up to a point, but gives us mathematical induction only for stratified formulas. Unfortunately some useful formulas are not stratified. For example, we cannot prove in NF, and are therefore forced to add as an axiom, the *axiom of counting*:

If n is a natural number then it has exactly n predecessors.

In an effort to overcome this inconvenience Quine has proposed a second system called ML which stands to NF in somewhat the same relation that NBG stands to ZF. The second system takes $\mathbf{V}$ as a primitive constant and has the following axioms.

Axiom of extensionality: As before.

Axiom schema of class existence. If $\Phi(x)$ is a formula in which y does not occur free then this is an axiom:

$$\exists y \forall x(x \in y \leftrightarrow y \in \mathbf{V} \wedge \Phi(x)).$$

Axiom schema of set existence. If $\Phi(x, v_1, v_2, \ldots, v_n)$ is a stratified formula in which no variables other than those listed occur free, then this is an axiom:

$$\forall v_1, v_2, \ldots, v_n \exists y \in \mathbf{V} \\ \forall x(x \in y \leftrightarrow \Phi^{(\mathbf{V})}(x, v_1, \ldots, v_n)).$$

In contrast to NF, ML is not finitely axiomatizable if it is consistent. If Φ is any formula in which $\mathbf{V}$ does not occur then

$$\mathrm{NF} \vdash \Phi \text{ if and only if } \mathrm{ML} \vdash \Phi^{(\mathbf{V})}.$$

Thus ML is consistent if and only if NF is consistent. However, it is still not known whether ML and NF are consistent if ZF is consistent. If we define the natural numbers in superficially the same way as in NF, we get mathematical induction even for non-stratifiable formulas, obviating the need for an axiom of counting. However, this time we cannot prove that the class of natural numbers is a set (if ML is consistent). So now we have to add *this* as an axiom. Thus neither NF nor ML suffices without supplementation as a basis for mathematics.

See also: Logical and mathematical terms, glossary of

MICHAEL POTTER

References and further reading

Ackermann, W. (1956) ‘Zur Axiomatik der Mengenlehre’ (On the Axiomatization of Set Theory), *Mathematische Annalen* 131: 336-45.(The source for the system A mentioned in §5.)

Aczel, P. (1988) *Non-Well-Founded Sets*, Stanford, CA: Center for the Study of Language and Information, Stanford University.(An account of anti-foundation axioms, aimed at the mathematically aware.)

Boolos, G. (1989) ‘Iteration Again’, *Philosophical Topics* 17 (2): 5-21.(An interesting but fairly technical discussion of the iterative and limitation-of-size conceptions of set.)

Bourbaki, N. (1968) *Theory of Sets*, New York: Addison-Wesley.(A development of set theory in a system equivalent to ZF^- together with a strong form of the axiom of choice. The system is notable for its employment of the axiom schema of selection and union referred to in §3.)

Forster, T.E. (1992) *Set Theory with a Universal Set*, Oxford: Oxford University Press.(A survey of work on NF, well written but aimed at logicians.)

Fraenkel, A.A. (1922) ‘Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre’ (On the Foundations of Cantor-Zermelo Set Theory), *Mathematische Annalen* 86: 230-7.(This paper is the reason Fraenkel’s name is associated with the standard formulation of set theory.)

Fraenkel, A.A., Bar-Hillel, Y. and Levy, A. (1958) *Foundations of Set Theory*, Amsterdam: North Holland; revised 2nd edn, 1973.(Detailed discussions of most of the systems of set theory mentioned here and a wealth of historical information and references.)

Gödel, K. (1940) *The Consistency of the Continuum Hypothesis*, Princeton, NJ: Princeton University Press, 1994. (The definitive formulation of NBG.)

Hallet, M. (1984) *Cantorian Set Theory and Limitation of Size*, Oxford: Clarendon Press.(A detailed study of the early history of set theory.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press.(It is possible to trace the historical development of ZF and NBG through the papers reprinted here.)

Kelley, J.L. (1955) *General Topology*, Princeton, NJ: Van Nostrand.(The appendix includes the first published version of MK.)

Kuratowski, K. (1921) ‘Sur la notion d’ordre dans la théorie des ensembles’ (On the Notion of Order in Set Theory), *Fundamenta Mathematicae* 2: 161-71.(Mentioned in §3.)

Lewis, D.K. (1991) *Parts of Classes*, Oxford: Blackwell.(An entertaining, non-technical discussion of a way of separating the hierarchical from the non-hierarchical part of set theory.)

Morse, A. (1965) *A Theory of Sets*, New York: Academic Press.(A treatment of the theory of classes based on a version of MK.)

Neumann, J. von (1925) 'Eine Axiomatisierung der Mengenlehre' (An Axiomatization of Set Theory), *Journal für die reine und angewandte Mathematik* 154: 219-40; trans. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967.(The first formulation of an axiomatization similar to what is now called NBG.)

Potter, M.D. (1990) *Sets: An Introduction*, Oxford: Oxford University Press.(An account of set theory based on a Scott-type axiomatization; intended for the mathematically-minded.)

Quine, W.V. (1937) 'New Foundations for Mathematical Logic', *American Mathematical Monthly* 44: 70-80. (Principally of historical interest. For up-to-date information about NF consult Forster (1992).)

Quine, W.V. (1940) *Mathematical Logic*, Cambridge, MA: Harvard University Press; revised edn, 1951.(Very technical, but includes a wealth of information.)

Quine, W.V. (1963) *Set Theory and its Logic*, Cambridge, MA: Harvard University Press; revised edn, 1969. (Rather idiosyncratic but extremely clear and well written.)

Scott, D. (1974) 'Axiomatizing Set Theory', in *Axiomatic Set Theory*, Proceedings of the Symposium of Pure Mathematics 13 (2), Providence, RI: American Mathematical Society, 207-14.(The primary source for Scott's axiomatization of ZFrsquo; described in §5.)

Skolem, T. (1923) 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre', in *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, Helsinki: Akademiska Bokhandeln, 217-32; trans. S. Bauer-Mengelberg, 'Some Remarks on Axiomatized Set Theory', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 291-301.(The first precise formulation of the axiom schema of replacement and of Zermelo's notion of 'definite property'.)

Zermelo, E. (1908) 'Untersuchungen über die Grundlagen der Mengenlehre I', *Mathematische Annalen* 65: 261-81; trans. S. Bauer-Mengelberg, 'Investigations in the Foundations of Set Theory', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 200-15.(The first published axiomatization of set theory.)

Sextus Empiricus (*fl. c.* AD 200)

Sextus Empiricus is our major surviving source for Greek scepticism. Three works of his survive: a general sceptical handbook (Outlines of Pyrrhonism), a partly lost longer treatment of the same material, and a series of self-contained essays questioning the utility of the individual liberal arts.

Little is known about the life of Sextus Empiricus, our major source for Pyrrhonism. He was a doctor of the Empirical school, a sceptically-minded medical sect (see Hellenistic Medical Epistemology), and wrote some (lost) medical treatises. He also wrote three treatises on scepticism. *Outlines of Pyrrhonism*, is a three-volume handbook of sceptical arguments. The first covers the nature of Pyrrhonism and its distinction from other philosophies, the proper interpretation of the sceptic’s characteristic expressions (see Pyrrhonism §§3, 6), and the general form of sceptical argument (see Pyrrhonism §2). The second deals with the Pyrrhonian attack on dogmatic logic and epistemology, in particular with the issue of the criterion, and the nature of signs and proof (see Pyrrhonism §4). The third is on physics, attacking dogmatic notions of cause, motion, generation, space and time, number, and so on (see Pyrrhonism §5), and ethics (see Pyrrhonism §2).

A fuller treatment of the same material is standardly known as Books VII-XI of *Against the Professors* (sometimes called *Against the Mathematicians*). These break down into three subsections: *Against the Logicians* (Books VII-VIII), *Against the Physicists* (Books IX-X) and *Against the Ethicists* (Book XI), which contain the material to be found in Books II and III of *Outlines of Pyrrhonism*, as well as a reworking of some of the contents of *Outlines of Pyrrhonism* I 210-41. As Janáček (1963) has shown, *Against the Professors* VII-XI is incomplete, lacking at least one preliminary book corresponding to the bulk of Book I of *Outlines of Pyrrhonism*.

Finally, Sextus composed a series of treatises against the dogmatists’ conceptions of the liberal arts (grammar, rhetoric, geometry, arithmetic, astrology and music: *Against the Professors* I-VI), the scepticism of which seems a good deal more moderate in tone, and whose arguments owe much to Epicureanism.

Sextus makes no claims to originality as a thinker, and he clearly owes a great deal to Aenesidemus, and to subsequent Pyrrhonist systematizers, especially Agrippa. However, the Pyrrhonism he outlines is clear, consistent and rigorous; and, while the quality of the argumentation it contains is, as Sextus himself admits (*Outlines of Pyrrhonism* III 280), uneven, the Sextan version of scepticism became canonical. Diogenes Laertius (IX 116) mentions him as a prominent Pyrrhonian; and in the fourth century St Gregory of Nazianzus excoriates him in the same breath as Pyrrho himself, as the originator of ‘the vile and malignant disease’ of scepticism (*Orations* XII 21).

In 1564, Henri Étienne (better known as ‘Stephanus’) published a Latin translation of *Outlines of Pyrrhonism*; it was read with eagerness by Montaigne, and the ghost of Sextan argument can often be discerned behind Montaigne’s elegant and discursive prose. More importantly still, Greek scepticism came by this route to influence Descartes, and to provide the ultimate stimulus for his search for an epistemological foundation of knowledge and science proof against sceptical assault; as such, Sextus has had, albeit indirectly, a profound impact on the development of modern philosophy.

R.J. HANKINSON

List of works

Sextus Empiricus (*c.* AD 200) *Outlines of Pyrrhonism*, trans. J. Annas and J. Barnes, *Outlines of Scepticism*, Cambridge: Cambridge University Press, 1994.(Fine translation with introduction and notes.)

Sextus Empiricus (*c.* AD 200) *Against the Professors*, trans. R.G. Bury, *Against the Logicians*, *Against the Physicists*, *Against the Ethicists* and *Against the Professors*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 3 vols, 1935, 1936, 1949.(Parallel Greek text and English translation with minimal notes.)

References and further reading

Diogenes Laertius (*c.* early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London:

Heinemann, 1925, 2 vols.(Parallel Greek text and English translation; IX 61-116 is devoted to Pyrrhonism.)

House, D.K. (1980) ‘The Life of Sextus Empiricus’, *Classical Quarterly* 30: 227-38.(Stresses how little we actually know about Sextus Empiricus.)

Janáček, K. (1963) ‘Die Hauptschrift des Sextus Empiricus als Torso erhalten?’ (Does Sextus Empiricus’ Principal Work survive as a Torso?), *Philologus* 107: 271-7.(Establishes that our text of *Against the Professors* is incomplete.)

Sexuality, philosophy of

The philosophy of sexuality, like the philosophy of science, art or law, is the study of the concepts and propositions surrounding its central protagonist, in this case 'sex'. Its practitioners focus on conceptual, metaphysical and normative questions.

Conceptual *philosophy of sex analyses the notions of sexual desire, sexual activity and sexual pleasure. What makes a feeling a sexual sensation? Manipulation of and feelings in the genitals are not necessary, since other body parts yield sexual pleasure. What makes an act sexual? A touch on the arm might be a friendly pat, an assault, or sex; physical properties alone do not distinguish them. What is the conceptual link between sexual pleasure and sexual activity? Neither the intention to produce sexual pleasure nor the actual experience of pleasure seems necessary for an act to be sexual. Other conceptual questions have to do not with what makes an act sexual, but with what makes it the type of sexual act it is. How should 'rape' be defined? What the conceptual differences are, if any, between obtaining sex through physical force and obtaining it by offering money is an interesting and important issue.*

Metaphysical *philosophy of sex discusses ontological and epistemological matters: the place of sexuality in human nature; the relationships among sexuality, emotion and cognition; the meaning of sexuality for the person, the species, the cosmos. What is sex all about, anyway? That sexual desire is a hormone-driven instinct implanted by a god or nature acting in the service of the species, and that it has a profound spiritual dimension, are two - not necessarily incompatible - views. Perhaps the significance of sexuality is little different from that of eating, breathing and defecating; maybe, or in addition, sexuality is partially constitutive of moral personality.*

Normative *philosophy of sex explores the perennial questions of sexual ethics. In what circumstances is it morally permissible to engage in sexual activity or experience sexual pleasure? With whom? For what purpose? With which body parts? For how long? The historically central answers come from Thomist natural law, Kantian deontology, and utilitarianism. Normative philosophy of sex also addresses legal, social and political issues. Should society steer people in the direction of heterosexuality, marriage, family? May the law regulate sexual conduct by prohibiting prostitution or homosexuality? Normative philosophy of sex includes nonethical value questions as well. What is good sex? What is its contribution to the good life?*

The breadth of the philosophy of sex is shown by the variety of topics it investigates: abortion, contraception, acquaintance rape, pornography, sexual harassment, and objectification, to name a few. The philosophy of sex begins with a picture of a privileged pattern of relationship, in which two adult heterosexuals love each other, are faithful to each other within a formal marriage, and look forward to procreation and family. Philosophy of sex, as the Socratic scrutiny of our sexual practices, beliefs and concepts, challenges this privileged pattern by exploring the virtues, and not only the vices, of adultery, prostitution, homosexuality, group sex, bestiality, masturbation, sadomasochism, incest, paedophilia and casual sex with anonymous strangers. Doing so provides the same illumination about sex that is provided when the philosophies of science, art and law probe the privileged pictures of their own domains.

1 Conceptual analysis

The philosophy of sex investigates conceptual, metaphysical and normative questions, although the boundaries between these three are hardly firm. Metaphysical and normative philosophy of sex are well developed, stretching back to Plato and Augustine (see Plato §12); sexual ethics has a famous history, and the contemplation of the place of sexuality in human nature is central to Christianity. The analysis of sexual concepts, by contrast, is in its infancy. The subjects of analysis are these core concepts and the logical relationships among them: sexual desire, sensation, pleasure, act, arousal and satisfaction. Derivative sexual concepts, which presuppose an understanding of the core concepts, are also the subject of analysis. Among these are rape, sexual harassment, sexual orientation, sexual perversion, prostitution and pornography (see Pornography).

Consider adultery. It can be defined as a sexual act that occurs between two persons, at least one of whom is married but not to the other. The definition should also mention, as a necessary condition, willing and knowledgeable consent. Suppose *X* coerces *Y*, who is married to *Z*, into coitus. *Y* did not commit adultery, because

Y did not have the proper frame of mind; *Y* never intended to commit adultery or engage in intercourse at all. Or suppose *X* and *Y* engage in coitus, both believing, on the basis of good evidence (but falsely), that *X*'s spouse *Z* died years ago; or the unmarried *X* has good reason to think (but falsely, due to *Y*'s deception) that *Y*, too, is not married. Has *X* committed an adulterous sexual act, unwittingly and so, perhaps, not culpably; or is *X*'s lack of *mens rea*, *X*'s ignorance of the true state of affairs, incompatible with adultery?

We cannot fully understand the derivative sexual concept 'adultery' until we have defined 'sexual act'. If *X* and *Y* send to each other sexually arousing messages through the Internet, have they engaged in a sexual act ('cybersex')? Is their exchange of messages sexual enough, quantitatively or qualitatively, to be adulterous, if one of them is married? Here we can see the intertwining of conceptual and moral matters. A lack of clarity about 'sexual act' allows the exoneration of adultery by a convenient redescription of what occurs between *X* and *Y* - it was only 'fooling around', not 'real' sex. Another, quite opposite, manoeuvre, is possible. Theologians often define adultery in the spirit of Matthew 5: 28, making a sexual fantasy sufficient, even in the absence of physical contact: *X* commits adultery if *X* thinks lustful thoughts about *Y*.

2 Sexual activity

Our interest in defining 'sexual act' is not merely philosophical; it is also practical. Precise definitions of 'sexual act' are needed for social scientific studies of sexual behaviour and orientation (to be used in the consideration of questions about, for example, who engages in homosexual acts and whether this correlates with genetics, and how often people engage in sex) and for legislation in the areas of child abuse, rape, harassment and adultery.

Sexual acts might be defined as those involving sexual body parts. The sexual parts of the body are first catalogued; acts are sexual if and only if they involve contact with one of these parts. 'Sexual act', on this view, is logically dependent on 'sexual part'. But do we clearly understand 'sexual part'? Two people might shake hands briefly, without the act being sexual; they could, alternatively, warmly press their hands together and feel a surge of sexual pleasure. Sometimes, then, the hands are used nonsexually and sometimes they are used sexually. Are the hands a sexual part? Whether the hands are a sexual part depends on the activity in which they are engaged. Hence, instead of an act's being sexual because it involves a sexual body part, a body part is sexual because of the sexual nature of the act in which it is used. We might say that a genital examination is not a sexual act even though the genitals are touched; hence contact with a sexual part is not sufficient for an act to be sexual. But we could also say that not even the genitals are sexual parts in the requisite sense; for in the medical context the genitals are not being treated as sexual parts.

The morality of sexuality has been understood by some in terms of its procreative function (see §5). Alternatively, the procreative nature of sexual activity might be employed analytically rather than normatively. Sexual acts, on this view, are those having procreative potential in virtue of their biological structure. The principal case of such an act is heterosexual intercourse. This analysis, then, is too narrow, if taken as stating a necessary condition. Here is a more plausible formulation: sexual acts are (1) acts that are procreative in structure and (2) any acts that are the physiological or psychological precursors or concomitants of acts that are sexual by (1). This version casts a wider net, but not wide enough. Masturbation, which is not a procreative act and not often a precursor or concomitant of coitus, turns out not to be a sexual act. Another emendation suggests itself: sexual acts are also (3) acts that bear a close physical resemblance to the acts judged sexual by (1) or (2). This vague condition does not save the proposed analysis. Some sexual perversions (such as fondling shoes) are sexual even though they bear no reasonable resemblance to coitus or its concomitants. This analysis also suggests that homosexual acts, all of which are nonprocreative, are sexual just because they sufficiently resemble heterosexual acts. That seems to be the wrong reason for the right conclusion.

Another view is that both homosexual and heterosexual acts are sexual in virtue of the type of pleasure or sensation they produce. Thus it seems reasonable to propose that sexual acts are those that produce sexual pleasure (Gray 1978). But if pleasure is the criterion of the sexual, pleasure cannot be the gauge of the nonmoral quality of sex acts. The couple who have lost sexual interest in each other, and who engage in routine coitus from which they derive no pleasure, are still performing a sexual act. We are forbidden, by this analysis, from saying that they engage in sex but that it is (nonmorally) bad sex. Rather, we can say, at most, that they tried to engage in sex, and failed. Furthermore, in this analysis 'sexual act' is logically dependent on 'sexual pleasure', so we cannot say that sexual pleasure is the pleasure produced by sexual acts. Then how might we distinguish sexual pleasure or

sensations from others? This problem also arises for a more complex analysis: sexual acts are those acts that tend to satisfy sexual desire, where sexual desire is taken to be the desire for the pleasure of physical contact (Goldman 1977). ‘Pleasures of physical contact’ might not specify sexual pleasure accurately enough. An additional complication is that a gender difference in the experience and conceptualization of sexual pleasure might exist. Additionally, someone might experience sexual desire yet have no idea what to do as a result of having it, no idea that physical contact, or what kind of physical contact, is the next, but hardly mandatory, step. Sexual desire, as argued by Jerome Shaffer (1978: 186-7), might not be a desire *for* something or *that* something at all. What, then, are the features of a desire that make it sexual? Sexual desire is distinguished, on Shaffer’s account, by being accompanied by sexual excitement and arousal. We can, in turn, ask what sexual excitement and arousal are, as opposed to other kinds of excitement and arousal. For Shaffer, sexual arousal is ‘directly sexual in that it involves the sexual parts, viz., the genital areas’. Have we gone full circle?

Finally, sexual acts might be understood as those involving a sexual intention. But an intention to produce or experience sexual pleasure, for example, might be neither necessary nor sufficient for an act to be sexual. A couple engaging in coitus, both parties intending only that fertilization occur and neither concerned with sexual pleasure, performs a sexual act. Maybe this is not the correct sexual intention. But the intention to procreate is not it: gays and lesbians experience desire and arousal and engage in sex without any procreative intent. Furthermore, intentions are arguably irrelevant in making sexual acts sexual. Rape can be sexual whether the rapist intends to get sexual pleasure from it, to humiliate his victim, or to assert his masculinity. From the fact that in some rapes, rapists intend to degrade their victims, to dominate and exert power over them, it does not follow that the act is not sexual. Indeed, the rapist might have chosen a sexual act quite on purpose as his method to humiliate and degrade. His victim is degraded exactly by the sexual nature of the act endured; the victim experiences a shame that accompanies a forced sexual act but would not accompany sexless assault.

3 Social constructionism

What, then, are sexual acts? Maybe they have no transcultural or ahistorical essence, and the analytic project is doomed. Acts involving the same body parts are sometimes sexual, sometimes not. Some touches and movements are deemed sexual in one culture but not in others; the fragrances, mannerisms and costumes that are sexually arousing vary among places and times. No lowest common denominator exists that makes all sexual acts sexual. Bodily movements acquire meaning - as sexual, or as something else - by existing within a culture that attaches meaning to them. There are, then, only variable social definitions of the sexual.

Such is the view known as social constructionism (or anti-essentialism). As one proponent puts it, ‘the very meaning and content of sexual arousal’ varies so much among genders, classes, and cultures that ‘there is *no* abstract and universal category of "the erotic" or "the sexual" applicable without change to *all* societies’ (Padgug 1979: 54; original emphasis). Nancy Hartsock elaborates:

> We should understand sexuality not as an essence or set of properties defining an individual, nor as a set of drives and needs (especially genital) of an individual. Rather, we should understand sexuality as culturally and historically defined and constructed. Anything can become eroticized.
>
> (1983: 156)

Hartsock’s expression ‘anything can become eroticized’ must mean ‘anything can be linked to sexual arousal and pleasure’. That might be true; after all, unusual items bring paraphiliacs sexual joy. If so, however, there seems to be a common denominator after all, an essential even if narrow core to the sexual: an unchanging, culturally invariable subjective experience of sexual pleasure.

The history of sexuality is the history of our discourse about sex, as Michel Foucault might have put it. We create things by using words. There really is no such thing as masturbatory insanity or nymphomania - no medical condition, no psychological character trait, no underlying pathology. Well, there is, but only because we have picked out some behavioural patterns and made up a word to name them, not because masturbatory insanity and nymphomania have, like the moon, an existence independent of our words, our observations, and our evaluations of it. Social facts, such as the existence of ‘peasants’, ‘witches’ and ‘yuppies’ (Edward Stein’s examples), have an odd, plastic, fuzzy nature. Similar considerations apply to ‘perversion’, ‘philanderer’ and ‘homosexuality’. Thus the title of David Halperin’s social constructionist monograph, *One Hundred Years of Homosexuality* (1990). It did

not exist before the word 'homosexual' was coined by Károly Mária Benkert in 1869.

Adrienne Rich asserts that even heterosexuality is a 'man-made institution'. Heterosexuality is 'forcibly and subliminally imposed on women', 'designed to keep [them] within a male sexual purlieu' (1986: 34-57). Ti-Grace Atkinson similarly claims that 'sexual intercourse is a political construct, reified into an institution' (1974: 13). People have been engaging in sexual intercourse because doing so serves not their own individual needs, but a political purpose: to keep the species (and the system) going by replenishing the population. People have been groomed by society to perform the act of intercourse, their desire-pleasure apparatus manipulated, even created, to serve this purpose. Hence, when the development of brave-new-world reproductive technology eliminates the social function of coitus, individuals' 'sexual "drives" and "needs" would disappear' as well (20), there not being any reason to keep the grooming in operation. The end of biological reproduction is also the end of sex, or at least of sex as we know it. Whether that possible state would be a cause for celebration or despair is unclear.

4 The metaphysics of sex

In sex we are vulnerable to another's enticing words and seductive touch. Our wills are weak, so we are dominated as much by our desires as by the other's physical strength or alluring beauty. Engulfment or invasion of our bodies by the other's gaze and flesh is hazardous. In exchange for exquisite pleasure, we make ourselves susceptible to embarrassment, anguish, betrayal (see Vulnerability and finitude §2). This psychology of sex provides reason for taking sex, and so sexual ethics, seriously. Its consequences - transmission of disease, the existence of a new human being - do so as well. If procreation is a couple's contribution to God's ongoing work of creation, if it is the sacred ground where humans and God engage in a shared project, then sexuality must be protected by stringent ethics. Or if sexual personality resides at the core of moral personality, if the training of sexuality impinges on developing character in such a way that the failure to learn to control the pursuit of sexual pleasure undermines the achievement of virtue, then the moral education of the body is crucial.

In light of sexuality's intricate psychology and far-reaching consequences, sexual activity might be justifiable only by weighty nonsexual considerations. Consider the hostility of Christianity to sex, as in Augustine's *De nuptiis et concupiscentia (On Marriage and Concupiscence)*, where we hear the strains of Plato's *Phaedrus*:

> A man turns to good use the evil of concupiscence…when he bridles and restrains its rage…and never relaxes his hold upon it except when intent on offspring, and then controls and applies it to the carnal generation of children…not to the subjection of the spirit to the flesh in a sordid servitude.
>
> (I 9)

Neither is Immanuel Kant kind to sexuality: 'Sexual love makes of the loved person an Object of appetite…. Taken by itself it is a degradation of human nature' (1780-81: 163). If sexual desire objectifies, in virtue of pushing us to seek pleasure without regard for our partners, if sexual urges engender deception and manipulation, then sexuality is morally suspicious. Only special circumstances could make acting on these desires morally right. For these gloomy reasons, many conservative and religious thinkers, but also some feminists, believe that sexual activity is redeemed only by love or marriage.

In contrast, sexual liberals suppose that sexuality is a wholesome bonding mechanism that allows persons to overcome the psychological and moral tension between egoism and benevolence (see Egoism and altruism). Sexual activity involves pleasing the self and the other at the same time; these exchanges of pleasure generate gratitude and affection. Further, sexual pleasure is a valuable thing in its own right, the pursuit of which does not require external or nonsexual justification. And sexuality is a cardinal affirmation of the goodness of bodily existence. There is no contradiction in presuming that a virtuous person can lead a life in which sexual pleasure is sought for its own sake - in moderation, of course. Weak, not stringent, moral rules apply to sex.

The claim that sexual pleasure is valuable does not mean we should not condemn sexual misconduct. We often do, however, pardon sexually motivated misconduct when we would not excuse similar conduct motivated otherwise (see Pausanias in Plato's *Symposium* 182e-183c). Does sex warrant this exculpation? If the most intense way we relate to another person is sexually, then maybe forgiveness for sexual offences should be graciously forthcoming. But sexuality is hardly unique in the depth of the personal relationship it elicits; think about mutual hatred. Nor is there much intensity in the dull coitus of a long-married couple. According to another line of thought, sexuality has

a peculiar ability to thwart reason: sexual impulses make us temporarily deranged (recall the dark horse of Plato's *Phaedrus*). We are to be excused because in sexual matters we cannot control ourselves. But the lures of politics, ambition and money are just as powerful and devilish as the anticipations of the flesh. In none of these respects does sexuality seem unique or significant enough to deserve special attention.

Perhaps sexual desire - as a component of love and as opposed to mere 'horniness' - latches on to particular objects (I want *Jennifer*), in a way hunger does not (I want a sausage, and any fat, juicy one will do). Genuine *eros* makes us desire a particular person; crude desire is satisfiable by fungible bodies. But the distinction between *eros* and lust is a fine one, and in many instances doubtful; we only deceive ourselves that this person is not replaceable by others in our affections. As Roger Scruton puts it, 'a metaphysical illusion resid[es] in the heart of sexual desire' (1986: 95). Sexual passion misleads us; it makes it *appear* that we are ontologically more than we are, transcendental selves rather than mere material beings (1986: 130).

Plato, in the *Symposium*, also issued a warning: what we think we are seeking is not really what we want; our *eros* for bodies is really *eros* for truth and beauty. Augustine similarly thought that the search for God was hidden beneath the search for sensual pleasure in another's body. And for Arthur Schopenhauer, the beauty of the object of sexual desire was nature's way of tricking us (men?) into thinking that the satisfaction of our erotic love for our beloveds is for our own individual good. To the contrary, sexual love benefits only the species, for the good of which nature makes mere use of us (1844: 538-40) (see Schopenhauer, A. §4). This naturalist vision of sex is not far removed from Atkinson's anti-essentialism; both see its purpose in terms of the species and not of the individual.

5 Aquinas and natural law

For Aquinas, sexual acts are morally wrong in two different ways (*Summa theologiae* IIaIIae.154.1) (for an introduction to his ethical views, see Aquinas, T. §13). First, 'when the act of its nature is incompatible with the purpose of the sex-act [procreation]. In so far as generation is blocked, we have unnatural vice, which is any complete sex-act from which of its nature generation cannot follow.' Aquinas gives four examples (IIaIIae.154.11) of sexual acts that are unnatural vice because not procreative: 'the sin of self-abuse', 'intercourse with a thing of another species', acts 'with a person of the same sex', and acts in which 'the natural style of intercourse is not observed, as regards the proper organ or according to other rather beastly and monstrous techniques'. Second, 'the conflict with right reason may arise from the nature of the act with respect to the other party', as in incest, rape, seduction and adultery.

Sexual sins in the first category are the worst: 'unnatural vice flouts nature by transgressing its basic principles of sexuality, [so] it is in this matter the gravest of sins' (IIaIIae.154.12). Aquinas is replying to an interlocutor who argues that unnatural vice is *not* the morally worst sex. 'The more a sin is against charity', says the interlocutor, 'the worse it is. Now adultery and seduction and rape harms our neighbour, whereas unnatural lust injures nobody else, and accordingly is not the worst form of lust.' Aquinas rejects this thinking. Seduction, rape and adultery violate only 'the developed plan of living according to reason' that derives from humans living in society, while 'unnatural sins', which violate the plan of creation, are an 'affront to God'. If some sexual acts are unnatural, they are morally wrong, in Thomistic ethics, just for that reason. To the list of reasons sexual acts might be wrong - they are dishonest, cruel, unfair, manipulative, coercive, exploitative, selfish or negligently dangerous - a Thomist adds 'unnatural'. Not so the sexual libertarian. Mutual consent is, in the absence of third-party harm, sufficient for the morality of sexual acts, and no law of God or nature need supplement this basic principle of proper relations among humans.

In arguing that sexual behaviour ought to conform to human nature within God's plan, one must be able to justify particular conceptions of that nature and his intentions. Augustine and Aquinas knew, or thought they did, that God wanted sexuality to be the mechanism of procreation. Aquinas displays confidence in his account of human nature:

> It is evident that the bringing up of a human child requires the care of a mother who nurses him, and much more the care of a father, under whose guidance and guardianship his earthly needs are supplied and his character developed. Therefore indiscriminate intercourse is against human nature. The union of one man with one woman is postulated, and with her he remains, not for a little while, but for a long period, or even for a

whole lifetime.

(IIaIIae.154.2)

Christine Gudorf (1994), however, argues for a new, non-Augustinian yet Christian understanding of the body. Basing her view in part on the existence of the clitoris, an organ that on her view has only a pleasure-producing function and no procreative function, Gudorf argues that God designed the human body foremost for sexual pleasure.

Although he appeals to biology and not the Lord's plan, Michael Levin concludes, in agreement with Aquinas, that homosexual acts 'involve the use of the genitals for what they aren't for' (1984: 253). Homosexual anal intercourse is unnatural because being inside another man's rectum is not what a man's penis is 'for'; it is for penetrating a woman's vagina. This kind of Thomism is susceptible to mockery (Does a man's masturbation misuse the penis, or the hand? Does heterosexual cunnilingus misuse the tongue?) but it has also received serious discussion (see Murphy 1987). In deriving ethical judgments directly from nature, the problem is to come up with a coherent, plausible account of the 'natural'. Whereas Aquinas claims that a man's 'indiscriminate intercourse is against human nature', much recent biology claims that promiscuity in men is perfectly natural, the result of evolutionary mechanisms. In this way, science has the power to turn philosophical or theological reliance on nature on its head (see Evolution and ethics §§2-3). Thus, if the current research suggesting that homosexual orientation has a substantial genetic basis is vindicated, the Western religions might have to concede that such sexual desires and behaviour are, after all, part of the design of nature. Patricia Jung and Ralph Smith (1993) offer a Christian defence of loving, homosexual marriages, in part based on the idea that being homosexual is little different from being left-handed.

6 Kant's sexual ethics

Most philosophers have had something to say about sex; with a little digging, one can uncover the unsystematic sexual thoughts of Aristotle, Descartes and Hume. Others - such as Schopenhauer, Kierkegaard and Sartre - took sexuality more seriously. And for some, most notably Plato and Freud, the sexual was nearly the heart and soul. But in Kant's sexual philosophy, the conceptual, the metaphysical and the ethical are most provocatively combined; and in Kant, contemporary philosophical problems and disputes about sex can be glimpsed clearly.

Aquinas' fame (or notoriety) rests with his natural law sexual ethics; Kant is important as the author of a sexual ethics of respect (see Respect for persons). For Kant, human sexual interaction involves one person's merely using another for the sake of pleasure:

> there is no way in which a human being can be made an Object of indulgence for another except through sexual impulse…. Sexual love makes of the loved person an Object of appetite…. Sexual love…by itself and for itself…is nothing more than appetite…. As an Object of appetite for another a person becomes a thing and can be treated and used as such by every one. This is the only case in which a human being is designed by nature as the Object of another's enjoyment.

(1780-81: 163)

If all sexual acts - not only rape, or those in which consent is absent - are objectifying and instrumental, is not celibacy required? Kant thinks not. Following Paul ('The wife hath not power of her own body, but the husband: and likewise also the husband hath not power of his own body, but the wife', 1 Corinthians 7: 4), Kant lays down a stringent rule:

> The sole condition on which we are free to make use of our sexual desire depends upon the right to dispose over the person as a whole…. If I have the right over the whole person, I have also the right…to use that person's *organa sexualia* for the satisfaction of sexual desire. But how am I to obtain these rights over the whole person? Only by giving that person the same rights over the whole of myself. This happens only in marriage. Matrimony is an agreement between two persons by which they grant each other equal reciprocal rights, each of them undertaking to surrender the whole of their person to the other with a complete right of disposal over it…. If I yield myself completely to another and obtain the person of the other in return, I win myself back…. In this way the two persons become a unity of will…. Thus sexuality leads to a union of human beings, and in that union alone its exercise is possible.

(1780-81: 166-7)

Mary Ann Gardell, instead of sensing the Pauline ‘marriage debt’ in the Kantian exchange of rights, reads Kant as benignly claiming that ‘marriage transforms an otherwise manipulative masturbatory relationship into one that is essentially altruistic in character’ (1987: 11). But Kant speaks of marriage as a contract, an exchange of rights of access to the body. So he might be claiming that the marital pledge, the voluntary assumption of the terms of an agreement, assures that the spouses are not treating each other only as means, but also as ends, in the marriage bed. Or Kant might be justifying marital sex by abolishing the mere possibility of instrumentality: after the ontological union of two persons into one by marriage, there cannot *be* any use of one person by another (Baker and Elliston 1984: 26-7). This is why Kierkegaard found such views not benign but pernicious. ‘All pleasure is selfish. The pleasure of the lover…is not selfish with respect to the loved one, but in union they are both absolutely selfish, inasmuch as in union and in love they constitute one self’ (1845: 56).

Kant’s idea that marriage cleanses sex of instrumentality apparently implies that homosexual marriage would similarly cleanse same-sex sexuality. Kant sidesteps this conclusion, asserting that homosexuality is one of the *crimina carnis contra naturam*:

> Onanism…is abuse of the sexual faculty…. By it man sets aside his person and degrades himself below the level of animals…. Intercourse between *sexus homogenii*…too is contrary to the ends of humanity; for the end of humanity in respect of sexuality is to preserve the species without debasing the person.
>
> (1780-81: 170)

The homosexual ‘no longer deserves to be a person’. Kant, following Augustine and Aquinas, condemns nonprocreative sex as unnatural, even if it is, in his own sense, noninstrumental.

Kant’s notion of marriage, in which a person obtains rights over a person and their genitals, might itself reduce the spouses to sex objects, unless the voluntary agreement of the spouses to the arrangement is sufficient to eliminate mere use. But if we emphasize the voluntary nature of the exchange of rights that occurs in marriage, Kant’s contention that sex is permissible only in marriage is undermined, because two people - gay or straight - can grant each other reciprocal rights to dispose over their persons for a limited period of time (as in the casual sex of one evening). Nothing in the idea of an exchange of rights seems to entail that the exchange must be forever or exclusive. Is there something irreversible about *this* exchange of rights? If not, Kant’s defence of monogamous, lifelong marriage in terms of a ‘unity of will’ is no more convincing than Aquinas’ appeal to human nature.

7 Contemporary Kantians

From the various strands of Kant’s sexual philosophy, twentieth-century sexual conservatives and liberals have fashioned their own brands of Kantian sexual ethics.

Libertarian sexual ethics. If oral and anal sex, gay and lesbian sex, bisexual and group sex, contraceptive coitus, wearing lingerie and cosmetics, adultery, prostitution, making or viewing pornography, and the paraphilias - the things often condemned by conservative sexual ethics - can be carried out without harm befalling the participants or others, by consenting adults who know what they want, how could they be morally wrong? According to libertarian sexual ethics, as long as the persons involved are participating voluntarily, they are not merely using each other for their own purposes; the free and informed consent of each to the acts that occur is sufficient to eliminate mere use and thus to make sexual activity, of whatever flavour, morally permissible (see Libertarianism). The paradigmatically wrong sexual act is not buggery, but rape, in which the absence of consent makes the mere use obvious. Consensual participation in sexual activity implies that each person is respecting the other as an autonomous agent capable of making up their mind about the value of the activity. Furthermore, as Alan Goldman claims, sexual partners, by recognizing each other’s ‘subjectivity’, can satisfy Kant’s ‘second formulation’:

> Even in an act which by its nature ‘objectifies’ the other, one recognizes a partner as a subject with demands and desires by yielding to those desires, by allowing oneself to be a sexual object as well, by giving pleasure or ensuring that the pleasures of the act are mutual.
>
> (1977: 87)

It follows that neither marriage nor heterosexuality nor love are necessary for salvaging sex, and the acquisition of

the right of bodily access can be temporary and reversible.

The Vatican. Although the Church embraces natural law ethics, it has also invoked Kant. In his 1968 encyclical 'Humanae Vitae', Paul VI argued against the permissibility of the use of contraceptive devices. 'Each and every marriage act…must remain open to the transmission of life' (sect. 11), and therefore 'conjugal acts made intentionally infecund' (sect. 14) are immoral. In Kantian style, he warned that the use of contraceptive devices makes husbands 'lose respect' for their wives. Since using contraception implies the act is primarily for pleasure, the husband sees his wife as 'a mere instrument of selfish enjoyment' (sect. 17). Karol Wojtyla (later John Paul II) has expressed a similar thought:

> When the idea that 'I may become a father'/'I may become a mother' is totally rejected in the mind and will of husband and wife nothing is left of the marital relationship, objectively speaking, except mere sexual enjoyment. One person becomes an object of use for another person.
>
> (1960: 239)

Rejecting the libertarian view that consent is sufficient, John Paul also asserted that only love eliminates the sexual use of one person by another, since love is a unification of persons achieved through the mutual gift - rather than a Kantian contractual exchange - of their selves. Marriage might by its nature be indissoluble (irreversible) if, once a person makes a gift of self to another, that gift cannot, for logical reasons, be taken back or returned.

Conservative sexual philosophy. John Finnis (1993: 12) argues that there are morally worthless sexual acts in which 'one's body is treated as instrumental for the securing of the experiential satisfaction of the conscious self'. In masturbating or in being sodomized, the body is just a tool of satisfaction and, as a result, one undergoes 'disintegration'. 'One's choosing self [becomes] the quasi-slave of the experiencing self which is demanding gratification.' As in Kant, the worthlessness and disintegration attaching to masturbation and sodomy attach to 'all extramarital sexual gratification'. This is because only in married, heterosexual coitus do 'their reproductive organs…make them a biological (and therefore a personal) unit'. Finnis begins with the Kantian intuition that sexual activity involves treating the body instrumentally, and he concludes with the Kantian intuition that sex in marriage avoids disintegrity since the couple is a unit: 'the orgasmic union of the reproductive organs of husband and wife really unites them biologically'.

8 Consent and coercion

Sexual activity between an adult and a child - 'intergenerational' sex - is considered morally wrong in many cultures. The arguments against it claim that it is physically and psychologically harmful to the child and violates canons of consent. Children are not able to consent to sex in the same way they are not able to consent to surgery, or to what school they attend; in these areas parents are empowered to decide for them. Defenders of intergenerational sex reply that depending on the age of the child and the acts performed, sex is not physically harmful, and it is psychologically harmful only to the extent that society condemns it. In addition, children are able to consent to sex in the same way they are able to give consent, or withhold it, to what they eat, wear, or the films they see. Thus whether children are able to give meaningful consent to sex depends on its significance - whether sex is more like eating pizza or contributing to the 'grand design'. (Analogous views are possible about sex between humans and animals: Is it harmful to the nonhuman creatures? Can they consent? Does it matter?) Intergenerational sex thus suggests separate and pressing issues about the nature of consent.

In the 1920s, Bertrand Russell admonished: 'the intrusion of the economic motive into sex is always…disastrous. Sexual relations should be a mutual delight, entered into solely from the spontaneous impulse of both parties.' In employing this principle, Russell found prostitution lacking. He also observed, prophetically, that it applied elsewhere, to married women who have economic reasons for acquiescing to the sexual demands of their husbands (1929: 121-2; see Engels 1884). Russell was voicing something close to Robin Morgan's feminist definition of rape:

> How many millions of times have women had sex 'willingly' with men they didn't want to have sex with? … How many times have women wished just to sleep instead or read or watch the Late Show? … Most of the decently married bedrooms across America are settings for nightly rape.
>
> (1977: 166)

The fundamental idea is that genuinely consensual participation in sex, without a hint of coercion, requires substantial economic, social and psychological equality between the persons involved. A society characterized by both poverty and wealth is one in which people are exposed to economic coercion, even if it appears that they freely consent to what they do. If some groups have less economic and social power than others, their members will be exposed to sexual coercion, among other kinds. Women's consent to heterosexual sexual activity might be largely chimerical also in virtue of systematic male dominance. If women's consent to heterosexual acts within an oppressive patriarchal society is not genuine, this will be especially true of prostitution, which can be seen as rape perpetrated by men employing economic or psychological power rather than physical assault or the threat of a knife.

An extreme view is that the presence of any pressure is coercive and morally objectionable. In 'Nonviolent Sexual Coercion' (1991), Charlene Muehlenhard and Jennifer Schrag list, among other things, 'status coercion' (women are coerced into sex or marriage by a man's occupation) and 'discrimination against lesbians' (which compels women into sexual relationships with men) as objectionable forms of coercion. Depending on the kind of case we have in mind, it might be more accurate to say either that some pressures are not coercive (a husband who constantly nags for sex) or that some pressures are coercive but not morally objectionable (a woman who marries for economic reasons). This last point might hold as well for sexual objectification: is it all wrong? Why is sexual objectification condemned, when other objectification is not?

Other questions about consent to sex are analogous to those that arise in medicine, business and law (see Medical ethics, §3). How *specific* must consent be in order that a person engage voluntarily in subsequent sexual behaviour? Because consenting is opaque, when one person agrees 'to have sex' with another, the first has not necessarily consented to any sensual caress or coital position the second has in mind. How *explicit* must consent be in order to be an indication of voluntary choice? Whether consent can be implied reliably by nonverbal behaviour, and whether anyone should take nonverbal cues as showing decisively that another has consented to sex, are questions that arise in discussions of acquaintance rape. And how *informed* must consent be? Do sexual partners have a duty to reveal their marital or immunological status? Conventional wisdom thinks so, but there is dissent (see Mayo 1996).

9 Utilitarianism

It seems likely that the John Stuart Mill who wrote *On Liberty* would have defended libertarian sexual ethics on general utilitarian grounds, insisting that the consent of the participating parties is morally sufficient as long as significant harm to third parties is avoided (see Utilitarianism). Do as you will in your experiments in living, he might have said; neither the law nor public opinion should be brought to bear unless, say, violations of assignable duties occur. Kantians and Thomists agree that the interests of third parties and society might be adversely affected by sexual activity, even if the sexual act occurring between two persons embodies genuine respect for each or is fully natural.

The difficult question concerns which of the ubiquitous effects on third parties or society are significantly harmful. Is a person harmed by becoming nauseous when noticing two homosexuals (or heterosexuals) kiss in public? Is a spouse harmed by the infidelity about which he or she knows nothing? Is someone harmed merely by knowing that immoral or offensive conduct is occurring? This area of social philosophy has been especially contentious, since decisions about 'harm' have profound implications. A narrow notion of harm yields a permissive (sexual and nonsexual) ethics and provides little justification for using the criminal law to interfere with sexual behaviour; a broad notion of harm implies the opposite. Mill and James Fitzjames Stephen squared off against each other over this issue in the nineteenth century; in the twentieth century, the debate was revived, initially by Patrick Devlin and H.L.A. Hart (see Law and morality §§1-3; Law, limits of). Utilitarian sexual ethics can in any event be restrictive *or* permissive, depending on the truth of empirical assertions about the consequences of sexual behaviour, or about the consequences of trying to prevent it. In addition to depending on contestable notions of wellbeing, the empirical claims underlying utilitarian judgments are difficult to verify. Similar problems plague claims that the existence of pornography helps maintain a social environment inimical to the wellbeing of women (see Pornography).

10 Sadomasochism and love

Sadomasochism has long been scorned by moralists. Nonlibertarian Kantians think that sadomasochistic sexual activity is wrong because it violates the personhood of the participants; humiliating and demeaning attitudes retain their character even when the sexual acts are consensual. A utilitarian might conclude that this activity, or legally permitting it, does more harm than good (as in some utilitarian arguments against permitting active euthanasia). And for a moral paternalist, it is wrong for one person to harm another, even if the latter is willing, or for persons to allow themselves to be harmed. For other utilitarians and Kantians, however, sadomasochism can be engaged in lovingly, or while fully respecting one's sexual partner, and without significant harm being done to nonparticipants.

Sadomasochism has been gaining in visibility, especially in homosexual culture. The radical libertarian sons of Kant have even given birth to the daughters of de Sade. Thus the debate over sadomasochism is especially heated among feminists, for whom there are strong partisan differences in the moral and political assessment of both heterosexual and homosexual sadomasochism (see Feminist ethics §3). About lesbian sadomasochism in particular, Claudia Card says, on the positive side, that its 'participants generally wish each other well and respect each other's choices'. But Card also worries about such activities: 'the only things distinguishing the behavior of [a top] from battery and other abuse may be the motivations of the parties and the consent of the [bottom]' (1995: 221) - and, for Card, neither the consent nor the motivation (for example, sexual pleasure) seem to turn battery and abuse into anything other or better than battery and abuse. But what more could we want in distinguishing respectful sadomasochism from battery, other than the consent to experience mutual pleasure? To say that consent is not enough, that the persons must be married, committed to each other, 'in love', or must avoid uncommonly repugnant acts, is to jeopardize all unconventional sex, not just sadomasochism. The task for a defender of this position is to deny that marriage is essential for legitimate sexuality and to bless a great deal of human sexual interaction on the basis of consent (such as homosexuality), without adopting a libertarianism so thin that it condones sadomasochism. There are well-founded doubts that that project can be successfully completed.

Unwoolly accounts of the ingredient that is required beyond consent are hard to come by, but the idea that something is required has often recommended itself:

> Avoiding deceit and coercion are only the core of treating others as persons in sexual relationships. In avoiding these we avoid…obvious ways of using as (mere) means. But to treat another as a person in an intimate, and especially an intimate sexual, relationship requires far more. These further requirements reflect the intimacy rather than the specifically sexual character of a relationship. However, if sexual relationships cannot easily be merely relationships between consenting adults, further requirements for treating another as a person are likely to arise in any sexual relationship. Intimate bodies cannot easily have separate lives.
>
> (O'Neill 1985: 269-70)

If the quest for sexual pleasure threatens the self's wholeness (Finnis), the other's personhood (John Paul II), and society's viability (Devlin), if it insults the human spirit (Augustine), the natural order of things, and God (Aquinas), then love - be it a dose of *agapē* or *caritas* (an uplifting sort of *eros*), or the trendy 'intimacy' - is necessary, and perhaps (as in some versions of romanticism) also sufficient, for sexual relations to be licit.

The thought that love is the magical ingredient has immediate application to homosexuality: gay and lesbian sex can be justified within loving, monogamous homosexual marriages. Whether sadomasochism can genuinely express or involve love seems to be an empirical question, as long as we have a workable operational definition of 'love'. What kind of domestic arrangement do these sex partners have - is it much different, aside from the sadomasochistic sex, from Ozzie and Harriet's? Furthermore, love as the extra ingredient makes defending the use of contraception easy: sex acts, including the nonprocreative, express and bolster the love spouses have for each other. Given these rhetorical benefits, it might not be surprising that natural science cements the link between sex and love: 'Nature has emotionally enriched the human reproductive impulse through love, and in doing so she has immensely increased our enjoyment of both' (Walsh 1994: 371). In one biological swoop, a perplexing question has been answered, namely, 'why species survival, the means of impregnation, and emotional/erotic relationships should ever have become so rigidly identified with each other' (Rich 1986: 35; see Hume 1739/40: section 11). If you prefer God to nature, try Paul VI's answer: there is an 'inseparable connection, established by God, which man on his own initiative may not break, between the unitive significance and the procreative significance which are both inherent to the marriage act' (1968: §12).

See also: Family, ethics and the; Friendship; Genetics and ethics; Kantian ethics; Love; Morality and emotions; Reproduction and ethics

ALAN SOBLE

References and further reading

The most challenging items on this list are the books by Boswell, Foucault, Halperin, Hartsock, MacKinnon, Price and Scruton, plus some of the essays in the collections edited by Murphy, Soble, Stein and Stewart.

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, ed. T. Gilby, Cambridge: Blackfriars, 1968, vol. 43, IIaIIae.153-4.(Classic natural law sexual ethics.)

Atkinson, T-G. (1974) *Amazon Odyssey*, New York: Links.(A famous piece of radical feminism that condemns the political 'institution' of heterosexuality.)

Augustine (419-21) *De nuptiis et concupiscentia (On Marriage and Concupiscence)*, in *Works*, vol. 12, Edinburgh: T. & T. Clark, 1874.(Early Christian sexual ethics and metaphysics: concupiscence is sinful; sexual intercourse is forgiven in marriage; children thereby produced contract original sin. See also *The City of God* [vols 1-2] and *Confessions* [vol. 14].)

Baker, R. and Elliston, F. (eds) (1975) *Philosophy and Sex*, B uffalo, NY: Prometheus Books; 2nd edn, 1984, editorial introduction, 11-36.(Introductory textbook, which includes papers on marriage, abortion, contraception, pornography, homosexuality and paedophilia, and contains an enormous bibliography compiled by W. Vitek.)

Boswell, J. (1980) *Christianity, Social Tolerance and Homosexuality*, Chicago, IL: University of Chicago Press. (A book of history rich in intelligent philosophical discussion of the definition of homosexuality, social constructionism, and Aquinas' natural law.)

Card, C. (1995) *Lesbian Choices*, New York: Columbia University Press.(Discusses lesbian ethics, culture, friendship and sexuality.)

Davis, M. (1983) *Smut: Erotic Reality/Obscene Ideology*, Chicago, IL: University of Chicago Press.(Insightfully divides the metaphysical terrain into Jehovanist, Naturalist and Gnostic philosophies of sex.)

Engels, F. (1884) *Der Ursprung der Familie, des Privateigentums und des Staats*, trans. *The Origin of the Family, Private Property and the State*, New York: International Publishers, 1972.(The Marxist alternative to Augustine's theory of original sin.)

Finnis, J.M. (1994) 'Law, Morality, and "Sexual Orientation"', *Notre Dame Law Review* 69 (5): 1049-76.(Lays out a Catholic case against the toleration of homosexuality by the law; criticizes M. Nussbaum's understanding of Greek homosexuality. Occasionally woolly.)

Finnis, J.M. and Nussbaum, M. (1993) 'Is Homosexual Conduct Wrong? A Philosophical Exchange', *The New Republic* 15 November: 12-13.(Finnis presents a Kantian-Catholic argument for the moral inferiority of homosexuality; Nussbaum finds the argument unconvincing.)

Foucault, M. (1978, 1985, 1986) *The History of Sexuality*, vol. 1, *An Introduction*, New York: Vintage Books; vol. 2, *The Use of Pleasure*, New York: Pantheon; vol. 3, *The Care of the Self*, New York: Vintage Books. (Provocative ruminations on sexuality and its history by the purported instigator of the academic furore over social constructionism.)

Gardell, M.A. (1987) 'Sexual Ethics: Some Perspectives From the History of Philosophy', in E.E. Shelp (ed.) *Sexuality and Medicine*, Dordrecht: D. Reidel, vol. 2, 3-15.(A brief introduction to the history of sexual ethics.)

Goldman, A. (1977) 'Plain Sex', *Philosophy and Public Affairs* 6 (3): 267-87; repr. in A. Soble (ed.) *The Philosophy of Sex*, Savage, MD, Rowman and Littlefield, 2nd edn, 1991, 73-92.(This straightforward, sensible paper combines conceptual analysis of 'sexual activity' and 'sexual desire' with Kantian, libertarian sexual ethics.)

Gray, R. (1978) 'Sex and Sexual Perversion', *Journal of Philosophy* 75 (4): 189-99; repr. in A. Soble (ed.) *The Philosophy of Sex*, Totowa, NJ: Rowman and Littlefield, 1980, 158-68.(Conceptual analysis of 'sexual perversion' and 'sexual act'.)

Gudorf, C.E. (1994) *Body, Sex and Pleasure: Reconstructing Christian Sexual Ethics*, Cleveland: Pilgrim Press. (Sexual pleasure is a gift from God; the fundamental ethical principle governing sexual relations is 'mutuality'.)

Halperin, D. (1990) *One Hundred Years of Homosexuality*, New York: Routledge.(Scholarly treatment of homosexuality, social constructionism, and Plato's *Symposium*.)

Hartsock, N.C.M. (1983) *Money, Sex and Power: Toward a Feminist Historical Materialism*, New York:

Longman.(Develops a feminist-Marxist theory; ch. 7, ‘Gender and Power’, discusses sexuality.)
Herman, B. (1993) ‘Could it Be Worth Thinking About Kant on Sex and Marriage?’ in L.M. Antony and C. Witt (eds) *A Mind of One’s Own*, Boulder, CA: Westview, 49-67.(Taking the answer to this question to be ‘Yes, it could’, this paper explores intriguing similarities between Kant’s sexual philosophy and that of C. MacKinnon.)
Hume, D. (1739/40) ‘Of the Passions’, in *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1986, book 2, part II.(Focuses on love and hate; section 11 discusses sexual desire as a ‘compound passion’.)
Jung, P. and Smith, R. (1993) *Heterosexism: An Ethical Challenge*, Albany, NY: State University of New York Press.(Argues that an unbiased, that is, nonheterosexist, reading of Scripture points in the direction of Christian homosexual marriage.)
Kant, I. (1780-81) *Lectures on Ethics*, trans. L. Infield, New York: Harper & Row, 1963.(Traditional sexual metaphysics and ethics from someone whose accomplishments in other areas boggle the mind.)
Kierkegaard, S.A. (1845) *Stadier Paa livets Vej*, trans. W. Lowrie, *Stages on Life’s Way*, Princeton, NJ: Princeton University Press, 1945.(Includes ‘In Vino Veritas’ (‘The Banquet’) (25-93), Kierkegaard’s [or William Afham’s] splendid variation on the *Symposium* theme. To be read right after *Either/Or*.)
Levin, M. (1984) ‘Why Homosexuality is Abnormal’, *The Monist* 67 (2): 251-83.(Builds a secular case for the nonmoral inferiority of homosexuality.)
Linden, R.R., Pagano, D., Russell, D.E.H. and Star, S.L. (eds) (1982) *Against Sadomasochism: A Radical Feminist Analysis*, East Palo Alto, CA: Frog in the Well.(Collection of essays aimed at refuting the defence of lesbian sadomasochism made in the work edited by *SAMOIS*, *Coming to Power*.)
MacKinnon, C.A. (1989) *Toward a Feminist Theory of the State*, Cambridge, MA: Harvard University Press.(The ‘unmodified feminist’ legal and social analysis of heterosexuality, rape and pornography.)
Mayo, D.J. (1997) ‘An Obligation to Warn of HIV Infection?’, in A. Soble (ed.) *Sex, Love and Friendship*, Amsterdam: Editions Rodopi, 447-53.(Argues that HIV-positive persons do not have a moral duty to warn sexual partners of the health hazards of contemplated sexual acts.)
Morgan, R. (1977) *Going Too Far: The Personal Chronicle of a Feminist*, New York: Random House.(Her collected *oeuvre*; includes ‘Theory and Practice: Pornography and Rape’.)
Muehlenhard, C.L. and Schrag, J.L. (1991) ‘Nonviolent Sexual Coercion’, in A. Parrot and L. Bechhofer (eds) *Acquaintance Rape: The Hidden Crime*, New York: John Wiley, 115-28.(A catalogue of ways women are [or might be] coerced into sexual activity without being forced violently.)
Murphy, T.F. (1987) ‘Homosexuality and Nature: Happiness and the Law at Stake’, *Journal of Applied Philosophy* 4 (2): 195-204.(A thorough critique of M. Levin’s ‘Why Homosexuality is Abnormal’; the essay focuses on the arguments from nature and happiness and rebuts their purported legal implications.)
Murphy, T.F. (ed.) (1994) *Gay Ethics: Controversies in Outing, Civil Rights and Sexual Science*, New York: Haworth Press.(Introductory-to-advanced collection of essays on such issues as outing, gay marriage, gays in the military, affirmative action, and the science of sexual orientation.)
Nagel, T. (1969) ‘Sexual Perversion’, *Journal of Philosophy* 66 (1): 5-17; revised version repr. in *Mortal Questions*, Cambridge: Cambridge University Press, 1979, 39-52.(Sophisticated analysis of ‘sexual perversion’ that draws on Sartre and Grice in formulating an account of natural human sexual psychology.)
Noonan, J.T. (1965) *Contraception: A History of Its Treatment by the Catholic Theologians and Canonists*, Cambridge, MA: Harvard University Press, enlarged edn, 1986.(Excellent history of the philosophical and theological arguments surrounding the use of methods to prevent conception during heterosexual intercourse.)
Nussbaum, M. (1995) ‘Objectification’, *Philosophy & Public Affairs* 24 (4): 249-91.(Analyses seven distinct types of sexual objectification; some are benign, others malignant. Occasionally pornographic.)
O’Neill, O. (1985) ‘Between Consenting Adults’, *Philosophy & Public Affairs* 14 (3): 252-77.(An essay on what it is - and what it is not - to treat someone as a person, with respect; the account is applied to sexual relations.)
Padgug, R. (1979) ‘Sexual Matters: On Conceptualizing Sexuality in History’, *Radical History Review* 20 (Spring/Summer): 3-23; repr. in E. Stein (ed.) *Forms of Desire*, New York: Routledge, 2nd edn, 1992, 43-67. (Social constructionism with a Marxist gloss: hunger is hunger, but not all hunger is the same, as Marx put it in the *Grundrisse*.)
Paul VI (1968) ‘Humanae Vitae’, *Catholic Mind* 66 (September): 35-48; repr. in R. Baker and F. Elliston (eds) *Philosophy and Sex*, Buffalo, NY: Prometheus Books, 2nd edn, 1984, 167-83.(This papal encyclical is Paul VI’s post-Vatican II affirmation of the doctrine that the use of contraception is morally illicit.)
Pineau, L. (1989) ‘Date Rape: A Feminist Analysis’, *Law and Philosophy* 8: 217-43.(A lucid discussion of what it is for one person to have a *reasonable belief* that another person has consented to engage in sexual activity.)

Plato (*c*.386-380 BC) *Symposium*, trans. R. Waterfield, Oxford: Oxford University Press, 1994.(Plato's brilliant, moving, sometimes hilarious, and endlessly discussable account of *eros*.)

Price, A.W. (1989) *Love and Friendship in Plato and Aristotle*, Oxford: Clarendon Press.(Of the vast writings on love, sex and friendship in Plato and Aristotle, this absorbing treatise is one of the best places to start.)

Primoratz, I. (1993) 'What's Wrong With Prostitution?' *Philosophy* 68 (264): 159-82.(The answer according to this essay is 'Not much'. An uncompromising libertarian view of prostitution, which rebuts feminist objections.)

Rich, A. (1986) 'Compulsory Heterosexuality and Lesbian Existence', in *Blood, Bread and Poetry*, New York: W.W. Norton, 23-75.(A classic item of radical feminist sexual philosophy.)

Ruse, M. (1988) *Homosexuality: A Philosophical Inquiry*, New York: Blackwell.(A comprehensive philosophical treatment of ethical and scientific questions about homosexuality.)

Russell, B.A.W. (1929) *Marriage and Morals*, London: Allen & Unwin.(A little book of liberal common sense that caused Russell untold trouble.)

SAMOIS (ed.) (1981) *Coming to Power*, Palo Alto, CA: Up Press; 2nd edn, Boston, MA: Alyson Publications, 1982.(A brave collection of essays defending lesbian sadomasochism; especially noteworthy is G. Rubin's 'The Leather Menace'. *SAMOIS* is the name of a lesbian/feminist sado-masochism collective.)

Schopenhauer, A. (1844) *Die Welt als Wille und Vorstellung*, vol. 2, Leipzig: Brockhaus; trans. E.F.J. Payne, *The World as Will and Representation*, vol. 2, Indian Hills, CO: Falcon's Wing Press, 1958.('The Metaphysics of Sexual Love' appears as ch. 44 in the supplements to book 4. A cynical - or just hard-nosed - naturalistic view of human sexuality in which we are mere pawns in a larger game and our personal dreams and intentions are illusions.)

Scruton, R. (1986) *Sexual Desire: A Moral Philosophy of the Erotic*, New York: Free Press.(A conservative theory of sex thick with Kantian themes; Scruton's favourite notion is 'individualising intentionality' (see 384). Readers are also recommended to see M. Nussbaum's tart review of this essay, 'Sex in the Head', *New York Review of Books*, 18 December 1986, 49-52.)

Shaffer, J. (1978) 'Sexual Desire', *Journal of Philosophy* 75 (4): 175-89; repr. in A. Soble (ed.) *Sex, Love and Friendship*, Amsterdam: Editions Rodopi, 1997, 1-12.(Sophisticated conceptual analysis of 'sexual desire'; a gem of an example of what the philosophy of sex - or philosophy - can be.)

Singer, I. (1984-7) *The Nature of Love*, Chicago, IL: University of Chicago Press, 3 vols.(A magnificent history of the philosophy of love and sex, from Plato right through twentieth-century psychology.)

Soble, A. (ed.) (1980) *The Philosophy of Sex*, Totowa, NJ: Rowman and Littlefield; 2nd edn, Savage, MD: Rowman and Littlefield, 1991.(Introductory-to-advanced textbook; topics covered in the 2nd edn include pornography, adultery, prostitution, rape, homosexuality, sadomasochism and perversion.)

Soble, A. (ed.) (1989) *Eros, Agapē and Philia*, New York: Paragon House.(Introductory-to-advanced textbook, a mixture of historical [Plato, Aristotle, Montaigne, Descartes, Pascal] and contemporary pieces [G. Vlastos, L. Kosman, D. Hamlyn, S. Mendus, W. Newton-Smith].)

Soble, A. (1996) *Sexual Investigations*, New York: New York University Press.(A work of sexual ethics, politics and conceptual analysis.)

Soble, A. (ed.) (1997) *Sex, Love and Friendship*, Amsterdam: Editions Rodopi.(Introductory-to-advanced collection of sixty contemporary essays that discuss Plato, Aristotle, Spinoza, Kant, Sade, Kierkegaard, Heidegger, and diverse topics, including sexual objectification, abortion, sexual perversion, love and sex, pornography and gay rights.)

Stein, E. (ed.) (1990) *Forms of Desire*, New York: Garland; 2nd edn, New York: Routledge, 1992.(Advanced collection of essays explaining, defending and criticizing social constructionism.)

Stewart, R.M. (ed.) (1995) *Philosophical Perspectives on Sex and Love*, New York: Oxford University Press. (Comprehensive advanced textbook on sex, love [of various kinds] and friendship; sexual topics discussed include pornography, prostitution, sexual objectification, perversion, homosexuality, sadomasochism, rape and adultery.)

Sullivan, A. (1995) *Virtually Normal: An Argument About Homosexuality*, New York: Alfred A. Knopf.(An iconoclastic neoconservative political philosophy applied to questions revolving around homosexuality; Sullivan argues that the linchpin of gay liberation is access to civil marriage.)

Vannoy, R.C. (1980) *Sex Without Love: A Philosophical Exploration*, Buffalo, NY: Prometheus Books.(Argues that sexual activity is morally and nonmorally sound without being attached to love; provides conceptual analyses of 'sexual perversion' and 'love'.)

Verene, D. (ed.) (1972) *Sexual Love and Western Morality*, New York: Harper & Row; 2nd edn, Boston, MA: Jones and Bartlett, 1995.(Introductory textbook; 2nd edn includes Plato, Aristotle, Augustine, Aquinas, Luther,

Kant, Schopenhauer, Nietzsche, Hegel, Sartre, Freud, Sade and Foucault.)

Walsh, A. (1994) 'Love and Sex', in V.L. Bullough and B. Bullough (eds) *Human Sexuality: An Encyclopedia*, New York: Garland, 369-73.(An introduction, with an evolutionary bent, to the relationship between love and sex.)

Weithman, P. (1997) 'Natural Law, Morality, and Sexual Complementarity', in D. Estlund and M. Nussbaum (eds) *Sex, Preference, and Family: Essays on Law and Nature*, New York: Oxford University Press, 227-46.(A meticulous explication, analysis and critique of John Finnis' sexual ethics.)

Wertheimer, A. (1996) 'Consent and Sexual Relations', *Legal Theory* 2 (1): 89-112.(Sharp, insightful conceptual analysis of the moral and legal force of 'consent' in, especially, sexual contexts.)

Wojtyla, K. (later John Paul II) (1960) *Mitość I Odpwiedzialność*, Krakow: Wydawnicto, Znak; trans. H.T. Willetts, *Love and Responsibility*, New York: Farrar, Strauss & Giroux, 1981.(A Kantian-Catholic sexual ethics, written before Wojtyla, a professor of philosophy, was elected pontiff; his *official* view is expressed in, for example, 'Evangelium Vitae', *Origins* 24 (42): 689-727, 1995.)

Shaftesbury, Third Earl of (Anthony Ashley Cooper) (1671-1713)

Shaftesbury, whose influence on eighteenth-century thought was enormous, was the last great representative of the Platonic tradition in England. He argued that by natural reason we can see that the world is an intelligible, harmonious system. In reflecting on our character traits we will inevitably approve of those which contribute to the good of humanity and of the whole system. These same personal qualities are also needed for a happy life, so virtue and happiness go hand in hand.

Shaftesbury is often seen as the founder of the moral sense or 'sentimentalist' school in ethics, whose members held that morality was based on human feeling rather than on reason. Although leading sentimentalists, such as Hutcheson and Hume, made use of many of his ideas, Shaftesbury himself has more in common with the rationalists, who held that there are eternal moral truths which we can know by the use of reason.

1 Life and influences

Anthony Ashley Cooper, Third Earl of Shaftesbury, was born in February 1671. Superintendence of his education was handed over to no less a personage than John Locke, with whom Shaftesbury seems to have had an ambivalent relationship. While honouring him as his 'friend and foster father', Shaftesbury came not only to reject, but even to detest, much of Locke's philosophy. He played a full part in English political life, first in the Commons and then in the Lords, though ill health forced occasional retreats to the Continent. He devoted the latter portion of his life to writing; his published work was eventually brought together in his three-volume *Characteristics of Men, Manners, Opinions, Times* (1711), a work very widely read in the eighteenth century.

Shaftesbury was deeply influenced by Greek and Roman thought, in which he distinguished two strands. The first presupposes that we live in an ordered and intelligible universe; the second that everything came into being by chance and that nothing has any meaning. It is to the former strand that Shaftesbury owes allegiance. It proceeds through Plato, Aristotle, the Neoplatonists and Platonized Christianity to the Cambridge Platonists of the seventeenth century (see Cambridge Platonism; Neoplatonism). Shaftesbury's notebooks reveal in particular the marked influence of the Stoics, especially Epictetus and Marcus Aurelius (see Stoicism).

2 Teleology

Shaftesbury's conception of the world is thoroughly teleological. By the use of unaided natural reason, we can come to see that the universe forms a well-ordered, intelligible, harmonious system, in which humans have their proper place. To understand the nature of anything, including ourselves, is to know what functions are natural and proper to it, and which ones are unnatural and perverted. This knowledge is practical as well as theoretical. In the Platonic tradition to which Shaftesbury belongs, to know the good is to love it. Shaftesbury is thus an apostle of intellectually disciplined enthusiasm; rational beings cannot help but be moved once they are aware of this universal harmony, which Shaftesbury celebrates in glowing and lyrical terms. To live in accordance with nature, to have one's inner harmony attuned to the outer, is to be both virtuous and happy. Virtue is its own reward; the virtuous life is in itself the most fulfilling there is.

The passions that disturb our lives stem from false values, from a wrong assessment of what is good and bad. Once we see that our good is to be found in the part we play in the good of the whole then we can face hardship and even death with equanimity. For we find our true good not in externals, such as wealth or reputation, which can be taken from us, but in a balanced and harmonious mind, which cannot be overthrown by anything except our folly. Moreover, the goods of the mind are superior to any goods we find in nature or in human art. As Shaftesbury explains in *The Moralists, A Philosophical Rhapsody* (1709), mind is not only the source of harmony, beauty and goodness (which for him are essentially one and the same), but is itself better and more beautiful than its products. To value any external good more highly than inner harmony is to prefer a secondary beauty to its primary source, to prefer the shadow to the reality.

Shaftesbury stresses the immanence of the divine mind rather than its transcendence. Shaftesbury often pictures God's relation to the world as more like that of the soul to the body than it is like that of the watchmaker to his watch. We see this in *The Moralists* (part III, section 1) where we find an unusual variant of the argument from design which starts from the issue of personal identity (see God, arguments for the existence of §§4-5). Among all

the changes to both body and mind which any person undergoes in the course of life, it is undeniable that there is something which unites these stages and makes them all properties of the same person. We cannot give a coherent account of the metaphysical status of this self, but that there is such a continuing self which governs and superintends one's life is undeniable. Since nature is also self-regulating and orderly, it too must be governed by a mind, which bears the same relation to the world and its activity as our mind does to our body and actions, even if we can give no clear indication of its metaphysical status.

In spite of his token orthodoxy Shaftesbury is a true parent of the Enlightenment in his willingness to question authority and in his advocacy of liberty. In thought and discussion there should be freedom of speech (at least among educated gentlemen!). No subject is too grave that it may not be subject to jest; for raillery, Shaftesbury contends, is a test of truth. Error cannot withstand mockery. Art and culture only flourish in conditions of freedom. Genuine patriotism can only spring from an identification with the body politic which cannot be found in tyrannies.

3 Ethics

The core of Shaftesbury's philosophy, however, is his moral theory, which is most fully and systematically expounded in his *Inquiry Concerning Virtue or Merit* (1699), his earliest work. In the first part he delineates what goodness and virtue are. Every creature, and every species, is part of a wider system to which it contributes. Each system is itself part of a wider system, until we reach the universe, the complete system which incorporates all the subsystems. Each individual (or species) is to be judged by its contribution to the good of its system and, ultimately, by its contribution to the good of the whole.

In judging an individual creature we are concerned with its character, with what Shaftesbury calls its affections - its, desires, motives and enjoyments. A good creature is one whose affections will, in the normal course of events, cause it to act in a way that will be for the good both of itself and of the system of which it is a part. While animals can be good or bad, humans alone can be virtuous or vicious, for humans alone are self-conscious and capable of reflecting on their own actions and affections so that these can in their turn become the object of an affection, this time of approval or disapproval. We are naturally and ineluctably led to approve of what is natural and honest and to disapprove of what is dishonest and corrupt. Our capacity for reflection, however, enables us to be self-governing, and thus virtuous or vicious, for we can choose whether to indulge our good or our bad inclinations (see Autonomy, ethical).

We might note that Shaftesbury departs from the Stoic model in one respect. For those in the Socratic tradition the four cardinal virtues are wisdom, courage, temperance and justice. Benevolence, pity or compassion receive less emphasis. Epictetus concedes that we should behave sympathetically towards the unfortunate; but we should not disturb our stoic tranquillity by pitying them. Shaftesbury, by contrast, puts the social virtues of love, friendship and universal benevolence at the forefront of his theory (see Virtues and vices §1).

Shaftesbury proceeds in the second half to ask what reason we have to be virtuous. He assumes without argument that the only reason we can offer is that it is in our interests - a virtuous life will be a happy one. He is thus a rational egoist, for he holds that questions of the rational justification of a way of life must appeal to self-interest. But he is not a psychological egoist, for he holds, as against Hobbes, that agents have altruistic as well as egoistic motives (see Hobbes, T. §8). Nor is he an ethical egoist, for morality requires us to be motivated by concern for others (see Egoism and altruism §1).

Shaftesbury then argues, ingeniously if not always persuasively, that the qualities of character needed to live a good life are the very same as those needed to live a virtuous one. In particular, proper development of the social affections, which are directed to the good of others, gives a degree and type of satisfaction not otherwise available.

It is customary to view Shaftesbury as the source or founder of the moral sense or sentimentalist school of ethics, whose foremost members were Hutcheson and Hume (see Moral sense theories). It is true that they were much influenced by him, but in the great eighteenth-century debate as to whether morality is founded on reason or on sentiment, Shaftesbury should, I think, be counted on the rationalist side in so far as he holds that there are eternal moral truths, existing independently of us, and revealed to us by the use of reason.

DAVID McNAUGHTON

List of works

Shaftesbury (1995) *The Shaftesbury Collection*, Bristol: Thoemmes, 4 vols. (Comprehensive recent collection.)

Shaftesbury (1699) *An Inquiry Concerning Virtue or Merit*, ed. D. Walford, Manchester: Manchester University Press, 1977. (In the first part of his major work on ethics Shaftesbury discusses what virtue is, and in the second part he argues that it is in our interest to be virtuous.)

Shaftesbury (1709) *The Moralists, A Philosophical Rhapsody*, in *Characteristics of Men, Manners, Opinions, Times*, vol. 2, ed. J.M. Robertson, London: Grant Richards, 1900; repr. in *The Shaftesbury Collection*, vol. 2, Bristol: Thoemmes, 1995. (A philosophical dialogue, in which the claim that the world is the product of an intelligent and beneficent Mind is examined and defended.)

Shaftesbury (1711) *Characteristics of Men, Manners, Opinions, Times*, ed. J.M. Robertson, London: Grant Richards, 1900, 2 vols; repr. in *The Shaftesbury Collection*, vols 1 and 2, Bristol: Thoemmes, 1995. (Contains all important works published by Shaftesbury in his lifetime.)

Shaftesbury (1900) *The Life, Unpublished Letters, and Philosophical Regimen of Anthony, Earl of Shaftesbury*, ed. B. Rand, London: Swan Sonnenschein; repr. in *The Shaftesbury Collection*, vol. 3, Bristol: Thoemmes, 1995. (Contains Shaftesbury's important personal notebooks and a fairly full selection of his letters.)

Shaftesbury (1914) *Second Characters, or The Language of Forms*, ed. B. Rand, Cambridge: Cambridge University Press; repr. in *The Shaftesbury Collection*, vol. 4, Bristol: Thoemmes, 1995. (Contains Shaftesbury's uncompleted proposed work on aesthetics, including his long and fascinating instructions to the artist on the composition of 'The Judgement of Hercules', which he commissioned while in Italy.)

References and further reading

Darwall, S. (1995) *The British Moralists and the Internal 'Ought'*, Cambridge: Cambridge University Press, 176-206. (A lucid and authoritative discussion of Shaftesbury's ethics in relation to his predecessors.)

Fowler, T. (1882) *Shaftesbury and Hutcheson*, London: Sampson Low, Marston, Searle, & Rivington, 1-167. (A clear summary of Shaftesbury's main views with numerous quotations.)

Grean, S. (1967) *Shaftesbury's Philosophy of Religion and Ethics*, Athens, OH: Ohio University Press. (Discusses whole of Shaftesbury's work with particular emphasis on his advocacy of enthusiasm.)

Klein, L. (1994) *Shaftesbury and the Culture of Politeness*, Cambridge: Cambridge University Press. (Examines political context of Shaftesbury's writings in some detail and places emphasis on rhetorical structure of his writings.)

McNaughton, D. (1996) 'British Moralists of the Eighteenth Century: Shaftesbury, Butler and Price', in S. Brown (ed.) *Routledge History of Philosophy*, vol. 5, *The British Enlightenment*, London: Routledge, 203-27. (Expansion of the material in this entry.)

Taylor, C. (1989) *The Sources of the Self: The Making of the Modern Identity*, Cambridge, MA: Harvard University Press, 248-59. (A brief sketch, placing Shaftesbury firmly in his contemporary philosophical context.)

Trianosky, G. (1978) 'On the Obligation to be Virtuous: Shaftesbury and the Question, Why be Moral?', *Journal of the History of Philosophy* 16: 289-300. (A penetrating discussion of book II of *An Inquiry Concerning Virtue or Merit*.)

Voitle, R. (1984) *The Third Earl of Shaftesbury*, Baton Rouge, LA: Louisiana State University Press. (The definitive biography, with an excellent bibliography.)

Shah Wali Allah (Qutb al-Din Ahmad al-Rahim) (1703-62)

Shah Wali Allah of Delhi, the greatest Muslim scholar of eighteenth-century India, made an immense contribution to the intellectual, economic, social, political and religious life of the Muslim community in India, the effects of which persist to the present day. He lived during a time when the Muslim empire was losing ground on the Indian subcontinent, with the Muslim community divided and at odds. Seeking to give theological and metaphysical issues a new rational interpretation and labouring to harmonize reason and revelation, he tried to reconcile the various factions of the Indian Muslims, thereby protecting the empire from collapse.

Shah Wali Allah contended that the root cause of the downfall of the Indian Muslims was their ignorance of the sacred scripture of Islam. He initiated a movement with the theme 'Back to the Qur'an', and translated the Qur'an into Persian to facilitate its understanding among all the Muslims of India. It is believed to be the first complete translation of the Qur'an from the Arabic by an Indian Muslim scholar.

1 Life

Qutb al-Din Ahmad ibn 'Abd al-Rahim, known as Shah Wali Allah, was born in AH 1114/AD 1703 near Delhi, a member of a distinguished intellectual and religious family. He received a highly structured education and spiritual instruction at the *madrasa* (religious school) established by his father, Shah 'Abd al-Rahim, at Delhi. As well as the Qur'an, he studied Arabic and Persian grammar and literature and the higher philosophical, theological, metaphysical, mystical and juridical texts. He graduated from the school when he was barely fifteen years old; in the same year, his father initiated him into the famous Naqshbandi order. He began his career as a teacher at the Madrasa-e-Rahimiyya under the tutelage of his father; after the death of the latter in AH 1131/AD 1719, Shah Wali Allah became the head of the *madrasa*, teaching all the current sciences at the school for about twelve years. During the same period he continued his own studies, growing in stature as a teacher and attracting students to his circle.

In AH 1143/AD 1731, Shah Wali Allah went on the *hajj* (pilgrimage to Mecca), after which he remained in Mecca and Medina, the sacred cities of Islam, for about fourteen months, studying *hadith* (accounts of the Prophet) and engaging in intellectual discussions, meditation and spiritual perfection. During this time, he saw the forty-seven spiritual visions which form the subject matter of his famous mystical work *Fuyud al-haramayn (Emanations or Spiritual Visions of Mecca and Medina)*. After making his second *hajj*, Shah Wali Allah returned home to Delhi in AH 1144/AD 1732. He spent the rest of his life teaching *hadith* literature and metaphysics and writing. All but one or two of his works were produced during his later years. He died in AH 1176/AD 1762.

2 Intellectual and metaphysical contribution

Shah Wali Allah wrote in both Arabic and Persian. He published between fifty and seventy works, including five collections of letters and epistles. His writings played a major role in the intellectual and spiritual life of the Muslims in the Indo-Pakistan subcontinent, a role which continues today. Some of these works have greatly changed the Muslim approach to the study of the Qur'an.

In addition, Shah Wali Allah tried to reshape Islamic metaphysics in greater conformity with the teachings of the Qur'an and the *sunna* of the Prophet. He adopted a more rational approach to the controversial issues of metaphysics, which led to greater harmony among subsequent Islamic metaphysical thinkers. He was careful to give a balanced criticism of some of the views of his predecessors and contemporaries. His constructive and positive approach to those issues was always considered a sincere attempt at reconciliation.

Shah Wali Allah made the first attempt to reconcile the two (apparently) contradictory doctrines of *wahdat al-wujud* (unity of being) of Ibn al-'Arabi and *wahdat al-shuhud* (unity in conscience) of Shaykh Ahmad Sirhindi. Shaykh Muhyi al-Din ibn al- 'Arabi, the advocate of *wahdat al-wujud*, was of the opinion that being in reality is one and God. All other actual and possible beings in the universe are manifestations and states or modes of his Divine Names and Attributes. By the act of creation through the word *kun* (be), Ibn al-'Arabi means the descent of Absolute Existence into the determined beings through various stages. This gradual descent of the Absolute Existence is called *tanazzulat al-khamsa* (five descents) or *ta'ayyunat al-khamsa* (five determinations) in Sufi terminology. On the other hand, according to Shaykh Ahmad Sirhindi, the exponent of the doctrine of *wahdat*

al-shuhud, God and creation are not identical; rather, the latter is a shadow or reflection of the Divines Name and Attributes when they are reflected in the mirrors of their opposite non-beings (*a'dam al-mutaqabila*). Shah Wali Allah neatly resolved the conflict, calling these differences 'verbal controversies' which have come about because of ambiguous language. If we leave, he says, all the metaphors and similes used for the expression of ideas aside, the apparently opposite views of the two metaphysicians will agree. The positive result of Shah Wali Allah's reconciliatory efforts was twofold: it brought about harmony between the two opposing groups of metaphysicians, and it also legitimized the doctrine of *wahdat al-wujud* among the *mutakallimun* (theologians), who previously had not been ready to accept it.

Shah Wali Allah wrote about thirteen works on metaphysics, which contain his constructive and balanced metaphysical system. One of the most important is *al-Khayr al-kathir (The Abundant Good)*. This work is divided into ten chapters, each called a *khizana* (treasure). The first four chapters deal with the reality of *wujud* (being), knowledge of God, the relationship between God and the universe, and human knowledge. From the discussion of human knowledge, Shah Wali Allah turns to the discussion of the reality of prophecy and the prophethood of Muhammad. In the seventh *khizana*, he deals with the rules and principles of sainthood and mysticism. The eighth and ninth chapters contain details about practical aspects of Islam, the *shari'a*, as well as the eschatological view of Islam. In the tenth *khizana*, Shah Wali Allah explains his theological view which, according to him, is in full accord with Ash'arite theology.

Altaf al-quds fi ma'rifat lata'if al-nafs (The Sacred Knowledge) is another metaphysical work concerned with the inner dimensions of human personality. Here Shah Wali Allah deals with the important questions of mystical intuition (*kashf*) and inspiration (*ilham*). He examines systematically the reality of both the external and internal perceptive qualities of a human being as the heart, the intellect, the spirit, the self, the secret (*al-sirr*) and the ego. A separate chapter is devoted to the metaphysical teachings of Shaykh Junaid Baghdadi, wherein he presents a brief historical account of mysticism. The last chapter deals with the subtle question of 'thoughts and their causes'. Shah Wali Allah specifies various external and internal causes which affect the human mind and produce thoughts.

Sata'at (Manifestations) is a systematic division of *wujud* (being), representing Shah Wali Allah's view concerning the *tashkik al-wujud* (hierarchy or gradation of being). Existence, in relation to determined being, is composed of existence and essence and has many grades, stages and modes. The particular beings in the universe provide the foundation for the claim of the *tashkik* (gradation) and *kathrat* (multiplicity) of being. Each grade or stage covers a certain area of determination and each stage is related to the next, not in a way that a material being is connected to another material being, but in *ma'nawi* (conceptual) manner. He describes the relationship between the various stages of being as like that between the lights of various lamps in a single room. The lights of these lamps are apparently mingled and are one, and are difficult to differentiate from one another; but in reality, they are distinguishable from one another because of the number of the lamps.

Shah Wali Allah's 'magnum opus' is his *Hujjat Allah al-baligha (The Profound Evidence of Allah)*. This comprehensive work deals with both intellectual and practical aspects of Islam. The first part deals with metaphysics, scholastic theology, the gradual development or evolution of human society and the philosophy behind the divine injunctions. The second part is devoted to ethics, politics, rituals and the social life of Islam.

Al-Tafhimat al-ilahiyya (Instructions or Clear Understanding) is one of his most comprehensive metaphysical works. The work is divided into sections called *tafhim* (instruction). Both Arabic and Persian languages are used for the expression of ideas and concepts in this work. These *tafahim* (plural of *tafhim*) are actually Shah Wali Allah's mystical visions and experiences, and his letters and articles written to various people at various times in different contexts. The famous epistle called *Maktub al-madani (Madinian Epistle)* to Isma'il Afandi is a part of the second volume of the book. This article is a detailed description of *wahdat al-wujud* and *wahdat al-shuhud*, along with Shah Wali Allah's attempt at reconciliation concerning this controversial issue. In addition to the ontological discussions, the work also includes the author's cosmological, anthropological and theological views.

Another important metaphysical work is *al-Budur al-bazigha (The Full Moons Rising in Splendour)*. The introduction deals with basic metaphysical issues such as *wujud* in general, the unity of God, the essence and existence of God and the relationship between God and the universe. Shah Wali Allah considers the universe to be a manifestation of the Divine Attributes. In the first chapter, he deals with the study of humanity with respect to its social and rational being. The second chapter is devoted to humanity's relationship with the Creator. At the end of

the work, Shah Wali Allah describes in detail the reasons and causes for the development and evolution of the various *shara'i'* (religions or religious laws) and *milal* (religious communities).

Shah Wali Allah also tried to provide a basis for bringing the four schools of law closer to each other. His commentaries on the *Mu'atta* (a collection of the Prophet's sayings) of Imam Malik, called *al-Musawwa* (Arabic) and *al-Musaffa* (Persian), were written with a view to finding common orthodox ground for the reconciliation of different schools of Islamic law. Likewise, he wrote *'Aqd al-jid fi akham al-ijtihad wa'l-taqlid* with the proposal that the door of *ijtihad* (judgement) is open. According to him, the experts of Islamic knowledge (*'ulama'* (religious scholars) and*mujtahidin* (legists) have the right to respond effectively to new situations instead of being perpetually bound to previous solutions.

3 Political contribution

A hallmark of Shah Wali Allah was his ability to reconcile opposing points of view to the satisfaction of each side. Standing behind this aspect of his teachings is the unity of the Muslim community or *umma*. His powerful abilities as a reconciler enabled him to provide common ground and a strong basis for co-operation and harmony between the Sunni and Shi'i.

Shah Wali Allah lived during a time of political and moral decline, chaos and destruction in the Mughul empire. His vantage point near the centre of the Muslim state gave him a clear view of the situation. He did his best to bring stability to the tottering empire and protect the Indian Muslims from disaster. Through his writings, especially his letters, he appealed to the Muslim rulers, nobles and intelligentsia to be aware of the dreadful situation and its possible consequences. His correspondence reveals many factors of Indian politics in the eighteenth century. His detailed letter to Ahmad Shah Abdali, the founder and ruler of Afghanistan, contained a comprehensive picture of the political situation in India. Ahmad Shah Abdali heeded Shah Wali Allah's call to invade India and restore Muslim power to the country, culminating in the defeat of the Marathas and their allies at the battle of Panipat in 1761. Shah Wali Allah himself left a rich intellectual legacy in the form of literary works, well-trained disciples including his four sons - who also became eminent scholars - and one of the greatest educational institutions of the time.

See also: Islam, concept of philosophy in; Islamic philosophy, modern; Islamic theology; Mystical philosophy in Islam

HAFIZ A. GHAFFAR KHAN

List of works

Shah Wali Allah (1703-62) *Altaf al-quds (The Sacred Knowledge)*, ed. D. Pendlebury, trans. G. Jalbani, *The Sacred Knowledge*, London: Octagon, 1982.(A general account of the metaphysics of Shah Wali Allah.)

Shah Wali Allah (1703-62) *al-Khayr al-kathir (The Abundant Good)*, trans. G. Jalbani, Lahore: Ashraf, 1974. (Comprehensive discussion of the links between metaphysics and theology.)

Shah Wali Allah (1703-62) *Hujjat Allah al-baligha (The Profound Evidence of Allah)*, Lahore: Shaikh Ghulam Ali and Sons, 1979.(A detailed discussion of the links between theoretical and practical philosophy.)

Shah Wali Allah (1703-62) *Sata'at (Manifestations)*, trans. into Urdu by S.M. Hashimi, Lahore: Idarah Thaqafat Islamiyya, 1989; trans. into English by G. Jalbani, *Sufism and the Islamic Tradition: the Lamahat and Sata'at of Shah Waliullah*, London.(A systematic and highly influential account of being.)

Shah Wali Allah (1703-62) *Lamahat (Flashes of Lightning)*, Hyderabad: Shah Wali Allah Academy, 1963; trans. G. Jalbani, *Sufism and the Islamic Tradition: the Lamahat and Sata'at of Shah Waliullah*, London, 1980.(One of the important writings on Sufism.)

Shah Wali Allah (1703-62) *Fuyud al-haramayn (Emanations or Spiritual Visions of Mecca and Medina)*, Delhi: Matba' Ahmadi, no date.(A collection of pure mystical and metaphysical experiences and visions received during his stay in Mecca and Medina.)

Shah Wali Allah (1703-62) *al-Tafhimat (Instructions or Clear Understanding)*, Dabhail, 1936, 2 vols.(One of the most comprehensive metaphysical works.)

Shah Wali Allah (1703-62) *al-Budur al-bazighah (The Full Moons Rising in Splendour)*, Dabhail: Madinah Barqi Press, 1354 AH.(Important metaphysical work.)

References and further reading

Hermansen, M. (1986) ‘Shah Wali Allah of Delhi’s *Hujjat Allah al-baligha*: Tension Between the Universal and the Particular in an 18th Century Islamic Theory of Religious Revelation’, *Studia Islamica* 63: 143-57.(A clear account of Shah Wali Allah’s major work.)
Kemal, R. and Kemal, S. (1996) ‘Shah Waliullah’, in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 37, 663-70.(Account of the life, times and influence of the philosopher.)
Malik, H. (1973) ‘Shah Wali Allah’s Last Testament’, *Muslim World* 63: 105-18.(A useful summary of his basic philosophical principles.)
Rizvi, S. (1980) *Shah Wali Allah and His Times*, Canberra: Ma‘rifat Publishing House. (A discussion of the links between Shah Wali Allah’s philosophical views and the renewal movement in India.)

Shao Yong (1012-77)

One of the founders of neo-Confucianism, Shao Yong was a Chinese philosopher best known for his use of numerical ideas to illustrate natural patterns of change. His thought encompassed a variety of concerns including knowledge, language and self-cultivation, and has received differing interpretations.

Shao Yong (styled Yaofu, called Kangjie) was considered one of the five founders of the Song dynasty (960-1279) philosophical movement called the learning of the Way, widely known as neo-Confucianism. His family was from north China, and he established his home in Loyang. While friendly with scholars and officials, Shao never took the examinations and declined several official recommendations. His major works were *Huangji jingshishu (Book of the Supreme Ultimate Ordering the World)* and *Yichuan jirangji (Collection of Yichuan's Beating on the Ground)*, a collection of poetry. His supposed authorship of *Yuqiao wendui (Conversation between the Fisherman and Woodcutter)* and *Wuming gongzhuan (Biography of Mr No Name)* has not been verified.

Combining aspects from different traditions, Shao focused on natural patterns of change in the universe and the experience of sagely knowledge. Called the 'Prior to Heaven learning', his ideas regarding universal regularities belonged to the learning of images and numbers, based on the *Yijing (Book of Changes)* and emphasizing correlations. His views on knowledge, language and self-cultivation drew on Daoist thinking (see Daoist philosophy). While accepting Confucian thought, Shao stressed that political history, individual behaviour and the natural universe shared identical patterns. Buddhist influence is suggested in his ideas on sagely perception (see Buddhist philosophy, Chinese).

Shao viewed the universe as characterized by constant change, occurring according to regular patterns. Composed of *qi* (matter-energy) (see Qi), the universe and cosmological development were described in terms of theoretical levels, from the extrasensory original unity to all the things/events of human experience. Although cycles of twelve and thirty were important, most critical were the binary and quaternary patterns of change. Original unity (the Supreme Ultimate) gives rise to *yin* and *yang* which in turn interact to produce divisions of four, eight (the trigrams) and eventually sixty-four (the hexagrams) (see Yijing; Yin-yang). Production and completion, and movement and response, illustrate the binary pattern while the four seasons, four directions and four classics illustrate the quaternary pattern.

Representable by numbers and images, patterns were unknowable without things, the two largest of which were the heavens and earth. Changes in one realm were correlated with those in the other, according to a quaternary pattern that systematically classified all things/events. For instance, the heavens were symbolized by three, five, seven, nine and the sun, moon, stars and celestial markers, while earth was symbolized by two, four, six, eight and water, fire, earth and stone. The relationships among the categories explained all kinds of activity, and numbers reflected the regularity of change.

Shao urged people to overcome their limited perspectives, but only a sage perceived things from the viewpoint of the whole. Admired as a principled Confucian recluse, eccentric, teacher and poet, Shao was respected for his numerological formulations, predictive abilities and ideas about the sage. Shao's numerical ideas have been misinterpreted as bad astronomy and as a deterministic view of human behaviour.

See also: Daoist philosophy; Neo-Confucian philosophy; Yijing

ANNE D. BIRDWHISTELL

List of works

Shao Yong (1050s-70s, self-preface 1066) *Yichuan jirangji (Collection of Yichuan's Beating on the Ground)*, Sibu congkan edition, Daozang edition.(Collection of most of Shao's extant poems written over a period of about thirty years.)

Shao Yong (1060s-70s) *Huangji jingshishu (Book of the Supreme Ultimate Ordering the World)*, Sibu beiyao edition, Siku quanshu edition, Daozang edition.(Contents emphasize cosmological, historical and numerous other topics and include writings by Shao, his (possible) comments recorded by others, and charts, diagrams and writings by others.)

Shao Yong (11th-12th century) *Wuming gongzhuan (Biography of Mr No Name)*, Siku quanshu edition; trans. A.J.

Berkowitz, ‘Biography of the Gentleman with No Name’, in V.H. Mair (ed.) *The Columbia Anthology of Traditional Chinese Literature*, New York, Columbia University Press, 1994, 751-6.(Translation of Shao Yong’s autobiography; but attribution uncertain, possibly authored by Shao’s son or students.)

Shao Yong (11th-12th century) *Yuqiao wendui (Conversation between the Fisherman and the Woodcutter)*, Siku quanshu edition; trans. K. Lundbaek, *Dialogue Between a Fisherman and a Woodcutter,* Hamburg: C. Bell Verlag.(Includes a short helpful introduction and the Chinese text to accompany the English translation. Attribution uncertain, possibly authored by Shao’s son or students.)

References and further reading

Berkowitz, A.J. (1083) ‘On Shao Yong’s Dates (21 January 1012-27 July 1077)’, *Chinese Literature: Essays, Articles, Reviews* 5: 91-4.(Analysis of various writings to determine Shao Yong’s dates.)

Birdwhistell, A.D. (1982) ‘Shao Yung and His Concept of Objective Observation (fan-kuan)’, *Journal of Chinese Philosophy* 9 (4): 367-94.(Theoretical analysis of Shao’s concept of objective observation, or reflective perception.)

Birdwhistell, A.D. (1989) ‘The Philosophical Concept of Foreknowledge in the Thought of Shao Yung’, *Philosophy East and West* 39 (1): 47-65.(Epistemological analysis of Shao’s concept of foreknowledge.)

Birdwhistell, A.D. (1989) *Transition to Neo-Confucianism: Shao Yung on Knowledge and Symbols of Reality*, Stanford, CA: Stanford University Press.(Emphasizes epistemological analysis of Shao’s philosophy; has general comments on historical context and extensive bibliography.)

Black, A.H. (1989) *Man and Nature in the Philosophical Thought of Wang Fu-chih*, Seattle, WA: University of Washington Press.(Includes extensive treatment of Wang’s seventeenth-century views of Shao Yong and Song thought.)

Freeman, M.D. (1982) ‘From Adept to Worthy: The Philosophical Career of Shao Yung’, *Journal of the American Oriental Society* 102: 477-91.(Discussion of Shao’s life and thought from perspective of intellectual history.)

Ryan, J.A. (1993) ‘The Compatibilist Philosophy of Freedom of Shao Yong’, *Journal of Chinese Philosophy* 20 (3): 279-91.(Argues for a compatibilist position on freedom, with comparison to Leibniz.)

Ryan, J.A. (1996) ‘Leibniz’ binary system and Shao Yong’s *Yijing*’, *Philosophy East and West* 46 (1): 59-90. (Compares Shao Yong and Leibniz from an implicit positivist position, thus seeing Shao’s position as proto-science.)

Smith, Jr, K., Bol, P.K., Adler, J.A. and Wyatt, D.J. (1990) *Sung Dynasty Uses of the I Ching*, Princeton, NJ: Princeton University Press.(Discusses the thought of Shao and three other thinkers in a Song historical context.)

Wyatt, D.J. (1984) ‘Shao Yung: Champion of Philosophical Syncretism in Early Sung China’, unpublished Ph.D. dissertation, Harvard University.(General introduction to Shao’s life and thought.)

Wyatt, D.J. (1985) ‘Chu Hsi’s Critique of Shao Yung: One Instance of the Stand Against Fatalism’, *Harvard Journal of Asiatic Studies* 45 (2): 649-66.(Analysis from perspective of intellectual history.)

Shem Tov family

The ibn Shem Tov family included four Jewish intellectuals of fifteenth century Spain whose philosophical, theological, homiletical and polemical works followed the persecution of 1391 and the ensuing mass apostasy of Jews. Responding to these traumatic events, the Shem Tovs rethought the place of philosophy in traditional Judaism. Although the pater familias *reacted sharply to the spiritual crisis by criticizing Maimonides and endorsing Kabbalah, his offspring charted a more moderate course that enabled Jewish intellectuals to cultivate philosophy and the kindred arts and sciences while asserting the ultimate primacy of their revealed faith over philosophy, and its philosophical superiority to Christianity.*

1 Shem Tov ibn Shem Tov

Shem Tov ibn Shem Tov (*c.* 1380-1441) witnessed the destruction of Spanish Jewry in his youth. Like Hasdai Crescas and Shlomo Alami, he responded by comprehensively re-examining Judaeo-Hispanic culture. In *Sefer ha-'Emunot (The Book of Beliefs)* he held Jewish rationalists responsible for the mass defections that followed the persecutions. He charged his chief targets, Abraham ibn Ezra and Moses Maimonides, with denying cardinal tenets of rabbinic Judaism including divine retribution, bodily resurrection, personal providence, miracles and individual immortality. They had accorded Jewish Law merely instrumental value, had made human perfection dependent on knowledge of philosophy, equated prophecy with natural human knowledge and reduced the Torah to a popular allegory of philosophic profundities.

Shem Tov conceded to philosophy sound knowledge of the natural world, knowledge that can aid us in grasping the plain sense of the Torah, but he found philosophy inherently unreliable in metaphysics and subordinated it to revelation, whose salvific truths are beyond the ken of natural reason (see Revelation). Revelation itself requires authoritative interpretation through a tradition that, for Shem Tov, includes the teachings of Kabbalah (see Kabbalah). In a work still extant in manuscript on the subject of the Sefirot - originally Neo-Pythagorean hypostases but later understood to be aspects of the divine pleroma - he places Kabbalistic theosophy above philosophy.

Shem Tov's endorsement anticipated by over a century the rise of Kabbalah to centrality in Jewish theology, but he failed to diminish the stature of Maimonides or to discredit Judaeo-Hispanic rationalism. His own sons, Joseph (*c.*1400-*c.*1460) and Isaac (d. *c.*1489) and his grandson, Shem Tov ben Joseph (*fl.* 1480s), perpetuated Jewish Aristotelianism and defended Maimonides, although they agreed that philosophy cannot secure salvation alone; it must be perfected by revelation and tradition. They accepted Kabbalah as integral to Judaism but did not displace philosophy, instead incorporating Kabbalistic ideas into their rationalism. For example, they identified the Sefirot with the 'attributes of action' allowed to God in Maimonidean negative theology.

2 Joseph and Isaac ibn Shem Tov

Joseph, the most illustrious of the Shem Tovs, was a court physician and auditor of accounts in the courts of King John II and Henry IV of Castile until he was deposed in 1456. A prolific exegete, Joseph wrote commentaries, extant in manuscript, on Aristotle's *On the Soul* and *Nicomachaean Ethics*, Averroes' now lost epistle on the possibility of conjunction (in both longer and shorter formats), his epistle on the nature of material intellect, his commentary on Alexander of Aphrodisias' *On the Intellect*, and Moses Narboni's commentary on al-Ghazali's Avicennan summary, *Maqasid al-falasifa (The Intentions of the Philosophers)*. No longer extant are commentaries on Porphyry's *Isagōgē* and Aristotle's *Economics*. The surviving manuscripts show a deep familiarity with Judaeo-Muslim Aristotelianism and, strikingly, with Christian scholasticism.

Joseph's younger brother Isaac, a popular teacher of Aristotelian philosophy in Aguilar di Campaha, Castile, wrote four supercommentaries on Averroes's middle commentary on Aristotle's *Physics*, as well as treatments (now lost) of Averroes' long commentaries on *On the Soul* and *On Generation and Corruption* and on al-Ghazali's *Maqasid al-falasifa*. In the *Physics* supercommentaries, Isaac rebutted Crescas' attack on the twenty-five theses of Aristotelian philosophy which Maimonides had laid out as the premises of Aristotelian proofs of the existence, unity and incorporeality of God. Crescas' aim had been to liberate Judaism from Aristotelian philosophy. A loyal Maimonidean, Isaac dismissed Crescas' critique as ill-taught philosophy. He failed to grasp the originality and

subtlety of Crescas' views, but pondered the origin of the universe and other metaphysical questions in his lost *'Etz ha-Da'at (The Tree of Knowledge)*.

3 Shem Tov son of Joseph ibn Shem Tov

The last known member of the family, Shem Tov, son of Joseph ibn Shem Tov, also taught philosophy in commentaries. He wrote a supercommentary on Averroes' middle commentary on Aristotle's *Physics* and *On the Soul*, Hebrew philosophical works on the distinction between matter and form, and a work on teleology. The technical virtuosity of these works, written to preserve a worldview that had come under increasing attack, attests to the abiding appeal of philosophy in Spain until the very end of the Jewish presence there, but they show little novelty. Intellectually, the movement was a spent force. The real creativity of the Shem Tovs (and of other Jewish thinkers of their day) lay in theology and philosophical hermeneutics.

4 Theology, preaching, hermeneutics and polemics

The major vehicle of Jewish theology in fifteenth-century Spain was commentary on Maimonides' *Guide to the Perplexed* (see Maimonides, M.) Joseph ibn Shem Tov's commentary on the highly controversial 68th chapter of Part I survives in manuscript. Isaac commented on the entire *Guide* (the commentary on Part I is extant), and Joseph's son Shem Tov's linear commentary on the entire *Guide* is still printed in traditional editions. By seeking to fathom Maimonides' original intent, these commentaries ranged over the full repertoire of medieval Jewish philosophy, including God's existence and attributes, creation, providence, miracles, prophecy, human knowledge and the ultimate human goal. The works also defended Maimonides against charges of heresy and inconsistency. Like other Maimonideans (Abraham Shalom, Abraham Bibago and Moses Alashkar), the Shem Tovs ascribed their own moderate views to Maimonides and projected upon him a synthesis they had developed from their own Maimonidean reflections on Averroes (see Ibn Rushd), Avicenna (see Ibn Sina), al-Ghazali, Ibn Bajja and Ibn Tufayl.

Joseph ibn Shem Tov's systematic treatises develop his moderate synthesis of religion and philosophy. In *Kevod Elohim (The Glory of God)*, he follows Meir Alguades' Hebrew translation of the Latin *Nicomachaean Ethics*, seeking to harmonize Aristotle's idea of happiness with rabbinic values. Taking his cue from Thomas Aquinas, Joseph argues that Aristotle delineates the supreme temporal good, whereas eternal happiness can be known only through revelation. Like Aquinas, Joseph understood this ultimate felicity as a beatific vision by the immortal soul of God's essence; but only those who live by the precepts of the Torah can attain such eternal bliss. Joseph's narrowing of philosophical universality by the specification of this requirement clearly aimed to strengthen the faith of despairing Spanish Jews.

In his *Moznei ha-'Iyyun (The Scales of Speculation)*, Joseph developed these themes further and sought to explain the mass apostasy of Spanish Jews as the outcome not of philosophical study but of weak character and superficial intellectual commitment, which had allowed his countrymen to succumb to the material blandishments of Christianization. From Talmudic sources, he showed that there was no religious prohibition against philosophy, which was in fact highly beneficial in understanding of the revealed canon. Addressing another oft-cited cause of mass apostasy, Joseph composed *Da'at 'Elyon (Knowledge of the Supernal)* attacking the determinism of Abner of Burgos, a famous Jewish apostate and Christian theologian (see Voluntarism, Jewish §1). Finally, he wrote a commentary on Yeda'aiah Berdersi's *Behinat 'Olam (Examination of the Universe)*, a popular fourteenth-century poetic summary of Jewish Aristotelianism. These two works are no longer extant.

The Shem Tovs all engaged in sacred hermeneutics. The elder Shem Tov commented on Mishnah Avot and the Passover Haggadah; Joseph commented on Lamentations, Job and Genesis, and on Tractate *Ketubot* of the Babylonian Talmud. Shem Tov ben Joseph published homilies on the Torah based on his sermons. Preaching was the major vehicle for the dissemination of Jewish theology, and Joseph was not only a popular preacher (fifteen of his sermons survive in manuscript) but the author of the earliest known Hebrew manual on homiletics, *'Eyn ha-Qore (The Eye of the Reader)*. Inspired by a humanist concern for eloquence, it sought to improve the quality of Jewish preaching and counter the appeal its Christian counterpart.

Like other polemicists of his day, Joseph defended Judaism as a rational religion grounded in the revelation of God's perfect wisdom. Christianity was philosophically inferior, based on irrational beliefs and paradoxes. To aid

the resistance to conversionary pressures and rebut Christian claims to spiritual superiority, he translated Hasdai Crescas' *Bittul 'Iqqarey ha-Notzrim (Refutation of Christian Principles)* from the Catalan original into Hebrew and wrote a commentary that preserves the only extant version of this influential text (see Crescas, H.). He also commented on Profiat Duran's epistle *Al Tehi ka-Avotekha (Be Not like Your Fathers)*, known to Christians under the distorted title *Alteca Boteca*. Duran had mocked the apostasy of the epistle's addressee by satirizing Christian dogmas (see Duran, P.). Joseph's commentary sought to ensure that Jews understood the irony of the original polemical work, but it had the unintended effect of alerting the Inquisition to the work's genuine intent.

See also: Aristotelianism, medieval; Maimonides, M.

HAVA TIROSH-SAMUELSON

References and further reading

Davidson, H. (1983) 'Medieval Jewish Philosophy in the Sixteenth Century', in B.D. Cooperman (ed.) *Jewish Thought in the Sixteenth Century*, Cambridge, MA: Harvard University Press, 106-45.(The first part of this essay portrays the conservative stance of fifteenth-century Jewish philosophers, among them Joseph ibn Shem Tov.)

Gottlieb, E. (1976) 'Al Darko shel Shem Tov ibn Shem Tov La-Qabbalah' (The Path of R. Shem Tov ibn Shem Tov towards Kabbalah), in J. Hacker (ed.) *Studies in the Kabbalah Literature*, Tel Aviv: Tel Aviv University, 347-56.(Discusses the shift of Shem Tov ibn Shem Tov from rationalism to Kabbalah and lists the manuscripts of his Kabbalistic works.)

Guttmann, J. (1913) 'Die Familie Shem Tob in ihren Beziehungen zur Philosophie' (The Shem Tov Family and its Relationship to Philosophy), *Monatschrift für Geschichte und Wissenschaft des Judentums* 57: 177-95, 326-40, 419-51.(To date, this is still the only comprehensive study of the entire Shem Tov family.)

Ibn Shem Tov, Shem Tov (*c*.1420?) *Sefer ha-'Emunot (The Book of Beliefs)*, Ferrara, 1556; photomechanical reproduction, Jerusalem, 1969.(Attacks Ibn Ezra and Maimonides, whom Ibn Shem Tov holds responsible for defections from the faith following Christian persecutions.)

Ibn Shem Tov, Joseph (1442) *Kevod Elohim (The Glory of God)*, Ferrara, 1556; photomechanical reproduction, Jerusalem, 1969.(Work on ethics, following Aristotle's *Nicomachean Ethics*.)

Ibn Shem Tov, Joseph (1451) *Sefer Bittul 'Iqqarey ha-Nozrim*, ed. D.J. Lasker, *Sefer Bittul Iqqarei ha-Nozrim: Translation of (Crescas' work by) Joseph Ben Shem Tov*, Ramat Gan and Beer Sheva: Bar Ilan University Press and Ben Gurion University, 1990; trans. D.J. Lasker, *The Refutation of the Christian Principles by Hasdai Crescas*, Albany: State University of New York Press, 1992.(Translation of the original by Hasdai Crescas, on the refutation of Christian principles.)

Ibn Shem Tov, Shem Tov ben Joseph (1489) *Derashot 'al ha-Torah (Sermons on the Torah)*, Salonica, 1525.(Religious work by the last known member of the family.)

Lasker, D.J. (1977) *Jewish Philosophical Polemics Against Christianity in the Middle Ages*, New York: B'nai B'rith Publications.(An excellent study of the Jewish-Christian polemics which climaxed in the fifteenth century; explains how Jewish polemicists used Aristotelian philosophy against Christian claims to spiritual superiority.)

Regev, S. (1983) 'Teologiah u-Mysticism Rationali be-Kitvey R. Yosef ibn Shem Tov' (Theology and Rational Mysticism in The Writings of R. Joseph ben Shem Tov), unpublished Ph.D. dissertation, Hebrew University.(In Hebrew with an English summary. The most systematic analysis of Joseph ibn Shem Tov's works and his moderate rationalism; analyzes his theory of prophecy and its Averroian epistemological background; comprehensively discusses the relationship between philosophy and religion in his theological compositions, and lists all the known manuscripts of his works.)

Regev, S. (1986) 'Ha-Mahshavah ha-Rational-Mystit ba-Hagut ha-Yehudit ba-Me'ah ha-Tet-Vav' (Rational-Mystical Thought in Jewish Philosophy of the Fifteenth Century), *Jerusalem Studies in Jewish Thought* 5: 155-89.(Presents Joseph ibn Shem Tov as a typical example of Jewish philosophy in the fifteenth century, and shows how he sought to reconcile rationalist philosophy and Kabbalah.)

Regev, S. (1988) 'Le-Ba'ayt Limud ha-filosofia be-Hagut ha-Me'ah ha-Tet-Vav: R. Yosef ibn Shem Tov ve-R. Abraham Bibago' (On the Problem of the Study of Philosophy in 15th-Century Thought: R. Joseph ibn Shem Tov and R. Abraham Bibago), *Daat* 16: 57-86.(Considers the debate on the status of philosophy in Jewish culture during the fifteenth century in the wake of the mass apostasy of Iberian Jews. By focusing on Ibn Shem Tov and Bibago, who wrote commentaries on Averroes, the author shows how the philosophers argued for the superiority of Kabbalah over philosophic knowledge.)

Saperstein, M. (1989) *Jewish Preaching: 1200-1800*, New Haven, CN: Yale University Press, 167-79, 180-201.

(This anthology includes scholarly translations of sermons by Joseph ibn Shem Tov and his son, Shem Tov, locating them in the context of Jewish biblical exegesis and homiletics.)

Schwartzmann, J. (1991) 'Perushu shel Yitzhaq ibn Shem Tov le-'Moreh Nevukhim' (Isaac ibn Shem Tob's Commentary on 'The Guide of the Perplexed'), *Daat* 26: 43-59.(Based on an unpublished doctoral dissertation, this essay summarizes the major themes of Isaac ibn Shem Tov's religious philosophy in ontology, cosmology, theology, rational psychology and the interactions of reason and revelation.)

Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Cambridge: Cambridge University Press, 381-4.(The chapter on Jewish philosophy in the fifteenth century includes important summaries of Joseph and Isaac ibn Shem Tov's chief works and ideas.)

Tirosh-Rothschild, H. (1997) 'Human Felicity - 15th Century Perspectives on Happiness', in B. Cooperman and A. Seeff (eds) *Iberia and Beyond: Hispanic Jews Between Two Cultures*, Willmington, DE: University of Delaware Press. (Discusses Joseph ibn Shem Tov's view of human happiness, the reception of Aristotle's *Ethics*, Jewish-Christian polemics and Joseph's debt to Aquinas.)

Wolfson, H.A. (1977) 'Isaac ibn Shem-Tob's Unknown Commentaries on the *Physics* and His Other Unknown Works', in I. Twersky and G.H. Williams (eds) *Studies in the History of Philosophy and Religion*, Cambridge, MA: Harvard University Press, vol. 2, 479-90.(This essay was the first to identify Isaac ibn Shem Tov as the author of the four commentaries on Aristotle's *Physics* that rejected Crescas's critique of Aristotle.)

Shestov, Lev (Yehuda Leib Shvartsman) (1866-1938)

A major though atypical figure of the Russian Religious-Philosophical Renaissance, Shestov taught that reason and science can neither explain tragedy and suffering, nor answer the questions that matter most. A maximalist, a subjectivist and an anti-dogmatist, Shestov regarded philosophical idealism as an attempt to gloss over the 'horrors of life' and attacked morality and ethics as inherently coercive. He maintained that science ignores the contingent, the unique and the ineffable, that philosophy cannot be a science, and that necessity depersonalizes and dehumanizes the individual. Philosophy and revelation are incompatible because God is not bound by reason, nature or autonomous ethics. To God 'all things are possible', even undoing what has already happened. God even restored Job's dead children to him - the same children, not new ones, Shestov insisted.

In Dobro v uchenii gr. Tolstogo i Fr. Nitshe (The Good in the Teaching of Tolstoy and Nietzsche) (1900) and Dostoevskii i Nitshe(Dostoevsky and Nietzsche) (1903) Shestov attacked philosophical idealism and attributed his subjects' philosophies to a defining personal experience: Tolstoi's horror at urban poverty, Nietzsche's illness, and Dostoevskii's Siberian exile, respectively. These books established Shestov as a major literary critic and interpreter of Nietzsche. Around 1910 he turned to philosophy and religion. In his magnum opus Athènes et Jerusalem (1938) Shestov preached a religious existentialism centred on the living God of Abraham, Isaac and Jacob, and argued that evil came into the world with knowledge. Adam and Socrates were fallen men because they opted for knowledge over life and faith. Socrates, Aristotle, the Scholastics, Descartes, Spinoza, Kant, Hegel and Husserl were all faith-destroying. Shestov preferred the anti-rationalism of Dostoevskii, Nietzsche, Tertullian, Luther, Pascal and Kierkegaard. He wanted to restore the primordial freedom of Adam before the Fall. Although Shestov quoted the Gospels and certain Christian theologians approvingly, he was not a Christian. Neither was he an adherent of traditional Judaism.

A brilliant stylist, Shestov used reason and knowledge to combat reason and knowledge. He distinguished between the empirical realm where they applied and the metaphysical realm where they did not. But since he philosophized only about the metaphysical realm he comes across as an irrationalist.

1 Life

Son of a wealthy Jewish family, Shestov was born Yehuda Leib Shvartsman in Kiev. As a student, he was primarily interested in political and social questions. Because of his radical political views, he had to transfer from a gymnasium in Kiev to one in Moscow. He studied science and mathematics and then law at Moscow University, but completed his studies at Kiev University. His doctoral dissertation was on factory legislation; because it described the appalling condition of the workers it was rejected by the Board of Censors in Moscow as revolutionary. Shestov graduated with the degree of Candidate, not Doctor, of Law in 1889. He was licensed to practise law, but never did. He joined the army and then entered his father's textile business. In 1895, he had a nervous breakdown brought on by a combination of business pressures (he saved the firm from bankruptcy) and unhappy romantic involvements with two Gentile women, one after another, whom his father would not allow him to marry. To recover his health, Shestov went abroad in the spring of 1896. In Rome, he fell in love with a Russian Orthodox medical student, Anna Elezarovna Berezovskaia. They lived together and had two daughters. In 1907, they were married in London. Shestov did not tell his parents about Anna and the children until 1914, living apart from them, sometimes in different cities, for months at a time. Shestov also had an illegitimate son, who lived in Moscow with his mother. Shestov supported them and was a real father to his son.

In 1898, Shestov returned to Russia, bringing two completed manuscripts with him: *Shekspir i ego kritik Brandes (Shakespeare and his Critic Brandes)* and *Dobro v uchenii gr. Tolstogo i Fr. Nitshe (The Good in the Teaching of Tolstoy and Nietzsche)*. He lived in St Petersburg and Moscow and then resumed working in his father's business in Kiev. Shestov's acquaintance with Merezhkovskii, Rozanov, Diaghilev and Ivanov, and his lifelong friendships with Berdiaev, Bulgakov and Remizov date from this time (see Russian Religious-Philosophical Renaissance §1). Early in 1906 he came close to another nervous breakdown. He was then forty years old and still unable to live as he pleased. *Apofeoz bezpochvennosti (The Apotheosis of Groundlessness)* (1905) and *Nachala i kontsy (Beginnings and Endings)* (1908) reflect the hopelessness and despair that overwhelmed him. Between 1910 and 1914 he lived in Switzerland with his family and worked on *Sola Fide (By Faith Alone)* and *Vlast' kliuchei (The Power of the*

Keys). When the First World War broke out, they returned to Russia and settled in Moscow. In mid-1918, they fled to Kiev, then under German control. He gave a course on Greek philosophy at the University of Kiev and was awarded an honorary doctorate by the University of Simferopol. In February 1919, Kiev was 'liberated' by the Red Army.

The Bolsheviks considered Shestov a revolutionary writer and offered to publish *Potestas Clavium* if he would add a brief introduction defending Marxist doctrine, but he refused. In the autumn of 1919 the Shestovs emigrated, reaching Paris by a round-about route in 1920. Shestov was almost unknown there, but three articles, published in French, on Dostoevskii, Tolstoi and Pascal, brought him a European-wide reputation. The German Nietzsche Society elected him its honorary president and he was invited to lecture all over Europe. He was also elected to the Kant Society. His books were translated into French, German and other European languages. He taught at the Institut des Etudes Slaves and in the extension division of the Sorbonne, and associated with leading European intellectuals, including Husserl, who became one of his best friends despite their intellectual differences. It was at Husserl's urging that Shestov read Kierkegaard. Shestov's last article, 'In Memory of a Great Philosopher', was a tribute to Husserl, who had just died.

Politically Shestov was a liberal or a moderate socialist. After all, politics and economics belonged to the empirical realm. In 'Chto takoe bol'shevizm?' (What is Bolshevism?) (1920), he described Soviet rulers as reactionary, destructive and parasitical, emphasized their hostility to freedom, and pointed to the 'paradox of idealists who believe only in crude physical force'. In 'Menacing Barbarians of Today' (1934), Shestov compared the Nazis to the Tartars and maintained that civilization entailed the triumph of the spirit over force.

There is a personal dimension to Shestov's lifelong campaign against reason and morality - the double life he led for seventeen years and long-standing tensions with his father. At 12 or 13 Shestov was kidnapped by radicals and held for ransom. Apparently, this was a hoax to raise funds and Shestov was involved in the plot. Perhaps suspecting this, his father refused to pay. After six months Shestov was released unharmed. The battle of wills between father and son is significant, especially since Shestov lost. It is almost as if his writings were an outlet for his rage, not only against his father but against authority and 'necessity' in any form. Shestov's interpretation of the authors he treats is highly personal. He highlighted episodes or statements that resonated with his own experience, repeated them over and over again, and omitted others of equal or even greater importance.

2 Early writings

Shestov's early writings treated issues that were to occupy him all his life. In his first book, *Shekspir i ego kritik Brandes*, Shestov discussed the meaning of tragedy, the limits of positivism and Kantianism, and the indifference of science to life, giving as an example a brick that falls on someone's head and kills him. The law of gravity can explain why the brick fell, he said, but not why it fell on that particular man, nor does it care. Disputing Brandes' contention that tragedy is meaningless, Shestov argued that tragedy has a higher purpose - spiritual development or moral edification. He rejected this idealist interpretation in his books on Nietzsche, but came back to it, in a way, later on. A tenet of Shestov's religious existentialism is that suffering is the beginning of awakening. He also stated that philosophy begins in despair, not in wonder, as the Greeks believed.

Shekspir i ego kritik Brandes begins and ends with a quotation from Nietzsche, but Shestov had not yet assimilated Nietzsche's ideas. On first reading, Shestov did not understand *Beyond Good and Evil*. He found *The Genealogy of Morals* profoundly disturbing and searched for ways to combat Nietzsche's 'terrible, pitiless thought'. But then he learned about Nietzsche's sickness; 'on this day, I understood'. Reading Nietzsche sensitized Shestov to 'Nietzschean' qualities in Dostoevskii and accelerated Shestov's rejection of philosophical idealism and moralism (see Dostoevskii, F.; Nietzsche, F.).

In *The Good in the Teaching of Tolstoy and Nietzsche*, Shestov contrasted Nietzsche's passionate search for truth 'beyond good and evil' with Tolstoi's 'cowardly' submission to idealist pieties (see Tolstoi, L.N.). As Shestov tells it, Tolstoi was appalled by the poverty he saw in Moscow, but rather than admit his helplessness, he retreated to his estate to perfect himself. From then on, Tolstoi preached the identity of goodness and God, and tried to impose his self-denying morality on others. By contrast, Nietzsche's sickness led him to realize the futility of lofty ideals. The sickness was unjustified; Nietzsche had lived a virtuous life, but that was Shestov's point - great suffering is inexplicable and arbitrary. From personal experience Nietzsche learned that nature is cruel and that

God (goodness) is dead. But then Nietzsche lost his nerve and created a new idol - the Superman. Nevertheless, Shestov concluded: 'Nietzsche has shown the way. We must seek what is *higher* than compassion, what is *higher* than good. We must seek God' ([1900] 1969: 140).

In his next book, Shestov treated Dostoevskii and Nietzsche as 'underground' thinkers and spiritual twins. Dostoevskii's defining experience was hard labour and exile in Siberia, followed by his discovery of his own egoism upon his return to St Petersburg. In *Zapiski iz podpol'ia (Notes From Underground)* (1864) the protagonist declares that the whole world could 'go to pot' as long as he had his regular cup of tea. In *Crime and Punishment* Raskol'nikov asserts that there are two moralities, one for ordinary people, the other for extraordinary people like himself. Therefore, he was entitled to murder a 'useless' old woman because he needed her money. Nietzsche's 'underground' was illness, physical pain and loneliness. Like the protagonist of *Notes From Underground*, Nietzsche honestly and openly admitted that the happiness of humanity did not interest him. Both writers realized that egoism negates reason, conscience and idealism. Theirs was a '*morality of tragedy*', as distinct from a 'morality of *commonplaceness*'. Circumstances, not innate nobility, forced them into it.

Shestov believed that social changes will never banish tragedy from life and that to deny suffering is to deny reality. Philosophy's task, therefore, was not to preach humility, submission, renunciation or self-sacrifice, but to teach people not 'to transfer all the horrors of life into the sphere of the *Ding an sich*'. In *Apofeoz bezpochvennosti*, a compilation of 168 numbered sections, Shestov wrote: 'The business of philosophy is to teach men to live in uncertainty - man who is supremely afraid of uncertainty, and who is forever hiding behind this or the other dogma' ([1905] 1920: §11). In 'Creation From the Void', an essay published in *Nachala i kontsy (Beginnings and Endings)* (1908), Shestov praised Chekhov precisely because he did not offer a consoling ideal; his protagonists were people who lost all hope. In 'The Gift of Prophecy', Shestov pointed to prophecies of Dostoevskii, Tolstoi and Solov'ëv that did not come true. In *Velikie kanuny (Great Vigils)* (1911), Shestov interpreted the life and work of Sologub, Ibsen, William James and Tolstoi (as creator and destroyer of worlds) and criticized Kantian and positivist theories of knowledge.

Nietzsche was one of Shestov's lifelong reference points. He seconded Nietzsche's attack on Socratic rationalism (for different reasons), and claimed that it was Spinoza who murdered God. Shestov praised 'extraordinary' people who were given a second pair of eyes by the angel of death, or who thought in a second dimension. His protest against undeserved suffering turned into a conviction that for salvation 'works' don't matter. Shestov viewed Luther as a tragic figure like Nietzsche. He interpreted Nietzsche's 'will to power' as a variant of Luther's 'sola fide' and emphasized their mutual hostility to reason, righteousness and law (see Luther, M.). Dostoevskii was another lifelong reference point. When discussing people's willingness to give up their freedom for bread, security, and happiness, Shestov often alluded to 'The Grand Inquisitor' (see Dostoevskii, F.M. §4). When Shestov claimed that all metaphysical systems begin with freedom and end in necessity, he probably had Shigalëv's confession (in *The Possessed*) in mind. But he never invoked Dostoevskii's ideal of a community of faith. There is no church or synagogue in Shestov's philosophy. Focused on the solitary individual, Shestov rejected Solov'ëv's ideals of 'all-unity' and a supra-personal 'Godmanhood' (see Solov'ëv, V.S.). Moreover, Solov'ëv's orientation was too rational for him.

3 A leap into faith

Shestov's leap into faith is evident in *Sola Fide* (written 1911-14), *Vlast' kliuchei (The Power of Keys)* (1923), and *Na vesakh Iova (In Job's Balances)* (1929). Shestov considered *Sola Fide* a personal statement and did not publish it in his lifetime. The two sections, 'Greek and Medieval Philosophy' and 'Luther and the Church', reveal Shestov's immersion in Western philosophy and religion from the pre-Socratics to Husserl as well as Shestov's turn to the Bible. Describing Christian disputations on law and grace and on faith and works, Shestov pointed out that theologians attracted to Hellenic rationalism, ignored or denied miracles and mystery. *Vlast' kliuchei* reflects Shestov's horror at the First World War, not only because of the unprecedented mass slaughter, but because his beloved son was killed at the front. The title is an allusion to the keys that open and close the gates of heaven that Jesus gave to Peter. But, Shestov argued, Socrates and his successors also claimed the keys to heaven, substituting autonomous reason and autonomous virtue for the gods and maintaining that the virtuous man can be happy in a torture device (the bull of Phalaris). The Medieval Church wrested the keys away from the pagan philosophers; making the fatal mistake of trying to reconcile faith and reason, it ended up subordinating Scripture to reason. Now

the keys are being claimed by science, a development Shestov deplored. He blamed the war on Europeans' sacrilegious attempt to deify themselves and their reason in place of God. In 'Memento Mori', Shestov attacked Husserl for relying on self-evidences and logic, arguing that the law of identity, A = A, and the law of non-contradiction apply only to 'the middle zones of human and universal life, which do not at all resemble the polar and equatorial zones. The truths of one zone are not the truths of the other'. God's truths are not man's truths.

The epigram to *In Job's Balances* is Job's lament that his grief and calamity were heavier than the sands of the sea (Job 4: 2-3). Shestov's point was that human suffering weighs more than 'objective' truth. The foreword, 'Science and Free Inquiry', was about the limitations of science. Part I, 'Revelations of Death', comprises two essays: 'The Conquest of the Self-Evident: Dostoevsky's Philosophy', and 'The Last Judgment: Tolstoi's Last Works'. Dostoevskii and Tolstoi, Shestov argued, blunted their unique visions, by trying to fit them into the framework of commonplace truths. Dostoevskii transformed his insight that people actually like to suffer into the idea that suffering must 'buy' something, or redemption through suffering. Tolstoi, however, finally realized when he wrote 'The Death of Ivan Il'ich', 'Master and Man' and 'Father Sergei', that 'works' don't matter, that goodness isn't God. Shestov regarded death as an escape from 'omnitude' (allness, the One, total unity) to self-consciousness because everyone faces death alone.

Part II, 'Revolt and Submission', is comprised of fifty-two numbered sections. In the section about medieval debates on why God became a man, Shestov opposed the usual explanation, that there was no other way to save man, with his own: 'supernatural interference was only necessary because man had to be supported in his "mad endeavor", in his *unreasonable* audacity and self-affirmation' ([1929] 1975: 177; emphasis added).

Part III, 'On the Philosophy of History', comprises the essays 'Children and Stepchildren of Time: Spinoza in History', 'Gethsemane Night: Pascal's Philosophy' and 'Words That are Swallowed Up: Plotinus's Ecstasies'. Shestov held that there is no immanent Idea, or development in history, as Hegel maintained; it is all contingency. Pascal's method of seeing 'by lamentation' (Pascal was always in pain) he considered superior to Spinoza's dictum 'not to laugh, not to weep, not to curse, but to understand'. Spinoza was referring to human action and to the passions; Shestov, to misfortune. The essay on Plotinus emphasized the ecstatic insights that enable one to soar above knowledge, and Plotinus' resistance to the world-renunciation preached by the Gnostics. 'Suddenly' was one of Shestov's (and Dostoevskii's) favourite words; he used it to describe both tragedy and epiphany and associated 'suddenly' with the absurd, in other words, not part of the regular order.

4 Religious existentialism

In *Athènes et Jerusalem* (1938) Shestov summed up and developed ideas set forth in his previous works, including *Kierkegaard et la Philosophie Existentielle* (1936). His major theme was that philosophy and biblical revelation are incompatible. Therefore people must make an either/or choice between faith and knowledge. To this theme Shestov subordinated others: knowledge enslaves man; evil came into the world with knowledge; salvation is by faith alone; man must choose between the tree of knowledge and the tree of life; and God's freedom is infinite. Quoting Peter Damian's assertion that God can 'make what happened not to have happened' and Kierkegaard's *Fear and Trembling*, Shestov proclaimed that God can violate or suspend the laws he has established, be they the laws of nature or the Ten Commandments. For example, God commanded Abraham to kill Isaac. God can restore Job's children and undo Socrates drinking the hemlock and Nietzsche's madness (see Kierkegaard, S.A.).

Shestov associated the impossible with the absurd, attacked religions of consolation and claimed that both Christian and Jewish theologians have abandoned Scripture. As an example, he contrasted Isaiah's prophecy that the lion will lie down with the lamb with Aquinas' dictum, 'grace does not abolish nature'. One of Shestov's favourite biblical quotations was 'What things soever ye desire…ye shall have them' (Mark 11:24).

According to Shestov reason and morality demand resignation while the absurd and faith sanction daring. In this context, he quoted Kierkegaard on the difference between a 'knight of faith' and a 'knight of resignation', as well as Kierkegaard's admission that if he himself really had faith he would not have broken his engagement to Regina.

Shestov maintained that Western philosophy has struggled against the 'I', against individual existence, from its very beginning. Thales declared that all is God. Anaximander considered voluntary separation from the 'all' as a great sin, which doomed the individual to the greatest punishment, to death. This dreadful law inseparably links

birth and death. Necessity, uniformity, regularity and the depersonalizing philosophy of Descartes, Spinoza and Hegel all stem from it. By contrast, the Living God of Abraham, Isaac and Jacob, created man 'in his own image and likeness'. For man, as for God, everything is possible, including immortality, provided that he tear himself away from the seduction of reason. Not all questions can be answered, Shestov opined; God hides his truths in mystery.

To Shestov, the sin of Adam and Eve was choosing knowledge over life, not disobedience, as is usually maintained. Before the Fall they did not have to choose between good and evil, because there was no evil. Everything that God created was good. But the serpent deceived Eve. Knowledge does not increase human power but diminishes it. Evil and sin came into the world with knowledge. Shestov defined 'religious philosophy as the final supreme struggle to recover original freedom and the divine "very good", which was hidden in this freedom and which, after the Fall, was split into powerless good and our destructive evil' ([1938] 1966: 70). Now, instead of the creative 'Let there be!', there is only the plane of the petrified 'is'. And in the same spirit: faith abolishes reason. Faith is given to man to make him master in the world established for him by the Creator. Faith enables man to break through the self-evidences of reason and reach the freedom in which the impossible becomes the reality.

People subordinate God to reason and morality because they fear an omnipotent and capricious God, Shestov argued. Fear of the unknown leads people to take refuge in dogmatism. Luther created new catechisms and made Scripture the ultimate authority. Nietzsche deferred to necessity in the form of *amor fati*. Kierkegaard had all sorts of 'you musts'. Shestov denied any intention of attributing 'authority' to the Bible; he believed that Scripture 'decisively rejects the idea of authority'. God does not coerce man. Distinguishing between God's truth (freedom) and coercive truth (dogma), Shestov maintained that coercive truth destroys communication and leads to eternal hatred.

The utopianism and antinomianism of Shestov's *magnum opus* place him in the mainstream of the Religious-Philosophical Renaissance. On the other hand, despite his paeans to biblical revelation, his discussion of the Incarnation and his invocations of Christian theologians, he was not a Christian.

See also: Existentialism; Nietzsche: impact on Russian thought; Russian Religious-Philosophical Renaissance; Tragedy

BERNICE GLATZER ROSENTHAL

List of works

Shestov, L. (1911) *Sobranie sochinenii (Collected Works)*, St Petersburg: Shipovnik, 6 vols.(Contains all listed works to 1911.)

Shestov, L. (1898) *Shekspir i ego kritik Brandes (Shakespeare and his Critic Brandes)*, St Petersburg: Mendelevich.

Shestov, L. (1900) *Dobro v uchenii gr. Tolstogo i Fr. Nitshe - filosofiia i propoved'*, St Petersburg: Stasiulevich; 1st edn repr. Paris: YMCA-Press, 1971; trans. S. Roberts, introduction by B. Martin, *The Good in the Teaching of Tolstoy and Nietzsche*, part I of *Dostoevsky, Tolstoy and Nietzsche*, Athens, OH: Ohio University Press, 1969.

Shestov, L. (1903) *Dostoevskii i Nitshe - filosofiia tragedii*, in *Mir iskusstva*; St Petersburg: Shipovnik; 1st edn repr. Paris: YMCA-Press, 1971; trans. S. Roberts, *Dostoevsky and Nietzsche: The Philosophy of Tragedy*, Part II of *Dostoevsky, Tolstoy and Nietzsch*, Athens, OH: Ohio University Press, 1969.

Shestov, L. (1905) *Apofeoz bezpochvennosti (The Apotheosis of Groundlessness)*, St Petersburg: Obshchestvennaia pol'za; repr. of 2nd edn, Paris: YMCA-Press, 1971; trans. S.S. Koteliansky, *All Things Are Possible*, London: M. Secker, 1920, introduction by D.H. Lawrence; repr. with additional introduction by B. Martin, Athens, OH: Ohio University Press, 1977.(The 1977 edition does not include the essays on Julius Caesar and Dmitrii Merezhkovskii.)

Shestov, L. (1908) *Nachala i kontsy (Beginnings and Endings)* St Petersburg: Stasiulevich; 1st edn repr. Ann Arbor, MI: Ardis, 1978; *Penultimate Words and Other Essays*, London: M. Secker, 1920; also published as *Anton Chekhov and Other Essays*, introduced by S. Monas, Ann Arbor, MI: University of Michigan Press, 1966.(Neither translation includes 'Pokhvala gluposti' ('In Praise of Folly'), a review of Berdiaev's book *Sub specie aeternitatis*.)

Shestov, L. (1910) *Velikie kanuny (Great Vigils)*, St Petersburg: Shipovnik.

Shestov, L. (1920) 'Chto takoe bol'shevizm?' (What is Bolshevism?), Berlin: Skify.
Shestov, L. (1923) *Vlast' kliuchei (The Power of the Keys)*, Berlin: Skify; trans. B. Martin, *Potestas Clavium*, Athens, OH: Ohio University Press, 1968. (The translation does not include 'Vyacheslav velikolepnyi' (Viacheslav the Magnificent) about Viacheslav Ivanov.)
Shestov, L. (1929) *Na vesakh Iova*, Paris: Sovremennye zapiski; repr. Paris: YMCA-Press 1973; trans. C. Coventry and C.A. Macartney, *In Job's Balances*, London: Dent & Sons, 1932; repr. Athens, OH: Ohio University Press, 1975, introduction by B. Martin.(The 1975 edition was translated from the German and collated with the Russian.)
Shestov, L. (1934) 'Menacing Barbarians of Today', in *The Aryan Path*, August, Bombay; 'Ugrozy sovremennykh varvarov', *Vestnik R. Kh. D.*, Paris (1976), 119.(*The Aryan Path* was a monthly philosophy journal.)
Shestov, L. (1936) *Kierkegaard et la Philosophie Existentielle*; trans. T. Rageout and B. de Schloezer, Paris: Vrin; *Kirgegard i ėkzistentsial'naia filosofiia*, Paris: Dom knigi i sovremennye zapiski, 1939; repr. Moscow: Progress-Gnozis 1992, includes Russian translation of Czeslav Milosz' essay 'Shestov, or the Purity of Despair' (1973) from the French rather than the original English; trans. E. Hewitt, *Kierkegaard and the Existential Philosophy*, Athens, OH: Ohio University Press, 1969.
Shestov, L. (1938) *Athénes et Jerusalem*, trans. B. de Schloezer, Paris: Vrin and *Athen und Jerusalem*, trans. H. Ruoff, Graz: Schmidt-Dengler; *Afiny i Ierusalim*, Paris: YMCA-Press, 1951; trans. B. Martin, *Athens and Jerusalem*, Athens, OH: Ohio University Press, 1966.(The 1966 translation contains bibliographical footnotes.)
Shestov, L. (1964) *Umozrenie i otkrovenie*, Paris: YMCA-Press; trans. B. Martin, *Speculation and Revelation*, Athens, OH: Ohio University Press, 1982. (Includes Shestov's essay on Husserl, 'In Memory of a Great Philosopher', trans. George L. Kline.)
Shestov, L. (1966) *Sola fide-tol'ko veroiu (By Faith Alone)*, Paris: YMCA-Press.(Written 1910-14.)
Shestov, L. (1993) *Lev Shestov*, Moscow: Nauka, 2 vols, A.V. Akhutina, ed. and author of introduction, 'Odinokii myslitel''.(Vol. 1 contains *Vlast' kliuchei* and *Afiny i Ierusalim*. Vol. 2 contains Na vesakh Iova.)

References and further reading

Baranova-Shestova, N. (1983) *Zhizn' L'va Shestova (The Life of Lev Shestov)*, Paris: La Presse Libre, 2 vols. (Shestov's life through correspondence and memoirs.)
Erofeev, V. (1975) 'Ostaëtsia odno: Proizvol' (Only One Thing Remains: Arbitrariness), in *Voprosy literatury* 10: 153-88.(Argues that the idea of justice is implicit in Shestov's protest but that this idea contradicts his conception of God as arbitrariness and undermines his entire religious philosophy.)
Fondane, B. (1982) *Rencontres avec Léon Chestov*, Paris: Plasma.(An intellectual biography and reminiscences of Shestov by his only disciple. Fondane died in a Nazi concentration camp in 1944; the texts were collected and edited by Natalia Baranoff and Michel Carassou.)
Gal'tseva, R.A. (1992) *Ocherki russkoi utopicheskoi mysli xx veka (Essays in twentieth-century Russian Utopian Thought)*, Moscow: Nauka, ch. 2, 77-119. (Views Shestov as an existentialist before existentialism and contrasts his sympathy for the defeated and the suffering with Nietzsche's immoralism.)
Hill, K.R. (1976) 'The Early Life and Thought of Lev Shestov: 1866-1903', unpublished MA thesis, University of Washington.(Detailed analysis of Shestov's first three books and early articles.)
Hill, K.R. (1980) 'On the Threshold of Faith: An Intellectual Biography of Lev Shestov from 1901 to 1920', Ph.D. thesis, University of Washington.(Argues that Shestov never found the faith he so avidly sought.)
Kline, G.L. (1968) 'Religious Existentialists Shestov and Berdyaev', in *Religious and Anti-Religious Thought in Russia*, Chicago, IL: University of Chicago Press, 73-102.(Discussion of similarities and divergences in their religious existentialism.)
Kline, G.L. (1975) 'Spor o religioznoi filosofii: L. Shestov protiv V. Solov'ëva' (Arguments about Religious Philosophy: L.Shestov against V. Solov'ëv), in *Russkaia religiozno-filosofskaia mysl' xx veka*, ed. N. Poltoratzky, Pittsburgh, PA: University of Pittsburgh, 37-53.(Examines Shestov's 'existentialist' and 'irrationalist' critique of Solov'ëv's idealist speculative system.)
Maia Neto, J. (1995) 'After the Christianization of Pyrrhonism: Shestov's Irrationalism', in *The Christianization of Pyrrhonism. Scepticism and Faith in Pascal, Kierkegaard, and Shestov*, Dordrecht: Kluwer, 55-109.(Argues that Shestov's philosophy is informed by his responses to Neo-Kantianism.)
Shein, L. (1991) *The Philosophy of Lev Shestov (1866-1938), A Russian Religious Existentialist*, Lewiston, NY: Mellen.(Published posthumously. Shestov's early works, including *Sola fidei*. Vol. 2 would have included works written in emigration.)
Steinberg, A. (1991) 'Shestov', *Druz'ia moikh rannikh let (1911-28) (Friends of My Early Years (1911-28))*, Paris: Sintaksis, 217-65.(Personal recollections.)

Valevicius, A. (1993) *Lev Shestov and his Times. Encounters with Brandes, Tolstoi, Dostoevsky, Chekhov, Ibsen, Nietzsche, and Husserl*, New York, NY: Peter Lang.(Argues that Shestov developed his thought in response to the above.)

Wernham, J. (1968) *Two Russian Thinkers: Berdyaev and Shestov*, Toronto, Ont.: University of Toronto Press, 57-109.(Views them as existentialists; emphasizes their responses to Dostoevskii.)

Zakydalsky, T. (1994) 'Lev Shestov and the Revival of Religious Thought in Russia', in *Russian Thought After Communism: The Recovery of a Philosophical Heritage*, ed. J.P. Scanlan, Armonk, NY: M.E. Sharpe, 153-164. (Discussion of Russian studies of Shestov from the thaw to the present.)

Shinran (1173-1263)

Shinran lived in thirteenth-century Japan, an age of socio-political turmoil, when the old order represented by imperial rule, aristocratic culture and monastic Buddhism was in the process of internal disintegration, and a vibrant age of military clans, popular culture and new schools of Buddhism, appealing to the disenfranchised, was beginning to emerge. Although Shinran's name is not found in the historical records of the period, he left many writings, including original works, commentaries, poetry and letters that contain religious and philosophical insights which had a great impact on subsequent Japanese life. His place in history was secured when in 1921 a collection of his wife's letters, attesting to their relationship over the years, was discovered in the archives of Nishi Hongwanji in Kyoto.

Shinran inherited the pious strain of Pure Land devotional Buddhism, a strain that existed in various subsidiary forms within the great monastic schools such as Tendai, Shingon, Hossō, Kegon and Sanron. In 1175, however, his teacher Hōnen made a radical break from tradition and declared the establishment of an independent Pure Land (Jōdo) School. As a consequence, the movement suffered societal criticism and political persecutions. Among Hōnen's followers, it was Shinran who deepened and expanded the Pure Land insights within the context of Mahāyāna Buddhist thought. His basic teaching may be summarized by such concepts as self-power and Other Power, the various stages of religious life, Amida Buddha as Other Power, and birth in the Pure Land.

The terms 'self-power' and 'Other Power' were first used by Tanluan (476-542) and developed by both Hōnen and Shinran as central concepts in Pure Land praxis. Self-power refers to the belief in one's capacity to attain enlightenment; it is rooted in the unconscious attachment to the ego-self in which every thought, act and speech aims at self-gratification. In contrast, Other Power, according to Shinran, is 'that which is free of calculation'. While self-power is manifested within the subject-object dichotomous mode, Other Power is an awakening to a non-dichotomous reality, such that one sees things, including the self, free of myopic self-centredness. Religiously, Other Power is symbolically represented by Amida, the Buddha of Immeasurable Life and Light. Ultimately, self-power cannot be recognized apart from appreciating Other Power, and the latter cannot be fully realized without exerting the former to its utter, breaking limits. The contrast between the two is not unrelated to Heidegger's distinction between calculative and meditative thinking (see Heidegger, M.), although they are not religious categories; calculative thinking is epitomized in science and technology, and meditative thinking in poetry as a form of disclosure.

The aim of the Pure Land path, then, is abrogating self-power for Other Power as the center of one's life. This transformation involves the proper understanding of Amida and Pure Land, whereby the 'other' is not to be understood in the conventional subject-object sense. Shinran explicates the two in the opening lines of the chapter on True Buddha and True Land of his major work, *Kyogyoshinsho (Teaching, Practice, Faith and Realization)*, as follows:

> Reverently contemplating the true Buddha and the true land, I find that the Buddha is the Tathāgata of inconceivable light and that the land also is the land of immeasurable light.
>
> (*Kyogyoshinsho*, in Ueda 1987, III: 395)

While the Buddha and Pure Land may appear to be objectified realities seen within a subject-object framework, true appreciation for both occurs when the reality of karma-bound self is illuminated through their working. It is not the case that there exists a being called Buddha who emanates rays of light, nor is there a place called Pure Land radiating light in all directions; rather, light which is regarded as 'the form taken by wisdom' (*prajnā*) illuminates the 'darkness of ignorance' (*avidyā*) within, thereby liberating the self from its binding powers. When this light is personified, it is called Amida Buddha; and when it is localized, the Pure Land. This dynamic working of reality in the midst of *samsāra* accords with the basic definition of Buddhahood traditionally given in East Asia: 'self-awakening and awakening of others, the activity of awakening unfolding without end' (*jikaku-kakuta kakugyō-gūman*).

The transformation of a samsaric being of ignorance to a person of true awakening in the Pure Land mode involves a spiritual evolution which Shinran called the 'transformation through the three stages' (*sangan-ten'yu*), based on

the praxis stipulated in three basic vows among the forty-eight vows fulfilled by Amida Buddha: the 19th, 20th and 18th vows. To adapt the language of Kierkegaard's stages of life's way (see Kierkegaard, S.A.), the 19th may be called the ethical, the 20th religiousness 'A', and the 18th religiousness 'B'. The 19th vow requires of a devotee a deep commitment to enlightenment, good works and sincere desire for birth in the Pure Land. When such self-power activities are recognized ultimately as fruitless for the truly religious life, one turns to the 20th vow, devised for those resigned to their lack of spiritual capacity and relying wholly on the sole invocation of *nembutsu* (the holy name of Amida), a religious act said to be 'planting the roots of merits'. But sooner or later, this act too is recognized for what it is: another calculation (*hakarai*) of self-power, unconsciously rooted in the ego-self and prideful of cumulative invocation. The realization of the futility of spiritual fulfillment according to the prescriptions in the 19th and 20th vows opens up the self to the working of Other Power, a transformation experienced within the parameters of the 18th vow. The person now experiences a 'leap' (*ōchō*) not from the relative to the relative but from the relative to the absolute; the primary impetus originates with the absolute, whether it is called true compassion, Other Power or Amida Buddha (see Absolute, the). As a consequence, one is endowed with a trusting heart and mind (*shinjin*).

Such a person attains 'birth in the Pure Land', which for Shinran had two implications. First, at the moment of being endowed with a trusting heart one attains a new life, here and now. It is said that although our defiled bodies remain within *samsāra*, our hearts and minds play in the Pure Land. Second, when the karmic life on this earth is exhausted at death, one attains birth in the Pure Land, which is simultaneous with attaining Buddhahood. This means that now, as a fully awakened being of wisdom and compassion, one returns to the world of *samsāra* to work for the deliverance of all existence. 'Birth in the Pure Land' understood in these two ways manifests the dynamic relationship between time and timelessness. Since the Pure Land symbolizes timelessness (eternity), it is inseparable from time (*samsāra*) and manifests itself in time. A person thus can have an experience of timelessness while remaining in time as a karma-bound being. But when karmic life is exhausted and Buddhahood realized (timelessness), one cannot help but identify with every moment of time (*samsāra*) (see Eternity).

Going to the Pure Land and returning to *samsāra* are the fullest expressions of the life of *bodhisattva*, philosophically expressed as the simultaneous activity of ascent and descent. Both are made possible not by any human calculation but by the working of Other Power. This powerful working leads to the releasement of self-power, manifested in a person as humility, repentance and gratitude which form the core of a truly ethical life, established on the basis of interdependence and interconnectedness of all existence, both animate and inanimate.

See also: Buddhist philosophy, Japanese

TAITETSU UNNO

List of works

Shinran (1173-1263) *Kyogyoshinsho (Teaching, Practice, Faith and Realization)*, ed. and trans. Ueda Yoshifumi, *The True Teaching, Practice and Realization of the Pure Land Way*, Shin Buddhism Translation Series, Kyoto: Hongwanji International Centre, 1983-90, 4 vols.(This series contains Shinran's complete works, including his letters. Alternate translations are found in the Ryukoku Translation Series, published at Kyoto: Ryukoku University Translation Centre, 1961-86.)

References and further reading

Bloom, A. (1965) *Shinran's Gospel of Pure Grace*, Tucson, AZ: University of Arizona Press.(Outdated but still the only good one-volume work on Shinran available in English.)

Ueda Yoshifumi and Hirota, D. (1989) *Shinran: An Introduction to His Thought*, Kyoto: Hongwani International Center.(Outline of Shinran's thought, including original text and interpretations, based on the publications of the Shin Buddhism Translation Series.)

Unno Taitetsu (1996) *Tannisho: A Shin Buddhist Classic*, 2nd revised edn, Honolulu, HI: Buddhist Study Center Press.(An introduction to the thought of Shinran as recorded by one of his disciples.)

Shintō

Shintō means the 'way of the kami *(gods)' and is a term that was evolved about the late sixth or early seventh centuries - as Japan entered an extended period of cultural borrowing from China and Korea - to distinguish the amalgam of native religious beliefs from Buddhism, a continental import. Shintō embraces the most ancient and basic social and religious values of Japan. It is exclusively Japanese, showing no impulse to spread beyond Japan. The exportation of Shintō would in any case be exceedingly difficult since its mythology is so closely bound to the creation of Japan and the Japanese people, and since many of its deities are believed to make their homes in the mountains, rivers, trees, rocks and other natural features of the Japanese islands.*

Shintō comprises both great and little traditions. The great tradition, established in the mythology that was incorporated into Japan's two oldest extant writings, Kojiki (Record of Ancient Matters) and Nihon Shoki (Chronicle of Japan), both dating from the early eighth century, is centred on the imperial institution. According to the mythology the emperorship was ordained by the sun goddess, Amaterasu, who sent her grandson from heaven to earth (Japan) to found a dynasty 'to rule eternally'. The present emperor is the 125th in a line of sovereigns officially regarded, until Japan's defeat in the Second World War, as descended directly from Amaterasu.

Shintō's little tradition is a mixture of polytheistic beliefs about kami*, manifested in nature worship (animism), ancestor worship, agricultural cults, fertility rites, shamanism and more. Lacking a true scriptural basis, Shintō derives from the faith of the people, and from earliest times has had its roots firmly planted in particularistic, localistic practices. Thus it has always been strongest in its association with such entities as families, villages and locales (for example, mountains thought to be the homes of certain* kami *or, indeed, to be the* kami *themselves).*

1 *Kami*, people and nature

In Shintō belief there is no sharp distinction between *kami* and people - indeed some people, such as emperors, great heroes and ancestors, are regarded as *kami* - and both *kami* and people are joined together by their common habitation in the natural environment of the Japanese islands (although some *kami* in the mythology, including the sun goddess Amaterasu, reside in heaven). The human, natural and sacred realms merge. As the great eighteenth-century Shintō scholar Motoori Norinaga observed, '…the *kami* are spirits that abide in and are worshipped at the [Shintō] shrines. In principle human beings, birds, animals, trees, plants, mountains, oceans - all may be *kami*. According to ancient usage, whatever seemed strikingly impressive, possessed the quality of excellence, or inspired a feeling of awe was called *kami*' (quoted in Agency for Cultural Affairs 1972: 38).

Through Shintō, the Japanese people express their respect for all life as well as their appreciation of the beauties and powers of nature. And because certain elements and phenomena of nature, such as the sun, the moon, the wind and mountains, are regarded as *kami* and because nature as a whole is the principal home of the *kami*, nature is seen as 'good'. The Shintō view of nature has given a powerful stimulus to the arts and aesthetic tastes of the Japanese. Regarding nature as both beautiful and good, the Japanese have celebrated its qualities - especially the perishable beauties of the changing seasons, such as the cherry blossoms of spring and the maple leaves of autumn - in countless poems, paintings and other artistic media. Although Buddhism has also influenced the Japanese understanding of nature, the natural environment of Japan is first and foremost a world of Shintō (see Aesthetics, Japanese).

2 Worship of the *kami*

Kami are, for the most part, beneficent entities, requiring only that they be properly and respectfully attended to. Through worship, the people provide this attention and express their gratitude to the *kami*. The principal form of worship is the presentation of offerings to the *kami*, which can be rendered privately at home or publicly at a Shintō shrine. Items used as offerings include rice, other foods, *sake* and money.

People also say prayers and address petitions to the *kami*. The formal prayers, recited in archaic language by priests at shrines, typically praise the *kami* and express thanksgiving for the benefits they have bestowed. They may also include accounts of the lineages of the *kami* and describe the offerings being made to them. In their personal prayers and petitions, people make requests for such things as good health and personal success.

3 Purification

Shintō is a religion of rituals (see Ritual), and one of the central rituals is purification. The importance attached to purification in Shintō is readily observable, for example, in the procedures followed in preparing for worship and in worshipping. In approaching a shrine, one first performs ‘external’ or ‘outer’ purification - thus making oneself ready to enter the presence of the *kami* - by rinsing the mouth and hands with water at an ablution basin. Inside the shrine, the priest performs ‘internal’ or ‘inner’ purification in the form of exorcism by reciting prayers and waving a wand. Reflecting primitive, pre-ethical attitudes, Shinto regards both manifest defilements - such as bleeding, disease and death - and acts of evil or wrongdoing as ‘pollutions’ that must be ritually cleansed. Purification effects renewal: through purification, things are returned to their original states and proper identification with the *kami* is re-established.

Manifestations of the concern for purity and purification, deriving primarily from Shintō, can be observed in countless aspects of Japanese behaviour and life even today. The great attention given by the Japanese to bathing and general cleanliness is the most obvious of these manifestations, but they can also be found in such practices as the scattering of salt - a form of ritual purification - by sumo wrestlers before their bouts and the constant ‘cleansing’ of the utensils used in the tea ceremony. Although the tea ceremony, as it evolved during the medieval age, is generally thought to derive its spiritual essence from Buddhism, especially Zen, it is also thoroughly infused with Shintō or, more precisely, Shintō-like rituals of purification. We see this not only in the cleansing of the utensils but also in the precise and careful handling of water and fire, both of which are important agents of purification.

4 Shintō and Buddhism

According to the *Nihon Shoki (Chronicle of Japan)* in the early eighth century, Buddhism was introduced to Japan from Korea in 552. Undoubtedly the Japanese knew about Buddhism from a much earlier time, but it was in the middle or late sixth century that Buddhism was formally established in Japan, becoming in fact a principal carrier of continental culture during the long period of borrowing from China and Korea from the late sixth to mid-ninth centuries. It appears, as mentioned above, to have been the introduction of Buddhism to Japan that prompted the Japanese to coin the word Shintō for their native religious beliefs and practices. The construction of Buddhist temples in Japan also apparently inspired the Japanese to build sacred structures - which we call shrines - for Shintō (although there may have been some shrine construction from an earlier period). The continental architecture of Buddhism influenced Shintō architecture in various ways, particularly in later centuries. However, the structures of Shintō shrines (see §6) are based primarily on native styles and tastes that set them quite distinctively apart from Buddhist temples.

In addition to the architectural wonders of its temples, Buddhism also brought to Japan new types of religious paintings and statuary. Buddhist iconography is extremely rich and varied, and became a major component of art in Japan in succeeding centuries. Yet this is an area in which Buddhism exerted little influence on Shintō, for Shintō, with only minor exceptions, has not conceived of its deities in visual terms: that is, it has made virtually no attempt to portray the *kami* as entities based on, say, anthropomorphic or zoomorphic prototypes.

Buddhism’s view of the world could scarcely differ more from that of Shintō. Whereas Shintō celebrates life - and hence this existence - in all its manifold forms and creative vitality, Buddhism looks upon the world with the deepest pessimism as a place of incessant suffering. Suffering is caused because all things in the world are in flux, in a constant process of change (see Mujō). People seek to acquire things, to hold onto them, but this is impossible because of the impermanence of everything: nothing is substantial, nothing real. The result of human acqusitiveness is, inevitably, suffering. Although the various sects of Buddhism differ in the methods and the paths they recommend to their followers, they share the goal of achieving release from the cycle of life and death and hence from suffering (see Suffering, Buddhist views of origination of).

It is certainly an oversimplification to say, as has often been said, that Shintō is concerned with life and Buddhism with death, but there is a rough truth to this statement. Although Shintō mythology describes an underworld to which people must go upon death, essentially it has developed no theology about the afterlife. Death is a form of pollution, which like other pollutions requires ritual purification. Very likely a major reason why Shintō did not develop a theology of the afterlife is because Buddhism, upon its arrival in Japan, pre-empted this field. In any

case, Buddhism handles death - through funerals and periodic rites of mourning - to the point where its priests have in recent times been unflatteringly labelled the ‘undertakers of Japan’.

One reason, then, why Shintō and Buddhism have been able to coexist in Japan through the centuries is because they deal with essentially different realms, roughly categorizable as life and death. Another reason is that they have been partially syncretized. Syncretism is characteristic of East Asian religion and thought in general, and is certainly observable in Buddhism where, for example, we find the ninth-century founder of Shingon (True Word) Buddhism, Kūkai, writing a tract entitled *Jūjūshinron (The Ten Stages of Religious Consciousness)*, in which he ranks Confucianism, Daoism, and various sects of Buddhism in a hierarchy of ascending ‘consciousness’ culminating, at the highest level, in Shingon.

In both India and China, Buddhism recognized the ‘deities’ of other religions or systems of belief, including Confucianism and Daoism in China, by considering them as avatars of buddhas and bodhisattvas (see Buddhist philosophy, Indian; Buddhist philosophy, Chinese). It did this according to the concept of *honji suijaku* (original substance manifests traces), which became particularly important in Japan in regard to the association of Buddhism and Shintō. From the standpoint of Japanese Buddhists, *honji suijaku* meant that the deities of Buddhism (the substance) manifested themselves as the *kami* (traces) of Shintō. Thus, for example, the sun goddess Amaterasu was thought to be the ‘trace’ of Shingon’s cosmic Buddha, known in Japanese as Dainichi (Great Sun).

Fundamental to *honji suijaku* thought during the early centuries after its introduction to Japan was the belief that the buddhas and bodhisattvas were primary and the deities of Shintō were secondary. In the medieval age (1185-1573), however, there arose a movement within Shintō - as part of a larger Shintō revival - that sought to reverse this relationship, declaring the Shintō deities to be primary and the buddhas and bodhisattvas secondary. A good example of this reversal of the *honji suijaku* relationship can be observed in the late thirteenth century, when two attempts by forces of the Mongol dynasty of China to invade Japan, in 1274 and 1281, were defeated by typhoons that devastated the Mongol fleets. The Japanese believed these storms were *kamikaze*, ‘winds of the gods’, that were sent to protect Japan in times of greatest peril. Whereas for centuries the guardian deities of Buddhism had been looked upon as the principal protectors of the state, the function of state protection was thereafter shifted to the *kami* of Shintō.

5 The imperial institution

The mythology of Shintō proclaims that the sun goddess Amaterasu mandated her grandson - his name was Ninigi - to descend to Japan and establish a dynasty to rule the country eternally (although the actual dynastic founder and first emperor was Ninigi’s grandson, Jimmu). In any case, belief that Japan would be ruled forever by a single royal line became one of the most powerful myths in Japanese history, a myth that retained its potency as a tool of rule until it was shattered by Japan’s catastrophic defeat in the Second World War. Although through most of Japanese history the emperor has in fact functioned more as a ‘sacred legitimizer’ than an actual ruler (he has reigned rather than ruled), he, more than any other person or thing, has through the centuries symbolized ‘Japan’.

The significance of the imperial institution to Japan and the Japanese was given important new definition during the Shintō revival in the early medieval period. Following upon the assertion that the *kami* of Shintō, rather than the deities of Buddhism, provided protection for the state at the highest level, the fourteenth-century courtier Kitabatake Chikafusa declared in the opening lines of his history of Japan, *Jinnō Shōtōki (Chronicle of the Direct Descent of Gods and Sovereigns)*:

> Great Japan is the divine land. The heavenly progenitor founded it, and the sun goddess bequeathed it to her descendants to rule eternally. Only in our country is this true; there are no similar examples in other countries. This is why our country is called the divine land.
>
> (Varley 1980: 49)

Here Japan is extolled as a country superior to others because it has been ‘ruled’ from its founding by an unbroken dynastic line of sovereigns. As Chikafusa observed, other countries- such as India and China - have often been without sovereigns or have suffered frequent dynastic changes. The belief in the superiority of Japan and the Japanese centred on Japan’s ‘unique’ imperial institution, implicit in Chikafusa’s declaration, was sustained into the modern age, becoming the core of the *kokutai* (national polity) ideology that was officially promoted by

Japanese governments until the end of the Second World War (see Sovereignty).

6 The shrine

Probably the best known symbol of Shintō is the *torii* or 'bird perch' entranceway to a shrine. This entranceway supposedly has its origins in the mythology, where we read that a perch was erected for a cock to crow and signal the commencement of entertainment organized by the other gods in order to lure Amaterasu out of a cave in which she had secluded herself.

Shintō shrines are of a great variety, constructed in many diverse locations. At the base of society, in the villages and towns, shrines are found primarily in natural settings removed from most human habitation. These shrines seem to merge with nature in the same way that the human realm merges with the natural and sacred realms (see §1).

Shrines are the homes of *kami* and the places where offerings are made and prayers recited. Although some shrines are quite elaborate, comprising many buildings and covering extensive areas, the shrine at its simplest has an inner sanctuary, where the symbol of the resident *kami* (or symbols, if the shrine houses more than one *kami*) is housed, and an outer area for offerings and prayers. All shrines also have an ablution basin (see §3) where worshippers perform external purification. Worshippers do not enter the shrine buildings; only priests are allowed in them.

Although it may be difficult to perceive it in those located in congested cities, the shrine is always designed to evoke a sense of mystical unity with the *kami* and indeed with the land of Japan itself, the home of the *kami*. Shintō, represented by the shrine, constitutes the very basis of Japanese civilization.

See also: Aesthetics, Japanese; Buddhist philosophy, Japanese; Japanese philosophy; Motoori Norinaga; Religion, philosophy of

PAUL VARLEY

References and further reading

Agency for Cultural Affairs (ed.) (1972) *Japanese Religion*, Tokyo: Kodansha.(Introduction to Japanese religion.)

Grapard, A.G. (1983) 'Shintō', *Kodansha Encyclopedia of Japan*, vol. 7, Tokyo: Kodansha.(Good introduction to the subject.)

Ono Sokyo (1962) *Shintō, the Kami Way*, Tokyo: Charles E. Tuttle Co.(More detailed description of Shintō.)

Varley, H.P. (trans.) (1980) *A Chronicle of Gods and Sovereigns: The Jinnō Shōtōki of Kitabatake Chikafusa*, New York: Columbia University Press.(Translation of Kitabatake's history of Japan.)

Shōtoku Constitution

The Shōtoku Constitution is the earliest fundamental political document of Japan. Promulgated in AD *604, it is ascribed to the regent Shōtoku, who was also a devout Buddhist and philosopher. The document reflects the influences of Confucianism, Buddhism, Daoism and Legalism in its various provisions; it is strongly marked by Chinese thought rather than being influenced by Shintō. Not a constitution in the modern sense, the document is rather a set of ideals, guiding principles and basic requirements for those in government. As well as helping to lay the foundation for a unified Japan, the Constitution also marks the beginning of a period of assimilation of Chinese culture and philosophy.*

The Shōtoku Constitution was, according to tradition, promulgated in AD 604. Also called the Constitution in Seventeen Articles, it is ascribed to Shōtoku (574-622), a philosopher and a devout Buddhist as well as a statesman, who served as regent of Japan under the Empress Suiko. Before this time, Japan had not existed as a nation but had been divided into autonomous regions and clans, whose chiefs vied with one another for hegemony. With his Constitution, Shōtoku sought to establish Japan as a unified nation under the sole authority of the imperial throne, and to ground its conduct of government upon universal principles which he found in Buddhism and classical Chinese philosophy.

The Shōtoku Constitution is not a constitution in the modern and technical sense of a set of principles formally limiting the extent of legislation and formally grounding the legality of other laws. It is instead a set of ideals, guiding principles and basic requirements for the officials who exercise governmental powers.

The Constitution consists of seventeen articles, as follows:

Article 1 declares harmony to be a principle of governance within a hierarchical order and promotes the discussion of public affairs.
Article 2 exhorts the reverence of the Three Treasures of Buddhism, namely the Buddha, the Law and the Community of Practitioners, as ‘the final refuge of all living beings’ and ‘the ultimate foundation of all nations’.
Article 3 compares the lord to Heaven, the vassal to Earth and enjoins the vassal to obey the imperial commands scrupulously.
Article 4 enjoins the ministers and functionaries to base their conduct upon decorum.
Article 5 warns against the settling of law-suits under the influence of bribery.
Article 6 emphasizes the importance of moral integrity on the part of officials, and admonishes them not to resort to self-promoting deceptions.
Article 7 lays down the policy of honouring the wise in the management of bureaucracy.
Article 8 seeks to impress officials with the importance of diligence and punctuality.
Article 9 declares trustworthiness among vassals to be ‘the basis of righteousness’ on which the success or failure of government depends.
Article 10 calls for the cessation of anger, resentment and self-righteousness: ‘For everybody has a heart, and every heart has its own attachment. Their right is our wrong, our right, their wrong.’
Article 11 stresses the importance of reward and punishment according to deserts.
Article 12 forbids provincial governors and district administrators to levy taxes for themselves, and reserves taxation as a prerogative of the sovereign.
Article 13 commands the officials to execute public affairs responsibly, stating that their absence, for reasons official or personal, does not excuse them from their duties.
Article 14 points out the evils of envy at the sight of the promotion of those with superior intelligence or ability.
Article 15 calls for the setting aside of private motives, which otherwise would lead to resentment and discord.
Article 16 states that the conscription of labour for public works should not interfere with the peasants’ agricultural production.
Article 17 points out the need of public discussions of weighty public affairs.

In this set of principles the influence of Confucianism is prominent. Harmony, decorum and trustworthiness, which are the bases of Articles 1, 4 and 9, are central elements of Confucian ethics (see Confucian philosophy, Japanese; Confucian philosophy, Chinese). Article 2 is striking in its commitment to Buddhism; Buddhists vow to take

refuge in the Three Treasures (see Buddhist philosophy, Japanese). Articles 5, 11 and 12, concerned respectively with the fair enforcement of laws, reward and punishment, and taxation, betray the influence of Legalism (see Legalist philosophy, Chinese). The call to abandon desires in Article 5 and the relativity of conventional valuation in Article 10 sound Daoist (see Daoist philosophy). In contrast with the marked presence of Chinese thought in the Constitution, the influence of Shintō, the indigenous folk religion, is characteristically absent.

The philosophical interest and significance of the Shōtoku Constitution lie primarily in the following three areas:

(1) Exactly what universal principles of East Asian continental thought were applied, and how were they applied, to the political situation in the southwestern half of the Japanese archipelago at the formative stage of Japan in the early seventh century? Did the urgent practical need for political reform dictate the utilizations of ideas and ideals of political philosophy as instruments, or did the political situation serve as the medium through which to implement the ideas and ideals?
(2) How are the ideas and ideals culled from the different sources of Mahāyāna Buddhism, Daoism, Confucianism and Legalism related one to another in the ways in which they inform the Constitution? Do the influences of Confucianism and Buddhism, for instance, simply coexist side by side, or are they reconciled with each other and integrated into some coherent pattern? If so, what is that pattern?
(3) How did the Constitution, the earliest politically authoritative articulation of an ideology, set a precedent for the subsequent development of Japanese thought and culture?

On these three issues, no clear scholarly consensus has yet emerged. However, scholars have to a considerable extent overcome healthy scepticism concerning the prior issues about the date and the authorship of the Constitution, and have canvassed a vast body of Chinese literature, including Chinese versions of Mahāyāna *sūtras*, in order to identify the sources of verbal borrowings, echoes and allusions with which the text of the Constitution abounds. The entire text of the Constitution, written in highly stylized classical Chinese, has been preserved as a quotation in the *Nihon shoki (Chronicles of Japan)*, and is there attributed to Shōtoku.

Among other achievements of the Shōtoku regency are the construction of Buddhist temples, the writing of commentaries on three Mahāyāna *sūtras*, the institution of the Twelve Cap-ranks of bureaucracy and the inauguration of official envoys to China. The Constitution, together with these achievements, contributed much to the transition of Japan from a divided tribal society to a united nation, the adoption and assimilation of Chinese civilization including philosophy, and the transformation of Japan into a land of Buddhism.

See also: Buddhist philosophy, Japanese; Bushi philosophy; Legalist philosophy, Chinese; Confucian philosophy, Japanese; Daoist philosophy; Japanese philosophy; Shintō

YUKIO KACHI

References and further reading

The text of the Shōtoku Constitution is in *Nihon shoki (Chronicles of Japan)* (AD 720), ed. Ienaga Saburō, Inoue Mitsusada, Ōno Susumu and Sakamoto Tarō, Tokyo: Iwanami, 1968, vol. 2, 180-7; trans. in W.T. de Bary, D. Keene and Tsunoda Ryūsaku, *Sources of Japanese Tradition*, New York: Columbia University Press, 1958, 49-53. (The latter gives an English translation of the Shōtoku constitution.)

Anesaki, M. (1930) *History of Japanese Religion*, London: Kegan Paul, Trench, Trubner, 57-65.(Sees the Constitution, an expression of Shōtoku's political philosophy, as an application of Buddhist ideals.)

de Bary, W.T (ed.) (1969) *The Buddhist Tradition in India, China and Japan*, New York: The Modern Library, 255-65.(Articulates the views of Anesaki (1930).)

Kitagawa, J.M. (1987) *On Understanding Japanese Religion*, Princeton, NJ: Princeton University Press.(Studies of the patterns of blending of Buddhism, Chinese schools of thought, and Shintōin Japan.)

Sansom, G. (1958) *A History of Japan to 1334*, Stanford, CA: Stanford University Press, 40-81.(Includes a history of Japanese thought from the fifth to the seventh century, inclined to deny Shōtoku's authorship of the Constitution.)

Sonoda, K. and Brown, D.M. (1993) 'Early Buddha Worship', in J.W. Hall, M.B. Jansen, Madoka Kanai and D. Twitchett (eds) *The Cambridge History of Japan*, vol. 1, *Ancient Japan*, Cambridge: Cambridge University Press.(An up-to-date and judicious survey of scholarly opinions on Shōtoku's authorship of the 'Injunctions' and also on the issue of priority among the diverse strands of thought involved. The context of this survey is the placement of Shōtoku in the early phase of the history of Buddhism in Japan.)

Umehara, T. (1993) *Shōtoku Taishi (Prince Shōtoku)*, Tokyo: Shūei-sha, 4 vols.(A close study of the structure of the Constitution with a view to determining the pattern of conceptual integration underlying the document; see especially vol. 2, 340-533.)

Shpet, Gustav Gustavovich (1879-1937)

Gustav Shpet was the first Russian philosopher to take up Edmund Husserl's idea of pure phenomenology as prima philosophia *and develop it in several directions. Thus, in his most important phenomenological work Iavlenie i smysl (Appearance and Sense) (1914), he outlined his 'phenomenology of hermeneutic reason' on the basis of Husserl's* Ideas 1. *In this theoretical framework he formulated, between 1914 and 1918, hermeneutic and semiotic problems, which in the 1920s he elaborated more specifically within the fields of philosophy of language and theory of art. In doing so he took up ideas from other philosophical movements, particularly Dilthey's hermeneutics and Wilhelm von Humboldt's philosophy of language.*

Shpet's reception of phenomenology has to be seen in the context of Russian intellectual and cultural life during the first two decades of the twentieth century. The Platonic 'Moscow Metaphysical School' (which included V. Solov'ëv and S. Trubetskoi) provided the intellectual atmosphere in which his turn to Husserl's phenomenology took place, and his ideas on theories of language and signs are close to ideas of the contemporary movement of Russian Formalism. Through his phenomenological and structural theories he influenced Prague structuralism via the 'Moscow Linguistic Circle', and is seen as a precursor to Soviet Semiotics.

1 Life

Gustav Shpet, founder of a phenomenological movement in Russia, was born in Kiev. He studied there at the Vladimir University from 1901-5 and finished his studies with a monograph entitled *Problema prichinnosti v Iume i Kante (The Problem of Causality in Hume and Kant)*. In 1907 he moved to Moscow, and taught at Moscow University from 1910 onwards. During a stay in Göttingen (1912-13) where he studied with Husserl he turned to Husserl's transcendental phenomenology. His first phenomenological publication *Iavlenie i smysl (Appearance and Sense)* (1914) marked the beginning of a productive reception of Husserl's phenomenology in Russia. In 1916 he had his viva on his Master's thesis *Istoriia kak problema logiki (History as a Problem of Logic. Part I)*. In 1918 he finished his manuscript on *Germenevtika i eë problemy (Hermeneutics and its Problems)*, in which he critically presented the problems of hermeneutics as they have developed throughout history from antiquity (especially in Origen and Augustine) to modern times, thereby at the same time elaborating the basic outline of his 'hermeneutical philosophy': a philosophy which is caught in the field of tension exerted, on the one side, by Husserl's 'Phenomenology of Reason' and, on the other, by Dilthey's 'Philosophy of Life'.

After the Revolution of 1917, Shpet was active in many different fields. He received a professorship of philosophy at Moscow University. In 1920 he joined the 'Moscow Linguistic Circle' (MLK), a centre of Russian Formalism (see Russian literary Formalism), and in 1921 was appointed director of the Institute for Scientific Philosophy, a new research institute at Moscow University. Expelled from there in 1923 for political reasons, he concentrated his activities on the State Academy of the Arts (GAKhN), whose vice-president he was until 1929, and whose Department of Philosophy he temporarily led.

His most important contributions to the theory of art and language are his *Ėsteticheskie fragmenty (Aesthetic Fragments)*, published in 1922 and 1923 in Petrograd, and *Vnutrenniaia forma slova (The Internal Form of the Word)* (1927). *Ėsteticheskie fragmenty* includes a phenomenology of 'living discourse' and an analysis of those rules which determine the constitution of meaning in poetic discourse. These phenomenological and structural analyses of language, which aim to construct a poetics, were further developed through a critical assessment of Humboldt's philosophy of language in Shpet's last substantial work *The Internal Form of the Word*.

During a 'cleansing' of the GAKhN in 1929, Shpet was forced to retire, but was still able to work for a while as a translator, editor and critic. It was then that he translated Dickens, Byron and Shakespeare into Russian. In March 1935 he was arrested by the NKVD and charged with having led an anti-Soviet group during his time at the GAKhN in the 1920s. After a lengthy detention he was exiled for five years to Eniseisk, later to Tomsk. Here in 1937 he finished his Russian translation of Hegel's *Phenomenology of Spirit*. In October of the same year he was again arrested and shot by the NKVD.

2 Development towards phenomenology

Representative of Shpet's notion of philosophy before his turn to phenomenology (and of expectations he then had

for a future reform of philosophy and psychology) is his article ‘Odin put’ psikhologii i kuda on vedët’ (One Method of Psychology and Where It Leads To), published in 1912. He criticized the prevailing trend of an experimental and explanatory psychology for having replaced ’lively and concrete facts’ with ’empty schemata and abstractions’. Only a descriptive psychology which focuses on the pure data of consciousness would be able to fathom psychic life in its concreteness and totality. He saw the basis for such a new direction in psychological thinking in Wilhelm Dilthey’s *Ideas on a Descriptive and Analytical Psychology* (1894) (see Dilthey, W.) and he argued that there was a particular style of philosophizing, that would correspond to this new type of philosophy which takes into account the totality of psychic life: a ’realistic metaphysics’ whose task it would be to grasp ’the real in its true essence and its totality’. Shpet thought that such a philosophy, which draws on the evident facts of ’inner experience’, had been realized in important movements of nineteenth and early twentieth century Russian philosophy. Philosophers of the ’Moscow Metaphysical School’ are mentioned as typical exponents of this trait in Russian thought (see Solov’ëv. V.S.).

Another, no less important, influence on Shpet’s reception of Husserl was his interest in the logic of the historical sciences. During his stay in Göttingen (1912-13) he found in Husserl’s phenomenology the theory he had been looking for, and his hermeneutical interest motivated him to try to develop Husserl’s ’Phenomenology of Reason’, as outlined in *Ideas Pertaining to a Pure Phenomenology, volume 1* (referred to as *Ideas 1*) into a theory of hermeneutic reason which focuses on the problem of understanding signs.

Although the ideas Shpet encountered in Göttingen primarily concerned *transcendental* phenomenology - the seminar on ’Nature and Spirit’, which Shpet attended with other influential phenomenologists like Roman Ingarden and Hans Lipps, certainly met his hermeneutical interests - on the other hand, the *ontological* trend in the intellectual atmosphere among Husserl’s fellow students in Göttingen also has to be taken into account (see Phenomenological movement §§3-4).

3 Concept of phenomenology

Shpet’s encounter with Husserl’s phenomenology in the light of his expectation of a reform of philosophy and psychology leads to a peculiar notion of phenomenology, which is documented in *Iavlenie i smysl (Appearance and Sense)* (1914). On the one hand, the Russian phenomenologist tries to reconstruct Husserl’s noetic-noematic studies within the framework of an ontological question, based on the Neoplatonism of the ’Moscow Metaphysical School’; on the other he demonstrates the incompleteness of Husserl’s analyses of objects, as given in *Ideas 1*, and completes these analyses with his own. The ’noematic sense’ intended in acts of consciousness, as presented by Husserl, presupposes for Shpet a class of intentional experiences hardly dealt with in *Ideas 1*; acts of consciousness through understanding, which play a role in the constitution of all classes of concrete objects. The structure of these ’hermeneutic acts’ is illustrated by a range of phenomena which are only of minor importance in *Ideas 1*: by the mode of appearance of items of practical use, the specific character of historical sources and the understanding of linguistic utterances. Thereby Husserl’s ’Phenomenology of Reason’ meets Shpet’s endeavour to find a basis for historical cognition in scientific logic, and eventually to prepare a grounding of the humanities, in other words, to prepare an analysis of their conceptual frameworks and most important methods.

Appearance and Sense marks the early phase of Shpet’s reception of Husserl. Here he outlines his project of a ’phenomenology of hermeneutical reason’. The ensuing works on hermeneutics, philosophy of language and theory of art, published or written between 1916 and 1927, can be seen as a further development of such a hermeneutical phenomenology, whose leading idea is the correlation of signs (as combinations of expression and meaning) and sign-interpreting consciousness.

Shpet also characterizes his project as a semiotic ’Philosophy of Culture’ in which language, art, myths and manners are to be described as systems of signs. He develops the basic model of a sign out of Husserl’s concept of linguistic expression, which acts as prototype for all other forms of signs. The idea of a ’purely logical grammar’, which formulates laws for the grammatical meanings of natural languages, should be applied analogously to all the other cultural systems.

4 Hermeneutics, philosophy of language and poetics

The concrete form of Shpet’s phenomenology of hermeneutical reason in his philosophy of language and his

poetics was also much influenced by Dilthey's 'Philosophy of Life'. In his hermeneutical philosophy, as outlined in the manuscript *Germenevtika i eë problemy (Hermeneutics and its Problems)*, he worked with Schleiermacher's, Boeckh's and Dilthey's theories of understanding (see Hermeneutics §§2-4; Schleiermacher, F.D.E.). Above all, he tried to deepen and refine Dilthey's late grounding of the humanities - then the culmination in the development of hermeneutics - with insights in the domain of semiotic theories, which he found not only in Husserl's first *Logical Investigation* on *Expression and Meaning*, but also in other semantic works of the Brentano School (Marty and Meinong especially). A combination of Husserl's semantics with Dilthey's hermeneutics would be an enrichment for both sides, as Shpet wrote at the end of the manuscript. The theory of understanding could find a new answer to the question of the mutual relation of the different methods of interpretation, whereas semantics would experience in this combination a 'philosophically lively and concrete embodiment'.

This actualization of Husserl's semantics, with a hermeneutical intention, has left its traces in Shpet's *Ėsteticheskie fragmenty*. With this tripartite work he entered the contemporary discussions on literary theory, as initiated by Russian Formalism. He was particularly concerned with the definition of the specific character of poetical discourse as opposed to others, be they scientific, rhetorical or everyday discourses. If one puts this question phenomenologically, one has to ask under what condition a linguistic utterance appears as artistic or poetic to a listener or reader. Since a poetic utterance is experienced only as a contrast to everyday use of language, one has to analyse first the reception of everyday language. This is the procedure Shpet follows in the second part of *Ėsteticheskie fragmenty*. In the description here of the various forms and aspects of the understanding of language, the difference between understanding the message and understanding the author of the message plays a pivotal role.

Here only the description of the understanding of language is called a phenomenological analysis. In opposition to this, Shpet presented his description of the structure of linguistic expression as 'ontology of the word', which he, in turn, subsumed under a general theory of semiotics. In this confrontation between a phenomenological inquiry, which is confined to the side of experience, and an ontology, which focuses on the object, the ever increasing influence of Husserl's early concept of phenomenology on Shpet becomes visible. In Shpet's 'ontology of the word' a particular concept of structure is of central importance. 'The structure is a concrete construction whose individual parts can vary in their extent and even in their quantity, but not a single part of the whole *in potentia* can be removed without destroying this whole' (1922-3: II, 11). By 'structure of the word' Shpet did not mean the morphological, syntactic or stylistic construction, in short, not the arrangement of linguistic units 'in the plain', but 'the organic, depth-wise, as it were, arrangement of the word - from the sensually conceivable wording to the eidetic object' (1922-3: II, 11). Therefore the word-structure consists not only of the relations between wording and meaning, but also of the relations between meaning and 'object' (the latter, as ideal object, being ontologically distinguished from concrete individual things).

When Shpet spoke of the structure of the *word*, he took it in the wide sense of the Russian expression for 'word', namely 'slovo'. 'Slovo' is used not only for words but also for groups of words, sentences and combinations of sentences, even for literary texts and for the whole of a natural language. Shpet used it with all these different meanings, yet was mainly concerned with the 'communicating word': meaningful discourse able to convey something to another person. Thereby Shpet took up Plato's definition of predicative statements, as the 'shortest and most simple *logos*'. Shpet saw its structure as follows:

In a simple predication the expression in subject position denotes a concrete, individual object; the predicate indicates a property belonging to this object. How such a sentence, consisting of a proper name (or other denominations) and a predicate, can have a sense and refer to objects, is clarified by Shpet through an analysis of the use of these expressions. In denoting speakers refer to a thing, in predication they say something about it. What can be said about this thing, and, conversely, which predications are impossible, is determined by the species to which the thing belongs. Therefore an act intending this species, which Shpet called also the 'eidetic object', is constitutive for the construction of a meaningful predication.

The central point of Shpet's theory of predication is that the meaning of the predicate can only be given in the act of predication, in so far as an assertion of meaning only happens in the interplay of the denoting of individuals and the act intending an eidetic object.

With these definitions Shpet outlined the 'word's' structure, which is common to factual or scientific

communicating, to rhetorical and to poetic discourse. The key to the specification of this general structure in the artistic and literary usage and formation of words, sentences and combinations of sentences, is to be found in Shpet's theory of linguistic functions; a theory which he developed through a critical assessment of Husserl's and Marty's philosophy of language. He started from three different functions of language which can be fulfilled by each type of discourse. Thereby in each case one function dominates the others. These three functions are the factual-communicating, the expressive and the poetic, in other words, the function working through the creative formation of language. (The idea of the creative character of language, which is most obvious in poetic discourse, is common to both Russian Formalism and Shpet's poetics, and can be traced back to Wilhelm von Humboldt's concept of language as *'Energeia'* (see Humboldt, W. von).

Depending on which of the three functions is dominant, discourse is either scientific (concerned with factual communication), rhetorical (concerned with influencing the emotions), or poetic (primarily concerned with the arrangement of linguistic expressions as such). The predominance of one of the three functions implies in each case a different mutual relation between the parts of the word-structure. Whereas, for example, in everyday language the arrangement on the level of expression aims primarily at structuring the expressed meaning, and thereby at the communication of facts, in poetic discourse all levels gain a relative importance on their own. The rhythmic forms and syntactic particularities of this discourse should be noticeable as such. On the other hand, the meaning expressed in poetic discourse is more dependent on those external forms of language: whereas the meaning of a factual - above all scientific - communication is not affected by each change of wording and syntactic arrangement, the 'poetic meaning' is far more sensitive to such changes.

5 Turning away from Husserl's concept of language

By giving pure logic, which deals with the condition of the possibility of science, a phenomenological foundation, Husserl excluded important aspects of living discourse. Shpet's project was more extensive than Husserl's in that he analysed scientific discourse as a possible form of discourse next to poetic, rhetoric and everyday discourse. This widening of the horizon entails a turning away from (not only a modification of) Husserl's concept of language. This becomes obvious when Shpet questioned a central presupposition of *Logical Investigations* - that scientific discourse can be marked off from living discourse. These two kinds of discourse are only tendencies, as Shpet emphasized; they are not fully realized in any empirical speech sample: 'Figurativeness is not only a trait of "poetry"… it is a general property of language, which belongs to scientific discourse, too'. 'The figurative and the literal can only be distinguished by tendency' (1922-3: III, 32).

The thesis of the irreducibility of figurative-ambivalent discourse has to do with Shpet's emphasis on the fact that thought is inseparably bound to language. With this concept of language as 'formative organ of thought', as outlined by Shpet in his interpretation of Humboldt of 1927, he turned away most clearly from Husserl's *Logical Investigations*, according to which 'the fact of being expressed is arbitrary for the meanings'.

Translated from the German by P. Schnyder
ALEXANDER HAARDT

List of works

Shpet, G. (1907) *Problema prichinnosti v Iume i Kante (The Problem of Causality in Hume and Kant)*, Kiev: Universitet St Vladimira.(Shpet's first work, a monograph.)

Shpet, G. (1912) 'Odin put' psikhologii i kuda on vedët' (One Method of Psychology and Where It Leads To), in *Filosofskii sbornik L.M. Lopatinu*, Moscow, 245-64.(Representative of Shpet's notion of philosophy before his turn to phenomenology.)

Shpet, G. (1914) *Iavlenie i smysl. Fenomenologiia kak osnovnaia nauka i eë problemy*, Moscow; trans. T. Nemeth, *Appearance and Sense. Phenomenology as the Fundamental Science and its Problems*, Dordrecht, Boston, London: Kluwer Academic Publisher, 1991.(Shpet outlines his 'phenomenology of hermeneutical reason'.)

Shpet, G. (1916) *Istoriia kak problema logiki (History as a Problem of Logic)*, Moscow.(Shpet's master's thesis.)

Shpet, G. (1918) *Germenevtika i eë problemy (Hermeneutics and Its Problems)*, in *Kontekst*, Moscow; German translation, A. Haardt and R. Daube-Schackat (eds), *Die Hermeneutik und ihre Probleme*, Freiburg, Munich: Alber, 1993.(Critical presentation of the problems of hermeneutics as they have developed throughout history.)

Shpet, G. (1922-23) *Ėsteticheskie fragmenty I-III (Aesthetic Fragments)*, Petrograd: Kolos; 2nd edn in *Sochineniia*

(Works), Moscow: Pravda, 1989, 343-472.(Shpet's contribution to the theory of art and language.)
Shpet, G. (1927) *Vnutrenniaia forma slova (The Internal Form of the Word)*, Moscow: GAKhN.(A critical assessment of Humboldt's philosophy of language.)

References and further reading

Haardt, A. (1993) *Husserl in Rußland. Phänomenologie der Sprache und Kunst bei Gustav Shpet und Aleksej Losev (Husserl in Russia. Phenomenology of Language and Art in Gustav Shpet and Aleksei Losev)*, Munich: Fink.(Essential background reading for an understanding of the Phenomenological Movement in Russia).

Sidgwick, Henry (1838-1900)

Henry Sidgwick was a Cambridge philosopher, psychic researcher and educational reformer, whose works in practical philosophy, especially The Methods of Ethics (1874), brought classical utilitarianism to its peak of theoretical sophistication and drew out the deep conflicts within that tradition, perhaps within the age of British imperialism itself. Sidgwick was profoundly influenced byJ.S. Mill, but his version of utilitarianism - the view that those social or individual actions are right that maximize aggregate happiness - also revived certain Benthamite doctrines, though with more cogent accounts of ultimate good as pleasure, of total versus average utility, and of the analytical or deductive method. YetSidgwick was a cognitivist in ethics who sought both to ground utilitarianism on fundamental intuitions and to encompass within it the principles of common-sense ethics (truthfulness, fidelity, justice, etc.); his highly eclectic practical philosophy assimilated much of the rationalism, social conservatism and historical method of rival views, reflecting such influences as Butler, Clarke,Aristotle, Bagehot, Green, Whewell and Kant. Ultimately, Sidgwick's careful academic inquiries failed to demonstrate that one ought always to promote the happiness of all rather than one's own happiness, and this dualism of practical reason, along with his doubt about the viability of religion, led him to view his results as largely destructive and potentially deleterious in their influence.

1 Life

Henry Sidgwick's entire life fell within the reign of Queen Victoria and his entire career within the domain of Cambridge University, where he went from a brilliant undergraduate performance at Trinity College to become Knightbridge Professor of Moral Philosophy in 1883. He began his career as a classicist, but throughout his life extended his research and teaching into new disciplines: primarily moral theory, epistemology and metaphysics, political economy and political theory. His impact on the moral sciences at Cambridge was profound, and he was the guiding spirit behind Newnham College, one of England's first women's colleges. A long-standing member of the elite Cambridge discussion group, the Apostles, Sidgwick taught in a style that was, as W.R. Sorley put it, 'a training in the philosophical temper - in candour, self-criticism and regard for truth'. His best-regarded work has always been *The Methods of Ethics* (1874), but he also wrote major treatises on *The Principles of Political Economy* (1883) and *The Elements of Politics* (1891), as well as a primer on the history of ethics and numerous essays and reviews. Various posthumous works also appeared, thanks to the efforts of Eleanor Mildred Sidgwick (née Balfour), whom he had married in 1876 and who collaborated with him on many projects.

The sole interruption in Sidgwick's Cambridge career came in 1869, when he resigned his fellowship because he could no longer in good conscience subscribe to the Thirty-Nine Articles as required by law. Although he kept a lectureship and the tests were abolished in 1871, this was a formative event in his life, both stimulating and prefiguring his struggles with the foundations of ethics - as shown by his pamphlet on *The Ethics of Conformity and Subscription* (1870). His years of 'storm and stress' over religious questions led him from historical biblical criticism to philosophy and psychic research. He was a founder and the first president of the Society for Psychical Research, and his scientific investigations of purported psychical phenomena complemented his philosophical investigations into the 'deepest problems of human life', since he thought that evidence for personal survival of death might provide rational grounds for religious belief and ethical conduct.

2 Ethics and utilitarianism

In composing *The Methods*, Sidgwick thought of himself as emulating Aristotle, whose *Ethics* had 'reduced to consistency by careful comparison' the common-sense morality of Greece, treated not from the outside but as what he and others thought upon reflection. He wants to 'do the same for *our* morality here and now, in the same manner of impartial reflection on current opinion'. His aim is less practice than knowledge, merely 'to expound as clearly and as fully as my limits will allow the different methods of Ethics that I find implicit in our common moral reasoning; to point out their mutual relations; and where they seem to conflict, to define the issue as much as possible'.

Sidgwick's reflections often clash with earlier traditions, including utilitarianism, the empiricism, psychological egoism and reductionism of which he rejected. He maintains that the basic concept of morality - 'ought' or 'right' - is unique and irreducible, *sui generis*. Also, moral approbation is 'inseparably bound up with the conviction,

implicit or explicit, that the conduct approved is "really" right - that is, that it cannot, without error, be disapproved by any other mind'. He describes these 'dictates' or 'imperatives' in a largely internalist fashion as 'accompanied by a certain impulse to do the acts recognized as right', though other impulses are likely to conflict. Despite such debts to Butler, Clarke andKant, Sidgwick saw the issue of free will v. determinism (and other metaphysical issues) as largely irrelevant to ethics, since it is usually impossible, in deliberation, to regard the mere absence of adequate motive as 'a reason for not doing what I otherwise judge to be reasonable' (see Moral motivation §1; Praise and blame).

Judgments of ultimate good, rather than right, do not involve definite precepts to act or the assumption that we are capable of doing so, and leave it open 'whether this particular kind of good is the greatest good that we can under the circumstances obtain'. 'Ultimate good on the whole,' Sidgwick suggests, if 'unqualified by reference to a particular subject, must be taken to mean what as a rational being I should desire and seek to realize, assuming myself to have an equal concern for *all* existence' - in contrast to taking my own existence alone to be considered, as in 'ultimate good on the whole for me'. With greater Benthamite consistency thanMill, he argues that happiness (whether for the egoist or the utilitarian) should be interpreted hedonistically in terms of pleasure or desirable consciousness, and that this yields the best account of ultimate good (see Good, theories of the; Happiness; Hedonism).

Although not preoccupied with the moral faculty as such, Sidgwick is specially concerned with moral reasoning. A 'method' of ethics is not simply a principle or theory, but a rational procedure 'for determining right conduct in any particular case', for determining the rightness of an individual's actions by seeing whether the acts in question possess some property, which, on the basis of principle, provides the ultimate reason for the rightness of acts. The plain man, he holds, uses a jumble of different methods, insufficiently reflected upon. Though many think that conscience delivers immediate judgments on the rightness of particular acts ('perceptional' or 'ultra' intuitionism), Sidgwick has 'no doubt that reflective persons, in proportion to their reflectiveness, come to rely rather on abstract universal intuitions relating to classes of cases conceived under general notions'. But this 'dogmatic' intuitionism, on which 'the practically ultimate end of moral actions' is their 'conformity to certain rules or dictates of Duty unconditionally prescribed' and discernible with 'really clear and finally valid intuition', can also be subjected to further reflection and found wanting. In synthesizing its precepts, and 'without being disposed to deny that conduct commonly judged to be right is so, we may yet require some deeper explanation *why* it is so', yielding a third phase of Intuitionism:

> which, while accepting the morality of common sense as in the main sound, still attempts to find for it a philosophic basis which it does not itself offer: to get one or more principles more absolutely and undeniably true and evident, from which the current rules might be deduced, either just as they are commonly received or with slight modifications and rectifications.
>
> (*The Methods of Ethics* 1907: 102)

Thus Sidgwick famously reconciles intuitional or common-sense morality - the nineteenth-century opponent to utilitarianism - with utilitarian principles. There are self-evident practical principles, but they 'are of too abstract a nature, and too universal in their scope, to enable us to ascertain by immediate application of them what we ought to do in any particular case'. As reflection is followed out, apprehension changes; the 'self-evidence' of particular maxims of duty fades beside that of such abstract principles as 'what is right for me must be right for all persons in precisely similar circumstances' and 'I ought to prefer the greater good of another to my own lesser good'. The dogmatic intuitionist discovers that such rules as veracity, fidelity and justice require both qualification and further systematization by higher principles, which will also resolve the conflicts between and variable formulations of such rules. Utilitarianism can sustain the 'general validity' of such rules and correct for 'the defects which reflection finds in the intuitive recognition of their stringency'; it also 'affords a principle of synthesis, and a method for binding the unconnected and occasionally conflicting principles of common moral reasoning into a complete and harmonious system'. If reflection thus reveals utilitarianism as the view to which common morality 'naturally appeals for that further development of its system which this same reflection shows to be necessary, the proof of Utilitarianism seems as complete as it can be made' (see Intuitionism in ethics §2; Moral Justification; Utilitarianism).

Sidgwick allows that current morality should not be accepted *en bloc* as the middle axioms of utilitarianism. But

he insists that in both ethics and politics, utopian speculation is an ‘illimitable cloudland’, and we must start

> with the existing social order, and the existing morality as a part of that order: and in deciding the question whether any divergence from this code is to be recommended, must consider chiefly the immediate consequences of such divergence, upon a society in which such a code is conceived generally to subsist.
>
> (*The Methods of Ethics* 1907: 474)

Besides, it is not self-evident that universal benevolence ‘is the right *means* to the attainment of universal good’, since this end may be ‘self-limiting; may direct its own partial suppression in favour of other impulses’. Indeed, Sidgwick notoriously countenanced the idea of an ‘esoteric morality’, since in less than ideal utilitarian circumstances, a utilitarian could reason that ‘some of his conclusions should be rejected by mankind generally; or even that the vulgar should keep aloof from his system as a whole, in so far as the inevitable indefiniteness and complexity of its calculations render it likely to lead to bad results in their hands’ - a view that, Sidgwick admitted, is itself perhaps best kept esoteric. Such arguments continue to inspire both utilitarians and their critics, and suggest the difficulties of finding in Sidgwick’s focus on method any great confidence in the normal person’s capacity for moral self-direction. A ‘method’ straddles the distinction between ethical standards and decision-procedures; it admits of indirect (or two-level) utilitarianism, and only in this way addresses act v. rule utilitarianism, motive utilitarianism, and so on (see Consequentialism; Moral expertise).

It is ironic that Sidgwick assimilated both the epistemology and much of the social conservatism for which utilitarians had long castigated such intuitionists as Whewell: ‘Adhere generally, deviate and attempt reform only in exceptional cases in which… the argument against Common Sense is decisive’. As D.G. Ritchie noted,Sidgwick’s ethics and politics were decidedly ‘tame and sleek’ - the ‘method is Bentham’s; but there is none of Bentham’s strong critical antagonism to the institutions of his time’. Sidgwick even downplayed his own critical insights, if they veered too far from common sense. *The Methods* provides a clear and original defence of classical against average utility calculations in relation to optimal population size and future generations, but *The Elements* elides the issue in the search for common ground, a manoeuvre also made with respect to women’s equality and the utilitarian critique of nationalism.

3 The dualism of practical reason

Sidgwick’s philosophical intuitionism houses a number of presumptively compatible, abstract, self-evident truths: a universalizability principle, a principle of rational prudence to the effect that, *ceteris paribus*, one should be equally concerned with all temporal parts of one’s life, and the principle of rational benevolence, the basis for utilitarianism (somewhat questionably derived by enjoining aiming at good generally and taking the good of each as equally important). He anticipates many recent utilitarians in interpreting the universalizability principle, derived fromKant, as purely formal, meaning simply that whatever ‘is right for me must be right for all persons in similar circumstances’ and therefore as compatible with utilitarianism (see Impartiality). However, such concordances notwithstanding, the conflict between the methods of utilitarianism and egoism, unlike that between intuitional morality and utilitarianism, is real. In an essay on *The Methods*, he explains that along with

> (a) a fundamental moral conviction that I ought to sacrifice my own happiness, if by so doing I can increase the happiness of others to a greater extent than I diminish my own, I find also (b) a conviction - which it would be paradoxical to call ‘moral’, but which is none the less fundamental - that it would be irrational to sacrifice any portion of my own happiness unless the sacrifice is to be somehow at some time compensated by an equivalent addition to my own happiness.
>
> (‘Some Fundamental Ethical Controversies’ 1889: 483)

Each of these convictions has as much clearness and certainty ‘as the process of introspective reflection can give’ as well as a preponderant, if implicit, assent ‘in the common sense of mankind’.

Excepting the criterion of mutual consistency, both convictions could pass the most demanding tests for self-evident truths: careful reflection, clarity and precision, and consensus among competent judges.Sidgwick therefore regards this as a ‘fundamental contradiction in our apparent intuitions of what is Reasonable in conduct’ and concludes, in the first edition of *The Methods*, that unless a way be found to reconcile ‘the Individual with the Universal Reason’, the ‘Cosmos of Duty is thus really reduced to a Chaos: and the prolonged effort of the human

intellect to frame a perfect ideal of rational conduct is seen to have been foredoomed to inevitable failure'.

WhetherSidgwick regarded egoism as a matter of ethics or simply of rationality is hardly to the point; for him, the problem was to frame a rational way to live. He would have been dismayed to learn that the consistent (agent-relative) egoist would still challenge moral theorists at the close of the twentieth century.

In later editions and works, Sidgwick downplayed this dilemma, but his journal and letters reveal that he was always troubled by it. He had turned to ethics hoping to establish it as independently rational, perhaps a show of God's order; his inquiry drove him to the conclusion that ethics was incoherent unless the utilitarian could infer 'the existence of Divine sanctions' that would 'suffice to make it always every one's interest to promote universal happiness to the best of his knowledge'. Evidence for the personal survival of death might help, but by 1887 he was concluding 'that we have not, and are never likely to have, empirical evidence of the existence of the individual after death'. In a singular revelation, he explained that

> When I was writing my book on Ethics, I was inclined to hold with Kant that we must *postulate* the continued existence of the soul, in order to effect that harmony of Duty with Happiness which seemed to me indispensable to rational moral life. At any rate I thought I might *provisionally* postulate it, while setting out on the serious search for empirical evidence. If I decide that this search is a failure, shall I finally and decisively make this postulate? Can I consistently with my whole view of truth and the method of its attainment?
>
> (*Henry Sidgwick, A Memoir* 1906: 467)

His crisis of faith had proved enduring.

4 Epistemology and other works

Sidgwick also explored the epistemological issue of whether such a postulate was unwarranted. If physical science is indeed based on self-evident premises, then practical philosophy should seek such a foundation as well. However,

> We find that in our supposed knowledge of the world of nature propositions are commonly taken to be universally true, which yet seem to rest on no other grounds than that we have a strong disposition to accept them, and that they are indispensable to the systematic coherence of our beliefs, - it will be more difficult to reject a similarly supported assumption in ethics, without opening the door to universal scepticism.
>
> (*The Methods of Ethics* 1907: 509)

His intuitionism was scarcely unthinking or mystical - he was only converted to it by the claim that mathematics could find no other basis - and was based on the view that inferential knowledge presupposes significant non-inferential knowledge. And as such works as 'Criteria of Truth and Error' show, he rejected any 'simple infallible criterion' for ultimate knowledge and favoured a sophisticated fallibilist criterion for 'the humbler task' of 'excluding error'. For this, he advanced a threefold test: intuitive or Cartesian verification (clarity and certainty on careful examination), discursive verification (system and coherence) and ecumenical verification (consensus). The second test, he allows, 'is of special and pre-eminent importance', since 'the ideal aim of philosophy is systematization - the exhibition of system and coherence in a mass of beliefs which, as presented by Common Sense, are wanting therein', but, he adds, 'the special characteristic of *my* philosophy is to keep the importance of the others in view'.

Sidgwick has been cast as both an early practitioner of Rawlsian reflective equilibrium and a prescient critic of it, but such attempts to dress his moral methodology for contemporary debates often read him selectively and out of his historical context. He was moved by the intellectual, cultural and political tensions of the Victorian era; religious scepticism, moral scepticism and the threat of egoism in theory and practice came to form an unholy trinity against which he would contend in one arena after another. Even *The Methods* does not fully convey the persistence and ingenuity with which he engaged 'the deepest questions of human life' or the way in which his meta-ethics were at once intuitionist and fallibilist, commonsensical and dialectical.

Indeed, Sidgwick's work is rarely appreciated even as a comprehensive account of utilitarianism, covering political, legal and economic theory. *The Principles* gives a penetrating treatment of distributive justice from the utilitarian standpoint, and *The Elements* and *The Development of European Polity* carry the classical tradition far

beyond the simplicities of Austinian jurisprudence and ahistorical Benthamism, with a complex approach to the art of politics that balances analytic, historical and comparative methods. Such works reveal how, for Sidgwick, egoism was hardly a potentially cogent alternative for reformulating morals and politics, but was rather the view capable of inspiring most of what he found politically abhorrent: class or party conflict at home and neo-Machiavellianism in foreign policy.

In the end, Sidgwick's inquiries into epistemology and politics were, in his eyes, no more successful than his inquiries into religion, ethics and psychic research. Yet beyond his enduring accomplishments in clarifying utilitarianism and advancing substantive moral and political theory, it is perhaps the very tension between his subtle, penetrating scepticism and his longing for a rational faith and morality that make him one of the most fascinating representatives of his age.

See also: Common sense school; Commonsensism; Ethics; Mill, J.S. §§8-12; Morality and ethics; Moral scepticism; Teleological ethics; Universalism in ethics

BART SCHULTZ

List of works

The Sidgwick Papers, Wren Library, Trinity College, Cambridge, form the most extensive collection of original manuscript materials, but there are many other archival resources. The most comprehensive edition of Sidgwick's works, including all of his major works (with both the first and seventh editions of *The Methods*) and a wide range of his previously unpublished lectures and correspondence, is a database: (1997) 'The Complete Works and Select Scholarly Correspondence of Henry Sidgwick', ed. B. Schultz, Past Masters series, Charlottesville, VA: InteLex Corporation.

Sidgwick, H. (1870) *The Ethics of Conformity and Subscription*, London: Williams & Norgate.(This extremely important pamphlet marksSidgwick's first serious attempt to show how common-sense morality is incomplete and requires something like utilitarianism.)

Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan; later edns, 1877, 1884, 1890, 1893, 1901, 1907. (Sidgwick's masterpiece, in which classical utilitarianism is given its most sophisticated formulation and reconciled with common-sense morality, though not with egoism. *The Methods* set the agenda for much of the substantive ethical theory and metaethics of the twentieth century. Changes made for the second and third editions were also published separately as supplements. Sidgwick wrote a number of commentaries on his book, including 'Some Fundamental Ethical Controversies' *Mind* 56 (1889); 'Prof. Calderwood on Intuitionism in Morals' *Mind* 4 (1876); and 'The Establishment of Ethical First Principles' *Mind* 13 (1879).)

Sidgwick, H. (1883) *The Principles of Political Economy*, London: Macmillan; later edns, 1887, 1901.(Sometimes dismissed as inappreciative of the marginalist revolution, Sidgwick's political economy was in fact informed by the work ofJevons, Marshall and Edgeworth, and his discussion of distributive justice and the role of the state provides an essential complement to the arguments of *The Methods*.)

Sidgwick, H. (1885) *The Scope and Method of Economic Science*, London: Macmillan. (Sidgwick's presidential address to the Economic Science and Statistics section of the British Association is a clever exposition of the limits of *laissez-faire* and the absurdities of the grandiose sociologies of Comte andSpencer (reprinted in *Miscellaneous Essays*).)

Sidgwick, H. (1886) *Outlines of the History of Ethics for English Readers*, London: Macmillan; later edns, 1888, 1892, 1896, 1902, 1931.(May well remain the best work of its kind; with succinctness and accuracy,Sidgwick surveys the entire range of philosophical ethics from its origins down to his own time.)

Sidgwick, H. (1891) *The Elements of Politics*, London: Macmillan; later edns, 1897, 1908, 1919; student edn, ed. B. Schultz, Cambridge: Cambridge University Press, 1998.(Sidgwick's second most important book, this massive volume provides another vital complement to *The Methods* by setting out Sidgwick's utilitarian politics and distinguishing his analytical approach from that of Bentham,Austin and the Mills. The student edition contains helpful secondary material, as well as new primary sources.)

Sidgwick, H. (1898, 1909) *Practical Ethics: A Collection of Addresses and Essays*, London: Swan Sonnenschein. (Essays and addresses delivered before various 'Ethical Societies' which aimed at 'the intelligent study of moral questions with a view to elevate and purify social life'. 'The Ethics of Religious Conformity' and 'Clerical Veracity' show the significance of this issue in Sidgwick's life and thought.)

Sidgwick, H. (1902) *Lectures on the Ethics of T.H. Green, H. Spencer and J. Martineau*, ed. E.E. Constance Jones, London: Macmillan.(Sidgwick came to think that *The Methods* did not do enough to address critically the transcendentalist and evolutionist positions, and this posthumous work thus provides a useful supplement to

The Methods by giving detailed criticisms of key representatives of these schools as well as further material on Martineau.)

Sidgwick, H. (1902) *Philosophy, Its Scope and Relations*, ed. J. Ward, London: Macmillan. (A succinct, precise articulation of his epistemology and its bearing on cognate fields of inquiry, mainly psychology, history and sociology.)

Sidgwick, H. (1903) *The Development of European Polity*, ed. E.M. Sidgwick, London: Macmillan.(A lucid historical survey of political doctrines, this posthumous volume provides the inductive, historical approach to politics that Sidgwick regarded as a necessary supplement to the analytical and deductive approach of *The Elements*.)

Sidgwick, H. (1904) *Miscellaneous Essays and Addresses*, ed. E.M. Sidgwick and A. Sidgwick, London: Macmillan.(Contains many ofSidgwick's best short works, including discerning literary studies of Clough andShakespeare, sober critiques of Arnold and perfectionist educational theories, various probing assessments of socialism, and a witty essay on 'Bentham and Benthamism'.)

Sidgwick, H. (1905) *Lectures on the Philosophy of Kant and Other Philosophical Lectures and Essays*, ed. J. Ward, London: Macmillan.(Contains some of Sidgwick's most significant metaphysical and epistemological work, and conveys some sense of his plan for his projected study of 'Kant and Kantism in England'. Includes 'Criteria of Truth and Error'.)

Seeley, J. (1896) *Introduction to Political Science*,ed. H. Sidgwick, London: Macmillan.(In editing the work of his colleague Seeley, an ideologue of British imperialism, Sidgwick perhaps unwittingly demonstrated the severe limitations of academic Victorian political theory.)

References and further reading

Broad, D.D. (1930) *Five Types of Ethical Theory*, London: Routledge & Kegan Paul. (An extensive, sympathetic, philosophically sophisticated account that is marred by various confusions concerning Sidgwick's notion of a method, which is treated as equivalent to a theory.)

Frankena, W. *et al.* (1974) '*Monist* Symposium: Henry Sidgwick', *Monist* 58. (Commemorates the centenary of the publication of *The Methods*; Frankena's paper is especially important for understanding Sidgwick's internalism and Singer provides a provocative defence of intuitionism against reflective equilibrium.)

MacIntyre, A. (1990) *Three Rival Versions of Moral Enquiry*, Notre Dame: Notre Dame University Press. (Contains a critique of Sidgwick's 'encyclopedist' version of enquiry from the perspective of a traditionalist.)

Moore, G.E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press. (Moore was an ungenerous critic who largely begged the question of the dualism of practical reason, but was effective in diminishing Sidgwick's reputation.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Oxford University Press.(Important work in moral theory revealing a deep indebtedness to *The Methods*, with many astute asides on Sidgwick's critique of common-sense morality, his account of ultimate good and his view of the person.)

Ritchie, D.G. (1891-2) 'Review: *The Elements of Politics*', *International Journal of Ethics* 2: 254-7.(A witty, critical and very widely-cited review of Sidgwick's *magnum opus* on politics.)

Schneewind, J.B. (1977) *Sidgwick's Ethics and Victorian Moral Philosophy*, Oxford: Clarendon Press.(The most important work onSidgwick's ethics, or indeed, on Sidgwick. Schneewind situates Sidgwick's work within the broader context of English moral theory, and is especially concerned to construe *The Methods* as a piece of systematic moral theory rather than simply as a brief for utilitarianism. Contains an excellent bibliography.)

Schneewind, J.B. and Schultz, B. (1998) 'Henry Sidgwick, A Bibliography', *The Cambridge Bibliography of English Literature*, Cambridge: Cambridge University Press, 3rd edn.(The most comprehensive bibliography of works by or on Sidgwick published before 1920.)

Schultz, B. (ed.) (1992) *Essays on Henry Sidgwick*, Cambridge: Cambridge University Press.(A wide-ranging collection of articles, many new, covering most of the current interpretive controversies.)

Schultz, B. (1996) 'Sidgwick, Henry', *Dictionnaire d'éthique et de philosophie morale*, ed. M. Canto-Sperber, Paris: Presses Universitaires de France, 1391-99.(A short but up-to-date introduction that both surveys Sidgwick's philosophy and points to his larger cultural significance, including his connections to such controversial figures as J.A. Symonds and F.W.H. Myers.)

Schultz, B. (1998) *Eye of the Universe: Henry Sidgwick and the Victorian Quest for Certainty*.(The only comprehensive treatment of Sidgwick's practical and theoretical philosophy, this book brings out the overall unity of Sidgwick's writings and the central role of scepticism and the dualism of practical reason in his life and work.)

Sidgwick, E.M. and Sidgwick, A. (1906) *Henry Sidgwick, A Memoir*, London: Macmillan.(The essential

biographical treatment, with a good bibliography; Sidgwick's widow and brother drew from his letters and journal, though they excluded excessively technical or personal material.)

Williams, B. (1982) 'The Point of View of the Universe: Sidgwick and the Ambitions of Ethics', *Making Sense of Humanity*, Cambridge: Cambridge University Press.(A wonderful introduction to Sidgwick, with a penetrating critical overview that gives special currency to the interpretation of Sidgwick as a 'Government House' utilitarian, but compares him favourably to Moore.)

Siger of Brabant (*c*.1240-*c*.1284)

Born probably circa *1240 in the Duchy of Brabant, Siger of Brabant studied philosophy in the arts faculty at the University of Paris and became regent master there in the 1260s. Various positions which he defended were included in Bishop Etienne Tempier's condemnations of 1270 and 1277, and he appears to have fled France when cited to appear before the French inquisition. He probably spent his final days in Italy, and died there before November 1284.*

As a professional teacher of philosophy, Siger regarded it as his primary mission to lecture on Aristotle and other philosophers and to present their views on the points at issue. Early in his career he defended some of the positions condemned by Bishop Tempier, but after 1270 he often nuanced his exposition of such views by noting that he was only presenting the views of the philosophers or of Aristotle, or that he was proceeding philosophically in these discussions. Often regarded as a leading Latin 'Averroist', he agreed with Averroes that there is only one human intellect, though he eventually reversed his view on this. His personal philosophy is strongly Aristotelian, but with various elements derived from Neoplatonism.

On the relationship between essence and existence in created beings, Siger denies that existence is something added to a thing's essence and holds that the existence of such entities belongs to their essence. He holds that one can demonstrate God's existence, but insists that Aristotle's physical argument for a First Mover must be completed by metaphysical argumentation. While denying that human beings in this life enjoy any direct knowledge of the divine essence, he seems open to Averroes' view that they can reach some knowledge thereof.

1 Life and works

Siger studied philosophy in the arts faculty at the University of Paris and became regent master in that faculty in the 1260s, perhaps *circa* 1263-5. Some of his views were included in a condemnation of thirteen propositions by Etienne Tempier, Bishop of Paris, in December 1270, and many more were included in Tempier's much broader condemnation of 219 propositions on 7 March 1277. Little is known about Siger's final years, but it is certain that in November 1276 he and two colleagues from the arts faculty were cited to appear before the French inquisition in January 1277. Apparently he and the others had already departed from France. He probably spent his final years in Italy and died some time before November 1284, murdered perhaps by a secretary. Although his name has long been associated with the radical Aristotelian movement in the arts faculty at Paris known by some as Latin 'Averroism' (see Averroism), he was never found guilty of heresy. Scholars have been intrigued by the fact that Dante Alighieri places him in Paradise, and has Thomas Aquinas introduce him.

Most of his surviving writings date from after 1270. Many of his works were reports taken down by others at his lectures, although some of these were later revised by Siger. Other works were originally composed by Siger himself, and some of these were probably based on prior scholarly disputations. A number are commentaries on works by Aristotle, or in one case, on the *Liber de causis*, written in the form of questions occasioned by the text (see Liber de causis).

2 Philosophy and religious belief

By Siger's time at Paris, the faculty of arts had become in fact a faculty of philosophy. This was owing in large measure to the recovery in Latin translation in the twelfth and thirteenth centuries of a large amount of philosophical literature of non-Christian origin originally written in Greek or in Arabic (see Translators). There had been initial reservations and even prohibitions early in the thirteenth century on the part of Church authorities at Paris against incorporating some of this newly available material into teaching in the arts faculty at the University. By 1255, however, the situation had changed to the point where the statutes for that faculty required students to read all the known works of Aristotle. Thus Siger regarded it as his primary mission to lecture on Aristotle and the other philosophers, and gave great weight to them.

Siger did this with such enthusiasm in two early works that some of his positions alarmed Bishop Tempier and were condemned in 1270. Thus in his *Quaestio utrum haec sit vera* he had accepted the eternity of the human species (and hence of the world) (see Eternity of the world, medieval views of), and in his *Quaestiones in Tertium De anima (Questions on Book III of Aristotle's On the Soul)*, the view associated with Averroes that the human

intellect is only one for the entire human race and exists apart from individual human beings (see Ibn Rushd). Particularly offensive to ecclesiastical authorities was the implication following from this that individual souls are not immortal (see Soul, nature and immortality of the).

In addition to Tempier's condemnation of these views, Thomas Aquinas sharply attacked the Averroistic and Sigerian defence of the unicity of the intellect in his *De unitate intellectus*, also of 1270. In subsequent works, when dealing with theologically sensitive issues, Siger became more circumspect and struggled to find a satisfactory resolution for the faith-reason problem. When dealing with points defended by Aristotle or Averroes or other philosophers which contradict Christian teaching, he often remarks that he is simply presenting such positions as the view of the philosophers, or according to the mind of the philosophers, or that he is speaking philosophically. In his final works, especially his *Quaestiones super Librum de causis (Questions on the Liber de causis)*, he moves still further in the direction of religious orthodoxy. In sum, while he never works out a fully satisfactory solution to the faith-reason problem, he appears to be sincere in his repeated protestations of religious belief and in his gradual movement toward more orthodox positions. At the same time, along with other colleagues in his faculty, especially Boethius of Dacia, he strongly defends the autonomy of philosophy within its own sphere, and the right of the philosopher, even though Christian, to draw conclusions from philosophical principles wherever they may lead. If these conclusions contradict religious belief, the Christian, being aware of the fallibility of human reason, should regard the teachings of faith as true. Siger does not defend the 'double-truth theory', according to which two contradictory propositions, one taught by philosophy and one taught by faith, could both be true at the same time (see Aristotelianism, medieval).

3 Metaphysics: the science of being as being

Siger is familiar with Aristotle's description in *Metaphysics* IV of a science that studies being as being and his contrast between this and more particular sciences which study the attributes of only particular parts of being. Siger also recalls Aristotle's presentation of three theoretical sciences in *Metaphysics* VI, physics, mathematics, and a divine science that studies separate and immobile (and divine) entities (see Aristotle). Avicenna denies that God can be regarded as the subject of this science, for he maintains that it belongs to metaphysics and not to physics to demonstrate God's existence. Since no science can demonstrate the existence of its subject, the subject of metaphysics is not God, but being as being (see Ibn Sina). Averroes affirms that it belongs to physics to demonstrate the existence of the First Mover or God, and that in referring to being as being as the subject of metaphysics, Aristotle really means that its subject is the primary instance of being, or God (see Ibn Rushd). In agreement with Avicenna, Siger holds that the subject of metaphysics is not God but being as being. However, he adopts an intermediary position on the issue of a physical or metaphysical proof of God's existence. The middle term used by Aristotle to prove God's existence in his *Metaphysics* is taken from his *Physics* (see God, arguments for the existence of).

4 Essence and existence

Much contested during Siger's time and thereafter was the issue of the relationship between essence and existence in created entities (see Existence). It was generally granted that they are identical in God. To account for the metaphysical structure of finite beings, Aquinas defends a composition of two distinct principles in such beings, an essence and a corresponding act of being (*esse*), often referred to by others as existence. Giles of Rome defends a seemingly more extreme version of this approach, and by 1276 Henry of Ghent sharply criticizes the real distinction between essence and existence, especially as presented by Giles; he postulates an 'intentional' distinction between them, that is, one that is less than real but more than purely mental.

In his *Quaestiones in Metaphysicam (Questions on the Metaphysics)* Siger asks whether in caused beings existence (*esse*) belongs to their essence. Against this claim he cites Boethius and Avicenna, and then as a leading critic of distinction between them, Averroes. Among contemporaries he presents Albert the Great as holding that existence is added to a thing's essence, and rejects this position. Siger presents as intermediary the view of Aquinas (in the latter's commentary on the *Metaphysics*) that existence is something added to a thing's essence, which does not belong to its essence and yet is not an accident but is, as it were, 'constituted' from the principles of the essence. Siger agrees with Aquinas that existence is not an accident, but complains about the way he formulates his position. For Siger, if existence belongs to a thing it must either be a part of that thing's essence, like matter and form, or something composed of those principles, or an accident. For Aquinas to say that it is added to the essence

without being any of these is for him to postulate a fourth nature among entities. Siger concludes that the existence of such entities belongs to their essence and is not something superadded. The terms which he now uses to signify essence and existence, ‘thing’ (*res*) and ‘existing entity’ (*ens*), do not signify different intentions but rather one and the same essence in different ways, one as potentiality (*res* = essence), and the other as actuality (*ens* = existing entity).

In responding to various arguments he had offered for real distinction between essence and existence, Siger notes that one especially moved Aquinas. Everything other than the First Being (God) must be composed. Because created pure spirits lack matter-form composition, they must be composed of essence and *esse*. To this Siger proposes two responses, though he does not assert the first definitively. First, it is enough to say that things other than God recede from him as pure actuality and are multiplied by approaching potentiality. This suffices to distinguish them from God without postulating different ‘essences’ within them. Second, even if one concedes Aquinas’ claim that to be distinguished from God created entities must somehow be composed, Siger finds it sufficient to appeal to substance-accident composition in created spirits. Even in them, there is a distinction between their substance and the intelligible species they use in thinking.

Interestingly, in his final work, *Quaestiones super Librum de causis (Questions on the Liber de causis)*, while discussing the kind of infinity enjoyed by the First Cause, Siger notes that in created spiritual entities existence (*esse*) is not limited by matter. Yet their existence is limited to their nature which receives it and is related to it as potentiality to actualization. Perhaps Siger now accepts Aquinas’ position on essence and the act of being.

5 Philosophical knowledge of God

Reference has been made to Siger’s insistence on incorporating Aristotle’s physical argument for a First Mover into metaphysical argumentation for God’s existence. For Siger, one cannot stop with a purely physical proof but must in metaphysics establish the existence of a supreme being which is unique, and which is the cause of being for all other entities. At different points in his *Quaestiones in Metaphysicam* he attempts to show that the First Mover is the creative cause of all other entities, including separate - that is, immaterial - substances. He also offers a series of arguments to show that there is only one first and uncaused cause (see Causation).

Presupposed for all his reasoning from effects to God as the first cause is Siger’s conviction that in this life human beings do not enjoy a direct knowledge of God’s essence. However, on one occasion in his *Quaestiones in Metaphysicam*, Siger comments favourably on Averroes’ view that human beings can arrive at some kind of knowledge of the divine essence. Siger himself here suggests that one who is truly expert in philosophy may be able to form an adequate idea of the divine essence. Also in his *Quaestiones super Librum de causis* he offers a series of arguments for and another series against the claim that the essence of the First Cause can be grasped by the human intellect. Unfortunately, his resolution of this question is missing from the text.

6 Procession of created reality from God

In his *De necessitate et contingentia causarum (On the Necessity and Contingency of Causes)*, Siger presents God as the immediate, necessary and eternal cause of the first intelligence, but as only the mediate cause of other effects. He supports this by appealing to the Neoplatonic axiom that from the one simple being (God) only one thing can proceed immediately. This view is at odds with traditional Christian belief both in affirming the eternal existence and the necessary production of various created realities, and in denying that God is the immediate creator of anything but the first intelligence. However, Siger also states twice in the course of his discussion that this is so according to the mind of the philosophers. Hence he may not have adopted this as his own position, although he does offer a sympathetic presentation of necessary emanation in one version of his *Quaestiones in Metaphysicam*. However, he clearly rejects the theory of mediate emanation or creation in his *Quaestiones super Librum de causis*.

7 Eternity of the universe

While their religious faith required Christian participants in this discussion to believe that the world began to be, much disagreement existed concerning whether this could be demonstrated philosophically. Bonaventure, at least as he is usually interpreted, defends this possibility. Aquinas denies that human reason can prove this and in his final treatment of the issue goes so far as to hold that an eternally created world is possible, philosophically

speaking. As already noted, in a work written before 1270, Siger defends the eternal existence of the human race. After the condemnation of 1270, when he presents argumentation for an eternal world or some eternally produced creature, he usually qualifies this by saying that this is so according to the mind of Aristotle or according to the mind of the philosophers and is careful not to defend the eternity of the world in his own name. Like Aquinas, he denies that human reason can prove that the world began to be (see Eternity of the world, medieval views of).

8 Unicity versus plurality of the human intellect

In his *Quaestiones in Tertium De anima* from the 1260s, Siger defends the view that there is only one separate intellect (including both the agent and the receiving or possible intellect) for the entire human race. The agent and possible intellects are two powers of one separate substance, and not two separate substances, as Averroes holds in his Long Commentary on the *De anima*. After the condemnation of 1270 and Aquinas' attack that same year, Siger deals with this issue very carefully. In his *De anima intellectiva (The Intellective Soul)* of *circa* 1273, after much discussion he declares himself unable to resolve this issue on purely philosophical grounds. He concludes that in such a matter one must follow the teaching of faith. In his final work, his *Quaestiones super Librum de causis*, he strongly argues on philosophical grounds that the intellective soul is a perfection of each individual human body, and that it is multiplied as are human beings themselves. His position is now perfectly orthodox.

See also: Aquinas, T.; Aristotelianism, medieval; Averroism; Boethius of Dacia; Causation; God, concepts of; Ibn Rushd

JOHN F. WIPPEL

List of works

Siger of Brabant (*c.*1265) *Compendium magistri Sugeri super De generatione et corruptione (Compendium of Master Siger on Aristotle's On Generation and Corruption)* ed. B. Bazán, *Siger de Brabant. Écrits de logique, de morale et de physique*, Louvain: Publications universitaires, 1974.(Commentary on Aristotle.)

Siger of Brabant (*c.*1269 or earlier) *Quaestiones in Tertium De anima (Questions on Book III of Aristotle's On the Soul)*, ed. B. Bazán, *Quaestiones in Tertium De anima, De anima intellectiva, De aeternitate mundi*, Louvain: Publications universitaires, 1972.(Questions on *On the Soul*.)

Siger of Brabant (before 1270) *Sophisma Omnis homo de necessitate est animal (The Sophisma 'Every Human Being is Necessarily an Animal')*, ed. B. Bazán, *Siger de Brabant. Écrits de logique, de morale et de physique*, Louvain: Publications universitaires, 1974.(Edition of the sophisma.)

Siger of Brabant (before 1270) *Quaestio Utrum haec sit vera: homo est animal nullo homine existente (Is the Proposition 'A Human Being is an Animal' True When No Human Being Exists?)*, ed. B. Bazán, *Siger de Brabant. Écrits de logique, de morale et de physique*, Louvain: Publications universitaires, 1974.(Edition of the question.)

Siger of Brabant (*c.*1270-1) *Quaestiones in Physicam (Questions on Aristotle's Physics)*, ed. A. Zimmermann in B. Bazán (ed.) *Siger de Brabant. Écrits de logique, de morale et de physique*, Louvain: Publications universitaires, 1974.(Edition of the questions.)

Siger of Brabant (*c.*1272) *De necessitate et contingentia causarum (The Necessity and Contigency of Causes)*, ed. J.J. Duin, *La doctrine de la providence dans les écrits de Siger de Brabant*, Louvain: Institut supérieur de philosophie, 1954, 14-50.(Work on causation.)

Siger of Brabant (*c.*1272) *De aeternitate mundi (The Eternity of the World)*, ed. B. Bazán, *Quaestiones in Tertium De anima, De anima intellectiva, De aeternitate mundi*, Louvain: Publications universitaires, 1972.(Siger's work on eternity.)

Siger of Brabant (*c.*1272) *Quaestiones logicales (Logical Questions)*, ed. B. Bazán, *Siger de Brabant. Écrits de logique, de morale et de physique*, Louvain: Publications universitaires, 1974.(Edition of the questions.)

Siger of Brabant (*c.*1272) *Impossibilia*, ed. B. Bazán, *Siger de Brabant. Écrits de logique, de morale et de physique*, Louvain: Publications universitaires, 1974.(Edition of the work.)

Siger of Brabant (*c.*1272-3) *Quaestiones naturales (Natural Questions)* (Paris), ed. B. Bazán, *Siger de Brabant. Écrits de logique, de morale et de physique*, Louvain: Publications universitaires, 1974.(Edition of the questions.)

Siger of Brabant (*c.*1273) *De anima intellectiva (The Intellective Soul)*, ed. B. Bazán, *Quaestiones in Tertium De anima, De anima intellectiva, De aeternitate mundi*, Louvain: Publications universitaires, 1972.(Siger's work on the intellect.)

Siger of Brabant (*c.*1273-4) *Quaestiones morales (Moral Questions)*, ed. B. Bazán, *Siger de Brabant. Écrits de*

logique, de morale et de physique, Louvain: Publications universitaires, 1974.(Edition of the questions.)

Siger of Brabant (*c.*1273-4) *Quaestiones naturales (Natural Questions)* (Lisbon), ed. B. Bazán, *Siger de Brabant. Écrits de logique, de morale et de physique*, Louvain: Publications universitaires, 1974.(Edition of the questions.)

Siger of Brabant (*c.*1273-4) *Quaestiones in Metaphysicam (Questions on Aristotle's Metaphysics)* (Munich and Vienna), ed. W. Dunphy, Louvain-la-Neuve: Institut Superieur de Philosophie, 1981.(These may repeat one or more sets of oral lectures.)

Siger of Brabant (*c.*1273-4) *Quaestiones in Metaphysicam (Questions on Aristotle's Metaphysics)* (Cambridge and Paris), ed. A. Maurer, Louvain-la-Neuve: Institut superieur de philosophie, 1983.(These may repeat one or more sets of oral lectures.)

Siger of Brabant (*c.*1275-6) *Quaestiones super Librum de causis (Questions on the Liber de causis)*, ed. A. Marlasca, *Les Quaestiones super Librum de causis de Siger de Brabant*, Louvain: Publications universitaires, 1972.(Edition of the questions.)

References and further reading

Bianchi, L. (1990) *Il vescovo e i filosofi. La condanna Parigina del 1277 e l'evoluzione dell'aristotelismo scolastico (The Bishop and the Philosophers: The Condemnation of 1277 at Paris and the Evolution of Scholastic Aristotelianism)*, Bergamo: Pierluigi Lubrina Editore. (Contains many references to Siger and a very full bibliography concerning the events surrounding the condemnation of 1277.)

Hissette, R. (1977) *Enquête sur les 219 articles condamnés à Paris le 7 mars, 1277 (An Investigation Concerning the 219 Articles Condemned at Paris on 7 March, 1277)*, Louvain: Publications Universitaires, Paris: Vander-Oyez.(Indispensable for the condemnation of 1277 and also for Siger's implication therein.)

Mandonnet, P. (1908-11) *Siger de Brabant et l'averroïsme latin au XIIIe siècle (Siger of Brabant and Latin Averroism in the Thirteenth Century)*, Les Philosophes Belges VI-VII, Louvain: Institut supérieur de philosophie, 2 vols. (Though now badly dated, this was a seminal work for much earlier twentieth-century scholarship on Siger.)

Putallaz, F.-X. (1995) *Insolente liberté - controverses et condamnations au XIII^e^ siècle (Insolent Liberty: Controversies and Condemnations in the Thirteenth Century)*, Fribourg: Éditions Universitaires de Fribourg Suisse.(See especially pages 15-50 on Siger's explanation of the roles of the intellect and the will in accounting for human choice.)

Van Steenberghen, F. (1931-42) *Siger de Brabant d'après ses oeuvres inédites (Siger of Brabant According to His Unedited Works)*, Les Philosophes Belges XII-XIII, Louvain: Institut supérieur de philosophie, 2 vols. (The major work of its day calling for a re-evaluation of Siger, although now surpassed by its author's later works.)

Van Steenberghen, F. (1977) *Maître Siger de Brabant (Master Siger of Brabant)*, Louvain: Publications universitaires, Paris: Vander-Oyez.(The most definitive work yet to appear on Siger's life, career, works and doctrine.)

Van Steenberghen, F. (1991) *La philosophie au XIIIe siècle (Philosophy in the Thirteenth Century)*, 2nd edn, Louvain: Éditions Peeters.(See Chapter VIII especially for the author's most recent overview of Siger.)

Wippel, J.F. (1995) *Medieval Reactions to the Encounter Between Faith and Reason*, Milwaukee, WI: Marquette University Press.(See pages 14-18 on Siger's implication in the 1270 and 1277 condemnations at Paris, and pages 33-59 on Siger's views on the faith-reason issue.)

Signposts Movement

*The symposium Signposts (*Vekhi*, sometimes translated* Landmarks*), published in 1909, was a* succès de scandale *which provoked a long debate of extraordinary intensity and scope on the nature and outlook of the Russian intelligentsia. The discussion continued after the 1917 Revolution among intellectuals in exile, and was resumed in Russia with the republication of the volume in the Gorbachev era.*

The contributors to Signposts were the philosophers N.A. Berdiaev (1874-1948), S.N. Bulgakov (1871-1944) and S.L. Frank (1877-1950); M.O. Gershenzon (1869-1925), a well-known critic and historian of literature; A.S. Izgoev (pseudonym of A.S. Lande, 1872-1935), a journalist active in the liberal Constitutional Democratic (Kadet) Party; B.A. Kistiakovskii (1868-1920), a specialist in constitutional law; and P.B. Struve (1870-1944), an eminent economist and editor, and a member of the Kadet Party's Central Committee. They argued that the 1905 Revolution had revealed the despotic potential in the intelligentsia's traditional materialist faith, and urged it to re-examine its values, which were based on a misunderstanding of human nature and threatened the existence of Russian culture.

The shock caused by the volume owed much to the fact that, unlike most of the intelligentsia's critics, the authors were not of the political right. The majority were former Marxists who had moved to forms of liberalism based on idealist positions in philosophy. But although many intelligentsia groups were also reassessing their values in the wake of the 1905 Revolution, for political reasons liberals closed ranks with the left in an overwhelming condemnation of Signposts as a betrayal of the cause of freedom. The famous Signposts debate (which was pursued in exile after 1917 as a discussion on the meaning of the Russian Revolution) was not a dialogue, but rather a succession of dogmatic professions of faith by mutually hostile political groups. Western historians have commonly stressed the symposium's significance as a prophetic indictment of Russian radical messianism from the standpoint of liberal pragmatism. But Signposts was not an ideological unity: as well as its dominant liberal Westernism it contained a strong strand of nationalistic messianism which had affinities with traditions on both the Russian left and the right. In this respect the volume reflects a fundamental tension in Russian thought between dogmatic utopianism and radical humanism.

1 From Marxism to idealism

Signposts' critique of the intelligentsia's outlook was grounded in the revival of metaphysical idealism in Russia at the beginning of the twentieth century. A major part in this was played by Struve, Berdiaev, Bulgakov and Frank, who had been leading exponents of Marxism in Russia in the 1890s, stressing its superiority over previous radical ideologies because of its 'scientific' demonstration of the historical inevitability of Russia's progression through capitalism to socialism. But, influenced by German Marxist revisionism, they began to lose faith in the predictive powers of Marx's revolutionary dialectic, and came to believe that the phenomenal world of law and necessity did not exhaust reality: there was also an inner world of goals and ideals. In 1901 a long preface by Struve to Berdiaev's *Sub''ektivizm i individualizm v obshchestvennoi filosofii (Subjectivism and Individualism in Social Philosophy)* outlined the general position on which all four were then converging as a result of their critique of the empirical and logical defects of deterministic materialism: a Kantian dualism that placed human goals and values in a transcendental realm, the source of the categorical imperative that the individual be regarded always as an end in himself and never as a means (see Kant, I. §9; Neo-Kantianism, Russian §5). They argued that in their pursuit of material goals the Russian radical intelligentsia had ignored mankind's metaphysical thirst for good, truth and beauty, and neglected what should be the ultimate aim of the fight for progress: the creation of a new human type, spiritually reborn, the bearer of ideal values.

In finding the positivist and materialist tenets of the Russian radical tradition inadequate as explanations of the world or guides to action, the four thinkers represented a widespread mood that found expression in the Russian Religious and philosophical renaissance and the cultural ferment of the 'Silver Age' (see Russian Religious-Philosophical Renaissance). In 1903 they contributed with eight other writers to the symposium *Problemy idealizma (Problems of Idealism)*, which sketched out idealist positions in the fields of ethics, sociology, the philosophy of history and law, in which to ground a systematic defence of the proposition that the individual was not just a cog in society but the source of its values. The contributors stressed the symposium's practical

purpose: like the new movements in literature and the arts, it was a response to a ‘striving for moral renewal’ among large numbers of the intelligentsia who believed that commitment to narrow political ideologies had stunted their personal growth and distorted their moral vision: the ’new man’ of the future, they hoped, would be a many-sided ’integral personality’, morally, intellectually and aesthetically developed. But while some contributors held that such a type would possess an ‘integral worldview’, in which knowledge and faith would be harmonised in the perception of the historical process as a movement to a transcendent goal, others were moving towards a humanistic idealism incompatible with a dogmatic metaphysics. The former included Berdiaev and Bulgakov, the latter Struve and Frank. Their diverging paths would lead to irreconcilable positions in *Signposts* on the question of the goals of progress in Russia and the intelligentsia’s role in their attainment.

2 Utopianism versus humanism

In his book *Sub’’ektivizm i individualizm v obshchestvennoi filosofii*, Berdiaev declared the need for ’an organically whole, positive view of the world; at this historical moment scepticism can only serve reactionary purposes’ (1901: 24). He and Bulgakov both believed that disenchantment with the millenarian promises of Marxism had left a vacuum that must be filled by a new faith to inspire the fight against autocracy. They found the source of this faith in the messianic vision of Vladimir Solov’ëv, who was a major influence on the revival of idealism in Russia at that period. Solov’ëv taught that alienation and social conflict were the results of the divorce of reason from religious faith. The spiritual unity of the psyche would be restored only when philosophy, science and theology were united in the pursuit of the supreme aim of knowledge: an understanding of the ’total-unity’ of all existence through communion with its source, God (the Absolute) to which it is destined to return. The historical process was a movement to this final goal: an era when mankind would live an ’integral life’ within an ’integral society’ - a utopia both humanistic and theocratic, uniting East and West, Church and state. Like the Slavophiles, Solov’ëv believed that Russia’s destiny was to fulfil a religious mission: to be the first bearer of the new life among mankind (see Slavophilism; Solov’ëv, V.S.).

Berdiaev and Bulgakov saw no contradiction between their defence of the individual’s right to self-direction against the encroachments of collectivistic social doctrines and their belief in a form of historical determinism that drew heavily on Solov’ëv. In *Problemy idealizma* Bulgakov defines historical progress as the march of the Absolute towards its goal. All nations and individuals have their particular historical mission, ’foreordained in the moral world order’, which can be discovered by consulting conscience, the voice of the Divine Will within each individual. Bulgakov’s religious evolution was already leading him away from historical to theological concerns (he became an Orthodox priest in 1918). Berdiaev on the other hand was involved in all aspects of the Russian cultural renaissance. He was active in discussions in which representatives of the Church and the intelligentsia sought a *rapprochement* on questions of morality and culture, and until 1908 was part of the circle of Symbolist writers and religious thinkers grouped around D.M. Merezhkovskii and his wife, the poet Zinaida Gippius. He shared their enthusiasm for Nietzsche’s ideas and adopted Merezhkovskii’s theory of the imminent coming of a ‘New Religious Consciousness’ based on a ’Third Testament’ which would teach a morality reconciling Christ and AntiChrist, the flesh and the spirit, Nietzschean self-affirmation and Christian brotherhood (see Nietzsche: impact on Russian thought §1). He greeted the 1905 Revolution as an apocalypse that signalled the advent of a religious utopia, but, disillusioned by its failure to achieve the goal of political liberty, concluded that the attainment of freedom was not possible in the material world, and broke with Merezhkovskii’s group, taking up a stance of spiritual revolt against the irredeemable philistinism of human societies, a position much influenced by Nietzsche, and which he described as mystical anarchism. He warned the intelligentsia that Russia’s cultural backwardness could not be overcome by a historical leap - a view that was hard to reconcile with the belief that he also expressed at that period, that Russia was destined to bring about a universal spiritual regeneration by assimilating Western culture and combining it with Orthodox spirituality.

Berdiaev and Bulgakov were intransigent utopians who (as their memoirs reveal) had been attracted to Marxism by its promise of a wholly emancipated mankind. In contrast, Struve records that he had been drawn to Marx’s economic analysis by the light it shed on the practical problems of achieving civil and political liberties in an undeveloped economy. In 1900 he moved from Social Democracy to liberalism, and in 1906 became a member of the Central Committee of the newly formed Kadet Party and subsequently editor of a leading liberal journal, *Russkaia mysl’ (Russian Thought)*. He was followed to liberalism by Frank, who had become a close friend. Both regarded their political allegiance as an expression of their religious humanism, basing their defence of gradualism

and compromise on their belief in the inseparability of the material and spiritual elements in human creativity, and its roots in the continuity of national life. They stressed the interdependence of freedom and culture, which they defined as the embodiment of eternal values in the economic and social structures, artefacts and everyday activities of historical existence. They regarded the Russian left's faith in the principle of destruction as anti-cultural and thereby antilibertarian, and devoted their journalism between 1905 and the First World War principally to a campaign against all forms of radical and religious messianism that, in the expectation of an historical leap to a state of ideal freedom, ignored the importance of historical continuity and of the everyday tasks of social construction and cultural creation.

3 *Signposts*

While the four ex-Marxists often aired their differences in print, they were agreed in holding the intelligentsia principally to blame for the anarchy and destruction of the 1905 Revolution and its failure to attain its objectives. For this reason the literary historian M.O. Gershenzon invited them to contribute to a symposium on that theme, together with two liberals concerned with similar issues, B.A. Kistiakovskii (who had contributed to *Problemy idealizma*) and A.S. Izgoev.

In his introduction to *Signposts* Gershenzon outlines the reasons for the project. The intelligentsia's traditional values had been tested by the Revolution and found wanting. The volume's aim was to contribute to the process of self-examination that had ensued among the intelligentsia by offering its own diagnosis of their malaise. Despite differences of principle among the authors, they shared a common platform: recognition of the theoretical and practical primacy of the individual's inner spiritual life over the external forms of community. They believed that the intelligentsia's outlook, which rested on the opposite principle - recognition of the unconditional primacy of social forms - was erroneous in theory and harmful in practice.

Although the essays were written without prior consultation among the authors, they had a number of themes in common, chief among which was a critique of the intelligentsia's moral relativism and its utilitarian approach to all values as means to political ends. Kistiakovskii points to the traditional contempt of Russian radicals for legal principles, which they perceived as impediments to revolutionary change. Berdiaev notes their indifference to philosophical truths which could not be used for social purposes. Frank defines their outlook as 'nihilistic moralism': a rejection of objective and universally binding moral standards combined with a deification of the subjective interests of the masses which (identified with the goals of a specific class or party) had been invoked to sanction violence and criminality during 1905. Gershenzon pointed out that the intelligentsia's preoccupation with revolutionary agitation had led them to neglect the skills and capacities needed for the task of social construction. Izgoev noted that the supreme criterion of virtue among young radicals was readiness to die for the cause; Bulgakov argued that the left's idolatrous worship of the revolutionary hero was the root cause of its preference for extreme positions over compromise in all cases. But he and Berdiaev differed from the other contributors in discerning positive qualities in the intelligentsia's maximalism, arguing that its ascetic hostility to bourgeois culture and values expressed an unconscious religious thirst for the Kingdom of God on Earth which had close affinities with the religious and cultural messianism of the Slavophiles, Dostoevskii, and Vladimir Solov'ëv. They placed their hopes for Russia's political and cultural progress on a future synthesis of the religious and secular traditions of Russian thought whereby the intelligentsia, having renounced its atheism, would apply itself to the elucidation of Russia's spiritual mission in the world.

Struve's essay, however, rejects the theory of the intelligentsia's unconscious religiosity as a myth. He and Frank unambiguously condemn its ideological intransigence, in which they see the seeds of despotism and mob rule. Struve argued that the intelligentsia's traditional opposition to the state had led it to ignore the problem of education in politics and social reform: it had appealed to the instincts and passions of the masses in 1905 instead of using the concessions forced out of the government as a means of schooling the people in the habits of democracy. Frank attacks the cult of 'simplicity' common to Russian radicalism and Tolstoism: a puritanical hostility to all cultural values that were alien to the traditional life of the Russian peasant. Berdiaev and Bulgakov discerned an admirable otherworldliness in the intelligentsia's aversion to the bourgeois culture of the West; Struve and Frank looked forward to its embourgeoisement, when it would cease to exist as a separate cultural category with its own traditions and fanatical faiths, and would turn its attention from the redistribution of material goods to the creation of cultural values.

As Struve remarked after *Signposts'* publication, it reflected the traditional division in Russian thought between Westernizers and Slavophiles. (The latter included Gershenzon who, unlike the other contributors, had little interest in politics but shared the Slavophiles' concern with spiritual 'integrality'.) *Signposts* pointed simultaneously in opposing directions.

4 The debate

The symposium was the sensation of the year. It was pronounced on by public figures and passionately discussed in newspapers and journals throughout the Russian Empire. It generated three counter-symposia; meetings and speeches were devoted to it, and it soon ran into five editions.

To the authors' embarrassment, the volume was fulsomely praised by the conservative right which interpreted it as a political attack on the left. The great majority of the intelligentsia took the same view, representing the authors as the government's agents or its unconscious tools. In the words of the Socialist Revolutionary leader Viktor Chernov *Signposts* was 'the most reactionary book of the decade'. A Kadet symposium on the volume argued that while some of its criticism was valid, the intelligentsia's shortcomings were a consequence of Russia's economic backwardness and the Signposts authors themselves were not immune from its diseases. Among Russian writers only the Symbolist poet and novelist Andrei Belyi and the essayist Vasilii Rozanov publicly expressed solidarity with its critique of the values of the left, whose battle against metaphysics, aesthetics and religion had, in Rozanov's view, resulted in the loss of its 'spiritual sight'.

The intelligentsia's overwhelming hostility to *Signposts* has been cited by historians as evidence of its dogmatic adherence to its traditional revolutionary mystique. But the experiences of 1905 had led to a widespread questioning of that mystique in the radical movement, notably among the leadership of the Socialist Revolutionary Party, who had begun a fundamental re-examination of their moral assumptions, while writers sympathetic to the left, such as D. Merezhkovskii, commented on the role of religious and philosophical ideas in the transformation of its outlook. The closing of the intelligentsia's ranks against *Signposts* was due to the belief that their criticism came from a hostile camp and would strengthen conservative political forces. In the Kadets' case there was a tactical reason (the desire to preserve a coalition with the left) for distancing themselves from *Signposts*. The party's leader, Miliukov, even found it necessary to undertake a lecture tour around Russia for the purpose of attacking the volume: this did not prevent Lenin from labelling *Signposts* 'an encyclopedia of liberal apostasy'.

The *Signposts* authors failed to spark off a discussion of the intelligentsia's self-image on their own terms; but the volume may be seen as part of a wider mood of self-examination among the Russian intelligentsia which sprang from the philosophical and religious revival of the early twentieth century, and which, in the last decade before the Revolution, was reflected in a growing moral opposition to the outlook of the Bolsheviks from other parties on the left.

5 The sequels

In 1918, as the Bolshevik dictatorship began to hunt down its critics, while civil war raged in the south of Russia and the countryside was ravaged by anarchy and famine, *Signposts*' predictions of a national catastrophe seemed tragically prophetic. Struve, living clandestinely in Moscow, conceived the idea of a sequel which could discuss the causes and meaning of the 1917 Revolution. He brought together all the *Signposts* authors except Gershenzon and Kistiakovskii, and six others, including P.I. Novgorodtsev, who had been the editor of *Problemy idealizma*, and the Symbolist poet Viacheslav Ivanov. All had been prominent in Russian intellectual life before the Revolution.

The new symposium's title, *Iz glubiny (Out of the Depths)*, was taken from the 130th Psalm. It was confiscated before it could be distributed and only two copies found their way to the West. It was virtually unknown until it was reprinted in Paris in 1967. The authors interpret the Revolution as a vindication of *Signposts'* warnings. The essays fall into two main categories: those primarily concerned with what Struve called the 'ultimate religious questions' of human existence, including such topics as the 'Russian soul'; and those that focus on more concrete social and political issues. The two approaches are exemplified in the contributions of Berdiaev and Struve respectively. Struve points to the contradiction between socialism's rational theory and destructive practice, while Berdiaev contends that the Revolution revealed a dark side of the Russian soul: the anarchic nihilism first

diagnosed by Gogol’ and Dostoevskii.

All the contributors express the hope for a spiritual rebirth of the Russian people based on national and religious values, but are unspecific on the form this should take. Berdiaev was then engaged in developing a philosophical interpretation of the Russian Revolution as the prelude to the end of European civilization, which would be apocalyptically transformed into a new universal culture which Russian thinkers like himself (whose spiritual make up, he believed, represented a synthesis of East and West) were destined to reveal to the world. In his book *The Philosophy of Inequality*, published in Berlin in 1923, he would attack the ’tyranny of democracy’, and call for ’total freedom’, which he admits is unrealizable on earth. In *Out of the Depths* as in *Signposts*, the aims of religious utopians conflict with those of liberal humanists, such as Struve, who defends the right to property and economic liberty and the principles of individual responsibility and initiative, although (as in earlier writings on this theme) he emphasizes the importance of a strong and unified state as the guardian of the national culture.

In 1921 the debate over *Signposts* was taken up in emigration with the publication in Prague of a volume of essays entitled *Smena vekh (Change of Signposts)*, the manifesto of a short-lived movement later known as ’National Bolshevism’. Its proponents, mainly former Kadets who claimed to be inspired by the nationalist strand in *Signposts*, called for reconciliation with the Bolsheviks on the grounds that in its sweeping destruction the Revolution had been an authentically Russian phenomenon, an expression of the people’s instincts. A counter-symposium, *O smene vekh (On Change of Signposts)* (1922), whose authors included Izgoev, appeared inside Russia, protesting against what they saw as this group’s distortion of the principles of *Signposts*: as the embodiment of revolutionary nihilism, Bolshevik ideology could not be the basis of a new national culture. Struve (by then in emigration) also strongly repudiated the new movement as intellectually and morally contemptible.

This marked the end of the public debate over *Signposts* in emigration, although the work frequently figured in émigré literature. Its central themes were echoed in the work of two *Signposts* authors in particular: as editor of a succession of émigré periodicals, Struve devoted much of his writing to the question of the meaning of the Russian Revolution, while Berdiaev (who as a philosopher became the best known of the *Signposts* authors in the West) produced interpretations of Russian thought for a Western audience.

Mention of *Signposts* in official Soviet historiography was confined to a recapitulation of Lenin’s attack on it, but in the post-Stalin period Russian dissidents who were rediscovering pre-revolutionary Russian philosophy and religious thought found its critique of the intelligentsia’s values highly relevant to their own time. In 1974 the exiled Solzhenitsyn, with a number of intellectuals who were still in Russia, published a collection of essays modelled on *Signposts* whose title *Iz-pod glyb (From under the Rubble)* was a phonetic echo of the Russian words for *Out of the Depths*. Although differing from *Signposts* on some issues, the collection cites it as a vision of the future which sheds light on the spiritual bankruptcy of the Soviet elite.

One of the first fruits of the new press freedoms under Gorbachev was the republication of *Signposts* and *Out of the Depths*. The lively interest of post-Soviet Russian intellectuals in *Signposts* (which has since gone into several reprints) suggests that it may now play a role in the intelligentsia’s self-examination that was denied it in 1909. Recent commentaries on *Signposts* in the West have linked it with Mikhail Bakhtin’s subsequent critique of dogmatic thought, as part of a minority current in the Russian intellectual tradition which refused to regard all moral questions as political ones, and which is enjoying a significant revival today.

See also: Liberalism, Russian §3; Marxist philosophy, Russian and Soviet §2

AILEEN KELLY

List of works

Berdiaev, N.A. (1901) *Sub’’ektivizm i individualizm v obshchestvennoi filosofii. Kriticheskii ėtiud o N.K. Mikhailovskom, s predisloviem P.B.Struve (Subjectivism and Individualism in Social Philosophy. A critical study of N.K. Mikhailovskii, with a preface by P.B. Struve)*, St Petersburg: Popova.(A critique of the radical intelligentsia’s philosophical outlook from an idealist perspective.)

Novgorodtsev, P.N. (ed.) (1903) *Problemy idealizma. Sbornik statei (Problems of Idealism. A Collection of Articles)*, Moscow: Moskovskoe psikhologicheskoe obshchestvo.(Essays on aspects of the future Signposts movement’s philosophy.)

Vekhi. Sbornik statei o russkoi intelligentsii (1909) Moscow: I.N. Kushnerev; 5th edn, 1910; trans. M.S. Shatz and

J.E. Zimmerman, *Signposts. A Collection of Articles on the Russian Intelligentsia*, Irvine, CA: Charles Schlacks Jr, 1986.(Contains an informative introduction and a bibliography of the debate.)

Iz glubiny: Sbornik statei o russkoi revoliutsii (1918) Moscow; 2nd edn Paris: YMCA-Press, 1967; trans. W.F. Woehrlin, *Out of the Depths. A collection of articles on the Russian Revolution*, Irvine, CA: Charles Schlacks Jr., 1986.(Distribution of the 1918 printing was suppressed for political reasons. The collection survived only because two copies reached the West. The 1967 edition, based on one of these copies, represents its first publication.)

Smena vekh: Sbornik statei (Change of Signposts: A Collection of Articles) (1921) Prague: Politika.(A pro-Bolshevik response to *Signposts*.)

O smene vekh (On Change of Signposts) (1922) Petrograd: Logos.(A reassertion of the principles of *Signposts* against *Smena vekh*.)

Solzhenitsyn, A. *et al.* (1974) *Iz-pod glyb: Sbornik statei*, Paris: YMCA-Press; trans. A.M. Brock *et al*., *From under the Rubble*, Boston: Little, Brown, 1975.(A collection of essays by Russian intellectuals modelled on *Signposts*.)

References and further reading

Burbank, J. (1986) *Intelligentsia and Revolution. Russian Views of Bolshevism 1917-1922*, Oxford: Oxford University Press.(Chs 5 and 6 on the intellectual background to *Out of the Depths* and *Change of Signposts*.)

Kelly, A.M. (1998) *Towards Another Shore: Russian Thinkers between Necessity and Chance*, New Haven, CT: Yale University Press.(Includes an essay on the conflict between utopian and anti-utopian positions in *Signposts*.)

Kindersley, R. (1962) *The First Russian Revisionists*, Oxford: Clarendon Press.(On the early thought of the *Signposts* authors.)

Mendel, A.P. (1961), *Dilemmas of Progress in Tsarist Russia*, Cambridge, MA: Harvard University Press.(A subtle study of the early philosophical evolution of Berdiaev, Bulgakov and Struve.)

Morson, G.S. (1993) ’Prosaic Bakhtin: Landmarks, Anti-Intelligentsialism, and the Russian Counter-Tradition’, *Common Knowledge* 2 (1): 35-74.(Approaches *Signposts* as part of a tradition antithetical to the intelligentsia’s.)

Oberländer, E. (1968) ’Nationalbolschewistische Tendenzen in der russischen Intelligentz: Die ’Smena Vech’-Diskussion 1921-1922’ in *Jahrbücher für Geschichte Osteuropas* 16 (2): 194-211. (On the *Change of Signposts* controversy.)

Oberländer, E. (1965) *Die Vechi-Diskussion 1909-12*, Köln: Ph.D. dissertation.(A summary of positions in the debate.)

Read, C. (1979) *Religion, Revolution and the Russian Intelligentsia 1900-1912: the Vekhi Debate and its Intellectual Background*, London: Macmillan.(Focuses on the role of religious belief in the Signposts debate.)

Schapiro L. (1955) ’The Vekhi Group and the Mystique of Revolution’ in *Slavonic and East European Review* 34 (82): 56-76.(Interprets *Signposts* as a liberal critique of the Russian left).

Silvestri, Francesco (1474-1528)

*A Thomist philosopher and theologian, Silvestri composed, along with Aristotelian commentaries and polemical works, a vast commentary on Aquinas' Summa Contra Gentiles which, from the first, has been recognized as its classic exposition. Silvestri imitated the method (*expositio formalis*) of Cajetan's commentary on Aquinas' Summa Theologiae, but disagreed with Cajetan on key points of Thomistic doctrine and proposed interpretations generally closer to the letter of Aquinas. Chief among these are the doctrines of analogy, abstraction and the rational demonstrability of the soul's immortality.*

Francesco Silvestri was born in Ferrara and, hence, is commonly known as Ferrariensis (or as Francis Sylvester of Ferrara). He entered the Dominican Order in 1488 and completed his initial studies in Bologna in 1497. In 1498 he was appointed lector in Mantua. From 1503-8 he lectured in Milan and from 1509-11 in Bologna. He spent 1511-3 as bachelor of the *Sentences* in Milan and subsequently returned to Bologna where he graduated as master of theology in 1516. He was then elected prior of the convent in Ferrara, and also taught at the university. In 1519 he was appointed inquisitor in Bologna and in 1521-4 he was regent master of the Dominican *studium generale* there. In 1524 he was elected prior of the convent in Bologna and, soon after, Pope Clement VII named him vicar general of the order. He was elected master general at the general chapter held in Rome in 1525 and continued in this office until his death.

Silvestri's doctrine of analogy is presented in his commentary on Aquinas' *Summa contra gentiles* as a critical supplement to that of Cajetan (§2). Unlike the latter, Silvestri is more concerned with accurate exegesis of Aquinas' text than the refutation of Scotism and is, accordingly, more attuned to the determination of analogy's logical structure than to its metaphysical implications. Starting from Aquinas' affirmation that analogy is the predication of terms according to an order of priority and posteriority (see Aquinas §9), he makes a clear distinction between such an order (1) as operative in the psychological process of naming, and (2) as holding in reality and representing the relative positions of things in the hierarchy of being.

Two features are essential to every instance of analogy: first, secondary analogates bear some kind of reference to the primary analogate (*ordo ad unum*); second, the proper notion (*ratio propria*) of the primary analogate enters the definition of secondary analogates. From this follows a division of analogy foreign to Aquinas: analogy may be founded on the relation (or resemblance) which holds between either a plurality of things and one other thing (*duorum ad tertium*) or between one thing and another (*unius ad alterum*). The latter kind is further divided into analogy founded on a direct, fixed relation between one thing and another (analogy of proportion) and analogy founded on an indirect, loose relation between one thing and another (analogy of proportionality). Unlike Aquinas and Cajetan, Silvestri distinguishes between these two latter types of analogy in psychological rather than mathematical terms. The analogical predication of names of God and creatures cannot be of the kind *duorum ad tertium* because there is no prior 'third thing' to which both God and creatures are related (or that both God and creatures resemble), so as to justify the imposition of common terms. It can only be the analogy of proportionality that permits one to talk meaningfully about God and that, unlike analogy of proportion, safeguards the divine transcendence (see Language, Renaissance philosophy of §4).

Silvestri proposed a personal theory of abstraction in a critique of Cajetan's explanation (*In Contra Gentiles Commentaria* II c.77). He argues that Cajetan attributes to the phantasm features of the intelligible object (universality, immateriality), transforming it into a direct object of intellection and rendering the agent intellect redundant. For Silvestri the agent intellect performs a 'radical and fundamental' illumination. It acts upon both the phantasm and the possible intellect, and acts in conjunction with the imagination (*phantasia*) to produce the impressed intelligible species.

Silvestri is original on several other epistemological and metaphysical issues. The possible intellect has the capacity to form a proper concept of the singular sensible object in addition to, and after, the formation of the universal concept which represents the common nature. Numerical multiplication is possible within species of immaterial substances. The individuation of material substances is effected by matter in so far as it is quantified by determined dimensions.

Silvestri has been considered fundamentally a disciple, albeit at times critical, of Cajetan. He should rather be counted a member of the Bolognese Thomist school founded by Peter of Bergamo (d. 1482) and represented during the early sixteenth century by such figures as Sylvester Prierias (1456-1527), Chrysostom Javelli (1470-1543) and Bartholomeo de Spina (d. 1545). Generally opposed to Cajetan on speculative issues, their differences came to the fore during the Pomponazzi affair (see Pomponazzi, P. §2). This is the context of Silvestri's own defence of the rational demonstrability of the soul's immortality. Especially in need of correction is the explanation of the Bolognese Thomists' opposition to Cajetan (and Pomponazzi) in terms of an inability to distinguish between philosophy and Aristotelianism (Laurent 1938; Nardi 1958). This distinction is, in fact, stressed by all of them, including Silvestri.

See also: Aquinas, T.; Aristotelianism, Renaissance; Cajetan; Pomponazzi, P.

MICHAEL TAVUZZI

List of works

Silvestri, F. (1517) *Adnotationes in libros Posteriorum Aristotelis et S. Thomae*, Venice.(Commentary on Aristotle's *Posterior Analytics* and Aquinas' commentary on the same.)

Silvestri, F. (1524) *In libros s. Thomae Aquinatis Contra Gentiles Commentaria (Commentary on Aquinas' Summa contra gentiles)*, Venice; in T. Aquinas, *Opera Omnia*, Editio Leonina, vols 13-15, Rome: Typis Riccardi Garroni, 1918-30.(Begun in 1508 and completed in 1517, the first edition of *In Contra Gentiles Commentaria* appeared in 1524.)

Silvestri, F. (1525) *Apologia de convenientia institutorum Romanae Ecclesiae cum evangelica libertate (Apology for the Appropriateness of the Institutions of the Roman Church for Evangelical Freedom)*, Rome; ed. G. Sestili, Rome: F. Pustet, 1906.(Polemical work against Luther.)

Silvestri, F. (1535) *Quaestiones in III libros De Animae (Questions on the Three Books of Aristotle's On the Soul)*, Venice.(Focuses on the most controversial aspects of Aristotle's *On the Soul*.)

Silvestri, F. (1535) *Quaestiones in libros Physicorum (Questions on Aristotle's Physics)*, Venice.(Focuses on disputed points of Aristotle's *Physics*.)

References and further reading

Bellerate, B.M. (1960) 'L'analogia tomista nei grandi commentatori di S. Tommaso' (Thomistic Analogy in the Great Commentators on Aquinas), Rome: Salesian University, 73-95.(Abstract of a Ph.D. thesis that gives the most detailed account of Silvestri's theory of analogy.)

Giacon, C. (1944) *La seconda scolastica (The Second Scholasticism)*, Milan: Bocca, vol. 1, 37-52, 80-2, 91-114, 130-2, 146-62.(Overview of Silvestri's developments of Aquinas and disagreements with Cajetan.)

Hegyi, J. (1959) *Die Bedeutung des Seins bei den klassischen Kommentatoren des heiligen Thomas von Aquin: Capreolus, Silvester von Ferrara, Cajetan (The Meaning of Being in the Classic Commentators on Aquinas: Capreolus, Silvestri and Cajetan)*, Pullacher Philosophische Forschungen 4, Pullach bei München: Berchmanskolleg, 53-106.(Account of Silvestri's metaphysics.)

Lauchert, F. (1912) *Die italienischen literarischen Gegner Luthers*, Freiburg: Herder, 269-78.(Account of Silvestri's anti-Luther polemics.)

Laurent, M.-H. (1938) 'Introductio', in Cajetan, *Commentaria in De Anima Aristotelis*, ed. P.I. Coquelle, Rome: Institutum Angelicum, vol. 1, xlvii-li.(Influential, but erroneous, evaluation of Silvestri.)

Lohr, C. (1988) 'Silvestri (de Silvestris) Ferrariensis, Franciscus O.P.', *Latin Aristotle Commentaries: II, Renaissance Authors*, Florence: Olschki, 422-3.(Not entirely accurate bio-bibliography, lists editions of commentaries on Aristotle.)

Mortier, A. (1911) *Histoire des maîtres généraux de l'ordre des Frères Prêcheurs*, Paris: Picard, vol. 5, 260-84.(Account of Silvestri's activities as Dominican master general.)

McInerny, R.M. (1961) *The Logic of Analogy*, The Hague: Nijhoff, 23-31.(Brief account of Silvestri on analogy.)

Nardi, B. (1958) *Saggi sull'Aristotelismo Padovano (Essays on Paduan Aristotelianism)*, Florence: Sansoni, 222-3. (Appropriates Laurent's evaluation of Silvestri.)

Sestili, G. (1923) 'Francesco Silvestri', *Gli scienziati italiani*, Rome: Nardecchia, vol. 1, 128-37.(Most thorough bio-bibliography, not entirely reliable.)

Simmel, Georg (1858-1918)

Georg Simmel was a prolific German philosopher and sociologist, who was one of the principal founders of sociology in Germany. His philosophy and social theory had a major impact in the early decades of the twentieth century, both among professional philosophers and sociologists and within the cultural and artistic spheres. This is true of his foundation for sociology, his philosophy of art and culture, his philosophy of life and his philosophy of money. His thought ranged from substantive issues within the philosophical tradition to a concern with the everyday world and its objects.

Although he insisted on many occasions that he wished to be recognized as a philosopher and despite his extensive philosophical writings, Simmel is today still largely known as one of the founders of German sociology at the turn of the nineteenth century.

Simmel taught at Berlin University from 1885 to 1914 until he obtained a chair in philosophy at Strasbourg University, which he held until his death. Although he studied philosophy (under Eduard Zeller and Friedrich Harms), history and ethnology at Berlin University, Simmel regarded his most significant teacher to be the *Völkerpsychologist* Moritz Lazarus. Simmel's writings, especially in the early period, manifest a wide range of influences: Darwin, Herbert Spencer, Gustav Fechner, Lazarus, Kant, Schopenhauer and others. Not surprisingly he was subsequently categorized as a positivist, pragmatist, Neo-Kantian, the 'German Bergson', a philosopher of life and a relativist.

His doctoral dissertation of 1881 on Kant's monadology indicated the commencement of a critical confrontation with Kantian philosophy, but despite his later contact with leading Neo-Kantians, such as Heinrich Rickert, it is not easy to subsume Simmel's philosophy under that of the Neo-Kantian tradition (see Neo-Kantianism, §2). His early monographs displayed seemingly diverse theoretical interests. *Über sociale Differenzierung (On Social Differentiation)* (1890) contained Simmel's first attempt to ground sociology methodologically as an independent discipline, as well as applications of his sociological method. *Die Probleme der Geschichtsphilosophie (The Problems of the Philosophy of History)* (1892) argued against the existence of universal historical laws, but it was its second (totally rewritten) edition of 1905 which had a significant impact upon methodological discussions within the social sciences on historical method and problems of interpretive understanding. The two-volume *Einleitung in die Moralwissenschaft (Introduction to Moral Science)* (1892-3) was devoted to rejecting universal ethical concepts and principles as an a priori formalistic grounding for moral philosophy in favour of a critical, psychological, sociological and historical investigation of the diversity of ethical norms in their everyday practice.

If several of these early works did not receive substantial acclaim, the same cannot be said of Simmel's most sustained philosophical study *Philosophie des Geldes (The Philosophy of Money)* (1900). If, as stated in its preface, one of philosophy's aims is to bring together the fragments of knowledge into a totality, then money serves as an instance for the presentation of relations between the most diverse and often fortuitous objects and for the possibility of discovering in each of the details of life the totality of its meaning. The philosophy of value appropriate for money as the universal equivalent of all values is some version of relativism or relationism. Simmel's wide ranging philosophy of money makes important contributions to the theory of value, to the philosophy of social action within a means/end framework, to a theory of the individual person and individual freedom, and to the aesthetics of modern society. Money as symbol of the eternal flux of interrelations between the most diverse phenomena serves to generate a theoretically grounded relativism. Many of Simmel's contemporaries viewed this rich synthesis as a philosophy of specifically 'modern' times and its author as possessing an instinct for the philosophical exploration of modernity.

In the period after publication of *The Philosophy of Money*, Simmel intended working on a volume on the philosophy of art and aesthetics, a project which, despite an extensive number of essays in this area on landscape, the picture frame and other aesthetic themes, never came to fruition. However, he did publish his highly popular and influential introductory lectures on *Kant* (1904) and on *Schopenhauer und Nietzsche* (1907) - both volumes serving as testimony to the impact of these philosophers upon Simmel's own development.

The work which did have a major influence was his *Soziologie (Sociology)* (1908), the culmination of more than a decade's concern with explicating his version of sociology. There, numerous substantive applications of his sociological method were preceded by Simmel's most elaborated foundation of sociology as the study of the forms of sociation and social interaction. The concept of interaction or reciprocal effect (*Wechselwirkung*) which pervades his philosophical writings lies at the centre of his foundation of sociology, focusing on the analysis of forms that social relations take. Human sociation is created in social interactions and Simmel's programme for sociology opens up the study of any and all forms of sociation as a basis for understanding how society is possible. This accounts for the extensive breadth of Simmel's sociological - and philosophical - investigations, an insistence upon the fundamental interrelatedness of all phenomena, a relational sociology and philosophy.

The necessary abstraction and separation of form from content which his sociology requires is itself indicative of more fundamental separations pervading his philosophy as a whole, but most fully developed in his later writings, namely between subjective and objective culture and between life and form. Explorations of the dialectical relationship between subjective and objective culture are at the heart of Simmel's philosophy and sociology of culture; the investigation of the opposition between life and form is at the centre of his philosophy of life (*Lebensphilosophie*). (See Lebensphilosophie).

Simmel's theory of culture radicalized the widening gap between subjective, individual culture and a more rapidly expanding objective culture - both material and non-material. This problematic of cultural alienation, in which human subjects must still realize themselves within an objective culture from which they are estranged, had a significant influence upon cultural and social theorists (especially in the tradition of critical theory). At the same time, essay collections such as *Philosophische Kultur (Philosophical Culture)* (1911) revealed a concern to develop not merely a philosophy of *culture* but also a *philosophical* culture capable of analysing this and other problematics.

Simmel's later writings also developed a philosophy of life grounded in the opposition of life and form in which life as a dynamic dimension struggled against rigidified form in a dialectical tension. This suggested affinities with Bergson's vitalism (see Bergson, H.-L.), but also significantly influenced contemporary artistic movements, notably German Expressionism.

DAVID FRISBY

List of works

Simmel, G. (1988-) *Gesammelte Werke (Complete Works)*, ed. O. Rammstedt, Frankfurt: Suhrkamp, 24 vols.(A recent critical edition of Simmel's complete works.)

Simmel, G. (1890) *Über soziale Differenzierung (On Social Differentiation)*, ed. H.J. Dahme, Frankfurt: Suhrkamp, 1989.

Simmel, G. (1892) *Die Probleme der Geschichtsphilosophie*, Leipzig: Duncker & Humblot, 1905; trans. and ed. G. Oakes, *The Problems of the Philosophy of History. An Epistemological Essay*, New York: Free Press, 1977.

Simmel, G. (1892-3) *Einleitung in die Moralwissenschaft. Eine Kritik der ethischen Grundbegriffe (An Introduction to Moral Science. A Critique of Basic Ethical Concepts)*, ed. K.C. Köhnke, Frankfurt: Suhrkamp, 1989, 1991.

Simmel, G. (1900, 1907) *Philosophie des Geldes*, Leipzig: Duncker & Humblot; trans. T. Bottomore and D. Frisby, *The Philosophy of Money*, London: Routledge, 1978.

Simmel, G. (1904) *Kant*, Leipzig: Duncker & Humblot.

Simmel, G. (1906a) *Kant und Goethe*, Berlin: Bard, Marquardt.

Simmel, G. (1906b) *Die Religion*, Frankfurt: Rütten & Loening.

Simmel, G. (1907) *Schopenhauer und Nietzsche*, Leipzig: Duncker & Humblot; trans. H. Loiskandl, D. Weinstein and M. Weinstein, *Schopenhauer and Nietzsche*, Amherst, MA: University of Massachussetts Press, 1986.

Simmel, G. (1908) *Soziologie*, ed. O. Rammstedt, Frankfurt: Suhrkamp, 1992; partly trans. K.H. Wolff, *The Sociology of Georg Simmel*, New York: Free Press of Glencoe, and London: Collier-Macmillan, 1964.

Simmel, G. (1910) *Hauptprobleme der Philosophie (Key Problems of Philosophy)*, Leipzig: Göschen.

Simmel, G. (1911) *Philosophische Kultur. Gesammelte Essais (Philosophical Culture. Collected Essays)*, Leipzig: W. Klinkhardt. (1922) *Zur Philosophie der Kunst. Philosophische und kunstphilosophische Aufsätze (On the Philosophy of Art. Articles on Philosophy and Philosophy of Art)*, ed. G. Simmel, Potsdam: Kiepenheuer.

Simmel, G. (1987) *Das individuelle Gesetz. Philosophische Exkurse (The Individual Law. Philosophical*

Excursions), ed. M. Landmann with afterword by K.C. Köhnke, Frankfurt: Suhrkamp.
Simmel, G. (1959) *Georg Simmel, 1858-1918. A Collection of Essays with Translations and a Bibliography*, ed. K.H. Wolff, Columbus, OH: Ohio State University Press.
Simmel, G. (1968) *The Conflict in Modern Culture and Other Essays*, trans. and ed. P.K. Etzkorn, New York: Teachers College Press.
Simmel, G. (1971) *On Individuality and Social Forms. Selected Writings*, ed. D.N. Levine, Chicago, IL: University of Chicago Press.
Lawrence, P.A. (ed.) (1976) *Georg Simmel: Sociologist and European*, trans. D.E. Jenkinson *et al.* and with intro. by P.A. Lawrence, Sunbury: Nelson.
Simmel, G. (1980) *Essays on Interpretation in Social Science*, G. Oakes, Manchester: Manchester University Press.
Frisby, D. and Featherstone, M. (eds) (1996) *Simmel on Culture: Selected Writings*, with intro. by D. Frisby, London: Sage.

References and further reading

Frisby, D. (1984) *Georg Simmel*, London: Tavistock.(An introduction to his social theory.)
Frisby, D. (1992) *Simmel and Since*, London: Routledge.(Essays on Simmel's social theory.)
Frisby, D. (ed.) (1994) *Georg Simmel. Critical Assessments*, London: Routledge, 3 vols. (Assessments of most aspects of Simmel's work and key early reviews.)
Kaern, M., Phillips, B. and Cohen, R.S. (eds.) (1990) *Georg Simmel and Contemporary Sociology*, Dordrecht, Boston, MA, and London: Kluwer.(A valuable collection, also containing essays on aspects of Simmel's philosophy.)
Köhnke, K.C. (1996) *Der junge Simmel* (The Young Simmel), Frankfurt: Suhrkamp. (The first systematic examination of Simmel's early philosophical writings.)
Kracauer, S. (1995) 'Georg Simmel' in *The Mass Ornament*, Cambridge, MA: Harvard University Press, 225-7. (A brief introduction to Simmel's philosophy by one of his students.)
Poggi, G. (1993) *Money and the Modern Mind*, Berkeley, CA, and London: University of California Press.(An introduction to Simmel's *Philosophy of Money*.)
(1991) 'A Special Issue on Georg Simmel', *Theory, Culture & Society* 8 (3).(Contains translations and essays largely on culture.)
Weingartner, R.H. (1962) *Experience and Culture: The Philosophy of Georg Simmel*, Middletown, CT: Wesleyan University Press.(Looks especially at his later philosophy.)
Weinstein, D. and Weinstein, M. (1993) *Postmodern(ized) Simmel*, London: Routledge.(Relates Simmel's philosophy to postmodern theorists.)

Simplicity (in scientific theories)

In evaluating which of several competing hypotheses is most plausible, scientists often use simplicity as a guide. This raises three questions: what makes one hypothesis simpler than another? Why should a difference in simplicity make a difference in what we believe? And how much weight should simplicity receive, compared with other considerations, in judging a hypothesis' plausibility? These may be termed the descriptive, the normative, and the weighting problems, respectively. The aesthetic and pragmatic appeal of more simple theories is transparent; the puzzle is how simplicity can be a guide to truth.

1 Curve-fitting, parsimony and unification

Considerations of simplicity play a role in the *curve-fitting problem*, depicted in the figure. Suppose a scientist wants to discover what general relationship obtains between the independent variable *x* and the dependent variable *y*, and pursues this question by gathering a set of observations, each of which corresponds to a point.

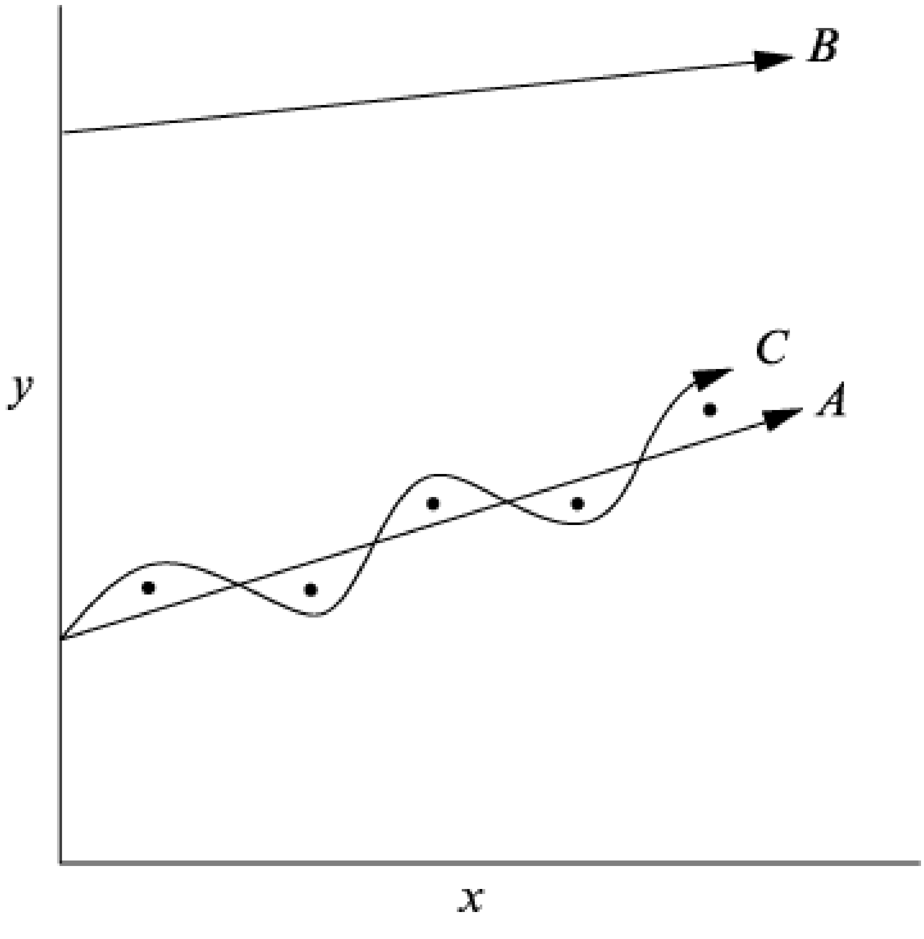

What role might the data play in evaluating different competing hypotheses, each of which corresponds to a curve? If observational error is impossible, then any curve that fails to pass exactly through the data points must be false. However, error is always possible. This means that the true curve may fail to pass exactly through the data points. How, then, are the data to be used?

The standard scientific procedure is to use the concept of 'goodness-of-fit'. Consider Curve *A*. For each data point, compute the squared distance from the observed *y*-value to the *y*-value that Curve *A* predicts. The 'sum of squares' (SOS) measures how far the curve is from the data. If SOS measures how well the data support the curve in question, then the data in the figure support Curve *A* better than they support Curve *B*.

The rationale for SOS as a measure of evidential support is not far to seek. Curves that are close to the data are more likely, in the technical sense defined by R.A. Fisher (1925). Curve *A* is more likely than Curve *B* (given standard assumptions about the probability of error) because *A* confers on the data a higher probability than *B*

does: $\mathrm{P}(\mathrm{Data} \mid \mathrm{Curve}\ A) > \mathrm{P}(\mathrm{Data} \mid \mathrm{Curve}\ B)$. It is important in this context not to confuse likelihood and probability. The conclusion is not that *A* is more probable - that is, that $\mathrm{P}(\mathrm{Curve}\ A \mid \mathrm{Data}) > \mathrm{P}(\mathrm{Curve}\ B \mid \mathrm{Data})$.

Although likelihood as measured by SOS helps discriminate between Curve *A* and Curve *B*, it says nothing concerning how *A* compares with Curve *C*. They have the same SOS values. It is at this point that simplicity is said to be relevant. Curve *A* is smooth and Curve *C* is bumpy. Curve *A* seems more simple. If simplicity guides our judgements about which curve is more plausible, we will conclude that Curve *A* is superior to Curve *C*, given the data at hand.

If curve-fitting is subject to the twin criteria of goodness-of-fit and simplicity, the descriptive, normative, and weighing problems must all be addressed. The last of these is quite fundamental. In general, the goodness-of-fit of a hypothesis can be improved by making it more complicated. In fact, with *n* data points, *perfect* goodness-of-fit (an SOS of zero) can always be secured by a polynomial of degree *n* (that is, an equation of the form $y = a_1 + a_2x + a_3x^2 + \ldots + a_nx^{n-1}$). If there are just two data points, a straight line will fit them perfectly; if there are three, a parabola can be found with an SOS of zero, and so on. The practice of science is to view goodness-of-fit as one consideration relevant to evaluating hypotheses, but not the only one. The problem is to understand that practice.

Simplicity considerations play a role in other inferential contexts. Ockham's razor, the principle of parsimony, says that hypotheses that postulate fewer entities, causes or processes are better than hypotheses that postulate more. In addition, the preference for simpler hypotheses finds expression in the idea that a unified theory that covers two bodies of phenomena is sometimes more plausible than a disunified theory that treats them as separate and independent. An adequate account of the role of simplicity in curve-fitting should bring in its wake an understanding of the value that science seems to place on parsimonious and unified theories.

2 Bayesian approaches

Bayes' theorem shows how the posterior probability $\mathrm{P}(H \mid O)$ of a hypothesis - the probability that *H* has, given that the observation *O* has been made - is related to the probability $\mathrm{P}(H)$ the hypothesis had prior to the observation: $\mathrm{P}(H \mid O) = \mathrm{P}(O \mid H)\mathrm{P}(H)/\mathrm{P}(O)$ (see Probability, interpretations of §5; Statistics §2). When two hypotheses, H_1 and H_2, are evaluated in light of the observations *O*, the theorem entails that $\mathrm{P}(H_1 \mid O) > \mathrm{P}(H_2 \mid O)$ if and only if $\mathrm{P}(O \mid H_1)\mathrm{P}(H_1) > \mathrm{P}(O \mid H_2)\mathrm{P}(H_2)$. That is, if H_1 has the higher posterior probability, then it must have the higher likelihood or the higher prior probability (or both).

Let us apply the Bayesian framework to the curve-fitting problem represented in the figure. How could Curve *A* have a higher posterior probability than Curve *C*, relative to the data depicted? There are two possibilities to explore, and one of them will not help. We have already noted that the two curves have the same likelihoods. This means that a Bayesian must argue that Curve *A* has the higher prior probability.

However, if prior probabilities are to be assigned to curves, a problem arises. The straight lines in the *x*-*y* plane are not denumerable. Apparently, they are equally simple. If their prior probabilities are to reflect just their simplicity, then they must be assigned the same prior probability - namely, zero. But hypotheses assigned a prior probability of zero cannot increase in probability when favourable evidence is obtained.

An alternative approach would be to assign probabilities, not to specific curves, but to families of curves. For example, the straight lines comprise a family of curves (LIN), each having the form $y = a_1 + a_2x$, while (PAR) is the family of parabolas, each having the form $y = a_1 + a_2x + a_3x^2$. In these equations, the a_i terms are *adjustable parameters*; members of the family are specified by assigning values to these parameters.

If the assignment of prior probabilities is limited to a denumerable set of such families, with simpler families receiving higher values, it is possible to construct an infinite series of positive prior probabilities that converges on a sum that is not greater than unity. This is the approach that Harold Jeffreys (1957) adopts.

It is important to take care that the families in this series are disjoint. For example, since (LIN) is included in (PAR), (LIN) cannot have a higher probability. With nested families of curves, more complex families have higher probabilities. One solution is to define the family (PAR′), which is the subset of (PAR) in which all adjustable parameters are nonzero. No contradiction ensues if one assigns (LIN) a higher prior than (PAR′).

However, a problem now arises, one that concerns Bayesianism in general. If prior probabilities merely report an agent's subjective degrees of belief, then there is no reason why we should assign (LIN) a higher prior than (PAR′) rather than doing the reverse. And once this normative question is addressed, it must puzzle the Bayesian why (LIN) should be more probable than (PAR′). After all, (LIN), in effect, sets the parameter $a_3 = 0$, while (PAR′) stipulates that $a_3 \neq 0$. What could show that the former is more probable than the latter?

Bayesians may reply that their goal is to describe the practice of science, not to justify it. Another option is to abandon the goal of analysing the complexity of families of curves and to focus exclusively on some denumerable set of curves with no adjustable parameters. A third would be to theorize about the likelihoods of families of curves as a measure of their simplicity (Rosenkrantz 1977).

3 Popper

Karl Popper (1959) turns the Bayesian approach on its head. For him, the goal of science is to find highly *im*probable hypotheses. Hypotheses should be valued for their falsifiability - for their saying more, rather than less. Popper conjoins this requirement with the idea that we should reject hypotheses that have been falsified. So the two-part epistemology he proposes is that we should prefer hypotheses that are unfalsified though highly falsifiable. This means that we should prefer (LIN) over (PAR) if neither has been falsified by the data. (LIN) is more falsifiable, says Popper, because at least three data points are needed to falsify it; for (PAR) to be falsified, at least four observations are required.

One problem for Popper's proposal arises when it is applied to equations that contain no adjustable parameters. If we consider, not (LIN) and (PAR), but some specific straight line and some specific parabola, then each can be falsified by a single observation. All specific curves are equally falsifiable.

A further limitation in the Popperian approach should be noted. A purely deductivist methodology cannot explain why scientists reasonably decline to reject curves that fail to pass exactly through the data points. The three curves in the accompanying figure are all 'falsified' by the data. However, as Popper sometimes intimates, a more adequate assessment of the fit of hypotheses to data is afforded by the concept of likelihood; rather than considering the yes/no question of whether a hypothesis has been falsified, we need to consider the quantitative issue of the degree to which a hypothesis says that the observations were to be expected. It is arbitrary to impose on that continuum a cut-off point that separates the stipulated categories of 'falsified' and 'unfalsified'. However, once likelihood is substituted for the concept of deductive falsification, the weighting problem arises as a significant issue.

4 Akaike

The statistician, H. Akaike and his school have developed a set of theorems that show how the predictive accuracy of a family of curves may be estimated (1973).

The predictive accuracy of a family reflects how well it does in predicting new data when its parameters are estimated from an old data set. Consider (LIN) as an example. First find the member of (LIN) that best fits the available data, D_1. This straight line is the likeliest member of that family; call it $L(\text{LIN})$. Then obtain a new set of data, D_2, from the true (and unknown) underlying curve, and measure how well $L(\text{LIN})$ fits D_2. Now imagine repeating this process, first finding the best fitting member of (LIN) with respect to one data set and then using that curve to predict new data. The average performance of (LIN) defines its predictive accuracy, which also might be termed its closeness to the truth.

Akaike's theorem says that the following provides an *unbiased* estimate of the distance from the truth of family *F*:

$$\text{Estimate (Distance from the truth of family } F) = \text{SOS}[L(F)] + 2k\sigma^2 + \text{constant}$$

Here *k* is the number of adjustable parameters in *F*, and σ^2 is the error variance - the degree of dispersion of observations around the true curve. The third addend disappears when hypotheses are compared with each other, and so may be ignored.

The second addend in the displayed equation gives simplicity its due: the complexity of a family is measured by how many adjustable parameters it contains. It is important to note that the number of adjustable parameters is not

a syntactic feature of an equation. Although the equations ‘ $y = ax + bx$’ and ‘ $y = ax + bz$’ may each seem to contain two parameters (*a* and *b*), this is not the case. The former equation can be *reparameterized*. Let $a' = a + b$; now the first equation can be restated as ‘ $y = a'x$’. In fact, the first equation contains one parameter, while the second contains two.

The second term in Akaike’s theorem, besides adverting to the number of adjustable parameters, also mentions the error variance. When this variance is large, this second term plays a larger role in estimating the family’s distance from the truth; when observations are largely free of error, it makes only a negligible contribution. Simplicity cannot matter when observational error is impossible; it matters more and more as the data become noisier.

Akaike’s theorem applies to one of the most famous examples of the role of simplicity considerations in hypothesis evaluation - the dispute between Copernican and Ptolemaic astronomy. In Ptolemy’s geocentric model, the relative motion of the earth and the sun is replicated within the model for each planet; the result is a system containing a very large number of adjustable parameters. Copernicus decomposed the apparent motion of the planets into their individual motions around the sun and a *common* sun-earth component, thereby drastically reducing the number of adjustable parameters. In *De revolutionibus*, Copernicus argued that the weakness of Ptolemaic astronomy derives from its failure to impose principled constraints on the separate planetary models. Thomas Kuhn (1957) and others have claimed that the greater unification and ‘harmony’ of the Copernican system is a merely aesthetic feature. The Akaike framework offers a quite different diagnosis: since the Copernican system fitted the data then available about as well as the Ptolemaic system did, Akaike’s theorem entails that the former had an *astronomically* superior degree of estimated predictive accuracy.

See also: Confirmation theory; Inductive inference; Inference to the best explanation; Scientific realism and antirealism; Statistics; Theoretical (epistemic) virtues

ELLIOTT SOBER

References and further reading

Akaike, H. (1973) ‘Information Theory and an Extension of the Maximum Likelihood Principle’, in B. Petrov and F. Csaki (eds) *Second International Symposium on Information Theory*, Budapest: Akademiai Kiado, 267-81. (Develops the Akaike information criterion for estimating a model’s predictive accuracy.)

Fisher, R.A. (1925) *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd.(Presents likelihood as a measure of the degree to which a set of observations supports a hypothesis.)

Forster, M. and Sober, E. (1994) ‘How to Tell when Simpler, More Unified, or Less *Ad Hoc* Theories will Provide More Accurate Predictions’, *British Journal for the Philosophy of Science* 45: 1-36.(Provides exposition and application of Akaike’s ideas, as well as criticism of some other proposals.)

Hesse, M. (1967) ‘Simplicity’, in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York: Macmillan, vol. 7, 445-8.(A useful overview of philosophical accounts.)

Howson, C. (1988) ‘On the Consistency of Jeffreys’s Simplicity Postulate and its Role in Bayesian Inference’, *Philosophical Quarterly* 38: 68-83.(Defends Jeffreys’ approach against Popper’s criticisms, but concludes that it is ultimately unworkable.)

Jeffreys, H. (1957) *Scientific Inference*, Cambridge: Cambridge University Press, 2nd edn.(Presents a Bayesian solution to the problem of understanding the role of simplicity in hypothesis evaluation.)

Kuhn, T. (1957) *The Copernican Revolution*, Cambridge, MA: Princeton University Press.(Argues that the appeal to ‘harmony’ that Copernicus made in defence of his hypothesis invoked a merely aesthetic consideration.)

Popper, K. (1959) *The Logic of Scientific Discovery*, London: Hutchinson.(Defends the idea that the simplicity of a hypothesis reflects its degree of falsifiability, and for this reason is relevant to the task of hypothesis evaluation.)

Rosenkrantz, R. (1977) *Inference, Method, and Decision*, Dordrecht: Reidel.(A Bayesian proposal in which simplicity is connected with the idea of average likelihood.)

Sober, E. (1988) *Reconstructing the Past: Parsimony, Evolution and Inference*, Cambridge, MA: MIT Press. (Explores why common cause explanations are superior to separate cause explanations in the context of the evolutionary problem of phylogenetic inference.)

Sober, E. (1994) ‘Let’s Razor Ockham’s Razor’, in E. Sober (ed.) *From a Biological Point of View*, Cambridge: Cambridge University Press, 136-57.(Argues that invoking the principle of parsimony is a surrogate for an assumed background theory, which, once stated explicitly, renders appeal to parsimony superfluous.)

Simplicity, divine

To be complex is to have many parts. To be simple is to have few. Theists of all religious traditions have asserted that God is completely simple - that is, has no parts of any sort. A contemporary statement of this doctrine of divine simplicity would add that God is identical with each of his intrinsic attributes. Thus if God is omnipotent and omniscient, for example, then he is identical with omnipotence and omniscience (and so these attributes are also identical). The doctrine's popularity across traditions may rest on theists' shared convictions that God is the ultimate reality, perfect, and creator of all that is not himself, and on the fact that many theists have thought that each of these claims entails the doctrine of divine simplicity. It may also rest on strands of mystical experience common to many traditions.

The doctrine has played an important part in ontological arguments for the existence of God because of its assertion of the identity of God with God's nature (if the latter in some sense exists, so must the former). Indeed, divine simplicity is near the conceptual core of classical theism; it is one chief reason classical theists think God immutable, impassible, timeless and wholly distinct from the universe. Many who oppose divine simplicity do so because of these other doctrines it entails. Other recent critics have charged that divine simplicity makes God a mere abstract entity or requires him to have all his attributes necessarily or all contingently.

1 The popularity of simplicity

The claim that God is simple asserts that God has no parts of any sort. Different thinkers have had different inventories of kinds of part. But the doctrine of divine simplicity claims that *whatever* kinds of part there are, God lacks them. (This is why Trinitarian Christians have been able to accept divine simplicity: whatever the Trinity's persons are, they are not parts of God (see Trinity).) So thinkers' beliefs in different kinds of part need not give them different versions of divine simplicity. Medieval friends of divine simplicity treated intrinsic accidental attributes and essential attributes distinct from their subject as among its parts (see, for example, *Summa theologiae* Ia 3, 3 and 6). So the doctrine of divine simplicity led them to say that God lacks these, that is, that God is not distinct from his essential attributes and has no intrinsic accidents. Given divine simplicity, then, if God is essentially omnipotent, God is identical to omnipotence. 'Omnipotence' *seems* to pick out an essential attribute God has, but *really* refers just to God. Further, if God is essentially omniscient, God is identical to omniscience, and so omniscience is identical to omnipotence.

But there is a question as to how we interpret these attribute-identities. Most medieval thinkers before Duns Scotus held that, save for 'proper accidents' (of which a simple God has none), if necessarily whatever is *F* is *G* and whatever is *G* is *F*, *F*ness and *G*ness are identical. But these thinkers used a concept of necessity weaker than broadly logical necessity (see Knuuttila 1993). As they use the term 'necessary', it can be the case that necessarily whatever is *F* is *G* and vice versa even if it is *logically* possible that something be *F* but not *G*. On their modal views, then, it can be logically possible that something be *F* but not *G* and yet be the case that *F*ness and *G*ness are identical. So if pre-Scotus medieval philosophers say that intrinsic divine attributes are identical, they may make a weaker claim than we would make by these words. If *we* say that *F*ness is identical to *G*ness, we imply that nothing can be *F* but not *G*. If medieval philosophers who say that *F*ness is identical to *G*ness do not imply this, then our concept of identity may not correctly represent their claim. Thus medieval versions of divine simplicity may not identify all of God's intrinsic attributes, but only those *F*s and *G*s for which they would say that nothing can be *F* but not *G* or *G* but not *F*. So perhaps divine simplicity's medieval partisans would let God have logically contingent attributes distinct from his essence (Stump and Kretzmann 1985). Still, many medieval arguments which deny that God has intrinsic accidents work as well against the claim that God has logically contingent attributes (see, for example, *Summa theologiae* Ia, 3, 6). So it is likely that the modal theory medieval philosophers employ betrays their intent, and that they *meant* to identify God with all of God's intrinsic attributes.

Divine simplicity has been in the mainstream of Western theism at least since Aristotle (§16). It moved from Middle Platonism into Judaism through Philo of Alexandria. The doctrine entered Christianity through such Church Fathers as Athenagoras, Irenaeus and Clement of Alexandria (see Platonism, medieval §§1-2). It became orthodox in Augustine (§7) and reigned unchallenged through to the time of Duns Scotus. The key to Plotinus' theology, and so to the change from Middle Platonism to Neoplatonism, is his making divine simplicity the core of

his account of the One, the supreme God; as Stoicism and Epicureanism faded away, versions of Platonism embodying divine simplicity became the theology of late paganism (see Plotinus §3; Neoplatonism §3). Divine simplicity was the main point at issue in one of the earliest Islamic theological debates, that over the status of the divine attributes. This debate later became that between orthodox Islamic theologians and partisans of Aristotle, and so enlisted Ibn Sina, al-Ghazali and Ibn Rushd (see Islamic theology §3). Given the ties between medieval Arabic and Jewish philosophy, divine simplicity naturally became a prime topic of medieval Jewish theology as well. Mystics of all periods, East and West, have claimed divine simplicity as a direct deliverance of their experience.

2 The case for simplicity

One may well wonder why (save for mystical experience) theists of so many traditions and times concur in the doctrine of divine simplicity, or even care about so abstract a claim. One reason is that basic theistic intuitions seem to converge on divine simplicity. The intuition that God is the ultimate reality (see God, concepts of §1) suggests divine simplicity, for anything composed of parts (or having distinct attributes) is less basic than they and depends on them (*Summa contra gentiles* I, 18). The intuition that God is creator of all that is distinct from himself suggests divine simplicity, for if God was composed of parts or had essential attributes distinct from himself, these would be distinct from him, yet he could not have created them without creating himself. The intuition that God is perfect suggests divine simplicity, for (at least to classical theists) it suggests that God is maximally self-sufficient. Whatever has parts or essential attributes distinct from itself depends on these and so is less self-sufficient than a simple being.

A second reason for divine simplicity's popularity has been that it is a fertile source of further claims about God; it yields a strong, interesting theological theory. First, if God is simple, then:

(N) necessarily, whatever is God is simple.

For if God is simple, the attribute of being divine is identical to the attribute of being simple. (N) entails that whatever is God is:

Unique. There can be at most one God. For suppose that there are two simple beings, *A* and *B*. *A* and *B* must differ in some attribute, else they would be identical. So suppose that *A* and *B* differ in that *A* has and *B* lacks an attribute *F*. If *F* is the attribute of being simple, then if *A* and *B* are both simple, both *A* and *B* have it. But *A* and *B* *differ* in *F*. So *F* is not being simple. But then if *A* has *F*, *A* has two distinct attributes, *F* and being simple. Whatever has two distinct attributes is not simple; all of a simple being's attributes are identical, as all are identical with the simple being itself. Hence if *A* has *F*, *A* is not simple. So only *B* is simple. Thus there can be at most one simple being. So given (N), there can be at most one God: divine simplicity guarantees monotheism. This helped draw Jews and Muslims to divine simplicity, given their creedal emphasis that God is one; Maimonides was so enthusiastic as to argue that monotheism entails divine simplicity (see *The Guide for the Perplexed*). The tie to monotheism may also have commended divine simplicity to Trinitarian Christians facing charges of tritheism (see Monotheism).

Immaterial. Whatever has matter in its make-up contains that matter's parts. Divine simplicity thus guarantees God's complete distinctness from the universe, or his transcendence.

Immutable. Whatever changes becomes different in a part or attribute (else there was no change), yet is the same in some parts or attributes (else there was not change, but a thing's disappearing and being replaced by another). So what has no parts or distinct attributes cannot change (see Immutability §§1-2). Aquinas and many others argue that any immutable substance is:

Timeless. It has no past or future, its life neither containing nor located in any series of earlier and later events (*Summa theologiae* Ia, 9, 1 and 10, 2). So if this is right, divine simplicity entails timelessness (see Eternity).

If nothing created God, divine simplicity guarantees that God is:

A se. God is entirely independent. For if God has no parts or distinct attributes, nothing can make a causal contribution to his existence except by making him *ex nihilo*, or out of no prior material. If God is *a se*, he is:

Impassible. He is completely unaffected by other things. If other things affected God, putting him in a state *S*, God would depend on those things for *S* and so not be wholly *a se*. Further, given divine simplicity, God would be identical with *S*. So whatever caused *S* to exist would have caused God to exist.

Given appealing assumptions, divine simplicity entails that God is:

Necessarily existent. If he exists, he cannot fail to exist. Assume that:

(a) God exists,
(b) if *P* is actual, then necessarily, *P* is at least possible (the Brouwer axiom), and
(c) entities have attributes only if they actually exist.

Given (a) and (b), then necessarily, it is possible that God exists. On (c), if possibly God exists, this is because some entity has an attribute which makes this true: presumably because God's nature exists and is possibly exemplified. So given (a)-(c), necessarily, God's nature exists. On divine simplicity, God is identical to God's nature. So if God's nature exists necessarily, God exists necessarily. So given (b), (c) and divine simplicity, if God exists, God exists necessarily (see Necessary being §§1-2).

Divine simplicity explains the common theistic claim that God is present as a whole wherever he is present (see Omnipresence §4): what has no parts is present as a whole if at all. Again, given divine simplicity, God's psyche is peculiarly simple. This can help to explain God's moral perfection. Traditionally, this includes *impeccability*, or inability to sin (see, for example, *Summa theologiae* Ia, 25, 3 *ad.* 2). If God is simple and wants to do good, then God is impeccable. For given divine simplicity, all God's desires are identical with God's desire to do good. So if nothing can be both a desire to do good and a desire to do evil, God cannot have a desire to do evil. But his necessary omniscience guarantees that he cannot do evil through ignorance, and his necessary omnipotence guarantees that nothing can force him to do evil. So divine simplicity, a desire to do good, necessary omnipotence and necessary omniscience yield impeccability. Again, if God's psyche is simple, his knowing is identical to his willing. This generates appealing theories about God's relation to morality and modality (Mann 1989), as does divine simplicity's implication that God is identical to God's nature (Kretzmann 1983).

Divine simplicity implies all these conclusions, but most do not in turn entail divine simplicity. Thus divine simplicity is one of the strongest, most basic assertions of theologies in which it appears; it is near the conceptual roots of the classical theist's concept of God (see God, concepts of §6).

A third reason for divine simplicity's popularity has been its connection with efforts to prove God's existence. Historically, Aristotelians inferred divine simplicity directly from their cosmological arguments (for example, *Summa theologiae* Ia, 3, 1-7) (see God, arguments for the existence of §1). Given actualism, 'possibly God exists' entails 'God's nature exists', and given divine simplicity, God is identical to God's nature. So given divine simplicity plus actualism:

(P) 'possibly God exists' entails 'God exists'.

(P) is a key premise of many ontological arguments for God's existence (see God, arguments for the existence of §§2-3). Good prospects (as they thought) for proving a simple being's existence likely gave Western religious thinkers further reason to embrace divine simplicity; by so doing, they believed, they could help give theism rational warrant.

3 Challenges to simplicity

Islam rejected divine simplicity in reacting against Aristotelianism; some Jewish philosophers (for example, Crescas) also rejected it. Within Christianity, divine simplicity's first real challenge was Duns Scotus' positing of a 'formal distinction' between divine attributes, faculties, and so on. Other challenges arose in Leibniz, Spinoza, Hegel, and nineteenth- and twentieth-century theology. Some reject divine simplicity because it implies other doctrines they question: it has been a casualty in revolts against divine immutability, impassibility, timelessness and distinctness from the universe. Thus one must in part pursue the case against divine simplicity in cases against these other doctrines.

Recent discussion in analytic philosophy has focused on divine simplicity itself. Some philosophers (for example,

Plantinga 1980) reject divine simplicity on the grounds that since it entails that God is identical to God's attributes, it entails that God *is* an attribute, an abstract object. But this need not follow: it may follow only that the one thing which is both God and an attribute has divine attributes and also some attributes typical of abstract objects, a claim theists endorse anyway.

Some philosophers (for example, Morris 1985) raise a problem of modal uniformity: if all God's attributes are identical, God has all attributes necessarily or all contingently. But God seems to have both necessary and contingent attributes: he is necessarily omnipotent, but only contingently the creator of Adam (since he need not have created Adam). Perhaps divine simplicity need not face this problem; as we have seen, the medieval philosophers may not have meant to identify all of God's attributes. But if they did, advocates of divine simplicity can perhaps counter by distinguishing attributes from predicates. Even if God has just one intrinsic attribute, that attribute, coupled with God's *relations* to other things, may let God satisfy some predicates necessarily and others contingently. God necessarily satisfies '…is omniscient', because every possible world contains only truths God would know if he created that world. God contingently satisfies '…is Creator of Adam' because only in some possible worlds is there an Adam. The difference in modality rests on entities outside God and their relations to God.

If God cannot have logically contingent intrinsic attributes but has created Adam, it may follow that the inner intrinsic state of God by which he knows and has created Adam might instead have not known and not created Adam, without itself being intrinsically any different. Is this conceivable? Consider a complex entity consisting of God plus the world he makes. If this complex entity contains the world, it is different if the world is different, but it does not follow that God himself is different. On an 'externalist' view of divine mental content, this complex entity could be God's mind. On the more usual picture, God's mind is wholly internal to God, and a difference in God's mind is prior to and explains a difference in the world. On an 'externalist' picture of God's mind, a difference in God's mind *includes* a difference in the world, but may not include a difference in God himself. If divine intrinsic simplicity does preclude God's having logically contingent intrinsic attributes, perhaps its price is that we are always in God's mind.

See also: Natural theology; Negative theology

BRIAN LEFTOW

References and further reading

Aquinas, T. (1259-65) *Summa contra gentiles* I, trans. A. Pegis, Notre Dame, IN: University of Notre Dame Press, 1975.(Classic statement of classical theism built around divine simplicity.)

Aquinas, T. (1266-73) *Summa theologiae*, trans. Dominican Fathers, New York: Benziger Brothers, 1948.(Classic statement of divine-simplicity-centred classical theism.)

Aristotle (mid 4th century BC) *Physics*, trans. R. Hardie and R. Gaye, in R. McKeon (ed.) *The Basic Works of Aristotle*, New York: Random House, 1941. (Gives his argument for God's existence, and ties it to divine simplicity.)

Athenagoras (177) *An Embassy for the Christians*, trans. B. Pratten, in A. Roberts and J. Donaldson (eds) *Ante-Nicene Fathers*, vol. 2, Peabody, MA: Hendrickson, 1994.(Second-century Christian adopts divine simplicity; see page 132.)

Clement (194-201) *Stromata*, trans. A. Coxe, in A. Roberts and J. Donaldson (eds) *Ante-Nicene Fathers*, vol. 2, Peabody, MA: Hendrickson, 1994.(Rambles through many topics, including divine simplicity; see page 464.)

Hughes, C. (1989) *On a Complex Theory of a Simple God*, Ithaca, NY: Cornell University Press.(Excellent extended critique of divine simplicity. Rigorous but not technical.)

Irenaeus (182-8) *Against Heresies*, trans. A. Coxe, in A. Roberts and J. Donaldson (eds) *Ante-Nicene Fathers*, vol. 1, Peabody, MA: Hendrickson, 1994.(Early Christian adopts the doctrine of divine simplicity; see page 374.)

Knuuttila, S. (1993) *Modalities in Medieval Philosophy*, New York: Routledge. (Gold mine of information on evolving concepts of necessity. Occasionally technical.)

Kretzmann, N. (1983) 'Abraham, Isaac and Euthyphro', in D. Stump (ed.) *Hamartia*, Toronto, Ont.: Mellen, 27-50.(Divine simplicity and meta-ethics. Very clear.)

Maimonides, M. (*c*.1190) *The Guide for the Perplexed*, trans. M. Friedlander, New York: Dover, 1956.(Jewish version of classical theism; extended treatment of divine simplicity and its implications.)

Mann, W. (1982) 'Divine Simplicity', *Religious Studies* 18 (4): 451-71.(Reply to Plantinga, exposition of divine

simplicity. Criticized by Morris (1985).)

Mann, W. (1983) 'Simplicity and Immutability in God', *International Philosophical Quarterly* 23 (3): 267-76. (Rebuts a number of objections to divine simplicity.)

Mann, W. (1988) 'God's Freedom, Human Freedom, and God's Responsibility for Sin' in T. Morris (ed.) *Divine and Human Action*, Ithaca, NY: Cornell University Press, 182-210.(Clear discussion of surprising connections between these issues and divine simplicity.)

Mann, W. (1989) 'Modality, Morality and God', *Noûs* 23 (1): 83-99.(Elegant discussion of divine simplicity's implications for meta-ethics and God's relation to necessary truth.)

Miller, B. (1996) *A Most Unlikely God*, Notre Dame, IN: Notre Dame University Press.(Defends divine simplicity, skilful but idiosyncratic.)

Morris, T. (1985) 'Of God and Mann', *Religious Studies* 21 (3): 299-318. (Forceful, clear reply to Mann (1982); poses modal uniformity problem.)

Philo (first half of first century AD) *Legum allegoria*, in F. Colson and G. Whitaker (trans) *Philo*, vol. 1, Cambridge, MA: Loeb Classical Library, Harvard University Press.(Jewish Platonist adopts divine simplicity; see page 225.)

Plantinga, A. (1980) *Does God Have a Nature?*, Milwaukee, WI: Marquette University Press.(Criticizes the doctrine of divine simplicity during a broader treatment of God's relation to necessary truth.)

Stump, E. and Kretzmann, N. (1985) 'Absolute Simplicity', *Faith and Philosophy* 2 (4): 353-82.(Careful reading of Aquinas' version of divine simplicity as allowing contingent intrinsic attributes. Relates divine simplicity to divine freedom and issues in philosophical theology.)

Wolfson, H. (1977) *The Philosophy of the Kalam*, Cambridge, MA: Harvard University Press.(Largely about the medieval Islamic debate over divine simplicity.)

Wolfson, H. (1979) *Repercussions of the Kalam in Jewish Philosophy*, Cambridge, MA: Harvard University Press. (Largely about the medieval Jewish debate over divine simplicity.)

Wolterstorff, N. (1991) 'Divine Simplicity, in J. Tomberlin (ed.) *Philosophical Perspectives V: Philosophy of Religion*, Atascadero, CA: Ridgeview, 531-52. (Accessible account of divine simplicity and the medieval writers' treatment of attributes as parts of their subjects.)

Simplicius (*fl.* first half 6th century AD)

Simplicius of Cilicia, a Greek Neoplatonic philosopher and polymath, lived in the eastern part of the Roman Empire. He is the author of the most learned commentaries on Aristotle produced in antiquity, works which rest upon the accumulated accomplishments of ancient Greek philosophy and science. In them he gives numerous illuminating references and explanations that not only lead to a fuller understanding of Aristotle, but also allow one to reconstruct the history of the interpretation and criticism of Aristotelian doctrines in antiquity. The main principle that guides Simplicius' exegesis is the conviction that most Greek philosophers, including some Presocratics, can be brought into agreement with Neoplatonism. Simplicius adduces copious quotations to prove his point, thereby supplying us with substantial fragments from lost works of thinkers like Parmenides, Empedocles, Anaxagoras, Eudemus and the Stoics. A devout pagan, Simplicius sought to defend traditional Greek religion and philosophy against the oppressive dominance of Christianity. His commentaries have influenced the reception and interpretation of Aristotle's philosophy ever since.

1 Life and writings

The circumstances of Simplicius' life are largely unknown. He was probably slightly older than his great rival, the Christian commentator and theologian John Philoponus. Simplicius studied first at Alexandria under Ammonius, son of Hermeas, the influential philosophical head of the school there, and later under Damascius, the last head of the Neoplatonic school at Athens. The course of Simplicius' life and career underwent dramatic change in 529, when, in the wake of the empire's anti-pagan policies, the Emperor Justinian I closed the Athenian school of philosophy. During the crisis, seven Athenian philosophers, among them Simplicius, Damascius and Priscian, emigrated to Persia where Chosroës I enjoyed a reputation as a philosopher-king. Their sojourn was disappointing and brief. The philosophers left the Persian court when, in a peace treaty between Persia and Byzantium in 532, Chosroës' diplomacy secured them the liberty to live on Roman territory without fear of persecution, although they were presumably barred from public teaching. Arabic sources, in conjunction with inconclusive evidence in Greek texts, suggest however that some of them were active in the Mesopotamian town of Carrhae (Harran), a pagan stronghold near the Persian border. Whether Simplicius belonged to this group, or returned to Athens, or indeed went elsewhere, is a matter of considerable controversy (see also Aristotle Commentators).

Simplicius wrote his major works after the Persian adventure. With the exception perhaps of the relatively late commentary on Aristotle's *Categories*, these are pieces of scholarship to be read, rather than lectures delivered in class. There seem to be no further treatises extant in manuscript form. Works on Iamblichus and on Hermogenes' rhetoric are lost, and perhaps even a commentary on Aristotle's *Metaphysics*, which is referred to in the commentary on Aristotle's *On the Soul*. This latter work bears Simplicius' name, but is of inferior quality and has plausibly been attributed to Simplicius' contemporary Priscian. If this is correct, evidence for a *Metaphysics* commentary by Simplicius becomes slim.

2 Hermeneutics and ethics

Simplicius makes his hermeneutic position explicit in the commentary on the *Categories*, a set book for philosophical beginners. The scholars of the sixth century (Ammonius, Philoponus, Olympiodorus, Elias and Simplicius) all discuss, in the introductions to their lectures on the *Categories*, ten questions regarding Aristotle's philosophical corpus as a whole, its structure and its end. One of these questions concerns the qualifications and attitude of the proper exegete of Aristotle. The stock answer is that the interpreter, in addition to being knowledgeable, must strive to discern the truth and, if necessary, prefer it to Aristotle. Philoponus seems to have been the only one to put this maxim to practical use. Characteristically, Simplicius glosses over this requirement and emphasizes instead that rather than voice his own philosophical convictions the exegete should seek out the agreement between Plato and Aristotle. Likewise, the student should refrain from quibbling with Aristotle. Simplicius is convinced that the 'truth' can be found in the harmony of Plato, Aristotle and other inspired thinkers. Thus, he is prepared to accept doctrines on the basis of authority, especially if they can be shown to agree with the fundamental ideas of Neoplatonism (see Neoplatonism).

Simplicius' earliest surviving work, *On Epictetus' Manual*, is a commentary on the *Manual (Enchiridion)* of the Stoic philosopher Epictetus. In general, Neoplatonists scorned the materialist outlook of the Stoics (see Stoicism

§3). Stoic ethics, however, conforms to the lowest level of the Neoplatonic system of moral instruction, which requires the practice of, in ascending order, political, purificatory and contemplative virtues. In the Preface, Simplicius explains that Epictetus' aphorisms and maxims are beneficial to those who want their bodies and desires to be ruled by rationality. A training in the political virtues resulting in a moderation of one's emotions initiates the ascent of the embodied soul through philosophy. The aim is the pure contemplation of intelligible Forms. This separation of the soul from the concerns of the body is also the road to salvation, and Simplicius conceives of his works simultaneously as intellectual and spiritual exercises. In a political atmosphere that increasingly suppressed and eradicated the remnants of paganism, his commentaries also protested against what he perceived as the blasphemy of the new religion.

3 Physics

The text that exemplifies this aspect most clearly is the commentary on Aristotle's *On the Heavens*. In one respect, it is a scholarly exegesis of Aristotle's cosmological treatise, enriched with invaluable information on ancient astronomy. A prominent feature, however, is the polemic against the renegade John Philoponus, a Christian Neoplatonist trained in Alexandria. At the very time when the Athenian philosophers were struggling for survival, Philoponus launched an attack on the Aristotelian-Neoplatonic doctrine of the eternity of the world by rejecting, among other points, Aristotle's theory of the existence of a celestial fifth element, aether (*On the Heavens* I.2-4). In an unprecedented spirit of criticism, Philoponus called the cogency of Aristotle's reasoning into doubt and argued that the supra- and sub-lunary regions are in fact of the same corruptible nature. In his reply, Simplicius offers the last defence in antiquity of Aristotle's controversial theory of aether. However, in the course of his counter-attack Simplicius loses sight of the crucial point - Aristotle's postulate of a fifth elementary body that naturally displays circular motion. Simplicius interprets Aristotle as meaning something approximating Plato's theory (*Timaeus* 40a, 58c-d) that the heavens consist of a mixture of the purest grades of the four elements, in particular luminous fire (*On Aristotle's On the Heavens* 12.27-13.3; 85.7-15). Moreover, Simplicius holds that celestial motion is not caused by nature but rather by the agency of a celestial soul (78.17-80.13), a position Aristotle at one point explicitly rejected (*On the Heavens* II.1, but see also 285a29). Despite considerable reinterpretations such as these, the commentary turns into a grandiose reaffirmation of the divinity and eternity of the heavens and the universe as a whole. The sincere prayer to the creator-intellect with which Simplicius ends his work underscores the vital link, so characteristic of Neoplatonism, between philosophy and spirituality.

A less elevated work is the monumental commentary on Aristotle's *Physics*, which modern scholars have culled primarily for its invaluable reports and quotations from earlier philosophers, most importantly the Presocratics. Recent scholarship, however, has shown that here too Simplicius occasionally comes forward with original interpretations of fundamental concepts of Aristotle's natural philosophy. For example, ever since Plotinus (§4), many Neoplatonists had interpreted Aristotle's prime matter as incorporeal, formless and devoid of magnitude, an elusive entity capable of accepting the property of indefinite extension and thus becoming the substrate for all physical forms. In the sixth century, both Philoponus and Simplicius revised this idea and paved the way for a conception of matter as mass. Philoponus polemically rejected the notion of prime matter as incoherent and superfluous, positing three-dimensional extension as the lowest level of the physical world. Without entering into any kind of polemic, Simplicius chose to interpret the Aristotelian prime matter in a very similar way, calling it an extension and an indefinite diffusion (*Physics* commentary, 537.22-538.14).

This notion of prime matter is not easily distinguished from that of space. Like many thinkers before him, Simplicius finds fault with Aristotle's definition of a thing's place (*topos*) as the adjacent boundary of the containing body (*Physics* 212a2-6). However, he also rejects the notion of space (*chōra*) as empty extension, and in the so-called 'corollary on place' (*On Aristotle's Physics* 601-45) he conflates, much like Aristotle, the idea of 'extension' with that of 'location'. On the one hand, Simplicius speaks of place *qua* extended space as a substance that is both animate and immobile (623.19-21). On the other hand, he defines place as incorporeal but perceptible in bodies, an ordering of the position of each thing relative to others (642.14-21). Significantly, it is place that facilitates order in the physical universe, since it provides the physical images of the intelligible paradigms with structure and allows them to be similar to the paradigms (626.28-32).

A further example of Simplicius stepping beyond the bounds of mere exegesis is his discussion of time. Aristotle began his analysis in *Physics* (4.10) by presenting two arguments suggesting, paradoxically, that time is unreal.

The premises (1) that the past exists no longer, (2) that the future exists not yet, and (3) that the present, being an instant, cannot be considered a 'part' of time (for as the temporal analogue of a point, it cannot serve as a measure of any length of time) seem to entail the unreality of time. Moreover, the view of the present as an instant renders mysterious how the present ever ceases to exist, for this can neither take place *at* the present instant nor *at* some later instant. In the so-called 'corollary on time' (*On Aristotle's Physics* 773-800) Simplicius points out that Aristotle's paradoxes arise only if one regards as existent that which is in fact always and only in a process of becoming. He reiterates the view of Damascius that time does not exist in actuality, as a sensible object might exist, but that 'it has its being in becoming'. His solution that time is never actual, but always potential, may seem to yield too much to the claim that time is unreal. However, on Simplicius' Neoplatonic view, passing time is only an image of 'primary time', which is a stable intermediate between the fleeting time of the physical realm and unparticipated eternity (784.17-785.10).

4 Logic and mathematics

On Aristotle's Categories is probably the latest authentic work of Simplicius we possess. One question that particularly interested Neoplatonists was the subject matter, or *skopos*, of Aristotle's treatise: is it about words, things or concepts? (See Porphyry §2; Categories §1; Aristotle §7). Simplicius explains that the treatise concerns itself with human language and, more precisely, with its simple, primary and generic utterances (*phōnai*) in so far as they signify concrete objects (*pragmata*). It instructs us about such objects, but also about the concepts (*noēmata*) through which the utterances gain signification (*On Aristotle's Categories* 13.11-15). In the course of this study of language the soul is once again being prepared in an ideal way for its ascent to the noetic world.

Like most Neoplatonists, Simplicius had a keen interest in mathematics. *On Aristotle's Physics* contains an extensive discussion of geometry including an extract from Eudemus' *History of Geometry* concerning the quadrature of lunes by Hippocrates of Chios. These passages bear witness to the oldest Greek mathematical work and present important material for a reconstruction of the history of pre-Euclidean mathematics. Fragments of a commentary by Simplicius on the definitions, postulates and axioms of book I of Euclid's *Elements* are extant in Arabic.

5 Influence

Through his work, Simplicius shaped the reception and interpretation of Aristotle for centuries. The bibliographies of Ibn an-Nadim and Ibn al-Qifti show that his commentaries were at least partially known in the Arab world (see Aristotelianism in Islamic philosophy). The first Latin translations by Robert Grosseteste and William of Moerbeke date from the thirteenth century and had an immediate impact on Thomas Aquinas and his circle. Later, Simplicius is one of the pillars of Renaissance Aristotelianism (Galileo, in the *Dialogue*, names his Aristotelian straw man 'Simplicio'), and even today his work is indispensable for any serious student of Aristotle or the history of Greek philosophy.

CHRISTIAN WILDBERG

List of works

Simplicius (*c*.533) *On Epictetus's Manual*, ed. I. Hadot, *Simplicius - Commentaire sur le 'Manuel' d'Épictète. Introduction et édition critique du texte grec*, Leiden: Brill, 1996.(The volume includes a long general introduction to Simplicius as well as a good bibliography.)

Simplicius (*c*.536) *On Aristotle's On the Heavens*, ed. I.L. Heiberg, *Commentaria in Aristotelem Graeca VII*, Berlin: Reimer, 1894.(A commentary on Aristotle's cosmological treatise and vigorous defence of the pagan notion of the divinity of the heavens.)

Simplicius (after 538) *On Aristotle's Physics*, ed. H. Diels, *Commentaria in Aristotelem Graeca IX- X*, Berlin: Reimer, 1882-5; parts available in translation, as follows: B. Fleet, *Simplicius on Aristotle's Physics 2*, London: Duckworth, 1997; J.O. Urmson and L. Siorvanes, *Simplicius, Corollaries on Place and Time*, London: Duckworth, 1992; J.O. Urmson and P. Lautner, *Simplicius, On Aristotle on the Void*, London: Duckworth, 1994; D. Konstan, *On Aristotle Physics 6: Simplicius*, London: Duckworth, 1989; C. Hagen, *On Aristotle's Physics 7: Simplicius*, London: Duckworth, 1994; C. Wildberg, 'Simplicius, Against Philoponus on the Eternity of the World', in D. Furley and C. Wildberg, *Place, Void, and Eternity*, London: Duckworth, 1991, 95-134.(Diels gives Greek text only; the others are translations with notes.)

Simplicius (after 538) *On Aristotle's Categories*, ed. C. Kalbfleisch, *Commentaria in Aristotelem Graeca VIII*,

Berlin: Reimer, 1907; trans. and ed. I. Hadot, *Simplicius: Commentaire sur les Catégories*, Leiden: Brill, fasc. I and III, 1990, fasc. II, 1996; trans. *Simplicius, Commentaire sur le Catégories d'Aristote, traduction de Guillaume de Moerbeke*, ed. A. Pattin and W. Stuyven, *Corpus Latinum Commentariorum in Aristotelem Graecorum III*, Paris: Éditions Béatrice-Nauwelaerts, 1971.(The Hadot text contains French text with comments by different authors; the Patton and Stuyven text is the standard translation of the medieval Latin translation.)

Simplicius (*c*.540) *On Aristotle's On the Soul*, ed. M. Hayduck, *Commentaria in Aristotelem Graeca XI*, Berlin: Reimer, 1882; trans. in part, J.O. Urmson and P. Lautner, *Simplicius, On Aristotle's On the Soul 1.1-2.4*, London: Duckworth, 1995.(The authorship of this work is disputed, the date uncertain.)

References and further reading

Cameron, A. (1969) 'The Last Days of the Academy at Athens', *Proceedings of the Cambridge Philological Society* 195, new series, 15: 7-29.(Argues that the school continued to exist in Athens after 529 and that Simplicius returned there; for the opposing view see Hadot's *Commentaire sur le 'Manuel' d'Épictète*.)

Gätje, H. (1982) 'Simplikios in der arabischen Überlieferung' (Simplicius in the Arabic Tradition), *Der Islam* 59: 6-31.(A discussion of bibliographical references to Simplicius in Arabic texts.)

Hadot, I. (ed.) (1987) *Simplicius - Sa vie, son œuvre, sa survie (Simplicius - His Life, Work and Influence)*, Berlin and New York: de Gruyter.(A collection of articles on Simplicius' biography, his doctrine, the textual transmission of his work and its influence in the Renaissance.)

Hadot, I. (1990) *'The Life and Work of Simplicius in Greek and Arabic Sources'*, in R.R.K. Sorabji (ed.) *Aristotle Transformed*, London: Duckworth, 275-303.(This article owes much to the pioneering work by Michel Tardieu; the whole volume is an excellent guide to philosophy in late antiquity, offering an extensive bibliography.)

Hoffmann, P. (1987) 'Simplicius' Polemics. Some Aspects of Simplicius' Polemical Writings against John Philoponus: From Invective to a Reaffirmation of the Transcendency of the Heavens', in R.R.K. Sorabji (ed.) *Philoponus and the Rejection of Aristotelian Science*, London: Duckworth, 57-83.(A discussion of Simplicius' rhetoric of invective and piety in the commentary on Aristotle's *On the Heavens*.)

Knorr, W.R. (1986) *The Ancient Tradition of Geometric Problems*, Boston, MA: Birkhäuser.(Chapter 2 offers a discussion of the mathematical passages in the *Physics* commentary.)

Sonderegger, E. (1982) 'Simplikios: Über die Zeit. Ein Kommentar zum Corollarium de tempore' (Simplicius: On Time. A Commentary on the 'Corollary on Time'), *Hypomnemata* 70, Göttingen: Vandenhoeck. (A detailed analysis of Simplicius' 'Corollary on Time' in the *Physics* commentary.)

Sorabji, R.R.K. (1988) *Matter, Space and Motion*, London: Duckworth, ch. 1.(A discussion of Simplicius' interpretation of Aristotle's prime matter as extension.)

Tardieu, M. (1986) 'Sabiens coraniques et "Sabiens" de Harran' (The Sabieans of the Koran and the Sabians of Harran), *Journal Asiatique* 274: 1-44.(A masterly article establishing the thesis that the Athenian Neoplatonists continued their work in Carrhae/Harran.)

Sin

The most archaic conception of human fault may be the notion of defilement or pollution, that is, a stain or blemish which somehow infects a person from without. All the major religious traditions offer accounts of human faults and prescriptions for dealing with them. However, it is only when fault is conceived within the context of a relationship to a personal deity that it makes sense to speak of it as an offence against the divine will. The concept of sin is the concept of a human fault that offends a good God and brings with it human guilt. Its natural home is in the major theistic religions of Judaism, Christianity and Islam.

These religious traditions share the idea that actual or personal sins are individual actions contrary to the will of God. In the Hebrew Bible, sin is understood within the context of the covenantal relation between Yahweh and his chosen people. To be in covenant with Yahweh is to exist in holiness, and so sin is a deviation from the norms of holiness. In the Christian New Testament, Jesus teaches that human wrongdoing offends the one whom he calls Father. The Qur'an portrays sin as opposition to Allah rooted in human pride.

According to Christian tradition, there is a distinction to be drawn between actual sin and original sin. The scriptural warrant for the doctrine of original sin is found in the Epistles of Paul, and the interpretation of Paul worked out by Augustine in the course of his controversy with the Pelagians has been enormously influential in Western Christianity. On the Augustinian view, which was developed by Anselm and other medieval thinkers with considerable philosophical sophistication, the Fall of Adam and Eve had catastrophic consequences for their descendants. All the progeny of the first humans, except for Jesus and his mother, inherit from them guilt for their first sin, and so all but two humans are born bearing a burden of guilt. The Augustinian doctrine of original sin is morally problematic just because it attributes innate guilt to humans. It was criticized by John Locke and Immanuel Kant.

1 Actual sin

Actual or personal sins are individual human actions that are offensive to God. Since God is conceived of in the major theistic religions as morally perfect, all wrongdoing will be offensive to God and hence will be sinful. Most theists hold that some things would be morally wrong even if there were no God; only theological voluntarists suppose that nothing would be morally wrong if God did not exist (see Religion and morality §1; Voluntarism). On the majority view, much of morality is independent of the existence and will of God; divine prohibitions do not make such things as murder, torture or rape wrong. Divine commands to refrain from doing such things serve to reinforce an independent morality. Murder becomes, so to speak, doubly wrong when forbidden by God. It is a wrong against the victim and it is a wrong against God, but it would still be wrong even if, on account of the nonexistence of God, it were not sinful. Actions of this kind are such that their moral wrongness is independent of their sinfulness.

However, not all actions are of this kind. It is commonly thought that we have a moral duty to express gratitude to our benefactors. If God created us and conserves us in existence and if the gift of life is on the whole good, then we have a duty to express gratitude to God. It would be morally wrong and hence sinful not to do so. But if there is no God, we do not owe such a debt of gratitude, and so it would be neither morally wrong nor sinful not to express gratitude to God. So some actions are such that both their moral wrongness and their sinfulness depend on the existence and actions of God. For such actions moral wrongness is not independent of sinfulness.

A third possibility worth considering is that there are distinctively religious duties which are not also moral duties. It seems to be morally indifferent whether one worships God on one day rather than another, or using one ritual rather than another. But God may have prescribed the form worship is supposed to take. If so, it would be offensive to God and hence sinful not to worship in the prescribed manner. Yet it is at least arguable that it would not be immoral to fail to worship in the prescribed manner even if, because of a duty of gratitude to God, it would be morally wrong not to express reverence to God in some manner. So perhaps there are also actions or omissions that are sinful even though they are not morally wrong.

There are thus at least two varieties of actual sin. There are actions or omissions that are morally wrong whether or not God exists but are also sinful if God exists. And there are actions or omissions that are neither morally wrong

nor sinful if God does not exist but are both morally wrong and sinful if God exists. In addition, there may be actions or omissions that are neither morally wrong nor sinful if God does not exist but are sinful yet not morally wrong if God exists.

A distinction can be drawn between objective and subjective actual sin. If people do what is objectively offensive to God, they sin objectively and acquire objective guilt. If people do what they believe to be offensive to God, they sin subjectively and acquire subjective guilt. If God exists, there will be cases in which people sin both objectively and subjectively, cases in which people sin objectively but not subjectively, and cases in which people sin subjectively but not objectively. The guilt of actual sin renders the sinner liable to punishment by God if appropriate conditions on responsibility are satisfied. A person who sins both objectively and subjectively and whose true beliefs about what is offensive to God are well warranted may deserve severe punishment. A person who sins objectively but not subjectively and whose lack of true beliefs about what is offensive to God are not the result of culpable ignorance may deserve little or no punishment. In general, the relations among actual sin, guilt before God, responsibility and desert of punishment in the theistic religions are similar to and no more problematic than the relations among wrongdoing, moral guilt, responsibility and desert of sanctions in common morality. This is not the case for Augustinian original sin.

2 Augustinian original sin

The roots of the doctrine of original sin lie in the story of the Fall of Adam and Eve. When they disobey a divine command and eat the forbidden fruit, God punishes them with toil, suffering and death. Being subject to such things is part of a legacy we inherit from them. But the story does not force us to conclude that they are also to be thought of as punishments in our case. Nor does it clearly suggest that we have also inherited from the first humans a burden of guilt. That suggestion derives from the Epistles of Paul.

Seeking to appropriate the contents of the Hebrew Bible for Christianity, Paul reads it as full of foreshadowings of things that only come to fruition in the life and death of Jesus Christ. In Romans, he powerfully contrasts Adam and Christ, beginning as follows: 'Therefore, just as through one man sin entered the world and with sin death, death thus coming to all men in as much as all sinned' (5: 12). He concludes: 'Just as through one man's disobedience all became sinners, so through one man's obedience all shall become just' (5: 19). The obedience (good action) of Christ is said to be that through which all shall become just (have positive moral status); this completely contrasts with the disobedience of Adam, through which all became sinners. The implication is that everyone acquired the negative moral status of being sinners through Adam's disobedience, which would be easily explained on the assumption that guilt is somehow transferred from Adam to his progeny in a way that parallels the transfer of justice from Christ to those who benefit from his atonement.

Augustine (§§6, 9) explicitly makes this assumption. He tells us that 'when the first couple were punished by the judgment of God, the whole human race, which was to become Adam's posterity through the first woman, was present in the first man' (*De civitate Dei*, book 13: 271). All humans were present in Adam because human nature was present in Adam's semen. So we are all born in the condition of being soiled by sin, doomed to death and justly condemned (such is our nature), and this condition is part of what we would today refer to as our genetic endowment because it propagates biologically from Adam to his posterity by means of semen. As *The New-England Primer* neatly summarizes this doctrine of innate sin and guilt: 'in Adam's fall, we sinned all'. When he tries to defend it against his Pelagian opponent Julian of Eclanum, who rejects it, Augustine frequently appeals to the authority of Pauline texts such as Romans 5: 12 and 5: 19 (see Pelagianism).

Many scholars doubt the legitimacy of the appeal to Romans 5: 12. As they see it, Augustine misinterpreted that verse because he read it in Latin rather than in the original Greek. Apparently he took it to say that through one man sin entered the world and with sin death, death thus coming to all men, *in whom* all sinned; he mistakenly supposed that the final clause referred back to the one man, Adam, which led him to conclude that Adam's fall brought about not only universal death but also universal sin. But even if this conclusion is not supported by Romans 5: 12, it does seem to be supported by Romans 5: 19, and so the Augustinian doctrine of original sin appears to have a basis in scriptural texts. It was accepted by most Western Christian thinkers for more than a millennium, and it is still accepted by many Christians. It bears a striking resemblance to the archaic account of pollution that infects a person from without.

The Augustinian doctrine of original sin was elaborated in the work of medieval theologians such as Anselm of Canterbury (§8). According to Anselm, human beings are metaphysical composites that include both a nature, which makes each - like others - human, and a principle of individuation, which makes each a particular person distinct from all others. Original sin is sin that one contracts with human nature at the very origin of one's existence as a person; it is innate and unavoidable. It consists of a will that lacks a proper upright orientation because it is not subject to the will of God. Anselm describes the process by which their descendants derive original sin from Adam and Eve in terms of a principle of causal transmission. It states that 'as what is personal passes over to the nature, so what is natural passes over to the person' (*De conceptu virginali et peccato originali*: 202). By the first half of this principle, the sin Adam and Eve committed when they disobeyed God caused human nature to become sinful; by its second half, sinful human nature in turn causes their descendants to be sinful from the very first moment they possess it. And Anselm draws a shocking conclusion from the Augustinian doctrine. It is that infants who die unbaptized, before having committed any personal sins of their own and so with only the stain of original sin on their souls, are condemned by God, who is morally perfect, to exclusion from the kingdom of heaven (see Limbo §1). Many of Anselm's successors, including Thomas Aquinas, Martin Luther, John Calvin (§4) and Jonathan Edwards, played other variations on this Augustinian theme.

3 Modern philosophical critiques

Important philosophers of the modern era have been critical of the Augustinian doctrine of original sin. Locke (§1) tried to undermine its scriptural support by reinterpreting the Pauline texts cited by Augustine, and Kant (§11) proposed to substitute for it his doctrine of radical evil.

Locke's procedure is to quote a verse, then to offer his own paraphrase of it, and finally to argue in support of the paraphrase in an appended note. The paraphrase of Romans 5: 12 goes as follows:

> Wherefore to give you a state of the whole matter from the beginning. You must know, that as by the act of one man Adam the father of us all, sin enterd into the world, and death, which was the punishment annexed to the offence of eating the forbidden fruit enterd by that sin for that Adams posterity thereby became mortal.
>
> (1707: 523)

Having, in effect, substituted 'became mortal' for 'sinned' at the end of the verse, Locke has blocked both the inference that sin is inherited as well as death and the inference that death is a punishment for sin in Adam's posterity. In the note, he tries to justify this substitution by claiming that Paul is using metonymy, that is, substituting the cause for the effect, sin in Adam being the cause of his mortality and, through him, the cause of the mortality of his posterity.

Similarly, Locke's paraphrase of Romans 5: 19 is this:

> For as by one mans disobedience many were brought into a state of mortality which is the state of sinners soe by the obedience of one shall many be made righteous. i e be restord to life again as if they were not sinners.
>
> (1707: 527)

The justificatory note is terse: '*Sinners*. Here St Paul uses the same metonymie as above ver. 12 putting *sinners* for *mortal* whereby the Antithesis to righteous is the more lively' (1707: 527). We could always paraphrase away talk of Adam's disobedience making his progeny sinners in favour of talk of Adam's disobedience making his progeny mortal, thereby undercutting the scriptural basis for Augustinian original sin, if Locke's metonymy gambit were successful across the board. However, not many biblical scholars are persuaded that it is successful.

Kant flatly rejects the Augustinian doctrine of inherited sin and guilt:

> However the origin of moral evil in man is constituted, surely of all the explanations of the spread and propagation of this evil through all members and generations of our race, the most inept is that which describes it as descending to us as an *inheritance* from our first parents.
>
> (1793: 35)

Yet Kant holds that there is, in a sense, radical evil in human nature because there is in all humans, as far as we can tell, a morally evil propensity to evil. It is this propensity that is the proxy for original sin in his philosophy of

religion; he even calls it *peccatum originarium*.

A propensity, Kant tells us, is a predisposition to crave a delight which, when once experienced, arouses in the subject an inclination to it. Those who have a propensity for chocolate, for example, do not desire chocolate before they first taste it, but once they have sampled it they develop a craving for it. Kant describes propensities of this sort as physical because they belong to their possessors considered as beings determined by natural laws. Whatever is determined by natural laws is morally indifferent, and so physical propensities are morally indifferent. If all propensities were physical, then even if there were a propensity for evil in humans it would not itself be morally evil. So there must be nonphysical propensities if a propensity for evil is to make its human possessors morally evil. For Kant, nothing is morally evil but free action and what derives from it, and so a propensity for evil that is itself morally evil has to be a product of the exercise of freedom. Though the propensity for evil can, according to Kant, be represented as innate, it should not be represented merely as innate; it should also be represented as brought by humans upon themselves. It can be represented as innate in its possessors because, as the underlying ground of all morally evil actions, it is to be thought of as present in them antecedent to all such actions and so represented temporally as present in them as far back as birth. It should be represented as brought by humans upon themselves because, being morally evil, it has to be a product of freedom for which its possessors are accountable. And it can be represented as brought by humans upon themselves because it can be thought of as the product of an atemporal act of noumenal freedom.

On Kant’s view, there is radical evil in human nature only in the sense that, as far as we can tell on the basis of the available evidence, all humans have brought upon themselves a morally evil propensity for evil. This propensity is not transmitted to us causally from our remote ancestors; it is not part of the genetic endowment that comes to us through sexual procreation. When Kant endorses the formula ‘in Adam all have sinned’, he means only to say that the story of the sin of Adam and Eve represents in a symbolic fashion something that actually occurs in every human life. Whether the Kantian account of radical evil is an improvement on the Augustinian doctrine of original sin depends heavily on the plausibility of Kant’s assumption that there are atemporal acts of noumenal freedom. Many philosophers do not consider it even remotely plausible.

4 Contemporary philosophical critiques

Original sin has not been a major issue in the philosophy of religion of the twentieth century. However, some Christian philosophers have added to the criticism of the Augustinian doctrine of innate guilt.

Richard Swinburne (1989) acknowledges that humans have an innate proneness to sin. It stems from the strong selfish desires that are part of our evolutionary heritage and it is genetically transmitted. Although he describes this proneness to wrongdoing as original sinfulness, he insists that the bad desires in which it consists incline without necessitating and so do not inevitably issue in actual wrongdoing. And he emphatically rejects the doctrine of original guilt according to which all of Adam’s descendants are guilty of Adam’s original sin. He argues that one cannot be guilty in a literal sense for the sins of another unless one had some obligation to deter that person and failed to do so. Since no one alive today could have had an obligation to deter Adam and Eve, we cannot be guilty for their first sins. Swinburne also cites Scripture to support his view: ‘The son shall not bear the iniquity of the father, neither shall the father bear the iniquity of the son; the righteousness of the righteous shall be upon him, and the wickedness of the wicked shall be upon him’ (Ezekiel 18: 20). As he sees it, the guilt we acquire when we fail in our obligations to deter others from wickedness is the only literal exception to this sound prophetic claim.

The doctrine of original guilt was taught by the federal theology of the Reformation. According to federalism, Adam was, by covenant with God, the federal head or representative of the whole human race. All of Adam’s posterity stood their probation in him, and so guilt for his sin is imputed to them by God in consequence of his having fallen while acting as their covenant head. Federal theologians do not deny that Adam’s progeny were seminally present in him or that a proneness to sin is biologically transmitted from Adam to his posterity. However, they insist that guilt for Adam’s first sin extends to his descendants not by way of biological transmission but by way of divine imputation. The Puritan theologian and philosopher Jonathan Edwards (1758) defended the account of original guilt offered by the federal theology. In a critical examination of his views, William Wainwright (1988) explores various models of the moral or legal relations that might be involved in Adam being the covenant head or representative of his posterity. Some of these models permit the transfer of liability from one person to another, but none of them allows for the transfer of guilt. According to Wainwright,

the reason for this is that one must have committed an act to be guilty of it and one cannot commit another's act. His conclusion is that 'even though liability can be transferred from one person to another, guilt cannot. Adam's posterity cannot be guilty of Adam's fault unless Adam's act is somehow *literally* their own' (1988: 47). But, of course, none of the acts of any of one's ancestors is literally one's own.

See also: Hell; Purgatory; Salvation

PHILIP L. QUINN

References and further reading

Adams, M.M. (1991) 'Sin as Uncleanness', in J.E. Tomberlin (ed.) *Philosophical Perspectives 5: Philosophy of Religion*, Atascadero, CA: Ridgeview Publishing.(Treats sin as ultimately an impropriety in the relation of created persons to God rather than a matter of moral wrongdoing.)

Anselm of Canterbury (1100) *De conceptu virginali et peccato originali (The Virgin Conception and Original Sin)*, trans. J.M. Colleran, Albany, NY: Magi Books, 1969. (Presents the elaboration of the Augustinian doctrine of original sin discussed in §2.)

Augustine of Hippo (426) *De civitate Dei (City of God)*, trans. G.G. Walsh, D.B. Zema, G. Monahan and D.J. Honan, Garden City, NY: Doubleday, 1958.(Book 13 contains the account of original sin discussed in §2.)

Edwards, J. (1758) *Original Sin*, ed. C.A. Holbrook, New Haven, CT and London: Yale University Press, 1970. (Contains the account of original sin referred to in §4.)

Kant, I. (1793) *Religion innerhalb der Grenzen der bloßen Vernunft (Religion within the Limits of Reason Alone)*, trans. T.M. Greene and H.H. Hudson, New York: Harper & Brothers, 1960.(Book 1 presents the account of radical evil discussed in §3.)

Locke, J. (1707) *A Paraphrase and Notes on the Epistles of St Paul*, ed. A.W. Wainwright, Oxford: Clarendon Press, 1987.(Volume 2 contains the Lockean interpretation of Romans 5: 12 and 5: 19 discussed in §3.)

Pagels, E. (1988) *Adam, Eve, and the Serpent*, New York: Random House.(A readable historical and critical study of Augustine's doctrine of original sin.)

Quinn, P.L. (1992) 'On Demythologizing Evil', in F. Reynolds and D. Tracy (eds) *Discourse and Practice*, Albany, NY: State University of New York Press. (Expansion of the material of §§2-3 of this entry.)

Ricoeur, P. (1967) *The Symbolism of Evil*, New York: Harper & Row.(Part 1 is an important study of defilement, sin and guilt from a phenomenological perspective.)

Swinburne, R. (1989) *Responsibility and Atonement*, Oxford: Clarendon Press. (Chapter 9 contains the criticism of the doctrine of original guilt discussed in §4.)

Wainwright, W.J. (1988) 'Original Sin', in T.V. Morris (ed.) *Philosophy and the Christian Faith*, Notre Dame, IN: University of Notre Dame Press.(Presents the criticism of Edwards on original sin discussed in §4.)

Watts, I. (1790) *The New-England Primer*, Newburyport, MA: John Mycall.(There are many editions of this work, which is cited in §2.)

Williams, N.P. (1927) *The Ideas of the Fall and of Original Sin*, London: Longmans, Green & Co.(A classic historical and critical study of the doctrine of original sin.)

Sirhak

Sirhak refers to the reformist scholarship and thought in Korea during the latter half of the Chosôn (Yi) Dynasty (1392-1910). The term was coined in the twentieth century to refer to the writings of individual scholars from the eighteenth to the mid-nineteenth centuries who were critical of the existing political, social and economic conditions.

As Korea recovered slowly from the devastations suffered from the two major foreign invasions by the Japanese (1592-8) and the Manchus (1636), the ruling Chosôn dynasty was faced with many serious political, social and economic problems, due largely to the political infighting and ineptitude of the ruling circle. At the same time, the traditional emphasis on strict adherence to neo-Confucian orthodoxy created a climate of intellectual rigidity, as the overriding philosophical concerns centred mostly on abstract and metaphysical issues such as debates on the primacy of the principle (*i*; in Chinese, *li*) and the material force (*ki*; in Chinese, *qi*), and on ritual formalities (see Confucian philosophy, Korean). It was in reaction to this that *sirhak* (practical learning) scholarship arose. Turning away from the conventional emphasis on abstract issues, the *sirhak* scholars dealt with more practical questions ranging from the classics and political economy to science and agriculture. In general, their encyclopedic writings emphasized four broad areas: the role of the government in promoting and protecting the welfare of the people, the practical use of knowledge, empirical evidence for knowledge and consciousness of Korea's identity.

Mindful of the well-being of the people, the *sirhak* scholars criticized the government and its policies for failing to protect the people adequately and proposed wide-ranging reforms of various institutions and practices. Seeing inequity in land holdings as the primary source of social ills, many proposed land reforms. Yu Hyôngwôn (1622-73) and Yi Ik (1681-1763) made proposals that would assure that farmers had their land to cultivate, while Chông Yagyong (1762-1836) called for a 'village land system' whereby villagers would jointly own the land and work together, sharing the harvest in proportion to the amount of labor each put in. Along with Yu Suwôn (1694-1755) and other *sirhak* scholars, they severely criticized the rigid social structure that granted privileges to the elite *yangban* social class based on birth. The abusive practices of the civil service examination system were also blamed for contributing to the social ills of the time. Through his satirical novellas, which resemble those of Voltaire, Pak Chiwôn (1737-1805) condemned the hypocritical life led by the *yangban*. Envisaging a more egalitarian society, the *sirhak* scholars wanted to do away with privileges based on birth and to make government more responsive to the needs of the people (see Equality).

The *sirhak* scholars believed in the practical application of knowledge in daily life. Agriculture was of special interest to many *sirhak* scholars, and Pak Sedang (1629-1703), Hong Mansôn (1643-1715), Sô Hosu (1736-99) and Sô Yugu (1764-1845) wrote extensively on soil conditions, fertilizer, irrigation, horticulture, sericulture, livestock, forestry and similar subjects in order to improve the condition of the peasants. Unlike the conventional Confucian scholar-officials, many *sirhak* scholars recognized the importance of trade and commerce. For Pak Chega (1750-1805), prudent consumption was necessary to create demand for production; he wanted the government to encourage domestic as well as international trade. The introduction of Western science and technology also stimulated the thinking of many *sirhak* scholars. Kim Sôngmun (1658-1735) and Hong Taeyong (1731-83) wrote extensively on mathematics and explained terrestrial movement within a solar system. Based on knowledge acquired from China, Chông Yag-yong devised cranes that were used to construct the new city of Suwôn and developed inoculations for measles and smallpox.

There was also a growing interest in historical and geographical studies of Korea. An Chôngbok (1712-91) rewrote Korean history, emphasizing the legitimacy of the Korean rulers starting from Tangun. In his monumental study of Korean history, Yi Kûngik (1736-1806) tried to present an objective history unhampered by factional biases. Yu Tûkkong (1748-1807) was the first historian to treat the kingdom of Parhae (Po-hai in Chinese) (699-926) as an integral part of Korean history. In historical geography, Sin Kyôngjun (1712-81), Han Chinsô (1777-?) and Chông Yagyong traced Korea's boundaries at various historical periods, while Yi Chunghwan (1690-1752) presented an authoritative study on the cultural and economic characteristics of various regions of Korea. Kim Chôngho (*fl.* 1834) produced several maps of Korea that are remarkably detailed and accurate even by modern standards. These studies enhanced consciousness of Korea's own identity.

Many *sirhak* scholars emphasized that knowledge is founded on empirical evidence. In his studies of the classics, Chông Yagyong wanted to go back to the texts of Han times or earlier for verification, in the belief that the classical canons are not necessarily the final words we must accept in faith, but merely the records of the ancient sages in search of betterment. Perhaps the most articulate proponent of empiricism in Korea was Ch'oe Hangi (1803-79). Upholding the primacy of *ki* (material force) over *i* (principle), Ch'oe emphasized that knowledge is acquired through the perceptions of our senses and that even the Mencian notion of humanity, righteousness, propriety and wisdom is derived from our experience, not from innate knowledge (see Mencius). Ch'oe Hangi is often regarded as an intermediary between tradition and modernity in Korea.

As the proposals of the *sirhak* scholars contained certain elements of the modernistic ideas, many Korean scholars tried to characterize them as the forerunners of modern Korea. However, the *sirhak* scholars envisioned their reform ideas within the bounds of the Confucian worldview, and regarded themselves as the heirs of the Confucian tradition.

See also: Chông Yagyong; Confucian philosophy, Korean

YÔNG-HO CH'OE

References and further reading

Asea munje yôn'guso (ed.) (1974) *Sirhak sasang ûi t'amgu (Search for Sirhak Thought)*, Seoul: Hyônmasa.(A collection of the studies on five representative *sirhak* scholars and also on *sirhak* scholars' criticism of Catholicism, by five Korean scholars.)

Baker, D. (1990) 'Sirhak Medicine: Measles, Smallpox and Chông Tasan', *Korean Studies* 14: 135-66.(An interesting study of how Chông Yagyong approached Chinese and Western medicine within *sirhak* tradition.)

Chông Sông-ch'ôl (1974) *Sirhakp'a ûi ch'ôrhak sasang kwa sahoe chôngch'i chôk kyônhae (Philosophical Thought and Socio-Political View of the Sirhak School)*, Pyongyang: Sahoe kwahak ch'ulp'ansa.(An extensive survey of *sirhak* focusing on Yi Ik, Hong Taeyong, Pak Chega and Chông Yagyong from a Marxist standpoint by a distinguished North Korean scholar.)

Kalton, M.C. (1975) 'An Introduction to Silhak', *Korea Journal* 15 (5): 29-46.(A useful overview of *sirhak* by an American specialist on Korean intellectual history.)

Kûm Changt'ae (1987) *Hanguk sirhak sasang yôngu (Study of the Thought of Practical Learning in Korea)*, Seoul: Chimmundang.(After giving an overview of *sirhak* thought, a well-known Korean specialist examines the philosophical ideas of Chông Yagyong and Ch'oe Han'gi.)

Lee, W. (1975) 'Korean Intellectual Tradition and the *Sirhak* Thought', in H.W. Kang (ed.) *Traditional Culture and Society of Korea: Thought and Institutions*, Honolulu, HI: University of Hawaii Press.(A Korean specialist examines *sirhak* within the context of Korean intellectual history.)

Palais, J.B. (1996) *Confucian Statecraft and Korean Institutions: Yu Hyôngwôn and the Late Chosôn Dynasty*, Seattle, WA: University of Washington Press.(A massive study of a representative *sirhak* scholar, Yu Hyôngwôn, examining comprehensively various ideas of Yu with a comparative perspective in traditional East Asian history.)

Yôksa hakhoe (ed.) (1974) *Sirhak yôngu immun (Introduction to the Studies of Sirhak)*, Seoul: Ilchogak.(A collection of short studies on nine sirhak scholars with Korean translation of their representative works.)

Yu Wôndong (1983) *Hanguk sirhak kaeron (General Discourse on Sirhak in Korea)*, Seoul: Chôngôm munhwasa. (A Korean economic historian examines the reform proposals of ten *sirhak* scholars.)

Situation ethics

'Situation ethics' accords morally decisive weight to particular circumstances in judging whether an action is right or wrong. Thus we should examine critically all traditional rules prohibiting kinds of actions. Proponents of these views have exerted their greatest influence in Europe and North America in the twentieth century, although such influence waned by 1980. The views received extensive scrutiny in Christian communities. Three quite different warrants were offered for privileging discrete situations. First, we should remain dispositionally open to God's immediate command in a particular time and place (theological contextualism). Second, we should take the actual consequences of particular actions as morally decisive (empirical situationism). Third, we should be ready to perform actions that compromise moral ideals when doing so improves matters in ways a given situation, with its distinctive constraints, makes viable (mournful realism).

St Augustine's famous dictum, 'Love, and do what you will', summarizes the views that many proponents of situation ethics commend. Doubtless it would distress Augustine to see his dictum employed to support views he himself repudiates. He holds that a motive of love is a necessary condition for a good action, yet denies that it is a sufficient condition. Indeed, he claims that certain actions are wrong in themselves. Those forbidden absolutely include murder, lying, and adultery (see Deontological ethics).

Proponents of situation ethics, for their part, adopt his dictum, but reject his stipulation that there are certain things that are always wrong. They attend instead to the distinctive features of the case at hand. Situation ethics is thus a movement that protests generally against the imposition of unchanging moral absolutes that prohibit everywhere certain classes of conduct. Yet deeper currents contributed to the widespread debates the movement occasioned earlier in the twentieth century (roughly until 1980), both in Europe and North America. Some currents came from secular communities, where existentialist thinkers stressed the importance of 'authentic' personal decisions, and attacked rules that blocked spontaneity and awareness of change (see Existentialist ethics). Other currents came from Christian communities, where so many of the debates occurred, and where three warrants for concentrating on particular situations proved especially influential.

One warrant - theological contextualism - stems from the claims that Søren Kierkegaard advances in *Frygt og Baeven* (*Fear and Trembling*) (1843). He examines the biblical narrative of Abraham's near-sacrifice of Isaac. Kierkegaard insists that Abraham cannot 'love' Isaac if he disobeys God's command, and that Abraham paradigmatically displays a dispositional openness to God's immediate command. We can suppose for the best of human reasons that a given action is wrong, but if God commanded it in a certain context, we should have to change our minds. Twentieth century religious thinkers who extend a case for immediacy include Karl Barth and Emil Brunner. The particular situation has priority over 'code morality' in that it must be *God* who commands in the present moment. We may jeopardise divine sovereignty and be diverted from a personal confrontation with God by claiming to know in advance what we should do and forbear so that we only interpret, apply, and execute this knowledge. The enterprise of 'casuistry' with its subsumption of cases is thought, fairly or not, to afford an opportunity for such diversion (see Casuistry). Commended in its place is a kind of act-deontology. God's immediate command, self-interpreting in a particular situation, may sometimes lead us to suspend our antecedent moral judgments.

A second warrant - empirical situationism - jettisons any appeal to immediate divine commands, and proceeds instead to apply directly to each situation the general injunction 'to love your neighbour as yourself'. In *Situation Ethics* (1966), Joseph Fletcher contends that we should determine what action is right in a particular situation by confronting our loving wills with the contextual details of the situation itself. He allies his 'situationism' formally with modified act-utilitarianism (see Utilitarianism). Neighbour-love, like utility, is the sole governing criterion that should always guide action. Other subsidiary precepts and rules, including the traditional prohibitions that Augustine takes as absolutes, Fletcher affirms only as 'illuminative maxims'. They are at most 'statistically preponderant generalizations', based on past cases and cumulative experience that inform us of the kinds of action that typically produce good or bad consequences. But we should suspend or reject maxims whenever it seems more loving to do so, that is, whenever we effect more good by breaking them than by adhering to them. If a traditional maxim proves unloving in this situation, we should disregard it, without remorse, and without a sense of

conflicting obligations.

A final warrant - mournful realism - accents the moral compromises a particular situation may exact. In *The Nature and Destiny of Man* (1949), Reinhold Niebuhr laments that a particular situation may present us with unwanted choices, and induce us to consider doing things we may judge repellent, or at least things we would never do for their own sake. But he justifies making choices we do not want to make, when the alternative is to acquiesce to corrupt forces that threaten altogether the conditions for a tolerable community.

This third warrant is like the second in several ways. It judges that immediate divine commands are not possibilities to be taken routinely into account. Neighbour-love remains insuperable. No one subsidiary precept or rule binds absolutely, whatever the consequences. Contextual details may assume effective priority in determining what actions are appropriate. There are, however, some differences. Fletcher is too complacent when he maintains that the greatest moral good possible in a situation is the most loving thing to do. For what is possible often demands compromise that leaves us morally dissatisfied. Compromise never becomes *all right*. Fletcher also goes wrong in weighing all considerations from the ground up each time we decide. While subsidiary precepts and rules are not exceptionless or always decisive, they should never be omitted from consideration in any situation in which they are relevant. Some carry a presumption in their favour. They are not to be disregarded but only overridden, with compunction, and with a sense of competing obligations.

Although situation ethics as a self-conscious movement with its own internal debates faded, the importance of the issues it addressed persists. It still has lessons to teach.

See also: Intuitionism in ethics; Moral judgment; Moral realism; Moral sense theories

GENE OUTKA

References and further reading

Augustine (416) *In ioannis evangelium tractatus (Homilies on the Gospel of John)*, ed. and trans. J. Burnaby, *Ten Homilies on the First Epistle of St John*, in *Augustine: Later Works*, Library of Christian Classics, Philadelphia, PA: The Westminster Press, London: SCM, 1955, vol. 8, 316.(Love is the theme of his exhortations throughout, as it is the theme of this epistle.)

Fletcher, J. (1966) *Situation Ethics*, Philadelphia, PA: The Westminster Press.(A popular summary of key propositions.)

Frankena, W.K. (1964) ‘Love and Principle in Christian Ethics’, in A. Plantinga (ed.) *Faith and Philosophy*, Grand Rapids, MI: Eerdmans, 203-25.(An essay that shapes the terms of the debate.)

Kierkegaard, S.A. (1843) *Frygt og Baeven*, trans. H.V. Hong and E.H. Hong, *Fear and Trembling*, Princeton, NJ: Princeton University Press, 1983.(Defends the possibility of a ‘teleological suspension of the ethical’.)

Mahoney, J. (1987) *The Making of Moral Theology: A Study of the Roman Catholic Tradition*, Oxford: Clarendon Press, ch. 5.(Describes situation ethics in Roman Catholic discussions.)

Niebuhr, R. (1949) *The Nature and Destiny of Man*, New York: Charles Scribner’s Sons.(The full dress account of Christian realism.)

Outka, G. (1972) *Agapē; An Ethical Analysis*, New Haven, CT: Yale University Press, chaps 4, 6.(Discusses senses of situation ethics.)

Outka, G. and Ramsey, P. (eds) (1968) *Norm and Context in Christian Ethics*, New York: Charles Scribner’s Sons; London: SCM.(Contains original theological and philosophical essays covering a spectrum of Protestant and Roman Catholic views.)

Skinner, Burrhus Frederick (1904-90)

B.F. Skinner advocated a philosophy of psychology, called 'radical behaviourism', as well as a substantive psychological theory, 'scientific behaviourism'. Radical behaviourism restricted psychology to establishing lawful links between the environment and behaviour, rejecting the 'mind' as a 'needless way station' mediating the two. Scientific behaviourism proposed specific links, the laws of 'operant conditioning', whereby behaviours are 'reinforced' by the consequences they have had in an animal's past. Although Skinner brought to psychology new standards of experimental rigour, and managed to train animals to do remarkable things, there are serious limits to the range of behaviours scientific behaviourism can explain. Both it and radical behaviourism have been obviated by the development of a computational theory of mind.

Influenced by reading Bertrand Russell, Skinner become an adherent of Watson's thesis that 'everything that can be known about man is discoverable by the method of external observation' (1925), that is, that none of our knowledge depends upon introspection; a doctrine called 'methodological behaviourism'. Skinner went further and concluded from this that any serious psychology should therefore confine itself to finding lawful relations between externally observable stimuli and responses, rejecting internal mental states as 'needless way stations' (1963), a doctrine called 'radical behaviourism' (see Behaviourism, methodological and scientific). Skinner often identified mentalism with dualism, which he thought entailed that the mental was essentially 'private'. Alternatively, he sometimes endorsed a version of 'analytical' behaviourism, according to which the meaning of mental talk is be *analysed* in terms of (dispositions to) overt, publicly observable behaviour. Sometimes he even seemed to abandon behaviourism for mere physicalism, allowing psychology to refer to mental events so long as such terms as 'being hungry' or 'thinking to oneself' were physical events in the brain.

Skinner made two very important contributions to the more substantive theory, scientific behaviourism. First, he broadened Pavlov's work (1927) on 'classical conditioning' to include not only elicited, but spontaneous behaviours (what he called 'operants', since they operated in the environment), subsuming them all under Thorndike's 'law of effect' (1911). This says, roughly, that the probability of a response *R*, given a stimulus *S*, is increased/decreased if *R* following *S* has been paired with a positive/negative reinforcement *F* (for example, a reward or a punishment) in the past. Thus, for example, if a baby says 'Ma' [*R*] in front of its mother [*S*], and the mother hugs the child [*F*], the probability of the child saying 'Ma' in front of the mother again increases. Skinner believed that by suitable patterns of conditioning one could explain all the intelligent behaviour of people and animals, which, moreover, could be 'shaped' into almost any form. He wrote popular works expounding this view, most notably his utopian novel, Walden Two, and a philosophical tract, Beyond Freedom and Dignity, in which he draws the kinds of consequences for the ordinary conception of freedom that are drawn by many other determinists (see Free will).

Second, Skinner brought to light surprising facts about the patterns (or 'schedules') of reinforcement, for example, that they were more effective when they were intermittent, rather than constant, and that punishments were often ineffective in delicately shaping behaviour (see Ferster and Skinner 1957; Skinner 1969). These findings remain interesting even for those who reject behaviourism.

Problems with scientific behaviourism arose primarily in explaining the very behaviour of rats in mazes that was supposed to be the theory's parade example. Rats can learn 'latently' (without reinforcement, when they have been sated) or 'passively' (without emitting responses, as when they are wheeled through a maze in a wagon), and can improvise new, appropriate behaviour that has not been reinforced (for example, rats trained to run a maze will swim it when it is flooded). Both this and other ethological evidence suggests that animals in general navigate not by associating stimuli with responses, but by exploiting some sort of internal 'map' (Tolman 1948), and/or computation on vectors that indicate how long they have moved in a certain direction (Gallistel 1990).

One of Skinner's reasons for spurning the mental had to do with a philosophically motivated rejection of 'intervening variables' in an explanation. His complaint here was not that the mind did not *exist*, but that it was *irrelevant* to scientific explanation.

> We cannot account for the behaviour of any system while staying wholly inside it; eventually we must turn to forces operating on the organism from without. Unless there is a weak spot in our causal chain so that the

> second link is not lawfully determined by the first, or the third by the second, then the first and third links must be lawfully related.
>
> (1953: 35)

Hempel (1965) discussed this issue generally as 'the theoretician's dilemma' (see Explanation).

The argument assumes that in the usual case there is a three-term causal relation consisting of some set *E* of environmental events causing some set *P* of psychological/neural events, which in turn cause some set *R* of behavioural responses. The transitivity of 'cause' suggests that the middle link can be dropped. However, this begs several questions. It assumes that explanation is concerned only with predicting observable phenomena, for example, that it is only *R* that needs explaining. But what if we want to know what explains *P*, for example, why some organism is thinking to itself or wants to eat? Surely *P* cannot be dropped in that case.

More crucially, there is the further important assumption, inherited from classical empiricism, that most interesting causal chains start in the environment. However, some behaviour may result in part from endogenous features of an individual's neurochemistry or from maturational patterns determined by biological evolution. Some species of bird are disposed to learn about the positions of stars *only* when they are young and unable to fly, and become *unable* to learn about them when mature (Emlen 1969). One of Skinner's most penetrating critics, Chomsky (1959), argued that the remarkable rapidity with which children learn their first language requires that substantial knowledge of 'universal grammar' be innate (see Chomsky, N.; Language, innateness of).

Scientific behaviourism brought to psychology a higher standard of experimental rigour than it had previously enjoyed, and led to more disciplined tests for even non-behaviourist claims. However, apart from his findings about schedules of reinforcement, few of Skinner's claims remain influential. Scientific behaviourism has been eclipsed by the rise of cognitive and computational theories in psychology that, moreover, obviate the strictures of radical behaviourism, since these theories are compatible with materialism, and are entirely open to public verification.

See also: Behaviourism, analytic; Behaviourism, methodological and scientific; Operationalism

OWEN FLANAGAN
GEORGES REY

List of works

Skinner, B. (1938) *The Behavior of Organisms: An Experimental Analysis*, New York: Appleton-Century-Crofts. (Skinner's first major presentation of his views.)

Skinner, B. (1948) *Walden Two*, New York: Macmillan.(Skinner's utopian novel about a community organized according to behaviourist principles.)

Skinner, B. (1953) *Science and Human Behavior*, New York: Macmillan.(An early popular presentation of behaviourism.)

Skinner, B. (1957) *Verbal Behavior*, New York: Appleton-Century-Crofts.(Skinner's sketch of a behaviouristic theory of language.)

Ferster, C. and Skinner, B. (1957) *Schedules of Reinforcement*, New York: Appleton-Century-Crofts.(A technical presentation of the findings regarding the effects of different temporal patternings of reinforcement.)

Skinner, B. (1963) 'Behiorism at Fifty', *Science* 140: 951-58; repr. with useful commentaries by contemporary philosophers and psychologists in *Behavioral and Brain Sciences* 7 (4): 615-21.(Revealing discussion of much of the philosophical motivation for behaviourism.)

Skinner, B. (1969) *Contingencies of Reinforcement: A Theoretical Analysis*, New York: Appleton-Century-Crofts. (Late theoretical presentation of the main scientific behaviourist results.)

Skinner, B. (1971) *Beyond Freedom and Dignity*, New York: Knopf.(Draws consequences of radical behaviourism for the ordinary conception of the person.)

Skinner, B. (1974) *About Behaviorism*, New York: Random House.(A late defence of the main claims of behaviourism.)

References and further reading

Chomsky, N. (1959) 'Review of *Verbal Behavior* (by B.F. Skinner)', *Language* 35: 26-58; repr. in J. Fodor and J. Katz (eds) *The Structure of Language: Readings in the Philosophy of Language*, Englewood Cliffs, NJ:

Prentice Hall.(Forceful attack on Skinner's account of language acquisition from a foremost linguist whose work set the stage for the cognitive revolution that has largely replaced behaviourism in psychology.)

Dennett, D. (1978) 'Skinner Skinned', in *Brainstorms*, Cambridge, MA.: MIT Press.(An excellent philosophical examination and critique of Skinner's attack on mentalistic explanation.)

Emlen, S. (1969) 'The Development of Migratory Orientation in Young Indigo Buntings', *Living Bird* 8: 113-26. (Evidence of quite specific, purely biological constraints on learning.)

Flanagan, O. (1991) *The Science of the Mind*, Cambridge, MA: MIT Press, 2nd edn.(Exposition and critique of Skinner's overall conception of behaviourism as a philosophy of science and mind.)

Gallistel, C. (1990) *The Organization of Learning*, Cambridge, MA: MIT Press.(Extremely rich discussion of recent research on animal learning and navigation, how it seems to require postulation of internal computations, and resists behaviouristic explanation.)

Gleitman, H. (1995) *Psychology*, New York: Norton, 4th edn.(One of the best textbooks in the field; chapter 4 presents an extended discussion of the experimental literature for and against behaviourism.)

Hempel, C. (1965) 'The Theoretician's Dilemma: A Study in the Logic of Theory Construction', in *Aspects of Scientific Explanation*, New York: Free Press.(Exploration of the puzzle about the role of 'intervening variables' in a theory.)

MacCorquodale, K. (1970) 'On Chomsky's Review of Skinner's *Verbal Behavior*', *Journal of the Experimental Analysis of Behavior* 13: 83-99.(A defence of Skinner's *Verbal Behavior* against Chomsky's 1959 criticisms.)

Pavlov, I. (1927) *Conditioned Reflexes*, Oxford: Oxford University Press.(The theory of 'classical conditioning', on which Skinner's theory of 'operant conditioning' was partly based.)

Pinker, S. (1994) *The Language Instinct*, New York: Morrow.(Superb popular introduction to the revolution occasioned by Chomsky's work in linguistics.)

Rey, G. (1997) *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Oxford: Blackwell.(A defence of a computational approach to mentalism; chapter 4 includes a summary of the main problems with scientific behaviourism as it concerns philosophy of mind.)

Thorndike, E. (1911) *Animal Intelligence: Experimental Studies*, New York: Macmillan.(The earliest statement of behaviouristic principles, particularly the 'law of effect'.)

Tolman, E. (1948) 'Cognitive Maps in Rats and Men', *Psychological Review* 55: 189-208.(Early proposal of an alternative to behaviourism as an explanation of navigation.)

Watson, J. (1925) *Behaviorism*, New York: Norton.(Early statement of behaviouristic ideas that influenced both Russell and Skinner.)

Skovoroda, Hryhorii Savych (1722-94)

Skovoroda was the first truly independent philosopher produced by Ukraine and the last brilliant exponent of its Baroque culture. Departing from the Aristotelian tradition of the Kiev Academy, he constructed an original synthesis of ancient and patristic thought. Because of his classical aloofness from history, the sociopolitical trauma of Ukraine's absorption by Russia, which he witnessed, is hardly reflected in his writings. His worldview foreshadows the Romantic and religious tendencies in nineteenth-century Ukrainian and Russian thought.

Hryhorii Skovoroda (born in Chornukhy, died in Pan-Ivanivka, Ukraine) interrupted his studies at the Kiev Academy first in 1741-4 to sing in the court choir in St Petersburg and then in 1745-50 to serve as music director at the Russian imperial mission in Tokai, Hungary. Returning to Ukraine, he taught poetics at Pereiaslav College and then resumed his studies in Kiev (1751-3). After serving as private tutor, he taught poetics (1759-60), syntax and Greek (1762-4) and Christian ethics (1768-9) at Kharkiv College. Dismissed because of his unorthodox teachings, he spent the rest of his life wandering from friend to friend and writing. Most of his works belong to this period: over a dozen verses, thirty fables, two treatises, twelve dialogues and over 100 letters. His manuscripts were copied and circulated among his friends; none was published during his lifetime.

Skovoroda's style of writing is largely responsible for the widely differing interpretations of his thought. He conveyed his ideas mostly through imagery, similes, metaphors, fables, myths, parables and proverbs. Even philosophical terms are often defined by images, leading to ambiguity in his doctrines. Little clarity can be gained by seeking logical links among his doctrines: he did not arrange them into a system and rarely made explicit logical connections. His favourite devices of persuasion are analogies and illustrations, not arguments. Often, he ignores the implications of his teachings - the obvious questions and difficulties that follow from his statements. Thus, some scholars have viewed Skovoroda as an eclectic without a coherent body of doctrine. Most, however, agree that there is an inner unity to his thought. But then they differ on how his ideas fit together: some take his metaphysical teachings to be central and treat him as a theologian or mystic, others claim that he is a moral thinker. There are also diverse interpretations of his principal doctrines such as God's relation to the world, personal immortality and the nature of matter.

The purpose of philosophy for Skovoroda is practical - to show the way to happiness; hence the key questions are 'What is happiness?' and 'How can it be attained?'. He defines happiness as an inner state of joy, peace and confidence. To reach this state one must know some basic truths about oneself and the world. Skovoroda limits his metaphysical and anthropological teachings to those he regards as necessary for happiness. He shows no interest in theoretical questions that have no immediate bearing on existence.

Skovoroda divides reality into three worlds: the macrocosm or universe and two microcosms - man and the Bible. All three have a similar dualistic structure: they consist of an inner ideal nature that is spiritual, eternal and immutable and an outer sensible nature that is material, transitory and changing. The inner nature is higher; it imparts being to the outer as a tree supports its shadow. In the macrocosm, the inner nature is God and the outer is the physical universe. This position is panentheist, rather than pantheist, for God is in things through his ideas, not immediately. Although parts of the universe are transitory, the universe as a whole shares in God's eternity; hence matter is eternal.

In man, the inner nature is the spirit, soul, heart or true man; the outer is the body. It is unclear whether Skovoroda takes the true man to be a universal being common to all human beings or something unique in each individual. In any case, the soul can know itself more easily than it can know other souls; thus, all knowledge of spiritual reality begins with self-knowledge.

The symbolic microcosm or Bible consists of images of all the things in the macrocosm. The Bible's inner nature is the symbolic meaning, the higher truth known by the wise men of all ages; its outer nature is the literal meaning, which is often mistaken for truth itself. The purpose of biblical stories is to lead the reader to higher knowledge; the people and events described in the stories are insignificant and the miracles, which are contrary to nature, are fictional. Skovoroda interprets biblical symbols so as to show that the Bible's true meaning is expressed in his philosophy.

From this metaphysical scheme Skovoroda draws a number of practical conclusions. Since what is most valuable in us is eternal, we have no reason to fear death. Nor have we any reason to fear privation in this life, for Providence guarantees the necessities of happiness for all. The principle that what is necessary (for happiness) is easy and what is difficult is unnecessary serves as a criterion of rational need and reasonable expectation. Besides absence of fear, happiness requires self-fulfilment - the active pursuit of one's God-given innate calling or congenial task (*srodnoe delo*). To work in one's vocation, regardless of hardships, is to be happy, while to assume an unnatural task is to suffer hellish torment regardless of external rewards. Furthermore, since vocations are distributed by God in such a way as to ensure social harmony, to evade one's vocation is to introduce disorder into society and to harm others. The doctrine of congenial work is central to Skovoroda's moral system and the foundation of his optimistic attitude to life.

Skovoroda's influence on Ukrainian or Russian philosophy was negligible. His poetry was popular among the common people and valued by the founders of Ukrainian vernacular literature in the nineteenth century. During periods of cultural revival in the twentieth century, Ukrainian writers found inspiration in Skovoroda's ideas and independent personality.

TARAS D. ZAKYDALSKY

List of works

Skovoroda, H. (1973) *Hryhorii Skovoroda: Povne zibrannia tvoriv u dvokh tomakh (Hryhorii Skovoroda: The Complete Collection of his Work in Two Volumes)*, eds V. Shynkaruk *et al*., Kiev: Naukova dumka; Ukrainian edition 1994.(The first full collection of Skovoroda's works, based on manuscripts and preserving the language he used.)

Skovoroda, H. (*c*.1773) 'Razgovor piati putnikov o istinnom shchastii v zhizni'; abridged trans. G.L. Kline, 'A Conversation Among Five Travellers Concerning Life's True Happiness', in *Russian Philosophy*, vol. 1, eds J. Edie *et al*., London and New York: Quadrangle Books, 1965, 26-57.(A key source for Skovoroda's moral doctrine.)

Skovoroda, H. (1990) *Fables and Aphorisms*, trans. D. Chopyk, New York: Peter Lang.(An entertaining supplement to Skovoroda's dialogues.)

References and further reading

Bird, T. and Marshall, R. Jr. (eds) (1994) *Hryhorij Savyc Skovoroda: An Anthology of Critical Articles*, Edmonton and Toronto: Canadian Institute of Ukrainian Studies Press.(Recent studies of Skovoroda's language, style, literary influence and philosophical ideas, with bibliography.)

Ivan'o, I. (1983) *Filosofiia i styl' myslennia H. Skovorody (H. Skovoroda's Philosophy and Style of Thought)*, Kiev: Naukova dumka.(An account of Skovoroda's cultural context, symbolism and literary style in relation to his ideas.)

Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, 2 vols; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, vol. 1, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, 53-69.(An exposition and critique of Skovoroda's philosophy.)

Slavery

The moral, economic and political value of slavery has been hotly disputed by philosophers from ancient times. It was defended as an institution by Plato and Aristotle, but became increasingly subject to attack in the modern period, until its general abolition in the Western world in the nineteenth century.

In the twentieth century our belief that slavery is fundamentally unjust has become a benchmark against which moral and political philosophies may be tested. Both utilitarians and contractarian philosophers have argued against slavery in general and the enforceability of slavery contracts more specifically, although for very different moral reasons. Others have argued that only by viewing slavery from the standpoint of the slave can its moral significance be understood.

1 Classical views of slavery

Slavery was widely practised in the ancient world, in different forms. In the Old Testament, the practice was generally limited to the women and children of the enemies of the Israelites, and was said to be instituted by God (Genesis 9: 25). The men of defeated cities were all killed (Deuteronomy 20: 13) as, often, were the women and children (2: 34, 3: 6). But women and children could be spared and enslaved by the conquerors (2: 14). Slavery also appears to be sanctioned in the New Testament, where slaves are told to obey their masters (Ephesians 6: 5; Colossians 3: 22; Titus 2: 9).

The Greeks, in contrast, often acquired adult male slaves by military means, and even conceived war as a form of hunting for slaves (Aristotle, *Politics* I.7.1255b37-40). There has been some debate over whether Plato envisaged the ideal state he describes in *Republic* as including the practice of slavery (see Wild 1953 versus Vlastos 1968), but the reference to slaves at *Republic* 433e would appear to leave no doubt in the matter. Certainly, they were included in the state described in Plato’s *Laws* (for example, 773e).

Plato never offers a whole theory of slavery, but it is clear from what he does say that the true slave is one who lacks reason (*Laws* 966b). This Platonic conception of the slave may also be found in what is certainly the most extensive philosophical treatment of slavery in antiquity: Book I of Aristotle’s *Politics*. Aristotle seeks to defend the practice of slavery as a just social institution, on the ground that it is not merely supported by existing laws and conventions, but also by nature itself. Aristotle forms his argument as a direct reply to those who argued that slavery is unnatural (*Politics* I.6). On the contrary, Aristotle is convinced that there do exist natural slaves, who are ‘from birth marked out by nature as slaves’ (I.5.1254a21-4). Like Plato, Aristotle locates the naturalness of slavery in the natural slave’s rational deficiency: the slave, Aristotle says, wholly lacks the deliberative part of the soul (I.13.1260a12), and may even be conceived as no more than a part of the natural master’s body (I.6.1255b11-12). Natural slave and natural master are both benefited by the institution of slavery: the master through access to the labour of the slave, the slave through access to the deliberative ability of the master. Aristotle concedes that purely conventional slavery - where those naturally suited to be masters are enslaved, or those naturally suited to be slaves become masters - is morally indefensible.

Both Plato and Aristotle are highly critical of what they regarded as the excessively lenient treatment slaves received in democracies (*Republic* 563b; *Politics* VI.4.1319b28). Modern thinkers may well wonder at such an assessment in the context of democratic Athens, in which slaves were treated - by modern lights - very harshly indeed: slaves could be beaten at the whims of their owners (MacDowell 1978). But Aristotle also thinks that Plato’s conception is excessively harsh, for Plato imagines that even the use of admonition ‘spoils’ the slave (*Laws* 773e), whereas Aristotle argues that the slave is even more appropriately admonished (as opposed to simply commanded) than is a child (*Politics* I.13.1260b5-7). It is problematical what Aristotle may have thought merited the rationalizations of admonition in the slave given that Aristotle makes his case for the naturalness of slavery precisely on the ground that the slave lacks reason (Smith 1991).

2 Later views of slavery

Later views of slavery through the Middle Ages combined this Aristotelian naturalism with the Christian concept of original sin. Augustine placed special emphasis on the latter; for him the ‘prime cause’ of slavery was sin, not nature (*City of God* XIX, 15). Aquinas leaned in the other direction (*Summa theologaie* Ia.96.4). Could slavery

have existed in man's natural state? By likening slavery to old age, Aquinas barely managed to avoid an affirmative answer. Slavery is occasioned by sin, but like old age and death it derives its naturalness from the rational order of being (3.suppl.52.1).

Jean Bodin dissented from early modern apologies for slavery, attacking Aristotelian and Augustinian views on empirical grounds (1606: I.5). Thomas Hobbes, who shared Bodin's preference for absolute sovereignty, took a more hypothetical view. Despite his vociferous attacks on Aristotle and 'the Schoolmen', his own argument that slavery is irrational rests, in part, on the acknowledgement that some people are better at rational deliberation than others. The problem with Aristotle's view is that lacking absolute sovereign power, this particular difference may be less salient than it seems in a more stable society. Hence, Hobbes' 'ninth law of nature' urges individuals in the 'state of nature' not to be proud and to acknowledge one another as natural equals (1651: ch. 15). For less-able deliberators the alternatives to consenting to an absolute sovereign (living as outlaws or being slaves) are worse (Kidder 1983: 142).

Despite Locke's standing as a theorist of individual liberty and his denunciation of slavery as 'so vile and miserable an estate', (1690: I.i), he gave a qualified defence of slavery as part of his conception of a just war. When an unjust aggressor wages war, the just conqueror may enslave the former (II.xxiii.85). But even where an unjust act of aggression occurs, further conditions still must be met. The conqueror cannot enslave his own conquering forces, cannot take more than is needed to repair the damages incurred in waging war, and cannot enslave noncombatants. This last qualification extends to future generations and rules out the confiscation of land because land belongs to future generations by virtue of their inheritance rights (Farr 1986: 272-3) (see Locke, J. §10).

Enlightenment philosophers, while not unanimous in their condemnation of slavery, shifted the balance of arguments more in that direction. Montesquieu, sometimes ironically, rejected Aristotle's view but added that in warm climates the threat of slavery may be needed to overcome laziness (1748: 15.7). Unsatisfied with this argument, however, he then wondered whether the laws in such climates are responsible for human laziness, and whether perhaps slavery would not be needed if there were better laws to make people less lazy (15.8). Rousseau made a similar point more forcefully. Human nature can be servile when it is enslaved, but Aristotle 'mistook the effect for the cause' (Rousseau 1762: I.ii). Neither nature nor conquest justifies slavery.

The moral stance taken by the *philosophes* against slavery gave way in the nineteenth century to a complex array of historical interpretations. Hegel argued that in the dialectical struggle for recognition the master's drive for autonomy breeds dependency upon the slave at the same time as the slave's own labour provides him with a mind of his own (Hegel 1807). Slavery, on this interpretation, is a necessary stage in the history of human emancipation. African slavery is especially necessary and should be eliminated only gradually given the African's stage of moral development (1822-30: 183).

Hegel's emphasis on the transformative power of labour captured Marx's imagination. He saw the wage worker under capitalism as engaged in something akin to the dialectical struggle for recognition between master and slave. As the 'wage-slave' shapes the product of his labour he contributes to the creation of a more self-conscious working class ([1844] 1978: 112) (see Alienation; Marx. K. §4). In contrast, Nietzsche (§8) argued that the new morality of the slave was not a form of greater, emancipatory self-consciousness, but *ressentiment* and a base preoccupation with mere existence (1887: I.10). Eventually, however, these 'last men' and their democratic societies would give way to a more noble life (1886: §202).

John Stuart Mill, against the background of a harmonious philosophy of history, focused more narrowly on the morality of voluntary slavery contracts (1859: ch. 5). Despite his strong stance against paternalism, Mill thought that slavery contracts should not be enforced for two reasons: slavery is always more harmful than freedom and freely choosing not to be free is incoherent. Scholars have questioned both arguments; under certain difficult circumstances, it is neither unreasonable nor most harmful for individuals to agree to be the slaves of benevolent slave-holders (Callahan 1985: 225).

More orthodox utilitarians since Mill have been troubled by the argument that under some conceivable circumstances they would have to permit even non-voluntary slavery. In response they claim that utilitarianism is a morality for the world as it actually is and one that respects all individuals equally (see Utilitarianism). Under

actual circumstances utilitarian moral reasoning would not justify slavery (Hare 1979). In contrast, contractualists have argued that the problem is not what could conceivably be justified on utilitarian grounds, but the administrative kind of reasons utilitarians give in actual cases. According to contractualists, slavery is wrong because the slave-holders have no moral standing and the practice of slavery would not be accepted by the slaves themselves as a reasonable, permanent form of social cooperation (Rawls 1971a: 260-6), even if in dire circumstances it was temporarily acceptable to them (Rawls 1971b: 248) (see Contractarianism §7-8).

Finally, some philosophers have argued that in addition to the question of justification, one can look back through the experience of slavery from the standpoint of the slaves, not the slave-holders, to understand more fully complex moral concepts such as paternalism, forgiveness, citizenship and resistance (McGary and Lawson 1992).

STEPHEN L. ESQUITH
NICHOLAS D. SMITH

References and further reading

Aquinas, T. (1266-73) *Summa theologaie*, trans. Fathers of the English Dominican Province, London: Burns Oates & Washburne Ltd, 1912-36.(Passages in Ia and 3, supplement, suggest an Aristotelian interpretation of original sin as the 'prime cause' of slavery.)

Archard, D. (1990) 'Freedom not To Be Free: The Case of the Slavery Contract in J.S. Mill's *On Liberty*', *The Philosophical Forum*, October 40 (160). 453-65.(Defends and clarifies the limits of Mill's arguments against enforcing voluntary slavery contracts.)

Aristotle (*c*.mid-4th BC) *The Politics*, ed. S. Everson, Cambridge: Cambridge University Press, 1988.(Book I offers Aristotle's theory of natural slavery.)

Augustine (413-27) *City of God*, trans. M. Dods, New York: Random House, 1950.(Chapter XIX outlines a defence of slavery in terms of original sin.)

Bodin, J. (1606) *The Six Bookes of a Commonweale*, ed. K.D. Macrae, Cambridge: Harvard University Press, 1962.(Book I, chapter 5 offers an empirical argument against slavery.)

Callahan, J.C. (1985) 'Enforcing Slavery Contracts', *The Philosophical Forum* 16 (3): 223-36.(Carefully reviews the most prominent arguments for and against enforcement of voluntary slavery contracts.)

Davis, D.B. (1966) *The Problem of Slavery in Western Culture*, Ithaca, NY: Cornell University Press. (Comprehensively reviews the history of the idea and practice of slavery in the West.)

Farr, J. (1986) ''So Vile and Miserable an Estate': The Problem of Slavery in Locke's Political Thought', *Political Theory* 14 (2): 169-88.(Clarifies the Lockean conditions under which slavery is justified against an unjust aggressor.)

Fortenbaugh, W.W. (1977) 'Aristotle on Slaves and Women', in J. Barnes, M. Schofield and R. Sorabji (eds) *Articles on Aristotle 2: Ethics and Politics*, London: Duckworth, 135-39. (Seeks to defend the logical coherence of Aristotle's account in *Politics*.)

Garver, E. (1994) 'Aristotle's Natural Slaves: Incomplete 'Praxeis' and Incomplete Human Beings', *Journal of the History of Philosophy* 32 (2): 173-95.(Seeks to defend the logical coherence of the Aristotelian account.)

Glausser, W. (1990) 'Three Approaches to Locke and the Slave Trade', *Journal of the History of Ideas* 51 (2): 199-216.(Argues that Locke's position on slavery was not coherent.)

Hare, R.M. (1979) 'What is Wrong with Slavery?', *Philosophy and Public Affairs* 8 (2): 103-21.(Defends utilitarianism against the charge that it could possibly justify slavery.)

Hegel, G.W.F. (1807) *Phenomenology of Mind*, trans. J.B. Baillie, London: Allen & Unwin, 1966.(Book IVA discusses the master-slave dialectic.)

Hegel, G.W.F. (1822-30) *Lectures on the Philosophy of World History: Introduction*, trans. H.B. Nisbet, Cambridge: Cambridge University Press, 1975.(Argues for the gradual elimination of African slavery.)

Hobbes, T. (1651) *Leviathan*, New York: MacMillan, 1962.(Chapter 15 suggests a prudential argument for submitting to absolute sovereignty.)

Hunting, C. (1978) 'The '*Philosophes*' and the Question of Black Slavery: 1748-1765', *Journal of the History of Ideas* 39 (4): 405-18.(Defends the '*philosophes*' as critics of black slavery.)

Kidder, J. (1983) 'Acknowledgements of Equals: Hobbes's Ninth Law of Nature', *The Philosophical Quarterly* 33 (131): 133-46.(Reduces the distance between Aristotle's natural theory of slavery and Hobbes's views on equality.)

Locke, J. (1689/90) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1960. (In I.i, argues against slavery as a violation of natural rights.)

MacDowell, D. (1978) *The Law in Classical Athens*, Ithaca, NY: Cornell University Press.(Provides a good summary of the laws concerning slavery in ancient Athens.)
McGary, H. and Lawson, B.E. (1992) *Between Slavery and Freedom: Philosophy and American Slavery*, Bloomington, IN: Indiana University Press.(Analyses moral concepts related to slavery from the standpoint of the slave.)
Marx, K. (1844) *Economic and Philosophic Manuscripts of 1844*, trans. M. Milligan, in *The Marx-Engels Reader*, ed. R.C. Tucker, New York: W.W. Norton, 1978. 2nd edn.(Draws an analogy between slavery and wage labour.)
Mill, J.S. (1859) *On Liberty and Other Writings*, ed. S. Collini, Cambridge: Cambridge University Press, 1989. (Chapter 5 argues against the paternalistic enforcement of voluntary slavery contracts.)
Montesquieu, C. de Secondat, Baron de (1748) *The Spirit of the Laws*, trans. and ed. A.M. Cohler, B.C. Miller and H.S. Stone, New York: Cambridge University Press, 1989.(Chapter 15 argues ironically and with some hesitation against the Aristotelian theory of natural slavery.)
Nietzsche, F. (1886) *Beyond Good and Evil, in The Basic Writings of Nietzsche*, trans. W. Kaufmann, New York: Random House, 1968.(Section 202 foresees the eventual passing of slave morality.)
Nietzsche, F. (1887) *On the Geneology of Morals, in The Basic Writings of Nietzsche*, trans. W. Kaufmann, New York: Random House, 1968.(Contains a criticism of slave morality in I.10.)
Plato (*c*.mid-4th BC) *Republic*, in *Plato: Collected Dialogues*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1963.(At 433e, appears to take it for granted that the ‘ideal state’ would use slaves.)
Plato (*c*.mid-4th BC) *Laws*, in *Plato: Collected Dialogues*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1963.(At 773e, shows that slaves also will be included in the ‘second-best state.’)
Plato (1971a) ‘Justice as Reciprocity’, in *Mill: Utilitarianism*, ed. S. Gorovitz, Indianapolis, IN: Bobbs-Merrill. (Argues from hypothetical consent against the practice of slavery.)
Rawls, J. (1971b) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(Notes the circumstances in which slaves themselves might prefer slavery to other likely outcomes.)
Rousseau, J.-J. (1762) *On the Social Contract*, ed. R.D. Masters, trans. J.R. Masters, New York: St Martin’s Press, 1978.(Argues against Aristotle’s theory of natural slavery.)
Salter, J. (1992) ‘Adam Smith on Feudalism, Commerce, and Slavery’, *History of Political Thought* 13 (2): 219-41.(Reviews Smith’s views on the decline of slavery and economic development.)
Schutrumpf, E. (1993) ‘Aristotle’s Theory of Slavery - A Platonic Dilemma’, *Ancient Philosophy* 13 (1): 111-23. (Argues that Aristotle’s confusion over slavery is a legacy from Plato.)
Smith, N.D. (1991) ‘Aristotle’s Theory of Natural Slavery’ in D. Keyt and F.D. Miller, Jr (eds) *A Companion to Aristotle’s Politics*, Oxford: Blackwell, 142-55.(Argues that Aristotle’s theory of slavery in *Politics* is logically incoherent.)
Smith, S.B. (1992) ‘Hegel on Slavery and Domination’, *The Review of Metaphysics* 46 (1): 97-124.(Argues for a theological reading of Hegel’s critique of Aristotle’s theory of natural slavery.)
Vlastos, G. (1968) ‘Does Slavery Exist in the Republic?’, *Classical Philology* 63 (4): 291-5.(Argues that Plato included slavery in his ‘ideal state.’)
Wild, J. (1953) *Plato’s Enemies and the Theory of Natural Law*, Chicago. IL: Chicago University Press.(Denies that Plato included slavery in his ‘ideal state.’)

Slavophilism

In the Slav countries outside Russia the term 'Slavophilism' is a generic name for all advocates of the 'Slav idea', irrespective of their philosophical views and political commitments. In Russia, however, this term is used, as a rule, to denote one specific ideology, elaborated in the 1840s by the former members of the Schellingian circle of 'Lovers of Wisdom': Ivan Kireevskii (1806-56) and Aleksei Khomiakov (1804-60). Among its other followers, the most creative were the former Hegelians - Konstantin Aksakov (1817-60) and Iurii Samarin (1819-76). Despite some individual differences, all these thinkers shared a coherent view of the world which was expressed in their philosophical, theological and historical ideas. Their importance was not immediately recognized, but after Dostoevskii and Solov'ëv it became clear that they were the most important part of Russia's 'philosophical awakening' in the first half of the nineteenth century.

The words 'Slavophiles' and 'Slavophilism' were originally intended as gibes. The same was true of the words 'Westernizers' and 'Westernism'. All these terms, however, could be interpreted positively and were finally accepted by both sides of the 'Slavophile-Westernizer' controversy. But in the case of the Slavophiles it was a very reluctant acceptance: they felt that the term 'Slavophilism' failed to express the essential nature of their philosophical and religious position. In addition, this term contained a rather misleading suggestion as to their solidarity with non-Russian Slavs: in fact they focused their attention on the 'truly Christian' and 'purely Slav' spiritual heritage of pre-Petrine Russia. An interest in the fates of non-Russian Slavs began to play a role in their ideology only at the time of the Crimean War. This shift of focus transformed the original Slavophilism into a form of imperial Russian Pan-Slavism.

1 Theory of integral knowledge and integral personality

The first sketch of the Slavophile worldview was Kireevskii's unpublished article 'A Reply to Khomiakov' (1839). The most elaborate presentation of its philosophical foundations is Kireevskii's essay 'On the Necessity and Possibility of New Principles in Philosophy', published posthumously in 1856.

Kireevskii's philosophical views revolve around the concept of 'integral knowledge' and 'integral personality', opposed to the disintegrating rationalism and individualism of the West. In the theory of knowledge he accused rationalism of taking an anti-ontological position, separating the knower from the known and thus transforming reality into a scattering of isolated fragments, united only by a cobweb of abstract relations. The power of reason, he claimed, is purely formal and, therefore, cannot attain the 'substantive' penetration of reality; the substantial can be grasped only by substantiality, that is to say, by personality as an integral whole. True knowledge is a result of the harmonious cooperation of all the powers of the personality; the precondition of such knowledge is wholeness of spirit and a living, direct bond between the knower and reality. Hence it must be a kind of revelation, an immediate understanding available through faith. But authentic faith (and therefore knowledge) cannot be experienced by an isolated individual; it must be rooted in the supraindividual, corporate consciousness of community. This supraindividual, communal consciousness had been destroyed in the socially atomized West; it was preserved, however, by the Orthodox Church. In Kireevskii's view, membership of the Orthodox Church made it possible for the Russian thinkers to create an entirely new, truly Christian philosophy, breaking with the legacy of rationalism and providing an effective remedy for the European spiritual crisis.

In his philosophy of man Kireevskii developed the concept of 'integral personality', setting it against the Western concept of the 'autonomous individual'. He described autonomy as a divisive, destructive principle: the autonomy of the individual brings about the atomization of society, the autonomy of reason destroys the harmony of the soul, splitting human personality into separate forces, each of them claiming autonomy for itself. In this way human life becomes fragmented into different aspirations: knowledge, politics and economic activity separate themselves from religion and morality, the state separates itself from the Church, aesthetic values become divorced from moral values, and so forth. Thus the striving for autonomy is incompatible with personal wholeness. The latter can be achieved only through the bringing together of all spiritual forces, through the religious concentration of spirit.

Kireevskii presented his philosophy of 'integralism' as rooted in the Russian Orthodox tradition and derived ultimately from the teaching of the Holy Fathers of Eastern Christianity. There was some truth in this claim: it is easy to see an affinity between Slavophile antirationalism and the programmatic anti-intellectualism of Eastern

apophatic theology. Nevertheless it is fair to say that the direct source of Kireevskii's philosophy should be seen rather in the ideas of the German conservative Romantics. At the same time it must be stressed that the main target of Kireevskii's critique of rationalism was not the abstract rationalism of the Enlightenment (as was the case with the early Romantics) but the dialectical, historicized rationalism of Hegel. In a sense, Slavophile philosophy was a Russian response to the crisis of Hegelian 'Absolute Idealism'. Kireevskii, along with Khomiakov, saw Hegel as the greatest and most consistent rationalist philosopher, in whom Western thought attained its culminating point; hence the crisis of Hegelianism was for them the crisis of Western philosophy as such. They sympathized with the aged Schelling whose 'philosophy of revelation' was consciously opposed to Hegelian panlogism. They were convinced, however, that a genuinely new and truly Christian philosophy could spring only from the Russian soil (see Hegelianism, Russian §4; Schellingianism).

According to the Slavophile philosophers, Hegelian Absolute Idealism was proof that rationalism provides only an abstract, purely formal, empty knowledge, and attempts to create 'a world without a substratum'. Consequently, their own philosophical standpoint was a consistent ontologism: that is, the view that knowledge is a part and function of an existential penetration of reality. This standpoint was almost universally accepted by Russian religious philosophers of the early twentieth century. Some of them (Berdiaev, N.A. and Ern) saw in it a distinctive feature of Russian philosophy.

2 Philosophy of history

The best-known part of Slavophile ideology was their philosophy of history. It revolved around two interconnected antitheses: (1) between the 'ancient' (pre-Petrine) and the 'modern' (post Petrine) Russia, and (2) between Russia and Europe.

According to Kireevskii, Western civilization was composed of three constitutive elements: (1) Christianity, (2) the young barbaric peoples which destroyed the Roman Empire, and (3) the heritage of the ancient pagan world. The main difference between Russia and Western Europe lay in the fact that Russia lacked this ancient Roman heritage. In his pre-Slavophile period Kireevskii saw this difference as a great misfortune for his country; by the end of the 1830s, however, he came to see it as a blessing. Following some conservative German Romantics (especially Franz von Baader and Adam Müller) he saw the essence of Roman civilization in its rationalism, as embodied in the rational and formalized system of Roman law (see Romanticism, German). Due to this, the ancient Roman society was an aggregate of rationally thinking individuals, motivated solely by individual profit; the only social tie was a bond of interest; the only unity, the unity of a party. There was no real community, but only an association of isolated human beings, lacking any common faith, convictions and customs. In such a situation the only regulator of human relations was legal convention backed by the external force of the state machine.

Kireevskii believed that the juridical rationalism of ancient Rome had exerted a fatal influence upon Western Christianity and, consequently, upon the entire course of Western history. He developed this view with the aid of Khomiakov, the Slavophile theologian and historian of the Church. Rationalism, inherited from ancient Rome, was, they argued, the prime disintegrating factor in the history of Western Christianity. The Roman Catholic Church broke away from the real 'catholicity' of tradition by arbitrary changes of dogma and attempts to prove them in a rational, scholastic way. The Church became a hierarchically organized juridical institution, external and alien to her body of believers. The religious individualism of the Reformation was a justified protest against Catholic authoritarianism; its consequence, however, was atheism and the apotheosis of egoism, which laid the spiritual foundation for the modern, rational and industrial European civilization. Thus, the ancient Roman heritage, the increasing rationalism of the Western Churches, the secularizing effect of the Enlightenment, the French Revolution and, finally, modern industrialism were seen by the Slavophiles as links in one historical process, replacing organic communal ties by legal conventions and paving the way for a completely soulless, mechanical civilization.

Old, pre-Petrine Russia represented an entirely different form of social development. Russian Orthodoxy, not infected by Western rationalism, preserved the teachings of Christ in all their purity. In contrast to the West, whose political systems were formed, as a rule, as a result of conquest, the old Russian state was created by a voluntary act of the people; thus in Russia the relationship between the rulers and the ruled could be based upon mutual confidence and solidarity. The basic cell of the old-Russian social structure was the *mir*, or *obshchina* - a self-governing unit based upon communal ownership of the land, respecting old traditions and managing its affairs

according to the principle of unanimity. These elementary cells were united into larger structures ruled by the same basic principles. All ’Holy Russia’ was one great *mir*, one great community of land, customs and faith. Thus, unlike the West, it was a society where the integrating factors were not mere interests but moral convictions, not things but values.

Konstantin Aksakov, the most original thinker among the younger Slavophiles, developed these views in the direction of an extreme communitarianism, combining a kind of Christian anarchism with an apologia for autocracy. For him the contrast between pre-Petrine Russia and the West was a contrast between internal and external truth. By ’internal truth’ he meant the voice of conscience and a living tradition freely expressed in community life; by ’external truth’ he meant juridical and political forms. In his view law and state represented the sphere of society’s self-alienation; they demanded obedience irrespective of moral convictions, under the threat of physical coercion, and were therefore incompatible with morality, which they tried to replace with mere legality. The worst form of government was the modern republic since it demanded from the people an active political participation, that is an active involvement with ’external truth’.

In contrast to the West, pre-Petrine Russia accepted law and the state merely as a sad necessity, never as a principle. Autocracy in the sphere of politics was not only compatible with freedom but even an essential condition of it: it was so because true freedom is not political freedom but freedom from politics, and such a freedom could exist only under an autocratic ruler who could carry the burden of politics on his own shoulders. In Muscovite Russia the relations between the state and the land consisted in mutual non-interference: the people never tried to rule and the monarchs (with the regrettable exception of Ivan the Terrible) did not interfere with the communal life of the land. The idea of the absolute state, the state which can arbitrarily change or destroy the living principles of society, was introduced into Russia by Peter the Great. After his violent Westernizing reforms the educated classes, the ’servants of the state’, betrayed their national heritage and only the plain folk preserved in its life the old ’internal truth’. This diagnosis substantiated Aksakov’s view that the peasant commune, integrated by moral principles and common ownership of the land, was the only remedy against the deep moral crisis resulting from Westernization, which threatened Russia with a destructive revolution. Hence Russia should regenerate itself through returning to its old ways and bridging the gap between the educated classes and the common people. This was the conclusion of Aksakov’s long memoir *O vnutrennem sostoianii Rossii (On the Internal State of Russia)*, which he submitted to the new Tsar, Alexander II, in 1855 on his accession to the throne.

3 Ecclesiology and the concept of *sobornost’*

An important part of the Slavophile doctrine was its ecclesiology, developed by Khomiakov. It was an all-embracing theory of the Church as a divine-human organism. Its central concept, corresponding roughly to Kireevskii’s ’integrality’, was *sobornost’*. This term can be associated with both Church sobor (council) and the verb sobirat’ (to gather, unite). According to Khomiakov, the word *sobornost’*, which often occurs in old texts in Church-Slav, exactly corresponded to the Greek concept of ’catholicity’, which meant universal unity, the unity of all, personified in the Church council as an organ of the Holy Spirit.

Sobornost’, or catholicity, conceived in this way was a category which had an ecclesiological, social and epistemological meaning alike. Khomiakov derived its epistemological sense from the thesis that knowledge is rooted in will and faith, and these in turn depend on the strength of the bonds that link the individual to the Church community. The individual acquires knowledge of the truth only by uniting in love with the universal Church and thus becoming an organ of *sobornost’ soznaniia*, that is to say, supraindividual consciousness stemming from the charismatic unity of life. Separation from that community means entering the fatal road of rationalism and individualism. This was the case of the Church of Rome: by arbitrarily changing the Creed (the insertion of the Filioque) it raised the private opinion of Western bishops to the rank of a dogma, violating thereby the fundamental principle of catholicity. From that moment on unity and freedom became incompatible concepts in the West. The history of Europe changed into the history of ’unity without freedom’ (Catholic authoritarianism) and ’freedom without unity’ (Protestant individualism). The secret of the harmonious reconciliation of unity and freedom was preserved by the Orthodox Church, which remained faithful to the old Christian tradition of *sobornost’*. Hence the Orthodox Church had become the only universal Church, the only depository of the true ’catholicity’.

Khomiakov’s ecclesiology, systematically expounded in his treatise *The Church is One* and developed in a number

of polemical pamphlets and private letters (including letters to William Palmer, a member of the Oxford Movement whom he tried to convert to Orthodoxy), greatly differed from the standard views of the Russian Church hierarchy of his time. It was radically *pneumatic*, since it defined the essence of the Church as neither an institution, nor a doctrine, but as the charismatic community of tradition that covers all the faithful. It was *immanentist*, since it saw the inner consciousness of the Church as the only criterion of truth and categorically rejected the idea of authority. Its holistic and historicist character expressed itself in the denial that any local church (let alone an individual bishop) could be treated as the authoritative teacher. The whole truth, Khomiakov argued, is known only to the Church taken as a whole, including the laity; it crystallizes in an organic historical process, hence even the decisions of the Church councils need to be confirmed through becoming a part of the universally accepted tradition.

It is understandable that the hierarchy of the Russian Church saw in Khomiakov’s theological and ecclesiological works many dangerous ideas and (until 1879) strictly prohibited their publication in Russia. The Slavophiles, however, embraced these ideas with enthusiasm. Samarin wrote in 1867 that future generations would probably treat Khomiakov as a thinker who deserved to be called a Doctor of the Church. This prediction proved to be quite accurate. Despite dissenting voices (among them that of Father Pavel Florenskii), the prominent orthodox theologians of the early twentieth century (such as Father S.N. Bulgakov, Father G. Florovskii and many others) perceived Khomiakov’s views as the best insight into the true essence of Orthodoxy, an ’eyewitness account’ (Florovskii) of the reality of their Church. In many works Khomiakov is described as an important forerunner of the theological renewal in the Orthodox Church in the twentieth century (see Russian Religious-Philosophical Renaissance), or even as a Christian thinker who anticipated some ideas of the Second Vatican Council.

4 Ideological evolution and different continuations

Classic Slavophilism can be defined as a backward-looking utopianism, a Russian variant of conservative Romanticism, setting itself in opposition to the institutions and values of modern, capitalist civilization. Its utopian quality was not a matter for concern, since under the autocratic rule of Nicholas I no currents of Russian thought were allowed to be expressed in, and tested by, political activities. However, the loosening of political controls under Alexander II made it possible and necessary for the Slavophiles to pass from theory to practice. In these new conditions, Slavophilism was transformed into the right-wing variant of gentry reformism, on the one hand, and into Pan-Slavism, on the other. In both cases Romantic utopianism and emphasis on ’truly Christian’ values were replaced by an open commitment to nationalism, accepting, at least tacitly, the need for capitalist modernization of the Russian state.

The main Slavophile ideologists of the gentry-supported movement for reforms were Iurii Samarin and Aleksandr Koshelëv. They took an active part in the emancipation of serfs and because of this earned the reputation of ’liberals’, although their attitude towards Western liberalism remained intransigently hostile. The difference between them was that Samarin feared all forms of representative institutions on the national level while Koshelëv wanted an all-Russian Land Assembly with advisory power to be convened in Moscow, to serve as a counterpoise to the St Petersburg bureaucracy. Both men greatly contributed to the preservation of the village commune, defending it as a bulwark against proletarianization and as a supplier of cheap labour for the estates of the gentry. Samarin also became a leading ideologist of the russification policies in Russia’s ’western borderlands’ and in the Congress of Poland.

The leading figure in the transformation of Slavophilism into Pan-Slavism - an ideology providing ’Slav’ arguments for Russia’s territorial expansion - was Ivan Aksakov (younger brother of Konstantin Aksakov). He was an influential publicist, demanding the conquest of Constantinople and the establishment of a powerful, Russian-led Slav federation. His nationalism was anti-Semitic and violently anti-Polish, defining Poles as ’renegades of Slavdom’. More sophisticated and philosophically developed was the Pan-Slavism of Nikolai Danilevskii (1822-55), the author of *Russia and Europe* (1869). He grounded the ’Slav idea’ on a naturalistic foundation and developed an original theory of ’historico-cultural types’ of civilization, anticipating to a certain extent Spengler’s philosophy of history. But Danilevskii’s links with the founders of Slavophilism were very tenuous: he rejected their political Romanticism, especially the idea of a ’Christianization of politics’, claiming that Christian morality applied only to individuals, and that the relation of states and nations could only be based on self-interest and struggle for power (see Pan-Slavism §3).

The impact of Slavophile ideas on Russian thought was diverse and profound. The Slavophiles' views on the Russian peasant commune influenced Herzen and through him other ideologists of Russian populism. The Slavophile critique of the West was taken up by Dostoevskii and set the tone for all currents of Russian anti-Westernism. Slavophile philosophy, especially Kireevskii's 'integralism', was the main influence on Vladimir Solov'ëv in the first phase of his philosophical development; in later years he broke with the epigones of Slavophilism protesting against their nationalism and anti-Catholicism, but, none the less, continued to develop quite a few of their philosophical ideas. Russian religious philosophers of the early twentieth century had, as a rule, no doubts that Kireevskii and Khomiakov laid the foundation of a distinctively Russian tradition in philosophy.

See also: Enlightenment, Russian

ANDRZEJ WALICKI

References and further reading

Berdiaev, N. (1912) *Aleksei Stepanovich Khomiakov*, Moscow. (The classic work on Slavophile philosophy by a famous Russian philosopher of the early twentieth century.)

Bolshakoff, S. (1945) *The Doctrine of the Unity of the Church in the Works of Khomiahov and Moehler*, London. (A detailed analysis of Khomiakov's ecclesiology. The author compares Khomiakov's doctrine with the views of J.A. Moehler, a Romantic Catholic theologian from Tübingen.)

Christoff, P.K. (1961, 1972, 1982, 1991) *An Introduction to Nineteenth-Century Russian Slavophilism*, vol. 1, *A.S. Xomjakov*, The Hague; vol. 2, *I.V. Kireevskii*, The Hague; vol. 3, *K.S. Aksakov*, Princeton, NJ: Princeton University Press; vol. 4, *Iu.F. Samarin*, Boulder, CO: Westview Press.(The most comprehensive historical account of the Slavophile movement in any language.)

Danilevskii, N. Ia. (1869) *Rossiia i evropa (Russia and Europe)*, New York: Johnson Reprint Corporation, 1966. (Referred to in §4.)

Edie, J.M., Scanlan, J.P. and Zeldin, M.B. (eds) (1976) *Russian Philosophy*, vol. 1, 2nd impression, Knoxville, TN: University of Tennessee Press.(Representative selection from the works of I. Kireevskii, S. Khomiakov and K. Aksakov.)

Gleason, A. (1972) *European and Muscovite: Ivan Kireevskii and the Origins of Slavophilism*, Cambridge, MA. (Deserves special mention in the literature on Kireevskii.)

Gratieux, A. (1939) *A.S. Khomiakov et le mouvement Slavophile (A.S. Khomiakov and the Slavophile Movement)*, 2 vols, Paris.(The first comprehensive monograph of Slavophilism in a Western language examining Slavophile thought from a Catholic viewpoint.)

Leatherbarrow, W.J. and Offord, D.C. (eds) (1987) *A Documentary History of Russian Thought From the Enlightenment to Marxism*, Ann Arbor, MI: Ardis.(Includes selections from the prominent Russian Slavophiles.)

Müller, E. (1966) *Russischer Intellekt in europäischer Krise: Ivan V. Kireevskij*, Cologne.(A recommended work on Kireevskii.)

Raeff, M. (ed.) (1983) *Russian Intellectual History: An Anthology*, 2nd impression, Sussex: Harvester Wheatsheaf. (Collects writings that have helped shape the social and political consciousness of Russia.)

Walicki, A. (1975) *The Slavophile Controversy: History of a Conservative Utopia in Nineteenth-Century Russian Thought*, Oxford: Clarendon Press.(Analyses the Slavophile philosophy from a comparative European perspective; it deals also with the antecedents of Slavophilism, with the confrontation of Slavophilism and Westernism, and with the different continuations of the Slavophile thought.)

Slovakia, philosophy in

Until as late as 1918, social and national circumstances were not favourable to the development of philosophy in Slovakia. The enforced retardation of the country had an obviously negative impact on intellectual and cultural life, and stood in the way of the possibility of diversity. That is why the first important Slovak philosophers such as Bayer and Caban emerged only as recently as the seventeenth century. Following this, philosophers of the Enlightenment such as Karlovský, Laurentzy, Steigel and Feješ started to criticize the anti-scientific ideas which still survived. In the first phase of the National Revival Movement, Jan Kollár created a new philosophy of history by postulating the cultural unity of the Slavs. A specific contribution to this theory came from L. Štúr and his followers (Hurban, Hodža, Kellner) who applied it to practical conditions and stressed the necessity of national emancipation for the Slovak nation. Their influence is evident in all further developments of the national movement and its philosophy.

The end of the nineteenth century saw the replacement of this idea by the philosophy of Thomas Masaryk. Slovak cultural life and philosophy, however, only began to develop fully in the context of an independent Czechoslovak Republic after 1918. The mainstream philosophy of the time was rationalistic, evident in the thought of such thinkers as Sv. Štúr, Koreň and Osuský. Philosophy after 1945 was marked by the achievements of Hrušovský, his students and colleagues, whose initial form of neopositivism took on a Marxist dialectical structurology. The importance of Marxism has recently been gradually declining and the Slovak philosophic scene is independently evolving in an atmosphere more in line with other European countries, particularly since the formation of the Slovak Republic in 1993.

1 History and evolution

As early as the tenth century, Slovakia lost its political independence during the formation of new states in the Danube valley. Subsequently, until 1918, Slovakia was politically subject to Hungary and later was part of the Austro-Hungarian empire. Under these circumstances the Slovak nation had lived through centuries of almost uninterrupted national, economic and cultural oppression by foreign rulers. The feudal atmosphere retarded cultural development, as well as the spread of higher education. Many Slovaks thirsting for knowledge went abroad - many of these never returned.

At the end of the sixteenth century the situation began to change as waves of humanism and later-Reformation thought arrived in Slovakia. The first humanistic scholars from Slovakia are connected with this period. These worked abroad, however, and included Ján Sambocký of Trnava (d. 1584), translator of Plato’s dialogues, Vavrinec Benedikti of Nedožiery (d.1615), educational reformer and propagator of the ideas of Petrus Ramus, and Ján Jessenius (d. 1621), physician and supporter of Renaissance Neo-Platonist natural philosophy.

2 The seventeenth to nineteenth centuries

The Reform School (*kolegium*) in the East Slovak town of Prešov became the country’s first important philosophical centre. This school tried to overcome scholasticism by promoting Bacon’s inductive empirical methods, thereby hoping to achieve closer connection between philosophy and the natural sciences (see Bacon, F.). The leading representatives, who were influenced also by the works of Comenius, were Ján Bayer (d. 1674) and Izák Caban (d. 1707). At the Jesuit university in Trnava, centre of the anti-reform movement, some professors, including František Kéri (d. 1768), Andrej Jaslinský (d. 1784), and Ján B. Horvát (d. 1800), were impressed by the methods of current natural science and worked to free Slovak philosophy from orthodox neoscholasticism.

The Enlightenment reached its peak in Slovakia between 1789 and 1820. Here it showed some indications of compromise (particularly in relation to religion) when compared to its western European counterpart. Nevertheless, it also fulfilled an emancipatory function through figures such as Žigmund Karlovský (d. 1821), Ján Laurentzy (d.1819), Michal Steigel (d.1829), and Ján Feješ (d.1823).

National Revival, developing at first in parallel and in cooperation with similar Czech feelings, brought a change in the orientation of philosophy. Abstract metaphysics, ontology and logic were no longer at the forefront. Their place was taken by philosophy of history, an interest in theories of nationhood (in connection with philosophy of language, arts and culture) and, in general, philosophy conceived as the theory of particular social practices.

Reform tradition, German neohumanism and philosophers such as Herder, Fries and Hegel stimulated new thought. Theoreticians of the National Revival saw ideological support for their efforts in the ‘Slavonic’ emergence of Herderian-Hegelian philosophy of history as the highest degree of human development. Ján Kollár (d. 1852) built on this philosophy to produce his own idea of the mutual cultural evolution of Slavs. L’udovít Štúr (d. 1856) developed the idea in a Hegelian spirit by stressing the necessity of a specific emancipatory role for the Slovak nation, thus influencing the whole future course of the Slovak national liberation effort. Štúr’s contemporaries and followers put forward similar ideas, and this concept achieved a kind of Slavonic messianicism in the thought of some philosophers.

A change is evident in the second half of the nineteenth century with thinkers interested in the particular needs of national society. This work was undertaken by scientists such as D. Štúr, A. Stodola and J.A. Wagner on the one hand, while on the other there was later, from the beginning of the twentieth century, a group of intellectuals, including V. Šrobár, A. Štefánek and J. Lajčiak, gathered round the magazine *Hlas* (Voice) which was inspired by the philosophy of Masaryk.

3 Twentieth century philosophy

When Slovakia became part of the independent Czechoslovak Republic in 1918, a new situation arose. Slovak intellectual life flourished and became the potential repository of a new philosophical culture. Contact with the more progressive Czech philosophy became more intensive (Czech philosophers lectured at the newly founded university in Bratislava). Direct contacts with contemporary European thinkers also increased. The influence of Masaryk’s philosophy continued to be strong, particularly in the work of Svätopluk Štúr who also made use of ideas from Benedetto Croce and J.L. Fischer in his search for a synthetic type of rationalism. Classical positivism prevailed in the work of Jozef Koreň and Samuel Š. Osuský, ideologically influenced by the Protestant tradition.

An important event in Slovak intellectual life was the establishment in 1937 of the Association for Scientific Synthesis which brought together researchers from different scientific fields. This association emphasized the use of methodological principles taken from neopositivism and structuralism. The leading figure was Igor Hrušovský who, even after accepting the basic theses of Marxism after 1945, remained the dominant figure of Slovak philosophy, and created a new concept of dialectical structurology.

When the Communist Party regained power in 1948, a Marxist orientation was enforced and subsequently dominated public Slovak philosophy. Historians of philosophy such as E.Várossová, T. Münz, J. Bodnár, J. Kocka, M. Burica, M. Zigo and others began to free themselves from ideological oppression, and many published valuable works. These were followed by works of authors from other philosophical disciplines, and Slovak translations of philosophical classics achieved important success.

After 1968, in the period of the totalitarian regime, philosophers Milan Šimečka and Miroslav Kusý worked in the dissident movement. Their works, analysing the problems of the time, were published illegally or abroad. In the 1980s, the younger generation of Slovak philosophers began publishing the results of their research on the history of philosophy, the problems of phenomenology, existentialism, analytic philosophy, the Frankfurt School and also on postmodern trends. However, it was only after the fall of the totalitarian regime in 1989 that complete freedom of philosophical activity became possible.

See also: Czech Republic, philosophy in

JOSEF ZUMR

References and further reading

Bodnár, J. (ed.) (1988) *Geschichte des philosophischen Denkens in der Slowakei (History of philosohical thought in Slovakia)*, Bratislava: Veda.(Detailed history of earlier Slovak philosophy until 1900.)

Král, J. (1934) *La philosophie en Tchécoslovaquie (Philosophy in Czechslovakia)*, Prague: Bibliothèque des problèmes sociaux, 2.(Brief summary of historical development of philosophy in Bohemia, considering also Slovakia.)

Lobkowicz, N. (1961) *Marxismus-Leninismus in der ESR, Die tschechoslowakische Philosophie seit 1945*, Dortrecht: Reidel.(Detailed analysis of Czech and Slovak Marxism considering also other trends.)

Pichler, Tibor and Gašparíková, Jana (eds) (1993) ‘Language, Values and the Slovak Nation’, *Slovak Philosophical Studies I*, Cultural Heritage and Contemporary Change IV, Washington, DC: The Council for

Research in Values and Philosophy.(The history of the development of the sense of national identity, written by a team of authors.)

Zumr, J. (1971) 'Philosophie der Gegenwart in der Tschechoslowakei' (Contemporary philosophy in Czechoslavakia), in R. Klibansky (ed.) *Contemporary Philosophy*, Florence: La nuova italia editrice, 455-73. (Brief summary of Czech and Slovak thought, 1930-69.)

Smart, John Jamieson Carswell (1920-)

J.J.C. (Jack) Smart was born in England and studied at Glasgow and Oxford universities before moving to Australia to take up the Chair of Philosophy at Adelaide University. He was one of the earliest and most influential advocates of the mind-brain identity theory, the view that mental states are identical with brain states. He also played a major role in articulating and defending realism in science, the four-dimensional view of time and act utilitarianism in ethics.

1 Biography

J.J.C. Smart was born on 16 September 1920 in Cambridge, England. At Oxford he was influenced by Gilbert Ryle though he subsequently abandoned Rylean behaviourism in favour of a robust realism about mental states and scientific entities. He took up the Chair of Philosophy at Adelaide University in 1950, and moved to La Trobe University in 1972 before going to the Chair of Philosophy at the Institute of Advanced Studies at the Australian National University in 1976. He has made a number of visits to the United States and probably has had more philosophical influence there and in Australia than in Britain.

A distinctive feature of his writing is the clarity and directness with which he advances his views. This is a trait that he shares with another important Australian philosopher, David Malet Armstrong, and together they have influenced the way a generation of philosophers in Australia write, as well as influencing the doctrines they espouse.

2 Mind-brain identity theory

In a landmark article, ‘Sensations and Brain Processes’, first published in 1959, Smart defends the view (put forward earlier by his Adelaide colleague U.T. Place) that sensations are brain processes. Sensations are not correlated with or causally connected to, but literally identical with brain processes in the same way that lightning is identical with an electrical discharge. Later he extended the view to encompass intentional states like belief and desire and mental states in general.

An important feature of his version of the identity theory is his topic neutral analysis of sensation reports. He grants that when I report a sensation in a sentence like ‘I am in pain’ or ‘I am having a yellow after-image’, I am not making a claim about my brain as such. I may not even know that I have a brain. I am, he argues, making a claim roughly to the effect that something is going on in me like what goes on in me when a pin is stuck in me or a whip is applied to me or … (in the case of pain), and like what goes on in me when I see a lemon or a sunflower or … (in the case of a yellow after-image). The nature of these ‘somethings that go on in me’ is then a matter for empirical investigation, and his view is that they will turn out to be certain processes in the brain. With the benefit of hindsight we can now see his topic neutral analyses as a precursor of functionalism, for both see mental states as inner states whose mental nature is a matter of their relational properties (see Mind, identity theory of; Functionalism).

3 Realism, time and ethics

The view that mental states are brain states is a species of realism. Mental states are not ‘convenient fictions’ for the prediction of behaviour. Similarly, Smart argues that the sub-microscopic particles of physics are real. Electron theory is not merely a device for predicting the tracks in Wilson cloud chambers (see Fictionalism; Scientific realism and antirealism). Electrons rather are the causes of these tracks. Otherwise, Smart argues, the regular appearance of the tracks would have to be regarded as a cosmic coincidence. More generally, he argues that only the realist can explain why positing sub-microscopic particles should have such great instrumental value. Smart insists that the situation with electrons and macroscopic entities like tables and the sun is essentially the same: none is to be reduced to patterns of observations or patterns of experiences. They are all independently existing entities to be believed in as the best explainers of the patterns.

Smart views time as a fourth dimension akin to the three spatial ones, and sees this as the only view to hold in light of the Minkowski interpretation of relativity theory. Objects are extended in time as well as space, and accordingly have temporal parts as well as spatial ones. A tree’s growing is a matter of later temporal parts of the tree being

taller than earlier ones, and the problem of identity over time - of persons, trees, schools or whatever - becomes the problem of finding the right relation to unify different temporal parts to make up a single, temporally extended person, tree, school, or whatever. On this view the present has no special ontological status: what I call ‘now’ is simply the time of my calling, and the past is simply times before then, and the future times after then.

Smart’s position in meta-ethics is a version of noncognitivism: ethical statements are expressions of overriding attitudes, and words like ‘good’ and ‘right’ are terms of commendation, not terms that describe (see Emotivism). His position in normative ethics is an uncompromising version of act utilitarianism. The right act is that act out of those available to the agent that would produce the most happiness (or, better, has the greatest expectation of doing so). He famously criticizes rule utilitarianism - the view that the right act is the act in accord with the rule the following of which would produce the most happiness - as involving a kind of ‘rule worship’ inconsistent with utilitarianism’s guiding thought that it is outcomes that matter. In Smart’s view the value of keeping promises, distributing goods equitably and not punishing innocent people, is instrumental only. Thus, in (arguably rare) cases where punishing someone innocent would have good results - perhaps it is generally believed the prisoner is guilty so that punishing him would have a deterrent effect on potential wrongdoers - it is straighforwardly right to do so (see Utilitarianism).

See also: Australia, philosophy in

FRANK JACKSON

List of works

Smart, J.J.C. (1961) *An Outline of a System of Utilitarian Ethics*, Melbourne: Melbourne University Press. Revised edn in J.J.C. Smart and B. Williams, *Utilitarianism: For and Against*, Cambridge: Cambridge University Press, 1973, 1-74.

(1963) *Philosophy and Scientific Realism*, London: Routledge & Kegan Paul.(One chapter of this book is largely based on ‘Sensations and Brain Processes’.)

(1968) *Between Science and Philosophy*, New York: Random House.

(1984) *Ethics, Persuasion and Truth*, London: Routledge & Kegan Paul.

(1987) *Essays Metaphysical and Moral*, Oxford: Blackwell.

(1989) *Our Place in the Universe*, Oxford: Blackwell.

Smart, J.J.C. and Haldane J. (1996) *Atheism and Theism*, Oxford: Blackwell.

References and further reading

Pettit, P. *et al*. (eds) (1987) *Metaphysics and Morality*, Oxford: Blackwell.(Good collection of papers on Smart’s work, with a list of his publications up to 1986.)

Smith, Adam (1723-90)

Despite his reputation as the founder of political economy, Adam Smith was a philosopher who constructed a general system of morals in which political economy was but one part. The philosophical foundation of his system was a Humean theory of imagination that encompassed a distinctive idea of sympathy. Smith saw sympathy as our ability to understand the situation of the other person, a form of knowledge that constitutes the basis for all assessment of the behaviour of others. Our spontaneous tendency to observe others is inevitably turned upon ourselves, and this is Smith's key to understanding the moral identity of the individual through social interaction. On this basis he suggested a theory of moral judgment and moral virtue in which justice was the key to jurisprudence. Smith developed an original theory of rights as the core of 'negative' justice, and a theory of government as, primarily, the upholder of justice. But he maintained the political significance of 'positive' virtues in a public, non-governmental sphere. Within this framework he saw a market economy developing as an expression of humanity's prudent self-interest. Such self-interest was a basic feature of human nature and therefore at work in any form of society; but commercial society was special because it made the pursuit of self-interest compatible with individual liberty; in the market the poor are not personally dependent upon the rich. At the same time, he recognized dangers in commercial society that needed careful institutional and political management. Smith's basic philosophy is contained in The Theory of Moral Sentiments (1759), but a major part concerning law and government was never completed to Smith's satisfaction and he burnt the manuscript before he died. Consequently the connection to the Wealth of Nations (1776) can only be partially reconstructed from two sets of students' notes (1762-3 and 1763-4) from his Lectures on Jurisprudence at Glasgow (Smith [1762-6] 1978). These writings are complemented by a volume of essays and student-notes from lectures on rhetoric and belles-lettres.

Although a philosopher of public life and in some measure a public figure, Adam Smith adhered to the Enlightenment ideal of privacy to a degree rarely achieved by his contemporaries. He left no autobiographical accounts and, given his national and international fame, the surviving correspondence is meagre. The numerous eyewitness reports of him mostly relate particular episodes and individual traits of character. Just as there are only a few portraits of the man's appearance, there are no extensive accounts of the personality, except Dugald Stewart's 'Life of Adam Smith' (1793), written after Smith's death and designed to fit Stewart's eclectic supplementation of common sense philosophy. While Smith was a fairly sociable man, his friendships were few and close only with men who respected his desire for privacy. David Hume was pre-eminent among them.

1 Life

Adam Smith's date of birth is unknown, but he was baptized in Kirkcaldy in Fife, Scotland, on 5 June 1723. His father, a customs officer, died some months earlier and Smith was brought up as an only child, a circumstance shared with several contemporary Scots literati, as has been emphasized in recent psycho-history (Camic 1983). Smith evidently remained close to his mother., and some will see this as an explanation for the fact that his attraction to female company never stretched to marriage. Otherwise, we have little evidence of the nature of Smith's intimate life.

As was common, Smith attended the local parish school and then went to Glasgow University in 1737, where he was taught moral philosophy by Francis Hutcheson. In 1740 he went to Balliol College, Oxford, lingering there for six years. He compensated for the lack of organized education - later condemned in the *Wealth of Nations* (1776; henceforth *WN*) - by extensive private study, mainly in Greek, Latin and French literature, turning himself into an extremely well-read man. From 1746-8 Smith seems to have stayed with his mother in Kirkcaldy; but during the following three winters he gave public (non-university) lectures in Edinburgh, sponsored by Henry Home (later Lord Kames) and his circle. The lectures were on rhetoric and belles-lettres, to which was eventually added a series of lectures on jurisprudence (published subsequently as *Lectures on Jurisprudence* ([1762-6] 1978; henceforth *LJ*)). This performance had the desired effect; in 1751 Smith was appointed professor at Glasgow University, first in logic and, after one year, in moral philosophy. Until his resignation in the middle of 1763-4, Smith gave the basic course in moral philosophy that had been founded by Gershom Carmichael in the 1690s and developed by Francis Hutcheson in the 1730s and 1740s. It encompassed natural religion, ethics and jurisprudence. In addition, Smith gave an advanced class on rhetoric and belles lettres. Hardly a trace of Smith's lectures on

natural religion has survived, but those on ethics formed the basis for his first major work, the *Theory of Moral Sentiments* (1759; henceforth *TMS*). Two student-reports on the lectures on jurisprudence, which include economics, have survived from 1762-3 (incomplete) and 1763-4 (henceforth *LJ*(*B*) and *LJ*(*A*), respectively); see Smith 1762-6: there is also the fragment of a report from earlier in Smith's career ('Anderson Notes'). Finally, there is a student's report on the lectures on rhetoric and belles-lettres from 1762-3 (henceforth *LRBL*; see Smith 1762-3).

Early in 1764 Smith became travelling tutor to the Duke of Buccleuch, a lucrative post that afterwards gave him a pension and an influential connection for life. Travelling in France and Switzerland until 1766, Smith met many of the French Enlightenment's leading thinkers, notably François-Marie Arouet Voltaire and the circle of physiocratic economists led by François Quesnay and Anne Robert Jacques Turgot. Smith admired Voltaire, especially as a dramatist; and he shared the conceptual, empirical and normative concerns of physiocracy, though he reached different conclusions about these matters. Like the physiocrats, he understood the relationship between economics and civil society by postulating ideal-typical stages of social development; and like them, he rejected the common mercantilist analysis of wealth in terms of money and advocated a free economic system to create such wealth. Smith, however, rejected the physiocratic idea of land as ultimate source of wealth.

Smith had begun working on these matters during his Glasgow years, as *LJ* and the 'Early Draft' of the *WN* (reprinted in *LJ*) show. But it took him nearly another decade to develop his own grand theory. He worked until 1773 in seclusion at his mother's house in Kirkcaldy and then spent three years in London revising his manuscript. The *WN* appeared in the spring of 1776 to resounding praise from David Hume and Edward Gibbon, among many, and quickly turned the distinguished moral philosopher into a famous political economist.

1776 was in another respect a turning point for Smith, for in the late summer David Hume died. This led Smith publicly to indicate his religious outlook more clearly than before. He was already assumed, as we see from the cool attitude of James Boswell (once Smith's student) and Samuel Johnson, to be a sceptic concerning revealed religion. But when Smith publicly maintained that the infidel Hume had approached 'as nearly to the idea of a perfectly wise and virtuous man, as perhaps the nature of human frailty will permit', he was widely taken to have declared his own atheism (1777: xlix). In fact, Smith never made explicit his religious beliefs. Much of his work appears to be in the mould of common deism, but often it is unclear whether Smith is analysing deism as a psychological disposition or whether he is accepting it as a doctrine. In the end his obdurate silence suggests he had accepted the basic lesson of agnosticism. He seems cooler toward religion as he gets older, and his 'obituary' for Hume is a milestone on this road.

In 1777 Smith was appointed Commisioner of Customs in Edinburgh, a lucrative office that he discharged conscientiously until his death. In addition to being in public office, Smith was now a public man of letters, sought by government and politicians for advice on matters of state and policy, such as relations with America and Ireland, trade and taxation. He had been elected to the Royal Society and to the Johnsonian 'The Club' during his stay in London in 1773-6. When he paid a rare visit to London, he was received in the literary circles around the Oyster Club, and although he never ventured abroad again, he remained a name in the Parisian salon-world, where he had been so well received in the '60s. He was also part of the cultural life of the high Enlightenment in Edinburgh, sought out by literary tourists and opening his house to a circle of friends once a week (see Enlightenment, Scottish). True to his suggestion that a citizen militia served a moral purpose, he was an officer in the Edinburgh militia. Smith lived with his mother, his cousin, Janet Douglas, and a nephew of the latter, David Douglas, whom Smith made his heir.

During these years Smith kept writing. He revised the *TMS* several times, including a major recasting (6th edn 1790) shortly before his death. The *WN* was revised for the second and, especially, the third editions (1778 and 1784) and saw two more life-time editions. But Smith also undertook new projects. One was a 'sort of theory and history of law and government', announced already in the first edition of *TMS* and still aspired to in the last edition of that work. The *WN* had taken account of 'police, revenue and arms', but the part concerning justice, the theory of jurisprudence, was still lacking (*TMS*, Advertisement: v). Smith's other project was 'a sort of Philosophical History of all the different branches of Literature, of Philosophy, Poetry and Eloquence' (Smith 1977: 287). A few days before he died, Smith requested his sixteen manuscript volumes to be burnt. We can form some idea of the former project from the lectures on jurisprudence and parts of *TMS* and *WN*; the latter was obviously related to the

early *Essays on Philosophical Subjects* (henceforth *EPS*), published posthumously in 1795, and to the Glasgow lectures on rhetoric and belles lettres. But the systematic coherence of Smith's work is a matter of reconstruction - some scholars even deny that it has such coherence, despite Smith's declared intentions. Smith died in his house in the Canongate on 17 July 1790 and is buried in the Canongate cemetry.

2 Imagination: scientific and moral order

The single most important influence on Adam Smith was David Hume's philosophy, especially the theory of the imagination as the active mental power that fashions a specifically human world within nature (see Hume, D.). Like Hume, Smith saw imagination as that which enables us to create connections between the perceived elements of both the physical and the moral world, ranging from particular events and things to the cosmos and the system of humanity. For Smith there are two fundamentally different kinds of imagination, one concerning other persons as agents and one concerning things - human beings included - and events. The former is the basis for personal identity and sympathy (in Smith's special sense) and, thus, for the moral world. The latter is the basis for all theoretical activity of the mind, including science and the arts. The activity of the imagination is a spontaneous search for order and coherence in the world; satisfaction of it carries its own pleasure, while frustration brings 'wonder and surprise' and, if prolonged, anxiety and unease. Smith talks of this imaginative striving in aesthetic terms as a concern with beauty and harmony.

The 'theoretical' imagination is invoked for a number of different explanatory purposes. It accounts, according to Smith, for our ability to order things and events so that we can orient ourselves in life. It explains the 'aesthetics' of ordinary living, such as our tendency to order things for no other purpose than the order and arrangement that please, and our desire for machinery, gadgets and other 'systems'. Works of art, as well as of technology, are works of imaginative order. Not least, philosophy and science are products of the imagination's attempt to create order in the chaos of experience. This is expressed in the recurring machine analogies for the natural world and for society; and it is reflected in the human mind's tendency to underpin the perceived orderliness of the world by assuming an orderer with a purpose. In other words, science, deism and natural providence are all parts of the explanatory web that the imagination creates to satisfy its desire for order.

Within this framework Smith writes his remarkable histories of science, notably the essay on the history of astronomy. The basic thesis is that empirical evidence can only play a role as evidence if it fits into an orderly system of beliefs. Smith exhibits a sceptical distance in relation to all absolute truth-claims in both science and religion, epitomized in the famous conclusion to 'The History of Astronomy' in *EPS*:

> [Newton's] principles, it must be acknowledged, have a degree of firmness and solidity that we should in vain look for in any other system. The most sceptical cannot avoid feeling this… And even we, while we have been endeavouring to represent all philosophical systems as mere inventions of the imagination, to connect together the otherwise disjointed and discordant phaenomena of nature, have insensibly been drawn in, to make use of language expressing the connecting principles of this one, as if they were the real chains which Nature makes use of to bring together her several operations.
>
> ([1795] 1980, IV.76: 105)

When we introduce order into our observations of other people, as opposed to the rest of nature, our imagination takes on a special quality. We are able to 'make sense' of the behaviour of others through imaginative identification - what Smith calls sympathy, a central concept in his philosophy. 'Sympathy'here has no evaluative connotations; it does not mean that one person accepts the motives and actions of another as good and right, nor is it a motive for action. Sympathy is the ability to see the point of view of another person so as to be able to accept or reject it as appropriate in *that* person's situation. This native ability is the key to Smith's social theory of the self and, through that, to his theory of the moral world.

Inclination and need lead people to interact; interaction depends on observation and appreciation of others and their situation, and this is facilitated by the universal tendency sympathetically to adopt the position of the other and compare reactions. Observation of others will bring awareness that oneself is the object of other people's observation and assessment. Society is the mirror in which individuals see their own nature. Further, in observing that one is being observed one has the basis for such mutual adjustment of behaviour as is necessary for all social living. Each individual is generally led to pre-empt this assessment by others through self-observation and

self-assessment. In short, people internalize the spectator, and the internal spectator has the force to prompt such adjustment of behaviour as would otherwise be demanded by external spectators in order to satisfy the inclination to, or the need for, agreement or conformity.

The process of mutual adjustment through the sympathetic search for a common standpoint often fails, of course, leading to moral and social disorder. Thus we are led to imagine an ideal judgment and an ideal judge, transcending the limitations of knowledge, bias, and so forth of those actually involved. Once we establish a dialogue with this imagined ideal of an impartial spectator, we have a moral conscience.

With this social theory of moral personality Smith rejects the idea of Hutcheson and others that moral agency hinges on a special moral sense, offering instead explanations based on empirical features of the mind. Like Hume, Smith rejects the suggestion of Samuel Clarke and others that moral judgment and motivation are forms of rational inference. And he ignores religious ideas of conscience as infusion (inspiration) by the deity or a response (fear) to the deity's might perceived. For Smith, formation of the moral self begins with others.

3 Moral theory

By means of this theory of the self, Smith produces a complex analysis of moral life. Moral judgment is inevitably an assessment of propriety - first of the motives of an agent in a given situation, then of the motives of the person who is the 'object' of the agent's actions. But our inclination towards orderliness leads us to classify actions just as our search for impartiality suggests general and impersonal points of view; and on this basis we develop general rules of behaviour which, once given social acceptance, significantly influence our assessment of the propriety of particular motives and actions. Sympathy ties us to the particularity of the situation, while the impartial spectator calls for the generality of rules. Similarly, while we have great difficulty making a moral assessment of motives and associated behaviour on the basis of their outcome, we tend to allow utility and disutility to influence our perception of propriety.

In a different mode from that of the analysis of moral judgment, Smith gives a complex account of moral virtue. At one level he looks on the moral and social ideas that make the world go round from the elevated standpoint of the stoic sage (see Stoicism §19). Smith was attracted by the stoic ideal of tranquility as the end of moral life; yet his account of moral psychology showed everything distinctive about the life of the human species to be due to people's inability to live in tranquility. The exercise of our productive powers, as portrayed in the *WN* and in the *TMS* as emulative vanity, is only the most dramatic illustration of an inescapable restlesness pervading our lives. In Smith's view a dialectic tension between tranquility and activity is bound to be a permanent feature of human life, and the implication is clearly that it would be futile for the philosopher to attempt to defend the one over the other.

At another level, Smith looks at morality from an historical point of view, and combines this with a theory of the cardinal virtues. The features of personality and courses of action people approve and disapprove of in others and in themselves vary from one culture and period to another. Accordingly, a major task for philosophy is to look at humanity historically and comparatively. This reveals that in one sphere of human response there is a great deal that is common to all humankind, namely in the sphere of *moral* approval and disapproval (as opposed to, say, the aesthetic or religious response). The moral philosopher can identify a number of basic virtues and vices, the tone and composition of which may vary significantly, but which are nevertheless universally recognizable and comparable.

One way of organizing this field is by means of the cardinal virtues, though Smith, like Hume, is more interested in distinguishing 'positive' from 'negative' virtues, - benevolence from justice. The distinction is based upon spectator reactions. When the spectator, whether the actual or the imagined impartial one, sympathetically enters into the situation of an agent, the result is approval or disapproval of the agent's judgment and action. When the agent tries to promote the good of someone, whether of self or of other, the spectatorial approval or disapproval tends to vary from person to person. For while we tend to agree on what is good in broad outline, we have great difficulty agreeing on what is good for particular persons in specific situations, unless we are connected in some moral community. By contrast, we tend to agree on what is harmful not only in general but in each case, and the pattern of reaction to harmful behaviour therefore has a high degree of uniformity, known as resentment.

4 Law and politics

The negative virtue of avoiding harm or injury is, according to Smith, *justice*, which is the foundation of law and the subject of jurisprudence. The positive virtues encompass the life of the private sphere, but, in addition, they are the basis for a public sphere that is neither political nor administrative in modern terms. By maintaining a system of justice, the government protects both private life and the non-political public sector of human activity, and one of the most striking features of the modern world of commerce is, in Smith's eyes, that by doing so government has made it possible for economic behaviour to be transformed from private to public.

The personal attributes and actions that are protected in each person when others show them justice, that is abstain from injuring them, are their rights. A *right* is a sphere of freedom to be or do or have something that the individual can maintain against all others because the spectatorial resentment towards infringement of this sphere is so strong that it has been institutionalized in the form of the legal system. This line of argument puts Smith in a tradition of thinking about rights that goes back to Grotius, Hobbes and some religious covenanters. In this tradition the primary moral characteristic of the individual is self-assertion *vis-à-vis* the rest of humanity, so that all common or social morality arises through 'negotiation' between conflicting claims. Natural *law* is a secondary concept to that of *right,* in contrast to the ideas of the mainstream of thinking about natural law. While Hume was in some respects close to this tradition, he never found a way of accommodating the concept of rights within his sentimentalist theory of morality. This was left to Smith's theory of the spectatorial regulation of our moral sentiments (see Natural law; Rights).

'Rights', 'injury', and 'personality' are linked. The imagination depends on social experience and hence varies from one stage of society to another; consequently the idea of moral personality must vary. This is the core of Smith's historical approach to justice and law. Smith rejects the idea of a state of nature as a device for understanding human nature. The moral life of the species is unavoidably social, since only the social mirror lends us humanity. All consideration of our moral characteristics must therefore include the social setting; this applies not least to rights as the primary characteristic. Even so, certain minimal rights are universal to all social living. A social group is only viable if it recognizes rights to physical, moral and some kind of social personality; these may accordingly be considered universal, 'natural' rights. Beyond this minimum, we have to look to the historical circumstance in order to understand rights.

In Smith's view, humanity can be divided into four broad stages of social development according to the extension of the concept of the person and, consequently, the scope of rights recognized. Hunters and gatherers recognize little beyond what immediately sustains the physical and moral person (food, shelter, personal freedom and social recognition). Dramatic extensions of personhood are produced by nomadic 'shepherds', with the recognition of property in food and tools much beyond what is required for each individual and their immediate dependents; and, further, in the agricultural stage, with recognition of property in land. The most abstract extension of the concept of personality occurs in commercial society, with the full development of contractual entitlements and ownership of purely symbolic property (paper money, credits) as parts of what a person is.

Each of these developments requires stronger government to protect its new rights, and, accordingly, the four-stages theory also accounts for the growth of government and law. However, the four stages are not to be understood as the actual steps of the past; they are 'ideal types'. The historical past has deviated greatly from this conjectural model, mainly due to forceful and tyrannical behaviour by rulers and conquerors. The modern world was thus, for Smith, decisively shaped by the emergence of commerce before agriculture was properly developed, due largely to an alliance of monarchs and city-burghers forged for the purpose of breaking the power of feudal nobility. The resulting alliance of modern government with commerce and finance is the unholy alliance targeted in *WN*.

The *WN* is the greatest working-person's tract ever written. Central to its argument is that in the modern world even the working poor can enjoy personal liberty because the modern economy enables them to sell their labour without selling themselves. With the deepening of the division of labour, each piece of work becomes more 'abstract', less tied to personal abilities, and more easily assessed in monetary terms. The labourer can therefore sell his labour-power to anyone, without the personal dependence of traditional service-relationships. But this will only be effective if the labourer's freedom to sell his labour is protected by government as a right, against traditional monopolistic restrictions.

The liberation of labour from 'servility' is central to the transformation of productive life from the private, familial sphere to the public market, from household oeconomy to political economy. This can only take place when the immoral exaggeration of the virtue of prudent self-interest, namely avarice, can be rendered innocuous in the eyes of society. This happens when the problem of distribution begins to be solved through the market. In pre-commercial society, owners can only use their property by consuming it, and this they can do only by maintaining dependents, typically in the extended family. In commercial society they can spend their riches via the anonymous other of the public market. Accumulation of wealth becomes not only compatible with but dependent upon the freedom of the poor, which leads to a distribution of goods that, while unintended, is as equal as is realistic among humanity.

> [The rich] are led by an invisible hand to make nearly the same distribution of the necessaries of life, which would have been made, had the earth been divided into equal portions among all its inhabitants, and thus without intending it, without knowing it, advance the interest of the society, and afford means to the multiplication of the species.
>
> ([1759] 1976, IV.1.10: 184-5)

There is a price for this freedom, namely the danger of moral corruption. The rich are constantly tempted to undermine the system that creates their wealth by seeking protection against competition. And the poor are exposed to the stupification of mechanical work, rendering them incapable of taking charge of their lives and making them useless as citizens and soldiers. In both areas protection is to be sought from government; in the former through anti-monopoly legislation and policy, in the latter through basic education funded in part by the public and through freedom of religious worship, both of which tend to establish a moral community of spectators within which the labourer can develop a moral character.

In addition to the economy that springs from self-interest, and the system of law that institutes the virtue of justice, Smith thus operates within a purely political sphere. Politics concerns 'police, revenue and arms', which means various municipal services and major construction works that the market cannot provide, and which facilitate commerce, plus defence. But under 'revenue' Smith also includes educational and cultural policies that are concerned with 'positive' virtues, or at least with facilitating such virtues. Smith is wary of stretching this function of government too far, because of the danger of tyranny in enforcing specific ideas of the good life. This, however, is not the limit of the role of positive virtue in public life: it extends also into an important public, non-political sphere.

Smith has much to say about the importance of a wide variety of social virtues - liberality, probity, generosity, courage, leadership, indeed justice in a distributive sense; or 'public spirit', as he often sums it up. But while these are virtues associated with public office, they are not the virtues that provide the very rationale of government in the way that justice does. Nor are these desirable qualities simply the private virtues of public figures, though they are certainly also that. Integral to Smith's society was a public sphere that was in some sense political, yet was not governmental. This was the world of the local elite, lairds and landlords with significant leadership in local affairs that today are matters of public policy. At the pinnacle of this world were the members of Parliament, whose work overwhelmingly consisted in dealing with private members' bills - often concerned with local matters - rather than with the business of executive government. The proper basis for these public roles was provided by the kind of positive virtues indicated above. While Smith wanted to exclude, or at least severely limit, these virtues in the function of government proper, he never doubted their essential role in a non-political public sphere. Today this sphere has largely been subjected to politics in a development that Hegelians might think of as the politicization of civil society; but it is anachronistic to use this sense of civil society in reading Smith: the public virtues and behaviour he is concerned with are not defined *vis-á-vis* the state.

Smith was not simply an advocate of a particular political and economic scheme. From his Humean perspective, there was little point in advocating things unless they were within the limits of people's imagination. The task was to extend that imagination by informing it of the full complexity of its situation. People who are thus informed may become impartial in judging their society. Smith had educated himself to be an impartial spectator who saw the injuries or injustices done to classes of people in the name of political exigencies that had long since passed. An appeal to the universal virtue of negative justice and an explanation of how and why it was being infringed in this particular historical situation was therefore the backbone of his history and theory of law and government; but he

did not thereby reduce all moral life to the administration of justice. At his best he combined the psychological perspicuity of Jane Austen, the historical richness of Gibbon and the philosophical acumen of Hume.

See also: Economics and ethics; Economics, philosophy of; Justice; Market, ethics of the; Moral sense theories; Property; Work, philosophy of

KNUD HAAKONSSEN

List of works

Smith, A. (1759) *The Theory of Moral Sentiments*, ed. D.D. Raphael and A.L. Macfie, Oxford: Clarendon Press, 1976, vol. 1.(The Glasgow Edition of the Works and Correspondence of Adam Smith.)

Smith, A. (1762-3) *Lectures on Rhetoric and Belles-Lettres*, ed. J.C. Bryce, Oxford: Clarendon Press, 1983; vol. 4.(The Glasgow Edition of the Works and Correspondence of Adam Smith.)

Smith, A. (1762-6) *Lectures on Jurisprudence*, ed. R.L. Meek, D.D. Raphael and P.G. Stein, Oxford: Clarendon Press, 1978, vol. 5.(The Glasgow Edition of the Works and Correspondence of Adam Smith.)

Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, ed. R.H. Campbell and A.S. Skinner, Oxford: Clarendon Press, 1976, vol. 2.(The Glasgow Edition of the Works and Correspondence of Adam Smith.)

Smith, A. (1777) 'Letter from Adam Smith, LL.D. to William Strahan, Esq.', Appendix to *The Life of David Hume, Esq., Written by Himself*, London: Strahan & Cadell; repr. in D. Hume, *Essays Moral, Political and Literary*, ed. E.F. Miller, Indianapolis, IN: Liberty Classics, 1985, xxxi-xli.

Smith, A. (1795) *Essays on Philosophical Subjects*, ed. W.P.D. Wightman and J.C. Bryce, Oxford: Clarendon Press, 1980, vol. 3.(The Glasgow Edition of the Works and Correspondence of Adam Smith.)

Smith, A. (1977) *Correspondence of Adam Smith*, ed. E.C. Mossner and I.S. Ross, Oxford: Clarendon Press; vol. 6.(The Glasgow Edition of the Works and Correspondence of Adam Smith.)

Smith, A. (1976) 'Anderson Notes', in R.L. Meek, 'New Light on Adam Smith's Glasgow Lectures on Jurisprudence', *History of Political Economy*, 8: 439-77.(From John Anderson's Commonplace Book, Andersonian Library, University of Strathclyde, volume 1.)

References and further reading

Brown, V. (1994) *Adam Smith's Discourse: Canonicity, Commerce and Conscience*, London: Routledge.(Uses the literary theory of Bakhtin in an analysis that suggests less systematicity in Smith's work than argued in other recent work.)

Camic, C. (1983) *Experience and Enlightenment: Socialization for Cultural Change in Eighteenth-century Scotland*, Edinburgh: Edinburgh University Press.

Campbell, R.H. and Skinner, A.S. (1982) *Adam Smith*, London: Croom Helm.

Campbell, T.D. (1971) *Adam Smith's Science of Morals*, London.(Stresses the scientific intention of Smith's work.)

Copley, S. and Sutherland, K. (eds) (1995) *Adam Smith, The Wealth of Nations: New Interdisciplinary Essays*, Manchester: Manchester University Press.

Cropsey, J. (1957) *Polity and Economy: An Interpretation of the Principles of Adam Smith*, The Hague: Nijhoff. (Smith in the tradition of modern political thought from Machiavelli onwards.)

Fitzgibbons, A. (1995) *Adam Smith's System of Liberty, Wealth and Virtue: The Moral and Political Foundations of the Wealth of Nations*, Oxford: Clarendon Press.(The ethics of the economics.)

Haakonssen, K. (1981) *The Science of a Legislator: The Natural Jurisprudence of David Hume and Adam Smith*, Cambridge: Cambridge University Press.(Reconstruction of Smith's philosophical system with jurisprudence at the centre.)

Haakonssen, K. (1996) *Natural Law and Moral Philosophy: From Grotius to the Scottish Enlightenment*, Cambridge: Cambridge University Press.(Smith's theory of rights (chapter 4) and the role of his project in the later Scottish Enlightenment.)

Hollander, S. (1973) *The Economics of Adam Smith*, Toronto, Ont.: Toronto University Press.(Difficult for non-economists.)

Jones, P. and Skinner, A.S. (eds) (1992) *Adam Smith Reviewed*, Edinburgh: Edinburgh University Press.

Justman, S. (1993) *The Autonomous Male of Adam Smith*, Norman, OK: University of Oklahoma Press.(The tension between the femininity of virtue and the masculinity of stoicism.)

Malloy, R. P. and Evensky, J. (eds) (1994) *Adam Smith and the Philosophy of Law and Economics*, Amsterdam: Kluwer.

Minowitz, P. (1993) *Profits, Priests and Princes: Adam Smith's Emancipation of Economics from Politics and Religion*, Stanford, CA: Stanford University Press.(Thesis suggested by subtitle.)
Mizuta, H. and Sugiyama, C. (eds) (1993) *Adam Smith: International Perspectives*, New York: St Martin's Press.(The international reception of Smith.)
Muller, J.Z. (1993) *Adam Smith in His Time and Ours: Designing the Decent Society*, New York: The Free Press.(General, introductory survey.)
Rae, J. (1895) *Life of Adam Smith*; repr. with additional material by J. Viner, New York: Kelley.(Standard biography for the past century.)
Raphael, D.D. (1985) *Adam Smith*, Oxford: Oxford University Press.(Introductory survey of the whole of Smith's life and works.)
Ross, I.S. (1995) *The Life of Adam Smith*, Oxford: Clarendon Press.(General biography.)
Scott, W.R. (1937) *Adam Smith as Student and Professor*, Glasgow: Jackson.
Shapiro, M.J. (1993) *Reading 'Adam Smith': Desire, History and Value*, London: Sage.
Skinner, A.S. (1979/1996) *A System of Social Science: Papers Relating to Adam Smith*, Oxford: Clarendon Press.(Stresses the systematic coherence between the *WN* and the rest of Smith's work.)
Skinner, A.S. and Wilson, T. (eds) (1975) *Essays on Adam Smith*, Oxford: Clarendon Press.
Stewart, D. (1793) 'Account of the Life and Writings of Adam Smith LL.D.', in D. Stewart, *Collected Works*, ed. Sir William Hamilton, 1854-60, 11 vols.; facsimile repr. with an introduction by K. Haakonssen, Bristol: Thoemmes, 1994, vol. 10, 1-98.
Teichgraeber, R.F. (1986) *'Free Trade' and Moral Philosophy. Rethinking the Sources of Adam Smith's Wealth of Nations*, Durham, NC: Duke University Press.(Close study of links to Hutcheson and Hume.)
Werhane, P.H. (1991) *Adam Smith and His Legacy for Modern Capitalism*, New York: Oxford University Press.
Winch, D. (1978) *Adam Smith's Politics: An Essay in Historiographic Revision*, Cambridge: Cambridge University Press.(The nature of Smith's *political*, as opposed to economic, theory.)
Winch, D. (1996) *Riches and Poverty: An Intellectual History of Political Economy in Britain, 1750-1834*, Cambridge: Cambridge University Press.(Approaches Smith via the debate on the connections between luxury and inequality, stressing the realistic qualities of his science of the legislator.)
Wood, J.C. (ed.) (1983-84) *Adam Smith: Critical Assessements,* London: Croom Helm, 4 vols.

Social action

Most of our actions take place in a social context and are, accordingly, in one way or another, dependent on the existence of other persons and their relevant actions, social institutions, conventions, or the like (for example, saluting, voting, drawing money from one's bank account, using lipstick, buying something). But people also perform actions jointly or collectively, to achieve some joint goal. Thus they may jointly sing a duet, play tennis, build a house, or conserve energy. This is collective social action in its most central sense. Such action is based on the participants' mutually known joint intention ('joint plan') to perform it. In weaker kinds of collective social action the participants are interdependent - as to their actions or thoughts - in some other ways.

This entry shall concentrate on collective social action, indeed on the central case: joint action based on joint intentions. Thus, you and I may form the joint intention to paint the house together and agree upon some method of doing it, for instance, that you paint the front while I paint the back of the house. We also agree upon which paint to use and other similar things. Carrying out our joint intention, we come to paint the house jointly. This kind of intentional joint action can be characterized more generally as follows, taking a joint action to be an action divisible into single-agent parts, that generate a purported outcome: the participants have formed a joint plan for a joint action; the plan is taken to involve a relevant joint intention, entailing for each participant the intention to perform their part of the joint action. The participant is assumed to believe that the various conditions for the success of the joint action will be fulfilled at least with some probability, and also to believe that this is mutually believed by the participants. The performance of a joint action can be regarded as agreement-based if the plan has been accepted by the participants and if they have appropriately communicated their acceptances to the others so that a joint obligation to perform the joint action has come about. Joint action in the fullest sense can be argued to be based on either explicit or implicit agreement (in a wide sense) (see Intention).

Over and above the standard sense of joint action based on the participants' joint intention we may also speak of joint action in a somewhat wider sense not necessarily involving joint intention. For example: some persons, seeing a car starting to slide down a hill, together start pushing it up the hill. I am passing and form the intention to push the car together with the others, an intention with a collective action as its content, and take part in the pushing. There may even be a belief among all the pushers that I also am participating. All this can take place without these car-pushers (or at least all of them) having a joint intention (joint plan) to push the car together with the others. So understood, we do not have here more than joint action in a wide sense - namely joint activity based on shared collective intention but not full-blown joint intention. The participants' intentions, when the participants are mutually aware of them, can still be regarded as social attitudes ('we-attitudes' in the sense of Tuomela 1994). Ideally, a person has such a social attitude (for example, an intention to do something together) relative to their group if and only if they (1) have (or share) this attitude, (2) believe the group members have it, and also (3) believe that there is a mutual belief that the members have this attitude. The broadest sense of joint action can be taken to be collective action performed because of a shared social attitude.

Joint action in this wide sense excludes such 'mere co-action' as people acting simultaneously in the direction of the same goal but (possibly) independently of each other (for instance, Max Weber's example of people simultaneously opening their umbrellas when it starts to rain). Such co-action does not deserve to be called social action at all.

We can, however, include among social action the following two kinds of collective action with social features - although they do not quite amount to collective social action in our previous sense:

(1) Collective action which is merely accompanied by (but not performed because of) a social attitude (we-attitude). The participants may be performing an action (for example, selling shoes at a department store) either separately or 'together' as their parts of a collective project (or believed project), planned or structured by them or by somebody else.
(2) Collective action consisting of separate individual actions directed towards and performed because of a shared, possibly merely personal goal (the realization of which normally requires several agents' participation).

In the literature on collective action either social action in the strict sense or cases of the kind (1) and even of the

kind (2) may be involved. In those cases there is often a public good to which individual actions contribute, although they normally are relatively costly to the contributors. It is worth noting, too, that the interaction (either ‘parametric’ or ‘strategic’ interaction) is typically present in joint action in the wide sense and sometimes also in cases of the kinds (1) and (2).

The participants’ preference structures in a joint action can be fully cooperative (for example, carrying a table) or they can be to some extent noncooperative (for example, chess, buying and selling). Furthermore, joint action can be physical (for example, carrying a table) or it can involve a conventional or normative element such as transfer of rights (for instance, toasting a national victory, making a business deal, getting married).

Joint action as collective action performed because of a social attitude and social action in the broader senses (1) and (2) are all cases of collective social action and seem to make the classification exhaustive.

See also: Action; Social norms

RAIMO TUOMELA

References and further reading

Coleman, J. (1990) *Foundations of Social Theory*, Cambridge, MA and London: The Belknap Press of Harvard University Press.(Gives a comprehensive classification of social action using conceptual and mathematical tools familiar from economics.)

Elster, J. (1989) *The Cement of Society*, Cambridge: Cambridge University Press. (Contains classifications of collective action and discussions of the kinds of motivation of collective action.)

Holmstroem-Hintikka, G. and Tuomela, R. (eds) (1997) *Contemporary Action Theory*, vol. 2 *Social Action*, Synthese Library, Dordrecht and Boston: Kluwer.(Volume two contains papers about social action.)

Miller, S. (1992) ‘Joint Action’, *Philosophical Papers* XXI: 275-97.(Presents an account of social action based on collective ends.)

Tuomela, R. (1994) *The Importance of Us: A Philosophical Study of Basic Social Notions*, Stanford, CA: Stanford University Press.(Gives a comprehensive account of social action, including actions performed by social groups; special emphasis is given to the view that full-blown social action involves explicit or implicit agreement to perform the action.)

Social choice

Social choice theory is the branch of economics concerned with the relationships between individual values, preferences and rights and collective decision making and evaluation. Social choice theory therefore provides connections between the formal analysis of rational choice, the debate on political process, and ethics. A central theme in social choice theory has been the aggregation of individual preferences into either a social decision rule or a social evaluation rule. The most famous result in social choice theory - Arrow's impossibility theorem - is that such aggregation is impossible if individual preferences are conceived as ordinal in nature, and if the aggregation procedure is to satisfy certain apparently reasonable conditions. This result implies that neither a voting system nor a system of moral evaluation can be found that satisfies all of the required conditions. Further impossibility theorems arise from attempts to model the role of individual rights.

Much of social choice theory is concerned with interpreting, extending and questioning these impossibility theorems in a variety of contexts. This discussion has generated an extensive interchange at the margins of economics and ethics on topics such as the commensurability of values and the relationship between morality and rationality.

1 Impossibility theorems

Arrow's theorem concerns the aggregation of a set of individual orderings to form a social or collective ordering. The word 'ordering' is used in its technical sense to indicate an ordinal relation that is reflexive, transitive and complete. The objects of the ordering are not specified by the theorem but, following Arrow, social choice theorists refer to orderings of social states; that is, alternative complete descriptions of the world. Similarly, the nature of the individual orderings is not specified by the theorem, but social choice theorists refer to individual preferences. On this standard interpretation, the theorem concerns the possibility of aggregating a set of individual preference orderings over states of the world into a single social ordering.

Arrow (1951) investigated four conditions that impose requirements on the process of aggregation. Condition U, the condition of unrestricted domain, requires the process of aggregation to operate for arbitrary sets of individual orderings. Condition P, the weak Pareto principle, requires that if every individual ranks α above β, then society must rank α above β. Condition I, the independence of irrelevant alternatives, requires that the social ordering over any pair of alternatives depends only on individual orderings over that pair of alternatives. Condition D, nondictatorship, requires that there be no individual whose ordering becomes the social ordering regardless of the orderings of all other individuals. The four conditions were viewed by Arrow as necessary but insufficient conditions for rational social choice. The theorem is that there exists no process of aggregation which satisfies all four conditions.

Much of normative economics is welfarist in the sense that it takes individual utilities to be the only admissible information, and Paretian in the sense of Arrow's condition P. In fact welfarism can be thought of as the combination of Arrow's conditions U, P, and I; so the Arrow theorem can be seen as demonstrating the impossibility of nondictatorial welfarism.

A second impossibility theorem highlights the tension between the weak Pareto condition and a weak concept of individual rights. The relevant concept of a right is that an individual is decisive in the choice between a particular pair of alternative social states - a pair that might differ only in some respect deemed to be private to that individual - so that individual A has a right with respect to social states α and β if and only if A's ranking of α and β is causally sufficient to determine the social ranking of α and β.

Sen (1970a) shows the impossibility of a Paretian liberal by proving that no social choice rule can simultaneously satisfy the weak Pareto condition and grant at least two individuals rights over (distinct) pairs of alternatives.

2 Interpretation and critique

Sen (1977) offers four settings in which the Arrow result may be considered: (1) the aggregation of individual interests into overall ethical judgments of social welfare, (2) the aggregation of individual ethical judgments into overall ethical judgments of social welfare, (3) the aggregation of individual interests into social decisions, (4) the

aggregation of individual ethical judgments into social decisions. Settings (1) and (2) cast social choice theory in terms of ethics, while settings (3) and (4) cast social choice theory in terms of politics. Alternatively, settings (1) and (3) focus on the aggregation of private interests or utilities, while settings (2) and (4) focus on the aggregation of broader judgments of overall value. While the theorem applies equally in each of these settings, its interpretation and significance varies.

The now standard interpretation of the theorem in setting (1) is that simple personal preference orderings cannot in themselves provide sufficient information to ground social evaluations. Further information may take either of two forms; it might be more detailed information concerning the welfare of individuals, or it might relate to non-welfare aspects of individuals. Within the confines of welfarism, the two most obvious types of further information arise from the cardinality of individual utility and the interpersonal comparability of utility. In fact, if cardinality is allowed, but incomparability maintained, the Arrow result is not seriously affected. But once interpersonal comparability is allowed, the nature of the feasible social orderings depends upon the measurability of utility and the extent of interpersonal comparability. Thus, if ordinalism is retained alongside full comparability, social orderings such as that generated by the maximin criterion are feasible. Of particular interest is the simple utilitarian ordering in which social states are ranked by reference to the simple sum of individual utilities. Such utilitarianism requires at least cardinal scale measurability and at least unit comparability.

If welfarism is breached, so that further categories of information are allowed, the feasibility of social orderings depends on the answers to similar questions concerning the measurability of the new aspect of wellbeing, and its commensurability both across individuals and with utility.

In setting (2) we move from individual utility to individual judgments of social welfare, and it is here that the Arrow theorem is most persuasive. There can be no appeal to additional information in this case, since individual judgments may be taken to be all-things-considered judgments. However, the novelty and practical significance of the theorem may be relatively slight in this setting. It is hardly a surprise to find that different value systems (individuals' judgments of social welfare) are, in general, incompatible with each other.

In settings (3) and (4) we move from ethics to politics in viewing social decision-making rules based on individual interests or individual ethical judgments; we are concerned with the more practical matters of the design of voting systems or committee procedures. One general point to be made is that in these settings we are often not concerned with generating a full social ordering over the alternatives, but with selecting a winner (or set of joint winners). This fact allows us to weaken the requirements placed on the aggregation process slightly and so formally escape from the impossibility theorem. But this escape is more apparent than real. If we require the choices made by our voting rule to satisfy even rather weak consistency conditions, new impossibility results arise.

In these settings the Arrow theorem extends and generalizes the results on specific voting mechanisms associated with Condorcet and Borda, most famously that simple majority voting over at least three alternatives can yield intransitivities or cycles - so that even if all voters have well-behaved preferences over the alternatives, and each votes in accordance with those preferences, alternative α can be majority-preferred to alternative β, alternative β majority-preferred to alternative γ, and alternative γ majority-preferred to alternative α. A further result which follows from the Arrow theorem when interpreted in terms of social decision making relates to the manipulability of voting mechanisms. Broadly, the result is that all feasible voting schemes will, in general, provide incentives for individuals to vote strategically; that is, to misrepresent their true preferences in an attempt to manipulate the outcome of the vote.

The debate on the impossibility of a Paretian liberal has been voluminous. Topics considered include a variety of conceptions of individual rights - the contrast between the right to determine an aspect of a social outcome and the right to determine the individual's own strategic behaviour, for example; the extent to which preferences might be laundered to exclude the meddlesome preferences of one individual over matters deemed to be private to another individual; the extent to which the ability to trade rights might eliminate the results, and so on. Whatever the modifications to Sen's original structure, it seems that if aspects of individuals other than their preferences or values are to be included in the process of social evaluation or decision making, further conflicts await.

A fundamental criticism of the basic enterprise of social choice theory first put by Buchanan (1954) argues from an individualistic perspective that it is simply inappropriate to investigate social evaluation or social decision making

against criteria of collective rationality. This approach suggests that while individuals may be able to articulate evaluative orderings over alternative states of the world, there is no sense in the idea of a social or aggregate evaluative ordering. Furthermore, if collective decision making rather than evaluation is the focus of attention, we must attend to the question of how individuals might agree on decision-making procedures. In a commonly used example, we might agree to a process of simple majority voting in certain contexts without any commitment to the idea that the outcome of this political procedure is in any aggregate sense best. The claim is that the appropriate framework for the study of collective action problems is provided by the model of voluntary exchange between individuals. This position is strongly linked to a generally contractarian view of both politics and ethics (see Contractarianism).

3 Economics and ethics

Social choice theory has stimulated considerable work in the intersection of economics and ethics. In part this is because of the detailed, formal attention paid to questions such as the measurability and comparability of welfare, or the specification of individual rights; in part because of the implications of the theory for the classic debates between consequentialist and non-consequentialist moralities, or between liberalism, utilitarianism and contractarianism; and in part because of the illumination of the relationships between rationality and morality. It is not too difficult to justify the claim that social choice theory has had a major impact on both normative economics and moral theory - but most especially on the interaction between the two.

The debate in normative theory is in sharp contrast to the earlier orthodoxy within economics, when economics was seen as a predominantly technical study divorced from value judgments. Traditional neoclassical economics of the type captured in the textbooks of the mid twentieth century might be caricatured as depending on the twin ideas of a simple conception of self-interested, utility-maximizing individual rationality and an equally simple welfarist conception of the public interest. Initially social choice theory challenged the assumed link between these two ideas, but much of the continuing contribution of social choice theory lies in the more detailed re-examination of each idea. In the same way, while many of the relatively early contributions to social choice theory might be seen as technical exercises in aggregation theory, the more lasting contributions are those which engage with the substantive issues in ethics or political philosophy. In this way social choice theory has provided the basis for contributions which draw on the technical apparatus of economics in addressing topics of perennial interest to philosophers, while at the same time opening up the possibility of a reverse flow of ideas from philosophy into economics.

See also: Economics and ethics; Rational choice theory; Rights; Utilitarianism; Welfare

ALAN HAMLIN

References and further reading

Arrow, K.J. (1951) *Social Choice and Individual Values*, New Haven, CT: Yale University Press, rev. edn, 1963. (Classic source for the Arrow impossibility theorem.)

Buchanan, J.M. (1954) 'Social Choice, Democracy and Free Markets', *Journal of Political Economy* 62: 114-23. (Individualist criticism of social choice approach - included in Rowley, 1993.)

Hamlin, A.P. (ed.) (1996) *Ethics and Economics*, Aldershot: Edward Elgar.(A two volume collection of major contributions with a substantive introduction.)

Rowley, C.K. (ed.) (1993) *Social Choice Theory*, Aldershot: Edward Elgar.(A three volume collection of major contributions to social choice theory.)

Sen, A.K. (1970a) 'The Impossibility of a Paretian Liberal', *Journal of Political Economy* 78: 152-7.(A classic impossibility theorem - included in Rowley, 1993.)

Sen, A.K. (1970b) *Collective Choice and Social Welfare*, San Francisco, CA: Holden-Day.(Still the best book-length introduction to the ideas and techniques of social choice theory.)

Sen, A.K. (1977) 'Social Choice Theory: A Re-examination', *Econometrica* 45: 53-89. (Review of literature emphasizing the informational context - included in Rowley, 1993.)

Sen, A.K. (1987) *On Ethics and Economics*, Oxford: Blackwell.(Introductory lectures with an extensive bibliography.)

Social democracy

The idea of social democracy is now used to describe a society the economy of which is predominantly capitalist, but where the state acts to regulate the economy in the general interest, provides welfare services outside of it and attempts to alter the distribution of income and wealth in the name of social justice. Originally 'social democracy' was more or less equivalent to 'socialism'. But since the mid-twentieth century, those who think of themselves as social democrats have come to believe that the old opposition between capitalism and socialism is outmoded; many of the values upheld by earlier socialists can be promoted by reforming capitalism rather than abolishing it.

Although it bases itself on values like democracy and social justice, social democracy cannot really be described as a political philosophy: there is no systematic statement or great text that can be pointed to as a definitive account of social democratic ideals. In practical politics, however, social democratic ideas have been very influential, guiding the policies of most Western states in the post-war world.

1 Social democracy *c*.1880-1940

In its earliest phase, social democracy became the leading form of socialism in Western societies (see Socialism). It emerged from a fusion between the ideas of revisionist Marxists such as Eduard Bernstein and those of 'new' liberals such as L.T. Hobhouse who had come to believe that some form of socialism was necessary in order to realize liberal ideals. It differed from the Marxism promulgated by Lenin and his followers in the international communist movement in three main respects. First, in place of the belief that the economic collapse of capitalism was inevitable, the social democrats held that capitalism could only be transformed by the political will of the people; it was important, therefore, to emphasize not only class struggle, but the aims and ideals of socialism. Second, the transformation would not occur in a single revolutionary moment but over an extended period of time, with changes in economic structure being guided by social democratic parties holding political power. Third, the social democrats accepted the procedures and constraints of parliamentary democracy: confident that they could, in due course, win majority support, they saw the route to power lying through open elections.

In this phase, therefore, the final aim of social democracy was to replace private ownership of industry with state or social ownership, but the means were to be those of parliamentary democracy.

2 Social democracy *c*.1940-

In the second, mainly post-war, phase, social democrats came to believe that their ideals and values could be achieved by reforming capitalism rather than abolishing it. They favoured a mixed economy in which most industries would be privately owned, with only a small number of utilities and other essential services in public ownership. Their reformism expressed itself in four main directions.

(1) Under the influence of Keynes, they believed that, by controlling the level of public expenditure, the state could also control the level of aggregate demand, thereby preventing the expansion-recession cycles of uncontrolled capitalism (see Keynes, J.M. §2). In this way, steady economic growth and full employment were achievable.
(2) The state should provide welfare services (health, education and so forth) free of charge to all who needed them, and should also create a social insurance scheme for those temporarily unable to find work. Together these initiatives would create a social minimum that every citizen would be guaranteed regardless of personal circumstances.
(3) Workers' rights should be protected by legislation and the formation of trade unions, and the state should encourage workers' participation in the management of firms through co-determination schemes and the like. There should be (some degree of) economic democracy alongside democracy in the state itself.
(4) The first three measures would together serve to create considerably more social equality than was possible under unreformed capitalism. But, in addition, the state could legitimately use its powers of taxation and public provision to correct the distribution of resources still further, for instance by taxing incomes at progressive rates and using the proceeds to provide subsidies for the poor (such as cheap housing).

Social democrats argued that in these circumstances the old debate about the ownership of industry had become an irrelevance. Properly controlled, capitalism could be relied upon to produce the resources that social democratic governments could then use to advance their ideals practically. These ideals were a widening of democracy beyond the state to encompass economic and social organizations; social justice, in the sense of equal opportunity to achieve positions of advantage, together with the guaranteed social minimum as outlined above; and, less tangibly, a greater sense of community among citizens to offset the atomizing effects of the market economy.

In the last quarter of the twentieth century, the most powerful challenge to these ideals has come not from socialists but from libertarians, who argue that regulation of the market and redistribution of its proceeds are not only morally illegitimate but also likely to be self-defeating (see Libertarianism). Their cause has been aided by increasingly intense international economic competition, which has narrowed the range of policy options open to social democratic governments. Yet the ideals of social democracy have retained their popularity, and much recent political philosophy can be seen as an attempt to formalize and ground these ideals (see Dworkin, R.; Rawls, J. §4).

See also: Democracy; Justice

DAVID MILLER

References and further reading

Bernstein, E. (1899) *Die Voraussetzungen des Sozialismus und die Aufgaben der Sozialdemokratie (The Premises of Socialism and the Tasks of Social Democracy)*, Dietz: Stuttgart; trans. and ed. H. Tudor, *The Preconditions of Socialism*, Cambridge: Cambridge University Press, 1993.(Seminal text for the earlier generation of social democrats.)

Crosland, C.A.R. (1956) *The Future of Socialism*, London: Cape.(Important statement of social democratic principles by a leading member of the British Labour Party.)

Gay, P. (1952) *The Dilemma of Democratic Socialism*, New York: Columbia University Press.(A full-length study of Bernstein's philosophical and economic thought.)

Levine, A. (1988) *Arguing for Socialism*, London: Verso. (Recent attempt by a philosopher to restate the basic arguments for democratic socialism.)

Tilton, T. (1991) *The Political Theory of Swedish Social Democracy*, Oxford: Clarendon Press.(Surveys the thinking of the leading figures in what has perhaps been the most successful of all social democratic parties.)

Vaisey, J. (1971) *Social Democracy*, London: Weidenfeld & Nicolson.(An introductory survey of social democratic ideas and movements.)

Wright, A. (1986) *Socialisms*, Oxford: Oxford University Press.(Accessible introduction that places social democracy within the socialist tradition as a whole.)

Social epistemology

Social epistemology is the conceptual and normative study of the relevance to knowledge of social relations, interests and institutions. It is thus to be distinguished from the sociology of knowledge, which is an empirical study of the contingent social conditions or causes of what is commonly taken to be knowledge. Social epistemology revolves around the question of whether knowledge is to be understood individualistically or socially.

Epistemology has traditionally ascribed a secondary status to beliefs indebted to social relations - to testimony, expert authority, consensus, common sense and received wisdom. Such beliefs could attain the status of knowledge, if at all, only by being based on first-hand knowledge - that is, knowledge justified by the experience or reason of the individual knower.

Since the work of the common sense Scottish philosopher Thomas Reid in the mid-eighteenth century, epistemologists have from time to time taken seriously the idea that beliefs indebted to social relations have a primary and not merely secondary epistemic status. The bulk of work in social epistemology has, however, been done since Thomas Kuhn depicted scientific revolutions as involving social changes in science. Work on the subject since 1980 has been inspired by the 'strong programme' in the sociology of science, by feminist epistemology and by the naturalistic epistemology of W.V. Quine. These influences have inspired epistemologists to rethink the role of social relations - especially testimony - in knowledge. The subject that has emerged may be divided into three branches: the place of social factors in the knowledge possessed by individuals; the organization of individuals' cognitive labour; and the nature of collective knowledge, including common sense, consensus and common, group, communal and impersonal knowledge.

1 Reliance on testimony

The epistemological status of testimony and received wisdom is the topic in social epistemology which is most extensively treated in classical texts and contemporary discussion, and the only topic for which we have a detailed history (Coady 1991) (see Testimony). The central question here is whether testimony is a source of individual knowledge not derived from a non-social source like perception. According to *strong* individualism, to whichJohn Locke subscribes, all knowledge must be first-hand; no belief can be epistemically justified on the basis of testimony (where epistemic justification is the kind of justification required for knowledge - see Justification, epistemic §1). The objection to this view is that it entails a broad scepticism, excluding my knowledge of such matters as my own date of birth and the proposition that a cloudless sky is usually blue. According to *weak* individualism, a belief may be justified on the basis of testimony, but any such belief must ultimately be justified on the basis of non-testimonially justified beliefs. There are at least three versions of weak individualism:

(1) On the *inductive* version, to which David Hume subscribes, a belief based on testimony (call it a 'testimonial belief') is justified on the basis of the non-testimonially justified belief that the testimony is trustworthy or reliable (that is, tends to be true), which in turn is justified by induction from non-testimonially justified beliefs (see Testimony §3). The most significant objection to the inductive version is that we rely on testimony for many of the beliefs that would have to serve as the non-testimonial basis of the induction to the reliability of testimony. It is practically impossible to check at first-hand more than a tiny fraction of the reports given by testimony. Hence, the basis for an induction is too slim to justify the belief in reliability.
(2) On the a priori version of weak individualism, a testimonial belief is justified by appeal to an epistemic parity between my own beliefs and those of others. The idea that one person is epistemically no better than another is ancient and was assumed by Sextus Empiricus in the following maxim, which he used to argue for Pyrrhonian scepticism: where others disagree with me, I ought to suspend judgment. The present appeal to parity reverses the sceptical strategy of Sextus. A priori, I have no more reason to trust my own beliefs than I do those of others. Consequently, if I may trust my own beliefs, I may also trust others' beliefs, and thus I am justified in my testimonial beliefs. An objection to this version is that it equivocates on 'trust'. Perhaps I must trust others in the same sense in which I must trust myself. But why suppose that it follows that others' beliefs are trustworthy or reliable in a sense relevant to *my* being justified in having the same beliefs, rather than in a sense relevant to *their* being justified in these beliefs? There is no reason a priori to suppose that my epistemic position is similar to theirs in a way that makes these beliefs justified for me as they are for them. Thus it does not follow that my testimonial

beliefs are justified.

(3) On the coherence version, a testimonial belief is justified by its coherence with non-testimonially justified beliefs. The trouble with this proposal is that the objection to the inductive version shows that there is very little coherence between any given testimonial belief and the non-testimonially justified beliefs that might form the basis of an induction to the reliability of the testimony. No doubt testimonial beliefs are often unsurprising and in this sense fit with non-testimonial beliefs. But the epistemic force of this lack of surprise depends on antecedent generalizations about the world, and these in turn depend on prior testimony. Nor can the proponent of the coherence version retreat to the idea that testimonial beliefs are justified by the coherence of all beliefs together, testimonial as well as non-testimonial. For this is no longer a version of weak individualism, since testimonial beliefs are no longer justified on the basis of non-testimonially justified beliefs alone. True, it still counts as a version of individualism. But since we currently have no idea how all these beliefs are supposed to cohere, it seems we lack an individualist account of the justification of testimonial beliefs.

What is the alternative to individualism? Reid proposed that we take the justification of testimonial beliefs to be underived, on a par with the justification of perceptual beliefs (see Reid, T. §7). The justification of testimonial beliefs is governed by a first (that is, underived) principle, ‘That there is a certain regard due to human testimony in matters of fact, and even to human authority in matters of opinion’ (Reid [1785] 1969: 640). Conformity to this principle is made possible by certain innate dispositions of the human constitution: the principle of veracity, which disposes us to tell the truth, and the principle of credulity, which disposes us to believe what is said. First principles are not susceptible to proof, though they can be supported by recognizing their similarity to other first principles, under various desiderata, including common assent, conformity to the human constitution and practical necessity. In the terms of subsequent epistemology, Reid appears to have proposed that testimonial beliefs are *prima facie* justified in a manner analogous to the *prima facie* justification of perceptual beliefs. They are, at least early on, basically (that is, non-inferentially) justified and can be defeated by the subject’s other testimonial and non-testimonial beliefs.

2 Conditions of knowledge

Reid incorporates social factors into the conditions of justification by adding a principle of testimony. Others have proposed that the conditions of knowledge are in some sense social throughout. In particular, the content of the concept of knowledge is inherited from the social function of the concept of knowledge. Not all social functions of a concept are manifested in the content of the concept, but if the concept of knowledge has what we might call a social *epistemic* function, then perhaps it also has a social content. There are several examples of this approach.

Edward Craig (1990) has argued that the concept of knowledge is used in testimony as a tag to identify good informants. The concept of knowledge is not identified with the notion of a good informant itself, since knowing does not require being able to communicate to others the proposition believed, as does being a good informant. Rather, the proposal is that the concept of knowledge is an ‘objectivization’ of the notion of a good informant - a concept in which some of the features of a good informant that serve specific uses are stripped away. An analogous idea is that the concept of justification is used to identify *justifiers* - people able to justify a claim to an audience. On the most straightforward version of this view, suggested by David Annis (1978), one is justified in a belief (or claim) (relative to an audience) just in case one is able to justify it to the audience - or at least to hold one’s own in a conversation in which the claim is challenged. With regard to all such approaches, one worries that they have singled out a particular use of the concept of knowledge at the expense of other uses, uses that demand a different conceptual content (see Justification, epistemic §7).

3 The organization of cognitive labour

Traditional epistemology has little to say on the matter of how individuals should coordinate and institutionalize their inquiry to enhance their prospects of obtaining individual or group knowledge. In recent years, epistemologists have taken an interest in this matter, especially as it applies to science. One question concerns the obligations of individuals to disseminate information to or withhold it from others. A.I. Goldman (1992) has argued for *epistemic paternalism*, the view that individuals should sometimes withhold information to enhance the prospects that recipients will arrive at knowledge or true belief - as, for example, when evidence of a crime is ruled inadmissible in a court of law to prevent the jury from being misled. A second question is what features

interpersonal argumentation should have if it is to serve the development of knowledge. A third question is whether, to enhance the prospects of scientific group knowledge, individuals (or subgroups) should pursue lines of inquiry that are unpromising or based on improbable theories. A fourth question is whether individuals' cognitive biases induced by their social interests, such as ambition for professional credit, may steer individual research and belief-formation in such a way as to enhance the prospects of proper group belief-formation.

4 Collective knowledge

The key question of collective knowledge - of common, group and communal knowledge - is whether it reduces to individual knowledge or amounts to something over and above individual knowledge. A related question is whether individual knowledge must in some way be assimilated to communal knowledge (that is, the knowledge possessed by a community). The pragmatist C.S. Peirce argued for an affirmative answer to the latter question on the ground that the aim of proper method must be understood socially (see Peirce, C.S. §2). Peirce (1955) took the aim of proper method not to be true belief but relief from doubt. He then argued that the aim must be common rather than individual relief from doubt, since the 'social impulse' will drive people to think that others' opinions are as good as their own, and thus differences of opinion will induce doubt. One might question whether it follows from the claim that relief from doubt depends psychologically on common relief from doubt, that the aim of proper method is common relief from doubt. But Peirce could reply that there is no point in taking relief from individual doubt as the aim when we will in any case have to aim at common belief. Clearly, Peirce's argument for a common aim of proper method carries more weight for beliefs that are publicized, as in the case of science, than for beliefs that tend not to be exposed to the doubts of others (for example, beliefs about the details of one's private life).

It is natural to think that Peirce's argument, if successful, undermines a sharp distinction between individual and common knowledge, since a lone individual evidently cannot employ a method that aims at common beliefs; the coordinated efforts of many individuals are required. But whether Peirce's argument goes this far depends on what a proper method involves. If it involves only accumulating the evidence that would suffice to dispel doubt, rather than including the social actions that actually dispel doubt, then an individual on their own may still employ a proper method.

There are other accounts of individual justification that in some way assimilate it to communal justification. In *On Certainty*, Wittgenstein suggests that an individual's belief is justified by appeal to certain communally accepted claims. On a multi-perspectival view of justification, an individual is justified in believing *p* only if *p* is accepted from each perspective represented in the individual's community. Another view is that a belief is justified only if it conforms to communally accepted rules (see Contextualism, epistemological).

Even if individual knowledge is not assimilated to communal knowledge, there remains the question of whether common or group knowledge stands over and above individual knowledge or, instead, reduces to it. *Common* or *mutual* knowledge does reduce to individual knowledge: it obtains when two or more individuals know a proposition, know that each of the others knows, and so on. Common knowledge has proved an essential ingredient in the analysis of social conventions and group knowledge, but whether it amounts to group knowledge is a matter of controversy.

The simplest individualistic account of group justification is a *summative* account: a group is justified in believing a proposition just in case most (or all) of its members are justified in believing the proposition. Opposed to an individualistic account is a *joint* account, on which group justification is not a matter of group members being justified in believing *p*, but of members jointly possessing a good reason to believe *p*, where members jointly possess a reason for *p* just in case they openly express a willingness to treat it as a reason for *p*. The joint account characterizes group justification in terms of the actions and attitudes of the group members. It thus avoids requiring for group justification that there be some group mind over and above the minds of the members of the group.

5 The social construction of knowledge

There are three senses in which it has been claimed that knowledge is socially constructed (see Constructivism). (1) Some sociologists of science (for example, Bruno Latour 1987) have proposed that the facts reported in scientific journals are fabricated by the manipulation of inscriptions involving 'modalities' (that is, ways of

embedding sentences in other sentences). It has also been proposed (Joseph Rouse 1987) that many scientific laws strictly apply only to artificially created laboratory phenomena and extend to phenomena in nature only by analogy; thus, a law implicitly refers to the social conditions in which laboratory phenomena are created. (2) It has been proposed (again by Rouse) that the conditions of scientific communal knowledge are social. The argument is that the standards of knowledge to which scientific theories and experimental results are held (for example, the degree of confirmation required) depend on the uses to which the theories are put, and these uses are in turn determined by research needs, hence by the social conditions that determine those needs. (3) Finally, it has been claimed, particularly by those associated with the ‘strong programme’ in the sociology of science, that scientific theoretical and observational communal knowledge, or at least beliefs that are esteemed knowledge, are caused by social, economic and political interests. This sociological causal claim contrasts with *rationalism* about the causes of scientific theory choice, according to which theory choices are caused by rational reasoning on the basis of observations and theoretical desiderata, except where noncognitive factors (such as desires, emotions, peer pressure and the like) interfere with rational choices and pre-empt them.

One argument for the sociological causal claim is a priori: theories are underdetermined by observations and so choice is not caused by observations alone but by social interests (see Underdetermination). Nor can rationalists appeal to theoretical desiderata like simplicity and explanatory power to explain theory choice in the face of underdetermination, since these desiderata are merely rationalizations of theory choice, not rationally supported independently of choice. This objection to rationalism underestimates the theory-independent grounds for desiderata like simplicity - such grounds as that simpler theories carry less information than complex theories and thus accepting them entails less exposure to the risk of error (see Theoretical (epistemic) virtues). These grounds for deciding between theories are sometimes said to be theory-laden, leaving the choice of theory once again unexplained. But even if this is so, the way in which these grounds are selected may depend on metaprinciples relating grounds and theories.

Sociologists have also supported the sociological causal claim by appeal to historical case studies of the role of interests in theory choice (see Gender and science §3; Marxist philosophy of science; Postcolonial philosophy of science; Sociology of knowledge). But the poverty of the historical materials makes it difficult to establish that interests cause, and are not merely correlated with, theory choice in these cases. However, a correlation between interests and theory choice, even if not shown to be causal, may be hard to reconcile with rationalism, since we would not expect interests to be correlated with rational choices.

See also: Feminist epistemology; Naturalized epistemology

FREDERICK F. SCHMITT

References and further reading

Annis, D. (1978) ‘A contextualist theory of justification’, *American Philosophical Quarterly* 15: 213-19.(Argues that justification is relative to an audience and to the social role of the subject.)

Brown, J.R. (1987) *The Rational and the Social*, London: Routledge.(Critical review of the philosophical claims of the ‘strong programme’ in the sociology of scientific knowledge.)

Coady, C.A.J. (1991) *Testimony: A Philosophical Study*, Oxford: Oxford University Press.(Comprehensive historical and philosophical study of epistemological issues about testimony.)

Craig, E. (1990) *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*, Oxford: Clarendon Press.(Explains various features of the concept of knowledge by deriving it from the notion of a good informant.)

Gilbert, M. (1989) *On Social Facts*, London: Routledge.(The most systematic account of collective concepts yet attempted, with an extensive discussion of common knowledge and group belief.)

Goldman, A.I. (1992) *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge, MA: MIT Press.(Includes articles on epistemic paternalism and - with M. Shaked - on the role of credit in fostering true belief in science.)

Hume, D. (1748) *An Enquiry Concerning Human Understanding*, in *Hume’s Enquiries*, ed. P.H. Nidditch and L.A. Selby-Bigge, Oxford: Oxford University Press, 1975. (The discussion of testimony appears in §88.)

Kitcher, P. (1990) ‘The division of cognitive labor’, *Journal of Philosophy* 87: 5-22. (Argues that pursuing lines of inquiry based on improbable theories can foster the cognitive goals of science.)

Latour, B. (1987) *Science in Action*, Cambridge, MA: Harvard University Press. (Argues for the first kind of social constructivism defined in §5 above.)

Lehrer, K. and Wagner, C. (1981) *Rational Consensus in Science and Society*, Dordrecht: Reidel.(Contains an account of the conditions in which an individual is committed to accepting the consensus of a group.)
Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. A.C. Fraser, New York: Dover, 2 vols, 1959. (The remarks about testimony appear at Book I, page 58 and Book IV, chapters 15 and 16, §§10 and 11.)
Longino, H. (1990) *Science as Social Knowledge*, Princeton, NJ: Princeton University Press. (A multiperspectival account of scientific knowledge.)
Peirce, C.S. (1955) *'The fixation of belief'*, in *Philosophical Writings of Peirce*, ed. J. Buchler, New York: Dover. (The argument for common relief from doubt as the aim of proper method appears on pages 12-13.)
Reid, T. (1785) *Essays on the Intellectual Powers of Man*, ed. B. Brody, Cambridge, MA: MIT Press, 1969.(The material on testimony appears in Essay VI, chapter 5.)
Rouse, J. (1987) *Knowledge and Power: Toward a Political Philosophy of Science*, Ithaca, NY: Cornell University Press.(Defends the first two versions of social constructivism about knowledge mentioned in §5 above.)
Schmitt, F.F. (ed.) (1987) *Synthèse* 62, special issue on social epistemology.(Collection of articles on diverse topics in social epistemology.)
Schmitt, F.F. (ed.) (1994) *Socializing Epistemology*, Lanham, MD: Rowman & Littlefield. (An anthology containing articles on topics in all branches of social epistemology, with an introduction that expands on §§1 and 5 of this entry and an extensive bibliography.)
Sextus Empiricus (*c*.200) *Outlines of Pyrrhonism*, in *Sextus Empiricus*, vol. 1, trans. R.G. Bury, Loeb Classical Library, Cambridge, MA: Harvard University Press, 1933. (The maxim of epistemic parity among persons is assumed in the second through fifth and in the tenth modes in book I, chapter 14.)
Solomon, M. (1994) 'Social empiricism', *Nous* 28: 325-43.(Argues that biases resulting from scientists' training and professional preoccupations may foster the cognitive goal of empirical adequacy.)
Wittgenstein, L. (1969) *On Certainty*, ed. G.E.M. Anscombe and G.H. von Wright, trans. D. Paul and G.E.M. Anscombe, New York: Harper.(Suggests that beliefs or claims are justified on the basis of communally accepted propositions.)

Social laws

Social science has always aspired to be like natural science (Hawthorn 1976). And since natural science claims to discover laws of nature, social science has always claimed to discover laws of society. There are two important problems raised by such social laws. What makes the laws social in the appropriate sense? And if they really obtain, does that mean that human beings are not as autonomous as one might have thought: that we are pawns in a game that the laws control?

We may take laws here in a more or less unanalysed fashion as reliable regularities, usually reliable causal regularities, in the succession of events on one another. The statement of a causal law will postulate that whenever an event of a certain type is realized then an event of another type is also going to be realized, perhaps so close in time that they overlap, perhaps at some temporal distance. Or, more plausibly, it will postulate a probabilistic and bounded version of that sort of linkage. The linkage will be probabilistic to the extent that the first sort of event is held to raise the chance of the second sort occurring, not to make it inevitable. And it will be bounded so far as it is assumed to obtain only under certain boundary conditions: only when other things, usually other unspecified things, are equal.

What are the laws of the social realm that the social science project envisages? They cannot by all accounts be just the sorts of laws that empirical psychology, or the sort of psychology employed in microeconomics, postulates. They cannot just be regularities in the way people's individual beliefs and desires, probabilities and utilities, are formed and develop. They must be distinctively social laws (Brown 1984).

A law will be a distinctively social law, we may presume, just where the properties that it involves are, or at least include, some social properties. But what is it for a property to be social? There are many rival accounts (Ruben 1985). One simple one holds that a property is social just where it cannot be realized unless a number of intentional subjects, usually interactive intentional subjects, hold by certain attitudes or exhibit certain actions. The property of an individual that they are a judge or a celebrity, the property of a place that it is a country or a football field, the property of a group that its members meet on Sundays or that they generally go to sleep before midnight: these are all social properties on this generous account.

There is little doubt but that there are social laws, under this account of social laws (Pettit 1993). It is a putative law, however probabilistic and bounded, that urbanization leads to secularization, that an increase in unemployment leads to a rise in crime, and that commercial firms tend to approach the profit levels of their rivals. Such laws indeed are social in quite a strong sense. While the properties involved cannot be realized without a number of people displaying certain intentional actions or attitudes - that is what makes the properties social - still the laws are not just an extension of the body of psychological laws. They do not introduce factors that might be thought to interact with the variables of established psychological laws, thereby extending the laws causally. Nor do they involve patterns that are logically secured just by the fact that certain psychological laws obtain: they are logically as well as causally discontinuous with psychological laws.

Do social laws, so understood, threaten the sense that we human beings have of ourselves as relatively autonomous thinkers and agents? Those who espouse a collectivist or structuralist vision of social science, for example many in the tradition of Durkheim, tend to say that they do; they tend to take social science, in particular the discovery of social-structural laws, as a challenge to our ordinary common-sense view of ourselves. The idea is that whereas common sense represents us as forming attitudes and actions in response to familiar, more or less rational pressures, social science reveals that we are destined, perhaps even determined, to behave in the manner that certain aggregate social laws require. We are the unknowing dupes of those impersonal dictators (see Structuralism in social science §1).

But this vision of dictatorship by social law - this version of the 'death of the subject' - is uncompelling. It is more plausible to think of social laws as related to intentional or psychological laws - the laws that ensure our relative autonomy - in an architecture of the following kind (Pettit 1993). A higher-level social law relates an antecedent which is variously realizable at the intentional level to a consequent that is variously realizable at the intentional level: so urbanization or secularization may come about via any of an indeterminate number of actions, and any of

an indeterminate variety of attitudes, on the part of individuals. If the antecedent is related in a law-like way to the consequent, that is because the intentional realizers of the antecedent, no matter what they are, are more or less bound to lead, under the intentional laws, to the realization of the consequent. The realization of the antecedent ‘programmes’ for the realization of the consequent helps to ensure the realization of the consequent in a manner that depends on more or less familiar psychological regularities, not in a manner that imposes an alien order on the intentional realm.

This architecture is attractive, because it would enable us to see how social science may discover surprising laws without undermining our image of ourselves as more or less autonomous agents. It would also enable us to see how such laws can be logically discontinuous from the body of psychological laws. The laws may obtain, not just in virtue of psychological laws, but also in virtue of the boundary conditions under which those laws rule. If unemployment leads to crime, that may be a result, not just of the way human psychology works, but also of contingent features of the world in which it works: in a world where the unemployed could always have recourse to wild fruit and warm beaches, for example, the regularity might well break down.

See also: Natural law; Social sciences, philosophy of

PHILIP PETTIT

References and further reading

Brown, R. (1984) *The Nature of Social Laws*, Cambridge: Cambridge University Press.(Discusses the history of the notion of social law.)

Hawthorn, G. (1976) *Enlightenment and Despair: A History of Sociology*, Cambridge: Cambridge University Press.(Discusses the aspiration of social science to be science and discover genuine laws.)

Pettit, P. (1993) *The Common Mind: An Essay on Psychology, Society and Politics*, New York: Oxford University Press.(Chapter 3 characterizes social laws and argues in support of the programme-architecture.)

Ruben, D.-H. (1985) *The Metaphysics of the Social World*, London: Routledge. (Discusses the meaning of ‘social’ thoroughly.)

Social norms

A social norm may be defined as the rule of a particular social group. That men are to open doors for women, for instance, may be the rule of a particular group. But what is it for a group to have a rule, according to our everyday understanding? Philosophers have disagreed on this point. An important general issue is whether a group's having a rule is a matter of some or all of its members individually conforming to the rule or individually accepting the rule as a standard of behaviour for the group. An alternative type of account invokes the idea of the group's members jointly, rather than individually, accepting the rule, in effect agreeing to conform to it. It can be argued that this less individualistic account better explains the way in which people criticize deviations from social norms.

1 Lewis' game-theoretic account of social convention

Lewis (1969), following Schelling (1960), proposed an account of (social) convention such that underlying every convention is a recurring 'coordination problem' that will be solved by members' regular conformity to a certain pattern of behaviour (see Lewis, D.K.). A typical coordination problem occurs when a telephone conversation is cut off. Each party wants the call reconnected. If both call back, however, or if both wait, no reconnection will take place. There is a need for coordination of behaviour in order that the result both prefer occurs.

Lewis' proposal is roughly this: by definition, there is a convention when and only when a suitable regularity in behaviour arises within a coordination problem, accompanied by expectations of conformity. He argues that conventions in his sense are 'norms': we believe that those with such a convention ought to conform to it, since otherwise they go against their own and the other parties' preferences (1969: 97). That may be so, but the question remains whether conventions according to Lewis are social norms in the sense of group rules (see Decision and game theory).

Critics argue that Lewis has not captured the vernacular concept of a social convention, as he hoped (see, for instance, Gilbert 1989, ch. 6). Among other things, conventions do not only arise in the context of coordination problems. Further, it can be argued that from an intuitive point of view social conventions are a class of group rules. According to Lewis' definition, conventions do not amount to group rules intuitively. Among other things, if a convention were a group rule, the parties would understand they had reason to conform irrespective of their personal preferences and expectations.

2 Hart's practice theory of social rules

Hart's account of social or group rules runs roughly as follows: for there to be a social rule, a certain pattern of behaviour must be regularly conformed to within a given group. In addition, members of the group must regard the pattern as a standard by which group members' behaviour may be judged correct or incorrect. They must believe that, all equal, members ought to conform to it. Further, group members must believe they are justified in pressuring one another to conform. In typical cases, such pressure will be imposed on deviants and those who threaten deviance (see Hart, H.L.A.).

These conditions seem not to rule out cases like the following. The individual members of a group believe that lying is wrong. They also believe that members are justified in pressuring one another not to lie. In the situation as described, however, the rule proscribing lying does not seem to be a genuine group rule, as opposed to a rule that everyone individually accepts (and deems it appropriate for everyone to enforce) (Dworkin 1977; Hart[1961] 1994, postscript). What is missing from this case?

As Hart himself observes, when a group has a rule the members take themselves to be justified in the imposition of pressure on others solely because of actual or threatened violation of a rule of their group. This is so whether or not they take the rule to be a moral one. (Presumably the rule that men are to open doors for women might not be perceived as a moral one.) This suggests that an adequate account of social rules will make clear how, by their nature, they justify the imposition of pressure for conformity.

3 Gilbert's joint acceptance account of social rules

If we are looking for something other than a moral rule to ground an entitlement to exert pressure on others, an obvious candidate is something akin to an agreement. If you and I have agreed to act in a certain way in the future, and I then act differently, we will both understand that I have failed in an obligation I had to you, and that you therefore have grounds for reproving me, whether or not you choose to do so.

Of course group members do not in general get together and agree to conform to certain standards. Commitments of the sort involved in an agreement can be set up, however, without an agreement proper being made. Language can play a role here, as when members accept appeals to ‘our’ rule, with the attendant commitments that this may imply (Gilbert 1989, ch. 4).

Gilbert (1989, ch. 6) suggests that a social convention as ordinarily understood is a jointly accepted fiat, while a social rule is a jointly accepted rule, that may or may not be of the fiat form. For members of a population jointly to accept a rule is for them to enter a ‘joint commitment’ to accept that rule together, that is, as a body. A joint commitment is one that is seen by the participants precisely as joint: it is not the sum of the different individual members’ independent commitments. Rather, it is our commitment. All are committed through it, but these ‘individual commitments’ have a single source in the joint commitment, on which they depend. Each member is beholden to the others for release from the commitment. A joint commitment may be said to involve mutual obligations of a particular sort, and to provide grounds for reproof if the commitment is not complied with (Gilbert 1996).

Given this appeal to the notion of a joint commitment (which she sees as a fundamental intuitive concept) Gilbert argues, in effect, for a more holistic, less individualistic account of social norms than is envisaged by Hart or Lewis. Rules jointly accepted by group members are more aptly referred to as the rules of a group than are rules merely conformed to or individually accepted by group members.

See also: Holism and individualism in history and social science; Social laws; Social action

MARGARET GILBERT

References and further reading

Dworkin, R. (1977) *Taking Rights Seriously*, London: Duckworth.(A collection of Dworkin’s articles in jurisprudence including ‘The Model of Rules II’ in which he criticizes Hart’s account of social rules.)

Gilbert, M. (1989) *On Social Facts*, London: Routledge; 2nd edn, Princeton, NJ: Princeton University Press, 1992.(A theory of the structure of everyday collectivity concepts including the concept of a social group and a social convention.)

Gilbert, M. (1996) *Living Together: Rationality, Sociality, and Obligation*, Lanham, MD: Rowman and Littlefield. (A collection of Gilbert’s articles relating to *On Social Facts* (above). Chapter 12, ‘Agreements, Coercion, and Obligation’, argues for a particular understanding of the obligations of agreements and related phenomena, in terms of a ‘joint commitment’ underlying agreements.)

Hart, H.L.A. (1961) *The Concept of Law*, Oxford: Clarendon Press, 2nd edn with postscript, 1994.(A classic account of the nature of law in terms of a union of primary and secondary rules, with special reference to the nature of social rules in general.)

Hart, H.L.A. (1982) *Essays on Bentham*, Oxford: Clarendon Press.(Chapter X provides a new account of what it is for there to be a social rule to the effect that a particular ‘commander’ should be obeyed.)

Lewis, D.K. (1969) *Convention*, Cambridge, MA: Harvard University Press.(A game-theoretic account of social conventions, according to which they involve coordination problems, intended to throw light on the idea that language involves convention.)

Raz, J. (1975) *Practical Reason and Norms*, London: Hutchinson; Princeton, NJ: Princeton University Press, 1992.(Chapter 2 focuses on ‘mandatory norms’, offers a critique of Hart, and argues in relation to the role of norms in practical reasoning that mandatory norms are ‘exclusionary reasons’, that is, roughly, reasons to disregard other reasons.)

Schelling, T. (1960) *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.(Suggests that ‘coordination games’ underlie the stability of institutions and traditions. Influenced Lewis and others.)

Ullman-Margalit, E. (1977) *The Emergence of Norms*, Oxford: Oxford University Press. (Focuses on three game-theoretically defined structures, including coordination problems, and argues they are apt to give rise to norms.)

Social relativism

People in different societies have very different beliefs and systems of belief. To understand such diversity is a prime task of the student of society. The task is especially pressing when alien beliefs seem obviously mistaken, unreasonable or otherwise peculiar. A popular response is social relativism. Perhaps beliefs which seem mistaken, unreasonable or peculiar viewed from our perspective, are by no means mistaken, unreasonable or peculiar viewed from the perspective of the society in which they occur. Different things are not just thought true (reasonable, natural) in different societies - rather, they are true (reasonable, natural) in different societies. Relativism recognizes diversity and deals with it even-handedly.

Relativism has absurd results. Consider the view that what is true in society A need not be true in society B. So if society A believes in witches while society B does not, there are witches in A but not in B. Relativism regarding truth drives us to different 'worlds', one with witches in it and another without. This seems absurd: people who live in different societies do not in any literal sense live in different worlds. The challenge is to do justice to social diversity without falling into absurdities such as this.

1 Social diversity

Social diversity is one of the first and most familiar facts to strike the student of human beings and the societies in which they live. The foregoing treats of the social diversity of belief. Along with different beliefs come differences in behaviour, customs, rituals, traditions, conventions, institutions, political and legal arrangements, and so forth. For instance, some people enact ritual dances to ensure that rain will fall on the crops, others think that pregnancy has nothing to do with sexual intercourse, yet others blame magical influences for their misfortunes. Here capital punishment is frowned upon, there it is not. Here a particular religion is state-sponsored, there it is not. Many things other than belief-systems are 'relative to' (that is, distinctive of) different societies.

Yet there is good reason to focus on beliefs. The other differences are all constituted by, or sustained by, or at least associated with, differences in belief. Canons of rationality or of value are generally accepted beliefs about what is reasonable or valuable. It is the custom in a society to bury the dead only if sufficiently many members of that society think it is the thing to do and act accordingly. Traditions and institutions are sustained by the beliefs of the participants in them. Actually, there are deep questions regarding exactly how, and to what extent, social phenomena are 'dependent' on beliefs and belief-systems. We cannot enter into those controversies here, but rather focus on the diversity of beliefs, as is customary.

2 Social relativism

An early response to the diversity of beliefs was to distinguish 'primitive societies' (usually African) from advanced ones (usually European). The bizarre beliefs of 'primitive peoples' were just false, the product of irrational, illogical, even 'pre-logical' processes of thought. Then a reaction set in. You cannot really understand a 'primitive' society by viewing it from the perspective of your own and deeming its beliefs and practices false or irrational or pre-rational. Each society should be judged on its own terms. Beliefs deemed false and irrational by our lights may be perfectly sound and rational seen from the perspective of the society which sustains them. Just as belief-systems are socially and culturally bound, so are honorifics like 'true' or 'reasonable' with which they are evaluated. This is social relativism.

Social relativism is not merely the recognition of social diversity or of the difficulty of understanding the arrangements of one society from the perspective of another. Social relativists add that there is nothing to choose between different beliefs and practices. There are no neutral canons of truth or reason which would enable us to do this. Local mistakes remain possible. An incompetent witch-doctor may consult the poison oracle wrongly or have a false belief about the origin of a particular misfortune. An incompetent priest may administer the sacrament wrongly or make a theological mistake. But neither witchcraft nor Christianity can be globally false or irrational. Truth and rationality are themselves relative to societies. No belief or system of belief can be deemed false or irrational *simpliciter*.

3 Proliferation

One worry about social relativism concerns its tendency to proliferate. Societies are pluralistic and ill-defined. One group may enact magical rituals and also spread fertilizers to ensure good crops. Another may keep their cars well-maintained and also bless them with holy water to ensure safe travel. There exists both a Society of Friends and a Society of Accountants. This is no mere pun on the word ‘society’: the beliefs, traditions, practices, and so on, of these two groups are as diverse as one might wish.

Social diversity proliferates horizontally, across different societies at the same time. It also proliferates vertically, down the same (or different) societies at different times. For, as L.P. Hartley said at the beginning of *The Go-Between,* ‘The past is a foreign country - they do things differently there’. Historical diversity is as familiar a fact to the student of history, as social diversity is to the student of society.

One example of historical diversity is particularly important. Some have been tempted to take science and technology as the arbiters of truth and reason, against which all is to be judged. If we want good crops, we know that fertilizing the soil is better than uttering magical incantations over it. We know this because science teaches it. If we have to choose between ‘African traditional thought and western science’, to cite the title of a famous contribution to these debates, by Robin Horton (reprinted in Wilson 1970), it is clear how we should choose. But is ‘western science’ the monolithic system of belief and practice which this seems to suppose? Thomas Kuhn said not. A glance into the history of science reveals the same kind of diversity of belief and practice, as we see today as we look across different societies. Furthermore, Kuhn seemed to say, there is as little to choose between the different ways of practising ‘western science’, as there is between the different beliefs and practices we find today in different societies. Far from being a way out of the relativistic impasse, science viewed historically is a part of it.

One response to social diversity is to maintain that, beneath all the diversity and under-pinning it, there are continuities. All social groups are interested in finding solutions to the basic problems of life, usually detailed as food, shelter and procreation. To that end, all social groups are interested in explaining and, so far as they are able, controlling the phenomena of nature. In pursuit of these basic ends, all social groups have an interest in acquiring beliefs that are true and explanations that work.

Thoroughgoing relativists remain unimpressed. They see diversity even concerning the ‘basic problems of life’. There are peoples who (it seems) believe that pregnancy has nothing to do with sexual intercourse. A common interest in explanation represents continuity only if proffered explanations are of the same type and are adjudicated by common standards. A common interest in true beliefs represents continuity only if there is a common notion of truth.

Like many general sceptical theses, thoroughgoing relativism threatens to self-destruct. ‘All truths are relative to some S’ will itself, if true, only be true relative to some S. One can ask in which S relativism is true. And the best answer seems to involve the society or sub-culture of social anthropologists and their various philosophical hangers-on. Alternatively, one can view relativism as an absolute, ahistorical, trans-cultural truth - in which case, it is false.

4 Idealism

Relativism about truth flies in the face of platitudes. For example, it is true that bloodletting cures diseases if, and only if, bloodletting really does cure diseases. Or generally, something or other is true if, and only if, it is really the case. Such platitudes seem to leave no room for the notion that something might be true for A but not for B. These platitudes assume that what is the case does not in general depend upon what people think is the case. It is to suppose that the world, its objects and their features do not in general depend on our ideas, on our thought or talk. Suppose instead that what is true for one society need not be true for another. It follows that inhabitants of different societies inhabit different worlds. In society A it is believed that diseases are brought about by ill-wishing magicians, in society B it is believed that diseases are brought about by bacteria, viruses and other ‘natural causes’. Some people might think that both beliefs cannot be true (though both might be false). But this is to fall back upon a monolithic, societally-neutral notion of truth. No, it is true in society A that diseases are brought about by ill-wishing magicians - not just thought to be true (that is, believed) but true. It follows that there really are in the world ill-wishing magicians - for inhabitants of society A. Once monolithic truth is jettisoned, so is a monolithic ‘world’ to which monolithic truths ‘correspond’. No, different peoples inhabit different worlds. The world of the

Kalahari bushman is very different from the world of the industrialized European. The world of the Aristotelian scientist is very different from the world of the Newtonian scientist. If truth is socially constructed, so is reality. The world, its objects, and their features depend for their very nature upon what human beings think about the world, its objects, and their features. (And why just human beings? Avoid human chauvinism, and you will have to agree that the world of the chimpanzee or honeybee is worlds apart from the world of Albert Einstein.) This is idealism (see Idealism; Realism and antirealism).

Relativistic and idealistic ways of speaking are popular, and may be innocuous. ‘That’s true for you but not for me’ may just be a picturesque way of saying that you think that true but I do not. ‘Witches really existed in seventeenth-century Europe’ may just be a picturesque way of saying that seventeenth-century Europeans thought that witches really existed. But if we take such locutions literally, they involve us in idealism.

What of logic and rationality - might they be socially relative? Might a people believe that there are witches - and also believe that there are not? Might there be a ‘pre-logical’ society where the logical law of contradiction breaks down? The matter of rationality is more difficult. Beliefs can be rational yet false, and irrational yet true. Our ancestors once believed that the earth was flat - their belief was false, but reasonable given the evidence and argument available to them. Should we now find a society of flat-earthers, we might declare their belief false yet reasonable, for much the same reasons. Alternatively, we might suspect that they do not really believe this and that we have mistranslated utterances which led us to think that they do. The enterprise of translation, however, seems to presuppose that sincere utterances are thought true, and not thought untrue at the same time.

Cultural relativism and idealism are fuelled by a laudable ambition to be even-handed about the beliefs and practices of others. But we can understand the beliefs and practices of others ‘on their own terms’ without dismantling the rational unity of humankind. Indeed, the latter is curiously self-defeating. What is more patronizing than the declaration that this or that is true and real - but just for you? The project is ‘understanding a primitive society’ (to cite the title of another famous contribution to these debates by Peter Winch, reprinted in Wilson 1970). Do you advance that project by claiming that in ‘primitive societies’ understanding means something quite different to what it means in ours?

See also: Epistemic relativism; Moral relativism; Rationality and cultural relativism; Rationality of belief; Relativism

ALAN MUSGRAVE

References and further reading

Hollis, M. and Lukes, S. (eds) (1982) *Rationality and Relativism*, Oxford: Blackwell. (A useful collection of readings, containing a bibliography.)

Sumner, W.G. (1907) *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*, Boston, MA: Ayer, 1979.(Sumner’s masterpiece, in which he argues that each society has its ultimate ‘mores’.)

Wilson, B.R. (ed.) (1970) *Rationality*, Oxford: Blackwell.(A more recent collection of readings, also containing a helpful bibliography.)

Social science, contemporary philosophy of

Some philosophers think that the study of social phenomena must apply methods from natural science. Researchers should discover causal regularities (whenever C operates, E occurs) and fit them into systematic theories. Some philosophers hold that social phenomena call for an entirely different approach, in which researchers seek to interpret fully the meaning of people's actions, including their efforts to communicate and cooperate. On this view, the nearest that researchers will come to regularities will be to discover rules (whenever the situation is S, everyone must do A). The nearest that they will get to systematic theories will be systematic expositions of rules, like the rules of a kinship system.

Besides the naturalistic school and the interpretive school, the philosophy of social science harbours a critical school. This finds researches endorsed by the other two schools shot through with bias. It inclines to agree with the interpretive school in resisting naturalistic methods. However, its charges against naturalistic researches extend to interpretations. For interpretations may give untroubled pictures of societies in deep trouble, or picture the trouble in ways that serve the interests of the people who profit from it, for example, by leaving current rules about taking workers on and laying them off unquestioned. Here the critical school may itself use naturalistic methods. If it contends that ignoring ways of reassigning authority over employment increases the chances of private enterprises' retaining their present authority, the critical school is talking about a causal connection. There is no rule that says anyone must increase the chances.

Yet the researches sponsored by the three schools are complementary to the degree that researches into regularities and into rules are complementary. Settled social rules have counterparts in causal regularities, which may be expressed in similar terms, although the evidence for regularities need not include intended conformity. Some regularities are not counterparts of rules, but involve rules notwithstanding. If the proportion of marriages in Arizona ending in divorce is regularly one-third, that is not (as it happens) because one-third of Arizonans who marry must divorce. Yet marriage and divorce are actions that fall under rules.

The three schools do more than endorse studies of rules or regularities. The critical school denies that any study of social phenomena can be value-free, in particular on the point of emancipating people from the oppressions of current society. Either researchers work with the critical school to expose oppression; or they work for the oppressors. The interpretive school brings forward subjective features of human actions and experiences that overflow the study of rules. These features, too, may be reported or ascribed correctly or incorrectly; however, the truth about them may be best expressed in narrative texts more or less elaborate.

Postmodernism has generalized these themes in a sceptical direction. Every text can be read in multiple, often conflicting, ways, so there are always multiple, often conflicting, interpretations of whatever happens. Every interpretation serves a quest for power, whether or not it neatly favours or disfavours an oppressive social class. Such contentions undermine assumptions that the three schools make about seeking truth regarding social phenomena. They do even more to undermine any assumption that the truths found will hold universally.

The assumption about universality, however, is a legacy of the positivist view of natural science. Positivism has given way to the model-theoretic or semantic view that science proceeds by constructing models to compare with real systems. A model - in social science, a model of regularities or one of rules - that fits any real system for a time is a scientific achievement empirically vindicated. Renouncing demands for universality, the philosophy of social science can make a firm stand on issues raised by postmodernism. It can accept from postmodernists the point that scientific success happens in local contexts and only for a time; but resist any further-reaching scepticism.

1 The point of distinguishing schools

Why make anything of a distinction between schools? One could look upon social science as embracing a great number of inquiries on a variety of topics. Some inquiries deal with regularities; some, with rules; some, as much with the one as with the other. They may all be subject to criticism for ignoring other matters, including sometimes their service to current oppressive arrangements.

However, the different contentions of philosophers on the subject of social science demand something like the

three-schools distinction. Among the contentions arise exclusive claims for naturalism; interpretation; critical social theory. The three-schools distinction distributes these exclusive claims under appropriate heads, then confronts them. This is worth doing not only to give the contentions in question due attention. It is in order because social scientists themselves carry on inquiries in different styles answering to the differences between the schools.

The three-schools distinction directs attention to important issues. Is any one school to prevail to the exclusion of the others? If all three survive, because all of them correspond to distinctive and worthwhile inquiries, are the schools and inquiries quite separate? Connections - close and extensive connections - between rules and regularities make the schools allies; and individual inquiries frequently mix the several styles. The schools remain allies, furthermore, when aspects of social phenomena come into the picture that demand philosophical attention though they do not explicitly have to do with rules or regularities.

Chief among these further aspects are, first, the basic subjective features - the temporal structures - of human action and experience, in both of which agents behold the present in a perspective established at once by the immediate past and the foreshadowed immediate future. Second, there is the interpretation of actions as texts the true interpretation of which changes, like the significance of judicial decisions, as history goes on and the perspective of interpretation changes. This is the concern of hermeneutics (see Hermeneutics). Both of these matters can be assigned to the interpretive school, along with a special interest in narratives, without making the school incoherent.

Similarly, the naturalistic school can digest the model-theoretic view of science, without in any way abdicating from advocacy of naturalistic methods (see Model theory). On the contrary, the advocacy becomes more plausible, without encouraging exclusive claims. Within the three-schools framework, the model-theoretic approach, once installed in the naturalistic school, invites application also in the interpretive school. The critical school has for its part a visible opportunity to employ models either of rules or of regularities.

With these new matters accommodated, the relations between the schools need re-examination; and re-examination strengthens the case for complementarity. Moreover, the schools framework furnishes a venue for assessing the impact of postmodernism. How far can it be accommodated, and where? Once the relations within and between the schools have been assessed again, with the lessons about postmodernism allowed for, the relations can be redescribed as relations of inquiries - inquiries concerned with rules, concerned with regularities, concerned with texts, concerned with subjective experience. If one likes, the distinction between the schools, having served all the purposes mentioned, can then be left behind as a historical curiosity.

2 Naturalism recast

John Stuart Mill was a strong early advocate of naturalistic methods in the social sciences; but in the twentieth century its chief advocates were for a long time philosophers grouped at least roughly with logical positivists. During its reign in the philosophy of natural science, positivism held that perfected scientific theories embraced universal laws holding for all times and places. This is an exacting prescription for the natural sciences. In the social sciences it has been impossible to fulfil, and a major embarrassment for the naturalistic school (see Positivism in the social sciences).

The prescription has vanished from the philosophy of natural science with the eclipse of positivism. Yet the revisions that would recast naturalistic doctrines regarding social science to incorporate the now-prevailing model-theoretic view of science have hardly begun. According to this view, as set forth by R.N. Giere (1988), the business of science is accomplished, first, by setting up models with predicates that hold true precisely only of the models and there as a matter of definition; and then showing that the models resemble one range or another of phenomena closely enough to be more useful than other models in prediction and explanation. The predicates may incorporate universally quantified statements (‘In our model of a perfectly competitive market, F’s [rises in price] will always be followed by G’s [increases in supply]’). However, these need hold only for the model or only for the model and (more or less loosely) for some range of phenomena, perhaps just one, strictly circumscribed (markets for pork bellies, given time to raise more pigs). They may also - another important relaxation in scientific standards - go no further in causal content than to specify causal conditions, necessary or sufficient only in the presence of other conditions, and leaving open questions about underlying causal mechanisms. One could know

that the number of pork bellies coming to market in a given year increased only after a rise in prices for pork bellies in some earlier year, without having worked out how one thing led to the other or how quickly.

At one stroke, the model-theoretic approach relieves social science of the burden of finding regularities that hold for all societies. There may be such regularities; in spite of many disappointments it may be worth looking for them; but social science is already working in perfect keeping with the current conception of natural science whenever it works out a model that fits the regularities of just one society, for example, one with money and an organized national market for agricultural commodities. (Indeed, social science does not even have to produce models to work in keeping with scientific standards: it does so just by identifying rules or regularities not identified before, or even just by deploying statistical skills to establish whether the rates of crime or of unemployment are changing.)

Models may originate in attempts, given a mass of information about various societies or various social practices, to draw a simple picture of what those cases have in common. They may be produced with a direct aim of giving a simplified or idealized account of one real case, for example, the economy of Ireland in the late nineteenth century. But then the perspective in which the model is treated is the same as if it condensed a variety of cases. It is carried from one range of phenomena to another to see where it fits.

When statements that would otherwise be taken to be statements by a theory about the world become statements describing a model, it is no longer worrisome whether they are strictly speaking true. We make them true by definition, by postulating a model that answers to the description; and we can make the model as abstract and idealized as we like. If science is in the business of producing models, then it can easily be accepted that the models will often be idealized and will not quite fit any real phenomena. It suffices for a model to come quite close, as it may do in the social sciences as well as in the natural ones, to fitting the phenomena that we want to predict and explain. We do not have to leave the meaning of being close unduly vague. We can count a model as having been improved on when we advance to making more precise predictions of the variables that specially interest us.

The model-theoretic view is also called the semantic conception because the relation of statements describing models to the models is a matter of semantics (while positivism, preoccupied with universally quantified forms of statements, inclined to emphasize the syntax of theories). Models are non-linguistic, even when they are defined in language (maybe mathematical language), without constructing anything tangible to embody the definition. A tangible model can often be furnished if it is demanded. When it is not furnished, diagrams can serve as one-dimensional or two-dimensional physical models. However, models are sometimes clearly not physical at all; logicians use the integers or the positive numbers as models. In any case, models actually produced would belong with items that language refers to rather than with items of language, just as toy trucks or table-top maps would belong with sharks and trees, not with the words ‘toy truck’, ‘map’, ‘shark’, ‘tree’.

An important philosophical issue on which different positions have arisen in the naturalistic school turns on whether all the findings of social science about groups can be reduced to findings about individual persons. Mill held that they can; chief to say that they cannot was the French philosopher and sociologist Émile Durkheim (§2), who contended strongly for a social science oriented to the discovery on the social level of facts that constrained people’s actions and did not simply combine them. Important as this issue is, it does not compel us to divide the naturalistic school into two. Division on the issue is entirely compatible with the pursuit of causal regularities, though on one view these will be (always) reducible to findings about individual persons and on the other (sometimes) irreducibly regularities about groups.

3 The several concerns of the interpretive school

Giambattista Vico (§§4-6), the pioneer of the interpretive school, held that human beings should have every confidence in their ability to understand social phenomena, because the phenomena are all of their own making. ‘Understand’ can be taken here in the broad sense in which we understand celestial phenomena better when we grasp what the red shift implies about the movement of the stars. However, it can also be taken in the sense in which we understand another person’s actions when we appreciate the hopes and fears that led to them. Max Weber (§3), highlighting ‘understanding’ of this sort (*Verstehen*), thought it went hand in hand with explanation. Indeed, any sort of understanding would, though some champions of interpretation, drastically shrinking the scope

of the term, use ‘explanation’ only for attempts at explanation in the naturalistic school.

The study of rules, which are surely of human making, invites Vico’s confidence as much as other concerns of the champions of interpretation. Widely practised, it is continually reinvigorated by new techniques like game theory and by recurrent demands for treating the state and other institutions as systems of rules (March and Olsen 1989; Ostrom 1990). It may serve as a distinguishing mark of the interpretive school. Concern with the study of rules, however, does not do much to identify the other concerns agitating the school, which in its eyes may be much more profound matters.

One such matter, which has found philosophical expression in Edmund Husserl and Martin Heidegger and their followers, is concern that the basic subjective features of human experience receive attention and respect. What does an action mean to the person doing it? This is not a question that is to be answered just by referring to the rules under which the action falls. Simple actions have subjective meanings for the people doing them, consisting of temporal structures in which their awareness of what they are doing shifts from one aspect of it to another (Carr 1986). What agents are aware of as just having done in starting up the action (moving to the door when the doorbell rings) gives way to what they are aware of doing now (turning the doorknob and swinging the door open) and this in turn to what they will be aware of having done when they finish (welcoming the person who has come to the door). Even simpler temporal structures may be embedded in these; they in turn are embedded in more elaborate ones.

A structure does not have to be very elaborate to take on aspects of a narrative. Agents may tell themselves a story as they enact it about how they started, how far they have gone, and towards what outcomes they are headed. (I left a sister behind in East Germany twenty years ago; it is she, I believe, who has just rung the doorbell; if all goes well in the reunion I shall take her into my household.) No references to rules may appear in the telling. The stories may have engrossing meaning for the agents without being for them a story about rules at all.

Narratives are examples of texts. Complex events and actions become texts themselves when the notion of texts is extended in a way that hermeneutics encourages. What do texts mean, judiciously interpreted? This is again a question that does not prima facie make references to rules prominent. It would seem, certainly at first sight, to have much more to do with exploring the larger narratives surrounding a text. (The sister’s coming is a story that fits in a number of intercrossing ways into other stories - family histories; the recent history of Germany; the history of the Cold War.) Hermeneutics as Hans-Georg Gadamer (§3) conceives it is concerned with interpreting texts under a ‘fusion of horizons’. When the texts are actions, only as much place goes to what the actions meant to the agents from what they could see at the time of doing them (with the narratives going on then) as is consistent with giving due weight also to the interpreters’ own horizon of meaning (and the narratives now current). That horizon embraces not only what has happened since the action was done, but also present expectations regarding its further consequences.

A clear example is an opinion delivered by the US Supreme Court, in *Planned Parenthood of Southeastern Pennsylvania* v. *Casey* (1991), a case on abortion. In that opinion Justice Sandra O’Connor held that whatever stand one might have taken at the time that *Roe* v. *Wade* (1973), the crucial previous case, was decided, one must recognize that millions of US women had been assuming for some twenty years that abortion was their legal right. This made a crucial difference now both to the issue and to how the previous case was to be understood.

Concern with the study of rules can be made more congenial to the other concerns at stake in interpretation by fully expressing the complexity of the concern. The interpretive school will be interested (since this is essential to discovering which rules are socially settled ones) in rules that elicit observable conformity. But social scientists who fall in with the school are not to be thought preoccupied with conformity. They are as much interested in the abuse of rules, and in violations. They are interested in rules half-formed and merely advisory precepts. Not least in importance, they are interested in the variable space that a structure of settled social rules may leave for spontaneous, inventive, and innovative action. (Professors have been lecturing in academic gowns; but some begin coming to class in sweaters and blue jeans; they get away with it.) Thus the interpretive school endorses, not so much questions about ‘What rules do given actions fall under?’, but rather ‘In what ways are social phenomena related to the presence or absence of rules?’. That allows considerable room for satisfying other interpretive concerns in the very course of studying rules. Subjective experience and narratives at all levels of generality reflect the impact of current rules, sometimes to a point of saturation or obsession, sometimes to a point of outrage and

defiance.

It is important philosophically to make the point that if social science disregards the subjective forms of experience utterly (as naturalistic inquiries often do), or disregards any of the ways in which their content escapes the study of rules, social science is bound to give an impoverished picture of human life. Yet it would not follow from this that social science could make anything of the subjective forms of experience. Social science seeks generalizations, whether they are generalizations identifying rules or generalizations identifying regularities. For a closer, more exciting approach to the subjective forms of experience people might have to go to biography, fiction, drama, poetry.

In fact social science can make something of the subjective forms of experience. Naturalistic social science conducts surveys, sometimes more penetrating in conception, sometimes less, of people’s feelings. Are people confident enough about holding their jobs to feel satisfied with the government? Confident enough to buy consumer durables? Matters like these are not mere matters of curiosity. Generalized predictions of decline or recovery in the economy may be founded upon them; and incorporating some of them in the models that political scientists have developed for explaining election results may be indispensable. Models taking into account only ‘objective’ variables like income, the rate of unemployment, housing starts, and so on, may not work.

Naturalistic social science working hand in hand with interpretive social science has something to gain from investigating narratives. Narratives shared by whole communities figure in explanations of how the respective communities are formed and maintained. It is already itself a generalization of social science that communities cannot be formed or maintained without such narratives. This generalization, moreover, opens up a branch of inquiry into comparisons of narratives. What sorts of narratives are effective with what sorts of peoples and in what sorts of contexts? If the Japanese had other equally strong sources of community feeling, they might be readier to accept the hypothesis that their imperial family was Korean in origin. Effective narratives might share certain plot-elements across cultures (like Claude Lévi-Strauss’ myth-forms, of which descent of the emperors from a divinity - the Sun-Goddess, in the Japanese case - may be an instance) (see Lévi-Strauss, C. §§5-6).

4 Critical social theory: effective with or without reduction

Does asserting that critical social theory ask naturalistic questions about regularities and interpretive questions about rules imply a ‘reduction’ of it to a combination of naturalistic and interpretive views? The term ‘reduction’ has unwanted overtones. Nevertheless, even in the strong form of saying that in its methods critical social theory is either naturalistic or interpretive, the assertion is strictly compatible with honour for other characteristic features of critical social theory. Distinctive questions, pursued with a distinctive commitment to emancipation, do not imply distinctive sorts of questions or distinctive sorts of methods.

A systematic enumeration of the components of critical social theory would include: (1) distinctive questions, for example, about ideological bias; (2) the thesis that there is no attempt at social science without commitments to political values; (3) commitment, inherited from Karl Marx (§12), to the project of social emancipation; (4) distinctive critical findings respecting both current social arrangements and the limitations of conventional social science. The components may include (5) the thesis that the very phenomena (objects) that critical social theory conceives to be suitable for study in the social sciences differ from those that conventional social science purports to study, and this thesis might extend to implying that the standards of inquiry must differ, too. Further possible components, like an abstract philosophical argument for presupposing a certain ideal of communication (the ideal speech-situation (see Habermas, J.)) as a condition of productive discourse, can be disregarded for present purposes.

Whether much remains of (5), the thesis about studying distinctive phenomena, when due allowance has been made for the other components, and for the full operation of interpretive concerns in social science, is doubtful. For these things do bring in the subjective forms of experience; and value-commitments, displayed in the simplest actions, which are done to reach goals that the agents value. Thus, if critical social scientists wish to say that social phenomena at every level are loaded with value-commitments (which naturalistic social science sometimes ignores), this point is accommodated once interpretive social science gets its place alongside social science with a naturalistic cast.

Suppose, however, this reflection is not enough to convince the proponents of critical social theory that social

phenomena as they view them are treated adequately by the combination of the inquiries that the other schools have in view. (The combination might not be sufficiently intimate or harmonious.) We could have a robust critique of current society and current social science resting on the first four components of critical social theory alone. Call this ‘para-critical social theory’. If it would accomplish substantially the same things in practical political criticism and policy guidance, how much importance attaches to the thesis about distinctive objects?

Some critical social theorists, at some stages of their thinking, would not temporize with social science done under other auspices. They would invoke the thesis of ubiquitous commitment to political values (2); and add that much of current social science is committed not to emancipation but to oppression. Social science that is not committed to emancipation should be eliminated, by vigorously criticizing it, by setting new examples, by supersession. But para-critical social theory could do this work, too.

True, even if peaceful supersession is all that is in prospect, elimination would not be cheerfully accepted in every quarter. A lot of researchers dissent from the position that a thorough-going social transformation is desirable; or favour social transformations that are not (like the project of emancipation) socialist in inspiration. Moreover, a reasonable critical social theorist would concede that criticisms from these quarters of policies at first sight attractive to socialists and emancipators sometimes make useful points, for example, in advancing market solutions to resist unnecessary expansions of bureaucracy.

The demand for elimination is excessive in other ways. The thesis about inevitable commitment to political values is an indispensable argument for elimination. This thesis, however, cannot be put through unless omitting to invoke certain values counts as favour for values opposed to them. Grant this; otherwise there are too many examples of studies or findings that do not mention the values on one side or another, much less explicitly endorse them. It does not follow that the findings coming forward with these omissions deserve no respect. One can think party competition under a capitalist regime no better than a masquerade. Findings about the effects on party competition of single-member district representation as against various schemes for proportional representation advance knowledge notwithstanding. Such findings may even turn out to have useful applications in a transformed society, supposing that such a society will leave room for contested elections and party competition.

If it will not, many political scientists would argue that the project of emancipation will lead to less freedom rather than more, and less democracy. They would not be alone. Postmodernist philosophers, reflecting on how the Marxist project of emancipation miscarried in Bolshevik hands to produce a totalitarian regime, would decry any tendency to eliminate dissent in the name of the project, or reduce the variety of inquiries, whether the elimination is to occur after the project is carried to an end or beforehand. Critical social theorists could agree. They did not contemplate repressing other sorts of inquiry. They carried on a variety of inquiries themselves. Habermas, their chief heir, has made an explicit endeavour to find a place for all the sorts of inquiry sponsored by the three schools.

5 The impact of postmodernism

Disillusion with the project of emancipation surfaced in critical social theory itself as an aspect of disillusion with the Enlightenment. The Enlightenment fostered expectations of progress in many dimensions - not just in emancipation, but also in science, in technology, in material comforts, in rational and efficient social arrangements. It thus sponsored a number of what the postmodernist philosopher Jean-Francois Lyotard (§2) calls ‘grand narratives’, all of which postmodernists categorically repudiate. Michel Foucault did the same. Without ranking himself among postmodernists, Foucault was nevertheless a major source of ideas for them and an ally. Together he and Lyotard do as much as anyone to bring postmodernist ideas to bear upon the philosophy of social science.

Foucault and Lyotard draw points from the philosophy of language and the theory of knowledge to attack, along with grand narratives, ‘globalizing’ discourses of all kinds, and with them any claim to speak for, or even to aim at, a unified and comprehensive scientific view of the world. The most that can be hoped for, they say, are limited narratives and local bodies of knowledge, confined to transient social practices. Foucault has made ambitious and arresting studies of some of these, like psychiatry, clinical medicine, and penology. To combine his terminology with Lyotard’s, such practices are so many language games in which genres of discourse figure, perhaps more than one in a given instance, and employ locutions drawn from various ‘regimes’ of locutionary acts (phrases). Thus

clinical medicine employs declaratives (from one regime), interrogatives (from another), and imperatives (from a third) in discourse that is in genre both cognitive and therapeutic. The rules from the different regimes apply, and so do the rules distinctive of the different genres. In addition, the practice itself brings in rules, for example, a rule requiring that doctors make their diagnoses public before pathologists report on dissections in the same cases.

The knowledge gained within such a practice cannot be generalized. It would be out of place in another practice; and the findings of different practices cannot as they stand be accommodated at more general levels of knowledge. Foucault, for a time at least, went along with Nietzsche in treating every claim to knowledge as a demand for power. Disputes in social science, like disputes about knowledge in other connections, become conflicts of power, or of 'wills to truth'. Lyotard comes back repeatedly to disputes about knowledge turning upon issues (*différends*) that cannot be settled without (he holds) short-changing the ideas on one side or another. Researchers from one practice confront researchers from another; no general discourse is automatically available to mediate between them.

Some postmodernists might be content with taking guerrilla action against received views, darting out to attack generalizing pretensions whenever they parade by. Given the incessant tendency for styles of inquiry to congeal into rule-bound orthodoxy, disrupting them in this way can be useful and important. It does something to restore the nonchalance about rules and creative playfulness that in the young can be mistaken for mere fecklessness.

The guerrilla role would not suit Foucault, since at stake in his theory of knowledge are his own notable contributions to the histories of practices. Let us distinguish (1) what he does in these accounts; (2) what he describes himself as doing; (3) what he has to say in general about the character of knowledge and possibilities in the social sciences. Some of what he says in (2) may be at odds with (1); some of what he says in (3) may be at odds with (1) and (2) as well, at least in part. Yet the accounts brought forward under (1) come with claims of truth, or at least some approximation of truth. It is, for example, set forth as a truth that punishment as a practice of awe-inspiring spectacles gave way late in the eighteenth century to the practice of systematic imprisonment and surveillance. Moreover, an abundance of passages falling under (2) support saying that he describes himself as aiming to discover truths (in some sense opposed to overall scepticism) in his researches into practices.

Suppose that different views do mask the wills to truth and power of different persons and groups. Suppose, as Foucault maintains repeatedly, the views of the powerful prevail and other views are subjugated, though now and then there is a change in who holds power or a change in the assortment of views. A range of situations answers to this description, from totalitarian domination at one pole to an equal-power situation at the other. Could a general consensus about truth be brought about in an equal-power situation, unlikely though this might be?

It accords with the weight of what Foucault says in (3) that there could be a local and temporary resolution of the conflict. It might take the form of mutual acknowledgements among the contenders of there being 'some truth' in what other groups hold, and a joint statement listing those partial truths. In the intermediate situations, which are much more likely, local and temporary resolutions in the form of mutual acknowledgements could occur, though they would wring only limited concessions from the views of the powerful.

This not only accords with what Foucault says; it is essential to making sense of his campaign on behalf of 'subjugated knowledges'. Some headway - no doubt short of carrying all the way to the acknowledgements of an equal-power situation - can be made in redressing received patterns of domination in favour of giving at least some attention to the subjugated views. The defects of prevailing views will in consequence abate. If the prevailing view is that solitary confinement induces penitence in prisoners, the subjugated knowledge of the prisoners that solitary confinement enrages and embitters them could be invoked to modify the prevailing view.

Lyotard goes further in making room for advances in knowledge. He may exaggerate the frequency of *différends*. He does not consider at any length how different practices may intersect, as it seems they must do for there to be an issue between them; or how different practices may get along comfortably enough even when inconsistent. Consider astronomers in everyday life speaking of sunrise and sunset, and in poetical moments entertaining the idea of an animate moon ('queen and huntress, chaste and fair'). Yet Lyotard argues that *différends* can be surmounted by conceptual-linguistic innovation that is fair to each of them. Such innovation may have been required to straighten out the foundations of the calculus; and to reconcile quantum physics with logic and the principle of contradiction. One may doubt whether it is required to reconcile naturalistic social science with

interpretive, though the spectacle of Weber and Durkheim struggling for a view of social science integrating both might give pause on this point. In any case, Lyotard's idea that even *différends* can be surmounted with conceptual innovation amplifies the conception of local resolutions between conflicting truths in the form of mutual acknowledgements; and makes the conception more credible, wherever it is to be applied.

6 The relations between the schools re-examined

It is not just the furore created by postmodernism that threatens to unsettle the tidy picture of social science given by the three schools framework. A more explicit allowance for the hermeneutical concern with texts and for even more familiar subjective forms of experience has made it harder to see how interpretive studies bear affinities with naturalistic ones. It has made it even harder to see how naturalistic studies are an indispensable supplement. Yet anything lost in this regard can easily be made up.

Postmodernists like Foucault and Lyotard bring back the study of rules, in Foucault's case, with an emphasis more thorough-going than any other philosopher who has touched on the philosophy of social science. But if there can be models of systems of rules just as there can be models of regularities, then how far a model fits a real system, in regard to any universally quantified statement that figures in the model, is a matter for quantitative - statistical - study just as much with rules as with regularities.

With these considerations, the affinity of naturalism with interpretation returns. Not only does the affinity return; it returns in a form that is ready to receive the insistence of postmodern writers on truth being local, a matter of limited context that is always subject to supersession. For a model, if it has any success at all in matching a range of phenomena, may match only the phenomena of one society or one of its subgroups; and its success, moreover, is likely to be temporary. Not only will phenomena change, but what social scientists are looking for as important points of resemblance will change, too.

To which of the three schools should postmodernists be assigned? The rules that Foucault and Lyotard would have studied are rules of discourse, passing into rules of the practices in which the genres of discourse are embodied. Studying them would embrace the study of texts, in both literal and extended senses of the notion of text. Thus Foucault and Lyotard hook up with the interest in texts of the interpretive school and its hermeneutical component. Both Foucault and Lyotard, furthermore, are specially concerned with narratives, and narratives are a species of texts much attended to in the interpretive school.

On the other hand, neither Foucault nor Lyotard has much use for the forms of subjective experience more basic than narratives. Their animus against 'subjects' is directed in some passages against 'transcendental subjects' - ideal observers or agents who Foucault and Lyotard think are implicitly postulated in references to comprehensive knowledge. (Some readers would charge them with acting as transcendental subjects themselves when they make sweeping pronouncements.) In other passages, however, the animus seems to tell against having subjects at all: it is not persons who know or act or contend with one another, but discourses and their ingredients. The interpretive school cannot easily digest this point.

A more congenial home for it can be found in the critical school. The critical school has been all along inclined (though doing so belies some of its own work) to repudiate naturalistic inquiries. What needs to be raised to the critical level in its view has been interpretive social science. Foucault and Lyotard are interpretive enough to qualify; and their campaigns against orthodox views and grand narratives have precedents in the campaigns of the critical school itself against the Enlightenment and against ideology. Foucault and Lyotard generalize the critique: it is not just power reflected in the ideology of the ruling social class that dominates uncritical thinking; it is power everywhere, exerted by various groups and various practices contending with each other, that manifests itself in claims to truth. We can, so far as we can make sense of the idea, think of subjects, both transcendental and personal, as dropping out in the course of generalizing the critique.

Another way of generalizing the critique, which again would leave the three schools standing, along with the variations in them of styles of inquiry, would be simply to demand that every inquiry in social science choose questions that are more useful rather than less so. The standard of usefulness might refer to the urgency of current social problems. On a less intensely moralizing approach, running less danger of reducing the variety of inquiries, the standard might refer also to anomalies or gaps in current theories, or even to unsatisfied curiosity. Broad or narrow, the standard would do more to discriminate between inquiries than postmodernism is able to do given a

commitment to variety hard for it to modify. Inquiries that would make only a trivial difference to existing knowledge would fail the test of curiosity.

Some things that Foucault and Lyotard say, along with things voiced elsewhere on the postmodernist scene, are perhaps too devastatingly sceptical in intention or implication to be reconciled with the pursuit of social science in any sense. We may thus have to allow in the philosophy of social science for a school that (unlike any of the three schools in the framework) holds that social science is not possible. We do not have to join it.

See also: Explanation in history and social science; Holism and individualism in history and social science; Naturalism in social science; Post-structuralism in the social sciences; Social laws; Social norms; Social science, history of philosophy of; Structuralism in social science; Value judgments in social science

DAVID BRAYBROOKE

References and further reading

Adorno, T. and Horkheimer, M. (1944) *Dialectic of Enlightenment*, trans. J. Cumming, London: Verso, 1973. (Attacks the Enlightenment within critical social theory, but already in a postmodernist vein.)

Braybrooke, D. (1987) *Philosophy of Social Science*, Englewood Cliffs, NJ: Prentice Hall.(Expounds in detail the distinction between the three-schools and their relations.)

Braybrooke, D. (1990) 'How Do I Presuppose Thee? Let Me Count the Ways: The Relation of Regularities to Rules in Social Science', *Midwest Studies in Philosophy* 15: 80-93. (Improves on the 1987 book in treating the relation of regularities to rules.)

Braybrooke, D., Brown, B. and Schotch, P.K. (1995) *Logic on the Track of Social Change*, Oxford: Clarendon Press.(Improves on the 1987 book in defining rules.)

Carr, D. (1986) *Time, Narrative, and History*, Bloomington, IN: Indiana University Press.(Treats, in a phenomenological perspective, but with a clarity of argument impressive to analytical philosophers, the concept of narrative.)

Durkheim, É. (1895) *Les règles de la méthode sociologique*, trans. S.A. Solovay and J.H. Mueller, *The Rules of Sociological Method*, Chicago, IL: University of Chicago Press, 1938.(Puts forward a classical case against an individualistic approach to social phenomena.)

Fay, B. (1996) *Contemporary Philosophy of Social Science*, Oxford: Blackwell.(Relates the interpretive view of social science simultaneously to current conflicts between cultures and to central current debates in philosophy.)

Foucault, M. (1976) 'Two Lectures' in C. Gordon (ed.) *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977*, trans. K. Soper, Brighton: Harvester Wheatsheaf, 1980, 78-108.(With next article referred to, concentrates in short texts Foucault's leading ideas.)

Foucault, M. (1977) 'Truth and Power', in C. Gordon (ed. and trans.), *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977*, Brighton: Harvester Wheatsheaf, 1980, 109-33.(With previous article referred to, concentrates in short texts Foucault's leading ideas.)

Gadamer, H.-G. (1960) *Truth and Method*, original trans. W. Glen-Doepel, revised by J. Weisheimer and D.G. Marshall, New York: Crossroad, 2nd edn, 1989.(Develops at length the perspective of hermeneutics.)

Giere, R.N. (1988) *Explaining Science*, Chicago, IL: University of Chicago Press. (Offers the most accessible treatment of the model-theoretic view of science.)

Habermas, J. (1967) *On the Logic of the Social Sciences*, trans. S.W. Nicholson and J.A. Stark, Cambridge, MA: MIT Press, 1988.(Enlarges a perspective received from critical social theory to embrace a variety of approaches in social science.)

Hempel, C.G. (1965) *Aspects of Scientific Explanation*, New York: Free Press.(Sets forth refined positivist views with unparalleled lucidity.)

Hollis, M. (1994) *The Philosophy of Social Science: An Introduction*, Cambridge: Cambridge University Press. (Explores in a lively and imaginative way the relations between naturalistic and interpretive social science.)

Lévi-Stauss, C. (1987) *Anthropology and Myth*, Oxford: Blackwell.(One source for Lévi-Strauss' theory of shared myth-form.)

Little, D. (1989) *Understanding Peasant China: Case Studies in the Philosophy of Social Science*, New Haven, CT: Yale University Press.(Looks into a number of works by social scientists in very different styles and extracts implications for philosophical issues.)

Lyotard, J.-F. (1988) *The Differend: Phrases in Dispute*, trans. G. van den Abeele, Minneapolis, MN: University of Minnesota Press.(Ranked by Lyotard as his fundamental contribution to postmodernist philosophy; very

difficult reading.)
Lyotard, J.-F. (1992) *The Postmodern Explained*, trans. D. Barry *et al.*, Minneapolis, MN: University of Minnesota Press.(Briefly and clearly surveys the main points of Lyotard’s contributions to postmodernism.)
March, J.G. and Olsen, J.P. (1989) *Rediscovering Institutions*, New York: Free Press.(Champions the conception of social institutions as systems of rules.)
Marx, K. (1845-6) *The German Ideology: Part I*, trans. S. Ryazanskaya, Moscow: Foreign Languages Publishing House, 1972.(Serves with *The Communist Manifesto* as a prime source for the Marxist conception of ideology.)
Marx, K. and Engels, F. (1848) *The Communist Manifesto*, trans. S. Moore, New York: International Publishers, 1948.(Not only a manifesto, but a brilliant synopsis of Marxist theory.)
Mill, J.S. (1843) *A System of Logic*, London: Longman, Green, 1930.(Book VI discusses with great care ‘The Logic of the Moral [Social] Sciences’.)
Ostrom, E. (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge: Cambridge University Press.(Gives a more rigorously worked-out account of institutions as systems of rules.)
Popper, K.R. (1957) *The Poverty of Historicism*, London: Routledge, 2nd edn, 1961. (In the course of attacking unscientific historical prophecy, provides a classical expression of the naturalistic view of social science.)
Rosenau, P.M. (1992) *Post-Modernism and the Social Sciences*, Princeton, NJ: Princeton University Press. (Surveys in unmatched detail, critically but not unfairly, the whole range of postmodernist thinking as related to the social sciences.)
Runciman, W.G. (1972) *A Critique of Max Weber’s Philosophy of Social Science*, Cambridge: Cambridge University Press.(Gives a lucid, brief account of Weber’s ideas and a clarified revision of them.)
Vico, G.B. (1744) *The New Science*, trans. T.G. Bergin and M.H. Fisch, Ithaca, NY: Cornell University Press, revised edn, 1984.(Protests with great vigour against all attempts to displace interpretive concerns in favour of a naturalistic treatment of social phenomena.)
Weber, M. (1904, 1905, 1917) *Max Weber on the Methodology of the Social Sciences*, three essays trans. and ed. E.A. Shils and H.A. Finch, Glencoe, IL: Free Press, 1949. (Reflects on the interrelation of interpretation and causal inquiry and on the demand for value-free social science.)
Winch, P. (1990) *The Idea of a Social Science and Its Relation to Philosophy*, London: Routledge, 2nd edn. (Champions the interpretive view with emphasis on the study of rules.)
Wright, G.H. von (1971) *Explanation and Understanding*, Ithica, NY: Cornell University Press.(Classically lucid exposition of the difference between ‘explanation’ and ‘understanding’ as key notions in natural and interpretive social science respectively.)

Social science, history of philosophy of

The history of social science can conveniently be divided into four uneven periods, starting with the beginnings of both western science and philosophy in the ancient Greek polis *(city or state). It is fair to say, with qualifications, that the debate generated by the so-called Sophists, professional teachers of rhetoric in fifth-century Athens, established what would become the central questions for the future. The fundamental issue could be put thus: is society 'natural' or is it 'conventional', a historical product of human activities which vary across time and space? The Sophists, often abused in our standard histories, supported the conventional view. They held that even if it was anthropologically necessary that* Homo sapiens *live in societies, nature was silent about the character and ends of society. They thus defended what might be called 'cultural relativism'. By contrast Aristotle argued that some men were 'naturally' slaves and that all women were 'naturally' inferior; therefore slavery and patriarchy were dictated by nature, a view that prevailed well into the early modern period.*

Beginning in the sixteenth century we find a host of thinkers who reconceived the problem first raised by the Sophists. Many of them, for example, Hobbes, Rousseau and Adam Smith, held that 'by nature' humans had similar capacities and powers. Inequalities of power were 'artificial', wholly the result of historically established conventions. These writers also rejected the idea that society was a kind of natural community. For many of them, society existed by consent, the result of a contract.

The rejection of Aristotelianism was inspired by the Copernican revolution and the new physics of Galileo and Newton. This produced a self-conscious effort by early modern writers to articulate the idea of human science, modelled on the new physics. This critical idea was well put by the physiocrat Francois Quesnay: 'All social facts are linked together in the bond of eternal, immutable, ineluctable, inevitable laws, which individuals and government would obey if they were once known to them' (Randall 1940: 323).

The third period, roughly the nineteenth century, is then a battleground over both the idea of science and the idea of a human science. The paradigm provided by celestial mechanics was nearly overwhelming; even so, there was disagreement as regards its character, especially as regards the question of causality and explanation. Until very recently, 'positivists' have tended to prevail. That is, writers have followed Auguste Comte, who gave us the terms 'positivism' and 'sociology', and who held there were social laws which were to be analysed as 'relations of invariable succession': whenever this, then that.

As regards the possibility of a human science, consciousness and the problem of a free will raised the biggest questions. Materialists found nothing special about either; idealists did. Indeed a surprising amount of the most recent debates in the philosophy of the social sciences have their roots in these issues. If, as positivists insist, activity is governed by law, then what of human freedom? On the other hand, if humans have collectively made society and thus can remake it, then what is the nature of a human science?

1 Pre-modern philosophy of the human sciences

If we are to offer a history of the philosophy of the social sciences which begins prior to the twentieth century, we must admit some anachronism. Prior to this time, perhaps the only term in such a title with its present meaning is 'history'. Only in the nineteenth century was philosophy separated from 'science', and only in the early years of the twentieth century did 'social science' emerge as a set of disciplinary practices which could become the object of 'philosophy'. This is more than a quibble over terms, since issues regarding the nature and methods of acquiring knowledge of the human condition have again and again been contested.

In what follows, we admit the anachronism and sketch what earlier writers said about such inquiry. Moreover, since the lines between a philosophy of social science, a social philosophy, and a social science are also blurred, we consider these matters rather broadly. Until lines were drawn, there were no boundaries - and even today the boundaries are fuzzy.

It is well known that both western science and philosophy began in the West in ancient Greece. The so-called 'nature philosophers', Thales to Democritus, had included in their inquiry concerns about humankind and 'society'. But this was the special concern of the Sophists, who emerged in fifth-century democratic Athens as teachers who prepared men for political careers with instruction not only in rhetoric but also in 'wisdom' (*sophia*)

and in ‘virtue’ (*aretē*). It was in this context that the *physis*/*nomos* debate arose.

Nomoi (norms, laws, customs or conventions) establish what behaviour is to be sanctioned by the community. *Physis* (nature) was the object of inquiry of the Ionian ‘scientists’ but it also referred to the particular characteristics of things, for example, the nature or ‘essence’ of man. Aristotle summarized the debate: ‘Some people [for example, the Sophists] think that all rules of justice are merely conventional, because whereas a law of nature is immutable and has the same validity everywhere, as fire burns both here and in Persia, rules of justice are seen to vary’ (1968: 295). The implications are many: If humans are alike ‘by nature’, then why do *nomoi* treat them differently? Do *nomoi* originate from ‘agreements’, from the acts of legislators, or from gods? What is the *physis* of man? Is understanding *physis* sufficient for human society, or should we follow the sophist Protagoras, who argued that since *physis* could offer no clear prescriptions, human societies require humanly constructed *nomoi*? And if, as he also argued, all men are capable of sharing in the political (social) virtues, then perhaps the Athenians were right to allow every man to offer his advice regarding the *nomoi* of the city? (See Protagoras.)

For complicated historical reasons, the Sophists acquired a very bad name. Against them, Plato insisted that experience could not sustain an understanding of human society, still less an objective set of norms for the good human life. Plato did indeed raise some challenging problems for an empirical philosophy of humankind and elements of what we call ‘platonism’ still reverberate in discussions on the human sciences. Against the Sophists, Aristotle sought to found the human sciences in nature. He insisted that ‘the *polis* exists by nature’ and ‘a man that is by nature… cityless is either low in the scale of humanity or above it’ (1977: 9).

Although Aristotle offers immense insight relevant to the idea of a human science, particularly his overall ‘naturalistic’ approach to inquiry, his most striking and probably most influential appeals to nature were, perhaps, those that reinforced the prejudices of his time. For example, ‘to rule or to be ruled (*archai kai archesthai*) are conditions not only inevitable but also expedient; in some cases things are marked out from the moment of birth to rule or to be ruled’ (1977: 19). Those to be ruled were ‘barbarians’ and all women.

Most of the current vocabulary of the social sciences comes from Greek and Latin, however we must be careful about changing meanings. For example, ‘democracy’ and ‘economics’ share but faint resemblences to their original senses. Many other critical terms derive from Latin appropriations of Greek terms, for example, republic (*res publica*, *politeia*) and citizen (*civitas*, *politeian*). By the time of Polybius’ *Histories* (*c*.150BC), the Roman republic was huge by Greek standards. Similarly, Roman citizens bore little resemblance to the Greek original. Roman citizens had what we would call ‘civil rights’, but lacked utterly what defined a citizen for a Greek: as Aristotle had put it, a right ‘to participate in ruling and judging’.

We must forego examination of both Roman and medieval discussions of matters relevant here, except to notice that because in the Augustinian eschatology, redemption depended on grace rather than knowledge, historical modes of explanation were rejected. The Arabic reintroduction of Aristotle found expression, especially in the writings of St Thomas Aquinas and Marsiglio of Padua. With Aquinas, it was impossible ‘for the truth of faith to be contrary to principles known by natural reason’. Aquinas and Marsiglio shared with Aristotle the idea that existing things are arranged in natural hierarchies, including therefore humans and their communities. All had ‘natures’ which seek their *telos* or end. Finally, Renaissance humanism sought to recover, in Christian terms, the Roman concept of *virtus*, total human excellence. For early modern thinkers, Rome became an important case study.

2 The late Renaissance and early modern period

A great structural change had begun sometime in the late medieval period. It would ultimately produce modernity: the modern state and a capitalist social order. This would be the condition and problem for all subsequent social thinkers. These remarkable changes forced reconsideration of the problems of the human condition and especially, reconsideration of what are the prime philosophical questions of relevance here: what is the object of inquiry in the human sciences and how do we study it?

We identify two major shifts, which both represent efforts to break with the past. First, there was a rejection of the set of ideas which presumed that human associations were ‘natural communities’ which existed for the sake of realizing human virtues, secular or saintly. The introduction of the terms ‘society’ and ‘state’ suggest this. Several points of difference between the new states and older organizations can be noted. These new states had

‘governments’, itself a new term, which became increasingly concerned to maintain their autonomy not only versus other organizations of similar types, but also as regards organizations within their territories. With the development of the idea of government, the problem of legitimation arose in new forms. Similarly, *societas* acquired a new sense. Originally, a voluntary coming together for some purpose, ‘civil society’ could be used instead of ‘community’, or ‘republic’ which carried ancient connotations.

The second huge shift regarded the aims of ‘science’ (knowledge) and a new concern with method. The modern writers sought ‘inductively grounded’ practical knowledge. God was still maker of the universe, but he was to be known through his works, not through contemplation and faith. Similarly, if everything was ‘governed’ by law, ‘prevision’ and thus control was possible. Method-conscious Descartes well expressed the Faustian aspect of the new learning:

> it is possible to attain knowledge which is useful in life, and that, instead of that speculative philosophy which is taught in the Schools, we may find a practical philosophy by means of which, knowing the force and the action of fire, water, air, the stars, heavens and all other bodies that environ us, we can in the same way employ them in all those uses to which they are adapted, and thus render ourselves the masters and possessors of nature.
>
> ([1637] 1931: 119)

3 Machiavelli and Bodin: redefining the polity

Exemplifying these changes, Machiavelli was perhaps the first modern ‘social scientist’. Proceeding ‘inductively’ and convinced that men of virtue could tame *fortuna*, he began his *Discourses* with the complaint that for all their admiration of the ancient writers, his generation had failed to appreciate the examples provided by antiquity. But he departed sharply from ancient writers as regards the aims of political organization. It was, he insisted, securing the ‘liberty’ (independence) and ‘security’ of the body politic. Today an uncritical given, he seems to be the first to hold that this was the defining condition of associated life.

Machiavelli suggested a choice between two sorts of republics, one that ‘desires to extend its empire, as Rome’, the other, as Sparta or Venice, that ‘confines itself merely to its own preservation’. But in the new environment, the only real choice was Rome.

It is in this context, as well, that his more famous *Prince* should be understood. As Machiavelli had said in his *Discourses*: ‘A wise mind will never censure any one for having employed extraordinary means for the purpose of establishing a kingdom or constituting a republic’ (1950: 138-9). The message of the *Prince* is technical, not moral. For the Greek, the citizen was the government, for citizens ruled. But Machiavelli assumed that the state (government) and ‘the people’ are distinct. Public persons constitute a government which exists to serve the governed persons in their private lives. Government may do this well or badly, but it cannot do it well if decisions are persistently opposed, and if it cannot act for the body as a whole. As he put the matter: government is the management of citizens so that ‘they are neither able nor disposed to oppose you’ (1950: 358). Accordingly, the governors of any well-ordered state will and should quite normally employ measures which, from the standpoint of common morality, are immoral.

What, however, was the character of this new kind of entity? Writing in 1576, Bodin answered:

> Ancient writers have called Common weals, Societies of men assembled to live well and happily together. Which as it may serve for a description of a Citie, so can it not stand for a true definition of a Commonweale.
>
> ([1576] 1962: 3)

It was not necessary that members of a commonwealth live ‘happily’ or even that they be well governed. Moreover, as individuals pursue many ends, so do commonwealths. The proper definition of a commonwealth required only that members be ‘governed by a puissant sovereign of one or many rulers; albeit that they differ in lawes, language, customes, religions, and diversity of nations’ ([1576] 1962: 49). Missing entirely is the idea that a polity is a natural community, a ‘family’ with common ends.

4 Two views of society: Hobbes, Locke and Adam Smith; Rousseau and Vico

The *nomos/physis* debate anticipated what we can now clearly identify as the two fundamentally different and still competing metatheoretical assumptions regarding human associations, one ‘holistic’, the other ‘individualist’. The former idea will not lose its persuasiveness, recurring in variant forms in Vico, Rousseau, Herder, Hegel’s variant descendants, and Durkheim. The newer conception, associated with liberal or ‘utilitarian’ social theory is presupposed by Hobbes, Locke, Adam Smith, Bentham, Mill, Herbert Spencer, Pareto and, in its modern recent variations, rational choice theory (including here modern microeconomics). The choice between the two entails a host of other important differences.

The idea that society is but an aggregation of individuals united by agreements to accept the authority of certain conditions for acting is brilliantly developed by Thomas Hobbes, who also broke decisively with the past in postulating both a universal human nature and the idea that ‘by nature’ humans differ little in terms of their powers and abilities. Postulating endless desire, individuals everywhere and everywhen ‘seek power after power’. Accordingly, the ‘natural condition of mankind’ is a condition of war. But if individuals alienate their ‘right to all things’ to a sovereign power, peace can be secured. Since this is the ‘rational’ thing for mutually self-interested individuals to do, civil society is ‘explained’ and sovereign power is justified as the cement of society. Hobbes was, accordingly, the first ‘law and order’ theorist.

Hobbes, Locke and Rousseau are thought of as ‘contract’ theorists, but the differences are critical. Perhaps Locke’s greatest achievement, usually not acknowledged, was his ‘explanation’ of private property, which, it seems, is the only one available that is also a serious defence (see Locke, J. §10). Locke believed that God had ‘given the earth to Mankind in common’, but he rejected Hobbes’ idea that the right to property was merely conventional. If our bodies and, thus, our labour was ours, what we ‘mix our labour’ with is ours. Although this is a highly problematic idea, neither Proudhon, who believed that property was theft, nor Marx would have cause to disagree. The problem comes when the two moral conditions imposed on the free appropriation from nature are put into question. We should not waste, nor can we appropriate when such appropriation puts others at a disadvantage. Locke argued that money was introduced by our consent. This allowed unlimited appropriation. What is produced in excess of need can be sold to those who are unable to appropriate freely from nature. Since it is their right to sell their labour power to landowners, no one, presumably, is put at any disadvantage. Class inequality, accordingly, was an unintended consequence of the voluntary introduction and use of money. And all this prior to the ‘consent’ which constituted civil society.

Rousseau’s views stand in marked contrast. He shared with Vico the view that one could not ‘explain’ society in terms of the ‘contract’ and he further suggested that the psychological qualities attributed to mankind by Hobbes and the liberals who followed him were social constructions (see Rousseau, J.-J.). In his *Discourse on the Origins of Inequality* (1755), Rousseau bitterly criticized his predecessors on just these grounds. For him, ‘society’ emerged ‘naturally’, the result of the development of language. The true founder of civil society, he ironically wrote, was the person who having staked out an area of land, declared, ‘This is mine’ and found people ‘simple enough’ to believe him. Rousseau seems to have been the first to argue that human history is largely the product of unintended consequences, and he is one of the very few who insisted that the changes were not progressive. For him, humans had unwittingly forged chains which imprisoned them. This was largely the result of the interdependence which genuine humanness had brought. The solution was the social contract. Since sovereignty was for him inalienable, and governments had only executive power, Rousseau, a radical democrat, insisted that unless people take collective control of the conditions of their lives, tyranny is quite inevitable (see Contractarianism §6).

We can here compare Vico’s *The New Science* (1725). For Vico, culture and its artefacts are all creations of human consciousness. He thus insisted, anticipating Dilthey, that humans can understand themselves and all that they have created. Like Rousseau, but no democrat, Vico rejected ‘contract’ theory as incapable of explaining society. He argued that history, which was cyclical, set different problems for different ages (see Vico, G.). Marx was impressed by Vico’s view that an ‘inductive’ historical ‘science’ could emancipate people by emancipating them from superstition and by displaying their true nature.

Adam Smith’s rational individualism puts him in direct opposition to Vico and Rousseau (see Smith, A.). If for Hobbes statutory law turns the violent competition of ‘the natural condition’ into peace, and if for Rousseau, the division of labour is the fundamental cause of human self-alienation, for Smith, by means of ‘the invisible hand’,

the division of labour gives ‘self-love’ distinctly beneficial unintended consequences. Indeed, Smith made the effort to explain the division of labour with one universal psychological principle: the ‘propensity to truck, barter and exchange’. By this time, Smith could take private property for granted, along with the idea, still uncritically assumed, of the market.

If Smith was the founder of political economy, it is also important to notice that he did not conceive of its domain narrowly. When *The Wealth of Nations* was published in 1776, ‘political economy’ referred to ‘the art or practical science of managing the resources of nations’. This is why some 60 per cent of *The Wealth of Nations* is historical and sociological, and why from the point of view of modern economic thinking this portion is utterly irrelevant.

John Stuart Mill’s 1829 essay, ‘On the definition of political economy; and the method of investigation proper to it’ (1844) settled several of the then unsettled questions. Thereafter, political economy would be thought of as an abstract, ahistorical science which treats the ‘laws’ that regulate the production, accumulation and distribution of commodities (see Economics, philosophy of).

5 Progress and positivism: Comte and Herder

Auguste Comte was an heir to the French Enlightenment, but between Turgot and Montesquieu and Comte’s *Cours de Philosophie Positiv* (*The Positive Philosophy of Auguste Comte* [1830-42]) came the French Revolution and the rumblings of anarchism and communism. Not only was the thought of Rousseau and Montesquieu rendered irrelevant, but the events raised the question of whether there could be both order and progress. Comte said yes - if politicians would accept the truths of science, understood by him in distinctly positivist terms (see Comte, A.).

Comte held that there are three fundamental ‘epochs of the mind of the race’, three stages: the theological or fictitious, the metaphysical or abstract, and the scientific or positive. The theological mind ‘supposes all phenomena to be produced by the immediate action of supernatural beings’. The metaphysical mind ‘supposes, instead of supernatural beings, abstract forces… inherent in all things, and capable of producing all phenomena’. The positivist mind, by contrast, has not only ‘given over the vain search for Absolute notions, the origin and destination of the universe’, but also ‘the causes of phenomena’. Instead, it ‘applies itself to the study of their laws - that is, their invariable relations of succession and resemblance’ (Comte [1830-42] 1875: I, 2).

There is here a clear contrast between a positivist sense of law and causality and a distinctly realist conception - causes as productive powers. A realist conception, ‘metaphysical’ for the positivist, is common to ordinary thinking, and shorn of what are, today, unwelcome teleological commitments, it has been given increasing attention since the 1970s. According to positivists, we do not observe productive powers ‘inherent in things’. Following Hume, we ‘observe’ relations of succession: if this, then that. A realist and teleological view was argued by Aristotle and in the tradition which followed him. It is beautifully illustrated, with many of its problems, in the writings of Johann Herder, one of those who articulated a philosophy of human science which fell by the wayside.

Influenced by Rousseau and Montesquieu, Herder expressed a desire to be ‘the Newton of history’. He begins with the realist idea that ‘whenever we observe a power (*Kraft*) in operation, it is inherent in some organ and in harmony with it’. But ‘power as such is not open to investigation, at least not by our senses. It exists for these by its manifestations… ’. This is then applied: ‘The human essence - *Humanität* - is not ready made, yet is potentially realizable’ ([1794-91] 1969: 266). As with individuals, linguistically interacting people - *Völker* - have environments, ‘climates’, with which they epigenetically transact. Moreover, ‘whatever the climatic influence, every human being, every animal, every plant has a climate of its own; for each absorbs and adapts to external influences in its own organic manner’ (Barnard 1965: 120). A multiplicity of equally valuable *Völk*, each in transaction with its ‘climate’, has created a human world, many concrete realizations of our shared *Humanität*. Indeed, there are no races since ‘complexions run into each other… *in toto* they are, in the final analysis, but different shades of the same great picture which extends though all ages and all parts of the earth’ (Herder [1794-1] 1969: 284). Herder’s vision was anarchist and cosmopolitan, not statist and chauvinist: each *Völk* has its own ‘spirit’, its own ‘genius’. Herder emphatically rejected Eurocentric notions of progress, which for him were but ‘figments’, constructed from ‘embroidered or invented facts’.

Herder lost, Comte’s Eurocentrism and positivist conception of a human science carried the day. Comte articulated what are, appropriately, the defining ideas of positivism: not only should we reject the search for causes as

productive powers, but also, 'what is now understood when we speak of an explanation of facts is simply the establishment of a connection between single phenomena and some general facts… ' (Comte 1875: I, 2) - what is now called 'the covering law model of scientific explanation'.

But not all of Comte's views carried the day. Since humans are essentially social, there is no science between physiology and sociology. Second, not only is society not 'decomposible into individuals' but 'there can be no scientific study of society, either in its conditions or its movements, if it is separated into portions, and its divisions are studied apart' (Lenzer 1975: 228). Comte would not have approved of the present disciplinary division of the human sciences, nor its characteristic 'methodological individualism' (see Positivism in the social sciences §1).

6 Mill's *Logic*

This is not true of J.S. Mill, who was writing his *System of Logic* (1843) when he read Comte. While he accepted the inevitability of progress, he was dissatisfied with Comte's law of progress. If it was a law, then, it had to be 'invariant' and therefore there had to be as Comte had said, 'a necessarily invariable development of all humanity'. Mill could not see how this was possible.

Mill distinguished 'laws of nature' from 'empirical laws'. With Comte, Mill would leave to the metaphysicians any talk about 'the ultimate mode of production of phenomena, and of every other question regarding the nature of "Things in themselves"' ([1843] 1930, III, ch. V). But if 'empirical laws' are 'uniformities which observation and experience has shown to exist', what makes them different from laws? Mill distinguishes 'ultimate' from 'derivative laws'. Derivative laws are the result of 'a collocation of causes'. Derivative laws are 'resolved' into ultimate laws and are, by subsumption, explained by them. The law of inertia is 'ultimate', an unexplained explainer. Kepler's laws are derivative.

This may not do, especially if the covering law model of explanation is rejected. One might ask, for example, whether there are any empirical uniformities which are not the product of a collocation of causes? And, if so, perhaps the problem resides in the analysis of causality?

Mill affirmed that there were 'laws of society', but they 'can be nothing but the laws of the actions and passions of human beings' ([1843] 1930, VI, ch. VII). In addition, then, to the universal 'laws of the mind', there were 'derivative laws', the object of 'ethology', Mill's science of character. History, then, can only be explained as 'collocations of causes', a point made by Weber who with greater consistency also emphatically rejected the covering law model of explanation.

7 Karl Marx: one human science?

Although every author discussed in this entry has been variously interpreted, the situation is especially difficult as regards Marx's philosophy of the human sciences (see Marx, K.). Not only were many of his important texts unknown until this century, but different interpretations give rise to different Marxist politics. Two widely diverging interpretations are first, the conventional understanding, shared by the mainstream on both sides of the historic confrontation between capitalism and Marxian socialism, and the second which became prominent in the 1970s.

On the 'orthodox' view, Marxism is a positivism and a materialism. As a positivism, it has a deterministic theory of history in which societies must pass through stages. As a materialism, it holds that 'the economic structure of society [is] the real foundation on which arises the legal and political superstructure and to which corresponds definite forms of social consciousness'. A 'dialectic' then explains change. Thus, 'at a certain stage of development, the material productive forces of society come into contact with existing relations of production…. From forms of development of the productive forces these relations turn into fetters. Then begins an era of social revolution' ([1859] 1904:12). As in Comte, agency drops out.

On the alternative interpretation, Hegel is taken far more seriously, but on this reading, Marx rejects the idealism/materialism problematic, offering instead a naturalism that acknowledged the active constitutive side of idealist epistemology while holding to the causal demands of the independently existing 'material' reality. On this view, Marx rejected the reified abstractions of idealism. Thus, superstructural properties, for example, law and ideology, have no independent existence and, versus the 'ideologists', there can be no independent histories of law

and ideas. There are only 'real individuals, their activity and the material conditions under which they live, both those which they find already existing and those produced by their activity' (1947: 7). This text, from *The German Ideology*, published only in 1883, along with the much later published *Paris Manuscripts* and *Grundrisse* offer the best textual support for this strongly agency-centred reading of Marx's philosophy of social science. It is summarized in the oft-quoted text from Marx's *Eighteenth Brumaire*: 'Men make their own history, but they do not make it just as they please; they do not make it under circumstances chosen by themselves, but under circumstances directly encountered, given and transmitted from the past' ([1852] 1963: 15). Similarly, 'stage theory' is rejected: As Marx wrote in response to a reviewer of *Capital*: 'my critic… feels that he absolutely must metamorphose my historical sketch of the genesis of capitalism in Western Europe into an historico-philosophical theory of the general path every people is fated to tread, whatever the historical circumstances in which it finds itself… But I beg his pardon' ([1877] 1942: 354).

Following on this interpretation is a strongly realist reading of *Capital*. Volume 1 provides an abstract model of capitalist reproduction, shorn of all the particularities and contingencies which shape concrete capitalist societies. For example, it is a system-law that by virtue of the structural relations which define capitalism, there will be a tendency for the falling rate of profit. But this tendency may not be actualized, just as the inertial tendency of a planet to move rectilinearly is not realized. There are other causes at work in the world: For example, capitalists, struggling to reduce costs (and thus reducing the rate of exploitation) will move capital where labour is cheaper.

Finally, since conscious beings make history, it becomes critical to consider how their understanding enters into social reproduction/transformation. In all societies, production is social and people are more or less interdependent. But the form which products take when production is organized for generalized exchange is the commodity. Under this condition, however, 'the relations connecting the labour of one individual with that of the rest appear, not as direct social relations between persons at work, but as what they really are, material relations between persons and social relations between things' ([1867] 1970: I, 1). Such 'fetishism' is essential to capitalist reproduction, and on this reading, its analysis is continuous with his earlier analysis of 'alienation'.

8 Dilthey and historicism

Marx and Engels asserted that 'we know only one science, the science of history'. But as 'naturalists', this did not require a fundamentally different epistemology. The writings of Wilhelm Dilthey are in the background of most subsequent 'anti-naturalisms' in the philosophy of the social sciences.

For him, the natural sciences had emancipated themselves from metaphysics - in positivist fashion. This would not do for *Geisteswissenschaften*, a term introduced into German as a translation of Mill's 'moral sciences'. In the *Geisteswissenschaften* we are not dealing with 'representations' of an unknowable external reality, but with the objects of consciousness themselves. As Iggers writes: 'Understanding [*Verstehen*] is possible… because life "objectivates" itself in such institutions as the family, civil society, state and law, art, religion and philosophy. As products of life and spirit, these institutions can be understood' (1983: 139).

This 'solution' to the problem of the foundations of knowledge in the human sciences is not without difficulties, which is acknowledged by Dilthey. Thus, 'this task led me to the most general problems; a seemingly insoluble contradiction arises if we pursue historical consciousness to its last consequences. The finitude of every historical phenomenon, whether it be a religion, an ideal, or a philosophic system, hence the relativity of every sort of human conception about the connectedness of things, is the last word of the historical world view' (Iggers 1983: 143-4).

For Hegel, 'spirit' was 'objective' and constituted a specific stage in historical development. This could overcome both subjectivism and relativism, but while Dilthey seems to have moved in this direction, it is doubtful that he finally did overcome the 'seemingly insoluble contradiction'.

9 Max Weber: human science as concrete

A consequence of the Cold War was the view that Marx and Weber polarize on the idealism/materialism dichotomy. There is good reason, however, to believe that Weber, influenced by Neo-Kantian philosophy also rejected this dichotomy. For him, *Verstehen* (understanding) was unproblematic, 'a transcendental assumption' of both everyday life and of any attempt to understand it. Like Marx, Weber rejected Hegelian versions of history in which concrete reality 'emanated' from abstract ideas which had independent existences. Finally, while Weber

rejected the reductionist monocausal understanding of history put forward by the Second International, one of the most important aims of the journal he edited was 'the advancement of the economic interpretation of history'.

Weber made a central distinction between two kinds of sciences: abstract or nomothetic and concrete or historical, a distinction which in variant forms begins with Otto Ranke and is found in Dilthey, Rickart, Menger, Simmel and Windelband, who gave us the pair, 'nomothetic/idiographic'.

For Weber, the two kinds of science satisfy different interests and require different methodologies, but crucially the distinction does divide the natural and human sciences. The purpose of 'nomological science' is 'generalized abstraction' and the elimination of 'purely contingent facts'. Its concepts are 'generic'. Thus, the laws of pure mechanics are true of everything; but such knowledge is none the less 'worth knowing'. The human sciences could be nomological, but assertions true of all societies would not be interesting. Indeed, 'the theses that the ideal of science is the reduction of empirical reality to "laws" is meaningless'. The concrete sciences have as their aim 'the descriptive reproduction of reality in its full actuality'. We want knowledge of things 'which we regard as essential because of their individual peculiarities' (Weber [1903-6] 1975: 57). In the human sciences, we need, not generic concepts, but ideal-type concepts, rich in intention and particular in extension. An ideal-type is 'like a utopia which has been arrived at by the analytical accentuation of certain elements of reality', including 'typical modes of action' (Weber[1904] 1949: 90). The construction of such concepts is the main task of sociology. Ideal-types solve the problem of causal explanation because they enable us 'to become aware of the characteristic meaning of single, concrete cultural elements together with their concrete causes and effects' (Weber [1904] 1949: 65). It follows also that explanation is not by subsumption under law.

10 Spencer and Engels: two monisms

While Darwin's *Origin of Species* (1859) surely established 'evolution' as a prominent way of thinking, Herbert Spencer was already insisting that evolution ensured that 'progress… is not an accident, but a necessity'. Spencer 'induced' one evolution 'going on everywhere after the same manner….There is a change from an incoherent homogeneity to a coherent heterogeneity, accompanying the dissipation of motion and the integration of matter' ([1862] 1976: 325). As regards society, this issued in an ever-perfecting functional harmony: 'As surely as the tree becomes bulky when it stands alone, and slender if one of a group… so surely must the human faculties be moulded into complete fitness for a social state… so surely must man become perfect' ([1850] 1897: 32).

Spencer's views could be classified as both positivist and monistic naturalist. They compare in this regard to those of his contemporary, Friedrich Engels. The 'old materialists' (Dühring, Büchner) could not explain novelty. This could be overcome with a 'dialectical' materialism: 'the science of the general laws of motion and development in nature, human society and thought' (Engels [1885] 1939: 155). These laws 'reduce' to three: 'the law of the transformation of quantify into quality, and vice versa; the law of the interpenetration of opposites; [and] the law of the negation of the negation' ([1927] 1940: 26). These are all in Hegel, but his 'mistake lies in the fact that these laws are foisted on nature and history and laws of thought, and not deduced [inferred] from them' ([1927] 1940: 26).

Engels, like Spencer, is aware that his super-generalizations are abstract and do not 'say anything concerning the particular processes of development' ([1885] 1939: 154), nor for the same reason, do they say anything about the particular mechanisms of these processes. But if so, then as with Spencer, the most obvious problem is their vacuousness.

11 Pareto and Durkheim: two positivisms

Gustav Schmoller lost the *Methodenstreit* (dispute of methods) to Karl Menger and the new neoclassical economists. As Schumpeter (1954) has argued, for all the technical innovations, marginalism did not represent a paradigm shift from earlier classical political economy. Three differences, all consequences of a more self-conscious positivism, can be noted. First, hints of realism as regards causality largely disappear. Second, as Pareto put the matter, 'establishing a theory is something like passing a curve through a number of fixed points' ([1906] 1971: 31); and third, 'normative propositions' must be sharply distinguished from 'existential propositions'.

Durkheim and Pareto both accepted positivist versions of law and theory. In consequence, both contributed to the

idea that quantitative social research was the main task of the social scientist. Both thought of a society as a system although they differed fundamentally on its character.

Pareto argued that ‘the form of society is determined by all the elements acting upon it, and it, in turn, reacts upon them’ ([1915-19] 1935: para. 2060). The economic system, the domain of ‘rational’ behaviour, is a subsystem of the social system, embedded in a natural environment. But features of this are accounted for indirectly, since their effects are felt on individuals, the units of the social system. Individuals are possessed of ‘sentiments’, Pareto’s term for the various manifestations of non-rational behaviour, so that action is a joint outcome of the ‘rational’ and ‘non-rational’. Generalizing general equilibrium theory, Pareto warmly endorsed the dream of positivist social science:

> In order thoroughly to grasp the form of society in its every detail it would be necessary to know what all the very numerous elements are, and then to know how they function - and that in quantitative terms… The number of equations would be equal to the number of unknowns and would determine them exhaustively.
>
> ([1915-19] 1935: para. 2062)

In contrast to Weber (and more recent complexity theory), the social system, like the system of planets, is a closed system. Accordingly, the ‘laws’ of society can be mathematical functions of variables. This vision, usually unspoken, still pervades inquiry in social science.

Émile Durkheim saw no usefulness in economics, nor for that matter in any of the ‘utilitarian’ sorts of theory which Hobbes had initiated. He agreed with Comte, also, that ‘since it does not call for one social form rather than another’, psychology cannot explain any of them ([1895] 1938: 108). Instead, there are irreducible ‘social facts’ - that is, ‘ways of acting or thinking with the peculiar characteristic of exercising a coercive influence on individual consciousness’ ([1895] 1938: liii). Social facts are ‘external’ and depend on the collective consciousness which has its own laws. This road, which connects also to Platonism, leads directly to twentieth-century structuralism.

Durkheim was a functionalist, but he was also a critic of the functionalism which he attributed to Comte and Spencer (see Functionalism in social science). Once having shown that some social form was ‘useful’ in so far as it satisfied some ‘need’, they were content. Missing was an account of ‘how it originates or why it was what it is’ ([1895] 1938: 90). Moreover, a social fact can exist without being useful at all. But Durkheim did not escape the progressivism of Comte and Spencer. Societies, like organisms, have a lawful ‘development’. ‘*Anomie*’ was ‘abnormal’, the product of the breakdown of religion and authority which had subordinated modern industry. Indeed, ‘what is needed if social order is to reign is that the mass of men be content with their lot… And for this, it is absolutely necessary that there be an authority whose superiority they acknowledge and which tells them what is right’ (1959: 200).

12 The twentieth century

Employing available intellectual materials and exploiting the new opportunities in the new institutions of higher education in the USA, Americans constructed the prevailing disciplinary division of the human sciences (Manicas 1987). By the early 1920s these were reconstituted on a manifestly positivist reading of ‘hard science’. This was then profoundly reinforced by ‘logical positivism’, imported from Vienna.

By the 1950s, the philosophical debate had polarized between positivists and a minority of anti-naturalists who rejected the idea of a human science, at least as understood by positivists (Natanson 1963). Thus Popper, Hempel, Brodbeck and others continued to defend the relevance of the covering law model for social science, even while it was being attacked by sympathetic critics. Except for marginalized Marxists and a few non-Marxist dissenters, for example C.W. Mills (1959), it was business as usual in the departments of the academy.

Kuhn (1967) devastated foundationist epistemology and generated a radical sociological challenge to it (Barnes 1982), but he did not raise questions about either causality or explanation. A third alternative, a non-positivist, realist understanding of science, did. Hinted at by Toulmin, Scriven, Sellers, Hanson and others, it was given perhaps its first clear expression by Harré (1970) and then by Bhaskar (1975). It was applied to social science by Harré and Secord (1973) and influenced by the fresh readings of Marx, by Keat and Urry (1975), Benton (1977), Bhaskar (1979) and Outhwaite (1983).

This enabled a reconsideration of the ancient ontological debate regarding the nature of society. It continued to be polarized across two dimensions, on the one hand between 'methodological individualists' standing in the tradition of Hobbes (Gellner 1956; Watkins 1957; Homans 1967) and functionalist 'holists', generally influenced by Durkheim (Mandelbaum 1955). On the other hand between 'subjectivist' approaches in, for example, Schütz (1966) and Garfinkel (1967) and variant 'objectivist' structural approaches which included the very different views of, for example, Lévi-Strauss (1962), Althusser (1965) and Alexander (1987).

The more or less explicitly realist efforts to reconcile these turned on re-examining the status of 'agents' as conscious beings engaged in practical activity, an extended gloss on Marx's remark that persons make history but not with materials of their choosing. On Bhaskar's 'transformational model of social action', social forms predate the actions of agents which reproduce and transform them. Giddens' 'structuration' view (1976; 1980), less explicitly realist, defines social structure in terms of the 'rules and resources' available to agents in acting. Social structure thus has but 'virtual existence'.

Some of the same polarities were confronted in re-examining the question of culture (Foucault 1969; Bourdieu and Passeron 1970; Williams 1977). Feminist voices (Harding and Hintikka 1983) were also increasingly important. In turn, 'cultural studies' was being influenced by developments in 'postmodern' theory, including here a 'decentring of the subject' and a strongly anti-naturalistic turn towards 'discourse' (Lacan 1977; Derrida 1970). As regards the philosophy of the social sciences, by the 1990s the hegemony of positivism could no longer be taken for granted.

See also: Explanation in history and social science; Holism and individualism in history and social science; Ibn Khaldun §2; Naturalism in social science; Nature and convention; Social laws; Social science, contemporary philosophy of; Society, concept of

PETER T. MANICAS

References and further reading

Alexander, J. (1987) *Theoretical Logic in Sociology*, 4 vols, Berkeley, CA: University of California Press. (Alexander begins his extended review with an account of positivism, proceeds in vol. 2 to discuss Marx and Durkheim and then in the following volumes, what he sees to be two efforts at synthesis: Weber and Parsons. Extensive bibliography.)

Althusser, L. (1965) *For Marx*, trans. B. Brewster, New York: Vintage, 1970.(The critical book in provoking the idea that there is a decisive 'break' in the thought of Marx; introduces the idea of 'structure in dominance'.)

Aristotle (*c.* mid 4th century BC) *Politics*, trans. H. Rackham, Cambridge, MA: Harvard University Press, 1977. (This is certainly the most important of Aristotle's work as regards the nature and organization of human associations.)

Aristotle (*c.* mid 4th century BC) *Nichomachean Ethics*, trans. H. Rackham, Cambridge, MA: Harvard University Press, 1968.(Contains much of interest as regards Aristotle's influential views on human nature and conduct. Still relevant.)

Barnard, F.M. (1965) *Herder's Social and Political Thought*, Oxford: Clarendon Press.(A very useful selection of key texts.)

Barnes, B. (1982) *T.S Kuhn and Social Science*, New York: Columbia University Press.(Barnes, an originator of the so-called 'strong programme' in the sociology of science, shows why Kuhn's work was important to a sociological understanding of social science.)

Benton, T. (1977) *Philosophical Foundations of the Three Sociologies*, London: Routledge.(Discusses 'positivism', 'humanism' (or anti-naturalism in social science) and an alternative which derives from Marx.)

Berlin, I. (1976) *Vico and Herder*, New York: Viking.(An accessible and sensitive comparison.)

Bhaskar, R. (1975) *A Realist Theory of Science*, Atlantic Highlands, NJ: Humanities Press, 2nd edn, 1978.(A systematic effort at 'realist' theory of science, versus Humean causality and the covering law model of explanation.)

Bhaskar, R. (1979) *The Possibility of Naturalism*, Atlantic Highlands, NJ: Humanities Press.(Applies 'realism' to develop a 'critical naturalist' theory of social science in which 'society is both the ever-present *condition* (material cause) and the continually reproduced *outcome* of human agency' (p. 43).)

Bodin, J. (1576) *Six Books of a Commonweale*, ed. with an intro. by K.D. McRae, Cambridge, MA: Harvard University Press, 1962.(A facsimile reprint of the English translation of 1606. Offers the earliest radical reconceptualization of the political association as that had been derived from the Greeks. Spellings of the

citations in §3 have been modernized from the old English.)
Bourdieu, P. and Passeron, C. (1970) *Reproduction in Education, Society and Culture*, trans. R. Nice, London: Sage, 1977.(An alternative to the dominating assumption that 'culture' and 'society' are analytically distinguishable.)
Carver, T. (1983) *Marx and Engels: The Intellectual Relationship*, Brighton: Wheatsheaf Books.(The best treatment of the very critical differences in their respective philosophies.)
Comte, A. (1830-42) *The Positive Philosophy of August Comte*, freely trans. and condensed by H. Martineau, London: Trubner and Co., 2 vols, 2nd edn, 1875. (Comte's philosophy is the first and perhaps most influential effort to define social science as science, conceived as the search for invariant lawful regularities.)
Darwin, C. (1859) *Origin of Species*, New York: Cambridge University Press, 1981. (The classic of modern evolutionary theory.)
Derrida, J. (1970) 'Structure, Sign and Play in the Discourse of the Human Sciences', repr. in *A Postmodern Reader*, ed. J. Natoli and L. Hutcheon, Albany, NY: State University of New York Press, 1993.(Although so-called 'post-modernism' is a highly contestable notion, Derrida's work is surely critical. This is an accessible essay which argues against most of the deep assumptions of mainstream social science.)
Descartes, R. (1637) *Philosophical Works of Descartes*, vol. 1, trans. E.S. Haldane and G.R.T. Ross, New York: Dover Publications, 1931.(Descartes' *Discourses* powerfully influenced thinkers of the modern age who sought to break away from the inherited legacy of Aristotle and the Schools.)
Dilthey, W. (1910) *Pattern and Meaning in History*, trans. unacknowledged, ed. H.P. Rickman, New York: Harper & Row, 1962.(A useful, if brief collection.)
Durkheim, É. (1895) *The Rules of Sociological Method*, New York: Free Press, 1938.(A book which, perhaps, most influenced the main strands of post Second World War sociological thought.)
Durkheim, É. (1958) *Socialism and Saint Simon*, Yellow Springs, OH.: Antioch Press; London: Routledge & Kegan Paul, 1959.(From lectures given from November 1895 to May 1896. A critique of the 'founders' of socialist theory which in some ways, at least, attempts a reproachment of Comte and Marx.)
Engels, F. (1885) *Herr Eugen Dühring's Revolution in Science*, New York: International Publishers, 1939.(This book, along with *The Dialectics of Nature* (below) established the dominating conception of Marxian social science as a 'dialectical materialism'. See Carver 1983.)
Engels, F. (1927) *Dialectics of Nature*, New York: International Publishers, 1940.(The best statement of Engels' philosophy of science.)
Foucault, M. (1969) *The Order of Things: An Archaeology of the Human Sciences*, New York: Pantheon, 1970. (In this influential, but very difficult book, Foucault, another of those who have propelled 'postmodernism', aims to show how different 'discourses' - of life, labour and language - get 'structured' and why we view them as providing the universals of social life - a view which he emphatically rejects.)
Frisby, D. and Sayer, D. (1986) *Society*, London and New York: Tavistock.(The best introductory account of the 'career' of the concept of society. Though brief, all the major figures are intelligently and clearly explicated.)
Garfinkel, H. (1967) *Studies in Ethnomethodology*, Englewood Cliffs, NJ: Prentice Hall.(A fundamental critique of the 'normative model' which has dominated sociology since Parsons.)
Gellner, E. (1956) 'Holism Versus Individualism', repr. in *Readings in the Philosophy of the Social Sciences*, ed. M. Brodbeck, New York: Macmillan, 1968. (This volume well represents the hegemony of positivist philosophy of social science during this period.)
Giddens, A. (1976) *New Rules of Sociological Method*, London: Hutchinson.(In this rather brief book, Giddens discusses critically the main themes and some of the main figures of philosophy of social science: Schütz, Garfinkel, hermeneutics, Durkheim, Parsons and Marx and concludes with a criticism of positivism.)
Giddens, A. (1980) *The Constitution of Society*, Berkeley, CA: University of California Press.(Giddens' most thoroughgoing statement of his 'structuration' theory - his effort to resolve the continuing dichotomies of social theory. Important.)
Harding, S. and Hintikka, M. (eds) (1983) *Discovering Reality*, Boston, MA: Reidel.(One of the better collections of interventions in epistemology by feminists.)
Harré, R. (1970) *The Principles of Scientific Thinking*, Chicago, IL: University of Chicago Press.(Perhaps a 'Copernician revolution' in the philosophy of science; it provoked a range of new 'realist' thinking.)
Harré, R. and Secord, P. (1973) *The Explanation of Social Behavior*, Oxford: Blackwell.(The 'ethnogenic' approach of this book offered a clear alternative to most mainstream conceptualizations.)
Herder, J.G. (1794-91) *J.G. Herder on Social and Political Culture*, trans. ed., and with an intro. by F.M. Barnard, Cambridge: Cambridge University Press, 1969.(Very useful for the general student.)
Hobbes, T. (1651) *Leviathan*, New York: E.P. Dutton, 1950.(The source of much modern political theory.

Indispensable.)
Homans, G. (1967) *The Nature of Social Science*, New York: Harcourt Brace.(An important work by a recent 'behaviouralist' critic of 'holistic' (structuralist) views of sociology.)
Iggers, G.G. (1983) *The German Conception of History: the National Tradition of Historical Thought from Herder to the Present*, Middletown, CT: Wesleyan University Press.(An excellent overview of this critical history.)
Keat, R. and Urry, J. (1975) *Social Theory as Science*, London: Routledge & Kegan Paul.('Realist' theory.)
Kerferd, G.B. (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press.(Gives a sympathetic account of the central issues discussed in §1, above.)
Kuhn, T. (1967) *The Structure of Scientific Revolution*, Chicago, IL: University of Chicago Press, enlarged edn, 1970.(A book which upset the long legacy of empiricist, positivist thinking about the nature of science, including social science. Essential reading.)
Lacan, J. (1977) *Ecrits, A Selection*, trans. J. Sheridan, New York: Norton, 1977.(A 'decentring of the subject' which draws on Freud's theory of the unconscious. Often cryptic.)
Lenzer, G. (ed.) (1975) *Auguste Comte and Positivism: The Essential Writings*, New York: Harper, 1975.(A very convenient student's edition. Judicious editing.)
Lévi-Strauss, C. (1962) *The Savage Mind*, Chicago, IL: University of Chicago Press, 1966.(A strongly 'structuralist' but anti-functionalist approach to social science.)
Locke, J. (1690) *Two Treatises of Civil Government*, with an intro. and *appartus criticus* by P. Laslett, 2nd edn, Cambridge: Cambridge University Press, 1967.(The foundation of modern constitutionalism and liberalism.)
Machiavelli, N. (1515, 1527) *The Prince and the Discourses*, New York: Modern Library, 1950.(*The Discourses* are less well known than *The Prince* since it was the latter which has shocked readers; but contrary to widespread opinion, for reasons set out in *The Discourses*, most leaders of modern states have learned from Machiavelli.)
Makkreel, R.F. (1975) *Dilthey: Philosopher of the Human Studies*, Princeton, NJ: Princeton University Press. (Perhaps the most important recent commentary on Dilthey.)
Mandelbaum, M. (1955) 'Societal Facts', *The British Journal of Sociology* 6 (4): 305-17.(A 'classic' defence of 'holism' against 'methodological individualism'.)
Mandelbaum, M. (1971) *Reason, Man and History*, Baltimore, MD: Johns Hopkins University Press.(Perhaps the best review of nineteenth-century writers relevant to the human sciences. Important and accessible.)
Manicas, P.T. (1987) *A History and Philosophy of the Social Sciences*, Oxford: Blackwell.(A fuller treatment of the history beginning with the eighteenth century.)
Marx, K. (1852) *The Eighteen Brumaire of Louis Bonaparte*, New York: International Publishers, 1963.(Probably the best concrete discussion of class and its role in social change in the corpus of Marx's volumious writings.)
Marx, K. (1859) *A Contribution to the Critique of Political Economy*, Chicago, IL: Charles H. Kerr, 1904.(The 'Preface' has been extremely influential in defining the character of Marx's 'historical materialism'.)
Marx, K. (1867) *Capital*, vol. 1, London: Lawrence & Wishart, 1970.(Certainly one of the most influential books of social science, very much incorporated into current theory - by Marxists and non-Marxists alike.)
Marx, K. and Engels, F. (1883) *The German Ideology*, New York: International Publishers, 1947.(For many interpretators, this is the most critical text in gaining an understanding of 'historical materialism'.)
Marx, K. and Engels, F. (1877) *Marx and Engels Selected Correspondence*, New York: International Publishers, 1942.(One of several collections.)
Mill, J.S. (1844) *Essays on Some Unsettled Questions of Political Economy*, 2nd edn, Clifton, NJ: Augustus Kelly, 1974.(Now little read, these lively essays show that Mill would not have approved of subsequent developments in the philosophy of 'economics'.)
Mill, J.S. (1843) *System of Logic*, 8th edn, London: Longman, 1930.(A monumental, influential and still useful effort to set out the 'logic' of the human sciences. Introduces his famous 'methods'.)
Mills, C.W. (1959) *The Sociological Imagination*, Harmondsworth: Penguin.(The title, now very much in the literature, is often hardly taken seriously; a hardnosed attack on 'abstracted empiricism' and 'grand theory', both still very much in evidence.)
Natanson, M. (ed.) (1963) *Philosophy of the Social Sciences*, New York: Random House.(The editor made his selections based on the assumption that there were two polar positions underlying the social sciences, an 'objective' versus a 'subjective' point of view. Contains most of the most important essays which define this debate. Very influential.)
Outhwaite, W. (1983) *New Philosophies of Science: Realism, Hermeneutics and Critical Theory*, New York: St Martins Press, 1987.(Parallels in some ways Benton 1977.)

Pareto, W. (1915-19) *Mind and Society*, New York: Harcourt, 1935.(This massive and often tedious work is now largely forgotten even while there is a considerable residue of its driving vision.)

Pareto, W. (1906) *Manual of Political Economy*, New York: Kelly, 1971.(A clear statement of Pareto's philosophy of social science.)

Passmore, J. (1957) *A Hundred Years of Philosophy*, Harmondsworth: Penguin, 1968. (Certainly the best account of the often ignored late nineteenth-century 'philosopher/physicists' whose views powerfully influenced the idea of science which was appropriated by the emerging modern social sciences. Passmore is a reliable expositor of all the figures he treats.)

Randall, J.H. (1940) *The Making of the Modern Mind*, Cambridge, MA: Riverside Press.(Randall gives a superb account of how thinking about society shifted from the medieval period to modernity. Very readable.)

Ross, D. (1991) *The Origins of American Social Science*, Cambridge: Cambridge University Press.(A detailed study of the critical period, 1870-1929, of the constitution of US disciplinary social science.)

Rousseau, J.-J. (1755) *The Social Contract and Discourses*, trans. G.P.H. Cole, New York: E.P. Dutton, 1950. (Rousseau's writings are essential, but are less often read in terms of philosophy of social science.)

Sayer, A. (1984) *Method in Social Science: A Realist Approach*, 2nd edn, London: Routledge, 1992.(A very useful realist textbook on method.)

Schumpeter, J.A. (1954) *History of Economic Analysis*, New York: Oxford University Press.(A rich historical source, but not helpful as regards the *Methodenstreit*.)

Schütz, A. (1966) *The Collected Papers*, 3 vols, The Hague: Niijhoff.(Schütz was central in the effort to restore 'subjectivity' to social scientific understanding - versus the reigning positivism. Difficult.)

Shanley, M.L. and Pateman, C. (eds) (1991) *Feminist Interpretations and Political Theory*, College Park, PA: Pennsylvania State University Press.(An excellent collection of feminist re-examinations of many of the figures discussed in this entry.)

Skinner, Q. (1978) *The Foundations of Modern Political Thought*, 2 vols, Cambridge: Cambridge University Press, 1978.(A detailed study with attention to the development of the concept of the modern state, a datum for all subsequent social scientific thinking.)

Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, eds R.H. Campbell, A.S. Skinner and W.B. Todd, Oxford: Clarendon Press, 2 vols, 1976.(A classic, but as much sociological and historical as economic.)

Spencer, H. (1850) *Social Statics*, New York: Appleton, 1897.(Spencer's best effort to defend '*laissez-faire*'.)

Spencer, H. (1862) *First Principles*, Westport, CT: Greenwood Press, 1976.(Aims to show that nature and history may be understood in terms of fundamental, evolutionary principles.)

Vico, G.B. (1725) *The New Science*, trans. T.G. Bergin and M. Fisch, Ithaca, NY: Cornell University Press, 1948. (Usually ignored, Vico's work is strikingly modern in many ways.)

Watkins, J.W.N. (1957) 'Historical Explanation in the Social Sciences', *British Journal for the Philosophy of Science* 8: 104-17.(Perhaps the 'classic' statement of 'methodological individualism'.)

Weber, M. (1903-06) *Roscher and Kneis: The Logical Problems of Historical Economics*, trans. with an intro. by G. Oakes, New York: Free Press, 1975.(This book contains Weber's most systematic effort to think through problems in the philosophy of social science, but it is not often read. Hard going.)

Weber, M. (1904) '"Objectivity" in Social Science', in *The Methodology of the Social Sciences*, trans. E.A. Shils and H.A. Finch, New York: Free Press, 1949.(This volume contains the more famous of Weber's discussions of social science.)

Williams, R. (1977) *Marxism and Literature*, Oxford: Oxford University Press.(A very influential effort to overcome the 'reductionism' of much Marxist theory.)

Social science, methodology of

Each of the sciences, the physical, biological, social and behavioural, have emerged from philosophy in a process that began in the time of Euclid and Plato. These sciences have left a legacy to philosophy of problems that they have been unable to deal with, either as nascent or as mature disciplines. Some of these problems are common to all sciences, some restricted to one of the four general divisions mentioned above, and some of these philosophical problems bear on only one or another of the special sciences.

If the natural sciences have been of concern to philosophers longer than the social sciences, this is simply because the former are older disciplines. It is only in the last century that the social sciences have emerged as distinct subjects in their currently recognizable state. Some of the problems in the philosophy of social science are older than these disciplines, in part because these problems have their origins in nineteenth-century philosophy of history. Of course the full flowering of the philosophy of science dates from the emergence of the logical positivists in the 1920s. Although the logical positivists' philosophy of science has often been accused of being satisfied with a one-sided diet of physics, in fact their interest in the social sciences was at least as great as their interest in physical science. Indeed, as the pre-eminent arena for the application of prescriptions drawn from the study of physics, social science always held a place of special importance for philosophers of science.

Even those who reject the role of prescription from the philosophy of physics, cannot deny the relevance of epistemology and metaphysics for the social sciences. Scientific change may be the result of many factors, only some of them cognitive. However, scientific advance is driven by the interaction of data and theory. Data controls the theories we adopt and the direction in which we refine them. Theory directs and constrains both the sort of experiments that are done to collect data and the apparatus with which they are undertaken: research design is driven by theory, and so is methodological prescription. But what drives research design in disciplines that are only in their infancy, or in which for some other reason, there is a theoretical vacuum? In the absence of theory how does the scientist decide on what the discipline is trying to explain, what its standards of explanatory adequacy are, and what counts as the data that will help decide between theories? In such cases there are only two things scientists have to go on: successful theories and methods in other disciplines which are thought to be relevant to the nascent discipline, and the epistemology and metaphysics which underwrites the relevance of these theories and methods. This makes philosophy of special importance to the social sciences. The role of philosophy in guiding research in a theoretical vacuum makes the most fundamental question of the philosophy of science whether the social sciences can, do, or should employ to a greater or lesser degree the same methods as those of the natural sciences? Note that this question presupposes that we have already accurately identified the methods of natural science. If we have not yet done so, the question becomes largely academic. For many philosophers of social science the question of what the methods of natural science are was long answered by the logical positivist philosophy of physical science. And the increasing adoption of such methods by empirical, mathematical, and experimental social scientists raised a second central question for philosophers: why had these methods so apparently successful in natural science been apparently far less successful when self-consciously adapted to the research agendas of the several social sciences?

One traditional answer begins with the assumption that human behaviour or action and its consequences are simply not amenable to scientific study, because they are the results of free will, or less radically, because the significant kinds or categories into which social events must be classed are unique in a way that makes non-trivial general theories about them impossible. These answers immediately raise some of the most difficult problems of metaphysics and epistemology: the nature of the mind, the thesis of determinism, and the analysis of causation. Even less radical explanations for the differences between social and natural sciences raise these fundamental questions of philosophy.

Once the consensus on the adequacy of a positivist philosophy of natural science gave way in the late 1960s, these central questions of the philosophy of social science became far more difficult ones to answer. Not only was the benchmark of what counts as science lost, but the measure of progress became so obscure that it was no longer uncontroversial to claim that the social sciences' rate of progress was any different from that of natural science.

1 Testability

As in natural science, for a long time one leading aim of many philosophers of social science was to improve and ensure the ‘scientific’ character of theories in social and behavioural science by increasing their testability. The demand for testability was characteristic of logical positivist philosophy of social science. Philosophers like Neurath insisted that much which passes for social science was pseudoscience because there was no empirical evidence which could bear on its truth or falsity (see Positivism in the social sciences §2). Even more vigorously, Karl Popper and his followers long wielded the requirement that scientific theories be in principle falsifiable as a litmus test to find fault with Freudian psychodynamic theory, Marxian economics, and functional sociological theory. Operationalists following Bridgman insisted that for any term introduced in a scientific theory, there be specified operations to tell whether the term applied to something or not (see Operationalism). Empiricist-oriented social scientists, behaviourists in psychology or sociology like B.F. Skinner or G. Homans, neoclassical economists like P. Samuelson, or Milton Friedman, biologically inspired anthropologists like M. Harris, and a whole host of behaviouralists in political science, attempted to reorient theory in order to meet the positivist demand that it be possible to state explicitly what sort of evidence would confirm or disconfirm it (see Behaviourism in the social sciences §§1, 3). One result of the demand for testability was a suspicion of theoretical terms, neologisms, and metaphorical expressions invoked without definitions that could link them to observations providing these terms with empirical content.

Positivism as a philosophy of science receded largely because the positivists and their students were unable to provide an account of empirical testability which satisfied their own standards of adequacy. All specific formulations were found either too narrow - excluding as untestable such scientifically indispensable concepts as ‘positive charge’ or ‘acid’, or too broad - including as testable such pseudoscientific concepts as ‘the absolute’. With no adequate criterion of testability, the demand that theories satisfy strictures on empirical verifiability came gradually to be surrendered in the philosophy of social science. With the surrender of the search for such principles, the inclination to stigmatize some social theories as meaningless declined and the prescriptive role of the philosophy of social science weakened as well. Nevertheless, there remains among some philosophers and many social scientists a healthy suspicion of explanatory terms and theories embodying them, when these theories lack significant predictive power.

Among some social scientists, especially economists, the supposition that there is a test for scientific as opposed to non-scientific discourse or metaphysics or pseudoscience lasted far longer than it did among empiricist philosophers of science. And though the demand for testability may have had a salutary effect on the minimization of cant and circumlocution, it could no more be honoured by significant theory in social science than it could in physics or chemistry. This has made much of the official philosophy of science of disciplines like neoclassical economics particularly schizoid. Among economists Popper’s demand that theory be falsifiable has long held sway in the methodological volumes while being steadily ignored in actual economic theorizing. The demand for falsifiability, like that of testability, has been shown to be a ‘will-o’-the wisp’, in large part because falsifying evidence always impugns a large set of hypotheses, not just the particular one the scientist sets out to test. Falsification casts doubt as well on all the auxiliary hypotheses governing the behaviour of measuring instruments, establishing the character of the boundary or initial conditions under test, and assuring the reliability of the data of a test as a reflection of the phenomena under study.

The failure of various principles of testability to be reflected in the actual scientific practice of economics and other disciplines, increasingly resulted in a greater interest by philosophers in the actual character of theory appraisal as pursued by social scientists in the several sciences than in the methodological expressions of opinion by social scientists. This increased attention to what social scientists do, as opposed to what they say they should do, has characterized post-positivist philosophy of social science.

2 Methodological individualism versus functionalism

The question of what the objects of social theories are emerged from the controversy over testability, but has taken on a life of its own. Even before Durkheim, historians had adverted to groups, institutions, and organizations whose existence and behaviour could not be exhaustively analysed into or explained by appeal to the existence and behaviour of the individual agents who participate in them. From Durkheim onward such hypostasis was based on the explanatory role and apparently autonomous coercive powers such wholes seemed to have (see Holism and individualism in history and social science).

To those who were sceptical about the notion that there were, in Durkheim's terms, social facts above and beyond individual ones, the slogan that the whole is greater than the sum of its parts was simply mystery-mongering provocation. Foremost among social scientific opponents of this holism were the economists, whose own theory rigorously rejected any dispensation from the obligation to explain all economic phenomena by derivation from the rational choices of individual agents. Empiricist philosophers of social science make common cause with such social scientists: since social wholes, facts, institutions, and so on, cannot be directly observed, and since ontologically their existence depends wholly on the existence of agents who compose them, it is tempting to claim that such wholes must be reducible by definition or by explanation to the behaviour of individuals. Allowing such wholes autonomous existence is not only metaphysically bootless, but epistemically gratuitous.

With the demise of demands for testability much of the force of the purely epistemological argument against holism must fade. Thus contemporary debates about methodological individualism versus holism turn on the explanatory indispensability of the appeal to trans-individual entities within social science and biology.

When asking the question 'why biology?' it is important to bear in mind that over the second half of the twentieth century lessons and morals drawn from biology and the philosophy of biology have bulked large in disputes within the philosophy of social science. Just as philosophers have become more interested in the actual practice of social science, so too have they acquired an interest in biology's methodological agenda. The reasons for this interest overlap those of the philosopher of social science. Examination of biological methodology must shed light on the adequacy of a philosophy of science drawn from the examination of physics, and differences between biology and physics will show that prescriptions drawn from physics are inapplicable in biology. This will reinforce the argument that such inapplicability in social science shows the irrelevance of methodological strictures drawn from physical science beyond its official writ. This is a point to which we return below.

Meanwhile, as in social science, evolutionary biology seems to countenance holistic theories, ones which postulate the existence and causal role of family groups, inter-breeding populations, species and ecosystems independent of the individual organisms that compose them. If such postulation is legitimate in biology, the argument proceeds, it should be equally legitimate in social science. On the other hand, some methodological individualists argue the very reverse: first they show that evolutionary explanation in biology is implicitly or explicitly individualist, and then they argue that it must be so in social science as well.

Whether the argument from biology tells in favour of or against the methodological individualist in social science hinges on the correct understanding of teleological explanation and functional analysis in biology. This is another reason why philosophers of social science closely follow or participate in debates among philosophers of biology.

Durkheim was among the earliest explicitly to attribute functions to social institutions above and beyond those individuals ordain or recognize. Durkheim and functionalist social scientists who follow him hold that social wholes exist in order to fulfil certain functions and that the fact that they fulfil such functions can actually explain their existence. Their argument for this view simply followed biological precedent: such explanations are widespread and universally accepted in life science. For example, the presence of the heart is explained by citing its function, the pumping of blood. By contrast the presence of the appendix in humans is problematical, just because it has no function. *Mutatis mutandis* in social theory: social structures exist in order to curb anomie, minimize alienation, optimize social integration, or fulfil some other function vital to the society's survival or well being.

At least since Spinoza philosophers and others have been suspicious of functional and other forms of purposive explanation just because they seem to reverse the order of causation (see Functional explanation; Functionalism in social science). To explain the presence of the heart by appeal to its pumping the blood is to cite an effect to explain its cause. As Spinoza said, this is to reverse the order of nature. Worse, many things with functions do not fulfil their purposes. For instance, most sperm cells fail to fertilize any egg, yet that is their function. Can we explain the presence of a particular sperm cell by adverting to a function it fails to perform? Add to this the difficulty of identifying functions, both in the biological case and the social one: what exactly is the function of the peacock's plumage, or matrilineal marriage rules in a patrilateral society? Many empiricist philosophers of biology have looked to the theory of natural selection to legitimize, eliminate and minimize the ontological excesses of functional explanation in biology, and by parity of reasoning the problematical role of such explanation in social science. The theory of natural selection is sometimes invoked to legitimate functional explanation and sometimes

to show that teleology is a mere appearance - an overlay we place on purely causal processes. Either way the analysis is roughly the same: the appearance or the reality of adaptation is the result of a long and slow process of random variation in hereditary traits subjected to the culling of nature in the struggle for survival. In a relatively constant environment over a long period of successive generations, variants that are fortuitously advantageous will be selected and will increase their proportion in the whole of a population until the variants become ubiquitous. Thus, to say mammals have hearts in order to pump blood is true and explanatory because the heart exists in mammals now as the result of a long process of selection for blood pumps.

Once teleology is legitimated in biology, it is mere cavilling to withhold it from the social sciences. The trouble is that proponents of its legitimacy in social science are obliged to establish close analogies between mechanisms of biological and social evolution. Here the problem for the methodological holist is that functional explanations in biology are widely held to substantiate a version of methodological individualism. By and large there is agreement among evolutionary biologists that selection does not operate at the level of the group, but only at the level of the individual, that the mechanism of hereditary transmission operates exclusively at the level of the individual and that group properties must be fully explained by appeal to the properties and interactions of individuals. Functional explanation in biology seems resolutely individualist.

This result is disturbing for holists. Their best argument for the existence of social wholes and their properties is the explanatory role of social roles and their properties. But these properties - say kinship rules - are usually accorded a functional role, for example, that of enhancing social integration, and their existence is explained by citing these functions. If functions have explanatory force only to the extent that their existence can in turn be given a thoroughly individualist foundation, then for the holist the game is not worth the candle. The best a holist can hope for is that the drift of the debate in the philosophy of biology, which has long favoured the individualist, may turn in the other direction. Short of that, holists have no other recourse but to repudiate the relevance of biology in the justification of social methodology. To do this is of course to surrender ‘naturalism’ - the doctrine that the methods of natural and social sciences should be broadly similar. This is an approach many social scientists have further independent reason to contemplate, as we shall now see.

3 Intentionality laws and the philosophy of psychology

As noted above, perhaps the central questions in the philosophy of social science are whether the social sciences do or should employ the same methods as the natural sciences. The thesis that the social sciences should and can do so is traditionally labelled naturalism (see Naturalism in social science). Its denial has been borne by a succession of movements in social science and its philosophy: structuralism, semiotics, ethnomethodology, hermeneutics, interpretationalism, and more recently deconstruction, postmodernism, and rhetoric (see Post-structuralism in the social sciences §2; Structuralism in social science). These approaches differ from each other in crucial ways. But they share in common a rejection of naturalism.

The issue on which naturalism and anti-naturalism have longest contended is that of the nature of human action and its explanation. All parties agree that social science shares a presumption, along with history and common sense, that much human behaviour - the displacement of the human body - is action, that action is explained by the joint operation of desires and beliefs, and that action and its consequences are the concern of the social sciences. This is a theory which, on the one hand, is so obvious as to go unmentioned in history and biography, and, on the other hand, has been transformed into the economist’s formalization as rational choice theory: each individual has a set of preferences - that is desires - which set is complete, transitive, and continuous, and a set of expectations - beliefs - about available means of satisfying these preferences, and each individual is rational - that is, chooses that action which, in the light of expectations, will attain the object of strongest desire (see Rational choice theory). If explanation of human action is common sense and all the social sciences appeal to this principle, formalized or not, then the vindication of naturalism or its refutation turns on whether some version of the theory of rational choice is a causal law or not. Similarly the adequacy of anti-naturalism hinges on whether such explanations’ force lies in some non-causal power to illuminate their explanans.

This is because scientific explanation is held crucially to involve derivation of the explananda from causal laws and initial or boundary conditions. Though this analysis of explanation in science has been significantly qualified and circumscribed over the years, the notion that explanation involves systematizing laws and/or causality remains a fixed point in the philosophy of natural science. Accordingly, if a principle of rational action is indispensable in

social explanation, then naturalism requires that the principle's explanatory force rests on a causal or nomological understanding. It will have to be a causal law that:

L (x) (if x desires d, and x believes that all things considered doing action a is the most efficient means of attaining desire d, then x does a).

Or if this is not a law, it had better be improvable in the direction of a law of human behaviour. The requirement that a principle of rational choice must be treated as a causal regularity has been common to naturalists in the social sciences at least since it was implicitly embraced by Max Weber. Treating rational choice theory as at least a body of incipient laws is especially crucial to any attempt to make sense of the actual practice of economic theory. Given the predictive weakness of economic theory and its resolute attachment to rational choice theory as its explanatory core, it becomes crucial to economists' claims for their science to show that this core can be retained as a body of laws or approximations to them in the face of apparent empirical disconfirmation.

However the debate on whether a principle like L embodies a law or not is a chapter of what was once the philosophy of mind and has become the philosophy of psychology. For, as Hume long ago noted, for L to be a law, beliefs, desires, and actions must behave in the way that causes and effects recognized in the natural sciences behave. They must be logically independent of one another and it must be possible to establish that each of a particular package of belief, desire and consequent action, obtain without having to establish that the other two obtain. However, the logical connection between these states is something that has long been recognized in the philosophy of psychology. Beliefs and desires have what Brentano unmetaphorically described as 'aboutness', or content or intentionality (see Intentionality §1). Beliefs are about actual or (in the case of false beliefs) non-actual states of the world, as are desires; actions are distinct from mere bodily motion only because they are the effects of desires and beliefs. Thus, action too is imbued with intentionality. Intentionality is an obstacle to a naturalistic treatment of the psychological for two reasons. First, beliefs and desires are identified and distinguished from one another by their intentional content; but the only way to establish the content of a person's beliefs, given their actions, is to have prior knowledge of all their desires and all other beliefs, and vice versa. Moreover, in order to infer desire and belief from action one must be able to distinguish the action which the combination of the desire and belief cause from mere behaviour, and this cannot be done unless we can establish that the body's motion constitutes action - that is movement caused by belief and desire. We are thus caught in an intentional circle. There is no way independent of the effects of action to establish that its causes obtain, and vice versa. But causes and effects must be logically distinguishable from each other; more than that they must be methodologically distinguishable. Otherwise there will be no way to test the claim that some combination of desire and belief causes some particular action, and consequently no way to systematically apply our theory of rational choice in the prediction of actions with any hope of improvement over common sense.

Of course if we could establish when a certain combination of desires and beliefs obtain independent of the actions they cause, for example by reading mental states off from the neurological states of the brain, we could in principle identify causes and effects independently. Alas, philosophers and psychologists from Descartes to Skinner have rejected this course as unavailing either in principle or in practice. Despite the scientific conviction that the mind is the brain, no one has yet overcome Descartes' arguments against this claim, nor shown to general satisfaction how intentional states could be physical states. And even were this obstacle overcome, as Skinner long argued, neuroscience can provide no real help in establishing the boundary conditions to which intentional generalizations like those of rational choice are applied for explanation and prediction.

In the latter half of the twentieth century the philosophy of social science was dominated by the dispute about whether the intentional explanations of common sense, history, cultural anthropology, a good deal of political science and sociology, or their economic and psychological formalization, are or could be causal laws. Naturalists proclaimed that they were, but with the coming of Wittgenstein's influence in the philosophy of mind, this view came increasingly under attack. The gist of this attack on naturalism was that since intentional states and actions are logically connected with one another, explanations that appealed to them could not be causal. Instead the explanatory force of such explanations had to have different non-naturalistic foundations. As Wittgenstein wrote in the *Philosophical Investigations*, 'In psychology there are experimental methods and conceptual confusions' (1953, pt XIV: 2325). The conceptual confusion was to treat belief and desire as the causes of action, and to approach the generalization which links them as an empirical generalization which might be improved and refined

so as to become a law. In fact, beliefs and desires are linked to actions as their reasons, and the linkage is established by rules. It is to this confusion of rules for regularities that the sterility and vacuity of much social science is attributable. Rules are learned by asking the right questions, not by making experimental observations of behaviour. When we attempt to apply empirical methods to what is in essence a conceptual inquiry the result is bound to be unsatisfactory.

Opponents of naturalism before and since Wittgenstein have been animated by the notion that the aims of social science are not causal explanation and improving prediction, but uncovering rules that make social life intelligible to its participants (see Explanation in history and social science). For these purposes there is no alternative to ‘folk psychology’, the account of action and its sources implicit in our every day beliefs and ubiquitous in all cultures. Folk psychology is what enables us to ‘interpret’ the behaviour of others, to show it rational or reasonable by our own lights. If we fail to so understand the actions of others, then by and large the fault is not in our ‘theory’ but in our application of it. We have misdiagnosed the beliefs and the desires of those we seek to understand. From this view the goal of social inquiry is to be understood on the model of the cultural anthropologist ‘going native’ in order to learn the meaning of day-to-day life for the participant of the culture under study (see Anthropology, philosophy of §1). These meanings are embodied in rules. Indeed the principle of rationality is but a rule which may characterize only western culture. Note that rules may be broken, and when they are there are further rules which dictate the sanctions to be imposed, and so forth. The social scientist’s objective is to uncover these rules which render what happens in a society intelligible, though they never make social life predictable beyond limits set by common-sense folk psychology.

That social science is a search for intelligibility will explain why its theories ought not be construed causally and why it neither embodies nor needs explanatory and predictive laws. Thus, the failure of empirically inspired social science to uncover generalizations about human action that seem reasonable candidates for laws or even their rough and improvable precursors is explained not by the complexity of human behaviour and its intentional causes, but by the fact that such a search for laws misunderstands the aim of social science and the role which concepts like rationality sometimes play in attaining this aim. This aim is an interpretation of actions and events which gives them meaning - sometimes the participant’s meaning, sometimes a ‘deeper’ meaning, but always one which presupposes a significant overlap in most of the broader beliefs and desires of the participants and observers as well. However, by adapting the arguments of Quine, anti-naturalists have concluded that the sort of interpretation sought is inevitably underdetermined by the evidence, defeasible, and is a construction subject to negotiation among interested parties.

Under various names this anti-naturalist view of the aims and claims of social science has waxed and waned in fashion along with naturalism from Dilthey and Weber to the present. There is, however, a third alternative to asserting or denying that beliefs and desires work together to explain behaviour as action through causal laws. Eliminativism is an empiricist philosophy of social, behavioural and cognitive science which adopts the interpretationalist conclusion that under their intentional characterizations beliefs and desires cannot be linked to behaviour via laws; so much the worse for beliefs and desires (see Eliminativism). Since our goal in social science is causal explanation and improving precision and scope of prediction, we should surrender any hope of explaining behaviour as action caused by intentionally characterized psychological attitudes. Instead we should adopt a neuroscientific or some other sort of non-intentional perspective on individual behaviour, and we should seek aggregate generalizations about social processes while remaining agnostic as to their psychological foundations.

Eliminativism has found few defenders in either social science or its philosophy. Faced with the Scylla of interpretationalism and the Charybdis of eliminativism, most naturalists among social scientists and philosophers participating in this debate have challenged a presumption shared by both eliminativists and interpretationalists: the claim that on its naturalistic interpretation rational choice theory cannot reasonably be construed as embodying empirical generalizations with any prospect of being improved into causal laws. Instead naturalists point with pride to relatively controversial and certainly limited successes whether in economics or political science or sociology. More to the present point, naturalists dispute the philosophical theses from which the interpretationalist and the eliminativists draw their pessimistic conclusions about the prospect for intentional laws. Some naturalists dispute the requirement that the description of causes be logically distinct from those of their effects. Others argue that despite ineliminable *ceteris paribus* clauses, the status of the rationality principle as a general law is unimpeachable, or at least no more problematical than *ceteris paribus* laws elsewhere, in biology for example.

These philosophers of social science continue to seek the obstacles to predictive power in the social sciences in the complexity and interdependence of human behaviour, and in the openness of social processes to interference by exogenous forces. Others seek the limitations on predictive strength of social science in the reflexivity of social phenomena.

4 Reflexivity and historicism

One set of problems that the social sciences do not share with the natural sciences stem from the fact that the subjects of social science are themselves epistemic agents, that can be influenced by their own beliefs about the generalizations, the input and the output of social theory (see Social science, prediction in). It is well known that the publication of an economic prediction can result in its being falsified by the action of agents who act on the prediction, or again that the publication of polling results can make their predictions self-fulfilling. This fact has led to an entire movement in macroeconomics, 'rational expectations' theory, which accords the subjects of economic theory - individual rational agents - at least as much knowledge of economic theory as it accords government policy makers. Under this assumption it can be shown that implementing certain policies predicated on the truth of Keynesian theory will cause that theory to be falsified.

Theories and predictions whose dissemination can effect their confirmation are known as 'reflexive' ones. Some philosophers and social scientists have held that theories in social science must be different in kind from those of natural science because the former are potentially reflexive when the latter are not. One extreme version of this view denies that social theory can improve prediction beyond certain narrow limits, and so constitutes a serious obstacle to a naturalistic social science uncovering a succession of improvable causal generalizations. Among advocates of critical theory the doctrine is held not so much to limit knowledge, but to burden it with a special responsibility. Critical theorists hold that since theory can influence social behaviour it ought to, or at least the social scientist has the responsibility of framing and disseminating theory in a way that will emancipate people by showing them the real meaning of social institutions and freeing them from false and enslaving beliefs about themselves and their communities (see Critical theory). Critical theorists adopt an empiricist account of natural science, but they insist that the reflexive character and normative aims of social science make for important differences in its methods and epistemology: social science is a search for emancipative intelligibility, not laws, or even rules. Its methods are social criticism of ideologies masquerading as fixed truths, mere interpretations which are to be unmasked as social constructs, not the inevitable result of natural forces.

As critical theory and other anti-naturalistic doctrines give way to more radical doctrines, even its concessions to the empirical character of natural science are withdrawn, in favour of an inference from the non-empirical character of social science to the ideological, negotiated non-cumulative character of natural science. Postmodernism, helping itself to a radical interpretation of Kuhn's theory of scientific change, not only denies that the social sciences have a methodology or an epistemology, but also that natural science yields objective information about the way the world or even observable phenomena are.

These more extreme anti-naturalistic doctrines share with nineteenth-century philosophy of history a commitment to 'historicism' (see Historicism). The natural sciences have traditionally been viewed as ahistorical in a number of different senses. Natural laws are typically time symmetrical: given laws and initial conditions for a closed deterministic system, we can retrodict past events as well as future ones. There is no causal action at a temporal distance: a past event can only influence a future one transitively through their intermediate effects at each intervening event between the cause and its ultimate effect. If there are causal laws their writ runs across all places and all times: a regularity confirmed in one spatio-temporal region and disconfirmed in another is not a law in either region, it is at best a local empirical regularity to be explained by some exceptionless universal law.

Historicism in social science involves the denial that one or another of these three theses characterizes the social sciences. Typically historicists argue that social processes reflect the operation of asymmetrical principles which mandate a fixed order of events: thus capitalism could not have come into existence without the prior appearance of feudalism; and adult neurosis could not have come into existence without the prior frustration of infantile sexuality. Historicism sometimes also embodies the thesis that each historical epoch operates in accordance with its own distinct explanatory laws, and that sometimes the discovery of such laws can usher in a new era with new laws - whence the connection between historicism and reflexivity.

Reflexivity raises important methodological questions: Can we minimize the obstacles to empirical testing that self-fulfilling prophesies and suicidal predictions make? If not, how do we assess the cognitive status of explanations embodying reflexive theories? Historicism raises metaphysical questions of an even more fundamental sort: Suppose historicism is true, what is it about human action, social institutions, and large scale historical events that makes them so different from causal processes governed by symmetrical causal laws, and how can events reach out to produce much later ones without going through normal causal chains? These questions either go unanswered in contemporary debates or lead us back to problems about the intersection of the philosophy of social science and the philosophy of psychology: problems about the nature of intentionality and the mind. For the answers given to these questions about the historical character of social inquiry invariably appeal to agency, action, will and thought.

5 Facts, values and dangerous knowledge

One traditional set of questions distinctive of the social sciences reflects their special relevance to normative questions about individual and social policy (see Value judgments in social science). Well confirmed theories of human behaviour provide the means to ameliorate, or to worsen human life. This fact raises questions about how this knowledge should be employed. In addition, in choosing which research questions to examine, and which hypotheses to test, the social scientist makes value judgments about the interest, importance, significance of alternatives. However these questions are in principle no different from those raised by advances in theoretical physics. Different questions are raised by the fact that well-confirmed social theories would enable us to control and manipulate individual and aggregate human behaviour. A further distinctive problem is raised by the need to employ human subjects in order to test some theories of human behaviour. Whether it is permissible to treat humans in the way we treat laboratory animals, or even animals in field studies, is an issue social scientists face along with medical researchers.

Additionally, there is the problem of potentially dangerous knowledge. Some inquiries are best left unmade, or so it is alleged. Studies of the correlation of criminality and chromosomal abnormality, or the heritability of intelligence, or the dynamics of jury deliberation, will be condemned on the ground that disseminating the findings may be harmful whether they are scientifically significant or not. Over the last half century this claim has been made with special force in regard to studies of the statistical heritability of IQ among various groups. Some philosophers have attacked such studies on cognitive grounds, arguing that IQ is not a measure of intelligence, and that heritability shows little about genetic inheritance. They have also held that the examination of such questions should be forsworn because merely pursuing them is socially inflammatory and the results, even of well-conceived and well-executed studies, are likely to be misused. Whether this prospect raises special questions beyond those faced in research on toxic chemical substances or harmful disease vectors that might be inadvertently released is one the philosopher of social science needs to debate.

Beyond the normative issues to which social science is relevant, there is the further debate among philosophers of social science about whether claims in the social sciences are themselves explicitly, implicitly, inevitably evaluative, normative, prescriptive, or otherwise ‘value laden’. Arguments for this claim often turn on the allegedly evaluative meaning of certain descriptive and explanatory concepts indispensable to varying social theories. For example, the term ‘rational’ has positive connotations, as do expressions like ‘functional’, or ‘adaptational’. No matter what meaning is stipulated for these terms they may nevertheless convey attitudes or obstruct certain questions without the social scientist recognizing their normative role. More fundamentally, the very treatment by social science of institutions and norms as subjects of objective scientific exploration may unwittingly suggest the naturalness, inevitability and immutability of social arrangements which are in fact artificial, constructed and subject to negotiation. As noted above, it is the responsibility, some will hold, of social scientists to unmask the character of oppressive or exploitative institutions; this is made more difficult by the failure to recognize the implicit normative dimension of social theories and methods.

On a traditional view of the normative dimension of social science *qua* science there is a distinction between facts and values, and by scrupulousness about this distinction the social scientist *qua* scientist can and should avoid making normative claims implicitly or explicitly. The normative force of this claim rests on the view that scientists should be as objective as possible. For the value of scientific knowledge and its further accumulation is jeopardized by the appearance of partiality. Radically opposed to this view is the claim that objectivity is

impossible in any science, social or natural, that ‘knowledge’ is a coercive term that obscures the partiality and negotiated social constructions of physical science, and in any case, as Kuhn is taken to have shown, there is no real accumulation or progress in the history of science. Ironically, on this radical view, the problem of normative versus positive theory is not at all a distinctive one for the philosophy of social science. Either it is a problem for the philosophy of all the sciences or for none of them.

See also: Economics, philosophy of; Social science, contemporary philosophy of; Social science, history of philosophy of

ALEX ROSENBERG

References and further reading

Beauchamp, T., Faden, R.R., Wallace, R.J. and Walters, L. (1982) *Ethical Issues in Social Science Research*, Baltimore, MD: Johns Hopkins University Press.(An introduction to moral problems of controlled inquiry.)

Churchland, P. (1981) ‘Eliminative materialism and the propositional attitudes’, *Journal of Philosophy* 78: 67-90. (Argues for elimination of intentional variables in explanation of behaviour.)

Dennett, D. (1969) *Content and Consciousness*, London: Routledge & Kegan Paul. (Influential treatment of the problem of intentionality in psychology.)

Durkheim, É. (1951) *The Rules of the Sociological Method*, New York: Free Press. (Arguments presented for holism and functionalism by the founder of empirical sociology.)

Friedman, M. (1953) *Essays in Positive Economics*, Chicago, IL: University of Chicago Press.(The *locus classicus* for instrumentalist accounts of social science.)

Geuss, R. (1981) *The Idea of Critical Theory*, Cambridge: Cambridge University Press.(An introduction to the role of reflexivity in critical theory’s philosophy of social science.)

Harris, M. (1979) *Cultural Materialism*, New York: Random House.(Defends empiricist approaches in anthropology.)

Hausman, D. (1992) *The Inexact and Separate Science of Economics*, Cambridge: Cambridge University Press.(A complete treatment of the philosophical problems of economics, includes an introduction to the philosophy of science with a lengthy bibliography.)

Hempel, C. (1965) *Aspects of Scientific Explanation*, New York: Free Press.(A collection of influential essays in the philosophy of social and natural science.)

McIntyre, L. and Martin, M. (1994) *Readings in the Philosophy of Social Science*, Cambridge, MA: MIT Press. (The newest and most complete collection of important papers in the philosophy of social science.)

Neurath, O. (1941) ‘Philosophy of social science’, in *International Encyclopedia of Social Science*, Chicago, IL: University of Chicago Press.(An early logical positivist exposition.)

Popper, K. (1957) *The Poverty of Historicism*, London: Routledge & Kegan Paul.(An attack on historicism and anti-empiricist philosophies of social science, expounds Popper’s doctrine of falsifiability.)

Rosenberg, A. (1988) *The Philosophy of Social Science*, Boulder, CO: Westview Press.(An introduction to the philosophy of social science.)

Skinner, B.F. (1953) *Science and Human Behavior*, New York: Macmillan.(A defence of behaviourism and eliminativism in psychology.)

Sober, E. (1993) *The Philosophy of Biology*, Boulder, CO: Westview Press.(An introduction to the philosophy of biology, with relevance to social science.)

Taylor, C. (1964) *The Explanation of Behaviour*, London: Routledge & Kegan Paul. (A defence of an anti-empiricist philosophy of social science.)

Weber, M. (1949) *The Methodology of the Social Sciences*, New York: Free Press.(A historically influential argument for compatibility of causal and intentional explanation in social science.)

Winch, P. (1958) *The Idea of a Social Science*, London: Routledge & Kegan Paul. (The earliest sustained application of Wittgenstein’s views to attack the possibility of empirical social science.)

Wittgenstein, L.J.J. (1953) *Philosophical Investigations*, ed. G.E.M. Anscombe and R. Rhees, trans. G.E.M. Anscombe, Oxford: Blackwell.(The most polished and worked over of all Wittgenstein’s later work; it contains the presentation of his ideas on meaning and philosophical psychology with which he was most nearly satisfied.)

Social sciences, philosophy of

Although some of the topics and issues treated in the philosophy of social science are as old as philosophy itself (for example, the contrast between nature and convention and the idea of rationality are dealt with by Aristotle), the explicit emergence of a subdiscipline of philosophy with this name is a very recent phenomenon, which in turn may itself have stimulated greater philosophical activity in the area. Clearly, this emergence is tied to the development and growth of the social sciences themselves.

1 Historical approach

There are, perhaps, four distinct ways in which to gain an understanding of the subdiscipline. These ways are, of course, complementary. First, just as with most other areas of philosophy, one might approach the philosophy of the social sciences historically, by studying major schools or philosophers of an earlier period. There is much to recommend this approach (see Social science, history of philosophy of). There are a number of classical texts (by Weber and Durkheim, for example) of which any interested student of the philosophy of the social sciences should be aware, much as there is in epistemology or ethics. This provides an interesting contrast with the philosophy of the natural sciences; far less could be said in favour of gaining an understanding of the latter in this way.

Compared with other areas of philosophy, the history of the philosophy of the social sciences is somewhat truncated, since it can only begin properly with the earliest attempts at social science, in the late eighteenth and early nineteenth centuries, first in the Scottish Enlightenment and subsequently in Germany. Prior to this period, there had been speculation about the nature of society, some of it quite rich and rewarding (Hobbes and Vico provide two examples of this), but it is only in the period of the Scottish Enlightenment and after that writers begin to reflect the first systematic attempts to study and understand society.

There is no clear line of demarcation between philosophers of social science and of society on the one hand and social theorists on the other, especially in this early period. Conventionally, to select only a few examples, G.W.F. Hegel, Wilhelm Dilthey, F.H. Bradley and T.H. Green are considered to be examples of the former, and Adam Smith, Karl Marx, Émile Durkheim, and Max Weber, are considered as examples of the latter, but the line is sometimes somewhat arbitrary (see Hegel, G.W.F.; Marx, K.; Dilthey, W.; Bradley, F.H.; Green, T.H.; Smith, A.; Durkheim, É.; Weber, M.).

2 Problems

A second way in which to gain an understanding of the philosophy of social science is through the study of the issues and problems that these writers, and their contemporary counterparts, address (see Social science, methodology of). Many of these problems arise in ordinary as well as in more scientific discussions of and thought about the social realm. It is not only social scientists who think about the social world; all of us do, a great deal of the time. Even in those cases in which the social scientist introduces neologisms, for example, ‘demand curves’ or ‘anomie’, they seem closely connected to, and sometimes only a refinement of, concepts already grasped by the lay person.

This nonscientific reflection arises quite apart from any specialized scientific work. It is, to a certain extent, misleading to think of the field as only the philosophy of the social *sciences*. Since so much of the motivation for critical discussion of the problems in this area comes from philosophical reflection on these quite ordinary modes of thought and understanding, the field should perhaps be called ‘the philosophy of *society*’, to reflect this nonscientific, as well as the scientific, interest in those problems.

Most of the things that social science is about, social structures (like families or society itself), norms and rules of behaviour, conventions, specific sorts of human action, and so on, are items that find a place in the discourse of the ordinary lay person who has as good a grasp of common talk about social class and purchase, voting and banking, as does the social scientist. This raises, in a direct way, metaphysical questions about the nature of these things. Are these social structures anything more than just individuals and their interrelations? Many philosophers, in the grip of the ideal of the unity of science, have held out the prospect that social science can be derived from, and is therefore reducible to, psychology (the latter eventually being reducible to chemistry and physics). For such thinkers, the world is ultimately a simple place, with only many different ways in which to speak about it. Other

thinkers have been struck by the reality and integrity of the social world, and how it seems to impress itself on the individual willy-nilly (see Society, concept of; Social norms; Holism and individualism in history and social science).

What is an action, and how does it differ from the mere movement of one's body? It seems hard to say in what this difference consists in a way that remains plausible and true to what action is like. Whatever an action is, what makes some actions *social* actions? One might think that an action is social in virtue of its causal consequences on others. Another line of thought holds that an action is social in virtue of its intrinsic character, quite apart from the question of its effects. Much of the philosophical discussion of action arose in the philosophy of history, over the explanation of historically important action, but has now been absorbed into a separate area of philosophy, the theory of action (see History, philosophy of; Action; Social action).

The alleged contrast between nature and convention occurs to those who think about humankind and its development, whether they be scientists and philosophers or not. Anyone who has travelled widely and noticed the social differences between peoples and cultures may have wondered whether all social practice was rational in its own terms, wherever found and no matter how apparently peculiar by our home-grown lights. Or perhaps, on the other hand, there are some universal standards of rationality, in the light of which evaluation of social practices and criticism of some of them can be mounted (see Nature and convention; Rationality and cultural relativism; Social relativism).

The relationship between scientific theory and ordinary modes of thought is, of course, interactive, since many of the concepts or issues that have become part of ordinary lore have their roots in earlier scientific theory (our modern, and by most accounts, confused, concept of race might be an example of this; see Race, theories of).

Another set of problems arise in thinking through the nature of the social scientific enterprise itself. What standards must full explanation in social science meet? Causal explanation is a mode of explanation in natural science that is, relatively speaking, well understood. Explanations of a ritual or practice in society do not appear to be causal explanations, nor do explanations of human action. The first are often functional explanations (for example, a certain ritual exists because it produces such-and-such) and this appears to be an explanation of something by its effects rather than by its causes. Explanations of human action are intentional explanations, whereby an action is explained by the goal or end at which it is directed. This also appears not to be causal. But perhaps appearances are deceptive, and these can be recast as causal explanations after all (see Explanation in history and social science; Functionalism in social science).

Natural scientists believe that their work is ethically neutral. To be sure, their work can be put to good and bad uses, but this presumably reflects on the users rather than on the content of the science itself. The relationship between social science and the values of the social scientist seems far more immediate and direct than this, and this alleged contrast has been the subject for continuing discussion and debate (see Value judgments in social science).

Is social science like natural science in important ways? In the developed natural sciences, there are controlled experiments and predictions. Neither seem available to the social scientist. Natural scientists attempt to formulate the laws that govern the phenomena they study. Is this a reasonable goal for the social scientist? Certainly, there are not many candidate laws for the social sciences one can think of. Does the social scientist use statistical evidence in the same way as the natural scientist? (See Experiments in social science; Social science, prediction in; Social laws; Statistics and social science.) Finally, in natural science, we distinguish between theory and observation in a relatively sharp way, and we believe that a rational person should accept that theory which is best confirmed by observations. It is not clear that we can make the same distinction in the social sciences, nor that theory is supported by observation in just the same way. Our observations of the social world seem even more coloured by the theory we employ than is the case in the natural sciences (see Theory and observation in social sciences.)

3 Contemporary movements

A third way in which to approach the subject is through the study of either contemporary movements and schools of philosophy, or specific philosophers, who bring a specific slant to the subdiscipline. Controversy marks the natural as well as the social sciences, but observers have noted that there seems to be even less consensus, even less of an agreed paradigm at any particular time, in the latter than in the former.

Critical reflection on society, or on social science, or both, is very different in France and Germany from the way it is in the English-speaking world. The problems are the same, but the traditions and the manner in which the discussions proceed are markedly distinctive. The hope is that each tradition may learn something from the other (see Social science, contemporary philosophy of; Behaviourism in the social sciences; Critical realism; Evolutionary theory and social science; Lévi-Strauss, C.; Naturalism in social science; Positivism in the social sciences; Post-structuralism in the social sciences; Scientific realism and social science; Sociology of knowledge; Structuralism in social science; Symbolic interactionism; Systems theory in social science; Bourdieu, P.; MacIntyre, A.; Schütz, A.).

4 Specific social sciences

Fourth and finally, one might approach the philosophy of the social sciences by studying the philosophical problems that arise specifically within each of the social sciences. Some, although not all, of the social sciences have thrown up philosophical industries all their own. Economics is the most salient example. In many ways, it is the most developed of all the social sciences, and this may be the reason why some of the best-defined controversies in the philosophy of social science arise from within it. Questions about the philosophical foundations of economics touch on the philosophically central issues of rationality, choice and the nature of wants or desires and their connection with action (see Economics, philosophy of; Social choice; Rational choice theory). But other social sciences have also given rise to specific problems, including history, psychology, sociology, and anthropology (see Psychology, theories of; Sociology, theories of; Anthropology, philosophy of).

See also: Feminism and social science

DAVID-HILLEL RUBEN

References and further reading

Martin, M. and McIntyre, L. (eds) (1994) *Readings in the Philosophy of Social Science*, Cambridge, MA: MIT Press.(A useful collection of recent and contemporary articles, grouped around some of the main issues in the philosophy of social science.)

Root, M. (1993) *Philosophy of Social Science*, Oxford: Blackwell.(Argues that some of the most prominent research programmes in the social sciences flout the ideal of moral and political philosophy.)

Ruben, D.-H. (1998) ‘The Philosophy of Social Sciences’, in A. Grayling (ed.), *Philosophy 2: Further Through the Subject*, Oxford: Oxford University Press, vol. 2.(A discussion of the main problems in the philosophy of social science, intended for the philosophy student. Assumes some prior knowledge of philosophy.)

Social sciences, prediction in

Prediction is important in science for two reasons. First human beings have a practical interest in knowing the future. Therefore, all science is potentially predictive in the sense that its results may be used as a basis for expectations. Second, a test of our beliefs is the truth of the predictions we can derive from them. In the social sciences, however, predictions are often supposed to create specific philosophical and methodological problems, the roots of which are the following: the phenomena studied in the social sciences are so complex and so interrelated that it is practically impossible to formulate law-like generalizations about them; human beings are supposed to possess free will; and the predictions may themselves modify the phenomena predicted.

Some phenomena are deterministic; they obey invariable laws. Some other phenomena are stochastic; they obey laws that could be expressed only in statistical terms. Finally, some phenomena - most notably human actions - are claimed to be free. According to the incompatibilist philosophers, free actions are not subjects of either deterministic or stochastic laws (see Free will §2).

As regards predictability, we should note that (1) predictability does not imply determinism, because we may have statistical predictions concerning stochastic phenomena. (2) nor does determinism imply predictability, for the following reasons: some phenomena are deterministic but chaotic in the sense that changes which are undetectably small can sometimes produce large changes; some phenomena are so complex that it is humanly impossible to describe all the relations and interrelations relevant for predictions; and in some cases it is practically impossible to acquire the information relevant for making informed predictions. From (2) it follows that unpredictability does not imply freedom. All the limitations mentioned in (2) may be relevant in regard to the study of social phenomena. But they can also be relevant in other areas. For example, financial markets may function chaotically, but the same is true of meteorological phenomena. Individual human beings, as well as human societies, are extremely complex entities, but so are animals. Moreover, it is possible for phenomena having extremely complex internal structures to exhibit a predictable pattern in their general behaviour. We may be able to predict the movements of the whole without being able to describe or predict the movements of the parts. Some philosophers have claimed that we cannot find general laws of human societies because all societies are essentially open systems. Unlike in a laboratory experiment, or an astronomical observation, we cannot eliminate the possible influence of outside factors. Thus, we are never entitled to say that phenomena are caused by those particular factors we are experimenting with. Again, all this is equally true in many natural sciences (for example, in meteorology).

It seems that the only philosophical problem specifically relevant to social predictions is the question of free will. However, this supposed freedom does not prevent us making predictions in our everyday life. Usually, we predict without the help of any (systematic) scientific knowledge: we ascribe certain aims and beliefs to our fellow beings, derive rough predictions of how they are going to behave, and form our own expectations accordingly. Similarly, social scientists may postulate general motives, for example, profit maximization, and shared beliefs, for example, correct beliefs about present market prices, deriving useful predictions regarding general trends. Even if the postulates are not true in respect of every individual, individual idiosyncrasies may cancel each other out, or, in particular contexts, there may be pressures among individuals to conform to the postulates. For example, entrepreneurs who do not try to maximize their profits and to use adequate information are soon out of business. The structure of the choice situation may create a predictable pattern.

There is, however, one further problem peculiar to the social sciences only. Mutually held expectations are essential for our social life. Most institutions can work effectively only because people conform to certain predictable patterns, and they conform to these patterns partly because they expect that institutions (for example, banks, governments or courts) will work effectively. When forming expectations, human beings try to take into account information concerning the future, including the information produced by the social sciences. Thus, predictions about the future rate of inflation affect the actual rate of inflation by influencing the expectations of firms and trade unions. Social predictions may possess a self-modifying (reflexive) force: they may make themselves true or untrue. Social scientists may try to take into account the possible effect of their predictions, but the subjects of the predictions may, in turn, react to such attempts. It appears that in certain circumstances a social

theory cannot describe and predict the consequences it has for the people it describes and predicts. This result does not presuppose a metaphysical notion of free will. People may react to the content of a theory T1 in a deterministic way, and this deterministic process can be predicted by another theory T2. T2 cannot predict its own effects, but they can be predicted by T3, and so on.

In some research contexts, the self-modifying potential of predictions can be ignored. First, some issues studied, for example, in psychology, are not really of a social nature. They are closely related to the physiological processes of human beings, setting the necessary conditions and limitations of human agency.

Second, in many cases the predictions produced by social scientists are not available or comprehensible to the subjects of the predictions, for various reasons. This solves the reflexivity problem, but only at an ethical cost. For it means that in such contexts, successful predictions presuppose an informationally superior position.

Third, the predicted social situation may be moving towards a unique equilibrium. Here, equilibrium is a technical notion, meaning that the relevant actors (individuals, firms, states and so on) do not simultaneously possess motive and ability to deviate from the predicted pattern. No new information can change the outcome. In a trivial case (for example, an idealized market situation), all or most agents have only a single feasible option.

The self-modifying effect is likely to appear in situations where no equilibria exist, or where there are several - for example, in imperfect markets, in multiparty politics and in multipolar international systems. In such contexts, the agents have an on-going motive to acquire information and to try to outguess each other’s future actions. Hence, the ability of the social sciences to predict is dependent on the distribution of knowledge in societies.

See also: Explanation in history and social science; Social laws

EERIK LAGERSPETZ

References and further reading

Bicchieri, C. (1987) ‘Rationality and Predictability in Economics’, *The British Journal for the Philosophy of Science* 38: 501-13.(A relatively technical critique of recent economic theories.)

Friedrichs, R.W. (1970) *A Sociology of Sociology*, New York: Free Press.(The sociological aspects of the prediction problem are discussed on pages 177-95.)

Goldman, A. (1970) *A Theory of Human Action*, Pinceton, NJ: Princeton University Press.(Chapter 6 provides an excellent discussion of the problems of prediction of human actions.)

Henschel, R.L. (1978) ‘Self-altering Predictions’, in J. Fobles (ed.), *Handbook of Futures Research*, Westport, CT: Greenwood Press, 99-123.(Contains an interesting discussion on the self-modifying effects of predictions; bibliography is excellent.)

Henschel, R.L. (1982) ‘The Boundary of the Self-fulfilling Prophecy and the Dilemma of Social Prediction’, *The British Journal of Sociology* 33 (4): 510-28.(A discussion on the extent of the prediction problem and on the conditions under which it is *not* likely to emerge.)

Hutchison, T.W. (1977) *Knowledge and Ignorance in Economics*, Oxford: Blackwell. (A non-technical and polemical discussion.)

Lagerspetz, E. (1988) ‘Reflexive Predictions and Strategic Actions’, *Social Science Information* 27: 307-20. (Emphasizes the ethical and political implications of the problem.)

MacIntyre, A. (1981) *After Virtue: A Study in Moral Theory*, London: Duckworth, 2nd edn, 1985.(Chapter 8 of the book is a good philosophical introduction to the subject.)

Rosenberg, A. (1993) ‘Scientific Innovation and the Limits of Social Scientific Prediction’, *Synthèse* 97: 161-82. (Concentrates on Karl Popper’s arguments, which are not discussed above, but relates them to our themes.)

Sen, A.K. (1986) ‘Prediction and Economic Theory’, *Proceedings of the Royal Society*, A 407: 3-23.(On the role of predictions in economics.)

Social theory and law

Social theory embodies the claim that philosophical analyses, reflections on specific historical experience and systematic empirical observations of social conditions may be combined to construct theoretical explanations of the nature of society - that is, of patterned human social association in general and of the conditions that make this association possible and define its typical character. Social theory, in this sense, can be defined broadly as theory seeking to explain systematically the structure and organization of society and the general conditions of social order or stability and of social change. Since law as a system of ideas can also be thought of as purporting to specify, reflect and systematize fundamental normative structures of society, it has appeared as both a focus of interest for social theory and, in some sense, a source of competition with social theory in explaining the character of social existence.

The relation of legal thought to social theory is, thus, in important respects, a confrontation between competing general modes of understanding social relationships and the conditions of social order. In one sense, this confrontation is as old as philosophy itself. But as an element in modern philosophical consciousness it represents a gradual working-out in Western thought, over the past two centuries, of the implications of various 'scientific' modes of interpreting social experience, all in one way or another the legacy of Enlightenment ideas.

From the late eighteenth century and throughout the nineteenth century, criteria of 'scientific' rationality were carried into the interpretation of social phenomena through the development of social theory. These criteria also significantly influenced the development of modern legal thought. The classic social theory of the late nineteenth and early twentieth centuries, which established an enduring vocabulary of concepts for the interpretation of social phenomena, treated law as an object of social inquiry within its scope. It sought scientific understanding of the nature of legal phenomena in terms of broad systems of explanation of the general nature of social relationships, structures and institutions.

In the late twentieth century the relationship between social theory and law has been marked by fundamental changes both in the outlook of social theory and in forms of contemporary regulation. On the one hand, social theory has been subjected to wide-ranging challenges to its modern scientific pretensions. It has had to respond to scepticism about claims that social life can usefully be analysed in terms of historical laws, or authoritatively interpreted and explained in terms of founding theoretical principles. On the other hand, the inexorable expansion of Western law's regulatory scope and detail appears, sociologically, as largely uncontrollable by moral systems and relatively unguided by philosophical principles. Hence, in some postmodern interpretations, contemporary law is presented as a system of knowledge and interpretation of social life of great importance, yet one that has ultimately evaded the Enlightenment ambition systematically to impose reason and principle - codified by theory - on agencies of political and social power.

1 Civil society and modernity

Social theory aims to develop systematic theoretical understandings of the nature of social phenomena or of 'society' treated as a specific object or focus of inquiry. Intellectual sources of social theory are found in speculations throughout the history of philosophy on the foundations of social order, the character of human association and the essence of human nature. Nevertheless, social theory as a distinct intellectual enterprise is best understood as originating with the Enlightenment. Emerging out of social philosophy, its concerns are closely implicated with general philosophical issues of individuality, rights, authority, responsibility and community. Its establishment as an intellectual field depended on three main developments: first, the identification of an analytically distinct field of the 'social'; second, the acceptance of the necessity of a grounding for social speculation in evidence about social experience in specific historical contexts; and, third, a conception of social change as theoretically explicable, suggesting a qualitative differentiation between 'modern' and 'pre-modern' societies or characteristics of social life.

The identification of a distinct field of the 'social' is associated with the conceptualization, especially derived from Hegelian thought, of civil society as an autonomous realm of social interaction, differentiation, organization and institutions. One consequence of this conceptualization was to identify such institutions as contract, property and social class and the processes and effects of the administration of civil justice as foci for theoretical analysis, not in

terms of their ethical implications or their significance for political theory but as components of a cohesive social fabric requiring explanation in terms of its integrative structure and historical dynamics.

The specification of the ‘social’ also entailed the development of distinctive approaches to analysis and, eventually, the establishment of an intellectual enterprise of sociology claiming its own theory, methods and aims. Beginnings of these processes can be seen in Montesquieu’s *L’Esprit des lois (The Spirit of the Laws)* (1748) which attempts to link typical social and political structures with characteristics of the types of natural and historical environments in which these structures emerge. Climate, geography and population size are related to law and political institutions, and law is seen as a consequence of natural conditions and the culture they inspire as well as a force acting upon them (see Montesquieu, C. §3).

A further element essential to the project of social theory but not present in Montesquieu is a conception of modernity; that is, a conception of a general historical transition from characteristic earlier forms of society or of social relationships to a specifically modern form, which could be generalized as both qualitatively distinct from earlier forms and also typical of the most advanced or developed actual societies of the present. The theme of such a transition pervades the major literature of nineteenth-century social theory. One of its most striking expressions is Ferdinand Tönnies’ distinction (1887) between *Gemeinschaft* and *Gesellschaft*; that is, between, on the one hand, multi-faceted, long-term, often emotionally or traditionally sanctioned social relationships typical of close-knit community or family life and, on the other hand, instrumental, specific and limited, often short-term and frequently changing social relationships based on agreement or the negotiation of rapidly changing patterns of social interaction. Tönnies’ concepts abstractly present ubiquitous elements of social relationships unrelated to any particular historical context, but the inspiration for their formulation was undoubtedly a sense of the increasing prominence of *Gesellschaft* relations in the conditions of mobile, diverse and dynamic ‘modern’ society.

Modernity is characterized variously in the most prominent forms of social theory developed by the end of the nineteenth century: for example, in Herbert Spencer’s emphasis on functional differentiation and reintegration in modern society; in Émile Durkheim’s characterization of complex industrial societies as morally diverse yet also morally integrated by functional interdependence; in Karl Marx’s identification of the anatomy and pathology of capitalist society as a distinct historical form; and in Max Weber’s ambivalent exploration of the progress of numerous forms of rationalization culminating in the bureaucratized technical consciousness and morally compartmentalized social environments of late nineteenth-century Western capitalist societies.

Many of these presentations of modernity in social theory rely on analyses of law and legal history. Law’s doctrinal history provided documentary materials that could be treated as empirical sources in plotting trajectories of social change. Both Tönnies and Durkheim used legal materials - including the jurist Henry Maine’s evolutionary legal history (see Jurisprudence, historical §2) - extensively in developing their accounts of the contrast between social modernity and pre-modernity. And Weber’s discussion of political and economic structures and modes of thought characteristic of developed capitalism is strongly flavoured by his sophisticated understanding of the nature of modern Western legal thought.

2 Science of law and science of society

A further condition for the emergence of social theory was an emphasis on the need for systematic empirical inquiry about social phenomena and an explicit conception of the nature of scientific method in social analysis (see Social sciences, philosophy of §1; Positivism in the social sciences §2). Here it is important to see both a parallel with developing ideas in legal analysis and a challenge to those ideas. Long before the idea of a rational ‘science of society’ began to establish itself (especially through the writings of Henri de Saint-Simon and Auguste Comte early in the nineteenth century), legal thought presented itself as a professional ‘artificial reason’ authoritatively conceptualizing social reality. English common law thought, as elaborated by such as Edward Coke and Matthew Hale in the seventeenth century, embodies conceptions of community that underlie legal doctrine’s interpretations of the normative structure of society.

Late eighteenth-century and nineteenth-century movements for scientific rationalism in legal thought did not necessarily suggest the need for theory that would locate law in a wider terrain of empirical social knowledge. The initial demand was, at least in the English context, for more systematic conceptual analysis of legal doctrine than that allowed by the pragmatic methods of judge-made common law. Nevertheless, utilitarian reformers demanded

that romantic or mythical common law conceptions of community and of an ancient unchanging constitutional order be replaced with more realistic understandings of the nature of modern government, of the sources of legal authority and of the political processes of creation and application of law.

Comte's identification of a specific science of society, which in 1838 he named sociology, is almost contemporaneous with the jurist John Austin's powerfully influential outline of a modern 'science of law', requiring no support from the moralistic speculations of natural law theory (see Austin, J.). The appeal to 'science' made by many nineteenth-century jurists was usually to some general model of rigorous, rational, objective and systematic thought, such as geometry. The contribution of an emerging social theory, during the nineteenth century, was to help bring into legal analysis a self-consciousness about empirical resources for legal speculation and, eventually, a limited recognition that many resources necessary to an adequate theoretical understanding of law must be found, beyond the scope of the lawyer's traditional structures of reason, in systematically ordered data derived from the observation and recording of social experience.

Social theory in the late nineteenth and twentieth centuries shows an increasing attention to the dilemma of the meaning of 'science' in social analysis. The Durkheimian injunction at the end of the nineteenth century to treat social facts as if they were things (objects of science) to be analysed by means of the observational techniques of natural science (see Durkheim, É. §2) was confronted and challenged by contemporaneous claims that the methods of the natural and the social sciences are necessarily radically opposed and that the latter require distinctive methods of interpreting the subjective meaning of social acts for those involved in them if the patterns of social life are to be intelligible to the scientist (see Weber, M. §3).

The often-quoted dictum that sociology was born in a state of hostility to law can be understood as a recognition that in establishing its empirically-oriented, systematic, generalizing modes of analysis of social life, social theory challenged the prerogatives of such established enterprises of social interpretation as those of law. It sought to replace the normative, evaluative social interpretations that legal analysis offered with explanatory, descriptive and ostensibly non-judgmental analyses of social phenomena, including legal phenomena.

In so far as social theory founded itself on conceptions of modernity and sought to explain patterns of social development theoretically, it challenged common law assumptions about the timeless stability of the community whose immanent normative framework the common lawyers purported to understand and interpret. And, in its unending questioning of the nature of 'science' in social analysis, social theory provided material that would eventually challenge lawyers' claims to possess a secure and autonomous 'science of law'. As long as law was admitted to be a social phenomenon, legal theory would eventually have to take account of debates on social interpretation and analysis that social theory presented. In this way, the advent of modernity in law and in social theory raised the possibility that legal science would become dependent on and subordinate to social theory.

3 Law in classic social theory

The bodies of social theory that have exerted most long-term and widespread influence in social analysis and still provide the main vocabulary of concepts for considering the social character of modern Western law are the systems of thought associated with Karl Marx, Émile Durkheim and Max Weber (see Marxism, Western). Their contribution to legal scholarship is especially significant because each devotes considerable attention to law within a wider project of social theory, and because each views law, at least in part, as comprising distinctive systems of doctrine or modes of reasoning. They thus indicate the prospect of directly confronting, in social theory, juristic understandings of the nature and significance of legal doctrine.

In these respects, the contributions of this classic social theory to legal scholarship are broader than is suggested by its use in much of the research in sociology of law. Empirical sociology of law is often behaviourally oriented (focusing on the explanation of activity in legal settings - for example, police practices - rather than on legal discourse or doctrine itself). It tends to be strongly influenced by the present organization of social science disciplines and therefore to reflect their perspectives. Partly as a consequence of this organization which largely excludes law, treating it as an independent and external discipline, sociology of law is often viewed as peripheral by the social science disciplines (especially academic sociology and anthropology) to which it relates. At the same time, sociology of law's respect for disciplinary boundaries of academic social science has often deterred it from engaging effectively with the disciplinary outlook of lawyers and the concerns of legal philosophy.

Marxist theory's main influence on legal analysis has been to highlight the need to relate legal doctrine and legal thought systematically to specific social interests which law promotes and to see law as a focus of political struggle often obscured by the technicality of professional legal discourse. While the link between law and social class interests was recognized long before Marx wrote, Marxist theory has focused attention on the devices by which law's class character may be obscured. It radically challenges the idea that law can attain a level of principled generality that gives it a moral and political autonomy above the endless struggle and compromise of conflicting economic and social interests. Marxist theory does not generally engage with the arguments over legal principle and legal foundations that legal philosophy pursues. Rather it regards legal philosophy's moral arguments as being largely conducted on the terrain of ideology. Marxist social theory's most important contribution to legal analysis is the effort to develop a coherent theory of the nature of ideology; a theory that can be used consistently to relate patterns of legal ideas to broader social, economic and political developments while also explaining the mechanisms and conditions under which legal discourse in advanced capitalist societies consistently proclaims its own moral, intellectual and political autonomy from the conflicting social forces that it purports to regulate.

By contrast, Weber, trained in law and viewing his legal studies as central to his overall project of interpreting the social character of Western capitalism, emphasized aspects of modern law different from the mystificatory or ideological elements and repressive structures that Marx had treated as essential. For Weber, modern law's importance lies primarily in its distinctive forms of rationality, which correspond to certain rational requirements of modern Western state structures. This rationality also underpins forms of purposive social action which are especially highly developed in and important to modern capitalist societies, and it contributes to the stability and legitimacy of modern political systems. Weber's acute sensitivity to changing forms of legal reasoning and decision-making allows him to provide important insights into the social and political conditions of existence of the *Rechtsstaat* (see Rule of law (Rechtsstaat) §1). Unlike Marx, who views legal ideas consistently from the external standpoint of a demystifier and critic of ideology, Weber interprets these ideas in lawyers' terms, understanding both their social importance and the technical processes by which they evolve. Yet he contextualizes this interpretation, explaining, for example, intimate relationships between changing forms of legal reasoning and conditions for the expansion of capitalistic economic activity; and between these legal forms and the kinds of legitimacy claims that political and other authorities are able to make in specific historical conditions.

Because Weber's social theory is informed by a deeper understanding of the technicalities of legal thought than is Marx's, it offers a more nuanced view of the ambiguities of legal modernity. Its complexity of outlook and its ambivalence about modernity in general, and about the rationalization processes at work in modern law, also make the consequences of Weberian social theory for legal philosophy's concerns harder to define than is the case with Marxist theory. In general, despite its richness and a large secondary literature on its legal aspects, Weber's work still remains insufficiently assimilated in legal scholarship. The explanation lies partly in his methods, which reject the possibility of discovering scientific laws in history and celebrate in often dense, much-qualified argument the variety of perspectives that can be brought to light by using provisional, pragmatically chosen interpretive concepts. Beyond this, however, Weber poses profound challenges - rather like Maine in an earlier era - to the jurist to ignore disciplinary demarcations and recognize law's place as one strand of understandings in a vast, complex, cultural matrix of values and beliefs, modes of reasoning and interpretation.

Among the classic social theorists of law, Durkheim has so far exerted least direct influence on legal theory, perhaps because the emphasis of his legal thought, with its almost complete silence on questions about the relations of law and power (social, economic or political), has seemed naïve to legal scholars who absorbed some of the claims of Marx or Weber about the political or economic conditions of existence of law. The emphasis of Durkheim's legal theory is on law's intimate relations with morality and on the problem of defining the moral foundations of law in modern, secular, functionally and socially differentiated societies. In these societies, shared values, if they can be held to exist, are necessarily greatly limited in scope and significance by comparison with the situation in simpler, more socially uniform and close-knit societies, or those where strong and universal religious commitment underpins unifying social values.

Law was an important focus of study for Durkheim and his school of associates but the more limited impact of his ideas on legal theory by comparison with those of Marx and Weber is understandable in so far as the Durkheimian approach seemed to deny what legal scholars, no less than social analysts, came to sense as one of the major

characteristics of modernity - the disintegration of moral unities and the emergence of law as a mechanism by which modern society might endlessly remake itself. The Durkheimian idea of law as a reflection of moral conditions has seemed less appropriate than Marxist and Weberian perceptions of law as a shifting element in complex articulations of economic and political interests in rapidly changing modern society. It has generally seemed more useful to view law as an instrument of political action and a system of ideas integrated with dynamic social forces determining the pace and direction of change and the conditions of social stability.

4 Law and recent social theory

During the last decades of the twentieth century, social theory increasingly rejected or radically modified the three assumptions or claims - a distinct realm of the 'social', distinctive 'scientific' methods of analysis of social phenomena, and the experience of modernity as the ever-present context of analysis - underpinning its original development. This phase of the relationship between social theory and Western legal thought is characterized by a more ambiguous theoretical recognition of the field of the 'social', by a widespread hesitancy about theoretically defending not only distinctive methods of social analysis but also the validity of the social knowledge that they produce, and by a correspondingly greater difficulty in identifying a distinct field of social theory. Thus, relationships between legal scholarship and social theory are increasingly replaced by complex connections between legal theory and various theoretical projects associated with a wide range of intellectual traditions that are not unified by any shared conception of the nature or scope of the 'social' as a research focus.

These developments can be considered in terms of their impact on legal analysis. Of all forms of social theory, Marxist theory probably has had the most direct influence on the study of modern law. Its appeal has depended on its claim to provide a materialist explanation of legal developments; an explanation that showed them as the predictable outcome of a theoretically explicable historical process. Radical legal scholarship in various Western countries, especially from the 1970s, used Marxist theory to develop the idea of law as a site of political struggle, and especially of ideological struggle, following the insights of the Marxist philosophers Antonio Gramsci and Louis Althusser.

The idea of law as a site of ideological conflict, and more generally an emphasis on law's rhetorical significance and its power to shape consciousness, has been retained and developed in current radical or critical legal theory. But, reflecting recent tendencies in social theory, critical approaches in contemporary legal scholarship - in no way limited to those associated with the international Critical Legal Studies movement - now typically resist any idea of objective laws of history or general long-run mechanisms of social change. While ideology remains an important concept, ideology is not to be considered as a distinct object of social analysis but rather as an aspect of intersubjective practices. The idea of society - or any other large-scale social phenomenon - as an object of inquiry has been widely discarded in favour of the study of specific social and political practices and the rhetorics (in which legal ideas often play a prominent part) that accompany them.

Many related tendencies in contemporary social theory are encapsulated in these kinds of scepticism. A distrust of what J.-F. Lyotard (1979) has called 'grand narratives' - theories or perspectives purporting to encapsulate in general terms large swathes of human experience - is founded on a recognition that the ultimate validity of these narratives is necessarily untestable, as well as on a widespread conviction that knowledges adequate to confront the complexities and variety of social experience will necessarily be local rather than general. More specifically, the idea of 'science' as a mark of appropriate social analysis has lost much of its resonance and become associated - perhaps unjustly - with positivist social science and with a lack of sensitivity to the unquantifiable, subjective, transient, immanent but powerful cultural ambiences of social life. It has been associated with the contested claim that clear lines can be drawn between valid and invalid social knowledges, or between truth and falsity in interpretation of social life and social phenomena. Naturalistic models of science have been displaced in favour of models of practice - hermeneutics, literary interpretation, arts of rhetoric, aesthetics, semiotics and linguistic analysis - that offer the prospect of escape from social-scientific protocols of testability and relevance, and promise access to ineffable knowledges that capture the intricacy and indeterminacy of contemporary social experience.

In legal analysis, the issue of the nature of interpretation of legal doctrine remains central and it does so in conditions in which legal regulation proliferates inexorably, assuming a very wide variety of forms. As legal scholarship has accommodated itself to the increasing difficulty of portraying law as an integrated doctrinal system, its concern has focused more sharply on problems of understanding practices of interpretation of doctrine

in these conditions. Classic social theory offered, in the main, accounts of the social origins and consequences of particular historical interpretations of legal materials. But it is often charged with failing to explore legal interpretive or discursive processes themselves. Hence literary theory and other bodies of ideas concerned with the nature of interpretation and its contexts have tended to usurp space previously occupied by social theory as a resource for supplementing and illuminating legal inquiry.

5 Images of contemporary law

The most important forms of late twentieth-century social theory that have been imported into legal scholarship have achieved influence mainly because they relate directly to contemporary problems of legal analysis in grappling with law's perceived unprincipled doctrinal complexity and bulk. Michel Foucault's writings describe the development of particular sociological and psychological discourses that have brought new forms of scientific rationality into governmental regulatory practices (see Foucault, M.). Foucault sees this development as part of the very process that, by merging disciplinary power with claims to expert knowledge, has created the immense diversity of contemporary regulation and made the idea of legal sovereignty sociologically problematic. Law, which Foucault treats as the coercive instrument of sovereign authority, has found its prerogatives of social control engulfed and usurped by numerous new regulatory forms and mechanisms that gain their authority specifically from technical knowledge and practice, developed especially in the social and human (for example, psychiatric) sciences.

Other writers emphasize contemporary law's embodiment in systems that not only float free of the moral groundings that Durkheim sought for legal regulation, but also confound 'modern' views of law as a steering mechanism of social life. Thus, Niklas Luhmann (1989) insists that the complex differentiation of society into a range of distinct, self-sustaining systems of communication condemns law to inhabit only its own discursive system, rather than to have the ability to steer all of society's systems - economic, scientific, cultural, political, for example - in the way that modernity's aspirations for rational government proposed.

Contemporary social theory's image of law and society tends to be one of fragmentation and diversity and of the ungroundedness of social knowledge; an image of specific social spheres governed by values and understandings that have no foundations except in the experience and perspectives of the local milieux in which they arise. The old, 'modern' project of a science of society is seen as embodying untenable assumptions about the generalizability of social experience, and about the possibility of agreement on truth claims underpinning universal systematic knowledge of the nature of society. Thus, the tenets of modernity have increasingly been replaced by postmodern conceptions of law and society (see Postmodernism).

It would, however, be wrong to see the Enlightenment project of social theory as dead. Jürgen Habermas' highly influential work has, in large part, been concerned to find the basis of an underlying rational and moral foundation of social life - in his terms, a communicative rationality - despite the recognition that, in many vital respects, the fragmentation of social life in contemporary Western societies into disparate systems has destroyed possibilities for this kind of moral and rational grounding (see Habermas, J. §4; Communicative rationality; Legal reasoning and interpretation §4). Habermas' analysis of law treats it as precariously balanced between its technical role in mediating economic, political and other systems within society and its moral potential as a set of institutions expressing the values and understandings that arise in the lived experience and interactions of individuals.

Whether or not Habermas' approach is the most appropriate means of pursuing the search for general theory postulating an underlying rationality of legal and social experience, there are good grounds for claiming that this search is unlikely to cease with a postmodern recognition of its complexity and contradictions. Nor is the invocation of science in the interpretation of social phenomena necessarily rendered inappropriate by critiques of positivist science, 'scientific' interpretations of history, or other kinds of 'grand narratives'. The search for a science of society applicable to legal analysis is best understood as an ongoing commitment to make explicit the partial character of all existing understandings of social (including legal) phenomena and to subject these understandings to unending systematic questioning in the light of broadening social observation and ongoing historical experience. Viewed in this way legal and social theory remain interdependent and the project of social theory appears no less vital, if very much more complex and ambiguous and less coloured by optimism, than in the heyday of the classic social theorists.

See also: Critical Legal Studies §§3-4; Law, philosophy of; Legal discourse

ROGER COTTERRELL

References and further reading

Cain, M. and Hunt, A. (eds) (1979) *Marx and Engels on Law*, London: Academic Press.(Collects and introduces Marx's and Engels' essential texts on law.)

Foucault, M. (1975) *Surveiller et punir: naissance de la prison*, Paris: Gallimard; *Discipline and Punish*, trans. A. Sheridan, London: Allen Lane, 1977. (Presents Foucault's analysis of aspects of disciplinary regulation exemplified by the emergence of the prison.)

Habermas, J. (1986) 'Law as Medium and Law as Institution', in G. Teubner (ed.), *Dilemmas of Law in the Welfare State*, Berlin: de Gruyter.(Sets out a view of law in terms of Habermas' developed social theory.)

Kelly, D.R. (1990) *The Human Measure: Social Thought in the Western Legal Tradition*, Cambridge, MA: Harvard University Press.(Explores interpretations of the character of society developed through the traditions of Western legal thought.)

Luhmann, N. (1989) 'Law as a Social System', *Northwestern University Law Review* 83: 136-50.(Outlines Luhmann's view of the social character of law as a communication system.)

Lukes, S. and Scull, A. (eds) (1983) *Durkheim and the Law*, Oxford: Martin Robertson.(Collects and introduces many of Durkheim's essential texts on law.)

Lyotard, J.-F. (1979) *La Condition postmoderne: rapport sur le savoir*, Paris: Éditions de Minuit; *The Postmodern Condition: A Report on Knowledge*, trans. G. Bennington and B. Massumi, Manchester: Manchester University Press, 1984.(A *locus classicus* of claims about the status and situation of knowledge and theory in postmodernity.)

Montesquieu, C. de (1748) *L'Esprit des lois (The Spirit of the Laws)*, trans. T. Nugent, London: Collier Macmillan, 1949.(A pioneer text of the emerging project of social theory, reinterpreting established structures of Western legal thought. This edition contains an important introduction by Franz Neumann.)

Tönnies, F. (1887) *Community and Association*, trans. C.P. Loomis, London: Routledge & Kegan Paul, 1955. (Develops the contrast between *Gemeinschaft* and *Gesellschaft* referred to in §1 above.)

Weber, M. (1922) *Wirtschaft und Gesellschaft (Economy and Society)*, trans. E. Fischoff *et al.*, Berkeley, CA: University of California Press, 1978. (Weber's posthumous *magnum opus*, including his extensive writings on sociology of law.)

Socialism

While socialist ideas may retrospectively be identified in many earlier forms of protest and rebellion against economic injustice and political oppression, socialism both as a relatively coherent theoretical doctrine and as an organized political movement had its origins in early nineteenth-century Europe, especially in Britain, France and Germany. It was, above all, a critical response to early industrial capitalism, to an unregulated market economy in which the means of production were privately owned and propertyless workers were forced to sell their labour power to capitalists for often meagre wages. The evils of this system seemed manifest to its socialist critics. Not only was the relationship between workers and capitalists inherently exploitative, and the commodification of labour an affront to human dignity, but it generated widespread poverty and recurrent unemployment, massive and unjust inequalities of wealth and economic power, degrading and soul-destroying work, and an increasingly atomized and individualistic society.

Socialists were not alone in criticizing some of these features of industrial capitalism and its accompanying ideology of economic liberalism. In particular, antipathy towards individualism was also a characteristic of conservative thought. But whereas conservatives found their inspiration in the hierarchically structured organic communities of the past, and were deeply hostile to the political radicalism of the French Revolution, socialists looked forward to new forms of community consistent with the ideals of liberty, equality and fraternity. For them, the evils of capitalism could be overcome only by replacing private with public or common ownership of the means of production, abolishing wage labour and creating a classless society where production geared to capitalist profits gave way to socially organized production for the satisfaction of human needs. In such a society, the human potential for a genuinely 'social' mode of existence would be realized, with mutual concern for others' wellbeing rather than unbridled pursuit of self-interest, with cooperation for common ends rather than competition for individual ones, and with generosity and sharing rather than greed and acquisitiveness - a truly human community.

For most nineteenth-century socialist theorists, the historic task of creating such a society was assigned to the organized industrial working class; most notably by Marx, the pre-eminent figure in the history of socialism. It was Marx who (along with Engels) provided the socialist movement not only with a theoretically sophisticated economic analysis of capitalism and a biting critique of its social consequences, but also, through his scientific, materialist theory of historical development, with the confident belief that the inherent contradictions and class antagonisms of capitalism would eventually give birth to a socialist society.

In marked contrast to such earlier optimism, contemporary socialists are faced with the continued resilience of capitalist societies and the collapse of at least nominally socialist regimes in the USSR and elsewhere, regimes in which state ownership and centralized planning have been accompanied by political repression and economic failure. For those who reject the idea that a suitably regulated form of welfare capitalism is the most that can be hoped for, the task is to construct some alternative model of a socialist economy which is preferable to this yet avoids the evils of centralized state socialism.

1 Defining socialism

The term 'socialism' was first used by Owenites in Britain and by Saint-Simonians in France in the 1820s and 1830s, and soon became widely adopted to refer both to a body of ideas critical of capitalism and to the future society that would or should replace it. But disputes about the meaning of 'socialism' - itself sometimes contrasted with 'communism', at other times taken to include it as a subcategory (see Communism) - have been endemic to its history, even more so than with its two main ideological counterparts, liberalism and conservatism. An enormous variety of theoretical positions and political movements have been termed 'socialist', by proponents and critics alike. Particular versions may be indicated by some qualifying term - as in state socialism, market socialism, guild socialism, revolutionary socialism, scientific socialism, ethical socialism, even national socialism (fascism); but there is no agreed classification of types since the relevant basis for this is itself subject to dispute. For some, what is crucial is the political means through which the desired future is to be achieved - for example, revolution or reform; for others, the specific nature of socialist economic institutions - for example, state planning or decentralized producer democracy; and so on.

An especially significant issue is whether socialism is to be defined normatively or institutionally: in terms of a set of values or ideals which socialists aim to realize, and which provide the basis for their critique of capitalism; or of the specific character of the economic institutions of a socialist society. The latter option has the obvious disadvantage of leaving unmentioned just why this institutional form should be seen as preferable to capitalism, and what makes it a system worth fighting for. The former option, by leaving the institutional requirements for socialism entirely unspecified, makes it a purely empirical question whether, as many who are now termed 'social democrats' would claim, capitalism can be modified so as to realize socialist values: there would then be no logical contradiction in calling such a capitalist society 'socialist'. But given the socialist tradition's opposition to capitalism on the grounds that no such modification is possible, it seems preferable to regard this social democratic thesis, if true, as a refutation of socialism, rather than as consistent with it (see Social democracy).

So it seems best to include in the definition of socialism both normative and institutional elements. For the socialist, the economic institutions of capitalism embody certain features and/or generate certain consequences that are objectionable from the standpoint of certain values; and there are possible alternative forms of economic organization which would either fully realize those values, or at least be markedly preferable to capitalism when judged in these terms. This, of course, is little more than a definitional schema, and leaves room for many varieties of socialism with respect both to the specific values involved and the specific form which a socialist economy might take.

On the latter question, while socialists have typically argued for the replacement of private by social, public or common ownership of the means of production, they have differed about what exactly this should involve - ownership by the state, by functional associations or local communities, by the members of producer cooperatives, and so on - and indeed about whether it is ownership or control that is crucial. Likewise, different solutions to the problems of economic coordination and allocation have been proposed: centralized planning by the state, decentralized planning, or even a market system shorn of its distinctively capitalist property relations. These different proposals are themselves often related to different views as to precisely what it is about capitalism that is objectionable and/or causally responsible for its ills: whether all relations of market exchange are undesirable, or only those involving the sale and purchase of labour power; whether it is private ownership of the means of production that is chiefly responsible for unjust inequalities in the distribution of economic goods, or the operation of a competitive market; and so on.

On the former question, of the specific character of socialist values, perhaps the main source of variation is the attitude taken towards political liberalism (see Liberalism). Some socialists have seen their task as engaging in an immanent critique of liberal democracy, broadly endorsing its declared values of freedom and equality but trying to show that their distinctively liberal interpretation is unduly narrow and restrictive: for example, by arguing for the extension of individual rights to include social and economic ones, and of democracy to include control over economic decisions. From this standpoint, liberal democracy is to be transcended - in the Hegelian sense of 'going beyond yet preserving' - rather than totally rejected; and the ideal of community is understood to involve harmonious relations between individuals who respect and enjoy one another's freedom. For others, by contrast, political liberalism - including its emphasis on the rule of law - is no more than an ideological facade of capitalism, so that, for example, its conception of legally enforceable individual rights has no place in a socialist society; and the value of community is understood in a more holistic manner.

With these broad points about the definition and variations of socialism in mind, we can proceed to examine some of the main arguments for socialism and the critical responses to these, focusing in turn on debates about economic efficiency, human wellbeing, democracy and power, and distributive justice. While the case for socialism typically begins by attributing various ills to capitalism, it depends also on being able to show that there is some alternative system in which these would be absent or greatly reduced. Correspondingly, critics of socialism may deny either that the supposed ills are properly regarded as such, or that they are attributable to capitalism intrinsically rather than to contingent features of particular capitalist societies; and/or they may argue that the proposed socialist alternative fails to overcome these ills, or that it does so only at the cost of producing further ones of its own.

2 Efficiency, planning and markets

Socialist critics of capitalism have frequently pointed to its economic failings. Its reliance on anarchic market forces makes it liable to cycles of booms and slumps, subjects workers to recurrent periods of unemployment and

permanent insecurity, and fails systematically to make effective use of human and natural resources. For all its dynamism in producing goods to satisfy increasingly exotic consumer preferences, it leaves many basic needs unmet. It underprovides public goods, cannot incorporate the social costs of economic activity and is destructive of environmental and other preconditions of long term economic development. For many socialists the solution has been seen to lie in a centrally planned economy in which resources can be effectively and rationally directed to satisfy human needs while also meeting other social and ethical objectives.

This case against the market and in support of planning has been met by three main kinds of argument: neoclassical, motivational and epistemic. Neoclassical economic theorists claim that a market economy is demonstrably ‘efficient’ in the sense of being Pareto-optimal: the outcomes it produces cannot be departed from without at least some individual(s) experiencing a reduced level of preference satisfaction (see Economics and ethics §3). Admittedly this can be shown only for ‘ideal’ markets: for example, where there are no public goods or externalities. But such cases of (actual) ‘market failure’, it is claimed, can be dealt with either by limited forms of state intervention, or by assigning property rights to currently unowned goods and resources.

The motivational argument is that socialism must fail because it makes unrealistic demands on the potential altruism of economic agents. Workers must be expected to perform their tasks without the promise of differential material rewards, and planners to put aside their own interests in making economic decisions. But humans are by nature primarily self-interested beings: the supply of altruism is strictly limited and cannot be relied upon as the basis for economic organization. The market takes advantage of this fact by providing incentives for workers and managers that ensure the allocation of resources to those spheres of production where they will be most beneficially employed. Without such incentives, socialist alternatives to the market will inevitably fail.

The epistemic argument, developed in the Austrian school of economics and especially by Hayek, is that only the market can solve the economic problem of ignorance (see Hayek, F.A. von §2). This problem is due to the division of knowledge in society - its dispersal among different economic agents; and to the practical nature of much of this knowledge (knowledge ‘how’ rather than knowledge ‘that’), which cannot be articulated in propositional form, and hence cannot in principle be acquired by a centralized planning agency. Thus no economic plan can gain access to and utilize all the knowledge relevant to economic decisions. By contrast, the price mechanism manages to distribute the knowledge relevant for economic coordination while allowing individual agents to rely on their own ‘local’ knowledge in the decisions that concern them.

Two main responses to these arguments may be made by socialists. The first is broadly to accept them, but to argue that they do not amount to a defence of capitalism but only of the market, which may itself take either capitalist or non-capitalist forms. This is the line taken by market socialists. The market is, *ceteris paribus*, the most efficient and dynamic economic mechanism, but it can be constructed in a form compatible with socialism by eliminating its distinctively capitalist features, notably private ownership in the means of production and wage labour. Among numerous versions of market socialism, probably the closest to traditional socialist aspirations involves social ownership in the form of workers’ cooperatives whose assets are either directly owned by their members or leased to them by the state.

The second response is to challenge those arguments for the economic superiority of the market. For example, the neoclassical concept of efficiency may be criticized for defining this in relation to the satisfaction of individual wants or preferences, whatever their specific content or rationale. To show that markets are efficient in this sense is to show very little: it ignores the distinction between wants and needs, and fails to discriminate between preferences in terms of their contribution to human flourishing or wellbeing.

To the motivational argument that socialism makes unrealistic demands on altruism, there are several possible replies. One is to say that egoism is not a part of human nature, but the product of living in a capitalist society: a far greater degree of altruism can be expected in the different institutional context of socialism, either because such altruism is itself natural to humans, but inhibited by capitalism, or because human motivations are highly malleable in this respect. Alternatively, the ways in which self-interest and altruism are typically conceived and contrasted in these debates may be challenged. Actions oriented towards the wellbeing of others, and the relationships within which these take place, are often themselves a significant source of wellbeing to those concerned (see Human nature §3). The kinds of self-interest the market relies upon are unduly narrow and restrictive in character: socialism does not require self-sacrificial forms of altruism, but instead makes possible the satisfaction of a more

extensive and fulfilling range of interests. But the problem remains of how workers in a non-market economy are to be motivated to move to where they are most needed.

Against the epistemic argument for the market it may be objected that the market's informational virtues are much exaggerated. The price mechanism by no means distributes all the information relevant for economic coordination: indeed competitive pressures often generate disincentives to communicate relevant information about other producers' plans, scientific and technical innovations, and so on. Nor is the information required for coordination the only information relevant for economic activity, as the frequent exclusion of environmental impacts from market decisions illustrates. Admittedly, even if such criticisms of the informational virtues of the market can be sustained, there remains a strong epistemic case against centralized economic planning. Yet there may be institutional forms superior both to markets and to centralized planning for at least some tasks, such as those involved in the organization of scientific communities. Whether these provide plausible models for economic coordination is unclear.

3 Wellbeing and the human good

The replies to both the neoclassical and motivational arguments for the market involve claims about the nature of the human agent and of human wellbeing. In doing so they express a recurrent concern on the part of its socialist critics with the market's failure to establish the necessary conditions for human flourishing, and its encouragement of attitudes, motivations and character traits damaging both to those who acquire these and to those affected by them. For example, it is claimed that the market issues in such vices as competitiveness, avarice, egoism, possessiveness and vanity, all at the expense of the proper virtues of character and the humanly beneficial social relationships of community; that the pleasures of private consumption are mistakenly privileged over the more fulfilling demands of public life; and that capitalist forms of production deny to workers the exercise both of their active and creative capacities in work and of their deliberative capacities in the democratic control of the productive process (see Work, philosophy of §2).

Marx's critique of capitalism as a condition of 'alienation' is typical here (see Alienation), but similar arguments against capitalism for its development of particular kinds of human character and social relationships incompatible with human wellbeing are to be found amongst other socialist theorists, from Morris and Tawney in the British tradition to Fromm and Marcuse in the continental. In current philosophical terminology such criticism is 'perfectionist' in form: it claims that social and political institutions should be judged by reference to a specific conception of the human good (see Perfectionism). It is often broadly Aristotelian in content, seeing human wellbeing as consisting of the development and exercise of species powers and capacities. Hence it rejects, *inter alia*, the utilitarian, preference-based accounts of wellbeing that often underpin the neoclassical concept of efficiency: economic institutions are to be assessed not by their ability to satisfy given preferences, but in terms of the nature and value of the preferences they themselves encourage and make it possible to satisfy (see Welfare §1).

Two kinds of response may be made to such claims. The first, which is typical of much modern liberal thought, is to reject perfectionist arguments altogether. It is a mark of liberal institutions that they be neutral between different conceptions of the good (see Neutrality, political). Such neutrality is especially desirable given the pluralistic character of modern societies, in which diverse and irreconcilable conceptions of the good are espoused. In this situation, perfectionism must imply the imposition of a contested conception of the good by coercive means, and hence a political practice which is paternalistic or even authoritarian. By contrast, the market is consistent with liberal neutrality. It provides an economic framework in which individuals with quite different ends and beliefs about the good can pursue these through mutually beneficial free exchanges.

The second response accepts the legitimacy of perfectionist arguments but denies that markets are incompatible with the human good. This may be argued either by denying that the market necessarily has the ill effects that socialists attribute to it, or by attributing to the market effects deemed highly desirable in terms of some alternative conception of the good. In the latter case it may be claimed, for example, that the market has the great merit of fostering individual autonomy; although socialists may respond to this by arguing that the development of such autonomy in fact requires the existence of certain kinds of social relationships which the market tends to undermine.

4 Democracy, power and freedom

The claim that socialism produces the conditions for the realization of liberal political values has been central to arguments for socialism that appeal to the values of democracy and freedom. The argument can take both weaker and stronger forms. The former asserts that many of the standard liberal rights and freedoms are empty without the material conditions for their effective exercise, which the market systematically fails to guarantee. This view is sometimes stated in terms of a criticism of ‘negative’ conceptions of liberty as the absence of coercion, proposing instead that these material conditions should be included in the definition of liberty itself (see Freedom and liberty §3). Alternatively it may simply be argued that if (negative) liberty is valuable, then so too are the conditions for its exercise. In response to the objection that the market’s unequal distribution of material conditions is the unintended consequence of a spontaneous order, and hence should not be seen as a constraint on freedom, the socialist may argue that the market is not the outcome of natural events but of social decisions aimed at its creation and/or maintenance, that its distributive consequences are both foreseeable and alterable, and that there is no justification for restricting one’s conception of liberty to the absence of intentional constraints.

The stronger form of the claim that socialism realizes liberal values is that it represents their consistent application to the economic sphere. A version of this position can be found in Marx’s early writings, in which the ‘ideal’ community of the liberal democratic polity is contrasted with the egoistic realm of modern civil society; correspondingly, the rights of citizens exercised through their participation in the political community are set against the ‘rights of man’ exercised by the private individuals of civil society. The project of socialism is then expressed as bringing the ideal world of the polity down to reality through the democratic transformation of economic life. This strategy of pointing to the divergence between the rights and freedoms of the ideal liberal political order and their absence in the economic sphere has also been employed in other contexts: for instance, by contrasting democratic participation in the political system with the authoritarian nature of power relationships within the capitalist firm.

To such claims that socialism represents the completion of the democratic project, two main responses may be made: the liberal and the radical. The first denies that there is a conflict between the capitalist economic order and the liberal polity: rather, the existence of a market order, in which individuals enter freely into voluntary contractual relations, is itself a condition of the political rights and freedoms that define the liberal polity. It is not a sufficient condition: authoritarian states are compatible with capitalism. However, it is a necessary condition: political liberties are only to be found in free market economies. The socialist order, by concentrating economic, social and political power, destroys the space of civil society in which individuals enter into voluntary relations independent of the state, and which provides the cultural and social conditions for opposition to state power (see Civil society). This objection applies primarily to centrally planned socialist economies with state ownership of the means of production, and is often endorsed by market socialists as a further reason to support their own proposals. Alternatively, it may be argued that the existence of such concentrations of economic and political power is compatible with restraint in its exercise, and that the absence of liberal political rights in state socialist societies such as the erstwhile USSR can be explained by reference to factors other than the absence of markets - for example, to the lack of liberal democratic institutions prior to their revolutionary transformation.

The second, radical response is concerned not so much with the cogency of the socialist critique of liberalism but with its completeness. In effect it extends the socialist criticism of liberalism to socialism itself. By placing its emphasis on the economic sphere, socialism has been blind to other sources of power, in particular those concerned with race and gender. Thus the feminist criticism of traditional socialism: that by confining its attention to the ‘public’ sphere of economic and political life it fails, like liberalism, to address the primary origin and location of women’s oppression, in the ‘private’ sphere of the family and in ‘personal’, including sexual, relations between men and women (see Feminist political philosophy §4). Two replies may be made by socialists to this kind of critique. The first is to claim that these other asymmetries of power can themselves be explained in terms of the organization of economic production. The second is to accept the existence of diverse sources of social power and to re-conceptualize the traditional socialist project in the economic sphere as simply one component of a wider programme of human emancipation.

5 Distributive justice

The final set of arguments for socialism to be considered here are distributive ones. Capitalism is criticized for the unequal and/or unjust distribution of material, social and cultural goods. But this criticism takes a number of

different forms. For some socialists, the preferred distributive principle is strict equality, either contrasted with justice or regarded as its proper interpretation. Others espouse some not necessarily egalitarian principle of justice: distribution according to need is the most common, as in the well-known principle, ‘from each according to their ability, to each according to their needs’; but appeal may also or instead be made to a principle of desert, such as reward in proportion to contribution. While the last of these is rarely proposed as the sole distributive principle for a socialist society, it has often figured in socialist criticisms of capitalism, especially in the claim that capitalist profits are undeserved and hence unjust. But there is disagreement among socialists about whether the distribution of goods in a non-capitalist market economy would be unjust from this standpoint.

Defenders of capitalism have sometimes argued that it can be shown to be just by reference to a principle of desert: for example, by arguing that profits are a deserved reward for risk-taking. But more commonly they reject altogether the legitimacy of desert-based, need-based, egalitarian or any other so-called ‘patterned’ principles of justice in judging economic systems, on the grounds *inter alia* that any attempt to realize such patterns will involve unjustifiable and systematic coercion by the state. Instead, it is argued, justice should be understood as a purely ‘procedural’ concept: distributions are just if they are the outcome of fair or appropriate procedures, whatever the resulting pattern may be. An especially favoured procedure is voluntary exchange between free and equal parties. It is then claimed that since capitalist market transactions consist exclusively of such exchanges, capitalism is a just system (see Libertarianism; Nozick, R. §2).

For socialists, such procedural definitions illicitly reduce the concept of justice to that of (negative) liberty. But it may also be argued that even if adopted, they would fail to show that the outcomes of the transactions between capitalists and workers are just. For although the sale of labour power in return for wages is a transaction between formally or legally free and equal parties, there is an absence of substantive freedom on the part of workers, coupled with marked asymmetries of power. In addition, socialists may point to the historical origin of most current capitalist holdings of private property in past acts of theft, fraud, violence or state coercion.

While arguments for socialism couched in terms of distributive justice and equality have been widely employed, they have often been criticized by Marxists, partly for assuming that the crucial defect of capitalism lies in the improper distribution of goods in the sphere of exchange, rather than in the power relations in the sphere of production consequent upon capitalist property relations. Hence emphasis is placed instead on the concept of exploitation, understood as the extraction by capitalists of surplus value from workers which accrues to the capitalist in the form of profits - something that cannot be remedied by higher wages but only by the abolition of wage labour. Whether this concept of exploitation and its associated theory of value can be sustained is much disputed amongst socialist theorists, some of whom have argued instead for a purely distributional interpretation of the concept: for example, that someone is exploited if they would be better off than they now are had there been an initially egalitarian distribution of ownership rights over the means of production.

6 The future of socialism

The near universal collapse of nominally socialist regimes since the late 1980s has led many critics of socialism to proclaim its death. But these regimes have always had their socialist critics also, for many of whom these historic events may prove welcome, not least in undermining the previously hegemonic status of Marxist-Leninism in socialist theory and practice, and the marginalization of other significant traditions of socialist thought. Thus one response to the political and economic failures of state socialism has been to return to some of these earlier, non-centralist socialist traditions, and attempt to re-work them for contemporary purposes. One such attempt is that of the ‘associational socialists’ who, taking their inspiration from nineteenth- and early twentieth-century guild socialism and syndicalism, propose independent and self-governing functional associations as the units of political and economic authority, rejecting both state and market.

By contrast, as noted earlier, the project of ‘market socialism’ is to construct a non-capitalist market system operating within liberal democratic political institutions. This position has the virtue of being able to provide a relatively well-articulated alternative to capitalism which takes account of the powerful objections to centralized planning. Yet its acceptance of market forces, of individual self-interest and relationships of exchange and competition, makes it appear to non-market socialists a poor substitute for the ‘truly human’ community to which they aspire. To this its proponents may reply that one should not aim at a single, monolithic ideal of community for society as a whole, but accept instead a more differentiated conception of social existence in which different forms

of human wellbeing are realized within different spheres or domains, of which the economic is but one. Yet it remains unclear whether market economies are compatible with, or inimical to, the flourishing of significant forms of community outside the economic sphere (see Communty and communitarianism).

Further, like any market system, market socialism is subject to a range of objections from an ecological or environmental perspective. The market fails to take account of the environmentally destructive consequences of economic growth; it is unable to incorporate the interests of those who cannot engage in market transactions - whether the poor, members of future generations, or non human beings; and it encourages people to misidentify the primary source of wellbeing as the endless pursuit of consumer satisfactions. Indeed, whilst the environmental movement has presented a serious challenge to the tendency of much socialist thought to conceive of human emancipation as requiring the subordination of nature to human ends, it has also given new life to many traditional socialist objections to the market: its remarkable ability to generate collectively irrational outcomes from individually rational behaviour, including the underproduction of public goods and the overproduction of public ills. For both political ecologists and socialists, there is a vast range of social problems which require collective rather than individual action for their solution, and a continuing need for forms of ethical and political commitment that the market both fails to recognize and may often undermine (see Green political philosophy).

See also: Bakunin, M.A.; Bernstein, E.; Engels, F.; Gramsci, A.; Kautsky, K.J.; Lassalle, F.; Lenin, V.I.; Luxemburg, R.; Market, Ethics of the §§1-3; Marx, K.; Marxism, Western; Marxist philosophy, Russian and Soviet; Political philosophy, history of; Proudhon, P.-J.

RUSSELL KEAT
JOHN O'NEILL

References and further reading

Berki, R. (1975) *Socialism*, London: Dent.(A sophisticated analysis including discussion of Third World socialism and historical uses of the terms 'socialism', 'communism' and 'social democracy'.)

Bottomore, T. (1990) *The Socialist Economy*, New York: Harvester Wheatsheaf.(A short and readable survey of different theoretical and practical forms of socialist economy and debates about their defensibility.)

Buchanan, A. (1985) *Ethics, Efficiency and the Market*, Oxford: Clarendon Press. (Includes a relatively non technical account of neoclassical defences of the market and their problems.)

Cohen, G.A. (1988) *History, Labour and Freedom*, Oxford: Clarendon Press.(Contains several essays providing sympathetic analytical reconstructions of central normative concepts in Marx's work, including exploitation, coercion and freedom.)

Cole, G.D.H. (1953-60) *A History of Socialist Thought*, London: Macmillan, 5 vols.(A monumental and still unsurpassed history of socialist thought, especially in Europe.)

Fried, A. and Sanders, R. (eds) (1992) *Socialist Thought: A Documentary History*, New York: Columbia University Press.(An extensive collection of extracts from the work of major nineteenth- and twentieth-century socialist theorists, including non-European writers.)

Gray, J. (1993) *Beyond the New Right*, London: Routledge.(An accessible account of epistemic objections to socialism combined with a perfectionist defence of the market pointing also to its limitations.)

Hayek, F.A. (1973-9) *Law, Legislation and Liberty*, London: Routledge, 3 vols.(Arguably the major work of the most influential twentieth-century defender of markets and critic of socialism.)

Hirst, P. (ed.) (1989) *The Pluralist Theory of the State: Selected writings of G.D.H. Cole, J.N. Figgis and H.J. Laski*, London: Routledge.(A selection of writings by associational theorists, with a useful editorial introduction.)

Kolakowski, L. (1978) *Main Currents of Marxism*, Oxford: Oxford University Press. (An influential, informative and highly critical analysis of the main forms of Marxist thought through the nineteenth and twentieth centuries.)

Levine, A. (1988) *Arguing for Socialism*, London: Verso.(Argues that socialism can be justified by reference to values central to liberal political thought.)

Lichtheim, G. (1983) *A Short History of Socialism*, London: Fontana.(A classic history of socialist thought and the socialist movement in Europe and the USA.)

McLellan, D. (ed.) (1977) *Karl Marx: Selected Writings*, Oxford: Oxford University Press.(A useful collection of Marx's work, including his early critiques of alienation and the separation of state and civil society.)

Miller, D. (1989) *Market, State and Community*, Oxford: Oxford University Press.(A defence of market socialism responding both to traditional socialist objections to the market and contemporary liberal objections to

socialism.)
Phillips, A. (1993) *Democracy and Difference*, Polity Press: Cambridge.(A sympathetic but critical discussion of socialism's failure to develop an account of democracy which recognizes forms of oppression based on gender, race, and so on.)
Polanyi, K. (1957) *The Great Transformation*, Boston, MA: Beacon Press.(A broadly socialist but non-Marxist historical account of the emergence of the market economy.)
Roemer, J.E. (1988) *Free to Lose: An Introduction to Marxist Economic Philosophy*, London: Century Hutchinson.(A rational choice reinterpretation of Marxist theory, presenting exploitation as the outcome of unequally distributed economic assets.)
Rubel, M. and Crump, J. (eds) (1997) *Non-Market Socialism in the Nineteenth and Twentieth Centuries*, London: Macmillan.(A collection of sympathetic essays on the main currents of non-market socialism.)
Ryle, M. (1988) *Ecology and Socialism*, London: Century Hutchinson.(Argues that neither ecological politics nor socialism can do without one another.)
Sypnowich, C. (1990) *The Concept of Socialist Law*, Oxford: Clarendon Press.(Argues that legal rights and the rule of law have a positive role in socialism, rather than being tied to market societies.)
Wright, A. (1987) *Socialisms: Theories and Practices*, Oxford: Oxford University Press. (Perhaps the best brief introduction, emphasizing the diversity of socialist thought and sympathetic to the ethical socialism of the British tradition.)

Society, concept of

The term 'society' is broader than 'human society'. Many other species are described as possessing a social way of life. Yet mere gregariousness, of the kind found in a herd of cattle or a shoal of fish, is not enough to constitute a society. For the biologist, the marks of the social are cooperation (extending beyond cooperation between parents in raising young) and some form of order or division of labour. In assessing the merits of attempts to provide a more precise definition of society, we can ask whether the definition succeeds in capturing our intuitive understanding of the term, and also whether it succeeds in identifying those features of society which are most fundamental from an explanatory point of view - whether it captures the Lockean 'real essence' of society.

One influential approach seeks to capture the idea of society by characterizing social action, or interaction, in terms of the particular kinds of awareness it involves. Another approach focuses on social order, seeing it as a form of order that arises spontaneously when rational and mutually aware individuals succeed in solving coordination problems. Yet another approach focuses on the role played by communication in achieving collective agreement on the way the world is to be classified and understood, as a precondition of coordination and cooperation.

1 Defining social interaction

Max Weber defines social action as 'action in which the meaning intended by the agent or agents involves a relation to another person's behaviour' (1978: 7), and a social relationship exists 'when several people reciprocally adjust their behaviour to each other with respect to the meaning they give to it' (1978: 30) (see Social action). Thus several people acting in a similar way need not be engaging in social action, but two cyclists who try to give way to each other are. Weber excludes non-human animals from the realm of the social on the grounds that they are not capable of the intentions that might give meaning to their behaviour. (His paradigm of 'meaningful action' is the rational pursuit of a consciously chosen goal.) In that respect, Weber's definition of the social may seem unduly narrow, but it also seems to be over-inclusive. A population of individuals who took account of each other's behaviour only to the extent of seeking to avoid collisions would not on that account qualify as living a social way of life, not at least by any intuitive criterion. We may also find it odd that cooperation is not a necessary feature of society on Weber's account.

Alfred Schütz (1932) seeks to improve on Weber's definitions, telling us that we have a social relationship where there is mutual awareness of the way each sees the other (see Schütz, A.). This requirement identifies an important feature of human interaction, but it is still not clear that it captures what makes it 'social' interaction. A sophisticated form of the predator-prey relationship, in which the predator is aware that the prey is aware of it and vice versa, would seem to qualify as a social relationship by this criterion. More promising is Schütz's definition of 'social interaction', which involves, in addition to the requirement of mutual awareness, an attempt by one party to affect the consciousness of the other. Thus social interaction is defined as an attempt to affect another's state of mind in which the other is aware of one's intentions and is also aware that one knows that they are aware of one's intentions. Schütz also notes that this is what we call communication. The work of H.P. Grice develops this approach to communication, invoking yet more complex levels of mutual awareness. Though Grice himself is not concerned to define the social, the kind of mutual awareness to which he appeals is clearly not confined to linguistic interaction, and the idea that communication is a central feature of social interaction has proved fruitful (see §3 below).

2 Game theory and social order

The effect of an exclusive focus on social interaction as a relationship between individuals is to make social order seem a merely incidental and inessential feature of social life. Intuitively that is odd. More recent individualist approaches employing the resources of game theory (see Rational choice theory §2) have taken it for granted that both social order and cooperation are defining features of social life, and have seen it as their task to demonstrate that these features can be understood as the outcome of interactions between individuals pursuing their own individual purposes.

In some respects, game theory represents a further refinement and idealization of the assumption of mutual

awareness. Individuals are assumed to be instrumentally rational in that they are seeking to maximize the satisfaction of their preferences. This need not mean that they are selfish, but it does mean that the goals they are pursuing are their own and not, for example, a collective goal. It is also ‘common knowledge’ between them that they are rational. (This amounts to saying that each knows that the others know this, and knows that the others know that they know this, and so on *ad infinitum*.) It is also assumed that there is common knowledge of the possible choices of action open to each party and of the payoffs attached to possible outcomes. There is to be no communication between individuals, but each is able to assess what it would be rational for others to do and can take that into account in deciding what it is rational for them to do. The claim is that, given these assumptions, where there is a mutual interest in coordination (for example, in everyone driving on the same side of the road), individual rational choices will give rise to uniformities in behaviour without the need for explicit agreement or social sanctions. Lewis (1969) makes this thought the basis of his definition of ‘convention’ as - very crudely - a regularity in behaviour that everyone keeps to because it is in their interest to do so given that everyone else does (see Social norms §1). Using this account of convention, Lewis seeks to account for linguistic and other forms of communication as forms of game-theoretically rational action with the complexity of levels of mutual awareness to which Grice appeals.

The game-theoretic approach promises not just an explanation of social order but an account of what social order is, for clearly not just any regularity in the behaviour of a population will count as an example of social order. Social order arises, on this view, when rational and sufficiently mutually aware individuals succeed in solving coordination problems. However, it is controversial whether the rational agents of game theory can ever do this, and it is in any case widely conceded that there are aspects of social order that cannot be understood in this way. There are many circumstances in which, though the collective interest would be served by everyone obeying a rule, it is in the individual’s interest to break it. In these circumstances game theory predicts that no one will obey the rule and so all will lose out. This is known as the ‘free-rider’ problem. It does not help to appeal to sanctions against rule breakers, for we then need to explain how it could be rational for one individual to seek to impose sanctions on another, and the free-rider problem arises all over again. Thus the game-theoretic approach fails to explain a form of cooperation that is central to human social life.

One way of responding to this failure is to modify our assumptions about individual motivation, for example by building in a degree of concern for the interests of others, though interestingly it is not clear that this is enough. Elster (1989) argues that we also need to invoke a respect for social norms, but this suggestion faces a dilemma. Either we appeal to a respect for norms in general along the lines of Kant’s categorical imperative (see Kant, I. §11), whose implications threaten to be insufficiently determinate to help with our problem, or else we appeal to the respect individuals feel for the norms of their particular society, in which case we beg the question of the explanation of social order by treating social norms as already given.

3 Society as a system of communication

For G.H. Mead and for the Durkheimian tradition, especially as it has been developed in anthropology, the key feature of social interaction is communication (see Mead, G.H.; Durkheim, É.). Not just language but action in general is seen as having communicational or expressive significance, and its function is not simply to transmit useful information but to get agreement on the way the social world is to be classified and understood. A shared way of interpreting and defining reality, sustained by its expression in language, ceremony and ritual, is on this view a precondition, not only of mutual intelligibility, but of the kind of coordination and cooperation between individuals that game theory has such difficulty in explaining. Communication is also an important feature of animal social life, and it is possible to see it as having a similar function there too, even if the forms of expression and the schemes of classification involved are much more restricted.

Even so, it is clear that not just any system of communication constitutes society. The simple transmission of information, as in alarm calls in flocks of birds, is not enough. Nor is it enough that communication yields a shared scheme of classification or definition of reality. The defence of territorial boundaries, for example, yields an agreed definition of reality that is consistent with relationships between individuals being wholly antagonistic. It seems best to say that we have something we can call society only when the agreed definition of reality provides for the possibility of cooperation; though if we are to avoid seeing cooperation between parents in raising offspring as in itself amounting to a social way of life, we must add that the agreement on ways of defining reality must

extend beyond two individuals. This is not an arbitrary restriction, for beyond that number agreement becomes significantly more difficult to achieve. Inevitably, questions of classification will sometimes be a matter for dispute, making claim and counterclaim endemic to social life. Thus while cooperation is, by this definition, a necessary feature of social life, so too is a distinctively social form of conflict.

For the Durkheimian tradition, the key distinguishing feature of *human* society is the normative dimension, the existence of shared standards of correct thought and behaviour. This means not simply the fact of general conformity to standards of correct thought and behaviour but the fact of those standards being collectively policed. For Durkheim, the authority behind social norms is the authority of the collective viewed as an organized group, but Mead's idea of 'the generalized other' suggests a different view. In the absence of shared standards of correctness and legitimacy, agreement emerges, if it emerges at all, from the play of power in disputes between interested parties, supported perhaps by friends and allies. The existence of shared standards means that parties in dispute can appeal to the disinterested jury of others in general, and that they will tend to accept the verdict of that jury. Some standards, of course, are likely to be policed in a more organized way (for example the criminal law), but more generally, and more fundamentally, shared standards are policed by the authority of others in general. For Mead, it is by taking up the point of view of others in general that an individual comes to possess their own sense of what is correct and legitimate, and also a sense of themselves as an individual.

If communication is to be seen as the basis of society, it cannot be viewed as a form of instrumentally rational action of the sort modelled by game theory (see §2 above). Communication must be capable of operating independently of, and at a more primitive level than, the degree of rationality and mutual awareness assumed by game theory. As social beings, we must be supposed to possess an urge to express our own view of things, and also a tendency to be influenced by the way others see things, that is prior to our rationality and which constitutes a distinctive source of motivation in its own right. Thus Jürgen Habermas (§2) distinguishes between instrumental action which is 'oriented to success', and communicative action which is 'oriented to reaching understanding'. Habermas also stresses that, in a context of shared standards of correctness, acts of communication take on the force of legitimacy claims. The anthropologist Mary Douglas (1977) makes a similar distinction between instrumental action and expressive action, but emphasizes the competitive aspect of social communication. Our expressive activity reflects a need to give meaning to the world, a need, in fact, to construct an intelligible world as a condition of remaining sane. But my attempts to win agreement to my interpretations of reality may conflict with yours. To see society as a system of communication is to see it as a battleground of conflicting interpretations of reality in which the price of failure to win agreement is exclusion from social interaction.

4 Social groups

For methodological individualists like Weber, talk of social collectivities can only be a way of talking about 'actual or possible social actions of individual persons' (see Holism and individualism in history and social science). Elster (1989) offers one way of putting flesh on this idea, defining a society as a geographical area for which there is a local maximum of the number of transactions between the individuals in the area divided by the total number of transactions in which those individuals are involved. Shorn of the reference to a geographical area, this could serve as a definition of social class, for class boundaries are certainly barriers to social interaction, but most social groups are thought of as having a greater degree of unity than this.

For Georg Simmel (1908), the unity of a social group consists in the fact that its members think of it as possessing a unity. Gilbert (1989) develops this idea in a way that recalls Rousseau's distinction between the general will and the will of all (see Rousseau, J.-J. §3). Appealing to what she argues is the central use of the first person plural 'we', she defines a social group as a 'plural subject' constituted by the fact that each of a number of individuals finds it appropriate to use 'we' in attributing purposes, activities, beliefs, and so on, to themselves jointly rather than severally. The idea of a plural subject offers us a less individualist conception of social cooperation than that which emerges from game theory, though seen from a communication system perspective (see §3 above), the idea that 'we' constitute a plural subject is just one of the agreed ways of defining reality which can serve to facilitate coordination and cooperation.

See also: Anthropology, philosophy of; Bushi philosophy; Confucian philosophy, Chinese; Confucian philosophy, Japanese; Ibn Khaldun; Neo-Confucian philosophy; Watsuji Tetsurō

ANGUS ROSS

References and further reading

Douglas, M. (1977) 'Why do people want goods?' in S. Hargreaves Heap and A. Ross (eds), *Understanding the Enterprise Culture*, Edinburgh: Edinburgh University Press, 1992.(An example of the Durkheimian approach described in §3.)

Douglas, M. (1987) *How Institutions Think*, London: Routledge, chaps 1-4.(An example of the Durkheimian approach described in §3.)

Durkheim, É. (1895) *Rules of Sociological Method*, trans. W.D. Halls, New York: Free Press, 1982.(The classic statement of Durkheim's position.)

Elster, J. (1989) *The Cement of Society*, Cambridge: Cambridge University Press. (An example of the game-theoretic approach described in §2. Also contains the definition of 'social group' referred to in §4.)

Gilbert, M. (1989) *On Social Facts*, London: Routledge.(The account of social groups as plural subjects described in §4.)

Grice, H.P. (1957) 'Meaning', *Philosophical Review* 66: 377-88.(Definition of 'communication' in terms of mutual awareness, referred to in §1.)

Habermas, J. (1981-2) *The Theory of Communicative Action*, vols 1 and 2, trans. T. McCarthy, Boston, MA: Beacon Press, 1984-7.(Seeks to integrate a Weberian social action approach with a Durkheimian communication system approach, see especially chapters III and V.1.)

Lewis, D.K. (1969) *Convention*, Cambridge, MA: Harvard University Press.(An example of the game-theoretic approach described in §2.)

Mead, G.H. (1934) *Mind, Self and Society*, Chicago, IL, and London: University of Chicago Press.(A seminal account of the communication perspective.)

Schütz, A. (1932) *The Phenomenology of the Social World*, London: Heinemann, 1972.(Definition of 'social relationship' and 'social interaction' in terms of mutual awareness, referred to in §1. See especially chapter 4.)

Simmel, G. (1908) 'How is society possible?' in *Georg Simmel: On Individuality and Social Forms*, ed. D.N. Levine, Chicago, IL and London: University of Chicago Press, 1971.(Contains definition of 'social group' referred to in §4.)

Ullman-Margalit, E. (1977) *The Emergence of Norms*, Oxford: Oxford University Press.(An example of game-theoretic approach described in §2.)

Weber, M. (1978) 'The nature of social action' in *Max Weber: Selections in Translation*, ed. W.G. Runciman, Cambridge: Cambridge University Press, ch. I, §1. (Seminal definition of social action quoted in §1.)

Socinianism

Socinianism was both the name for a sixteenth- and seventeenth-century theological movement which was a forerunner of modern unitarianism, and, much less precisely, a polemic term of abuse suggesting positions in common with that 'heretical' movement. Socinianism was explicitly undogmatic but centred on disbelief in the Trinity, original sin, the satisfaction, and the natural immortality of the soul. Some Socinians were materialists. Socinians focused on moralism and Christ's prophetic role; the elevation of reason in interpreting Scripture against creeds, traditions and church authority; and support for religious toleration. The term was used polemically against many theorists, including Hugo Grotius, William Chillingworth, the Latitudinarians, and John Locke, who emphasized free will, moralism, the role and capacity of reason, and that Christianity included only a very few fundamental doctrines necessary for salvation.

1 The Socinian movement

Socinianism was founded by Laelius and Faustus Socinus, and was the name given to an unorthodox evangelical theological movement which originated in Italy and spread first to Poland, and then to the Netherlands and England. At its height the academy of the Socinian Minor Church in Racov had over a thousand students, and its press published works which were disseminated in many languages and countries. In many countries, however, Socinians were punished or executed as 'heretics' or forced into exile, and the Socinian church was crushed in Poland in 1660. Many Socinians then settled in the Netherlands. The leading Socinian authors in the generations after Faustus Socinus were Johann Crell, Johann Völkel, Johann Von Wolzogen, Andreas Wiszowaty and Samuel Przypkowski.

2 The Trinity and satisfaction

Socinianism was explicitly undogmatic, and many of its beliefs changed over time, but throughout it centred on disbelief in the Athanasian Trinity of three co-eternal consubstantial persons in one Godhead in favour of the belief that Jesus was a divinely inspired man who brought the definitive revelation or Word of God, and that the Holy Spirit was a power or energy of God. Jesus was distinguished from the rest of mankind by his perfect holiness of life, by being conceived by the Holy Spirit and by being endued with divine wisdom. At his resurrection he had been given power to rule over the world. Some Socinians argued that a plurality of infinite persons in one God was impossible. Intimately connected to denial of the Trinity was denial of orthodox Calvinist accounts of Christ's satisfaction. It was held by many Socinians that Christ had not paid an equivalent for humanity's sins in order to satisfy God's justice; his mission was centrally prophetic rather than priestly. God forgave sins, rather than receiving payment for them.

3 Original sin, mortalism and materialism

Socinians denied original sin and held the 'mortalist' belief that human souls were not naturally immortal. For Socinians, Adam's sin had not corrupted the power of his posterity to choose good over evil. They maintained that humans had free will, and opposed predestination, in part to protect God from responsibility for sin. They simultaneously maintained that humans were sinful, by emphasizing both that individuals were responsible for their own sins and that all generations had engaged in extensive sinning. They depicted naturally mortal humans as left to their natural mortality unless saved by the power of the Spirit in following the commandments of Jesus. They denied both the natural immortality of the soul and the resurrection of the earthly body in favour of belief that there would be a selective resurrection of the saved providing them with immortality in celestial bodies. The eternity of punishment for sin was denied in favour of belief in an eternal fire causing annihilation of the damned. Some Socinians combined mortalist belief with materialism and held that matter was coexistent and coeternal with God; Faustus Socinus himself held that the world had been made from pre-existent matter. Many Socinians depicted Christ as bringing information about an afterlife and a promise of immortality inaccessible by reason alone, with Christ's resurrection itself serving as a central testimony to immortality.

4 Reason and toleration

Socinians placed a heavy emphasis on the role of reason or 'right reason' as the sole and final interpreter of

Scripture, and on Scripture as the definitive revelation of God's will. Scripture was said to be perspicuous to reason and Socinians attacked the authority of traditions, church authority and creeds. By the mid-seventeenth century Joachim Stegmann was arguing that nothing in the Bible exceeded human understanding. Many in the later generations of Socinianism suggested that contemporary orthodoxy was overly influenced by the Platonic and Neoplatonic philosophy of the early church, depicting this influence as buttressing belief not just in the Trinity but also in the natural immortality of the soul. Elevating reason against church authority, and personally attacked as 'heretics' for their beliefs, Socinians supported religious toleration and argued that the state was dedicated to purely natural ends. Many were pacifists.

5 'Socinianism' as an accusation and John Locke

The polemic accusation of Socinianism was regularly made in the seventeenth century against many thinkers, including Grotius, Chillingworth and Locke. Grotius' *De Veritate Religionis Christianae* omitted original sin and the atonement, but he published elsewhere in defence of the satisfaction and he believed in the resurrection of the same bodies (see Grotius, H.). William Chillingworth and the Latitudinarians emphasized the limited fundamentals of Christianity, the role of reason, and the centrality of morality in Christianity, but explicitly endorsed the Trinity, the atonement, and the immortality of the soul (see Chillingworth, W.; Latitudinarianism). The most important philosopher accused of Socinianism was John Locke. A series of writers declared his *Reasonableness of Christianity* 'all over Socinianized'. Leibniz described Locke as 'inclined' to Socinianism for his recognition of the possibility that matter might think, because this rendered it impossible, in Leibniz' view, to prove the soul's natural immortality. Bishop Edward Stillingfleet argued that Locke's epistemology 'favoured' Socinianism for his arguments about the limitation of our knowledge about substance, which undercut Trinitarian explications of the consubstantial Godhead.

Locke's library included many Socinian works, some of which he read carefully. His *Reasonableness of Christianity* was close to Socinianism in a number of ways. It interpreted Adam's Fall as not involving a 'corruption of human nature in his posterity'; discussed Christ's mission as essentially prophetic and especially kingly rather than priestly; avoided the satisfaction; emphasized Christ's promise of and testimony to immortality; and asserted a credal minimalism of belief in Jesus as the Christ without any mention of the Trinity. Yet he never directly attacked the Trinity or satisfaction, and his comments were compatible with a non-essential Trinitarianism or with kinds of Arianism, and with belief in a satisfaction. Locke's *Paraphrase* included one probably anti-Socinian textual explication. His arguments in the *Essay* may have been constructed in part to undercut Trinitarians' extended explanations of the Trinity and, by its final edition, to allow intellectual space for mortalism. These arguments in the *Essay* were, however, compatible with Socinianism rather than definitively Socinian. They could be combined with anti-Socinian views and were epistemological rather than ontological in orientation as observations about the limitations of human understandings of substance, matter, thought and God's power (see Locke, J. §7).

See also: Deism; Natural law

JOHN MARSHALL

References and further reading

Bibliotheca Fratrum Polonorum (The Library of the Polish Brethren) (1688), Amsterdam.(Collection of all major Socinian works in 8 volumes; the first 2 volumes are of works by Faustus Socinus.)

Fix, A. (1991) *Prophecy and Reason*, Princeton, NJ: Princeton University Press.(Chapter 6 provides a very informative introduction to Socinianism in the Netherlands.)

Grotius, H. (1622) *De Veritate Religionis Christianae (The truth of the Christian religion)*, London, 1624.(An eirenic work which avoided disputed dogmatic issues and was accused of Socinianism.)

Jolley, N. (1984) *Leibniz and Locke*, Oxford: Oxford University Press.(Chapters 2 and 3 provide an excellent account of Leibniz's accusation of Socinianism against Locke and its contexts.)

Locke, J. (1689) *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Clarendon Edition, Oxford: Oxford University Press, 1975.(Locke's magnum opus, accused of inclining to Socinianism.)

Locke, J. (1695) *The Reasonableness of Christianity*, London; 2nd edn, 1696; ed. J. Higgins-Biddle, Clarendon Edition, Oxford: Oxford University Press, 1998.(A work arguing that the sole belief necessary to become a Christian was that Jesus was the Christ, emphasizing morality, and stressing Christ's role as king and not as

priest; accused of Socinianism.)

Locke, J. (1700-4) *A Paraphrase and Notes on the Epistles of St Paul*, London, 6 vols, 1705-7; ed. A.W. Wainwright, Clarendon Edition, Oxford: Oxford University Press, 2 vols, 1987.(A work of paraphrase and exegesis of Paul's epistles; silent on the Trinity but one textual explication seems anti-Socinian on Christ.)

McLachlan, H. (1951) *Socinianism in Seventeenth Century England*, London: Oxford University Press.(Broad account of the influence of Socinianism in England.)

Katechizm zbory tych ludzi, ktory w Krolestwie Polskim… (1605), Racov; trans. as *The Racovian Catechism*, London: Thomas Rees, 1818.(The first English translation appeared in 1652; the closest work to a standard account of Socinian beliefs.)

Stegmann, J. (1644) *De iudice et norma controversarum fidei (Of judgment and standards in controversies of faith)*, Amsterdam.(A work indicating that nothing in the bible exceeded human understanding.)

Sullivan, R. (1982) *John Toland and the Deist Controversy*, Harvard, MA: Harvard University Press.(Chapter 3 provides a useful account of the varieties of Socinianism in late seventeenth-century England.)

Wallace, R. (1850) *Anti-Trinitarian Biography*, London: Whitfield, 3 vols.(Excellent survey of Socinianism and of individual Socinian thinkers.)

Wilbur, E.A. (1946-52) *A History of Unitarianism*, Cambridge, MA: Harvard University Press, 2 vols.(The standard account of unitarianism.)

Williams, G.H. (1980) *The Polish Brethren*, Missoula, MT: Scholars Press, 2 vols.(Collection of Socinian documents.)

Wootton, D. (1989) 'John Locke: Socinian or Natural Law Theorist?', in J. Crimmins (ed.) *Religion, Secularization and Political Thought*, London: Routledge.(A challenging reading of Locke's thought in the light of Socinianism.)

Sociobiology

Following Darwin, biologists and social scientists have periodically been drawn to the theory of natural selection as the source of explanatory insights about human behaviour and social institutions. The combination of Mendelian genetics and Darwinian theory, which did so much to substantiate the theory of evolution in the life sciences, however, has made recurrent adoption of a biological approach to the social sciences controversial. Excesses and errors in social Darwinism, eugenics and mental testing have repeatedly exposed evolutionary approaches in the human sciences to criticism.

Sociobiology is the version of Darwinism in social and behavioural science that became prominent in the last quarter of the twentieth century. Philosophical problems of sociobiology include challenges to the explanatory relevance of Darwinian theory for human behaviour and social institutions, controversies about whether natural selection operates at levels of organization above or below the individual, questions about the meaning of the nature-nurture distinction, and disputes about Darwinism's implications for moral philosophy.

1 Altruism

Sociobiology is a term coined by E.O. Wilson to name that discipline which applies the synthetic theory of evolution - the conjunction of Darwin's claim that nature selects over random variations and the Mendelian theory of genetic inheritance (see Evolution, theory of §1; Darwin, C.R.; Genetics §1) - to explain aspects of and differences in individual behaviour and the behaviour of groups of individuals. Sociobiology seeks to explain behaviour by showing that its emergence is to be expected as a strategy adapted to secure the fitness of the individuals that manifest the behaviour and their genes.

Sociobiology's promise, its explanatory strategy and the controversies surrounding it are aptly illustrated in the discussion of human altruism. Wilson (1976: 3) identified altruism as the 'central' theoretical problem of sociobiology: 'how can altruism, which by definition reduces personal fitness, possibly evolve by natural selection?' Altruism, and cooperation in general, are obvious features of human behaviour. Indeed, sociality requires them. But Wilson (1976: 578) notes that altruism is 'self-destructive behaviour performed for the benefit of others'. Therefore altruism reduces fitness. Accordingly, through evolution it should have been expunged not enhanced. So a characteristic feature of human behaviour appears to be not just exempt from the constraint of natural selection, but perhaps inconsistent with it. Once sociobiologists began to show how natural selection might in fact explain the emergence of altruism as a fitness-maximizing strategy, both its theoretical promise and its interest to philosophers grew rapidly.

The problem of cooperation among humans is one Darwin broached in *The Descent of Man* (1871). His answer to the question of how other-regarding behaviour could evolve made an appeal to the group as the unit of selection. Groups in which members cooperated or otherwise behaved altruistically, are likely in the long run to do better in meeting their members' needs than will groups of egoists. Thus, to the extent that behaviour can be controlled or at least encouraged by individual genetic inheritance, in the struggle for survival between groups, groups of altruists will be selected for. The trouble with this argument is that a group of altruists is vulnerable to 'invasion' by an egoist. Suppose mutation turns a single member of a group of altruists into an egoist who free-rides on the cooperation of others, but does not reciprocate. Note that if altruistic behaviour is genetically fixed, altruists will not retaliate against egoists. The egoist of course does better than the altruists, and amasses more resources. Thus its fitness increases, and it will have more and/or stronger offspring than will the average among altruists. Result: over the long run, the proportion of egoistic descendants will increase, since egoism is a genetically coded mutation that is transmitted down the original mutant's line of descent. After enough generations, there are nothing but egoists left in the original altruistic group.

This argument from individual selection's power to overwhelm group selection had two important consequences: it reinforced evolutionary biologists' commitment to methodological individualism, and it made Darwinism seem irreconcilable with the basic social disposition of altruism. It was W.D. Hamilton (1964), who realized how evolutionary theory and genetics could accommodate altruism, and thus gave sociobiology its late-twentieth-century impetus. Hamilton argued that nature will select for that strategy which leaves the largest number of copies of the gene that codes for it. In the case of sexually reproducing organisms this will usually be

direct offspring - sons and daughters. Whence the altruism of parents to offspring. Moreover, a strategy that sacrifices a son or daughter to save two siblings, three nephews or nieces, or a parent will ensure the survival of just as many or more copies of the individual's genes as would a strategy of saving a son or daughter. Accordingly, nature will select for 'inclusive fitness' which reflects the organisms' fitness and the fitness of each of its kin - the other organisms with whom it shares copies of the same genes, corrected by some coefficient of relatedness. If nature engages in 'kin selection' by selecting for 'inclusive fitness', then, altruism may emerge as an adaptive strategy for an individual that is part of a group of kin.

A problem arises immediately: why does an altruist provide resources at its own expense to unrelated organisms? This question is crucial for sociobiology. As human society evolves it is characterized more and more by cooperation, and less and less by kinship. So, again, either features of human society cannot be accounted for on the basis of selection for the fittest individual strategies, or else sociobiology needs to find another explanation for the emergence of altruism than its inclusive fitness. Sociobiological solutions to this problem have increased its interest for philosophers.

We may hypothesize that altruism emerges as a reciprocal strategy, in which individuals are cooperative in order to secure cooperation from others on later occasions. 'Reciprocal altruism' (the term was introduced by Richard Trivers (1971)) is, however, vulnerable to disappointment in the same way that unconditional altruism is. For the optimal strategy for a fitness maximizer is to accept altruistic pay-offs, but to decline to make them. Accordingly, altruism in the expectation of reciprocation will be displaced by egoism. Trivers and others had noticed that among animals, including humans, the problem of reciprocal altruism reflects what economists have called a 'prisoner's dilemma'. In this setting, the pay-off to free-riding, and the cost of being taken for a sucker, exclude altruism as an available strategy to the rational agent maximizing utility. To the extent that nature selects for maximizing fitness, it will discourage altruism and encourage free-riding whenever the pay-offs in nature are like those in the prisoner's dilemma. (see Decision and game theory §§3, 6). The prisoner's dilemma arises whenever there is a chance to be taken for a sucker combined with an opportunity to free-ride: possible cases include signalling the presence of predators or silently fleeing instead, sharing food or hogging it, encroaching on unprotected territory versus respecting it, procreating without remaining to raise the off-spring. In short, opportunities for altruism by one agent are opportunities for free-riding by another, and it appears that the costs and benefits make the latter an adaptational strategy for individuals and the former a maladaptive one.

However, the prisoner's dilemma is a situation animals face over and over again with a relatively small number of other animals. It is in the search for an optimal strategy for repeated or 'iterated' prisoner's dilemma situations that Hamilton and Axelrod (Axelrod 1984) found a potential solution to Wilson's problem of how altruism emerged. They argued that, under certain circumstances, in iterated prisoner's dilemma encounters, the optimal strategy for the individual is one called 'tit-for-tat': cooperate in game one, and then in each subsequent round do what the other player did in the previous round. Thus, if the other player tries to free-ride in round one, the best response is to refuse to cooperate in round two. If in round two the other party changes strategies and cooperates, in round three one should return to cooperating as well. In the long run no strategy generates a higher pay-off than tit-for-tat. Over the long haul, nature will select fitness maximizers employing this strategy, and as a result cooperation or reciprocal altruism will become fixed in populations. Whence altruism is shown to be compatible with natural selection, and may even have emerged through its operation.

This approach to explaining human behaviour raises issues of philosophical interest: the foundations of rational choice theory, the nature of social conventions, the analysis of teleology.

2 Anthropomorphic and adaptationalist fallacies

Sociobiologists have appeal to fitness-maximizing strategies to explain behaviour like altruism, promiscuity, xenophobia, homosexuality; to explain institutions like the incest taboos, preferential marriage rules, sex role differences, social inequality. Each of these is hypothesized to be the result of natural selection by environmental factors for individual fitness maximization. However, it has seemed absurd to opponents of sociobiology that behaviour which is by definition the result of intentions - desires and beliefs of rather complex sorts - could be intelligibly attributable to nonhumans as far down the phylogenetic ladder as birds and fish. Relatedly, it seems equally absurd to suppose that genes could be causally implicated in the occasions on which such intentions are acted on by humans. Both of these assumptions have been attributed to sociobiology and subjected to

philosophical scrutiny.

The argument runs: ‘altruism’ is defined as action that advantages another with the motive of doing so. It is a concept which carries a definitional commitment to an intentional cause. But this motive forms no part of the sociobiologist’s stipulative definition of fitness, as Wilson, for example, gives it. The term is then employed to explain both human and nonhuman non-intentional dispositions, and is applied to explain human intentional altruism by a gratuitous anthropomorphism of nonhuman dispositions.

Arguments like this are too strong. Few sociobiologists suppose that the occurrences of intentional actions are explainable by natural selection. Their interest in any case is in explaining the emergence of dispositions to behave in certain ways under certain conditions. Such explanations are neutral about the immediately prior causes that trigger the exercise of the dispositions explained. Indeed, if selection works on behaviour as it does on structure, it is likely that it will select a variety of different causes with similar effects, just because these effects have the same selective advantage. In human altruism these causes may include intentional states, while in nonhuman ones it may involve some much simpler mechanism. Sociobiology aims to explain a genus - the disposition to altruism, whether intentional or not, of which the former is a species. Similarly, nature will select not just individual genes, but packages of them, and in fact competing packages of differing genes so long as these have the same or equally adaptive effects in human and infrahuman dispositions. No one supposes that there is a single gene or a small number of genes that code for intentional behaviour of any sort, or even a large number of genes that do more than provide for the disposition to engage in intentional actions of an adaptive sort.

But why suppose that selection operates so pervasively as to make adaptational explanations appropriate? This is a more fundamental challenge to sociobiology. Like other applications of the theory of natural selection, sociobiology has been accused of an obsession with identifying every salient behaviour pattern or institution as the adaptational product of pervasive selection. This commitment to adaptationalism is alleged to reflect an unfalsifiable assumption that deprives sociobiology of its scientific status (see Demarcation problem; Scientific method). Many sociobiologists along with most evolutionary theorists will plead guilty to the charge of adaptationalism, but defend the practice as one demanded by the theory of natural selection. The adaptational imperative of the theory is no different from principles governing hypothesis selection in any other theory. It can be misused to defend the theory come what may, but it can hardly be eschewed altogether in a theory whose aim is to apply Darwinism to the human sciences.

3 The levels of selection

Among evolutionary biologists the notion that species, populations or interbreeding groups could be selected for has been unpopular for reasons sketched out above: such groups are liable to ‘invasion’ by individuals who can take advantage of group traits while not contributing to their maintenance, and group selection is always swamped by individual selection. Some sociobiologists have gone further and argued that not even the individual organism is properly viewed as the level on which selection operates. As G.C. Williams has held, considerations of theoretical parsimony require that adaptational significance should not be accorded to a trait at any level above that necessary to explain the persistence of the trait in question. Following this principle, it has been argued that the locus of selection is always the individual gene - so that if altruism among animals is to be given an evolutionary explanation it must be shown how it can be adaptive for the individual gene borne by the individual altruist (see Genetics §5).

This controversy has drawn philosophers concerned with reductionism into active debates in sociobiology. Genic reductionists argue that attributing purposes or functions in biology is always dangerous, and the risks of anthropomorphism are minimized by restricting their attribution to the lowest level of organization compatible with the evidence: this level is argued to be the individual gene or allele. Holists have argued that some changes at the level of populations cannot be explained by appeal to properties of the individuals or genes themselves.

For instance, given a certain proportion of altruists and egoists in a population, whether the population expands, and how fast are functions of the distribution of altruists and egoists in the subpopulations out of which the whole group is composed. If the structure of the group includes subpopulations in which altruism is heavily represented, then the population as a whole may increase faster than otherwise. But population structure is a group trait, which can be selected for independently of the selection for traits of individual organisms or their particular genes.

Philosophical scrutiny of the claim that the gene is the unique locus of selection has brought sociobiology into contact with fundamental problems in the analysis of probabilistic causation (see Causation §4). Suppose a gene's having a property raises its chances of leaving copies in some contexts but not in others. Then the question arises of how to describe the overall causal power of the trait in all contexts. When we add in the fact that in a structured population of genes, the trait may have probabilistic effects different from those it has when identically but randomly distributed in an unstructured population, the problem of probabilistic causality makes contact with the problem of determining the level of selection.

4 Nature versus nurture

The rise of sociobiology renewed interest in the substantive questions of the relative causal importance of heredity and environment in the distribution of traits. Almost from the first, sociobiology was stigmatized as guilty of 'biological determinism' - the thesis that traits of interest are not subject to human modification because they are genetically fixed. To some extent this charge betrays a fundamental misunderstanding of the synthetic theory of evolution. For Darwinian theory holds that environmental factors select from among hereditary variations; and genetic theory holds that traits - phenotypes - are the product of genetic and environmental factors. Genes alone can fix nothing.

Nevertheless the question remains of how to apportion causal responsibility between environmental and genetic factors. Biologists have done so by linking causality to experimental methods. In an experiment where one can control both environmental variables and the presence or absence of a gene, differences among outcomes enable one to identify the effects of each, and sometimes even apportion the causal responsibility between the environment and the gene. If the presence of a gene has a greater impact on reproductive success than the presence of an environmental factor, then there is some reason to credit it with a predominantly causal role, even though the environmental factor was also causally necessary for any reproduction.

This operational definition of greater or lesser causal responsibility makes it clear that such judgements are always relativized to sets of genes and sets of environmental variables. To say that a trait is mainly under the control of genes will be to make implicit appeal to a range of environmental variables whose presence or absence has lesser effects on the trait's appearance. In other environments, the presence or absence of the same gene may have lesser impact. This difference has led some writers to condemn claims that some traits are innate genetically, or biologically determined, as false or unfounded. For there may always be crucial environmental factors we can almost never vary whose influence on the presence or absence of the trait will be greater than that of the alleged genetic factors.

5 Ideological and moral implications

Some sociobiologists have argued that a Darwinian approach to the evolution of human behaviour can explain the existence of *mores* and moral codes, and the differences among them as the consequence of selection for fitness maximizing over variation in human and prehuman behavioural dispositions. Some writers have gone further (following Herbert Spencer) and have suggested that Darwinian theory might even underwrite or justify some normative claims (see Evolution and ethics).

Few philosophers have expressed sympathy with the latter view, noting that it is a discredited form of naturalism (see Naturalism in ethics §1). Moreover, it has been noted that a theory which purports to explain the moral codes and the social mores which characterize different societies as the result of natural selection can easily become rationalizations for these codes and *mores*, can effectively serve as an intellectual obstacle to changes even when the morals and *mores* are in fact not resistant to fundamental changes in them. Scientists' development of evolutionary explanations for human inequalities and differences can and have been misused to argue that these inequalities are biologically determined, and so cannot be ameliorated or eradicated by environmental intervention. As such, sociobiology can serve an unintended ideological function independent of its cognitive status. Philosophers with differing ideological convictions have sought to uncover the improper ideological exploitation of sociobiological theory.

See also: Behaviourism in the social sciences; Evolutionary theory and social science; Human nature; Methodological individualism; Reduction, problems of; Species; Unity of science

ALEX ROSENBERG

References and further reading

Alexander, R. (1979) *Darwinism and Human Affairs*, Seattle, WA: University of Washington Press.(Develops an account of social institutions as reflecting genetic and environmental factors.)

Axelrod, R. (1984) *The Evolution of Cooperation*, Ann Arbor, MI: University of Michigan Press.(An accessible account of the problem of altruism and the role of game theory, and computer modelling in sociobiology. Includes original Hamilton and Axelrod paper.)

Darwin, C. (1859) *On the Origin of Species by means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, London: John Murray; repr. Cambridge, MA: Harvard University Press, 1964. (Darwin's contribution to the discussion of group selection and human cooperation.)

Darwin, C. (1871) *The Descent of Man and Selection in Relation to Sex*, London: John Murray, 2 vols; repr. Princeton, NJ: Princeton University Press, 1981.(Darwin's main work on human origins, with a long account of sexual selection and an attempt to apply that theory to explain human characters.)

Dawkins, R. (1989) *The Selfish Gene*, Oxford: Oxford University Press, 2nd edn.(A strong defence of the gene as the unit of selection and of its explanatory relevance to human behaviour.)

Hamilton, W.D. (1964) 'The Genetic Evolution of Social Behaviour', *Journal of Theoretical Biology* Parts I and II, 7: 1-52.(Develops the theory of kin-selection later exploited by Axelrod.)

Kitcher, P. (1985) *Vaulting Ambition*, Cambridge, MA: MIT Press.(A sustained philosophical attack on sociobiology.)

Maynard Smith, J. (1982) *Evolution and the Theory of Games*, Cambridge: Cambridge University Press. (Introduction to biological exploitation of game theory.)

Richarson, P. and Boyd, R. (1985) *Culture and the Evolutionary Process*, Chicago, IL: University of Chicago Press.(A sophisticated treatment of the interaction of environment, heredity and learning in human social evolution.)

Sober, E. (1983) *Conceptual Issues in Evolutionary Biology*, Cambridge, MA: MIT Press; repr. 1994.(An anthology of important papers in the philosophy of biology touching on many of the issues broached above. The 2nd edition of 1994 contains further papers.)

Sober, E. (1993) *The Nature of Selection*, Chicago, IL: University of Chicago Press.(A discussion of the controversy surrounding the units of selection and the nature of evolutionary theory.)

Trivers, R. (1971) 'The Evolution of Reciprocal Altruism', *Quarterly Review of Biology* 46: 35-57.(First development of theory of reciprocal altruism.)

Williams, G.C. (1966) *Adaptation and Natural Selection*, Oxford: Oxford University Press.(Opposes attributions of selection at levels above the gene, subject to close philosophical examination by Sober and others.)

Wilson, E.O. (1976) *Sociobiology: The New Synthesis*, Cambridge, MA: Harvard University Press.(Introduction to sociobiology among both humans and infrahuman species.)

Wilson, E.O. (1978) *On Human Nature*, Cambridge, MA: Harvard University Press.(Defence and extension of theories of *Sociobiology: The New Synthesis*.)

Sociology of knowledge

Sociologists of knowledge contribute to the enterprise of generating a naturalistic account of knowledge by describing and explaining the observed characteristics of shared cultures. They assume that knowledge can be treated as an object of empirical investigation (rather than mere celebration or condemnation). Because science is understandably taken as our best example of knowledge, the sociology of scientific knowledge plays a pivotal role in the field. It is argued that our natural reasoning capacities, and our sense experience, are necessary but not sufficient conditions for scientific knowledge. Sociologists looking for the causes of its content and style focus on the contribution of conventions and institutions.

The aim of sociologists of knowledge, is to identify such features as (1) the general character of the processes by which new cultural members are ‘socialized’, that is, trained and educated; (2) the specific institutions and authorities charged with this task in particular cases; (3) the mechanisms by which a body of culture is kept relatively stable and hence available for use; (4) the precise circumstances and purposes associated with its employment on particular occasions; (5) the processes by which change is managed and its locus and extent negotiated; (6) the distribution of taken-for-granted beliefs according to status and membership criteria, for example, professional or amateur, male or female, doctor and patient, scientist or technician.

Sociologists emphasize that accepted systems of belief cannot be adequately accounted for by the reasonableness of our individual cognitive faculties, or our sensory experience of the environment. Reason and experience are vital parts of the story of knowledge (social learning and cultural transmission would be unintelligible without them), but they are necessary, not sufficient, conditions. There is always the need to introduce further elements such as the traditional background of practices on which experience impinges, and the highly consequential fact that individuals must make allowances in their calculations for the calculations of others. Individual reason operates in a social context. Furthermore, our individual conclusions and reactions must be coordinated and collectively assessed before they count as knowledge rather than subjective belief, idiosyncrasy or error.

These empirical features are conspicuously visible in scientific culture, hence the central role played by the sociology of scientific knowledge within the general field. This also explains why, wittingly or unwittingly, philosophers of science have done much of the theoretical groundwork for the discipline. In particular they have emphasized that scientific conclusions are under-determined by observational or experimental input, and by the canons of formal logic or the powers of individual reason. Whether it be Duhem’s insight that hypotheses cannot meet the test of experience in isolation, Popper’s insistence on the conjectural character of all knowledge claims, Carnap’s discovery of an infinity of inductive strategies, Kuhn’s highlighting the role of dogmatically accepted paradigms, or Hesse’s (1974) development of a Bayesian approach with its dependence on a negotiable network of shared classifications and prior probabilities, the conclusions all point in the same direction. Scientific knowledge - our best example of knowledge - depends for its structure, stability, content and dynamics on phenomena best identified, and investigated, as social institutions (see Duhem, P.M.M. §4; Carnap, R.; Popper, K.R.; Kuhn, T.S. §2).

Not all philosophical work is relevant in this way. Some gives every appearance of being deliberately fashioned to inhibit naturalistic inquiry. There are philosophers who effectively define ‘rationality’ in terms of its supposed opposition to causality, and particularly social causality. They assume social considerations can only illuminate the general conditions encouraging or discouraging the progress of knowledge, and explain distortions, or deviations from the demands of the scientific method. The task of the sociologist is then confined to explaining, say, the deleterious effect of totalitarian state interference with the freedom of science. This limited agenda may be called the ‘weak programme’ in the sociology of knowledge, in contrast to the ‘strong programme’ suggested by under-determination, which seeks the social dimension to be found in all bodies of knowledge however they are currently evaluated.

Some of the best work in the sociology of knowledge, concentrating on the hardest cases, has been done by historians of science. Ignoring the proscriptions of philosophers who would restrict them to the weak programme, they have provided an abundance of high-quality studies, readable as contributions to the strong programme. These exhibit the problematic, contextual character of the interpretation of experimental findings and, in many cases,

press the argument further and identify plausible causes for the preferences and opinions of the actors involved, for example, contingent commitments to a theory or research method, or the utility the result had for a range of goals and purposes, some of them narrowly professional, others broader and even overtly political. These studies often identify an interest at work. It is important to realize that interests can be generated internally, within the workings of the social system of science itself. The sociology of knowledge carries no prior commitment to what is sometimes called 'externalism', rather than 'internalism', in its methods. The richness of this empirical material precludes summary here, but it spans early Royal Society work on the air pump, the attractions of the corpuscular philosophy, nineteenth-century anatomy and evolutionary theory, cellular biology, geological controversies, non-Euclidean geometry, astronomical discoveries, statistical techniques, genetic theory, and modern particle physics.

Sadly, many of the critical discussions of the sociology of knowledge are blighted by ill-informed and hostile stereotypes. These sustain a number of oft-repeated arguments that are widely taken to discredit the enterprise. For example, the sociology of knowledge is said to be 'self-refuting'. The charge begs the question, because it simply takes for granted the premises of the weak programme: that causal determination equals error or distortion. Again, a distinction is often drawn between the 'context of discovery' and the 'context of justification'. Is not the sociologist concerned with the origin of beliefs, and is that not irrelevant to their epistemological status? In fact, sociologists have little to offer on the origin of ideas, but much to say about their evaluation and subsequent elaboration; so the criticism gets things the wrong way round. Nor, as is often alleged, need the sociology of knowledge derive from, or issue in, a negative stance towards the beliefs under study, as if the only purpose were to 'unmask'. To exhibit, the social construction of science is no more to criticize it than studying the physiology of the eye is to criticize vision.

See also: Feminist epistemology; Foucault, M.; Naturalism in social science

DAVID BLOOR

References and further reading

Barnes, B. (1982) *T.S. Kuhn and Social Science*, London: Macmillan.(Provides an excellent sociological reading of Kuhn, and brings out the connection with Hesse's work.)

Bloor, D. (1991) *Knowledge and Social Imagery*, Chicago, IL: University of Chicago Press.(A defence of the strong programme with a discussion of some of the main philosophical objections.)

Brown, J.R. (1984) *Scientific Rationality: The Sociological Turn*, Dordrecht: Reidel. (A selection of papers both by those for and those against the sociology of knowledge in its strong form.)

Collins, H. (1992) *Changing Order*, Chicago, IL: University of Chicago Press. (Valuable study centring on the dispute over gravity wave detection and the problematic character of replication.)

Desmond, A. (1989) *The Politics of Evolution*, Chicago, IL: University of Chicago Press.(Fascinating analysis of the social basis of the scientific elite's fears about 'materialistic' and (pre-Darwinian) evolutionary ideas.)

Fleck, L. (1935) *Genesis and Development of a Scientific Fact*, Chicago, IL: University of Chicago Press, 1979. (Outstanding and pioneering work in the sociology of knowledge. It was mentioned in passing by Reichenbach, and then forgotten until it acted as a source of inspiration for Kuhn.)

Harwood, J. (1994) *Styles of Scientific Thought*, Chicago, IL: University of Chicago Press.(A study of the German genetics community between 1900 and 1933, describes and explains the stylistic and substantial differences between German and US genetics, and differences within the German community.)

Hesse, M. (1974) *The Structure of Scientific Inference*, London: Macmillan. (Sophisticated Bayesian model of knowledge, particularly useful to the sociologist because of its 'network' model of a classificatory system.)

Hollis, M. and Lukes, S. (1982) *Rationality and Relativism*, Oxford: Blackwell. (Arguments for and against a 'relativist' account of knowledge of the kind associated with the sociology of knowledge.)

Pickering, A. (1984) *Constructing Quarks*, Edinburgh: Edinburgh University Press. (A detailed history of elementary particle physics.)

Shapin, S. (1982) 'History of Science and its Sociological Reconstructions', *History of Science* 20: 157-211. (Though now dated, it still provides the best introduction to the literature on the sociology of knowledge. Contains a bibliography of 149 items.)

Shapin, S. and Schaffer, S. (1985) *Leviathan and the Air-Pump*, Princeton, NJ: Princeton University Press.(An account of the disputes between Hobbes and Boyle over the interpretation of the air-pump experiments, emphasizing the unity of the problem of natural order and social order.)

Sociology, theories of

Throughout the history of sociology, three types of theorizing have co-existed, sometimes uneasily. 'Theories of' provide abstract models of empirical processes; they function both as guides for sociological research and as sources for covering laws whose falsification or validation is intended to provide the basis for a cumulative science. 'Presuppositional studies' abstract away from particular empirical processes, seeking instead to articulate the fundamental properties of social action and order; meta-methodological warrants for the scientific investigation of societies; and normative foundations for moral evaluations of contemporary social life. 'Hermeneutical theory' addresses these basic sociological questions more indirectly, by interpreting the meanings and intentions of classical texts.

The relation between these three forms of theorizing varies historically. In the post-war period, under the institutional and intellectual influence of US sociologists like Parsons and Merton, presuppositional and hermeneutical issues seemed to be settled; 'theories of' proliferated and prospects seemed bright for a cumulative, theoretically-organized science of society. Subsequent social and intellectual developments undermined this brief period of relative consensus. In the midst of the crises of the 1960s and 1970s, presuppositional and hermeneutical studies gained much greater importance, and became increasingly disarticulated from empirical 'theories of'. Confronting the prospect of growing fragmentation, in the late 1970s and early 1980s there appeared a series of ambitious, synthetical works that sought to reground the discipline by providing coherent examples of how the different forms of sociological theory could once again be intertwined. While widely read inside and outside the discipline, these efforts failed in their foundational ambitions.

As a result of this failure, over the last decade sociological theory has had diminishing influence both inside the discipline and without. Inside social science, economic and anthropological theories have been much more influential. In the broader intellectual arena, the most important presuppositional and hermeneutical debates have occurred in philosophy and literary studies. Sociological theorists are now participating in these extra-disciplinary debates even as they have returned to the task of developing 'theories of' particular institutional domains. The future of specifically sociological theory depends on reviving coherent relationships between these different theoretical domains.

1 The forms of sociological theory

Theories of. Following what is taken to be the natural science approach, theorists have produced sets of highly general causal or descriptive propositions aimed at modelling empirical processes (Stinchcombe 1968). These models have had a dual purpose. On the one hand, they are guides to more specific empirical studies of particular social processes. On the other hand, they are designed to produce putative covering laws which such particular empirical studies will falsify or validate.

Max Weber's famous essay on bureaucracy can be viewed as providing a classic example of theorizing along these lines. Weber presented a series of propositions about the structure and processes of bureaucratic authority that has provided the central reference for empirical studies, complementary elaborations, and competing general models in the sociological study of organizations over the last fifty years. In an effort to test the model's insistence on rationality, formal rules and hierarchical command, for example studies emerged that emphasized informal organization and decentralized, problem-oriented decision making; the limited, primarily short-term rationality of bureaucratic action; and even the ceremonial, merely ritualistic qualities of bureaucratic processes. Yet, because Weber's original essay has been accepted as an empirically grounded 'theory of', the field of organizational studies has organized and conceptualized itself as being guided by an accumulative, increasingly elaborated, essentially progressive theoretical model.

While few theoretical models have exercised comparable hegemony, in many empirical sub-fields of sociology one can find 'theories of' that have played similar roles for more limited periods of time.

Presuppositional studies. Rather than developing empirical generalizations to guide studies or provide covering laws, another kind of theoretical effort generalizes away from particular empirical realms to consider fundamental issues that are held to undergird them (Alexander 1982).

Some of these theoretical efforts discuss issues like action and order: is social action practical and goal-oriented or is it normatively guided and oriented by aesthetic, emotional, or moral concerns? Does the patterned nature of social activity derive from controls exercised over individuals by institutions, whether coercive or moral, or does it emerge from pragmatic negotiations between actors as they spontaneously confront the unpredictable contingencies of everyday life?

Some presuppositional studies address more meta-methodological issues: should sociology be modelled after the nomological sciences that aim at producing covering laws and abstract away from the particular and idiographic, or should it be more hermeneutical and aim at producing rich ethnographic interpretations of particular institutions and events?

Other presuppositional studies take the form of critical ideological and moral inquiry: is capitalism an oppressive social system? Is modernity healthy? What are the social obligations of scientists when they are faced with injustice or abuse?

If we consider the relationship of presuppositional studies to the earlier example of bureaucracy, we find an illuminating point of contrast with the ‘theories of’ approach. Rather than arguing on the basis of empirical studies that informal ties exist alongside formal rules, presuppositional theorists have argued that, because emotion is central to action and cooperation basic to order, bureaucratic institutions simply cannot be conceived or explained in the way Weber proposed. Arguing that Weber was, in fact, generalizing from only distinctively Prussian authority structures rather than from bureaucratic organizations as such, other presuppositional theorists have pushed for a case-study approach to organizations in different countries and different historical milieu. Others, focusing on the moral dimension, have condemned both the idea and the practice of bureaucratic organization for ignoring the possibility of democratic self-regulation.

Hermeneutical. ‘Theories of’ and presuppositional studies are highly abstract enterprises, each broaching the question ‘what is society?’ in a direct way. By contrast, the third form of sociological theorizing is not abstracting but idiographic and concrete. Hermeneutical theory seeks to explain human society by explicating, through interpretive argument, the meaning of sociological texts whose consequentiality has given them a classical status (Levine 1995).

These textual efforts are neither historical nor literary. Rather than being primarily directed to questions of intellectual and social context or to questions of genre and aesthetic form, they engage the hermeneutical method to investigate the traditional questions of sociology, albeit in an indirect and idiographic rather than direct and abstracting manner. By reinterpreting canonical texts, hermeneutical readings challenge extant theories and empirical studies that derive their authority from the covering laws or the presuppositions espoused by these classics. By producing new and authoritative versions of canonical texts, the hermeneutical ambition is to re-orient theoretical and empirical practice.

Once again, Weber’s organizational studies provide an illustration. Arguing that Parsons’ translation of Weber’s term, *Herrschaft*, as ‘imperative coordination’ implied that Weber intended a consensual rather than coercive model of authority, hermeneutical theorists have suggested that ‘domination’ is closer to the meaning Weber intended (Roth 1968). This translation challenge has, in fact, been part of the broader theoretical debate between functionalist and conflict-oriented ‘theories of’ various institutional realms (see Functionalism in social science). Within the specific field of organizational studies, hermeneutical students of Weber’s corpus have argued that he envisaged his essay on bureaucracy primarily as a historical and comparative explanation of the development of Western authority structures over time rather than as an effort to model the processes involved in contemporary organizations as such. Making connections to the emphases on charisma and morality in Weber’s other historical studies, these hermeneutical theorists have, indirectly, tried to reframe his original bureaucracy model as one that can incorporate without contradiction the informal, emotional, and normative elements of organizational life.

Very similar kinds of hermeneutical arguments have been made in regard to the classical writings of Marx, Durkheim, and Parsons. Through translations and explications of his earlier writings, Marx has been reinterpreted as being more concerned with the qualitative, cultural, and psychological problems of capitalism than with the quantitative and economic. By exposing the increasing influence of religion in Durkheim’s later and posthumously published work, Durkheim has been reinterpreted as aiming to produce a symbolic rather than morphological

sociology. By emphasizing the disjuncture between the early and the later Parsons, hermeneutical theorists have indirectly argued against systems theory and in support of more pragmatic presuppositions about action and order (see Systems theory in social science §§1-2).

While these three forms of theoretical activity in sociology represent distinct and long-standing traditions, each of them has been the object of polemical challenges from champions of the other types. In practice, however, the three types have often interpenetrated one another. Thus, rather than taking as data only empirical studies, 'theories of' particular institutions have often justified their abstract models by making textual interpretations, challenging or elaborating classical works that earlier sought to explain the same institution. Presuppositional studies, for their part, are often deeply implicated in evaluative statements about the moral worth of different approaches to action and order. Similarly, ideological debates are replete with explanatory claims about the empirical causes and effects of particular institutions.

The classical example of such a blurred genre is Talcott Parsons' *The Structure of Social Action* (1937), perhaps the single most influential theoretical work of post-war sociology. Parsons presented the work primarily as a hermeneutical one, describing it as an investigation of the writings of a group of important European social thinkers, Pareto, Marshall, Durkheim, and Weber. In the lengthy first section of his work, however, Parsons developed a series of highly abstract presuppositional arguments; while ostensibly provided to facilitate his subsequent textual interpretation, this discussion became a fundamental, *sui generis* theoretical argument in itself. Throughout his subsequent exegetical discussions, Parsons also produced pointedly polemical 'theories of', using his textual interpretations to challenge existing empirical models of various institutions.

2 Sociological theory in the post-war world

While the three forms of sociological theorizing have, indeed, always interpenetrated in significant ways, over time there have been remarkable shifts in the relative importance of each genre in the theoretical field. These changes are caused not only by intellectual developments within sociology but also by developments outside of it, by shifts in the intellectual life of society and in the institutions of society itself. These shifts have had significant effects on the relation of sociological theory to theory in other social scientific disciplines and to theoretical efforts in philosophy and aesthetics.

Despite the fundamental and continuing role of theories produced by classical sociological theorists - Marx, Weber, Durkheim, Simmel and Mead - the story line of 'modern' sociological theory begins in 1937 with the publication of Parsons' *The Structure of Social Action*. Despite its intrinsic intellectual power, this blurred genre made virtually no impact at the time. Under the impact of the Great Depression and massive social movements on the Left and Right, US and European sociology was highly fragmented. No single institutional or intellectual centre dominated, and within each sociological community radically divergent theories struggled for recognition. In the wake of the world-historical realignment that followed the Second World War, the social and intellectual situation was fundamentally changed. While European theorizing retained its distinctive forms and practices, the social stability, institutional power, and ideological hegemony of the USA allowed it to emerge as the arbiter of sociology in its professionalized, disciplinary form.

In this new context, Parsons emerged as the dominant theorist of his time. A new generation of sociologists came to consider his (1937) work as having established the canonical foundations of a distinctively modern sociological theory. While this theory was termed 'structural-functionalism', it might more accurately be considered a version of Kantianism (Munch 1981). Rather than using biological models to establish an organicist model of society, Parsons actually devoted his efforts to describing the institutional, psychological, and cultural foundations of differentiated and pluralistic societies that could process economic, political, religious, and ethnic conflicts in democratic, fundamentally cooperative ways (Turner and Holton 1986). Within this Parsonian framework, during the 1950s and early 1960s 'theories of' dominated other types of theoretical practice, and sociology increasingly was viewed as a maturing and cumulative science of society. This optimistic scientism was encouraged by the impression that the post-war explosion of large scale quantitative empirical studies were conducted within the rubric of functionalist thought, an illusion encouraged by the institutionally-oriented, 'middle level' functionalist models of Parsons' discipline (Merton 1967).

Because of its high degree of internal consensus, its cumulative character, and its high degree of intellectual

creativity, in the post-war period US sociological theory achieved a high level of influence, both inside the discipline of sociology and outside. Historians, political scientists, and anthropologists, for example, drew heavily upon the premises of modernization theory, which was the historical foundation of Kantian functionalism. Literary critics drew upon sociological theory to investigate the origins of modern genres. While philosophy remained largely in the analytical mode, more historical and socially-oriented philosophical efforts also drew upon these sociological ideas.

The social and intellectual upheavals that rocked Western societies between the mid 1960s and the 1970s had the effect of displacing not only Kantian functionalism but the approach that emphasized 'theories of'. Pragmatic, phenomenological, and behaviouristic traditions re-emerged as 'micro-sociological theories', powerfully challenging Parsonian theory in various domains (Homans 1961; Garfinkel 1967). Conflict theories and Marxism became increasingly influential, constituting 'macro' challenges to functionalism's optimistic and relatively consensual social models (Collins 1975). As both the models and the canon established by Parsons were put to the test, sociological theorists became increasingly involved in presuppositional and hermeneutical studies (see Pragmatism; Behaviourism in the social sciences).

By the end of the 1970s, Parsons' challengers had triumphed. Sociology was proclaimed a 'multi-paradigmatic' field which, once again, was devoid of either an institutional or theoretical centre. Paradoxically, even as this dramatic sea change reached its climax, the historical context facilitating it dramatically changed. The waves of anti-Establishment protests subsided, backlash movements against them asserted increasing strength, and Western societies began to experience conservative retrenchments. During the same period, state-communist societies began to crumble under their own weight.

These shifts placed sociological theory in a precarious position, even when its prestige was at its greatest height. Responding to the de-centred situation of sociology and to the broader sense of imminent transition in the intellectual and social world at large, the most influential theoretical works that emerged at this time were marked by their refusal to endorse the notion of pluralistic and relativistic social science. To the contrary, these highly ambitious theoretical efforts stepped outside their respective traditions and created grand syntheses (Bourdieu 1977; Giddens 1984; Habermas 1984; Alexander 1982; Collins 1975). Despite marked differences, they had a common aim: to found a new era of intellectual cooperation and social progress in a post-positivistic frame. Indeed, with their combination of hermeneutical, presuppositional, and explanatory ambitions, these efforts closely resembled Parsons' earlier effort to overcome intellectual and social fragmentation in 1937.

But the post-war world that had allowed Parsons' earlier effort to succeed had been dramatically and fundamentally changed. Despite the fact that these ambitiously synthetical works were widely read and admired both inside and outside the discipline of sociology, they turned out to have brought the curtain down on an era rather than ushering in a new one. In a *fin-de-siècle* world torn between deeply pessimistic moods of fragmentation and decline and wildly optimistic hopes of utopian rebirth, none of these synthetic efforts to recreate a new version of Neo-Kantian sociological theory succeeded in gaining hegemony or even in creating traditions that could convincingly establish coherent new lines of theoretical and empirical work.

3 Sociological theory today

The failure of these synthetic efforts, and the continuing impact of the social and intellectual conditions that gave rise to them, has given rise to developments that have had significant repercussions for sociological theory.

Self-identified 'sociological theorists' have been increasingly less able to exert influence over the explanatory models and canonical texts inside the discipline of sociology. Instead, sociological theory is now increasingly viewed merely as one sub-field among many, one devoted to hermeneutical reconstructions of the origins of classical texts, on the one hand, and to 'meta-theoretical' commentaries and arguments about presuppositions, on the other (Ritzer 1992). With the exception of the work of Bourdieu, whose influence, while substantial, remains relatively limited, there exists at present no over-arching general theory that combines hermeneutical and presuppositional arguments with sociological 'theories of' various institutional spheres. While empirical sub-fields continue to be organized theoretically, and the three different types of sociological theorizing flourish within each, these discussions are hermetically cut off from the more generalized theoretical debates in the discipline at large.

Distinctively sociological theories have lost their influence *vis-à-vis* theorizing in other social scientific

disciplines. This is demonstrated by two contradictory developments. The model of society as based upon exchanges between rational actors, derived from economics, has become an international cross-disciplinary tradition that already exercises a deep influence in political science and, increasingly, in sociology (Coleman 1990). At the same time, cultural models of symbolic action and ritual order, which emerged in anthropology (Geertz 1973), have had far-reaching theoretical and empirical effects in history and sociology.

The most influential presuppositional and hermeneutical studies, and even some of the most important theoretical discussions about the nature of contemporary societies (that is, 'theories of'), now occur outside the social sciences in philosophy and literary studies. Under the influence of Michel Foucault and Jacques Derrida, deconstructionism and postmodernism have established an anti-universalist position that has had widespread repercussions throughout Western intellectual life, including the social sciences (Bauman 1995; Seidman 1991). In opposition to this position, Neo-Kantian theorizing about the possibilities for universalism and democracy has been rejuvenated under the influence of John Rawls and Jürgen Habermas (§1) (Cohen and Arato 1992). Between these two poles there have emerged arguments for new forms of pluralistic, identity-based democratic associations, arguments derived from Aristotle, Hegel, and US pragmatism and associated most significantly with the writings of Charles Taylor (§5), Richard Rorty, and Michael Walzer (Boltanski and Thevenot 1991; Warnke 1992).

4 Prospects

In considering the future development of sociological theory, it is important to observe two apparently contradictory developments within the work of those who participated in the creation of synthetic general theories in the 1980s. Leading sociological theorists are now participating in the broader philosophical, political, and literary debates outside of the discipline itself (Calhoun 1992). At the same time, they have also turned towards empirically-focused topics and are developing 'theories of' particular institutional spheres. Only if and when these very different kinds of practices can be brought back together (Mouzelis 1995) will sociological theory regain the coherence, vitality, and intellectual power that allowed it to exercise such wide influence in an earlier day.

JEFFREY C. ALEXANDER

References and further reading

Alexander, J. (1982) 'Positivism, presuppositions, and current controversies', in *Theoretical Logic in Sociology*, London: Routledge & Kegan Paul, vol. l.

Alexander, J. (1995) *Fin-de-Siècle Social Theory: Relativism, Reduction, and the Problem of Reason*, London: Verso.(Criticizes the alarming growth of relativism in recent social theory.)

Bauman, Z. (1995) *Life in Fragments: Essays in Postmodern Morality*, Oxford: Blackwell. (Argues that our new era may yet prove to represent a dawning, rather than a twilight, for ethics, emancipated from the false consciousness entailed by modernity.)

Boltanski, L. and Thevenot, L. (1991) *De la justification: Les economies de la grandeur*, Paris: Gallimard.

Bourdieu, P. (1972) *Esquisse d'une théorie de la pratique: précédée de trois études d'ethnologie kabyle*, Geneva: Droz; trans. R. Nice, *Outline of a Theory of Practice*, Cambridge: Cambridge University Press, 1977.(Bourdieu discusses the main outline of his social theory. Complex. The English version differs substantially from the original *Esquisse*.)

Calhoun, C. (ed.) (1992) *Habermas and the Public Sphere*, Cambridge, MA: MIT Press.(An important example of how contemporary discussions reach well beyond, but still include, sociological theory.)

Cohen, J.L. and Arato, A. (1992) *Civil Society and Political Theory*, Cambridge, MA: MIT Press.(Looks at how Neo-Kantian theorizing about the possibilities for universalism and democracy has been rejuvenated under the influence of Rawls and Habermas.)

Coleman, J. (1990) *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.

Collins, R. (1975) *Conflict Sociology*, New York: Academic Press.

Garfinkel, H. (1967) *Studies in Ethnomethodology*, Englewood Cliffs, NJ: Prentice Hall.

Geertz, C. (1973) *The Interpretation of Cultures*, New York: Basic Books.(Collection of essays forwarding a view of what culture is, what role it plays in social life and how it ought to be properly studied.)

Giddens, A. (1984) *The Constitution of Society*, London: Blackwell.

Habermas, J. (1981) *Theorie des kommunikativen Handelns*, Frankfurt: Suhrkamp; trans. T. McCarthy, *The Theory of Communicative Action*, Boston: Beacon Press, 1984/1987.(His two-volume *magnum opus*, a systematic treatise on the foundations of social theory and one-sided process of rationalization in modern

societies.)
Homans, G.C. (1961) *Social Behavior: Its Elementary Forms*, New York: Harcourt, Brace, and World.
Levine, D.N. (1995) *Visions of the Sociological Imagination*, Chicago, IL: University of Chicago Press.
Merton, R.K. (1967) *On Theoretical Sociology*, New York: Free Press.
Mouzelis, N. (1995) *Sociological Theory: What Went Wrong? Diagnoses and Remedies*, London: Routledge. (Analysis of the central problems of sociological theory and the means to resolve them.)
Munch, R. (198l) 'Talcott Parsons and the theory of action, I: the structure of the Kantian core', *American Journal of Sociology* 86: 709-39.
Munch, R. (1987) *Theory of Action*, London: Routledge & Kegan Paul.
Parsons, T. (1937) *The Structure of Social Action*, New York: Free Press.
Ritzer, G. (ed.) (1992) *Metatheorizing*, Los Angeles, CA and London: Sage.
Roth, G. (1968) 'Introduction', in *Max Weber*, *Economy and Society*, New York: Bedminster Press, xxxii-cx. (Examines the translation of the term *Herrschaft* and Weber's intended meaning.)
Seidman, S. (1991) 'The End of Sociological Theory: the Postmodern Hope', *Sociological Theory* 8 (Fall).
Stinchcombe, A.L. (1968) *Constructing Social Theories*, New York: Harcourt Brace.
Sztompka, P. (1979) *Sociological Dilemmas: Toward a Dialectical Paradigm*, New York: Academic Press.
Turner, B.S. and Holton, R.J. (1986) 'Against Nostalgia: Talcott Parsons and a Sociology for the Modern World', in *Talcott Parsons on Economy and Society*, London: Routledge, 207-34.
Warnke, G. (1992) *Justice and Interpretation*, Cambridge, MA: MIT Press.(An illuminating discussion of major currents within the new 'interpretive' philosophy by a scholar who remains committed to a more Neo-Kantian project.)
Weber, M. (1968) 'Bureaucracy', in *Economy and Society*, New York: Bedminster Press, 956-1005.(Detailed examination of the basic concepts of social and economics, as applied to history.)

Socrates (469-399 BC)

Socrates, an Athenian Greek of the second half of the fifth century BC, wrote no philosophical works but was uniquely influential in the later history of philosophy. His philosophical interests were restricted to ethics and the conduct of life, topics which thereafter became central to philosophy. He discussed these in public places in Athens, sometimes with other prominent intellectuals or political leaders, sometimes with young men, who gathered round him in large numbers, and other admirers. Among these young men was Plato. Socrates' philosophical ideas and - equally important for his philosophical influence - his personality and methods as a 'teacher' were handed on to posterity in the 'dialogues' that several of his friends wrote after his death, depicting such discussions. Only those of Xenophon (Memorabilia,Apology, Symposium) and the early dialogues of Plato survive (for example Euthyphro, Apology, Crito). Later Platonic dialogues such as Phaedo, Symposium and Republic do not present the historical Socrates' ideas; the 'Socrates' appearing in them is a spokesman for Plato's own ideas.

Socrates' discussions took the form of face-to-face interrogations of another person. Most often they concerned the nature of some moral virtue, such as courage or justice. Socrates asked what the respondent thought these qualities of mind and character amounted to, what their value was, how they were acquired. He would then test their ideas for logical consistency with other highly plausible general views about morality and goodness that the respondent also agreed to accept, once Socrates presented them. He succeeded in showing, to his satisfaction and that of the respondent and any bystanders, that the respondent's ideas were not consistent. By this practice of 'elenchus' or refutation he was able to prove that politicians and others who claimed to have 'wisdom' about human affairs in fact lacked it, and to draw attention to at least apparent errors in their thinking. He wanted to encourage them and others to think harder and to improve their ideas about the virtues and about how to conduct a good human life. He never argued directly for ideas of his own, but always questioned those of others. None the less, one can infer, from the questions he asks and his attitudes to the answers he receives, something about his own views.

Socrates was convinced that our souls - where virtues and vices are found - are vastly more important for our lives than our bodies or external circumstances. The quality of our souls determines the character of our lives, for better or for worse, much more than whether we are healthy or sick, or rich or poor. If we are to live well and happily, as he assumed we all want to do more than we want anything else, we must place the highest priority on the care of our souls. That means we must above all want to acquire the virtues, since they perfect our souls and enable them to direct our lives for the better. If only we could know what each of the virtues is we could then make an effort to obtain them. As to the nature of the virtues, Socrates seems to have held quite strict and, from the popular point of view, paradoxical views. Each virtue consists entirely in knowledge, of how it is best to act in some area of life, and why: additional 'emotional' aspects, such as the disciplining of our feelings and desires, he dismissed as of no importance. Weakness of will is not psychologically possible: if you act wrongly or badly, that is due to your ignorance of how you ought to act and why. He thought each of the apparently separate virtues amounts to the same single body of knowledge: the comprehensive knowledge of what is and is not good for a human being. Thus his quest was to acquire this single wisdom: all the particular virtues would follow automatically.

At the age of 70 Socrates was charged before an Athenian popular court with 'impiety' - with not believing in the Olympian gods and corrupting young men through his constant questioning of everything. He was found guilty and condemned to death. Plato's Apology, where Socrates gives a passionate defence of his life and philosophy, is one of the classics of Western literature. For different groups of later Greek philosophers he was the model both of a sceptical inquirer who never claims to know the truth, and of a 'sage' who knows the whole truth about human life and the human good. Among modern philosophers, the interpretations of his innermost meaning given by Montaigne, Hegel, Kierkegaard, and Nietzsche are especially notable.

1 Life and sources

Socrates, an Athenian citizen proud of his devotion to Athens, lived his adult life there engaging in open philosophical discussion and debate on fundamental questions of ethics, politics, religion and education. Going

against the grain of the traditional education, he insisted that personal investigation and reasoned argument, rather than ancestral custom, or appeal to the authority of Homer, Hesiod and other respected poets, was the only proper basis for answering these questions. His emphasis on argument and logic and his opposition to unquestioning acceptance of tradition allied him with such Sophists of a generation earlier as Protagoras, Gorgias and Prodicus, none of whom was an Athenian, but all of whom spent time lecturing and teaching at Athens (see Sophists). Unlike these Sophists Socrates did not formally offer himself or accept pay as a teacher. But many upper-class young Athenian men gathered round him to hear and engage in his discussions, and he had an inspirational and educational effect upon them, heightening their powers of critical thought and encouraging them to take seriously their individual responsibility to think through and decide how to conduct their lives. Many of his contemporaries perceived this education as morally and socially destructive - it certainly involved subverting accepted beliefs - and he was tried in 399 BC before an Athenian popular court and condemned to death on a charge of ‘impiety’: that he did not believe in the Olympian gods, but in new ones instead, and corrupted the young. Scholars sometimes mention specifically political motives of revenge, based on guilt by association: a number of prominent Athenians who were with Socrates as young men or were close friends did turn against the Athenian democracy and collaborated with the Spartans in their victory over Athens in the Peloponnesian war. But an amnesty passed by the restored democracy in 403 BC prohibited prosecution for political offences before that date. The rhetorician Polycrates included Socrates’ responsibility for these political crimes in his *Accusation of Socrates* (see Xenophon, *Memorabilia* I 2.12), a rhetorical exercise written at least five years after Socrates’ death. But there is no evidence that, in contravention of the amnesty, Socrates’ actual accusers covertly attacked him, or his jurors condemned him, on that ground. The defences Plato and Xenophon constructed for Socrates, each in his respective *Apology*, imply that it was his own questioning mind and what was perceived as the bad moral influence he had on his young men that led to his trial and condemnation.

Socrates left no philosophical works, and apparently wrote none. His philosophy and personality were made known to later generations through the dialogues that several of his associates wrote with him as principal speaker (see Socratic dialogues). Only fragments survive of those by Aeschines of Sphettus and Antisthenes, both Athenians, and Phaedo of Elis (after whom Plato’s dialogue *Phaedo* is named). Our own knowledge of Socrates depends primarily on the dialogues of Plato and the Socratic works of the military leader and historian Xenophon. Plato was a young associate of Socrates’ during perhaps the last ten years of his life, and Xenophon knew him during that same period, though he was absent from Athens at the time of Socrates’ death and for several years before and many years after.

We also have secondary evidence from the comic playwright Aristophanes and from Aristotle. Aristotle, although born fifteen years after Socrates’ death, had access through Plato and others to first-hand information about the man and his philosophy. Aristophanes knew Socrates personally; his *Clouds* (first produced *c*.423 BC) pillories the ‘new’ education offered by Sophists and philosophers by showing Socrates at work in a ‘thinkery’, propounding outlandish physical theories and teaching young men how to argue cleverly in defence of their improper behaviour. It is significant that in 423, when Socrates was about 45 years old, he could plausibly be taken as a leading representative in Athens of the ‘new’ education. But one cannot expect a comic play making fun of a whole intellectual movement to contain an authentic account of Socrates’ specific philosophical commitments.

However, the literary genre to which Plato’s and Xenophon’s Socratic works belong (along with the other, lost dialogues) also permits the author much latitude; in his *Poetics* Aristotle counts such works as fictions of a certain kind, alongside epic poems and tragedies. They are by no means records of actual discussions (despite the fact that Xenophon explicitly so represents his). Each author was free to develop his own ideas behind the mask of Socrates, at least within the limits of what his personal experience had led him to believe was Socrates’ basic philosophical and moral outlook. Especially in view of the many inconsistencies between Plato’s and Xenophon’s portraits (see §7 below), it is a difficult question for historical-philosophical interpretation whether the philosophical and moral views the character Socrates puts forward in any of these dialogues can legitimately be attributed to the historical philosopher. The problem of interpretation is made more difficult by the fact that Socrates appears in many of Plato’s dialogues - ones belonging to his middle and later periods (see Plato §§10-16) - discussing and expounding views that we have good reason to believe resulted from Plato’s own philosophical investigations into questions of metaphysics and epistemology, questions that were not entered into at all by the historical Socrates. To resolve this problem - what scholars call the ‘Socratic problem’ - most agree in preferring

Plato to Xenophon as a witness. Xenophon is not thought to have been philosopher enough to have understood Socrates well or to have captured the depth of his views and his personality. As for Plato, most scholars accept only the philosophical interests and procedures, and the moral and philosophical views, of the Socrates of the early dialogues, and, more guardedly, the Socrates of ‘transitional’ ones such as *Meno* and *Gorgias*, as legitimate representations of the historical personage. These dialogues are the ones that predate the emergence of the metaphysical and epistemological inquiries just referred to. However, even Plato’s early dialogues are philosophical works written to further Plato’s own philosophical interests. That could produce distortions, also; and Xenophon’s relative philosophical innocence could make his portrait in some respects more reliable. Moreover, it is possible, even probable, that in his efforts to help his young men improve themselves Socrates spoke differently to the philosophically more promising ones among them - including Plato - from the way he spoke to others, for example Xenophon. Both portraits could be true, but partial and needing to be combined (see §7). The account of Socrates’ philosophy given below follows Plato, with caution, while giving independent weight also to Xenophon and to Aristotle.

2 Life and sources (cont.)

Xenophon’s *Apology* of Socrates, *Symposium* and *Memorabilia* (or *Memoirs*) may well reflect knowledge of Plato’s own *Apology* and some of his early and middle period dialogues, as well as lost dialogues of Antisthenes and others. Xenophon composed the *Memorabilia* over many years, beginning only some ten years after Socrates’ death, avowedly in order to defend Socrates’ reputation as a good man, a true Athenian gentleman, and a good influence upon his young men. The same intention motivated his*Apology* and *Symposium*. Anything these works contain about Socrates’ philosophical opinions and procedures is ancillary to that apologetic purpose. Plato’s *Apology*, of course, is similarly apologetic, but it and his other early dialogues are carefully constructed discussions, strongly focused upon questions of philosophical substance. Plato evidently thought Socrates’ philosophical ideas and methods were central to his life and to his mission. Xenophon’s and Plato’s testimony are agreed that Socrates’ discussions consistently concerned the *aretai*, the recognized ‘virtues’ or excellences of character (see Aretē), such as justice, piety, self-control or moderation (*sōphrosynē*), courage and wisdom; what these individual characteristics consist in and require of a person, what their value is, and how they are acquired, whether by teaching or in some other way. In his *Apology* and elsewhere Plato has Socrates insist that these discussions were always inquiries, efforts made to engage his fellow-discussants in coming jointly to an adequate understanding of the matters inquired into. He does not himself know, and therefore cannot teach anyone else - whether by means of these discussions or in some other way - either how to be virtuous or what virtue in general or any particular virtue is. Furthermore, given his general characterization of virtue (see §§4-5), Plato’s Socrates makes a point of suggesting the impossibility in principle of teaching virtue at all, by contrast with the Sophists who declared they could teach it. Virtue was not a matter of information about living or rote techniques of some sort to be handed on from teacher to pupil, but required an open-ended personal understanding that individuals could only come to for themselves. Xenophon, too, reports that Socrates denied he was a teacher of *aretē*, but he pays no attention to such issues of philosophical principle. He does not hesitate to show Socrates speaking of himself as a teacher (see *Apology* 26, *Memorabilia* I 6.13-14), and describes him as accepting young men from their fathers as his pupils (but not for a fee), and teaching them the virtues by displaying his own virtues to them for emulation, as well as through conversation and precepts. Perhaps Socrates did not insist on holding to strict philosophical principles in dealing with people on whom their point would have been lost.

In his *Apology* Plato’s Socrates traces his practice of spending his days discussing and inquiring about virtue to an oracle delivered at the shrine of Apollo at Delphi. Xenophon also mentions this oracle in his *Apology*. A friend of Socrates’, Chaerephon, had asked the god whether anyone was wiser than Socrates; the priestess answered that no one was. Because he was sure he was not wise at all - only the gods, he suspected, could actually *know* how a human life ought to be led - Socrates cross-examined others at Athens with reputations for that kind of wisdom. He wanted to show that there were people wiser than he and thus discover the true meaning of the oracle - Apollo was known to speak in riddles requiring interpretation to reach their deeper meaning. In the event, it turned out that the people he examined were not wise, since they could not even give a self-consistent set of answers to his questions: obviously, true knowledge requires at least that one think and speak consistently on the subjects one professes to know. So he concluded that the priestess’s reply had meant that of all those with reputations for wisdom only he came close to deserving it; he wisely did not profess to know these things that only gods can know, and that was

wisdom enough for a human being. Because only he knew that he did *not* know, only he was ready earnestly to inquire into virtue and the other ingredients of the human good, in an effort to learn. He understood therefore that Apollo's true intention in the oracle had been to encourage him to continue his inquiries, to help others to realize that it is beyond human powers actually to *know* how to live - that is the prerogative of the gods - and to do his best to understand as far as a human being can how one ought to live. The life of philosophy, as led by him, was therefore something he was effectively ordered by Apollo to undertake.

We must remember that Socrates was on trial on a charge of 'impiety'. In tracing his philosophical vocation back to Apollo's oracle, and linking it to a humble recognition of human weakness and divine perfection, he was constructing a powerful rebuttal of the charges brought against him. But it cannot be literally true - if that is what he intended to say - that Socrates began his inquiries about virtue only after hearing of the oracle. Chaerephon's question to Apollo shows he had established a reputation in Athens for wisdom before that. That reputation cannot have rested on philosophical inquiries of another sort. In Plato's *Phaedo* Socrates says he had been interested as a young man in philosophical speculations about the structure and causes of the natural world, but he plainly did not take those interests very far; and in any event, his reputation was not for that kind of wisdom, but wisdom about how to lead a human life. In fact we do not hear of the duty to Apollo in Xenophon, or in other dialogues of Plato, where we might expect to find it if from the beginning Socrates thought Apollo had commanded his life of philosophizing. However, we need not think Socrates was false to the essential spirit of philosophy as he practised it if in looking back on his life under threat of condemnation for impiety he chose, inaccurately, to see it as initially imposed on him by Apollo's oracle.

Despite its impressiveness, Socrates' speech failed to convince his jury of 501 male fellow citizens, and he died in the state prison by drinking hemlock as required by law. His speech evidently offended the majority of the jurors by its disdain for the charges and the proceedings; Xenophon explains his lofty behaviour, which he thinks would otherwise have been lunatic - and damaging to his reputation - by reporting that he had told friends in advance that as a 70-year-old still in possession of his health and faculties it was time for him to die anyhow, before senility set in. Furthermore, his 'divine sign' - the 'voice' he sometimes heard warning him for his own good against a contemplated course of action - had prevented him from spending time crafting a defence speech. (This voice seems to have been the basis for the charge of introducing 'new' gods.) So he would do nothing to soften his manner in order to win his freedom. Even if this story is true, Plato could be right that Socrates put on a spirited, deeply serious defence of his life and beliefs - one that he thought should have convinced the jurors of his innocence, if only they had judged him intelligently and fairly.

3 Socratic elenchus, or refutation

In cross-examining those with reputations for wisdom about human affairs and showing their lack of it, Socrates employed a special method of dialectical argument that he himself had perfected, the method of 'elenchus' - Greek for 'putting to the test' or 'refutation'. He gives an example at his trial when he cross-examines Meletus, one of his accusers (Plato, *Apology* 24d-27e). The respondent states a thesis, as something he *knows* to be true because he is wise about the matter in question. Socrates then asks questions, eliciting clarifications, qualifications and extensions of the thesis, and seeking further opinions of the respondent on related matters. He then argues, and the respondent sees no way not to grant, that the original thesis is logically inconsistent with something affirmed in these further responses. For Socrates, it follows at once that the respondent did not know what he was talking about in stating his original thesis: true knowledge would prevent one from such self-contradiction. So the respondent suffers a personal set-back; *he* is refuted - revealed as incompetent. Meletus, for example, does not have consistent ideas about the gods or what would show someone not to believe in them, and he does not have consistent ideas about who corrupts the young, and how; so he does not know what he is talking about, and no one should take *his* word for it that Socrates disbelieves in the gods or has corrupted his young men. In many of his early dialogues Plato shows Socrates using this method to examine the opinions of persons who claim to be wise in some matter: the religious expert Euthyphro on piety (*Euthyphro*), the generals Laches and Nicias on courage (*Laches*), the Sophist Protagoras on the distinctions among the virtues and whether virtue can be taught (*Protagoras*), the rhapsodist Ion on what is involved in knowing poetry (*Ion*), the budding politician Alcibiades on justice and other political values (*Alcibiades*), the Sophist Hippias on which was the better man, Odysseus or Achilles (*Lesser Hippias*), and on the nature of moral and aesthetic beauty (*Greater Hippias*). They are all refuted - shown to have mutually inconsistent ideas on the subject discussed (see Plato §§4, 6, 8-9).

But Socrates is not content merely to demonstrate his interlocutor's lack of wisdom or knowledge. That might humiliate him into inquiring further or seeking by some other means the knowledge he has been shown to lack, instead of remaining puffed up with self-conceit. That would be a good thing. But Socrates often also indicates clearly that his cross-examination justifies him and the interlocutor in rejecting as false the interlocutor's original thesis. Logically, that is obviously wrong: if the interlocutor contradicts himself, at least one of the things he has said must be false (indeed, all of them could be), but the fact alone of self-contradiction does not show where the falsehood resides. For example, when Socrates leads Euthyphro to accept ideas that contradict his own definition of the pious as whatever pleases all the gods, Socrates concludes that that definition has been shown to be false (*Euthyphro* 10d-11a), and asks Euthyphro to come up with another one. He does not usually seem to consider that perhaps on further thought the additional ideas would seem faulty and so merit rejection instead.

Socrates uses his elenctic method also in discussion with persons who are not puffed up with false pride, and are quite willing to admit their ignorance and to reason out the truth about these important matters. Examples are his discussions with his long-time friend Crito on whether he should escape prison and set aside the court's death sentence (Plato, *Crito*), and with the young men Charmides, on self-control (*Charmides*), and Lysis and Menexenus, on the nature of friendship (*Lysis*). Socrates examines Crito's proposal that he escape on the basis of principles that he presents to him for his approval, and he, together with Crito (however half-heartedly), rejects it when it fails to be consistent with them. And he examines the young men's successive ideas about these virtues, rejecting some of them and refining others, by relying on their own acceptance of further ideas that he puts to them. Again, he is confident that the inconsistencies brought to light in their ideas indicate the inadequacy of their successive proposals as to the nature of the moral virtue in question.

In many of his discussions, both with young men and the allegedly wise, Socrates seeks to know what some morally valuable property is - for example, piety, courage, self-control or friendship (see §5). Rejecting the idea that one could learn this simply from attending to examples, he insisted on an articulated 'definition' of the item in question - some *single* account that would capture all at once the presumed common feature that would entitle anything to count as a legitimate instance. Such a definition, providing the essence of the thing defined, would give us a 'model' or 'paradigm' to use in judging whether or not some proposed action or person possesses the moral value so defined (*Euthyphro* 6d-e). Aristotle says (in *Metaphysics* I, 6) that Socrates was the first to interest himself in such 'universal definitions', and traces to his interest in them Plato's first impetus towards a theory of Forms, or 'separated' universals (see Plato §10).

In none of his discussions in Plato's early works does Socrates profess to think an adequate final result has actually been established - about the nature of friendship, or self-control, or piety, or any of the other matters he inquires about. Indeed, on the contrary, these works regularly end with professions of profound ignorance about the matter under investigation. Knowledge is never attained, and further questions always remain to be considered. But Socrates does plainly think that progress towards reaching final understanding has taken place (even if only a god, and no human being, could ever actually attain it). Not only has one discovered some things that are definitely wrong to say; one has also achieved some positive insights that are worth holding onto in seeking further systematic understanding. Given that Socrates' method of discussion is elenctic throughout, what does he think justifies this optimism?

On balance, our evidence suggests that Socrates had worked out no elaborate theory to support him here. The ideas he was stimulated to propound in an elenctic examination which went against some initial thesis seemed to him, and usually also to the others present, so plausible, and so supportable by further considerations, that he and they felt content to reject the initial thesis. Until someone came up with arguments to neutralize their force, it seemed the thesis was doomed, as contrary to reason itself. Occasionally Socrates expresses himself in just those terms: however unpalatable the option might seem, it remains open to someone to challenge the grounds on which his conclusions rest (see *Euthyphro* 15c, *Gorgias* 461d-462a, 509a, *Crito* 54d). But until they do, he is satisfied to treat his and his interlocutor's agreement as a firm basis for thought and action. Later, when Plato himself became interested in questions of philosophical methodology in his *Meno*, this came to seem a philosophically unsatisfactory position; Plato's demand for justification for one's beliefs independent of what seemed on reflection most plausible led him to epistemological and metaphysical inquiries that went well beyond the self-imposed restriction of Socratic philosophy to ethical thought in the broadest sense. But Socrates did not raise these questions. In this respect more bound by traditional views than Plato, he had great implicit confidence in his and

his interlocutors' capacity, after disciplined dialectical examination of the issues, to reach firm ground for constructing positive ideas about the virtues and about how best to lead a human life - even if these ideas never received the sort of final validation that a god, understanding fully the truth about human life, could give them.

4 Elenchus and moral progress

The topics Socrates discussed were always ethical, and never included questions of physical theory or metaphysics or other branches of philosophical study. Moreover, he always conducted his discussions not as theoretical inquiries but as profoundly personal moral tests. Questioner and interlocutor were equally putting their ways of life to what Socrates thought was the most important test of all - their capacity to stand up to scrutiny in rational argument about how one ought to live. In speaking about human life, he wanted his respondents to indicate what they truly believed, and as questioner he was prepared to do the same, at least at crucial junctures. Those beliefs were assumed to express not theoretical ideas, but the very ones on which they themselves were conducting their lives. In losing an argument with Socrates you did not merely show yourself logically or argumentatively deficient, but also put into question the very basis on which you were living. Your way of life might ultimately prove defensible, but if you cannot now defend it successfully, you are not leading it *with* any such justification. In that case, according to Socrates' views, your way of life is morally deficient. Thus if Menexenus, Lysis and Socrates profess to value friendship among the most important things in life and profess to be one another's friends, but cannot satisfactorily explain under pressure of elenctic investigation what a friend is, that casts serious doubt on the quality of any 'friendship' they might form (Plato, *Lysis* 212a, 223b). Moral consistency and personal integrity, and not mere delight in argument and logical thought, should therefore lead you to repeated elenctic examination of your views, in an effort to render them coherent and at the same time defensible on all sides through appeal to plausible arguments. Or, if some of your views have been shown false, by conflicting with extremely plausible general principles, it behoves you to drop them - and so to cease living in a way that depends upon accepting them. In this way, philosophical inquiry via the elenchus is fundamentally a personal moral quest. It is a quest not just to understand adequately the basis on which one is actually living, and the personal and moral commitments that this contains. It is also a quest to change the way one lives as the results of argument show one ought to, so that, at the logical limit of inquiry, one's way of life would be completely vindicated. Accordingly, Socrates in Plato's dialogues regularly insists on the individual and personal character of his discussions. He wants to hear the views of the one person with whom he is speaking. He dismisses as of no interest what outsiders or most people may think - provided that is not what his discussant is personally convinced is true. The views of 'the many' may well not rest on thought or argument at all. Socrates insists that his discussant shoulder the responsibility to explain and defend rationally the views he holds, and follow the argument - reason - wherever it may lead.

We learn a good deal about Socrates' own principles from both Plato and Xenophon. Those were ones that had stood up well over a lifetime of frequent elenctic discussions and had, as he thought, a wealth of plausible arguments in their favour. Foremost is his conviction that the virtues - self-control, courage, justice, piety, wisdom and related qualities of mind and soul - are essential if anyone is to lead a good and happy life. They are good in themselves for a human being, and they guarantee a happy life, *eudaimonia* - something that he thought all human beings always wanted, and wanted more than anything else. The virtues belong to the soul - they are the condition of a soul that has been properly cared for and brought to its best state. The soul is vastly more important for happiness than are health and strength of the body or social and political power, wealth and other external circumstances of life; the goods of the soul, and pre-eminently the virtues, are worth far more than any quantity of bodily or external goods. Socrates seems to have thought these other goods *are* truly good, but they only *do* people good, and thereby contribute to their happiness, under the condition that they are chosen and used in accordance with virtues indwelling in their souls (see Plato, *Apology* 30b, *Euthydemus* 280d-282d, *Meno* 87d-89a).

More specific principles followed. Doing injustice is worse for oneself than being subjected to it (*Gorgias* 469c-522e): by acting unjustly you make your soul worse, and that affects for the worse the whole of your life, whereas one who treats you unjustly at most harms your body or your possessions but leaves your soul unaffected. On the same ground Socrates firmly rejected the deeply entrenched Greek precept to aid one's friends and harm one's enemies, and the accompanying principle of retaliation, which he equated with returning wrongs for wrongs done to oneself and one's friends (*Crito* 49a-d). Socrates' daily life gave witness to his principles. He was poor, shabbily dressed and unshod, and made do with whatever ordinary food came his way: such things matter little. Wealth, finery and delicacies for the palate are not worth panting after and exerting oneself to enjoy. However,

Socrates was fully capable of relishing both refined and plain enjoyments as occasion warranted (see §7).

5 The unity of virtue

The Greeks recognized a series of specially prized qualities of mind and character as *aretai* or virtues. Each was regarded as a distinct, separate quality: justice was one thing, concerned with treating other people fairly, courage quite another, showing itself in vigorous, correct behaviour in circumstances that normally cause people to be afraid; and self-control or moderation, piety and wisdom were yet others. Each of these ensured that its possessor would act in some specific ways, regularly and reliably over their lifetime, having the justified conviction that those are ways one *ought* to act - *agathon* (good) and *kalon* (fine, noble, admirable or beautiful) ways of acting. But each type of virtuous person acts rightly and well not only in regularly recurring, but also in unusual and unheralded, circumstances; the virtue involves always getting something right about how to live a good human life. Socrates thought these virtues were essential if one was to live happily (see §4). But what exactly were they? What was it about someone that made them just, or courageous, or wise? If you did not know that, you would not know what to do in order to acquire those qualities. Furthermore, supposing you did possess a virtue, you would have to be able to explain and defend by argument the consequent ways in which you lived - otherwise your conviction that those are ways one *ought* to act would be shallow and unjustified. And in order to do that you would have to know what state of mind the virtue was, since that is essential to them (see Plato, *Charmides* 158e-159a). Consequently, in his discussions Socrates constantly asked for ‘definitions’ of various virtues: what is courage (*Laches*); what is self-control or moderation (*Charmides*), what is friendship (*Lysis*) and what is piety (*Euthyphro*). As this context shows, he was asking not for a ‘dictionary definition’, an account of the accepted linguistic understanding of a term, but for an ethically defensible account of an actual condition of mind or character to which the word in common use would be correctly applied. In later terminology, he was seeking a ‘real’ rather than a ‘nominal’ definition (see Definition; Plato §§6-9).

Socrates objected to definitions that make a virtue some external aspect of a virtuous action (such as the manner in which it is done - for example its ‘quiet’ or measured quality in the case of moderation, *Charmides* 160b-d), or simply the doing of specific types of action, described in terms of their external circumstances (such as, for courage, standing one’s ground in battle; *Laches* 190e-191d). He also objected to more psychological definitions that located a virtue in some non-rational and non-cognitive aspect of the soul (for example, in the case of courage, the soul’s endurance or strength of resistance) (*Laches* 192d-193e). For his own part, he regularly shows himself ready to accept only definitions that identify a virtue with some sort of knowledge or wisdom about what is valuable for a human being. That ‘intellectualist’ expectation about the nature of virtue, although never worked out to his satisfaction in any Platonic dialogue, is central to Socrates’ philosophy.

Given that in his discussions he is always the questioner, probing the opinions of his respondent and not arguing for views of his own, we never find Socrates stating clearly what led him to this intellectualism. Probably, however, it was considerations drawn from the generally agreed premise that each virtue is a condition motivating certain voluntary actions, chosen because they are good and fine or noble. He took it that what lies behind and produces any voluntary action is the idea under which it is done, the conception of the action in the agent’s mind that makes it seem the thing to do just then. If so, each virtue must be some state of the mind, the possessor of which constantly has certain distinctive general ideas about how one ought to behave. Furthermore, since virtues get this *right*, these are true ideas. And since a virtuous person acts well and correctly in a *perfectly reliable* way, they must be seated so deeply in the mind as to be ineradicable and unwaveringly present. The only state of mind that meets these conditions is knowledge: to know a subject is not just to be thoroughly convinced, but to have a deep, fully articulated understanding, being ready with explanations to fend off objections and apparent difficulties and to extend old principles into new situations, and being prepared to show with the full weight of reason precisely why each thing falling under it is and must be so. Each virtue, then, must be knowledge about how one ought to behave in some area of life, and why - a knowledge so deep and rationally secure that those who have it can be counted upon never to change their minds, never to be argued out of or otherwise persuaded away from, or to waver in, their conviction about how to act.

In Plato’s *Protagoras* Socrates goes beyond this, and identifies himself with the position, rejected by Protagoras in their discussion, that the apparently separate virtues of justice, piety, self-control, courage and wisdom are somehow one and the same thing - some single knowledge (361a-b). Xenophon too confirms that Socrates held

this view (*Memorabilia* III 9.5). Protagoras defends the position that each of the virtues is not only a distinct thing from each of the others, but so different in kind that a person could possess one of them without possessing the others (329d-e). In opposing him, Socrates sometimes speaks plainly of two allegedly distinct virtues being ‘one’ (333b). Given this unity of the virtues, it would follow that a person could not possess one without having them all. And in speaking of justice and piety in particular, Socrates seems to go further, to imply that every action produced by virtue is equally an instance of *all* the standardly recognized virtues: pious as well as just, wise and self-controlled and courageous also. Among his early dialogues, however, Plato’s own philosophical interests show themselves particularly heavily in the *Protagoras*, so it is doubtful how far the details of his arguments are to be attributed to the historical Socrates. The issues raised by Socrates in the *Protagoras* were, none the less, vigorously pursued by subsequent ‘Socratic’ philosophers (as Plutarch’s report in *On Moral Virtue* 2 demonstrates). And the positions apparently adopted by Plato’s Socrates were taken up and ingeniously defended by the Stoic philosopher Chrysippus (see Stoicism §16). As usual, because of his questioner’s role, it is difficult to work out Socrates’ grounds for holding to the unity of virtue; and it is difficult to tell whether, and if so how, he allowed that despite this unity there were some real differences between, say, justice and self-control, or courage and piety. Apparently he thought the same body of knowledge - knowledge of the whole of what is and is not good for human beings, and why it is so or not - must at least underlie the allegedly separate virtues. If you did not have that vast, comprehensive knowledge you could not be in the state of mind which is justice or in that which is courage, and so on; and if you did have it you would necessarily be in those states of mind. It seems doubtful whether Socrates himself progressed beyond that point. Efforts to do that were made by Chrysippus and the other philosophers referred to above. And despite denying that all virtues consist in knowledge, Plato in the *Republic* and Aristotle in *Nicomachean Ethics* VI follow Socrates to the extent of holding, in different ways, that you need to have all the virtues in order to have any one.

6 Weakness of will denied

In Plato’s *Protagoras* Socrates also denies the possibility of weakness of will - being ‘mastered’ by some desire so as to act voluntarily in a way one knows is wrong or bad (see also Xenophon, *Memorabilia* III 9.4, IV 5.6.) All voluntary wrongdoing or bad action is due to ignorance of how one ought to act and why, and to nothing else. This would be easy to understand if Socrates were using ‘knowing’ quite strictly, to refer to the elevated and demanding sort of knowledge described in §5 (sometimes called ‘Socratic knowledge’). Someone could know an action was wrong or bad, with full ‘Socratic knowledge’, only if they were not just thoroughly convinced, but had a deep, fully articulated understanding, being ready with explanations to fend off objections and apparent difficulties, and prepared to show precisely *why* it was so. That would mean that these ideas were seated so deeply in the mind as to be ineradicable and unwaveringly present. Accordingly, a person with ‘Socratic knowledge’ could not come to hold even momentarily that the action in question would be the thing to do, and so they could never do it voluntarily.

However, Plato’s Socrates goes further. He explains his denial of weak-willed action by saying that a person cannot voluntarily do actions which, in doing them, they even *believe* to be a wrong or bad thing to do (*Protagoras* 358c-e). He gives a much-discussed, elaborate argument to establish this stronger conclusion, starting from assumptions identifying that which is pleasant with that which is good (352a-357e). These assumptions, however, he attributes only to ordinary people, the ones who say they believe in the possibility of weak-willed action; he makes it clear to the careful reader, if not to Protagoras, that his own view is simply that pleasure is a good thing, not ‘the’ good (351c-e; see 354b-d). Although some scholars have thought otherwise, Socrates himself does not adopt a hedonist analysis of the good in the *Protagoras* or elsewhere either in Plato or Xenophon; indeed, he speaks elsewhere against hedonist views (see Hedonism). The fundamental principle underlying his argument - a principle he thinks ordinary people will accept - is that voluntary action is always ‘subjectively’ rational, in the sense that an agent who acts to achieve some particular sort of value always acts with the idea that what they are doing achieves *more* of that value than alternatives then thought by them to be available would achieve. If someone performs an overall bad action because of some (lesser) good they think they will get from it, they cannot do it while *believing* it is bad overall. That would mean they thought they could have got *more* good by refraining, and their action would violate the principle just stated. Instead, at the time they acted (despite what they may have thought before or after acting), they believed (wrongly and ignorantly) that the action would be good overall for them to do. Thus ignorance, and only ignorance, is responsible for voluntary error. Weakness of will - knowingly

pursuing the worse outcome - is psychologically impossible: ‘No one does wrong willingly’.

The details of this argument may not represent explicit commitments of the historical Socrates. None the less, his denial of weakness of will, understood as presented in Plato’s *Protagoras*, was the centre of a protracted debate in later times. First Plato himself, in *Republic* IV, then Aristotle in *Nicomachean Ethics* VII, argued against Socrates’ conclusion, on the ground that he had overlooked the fact that human beings have other sources of motivation that can produce voluntary actions, besides their ideas about what is good or bad, or right or wrong to do. ‘Appetites’ and ‘spirited desires’ exist also, which can lead a person to act in fulfilment of them without having to adopt the idea, in their beliefs about what is best to do, that so acting would be a good thing (see Plato §14; Aristotle §20, 22-23). The Stoics, however, and especially Chrysippus, argued vigorously and ingeniously in defence of Socrates’ analysis and against the Platonic-Aristotelian assumption of alternative sources of motivation that produce voluntary action on their own (see Stoicism §19). In fact, during Hellenistic times it was the Socratic, ‘unitary’ psychology of action that carried the day; the Platonic-Aristotelian alternative, dominant in the ‘common sense’ and the philosophy of modern times, was a minority view. The issues Socrates raised about weakness of will continue to be debated today.

7 Socrates’ personality

Socrates drew to himself many of the brightest and most prominent people in Athens, securing their fascinated attention and their passionate friendship and support. His effectiveness as a philosopher, and the Socratic ‘legend’ itself, depended as much on the strength and interest of his personality as on the power of his mind. Plato’s and Xenophon’s portraits of Socrates as a person differ significantly, however. Plato’s Socrates is aloof and often speaks ironically, although also with unusual and deeply held moral convictions; paradoxically, the depth and clarity of his convictions, maintained alongside the firm disclaimer to know what was true, could seem all the stronger testimony to their truth, and made them felt the more strongly as a rebuke to the superficiality of one’s own way of living. In Xenophon, Socrates is also sometimes ironical and playful, especially in the *Symposium*, but his conversation is usually direct, even didactic, and often chummy in tone; his attitudes are for the most part conventional though earnest; and there is nothing to unsettle anyone or make them suspect hidden depths. It is much easier to believe that the Socrates of Plato’s dialogues could have had such profound effects on the lives of the brightest of his contemporaries than did the character in Xenophon. That is one reason given for trusting Plato’s more than Xenophon’s portrait of the historical personage. But perhaps Socrates used the more kindly and genial manner and conventional approach depicted by Xenophon to draw out the best in some of his young men and his friends - ones who would have been put off by the Platonic subtleties. The historical Socrates may have been a more complex person than even Plato presents.

Plato and Xenophon both represent Socrates as strongly attracted to good-looking young men in the ‘bloom’ of their middle to late teens, just the period when they were also coming of age morally and intellectually. In both he speaks of himself as unusually ‘erotic’ by temperament and constantly ‘in love’. But he explains his ‘erotic’ attachments in terms of his desire to converse with bright and serious young men, to question them about virtue and how best to live a human life, and to draw out what was best in their minds and characters. In Xenophon he describes his love as love for their souls, not their bodies, and he vigorously condemns sexual relations with any young man: using him that way disgraces him and harms him by encouraging a loose attitude as regards physical pleasures *Symposium* 8). The overheated sexuality of Plato’s own accounts (*Symposium* and *Phaedrus*) of *erōs*, sexual love, for a young man’s beauty as motivating an adult male to pursue philosophical truth into an eternal realm of Forms (see Plato §12) is to be distinguished sharply from Socrates’ ideas, as we can gather them from Xenophon and from Plato’s own early dialogues.

Xenophon emphasizes Socrates’ freedom from the strong appetites for food, drink, sex and physical comfort that dominate other people; his *enkrateia* or self-mastery is the first of the virtues that Xenophon claims for him (*Memorabilia* I 2.1). He was notorious for going barefoot even in winter and dressing always in a simple cloak. Socrates’ self-mastery was at the centre of Antisthenes’ portrayal, and is reflected also in several incidents reported in Plato, such as his serene dismissal of the young Alcibiades’ efforts to seduce him sexually (Plato, *Symposium* 217b-219e), or, perhaps when engrossed in a philosophical problem, his standing in the open (during a break in the action while on military service) from morning to night, totally indifferent to everything around him (*Symposium* 220c-d). This ‘ascetic’ Socrates, especially as presented by Antisthenes - rejecting conventional

comforts and conventional behaviour - became an inspiration for the 'Cynics' of later centuries (see Cynics).

8 Socrates in the history of philosophy

Looking back on the early history of philosophy, later philosophers traced to Socrates a major turn in its development. As Cicero puts it: 'Socrates was the first to call philosophy down from the heavens… and compel it to ask questions about life and morality' (*Tusculan Disputations* V 10-11). Previously it had been concerned with the origins and nature of the physical world and the explanation of celestial and other natural phenomena. Modern scholarship follows the ancients' lead in referring standardly to philosophers before Socrates collectively as 'Presocratics' (see Presocratic philosophy). This includes Democritus, in fact a slightly younger contemporary of Socrates; Cicero's verdict needs adjustment, in that Democritus, independently of Socrates, also investigated questions about ethics and morality. With the sole exception of Epicureanism, which developed separately out of Democritean origins, all the major movements of Greek philosophy after Socrates had roots in his teaching and example. This obviously applies to Plato, whose philosophical development began with a thorough reworking and assimilation of Socratic moral inquiry, and through him to Aristotle and his fellow members of Plato's Academy, Speusippus and Xenocrates and others, as well as to later Platonists. Among Socrates' inner circle were also Aristippus of Cyrene, who founded the hedonist Cyrenaic school (see Aristippus the Elder; Cyrenaics), and Antisthenes, an older rival of Plato's and major teacher in Athens of philosophical dialectic. Both of these figure in Xenophon's *Memorabilia* (Antisthenes also in his *Symposium*), where they are vividly characterized in conversation with Socrates. Another Socratic, Euclides, founded the Megarian school (see Megarian school). These 'Socratic schools' developed different themes already prominent in Socrates' own investigations, and competed in the claim to be his true philosophical heirs (see Socratic schools; Dialectical School).

In the third to first centuries BC, both the Stoics and their rivals the Academic sceptics claimed to be carrying forward the Socratic tradition. In both cases this was based upon a reading of Plato's dialogues and perhaps other eye-witness reconstructions of Socrates' philosophy. The Academic Arcesilaus interpreted the Platonic Socrates as a sceptical inquirer, avidly searching but never satisfied that the truth on any disputed question had been finally uncovered. He could point to much about Plato's Socrates in support: his modest but firm denial that he possessed any knowledge, and his constant practice of inquiring into the truth by examining others' opinions on the basis of ideas which they themselves accepted, without formally committing himself to these ideas even when he was the one to first suggest them. Arcesilaus, however, applied his sceptical Socratic dialectic to more than the questions of ethics and human life about which Socrates himself had argued, making it cover the whole range of philosophical topics being investigated in his day. The Stoics read the dialogues (especially the *Euthydemus* and *Protagoras*) quite differently. They found Socrates espousing a complete doctrine of ethics and the psychology of human action. He posed his questions on the basis of this doctrine, leaving the respondent (and the reader) to recover for themselves the philosophical considerations underlying it. They thus emphasized the conceptions of virtue as knowledge, of virtue as unified in wisdom, and of voluntary action as motivated always by an agent's beliefs about what is best to do, that emerged through Socrates' examination of Protagoras (see §§6-7). They thought these constituted a positive, Socratic moral philosophy, and in their own moral theory they set out to revive and strengthen it with systematic arguments and with added metaphysical and physical speculations of their own. Later Stoics regularly referred to Socrates as a genuine wise man or 'sage', perhaps the only one who ever lived. He had brought to final, systematic perfection his knowledge, along Stoic lines, of what is good and bad for human beings, and what is not, and therefore possessed all the virtues and no vices, and lived unwaveringly the best, happy life, free from emotion and all other errors about human life. It is a tribute to the complexity and enigmatic character of Socrates that he could stand simultaneously as a paragon both of sceptical, non-committal inquiry and life led on that uncommitted basis, and of dogmatic knowledge of the final truth about all things human.

The figure of Socrates has continued to fascinate and to inspire ever-new interpretations of his innermost meaning. For Montaigne, he proved that human beings can convincingly and attractively order their own lives from their own resources of mind, without direction from God or religion or tradition. In the nineteenth century Kierkegaard and Nietzsche offered extensive interpretations of him, both heavily dependent upon Hegel's absolute-idealist analysis. Hegel interpreted Socrates as a quintessentially negative thinker, aiming at making people vacillate in their superficial moral beliefs and endorse none of them wholeheartedly, thus hinting that the truth, although universal and objective, lies deep within the freedom of their own subjectivity. For Kierkegaard he represents, on

the contrary, the possibility of living wholeheartedly by occupying an unarticulated position somehow beyond the negative rejection but expressed through it: ‘infinite absolute negativity’. In *Die Geburt der Tragödie (The Birth of Tragedy)* Nietzsche treats Socrates principally as having poisoned the ‘tragic’ attitude that made possible the great achievements of classical culture, by insisting that life should be grounded in rational understanding and justified by ‘knowledge’; but his fascinated regard for Socrates led him to return to him repeatedly in his writings. Socrates was paradigmatically a philosopher whose thought, however taken up with logic and abstract argument, is inseparable from the search for self-understanding and from a deeply felt attachment to the concerns of human life. His power to fascinate and inspire is surely not exhausted.

JOHN M. COOPER

References and further reading

Aristophanes’ (*c*.423 BC) *Clouds*, ed. K.J. Dover, *Aristophanes Clouds*, Oxford: Clarendon Press, 1968; trans. A.H. Sommerstein, Warminster: Aris & Phillips, 1982; trans. A.H. Sommerstein, *Aristophanes: Lysistrata, The Acharnians, The Clouds*, Harmondsworth: Penguin, 1973.(Classic comic portrayal of Socrates. Dover is an edition of the Greek text with explanatory notes; Sommerstein’s 1982 translation also includes the Greek text and explanatory notes.)

Aristotle (*c*. mid 4th century BC) *Metaphysics* I 6, XIII 4, *Nicomachean Ethics* VII 2-3, *Sophistical Refutations* 34, *Magna Moralia* I 1, *Poetics* 1, in J. Barnes (ed.) *The Complete Works of Aristotle*, revised Oxford Translation, Princeton, NJ: Princeton University Press, 1984, 2 vols.(Important testimony on Socrates. The *Magna Moralia* is possibly by a follower of Aristotle, and is of uncertain date.)

Benson, H.H. (ed.) (1992) *Essays on the Philosophy of Socrates*, New York: Oxford University Press.(Reprints fifteen of the best journal articles of the 1970s and 1980s on Socrates and Socratic philosophy; broad coverage of topics, full bibliography.)

Cicero, M.T. (late 45 BC) *Tusculan Disputations*, trans. J.E. King, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1927.(Parallel Latin text and English translation.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae (The Fragments of Socrates and the Socratics)*, Naples: Bibliopolis, 4 vols.(Volumes 1 and 2 contain the surviving Greek and Latin testimonia of Socrates - other than in Plato and Xenophon - and of Antisthenes, Aeschines, Aristippus and other ‘minor’ Socratics; volume 4 has notes in Italian.)

Grote, G. (1875) *Plato and the Other Companions of Socrates*, London: Murray, 3 vols.(Judicious, perceptive older account of Socrates in Plato’s works and of the other Socratics; still valuable.)

Hegel, G.W.F. (1833-6) *Lectures on the History of Philosophy*, trans. E.S. Haldane and F.H. Simpson, London: Routledge & Kegan Paul, 1892.(Volume 1, pages 384-487 concern Socrates and the ‘minor’ Socratics.)

Kierkegaard, S. (1840-1) *Om Begrebet Iron*, trans. L.M. Capel, *The Concept of Irony*, London: Collins, 1965. (Subtitled ‘with constant reference to Socrates’.)

Maier, H. (1913) *Sokrates*, Tübingen: Mohr.(Authoritative German treatment of the ‘Socratic problem’, arguing for the reliability of Plato’s testimony as against that of Xenophon.)

Nietzsche, F. (1872) *Die Geburt der Tragödie*, trans. W. Kaufmann, *The Birth of Tragedy*, New York: Vintage Books, 1967.(Sections 7-13 especially concern Socrates.)

Plato (*c*.390s-380s BC) *Apology*, *Euthyphro*, *Crito*, *Charmides*, *Laches*, *Lysis*, *Ion*, *Lesser Hippias*, *Protagoras*, *Greater Hippias*, in *Plato: Complete Works*, ed. J.M. Cooper, Indianapolis, IN: Hackett, 1997.(These are the works of Plato usually categorized as ‘early’ or ‘Socratic’ dialogues.)

Plato (*c*.380s-365 BC) *Meno*, *Gorgias*, *Phaedo*, *Symposium*, *Republic* I, *Phaedrus*, *Euthydemus*, *Theaetetus*, in *Plato: Complete Works*, ed. J.M. Cooper, Indianapolis, IN: Hackett, 1997.(The philosophical ideas in these dialogues cannot in general be assumed to be those of the historical Socrates; but they do contain biographical information and characteristic portrayals of his philosophical personality and manner.)

Reeve, C.D.C. (1989) *Socrates in the Apology*, Indianapolis, IN: Hackett.(Commentary on Plato’s *Apology* in the light of Socrates’ general philosophy; accessible to the general reader.)

Vlastos, G. (ed.) (1971) *The Philosophy of Socrates*, Garden City, NY: Doubleday.(Collection of essays: see especially G. Vlastos, ‘The Paradox of Socrates’, A.R. Lacey, ‘Our Knowledge of Socrates’, and M.F. Burnyeat, ‘Virtues in Action’.)

Vlastos, G. (1991) *Socrates, Ironist and Moral Philosopher*, Ithaca, NY: Cornell University Press. (Comprehensive account based on a lifetime’s engagement with Socrates’ philosophy; good bibliography.)

Xenophon (*c*.385 BC) *Apology of Socrates to the Jury*, trans. O.J. Todd, *Socrates’ Defence to the Jury (Apology)*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1922.(Parallel

Greek text and English translation.)
Xenophon (*c*.370s BC) *Symposium*, trans. O.J. Todd, *The Banquet*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1922.(Parallel Greek text and English translation.)
Xenophon (*c*.360s BC) *Memorabilia*, trans. E.C. Marchant, *Memoirs*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1923.(Parallel Greek text and English translation.)
Zeller, E. (1875) *Die Philosophie der Griechen*, 2. Teil, 1. Abteilung, *Sokrates und die Sokratiker*, trans. O.J. Reichel, *Socrates and the Socratic Schools*, London: Longmans, Green, 1885.(Exhaustive scholarly discussion and interpretation, which is still useful; part 2, section 1 concerns Socrates.)

Socratic dialogues

After Socrates' death in 399 BC, a number of his followers composed imaginary dialogues between Socrates and various persons, usually historical. In addition to the dialogues of Plato there were works by Antisthenes, Aeschines, Phaedo, Euclides and, somewhat later, Xenophon. Only the writings of Plato and Xenophon have survived intact. The portrayal of Socrates varied from author to author, but the charismatic personality and skilful questioner is recognizable in each version. The connection between love and Socratic philosophy was frequently illustrated in Socrates' relationship to Alcibiades. The erotic theme was also represented in the role played by Aspasia, Pericles' mistress, in at least four authors, including Plato. Historical fact and even chronological possibility were regularly disregarded. This was essentially a genre of fiction.

1 The dialogue as a fictional genre

Socratic dialogues or 'Conversations with Socrates' (*Sōkratikoi logoi*) are recognized as a literary genre by Aristotle (*Poetics* 1), who is reported to have named a certain Alexamenos of Teos as the first author of such works. We know nothing further of Alexamenos, but we have remains from dialogues by Antisthenes, Aeschines, Phaedo and Euclides, in addition to the extant works of Plato and Xenophon. There must have been others, but the titles listed by the biographer Diogenes Laertius (II 12-16), of works by Crito, Simon the shoemaker, Glaucon, Simmias and Cebes, are probably all spurious.

The conversations are fictitious. Socrates' interlocutors are typically historical, but even here there are exceptions. The unwashed Pythagorean Telauges, after whom one of Aeschines' dialogues is named, is likely to be an imaginary stereotype. And in Plato, Ion as typical rhapsode and Callicles as the unscrupulous politician, ideologically 'liberated' by the Enlightenment, may well be fictitious characters. The contrast with history is clearest in the case of Xenophon, who wrote historical works as well as Socratic dialogues. When Xenophon in the *Memorabilia* or the *Oeconomicus* says 'I was there when Socrates said…,' no one is expected to believe him. He begins his *Symposium* with the remark that he was present at the occasion. But the fictive date is set in 421 BC, when Xenophon would have been a child, and Xenophon is in fact not mentioned again in the dialogue. His claim to have been present seems to be only a device for introducing the narrative form without specifying a narrator.

In Aeschines' *Aspasia*, the occasion itself is fantastic. When asked by Callias to recommend a moral tutor for his son, Socrates proposes Aspasia, Pericles' mistress and a symbol of loose living in Attic comedy. The episodes are all chronologically dubious. Thus Aspasia is represented as interrogating Xenophon and his wife, although Aspasia must have been long dead before Xenophon was old enough to get married. In Aeschines' *Alcibiades* we find Alcibiades bursting into tears and laying his head on Socrates' knees, which is no more likely to be historical than the bedroom scene reported by Alcibiades in Plato's *Symposium*. In Phaedo's dialogue *Zopyrus*, an oriental sage of that name arrives in Athens and offers to read Socrates' character from his appearance. The diagnosis is surprising: Socrates is described as a sensualist. Both the episode and Zopyrus himself seem to have been invented by Phaedo in order to permit Socrates to respond: 'Such are my native tendencies, but they have been overcome by philosophy.'

Plato is probably unique in creating the illusion of historical reality for his dialogues (although Xenophon has tried to imitate this in his own *Symposium*). Historians have been misled therefore into dating Parmenides and Zeno by their trip to Athens, imagined by Plato in the *Parmenides*. Here again it seems clear that the occasion was invented for dramatic and philosophical reasons, since in Plato's eyes only Parmenides deserved the honour of refuting Socrates and criticizing the doctrine of Forms. The fictional character of the genre is even more obvious in Plato's *Menexenus*. There Socrates delivers a funeral oration which alludes to events of 386 BC (thirteen years after his own death), but the speech is attributed to Aspasia, who made it 'from the scraps left over from the funeral oration she composed for Pericles in 431 BC (*Menexenus* 236b).

2 Literary interaction and common themes

By introducing Aspasia as author of funeral orations, Plato is developing a motif from Aeschines' *Aspasia*, where Aspasia is said to have taught rhetoric to Pericles and sharpened his tongue on the whetstone of Gorgias (fr. 22 Dittmar, A65 Giannantoni). We can trace this literary contact between Aeschines and Plato in several dialogues,

and in both directions. Literary interaction on a larger scale is found in the erotic themes that are common to all the known Socratic authors. Thus dialogues named *Alcibiades* and *Aspasia* were composed both by Antisthenes and by Aeschines. Euclides is also credited with an *Alcibiades*, and with an *Eroticus* as well (unless these are two names for the same work). And, as we have seen, Phaedo's *Zopyrus* dealt explicitly with Socrates' erotic temperament. As generally with Socrates, we cannot know just what historical reality underlies this well-defined Socratic literary theme. What we can see is that Plato's treatment of love (*erōs*) in the *Lysis* and *Symposium*, and the erotic allusions to Alcibiades in the *Gorgias* and *Protagoras*, are firmly rooted in the Socratic literary tradition. Diotima in the *Symposium* seems to be Plato's answer to Aeschines' Aspasia, who gives a Socratic lesson on virtue to Xenophon and his wife.

There are many examples of Platonic influence on Xenophon, who seems also to have absorbed material from Antisthenes and Aeschines. There is no trace of Xenophon's Socratic writings in Plato.

3 The historical Socrates?

Parallels between different Socratic authors may allow us to catch a glimpse of the historical Socrates. Thus there is an interesting parallel to the Platonic Socrates' avowal of ignorance in Aeschines' *Alcibiades*, where Socrates insists that it was not by any art (*technē*) or knowledge (*mathēma*) that he thought he could help Alcibiades in his quest for virtue, but only by love (fr. 11 Dittmar, A53 Giannantoni). The same dialogue depicts an elenchus, in which Socrates succeeds in bringing Alcibiades to a consciousness of his own ignorance and need for improvement - not by a systematic argument, however, but by presenting him with the model of Themistocles, in comparison to which Alcibiades feels immensely inferior. This confirms the suggestion given in several Platonic passages (for example, *Apology* 24-7 and *Laches* 187e-188b), that the original elenchus was more a testing of persons than of propositions.

4 External form

The earliest Socratic dialogues are likely to have been in simple mime form (without narration or dramatic proem) as in Plato's *Crito*, *Ion*, and *Hippias Minor*. Aeschines' *Alcibiades* is more complex: it is narrated by Socrates and located in a gymnasium, like Plato's *Charmides* and *Lysis*. Antisthenes may have introduced himself into the dialogue as narrator or interlocutor, as Xenophon does (and later Cicero); whereas Plato and (as far as we know) Aeschines never appear in their own works. From the literary point of view, Aeschines was probably the most gifted Socratic author after Plato. But this was originally a minor genre, comprising short works like the *Ion* and *Hippias Minor*. With large compositions like the *Gorgias* and *Protagoras* Plato transformed the Socratic dialogue into a major art form, the vehicle for his own philosophy.

Writing after Plato, Xenophon composed the *Memorabilia* as a large work by putting together many mini-dialogues. In form, Xenophon's *Symposium* is loosely modelled after Plato's work of that name, although the content is altogether different.

See also: Antisthenes §1; Socrates §§1-2

CHARLES H. KAHN

References and further reading

For the Socratic dialogues of Plato and Xenophon, see 'List of works' under Plato and Xenophon.

Aeschines (420-380 BC) Fragments in *Aischines von Sphettos*, ed. H. Dittmar, Berlin: Weidmann, 1912.(The basic study of Aeschines, with a full collection of his fragments, as known in 1912.)

Ehlers, B. (1966) *Eine vorplatonische Deutung des sokratischen Eros: Der Dialog Aspasia des Sokratikers Aischines (A pre-Platonic Interpretation of Socratic Love: The Dialogue Aspasia by Aeschines Socraticus)*, Munich: C.H. Beck'sche Verlagsbuchhandlung.(Convincing reconstruction of Aeschines' *Aspasia* and *Alcibiades*, in German.)

Giannantoni, G. (1991) *Socratis et Socraticorum Reliquiae (The Fragments of Socrates and the Socratics)*, Naples: Bibliopolis, 4 vols.(Volumes 1-2 include the fragments, in Greek or Latin, of the lost dialogues of the Socratics.)

Kahn, C.H. (1996) *Plato and the Socratic Dialogue*, Cambridge: Cambridge University Press.(Chapter 1 discusses Plato's relationship to the Socratic literary genre.)

Momigliano, A. (1971) *The Development of Greek Biography*, Cambridge, MA: Harvard University Press.(Best

discussion in English of the non-historical character of the Socratic literature.)
Rossetti, L. (1980) ’Ricerche sui dialoghi socratici di Fedone e di Euclide’ (Research on the Socratic Dialogues of Phaedrus and Euclides), *Hermes* 108: 183-200.(The fullest collection of evidence for the dialogues of Phaedo.)

Socratic schools

*For approximately one and a half centuries after Socrates' death in 399 BC, several Greek philosophical schools and sects each claimed to be the true intellectual heirs of Socrates. Later doxographers emphasized the Socratic pedigree of each of these schools by establishing an uninterrupted succession (*diadochē*) between its alleged founder, who was invariably a member of Socrates' own entourage, and the philosophers who succeeded him as leaders of the school.*

*Leaving aside Plato, the founder of the Academy, the members of the Socratic circle who left a succession behind them are Antisthenes, Aristippus of Cyrene, Euclides of Megara, and Phaedo of Elis, considered respectively the founders of Cynicism, and of the Cyrenaic, Megarian and Elian schools. It is these groupings, plus several of their offshoots, that are conventionally known as the '*Socratic schools*'. All can be seen as, in their own ways, developing Socrates' ethical outlook, and several were concerned with exploring the logical and metaphysical implications of his dialectical principles.*

1 Cynics, Cyrenaics, Megarians and Dialecticians

The Cynic movement appears intermittently for a period of about ten centuries, from the fourth century BC to the sixth century AD. Its archetypal figure was Diogenes of Sinope, whose bohemian lifestyle and beggarly appearance exemplified the unity of principle and practice governing early Cynicism. Although Antisthenes (§1) was looked back on as the founder of Cynicism, his connection with the movement is problematic. Lacking any institutional structure, it could not, in any formal sense, have a founder. Nor can he be considered the forerunner of the kind of philosophical instruction practised by the Cynics, since unlike Antisthenes they undermined curricular education, preferred practical example to philosophical argumentation and adopted informal means of instruction (see Cynics §1).

However, there is a sense in which the claims that Antisthenes is the predecessor of the movement and that the Cynics as a Socratic school are legitimate. Antisthenes is often considered the closest associate of Socrates, as reflected in both his logic and his ethical doctrine. In ethics, he shared the intellectualism of the Platonic Socrates, in that he considered virtue an understanding which, once acquired, amounts to wisdom and cannot be lost. But he also stressed the importance of physical and mental exercise, and the strength of character by which one overcomes one's weaknesses and achieves virtue. This moderated version of Socratic intellectualism provides a substantial common ground between the doctrine of Antisthenes and those of various Cynics.

The Cyrenaic school clearly counted as Socratic because founded by an associate of Socrates, Aristippus of Cyrene, whose own descendants succeeded him as the leaders of the school. Philosophically however, its doctrines have often been considered un-Socratic. For in ethics the Cyrenaics held various versions of hedonism, and several of them maintained that the bodily pleasure experienced at the present moment is the moral end. Again, in epistemology they developed a radical scepticism regarding our knowledge of the properties of external objects, whereas much of the evidence for Socrates suggests that, at least in the period in which Aristippus would have known him, he was concerned primarily with ethics. None the less, there were ways in which the Cyrenaics could reasonably claim that they remained faithful to the spirit of Socratic philosophy (see Cyrenaics §5; Aristippus the Elder).

The Socratic pedigree of the Megarian school is secured through Socrates' friend Euclides, who founded the school in his native Megara, although it is widely held to have been influenced also by the Eleatic philosopher Parmenides. Its doctrines were ethical and metaphysical, and also concerned philosophical methodology and logic. Euclides taught an ethical monism related to the doctrine of the unity of virtue held by the Platonic Socrates and, perhaps, to the belief of Xenophon's Socrates in a providential universe that, presumably, is wholly good (see Megarian school).

The Megarians overlapped for about fifty years with the rival school of the Dialecticians, linked with Socrates via its founder, Clinomachus of Thurii, himself a pupil of Euclides. Although this school's speciality was apparently the development of dialectical skills for their own sake, it was classified as one of the ten ethical sects which developed the ethical part of philosophy that originated with Socrates (see Dialectical school).

2 The Elian and Eretrian schools

The Elian school was founded by Socrates' associate, Phaedo of Elis, soon after Socrates' death. Its founder is recorded as the author of several Socratic dialogues, of which only the *Zopyrus* and *Simon* were certainly his own works (see Socratic Dialogues §1). The evidence suggests that both dialogues explored ethical subjects. The *Zopyrus*, named after a fifth-century physiognomist, probably aimed to modify the principle that there is an intrinsic relation between natural disposition and bodily form by arguing that the first can be entirely transformed - as in Socrates' own life - by the power of philosophy. *Simon* may have discussed various conceptions of virtue and its relation to pleasure, and perhaps defended a position according to which certain joys or pleasures are compatible with virtue. The themes linking the two works and securing a Socratic pedigree for Phaedo are the healing and reformative power of philosophy, its appropriateness for every person in every condition, the gradual and imperceptible effects of good and evil habits, and the importance of spiritual freedom with regard to external circumstances.

Phaedo's immediate successor was Plestanus of Elis, otherwise unknown. Anchipylus and Moschus, also from Elis, are listed as members of the Elian school, although Cynic features are attributed to them as well. But the most important philosophical heir of Phaedo was Menedemus of Eretria, after whom the school was relabelled the 'Eretrian school'. His criticisms of Plato, of Xenocrates, of the Cyrenaic Paraebates, of the Megarian Alexinus and of Aeschines suggest that he too was bidding for the mantle of Socrates. Even the testimony that he wrote nothing and did not adhere firmly to any doctrine may point to deliberate imitation of Socrates. However, a number of logical, metaphysical and ethical tenets are attributed to him. Reportedly, he accepted affirmative propositions and simple propositions but disallowed negative, complex or conditional statements. Although the point of this position is unclear, it may have been related to his beliefs that each thing can only be called by its own name, and that nothing must be at once one and many. These led him to remove the verb 'to be' from sentences such as 'that man is pale' and to remodel them with a periphrasis involving no use of the copula. In these respects, as well as in the commendation of tautologies, Menedemus' doctrine approaches Antisthenes' logic (see Antisthenes §4). Although he argued with great keenness and occasionally used paradoxes, he rejected Megarian eristics; on this account too Menedemus could claim to be faithful to the spirit of Socrates.

In ethics, Menedemus maintained that virtue is one thing called by many names, and probably implied that names which conventionally designate the different virtues are in fact synonyms - a point on which the Platonic Socrates is notoriously ambiguous. He espoused Euclides' position concerning the unity of the good (see §1) and he probably identified the good with virtue. He shared the intellectualism of other Socratics in that he placed the supreme moral good in the soul and believed that it can be achieved only by means of philosophical education which, therefore, is the single most important activity. But his bodily habits indicate that, like Antisthenes and the Cynics (see §1), he attributed moral importance to physical training as well. His proneness to superstition may suggest that he believed in a providential universe and, perhaps, in a divine creator. If so, his beliefs are comparable to those of Xenophon's Socrates and of Euclides. He took an active part in politics (a fact much resented by the Cynics), but similarly to many Socratics he kept his dignity, frankness and spiritual independence in the face of the powerful men of his day. His end, sadly, is reminiscent of Socrates' own: he was unjustly denounced for treason but, unlike Socrates, he left Eretria and died in exile by his own hand.

3 How Socratic are the Socratic schools?

Apart from the fact that the Socratic schools are founded by the entourage of Socrates, one should perhaps not attempt to find one single common thread unifying them. The links between them consist rather in family resemblances, between Socratics of the same generation or of different generations, many of them attributable with greater or lesser plausibility to Socrates himself. For example: Antisthenes and Aristippus stressed the importance of self-mastery and self-control regarding pleasure; Antisthenes and Phaedo considered some pleasures entirely compatible with virtue; Aristippus and the Cynics adopted the political attitudes of cosmopolitanism and of detachment from obligations to any particular city; Menedemus veered towards the ethics of the Cynics, and also held similar positions to Antisthenes in logic; the Megarians and the Dialecticians also worked on similar logical topics, such as modality, although their inquiries differed in scope and depth; Antisthenes, Stilpo, Menedemus and, perhaps the Dialecticians, subscribed to various forms of the doctrine that each thing has only one essence and that there is only one *logos* describing it (see Logos §1).

Some of the schools appear more formally organized than others. Aristippus and Phaedo probably ran proper schools, Antisthenes had a steady number of followers identified as ‘the Antisthenians’; the Cynics and the later Cyrenaic sects were dispersed; and Menedemus was entirely informal in his teaching. Nevertheless, the label ‘Socratic’ is to a certain extent justified. The doctrinal links both with each other and with the views attributed to the historical Socrates are symptomatic of the competition, widespread in the fourth century BC and after, to recover and expound Socrates’ authentic teachings.

See also: Plato; Socrates §8; Xenophon

VOULA TSOUNA

References and further reading

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press, 1925, 2 vols. (Greek text with English translation; see especially book II on the Socratics and book VI on Antisthenes and the Cynics.)

Giannantoni, G. (1990) *Socratis et Socraticorum Reliquiae (The Fragments of Socrates and the Socratics)*, Naples: Bibliopolis.(The standard collection of Greek and Latin testimonies on the Socratic schools.)

Vander Waerdt, P. (1994) *The Socratic Movement*, Ithaca, NY: Cornell University Press.(Collection of essays.)

Zeller, E. (1869-82) *Die Philosophie der Griechen (The Philosophy of the Greeks)*, 3rd edn, Leipzig: Reisland; relevant part available in English trans. by O.J. Reichel, *Socrates and the Socratic Schools*, Leipzig: Fues’ Verlag (Reisland), 3rd edn, 1988.(Still in most ways the best overall account.)

Solidarity

Solidarity exists among a group of people when they are committed to abiding by the outcome of some process of collective decision-making, or to promoting the wellbeing of other members of the group, perhaps at significant cost to themselves. Many regard solidarity as an important political ideal on the grounds that it is related to community and fraternity, and conducive to social cohesion and stability. Some individualists, however, believe that it is incompatible with autonomy on the grounds that full autonomy requires one always to take the final decision oneself about what one should do.

Solidarity is generally seen in two ways. First, as a commitment to other members of a group to abide by the outcome of their collective decision-making. Second, as a concern for other members of a group, which may require an unwillingness to receive a benefit unless the others do, or an unwillingness to receive a benefit when this will harm them. This commitment to the wellbeing of others is sometimes conceived in terms of the recognition of special obligations between the members of a group, which exist in virtue of their being members of it.

Solidarity has been of interest to social theorists concerned with the issue of what bonds unite people together in societies (see, for example, Durkheim 1983) (see Durkheim, É. §3). An understanding of the causes of solidarity can provide an insight into what makes societies stable, and what leads them to disintegrate. Because of its potential contribution to social stability and cohesiveness, solidarity has been regarded as instrumentally valuable.

Many political philosophers, particularly communitarians, have supposed that solidarity is also intrinsically valuable, believing that there is a conceptual connection between it and community or fraternity (see Community and communitarianism). For some communitarians, part of what it is to be a member of a community is to be in a reciprocal relationship of solidarity with the other members.

Socialist theorists have had special reasons for valuing solidarity because they have regarded it as a weapon against exploitation by capitalists, both in the struggle for better wages and conditions within the capitalist system, and in the fight against that system (see Socialism). Reformist socialists have thought that strikes are likely to be effective only if there is solidarity among the workers. Some revolutionary socialists have thought that capitalism will be overthrown only if there is solidarity among the working class in general.

The existence of solidarity can provide a way of out situations which have the structure of a ‘prisoner’s dilemma’ (see Rational choice theory §2). In the original prisoner’s dilemma, two prisoners who have committed a certain crime are interviewed separately and offered a deal. If only one confesses, that prisoner alone will be released, whereas the other will receive a harsh sentence. If both confess, each will receive an intermediate sentence. If both refuse to confess, they will get a milder sentence for some different crime which can be proved independently of either confessing. If each prisoner is concerned solely with their own interest, it is rational for them to confess, even though if neither confesses the outcome will be better for each. But if each is unwilling to receive a benefit when this will harm the other, then each will refuse to confess, and both will be better off.

Solidarity will not, however, provide a way out of every such situation, because in many-person cases, it is often the case that one’s own actions, even when motivated by a consideration of the good of others, will make no appreciable difference to the overall outcome. Consider, for example, the question of whether I should modify my car to reduce emissions from it. In such cases, it is not clear that solidarity will or should make any difference to my reasoning, except in the special case in which my community can be understood to have made some collective decision, perhaps to reduce emissions from cars, and solidarity is understood as a commitment to abide by such decisions. This kind of problem apparently besets Marxist analyses of how the revolution against capitalism will occur. One person’s contribution will make no difference to whether the revolution will occur, so what reason does any individual have to expend energy in attempting to bring it about, even if they have a sense of solidarity with other working-class people?

Not everyone values solidarity, however, and even those who do value it may disagree about when it is valuable. It might be argued that solidarity is valuable only when channelled in pursuit of morally permissible or worthwhile goals. More fundamentally, someone might doubt whether solidarity is ever valuable because they believe that it is

incompatible with personal autonomy (see Autonomy, ethical). R.P. Wolff (1970), for example, argues that being autonomous always involves taking the final decision about what one should do, and that it is incompatible with any form of commitment requiring that one should act in accordance with the outcome of collective decision-making. First, however, not all conceptions of autonomy require that one should always retain the final decision about how to act. Second, even if autonomy is valuable and is incompatible with solidarity, this does not show that solidarity has no value, for perhaps there can be genuine values that are incompatible with one another (see Moral pluralism). Third, a commitment to act in accordance with the outcome of a process of collective decision-making need not mean that one accepts that outcome uncritically. One might, for example, simply give the fact that it is such an outcome considerable weight in making the final decision about what one should do.

There are also disputes about the sort of relationships in which solidarity can be realized. Some maintain that solidarity is feasible only for small groups, whereas others believe that it is realizable among large groups. A society in which everyone willingly accepted and acted from what John Rawls (1971) has termed ‘the difference principle’ might be thought to embody solidarity (see Rawls, J. §§1-2). The difference principle holds that inequalities in a society are permissible only if they benefit (or at least, do not harm) the worst-off. If the best-off were willing to forego benefits when by accepting them they would harm the worst-off, then solidarity would exist between them. Richard Rorty (1989) has argued that it makes sense to aim at solidarity even at the global level (see Rorty, R. §3). In his view solidarity would require not a belief in a shared human essence, but rather an ability to see traditional differences (for instance, those between tribes, nations, religions and races) as unimportant compared to our shared capacity to experience pain and humiliation.

See also: Family, ethics and the; Friendship; Honour

ANDREW MASON

References and further reading

Benn, S.I. (1988) *A Theory of Freedom*, Cambridge: Cambridge University Press, ch. 12.(Traces the connections between notions such as comradeship, mutual concern, community and autonomy, and raises the question in what sort of relationships each of these can be realized.)

Durkheim, É. (1983) *De la division du travail social*, trans. W.D. Halls, *The Division of Labour in Society*, London: Macmillan.(Distinguishes different kinds of solidarity and examines their causes.)

Dworkin, G. (1988) *The Theory and Practice of Autonomy*, Cambridge: Cambridge University Press, ch. 1. (Criticizes conceptions of autonomy which are logically incompatible with commitment.)

Elster, J. (1985) *Making Sense of Marx*, Cambridge: Cambridge University Press, 6.2.(Raises the issue of how Marxists are to explain the emergence of revolutionary motivation.)

Rawls, J. (1971) *A Theory of Justice*, Oxford: Clarendon Press, 60-65, 76-80, 105-6.(Defends the ‘difference principle’ as an interpretation of fraternity.)

Rorty, R. (1989) *Contingency, Irony, and Solidarity*, Cambridge: Cambridge University Press, ch. 9.(Offers an account of how solidarity should be understood at the global level.)

Wolff, R.P. (1970) *In Defence of Anarchism*, New York: Harper & Row.(Defends the idea that full autonomy is incompatible with a commitment to abiding by collective decision-making.)

Solipsism

'Solipsism' (from the Latin solus ipse *- oneself alone) is the doctrine that only oneself exists. This formulation covers two doctrines, each of which has been called solipsism, namely (1) that one is the only self, the only centre of consciousness, and, more radically, (2) that nothing at all exists apart from one's own mind and mental states. These are not always distinguished from corresponding epistemic forms: for all we know, (1) or (2) might be true.*

A more recent coinage is 'methodological solipsism', which has a quite different meaning: that the content of an individual's thoughts is fully determined by facts about them, and is independent of facts about their environment.

Philosophical interest in solipsism does not arise from the fact that some significant philosopher has advocated it, for none has done so. Rather, solipsism has played dialectical roles, as for instance when philosophers try to argue that their opponents' positions, consistently pursued, lead to solipsism - and are thereby shown to be harbouring an absurdity. It plays another such role when philosophers ask whether there are any rational considerations capable of refuting it. More extreme forms of scepticism are sometimes treated in the same way - unsurprisingly, since the epistemic versions of solipsism are themselves extreme forms of scepticism.

That one is, for all one knows, the only centre of consciousness, is precisely the traditional problem of other minds (see Other minds). It can be based on the thoughts that one has no direct awareness of the mental states of others, and that to infer that they must have mental states because of the outward resemblance between their bodily behaviour and one's own is to make an inference based on one case only - not normally regarded as sound inductive procedure. Another, more radical, argument is that one is unable even to form any concept of a state of consciousness that is not one's own. This approach is discussed by Wittgenstein - who was not, however, recommending that we should accept its conclusion, but rather that we should abandon the natural, though mistaken, conception of a state of consciousness which leads us into it. On a right understanding of the matter, ascriptions of mental states to ourselves and to others are co-ordinate achievements. Solipsists wrongly think they can ascribe such states to themselves and *then* consider the question whether other human bodies have associated mental states as well (see Criteria; Private language argument).

The second type of solipsism, which finds problematic not just the existence of other minds but of *everything* other than one's own mind and mental states, is a very close relation of scepticism about the existence of the external world (see Scepticism). Although closely related, it is not the same thing, since a number of philosophers from widely varying traditions have denied the existence - or at least the independent existence - of a material world, while showing no tendency to assert that theirs was the only mind. Berkeley is a prime example, and Dharmakīrti, having portrayed the material world as illusory, wrote a treatise explaining why solipsism was not an inevitable consequence.

The expression 'methodological solipsism' is a recent coinage popularized by, in particular, Hilary Putnam and Jerry Fodor. This phrase refers to the view that the content of an individual's thoughts is fully determined by intrinsic, non-relational facts about that individual, hence is not affected by the nature of the objects that form their environment (see Content: wide and narrow; Methodological individualism).

EDWARD CRAIG

References and further reading

Dharmakīrti (mid 7th century) *Samtānāntarasiddhi (Proof of other continua)*; English trans. in Kitagawa, Hidenori, *Indo koten ronrigaku no kenkyū: Jinna (Dignāga) no taikei*, Tokyo: Suzuki Gakujutsu Zaidan, 2nd edn, 1973, Appendix A.(Dharmakirti's defence against the threat of solipsism. The 'continuum' of the Sanskrit title is a Buddhist word for mind in the sense of a continuous series of mental events.)

Russell, B. (1948) *Human Knowledge: its Scope and its Limits*, London: George Allen & Unwin.(Part III Chapter 2, 'Solipsism', introduces a radical version of solipsism and considers reasons against it.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(Contains an extended discussion of the idea that mental states are inner objects, awareness of which gives us our primary understanding of the mental - see approximately Sections 243-317; suggestive, but difficult.)

Soloveitchik, Joseph B. (1903-93)

Joseph B. Soloveitchik was a Jewish philosopher in the fullest sense. For such thinkers, the task of building intellectual and spiritual bridges between their particular traditions and other cultures permeates and shapes all their philosophic commitments and endeavours.

Medieval philosophers sought to integrate the competing knowledge claims of natural reason and authoritative revelation. Soloveitchik, by contrast, ignored metaphysics and epistemology, focusing instead on refuting the alleged incompatibility between Judaism and the active, human-centred ethos of modernity. His major concern was not the truth of religion but the relevance and significance of religious human types and ideals in modern Western culture.

1 *Halakhic* man as religious hero

Unlike Buber, who introduced traditional Judaism to modern Western culture by focusing on such universal themes as religious existentialism, pietism and mysticism, Soloveitchik introduced a new human archetype into the philosophic discussion: the *talmid hakham*, the Talmudic scholar, the hero of the eastern European (especially Lithuanian) Jewish intellectual tradition. Although such a figure seems *prima facie* to be the very obverse of the modern, triumphant man of action, Soloveitchik set out to reveal the implicit modern human spirit of the Talmudic scholar. The culture of Talmudic learning, Soloveitchik showed, was unique in its conceptual methodology and distinctive in its consuming devotion to study. Its values of intellectual rigour, integrity, boldness and creativity cultivated a human type that personified a special variety of human independence, autonomy and self-worth - the principal characteristics of the hero of modern Western culture.

Soloveitchik’s methodology, which he pursues in later writings, consists in constructing contrasting human types, in this case cognitive man versus *homo religiosus*. The former is an archetype of the modern scientist - this-worldly, action-oriented, confident, triumphant, at home in the world. *Homo religiosus*, on the other hand, is guilt-driven and grace-oriented, a spiritual manic-depressive who oscillates between a disgust for this imperfect world and a longing for the transcendent perfection of the divine reality.

Unlike his countertype, the man of natural religion, who is torn between self-negation and self-glorification, *halakhic* man possesses an objective normative way of life which fortifies him against the wild fluctuations of subjectivity, giving him a permanent sense of legitimacy and self-worth. The performance of *mitzvot* (commandments) entails a sense of covenantal responsibility which confirms the doer, as God’s, that is, the covenantal partner - with dignity and legitimacy.

In Soloveitchik’s conceptual scheme, not only law and the Sinaitic covenant but also the biblical narrative of creation itself constitutes a mandate enjoining human autonomy and dignity. Soloveitchik makes a bold exegetical move by extending the idea of *imitatio Dei* (which is applied to ethical attributes in rabbinic literature to include the attribute of God as creator). Creative human behaviour becomes metaphysically similar to divine creativity.

Spinoza had claimed that the Judaic legal tradition fostered a slave morality that inhibited the development of personal moral responsibility. Soloveitchik countered both this and Kant’s critique of Judaism for holding to what he deemed a heteronomous ethics by arguing that *halakhic* man appropriates the heteronomous, revealed norm as an autonomous, self-created norm (see Kant, I.). Soloveitchik himself creatively showed how modern existentialist typologies could find roots in *halakhah*.

2 ‘The Lonely Man of Faith’

In later works, Soloveitchik’s protagonist changes from the heroic *halakhic* man (in many ways a counterpart to the archetypal ideal of American Pragmatism) into a more dialectical, existentialist figure: the lonely man of faith. The empathy of Soloveitchik’s prose shift from the active, outward-directed virtues of creativity and autonomy, to the more introspective, radically personal experiences of loneliness, crisis and defeat.

The essay ‘The Lonely Man of Faith’ (1965) discusses two human types derived from the two creation stories in Genesis: ‘Adam the first’ is created in the image of God, together with Eve, and ordered by God to ‘fill the earth and master it’ (Genesis 1: 28); ‘Adam the second’ is fashioned alone from the dust of the earth by God, who

responds to his existential predicament ('It is not good for man to be alone' (Genesis 2: 18)) by creating another human being.

While this typological structure resembles the dichotomy between the cognitive-man and *homo-religiosus* of *Ish ha-Halakhah (Halakhic Man)* (1944), the superior human synthesis, now called covenantal man, is less weighted in favour of assertiveness and adequacy. Here we find the strong countervailing influence of existential loneliness and defeat.

Adam the first fulfils the mandate of the impersonal God of Being by pursuing science and technology and by gaining dominion over an impersonal nature. Adam the second responds to an inner, existential need by seeking relationship with another individual or indeed, with a personal God (Adonai, the Lord, the God who is named by the Tetragrammaton).

Soloveitchik argues that *halakhic* culture addressed the second Adam's existential quest through the covenant and the *mitzvot*. The individual can transcend temporality and personal death through participation in the normative history of the Jewish people. The existential need for relationship can be met by encountering a God who addresses one as thou and with whom one can share in a relationship of covenantal mutuality. The fundamental point of the covenantal faith community of Adam the second is that it seeks redemption through sacrifice, even in defeat and retreat. The underlying crisis calling for redemption is loneliness rather than sin or inadequacy.

The man of faith is far more complex than his heroic predecessor, *halakhic* man. Like the latter, he knows that achieving and expressing individual dignity and adequacy are integral constituents of the total faith gesture. Human majesty and creativity are forms of *imitatio Dei*. Faith thus requires that the man of faith be creative, free and assertive, as well as obedient, subservient and sacrificial. The God who commands the former also commands the latter (see Faith).

See also: Halakhah; Jewish philosophy, contemporary

D. HARTMAN

List of works

Soloveitchik, J. (1944) *Ish ha-Halakhah*, trans. L. Kaplan, *Halakhic Man*, Philadelphia, PA: Jewish Publications Society, 1983.(His early analysis of the notion of the *halakhic* individual as hero.)

Soloveitchik, J. (1944) *Ish ha-Halakhah Gahu ve-nistar*, trans. L. Kaplan, *The Halakhic Mind: An Essay on Jewish Tradition and Modern Thought*, New York: Free Press, 1986.(Analysis of the nature of Jewish law with its implications for human action.)

Soloveitchik, J. (1964) 'Confrontation', *Tradition* 6 (2): 5-29.(Discussion of modernity versus tradition.)

Soloveitchik, J. (1965) 'The Lonely Man of Faith', *Tradition* 7 (2): 5-67.(Crucial essay of his later period stressing existentialist strategies.)

Soloveitchik, J. (1979) *Reflections of the Rav*, ed. A.R. Besdin, Jerusalem: World Zionist Organization.(Collection of his papers.)

Soloveitchik, J. (1989) *Man of Faith in the Modern World: Reflections Of The Rav, II*, ed. A.R. Besdin, Hoboken, NJ: Ktav.(Adaptation and translation of lectures.)

References and further reading

Fox, M. (1989) 'The Unity and Structure of Rabbi Joseph B. Soloveitchik's Thought', *Tradition* 24 (2): 44-65.(An interesting but forced attempt at unifying all of Soloveitchik's thought.)

Hartman, D. (1985) *A Living Covenant*, New York: Free Press, 60-109, 150-60.(A critique of Soloveitchik's *halakhic* anthropology.)

Leaman, O. (1996) 'Jewish Existentialism: Rosenzweig, Buber and Soloveitchik', in D. Frank and O. Leaman (eds) *History of Jewish Philosophy*, London: Routledge, 799-819.(Examination of the links between Soloveitchik and other Jewish existentialists.)

Singer, D. and Sokol, M. (1982) 'Joseph Soloveitchik: Lonely Man of Faith', *Modern Judaism* 2: 227-72.(A psychological analysis of the inner tensions in Soloveitchik's thought.)

Spero, S. (1996) 'Rabbi Joseph Dov Soloveitchik and the Philosophy of Halakha', *Tradition* 30 (2): 41-64.(A critique of Soloveitchik's understanding of the relationship of nature and *halakhah*.)

Solov'ëv, Vladimir Sergeevich (1853-1900)

It has been widely acknowledged that Vladimir Solov'ëv is the greatest Russian philosopher of the nineteenth century; his significance for Russian philosophy is often compared to the significance of Aleksandr Pushkin for Russian poetry. His first works marked the beginning of the revolt against positivism in Russian thought, followed by a revival of metaphysical idealism and culminating in the so-called Religious-Philosophical Renaissance of the early twentieth century.

Unlike the Russian idealists of the Romantic epoch, Solov'ëv was a professional, systematic philosopher. He created the first all-round philosophical system in Russia and thus inaugurated the transition to the construction of systems in Russian philosophical thought. At the same time he remained faithful to the Russian intellectual tradition of reluctance to engage in purely theoretical problems; his ideal of 'integrality' postulated that theoretical philosophy be organically linked to religion and social practice. He saw himself not as an academic philosopher, but rather as a prophet, discovering the way to universal regeneration.

One of the main themes of Solov'ëv's philosophy of history was Russia's mission in universal history. Owing to this he was interested in the ideas of the Slavophiles and, in the first period of his intellectual evolution, established close relations with the Slavophile and Pan-Slavic circle of Ivan Aksakov. He was close also to Dostoevskii, on whom he made a very deep impression. At the beginning of the 1880s he began to dissociate himself from the epigones of Slavophilism; his final break with them came in 1883, when he became a contributor to the liberal and Westernizing Vestnik Evropy (European Messenger). The main reason for this was the pro-Catholic tendency of his thought, which led him to believe that Russia had to acknowledge the primacy of the Pope. In his view, this was a necessary condition of fulfilling Russia's universal mission, defined as the unification of the Christian Churches and the establishment of a theocratic Kingdom of God on earth.

In the early 1890s Solov'ëv abandoned this utopian vision and concentrated on working out an autonomous ethic and a liberal philosophy of law. This reflected his optimistic faith in liberal progress and his confidence that even the secularization of ethics was essentially a part of the divine-human process of salvation. In the last year of his life, however, historiosophical optimism gave way to a pessimistic apocalypticism, as expressed in his philosophical dialogue Tri razgovora (Three Conversations) (1900), and especially the 'Tale of the Antichrist' appended to it.

1 The philosophy of All-Unity

Vladimir Solov'ëv was the son of Sergei Solov'ëv, a leading Westernizing historian and professor at Moscow University. After the brilliant defence of his master's thesis, *Krizis zapadnoi filosofii: protiv pozitivistov (The Crisis of Western Philosophy: Against the Positivists)* (1874), he began lecturing at St Petersburg University. His academic career, however, proved to be short-lived. The assassination of Alexander II (March 1881) prompted him to give a public lecture in which, while condemning the revolutionary terrorists, he appealed to the new tsar to spare their lives. As a result he was forbidden to lecture in public and shortly afterwards resigned from the university.

The general outline of Solov'ëv's philosophy follows the Neoplatonic scheme of self-enriching alienation. Conceiving the world as the divine absolute in the process of becoming, he distinguished in it three moments, corresponding to the three persons of the Holy Trinity: the moment of a static, undifferentiated unity, the moment of individuation and differentiation and, finally, the highest moment of a free, differentiated unity. In his *Filosofskie nachala tselnogo znaniya (Philosophical Principles of Integral Knowledge)* (1877) and in his doctoral dissertation, *Kritika otvlechënnych nachal (The Critique of Abstract Principles)* (1880), Solov'ëv applied this scheme to human history and knowledge. In the evolution of humankind, he argued, the first phase (substantial monism) was represented by the Eastern world (including nineteenth-century Islam), and the second phase (atomism) by Western European civilization. During the first phase the three spheres of human activity - creativity, knowledge and social practice - were entirely subordinated to religion. In the sphere of creativity, technology (the material grade) was fused with art (the formal grade) and mysticism (the highest, absolute grade) in a mystical *theurgy*. In the sphere of knowledge, positive science (the material grade) was fused with abstract philosophy (the formal grade) in an undifferentiated *theosophy*. In the sphere of social practice, the economic self-government or

zemstvo (the material grade) was fused with the state (the formal grade) and the Church (the absolute grade) in a homogeneous *theocracy*. In the second evolutionary phase the different grades within each sphere strove for absolute autonomy and mastery over one another. In the resulting struggle matter conquered spirit: the final outcome of Western civilization was materialistic socialism (the true scion of capitalism) in the social sphere, positivism in the sphere of knowledge, and utilitarian realism in the sphere of creativity. Now the time had come for the third phase - that of *free unity* - in which the separate spheres and grades of human creativity, knowledge and social practice would once more be united, but without losing their distinctiveness: as a free theurgy, a free theosophy and a free theocracy. The mission of inaugurating this universal regeneration and reintegration of human life was to be fulfilled by the Russians.

In the sphere of knowledge Solov’ëv distinguished three types of philosophy - naturalism (empiricism), rationalism and mysticism. Empiricism and rationalism took different paths to arrive at the same result - the denial of the substantial reality of both the external world and the knower himself. Following the elderly Schelling and the Slavophiles, Solov’ëv saw Hegelianism as the last word in rationalism (see Schelling, F.W.J.; Slavophilism). He accused Hegel of creating a merely ’negative’ philosophy, dealing only with empty concepts, and, like Schelling, set against this a programmatically ’positive’ philosophy, rehabilitating ’the given’. It involved a rehabilitation of nature, as a manifestation of the divine absolute, and, above all, a philosophical interpretation of the Christian revelation.

The central idea in Solov’ëv’s system was that of positive ’All-Unity’. He also used Kireevskii’s concept of ’integral wholeness’ (see Slavophilism §1) but interpreted it differently, stressing that free unity presupposes the differentiation of the original wholeness and autonomization of different spheres of life and knowledge. Therefore, Western civilization, including Western philosophy, was for him not a ‘wrong path’ (as it was for the Slavophiles), but a necessary phase of human development.

2 Godmanhood and Sophia

Solov’ëv’s reinterpretation of Christianity revolves around the concept of Godmanhood, developed by him in a series of extremely successful lectures in St Petersburg *Chteniia o bogochelovechestve (Lectures on Godmanhood)* (1877-81). The concept of ’God made man’, Solov’ëv asserted, does not assume either a dualistic belief in the transcendence of God, or a pantheistic belief in his immanence in the world. God is both immanent and transcendent, and the mediating principle that allows the world to become transfused by the divine spirit is Man. The ultimate purpose of the universe is to achieve the stage of a free, innerly differentiated All-Unity. The cosmic history began with a dramatic event: the soul of the world, or Sophia, had once fallen away from the divine Logos, plunging the world into a chaotic struggle of hostile elements. The second act of the cosmic drama was the appearance of man. Since then Sophia, having identified itself with ’ideal humanity’, began to ascend to a renewed unity with God in Godmanhood. The union of Logos with Sophia, of God with man, was accomplished in Jesus Christ. The incarnation of God in Jesus was thus the central event in the entire cosmic process. Even this, however, was only the beginning of the soteriological process. To finish the work of salvation, that is, to achieve the universal regeneration and transfiguration of the world, it was necessary to pass through all stages of human history and to realize the destiny of man in the Kingdom of God on earth.

The concept of Sophia in Solov’ëv’s philosophy is somewhat ambiguous and subject to different interpretations. In some of his writings Solov’ëv defined Sophia, or the Wisdom of God, as ’the substance of the Holy Ghost’, the passive, female element in God, ‘external Womanhood’. In the autobiographical poem ’Three Meetings’ he identified Sophia with the mysterious feminine being which visited him in his mystic visions. He was thoroughly acquainted with a vast mystical and theosophical literature on Sophia: in the Bible (Solomon’s Proverbs), in the Kabbalah (where Sophia is presented as a woman) (see Kabbalah), in the writings of the gnostics (where it is ‘divine matter’ and the ’feminine aeon’) (see Gnosticism §4), in Jakob Boehme (who identified it with ’eternal virginity’), in Emanuel Swedenborg and Louis-Claude Saint-Martin. He thought that the idea of Sophia was particularly close to the traditions of Eastern Christianity, and saw a proof of this in an old icon with the image of Sophia in the cathedral of Novgorod. In the most general terms it may be said that he needed the concept of Sophia to define divinity within the world, unity in the creation, and the link between the world and the transcendent God.

As ’the feminine principle in God’ Sophia was bound up with Solov’ëv’s theory of sexual love, expounded in *Smysl liubvi (The Meaning of Love)* (1892-4). It claimed that the real significance of love consists not in

procreation, which is merely the absurd multiplication of mortal beings, but in the human striving for the likeness of God in a restored androgynous unity.

Solov’ëv’s vision of universal reintegration can also be found in his aesthetics. As a motto for his essay ‘Beauty in Nature’ (1889) Solov’ëv chose Dostoevskii’s words ’Beauty will save the world’. Natural beauty, he maintained, is a manifestation of the concrete operations of the Absolute in the material world; by ’transilluminating’ and spiritualizing matter, beauty helps to raise the fallen World Soul and to introduce an element of the divine into reality. In the essay ’The Overall Meaning of Art’ (1890) Solov’ëv applied this conception to artistic beauty: art, he argued, is an instrument of reintegration, a theurgic force capable of ’transilluminating’ and transforming the human world.

3 Theocratic utopia

Russia’s mission in universal history was defined by Solov’ëv as the realization of a ’free theocracy’ - a social system based upon Christian principles and thus establishing the Kingdom of God on earth. Solov’ëv’s Slavophile friends warmly sympathized with this vision but remained sceptical about his ecumenical ideas. When Solov’ëv arrived at the conclusion that Russia, in order to fulfil her mission, should recognize the authority of the Pope, his break with the Slavophile circles became unavoidable.

To avoid censorship, Solov’ëv’s vision of ’free theocracy’ had to be expounded in books published abroad, such as *Istoriia i budushchnost’ teokratii (The History and the Future of Theocracy)* (Zagreb, 1887), *L’Idée Russe (The Russian Idea)* (Paris, 1888) and *La Russie et l’Église Universelle (Russia and the Universal Church)* (London 1889). In Russia these books were received as Roman Catholic propaganda and consequently often aroused hostility. This was not just a misunderstanding; in 1886 Solov’ëv recognized the Pope as supreme judge in matters of religion, and three years later, in *Russia and the Universal Church*, publicly defended this view. It was known, as well, that his vision was supported by the influential Croatian bishop, Josip Strossmayer (who commended Solov’ëv’s ideas to Pope Leo XIII). On the other hand, however, it needs to be stressed that the French Jesuits ceased to support his project on the grounds of its heterodox ideas (especially in the theosophical part of his views).

Solov’ëv’s conception of ’free theocracy’ rested on his view that the structure of legitimate authority had to be triune, corresponding to the three persons of the Holy Trinity. The Kingdom of God on earth, he argued, will be established when humanity becomes unified under the authority of one high priest (representing God the Father) and one emperor (representing the Son), and when genuine prophets, ’the free breath of the Holy Ghost’, appear constantly among the common people to mediate between the temporal and spiritual authorities. This vision was to materialize in the unification of the human race under the spiritual rule of the Pope and the secular rule of the Russian tsar. In this way Russia would take up the old idea of a Christian Empire, formulated by Constantine the Great, Charlemagne and the greatest Catholic writer, Dante Alighieri. In the realization of this task Russia would be supported by the two ’theocratic nations’ which shared in her historical fate: the Jews and the Poles.

Solov’ëv did not proclaim that members of the Orthodox Churches should simply convert to Roman Catholicism. On the contrary: he combined sharp criticism of religious particularism (or even the nationalism of the Greek and Russian clergy), with an emphasis on the unique spiritual values of Eastern Christianity. The Roman Pope was, in his view, the legitimate head of the Universal Church, but this did not mean equating the Universal Church with the Roman Catholic Church. The Universal Church was a broader notion, embracing both Western and Eastern Christianity. Solov’ëv concluded from this that the recognition of the authority of the Pope would liberate the Orthodox Church from humiliating dependence on the imperial power without entailing any danger for the existence of the distinctively Eastern religious culture.

Solov’ëv saw no contradiction between his theocratic ideal and his growing sympathy with liberal political ideas. He stressed that his theocracy was to be ’free’, hence liberal and universalist, and thus incompatible with nationalism (which he distinguished from patriotism). In a series of articles published in the liberal *Vestnik Evropy (European Messenger)* and reprinted in the book entitled *Natsional’nyi vopros Rossii (The Problem of Nationalities in Russia)* (two volumes, 1891), he firmly condemned all forms of persecution and discrimination against national and religious minorities.

At the beginning of the 1890s Solov’ëv became disillusioned with theocratic utopianism and as a result drew even

closer to the liberals. But he remained faithful to the view that, in order to resist the evils of national egoism, Russia should embrace liberal principles without abandoning her imperial calling.

4 Ethics, philosophy of law and theory of knowledge

The de-utopianization of Solov'ëv's thought brought about a marked secularization of his views on ethics. His comprehensive treatise *Opravdanie dobra (The Justification of the Good)* (1897) was to be a system of independent ethics, in the sense of the independence of ethical conduct from metaphysical theories and religious dogmas. It is divided into three parts: (1) The Good in Human Nature, (2) The Good from God and (3) The Good in the History of Humankind. In Part I Solov'ëv tried to give his ethic empirical foundations by deriving it from feelings of shame, compassion, and religious adoration; he stressed, however, that empirical ethics must be supplemented by rational ethics whose highest principle is the categorical imperative of Kant (see Kant, I. §9). In Part II he dealt with moral progress, explaining it as the development of man's divine features, or the realization of the idea of Godmanhood. The two parts of this process were (1) the realization of Godmanhood in an individual, and (2) the realization of Godmanhood in the collective life of humankind.

With this conclusion Solov'ëv passed to Part III of his ethical system - to 'objective ethics', or the institutionalized forms of the historical education of humanity, as opposed to 'subjective ethics', concentrating on purely individual perfection and salvation. The emphasis on objective ethics, embodied in the institutions of the state, was also an essential feature of his theocratic ideal. But the difference was profound: according to Solov'ëv's objective ethics Church and state were to be formally separated and the task of securing the moral character of social and political life was to be fulfilled not by a theocratic government, but by an impersonal system of legal rules. In this way the realization of the idea of Godmanhood in history was made dependent on man's maturity, on his full moral autonomy, incompatible with any form of tutelage in the spiritual sphere.

This Kantian reinterpretation of the idea of Godmanhood found expression in Solov'ëv's philosophy of law, elaborated in the relevant chapters of *The Justification of the Good* and in a separate booklet *Law and Morality* (1897). This legal philosophy, defining law as the compulsory and enforceable minimum morality, was the earliest Russian variant of a 'new liberalism' concerned not only with negative freedom, but with positive freedom as well. In Solov'ëv's conception, the modern rule-of-law state was given a new function: that of ensuring for everyone a minimum of positive freedom and justice, that is equal opportunity and the minimum of welfare necessary for a 'dignified existence'. Solov'ëv saw 'dignified existence' as a new human right, as something legally claimable, and interpreted it very broadly. In his view, 'the right to a dignified existence' included not only rights to employment, to proper working conditions and proper earnings, but also to rest and free time. He anticipated modern ecological movements by postulating that enterprising individuals should not be given a free hand in conquering and exploiting nature, because the preservation of the natural environment is a necessary condition of a worthwhile existence.

Solov'ëv made it clear that even the best realization of this ideal would not have amounted to the millenarian Kingdom of God. He stressed that he did not want to legislate absolute morality but only to give legal backing to a moral minimum.

Another consequence of the de-utopianization of Solov'ëv's views was his increased willingness to treat interest in theoretical questions as autonomous, independent of social and religious concerns. This shift in his thought is shown in his three epistemological articles under the title *Osnovy teoreticheskoi filosofii (Foundations of Theoretical Philosophy)* (1897-9). His reflections on epistemological issues led Solov'ëv to reject the Cartesian tradition of 'spiritualist dogmatism' that assumed the substantiality of the knower. He also substantially modified his conception of 'integral knowledge' by de-emphasizing its mystical character and stressing instead 'purposeful cognition', or a 'vital act of decision' that directs the consciousness towards absolute truth and transforms immediate sense impressions into the material of the complex process of active cognition. To some extent the concept of 'purposeful cognition' recalls the phenomenological notion of 'intention' and the 'intentional act'. Unfortunately, further elaboration of these ideas was prevented by Solov'ëv's premature death.

5 Solov'ëv's continuators

During his lifetime Solov'ëv won recognition in the milieu of professional philosophers grouped around the

periodical *Problems of Philosophy and Psychology*. He had also two direct and very talented disciples, brothers Sergei and Evgenii Trubetskoi. His influence, however, only became widespread, reaching far beyond professional philosophers, in the first decade of the twentieth century, due to the powerful revival of the interest in religious ideas in Russian thought. It is no exaggeration to say that an entire generation of Russian idealist philosophers and religious thinkers was schooled in his philosophy. Sergei Bulgakov, Semën Frank, Father Pavel Florenskii and Lev Karsavin based their work on Solov'ëv's metaphysics of All-Unity and developed different variants of 'Sophiology'. Prince Sergei Trubetskoi made use of Solov'ëv's ideas in his polemics against 'epistemological individualism', in which he referred to the concept of a supra-individual consciousness identified by him with Sophia. The intuitivist Nicholas Lossky tried to use Solov'ëv's ideas in a different way, combining his notion of All-Unity with a monadistic pluralism. Messianic and eschatological motifs in Solov'ëv's thought exercised a strong, formative influence on Nikolai Berdiaev and on the critic and novelist Dmitrii Merezhkovskii who formulated the ideas of a 'new religious consciousness' (see Russian religious-philosophical renaissance §3). Solov'ëv's seminal influence can also be traced in the work of the Russian symbolist poets - Viacheslav Ivanov, Andrei Belyi, and Aleksandr Blok (see Russian religious-philosophical renaissance §4). Solov'ëv's conception of the 'right to a dignified existence' was taken up and developed by Pavel Novgorodtsev, the leading representative of the neo-idealist current in Russian philosophy of law, closely bound up with Russian liberalism (see Liberalism, Russian §3). Another philosopher of Russian 'new liberalism', Sergei Hessen, elaborated a Neo-Kantian reinterpretation of Solov'ëv's philosophy of All-Unity, stressing the idea of the autonomy of different spheres of culture and setting it against utopian motifs in Solov'ëv's thought.

In the Soviet Union interest in Solov'ëv was deliberately suppressed. The first Soviet edition of Solov'ëv's selected works appeared only under Gorbachev's *perestroika*, in 1988. In the West interest in Solov'ëv is still limited but constantly growing. He is best known in Germany, where his selected works were first published in 1914-17 (Jena, three volumes) and 1921-2 (Stuttgart, four volumes), paving the way for the ambitious undertaking of a German edition of his collected works, started in 1953. In France Solov'ëv is known mostly in Catholic circles interested in ecumenical ideas. At the end of 1991 a transnational 'Vladimir Solovyov Society' (with offices in Norwich, Vermont, and in Moscow) was called into being at a meeting in Moscow between American and Russian scholars.

ANDRZEJ WALICKI

List of works

Solov'ëv, V.S. (1966-70) *Sobranie sochinenii (Collected Works)*, phototype edn, Brussels, 12 vols.(Solov'ëv's collected works in Russian.)

[Solowjew. W.S.] (1977-9) *Deutsche Gesamtausgabe der Werke von Wladimir Solowjew (German edn of Solov'ëv's Collected Works)*, ed. W. Szylkarski, W. Lettenbauer and L. Müller, Munich: E. Wewel, 8 vols.

[Solowjew. W.S.] (1877-81) *Chteniia o bogochelovechestve*; trans. P. Zouboff, *Lectures on Godmanhood*, London: Dennis Dobson, 1948.

[Solowjew. W.S.] (1889) *La Russie et l'Église Universelle*, Paris, trans. H. Rees, *Russia and the Universal Church*, London: The Centenary Press, 1948.

[Solowjew. W.S.] (1892-4) *Smysl lyubvi*; revised trans. T.R. Beyer (ed.), *The Meaning of Love*, West Stockbridge, MA: Lindisfarne Press, 1985.

[Solowjew. W.S.] (1897) *Opravdanie dobra: Nravstvennaya Filosofiya*, Moscow; trans. N. Duddington, *The Justification of the Good: An Essay in Moral Philosophy*, London: Constable, 1918.

[Solowjew. W.S.] (1937) *God, Man and the Church: The Spiritual Foundations of Life*, trans. D. Attwater, Greenwood, SC: Attic Press, 1974.

[Solowjew. W.S.] (1978) *La Sophia et les autres écrits Français (Sophia and Other Works in French)*, Lausanne: Editions l'Age d'Homme. (Solov'ëv's works in French published by Father F. Rouleau.)

[Solowjew. W.S.] (1988) *Sochinenia (Works)*, Moscow: Mys'l, 2 vols.(Soviet edition selected from his collected works (1966-70), with the introduction by A.V. Gulyga.)

References and further reading

Dahm, H. (1975) *Vladimir Solovyev and Max Scheller: Attempt at a Comparative Interpretation*, Dordrecht and Boston, MA: Reidel. (A good example of the German contribution to Solov'ëv scholarship. Contains a bibliography.)

Losev, A. (1990) *Vladimir Solov'ëv i ego vremia (Vladimir Solov'ëv and his Times)*, Moscow: Progress.(The best work on Solov'ëv produced in the Soviet Union.)
Mochulsky, K. (1951) *Vladimir Solov'ëv: Zhizn' i uchenie (Vladimir Solov'ëv: Life and Teaching)*, Paris: YMCA-Press.(A standard monograph of Solov'ëv by a Russian émigré scholar.)
Solov'ëv, S.M. (1977) *Zhizn' i tvorcheskaia evoliutsiia Vladimira Solov'ëva (Life and Intellectual Evolution of Vladimir Solov'ëv)*, Brussels: Foyer Oriental Chretien.(A presentation of Solov'ëv's religious philosophy from a Catholic standpoint. The work of his nephew who had converted to Catholicism and became a priest.)
Stremoukhoff, D. (1979) *Vladimir Soloviev and His Messianic work*, Belmont: Nordland.(Translated from the French edition of 1935, this is the most comprehensive monograph in English.)
Trubetskoi, E.N. (1913) *Mirosozertsanie Vl. S. Solov'ëva (The Worldview of V.S. Solov'ëv)*, Moscow: Put', 2 vols; new edn, Moscow: Medium, 1995.(The classic and most comprehensive work on Solov'ëv.)
Walicki, A. (1987) *Legal Philosophies of Russian Liberalism*, Oxford: Clarendon Press.(Solov'ëv's philosophy of law is analysed in detail in ch. 3.)
Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, vol. 2; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, London: Routledge & Kegan Paul and New York: Columbia University Press, 1953, vol. 2, chs 16 and 17.(A thoughtful discussion of Solov' ëv's philosophy from an Orthodox Christian point of view.)

Sophists

The Sophists were itinerant educators, the first professors of higher learning, who appeared in Greece in the middle and later fifth century BC*. The earliest seems to have been Protagoras, who was personally associated with the statesman Pericles. The next most eminent was Gorgias, an influential author and prose stylist. The Sophists succeeded in earning very large sums for their instruction. They lectured on many subjects, including the new natural philosophy, but their most important teaching was in rhetoric, the art of influencing political assemblies and law courts by persuasive speech. In conservative circles their great influence was regarded with hostility, as corrupting the young.*

The term *sophistēs* is an agent noun from the Greek verb *sophizesthai*, ‘to be skilful’ (*sophos*). So a *sophistēs* was originally an expert in any form of wisdom or skill (*sophia*). The term was thus applied to poets, philosophers and sages of all sorts. But by the late fifth century BC it had received a special application to a new class of men: the professional educators, known to Plato and posterity as the ‘Sophists’.

According to Plato, Protagoras (§2) was the first to identify himself as a Sophist and to offer training in public affairs for a fee. He professed to teach men ’to manage their households well and to administer the business of the city’ (*Protagoras* 318e). For this, Protagoras and the other Sophists could expect to be paid handsomely. Their instruction was particularly attractive to ambitious young men in pursuit of a political career. Skill in public speaking was of supreme importance, above all in Athens, where political decisions were made in public assemblies and in very large law courts. Some Sophists, such as Gorgias, claimed only to teach rhetoric, that is, skill in public speaking. Typically, however, the goal of Sophistic training was conceived in terms of civic virtue or excellence (see Aretē). Hence the urgency of the question, much debated in the dialogues of Plato (§11), whether virtue can be taught.

The Sophists were travelling professors who founded no permanent schools. They came from all parts of Greece but were frequently active in Athens, the richest and most powerful Greek city of the time. They were often public figures: Gorgias, Hippias and Prodicus came to Athens as political representatives of their native cities, and Protagoras wrote the lawcode for the Periclean colony of Thurii. Their closest Athenian colleague was Socrates, who was popularly regarded as a Sophist, although he insisted that he was not a teacher and took no fees. But Socrates belonged to the same intellectual circles and was believed to exercise a similar influence on the young (see Socrates §1,2,7).

The content of Sophistic education was very diverse. Although the principal subject was political eloquence, training also included the study of Homer and the other poets as presenting traditional paradigms of excellence. Rhetoric was taught by example, and by memorizing and imitating such display pieces as the *Encomium of Helen of Troy* and the *Defence of Palamedes* by Gorgias, and the ’Choice of Heracles’ by Prodicus. Some Sophists also taught mathematics, astronomy and natural philosophy. All of them were acquainted with, and served to propagate and popularize, the new naturalistic worldview that had begun to replace traditional religious and mythopoetic ways of thought. The new outlook had begun in sixth-century Miletus and been developed by Xenophanes, Heraclitus and Parmenides, as well as by fifth-century cosmologists such as Anaxagoras and Empedocles. This is the world view of the fifth-century Enlightenment, whose influence can be seen in Euripides, Thucydides and many of the Hippocratic treatises (see Hippocratic medicine).

The Sophists played a key role not only in the transmission but also in the development of this new rationalism in the second half of the fifth century. The theological scepticism and epistemological relativism expressed in Protagoras’ thesis that ’man is the measure of all things’ mark a radical break with the traditional Greek worldview centred on the gods. And the psychological cynicism displayed in Gorgias’ defence of Helen of Troy implies a similar alienation from traditional Greek morality. All this lies behind Aristophanes’ hostile caricature of the new education in his comedy the *Clouds*, in the figure of the Unjust Argument who defends adultery and self-indulgence as ’necessities of nature’ and urges his auditor to ’follow nature, frolic, laugh, consider nothing shameful’ (*Clouds* 1075-8).

Aristophanes is alluding here not only to Gorgias’ amoral defence of Helen’s adultery, but more generally to the

antithesis between nature and convention (*physis* and *nomos*) that was made popular in Sophistic rhetoric and used by some of the Sophists to undermine the authority of the traditional moral code, treated now as a mere human convention (see Antiphon; Callicles; Physis and nomos). So in Aristophanes' *Clouds* the pupil of the Unjust Argument will learn 'to despise the established norms (*nomoi*)' and to treat the rule against father-beating as a matter of convention that can be changed by persuasive speech. All of this is presented by Aristophanes as the fruit of the Sophistic enterprise of making the weaker argument the stronger.

Aristophanes is a representative of the popular, conservative reaction against the new rationalism, which was seen as a threat to traditional values. In this regard Socrates and the Sophists were tarred with the same brush, and in the *Clouds* Socrates is in fact presented as the principal Sophist. It was Plato who succeeded in distinguishing the philosophers from the Sophists and thus separating Socrates from the professional educators, despite the fact that both belonged to the new Enlightenment tradition. For Plato, the sophist is a false image or imitation of the philosopher. For Aristotle, the sophist is a specialist in invalid but persuasive argument. It is this Platonic-Aristotelian conception of sophistry that has been transmitted to the present day.

Since the Sophists founded no schools, their influence extended beyond their own lifetime only through their writings and the work of their pupils, notably the rhetoricians Isocrates and Alcidamas, who had studied with Gorgias. Among the followers of Socrates, Antisthenes §1 was also heavily influenced by Sophistic modes of thought, and he too is alleged to have been a pupil of Gorgias, presumably before he attached himself to Socrates.

See also: Dissoi logoi

CHARLES H. KAHN

References and further reading

Aristophanes (*c*.455-386 BC) *Clouds*, trans. W. Arrowsmith, Ann Arbor, MI: University of Michigan Press, 1962. (Comic presentation of Socrates as a typical Sophist, headmaster of a Thinking-School.)

Classen, C.J. (ed.) (1976) *Sophistik*, Darmstadt: Wissenschaftliche Buchgesellschaft.(An anthology of scholarly articles, some in English; includes extensive bibliography.)

Diels, H. and Kranz, W. (1952) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, vol. 2, 252-416.(The standard collection of the ancient sources; includes Greek texts with translations in German.)

Dodds, E.R. (1956) 'Rationalism and Reaction in the Classical Age', in *The Greeks and the Irrational*, Berkeley and Los Angeles, CA: University of California Press, ch. 6.(A brilliant study of the moral and cultural context of the Sophistic movement.)

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, vol. 3, Cambridge: Cambridge University Press; part of vol. 3 repr. as *The Sophists*, Cambridge: Cambridge University Press, 1971.(A full, scholarly account.)

Kerferd, G.B. (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press.(A briefer, more personal interpretation than Guthrie (1969).)

Sprague, R.K. (ed.) (1972) *The Older Sophists*, Columbia, SC: University of South Carolina Press.(Full English translation of the fragments and testimonia from Diels and Kranz (1952).)

Sorel, Georges (1847-1922)

The French social theorist Georges Sorel is best known for his controversial work Réflexions sur la violence (Reflections on Violence), first published in 1908. He here argued that the world could be saved from 'barbarism' through acts of proletarian violence, most notably the general strike. This, he believed, would not only establish an ethic of the producers but would also serve to secure the economic foundations of socialism. Moreover the inspiration for these heroic deeds would be derived from a series of 'myths' that encapsulated the highest aspirations of the working class. More broadly Sorel should be seen as an innovator in Marxist theory and the methodology of the social sciences.

Born in Cherbourg,Sorel was educated at the prestigious *École Polytechnique* in Paris before gaining employment as a government engineer. Remaining in this position until 1893, it was not until the late 1880s that Sorel's writings began to appear. Upon retirement he took up residence in the Paris suburb of Boulogne-sur-Seine and thence became a central figure in the intellectual world of the Parisian Left Bank. Associated with the intriguing figure of Charles Péguy, Sorel quickly established a wide range of international correspondents (for example, Benedetto Croce and Antonio Labriola) and over a period of thirty years was intimately associated with an extensive series of reviews that did much to introduce new ideas into France.

Fascinated by what he termed the 'ethics of socialism', it was during a period of intense trade union strike activity and government repression that Sorel published *Réflexions sur la violence (Reflections on Violence)* (1908a). Heavily influenced by Vico, Nietzsche and Bergson (as well as by Marx), Sorel here argued that it would be through acts of violence that the proletariat would establish the economic foundations of socialism while simultaneously engendering a new morality of the producers. If the working class, and more narrowly the revolutionary syndicalist movement, was to be the vehicle of this radical transformation of society, its inspiration was to be a series of 'myths' associated principally with what was characterized as the proletarian epic of the general strike. Arguing against what he termed 'the intellectualist philosophy' Sorel's point was that 'people who are living in the world of "myths" are secure from all refutation' and therefore are capable of heroic deeds. It was such heroism on an almost Homeric scale that was required if the 'total ruin of institutions and of morals' was to be reversed.

To understand why Sorel adopted this stance it is necessary to be familiar with at least two dimensions of his thought. First, an intense moralism characterizes his writings throughout and underpins and explains his frequent changes in political position. Sorel was convinced that he lived in a world of moral decline and decadence and this he came increasingly to associate with the Western bourgeoisie. Sorel's own preference was for what was, in effect, a Catholic and Proudhon-inspired morality of extreme severity, leading him to argue (for example) that the 'world becomes more just to the extent that it becomes more chaste'. Always a pessimist, he saw the image of the Wandering Jew, 'condemned to march forever without knowing rest', as the symbol of man's highest aspirations. What mattered was that the working class should not be contaminated by this decadence.

Second, if Sorel was one of the people most responsible for the introduction of the ideas of Marx into France during the 1890s, then he also engaged in the process of an extensive reformulation and reinterpretation of what he took to be the central tenets of Marxist philosophy. Quickly abandoning the idea that Marx had discovered the laws that 'determined' the development of capitalism, he first redefined Marx's key notions (for example, the theory of value) as metaphysics, then as 'social poetry', before finally hitting upon the idea that they were best seen as 'myths' capable of inspiring action in the proletariat. In political terms this initially meant support for social democracy (and agreement with Eduard Bernstein), but the outcome of the Dreyfus Affair convinced Sorel that the 'revolutionary idea' was indispensable for Marxism. Crucially, however, the revolution (if it ever occurred) was not to take the form of the collapse of capitalism but of the transvaluation of all values (a term he borrowed fromNietzsche to denote his rejection of orthodox Marxism and its emphasis upon economics).

By 1909 Sorel had lost faith in the capacity of the French working class to effect the social transformation he so desired. There followed a brief association with intellectuals on the extreme Right, opposition to the First World War and the plutocratic civilization it represented, before a final enthusiasm for Lenin's Bolshevik Revolution and the hope that this might bring the end of 'the arrogant bourgeois democracies, today shamelessly triumphant'.

Yet beneath this lay a deeper interest in the philosophical issues of the day. Educated in the positivist tradition of Comte, Sorel repeatedly sought to defend science as a progressive, experimental activity while refuting the universalistic claims of the scientism he associated with the positivism he had imbibed in his youth. His extensive writings on both religion and science reveal a thinker determined to defend the validity of our different forms of knowledge, embracing a methodological and epistemological pluralism that would itself avoid the perils of subjectivism. In this he drew inspiration not just from Vico's notion of *verum ipsum factum*, that we can only have knowledge of what we have made ourselves, but in turn from the conventionalism of Poincaré, the intuitivism of Bergson and the pragmatism of William James.

See also: Revolution; Socialism

JEREMY JENNINGS

List of works

Sorel, G. (1889) *Le Procès de Socrate* (The Trial of Socrates), Paris: Alcan.(An early statement of Sorel's moral position.)

Sorel, G. (1906) *Le Système historique de Renan* (The Historical System of Renan), Paris: Jacques.(Sorel's most extensive discussion of historical methodology and religion.)

Sorel, G. (1908a) *Réflexions sur la Violence*, Paris: Pages libres; trans. T.E. Hulme, *Reflections on Violence*, New York: Collier Books,1972.(Sorel's standard work.)

Sorel, G. (1908b) *La Décomposition du Marxisme (The Decomposition of Marxism)*, Paris: Rivière.(Sorel's most succinct analysis of Marxism.)

Sorel, G. (1908c) *Les Illusions du progrès*, Paris: Rivière; trans. J.L. Stanley, *Illusions of Progress*, Berkeley and Los Angeles, CA: University of California Press, 1969.(A full-blooded attack on the Enlightenment as bourgeois ideology.)

Sorel, G. (1921) *De l'utilité du pragmatisme (The Utility of Pragmatism)*, Paris: Rivière. (Sorel's final philosophical statement.)

Sorel, G. (1976) *From Georges Sorel: Essays in Socialism and Philosophy*, ed. J.L. Stanley, New York: Oxford University Press.(An excellent selection of Sorel's writings.)

References and further reading

Andreu, P. (1982) *Georges Sorel: entre le noir et le rouge (Georges Sorel: Between the Black and the Red)*, Paris: Syros.(Published originally in 1953, this remains a standard work of reference.)

Jennings, J. (1985) *Georges Sorel; The Character and Development of his Thought*, London: Macmillan.(An account of the full range of Sorel's ideas from the 1880s until his death.)

Sand, S. (1985) *L'illusion du politique: Georges Sorel et le débat intellectuel 1900 (The Illusion of Politics: Georges Sorel and intellectual debate around 1900)*, Paris: La Découverte.(Brilliant exegesis of Sorel's encounter with Marxism.)

Stanley, J. (1981) *The Sociology of Virtue: The Political and Social Theories of Georges Sorel*, Los Angeles, CA: University of California Press.(A challenging interpretation of Sorel's thought.)

Sôsan Hyujông (1520-1604)

Sôsan was a Korean Sôn Buddhist monk who sought to establish the equality of various ideas and systems. His philosophical perspective conferred equality on Confucianism, Daoism and Buddhism alike, and within Buddhism he denied there was any inherent conflict between the Kyo and Sôn schools. However, he viewed Sôn (meditation) as being the most advanced form of practice.

Sôsan Hyujông, popularly known as 'Sôsan,' was a Korean Sôn (Zen, in Japanese) Buddhist monk of the Chosôn Dynasty (1393-1910) who espoused 'liberal' views. Why were Sôsan's views so liberal? The answer may lie in the fact that he was an extremely pure Sôn monk. True to the spirit of Sôn, he perceived all things and all beings as equal. Perhaps we can clarify this point through the well-known Buddhist simile of an ink-seal, used to stamp three different elements: clay, water and air. In both clay and water, traces of the seal remain, but in air the seal makes no imprint whatsoever. Sôsan was like the air: whatever happened to him, he was not influenced by it but remained the same. This was perhaps his most valuable asset, as it afforded him much flexibility in both thought and action.

Sôsan is important in the history of Korean Buddhism for three reasons. First, during the Japanese invasion of Korea in 1592, he organized an army of Buddhist monks to participate in the country's defence. Though this force helped to repel the invaders, he was later criticized for breaking the Buddhist precept against killing. While believing that his actions had been necessary, Sôsan himself allegedly felt a certain amount of remorse. After the war, King Sônjo recognized his contribution with an award, and the two men subsequently developed a friendship.

Second, Sôsan pioneered an overarching philosophical perspective which stressed the essential equality of all three of the religious systems then active in Korea: Confucianism, Daoism and Buddhism. One commentator has explained Sôsan's view of the relationship of the three religions in terms of rice farming: while the practice of Confucianism is akin to planting the seeds, and that of Daoism is akin to nourishing them, the practice of Buddhism is akin to uprooting sick shoots and replacing them with healthy ones. Sôsan thus viewed all three thought systems as cooperative and complementary, and believed that each should be allowed to perform its own special function free of interference from the others. Due to his influence, communication was greatly expanded among the leading thinkers of his day.

Third, Sôsan espoused a similar doctrine of non-interference within Buddhism itself, particularly in regard to the longstanding antagonism between the scriptural (Kyo) and meditational (Sôn) schools. Denying the view that there was an inherent conflict between these schools, Sôsan instead forged a conceptual link between them by equating Kyo with the Buddha's words and Sôn with his mind. How can they be contradictory, he asked, since both originate from the same source? He did feel, however, that once a practitioner had achieved a firm grasp of scripture he should continue on to the next stage, which was meditational practice. Thus Sôsan still viewed Sôn, or meditation, as the more advanced form of practice. In this sense he deserves to be called a typical Chan master, in the tradition of the Linji school.

See also: Buddhist philosophy, Korean

SUNG BAE PARK

List of works

Han'guk pulgyo chônsô (The Collected Works of Korean Buddhism), Seoul: Dongguk University press, 1979, vol. 7, 616-751; trans. in P.H. Lee (ed.) *Source Book of Korean Civilization*, vol. 1, *From Early Times to the Sixteenth Century*, New York: Columbia University Press, 1993.(Lee's translation is not entirely complete, but does include all the important intellectual works. Lee's work is a well-written, comprehensive anthology of Korean civilization.)

References and further reading

The Korean Buddhist Research Institute (ed.) (1994) *Buddhist Thought in Korea*, Seoul: Dongguk University Press.(A collection of articles by various authors on various aspects of Korean Buddhist thought.)

Kim Young-tae (1975) *Sôsan Daesa ûi Saeng'ae wa Sasang (Venerable Sôsan's Life and Thought)*, Seoul: Pagyongsa.(A biography of Sôsan, including a discussion of his thought.)

Soto, Domingo de (1494-1560)

The sixteenth-century Spanish Dominican, Domingo de Soto, was a mainstay of the Thomistic revival begun at Salamanca by Vitoria. After study at Paris (where he was taught by the nominalist John Major) and Alcalá, Soto taught both philosophy and theology. He was influential within the Dominican Order and the Catholic Church; he served as Emperor Charles V's theologian at the Council of Trent and played an active role in the development of the Council of Trent's teaching on original sin. Besides his theological writings, Soto composed philosophical works chiefly in logic, natural philosophy and juridical theory. In logic, he authored an exposition of the Summulae of Peter of Spain and a commentary, by way of questions, on three of Aristotle's works. His natural philosophy anticipated later scientific approaches, while in his philosophy of law Soto presented a basically Thomistic doctrine updated to confront sixteenth-century issues.

1 Life and works

Born in 1494 at Segovia in Spain, Domingo de Soto studied at the University of Alcalá between 1512 and 1516. In 1517 he received a master of arts degree at the University of Paris, where he came under nominalist influence. Between 1517 and 1519, while teaching philosophy at Paris, he studied theology there, principally under the Scottish master, John Major. At this time he may have attended lectures by Francisco de Vitoria at the Dominican house of St Jacques in Paris. Returning to Spain, he continued the study of theology, as well as teaching philosophy, at Alcalá, where in 1525 he received his doctorate in theology. That same year he joined the Order of Preachers, the Dominicans. From 1525 to 1532 he was teaching philosophy and theology in the Dominican convent of San Esteban at Salamanca. In 1529 he published at Burgos the first edition of his *Summulae*. This work, which he would revise extensively in 1539, was a logic textbook partly based upon the famous *Summulae logicales* of Peter of Spain. In 1532 he assumed the 'Chair of Vespers' in theology at the University of Salamanca, which office he held until 1549. During this period he served two terms as prior of San Esteban.

Soto was chosen by Charles V to be an imperial theologian at the first sessions of the Council of Trent; between 1545 and 1547 he played an active role in the development of the Council's teaching on original sin. He also wrote an important treatise, *De natura et gratia, libri tres (Three Books on Nature and Grace)*, published at Venice in 1547. From 1548 to 1550 he was confessor to Charles V in Germany. At this time Charles offered him the bishopric of his native Segovia, which Soto respectfully but firmly refused. Back in Spain, the following year he presided over the Junta de Vallodolid, made up of fourteen theologians selected to judge the matter of the Spanish conquest of the Indians of the New World. The principal business before the Junta was a debate between the humanist Gino de Sepúlveda, the defender of the Spanish role, and its most severe critic, Bartolomé de las Casas, Bishop of Chiapa in Mexico (see Latin America, colonial thought in §1). Soto composed an objective summary of the debate which presented both positions fairly but did not take sides.

In 1552, Soto succeeded Melchior Cano in the 'Chair of Prime' in theology at Salamanca, a chair earlier occupied by Vitoria. In this position, he continued Vitoria's renewal of Thomism and its development to meet new problems presented by the Reformation, the discovery of America, and the rise of a free-market system in Spain. In 1554 Soto published a third edition of his *Summulae*, in 1555-6 a commentary on the fourth book of the *Sentences* of Peter Lombard, and in 1553-4 his famous *De justitia et jure, libri decem (Ten Books on Justice and Law)*. Retiring from his chair in 1556, he once more became prior of San Esteban, in which office he died on 15 November 1560.

Soto's philosophical work is in three main parts: logic, natural philosophy, and juridical theory or philosophy of law. In addition, his theological writings touch on diverse philosophical issues. For example, in his treatise on nature and grace, he dealt at length with the nature of human beings, the soul's relation to the body, the intellect's need for the senses, and the freedom of the human will *vis-à-vis* the causality of God. At different points in his commentary on the fourth book of the *Sentences*, which is mainly concerned with sacramental theology, he discussed signs and things signified, conditions required for a definition, the existence of accidents in material substances, the possibility of accidents existing without a subject in which they inhere, the transcendental and non-generic character of being as shared between substance and accidents, the immortality of the human soul, as well as the communication of temporal authority by God first to the people and then to the sovereigns they choose.

2 Logic

In logic, Soto's revised *Summulae* of 1539 and 1554 comprises five books, plus two treatises on insolubles and obligations. Book 1 is on terms, which Peter of Spain defined as 'words which are by convention significative'. This occasions for Soto, in Lecture 1 of Chapter 2, an extended treatment of signification and signs. The latter are divided into natural, conventional and customary. Natural signs in turn are divided into formal and instrumental. Book 2 is on propositions. Here he deals with nouns and verbs as they compose propositions. He then treats of definition, division and argumentation. Chapters 5 to 8 of Book 2 treat of categorical and hypothetical propositions, while further chapters cover supposition, relatives, induction, ampliation and appellation. Book 3 concerns the opposition, equipollence and conversion of propositions. Book 4 treats exponible propositions and their species such as exclusives, exceptives, reduplicatives and so on. Book 5 examines the syllogism, its figures and moods, its essential conditions and defects.

In the first volume of his work *In Dialecticam Aristotelis (On the Logic of Aristotle)* (1543), Soto begins with questions about the scientific character and subject matter of logic. Following this he comments on the *Isagōgē* of Porphyry and treats the notions of genus, species, difference, property and accident. In a second volume, he deals with *antepredicamenta* (which are the definitions of equivocal, univocal and analogical terms, the division of terms into simple and complex, and so on), the ten categories of substance and accidents, as well as *postpredicamenta* (for example, opposition and its species). In the same volume, he comments on twenty-seven chapters of Book 1 of the *Posterior Analytics*, raising special questions with regard to science and demonstration. Then more briefly he explains eighteen chapters of Book 2 of the *Posterior Analytics*, without raising any special questions (see Language, Renaissance philosophy of §1; Logic, Renaissance §2; Aristotelianism, Renaissance).

3 Natural philosophy

In his work *Super VIII libros Physicorum Aristotelis (On the Eight Books of Aristotle's Physics)* (1545), Soto was concerned, according to Marcial Solana (1940), not so much to know what Aristotle thought as to know the truth with respect to nature, its principles and causes, movement, place, time, infinity, continuity, and so on. In at least one passage, Soto has, in the estimate of Pierre Duhem (1913), anticipated Galileo (§1) on the free fall of bodies. The core of this passage reads: 'When a body falls from a height through a uniform medium, it is moved with more velocity at the end of its fall than at the beginning. But the motion of projectiles is slower at the end than at the beginning. And in this way the first [motion] "gets faster with uniform acceleration" (*uniformiter difformiter intenditur*) while the second ''declines with uniform deceleration'' (*uniformiter difformiter remittitur*)' (*Super VIII libros Physicorum*, VII q.3). However, Soto did not anticipate the careful experimentation and the mathematics on which the fame of Galileo rests.

4 Philosophy of law

Soto's *De justitia et jure, libri decem* presents a basically Thomistic doctrine updated to confront sixteenth-century issues. In Book 1 he explicates the notion of *jus*, which first means 'law', but also means 'right' in the sense of what is right or just. In this latter sense, *jus* is the object of the virtue of justice. Law is defined as 'an ordinance of reason promulgated for the common good by one who has charge of the republic'. Law is divided into eternal, natural and positive (divine and human). Eternal law is the divine reason which from eternity has ordered all things to their respective ends. Natural law is an impression of the eternal law upon human beings to guide them in ordering their actions to the end for which they were created (see Natural Law §1). The natural law is made up of precepts which are immediately evident and unchangeable and precepts inferred from these as conclusions which are subject to change by addition or by changing circumstances (*mutatio materiae*). Divine positive law comprises prescriptions and proscriptions of the Old and New Testaments. Human positive law is either civil or canon law, plus the law of nations (*jus gentium*). The law of nations is positive law which human beings inasmuch as they are rational have everywhere established for themselves. For this establishment, unlike that of civil law, there was no need of a general assembly or an explicit enactment. Instead, by reason alone, human beings everywhere are in substantial agreement on the main facts of dominion, private property, exchanges, buying and selling, war and peace, slavery in some instances, keeping faith even with enemies, the immunity and protection of ambassadors, and so on. After treating distributive justice in Book 3, Soto deals in Book 4 with commutative justice. Here he looks at ownership and its transfer. In Book 5, he treats injustice of various kinds and its punishment. Here he supports the right of the state to exact the death penalty and to wage a just war. In the remaining books, Soto deals in more detail with property rights and their transfer, partnerships, sales contracts, just prices, deposits, loans,

interest and usury. On numerous occasions he discusses the rights of individuals *vis-à-vis* those of the community (see Law, philosophy of; Political philosophy, history of §8).

See also: Aristotelianism, Renaissance §§2-3, 7; Law, philosophy of; Logic, Renaissance

JOHN P. DOYLE

List of works

Soto, D. de (1529) *Summulae*, Burgos; 2nd edn, Salamanca: Andreas de Portonariis, 1539; 3rd edn, 1554; repr. Hildesheim: Olms, 1980.(A logic textbook partly based on the *Summulae logicales* of Peter of Spain and extensively revised in 1539.)

Soto, D. de (1543) *In Dialecticam Aristotelis Commentarii (Commentary on the Logic of Aristotle)*, Salamanca: Juan de Junta; Venice, 1587; repr. Frankfurt: Minerva, 1967.(Soto's commentary on Porphyry's *Isagōgē*, and Aristotle's *Categories* and *Posterior Analytics*.)

Soto, D. de (1545) *Super VIII libros Physicorum Aristotelis commentaria (Commentary on the Eight Books of Aristotle's Physics)*, Salamanca, 1st complete edn, 1551.(An independent commentary, by way of questions.)

Soto, D. de (1553-4) *De Justitia et jure, libri decem (Ten Books on Justice and Law)*, Salamanca, 2nd edn, 1559; Spanish trans. M. González Ordoñez, *De justitia et jure. De la justicia y del derecho*, Madrid: Instituto de Estudios Politicos, 1967-8, 5 vols.(Soto's main work of jurisprudence; the Madrid edition provides Latin text with Spanish translation, and an introduction by V.D. Carro.)

References and further reading

Ashworth, E.J. (1974) *Language and Logic in the Post-Medieval Period*, Dordrecht and Boston, MA: Reidel.(A general study which includes extensive discussion of Soto's logical doctrine.)

Ashworth, E.J. (1990) 'Domingo de Soto (1494-1560) and the Doctrine of Signs', in G.L. Bursill-Hall, S. Ebbesen and K. Koerner (eds) *De Ortu Grammaticae: Studies in Medieval Grammar and Linguistic Theory in Memory of Jan Pinborg*, Amsterdam: John Benjamins, 35-48.(Soto's semiotic theory discussed.)

Ashworth, E.J. (1988) 'Changes in Logic Textbooks from 1500 to 1650: The New Aristotelianism', in E. Kessler, C.H. Lohr and W. Sparn (eds) *Aristotelismus und Renaissance: In memoriam Charles B. Schmitt*, Wiesbaden: Harrassowitz, 75-87. (Discusses the change between Soto's *Summulae* of 1529 and 1539 on pages 80-1.)

Barrientos García, J. (1985) *Un siglo de moral económica en Salamanca (1526-1629): I, Francisco de Vitoria y Domingo Soto (A Century of Economic Morality in Salamanca)*, Salamanca: Ediciones Universidad de Salamanca.(Contains extensive discussion of the economic philosophy of Soto.)

Beuchot, M. (1988) 'Un libro de texto para la Nueva España: La Lógica de Domingo de Soto' (A Textbook for New Spain: the Logic of Domingo de Soto), *Revista de Filosofia de la Universidad Ibero-Americana* 21: 62, 152-67. (Treats the influence of Soto's logic in sixteenth-century Mexico.)

Carro, V.D. (1943) *Domingo de Soto y su doctrina jurídica (Domingo de Soto: His Juridical Doctrine)*, Salamanca: Hijos E. Minuesa.(Treatment of Soto's legal theory.)

Duhem, P.M.M. (1913) 'Les précurseurs parisiens de Galilée: Dominique Soto et la scolastique parisienne' (Galileo's Parisian Forebears: Domingo de Soto and Scholasticism in Paris), in *Études sur Léonard de Vinci*, chap. 3, 261-583.(Referred to in §3; Soto as anticipating Galileo.)

Hamilton, B. (1963) *Political thought in sixteenth-century Spain: A Study of the Political Ideas of Vitoria, Soto, Suarez and Molina*, Oxford: Clarendon Press.(Soto's political theory treated in the course of a wider study.)

Lohr, C.H. (1987) *Latin Aristotle Commentaries: II Renaissance Authors*, Florence: Olschki, 430-1.(Contains biographical and bibliographical information on Soto.)

Muñoz Delgado, V. (1964) *Lógica formal y filosofía en Domingo de Soto (1494-1560) (Formal Logic and Philosophy in Domingo de Soto)*, Madrid: Edita Revista Estudios.(A good treatment of Soto's logic.)

Risse, W. (1964) *Die Logik der Neuzeit. I Band. 1500-1640 (The Logic of the Modern Period)*, Stuttgart and Bad Cannstatt: Frommann, vol. 1, 329-36.(Soto's logic treated within its historical context.)

Solana, M. (1940) *Historia de la filosofía española: Época del renacimiento (siglo XVI) (History of Spanish Philosophy. The Time of the Renaissance (16th Century))*, Madrid: Asociación Española para el Progreso de las Ciencias, vol. 3, 91-130. (Referred to in §3. A good introduction to Soto.)

Wallace, W.A. (1981) *Prelude to Galileo: Essays on Medieval and Sixteenth-Century Sources of Galileo's Thought*, Dordrecht and Boston, MA: Reidel, esp. 99-109. (Discussion of Soto's adumbration of Galileo on the free fall of bodies.)

Soul in Islamic philosophy

The discussion of the human soul, its existence, nature, ultimate objective and eternity, occupies a highly important position in Islamic philosophy and forms its main focus. For the most part Muslim philosophers agreed, as did their Greek predecessors, that the soul consists of non-rational and rational parts. The non-rational part they divided into the plant and animal souls, the rational part into the practical and the theoretical intellects. All believed that the non-rational part is linked essentially to the body, but some considered the rational part as separate from the body by nature and others that all the parts of the soul are by nature material. The philosophers agreed that, while the soul is in the body, its non-rational part is to manage the body, its practical intellect is to manage worldly affairs, including those of the body, and its theoretical intellect is to know the eternal aspects of the universe. They thought that the ultimate end or happiness of the soul depends on its ability to separate itself from the demands of the body and to focus on grasping the eternal aspects of the universe. All believed that the non-rational soul comes into being and unavoidably perishes. Some, like al-Farabi, believed that the rational soul may or may not survive eternally; others, like Ibn Sina, believed that it has no beginning and no end; still others, such as Ibn Rushd, believed that the soul with all its individual parts comes into existence and is eventually destroyed.

1 The existence of the soul

All Muslim philosophers concerned themselves with the subject of the soul. The most detailed and most important works on this subject are those of al-Kindi, al-Farabi, Ibn Sina and Ibn Rushd. Muslim philosophers recognized that the first issue that confronts the human mind with regard to the soul is its existence. That is why, at the very beginning of his inquiry about the soul in *al-Shifa' (Healing)*, Ibn Sina (§6) asserts that we infer the existence of the soul from the fact that we observe bodies that perform certain acts with some degree of will. These acts are exemplified in taking nourishment, growing, reproducing, moving and perceiving. Since these acts do not belong to the nature of bodies, for this nature is devoid of will, they must belong to a principle they have other than bodies. This principle is what is called ‘soul’.

This argument is intended to prove the existence of the animal soul, which includes the plant soul. The soul is the source of acts performed by the will, not inasmuch as it is ‘a substance’ (an independent entity), but inasmuch as it is ‘the principle of such acts’. The rational soul, on the other hand, need not look outside itself to infer its existence. It is aware of its existence with immediacy, that is, without any instruments. Ibn Sina’s example of the suspended man is intended to prove that the rational soul is aware of itself apart from any body. His argument boils down to the view that, even if the adult rational soul is not aware of anything material, not even its body, it remains aware of its own existence.

2 The nature of the soul

While Islam made it incumbent on Muslim philosophers to occupy themselves extensively with the study of the soul and to make certain statements that in some cases appear consistent with Islamic beliefs, Greek philosophy had the upper hand in forming the real convictions of Muslim philosophers with regard to the nature of the soul. Unless otherwise specified, reference to the soul here is limited to the terrestrial soul to the exclusion of the celestial one, since Muslim philosophers concerned themselves primarily with the former. It must be pointed out at the outset that ‘soul’ (*nafs*) was used in more than one sense in Islamic philosophy; the term was used to refer to the plant or vegetative part of a living being, the animal or sensitive part, the rational part and finally the totality of all three parts. The first two are the non-rational soul and the totality is the human soul. To add to the confusion, ‘human soul’ is used only in the sense of this fourth type of soul. The plant, animal and rational souls are also called powers or parts of the soul. Only from the context can one understand whether a Muslim philosopher was using ‘soul’ in the broad sense to mean the human soul (the totality of the parts of the soul), or in the narrow sense to mean a specific part of the human soul.

Inasmuch as it has a certain relation to a body, the soul is a form for that body, that is, the perfection of that body. It is a form because a natural body is composed of matter and form, which in the case of animals are body and soul. Since it has been shown that the soul is the source of will and therefore is not matter, it remains a form. Perfection is of two types, primary and secondary. A primary perfection is what makes a thing actually a species,

as shape does for the sword, or a genus, as sensation and movement do for animals. A secondary perfection is an act necessitated by the nature of the species or genus, such as cutting for the sword and touching for animal. The soul is a primary perfection of a natural body capable of performing the secondary perfections necessitated by this primary perfection. Together with its body, the soul constitutes a material substance. This substance can be the subject of plant, animal or human life.

The soul is a perfection inasmuch as it makes a natural body into a plant, an animal or a rational being. However, to define the soul as a perfection does not give us a clue as to what the soul is in itself, but only inasmuch as it has a relation to the body. The body is, therefore, an essential element in the definition of the soul. Without relating to a body, the thing we call ‘soul’ is not a soul and does not require the body as an essential part of its definition. Note, however, that in spite of this assertion, perhaps for the lack of any better term, Muslim philosophers use ‘soul’ also to refer to the rational soul after it separates from the body and reaches a complete state of purity from matter.

In its first or lowest stages of relating to the body, the soul is the plant soul, which is a primary perfection for an organic natural body inasmuch as this body can take nourishment, grow and reproduce. The plant soul is the power human beings and other animals share with plants. If the body with a soul is an animal, the soul develops into the animal soul, which is a primary perfection for an organic natural body inasmuch as this body has sensation and movement through will. While this soul includes the plant soul, it has also a sensitive power and a locomotive one. The sensitive power has both external and internal senses. The external senses are, in priority of existence, touch, taste, smell, hearing and sight. The first three are said to be necessary for survival and the last two for well being. In *Talkhis kitab an-nafs (Middle Commentary on Aristotle’s On the Soul)*, Ibn Rushd (§3) asserts that the five external senses may be in potentiality, as in infancy and sleep, or in actuality, as in daily seeing or hearing. He also argues that there cannot be any external sense other than these five because there would be no function for it, since there is no external sensation other than the objects of the five senses mentioned above. Most Muslim philosophers mention three types of internal senses: common sense, imagination and memory. Ibn Sina (§3) enumerates five internal senses: common sense, representational power, imagination, estimative power and memory. On the whole, the philosophers agree on the function of the common sense, imagination and memory; the function that Ibn Sina limits to the representational and estimative powers, other Muslim philosophers allocate to the imagination.

The common sense is an internal power in which all the objects of the external senses are collected. Contrary to the external senses, which can grasp only one type of sensation, as sight grasps light and hearing grasps sound, the common sense can grasp all external sensations, such as that honey is of such and such a colour, texture and smell. The representational power preserves the sensations of the common sense even after sensible things disappear. The imagination selects at will to combine some of the objects of the representational power with each other and to separate the rest. It makes its judgment about external things, but in the absence of these things. That is why it functions best when the external senses, which represent external things, are not at work, as in sleep. Ibn Rushd points out that animals such as worms and flies that do not act except in the presence of sensible things are devoid of imagination. The imagination is called such inasmuch as it is an animal instrument; it is called cognitive inasmuch as it is a rational instrument. The estimative power grasps non-sensible notions of sensible things, such as the sheep’s notion that the wolf is to be avoided. This notion is about a sensible thing but is not grasped through the external senses, as is the colour or shape of a wolf. Memory preserves the notions of the estimative power. The imagination acts on the objects of memory in the same way it acts on those of the representational powers. Like the objects of the external senses, those of the internal senses are particular and material. The difference is that they can be experienced in the absence of external things and are to some degree abstracted from matter.

The locomotive power branches into that which causes movement and that which actually moves. The former, the desiderative power, subdivides into the appetitive and the irascible. The appetitive causes movement toward what is imagined to be necessary or beneficial in the pursuit of pleasure. The irascible causes avoidance of what is imagined to be harmful or an impediment in the pursuit of dominance. The power that actually moves uses the nerves to relax the muscles at the demands of the appetitive power or tighten them at the demands of the irascible one.

3 The rational soul

The rational soul, which is defined as a primary perfection for an organic natural body inasmuch as this body can

act by rational choice and grasp the universals, is divided into the practical and the theoretical intellects. The practical intellect seeks knowledge in order to act in accordance with the good in its individual body, its family and its state. It must, therefore, know the principles for properly managing the body, the family and the state, that is, ethics, home management and politics. The practical intellect is the rational soul turning its face downward. The function of the theoretical intellect is to know just for the sake of having the universals (the realities or natures of things). Some of these natures, such as God and the intellect, cannot attach to movement; knowledge of them is metaphysics. Other natures, such as unity, can attach to movement but do not; knowledge of them is mathematics. Still other natures, such as humanity and squareness, can attach to movement either in reality and thought, such as humanity, or in reality but not in thought, such as squareness. Knowledge of these is physics.

The theoretical intellect is the rational soul with its face upward. The practical intellect looks up to the theoretical one and moves its body accordingly. In this, the practical intellect is similar to the celestial soul that looks up to the intellect of its sphere and moves its sphere accordingly. Thus, like the celestial soul, the practical intellect is the link between intellect as such and matter.

On the whole, Muslim philosophers followed al-Kindi's division of the theoretical intellect into the material intellect (*al-'aql al-hayulani*), the habitual intellect (*al-'aql bil-malaka*), the actual intellect (*al-'aql bi'l-fi'l*) and the acquired intellect (*al-'aql al-mustafad*). The material intellect is a blank slate with the potentiality for grasping the intelligible forms or universals. Ibn Sina points out that it is referred to as material, not because it is actually material but because it resembles matter in accepting the form. The habitual intellect grasps the universals, as one acquires the skill to write; in other words, this intellect has the ability to use the universals but does not always do so. The actual intellect grasps the universals in actuality and is always ready to use them. While Muslim philosophers differed slightly with regard to their accounts of the acquired intellect, their general view is that it is the highest human state, the point of contact with the divine, the agent intellect (the intelligence of the moon, the lowest celestial intellect), which makes it possible for the theoretical intellect to acquire the universals in the purest form (see Epistemology in Islamic philosophy §4).

4 The ultimate objective of the soul

Al-Farabi asserts that even though the soul is of different parts, it is a unity with all its parts working for one final end, happiness. While the plant soul, for example, serves a specific function, it also serves the powers that are higher than it in rank, the animal powers. Without nourishment, growth and reproduction, the animal powers cannot perform their necessary functions. Similarly, while the function of the animal powers is to have sensation and movement, by performing this function they also promote the functions of the powers above them, the rational ones. The operations of the animal powers, especially those of the senses, are particularly important for the attainment of the final end. The external senses strip the forms from material objects and convey them to the internal senses. The more they are transferred internally, the less mixed with matter do they become. Since the innermost sense they reach is the imagination, they are there in their purest material existence (see Imagination).

The role of the objects of the imagination is not always clearly defined in Islamic philosophy. Occasionally it is said by somebody like Ibn Sina to be one of preparation for the theoretical intellect to receive the universals from the agent intellect. At other times Ibn Sina, like other Aristotelians such as Ibn Rushd, takes these objects to be the ingredients out of which the universals are made after the last process of purification (see Epistemology in Islamic philosophy). It seems, however, that in either case the light of the agent intellect is needed to complete the process. In the former case, this light gives the intelligible forms to the theoretical intellect when this intellect is prepared. In the latter case, it sheds itself on the objects of the imagination, which are then reflected on the theoretical intellect without their matter. Since the theoretical intellect is in its first stages in potentiality, it cannot act on the objects of the imagination directly; hence the need for the agent intellect, which is pure actuality. The role of the practical intellect in all this is to put order into the body. This sets free the theoretical intellect from preoccupation with the body and helps the powers whose function is necessary for theoretical knowledge to function unhampered.

Muslim philosophers adhered to the view that the acquired intellect is one with its objects, for they thought the knower and the known are one, as did their Greek predecessors. This means that the highest human state is one in which unity with the universals or the eternal aspects of the universe is reached. This state is described as happiness because in it eternity, an aspect of the objects of the acquired intellect, is attained.

5 Eternity of the soul

When Muslim philosophers assert that the soul comes into existence simultaneously with the coming into existence of the body, some, such as Ibn Sina (§6), who believe that the rational soul is in essence non-material, are thinking only of the non-rational soul. Others, such as Ibn Rushd (§3), who believe that the rational soul is originally not separate from matter, contend that the whole human soul comes into existence. The latter believe that since the rational soul grasps the universals from particular sensibles, and since such sensibles are material and have a temporal beginning, this soul must also be material and must have a temporal beginning. Those who attribute non-materiality to the essence of the rational soul, such as al-Kindi and Ibn Sina, assert that this soul pre-exists the body. While all of them agree that the non-rational soul is destroyed after the destruction of the body, they differ with regard to the end of the rational soul.

Al-Kindi and Ibn Sina, for example, strongly adhere to the view that all rational souls are indestructible because by nature they are simple. Al-Farabi reminds us that the reason for eternal existence is the rational soul's knowledge of the eternal aspects of the universe. From this he draws the conclusion, as did Alexander of Aphrodisias before him, that only those rational souls that have this knowledge at their separation from the body are indestructible. Other rational souls are eventually destroyed. Ibn Sina finds in the grasping of the universals the grounds for happiness, not the eternity of the soul. Ibn Rushd seems to hold that only the acquired intellect can be indestructible; but the acquired intellect, he argues (as does his teacher Ibn Bajja), is divine and numerically one in all. Ibn Rushd was attacked for this view because it denies eternal existence of individual souls (see Averroism; Soul, nature and immortality of the).

See also: Aristotelianism in Islamic philosophy; Epistemology in Islamic philosophy; Ethics in Islamic philosophy; Ibn Rushd §3; Ibn Sina §6; Mystical philosophy in Islam; Soul, nature and immortality of the

SHAMS C. INATI

References and further reading

al-Farabi (*c*.870-950) *al-Madina al-fadila (The Virtuous City)*, trans. R. Walzer, *Al-Farabi on the Perfect State: Abu Nasr al-Farabi's Mabadi' Ara' Ahl al-Madina al-Fadila*, Oxford: Clarendon Press, 1985.(English and Arabic of the most comprehensive and best known philosophical work of al-Farabi.)

Ibn Rushd (1180) *Tahafut al-tahafut (The Incoherence of the Incoherence)*, trans. S. Van Den Bergh, London: Luzac, 1954.(A response to a number of issues raised by al-Ghazali against philosophers. One of the three most important of these issues is that of the soul and its fate.)

Ibn Rushd (*c*.1174) *Talkhis kitab an-nafs (Middle Commentary on Aristotle's On the Soul)*, ed. A.F. al-Ahwani, Cairo: Maktabat an-Nahda, 1950.(Includes also four other essays: Ibn Bajja's *Risalat al-ittisal (Essay on Conjunction)*, Ishaq Ibn Hunayn's *Kitab fi an-nafs (Book on the Soul)*, Ibn Rushd's *Risalat al-ittisal (Essay on Conjunction)* and al-Kindi's *Risalat al-'aql (Essay on Intellect)*.)

Ibn Sina (980-1037) *an-Nafs (The Soul)*, ed. F. Rahman, *Avicenna's de Anima*, London: Oxford University Press, 1959.(The most important and detailed philosophical treatise on the soul in Islamic philosophy, the sixth part of the *Physics* of *al-Shifa'*. An Arabic edition of the text is included.)

Ibn Sina (980-1037) *Ahwal an-nafs (The States of the Soul)*, ed. A.F. al-Ahwani, Cairo: Dar Ihya' al-Kutub al-'Arabiyya, 1952.(Includes *Risala fi an-nafs wa-baqa'ha wa-ma'adiha (Essay on the Soul, Its Permanence and Its Second Life), Mabhath 'an al-qiwa an-nafsaniyya (Inquiry about Psychic Powers), Risala fi ma'rifat an-nafs an-natiqa (Essay on Knowing the Rational Soul)* and *Risala fi al-kalam 'ala an-nafs an-natiqa (Essay on an Inquiry Concerning the Rational Soul)*.)

Ibn Sina (980-1037) *an-Najat (Deliverance)*, ed. F. Rahman, *Avicenna's Psychology*, London: Oxford University Press, 1952.(The psychology of *an-Najat* is an abridgement by Ibn Sina of his encyclopedic work *al-Shifa' (Healing)*.)

Ibn Sina (980-1037) *Rasa'il ash-shaykh ar-ra'is fi asrar al-hikma al-mashriqiyya (Essays of the Master of the Head on the Secrets of Oriental Wisdom)*, ed. M. Mehren, Traités mystiques d'Avicenna, Leiden: Brill, 1889-99.(Ibn Sina's 'oriental philosophy'.)

Inati, S.C. (1996) *A Study of Ibn Sina's Mysticism*, London: Kegan Paul International. (Includes a detailed analysis of Ibn Sina's notion of the soul and a translation of the fourth part of *al-Isharat wa-'l-tanbihat (Remarks and Admonitions)*.)

al-Kindi (before 873) *Rasa'il al-Kindi al-falsafiyya (Al-Kindi's Philosophical Treatises)*, ed. M. Abu Rida, Cairo:

Dar al-Fikr al-‘Arabi, 1953.(Includes al-Kindi’s most relevant works on the subject of the soul, *al-Qawl fi an-nafs (Discourse on the Soul)*, *Fi an-nafs (On the Soul)* and *Fi mahiyyat an-nawm war-ru’ya (On the Essence of Sleep and Internal Vision)*.)

Soul, nature and immortality of the

For the Greeks, the soul is what gives life to the body. Plato thought of it as a thing separate from the body. A human living on earth consists of two parts, soul and body. The soul is the essential part of the human - what makes me me. It is the part to which the mental life of humans pertains - it is the soul which thinks and feels and chooses. Soul and body interact. Bodily states often cause soul states, and soul states often cause bodily states. This view is known as substance dualism. It normally includes the view that the soul is simple, that it does not have parts. If an object has parts, then one of those parts can have properties which another part does not. But for any experience that I have, an auditory or visual sensation or thought, it happens to the whole me. Plato also held that at death, soul and body are separated; the body decays while the soul departs to live another life. Aristotle, by contrast, thought of the soul simply as a 'form', that is, as a way of behaving and thinking; a human having a soul just is the human behaving (by moving parts of the body) and thinking in certain characteristic human ways. And just as there cannot be a dance without people dancing, so there cannot be ways of behaving without embodied humans to behave in those ways. Hence, for Aristotle, the soul does not exist without the body.

Christian theology, believing in life after death, found it natural to take over Plato's conception of the soul. But in the thirteenth century, St Thomas Aquinas sought to develop an Aristotelian conception modified to accommodate Christian doctrine. The soul, Aquinas taught, was indeed a form, but a special kind of form, one which could temporarily exist without the body to which it was naturally fitted. It has always been difficult to articulate this view in a coherent way which makes it distinct from Plato's. Descartes restated Plato's view. In more modern times, the view that humans have souls has always been understood as the view that humans have an essential part, separable from the body, as depicted by Plato and Aquinas. The pure Aristotelian view has more normally been expressed as the view that humans do not have souls; humans consist of matter alone, though it may be organized in a very complicated way and have properties that inanimate things do not have. In other words, Aristotelianism is a kind of materialism.

If, however, one thinks of the soul as a thing separable from the body, it could still cease to exist at death, when the body ceases to function. Plato had a number of arguments designed to show that the soul is naturally immortal; in virtue of its own nature, because of what it is, it will continue to exist forever. Later philosophers have developed some of these arguments and produced others. Even if these arguments do not show it (and most philosophers think that they do not), the soul may still be naturally immortal; or it may be immortal because God or some other force keeps it in being forever, either by itself or joined to a new body. If there is an omnipotent God, he could keep it in existence forever; and he might have revealed to us that he is going to do so.

1 The existence and nature of the soul

The form of a thing is its shape, appearance, pattern of reaction, and (in the case of a living thing) its way of changing over time and acting, in fact, all the properties that make it the kind of thing it is. The form of a living thing is, for Aristotle, its soul. Plants have 'vegetative' souls, animals have 'sensitive' souls and humans have rational or intellectual souls; a human soul is not just a way of behaving publicly, but also a way of thinking. Forms for Aristotle are universals, in the sense that the same form can be instantiated in many different things. The form of a table, for example, can be instantiated in many different tables. What individuates, what makes the particular table the one it is, is the matter of which it is made. So too, claims Aristotle, what makes a particular human the one they are is the matter of which they are made. No form, he claims, can exist apart from the thing (a particular substance) in which it is instantiated. So a human soul cannot be separated from its body. At least, a whole soul cannot be separated, for Aristotle considers that a soul has parts, and he adds (*On the Soul* 413a) that some parts of the soul may be separable. He may have in mind the part or parts responsible for thought.

Aquinas (§10-11) claimed that the soul does not need any bodily organ for thinking, and that is a main reason why he supposes that the whole soul can continue to exist without a body. In accord with the normal articulation of the Christian doctrine of life after death, Aquinas holds that the soul does continue to exist on the destruction of the body until it is reunited with it in the general resurrection. But what survives, claims Aquinas, is not the whole human being, but that essential part of it with limited functions, whose nature it is to be united with the body, and make it live and have its normal human operations of thought, perception and bodily action. Yet what makes such

a separated soul my soul as opposed to yours? This question cannot be answered within a pure Aristotelian scheme. For Aristotelian forms are universals. There is just the form of humanity, which can be instantiated in many different individuals, and the separated soul is not united with matter. Aquinas says that each form has an ‘inclination’, is ‘fitted’, to be reunited with a certain body, the one from which it has been separated. Duns Scotus objected to this answer on the ground that there cannot be a bare inclination; an inclination has to be grounded in some actually existing feature of the soul. To take a modern analogy, if some substance has an inclination to liquefy at a certain temperature, that must be grounded in its chemical constitution. Hence Scotus holds that souls are not mere universals, but individual forms or essences, that is, that there is an essence of Socrates as well as of humanity. Anyone who holds that souls can exist separately from bodies must give an account of what makes a soul this particular soul, and the same soul as that of some earlier person. For Plato and for Descartes, this is something ultimate, just as perhaps what makes this chunk of matter different from that chunk of matter is something not analysable further. So too for Scotus, even though he thinks of souls as forms of bodies (*Ordinatio* II, d.3).

If Aquinas’ view is to be spelled out coherently, it must be done in the same way. What did happen to a soul in the past, namely that it was united to a certain body, and will happen to it in future cannot make it the soul it is now. That must be something internal to it now. Religious believers who believe that humans can exist without their bodies, even if only temporarily, must hold that. So too must any believer who holds that there is life after death, even if souls do not exist separately from bodies. For if I come to live again, the question arises as to what makes some subsequent human me, for my body will be largely if not entirely destroyed. If the answer is given that (most of) the atoms of my original body will be reassembled into bodily form, there are two problems. First, many of the atoms may no longer exist; they may have been transmuted into energy. And second, what proportion of the atoms do we need? Sixty per cent, seventy per cent, or what? If it is mere atoms which make some body mine and so some living human me, then no body will be fully mine unless it has all my atoms. Yet some of my atoms, even if not destroyed, will have come to form other human bodies. It will be a matter of degree whether a human survives, unless what makes a human the individual human they are is something other than the body, namely the soul (see Resurrection).

But why suppose that there is a separate soul? Descartes (§§5, 8), in his *Meditations*, gave a famous argument which must be the basis of any justification of substance dualism. He argues that I know infallibly that I exist as a thinking thing, but I do not know infallibly that I have a body; maybe that is all a dream. Hence the body is a separate thing from what is essential to me, namely my soul. A traditional criticism of this argument is that the mere fact that I do not know infallibly that my body exists does not show that it is separable from me; maybe I do not know what is involved in my present existence. After all, the argument has the same form as this clearly invalid argument: I know infallibly that I am Bloggs; I do not know infallibly that I am the son of John and Mary; hence the son of John and Mary is not the same thing as Bloggs. The argument can, however, be rephrased to meet this objection, not as an argument from knowledge and ignorance, but as one from logical possibility and impossibility. Descartes, in writing of what is ‘imaginable’, is often near to phrasing the argument in this way. Here is a rephrased version:

(1) I exist as a thinking thing now.
(2) It is logically possible that I go on thinking and so existing, even if my body is suddenly destroyed (and this remains logically possible, whatever else might be the case now inside or outside my body compatible with my existing now as a thinking thing).
(3) It is not logically possible that any thing continue to exist unless some part of it continue to exist.

Hence I must already have another part beside my body, namely my soul. (For a more detailed presentation of this version of Descartes’ argument, see Swinburne [1986] 1997: 145-60, 322-32.)

Most objectors to the argument in this form will deny (2) (see, for example, Shoemaker and Swinburne 1984: 141-8). They would claim that it is not logically possible that I survive without my body. It certainly seems logically possible. Our only grounds for supposing something to be logically possible are that we can make sense of it; we can spell out what it would be like for it to be true. Our only grounds for supposing something to be logically impossible are that we can derive a contradiction from it. It does look as if each of us can tell a coherent story of losing our body and yet continuing to think, and it does not look as if there is an obvious contradiction

involved. So to some philosophers it looks as if the conclusion is sound.

Some claim that the causal interaction which substance dualism postulates between the soul and the body is unintelligible unless we can have some grasp of a mechanism through which brain states cause soul states and vice versa. A defender of substance dualism will respond that we often know very well that certain states cause others without having the least idea how this could occur. Before the theory of electromagnetism was developed, people knew that the motion of a magnetic charge caused an electric current, but had no idea how these very diverse states could influence each other.

2 The immortality of the soul

Even if the soul is simple and separable from the body, it does not follow that it will continue to exist after death, let alone exist forever with a mental life, with thoughts, feelings and sensations. It might instead exist in the way that our souls exist while we have a dreamless sleep. There have been a large number of arguments for the natural immortality of the soul, purporting to show that it is the nature of the soul to be immortal; it will naturally go on forever, without God or any external agency needing to intervene to give it special powers to continue in existence. Most of these are arguments which try to show that the soul has such a peculiar nature that the normal forces which cause things to cease to exist would not affect it.

One well-known argument is that given by Plato (§§10-11, 13) (*Phaedo* 78b) and repeated by, among others, Berkeley (§8) (1710, section 141) and Joseph Butler (§6) (1736 I, ch. 1), from the simplicity of the soul. It runs as follows. We only know of things ceasing to exist when they have parts; and they cease to exist only when the parts are separated from each other. A house ceases to exist when the bricks are taken away from each other. But the soul has no parts, and so we know of nothing which in the normal course of things would cause it to cease to exist, barring divine intervention. So it is reasonable to suppose that it is naturally immortal. But this argument is weak. We do know of things ceasing to exist without any parts being separated from each other. An atom ceases to exist when it is turned into energy. Further, it does look as if the functioning of the soul, its having a mental life, depends on the functioning of the brain. (Does not damage to the brain render a person un-conscious?) We need more positive argument to show that the soul could function without the brain sustaining it. And if it turns out that the soul is incapable of functioning without the brain, why suppose that it even exists?

Plato gives several other arguments for the natural immortality of the soul. They include the argument (*Phaedo* 73a-78a; *Meno* 81b-86b) that humans know many things and have many concepts which they have not learned or acquired on earth. In the *Meno*, Plato describes a slave-boy being led by Socrates to assert various truths of mathematics which no one has taught him. Hence, Plato argues, he must have learned them in a previous existence. But if souls exist before birth and through a life of embodiment on earth, it is natural to suppose that they will continue to exist after death. An objector, however, may reasonably argue that the slave-boy did not know the truths in question until Socrates helped him to become aware of them. The argument from pre-existence did not appeal to Christian philosophers, for Christians have normally held that souls come into existence at conception or birth, though they continue to exist after death. Aquinas appealed to the fact that humans naturally desire to exist forever and that ‘it is impossible that natural appetite should be in vain’ (*Summa contra gentiles* II, 79.6). Yet it is far from obvious why that is impossible. He also appealed to the ability of the soul to ‘apprehend incompatible things’, that is, to understand the eternal truths of mathematics and to have contact with God, as showing that the soul’s own nature is ‘incorruptible’, or naturally immortal (*Summa contra gentiles* II, 79.5). But it is in no way obvious that the finite cannot understand the infinite or that the mortal cannot understand the immortal.

Arguments to the *natural* immortality of the soul are very unappealing today. Some philosophers find arguments to show that the world is such that something external to the soul will guarantee its immortality more persuasive. An argument of Kant captures much more ordinary thinking. It is not strictly speaking, Kant claims, a theoretical argument which proves that the soul is immortal, but a ‘practical argument’ to this effect: we are obliged to act morally, and it only makes sense to do so if we suppose that we, and so our souls, will continue to live forever. Kant regards the various moral obligations under which we stand as different aspects of the obligation to realize the supremely good state of affairs, the *summum bonum*. The *summum bonum* involves our own moral perfection, something towards which we make only slow progress in this life and need endless duration finally to achieve. For moral actions to make sense, we must regard the *summum bonum* as attainable and so regard ourselves as having an endless life. Kant backs up his argument with the further consideration that the *summum bonum* requires that

happiness be in proportion to good moral behaviour, both for ourselves and others. But in this world the good are not always happy. So again for the *summum bonum* to be attainable we need another world in which 'the greatest happiness is… combined…with the highest degree of moral perfection' (Kant 1788, bk 2, ch. 2, section 5). This, Kant claims, requires the operation of God, who alone can bring about such coincidence. Now Kant has a very high conception of the status of moral obligation. Others may not regard its claims so highly. Yet even if we conceive of it as Kant did, it is in no way obvious that to make sense of our doing some moral action, we need to think of it as a step towards our own moral perfection which will eventually be achieved. When conscious of an obligation to help the starving, I give money to a charity; the point of my action is achieved by my giving the money, independently of whether it improves my character. Likewise its purpose is achieved whether or not I am the happier for achieving it. Holiness of will combined with supreme happiness may be good, but acting morally seems to have its own point quite apart from whether all that results from it. Moral actions are in general fully worth doing for totally mundane reasons, even if there are also reasons connected with an after-life for doing them.

Kant's argument has very considerable similarity to the argument of Eastern religions from the law of karma, which they affirm to be a basic principle governing the universe, that all actions get their proper reward (see Karma and rebirth, Indian conceptions of; Reincarnation). Good actions lead to happiness, and bad actions to unhappiness. Since so often this does not happen to a soul during a given earthly life, it must have another life in which morality and reward can be evened out. Although karma does not require another life on this earth, Eastern religions believe that we are normally reincarnated on earth. The good are reincarnated in more desirable states; the good poor man has a more wealthy next life, and the bad poor man may be reincarnated as a mere animal. But Eastern religions do not all hold that this necessarily goes on forever. Buddhism claims that our supreme goal should be to break the cycle of death and rebirth, by living such holy lives that after death we are not reincarnated at all, but depart to *nirvāṇa* (nothingness). It is disputable whether this means 'annihilation' or, on the contrary, a totally blissful state (see Nirvāṇa). However, before we could accept this argument, we would need good grounds for believing that the law of karma operates.

If the soul is not naturally immortal, then the action of some supernatural law (karma) or agent (God) is required to make it immortal. Much Christian tradition affirms that the free action of God is required for this purpose and that he is under no obligation so to act. In that case, in order to have grounds to believe in immortality, we will need his assurance that he will so act. At this point Christian tradition normally produces, not a Kantian argument, but the claim that it is an item of revelation. God has told us that he will so act, that he will bring the dead to life again - the good dead to an everlasting good life, and the bad dead to an everlasting bad life. (There are, however, differences within Christian tradition about what will happen to the bad, and to those who die before they become clearly good or bad (see Heaven; Hell; Limbo; Purgatory; Revelation §§2-3).)

As well as such very general arguments for life after death, there are some empirical arguments. Eastern religions occasionally produce children who claim to be able to recall events of a previous life, events which they could not have known to have happened by any normal means, but which subsequent investigation reveals to have happened. Some child describes the house at some place where 'he' lived in his previous life, and then such a house is found. But there are doubts about the worth of such evidence. Relatively few children make such claims; the story which they tell about the previous life can be vague; and there are doubts about whether ordinary means of their having obtained the information about that life have really been excluded. Then there are arguments from what mediums declare is told to them, while they are in a trance, by persons already dead. The evidence that those persons give is supposed to be the knowledge that mediums acquire of details of those persons' past lives, which, it is claimed, they could not have acquired by any normal means. But there are similar doubts about the worth of evidence of this kind. The knowledge acquired may be vague and might have been obtained by some normal route. Even if these doubts could be satisfactorily resolved, a hypothesis of some abnormal way of acquiring information about the past might seem less radical than a hypothesis of the present existence of persons who have died. Finally, and most interestingly, there has been the discovery of 'after-death experiences'. Some ten to fifteen per cent of subjects resuscitated after being clinically as good as dead report having had 'transcendental experiences' during the period in which they were virtually dead. These experiences consist of a very vivid awareness of approaching a border or frontier and seeing into a strange, normally immensely pleasing world and then being called back. This coincidence of kinds of experience among people of various backgrounds is interesting; but the experiences have relatively little common detail, so as hardly to provide by themselves very strong evidence that what was observed

was really there. These observations would require support from other considerations. In any case, none of the empirical evidence by itself gives any reason at all for supposing the subsequent existence to be everlasting.

See also: Alexander of Aphrodisias §2; Eschatology; Mendelssohn, M.; Psychē; Salvation; Soul in Islamic philosophy

RICHARD SWINBURNE

References and further reading

Aquinas, T. (1259-65) *Summa contra gentiles* II, trans. J.F. Anderson, *On the Truth of the Catholic Faith*, Garden City, NY: Doubleday., 1956.(Aquinas' account of the nature of 'intellectual substances', that is, souls, and defence of the view that they can be united to human bodies as their forms; see chapters 46-89. Fairly difficult for those unfamiliar with Greek or medieval philosophy.)

Aristotle (*c*. mid 4th century BC) *On the Soul* (*De Anima*) 412a-418a; bks 2-3 trans. and notes by D.W. Hamlyn, Oxford: Clarendon Press, 1968.(Aristotle's account of the nature and attributes of the soul. Fairly difficult for those unfamiliar with Greek philosophy.)

Berkeley, G. (1710) *The Principles of Human Knowledge*, London: J.M. Dent & Sons, 1975.(Section 141 gives Berkeley's argument for the immortality of the soul.)

Butler, J. (1736) *The Analogy of Religion*, London: George Bell & Sons, 1902. (Butler's argument for the reasonableness of belief in life after death; see part 1, chapter 1, 'Of a Future Life'.)

Churchland, P.M. (1984) *Matter and Consciousness*, Cambridge, MA: MIT Press. (Simple and fair introduction to the mind-body problem by a firm materialist.)

Descartes, R. (1641) *Meditations on First Philosophy*, in E.S. Haldane and G.R.T. Ross, *The Philosophical Works of Descartes*, Cambridge: Cambridge University Press, 1972, vol. 1.(Descartes' easy-to-read argument for substance dualism; see particularly meditations 2 and 6.)

Duns Scotus, J. (*c*.1305) *Ordinatio* II d.3, vol. 7 of *Omnia opera*, Vatican City: Scotistic Commission, 1950-. (Scotus' account of the nature of the soul. There is no available English translation.)

Edwards, P. (ed.) (1992) *Immortality*, New York: Macmillan.(A large collection of short excerpts from various writers from Plato onwards.)

Kant, I. (1788) *Critique of Practical Reason*, trans. T.K. Abbott, London: Longmans, Green & Co., 1909.(Kant's argument that we can only make sense of morality if we suppose the soul to be immortal; see book 2, chapter 2, sections 4-5.)

Plato (*c*.386-380 BC) *Meno*, trans. W.K.C. Guthrie, *Protagoras and Meno*, Harmondsworth: Penguin, 1956. (Includes the 'slave-boy' argument.)

Plato (*c*.386-380 BC) *Phaedo*, trans. with notes by D. Gallop, Oxford: Clarendon Press, 1975.(Socrates justifies his view that his soul will continue to live after the death of his body.)

Shoemaker S. and Swinburne R. (1984) *Personal Identity*, Oxford: Blackwell.(A debate between a materialist and a dualist about personal identity.)

Swinburne, R. (1986) *The Evolution of the Soul*, Oxford: Clarendon Press; revised edn, 1997.(A full modern statement of the dualist position.)

South Slavs, philosophy of

Philosophy as a distinct intellectual activity emerged in the coastal towns of the Adriatic during the Renaissance. Philosophers from this area wrote in Latin and taught philosophy in Italy, Germany and Austria. The first popular philosophical works in the vernacular did not appear until late in the eighteenth century, and it was almost one hundred years later that the first chairs in logic and philosophy were founded in the universities. Academic philosophy, usually derived from German or British models, was thus brought into the intellectual life of South Slavs. Local academic philosophers, mostly educated in Germany, brought home a variety of philosophical approaches, and by the early twentieth century many schools of thought flourished: Neo-Kantianism flourished in Beograd, while Neo-Thomism was particularly strong in Ljubljana and Zagreb (until 1941 the Meinongian phenomenological approach also flourished in Ljubljana). In their quest for originality, local philosophers constructed various eclectic philosophical systems.

The Soviet Marxism-Leninism, imposed in all Yugoslav universities after 1945, was replaced in the late 1950s by a Neo-Marxist approach based on the concept of praxis. This was a concept of a purposeful human activity which, among other things, results in social change (for example, the development of socialism). These Neo-Marxists engaged in vigorous debates on the nature of truth and knowledge, on human freedom, alienation, socialism and humanism as well as on social and political issues. Their work reached an international audience through the international journal Praxis*, published in Zagreb, and through their international summer school of philosophy at Korčula. As their support for the Yugoslav communist regime turned into open criticism, the regime halted the publication of* Praxis *and in 1974 forced the Beograd Neo-Marxists from their teaching posts.*

The Neo-Marxist interest in non-Marxist philosophy facilitated the reception of various non-Marxist approaches. From the early 1960s, existentialism, structuralism, hermeneutics and phenomenology - in particular the philosophy of Heidegger - exerted considerable influence on Neo-Marxists as well as on philosophers who had gradually abandoned this approach. In the early 1970s, a younger generation of thinkers in Beograd, Zagreb, Ljubljana and Zadar developed an interest in analytic philosophy. Since that time no approach or trend in philosophy has emerged as the dominant one in any of the academic centres in the countries which were formerly part of Yugoslavia.

1 Philosophy in Latin

Until the Renaissance, philosophy among the South Slavs (on the territory of the former Yugoslavia) was practised only within the theological framework of Roman Catholicism and Eastern Orthodoxy. In the fifteenth century, the Ottoman conquest and the resulting intermittent warfare had greatly impeded intellectual activity. At the same time Renaissance humanism penetrated the coastal towns of the Adriatic (which were not exposed to the Ottoman onslaught). In the Republic of Ragusa (Dubrovnik) especially, an indigenous literature flourished in the local Serbo-Croat dialect, as well as in Italian and Latin. The growth of interest in philosophy nurtured philosophers of South Slav origin who subsequently taught philosophy and published their philosophical works in Latin in various countries of Europe, principally in Italy, Austria and Germany.

Among the first of the South Slav humanist philosophers was Juraj Dragišić (Georgius Benignus de Salviatis), born in Bosnia in 1450 and educated in Dubrovnik, Paris, Oxford and Florence. A Franciscan monk, he was a courtier to Lorenzo di Medici and later a professor of philosophy at the University of Pisa. Philosophically, he was inclined towards Platonism, and in his polemical works he defended Pico della Mirandola as well as Savanarola. His main works were in logic (*Artis dialecticae praecepta vetera ac nova* 1520) and metaphysics (*De natura caelestium spirituum* 1499). The humanist Matija Hvale, born in what is now Slovenia, was active at the University of Vienna in the early sixteenth century.

Matija Vlačić (Mattias Flacius Illyricus) born in Labin in Istria, was educated in Germany where he became one of the leading Lutheran theologians and polemicists of the sixteenth century, firmly opposed to any compromise with Roman Catholicism. In his theological treatise *Clavis Scripturae Sacrae (A Key to the Holy Scriptures)* he not only offered a systematic linguistic exegesis of the Bible, but expounded a succinct philosophy of language and of translation. Language is a sign or image of things, that is, ‘glasses, as it were, through which we look at things’. The meaning of words is determined by the historical context in which they are used, and is thus subject to change.

This contextual view of meaning guides Vlačić's own Biblical exegesis, which Wilhelm Dilthey considered the forerunner of modern hermeneutics (1914-90 (5): 325).

Also in the sixteenth century, Franjo Petrić (Franciscus Patricius) was born in Cres, Dalmatia, educated in Germany and Italy and taught philosophy at the Universities in Ferrara and Rome. His youthful *La citá felice (The happy city)* belongs to the tradition of Renaissance utopianism exemplified by Campanella's *City of the Sun*. His philosophical system, expounded in his *Nova de universis philosophia (The new universal philosophy)*, is an attempt to synthesize Catholicism and Platonism: the light, as a spiritual force, is the first cause, the source of all movement and life in the universe; the spiritual thus permeates the whole universe and in this the human microcosm reflects the universe's macrocosmic order. In this, as in his earlier works, he rejected Aristotle's philosophy as false and godless, which resulted in his work being briefly placed on the Roman Catholic Index of prohibited books.

Rudjer Josip Bošković, (Roger Joseph Boscovich), a native of Dubrovnik who joined the Jesuit order, was educated and taught philosophy in Rome. A scientist and mathematician, he is today best known for his theory of non-extended atoms or *puncta*. In Article 164 of his *Theoria philosophiae naturalis (A Theory of Natural Philosophy)* (1763), he claims that ' Matter is composed of perfectly indivisible, non-extended, discrete points'. Having no mass, these points do not exert force in the Newtonian sense but still interact with each other. They repel each other at very small distances, reaching an equilibrium, beyond which the points attract and then again repel each other. This process is governed by Bošković's law of oscillation. Within the framework of their spatial relations, pairs of oscillating points form minute composite particles. Assemblies of these first-order particles form larger particles which, in turn, form seemingly hard bodies. The collision of impenetrable atoms, he observed, would introduce non-continuous or sudden change in their velocity (Art. 18). In order to avoid this violation of the law of continuity, he conceived atoms as non-extended and impenetrable; in contrast, observable, hard bodies are held to be extended and penetrable. Thus Bošković attempted to uphold the law of continuity which he derived from Leibniz and held to be an axiom. As L.L. Whyte (1961) shows, his deductive kinematic theory of non-extended atoms was highly influential until experimental research on the composition of atoms began early in the twentieth century.

2 Non-academic philosophy in the vernacular

At the end of the eighteenth century, the first philosophical works began to appear in as yet unstandardized South Slav vernaculars. These were works of non-academic philosophy, either popularizing philosophical ideas or presenting them in verse form. Dositej Obradović, a defrocked Serbian Orthodox monk who had attended philosophy lectures at the universities of Halle and Leipzig, was one of the first to popularize the ideas of Enlightenment. But the most original work in the vernacular was a philosophical epic in six cantos, *Luča Mikrokosma (The Ray of Microcosm)* (1845), by Petar Petrović Njegoš, Bishop-Prince of Montenegro. The epic presents a thoroughly dualistic metaphysics and a rather unorthodox cosmogony. To the duality of matter and spirit corresponds the duality of good and evil (although the source of evil is not located exclusively in the matter). The world of light borders on the world of darkness - Hades - which as a sun has its own globe of darkness. God created the heavens, the realm of light, by a stroke of shining light which destroyed the reigning realm of darkness. Satan, a duke of angels, rejected this view of the creation and rebelled against God in the name of the equality of all spirits. Upon his defeat, he was thrown into Hades while Adam (who repented of his association with Satan) and his progeny were settled on Earth, a half-way house between the heavens and Hades. Caged in the body, the human soul is governed by the conflicting laws of good and evil, and is thus condemned to the realm of tears and suffering. Both in its epic style and its pessimistic dualism, this metaphysical poem is unique in South Slav literature and philosophy.

Throughout the nineteenth century, liberal political ideas were disseminated through political pamphlets and popular encyclopedias such as Vladimir Jovanović's incomplete Politički rečnik (Political dictionary) published in Novi Sad, then in Austro-Hungary. Svetozar Marković, a socialist journalist, popularized socialist political ideas (mainly drawn from the works of Chernyshevski and Pisarev) through his newspapers and pamphlets published in Serbia. The future king of Serbia, Petar Karađorđević, while still in exile, published his (now standard) translation of Mill's *On Liberty*.

Božidar Knežević, a poor secondary school teacher, was educated at Beograd, where in 1898 he published his

visionary treatise *Principi istorije (Principles of history)*. Since everything that exists, he argued, exists only in history, history takes over the fields of other sciences and offers the highest human understanding. In addition, history binds all peoples and leads to their reconciliation and overall harmony. Knežević's optimism and belief in the progress of the human mind is tempered with his belief that the total quantity of time available to the living is limited: human civilization and even human life is thus bound to disappear. Proportion, he boldly states, is the telos of history. As both nature and humans strive after this ideal, proportion is used to explain the nature of truth, reason, good, progress, beauty, justice and freedom. Once elements achieve proportion and balance with each other, 'they live simultaneously' in a great organic whole in which one can ultimately arrive at 'complete morality, freedom, justice and truth'. Whereas academic philosophers repudiated this system as incoherent, many Serb avant garde poets and writers found in it a congenial vision of the universe in which everything, including poetry and beauty, had its own rightful place in a world striving after proportion.

3 Academic philosophy before the Second World War

From their inception, the first South Slav institutions of higher learning have established chairs of philosophy or logic: in 1863, the Beograd Great School (since 1905 the University of Beograd), in 1874 the University of Zagreb, and in 1919 the University of Ljubljana. These, as well as the universities which were founded later, have remained centres of academic philosophy.

The first professor of logic in Beograd, Konstantin Branković, based his handbooks dealing with philosophy and logic on the works of the German academic W.T. Krug. His successor, Alimpije Vasiljević, used J.S. Mill's *Logic* for the same purpose. The first philosophy professor at Zagreb, Franjo Marković (a pupil of Herbart) called for the study of Croatian philosophy, arguing that 'only the people to have acquired a homeland of thought have firmly come to own their real homeland'.

Albert Bazala, Marković's successor, was the author of *Povjest filozofije (A history of philosophy)* (1906), one of the first original works in history of philosophy in Serbo-Croat, in which Kant's work is presented as the highest achievement of modern philosophy. He continued Marković's research in the history of South Slav as well as Croatian philosophy. In his *Metalogički korjen filozofije (The metalogical root of philosophy)* Bazala argued that the human mind is not a static faculty but 'a dynamic-teleological activity'. As a dynamic activity, cognition requires that its object be sustained in time and organically related to a whole; therefore, there are no singular truths about discrete objects. In opposition to this, Bazala's colleague Pavao Vuk-Pavlović expounded a dualist theory in his *Spoznaja i spoznajna teorija (Knowledge and theory of knowledge)* (1926), where the truth-value of cognition is found in 'an experience of the logically evident' which may constitute singular truths. The relation between cognition and its objects is a subject-matter not of science but of a meta-empirical or metaphysical investigation. In defending his philosophical humanism, Vuk-Pavlović argued against Bazala's determinism and emphasized the autonomy and creativity of each human being.

France Veber (Franz Weber), the first professor of philosophy at the University of Ljubljana, was also the first academic philosopher to write in Slovene. A pupil of Meinong, his *Sistem filozofije (A system of philosophy)* developed a theory of mental presentation and of objects. The presentations (*Vorstellungen*) are the basic ingredients of all experiences, including the sensory ones. Each sensory experience can be evaluated in terms of its function either of 'attainment' or 'presentation'. The attainment is primary in sensory experiences of touch; in consequence, such experiences are closer to 'encountering reality' than any others. In visual experiences, presentation is primary, and thus reality is represented, not 'encountered'. From 1930 onwards, Veber developed a philosophical anthropology and a philosophy of God as well as a new dynamic ontology, eschewing his earlier dichotomy of object-mental presentation. Several of his students developed his earlier theory of presentation, but as the communist regime removed him from his chair in 1945, his work was ignored in Slovenia until 1987.

Branislav Petronijević from Beograd was educated in Germany and inspired by the writings of Lotze and Von Hartmann. In his principal work *Prinzipien der Metaphysik (Principles of metaphysics)* (1904), he argued that being consists of simple, discrete qualitative points which are mutually related in various ways to form a unity. His metaphysical system was thus conceived as a synthesis of the metaphysics of Spinoza and Leibniz into 'mono-pluralism'. Our immediate experience not only presents reality as is but is also the source of basic logical and metaphysical axioms (for example, the laws of identity, of contradiction and of excluded middle): all are derived from the data of immediate experience. His eclectic 'empirio-rationalist' epistemology was thus based on a

rather complex derivation of logical laws from immediate experience.

Ksenija Atanasijević from Beograd was the first recognized woman philosopher among South Slavs and a leader of Beograd's women's movement. Educated in Paris, she wrote in French and German, producing studies on the philosophy of Epicurus and of Giordano Bruno as well as on the history of South Slav philosophy. Her principal work *Filosofski fragmenti (Philosophical fragments)* (1929), attempted to reconcile her metaphysical belief in the inevitability of evil with a justification of her activist, humanist stance.

Until the outbreak of war in 1941, philosophers at Zagreb, Ljubljana and Beograd pursued a wide variety of philosophical interests. In Ljubljana, the Neo-Thomist approach had been developed since the 1860s. Additionally, philosophers in Zagreb had interest in logical positivism, in Marxism and in the contemporary philosophy of education. In Beograd, philosophers had developed interests in Neo-Kantianism, pragmatism and Bergson's philosophy, as well as in Eastern Orthodox philosophical traditions. Toma Živanović and Đorđe Tasić, Beograd professors of law, wrote in German, French and Italian on the relation of morality and law and on the fundamental grounds of law. Their colleague Slobodan Jovanović critically examined the political philosophy of Plato, Machiavelli, Burke and Marx; his essays, published in the 1930s, are minor classics of Serbian prose. Partly due to this divergence of interests, there was almost no contact or dialogue between these three academic centres during this period.

4 Philosophy after 1945

With the communist takeover in 1945, Soviet Marxism-Leninism became the official philosophy in Yugoslavia. Non-Marxist academic philosophers were faced with conversion to the official philosophy or dismissal. After the expulsion of the Yugoslav Communist Party from the Soviet-dominated international communist organization in 1948, Yugoslav Marxists were called upon to revise the official philosophy. Responding to this call, the post-war generation of Yugoslav philosophers developed a Neo-Marxist approach based on Marx's concept of praxis (mentioned in his early writings) and inspired by the writings of the Frankfurt School members as well as French existentialists (see Existentialism; Frankfurt School). The concept of praxis was chosen partly because it symbolized the Marxist commitment to bringing about social change through purposeful action.

By 1960 the orthodox historical and dialectical materialism was replaced at all Yugoslav universities by this less rigid approach. From then until the early 1980s, Yugoslav philosophers vigorously debated the nature and scope of praxis, of knowledge and truth, of freedom, alienation, socialism and humanism as well as a variety of social and political issues. These debates were carried out not only in the newly founded philosophical journals (*Filozofija* in Beograd, *Pregled* in Sarajevo, *Perspektive* in Ljubljana) but also in the mass media. During these debates the concept of praxis came to be interpreted in the following distinct but ultimately compatible ways: (1) as a free and creative activity which defines a specifically human existence - this primarily anthropological interpretation was first advocated by Gajo Petrović in Zagreb; (2) as an object-directed activity firmly anchored in human history, providing an ontological ground not only for the human universe but also for the world of nature (Milan Kangrga, Zagreb); (3) as a source and criterion of human knowledge from which one could further derive the concepts of a subject and an object (Mihailo Marković, Beograd); (4) as a social activity, aiming at abolishing and transcending the limitations of the present social reality (this axiological interpretation was advocated both by Marković and Petrović).

Yugoslav Neo-Marxists organized an international summer school of philosophy on the island of Korčula and, from 1964, published the international journal *Praxis* in Zagreb. By the late 1960s, they had changed from ardent supporters to open critics of the Yugoslav communist regime; as a result, *Praxis* was forced to cease publication and Beograd Neo-Marxists were forced out of their teaching posts in 1975. This ended the Neo-Marxist domination of Yugoslav philosophy. The dialogue which they had initiated between Yugoslav philosophers from different academic centres - the first such sustained dialogue - ended in 1983 with a quarrel between Neo-Marxists from Zagreb and Beograd concerning the publication abroad of *Praxis International* (this journal ceased in 1994).

As Praxis-Neo-Marxists cultivated interests and contacts outside Marxist philosophy, Yugoslav academic philosophers became acquainted with various non-Marxist approaches to philosophy. Among the first to promote Heidegger's approach to philosophy was Vanja Sutlić from Zagreb. In his principal work *Bit i suvremenost (Essence and contemporaneity)* (1967) he endeavoured to reinterpret the concept of praxis in Heideggerian terms:

praxis was presented as the production of *Sein* from *Dasein* (see Heidegger, M. §4). Abdulah Šarčević from Sarajevo engaged in a multifaceted critique of bourgeois society and of modernity. His *De Homine* (1986) rejected the domination of Western technology-laden thought over non-Western (such as Islamic) approaches to philosophy, and pointed out how, with its own narrow boundaries, scientific and technological progress has failed to humanize our society and has failed to humanize nature. Ivan Foht from Sarajevo developed an aesthetics in which our understanding as well as the ontological status of a work of art depends primarily on its form and not on its content. Art, he argued in his *Savremena estetika muzike (Contemporary aesthetics of music)* (1980) is a highly autonomous human activity; music, in particular, has an atemporal validity. Foht's search for an ontological definition of art, inspired in part by Ingarden, is unique in post-1945 Yugoslav philosophy.

Equally unique is the philosophical development of Mihailo Đurić from Beograd. His work in philosophy started in the early 1950s with studies in Ancient Greek philosophy and the early Wittgenstein. In the 1960s he published in Serbo-Croat a selection of Max Weber's writings, which was highly influential. In the 1970s he repudiated attempts to develop Marxist sociology or economics, arguing that Marx was neither a philosopher nor a scientist but a visionary and a critic of the inhumanity of bourgeois society. About the same time he developed an interest in Heidegger and his thesis on the crisis and the end of European metaphysics. Since the 1980s he has concentrated on the contemporary relevance of Nietzsche's philosophy: as an open workshop of thought, Nietzsche's philosophy points the way out of the crisis of metaphysics and of Western philosophy in general (see Nietzsche, F.).

In the late 1970s, Slavoj Žižek from Ljubljana developed a unique Lacanian psychoanalytic critique of totalitarian ideology and thought (see Lacan, J.). In his later writings, published in French and English, he applied this Lacanian approach to other ideologies, including the liberal-democratic one, and to popular culture and film. In his *Looking Awry* (1991) he advocated the maxim of psychoanalytic ethic 'avoid as much as possible any violation of the fantasy space of the other', and argued that nationalism is the way subjects of a given nation organize their collective enjoyment through national myths.

Interest in logical positivism and the analytic approach to philosophy developed in the 1920s in Zagreb and in Beograd. Zvonko Richtman and Rikard Podhorski from Zagreb defended a logical positivist approach to science (see Analytical philosophy; Logical positivism). In Beograd, Dragiša M. Đurić developed a highly reductive phenomenalism (which reduced the concept of time to the phenomenon of now), while Kajica Milanov discussed Frege's and the logical positivists' analyses of meaning (see Phenomenalism). In the early 1960s, Yugoslav Neo-Marxists revived interest in the analytic philosophy of meaning and meta-ethics within their Neo-Marxist framework, while Jovan Ćulum from Beograd in his *Filozofske beleške (Philosophical notes)* (1967) used analytic techniques to criticize the Hegelian and Marxist approach to philosophy. At the same time, Svetlana Knjazev-Adamović and Staniša Novaković wrote about analytic philosophy of knowledge and of science, while Aleksandar Kron started his work in symbolic logic outside any Marxist framework. In his later *Hipoteze i znanje (Hypotheses and knowledge)* (1984) Novaković argued against the view that the acceptance or rejection of hypotheses in modern science is based exclusively on empirical evidence.

In the early 1970s, analytic philosophers in Beograd engaged in a polemic concerning the analysis of moral terms. Igor Primorac forcefully defended a deontological theory. His arguments, appealing to our moral intuitions and ordinary usage of moral terms, were criticized by Knjazev-Adamović and others. In the early 1990s, she and Jovan Babić extended the debate in the journal *Theoria* to issues in normative ethics, including that of pacifism in times of war. A similar debate concerning scepticism and knowledge was sparked in 1974 by a series of articles by Neven Sesardić from Zagreb. Sesardić rejected both the argument from the evil demon and from illusion as invalid. In response, Aleksandar Pavković in his *Razlozi za sumnju (Reasons for doubt)* (1988) argued that, suitably reconstructed, these sceptical arguments could be rendered valid (see Perception, epistemic issues in). Živan Lazović from Beograd argued for the contextualism of principles in *O prirodi epistemičkog opravdanja* (On the nature of epistemic justification), an alternative to both foundationalism and coherentism which emphasizes the constitutive role of the established epistemic principles, techniques and procedures (see Justification, epistemic).

In his *Fizikalizam (Physicalism)* (1984), the first examination of contemporary materialism in Serbo-Croat, Sesardić defended the mind-body identity thesis, arguing that the replacement of intentional terms by descriptions of behavioural dispositions will facilitate scientific discovery of the relations between behavioural dispositions and

structural states of the brain (see Mind, Identity theory of). In contrast, in his *Materija, svest, saznanje (Matter, mind and knowledge)* (1990) Nikola Grahek from Beograd argued that a materialist view of the mind such as Sesardić's, faces serious problems in explaining introspection of the subject's own mental states. The debate concerning materialism and functionalism, which started in 1980s in the philosophical journals in Beograd, Zagreb and Ljubljana, appears to be continuing (see Materialism in the philosophy of mind).

Writing on the philosophy of language, Nenad Miščević (formerly from Zadar, now at Maribor) shares Sesardić's scientific naturalism. Cognitive science and linguistics, he argued in his 1987 work *Od misli do jezika (From thought to language)*, would provide the best basis for a comprehensive theory of language. Matijaž Potrč from Ljubljana has also extensively discussed current analytic approaches to the philosophy of language. Miloš Arsenijević, in his unique treatment of Zeno's paradoxes and contemporary philosophy of space and time, *Prostor, vreme, Zenon (Space, time, Zeno)* (1986), has attempted to solve the paradoxes by arguing that Zeno's premises are tenable each on their own but are not cotenable.

The break-up of Yugoslavia in 1991 has, perhaps temporarily, put an end to the debates among philosophers from Zagreb, Beograd, Ljubljana and Sarajevo. Many of them, however, continue to participate in philosophical debate outside their newly established states.

See also: Marx, K.; Marxist philosophy, Russian and Soviet

ŽIVAN LAZOVIĆ
ALEKSANDAR PAVKOVIĆ

References and further reading

Arsenijević, M. (1986) *Prostor, vreme, Zenon (Space, time, Zeno)*, Beograd and Zagreb: Filozofsko Društvo Srbije & Liber.(A systematic examination of various attempts to solve Zeno's paradoxes, arguing, at the end, that Zeno's premises are not cotenable.)

Arsenijević, M. (1989) 'How many physically distinguished parts can a limited body contain', *Analysis* 1 (49): 36-42.(An exercise in a metaphysical speculation over divisibility of bodies.)

Arsenijević, M. (1993-4) 'Mathematics, Infinity and the Physical World', *Dialektik* 89-107.(A new approach to the ancient question of the relation of mathematics to reality.)

Atanasijević, K. (1923) *The Metaphysical and Geometrical Doctrine of Bruno as Given in His work 'De Triplici Minimo'*, trans. and intro. G.V. Tomashevich, St Louis, MO: Warren H. Green, 1972.(Attempt at a modern interpretation of Bruno's work.)

Atanasijević, K. (1929) *Filosofski fragmenti (Philosophical fragments)*, Beograd: Geca Kon.(A humanist approach to various questions of normative ethics.)

Babić, J. (1992) 'Primenjena etika I' (Applied ethics I), *Theoria* 34 (1): 97-105. (The debates on this topic, continued in the following issue.)

Bazala, A. (1906) *Povjest filozofije (A history of philosophy)*, Zagreb: Liber, 1989, 2nd edn. (A Kantian exposition of the history of philosophy with Kant's philosophy presented as the highest achievement of Western Philosophy.)

Boscovich (Bošković), R.J. (1763) *Theoria philosophiae naturalis: A Theory of Natural Philosophy*, trans. J.M. Child (1921), London: MIT Press, 1966.(Contains a useful introduction by the translator.)

Campanella, T. (1602) *La Città del Sole: Dialogo Poetico (The City of the Sun: A Poetical Dialogue)*, trans. with notes by D.J. Donno, Berkeley and Los Angeles, CA, and London: University of California Press, 1981.(A scientifically ordered utopia presided over by metaphysicians and scientists.)

Ćulum, J. (1967) *Filozofske beleške (Philosophical notes)*, Beograd: Nolit.(Elegant essays in analytic philosophy, arguing for the untenability of Hegelian metaphysics and its Marxist derivatives.)

Dilthey, W. (1914-90) *Gesammelte Schriften (Collected Works)*, Göttingen: Vandenhoeck & Ruprecht, 20 vols. (The standard complete works.)

Dragišić, J. (Salviatis, Georgius Benignus de) (1499) *De natura caelestium spirituum (On the nature of celestial spirits)*, Florence: Bartolommeo di Libri.(A Neoplatonist treatise of metaphysics. His 1520 work *Artis dialecticae praecepta vetera ac nova (Principles old and new of the dialectic art)* was a discourse on logic and rhetoric.)

Đurić, M. (1985) *Nietzsche und die Metaphysik (Nietzsche and metaphysics)*, Berlin and New York: de Gruyter. (Holds that Nietzsche's philosophy points the way out of the crisis in contemporary Western metaphysics.)

Đurić, M. and Žunjić, S. (eds) (1993) *Die Serbische Philosophie Heute (Serb philosophy today)*, Munich: Slavica

Verlag Dr Anton Kovač.(A collection of essays by eleven Serb philosophers, with a detailed historical introduction.)

Foht, I. (1980) *Savremena estetika muzike (Contemporary aesthetics of music)*, Beograd: Nolit.(A collection of essays on the main trends in contemporary aesthetics of music written from a phenomenological point of view.)

Grahek, N. (1990) *Materija, svest, saznanje (Matter, mind and knowledge)*, Beograd: Filozofsko društvo Srbije. (A systematic refutation of various contemporary materialists' arguments, presenting an argument for the irreducibility of the mental.)

Grahek, N. (1995) 'The Sensory Dimension of Pain', *Philosophical Studies* 79: 167-81.(An argument for the irreducibility of the sensory content of pain feelings.)

Kangrga, M. (1966) *Etika i sloboda (Ethics and freedom)*, Zagreb: Naprijed.(An influential Neo-Marxist approach to ethics.)

Knežević, B. (1898) *Principi istorije (The Principles of History)*, Beograd: B. Knežević, 2 vols.(A highly speculative attempt at the formulation of the most general laws of history.)

Knežević, B. (1921) *History, the Anatomy of Time; the Final Phase of Sunlight*, trans. and intro. G.V. Tomashevich, New York: The Philosophical Library, 1980.(Translations from his *Dva zakona* first published in 1921.)

Lazović, Ž. (1994) *O prirodi epistemičkog opravdanja (On the nature of epistemic justification)*, Beograd: Filozofsko društvo Srbije.(A defence of a contextualist theory of epistemic justification.)

Marković, F. (1881) 'Filosofijske struke pisci hrvatskog roda s onkraj Velebita u stelječih XV. do XVIII' ('The philosophical writers of Croatian birth from the other side of Velebit from the fifteenth to the eighteenth century') in *Izvještaj Sveučilišta u Zagrebu (The Annual Report of the University of Zagreb)*, Zagreb: Sveučilište u Zagrebu, 1882 (reprinted in *Prilozi za istraživanje hrvatske filozofske baštine (Contributions to the study of the Croatian philosophical heritage)*, 1 (1-2, 1975): 255-81.(Marković's speech as Rector of the University of Zagreb in 1881, with a short history of philosophy of the Adriatic littoral.)

Marković, M. and Cohen, R.S. (1975) *Yugoslavia: The Rise and Fall of Socialist Humanism. A History of the Praxis Group*, Nottingham: Bertrand Russell Peace Foundation.(A polemical account of the Praxis group, co-authored by one of its members.)

Miščević, N. (1987) *Od misli do jezika (From thought to language)*, Rijeka: Dometi.(An examination of contemporary analytic theories in the philosophy of language.)

Miščević, N. (1996) 'Computation, Content and Cause', *Philosophical Studies* 2 (82): 241-63.(A discussion of intentionality and causation.)

Njegoš, P.P. (1845) *Luča Mikrokozma - The Ray of Microcosm*, trans. A. Savić-Rebac, Beograd: Vajat, 1989. (This translation and foreword were originally published in *Harvard Slavic Studies* 3 (1957).)

Novaković, S. (1984) *Hipoteze i saznanje (Hypotheses and knowledge)*, Beograd: Nolit. (Scientific hypotheses, it argues, are not rejected or accepted on the basis of empirical evidence alone.)

Pavković, A. (1988) *Razlozi za sumnju (Reasons for doubt)*, Beograd: Istraživačko-izdavački centar SSO.(A reconstruction of arguments from illusion, from dreams and from the evil daemon.)

Pavković, A. (ed.) (1988) *Contemporary Yugoslav Philosophy: The Analytic Approach*, Dordrecht and New York: Kluwer.(Seventeen essays by Yugoslav analytic philosophers; includes a short historical introduction and a select bibliography.)

Pavković, A. (1990) 'Two Thaws in Yugoslav Philosophy', in M. Pavlyshyn (ed.) *Glasnost in Context*, Oxford: Berg.(A short outline of post-1945 Yugoslav philosophy in its political context.)

Petrić, F. (Patricius, Franciscus) (1593) *Nova de universis philosophia* (The new universal philosophy), Venice: NP.(A synthesis of Catholicism and Platonism. *La citá felice*, by the same author, described a utopia in the tradition of Campanella's *City of the Sun*.)

Petronijević, B. (1904) *Prinzipien der Metaphysik (Principles of metaphysics)*, Heidelberg: C. Winter.(Holds that the universe consists of simple, discrete qualitative points and our immediate experience is the source of basic logical and metaphysical axioms.)

Petrović, G. (1967) *Marx in the Mid-Twentieth Century*, Garden City, NY: Anchor Books. (An unorthodox interpretation of Marx by a prominent Praxis philosopher from Zagreb.)

Petrović, G. and Marković, M. (eds) (1979) *Praxis: Yugoslav Essays in the Philosophy and Methodology of the Social Sciences*, Dordrecht: Reidel.(A selection of Neo-Marxist papers from the journal *Praxis* dealing with the philosophy of social sciences.)

Potrč, M. (1992) 'The Sensory Basis of Content in Veber', *Slovene Studies* 13 (1): 71-90.(A systematic examination of Veber's theory of representation.)

Primorac, I. (1978) *Kazna i prestup* (Punishment and offence), Beograd: Mladost.(A defence of a deontological

and retributivist theory of punishment.)

Šarčević, A. (1986) *De Homine: Mišljenje i moderni mit o čovjeku (On man: thought and the modern myth about man)*, Sarajevo: Veselin Masleša.(A critique of Western philosophy, science and technology and its failure to humanize society.)

Sesardić, N. (1984) *Fizikalizam (Physicalism)*, Beograd: Istraživačko-izdavački centar SSO. (The first defence of the mind-body identity thesis in Serbo-Croat.)

Sesardić, N. (1993) 'Heritability and Causality', *Philosophy of Science* 3 (60): 396-418.(A discussion of recent theories of heritability and causality.)

Stojanović, S. (1973) *Between Ideals and Reality: A Critique of Socialism and its Future*, New York: Oxford University Press.(A prominent Beograd Praxis philosopher presents his vision of socialism.)

Sutlić, V. (1967) *Bit i suvremenost (Essence and contemporaneity)*, Sarajevo: Veselin Masleša.(An early Heideggerian interpretation of Marxist concept of praxis.)

Urbančič, I. (1993) 'Philosophy with the Slovenes', *Nationalities Papers* 21 (1): 127-137.(Thorough post-communist account of the Slovene philosophy.)

Vlačić, M. (Flacius Illyricus, Matthias) (1567) *Clavis Scripturae Sacrae (A Key to the Holy Scriptures)*, Basel: Per Iaonnem Oporinum et Eusebium Episcopium.(A systematic exegesis of the Bible based on a philosophy of language anticipating modern hermeneutics.)

Vuk-Pavlović, P. (1926) 'Spoznaja i spoznajna teorija' ('Knowledge and theory of knowledge') in *Duševnost i umjetnost*, Zagreb: Liber, 1989.(A call for a meta-empirical or metaphysical investigation of the relation of cognition and its objects.)

Weber (Veber) F. (1921) *Sistem filozofije (A system of philosophy)* Ljubljana: K&B.(A system of philosophy based on a complex theory of mental presentation of objects.)

Weber (Veber) F. (1987) 'Empfindungsgrundlagen der Gegenstandstheorie' (The experiental grounds of the theory of objects), *Conceptus* 21: 75-87.(A lecture on the Meinongian theory of objects presented in Graz in 1954.)

Whyte, L.L. (ed.) (1961) *Roger Joseph Boscovich, S.J., F.R.S., 1711-1787*, London: Allen & Unwin.(Very good essays on his philosophical and scientific contributions; comprehensive bibliography.)

Žižek, S. (1991) *Looking Awry: An Introduction to Jacques Lacan through Popular Culture*, London: MIT Press. (A humorous and Lacanian examination of various manifestations of popular culture, including film.)

Sovereignty

In legal and political philosophy sovereignty is the attribute by which a person or institution exercises ultimate authority over every other person or institution in its domain. Traditionally, the existence of a final arbiter or legislator is said to be essential if people are to live together in peace and security. The example brought most readily to mind by the word 'sovereign' is the individual monarch, and the theory of sovereignty was at one time closely linked with the defence of monarchy. But leading theorists of sovereignty, like Jean Bodin and Thomas Hobbes, recognize that authority can be exercised by sovereign bodies of people; and later writers, like Rousseau and Austin, locate sovereignty in the people, to whom the officials of more democratic institutions are ultimately accountable.

Traditionally, too, it is deduced from the nature of the state or law that the sovereign's authority must be absolute, not limited by conditions; perpetual, not merely delegated for a time; and indivisible, not distributed between different persons or institutions. It is further deduced that the sovereign must be independent from external domination as well as internally supreme. All these inferences have been subjected to criticism, not least because they can be difficult to reconcile with the actual practice of states and legal systems.

1 Sovereignty and utility

It seems to have been no coincidence that the two foundation texts of the modern theory of sovereignty were both published in conditions of civil conflict, Jean Bodin's *Six livres de la république* (1576) at the height of the French wars of religion, Thomas Hobbes' *Leviathan* (1651) in the immediate aftermath of the English Civil War (see Bodin, J. §2; Hobbes, T.). For the early appeal of the concept delineated in these works lay largely in its apparent solution to the problem of anarchy. Hobbes' treatment is particularly illuminating in this respect. He begins by arguing that given certain features of human nature the life of people outside the commonwealth is bound to be, as he famously put it, solitary, poor, nasty, brutish and short. To live in peace and security people have to submit to a power that is able to compel them to fulfil their obligations.

To begin with people live in small family units, but their existence at this stage is characterized by internecine feuds, not by peace and security. The emergence of more powerful groupings of individuals or families may achieve a greater measure of stability, but Hobbes maintains that however large and powerful these groups may be they cannot ultimately enable people to live peacefully and securely. The balance of power may shift through regrouping, and within each group people will continue to pursue their individual interests. The appearance of a common enemy may draw them together, but only temporarily, and even then they will argue about the best use of their resources. In short, the collective life of people will of itself be no different from the solitary life.

To live in peace and security people need to establish a commonwealth, and for Hobbes that means submitting to a 'Soveraigne Power'. Each subject individual must surrender their personal power to do what they think best for their own preservation and accept the determination of another, until one person alone is left with the authority to decide what should be done. That person, whether an individual or an assembly of individuals, will then have the sovereign power necessary to coerce the rest of the people into fulfilling their obligations. The person may in fact have been the head of a family or larger grouping, but their judgment will now be final and decisive, and their exercise of power will not be limited to the duration of an emergency. Their sovereignty will continue for as long as they are able to protect their people, and the people will have peace and security for as long as they have a sovereign.

2 Sovereignty and logic

If it was this sort of argument for the expediency of submitting to a final arbiter that first attracted exponents of sovereignty, and if this sense of the practical necessity of an ultimate authority has remained important throughout the history of the concept, a more prominent feature of its presentation has been an emphasis on logical analysis. Hobbes himself was committed to a mode of exposition that starts with the definition of a term, then proceeds to explore its implications. So from the definition of sovereignty it may be deduced that it must be absolute and indivisible. It must be absolute because conditions could only be meaningfully imposed on a sovereign's exercise of power if there were some other person with authority to determine when they were being violated, in which case

the sovereign would not be the final arbiter. Similarly sovereignty must be indivisible because otherwise there might be disagreements between the different bearers of authority, in which case there would again be no final arbiter.

For Hobbes the definition of the sovereign's authority is effectively the definition of the state. So too for Bodin the concept of absolute and indivisible sovereignty is logically entailed by the very idea of the state, which he defines as the just government, with sovereign power, of several families and their belongings. Bodin claims to have broken new ground by providing a clear analysis of the sovereign authority that must be inherent in any genuine state. He identifies two aspects in particular. First, sovereignty must be perpetual and not the kind of power that is entrusted to someone for a time by people who may take it back again. There cannot be a state where authority is merely delegated to the ruler, and where there is a state it is the fact that lesser magistrates have only delegated authority that distinguishes them most clearly from the sovereign. Second, Bodin goes to greater length in stressing that sovereignty must be absolute and not the kind of power that is hedged in with obligations and conditions. In essence, he concludes, sovereignty is the power to lay down laws that will bind the subjects in a commonwealth but not the sovereign themselves. The sovereign must be able to legislate without their subjects' consent, and cannot be bound by the laws of predecessors, still less by their own laws. It would be logical nonsense to say that a sovereign legislator was bound by the laws of a sovereign.

Bodin declares that law is nothing else than the command of the sovereign in the exercise of sovereign power, a view maintained more recently by analytical jurists in an effort to anatomize law properly so called (see Legal positivism §§1-2). In *The Province of Jurisprudence Determined* (1832) John Austin defines the positive laws of men as the general commands of sovereigns backed up by the threat of sanctions (see Austin, J.). Laws emanate from a sovereign power whenever the majority of the people in a reasonably populous society are in the habit of obeying the legislator, and the legislator in turn is not in the habit of obeying anyone else. Expressed like this the concept of sovereignty has purely analytical significance, and Austin makes no attempt in his legal theory to explain why or when a legislator should be obeyed. He is content to specify that there must be a 'habit' of obedience on the part of the 'bulk' of the people to a superior who is both 'determinate' and 'common' to them all.

Austin also reinforces the point that the legislator must not themselves be in the habit of obeying anyone else. This standard assertion made in one form or another that sovereignty is logically illimitable has attracted widespread criticism. It can seem contradictory for Bodin to write as he does of the sovereign being bound to legislate in accordance with divine and natural laws. If one response is to agree with Austin that this type of requirement imposes moral, not legal, constraints, the same cannot so easily be said with respect to Bodin's claim that the sovereign should also legislate in accordance with the fundamental laws of the state, or to Hobbes' concession that the subject need not obey a sovereign intent on his destruction. It may instead be replied that the legislator is sovereign only so long as they act like a sovereign. But who must decide when they do not, and on what basis? Must it then be said that the state and the law have ceased to exist? To many critics it seems more plausible to recognize that legislators in theory may and in practice often do operate under constitutional constraints which are both fundamental and properly legal.

3 Sovereignty and republicanism

Another standard assertion that has been widely criticized is that sovereignty is logically indivisible. Initially the assertion was made in opposition to the republican ideal of a blend of monarchic, aristocratic and democratic modes of government (see Republicanism). While Bodin and Hobbes do not exclude the possibility of genuinely sovereign rule by popular assemblies, instead merely exhibiting a preference for monarchy, they consider it absurd to talk of a mixed constitution. Restated in the juridical language of sovereignty the republican theory of balanced interests does seem rather less coherent. Yet it is equally apparent that the theory of sovereignty is difficult to apply to the practice of states in which legislative authority is distributed among distinct institutions, particularly where the structure of the state is federal.

The author who did most to reconcile the languages of sovereignty and republicanism was Jean-Jacques Rousseau, writing in the century between Hobbes and Austin. Although Rousseau does not entirely agree with Hobbes' pessimistic account of the nature of human beings, he does accept that discord is an inevitable consequence of the human condition and that the solution lies in the establishment of civil societies. Indeed, for Rousseau, people do not surrender their freedom on entering into civil society but find there the possibility of fulfilment in a life of

virtue and genuine liberty. In his *Du contrat social* (1762) Rousseau makes the famous observation that people are born free but are everywhere in chains, before going on to show how that bondage might be avoided in a properly constituted republic. The crucial requirement is that people must not transfer their sovereignty to the ruler but must retain it for themselves. By adopting the paradoxical position that the people as sovereign legislate for themselves as subjects, Rousseau is able to argue that laws will never be oppressive when enacted by the general will.

In most respects the concept of sovereignty deployed here remains traditional. In particular, sovereign power is still conceived to be absolute. Though Rousseau admits that it must be exercised in keeping with natural law (see Natural law §4) and for the common good, both concessions are consistent with the usual qualifications. In the first case, Rousseau denies that there can be any judge of the conformity of legislation with natural law other than the general will itself. In the second case, he takes the failure to legislate for the common good to show not so much that a sovereign people has exceeded the limits of its authority as that the law is not the enactment of a sovereign people after all. For the general will is not to be equated with the will of the majority, nor even with the will of all. The general will consists in an interpretation of the common good of the society, not in an amalgamation of individual interests.

Sovereignty thus remains unlimited in the usual sense, and Rousseau also understands it to be indivisible. However, now that legislative sovereignty remains with the people the concept no longer excludes the possibility of mixed government. While the people must assent to legislation for the common good, they can be governed by separate agencies which may take any form, though some forms of government will be less likely to engender corruption than others. A good ruler will frame sound laws to introduce for the people’s assent, and by enforcing sound laws will help to shape and preserve a virtuous people capable of exercising the general will. The good ruler will not attempt to encroach on the people’s sovereignty, though Rousseau accepts that in practice this will be likely to happen.

Accordingly, if the problem with Rousseau’s version of sovereignty is that it will seldom exist in reality, it does have the advantage that it could in principle be discovered in any state, including one with a mixed constitution or federal structure. Moreover, the problem can be avoided and the advantage retained if the notion of the general will is abandoned. Austin, for example, is able to identify undivided sovereigns in complex states by focusing on the electors of representative assemblies rather than on the assemblies themselves. Yet the difficulty that still remains is the need to regard the same persons as both rulers and ruled. On this analysis, sovereignty ceases to be an attribute of those who appear to legislate and rule and becomes instead an abstract quality of those who appear to obey and be ruled but who are in fact the ultimate source of authority. As such, the concept of sovereignty must compete with other attempts to explain what it is that gives unity and integrity to states or legal systems.

4 Sovereignty and independence

On the sovereignty model the unity and integrity of states or legal systems is taken to depend not only on the absence of an internal or domestic rival to the unity and superiority of the sovereign, but also on the absence of an external or foreign rival. Where external independence is regarded as a further facet of the internal superiority of the sovereign, the constraints of international law can no more be considered properly legal than the constraints of natural law, and international institutions can at best be considered to have delegated authority. With corollaries like these, the connection between sovereignty and independence has understandably been questioned, especially in an era like the present in which the legitimacy of state activity is increasingly assessed according to the standards of wider communities. Keeping the two aspects of sovereignty distinct is therefore urged as an essential means of avoiding confusion, yet what can scarcely be denied is that sovereignty has historically been associated with the idea of the independent state.

Historians have found the earliest expression of the concept in the observations of the Roman jurist Ulpian, that the *imperium* of the Roman people was transferred to the emperor, that the emperor was not bound by the laws, and that what pleased the emperor had the force of law. But although these statements were not unknown in medieval Europe, it was not until the sixteenth century that the theory of sovereignty was fully developed. Consequently it may be said that the notion of absolute, perpetual and indivisible sovereignty made its first appearance under a monolithic empire and reappeared only with the emergence of the modern nation state. If this is correct, it may well transpire that the concept has had its day. If the idea of the nation state continues to decline as a focus for legal and political identity, the concept of sovereignty seems likely to decline with it.

See also: Absolutism; Authority; Constitutionalism; De; General will; Law, philosophy of; Rule of law (Rechtsstaat); ShintōČsect;5; Shōtoku Constitution; State, the

J.D. FORD

References and further reading

Austin, J. (1832) *The Province of Jurisprudence Determined*, H.L.A. Hart (ed.), London: Weidenfeld & Nicolson, 1954.(Deals with sovereignty especially in Lecture VI.)

Bodin, J. (1576) *Six livres de la république (Six Books of a Commonweal)*, trans. R. Knolles, 1606; K.D. McRae (ed.), Cambridge, MA: Harvard University Press, 1962.(Deals with sovereignty especially in Book I, chapters 8-10.)

Hinsley, F.H. (1986) *Sovereignty*, 2nd edn, Cambridge: Cambridge University Press. (Surveys the history of the concept while presenting a historical thesis about the conditions for its emergence; also contains a useful bibliography.)

Hobbes, T. (1651) *Leviathan*, R. Tuck (ed.), Cambridge: Cambridge University Press, 1991.(Deals with sovereignty especially in Part II.)

Jouvenel, B. de (1957) *Sovereignty*, Cambridge: Cambridge University Press.(Develops a philosophical thesis about the place of sovereignty in political science, providing historical observations along the way.)

King, P. (1974) *The Ideology of Order*, London: Allen & Unwin.(Like most monographs on the leading authors referred to, comments on their conception of sovereignty; this study of Bodin and Hobbes pays special attention to the theme.)

Merriam, C.E. (1900) *History of the Theory of Sovereignty since Rousseau*, New York: Columbia University Press.(Provides a more detailed survey than Hinsley of the history of the theory in the nineteenth century.)

Rousseau, J.-J. (1762) *Du contrat social (The Social Contract)*, trans. M. Cranston, Harmondsworth: Penguin, 1968.(Reflects on sovereignty.)

Space

In some of its uses, the word 'space' designates an empty or potentially empty expanse among things, for example, when a driver finds a space in a crowded parking lot, or when a typesetter increases the space between words on a page. In other uses, 'space' is meant to stand for a boundless extension which supposedly contains everything, or every thing of a certain sort. The former sense is well-grounded in ordinary experience and can be traced back to the etymology of the word (from the Latin word spatium, *meaning 'race-track', or generally 'distance', 'interval', 'terrain'). The latter sense originated in scholarly circles - possibly as late as the fourteenth century - by a bold extrapolation of the former; it does not refer to anything that can be exhibited in sense-perception; and yet, through the influence of Newtonian science on Euro-American common sense, it has become so entrenched in ordinary usage that it is normally viewed as the primary meaning of 'space', from which all others are derived.*

According to Cornford, the 'invention of space' as a boundless, all-encompassing container happened in the fifth century BC*. However, it is more likely to have occurred in the late Middle Ages. At any rate, the idea was rampant in Cambridge in the 1660s, when Newton made it a fundamental ingredient in his framework for the description of the phenomena of motion. In a posthumous paper, Newton stressed that space evades the traditional classification of entities into substances and attributes, and has 'its own manner of existence'. Until the publication of this paper in 1962, philosophers took Newtonian space for a substance, and most of them thought this to be utterly absurd. In view of the role of all-encompassing space in Newtonian physics, Kant opted for regarding it as a precondition of human knowledge, contributed once and for all by the human mind. Newton had written that the points of space owe their individual identity to the relational system in which they are set. Nineteenth-century mathematicians vastly extended this concept of space by conceiving many such relational systems. They thus made it possible for relativity theory to substitute four-dimensional spacetime for Newtonian space and time, and for current string theory to countenance a ten-dimensional physical space. These developments confirm the productivity, but not the fixity, of the knowing mind.*

1 The emergence of space

The invention of space as a boundless, all-encompassing expanse was traced back by Cornford (1936) to the fifth century BC. He argued that Greek geometry, which developed at that time, presupposed such an expanse, and that this was identical with the void postulated by the atomist philosophers. However, the atomists' void, though boundless, is not an expanse that *contains* everything, but rather the nothingness *outside* all things. The relation between space in this sense and Greek geometry is less straightforward. The Greeks conceived of geometry as a study of figures, their properties and relations, not as the science of space. Greek geometrical discourse involved the tacit assumption - eventually expressed in Euclid's postulate - that any two coplanar straight lines meet at a point unless they are both perpendicular to a third straight line. Obviously, if two coplanar straights are almost perpendicular to a third one, their intersection can be enormously far away, and an infinite expanse is needed to make room for all such intersections. However, Aristotle did not see this requirement as a constraint on physical reality; he argued that such an infinite expanse is inconceivable and shrewdly remarked that any construction involved in the solution of a geometrical problem can be made to fit - by rescaling - within the finite expanse beneath the firmament (inasmuch as 'any magnitude whatsoever can be divided in the same ratio as the greatest magnitude' (*Physica* 207b31-2)).

In the late Middle Ages, the idea of space springs up in several writers in connection with God's omnipotence. Catholic authorities condemned the proposition that 'God cannot move the firmament in a straight line'. Yet most theologians were loath to admit the existence of a void beyond the firmament. Thomas Bradwardine came up with the idea of an 'infinite imaginary vacuum' which God can make real (in part), should he decide to move the world. Nicole Oresme also speaks of an 'imaginary infinite and motionless space outside the world' (*c*.1375: 368), but he asserts its actual existence: 'Outside the firmament there is a bodyless empty space in a manner that differs from that of any full and corporeal space' (*c*.1375: 176). According to Oresme, the speed of moving bodies 'is measured in regard to' the said 'motionless imaginary space' (*c*.1375: 372).

By 1600, space had become a familiar ingredient of natural philosophy. In Bruno's words:

> Space is a continuous three-dimensional natural quantity, in which the magnitude of bodies is contained, which

is prior by nature to all bodies and subsists without them but indifferently receives them all and is free from the conditions of action and passion, unmixable, impenetrable, unshapeable, non-locatable, outside all bodies yet encompassing and incomprehensibly containing them all.

(1591: 1.8)

2 Newtonian space

Newton's idea of space stems from this tradition, but it also benefited from the conversion of geometry into a science of space by Descartes. Geometrical figures were henceforth conceived as sets of spatial points meeting specific conditions, each point being identified by its coordinates, that is, its (oriented) distances from three pairwise intersecting infinite planes. By bestowing on the class of oriented lengths on a straight line the structure of an algebraic field (actually indistinguishable from our real number field $\mathfrak{R}$), Descartes managed to express geometrical properties and relations by systems of algebraic equations. He maintained that bodies are really nothing but their extension, because their geometrical properties and relations are the only ones that we can grasp clearly and distinctly (see Descartes, R. §5). But geometry was still too weak to meet the demands of a total geometrization of physics, and Newton began his career as a physicist by 'doing away' with Descartes' 'fictions' (1962: 92). He neatly distinguished bodies from the infinite space in which God created them, which is a repository of points that satisfy the theorems of Euclidean (that is, Cartesian) geometry.

Newton sought to ascertain the forces of nature by studying the motions of bodies in space. The motion of a body can be accurately described by means of the varying coordinates of some suitably chosen points in it. The set of three planes from which such coordinates are measured defines a reference frame. Newton was aware that any usable reference frame must be anchored to a rigid body. But a description of motions referred to such a frame cannot exhibit real forces unless that body rests in space or is known to move in it in a definite way. Newton acknowledges two sorts of force: the force of inertia inherent in each body causes it to move in space with constant velocity in a straight line; and the impressed forces acting on a body cause a change - an acceleration - in its state of motion. An inertial frame is a reference frame anchored to a body free from impressed forces. Newton assumes that every impressed force issues from a body and that if a body *A* acts upon a body *B* with a force f, *A* is simultaneously acted upon by a force issuing from *B*, opposite to f but equally strong. In principle, this assumption should enable one to pick out the inertial frames. The very geometry of space makes it impossible to identify an inertial frame as being at rest, but this has no effect on the study of impressed forces, for the acceleration they cause on a given body displays the same magnitude and direction in every inertial frame, regardless of its state of motion (see Mechanics, classical §2; Newton, I. §3).

3 Philosophies of space

What *is* space? The successes of Newtonian physics forced philosophers to face this question, although it proved impossible to bring space under one of the traditional metaphysical categories. Much needless discussion might have been forestalled had Newton's manuscript 'De gravitatione et aequipondio fluidorum' (On the Gravity and Equilibrium of Liquids) not remained unpublished until 1962. In it he boldly asserts that space is neither a substance nor an attribute of a substance, but has 'its own manner of existence' (1962: 99, 132). There follows a characterization which is reminiscent of Bruno, but which includes an unprecedented feature: according to Newton, each point of space is the particular point it is by virtue of the relations it has to the other points, and the only source of its individuality (*individuationis principium*) is the post it holds in the system of such relations. Thus, Newton's concept of space provides the prototype for what is now known as a (categoric) mathematical structure, which can be roughly described as a collection of objects fully specified by a list of mutual relations.

A conception of space as a purely relational system or mathematical structure was also put forward by Leibniz in his polemic against the view of space as a substance, which he imputes to Newton. Leibniz characterizes space as the abstract order of coexisting things. If we forget the peculiarities of each thing and retain only its '*situation* or distance' to the other things we obtain the notion of the thing's place, which may be taken by anything. 'And that which comprehends all those places, is called *Space*' (Leibniz 1716: §47). But Leibniz abides by the traditional view (rooted in the medieval notion that God must be capable of annihilating any particular creature without essentially affecting the others) according to which everything that exists is either a substance or an attribute of a substance. Since space is neither, he maintains that it is no more than a well-grounded phenomenon, lacking

genuine reality (see Leibniz, G.W. §11). From then on, the debate raged between Leibnizian relationalists and 'Newtonian' substantialists, the former insisting that space must be based on and can only be known through the relations among bodies, the latter pointing out that the one scientifically effective description of bodies and their motions - namely, Newton's - required a logically and ontologically independent space. In particular, Berkeley maintained that space without bodies, 'infinite, immoveable, indivisible, insensible, without relation and without distinction', was 'mere nothing' (Berkeley 1721: §53). On the other hand, Euler (1748), though opposed to Newtonian action-at-a-distance, argued for absolute space as a condition of scientific dynamics.

The tension between relationalism and substantialism is clearly visible in Kant's philosophical development. In his first book he ventured a Leibnizian view based on Newtonian physics: space has three dimensions because bodies attract each other with a force proportional to the inverse square of their distances; if matter had different dynamical properties, space would sport 'other properties and dimensions'; the 'highest geometry' would deal with 'all these possible kinds of space' (Kant 1746: §10). But in 1768 Kant alighted on an idea which, he thought, implied the ontological priority of space over bodies. Take two bodies which are exact mirror images of each other (say, a right and a left shoe, or two molecules of, respectively, laevorotary and dextrorotary glucose). An exhaustive description of the mutual distances between all the component parts of one such body would be identical to a similar description of the other. Therefore, the intrinsic difference between them must depend on the relation of each to space. Thus, space is involved in the constitution of each body. Yet space does not meet the traditional requirements of thinghood: its actual existence would require the 'complete synthesis' of infinitely many parts - an impossible task. Moreover, each part of space is infinitely divisible into further parts, so when all relational bonds are removed no substantive foundation remains on which they might rest. Kant solved this philosophical difficulty - and many others - by conceiving space as a 'form of sensibility': 'not something objective and real, not a substance, nor an attribute, nor a relation, but a subjective and ideal sort of scheme for mutually coordinating all external sensations in every way, which issues from the nature of the mind according to a stable law' (Kant 1770: §15D). Since bodies depend on space for their very being, they do not exist in and of themselves, apart from the mind, which is the source of spatial order; they are mere appearances of an underlying reality, which eludes our senses. In the *Critik der reinen Vernunft (Critique of Pure Reason)* (1781, 1787), Kant acknowledges that such underlying reality, though undeniable, is also inaccessible to our intellect, which can only be guaranteed to work within the bounds of sense experience. Space is now described with greater precision as a 'form of appearance', a condition imposed by the mind on sense impressions which *allows* them to be ordered in certain relations by the understanding and thereby to be grasped as presentations of objects. This provides a welcome explanation of the purported fact - which Kant took for granted - that all bodies comply with the truths of geometry and that these truths can be known a priori, ahead of experience. For, as Kant saw, geometric theorems do not simply flow from definitions and the truths of logic (they are 'synthetic', not 'analytic', statements) (see Kant, I. §§4-5).

Although Kant was the most influential philosopher of modern times, his doctrine of space was not adopted by any major thinker except Schopenhauer. Thanks to Kant's critique, the rigid substance- attribute ontology carried no weight in the nineteenth century. Therefore, the primacy of space over bodies no longer appeared to threaten reason and even morality, as it had in the eyes of Leibniz, Berkeley and Kant. But the formulation of new geometries that contradicted the Euclidean canon (see §4 below) undermined Kant's view of geometry as a non-empirical science providing the descriptive framework required for empirical research. Kant's view is not endangered by the existence of consistent axiomatic theories about points, lines and planes containing the negation of some of Euclid's theorems. Indeed, Kant's claim that geometric theorems are synthetic a priori statements entails that there must be such theories. But after several alternative geometries became known, most researchers felt it was only fair that experience should be called upon to decide between them.

In his youth, Russell (1897) proposed a curious compromise on this issue. He retained the classical affine structure of space - supporting the notion of 'constant direction', and hence of straight line - as an apanage of reason, while leaving to experience the choice between three distinct metric structures - supporting alternative concepts of distance and size - which Klein had proved to be compatible with the said affine structure, namely, the Euclidean, in which the three internal angles of a plane triangle are always equal to two right angles or 180°, and two non-Euclidean ones, in which the said three angles always add up to, respectively, less than or more than 180°. Later Russell called this arrangement 'somewhat foolish' because it excludes the Riemannian geometries of

variable curvature employed in Einstein's gravitational theory (see Relativity theory, philosophical significance of §4; Spacetime §3).

In an unpublished fragment written in about 1893, C.S. Peirce emphasized the utter unlikelihood that, among the infinitely many values in any neighbourhood of $180°$, the sum of the three internal angles of a triangle should always settle exactly at $180°$ (Peirce 1931-58: §1.130). Similar reflections may have inspired, in part, Poincaré's innovative thesis on the conventionality of geometry. For him, any consistent geometry can be used in physics, provided it is comfortable. Euclidean geometry meets this requirement because the motions of ordinary solids agree fairly well with the translations and rotations of rigid figures in Euclidean three-dimensional space. Thus, Euclidean geometry, in which all triangles have internal angles adding up to precisely $180°$, constitutes a useful idealization of experience, but it would be ridiculous to claim truth for it (see Poincaré, H.J. §3).

In Newtonian physics, the geometry of space plays a dynamical role through the law of inertia: a force-free particle moves along a straight line, traversing equal distances in equal times. Thus, we may say that space fixes the norm of default dynamical behaviour. Why does matter comply with it? The very idea that uniform, inert space can thus be 'read' at each point by bodies as an infinitely rich but perfectly definite network of routes seems, at first blush, nonsensical. Mach (1883) proposed that inertial behaviour was caused by the presence of very distant masses. From this standpoint he criticized the water-bucket experiment devised by Newton to show the difference between relative motion and 'true' motion in absolute space. In this experiment, a bucket filled with water is made to rotate. Initially the rotation of the bucket is not communicated to the water; the latter therefore rotates relative to its container; its surface remains flat. But as soon as the water acquires the bucket's rotation, that is, as soon as it is again at rest relative to its container and moves with it in space, the surface takes a concave shape. Mach notes that 'moving in space' is here just Newton's idiom for 'moving relatively to the bulk of matter'. He wrily comments that the outcome of the experiment might be quite different if the bucket had lead walls that were one mile thick. Mach did not provide a viable alternative to Newtonian dynamics, but his view of inertia deeply influenced Einstein's quest for a relativistic theory of gravitation (see Relativity theory, philosophical significance of §6; Einstein, A. §3). Einstein, indeed, worked with two ideas that Mach never dreamed of using: he substituted spacetime for space-plus-time as the arena of nature, and Riemannian for Euclidean geometry. This was made possible by the mathematical developments sketched below.

4 General spaces

Bourbaki defined the mathematical concept of 'structure' for the sake of ordering and classifying the rich and intricate concept jungle generated by mathematicians since 1800 (Bourbaki 1970). As noted above, space - as understood by Newton in the wake of Descartes (and Euclid) - furnished a prototype for that concept. However, not every structure is called a 'space' by mathematicians. They reserve the word for structures which somehow resemble Newtonian space because they involve a notion of distance ('metric' spaces), of straightness ('affine' and 'projective' spaces), or of neighbourhood and continuity ('topological' spaces). The conception of such general spaces owes much to Grassmann's *Die lineale Ausdehnungslehre (The Theory of Linear Extension)* (1844), Riemann's lecture 'Über die Hypothesen, welche der Geometrie zugrunde liegen' (On the Hypotheses which lie at the Foundations of Geometry) (1854) and Klein's so-called Erlangen Programme (1872). We shall briefly comment on the latter two (for applications, see Spacetime).

New geometries flourished in the nineteenth century. Most famous outside mathematics is Lobachevskii's non-Euclidean geometry, which took over all the explicit and tacit assumptions of Euclidean geometry except Euclid's postulate, which was replaced by its denial. Much more influential, however, was the projective geometry of Poncelet (1822), which added a 'point at infinity' to each Euclidean straight. This point is the intersection of a family of parallels and can be reached by travelling along any of them in either direction. Because of this last feature, the neighbourhood systems - the topology - of projective and Euclidean space are drastically different. The power and elegance of projective geometry led nineteenth-century geometers to view ordinary geometry as the theory of a truncated projective space. Klein proposed to bring all past and future geometries under a single unifying standpoint. This was supplied by the notion of a transformation group acting on a space. Let S be the set of all points in Euclidean space. A transformation of S is a bijection of S onto itself, that is, a mapping f which assigns to every $p \in S$ an exclusive 'value' $f(p) \in S$ so that every $q \in S$ is thus assigned to some $p \in S$. It is clear that (1) the identity mapping i that assigns each point $p \in S$ to itself is a transformation of S; (2) if f and g

are transformations, the composite mapping *gf* which assigns to each $p \in S$ the value assigned by *g* to *f*(*p*) is also a transformation; (3) for every transformation *f* there is an inverse transformation *f′* such that $f'f = ff' = i$. Conditions (1)-(3) entail that the set of all transformations of *S* constitutes a group in the standard algebraic sense. Call it T_S. Let $f \in T_S$ and let *R* be an *n*-ary relation on *S*. We say that *f* preserves *R* and that *R* is invariant under *f* if whenever *R* holds between $p_1, \ldots, p_n \in S$ it also holds between $f(p_1), \ldots, f(p_n)$. If every transformation of a group *G* preserves *R* we say that *R* is an invariant of *G*. Though no significant geometrical relation is an invariant of T_S, distinct sets of geometrical relations turn out to be the invariants of different subgroups of T_S (a subgroup is a subset of a group which is a group in its own right). Thus, a short reflection will show that all transformations which map continuous paths onto continuous paths form a group; that collinearity of points and the mutual intersection between straights are invariants of a subgroup of that group; and that congruence of segments and angles are invariants of a subgroup of the latter. The invariants of a group determine a structure which is studied by a geometry. Different groups acting on the same arbitrary set of points will organize them into very different structures. In particular, Klein showed that Euclidean and Lobachevskian geometry study the invariants of different subgroups of the projective group. Poincaré concluded that the choice between them is conventional, for ‘the existence of one group is not incompatible with that of another’ (1887: 90).

While Klein sought to reorganize geometry as a branch of pure mathematics, Riemann’s chief aim was to make it more responsive to the needs of physical research, to replace the Euclidean-Newtonian straitjacket with something much more pliable. Physicists write nature’s laws as differential equations involving time and three spatial coordinates. It may seem that they thereby assume that physical space is homeomorphic (topologically equivalent) to $\Re^3$, the structured set of ordered triples of real numbers. However, the stated assumption is unnecessarily strong. Physicists need only coordinatize a finite region surrounding the objects under study. They must therefore assume that each point of space has a *neighbourhood* homeomorphic to $\Re^3$. But such neighbourhoods may be pieced together in a global structure that is very different from $\Re^3$, just as the patches mapped by cartographers onto the *flat* pages of an atlas jointly form the surface of the earth, which is homeomorphic to a *sphere*. Physical space should therefore be regarded as a threefold instance of what Riemann called ‘*n*-fold extended quantities’. A good explication of this notion is supplied by our current concept of a ‘real *n*-dimensional differentiable manifold’, that is, a topological space *M*, every point of which has a neighbourhood homeomorphic with $\Re^n$ which is mapped bijectively and bicontinuously onto an open region of $\Re^n$, so that (1) every point of *M* lies in the domain of one of these mappings or ‘coordinate systems’, and (2) if two such coordinate systems, *g* and *h*, have overlapping domains, the ‘coordinate transformations’ gh^{-1} and hg^{-1} are differentiable. (Here h^{-1} designates the inverse of *h*, that is, the mapping that assigns to each real number *n*-tuple in the range of *h* the point $p \in M$ such that the said *n*-tuple is *h*(*p*); differentiability is required in order to write laws as differential equations.) A particle’s trajectory in *M* can be described by a mapping *f* of a time interval into *M*. The mapping *f* must be differentiable in the following sense: if its range lies within the domain of a coordinate system *g*, the composite mapping *gf* (which yields the *g*-coordinates of successive positions of the particle) is differentiable. The quantitative description of motion requires that such trajectories have definite lengths. Riemann thought that the length of a real trajectory depends on the interplay of forces acting on and about it. The success of the standard definition of lengths expressed in Pythagoras’ theorem betokens its approximate validity on the human scale. But it could well break down on a larger or smaller scale. Physicists should therefore wield a more flexible definition, enabling them to give - whenever experience makes it advisable - a more accurate description of motions. This is secured by Riemann’s concept of a ‘metric’ on the manifold *M*. Roughly, a metric is a rule for measuring, at each $p \in M$, the velocity of every conceivable trajectory through *p* in such a way that the velocity varies smoothly, if at all, along the trajectory. Along a trajectory of velocity 1, the integral of the velocities measures the length of the trajectory. Since a metric based on Pythagoras’ theorem had until then proved suitable, Riemann thought that for the time being one ought only to allow metrics that agree to first order with the Pythagorean metric on a neighbourhood of each point. Such metrics are called ‘Riemannian’.

See also: Cosmology and cosmogony, Indian theories of; Geometry, philosophical issues in

ROBERTO TORRETTI

References and further reading

Aristotle (*c.* mid-4th century BC) *Physica*, ed. and with notes by W.D. Ross, Oxford: Clarendon Press, 1956. (Aristotle’s lectures on the principles of natural philosophy; regarded as the fundamental treatise on the subject

at the time Galileo and Descartes went to school.)

Berkeley, G. (1721) *De motu* (On Motion), London: Tonson.(A spirited defence of the relativity of motion, against Newton and the Newtonians.)

Boi, L. (1995) *Le problème mathématique de l'espace: Une quête de l'intelligible (The Mathematical Problem of Space: A Quest For Intelligibility)*, Berlin: Springer.(A careful and intelligent exposition of the subject touched on in §4, rich in mathematical detail and philosophical insights.)

Bourbaki, N. (1970) *Théorie des ensembles* (Set Theory), Paris: Hermann.(Lays the groundwork for Bourbaki's systematic ordering of mid-twentieth-century mathematical lore around the concept of structure.)

Bruno, G. (1591) *De immenso et innumerabilibus (On the Immense and the Uncountables)*, in *Opera Latina Conscripta*, Naples: Morano, 1879, vol. 1, 1.(Expanded Latin version of Bruno's great Italian dialogues *De l'infinito, universo e mondi (On the Infinite, the Universe and the Worlds)* (1584), which argued for the existence of an infinite number of worlds; cited here for its description of space.)

Buroker, J.V. (1981) *Space and Incongruence: The Origin of Kant's Idealism*, Dordrecht: Reidel. (Stresses the significance of the philosophical problem of space in the development of Kant's critical philosophy; an original and very illuminating study.)

Cornford, F.M. (1936) 'The Invention of Space', in *Essays in Honour of Gilbert Murray*, London: Allen & Unwin, 215-35; repr. in M. Capek (ed.) *The Concepts of Space and Time*, Dordrecht: Reidel, 1976, 3-16. (According to Cornford, boundless all-encompassing space was invented in Greece during the fifth century BC.)

Earman, J. (1989) *World-Enough and Space-Time: Absolute versus Relational Theories of Space and Time*, Cambridge, MA: MIT Press.(Twentieth-century perspectives on the classical debate between Leibniz and Clarke.)

Euler, L. (1748) 'Réflexions sur l'espace et le tems' (Reflections on Space and Time), *Mémoires de l'Académie des Sciences de Berlin* 4; repr. in *Leonhardi Euleri Opera omnia*, series 3, vol. 2, 376-83. (Space and time are indispensable presuppositions of mechanics, required by the law of inertia.)

Grant, E. (1981) *Much Ado About Nothing: Theories of Space and the Vacuum from the Middle Ages to the Scientific Revolution*, Cambridge: Cambridge University Press.(Very learned and instructive; no mathematics at all.)

Grassmann, H. (1844) *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik (The Linear Theory of Extension: A New Branch of Mathematics)*, Leipzig: Otto Wigand; 2nd edn, 1862.(The first, very difficult version of the author's theory of extension; it develops *n*-dimensional vector algebra as part of a much more comprehensive, ambitious and original system of geometric algebra.)

Hall, A.R. (1980) *Philosophers at War: The Quarrel Between Newton and Leibniz*, Cambridge: Cambridge University Press.(Good historical narrative of the various aspects of this regrettable dispute.)

Kant, I. (1746) *Gedanken von der wahren Schätzung der lebendingen Kräfte (Thoughts on the True Measure of Live Forces)*, Königsberg: Dorn.(Kant's MA thesis, on the question, disputed between Cartesians and Leibnizians, whether the 'live force' of bodies in motion equals mass times speed or mass times squared speed.)

Kant, I. (1768) 'Von dem ersten Grunde des Unterschiedes der Gegenden im Raume' (On the Foundation of the Difference between Regions in Space), *Wochentliche Königsbergsche Frag- und Anzeigungs-Nachrichten* 6-8. (In this short paper published in an obscure local journal, Kant made the turnabout in the philosophy of space which eventually led to his critique of reason; in it he argues that the difference between incongruent bodies which are exact mirror images of each other can only be grasped in terms of the relation of each body to absolute space.)

Kant, I. (1770) *De mundi sensibilis atque intelligibilis forma et principiis (On the Form and Principles of the Sensible and the Intelligible World)*, Regiomonti: Kanter.(Kant's inaugural dissertation on taking the chair of metaphysics in Königsberg; argues for the thesis that real things, as such, are neither spatial nor temporal.)

Kant, I. (1781) *Critik der reinen Vernunft (Critique of Pure Reason)*, Riga: Hartknoch.(This enormously influential book brought about Kant's 'Copernican revolution' in philosophy; this first edition is often denoted by the letter A.)

Kant, I. (1787) *Critik der reinen Vernunft, Zweyte hin und wieder verbesserte Auflage (Critique of Pure Reason, Second Improved Edition)*, Riga: Hartknoch.(Edition B of Kant's book; emphasizes more strongly than A the productive role of the human intellect in the constitution of human knowledge.)

Klein, F. (1872) *Vergleichende Betrachtungen über neuere geometrische Forschungen (Comparative Reflections about Modern Geometrical Inquiries)*, Erlangen: A. Düchert.(A programme for unifying the newly developed forms of geometry from the point of view of transformation groups and group invariants.)

Koyré, A. (1957) *From the Closed World to the Infinite Universe*, Baltimore, MD: Johns Hopkins University

Press.(A lively narrative of the emergence of space.)
Leibniz, G.W. (1716) ‘Mr. Leibnitz’s Fifth Paper, being an Answer to Dr Clarke’s Fourth Reply’, in *The Leibniz-Clarke Correspondence*, ed. with intro. and notes by H.G. Alexander, Manchester: Manchester University Press, 1956.(Part of a polemical correspondence, on the nature of space and time and related subjects, between Leibniz and Newton’s spokesman Samuel Clarke.)
Mach, E. (1883) *Die Mechanik in ihrer Entwicklung historisch-kritisch dargestellt (The Development of Mechanics: A Historico-Critical Exposition)*, Leipzig: Brockhaus; trans. T. McCormack, *The Science of Mechanics*,1893; La Salle, IL: Open Court, 1974.(A critical history of mechanics from Galileo to Hamilton and Jacobi, somewhat weak in its appreciation of the power of mathematical ideas; Mach’s criticism of Newton’s allegedly needless ‘metaphysics’, especially in the matter of space and time, had some influence on Einstein.)
Newton, I. (1962) ‘De gravitatione et aequipondio fluidorum’ (On the Gravity and Equilibrium of Liquids), ed. and trans. A.R. Hall and M.B. Hall, in *Unpublished Scientific Papers of Isaac Newton*, Cambridge: Cambridge University Press, 89-121; trans. 121-56.(A long paper, posthumously edited, containing much criticism of Descartes and elucidations of the fundamental concepts of mechanics.)
Oresme, N. (*c*.1375) *Le Livre du ciel et du monde (The Book of Heaven and the World)*, ed. A.D. Menut and A.J. Denomy, Madison, WI: University of Wisconsin Press, 1968.(Late medieval treatise of cosmology. Contains a facing English translation by A.D. Menut.)
Peirce, C.S. (1931-58) *Collected Papers*, ed. C. Hartshorne, P. Weiss and A.W. Burks, Cambridge, MA: Belknap Press, 8 vols.(A selection of papers and excerpts from manuscripts, organized thematically; though not altogether reliable, this edition is still probably more useful for the new reader than *The Writings of Charles S. Peirce: A Chronological Edition*, ed. M. Fisch *et al*., Bloomington, IN: Indiana University Press, 1982-, which will eventually supersede it.)
Poincaré, H. (1887) ‘Sur les hipothèses fondamentales de la géométrie’ (On the Fundamental Hypotheses of Geometry), *Bulletin de la Société mathématique de France* 15: 203-16; repr. in *Oeuvres*, Paris: Gauthier-Villars, 1916-56, vol. 11, 79-91.(Poincaré’s attempt to state without redundancy all the necessary assumptions of plane geometry; provides important cues for understanding Poincaré’s conventionalist philosophy of geometry.)
Poncelet, J.V. (1822) *Traité des propriétés projectives des figures (Treatise on the Projective Properties of Figures)*, Paris: Bachelier.(First systematic treatise of projective geometry.)
Riemann, B. (1854) ‘Über die Hypothesen, welche der Geometrie zugrunde liegen’ (On the Hypotheses which lie at the Foundations of Geometry), *Abhandlungen der Kgl. Gesellschaft der Wissenschaften zu Göttingen* 13: 133-52; trans. with mathematical commentary in M. Spivak, *A Comprehensive Introduction to Differential Geometry*, Berkeley, CA: Publish or Perish, 1979, vol. 2, 135-.(A lecture, on the occasion of Riemann’s habilitation as *Privatdozent*; the starting point of modern physical geometry.)
Rosenfeld, B.A. (1988) *A History of Non-Euclidean Geometry: Evolution of the Concept of a Geometric Space*, trans. A. Shenitzer, New York: Springer.(Excellent, moderately technical exposition.)
Russell, B.A.W. (1897) *An Essay on the Foundations of Geometry*, Cambridge: Cambridge University Press. (Russell’s youthful attempt at solving all the main philosophical problems of geometry; ingenious and befuddled.)
Schutz, B.F. (1980) *Geometrical Methods of Mathematical Physics*, Cambridge: Cambridge University Press.(A very readable, suitably technical but not overwhelmingly rigorous introduction to Riemannian methods and ideas.)

Spacetime

Spacetime is the four-dimensional manifold proposed by current physics as the arena for Nature's show. Although Newtonian physics can very well be reformulated in a spacetime setting, the idea of spacetime was not developed until the twentieth century, in connection with Einstein's theories of special and general relativity. Due to the success of special relativity in microphysics and of general relativity in astronomy and cosmology, every advanced physical theory is now a spacetime theory. Spacetime *is undoubtedly an artificial concept, which our hominid ancestors did not possess, but the same is true of Newtonian* space *and* time*.*

1 A world of events

Newtonian physics conceives the course of nature as the motion and change in time of bodies located in space (see Mechanics, classical §1; Space §2). Einstein's special theory of relativity implies that the shape of bodies and the order of succession of their changes varies with - 'is relative to' - the reference frame adopted for describing them (see Relativity theory, philosophical significance of §2). Hermann Minkowski (1909, 1915) showed that a unique - 'absolute' - description of natural phenomena consonant with Einstein's theory can be achieved by analysing them into instantaneous events-at-a-point. Such events typically constitute the history of ordinary lumps of matter but they may also occur - in the guise, say, of oscillations of the electromagnetic field - in a perfect vacuum. *Spacetime* is our name for the four-dimensional continuum of such events (Minkowski called it *the world*). It can be regarded - in infinitely many alternative ways - as compounded of one-dimensional time and three-dimensional space. Indeed, each one of the inertial reference frames privileged by classical physics and special relativity defines a distinct way of splitting spacetime into time and space. No wonder, then, that the spatial groupings ('bodies'), and the time series into which events coalesce relatively to each such reference frame differ, from one frame to another. At first Einstein viewed Minkowski's approach somewhat disparagingly, as a 'formalism' which made calculations easier but contributed no genuine insight. However, Einstein's subsequent development of a theory of gravity ('general relativity') is inconceivable without Minkowski's achievement (see Einstein, A. §3).

2 Minkowski spacetime

Should we decide to regard all natural phenomena as embedded in some sort of continuum, a four-dimensional spacetime is a much more obvious choice than a three-dimensional space lasting through time. Imagine a butterfly flying: one may eventually succeed - after much work - in *analysing* this phenomenon into a succesion of infinitely many instantaneous postures, but it is originally *given* as a single coherent spacetime flow. Nevertheless, Minkowski's conception of spacetime does not stem from this intuitive idea, but rather from reflecting about the numerical representation of phenomena in the new kinematics proposed by Einstein (1905). Let P_1 and P_2 be two points marked on a rigid rod *R* at rest in an inertially moving laboratory *L*. Let E_1 and E_2 be events at P_1 and P_2, respectively. Let *L* be furnished with a time coordinate *t*, defined by Einstein's radar method, and a Cartesian system of space coordinates *x*, *y*, *z*. Let (x_1, y_1, z_1) and (x_2, y_2, z_2) be the coordinates assigned in this system to P_1 and P_2, and let t_1 and t_2 be the times of occurrence of E_1 and E_2. Then, the lapse of time between E_1 and E_2 is given by $\tau = | t_1 - t_2 |$, and the distance λ between P_1 and P_2 is given at all times by the positive square root of $(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2$. Consider now a laboratory L' furnished with a time coordinate t' and a Cartesian system x', y', z', based on the same measurement units as the unprimed coordinates of *L*. For simplicity's sake, we assume that

(i) at some instant, the origin of the primed and the unprimed Cartesian systems coincide;
(ii) at that instant the axes of the unprimed system are aligned with the homonymous axes of the primed system;
(iii) this coincidence occurs at time $t = t' = 0$.

Suppose that *L* moves relatively to L' in the direction of the x'-axis, with constant speed v. Let (x'_1, y'_1, z'_1) and (x'_2, y'_2, z'_2) be the coordinates assigned in the primed Cartesian system to two simultaneous positions of P_1 and P_2 in L'. The distance λ' between these two positions is equal to the positive square root of $(x'_1 - x'_2)^2 + (y'_1 - y'_2)^2 + (z'_1 - z'_2)^2$. However, unless $v = 0$, $\lambda' < \lambda$ ('length contraction'), for $\lambda' = \lambda\sqrt{1-(v/c)^2}$ (where c denotes the constant speed of light *in vacuo*, expressed in the units employed in the definition of our coordinate systems). Relative to L', the time lapse τ' between E_1 and E_2 is equal to

$| t'_1 - t'_2 |$. Unless $v = 0$, $\tau' > \tau$ ('time dilation'), for $\tau' = \tau / \sqrt{(1 - v/c)^2}$. Thus the two fundamental kinematic quantities of length and duration turn out to be relative to the inertial frame of reference chosen. On the other hand, as Minkowski observed, the following equation holds always, even if we lift the simplifying restrictions (i)-(iii) and we allow *L* to move in L' with constant speed in *any* fixed direction:

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 - c^2(t_1 - t_2)^2 = (x'_1 - x'_2)^2 + (y'_1 - y'_2)^2 + (z'_1 - z'_2)^2 - c^2(t'_1 - t'_2)^2 = \sigma_{12}$$

The invariant quantity σ_{12} is the *spacetime interval* between events E_1 and E_2. If $\sigma_{12} < 0$, the last term must be larger than the sum of the other three, in which case σ_{12} is *timelike*. Likewise, σ is *spacelike* if $\sigma_{12} > 0$. Finally, σ_{12} is said to be *lightlike* if $\sigma_{12} = 0$, for only in this case can one of the events be the emission and the other the reception of a light signal that travels *in vacuo*. The kinematics of Einstein (1905) can therefore be conceived as the geometry of a four-dimensional continuum, a locus of events - Minkowski's 'world' or spacetime - characterized by the display equation. Since the change of the primed into the unprimed coordinates can be effected by any transformation of the Poincaré-Lorentz group (see Relativity theory, philosophical significance of §1), the said geometry can be characterized, from Klein's standpoint, (see Space §4), by the action of the Poincaré-Lorentz group on a manifold homeomorphic to $\mathbb{R}^4$. It can also be characterized, from Riemann's standpoint (see Space §4), by conferring on such a manifold a Riemannian metric η that can be - inelegantly but accurately - described as follows:

(a) if E_1 and E_2 are simultaneous in a given inertial lab *L*, then σ_{12} is the squared length assigned by the metric η to a straight line joining the locations of E_1 and E_2 in the relative space of *L*;
(b) if E_1 and E_2 are events in the history of an inertial particle *p*, then σ_{12} is the squared length assigned by the metric η to the spacetime trajectory - the 'world-line' - of *p* between E_1 and E_2.

Because spacetime intervals can be negative or zero, the said metric η is not a standard Riemannian metric and is sometimes described as *semi*-Riemannian. In a manifold endowed with a Riemannian metric, one can discriminate between curves with or without a steady direction. The former are called *geodesics*. Thus, on a plane the geodesics are straight lines, on a sphere they are great circles. In Minkowski spacetime, the world-lines of inertial particles and light-signals *in vacuo* are geodesics. At each point of the manifold the metric determines a quantity called the *curvature scalar*, which is constant $= 0$ on a plane, $= 1/r$ on a sphere of radius *r*. In Minkowski spacetime the curvature scalar is 0 everywhere, so we say that it is *flat*. (It is perhaps worth noting that the flatness or deviation from flatness of a Riemannian manifold $\mathcal{M}$ is measured by the so-called *curvature tensor*, whose value at each point *p* of $\mathcal{M}$ is expressed - relative to a coordinate system defined on a neighbourhood of *p* - by an *array* of real numbers. The curvature scalar generally conveys less information than the curvature tensor, on which it depends. However, if every term of the array equals 0 everywhere, the curvature scalar is also constant and equal to 0; and if $\mathcal{M}$ is two-dimensional the said array has only one term, equal to the curvature scalar.)

3 Curved relativistic spacetimes

Prompted by the radical incompatibility between special relativity and Newtonian gravity, Einstein worked from 1907 on a theory of gravitation. He was guided by the assumption - known as the *equivalence principle* - that an observer in a closed laboratory has no way of telling whether the lab moves inertially, or falls freely in a uniform gravitational field. If the world were filled by a uniform gravitational field it could therefore be regarded as a Minkowski spacetime in which freely falling particles are constrained by gravitational forces to stick to geodesics. But real gravitational fields are not uniform, and the equivalence of free fall and inertial motion holds only approximately. This suggested to Einstein the following bold analogy: just as the Riemannian metric of a sphere is approximated on a small neighbourhood of each point by the Pythagorean metric of the Euclidian plane (see Space §4), so the (semi-)Riemannian metric of real spacetime is approximated on a small neighbourhood of each event by the flat Minkowski metric η. But the actual metric is not flat: it deviates more or less from flatness according to the presence or absence of matter. Freely falling particles follow geodesic world-lines as determined by this metric. Thus, gravitational phenomena depend, *through the geometry*, on the distribution of matter. The theory finally attained by Einstein (1915, 1916), known as 'general relativity', roughly fits this scheme (although its equations admit solutions in which the density of matter is equal to 0 everywhere and the law of geodesic free fall generally holds only for uncharged spinless particles). Celebrated until *c.*1950 chiefly for its beauty, it has since then enjoyed a spectacular run of experimental successes which place it far ahead of other available theories of

gravity (on the other hand, it is incompatible with the currently accepted quantum theories of non-gravitational forces). The theory turns on a system of differential equations, the *Einstein field equations* (EFE), in which the spacetime metric is the unknown and a particular distribution of matter is assumed. Under different assumptions, the equations yield diverse solutions, which is why one usually speaks, in the plural, of relativistic spacetimes.

4 A wealth of worlds

Philosophers have been greatly exercised by the many surprising alternative ways of conceiving the overall structure of spacetime offered by general relativity. Assuming that the density of matter ρ is always and everywhere roughly the same, Einstein (1917) obtained the first cosmological solution of the EFE: a spacetime that can be decomposed into a succession of geometrically identical slices of simultaneous events, homeomorphic to the three-dimensional sphere $\mathbf{S}^3$. This was hailed as a solution of Kant's antinomy concerning the world's size, for Einstein's spherical space is finite and yet, being boundless, it poses no questions about what lies beyond it. But Einstein's world clashes with the redshift noticeable in radiation from extragalactic sources, which would indicate that ρ is decreasing. It is, moreover, an unstable limiting case of a family of solutions found by Friedmann (1922, 1924), under the less stringent assumption that ρ is spatially uniform but may vary with time. A generic Friedmann world will either

(i) contract indefinitely,
(ii) expand indefinitely,
(iii) expand for a finite time and then contract.

Cases (ii) and (iii) could agree, to a first approximation, with the world as we see it. They share this feature: ρ increases beyond all bounds within a finite time as we go back into the past along any world-line of matter. Since the metric is undefined unless ρ is finite, spacetime cannot contain a point at which $\rho = \infty$. Hence at present only a finite time *T* has elapsed along each world-line of matter. Friedmann called *T* ' the time since the creation of the world' (1922: 384), and some writers believe that his work lends scientific support to Genesis 1: 1. Be that as it may, one should bear in mind that *every* given moment in a Friedmann world is preceded by *some* time, no matter how short, Thus, his cosmology solves Kant's antinomy concerning the world's duration, for Friedmann time is finite and yet, having no beginning, it raises no questions about what was the case before it.

Less realistic but no less exciting is a family of solutions of the EFE discovered by Gödel (1949a), characterized by the presence of closed (potential) world-lines. If one of these happens to be the world-line of a conscious particle, it will re-experience its past infinitely many times as it lives on into the future (see Time travel). Opinions on this matter are divided between those who believe with Gödel (1949b) that the mere conceivability of such relativistic worlds should affect our understanding of time, and those who think that those features of an abstract mathematical structure which in a realistic model constitute a plausible representation of time may well not have any significant analogy with time in other models (see Time §3).

Earman (1995) surveys the whole range of such relativistic 'monsters'. Particularly interesting for the philosopher are the 'observationally indistinguishable spacetimes' discovered by Glymour (1972, 1977). Let us say, with Malament (1977), that two solutions of the EFE, $\mathcal{W}_1$ and $\mathcal{W}_2$, are *weakly observationally indistinguishable* if for every event x_1 in $\mathcal{W}_1$ there is an event x_2 in $\mathcal{W}_2$ such that the spacetime region $\mathcal{I}^-(x_1) \subset \mathcal{W}_1$ from which slower-than-light signals can reach x_1 is isometric with the region $\mathcal{I}^-(x_2) \subset \mathcal{W}_2$ from which slower-than-light signals can reach x_2. This entails that observers in $\mathcal{W}_1$ cannot learn from experience whether they live in $\mathcal{W}_1$ or in $\mathcal{W}_2$, although it may occur that $\mathcal{W}_1$ and $\mathcal{W}_2$ do not share some very important global properties (for example, the existence of closed world-lines).

ROBERTO TORRETTI

References and further reading

Ciufolini, I. and Wheeler, J.A. (1995) *Gravitation and Inertia*, Princeton, NJ: Princeton University Press.(A recent, quick-paced exposition of general relativity, with all the brilliance and keen sense of essentials which are normal in books coauthored by Wheeler.)

Earman, J. (1989) *World-Enough and Space-Time; Absolute versus Relational Theories of Space and Time*, Cambridge, MA: MIT Press.(Discusses absolutism and relationalism in both the Newtonian and the relativistic

setting.)
Earman, J. (1995) *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*, New York and Oxford: Oxford University Press.(A critical survey of spacetime singularities - naked and clothed - supertasks, horizons, time travel, and so on, by a distinguished philosopher well-versed in physics and mathematics.)
Einstein, A. (1905) 'Zur Elektrodynamik bewegter Körper' (On the Electrodynamics of Moving Bodies), *Annalen der Physik* 17: 891-921.(Einstein's first - and principal - paper on special relativity.)
Einstein, A. (1915) 'Die Feldgleichungen der Gravitation' (The Field Equations of Gravitation), *K. Preußische Akademie der Wissenschaften. Sitzungsberichte*, 844-7. (First publication of the Einstein field equations.)
Einstein, A. (1916) 'Die Grundlage der allgemeinen Relativitätstheorie' (The Foundation of the General Theory of Relativity), *Annalen der Physik* 49: 769-822.(Einstein's detailed presentation of his theory of gravity.)
Einstein, A. (1917) 'Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie' (Cosmological Considerations on the General Theory of Relativity), *K. Preußische Akademie der Wissenschaften; math.-phys. Cl. Sitzungsberichte*, 142-52.(In this trailblazing paper Albert Einstein founded modern cosmology.)
Ellis, G.F.R. and Williams, R.M. (1988) *Flat and Curved Space-Times*, Cambridge: Cambridge University Press.(Outstanding introduction to the mathematics of relativistic spacetime; requires high-school geometry and algebra but no higher mathematics; an excellent reading list directs the student to more advanced books.)
Friedmann, A. (1922) 'Über die Krümmung des Raumes' (On the Curvature of Space), *Zeitschrift für Physik* 10: 377-86.(Alexander Friedmann's first paper on relativistic cosmology.)
Friedmann, A. (1924) 'Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes' (On the Possibility of a World with Constant Negative Curvature of Space), *Zeitschrift für Physik* 21: 326-32.(Alexander Friedmann's second paper on relativistic cosmology.)
Friedman, M. (1983) *Foundations of Space-Time Theories: Relativistic Physics and Philosophy of Science*, Princeton, NJ: Princeton University Press.(Currently, the standard philosophical treatment of spacetime theories.)
Glymour, C. (1972) 'Topology, Cosmology and Convention', *Synthèse* 24: 195-218. (The possibility of indistinguishable spacetimes was first announced here.)
Glymour, C. (1977) 'Indistinguishable Space-times and the Fundamental Group', *Minnesota Studies in the Philosophy of Science* 8: 50-60.(Glymour's principal publication on the subject.)
Gödel, K. (1949a) 'An Example of a New Type of Cosmological Solution of Einstein's Field Equations of Gravitation', *Reviews of Modern Physics* 21: 447-50.(Gödel's cosmological solution of the Einstein field equations, which allows closed timelike curves.)
Gödel, K. (1949b) 'A Remark about the Relationship Between Relativity Theory and Idealistic Philosophy', in P.A. Schilpp (ed.) *Albert Einstein, Philosopher-Scientist*, Evanston, IL: Northwestern University, 555-62.(Gödel's philosophical comments on his cosmological solution of the Einstein field equations.)
Grünbaum, A. (1973) *Philosophical Problems of Space and Time*, Dordrecht: Reidel, 2nd edn enlarged.(Detailed coverage of philosophical debate until the date of publication.)
Grünbaum, A. (1991) 'Creation as Pseudo-explanation in Current Physical Cosmology', *Erkenntnis* 35: 233-54.(Discusses creationism in connexion with relativistic cosmology.)
Malament, D. (1977) 'Observationally Indistinguishable Space-times', *Minnesota Studies in the Philosophy of Science* 8: 61-80.(A very clear presentation of the subject, which in some ways improves on Glymour's original work, listed above.)
Minkowski, H. (1909) 'Raum und Zeit' (Space and Time), *Physikalische Zeitschrift* 10: 104-11.(Minkowski's classical lecture of 1908, in which he presented his conception of spacetime to a broad scientific audience.)
Minkowski, H. (1915) 'Das Relativitätsprinzip' (The Principle of Relativity), *Jahresbericht der Deutschen Mathematiker-Vereinigung* 24: 372-82.(Posthumously published text of the first public lecture on Minkowski's formulation of special relativity theory, delivered in Göttingen in 1907.)
Stein, H. (1967) 'Newtonian Spacetime', *Texas Quarterly* 10: 174-200.(This remarkable essay, written for non-mathematicians, shows how much more natural it is to formulate Newtonian dynamics in a spacetime setting, rather than in the traditional space-and-time setting.)
Torretti, R. (1983) *Relativity and Geometry*, Oxford: Pergamon Press; repr. New York: Dover, 1996.(An essay on spacetime geometry by the author of this article.)

Spain, philosophy in

Historians have argued about precisely when to date the commencement of Spanish history proper, rendering dubious any reference to Spain as such in the period prior to the official constitution of nationality. If this is the case, one can not really speak of philosophy in Spain before 1474, although it remains a fact that philosophy had been practised on the Iberian Peninsula from the earliest times. During the period of the Roman Empire, distinguished philosophical figures included Lucius Annqeus Seneca; under Visigothic rule, Saint Isidore of Seville came to the fore; and the Islamic Empire featured some of the most eminent philosophers of the Arabic and Judaic traditions, such as Ibn Hazm, Averroes, Ibn Gabirol, Yehuda Ha-Levi and Maimonides. There is no doubt that the centre of philosophical activity within the peninsula during the Middle Ages was the so-called School of Translators of Toledo, where numerous thinkers from many countries gathered. Together with Spanish scholars such as Domingo Gundisalvo and Juan Hispano, they collaborated in making Greek philosophy available to the countries of Europe; instrumental in this process were Gerard of Cremona, Daniel of Morlay, Alexander Neckham and Michael Scot.

After Spanish nationality was constituted under the Catholic Monarchs (1474-1516) on the basis of a single, unified faith, philosophy was destined to become closely linked with religion. During the sixteenth century, this gave rise to a burgeoning of philosophy of the very highest order, which followed two separate paths: that of the Erasmian-style Renaissance, featuring Luis Vives, which developed in line with the vanguard of the European Renaissance; and that of Spanish Scholasticism, which was fuelled by the thrust of the Counter-Reformation on the one hand, and by the discovery of America on the other. After the reigns of Charles I and Philip II (the chief protagonists in the creation of the empire 'in which the sun never set'), the seventeenth, eighteenth and nineteenth centuries witnessed a relentless decline which, towards the beginning of the twentieth, seemed to come to an end. The Generation of 1898, with its revolutionary secular theories, provided the catalyst for a philosophical recovery whose greatest protagonists were Miguel de Unamuno, José Ortega y Gasset and Xavier Zubir. These thinkers were succeeded by the philosophers who went into exile after the Civil War of 1936-9: José Ferrater Mora, José Gaos, María Zambrano, Joaquín Xirau, and Juan David García Bacca.

1 The Middle Ages

One cannot speak of a strictly Spanish philosophy until after the founding of the Spanish State in 1474. Nevertheless, one cannot dismiss the importance of Hispano-Arabic and Hispano-Judaic philosophy prior to that date, as well as certain figures from within the Christian community who attained great prominence during Medieval times. An outstanding figure of this period was Ramón Llull, Mallorcan by birth, who developed a complicated combinatory logic - called the 'Ars Luliana' - which offered an ingenious foretaste of our current axiomatized logics. He was the author of an immense and highly diverse series of works, which sparked off a long tradition of Lullism that has continued to the present. His thought contains the seeds of the *mathesis universalis*, which served as the basis for European Rationalism as developed later by Descartes and Leibniz. The study of Llull has continued to the present day in the journal *Estudios Lulianos*, and at the Ramundus Llullus Institute in Fribourg-am-Brisgovia. Llull was the first European philosopher to write in a Romance language, since part of his work is in Catalan, although most is in Latin and some even in Arabic (see Llull, R.).

2 The Renaissance

The founding of the Spanish State brought about by the marriage of Ferdinand V of Aragon and Isabel I of Castile - the so-called Catholic Monarchs - took place during a time of crisis for Christianity, the effects of which were felt most acutely in the Catholic countries. In Spain, this was expressed initially in the widespread diffusion of the doctrines of Erasmus of Rotterdam which took on a novel and highly radical form, propounded by men as renowned as Juan and Alfonso de Valdés and Andrés Laguna. From a philosophical point of view, the single most important figure was Luis Vives who lived the greater part of his life in Bruges, although he also spent some time in Oxford after being appointed private tutor to princess Mary Tudor and reader to queen Catherine of Aragon during her marriage to Henry VIII. Vives was one of the most eminent representatives of the Renaissance spirit: his critique of Scholasticism, his rehabilitation of the classics, the attention he devoted to anthropological problems and his practical vision of philosophy were unparalleled. Perhaps his most important book is *De anima et vita (On*

the soul and on life) (1538), a precursor to differential psychology and vocational guidance, since he recognizes the need to observe and distinguish the differences in aptitude and disposition between individuals with a view to choosing their office or profession. Vives also developed an original theory of the passions very much along the modern lines which Descartes was later to pursue. All in all, Vives was a cosmopolitan and universal spirit who showed himself to have advanced beyond European thinking in his *De concordia et discordia in humano genere (On harmony and discord in humankind)* (1529), as well as in his social attitudes regarding aid and protection for the needy classes, as he revealed in *De subventione pauperum (On poor relief)* (1526). Generally speaking, Erasmianism was also a form of Europeanism which endorsed the notion of the Carlist Empire as a ‘universitas christiana’, not an ‘imperium mundi’ as it was conceived of by the pagan emperors. One important representative of this Europeanism was Andrés Laguna, doctor to king Charles V and author of *Discurso de Europa (Discourse on Europe)* (1542), in which he argued for a coalition of the Christian monarchs to form a unified force designed to counter the threat of the Ottoman Empire.

3 The Golden Age

The Spanish Renaissance, which began as a reaction to the crisis faced by European Christianity, gradually changed in character during the seventeenth century when Protestantism consolidated its advance throughout the Germanic and Anglo-Saxon countries, and also as the possibilities opened up by the discovery and colonization of America were exploited. As a result, two key philosophical fronts emerged. The first, whose common denominator was the Roman Catholic Counter-Reformation, was to give rise to the movement known as Spanish scholasticism. The greatest practitioners of this scholasticism were two religious orders, the Dominicans and the Jesuits. The main centre of activity for the Dominicans was the University of Salamanca, aided by their nearby monastery of St. Stephen, and among the main representatives were Melchor Cano, Domingo de Soto and Francisco de Vitoria. Cano’s most important work is *De Locis teologicis (On theological standpoints)* (1563), while Soto’s famous treatise is entitled *De iustitia et de iure (On justice and law)* (1553). Both helped to prepare the ground which formed the bedrock for the Council of Trent (1545-63), where the fundamentals of Catholic doctrine for the next four centuries were formulated.

As for Francisco de Vitoria, he was chiefly occupied with problems stemming from the conquest and colonization of America, concerns which he voiced in his books *De indis prior (The question of the Indies)* (1539), *De iure belli (The law of war)* (1539) and *De potestate civili (On civil power)* (1528). In these works he developed his theory of the ‘just war’, as well as that of Spain’s legal rights to the conquest and colonization of America, implementing a doctrine which would later form the basis of future international law. Here Vitoria introduced the concept of a ‘world community’ to which all people belong as a consequence of their fundamentally social nature which exists above and beyond their division into nation states. This world community was governed by natural law (*ius naturale*) and, in those matters beyond its sphere, by the law of peoples (*ius gentium*) which functioned in accordance with various principles concerning international cohabitation. According to Vitoria:

> There can be no doubt that the whole world, which is, in a certain sense, a republic, has the right to dictate just and profitable laws to all of its members, similar to those laid down in the law of peoples…. From this it follows that whoever breaks the law of peoples, whether in peacetime or in war, is committing a mortal sin, and that in matters of import, such as that concerning the immunity of ambassadors, it is forbidden for any republic to refuse to comply with the law of peoples…. Just as the majority in a republic has the right to appoint a king, so too the majority of Christians, whether or not the minority be in agreement, has the right to name a sovereign leader, who must be obeyed by all.

(Vitoria 1557: Ch. 8)

Even if they only appear in embryonic form, the rubrics of a future ‘society of nations’ are clearly and succinctly mapped out in this passage, providing an ingenious antecedent to what would only be put into practice some four centuries later, as a consequence of which Vitoria might be viewed as the founder of international law.

The impact of the discovery of America on the price of goods, and even on the value of currency, completely disrupted economic relations within the Iberian Peninsula, most severely in the city ports such as Seville. This led some theologians to speculate on the laws of supply and demand, as well as on the nature and conditions regarding the borrowing and lending of money, and eventually led them to state for the first time the ‘theoretical principles

of quantitative analysis' (Vilar 1974; Grice-Hutchison 1952). Martín de Azpilcueta and Tomás de Mercado developed their work along these lines in the sixteenth century; the former wrote the famous *Comentario resolutorio de cambios (A resolutory commentary on exchange)* (1556) while the latter was author of the equally famous *Suma de tratos y contratos (Summa of deals and contracts)* (1569).

Juan de Mairena was an outstanding Jesuit, an advocate of highly advanced social doctrines such as the common ownership of goods, and of theories such as that of regicide which he set out in *De rege et de regis institutione (On the king and his training)* (1599) - this work was banned in France, because it was thought to have played a part in the assassinations of kings Henry III and IV. Worth mentioning within the context of the Jesuits is the polemic known as 'de auxiliis' which took place between Luis de Molina and Domingo Báñez (the latter was a Dominican). This was sparked off by Molina's work *Concordia liberi arbitrii cum gratiae donis, divina praescientia, providentia, praedestinatione (The harmonious concordance of free will with the gifts of grace, divine prescience, providence and predestination)* (1588) in which he posed the question of the difficulty of reconciling divine grace and the provident omniscience of God with human liberty. Molina's doctrine availed itself of the traditional scholastic notion known as 'simultaneous concurrence', but Báñez's counter-argument in his *De vera et legitima concordia liberi arbitrii cum auxiliis gratiae Dei (On the true and legitimate concordance of free will with the aid of God's grace)* (1595) led Molina to formulate the theory of the 'middle science', the theoretical core of which were the so-called 'futuribles' (conditional futures which, despite never having occurred, can be freely determined by man and known as such by God). Báñez restricted himself to re-formulating his own theory of 'physical premonition', incorporating subtle new disquisitions such as those known as the 'science of vision' and the 'science of simple understanding'.

However, from a philosophical point of view, the most important Jesuit philosopher was Francisco Suárez, whose major works include *Disputationes metaphysicae (Debates on metaphysics)* (1597) and *De legibus ac de Deo legislatore (On laws and on God as legislator)* (1612). The first of these (although belonging to Thomist scholasticism, deviated from it significantly to the point of constituting an independent doctrine which became known as *Suarism*), takes original standpoints on issues such as the formal distinction between essence and existence, the uniqueness of the principle of individuation, and the consideration of 'real being' as the starting point for metaphysics. In his work he examined the transcendental properties of being and its grounding principles, which he then applied to the two fundamental classes of being: the infinite or uncreated (God) and the finite or created (humans). In so doing, he abandoned the medieval style of the 'commentary' and adopted instead a form of argument which proceeded according to its own intrinsic necessities, giving his work an unmistakable air of 'modernity'. This is most evident in his construction of a systematic metaphysical corpus which existed independently of Aristotelian-style scholastic commentary. It resulted in the book appearing in a host of editions as soon as it was published, and it even found its way onto the syllabus in numerous Protestant universities throughout Europe. Its influence on modern authors - Descartes, Spinoza, Leibniz, Vico - cannot be disputed.

4 Decline and the seeds of recovery

The pinnacle of Spanish Scholasticism which was reached with Suárez marked the end of a glorious epoch. The next two centuries were centuries of decline; in the seventeenth, eighteenth, and a significant portion of the nineteenth centuries, philosophical Spain lost its nerve, but a faint intellectual flame was kept alight by a few minority groups who thus saved the tradition. In the eighteenth century the *Novatores* (renovators) movement, headed by Andrés Piquer, gradually laid the foundations for a future renaissance which never materialized. A few enlightened thinkers such as B.J. Feijoo, G.M. de Jovellanos and Francisco Cabarrús helped to check the landslide, but the French invasion under Napoleon Bonaparte put an end to everything. The subsequent civil wars - the so-called 'Carlist Wars' - left an intellectual wasteland behind them. It was only in the middle of the nineteenth century that Julián Sanz del Río managed to turn the tide: his trip to Germany in 1843 and his contact with the few remaining disciples of Christian Krause in Heidelberg, allowed him to take back a few seeds of Krausist idealism to his Iberian homestead where they blossomed into a movement of spiritual and cultural renewal. This had its most important centre of diffusion at the famous Institución Libre de Enseñanza (Institute of Free Education), founded in 1876 by Francisco Giner de los Ríos and other Krausist teachers, among whom Nicolás Salmerón and Gumersindo de Azcárate were the most distinguished.

The philosophy of Giner de los Ríos was the result of the fruitful marriage of Krausist idealism and the positivism

which dominated the epoch, a strikingly original movement known in the history of Spanish philosophy as ‘Krauso-positivism’. The tangible gains of this philosophy were most apparent in the sphere of education. From the Institución Libre de Enseñanza, which was dedicated to primary and secondary education, emerged organizations such as the Junta para Ampliación de Estudios e Investigaciones Científicas (Group for the broadening of scientific studies and research) in 1907, of which Santiago Ramón y Cajal, winner of the Nobel Prize for Medicine in 1906, and José Castillejo, an eminent Romanist, were the principals. After its foundation in 1910, the Residencia de Estudiantes was run by Alberto Jiménez Fraud, whose pioneering spirit and aesthetic sensibility made it the cradle of modernism and the avant garde. Among others, Juan Ramón Jiménez, Frederico García Lorca, Rafael Alberti and Luis Buñuel studied there. Following in the wake of this great wave of renewal, mention should also be made of the foundation of the Instituto-Escuela (1918), the Residencia de Señoritas (1915) and the Misiones Pedagógicas (1931), organized and run by Manuel Bartolomé Cossío, the successor to Giner de los Ríos.

5 The Generation of 1898

The stimulus towards recovery provided by Krausism and institutionism eventually came to fruition in a powerful regenerative movement which was to bear its first great literary and philosophical fruits with the so-called Generation of 1898. From a philosophical perspective, the most eminent of this group is Miguel de Unamuno, a precursor of existentialism who made the contradiction between reason and life the crux of a philosophical conception of the world marked by anguish (*Congoja*) in the face of death. This led to his most famous work - *Del sentimiento trágico de la vida (The tragic sense of life and peoples)* (1912) - where he describes the struggle to attain personal immortality in the face of the uncertainties of reason. That struggle takes the form of an agonistic interplay between life and death in its religious version, represented by him in his book *La agon'a del cristianismo (The agony of Christianity)* (1925). As with the French existentialists, this nexus of problems was to be carried over into a literary oeuvre which consisted of a variety of novels, poetry and theatrical works (see Existentialism).

6 The School of Madrid

The secularizing philosophical theories of the Generation of 1898 broke with the philosophical tradition marked by its links with Catholicism, and this opened up the way for an important transformation of the entire discipline. Within this rebuilding process the so-called Generation of 1914 played a role of utmost importance. Most prominent here was José Ortega y Gasset who, after his education at the Neo-Kantian school of Marburg, reacted against its teachings to found a philosophy of his own. This doctrine is to be found in works such as *Meditaciones del Quijote (Meditations on Quixote)* (1914) and *El tema del nuestro tiempo (The modern theme)* (1923), among others. Ortega suggested the following maxim as the key to his philosophy: ‘ I am myself and my circumstances, and if I cannot come to terms with these, then I can never come to terms with myself’ ([1914] 1983 (1): 322).

The framework consisting of the self and its circumstances was specifically *life as radical reality*, by means of which Ortega attempted to overcome every form of rationalism. ‘Reason cannot, and need not aspire to supplant life’, he said in 1914; this approach to the matter was fully developed in *El tema de nuestro tiempo*, in which he proposed that culture be replaced by life, since it was the latter that constituted primary reality. By life Ortega here meant individual human life:

> This radical reality on the basis of the strict observation of which we must ultimately establish and secure all of our knowledge of anything, is our life, human life. Human life as radical reality is only the life of every individual, it is only my own life. For the sake of linguistic convenience I will sometimes call it ‘our life’, but it must always be understood that with this expression I am referring to the life of every individual and not to that of others nor to an assumed plural and collective life. What we call ‘other people’s lives’, the life of our friend, of the one we love, is already an added element which encroaches on the scene which is my life. By calling it ‘radical reality’ I do not mean that it is the only, nor even the highest, most respectable, sublime or supreme reality, but simply that it is the root - hence the term radical - of all of the others, in the sense that the latter, of whatever nature, must, for us to experience them as reality, somehow make themselves present or, at least, announce themselves by causing the boundaries of our own life to tremble. Therefore this radical reality - my life - is so little ego-oriented, so utterly non-solipsistic, that it is, quintessentially, the space or scene which is offered to me and opened up so that any other reality can manifest itself therein. Hence no knowledge of anything is sufficient, that is to say sufficiently profound or radical, if it does not begin by pinpointing the place

> and specifying the way, within the sphere that is our life, wherein and whereby that something appears, shows itself, bursts forth or rises up, in short, exists.
>
> (Ortega [1923] 1983 (7): 99-101)

As one of the finest commentators on Ortega's philosophy has said, what we are witnessing here is a metaphysical revolution characterized by the 'real correlation of self and world, of thought and its object, of love and what is loved: (that is) it is exactly what we call life. Living is feeling things, thinking things; it is the coexistence of self and world. Life is the primary, fundamental fact, absolutely present and evident, upon which all philosophy must base itself. Life, on a deeper and more essential plane, is the 'cogito' of the new philosophy' (García Morente 1987: 67).

This discovery of life as metaphysical reality, demanded a method through which it could be rendered accessible: the method was that of 'vital reason', which was conceived of as the same thing as living since 'reason is only a form and function of life'. What we have here is a new conception of reason that goes beyond the rationalism of 'pure reason', as an expression of modernity. According to this new formulation, reason ' is nothing more than a tiny island floating on the sea of primary vitality; far from supplanting the latter, it must depend on it for support and sustenance, just as each one of the members of the body is given life by the body as a whole' (Ortega [1923] 1983 (3): 177-8).

We are faced with a philosophical programme which implies a reversal of the tenets of traditional Western culture - as opposed to a culture which unhouses life, a culture which gains sustenance from it, injecting it with new values; as opposed to the old irony of Socrates, the irony of Don Juan, who reaffirms spontaneity, sincerity and joyousness. This was the *modern theme* which consisted in ' making reason bow before vitality, placing it within the sphere of the purely biological, subordinating it to spontaneity'; a programme which, in its philosophical essence, might be formulated thus: ' pure reason must abdicate in favour of vital reason' (Ortega [1923] 1983 (3): 178).

The vitalism so described did not imply any form of irrationalism, but constituted a new mode of understanding which Ortega named 'ratio-vitalist', within which there took shape a new theory of knowledge, essential to which was the concept of 'perspective'. Ratio-vitalism was, then, a form of perspectivism.

Ortega's philosophical theories were to attract some important philosophical figures, such as Manuel García Morente, José Gaos María de Maetzu, Xavier Zubiri, María Zambrano, Luis Recasens Siches and Lorenzo Luzuriaga who, in 1933, would give rise to what can be referred to as the School of Madrid, even if the brevity of its existence (a result of the civil war) prevented it from producing what seemed destined to be a wealth of fine philosophy.

Although he belonged in principle to the School of Madrid, Xavier Zubiri was to follow his own highly original philosophy. His starting point was one of his own theories, according to which the human is a 'reality animal' characterized by what Zubiri called 'sentient intelligence' - that is to say, a sensuous intellection and an intellectual sensing - by means of which he is able carry out his most proper function, which consists in the 'actualization of the real in so far as it is real'. This theory, which had already appeared in embryonic form in his first work, *Naturaleza, Historia, Dios (Nature, history, God)* (1944), was to be consistently developed and elaborated in his later works *Sobre la esencia (On essence)* (1962) and, above all, in his trilogy on *Sentient Intelligence* (1980).

7 Philosophy in exile

Despite the fact that the civil war (1936-9) destroyed what ought to have been the natural continuity of the School of Madrid, its members (the majority of whom went into exile) proceeded to produce a magnificent body of philosophical work in the different countries which granted them asylum. Examples include José Gaos, who formulated his own particular theory which he described as 'a philosophy of philosophy', and which has had an obvious impact on the history of Hispanic thinking; María Zambrano, who expounded and developed her method of 'poetic reasoning'; Luis Recasens Siches, who made use of Ortega's vital reason in order to produce a highly individual philosophy of law; Manuel Granell, who carried out similar work in the field of logic; José Ferrater Mora, author of a famous dictionary of philosophy, among other publications; and finally, Eduardo Nicol, an

original and independent thinker who developed a meditation on the nature and future of philosophy.

Probably the most original thinker of all of those exiled from Spain was Juan David García Bacca, author of a multifaceted and highly varied oeuvre which revolved principally around a metaphysical meditation on the artificial world created by man, while within the sphere of Marxist socialism and related fields particularly significant is Fernando de los Ríos who later went on to write more historical studies, although his leanings were always towards social and political philosophy. His *Mi Viaje a la Rusia Soviética (My journey to Soviet Russia)* (1921) was decisive in keeping the Spanish Socialist Party away from the Third International. Within the tradition of more rigorously orthodox Marxist thought, mention should be made of the then young thinker Adolfo Sánchez Vásquez whose philosophical reflections have been directed at its more practical aspects, especially those relating to art and aesthetics, although his most important work continues to be *Filosof'a de la praxis (The philosophy of praxis)* (1967).

8 Independent thinkers

From the end of the nineteenth and during the first half of the twentieth century, a number of independent figures - difficult to categorize in terms of any one school or movement - appeared in Spain. The most striking of these was Jorge Ruiz de Santayana, better known as George Santayana, the name with which he signed his writings, since they all appeared in English in the USA, where he was professor at Harvard. After 1912 he lived in various European countries - England, Spain, Italy - before finally settling down in Rome where he remained until his death in 1952. He was the author of a highly diverse body of work, without doubt the most important of which was *The Realms of Being* (1942). His philosophy is best categorized as a form of sceptical materialism marked by a heightened aesthetic sensibility. Although an atheist, he placed a great deal of importance on religion from a symbolic point of view (in fact he lived and died in a monastery) which has led people to speak of an aesthetic naturalism.

A second independent thinker was the priest Angel Amor Ruibal, author of the dense and colossal work *Los problemas fundamentales de la filosof'a y el dogma (The fundamental problems of philosophy and dogma)* most of which remained unedited during the author’s lifetime, but which has been gradually published since his death in 1930. It has now been shown that these writings contain an original philosophical system which might be termed correlationism or transcendental relativism, since it is not a pure and simple relativism, but rather a universal atomism in which the value of the individual is retained. It sees the universe as being composed of a series of interrelated fundamental systems, in which beings constitute ‘links’ which subordinate one another until they interconnect - like a vertically structured chain running from the lowest being to the highest - in a continuous universal interaction which is monistic in nature, since the parts acquire meaning only through their relation to the whole.

A third thinker of importance is Eugenio d’Ors who, at the start of his career, was noted for his close links with the Catalonian Government and in his later years for those he established with General Franco’s dictatorship, but who led a solitary existence during the greater part of his life. Within his varied output he developed a sort of ‘figurative thinking’ upon which he constructed certain ‘figures of intelligence’, the object of which was to arrive at a ‘harmonious reason’, which he located precisely half way between life and pure reason. This in turn provides the basis for what he called a ‘Keplerian reform of philosophy’, in which the principle of contradiction was replaced by the principle of participation, and that of sufficient reason by that of required function. All of this gave rise to an understanding of the universe as a syntactical structure.

9 From dictatorship to democracy

Franco’s victory in the civil war of 1936-9 determined the subsequent course taken by philosophy in Spain. First, it forced the majority of thinkers and teachers of philosophy into exile; second, it imposed on secondary and university education a sort of ‘official philosophy’, the basis of which was the most antiquated scholasticism. It should be stressed, however, that the later peaceful transition from the dictatorship period to democracy owed much to those philosophers who managed to retain an independent sense of judgment, by adhering to their liberal principles as far as the regime would allow. Among these is José Luis Aranguren who, starting out from a Roman Catholic standpoint, allowed his thinking to open out onto increasingly broad perspectives, incorporating the latest ideas to emerge from existentialism, from Anglo-Saxon thought and from the dialectical strains of Marxism.

Another is Pedro Laín Entralgo, who was deeply involved in work as a historian of medicine, which gradually opened up the way for him to formulate a philosophy of hope and a medical anthropology in which the reality of the ‘other’ is always present. Also of importance is Enrique Tierno Galván, a committed socialist who played an active role in politics - he became Mayor of Madrid - which did not prevent him from producing an important body of philosophical work, to which his book *Razón mecánica y razón dialéctica (Mechanical reason and dialectical reason)* (1969) is testimony.

Among those who kept the sacred torch of Ortega’s philosophy burning in Spain are Paulino Garagorri, an important exegete and compiler of Ortega’s work, and Julián Marías, perhaps the most faithful disciple of the master whose philosophy he has systematized in his magnum opus *Ortega* (1962, 1983). It was this group of thinkers who made it possible for the first sparks of rebellion against Franco’s regime to flare up in 1956.

See also: Collegium Conimbricense; Molinism

Translated by Dominic Moran and Gemma Belmonte Talero
JOSÉ LUIS ABELLÁN

References and further reading

Abellán, J.L. (1967) *Filosofía Española en América 1936-1966 (Spanish Philosophy in America)*, Madrid: Ediciones Guadarrama.(Currently the only existing work on the Spanish philosophical exile resulting from the Spanish Civil War.)

Abellán, J.L. (1978) *Panorama de la Filosofía Española Actual (A survey on the Current Spanish Philosophy)*, Madrid: Espasa-Calpe.(A panoramic and critical view of the Spanish philosophy after the 1936-9 Civil War.)

Abellán, J.L. (1979-92) *Historia Crítica del Pensamiento Español (A Critical History of Spanish Thought)*, Madrid: Espasa-Calpe, 7 vols.(Spanish philosophy from the Middle Ages up to the present, from the perspective of the history of ideas.)

Amor Ruibal, A. (1914-36) *Los Problemas Fundamentales de la Filosofía y el Dogma (Fundamental Problems of Philosophy and Dogma)*, Madrid: Victoriano Suárez, 10 vols. (A historical-critical presentation of the author’s philosophical system.)

Azpilcueta, M. de (1556) *Comentario Resolutorio de Cambios (A Resolutary Commentary on Exchange)*, Salamanca: Andrea de Pontinariis.(Develops a theory of money as a starting point for the moral regulation of the economics between individuals and nations.)

Bañez, D. (1595) *De Vera et Legitima Concordia Liberi Arbitrii cum Auxiliis Gratiae Dei (On the True and Legitimate Concordance of Free Will with the Aid of God’s Grace)*, unedited manuscript, Inquisition Section at the National Historical Archive, Madrid: File 4.437.(A fundamental text for understanding the position of the Dominican Order regarding the controversial ‘de auxiliis’ about divine grace.)

Cano, M. (1563) *De Locis Teologicis (On Theological Standpoints)*, Salamanca: Matías Gast.(A basic text about the different locations in the Holy Scriptures where the revealed truths abide, which served as the doctrine of the Counter-Reformation.)

Estudios Lulianos (Lullian Studies) (1957-), Palma de Mallorca.(A journal dedicated to the study and research on Ramón Llul - essential work for researching this author.)

García Morente, M. (1987) *Escritos Desconocidos e Inéditos (Unknown and Unedited Writings)*, Madrid: BAC. (A posthumous publication with some of the author’s most important philosophical writings.)

Guy, A. (1983) *Histoire de la Philosophie Espagnole (A History of Spanish Philosophy)*, Toulouse: Université de Toulouse-Le Mirail.(A complete history of Spanish Philosophy from the Middle Ages to the present; it gives priority to the twentieth century, to which half the volume is dedicated.)

Grice-Hutchinson, M. (1952) *The School of Salamanca. Readings in Spanish Monetary Theory*, Oxford: Clarendon Press.(Serious studies and investigations of the monetary discoveries of sixteenth-century Spanish theologians.)

Laguna, A. (1542) *Discurso de Europa (Discourse of Europe)*, Colonia: Lupum Ioannes Aquensis.(Fundamental text for knowing the Europeanism characteristic of the Erasmists who became ideological speakers during the reign of Charles V.)

Mairena, J. de (1599) *De Rege et de Regis Institutione (On the King and his Training)*, Petrum Rodericum: Toledo.(An important work on the political philosophy which defends the legitimity of regicide and provoked great controversies in its time.)

Marías, J. (1962) *Ortega. I. Circunstancia y vocación (Ortega. I. Circumstances and Vocation)*, Madrid: Revista de Occidente.(The first systematic treatment of Ortega’s philosophy.)

Marías, J. (1970) *Antropología Filosófica (Philosophical Anthropology)*, Madrid: Revista de Occidente.(The most complete presentation of what the author calls the 'empirical structure of human life', the central point of his philosophical ideology.)

Marías, J. (1983) *Ortega. II. Las Trayectorias (Ortega. II. The Trajectories)*, Madrid: Alianza Editorial.(In this volume, the author culminates and resolves the dilemmas that arise in the first volume.)

Mercado, T. de (1569) *Suma de tratos y Contratos (Summa of Deals and Contracts)*, Salamanca: Matías Gast. (Observance of marine commercial transactions in Seville resulted in the author's formulation of the first laws on economic and financial exchange, as well as his moral ideology.)

Molina, L. de (1588) *Concordia Liberi Arbitrii cum Donis, Divina Praesciencial, Providentia, Praedestinatione (The Harmonious Concordance of Free Will with the Gifts of Grace, Divine Prescience, Providence and Predestination)*, Lisbon: Antonio Ribeiro.(Molina's formulation of his doctrine on 'middle knowledge', intended to make human freedom compatible with divine omnipresence.)

Ortega y Gasset, J. (1914) *Meditaciones del Quijote (Meditations on Quixote)*, Madrid: Residencia de Estudiantes.(Ortega's first book, essential for understanding his philosophy.)

Ortega y Gasset, J. (1923) *El Tema de Nuestro Tiempo (The Modern Theme)*, Madrid: Calpe. (Presentation of Ortegan 'ratiovitalism', central to his doctrine on 'vital reason'.)

Ortega y Gasset, J. (1983) *Obras Completas (Complete Works)*, Madrid: Alianza Editorial.(The works of this twentieth-century philosopher can serve as a point of reference for the whole of this entry.)

Ríos, F. de los (1921) *Mi Viaje a la Rusia Soviética (My Journey to Soviet Russia)*, Madrid: Caro Raggio. (Describes the difference that separates the Spanish socialism from the Third International.)

Sánchez Vázquez, A. (1967) *Filosofía de la Praxis (The Philosophy of Praxis)*, Mexico: Grijalbo.(An interpretation of Marxism in relation to practical questions, especially about the concept of 'praxis' as a central category.)

Santayana, G. (1942) *The Realms of Being*, New York: Scribner.(Defines and describes the four reigns of the being which Santayana considers fundamental: matter, essence, truth and spirit.)

Soto, D. de (1553) *De Iustitia et Iure (On Justice and Law)*, Salamanca: Ioannis Aute. (A theological text fundamental for understanding the differences that developed over the centuries between eternal law, natural law and positive law resulting in the foundations of the philosophy of law.)

Suárez, F. (1597) *Disputaciones Metaphisicae (Debates on Metaphysics)*, Salamanca: Juan y Andrés Renaut.(A combination of Spanish Reinassance scholastics and modern philosophical thought in Europe.)

Suárez, F. (1612) *De Legibus ac de Deo Legislatore (On Laws and on God as Legislator)*, Coimbra: Diego Gómez Loureyro.(The practical aspects of Suárez's metaphysical work are addressed in this great theological-political treaty, elaborated in ten books.)

Tierno Galván, E. (1969) *Razón Mecánica y Razón Dialéctica (Mechanical Reason and Dialectical Reason)*, Madrid: Tecnos.(A discussion from the Marxist viewpoint of the relations between scientific determinism and dialectics, as well as of its practical consequences.)

Unamuno, M. de (1912) *Del Sentimiento Trágico de la Vida (The Tragic Sense of Life and Peoples)*, Madrid: Renacimiento.(A major work of Unamuno's philosophical ideology, in which he explains the irreductible opposition between 'reason' and 'life'.)

Unamuno, M. de (1925) *La Agonía del Cristianismo (The Agony of Christianity)*, Madrid: Renacimiento.(The religious version of the 'tragic feeling' achieves its highest dramatism in this book.)

Vilar, P. (1974) 'Los Primitivos Españoles del Pensamiento Económico. "Cuantitativismo y Bullonismo"' (The Primitive Spaniards of Economical Thought: Cuantitativism and Bullonism) in *Crecimiento y Desarrollo (Growth and Development)*, Barcelona: Crítica, 135-62.(A lucid presentation of 'Cuantitativism', developed by Spanish thinkers of the sixteenth century.)

Vitoria, F. de (1557) *Relectiones Theologicae (Theological Reflections)*, Lyon: Jacobo Boyer.(All the texts that constitute this work, which includes the three mentioned in §3 above, form the first scientific proposal about the philosophy of law and international law.)

Vives, J.L. (1526) *De Subventione Pauperum (On Poor Relief)*, Brujas: Hubert de Crook.(One of the first formulations of doctrines of social welfare found in European thought.)

Vives, J.L. (1529) *De Concordia et Discordia in Humano Genere (On Harmony and Discord in Humankind)*, Amberes: Michele & Hillenium.(The authors' pacifism reaches a maximum radical level in this work.)

Vives, J.L. (1538) *De Anima et Vita (On the Soul and on Life)*, Basilea: Robert Winter.(This text is one of the first foundations of scientific psychology.)

Zubiri, X. (1944) *Naturaleza, Historia, Dios (Nature, History, God)*, Madrid: Editorial Nacional.(The first presentation, albeit in an embryonic form, of Zubiri's idea of man as an 'animal of realities'.)

Zubiri, X. (1962) *Sobre la Esencia (On Essence)*, Madrid: Sociedad de Estudios y Publicaciones.(Develops the fundamental concepts of Zubiri's philosophical system - a key work to understand this author.)
Zubiri, X.(1980) *Inteligencia Sentiente (Sentient Intelligence)*, Madrid: Alianza Editorial, 3 vols.(In this trilogy the author's philosophy and his development of what he calls 'feeling intelligence' reach their apex.)

Species

The diversity of life is not seamless but comes in relatively discrete packages, species. Is that packaging real, or an artefact of our limited temporal perspective on the history of life? If all living forms are descended from one or a few ancestors, there may be no real distinction between living and ancestral forms, or between closely related living animals.

Received wisdom holds that species are the 'units of evolution', for it is they that evolve. They are the upshot *of evolutionary processes, but, if species and not just their component organisms compete with one another, they are also important* agents *in the evolutionary process. If so, species are real units in nature, not arbitrary segmentations of seamless variation.*

The 'species problem' has been approached from two angles. One focus has been on specific taxa of the tree of life. What would settle whether some arbitrarily chosen organism is a member of homo sapiens *or* canis familiaris*? This is sometimes known as the 'species taxon' problem. An alternate way of approaching diversity has been to ask what all species have in common. What do all the populations we think of as species share? This is the 'species category' problem.*

One idea is to group organisms into species by appealing to the overall similarity. This 'phenetic' conception is in retreat. Most contemporary species definitions are relational, the animals that compose pan troglodytes *are a species, not because they are all very similar (they are very like the pygmy chimps as well) but because of their relations amongst themselves and with their ancestors. The most famous relational definition is the 'biological species concept', according to which conspecific organisms are organisms that can interbreed, however different they look.*

Relational species definitions aim to define a category of theoretical and explanatory interest to evolutionary and ecological theory. Given that there are many explanatory interests, one problem in evaluating these accounts is to determine whether they are genuinely rivals.

1 Species as historical entities

There have been three main views of the species category. So-called 'phenetic' species concepts define species membership and difference by appeal to some measure of overall morphological, genetic or behavioural similarity. This view has slid from favour. First, there are many measures of similarity at any one time. These give different results and choice between them is arbitrary. Second, the similarity conception of species entails that the segmentation of the tree of life is arbitrary. As the population gradually changes there are many ways of dividing up the twig, many different similarity groupings, on the tree of life leading to us. The distinction between *homo sapiens* and *homo erectus* emerges as nothing but a curator's convenience. Third, the collections of organisms so recognized would be of no particular evolutionary significance. If species are just collections of similar organisms, measured by one of the many different similarity measures, the species category is not a natural kind.

The alternatives accept some version of the Ghiselin (1974) and Hull (1978) proposal that particular species are defined by their histories. 'Platypus' names a segment of the tree of life. Species come into existence, change, and go extinct. Whereas any appropriately latticed arrangement of carbon atoms is a diamond, a *rattus rattus* lookalike produced through convergent evolution near Alpha Centauri is not a member of the rat species. But to say that species are historical kinds is one thing, to say what historical kinds, another. Why do we regard *canis familiaris* as a single species rather than as a group of sister species? What do we need to find out to determine whether the Neanderthals were a separate species of *homo*, or a mere subspecies of *homo sapiens*?

The biological species concept offers one answer to this question: a species is a population of actually or potentially interbreeding organisms. Coexisting with the biological species concept are a plethora of *phylogenetic species concepts*. These all identify species with segments of lineages of ancestral/descendent populations. Species are segments of a phylogenetic tree. Which segments? Proponents of phylogenetic definitions differ in their proposals for specifying the depth of the lineage in time, and its cross-section at a time.

One of our problems is to determine the extent to which the biological species concept genuinely competes with

the variants of the phylogenetic concept. I, like many, think the phylogenetic and biological species concepts divide the labour between them. Species are chunks of the genealogical nexus between speciation events. The biological species concept gives us an account of speciation events, that is, the process through which a single ancestral species becomes two or more descendent species.

2 The biological species concept

The biological species concept takes the reproductive community to be central to the role of species as evolutionary units. Adaptation and speciation require some isolating mechanism so that an incipient species, a small population in a new selective regime, can preserve in its gene pool the evolutionary innovations that develop within it. An unprotected gene pool will be diluted by migration. The special genetic information will disappear if there is substantial flow between it and the parent population. There can be no special suite of adaptations without some form of isolation; no protection of that suite of adaptations without entrenching that isolation which erects permanent barriers between old and descendant populations. The biological species concept identifies a category of populations - reproductively isolated populations - that can evolve distinctive adaptations.

This proposal faces problems important enough to show that the biological species concept is at best an incomplete account of the nature of species.

First, the notion of a reproductively isolated community is an idealisation, and we can legitimately choose different idealisations. We have to bear in mind two problems. Lineages can be genuinely separate despite some gene flow. There are occasional ‘freak’ crossings. Major Mitchell cockatoos occasionally hybridise with galahs, but these two lineages are distinct. Lineage-crossing is common amongst plants, amongst whom gene flow across species boundaries is common. There are more problematic cases. As a consequence of human modification of river ecologies in New Zealand, black stilts now hybridise with pied stilts sufficiently often to threaten the survival of the black stilt lineage - if it is a lineage. But we are faced not just with freak crossings but pseudo-division. Groups which merely happen to breed only amongst themselves do not constitute a new lineage. Royal families are members of *homo sapiens* however unsullied their blood lines. So the reproductive criterion yields no unique segmentation of organisms into species. The notion of ’potentially interbreeding’ neither is nor can be made fully precise (Kitcher 1989; O’Hara 1993). There is no unique count of protected genetic pools. Gene flow really does come in degrees.

Second, the limitation of gene flow is just one factor that make a lineage ’cohesive’. A buffered gene pool is not all that matters in explaining the distinctness and common fate of a group of populations. Both phylogeny, shared environment, and exposure to a common selective regime must equally be part of the cohesiveness of a species. Ehrlich and Raven (1969) show that in many species gene flow is very limited between local populations (’demes’). So we can have cohesiveness with little flow.

A third problem concerns time. The biological species concept seems to give no good way of segmenting a lineage over time that recognizes that the *Erectus/Sapiens* distinction is a real division in a continuous lineage. It gives no good way of deciding whether organisms living at different times are members of the same species. Perhaps we might suppose that an organism ceases to be conspecific with a member of a later generation if it would not recognize that changed organism as a potential mate. But then, in a gradually evolving lineage the segmentation into species would depend on the choice of baseline. The biological species concept needs supplementation. The interbreeding criterion should not be applied to organisms at different times. If Abe and Adolf are members of different generations they are in the same species, if Adolf has descended from organisms conspecific with Abe by the interbreeding criterion and no speciation event has intervened in the genealogical tree.

Finally, the biological species concept is mute in the face of obligatorily asexual organisms. These are not so rare or unimportant that they can be fudged away as a minor exception.

3 Phylogenetic species concepts

The basic idea of phylogenetic species concepts is to identify a species with a segment of a phylogenetic tree between two speciation events, or between speciation and extinction. Two organisms are members of the same species if they are part of the same lineage, and no speciation event has separated them. On this view, two conspecific organisms could be morphologically very different from one another, so long as the lineage of which

they are part has not split in the course of its evolutionary transformation. Phylogenetic species definitions therefore owe us an account of speciation, extinction, and a method for counting lineages. Granted satisfactory solutions to these problems, phylogenetic accounts have clear attractions. They may apply to both sexual and asexual species. If a suitable way of individuating lineages can be developed, a phylogenetic account might be noncommittal on the causes of ‘cohesion’. There is some advantage to that if in different lineages, gene flow, stabilising selection and phylogenetic inertia have different weights in establishing cohesion. The idea is that the phylogenetic species concept is founded in the objective branching *pattern* of nature rather than in a theory of evolutionary *process*. Humans are not conspecific with protists even though we descend from them because there exist punctuations of the tree in which lineages split into descendent lineages.

But it is not clear that this potential advantage is actual. Consider a minor evolutionary pattern: the fissuring or pseudo-fissuring of the lineage of the Australian brushtail possum into many Australian populations and assorted populations in the New Zealand archipelago. Which of these populations are new species? The phylogenetic species concept - in some incarnations - gives the crazy result that most are. Some proposals for counting lineages rely on the idea of a population having a separate historical fate or entering a new adaptive zone (Van Valen 1976; Wiley 1978). If that were right, the New Zealand possum population would be a separate species, as their adaptive zone and fate is distinct from the Australian population. So would be tiny populations that persisted for a few generations on small islands or isolated fragments of bushland. Phylogenetic identification of independent lineages seem to either count too finely, be excessively vague, or to tacitly depend on the biological species concept, and thus to inherit its problems with asexual species.

So the most obvious way of delimiting chunks of a tree is through appeal to reproductive isolation. Yet this ignores asexual species. Even if asexuality were rare, the problem would be serious. Reproductive isolation is alleged to be constitutive of being a species, not merely a good symptom of it (Kitcher 1989). A heroic solution would be to deny that asexual organisms are in species. This is heroic, for we would then have to explain just why ’pseudo-populations’ of, for example, some whiptail lizards seem to be species. Moreover, and more seriously the distinction between asexuality and sexuality is not sharp. Rather, asexuality is the endpoint of a continuum of degrees of gene flow whose other point is the promiscuity of plant hybridisation (Templeton 1989).

An ideal response would be to find a genuine equivalent in asexual organisms of reproductive isolation amongst sexual ones. We could try to find string to bundle the clones together, and an equivalent of the attainment of isolation to underwrite the segmentation of the bundles into two. Some version of Templeton’s ’cohesion’ species concept may do the job: Templeton argues that in most species ’cohesion’ - the clumping in the space of possible organisms which allows us to recognize species in the first place - is produced by ecology and selection, not just gene flow. Asexual species might be cohesive just in virtue of these other forces, though it remains unclear what the asexual equivalent of speciation would be.

Kitcher (1989) urges on us a more radical moral: multiply species concepts. He thinks it plain that asexual organisms clump into species. The phenetic concept of morphological distinctness is *constitutive* of the species category for asexual organisms. Why then not say that it is constitutive of at least *a species category* for sexual organisms? If Kitcher is right about asexual organisms, the same conclusion is not forced elsewhere. In both sexual and asexual lineages, clumped diversity is a residue of a historical process. But with sexual species that historical process is one in which gene exchange and a protected gene pool played a central role. That process is important enough to mark by regarding sexual organisms as clumped into species in a different sense than are asexual organisms. It is not at all obvious that a group of similar asexual clones form an evolutionary unit; the sexually reproducing population does.

We might have good reasons to consider organisms grouped by their intrinsic similarities of morphology, physiology or behaviour. It is perfectly appropriate to formulate hypotheses about the molecular mechanisms of, say, the retroviruses whether or not they all spring from the same stock, a hypothesis about the mechanism of reverse transcription could still be true and their distinctive adaptation might have a single molecular basis. Not all biology is historical biology. Even so, the validity of physiological groups in functional biology does not show that there is a notion of species defined by appeal to a shared physiology.

What emerges from all this? The most plausible account of species seems to be an evolutionary account: species are lineages between speciation events. The biological species concept re-emerges as an account of the process by

which a lineage divides into two. Lineages split when their components become reproductive isolated, one from another. Lineages converge when two formerly isolated lineages become a reproductive community through, for example, hybridisation. Two central problems remain to be solved. First, the inherent roughness of the notion of reproductive isolation means that the identification of species is in principle, not just in practice less sharp than some have claimed. Second, the status of asexuality remains problematic.

See also: Evolution, theory of; Taxonomy

KIM STERELNY

References and further reading

Ehrlich, P. and Raven, P. (1969) 'The Differentiation of Populations', *Science* 165: 1228-32.(A classic statement of a problem for the biological species concept: many species seem to preserve their distinctive suite of morphological, physiological and behavioural traits - their cohesion - with very little gene flow within them.)

Ereshevsky, M. (ed.) (1992) *The Units of Evolution: Essays on the Nature of Species*, Cambridge, MA: MIT Press.(An excellent collection of most of the central papers on the species problem.)

Ghiselin, M. (1974) 'A Radical Solution to the Species Problem', *Systematic Zoology* 23: 535-44.(The first clear statement of the idea that species are individuals not sets, an idea developed further in Hull 1978.)

Hull, D. (1978) 'A Matter of Individuality', *Philosophy of Science* 45: 335-60.(One of the two classic defences of the idea that a species' essential property is its history. This idea was originally formulated as the claim that a species is not a set of organisms but is instead, like the British Empire, a single thing though one scattered in space and time. It is now widely agreed that this is not the best formulation of that insight. Species may be kinds or sets, but if so the membership condition is historical.)

Kitcher, P. (1989) 'Some Puzzles about Species', in M. Ruse (ed.) *What the Philosophy of Biology Is*, Dordrecht: Kluwer, 183-208.(A defence of the idea that the notion of species is equivocal. We have many good species concepts, and we should not suppose that one is the 'right' concept.)

Mayr, E. (1988) *Towards a New Philosophy of Biology*, Cambridge, MA: Harvard University Press.(Mayr is the most famous defender of the biological species concept. Parts VI and VII defends his views on species and speciation against all comers.)

O'Hara, R.J. (1993) 'Systematic Generalization, Historical Fate and The Species Problem', *Systematic Biology* 42: 231-46.(A very clear demonstration of the inherent imprecision of the notion of reproductive isolation, and of the consequences of that imprecision for the notion of species.)

Ridley, M. (1989) 'The Cladistic Solution to the Species Problem', *Biology and Philosophy* 4: 1-16.(Defends the simplest possible version of a phylogenetic species definition.)

Templeton, A. (1989) 'The Meaning of Species and Speciation: A Genetic Perspective' in D. Otte and J. Endler (eds) *Speciation and Its Consequences*, Sunderland, MA: Sinauer, 3-27.(Following on from Ehrlich and Raven 1969 in arguing that species must be defined by appeal to all the factors that make species cohesive, not just gene flow.)

Van Valen, L. (1976) 'Ecological Species, Multispecies and Oaks', *Taxon* 25: 233-9. (Defends a version of the phylogenetic species definition in which ecological factors play the central role in counting lineages.)

Wiley, E. (1978) 'The Evolutionary Species Concept Reconsidered', *Systematic Zoology* 27: 17-26.(Wiley defends a version of the phylogenetic species definition in which the notion of a shared historical fate plays the central role in counting lineages.)

Speech acts

Making a statement may be the paradigmatic use of language, but there are all sorts of other things we can do with words. We can make requests, ask questions, give orders, make promises, give thanks, offer apologies and so on. Moreover, almost any speech act is really the performance of several acts at once, distinguished by different aspects of the speaker's intention; there is the act of saying something, what one does in saying it, such as requesting or promising, and how one is trying to affect one's audience.

The theory of speech acts is partly taxonomic and partly explanatory. It must systematically classify types of speech acts and the ways in which they can succeed or fail. It must reckon with the fact that the relationship between the words being used and the force of their utterance is often oblique. For example, the sentence 'This is a pig sty' might be used nonliterally to state that a certain room is messy, and further to demand indirectly that it be tidied up. Even when this sentence is used literally and directly, say to describe a certain area of a farmyard, the content of its utterance is not fully determined by its linguistic meaning - in particular, the meaning of the word 'this' does not determine which area is being referred to. A major task for the theory of speech acts is to account for how speakers can succeed in what they do despite the various ways in which linguistic meaning underdetermines use.

In general, speech acts are acts of communication. To communicate is to express a certain attitude, and the type of speech act being performed corresponds to the type of attitude being expressed. For example, a statement expresses a belief, a request expresses a desire, and an apology expresses a regret. As an act of communication, a speech act succeeds if the audience identifies, in accordance with the speaker's intention, the attitude being expressed.

Some speech acts, however, are not primarily acts of communication and have the function not of communicating but of affecting institutional states of affairs. They can do so in either of two ways. Some officially judge something to be the case, and others actually make something the case. Those of the first kind include judges' rulings, referees' decisions and assessors' appraisals, and the latter include sentencing, bequeathing and appointing. Acts of both kinds can be performed only in certain ways under certain circumstances by those in certain institutional or social positions.

1 Levels of speech acts

How language represents the world has long been, and still is, a major concern of philosophers of language. Many thinkers, such as Leibniz, Frege, Russell, the early Wittgenstein and Carnap, have thought that understanding the structure of language could illuminate the nature of reality. However noble their concerns, such philosophers have implicitly assumed, as J.L. AUSTIN complains at the beginning of *How to Do Things with Words* (1962), that 'the business of a (sentence) can only be to "describe" some state of affairs, or to "state some fact", which it must do either truly or falsely'. Austin reminds us that we perform all sorts of 'speech acts' besides making statements, and that there are other ways for them to go wrong or be 'infelicitous' besides not being true. The later Wittgenstein also came to think of language not primarily as a system of representation but as a vehicle for all sorts of social activity. 'Don't ask for the meaning', he admonished, 'ask for the use'. But it was Austin who presented the first systematic account of the use of language. And whereas Wittgenstein could be charged with conflating meaning and use, Austin was careful to separate the two. He distinguished the meaning (and reference) of the words used from the speech acts performed by the speaker using them.

Austin's attention was first attracted to what he called 'explicit performative utterances', in which one uses sentences like 'I nominate', 'You're fired', 'The meeting is adjourned' and 'You are hereby sentenced… ' to perform acts of the very sort named by the verb, such as nominating, firing, adjourning, or sentencing (see Performatives). Austin held that performatives are neither true nor false, unlike what he called 'constatives'. However, he came to realize that constatives work just like performatives. Just as a suggestion or an apology can be made by uttering 'I suggest… ' or 'I apologize… ', so an assertion or a prediction can be made by uttering 'I assert… ' or 'I predict… '. Accordingly, the distinction between constative and performative utterances is, in Austin's general theory of speech acts, superseded by that between saying something and what one does in saying it. This broader distinction applies to both statements and other sorts of speech acts, and takes into account the fact

that one does not have to say 'I suggest… ' to make a suggestion, 'I apologize… ' to make an apology, or 'I assert… ' to make an assertion.

The theory of speech acts aims to do justice to the fact that even though words (phrases, sentences) encode information, people do more things with words than convey information, and that when people do convey information, they often convey more than their words encode. Although the focus of speech act theory has been on utterances, especially those made in conversational and other face-to-face situations, the phrase 'speech act' should be taken as a generic term for any sort of language use, oral or otherwise. Speech acts, whatever the medium of their performance, fall under the broad category of intentional action, with which they share certain general features (see Action; INTENTION §5). An especially pertinent feature is that when one acts intentionally, generally one has a set of nested intentions. For instance, having arrived home without one's keys, one might push a button with the intention not just of pushing the button but of ringing a bell, rousing one's spouse and, ultimately, getting into one's house. The single bodily movement involved in pushing the button comprises a multiplicity of actions, each corresponding to a different one of the nested intentions. Similarly, speech acts are not just acts of producing certain sounds.

Austin identifies three distinct levels of action beyond the act of utterance itself. He distinguished the act *of* saying something, what one does *in* saying it, and what one does *by* saying it, and dubs these the 'locutionary', the 'illocutionary' and the 'perlocutionary' act, respectively. Suppose, for example, that a bartender utters the words, 'The bar will be closed in five minutes.' He is thereby performing the locutionary act of saying that the bar (that is, the one he is tending) will be closed in five minutes (from the time of utterance). Notice that what the bartender is saying, the content of his locutionary act, is not fully determined by the words he is using, for they do not specify the bar in question or the time of the utterance. In saying this, the bartender is performing the illocutionary act of informing the patrons of the bar's imminent closing and perhaps also the act of urging them to order a last drink. Whereas the upshot of these illocutionary acts is understanding on the part of the audience, perlocutionary acts are performed with the intention of producing a further effect. The bartender intends to be performing the perlocutionary acts of causing the patrons to believe that the bar is about to close and of getting them to order one last drink. He is performing all these speech acts, at all three levels, just by uttering certain words.

There seems to be a straightforward relationship in this example between the words uttered ('The bar will be closed in five minutes'), what is thereby said, and the act of informing the patrons that the bar will close in five minutes. Less direct is the connection between the utterance and the act of urging the patrons to order one last drink. Clearly there is no linguistic connection here, for the words make no mention of drinks or of ordering. This indirect connection is inferential. The patrons must infer that the bartender intends to be urging them to leave and, indeed, it seems that the reason his utterance counts as an act of that sort is that he is speaking with this intention. There is a similarly indirect connection when an utterance of 'It's getting cold in here' is made not merely as a statement about the temperature but as a request to close the window or as a proposal to go somewhere warmer. Whether it is intended (and is taken) as a request or as a proposal depends on contextual information that the speaker relies on the audience to rely on. This is true even when the connection between word and deed is more direct than in the above example, for the form of the sentence uttered may fail to determine just which sort of illocutionary act is being performed. Consider, by analogy, the fact that in shaking hands we can, depending on the circumstances, do any one of several different things: introduce ourselves, greet each other, seal a deal, or bid farewell. Similarly, a given sentence can be used in a variety of ways, so that, for example, 'I will call a lawyer' could be used as a prediction, a promise, or a warning. How one intends it determines the sort of act it is.

2 Communicative and conventional speech acts

The examples considered thus far suggest that performing a speech act, in particular an illocutionary act, is a matter of having a certain communicative intention in uttering certain words. Such an act succeeds, and the intention with which it is performed is fulfilled, if the audience recognizes that intention. Not by magic, of course - one must choose one's words in such a way that their utterance makes one's intention recognizable under the circumstances. However, as illustrated above, the utterance need not encode one's intention. So, in general, understanding an utterance is not merely a matter of decoding it.

A specifically communicative intention is a reflexive intention, of the sort characterized by H.P. Grice (1989) (see Communication and intention). This is an intention part of whose content is that it be recognized, indeed be

recognized partly on the basis that this is intended. Accordingly, it is an intention whose fulfilment consists in its recognition. This feature distinguishes acts of communication from most sorts of acts, whose success does not depend on anyone's recognizing the intention with which they are performed. One cannot succeed in running a marathon just by virtue of someone's recognizing one's intention to do so, but one can succeed in stating something, requesting something, and so on, by virtue of one's addressee recognizing that one is stating it, requesting it, or whatever. This is success at the illocutionary level. It is a further matter, a condition on the success of the perlocutionary act, whether the addressee believes what one states or does what one requests.

Now Austin did not take into account the central role of speakers' intentions and hearers' inferences. He supposed that the successful performance of an illocutionary act is a matter of convention, not intention. Indeed, he held that the use of a sentence with a certain illocutionary force is conventional in the peculiar sense that this force can be 'made explicit by the performative formula'. P.F. Strawson (1964) argues that in making this claim Austin was overly impressed by the special case of utterances that affect institutional states of affairs, and should have not taken them as a model of illocutionary acts in general. Austin was especially struck by the character of explicit performative utterances, in which one uses a verb that names the very type of act one is performing. For them he developed an account of what it takes for such acts to be performed successfully and felicitously, classifying the various things that can go wrong as 'flaws', 'hitches' and other sorts of 'infelicities'. It is only in certain conventionally designated circumstances and by people in certain positions that certain utterances can have the force they do. For example, only in certain circumstances does a jury foreman's pronouncement of 'Guilty' or 'Not guilty' count as a verdict, a legislator's 'Aye' or 'Nay' as a vote, and a baseball umpire's cry of 'Y'er out' as calling a runner out. In these cases it is only by conforming to a convention that an utterance of a certain form counts as the performance of an act of a certain sort. However, as Strawson argues, most illocutionary acts succeed not by conformity to convention but by recognition of intention. They are not conventional except in the irrelevant sense that the words and sentences being used have their linguistic meanings by virtue of convention (see Language, conventionality of).

Strawson's argument raises a serious problem for theories inspired by Austin's view. Consider, for example, the theory advanced by John Searle (1969), who proposes to explain illocutionary forces by means of 'constitutive rules' (conventions) for using 'force-indicating' devices, such as performative verbs and sentential moods. The problem is that the same sorts of illocutionary acts that can be performed by means of such devices can be performed without them. For example, one does not have to use a performative, as in 'I demand that you be quiet', or the imperative mood, as in 'Be quiet!', to demand that someone be quiet. Clearly a theory that relies on rules for using such devices is not equipped to explain the illocutionary forces of utterances lacking such devices. No such difficulty arises for a theory according to which most illocutionary acts are performed not with an intention to conform to a convention but with a communicative intention.

3 Types of speech acts

Pre-theoretically, we think of an act of communication, linguistic or otherwise, as an act of expressing oneself. This rather vague idea can be made more precise if we get more specific about what is being expressed. Take the case of an apology. If you utter '(I'm) sorry I didn't call back' and intend this as an apology, you are expressing regret for not returning a phone call. An apology just *is* the act of (verbally) expressing regret for, and thereby acknowledging, something one did that might have harmed the hearer. An apology is communicative because it is intended to be taken as expressing a certain attitude, in this case regret. It succeeds as such if it is so taken. In general, an act of communication succeeds if it is taken as intended. That is, it must be understood or, in Austin's words, 'produce uptake'. With an apology, this is a matter of the addressee recognizing the speaker's intention to be expressing regret for some deed or omission. Using a special device such as the performative 'I apologize' may of course facilitate understanding (understanding is correlative with communicating), but in general this is unnecessary. Communicative success is achieved if the speaker chooses their words in such a way that the hearer will, under the circumstances of utterance, recognize their communicative intention. So, for example, if you spill some beer on someone and say 'Oops' in the right way, your utterance will be taken as an apology.

In saying something one generally intends more than just to communicate - getting oneself understood is intended to produce some effect on the listener. However, our speech act vocabulary can obscure this fact. When one apologizes, for example, one may intend not merely to express regret but also to seek forgiveness. Seeking

forgiveness is, strictly speaking, distinct from apologizing, even though one utterance is the performance of an act of both types. As an apology, the utterance succeeds if taken as expressing regret for the deed in question: as an act of seeking forgiveness, it succeeds if forgiveness is thereby obtained. Speech acts, being perlocutionary as well as illocutionary, generally have some ulterior purpose, but they are distinguished primarily by their illocutionary type, such as asserting, requesting, promising and apologizing, which in turn is distinguished by the type of attitude expressed. The perlocutionary act is a matter of trying to get the hearer to form some correlative attitude and in some cases to act in a certain way. For example, a statement expresses a belief and normally has the further purpose of getting the addressee to form the same belief. A request expresses a desire for the addressee to do a certain thing and normally aims for the addressee to intend to and, indeed, actually do that thing. A promise expresses the speaker's firm intention to do something, together with the belief that by their utterance they are obliged to do it, and normally aims further for the addressee to expect, and to feel entitled to expect, them to do it.

Statements, requests, promises and apologies are examples of the four major categories of communicative illocutionary acts: *constatives*, *directives*, *commissives* and *acknowledgements*. This is the nomenclature used by Bach and Harnish (1979) who develop a detailed taxonomy in which each type of illocutionary act is individuated by the type of attitude expressed (in some cases there are constraints on the content as well). There is no generally accepted terminology here, and Bach and Harnish borrow the terms 'constative' and 'commissive' from Austin and 'directive' from Searle. They adopt the term 'acknowledgement', over Austin's 'behabitive' and Searle's 'expressive', for apologies, greetings, congratulations, and so on, which express an attitude regarding the hearer occasioned by some event that is thereby being acknowledged, often in satisfaction of a social expectation. Here are assorted examples of each type:

Constatives. Affirming, alleging, announcing, answering, attributing, claiming, classifying, concurring, confirming, conjecturing, denying, disagreeing, disclosing, disputing, identifying, informing, insisting, predicting, ranking, reporting, stating, stipulating.
Directives. Advising, admonishing, asking, begging, dismissing, excusing, forbidding, instructing, ordering, permitting, requesting, requiring, suggesting, urging, warning.
Commissives. Agreeing, guaranteeing, inviting, offering, promising, swearing, volunteering.
Acknowledgements. Apologizing, condoling, congratulating, greeting, thanking, accepting (acknowledging an acknowledgement).

Bach and Harnish spell out the correlation between type of illocutionary act and type of expressed attitude. In many cases, such as answering, disputing, excusing and agreeing, as well as all types of acknowledgement, the act and the attitude it expresses presuppose a specific conversational or other social circumstance.

For types of acts that are distinguished by the type of attitude expressed, there is no need to invoke the notion of convention to explain how it can succeed. The act can succeed if the hearer recognizes the attitude being expressed, such as a belief in the case of a statement and a desire in the case of a request. Any further effect it has on the hearer, such as being believed or being complied with, or just being taken as sincere, is not essential to its being a statement or a request. Thus an utterance can succeed as an act of communication even if the speaker does not possess the attitude they are expressing: communication is one thing, sincerity another. Communicating is as it were just putting an attitude on the table; sincerity is actually possessing the attitude one is expressing. Correlatively, the hearer can understand the utterance without regarding it as sincere - for example, take it as an apology, as expressing regret for something, without believing that the speaker regrets having done the deed in question. Getting one's audience to believe that one actually possesses the attitude one is expressing is not an illocutionary but a perlocutionary act.

4 Direct, indirect and nonliteral speech acts

As Austin observed, the content of a locutionary act (what is said) is not always determined by what is meant by the sentence being uttered. Ambiguous words or phrases need to be disambiguated (see Ambiguity) and the references of indexical and other context-sensitive expressions need to be fixed in order for what is said to be determined fully (see Demonstratives and Indexicals). Moreover, what is said does not determine the illocutionary act(s) being performed. We can perform a speech act (1) directly or indirectly, by way of performing another

speech act, (2) literally or nonliterally, depending on how we are using our words, and (3) explicitly or inexplicitly, depending on whether we fully spell out what we mean.

These three contrasts are distinct and should not be confused. The first two concern the relation between the utterance and the speech act(s) thereby performed. In indirection a single utterance is the performance of one illocutionary act by way of performing another. For example, we can make a request or give permission by way of making a statement, say by uttering 'I am getting thirsty' or 'It doesn't matter to me', and we can make a statement or give an order by way of asking a question, such as 'Will the sun rise tomorrow?' or 'Can you clean up your room?' When an illocutionary act is performed indirectly, it is performed by way of performing some other one directly. In the case of nonliteral utterances, we do not mean what our words mean but something else instead. With nonliterality the illocutionary act we are performing is not the one that would be predicted just from the meanings of the words being used, as with likely utterances of 'My mind got derailed' or 'You can stick that in your ear'. Occasionally utterances are both nonliteral and indirect. For example, you might utter 'I love the sound of your voice' to tell someone nonliterally (ironically) that you cannot stand the sound of their voice and thereby indirectly to ask them to stop singing.

Nonliterality and indirection are the two main ways in which the semantic content of a sentence can fail to determine the full force and content of the illocutionary act being performed in using the sentence. They rely on the same sorts of processes that Grice discovered in connection with what he called 'conversational implicature' (see Implicature), which, as is clear from Grice's examples, is nothing more than the special case of nonliteral or indirect constatives made with the use of indicative sentences. A few of Grice's examples illustrate nonliterality, such as 'He was a little intoxicated', used to explain why a man smashed some furniture, but most of them are indirect statements, such as 'There is a garage around the corner', used to tell someone where to get petrol, and 'Mr X's command of English is excellent, and his attendance has been regular', giving the high points in a letter of recommendation. These are all examples in which what is meant is not determined by what is said. However, Grice overlooks a different kind of case, marked by contrast (3) listed above.

There are many sentences whose standard uses are not strictly determined by their meanings but are not implicatures or figurative uses either. For example, if one's spouse says 'I will be home later', they are likely to mean that they will be home later that night, not merely some time in the future. In such cases what one means is an *expansion* of what one says, in that adding more words ('tonight', in the example) would have made what was meant fully explicit. In other cases, such as 'Jack is ready' and 'Jill is late', the sentence does not express a complete proposition. There must be something which Jack is being claimed to be ready for and something which Jill is being claimed to be late for. In these cases what one means is a *completion* of what one says. In neither sort of case is a particular word or phrase being used nonliterally and there is no indirection. They both exemplify what may be called 'impliciture', since part of what is meant is communicated not explicitly but implicitly, by way of expansion or completion.

5 Philosophical importance of speech act theory

The theory of speech acts has applications to philosophy in general, but these can only be illustrated here. In ethics, for example, it has been supposed that sentences containing words like 'good' and 'right' are used not to describe but to commend, hence that such sentences are not used to make statements and that questions of value and morals are not matters of fact. This line of argument is fallacious. Sentences used for ethical evaluation, such as 'Loyalty is good' and 'Abortion is wrong', are no different in form from other indicative sentences. Whatever the status of their contents, they are standardly used to make statements. This leaves open the possibility that there is something fundamentally problematic about their contents. Perhaps such statements are factually defective and, despite syntactic appearances, are neither true nor false. However, this is a metaphysical issue about the status of the properties to which ethical predicates purport to refer. It is not the business of the philosophy of language to determine whether or not there are such properties as goodness or rightness and whether or not the goodness of loyalty and the rightness of abortion are matters of fact. The above argument is but one illustration of what Searle calls the 'speech act fallacy'. He also identifies examples of the 'assertion fallacy', whereby conditions of making an assertion are confused with what is asserted. For example, one might fallaciously argue, on the grounds that because one would not assert that one believes something if one was prepared to assert that one knows it, that knowing does not entail believing. Grice identifies the same fallacy in a parallel argument, according to which

seeming to have a certain feature entails not actually having that feature (see Ordinary language philosophy).

For philosophy of language in particular, the theory of speech acts underscores the importance of the distinction between language use and linguistic meaning (see Pragmatics; Semantics). This distinction sharpens the formulation of questions about the nature of linguistic knowledge, by separating questions about capacities exercised in linguistic interaction from those specific to knowledge of language itself. A parallel distinction, between speaker reference and linguistic reference (see Reference), provokes the question of to what extent linguistic expressions refer independently of speakers' use of them to refer. It is common, for example, for philosophers to describe expressions like 'the car', 'Robert Jones' and 'they' as having different references in different contexts, but it is arguable that this is merely a misleading way of saying that speakers use such expressions to refer to different things in different contexts.

See also: Grice, H.P.; Language, philosophy of; Pragmatics; Semantics

KENT BACH

References and further reading

Austin, J.L. (1962) *How to Do Things with Words*, Cambridge, MA: Harvard University Press.(Develops the distinction between performative and constative utterances into the first systematic account of speech acts.)

Bach, K. (1994) 'Conversational Implicature', *Mind and Language* 9: 124-62.(Identifies the middle ground between explicit utterances and Gricean implicatures.)

Bach, K. and Harnish, R.M. (1979) *Linguistic Communication and Speech Acts*, Cambridge, MA: MIT Press.(Combines elements of Austin's taxonomy and Grice's theory of conversation into a systematic account of the roles of the speaker's communicative intention and the hearer's inference in literal, nonliteral and indirect uses of sentences to perform speech acts.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.(The essays on meaning and conversational implicature provide a framework for distinguishing speaker meaning from linguistic meaning and for explaining their relationship.)

Searle, J.R. (1969) *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.(Presents a theory of speech acts relying on the notion of constitutive rules.)

Strawson, P.F. (1964) 'Intention and Convention in Speech Acts', *Philosophical Review* 73: 439-60.(Applies Grice's account of meaning to support the claim that most speech acts are communicative rather than conventional, as Austin had suggested.)

Tsohatzidis, S.L. (ed.) (1994) *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, London: Routledge.(Collection of original essays on outstanding problems in the field, with useful bibliography.)

Spencer, Herbert (1820-1903)

Herbert Spencer is chiefly remembered for his classical liberalism and his evolutionary theory. His fame was considerable during the mid- to late-nineteenth century, especially in the USA, which he visited in 1882 to be lionized by New York society as the prophetic philosopher of capitalism. In Britain, however, Spencer's reputation suffered two fatal blows towards the end of his life. First, collectivist legislation was introduced to protect citizens from the ravages of the industrial revolution, andSpencer's spirited defence of economic laissez-faire *became discredited. Second, his evolutionary theory, which was based largely on the Lamarckian principle of the inheritance of organic modifications produced by use and disuse, was superseded by Darwin's theory of natural selection. Nearly a century after his death, however, there is renewed interest in his ideas, partly because the world has become more sympathetic to market philosophies, and partly because the application of evolutionary principles to human society has become fashionable once more.*

Spencer was born into a Nonconformist family in Derby, England, and never lost the individualistic temperament conferred by his upbringing. Trained as a civil engineer working for a railway company during the boom years of railway expansion, Spencer's fertile mind soon spread to social and economic issues, and he joined the *Economist* as a sub-editor in 1848. In 1851 the publication of his first book, *Social Statics*, established his reputation as a thinker of extraordinary power and originality. In it he deduced the features of a civilized society from the central principle of justice - the law of equal freedom - that 'every person has freedom to do all that he wills, provided he infringes not the equal freedom of any other man' (1851: 103). This principle of equal freedom, which has been regarded ever since as the cornerstone of classical liberalism, led Spencer to insist on a very narrow role for the state (see Liberalism; State, the §§1-2). The state was at best a necessary evil to prevent one person from violating the rights of another - that is, to defend natural rights. *Social Statics* is the definitive text of the minimal or 'nightwatchman' theory of the state.

Spencer subsequently embarked upon an ambitious task of creating a wholly new understanding of philosophy, at the heart of which was the theory of evolution. In *Social Statics* an idea of the evolutionary progress of humanity had been implicitly assumed but not developed.Spencer explicitly applied evolutionary theory in his essay 'The Development Hypothesis' ([1890] 1852). He rejected the religious notion of 'special creation' - that each of the 10 million species had been individually created by God - arguing instead, like Lamarck, that new species emerged as a result of modifications in existing species, brought about by exposure to new conditions. Spencer elaborated this evolutionary theory in his essay 'Progress: its Law and Cause' ([1890] 1857) where he introduced the word 'evolution'. FollowingVon Baer, Spencer claimed that progress was defined in terms of a change from the homogenous to the heterogeneous. His originality lay in the fact that he saw this 'law' at work across the whole range of natural and human phenomena. The ten volumes ofSpencer's *magnum opus*, the 'Synthetic Philosophy', traced the operation of evolutionary principles successively in psychology, metaphysics, biology, sociology and ethics.

In his *Principles of Psychology* (1855), Spencer put forward a Lamarckian explanation of mental development. For example, intelligence was a faculty developed as a result of cumulative modifications of the mind in successive generations of organisms responding to their environment. In *First Principles* (1862), he addressed the vexed question for his Victorian readership of the relation between science and religion. ForSpencer, science entailed the deduction of general laws, such as the conservation of energy, which were not empirical generalizations, but necessary truths about empirical phenomena. But these truths could never be completely grasped. We could never know, for example, *why* energy was conserved. Such ultimate questions about the nature of reality formed what Spencer called 'the Unknowable'. Religion yielded similarly unanswerable questions about fundamental issues - such as 'does God exist, and if so, how did He come into existence?' These issues were also part of the Unknowable; Spencer was an agnostic, not an atheist - he rejected anti-evolutionary religious doctrines, but God's existence was not incompatible with his theory of evolution.

In his *Principles of Biology* (1864, 1867) Spencer acknowledged the centrality of Darwinian natural selection, but insisted that Lamarckian modifications played their part in the evolution of organisms from simple to complex structures. In *Principles of Sociology* (1876-96) he explained how, as humanity advanced, a process of

differentiation of functions occurred, and society was gradually transformed from the 'militant' type, which was characterized by authoritarianism, uniformity and status, to the 'industrial' type, which was characterized by liberty, diversity and contract. He made use of the organic analogy (likening society to an organism) to sustain his *laissez-faire* theory, by interpreting organisms as made up of individualistic parts. Finally, in *Principles of Ethics* (1879-93) he gave the law of equal freedom an evolutionary dimension by linking it to the principle that each person ought to experience the full consequences of their actions, both good and bad - a principle which entailed that the fittest survived.

Spencer has left two enduring contributions to philosophy. The first is a highly cogent analysis of the philosophical foundations of classical liberalism. *Social Statics* is a masterpiece of argumentation, not far short of the stature of a Hobbesian or Lockeian text. The second is the evolutionary idea: it was Spencer, not Darwin, who was the founder of the philosophy later known as Social Darwinism, and who coined the term 'the survival of the fittest'. Although Spencer failed to answer many questions raised by his theory of evolution, it cannot be denied that his evolutionary vision marks an important stage in our understanding of social development.

See also: Darwin, C.R.; Evolution, theory of; Evolutionary theory and social science; Freedom and liberty; Holism and individualism in history and social science; Religion and science; Social science, history of philosophy of §10

TIM S. GRAY

List of works

Spencer, H. (1851) *Social Statics: or the Conditions Essential to Human Happiness and the First of them Developed*, London: Chapman.(Spencer's best work on political philosophy; hard but rewarding.)

Spencer, H. (1855) *Principles of Psychology*, London: Williams & Norgate; 4th edn, 1899. (A comprehensive account of mind, intelligence, the nervous system, feelings, perceptions, memory, reasoning, sympathy and realism; very dense.)

Spencer, H. (1861) *Education, Intellectual, Moral and Physical*, London: Everyman's Library, 1961.(Four very readable, instructive and popular essays extolling the value of natural methods in education.)

Spencer, H. (1862) *First Principles*, London: Williams & Norgate; 6th edn, 1900.(A work in two parts: 'The Unknowable', which shows how all science led ultimately (like all religion) to a belief in an Absolute that transcended human understanding; and 'Laws of the Knowable' - a statement of the fundamental scientific principles governing the world.)

Spencer, H. (1864, 1867) *Principles of Biology*, London: Williams & Norgate, 2 vols; 2nd edn, 1898, 1899.(A comprehensive and highly technical account of morphology, physiology, reproduction, growth, development, heredity, variation and evolution.)

Spencer, H. (1873) *The Study of Sociology*, Ann Arbor, MI: Ann Arbor Paperback, 1961.(A pioneering work of popularization for the discipline of sociology, providing an easy-to-read explanation of the nature and importance of the sociological method; still a standard text.)

Spencer, H. (1876-96) *Principles of Sociology*, London: Williams & Norgate, vol. 1, 1876; 3rd edn, 1885; vol. 2, part 4, 1879; part 5, 1882; vol. 3, part 6, 1885; parts 7 and 8, 1896.(At 2,300 pages, a massive but very readable study of anthropology, social structure, the family and ceremonial, political, religious, professional and industrial institutions.)

Spencer, H. (1879-93) *Principles of Ethics*, London: Williams & Norgate, vol. 1, part 1, 1879; parts 2 and 3, 1892; vol. 2, part 4, 1891; parts 5 and 6, 1893.(An evolutionary approach to moral philosophy, in which justice and beneficence are dissected for their social value; readable but much less impressive than *Social Statics*.)

Spencer, H. (1884) *The Man Versus The State*, repr. with abridged edn of *Social Statics*, London: Williams & Norgate, 1892.(Four highly polemical essays in which Spencer denounces the increasing collectivism of governmental policy in late nineteenth-century Britain; entertaining and thought-provoking.)

Spencer, H. (1890) *Essays: Scientific, Political and Speculative*, London: Williams & Norgate, 3 vols, revised edition.(A collection of highly accessible articles published by Spencer in leading Victorian periodicals, ranging over science, philosophy, aesthetics, ethics, psychology and politics. Contains 'The Development Hypothesis' (1852) and 'Progress: its Law and Cause' (1857) in vol. 1.)

References and further reading

Gray, J.N. (1982) 'Spencer on the Ethics of Liberty and the Limits of State Interference', *History of Political Thought* 3 (3): 456-81.(Gray argues that Spencer was an 'indirect utilitarian', justifying the law of equal freedom on grounds that it would lead to happiness.)

Gray, T.S. (1988) ‘Is Herbert Spencer’s Law of Equal Freedom a Utilitarian or a Rights-Based Theory of Justice?’, *Journal of the History of Philosophy* 26 (2): 259-78.(Argues that Spencer’s law of equal freedom drew its authority neither from utility nor from rights, but from the moral sense.)

Gray, T.S. (1996) *Herbert Spencer’s Political Philosophy: Individualism and Organicism*, Hampshire: Avebury. (A comprehensive rebuttal of the repeated charge that Spencer was contradictory in yoking together political individualism and social organicism.)

Hodgson, G. (1993) *Economics and Evolution*, Cambridge: Polity Press.(An authoritative book containing an impressive chapter on Spencer.)

Peel, J.D.Y. (1971) *Herbert Spencer, The Evolution of a Sociologist*, London: Heinemann.(This scholarly book is by far the best critical study of Spencer’s ideas. Peel sets these ideas in their historical and intellectual context, and shows their enduring influence.)

Steiner, H. (1982) ‘Land, Liberty and the Early Herbert Spencer’, *History of Political Thought* 3 (3): 515-33. (Steiner argues thatSpencer presents his law of equal freedom as a basic moral axiom - an intrinsic principle of right.)

Taylor, M. (1992) *Men Versus the State - Herbert Spencer and Late Victorian Individualism*, Oxford: Clarendon Press.(A major study of Spencer in relation to other individualists in Britain at the end of the nineteenth century.)

Speusippus (*c*.410-339 BC)

The Greek philosopher Speusippus was the second head of the Platonic Academy. Succeeding his uncle Plato on the latter's death, he developed his thought in interesting directions. He pursued further the tendency in the Academy to mathematicize reality that so annoyed Aristotle, postulating a complicated metaphysics, which started from a 'One' superior to being and all other qualities, and a material principle, 'multiplicity', from the union of which arose, first number, then geometrical entities, and then soul and the material world. In his hands, Plato's doctrines of first principles, of Forms, and of the union of Forms with matter, suffered transformations of which we have only imperfect reports. His later influence was greater on Neo-Pythagoreanism than on 'orthodox' Platonism, until Plotinus.

1 Life and works

Speusippus was the son of Plato's sister. He became the second head of the ACADEMY upon his uncle's death in 347 BC, and continued as head till his own death, when he was himself succeeded by Xenocrates. He may have gained the headship for family reasons, but he was a considerable philosopher in his own right, even if his original contributions have been much obscured by the loss of all his writings, and by the hostile criticisms of Aristotle. Other details of his life are hard to come by, since the surviving biography by Diogenes Laertius is most unreliable. Diogenes gives a list of thirty works by Speusippus. Most important, perhaps, were those *On Pleasure*, *On Philosophy*, *On the Gods*, *On the Soul*, and *On Pythagorean Numbers* and a treatise *On Congeners* (*homoia*) in ten volumes, which seems to have exercised the technique of division (*diaeresis*) on most of the species of the natural world. Speusippus apparently also pioneered the linking of Plato's philosophy with that of Pythagoras, which gave an impetus to the later 'Pythagoreanizing' of Platonism (see Neo-Pythagoreanism; Platonism, Early and Middle §1).

2 Metaphysics

Speusippus' concern was to carry on the work of Plato, but he developed Platonic doctrine in various interesting ways, which, although legitimate, seem to have found no other partisans in official Platonist circles before Plotinus (although certain Neo-Pythagorean speculations may perhaps be traced to his inspiration). Speusippus accepted the doctrine of two opposite principles, but modified it by laying particular emphasis on their status as 'seeds' or 'potencies' of all things. He argued that what is itself the cause of some quality in other things cannot have that quality itself in the same way, so that if the One is the cause of goodness and being for all other things, it cannot itself properly be termed good or even existent (fr. 42). The One is not even properly a creative principle - the Indefinite Dyad, or 'multiplicity', being credited with that role, as being the cause of all differentiation and individuation. The latter in turn should not be taken as 'evil', any more than the One is 'good' (fr. 44). Good and evil only arise at the level of actualized being, and perhaps only at the level of soul.

Aristotle condemns Speusippus for propounding an 'episodic' universe, with different first principles for every level of being (fr. 30). This is a satirical misrepresentation of the complexity of his position, however, which was based on the argument that a simple and unitary first principle working on an absolutely formless substratum could only produce one thing (for example, numbers), and that this could not account for the multifariousness of the actual cosmos. We must therefore postulate a process by which the product of a higher or more basic union becomes, somehow, the first principle of the next level of being, by uniting, once again, with the material principle.

This may all be less than convincing, but it is much less foolish than unquestioning reliance on Aristotle's testimony would make out. It depends, unfortunately, on accepting as evidence a passage from a work of Iamblichus, *Common Mathematical Theory* (ch. 4), which, although probably, is not indisputably Speusippan. Even Aristotle, however, if his evidence is properly understood, testifies to Speusippus' positing a first principle which is 'not yet existent' (fr. 43).

Exactly how Speusippus modified the Platonic theory of Forms is not clear, but Aristotle claims that he abandoned them in favour of numbers or 'mathematicals' (frs 31, 34). He is probably here only developing the tendency of the late Plato to mathematicize the Forms, while tidying up the confusion between the non-addible 'forms' of

numbers, and ’mathematicals’ or numbers in the ususal sense (see Platonism, Early and Middle §3). The Forms proper, as mathematical (or perhaps rather geometrical) entities, are closely allied with the soul, defined as ’the [F]orm of the omnidimensionally extended’ (fr. 54). The soul receives both number and geometrical extension from above, and these, synthesized, become its formal principle, which then combines once again with the material principle to form the physical world.

3 Ethics and logic

In ethics we have various reports of Speusippus’ views. He defined happiness (*eudaimonia*) as ‘the state of perfection in things natural’ (fr. 77), which seems to prefigure the Stoic principle of ‘living in accordance with nature’ (see Stoicism §17). The ideal to be aimed at is ’freedom from disturbance’ (*aochlēsia*) - an anticipation of the Epicurean ideal (see Epicureanism §10). This fits his attested belief that pleasure and pain are alike evils (fr. 80), the ideal being a state median between them.

In the sphere of logic, he was, as noted above, primarily an adherent of Platonic division (*diairesis*), but he developed certain distinctive theories. He held that in order to define anything adequately one must know all the differentiae of that thing in respect of everything with which it is not identical. This would logically involve knowledge of all the relations of the thing in question, however remote, with everything else in the world (fr. 63). This did not lead him to scepticism, however, but rather to a mighty effort to classify, by means of *diairesis*, as much of reality as he could, the results of which are attested in the remains of his *On Congeners*. Apart from this, his distinction between homonyms and synonyms (fr. 68) seems to have provoked Aristotle to his slightly different classification; and in the field of epistemology he seems to anticipate Aristotle in holding that first principles must be apprehended immediately if there is to be any knowledge at all (fr. 73) (see Aristotle §§4, 6).

Finally, mention may be made of Speusippus’ work *On Pythagorean Numbers*, a report of whose contents and a longish extract (on the Decad) are preserved in the pseudo-Iamblichean *Theology of Arithmetic* (fr. 28). In the first half of this work Speusippus, depending heavily on the writings of the Pythagorean Philolaus, discussed the ‘linear, polygonal, plane and solid numbers’, and also the five cosmic figures (of Plato’s *Timaeus*). The second half of the work he devoted entirely to the Decad, declaring it to be the most ’natural’ and perfect of numbers, and a model for the creator god in his creation of the world. Speusippus would seem thus to identify the ‘paradigm’ of the *Timaeus* with the Decad, while the creator or ‘demiurge’ is probably to be identified with the first principle of number. One can see here Speusippus’ ’Pythagoreanism’ coming out very strongly, as well as that mathematicization of reality to which Aristotle so strongly objected.

JOHN DILLON

References and further reading

Cherniss, H. (1945) *The Riddle of the Early Academy*, Berkeley and Los Angeles, CA: University of California Press.(A useful discussion of Form-numbers and other mysteries can be found in chapter 2, although from a somewhat perverse perspective.)

Dancy, R.M. (1989) ’Ancient Non-Beings: Speusippus and Others’, *Ancient Philosophy* 9: 207-44.(A good discussion of Speusippus’ metaphysics.)

Dillon, J. (1977) *The Middle Platonists*, London: Duckworth.(Introductory account of Speusippus is given at pages 11-22.)

Dillon, J. (1984) ’Speusippus in Iamblichus’, *Phronesis* 29: 325-32.(Critique of Tarán’s *Speusippus of Athens*.)

Iamblichus (early 4th century AD) *Common Mathematical Theory*, ed. N. Festa, *Iamblichi de communi mathematica scientia*, Leipzig: Teubner, 1891; repr. Stuttgart, 1975.(Greek text.)

Merlan, P. (1960) *From Platonism to Neoplatonism*, The Hague: Martinus Nijhoff, 2nd edn.(Speusippus can be found at pages 96-140.)

Speusippus (*c*.410-339 BC) Fragments, in L. Tarán, *Speusippus of Athens*, Leiden: Brill, 1981.(Collection of fragments, with copious introduction, commentary and bibliography, but somewhat too credulous of Aristotle. Includes *On Congeners*.)

Spinoza, Benedict de (1632-77)

A Dutch philosopher of Jewish origin, Spinoza was born Baruch de Spinoza in Amsterdam. Initially given a traditional Talmudic education, he was encouraged by some of his teachers to study secular subjects as well, including Latin and modern philosophy. Perhaps as a result of this study, he abandoned Jewish practices and beliefs and, after receiving stern warnings, he was excommunicated from the synagogue in 1656. Alone and without means of support, he Latinized his name and took up the trade of lens grinder with the intention of devoting his life to philosophy. He remained in Amsterdam until 1660, lived for the next decade in nearby villages, and in The Hague from 1670 until his death from consumption in 1677. During these years he worked continuously on his philosophy and discussed it with a small circle of friends and correspondents. His masterpiece, Ethica Ordine Geometrico Demonstrata (Ethics Demonstrated in a Geometrical Manner), was completed in 1675; but because of its radical doctrines, it was only published after his death.

The full scope of Spinoza's Ethics is not indicated by its title. It begins with a highly abstract account of the nature of substance, which is identified with God, and culminates in an analysis of human beings, their nature and place in the universe, and the conditions of their true happiness. Written in a geometrical form modelled after Euclid, each of its five parts contains a set of definitions, axioms and propositions which are followed by their demonstrations and frequently by explanatory scholia.

The defining feature of Spinoza's thought is its uncompromising rationalism. Like other philosophers of the time, Spinoza is a rationalist in at least three distinct senses: metaphysical, epistemological and ethical. That is to say, he maintains that the universe embodies a necessary rational order; that, in principle, this order is knowable by the human mind; and that the true good for human beings consists in the knowledge of this order and a life governed by this knowledge. What is distinctive of Spinoza's brand of rationalism, however, is that it allows no place for an inscrutable creator-God distinct from his creation, who acts according to hidden purposes. Instead, Spinoza boldly identifies God with nature, albeit with nature regarded as this necessary rational order rather than as the sum-total of particular things.

In its identification of God with nature, Spinoza's philosophy is also thoroughly naturalistic and deterministic. Since nature (as infinite and eternal) is all-inclusive and all-powerful, it follows that nothing can be or even be conceived apart from it: this means that everything, including human actions and emotions, must be explicable in terms of nature's universal and necessary laws. Moreover, given this identification, it also follows that knowledge of the order of nature specified through these laws is equivalent to the knowledge of God. Thus, in sharp opposition to the entire Judaeo-Christian tradition, Spinoza claims that the human mind is capable of adequate knowledge of God.

The attainment of such knowledge is, however, dependent on the use of the correct method. In agreement with Descartes and Thomas Hobbes (the two modern philosophers who exerted the greatest influence on his thought) and thoroughly in the spirit of the scientific revolution, Spinoza held that the key to this method lies in mathematics. This conviction is obviously reflected in the geometrical form of the Ethics; but it actually runs much deeper, determining what for Spinoza counts as genuine knowledge as opposed to spurious belief. More precisely, it means that an adequate understanding of anything consists in seeing it as the logical consequence of its cause, just as the properties of a geometrical figure are understood by seeing them as the logical consequence of its definition. This, in turn, leads directly to the complete rejection of final causes, that is, the idea that things in nature (or nature as a whole) serve or have an end, and that understanding them involves understanding their end. Not only did Spinoza reject final causes as unscientific, a view which he shared with most proponents of the new science, he also regarded it as the source of superstition and a major obstacle to the attainment of genuine knowledge.

The same spirit underlies Spinoza's practical philosophy, which is marked by his clinical, dispassionate analysis of human nature and behaviour. In contrast to traditional moralists (both religious and secular), he rejects any appeal to a set of absolute values that are independent of human desire. Since the basic desire of every being is self-preservation, virtue is identified with the capacity to preserve one's being, the good with what is truly useful in this regard and the bad with what is truly harmful. In the case of human beings, however, what is truly useful is knowledge; so virtue consists essentially in knowledge. This is because knowledge is both the major weapon

against the passions (which are the chief sources of human misery) and, in so far as it is directed to God or the necessary order of nature, the source of the highest satisfaction.

Apart from the Ethics, Spinoza is best known for his contributions to the development of an historical approach to the Bible and to liberal political theory. The former is contained in the Tractatus Theologico-politicus (Theological-Political Treatise), which he published anonymously in 1670 as a plea for religious toleration and freedom of thought. The latter is contained both in that work and in the unfinished Tractatus Politicus (Political Treatise) of 1677, in which Spinoza attempts to extend his scientific approach to questions in political philosophy.

1 The geometrical method

Although Spinoza uses the geometrical method in the *Ethica Ordine Geometrico Demonstrata (Ethics Demonstrated in a Geometrical Manner)* (1677a), he does not attempt to justify or even explain it. This has led many readers to view its argument as an intricate and fascinating chain of reasoning from arbitrary premises, which, as such, never touches reality. Nevertheless, Spinoza was very much aware of this problem and dealt with it both in an important early and unfinished work on method, *Tractatus de Intellectus Emendatione (Treatise on the Emendation of the Intellect)* (1677b) and in some of his correspondence.

At the heart of the problem is the nature of the definitions (and to a lesser extent the axioms) on the basis of which Spinoza attempts to demonstrate the propositions of the *Ethics*. In a 1663 letter to a young friend, Simon De Vries, who queried him about this very problem, Spinoza offers his version of the traditional distinction between nominal and real definitions. The former kind stipulates what is meant by a word or thought in a concept. Such a definition can be conceivable or inconceivable, clear or obscure; but, as arbitrarily invented, it cannot, strictly speaking, be either true or false. By contrast, the real definition, which supposedly 'explains a thing as it exists outside of the understanding', defines a thing rather than a term (Leibniz 1966 letter 9: 106-7). Consequently, it can be either true or false.

Since the definitions of the *Ethics* are typically introduced by expressions such as 'by… I mean that', or 'a thing is called', it would seem that they are of the nominal type, which gives rise to the charge of arbitrariness. It is clear from their use, however, that Spinoza regards them as real definitions. Like the definitions of geometrical figures in Euclid, they are intended to express not merely the names used, but the objects named.

The question, then, is how can one know that one has arrived at a true definition. Spinoza's answer reveals the depth of his commitment to the geometrical way of thinking, especially to the method of analytic geometry developed by Descartes. He appeals to the example of the mathematician, who knows that one has a real definition of a figure when one is able to construct it. The definition of a figure is thus a rule for its construction, what is usually called a 'genetic definition'. Spinoza develops this point in *On the Emendation of the Intellect* ([1677b] 1985 vol.1: 39-40) by contrasting the nominal definition of a circle as 'a figure in which the lines drawn from the centre to the circumference are equal' with the genetic definition as 'the figure that is described by any line of which one end is fixed and the other movable'. The point is that the latter definition, but not the former, tells us how such a figure can be constructed and from this rule of construction we can deduce all its essential properties.

Spinoza's claim, then, is that the principles that apply to mathematical objects and perhaps other abstract entities also apply to reality as such. Thus, we have a real definition, an adequate, true or clear and distinct idea of a thing (all of these terms being more or less interchangeable) in so far as we know its 'proximate cause' and can see how its properties necessarily follow from this cause.

But if knowledge of a thing reduces to knowledge of its proximate cause, then either we find ourselves involved in an infinite regress, which would lead to a hopeless scepticism, or the chain of reasoning must be grounded in a single first principle. Furthermore, this first principle must have a unique status: if it is to provide the ultimate ground in terms of which everything else is to be explained, it must somehow be self-grounded or have the reason for its existence in itself. In the scholastic terminology which Spinoza adopts, it must be *causa sui* (self-caused). Thus, Spinoza's rationalist method leads necessarily to the concept of God, which he defines as 'a being absolutely infinite, i.e., a substance consisting of an infinity of attributes, of which each one expresses an eternal and infinite essence' (*Ethics*: I, def. 6).

Given this concept, together with other essential concepts such as substance, attribute and mode (which are treated

in the definitions of part one), the argument of the *Ethics* proceeds in a deductive manner. Its goal is to enable us to understand reality as a whole in light of this concept in just the same way that the mathematician can understand all the essential properties of a geometrical figure in terms of its concept or genetic definition. At least this is the project of part one of the *Ethics* (On God). The remainder of the work is devoted to the demonstration of the most important consequences of this result in so far as they concern the human condition.

2 Substance-monism

The first fourteen propositions of the *Ethics* contain an argument intended to show that ‘Except God, no substance can be or be conceived’ (*Ethics*: I, prop. 14). Since it follows from an analysis of the concept of substance that whatever is not itself substance must be a modification thereof, Spinoza concludes in the next proposition that ‘Whatever is, is in God, and nothing can be or be conceived without God’. Together they express Spinoza’s substance-monism, which can be defined as the complex thesis that there is only one substance in the universe; that this substance is to be identified with God; and that all things, as modes of this one substance are, in some sense ’in God’.

The argument for this thesis is based largely on the analysis of the concept of substance, defined as ‘what is in itself and is conceived through itself, i.e., that whose concept does not require the concept of another thing, from which it must be formed’ (*Ethics*: I, def. 3). This definition is quite close to Descartes who likewise made a capacity for independent existence the criterion of substance; but the two philosophers drew diametrically opposed conclusions from their similar definitions. Although Descartes held that God is substance in a pre-eminent sense, he also maintained that there are two kinds of created substance - thinking things or minds, and extended things, both of which depend for their existence on God but not on each other. For Spinoza, by contrast, there is only the one substance - God - and thought and extension are among its attributes.

Spinoza argues indirectly for his monism by criticizing the two major alternatives that a substance-based metaphysics can provide: that there is a plurality of substances of the same nature or attribute and that there is a plurality of substances with different natures or attributes. Since Descartes is committed to both forms of pluralism, both parts of the argument cut against his views. In considering Spinoza’s position, however, it should be kept in mind that his target is not merely Descartes, but an entire philosophical and theological tradition which conceived of the universe as composed of a number of finite substances created by God (see Medieval philosophy). In spite of his radical critique of scholastic ways of thinking and his appeal to mathematics as the ideal of knowledge, Descartes remained in many ways a part of that tradition.

The argument against the first form of pluralism turns on the claim that ‘In nature there cannot be two or more substances of the same nature or kind’ (*Ethics*: I, prop. 5). This is at once one of the most important and controversial propositions in the *Ethics*. It is important because of its pivotal role in the overall argument for monism, controversial because of the demonstration offered in its support which is based on a consideration of the grounds on which two or more substances might be distinguished. This could be done either on the basis of their attributes, if they are substances of different types (for example, Descartes’ thinking and extended substances), or on the basis of their modifications, if they are distinct substances of the same kind (for example, particular minds or bodies). The claim is that neither procedure can distinguish two or more substances of the same kind; and since these are the only possible ways of distinguishing substances, it follows that such substances could not be distinguished from one another.

The unsuitability of the first alternative seems obvious and would be recognized as such by Descartes. Since a substance-type for Descartes is defined in terms of its attribute, it follows that two or more substances of the same kind could not be distinguished on the basis of their attributes. As obvious as this seems, however, it was criticized by Leibniz on the grounds that two substances might have some attributes in common and others that were distinct. For example, substance *A* might have attributes *x* and *y*, and substance *B* attributes *y* and *z*. Although a Cartesian would reject this analysis on the grounds that a substance cannot have more than one attribute, Spinoza (for whom God is a substance with infinite attributes) could hardly accept this Cartesian principle. Moreover, we shall see that it is essential to the overall argument for monism to eliminate the possibility suggested by Leibniz (see Leibniz, G.W. §§4-7).

Various strategies for dealing with this problem have been suggested in the literature, perhaps the most plausible

of which turns on the principle that if two or more substances were to share a single attribute, they would have to share all, and would, therefore, be numerically identical. Although Spinoza never argues explicitly in this way, it seems a reasonable inference from his conception of attribute as 'what the intellect perceives of a substance as constituting its essence' (*Ethics*: I, def. 4). This entails that attribute *y* of substance *A* is identical to attribute *y* of substance *B* just in case they express the same nature or essence - that is, are descriptions of the same kind of thing. But if they are things of the same kind, then any attribute *A* has will be possessed by *B* as well.

Against the second alternative Spinoza argues that, since by definition substance is prior to its modifications, if we consider substance as it is in itself then we cannot distinguish one substance from another. Although the claim is surely correct, the suggestion that we set the modifications aside seems to beg the question against Descartes. After all, two Cartesian thinking substances share the same nature or attribute and are distinguished precisely by their different modifications (thoughts).

Spinoza's reasoning at this point is unclear, but one could respond that, on the hypothesis under consideration, the substances must be assumed to be indistinguishable prior to the assignment of modifications. Moreover, it follows from this that the assignment of modifications could not serve to distinguish otherwise indistinguishable substances unless it is already assumed that they are numerically distinct. In other words, while we can distinguish two Cartesian substances by means of their modifications, we can only do so by presupposing that the distinct modifications belong to numerically distinct substances. But it is just this assumption to which the Cartesian is not entitled.

3 Substance-monism (cont.)

This, however, is only the first step in the argument for monism. It is also necessary to rule out the possibility of a plurality of substances of different kinds. Essential to this project is the demonstration that substance is infinite not merely 'in its own kind', that is, unlimited by anything of the same kind, but 'absolutely infinite', that is, all-inclusive or possessing all reality, which, for Spinoza, means infinite attributes. For example, Descartes' extended substance is infinite in the first sense because it is not limited or determined by anything outside itself (for example, empty space); but it is not infinite in the second and decisive sense because it does not constitute all reality. Moreover, this is precisely why the first sense of infinity is not sufficient to preclude a plurality of substances.

The basis of the argument for the absolute infinity of substance is the claim that 'The more reality or being each thing has the more attributes belong to it' (*Ethics*: I, prop. 9). This is a direct challenge to the Cartesian conception of substance as defined in terms of a single attribute, resting on the dual assumption that some things can possess more reality than others and that this superior degree of reality is manifested in a greater number of attributes. Unfortunately, in defence of this claim Spinoza merely refers to the definition of attribute as 'what the intellect perceives of a substance as constituting its essence' (*Ethics*: I, def. 4). Nevertheless, it does seem possible to understand Spinoza's point if we interpret attributes as something like distinct descriptions under which substance or reality can be taken. Consider, for example, a simple human action such as raising an arm. Although it may be possible to give a complete neurophysiological account of such an action in terms of impulses sent to the brain, the contraction of muscles and the like, one could still argue that no such account, no matter how detailed, is adequate to understanding it as an action. This requires reference to psychological factors such as the beliefs and intentions of an agent, which in Spinoza's metaphysics belong to the attribute of thought. Thus, we might say both that there is 'more reality' to an action than is given in a purely neurophysiological account and that this 'greater reality' can be understood as the possession of a greater number of attributes.

It follows from this that a being possessing all reality - that is, God, or the *ens realissimum* (most real being) of the tradition - may be described as possessing infinite attributes. It also follows that the Cartesian must either accept the possibility of a substance with infinite attributes or deny the possibility of God. And since the orthodox Cartesian could hardly do the latter, the former must be admitted.

Even granting this, however, at least two problems remain. One is how to understand the infinity of attributes. This might mean either that substance possesses infinitely many attributes, of which the human mind knows only two (thought and extension), or that it possesses all possible attributes, which is compatible with there being only two. Although scholars are divided on the point and there are indications from Spinoza's correspondence that he held

the former view, it is important to realize that the argument for monism requires only the latter. This argument turns on the claim that God is a substance that possesses all the attributes there are and, therefore, that there are none left for any other conceivable substance. Combining this with the proposition that two substances cannot share an attribute, it follows that there can be no substance apart from God.

The second problem is that the argument up to this point is completely hypothetical. It shows that *if* we assume the existence of God, defined as a substance possessing infinite attributes, then it follows that no substance apart from God is possible; but it has not yet established the existence of substance so conceived. Spinoza had, however, laid the foundation for this claim in *Ethics* (I, prop. 7) with the demonstration that existence pertains to the nature of substance; so it remains merely to apply this result to God. This is the task of proposition eleven, which contains Spinoza's arguments for the existence of God. Spinoza offers three separate proofs, the major one being his version of the ontological argument, which was first developed by Anselm and later reformulated by Descartes (see Anselm §4; Descartes, R. §6). Like his predecessors, Spinoza attempts to derive God's existence from the mere concept; but unlike them he makes no reference to God's perfection. Instead he appeals merely to the definition of God as a substance, from which it follows (by proposition seven) that God necessarily exists.

The nerve of the overall argument is, therefore, the proposition that existence pertains to the nature of substance or, equivalently, that its essence necessarily involves existence. Moreover, Spinoza's argument for this claim reveals the extent of his rationalism. From the premise that substance cannot be produced by anything external to itself (since such a cause would have to be another substance of the same nature, which has already been ruled out), he concludes that it must be the cause of itself, which is just to say that existence necessarily follows from its essence. Underlying this reasoning is what, at least since Leibniz, is usually termed the principle of sufficient reason, that is, the principle that everything must have a ground, reason or cause (these terms often being used interchangeably) why it is so and not otherwise. Although followers of Leibniz such as Christian Wolff attempted to demonstrate this principle, Spinoza, like most rationalists before Kant (including Leibniz), seems to have regarded it as self-evident (see Wolff, C. §6).

This characterization of substance as self-caused or self-sufficient being anticipates its identification with God. Quite apart from the question of its validity, however, perhaps the most interesting feature of Spinoza's argument for the necessary existence of his God-substance is that it is at the same time an argument for the non-existence of God the creator. Nevertheless, it should not be inferred from this that the result is purely negative or that Spinoza is concerned to deny the existence of God in every sense. On the contrary, his real concern in the opening propositions of the *Ethics* is to show that necessary existence and, therefore, the property of being a self-contained, self-explicated reality, is to be predicated of the order of nature as a whole rather than of some distinct and inscrutable ground of this order. And this also expresses the deepest meaning of his monism.

4 God and the world

Spinoza's monism does not, however, mean the end of all dualities. In fact, the identification of God with nature leads immediately to the distinction between two aspects of nature, which he terms *natura naturans* (active or generating nature), and *natura naturata* (passive or generated nature). The former refers to God as conceived through himself, that is, substance with infinite attributes, and the latter to the modal system conceived through these attributes (which includes, but is not identical to, the totality of particular things). Consequently, the task is to explain the connection between these two aspects of nature, a task which is the Spinozistic analogue to the traditional problem of explaining the relationship between God and creation.

Like the theologians, whose procedure he adopts even while subverting their claims, Spinoza divides his analysis into two parts: a consideration of the divine causality as it is in itself (or as *natura naturans*) and a consideration of it as expressed in the modal system (or as *natura naturata*). Given what we have already seen, the former holds few surprises. The basic claim is that 'From the necessity of the divine nature there must follow infinitely many things in infinitely many ways, (i.e., everything which can fall under an infinite intellect)' (*Ethics*: I, prop. 16). Spinoza here characterizes both the nature and extent of the divine causality or power. By locating this power in the 'necessity of the divine nature' rather than creative will and by identifying its extent with 'everything which can fall under an infinite intellect', that is, everything conceivable, Spinoza might seem to be denying freedom to God. Certainly he is denying anything like freedom of choice. To object on these grounds, however, is to ignore the conception of freedom which Spinoza does affirm and to appeal to the very anthropomorphic conception of the

deity against which his whole analysis is directed. To be free for Spinoza is not to be undetermined but to be self-determined (*Ethics*: I, def. 7); and God, precisely because he acts from the necessity of his own nature, is completely self-determined and, therefore, completely free.

The question of the relationship between God, so construed, and the ‘infinitely many things’ or modes that supposedly follow from God in ‘infinitely many ways’, which is perhaps the central question in Spinoza’s metaphysics, is greatly complicated by the fact that Spinoza distinguishes between two radically distinct types of modes. As modifications of the one substance, both types are dependent on and in a sense ‘follow from’ God, but they do so in quite different ways.

First, there are those modes that either follow directly from an attribute of substance or follow from one that does directly follow. These are termed respectively ‘immediate’ and ‘mediate eternal and infinite modes’. They are eternal and infinite because they follow (logically) from an attribute of substance, but they are not eternal and infinite in the same manner as substance and its attributes. Although Spinoza tells us very little about these modes in the *Ethics*, we know from his correspondence and the *Short Treatise on God, Man, and his Well-Being* (*c.*1660-5)) which contains the earliest statement of his system, that he regarded ‘motion’ or ‘motion and rest’, and ‘intellect’ or ‘infinite intellect’, as immediate eternal and infinite modes in the attributes of extension and thought respectively. As an example of a mediate eternal and infinite mode, he mentions only the ‘face of the whole universe’ (*facies totius universi*) which pertains to extension.

Given the highly schematic and fragmentary nature of the surviving accounts of these eternal and infinite modes, any interpretation is hazardous. Nevertheless, both their systematic function as mediators between God or substance and the particular things in nature (finite modes), and the names chosen for the modes of extension, suggest that the latter are best construed as the fundamental laws of physics. In fact, Spinoza’s characterization of motion and rest as an eternal and infinite mode may be seen as his attempt to overcome a basic difficulty in Cartesian physics. Having identified matter with extension, Descartes could not account for either motion or the division of matter into distinct bodies without appealing to divine intervention. On Spinoza’s view, however, this is not necessary, since extended substance has its principle within itself. Otherwise expressed, matter is inherently dynamical, a property which cannot be explained in terms of Descartes’ purely geometrical physics.

The ‘face of the whole universe’, which is identified with corporeal nature as a whole, may be understood in a similar fashion. By claiming that this is a mediate eternal and infinite mode of extension following directly from motion and rest, Spinoza is implying that the proportion of motion and rest in corporeal nature as a whole remains constant, even though it may be in continual flux in any given region. Moreover, this is equivalent to affirming the principle of the conservation of motion, which is a basic principle of Cartesian physics.

Viewing Spinoza’s account in the light of seventeenth-century physics also helps to understand his doctrine of finite modes, or the series of particular things, which is usually regarded as one of the more problematic aspects of his metaphysics. The problem is how to conceive the relationship between the series of these modes and God. If one assumes that, like the mediate eternal and infinite modes, they follow mediately from the attributes of God, then they too become eternal and infinite; but this is absurd, since it is of the essence of such modes to be transitory. If, however, one denies that they follow from God at all, then the dependence of all things on God, and with it Spinoza’s monism, is negated. Accordingly, it must be explained how, in spite of their finitude, particular things and occurrences depend on God and participate in the divine necessity.

Spinoza’s solution to this dilemma consists in claiming that the series of finite modes constitutes an infinite causal chain, wherein each finite mode is both cause and effect of others, *ad infinitum*, while the entire series (viewed as a totality) is dependent on the attributes of God and the eternal and infinite modes. Expressed in scientific terms, this means that every occurrence in nature is to be understood in terms of two intersecting lines of explanation. On the one hand, there is a set of general laws which for Spinoza are logically necessary (since they follow from the divine attributes); on the other hand, there must be a set of antecedent conditions. Both are required to explain a given phenomenon, say a clap of thunder. Clearly, no such explanation is possible without appealing to the relevant physical laws; but of themselves these laws are not sufficient to explain anything. It is also necessary to refer to the relevant antecedent conditions: in this case the state of the atmosphere. But, given these laws and the atmospheric state at t_1, it is possible to deduce the occurrence of thunder at t_2. And this means that nature is to be conceived as a thoroughly deterministic system.

Spinoza concludes from this that ‘In nature there is nothing contingent, but all things have been determined from the necessity of the divine nature to exist and to produce an effect in a certain way’ (*Ethics*: I, prop. 29). But because of this denial of contingency, he is sometimes accused of conflating determinism with the stronger thesis (usually termed ‘necessitarianism’) according to which the entire order of nature, that is, the infinite series of finite modes, could not have been different. The point is that determinism entails merely that, given the laws of nature and the set of appropriate antecedent conditions, any particular occurrence is necessary; but this leaves room for contingency, since it leaves open the possibility of a different set of antecedent conditions.

In response, one might distinguish between a consideration of finite modes, or a subset thereof, taken individually and a consideration of the series of such modes as a whole. The former fails to eliminate contingency, since any particular mode or subset of modes, viewed in abstraction from the whole, can easily be thought (or imagined) to be different. But the same cannot be said of the series taken as a whole. Since this series (considered as a totality) depends on God, it could not be different without God being different (which is impossible). The problem here is with the idea that the set of modes as a whole requires an explanation or grounding distinct from that of its constituent elements. Such a move is usually dismissed as a ‘category mistake’ (treating a collection as if it were a higher-order individual). But if it is a mistake, it is one to which Spinoza is prone in virtue of his rationalism; for nothing could be less Spinozistic than the idea that while particular events may be intelligible, the order of nature as a whole is not. And since for Spinoza making something intelligible involves demonstrating its necessity, this commits him to necessitarianism.

5 The human mind

Just as the target of part one of the *Ethics* is the dualism of God and created nature, so that of part two is the dualism of mind and body. Rather than holding with Descartes that the mind and the body are two distinct substances that somehow come together to constitute a human being, Spinoza maintains that they constitute a single individual expressed in the attributes of thought and extension. Since the fundamental modifications of thought are ideas (other modifications, such as desires and volitions, presuppose an idea of their object), while those of extension are bodies, this means that the human mind is an idea of a rather complex sort and that together with its correlate or object in extension (the body), it constitutes a single thing or individual. The great attractiveness of this view, particularly when contrasted with both Descartes’ dualism and Hobbes’ reductive materialism, is that it allows for the conception of persons as unified beings with correlative and irreducible mental and physical aspects (see Hobbes, T.). Unfortunately, this attractiveness is diminished considerably by its inherent obscurity. How can the mind be identified with an idea (even a very complex one)? And how can such an idea constitute a single thing with its object?

The place to begin a consideration of these questions is with Spinoza’s elusive conception of an idea, which he defines as ‘a conception of the mind that the mind forms because it is a thinking being’ (*Ethics*: II, def. 3). As he makes clear in his explication of this definition through the distinction between conception and perception, the emphasis falls on the activity of thought. To say that the mind has the idea of *x* is to say that it is engaged in the activity of conceiving *x*, not merely passively perceiving its mental image. Indeed, in one sense of the term, an idea for Spinoza just is the act of thinking. Moreover, this helps to remove at least some of the mystery in the identification of the mind with an idea. On the Spinozistic view, this means that the mind is identified with its characteristic activity, thinking; that its unity is the unity of this activity.

As acts of thinking, ideas may be identified with beliefs or ‘believings’; but this reflects only one dimension of Spinoza’s conception of an idea. For beliefs have propositional content and this, too, is an essential aspect of every idea. In short, ideas have both psychological and logical (or epistemological) properties. Moreover, although Spinoza is often charged with conflating these, he was well aware of the difference and of the importance of keeping them apart. This is evident from his appeal to the scholastic distinction, also invoked by Descartes, between the ‘formal’ and the ‘objective reality’ of ideas. The former refers to the psychological side of ideas as acts of thinking or mental events, the latter to their logical side or propositional content. Construed in the former way, ideas have causes which, in view of the self-contained nature of each attribute, are always other ideas. Construed in the latter way, they have rational grounds which likewise are always other ideas.

Spinoza differs from Descartes, however, in his understanding of the objective reality of ideas. For Descartes, talk

about the objective reality of an idea as it exists in someone's mind refers to that idea *qua* intentional object to which a 'real' (extra-mental) object may or may not correspond. For Spinoza, by contrast, the idea viewed objectively just is its object (a corresponding mode of extension) as it exists in thought. This is a direct consequence of Spinoza's mind-body monism and we shall see that it has important implications for his epistemology.

6 The human mind (cont.)

Our immediate concern, however, is with the implications of this conception of ideas for Spinoza's account of the mind-body relationship. Unquestionably, the key feature in this account is the principle that 'The order and connection of ideas is the same as the order and connection of things' (*Ethics*: II, prop. 7). Taken by itself, this might be viewed as the assertion of a parallelism or isomorphism between the two orders, thereby leaving open the possibility that the elements contained in these orders might be ontologically distinct, as, for example, in Leibniz's pre-established harmony (see Leibniz, G.W. §§4-7). In the scholium attached to this proposition, however, Spinoza indicates that he takes it to entail something more. Thus, he explains that just as thinking and extended substance are 'one and the same substance, comprehended now under this attribute and now under that. So also a mode of extension and the idea of that mode are one and the same thing, but expressed in two ways' (*Ethics*: II, prop. 7). In other words, rather than there being two series, one of extended things and the other of ideas, there is only a single series of finite modes, which may be regarded from two points of view, or taken under two descriptions. This also defines the sense in which Spinoza affirms an identity between mind and body. In claiming that mind and body constitute the same thing, he is not asserting that ultimately there is only one set of properties (which would make him a reductionist like Hobbes); rather, he is denying that the two sets of properties or, better, the two descriptions, can be assigned to two ontologically distinct things (as they are for Descartes).

Although the proposition makes a completely general claim about the relationship between the attributes of thought and extension, and their respective modifications, it also provides the metaphysical foundation for Spinoza's descent from the attribute of thought to its most interesting finite modification - the human mind. The descent is somewhat circuitous, however, since in subsequent propositions Spinoza stops to dwell on some topics that do not seem directly germane, such as the status of ideas of non-existent things; but the main line of the argument is clear enough. As a finite mode of thought, the essence of the mind must be constituted by an idea. Since the mind itself is something actual (an actual power of thinking), it must be the idea of an actually existing thing. And since this actually existing thing can only be a corresponding modification of extension, that is, a body, Spinoza concludes that 'The object of the idea constituting the human mind is the body, or a certain mode of extension which actually exists, and nothing else' (*Ethics*: II, prop 13).

Even if one accepts Spinoza's premises and the chain of reasoning leading to the conclusion, this 'deduction' of the human mind as the idea of the body raises at least two major questions. One is how to reconcile the identification of the body as the unique object of the mind with the capacity of the mind to know and, therefore, presumably, to have ideas of things quite distinct from the body with which it is identified. This will be discussed in the next section in connection with Spinoza's epistemology. The other question concerns how this enables us to understand what is distinctive about the human mind. As Spinoza himself remarks in the scholium to proposition thirteen, this result is perfectly general, applying no more to human beings than to other individuals. And, by way of accentuating the point, he adds that these other individuals are 'all animate [*animata*], albeit in different degrees'. Now, given the principles of Spinoza's metaphysics, it certainly follows that there must be 'in God' an idea corresponding to every mode of extension in the same way as an idea of a human body corresponds to that body. Thus, the claim that the human mind is the idea of the body may serve to determine its ontological status, but it does not enable us to understand its specific nature and activity. But unless Spinoza's deduction of the mind can accomplish this result, it cannot be judged a success, even on his own terms. Moreover, by suggesting that all individuals in nature are to some extent *animata*, he introduces a fresh element of paradox into his discussion. Indeed, this is particularly so if we take Spinoza to be claiming that something like a soul, or a rudimentary mind, must be attributed to all individuals.

Since the Latin '*animatus*' is cognate with the English 'animate', the sense of paradox can be lessened somewhat if we take the claim to be merely that all individuals are alive. Although this itself might seem bizarre, it becomes more plausible when one considers Spinoza's conception of life which, in the appendix to his early work

Descartes's 'Principles of Philosophy', he defines as 'the force through which things persevere in their being'. Since, as we shall see, Spinoza's conatus doctrine consists in the claim that each thing, in so far as it can, strives to preserve its being, it follows that every thing is in this sense 'alive'. And from this it is perhaps not too large a step to the conclusion that every thing has a 'soul' in the sense of an animating principle. But, of course, it does not follow from this that everything has a mental life that is even remotely analogous to that enjoyed by the human mind.

This makes it incumbent on Spinoza to account for the different degrees of animation and to explain thereby the superiority of the human mind to the 'minds' of other things in the order of nature. And he proceeds to do so by focusing on the nature of body. In essence, Spinoza's view is that 'mindedness' is a function of organic complexity; so the greater the capacity of a body (that is, brain and central nervous system) to interact with its environment, the greater the capacity of the mind to comprehend it. Thus, in a kind of speculative biophysics, Spinoza attempts to demonstrate that the human body is, indeed, a highly complex individual, which stands in a complex and reciprocal relationship with its environment. Although this account is extremely cryptic, it is also highly suggestive and points in the direction of an analysis of the phenomenon of life that goes far beyond the crude mechanism of the Cartesians (for whom the body is merely a machine). Perhaps more to the point, it also provides a theoretical basis for locating conscious awareness and rational insight on a continuum of mental powers, all of which are strictly correlated with physical capacities, rather than viewing them with Descartes as unique properties of a distinct mental substance.

7 Theory of knowledge

Although Spinoza does not assign to epistemological questions the priority given them by Descartes and the British empiricists, he certainly does not neglect them. In fact, his analysis of human knowledge, which follows directly upon his account of the mind, may be viewed as an attempt to show how the human mind, so conceived, is capable of the kind of knowledge presupposed by the geometrical method of the *Ethics*. In Spinoza's own terms, this means that what must be shown is nothing less than that the human mind is capable of adequate knowledge of the eternal and infinite essence of God.

Since adequate knowledge rests on adequate ideas, Spinoza must demonstrate that the human mind possesses such ideas. He defines an adequate idea as one which 'considered in itself, without relation to an object, has all the properties or intrinsic denominations of a true idea' (*Ethics*: II, def. 4). The term 'intrinsic' functions to rule out the extrinsic feature of a true idea: namely, agreement with its object. Thus, Spinoza's view is that truth and adequacy are reciprocal concepts: all adequate ideas are true (agree with their object) and all true ideas are adequate. Moreover, this enables him to dismiss the radical doubt regarding even our most evident conceptions envisaged by Descartes, which is supposedly overcome only by the manifestly circular appeal to God as the guarantor of truth. Since adequacy is an intrinsic feature of all true ideas, it serves as the criterion of truth. Consequently, someone who has a true idea immediately recognizes it as such and there is no longer room for Cartesian doubt. As Spinoza puts it with uncharacteristic elegance: 'As the light makes both itself and the darkness plain, so truth is the standard both of itself and of the false' (*Ethics*: II, prop. 43, scholium).

The intrinsic property through which truth manifests itself is explanatory completeness. An idea of *x* is adequate and, therefore, true just in case it suffices for the determination of all of the essential properties of *x*. For example, the mathematician's idea of a triangle is adequate because all of the mathematically relevant properties of the figure can be deduced from it. Conversely, the conception of a triangle by someone ignorant of mathematics is inadequate because it is incapable of yielding any such consequences.

It does not follow from this, however, that inadequate ideas are simply false. On the contrary, since every idea agrees with its object, every idea must in some sense be true. Specifically, 'All ideas, insofar as they are related to God, are true' (*Ethics*: II, prop. 32). Error or falsity arises because not every idea possessed by the human intellect is related by that intellect to God, that is, viewed as a determinate member of the total system of ideas. In other words, error or falsity is a function of incomplete comprehension, of partial truth being taken as complete. Spinoza illustrates the point by an example also used by Descartes: the imaginative, non-scientific idea of the sun as a disk in the sky located a few hundred feet above the earth. This idea is 'true' in so far as it is taken as an accurate representation of how the sun appears to us under certain conditions; but since it is not understood in this way by someone ignorant of optics and astronomy, such a person's idea is false in the sense of being inadequate or

incomplete. It is not, however, 'materially false' in the Cartesian sense that there is nothing in the realm of extension corresponding to it.

Correlative with the distinction between inadequate and adequate ideas is a contrast between two mutually exclusive ways in which ideas can be connected in the mind: either according to the 'common order of nature' or the 'order of the intellect'. The former refers to the order in which the human mind receives its ideas in sense perception or through imaginative association. Since these correspond exactly to the order in which the body is affected by the objects of these ideas, it reflects the condition of the body in its interaction with its environment rather than the true nature of an independent reality. And from this Spinoza concludes that all such ideas are inadequate. In fact, he argues that, in so far as its ideas reflect this order or, equivalently, are based on sense perception or imagination, the mind is incapable of adequate knowledge of either external objects in nature, its own body or even of itself. Spinoza also contends, however, that adequate knowledge of all three is possible in so far as the mind conceives things according to the order of the intellect, that is, the true order of logical and causal dependence, which, once again, is precisely why the correct method is so essential.

8 Theory of knowledge (cont.)

The central problem of Spinozistic epistemology is thus to explain how conception according to the order of the intellect is possible for the human mind. This is a problem because the possibility of such conception seems to be ruled out by the ontological status of the mind as the idea of the body. For how could a mind, so conceived, have any ideas that do not reflect the condition of its body? The gist of Spinoza's answer is that there are certain ideas that the human mind possesses completely and hence can conceive adequately because, unlike ideas derived from sense perception or imagination, they do not 'involve', or logically depend on, ideas of particular modifications of the body.

Spinoza's doctrine is particularly obscure at this point; but it is perhaps best approached by comparing it to the doctrine of innate ideas, which was appealed to by Descartes and later Leibniz to deal with a similar problem. Like Spinoza, they held that sensory experience cannot account for the possibility of knowledge of necessary and universal truth. Instead, they claimed that the source of such knowledge must lie in the mind and reflect its very structure. This was not understood in a naïve psychological sense, however, as if an infant were born with knowledge of the basic principles of mathematics. Rather, innate ideas were viewed more as dispositions that pertain essentially to the human mind, but of which individuals are not necessarily conscious (see Innate knowledge).

Although Spinoza's account of the mind as the idea of the body precludes the distinction drawn by Descartes between innate and adventitious ideas (that is, those that come from the mind and those that come from sensory experience), it does allow for an analogous distinction, which leads to much the same result. This is the distinction between ideas that are correlated with specific features of particular bodies and those whose correlates are common to all bodies or a large proportion thereof. The latter fall into two classes, corresponding to two levels of generality, which Spinoza terms respectively 'common notions' and 'adequate ideas of the common properties of things'. Their distinctive feature is that they do not arise in connection with an encounter with any particular kind of thing; and this enables Spinoza to claim that the mind possesses them in their totality and comprehends them adequately. Unfortunately, he does not provide examples of either class of these adequate ideas; but it seems reasonable to assume that the common notions include the axioms of geometry and first principles of physics (which are common to all bodies). Correlatively, since the adequate ideas of common properties of things correspond to properties that are common and peculiar to the human body and to other bodies by which it is affected (*Ethics*: II, prop. 39), it is likely that Spinoza was here referring to the basic principles of biology (or perhaps physiology). In any event, the crucial point is that the commonality of these ideas enables the human mind to grasp them completely, which is what is required for adequate knowledge.

The epistemological teaching of the *Ethics* culminates in the distinction between three kinds of knowledge (*Ethics*: II, prop. 40, scholium 2). The first is an experientially determined knowledge, which can be based either on the perception of particular things or on signs, which for Spinoza includes both sensory and memory images. The second is knowledge through reason, which is based on common notions and ideas of the common properties of things. Since the former mode of knowledge involves inadequate ideas and the latter adequate ones, this is just the contrast one would expect. At this point, however, Spinoza unexpectedly introduces a third kind of knowledge,

termed ‘intuitive knowledge’ (*scientia intuitiva*), which supposedly ‘proceeds from an adequate idea of the formal essence of certain attributes of God to the adequate knowledge of the essence of things’. He also attempts to clarify the difference between all three by comparing their respective treatments of the problem of finding a fourth proportional. Someone with the first kind of knowledge proceeds by rule of thumb, multiplying the second by the third and dividing the product by the first, thereby arriving at the correct answer without really understanding the principle at work. Someone with the second kind understands the principle and is, therefore, able to derive the result from the common property of proportionals as established by Euclid. Someone who possesses the third kind of knowledge, however, immediately grasps the result, without applying a rule or relying on a process of ratiocination.

Although both reason and intuition are sources of adequate knowledge, Spinoza recognizes two senses in which intuition is superior. First, whereas the province of reason concerns general truths based on common notions and consequently is abstract and general, that of intuition concerns the individual case and consequently is concrete and particular. This difference is not directly germane to Spinoza’s epistemology, but it plays a significant role in his practical philosophy. Second, and of more immediate relevance, knowledge from general principles alone remains ultimately ungrounded. Accordingly, as in Spinoza’s example, while the conclusion is inferred correctly from the principle, the status of the principle itself remains in question. Within the framework of Spinoza’s metaphysics, this question can be resolved only by grounding the principle in the nature of God, which is supposedly what is accomplished by intuitive knowledge.

But grounding our knowledge of things in the eternal and infinite essence of God obviously presupposes a knowledge of that essence; and even for a rationalist such as Descartes, this far transcends the capacity of the human intellect. This is not the case for Spinoza, however, given his unique conception of God. Since ‘it is of the nature of reason to regard things as necessary, not as contingent’ (*Ethics*: II, prop. 44), and since to conceive things in this way is to conceive them in relation to God, Spinoza in effect concludes that in so far as the mind has an adequate idea of anything at all, it must have an adequate idea of God. Moreover, since it presumably has been established that the human mind knows some things (has some adequate ideas), it follows that the mind has an adequate idea of God. Although initially this seems paradoxical in the extreme, it becomes much less so if one keeps in mind the nature of Spinoza’s God.

9 The emotions

Nowhere is Spinoza’s unique combination of rationalism and naturalism more evident than in his doctrine of the emotions, the topic of the third part of the *Ethics*. Appealing to the conception of the mind as the idea of the body, he defines the emotions or affects (*affectiones*) as ‘affections of the body by which the body’s power of acting is increased or diminished, aided or restrained, and at the same time, the ideas of these affections’ (*Ethics*: III, def. 3). Emotions are, therefore, directly related to the body’s capacity for action or level of vitality and have both a physiological and a psychological side. Moreover, given mind-body identity, these are seen as distinct expressions of the same thing. Thus, a conscious desire for some object and a corresponding bodily appetite are the same state considered under the attributes of thought and extension respectively.

Among other things, Spinoza’s conception of mind precludes the assumption of a distinct power of will through which the mind can exert control over the bodily appetites. Strictly speaking, there is nothing pertaining to the mind but ideas, that is, acts of thinking; but these ideas have a conative or volitional and affective as well as a cognitive dimension. In other words, to have an idea of *x* is to have not only a belief or propositional attitude with respect to *x*, but also some sort of evaluative attitude (pro or con). Moreover, this enables Spinoza to avoid concluding from his denial of will that the mind is powerless, condemned to being the passive observer of the bodily appetites. The mind, for Spinoza, is active in so far as it is the ‘adequate’, that is sufficient, cause of its states; and it is such in so far as it possesses adequate ideas, that is, in so far as its desires and hence its ‘decisions,’ are grounded in rational considerations - for example, when it desires a particular food because of the knowledge (adequate idea) that it is nutritious. Conversely, it is passive when its desires reflect inadequate ideas connected with sense perception and imagination.

But regardless of whether it is active or passive, the mind’s evaluative attitude is always an expression of its conatus, which is identified with the endeavour of each thing to persist in its own being. This endeavour pertains to the nature of every finite mode; but in human beings, who are conscious of this endeavour, it becomes the desire

for self-preservation. Spinoza thus agrees with Hobbes in regarding this desire as the basic motivating force in human behaviour. But rather than inferring this from observation, he deduces it from the ontological status of human beings as finite modes. This allows him to affirm not merely that this desire is, as a matter of fact, basic to human beings, but that it constitutes their very essence. Accordingly, one can no more help striving to preserve one’s being than a stone can help falling when it is dropped.

Perhaps under the influence of Hobbes, Spinoza also identifies this endeavour to preserve one’s being with a striving for greater perfection, understood as an increased power of action. Just as Hobbes insisted that individuals continually desire to increase their power because there can never be any assurance that it is sufficient for self-preservation, so Spinoza maintains that the endeavour of an organism to preserve its existence is identical to its effort to increase its power of acting or level of vitality. This is because anything that lessens an organism’s power, lessens its ability to preserve its being, while anything that increases this power enhances that ability.

The so-called primary emotions (pleasure, pain and desire) are correlated with the transition from one state of perfection or level of vitality to another. Thus, pleasure or joy (*laetitia*) is defined as the emotion whereby the mind passes to a greater state of perfection, and pain or sorrow (*tristitia*) is that by which it passes to a lesser state. Both reflect changes brought about in the organism through interaction with its environment. Although particular desires are directly related to pleasures, desire is none the less a distinct primary emotion because a desire for a particular object viewed as a source of pleasure is distinct from that pleasure.

The great bulk of part three of the *Ethics* is devoted to showing how the other emotions can be derived from the primary ones by means of combination and association with other ideas. Central to this project is the thesis that pleasure, pain and desire relate to present objects, which cause the affections of the body to which these emotions (as ideas) correspond. Accordingly, the derivative emotions are all species or combinations of pleasure, pain or desire, which are directed in various ways either to objects that do not presently affect the body or to those that are not themselves directly the cause of these affections. For example, love and hatred are defined respectively as pleasure and pain accompanied by the idea of an external cause. Similarly, hope and fear are understood as pleasures and pains that arise under conditions of uncertainty, when the image of some past or future thing is connected with an outcome that is in doubt. This analysis is extended with considerable subtlety, showing, among other things, how ambivalence is possible; and how, through various forms of association, the mind can come to feel love, hatred, hope or fear towards things with regard to which it has no direct desire or aversion. The key point, however, is that, like everything else in nature, these emotions do not arise capriciously, but in accordance with universal and necessary laws.

But these laws concern the mind only in so far as it is passive, that is, in so far as its ideas are inadequate and it is, therefore, only the partial or inadequate cause of its affections. Accordingly, at the end of his lengthy account of the passive emotions or passions, Spinoza turns briefly to the active emotions, which are connected with adequate ideas. Of the three primary emotions, only desire and pleasure have active forms, because only they can be grounded in adequate ideas. Active desire has already been described; it is simply rational desire. Active pleasure is a concomitant of all adequate cognition; for when the mind cognizes anything adequately it is necessarily aware of this and, therefore, of its own power or activity. And it is the awareness of its activity, not the nature of the object known, that is the source of pleasure. Finally, since pain or sorrow reflects a diminution of the mind’s power of activity (adequate ideas), Spinoza concludes that it can never be the result of this activity, but must always result from the mind’s inadequate knowledge and determination by external forces.

10 Moral theory

Spinoza’s moral theory is based on his analysis of the emotions and is formulated with the same clinical detachment as the remainder of the *Ethics*. In sharp contrast to Judaeo-Christian moralists and their secular counterparts, he proposes neither a set of obligations nor a list of actions, the performance of which make one morally ‘good’, and their omission or neglect morally ‘evil’: all such moral systems and concepts are based on inadequate ideas, particularly the ideas of free will and final causes. Instead, he is concerned to determine the means through which and the extent to which human beings, as finite modes, are capable of attaining freedom, understood as the capacity to act rather than to be governed by the passions. Morality in the traditional sense is, therefore, replaced by a kind of therapy, which is one reason why Spinoza is frequently compared with Freud. The concept of virtue is retained; but it is given its original meaning as power, which is itself understood in light of the

conatus doctrine as the power to preserve one’s being. In the same spirit, the good is identified with what is truly useful in this regard and the bad with what is truly harmful.

In spite of his amoralism, Spinoza does not equate virtue with the ability to survive or the good with what is in one’s self-interest, narrowly conceived. What matters is not mere living, but living well; and this means being active - that is, being, to the fullest extent possible, the adequate cause of one’s condition. And since being an adequate cause is a function of adequate ideas, virtue is directly correlated with knowledge. Knowledge, however, has a dual role in the Spinozistic scheme. It is the major weapon in the struggle against the passions, since it is through understanding our passions and their causes that we are able to gain some measure of control over them. But it is also itself constitutive of the good life, since our freedom is manifested essentially in the exercise of reason.

Nevertheless, Spinoza was under no illusions about the extent of the power of reason. Human virtue or perfection is merely relative and its attainment a rare and difficult feat. Thus, the first eighteen propositions of part four of the *Ethics*, which is significantly entitled ‘On Human Bondage, or the Powers of the Affects’, are concerned with the limits of the power of reason in its conflict with the passions. The basic point is that, as finite modes, the force through which human beings endeavour to preserve their being is infinitely surpassed by other forces in nature and, therefore, to some extent at least, they will always be subject to the passions. Moreover, knowledge itself has no motivating power simply *qua* knowledge, but only in so far as it is also an affect. Now knowledge is, indeed, an affect for Spinoza, since all ideas have an affective component, that is, possess a certain motivational force. But, as he attempts to demonstrate by means of an elaborate psychodynamics, the affective component of even an adequate idea is strictly limited and can easily be overcome by other (inadequate) ideas, which is why rational desires, based on a knowledge of what is truly beneficial, all too frequently give way to irrational urges.

After his analysis of human weakness, Spinoza turns to the question of what reason, limited as its power may be, prescribes. The basic answer, of course, is knowledge and, given Spinoza’s metaphysics and epistemology, this ultimately means knowledge of God. Thus, he concludes that ‘Knowledge of God is the mind’s greatest good; its greatest virtue is to know God’ (*Ethics*: IV, prop. 28). At this point, however, the discussion takes a surprising turn, one which indicates both the complexity of, and the inherent tensions in, Spinoza’s thought. For while this austere intellectualism suggests the picture of the isolated, asocial thinker, devoted exclusively to the life of the mind, what is affirmed instead is the essentially social character of human existence. For Spinoza, as for Aristotle, human beings are social animals; and the life lived under the guidance of reason is, at least to some extent, a social life (see Aristotle §22). This is not because human beings are intrinsically altruistic, but rather because, as relatively limited and weak finite modes, they are ineluctably interdependent. Thus, Spinoza argues that those who live under the guidance of reason desire nothing for themselves that they do not also desire for others (*Ethics*: IV, prop. 37). This reflects his undoubtedly idealized portrait of those devoted to the life of the mind. In so far as this devotion is pure (which it can never be completely), such individuals will not come into conflict because the good which they all seek, knowledge, can be held in common. In fact, not only will genuine seekers after truth not compete, they will cooperate; for in helping others acquire knowledge and the control of the passions that goes with it, one is also helping oneself. Moreover, although only the few capable (to some extent) of living under the guidance of reason may be able to grasp adequately and, therefore, internalize this truth, the need for cooperation applies to all; for all are members of the same human community of interdependent beings.

11 Moral theory (cont.)

Spinoza’s account of the specific virtues reflect his general principles. These virtues are identified with certain affects or emotional states and their value is regarded as a function of their capacity to promote an individual’s conatus. For this purpose the affects are divided into three classes: those that are intrinsically good and can never become excessive (the virtues); those that are intrinsically bad; and a large group that are good in moderation but bad if they become excessive. In identifying the virtues with affects that can never become excessive, Spinoza differs from Aristotle for whom virtues are regarded as means between two extremes.

Paramount in the group of virtuous affects is pleasure or joy. Since it reflects in the attribute of thought an increase in the body’s power of activity, it can never be harmful. This gives Spinoza’s thought a strongly anti-ascetic tone which stands in sharp contrast to the Calvinistic austerity of many of his countryman. Nevertheless, it is crucial to distinguish between genuine pleasure, which reflects the wellbeing of the organism as a whole, and titillation

(*titillatio*) or localized pleasure, which merely reflects that of a part. Although the latter can be good, it can also be quite harmful. Other affects in this mixed category include cheerfulness and self-esteem. In the latter case, the crucial factor is whether or not the affect is grounded in reason. Pride, or self-esteem, without any rational basis is obviously harmful and is to be avoided at all costs. But, in so far as self-approval arises from an adequate idea of one's power, it is the highest thing we can hope for, since it is simply the consciousness of one's virtue. Perhaps even more than his anti-asceticism, this indicates how much closer Spinoza is to the classical ideal of the virtuous life than he is to traditional religious morality.

In addition to pain, Spinoza assigns first place among the intrinsically harmful affects to hate. Closely associated with hate, and rejected in similarly unqualified terms, are affects such as envy, derision, contempt, anger and revenge. These might be termed the social vices, since they serve to alienate human beings from one another. It is also noteworthy that Spinoza here locates many of the traditional religious virtues: hope, fear, humility, repentance and pity. Since they all reflect ignorance and lack of power, they cannot be regarded as beneficial, or assigned any place in the life of reason. Nevertheless, in a concession to human frailty, Spinoza does acknowledge that because human beings rarely live in this manner, these affects have a certain pragmatic value as checks on our more aggressive tendencies.

The affects that can be either good or bad include - besides titillation - desire and love. If directed towards an object that stimulates or gratifies a part of the organism or one of its appetites at the expense of the whole, they can become excessive and hence harmful. And this is precisely what occurs in pathological states such as avarice, ambition, gluttony and above all, lust. In spite of his generally anti-ascetic attitude, Spinoza tended to regard sexual desire as an unmitigated evil. In sharp contrast to it stands the one kind of love that can never become excessive: the love of God.

12 The love of God and human blessedness

Although it does not enter into the account of virtue, the love of God plays a central role in the final part of the *Ethics*. The situation is complicated, however, by the fact that this part deals with two distinct topics, which may be characterized as mental health and blessedness; and the love of God is crucial to both. The question of health is the subject of the first twenty propositions, which are concerned with reason's unremitting struggle with the passions. Here Spinoza functions explicitly as therapist, providing his alternative not only to the religious tradition, but also to the training of the will advocated by the Stoics and Descartes (see Stoicism). Since what is crucial is that, as far as possible, one be moved by pleasurable thoughts (rather than reactive, negative affects), Spinoza's account amounts to an essay on the power of positive thinking. Moreover, since the ultimate positive thought is the love of God, this love serves as the chief remedy against the passions.

In spite of the religious language in which it is expressed, this claim is readily understandable in Spinozistic terms. Since by 'love' is meant simply pleasure accompanied by the idea of its cause, any pleasure accompanied by God as its cause counts as the love of God. But all adequate cognition is both inherently pleasurable, since it expresses the activity of the mind, and involves the idea of God as cause, since it consists in an idea of the object as following from God (the third kind of knowledge). Thus, the adequate knowledge of anything involves the love of God as its affective dimension. Moreover, since, in principle at least, it is possible to acquire an adequate idea of any modification of substance, it follows that this love can be occasioned by virtually anything. Consequently, it is its potential ubiquity, together with its superior affective force as expression of pure activity, that qualifies this love-knowledge of God as the supreme remedy against the passions.

In the second half of part five, the love of God is now explicitly characterized as 'intellectual' and paradoxically identified with the love with which God loves himself. And if this were not puzzling enough, Spinoza introduces the new discussion by proclaiming: 'With this I have completed everything which concerns this present life… so it is now time to pass to those things which pertain to the mind's duration without relation to the body' (*Ethics*: V, prop. 20, scholium). This sets the agenda for the final propositions, the basic concern of which is to show that 'The human mind cannot be absolutely destroyed with the body, but something of it remains which is eternal' (*Ethics*: V, prop. 23). It is within this context that Spinoza refers to the intellectual love of God.

Because of their apparent incompatibility with the central teachings of the *Ethics*, these final propositions remain a source of perplexity. Nevertheless, it does seem possible to make sense of Spinoza's thought here, if we see it in

the context of a shift of focus from the concern with reason (including the love of God) in its struggle with the passions, to a concern with the life of reason as the highest condition of which human beings are capable and, therefore, as constitutive of human blessedness. So construed, the abrupt change of tone is simply Spinoza's way of marking that shift rather than an indication of a lapse into a mysticism that is totally at variance with the spirit of his philosophy. Such a reading leads to an essentially epistemological rather than a metaphysical reading of the doctrine of the eternity of the mind. According to this reading, the human mind is 'eternal' to the extent to which it is capable of grasping eternal truth and ultimately of understanding itself by the third kind of knowledge, which, in turn, leads to the intellectual love of God. To be sure, this is not the eternal life promised by religion; but it is a state of blessedness or perfection in the sense that it involves the full realization of our capacities. Moreover, it is precisely the mode of life to which the *Ethics* as a whole points the way.

Such a reading is also supported by the final proposition in which Spinoza returns to the theme of virtue and links it with both health and blessedness. As he there puts it, 'Blessedness is not the reward of virtue, but virtue itself; nor do we enjoy it because we restrain our lusts; on the contrary, because we enjoy it, we are able to constrain them' (*Ethics*: V, prop. 41). There is not a trace of mysticism here, but merely the familiar Spinozistic emphasis on the connection between blessedness and knowledge on the one hand, and knowledge and power on the other. Accordingly, the point is that we do not acquire this knowledge by first controlling our lusts, but that we have the power to control them (virtue) only to the extent to which we already possess adequate knowledge. And it is in this knowledge, also characterized as peace of mind, that blessedness consists.

13 Politics

Spinoza's concern with political theory has its philosophical roots in his conception of human beings as social animals, which entails the necessity of living in a state under a system of laws; but it was also triggered by the political situation in his own time. In the Netherlands, the monarchist party was intent on overthrowing the republican form of government, and their allies, the Reformed clergy, desired to establish a state church. In spite of his commitment to a life of philosophical contemplation, Spinoza was keenly aware of this situation and the dangers it posed to freedom of thought and expression, which he regarded as essential. His philosophical response to this threat, as well as the statement of his own views about the nature and function of the state, are to be found in the *Tractatus Theologico-politicus (Theological-Political Treatise)* and the *Tractatus Politicus (Political Treatise)*. The former is a polemical work, intended, at least in part, as a response to the Reformed clergy, while the latter is a dispassionate essay in political science. But despite this difference in tone and some disagreement on substantive matters, both works argue for freedom as the supreme political value and both investigate the conditions under which it can be realized and preserved.

Spinoza's political thought is best approached by way of a comparison with Hobbes. Both thinkers view human beings as thoroughly determined parts of nature and as driven by the desire for self-preservation; both are amoralists in the sense that they hold that everyone has a 'natural right' to do whatever is deemed necessary for self-preservation; as a direct consequence of this view of natural right, both view the state of nature (the pre-political condition) as one of unavoidable conflict and insecurity; and, finally, both maintain that peace and security can be attained only if everyone surrenders all of their natural right to a sovereign power (which takes the form of a social contract). But whereas Hobbes concludes from his account the necessity of an absolute sovereign power, preferably in the form of a monarchy, Spinoza infers from substantially the same premises that the true end of the state is freedom, and, at least in the *Theological-Political Treatise*, that democracy is 'the most natural form of state'.

To some extent, these differences can be understood in terms of the different social and political conditions under which the two thinkers lived. There are also important philosophical differences, however, one of which is their respective conceptions of human reason. For Hobbes, reason has a merely instrumental value as a means to the attainment of ends dictated by desire. We have seen, however, that for Spinoza the goal is to transform desire through reason, which naturally leads him to focus on the conditions under which the life of reason can best be lived. Moreover, Spinoza seems to have arrived at his conclusions through a kind of internal critique of Hobbes. As he informed a correspondent in 1674:

> With regard to politics, the difference between Hobbes and me… consists in this, that I ever preserve the natural right intact so that the supreme power in a state has no more right over a subject than is proportionate to

> the power by which it is superior to the subject. This is what always takes place in the state of nature.
> (Spinoza 1966 letter 50: 269)

By suggesting that Hobbes did not keep natural right intact, Spinoza is implying that he did not consistently equate right with power. This is indeed true; but it does not explain how the identification of might with right enables Spinoza to arrive at his conclusions. The gist of the answer, as suggested by this passage, is that the identification applies also to sovereign power. In other words, rather than gaining absolute right over its subjects through the social contract, as Hobbes maintained, the sovereign's right is limited by its power; and since this power is inevitably limited, so too are the things that a sovereign may 'legitimately' demand of its subjects.

Among the things that a sovereign cannot require are acts so contrary to human nature that no threat or promise could lead a person to perform them. These include things such as forcing people to testify against themselves or to make no effort to avoid death. But Spinoza does not stop at such obvious cases. He also emphasizes the limitation of legislative power with respect to private morality; and he finds an argument for freedom of thought in the fact that a government is powerless to prevent it. More importantly, he points out that there are some things which a government can do by brute force, but in doing so inevitably undermines its own authority. And since a government cannot do these things with impunity, it does not have the 'right' to do them at all. Thus, he argues on entirely pragmatic grounds for the limitation of governmental power through the power of public opinion.

Spinoza's main concern as political theorist, however, is not to determine what the state cannot do, but rather what it should do in order to realize the end for which it was established. Moreover, while verbally agreeing with Hobbes in construing this end as peace and security, Spinoza understands these in a much broader sense. Accordingly, peace is not merely an absence of war or the threat thereof, but a positive condition in which people can exercise their virtue. Thus, the goal of the state is to create this condition, which is also the social condition necessary for the life of reason as depicted in the *Ethics*.

But the life of reason is only for the few, and political arrangements must concern the many. Moreover, since it is the many who determine the public opinion to which the government must pay heed if it is to rule effectively, it follows that there can be valid laws, approved at least tacitly by a majority, which are none the less inimical to true virtue. Spinoza was keenly aware of this problem; but his way of dealing with it indicates the tension between his democratic tendencies and his elitism that runs throughout his theory. Thus on the one hand he insists on the right of free expression, including the right to protest against laws deemed unjust, while on the other hand he emphasizes the necessity of total obedience to the existing law, no matter how contrary to reason it may be. The reason for this conservative turn, which is also reflected in the complete rejection of revolution as a political remedy, lies in his profound sense of the irrationality of the multitude. Given this irrationality, which poses a constant threat to the power of reason, Spinoza concludes with Hobbes that even under a tyrannical regime, obedience to the established authority is the lesser evil.

14 Scripture

Spinoza's revolutionary treatment of the Bible in the *Theological-Political Treatise* must also be understood within the framework of his political thought. In line with his concern to secure the freedom to philosophize, he launches a systematic attack on the authority of Scripture: its claim to be the revealed word of God. But rather than offering an external philosophical critique in the manner of the *Ethics*, he attempts, in a somewhat paradoxical fashion, to show from Scripture itself that it makes no such claim to authority. This strategy, in turn, rests on a new method of Biblical exegesis, one based on the Cartesian principle that nothing should be attributed to the text that is not clearly and distinctly perceived to be contained in it. In light of this principle, he rejects both the Calvinist doctrine that a supernatural faculty is required for interpreting the Bible and the older Jewish rationalism of Maimonides, which held that if the literal reading of a passage conflicts with reason, it must be interpreted in some metaphorical sense (see Maimonides, M.). Both of these approaches he regards as not only useless for interpreting the Bible, but as dangerous politically, since they lead to the establishment of spiritual authorities.

Applying his method, Spinoza argues that neither prophecy nor miracles, the twin pillars of biblical authority, are able to support the orthodox claims. The prophets are shown to differ from other individuals in their superior imaginations, not in their intellects. Similarly, biblical miracles are treated as natural occurrences, which only appeared mysterious to the biblical authors because of their limited understandings and, as such, have no probative

value. More generally, the Bible is viewed as a document which reflects the limited understandings of a crude people rather than the dictates of an omniscient deity. And by analysing Scripture in this way, Spinoza laid the foundation for the subsequent historical study of the Bible (‘higher criticism’), which endeavours to interpret it by the same methods applicable to any other ancient text.

Spinoza’s critique of the Bible is, however, largely directed against its speculative content and claim to be a source of theoretical truth. Thus, he affirms that in moral matters the Bible has a consistent and true teaching, which reduces essentially to the requirement to love one’s neighbour. Moreover, precisely because it appeals to the imagination rather than the intellect, it has the great virtue of presenting morality in a form which the multitude can grasp. Such a view of religion as morality for the masses is hardly original to Spinoza. It had already been expressed in the twelfth century by Averroes (see Ibn Rushd §4), and it found expression in many subsequent politically minded thinkers, including Machiavelli (see Machiavelli, N. §6). But, if not original to Spinoza, it is still an integral part of his political thought, since it enables him to ‘save’ religion while also protecting the autonomy of philosophy. And the latter is, of course, necessary for the life of reason as depicted in the *Ethics*.

See also: Aretè; Eudaimonia; God, concepts of; Monism; Substance; Will, the

HENRY E. ALLISON

List of works

Spinoza, B. de (1925) *Spinoza Opera*, ed. C. Gebhardt, 4 vols, Heidelberg: Carl Winter.(The standard edition of Spinoza’s writings.)

Spinoza, B. de (*c*.1660-5) *Korte Verhandelung van God, de Mensch, en Deszelfs Welstand (Short Treatise on God, Man, and his Well-Being)*, ed. and trans. A. Wolf, New York: Russell & Russell, 1963.(Written in Dutch, this is the earliest systematic statement of Spinoza’s philosophy, dating from the early 1660s but not published until 1862. The Wolf translation also has an introduction, commentary and life of Spinoza.

Spinoza, B. de (1663) *Renati Des Cartes Principiorum Philosphiae (Descartes’ ’Principles of Philosophy’)*, in *The Collected Works of Spinoza*, vol. 1, ed. and trans. E. Curley, Princeton, NJ: Princeton University Press, 1985.(An early work in which Spinoza expounds part of Descartes’ *Principles* for a student.)

Spinoza, B. de (1670) *Tractatus Theologico-politicus (Theological-Political Treatise)*, in *The Chief Works of Benedict de Spinoza*, vol. 1, trans. R.H.M. Elwes, New York: Dover, 1951.(A polemical work, which makes a major contribution to both political theory and interpretation of the Bible.)

Spinoza, B. de (1677a) *Ethica Ordine Geometrico Demonstrata (Ethics Demonstrated in a Geometrical Manner)*, in *The Collected Works of Spinoza*, vol. 1, ed. and trans. E. Curley, Princeton, NJ: Princeton University Press, 1985.(Spinoza’s major systematic work.)

Spinoza, B. de (1677b) *Tractatus de Intellectus Emendatione (Treatise on the Emendation of the Intellect)*, in *The Collected Works of Spinoza*, vol. 1, ed. and trans. E. Curley, Princeton, NJ: Princeton University Press, 1985.(An early fragment which contains an important account of Spinoza’s philosophical method.)

Spinoza, B. de (1677c) *Tractatus Politicus (Political Treatise)*, in *Benedict de Spinoza: The Political Works*, ed. and trans. A.G. Wernham, Oxford: Clarendon Press, 1958.(Although unfinished, it contains the most systematic statement of Spinoza’s political philosophy.)

Spinoza, B. de (1966) *The Correspondence of Spinoza*, trans. A. Wolf, London: Frank Cass.(Contains all the extant letters to and from Spinoza of philosophical significance.)

References and further reading

Allison, H.E. (1987) *Benedict de Spinoza: An Introduction*, New Haven, CT, and London: Yale University Press.(The contributor’s own general introduction to Spinoza.)

Bennett, J. (1984) *A Study of Spinoza’s Ethics*, Indianapolis, IN: Hackett Publishing Company.(A very challenging and controversial discussion of the central topics in the *Ethics*.)

Curley, E.M. (1988) *Behind the Geometrical Method, A Reading of Spinoza’s Ethics*, Princeton, NJ: Princeton University Press.(Clearly written, useful introduction to the *Ethics*, emphasizing the connections between Spinoza and Hobbes).

Delahunty, R.J. (1985) *Spinoza*, London: Routledge & Kegan Paul.(Contains a useful discussion of some of the major interpretive issues in the secondary literature.)

Donagan, A. (1989) *Spinoza*, Chicago, IL: University of Chicago Press.(Good, scholarly introduction to Spinoza’s thought.)

Gueroult, M. (1968) *Spinoza I, Dieu (Ethique, I)*, Paris: Aubier.(Probably the richest and most detailed study

available in any language of the first part of the *Ethics*.)
Gueroult, M. (1974) *Spinoza II, L'Ame (Ethique, II)*, Hildesheim: Olms.(The most detailed study of the second part of the *Ethics* available in any language.)
Hampshire, S. (1951) *Spinoza*, London: Faber.(A clearly written and still useful overview of Spinoza's thought).
Joachim, H.H. (1901) *A Study of the Ethics of Spinoza*, London: Oxford University Press; repr. New York: Russell & Russell, 1964.(One of the classic commentaries on the *Ethics*.)
Wolfson, H.A. (1958) *The Philosophy of Spinoza*, New York: Meridian Books, 2 vols.(The major study of the historical origins of Spinoza's thought; emphasizes the Medieval connections).

Split brains

Severing the direct neural connections between the two cerebral hemispheres produces a 'split brain'. Does it also multiply minds? The most extensive tests of the psychological results of this operation were conducted by Roger Sperry and his colleagues. He concluded that split-brain patients have 'Two separate spheres of conscious awareness, two separate conscious entities or minds, running in parallel in the same cranium, each with its own sensations, cognitive processes, learning processes, memories and so on'. Sperry's view faces both conservative and radical challenges. The conservative challenge is that the results of the tests do not imply that split-brain patients have two minds and are two persons. The radical challenge is that the operation does not multiply minds but, instead, reveals a startling fact: human beings with intact commissures already have two spheres of consciousness, house two minds, and are two persons.

For the purposes of this entry, a split-brain patient is one who has had a *complete* forebrain commissurotomy. In this operation (which has been replaced by less radical procedures), the *corpus callosum* and the other neural links ('commissures') between the two cerebral hemispheres were completely severed.

Patients underwent the operation for the relief of otherwise uncontrollable epilepsy, and it was considered a therapeutic success. Epileptic attacks disappeared, became less frequent or were confined to one hemisphere. Once the patients recovered from the operation, they were able to resume their normal lives; people who knew these patients before the operation would not notice any dramatic changes in their personality, intellect or everyday behaviour.

Observation under controlled conditions discloses a different picture. When input is limited to one cerebral hemisphere and response demanded of it, the behaviour of split-brain patients is decidedly abnormal, as the following simple example illustrates. 'Key ring' is flashed on a screen for a tenth of a second so that 'key' appears in the left visual field and 'ring' in the right. If split-brain subjects are asked to *say* what they saw, they respond that they saw 'ring' and show no sign of seeing 'key'. But, if they are asked, instead, to *retrieve* with their left hands the object named on the screen from an array of items concealed from sight, they will pick out a key while rejecting a ring. Asked to point with the left hand to the object named on the screen, they point to a key or a picture of a key and not to a ring or a picture of a ring. If they are allowed to use both hands to pick out the object named from an array of items hidden from sight, their left and right hands work independently, the right settling on a ring while rejecting a key and the left doing the opposite. Someone seems to have seen 'key'. Someone else seems to have seen 'ring'. No one seems to have seen 'key ring'. With suitable controls, input from the other sensory modalities, except taste, can also be confined exclusively to one hemisphere. When a response depends upon it, split-brain patients behave in similar abnormal ways.

The standard explanation of such behaviour is roughly as follows. The structure of the visual system assures that the left half of the field of vision is conveyed to the right hemisphere and vice versa. Normally, information about the contralateral visual field is supplied to each hemisphere by neural communication across the commissures and by subsequent eye movement. Since the commissures of split-brain patients are severed and the short exposure time serves as a control for eye movement, their right hemispheres see only the word 'key' and their left only the word 'ring'. In most people, speech production is localized in the left hemisphere; and so the oral response to the question reports only what the left hemisphere saw: the word 'ring'. The left hand is primarily controlled by the right hemisphere; so it retrieves the object the right hemisphere saw named - a key - and points to a key or a picture of a key. (Notice that this explanation presupposes speech comprehension in the mute right hemisphere.) Similarly, the right hand is primarily controlled by the left hemisphere, thus accounting for the independent search of items concealed from sight. The failure to elicit any response suggesting that 'key ring' was seen is that the contents of the visual field available to each hemisphere are not the same and, because of the severing of the commissures and the experimental controls, not communicated to the opposite hemisphere.

Behaviour of the sort illustrated in the 'key ring' example and its explanation fuel a natural, tantalizing line of inference.

(1) *The behaviour shows that the split-brain patient sometimes has a disunified consciousness.* No one has doubted that the behaviour associated with the left hemisphere in the 'key ring' example indicates that the subject

consciously saw ‘ring’. The behaviour associated with the right hemisphere seems to be equally clear and prototypical evidence that the subject has a conscious experience of seeing ‘key’; in fact, it is difficult to see how this can be denied short of a general scepticism about the consciousness of human beings who can comprehend, but not produce, language. So, in the example, the patient has simultaneous conscious experiences of seeing ‘key’ and of seeing ‘ring’, but none of seeing both.
(2) *This disunity of consciousness is a standing condition.* The cause of the disunity of consciousness, behaviourally evident in the ‘key ring’ example, is the severing of the neural connections between the two cerebral hemispheres. These remain severed and their neural functions unduplicated whatever the behaviour of split-brain patients. In the absence of controls to prevent it, the separate spheres of consciousness associated with the left and right hemispheres have highly similar contents. This overlap of content and other factors explain why the split-brain patient’s everyday behaviour does not dramatically display two separate spheres of consciousness.
(3) *Split-brain patients have two minds and are two persons.* Despite the significant differences between the two hemispheres, each sustains a range and complexity of psychological functions, including self-awareness, that is characteristically human. Examinations of hemispherectomy patients and their near functional kin - patients with severe strokes in a single hemisphere - confirm the observations of split-brain patients. Neither hemisphere has any better access to the conscious contents of the other than we do to those of other people; each has as direct access to its own experiences as we do to ours. So split-brain patients have two minds. If an embodied mind of characteristic human complexity is a person, then the split-brain patient is two persons since the patient embodies two of them.
(4) *If split-brain patients have two minds and are two persons, so do human beings with intact commissures.* Even with commissures intact, each hemisphere receives a separate neural representation of ‘key’ and ‘ring’ in conditions like the ‘key ring’ example. Why, then, do we not see the behaviour of a split-brain patient? The usual answer is that neural communication between the hemispheres ensures that the right hemisphere is made aware that the left is seeing ‘ring’ and, perhaps, brought to have such an experience itself and vice versa for the right hemisphere. But, then, we have a duplicate of the split-brain case. Communication between the two hemispheres provides a behavioural mask of two independent streams of consciousness - two minds - and two persons just as the duplication of content in everyday circumstances does in the split-brain patients. So, starting from incontestable neurological and behavioural facts about split-brain patients, one apparently arrives at the paradoxical conclusion that we are small collectives of two minds and two persons (see Mind, bundle theory of; Personal identity).

The line of inference just sketched can be used to define philosophical positions on split brains. For a variety of reasons, conservative challenges to Sperry (see above) hold that it should stop short of its third step. Eccles (1970) once claimed that there are no conscious mental phenomena known to be associated with the nonverbal right hemisphere and, later, that whatever conscious processes might be associated with it are subhuman in character. Others have argued that the disunity in consciousness split-brain patients sometimes exhibit is not a standing condition and does not imply two minds; yet others that, although split-brain patients have two minds, they constitute a single person because a single control structure governs both. Sperry (1968) has consistently held the middle ground, endorsing the line of inference as far as its third step. He has refused to join Bogen (1985) and Puccetti (1973) in taking the radical fourth step because, he claims, intact cerebral commissures are the physical basis of unity of consciousness and mind. Many other positions have been taken besides those mentioned. Thomas Nagel’s is particularly striking. He claims that there is no answer to the question of how many minds or persons split-brain patients contain, and that this shows that our ordinary concept of the unity of a person ‘may resist the sort of coordination with the understanding of humans as physical systems, that would be necessary to anything describable as an understanding of the physical basis of mind’ (1971).

The wide diversity of opinion has several sources. The data are unexpected and sometimes messy. Any attempt to deal with them faces a special version of the mind-body problem. One must decide, on some principled grounds, the relation of various anatomic, neurological and behavioural data to mentalistic descriptions; and the proper account of the mentalistic notions of prime concern - consciousness, mind and person - is, to put it mildly, controversial. We also still lack a detailed understanding of how brain structures are responsible for the psychological distinctions involved, for example, what specific role the *corpus callosum* plays, what a control structure is, and how to count centres of consciousness.

Besides raising the philosophical issues discussed above, research on split brains has provided much insight into

problems of deep physiological and psychological interest, for example, hemispheric specialization, which are not immediately tied to them. It has also provided a launching pad for a variety of 'thought experiments' in philosophical discussions of personal identity.

See also: Mind, bundle theory of; Consciousness; Personal identity

CHARLES MARKS

References and further reading

With the single exception noted, the individual works listed are understandable without any esoteric technical knowledge.

Benson, D.F. and Zaidel, E. (eds) (1985) *The Dual Brain: Hemispheric Specialization in Humans*, New York: Guilford.(A representative sample of the psychological and neuroscientific work inspired, in part, by the split-brain phenomena; some of the articles require some advanced technical knowledge.)

Bogen, J.E. (1985) 'The Dual Brain: Some Historical and Methodological Aspects', in D.F. Benson and E. Zaidel (eds) *The Dual Brain: Hemispheric Specialization in Humans*, New York: Guilford, 27-43.(History of thought on the two cerebral hemispheres and partial defence of his 'two persons' account of uncommissurotomized human beings; good bibliography. Some technical knowledge necessary.)

Eccles, J.C. (1970) 'The Brain and the Unity of Conscious Experience', in *Facing Reality: Philosophical Adventures of a Brain Scientist*, New York: Springer, 63-84.(Argument that right hemisphere processes are not conscious.)

Gillett, G. (1986) 'Brain Bisection and Personal Identity', *Mind* 95: 224-9.(Doubts about whether the right hemisphere is a conscious rational thinker.)

McKay, D.M. (1966) 'Cerebral Organization and the Conscious Control of Actions', in J.C. Eccles (ed.) *Brain and Conscious Experience*, Heidelberg: Springer, 422-44.(Argument that there is a single control structure for the operations of the disconnected hemispheres.)

Marks, C.E. (1981) *Commissurotomy, Consciousness, and Unity of Mind*, Cambridge, MA: MIT Press.(Defence of a conservative interpretation of the split-brain phenomena; extensive bibliography.)

Nagel, T. (1971) 'Brain Bisection and the Unity of Consciousness', *Synthèse* 22: 396-413; repr. in *Mortal Questions*, Cambridge: Cambridge University Press, 1979, 147-64.(Argument that the split-brain data present an obstacle to understanding the physical basis of mind.)

Puccetti, R. (1973) 'Brain Bisection and Personal Identity', *British Journal for the Philosophy of Science* 42: 339-55.(Argument that the split-brain data lead to the view that uncommissurotomized human beings have two streams of consciousness and are two persons.)

Puccetti, R. (1981) 'The Case for Mental Duality: Evidence from Split-Brain Data and Other Considerations', *Behavioral and Brain Sciences* 4: 93-123.(Amplification of the viewpoint of Puccetti (1973); with extensive commentary, the author's responses and a large bibliography.)

Sperry, R.W. (1968) 'Hemisphere Deconnection and Unity in Conscious Awareness', *American Psychologist* 23: 723-33.(Early presentation of the view that split-brain patients, but not those with unsplit brains, have two spheres of consciousness and two minds. Good non-technical introduction to the empirical data by the leading researcher.)

Sperry, R.W. (1977a) 'Consciousness, Personal Identity, and the Divided Brain', repr. with editorial updating of references, in D.F. Benson and E. Zaidel (eds) *The Dual Brain: Hemispheric Specialization in Humans*, New York: Guildford, 1985, 11-26.(Amplification and defence of the viewpoint of the previous article; excellent bibliography. Good introduction to empirical research.)

Sperry, R.W. (1977b) 'Forebrain Commissurotomy and Conscious Awareness', repr. in C. Travarthen (ed.) *Brain Circuits and Functions of the Mind: Essays in Honor of Roger W. Sperry*, Cambridge: Cambridge University Press, 1990, 371-88.(Also an amplification and defence of the viewpoint of Sperry's 1968 article, including his most careful consideration of rival views; excellent bibliography.)

Travarthen, C. (ed.) (1990) *Brain Circuits and Functions of the Mind: Essays in Honor of Roger W. Sperry*, Cambridge: Cambridge University Press.(A representative sample of the psychological and neuroscientific work inspired, in part, by the split-brain phenomena; some of the articles require some advanced technical knowledge.)

Wilkes, K.V. (1988) *Real People: Personal Identity without Thought Experiments*, Oxford: Oxford University Press, 132-67.(Expanded version of her earlier conservative assessment of the split-brain phenomena.)

Sport and ethics

Ethical controversies have formed some of the liveliest debate in the philosophy of sport. Some of the issues arise out of the very nature of sport as a rule-governed activity, especially since the breaking of those rules often presents opportunities for competitive advantage. Other debates concern over-emphasis on winning, which can lead to various forms of cheating. The 'problem of winning' is clearly related to the larger problem of competition itself, which has led to lively dialogue over whether competition in athletics inevitably causes alienation. Finally, one of the most provocative controversies has been over the question of performance-enhancing drugs and whether they should be banned.

1 Sport as rule-governed activity

The French writer, Albert Camus, once wrote that it was from his youthful involvement in sport that he 'learned everything he knew about ethics'. Whatever Camus meant by this cryptic remark, it is clear that ethical issues in sport have become one of the central concerns of the philosophy of sport. In part, this may be because in sport our ethical commitments are not merely articulated in speech; we *see* an athlete's ethical standards exhibited in the way they play.

Although some sporting activities such as jogging or recreational skiing, would seem to have few explicit rules, most are rule-governed to one extent or another. Adherence to these rules becomes an immediate ethical issue, especially since often an apparent advantage can be gained by cheating. Many sport philosophers have argued that since playing a given game can virtually be defined as adhering to its rules ('playing basketball' means conforming to the rules of basketball), someone who breaks those rules is not really playing that game at all.

This is complicated by the fact that in many sports certain instances of breaking the rules are so common that they are regarded as 'part of the game'. Consider personal fouls in basketball, offsides in football or soccer, or hitting a batter in baseball. Would we want to say that a soccer player who moves offside is not playing the game? This raises the yet further issue of certain strategies that involve unconcealed breaking of the rules. Consider, for example, 'taking a foul' in basketball when an opposing player has an easy shot. Is this cheating or smart strategy? If cheating, is the fouler still playing basketball? These and related issues are given perhaps the most thorough discussion in Warren Fraleigh (1984).

2 The problem of winning

Since all competitive sports entail that one person or team will win, the other lose, how could there be a 'problem' with winning? The obvious but troublesome answer is that the competitive drive to win has led many to take illicit steps in that pursuit. The 'overemphasis on winning' has been cited as the cause of everything from cheating within the game to unfair or even illicit steps to prepare for winning, such as recruiting violations, accepting unqualified athletes into college and the use of performance-enhancing drugs.

Public debate on this issue has tended towards two simplistic responses, encapsulated in the well-known maxims, 'Winning isn't everything, it's the only thing', and its contrary, 'It's not whether you win or lose but how you play the game'. The first assumes that there are no values or goals in competitive sport other than winning, which is clearly false. Though winning is *a* goal of every game, so also are such things as getting a good workout, having fun, and developing a sense of teamwork. On the other hand, no committed athlete will be satisfied with the second maxim, which implies that winning simply does not matter. Of course it does; every athlete feels exhilaration at victory and disappointment at defeat. The real issue is to recognize winning as *one* of the values in sport, but one among others, and to put it in proper perspective alongside the other legitimate values of participation. How to do that, of course, is a complicated issue.

3 Competition

The overemphasis on winning discussed above is evidently a function of the competitive nature of sport and the abuses competition can generate. Not surprisingly, there has been much criticism of competitive sport on these grounds, deriving often from Marxian criticisms of competition in capitalism. A central contention here is that competition, whether in the economic sphere or in sport, *causes* alienation in its various guises (see Market, ethics

of §3). No one can deny that too often in sport the competitive situation devolves into alienation. Every athlete has experienced alienation in sport, and it is an easy inference that such alienation is a function of the competition central to sporting games. On the other hand, not all competitive sporting events involve alienation; indeed, the vast majority of athletic games occur among friends and are occasions of friendship. So if many or even most athletic contests are friendly, it is hardly obvious that competition *per se* causes alienation.

Part of the problem may be that we tend to think of the paradigms of sport in terms of *professional* sport, where the aspirations of the players are complicated (some would say polluted) by financial factors. Although it is neither the case that all amateur sporting events are friendly, nor all professional ones alienated, it is nevertheless plausible - and compatible with the Marxian analysis - that once sport becomes an economic issue the likelihood of alienation is increased.

One way to think about this is to consider a teleological account of competitive sport. What is sport like when it ‘really works’, when it is the best that it can be? It has been argued that, at its best, sport is indeed an occasion of friendship, that competitive sport can offer a particularly intense encounter among friendly participants who wish to challenge each other to be the best that they can be. On this view, when competitive events do degenerate into alienation, that should be construed as a defective mode of sport, certainly not as ‘the way competition is’ or the way sport has to be. That should lead in turn to efforts to determine what causes competitive sport to devolve into alienation and how to avoid those conditions.

4 Performance-enhancing drugs

One of the major ethical controversies in sport philosophy involves the use by athletes of so-called performance-enhancing drugs, the best known of which are probably anabolic steroids. The claim and the belief of many athletes is that high doses of these drugs coupled with strenuous training will build significant muscle mass and therefore enhance the athlete’s abilities. There is some evidence, however, that such drugs, particularly in the doses employed by some athletes, are harmful to one’s physical and mental health, resulting in everything from acne and danger of cancer to pathological aggressiveness. This has led most sports organizations to ban the use of performance-enhancing drugs.

But should they be banned? This has become the subject of considerable controversy. Most writers have argued that they ought to be banned on several grounds. The main objections to their use have been danger to health, the issue of fairness, and the problem of coercion.

The danger-to-health argument hinges on the validity of the data suggesting that these drugs are indeed dangerous to health, and a certain commitment to paternalism, arguing that since such drugs are dangerous, athletes should be ‘protected from themselves’ by having the drugs banned (see Paternalism). The fairness argument turns on the clear sense that those athletes who do choose to use performance-enhancers, even at the risk of their health, will have an unfair advantage over those presumably more prudent athletes who choose not to use them. This leads in turn to the third argument, that athletes who wish to excel will be unfairly coerced into using the drugs if they want to compete with drug users.

All these arguments have been contested in a set of important and provocative articles by W.M. Brown (1980, 1984). Brown begins by pointing out the difficulty of determining nonarbitrarily just what a performance-enhancing drug is. What about aspirin for pains before a game? What about caffeine for stimulation? Do they not enhance performance? With regard to the presumed danger to health, Brown argues that even if it is shown conclusively that, say, anabolic steroids do constitute a health danger (in fact a disputed point), the very activities of most sports involve dangers to health much more severe than the putative dangers of performance-enhancers: various injuries and even, as in the case of sports such as rock climbing or automobile racing, risk of death. If ‘danger to health’ is a criterion for banning, it appears most sports should be banned.

Turning to the fairness argument, Brown rejoins that there are all sorts of unfairness inherent in most sports. Is it not ‘unfair’ to me as a normal sized male if my opponent in basketball is seven feet tall? Is it not ‘unfair’ if my cross-country ski opponent grew up in snowy Vermont while I grew up in Florida? Is it not ‘unfair’ of my opponent to practice for long hours or lift weights to gain an advantage when I do not? How, in short, do we separate nonarbitrarily the unfairness of performance-enhancing drugs from the myriad other unfairnesses that competitive sport entails?

Regarding the coercion argument, Brown retorts that the use of the term ‘coercion’ here is overblown. ‘No one forces athletes to seek Olympic gold,’ he says; the pressures on athletes to take performance-enhancing drugs are no different in principle from all the other pressures - and difficult decisions - that top level athletes face all the time. The debate continues.

See also: Sport, philosophy of

DREW A. HYLAND

References and further reading

Brown, W.M. (1980) ‘Ethics, Drugs, and Sport’, *Journal of the Philosophy of Sport* 7: 15-23.(The provocative article that began the controversy over performance-enhancing drugs, arguing that the decision should be the individual athlete’s.)

Brown, W.M. (1984) ‘Paternalism, Drugs, and the Nature of Sports’, *Journal of the Philosophy of Sport*11: 14-22. (A follow-up article in which Brown answers his critics and modifies his position slightly.)

Fraleigh, W. (1984) *Right Actions in Sport: Ethics for Contestants*, Champaign, IL: Human Kinetics Publishers. (A comprehensive look at problems in sport ethics)

Hyland, D. (1990) ‘Ethical Issues in Sport’, in *Philosophy of Sport*, New York: Paragon House, ch. 2.(Treats the issues discussed above in greater detail, including the problem of winning, competition and alienation, and the debate over performance-enhancing drugs.)

Various, *Journal of the Philosophy of Sport*.(The leading journal in the field, containing most of the important articles in sport ethics. See especially volume 11 for a symposium on performance-enhancing drugs.)

Sport, philosophy of

The philosophy of sport as a separate area of philosophy is largely a phenomenon of the second half of the twentieth century, although previous philosophers, back to the ancient Greeks, occasionally made reference to sport or used it as an example in a larger point. Within the philosophy of sport, a number of sub-areas have emerged as important: sport and ethics, questions concerning sport and society, the issue of self-knowledge in sport, the mind-body problem as it relates to sport, sport and art, and the controversy over the possibility of defining certain key terms within sport, such as sport, game, play and athletics.

Within the area of sport and society, several major debates have arisen. The first is about whether sport teaches values and, if so, whether the values taught are desirable or not. Second, considerable attention has been paid to how certain societal problems, such as sexism and racism, have manifested themselves in sport, and how they might be addressed within sport. Third, attention has been directed to the phenomenon of the athlete as cultural hero.

The relevance to sport of the age-old philosophical issue of self-knowledge is manifest, with a number of different construals of what counts as self-knowledge emerging as important. Attention has been paid to self-knowledge in the psychological sense, self-knowledge as manifested in Zen thought and self-knowledge in the Socratic sense, among others. The mind-body problem, also an old philosophic issue, has clear relevance within the domain of sport. Dualism, physicalism and phenomenological accounts have all been represented, the latter being the most dominant and persuasive so far.

1 Introduction: historical background

The history of the philosophic interest in sport in one sense goes back to the ancient Greeks, but in another sense is a phenomenon of the middle and late twentieth century. Greek culture in general placed great emphasis on what they called *gymnastikē*, the training of the body, making it, along with education in the arts (*mousikē*), one of the two foundations of education. Plato, in Books III-V of his *Republic*, reflects this interest by making education in gymnastics a core aspect of the education of those who will become the famous ‘philosopher-rulers’, and in his *Laws* (803c), the Athenian Stranger praises play as the very highest of human activities. In addition, Aristotle sometimes uses examples from athletics to make important points. Nevertheless, no Platonic dialogue, Aristotelian treatise or other philosophic work in ancient Greece focused explicitly on sport as a topic worthy of philosophic reflection in its own right.

This situation remained almost until the middle of the twentieth century. Indeed, for long periods of time, there was very little mention of sport or athletics by philosophers at all. In the late eighteenth and the nineteenth centuries, philosophers such as Schiller and Nietzsche praised play as a high human activity, and Nietzsche made it a guiding metaphor for his notion of the purposeless ‘happening’ of the will to power. Following him in the twentieth century, continental thinkers such as Heidegger, Gadamer, Foucault and Derrida used play as a metaphor for their guiding notions, and Wittgenstein used the concept of ‘language games’ as a core notion in his philosophy, but none of them treat sport as a philosophical theme in itself.

In 1938, however, the historian Johan Huizinga published his book *Homo Ludens: A Study of the Play Element in Culture* in which he argued the thesis that ‘civilization arises and unfolds as play’. Part of the importance of this book was that it legitimized play and sport as respectable topics for intellectual reflection and research. A few decades later, the US philosopher Paul Weiss published his *Sport: A Philosophic Inquiry* (1969), after which sport became a legitimate and identifiable area of philosophic investigation.

We shall discuss a number of the most important topics and issues in the philosophy of sport. Some of the most important topics not treated below at least deserve mention. We shall not here address the crucial connection between sport and ethics (see Sport and ethics). There is also a lively discussion in the philosophical literature concerning the definition of key terms in philosophy of sport. Can terms such as game, play, sport and athletics be defined? Or, in a Wittgensteinian mode, can we at best find a number of ‘family resemblances’ in and among these terms? Probably the most important single contribution to this debate has been Bernard Suits’ 1978 work *The Grasshopper: Games, Life, and Utopia*, a book-length study devoted to the thesis that ‘game’ can indeed be

defined. There is also an active discussion regarding the connection of sport and art, ranging from thinkers who virtually identify the two to those whose interest focuses on the aesthetic element in sport.

2 Sport and society

Huizinga established beyond doubt that sport is essentially a cultural phenomenon. Not surprisingly, many striking philosophical issues concerning sport centre upon the relation between sport and society. One crucial issue of controversy concerns the relation of sport and the teaching or inculcating of 'values'. Nearly everyone, from the most fanatical devotees of sport to its harshest critics, agrees that sport teaches values. The disagreement is over what values sport teaches. The litany that the partisans of sport offer is well known: sport teaches such virtues as self-discipline, teamwork, being a gracious winner and a fair loser, fair play and courage. The critics of sport are no less sure that sport teaches values: mindless obedience to authority, the desire to hurt or humiliate others, dog-eat-dog competitiveness, cheating and a willingness to ruin one's body in the name of winning a game.

A look at the empirical evidence attests that both are sometimes right. Depending on the situation, on the coaches and the players involved, sport can probably help inculcate the most desirable or the most deplorable qualities. The issue, much more complex, is how to structure our athletic situations so that the desirable qualities are encouraged and the deplorable ones resisted.

It is usually assumed that the values inculcated, whether positive or negative, derive originally from society. This raises the subsidiary and subtle question of whether there are any values or qualities genuinely peculiar to sport that could be productively redirected into the larger culture. Nevertheless, most of the values, it is assumed, come from society into sport. Two of the most problematic in this regard are racism and sexism, about which there is a lively literature in sport philosophy.

No one argues that sport is inherently racist. The racism that has long infected sport, it is plausibly assumed, is a reflection of the racism in society where it originates. And indeed, the history of racism in sport has closely paralleled that of society, from formal segregation early on, to the introduction of 'token' numbers of blacks in college and professional sports teams, to various 'quotas', explicit or implicit. In recent years, as blacks have come to dominate many of the most popular US sports, the issue of racism has taken on a different form. One of the early 'justifications' often offered for the segregation of blacks into their own leagues was that it was 'for their own good', that they were naturally inferior athletes who would only be humiliated by a direct comparison in competition with whites. As blacks were gradually integrated into sporting life, and especially as they began to be remarkably successful, the argument quickly reversed itself, but ironically, in order to serve the same racist ends. Blacks, it was suggested, are 'naturally superior athletes', only a few generations away from the primitive conditions of the jungle. The easy inference drawn from this claim is that because blacks' superior athletic ability is 'natural', it is not a consequence of the hard work, discipline, and perseverance attributed to whites who exhibit athletic ability. Whites get good by exhibiting hard work and other virtues. Blacks just do it 'naturally'. A second consequence often drawn was that the natural physical superiority of black athletes was matched by a corresponding intellectual superiority of whites. White players are 'smart', 'heady', 'know the game'. Blacks, by implication, lack the intelligence of white athletes.

Still, something calls for explanation here. Blacks have been enormously successful at athletics in proportion to their representation in society at large. Two basic explanations have been offered, one much more controversial than the other. The first, less controversial one emphasizes environmental factors. Perhaps the generally greater poverty of blacks has made them tend to be more competitive. Perhaps their exclusion from more common pursuits (business careers, law, medicine and so on) leads the most talented of them to focus their youthful energies on athletics. All such explanations make the athletic success of blacks not a function of inherent or natural differences but of social environment. And social environment can be changed.

The second, more controversial explanation argues that there are genetic or at least physiological differences between blacks and whites that explain blacks' athletic prowess. Claims have been made that blacks tend to have more fast-twitch muscles, that their muscle-to-fat ratio is superior to that of whites, or that their physiology tends towards longer limbs that enhance athletic ability. These claims are controversial on two grounds. First, the empirical evidence is highly disputed. Many of the 'experiments' conducted are replete with uncontrolled variables that make the results highly suspect. It is even doubted by many biologists whether the very concept of

‘race’ is a viable biological category. Second, claims regarding ‘natural’ differences are controversial in that claims of just this sort have long been used in the service of racism; blacks are ‘just different’, we have long been told, and ‘different’ too often implies ‘inferior’ (see Race, theories of).

The case of women’s participation in sport has in some ways paralleled the case of blacks. For all too long, women were virtually excluded from participation. Gradually, especially with the passage in the 1970s in the US of Title IX legislation demanding equal opportunity in sports for women, they began to have more opportunities to play sports and, accordingly, began to play in much greater numbers. Still, the perception remains widespread that because of physiological differences affecting speed, size and strength, women are inevitably inferior athletes to men.

Again the ‘nature versus nurture’ argument becomes germane. Is the gap both in participation and relative abilities largely explicable by appeal to the greater opportunities given to young male athletes and the greater encouragement given to them to compete athletically? Or do the physiological differences remain fundamental, and is the statistically significant extent to which men are stronger, faster and bigger than women a function of natural physiological differences?

Even if the latter argument is accepted, an important qualification must be admitted. The physiological differences may be natural, but the sports are invented. And most sports have been invented by men, in ways that highlight precisely those qualities - speed, size, strength - typically characteristic of male physiology. Suppose we used other criteria, perhaps balance and grace, as the criteria for determining athletic superiority. Suppose, for example, we used as a test of athletic ability one of the very few sports designed explicitly to highlight qualities exhibited paradigmatically by women: the balance beam in gymnastics. If we did, we would discover that women were by far superior ‘natural athletes’ to men (see Feminism §5).

A final issue regarding sports and society that has caused considerable controversy concerns the widespread tendency in contemporary culture to raise outstanding athletes to the status of cultural heroes. A culture’s heroes are often informative of the culture’s needs and values. What does the widespread elevation of athletes to the status of cultural heroes suggest about contemporary culture? It might suggest a nostalgia for the past. Physical prowess such as that exhibited by athletes, after all, is increasingly unnecessary in a culture dominated by technological developments that render physical strength, speed or size largely irrelevant. Or, to put a different slant on the same point, is our worship of athletes as heroes part of an effort to preserve the value of a set of characteristics - physical abilities exhibited by athletes - that we recognize are in danger of being lost? Or, finally, is the phenomenon of athletes as heroes an indication of a higher culture, one that raises to high status a set of activities of relatively little practical worth precisely because they are not merely practical? In any case, the elevation to hero status of athletes, especially since those outstanding athletes do not always exhibit the personal and moral qualities we desire, remains a controversial question in the philosophy of sport.

3 Sport and self-knowledge

‘Know thyself’ has been a core goal of philosophy since before Socrates. A claim is often made that athletic involvement, at least of the right kind, can lead to genuine self-knowledge. But what kind of self-knowledge? Several different senses of self-knowledge have received attention in the philosophy of sport. First is self-knowledge in a roughly psychoanalytic sense. The psychiatrist Arnold Beisser, in his book *The Madness in Sports* (1977), analyses troubled athletes and relates their choice of sport or their mode of involvement in sports to their respective pathologies. But implicit in his analysis is the inference that any athlete, pathologically troubled or not, can gain this sort of self-knowledge through reflection on their involvement in sport. A number of psychological categories come quickly to mind as clearly applicable.

Consider, for example, the social significance of one’s choice of sport. Suppose one is from an impoverished social background, and one’s favourite sports are golf, polo and squash. Does that suggest a desperate attempt to deny one’s social status, or a healthy refusal to be limited by it? Or suppose one is from a wealthy background and one chooses the same sports. Does that suggest an inability to escape from the social expectations of one’s class or a healthy acceptance of it? Note in each case that the issue itself does not supply the answer. Rather, it enables one to formulate the potentially fruitful question, which must be answered in each case individually.

There are numerous other categories of analysis that might prove fruitful. Consider the issue of team sports versus

individual sports. What does it suggest about oneself if one's preference is predominantly for one or the other? Or again, take the difference between competitive and non-competitive sports. Many people love to jog, swim or ski, but would never consider doing so competitively. Others get bored unless the activity can be transformed into a competitive game. Either way, reflection on one's choices might be psychologically revealing. Other sets of contrasts can be no less revealing: contact or non-contact sports; coeducational or single-sex sports; spectator sports or non-spectator sports and so on.

One's mode of involvement in sports can also be psychologically revealing. Certain people are notoriously poor losers or gracious winners, who 'come through in the clutch' or, to the contrary, 'fold' in key situations. Others are peculiarly injury-prone, or play well only when alienated from the opposition. All such phenomena can be occasions for self-knowledge.

But self-knowledge in the psychological sense is hardly the only sense of self-knowledge available in sport. A second construal that has gained considerable popularity in sport philosophy involves a convergence of the notion of 'peak experiences' popularized by Irving Mazlow and certain aspects of Zen Buddhist theory. The notion of a peak experience, originally conceived in terms of certain profound experiences such as love, friendship, or religious revelation, found a receptive audience among talented athletes. After long periods of preparation, these occasionally experience moments when a technically difficult activity, say shooting a basketball, hitting a baseball or skiing a steep slope, seems suddenly to happen easily, smoothly and without effort or analytic thought. Such experiences seem related to aspects of Zen Buddhist experience, especially as that has been applied to athletic activity. The key figure in the popularization of this notion is Eugen Herrigel, whose book, *Zen in the Art of Archery* (1953), set the stage for a host of other books applying the themes of Zen Buddhism to other sports. The key notion here seems to be that the 'self' in the sense of the ego or egoistic self must be 'let go', left behind in favour of a more profound experience of immersion that no longer separates the egoistic self from the experience in which it is involved.

Finally, considerable attention has been paid to self-knowledge in roughly the Socratic sense of 'knowing what I know and what I do not know'. In the sporting domain, this has been translated into something like 'knowing what I can do and cannot do', and this in turn has two dimensions. First, this sense of self-knowledge entails recognizing one's limits and staying within them, 'playing within oneself', as coaches sometimes exhort. But second, it also entails taking oneself to the limits of one's ability, doing everything that one can do. To the extent that an athlete fulfils both of these goals implicit in the Socratic sense of self-knowledge, they will be likely to attain the peak of athletic prowess.

4 Mind and body in sport

Most sporting activity is a consummate example of the conjoining of mental and physical activity. As such, it is a rich testing ground for one of the oldest of philosophical problems, the connection of physical and mental activity or, as it is sometimes expressed, 'the mind-body problem'. Three basic views have received some attention in the philosophy of sport literature. The first is the age-old position of the so-called dualism of mind and body. This view has mostly been criticized in philosophy of sport, often because the notion of somehow being two substantial entities (soul and body or mind and body) that mysteriously interact together is so far from the actual experience of most athletes. Second is the materialist or physicalist view, which has also received relatively little support among philosophers of sport, in part because of the standard reductionist criticisms, but probably more because it does not succeed, at least so far, in taking adequate account of the many experiential factors in sport. For this reason, the dominant and most well-received view has been the phenomenological approach that attempts to give an adequate description of the structures of the athletic experience *as experienced*, which is the main interest of most athletes and sport philosophers.

DREW A. HYLAND

References and further reading

Beisser, A. (1977) *The Madness in Sports*, Bowie, MD: Charles Press.(Referred to in §3. A psychiatrist relates his patients' pathologies to their sport involvement in ways that can be extended beyond the specific cases.)

Herrigel, E. (1953) *Zen in the Art of Archery*, New York: Vintage Books.(Referred to in §3. This is the first and best of those books that relate Zen teachings to sport experience.)

Huizinga, J. (1938) *Homo Ludens: A Study of the Play Element in Culture*, Boston, MA: Beacon Press, 1950. (Referred to in §1. The pioneering book that established the study of sport as a legitimate and important intellectual enterprise.)

Hyland, D. (1990) *Philosophy of Sport*, New York: Paragon House.(An introduction to the main issues. Chapter 5 outlines the debate about sport and aesthetics, not discussed above.)

Journal of the Philosophy of Sport, Champaign, IL: Human Kinetics Publishers.(The leading journal in the field, and an invaluable source of major articles on all aspects of the philosophy of sport.)

Morgan, W. and Meier, K.V. (1988) *Philosophic Inquiry in Sport*, Champaign, IL: Human Kinetics Publishers. (A large compendium of some of the best-known articles on all aspects of the philosophy of sport.)

Suits, B. (1978) *The Grasshopper: Games, Life and Utopia*, Toronto, Ont.: University of Toronto Press.(Referred to in §1. The main book in the controversy over whether key terms such as ‘game’ can be defined - he argues in the affirmative.)

Weiss, P. (1969) *Sport: A Philosophic Inquiry*, Carbondale, IL: Southern Illinois University Press.(Referred to in §1. The first book by a major philosopher devoted explicitly to the philosophy of sport. Both widely praised and criticized, the book did much to get philosophy of sport established as a discipline.)

Staël-Holstein, Anne-Louise-Germaine, Mme de (1766-1817)

Staël was a French woman of letters, whose corpus includes novels, plays, memoirs, criticism and works of historical and sociological observation. As a novelist and critic, Staël is perhaps best known as a forerunner of the Romantic movement in French literature, and for her typology of European culture. However, her extensive body of work also included a perfectibilist philosophy of history; an account of German character and literature containing an influential survey of contemporary philosophy; and a liberal defence of the French Revolution, in which Staël had been an active participant.

Staël was born in Paris on 22 April 1766, the daughter of Jacques Necker, the Genevan chief minister of Louis XVI. She supplemented an emotionally unrewarding marriage to Baron de Staël-Holstein, the Swedish ambassador in Paris, with a number of (often distinguished) lovers, and, in 1816, secretly married a Genevan soldier, John Rocca. Her salon became an important centre of intellectual and political activity, and her extensive entourage included the liberal writer and politician, Benjamin Constant; the economist and historian, Sismonde de Sismondi; and the scholar and translator, August Wilhelm von Schlegel. She welcomed the French Revolution, despite feelings of loyalty to the fallen monarchy and an attachment to a tradition of aristocratic liberty. A moderate republican emphasizing respect for the rule of law, Staël retreated to Coppet (the family château on Lake Geneva) ahead of the September massacres (1792). She travelled to England, returning to France in 1795. She supported the Directory and initially welcomed the Consulate, as instruments of social and political stability. However, her growing opposition to Napoleon eventually resulted in exile. From Coppet, she visited Germany and Italy, before, in 1812, fleeing Napoleonic Europe and making her way (through Austria, Russia, Finland and Sweden) to England. After the Bourbon Restoration, to which she became reconciled, Staël, ravaged by opium, returned to Paris, where she died on 14 July 1817.

Staël's nonfiction writings include studies of Rousseau, suicide, and the role of the passions, but her most important works are *De la littérature considérée dans ses rapports avec les institutions sociales (On Literature Considered in its Relation to Social Institutions)* (1800), *De l'Allemagne (On Germany)* (1810) and *Considérations sur les principaux événments de la Révolution (Considerations on the Principal Events of the French Revolution)* (1818).

In *De la littérature*, Staël propounded the claim that literature (expansively defined) reflects the historical influence of social institutions, laws and mores. The book includes digressions on a range of political and philosophical topics, as well as tracing the development of literature from the ancient world to the present. Staël advanced a perfectibilist view of historical development as exemplifying the progressive development of human faculties (promoting the medieval over the ancient world), and suggestively contrasted two 'completely distinct' European cultures, the light and elegant literature of the 'South' and the meditative and melancholy literature of the 'North'. This distinction was later identified with that between the 'classical' and 'romantic' (associated with France and Germany respectively), terms also connoting 'two eras in world history' (before and after the establishment of Christianity).

De l'Allemagne was banned by Napoleon and only properly published (in revised form and in England) in 1813. Surveying geography, customs, literature and religion, *De l'Allemagne* sought to introduce and interpret a foreign culture to a French audience, and undermine complacency about French cultural superiority. As part of this portrait 'of the German character and the distinctive spirit of their literature', Staël offered 'a simple and popular notion of their philosophical systems'. Staël provides synopses of Leibniz, Jacobi, Fichte, Friedrich Schlegel and Schelling, as part of a narrative largely structured around her account of the work of Kant. There are also thumbnail sketches of the state of, and connections between, French and English philosophy, and a general attempt to discredit sensationalism for undermining the possibility of morality. Although highly readable, doubts remain about the plausibility of Staël's picture of Germany as the paradigmatically contemplative nation, about the coherence of her Romantic enthusiasms, and about the accuracy and derivativeness of her account of German philosophy.

Considérations is a defence of the French Revolution, which established a tradition of liberal historiography. Staël sought to subvert the conservative myth of the Revolution as a rupture with fourteen centuries of unbroken order, by insisting that it was part of a continuing process (the latest episode in a thousand-year struggle for liberty).

Moreover, she attempted to identify the liberal current that she represented with part of the Revolution (the Constituent Assembly), and thereby deny responsibility for, or sympathy with, its other aspects (such as the Terror). *Considérations* also illustrates Staël's lifelong and determined defence of her father's reputation, and her Anglophilia. For Staël, England provided an enviable example of an enlightened nobility, working within the constitutional framework of bicameralism and the protection of individual rights, sharing political responsibility with a moderate bourgeoisie. England was also a Protestant country, and Staël, who considered Christianity indispensable to modern society, saw liberalism and Protestantism (which she sought to establish as the state religion of France) as linked by the shared assumption of moral equality.

Staël's political thought defended the value of privacy and of the liberty of the moderns (who prefer 'to fulfil their destiny apart from public affairs'). She challenged the identity of virtue and citizenship that underpinned republican models of freedom, arguing that 'liberty in modern times consists of whatever guarantees the independence of citizens from the power of the state', and rejected conceptions of sovereignty as unlimited and indivisible, insisting that political power should be limited and shared among different authorities.

Staël's literary success was considerable, and her intellectual and political importance was widely proclaimed. But although still acknowledged as an important precursor of the Romantic movement in French literature, and notwithstanding some recent feminist interest in her fiction, Staël has fallen into comparative neglect. It remains difficult to claim great philosophical significance for her work, but her accounts of historical development and of German philosophy are not without interest, and her political theory still merits attention.

See also: German idealism; History, philosophy of

DAVID LEOPOLD

List of works

Staël-Holstein, A.-L.-G., Mme de (1820-1) *Oeuvres complètes (Complete Works)*, ed. A. de Staël, Paris: Treuttel & Würtz, 17 vols.(First collected edition of Staël's work, reprinted Geneva, Slatkine Reprints, 1967.)

Staël-Holstein, A.-L.-G., Mme de (1800) *De la littérature considérée dans ses rapports avec les institutions sociales* (On Literature Considered in its Relation to Social Institutions), ed. P. van Tieghem, Geneva: Droz, 1959.(A good modern edition of Staël's text.)

Staël-Holstein, A.-L.-G., Mme de (1810) *De l'Allemagne (On Germany)*, ed. J. de Pange and S. Balayé, Paris: Hachette, 5 vols, 1958-60.(A good modern edition of Staël's text.)

Staël-Holstein, A.-L.-G., Mme de (1818) *Considerations sur le Révolution française* (Considerations on the French Revolution), ed. J. Godechot, Paris: Tallandier, 1983.(A good modern edition of Staël's text.)

Staël-Holstein, A.-L.-G., Mme de (1960-93) *Correspondance générale (Collected Correspondence)*, ed. B. Jasinski, Paris: Pauvert, 6 vols; Hachette, 1982-5; Klincksieck, 1993.(The most complete edition of Staël's correspondence.)

Staël-Holstein, A.-L.-G., Mme de (1987) *Major Writings of Madame de Staël*, trans. and with intro. by V. Folkenflik, New York: Columbia University Press.(Translated excerpts from a wide variety of Staël's works.)

References and further reading

Balayé, S. (1979) *Madame de Staël: Lumières et liberté (Madame de Staël: Enlightenment and Liberty)*, Paris: Klincksieck.(An important work by an eminent Staël scholar.)

Staël-Holstein, A.-L.-G., Mme de (1996) *Madame de Staël et les Français*, the Zaharoff Lecture for 1994-5, Oxford: Clarendon Press.(An engaging account of aspects of Staël's relation to France and French society.)

Besser, G.R. (1994) *Germaine de Staël Revisited*, New York: Twayne.(An introductory account of Staël's major works.)

Gutwirth, M. (1978) *Madame de Staël, Novelist*, Urbana, IL: University of Illinois Press.(A feminist study of Staël's fiction.)

Gwynne, G.E. (1969) *Madame de Staël et la Revolution française: Politique, philosophie, littérature (Madame de Staël and the French Revolution: Politics, Philosophy, Literature)*, Paris: Nizet.(An account of Staël's political thought.)

Herold, J.C. (1959) *Mistress to an Age. A Life of Madame de Staël*, London: Hamish Hamilton.(A biographical introduction.)

Isbell, J.C. (1994) *The Birth of European Romanticism. Truth and Propaganda in Staël's 'De l'Allemagne'*, Cambridge: Cambridge University Press.(A study of Romantic themes in *De l'Allemagne*.)

Stair, James Dalrymple, Viscount (1619-95)

An outstanding lawyer, senior judge, politician, and the founding father of modern Scots Law, Stair is also an interesting, if minor, philosopher of law of the seventeenth century. Stair believed that law is an inherently rational discipline and that its content can be derived from the principles of natural law which are self-evident to all humans. Stair led an active life at the heart of public affairs in seventeenth-century Scotland, finishing up as the chief judge of the supreme civil court. Born in Ayrshire, Scotland, he became a teacher at Glasgow University in 1641, was called to the Bar in 1648, became Judge in the Scots Cromwellian Court 1657, Vice President of the Court of Session 1660, Lord President of the Court of Session (Scotland's most senior judge) 1671, exiled to Holland 1682, and reappointed Lord President in 1689 subsequent to the 'Glorious Revolution'.

Stair was a philosopher and a devout Calvinist as well as politician and judge. His thought reflected all these elements, stressing the role of God as the upholder of the natural law and the need for order in human affairs as expressed through contemporary positive law. Stair never published a purely philosophical work. Instead his legal theory is expounded within the *Institutions of the Law of Scotland* (1681), a book expounding the national law of Scotland, and which presupposes that there is a divine and rational basis for that law, and indeed all man-made law. The *Institutions* continues to be a work of high authority within the modern Scots legal system. Stair rejects the notion, widely held even in the seventeenth century, that human laws lack any inherently rational basis, being rather based on the arbitrary will of humans as expressed through custom, court decisions and legislation. This apparent lack of rationality in the law extended according to its exponents both to the content of the law and to its organization.

Stair took it for granted that God was the ultimate source of human values and law, and his thought is very much in the line of classical natural law thought in Europe as derived from Aquinas (see Aquinas, T. §13) through to that of his near contemporary Hugo Grotius (see Grotius, H.), whose *De iure belli ac pacis (The Law of War and Peace)* (1625) seems to have been a major influence on Stair (see Natural law §1). The *Institutions* differs from the work of Grotius in two completely opposite ways: in its greater emphasis on God as the ultimate source of law and in its concentration on the actual positive law. The structure of Stair's thought has a definite Aristotelian flavour, derived from the strongly Aristotelian syllabus he both studied and taught at Glasgow University, and is organized according to Aristotle's doctrine of causes: formal, efficient and final. Stair argues that the formal cause of law is reason, its efficient cause the actions of God and man, its final cause the promotion of human welfare. This is seen in his definition of law, which is similar to that of Grotius, as 'the dictate of reason determining every rational being to that which is congruous and convenient for the nature and condition thereof' (1681: I.1.1).

Stair is ambiguous as to how we know the natural law, an ambiguity also found in earlier natural law thinkers, in asserting that humans have both the inclination or instinct to follow the natural law and are also led to follow the natural law by the light of reason. This contradiction, also found in earlier natural law theories, Stair seems to have made no effort to resolve or even to be aware of. To the obvious objection that, if this is so, then what need have humans of positive human law, Stair answers that there is no conflict: 'the principles of equity are the efficient cause of rights and laws' (1681: I.1.17) (Stair understands the term equity to be synonymous with the natural law, particularly its role as the measure of positive law; see Justice, equity and law). The three principles of equity are obedience to God, that man is free to act except as specifically limited by God's commands, and the obligation to fulfil one's engagements (contracts). The principles of positive law derived from these principles of natural law are society, property and commerce, a concise trinity of proto-bourgeois values reflective of the developing mercantile economy of the seventeenth century. From these principles in turn are derived the three different types of human rights: *liberty*, the right to life and to personal integrity; *dominium*, the right of ownership of possessions; and *obligation*, the right to freely make contracts and the duty to fulfil those contracts. Stair stresses that it is these three types of rights which are 'the formal and proper object of law'; this is perhaps his most original insight and anticipates one of the central principles of the later legal positivists.

Stair's notion that positive legal rights are the proper object of the law becomes his organizing principle, enabling him to reject the hallowed practice of the civilian tradition, derived from Justinian's *Institutes*, of dividing the law according to the categories of persons, things and actions (see Roman law §4; Gaius; Justinian). Instead, the

Institutions deals first with the constitution and nature of rights, second the conveyance of rights and third the enforcement of rights. The thesis that law can be a ‘rational discipline’ is interestingly tested in this work.

See also: Law, philosophy of

SCOTT C. STYLES

List of works

Stair, J.D. (1681) *Institutions of the Law of Scotland*, 2nd edn, substantially revised, 1693; ed. D.M. Walker Edinburgh: Edinburgh University Press, 1981.(D.M. Walker’s standard modern edition based on the 1693 edition, not on later revisions by eighteenth- and nineteenth-century editors.)

References and further reading

Campbell, A.H. (1954) *The Structure of Stair’s Institutions*, David Murray Lecture, no. 98, Glasgow: Glasgow University Publications.(Influence of Aristotle and Pufendorf in Stair’s thought.)

MacCormick, D.N. (1981) ’Stair As An Analytical Jurist’, in *Stair Tercentenary Studies*, Edinburgh: The Stair Society, 187-99.(Analytical concepts in Stair.)

MacCormick, D.N. (1982) ’Law, Obligation, and Consent: Reflections on Stair and Locke’, in *Legal Right and Social Democracy*, Oxford: Oxford University Press, 60-83.(Contrasts and compares Stair with Locke.)

Stein, P.G. (1981) ’Stair’s General Concepts’, in *Stair Tercentenary Studies*, Edinburgh: The Stair Society, pp. 187-99.(Intellectual influences on Stair.)

State, the

States are inescapable, powerful and fundamentally important in the modern world. They spend a substantial portion of their members' wealth; they tax, confiscate or compulsorily purchase private property; conscript; impose punishments, including capital punishment; defend their members from aggression and protect their rights; and provide educational, health and other essential social services.

States are also central to modern political philosophy, and figure in its main topics. For instance, the various theories of social justice concern which principle or principles of justice should be followed by states. Again, discussions of the rights of individuals, or of groups, presuppose states to make the preferred rights effective. The answers to traditional questions, such as whether one is morally obliged to obey the laws of a state, or whether freedom is reduced by the state or made possible by it, must depend in part on what a state is taken to be.

The principal features of the modern state are basically agreed upon (population, territory, effective and legitimate government, independence). But there are underlying assumptions needing notice, and many questions about the state, especially concerning its proper activities, are controversial and disputed. Moreover, the value of the state can be challenged, and its future doubted, especially in the light of increasing economic and political globalization and moral cosmopolitanism.

1 Main features of the state

'State' is sometimes used in the broad sense to mean any independent political organization, sometimes in the narrow sense, to mean one particular type of political organization - the modern state. The latter is the usage considered here. The modern state emerged gradually as European political institutions changed. The date of its initial recognition by political philosophers is much discussed. Some put it in the sixteenth century (Skinner 1989), others several centuries earlier (Black 1992). Machiavelli, Bodin and Hobbes were prominent among early theorists of the state.

The modern state has been defined in different ways by anthropologists, sociologists, political scientists, historians and lawyers, as well as by philosophers. There is much overlap between the definitions and broad agreement on the main features which a political organization must exhibit in order to belong to the class 'the state' (Green 1988: ch. 3; MacCallum 1987: chaps 3 and 5; Vincent 1987: ch. 1). (1) There is a population which reproduces itself and whose members are socially related. (2) There is territory. (3) There is a single government, which: (a) is a distinct body of rule, supported by a judicial, administrative and military machine; (b) is the ultimate prescriber and enforcer of law for all those within its jurisdiction; (c) claims exclusive control of the use of force within the territory and has preponderant control of its use; (d) claims authority for its existence and actions and is generally accepted as authoritative. (4) The state is legally and politically independent from other states, and recognized by other states as an independent or sovereign state (see Sovereignty §4).

These features are necessary conditions and together are sufficient. A state without a population is inconceivable, for example; as is one without a territory, except possibly in the short term. However there is considerable vagueness about what counts as meeting the conditions, and much variation in the degree to which they are met. Moreover, probably no state fully realizes all the characteristics ascribed to the abstraction 'the state'. For example, there is no agreement on what size the population or territory must be, how effective the government must be in enforcing its decisions, how coherent and consistent the legal system must be, how many inhabitants may reject the government's or the state's authority even to the extent of active resistance, or by how many other states recognition is required.

It should be noted that the state cannot be the same as its government. The government is the state's administrative organ, and acts in its name; it may be constrained in what it does by a formal constitution (see Constitutionalism). The state persists through changes of government and through changes in the form of constitution. Again, the state is not the same as its society, being only the political aspect of it, only one social institution among the many that constitute society (for example, religious, economic and family groups); states vary however in the amount of independence they permit to the other social institutions (this is part of what distinguishes liberal from totalitarian states). Although it is often urged that a state should be a nation, and every nation a state, seldom if ever do states

coincide with nations (if 'nation' means anything more than the body of those who are fellow nationals of a state), and being a nation is not a condition of being a state (see Nation and nationalism §1). Nor need a state rest on any particular principle of legitimacy, or take a particular constitutional form, although there is now an expectation that it should be in some sense democratic (see Legitimacy §4).

2 Controversial issues

It is important to appreciate that, given the historical specificity of the modern state and of the philosophical discussion of it, the terms in which the modern state has been and still is conceived include values which are neither permanent nor universal, but derive from the political, cultural, religious and economic assumptions of their times. For example, the early states were thoroughly patriarchal, and in their accounts of them philosophers took for granted that the members of the state were male heads of households (Pateman 1988). More generally, early theorists operated with a particular view of what an 'individual' is, which affected their understanding of what the state is (Siedentop 1983). Contemporary theorists need to examine their own assumptions to check that they are not treating ephemeral values as if they were those of human beings as such. For instance, a philosopher in the Anglo-American tradition may unconsciously take a more sceptical view of the state and what it is capable of than one in the continental European tradition, and a Western philosopher implicitly may make numerous assumptions about the place of religion in politics or the nature of the individual which are not held universally.

The value component of the idea of the state becomes even more obvious when one asks what the proper role of the state is. There is no agreement on the purpose or purposes of the state. Empirically, a huge range of tasks, not all compatible, has been given to - and sometimes taken away from - the state (see Law, limits of). Theoretically, there have been great divergences. For example, some have allowed very limited scope to the state, claiming that if it has greater powers it infringes the rights of individuals, particularly with regard to liberty and property (for example, see Nozick, R. §2). Others contend that the state must act extensively to enable individuals to enjoy their rights, which otherwise remain formal and empty, or in order to to achieve other desirable goals such as social justice (see Rawls, J.). Again, some view the state instrumentally as the means of protecting the material interests of its members; whereas others think it is intrinsically good because its citizens can exercise public virtue by participating in its affairs (see Republicanism).

Given the controversy over the role of the state, it seems hopeless to stipulate as a necessary feature of the state any particular function beyond protection of basic rights against force and fraud, resolution or control of dispute and conflict, and maintenance of its own ability to enforce its own decisions and to remain independent internationally. These would be regarded as realistic and morally acceptable by all except anarchists (Wolff 1970). But even when the state is defined minimally, its future may still be in doubt.

3 The future of the state

The state's role, whether limited or extensive, is difficult, and there is always dissatisfaction with its performance. The world system of states also is found inadequate because it is disfigured by war. Inevitably it is asked whether the purposes assigned to the state could be achieved better by other means; either without any political organization as the anarchist contends, or by a different political organization.

State theorists judge the state far preferable to anarchy, especially in modern political circumstances. The proposal to abolish it overlooks the problems, especially in complex societies, of social life without central authority, and ignores the substantial benefits which states do provide. Of course having a state involves risks (such as oppression by the government), but the appropriate response is to minimize the risks and maximize the benefits (as liberal democracy seeks to do). The anarchist 'solution' forgoes the best means to achieve that. Furthermore, there is one moral benefit which would be impossible under anarchy: the personal self-education and development which comes from participating in law making and taking responsibility for important public decisions and their enforcement (see Anarchism §3).

The main political alternative to the state is a cosmopolitan organization which strictly could not be a state, there being no other equal body, but which had the attributes of a state (making it unacceptable to anarchism). Its population would be humanity and its territory the world, and it would be superior in law and power to every other political body. States would lose their sovereignty and necessarily cease to be states (see International relations,

philosophy of §4).

This idea has been seriously discussed since the eighteenth century, and involves no impossibility. It is claimed that the political problems of the present system are so great it must give way. The cosmopolis could achieve the aims of states as well as or better than states. In addition it could achieve material and moral goals which states are incapable of - eliminating war, protecting the environment and eradicating world poverty. In particular, if justice is to be achieved on an international scale it cannot be left to states: states make exclusive claims to resources and territory, on behalf of their own members, which conflict with the distribution required by principles of international justice (Baldwin 1992) (see Justice, international §4). Since the ultimate moral criterion is the equality of claims of all human beings, the final political unit should reflect this.

There are two issues here. First, should a cosmopolitan morality be given priority over the moralities of individual states? A major reservation is that the principles and content of the universal cosmopolitan morality are not agreed among philosophers let alone accepted by all states. There is also an objection: if the cosmopolitan morality conflicts with the state's morality, why should the former take precedence in all cases? May not a state's values, rooted in its own culture, have independent and greater moral importance?

Second, even if the case for cosmopolitan morality were conceded, why should states be replaced by a cosmopolitan 'state'? It is uncertain that the cosmopolis would perform its functions adequately; or effectively continue what states had done (whatever that was thought to be); or enable its citizens to participate adequately. If it fell short, there could be no political body to correct or control it.

However, it is not necessary to abolish states in order for cosmopolitan morality to be practised. The latter could be combined with legal and political statism. Each state could accept that it had obligations to human beings other than its own members; although it would be left to the state itself to decide what the obligation amounted to and what action was required. Legally, states could continue as sovereign: morally, they would subject themselves to global standards and recognize and act on wider obligations and responsibilities.

Beginning from a world of states, the way forward is through states, whether this be via a system of cooperation between states or their replacement by a single global political organization. In either case, questions about the nature, purpose and best form of the state remain important.

PETER P. NICHOLSON

References and further reading

Baldwin, T. (1992) 'The Territorial State', in H. Gross and R. Harrison (eds) *Jurisprudence: Cambridge Essays*, Oxford: Clarendon Press, 207-30.(Examines the basis for the legitimacy of the territorial claims states make; suggests some principles.)

Barry, B. (1983) 'Self-government Revisited', in D. Miller and L. Siedentop (eds) *The Nature of Political Theory*, Oxford: Clarendon Press, 121-54; repr. in B. Barry, *Democracy, Power and Justice*, Oxford: Clarendon Press, 1989, 156-86.(Illuminating and analytically sharp discussion of 'nation' and 'nation-state'.)

Beitz, C.R. (1991) 'Sovereignty and Morality in International Affairs', in D. Held (ed.) *Political Theory Today*, Cambridge: Polity Press, 236-54.(Introductory survey.)

Black, A. (1992) *Political Thought in Europe 1250-1450*, Cambridge: Cambridge University Press.(An introductory history.)

Dunn, J. (1994) *Contemporary Crisis of the Nation State?*, Political Studies series, special issue no. 42, Oxford: Blackwell.(Includes contemporary and historical perspectives, and considers the theoretical implications.)

d'Entréves, A.P. (1967, 1969) *The Notion of the State: An Introduction to Political Theory*, Oxford: Clarendon Press.(Useful systematic survey of the basic ideas and issues, set in the perspective of the history of political thought.)

Fain, H. (1972) 'The Idea of the State', *Nous* 6 (1): 15-26.(Hostile view of the rights states assign one another.)

Green, L. (1988) *The Authority of the State*, Oxford: Clarendon Press, ch. 3.(Careful and accessible investigation of the distinctive features of the state, with robust responses to doubts about 'state' as a theoretical concept.)

Hoffman, J. (1995) *Beyond the State: An Introductory Critique*, Cambridge: Polity Press. (Valuably sceptical discussion of the worth of and need for states, investigating alternatives 'beyond' them.)

Linklater, A. (1982, 1990) *Men and Citizens in the Theory of International Relations*, London: Macmillan. (Philosophical and historical exploration of the tensions between moral obligations as a human being and rights

and duties as a member of a state.)
MacCallum, G.C. (1987) *Political Philosophy*, Englewood Cliffs, NJ: Prentice Hall.(Introductory level analyses, although compressed, of the main terms, including state, nation, government, authority and sovereignty; also includes a discussion of the main issues.)
Miller, D. (1995) *On Nationality*, Oxford: Clarendon Press.(Stresses the connection between state and nationality, and the consequent limit on cosmopolitanism.)
Pateman, C. (1988) *The Sexual Contract*, Cambridge: Polity Press.(Exposes the patriarchalism in the social contract and in modern forms of contract.)
Siedentop, L. (1983) 'Political Theory and Ideology: The Case of the State', in D. Miller and L. Siedentop (eds) *The Nature of Political Theory*, Oxford: Clarendon Press, 53-73.(Analyses 'state' in conjunction with 'individual' and claims both are historically specific to the modern Western idea of the state.)
Skinner, Q. (1989) 'The State', in T. Ball, J. Farr and R.L. Hanson (eds) *Political Innovation and Conceptual Change*, Cambridge: Cambridge University Press, 90-131.(Historical sketch of the emergence of the term 'state'.)
Thompson, J. (1992) *Justice and World Order: A Philosophical Enquiry*, London and New York: Routledge. (Clear, politically informed and judicious introductory level examination of the moral issues concerning the role of states in achieving a just world.)
Vincent, A. (1987) *Theories of the State*, Oxford: Blackwell.(Introductory survey of the main theories, with a helpful general chapter.)
Wolff, R.P. (1970, 1976) *In Defense of Anarchism*, New York: Harper & Row.(Spirited statement of philosophical anarchism, denying there can be any theoretical justification for the authority of the state.)

Statistics

The discipline of statistics encompasses an extremely broad and heterogeneous set of problems and techniques. We deal here with the problems of statistical inference, which have to do with inferring from a body of sample data (for example, the observed results of tossing a coin or of drawing a number of balls at random from an urn containing balls of different colours) to some feature of the underlying distribution from which the sample is drawn (for example, the probability of heads when the coin is tossed, or the relative proportion of red balls in the urn).

There are two conflicting approaches to the foundations of statistical inference. The classical tradition derives from ideas of Ronald Fisher, Jerzy Neyman and Egon Pearson and embodies the standard treatments of hypothesis testing, confidence intervals and estimation found in many statistics textbooks. Classicists adopt a relative-frequency conception of probability and, except in special circumstances, eschew the assignment of probabilities to hypotheses, seeking instead a rationale for statistical inference in facts about the error characteristics of testing procedures. By contrast, the Bayesian tradition, so-called because of the central role it assigns to Bayes' theorem, adopts a subjective or degree-of-belief conception of probability and represents the upshot of a statistical inference as a claim about how probable a statistical hypothesis is in the light of the evidence.

1 The classical tradition

Because classicists adopt a frequentist conception of probability (see Probability, interpretations of §3), the upshot of statistical inference within the classical framework is not, as it is within the Bayesian approach, the assignment of a probability to a hypothesis which measures how well the hypothesis is supported by the evidence. Instead, the classical approach focuses on finding rules or policies to guide the investigator's behaviour with respect to hypotheses - where the rationale for adopting such rules has to do with how often we can expect to make various sorts of mistake in the long run if we follow them. In this respect, the classical approach is similar in spirit to reliabilist programmes in epistemology (see Reliabilism).

Thus, for example, in classical significance tests the investigator proceeds by choosing a significance level (for example, 0.05) that corresponds to what they regard as an acceptable probability that they will mistakenly reject the hypothesis under test (the null hypothesis) even though it is true. The investigator then constructs a rejection region - a set of actions that would lead to rejection of the hypothesis - in such a way that the probability of rejecting the hypothesis even though it is true is less than or equal to the significance level. Intuitively, the rejection region will consist of outcomes that are unfavourable to the null hypothesis grouped together with the other possible outcomes that are even more 'extreme' or unfavourable to the null hypothesis. To illustrate, suppose that the investigator wishes to test the hypothesis that a coin is fair by flipping it twenty times and observing the number of heads, and there is no particular reason to think that, if the coin is biased, it will be biased towards heads rather than tails. The total probability of obtaining fifteen or more heads or five or fewer heads when flipping a fair coin twenty times is just under 0.05 and the observation of one of these outcomes will lead to rejection of the hypothesis that the coin is fair at a 0.05 significance level.

Within a classical framework, when we reject a hypothesis at the 0.05 level of significance, we are *not* entitled to infer that the hypothesis has at least a 0.95 probability of being false or at most a 0.05 probability of being true. Instead, the classical justification for the adoption of the above test is that, if it were repeated a large number of times, the characteristics of the test are such that we may expect to make a certain kind of mistake - rejecting the null hypothesis even though it is true - roughly 5 per cent of the time. According to classicists, this is a probability that has a perfectly straightforward frequency interpretation. It should also be understood that nothing in the formal theory of significance testing requires the particular choice of significance level adopted in this example. While 0.05 is a common choice, particularly within the social and behavioural sciences, investigators in many areas of science may and often do choose different significance levels. In general, this choice will correspond to the cost the researcher attaches to the error of rejecting the null hypothesis even though it is true. Presumably this cost will often vary from context to context - depending, for example, on the use to which the test result is to be put. This introduces what many critics regard as a subjective or arbitrary element: an outcome which would lead to

the rejection of a hypothesis at one significance level might not lead one to reject that hypothesis at some different (lower) significance level.

Suppose that one carries out a significance test and observes an outcome that does not fall within the rejection region. Properly speaking, an investigator can conclude in this case only that one has 'failed to reject' the null. The theory of hypothesis testing developed by Neyman and Pearson (1928) is motivated in part by the desire to construct an account of testing that permits decisions to 'accept' hypotheses as well as decisions to reject. Given that acceptance is a possibility, it is easy to see that significance testing focuses on only one of two possible errors one can make in testing a hypothesis. In addition to the possibility of mistakenly rejecting a true hypothesis (a so-called type I error), there is also the possibility of mistakenly accepting the hypothesis under test even though it is false (a so-called type II error). To specify the latter probability we must make reference to some alternative hypothesis against which the null is being tested. Thus, one of the distinctive features of the Neyman-Pearson account is that one tests a hypothesis by comparing it with some competing, rival hypothesis or set of such hypotheses (see Confirmation theory; Crucial experiments). Ideally a good test would minimize the probability of both a type I and a type II error, but unfortunately these two desiderata conflict, for reducing the former probability will in general increase the latter probability, and vice versa. The reader should consult the references for the solution to this difficulty recommended by Neyman and Pearson, but the fundamental point to appreciate is that, again, the basic rationale for adopting a test within this framework has to do with the frequencies with which one may expect to make various kinds of mistake if that testing procedure were to be repeatedly followed.

A similar rationale in terms of the error characteristics of repeated procedures underlies other forms of classical statistical inference. For example, a *confidence interval* is an interval constructed in such a way that it has a high probability (for example, 95 per cent) of containing some unknown parameter such as the bias of a coin. Since, for a classicist, is a parameter with a fixed value rather than a random variable, we cannot say in this case that the probability that falls within the constructed interval is 95 per cent. Instead, it is the confidence interval itself which is thought of as a random variable to which a probability attaches. The idea is that, if we were repeatedly to follow the same rule in constructing confidence intervals for , in the long run the intervals would contain about 95 per cent of the time.

2 The Bayesian tradition

The basic result on which the Bayesian approach relies is Bayes' theorem, a simple form of which is

$$to\mathrm{P}(\mathrm{H} \mid E) = \frac{P(E|H) \cdot P(H)}{P(E)} (1)$$

Bayesians interpret this theorem as telling us that the probability of a hypothesis H given evidence E - often called the posterior probability - is a function of the prior probabilities $\mathrm{P(H)}$ and $\mathrm{P(E)}$ (so called because they are the probabilities the investigator assigns to H and E prior to the collection of evidence) and the conditional probability or likelihood term $\mathrm{P(E \mid H)}$ (see Probability, interpretations of §5). In many statistical applications it will be more appropriate to use instead a continuous form of Bayes' theorem rather than the discrete form (1), but the basic structure of the inference will be the same. To illustrate, suppose that we are interested in estimating the unknown mean μ of some random variable which is normally distributed with known variance σ^2. Suppose that our prior probability regarding μ can itself be represented by a variable *X* ranging over different possible values of μ where *X* also follows a normal distribution with mean μ′ and variance σ'^2, and that we draw a sample of size *n* and that the sample mean is $\bar{y}$. When one combines these priors with the sample information, in accord with Bayes' theorem, the result is a posterior distribution for μ which is also normal and the mean μ'' and variance σ''^2 of which are given by

-

Here, unlike the classical approach, the result of the inference is the assignment of a probability distribution to μ'' and σ'' themselves. For example, we can use (2) to calculate the probability that μ falls within an interval centred around $\bar{y}$ while no such interpretation of a classical confidence interval is legitimate.

3 Criticisms of the classical tradition

One prominent Bayesian criticism focuses on the fact that, within classical approaches, the significance of evidence seems to depend on the test procedure or stopping rule by which the evidence is produced. For example,

a significance test like that described above, in which we flip a coin a fixed number of times (for example, twenty) and observe the number of heads which result, will have different error characteristics from a significance test in which we flip the coin until a fixed, predetermined number of heads are observed (for example, six) even though it may be the case that the observed result in both cases is six heads in twenty tosses. Indeed, the hypothesis that the coin is fair will be rejected at the 0.05 level if the observed outcome of six heads and fourteen tails is the result of a test conducted according to the second rule, but not if this outcome is the result of a test conducted according to the first rule.

Many Bayesians find it counterintuitive that the ‘same’ evidence should lead to different decisions about whether to reject depending upon which testing procedure is followed. If one is willing to assume, as many Bayesians are, that features of the stopping rule employed should not affect the likelihood term in a Bayesian analysis, then that approach will yield the result in the above example that the observed outcome of six heads, fourteen tails has the same evidential significance for the hypothesis that the coin is fair, regardless of which of the above rules was followed. Classicists respond by arguing, with considerable plausibility, that in many cases the testing procedure followed does seem relevant to the evidential import of the data it generates and that it is a defect in standard Bayesian analyses that they tend to ignore this.

Bayesians also argue that the classical notions of acceptance or rejection lack a clear meaning. Within a classical framework, to accept (reject) a hypothesis is not to regard it as true (false) beyond all reasonable doubt nor to assign it a high (low) probability. Bayesians argue that their picture, in which an investigator can have any degree of belief between zero and one and will continually change this as additional evidence accumulates, better fits the range of epistemic attitudes one finds in science and in ordinary life. Yet another Bayesian criticism focuses on what Bayesians regard as unacceptably subjective or arbitrary factors within the classical framework - for example, the choice of significance level. As we have emphasized, many of these features derive from the role that judgments about epistemic costs or values play within the classical framework.

While it is not possible to discuss these criticisms in detail, several further observations seem in order. The first is that, while the notions of acceptance and rejection may seem unclear, it is also by no means obvious that the Bayesian framework, with its demand for precise point-valued degrees of belief conforming to the probability calculus, adequately represents the structure of opinions and epistemic attitudes present in real life or in scientific investigation. The second point is that Bayesian strictures about acceptance and rejection within statistical contexts connect up with a more general debate within philosophy of science about whether scientists in fact ever accept hypotheses (or which may or may not be the same thing, believe them simpliciter) as opposed to assigning them degrees of belief strictly between zero and one and, if so, what such acceptance involves and whether, as a normative matter, it is ever rationally defensible (see Scientific realism and antirealism). Many writers - including some Bayesians - have argued that acceptance and rejection are ineliminable parts of scientific practice. The debate over the role that classicists assign to costs and benefits within hypothesis testing also echoes a more general debate within epistemology and philosophy of science over whether the stringency of standards for knowledge, justifiable belief and acceptance vary from context to context depending on, among other things, what is at stake in making various sorts of mistake and the investigator’s conception of the goals of inquiry (see Scientific method).

4 Criticisms of the Bayesian tradition

The fundamental classical objection to Bayesian approaches is the apparent subjectivity introduced by the choice of a prior distribution. One Bayesian response to this criticism, as noted above, is that subjectivity also enters into the classical programme at a number of points. Bayesians claim that their programme systematizes and makes explicit aspects of statistical inference that are unavoidably subjective, while avoiding certain objectionable kinds of subjectivity which are present in the classical programme, but often in an *ad hoc* and concealed form. The second Bayesian reply is that, as more and more data accumulates, the effects of different choices of prior distribution on one’s posterior distribution will often become less and less important. We can see this in the case of formula (2) - as sample size n grows, the sample data become more influential and the prior distribution less influential in determining the shape of the posterior distribution. Moreover, Bayesians point out that, when the choice of prior distribution does make a difference, the investigator can always produce a kind of robustness or sensitivity analysis, showing how (and to what extent) different choices of priors will lead to a different posterior

distribution, and leaving it up to other investigators to decide which priors to adopt.

See also: Induction, epistemic issues; Inductive inference; Probability theory and epistemology; Statistics and social science

JAMES WOODWARD

References and further reading

Cramér, H. (1971) *Mathematical Methods of Statistics*, Princeton, NJ: Princeton University Press.(An authoritative text at an advanced level in the classical tradition.)

Fisher, R.A. (1925) *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd, 14th edn.(see 1947)

Fisher, R.A. (1947) *The Design of Experiments*, New York: Hafner, 4th edn.(Along with 1925, provides a classic exposition of the basic ideas of significance testing and experimental design.)

Giere, R. (1977) 'Testing Versus Information Models of Statistical Inference', in *Logic, Laws and Life: Some Philosophical Complications*, Pittsburgh, PA: University of Pittsburgh.(Giere defends the classical approach, and this paper, as well as the 1973, contains responses to the criticisms of the classical approach discussed in §3 above.)

Giere, R. (1973) 'Empirical Probability, Objective Statistical Methods and Scientific Inquiry', in W.L. Harper and C.A. Hooker (eds) *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, vol. II, *Foundations and Philosophy of Statistical Inference*, 63-93.(A systematic philosophical defence of the Neyman-Pearson tradition.)

Hacking, I. (1965) *The Logic of Statistical Inference*, Cambridge: Cambridge University Press.(Discusses numerous criticisms of the classical approach. Highly recommended.)

Hodges, J.L. and Lehmann, E.L. (1970) *Basic Concepts of Probability and Statistics*, San Francisco, CA: Holden-Day.(Well-known textbook at a relatively elementary level which describes basic ideas in probability theory and statistical inference from a classical perspective.)

Howson, C. and Urbach, P. (1993) *Scientific Reasoning: The Bayesian Approach*, Chicago, IL: Open Court, 2nd edn.(Contains a lucid exposition of classical and Bayesian approaches to statistical inference, along with extensive criticism of the classical approach from a Bayesian perspective.)

Kyburg, H. (1974) *The Logical Foundation of Statistical Inference*, Dordrecht: Reidel.(Discusses many of the difficulties with the classical approach described in §3 and offers an account of statistical inference which is neither classical nor conventionally Bayesian.)

Lehmann, E.L. (1959) *Testing Statistical Hypothesis*, New York: John Wiley & Sons. (Well-known text in the classical tradition. Somewhat more technically sophisticated and demanding than Hodges and Lehmann 1970.)

Lindley, D.V. (1956) *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Cambridge: Cambridge University Press.(Systematic exposition of the Bayesian approach by one of its best-known defenders.)

Maher, P. (1993) 'Acceptance in Bayesian Philosophy of Science', in D. Hull, M. Forbes and K. Okruhlik (eds) *PSA 1992*, vol. 2, East Lansing, MI: Philosophy of Science Association.(Argues that acceptance of hypotheses plays an important role in science and that Bayesians should recognize this. See §3.)

Mayo, D. (1979) 'Testing Statistical Testing', in J. Pitt (ed.) *Philosophy in Economics*, Dordrecht: Reidel, 175-203.(A defence of the classical theory of hypothesis testing which explicitly addresses the Bayesian criticisms regarding stopping rules discussed in §3 above.)

(1996) *Error and Growth in Experimental Knowledge*, Chicago, IL: University of Chicago Press.(Extensive discussion of foundational issues in statistics and their relation to general philosophy of science. Defends classical and criticizes Bayesian approaches.)

Neyman, J. and Pearson, E.S. (1928) 'On the Use and Interpretation of Certain Test Criteria for the Purposes of Statistical Inference', *Biometrika* 20: 175-240, 263-94. (Basic ideas of the Neyman-Pearson theory of hypothesis testing.)

Nozick, R. (1993) *The Nature of Rationality*, Princeton, NJ: Princeton University Press.(Although not explicitly concerned with problems of statistical inference, contains relevant material on the role of costs and epistemic values in justification and acceptance. See especially pages 65-106.)

Rosenkrantz, R. (1977) *Inference, Method and Decision: Toward a Bayesian Philosophy of Science*, Dordrecht: Reidel.(A general defence of an 'objective' Bayesian approach to inductive inference, but pages 177-223 contain an interesting discussion of problems of statistical inference, including an argument from within a Bayesian framework that in the coin-tossing example the evidential import of data is affected by the stopping rule employed. See §3.)

Seidenfeld, T. (1979) *Philosophical Problems of Statistical Inference: Learning from R.A. Fisher*, Dordrecht: Reidel.(Extensive, technically informed discussion of foundational problems of statistical inference from a Bayesian perspective.)

Statistics and social science

There are a number of distinct uses for statistics in the social sciences. One use is simply to provide a summary description of complicated features in a population.

A second use of statistics is to predict (some) features of a unit or group in a population, given other features of the unit or group. For example, a company may charge lower health insurance rates for people who do not smoke, because smokers have a lower risk of lung cancer. Some companies could also charge lower health insurance rates for people who do not have a heavy cough, because the probability of having lung cancer is lower for such people. Predictions can be made by developing a probabilistic model of the joint distribution of incidence of smoking, lung cancer, and incidence of heavy coughs in the population.

A third use of statistics is to help predict the probable effects of adopting different policies. For example, the government may consider a number of alternative policies for reducing the rate of lung cancer. One policy would ban smoking. Another policy would make everyone who coughs take cough medicine. Both smoking and coughing are predictors of lung cancer. But because smoking is a cause of lung cancer, while coughing is an effect of lung cancer, the first policy seems as if it might achieve the desired effect, while the second does not. In order to answer policy questions we need to know not only how the variables are distributed in the actual population, but also how they are causally related. A causal model specifies the causal relations between features in a population, as well as specifying a probability distribution of the features. Statistical information together with causal information can be used to predict the effects of adopting a certain policy.

A fourth use of statistics is in helping decide which policies should be adopted in order to achieve specific goals. Such decisions are based not only on the probable effects of each policy, but also on assigning different utilities to each possible outcome. This use of statistics is a branch of decision theory.

1 Summary statistics

Consider a purely hypothetical example. Suppose a random sample of fifty units is taken from a population, and for each unit the values of three random variables (properties) are measured: *NCIG*, how heavy a smoker a person has been (on a scale from 0 to 7), *LC*, the presence of lung cancer, and *HC*, the presence of a heavy cough (on a scale from 0 to 7 of severity). (For this example, the sample was pseudo-randomly generated on a computer from an arbitrarily chosen population distribution.)

When the data are presented in the form of a table, the sheer volume of the information makes it uninformative, and it would be even less informative if the sample size were larger or the values of more random variables had been measured. For that reason, it is often useful to provide statistics which summarize important features of the sample. A number of summary statistics for the sample are shown in Table 1.

Table 1. Summary statistics

	NCIG	**LC**	**HC**
Minimum	0	0	0
Maximum	7	1	7
Median	3	0	3
Mean	2.84	0.42	3.04
Variance	4.30	0.25	5.63

A more informative visual presentation of the sample information about an individual variable such as *NCIG* is the histogram shown in Figure 1. The histogram shows the number of people in each category.

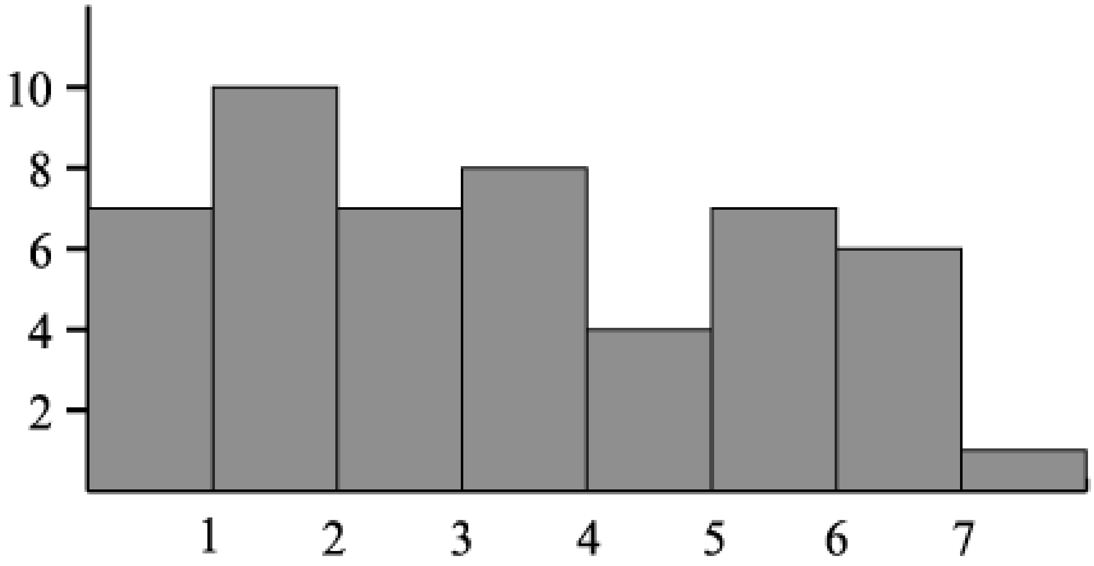

Figure 1

A histogram is useful for summarizing the distribution of a single variable, but it does not show the relationships between variables. One way of showing the relationship between a pair of multinomial variables is a contingency table, such as the one shown in Table 2. For each pair of values of *NCIG* and *LC* there is a box, and the number in the box represents the number of units in the population with that pair of values.

Table 2

NCIG									
		0	1	2	3	4	5	6	7
LC	0	6	9	3	0	0	2	3	7
	1	2	5	5	1	1	2	0	4

Devising visually informative displays of relationships between three or more variables is much more difficult than for two variables.

2 Probabilistic models

If *LC* is a variable that is difficult or expensive to measure, and *NCIG* and *HC* are relatively easy to measure, then given a new unit from the population, a doctor may wish to predict the value of *LC* for that unit from the more easily measured values of *NCIG* and *HC* for that unit. One simple method for predicting the value of *LC* for a new unit, given the values of *NCIG* = *x* and *HC* = *z* for that unit involves the following two steps:

First, use the sample to 'estimate' a population joint probability distribution P. (Strictly speaking, as explained below, statisticians usually speak of estimating the parameters of a family of distributions which may single out a unique probability distribution P, rather than of estimating P itself.)

Second, choose the value of *LC* for which

$$\mathrm{P}(LC \mid NCIG = x,\ HC = z)$$

is the highest (using an arbitrary rule to break ties).

One often attempts to 'estimate' a joint probability distribution from a sample by first choosing a family of probability distributions, and then progressively narrowing down the family until one probability distribution is

left. Deciding which family of distributions to start from, and which choices to make in the course of narrowing down the family of distributions is usually more of an art than a science. Let us now consider how the sample data might be used to obtain a joint probability distribution for *NCIG*, *LC*, and *HC* in the following way. First choose a family of distributions that is believed to contain the population distribution. This may be done by looking at the data, using background knowledge, or both. While the data may suggest a family of distributions, it does not generally single out a unique one. In the very simple example, *NCIG* and *HC* take on any of eight different values, and *LC* takes on one of two different values in the sample. In this case, it follows from the way that we have measured the variables that the distribution is a multinomial.

Having selected the family of distributions, a sample can be used to 'estimate' a unique distribution from the family of distributions. One way to do this uses the following formula, which is true of any multinomial probability distribution:

$to\mathrm{P}(\mathrm{NCIG}=\mathrm{x},\ \mathrm{LC}=\mathrm{y},\mathrm{HC}=\mathrm{z}) = \mathrm{P}\ (\mathrm{NCIG}=\mathrm{x}) \times \mathrm{P}(LC=y \mid NCIG=x) \times \mathrm{P}(HC=z \mid NCIG=x, LC=y)(1)$

Let H_1 be the family of distributions that can be represented in the form of equation (1). It is possible to pick out a unique multinomial distribution from the family of multinomial distributions by specifying P(*NCIG*), $\mathrm{P}(LC \mid NCIG)$, and $\mathrm{P}(HC \mid NCIG, LC)$. The quantities to be specified in order to select a unique distribution from a family of distributions are the parameters of the distribution family. The values of parameters can be fixed in one of two ways. A parameter can simply be assumed to have a value, perhaps on the basis of background knowledge. For example, the value of $\mathrm{P}(NCIG=2)$ could be fixed at 0.2; in that case $\mathrm{P}(NCIG=2)$ is a fixed parameter. If the value of a parameter is not fixed, but is instead estimated from the sample data, then the parameter is a free parameter. A free parameter is typically estimated by some formula that uses the values of the measured random variables in the sample. For example, an estimate of $\mathrm{P}(NCIG=2)$ in the population is given by the number of units in the sample with $NCIG=2$ divided by the number of units in the sample, and the other parameters can be calculated in an analogous way. Suppose that the free parameters are estimated in this way from the sample; let the resulting distribution be called P_1.

The same population parameter can be estimated by a number of different estimators. How should an estimator be chosen? One desirable property is that it be practically certain that the estimator can be made as close to the true value of the parameter as desired, as long as the sample is made large enough; such an estimator is 'consistent'. A second desirable feature of an estimator is that the expected value of the estimator equal the population value of the parameter; such an estimator is 'unbiased'. A third desirable property is that the probability of large differences between the value of the estimator on different samples of the same size is small; such an estimator has 'low variance'. Typically, the variance of an estimator decreases with increasing sample size. An estimator of a parameter which makes the sample data as likely as possible, given the constraints of the model is called a 'maximum likelihood' estimator; under a wide variety of conditions maximum likelihood estimators are unbiased and consistent. For a discrete distribution, the number of units in the sample with *NCIG* = *x* divided by the number of units in the sample is a maximum likelihood estimator.

How well does P_1 do in predicting the value of *LC*? In order to show how well P_1 predicts on a number of different samples, ten different samples of size fifty were pseudo-randomly generated on a computer from the same population probability distribution (the first sample was used to 'estimate' P_1). The number of errors in each sample made by P_1 in predicting the value of *LC* given the measured values of *HC* and *NCIG* is shown in Table 3.

Table 3

Sample	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Number wrong	3	7	13	14	17	11	13	12	12	9

H_1 is a very flexible family of distributions, and can reproduce any multinomial distribution exactly. A disadvantage of this is that it makes the 'estimate' of the population distribution vary a lot from sample to sample. That is why the number of wrong predictions of *LC* is relatively small in the sample which we used to 'estimate' (sample 1) but much larger on the other samples. Another way of looking at this is that the family of distributions

chosen has a large number of free parameters that have to be estimated with relatively few data points. This means that estimates from different samples vary quite widely. This in turn can lead to variations in the prediction of the value of *LC*, given the same values of *NCIG* and *HC*.

One way to reduce the variation in parameter estimates is to choose initially a family of models that has fewer free parameters. For example, let H_2 be the family of distributions that can be represented by the following formula:

$$to\ \mathrm{P}(\mathrm{NCIG}=\mathrm{x},\ \mathrm{LC}=\mathrm{y},\mathrm{HC}=\mathrm{z}) = \mathrm{P}\ (\mathrm{NCIG}=\mathrm{x}) \times \mathrm{P}(LC = y \mid NCIG = x \times \mathrm{P}(HC = z \mid NCIG = x)(2)$$

H_2 consists of those multinomial distributions in which

$$\mathrm{P}(HC = z \mid NCIG = x, LC = y) = \mathrm{P}(HC = z \mid NCIG = x),$$

that is, ***HC*** is independent of ***LC*** conditional on ***NCIG***. H_2 is properly contained in H_1. The free parameters of H_2 can be estimated using this formula and the sample by substituting the conditional frequencies in the sample for the corresponding conditional probabilities in the equation; call the resulting probability distribution P_2. Unlike H_1, not every multinomial distribution is contained in H_2. If the population distribution falls within H_2 (as it did in this example) then in general the predictions of *LC* from P_2 will be more accurate than the predictions of *LC* from P_1, because the number of parameters that need to be estimated has been reduced, and the estimates of the parameters tend to vary less from sample to sample. In the first sample, there are four units with $NCIG = 4$, where three of those four have $LC = 1$ and one has $LC = 0$. These four points are used to estimate fourteen parameters of P_1 (that is, $\mathrm{P}_1(HC \mid NCIG = 4, LC = 0)$ and $\mathrm{P}_1(HC \mid NCIG = 4, LC = 1)$) but only seven parameters of P_2 (that is, $\mathrm{P}_2(HC \mid NCIG = 4)$). Moreover, the predictions of *LC* are in general more accurate when the estimates of the population distributions are made using P_2 rather than P_1. Calculating the number of errors made in predicting *LC* from *HC* and *NCIG* on the same ten samples used before, the number of errors made by P_2 is shown in Table 4.

Table 4

Sample	1	2	3	4	5	6	7	8	9	10
Number wrong	5	11	12	10	17	7	10	8	9	14

Note that the number of errors made in sample 1 is larger for P_2 than for P_1. This is partially because sample 1 is the sample that was used to estimate the parameters of H_1 and H_2, and H_2 cannot reproduce the sample distribution exactly, while H_1 can. However, on all of the other samples, except samples 2 and 10, the prediction of *LC* from P_2 is at least as accurate as the prediction of *LC* from P_1.

While there are advantages to assuming that the population distribution lies in H_2, there are also possible disadvantages. If the assumption is false, the predictions of *LC* from P_2 may be worse than the predictions of *LC* made from P_1, because the parameter estimates could be biased. However, even if the population distribution does not lie in H_2, but it is close to a distribution that lies in H_2, the prediction of *LC* from P_2 may be better than the prediction made from H_1. Whether or not this is the case depends upon how close the population distribution is to a distribution that lies in H_2, and how much the variance of the estimator is reduced by assuming the distribution is in H_2. One way of judging whether the population distribution is a member of H_2 is by performing a significance test.

The simplest form of a significance test compares two hypotheses: in our case that the population distribution equals P_2 or that the population distribution equals P_1. The procedure either rejects P_2 or accepts P_2. There are two kinds of errors that can be made: rejecting P_2 when it is true, and accepting P_2 when it is false. In this context, the hypothesis is that P_2 is true is called the ‘null hypothesis’, the first kind of error is called a ‘type I error’, and the second kind of error is called a ‘type II error’. The power of the test is 1 - P(type II error). The goal is to minimize type I and type II errors. Unfortunately, these two goals usually conflict, so the actual procedure is to choose an acceptable probability of type I error, and then attempt to minimize the probability of type II error. An adaptation of the procedure can be used to test a given hypothesis against more than one alternative hypothesis. In that case, given a fixed acceptable probability of type I error, the goal is to find a test that minimizes the

probability of type II error against each of the alternative hypotheses. Although this procedure is widely employed in the social sciences to test probabilistic hypotheses, it has also come under severe criticism. The answer that one gets depends upon a number of free choices: the choice of which hypothesis is the null hypothesis, the choice of the alternative hypothesis, the choice of a test statistic, and the choice of a significance level. Moreover, given many parameters and small sample size, the probability of a type II error can be quite high.

The Bayesian approach to 'estimating' $\mathrm{P}(LC \mid NCIG, HC)$ is different from the procedure described above. First, choose some family of distributions, such as H_1. Let Θ represent the space of the set of parameters for the family of distributions. Next, for each in Θ, let *f*() be the value of a prior density function based on background knowledge. In the absence of specific background knowledge some variant of a ' non-informative prior' is often placed over Θ. Then the following equation can be used to calculate $\mathrm{P}(LC \mid NCIG, HC)$:

$$to\ \mathrm{P}(\mathrm{LC} \mid NCIG, HC) = \int_{\Theta} \mathrm{P}(LC \mid NCIG, HC, \theta)\ f(\theta \mid NCIG, HC)\mathrm{d}\theta$$

3 Causal models

Consider the following questions: what would the distribution of *LC* be if *NCIG* was forced to have the value 0 for each member of the population? What would the distribution of *LC* be if *HC* was forced to be 0 for each unit in the population?

These are questions about hypothetical populations, which may never exist. If P(*NCIG*, *LC*, *HC*) is the probability distribution in the actual population, let $\mathrm{P}_{\mathrm{Man}\,(NCIG=0)}\ (NCIG, LC, HC)$ and $\mathrm{P}_{\mathrm{Man}\,(HC=0)}\ (NCIG, LC, HC)$ be the probability distributions in the hypothetical population where *NCIG* has been forced to be zero, and the hypothetical population where *HC* has been forced to be zero, respectively.

How can $\mathrm{P}_{\mathrm{Man}\,(NCIG=0)}\ (LC)$ and $\mathrm{P}_{\mathrm{Man}\,(HC=0)}\ (LC)$ be inferred from P(*NCIG*, *LC*, *HC*)? Suppose hypothetically, that *NCIG* is a direct cause of *LC*, and *LC* is a direct cause of *HC*, but *NCIG* causes *HC* only via *LC*, and that these causal facts are known. This causal information can be represented in a directed acyclic graph of the following form:

$$NCIG \rightarrow LC \rightarrow HC$$

Figure 2

Here the arrow from *NCIG* to *LC* means that *NCIG* is a direct cause of *LC*, the arrow from *LC* to *HC* means that *LC* is a direct cause of *HC*, and the lack of an arrow from *NCIG* to *HC* means that *NCIG* is not a direct cause of *HC*.

Suppose that with each causal structure represented by a directed acyclic graph there is an associated set of probability distributions that are compatible with that causal structure. In the case of discrete distributions the rule can be stated very simply. If *G* is a directed acyclic graph representing a given causal structure, and every common cause of a pair of variables in *G* is also represented in *G*, assume that every probability distribution in a population with causal graph *G* can be factored in the following way:

$$\mathrm{P}(\mathbf{V}) = \prod_{V \in \mathbf{V}} \mathrm{P}(V \mid \mathbf{Parents}(V, G))$$

where V is the set of variables in *G*, and Parents(*V*,*G*) is the set of variables in *G* at the tail of a directed edge into *V*, for example, if $A \rightarrow B$ in *G* then *A* is a parent of *B* in *G*. If a distribution P can be factored according to this formula for directed acyclic graph *G*, then the distribution is said to satisfy the local directed Markov condition for *G*. In the case of Figure 2, the formula entails that

$$\mathrm{P}(NCIG, LC, HC) = \mathrm{P}(NCIG) \times \mathrm{P}(LC \mid NCIG) \times \mathrm{P}(HC \mid LC).$$

The local directed Markov condition is often, if not always explicitly, assumed in social science research. However, it is not universally true. In particular, if a directed graph is used to represent feedback, or if the properties of one unit affect the properties of another unit (via second-hand smoke, for example) the Markov condition may be false.

The following formula states a general rule for inferring the effects of forcing an 'ideal intervention' on variable *X* in a directed graph *G*, where an 'ideal intervention' forces some probability distribution $\mathrm{Man}(X)$ upon *X*, and has no other effect upon the system than through changing *X*.

$$\mathrm{P}_{\mathrm{Man}\,(X)}\,(\mathbf{V}) = \mathrm{P}_{\mathrm{Man}\,(X)}\,(X) \times \prod_{V \in \mathbf{V} \setminus \{X\}} \mathrm{P}(V \mid \mathbf{Parents}(V, G))$$

where $\mathbf{V} \setminus \{X\}$ is the set of all variables in V except *X*. Applying this equation to Figure 2, it follows that

$$\mathrm{P}_{\mathrm{Man}\,(NCIG)}\,(NCIG, LC, HC) = \mathrm{P}_{\mathrm{Man}\,(NCIG)}\,(NCIG) \times \mathrm{P}(LC \mid NCIG) \times \mathrm{P}(HC \mid LC),$$

so with a little arithmetic

$$\mathrm{P}_{\mathrm{Man}\,(NCIG)}\,(LC) = \mathrm{P}(LC \mid NCIG).$$

Similarly,

$$\mathrm{P}_{\mathrm{Man}\,(HC)}\,(NCIG, LC, HC) = \mathrm{P}(NCIG) \times \mathrm{P}(LC \mid NCIG) \times \mathrm{P}_{\mathrm{Man}\,(HC)}\,(HC),$$

so with a little arithmetic $\mathrm{P}_{\mathrm{Man}\,(HC)}\,(LC) = \mathrm{P}(LC)$. This accords with our intuition that manipulating a cause of *LC* changes the distribution of *LC*, but manipulating an effect of *LC* does not. (Formulation of this theorem in the form presented here is found in Spirtes, Glymour and Scheines 1993; and is stated in a form that does not use directed acyclic graphs in Robins 1986; and Wold and Strotz 1971.)

If the causal structure is not known, how can it be discovered? One approach would be to do a randomized experiment. A random sample could be taken from the actual population, and using some random device people could be assigned to a control group, who are forced to smoke a given number of cigarettes, or a treatment group, who are forced not to smoke. These two samples are just like samples from a hypothetical population in which everyone is forced to smoke a given number of cigarettes, and a hypothetical population in which everyone is forced not to smoke. This reduces the problem of determining the effect of *NCIG* on *LC* to a problem of inference about a population distribution from a sample distribution, which is the same question considered in §2.

In many cases, a randomized experiment cannot be performed, either for practical or for ethical reasons. The local directed Markov condition associates a given causal structure with a set of population probability distributions. If it is inferred from a sample that the population distribution does not lie in the set of distributions compatible with a given causal structure *G*, it could be concluded that *G* is false. This strategy obviously suffers all the problems of inferences about a population probability distribution from a sample distribution that were discussed in §2, but there are other problems in addition. If it is inferred from a sample that the population distribution does lie in the set of distributions compatible with a given causal structure *G*, it cannot be concluded that *G* is true, because the population distribution might also lie in the set of distributions compatible with other causal structures. For example, in the discrete case, every distribution that is compatible with any of the following three causal structures, is also compatible with the other two. Background information, such as time order, may be able to eliminate some of the alternatives, but if the background information is not available, using only the statistical data it is not possible reliably to decide between the three alternatives.

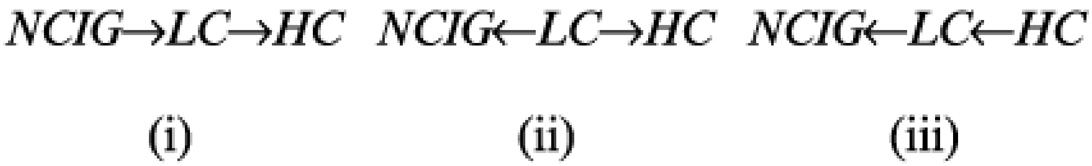

Figure 3

When the possibility of unmeasured common causes is considered, then the problem is even worse. For example, any distribution compatible with the causal models of Figure 3 is also compatible with either of the models depicted in Figure 4, as well as many others, where the 'T' variables are unmeasured.

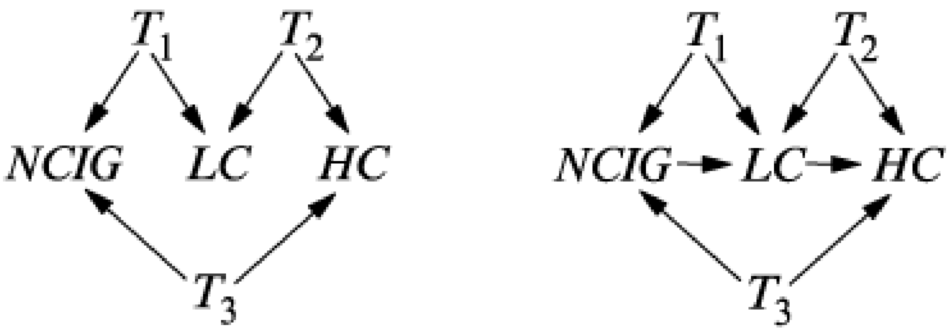

Figure 4

Again, strong background knowledge to the effect that there are no unmeasured common causes would eliminate the causal models in Figure 4 as a possibility. Another way to eliminate models such as those in Figure 4 is to make the (controversial) assumption that if the population distribution satisfies a given factorization, then it is entailed to satisfy that factorization for all parameterizations of the graph by the local directed Markov condition; call this the 'faithfulness' condition. Thus while there exist some parameterizations of the graphs in Figure 4 that can be factored according to equation 2, applying the local directed Markov condition to either of the graphs in Figure 4 does not entail that all parameterizations of these graphs can be factored according to equation 2. Given the assumption that a probability distribution is the marginal of a distribution that satisfies the local directed Markov condition and the faithfulness condition for the causal graph *G* that generated it, it is possible in some cases to narrow the set of graphs compatible with a given population distribution sufficiently to determine that *X* is a cause of *Y*, without knowing which occurred first. There is active research into devising algorithms for the automatic generation of causal models from sample statistics, background knowledge, and principles like the Markov condition, the faithfulness condition, and weaker assumptions.

For example, suppose that the causal structure *G* shown in Figure 5 generated a probability distribution P that satisfied the local directed Markov condition and the faithfulness condition for *G*, and that one knew P, but not *G*. If one admits the possibility of unmeasured common causes of variables, then P satisfies the local directed Markov condition and the faithfulness condition for an infinite number of different acyclic causal structures; however, all of those causal structures contain a directed path from *Z* to *W*, so it can be concluded from the local directed Markov condition, the faithfulness condition, and P, that *Z* is a cause of *W*. But in practice, one has knowledge of a sample distribution rather than the population distribution P, which makes the inference more difficult.

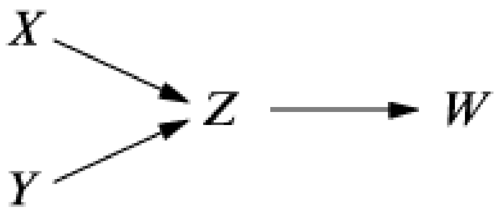

Figure 5

See also: Statistics

PETER SPIRTES

References and further reading

Bentler, P. (1989) *EQS: Structural Equations Program Manual*, Los Angeles, CA: BMDP Statistical Software Inc.(The user's manual for a statistical programme widely used in the social sciences, with a number of examples.)

Blalock, H.M. (1969) *Theory Construction: From Verbal to Mathematical Formulations*, Englewood Cliffs, NJ: Prentice Hall.(An elementary introduction to formulation social science models.)

Blalock, H.M. (ed.) (1971) *Causal Models in the Social Sciences*, Chicago, IL: Aldine-Atheston.(Collection of articles on problems of causal inference in the social sciences.)

Bollen, K. (1989) *Structural Equations with Latent Variables*, New York: Wiley. (Advanced undergraduate text describing construction, estimation, and testing of structural equation models.)

Freudenthal, H. (1965) *Probability and Statistics*, New York: Elsevier Publishing Co. (Undergraduate level introduction to statistics and probability.)

Fisher, R. (1990) *Statistical Methods, Experimental Design, and Scientific Inference*, Oxford: Oxford University Press.(A collection of classic works on experimental design.)

Lindley, D. (1985) *Making Decisions*, New York: Wiley.(Elementary introduction to Bayesian decision theory.)

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems*, San Mateo, CA: Morgan and Kaufman.(Explains recent advances in using Bayesian techniques in expert systems.)

Robins, J. (1986) 'A New Approach to Causal Inference in Mortality Studies with Sustained Exposure Periods - Application to Control of the Healthy Worker Survivor Effect', *Mathematical Modeling* 7: 1393-512.(Contains a non-graphical version of the manipulation theorem.)

Spirtes, P., Glymour, C. and Scheines, P. (1993) *Causation, Prediction, and Search*, New York: Springer Verlag Lecture Notes in Statistics, 81.(Graduate level text on causal inference from statistical data.)

Wold, H. and Strotz, J. (1971) 'Recursive versus Nonrecursive Systems: An Attempt at a Synthesis', in H. Blalock (ed.), *Causal Models in the Social Sciences*, Chicago, IL: Aldine-Atheston.(Contains a version of the manipulation theorem for linear systems.)

Steiner, Rudolf (1861-1925)

Rudolf Steiner held that humanity has passed through an astral and an etheric stage and has possessed intuitive and clairvoyant modes of consciousness. People, he held, once enjoyed psychic powers and existed in forms of matter more rarefied than those that characterize their current successors. Nevertheless, people still exist on four levels, the physical, the etheric, the astral and the spiritual, and earlier psychic skills are retrievable. Steiner, who called this theory 'anthroposophy', was particularly influenced by Goethe and theosophy.

Born on February 27, 1861, Rudolf Steiner studied from 1890 to 1897 at the Goethe-Archiv in Weimar, and from 1897-1900 was the editor of *Magazin: Monatschrift für Literatur (Magazine: Monthly Writings on Literature)*. In 1902, Steiner became general secretary of the new German section of the Theosophical Society (see Theosophy §2). He came to disagree with Annie Besant and Madame Blavatsky over attempts to build a 'spiritual science' on Eastern mysticism and over the promotion of Jiddu Krishnamurti as a messiah. In 1913, he founded the Anthroposophical Society. He died in 1925.

Steiner was initiated into an occult tradition by an adept whom he called 'The Master', and at the age of forty he began to speak publicly about what he believed to be clairvoyant experiences that had occurred to him since he was eight. He accepted the doctrines of reincarnation and karma, and the Hermetic teaching that humankind is the microcosmic clue to the macrocosm, so that discovering the true nature of humanity would also lead to a knowledge of the secrets of the universe. In 1896, he wrote his major work *Die Philosophie der Freiheit: Grundzuge einer modernen Weltanschauung (The Philosophy of Freedom: Essentials of a Modern Outlook)*, which contends that mechanistic science gives us only an abstract knowledge of natural uniformities. His model for a fuller human knowledge than that provided by mechanistic science was that of Goethe's 'primal plant', a self-evolving, self-developing organism (see Goethe, J.W. von §3). He claimed to have access to the Akashic records, which theosophists hold to be master records of all that has occurred since the beginning of the universe. These records allegedly exist in the form of impressions in the astral plane. Here, Steiner reported, he learned that humanity once possessed supranormal capacities that have since been lost.

Steiner taught that Christ came at the low point of the human descent into materialism and provided the opportunity for reascent. Uniting (South Asian) Indian and theosophical ideas with his interpretation of Christianity and the contents of his clairvoyant experiences, he constructed what he conceived of as a correct combination of modern science and religion. He reported perceiving nonhuman intelligences who affect human consciousness, some promoting spiritual advancement and some acting on behalf of human belief in a mechanistic materialism.

Steiner held that human nature has four levels: the physical body, the etheric body, the astral or soul body, and the spirit. The notion of an etheric body is connected with the notion of the aura, thought of as an envelope of vital energy that radiates from everything in nature and possesses a magnetic field. The etheric body is regarded as part of the aura, which is accessible only clairvoyantly. The notion of the astral body also links up with the notion of the aura, astral bodies also being said to be parts of auras. The astral body is viewed as the seat of emotions and as the carrier of consciousness in out-of-body and near-death experiences.

To each level of human nature, Steiner taught, there corresponds a level of human cognition: to the physical body corresponds perception; to the ethereal body, imaginative knowledge; to the astral body, inspirational knowledge; to the spirit, intuitive knowledge. He alleged that there are etheric forces at work among minerals and plants, and that awareness of these forces - through imaginative knowledge - provides the basis for a new science of 'biodynamic forming'. Inspirational knowledge is understood as knowledge of one's astral body, thought of as capable of existing independently of one's physical body and as what makes thought, memory and a sense of personal identity possible. Having intuitive knowledge involves experiencing one's entire life backwards, and, he held, includes encountering beings of higher intelligence.

KEITH E. YANDELL

List of works

Steiner, R. (1896) *Die Philosophie der Freiheit: Grundzüge einer modernen Weltanschauung (The Philosophy of*

Freedom: Essentials of a Modern Outlook), Dornach: Philosophisch-Antroposophischer Verlag Goetheanum; trans. R. Stebbing and ed. P.M. Allen, *The Philosophy of Spiritual Activity*, West Nyack: Rudolph Steiner Publications, 1963.(Originally published in English in 1916; argues that mechanistic science gives only formal knowledge and needs supplementation by Goethean epistemology.)

Steiner, R. (1925) *Mein Lebensgand*, Dornach: Schweitz; *The Course of My Life*, Blauvelt: Rudolph Steiner Publications, 1977.(Steiner's autobiography.)

Steiner, R. (1950) *Anthroposophische Grundlagen für die Arzneikunst (Foundations of Anthroposophy for the Art of Medicine)*, Dornach: Schweiz; *Anthroposophy: An Introduction*, London: Rudolph Steiner Press, 1983. (Four lectures given in Stuttgart in October 1922, forming an introduction to Steiner's adaptation of theosophy.)

Steiner, R. (1984) *The Essential Steiner*, ed. R.A. McDermott, San Francisco, CA: Harper & Row.(Steiner's publication list contains over two hundred and fifty items; this selection offers a manageable starting point and information as to how to proceed in following up various topics on which he wrote.)

References and further reading

Davy, J. (ed.) (1975) *Works Arising from the Life of Rudolph Steiner*, London: Rudolph Steiner Press.(Articles published to mark the fiftieth anniversary of Steiner's death.)

Stevenson, Charles Leslie (1908-79)

Stevenson's major contribution to philosophy was his development of emotivism, a theory of ethical language according to which moral judgments do not state any sort of fact, but rather express the moral emotions of the speaker and attempt to influence others.

What do we mean when we say that something is good or bad? On the face of it, we are describing, attributing to the thing some property, goodness or badness. What could these properties be? How do we find out about them? Much of philosophical moral theory explores various answers to these questions. Stevenson thought that questions about the nature of moral properties were misplaced. Our moral judgments do not, at least primarily, describe at all. Uttering moral sentences has a different function: to express emotions, and to influence or invite others to share them. All of his main contributions appeared in *Ethics and Language* (1944) and a collection of papers, *Facts and Values* (1963).

If I say 'Ann Arbor is in Michigan', I express my belief that Ann Arbor is in Michigan, but I do not say that I believe such a thing. For what makes what I said true? Not that I really do believe that Ann Arbor is in Michigan; only the fact that Ann Arbor really is in Michigan. Stevenson's theory of ethical language, in a nutshell, was that when I say 'Inequality is bad', I have expressed a certain negative moral attitude towards inequality, though I have not said that I have it.

Stevenson's main argument for his theory was that '[a] person who recognizes X to be "good" must ipso facto acquire a stronger tendency to act in its favor than he otherwise would have had' (1963: 13). Following Hume, Stevenson thought this showed that the 'belief' that something is good must really be no belief at all, but an emotive attitude, since beliefs do not of themselves give us any tendency to act (see Hume, D. §§3-4).

Besides expressing the speaker's attitude, Stevenson said, moral statements also 'create an influence', they invite the audience to share in the emotion expressed. Thus, 'x is good' is akin to 'Let us approve of x'. And, in context, ethical statements can come to have secondary descriptive content; in Victorian England, for example, calling a woman 'virtuous' implied that she was chaste.

Stevenson's theory was enormously influential. Taking its cue from Ayer's brief remarks on ethics in *Language, Truth and Logic* (1936), Stevenson's theory added sophistication and subtlety (see Ayer, A.J.). His view has important affinities both with the work of R.M. Hare and with some later theories, all of which share the central tenet that moral judgments express not beliefs but conative attitudes (see Prescriptivism). Stevenson in particular stressed that his work did not include any substantive moral judgments, but rather comprised 'analytic ethics', or what is now commonly called 'meta-ethics', the branch of moral theory that is *about* ethics and ethical language (see Analytic ethics). For by claiming that moral judgments serve to express emotions, he had not expressed his own moral emotions at all.

See also: Emotivism; Emotive meaning; Moral judgment; Moral scepticism; Morality and emotions

JAMES DREIER

List of works

Stevenson, C.L. (1944) *Ethics and Language*, New Haven, CT: Yale University Press.(Stevenson's only published monograph.)

Stevenson, C.L. (1963) *Facts and Values: Studies in Ethical Analysis*, New Haven, CT: Yale University Press.(A collection of papers. Essay 2 is especially useful as an introduction; essay 11 contains the most mature version of the theory.)

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz; 2nd edn, 1946, ch. 6.(Contains the germ of emotivism, which Stevenson developed into a full and sophisticated theory.)

Blackburn, S. (1984) *Spreading the Word*, Oxford: Oxford University Press, esp. ch. 6.(Develops a variation on emotivism designed to address prominent objections.)

Geach, P.T. (1960) 'Ascriptivism', *Philosophical Review* 69 (2): 221-5.(An influential, but somewhat technical,

objection to emotivism.)
Goldman, A. and Kim, J. (eds) (1978) *Values and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt*, Dordrecht and Boston, MA: D. Reidel.(A collection of critical essays; those on Stevenson's work are quite accessible. Contains a comprehensive bibliography of Stevenson's writing.)
Hare, R.M. (1952) *The Language of Morals*, Oxford: Oxford University Press.(A work contemporary with Stevenson's, with important similarities and contrasts.)

Stewart, Dugald (1753-1828)

Dugald Stewart was, after Thomas Reid, the most influential figure in the Common Sense School; he was a major influence on Victor Cousin and Théodore Jouffroy in France and on most academic philosophers in the United States. Along with Reid and Cousin, Stewart made the Scottish tradition the dominant philosophy in America for half a century. His Elements of the Philosophy of the Human Mind and Philosophy of the Active and Moral Powers of Man were his most important works and went through a number of printings. The abridged edition of his Active and Moral Powers was reprinted ten times from 1849 to 1868.

Stewart followed Reid in claiming that any philosophy which contravenes the principles of common sense must be false, and the problem is to discover and eliminate the premise which yields such results. He added the requirement that philosophical propositions must not change the meanings *of concepts in ordinary life, and he also added a new dimension to Reid's agency theory. More than any other writer he emphasized correctly the epistemic similarities between Reid and Immanuel Kant, but he followed Reid in avoiding Kant's distinction between phenomena and noumena.*

Stewart disagreed with Reid in avoiding the phrase 'principles of common sense' as misleading, rejected his mentor's realistic interpretation of universals and provided his own nominalistic alternative. He also modified to some extent, though quite cautiously, Reid's rigid inductivism and made some concessions to a realistic interpretation of scientific hypotheses. Stewart was equipped to discuss issues in the philosophy of science since he was well versed in mathematics and physics, having been professor of mathematics at Edinburgh for ten years before being named professor of moral philosophy. Stewart was arguably the first and finest philosopher of science in the Scottish tradition.

1 Background

Dugald Stewart was educated at the University of Edinburgh and in 1771 went to Glasgow, primarily to attend the lectures of Thomas Reid. He became, first, professor of mathematics and, ten years later, professor of moral philosophy at Edinburgh; his philosophical work, published in eleven volumes by Sir William Hamilton, shows him to have been no carbon copy of Reid, but a philosopher who added important new dimensions to the Scottish tradition and occasionally criticized and departed from it.

2 Reid's common-sense view

Thomas Reid formulated a number of what he called principles of common sense, including such judgments as that objects are perceived as in space, and events are perceived as in time. Space and time are not concepts learned from experience, since 'object' presupposes 'space' and 'event' presupposes 'time'. Thus space and time constitute nativistic epistemic input possessed by perceivers themselves. Other examples of nativistic input include intuitions to the effect that every event has a cause, qualities are directly perceived as properties of objects, objects exist independently of being known, and agents in some cases act freely. Principles of common sense can be identified (not proven) because they are universally held and are unavoidable in the sense that denying them is pointless because in so doing one never gets rid of them. The philosopher who momentarily denies them reaffirms them in the marketplace or even while denying them acts upon them unwittingly. That David Hume pointed out this state of affairs was, for Reid, a very honest if self-defeating thing for a sceptic to do. Because the knowledge incorporated in the principles is nativistic (though activated by appropriate experiences) and because the principles are universal and unavoidable, they always take precedence over any discursive argument which contravenes them. When philosophy comes into conflict with common-sense principles, so much the worse for philosophy. Any philosophy which contravenes the principles must contain a fallacy and should be rejected.

Reid rejected scepticism because it violated the principles of common sense, not only claiming that their truth-values are not reliable but that even if they were one could not know it to be the case. Absolute idealism would have presented Reid with an even more standard example of the rejection of his principles. Absolute idealists undermine the truth-values of ordinary propositions with the concept of internal relations. They do not deny the ordinary meanings of 'physical object', 'space' and 'time', but insist that such entities are never exemplified except as misleading appearances. Since everything is internally related there are no separate entities,

and since ordinary statements imply the existence of such entities such statements cannot be true of reality, only of appearances. As with the sceptic, the problem for the idealist, Reid thought, is to discover the premise that leads to such absurdities and reject it.

3 Stewart's contributions to this tradition

Stewart wholly agreed with Reid's Common Sense metaphilosophy, according to which sceptics and absolute idealists who violate the truth-values of ordinary statements are *ipso facto* mistaken. However he did not think that Reid succeeded in showing that George Berkeley rejected the truths of common-sense principles and he felt that a new dimension needed to be added to Reid's metaphilosophy. Stewart was impressed with Berkeley's insistence that his views squared completely with common sense. 'The cup is on the table' remains true, if true to begin with, whether the cup and table are construed as independently existing objects or as clusters of sensations. Or, again, Kepler's laws of planetary motion remain what they are if the planets are interpreted as independent objects or clusters of sensations in God's mind. Stewart concluded that Berkeley had escaped Reid's net, though he thought the good bishop contravened common sense by violating the meaning of ordinary terms. What we need, he believed, is an accurate analysis of the meaning of the word existence,

> which analysis would have at once shown, not only that we are irresistibly led to ascribe to the material world all the independent reality which this word expresses, but that it is from the material world that our first and most satisfactory notions of *existence* are drawn.
>
> ([1854-60] 1877, vol.3: 54)

Hence both the meaning-value and the truth-value of ordinary propositions are inviolable and cannot be contravened by philosophical analysis.

Stewart's remarks on Berkeley seem well taken and yield a supplementary ingredient to Reid's metaphilosophy, namely that it is just as illegitimate for a philosophical position to violate the *meanings* of ordinary propositions as it is to contravene their truth-value. He formulated this supplementary criterion very clearly and distinctly but unfortunately did not consistently apply it in his subsequent discussion of the meaning of 'causality' and Hume's denial of natural necessity. Even though he did not consistently maintain the meaning-inviolability of common sense principles, he succeeded in making the point in an admirably clear and telling way. With the exception of Asa Mahan at Oberlin College and some others, subsequent members of the tradition omitted to take account of Stewart's criterion. Indeed, even G.E. Moore, much later, failed entirely where Stewart had succeeded, by formulating only the Reidian truth-inviolability criterion. Moreover, it must be kept in mind that Stewart indicated that philosophical positions often result from semantical misadventures, and that philosophers often argue from misleading analogies, and that such claims are essentially related to, and reflect, his meaning-inviolability criterion. The fact remains that Stewart, above all others in the Scottish tradition, stressed the importance of meaning analysis, and was outdone in this respect only in the twentieth century, by various strands of analytic philosophy.

Stewart contributed to the Common Sense tradition in still another way. He saw more clearly than most others the striking similarity of Kant's transcendental aesthetic to Reid's views on space and time. Kant and Reid had the same goals, namely to show the inadequacy of John Locke's empirical construal of space and time (or the inadequacy of any empirical construal), and to attribute the experience of objects in space and time to laws or forms inherent in the structure of the human intellect itself (see Kant, I. §§3-7; Locke, J. §§2-6). These fundamental laws or forms, of course, are not independent of experience, since they are awakened into their proper functioning only by the occurrence of a contingent relevant sensation. Stewart concluded:

> This is precisely the ground on which Reid has made his stand… and I leave it to my readers to judge whether it was not more philosophical to state… the *fact*, in simple and perspicuous terms, as a demonstration of the imperfection of Locke's theory, than to have reared upon it a superstructure of technical mystery, similar to what is exhibited in the system of the German metaphysicians.
>
> ([1854-60] 1877, vol. 5: 118)

Unless, like Stewart, one sees the epistemic similarities of Reid and Kant (discounting their very great metaphysical dissimilarities), one will not be able to understand the American philosophical movement called

transcendentalism, with Ralph Waldo Emerson and Theodore Parker sharing nativistic epistemic points of departure but developing totally different metaphysical views. From the initial intuitionistic framework, Emerson went towards absolute idealism, while Parker embraced the Scottish realist tradition of Reid and Stewart. The two Americans did not disagree on the importance of intuitive truth; they simply had incompatible intuitions. Stewart would have thought Emerson's intuitions unacceptable because they were not universal and unavoidable.

A further contribution made by Stewart to the Scottish tradition was in the area of agency theory. He argued that the concept of cause is not applicable to human actions: in general people act from motives which explain the point of what they do but do not cause them to do it. Agents are the causes of their own actions, guided by motives. Stewart accepted that all actions have motives, but stressed that 'motive' is a metaphorical way of referring to agents' intentions or purposes in performing acts and not to substantive factors that necessarily cause their actions. His important contribution, however, was to reject any need for criteria to distinguish between cause-effect and motive-act explanations. He rejected James Gregory's proposal that causal relations hold universally whereas motive-act relations cannot be universalized. In addition, he rejected the criterion that causes and effects can be described independently of each other while motives and acts cannot since they are logically or conceptually connected. As for his general argument against the need for any criterion whatever, Stewart looked upon a criterion as a sign that something else is the case and thought a criterion unnecessary since agent-causality is *directly attested to* by consciousness. Consciousness attests to the *fact* (Stewart's emphasis) that people can decide among contending motives, can initiate an act, and could have done other than they did; and such agent causality is obviously not applicable to the physical world.

Stewart's claim that the concept of causality is not applicable to human agency (though it may be applicable to abnormal behaviour) has the great merit of clarity and succinctness as well as philosophical insight. Moreover, his rejection of criteria for distinguishing causal and motive-model explanations seems a praiseworthy addition to the tradition. Why should it be necessary to provide a *criterion* for what is *evident*?

4 Stewart's deviations from this tradition

Stewart was critical of Reid's use of the phrase 'principles of common sense'. The use of this phrase, he felt, is easily misunderstood. 'Common sense' is a phrase that already has a fixed meaning in ordinary parlance which is different from Reid's, and the result is constant misrepresentation. In ordinary language, 'common sense' ordinarily denotes a type of wisdom different from what can be acquired through education. To claim, then, as Reid did, that a philosophical analysis which contravenes the principles of common sense is thereby nullified, is to suggest, unfortunately, that learned philosophers are to be governed by the voice of the general population. As Stewart pointed out, Reid did not of course intend this outcome, but his use of 'common sense' encouraged it. Surely there must be some more appropriate name for the fundamental laws of human belief?

Stewart's criticism of Reid's use of 'common sense' seems plausible and, in any case, influenced the later Scottish realists. His criticism amounts to saying that Reid gave a technical definition of the expression 'common sense' and one that, in fact, violates its ordinary meaning. Subsequent figures in the Scottish tradition, among them Victor Cousin, Henry Tappan, Asa Mahan, and James McCosh, apparently agreed with Stewart's criticism, since they all used different locutions for Reid's fundamental principles - common consciousness, basic intuitions, and fundamental laws of human belief.

Stewart was also critical of Reid's realistic interpretation of universals. Philosophers of distinction, Stewart wrote, have managed to reject the messenger theory of perception, but unwittingly cling to Locke's theory of ideas in other operations of the mind. When Richard Price and Reid abstracted from particulars certain properties shared in common and then generalized such abstractions into classes, they were committing themselves to universals which either exist as entities of some sort or exist as images in the minds of those who use propositions containing such universals. Stewart was unimpressed because people are no more introspectively aware of the existence of universals apart from particulars than they are aware of impressions apart from the object perceived. He had an elaborate analysis to show that there is no reason to believe in the existence of anything in the mind distinct from the operations of the mind itself. The correct view, he maintained, is the nominalistic one, according to which general terms refer wholly to sets of particulars which exemplify them - any one of which in a given set can be given as an example or paradigm of the general term being used. Stewart put forward a nominalistic analysis that depended upon signs and language. Universals, he thought, refer to general signs wholly linguistic in nature, and

thus require no ontological or psychological entities as referents. Reasoning involving universals is sign-reasoning, in which general terms have no other existence than the linguistic one shared with all other parts of a well-constructed language.

Stewart's discussion of the nature of universals is a felicitous attack on Locke's theory of ideas; and Stewart's nominalism, phrased in terms of linguistic signs, is not an unsophisticated viewpoint - particularly his notion that the significance of a general term conceived as a sign consists in its successful use as a memory device that conserves experiences individually learned (and can always be checked, upon demand, by paradigm instances of its denotation). Certainly such a view eliminates any necessity of referring general terms to images in the mind, as did Reid. However, Reid apparently did not think that images in the mind were substantive entities like Locke's ideas or Hume's impressions. If not, was Reid's view that just as we should substitute 'sensing' for 'sensation', we should substitute 'imaging' for 'image'? Stewart was not sure, and wiped the board clean by giving a linguistic analysis of universals.

A further way in which Stewart departed from Reid and enriched the Scottish tradition was to weaken, albeit hesitantly, Reid's rigid inductivism. Reid interpreted Newton's '*hypotheses non fingo*' literally, and, indeed, found most of the mistakes in physics and philosophy to be based on unfortunate hypotheses, such as Descartes' vortices and Locke's intervening messengers. But according to Stewart, we must not, after all, take Newton's '*hypotheses non fingo*' too literally. Did not Newton himself suggest numerous hypotheses in his *Optics* and assume that his readers would understand that his 'Queries' had experimental consequences and thus could be scientifically tested? They are quite unlike Descartes' vortices, which are defended wholly on a priori grounds (see Descartes §11). It is likely that Newton, in damning the use of hypotheses, referred only to those like Descartes' vortices that are immune to falsification. He approvingly quoted Boscovich's view that hypotheses or theories are necessary anticipations of solid scientific advances since there can be no relevant observation without hypotheses as a guide for where to look and what to look for.

Notwithstanding the aforesaid, Stewart makes it clear that he is not endorsing the use of hypotheses in science in any way except as heuristic devices. Principles of science are not established until they have been fairly inferred from undoubted facts and not from the uninterrupted confirmation of the consequences of hypotheses. Stewart's reservation about giving a realistic interpretation to scientific hypotheses was simply the familiar criticism that the confirmation of hypotheses commits the fallacy of affirming the consequent (see Fallacies). This issue was the key one among anti-Copernican astronomers during the Renaissance. They wanted to interpret all astronomical hypotheses as mathematical fictions. Nicolas Reymer Baer in 1597 publicly defended a nonrealistic theory of astronomical hypotheses. Reymer had expounded the nonrealistic theory in its most extreme form: not only were astronomical hypotheses merely mathematical fictions, they would not be hypotheses if they were true! In his monumental work *Apologia Tychonis contra Nicolaum Ursum*, Kepler rejected this view, claiming that a logical fallacy is not a scientific mistake and that the available confirmation overwhelmingly allows a realistic interpretation of the Copernican hypothesis. And Galileo emphatically agreed with Kepler. Stewart could not set himself against such scientific and methodological stalwarts and agreed that the evidence in Galileo's day already justified his realistic claim, which science in the future continued to support even in more direct ways. But even at this point Stewart claimed that it was not only that Copernicus correctly predicted all the facts that made the acceptance of his hypothesis legitimate even in his own day, but that there were various analogies between sources of illumination that also made the hypothesis acceptable. But of course Stewart could not deny that analogies themselves are a form of hypothesis.

Stewart's dependence upon Boscovich further confirms the attraction that hypotheses held for him. Boscovich himself argued against an inductive view of science, since his own physical theory had concepts that referred to unobservable events. Since Stewart accepted Boscovich's methodology he again committed himself by implication to the legitimacy of hypotheses with unobservables as long as they had testable consequences. Stewart was drawn to Boscovich's physical theory of point centres. He was also drawn to Boscovich's argument that corpuscularian theory is incoherent, because compressible bodies can interact by contact since they do not instantaneously acquire a common velocity but by mutual compression are enabled to slow down or accelerate in a finite time; however, this concept of interaction cannot be applied to the ultimate corpuscles because they are absolutely incompressible. Hence there can be no transfer of motion between ultimate particles. Stewart was reassured by this argument because it undercut the atomistic materialism of David Hartley and Joseph Priestley. However, such reassurance

makes no sense unless Stewart acknowledged the existential import of the corpuscularian hypothesis. It would be absurd to accept, as Stewart did, Boscovich's theory of point centres of mutual influence and the corpuscularian theory as legitimate competitors and yet deny on the methodological level that a system with reference to unobservables is ever justified.

Stewart was perfectly aware of the tensions in his own thinking. On the one hand, he was a thorough inductivist in philosophy and psychology and respected Bacon's methodological dicta. On the other hand, he was sensitive to the realistic interpretations of Kepler and Galileo, who saw that hypotheses were not simply instrumental devices. He was pulled both ways, and his work emerges as a significant connecting link between the naïve inductivism of Reid and the later technical views on scientific theory and causality held by James McCosh (see McCosh 1867), the last first-rate figure in the Scottish tradition.

See also: American philosophy in the 18th and 19th centuries; Common Sense School; Commonsensism; Theories, scientific

EDWARD H. MADDEN

List of works

Stewart, D. (1854-60) *Collected Works*, ed. W. Hamilton, Edinburgh: Constable; ed. W. Hamilton, Edinburgh: Clark, 1877.(Of the eleven volumes in the *Collected Works*, the best known are *Elements of the Philosophy of the Human Mind* (3 vols), and *Philosophy of the Active and Moral Powers of Man* (2 vols).)

Stewart, D. (1803) *Account of the Life and Writings of Thomas Reid*, Edinburgh: Creech.

Stewart, D. (1841) *Esquisser de philosophie morale (Sketch of Moral Philosophy)*, Paris: Johanneau.

Stewart, D. (1868) *Philosophy of the Active and Moral Powers of Man*, ed. J. Walker, Boston, MA: Phillips Sampson.(Walker, of Harvard, abridged Stewart's book and improved it pedagogically by placing the long appendix on agency theory appropriately in volume 1, book. 2.)

References and further reading

Griffin-Collart, E. (1980) *La Philosophie écossaise du sense commun: Thomas Reid et Dugald Stewart (The Scottish Philosophy of Common Sense: Thomas Reid and Dugald Stewart)*, Bruxelles: Palais des Académies.

Madden, E.H. (1968) *Civil Disobedience and Moral Law*, Seattle, WA: University of Washington Press.(Portrays the dominance of Reid, Stewart, and Cousin in nineteenth-century American philosophy.)

Madden, E.H. (1981) 'Asa Mahan's Analysis of Synthetic Apriori Judgments', *Transactions of the C.S. Peirce Society* 17: 297-318.(Shows the similarities and differences between the Scottish tradition and Kant in a way that makes Stewart's efforts in this direction understandable.)

Madden, E.H. (1986) 'Stewart's Enrichment of the Commonsense Tradition', *History of Philosophy Quarterly* 3 (1): 45-63.(Considers in greater detail aspects discussed in §§2-3, giving a modified account.)

McCosh, J. (1867) *Intuitions of the Mind*, New York: Carter.(Provides the realistic interpretation of science for which Stewart was groping.)

McCosh, J. (1875) *The Scottish Philosophy, from Hutcheson to Hamilton*, New York: Carter.(Contains a rewarding chapter on Stewart.)

Meyer, D.H. (1974) *The Instructed Conscience*, Philadelphia, PA: University of Pennsylvania Press.(Portrays Stewart as 'the great systematizer' and shows the impact of the Scottish tradition in shaping the American national ethic.)

Olson, R. (1975) *Scottish Philosophy and British Physics, 1750-1880*, Princeton, NJ: Princeton University Press.(Argues that Stewart saw geometry as more basic than algebra and hence influenced British physicists such as James Clerk Maxwell and William M. Rankine.)

Robinson, D. N. (1989) 'Thomas Reid's Critique of Dugald Stewart', *Journal of the History of Philosophy* 27: 405-22.(An examination of Reid's critique of a pre-publication draft of Stewart's *Elements of the Philosophy of the Human Mind*.)

Stirner, Max (1806-56)

Max Stirner is the author of Der Einzige und sein Eigentum (The Ego and Its Own), first published in Germany in 1844 and best known for its idiosyncrasies of argument and idiom. Stirner condemns modernity as entrenched in religious modes of thought and envisages a positive egoistic future in which individuals are liberated from the tyranny of those ideas and social arrangements which restrict autonomy. The Ego and Its Own was an impulse to the decline of the Hegelian left as a coherent intellectual movement, and played an important role in the genesis of Marxism; Stirner has also been variously portrayed as a precursor of Nietzsche, an individualist anarchist and a forerunner of existentialism.

Stirner (born Johann Caspar Schmidt on 25 October 1806 in Bayreuth, Germany) had a largely unpropitious start to adult life, passing through university without distinction, before becoming a teacher at a respectable private girls' school in Berlin. However, in his spare time he became increasingly involved with 'the free' (a group of left Hegelians led by Bruno Bauer), before leaving his teaching post in 1844 and publishing *Der Einzige und sein Eigentum (The Ego and Its Own)*, his most important and influential work. Following a brief period of unremunerative notoriety, Stirner settled into a somewhat indigent and solitary lifestyle. Hack journalism, translation work and (before she left him) his second wife's dwindling inheritance failed to avert two brief spells in a debtors' prison. Stirner contracted a fever after an insect bite on his neck, and died on 25 June 1856.

Insisting on the relativity of rationality, truth and language, Stirner rejects traditional forms of expression and modes of exposition. The consequent employment of aphorism and metaphor, neologisms, relentless paronomasia and juxtaposition of words with formal similarities or related meanings to present his views makes *The Ego and Its Own* a stylistically striking, if idiosyncratic, text. The only limitation which Stirner places upon use of language and mode of argument is that it should serve our individual ends.

The Ego and Its Own is structured around a tripartite division of human experience into categories of 'realism' (where individuals are dominated by external forces), 'idealism' (where individuals are dominated by ideas), and 'egoism' (where individuals escape domination by dealing with things and ideas as they wish, setting their personal satisfaction above all else). This tripartite division is elaborated and exemplified in accounts of individual development (in which childhood, youth and adulthood correspond to stages of realism, idealism and egoism), of human history (in which the ancient, Christian and future worlds are portrayed as epochs of realism, idealism and egoism respectively) and in a racial (and racist) analogue of that historical account (in which the Caucasian race passes through the realist stage of 'Negroidity' and the idealist stage of 'Mongolism' before reaching the 'truly Caucasian' future in which the race is liberated from the hegemony of natural forces and ideas).

Part One aims to demonstrate that modernity, the epoch of idealism, fails to escape from that which it claims to have outgrown, namely religious modes of thought. Thus Stirner rejects the contemporary consensus that Ludwig Feuerbach (§2), the leading figure of the Hegelian left, had completed the critique of religion (see Hegelianism §2). For Stirner, the errors of religion are not overcome by a rejection of God as transcendent subject, but by opposing the subordination of the individual to 'spirit' in any form. Feuerbach's anthropological reduction had not revealed human nature as it was, but rather deified an abstract and prescriptive account of what being human involved, thus leaving the 'real kernel' of religion, the positing of an 'essence over me', intact. This continued immurement within religious thinking is portrayed as paradigmatic of modernity. Liberals, socialists (including the young Karl Marx) and contemporary critics of all kinds are convicted of separating individuals from their real natures and positing some fictitious essence as their goal. In contrast, Stirner seeks to rehabilitate the concrete and diverse 'un-man' (*unmensch*) in each of us, to whom this 'foolish mania to be something else' is completely alien.

Part Two elaborates Stirner's conception of egoism by relating it to 'ownness' (*Eigenheit*), a form of substantive individual autonomy valued above all else. 'Ownness' is violated by any desire or action which involves waiving or suspending individual judgment, and thus conflicts with any conception, however grounded, of obligation or duty. Even the legitimacy of self-assumed obligations, incurred, for example, by the act of promising, is denied. Perhaps most importantly (not least in establishing his anarchist credentials), Stirner maintains that a relationship of absolute hostility exists between the individual and the state, based on the incompatibility of 'ownness' and any obligation to obey the law. The notorious adjunct is that crime is applauded as an assertion of 'ownness' against its

chief usurper, weakening the ‘cement’ (respect for law) which holds the state together.

Morality, defined by its positing of duties on the individual, is also sacrificed to ‘ownness’. For those (including Stirner) who dispute the idea of an exclusive opposition between morality and immorality, it does not follow that the egoist is immoral. Nor is Stirner inconsistent in stressing the evaluative superiority of egoism over other modes of experience and action. Stirner’s rejection of morality is grounded in an affirmation of the nonmoral good of egoism and not in a repudiation of values: that is, he allows a realm of actions and desires, which although not moral (because they involve no obligations to others) are still to be assessed positively. Stirner’s conception of morality is in this sense a narrow one, and his rejection of its claims is not coextensive with a denial of the validity of all evaluative judgment.

The escape from social relations which conflict with ‘ownness’ is not the end of all contact between persons, only the end of binding rules for resolving conflicts between competing interests, and the end of constraints, other than ‘ownness’, upon the pursuit of individual enjoyment. Stirner consistently characterizes the resulting relationship between egoists and their objects (including other individuals) in proprietorial terms, although, in contrast to traditional juridical conceptions of ownership, ‘egoistic property’ largely collapses into a notion of instrumental treatment. The ‘union of egoists’ is established as the only form of association which does not violate ‘ownness’ and so can constitute a suitable vehicle for advancing individual interests. In the historical maturity of egoism, community survives only in the form of this constantly shifting alliance which enables egoists to unite without loss of sovereignty, a purely instrumental combination whose good is solely the advantages which individuals derive from the pursuit of their interests - there are no final ends and association is not valued in itself.

See also: Anarchism; Egoism and altruism

DAVID LEOPOLD

List of works

Stirner, M. (1842-7) *Kleinere Schriften (Short Works)*, ed. J.H. Mackay, Berlin: Bernhard Zack, 1914.(An interesting miscellany of Stirner’s articles.)

Stirner, M. (1844) *Der Einzige und sein Eigentum*, trans. S.T. Byington, *The Ego and Its Own*, ed. D. Leopold, Cambridge: Cambridge University Press,1994.(This excellent edition contains a translation of Stirner’s most important work, extensive annotations and a most substantial annotation.)

Stirner, M. (1852) *Geschichte der Reaktion (A History of Reaction)*, Berlin: Allgemeine Deutsche Verlags-Anstalt, 2 vols.(An incomplete work largely compiled from the writings of others, including Burke, Comte, Hengstenberg and Florencourt.)

References and further reading

Clark, J.P. (1976) *Max Stirner’s Egoism*, London: Freedom Press.(A short but useful critical discussion of *The Ego and Its Own*.)

Helms, H.G. (1966) *Die Ideologie der anonymen Gesellschaft (The Ideology of the Anonymous Society)*, Cologne: DuMont Schauberg.(A rather idiosyncratic interpretation of Stirner but contains an outstanding bibliography.)

Mackay, J.H. (1977) *Max Stirner: sein Leben und sein Werk (Max Stirner: His Life and Work)*, Berlin: Mackay Gesellschaft.(First published in 1898 this remains the standard biography of Stirner.)

Paterson, R.W.K. (1971) *The Nihilistic Egoist: Max Stirner*, Oxford: Oxford University Press.(An introductory study, especially helpful on the relation between Stirner and later thinkers.)

Stoicism

Stoicism is the Greek philosophical system founded by Zeno of Citium c.*300* BC *and developed by him and his successors into the most influential philosophy of the Hellenistic age. It views the world as permeated by rationality and divinely planned as the best possible organization of matter. Moral goodness and happiness are achieved, if at all, by replicating that perfect rationality in oneself, and by finding out and enacting one's own assigned role in the cosmic scheme of things.*

The leading figures in classical, or early, Stoicism are the school's first three heads: Zeno of Citium, Cleanthes and Chrysippus. It is above all the brilliant and indefatigable Chrysippus who can be credited with building Stoicism up into a truly comprehensive system. 'Early Stoicism' - the main topic of this entry - is in effect largely identical with his philosophy.

No formal philosophical writings of the early Stoics survives intact. We are mainly dependent on isolated quotations and secondary reports, many of them hostile. Nevertheless, the system has been reconstructed in great detail, and, despite gaps and uncertainties, it does live up to its own self-description as a unified whole. It is divided into three main parts: physics, logic and ethics.

The world is an ideally good organism, whose own rational soul governs it for the best. Any impression of imperfection arises from misleadingly viewing its parts (including ourselves) in isolation, as if one were to consider the interests of the foot in isolation from the needs of the whole body. The entire sequence of cosmic events is pre-ordained in every detail. Being the best possible sequence, it is repeated identically from one world phase to the next, with each phase ending in a conflagration followed by cosmic renewal. The causal nexus of 'fate' does not, however, pre-empt our individual responsibility for our actions. These remain 'in our power', because we, rather than external circumstances, are their principal causes, and in some appropriate sense it is 'possible' for us to do otherwise, even though it is predetermined that we will not.

At the lowest level of physical analysis, the world and its contents consist of two coextensive principles: passive 'matter' and active 'god'. At the lowest observable level, however, these are already constituted into the four elements earth, water, air and fire. Air and fire form an active and pervasive life force called pneuma *or 'breath', which constitutes the qualities of all bodies and, in an especially rarefied form, serves as the souls of living things.*

'Being' is a property of bodies alone, but most things are analysed as bodies - even moral qualities, sounds, seasons and so forth - since only bodies can causally interact. For example, justice is the soul in a certain condition, the soul itself being pneuma *and hence a body. A scheme of four ontological categories is used to aid this kind of analysis. In addition, four incorporeals are acknowledged: place, void, time and the* lekton *(roughly, the expressed content of a sentence or predicate). Universals are sidelined as fictional thought constructs, albeit rather useful ones.*

The world is a physical continuum, infinitely divisible and unpunctuated by any void, although surrounded by an infinite void. Its perfect rationality, and hence the existence of an immanent god, are defended by various versions of the Argument from Design, with apparent imperfections explained away, for example, as blessings in disguise or unavoidable concomitants of the best possible structure.

'Logic' includes not only dialectic, which is the science of argument and hence logic in its modern sense, but also theory of knowledge, as well as primarily linguistic disciplines like rhetoric and grammar. Stoic inferential logic takes as its basic units not individual terms, as in Aristotelian logic, but whole propositions. Simple propositions are classified into types, and organized into complex propositions (for example, conditionals) and complete arguments. All arguments conform to, or are reducible to, five basic 'indemonstrable' argument formats. The study of logical puzzles is another central area of Stoic research.

*The Stoics doggedly defended, against attacks from the sceptical Academy, the conviction that cognitive certainty is achieved through ordinary sensory encounters, provided an entirely clear impression (*phantasia*) is attained. This, the 'cognitive impression' (*phantasia katalēptikē*), is one of such a nature that the information it conveys could not be false. These self-certifying impressions, along with the natural 'preconceptions' (*prolēpseis*) which constitute human reason, are criteria of truth, on which fully scientific knowledge (*epistēmē*) - possessed only by*

the wise - can eventually be built.

Stoic ethics starts from oikeiōsis, *our natural 'appropriation' first of ourselves and later of those around us, which makes other-concern integral to human nature. Certain conventionally prized items, like honour and health, are commended by nature and should be sought, but not for their own sake. They are instrumentally preferable, because learning to choose rationally between them is a step towards the eventual goal of 'living in agreement with nature'. It is the coherence of one's choices, not the attainment of their objects, that matters. The patterns of action which promote such a life were systematically codified as* kathēkonta, *'proper functions'.*

Virtue and vice are intellectual states. Vice is founded on 'passions': these are at root false value judgments, in which we lose rational control by overvaluing things which are in fact indifferent. Virtue, a set of sciences governing moral choice, is the one thing of intrinsic worth and therefore genuinely 'good'. The wise are not only the sole possessors of virtue and happiness, but also, paradoxically, of the things people conventionally value - beauty, freedom, power, and so on. However geographically scattered, the wise form a true community or 'city', governed by natural law.

The school's later phases are the 'middle Stoicism' of Panaetius and Posidonius (second to first century BC) and the 'Roman' period (first to second century AD) represented for us by the predominantly ethical writings of Seneca, Epictetus and Marcus Aurelius.

1 School history, sources

The name 'Stoic' originates from the Stoa Poikile or 'Painted Colonnade' at Athens, where at least the school's first generation used to meet. At first called 'Zenonians', they came in time to be known as the 'Stoa people', or 'Stoics'. The founder Zeno of Citium and his successor as school head, Cleanthes, forged the system in large measure. But it was the third head, Chrysippus, at the end of the third century BC, who developed it into a truly global philosophical creed.

Stoicism was in many ways a technically updated version of Socrates' century-old philosophy. Its debt to earlier thinkers like Heraclitus and Plato is also manifest. How far it took account of Aristotle's work, on the other hand, is disputed. Contemporary negative influences included Epicureanism, to which it was diametrically opposed on most issues, and the sceptical critiques of Stoic positions launched from the New Academy.

Chrysippus is the main voice of Early Stoicism. His successors Diogenes of Babylon and Antipater made their own contributions and modifications, but are seen as continuing the same tradition. In the late second and early first century BC, there was some softening of positions and a renewed interest in the writings of Plato. This has led many scholars to classify Panaetius and Posidonius as leading a separate phase, known as 'Middle Stoicism'. Finally, 'Roman' Stoicism designates that of the early Roman Empire, whose main spokesmen for us are the Latin essayist Seneca and the Greek writers Epictetus and Marcus Aurelius. The great bulk of their interests lies in moral philosophy.

The chart is a guide to entries on individual Stoics (listed in capitals).

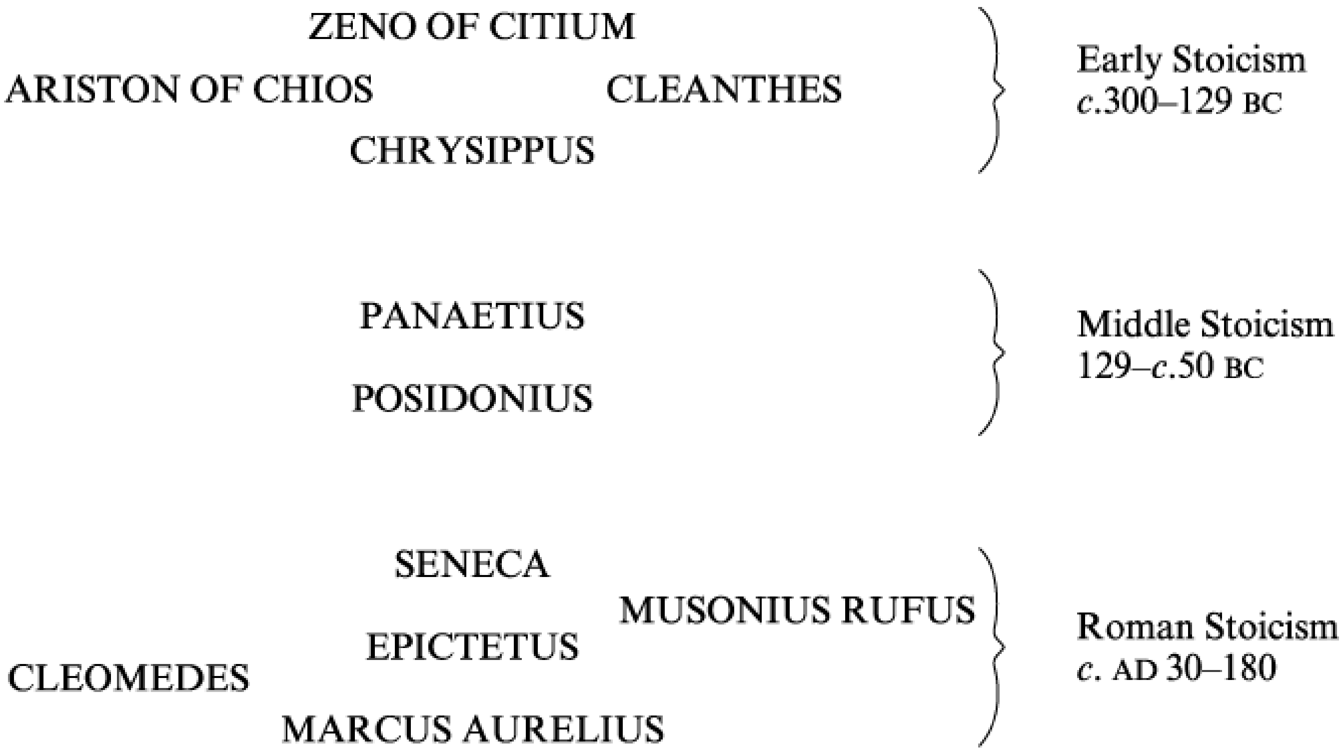

We possess little evidence about the school's institutional base, in Athens or elsewhere, and it is not even clear whether it ever occupied its own premises. Its cohesion was, at any rate, primarily doctrinal. There was extensive agreement between individual Stoics, alongside a good deal of in-school quarrelling over specific issues. The most publicized dispute occurred in the first generation, between Zeno and the independent-minded Ariston of Chios. After his death, Zeno was revered by most Stoics, who would not openly criticize him. Rather, their philosophical disagreements often took the outward form of disputes as to what Zeno had really meant.

No early Stoic text survives, apart from Cleanthes' short *Hymn to Zeus*. But modern scholarship has managed to reconstruct most of the system in considerable detail from secondary sources, which incorporate numerous verbatim quotations. Book VII of Diogenes Laertius' *Lives of the Philosophers* is a major source, as is the doxographer Arius Didymus. Cicero's philosophical treatises contain first-rate presentations of various parts of the system. And invaluable evidence is available even from entrenched critics of the Stoics, such as the Platonist Plutarch, the Pyrrhonist Sceptic Sextus Empiricus and the doctor Galen.

2 The parts of philosophy

The Stoics formally sanctioned what was to become thereafter the canonical division of philosophy into three parts: physics, logic and ethics. Stoicism is arguably the first fully and self-consciously *systematic* philosophy. In the next three paragraphs correspondences will be noted between these three main areas of philosophy and the individual topics covered in the following sections. However, some of these topics (especially those in §§6-9, 14, 20) may not fit quite as neatly into any one of the three areas as the schematization implies.

Logic (§§10-13), the science of rationality (derived from *logos*, 'reason'), included not only inferential logic in the modern sense (more correctly called 'dialectic' by the Stoics), but also theory of knowledge, in which generation after generation of Stoics defended the existence of cognitive certainty against sceptical attacks. All Stoics, with the exception of Ariston of Chios, regarded logic as a fundamental part of philosophy.

Physics (§§3-9) - the study of nature (*physis*) - was a largely speculative or theoretical discipline, concerned with understanding the world's rational structure, and therefore embracing such diverse disciplines as biology, psychology and theology within its scope. It drew on the findings of contemporary scientists, but was not itself any kind of empirical science. Indeed, opponents accused it of wilful blindness to inconvenient new scientific discoveries, such as the demonstration by Alexandrian doctors that the rational mind is in the head, not the chest. Physics presupposed logic, at least to the extent that its findings were largely developed in strings of syllogisms.

Ethics (§§14-19), the authentically Socratic core of Stoic philosophy, was the discipline which described how happiness could be achieved. It presupposed physics, which supplied an understanding of the world's rational structure and goodness and of the individual's place in it.

There was less agreement about how the three parts related to each other. One favoured model compared philosophy to an orchard in which logic was the protective outer wall, physics the soil and trees, and ethics the fruit. Posidonius favoured the analogy to a living animal, in which logic was the bones and sinews, physics the flesh and ethics the soul. These and other analogies probably agreed in making ethics the ultimate aim and crowning achievement of philosophy. The value of physics and logic was in a way instrumental - to acquire the understanding which would make a happy life possible. But that understanding, a perfected rationality, was itself so integral to the Stoic conception of happiness that to call it instrumental may be to underestimate the true unity of Stoic philosophy.

Most leading Stoics put the three parts of philosophy into the sequence logic-physics-ethics, but some favoured other orders. Here they are likely to have been specifying nothing more than the best order in which to teach it, with no implications which might threaten its conceptual unity.

3 The foundations of physics

Physics is the study of nature (*physis*), and to understand the true nature of the world it is necessary to start at the very lowest level of physical analysis. The world and its contents consist of passive 'matter' (*hylē*) and active 'god' (*theos*). These two 'principles' (*archai*) totally interpenetrate each other, and their interaction underlies all change. The active principle is quite literally god, a divine causal force which imbues the entire world with rationality.

Why is the second principle, matter, added? Not in order to make the world corporeal, since god (as well as matter) is already corporeal. It is a fundamental Stoic principle that only bodies are capable of causal interaction, and god could not shape the world if he lacked causal powers. Rather, matter is added because god is an entirely active causal power, and there must therefore be something passive on which he acts. Matter is thus a purely theoretical construct, with no properties of its own to contribute beyond its mere passivity.

Matter and god, then, are conceptual rather than empirical items. At the lowest *observable* level, at which cosmic processes of change can actually be studied, matter and god are already by their interaction constituted into the traditional four 'elements' - earth, water, air and fire. Of these, air and fire form an active and pervasive life force called *pneuma* or 'breath', which by its presence in things constitutes their qualities (see Pneuma §2). Earth and water, on the other hand, are essentially passive elements, which serve a primarily material role: it is only the pervasive presence of *pneuma* in them that shapes them into complex items, including living things. Thus at the lowest phenomenal level earth and water take over the role which pure matter had held at the lowest theoretical level, while air and fire, paired as *pneuma*, take over a role analogous to that of the second principle, god.

Pneuma, as the combination of warmth and breathed air that is fundamental to animal life, had already earned a central place in medical theory, to which Stoic physics was heavily indebted. The distinctive contribution of Stoicism, at any rate by the time of Chrysippus, was to extend the explanatory role of *pneuma* beyond individual animal life, and to make it the vital power of the world as a whole. Since the Stoic world (following Plato among others) is itself a living creature, this extension was not as surprising as it may at first seem. *Pneuma* is the vehicle of the divine 'reason' (*logos*) which pervades and governs the entire world (see Logos §1).

Pneuma is all-pervasive, but varies in its properties according to its degree of tension. In its most highly attuned form, a portion of it may serve as the soul (*psychē*) of an animal (see Psychē). Many Stoics held that the human soul, at least, survives the death of the body (although it must of course eventually perish in the conflagration, see §5). A lesser grade of it is called 'nature' (*physis*), in a special sense in which the vegetative life force of a plant may be called its 'nature'. Finally, a still less refined form of *pneuma* is present in any discrete object, even a stone or a cup, being what endows it with its cohesion as a single object: as such, it is called the thing's 'tenor' or 'state' (*hexis*).

Both the primary 'principles' and *pneuma* play a key role in securing one of the most characteristic of Stoic tenets, the identification of 'being' with corporeality. At *Sophist* 246-7 Plato had launched a devastating attack on crude

materialists who restrict being to bodies. Since they can neither deny that there *is* (for example) justice, nor equate justice with some body, Plato suggests, they must abandon their position in favour of a new existential criterion, namely that to be is to have the power to interact. Stoicism sets out to defend the materialist position from Plato's attack: the two criteria considered by Plato are in reality one and the same, since the only things that have the power to interact *are* bodies. Hence to be is, after all, to be a body.

Virtually everything in the Stoic world is a body (for the few exceptions, see §7-8). This should not be misunderstood as reductive materialism. At the lowest level, 'god' is a body, but not a *mere* body: life and intelligence are his irreducible properties. God has corporeality in addition to these vital properties simply because only thus can he have any causal role in moving and shaping matter. Once this link of corporeality to causality is accepted, it is extended to the vast majority of items traditionally thought incorporeal. Justice, for example, must have causal powers if it is to do any good. It must sometimes make the limbs and voice move in ways in which they would otherwise not have moved. Therefore justice is a body. This apparently paradoxical outcome is defused once we learn *what* body it is. Justice is simply the soul in a certain condition; and the soul is a body, being an individual portion of *pneuma* pervading the organism and therefore able to govern its movements. Similarly, (spoken) words are bodies because they are vibrating portions of air, and that is how they can affect the thoughts of a listener.

Stoic corporealism was much derided by ancient critics, but it has the considerable merit of explanatory economy.

4 The continuum

Matter is permeated by god, and the passive elements by the active ones which constitute *pneuma* (see §3). What is this permeation? Mere juxtaposition of particles, such as atomism posits, could never constitute the intimate causal link between god and the matter which makes the Stoic world an inherently and ideally intelligent being. The active body must permeate the passive body 'through and through'. Stoic theory distinguishes three grades of mixture. 'Juxtaposition', for example of mixed grains, conforms to the atomist model. 'Fusion' (*synchysis*) is a kind of interpenetration in which the ingredients irreversibly lose their distinctive properties and a single new stuff is generated. In between these lies 'blending' (*krasis*), which also involves total interpenetration, but with the ingredients retaining their own distinctive properties. A helpful Stoic example, the fire which is seen to permeate a red-hot piece of iron, may clarify how the two ingredients can be seen as literally coextensive, rather than alternating as in 'juxtapostion'. It is 'blending' that describes the relation of *pneuma* to the material substrate.

'Division is to infinity, or ''infinite'' according to Chrysippus (for there is not some infinity which the division reaches, it is just unceasing). And blendings, also, are through and through' (Diogenes Laertius, VII 150-1). The doctrine of total interpenetration depends on the infinite divisibility of body, because if each of the blended stuffs consisted of indivisibly small particles these could only be juxtaposed, and not further blended. Hence the Stoics are, like Aristotle, committed defenders of the infinite division, on both mathematical and physical grounds.

For example, critics of the continuum had argued that if a finite body is infinitely divisible it will consist of infinitely many equal parts, and hence, impossibly, be infinite in size. Chrysippus replied that a finite object does not consist of any particular number of ultimate parts, finite or infinite. On an old puzzle of Democritus', whether the two circular planes yielded by horizontally slicing a cone are equal (in which case why isn't it a cylinder?) or unequal, Chrysippus replied that they are 'neither equal nor unequal'. Unfortunately our sources are so depleted as to permit little more than speculation about his defence of this claim. The same applies to the traces of a Stoic solution to the celebrated paradox according to which motion through an infinitely divisible continuum is impossible because it would consist of an infinite, and therefore uncompletable, series of sub-motions (see Zeno of Elea §5). They replied that the moving object may complete a distance in a single undivided (though divisible) motion - possibly on the ground that divisions are thought constructs, of which only a finite number can be actually imposed on a distance.

5 Cosmology and theology

The Stoic world is a living creature with a fixed life cycle, ending in a total 'conflagration' (*ekpyrōsis*). Being the best possible world, it will then be succeeded by another identical world, since any variation on the formula would have to be for the worse. Thus the Stoics arrive at the astonishing conception of an endless series of identical

worlds - the doctrine of cyclical recurrence, according to which history repeats itself in every minute detail. (For the leading Stoic dissenter on this, see Panaetius §1.)

Although a conflagration terminates each world phase, this is not its collapse into ruin but its achievement of perfection. That is because 'creative fire' (*pyr technikon*) is the purest, the most divine and the most creative form of divine *pneuma*, to be associated more with the light and warmth that promote growth than with flame. In our present world phase, the strongest concentration of creative fire is the sun, which some Stoics identified as the world's divine 'commanding-faculty' (*hēgemonikon*). During the conflagration what exists is in effect pure intelligence, which plans the next world phase in every detail. There follows a process by which the fiery stuff differentiates and stratifies itself into the four elements. It already at this stage contains the 'seminal principles' (*spermatikoi logoi*), which may be viewed as blueprints for the individual organisms and other entities which will eventually emerge. Both the dominant role assigned to fire and its association with divine 'reason' (*logos*) owed much to the influence of Heraclitus (see Heraclitus §4; Logos §1).

By 'god' the Stoics meant, primarily, the immanent principle governing the world, variously also identified with 'creative fire', with 'nature' or with 'fate' (on which see §20). Second, the world itself was also called 'god'. But - characteristically of Greek religious thought - this apparent monotheism did not exclude polytheism. Individual cosmic masses were identified with individual gods: for example the sea and the air were linked with Poseidon and Hera respectively, and the remaining traditional gods were likewise assigned specific cosmic functions. By means of allegorical rationalization, Stoic theology incorporated and interpreted traditional religion, rather than replacing it. Etymology (sometimes highly fanciful) was one tool used in this process. For example, two common forms of the name Zeus were 'Zēn' and 'Dia', which could also mean respectively 'Life' and 'Because of': this made it easy to interpret the traditional head of the pagan pantheon as symbolizing the Stoic primary deity, who was the world's life-force and causal principle.

The world, then, is itself divine, and is from first to last providentially planned and governed by an immanent intelligence. This thoroughgoing teleology owed much to Plato's *Timaeus*, but also to his *Phaedo*, where (96-9) Socrates had been portrayed as advocating a teleological physics, while admitting his own incapacity to develop one. We can here glimpse one of the many ways in which Stoicism sees itself as working out in full technical detail what was already implicit in the thought and life of Socrates.

Since the world *is* god, in his most manifest form, there is no distinction in Stoicism between proving the existence of god and proving the perfect rationality of the world. These proofs, most of which are variants on the Argument from Design, generated massive controversy between the Stoics and their critics (see especially Cicero, *On the Nature of the Gods* II-III). They include the following lines of argument. First, the world (especially the heaven) is like a giant mechanism, vastly superior to the most elaborate human invention. If must therefore, *a fortiori*, have an intelligent designer. Second, the creation of such a world by mere accident (for which see Epicureanism §8) is as unlikely as producing a literary masterpiece by shaking out letters randomly onto the ground. Third, the world is full of beneficial structures which cannot have been invented by its inhabitants: for example, the cycle of seasons, the temporal uniformity of the heavenly rotations, symbiotic relationships between species, and the food chain - including that miraculous foodstore the pig, created as a living being purely in order to keep the meat fresh. Fourth, the world is supremely beautiful, and therefore the work of a consummate artist. And fifth, any imperfections are either merely apparent (for example, wild beasts, which encourage the virtue of courage in us), or inevitable concomitants of the best possible structure (for instance, an example borrowed from Plato, the fragility of the human head). Sometimes localized sufferings are justified by the greater good they serve, even if it is not always evident what that good is. Chrysippus remarked, 'If I knew that I was fated to be ill now, I would set out to be ill. So too the foot, if it could think, would set out to get muddy'.

The syllogisms used to further this theology contained some highly controversial premises, for example: if there is anything which human beings cannot create, whoever did create it is superhuman. Or again: if a thing has rational parts (as the world does - namely us), then the whole must be rational too.

6 The 'categories'

Aristotle developed a complex ontological scheme, his ten 'categories' (see Aristotle §7; Categories). The Stoics, following not Aristotle but the early Academy, recognized just two categories: that is, they divided all existing

things into absolute or *per se* items, and relative items. For example, a house or dog is something absolute, whereas to be a slave, or sweet, is relative: being a dog does not, as such, involve standing in any relation, but to be a slave is to be somebody's slave, and to be sweet is to taste sweet to one or more percipients.

However, the Stoics also developed a more original, fourfold ontological classification, commonly known to modern interpreters as the Stoic 'categories', although it is unlikely that they used this term for it, and it is safer to call them 'genera'. It is disputed what the origins and purpose of this scheme were, but its most interesting recorded uses are to account for individual identity over time, and to amplify the thesis described in §3 that 'being' belongs only to bodies.

According to this classification, any given item may be seen as: (1) a mere 'substrate' (*hypokeimenon*); (2) something 'qualified' (*poion*); (3) something 'disposed in a certain way' (*pōs echon)*; and (4) something 'disposed in a certain relation' (*pros ti pōs echon*).

To start with the first pair. To pick out a thing as (1), a 'substrate', is merely to individuate it as a (temporary) lump of matter, for example, by pointing at it. To mark it as (2), something 'qualified', is to distinguish it by one or more qualities which it possesses. Here the Stoics have a precise physical analysis of what a quality is: a portion of *pneuma* imbuing the thing, whether as its soul or as some other inherent property, for example, its colour or its weight. There is a further subdivision of the 'qualified' into (a), what is 'commonly qualified', for example, dog, human, wise, green, and (b), what is 'peculiarly qualified', for example, Fido, Socrates. These correspond, roughly, to what is signified by (a) common nouns and adjectives, and (b) proper names.

It is the first pair that is invoked to explain diachronic identity. Such identity had been challenged by the growing argument (*auxanomenos logos*), a puzzle much favoured by the Stoics' sceptical opponents in the New Academy. This argument objected that any 'growth', however slight, must generate a new individual, since a body with even one new particle added is not the same body as before; therefore there can be no enduring subject of growth; and hence no growth. The Stoic answer is that what endures as the subject of growth is not the material substrate (or 'substance', *ousia*), but the 'peculiarly qualified individual'. *Qua* 'substrate' Dion (to use the Stoics' favourite stock name) has little if any endurance, but *qua* 'peculiarly qualified individual' (*idiōs poios*), that is, *qua* Dion, he endures throughout his life. Although Dion's matter constantly changes, the individuating quality that makes him Dion is with him, unchanged, from birth to death. That there is such a lifelong peculiar quality is a thesis of Stoic physics (there is no evidence what they thought it consisted in; the uniqueness of fingerprints is a modern discovery), but the need to solve the growing argument elevates it to the status of a metaphysical necessity. It also plays a part in Stoic epistemology (see §12).

As well as being (1) a lump of matter, and (2) both (a) human, wise and so on and (b) Dion, he is also at all times (3) in some disposition or other, for example, sitting, and (4) in some relation, for example, the person on my right. Here (3) and (4) represent the final two genera. The third genus, 'disposed in a certain way', is widely used to analyse as corporeal items commonly believed to be incorporeal (see §3). For example, knowledge is analysed as 'the commanding-faculty of the soul disposed in a certain way' - where the soul itself is corporeal, being *pneuma*. The fourth genus has similar analytic uses. What is special about it as a mode of being is that it picks out, not all relative items, but those whose being what they are is wholly parasitic on that relation, so that they could cease to exist merely because their external correlate changes or perishes. Suppose Dion is Theon's accomplice. *Qua* Dion he may continue to exist despite Theon's desertion or death; but *qua* accomplice he cannot. Hence 'accomplice' is a fourth genus item, while 'Dion' is not. For an important debate about the status of virtue, using this fourth genus, see §16.

7 Space and time

After their heroic defence of corporealism (see §3), it may come as a disappointment that the Stoics do nevertheless admit four kinds of incorporeal: place, void, time and the *lekton* or 'sayable'. These do not have actual being, because none of them possesses any causal powers to act or be acted upon. Nevertheless, an account of the world is incomplete without them, and they are therefore proper subjects of discourse. This entitles them to be described, with a lesser label, as having 'subsistence'.

A thing's 'place' is conceived by the Stoics as a portion of three-dimensional (geometrical) space, coextensive with its own occupant. This is given in the Stoic definition of place as 'that which (i) is able to be occupied by

what-is and (ii) is occupied throughout, (iii) whether by some thing or by some things'. Here (i) gives place its genus - roughly, space; (ii) adds its differentia, namely that its capacity to be occupied must be being exercised; (iii) specifies that the 'being', that is, body, which occupies it may be comprised of one or more discrete individuals. Put non-technically, a place is a fully occupied portion of space, whether occupied by a single entity or by a collection of entities.

When on the other hand some space's capacity to be occupied is not being exercised, it is called 'void' (or 'vacuum'). Void is defined as 'that which is able to be occupied by what-is, but which is not occupied'. As a matter of fact, this failure to be occupied is held to be a possibility only for the infinite space surrounding the world, since the world itself is an absolute plenum, containing no vacuum at all. The extra-cosmic void comes to be occupied - at least, some of it does - only during the 'conflagration' (see §5), when the world in its pure fiery state is said to expand into it.

The whole universe, called the 'all' in the sense of 'sum total', consists of the world body plus the infinite surrounding void. But of this combination, it is only the world body which is called the 'whole'. A whole, to qualify as such, must have a single unifying *pneuma* (see §3), something with which the void, being altogether empty, could not possibly be imbued. Thus the terminological distinction captures the Stoic thesis that the void is not any interacting part of the cosmic organism. It provides the conditions for one special kind of cosmic change, but it is not itself a participant in any change.

Since void is undeniably an incorporeal, and since void and place are both generically the same, it is hardly imaginable that the Stoics could have found a way of re-analysing place as a body. In any case, an adequate account of place must make sense of motion, that is, of a place being entered or vacated by a body. If the place were itself identified with the occupying body, no such account would be feasible.

The third incorporeal on the Stoic list is 'time'. Yet, curiously, in our sources individual parts of time are often analysed, in typical Stoic fashion, as corporeal. For instance, days and nights are simply the world's atmosphere in such and such a condition, and hence both they and the longer periods of time composed of them are said to be bodies. And reasonably so, one may think, since these temporal items could easily be deemed to have active or passive causal powers - for example, to be caused by the movement of the sun, and in turn to cause the progression of life cycles, and so on.

It seems to be only time as a whole which is conceded to be an incorporeal. Why so? Time (in this sense) is defined as 'the dimension of the world's motion'. Probably then the idea is as follows. The regularity of a body's motion - including the rotation of the heaven, which paradigmatically displays the progress of time - presupposes fixed spatial and temporal intervals through which it takes place. If either of those intervals were (in the usual Stoic corporealizing fashion) identified with the moving body itself, its motion would be left altogether without objective coordinates. This consideration may be what motivates the thesis that not only the spatial interval - 'place' - but also the temporal interval - 'time' - subsists independently of the moving bodies which pass through it.

8 The *lekton*

The fourth incorporeal on the list, the *lekton* (plural *lekta*), 'sayable', is a key term in Stoicism. Plato in the *Sophist* (261-2) distinguishes two linguistic tasks: naming, which is to pick out a subject, and 'saying' (*legein*), which is, roughly, to attach a predicate to that subject. This is probably the background to the Stoic *lekton*, which seems to have originally meant, as distinct from a subject, the sort of thing that you can say about a subject. As in Plato, it is typically expressed by a verb like 'walks' or 'to walk'. The actual action, walking, is itself a body: it is analysed as the commanding faculty of the soul (itself *pneuma*, and therefore a body) functioning in a certain way. But in certain contexts the predicate expressed by the verb 'to walk' is not properly identified with some body. The Stoics noted at least two cases - wishing, and causation. (1) When we say 'I *want* to walk', the soul which wishes is a body, but what it wishes for is not either itself, or for some further body to be added to it: it is for some predicate to become true of it. (2) When, for example, fire *causes* wood to be burnt, the effect generated is not some new body, burnt wood, since the wood already exists: it is that some new predicate becomes true of the wood. Thus, both the objects of wishes and the effects of causes are incorporeal predicates, technically called *lekta*.

The most prominent role of the *lekton* is in logic. Stoic logic (see §§10-11) is less interested in predicates as such -

what they call ‘incomplete *lekta*’ - than in the complete propositions containing them, the primary bearers of truth. Normally speaking the predicate expressed by a verb, for example, ‘…walks’, serves a full linguistic function only when it has a subject term supplied by a noun, for example, when someone says ‘Dion walks’. The notion of a ‘complete *lekton*’ is used to distinguish such cases. Most commonly, a complete *lekton* is the proposition expressed by a complete declarative sentence, as in the example given, although other complete *lekta* include questions (‘Is Dion walking?’), commands (‘Walk, Dion’) and so on.

A complete *lekton* is said to be produced by attaching a predicate to a ‘nominative case (*ptōsis*)’. This generates an interpretative puzzle. A *lekton* is well attested to be an incorporeal, yet a nominative case, being a grammatical form, ought to be a word and hence a body (a spoken word being vibrating air; a written word probably ink). Can a complete *lekton* be incorporeal and yet have one part which is corporeal? It may be that the subject term, by being expressed, makes the *lekton* complete without actually becoming part of it. Or it may be that the Stoics posited, in addition to the predicate expressed by the verb, a further incorporeal incomplete *lekton*, namely the subject of the complete *lekton*, expressed by the subjective nominative noun of the corresponding uttered sentence. Both suggestions (which do not exhaust those available) are problematic, and neither gains unequivocal support from the surviving evidence.

Nor is it easy to find a link between the incorporeality of the *lekton* and that of place, void and time (see §7). These latter three have some sort of mind-independent reality. Can the same be said for *lekta*? This is controversial, but the causal role of *lekta* must lend them some degree of objectivity, since causal processes presumably go on in the same way whether or not anyone is there to observe or analyse them. The *lekton* is defined as ‘that which subsists in correspondence with a rational impression’ (that is, roughly, with a human thought (see §12). This could be taken to make it parasitic on the thought processes of rational beings. But it may alternatively mean no more than that a *lekton* is a formal structure onto which rational thoughts, like the sentences into which they can be translated, must be mapped. This latter possibility bears some comparison to space and time, which, although defined by reference to their potential or actual occupants, are the objective dimensions onto which the positions and motions of those occupants are to be mapped, and are not altogether parasitic on them for their reality. The analogy must not be pushed too far, since the *lekton* differs from the other three incorporeals in not being any kind of mathematically analysable extension. But rationality is as much an intrinsic feature of the Stoic world as dimensionality, and it would be entirely Stoic to hold that there are objective parameters onto which its rational structures can be mapped

9 Particular and universal

There seems to be no room for universals in a Stoic world, where to exist is to be a body and hence, it seems, a particular. The highest ontological class recognized by Stoicism is that of ‘something’ (*ti*, plural *tina*), a class so broad as to include both bodies and incorporeals. Yet even this class excludes certain ‘not-somethings’ (*outina*), which Stoics identify with universals. That a universal is a ‘not-something’ they demonstrate by the following syllogism, diagnosed as being invalid precisely because in the minor premise it illegitimately substitutes a universal, ‘man’, for ‘someone’ (*tis*, the masculine form corresponding to the neuter *ti*) in the major premise: ‘If someone is in Athens, they are not in Megara; but man is in Athens; therefore man is not in Megara’. So universals are apparently metaphysical outlaws, intractable even to basic logical laws.

On the other hand, universals have a fundamental place in the job of the dialectician. In the tradition bequeathed by Plato, the Stoics regard the science of dialectic (see §10) as largely concerned with the activities of division and definition: dividing genera into species and defining individual terms. And the things divided and defined are, of course, universals - as Plato had already indicated by equating them with Forms.

Hence a dilemma: universals are essential to dialectic, and yet are logically incoherent items with no place in a Stoic ontology. Their resolution of the dilemma comes in two parts, one logical, the other epistemological. Logically, the dialectician’s use of universals is justified by a re-analysis. A dialectician’s definition, for example, ‘Man is a rational mortal animal’, actually means ‘If something (*ti*) is a man, it is a rational mortal animal’. A similar analysis was offered for statements of division. Thus dialectical pronouncements which appear to have universals as subjects are legitimate only because they are disguised conditional statements about particulars (‘somethings’). (A partial analogy can be found in modern attempts to re-analyse as logically coherent ‘The average man has 2.4 children’.)

The epistemological side of the resolution is required because Plato had maintained that universals are, as such, proper objects of thought. This time the conditional re-analysis offers little help in formulating a Stoic response. Rather, their reply is that universals are indeed objects of thought, or ‘concepts’ (*ennoēmata*), but that far from being objective parts of the world’s furniture they are nothing more than thought constructs. The thoughts themselves are real enough: being ‘conceptions’ (*ennoiai*), they are simply our mental *pneuma* in a certain condition, and hence bodies; but the ‘concepts’ which constitute the intentional objects of those conceptions are not themselves bodies, or indeed anything at all. This position makes it appropriate to call the Stoic doctrine of universals an early variety of ‘conceptualism’ (see Nominalism §1).

10 Dialectic

Dialectic is the main branch of what the Stoics called ‘logic’, and amounts, roughly, to the science of argument. Although dialectic was both theorized and applied in innumerable treatises, the Stoics never lost sight of the Platonic conception of it as a fundamentally two-person activity, involving the interrogation of an interlocutor. ‘Dialectic’ actually means ‘the science of dialogue’. The science of rhetoric differed from dialectic precisely in being the science of producing a good monologue, and was treated as a separate branch of logic. There was no uniform Stoic view as to whether theory of knowledge (see §§12-13) counted as part of dialectic, or as a third branch of logic.

In so far as it is concerned with argument, dialectic has distinct parts dealing with (1) signifiers and (2) significates. The former are words (and hence bodies; see §8), and Stoic dialectic’s concern with language as such led it, among other things, to develop the first real grammatical theory in Western thought. It became the basis of all subsequent work on grammar in antiquity and far beyond.

The latter term, ‘significates’, designates *lekta* (see §8), the incorporeal meanings which in their complete form are expressed only by whole sentences. Stoic logic (to use the word now in its modern sense) concentrates on one species of *lekta*, the declarative ones, called *axiōmata*. This term, literally ‘judgments’, is more familiarly translated as ‘propositions’. Stoic logic is indeed the first fully developed propositional or sentential logic. In this it differs radically from Aristotelian logic, which is a logic of terms (see Logic, ancient). The origins of Stoic logic probably lie less in Aristotle than in the work of the Dialectical school (see Dialectical school; Diodorus Cronus; Philo the Dialectician).

11 Inferential logic

Propositions are the bearers of truth and falsity, and simple propositions are the atomic units of Stoic logic. One symptom of the latter is that in syllogistic the negation sign ‘Not…’ is properly prefixed to the entire proposition, rather than (as in Aristotle’s logic) to its predicate. On the other hand, propositions are classified partly according to the subject terms which they contain: ‘This individual is walking’ (ideally accompanied by pointing) is a ‘definite’ proposition; ‘Someone is walking’ is ‘indefinite’; and propositions expressed with a noun in the subject position, for example, ‘A human being is walking’, ‘Dion is walking’, are ‘intermediate’. This triple distinction, like most aspects of Stoic logic, is adopted ultimately for the sake of validating the arguments in which these simple propositions feature. Indefinite propositions are verified by the corresponding intermediate or definite ones, but not vice versa. Hence in a syllogism whose major premise begins ‘If *x* is walking…’ and whose minor premise has the form ‘But *y* is walking…’, validity will normally be preserved if ‘*x* is walking’ is indefinite, but neither if it is definite while ‘*y* is walking’ is indefinite or intermediate, nor if it is intermediate while ‘*y* is walking’ is indefinite.

Complex propositions are those compounded of two or more simple propositions with the help of connectives like ‘and’ or ‘if’. Here again, they are selected for their role in syllogistic theory. Thus they comprise: conjunctive propositions, for example, ‘It is day and it is light’, which are important mainly in negated conjunctions, for example, ‘Not: it is day and it is dark’; disjunctive propositions, for example, ‘Either it is day or it is night’; and conditionals, for example, ‘If it is day, it is light’.

Simple and complex propositions are the basic components of syllogisms. These are held either to have, or to be reducible to by four rules of analysis known as the *themata* (see Logic, ancient §6), one of the following five forms: (1) ‘If the first, the second; but the first; therefore the second’; (2) ‘If the first, the second; but not the

second; therefore not the first'; (3) 'Not: the first and the second; but the first; therefore not the second; (4) 'Either the first or the second; but the first; therefore not the second'; (5) 'Either the first or the second; but not the second; therefore the first'. These are considered irreducibly primitive argument forms, hence called the five 'indemonstrables'.

A great deal of work went into the diagnosis of argument validity (see Logic, ancient §5). Both formal rigour and inferential cogency were demanded. A valid argument, according to Stoic logic, is one in which the conclusion follows from the conjunction of the premises in just the way in which in a true conditional the consequent follows from the antecedent. Hence the notion of 'following' became a focal point of debate. Some Stoics adopted a truth-functional analysis of a true conditional: the second proposition follows from the first just if 'Not (the first and not the second)' is true (see Philo the Dialectician for the background to this). But Chrysippus adopted a much stronger criterion, called *synartēsis* or 'cohesion': 'If the first, the second' is true just if the negation of the second is incompatible with the first. This restricts 'following' to a strongly conceptual kind of implication.

The restriction in turn led Chrysippus to the view that many ordinary-language uses of 'if' do not represent real conditionals, and should properly be expressed by the negated conjunction: not 'If the first, the second', but 'Not both: the first and not the second'. He applied this, for example, to the rules of empirical sciences such as divination (see §20); and analogous doubts led at least some Stoics to deny the validity of inductive inference. It was also applied, importantly, to the individual steps of the Little-by-little argument or Sorites (see Vagueness §2). The Academic sceptics plagued the Stoics with this puzzle, aimed at challenging important philosophical distinctions by asking to be told the exact cut-off point. The archetypal Sorites is 'If two grains are not a heap, three aren't; if three aren't, four aren't;…' - and so on to the conclusion that if two grains are not a heap then 10,000 are not. Chrysippus, who wrote extensively on this problem, advised that at some point in the procedure there will be premises of this form which it is proper neither to affirm nor to deny. But he also seems to have authorized the Stoic practice of insisting that the individual steps be formulated not as conditionals, as above, but as negated conjunctions: for example, 'Not both: seven grains are not a heap and eight grains are a heap'. Part of the point is clearly that, whatever the grounds for asserting such a premise may be, there is no actual incompatibility between 'Seven grains are not a heap' and 'Eight grains are a heap', so that assent to the premise must remain optional.

A wide range of other logical puzzles exercised Chrysippus and other Stoics. Some turned on ambiguities, and they developed a sophisticated classification of ambiguity types. The most persistent thorn in their flesh, however, was the liar paradox: 'I am lying' is, if false, true, and, if true, false (see Semantic paradoxes and theories of truth §1). This, along with the Sorites, was wielded by the Academics as a potentially lethal weapon against Stoic logic. It appeared to undermine the most basic tenet - that every proposition is either true or false. Chrysippus wrote many books in refutation of the liar paradox, but if he had his own favoured solution it is impossible now to recover it from our sources.

Inferential logic is ubiquitously employed in Stoic texts on physics and ethics. These regularly purport to demonstrate their tenets, and a demonstration is itself defined as an argument in which evident premises serve to reveal a non-evident conclusion.

12 Cognitive certainty

The main Stoic epistemological theorist was Zeno of Citium, who developed his ideas in response to a series of challenges from the Academic sceptic Arcesilaus (see Arcesilaus §2). His key term is *katalēpsis* - 'apprehension' or 'cognition' - the infallible grasping of some truth, usually by use of the senses. Arcesilaus systematically questioned the grounding of this notion, and argued instead for *akatalēpsia*, 'inapprehensibility', or 'the impossibility of cognition'.

The starting notion is *phantasia*, literally 'appearance' but commonly translated as 'impression' or 'presentation'. To have an impression is simply for things to strike you as being a certain way. Whether or not you take the impression to be true depends on a further cognitive act, assent (*synkatathesis*), which you may give or withhold at will. Since mature human beings are rational, their impressions are called 'rational impressions', meaning that their content can be expressed in language. Strictly speaking it is the proposition associated with it that we are taking to be true when we assent to an impression.

Zeno symbolized an impression by spreading the fingers of one hand, and assent by pulling them together. The next stage, represented by a fist, was *katalēpsis*, literally ‘grasping’, that is, infallibly recognizing the truth. This is not so much successive to assent as an ideally successful *way* of assenting, based on a sufficiently lucid impression. If instead someone assents to an unreliable impression, that will not count as cognition, but as mere fallible ‘opinion’ (*doxa*).

One final stage in the hand simile is *epistēmē*, ‘knowledge’ - not everyday knowing (which is better identified with *katalēpsis*) but absolute, scientific knowledge, such as is possessed only by the wise. Zeno symbolized *epistēmē* by bringing the other hand over to grasp his fist firmly. The point is as follows. *Katalēpsis* is infallible, in that it successfully applies a simple guaranteed cognitive mechanism shared by virtually all human beings: the truth stares you in the face, and you assent to it without the slightest possibility of being wrong. However, although on this very ordinary model of knowing all of us know lots of things, only the genuinely wise know that they know. This is because the wise, having a complete set of mutually supporting cognitions, could never be argued into disbelieving one of them. The unwise, that is, most people, could be argued out of their assent to genuine cognitions, because they are likely to have also a number of false beliefs, adopted as if they were genuine cognitions. You may, for example, plainly witness a miracle but disbelieve it because, being an Epicurean sympathizer, you falsely believe that the gods do not intervene.

The interest of the Stoic-Academic debate lies largely in its concentration, not on the wise, but on the infallible cognition attributed to ordinary people. The Stoics remained convinced, perhaps like most of us, that in everyday encounters where the truth stares you in the face you cannot be wrong and would be insane to withhold assent. But this common-sense position proved extraordinarily vulnerable to Academic criticism.

Debate centred on the kind of impression to which assent might be fully justified. Zeno called this the *phantasia katalēptikē*, ‘cognitive impression’, and initially defined it as one which is ‘(i) from something real (*apo hyparchontos*), and (ii) moulded and impressed according to that real thing itself’. Here (i) is likely to mean, not ‘caused by something real’ (no impression, however delusory, could be either caused by something unreal or totally uncaused), but ‘representing something actual’ - where it is indifferent whether the thing represented is an object or a fact. What (ii) adds is that the impression depicts this thing in full graphic detail. Arcesilaus’ complaint was that such an impression would still not be unmistakably true, since an identical but false impression could occur. This led Zeno to add a third clause, ‘(iii) such as could not be from something unreal’, which we can take to mean ‘such as could not represent some non-actual object or fact’.

The problem now was whether there could be such an impression. The ensuing debate focused partly on the nature of external objects, with the Stoics asserting that each object - even an identical twin - has some unique feature (see §6 for the metaphysical justification of this), and the Academics asking in reply how we could ever be sure that the relevant feature was currently evident. The Stoics also tried to describe the phenomenological features of these cognitive impressions, suggesting that as a species of impression they somehow carried their own badge of identity ‘just as horned snakes differ from other snakes’. The Academic reply, now through the mouth of Carneades, included appeal to cases where false impressions are so graphic that they lead to action, exactly as allegedly cognitive ones do: can the latter then have any intrinsic features to differentiate them from the former?

The final word belongs to a group reported simply as the ‘later Stoics’ (Sextus Empiricus, *Against the Professors* VII 253-60), probably including Antipater, a contemporary of Carneades. They point out that in all but the most special circumstances we simply have no choice whether to assent to a cognitive impression. Once you have such an impression, you do believe it, and that is that. What is in our own hands is not whether, once we are in a position where the truth is unmistakable, we accept it, but whether we take the trouble to get ourselves into such a position in the first place (moving closer, turning on the light, and so on).

This is quite an effective reply to those, like the Academics, who recommend suspending assent. But it leaves the question why the irresistibility of certain impressions should imply their truth. The Stoics here had recourse to a teleological argument: we would not have been given this superb cognitive equipment for any other purpose than to learn the truth. But of course the Academics could easily question the empirical premises on which that teleology was founded, bringing the debate full circle.

13 Criteria of truth

To identify the ‘criterion (or criteria) of truth’ was a standard requirement of Hellenistic philosophers (see Hellenistic philosophy). The Greek word *kritērion* is literally a ‘discriminator’, and a common equivalent was *kanōn*, ‘yardstick’. A criterion of truth was expected to be something naturally available to every mature human being as a basis for distinguishing true from false. Since it was that which made progress towards philosophical understanding possible, its availability and use could not be restricted to those who were already philosophers. When it came to naming the criterion of truth, the Stoics differed among themselves, but in some form or other they all identified it primarily with the ‘cognitive impression’ (see §12), the concept on which the Academic attack on the criterion likewise focused.

A second criterion, widely used in Stoicism and formally named as a criterion at least by Chrysippus, was *prolēpsis*, inadequately translated as ‘preconception’. A *prolēpsis*, literally ‘prior grasp’, is any naturally acquired generic ‘conception’ (*ennoia*) of a thing. Two other terms which are in most contexts interchangeable with it are *koinē ennoia*, ‘common conception’ (that is, common to all human beings) and *physikē ennoia*, ‘natural conception’. These descriptions distinguish *prolēpsis* from artificially acquired conceptions, usually culture-dependent ones, most of which are not directly given in experience but depend on a synthetic mental process. A centaur, for example, is arrived at by combining natural conceptions, a giant by enlarging them, and so on. Some artificial conceptions are liable to be misleading, but others are an integral part of scientific understanding, for example, one’s conception of the centre of the earth, acquired ‘by analogy with smaller spheres’. Human reason is itself simply an ample stock of conceptions, some but not all of them natural ones.

What makes a *prolēpsis* a reliable guide to truth is precisely the fact that the natural conception has not been tampered with. But where does the *prolēpsis* itself come from? The Stoics sometimes sound like hard-line empiricists, as when they compare the mind of a new-born infant to a blank sheet of paper which will in due course have its stock of natural conceptions written on it by repeated sense impressions, classified and stored as ‘experience’. Here a *prolēpsis* is ‘natural’ in the sense of being mechanically imprinted on us, and hence unmediated by fallible reasoning. However, some texts indicate that at least basic moral notions are called natural for the quite different reason that they are dispositionally innate in us.

Many Stoic arguments proceeded from appeals to some *prolēpsis* or ‘common conception’. This was their version of the widespread philosophical practice of citing what are alleged to be ‘our intuitions’. It ran into the difficulty that such a practice always faces: separating genuinely natural conceptions from those infected by one’s culture or other beliefs becomes the new bone of contention. For example, both Stoics and Epicureans appealed to the *prolēpsis* of ‘god’, but while the Stoics regarded providence as an integral part of this *prolēpsis*, the Epicureans argued that god’s providentiality was a cultural imposition on the basic *prolēpsis*, motivated by human bafflement at the world’s workings (see Epicureanism §9). A very similar dispute launches Stoic ethics.

14 *Oikeiōsis*

Epicurus proposed a method for identifying the genuinely natural human value: consult a new-born baby. Inarticulate infants, and for that matter irrational animals, cannot possibly have been infected yet with the norms of society, and their actions tell us, louder than any words, that their sole motivation is the acquisition of pleasure and the avoidance of pain. Stoic ethics responds by adopting the same starting point but questioning the Epicurean analysis of infant behaviour.

The highly influential concept which the Stoics introduced to facilitate their own analysis is *oikeiōsis*, variously translated as ‘appropriation’, ‘familiarization’, ‘affiliation’ or ‘affinity’. Literally, this is the process of ‘making something one’s own’. An animal’s *oikeiōsis* is its natural impulse or inclination towards something which it regards as belonging to it.

A creature’s first *oikeiōsis*, Stoics argue, is towards itself and its own constitution, a priority which it displays by making self-preservation its dominant goal. Far from pursuing pleasure, it courts pain in order to preserve and develop its natural constitution, as when we see a toddler repeatedly fall in striving to walk, and an overturned tortoise struggling to regain its upright position. As the human child develops, its *oikeiōsis* is extended beyond itself: it treats its parents and siblings as belonging to it, and cares for them accordingly, in much the same way in which it already cares for itself. In due course this same other-concern is extended to cover a wider range of people, albeit in increasingly diluted measure. At an extreme it takes in the entire human race. (For a graphic Stoic

elaboration of this idea, see Hierocles.)

Oikeiōsis is a continuum, stretching from the instinctive self-preservation of the new-born infant to the other-regarding conduct which is equally natural in rational adults. Where most ancient ethical systems struggled to explain altruism as an extended form of self-interest, there is no such tension in Stoicism, where others already fall within the ambit of our natural affection in much the same way as we ourselves do. This rationally extended sense of what belongs to us does not yet amount to moral goodness, but it is its indispensable basis. Goodness lies in our understanding and collaborating with the ideally rational world plan. It is no wonder that our natural *oikeiōsis* towards the rest of the human race should be what grounds the project of completely integrating ourselves into that plan.

Oikeiōsis is an affinity founded on the shared rationality of the entire human race. The doctrine thus helped to foster Stoic cosmopolitanism and other widely admired humanitarian stances (see §18). Seneca (§1), for example, reminded his readers of their moral obligations even to their slaves. Conversely, however, the *oikeiōsis* doctrine also encouraged a hardening of attitudes to non-rational animals, with which humans were judged to stand in no moral relation at all.

15 The indifferents

Perhaps the most characteristic doctrine of Stoic ethics is that virtue alone is good, vice alone bad. Everything else traditionally assigned a positive or negative value - health or illness, wealth or poverty, sight or blindness, even life or death - is ‘indifferent’. By making this move, the Stoics authorized the use of the word ‘good’ in a distinctly moral sense - a usage which is still with us, although they themselves bought it at the high price of simply denying that the word, properly understood, has any other sense.

The inspiration of this doctrine is undoubtedly Socratic. In various Platonic dialogues (see especially *Euthydemus* 278-81, *Meno* 86-9), Socrates argued that most things traditionally called good - typified with largely the same examples as the Stoic ‘indifferents’ - are in their own nature intermediate between good and bad. If used wisely, they become good, if unwisely, bad. Hence wisdom is the only intrinsically or underivatively good thing (see Socrates §§4-6).

This Socratic argument encouraged the Cynic idea that only wisdom - or more generally, virtue - is good, and that such coveted possessions as reputation, health and physical comfort are literally irrelevant to the goodness, and hence the happiness, of one’s life (see Cynics). The Cynics acted on this by adopting a bohemian lifestyle, disdaining the values of society. And Zeno, the founder of Stoicism, had his first philosophical training from a Cynic. Zeno’s independent-minded colleague Ariston stuck close to the Cynic thesis (see Ariston of Chios §2). Zeno, on the other hand, modified it in a way which does more than anything else to account for the widespread success of Stoicism; Zeno’s subtly revised position leaves wealth, fame, and so on morally indifferent, while explaining why we are nevertheless fully justified in pursuing them.

Although being healthy does not make you happy, Zeno maintains, the natural thing to do in ordinary circumstances is nevertheless to stay healthy and avoid illness. We should not try to suppress this natural instinct, because to be happy - the ultimate goal to which we all aspire - is to be totally in tune with nature. Therefore the proper way to start out is to respect the preferences which nature dictates, opting where possible for affluence, high civic status, family values and other ‘natural’ desiderata. As you progress, you will learn when to vary the formula. It may be that in special circumstance the right way for you to fit in with nature’s plan is to be poor, or sick, or even to die. If you understand why one of these is the rational and natural thing for you, you will embrace it willingly, and thus further rather than hinder your project of perfect conformity with nature. But barring such special circumstances, the natural values to adopt coincide on the whole with the ordinary values of society.

This leads, in typical Stoic fashion, to a terminological jungle of epithets for the ‘indifferents’. The ‘things which accord with nature’ (*ta kata physin*), such as health, have a positive, albeit non-moral, ‘value’ (*axia*), and are therefore labelled ‘preferred’ (*proēgmena*), which means that in normal circumstances we should opt for them, they are ‘to be taken’ (*lēpta*). The ‘things which are contrary to nature’ (*ta para physin*), such as illness, earn a contrary set of technical terms: ‘disvalue’ (*apaxia*), ‘dispreferred’ (*apoproēgmena*), ‘not to be taken’ (*alēpta*).

The linchpin of Stoic ethics is the way in which it legitimizes a familiar scale of personal and social values, while

denying them any intrinsic worth. Their value is purely instrumental, because they are the subject matter of the choices by means of which we progress towards true moral understanding. We might compare the relative ‘values’ of, say, illness, fame and eyesight, in Stoic eyes to the relative values of cards in a card game. Learning how to choose between these, and even to sacrifice cards of higher value when the circumstances dictate, is an essential part of becoming a skilled player. But these choices matter only instrumentally: It would be absurd to compare the value of an ace to the value of being a good card-player. In Stoic eyes it is an equally grave error - although unfortunately one of which most people are guilty - to rank wealth or power along with moral goodness on one and the same scale.

The things which are naturally ‘preferred’ can be encapsulated in rules: honour your parents, take care of your health, cultivate friends, and so on. From the start - although again with the dissent of Ariston of Chios - the Stoics attached importance to rules or ‘precepts’ as the basis of moral progress. What a precept prescribes is a *kathēkon* (plural *kathēkonta*), a ‘proper function’ or ‘duty’, and many Stoic treatises were devoted to working out detailed lists of *kathēkonta*. A *kathēkon* is defined as ‘that which, when done, has a reasonable justification’: for a rational adult, what is reasonable and what is natural should coincide.

There are two main types of *kathēkonta*: circumstantial and non-circumstantial. Circumstantial *kathēkonta*, that is, those prescribed only in very special circumstances, include such abnormal acts as self-mutilation, giving away your property, and even suicide (something of a Stoic obsession, inspired by Socrates’ willing death). Non-circumstantial *kathēkonta* are, despite their name, not prescribed in literally all circumstances, since to each non-circumstantial *kathēkon* (for example, looking after one’s health), there is opposed a circumstantial one, (for example, in very unusual circumstances, getting ill). Rather, they are ‘non-circumstantial’ because they are what, other things being equal, you should do as a matter of course, and not as a response to your present circumstances.

16 Goodness

Kathēkonta are ‘intermediate’ patterns of behaviour - that is, available to everybody, wise and non-wise alike. Yet in advertizing them the Stoics regularly referred to the conduct of the ‘sage’, the idealized wise person whom they always held up as a model, despite admitting that the criteria for this status were so tough that few people, if any, ever attained them. What was possible for everybody, they insisted, was progress (*prokopē*) towards this state of wisdom, and that is why they stressed the continuity between the proper conduct of the non-wise and the ideally good conduct of the wise. When you actually become wise and virtuous, what are outwardly the very same kind of *kathēkonta* which you were already habitually performing are suddenly transformed by your new state of understanding, earning themselves the name *katorthōmata* or ‘right actions’.

Alongside this continuity in moral progress, there is also the sharpest possible discontinuity. One of the most notorious Stoic paradoxes was that all sins are equal. If you are not virtuous and wise, you are totally bad and foolish. The wise are totally happy, the foolish totally unhappy. Whatever strides you may have made towards virtue, you are no happier till you get there. They compared what it is like to be drowning: whether you are yards from the surface or only inches from it, you are still just as effectively drowning.

The motivation of this depressing thesis is not entirely clear. Stoic concern with the paradox of the Sorites (see §11) may have contributed to it, but the main driving force seems to be the conviction that actual goodness, if achieved, differs not in degree, but in kind, from the scale of natural values. At a certain point of moral development, you notice an emerging agreement or harmony between your individual choices and acts. It is, thereafter, not the choices and acts or their objects that matter any longer, but harmony for its own sake. Only from that point on do you have a conception of what goodness is: it is located in a perfect ‘agreement’ both within the individual and between that individual and cosmic nature.

What does this agreement consist in? Despite the Stoics’ extensive cataloguing and classification of the *kathēkonta* which the sage will perform, ultimately the wise are characterized, not by the actual success of their actions - which may not always be in their control - but by the morally perfect frame of mind with which they act - in other words, by virtue (see Aretē). Socrates had propounded that paradox that virtue is knowledge: all there is to being good is to *know* the right things. The Stoics develop this Socratic idea to the full. The word for knowledge - *epistēmē* - can also more specifically mean ‘science’, and they regard each virtue as a genuine science, complete with its own constituent theorems. The skill of living in harmony is a skill analogous to, although vastly more

difficult than, any branch of mathematics or medicine.

Plato had given four virtues canonical status: justice, wisdom, temperance or self-control (*sōphrosynē*), and courage. The Stoics adopt this list, and treat all other virtues as subordinate species, or perhaps branches, of the four. Are these, then, four entirely distinct sciences? No. Among Socrates' most enduringly influential doctrines was that of the Unity of the Virtues. On one widely accepted version, adopted by some early Stoics (see Ariston of Chios §3; Cleanthes), Socrates' thesis meant that the four virtues are all simply one and the same state of mind, albeit going under different names in different external circumstances. Others took the view that the unity of the virtues consists rather in their inseparability. For example, you could hardly count as just if you were not brave and temperate too, or else you might be deflected from just behaviour through intimidation or bribery. It is perhaps this thought that led Chrysippus to the following view: the four virtues are four separate sciences, each with its own defining set of theorems; but each virtue incorporates and uses the theorems of the other three as subsidiary theorems. The thesis that the virtues are distinct sciences was put by Chrysippus, in the technical language of Stoic metaphysics (see §6), by saying that they belong to the second category, 'quality', and not the fourth category, 'relative disposition'. That is, they differ from each other as distinct qualities or states of mind, not simply as one state of mind differentiated by the varying external situations with which it is confronted.

17 The goal

The 'goal' or 'end' (*telos*) is defined as 'that for the sake of which everything is done, while it is not itself done for the sake of any further thing'. This is identified with happiness (*eudaimonia*), or 'living well'. Both are commonplace to the Greek philosophical tradition. The partisan content arises when philosophers offer their formulas for what this end actually consists in. Zeno's formula was 'living in agreement' (*homologoumenōs zēn*). The history of Stoic ethics over the next two centuries is largely a history of successive attempts to work out what Zeno must have meant.

Zeno's vagueness was probably deliberate. The 'agreement' comprises both the perfect internal coherence and rationality (the '*-log-*' part of *homologoumenōs* means 'reason') of the good life - 'living in accordance with one concordant reason' - and its conformity with nature, the 'nature' in question being itself equated with both one's own individual nature and the nature of the world. Happiness is also identified as a 'smooth flow of life', and Zeno's real point was that only those with complete understanding of cosmic rationality can make their own aims and choices entirely one with those of nature, and thus never come into conflict with either their own or the world's rationality.

Pressure for clarification led either Zeno himself or Cleanthes to make the first addition to the formula, which now became 'living in agreement *with nature*'. Chrysippus substituted 'living in accordance with experience of what happens by nature'. What became clearer, as these and other formulations competed, was that the ideal life was defined in terms of things which were themselves morally indifferent - the 'things which accord with nature' (see §15). The challenge which the Stoics faced from their opponents in the Academy was how moral good could depend on a set of aims whose attainment was morally indifferent. The answer - compare §15 - was as follows. What matters is not necessarily *achieving* natural advantages like health, which cannot be guaranteed in all circumstances, and which in any case do not bring happiness. What matters is making the right rational choices - doing everything that lies in your power towards achieving what nature recommends. It is the consistency of those *efforts*, not of their results, that may ultimately become perfect agreement with nature, that is, happiness. Formulas for the goal designed to capture this emphasis included (Diogenes of Babylon) 'reasoning well in the selection and disselection of things which accord with nature', and (Antipater) 'to live continuously selecting things which accord with nature and disselecting things contrary to nature'. Zeno's original formulation had more to recommend it.

18 The cosmic city

Everybody without exception strives for a good and happy life, but only the wise achieve it. Most people in fact misapply the very words 'good' and 'happiness', which they mistakenly associate with morally indifferent states like wealth and honour. This simple point came to be extended by the Stoics to all the other things which are conventionally prized. Everybody wants to be rich, free, powerful, beautiful, loveable, and so on, but, paradoxically, only the wise achieve these goals. Everyone else is, whatever they may think, actually poor,

enslaved, powerless, ugly and unloveable. This is because real wealth is to have something of genuine worth (that is, virtue), or to lack nothing that you need; real freedom is to be in full control of your life (including the knowledge of when to accept death rather than ever be forced to do what you do not truly want to do); real power is to be able to achieve everything you want; real beauty is a quality of the soul not the body; and only the genuinely beautiful are genuinely loveable. These Stoic 'paradoxes' are of Socratic inspiration.

A primary motif of Stoic political thought is the extension of such paradoxes into the civic realm. Conventional political ambitions belong to the realm of the indifferent just as much as wealth and health do. Thus, while Stoicism actively promotes conventional political activity as a way of following human nature, it at the same time downgrades it in relation to true moral goodness. Everybody wants to have power, and would like if they could to be a king; but only the wise have power (only they can achieve everything they want) and kingship (defined as 'rule which is accountable to no one'). These Socratic redefinitions were extended even to humbler civic aims: only the wise are generals, orators, magistrates, lawyers, and so on.

An upshot of this was a corresponding downgrading of the civic institution within which such offices operated. A city, in the conventional sense of a human cohabitation with geographical boundaries, a legal code and so on, is an artificial construct. A city in the most correct sense is not constrained in these ways: in fact the world itself is the ultimate city, being a habitation common to humans and gods, united by their shared rationality.

The idea was of Cynic inspiration. The Cynics had already coined the expression 'citizen of the world', *kosmou politēs*, which the Stoics took over. In a way every human being is a citizen of the world, and this generous version of Stoic cosmopolitanism was to become enormously influential on the ideology of the Roman Empire, as well as leading some Stoics to challenge entrenched gender and class barriers. But on a narrower criterion - influenced by Zeno's early utopian work the *Republic* (see Zeno of Citium) - it is not all human beings, but only the wise, who participate in the real cosmic city. The cosmic city has its own law, a natural moral law defined as 'right reason (*orthos logos*) which commands what should be done and forbids what should not'. This notion of a cosmic moral law which transcends local legal codes exerted a powerful influence on later theories of natural law.

Although the Stoics encouraged political involvement in conventional cities, and were themselves prepared to act as advisors to monarchs, there is little reason to think that any Stoic before the late second century BC (see Panaetius §2) made a serious contribution to non-utopian political theory.

19 Passions

Everyone who has not achieved virtue is in a state of vice or moral badness. Most commonly - for example, in the work of Plato and Aristotle - vice was viewed as a state in which reason is dominated and deflected by strong irrational emotions, or 'passions' (*pathē*, singular *pathos*). But Socrates had established an enduring intellectualist alternative, according to which the soul has no irrational parts, and virtue is knowledge, so that its lack, vice, is simply ignorance: 'No one does wrong willingly. The Stoics are fully committed to developing Socrates' position, in particular the thesis that passions are really value judgments.

A passion is commonly thought of as disobedient to reason. Reason says that you should face some danger, but fear disobeys. Reason chooses to abstain from embezzlement, but greed wins out. This suggests that an emotion can hardly itself be a rational state. The Stoics accept the description of emotions as 'disobedient to reason', but redescribe what this amounts to.

An emotion is primarily a judgment - a false one. A fear may be the false judgment that some impending thing, say injury, is bad for you. The falsity lies in the fact that physical injury is actually not bad, just a 'dispreferred indifferent' (see §15) and therefore strictly irrelevant to happiness. Your belief that it is bad takes the form of an 'excessive impulse' to avoid it, and that impulse, as well as being a judgment, is like any intellectual state also a physical modification (in this case called a 'contraction') of the *pneuma* that constitutes the commanding-faculty of your soul. The new overtensioned and perturbed state of your mental *pneuma* is one that you cannot instantly snap out of. Were you to entertain the correct judgment that you should not shrink from the danger, your *pneuma* would not be able to respond. That is what makes the passionate state of fear 'disobedient to reason' - a status it can have while itself also being a piece of faulty reasoning. Chrysippus compared it to a runner who is going too fast and therefore cannot stop at will. (For a later Stoic's disagreement with Chrysippus on this issue, see Posidonius §5.)

The four main kinds of emotion are appetite, fear, pleasure and distress. Appetite and fear are faulty evaluations of future things as good and bad respectively. Pleasure and distress are corresponding mis-evaluations of things already present. Each has a variety of sub-species, and one of particular importance in Stoic discussions (see Seneca, *On Anger*) is anger, identified as a species of desire, namely the desire for revenge. Calling pleasure a passion and a vice may sound harsh, but the kind of pleasure envisaged here is one involving conscious evaluative attitudes, such that its sub-species include gloating and self-gratification. ('Pleasure' understood as that sensation of wellbeing which automatically accompanies certain states and activities is not a vice but an 'indifferent'. It is the view that pleasure - in this latter sense - and pain are indifferent that has given 'stoical' its most familiar modern meaning.)

It should not be inferred that a Stoic sage is feelingless. The wise lack the 'passions', which are overevaluations, but they do instead have the correct affective states, which the Stoics call *eupatheiai*, or 'good feelings'. Thus the sage has no 'appetites', but does have 'wishes', whose species include kindness, generosity, warmth and affection. Similarly, instead of 'fear' the wise have 'watchfulness', and so on.

The Stoics' conviction that emotional states, far from being mere irrational drives, are primarily specified by their cognitive content is one of their most valuable contributions to moral philosophy. Its most important implication in their eyes is that philosophical understanding is the best and perhaps the only remedy for emotional disquiet. In the short term strong emotions are disobedient to correct reasoning, but in the long term rational therapy can restructure the intellect and dispel all passions (see Emotions, nature of).

20 Fate

Socrates had been a firm believer in the powers of divination and in divine providence. Stoicism took over this outlook and developed it into a doctrine of 'fate' (*heimarmenē*), which by the time of Chrysippus had become a full-scale thesis of determinism.

That everything that happens is predetermined is a thesis which flows easily from all three branches of Stoic philosophy. Ethics locates human happiness in willing conformity to a pre-ordained plan (§17), and treats the use of divination as a legitimate means towards this goal. Physics provides the theory of the world's divinely planned cyclical recurrence, unvarying in order to maintain its own perfection (§5).

Physics also supplies a fundamental principle, regarded as conceptually self-evident, that nothing happens without a cause. This quickly leads to the conclusion that the world's entire history is an unbroken causal network. 'Fate is a natural everlasting ordering of the whole: one set of things follows on and succeeds another, and the interconnection is inviolable'. 'The passage of time is like the unwinding of a rope, bringing about nothing new and unrolling each stage in its turn.' A modern analogy might be the continual rerunning of a film.

Finally, logic offers the principle of bivalence: every proposition, including those about the future, is either true or false. Therefore, Chrysippus argued, it is true *now* of any given future event either that it will happen or that it will not happen. What does that present truth consist in? It can only lie in the causes now present, sufficient either to bring the event about or to prevent its happening. Therefore all events are predetermined by antecedent causes sufficient to bring them about and to prevent all alternatives from occurring. (Compare Aristotle §20 and Epicureanism §12, for escape routes from this argument.)

21 Responsibility

The greatest interest of this determinist position lies in the Stoics' attempts to meet the challenge it poses to moral responsibility (see Free will). They implicitly accept that a person is responsible for an action only if they could have done otherwise. But how could this latter be true in a Stoic world, where the actual action performed is causally determined and even predictable in advance? Chrysippus was the author of the main Stoic answers to this challenge. His task (see Cicero, *On Fate*) was to show that even in such a world 'could have done otherwise' makes sense: an action which I did not in the event perform may nevertheless have been *possible* for me, that is, my failure to perform it was not *necessary*. The strategy for securing this result included the following lines of argument.

(1) A 'possible' proposition is defined in Stoic sources as 'one which (i) admits of being true, and (ii) is not

prevented by external circumstances from being true'. Suppose that you have failed to pay a bill despite having the cash. Paying the bill was 'possible' for you. (i) It 'admits of being true': there is such a thing as paying a bill, unlike for example, being in two places at once. (ii) Nothing external to you prevented you: you did not lack the funds, you were not forcibly detained, and so forth. This account of possibility does allow that something *internal* to you may (indeed must) have prevented you from paying: for example, your meanness, forgetfulness or laziness. Still, it was possible for you to pay, in the sense that you had the *opportunity* to pay. Chrysippus seems to maintain that the 'could have done otherwise' notion of responsibility holds in his world, because alternative actions are 'possible' in just this sense: we regularly have the opportunity to do otherwise, and therefore have only ourselves to blame for what we actually do.

Stoicism resists the alternative that 'could have done otherwise' might entail our being actually *capable* of acting otherwise: surely the good, in order to claim credit for their conduct, do not have to be capable of wrongdoing, nor need the bad, if they are to be blamed, be capable of acting well.

(2) Much this same point was also expressed in terms of a causal distinction (one among many: the Stoics were accused of introducing a 'swarm of causes'). When a cylinder is pushed and rolls, the 'principal' cause of the rolling is its shape, the push being just the triggering or 'proximate' cause. From our point of view as agents, Chrysippus argues, fate functions as a series of mere triggering causes - the openings, provocations and so on with which the world presents us. Given our respective states of character, we are bound to respond to these prompts in just the ways we do. But the principal cause lies in our character. And that makes us, not fate, responsible for our actions.

Importantly, the Greek word for a cause, *aition*, literally means the 'thing responsible'. However, the Stoics' technical term for moral responsibility is *eph' hēmin*: our actions are 'within our power'. This is not a thesis of free will. What matters to them is not to posit an open future, but to establish the moral accountability of human action even within a rigid causal nexus.

(3) Divination might be thought to make the future necessary. Take an astrological law: if you were born at the rising of the dog star, you will not die at sea. Suppose you *were* born at the rising of the dog star. This is an unalterable fact about the past, and therefore, Chrysippus accepts, necessary. There is also a widely accepted law of logic that if the antecedent of a conditional is necessary so is its consequent. Therefore 'You will not die at sea' also becomes a necessary truth. One of Chrysippus' replies was that divinatory laws do not express conceptually indubitable truths, and are therefore not properly expressed as conditionals, but rather as negated conjunctions (see §11), in which this transmission of necessity does not occur.

(4) One remaining challenge was the Lazy Argument. Why, its proponents asked, should we bother to make decisions if the outcome is already fixed? Why call the doctor, if whether you will die or recover from your illness is already fated? Chrysippus' answer is that such sequences of events as calling the doctor and recovering are 'co-fated'. In most cases the outcome is fated *via* the means, not regardless of them.

Some landmark events, however, such as the day of your death, may be fated regardless of the means. Your character will cause you to decline numerous alternative actions to those you will choose, but even if, counterfactually, you were going to choose one of those alternatives, it would still be going to lead to your death on that same day. For example Socrates (in Plato's *Crito*) knew through a prophetic dream that he would die in three days' time, and his reasoned decision to stay and accept execution was willing cooperation with the rational world plan, where a bad person would have resisted by escaping but still died on that same fated day. Zeno and Chrysippus compared a human being to a dog tied to a cart: it can follow willingly, or be dragged.

In this way, morality is not simply argued to be compatible with determinism, but to require it. Only within a framework of rational predestination can moral choices have their true significance. There remains, however, the question why, in a world where it was pre-ordained that we would be precisely the kind of people we are, our choices should have any moral significance at all. The answer is that goodness belongs primarily to the world as a whole (identifiable with god). It is from this that moral qualities filter down to individuals and their actions, as a measure of their cooperation with or obstruction of the rational world plan.

22 Later fortunes

Stoicism's success ran high in the first century AD. It was perceived by writers like Seneca and Lucan as embodying the traditional Roman virtues whose decline was so widely lamented. Roman Stoics formed the main resistance to the emperor's rule, and, following the earlier model of the Stoic Cato, made the principled act of suicide into a virtual art form.

In a way Stoicism's crowning achievement was in AD 161, when its adherent Marcus Aurelius became Roman emperor. Here at last was a genuine philosopher-ruler. When Marcus established chairs of philosophy at Athens, these included one of Stoic philosophy. Nevertheless, Stoicism was already on the decline in the late second century, eclipsed by the revived philosophies of Plato and Aristotle. By then, however, it had entered the intellectual bloodstream of the ancient world, where its concepts remained pervasive in such diverse disciplines as grammar, rhetoric and law, as well as strongly influencing the thought of Platonist philosophers like Porphyry, and Church Fathers such as Clement of Alexandria.

Through the writings of Cicero (whose philosophical works, although not Stoic, embody much Stoicism) and Seneca, Stoic moral and political thought exercised a pivotal influence throughout the Renaissance (see Renaissance philosophy). Early modern philosophers who incorporated substantial Stoic ethical ideas include Spinoza and Kant. The recovery of Stoic physics, epistemology and logic, however, has been largely an achievement of the nineteenth and twentieth centuries.

See also: Cleomedes; Epictetus; Musonius Rufus

DAVID SEDLEY

References and further reading

Arnim, H. von (1903-5) *Stoicorum Veterum Fragmenta (Fragments of the Early Stoics)*, Leipzig: Teubner, with vol. 4, indexes, by M. Adler, 1924.(The standard collection of early Stoic fragments, in Greek and Latin, commonly abbreviated as *SVF*.)

Atherton, C. (1993) *The Stoics on Ambiguity*, Cambridge: Cambridge University Press.(Outstanding rediscovery of a sophisticated theory; technically demanding.)

Brunschwig, J. (ed.) (1978) *Les Stoïciens et leur logique (The Stoics and their Logic)*, Paris: Vrin.(Important collection of papers.)

Burnyeat, M.F. (1982) 'Gods and Heaps', in *Language and Logos*, ed. M. Schofield and M. Nussbaum, Cambridge: Cambridge University Press, 315-38.(Classic study of Stoic treatment of the Sorites; see §11.)

Cicero, M.T. (46 BC) *Stoic Paradoxes*, in *On Stoic Good and Evil: 'De finibus' 3 and 'Paradoxa Stoicorum'*, trans. M.R. Wright, Warminster: Aris & Phillips, 1991.(Latin text with English translation and notes; short declamations defending six Stoic moral paradoxes.)

Cicero, M.T. (45 BC) *On Ends*, book III, in *On Stoic Good and Evil: 'De finibus' 3 and 'Paradoxa Stoicorum'*, trans. M.R. Wright, Warminster: Aris & Phillips, 1991.(Latin text with English translation and notes; an outstandingly lucid defence of Stoic ethics.)

Cicero, M.T. (45 BC) *On the Nature of the Gods*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933.(Latin text with English translation; books II-III are a report and critique of Stoic theology.)

Cicero, M.T. (early 44 BC) *On Divination*, trans. W.A. Falconer, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1923.(Latin text with English translation; defence, followed by a critique, of Stoic belief in divination.)

Cicero, M.T. (mid 44 BC) *On Fate*, trans. R.W. Sharples, Warminster: Aris & Phillips, 1991.(Latin text with English translation; includes much of our best information on the Stoic doctrine of fate, see also §§20-1.)

Colish, M.L. (1985) *The Stoic Tradition from Antiquity to the Early Middle Ages. I, Stoicism in Classical Latin Literature* Leiden: Brill.(A very thorough survey of Stoic influence.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press, 1925, 2 vols. (Greek text with English translation; book VII, in volume 2, contains the lives of the leading Stoics, along with much doxographical and other information.)

Dragona-Monachou, M. (1976) *The Stoic Arguments for the Existence and Providence of the Gods*, Athens: National University Press.(Valuable study of Stoic theological arguments.)

Epp, R.H. (ed.) (1985) *Spindel Conference 1984: Recovering the Stoics*, suppl. vol. 23, *Southern Journal of Philosophy*.(Collection of articles on Stoicism, together with a comprehensive bibliography.)

Erskine, A. (1990) *The Hellenistic Stoa*, London: Duckworth.(A challenging reinterpretation of Stoic political thought and practice.)
Frede, M. (1974) *Die stoische Logik (Stoic Logic)*, Göttingen: Vandenhoeck & Ruprecht.(The best study of Stoic logic.)
Hülser, K. (1987) *Die Fragmente zur Dialektik der Stoiker (The Fragments on the Dialectic of the Stoics)*, Stuttgart and Bad Cannstatt: Frommann-Holzboog, 4 vols.(State-of-the-art collection, with German translation and commentary, of the texts bearing on Stoic dialectic; abbreviated as *FDS*.)
Inwood, B. (1985) *Ethics and Human Action in Early Stoicism*, Oxford: Oxford University Press.(Among the most valuable studies of Stoic ethics.)
Inwood, B. and Gerson, L.P. (1988) *Hellenistic Philosophy, Introductory Readings*, Indianapolis, IN: Hackett. (Includes a large body of primary texts on Stoicism, in translation.)
Kneale, W. and Kneale, M. (1962) *The Development of Logic*, Oxford: Oxford University Press.(Contains what is still among the best evaluations of Stoic logic.)
Long, A.A. (ed.) (1971) *Problems in Stoicism*, London: Athlone.(Collection of articles, some seminal.)
Long, A.A. (1974) *Hellenistic Philosophy*, London: Duckworth.(Includes what in most ways is still the best introductory study of Stoicism; very accessible.)
Long, A.A. (1996) *Stoic Studies*, Cambridge: Cambridge University Press.(Almost complete collection of articles by the leading post-war specialist in Stoicism.)
Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press. (Volume 1 contains Stoic sources in translation with commentary; volume 2 has the original texts.)
Mates, B. (1961) *Stoic Logic* (2nd edn; 1st edn 1953), Berkeley, CA: University of California Press.(Largely superseded by Frede (1974), but still an exceptionally lucid and helpful introduction.)
Nussbaum, M. (1994) *The Therapy of Desire*, Princeton, NJ: Princeton University Press.(An eloquent defence of the Stoic intellectualist treatment of emotions, see also §19.)
Pohlenz, M. (1959) *Die Stoa*, Göttingen: Vandenhoeck & Ruprecht, 2nd edn, 2 vols; 1st edn, 1948.(The classic study of the school, still not fully superseded; later editions do not alter the main text.)
Rist, J.M. (ed.) (1978) *The Stoics*, Berkeley, CA: University of California Press.(First-rate collection of papers; includes valuable introductory accounts of Stoic cosmology and logic.)
Sambursky, S. (1959) *Physics of the Stoics*, New York: Macmillan.(A classic attempt to link Stoic to modern physics.)
Sandbach, F.H. (1975) *The Stoics*, London: Chatto & Windus.(An outstandingly accessible introduction.)
Sandbach, F.H. (1985) *Aristotle and the Stoics*, Cambridge: Cambridge Philological Society.(Seminal but controversial critique of the view that Stoicism was indebted to Aristotle.)
Sedley, D. (1982) ‘The Stoic Criterion of Identity’, *Phronesis* 27: 255-75.(On the Growing Argument and the ‘categories’; see also §6.)
Sextus Empiricus (2nd century AD) *Against the Professors*, books VII-XI, trans. R.G. Bury, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1935-6.(Greek text with English translation; contains extensive reports of Stoic doctrine and argument.)

Strato (d. *c*.269 BC)

The third head of Aristotle's school, from c.*287 to* c.*269 BC, Strato has been regarded as substituting materialism for Aristotelian metaphysics, mechanism for teleology, atheism for theology and empiricism for intuition; and he has been blamed for the decline of Aristotle's school, or (less often) praised for his adoption of a more scientific outlook, especially in psychology. However, on some issues at least it may be more a matter of a selective emphasis of certain parts of Aristotle's teachings than of conscious and deliberate anti-Aristotelianism. None of Strato's writings survives.*

1 General

Strato, whose date of birth is unknown, was the third head of the Lyceum, the school founded by Aristotle. After teaching the future king Ptolemy II Philadelphus in Alexandria, he succeeded Theophrastus (§1) as head of the school on the latter's death *c*.287 BC and remained in office until his own death eighteen years later. In some respects the most innovative and original member of the school after Aristotle, he has been held responsible for its subsequent decline, on the grounds that he abandoned certain Aristotelian doctrines; but there is room for debate both about Strato's actual views - given our complete dependence on second-hand reports - and about their relation to Aristotle's (see Peripatetics).

Where logic and ethics are concerned we know little of the contents of Strato's works, as opposed to their titles, and according to Cicero his interest in ethics was slight (frs12-13, Wehrli 1969); these reports must be read in the context of the perspective on the history of philosophy which Cicero derived from Antiochus, but may still be accurate enough as to the actual facts. Indeed Strato was often referred to by ancient sources as 'Strato the natural philosopher' (*ho physikos*).

2 Mechanism

Strato has often been interpreted both as a materialist and mechanist; in the eighteenth century 'Stratonist' was applied to those who denied divine influence on processes in the natural world. Purpose and form are connected in Aristotle, and Strato was supposed to differ from him on both issues. Certainly Strato (like Theophrastus before him, as far as our evidence goes) shows little interest in the role of the formal cause or in such issues as the status of universals, and is more concerned with the interactions of the elements than the division into matter and form. He regarded heat and cold as the fundamental principles of physical change (frs 42-8, Wehrli 1969), following Aristotle (*Generation and Corruption* II.2). The influence of Strato's approach can be seen in works attributed to Aristotle but in fact written in the third century, such as the *Problems*, the *Mechanics* and *On Things Heard* (*De audibilibus*).

That Strato rejected teleological explanations as a matter of principle, however, is less certain. Cicero reports him as saying that things are created by natural weights and movements, not by god (fr. 32, Wehrli) and as placing in nature a divine power which causes coming-to-be, growth and decay but lacks consciousness and form or shape - Latin *sensus* and *figura* - which in the context probably means, as Repici (1988) has argued, that it is not an anthropomorphic divinity (fr. 33, Wehrli 1969, see also frs 34, 36, 37). None of this sounds un-Aristotelian; Aristotle himself does not hold that things are created by god rather than by natural processes, does not regard nature as a conscious force, and shows in his zoological works how purpose and necessity work together for the most part but not always. Plutarch attributes to Strato (fr. 35, Wehrli 1969) the view not only that the world is not a living creature (where Aristotle, again, would agree; *On the Heaven* II.2, sometimes appealed to for the contrary view, is in fact concerned with the heaven, not the whole world), but also that chance (or 'the spontaneous': *to automaton*) is primary, nature secondary; this has been explained by scholars as a distortion resulting from Plutarch's desire, in the context, to stress Strato's disagreement with Platonism, or as indicating that, while processes in the world are brought about by nature, there is no further source or explanation of the existence of movement in the world as a whole, such as the Unmoved Mover provides in Aristotle's theory.

We have, indeed, no instances of Strato's actually using teleological explanations of nature; but our evidence is scanty in any case. It may be appropriate to speak, not of a rejection of Aristotelian positions by Strato, but of a selective emphasis on certain aspects of Aristotle's teaching rather than others. Strato almost certainly rejected

Aristotle's doctrine of the Unmoved Mover, which Theophrastus had questioned and which Aristotle had not held throughout his career. He also (fr. 84, Wehrli 1969) rejected the fifth element, the incorruptible *aithēr* of which Aristotle held the heavens were made.

3 Microvoid and *horror vacui*

In the physics of the sublunary world Strato's greatest innovation was his readiness to allow the temporary existence of imperceptibly small and completely empty void spaces - 'microvoids' - within material bodies (frs 54-67, Wehrli 1969), using this to explain the compressibility of bodies and penetration, such as that of light through water. The idea that penetration of one body by another is to be explained in terms of 'pores' or passages is already present in Theophrastus and in book IV of Aristotle's *Meteorology* (which some have, questionably, attributed to Strato himself or to Theophrastus, rather than to Aristotle), but here the pores seem not to be void but to be filled with a substance different from that surrounding them. (Gatzemeier's (1970) view that no more than such a theory of *potential* void was to be attributed to Strato has been refuted by Furley (1985).) Where Aristotle had held that light was propagated as a change in a medium, Strato regarded light-rays as material.

There is no direct evidence for Strato's use of the principle of 'nature abhorring a vacuum' (henceforth *horror vacui*); fr. 5 (Gottschalk (1965)), from Theophrastus' *Meteorology* on the motion of the winds, is claimed for Strato rather than Theophrastus by Gottschalk, but the attribution has subsequently been challenged. The doctrines of *horror vacui* and of microvoids are logically distinct; but it seems likely that Strato combined them, by relating *horror vacui* only to a 'massed' or 'perceptible' vacuum. Such a connection was made by the Alexandrian physician Erasistratus, who had been a pupil of Theophrastus - like Strato - and used *horror vacui* to explain physiological processes (see Helllenistic medical epistemology §1); it also appears in the *Pneumatics* of the first-century AD technological writer Hero of Alexandria, where the influence of Strato is present though its extent is uncertain. But Strato's theory is in any case far removed from the atomist conception of discrete particles of matter moving within an otherwise empty space, and of indivisible minima of extension (see Epicureanism §§2-3; (Atomism, ancient).

4 Time, place, motion

Strato rejected Aristotle's view of time as the numbered aspect of motion, arguing that motion and time are continuous whereas number is discrete; the disagreement with Aristotle concerns the implications of 'numbered' rather than the actual continuity of time. Similarly, in defining time as quantity or measure both in motion and in rest, Strato may be differing from Aristotle more in the formulation than in the underlying doctrine (see also frs 75-9, Wehrli 1969). Strato rejected Aristotle's view of place as 'the first unmoved boundary of what surrounds', and defined it instead as the interval or extension delimited by the outermost surface of what is contained or the innermost surface of what contains it - which amounts to saying that the place of a thing is, not as for Aristotle what bounds it, but the space that it occupies (see also frs 55, 59-60, Wehrli 1969; Gottschalk 1965: 169). Unlike Aristotle, but like the Stoics, Strato held that all four elements, earth, air, fire and water, naturally move to the centre of the universe (frs 50-2, Wehrli 1969). He agreed with Aristotle that falling heavy bodies accelerate, and demonstrated this, in a notable combination of reasoning and empirical observation, from the facts that water which falls as a continuous stream breaks into separate droplets further down, and that a body dropped from a height makes an impression in the earth which it does not if dropped from a short distance (fr. 73, Wehrli 1969).

5 Soul

Strato emphasized the role of *pneuma*, 'breath' or 'spirit' (see Pneuma), in the functioning of the soul (see Psychē). Aristotle and Theophrastus had used *pneuma* to explain bodily processes, but it is debated how far they had identified specific channels through which it functioned. For Strato (frs 119-20, Wehrli 1969) soul-activities were explained by *pneuma* extending throughout the body; the 'ruling part' *(hēgemonikon)* he located (perhaps not under that name, which is Stoic) not in the chest, as in Epicureanism and Stoicism, but in the head, more precisely in the space between the eyebrows. Tertullian illustrates Strato's theory with the analogy of air in a musical pipe (Strato fr. 108, Wehrli 1969), the point in the context being that the same air can function in different ways; but he attributes the analogy also to Heraclitus and Aenesidemus. Strato may have been influenced in his theory of *pneuma* by developments in contemporary medicine and anatomy; Erasistratus investigated the function of the nerves by dissection and argued that they contained 'psychic' *pneuma* extending from the brain. But there is no

explicit evidence that Strato himself linked *pneuma* with the nerves. Sensation was actually felt, he argued, in the ruling part of the soul, rather than in the bodily extremities (frs 110-11, Wehrli 1969).

For Strato, all sensation involved thought (fr. 112, Wehrli 1969, noting what happens if our attention wanders while reading). Conversely, there is no thought not derived from sensation (fr.74, Wehrli 1969). Sextus Empiricus goes further and says that for Strato sensation and thought (*dianoia*) were identical (fr.109, Wehrli 1969), but this is an oversimplification; for Strato thought is *based on* sensation (and for Aristotle too it depends on images; *On the Soul* 432a13-14). Epiphanius' claim that for Strato (fr. 48, Wehrli 1969) every living creature was capable of thought may be as unreliable as many of his other statements. Plutarch too cites the claim that all sensation involves thought (fr. 112, Wehrli 1969) in the context of animal intelligence, though there is nothing in the report itself to suggest this. Strato's emphasis on *pneuma* as the means through which soul functions is a development, not a rejection, of Aristotle's own (arguably functionalist) view of soul as form. There is no place in Strato's theory of soul for the individual soul conceived as form separable, at least in part, from the body; but this is not the only possible interpretation of Aristotle's own position, even if many have adopted it for Platonizing, theological or anti-materialist reasons of their own. Strato brought some highly pertinent criticisms (frs 122-7, Wehrli 1969) against Plato's arguments for immortality in the *Phaedo*, pointing out for example that the fact that a soul cannot be dead does not show that the soul is everlasting.

6 Scientific method

In criticizing Plato's arguments for immortality Strato contrasted demonstrative argument with the Platonic theory of recollection (see Plato §§11, 13) to the disadvantage of the latter (frs 125-6, Wehrli 1969; fr. 14b, Gottschalk 1965). Isnardi Parente (1991) connects this with Strato's use of experimental method and with Hero's reporting that microvoid is shown to exist by 'demonstration from sense-perception' (*aisthētikēapodeixis*, Strato fr.1a p.112.13, Gottschalk 1965). Although the connection of the latter passage with Strato is not certain, this reconstruction of a Stratonian epistemology which based demonstrative argument on empiricist foundations, and which was influential for Hellenistic science (see Hellenistic medical epistemology), is plausible enough. However, the contrast often drawn in this respect between Aristotle himself and Strato depends on the questionable attribution to the former of a belief in intuition as a mode of cognition independent of the senses (see Aristotle §6).

See also: Aristotelianism, medieval

R.W. SHARPLES

References and further reading

None of Strato's own writings survives except for fragments cited by other authors.

Annas, J.E. (1992) *Hellenistic Philosophy of Mind*, Berkeley, CA: University of California Press.(Discusses Strato's doctrine of soul and *pneuma*.)

Aristotle (*c.* mid 4th century BC) *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984.(Includes the works falsely attributed to Aristotle in the ancient tradition.)

Furley, D.J. (1985) 'Strato's Theory of the Void', in D.J. Furley, *Cosmic Problems*, Cambridge: Cambridge University Press, 1989: 149-60.(A comprehensive account of the evidence for Strato's microvoid theory.)

Gatzemeier, M. (1970) *Die Naturphilosophie des Straton von Lampsakos (The Natural Philosophy of Strato of Lampsacus)*, Meisenheim an Glan: Anton Hain. (Detailed study of Strato's treatment of motion in various aspects of his philosophy; includes a full survey of earlier literature, and a listing of passages other than those in Wehrli which have been connected with Strato.)

Gottschalk, H.B. (1965) 'Strato of Lampsacus: Some Texts', *Proceedings of the Leeds Philosophical and Literary Society, Literary and Historical Section* 11 (6): 95-182.(Texts additional to those in Wehrli, not all explicitly attributed to Strato, with English commentary.)

Hackforth, R. (1955) *Plato's Phaedo*, Cambridge: Cambridge University Press. (Translation of Strato's criticisms of the *Phaedo*, pages 195-8.)

Isnardi Parente, M. (1991) *Filosofia e scienza nel pensiero ellenistico (Philosophy and Science in Hellenistic Thought)*, Naples: Morano.(On Strato's epistemology and scientific method, see pages 123-48.)

Repici, L. (1988) *La natura e l'anima: saggi su Stratone di Lampsaco (Nature and the Soul: Essays on Strato of Lampsacus)*, Torino: Tirrenia.(Detailed discussion of ancient reports of Strato's theory of soul, his physiology and his theory of nature; argues forcefully against false contrasts between Strato's views and Aristotle's.)

Sorabji, R. (1983) *Time, Creation and the Continuum*, London: Duckworth.(On Strato's theory of time, see pages

82, 377-9.)
Wehrli, F. (1969) *Die Schule des Aristoteles*, vol.5 *Straton von Lampsakos (The School of Aristotle: vol 5 Strato of Lampsacus)*, 2nd edn, Basle: Schwabe.(The standard collection of fragments and testimonia, with German commentary.)

Strauss, David Friedrich (1808-74)

The Christian faith rests upon two major beliefs: the existence of a God who created the universe, and the claim that in the historical person of Jesus of Nazareth this God in a unique way entered into world history. Like the two foci of an ellipse, these beliefs, which may be designated the metaphysical and the historical, constitute the fundamental foundation of the Christian faith. Strauss was the first major theological figure to openly challenge this foundation. The revolutionary Das Leben Jesu (Life of Jesus), which appeared in 1835, was Strauss' attempt to disprove these two fundamental beliefs from a point of view which no longer accepted the old orthodox dogmas. As an alternative explanation to a divine Christ, Strauss posited a mythological Jesus; in his lifelong search to find a substitute for theism, Strauss moved through a philosophy of nature to Hegelianism, atheism and finally Darwinism.

1 *Life of Jesus*

David Friedrich Strauss was born in the town of Ludwigsburg, a few miles north of Stuttgart, in the province of Württemberg, on 27 January 1808. He studied theology at the University of Tübingen (one of his teachers being Ferdinand Christian Baur), and also devoted considerable time to philosophy. He found Kant too involved and difficult to understand. Romanticism was the ruling sentiment of the day, and Strauss enthusiastically embraced the works of Goethe, Schiller and other Romantics of the age. But the leading philosopher of Romanticism at that time was Schelling, whose pantheistic philosophy of nature Strauss eagerly absorbed (see Schelling, F.W.J. §1). In doing so he rejected the idea of a God who was both transcendent and personal. 'I have still not been able to convince myself of the personality of God, as it is found in theism', he wrote to a friend. 'In that teaching God is a personal, self-conscious God; the consciousness, however, is nothing other than the expression of a particular life of the Spirit which develops *in time*' (1851 (Harris 1973: 11)).

It was in 1828 that Strauss first began a serious study of Hegel's *Phenomenology*. What drew him most of all to Hegel was the alternative explanation which Hegel provided for the creating source behind the universe - a God who was not the three-person Creator of the Bible, but an unconscious Spirit which in a threefold process enters into history, becomes conscious, and returns to itself as the Absolute Spirit (see Hegel, G.W.F. §8). In 1831 Strauss travelled to Berlin, hoping to attend lectures from Hegel himself. He did indeed meet the Master, but after attending two lectures was informed by Schleiermacher that Hegel had died of cholera on the previous day. Strauss decided to remain in Berlin, and it was at this time he conceived his plan for a 'Life of Jesus' in which the Gospels would be interpreted from a new philosophical viewpoint.

In his investigation of the New Testament, Strauss' primary concern was to evaluate the Gospels as *historical* documents, going behind the supernatural elements to arrive at the historical truth lying behind the sources. Here he had first to make the determining assumption that everything supernatural was *ipso facto* unhistorical and unauthentic. For he perceived that to allow even one supernatural element to remain would vitiate the philosophical framework he envisaged. Therefore, the interpretation was above all 'a-theistic', meaning that a supernatural God (if existent) was excluded from the reckoning, without there being any explicit *denial* that such a God existed (the atheistic interpretation). This a-theistic or purely historical viewpoint, later also espoused by Baur (who after 1835 became the leading Hegelian theologian), may be called the Tübingen theological perspective. Strauss was the first to apply this a-theistic viewpoint openly and consistently to the whole of the New Testament, and not just piecemeal to certain isolated passages. From this point, Strauss abandoned the supernatural element completely in favour of the philosophical principle that in history only the rational is real and the real rational.

Strauss was not the first to interpret the Biblical narratives in this way. The rationalists, pre-eminent among them H.E.G. Paulus, had sought to explain miracles without recourse to the supernatural. But what was new in Strauss was a completely different method of interpreting the historical element. The rationalists had explained each miracle by a variety of natural causes; H.S. Reimarus had accused the Biblical authors of conscious fabrication. But Strauss was the first to interpret the Gospel narratives as a whole neither by rationalism nor by fraud, but by a new explanation called the 'mythological' interpretation, in which he claimed that the Biblical writers *unconsciously* presupposed miraculous events in the life of Christ on the ground of supposed prophecies in the Old Testament. Thus, according to Strauss, the accounts of miracles in the Gospels were simply creations in the minds

of the Gospel writers, who reasoned that since such miracles and events were predicted in the Old Testament, then they must have happened, and accordingly represented them as having so happened.

These were the interpretive principles which Strauss used, and they were thus far quite independent of his Hegelian views. But when it came to explaining positively what had actually happened - for Strauss did not deny that there was a historical person called Jesus - he employed his Hegelian philosophical viewpoint. Jesus was the God-man, not in the old traditional sense of an actual unity between the divine Son of God (the second person of the Trinity) and the human body which he adopted through the virgin birth, but in the sense that the man Jesus was simply the representation of what every man actually was, the unity of the divine and human in every personality. ‘The infinite Spirit is real only when it discloses itself in finite spirits, just as the finite spirit is true only when it merges itself in the infinite. The true and real existence of the Spirit, therefore, is neither God by himself, nor man by himself, but the God-man’ ([1835-6] 1972: 777). As such everyone is a God-man, and Jesus is simply on a higher level with other men of genius, such as Napoleon, Goethe, Raphael and Mozart, with the distinction that the religious genius occupies a higher place, since the divine Spirit is present in this form of consciousness directly, and not merely indirectly.

These, then, were essentially the views which in 1835 caused an uproar in the theological circles of Germany. ‘The book stood there’, wrote the Hegelian Karl Schwarz, ‘with the hard indifference of fate; it was the final account drawn up in the criticism of the Gospel history, and the inventory read: Bankrupt! For this reason the effect of the work was enormous. An electric shock ripped through the whole of German theology’ (1869: 97). In Britain the book was translated into English by the young Mary Ann Evans, later to achieve fame as the novelist George Eliot.

Reaction was not long in following. Strauss was dismissed from his position as tutor in the Tübingen Seminary. For a time he taught classical languages in the school at Ludwigsburg. An attempt by the new democratic government at Zurich to appoint him professor of theology aroused such strong opposition throughout the Canton that he had to be pensioned off without having set foot in the city; the pension, however, secured his financial independence. Marriage to a famous singer, Agnese Schebest, in 1842, brought him only temporary happiness, a divorce being agreed five years later. In 1848, following the revolution, Strauss was appointed to the Württemberg Parliament as the liberal candidate for Ludwigsburg, but he could not agree with the radical majority and after a few months handed in his resignation.

2 Later works

For some years Strauss devoted his time to literary works, producing biographies and sketches of lesser known historical figures. A new version of the *Life of Jesus* was published in 1864, *Das Leben Jesu für das deutsche Volk bearbeitet (A New Life of Jesus)*, in which Strauss followed the same a-theistic approach, but abandoned his Hegelian reconstruction of the person of Jesus. In 1872 his final work was published, *Der alte und der neue Glaube (The Old and the New Faith)*, in which he espoused the new evolutionary views of Charles Darwin - the first theologian to do so. Thirty-seven years had passed since he had written his first *Life of Jesus* and his philosophical views had undergone a great change; no longer did he hold the Hegelian ideas he had endorsed in his youth. In the 1840s the Hegelian bubble had burst (see Hegelianism §2). Baur’s pupils, Strauss, Eduard Zeller, Albrecht Ritschl and others, had all slowly abandoned the Hegelian rhetoric. Ludwig Feuerbach (§2) drove the final nail into the Hegelian coffin, declaring that such metaphysical ideas were simply so much verbiage; finally Baur himself discarded his former Hegelian convictions. Feuerbach had convincingly demonstrated that there was no absolute, infinite Spirit; such ideas were just the projection of man’s mind into space. The secret of theology was nothing but anthropology; the knowledge of God nothing more than the knowledge of man! All Hegel’s metaphysical ponderings were in the end just grandiose illusions. What remained was nothing but man himself.

But where had the universe come from? Where had humankind come from? Strauss’ examination of the various possibilities began with Schopenhauer’s unconscious primeval Will, which evolved through consciousness and intelligence, through the plant and animal world, to humankind itself as the highest manifestation of that Will (see Schopenhauer, A. §§3-4). Eduard von Hartmann developed these ideas further, and asserted that everything must have originated from the Unconscious, for if God was conscious and aware of what he was doing when he created the world, then his creation would be an enormous and inexpiable crime. Strauss was fascinated by these solutions, and though he did not accept all of Schopenhauer’s ideas and was extremely critical of Hartmann, he saw in the evolutionary process the key for which he was looking in order to unlock the riddle of the universe. In Darwin’s

natural selection he found the confirmation he was seeking. In that direction lay the way the train of scientific thinking would go.

In espousing evolution and Darwin, Strauss was criticized by many of his supporters, but his keen sense of logic and consistency have certainly been vindicated by the twentieth-century acceptance of the evolutionary viewpoint. In adopting this viewpoint Strauss was, as always, ahead of the radical thinking of his time. The book went through six large editions in a year, a success which had never before attended any theological or philosophical work in Germany. Even Nietzsche's vituperative and rude attack in the following year (Nietzsche looked upon Strauss as a representative of a decrepit, bourgeois culture and morality) could not significantly detract from the enormous popularity the book had enjoyed (see Nietzsche, F.).

Strauss died two years later on 8 February 1874, having expressly forbidden a Christian funeral. His first *Life of Jesus* set in motion the whole 'quest' for the historical Jesus, which sought to discover the true Jesus of history stripped of his supernatural features. Strauss' criticism of the Gospel sources also provided the foundation on which Baur built up the historical criticism of the New Testament, with other scholars working on parallel lines in the Old Testament. Rudolf Bultmann's later demythologization of the New Testament, where everything supernatural is regarded as mythical, is simply a modern extension of Strauss' work (see Bultmann, R.). Strauss has no 'followers' as such today, but his a-theistic principles and views in various ways have permeated and influenced almost every area of theological scholarship.

See also: Atheism

HORTON HARRIS

List of works

Strauss, D.F. (1876-8) *Gesammelte Schriften*, Bonn: Emil Strauss, 12 vols.(The collected works of Strauss containing most of his major writings, apart from the *Life of Jesus*, the *Streitschriften* and the *Christliche Glaubenslehre* (The Doctrine of Christian Faith). A complete list is contained in H. Harris, *David Friedrich Strauss and his Theology*, below.)

Strauss, D.F. (1835-6) *Das Leben Jesu*, Tübingen: Osiander; trans. M.A. Evans (George Eliot, 1846), *Life of Jesus*, Philadelphia, PA: Fortress, 1972.(The work which made Strauss famous and initiated the 'quest' for the historical Jesus. Translation reprinted 1892, 1898 and 1972, the last with an introduction by Peter C. Hodgson.)

Strauss, D.F. (1837) *Streitschriften zur Verteidigung meiner Schrift über das Leben Jesu (Writings in Defence of my Book about the Life of Jesus)*, Tübingen: Osiander.(A section of these 'Writings' dealing with the Hegelians was translated and introduced by M.C. Massey, *In Defense of my Life of Jesus against the Hegelians*, Hamden, CT, Archon Books, 1983.)

Strauss, D.F. (1864) *Das Leben Jesu für das deutsche Volk bearbeitet*, Leipzig: Brockhaus; trans. *A New Life of Jesus*, London: Williams & Norgate, 1865.(A revision of the 1835 *Life of Jesus*, but no longer from the Hegelian perspective that Strauss had long previously abandoned.)

Strauss, D.F. (1865) *Der Christus des Glaubens und der Jesus der Geschichte. Eine Kritik des Schleiermacher'schen Lebens Jesu*, Berlin; trans. and with intro. by L.C. Keck, *The Christ of Faith and the Jesus of History*, Philadelphia, PA: Fortress, 1977.(Strauss' view of Schleiermacher's understanding of Jesus and of Schleiermacher's theology in general.)

Strauss, D.F. (1872) *Der alte und der neue Glaube*, Leipzig: Hirzel; trans. M. Blind, *The Old and the New Faith*, London: Asher, 1873.(Strauss' final major work comprising an examination of the validity of the Christian religion and the origin and purpose of the universe.)

References and further reading

Cromwell, R.S. (1974) *David Friedrich Strauss and his Place in Modern Thought*, Fairlawn, NJ: Burdick.(A well-written account of Strauss' life and historical background.)

Graf, F.W. (1982) *Kritik und Pseudo-Spekulation (Criticism and Pseudo-Speculation)*, Munich: Kaiser.(A massive account of Strauss' theology in the Hegelian period. Easily the best German work in modern times, with a full bibliography.)

Harris, H. (1973) *David Friedrich Strauss and his Theology*, Cambridge: Cambridge University Press.(This contains the most comprehensive theological assessment in English of Strauss' life and theology.)

Harris, H. (1975) *The Tübingen School*, Oxford: Oxford University Press; 2nd edn, Grand Rapids, MI: Baker Book House, 1990.(A comprehensive survey of the Tübingen School, which also includes a discussion of the

influence of Schleiermacher and Hegel.)
Lawler, E.G. (1986) *David Friedrich Strauss and his Critics*, New York: P. Lang. (Surveys early criticisms of Strauss.)
Sandberger, J. (1972) *David Friedrich Strauss als theologischer Hegelianer (David Strauss as a theological Hegelian)*, Göttingen: Vandenhoeck & Ruprecht.(Discusses Strauss' Hegelian period.)
Schwarz, K. (1869) *Zur Geschichte der neuesten Theologie (The History of Recent Philosophy)*, Leipzig, 4th edn; quotation in §1 trans. H. Harris, in Harris, 1973, 66-7. (A review of philosophy and theology in the first half of the nineteenth century from a Hegelian perspective.)

Strauss, Leo (1899-1973)

Leo Strauss was a German-Jewish émigré political philosopher and historian of political thought, who wrote some fifteen books and eighty articles on the history of political thought from Socrates to Nietzsche. Strauss was no ordinary historian of ideas; he used the history of thought as a vehicle for expressing his own ideas. In his writings, he contrasted the wisdom of ancient philosophers such as Plato and Aristotle with the foolhardiness of modern philosophers such as Hobbes and Locke. He thought that the loss of ancient wisdom was the reason for the 'crisis of the West' - an expression that was in part a reference to the barbarities of the Holocaust. He therefore sought to recover the lost wisdom. He studied the classics and was a great devotee of Plato and Aristotle. However, he developed unusual interpretations of classical texts.

Strauss was born in Kirchhain, Hessen, Germany. He studied at the Universities of Marburg and Hamburg where he came into contact with Edmund Husserl and the young Martin Heidegger. He left Germany in 1932 and eventually settled in the USA where from 1949 to 1968 he was professor of political science at the University of Chicago. He amassed a sizeable following of devoted students, who have played a significant role in US academic life and government.

According to Strauss, the fundamental issue that divides ancient and modern thinkers is the relative importance of reason and revelation in human life. Modern philosophers such as Hobbes (§§6-7) and Locke (§10), exalt reason and believe that a political order can be founded on purely rational and secular principles. But Strauss believed that this modern liberal project was doomed to failure. He thought that reason cannot provide the requisite support for moral and political life; what is needed is belief in a transcendent God who punishes the wicked and rewards the righteous. Strauss thought that ancient philosophers understood this very well and that modern philosophers were seriously misguided in thinking that rational self-interest was a sufficient ground of social life. In Strauss' view, the modern faith in reason is at the heart of the 'crisis of the West'. Reason has destroyed faith and in so doing has opened the door to barbarism. Ancient philosophers understood that reason and philosophy have a corrosive effect on the 'city', as Strauss called the state. By the same token, Strauss was committed to philosophy and had no intention of denouncing it out of hand. He therefore argued that philosophy must be kept hidden or secret, not simply to permit philosophers to avoid persecution, but for the sake of the people and the for wellbeing of the city.

In *Persecution and the Art of Writing* (1952), a paradigmatic work, Strauss argued that all the great philosophers of the past understood the dangers involved in their own philosophical activity. Accordingly, they peppered their work with riddles, hints, clues, allusions, confusing language, intentional contradictions and other forms of subterfuge designed to conceal their true thoughts. Strauss argued that every great book contained a dual teaching: an exoteric or public teaching and an esoteric or private message. The former was a salutary teaching or noble lie intended for the consumption of the many while the latter was the dangerous truth intended only for the few.

Strauss' discovery of esotericism led him to advance unusual interpretations of classic texts. For example, Strauss argued that Plato (§4) wrote dialogues in order to conceal his true thinking. And contrary to popular belief, Strauss denied that Socrates was Plato's mouthpiece. He thought that in *Republic* Thrasymachus, not Socrates, was Plato's true spokesman (Strauss 1964: 77). According to Strauss, the argument made by Socrates was simply Plato's exoteric teaching. In arguing that justice leads to happiness, Socrates was displaying his sophistic skill - that is to say, his ability to make the weaker argument appear the better. Strauss thought that the Socratic argument in favour of justice could not succeed. He was certain that Plato was wise enough to realize that Thrasymachus was right - that justice is a function of power, and that in acting justly one serves the interest of others and not oneself. Strauss surmised that Socrates must have taken Thrasymachus aside and explained to him that his views were true, but too dangerous to express publicly. In this way, Socrates managed to silence Thrasymachus without refuting him. For Strauss, the truth is a luxury meant only for the few who hunger for the reality behind the necessary myths and illusions of the city.

Strauss insists that although the truth is dark, even 'sordid', it is still the erotic object of the philosopher's quest. However, it would seem that if we accept Nietzsche's premises, we must also accept his conclusion - if the truth is dark, then we must renounce it and live according to the humanizing illusions, the life-giving myths we create for

ourselves. By accepting Nietzsche's premises and rejecting his conclusion, Strauss cultivates an elite that is more vulgar than wise (see Nietzsche, F.).

The devotion of Strauss' followers, coupled with his esotericism, has made him a figure of some controversy. But one thing is clear, Strauss' followers regard US liberalism as the embodiment of the legacy of modernity and its attendant dangers, and their aim is to rescue the USA from such modernity. Strauss taught them that this would be possible if they could win the ear of the powerful, hence their interest in government and public policy. Some find Strauss' elitism disconcerting. An elite that is radical, secretive and duplicitous, an elite that exempts itself from the moral principles it deems applicable to the rest of humanity, cannot be trusted with political power.

SHADIA B. DRURY

List of works

Strauss, L. (1952) *Persecution and the Art of Writing*, Chicago, IL: University of Chicago Press.(His best account of the esoteric/exoteric thesis.)

Strauss, L. (1953) *Natural Right and History*, Chicago, IL: University of Chicago Press.(His best account of the difference between ancient and modern thinkers.)

Strauss, L. (1959) *What is Political Philosophy?*, New York: Free Press.(Collection of well-known essays.)

Strauss, L. (1964) *The City and Man*, Chicago, IL: University of Chicago Press.(Discussion of Plato, Aristotle and Thucydides.)

Strauss, L. (1968) *On Tyranny*, Ithaca, NY: Cornell University Press.(A commentary on Xenephon's *Hiero* on *Tyrannicus*.)

References and further reading

Bloom, A. (1987) *The Closing of the American Mind*, New York: Simon & Schuster.(This best-selling book is a popularization of Strauss' thought by one of his students.)

Burnyeat, M.F. (1985) 'Sphinx Without a Secret', *New York Review of Books*, 32 (9): 30-6.(A criticism of Strauss' understanding of classical philosophy, especially Plato.)

Drury, S.B. (1988) *The Political Ideas of Leo Strauss*, New York: St Martin's Press.(Argues that Strauss' ideas are largely inspired by Nietzsche. Contains a comprehensive bibliography.)

New York Review of Books (1985) 'The Studies of Leo Strauss: An Exchange', 32 (15).(Responses from Strauss' admirers to Burnyeat.)

Strawson, Peter Frederick (1919-)

Strawson taught at the University of Oxford from 1947, becoming Waynflete Professor of Metaphysical Philosophy in 1968, and retiring in 1987. A sequence of influential books and articles established him as one of the leading philosophers in Oxford during that period. He had a crucial role in the transition there from the dominance of Austin and linguistic philosophy in the 1950s to the more liberal and metaphysical approaches in the 1960s and later. The principal topics about which he has written are the philosophy of language, metaphysics, epistemology and the history of philosophy.

Strawson became famous with 'On Referring' *(1950), in which he criticized Russell for misconstruing our ordinary use of definite descriptions. Strawson endorses the slogan 'ordinary language has no exact logic', a viewpoint which is explored in* Introduction to Logical Theory *(1952). He argues that the utility of formal logic in its application to ordinary speech does not imply that the meaning of ordinary language is captured by the semantics of standard formal systems.*

In Individuals *(1959), Strawson's most discussed work, his task is descriptive metaphysics. He attempts to describe the referentially basic subject matter of our thought. They are relatively enduring, perceptible and reidentifiable bodies. The other element in the basic framework is what Strawson calls persons, enduring entities with both material and psychological features. In* The Bounds of Sense *(1966), Strawson continued the development of his metaphysical and epistemological ideas, by combining a critical study of Kant's Critique of Pure Reason, with the defence of some transcendental claims similar to Kant's. To think of oneself as an enduring subject of experience requires that one recognize objects which are independent of oneself. So the major epistemological problem in the empiricist tradition, of building up to the external world from private experiences, cannot arise.*

Skepticism and Naturalism; Some Varieties *(1985a) studies the conflicts between fundamental opinions which are natural to us, such as that we know things, and philosophical viewpoints claiming that these opinions are mistaken. Strawson argues that scepticism about these natural views can and should be resisted. Throughout his career, Strawson has tried to describe the basic content of our thoughts and experiences, to counter scepticism about or revisions of such thoughts, to illuminate them by making analytical connections between their basic elements, as well as investigating language, our vehicle for expressing these thoughts. He has linked his explorations to the insights of philosophers of the past, while engaging in critical debate with the period's other leading philosophers, such as Austin, Quine, Davidson and Dummett.*

1 Life

Peter Frederick Strawson was born in London in 1919. He was educated at St John's College, Oxford, and, in 1948, became a Fellow at University College, Oxford, until, in 1968, he succeeded Gilbert Ryle as Waynflete Professor of Metaphysical Philosophy. He was knighted in 1977 and retired in 1987.

2 'On Referring'

The theory of reference has been central to the philosophy of language. Strawson initiated the debate about definite descriptions and has made a developing contribution to that debate and to one about proper names.

'On Referring' (1950) is one of Strawson's most discussed and influential articles. The target is Russell's theory of definite descriptions (see Russell, B.A.W. §9). According to it, a sentence of the form 'The F is G' (which I shall call 'S') is equivalent to 'there is an F, and only one F, and it is G' (which I shall call S*). Strawson argues that S is importantly different from S*. Thus, if there is no F then S* is false; whereas for S, in these circumstances, the question of its truth or falsity does not arise. Further, S* says that there only one F, whereas utterances of S do not claim that there is only one F, since it is quite possible to say 'The car won't start', despite there being lots of cars. Strawson suggests that a sentence like S has a context-independent meaning, but when used in the right context, 'The F' is used referentially, being somewhat similar to 'That F'.

The paper inspired an enormous debate. It was disputed (1) whether there are truth-value gaps, that is whether uses of S where there is no F are neither true nor false, and (2) whether Strawson's proposal fits many sentences containing 'the'. Strawson subsequently introduced the important concept of 'identifying reference'

('Identifying Reference and Truth Values' in *Logico-Linguistic Papers*). Roughly, hearers have knowledge of many objects around them. Speakers employ definite descriptions (at least sometimes), in identifying reference, when they exploit this knowledge to enable the hearer to know which object is being spoken of. According to Strawson, Russell's fundamental mistake is to deny this role to definite descriptions. Strawson also separates the issue of the validity of the notion of identifying reference from the issue of truth-value gaps, by allowing that whether there is a gap is influenced by other aspects of a speech situation (see Descriptions) (Neale 1990).

3 Logic and language

Introduction to Logical Theory (1952) is Strawson's first book. It aims, first, to explain standard formal logic. Second, it aims to give philosophical elucidations to some notions which logicians employ, crucially inconsistency. Third, and most originally, it aims to determine the relation between formal logic and ordinary language.

Strawson argues that logical appraisal applies to statements (singly or in groups) and not to sentences, because in different contexts individual sentences make different claims. He takes the notion of inconsistency as basic, one in terms of which to explain others, such as entailment. Inconsistency is itself explained by 'our own determination of the limits of application of words' (1952: 9). Strawson therefore adopts, and tries to elucidate, the view that entailments are, in some sense, linguistically grounded.

The most original aspect of the work is its discussion of logical constants and other elements of ordinary language. I shall mention two points. First, in Chapter 6, he introduces the notion of 'presupposition', to be distinguished from entailment. Roughly, S presupposes S* if the truth of S* is a necessary condition of the truth or falsity of S, whereas S entails S* if the truth of S* is a necessary condition for only the truth of S. This notion enables Strawson to claim, that 'The F is G' presupposes but does not entail 'The F exists'. Presupposition is a notion which has been utilized more extensively by grammarians than by philosophers (see Presupposition).

The second aspect is Strawson's account of the meaning of the logical constants. He argues that it is a mistake to equate the meaning of the formal logical constant with the meaning of the natural language expressions with which logicians pair them. The case most extensively argued for is the pair ' →' and 'if…then…'.

Strawson's view is that natural language expressions have a meaning because their use is governed by various context-relative and imprecise rules, some relating to entailment, others to reference in a context. The formal logician has devised a system to track various general entailments, but this system gives neither an exhaustive nor totally accurate account of the complexities of natural language (see Formal and informal logic).

4 Truth and meaning

Strawson published a series of articles about truth, most of them collected in *Logico-Linguistic Papers* (1971). The central idea is that Ramsey's redundancy theory is correct (according to which roughly 'S is true' is equivalent to 'S'), but that it can be supplemented by noting that the presence of the truth-predicate enables us to perform various distinctive speech acts. For example, we can use it to concede ('It is true that p, but…'). Initially Strawson was close to, and was taken to be, endorsing a 'performative theory' of truth. Strawson defended these ideas against Austin's version of the correspondence theory of truth. Austin endorsed a formula claiming, roughly, that a statement is true if the state of affairs with which it is demonstratively correlated (to which it refers) is of the type with which the statement is descriptively correlated (see Austin, J.L.; Truth, correspondence theory of; Truth, deflationary theories of).

Strawson offers many criticisms. Among the crucial problems he alleges are the following: (1) the formula can apply to only a limited class of descriptive sentences; it can hardly fit, for example, generalizations; (2) talk of a sentence being correlated with a state of affairs is vague; (3) even if, contrary to fact, it captures what must hold for a sentence to be true, when we say 'it is true that p' we are not *reporting* the fulfilment of that condition.

The prolonged debate about truth between Strawson and Austin was the most important discussion of truth in the period until Dummett and Davidson, in their different ways, significantly altered its direction. Strawson has contributed to both the subsequent debates about antirealism and about the truth-conditional theory of meaning, with its attendant idea of specifying the logical form of natural language sentences. About the former, Strawson

(1976) argues that antirealism is a plain distortion of our thought, for example, about other minds. It is a distortion, and it is needless, answering to no legitimate demands (see Realism and antirealism). Strawson's attitude to Davidson is more extensively presented. In 'Meaning and Truth' (1971) Strawson argues that the notion of truth is secondary to that of saying, and that saying itself involves the idea of belief expression, and concludes that truth-theoretic accounts of meaning are unsuccessful and a better elucidation of meaning would analyse it in terms of the expression of basic psychological states, along the lines sketched by H.P. Grice (see Communication and intention; Meaning and truth).

Strawson has also repeatedly criticized the views of Quine. In an article written with H.P. Grice, 'In Defence of a Dogma' (1956), it is argued that Quine's criticisms of the analytic/synthetic distinction rest on an illegitimate demand that semantic notions should be reducible to non-semantic ones, and that the agreement observed in applying the distinction indicates that it corresponds to something real. In other articles (for example, Strawson 1985) Strawson has argued that some of the categories which Quine himself applies to language presuppose the application of other semantic notions which Quine rejects, and he has also argued that we cannot engage with language in the way we do and at the same time be sceptical about the application of the notions which Quine rejects, for example, determinacy of meaning (see Analyticity; Radical translation and radical interpretation).

5 Subject and predicate

Strawson has written on a number of occasions about the distinction between subject and predicate. His constant aim has been to explain the linguistic features of the distinction in terms of the ontological differences between the items that different parts of language introduce. Both his discussion of this problem in part 2 of *Individuals* (1959) and in *Subject and Predicate in Logic and Grammar* (1974b) exemplify this. In the latter he begins by listing various suggested differences between subjects and predicates, including Geach's proposal that negation and compounding attach to predicates, and Quine's that subject expressions alone yield places for quantification, but argues that, even if they are not dubious, they cannot be regarded as fundamental, precisely because they are in need of explanation. Strawson suggests that the explanation is that a basic type of judgment, central to human thought, in which it is judged that a spatiotemporal particular possesses a general character, are expressed by subject/predicate sentences, and the role of subject expressions is to indicate the particular and one of the roles of the predicate is to indicate the general characteristic. (Predicates also have an assigning role, of representing that the general characteristic is being assigned to the particular.) Strawson suggests that the initial marks of the subject/predicate distinction can be derived from this fundamental characterization once the contrasting nature of empirical particulars and general characteristics is described. Thus, if C is a feature then lacking C is also a feature, whereas, if X is a particular object there is no such object as the other-than-X. This is why negation goes with predicates, and not with subject expressions (see Predication; Reference).

6 *Individuals*

Strawson's project in *Individuals* is 'descriptive metaphysics', as opposed, on the one side, to 'revisionary metaphysics', and on the other, to ordinary philosophical analysis. It differs from the latter by aiming to lay bare the fundamental and most general features of our conceptual scheme, from the former by aiming only to describe the structure of our thought and not revise it. In *Individuals*, and in more recent writings, Strawson has opposed attempts to revise the fundamentals of our conceptual scheme, and part of the point of the division is to show that fundamental investigations need not be motivated by a revisionary purpose.

Strawson starts with the idea that a hearer can know of which object a speaker is speaking. He talks of the hearer *identifying* the particular, and of the speaker making an identifying reference (see §2 above). It may be that some types of particulars are identifiable only on the basis of prior identification of another type of particular. The latter are then described as more basic in our scheme of thought. His picture is that identification by demonstration is possible for sensibly present particulars, and that other particulars can be identified by their relation to such sensibly present items. But we can have a grasp of this spatiotemporal framework only if we can *re*identify particulars over time. An epistemological conclusion follows; no one can think that all our criteria for reidentification are mistaken or dubious, since this claim undermines the intelligibility of our thought of the framework in terms of which the doubt is expressed. Strawson argues that within this framework bodies are, in the sense explained earlier, the basic particulars. They are observable, enduring, and form a relatively stable framework. They are to be contrasted with what Strawson calls 'private particulars', such as P's pain, which can

be identified only via their possessors; theoretical unobservables; and events, some of which are perceptually observable but which, as a type of thing, do not form a sufficiently stable framework. Finally, the reidentifiable entities necessary for our grasp of the spatiotemporal framework are bodies, and reidentifying events rests on reidentifying the bodies involved in them.

This complex leading argument in *Individuals* has inspired considerable debate, both about the truth of some of the claims, for example, whether events are secondary to bodies, and whether the framework of events is so irregular, and about the status of Strawson's claims, for example, how far the argument rests on contingencies of the human environment and how far is it supposed to be necessary (see Campbell 1994).

In Chapter 2, Strawson considers this last question by attempting to find a place for the concepts of an unperceived and reidentifiable particular, and of oneself, in the thought of an imagined creature whose experiences are restricted to a world of sound. Sounds are not intrinsically spatial, and so considering such a world illuminates the degree to which space is fundamental to our thought. Strawson argues that some ground is needed to distinguish between hearing the same sort of sound on different occasions and hearing the same individual sound. He employs a master-sound-sequence, which is constantly heard and which varies in regular ways, to substitute for spatial location, and thereby grounds a distinction analogous to the one required. It is conceded to be incomplete, and, moreover, it is problematic that there is a ground in this world for the experiencer to think of itself, for, as Strawson puts it, a 'non-solipsistic consciousness'. The purpose of this brilliant chapter is to promote understanding of the character of our own scheme, rather than to make the sound world genuinely intelligible (see Evans 1980).

7 Persons

Strawson begins Chapter 3 of *Individuals* with two questions. Why do we ascribe states of consciousness to anything? Why do we ascribe them to the same thing to which we ascribe physical predicates? Two theories would not accept this characterization of our concepts: according to the no-ownership theory consciousness is not really ascribed to any object at all, and according to Cartesian dualism we ascribe the two characteristics to two separate entities: egos and bodies (see Dualism). However, the no-ownership account is incoherent because it needs to explain how conscious states are associated with bodies, and it can do that only by picking out a range of experiences as being someone's. Against dualism Strawson brings a fundamental principle: it is a necessary condition of ascribing states of consciousness to oneself that one does, or is prepared to, ascribe them to others. To be available for identification, other subjects must have non-mental, that is physical, characteristics. Another subject cannot be picked out as the ego standing in that same relation to that body as I stand to mine, because this makes reference to others secondary to self-reference, contrary to Strawson's fundamental principle. Further, by what right is it assumed there is only one such ego per body? Strawson also stresses in 'Mind, Body and Soul' (1974a) that it is hard to understand the notion of non-spatial particulars; the way that we have of understanding the possibility of numerically distinct but qualitatively identical particulars of a spatial kind, namely that they occupy different *places* in a spatial framework, is not available for egos.

Rejecting these characterizations, Strawson suggests that the concept of a person is 'a primitive concept' (1959: 101), by which he means that it is a primitive practice to recognize oneself and others as objects with both physical and psychological characteristics. From this he draws the epistemological consequence that at least for some psychological properties the ways of telling 'that they apply to others must constitute logically adequate kinds of criteria' (1959: 105). Strawson is suggesting that there cannot be a genuine problem of other minds. Nothing so far says how the material and the psychological relate. In Chapter 3 of *Skepticism and Naturalism* (1985a) Strawson claims that it is more satisfactory to regard psychological and neural events as causally related, rather than as identical.

Strawson's claims have been widely discussed. In particular, his fundamental principle and his epistemological claim about logically adequate criteria have been questioned (see Ayer 1964). Whatever the resolution of these questions, Strawson has undoubtedly articulated some fundamental aspects of our thought about ourselves and others (see Persons).

8 *The Bounds of Sense*

The Bounds of Sense (1966) is an exposition and criticism of Kant's *Critique of Pure Reason* and a defence of certain claims about the necessary structure of our experience, which resemble those advanced by Kant.

Strawson rejects Kant's transcendental idealism. He provides a careful interpretation of it but argues that its incoherence is particularly clear in relation to the self, the location of which oscillates between the phenomenal world and the noumenal world. Strawson is, however, highly sympathetic to many others among Kant's positive and negative claims. In particular, he suggests that from Kant's arguments in the Transcendental Deduction a strong case can be reconstructed for saying that it is unintelligible to suppose that there is a self-conscious subject unless some of the experiences of the subject are of, and are taken to be of, independent objects other than itself. From Kant's arguments in the Analogies support can also be extracted for the claim that the subject's experience must allow the application of concepts of persisting and re-encounterable objects in a spatial framework, and of causal properties and powers.

Strawson defends the first necessity, that unity of experiences in a single subject requires experience of objects, by arguing that there is no point in ascribing a succession of experiences to a single subject unless for some of its experiences a distinction exists between the way, in that experience, it seems to the subject and the way the experienced item actually is. Then the notion of experience, as something distinct from but related to reality, has a point, and with it, the notion of the possessor of the experience, the subject. If, *per impossibile*, we had a subject none of whose experiences permit this distinction to be applied, then we have lost the idea that there is experience at all, and we have lost the idea of a subject. If Strawson's modified Kantian claims are correct, there is no genuine problem corresponding to that which gripped the empiricists, of showing that there is an external world on the basis of the character of private experiences. Indeed, part of the point of Strawson's enquiry is to outlaw that problem.

Strawson's brilliant treatment of Kant has been widely discussed. Two areas of debate may be singled out. One is that the force and soundness of transcendental arguments, as a style of philosophical argument, is disputable (see §9 below). The other is whether Strawson, despite his sympathetic approach, has provided the best account of the significance of Kant's idealism (see Transcendental arguments) (Walker 1978).

9 Epistemology

In *Individuals* and *The Bounds of Sense* Strawson developed a response to different types of scepticism by defending transcendental arguments. These are of two sorts. In *Individuals*, especially Chapter 3, which deals with other minds, he claims that in order to employ psychological concepts at all - which, of course, the sceptics themselves do - one must acknowledge the adequacy of our basic ways of telling that they apply to others. In *The Bounds of Sense* the transcendental claim is that the conceptual applications which the sceptic is prepared to make (roughly about themselves and their experience) require the application of the concepts about the application of which they are sceptical.

Against such arguments it has been objected that we cannot be confident that the intelligibility of the concepts requires what is claimed. There is the permanent fear of an overlooked possibility. It has also been disputed what the force of the final requirement is. Is it merely that the concepts are applied or is it that they are veridically applied (see Stroud 1968)? Finally, what other things, if any, need to be said in response to the sceptic, even if the transcendental claims are accepted?

In *Skepticism and Naturalism*, Strawson, developing themes in Hume and Wittgenstein, argues that it is simply natural for us to believe in bodies, and this psychological fact renders any chance of sceptical arguments generating real doubt impossible, which means that such arguments need not be taken seriously. As Strawson says, ‘Arguments on both sides are idle’ (1985a: 29). Strawson's new viewpoint seems to be sustained by doubts, of the kind mentioned, about the anti-sceptical value of transcendental arguments. The proposal has been much debated, the chief issue being whether the idleness of the sceptical arguments is itself ground for ignoring them (see Scepticism).

10 Philosophy

There are certain general and recurring themes in his philosophical writings (particularly his later writings) which need to be described. Although Strawson does not share Wittgenstein's conception of philosophy, regarding

philosophy as a rational and theoretical (and not primarily therapeutic) activity, he has evident sympathy with the conceptual conservatism implicit in the slogan that philosophy 'leaves everything as it is'. There is no ground in philosophy to criticize our fundamental opinions. An example is Strawson's treatment of the supposed conflict between moral responsibility and determinism in 'Freedom and Resentment' (1962). His basic claim is that our view of ourselves as responsible is an ingredient in a network of emotional responses which cannot be abandoned and which do not rest on any general conviction that might be refuted. Philosophy is, in this way, full of false forced choices. They have their sources in multifarious misconceptions. One misconception is scientism, the idea that science is the arbiter of all empirical truth. Another is a tendency to misinterpret common convictions: for instance, our tendency to refer to abstract objects seems problematic because all objects are regarded as necessarily material or quasi-material. Another is the failure to appreciate that there can be autonomous conceptual levels, evidenced in philosophical demands that all truths be reducible to some favoured base, say, the language of science, or a neutral sense-datum language. Strawson's idea here is that either the favoured language is itself a development of, and hence secondary to, the prior language (as in the case of science), or is mythical (sense-datum language). A central task of philosophy is to describe, and in the best way we can, explain the irrevocable fundamentals of human thought. The explanation consists in linking elements in the structure of concepts and experience, for example by displaying how our thought of ourselves links with our conception of the nature of experience. This structure of philosophical convictions has been enormously influential, for example, on and through the work of Gareth Evans, David Wiggins and John McDowell.

PAUL F. SNOWDON

List of works

Strawson, P.F. (1950) 'On Referring', *Mind* 59: 320-44; repr. in *Logico-Linguistic Papers*, London: Methuen, 1971.(Strawson's first, and very famous, discussion of definite descriptions and Russell's treatment of them.)

Strawson, P.F. (1952) *Introduction to Logical Theory*, London: Methuen.(Strawson's account of the fundamental notions of formal logic, and of the relation between formal logic and natural language. Chapter 9 is Strawson's influential attempt to 'dissolve' the problem of induction.)

Strawson, P.F. and Grice, H.P. (1956) 'In Defence of a Dogma', *Philosophical Review* 65: 141-58.(A jointly written reply to Quine's criticisms of the analytic-synthetic distinction.)

Strawson, P.F. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen.(Strawson's highly original essay in descriptive metaphysics. Part 1 tries to determine the role of bodies and of persons in our thought, and Part 2 links the subject-prediacte distinction to these roles.)

Strawson, P.F. (1966) *The Bounds of Sense*, London: Methuen.(Strawson's magisterial study of Kant's first Critique.)

Strawson, P.F. (1971) *Logico-Linguistic Papers*, London: Methuen.(Contains, among others, 'Identifying Reference and Truth Values' (1964) and 'Meaning and Truth' (1969).)

Strawson, P.F. (1974a) *Freedom and Resentment and Other Essays*, London: Methuen. (Contains 'Mind, Body and Soul' (1966) and 'Freedom and Resentment' (1962).)

Strawson, P.F. (1974b) *Subject and Predicate in Logic and Grammar*, London: Methuen.(A further discussion of the subject-predicate distinction, and of other aspects of natural language.)

Strawson, P.F. (1976) 'Scruton and Wright on Anti-Realism etc.', in *Proceedings of the Aristotelian Society* 77: 15-21.(A brief presentation of Strawson's views on antirealism.)

Strawson, P.F. (1979) 'Perception and Its Objects', *Perception and Identity*, ed. J. MacDonald, London: Macmillan.(A characteristic treatment of philosophical problems to do with perception.)

Strawson, P.F. (1985a) *Skepticism and Naturalism: Some Varieties*, London: Methuen.(Strawson's widely discussed study and rejection of philosophical revisions of ordinary thought.)

Strawson, P.F. (1985b) *Analyse et Metaphysique*, Paris: Vrin.(An account, in French, of Strawson's approach to philosophy.)

Strawson, P.F. (1992) *Analysis and Metaphysics*, Oxford: Oxford University Press.(Strawson's description of philosophy, overlapping with (1985b), but containing further topics.)

References and further reading

Ayer, A.J. (1964) *The Concept of a Person*, London: Macmillan.(Contains Ayer's famous and readable discussion of Strawson on persons.)

Campbell, J. (1994) *Past, Space and Self*, Cambridge, MA; MIT Press.(The first (difficult) chapter links with the

first chapter of *Individuals*.)
Evans, G. (1980) ‘Things Without the Mind’, in Z. Van Straaten (ed.) Philosophical Subjects, Oxford: Clarendon Press.(A brilliant critical discussion of the ‘sound worlds’ of *Individuals*, Chapter 2.)
Neale, S. (1990) *Descriptions*, Cambridge, MA: MIT Press.(A difficult discussion of descriptions, including Strawson’s views.)
Sen, P.B. and Verma, R.R. (eds) (1995) *The Philosophy of P.F. Strawson*, New Delhi: Indian Council of Philosophical Research.(A collection of articles on Strawson by, among others, Putnam, Dummett, Platts and Cassam, with replies by Strawson.)
Stroud, B. (1968) ‘Transcendental Arguments’, *Journal of Philosophy* 65: 241-56. (A critical examination of the claims made for transcendental arguments.)
Van Straaten, Z. (ed.) (1980) *Philosophical Subjects*, Oxford: Clarendon Press.(A book containing excellent but difficult articles on Strawson by Evans, McDowell, Ishiguro, Wiggins and others, with replies by Strawson.)
Walker, R.C.S. (1978) *Kant*, London: Routledge.(A readable book on Kant which discusses some of Strawson’s interpretations and arguments).
Special journal issue (1981) *Philosophia* 10: 141-328.(A collection of articles on many of Strawson’s views, with a long reply by Strawson himself.)

Structuralism

The term 'structuralism' can be applied to any analysis that emphasizes structures and relations, but it usually designates a twentieth-century European (especially French) school of thought that applies the methods of structural linguistics to the study of social and cultural phenomena. Starting from the insight that social and cultural phenomena are not physical objects and events but objects and events with meaning, and that their signification must therefore be a focus of analysis, structuralists reject causal analysis and any attempt to explain social and cultural phenomena one-by-one. Rather, they focus on the internal structure of cultural objects and, more importantly, the underlying structures that make them possible. To investigate neckties, for instance, structuralism would attempt to reconstruct (1) the internal structure of neckties (the oppositions - wide/narrow, loud/subdued - that enable different sorts of neckties to bear different meanings for members of a culture) and (2) the underlying 'vestimentary' structures or system of a given culture (how do neckties relate to other items of clothing and the wearing of neckties to other socially-coded actions).

Ferdinand de Saussure, the founder of structural linguistics, insists that to study language, analysts must describe a linguistic system, which consists of structures, not substance. The physical sound of a word or sign is irrelevant to its linguistic function: what counts are the relations, the contrasts, that differentiate signs. Thus in Morse code a beginner's dot may be longer than an expert's dash: the structural relation, the distinction, between dot and dash is what matters.

For structuralism, the crucial point is that the object of analysis is not the corpus of utterances linguists might collect, that which Saussure identifies as parole *(speech), but the underlying system (*la langue*), a set of formal elements defined in relation to each other and which can be variously combined to form sentences. Arguing that the analysis of systems of relation is the appropriate way to study human phenomena, that our world consists not of things but of relations, structuralists often claim to provide a new paradigm for the human sciences. In France, structuralism displaced existentialism in the 1960s as a public philosophical movement. Philosophically, proponents of structuralism have been concerned to distinguish it from phenomenology.*

1 Development and domain

Structural linguistics was developed by the Swiss linguist Ferdinand de Saussure in lectures published as the *Cours de linguistique générale* (1916), but the term 'structuralism' was coined by the Russian linguist and literary theorist Roman Jakobson in 1929. Jakobson was a member of the Prague Linguistic Circle which, between 1926 and 1938, drew upon Saussurian linguistics, Russian Formalist literary criticism, Husserl's phenomenology and Gestalt psychology to develop Prague Structuralism. Structuralism was opposed to atomism and was presented as a general method for the humanities and social sciences, although most of the circle's work focused on language, literature and art. A literary work is a structural whole; to study it is to analyse its internal relations (it is a hierarchically ordered system of structures of different kinds) and its place in the systems of a language and a culture, especially the system of literary genres and conventions.

Jakobson himself provided a link between Prague Structuralism and the structuralism which developed in post-Second World War France, first in anthropology (Claude Lévi-Strauss), then in literary and cultural studies (Jakobson, Roland Barthes, Algirdas Greimas, Gérard Genette), psychoanalysis (Jacques Lacan), intellectual history (Michel Foucault) and political theory (Louis Althusser). Although these thinkers never formed a school as such, it was under the heading of 'structuralism' that their work was exported and read in the UK, the USA and elsewhere in the late 1960s and 1970s, where it was influential first in anthropology and literary studies, and later in historical and cultural analysis.

In anthropology, structuralism insists on the reconstruction of underlying systems (of kinship, of totemic and mythic thought) through which a culture orders the world. Within the field of the history of thought it seeks to identify the system of possibilities or formation rules that permit the emergence, in a particular era, of certain disciplines and theoretical objects. In literary studies it promotes a poetics interested in the conventions that make literary works possible, and it seeks not to produce new interpretations of works but to understand how they can have the meanings and effects that they do. In cultural studies generally it encourages attempts to make explicit the rules and signifying procedures that govern fashion or other cultural practices.

According to structuralist theory, the identity of structuralism should come from contrasts within the system of modern thought, from the differences shared by a range of thinkers, rather than from a historical filiation. In fact, the term 'structuralism' is generally used to designate work that marks its debts to structural linguistics and deploys a Saussurian vocabulary, including *sign*, *signifier* and *signified*, *syntagmatic* and *paradigmatic*. There are many writings, from Aristotle to Chomsky, that share the structuralist propensity to analyse objects as the products of a combination of structural elements within a system, but if they do not display a Saussurian ancestry, they are often not deemed structuralist, whatever their affinities with the writings so designated.

Once structuralism came to be defined as a movement or theoretical stance based on a set of principles, some theorists began to distance themselves from it. Readers who could see that works by some 'structuralists' did not fit certain accounts of structuralism began to regard some of these thinkers (particularly Barthes, Lacan, and Foucault) as post-structuralists, going beyond a structuralism defined as a project aiming at mastery of objective structures (see Post-structuralism). In fact, many of the views associated with post-structuralism - such as the difficulty for any metalanguage to escape entanglement in the phenomena it purports to describe, the possibility for texts to create meaning by violating the conventions that structural analysis seeks to delineate, or the inappropriateness of positing a complete system because systems are always changing - are manifest even in the early work of such thinkers as Barthes, Foucault and Lacan who were at the time regarded as structuralists. Post-structuralism involves not so much the demonstration of the inadequacies or errors of structuralism as a turning away from the project of working out what makes cultural phenomena possible and intelligible, and a return to interpreting and narrating (sometimes called 'historicizing').

It is difficult to distinguish, in a principled way, structuralism from semiotics, the science of signs, which traces its lineage to Saussure and Charles Sanders Peirce. However semiotics is an international movement that has sought to adopt a scientific model and to cultivate links with, for example, zoology, while mostly eschewing the philosophical speculation and cultural critique that has marked structuralism in its French and related versions (see Semiotics).

2 Shared principles

Structuralism takes many different forms, but there are some general principles that could be identified as structuralist. Most of these are best stated differentially, as contrasts. Structuralism has defined itself against historicism, atomism, mechanism, behaviourism, psychologism, and humanism.

Structuralism is not hermeneutics (see Hermeneutics). Structural analysis seeks to understand the conditions of meaning, and in that sense it takes the cultural meaning of objects and events as a point of departure, as what requires explanation. Hermeneutics characteristically seeks to discover the meanings of a text or cultural phenomenon. Linguistic analysis does not try to tell us what sentences mean but to explain how these sequences are constructed and how they can have the meaning they do for speakers of a language. The linguistic model enjoins a concentration on poetics rather than interpretation (Culler 1975).

Structuralism is not phenomenology (see Phenomenological movement). Although the intuitions of subjects about the meaning or at least the well-formedness or deviance of cultural phenomena are often its point of departure, structuralism seeks explanation at the level not of structures of consciousness but of unconscious infrastructures, systems of relation that operate through subjects and work to constitute subjects but are not necessarily accessible to them. In *Le Visible et l'invisible* (1964), Maurice Merleau-Ponty imagined the possibility of linking a phenomenological critique of the distinction between subject and object (both subject and object are a posteriori constructs, notions derived from a unitary experience rather than experiential givens) to a structuralist account of the differential nature of meaning. Most theorists, however, have insisted on the fundamental difference in focus and approach between phenomenology and structuralism (Ricoeur 1969).

Cultural phenomena have a relational identity. They are defined by their differences from one another, not by any essential features. This basic principle of Saussurian linguistics has wide applicability.

The object of analysis is not the phenomena themselves but the underlying system which makes them possible or intelligible. Analysts cannot examine all the texts, artefacts or acts related to a given system, for the system, like a language, may have the capacity to produce an infinite set of objects or events. The goal is to identify the set of

contrastive elements which combine, according to rules to be discovered, to produce the cultural forms.

Structure, which is not form alone but the formal organization of content, plays a determining role in human affairs. One writer speaks of 'superstructuralism' to indicate that, in contrast with Marxism which treats language and culture as a superstructure determined by a material base, structuralism and post-structuralism share the conviction that the structures of language and meaning generally are not direct reflections of economic relations but themselves determine the parameters within which human beings live and act (Harland 1987).

The priority of synchronic over diachronic analysis. Saussure distinguishes the study of language as a system (synchronic analysis) from the study of changes in languages over time (diachronic) and argues that describing systems is a condition of understanding changes from one state or system to another. Generally, structuralism has maintained that culture should be studied as a series of synchronic systems, that much 'historical' analysis should be seen as synchronic analysis of moments in the past, and that the analysis of change requires prior understanding of the systems that change.

Structuralism does not narrate. In place of explanation which proceeds narratively, by linking a phenomenon to origins and ends, structural analysis explores conditions of possibility. The critique of structuralism known as post-structuralism often entails, in effect, a return to narrative as a mode of knowledge, but with no confidence in its authority, perhaps because the principled structuralist rejection of narrative explanation has been ignored rather than refuted.

For structuralism, the subject or self is a construct that needs to be explained, not a principle of explanation. It is not that the subject 'man' and human beings do not exist but that what counts as man or what counts as the subject depends upon a cultural system that requires analysis.

Some structuralists (for example, Lévi-Strauss and Lacan) claim to analyse the 'laws of thought' but most are agnostic on the question of whether the categories and rules of one system might be universals. Unlike transformational-generative grammar, which posits a rich universal grammar (a human language faculty that strongly constrains the properties of possible human languages), structural linguistics assumes that one language (linguistic system) may differ radically from the next. Nevertheless, most structural analyses treat meaning as the product of hierarchically-ordered sets of binary oppositions and thus, in effect, posit some universal principles of signification (see Chomsky, N.).

See also: Structuralism in linguistics; Structuralism in literary theory; Structuralism in social science

JONATHAN CULLER

References and further reading

Caws, P. (1988) *Structuralism: The Art of the Intelligible*, Atlantic Highlands, NJ: Humanities Press.(An account of the foundations of structuralism, with specific reflection on structuralism as philosophy. Clear and disinterested.)

Culler, J. (1975) *Structuralist Poetics: Structuralism, Linguistics, and the Study of Literature*, London: Routledge. (Discussion of the linguistic model and its applicability to literary and other domains. Presumes no prior knowledge of linguistics or philosophy.)

Harland, R. (1987) *Superstructuralism: The Philosophy of Structuralism and Post-Structuralism*, London: Methuen.(A broad and lively introductory survey.)

Merleau-Ponty, M. (1964) *Le Visible et l'invisible*, ed. C. Lefort, Paris: Gallimard; *The Visible and the Invisible*, trans. A. Lingis, Evanston, IL: Northwestern University Press, 1968.(Unfinished philosophical reflections exploring the possibility of deploying a structuralist perspective within phenomenology.)

Ricoeur, P. (1969) 'Structure et hermeneutique', in *Le Conflit des interpretations*, Paris: Seuil; *The Conflict of Interpretations*, trans. D. Ihde, Evanston, IL: Northwestern University Press, 1974.(One of several essays in this volume which seeks to distinguish structuralism from other approaches.)

Saussure, F. de (1916) *Cours de linguistique générale*, ed. T. de Mauro, Paris: Payot, 1973; *Course in General Linguistics*, trans. R. Harris, London: Duckworth, 1983.(The foundations of structural linguistics, structuralism, and semiotics.)

Wahl, F., Sperber, D., Ducrot, O., Todorov, T. and Safouan, M. (1968) *Qu'est-ce que le structuralisme? (What is structuralism?)*, Paris: Seuil.(Essays on structuralism in anthropology (Sperber), linguistics (Ducrot), literary criticism (Todorov), philosophy (Wahl) and psychoanalysis (Safouan), produced at the height of the

movement.)

Structuralism in linguistics

The term structural linguistics can be used to refer to two movements which developed independently of each other. The first is European and can be characterized as post-Saussurean, since Saussure is generally regarded as its inspiration. The central claim of this movement is that terms of a language of all kinds (sounds, words, meanings) present themselves in Saussure's phrase 'as a system', and can only be identified by describing their relations to other terms of the same language; one cannot first identify the terms of a language and then ask which system they belong to. Moreover, because a language is a system of signs, one cannot identify expression-elements (sounds, words) independently of the content-elements (meanings), so that a study of language cannot be divorced from one of meaning. The second movement is an American one, which developed from the work of Leonard Bloomfield and dominated American linguistics in the 1940s and 1950s. It attached great importance to methodological rigour and, influenced by behaviourist psychology, was hostile to mentalism (any theory which posits an independent category of mental events and processes). As a result, unlike the first movement, it excluded the study of meaning from that of grammar, and tried to develop a methodology to describe any corpus in terms of the distribution of its expression-elements relative to each other. Whereas the first movement provided a model for structuralist thought in general, and had a significant impact on such thinkers as Barthes, Lacan and Lévi-Strauss, the second made a major contribution to the development of formal models of language, however inadequate they may seem now in the light of Chomsky's criticisms.

1 The concept of structural linguistics

Though a wide range of linguistic studies raise issues which are in some sense structural, we shall focus on two uses of the term *structural linguistics* which refer to separate movements, each of which advanced strong hypotheses about the nature of language and the methodology of linguistics. However, though each of the movements attaches great importance to structural analysis, their reasons for doing so differ profoundly; hence, the need to distinguish them.

The origin of the first movement is to be found in the work of Ferdinand de Saussure, the fundamental contention of which is that, since linguistic items of all kinds (sounds, words, meanings) present themselves as a system, they can only be identified in terms of the relations in which they stand to other elements within the system (Lyons 1973: 6). An acceptance of this thesis, together with one, sometimes qualified, of the following three Saussurean theses, defines post-Saussurean structuralism: it is necessary to distinguish *la langue* from *la parole*, and a synchronic study from a diachronic one; the fundamental linguistic study is a synchronic one of *la langue*; and, finally, a study of expression-elements (sounds and words) cannot be divorced from one of meanings. Of course, within this framework much is underdetermined, leaving scope for different emphases and developments, two of which are considered in §§2 and 3 below.

The origins of the second movement are to be found in the work of the American linguist Leonard Bloomfield, and, in particular, in his book *Language* (1933). Strongly influenced by behaviourist psychology (see Behaviourism, methodological and scientific), Bloomfield proposed an analysis of a speech act in terms of the notions of a *stimulus* and a *response*. In response to a feeling of hunger (a stimulus) one thing I might do is reach for an apple; but if I have mastered a language, another thing that I can do is to utter 'Pass me the apple', which is itself not only a response to my hunger but also a stimulus designed to get you to respond by passing the apple. So language increases the range of responses enormously by bridging the gap between different nervous systems. A problem with this analysis, as Bloomfield concedes, is that in most cases we seem unable to identify independently either the stimulus or the response. Since he defines the meaning of an utterance as a function of the situation in which it is uttered and the response called forth, it is hardly surprising that he concludes that the study of meaning largely evades our grasp. What can be studied nevertheless are the utterances themselves but, to avoid a lapse into a discredited mentalism, or an appeal to a theory which is not available to us, sound methodology requires that the study of language does not appeal to meaning. The study of language becomes, therefore, a study of a set of utterances, or of a corpus - indeed, Bloomfield once defined a language as the set of utterances uttered by its speakers. Such a study concentrates on the formal relations between constituents of utterances. It argues that for each constituent there is a limited range of environments in which it can occur; for instance, 'cat' can occur in the environment 'The jumped', but not in 'John Jill'. So if an item's distribution is defined as the ranges of

environments in which it can occur nondeviantly, there seem to be promising ways of defining its type in terms of its distribution, and of making generalizations about the relations between types. More and more sophisticated approaches to the analysis of a corpus in terms of the distributional relations holding between its constituents were developed in America in the 1940s and 1950s, culminating in the work of Z.S. Harris which will be discussed in §4. So three main features of post-Bloomfieldian structuralism are: it was corpus based; an appeal to meaning was rigorously excluded from the methodology adopted; and the object of analysis was the identification of the constituents of the utterances in a corpus, and the construction of a compact description of the distributional relations holding between them. Paradoxically, whilst it is clear why both movements should be regarded as structuralist, each from the point of view of the other is radically misconceived. Arguably the first movement is the more interesting from a philosophical point of view because of its impact on structuralist thought in general; but the second movement undoubtedly developed much more detailed formal models than the first.

2 The Prague School and functionalism

The central thesis of post-Saussurean structuralism is that the terms of a language present themselves as a system; but precisely what kind of a system do they belong to? A group of linguists whose work was strongly associated with Prague and who became known as the Prague School developed a distinctive answer to this question. For them a language is a dynamic system with many subsystems which interact with each other; one of these concerns the structure of its functional sounds (phonology), another the structure of its words (morphology), another the structure of its sentences (syntax), and so on. Each of these has a stable core, surrounded by a less stable periphery, and is striving for a state of equilibrium which it never quite attains, since a change which stabilizes one part of the system can adversely affect another. The Russian linguist Jakobson, who played a major part in the development of this conception, likened the attempt to restore equilibrium to the moves made in a chess game after the loss of a piece, which may strengthen one part of a defence at the expense of another. The Prague School's dynamic conception of a language as a system of subsystems striving to reach equilibrium led it in turn to adopt a different conception of diachrony from that held by Saussure, for all that it agreed with him that a synchronic perspective is logically prior. Since a language is a dynamic system, the School argued that the force for change in many cases is a state of the system itself, whereas Saussure had argued that change is brought about by factors external to the system which never affect it directly.

An important area in which the School made a major contribution to our understanding of what the system involved was phonology, the study of the functional sound units of a language. It seems clear that all languages utilize a relatively small number of such units (called phonemes), so that 'cat', for instance, contains /c/, /a/ and /t/; but how are they to be characterized, and in what way do they form a system? Saussure was prevented from answering this question in part because he assumed that phonemes were not complex, and that oppositions arise from differences. However, Trubetzkoy and Jakobson questioned both of these assumptions. They argued that phonemes can be analysed in terms of so called distinctive features, which are the ways in which phonemes belonging to the same language are opposed to each other, and that the differences between them arise from these oppositions. For instance, one way in which /m/ differs from /n/ is that the lips are closed when pronouncing it, so that there is a feature (labial) present in one case which is absent in the other. Features of this kind the School called 'distinctive features', claiming that each phoneme can be described in terms of the presence or absence of a small number of such features. Moreover, since what is being described is a set of ways in which the phonemes of a language can be opposed to each other, differences rest on oppositions. This theory was to have a major impact on structuralist thinkers; for example, Lévi-Strauss asked whether there are other kinds of oppositions which give rise to different kinds of structures in domains such as kinship and myth (see Structuralism in social science).

Another novel feature of the Prague School was its attempt to combine a functionalist conception of language with its structuralist conception. The central idea is that a study of a language cannot be divorced from its uses: 'language is not a self-contained whole, hermetically separated from the extra-lingual reality, but in fact its main function is to react to and to refer to this reality' (Vahek 1966: 7). Just as one cannot understand chess fully unless one understands what it is players are trying to achieve, so one cannot understand a language if one does not understand its uses. Moreover, various aspects of use have specific linguistic means of expression; for example, it is natural if one wants someone to do something to use an imperative, if one wants information to use an interrogative, and so forth. Whether functionalism is ultimately consistent with the strong form of the principle of the arbitrariness of the sign embraced by Saussure is questionable. However, by recognizing the diverse uses of

language, the Prague School greatly extended the range of domains open to structural analyses. For instance, if an utterance may have an expressive function as well as a descriptive one, may we not, as a limiting case, have a text which is purely expressive but which can nevertheless be analysed structurally - a poem for instance?

3 The Copenhagen School and glossematics

Arguably, the Prague School's theory of oppositions, which positively characterizes one item in an opposition, is incompatible with the Saussurean theses that in language there are only differences without positive terms, and that language is a form not a substance. The work of the Danish linguist Hjelmslev, which takes these theses as its cornerstone, is in this respect a more faithful development of the Saussurean conception of a system.

The fact that a given domain is structured in different ways by different languages shows, Hjelmslev argues, that a language is a form not a substance, with its own autonomous principles of organization. In the domain of colour, for instance, we find that 'Welsh lacks the English boundary between *blue* and *grey*, and likewise the English boundary between *grey* and *brown*' (Hjelmslev 1943: 53); other examples of this phenomenon are systems of tense and pronouns. But how can an autonomous system be described without characterizing some elements positively, and so appealing to something external to it? This can be done, Hjelmslev argues, by characterizing the elements *only* in terms of the relations they can stand in to other items, so that it matters not at all what the elements are, provided that *they are different from each other*.

The resulting description is a linguistic algebra which Hjelmslev called a *linguistic schema*; whilst the result of interpreting a schema, that is of applying it to a domain, he called a *linguistic usage*. The difference between schema and usage corresponds to a more and a less abstract conception of *la langue*; and for Hjelmslev the primary object of linguistics is the schema. The proper way of studying this is to develop a deductive system and to define its entities relationally. For example, an item of a given kind *A* is said to presuppose an item *B* in a context *C* if *A* cannot occur without *B*, but *B* can occur without *A*; for instance, an indefinite article requires a count-noun in a noun-phrase, but not vice versa. Another relation distinguished by Hjelmslev is that of interdependence, that is, of mutual presupposition, which is a relation holding, for instance, between vowels and consonants, or mood and tense. Finally, there is a relation called 'solidarity' which obtains when *A*, though compatible with *B*, can occur without it, and vice versa; for instance, this relationship holds between *accusative* and *plural* within a case system. By appeal to such relationships one can, Hjelmslev argued, characterize a linguistic-schema in purely formal terms without recourse to phonetic, phenomenological or ontological premises. Moreover, he argued that such an abstract study underlines the theory of meaning as well as that of expression-elements. For just as in phonetics 'cat' can be further analysed into /c/, /a/, /t/, so the content of 'ram' can be analysed as 'he-sheep', that of 'horse' as 'he-horse', and so on; so that in formal terms we can ask what is the set of minimal constituents and operations on them which will enable us to describe the structure of the content of the words of a language. To distinguish his conception of linguistics as the study of form, that is of a science which 'would be an algebra of language, operating with unnamed entities' (1943: 79), Hjelmselv coined the term 'glossematics'. Whether so abstract a study could be the fundamental study of a natural language, and indeed whether Saussure thought it could be, seems doubtful. But Hjelmslev's work is important because his development of the Saussurean claim that language is a form led to a formal conception of linguistics, and encouraged the comparison of the structure of natural languages with other formal structures in a number of fruitful ways.

4 Post-Bloomfieldian linguistics and distributional relations

Post-Bloomfieldian linguistics developed sophisticated methodologies for the analysis of a linguistic corpus, and, as an example, we shall concentrate on one of those developed by Z.S. Harris. This methodology did not have to be used when framing hypotheses, Harris argued, but it is needed to check the validity of claims made; and though there is no uniquely correct methodology, 'what is essential is the restriction to distribution as determining the relevance of the inquiry' (Harris 1951: 6). Since distributional relations are ones of co-occurrence or of mutual substitutability, the restriction is tantamount to the Saussurean one of the study of syntagmatic and paradigmatic relations. So that, though a corpus-based study like Harris' has no use for the conception of *la langue*, it does restrict the field of inquiry in one important respect in an analogous way to that in which Saussure did.

A distributional analysis must proceed at two levels, phonemic and morphologic. At each of these levels it must first identify the relevant units, before going on to describe the distributional relations holding between them. To

analyse *This is my house* on the morphemic level we must identify its morphemic segments. What shows, for instance, that *house* is such a segment is the fact that it is replaceable by many other *prima facie* segments, such as *car*, *hat* and *dog*, all of which can occur in other environments, such as *She has a* , *big* , and so forth. Having identified the morphemic segments in a corpus, then to give as compact a description of the corpus as possible it is necessary to group the segments into morphemic classes, and these in turn into more inclusive classes. For instance, segments occurring in the environments *The* and *s* are classed as *N*; whilst those which occur in *ed*-past are classed as *V*, within which class we could, for instance, distinguish segments occurring before *N* (*V-trans*) from those which do not (*V-intrans*). The need for more inclusive classes is shown by the fact that $N + S$ can occur whenever *N* occurs except when *N* itself is $N + S$ - we have both *the dog* and *the dogs* but not *the dogss* - so it is necessary to distinguish $N + S$ from *N* as an N^2. More inclusive classes can be defined in analogous ways, making very compact descriptions of a sentence's constituents (its 'phrase structure') possible. *The boys run*, for instance, consists of a higher order *N* followed by a *V-intrans*; whilst the higher order *N* (an *NP* in more familiar terminology) consists of a determiner and an N^2 which in turn consists of an $N + S$.

The great achievement of the post-Bloomfieldian linguistics was to propose many detailed analyses of sentences in terms of their phrase structure. This made the development of a theory of syntax possible, which is something that Saussure had notoriously failed to do. It also enabled the formulation of a strong hypothesis about *all* theories of structural linguistics, namely that because of the restriction to distributional relations, the most powerful syntactic description of a sentence available to it is of that sentence's phrase structure. Whether the kind of methodology proposed by Harris is capable of yielding such descriptions is another matter. Because of the refusal to appeal to meaning it is fatally flawed from a Saussurean perspective. And whether or not that is so, it seems that to apply his methodology Harris would often have to assume he knew something which at that stage had still to be established; it is hard to see, for instance, how on his account one could identify any morphemic segment unless one had already identified some others. But though the methodology is flawed, the crucial objection - due to Chomsky - is that the linguistic descriptions proposed are inadequate; a phrase structure analysis cannot distinguish *John is eager to please* from *John is easy to please*, for all that in the first *John* is the agent, but not in the second. So despite its many insights and achievements, structural linguistics has inherent limitations.

See also: Moscow-Tartu school; Russian literary Formalism; Structuralism; Structuralism in literary theory; Syntax

DAVID HOLDCROFT

References and further reading

Bloomfield, L. (1933) *Language*, New York: Holt.(A classic text which proposes a behaviourist approach to linguistics modelled on contemporary psychology, and sets the agenda for American structuralism.)

Caws, P. (1988) *Structuralism: The Art of the Intelligible*, Atlantic Highlands, NJ: Humanities Press.(Part IB is an excellent introduction to the linguistic basis of structuralism.)

Ducrot, O. and Todorov, T. (1972) *Dictionnaire encyclopédique des sciences du langage*, trans. C. Porter, *Encyclopaedic Dictionary of the Sciences of Language*, Oxford: Blackwell, 1981.(Contains helpful essays on functionalism, glossematics and distributionalism.)

Harris, Z.S. (1951) *Structural Linguistics*, Chicago, IL: University of Chicago Press.(A classic; Chapter 2 on methodology is an important statement.)

Hjelmslev, L. (1943) *Omkring sprogteoriens grundloeggelse*, trans. F.J. Whitfield, *Prolegomena to a Theory of Language*, Madison, WI: University of Wisconsin Press, 1961.(The best introduction to Hjelmslev's thinking.)

Joos, M. (ed.) (1958) *Reading in Linguistics: The Development of Descriptive Linguistics in America since 1925*, New York: American Council of Learned Societies.(Among other papers in this classic collection, Bloomfield (1926), Harris (1946) and Wells (1947) provide a good introduction to American structural linguistics.)

Lyons, J. (1973) 'Structuralism and Linguistics', in D. Robey (ed.) *Structuralism, An Introduction*, Oxford: Clarendon Press, 1973, 5-19.(A brilliant introductory essay.)

Vahek, J. (1966) *The Linguistic School of Prague: An Introduction to its Theory and Practice*, Bloomington, IN, and London: Indiana University Press.(An impressive introduction by a member of the School, with a detailed bibliography.)

Structuralism in literary theory

'Structuralism' is a term embracing a family of theories that between them address all phenomena of the human world - notably language, literature, cookery, kinship relations, dress, human self-perception. In all these domains, structuralists claim, the observable, apparently separate elements are rightly understood only when seen as positions in a structure or system of relations.

The linguist Ferdinand de Saussure is generally recognized as the founder of the structuralist movement. For him semiology - the science of the meaning of natural languages - consists in determining the formal place of any signe *within the inclusive system of signs that is language (*langue*), that is, to see it as a 'difference' among the system of inseparably linked 'differences'. Literary significance is treated in a similar way. But both in linguistic and literary studies the existence of a complete and closed system has been largely anticipated, presupposed rather than confirmed, where no more than fragments of the supposed system could ever really be collected.*

This itself is a point of serious contention: for one thing, the meaning of any fragment of a would-be system seems, on the structuralist view, not to be defined if the full system is not accessible; for another, there is no way to approximate to the inclusive system to which apparent fragments belong. But if that is so, it is asked, then can structuralism - whether applied to literature or to language in general - be a science at all?

1 Historical overview

The earliest structuralist practitioners includeSaussure, the Russian linguist and theoristT.S. Trubetzkoy, and Roman Jakobson. The movement expanded rapidly to encompass a remarkably diverse company, of which the best-known (though not theoretically uniform) contributors include the anthropologist Claude Lévi-Strauss, the psychoanalyst Jacques Lacan, the narratologist Gérard Genette, the early Roland Barthes, the literary critic Michael Riffaterre, the theoretical linguistLouis Hjelmslev, the analyst of narrative and social structureA.J. Greimas, and the folklorist V. Propp. To list such a diverse group is to confirm structuralism's cross-breeding tendencies. This was already apparent in the strong Slavophone branch of the movement, which linked Russian Formalism with structuralism and saw the flowering of Prague linguistics and literary studies in the second and third decades of the twentieth century. Remarkably, Roman Jakobson was prominent in both Slavic movements and appears to have introduced Lévi-Strauss to structuralism.

Once it became clear that the fixed system of *langue* or literature was unlikely to be determined and the seminal idea of an underlying network of internally linked significative *relata* became more and more generously construed, it became commonplace to recognize conceptual homologues of the structuralist orientation: for instance, in the works of Louis Althusser, Michel Foucault, Mikhail Bakhtin, Northrop Frye, Noam Chomsky, Lucien Goldmann, René Wellek, Jean Piaget, Bourbaki, Pierre Bourdieu, and even Ernst Cassirer. In any case, the general eclipse of the movement's core (that is, the eclipse of the model of a science of 'differences' or *relata* functioning within a closed system of *relata*) has ensured the permanent presence of the structuralist theme itself - among the human studies (for instance, various literary, linguistic and cultural studies). For there is no commanding alternative model of description and explanation among these studies that has been able to displace the semiological (significative, symbolic, syntactive, self-referential or interpretable) dimension of the human world, or to account for it in a way that does not to some extent suggest a system of differentiated *relata* construed as a holist model of rationality (a closed system factored by constituent *relata*).

2 General features of structuralism

Structuralism is traced as a distinct movement largely through the genealogy of its originating theories. For example, it is often claimed that Northop Frye's *Anatomy of Criticism* (1957) contributed to English-language literary structuralism. But Frye is explicitly an Aristotelian; his theory of determinate literary genres is largely based on Aristotle's conception of a biological species applied to forms such as tragedy and related genres thatFrye believes fit the same scientific analogy. Frye's 'structuralism' cannot reasonably be traced back toSaussure or Jakobson or to any intermediary figures, and so it is rightly thought not to be structuralist in the original sense of the term, though it bears a remote resemblance to structuralism. A similar argument may be mounted to show that Noam Chomsky is not a structuralist. There is little to be gained from such disputes.

Structuralism was originally motivated by a repugnance for those fields of the human studies that could not reasonably claim the status of a bona fide science - that is, could not address what is distinctive about the human world and at the same time be capable of supporting a methodological rigour comparable to that of the natural sciences. Without ever proving it, the structuralists (notably, Lévi-Strauss) believed that the 'laws of thought' (which supposedly determined the system underlying the phenomena of the human world) were ultimately the same as the laws of physical nature. In fact, they could not provide a sustained account of how the underlying structure could be confirmably abstracted from or imposed a priori on the pertinent phenomena.

The difficulty is deeper than it first appears. Consider that distinctly human phenomena are formed, entrenched and made legible to the apt members of a human society through practices that enable new cohorts both to share those practices and to discern their significance. The aggregated body of such practices is plainly collective, in the sense that no single participating human can master all of them. Oddly, on the one hand the structuralists tend *not* to theorize about the causative conditions under which system-like structures may be deemed to regularize the variable phenomena of cultural life; on the other, there are many theories that admit something like the structuralist thesis but are palpably not structuralist in the relevant sense. Almost any account of societal life - Hegelian, Marxist, Durkheimian, Weberian, Diltheyan, phenomenological, neo-Kantian, hermeneutic, Frankfurt Critical, Wittgensteinian and possibly even Aristotelian - will bear some resemblance to the structuralist emphasis; but that hardly warrants marking them down as structuralists, in all but an unacceptably vacuous sense. It is currently well-nigh impossible to distinguish sharply between structuralist or semiological analyses of literary materials and, for instance, hermeneutic accounts - except for differences in their theoretical lineage.

In this sense, both Barthes (1977) and Bloom (1973) favour the imposition of interpretive codes on canonical texts and canonical readings - codes that claim no canonical validity of their own beyond the power of their sheer plausibility and fresh force; and yet, Barthes, but notBloom, will be recognized as a kind of structuralist despite his (welcome) liberties. In fact, Jakobson and Lévi-Strauss' analysis of Baudelaire's *Les Chats* (1962), which many claim to be the very paradigm of a structuralist analysis of poetry, has been seriously (and reasonably) challenged - not so much for the details of its analysis as for the mistaken picture of structuralist analysis it affords (Riffaterre 1970).

3 Treatment of literature

To speak of a structuralist account of literature - chiefly poetry and narrative prose - is to emphasize, however informally, one or more of the following themes:

(1) the underlying and invariant laws of mental activity, unconscious at the level of linguistic and other cultural phenomena (Lévi-Strauss' central theme (1967));
(2) a dominant binarism ofSaussure's kind (*signifiant*/*signifié*), applied congruently with Jakobson's analysis of metaphor and metonomy (1956);
(3) a formal system of interpretive analytic codes said to constitute the underlying structure of the domain in question (notably, in V. Propp (1928) and A.J. Greimas (1987));
(4) any combination of (1)-(3), strictly or loosely.

Any interpretive critic or theorist meeting conditions (1)-(4), exhibiting a relevant lineage, may reasonably be said to be a structuralist. In this sense, Barthes (here, the 'later' Barthes, author of *S/Z*) andFoucault *are* structuralists, in spite of the fact - though possibly because of it - that they are also poststructuralists, that is, critics of the extreme presumption of the canonical structuralists; they are themselves interpretive critics who attack the sense of closure, apodictic certainty, a priori necessity or the fixities of canonical structuralism.

Barthes and Riffaterre provide an interesting contrast in their development; for both may be said to exhibit the general marks of structuralist lineage: in *The Elements of Semiology* (1964), for instance, Barthes accepts a relatively straightforward Saussurean theory; but, in *S/Z* (1970), in 'From Work to Text' (1977) and in other places, he overthrows strict structuralism in a decidedly florid poststructuralist manner. Riffaterre begins early on as a critic of Jakobson's poetic model, but then adheres quite strictly to the closure and relative isolation of the poetic domain itself (1984). Tacitly, Barthes' model calls Riffaterre's closure into question; and Riffaterre's would deny the coherence ofBarthes' practice. Yet both are structuralist. Without giving a detailed comparison of their critical methods, it needs to be said that the underlying dispute that their opposed practices implicate is the one that

animates Lévi-Strauss' criticism of Jean-Paul Sartre (*Tristes Tropiques* 1955).Lévi-Strauss challenges the putatively superficial history of phenomenal change, favouring instead the 'diachronic' articulation of an underlying 'synchronic' system - and this yields a sense of history deeper than empirical history could ever afford. This is what Lévi-Strauss means by 'anthropology' and its justified displacement of 'history'. The difference between Barthes and Riffaterre is the difference between Lévi-Strauss' andSartre's conceptions - within the space of structuralism.

Riffaterre's isolation of the poetic world from the intrusions of contingent history (developed through his theory of reference) justifies in his eyes the rejection of historical hermeneutics, historical genre studies, associative and impressionistic literary criticism and the like. Riffaterre falls back on the implicit codes of interpretation (sets of formal homologues, he believes) rightly applied to the images of particular poems. Barthes, by contrast, invents new interpretive 'codes' *ad hoc* from the 'infinite' resources of our enveloping culture; here, the undeniable ingenuity of his interpretation (in *S/Z*) threatens (implicitly) to confound Riffaterre's exclusiveness. On the historicist side (ultimately, Foucault's theme), the pretence of constructing a fixed, unconscious system is itself no more than an artefact of a contingent *episteme* (a regime of reason or knowledge) that, in time, will also be superseded. On the 'anthropological' side (which Riffaterre advocates), there can be no methodological rigour in a practice of analysis that is entirely open to the vagaries of phenomenal history.

The formalism of the structuralist approach to literature is possibly more the result of Jakobson's conviction and influence than that of any other early structuralist. Jakobson himself theorized more about poetry than narrative prose. But his single most important contribution to a structuralist analysis of literature, abstracted in a relatively late summary (1960), marks the distinction of poetry but also accords implicitly with Propp's influential study on the narrative structure of the folktale, originally published in 1928. Both Barthes (on narrative) and Riffaterre (on poetry) derive fromJakobson, although Barthes is certainly no formalist and Riffaterre expressly objects toJakobson's theoretical model. The line of narrative analysis (which extends to psychoanalysis and film) may be fruitfully traced to Greimas, Genette,Todorov, Eco and Kristeva. Regarding the analysis of poetry, there is hardly any figure who compares favourably with Riffaterre (1978).

Jakobson's formula (1960) is both famous and arresting: 'The poetic function projects the principle of equivalence from the axis of selection into the axis of combination.' With this, Jakobson means to distinguish between metaphor (vertical equivalences of meaning drawn from a suitable fund: 'selection') and metonomy (horizontal equivalences affected by contiguities of sentential structure: 'combination'); and to indicate the executive direction of poetry *from* the 'associative' equivalences of an imaginative play with images and meanings *to* the 'syntagmatic' equivalences of structural role, stress and the like in sentential contexts. Ironically, Barthes' treatment of *S/Z* may resemble Jakobson's model of poetry more thanRiffaterre's treatment of poetry itself.

See also: Althusser, L.; Bakhtin, M.; Barthes, R.; Bourdieu, P.; Cassirer, E.; Chomsky, N.; Foucault, M.; Kristeva, J.; Lacan, J.; Lévi-Strauss, C. §2; Piaget, J.; Sartre, J.-P.; Structuralism; Structuralism in linguistics

JOSEPH MARGOLIS

References and further reading

Barthes, R. (1964) *Elements of Semiology*, trans. A. Lavers and C. Smith, London: Jonathan Cape, 1967. (Mentioned in §3. Barthes' view of Saussurean semiology.)

Barthes, R. (1970) *S/Z*, trans. R. Miller, New York: Farrar, Straus & Giroux, 1974.(Mentioned in §3. Barthes' freewheeling quasi-structuralist reading of Balzac's *Sarrasine*.)

Barthes, R. (1977) 'From Work to Text', in S. Heath (trans.) *Image, Music, Text*, New York: Wang & Hill. (Mentioned in §§2, 3. Perhaps the best-known of Barthes' short statements on reading literature, as a poststructuralist.)

Bloom, H. (1973) *The Anxiety of Influence*, London: Oxford University Press.(Mentioned in §2. Possibly the best-known American poststructuralist manifesto of critical reading.)

Caws, P. (1988) *Structuralism: The Art of the Intelligible*, Atlantic Highlands, NJ: Humanities Press International. (A reliable overview of the entire movement.)

Frye, N. (1957) *Anatomy of Criticism: Four Essays*, Princeton, NJ: Princeton University Press.(Mentioned in §2.Frye's Aristotelian model applied to literature.)

Greimas, A.J. (1987) *On Meaning*, trans. P. Perron and F.H. Collins, Minneapolis, MN: University of Minnesota Press.(A selection of Greimas' writings, taken from various volumes.)

Jakobson, R. and Halle, M.(1956) *Fundamentals of Language*, The Hague: Mouton.(Mentioned in §3. Contains Jakobson's discussion of metaphor and metonymy.)

Jakobson, R. (1960) 'Closing Statement: Linguistics and Poetics', in A. Sebeok(ed.) *Style in Language*, Cambridge, MA: MIT Press, esp. 350-77.(Mentioned in §3. Possibly the most succinct and accessible of Jakobson's statements on poetics.)

Jakobson, R. and Lévi-Strauss, C. (1962) '"Les Chats" de Charles Baudelaire', *L'Homme* 2 (1): 5-21.(The text ofJakobson and Lévi-Strauss' famous analysis.)

Lévi-Strauss, C. (1955) *Tristes Tropiques*, trans. J. and D. Weightman, London: Jonathan Cape, 1973.(Mentioned in §3. Autobiographical approach to Lévi-Strauss' reflections on the development of his method.)

Lévi-Strauss, C. (1967) *The Elementary Structures of Kinship*, trans. J.H. Bell, J.R. von Sturmer and R. Needham, Boston, MA: Beacon Press, 1969.(Mentioned in §3. Affords a sense ofLévi-Strauss' view on the laws of the human mind.)

Propp, V. (1928) *Morphology of the Folktale*, trans. L. Scott, revised L.A. Wagner, Austin, TX: University of Texas Press, 1968, 2nd edn.(Mentioned in §3. The famous structuralist analysis of the folktale.)

Riffaterre, M. (1970) 'Describing Poetic Structures: Two Approaches to Baudelaire's *Les Chats*', in J. Ehrmann (ed.) *Structuralism*, Garden City, NY: Anchor Books, esp. 188-230. (Mentioned in §2. Riffaterre's well-known criticism of Jakobson and Lévi-Strauss' reading of Baudelaire's poem.)

Riffaterre, M. (1978) *Semiotics of Poetry*, Bloomington, IN: Indiana University Press.(A good introduction to Riffaterre's critical practice.)

Riffaterre, M. (1984) 'Intertextual Representation: On Mimesis as Interpretive Discourse', *Critical Inquiry* 11 (1): 141-62.(Mentioned in §3. A clear summary of Riffaterre's theory of critical reading.)

Saussure, F. de (1983) *Course in General Linguistics*, trans. R. Harris, La Salle, IL: Open Court.(The essential source of structuralism.)

Structuralism in social science

Any school of thought in the social sciences that stresses the priority of order over action is 'structural'. In the twentieth century, however, 'structuralism' has been used to denote a European, largely French language, school of thought that applied methods and conceptions of order developed in structural linguistics to a wide variety of cultural and social phenomena. Structuralism aspired to be a scientific approach to language and social phenomena that, in conceiving of them as governed by autonomous law-governed structures, minimized consideration of social-historical context and individual as well as collective action. Structural linguistics was developed in the early part of the twentieth century primarily by the Swiss linguist, Ferdinand de Saussure. After the Second World War, it fostered roughly three phases of structural approaches to social phenomena. Under the lead of above all the French anthropologist Claude Lévi-Strauss, classical structuralism applied structural linguistic conceptions of structure with relatively little transformation to such social phenomena as kinship structures, myths, cooking practices, religion and ideology. At the same time, the French psychoanalyst Jacques Lacan appropriated Saussure's conceptual apparatus to retheorize the Freudian unconscious. In the 1960s, a second phase of structural thought, neo-structuralism, extended structural linguistic notions of order to a fuller spectrum of social phenomena, including knowledge, politics and society as a whole. Many of Saussure's trademark conceptions were abandoned, however, during this phase. Since the 1970s, a third phase of structuralism has advanced general theories of social life that centre on how structures govern action. In so refocusing structural theory, however, the new structuralists have broken with the conception of structure that heretofore reigned in structural thought.

1 Social structure

The idea of social structure articulates the intuition and experience that social existence is ordered and not merely random and chaotic. Roughly four conceptions of social structure dominate in the social disciplines. At one extreme lies voluntarism, which combines the thesis that all there is to social phenomena are individuals' actions (and mental states) with the idea that with every action individuals wilfully recreate whatever order exists in social life. On this view, there are no real social structures. A second position, considerably more popular than voluntarism, is empiricism, the view that social structures are patterns in actions and practices. These patterns are usually not thought of as determinants of human activity. Rather, they are by-products of actions and practices and the mental states responsible for these. A third conception, grammaticism, assimilates structure to grammatical rules conceived of as non-conscious principles governing speech. On this view, social structures are the sets of rules and principles that govern activities in the different domains of social life. These rules do not determine action, but structure its possibilities, and do so by being embedded in actors' practical understandings or cognitive processes. A final conception, determinism, construes structures as abstract systems that determine action and social phenomena without working through practical or cognitive understanding. These systems are highly counterintuitive, for they are composed of elements, relations and principles known to social scientists alone. Structuralism typically works with structures of this fourth kind.

2 Structural linguistics

The notions of structure and order applied by later structuralists to cultural and social phenomena were introduced first in linguistics by Ferdinand de Saussure in his *Cours de linguistique générale* (1916). Above all two Saussurian ideas influenced structural thought after the Second World War. The first was that language is an autonomous structure distinct from the world. Saussure conceived of language as a system of signs and of a sign as a combination of signifier and signified. A signifier is a material entity - a sound, image, pencil mark - with meaning. Its meaning is the signified associated with it. Saussure argued that the signifiers and signifieds that together constitute signs are defined solely through their differences with other signifiers and signifieds, respectively. This thesis draws signifiers, signifieds and thereby signs into closed, discrete systems. Language is an autonomous structure of elements defined internally through differences, which can be studied in isolation from the world to which it refers and the activities in which it is used (see Saussure, F. de).

Saussure's second telling idea was that abstract structures govern spatial-temporal events. The autonomous system of elements that constitutes language (*langue*) is an abstract (or 'synchronic') entity lying outside space and time

that governs spatial-temporal (or 'diachronic') acts of speaking and writing (*parole*). This famous distinction between *langue* and *parole*, between abstract and concrete linguistic entities, lines up with at least two further oppositions, namely, depth/surface and essential/accidental. Concrete speech acts are accidental and variable phenomena in comparison to the underlying, abstract and essential structure of language.

Saussure thus detached language from social-historical context as well as from the understandings and practices of actual speakers and writers. In studying language, consequently, both history and people can be bracketed (except for comparative purposes).

3 Classical structuralism

After the Second World War, structuralism broke out of the domain of language (and poetry as well as folklore) to focus upon cultural and social phenomena more generally. Its initial phase was marked by wholesale application of the Saussurian apparatus to social affairs and by the resulting assimilation of these affairs to language. Among the social sciences, classical structuralism enjoyed greatest impact in social anthropology, undoubtedly because of the brilliance of that theorist who, more than any other, epitomizes structural social investigation: Claude Lévi-Strauss.

Heralding structural linguistics as the long awaited breakthrough that set the human studies for the first time firmly on the road to scientific knowledge, Lévi-Strauss proposed to extend the field's methods and conceptions of order to encompass communication, or symbolic, systems generally. He wielded, moreover, an expansive notion of a communication system, one that embraced not only myths, religion, and ideology, but kinship structures and social organization as well.

Lévi-Strauss took from Saussure the idea that communication systems are closed, discrete systems that can be studied in isolation from both socio-historical context and the consciousness of actors. Each communication system and type thereof, moreover, possesses an underlying synchronic structure that governs spatial-temporal practices and contents even though it is unknown to the actors involved. Lévi-Strauss further retained the idea that differences and relations, especially binary oppositions, play a crucial role in defining structural elements.

Structures, for Lévi-Strauss, are scientifically constructed matrices of possible combinations of relations (between elements). In the case of a particular kinship system, for instance, an anthropologist might construct a matrix along whose horizontal axis lie 'warm' (friendly) and 'cold' (distant) brother-sister and husband-wife relations, and along whose vertical axis lie warm and cold father-son and mother's brother-sister's son relations. The kinship structure involved would then consist in the possible combinations of warm and cold relations that can characterize the kinship system. According to Lévi-Strauss, which combinations of relations are possible is determined by a rule of transformation (a law), which thus specifies the range of combinations of warm and cold kin relations possible in the system. In this way, autonomous and internally defined structures lying outside space and time govern spatial-temporal states of affairs encompassing individuals.

Most provocatively, Lévi-Strauss maintained not only that all communication systems of a given type (for example, kinship systems) share certain structures, but that the different communication systems found both in a given society and across societies generally are governed by homologous structures. The reason for this homology, he hypothesized, was that all communication (symbolic) systems are ultimately grounded in the logical properties of cognitive brain processes.

The grandeur and scope of Lévi-Strauss' vision captivated a generation of anthropologists and other social scientists, especially in France and the UK, and secured for him acclaim as one of the leading twentieth-century theoreticians of social life.

As Lévi-Strauss was extending Saussurian analysis to cultural phenomena, Jacques Lacan was focusing Saussure's signifier-signified model of the sign inward upon the Freudian unconscious. Although Lacan transcended Saussure's conceptions in important ways, and for this reason is labelled a 'post-structuralist' by some commentators, his theories of child development represent one of the most consequent applications of the Saussurian model of the sign to human beings. Like Lévi-Strauss, moreover, Lacan believed that to apply Saussurian conceptions to phenomena other than language is to analyse these phenomena as if they were language. This attitude, it should be noted, also characterizes those 'semiotic' analyses that, in applying Saussure's

conception of the sign to social and cultural phenomena, treat those phenomena as systems of signs, for example, some of Roland Barthes' work (see Barthes, R. §2; Semiotics).

Lacan's account postulated an initial pre-Oedipal, or 'imaginary', phase of development in which the identity of the child is bound up with its mother's body. The subsequent Oedipal, or 'symbolic', phase is inaugurated when the father (the law) divides the child from the mother's body and drives its desire for her underground, thus forming the unconscious. The child's identity now derives from difference (rejection as the mother's lover, absence of the mother's body); and possession of the mother can occur only in the realm of language. So the child's desire for its mother, severed forever from real possession, can only reverberate within the system of language, moving anchorlessly from signifier to signifier. Dreams offer an especially fertile realm of signifiers for the articulation of desire. In this realm, the movement of desire among signifiers, which Freud had referred to as 'dreamwork', is effected by the substitution and association of signifiers. Dreams are series of signifiers, and the unconscious generally a movement of signifiers, whose signifieds are not immediately obvious. Lacan's theory is summed up in his famous epigram, that the unconscious is structured like a language.

Lacan's controversial Saussurian theories of child development and the unconscious have had greater influence in literary studies than in the human sciences. They have also found greater resonance in movements such as academic feminism and post-structuralism that became prominent after structuralism proper. This reflects how Lacan undermines the structural notion of determinate associations between signifiers and signifieds by fracturing desire among a plurality of signifiers to which stable signifieds cannot be attached. It also bespeaks how Lacan, in following Freud and making other people crucial to children's identity and development, opens the door to acknowledging of the crucial impact of society and history upon systems of signifiers and signifieds.

4 Neo-structuralism

During the 1960s, a wide variety of thinkers adopted an array of theses forcefully promoted by classical structuralism: de-emphasis of history, opposition to basing social life in individual subjects, belief in abstract structures and postulation of a governing stratum that underlies the surface appearances of social life. These thinkers were not, however, in a strict sense structuralists. They did not (1) emphasize the role of difference and opposition in the constitution of abstract, underlying structures; (2) employ Saussure's terminology of signifier and signified; or (3) work with a clear distinction between synchrony and diachrony. These changes meant that, although neo-structuralists retained classical structuralism's notion of internally-defined governing structures, they no longer assimilated the social phenomena governed by these structures to language. These changes also facilitated their attempt to extend the scope of governing structures beyond Lévi-Strauss' communications systems to embrace political systems, the organization of knowledge, and societies as wholes. Accordingly, neo-structuralism exploded classical structuralism's confinement primarily to social anthropology and spread structural ways of thinking to political theory, history and sociology.

Perhaps the most prominent branch of neo-structuralism was so-called 'structural Marxism'. Led by the French thinker Louis Althusser, this school theorized society as a set of political, economic, cultural, educational and other structures. The structures comprising a given society were thought to form a total state of interrelatedness, moreover, to which Althusser gave the name 'mode of production'. For the structural Marxists, all social change is a matter of either the recalibration of the different structures constituting a given society (piecemeal change), or the lining up of these structures in a 'structural conjunction' that enables wholesale transformation of the mode of production (revolution). All social causality and change is thus rooted in the realm of interlocking structures. Individuals, as mere occupants of positions established by structures, are powerless to effect change. Significantly, Althusser abandoned Marx's claim that the economy determined all other social sectors, arguing instead that the economic structure determined the remainder of social life only 'in the last instance'. Although this represented a welcome loosening of economic determinism, Althusser's extreme structural determinism made his position unpalatable to most.

A second especially fertile ground for the broadening of structural thinking was the organization of human knowledge. Jean Piaget and Michel Foucault proposed that human knowledge forms systems governed by abstract principles. Piaget elaborated an account of the genetic stages though which children's knowledge of the world progresses, on each level of which a different set of principles structures knowledge. Foucault, meanwhile, espied three large-scale *epistemes* in Western history since the Middle Ages, in each of which different rules governed the

formulation and certification of knowledge of living beings, language and wealth. During each such period, in his view, knowledge of these matters formed systems governed by rules that were unknown to the individual thinkers and researchers who complied with them. But while Foucault retained the structural marginalization of the individual subject, he like Lacan contravened the structural slighting of history. The rules governing knowledge are 'historical a prioris' that change over time. His work thus broke down the distinction between synchrony and diachrony, and in this and other ways marked a shift into post-structuralism.

5 The new structuralists

Classical and neo-structuralism's de-emphasis of history and agency worked against them in the 1970s and 1980s. Social theorists increasingly found it implausible to isolate communication and sign systems, and social phenomena more broadly, from either speakers/writers/actors or the social-historical contexts of speech, script and action. This, together with (1) the post-structural assault on the differences and binary oppositions wielded by classical structuralists, (2) the earlier mentioned (§3) break-up of determinate associations between signifiers, signifieds and referents, (3) the inability of structuralists to theorize satisfactorily the connections between abstract structures and action-practices, and (4) the spreading European reaction against unself-critical scientistic approaches in the social sciences, dampened enthusiasm for both flavours of structural thought.

Remnants of these movements remain however. Cultural studies in the 1980s and 1990s still widely employed Saussure's figure of signifier-signified. The idea of abstract, underlying structures that govern social phenomena similarly persisted in the social sciences. That these structures are composed of relations also persevered as an influential notion.

Revisions of the latter two ideas formed the heart of a third phase of structural thinking. The 'new structuralists' - the French anthropologist Pierre Bourdieu, the English sociologist Anthony Giddens and the English philosopher Roy Bhaskar (see Critical realism) - prolonged structuralism's emphasis on governing structures. In constructing theories of social life in general, moreover, they perpetuated the structural Marxist attempt to theorize all of social life as governed by structures. Their pursuit of this aim, however, granted action unprecedented prominence: structures relate to social phenomena only by way of governing the actions that constitute these phenomena.

This move underlay the profoundest difference between new structuralism and the two prior phases of structural thought. Structures are not autonomous systems outside space and time whose relations to social phenomena and spatial-temporal doings are an unresolved enigma. Structures instead govern action in being the contents of the practical knowledge on the basis of which people act and interact. The new structuralists thus abandoned the hitherto reigning 'deterministic' conception of structure in favour of a 'grammatical' one (see §1). Indeed, on their view, not only does structure govern action, but structure exists and is able to govern action only because action draws on and thus renews structure. Structure is both the medium and result of action. This circular dependency was accompanied by retheorization of how structures 'govern' action. In Giddens, for example, structures accomplish this by opening and closing possibilities of action.

At the same time, the new structuralists reached back to classical structuralism in depicting the structures embedded in practical understanding as centrally composed of relations and differences. Giddens, for instance, theorized structures as sets of interrelated rules (of meaning and normativity) and resources (arising from commands over persons and objects). These sets are linked by relations of interconvertibility, and thereby form 'structural sets'; redolent of Lévi-Strauss' matrices, that govern systems of action and practice. Bourdieu, meanwhile, exhibited greatest continuity with classical structuralism, construing the structures that govern action and practice as systems of oppositions (such as those between up/down, hot/cold and wet/dry). He combated the hypostatization of abstract structure pervasive in the two earlier phases of structuralism, however, by basing the system of oppositions on opposed movements of the human body and on the biological difference between male and female.

Despite their abandonment of the deterministic notion of structure, the new structuralists perpetuated Saussurian thinking in their accounts of grammatical structure. The future of structural thinking presumably rests on the possibility of further novel transformations of traditional structural themes.

See also: Post-structuralism in the social sciences §1

THEODORE R. SCHATZKI

References and further reading

Althusser, L. (1965) *Pour Marx*, Paris: Librairie François Maspero S.A.; trans. B. Brewster, *For Marx*, London: Allen Lane, 1969.(A difficult but rewarding entry into structural Marxism.)

Barthes, R. (1957) *Mythologies*, Paris: Éditions du Seuil; trans. A. Lavers, New York: Hill & Wang, 1972.(An entertaining work illustrating the influence of structuralism on semiotics.)

Bourdieu, P. (1980) *Le sens pratique*, Paris: Éditions de Minuit; trans. R. Nice, *The Logic of Practice*, Stanford, CA: Stanford University Press, 1990.(A masterful but difficult theory of the relations between practice and structure combining a structural emphasis on oppositions with a strong practical sensibility.)

Foucault, M. (1966) *Les Mots et les choses*, Paris: Éditions Gallimard; trans. A. Sheridan, *The Order of Things*, London: Tavistock, 1970.(Foucault's analysis of the three *epistemes* of modernity and their limits.)

Giddens, A. (1979) *Central Problems in Social Theory*, Berkeley, CA: University of California Press.(A comprehensive work that surveys central questions while constructing answers through an innovative theory of action-structure.)

Lacan, J. (1966) *Écrits*, Paris: Éditions du Seuil; trans. A. Sheridan, *Écrits: A Selection*, London: Tavistock, 1977. (An extremely difficult collection of essays setting forth Lacan's account of child development and the unconscious.)

Lévi-Strauss, C. (1958) *Anthropologie structurale*, Paris: Librairie Plon; trans. C. Jacobson and B.G. Schoepf, *Structural Anthropology*, New York: Basic Books, 1963. (An accessible and recommended introduction to the dimensions and concerns of classical structuralism.)

Lévi-Strauss, C. (1962) *La Pensée sauvage*, Paris: Librairie Plon; *The Savage Mind*, Chicago, IL: University of Chicago Press, 1966.(A somewhat more difficult and synoptic presentation of classical structuralism.)

Piaget, J. (1970) *Genetic Epistemology*, trans. Eleanor Duckworth, New York: Columbia University Press.(A succinct introduction to structural epistemology.)

Robey, D. (ed.) (1973) *Structuralism: An Introduction*, Oxford: Clarendon Press.(A useful multidisciplinary overview of the subject.)

Saussure, F. de (1916) *Cours de linguistique générale*, C. Bally, A. Sechechaye and A. Riedlinger (eds), Paris: Payot; trans. W. Baskin, *Course in General Linguistics*, New York: Philosophical Library, 1959.(The original formulation of structural linguistics that underlies all subsequent structuralism, semiotics, and post-structuralism.)

Subject, postmodern critique of

The critique of the subject in late twentieth-century continental philosophy is associated primarily with the work of Foucault, Derrida, Lacan and Deleuze. Driven by philosophical, political and therapeutic concerns, these thinkers question the subject's ability to declare itself self-evidently independent of the external conditions of its own possibility, such as the language in which it expresses clear and distinct ideas, the body whose deceptions it fears, and the historical or cultural conditions in which it perceives reason or tyranny. Moreover, they fear that the ethical price of such insistence upon absolute self-possession is the exclusion and oppression of social groups whose supposed irrationality or savagery represent the self's own rejected possibilities for change and discovery. Their work draws upon Marxist, Freudian and Nietzschean insights concerning the dependence of consciousness upon its material conditions, unconscious roots, or constituting 'outside'. However, their use of these influences is guided by a common fidelity to Kant's search for the 'conditions of possibility' underlying subjective experience, as well as his scepticism regarding our capacity to know the self and its motivations as objects 'in themselves'.

1 Intellectual and historical context

The work of Foucault, Derrida, Lacan and Deleuze, the thinkers primarily associated with the critique of the subject in late twentieth-century philosophy, reflects the intellectual and the political milieu of 1950s and 1960s France, but it also challenges important currents of thought regarding the relation of philosophical reason to historical change or progress. The critique of the subject is simultaneously an attack upon the idea of a universal humanist subject whose reason is reflected in the thought of all 'civilized' cultures, and an attack upon traditional philosophical dichotomies such as reason/madness, consciousness/embodiment or civilization/savagery. To a certain extent, therefore, the critique of the subject is an attempt to renegotiate the conceptual relation between history and philosophy, questioning whether the philosophical or scientific subject is necessarily the primary factor in the intelligibility of historical events. It is both an event in the history of ideas and a challenge to the way in which events (including intellectual events) have been conceptualized by philosophical reason. Traditionally, the self-present, freely acting subject has been considered the pivot linking historicist and transcendental accounts of Western political and psychological experience. A philosophical perspective which questions the plausibility of such a subject reopens the question of how philosophers can make sense of their own historical and psychological experience.

Anti-subjectivist approaches to philosophy arose in reaction to Sartrean existentialism and the extreme capacity for freedom, self-creation and historical agency with which Sartre credited the conscious ego (see Sartre, J.-P.). But Hegel was an equally prominent ghost on the intellectual horizon of this time, due to Kojève's influential lectures and translation of the *Phenomenology of Spirit* (see Hegel, G.W.F. §5; Kojève, A.). Where for Kant the self remained ignorant of its own nature and motivations as thing-in-itself, Hegel argued that the self *was* capable of thorough self-knowledge and conscious freedom to the extent that it recognized social and historical structures as mediating the relation between itself as consciousness and as thing-in-itself. Anti-subjectivist thinkers focused on the role played by mediating social structures in the formation of self-consciousness while retaining Kant's suspicion that such self-knowledge could never be complete. Unlike Kant, however, they placed higher ethical and political significance upon the recognition of this *blind spot* in the constitution of human freedom than upon the belief in human autonomy made possible by such scepticism.

Both Hegel and Sartre believed that historical struggles for freedom constituted the objective basis for the increasing rationality and self-knowledge of the subject. Western political culture, liberal or Marxist respectively, represented for these thinkers the highest form of this self-consciousness to date. It is significant, therefore, that the anti-subjectivist reaction against Hegel and Sartre took place amid upheavals in the world-historical role of the French state. Well before the student uprisings and general strike of May 1968, anti-colonial revolutions in Southeast Asia and Africa explicitly challenged the traditional equation of universal human subjectivity with the political consciousness specific to Western 'man'. The Nazi Holocaust and the legacy of Western imperialism had made it increasingly difficult to perceive the subject described by Western philosophy as the unquestionable vanguard of humankind's march toward historically embodied reason and freedom. Indeed, it became clear how much of the West's material progress and cultural self-sufficiency were built upon the oppression and exclusion of ethnic populations elsewhere in the world.

Ironically, considering his personal history of Nazi involvement, one of the single most important influences upon postmodern critiques of the subject was Martin Heidegger's lifelong investigation into the nature of the subject's *subjectivity* - into what makes it a 'subject' and makes it possible for us to speak of 'subjectivity' as such. For Heidegger, the subject's autonomy and capacity for knowledge do not precede its interaction with objects and other humans but arise amid a pre-given world of involvements, projects, moods and historically given meanings. Heidegger insists that human 'being-there' (*Dasein*) can only be understood from *within* those involvements, through an investigation of the *Being* of both subject and object. Language indicates the various ways in which Being occurs in the world of human involvements. Both the etymology and the ordinary usage of 'being' offer archaeological evidence of the ways in which humans have historically understood the relation between their self-conception as subjects and the way in which both subjects and objects come to *be* (see Heidegger, M. §§2-4).

2 Thinkers and themes

Heidegger's effort to illuminate suppressed or latent philosophical alternatives to the structures through which subjects and objects currently appear to one another served as a starting point for Foucault's investigation into the relationship between history and traditional philosophical conceptions of the self-knowing subject (see Foucault, M.). The Foucauldian critique of the subject explores the historical conditions making possible various conceptions of subjectivity, agency and truth. In his most anti-Sartrean work, *L'Archéologie du savoir* (1969) (*The Archaeology of Knowledge*, 1972), Foucault questions whether historical events are best understood as expressions of human agency, especially the sort of agency thought to define the Western political subject. Like Derrida, Foucault also questions the ontological unity of the philosophical and literary author, exploring the way in which the agency and consciousness attributed by readers to a philosophical or literary author depend upon social and academic conventions regarding the relation between types of texts. In doing so, he draws attention to the different ways in which the concept of truth has established criteria for the assignment of agency and consciousness to particular actors, writers and processes at different points in history. Like Derrida and Deleuze, Foucault attempts to develop the broader philosophical implications of Nietzsche's inquiry (known as genealogy; see Genealogy) into the historical or political origins of central metaphysical concepts such as truth or the good. His goal is less to embed these structures and criteria in a more accurate and complete history than to emphasize their contingency and possible malleability.

Foucault is best known, however, for philosophical histories which describe the production of modern individualized self-consciousness through the everyday social impact of institutions such as the school, prison, hospital, social welfare apparatus, and psychiatric and psychoanalytic discourses concerning sexuality. His work considers the human body as a plastic medium which expresses the biological and productive power of a population insofar as it observes and controls itself under the direction of specific social institutions. Foucault locates modern 'subjectivity' at the nexus of these techniques of control and surveillance, which affect bodies in their most everyday existence and may become the sites of resistance to such institutions.

Analysis of the subject of desire and sexuality forms a major point of contact and contestation among Foucault, Lacan and Deleuze, in addition to 'French' feminists such as Irigaray and Kristeva. In questioning the nature of the subject's relation to history, and in examining the conditions which allow us to perceive reason in both subjectivity and historicity, anti-subjectivist thinkers were also forced to reconsider the nature of the psychological experience in which the relation between reason and madness is expressed. The debt which this philosophical conversation owes to psychoanalysis is enormous. At the turn of the century, Freud's clinical discoveries had undermined the concept of a pure rationality or transparent subjectivity and drawn attention to the complexity of distinctions between the normal and the abnormal or the intellectual and the sexual. Yet by the 1950s the psychoanalytic establishment had downplayed the potentially revolutionary implications of Freudian theory in favour of a practical emphasis on therapeutic techniques designed to *adapt* patients to a social reality considered given and unquestionable. Foucault's early work on the distinction between madness and sanity, and his later accounts of the changing concept of 'sexuality', argued for a historical understanding of the conditions of the possibility of the desiring subject. Yet, despite Freud's contribution to philosophical critiques of the transparent and ahistorical subject, Foucault regarded psychoanalysis as one in a long series of institutions for the identification, study and eventual normalization of those who either would not or could not apprehend themselves as subjects.

Rather than investigate the historical conditions of the possibility of subjective experience and sexual desire,

radical psychoanalyst Jacques Lacan remained interested in the structural and logical conditions making such experience possible and recognizable at *any* point in history. In an effort to restore the critical implications of psychoanalytic theory, Lacan stressed the structural and logical conditions underlying the individual personality as subject of language and desire. His major innovation in Freudian theory is to have interpreted psychoanalysis as a linguistic account of self-formation, and thus to have conceptualized sexual identification and desire as the result of inclusion in a speaking community. In contrast to Anglo-American efforts to think of the psychoanalytic subject as a self-contained (and more or less ‘normal’) manipulator of words, the Lacanian school expressed deep suspicion regarding the extent of the ego’s self-control. For Lacan, neurosis *results* from the ego’s refusal to acknowledge its dependence upon a (conflicted) social and linguistic field in order to maintain an imaginary bodily integrity. Lacan’s theory describes an inevitably *split* subject, whose activity is largely unconscious; a subject formed in the child’s struggle to be represented by language as a speaker. The sexual specificity and desire of the resulting subject are irrevocable marks of this split or insufficiency. Because Lacan considers sexual difference to be a transcendental condition of the possibility of psychological experience as such, his work was both influential upon and ultimately reworked by feminist thinkers who shared his scepticism regarding the existence of transparent (and sexually neuter) subjects (see Feminism and psychoanalysis).

Many of the thinkers under consideration here share a general conviction that the terms of dichotomies such as reason/madness, thought/embodiment, freedom/power, or self/other ultimately depend upon one another and are only maintained *as dichotomies* through an inescapable act of exclusion, intellectual or political. This principle has been most explicitly developed in the work of Jacques Derrida. Like Foucault and Lacan, Derrida was greatly influenced by Heidegger, but unlike the other two he engages directly with Heideggerian texts and problems. Derrida is best known for his critique of the philosophical subject as self-present in the language of its thought. Derrida argues that the history of philosophy has regarded thinking as a sort of transparent awareness of present meaning, comparable to the testimony of a spoken confession rather than the (potentially misunderstood or forged) evidence of a written signature. In fact, he contends, thought always bears witness to ambiguities, to metaphoric constructions, and to prior texts and social contexts which render it partially opaque even to thinkers themselves. ‘Deconstruction’, the name given to Derrida’s technique for the analysis of authoritative philosophical texts, disrupts apparent hierarchies and identities within texts and cultural phenomena by demonstrating the dependence of a superior or independent term upon a constituting exterior or Other. Deconstruction attempts to show that political, ethical, and philosophical subjectivity bears traces of production by a prior social or conceptual field - one traditionally considered to be the subject’s textual *product*, subordinate to him or her.

The ’French’ feminist critique of the subject as embodied and sexually differentiated expands upon the critique of universal subjectivity offered by psychoanalysis and deconstruction. Yet, because of its links to an evolving international political movement, it also attempts to explore the positive aspects of certain non-traditional accounts of subjectivity and community, especially those which give a central role to embodiment or the act of writing rather than the self-presence of consciousness. ’French’ feminism is usually identified with the work of Luce Irigaray, Hélène Cixous and Julia Kristeva, despite great differences in their individual philosophical projects and the spread of ‘continental’ feminism well beyond the borders of France. This philosophical genre is unique for its critique of the supposed sexual neutrality (and therefore *de facto masculinity*) of the universal philosophical subject. Continental-style feminist philosophy seeks to undo the apparent simplicity and givenness of the notion of sex, and to investigate the manner in which culture and language shape the sexed individual’s opportunities for pleasure, political agency and philosophical activity. Like critics of colonialism, continental feminists are particularly concerned to undermine the association of subjectivity with particular dichotomies such as masculine/feminine or Western/non-Western.

Finally, the work of Gilles Deleuze combines a Derridean emphasis on difference and differentiation, a Foucauldian conception of power and the social field, and the psychoanalytic notion of desire (thought of as productive, rather than split) to create a resolutely Nietzschean understanding of the subject as a multiplicity of forces. In fact, subjectivity is for Deleuze only *one* aspect or face of multiplicity. He regards neither the traditional, transparent subject nor the split psychoanalytic subject as the healthiest or most liberating vehicles for productive, desiring multiplication. His collaboration with radical psychoanalyst Félix Guattari produced a major theoretical study of the way in which psychoanalysis itself perpetuates a form of subjectivity appropriate to capitalist production. *Anti-Oedipus* hoped to find in schizophrenia’s noted inaccessibility to analysis the resources for a way

of thinking about history, desire and subjectivity which resists capitalist culture itself. Deleuze’s other works on cinema, literature and historical figures such as Nietzsche, Hume, Spinoza and Bergson offer concepts with which the philosopher may actively *think* of him or herself as multiple without falling into despair at his or her failure to live up to the Oedipalized subject, whether supposedly self-present or self-absent.

3 Debates

This field of continental philosophy is characterized by several key debates. Many of the critical positions described above are designed to undo traditional political and conceptual hierarchies which limit the political agency of the oppressed, as well as the political imagination of theorists. The question remains, however, whether critiques of the subject do not undermine the very idea of an ‘agent’ capable of political contestation, thus perpetuating the hierarchies they were meant to subvert. Similarly, do theoretical accounts of ‘constructed’ subjects (those of Lacan and Foucault) attribute so much power to social structures that it no longer seems possible to imagine subjects exercising the power which capitalist and disciplinary institutions have invested in their bodily activity and self-policing? Critiques of the humanist subject and of the subject of scientific certainty have raised the spectre of cultural incommensurability which provokes many of the arguments between French-influenced continental philosophers and descendants of the Frankfurt School such as Jürgen Habermas. Another heated debate concerns the purported ‘valorization’ of schizophrenia, hysteria and madness in the works of Foucault, Deleuze and the ’French’ feminists. In attempting to scrutinize the theoretical and political context in which distinctions are made between the sane and the insane, do they romanticize or inappropriately universalize their conception of mental illness?

Finally, it must be noted that the concept of ‘postmodernism’ itself is controversial (see Postmodernism). The term and its possible relevance or meaningfulness depend upon our ability to conceptualize continuity, progress or clear breaks between historical periods. Jean-François Lyotard uses the term ‘postmodern’ to indicate the historical condition making possible various aspects of *modernity*, including the distrust of totalizing historical narratives which has characterized much of late twentieth-century thought. Both Lyotard and the American literary critic and Marxist Fredric Jameson emphasize the difference between post-structuralism, a philosophical movement which questions the ahistoricism and rigidity of structuralist accounts of subjectivity and culture, and postmodernism, a sociological or historical characterization of the culture of late capitalist societies (primarily in the First World). Jameson is critical of the supposed distinction between ‘modern’ and ‘postmodern’ periods, but does acknowledge that the decentralization of productive forces and communications networks, not to mention the saturation of everyday life with advertising, have fragmented social consciousness in a new and potentially problematic way. Whether one regards the critique of the subject and the advent of late capitalism as the results of transcendental structures or historical events, the difficult concept of the ‘postmodern’ calls attention to the uneasy relation between historical frameworks of analysis and the philosophical frameworks which have traditionally made sense of history by relying upon the agency of a supposedly self-present subject.

See also: Alterity and identity, postmodern theories of; Deleuze, G.; Foucault, M.; Heidegger, M.; Lacan, J.; Postmodernism

LAURA HENGEHOLD

References and further reading

Butler, J. (1987) *Subjects of Desire: Hegelian Reflections in Twentieth Century France*, New York: Columbia University Press. (Considers desire as a major Hegelian theme adapted into twentieth-century French thought by Kojève and Sartre and radically transformed by Lacan, Deleuze and Foucault.)

Deleuze, G. (1962) *Nietzsche et la philosophie*, trans. H. Tomlinson, *Nietzsche and Philosophy*, New York: Columbia University Press, 1983. (Creative, influential interpretation of Nietzsche according to an ontology of active and reactive forces rather than subjects, intentions and objects.)

Deleuze, G. and Guattari, F. (1972) *L’Anti-Oedipe*, trans. R. Hurley. M. Seem and H. Lane, *Capitalism and Schizophrenia*, vol. 1, *Anti-Oedipus*, New York: Viking, 1977. (Major anti-psychoanalytic work linking capitalism, psychoanalytic ideology concerning the family and the sane or ‘analysable’ subject.)

Derrida, J. (1967) ‘La Structure, le signe et le jeu dans le discours des sciences humaines’, trans. A. Bass, ‘Structure, Sign, and Play in the Discourse of the Human Sciences’, in *Writing and Difference*, Chicago, IL: University of Chicago Press, 1978. (Excellent introduction to basic concepts and techniques of deconstruction

in the context of Lévi-Strauss's analysis of myths as 'centred structures'.)
Derrida, J. (1972a) 'The Ends of Man', in *Marges-de la philosophie*, trans. A. Bass, *Margins of Philosophy*, Chicago, IL: University of Chicago Press, 1982. (Explores the Eurocentric presuppositions of Western humanism and opposes deconstruction to humanist themes in phenomenology and existentialism.)
Derrida, J. (1972b) 'Signature Event Context', in *Marges de la philosophie*, trans. A. Bass, *Margins of Philosophy*, Chicago, IL: University of Chicago Press, 1982. (Difficult but crucial argument to the effect that iterability, the ability of signs and performances to be transplanted from one context to another, is a necessary condition of their ability to guarantee the authenticity of their use in any given context; first article in a long-running debate between Derrida and the proponents of speech act theory.)
Foucault, M. (1969a) *L'Archéologie du savoir*, trans. A. Sheridan Smith, *The Archaeology of Knowledge and the Discourse on Language*, London: Tavistock Publications, 1972. (Suggests a (non-Marxist) materialist framework for conceptualizing historical change without reference to the agency of individual or collective subjects.)
Foucault, M. (1969b) 'Qu'est-ce qu'un auteur?', trans. D. Bouchard and S. Simon, 'What is an Author?', in *Language, Counter-Memory, Practice: Selected Essays and Interviews*, ed. D. Bouchard, Ithaca, NY: Cornell University Press, 1977. (Proposes that the idea of 'authorship' is a complex effect of conventions linking texts and institutions.)
Foucault, M. (1976) *La Volonté de savoir*, trans. R. Hurley, *The History of Sexuality*, vol. 1, *An Introduction*, New York: Random House, 1978.(Extremely influential and accessible account of the social construction of sexuality through discourses linking social power to the individual's quest for self-knowledge and authenticity.)
Fraser, N. (1989) *Unruly Practices: Power, Discourse and Gender in Contemporary Social Theory*, Minneapolis, MN: University of Minnesota Press.(Presents debates regarding possible political implications of philosophical critiques of the subject).
Irigaray, L. (1984) *Éthique de la différence sexuelle*, trans. C. Burke and G. Gill, *An Ethics of Sexual Difference*, Ithaca, NY: Cornell University Press, 1993.(Attempts to conceptualize a respectful relation between the sexed philosophical subject and his or her 'other', with special attention paid to the imaginative frameworks employed by major figures in the history of philosophy.)
Jameson, F. (1991) *Postmodernism, or, The Cultural Logic of Late Capitalism*, Durham, NC: Duke University Press.(Influential cultural study placing 'postmodernism' as a phenomenon of architectural, artistic, literary and philosophical style in the context of worldwide social and economic changes.)
Kojève, A. (1947) *Introduction à la lecture de Hegel*, assem. R. Queneau, trans. J. Nichols, Jr, *Introduction to the Reading of Hegel*, ed. A. Bloom, Ithaca, NY: Cornell University Press, 1980.(Series of lectures which greatly influenced French reception of Hegel, focusing on individual subjectivity as the locus of historical agency and emphasizing the relationship between the beginning sections of *The Phenomenology of Spirit* and the early Marx.)
Lacan, J. (1966) *Écrits*, trans. A. Sheridan, *Écrits: A Selection*, New York: W.W. Norton, 1977. (Selected articles from the French edition; see especially 'The Mirror Stage as Formative of the Function of the I').
Lyotard, J.-F. (1979) *La Condition postmoderne: rapport sur le savoir*, trans. G. Bennington and B. Massumi, *The Postmodern Condition: A Report on Knowledge*, Minneapolis, MN: University of Minnesota Press, Theory and History of Literature Series vol. 10, 1984.(Explores the declining legitimacy of 'grand narratives' of liberation and scientific progress and its impact on late twentieth-century ideas concerning the role of knowledge in technological societies.)
Megill, A. (1985) *Prophets of Extremity: Nietzsche, Heidegger, Foucault, Derrida*, Berkeley, CA: University of California Press.(Clear, accessible exposition of major themes linking these four authors developmentally.)
Turkle, S. (1978) *Psychoanalytic Politics: Jacques Lacan and Freud's French Revolution*, London: Guilford Press, revised and updated 1992.(Sociological account of Lacanian intellectual scene.)
Whitford, M. (1991) *Luce Irigaray: Philosophy in the Feminine*, London: Routledge.(Detailed and influential account introducing psychoanalytic and philosophical issues in the work of one 'French' feminist philosopher.)
Young, R. (1990) *White Mythologies: Writing History and the West*, London: Routledge.(Places the critique of the subject in the context of post-Second World War anti-colonialist and structuralist thought.)

Sublime, the

The origin of the term 'the sublime' is found in ancient philosophy, where, for example, Longinus linked it with a lofty and elevated use of literary language. In the eighteenth century, the term came into much broader use, when it was applied not only to literature but also to the experience of nature, whereafter it became one of the most hotly debated subjects in the cultural discourse of that age.

The theories of Addison, Burke and Kant are especially significant. Addison developed and extended the Longinian view of the sublime as a mode of elevated self-transcendence, while Burke extended John Dennis's insight concerning sublimity's connection with terror and a sense of self-preservation. While Addison and Burke encompassed both art and nature in their approaches, Kant confined the experience of the sublime to our encounters with nature. In his theory, the sublime is defined as a pleasure in the way that nature's capacity to overwhelm our powers of perception and imagination is contained by and serves to vivify our powers of rational comprehension. It is a distinctive aesthetic experience.

In the 1980s and 1990s Kant's and (to a much lesser extent) Burke's theories of the sublime became the objects of a massive revival of interest, in the immediate context of a more general discussion of postmodern society. Kant's theory, for example, has been used by J.-F. Lyotard and others to explain the sensibility - orientated towards the enjoyment of complexity, rapid change and a breakdown of categories - that seems to characterize that society.

1 Origins

Boileau's translation of Longinus' treatise *Peri Hupsous* (first published in 1674) was probably the single most important factor in the establishment of the sublime as a critical and aesthetic concept. For Longinus, the sublime consists essentially in a loftiness and excellence of language which, rather than engaging reason, serves to transport the reader to a higher plane. Indeed, to express and even exceed the vastness of nature through lofty thought and literature is the ultimate human vocation. It is to transcend the baseness and ignobility intrinsic to merely finite being. This is why Longinus goes on to suggest that those who achieve the sublime in literature are more than mortal; they lift us towards the Deity.

Before considering the influence of Longinus' 'self-transcendent' notion of the sublime on the theorists of the eighteenth century, we must first look at a second fundamental source of the concept. It is to be found in the literary criticism of John Dennis, and especially in his *The Grounds of Criticism in Poetry* (1704). In this work, we are told that the sublime consists of great thoughts that move the reader to 'Enthusiasm'. For Dennis, loftiness (that is, great thought) is a necessary, but not (as it is for Longinus) sufficient condition of the sublime. We require in addition that the great thought should transport the soul by means of 'passion'. This leads to a crucial step. Dennis first suggests that ideas producing terror are especially suited to the sublime. In this respect, he distinguishes between terror *per se* and 'Enthusiastic Terror' - terror that is experienced from a position of safety and is consequently attended by joy. (The paradigm here, one presumes, would be terrible events encountered as the subject matter in literature.) Second, the reason for our pleasure in the sublime is shifted on to a reflective awareness of our present immunity from the terrible events under consideration. Self-preservation rather than self-transcendence provides the ground of our response, and while this emphasis on self-preservation is only an aspect of Dennis's theory of the sublime, it is one which he originates and which marks the logical opposite of Longinus' notion. Together, however, the themes of self-preservation and self-transcendence constitute the seed from which eighteenth-century theories of the sublime grew. I shall now consider this in the context of possibly the most influential pre-Kantian theorists: Addison and Burke.

2 Nature

Addison's major discussion of the sublime is to be found in his famous *Spectator* articles on the pleasures of the imagination. In article 412, the subject is introduced by a way of threefold distinction between the great, the uncommon and the beautiful. For Addison, the 'Great' consists not in mere bulk, but in the largeness of a whole view - such as the scenes of an open country, mountains or wide expanses of water. We will remember that, for Longinus, the state of sublime self-transcendence is achieved through the creation of and response to 'lofty' thought and literature - literature that expresses and exceeds nature's greatness. Addison, by contrast, argues that a

similar state of self-transcendence can come about when we encounter nature’s greatness *by visual perception alone*. Nature provides us with a kind of image of liberty, which takes us beyond the confines of the immediate. The mediation of literature or lofty thoughts is not a prerequisite for the sublime in this context.

While Addison extends Longinus’ self-transcendent approach to the sublime, Burke develops Dennis’ notion of self-preservation (see Burke, E.). He bases his argument on the idea that pleasure and pain are logically independent of one another; pleasure is not simply the absence or diminution of pain, and vice versa. The reasoning behind this is as follows. Most of the time, we are in a state of ‘indifference’ (or ‘tranquillity’) where neither pleasure nor pain preponderates. Given this, it is clearly possible to move from a state of indifference to a state of pleasure without the mediation of pain, and vice versa. These unmediated states Burke terms ‘positive pleasure’ and ‘positive pain’ respectively. Having established this logical independence, Burke suggests that while the diminution or removal of pain and danger does not yield positive pleasure, it can be characterized as an agreeable state of ‘delight’. Hence, when terrible things excite ideas of pain or - from a position of safety - danger in our minds, the terror we feel is moderated into a state of delight. A similar state can be occasioned in us, indeed, when we behold vast or obscure items, which test or strain our perceptual faculties without pushing us over the threshold of conscious pain.

Interestingly, Burke explains both these routes to the sublime in terms of a common causal structure. Just as labour is necessary to keep the ‘grosser’ parts of the body healthy, so stimulation is required for the ‘finer’ parts upon which the mental powers act. Indeed, it is on the very occasions that we experience moderated terror or pre-conscious pain that the finer parts receive healthy stimulation - thus diminishing the possible danger that inactivity poses to the system. Burke’s causal argument is far from compelling, however, and towards the end of the eighteenth century Kant offered a more convincing account - an explanation orientated towards the cognitive.

3 Kant

Kant’s most thorough treatment of the sublime is found in his *The Critique of Judgment* (1790). It unites the self-transcendent and self-preservatory approaches in the context of a much broader and more complex theory concerning our ultimate human vocation as rational human beings (see Kant, I.). He distinguishes two fundamental varieties or modes of the sublime. First, the ‘mathematical’ mode arises through our perceptual engagements with vast objects. While his exposition of this is notoriously intricate and often obscure, his basic strategy can be described as follows. As rational beings, when we are presented with some perceptually overwhelming object, we strive to comprehend it - to find some measure whereby we can make its overwhelming aspect intelligible. The use of mathematical rules is the most obvious means to this end. (A towering mountain in the distance, for example, does not perhaps seem quite as towering if we can define it as ‘no more than *x* metres in height’.) However, all rules of measure presuppose units of measurement, and these are, in the final analysis, ‘aesthetic estimates’, that is, based on what we can grasp ‘by the eye alone’ in direct perception. It is at this point that Kant’s reasoning becomes particularly difficult. He seems to hold that, when faced with particularly impressive and vast objects, our search for an appropriate measure leads us towards the idea of infinity itself. We cannot, however, comprehend this notion in perceptual and imaginative terms. It defeats all our attempts to form an aesthetic estimate. The vast object therefore serves a mediating function. In our struggle to find a measure for it, the sense of being perceptually or imaginatively overwhelmed by the infinite illuminates the extraordinary scope of our rational powers. For we can, despite our sensible limitations, at least comprehend the infinite *as an idea*. The experience of the mathematical sublime therefore involves a ‘mental movement’ from privation to exhilaration at the superiority of our rational being over all sensible limitations.

The second mode of the sublime - the ‘dynamical’ - involves a similar mental movement, in this case instigated by our experience of mighty or dangerous objects or phenomena (from a position of safety). Again, Kant’s presentation of the argument is often awkward, but in contrast to the account of the mathematical sublime its logical structure is more straightforward. When beholding a mighty object or phenomenon, we know that in relation to our puny natural being the mighty item has absolute physical authority. It could destroy us many times over. However, as rational beings, we know that we are more than simply entities of physical nature. Life, destruction and death are not blind mechanistic forces; they are given meaning in the context of our rational endeavours. Hence while the mighty object indeed makes us fearful, we can, as rational beings, nevertheless evince a kind of moral resistance to it. In comprehending and articulating the danger, we are elevated above it.

Thus the object serves a mediating role. It provokes an awareness of our sensible limitations, which in turn vivifies the extraordinary scope of our rational being.

The importance of Kant's theory is that it coherently articulates the sublime as an aesthetic experience of 'the containment of excess'. In both the mathematical and dynamical modes an overwhelming item in the natural world is contained by our rational comprehension of it. The fact that this involves a direct interplay between our receptive sensible capacities and rational insight (rather than intellectual comprehension alone) is what gives it aesthetic character. By characterizing it in these terms, Kant is also able to link it (albeit problematically) to general criteria of aesthetic judgment involving, fundamentally, disinterestedness and universality.

4 Postmodern revival

The 1980s and 1990s saw the development of a widespread discussion of contemporary society in relation to the term 'postmodernism'. One of the most remarkable aspects of this development was a massive concomitant revival of interest in the concept of the sublime. J.-F. Lyotard, for example, has not only published a closely detailed account of Kant's position (1994) but also more speculative articles applying the theory. These extremely influential works centre on a modified concept of the Kantian sublime that is analysed in terms of two varieties or modes - on the one hand, a nostalgic, backward-looking mode, which yearns for the lost metaphysical absolutes of Romanticism, and, on the other, a forward-looking mode ('novatio') associated fundamentally with avant-garde art.

In this latter variety the difficulty or *outré* character of the particular work serves to represent the infinite developmental possibilities of artistic creativity. This suggestion of infinite potential is congruent with the ever-accelerating surge of technological transformation - a surge which seems especially manifest in the societal changes brought about by new information technologies. This congruence is taken by Lyotard to indicate an important continuity between avant-garde modern culture and new postmodern sensibility.

Lyotard's application of Kant is at best oracular, and at worst obscure. He seems to focus in particular on the idea that the sublime hinges on a 'mental movement' from a negative to a positive state. But what are the criteria for the ascription of such states? On this point, Lyotard takes us no further than Kant himself. An alternative approach is the one proposed by the author of this entry - again, in a detailed study of Kant and in a number of papers applying a Kantian-type theory. The basic strategy is to eliminate all talk of 'mental movement' and to recentre the theory on logical presuppositions. No matter how overwhelming or excessive an object may be in relation to our powers of perception or imagination, it is, *qua* phenomenal item, of *finite* extent and power. This means that it can be comprehended or contained in rational terms. The sublime object is one whose presence makes this intellectual insight emotionally vivid.

The sublime, then, has become fashionable once again. Raw nature was once the major source of this experience, but now the accelerating complexity and instability of society itself is also acknowledged as a basis of the spectacle. This broadening scope is further reflected in a burgeoning quantity of writing on the relation between sublimity, psychoanalysis, politics and literature. Slavoj Žižek (1988), for example, has considered the sublime in almost all these contexts (and more besides), and reads the experience on the basis of Lacanian psychoanalysis. Thomas Weiskel (1976) has attempted to reconstruct the Kantian approach in relation to the reading and content of texts, on the basis of both Freudian theory and structural linguistics (see Structuralism in linguistics). There is much in philosophical terms that is obscure and/or problematic about all these psychoanalytic strategies. However, it is significant that they draw on Kant's theory to greater or lesser degrees. His approach - suitably modified - seems to offer the best philosophical tool for understanding the basic experiential structure of sublimity that underlies its many historical and cultural changes of emphasis and context.

See also: Nature, aesthetic appreciation of §1

PAUL CROWTHER

References and further reading

Addison, J. (1712) *Critical Essays from The Spectator*, ed. D.F. Bond, Oxford: Clarendon Press, 1970.(Highly readable; see especially essays 409 and 411-21.)

Burke, E. (1757) *A Philosophical Inquiry to the Origin of Our Ideas of the Sublime and the Beautiful*, Oxford:

Blackwell, 1987.(A basic and straightforward text.)

Crowther, P. (1989) *The Kantian Sublime: From Morality to Art*, Oxford: Clarendon Press.(A close but accessible account of Kant's theory in the context of his general aesthetics and moral philosophy.)

Crowther, P. (1993) *Critical Aesthetics and Postmodernism*, Oxford: Clarendon Press.(Part 2 offers clear critical discussions of Burke, Kant and Lyotard.)

Dennis, J. (1704) *The Grounds of Criticism in Poetry*, London.(A basic, historically significant text.)

Kant, I. (1790) *The Critique of Judgment*, trans. J.C. Meredith, Oxford: Clarendon Press, 1973.(The essential text, albeit one of great interpretative complexity.)

Longinus (1st century AD) *On the Sublime*, trans. A.O. Pritchard, in *Criticism: The Rational Statement*, ed. Charles Kaplan, New York: St Martin's Press, 1975.(A clear and important text, addressing the sublime primarily in relation to literature.)

Lyotard, J.-F. (1994) *Lessons on the Analytic of the Sublime*, trans. E. Rottenberg, Stanford, CA: Stanford University Press.(A clear, sophisticated and extremely close reading of Kant's position in *The Critique of Judgment*.)

Lyotard, J.-F. (1984) *The Postmodern Condition: A Report on Knowledge*, trans. G. Bennington and B. Massumi, Manchester: Manchester University Press.(The appendix is a difficult, speculative piece linking the sublime to questions of postmodernism.)

Weiskel, T. (1976) *The Romantic Sublime: Studies in the Structure and Psychology of Transcendence*, Baltimore, MD: Johns Hopkins University Press.(A reasonably accessible discussion of the sublime in relation to literature.)

Žižek, S. (1988) *The Sublime Object of Ideology*, London: Verso.(An influential but difficult psychoanalytic exposition of the sublime.)

Substance

For Aristotle, 'substances' are the things which exist in their own right, both the logically ultimate subjects of predication and the ultimate objects of scientific inquiry. They are the unified material objects, as well as the natural stuffs, identifiable in sense-experience, each taken to be a member of a natural species with its 'form' and functional essence. Entities in other categories - qualities, actions, relations and so forth - are treated as dependent on, if not just abstracted aspects of, these independent realities.

With the rise of mechanistic physics in the seventeenth century, the Aristotelian multiplicity of substances was reduced to universal matter mechanically differentiated. This move sharpened the issue of the relation of mind to the physical world. The consequent variety of ways in which the notion of substance was manipulated by materialists, dualists, immaterialists and anti-dogmatists encouraged later scepticism about the distinction between independent realities and human abstractions, and so idealism.

Twentieth-century conceptualism, like some earlier versions of idealism, rejects the distinction altogether, commonly ascribing the logical priority of material things in natural language to the utility of a folk physics, as if they were the theoretical entities of everyday life. As such, their identity and existence are determined only through applications of a theory outdated by modern science. Yet this 'top-down', holistic philosophy of language is belied by the detailed insights of traditional logic, which point clearly to a 'bottom-up' account of classification and identity, that is an account which recognizes the possibility of perceptually picking out material objects prior to knowledge of their kind or nature, and of subsequently classifying them. The idea that material things are theoretical entities, and that their individuation is accordingly kind-dependent, is a hangover from an atomistic approach to perception which calls on theory to tie sensory information together. A more accurate understanding of sensation as the already integrated presentation of bodies in spatial relations to one another and to the perceiver is consonant with the possibility denied by the idealist- namely, that, with respect to its primitive referents, language and thought are shaped around reality itself, the independent objects given in active sense-experience. That the coherence or discrete unity of material objects has a physical explanation does not mean that physics explains it away.

1 The canonical concept of substance

The definitive concept of substance is Aristotle's, but it arose within an existing debate about what fundamentally 'is' in the world - meaning thereby not only what ultimately *exists*, but also what ultimately *is* such or such, as the subject of predication. The term employed, ουσια (being), was later translated in Aristotle's use of it as *substantia*, substance. Another Aristotelian term applicable to substance, ηποκειμενον, was more usually translated *substratum*, substrate.

Aristotle (see *Metaphysics* I) identifies two extremes. The first, typified by atomism, takes matter to be 'that of which all things that are consist, and from which they come, and into which they are finally resolved'. For atomism, all change occurs through the motion and rearrangement of immutable material atoms (see Atomism, ancient). Plato (see *Republic* V-VII), on the other hand, held the sensible world, because in perpetual flux, to be less real than the eternal universal characteristics or Forms fleetingly manifested in that world. It is the Forms which, as the subject-matter of enduring knowledge, are in the strong sense. Aristotle's middle way (see *Metaphysics* VI) takes unitary 'bodies', both living things and homogeneous stuffs, to be the fundamental objects of knowledge and subjects of predication. Such substances comprise two elements, particularizing matter and universal form. An individual horse is the specific form of a horse embodied in *this* matter. Forms are the objects of universal science - teleological principles of activity and change within what is material and individual.

The Aristotelian theory has two aspects, as it figures in logic on the one hand, and in the theory of scientific knowledge and explanation on the other. In the former role, substance is the first of the 'categories' or 'things said' (see *Categories*). Substances, falling under such predicates as 'horse' or 'wood', *are* in the primary and most fundamental sense. Other 'things said', sometimes called 'accidents', are somewhat arbitrarily listed under nine or ten heads. Such entities as a horse's shape, colour, location or neigh *are* in a derivative sense. They can be subjects of predication, but are themselves 'said of' and exist 'in' something else (in this case, the horse), whereas substances are neither 'said of' nor exist 'in' anything else. The brownness or neigh of a horse exists just in so far

as the horse is brown or neighing, but the primitive noun-predicate ’horse’ simply names the given individual, the horse itself. The horse is thus a logically independent entity, whereas its neighing or brownness is dependent on a subject or substrate. Aristotle called the individual horse the ’primary substance’, which cannot be ’said of’ anything, while the specific ’secondary substance’ *horse* can be ’said of’ (but does not exist ’in’) the individual, and the generic secondary substance *animal* can be ’said of’ both the individual and the species.

Only substances can be the subjects of contraries, that is endure through change, change explicable in terms of their natures or essences. Aristotle’s schema for scientific definition by *genus* and *difference*, later called the Doctrine of Predicables, applies primarily to substances. Non-substantial entities have definable essences in a sense, but Aristotle held, with reason, that only substances fall into natural species with genuine essences. *Properties* are attributes necessarily connected to essence, while *accidents* in the strict sense are those attributes of a substance which neither figure in, nor flow from, its definition. The Doctrine of Categories and the Doctrine of Predicables both encourage a question to which Scholastic Aristotelians gave a variety of answers: what is it for accidents to exist in a substance? The variety reflected the extent to which different philosophers were prepared to treat accidents as ’real’ (that is, really distinct, if naturally dependent) entities, as opposed to taking the view, which Epicurus had held, that they are merely conceptually distinct aspects of independent bodies. William of Ockham and Francisco Suarez wrote famous contributions to this debate (see Aristotle §§12-14; Epicureanism §§2-3; Plato §§2, 3, 8).

2 Early-modern heretical concepts of substance

Modern European philosophy began with a proliferation of variant conceptions of substance in an extended attempt to replace the dominant Aristotelian metaphysics and science with a new authoritative system. Later scepticism about the notion of substance arose with the rejection of this dogmatic enterprise. The very multiplicity of conceptions of substance may suggest that substance-based metaphysics was arbitrary and confused. Yet the various directions taken all fall within an intelligible framework of argument. The crucially productive move was the expulsion of teleological explanations from science through a reversion to something like the ancient atomists’ conception of matter as a substance possessing mechanical properties sufficient to explain all physical phenomena (see Teleology). Materialists such as Thomas Hobbes (1651) concluded that matter or body, determinately modified, is all there is (see Hobbes, T. §3). The variety of sensible qualities by which we distinguish things is a mere ’diversity of seeming’ consequent on differences in structure and motion. The activity of substances is simply the motion of bodies in accordance with necessary laws. Sensation and thought are themselves species of motion. More popular than materialism, however, in that religious age, was René Descartes’ dualism (see Descartes 1641), which postulated incorporeal thinking substance in addition to extended substance (see Descartes §8). A respectable philosophical motive for dualism was that it by-passed the insoluble problem of explaining consciousness, including ’seeming’ itself, in mechanical terms.

This new physics might seem far from the theory of predication, but mechanists saw themselves as explaining the logical relation between substance and accidents: accidents are simply the various and changing sizes, shapes and motions of bodies, limits or determinations of extension rather than Suárez’s distinct real entities mysteriously inherent in a subject. Descartes (1644) took it that, analogously, particular modes of thought are determinations of thought in general, and that the formal idea of a *subject* of extension or thought refers to nothing over and above these attributes. The connection with predication also figured in the system of Spinoza (1677), whose monism constitutes not only a holistic, deterministic philosophy of nature, but also a logical model according to which the ultimate subject of predication is the universe as a whole, of which individuals are modes or aspects (see Spinoza, B. de §§2-3). Leibniz (1686 and 1714), too, in arguing for his anti-materialist monadism, associated two principles: the claim that our awareness of ourselves thinking gives us our only intelligible paradigm of variety within a unitary, active substance, and the logical principle that in every true proposition the predicate is conceptually contained in the subject (see Leibniz, G.W. §§4-5). Partly by means of the latter principle he attempted to explain the identity and diversity of immaterial substances in a non-spatial, non-material world, and his failure supplies an invaluable philosophical lesson. There is, as Aristotle supposed, a demonstrable tie between concrete substantial existence and existence in space.

Logical theory and philosophy of science were differently linked by John Locke (1689), who rejected all claims to know essences and so to understand the substance - accident relation. Substances, including matter and mind, are

known to us only through a plurality of disparate observable qualities and powers, which we take to flow from some common cause. Hence we define the primitive noun-predicates which are the names of substances in terms which include a place-marker for that unknown essence: ‘When we speak of any sort of substance, we say it is a *thing* having such or such qualities, as body is a *thing* that is extended, figured, and capable of motion’ (1689, II. xxiii, 3). ‘Substance’ is just a name for the unknown. Modes, on the other hand, such as a *triangle*, *democracy* or a *triumph* are not mind-independent unknowns, but constructs reflecting human interests and containing nothing more than the features by which they are defined (see Locke, J. §§4-5). Locke was not denigrating the notion of substance, but attacking the pretensions of dogmatic theories. Scepticism was taken further by David Hume(1739-40), however, for whom the philosophical conception of substance was deconstructed as a product of deranged imagination (see Hume, D. §2). After Hume, Immanuel Kant attempted to reinstate the notion on a new basis, advancing the ‘transcendental’ argument that, for objective experience to be possible, it must be experience of a substance (that is, matter) undergoing change in accordance with law. The very ineluctability of the category, however, he took as proof that it does not correspond to anything in independent reality, but is generated by the interplay of sense and intellect, the forms of judgment and the forms of sensibility (see Kant, I. §§5, 6).

3 A new role for substance: realism versus conceptualism

Humean empiricism revived in the twentieth century. For A.J. Ayer, for example, the philosophical notion of substance reflects only a ’primitive superstition’ about names, since nothing is independently there to be named except sense-data, the equivalent of Hume’s impressions (see Ayer 1936). But much present-day philosophy, although influenced by both Hume and Kant, in general denies both that there is anything ineluctable about the way we think about the world, and that anything is simply there to be named. W.V. Quine (1953) expresses this denial in the terms of Russellian logic holding that to be is to be the value of a bound variable. For Quine, ontology is a matter of deciding which types of entity to postulate and quantify over - material objects, events, properties, ’stages’, ’fusions’, ’sense-data’ or what you will - a decision that is not determined for us by reality. The difference between types of entity is a difference between criteria of identity - between what we count as the same individual again (1960). Here a rational decision is simply a pragmatic one (see Quine, W.V. §5; Ontological Commitment). Material objects may fill the fundamental place in evolved natural language, but that is a matter of practical convenience or utility. Indeed, the folk-physics embodied in natural language has been supplanted, unless for everyday purposes, by a physics with no room for the ’bodies’ canonically assigned to the category of substance.

Is the traditional notion of substance therefore dispensable, the manifestation of an illusion engendered by contingent features of natural, everyday language? To suppose so would arguably be to set aside a storehouse of philosophical insight capable of suggesting to us just what can most cogently be brought against a conceptualism like Quine’s. Conceptualism is a ’top-down’ philosophy, taking the structure of our thought and language to be the structure of an imposed theory of interpretation of sensory input such that (contrary to Humean empiricism) even input itself must be conceived of in terms of the theory. Concepts are holistically inter-related, slicing out entities as theoretical postulates of the whole scheme. There are no independently given individuals. Not all conceptualists approve the analogy between natural language and scientific theory. Wittgenstein (1953), for example, invites us to see the structure of language as a product of its employment in, above all, social interaction. But all conceptualists have agreed, since the first modern idealists, that we cannot compare our scheme of notions or beliefs directly with reality. An equally old, still popular conclusion is that we can therefore only aim for coherence, and the total scheme that best anticipates experience.

The stiffest opposition to this aspect of pragmatism has come largely from two directions. Drawing on Kant, Peter Strawson (1958) has constructed a transcendental argument aimed at demonstrating that the distinction between experience of independent objects and merely subjective experience - in effect the distinction between oneself and other things which even the sceptic takes for granted - presupposes that both the individual objects and the individual subject of experience exist, endure, are mobile, and interact in space (see Strawson, P.F. §8; Transcendental Arguments). Whether or not this argument rebuts scepticism (would it not be enough for these distinctions that the subject should *seem to itself* to be so related to bodies in space?), it does offer an explanation of the primacy of the category of substance in natural language, and of the unreality of the suggestion that, but for utility, entities of a different category - events, thing-stages, sense-data or whatever, might have played the basic role. Yet although Strawson disavows an idealism like Kant’s, and presents a powerful alternative to Quine’s

relativism, he does not clearly depart from a ’top-down’, conceptualist picture. Indeed he has talked of his thesis as if it concerned the necessary structure of any conceptual scheme that could be the vehicle of objective experience - the general scheme or order which sensory input must fit, as it were, for there to be experience of objects. Individual substances are necessarily basic, but are picked out only by means of sortal concepts applied to input.

A different line of argument, pioneered by Saul Kripke (1980) and Hilary Putnam (1975), might be taken to refute the idea that our conceptual scheme structures the world for us, rather than being structured by the world. This proposes that, like proper names, names of natural kinds owe their meaning not to some concept in our heads but to their naming what they name. Consequently, ’gold’ on the lips of a present-day physical chemist means the same as an ancient Chinese character, whatever differences exist between the users’ conceptions of, or theories about, the natural kind of stuff that both name. That their words mean the same could not be determined by an inspection of their ’ideas’ or definitions, but would depend on the identity, that is the actual (and, for at least one of them, unknown) common nature, of what each names (see Reference §3). Yet even if this point is taken to undermine the analogy between language and theory, and to demonstrate the possibility of a pre-theoretical naming of things, it does not speak directly to the question of the distinction between what is naturally, and what is only conceptually distinct. For it could be made almost as readily about the general names of natural events, processes or properties as about the names of substances. The term ’freezing’ has the same pre-theoretical grip on a natural kind of process as ’horse’ has on a natural kind of substance, yet the freezing of some water in a glass is not a naturally discrete whole in the way that a horse or, for that matter, a glass is. Although no process is properly freezing unless it shares the common nature of freezing, unknown to most of us, yet a particular instance of freezing is individuated and bounded by the concept, not by nature. Every event or process is part of indefinitely many wider events or processes, and there are no natural, only conceptual, parts and wholes. But a horse is a natural whole, and its hooves and tail are natural parts. Because it has nothing directly to say about boundaries, unity, individuation and identity, the Kripke-Putnam thesis is compatible with conceptualism, as Putnam himself has claimed. David Wiggins (1980) has presented a ’conceptualist-realist’ theory of the identity of substances drawing on Kripke’s insights, but remains firmly within the conceptualist fold.

To rebut conceptualism, therefore, more is needed than a Kantian transcendental argument or the recognition of natural kinds in the Kripke-Putnam sense. It must be shown that reality itself contributes to the way we take the world to be articulated or divided, and this contribution must be distinguished from the contribution of the mind as receptor or interpreter or speculator - from the contingencies of sense-experience, language or theory. The canonical doctrine of substance took it that primitive logical subjects are apprehended as naturally unitary, naturally whole, naturally enduring individuals, prior to their classification by us. Such ‘bodies’ are standardly contrasted with merely notionally distinct and unitary individuals in other categories. Any philosophy of experience, identity and classification must either endorse such a distinction or rule it out: that is, must be either realist or conceptualist.

What reasons are there for adopting a realist, ‘bottom-up’ theory of identity and classification? A non-substantial entity such as an event is not determinately individuated except in virtue of some general concept, commonly a correspondent to nominalized predicate, adjectival or verbal. If standard conceptualist theories of identity were correct, the same would apply to substances. Only in virtue of a quasi-theoretical concept of an *insect*, or insect of a certain type, would it be possible to single out the individual which is in turn egg, larva, pupa and imago. Yet it seems evident that a material object can be picked out without our knowing what kind of thing it is, and that we can then learn from observation what metamorphoses such an object will normally undergo, or can survive. The possibility of refined classification comes with sustained observation of previously identified individuals. To classify substances is to group given individuals. Genera are groups of these groups. There is therefore no question of the individuals of the genus being different from the individuals of the species. This peculiarity of substances lies behind the Aristotelian principle that there are no ultimate species or individuals except in the category of substance. If a class of non-substances, say, *public assemblies*, is divided for some purpose into peaceful ones and violent ones so as to include, say, *demonstrations* and *riots*, then the class of riots is as good a species as the class of assemblies, and the individuals of these classes have different principles of individuation. An assembly exists while people are assembled, but a riot exists only as long as they riot. If an assembly is composed of a demonstration followed by a riot, the assembly, the demonstration and the riot are three distinct individuals. With substances, however, species can be divided without creating new species or individuals, as the species of human

beings is divided almost indefinitely by natural languages. Bakers, adolescents, albinos, diabetics and rioters constitute neither distinct species nor individuals distinct from human beings. Rioters do not cease to exist when they cease to riot, nor children when they grow up. This is not an accidental feature of natural language, modifiable by stipulation. To stipulate identity-conditions for an entity, (say) a *bakens*, such that a *bakens* exists just as long as a baker is a baker, is only to individuate a non-substantial entity such as a baker's practice of baking. The notion of a temporal part or 'stage' has played a central role in twentieth-century conceptualist theories of identity, but what comes to an end when a kitten becomes adult is its kittenhood, a part or stage of its life, not a temporal part of a cat - there is no such thing. 'Kitten' is a compound predicate satisfied by naturally unitary individuals whose identity through time consists in their continuing so united, not in their satisfaction of that or any other predicate. Things which satisfy primitive noun-predicates such as 'cat' do so throughout their existence, but that is because these predicates mark membership of more or less natural groups by origin, morphology and structure. The individuality of the members of the kind is prior to such classification. The unity upon which their continuity depends is the natural discreteness and coherence of a material object: materiality and the possibility of such natural unity go together. In contrast, individual non-substances are notional unities sliced out by general terms; which is why modification of the terms in question breeds new individuals, and new species.

Other considerations relating to identity confirm the contrast between the category of substance and other categories of entity. For example, just as we can literally place a material object in new circumstances, so we can consider what would have happened to it if circumstances had been different. The very same individual might have had very different attributes - that is implicit in our ordinary understanding of a material thing as possessing indefinitely many potentialities. An event or state of affairs or action could not in this way be supposed qualitatively different and yet numerically identical: a possible Battle of Hastings won by the Saxons cannot be the same event as the battle actually won by the Normans.

Realism's strength lies not only in its philosophy of language but in an epistemology which explains how substantial individuals can be 'given' prior to conceptualization. It has traditionally been supposed that the deliverances of the different senses present only qualities which have somehow to be tied together conceptually or theoretically in order that we should conceive of the things or substances that possess them. In the Second Mediation Descartes saw this process as the intellect's employing an innate idea of extended substance in order to think the reality behind appearances, but idealists and some empiricists have seen it as an act of quasi-theoretical construction (see Kant, I. §3; Phenomenalism). Yet the model itself is wrong. That we have several senses does not mean that they are not integrated in their operation, presenting a unified field of bodies in space oriented in relation to our own perceived and perceiving body (that is, ourselves). Indeed, the briefest reflection on the phenomenology of perception supplies instances of such integration not only of sense with sense, but of sense with action. That we can hear a sound close to our ear does not mean that we hear our ear, or infer a relation to our ear from a quality of the sound, but that hearing is integrated with bodily awareness. That prismatic spectacles can invert all the objects in the visual field equally demonstrates that the way we feel our body to be, which includes our feeling it to be upright, can frame the deliverances of another sense. Eye and hand are coordinated in normal subjects because sight, bodily awareness and touch present a single sense-field integrated with the field of action. Here the philosophy and empirical psychology of perception converge, allowing us to understand the possibility of our pre-theoretical, preconceptual apprehension of just those natural individuals, traditional 'substances' that are logically fundamental in natural language. Although such objects are not also, as Aristotle supposed, fundamental in natural science, their unity and structure are there for physics to explain. It is false, then, that the world as we experience it, conceive of it and act within it in ordinary life owes its articulation to the structure of language, theory or 'conceptual scheme'. Rather, thought is shaped by the experienced world, and language owes the possibility of its deepest features, reference and predication, to the senses' grasp of really discrete individuals. That is a lesson that traditional theory of substance can still help to teach us.

See also: Being; Continuants; Identity; Matter; Matter, Indian conceptions of; Ontology; Phenomenological movement; Realism and antirealism; Scientific realism and antirealism; Theories, scientific

MICHAEL AYERS

References and further reading

Aristotle (*c.* mid 4th century BC) Categories, in *The Complete Works of Aristotle*, (Revised Oxford Translation), ed. J. Barnes, Princeton, NJ: Princeton University Press, vol. 1, 1984.(Sets out the logical doctrine of

substance.)
Aristotle (*c*. mid 4th century BC) *Metaphysics*, in *The Complete Works of Aristotle*, (Revised Oxford Translation), vol. 2, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984.(Explores ways in which substance is fundamental. Book I (A) considers earlier views, and VII (Z) argues that unitary bodies, that is individual living things, are substances.)
Ayer, A.J. (1936) *Language, Truth and Logic*, London: Victor Gollancz.(Chapter 1, 'The Elimination of Metaphysics', argues that theories of substance are confusions due to language.)
Ayers, M.R. (1991) *Locke*, London: Routledge.(Volume 1, Part 3 contains a theory of integrated perception of objects. Volume 2, Part 1 finds insights in Locke's and other theories of substance; Part 3 expounds a realist theory of identity.)
Descartes, R. (1641) *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes* vol. 2, trans. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1985.(Classic argument for mind-body dualism.)
Descartes, R. (1644) *Principles of Philosophy*, Part 1, §§ 51-65, in *The Philosophical Writings of Descartes*, vol. 1, trans. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1985. (Descartes' account of substances and their relation to their principle attributes and modes.)
Epicurus (*c*.305 BC) Letter to Herodotus (from Diogenes Laertius, *Lives of the Philosophers*) in *Epicurus: The Extant Remains*, ed. C. Bailey, Oxford: Oxford University Press, 1926.(Argues that bodies, in particular atoms, are substances, and that accidents are not distinct entities.)
Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.(Chapter 34 sets out Hobbes' materialist conception of substance and accident.)
Hume, D. (1739-40) *A Treatise of Human Nature*, ed. L.A. Selby Bigge, revised P.H. Nidditch, Oxford: Oxford University Press, 1978.(Book I, part 1, §7; part 4, §§ 3 and 5 criticize the traditional concept of substance, and the substance-mode relation.)
Kant, I. (1781) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1963.(Argues that substance is an ineluctable category of judgment, imposed by the mind on what is given to sensibility.)
Kripke, S.A. (1980) *Naming and Necessity*, Oxford: Oxford University Press, 1980. (Rejects accounts of proper names which assign them sense or treat them as abbreviated descriptions, and extends a causal theory of meaning from this to other areas of language; revolutionary.)
Leibniz, G.W. (1686) *Discourse on Metaphysics, in Philosophical Essays*, trans. and ed. R. Ariew and D. Garber, Indianapolis, IN: Hackett Publishing Company, 1989. (Expounds a notion of individual substances, whose concepts contain all their predicates.)
Leibniz, G.W. (1714) *The Principles of Philosophy, or, the Monadology*, in *Philosophical Essays*, trans. and ed. R. Ariew and D. Garber, Indianapolis, IN: Hackett Publishing Company, 1989.(Identifies 'monads', simple immaterial individuals, as the only substances.)
Locke, J. (1689) *Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Oxford University Press, 1975.(See Book II, chapter 23, Book III, chapter 6, and following, for an interpretation of the idea of substance as place-marker for unknown essences or natures.)
Plato (*c*.380-367 BC) *The Republic*, in *The Dialogues of Plato*, ed. B. Jowett, Oxford: Oxford University Press, 1892.(Books V-VIII give an account of what *is*.)
Putnam, H. (1975) *Mind, Language and Reality*, in *Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press.(Contains a series of papers arguing that the meaning of natural-kind names lies in their reference to reality, not in definitions 'in the head'; revolutionary.)
Quine, W.V. (1953) 'On what there is', in *From a Logical Point of View*, New York: Harper & Row.(An early statement of Quine's view that ontological commitment is expressed through qualification.)
Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Classic exposition of contemporary pragmatist conceptualism, arguing that ontology is relative to language or conceptual scheme.)
Rorty, R. (1980) *Philosophy and the Mirror of Nature*, Oxford: Blackwell.(An extreme version of relativistic conceptualism, with sympathetic interpretations of many modern exponents; widely read and influential.)
Spinoza, B. de (1677) *Ethics*, in *The Collected Works of Spinoza*, vol. 1, ed. E. Curley, Princeton, NJ: Princeton University Press, 1985.(Makes thought and extension attributes of the one substance, of which all individual minds and bodies are modes.)
Strawson, P.F. (1958) *Individuals*, London: Methuen.(Argues that objective experience would be impossible without material things such as are basic objects of reference in natural language.)
Suárez, F. (1597) *Metaphysical Disputations*, Book VII, trans. C. Vollert, *On the Various Kinds of Distinction*, Milwaukee, WI: Marquette University Press, 1947. (Holds that accidents are 'modally distinct' and, although

naturally dependent, in principle separable from substances in which they inhere.)

Wiggins, D. (1980) *Sameness and Substance*, Oxford: Blackwell.(Advances a theory of identity consonant with Kripke’s insights, which attempts to combine conceptualism and realism.)

William of Ockham (1322-7) *Summa Logica*, Part 1, trans. M.J. Loux, *Ockham’s Theory of Terms*, Notre Dame, IN: University of Notre Dame Press, 1974.(Chapters 19-62 consider which categories of abstract terms stand for really distinct particular beings. Not many do.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell. (Wittgenstein’s late masterpiece, identifying the source of philosophical theories in misunderstandings of language and its structures, and the source of those structures in the roles of language in life.)

Suchon, Gabrielle (1631-1703)

Seventeenth-century rationalist feminist philosopher and author of two books dealing with many moral, social and political issues, Suchon advocated liberty, knowledge, authority and the possibility of an unmarried life for women. In this she did not employ any ready-made philosophy but created her own. Freedom and knowledge are everyone's natural rights, understood as rights that get their strength from themselves only and are to be observed everywhere alike.

Self-taught philosopher, author of a *Traité de la morale et de la politique (Treatise on morals and politics)* (1693) and *Du Célibat volontaire (Of voluntary single life)* (1700), Gabrielle Suchon was born at Semur in Burgundy, France. Forced into a convent, she later went to Rome and got a Papal writ releasing her from her vows. Her family took her to court and the Dijon Parliament ruled that she had to go back to her convent. Evading the sentence she spent the rest of her life reading and writing, 'abandoned of all creatures, but not of my Creator, who gave me the grace not to abandon myself' (1693: XV).

Suchon's *Traité* on morality and politics is valuable as a work of metaphysics, epistemology and theology. Advocating 'liberty, knowledge and authority' for women, the book provides inventive theories about, among other things, 'privation' (described as being productive of effects), natural rights, original sin and constraint. Suchon is not a relativist. Her method implies that looking at women's position with respect to knowledge, liberty, marriage and the like, provides a deeper insight into what those things are in themselves and helps better to judge of their exact values.

Like many others, Suchon was scandalized that schools and universities were closed to women while male 'idiots' had access to them. Yet she considered that personal reflection and use of the seeds of light given by God were most important of all, learning acquired by reading being second. As for knowledge derived from tutors, it is an enjoyable luxury but no more. Furthermore, the original sin was not the transgression of an interdict against knowledge of good and evil: God wanted humankind to know, and Suchon seems not to acknowledge any ban on curiosity. But Adam and Eve should have sought wisdom through their own intellectual effort, rather than seeking a ready-made answer by the facile method of eating an apple. Women can acquire a valuable understanding of moral, political and metaphysical questions by thinking for themselves. Reason is the name of that which enables us to be self-taught people. Although this applies to men and women alike, it can be best demonstrated by first considering women's position and then drawing from this an indication of how things stand in themselves, and so incidentally for men as well.

Knowledge and freedom are everyone's natural rights, natural right being a 'right that holds its strength from itself only and is to be observed everywhere alike' (1693: 23). It is not contingent upon laws or capricious custom. Knowledge and freedom both come into being by being exerted, but there is an interaction between the two. And knowledge is 'beatitude commenced'.

It is a knowledge of values that Suchon ultimately seeks. Why is freedom so good? What different kinds of liberties are most important? What is marriage but an oppression and a burden only a saint's sister could bear? Bad enough in itself, submission becomes so insufferable when one is forced to obey a person with no sense that it requires a special grace to endure the miseries involved. Again Suchon writes this with women's marital duties in mind. But in this second work, *Of voluntary single life*, she considers, as she did in her first treatise, that what is true for women must also to some extent be true for men. After all, male authors have harped on the many vexations of married life.

Suchon makes the female the universal, and her strategy is supported by a linguistic feature in French. The word 'person', which refers to men, women and components of the Holy Trinity alike, is grammatically feminine. It is therefore quite easy for her to have a universal 'she' and constantly to shift from statements concerning women to statements concerning 'persons' and vice versa. Her work is dedicated to the Trinity, again a feminine word in French and, obviously enough, a female power under Suchon's pen.

Nothing is known yet about how Suchon managed to educate herself. Through the authors she quotes, one may surmise what books she had access to: ancient philosophers, (though not all), countless Church Fathers, perhaps

the Scriptures themselves (not obviously a possession for a lay person in Louis XIV's France), some modern authors such as Petrarch, Pascal and Poulain de la Barre (a Cartesian who had published *De l'égalité des deux sexes (On Equality of Both Sexes)* in 1673). Her convent's library may well have been decently furnished, and she did make a point about using whatever little money one has to buy books. The reception of her work is not much known either. Perhaps it was not too bad. Her two books were reviewed in *Le Journal des Savants* in 1694 and 1700, the second one also in *Nouvelles de la République des Lettres*. Long after her death, an entry 'Gabrielle Suchon' appeared in Papillon's *Bibliothèque des auteurs de Bourgogne* (1745).

MICHÈLE LE DOEUFF

List of works

Suchon, G. (1693) *Traité de la morale et de la politique (Treatise on morals and politics)*, Lyon: B. Vignieu; selections of part I, ed. S. Auffret, repr. Paris: Des Femmes, 1988.(Although now deprived of freedom, knowledge and authority, women have a natural capacity for and claim to them. Originally published under the pen name of G.S. Aristophile.)

Suchon, G. (1700) *Du Célibat volontaire ou la vie sans engagement (Of voluntary single life)*, Paris: J.& M. Guignard, 2 vols; selected chaps, ed. S. Auffret, repr. Paris: Indigo & Côté-femmes, 1994.(There are now two positions only for a woman, namely matrimony or the cloister. There should be a third one, lay single life, for women who do not want to submit to a husband's authority nor to a monastic rule. Containing sharp criticisms of matrimony and the way families trade their daughters away to the cloister, the book provides a utopian vision of single women's lives.)

References and further reading

Albistur, M. and Armogathe, D. (1977) *Histoire du Féminisme Français (A history of French Feminism)*, vol. 1, Paris: Des Femmes.(A very brief survey of Suchon's life and work.)

Barre, P. de la (1673) *De l'égalité des deux sexes*, Paris: Fayard, 1984; trans. A.L. as *The Woman as good as the man, or, the Equality of both sexes*, London, 1677; repr. ed. G.M. McLean, Detroit, MI: Wayne State University Press, 1988.(When considered independently of custom, both sexes are utterly equal; the matter, disregarding great men's authority, must be discussed from the point of view of reason and common sense only.)

Bertolini, S. (1977) *Gabrielle Suchon: une Écrivaine engagée pour une vie sans engagement (A writer committed to a life without commitment)*, unpublished dissertation, Geneva, Faculté des Lettres.(A scholarly essay on Suchon's *Of voluntary single life*, containing findings about her life and family background.)

Geffriaud-Rosso, J. (1988) 'Gabrielle Suchon: une troisième voie pour la femme?', in U. Döring (ed.) *Ouverture et Dialogue*, Tübingen: Gunther Narr Verlag, 669-78. (Appraisal of Suchon's views on single life.)

Hoffman, P. (1978) 'Le Féminisme spirituel de Gabrielle Suchon', *XVIIe Siècle* 121: 269-76.(Suchon regarded as a masochist follower of Poulain de la Barre.)

Papillon (1745) *Bibliothèque des auteurs de Bourgogne (Library of Bourgogne authors)*, Dijon: F. Desventes; repr. Geneva: Slatline, 1970.(Papillon, who had met Suchon, is a main source about her life. Entry contains some remarks about the reception of her books.)

Ronzeaud, P. (1975) 'La femme au pouvoir ou le monde à l'envers', *XVIIe Siècle* 108: 9-33.(Views Suchon's way of thinking as not theoretical but as a spiritual quest for a feminine identity.)

Suffering

Although sometimes identified with pain, suffering is better understood as a highly unpleasant emotional state associated with considerable pain or distress. Whether and how much one suffers can vary in accordance with any meaning attached to the associated pain or distress, or with expectations regarding the future. Because suffering can be affected by thoughts of meaning or of the future, some have focused on this dimension of suffering and asserted that only humans can suffer. But there is a very strong empirical case that many nonhuman animals suffer. The fact of suffering provokes moral concern, especially when suffering is caused unnecessarily, and raises ethical questions, mainly regarding the nature and extent of our obligations to those who suffer. Suffering is also an important source of personal or religious meaning in many people's lives.

1 Concept

What is suffering? Casual usage sometimes suggests that suffering is identical to *pain*, but surely the pain caused by an ordinary pinch on the arm is too mild to count as suffering. Moreover, pain, but not suffering, can be located in specific body parts. While pain and suffering differ, it is possible to overstate their differences. On the *sensation model*, pain is simply a kind of sensation, which varies in intensity, duration, location, and features that permit classifying it as a particular kind of pain (such as an ache or twinge). On this model, pain need not be at all unpleasant: we may or may not suffer even when we are in great pain, since pain does not necessarily involve an affective dimension. On the *attitude model*, however, pain is any sensation (or perhaps any feeling) we dislike for its own felt qualities, suggesting less conceptual distance between pain and suffering. Each model of pain faces certain theoretical challenges. It seems safe to say, however, that pain is at least typically (whether or not intrinsically) unpleasant, in which case at least most painful experiences involve some degree of affect - a feature shared with suffering, as we will see (see Bodily sensations).

Suffering also bears a close relationship to *distress* but cannot be identified with it. The mild distress of a professor who is late to class need not involve suffering, even if someone having an anxiety attack (a kind of distress) clearly does suffer. Very roughly, distress is an emotional state that can be caused by, or take the form of, various more specific mental states such as fear, anxiety and discomfort.

While any precise analysis of 'suffering' will be controversial, suffering may be understood roughly as *a highly unpleasant emotional state associated with considerable pain or distress*. The words 'associated with' bypass the issue of whether considerable pain or distress *causes* suffering or *is a form of* (or conceptually overlaps with) suffering. Perhaps terror is a form of suffering, not a mere cause. Perhaps excruciating pain in its affective dimension (as opposed to its bodily location) is suffering. One's judgments here will depend on one's specific analyses of these mental states.

Whether and how much one suffers can vary in accordance with attitudes or expectations about the associated pain or distress or about the context in which it occurs. Even the mild pain of a common headache can lead to great distress and suffering if the pain endures with no end in sight, or if the subject believes the headache is a sign of impending physical deterioration. On the other hand, soldiers have sometimes received major injuries yet apparently not suffered much, due, for instance, to the relief of expecting removal from combat, or to positive attitudes to the heroic context of their injury. Long-distance runners who experience pain and discomfort in a race may or may not suffer, or may suffer more or less, depending on such psychological factors as how they evaluate their efforts, and whether they are relaxed or fearful in their attitudes regarding the remainder of the course. Thus any meaning one attaches to one's situation, as well as one's expectations for the future, are important factors in whether and how much one suffers.

2 Subjects of suffering

Who can suffer? Because meaning and expectations for the future are important factors in human suffering, and because nonhuman animals are often thought incapable of assigning meaning and anticipating the future, it is sometimes asserted that only humans can suffer (see Moral standing §2).

This assertion is highly doubtful, however. First, to say that meaning and expectations for the future are important factors in human suffering is not to say either is a necessary condition of suffering. It seems plausible that pouring

scalding water on a human baby would cause suffering, even if the baby could assign no more meaning to the event than a dog could. And even if expectations for the future were necessary for suffering, the thesis that only humans can suffer would depend on the premise that animals have no sense of the future. But it is highly questionable whether the behaviour of ‘higher’ animals, as studied in cognitive ethology, can be adequately understood without attributing at least minimal awareness of time to them.

Because suffering is an emotional state, only sentient beings can possibly suffer. Behavioural evidence, physiological data, and evolutionary considerations together make a strong case that many animals - possibly including most or all vertebrates - are sentient. But there may be sentient animals who cannot suffer, either because they can experience only minimal pain and distress or because they are incapable of highly unpleasant emotional states.

3 Ethical importance

It is generally agreed that suffering tends to make one less well off, due to its intrinsic aversiveness and its instrumental disvalue, the way it interferes with our doing and getting things we want. Suffering is a harm.

Suffering therefore provokes moral concern and raises ethical issues. It often isolates the subject, posing special challenges, for example, to medical professionals and families dealing with infirm elderly persons. When suffering results from human action, it clearly becomes an ethical matter, and virtually every ethical theory or tradition includes some injunction against unnecessary causing of suffering (see Evil). But whose suffering counts? While humans uncontroversially fall under the scope of such an injunction, it is increasingly believed that sentient animals do as well (see Animals and ethics).

Points of controversy include the following. First, when is it ‘necessary’ - that is, justified - to cause suffering? For example, do the products of factory farming justify the suffering caused to the animals? Second, do we have positive obligations to those who suffer and, if so, to what extent? What are our obligations, if any, to famine victims, politically persecuted individuals, and women sold into spousal slavery? Third, what affective capacities and moral virtues are required for appropriate responses to suffering? Can one respond well to a grieving associate without vividly grasping their experience? Could an appropriate response be described without mentioning virtues, such as compassion and sensitivity?

4 Personal and religious meaning

Although a harm, suffering is for many people a source of personal or religious meaning. Those who have suffered extensively often report that their experience has enabled them to appreciate what is truly important in life as opposed to what is superficial and fleeting. Someone who has endured cancer may say that they are now less concerned with personal accolades and more involved in their children’s lives. The experience of suffering sometimes increases one’s sensitivity to others’ suffering. Someone evicted unexpectedly from their apartment and rendered temporarily homeless may better understand and care about the plight of the chronically homeless. It is also often asserted that suffering builds character. A confident individual for whom things always came easily may be devastated by a spouse’s death but, in the aftermath, begin to cultivate courage, perseverance, and humility. Moreover, when suffering provokes a compassionate response from others, this may strengthen the subject’s faith in their fellows. (Sometimes such positive experiences of meaning help to alleviate present suffering (see §1).)

Religious individuals often understand suffering to have a positive aspect or some other important meaning. Priests and monks of various religions may embrace self-denial, the suffering that comes with it, and the virtues such suffering encourages. Christians understand suffering to permit greater identification with Jesus, who is believed to have suffered for the benefit of all, providing a model of virtuous sacrifice. According to Buddhism, suffering results from desire or craving - whether for sensual pleasure, wealth, or even personal ideals - and all desire involves the idea of oneself as separate from other things. Suffering is eliminated by extinguishing desire, through right living, meditation, and the eventual achievement of *nirvāna*, a state of peace and insight in which one abandons the notion of a separate self (see Asceticism).

See also: Evil, Problem of; Good, theories of the; Hedonism; Pleasure; Suffering, Buddhist views of origination of; Vulnerability and finitude

DAVID DeGRAZIA

References and further reading

Cassell, E. (1991) *The Nature of Suffering and the Goals of Medicine*, New York: Oxford University Press.(An insightful and well-informed discussion of suffering and its relation to the goals of medicine.)

DeGrazia, D. (1996) *Taking Animals Seriously: Mental Life and Moral Status*, Cambridge: Cambridge University Press.(A philosophical examination of the mental life of nonhuman animals, including their suffering, and of basic ethical issues regarding animals.)

Melzack, R. and Wall, P.D. (1983) *The Challenge of Pain*, New York: Basic Books.(A thoughtful scientific investigation of the physiology, psychology and control of pain.)

Pitcher, G. (1970) 'The Awfulness of Pain', *Journal of Philosophy* 68: 481-92.(An excellent examination of the strengths and weaknesses of different models of pain.)

Rahula, W. (1959) *What the Buddha Taught*, New York: Grove Press; revised edns, 1974, 1987.(A relatively clear, though not philosophically precise, introduction to the essential ideas of Buddhism, including the Buddhist understanding of suffering.)

Rose, M. and Adams, D. (1989) 'Evidence for Pain and Suffering in Other Animals', in G. Langley (ed.) *Animal Experimentation: The Consensus Changes*, 42-71, New York: Chapman and Hall.(An excellent summary of empirical evidence for pain and suffering in nonhuman animals.)

Sumner, W.L. (1992) 'Welfare, Happiness, and Pleasure', *Utilitas* 4: 199-223.(An exceptionally clear examination of competing models of some mental states of interest to utilitarianism, including pain and suffering.)

Suffering, Buddhist views of origination of

The Sanskrit term pratītyasamutpāda *(Pāli,* paṭiccasamuppāda*) literally translates as 'arising [of a thing] after encountering [its causes and conditions]'. This term, conventionally translated as 'dependent origination', 'conditioned co-arising' or 'interdependent arising', signifies the Buddhist doctrine of causality. This doctrine is usually applied to explain the origin of suffering (*duḥkha*) as well as the means of liberation from it. According to the Buddhist tradition, the Buddha discovered the law of dependent origination during his meditation on the night he attained his awakening. According to traditional accounts, he saw all his former lives and the lives of all other beings, understood the principle governing transmigration, and found the way of liberation. He then formulated the so-called Four Noble Truths, the Eightfold Noble Path and the Law of Dependent Origination. The twelve elements of the chain of dependent origination were designed to explain the mechanism of entanglement of a sentient being in a wheel of consecutive lives, and, at the same time, to explain how this entanglement is possible without admitting the concept of a permanent principle, like 'self', 'ego', and the like. These twelve members are: (1) ignorance, (2) formations (volitional dispositions), (3) consciousness, (4) name and form, (5) six bases of cognition, (6) contact, (7) feeling, (8) desire, (9) attachment, (10) existence, (11) rebirth, (12) ageing and death. In addition to the twelvefold formula, there is also the so-called 'general formula' of dependent origination, which goes 'when this is, that arises; when this is not, that does not arise.'*

1 Canonical accounts

The Buddhist tradition has it that Prince Siddhārtha Gautama, after a prolonged meditation under the bodhi tree, attained enlightenment or awakening (*bodhi*), the ultimate understanding of the Four Noble Truths: (1) there is suffering, (2) there is a cause of the arising of suffering, (3) there is an extinction of suffering, and (4) there is a means to come to the end of suffering. The means is to follow the so-called middle path between the extremes of self-indulgence and self-mortification. This fundamental teaching was expounded in the Buddha's first sermon, the so-called 'Sermon of the Turning of the Wheel of the Law' (*Dhammacakkapavattana Sutta*) (see Buddha).

The second of the Noble Truths, the Noble Truth of the Arising of Suffering, is often explained in terms of the twelvefold law of dependent origination. The twelve members of the chain of dependent origination are: (1) ignorance, (2) formations (volitional dispositions), (3) consciousness, (4) name and form (body and mentality), (5) six bases of cognition (sense faculties), (6) contact (of sense faculty with sensible object), (7) (physical and mental) feeling, (8) desire, (9) attachment, (10) existence, (11) birth, (12) old age and death. Each member of this sequence is conditioned by what precedes it. The doctrine is commonly regarded as the fundamental teaching of the Buddha, together with his teaching on the Four Noble Truths and the Eightfold Noble Path. An often-quoted canonical passage says that 'He who sees dependent origination sees the Doctrine (*Dharma*; Pāli, *Dhamma*); he who sees the Doctrine sees dependent origination' (*Majjhima Nikāya* 1.190-1). Elsewhere it is said that one who sees the Doctrine sees the Buddha. Thus, to see the principle of dependent origination is equivalent to seeing the Buddha. From the very beginning, this doctrine was regarded as the most profound and difficult teaching of the Buddha.

Some traditional texts declare that the law of dependent origination exists independently of whether Tathāgatas, those who know the truth, appear in the world to discover it. According to these texts, the law of dependent origination is rediscovered each time by the present *buddha*, as an old deserted city hidden in the jungle is discovered by a traveller. Śāriputra, one the most prominent disciples of the Buddha, was converted to his teachings by means of a single stanza uttered by Aśvajit (Assaji), one of the five first disciples: 'All things are born of causes, and of these the Tathāgata has proclaimed the cause, and their extinction too: thus does the Great Ascetic speak' (see Takasaki 1987: 72). This stanza summarizes the law of dependent origination, the very core of the Buddha's teaching on the causal character of the phenomena of existence. It later became regarded as a kind of Buddhist credo.

Research by Lamotte (1980) has shown that the Buddhist sources are not unanimous as to the moment of the discovery of dependent origination by the Buddha and its place within his doctrine. According to some texts, the Buddha acquired a full comprehension of the twelve-membered chain of dependent origination before the attainment of enlightenment, while other texts place the event during or after the enlightenment.

The notion of suffering (*duḥkha*) has in Buddhism an existential dimension and extends over all aspects of human life in this world. Everything existing is produced from causes, is momentary and devoid of a 'self' (that is, of any permanent unchangeable principle), and therefore is painful. Since the Buddhist doctrine denies any permanent principle like a soul, self or ego that transmigrates by virtue of previously committed actions from one life to another, it is admitted that a 'human being', a 'person' is but a conventional designation of a conglomerate of various factors, which appear in a stream of five groups of psychical and physical existential phenomena, and which are mutually conditioned and regulated by the law of dependent origination. A full insight into the twelve members of the chain of dependent origination, and into the extinction of consecutive members, brings realization of the ultimate goal of Buddhism, attainment of buddhahood.

According to Frauwallner's hypothesis, it is possible to demonstrate that the chain of dependent origination has a composite character; the twelvefold chain could have been compiled by the Buddha himself from two shorter causal chains. Also, the so-called *Pratītyasamutpādasūtra*, a fundamental canonical exposition of the doctrine, is to be regarded as a later compilation (Frauwallner 1956). This text was the object of a lengthy commentary by the fifth-century Buddhist philosopher Vasubandhu.

One can also show that beside the twelvefold chain of dependent origination there existed in the canonical texts other causal sequences, in which are clearly visible the components known to us from the former sequence. The final formulation of the doctrine seems to have been completed only after a long process of compilation which may have lasted until the great schism within the Buddhist order, perhaps shortly before Aśoka (middle of the third century BC).

2 Scholastic interpretations

The doctrine of dependent origination has been variously modified and interpreted within the schools of Buddhism. Usually it is applied to explain an individual human life in three time periods: past, present and future. Vasubandhu's *Abhidharmakośa (Treasury of the Doctrine)* (3.20-) presents several classifications of the twelve members of dependent origination (see §1). Vasubandhu distributes the twelve items into three periods of time: (1) and (2) belong to the past, (3) to (10) to the present, and (11) and (12) to the future. Alternatively, there is a twofold division of the members, whereby (1) to (7) belong to past existence, and the remainder to future existence. From the point of view of karma, (1), (8) and (9) are regarded as defilements that predispose one to karma, (2) and (10) are regarded as actions (karma itself), and the remaining seven members are regarded as supports of the defilements and actions.

The Pāli canonical *Mahātaṇhāsaōkhayasutta (Great Discourse on Destruction of Desire)* juxtaposes the theory of dependent origination and another explication of entanglement in the painful world, namely the theory of *gandhabba* (Sanskrit, *gandharva*), an immaterial being that acts as a bridge for human beings from one incarnation to another. The text describes the consecutive stages of conception, embryonic development, birth, growing up and sensual contact with the outer world, which result in origination of the whole mass of suffering. Conception takes place when mother and father are sexually united, when the mother is ovulating and when a *gandhabba* is present. When the child is born, its senses mature, and it amuses itself with the pleasures of the senses. It becomes attached to pleasant experiences and vexed by unpleasant ones. This delight in feelings is attachment; conditioned by attachment, there is desire for further becoming; conditioning by this desire, there is birth; conditioned by birth, there is old age and death, and also sorrow, lamentation, pain, grief and despair. Such is the origin of this whole aggregate of suffering, without any permanent 'soul' or 'self' or any other unchangeable principle.

The fusion of the theory of an intermediate immaterial being and the theory of the twelve links of dependent origination is mentioned in Vasubandhu's *Abhidharmakośa* (ad 3.15a-b; 3.19). His *Abhidharmakośa* (ad 3.28ab) contains the Sarvāstivāda interpretation of dependent origination, which is outlined in as follows. A fool who does not comprehend that the world is nothing more than a collection of mutually dependent phenomena develops a false notion of a self. He performs bodily, vocal and intellectual actions with the intention of experiencing pleasure. He performs beneficial deeds in the hopes of securing a better future life. He may perform harmful deeds in order to secure short-term gains. In this way he accumulates karma conditioned by ignorance. His stream of consciousness, forced by his previous karma, goes from one destiny to another in the same way that a flame 'moves'; in reality there is no motion, but merely the arising of contiguous moments, each caused by its predecessors. At death of the body, the predisposition to survive continues, so the stream of consciousness seeks

out a new body in which to dwell. In this way karma gives rise to new consciousness. When it finds a new body, consciousness conditions the acquisition of body and mentality. When these mature, sense faculties arise. These come into contact with sensible objects, giving rise to pleasant or unpleasant feelings. From these feelings there arises craving. Pleasant experiences give rise to a craving for pleasure to continue, while unpleasant experiences give rise to a craving for discomfort to end. These cravings are the condition for the development of clinging to false views and belief in (or hope for) an enduring self. This clinging conditions desires for future existence, as a result of which there is birth. From birth there inevitably follows ageing and death (see Mejor 1991: 96-7).

3 Mādhyamika and Yogācāra

With Nāgārjuna (second century), the founder of the Mādhyamika school, the theorem of dependent origination acquired a central position. The full comprehension of dependent origination is tantamount to comprehension of the Four Noble Truths: 'Whoever comprehends dependent origination also comprehends suffering, its arising, its ceasing and the path' (*Mūlamadhyamakakārikā(Fundamental Treatise on the Middle Path)* 24.40). The ultimate elements of being arise from causes. They are mutually conditioned, are dependent on others, and thus, devoid of any kind of self-nature, they have no real existence of their own. Whatever exists that has originated dependently is said to be empty (*śūnya*). Therefore, dependent origination is tantamount to emptiness (*śūnyatā*) (see Buddhist concept of emptiness). All names are but conventional designations, for ultimately they refer to no real existents. Discrimination of the two truths, conventional and ultimate, shows the equality of the emptiness and dependent origination of the elements of being. Nāgārjuna says: 'We say that dependent origination is emptiness. That is dependence upon convention. That itself is the middle path' (*Mūlamadhyamakakārikā* 24.18).

In the Yogācāra school, the functioning of twelvefold dependent origination is applied to the transformation of consciousness. According to Vasubandhu's exposition of the doctrine of 'ideation-only' (*vijñaptimātra*), all distinctions between subject and object are due to the threefold transformation of consciousness, which itself depends on various causes. Thus, the transformation of consciousness is but an ideation, a false imagining of subject and object, and all this is a mere representation of consciousness. As the sixth-century Yogācārin philosopher Sthiramati states in his commentary on Vasubandhu's *Triṃśikā(Thirty Stanzas)*:

> No cognizable object exists in reality, because it is in its nature wrongly imagined. Further, consciousness does not exist substantially, because it is dependently originated. So it is assumed. Further, the dependent origination of consciousness is known under the term 'transformation'.
>
> (commentary on stanza 1; see Vasubandhu, ed. Levi 1925)

In other words, the process whereby the elements of being arise and accumulate is a process of obscuration or defilement of thought. And this is dependent origination in its ascending order; that is, beginning with consideration of ignorance, one realizes that it conditions volitions, and so forth through the cycle of conditions to birth and its inevitable consequences of ageing and death. The reverse process of purification of thought is dependent origination in descending order, leading to final liberation; that is, realizing that there would be no ageing and death without birth, no birth without the desire to continue being, and ultimately none of this without ignorance, one knows that one must eliminate the ignorance that takes the form of imposing a distinction between subject and object upon experience as a whole (see La Vallée Poussin 1913: 65).

In the Perfection of Wisdom (*Prajñāpāramitā*) literature, the contemplation of dependent origination is included in the practice of a *bodhisattva* who aspires to perfection of wisdom. In this literature, it said, for example, that the *bodhisattva* should aspire to realize that no phenomenon whatsoever is produced without a cause. On seeing this, the *bodhisattva* will know that nothing is permanent, and nothing is spared eventual extinction. On seeing this in turn, he will see no item of experience as a self, a being, a soul, a creature, a human being, a person, a youth, a personality, an agent, one with feelings, one who knows or one who sees. If the *bodhisattva* does not see experience in these terms, he will not see experience as either permanent or impermanent, as pleasant or unpleasant, or as 'self' or 'other' (see Conze 1984: 491).

See also: Buddhism, Mādhyamika: India and Tibet; Buddhism, Yogācāra school of; Buddhist philosophy, Indian; Buddhist philosophy, Chinese §13; Nirvāṇa

MAREK MEJOR

References and further reading

Bernhard, F. (1969) 'Zur Interpretation der *Pratītyasamutpāda*-Formel' (On the Interpretation of the Pratītyasamutpāda Formula), *Wiener Zeitschrift für die Kunde Südasiens* 12/13: 53-63.(Supports with textual evidence Frauwallner's hypothesis on the original form of the formula of dependent origination.)

Conze, E. (1984) *The Large Sutra on Perfect Wisdom with the Divisions of the Abhisamayālaōkāra*, Berkeley, CA: University of California Press.(English translation of a Mahāyāna text.)

De Jong, J.W. (1974) 'A propos de Nidānasaṃyukta', in *Mélanges de Sinologie offerts à Monsieur Paul Demiéville*, Paris, vol. 2: 137-49; repr. in De Jong, J.W., *Buddhist Studies*, ed. G. Schopen, Berkeley, CA: University of California Press, 1979.(An important survey of Buddhist texts dealing with the *pratītyasamutpāda*.)

Frauwallner, E. (trans.) (1956) *Die Philosophie des Buddhismus*, ed. W. Ruben, Berlin: Akademie-Verlag, 4th edn, 1994.(An anthology of Buddhist philosophical texts, translated and presented by one of the best scholars of the subject.)

Gupta, R. (1977) '"Twelve-Membered Dependent Origination": An Attempted Reappraisal', *Journal of Indian Philosophy* 5: 163-86.(An attempt towards interpretation of the meaning of the twelve-membered formula of dependent origination.)

Kalupahana, D.J. (1975) *Causality: The Central Philosophy of Buddhism*, Honolulu, HI: University Press of Hawaii.(A thorough study of the Buddhist concept of causality in the context of non-Buddhist doctrines, with a chapter devoted to the exposition of dependent origination based mostly on Pāli sources.)

Kalupahana, D.J. (1987) 'Pratītya-samutpāda', in M. Eliade (ed.) *Encyclopedia of Religion*, New York and London: Macmillan, vol. 11, 484-8.(The author singles out the *Kātyāyana-avavāda* from the Pāli *Saṃyutta Nikāya* (2.16-17) as a text 'which serves as the *locus classicus* of all subsequent interpretations of the Buddha's "middle path".')

Kochumuttom, T.A. (1982) *A Buddhist Doctrine of Experience: A New Translation and Interpretation of the Works of Vasubandhu the Yogācārin*, Delhi: Motilal Banarsidass.(A study of Vasubandhu's works representing the 'ideation-only' doctrine; contains discussions and translations of numerous fragments.)

La Vallée Poussin, L. de (1913) *Théorie des douze causes*, Receuil de travaux publiés par la Faculté de Philosophies et Lettres de l'Université de Gand, fasc. 40, Gand: Librairie Scientifique E. Van Goethem.(A classical study by one of the greatest masters of Buddhist studies. Contains original texts in Sanskrit and Tibetan.)

Lamotte, E. (1980) 'Conditioned Co-production and Supreme Enlightenment', in S. Balasooriya, A. Bareau, R. Gombrich, S. Gunasingha, U. Mallawarachchi and E. Perry (eds) *Buddhist Studies in Honour of Walpola Rahula*, London: Gordon Fraser, 118-32.(An attempt to classify Buddhist records on the Buddha's discovery of the law of dependent origination.)

Law, B.C. (1937) 'Formulation of *Pratītyasamutpāda*', *Journal of the Royal Asiatic Society* (April): 287-92.(Somewhat dated but still useful paper introducing the formula of dependent origination from Pāli sources.)

Mejor, M. (1991) *Vasubandhu's Abhidharmakośa and the Commentaries Preserved in the Tanjur*, Alt- und Neu-Indische Studien 42, Stuttgart: Franz Steiner Verlag.(Contains useful information on Vasubandhu and his works, and a comparative description of the commentaries in the Tanjur, with special reference to the formula of dependent origination.)

Muroji, Y. (1993) *Vasubandhus Interpretation des Pratītyasamutpāda. Eine kritische Bearbeitung der Pratītyasamutpādavyākhyā(Saṃskāra- und Vijñānavibhaōga)*, Alt- und Neu-Indische Studien 43, Stuttgart: Franz Steiner Verlag.(Two chapters from Vasubanhu's commentary on the basic text of *pratītyasamutpāda* doctrine.)

Murti, T.R.V. (1955) *The Central Philosophy of Buddhism: A Study of the Mādhyamika System*, London: Unwin, 2nd edn, 1980.(A classic study of the Madhyamaka system of thought; contains discussion of the interpretation of the *pratītyasamutpāda* doctrine.)

Nāgārjuna (*c.* AD 150-200) *Mūlamadhyamakakārikā(Fundamental Treatise on the Middle Path)*, trans. D.J. Kalupahana, *Mūlamadhyamakakārikāof Nāgārjuna. The Philosophy of the Middle Way. Introduction, Sanskrit Text, English Translation and Annotations*, Delhi: Motilal Banarsidass, 1991.(Useful volume, although some of the translator's interpretations are questionable.)

Schlingloff, D. (1988) 'The Buddhist Wheel of Existence', in *Studies in the Ajanta Paintings: Identifications and Interpretations*, Delhi: Ajanta Publications, 167-74, 383-4.(New interpretation of a famous Ajanta painting in cave 17, which represents the twelve members of the law of dependent origination.)

Takasaki, J. (1987) *An Introduction to Buddhism*, trans. R.W. Geibel, Tokyo: The TōhōGakkai.(An excellent, lucid exposition of Buddhist doctrine, based mostly on Vasubandhu's *Abhidharmakośa*.)

Tucci, G. (1930) 'A Fragment from the *Pratītya-samutpāda-vyàāhyā* of Vasubandhu', *Journal of the Royal Asiatic Society* 611-23; repr. in *Opera Minora*, Rome, 1971, Part 1, 239-48.(Sanskrit fragments of Vasubandhu's commentary from a Nepalese manuscript.)

Vasubandhu (5th century) *Abhidharmakośa (Treasury of the Doctrine)*, ed. and trans. L. de la Vallée Poussin, *L'Abhidharmakośa de Vasubandhu, trad. et annoté*, Paris/Louvain: Paul Geuthner, 1923-31, 6 vols; new edn ed. É. Lamotte, Mélanges chinois et bouddhiques 16, Brussels: Institut belge des hautes études chinoises, 1980, 6 vols.(This 'Treasury of Doctrine' by Vasubandhu, one of the most important thinkers in the the history of Indo-Buddhist thought, is a fundamental exposition of Buddhist philosophy; a masterpiece of scholarship.)

Vasubandhu (5th century) *Triṃśikā*, ed. S. Lévi, in *Vijñaptimātratāsiddhi*, Bibliothèque de l'École des Hautes Études 245, Paris: Libraire Ancienne Honoré Champion, 1925; trans. S. Lévi, *Matériaux pour l'étude du système vijñaptimātratā*, Bibliothèque de l'École des Hautes Études 260, Paris: Libraire Ancienne Honoré Champion, 1932.(The 1925 edition also contains the Sanskrit text of Vasubandhu's *Viṃśatikā(Twenty Stanzas)*, together with commentaries; the 1932 edition is a French translation of the two works by Vasubandhu, with a useful index of terms.)

Wayman, A. (1971) 'Buddhist Dependent Origination', *History of Religions* 10: 185- 203.(An attempt to show a historical evolution of the doctrine of dependent origination from Pāli, Sanskrit and Tibetan sources.)

Suicide, ethics of

Suicide has been condemned as necessarily immoral by most Western religions and also by many philosophers. It is argued that suicide defies the will of God, that it is socially harmful and that it is opposed to 'nature'. According to Kant, those who commit suicide 'degrade' humanity by treating themselves as things rather than as persons; furthermore, since they are the subject of moral acts, they 'root out' morality by removing themselves from the scene.

In opposition to this tradition the Stoics and the philosophers of the Enlightenment maintained that there is nothing necessarily immoral about suicide. It is sometimes unwise, causing needless suffering, but it is frequently entirely rational and occasionally even heroic. Judging by the reforms in laws against suicide and the reactions to the suicides of prominent persons in recent decades, it appears that the Enlightenment position is becoming very generally accepted.

1 The horror of suicide

The word 'suicide' is used here in the sense defined by Emile Durkheim (1897), the leading sociological writer on the subject. He refers to 'all cases of death resulting directly or indirectly from a positive or negative act of the victim himself, which he knows will produce this result' (see Durkheim, E. §3). The question of whether suicide can ever be morally justified has been extensively discussed both by secular philosophers and by religious moralists. According to Plato, Aristotle, St Augustine, St Thomas Aquinas, Kant and Hegel suicide is always morally wrong, although Plato inconsistently allowed some exceptions. According to another tradition which goes back to Epicurus and the Stoic philosophers, suicide is not only frequently justified but on occasions highly admirable. 'Against all the injuries of life', wrote Seneca, 'I always have the refuge of death'. The Stoic view was revived in the sixteenth century by Montaigne and it had the support of all the leading figures of the French Enlightenment, notably Montesquieu, Holbach, Diderot, Rousseau and Voltaire. David Hume's 'On Suicide', which could not be published during his lifetime, may be regarded as the classic statement of the Enlightenment viewpoint. In Germany in the nineteenth century the Enlightenment position was championed by Schopenhauer and also by Nietzsche, who even went so far as to suggest that chronically sick people should be encouraged to commit suicide. 'In a certain state', Nietzsche wrote in *Twilight of the Idols* (1888: §36), 'it is indecent to live longer - to go on vegetating in cowardly dependence on physicians and medications, after the meaning of life… has been lost, ought to prompt a profound contempt in society'.

In the past, the anti-suicide viewpoint was frequently stated with extreme ferocity. Entirely typical of sermons in the eighteenth and nineteenth centuries is *The Guilt, Folly and Sources of Suicide* by Samuel Miller D.D., delivered and published in New York City in 1805. According to Miller suicide is 'repugnant to every genuine feeling of human nature'. It is 'as degrading as it is criminal'. It is a crime of 'complicated malignity' against which 'man, depraved, afflicted, and covered with evil, requires to be guarded by restraints'. The person who commits suicide 'is as great a monster in morals as an atheist in religion, or as the most hideous assemblage of deformities in animal nature'. Writing in 1908, the English poet and novelist G.K. Chesterton (1874-1935) declared suicide to be the worst of all crimes. 'Not only is suicide a sin', according to Chesterton, 'it is *the* sin. It is the ultimate and absolute evil…. The man who kills a man, kills a man. The man who kills himself, kills all men; as far as he is concerned he wipes out the world'. It is of some interest to note that the Talmud contains a similar idea. We are told that 'there is none more wicked than one who has committed suicide'. This is so because the world was created 'for the sake of one individual' and 'thus he who destroys one's soul is considered as though he had destroyed the whole world' (see Bioethics, Jewish §3).

If we take Chesterton literally, then Cato, Brutus, Cassius, Thomas Chatterton, Samuel Romilly, Heinrich von Kleist, Gerard de Nerval, Ludwig Boltzmann, Vladimir Maiakovskii, Wilhelm Stekel, Stefan Zweig, Ernest Hemingway, Percy Bridgman, Sylvia Plath, Paul Celan, Marilyn Monroe and Arthur Koestler, to mention just a few, were criminals; and indeed far worse criminals than mass murderers. Hitler's extermination of millions of innocent men and women was a lesser crime than his own eventual suicide. Chesterton did not at all deplore the 'weird harshness' that Christianity has shown to the suicide: 'The suicide is ignoble… he is a mere destroyer; spiritually, he destroys the universe'. We are assured that there is much rational and philosophic truth in the

practice of driving stakes through the bodies of suicides and burying them at crossroads. The suicide 'was so bad that his bones would pollute his brethren' (Kant [1775/80] 1930: 151).

Miller and Chesterton may be dismissed as religious fanatics and marginal figures. Immanuel Kant was not a religious fanatic and can hardly be dismissed as a marginal figure, but his denunciations of suicide are just as furious. According to Kant 'suicide is in *no* circumstances permissible'. The man who commits suicide 'sinks lower than the beasts'. We 'shrink from him in horror'. 'Nothing more terrible can be imagined'. 'We look upon the suicide as carrion'. If a man attempts suicide and survives, he has in effect 'discarded his humanity' and we are entitled to 'treat him as a beast, as a thing, and to use him for our sport as we do a horse or a dog'.

None of these writers showed the slightest compassion for a person who committed suicide, and their ravings were directed even against philosophers who did show such compassion and who denied that suicide was necessarily wrong or criminal. Hume in particular was denounced by clerical enthusiasts. G. Clayton, author of *The Dreadful Sin of Suicide* (1812: 48n), called Hume's essay a 'source of incalculable evil', and in a huge treatise, *A Full Inquiry into the Subject of Suicide* (1790 vol. 2: 54), Charles Moore denounced Hume as 'a more pernicious and destructive member of society than even the profligate and abandoned liver'.

2 Humans as God's property

According to Sir William Blackstone (1723-80), the influential English jurist, the suicide is guilty of a double offence. The first of these is 'spiritual' and consists in 'evading the prerogative of the Almighty, and rushing into his immediate presence' (Blackstone 1765-9 IV: 189). We evade God's prerogative, according to some theologians, because we are God's property and he alone has the right to terminate our lives. This argument is found in Plato's *Phaedo* (61c) where Socrates remarks that humans are 'the possession of the deity' and ought not to kill themselves before the deity lays them under a necessity of doing so: 'if one of your slaves were to kill himself, without your having intimated that you wished him to die, should you not be angry with him, and should you not punish him if you could?' This argument is repeated, almost verbatim, by Aquinas and also by Kant.

Kant relies primarily on purely secular arguments, but at the end of his most extended discussion of suicide he also appeals to the Divine ownership of man. 'We have been placed in this world', he writes, 'under certain conditions and for specific purposes. But a suicide opposes the purpose of his Creator; he arrives in the other world as one who has deserted his post; he must be looked upon as a rebel against God'. Human beings 'are sentinels on earth and may not leave their posts until relieved by another beneficent hand' (Kant [1775/80] 1930: 154). This notion had already been put forward in Locke's *Second Treatise on Civil Government*. We are there told that all are the 'servants of one sovereign Master… they are His property' and must not 'quit their station wilfully' (Locke 1690: §6).

The views just quoted are frequently accompanied by pronouncements that God created human beings primarily for his glory. Zachary Pearce (1690-1774), Bishop of Rochester, may be taken as a typical representative of this position. In his 'Sermon on Self-Murder' (1736), he declares that God had two ends in view when he created humans. These 'could have been no other than… the promoting of His own Glory and Service, and of our true and real happiness'. We promote God's glory 'by the practice of every virtue'. Patience and submission to the Divine Will are virtues that are 'chiefly exercised in a state of adversity'. From all of this it clearly follows that suicide is impermissible. 'The greater the adversity, the more conspicuous these virtues appear', and it follows 'very plainly' that 'the most afflicted man is capable of advancing the honour and interests of his Maker'.

All these arguments are open to a number of serious objections, even if one allows that the universe was created by God. In the first place, the fact that God created human beings does not mean that they are his property. My parents created me, but I am not their property. Similarly, the assertion that God presented me with the gift of life does not imply that I may not dispose of it in any way I please. A donor surely has no right to dictate to the recipient what they are to do with the gift. It might also be observed that human beings did not request the gift of life. Schopenhauer in one place remarks that many a human being 'would have declined such a gift if he could have seen it and tested it beforehand' ([1818] 1883, vol. 3: 390).

Another objection relates to God's perfect goodness and wisdom. It is not easy to see why such a being would be so concerned about his own glory, gripped by what Hume called 'one of the lowest of human passions, a restless appetite for applause' ([1777] 1965: 300). Nor is it at all obvious that a human being in a state of extreme physical

or mental agony, especially with no relief or improvement in sight, would be promoting God's glory by continuing to live. Moreover, if God is good and kind he will hardly wish human beings to suffer for *his* sake.

Furthermore, although it is, according to Kant, 'God's intention to preserve life', this is apparently consistent with human life being destroyed by all manner of causes other than by suicidal acts. Why is it not equally consistent with the latter? Kant remarks that 'human beings are sentinels on earth and may not leave their posts until relieved by another beneficent hand' (Kant [1775/80] 1930: 154). Why cannot the desire to die on the part of a person whose 'patience', in Hume's words, has been 'overcome by pain and sorrow' qualify as one of the manifestations of such a 'beneficent hand'? In such a situation, Hume adds, 'I may conclude that I am recalled from my station in the clearest and most express terms' ([1777] 1965: 301). If it is remembered that 'relief by the beneficent hand' may take such forms as cancer or murder at times when the person very much *wishes to live*, it is not easy to see that committing suicide when the individual does not wish to live will not also qualify as relief by the same hand.

3 Adverse effects on survivors

As mentioned above, Sir William Blackstone condemned suicide as a 'double offence'. The first, the 'spiritual' offence, was discussed in the previous section. The second offence is 'temporal, against the king, who hath an interest in the preservation of all his subjects'. This argument is found in Aristotle's *Ethics*, and stated more fully by Aquinas who writes:

> It is altogether unlawful to kill oneself, because every part, as such, belongs to the whole. Now every man is part of the community, and so, as such, he belongs to the community. Hence by killing himself he injures the community.
>
> (*Summa theologiae* IIaIIae.64.5)

Some writers have been less concerned with harm to the community than with the grief and suffering that a suicide may bring to friends and family. 'It is not wisdom but barbarity', in the words of the great Italian poet-philosopher Giacomo Leopardi ([1827] 1893: 225), 'to reckon as nothing the grief and anguish of the home circle, the intimate friends and companions'.

A little reflection shows that such considerations fail to establish the conclusion that suicide is always wrong. The most they show is that a responsible individual would carefully take into account the effects on others before taking such a drastic step. One need not deny that an individual has obligations to society or to family and friends. However, apart from the fact that some people have neither family nor friends and that some people are justly aggrieved against the society in which they live, there are two objections to all arguments of this kind. In the first place, in any number of situations, a person's welfare may rightly take precedence over obligations to others. If by suffering a little, a person can avoid a great deal of suffering on the part of friends and family or the whole of society, his suffering is justified. If, on the other hand, the continuation of the person's life involves terrible pain, far exceeding the sufferings which their suicide would produce on other people, the person is not obliged to go on living. Hume discussed this topic at some length. Allowing for the sake of argument, he writes, 'that our obligations to do good (to society) were perpetual, they certainly have some bounds; I am not obliged to do a small good to society at the expense of a great harm to myself' (Hume [1777] 1965: 304). In a great many cases, furthermore, a person who commits suicide acts, at least in the long run, for the *benefit* of others. Although death may produce grief it also produces relief. 'Suppose that it is no longer in my power to promote the interest of society', Hume writes, 'suppose that I am a burden to it; suppose that my life hinders some person from being much more useful to society; in such cases, my resignation of life must not only be innocent, but laudable'. Since we are here dealing with utilitarian considerations it may not be amiss to observe that the suicide of a malevolent tyrant may bring great joy to large numbers, and not only to those he has harmed. Millions of people were happy to hear of Hitler's death, but the fact that he committed suicide gave greater satisfaction than if he had died of cancer or heart disease: it was, in effect, an admission of defeat.

4 An offence against nature

One of the most widely used arguments against suicide is that it is opposed to 'nature'. Thus, in a much-quoted judgment, Mr Justice Brown in 1562 declared suicide to be a crime on the ground that it is 'an offence against nature: because to destroy oneself is contrary to nature, and a thing most horrible'. In Aquinas the appeal to nature

is the first of his three reasons for condemning suicide. Suicide, he tells us, ‘is contrary to the inclination of nature and to the charity whereby every man should love himself’ (*Summa theologiae* IIaIIae.64.5). It follows from this that suicide is ‘always a mortal sin’. The argument is also found in numerous Protestant tracts of the eighteenth and nineteenth centuries. Samuel Miller employs his usual hell-fire rhetoric. ‘Suicide is repugnant to every genuine feeling of human nature,’ he tells us. ‘It is an outrage on the dignity of those faculties with which the Author of Nature has endowed us’. Although most writers who used this argument were believers in God and also employed theological arguments, the appeal to nature is meant to be of a purely secular character. Leopardi, a total atheist, expressed himself in similar language. ‘Nature itself’, he wrote, ‘unmistakeably teaches that it is not permitted to quit this world by our mere will and our own act’. It is implied that if humans were to cut off their arms or pluck out their eyes, this would be horrendously unnatural. Such acts, however, are not nearly as unnatural as to extinguish one’s very life. Such an act is ‘the one most opposed to nature which a person can commit’ (Leopardi [1827] 1893: 215).

The sufficient answer to this argument is that the suppression of an instinctive urge is not necessarily wrong. A surgeon who gets extremely hungry in the course of an operation can surely not be blamed for suppressing a passionate desire for a roast beef sandwich. Furthermore, instinctive desires may and frequently do conflict with one another, and then the right action, whatever it is, will have to frustrate our instincts. It may be granted that our desire to continue life is natural, but it is just as natural to avoid intolerable pain with no relief in sight. In such a situation suicide may be no less ‘in accord with nature’ than continuing to live. The best comment on the arbitrariness of the orthodox position has come from Henry Romilly Fedden in *Suicide - A Social and Historical Study* (1938). ‘A monk’, Fedden writes, ‘denies sex, the suicide disregards self preservation: both should be equally guilty, yet one is a saint and the other a sinner’ ([1938] 1972: 282).

5 Kant’s arguments

Kant appeals not only to the alleged fact that man is God’s property for his blanket rejection of suicide. He also offers a number of purely secular arguments. Two of these deserve some discussion. According to one of these, people who commit suicide are ‘abasing’ and ‘degrading’ their humanity by treating themselves as no more than things:

> Man can only dispose of things; beasts are things in this sense; but man is not a thing, not a beast. If he disposes of himself, he treats his value as that of a beast. He who so behaves, who has no respect for human nature makes a thing of himself.
>
> (Kant [1775/80] 1930: 151)

Kant is surely wildly wrong here. I am treating people as things and debasing their humanity if I try to dominate them so that they will, under the force of my superior will, automatically do what *I* want. Setting aside the notion of treating somebody as a thing, it is unquestionable that people frequently debase other human beings. I am debasing people if I humiliate them, if I get them to the point at which, to preserve their jobs which I control, they have to fawn and beg for mercy or to confess to wrongs they never committed. In such circumstances I have no regard for the others’ feelings, especially for their pride and dignity. In reply to Kant it must be emphasized that a great many cases in which people committed or attempted to commit suicide do not at all resemble debasements of this kind. If I commit suicide I am not necessarily the victim of the stronger will of somebody else. I am not indifferent to my own feelings or dignity, but on the contrary I may *compassionately* decide to terminate what I regard as my pointless (perhaps even degrading) suffering. I have not become a thing and I have not at all debased myself.

Kant’s other argument is based on the undeniable fact that if people commit suicide they can no longer perform any moral acts. ‘It cannot be moral’, he says, ‘to root out the existence of morality itself from the world’. The suicide ‘robs himself of his person. This is contrary to the highest duty we have towards ourselves, for it annuls the condition of all other duties’ (Kant [1775/80] 1930: 152). To this it must be replied that people who commit suicide do *not* root out the existence of morality itself from the world. They do not do so any more than when they die a natural death or are killed in battle. They ‘root out’ any new moral acts on their part, but presumably there will be other people left. They would root out ‘morality itself’ only if they wiped out the human race.

Kant’s argument involves a confusion between: (1) I ought to do my duty as long as I am alive, and (2) It is my

duty to go on living as long as possible. Kant's basic value judgment that doing one's duty is the highest good implies (1), but it does not imply (2); and only (2) could serve as a basis for condemning suicide.

It should be noted that Kant himself in various places rejects statement (2). In one place he remarks 'there is much in the world far more important than life' and that 'it is better to sacrifice one's life than one's morality'. Furthermore, it is entirely permissible and even laudable 'to risk one's life against one's enemies, and even to sacrifice it, in order to observe one's duties towards oneself'. Kant also fully endorses the right of 'the sovereign' to 'call his subjects to fight to the death for their country'. Those who die in battle, Kant goes on, are not suicides but 'victims of fate'. They are to be admired as 'noble and high-minded' in contrast with soldiers who run away to save their lives. Yet the deaths of the noble 'victims of fate' root out the existence of 'morality itself' from the world just as much as the deaths of people who commit suicide, while cowardly soldiers who saves their own lives thereby preserve the condition for further moral action. Hence the mere fact of not preserving the presupposition of future moral action cannot be a sufficient reason for condemning suicide.

6 The legal punishment of suicide

Until relatively recently suicide was regarded as a capital crime in the legal statutes of most Western countries. Unfortunately from the point of view of the punishers, it is not possible to carry out a death sentence in the case of successful suicides. This does not mean, however, that nothing has been done by the state or church to show their extreme disapproval. The suicide was declared a felon and their property confiscated. Regular burials were strictly forbidden and Christian rites denied. The body was usually taken to the crossroads, a stake driven through it and a stone placed over the face. The Prussian Code published in 1788 and confirmed in 1794 declares that 'the corpse of a suicide shall be duly executed, if, in the opinion of the judge, the act would operate as a deterrent'. Such laws are perhaps not as absurd as they appear at first. Most people, almost regardless of their philosophical or religious views, are concerned about the disposition of their body and repelled by the notion of its mutilation.

In the case of suicides whose attempts fail, the authorities were in a position to carry out the execution of the 'murderer'. It is not known how often this was actually done but we have an appallingly gruesome description of one such case that took place in England in 1860 (Carr 1981: 336).

Laws against suicide were abolished in France in 1790 largely as a result of the influence of Beccaria, Montesquieu, Diderot and Voltaire. Other European countries and several US states followed the lead, but in Britain people were prosecuted as recently as 1955. Several of these cases are described in Glanville Williams (1958). The 'criminality' of attempted suicide was finally abolished in Britain in 1961 following a recommendation (1959) by a commission appointed by the Archbishop of Canterbury. Much of the report repeats, in very strong language, some of the older Christian condemnations, but in the end, after quoting extensively from the writings of Hastings Rashdall, Dean Inge and Canon Peter Green, it is grudgingly admitted that, in certain situations, suicide is not immoral.

Assisting suicide remains a crime in thirty-two US states but, as is generally known, such prohibitions are commonly disregarded by humane physicians. Dr Jack Kevorkian, a Michigan pathologist, has openly defied this law. He has been tried several times and acquitted on each occasion. In two (April and May 1996), one in New York City, the other in San Francisco, Courts of Appeal found statutes forbidding physician-assisted suicide to be unconstitutional. Both cases are under appeal at the time of writing, but there is every indication that, before long, physical-assisted suicide will be regarded as a constitutional right.

See also: Bioethics; Death; Life and death; Medical ethics

PAUL EDWARDS

References and further reading

Alvarez, A. (1971) *The Savage God: A Study of Suicide*, London: Weidenfeld & Nicholson.(Historical account of religious and secular attitudes towards suicide, as well as a survey of theories about the causation of suicide.)

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, trans. the English Dominican Fathers as *The Summa Theologica*, London, 1912-36, 22 vols.(Aquinas discusses suicide in IIaIIae.64.5.)

Aristotle (*c.* mid 4th century BC) *Nicomachean Ethics*, trans. J.A.K. Thompson, London: George Allen & Unwin, 1953; trans. T.H. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985.(See 1138a6-14 for

Aristotle’s argument that the suicide acts unjustly towards the state.)

Battin, M.P. (1982) *Ethical Issues in Suicide*, Englewood Cliffs, NJ: Prentice Hall.(Critical survey of all major religious and secular arguments against the justifiability of suicide.)

Battin, M.P. and Mayo, D.J. (eds) (1980) *Suicide: The Philosophical Issues*, New York: St Martin’s Press.(Full discussion of all philosophical and psychological questions about suicide by writers from different fields.)

Beauchamp, T.L. and Perlin, S. (eds) (1978) *Ethical Issues of Death and Dying*, Englewood Cliffs, NJ: Prentice Hall.(Contains sections on suicide, euthanasia and the meaning of life and death; religious and secular positions are both represented.)

Blackstone, W. (1765-9) *Commentaries on the Laws of England*, London, 4 vols; 1st edn repr. Chicago, IL: Chicago University Press, 1979.(The first comprehensive description of the principles of English law.)

Carr, E.H. (1981) *The Romantic Exiles: A Nineteenth-Century Portrait Gallery*, Cambridge, MA: MIT Press. (Illustrates the lengths to which authorities were prepared to go in pursuance of the law against suicide.)

Chesterton, G.K. (1908) *Orthodoxy*, London: John Lane; repr. New York: Doubleday, 1991.(Condemnation of suicide based on the view that ‘cosmic patriots’ have taken an ‘oath of loyalty to life’. A person who commits suicide denies the ‘primary sacredness’ of the universe.)

Clayton, G. (1812) *The Dreadful Sin of Suicide*, London: Black, Parry & Kingsbury.(Sermon denouncing suicide delivered in London in 1812.)

Committee Appointed by the Archbishop of Canterbury (1959) *Ought Suicide to be a Crime?*, Report of a Committee Appointed by the Archbishop of Canterbury, London: Church Information Office.(Statement of the Anglican position on suicide, allowing that on some occasions suicide is not morally blameworthy.)

Donnelly, G. (ed.) (1990) *Suicide - Rights or Wrong?*, Buffalo, NY: Prometheus Books.(Anthology representing all major viewpoints on the morality of suicide.)

Durkheim, E. (1897) *Suicide: A Study in Sociology*, trans. J.A. Spaulding and G. Simpson, Glencoe, IL: Free Press, 1951.(Pioneer study of different types of suicide and of their causation, based on the statistical data available at the end of the nineteenth century. The main thesis is that the frequency of suicide varies inversely with the degree of integration of the social group to which the individual belongs.)

Dworkin, R. *et al.* (1997) ‘Assisted Suicide: The Philosophers’ Brief’, *New York Review of Books* 27 March 1997. (Submission to the US Supreme Court in connection with recent cases of assisted suicide.)

Fedden, H.R. (1938) *Suicide - A Social and Historical Study*, New York: Benjamin Blom, 1972.(Covers much the same ground as Alvarez, with greater emphasis on the irrational attitudes of Christian theologians.)

Hume, D. (1777) ‘On Suicide’, in A. MacIntyre (ed.) *Hume’s Ethical Writings*, New York: Macmillan, 1965.(The classic statement of the view that suicide is not necessarily wicked or even unwise.)

Hume, D. (1779) *Dialogues concerning Natural Religion*, ed. N.K. Smith, London: Thomas Nelson, 1935.(Classic statement of the main objections to the teleological and cosmological arguments for the existence of God.)

Kant, I. (1775/80) *Lectures on Ethics*, trans. L. Infield, London: Methuen, 1930.(A little-known volume containing Kant’s lectures on ethics, as recorded by his students.)

Lecky, W.E.H. (1869) *History of European Morals from Augustine to Charlemagne*, London, 2 vols.(Contains amusing titbits about Roman philosophers who preached that death was better than life, resulting in suicide epidemics.)

Leopardi, G. (1827) *Essays, Dialogues and Thoughts*, trans. P. Maxwell, London: Walter Scott, 1893.(In the ‘Dialogue between Plotinus and Porphyrius’, the former character argues that we must ‘accommodate ourselves to the inevitable conditions of our existence’ and ‘self-destruction is an atrocious and inhuman act’.)

Locke, J. (1690) *Second Treatise on Civil Government, in Locke’s Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1970.(Classic work of political philosophy. For Locke’s comments on suicide see §§6, 23 and 135.)

Miller, S. (1805) *The Guilt, Folly and Sources of Suicide*, New York.(Sermon delivered in New York city in 1805.)

Moore, C. (1790) *A Full Inquiry into the Subject of Suicide*, London, 2 vols.(Detailed critique of defenders of suicide such as Donne and Hume, conducted in extremely immoderate language.)

Nietzsche, F. (1888) *The Twilight of the Idols*, trans. R.J. Hollingdale, Harmondsworth: Penguin, 1969.(The quotation in §1 above comes from the section entitled ‘Expeditions of an Untimely Man’.)

Pearce, Z. (1736) ‘Sermon on Self-Murder’, London.(Pamphlet arguing that ‘it is our duty to receive evil, as well as good, at the hand of God’.)

Perlin, S. (ed.) (1975) *A Handbook for the Study of Suicide*, New York: Oxford University Press. (Cross-disciplinary volume including articles by philosophers, biologists and psychiatrists.)

Plato (*c*.386-380BC) *Phaedo*, trans. with notes D. Gallop, *Plato Phaedo*, Oxford and New York: Oxford University

Press, 1993.(Dialogue concerning the immortality of the soul.)
Rosen, G. (1975) ‘History’, in S. Perlin (ed.) *A Handbook for the Study of Suicide*, New York: Oxford University Press.(Contains much fascinating material on the nineteenth-century literature in England and France.)
Russell, B. (1932) ‘Illegal?’, in *Mortals and Others - American Essays 1931-1935*, ed. H. Ruja, London, Routledge & Kegan Paul, 1975.(Excellent article outlining Russell’s views on suicide.)
Schopenhauer, A. (1818) *The World As Will and Idea*, trans. R.B. Haldane and J. Kemp, London: George Routledge & Sons, 1883.(Section 69 of Volume 1 outlines his attitude to suicide, which is also discussed in his specific essay on the topic.)
Schopenhauer, A. (1891) ‘Of Suicide’, in *Studies in Pessimism*, trans. T.B. Saunders, London; repr. New York: Wiley & Sons, 1992.(Endorses Hume’s arguments against the Christian view of suicide as sin, but also argues that it is only an apparent and not a real release from the sufferings of life.)
Seneca (*c.* AD early 60s) *Seneca’s Letters to Lucilius*, trans. E.P. Barker, Oxford: Clarendon Press, 1932, 2 vols.(Classical statement of the view of the Stoic philosophers that we are permitted to escape the ills of the world by ending our lives. Seneca himself committed suicide.)
Sprott, S.E. (1961) *The English Debate on Suicide from Donne to Hume*, La Salle, IL: Open Court.(Detailed account of many seventeenth- and eighteenth-century publications defending the traditional Christian viewpoint.)
Voltaire, F.-M. de (1764-9) *Philosophical Dictionary*, London: J. & H.L. Hunt, 1824, 6 vols.(Huge, misleadingly entitled collection of mostly short pieces, many in dialogue form, on a great variety of topics.)
Westermarck, E.A. (1906-8) *The Origin and Development of the Moral Ideas*, London and New York: Macmillan, 2 vols.(History of major moral ideas written from a relativistic viewpoint.)
Williams, G. (1958) *The Sanctity of Life and the Criminal Law*, London: Faber & Faber.(Contains a highly informative chapter on the history of the laws concerning suicide. This work cannot be sufficiently praised for its humanity.)

Sunzi

Sunzi: The Art of Warfare (or Sunzi bingfa), a text traditionally ascribed to Sun Wu, a contemporary of Confucius, is the most widely read military classic in human history. Although it provides counsel on military strategy and tactics, it is fundamentally a philosophical text, reflecting a way of thinking and living that is distinctively Chinese. The received text has thirteen 'core' chapters, but in 1972 an additional six chapters of the original eighty-two-chapter text was recovered from a Han dynasty tomb. In the same tomb, a text called Sun Bin (or Sun Bin bingfa), ascribed to a later descendent of Sun Wu, was also recovered. This work elaborates on the substance of the Sunzi's military philosophy.

In the centuries leading up to the founding of the Chinese empire in 221 BC, death had become a way of life and military texts were among the most widely circulating documents. The *Sunzi* and *Sun Bin* date from what is aptly called the Warring States period (403-221 BC), a period in which the heat and ferocity of warfare among independent nation-states, caught up in the frenzy of zero-sum politics, was rising exponentially. Historical records report that, between 350 and 250 BC, the approximate dates for the *Sunzi* and *Sun Bin* respectively, average battle casualties rose from 30,000-40,000 to ten times that number.

Many factors contributed to the carnage, including widespread use of the crossbow, siege engines to attack increasingly important population centres, advances in metallurgy and the introduction of cavalry. But in the final analysis, the revolution was intellectual. During this period, loosely organized militia gave way to the use of fixed formations, creating unrelenting killing machines similar in concept to the Roman phalanx, the revival of which in modern Europe made the Napoleonic armies so formidable.

The key and defining idea in the *Sunzi* is *shi* (pronounced like the affirmative, 'sure!'), translated variously as 'circumstances', 'situation', 'power', 'strategic advantage' and 'military or political purchase'. *Shi* entails manipulating circumstances to create a strategic advantage. It separates the skilful manipulation of a situation from brute force, and military wisdom from mere physical prowess.

An exploration of *shi* provides a contrast between an agent-centred model of order which tends to emphasize linear efficient cause, and the more situationally-defined alternative. When *shi* is translated as 'strategic advantage', many readers move immediately to assign it to one side of the conflict or the other. *Shi*, however, refers to the synergistic relationship that obtains among all of the factors on both sides of the conflict (numbers, terrain, logistics, morale, weaponry and so on) as they converge on the battlefield to give one side the advantage over the other. It is the tension generated in the contest between surplus and deficiency that becomes the 'force of circumstances'. *Shi* is both the shape and the momentum that the encounter assumes: it is the tide of battle. It is the purchase and the leverage that gives troops the will to join the battle and to win it.

Shi is thus not a given; it must be created and carefully cultivated. A failure to cultivate *shi* will surely give the upper hand to the enemy, and even the best soldiers under the worst circumstances will turn and flee. What makes cultivation of a situation possible are the indeterminate elements. At the same time, these ever-present fluid factors make each battlefield unique, and render any formulaic or mechanical application of even a proven military stratagem ultimately unreliable. Indeed, each new situation that develops in the field brings with it the necessity for reconsideration, and a flexibility that must constantly be ready to take account of any eventuality. It is this indeterminacy that makes the range of available strategies limitless.

The capacity of the small, incipient and seemingly incidental to control the large by virtue of its pivotal position is an enduring theme in classical Chinese philosophy, and underlies the notion of getting the most from a situation while minimizing loss. The commander must search the circumstances to identify a 'trigger' that can alter the propensity of battle. In this sense of order, this initially indeterminate aspect, while relatively small, can become a motivating force for reorganization and renewal, thus constituting a critical turning point. Because uncertainty can provide the occasion for transformation, it can offer either 'danger' or 'opportunity', depending upon whether or not the commander is able to capitalize on the moment and make the most of it. In fact, in modern Chinese, the combination of 'danger' and 'opportunity' - *weiji* - means 'critical point'.

In the *Sunzi*, the prosecution of warfare must imitate nature, and nature is too complex to be linear and

scale-invariant. Reconfiguration can take place on the battlefield in the same way that renewal can occur in nature because both situations are fluid, and rich in disorder and surprise. Small, pointed alterations can have precipitous and cascading consequences. Rather than separating what orders from what is ordered, this dynamic sense of order locates the energy of change within the situation itself by assuming order to be richly vague, ever-changing and always unique. When understood in this way, order is reflexive: it is self-organizing and self-renewing. While the stabilizing regularity of a specific event anticipates the way in which it will continue to unfold, the chaotic aspect within the event itself defeats any notion of necessity or absolute predictability. The combination of pattern and uncertainty in nature and on the field of battle undermines the possibility of universalizable claims and renders precarious any globalizing generalizations. All the commander can depend upon is the *relative* stability of site-specific and emerging patterns of order, and enough knowledge of the situation to identify and act upon the positive uncertainties that present themselves. He must constantly weigh the stochastic variables at every level of the battle and be prepared to parlay them into large scale advantages.

In the sequel text of the same lineage, the *Sun Bin*, the meaning of *shi*, a rather complex idea, was subcontracted out to other technical terms, a cluster of images which became evocative in both communicating and stimulating military insights. Among the most important were *quan* (weighing up with the lever scales or steelyard), *zhen* (battle formation and display), *bian* (adaptability) and *zhi* (foreknowledge based upon information gathering). These two texts belong to a lineage of military philosophy which, in modern times, has attracted a cult following with their ideas being extended from the battlefield to all other forms of competition, notably diplomacy, the marketplace and the field of play.

See also: Chinese philosophy; Political philosophy, history of; War and peace, philosophy of

ROGER T. AMES

References and further reading

Ames, R.T. (1993) *Sun-tzu: The Art of Warfare*, New York: Ballantine.(A translation and philosophical introduction to this classic based upon a 1972 archaeological find, including the lost six chapters and encyclopedic references.)

Lau, D.C. and Ames, R.T. (1996) *Sun Pin: The Art of Warfare*, New York: Ballantine.(A translation and philosophical introduction to the newly recovered classic based upon a 1972 archaeological find, including encyclopedic references.)

Supererogation

Supererogatory actions are usually characterized as 'actions above and beyond the call of duty'. Historically, Catholic thinkers defended the doctrine of supererogation by distinguishing what God commands from what he merely prefers, while Reformation thinkers claimed that all actions willed by God are obligatory. In contemporary philosophy, it is often argued that if morality is to permit us to pursue our own personal interests, it must recognize that many self-sacrificing altruistic acts are supererogatory rather than obligatory. The need for some category of the supererogatory is particularly urgent if moral obligations are thought of as rationally overriding. There are three main contemporary approaches to defining the supererogatory. The first locates the obligatory/supererogatory distinction within positive social morality, holding that the former are actions we are blameworthy for failing to perform, while the latter are actions we may refrain from performing without blame. The second holds that obligatory actions are supported by morally conclusive reasons, while supererogatory actions are not. On this approach the personal sacrifice sometimes involved in acting altruistically counts against it from the moral point of view, making some altruistic actions supererogatory rather than obligatory. The third approach appeals to virtue and vice, holding that obligatory actions are those failure to perform which reveals some defect in the agent's character, while supererogatory actions are those that may be omitted without vice.

1 Historical antecedents

Early Christian writers like Ambrose and Tertullian often distinguished those things that God willed through commandments or precepts, and those things that he willed only through counsels. They held that we are obliged only to obey God's precepts, and that we are permitted to disregard God's counsels if we so choose, although following them is always better than disregarding them. Actions in conformity with the counsels are called 'supererogatory'. The 'Evangelical Counsels', as they were first called, are poverty, chastity, and obedience (Tertullian, *Exhortation to Chastity*, section 3). It was held that in giving permission, for example, to live a life of less than complete chastity, God exercises his will 'in a spirit of indulgence', as a concession to the weakness of human character.

Scholastic thinkers typically offered a rather different account, sometimes side by side with this more traditional one. Aquinas, for example, held that only the conduct required by the precepts is necessary for salvation (see Aquinas, T. §13). Conduct recommended by the counsels is either advice about the most efficient way to gain salvation (*Summa theologiae*), or advice about the way to go beyond what is required for salvation, in order to attain perfection (*Summa contra gentiles*). On Aquinas' teleological view of moral necessity, conformity to the precepts is thus obligatory, or morally necessary, because it is necessary for salvation. Conformity to the counsels is morally optional because it is optional (though perhaps useful) with respect to that same end.

According to Catholic doctrine, God chooses to reward acts of supererogation with the gift of greater merit than the agent requires for salvation. This surplus merit is stored in the 'Treasury of the Saints', owned by the Church, and available to be drawn on by others. This notion of 'congruent merit' or freely-given and transferable reward provided the theological underpinnings for, among other things, the practice of granting indulgences, or the cancelling of another's punishment for some or all of their sins (Heyd 1982: 26).

The Reformers, particularly Luther and Calvin, vehemently criticized the doctrine of supererogation; but much of the criticism that was focused specifically on this doctrine was directed less to the Thomistic position than to the older view formulated by Tertullian (Calvin 1536; see Mellema 1991; Heyd 1982: 26-9).

2 The contemporary argument for supererogation

The Reformation disputations over the coherence and the legitimacy of the notion of supererogation are echoed in contemporary secular debates. It is often argued that if one were always obliged to promote wellbeing, then the demands of morality would be far too strenuous. Our own personal projects would never get off the ground; and even once launched they would constantly be held hostage to the needs of others (Wolf 1982). On the other hand, it is said, if morality recognized a category of supererogatory action and assigned many of the more personally demanding acts of beneficence to this category, then we would have some opportunity to pursue our own interests without moral penalty (see Help and beneficence).

This argument assumes, first, that there are two distinct and competing perspectives: the moral point of view and the self-interested point of view; and that these can recommend conflicting courses of action in the same circumstances (see Egoism and altruism). Second, the argument locates the cause of the conflict in a too-narrow, tripartite conception of the moral realm. This conception recognizes only three categories of action: the forbidden, the required, and the morally indifferent (morally permissible) (Urmson 1958). After all, the reasoning goes, since presumably acts of beneficence are never morally indifferent, on the tripartite conception (unless they are forbidden) they will be obligatory. Hence the conflict with self-interest.

But does the mere prospect of such conflict, however extensive, really require the introduction of some notion of the supererogatory? Why not just accept that such conflicts can occur? Alternatively, why not take the possibility of such widespread conflict with personal concerns to de-legitimize the very conception of the moral on which that possibility turns?

These questions show that our two assumptions do not force the introduction of a category of the supererogatory. However, if we are willing to set aside the option of rejecting our working conception of the moral, we can introduce a third assumption that makes the need for a category of the supererogatory more urgent. This is the assumption that moral reasons are rationally overriding. Suppose that moral reasons always outweigh reasons of self-interest, and that beneficent action is (nearly always) obligatory. It follows that to the extent that beneficence and self-interest are in frequent and fundamental conflict our pursuit of personal projects will be contrary to reason. This conclusion seems entirely unacceptable. To see whether admitting the idea of supererogatory action will avoid this conclusion, however, we must understand how supererogation is to be defined.

3 Supererogation and positive social morality

Assuming that supererogatory action and obligatory action can both be supported by moral reasons, how are we to distinguish them? There are at least three possible approaches. The first is John Stuart Mill’s proposal to restrict the operation of the distinction to the realm of social morality. On this proposal (1861), both supererogatory actions and obligatory actions are actions that maximize utility, and hence actions that one ought to perform. The difference lies only in whether various external and internalized sanctions like blame and guilt are appropriate (useful) given failure to perform. Here Mill echoes Tertullian’s idea that it is God’s willingness to punish failure to perform that distinguishes obligation from supererogation. To avoid the unacceptable conclusion described in the previous section, this approach must hold that to say moral reasons are overriding is only a way of making the point that they are strongly reinforced by negative social sanctions. It will then follow that only obligation-making reasons will be ‘overriding’.

4 Supererogatory acts as morally optional

The second approach focuses attention not on social morality but on the character of the reasons that support beneficent acts. Suppose we accept the following as partial definitions of obligation and supererogation: an act is obligatory only if its omission is morally impermissible; and an act is supererogatory only if its omission is morally permissible. Suppose we further posit that it is morally permissible to act against some set of moral considerations if and only if those considerations are not morally conclusive. We then have a criterion for distinguishing obligatory action from supererogatory action: any action supported by morally conclusive reasons will be obligatory; any action supported by reasons that are less than morally conclusive, will be supererogatory (Dancy 1993).

We can complete this approach by adding an account of overridingness: moral considerations are rationally overriding (if and) only if they are morally conclusive. This approach thus allows for the rational pursuit of self-interest whenever considerations of personal cost (represented within morality) are powerful enough to render the balance of moral reasons inconclusive. Conversely, only when morality regards certain considerations as conclusive must reason do so as well. Hence one is rationally required only to do one’s duty (see Duty §3).

5 Supererogation, virtue, and vice

The third approach accepts the partial definitions of obligatory action and supererogatory action with which the previous approach began. But it proposes that we regard both obligatory action and supererogatory action as supported by morally conclusive reasons. This proposal is backed by the suggestion that considerations about

personal inconvenience are not really full-blown moral considerations, on a par with considerations about human wellbeing, fairness, loyalty, honesty, and so forth. Thus helping others can be 'the moral thing to do', this approach says, regardless of whether it happens to require self-sacrifice.

Here is the puzzle for this third approach: whenever an action is supererogatory, it must be morally permissible to refrain from acting altruistically. But how can it ever be morally permissible to act against morally conclusive reasons?

We might solve this puzzle by making a distinction between first- and second-order moral judgments. Think of first-order judgments as being about agents' actions, while second-order judgments are about the motivation or perhaps the character agents display in their choice of actions. Judgments about the moral conclusiveness of the considerations at stake are first-order judgments. Now in some cases, when the cost to the agent is negligible and the human value at stake is great, to disregard morally conclusive considerations would be to show a degree of callousness, indifference, or smallness of spirit so great that it could only be called a vice (see Virtues and vices §§4-5). This is why refraining from acting in accordance with the first-order reasons is impermissible in this case, and why acting in accordance with those reasons is therefore obligatory.

In other cases, for instance when the cost to the agent is substantial, although to disregard the first-order moral considerations would be to display a less-than-ideal level of motivation, it would hardly be to sink so low as to warrant the imputation of vice. This is why it is permissible to refrain from acting in accordance with the first-order reasons in this case, and why acting in accordance with them is therefore supererogatory (see Trianosky 1986).

Finally, we can add to this third approach a more nuanced account of overridingness. On this account, morally conclusive first-order considerations will be rationally overriding if and only if to disregard them is to display some vice (and hence to do what is morally impermissible). Thus this third approach makes room for the rational pursuit of personal concerns just to the extent that morality is lenient in its judgments about virtue and vice.

6 Supererogation and praiseworthiness

By providing various accounts of overridingness, all three approaches offer some way to interpret the attractive idea that in doing what is supererogatory one acts freely, while in doing what is required, one is bound. It is often thought that this freedom is the ground of the special merit or praiseworthiness that is characteristic of supererogatory action (Heyd 1988).

But are all supererogatory acts necessarily praiseworthy? After all, whichever one of these three approaches one takes, the definition of the supererogatory will remain a purely *deontic* one, on a par with the definitions of the obligatory, the permissible, and the forbidden. This means that an agent can perform supererogatory acts from any one of a variety of motives, some morally admirable and some not. (Even the third approach says nothing about the motives of those who do what is supererogatory. It evaluates only the motives of those who refrain. Compare Brandt's discussion of obligation (1969).) In some ways this consequence seems intuitively plausible, since people can and sometimes do go 'beyond duty' for the most self-serving of reasons.

Even on a deontic analysis of supererogation there remains a certain conceptual link between supererogatory action and praiseworthy motivation: to go beyond duty, not for ulterior, self-interested reasons but for the sake of the moral concerns that are at stake, is necessarily praiseworthy. Here we have a notion of acting from *non*-obligatory moral considerations that is a strict analogue of Kant's notion of acting from duty. In the case of obligation, action from this praiseworthy motivation is what Kant calls 'dutiful' action. In the case of supererogation, it is what Mill (1843) calls 'noble' action. But not all actions in accordance with duty are dutiful; and not all actions beyond duty are noble (Mellema 1991: 3).

GREGORY VELAZCO Y TRIANOSKY

References and further reading

Aquinas, T. (*c*.1259-65) *Summa contra gentiles (Synopsis [of Christian Doctrine] Directed Against Unbelievers)*, trans. V.J. Bourke, Notre Dame, IN: University of Notre Dame Press, 1975.(Defends the claim that the counsels recommend conduct required for perfection, but not for salvation.)

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, trans. Fathers of the English Dominican

Province, Westminster, MD: Christian Classics, 1981.(Defends the claim that the counsels recommend more efficient ways of attaining salvation.)
Augustine (394) *De sermone Domini in monte (Commentary on the Lord's Sermon on the Mount)*, trans. and ed. J.J. Jepson, Westminster, MD: The Newman Press, 1948, 1.1.2.(Develops an account of 'greater and lesser righteousness' as an account of obligation and supererogation.)
Brandt, R.B. (1969) 'A Utilitarian Theory of Excuses', *Philosophical Review* 78: 337-61.(Contains a discussion of obligation that is seminal for the approach to supererogation which holds that morally conclusive first-order considerations will be rationally overriding if and only if to disregard them is to display some vice - and hence to do what is morally impermissible.)
Calvin, J. (1536) *Institutes of the Christian Religion*, trans. F.L. Battles, Grand Rapids, MI: W.B. Eerdmans, revised edn, 1986, esp. 2.8.51, 56-57.(Central Reformation critique of the concept of supererogation.)
Cicero (late 44 BC) *De officiis (On Duties)*, in three books, trans. with notes by M. Griffin and E.M. Atkins, Cambridge: Cambridge University Press, 1991, 1.3.8.(Influential presentation of the Stoic distinction between duties suitable for the Sage, *officia perfecta*, and duties suitable for ordinary, 'frail and fallible' people, *officia media*.)
Dancy, J. (1993) *Moral Reasons*, Oxford: Blackwell, ch. 8.(Contains the seminal presentation of the view that any action supported by morally conclusive reasons will be obligatory, while any action supported by reasons that are less than morally conclusive will be supererogatory; somewhat technical.)
Heyd, D. (1982) *Supererogation: Its Status in Ethical Theory*, London: Cambridge University Press.(A very useful and accessible survey of historical and philosophical issues concerning supererogation.)
Heyd, D. (1988) 'Moral Subjects, Freedom, and Idiosyncrasy', in J. Dancy, J.M.E. Moravscik and C.C.W. Taylor (eds) *Human Agency, Language, Duty, and Value: Philosophical Essays in Honor of J.O. Urmson*, Stanford, CA: Stanford University Press.(A discussion of the idea that supererogatory action are performed freely in a way that obligatory action cannot be.)
Mellema, G. (1991) *Supererogation, Obligation, and Offence*, Albany, NY: State University of New York Press.(Clearly written, accessible, and theologically informed discussion of historical and philosophical questions about supererogation. Ch. 3 deals with Luther and Calvin's criticism of the doctrine.)
Mill, J.S. (1843) *A System of Logic*, selections repr. in *Utilitarianism and Other Essays*, ed. A. Ryan, Harmondsworth: Penguin, 1987.(6.12.7 provides a discussion of the pursuit of noble character.)
Mill, J.S. (1861) *Utilitarianism*, repr. in *Utilitarianism and Other Essays*, ed. A. Ryan, Harmondsworth: Penguin, 1987, ch. 5.(The *locus classicus* for the approach which suggests that while both supererogatory actions and obligatory actions are actions that maximize utility, and hence actions that one ought to perform, only obligation-making reasons are 'overriding'.)
Nagel, T. (1986) *The View from Nowhere*, Oxford: Oxford University Press, ch. 10, section 4.(Presents a contemporary version of Tertullian's argument. The context is a very clear and thorough discussion of the tension between the moral and the self-interested points of view, which can be held to recommend conflicting courses of action in the same circumstances.)
Raz, J. (1975) 'Permissions and Supererogation', *American Philosophical Quarterly* 12: 161-8.(Defence of an influential interpretation of supererogation, which is built on the premise that we regard both obligatory action and supererogatory action as supported by morally conclusive reasons.)
Scheffler, S. (1992) *Human Morality*, Oxford: Oxford University Press, 120-32.(Presents an accessible and suggestive discussion of the contemporary argument about supererogation, which holds both that there are two distinct and competing perspectives (the moral point of view and the self-interested point of view) and that this conflict is caused by a too-narrow conception of the moral realm.)
Slote, M. (1985) *Common-Sense Morality and Consequentialism*, London, Routledge & Kegan Paul, 82-6.(Presents an alternative approach to the problem of overly-demanding obligations of benevolence, on which deontic terms are replaced with the scalar, non-deontic terms, 'good', 'better', and 'best'. Pages 86 and following contain a very suggestive discussion of the relative merits of deontic and non-deontic terms in guiding the agent in what to do.)
Tertullian, Quintus Septimus Florens (*c*.AD 160-*c*.AD 220) *Exhortation to Chastity*, in *Treatises on Marriage and Remarriage: Ancient Christian Writers, The Works of the Fathers in Translation*, vol. 13, trans. W.P. Le Saint, Westminster, MD: The Newman Press, 1951.(Seminal early discussion of the counsels/precepts distinction.)
Trianosky, G. (1986) 'Supererogation, Wrongdoing, and Vice', *Journal of Philosophy* 83: 26-40.(Elaboration of the distinction between deontic judgments and judgments of virtue and vice (aretaic judgments).)
Urmson, J.O. (1958) 'Saints and Heroes', in A.I. Melden (ed.) *Essays in Moral Philosophy*, Seattle, WA: University of Washington Press.(Seminal contemporary discussion of the need for a category of

supererogation.)
Wolf, S. (1982) ‘Moral Saints’, *Journal of Philosophy* 79: 419-39.(Influential and engagingly written discussion of the conflict between morality and self-interest. Wolf concludes with suggestive remarks about the role of supererogation in resolving the conflict.)
Williams, B. (1985) *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press, ch. 10.(An influential and accessible critique of the conception of morality that underlies the putative conflict between the moral point of view and the self-interested point of view.)

Supervenience

Supervenience is used of the relationship between two kinds of properties that things may have. It refers to the way in which one kind of property may only be present in virtue of the presence of some other kind: a thing can only possess a property of the first, supervening kind because it has properties of the underlying kind, but once the underlying kind is fixed, then the properties of the first kind are fixed as well. The supervening features exist only because of the underlying, or 'subjacent' properties, and these are sufficient to determine how the supervening features come out. For example, a person can only be good in virtue of being kind, or generous, or possessing some other personal qualities, and an animal can only be alive in virtue of possessing some kind of advanced physical organization. Equally, a painting can only represent a subject in virtue of the geometrical arrangement of light-reflecting surfaces, and its representational powers supervene on this arrangement. A melody supervenes on a sequence of notes, and the dispositions and powers of a thing may supervene on its physical constitution.

Although the word supervenience first appears in twentieth-century philosophy, the concept had previously appeared in discussion of the 'emergence' of life from underlying physical complexity. The central philosophical problem lies in understanding the relationship between the two levels. We do not want the relationship to be entirely mysterious, as if it is just a metaphysical accident that properties of the upper level arise when things are suitably organized at the lower level. On the other hand, if the relationship becomes too close so that, for instance, it is a logical truth that once the lower-level properties are in place the upper-level ones emerge, the idea that there are two genuinely distinct levels becomes problematic: perhaps the upper-level properties are really nothing but lower-level ones differently described.

If this problem is dealt with, there may still remain difficulties in thinking about the upper-level properties. For example, can they be said to cause things, or explain things, or must these notions be reserved for the lower-level properties? Supposing that only lower-level properties really do any work leads to epiphenomenalism - the idea that the upper-level properties really play no role in determining the course of events. This seems to clash with common-sense belief in the causal powers of various properties that undoubtedly supervene on others, and also leads to a difficult search for some conception of the final, basic or lowest level of fact on which all else supervenes.

1 Varieties of supervenience

The basic instinct behind a supervenience claim is that although we talk in terms of two levels of fact, one of them is fundamental in that once it is fixed then so is the other. It is as if God had only to decide how the basic level is arranged and the other follows on automatically, just as a composer need only decide on the notes and the tempo and then the melody emerges without anything further being done. In this kind of case the relationship is simply one of composition, and there is perhaps nothing mysterious in the way that the melody supervenes upon the notes and tempo, any more than there is anything mysterious in a house supervening on an arrangement of bricks and mortar. But in other cases we are not talking of the supervenience of one thing upon other things, but of one kind of fact upon other kinds of fact, and the relationship seems not to be so simple. In the philosophy of mind, for example, physicalists will certainly want to say that the fact that a creature is alive or conscious supervenes upon complex neurophysiological facts. God needs only to create beings of sufficient neurophysiological complexity and thereby will have created living or conscious beings; no further act of breathing life into the being is necessary.

But it seems only metaphorical to say that life is 'composed' of neurophysiological complexity. The relationship is not simply that of a building to its building blocks. Similarly, one might think that the fact that a computer is running a particular programme supervenes upon microphysical facts about its components, yet it is hard to make sense of the idea that its running the programme is literally composed of some arrangement of hardware. The term 'supervenience' was itself originally applied to the way ethical properties relate to natural ones, and in this case too it is hard to see how a person's virtue can literally be composed out of their natural dispositions, although here it may be more appropriate to think in analogous terms, since at least their virtue may be said to have components, such as charity or honesty. In any event, a general definition is needed.

If the root idea is that once the basic level is fixed, nothing more need be done to fix the upper level, then at least it

should be true that if two things are identical in respect of their basic properties, then they must also be identical in respect of the upper-level properties. The other way round this is not necessarily so, for two things may be identical in the upper-level respect without being identical all the way down. For example, two computers might be running the same programme but with rather different configurations of hardware, or two people may be equally virtuous but because of somewhat different characters. This is frequently referred to as the 'variable realization' of the upper-level properties, and is compatible with their supervenience on the lower level. Returning to the root idea, we can now frame it as an impossibility claim:

(S) It is not possible that two things should be identical in respect of their lower-level properties without also being identical in respect of their upper-level properties.

Various questions arise. First, and most obviously, any such claim only makes sense given a background distinction between the two levels of property. If we cannot genuinely separate the basic or subjacent (sometimes called subvenient) level from the upper or supervening level, then no such claim can be formulated. And in interesting cases, such as the supervenience of the ethical on the natural, there will be philosophers who deny that such a separation can be achieved, seeing it as dependent on a doubtful fact-value distinction.

Similarly, there may be disputes about priority: some philosophers may hold that the semantic elements of a language supervene upon the intentions with which its speakers use its elements; others hold that the content of peoples' intentions supervene upon the semantics of the language in which they would be disposed to express them (this kind of dispute also generates the battle between methodological individualism and various kinds of holism). There may also be disputes about the extent of the basic class. Thus, to some philosophers it is important that the present content of someone's sayings should supervene only on their present brain states; to others it is important that it supervene upon the nature of the historically extended entire social and physical environment in which they live.

A second range of problems arises because we might worry about the strength of impossibility in any particular case. Is it a matter of logical impossibility, or metaphysical impossibility, or is it simply a question of the laws of nature? One might need to advance claims in different strengths, such as:

(E) It is not possible, as a conceptual matter, that two things should be identical in respect of their natural properties without being identical in respect of their ethical properties.
(M) It is not possible, as a metaphysical matter, that two things should be identical in respect of their physical properties without being identical in respect of their mental properties.
(P) It is not possible, given the laws of physics as they are, that two things should be identical in respect of their microphysical properties without being identical in respect of their macro-physical properties (such as weight or tensile strength).

Such formulations probably do not exhaust the intended significance of supervenience claims. For example, they do not capture the asymmetrical dependency, whereby the upper arises 'in virtue of' the lower (after all, denying variable realization, someone might think it impossible that two things should be identical mentally without being identical physically, but this would not mean that the physical supervenes on the mental). Nevertheless, they form the essential core.

The third kind of claim is perhaps the easiest to understand: it only demands that the laws of physics, as they actually obtain, connect the two levels in such a way that identity in one respect delivers identity in the other. In different possible worlds obeying different physical laws, the connection may be severed, with the weight or tensile strength of objects depending on other factors also (in fact weight, in our world, actually depends on other factors, such as the distribution of matter around the universe). The only problem with this kind of supervenience claim is that it inherits the difficulties surrounding the concept of a law of nature.

A claim such as (M) raises the stakes. Here we want to say that there is a more intimate connection between the physical and the mental. It is not just in the world as we have it that mental properties emerge out of physical properties. Rather, in any world we can properly describe, the mental must emerge out of the physical. There is simply no metaphysical possibility of a 'ghost stuff' or separate source or underlying basis of mentality. It has to occur in creatures of particular kinds of physical complexity. It offends against metaphysics, not just physics as we

have it, to postulate two physically identical creatures, one of whom is conscious, the other a zombie. Anybody denying this would be thinking in terms of a dualism of some metaphysically improper stamp. Notice that although many philosophers will subscribe to this thought, it is not altogether easy to formulate it. One distinction frequently made is between 'strong' supervenience, whereby the claim applies to any two things in any possible worlds, and 'weak' supervenience, whereby it only governs things within the same world. The difference is that across worlds we might imagine different laws of nature, and perhaps different constituent elements of nature, whereas one world must be governed by one set of laws and contain only whatever elements it has.

Finally according to (E) anyone denying the supervenience is in conceptual error, betraying some kind of incompetence with the concepts involved. This goes further than (M), unless we think that metaphysics is no more than a display of what is implicit in the concepts with which we think. It is, at least, clear that someone disposed to value two things differently, although admitting that they are identical in every other respect, is somehow confused or out of line, or ignoring the entire purpose of the notion of valuing a thing, compared with someone who denies (M), perhaps because of lingering Cartesian dualism (see Dualism).

2 The promises and problems of supervenience

Once a supervenience claim is properly formulated and its strength suitably identified, it offers the seductive promise of a path between full-scale reduction of upper-level to lower-level properties, and an uncomfortable dualism. Full-scale reduction is here the attempt to show that there is really nothing but the underlying level, and that talk in upper-level terms merely makes disguised reference to this one kind of reality. It is prompted by metaphysical unease but it is expensive, since it usually turns out that the requisite reductions distort what we mean or what we are referring to. For example, it is highly unnatural to say that when we describe a computer by saying which programme it is running we either mean anything about its electronics or are referring to its engineering configuration. We are doing something different altogether, yet we ought to be able to make software descriptions metaphysically innocuous even if we think that all that is fundamentally going on is captured by the full story about the machine's hardware. Allowing that the software characterization supervenes on the hardware may fulfil this promise. Supervenience promises the relief from metaphysical anxiety, but without the costs of reduction.

To fulfil this promise it is clear that we must understand why the impossibilities in question arise. Without such an understanding, the upper-level properties may still seem to be unexplained 'danglers' or unexplained arrivals on the scene (reduction at least promises this explanation, by collapsing the two levels into one). This in turn may be a useful constraint on philosophies of various areas: for example, understanding why (E) is true may be easier on some accounts of ethical commitment than others. But without such extra commentary, and in particular in the philosophy of mind and the philosophy of language, supervenience claims may easily be part of the problem rather than part of the solution (see Reductionism in the philosophy of mind).

See also: Causation; Supervenience of the mental

SIMON BLACKBURN

References and further reading

Blackburn, S. (1973) 'Moral Realism' and 'Supervenience Revisited', in *Essays in Quasi-Realism*, New York: Oxford University Press, 1993.(Both papers stress the importance of an exact definition of the supervenience of the ethical on the natural, and the difficulty of explaining the fact on various views of ethics.)

Grimes, T.R. (1988) 'The Myth of Supervenience', *Pacific Philosophical Quarterly* 69: 152-60.(A critical look at proposed attempts to explicate the concept.)

Hare, R.M. (1952) *The Language of Morals*, Oxford: Oxford University Press.(The standard introduction of the notion of supervenience into modern debate.)

Hellman, G. and Thompson, F. (1975) 'Physicalism: Ontology, Determinism and Reduction', *Journal of Philosophy* 72 (17): 551-64.(A classic investigation of the metaphysics of supervenience.)

Kim, J. (1984) 'Concepts of Supervenience', *Philosophy and Phenomenological Research* 45: 153-76.(The classic paper defining and separating various concepts of supervenience.)

Kim, J. (1990) 'Supervenience as a Philosophical Concept', *Metaphilosophy* 21: 3-27. (A somewhat more pessimistic revisitation of the concept.)

Klagge, J.C. (1984) 'An Alleged Difficulty Concerning Moral Properties', *Mind* 93: 370-80.(An attempted

rebuttal of Blackburn’s argument.)
Morgen, C.L. (1923) *Emergent Evolution*, London: Williams & Norgate; repr. New York: AMS Press, 1977. (‘Emergence’ especially of life from chemical and physical complex systems, labelled the phenomenon of supervenience for an older generation of philosophers of mind and of biological science.)

Supervenience of the mental

Phenomena of one kind 'supervene on' phenomena of another kind just in case differences with respect to the first kind require differences with respect to the second. G.E. Moore claimed that beauty supervenes on non-aesthetic properties: if one painting is beautiful and another is not, there must be some relevant non-aesthetic difference between them. Supervenience seems to offer the possibility that a property may depend on other properties, without being explicable in terms of them. Contemporary philosophers of mind have employed the idea to capture the relation that appears to obtain between mental and physical properties.

Various supervenience relations have figured in recent discussions in philosophy of mind, of which the most important is 'strong property' supervenience characterized as follows:

Q strongly supervenes on $\{P_1, P_2, ...\}$ if and only if any individuals *x* and *y*, in any nomologically possible worlds, that differ with respect to *Q* also differ with respect to some of the *P*s.

Thus, beauty supervenes on physical properties if and only if any two things in any two nomologically possible worlds (worlds compatible with the laws of physics) that differ in their beauty also differ physically.

Supervenience has been claimed to capture two ideas. One is that individuals that agree on all their *P* properties also agree with respect to *Q*. The other is that the possession of *Q depends* on the possession of *P* properties. Besides capturing the relation that plausibly obtains between evaluative and non-evaluative properties (see Supervenience), it has recently seemed apt for capturing views about the relations between the mental and the non-mental.

One important doctrine concerning the relationship between mental and physical properties that has been formulated in terms of supervenience is 'externalism'. This is the *denial* of the claim that intentional mental properties (for example, the property of thinking that water is wet) strongly supervene on neurophysiological properties: two people may be exactly alike with respect to their neurophysiological properties, but differ with respect to their intentional ones. This is not surprising in the case of certain, obviously *relational* intentional properties, such as *knowing* that France has no king: two people could be neurophysiological duplicates, but one might live in a world in which France had no king, but another in a world in which it did. Since knowledge entails the truth of what is known, the first person would, but the second would not *know* that France had no king.

More controversially, Hilary Putnam argued for externalism by proposing a thought experiment in which two individuals are exactly alike neurophysiologically but live in different environments, the first in one like ours containing *water*, that is, H_2O, and the other in one containing a superficially similar but *different* substance, XYZ. He concluded that in this scenario the two individuals would be *thinking about* different substances when they assert 'water is wet' (see Content: wide and narrow; Methodological individualism).

Supervenience figures in a more positive way in formulating another important doctrine in contemporary philosophy of mind, 'non-reductive physicalism'. Many philosophers (for example, Davidson 1980) think there are good reasons for supposing that mental properties, such as thinking about water, could not be *identified* or 'reduced to' (for example, defined in terms of) any physical properties, but that, nevertheless, the exemplifications of the mental properties depend on the exemplifications of physical ones. For example, although the property of thinking about water may not be reducible to physical properties, the occurrence of thoughts about water depend on the occurrence of physical events and states (see Anomalous monism; Property theory; Reductionism in the philosophy of mind). (Note that non-reductive physicalism is compatible with externalism since the properties on which intentional mental properties are said to depend - the 'supervenience base' - may include properties other than neurophysiological ones, for example, actual causal interactions with H_2O.)

This strong supervenience of mental properties on physical properties does not entail that mental properties are physical properties, or even that the instantiation of physical properties can provide an illuminating *explanation* of the instantiation of mental ones. Thus, supervenience appears not to entail the reducibility of the mental to the physical (see Reductionism in the philosophy of mind). But it is unclear whether it actually captures the physicalist idea that mental properties *depend* on physical ones. As formulated above, supervenience is compatible with there

being irreducible laws that link mental with physical properties; if so, then there might be different irreducible mental-physical laws in worlds that were otherwise physically identical. One can get closer to the spirit of physicalism by requiring in the above definition that all the laws involved in characterizing nomologically possible worlds include *only* physical laws.

Two significant philosophical questions about supervenience arise. (1) Why should we believe that mental properties supervene on physical properties? (2) What, if anything, explains why mental properties supervene on physical properties? Loewer (1994) suggests that supervenience follows from two reasonable assumptions: that mental properties are causally connected with physical properties, and that physics is causally closed and complete. The latter means that every physical change, if it can be explained at all, can be explained in terms of prior physical events. If mental properties failed to supervene on physical properties then there could be two nomologically possible worlds exactly alike physically, but in one, mental property *M* is instantiated while in the other, *M* is not instantiated. Since physical events in these worlds are fully accounted for by prior events the presence/absence of *M* is, contrary to our assumption, causally irrelevant (see Mental causation).

Although the preceding argument may provide reason to believe that the mental supervenes on the physical, it does not provide any explanation of the supervenience relation itself. Without such an explanation, supervenience may seem to be a mysterious brute fact. The explanations that have been suggested are that mental properties are identical to physical properties, or that they are functional properties that are realized by physical properties (see Functionalism; Mind, identity theory of). While these accounts do explain supervenience, they may be stronger than some advocates of non-reductive physicalism have in mind when they claim that mental properties cannot be reduced to or explained in terms of physical properties. These philosophers may have to accept supervenience as unexplained.

BARRY LOEWER

References and further reading

Blackburn, S. (1993) ‘Supervenience Revisited’, in *Essays in Quasi-Realism*, Oxford: Oxford University Press. (Discusses an argument that some forms of supervenience suggest non-realism of the supervening properties, and that strong supervenience suggests reduction of the supervening properties.)

Davidson, D. (1980) ‘Mental Events’, in *Essays on Actions and Events*, New York: Oxford University Press. (Davidson introduces the notion of supervenience into philosophy of mind to formulate his version of non-reductive physicalism.)

Kim, J. (1993) *Supervenience and Mind*, Cambridge: Cambridge University Press.(The most comprehensive treatments of the use of supervenience to formulate non-reductive physicalism.)

Loewer, B. (1994) ‘An Argument for Strong Supervenience’, in E. Savellos (ed.) *New Essays on Supervenience*, Cambridge: Cambridge University Press.(Argues that the causal completeness of physics and the causal relevance of the mental implies that the mental supervenes on the physical.)

Moore, G.E. (1922) ‘The Conception of Intrinsic Value’, in *Philosophical Studies*, London: Routledge & Kegan Paul.(First expresses the idea that evaluative properties supervene on natural properties.)

Suso, Henry (*c.*1295-1366)

Suso was a Dominican friar and mystic and, with his friend John Tauler, a student of Master Eckhart. The three form the nucleus of the Rhineland school of mysticism. As a lyric poet and troubadour of divine wisdom, Suso explored with psychological intensity the spiritual truths of Eckhart's mystical philosophy. His devotional works were extremely popular in the later Middle Ages.

A native of Swabia, centre of the *Minnesang* and chivalry, Suso became a Dominican novice at thirteen and experienced a radical conversion at eighteen. There followed a decade of intense contemplation, accompanied by severe asceticism and self-torture. During his theological studies at Cologne (1324-8), a centre for Christian Neoplatonism since Albert the Great, he met John Tauler and may have studied under Meister Eckhart. Returning to Constance, he served as lector, then prior at the friary school until 1348, after which he lived in Ulm until his death in 1366. These last decades involved preaching and pastoral care in Dominican convents and in the Beghard and Beguine communities.

Suso's disciple Elsbeth Stagel, a nun at Töss, compiled anecdotes from his life which became the nucleus for *Das Buch von dem Diener (The Life of the Servant)* (1362), the first autobiography in German literature. Suso presents himself as a spiritual knight in pursuit of divine Wisdom. Parts of the *Life* echo Eckhartian themes, but its focus on love and interior experience places this and his other works squarely in the Augustinian tradition of spirituality of the sort found in the writings of Bernard of Clairvaux. Traditional aspects of monastic spirituality, like intense prayer and penance, are blended with features more typical of fourteenth-century hagiography such as raptures, visions and a general emphasis on personal experiences.

His first work, *Das Büchlein der Wahrheit (The Little Book of Truth)* (1328), is an introduction to Eckhart's speculative mysticism and scholastic theology in the form of a dialogue between the disciple (Suso) and Eternal Truth. This didactic use of dialogue as a literary device, the first in German literature, recalls Boethius' *De consolatione philosophiae*. Suso defends Eckhart's views, particularly concerning the absorption of the soul into God, against charges of heresy and explores the spiritual ideal of detachment (*Gelassenheit*). The goal of the spiritual life is to lose oneself in God, but Suso differs from Eckhart in holding that mystical union preserves the distinction between Creator and creature. Moreover, Suso adapts key aspects of Eckhart's Neoplatonic mystical theology to his more devotional spirituality. All things are nothing before the transcendence of the divine being. The effective means to attain God is to transcend all thought and self-will. God as the divine ground of being (*Grund*) is pure simplicity, an 'interior fortress'. Human nature is 'poured out' and God pours in, culminating in the 'breakthrough' (*Durchbruch*) to the divine ground through the Son. Suso connects the soul to God through Eckhart's likening the nothingness of creatures to the abyss of the divine ground, that is, the Godhead beyond God. The remarkable fifth chapter speaks of Suso's own mystical raptures and ecstasies, while the seventh and final chapter explores *Gelassenheit*, the detachment and resignation necessary to attain the highest spiritual state.

His second work, *Das Büchlein der ewigen Weisheit (The Little Book of Eternal Wisdom)* (1327-34), abandons theological speculation completely. In its expanded Latin version, the *Horologium sapientiae (Clock of Wisdom)*, it became the most popular devotional book in the late Middle Ages. This lyrical dialogue, which owes a great deal to the Song of Songs, embodies practical devotion in the form of meditations on the Passion of Christ and on the spiritual value of suffering. The uniqueness of the work derives from its combination of chivalrous imagery with mystical symbolism. The Wisdom pursued by the spiritual knight is also the incomprehensible Good which is utterly inexpressible in language: all creatures 'are swept away into the Good, from whom they emanated'.

Das Briefbüchlein (The Little Book of Letters) (1362), condensed from the larger *Das grosse Briefbuch (The Great Book of Letters)* (*c.*1360), is a devotional manual in the form of pastoral epistles, similar in tone to Seneca's *Moral Epistles* (see Seneca). Suso's poetic gifts transform natural images into spiritual allegories. Suso acts as spiritual guide, offering practical advice to his disciples amid life's tribulations. Here and throughout his writings Suso's psychological insight and poetic subtlety aim at furthering progress on the spiritual path to the ultimate goal.

See also: Meister Eckhart; Mysticism, history of; Neoplatonism; Platonism, medieval

JOHN BUSSANICH

List of works

Suso, Henry (*c.*1295-1366) Works, ed. K. Bihlmeyer, *Heinrich Seuse. Deutsche Schriften*, 1907; trans. F. Tobin, *The Exemplar, with Two German Sermons*, New York: Paulist Press, 1989.(Bihlmeyer is the standard edition of Suso's works in Middle High German. Tobin's translation is accurate, highly readable, and the only complete English translation of the vernacular works from medieval German, with excellent introduction and detailed notes.)

Suso, Henry (1328) *Das Büchlein der Wahrheit (The Little Book of Truth)*, ed. K. Bihlmeyer, *Heinrich Seuse. Deutsche Schriften*, 1907; trans. F. Tobin, *The Exemplar, with Two German Sermons*, New York: Paulist Press, 1989.(Suso's first work, an introduction to Meister Eckhart's mysticism and speculative theology.)

Suso, Henry (1327-34) *Das Büchlein der ewigen Weisheit (The Little Book of Eternal Wisdom)*, ed. K. Bihlmeyer, *Heinrich Seuse. Deutsche Schriften*, 1907; trans. F. Tobin, *The Exemplar, with Two German Sermons*, New York: Paulist Press, 1989.(Devotional work.)

Suso, Henry (1327-34) *Horologium sapientiae (Clock of Wisdom)*, ed. P. Künzle, *Heinrich Seuses Horologium sapientiae,* Freiburg: Universitatsverlag, 1977.(The Latin version of *Das Büchlein der ewigen Weisheit*.)

Suso, Henry (1362) *Das Briefbüchlein (The Little Book of Letters)*, ed. K. Bihlmeyer, *Heinrich Seuse. Deutsche Schriften*, 1907; trans. F. Tobin, *The Exemplar, with Two German Sermons*, New York: Paulist Press, 1989. (Condensed from *Das grosse Briefbuch*, devotional manual in the form of pastoral epistles.)

Suso, Henry (1362) *Das Buch von dem Diener (The Life of the Servant)*, ed. K. Bihlmeyer, *Heinrich Seuse. Deutsche Schriften*, 1907; trans. F. Tobin, *The Exemplar, with Two German Sermons*, New York: Paulist Press, 1989.(Suso's autobiography.)

References and further reading

Haas, A. (1971) *Nim din selbes war. Studien zur Lehre von der Selbsterkenntnis bei Meister Eckhart, Johannes Tauler und Heinrich Seuse,* Freiburg: Universitatsverlag.(Excellent account of the theme of self-knowledge in the Rhineland mystics.)

Kieckhefer, R. (1984) *Unquiet Souls: Fourteenth-Century Saints and Their Religious Milieu*, Chicago, IL: University of Chicago Press.(Excellent discussion of Suso's devotional spirituality and personal mysticism within the context of late medieval spirituality.)

Filthaut, E.M. (ed.) (1966) *Seuse-Studien: Heinrich Seuse. Studien zum 600. Todestag, 1366-1966*, Cologne: Albertus Magnus Verlag.(Excellent collection of scholarly articles on spiritual, textual and literary aspects of Suso's works.)

Suárez, Francisco (1548-1617)

Francisco Suárez was the main channel through which medieval philosophy flowed into the modern world. He was educated first in law and, after his entry into the Jesuits, in philosophy and theology. He wrote on all three subjects. His philosophical writing was principally in the areas of metaphysics, psychology and philosophy of law, but in both his philosophical and theological works he treated many related epistemological, cosmological and ethical issues. While his basic outlook is that of a very independent Thomist, his metaphysics follows along a line earlier drawn by Avicenna (980-1037) and Duns Scotus (1266-1308) to treat as its subject 'being in so far as it is real being'. By the addition of the word 'real' to Aristotle's formula, Suárez emphasized Aristotle's division of being into categorial being and 'being as true', as well as Aristotle's exclusion of the latter from the object of metaphysics. Divided into a general part dealing with the concept of being as such, its properties and causes, and a second part which considers particular beings (God and creatures) in addition to the categories of being, Suárez's metaphysics ends with a notable treatment of mind-dependent beings, or 'beings of reason'. These last encompass negations, privations and relations of reason, but Suárez's treatment centres on those negations which are 'impossible' or self-contradictory. Inasmuch as such beings of reason cannot exist outside the mind, they are excluded from the object of metaphysics and relegated to the status of 'being as true'. In philosophy of law he was a proponent of natural law and of a theory of government in which power comes from God through the people. He was important for the early development of modern international law and the doctrine of just war. While his brand of Thomism was opposed in his own time and after by some scholastics, especially Dominicans, he had great authority among his fellow Jesuits, as well as other Catholic and Protestant authors. Outside scholasticism, he has influenced a variety of modern thinkers.

1 Life and works

Francisco Suárez was born at Granada in Spain. After preparatory studies, in 1561 he enrolled at the University of Salamanca, where he studied law until on 16 June 1564 he entered the Society of Jesus (the Jesuits). Three months later he began the study of philosophy. In October 1566, still at Salamanca, he went on to theological studies. Chief among his mentors in these was the Dominican, Juan Mancio (1497-1576), himself a successor of Francisco de Vitoria (*c.*1486-1546) in Salamanca's principal chair of theology.

Following his theological studies, Suárez began to teach philosophy in 1570, initially at Salamanca and then at Segovia. Ordained a priest in 1572, he continued to lecture in philosophy until, in 1574, at the Jesuit College, Valladolid, he began his main life's work as a theology teacher. Later he taught theology at Avila (1575), Segovia (1575), Valladolid again (1576), Rome (1580), Alcalá (1585) and Salamanca (1593). In 1597, he assumed the principal chair of theology at the University of Coimbra, in which he remained until his retirement in 1615.

Besides teaching, Suárez took part in theological and political disputes. The most famous of these was the debate *De auxiliis* (On the Helps [for Salvation]), between sixteenth-century Jesuits and Dominicans. The controversy concerned God's foreknowledge and causality, grace and human freedom. In this debate Suárez, along with fellow Jesuits, Robert Bellarmine (1542-1621) and Luis de Molina (1535-1600), allowed for divine prerogatives but championed human free will. On the core question of conditional future contingents (*futuribilia*), Suárez adopts Molina's doctrine of a 'middle knowledge' that falls between God's knowledge of merely possible things and God's knowledge of things other than himself that are, were, or will be actual. However, he modifies Molina's view that God knows all things other than himself by a 'super-comprehension' of them in his own essence, substituting for this a view that God knows these things immediately in themselves as they exist in a purely intentional or objective way. More political was a quarrel between the republic of Venice and the papacy about the limits of papal jurisdiction. In defence of the papal position, Suárez in 1607 composed a treatise, *De immunitate ecclesiastica a Venetis violata (On the Ecclesiastical Immunity Violated by the Venetians)*. Commending his effort, Pope Paul V stated that the work revealed its author to be 'an eminent and pious theologian'. From this came the honorific title with which Suárez passed into history - *Doctor eximius ac pius* (Eminent and Pious Teacher).

The majority of Suárez's writings are theological, often corresponding to specific parts of the *Summa theologiae* of Thomas Aquinas (§6), but there is also the extremely influential *De legibus seu de Deo legislatore (On Laws or on God the Legislator)*, the fruit of Suárez's teaching between 1601 and 1603, published at Coimbra in 1612. Outside

the Thomistic framework are the work on ecclesiastical immunity mentioned above and, at Coimbra in 1613, the *Defensio fidei catholicae adversus anglicanae sectae errores, cum responsione ad apologiam pro juramento fidelitatis et praefationem monitoriam serenissimi Jacobi Angliae Regis (Defence of the Catholic Faith against the Errors of the Anglican Sect, with a Reply to the 'Apology' for the 'Oath of Fidelity' and the 'Warning Preface' of James, the Most Serene King of England)*. Upon its publication, this work, which details Suárez's political philosophy, was condemned by James I and publicly burned in London, because in it Suárez had opposed the absolute right of kings and defended the indirect power of the papacy over temporal rulers, as well as the right of the citizens to resist a tyrannical monarch. Suárez even defended tyrannicide in the case of a monarch deposed for heresy by the pope.

2 Metaphysics

Also outside the Thomistic framework are the two volumes of Suárez's most important work, the *Disputationes metaphysicae (Metaphysical Disputations)*, which first appeared at Salamanca in 1597. A résumé of his own and previous thought on myriad questions, it is arranged in the form of fifty-four 'Disputations' dealing systematically with metaphysics. In this format, Suárez's volumes mark a radical departure from previous metaphysical treatises, which usually had been either short works, such as Aquinas' *De ente et essentia (On Being and Essence)*, or commentaries on the text of Aristotle.

The *Disputationes metaphysicae* is a feat of learning. After stating each problem, Suárez has searched the history of philosophy and theology for solutions offered to it. Every conceivable Greek, Arabic, patristic and especially scholastic writer seems to have been cited at least once. As many as twenty-two opinions have been cited in connection with a single question. The historian J. Iturrioz (see Martín, Ceñal, Hellin *et al.* 1948) has compiled a list of 7,709 of these citations, which refer to 245 different authors. Of these, Aristotle was mentioned most often (a total of 1,735 times), and Aquinas the next most (cited 1,008 times). No mere compiler of opinions, Suárez was an independent thinker who, as he reported positions and gave due regard even to those he opposed, was intent on presenting his own metaphysics.

The disputations begin with Suárez telling us that the object of metaphysics is 'being in so far as it is real being' (*Disputationes metaphysicae* 1). To explain this, he employs two distinctions, already familiar among scholastic authors. The first of these is between the 'formal concept' as an act of the mind, and the 'objective concept' as what that act immediately intends as its object. This latter may be some individual thing or some common or universal character (*ratio*). Again, it may be something actual, possible or merely objective. The second distinction falls between 'being (*ens*) taken as a participle' (which refers to an actual existent), rather than 'being taken as a noun' (which refers to whatever is not a mere fiction but is true in itself and apt really to exist). The object of metaphysics is then more exactly identified with the 'common objective concept of being as a noun' (*Disputationes* 2). This object, which reflects Avicenna's interpretation of Aristotle's *Metaphysics*, prescinds from existence and, precisely as common, transcends all genera, species and differences to encompass everything real, from extrinsic denominations (such as 'being right', 'being left', 'being known' or 'being willed'), through mere possibles (which at their core reduce to non-contradiction), to actual created substances and accidents, to the subsistent, purely actual and necessary reality of God. Over the range of such beings, the common concept of being is analogous with 'an analogy of intrinsic attribution'. In this analogy, a unified concept is shared in an ordered way by different beings (God and creatures, substance and accidents) inasmuch as the being of what is posterior depends upon the being of what is prior (see Language, Renaissance philosophy of §4).

After a general treatment (*Disputationes* 3) of the transcendental properties of every being as it is a being - namely, unity, truth and goodness - questions are raised under unity about individuation, universal natures, and various kinds of distinction (*Disputationes* 5-7). Rejecting the Aristotelian and Thomistic account of individuation in terms of 'quantified matter', as well as Scotistic 'thisness', Suárez comes close to nominalism and says that every thing is individual by its very entity. On universals, he again leans towards nominalism and denies any real common nature independent of individuals, yet he insists that the universalizing activity of the mind has a foundation in the likenesses of things. Restricting distinctions to real, rational and modal, he rejects the formal distinction, which Duns Scotus asserted to be present between 'formalities' prior to any operation of the intellect. In following disputations, discussion of truth, which centres on the conformity of formal with objective concepts, is balanced by discussion of falsity (*Disputationes* 8-9); and discussion of goodness is balanced by that of evil, which in standard

scholastic fashion is regarded as a privation of good (*Disputationes* 10-1). The twelfth disputation generally treats causes, while 13-25 deal specifically with material, formal, efficient and final causes. Concluding this first part is a consideration of causes in comparison with their effects and in relation one to another (*Disputationes* 26-7).

The second part opens with the main division and principal analogy of being between infinite and finite (*Disputationes* 28). The existence of God is then causally demonstrated in an expressly metaphysical way which, again reflecting Avicenna, employs the principle, 'Everything which comes to be, comes to be by another' (*Disputationes* 29) and follows the analogous common concept of being from lesser and lower being to a First Being. In the same disputation, Suárez rejects any physical demonstration, like that of Aristotle, adopted by Averroes and Aquinas, which would employ the principle, 'Everything which is moved is moved by another', to pass from motion to a First Mover. Suárez goes on to investigate the perfection, simplicity, immensity, immutability, wisdom and omnipotence of God (*Disputationes* 30). Disputation 31 begins treatment of finite being with a denial of the Thomistic distinction in creatures between essence and existence, which Suárez understands as falling between two 'things' (*res*). Rejecting Scotistic formal or modal distinctions, Suárez says that the only distinction here is one of reason with a foundation in reality. Disputation 32 considers substance and accidents in general, plus the analogy of being between created substances and accidents. Substance is treated in detail through the next four disputations while the different categories of accident are the subject matter of Disputations 37-53. Of particular interest is Disputation 47, where Suárez treats real relations, which divide into transcendental and predicamental. Transcendental relations - such as those between matter and form, accidents and substance, creatures and God - transcend the lines of categories, whereas predicamental relations are restricted to the accidental category of relation. Every such relation (for example, one of likeness) involves a real subject (one person), a real terminus (another person), and a real foundation (brown hair), but Suárez maintains in nominalist fashion that a predicamental relation is ultimately identical with its foundation, as this exists in both the subject and the terminus. The work concludes with a discussion of 'beings of reason', which divide into negations (including impossible objects), privations and reason-dependent relations - all of which fall outside real being, the object of metaphysics (*Disputationes* 54).

3 Philosophical psychology

While Suárez treated topics like abstraction, universal knowledge and personhood also within metaphysics, he regarded philosophical psychology as properly a part of physics (that is, the Aristotelian philosophy of nature). His work *De anima (On the Soul)*, which he was revising at the time of his death and which was again a systematic presentation rather than itself a commentary, returned to the medieval tradition of commentary on Aristotle's *On the Soul*. Divided into six books, it dealt successively with the nature, attributes, faculties, operations and ultimate status of the rational soul, which in the human composite is related to the body as form to matter.

The only substantial form of a human being, the rational soul is principle for a variety of vegetative, sensitive and intellectual activities (*De anima* I). The distinction among such activities indicates specifically distinct powers (vegetative, sensitive and intellectual) from which they immediately stem, as well as a real distinction between such powers and the soul itself (II c.1-2). Differing from the whole scholastic tradition, Suárez unites the common sense, imagination, memory and estimative sense in one internal sense power (III c.30). At the intellectual level, he distinguishes between the adequate object of the intellect as such (that is, being as being, which embraces all intelligible items), and the object proportionate to the human intellect in its present state (that is, sensible or material things). He also allows for a direct and immediate knowledge of the singular (IV c.1-2). The excellence of its intellectual operations indicates the intrinsic spirituality and the transcendence of the soul over the body (I c.9).

For the immortality of the soul, Suárez offers two main arguments (*De anima* I c.10). First, inasmuch as the human intellective power is incorporeal and spiritual it is also substantially simple. Therefore, of its nature it is incorruptible and thus immortal. Second, under the justice of God evil persons should be punished and good persons rewarded. However, that often does not happen in this life. Therefore, there must be another life in which it can happen and this can only be if the soul is immortal.

Both in this life and in any afterlife the soul alone without the body would be less than a full human being. To designate it as such, Suárez accepts the bold word of Cajetan (§§3-4): the soul by itself is a 'semi-person' (*De anima* VI c.1). This incomplete character of the soul alone may suggest a philosophical suasion for the resurrection of the body (VI c.9) (see Aristotelianism, Renaissance §§4-5).

4 Philosophy of law and society

For Suárez all law stems from the 'eternal law', which is 'a free decree of the will of God establishing the order to be observed either generally by all parts of the universe in relation to the common good… or especially to be observed by intellectual creatures in their free operations' (*De legibus* II c.3). In so emphasizing God's will, Suárez has distinguished himself from Aquinas who identified the eternal law with divine reason as it governs the created universe.

Immediately derived from the eternal law is the natural law which resides in human minds and enables human beings to discern moral good and evil (see Natural law). Suárez (*De legibus* II c.7) tells us that natural law first embraces general moral principles like: 'Good must be done and evil avoided' and 'Do not do to anyone else what you would not want done to you'. Next come principles more particular but still evident from their terms, such as: 'Justice should be observed', 'God should be worshipped', 'One should live with self-control', and so on. From these principles come conclusions more or less easily and broadly known, for example, that such things as adultery and theft are wrong. Requiring more reasoning and not easily known to all are conclusions like: fornication is intrinsically evil, usury is unjust, or lying is never justified.

While it may be affected extrinsically because of changed circumstances, no true principle of natural law can be diminished or dispensed from by any human law or power (*De legibus* II c.14 n.5). Rejecting both the opinion of William of Ockham, that God can dispense from the entire Decalogue and could indeed abrogate the whole natural law, and the opinion of Duns Scotus, that God can dispense from the precepts in the second table of the Decalogue (which regard human beings or other creatures), Suárez maintains that the whole Decalogue is indispensible even by the absolute power of God (c.15 n.16). Thus, even though it is rooted in the divine will, law is not arbitrary either for men or for God.

Close on natural law, the 'law of nations' (*jus gentium*) has the general character of positive law but differs from the civil law of particular states. Unwritten, it has been established not by a single state but by the customs of almost all nations (*De legibus* II c.19 n.7). It thus originates in human consensus and it can in principle, though not with ease, be changed.

In one sense, the *jus gentium* is 'the law of nations among themselves (*inter se*)' - a law which different nations are obliged to observe *vis-à-vis* one another. Such items as the immunity of ambassadors and free commerce, as well as the 'right of war' (*jus belli*), belong to the *jus gentium* understood in this way. In a second way, it is 'the law of nations within themselves (*intra se*).' This is the law which individual states commonly observe within their own borders. Items such as the division of goods or possessions, private property, buying and selling, and the use of money, belong to the *jus gentium* taken in this second way. The first way, which is effectively an international law, is most properly called 'the law of nations' (*De legibus* II c.19 n.10).

Like other Catholic thinkers of his day, Suárez did not regard war as intrinsically evil (see War and peace, philosophy of §2). Although it was deplorable and should be avoided wherever possible, at times war was the only option open for the preservation of the republic, which has a right and even an obligation to defend itself (*De legibus* II c.18 n.5). That any war be 'just', proper authority (legitimate, public and supreme) was required to declare it. Again, a just cause of sufficient gravity was needed. Furthermore, right conduct should be the rule at the beginning of the war, in its prosecution and in victory afterwards. As regards authority to declare and wage it, war may be based upon one state's right to punish or avenge an injury done to it by another. Each state, 'supreme in its own order' - that is, the temporal order - with no tribunal beyond, has the authority forcibly to redress injuries against itself (c.19 n.8).

Because civil power as such is not greater in Christian than in pagan princes, Christians can have no more reason than pagans for a just war. Neither can Christians make war against pagans solely because these lack faith. By Suárez's time, questions raised by the evangelization of the Native Americans were largely settled for Catholic thinkers. From Vitoria on, it was commonly held that they were human beings, masters of their own lives and possessions, and that it was not lawful without just cause to subjugate and despoil them - even in order to Christianize them. This was also Suárez's position, which he took, however, on an abstract level, almost without mention of the Native Americans. Against a possible application to them of Aristotle's division of human beings into those fitted by nature to rule and those who were by nature 'slaves' (*Politics* 1.5.1254a18-1255a2), Suárez's

view was simple and direct. It is incredible to say that all the people in any region or province have been born 'monstrous and in a way that contradicts the natural disposition' of human beings to be free (*De justitia* q.6). In fact, all persons, as made in the image of God, are equally capable of dominion over themselves and their possessions (*Defence of the Catholic Faith* III c.1).

At its origin, the state is natural; but it is also voluntary (see State, the). Free persons, naturally inclined to political association, must still agree to it. Hence, the state itself arises out of a contract, or 'a consensus', explicit or tacit, freely entered into by the community (*De legibus* III c.3). People are not forced by nature to choose any particular form of state and in fact different forms exist, with a natural equality among them all. In practice, the best kind of government is some form of monarchy (c.4). But what form a monarchy takes and how much power any monarch has will depend upon the terms of the initial grant of the people (*Defensio fidei catholicae* III c.2). Thus, civil authority or power, in any form, is ultimately from nature, and the God of nature, but immediately from the people.

Even though each state is 'supreme in its own order', the power and laws of one end where those of another begin. Again, while the state has power to enact laws from which there is no appeal to any other earthly tribunal, the power of any state, even within its own territory, is not absolute. Though ordinarily their gift of power is irrevocable, in principle the people retain authority over their government. Accordingly, as the common good demands, political power, even that of a monarch, may in different circumstances be changed or limited (*Defensio fidei catholicae* III c.3).

Other restrictions on state power occur inasmuch as it stops short at the private zone of families and individuals. For these are by nature prior to the state. In addition, human beings are not just citizens of this world. And while he admits a legitimate concern by the republic for its members' morality, Suárez tells us that, even at a natural level, each person aims at a final happiness which transcends the reach of civil law and power (*De legibus* III c.11).

While temporal power is ultimately from God, immediately its origin is natural and human; in contrast, ecclesiastical power is directly of divine origin, 'from the special promise and grant of Christ' (*Defensio fidei catholicae* III c.6 n.17). Although church and state are each 'supreme in [their] own order', the basic relation between the two is hierarchical, comparable to that between the soul and the body (*De legibus* IV c.9 n.3). The power of the many states of this world is directly and exclusively within the temporal order, aiming at a common temporal good of 'political happiness' (I c.13). The power of the one Christian church, pointed towards the eternal salvation of its members, is directly within the spiritual order. Indirectly, however, that church has power over a Christian state even in temporal matters (III c.6). Conversely, civil power, at least in Christian states, is indirectly dependent upon and should be at the service of the higher goal of the church (*Defensio fidei catholicae* III c.5). Suárez acknowledged in this a certain inequality between Christian and infidel princes, but he counts it a plus for Christian states and sovereigns that their power is raised to a new height in its subordination to the church (c.30) (see Political philosophy, history of §8).

5 Influence

Suárez exercised wide and deep influence on post-Renaissance Catholic scholasticism. Also, largely through the growth and agency of the Jesuit Order, Suarezian metaphysics spread from the Catholic schools of Iberia to various northern European locales. It penetrated the Lutheran universities of Germany where the *Disputationes metaphysicae* (of which seventeen editions appeared between 1597 and 1636) was pondered especially by those who preferred Melanchthon's attitude towards philosophy to Luther's. Indeed, in a number of seventeenth-century Lutheran universities it served as a textbook in philosophy. In much the same way, Suárez had major influence in the Reformed tradition of German and Dutch schools for both metaphysics and law, including international law. To sense Suárez's importance here for the provenance of modern international law, it is enough to recall that, for the famous Dutch jurisprudent, Hugo Grotius, the Jesuit doctor was a philosopher and theologian of such penetration 'that he hardly had an equal' (*Epistola* CLIV; quoted in Scorraille 1912-3, vol. 2: 437).

As Martin Heidegger later saw it, Suárez was the main source through which Greek ontology passed from the Middle Ages to usher in the metaphysics and the transcendental philosophy of modern times. Most likely, Suarezian metaphysics was that first learned by Descartes from his Jesuit teachers at La Flèche. On at least one occasion he refers to the *Disputationes*, of which he is believed to have owned a copy. Leibniz boasted that while

yet a youth he had read Suárez 'like a novel'. Schopenhauer, in his chief work, *Die Welt als Wille und Vorstellung (The World as Will and Representation)*, displays much acquaintance with the *Disputationes*, which he values as 'an authentic compendium of the whole scholastic tradition' (quoted in Grabmann 1926: 535). Similarly, Franz Brentano, in his 1862 work on the manifold meaning of being according to Aristotle (*Von der mannigfachen Bedeutung des Seienden nach Aristoteles*), has recommended the *Disputationes metaphysicae* to anyone who wants to learn of the diversity of medieval views on Aristotle. But perhaps most strikingly, for Christian von Wolff (whose *Ontologia* Immanuel Kant thought practically coterminous with pre-critical metaphysics) it was 'Francisco Suárez, of the Society of Jesus, who among scholastics pondered metaphysical questions with particular penetration' (*Philosophia prima sive ontologia* I.2.3 n.169; quoted in Gilson 1952: 117).

See also: Aquinas, T.; Aristotelianism, Renaissance; Collegium Conimbricense; International relations, philosophy of; Molina, L. de; Natural law; War and peace

JOHN P. DOYLE

List of works

Suárez, F. (1590-1617) *Opera omnia (Complete Works)*, Paris: Louis Vivès, 1856-66, 26 vols; two supplementary volumes of indices, 1878.(This is the latest and most accessible edition of Suárez's complete works.)

Suárez, F. (1584) *De justitia et jure (On Justice and Law)*, ed. J. Giers, in *Die Gerechtigkeitslehre des jungen Suárez: Edition und Untersuchung seiner Römischen Vorlesungen 'De Iustitia et Iure'*, Freiburg im Breisgau, 1958.(Suárez's early lectures at Rome on justice and law; cited in §4.)

Suárez, F. (1597a) *Disputationes metaphysicae (Metaphysical Disputations)*, in F. Suárez, *Opera omnia*, vols 25-6, Paris: Louis Vivès, 1856-66; repr. Hildesheim: Olms, 1965.(Suárez's systematic presentation of metaphysics.)

Suárez, F. (1597b) *Disputatio V: Individual Unity and its Principle*, trans. J.J.E. Gracia, in *Suárez on Individuation*, Milwaukee, WI: Marquette University Press, 1982.(In addition to its own intrinsic worth, this disputation is important for its possible influence on the young Leibniz.)

Suárez, F. (1597c) *On Formal and Universal Unity (Disputatio VI)*, trans. J.F. Ross, Milwaukee, WI: Marquette University Press, 1964.(This disputation contains Suárez's teaching on the problem of the universals which situates him between medieval nominalism and realism.)

Suárez, F. (1597d) *On the Various Kinds of Distinctions (Disputatio VII)*, trans. C. Vollert, Milwaukee, WI: Marquette University Press, 1947.(In connection with Suárez's own theory, this disputation presents an overview of medieval doctrines of distinction.)

Suárez, F. (1597e) *The Metaphysics of Good and Evil according to Suárez: Metaphysical Disputations X and XI and Selected Passages from Disputation XXII and Other Works*, trans. J.J.E. Gracia and D. Davis, München: Philosophia, 1989.(These disputations contain Suárez's basic metaphysical optimism, and the translation contains a good introduction plus a very useful glossary.)

Suárez, F. (1597f) *On Efficient Causality: Metaphysical Disputations 17, 18, and 19*, trans. A.J. Freddoso, New Haven, CT: Yale University Press, 1994.(These disputations give an excellent overview of the medieval Aristotelian tradition as well as Suárez's own teaching on efficient causality.)

Suárez, F. (1597g) *On the Essence of Finite Being as Such, on the Existence of that Essence and their Distinction (Disputatio XXXI)*, trans. and intro. N.J. Wells, Milwaukee, WI: Marquette University Press, 1983.(Contains Suárez's famed denial of the real distinction of essence and existence in creatures.)

Suárez, F. (1597h) *On Beings of Reason (De Entibus Rationis): Metaphysical Disputation LIV*, trans. and intro. J.P. Doyle, Milwaukee, WI: Marquette University Press, 1995.(This disputation was very influential for the development of late scholastic intentionality theory.)

Suárez, F. (1607) *De immunitate ecclesiastica a Venetis violata (On the Ecclesiastical Immunity Violated by the Venetians)*, in J.B. Malou (ed.) *Opuscula sex inedita*, Brussels and Paris: Greuse, 1859.(Treatise on the superiority of papal authority over that of the Venetian republic.)

Suárez, F. (1612) *De legibus seu de Deo legislatore (On Laws or on God the Legislator)*, in L. Pereña *et al.* (eds) *Corpus Hispanorum de Pace*, vols 11-7, 21-2, Madrid: Consejo Superior de Investigaciones Cientificas, 1971-81.(A Latin-Spanish edition of *De legibus*, with good commentary and excellent footnotes.)

Suárez, F. (1612-21) *Selections from Three Works of Francisco Suárez, S.J.*, trans. G.L. Williams *et al.*, in J.B. Scott (ed.) *Classics of International Law* 20, Oxford: Clarendon Press, 1944, 2 vols.(Translates passages from the *De legibus*, *Defensio fidei*, and a section from the treatise 'On Charity' which contains Suárez's teaching on war.)

Suárez, F. (1613) *Defensio fidei catholicae adversus anglicanae sectae errores, cum responsione ad apologiam*

pro juramento fidelitatis et praefationem monitoriam serenissimi Jacobi Angliae Regis (Defence of the Catholic Faith against the Errors of the Anglican Sect, with a Reply to the 'Apology' for the 'Oath of Fidelity' and the 'Warning Preface' of James, the Most Serene King of England), Coimbra; in F. Suárez, *Opera omnia*, vol. 24, Paris: Louis Vivès, 1856-66.(Outlines Suárez's political philosophy; the work's opposition to absolute monarchy lead to its public condemnation in England.)

Suárez, F. (1614) *Defensio fidei III. Principatus politicus o la soberania popular (Defence of Faith III: the Political Supremacy of Popular Sovereignty)*, trans. and ed. E. Elorduy and L. Pereña, Madrid: Consejo Superior de Investigaciones Cientificas, 1965.(A Latin and Spanish edition of part of the *Defensio fidei*, with good introduction and notes.)

Suárez, F. (1614) *De juramento fidelitatis (On the Oath of Fidelity)*, trans. and ed. L. Pereña, V. Abril and C. Baciero, with A. Garcia and C. Villanueva, in *Corpus Hispanorum de Pace*, vol. 19, Madrid: Consejo Superior de Investigaciones Cientificas, 1978.(A Latin and Spanish critical edition of part of the *Defensio fidei*, with notes, bibliography and indices.)

Suárez, F. (1620-1) *De anima (On the Soul)*, ed. and intro. S. Castellote, foreword by X. Zubiri, trans. into Spanish by C. Baciero and L. Baciero, vols 1-3, Madrid: Sociedad de Estudios y Publicaciones, 1978-91.(A critical Latin edition, with Spanish translation, of Suárez's *De anima* which incorporates the manuscript changes left by Suárez himself.)

References and further reading

Alejandro, J.M. (1948) *La gnoseología del Doctor Eximio y la acusación nominalista (The Epistemology of the 'Eminent Doctor' and the Accusation of Nominalism)*, Comillas: Pontificia Universitas Comillensis.(A defence of Suárez's epistemology against the accusation of nominalism.)

Aristotle (1995) *Politics*, trans. T.J. Saunders, Oxford: Clarendon Press.(Referred to in §4. This edition gives a translation of the first two books of Aristotle's *Politics* and a commentary.)

Courtine, J.-F. (1990) *Suárez et le système de la métaphysique (Suárez and the Plan of Metaphysics)*, Paris: Presses Universitaires de France.(Situates 'the Suarezian moment' in the history of metaphysics as the point from which an Aristotelian doctrine of transcendental being passes to a modern concern for pure objectivity.)

Cronin, T. (1966) *Objective Being in Descartes and in Suárez*, Rome: Gregorian University Press.(Suárez's doctrine of objective being is treated in relation to its scholastic sources and for its influence on Descartes.)

Davitt, T. (1960) *The Nature of Law*, St. Louis, MO: Herder, 1960, 86-108.(Presents Suárez's doctrine of law as voluntaristic.)

Doyle, J.P. (1967) 'Suárez on the Reality of the Possibles', *The Modern Schoolman* 44: 29-40.(Highlights the 'double negative' - that is, non-contradictory - character of possible being according to Suárez.)

Doyle, J.P. (1969) 'Suárez on the Analogy of Being', *The Modern Schoolman* 46: 219-49, 323-41.(Treats Suarezian analogy of being as it presupposes the common objective concept of being as a noun.)

Doyle, J.P. (1972) 'Heidegger and Scholastic Metaphysics', *The Modern Schoolman* 49: 201-20.(Contrasts Suarezian and Thomistic metaphysics in view of Heidegger's 'the forgetfulness of being' charge against medieval metaphysics.)

Doyle, J.P. (1982) 'The Suarezian Proof for God's Existence', in L.J. Thro (ed.) *History of Philosophy in the Making: A Symposium of Essays to Honor Professor James D. Collins on his 65th Birthday*, Washington, DC: University Press of America, 105-17. (Treats the Suarezian metaphysical proof for God's existence as it presupposes the common objective concept of being as a noun.)

Doyle, J.P. (1984) 'Prolegomena to a Study of Extrinsic Denomination in the Work of Francis Suárez, S.J.', *Vivarium* 22 (2): 121-60.(An introductory study of the minimal being falling under the common objective concept of being as a noun.)

Doyle, J.P. (1987-8) 'Suárez on Beings of Reason and Truth', *Vivarium* 25 (1): 47-75, 26 (1): 51-72.(Treats Suárez's doctrine of truth with regard to beings of reason, especially impossible objects.)

Doyle, J.P. (1992) 'Francisco Suárez: On Preaching the Gospel to People like the American Indians', *Fordham International Law Journal* 15 (4): 879-951.(Treats Suárez's legal and political theory, including his just war doctrine, within the context of the Spanish discovery and conquest of the New World.)

Gilson, E. (1952) *Being and Some Philosophers*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 2nd edn, 96-107.(Presents Suárez as developing the 'essentialism' which he inherited from the metaphysics of Avicenna and Duns Scotus. Referred to in §5.)

Grabmann, M. (1926) 'Die Disputationes Metaphysicae des Franz Suárez in ihrer methodischen Eigenart und Fortwirkung' (The *Disputationes Metaphysicae* of Francis Suárez: its Methodical Character and Influence), in *Mittelalterliches Geistesleben: Abhandlungen zur Geschichte der Scholastik und Mystik*, Munich: Max Hueber,

vol. 1, 525-60.(This work treats the context, the character and the division of the *Metaphysical Disputations*, as well as its influence from the date of its appearance up to the 20th century. Referred to in §5.)

Gracia, J.J.E. (ed.) (1991) 'Francisco Suárez', special issue of *American Catholic Philosophical Quarterly* 65 (3). (Contains articles on various aspects of Suárez's work.)

Hoeres, W. (1965) 'Francis Suárez and the Teaching of John Duns Scotus on *Univocatio entis*', in J.K. Ryan and B.M. Bonansea (eds) *John Duns Scotus (1265-1965)*, Washington, DC: Catholic University Press, 263-90. (Compares and contrasts the doctrines of Suárez and Scotus on the concept of being.)

Iriarte, J. (1948) 'La proyección sobre Europa de una gran metafísica, o Suárez en la filosofía de los días del barroco' (The spread over Europe of a great metaphysics…), *Razón y fe* 138: 229-65.(Treats the spread of Suarezian metaphysics through Europe in the seventeenth century.)

Martín, J.I., Ceñal, R., Hellin, J., Elorduy, E., Iturrioz, J., Rommen, H. *et al.* (1948) 'Suárez en el cuarto centenario de su nacimiento (1548-1617)' (Suárez on the Fourth Centenary of His Birth), special issue of *Pensamiento* 4.(An important collection of articles on Suárez commemorating the four-hundredth anniversary of his birth, with a laudatory introduction by the then Spanish minister of education, José Ibañez Martín. Referred to in §2.)

Robinet, A. (1980) 'Suárez dans l'oeuvre de Leibniz' (Suárez in the Work of Leibniz), *Cuadernos Salmantinos de Filosofía* 7: 269-84.(A cautious consideration of the possible influence of Suárez on Leibniz.)

Rommen, H. (1926) *Die Staatslehre des Franz Suárez, S.J. (The Political Doctrine of Francisco Suárez)*, München-Gladbach: Volksverein-Verlag.(An exposition of Suárez's views on political society including the relation between church and state.)

Ruiz Moreno, I. (1949) 'El derecho internacional y Francisco Suárez' (International Law and Francisco Suárez), in *Actas del IV Centenario del Nacimiento de Francisco Suárez (1548-1948)*, Burgos, 1949, vol. 2, 331-63. (Treats the role of Suárez in the development of international law.)

Scorraille, R. de (1912-3) *François Suárez de la Compagnie de Jésus (Francisco Suárez of the Society of Jesus)*, Paris: Lethielleux, 2 vols.(The definitive biography of Suárez. Referred to in §5.)

Wells, N.J. (1962) 'Suárez, Historian and Critic of the Modal Distinction between Essential Being and Existential Being', *The New Scholasticism* 36: 419-44.(An exposition and criticism of Suárez's view of one proposed distinction between essence and existence in creatures.)

Wells, N.J. (1980-1) 'Suárez on the Eternal Truths, I and II', *The Modern Schoolman* 58: 73-104, 159-74.(Critical presentation of Suárez's understanding of the eternal status of necessary truths.)

Wells, N.J. (1984) 'Material Falsity in Descartes, Arnauld, and Suárez', *Journal of the History of Philosophy* 22: 25-50.(Treats a particular doctrine as it passed into modern philosophy from Suárez through Descartes.)

Wundt, M. (1939) *Die deutsche Schulmetaphysik des 17 Jahrhunderts (The German 'School-metaphysics' of the 17th Century)*, Tübingen: Mohr.(Treats the influence of Suárez on seventeenth-century Lutheran and Calvinist thinkers in Germany and Holland.)

Swedenborg, Emanuel (1688-1772)

Swedenborg was an eighteenth-century Swedish dignitary of considerable learning who believed that he had the power to communicate with spirits and angels and that these beings would help him fulfil the task allotted to him by God, namely to reveal the hidden meaning of Scripture and to usher in the new Church. His thought attracted the critical attention of no less a figure than Immanuel Kant.

Swedenborg was born in Stockholm to a prominent Lutheran cleric. At the age of eleven, he entered the University of Upsala, where he developed a keen interest in mathematics and natural philosophy. He was to pursue scholarly interests for the rest of his life. In 1716, Charles XII made him an honorary appointment to the Swedish Board of Mines. Swedenborg rose to the salaried position of Assessor Ordinary. Under this title, Swedenborg served the King with distinction in many different capacities, including that of statesman, engineer and geologist.

Swedenborg retired from the Board in 1747 to devote himself to biblical exegesis. He believed that God had given him the power to communicate with angels and spirits so that he could reveal the hidden meaning of Scripture. In 1756, he published anonymously the *Arcana coelestia*, a massive line-by-line commentary on Genesis and Exodus.

Stories about Swedenborg's prophetic gift began to spread throughout Europe at this time. In 1756, during a dinner party in Gotenborg, Swedenborg claimed to see a fire spreading in Stockholm fifty miles away. He described the catastrophe in such minute detail that his vision could be checked against reports from the stricken capital which arrived in Gotenborg shortly thereafter. These reports apparently confirmed Swedenborg's vision on every point. Swedenborg had now made his name as a visionary. More stories added to his fame and excited interest in the *Arcana coelestia*, which was now known to be his work.

The problem is what to make of Swedenborg. Was he a prophet, a lunatic, or a man with questions too difficult to answer by the usual methods of scientific enquiry? Whatever else we might think, it is important to understand that Swedenborg saw himself as a Church reformer. He believed that the Church had lost sight of a fundamental truth, namely that God is one and operates through all of creation. The Nicene Council was the first indication of spiritual myopia, because it ruled that there have been three distinct divine persons from eternity. According to Swedenborg, this was to make three different gods out of the Father, Son and Holy Spirit (see Trinity). Worse, it encouraged us to represent the Trinity as a council of human dignitaries, or even as a gang of thugs such as the Triumvirate at Rome. The polytheism of the Nicene Council and its attendant heresies were transmitted to the Roman Catholic Church, and from that Church to all the reformed Churches of Swedenborg's day. Swedenborg's mission was to right such wrong thinking and to prepare us for salvation in the new Church. His strategy was simple: he had only to describe his encounters with angels and spirits. Thus he reported how angels set him on the right path and how the spirits of churchmen revealed themselves in an otherworldly burlesque show as the buffoons they had been all along. Swedenborg found followers, not least the young William Blake. The Swedenborgian church is alive to this day.

Swedenborg thought that philosophers had as little spiritual insight as churchmen. Since they could not see spiritual things, as Swedenborg did, they could never formulate a true metaphysics of immaterial substances. So Swedenborg complained that metaphysics is all conjecture. He illustrated his complaint with reports of his discussions with the spirits of metaphysicians. Once, he said, they discussed the mind-body union. Swedenborg watched as the spirits divided into three camps representing the three hypotheses of the day: occasionalism, pre-established harmony and real interaction. When the discussion became acrimonious, a neutral spirit arrived to settle things. He invited a representative from each camp to write his favourite hypothesis on a slip of paper. The three slips of paper were put in a hat, and the spirits drew lots. Not surprisingly, Swedenborg's way of understanding things won the day. The lot chosen from the hat was 'spiritual influx', which Swedenborg interpreted as his own view that all life and wisdom flows from God into the human soul and from the human soul into the body.

Swedenborg was right at least about rational psychology. Champions of the different systems of mind-body union claimed to offer nothing more than hypotheses. The only way to make progress in the debate was to discredit the rival systems, and one might have concluded that too many opposing metaphysicians had done too good a job of

that. Swedenborg was frustrated by the apparent futility of the debate. He assumed that metaphysicians themselves would lose patience, if they had not done so already, and succumb to temptation - namely to simplify things by denying the existence of anything immaterial (including the rational soul) and to think that material nature operates under its own steam. They would come to think that God and nature are one. Swedenborg apparently believed that Spinozism results when human reason applies itself to metaphysics (see Spinoza, B. de). In this, he was to anticipate Jacobi.

Swedenborg himself believed that everything in material nature flows out of the thoughts of angels and spirits and that the power of thought in angels and spirits flows ultimately from God. This is the basis of his 'correspondence' theory. Accordingly, every material thing 'corresponds' to the spiritual thing of which it is an effect. Thus cows correspond to angels thinking about the affections of nature-bound minds; sheep correspond to angels thinking about spiritually uplifted minds, and so on. The correspondence theory is in turn the basis of Swedenborg's biblical exegesis. Swedenborg believed that every word in Scripture expresses a correspondence.

Swedenborg had a wide influence. His ideas were of interest to Emerson, Jaspers and Kant (§2). Kant became convinced in the mid-1760s that Swedenborg's visions were somehow emblematic of fundamental errors in Kant's own metaphysics. In 1766 he published a devastating self-critique in the guise of a satirical review of Swedenborg's *Arcana coelestia*. The title of Kant's review was *Dreams of a Spirit-Seer*. The writing of this review was to move Kant by 1770 to treat our representations of space and time as subjective forms of sensibility - an important step in the direction of his considered views in the *Critique of Pure Reason* (1781).

See also: Dualism; Mysticism, history of; Soul, nature and immortality of the

ALISON LAYWINE

List of works

Swedenborg, E. (1756) *Arcana coelestia*, trans. J.F. Potts, *Heavenly Secrets*, New York: The Swedenborg Foundation, 1978, 12 vols.(Commentary on Genesis and Exodus.)

Swedenborg, E. (1758) *De coelo et ejus mirabilibus et de inferno*, trans. G.F. Dole, *Heaven and Hell*, New York: The Swedenborg Foundation, 1979.(Swedenborg's visions of heaven and hell.)

Swedenborg, E. (1768a) *De Amore conjugiali*, trans. W.F. Wunsch, *Marital Love*, New York: The Swedenborg Foundation, 1975.(On the nature of love.)

Swedenborg, E. (1768b) *Apocalypsis revalata*, trans. J. Whitehead, *Apocalypse Revealed*, New York: The Swedenborg Foundation, 1975, 2 vols.(The meaning of Apocalypse revealed.)

Swedenborg, E. (1771a) *Vera Christiana religio*, trans. J.C. Ager, *True Christian Religion*, New York: The Swedenborg Foundation, 1980, 2 vols.(The theology of the new Church.)

References and further reading

Emerson, R.W. (1875) *Representative Men*, Boston, MA: Houghton Mifflin.(Includes an interesting and sympathetic study of Swedenborg and his claim that spiritual things lie behind every material thing.)

Kant, I. (1766) *Träume eines Geistersehers, erläutert durch Träume der Metaphysik (Dreams of a Spirit-Seer, explained through dreams of Metaphysics)*, Königsberg: Johann Jacob Kanter.(A scathing, satirical review of the *Arcana coelestia*, important for later developments in Kant's *Inaugural Dissertation* of 1770. See Volume 2 of the Academy Edition of Kant's works.)

Laywine, A. (1993) *Kant's Early Metaphysics and the Origins of the Critical Philosophy*, Atascadero, CA: Ridgeview Press.(A study of the development of Kant's early metaphysics with an extended discussion of Swedenborg's significance for Kant.)

Toksvig, S. (1948) *Emanuel Swedenborg*, New Haven, CT: Yale University Press.(An informative and sympathetic intellectual and spiritual biography.)

Symbolic interactionism

Symbolic interactionism is in the main a US sociological and social psychological perspective that has focused on the reciprocal relationship between language, identity and society. Philosophically it has largely been associated with pragmatists such as James (1907), Mead (1934), Dewey (1922) and Pierce (1958), although in the European context it has affinities with hermeneutics and phenomenology. In addition, it has links with various 'dramaturgical' approaches to communication that emphasize the interactive processes underpinning the construction, negotiation, presentation and affirmation of the self. In brief, symbolic interactionism is premised on the supposition that human beings are 'active' and not 'reactive'.

Although it is not easy to spell out the central propositions of Symbolic Interactionism in a systematic way, nevertheless, most of its proponents are committed to an interactive view of self and society, that is, they take issue with those views that see the social world as a seamless unity that completely encapsulates and determines individual conduct.

1 The language matrix

Very simply, this assumption suggests that the mind, self and society cannot be explained as separate phenomena. Their existence depends upon the way in which persons talk, interact, and communicate. Language, spoken or otherwise, is the medium through which men and women come to know themselves as individuals and members of social groups. In other words, the distinction between subjectivity and objectivity is always a matter of convention and historical situation. A person can only bear witness to their inner states and experiences in a shared symbolic context.

2 Reflexivity: the self and language

A critical thing about language for symbolic interactionism is its reflexivity, that is, its self-referring properties. Language allows a person to see themselves from the point of view of others, but simultaneously, it opens up the possibility of internal dialogue and debate. Hence, thinking is not a solitary activity, but is of necessity social. 'I think' because others around me talk to me. Accordingly, even the most private aspect of a person's inner life is somehow framed and performed as a conversation between social actors. Thought processes, in this view, replicate the dynamics of everyday communication. It could be said, although very circumspectly, that for symbolic interactionists, there is no distinction to be made between private and public language.

For symbolic interactionists, like Blumer (1969), it is reflexivity that makes it possible for people to overcome the limits set by social and physiological forces. Language, in this sense, is both liberating and imprisoning. It is liberating because it enables a person to envisage and construct alternative futures and pasts, and it is imprisoning because it often seems to coerce our sense of self.

3 The self as social

Perhaps more than anything else it is the symbolic interactionist view of the self as a social process that distinguishes it from other approaches. Against those versions of self that see it in terms of some underlying agency, symbolic interactionism highlights the tentativeness of all claims about the self. Nothing can be asserted about the self that does not meet the test of the situation or context. Put differently, this means that every social encounter may involve a different kind of self activity. So instead of a unitary self, the typical self in symbolic interactionist literature is multiple and episodic. Logically this means that there are as many selves as there are situations - perhaps an absurd proposition.

However, the apparent situatedness of the self is premised on some kind of notion of biographical continuity. Who I claim to be now, depends to a large extent on my past history. Even though each new interpersonal encounter may evoke a new way of 'presenting' myself, I am none the less not detached from earlier encounters. I have not invented my gender, class and ethnic affiliations, nor have I acquired my ability to speak and write my language in a random haphazard way. What this means therefore, is that biography always enters into interaction episodes, but simultaneously, biography is modified and challenged in interpersonal situations. In short, the self is constituted in the dialogue between the contingencies of everyday life, and the consequences of past interactions with significant

others like parents, teachers, siblings and peers.

4 Motives, intentions and identities as social constructions

In positing a social origin and deployment of the self, the symbolic interactionist takes issue with theories of behaviour that assume that human action is governed by drive-like motives. Motivation can only be understood in terms of the accounts, intentions, excuses, rationalizations of persons who live in a definite cultural milieu. In other words, motives are the meanings attributed to subjective states by persons participating in social encounters. At the same time, although motives are social, from the point of view of the individual, they are ‘real’. The agent’s ‘definition of the situation’ is thus a crucial aspect of any ongoing interaction episode.

Motives, then, are meanings that we give to our actions, and in the final analysis, these meanings are only significant if they are validated socially. They can be understood solely within the rubric of what people do and say as members of a society. For example, it makes no sense to talk of ambition in a context in which competition or achievement are not part and parcel of everyday activity. On the other hand, in most mundane situations, motives are taken for granted. We impute motives as a normal feature of our relationship with others.

For symbolic interactionists, the question of motives is also a question of identity. Who I believe myself to be is connected with what I believe I desire or want. The identity that I value also may be the reason for my actions. Be that as it may, both motives and identities are fashioned in social interaction.

5 Society

The symbolic interactionist view of society does not give it some overarching explanatory power. Society is what people do when they interact with each other in a multiplicity of situations. What we call society consists of literally countless encounters between people who talk, agree, argue and negotiate meanings and identities. In general, all interaction is sustained by implicit understandings of ‘what is the case’. Each new encounter is gained by some kind of background knowledge about the manner in which interactions work. We know what to do when we greet a friend, because in the past we have found occasion to learn what is meant by ‘friend’ and the requirements of a ‘greeting’.

Basically what this entails is seeing society as a communicative network in which participants are engaged in a ceaseless interpretive exercise to maintain some sense of order and identity. In this respect, society is never at rest - it is always in the process of being produced in interaction. At any one time the members of groups, collectivities, or institutions create and produce the conditions of their membership. Even a most rigidly structured bureaucratic institution, such as the army, depends on the reciprocal interaction of soldiers, officers and other ranks who interpret and monitor each others’ behaviour.

This does not mean that social order is simply a matter of *ad hoc* negotiation. What is negotiated is constrained by what has been negotiated in the past. While symbolic interactionists believe that individuals ‘make’ history, they also argue that what is ‘made’ has consequences for future interaction. Thus we have a picture of society as a set of flowing and volatile self-other processes intermeshing with an already interpreted history that profoundly influences the alternatives perceived as being possible in the present (see Society, concept of §§1, 3).

There are a number of objections that may be raised in conceiving of society in this way. First, there are situations in which negotiation appears to be irrelevant. How do concentration camp inmates negotiate with the functionaries who torture and incinerate them? There is not much scope here for the construction of new identities and meanings. Second, it may be that the symbolic interactionist picture of society is far too optimistic, and that it operates with an over-romanticized conception of human potentiality, a conception that gives the ‘underdog’, the ‘rebel’, and the ‘deviant’ almost superhuman status. Third, in its eagerness to demonstrate how selves and motives are socially constructed, it relies too heavily on the reflexive fecundity of language.

It could be argued that these objections are not only applicable to Symbolic Interactionism, but also to most social constructionist positions that give explanatory primacy to language. On the other hand, some of these criticisms may be misconceived because they have not taken into account the fact that Symbolic Interactionism is not a codified theoretical position, but is rather an orientation that questions the reifications of mainstream social science, as well as mocking and even subverting the myths and pretensions of policy makers and cultural

ideologues.

See also: Mead, G.H.

ARTHUR BRITTAN

References and further reading

Blumer, H. (1969) *Symbolic Interaction*, Englewood Cliffs, NJ: Prentice Hall.(A key text.)

Burke, K. (1962) *A Grammar of Motives and a Rhetoric of Motives*, Cleveland, OH: World Publishing.(A seminal discussion of dramatism and its relationship to motivation.)

Denzen, N.K. (1992) *Symbolic Interactionism and Cultural Studies*, Oxford: Blackwell.(A contemporary attempt to integrate symbolic interactionism and postmodernism.)

Dewey, J. (1922) *Human Nature and Conduct*, New York: Holt.(An attempt to explain human behaviour as a function of communicative and social process.)

Goffman, E. (1969) *The Presentation of Self in Everyday Life*, Garden City, NY: Doubleday.(A first statement of Goffman's dramaturgical perspective.)

James, W. (1907) *Pragmatism*, New York: Longmans, Green and Co.(A clear statement of the central philosophical influence on symbolic interactionism.)

Mead, G.H. (1934) *Mind, Self and Society: From the Standpoint of a Social Behaviourist*, Chicago, IL: University of Chicago Press.(Perhaps the most important text in the symbolic interactionist canon.)

Pierce, C.S. (1958) *Selected Writings*, Chicago, IL: University of Chicago Press. (Important discussions about language and the self and the development of US semiotics.)

Perinbanayagam, R.S. (1985) *Signifying Acts: Structure and Meaning in Everyday Life*, Carbondale and Edwardsville, IL: Southern Illinois University Press.(An important contribution to the construction of symbolic interactionist theory.)

Rock, P. (1979) *The Making of Symbolic Interactionism*, London: Macmillan.(A clear historical analysis.)

Syntax

Syntax (more loosely, 'grammar') is the study of the properties of expressions that distinguish them as members of different linguistic categories, and 'well-formedness', that is, the ways in which expressions belonging to these categories may be combined to form larger units. Typical syntactic categories include noun, verb and sentence. Syntactic properties have played an important role not only in the study of 'natural' languages (such as English or Urdu) but also in the study of logic and computation. For example, in symbolic logic, classes of well-formed formulas are specified without mentioning what formulas (or their parts) mean, or whether they are true or false; similarly, the operations of a computer can be fruitfully specified using only syntactic properties, a fact that has a bearing on the viability of computational theories of mind.

The study of the syntax of natural language has taken on significance for philosophy in the twentieth century, partly because of the suspicion, voiced by Russell, Wittgenstein and the logical positivists, that philosophical problems often turned on misunderstandings of syntax (or the closely related notion of 'logical form'). Moreover, an idea that has been fruitfully developed since the pioneering work of Frege is that a proper understanding of syntax offers an important basis for any understanding of semantics, since the meaning of a complex expression is compositional, that is, built up from the meanings of its parts as determined by syntax.

In the mid-twentieth century, philosophical interest in the systematic study of the syntax of natural language was heightened by Noam Chomsky's work on the nature of syntactic rules and on the innateness of mental structures specific to the acquisition (or growth) of grammatical knowledge. This work formalized traditional work on grammatical categories within an approach to the theory of computability, and also revived proposals of traditional philosophical rationalists that many twentieth-century empiricists had regarded as bankrupt. Chomskian theories of grammar have become the focus of most contemporary work on syntax.

1 The need for structure

One central aim of *semantics* is to explain how the meaning of a sentence (or any other complex expression) is a function of the meanings of its parts (see Compositionality), a project that presupposes an understanding of how the parts are put together, that is, an understanding of its *syntax*.

Let us call any sequence of words a 'string'. The following strings can be put together from the list of words 'Bill', 'slept', 'soundly', 'last' and 'night':

(1) (a) Bill slept soundly last night
(b) last night Bill slept soundly
(c) *last soundly Bill slept night
(d) *night Bill soundly last slept.

There is a clear difference between, on the one hand, (1a) and (1b) and, on the other, (1c) and (1d). In ordinary talk, the former are *sentences*; the latter are not. (Asterisks indicate non-sentences.) In order for a string of words to be a sentence (in this ordinary sense) the words must be put together in a certain way. Syntax is, in part, the study of the rules (or conditions) that determine the way that sentences are structured. The order of words in a string does more than determine whether or not a string is a sentence; it also plays a role in determining meaning. Compare (2a) and (2b):

(2) (a) The dog chased the cat.
(b) The cat chased the dog.

The fact that (2a) and (2b) differ in meaning cannot be a consequence of their containing words with different meanings, since they contain precisely the same words. The difference in meaning is attributable to a difference in *word order*. In the terminology of traditional grammar, in English we understand a noun phrase preceding a verb (in the active voice) as the subject, and the noun phrase following that verb as its direct object. Not all languages mark the subject-object distinction in this way. Thus, for example Latin marks it by inflection and this allows for the possibility of much freer word order.

Initially, the following generalizations might seem reasonable: (a) sentence meaning is the product of word meaning and word order; (b) a difference in word order results in a difference in sentence meaning. But in fact neither is accurate. Although a change in word order frequently results in a change in meaning, sentences (1a) and (1b) above show that this is not always the case, so (b) is incorrect. To see that (a) is incorrect consider

(3) Mary said Bill left voluntarily.

This admits of two distinct interpretations (or readings), which can be paraphrased as (a) 'Mary said that Bill left and she said this voluntarily' and (b) 'Mary said that Bill left and that he did so voluntarily'. It is as if the word 'voluntarily' can be understood in connection with either the saying or the leaving. Sentence meaning, then, is the product of more than word meaning and word order; it is the product of word meaning and sentence structure, and there is more to the sentence structure than the linear order in which its words appear: the way in which the words are grouped is crucial. One way of cashing out this idea might be to say that 'voluntarily' can be understood as an attachment either to the sentence 'Mary said Bill left' or to the smaller sentence 'Bill left'. Using square brackets to group parts of the sentence, we can represent the two different interpretations of (3) as follows:

(4) (a) [Mary said [Bill left]] voluntarily.
(b) Mary said [[Bill left] voluntarily].

In a sense that we can make precise, the moral of an example such as (3) is that the *hierarchical* organization (grouping) of the words in a sentence is just as important as their *linear* organization. (In terms explained in §5 below, the *scope* of 'voluntarily' is different in (4a) and (4b).)

2 Syntactic categories and constituent structure

The use of syntactic theory in projects such as compositional semantics requires that the theory have some systematic way of describing the syntax of sentences. The ways in which grammatical theories can describe linguistic structure are varied, but the following exposition will illustrate the most common. Since we are not here concerned with the internal structure of words (their morphology), let us say that words are the 'ultimate constituents' of sentences. The two parts of the sentence 'Odysseus returned' correspond to the traditional distinction between subject and predicate. Of course, both subject and predicate expressions can be more complex. If we replace the subject expression 'Odysseus' in 'Odysseus returned' by 'the hero', 'no hero' or 'a hero as great as godlike Achilles', in each case the result is another sentence. Since all of these expressions are built upon nouns, they are said to belong to the category of noun phrase (NP).

If we replace the verb 'returned' in the sentence 'Odysseus returned' by 'loves Penelope', or 'shaves another customer', in each case the result is another sentence. Since expressions in predicate position are (typically) based on verbs, they are said to belong to the linguistic category of verb phrase (VP). Using 'S' for 'sentence', we can represent the syntactic structure of the sentence 'Odysseus returned' using a 'phrase structure tree' (or 'phrase marker') as follows:

(5)

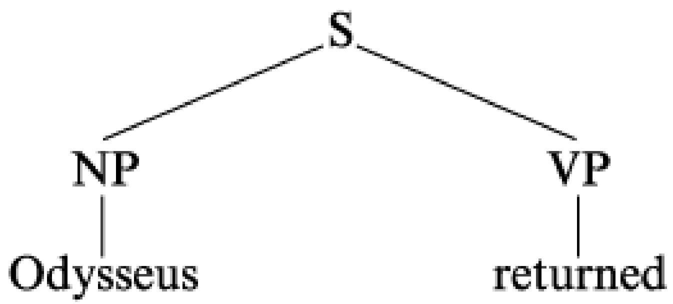

Alternatively, we can represent exactly the same syntactic information using a 'labelled bracketing':

(6) $[_{\mathrm{S}}[_{\mathrm{NP}}\,\mathrm{Odysseus}][_{\mathrm{VP}}\,\mathrm{returned}]]$.

(5) and (6) are notational variants of one another, they are equivalent descriptions of the same structure - 'structural descriptions' - that is, tree notation and bracket notation are two informationally equivalent ways of specifying syntactic structure.

We can begin to characterize a set of sentences by formulating a rule to the effect that an NP followed by a VP forms a sentence, S. The standard way of doing this is to use a 'phrase structure rule' (or 'rewrite rule'):

(7) $\mathrm{S} \Rightarrow \mathrm{NP} + \mathrm{VP}.$

This is read as 'S goes to NP VP'. All (7) says is that a sentence, or S, may be composed of an NP followed by a VP.

Let us turn now to the internal structures of NPs and VPs. Many NPs appear to be composed of words belonging to the traditional grammatical categories of article (for example, 'the', 'a') and noun (for example, 'cat', 'man'). Rather than using the category of article, let us use the broader category of *determiner* that includes not just 'the' and 'a', but also, for example, 'every', 'some', 'no', 'neither' and 'one'. We can represent the fact that an NP may be composed of a determiner (D) and a noun (N) using the following phrase structure rule:

(8) $\mathrm{NP} \Rightarrow \mathrm{D} + \mathrm{N}.$

Now we can provide a phrase structure tree for 'The hero returned' using the rules in (7) and (8):

(9)

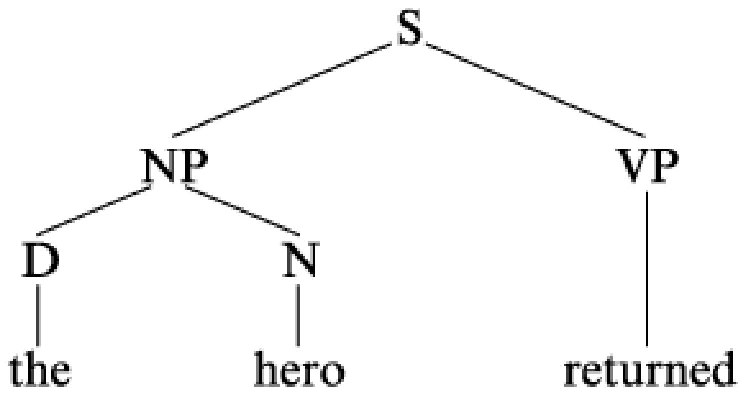

Or we can use a labelled bracketing:

(10) $[_{\mathrm{S}}[_{\mathrm{NP}}[_{\mathrm{D}}\mathrm{the}][_{\mathrm{N}}\mathrm{hero}]][_{\mathrm{VP}}\mathrm{returned}]].$

Some VPs may contain NPs as parts, for example, 'loves Penelope', and 'shaves another customer' (see above). Unlike verbs such as 'return' and 'snore', verbs like 'see' and 'respect' are *transitive*: they take NPs as direct objects. We have, then, at least two types of VPs to examine: those 'headed by' (that is, built around) intransitive verbs (V_is), such as 'return' and 'snore', and those headed by transitive verbs (V_ts), such as 'respect' and 'like'. Consequently, we need at least two phrase structure rules for VPs:

(11) $\mathrm{VP} \Rightarrow \mathrm{V_i}$
(12) $\mathrm{VP} \Rightarrow \mathrm{V_t} + \mathrm{NP}.$

The syntax of 'The hero returned' can now be spelled out in more detail:

(13)

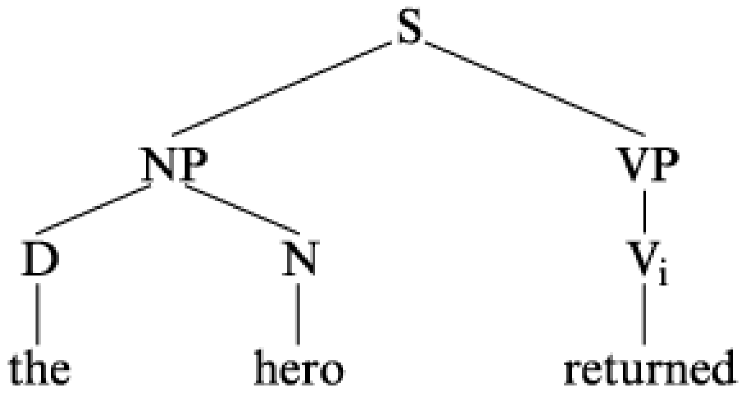

The syntax of 'The hero loved the goddess' is given by

(14)

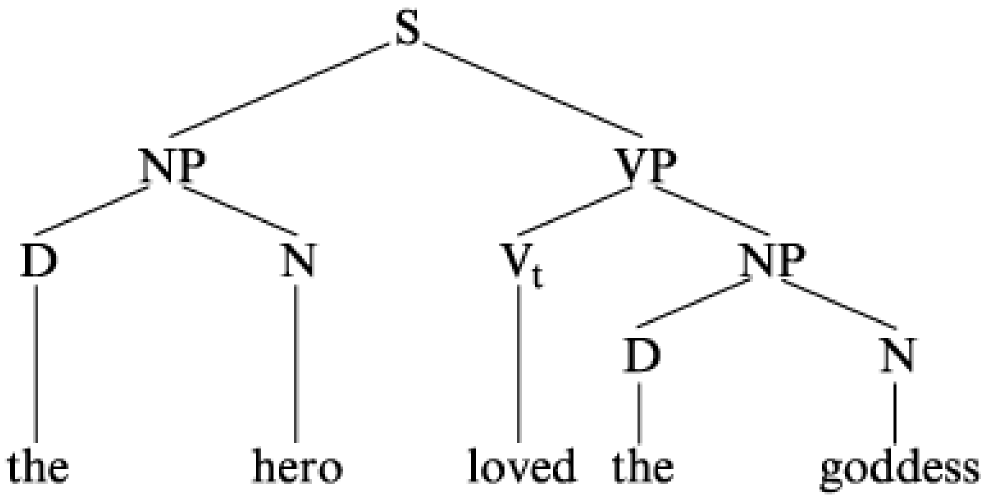

3 Phrase structure grammars

Not only do philosophical applications of syntactic theory require a set vocabulary for describing linguistic structure, they require linguistic theories to specify well-formed linguistic structures in a finite way. Contemporary syntactic theory offers just such resources.

One aim of what Chomsky has called ‘generative grammar’ is to articulate a finitely statable theory that generates all and only the phrase markers of a given language. With the resources made available in the previous section, we can construct a generative grammar for a fragment of English. For purposes of illustration, we can do this by articulating a ‘phrase structure grammar’. Such a grammar can be viewed as a formal system consisting of two parts: a ‘lexicon’ (a list of words together with a specification of the grammatical category of each word), and a set of phrase structure rules specifying how words from these categories may be put together to form sentences. In effect, then, a grammar is a theory, and like any other theory its usefulness lies in its predictive power. For any string of English words, the theory must say whether or not that string is a sentence. The following is a simple generative grammar:

Lexicon

$$\begin{aligned} \mathrm{PN} &= \{\mathrm{Fred, Mary, Bill}\} \\ \mathrm{D} &= \{\mathrm{a, every, the}\} \\ \mathrm{N} &= \{\mathrm{man, woman}\} \\ \mathrm{V_i} &= \{\mathrm{left, returned}\} \\ \mathrm{V_t} &= \{\mathrm{likes, respects}\} \end{aligned}$$

Phrase structure rules

$$\begin{aligned} \mathrm{S} &\Rightarrow \mathrm{NP + VP} \\ \mathrm{NP} &\Rightarrow \mathrm{PN}\ (\text{‘proper name’}) \\ \mathrm{NP} &\Rightarrow \mathrm{D + N} \\ \mathrm{VP} &\Rightarrow \mathrm{V_i} \\ \mathrm{VP} &\Rightarrow \mathrm{V_t + NP} \end{aligned}$$

Notice that each of the phrase structure rules has the following form:

(15) $\alpha \Rightarrow \beta_1 + \ldots + \beta_n.$

The important feature here is that in each rule of this type exactly one symbol α appears on the left-hand side of ‘$\Rightarrow$’. Such rules are called ‘context-free’ phrase structure rules, the idea behind the terminology being that they are blind to whatever symbols may appear on either side of α, that is, they are blind to the syntactic ‘context’ in which α appears. The systematic use of a set of context-free rules is a way of formalizing traditional work on

grammatical categories within an elegant approach to the theory of computability (or recursive functions); a context-free phrase structure grammar is equivalent to a categorial grammar in the sense of Ajdukiewicz (1935; see Ajdukiewicz, K. §5).

In a sense to be defined, the grammar just given generates the following phrase structure tree:

(16)

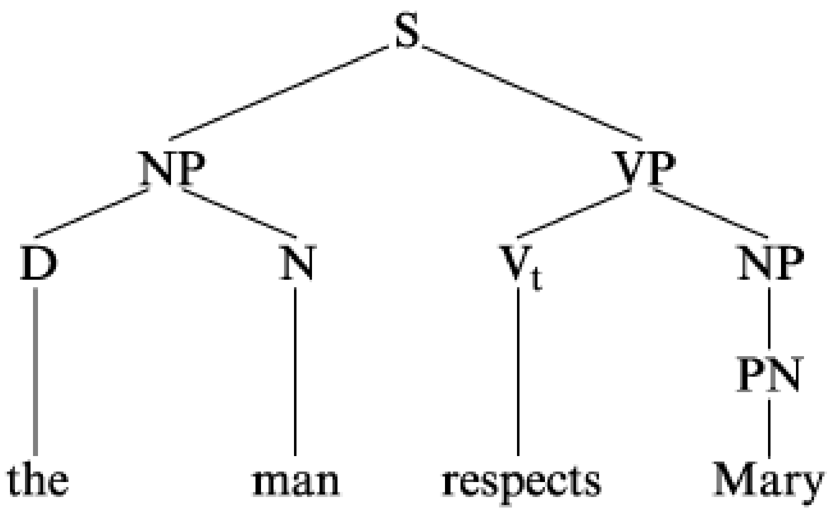

There is a specialized vocabulary in formal language theory (mathematical linguistics) for describing the parts of phrase structure trees and the relationships that obtain between the parts. We can tailor some of this vocabulary, rather informally, to suit our concerns:

(i) Each position in a tree is a *node*. A line connecting two nodes is a *branch*.

So, for example, in (1) there are twelve distinct nodes: S, NP, VP, D, N, V_t, PN, 'the', 'man', 'respects' and 'Mary'. Branches connect D and 'the', NP and D, and NP and N.

(ii) The S node at the 'top' of a tree is the *root node* of that tree. The words at the 'bottom' of a tree are its *leaf nodes* (or *terminal nodes*).

We are concerned with (a fragment of) English, so every leaf node will be a word of English.

(iii) A sequence of branches that connects two nodes is a *path*. Node β is *properly dominated* by node α if and only if the path from β back to the root node passes through α. The set of nodes properly dominated by node α is the *proper domain* of α.

Example: in (16) there is a path between D and 'the', and a path between S and D. Indeed, there is a path between any two distinct nodes. Additionally, D properly dominates 'the', NP properly dominates D, N, 'the' and 'man', and S properly dominates every node in the tree (apart from itself). (It dominates but does not properly dominate itself, just as a set is a subset but not a proper subset of itself.)

(iv) Each node in a tree corresponds to a *constituent* of the sentence in question.

So, for example, in (16) the NP node corresponds to the constituent 'the man'; and the VP node corresponds to the constituent 'respects Mary'. So whereas 'the', 'the man', 'respects' and 'respects Mary' are constituents of (16), 'man respects', 'man respects Mary' and 'the man respects' are not.

(v) If Γ is a phrase structure grammar and τ is a tree, then Γ *generates* τ if and only if
(a) the root node of τ is S;
(b) every terminal node of τ is in the lexicon for Γ; and
(c) every step from the root node to the string of terminal nodes is licensed by one of the phrase structure rules of Γ.
(vi) A *string* of Γ is a linear sequence of words taken from the lexicon for Γ. Γ *generates a string* Σ if and only if Γ generates a tree for Σ.

(vii) The set of strings generated by a grammar Γ is the *language* L_Γ of Γ, that is, the set of *sentences* of L_Γ.

An interesting feature of (vii) is that it appropriates the English words ‘language’ and ‘sentence’ and assigns them technical meanings. Later, the theoretical use of these words will be specified precisely to satisfy theoretical demands that arise when syntax and semantics come together.

The picture just presented is deliberately simplified. Nevertheless, it is widely held that context-free grammars of this form are inadequate to describe natural languages. In addition to context-free rules of the type we have just discussed, Chomsky (1957, 1965) posited ‘transformational’ rules that operated on the trees generated by context-free phrase structure grammars to produce trees that could not be produced in any satisfactory way by phrase structure rules alone. For example, Chomsky wanted to explain the syntactic and semantic relationships between the active and passive voice and accomplished this by postulating a transformational rule that derived a phrase structure tree for a passive sentence from the tree for its corresponding active. Much work by generative linguists in the 1970s sought to impose a rigorous set of constraints on possible transformational rules. Beginning with Chomsky (1981), generative linguists began to explore alternative ways of generating sentences using methods that depart from those of traditional phrase structure grammars.

4 Recursion and sentential verbs

The grammar above was finite, as was the language that it generated (it consists of a finite number of sentences). There are, however, finite grammars that generate infinite languages. This can be seen easily enough in connection with philosophically interesting verbs such as ‘believe’, ‘know’, ‘prove’, ‘remember’ and ‘say’, which can be used in ways that preclude classifying them as either V_is or V_ts. Consider the following sentences:

(17) Mary believes (that) Fred left.
(18) Mary said (that) Fred left voluntarily.

The VPs in these sentences seem to consist of a verb combined with a whole sentence. Since they seem to take whole sentences as their complements, these can be called ‘sentential verbs’ (V_s). Let us now add some V_ss to our lexicon:

V_s = {believe, know, doubt, say, suggest, realize, remember, prove}.

(Some V_ss are also known as ‘*psychological* verbs’ or ‘verbs of propositional attitude’, a label due to Russell - see Propositional attitude statements; §5 below.) We can now add the following phrase structure rule to produce a new grammar:

(19) $\mathrm{VP} \Rightarrow \mathrm{V_s} + \mathrm{S}.$

The resulting grammar generates the following phrase marker:

(20)

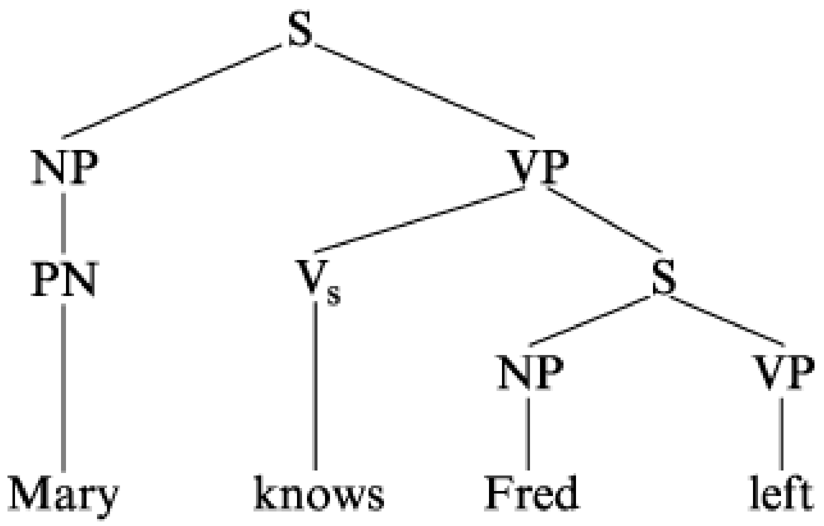

As this tree diagram reveals, the rules ‘ $\mathrm{S} \Rightarrow \mathrm{NP} + \mathrm{VP}$’ and ‘ $\mathrm{VP} \Rightarrow \mathrm{V_s} + \mathrm{S}$’ taken together give our grammar the capacity to generate trees in which S nodes are properly dominated by other S nodes. More precisely, our grammar generates trees in which an S node properly dominates a VP node that properly dominates another S node that properly dominates another VP node that…. So, for example, the grammar generates trees for both of the following:

(21) Bill knows Mary thinks Fred left.
(22) Mary doubts Fred knows Bill said Mary left.

Clearly, there is no end to the list of sentences generated: if our grammar generates a sentence ϕ, it also generates every sentence of the form

$[_{\mathrm{S}}[_{\mathrm{NP}}\ldots][_{\mathrm{VP}}[_{\mathrm{V\,s}}\ldots]\phi]]$.

Despite containing a finite set of rules, it generates an infinite set of sentences (the language generated is infinite). A structure such as (20) in which a node of category X properly dominates a distinct node also of category X is a *recursive* structure. We now say that our grammar *recursively generates* an infinite set of sentences. The recursive nature of natural language is further revealed once connectives such as ‘and’, ‘or’ and ‘but’ are introduced. A primary function of connectives is to join two sentences - they may also be used to connect, for example, NPs and VPs - in such a way that the result is a larger sentence, as in

(23) Bill left and Mary returned.
(24) Bill snores or Mary snores.

The following phrase structure rule can therefore be postulated:

(25) $\mathrm{S} \Rightarrow \mathrm{S\ CONN\ S}$,

where CONN is the category ‘connective’. This rule exploits the possibility of having three symbols on the right-hand side of ‘ $\Rightarrow$’, but it is still a context-free phrase structure rule of the form given in (15). Recursive structures are also found in (for example) NPs containing relative clauses such as ‘the man who loves the woman who found the child who…’.

5 Scope

There are a number of semantic issues of independent philosophical importance that can be illuminated by appealing to syntactic structure. The nature of scope in natural language is an instructive case study (see Scope). The following strings are all ambiguous:

(26) (a) everyone strives for some good
(b) if I know that *p* then necessarily *p*
(c) the number of planets is necessarily greater than seven
(d) Bill returned or Fred returned and Mary left
(e) George thought the King wasn’t the King
(f) Mary said Bill left voluntarily.

For example, someone uttering (26a) might be understood as saying that there is some particular good such that everyone strives for it or merely that everyone strives for some good or other. All of these examples involve ambiguities of scope that can be explicated within sophisticated syntactic theories. For present concerns, we restrict ourselves to a simple case mentioned earlier involving the scope of the adverb ‘voluntarily’ as it occurs in a string such as (26f), which is ambiguous between readings corresponding to the syntactic structures given earlier by (4a) and (4b). To provide a general characterization of scope in terms of phrase structure is straightforward once we introduce some new notions. First,

(viii) Node β is *immediately dominated* by node α if and only if β is properly dominated by α and there is no intervening node γ, that is, no node γ such that α properly dominates γ and γ properly dominates β.

In the following diagram, A properly dominates every other node; in addition it immediately dominates B and C,

but not D, E, F or G:

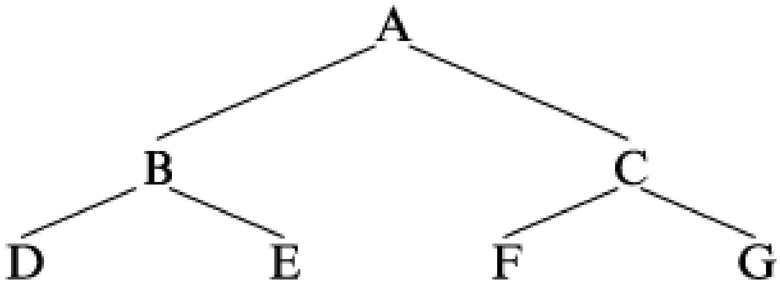

(ix) If distinct nodes β and γ are immediately dominated by node α, then α is a *branching node*.

A, B and C are branching nodes; D, E, F and G are not. We can now define 'scope' in a phrase structure tree:

(x) The *scope* of node α is (the constituent corresponding to) the first branching node properly dominating α. Node β is *within the scope* of node α if and only if the first branching node properly dominating α properly dominates β.

So in (20), the scope of 'Mary' is the entire sentence, whereas the scope of 'Fred' is the smaller sentence 'Fred left'.

(xi) If node β is within the scope of node α and node α is not within the scope of node β, then α has larger scope than β (and β has smaller scope than α).

Thus 'Mary' has larger scope than 'Fred' in (20). The scope of 'knows' in this phrase marker is the VP 'knows Fred left'. Thus the name 'Fred' occurs within the scope of 'knows', but the name 'Mary' does not. This is widely held to have important logical consequences. If Mary is Elizabeth - that is, if 'Mary' and 'Elizabeth' are two names of the same person - then the truth of (21) guarantees the truth of

(27) Elizabeth thinks Fred left.

Now suppose that Fred is Bert. Does the truth of (21) guarantee the truth of (28)?

(28) Mary thinks Bert left.

Following Frege, many philosophers believe it does not. (Mary may not know that Fred is Bert.) If this is correct, then we can describe the situation syntactically: the fact that 'Fred' appears within the scope of a 'psychological' verb (in this case, 'think') means there is no guarantee that 'Fred' can be replaced by a coreferring name to produce a sentence with the same truth-value as the original. A name occupying such a position is said to occur in a context that is 'referentially opaque' (see Propositional attitude statements §§1-2).

6 Ambiguities of scope

We turn now to an example involving what is usually called a 'structural ambiguity' or an 'ambiguity of scope'. The latter label (due to Russell) is more dominant in philosophy, logic, and mathematics. As Russell pointed out, much bad philosophy has resulted from inattention to scope ambiguities. Indeed, uncovering and avoiding them are skills all philosophers must acquire. Davidson has emphasized the importance of an understanding of the semantics of adverbs to a number of philosophical questions involving actions and events. Earlier, the fact that the following string is ambiguous was used to call attention to the grouping of words within a sentence:

(3) Mary said Bill left voluntarily.

It was suggested that the ambiguity in question was structural in that the words can legitimately be grouped in two different ways:

(4) (a) [Mary said [Bill left]] voluntarily.
(b) Mary said [[Bill left] voluntarily].

We can implement this suggestion by providing distinct phrase markers corresponding to (4a) and (4b). We will consider two rival ways of doing this. The word ‘voluntarily’ belongs to the category of adverb, so let us add to our lexicon the following:

(29) ADV = {voluntarily, hastily, discreetly}.

In line with the informal account of the ambiguity suggested earlier, the first attempt to characterize the syntax of adverbs involves positing the following phrase structure rule:

(30) $\mathrm{S} \Rightarrow \mathrm{S} + \mathrm{ADV}.$

The resulting grammar generates the following distinct phrase markers for (3):

(31a)

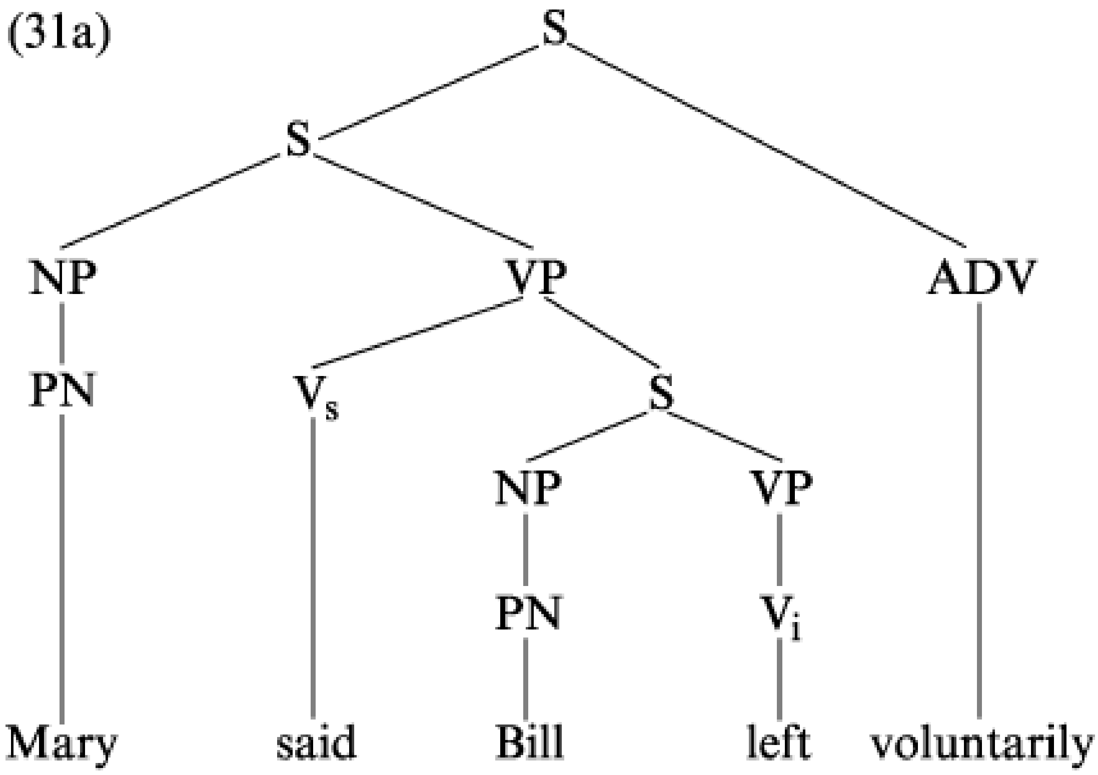

(31b)

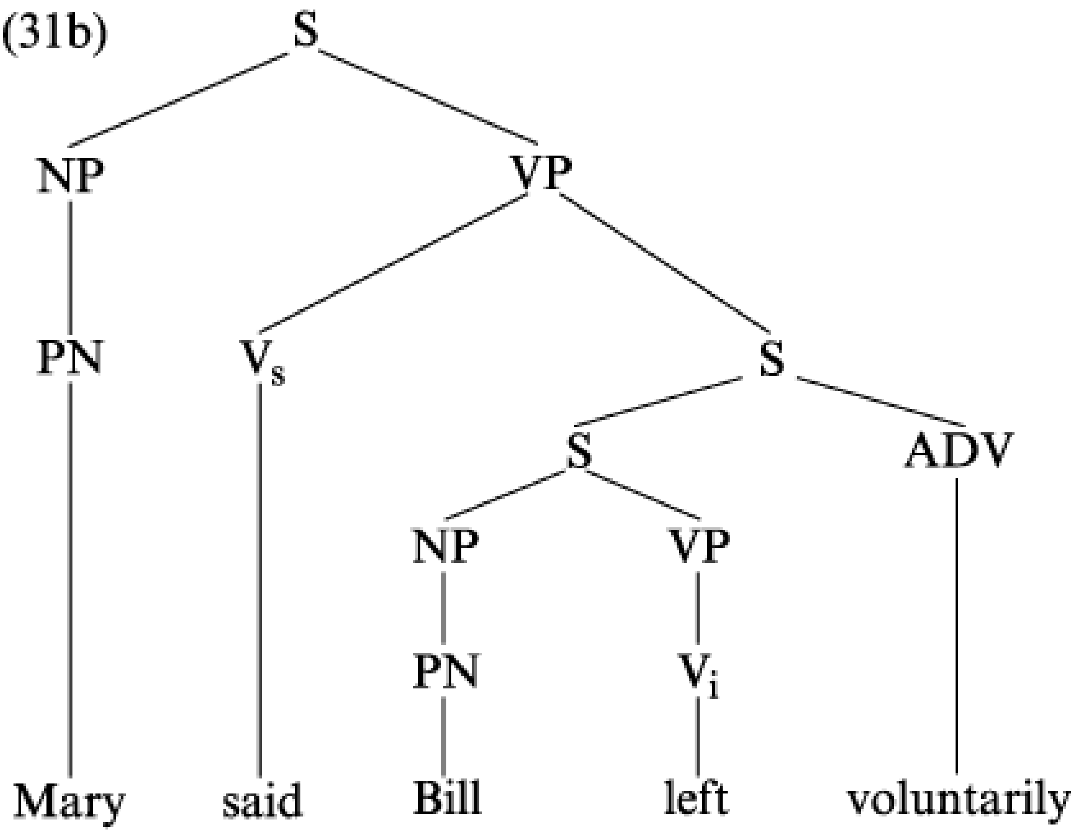

In (31a) the scope - as defined above - of ‘voluntarily’ is the entire sentence. Thus the verb ‘said’ is within its scope. By contrast, in (31b) the scope of ‘voluntarily’ is just the smaller sentence ‘Bill left voluntarily’. Thus the verb ‘said’ is not within its scope, but the verb ‘left’ is. On the assumption that, as the etymology suggests, an adverb modifies a verb in some way, the verb it modifies is the main verb of the sentence to which it is attached. Thus (31a) corresponds to the reading of (3) that might be loosely paraphrased as ‘Of her own volition, Mary said that Bill left’ and (31b) corresponds to the reading that might be paraphrased as ‘Mary said that Bill left as a matter of his own volition’. (Of course, until an account is provided of how syntactic structure contributes to meaning, the pairing of distinct phrase markers with distinct readings of a string is largely intuitive.)

We now make two observations about our definition of scope. First, by the definition provided, every constituent has a scope; but there are only certain types of constituents whose scopes will interest us (usually, connectives, quantifiers, verbs and adverbs). Second, notice that it is not always the case that for two constituents one is within the scope of the other.

It is worth asking whether this is the best way of characterizing the syntax of sentences containing adverbs. An alternative account emerges if we take the etymology of ‘adverb’ more seriously and view ADVs as attaching to VP nodes rather than S nodes. Only detailed empirical and logical work will tell which of these analyses is better. The best theory may tell us that there are two distinct types of ADV: one sort that attaches to an S node, another that attaches to VP. There is no need for us to settle this difficult matter here. Suffice to say, that if we find that claims about the semantic properties of adverbs are doing some philosophical work, we will have to look carefully at their syntax.

More than a few philosophical muddles have been due to confusions over the scope of operators in natural language sentences. Utilizing the resources of syntactic theory, it becomes possible for us to study scope relations in a more rigorous fashion, and to marshal arguments for and against certain proposals about the logical form of natural language. Indeed, whenever we encounter philosophical claims which rely upon the semantic properties of quantifiers, adverbs and so on, we will want to look carefully at their syntactic properties. This specific point is just an instance of a quite general principle concerning the relationship between the philosophy of language and theoretical syntax: the former cannot properly proceed independently of the latter.

See also: Analytical philosophy; Chomsky, N.; Language of thought; Logical form; Mind, computational theories

of; Scope

STEPHEN NEALE

References and further reading

Ajdukiewicz, K. (1935) 'Die Syntaktische Konnexität', *Studia Philosophica* 1: 1-27; trans. 'Syntactic Connexion', in S. McCall (ed.) *Polish Logic*, Oxford: Clarendon Press, 1967.(A full statement of the theory of categorial grammar.)

Chomsky, N. (1957) *Syntactic Structures*, The Hague: Mouton.(Early statement of the formal limitations of various approaches to syntax, mathematical properties of different types of grammars, and the theory of transformational grammar. Very readable.)

Chomsky, N. (1965) *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press.(Detailed statement of Chomskian methodology, and Chomsky's 'Standard Theory'.)

Chomsky, N. (1981) *Lectures on Government and Binding*, Dordrecht: Foris.(Detailed statement of major components of Chomsky's 'principles and parameters' approach to syntax that moves well beyond the transformational approach of his earlier work. Hard going.)

Chomsky, N. (1986) *Knowledge of Language*, New York: Praeger.(Very readable account of the philosophical and technical details of the 'principles and parameters' approach.)

Davidson, D. (1980) 'The Logical Form of Action Sentences', in *Essays on Actions and Events*, Oxford: Clarendon Press.(A proposal about the logical form of sentences involving action verbs.)

Lewis, D.K. (1983) 'General Semantics', in *Philosophical Papers*, vol. 1, Oxford: Oxford University Press.(Accessible exposition of a categorial grammar for a part of English and a corresponding compositional semantics.)

May, R. (1985) *Logical Form: Its Structure and Derivation*, Cambridge, MA: MIT Press.(First sustained attempt to connect work in Chomsky (1981) with traditional philosophical work on 'logical form'.)

Neale, S. (1994) 'Logical Form and LF', in C. Otero (ed.) *Noam Chomsky: Critical Assessments*, London: Routledge & Kegan Paul.(An attempt to connect Chomsky's syntactic programme to work on meaning, reference and logical form.)

Russell, B.A.W. (1905) 'On Denoting', *Mind* 14: 479-93; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 41-56.(The classic theory of definite descriptions that provided a paradigm for analysis of logical form.)

Sells, P. (1985) *Lectures on Contemporary Syntactic Theories*, Stanford, CA: Center for the Study of Language and Information, Stanford University.(Accessible overview and exposition of competing approaches to generative grammar.)

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, trans. D. Pears and B. McGuinness, London: Routledge & Kegan Paul, 1961.(An early, philosophically ambitious theory of logical form.)

Systems theory in social science

'Systems theory' is a label for two very different approaches to social analysis. The first was a post-1945 successor to traditional organicist theories of society that for some twenty years dominated US sociology and political science. It was never very popular outside the USA, and is now of largely historical interest; social scientists who aspire to develop a positive science of social interaction have for the past two decades rested their hopes on the individualist analyses provided by rational choice theory.

Systems theory in the first sense of the term flourished in US sociology and political science from the late 1940s to the late 1960s, and is especially associated with the names of Talcott Parsons in sociology and David Easton and Gabriel Almond in political science. A systems approach to social analysis was commonly, though not universally, associated with some form of functionalism, especially in the work of Parsons, the leading structural-functionalist of his day. It fell into disrepute along with functionalism, a victim of the changed political climate of the 1960s as much as of its purely intellectual weaknesses.

The second form of systems theory is associated especially with the name of Niklas Luhmann, and its leading critic is Jürgen Habermas. The second form is in vigorous life, but not well known in the USA.

1 The organicist background

Systems theory in this sense inherited the assumptions of traditional organicist theories of society. The belief that social analysis must be grounded in a distinctively holistic and organic approach to its subject matter was characteristic of conservative critics of eighteenth-century rationalism such as Edmund Burke; its philosophical defence was one of those few of Hegel's achievements that Marx unhesitatingly applauded (see Hegel, G.W.F.). In part, this was a way of staking out a terrain for sociology distinct from the terrain of economic analysis. The conservative implications of these ideas as understood by writers such as Burke and Hegel were not intrinsic to an emphasis on the organic character of society. Few thinkers were more conscious of the depth of social and economic conflict than Karl Marx, and his belief that social and economic relations formed a system of exploitation was one of the things that persuaded him that revolution rather than reformist tinkering was demanded.

At the end of the nineteenth century, liberals, too, rested their hopes on a new, more organic conception of society than they thought their predecessors - particularly the British utilitarians - had attained. In France, Émile Durkheim (§2) defended an extreme holism as to sociological method. So widely felt was the wish to ground ethics in sociology and to rescue liberal politics from exaggerated individualism that one might think it part of the *fin de siècle* spirit of the age (see Holism and individualism in history and social science).

This conception of the social organism fell into (something like) disrepute after the First World War; the obvious dissimilarities between a society and organisms such as fish, sheep or human beings had, of course, never been denied by nineteenth-century social thinkers, but they had perhaps not appeared to be worth mentioning in the course of arguments intended to emphasize that, whatever else they might be, societies were not established by freely contracting, independently formed, self-sustaining and autonomous individuals. Turn of the century liberal organicists had no such conservative purposes as, say, Burke had had, but the association of organicism with authoritarianism and irrationalism, however, was subsequently sufficient to damn it in the eyes of most interwar liberal thinkers.

2 The modern revival

Not every work that stresses the idea of system rests on such a form of systems theory. Kenneth Waltz's *The Theory of International Politics* (1979), is perhaps the most famous work in the field in the past two decades; it strongly emphasizes the opportunities and constraints that membership of an international system of states offers to and imposes on each state in the international system. It is, for all that, concerned with the international system as a balance of opposed forces, not with its quasi-organic qualities as something self-maintaining, self-reinforcing, and self-repairing. It is systems theory in this latter sense that was briefly popular, and so harshly criticized.

Several representative samples will stand for many. Talcott Parsons (1950) provided a framework for social

analysis that saw the social system as a system of interaction, or, as he put it an 'action-system'. The system as a whole was made up of four sub-systems: in effect, the economic, cultural, political and psychological aspects of a social order. It is important to emphasize that these 'sub-systems' were defined in terms of function, not concretely in terms of institutions. These functions included such things as socializing individuals into an acceptance of the local norms of sexual identity and moral conviction, or organizing their efforts at producing and distributing the means of life. Understanding how this was achieved was equally a matter of asking how social norms affected the orientation of individuals towards interactions with others: for instance, whether we treat each other instrumentally or in terms of some non-instrumental moral norm.

Much the same controversy beset systems theorists in political science, although none of them produced the elaborate categorical system that Parsons had produced; nor did they arouse the same intensity of discipleship and opposition. David Easton (1953) represented the political system as an input-output device as it were embedded in the wider social system; from the social system there flowed demands on the one hand and support on the other, and from the political system there flowed 'authoritatively allocated' values. Critics tended to complain not only that this was only a way of redescribing what everyone knew already, but that it misrepresented the nature of politics. Only in advanced modern societies was there a formal distinction between the realm of the political and that of the economic or the cultural, so this was not a picture of all forms of politics. Again, in some societies the political system is active rather than reactive, trying to mould the society rather than waiting to receive 'inputs'.

Almond's functional analyses of the political system were not entirely similar. In his account (*The Politics of the Developing Areas*, with James S. Coleman; *Comparative Politics Today*, with G. Bingham Powell) the political system must perform three output functions of 'rule-making, rule-enforcing and rule-adjudication', categories that bear perhaps too clear a resemblance to the three powers of the US legislature, executive and judiciary. Similarly, the 'input functions' of 'interest articulation' and 'interest aggregation' were sometimes derided as ornate synonyms for the work done by pressure groups and political parties in the USA. Enthusiasts for the approach responded by applying it to the then Soviet Union (David S. Lane, *Politics and Society in the USSR*) and employing it as a basis for the general theories of revolution (Chalmers Johnson, *Revolutionary Change*); unappeased critics continued to claim that such efforts showed that it was possible to restate the conventional wisdom in the new terminology, but not that there were new insights to be gained from its use.

What turned out to be intellectually unsatisfying and politically implausible were the unintended but hard to escape implications of the systems approach; that societies were 'naturally' orderly, that order was properly maintained by norms that integrated individuals into the social system, the selfishness required of economic actors would readily be held in check by altruistic and family-oriented norms, and more views of the same sort.

Systems theory was always criticized for its political deficiencies as much as its intellectual shortcomings. Among the complaints levelled at it, the best known were that it endorsed the US *status quo* by emphasizing the forces that made for social 'integration' and overlooking social conflict, that its adherents defined social and political development as a society's similarity to the USA, and that they removed the critical bite from democratic theory by elevating 'democratic stability' to the highest place among political values, at the expense of justice and equality. In general, it was widely argued the conflicts of interest and disparities of power were either ignored or made light of in the perspective so concerned to associate normality with stability.

3 Niklas Luhmann's systems theory

Just as the influence of Parsons' structural-functionalism began to wane in the USA, the German sociologist Niklas Luhmann emerged as the leading proponent of a new version of systems theory. Luhmann thought that the previous attempts to use systems theory in the social sciences applied cybernetic concepts too directly and suffered from the residual normative orientations of Durkheim and Parsons, which he, like structuralists and post-structuralists, denounced as so much 'old European humanism'. To be rigorous and consistent, systems theory had to drop all reference to actors and their self-interpretations, which were nothing but 'psychical systems' that form part of the environment for other systems. In this way, systems theory could replace the functionalist account of social integration through norms with the anonymous integration of interdependent parts and wholes and be generally applicable to every level of social analysis (Luhmann 1976).

Such a general and radical systems theory recognizes the fundamental fact of the large scale and complex forms of

social organization in modern societies, such as global markets and enormous bureaucracies, that work ‘behind the backs’ of social actors. Rather than being the product of intentional actions, such order is described in terms of functionally differentiated, yet interdependent, parts, whereby the internal operations of each maintain the system as a whole. Luhmann’s theoretical innovation was to emphasize how such social systems arise. Much like living organisms, social systems emerge through ‘autopoiesis’, that is, they are generated and maintained ‘exclusively’ by their own internal operations, and yet remain open to their environments. For Luhmann, systems are like Leibnizian monads, equipped with windows; ‘meaning’ is but one internal operation whose function is to reduce complexity, while ‘norms’ serve to restrict the possible operations of the system. Modern society can thus best be explained as a series of systems that have other systems as their environments, including law, the economy, organizations, science (including systems theory itself). Their internal complexity requires that they are externally related and do not form an overarching system.

According to Luhmann, systems may be formally defined in terms of their complexity. Although systems are always less complex than their environments, they must also constantly heighten their own complexity in order to adapt and remain stable. Once the inner-outer boundary is maintained and stabilized, a system becomes ‘operationally closed’, that is, it registers events in its environment only to the extent that they can be translated into its code and programming. This conception of ‘operational closure’ marks the most important difference between the theories of Luhmann and Parsons. Luhmann denies that modern societies are integrated in Parsons’ functionalist sense: there is no central or organizing system, whether state or society, but only the interdependencies between systems. In *Social Systems* (1984), Luhmann calls modern, highly complex and functionally differentiated societies ‘polycentric’, with no integrating centre or apex of power. As opposed to Parsons’ optimistic liberalism (1971), Luhmann sees democratic participation and control as an illusion in the face of overwhelming complexity and necessarily latent functional imperatives.

Luhmann does not limit his analysis to large scale social phenomena, but rather applies his systems theory to all levels of analysis, even to intimate relationships such as love. It is the lack of any restriction on the scope of such explanations that opens the theory to criticism, most thoroughly by Jürgen Habermas in his debate with Luhmann (1971). Luhmann’s analysis, according to Habermas, cannot recognize the coordinating effects of ordinary language communication within modern institutions. Such communication requires that actors are allowed back into social theory (Habermas 1970).

Habermas criticizes Luhmann’s systems theory on several levels. Theoretically, Habermas criticizes Luhmann for ignoring the continuing role of institutions, and thus of social integration, in ‘anchoring’ systemic mechanisms in the cultural life-worlds of their members. Because of the interaction of social and systemic integration, systemic mechanisms can have perverse effects, disrupting and ‘colonizing’ the domain of cultural reproduction. Methodologically, Habermas argues that action descriptions from the agents’ point of view remain a necessary condition for any social explanation. Normatively, Luhmann’s ‘methodological anti-humanism’ blinds the theory from the start to the possible influence of a society-wide and critical public sphere on complex, institutional processes.

See also: Functionalism in social science; Structuralism in social science

ALAN RYAN
JAMES BOHMAN

References and further reading

Alexander, J., Giesen, B., Münch, R. and Smelser, N. (1987) *The Macro-Micro Link*, Berkeley, CA: University of California Press.(Useful collection of essays on anti-reductionism, or linking individualist and collectivist explanation.)

Almond, G.A. and Coleman, J. (1960) *The Politics of Developing Areas*, Princeton, NJ: Princeton University Press.(Explanation of developing areas as input-output systems.)

Almond, G.A. and Powell, G.B. (1988) *Comparative Politics Today: A World View*, London: HarperCollins, 4th edn.(A study of comparative politics.)

Bohman, J. (1991) *New Philosophy of Social Science: Problems of Indeterminacy*, Oxford and Cambridge, MA: Polity/MIT Press.(Includes introduction to issues concerning collective explanations in the social sciences, including systems theory, with examples.)

Easton, D. (1953) *The Political System: A Systems Theory of Political Life*, New York: Alfred A. Knopf.(An early attempt by a political scientist to get some mileage of the idea of a ‘system’.)

Habermas, J. (1970) *The Logic of the Social Sciences*, trans. S. Nicholsen, Cambridge, MA: MIT Press, 1988. (Most complete statement of Habermas’ view of social science, including criticisms of Parsons, systems theory and functionalism.)

Habermas, J. and Luhmann, N. (1971) *Theorie der Gesellschaft oder Sozialtechnologie: Was leistet die Sozialforschung? (Theory of Society or Social Technology: What is the Purpose of Social Research?)*, Frankfurt: Suhrkamp Verlag. (Extensive debate between Habermas and Luhmann on social theory.)

Johnson, C. (1982) *Revolutionary Change*, Cambridge: Cambridge University Press, 2nd edn.(General theories of revolution.)

Lane, D.S. (1971) *Politics and Society in the USSR*, New York: Random House.(Uses systems theory as the basis for a general account of revolutions.)

Luhmann, N. (1976) *The Differentiation of Society*, trans. S. Holmes and C. Larmore, New York: Columbia University Press, 1982.(Best available collection of Luhmann’s essays, covering a variety of topics from law to science.)

(1984) *Soziale Systeme (Social Systems)*, Frankfurt: Suhrkamp Verlag.(Luhmann’s most systematic theoretical work, written in an intricate technical vocabulary.)

McCarthy, T. (1978) *The Critical Theory of Jürgen Habermas*, Cambridge, MA: MIT Press.(Includes an accessible summary of the Habermas/Luhmann debate and extensive bibliography.)

Parsons, T. (1950) *The Social System*, London: Routledge, 1991.(The classical statement of Parsonian systems theory; new edition with introduction by Brian Turner.)

Parsons, T. (1966) *Societies: Evolutionary and Comparative Perspectives*, Englewood Cliffs, NJ: Prentice Hall.(A short introductory treatment of its topic, markedly different from the above (Parsons 1950).)

Parsons, T. (1968) ‘Social Systems and Subsystems’, in *Encyclopedia of the Social Sciences*, New York: Macmillan and Free Press.(Less ‘Parsonian’ than many of its author’s accounts of the topic.)

Parsons, T. (1971) *The System of Modern Societies*, Englewood Cliffs, NJ: Prentice Hall.(Best introduction to Parsons’ use of systems theory.)

Waltz, K. (1979) *The Theory of International Politics*, New York: McGraw-Hill.(An analysis of international relations as a staff-maintaining system.)

Tagore, Rabindranath (1886-1941)

In the flurry of intellectual activity in nineteenth- and twentieth-century Bengal, Rabindranath Tagore became one of the best-known playwrights, poets, novelists, educators and philosophers, winning the Nobel Prize for literature in 1913. His thought drew on the English Romantics as well as Sanskrit and Bengali writers and movements.

Tagore was not a systematic philosopher. He termed his position 'a poet's religion' which valued imagination above reason. He moved between the personal warmth of human relationships to a theistic Divine and belief in an Absolute as a unifying principle. He advocated a thoughtful but active life, criticizing asceticism and ritualism.

Nineteenth- and twentieth-century Bengal witnessed a renewed flurry of intellectual activity as a response to the new Western presence in India. As a member of the 'Bengal renaissance', Rabindranath Tagore was both a prolific writer and an active educator. He left school aged fourteen, finding traditional education 'stifling', and in 1901 he founded a 'forest retreat' academy called Santiniketan which emphasized the arts and nature. In 1918 he founded Visvabharati, an international university promoting international understanding as well as Indian culture.

Tagore was brought up in a cosmopolitan home dominated by his father, Debendranath, who led a monotheistic and iconoclastic reform movement known as the Brahmo Samaj, founded in 1828 (see Brahmo Samaj). Debendranath was a religious and social reformer open to the ideas of the West. At home Rabindranath studied English, Sanskrit and Bengali literature and came to love the English Romantics Keats and Shelley, the classical Sanskrit writer Kālidāsa and the poetry of love for the Divine of the Bengali devotees of Kṛṣṇa, particularly the songs of the wandering singers known as the Bauls.

Although Tagore's writings reveal consistent themes and concerns, Tagore was not a systematic, philosophical thinker. He is best known for his plays, poetry, songs, essays and novels. He wrote his first poem at the age of eight. A collection of his devotional poems entitled *Gitanjali (Song Offerings)*, which he translated into English himself, brought him the Nobel Prize for literature in 1913. In four major religious-philosophical works, *Sādhanā: The Realization of Life* (1913), *Personality* (1917b), *Creative Unity* (1922) and *The Religion of Man* (1931), Tagore set forth his thought more fully. However, these are collections of essays and speeches: they are not solidly argued systematic treatises.

Tagore called his position 'a poet's religion', affirming that reality is best known through experience and emotion and that imagination is more important than reason (Sense perception, Indian views of). Reality is understood through recognition of the Absolute One behind the multiplicity, which is a 'creative principle of unity'. This means not only that the Divine is immanent in creation but that creation itself is a manifestation of the Divine. The world and its particulars are real because they are the Divine in playful manifestation. The underlying harmony and the inner interrelationship of each of the parts to the others is experienced as the expression of Divine joy and love. This affirmation of unity in Tagore's thought enabled contemporary Indian monists to read Tagore as a modern follower of Śaṇkara.

Tagore, however, rarely used the term 'Absolute' as he did not merely affirm an impersonal consciousness as the highest reality. Thus, early Western interpreters spoke of a Christian influence on Tagore. Besides the personal sense of the Divine he inherited through his father's Brahmo Samaj, his discovery at the age of fifteen of the love poems of the Bengali devotional movements, which sang of Kṛṣṇa and his loving relationship with his followers, was enough to redirect his understanding (see Gauḍīya Vaiṣṇavism). These poems provided inspiration both for Tagore's emphasis on the personal as the key to the universe and his philosophy of Divine joy and love. Ultimate reality, he believed, could not be divorced from personality; the Absolute is the supreme personality in intimate, personal and interdependent relationship with human beings. The Divine is not aloof, but involved with humanity. Their relationship of mutual love involves God's loving service for human wellbeing and the human response of realizing the infinite in the finite and becoming increasingly more united to the Divine.

The fullest expression of the Divine is in the human being. However, this does not mean that the individual does not exist as a separate entity, or that union with the Divine means absorption of the individual so as to lose its distinctive existence. Like the traditional philosophical position called *bhedābheda*, identity in difference, Tagore saw the human being as both identical with and different from the Divine. From what he describes as his first

religious experience at the age of eighteen, a vision lasting four days, he saw the personal as the central fact of the Divine-human-world relationship.

For Tagore, the poet, not the abstract philosopher, is the true seer. The poet reveals the Divine through song, paradox and aesthetic awareness. All art is the response of the soul to the Divine for it reveals beauty which is truth. Tagore defines beauty as the revelation of the harmony of the universe; conflict and chaos are the result of a limited vision which does not penetrate into the harmonious relationship of the whole but settles for usefulness and efficiency. Reason is limited to space and time, but the creative imagination found in all artistic endeavour connects humanity to the all and results in a response of active participation. Tagore, therefore, persistently criticizes the ascetic who flees from life as an escapist separated from the world. He also chides the religious devotee who is lost to the world in emotionalism. The traditional ascetic ideal, Tagore believed, would not further India's progress nor that of humanity. The active individual who transcends the egoistic desire for gain in the love of the Divine and its creation lives the good life. The poet stirs one to involvement in the world. The path to the realization of the Divine, *sādhanā*, includes all creative activity, which for Tagore was his writing, painting, composing and educating. 'Expression', he once wrote to a friend, 'is my religion'.

See also: Hindu philosophy; Monism, Indian; Vedānta

ROBERT N. MINOR

List of works

Tagore, R. (1912) *Gitanjali (Song Offerings)*, London: The India Society.(Collection of devotional poetry Tagore himself translated from his original Bengali poems.)

Tagore, R. (1913) *Sādhanā: The Realization of Life*, London: Macmillan. (Nontechnical, collected papers which discuss major themes in Tagore's thought.)

Tagore, R. (1917a) *Nationalism*, London: Macmillan.(Lectures given in the USA and Japan on Tagore's political philosophy.)

Tagore, R. (1917b) *Personality*, London: Macmillan.(Collection of lectures delivered in the USA.)

Tagore, R. (1922) *Creative Unity*, London: Macmillan.(Philosophical essays addressing the concepts of the human being, religion and art.)

Tagore, R. (1931) *The Religion of Man*, London: Allen & Unwin.(The Hibbert lectures delivered at Manchester University which survey the full range of Tagore's thought.)

References and further reading

Atkinson, D. (1989) *Gandhi and Tagore: Visionaries of Modern India*, Hong Kong: Asian Research Services.(An insightful introduction to Tagore's thought.)

Naravane, V. (1977) *An Introduction to Rabindranath Tagore*, Delhi: Macmillan.(A standard introduction to Tagore from the perspective of a monist philosopher.)

O'Connell, J. (1989) *Tagore's Heritage*, Toronto: Rabindranath Tagore Foundation. (A useful collection of essays analysing Tagore's thought and influence.)

Taine, Hippolyte-Adolphe (1828-93)

Hippolyte Taine dominated the intellectual life of France in the second half of the nineteenth century. He was seen as the leader of the positivist, empiricist, anti-clerical forces in a period characterized by dramatic advances in science and technology and inspired by the hope that scientific method could be applied to human affairs. Yet at the heart of his life and work was the rationalist, essentialist imperative of Spinoza and of Hegel: to demonstrate the world as system, as necessity, to 'banish contingency'. The story of his life is the story of the abandonment of this project: it is a long, painful learning experience ending in the acceptance of loss; his richly varied works can be seen as the products of this philosophical journey.

Taine (born in Vouziers in the Ardennes) was the Sartre of his time. As withSartre, philosophy (more specifically, metaphysics) was the mainspring of all his work; but (again like Sartre) this work was variegated and popular and quite different from that of the philosophers - Spinoza, Hegel, J.S. Mill - who were his models. Most of Taine's writing was first published in newspapers and periodicals, such as *Le Journal des Débats* or *La Revue des Deux-Mondes*, before being revised, collected and published in book form.

He produced literary criticism (the various collections of essays and *La Fontaine et ses fables (La Fontaine and his fables)* (1861)); literary history (one of the first histories of English literature in any language); books of travel and social observation (the *Voyages* (Travels) and *Notes sur Paris* (1867)); and a monumental political and intellectual history of France, *Les Origines de la France contemporaine (The Origins of Contemporary France)* (1875-93). He also wrote an iconoclastic account of the *spiritualiste* philosophers of the preceding generation - Victor Cousin, Maine de Biran,Théodore Jouffroy - and published an influential work of psychological philosophy, *De l'Intelligence (On Intelligence)* (1870). The *Vie et Correspondance (Life and Letters)* (1902-7) is one of the great collections of the century and contains, besides remarkable letters, a great deal of crucial unpublished early material.

The published works give an appearance of dispersion and fail to reveal how fundamental the quest for philosophical unity was for him. There are two main reasons for this. First, like most thinkers of his time he was deeply attached to scientific observation, empirical reality, inductive method, experience - to what he called 'facts'. Throughout, he starts from the concrete - specific works of art or literature, historical records, psychological phenomena - the range is great but the underlying motivation is always philosophical in the sense of an urge towards the general and the abstract. The key is always the preface or the conclusion. Whether he acknowledges it or not, his aim is always a philosophy of art or literature, a philosophy of history, a philosophy of mind and, ultimately, a metaphysics. His aspiration, like Hegel's, is enyclopedic, and the apparently disparate parts of his investigation are aspects of an envisaged totality. In modern terms, Taine is a critical theorist and not a literary critic.

The second reason involves the process of reception, the way in which, during their lifetime and subsequently, writers are constructed, by themselves and others so as to be non-contradictory. In his twenties, as a consequence of his loss of religious faith, Taine was deeply concerned with metaphysics. A remarkable unpublished text, written in 1848, called 'De la Destinée humaine' (On Human Destiny) describes his loss of religious faith, his systematic doubt, and his salvation throughSpinoza and pantheism. Soon after, he learned German to read Hegel (then untranslated). His voluminous writings of this period remained unpublished during his lifetime and he himself progressively attenuated the expression of his metaphysical ambitions. Thus, by the end of his life, he was seen as a positivist and a pluralist. Only the publication of the *Life and Letters* and subsequent studies of unpublished material and variants revealed the underlying idealism, the persistent urge to refute Kant, Hume andMill, and the need to 'banish contingency'. For some critics he then became a sort of closet idealist; for many he was just confused and 'flawed'. Only recently has his work been seen as a dramatic struggle between competing forces.

Taine is representative not only because he was the public spokesman for the positivist, deterministic, anti-clerical, anti-Romantic aspects of his society, but because he lived the tensions and contradictions of that period and left a precise record. His writings can be read as a succession of attempts to square the circle, to demonstrate the world as system, as necessary, not accidental, or to show, as he put it, 'that philosophy was possible'. 'De la Destinée

humaine' is full of youthful optimism. But the sceptic grows stronger: Spinoza, he complains, 'destroys the world' and ignores Time. Hegel, whom Taine in many ways resembles, seemed to offer salvation: he writes in the *Philosophes français*, 'I read Hegel every day for a whole year. I will probably never again experience sensations like those he gave me' (1857: 126; author's translation). But Taine (who had not read the *Phenomenology of Spirit*) concludes by finding Hegel 'abstract', positing order a priori, but not proving it. He too 'destroys the world'.

Thereafter Taine's life is a series of attempts to succeed where Hegel failed; each time the goal seems less attainable; each effort is followed by a period of illness and depression. This rhythm continues even after he has renounced philosophy and become a historian: every re-edition of *On Intelligence* has a modified ending. By the final (4th) edition of 1883, Hegel's name has disappeared and Taine seems resigned to the loss of the Absolute. Sartre saw that the book was not an 'empiricism' as people had thought, but was in fact 'a realist metaphysics *manqué*'.

Taine does not have a 'philosophy'. His achievement is to have struggled throughout his life to reconcile an inherited rationalist desire for system with a scrupulous respect for concrete experience, to reconcile Spinoza and Hegel with Kant, Hume and Mill. The attempt ended in failure. The *Derniers essais* (1894) are suffused with a sense of wasted potential. But the story of this painful and honest coming to terms with loss over a lifetime's philosophical effort is dramatic and moving. It is also modern - and more characteristic of the lives of philosophers than is usually acknowledged.

See also: Hegelianism §4

COLIN EVANS

List of works

Taine H.-A. (1857) *Les Philosophes français du XIXe siècle (French Philosophers of the Nineteenth Century)*, Paris: Hachette.(Iconoclastic account of previous generation of French philosophers.)

Taine H.-A. (1858) *Essais de critique et d'histoire (Critical and Historical Essays)*, Paris: Hachette.(The preface of this work is especially interesting.)

Taine H.-A. (1861) *La Fontaine et ses fables (La Fontaine and His Fables)*, Paris: Hachette. (Taine's doctoral thesis, written when a projected work on Hegel was declared unacceptable to the University.)

Taine H.-A. (1864) *Histoire de la littérature anglaise*, Paris: Hachette; trans H. Van Laun, *History of English Literature*, New York, Holt, 1872.(The introduction and the chapter on J.S. Mill are especially interesting.)

Taine H.-A. (1865a) *Philosophie de l'art*, Paris: Hachette; trans. J. Durand, *Lectures on Art*, New York, Holt, 1896.(A version of his lectures at the École des Beaux Arts, where he was Professor.)

Taine H.-A. (1865b) *Nouveaux essais de critique et d'histoire (New Critical and Historical Essays)*, Paris: Hachette.(Contains essays on Balzac, Marcus Aurelius and Racine.)

Taine H.-A. (1866) *Voyage en Italie*, Paris: Hachette; trans. J. Durand, *Italy*, 1870, 1871, New York, Holt.(Travel descriptions and personal responses to works of art.)

Taine H.-A. (1867) *Notes sur Paris, vie et opinions de M. Frédéric-Thomas Graindorge*, Paris: Hachette; trans. J.A. Stevens, *Notes on Paris*, New York, Holt, 1879.(In this work, Graindorge is a fictional American visitor to Paris.)

Taine H.-A. (1870) *De l'Intelligence*, Paris: Hachette; trans. T.D. Haye, *On Intelligence*, New York, Holt, 1899. (Taine revised each edition, changing especially the ending, to make the work less Hegelian.)

Taine H.-A. (1872) *Notes sur l'Angleterre*, Paris: Hachette; trans. W.F. Rae, *Notes on England*, New York, Holt. (Lively, detailed response to England and English society.)

Taine H.-A. (1875-93) *Les Origines de la France contemporaine*; trans. J. Durand, *The Origins of Contemporary France*, New York, Holt, 1885-94.(Taine's highly controversial history of France, designed as a 'medical consultation'.)

Taine H.-A. (1894) *Derniers essais de critique et d'histoire (Last Critical and Historical Essays)*, Paris: Hachette. (Contains moving, elegiac and confessional accounts of minor writers. The theme is '*Ubi sunt*'.)

Taine H.-A. (1902-7) *H. Taine, sa vie et sa correspondance*, Paris: Hachette, 4 vols; trans. R.L. Devonshire (vols 1 and 2) and E. Sparvel-Bayly (vol. 3. abridged), *Life and letters of H. Taine*, New York, Holt, 1902-8.(One of the great collections of the century. The editors made certain omissions, however, and the correspondence with Camille Selden was destroyed - see F. Léger below. Volume 1 contains extracts from 'De la Destinée humaine'.)

References and further reading

Charlton, D.G. (1959) *Positivist thought in France during the Second Empire*, Oxford: Oxford University Press. (Contains chapters on all the so-called positivist thinkers of the period, showing how there was always a tension with idealism.)

Evans, C. (1975) *Taine, essai de biographie intérieure*, Paris: Nizet, esp. 173-308. (Very long and detailed *doctorat d'état* thesis, adopting extreme diachronic approach; uses unpublished material; full bibliography.)

Evans, C. (1978) 'Taine and his fate', *Nineteenth-century French Studies* 6 (1-2): 118-28. (An account of Taine's critical reception.)

Léger, F. (1980) *La Jeunesse d'Hippolyte Taine*, Paris: Albatros.(A biography of Taine's life until the age of thirty.)

Léger, F. (1993) *Monsieur Taine*, Paris: Critérion. (A less academic account, but covers the whole life.)

Sartre, J.-P. (1936) *L'Imagination*, Paris: Presses Universitaires de France.(Offers a perceptive account ofTaine's psychology: see chapter 2, 'The Problem of the Image'.)

Weinstein, L. (1972) *Hippolyte Taine*, New York: Twayne.(The best full treatment of Taine's thought in English.)

Tanabe Hajime (1885-1962)

Tanabe Hajime was a central figure of the so-called Kyoto School, and is generally acknowledged to be one of the most important philosophers of modern Japan. He held Kant in high esteem, and used a Neo-Kantian critical methodology in his early studies in epistemology. In the 1920s he was chiefly influenced by Nishida Kitarō's original cosmological system. He adapted Nishida's idea of 'absolute nothingness' to political situations and, in so doing, contributed much to establishing the foundations of what became the most influential philosophical school in Japan up until the end of the Second World War.

In 1919, Tanabe was appointed associate professor under Nishida Kitarō at Kyoto University. In 1927 he was promoted to full professor, replacing Nishida soon after. In 1945 he retired to Kitakaruizawa, Gunma-ken, and dedicated his later life to meditation and writing.

Under the influence of Nishida, Tanabe wished to develop his own cosmological system. This desire grew stronger during his stay in Germany from 1922 to 1924, where he studied under Edmund Husserl at Freiburg University and became friends with Martin Heidegger. Although he did not accept much of their phenomenological method, he was stimulated by Heidegger's intention to remodel phenomenology into a cosmology by assimilating existentialism. Returning to Japan, he began to turn his interpretation of Kantian philosophy into a systematic cosmology. The fruit of these efforts was his work *Kanto no Mokutekiron (Kant's Teleology)*, published in 1924, in which he characterized Kantian reflective judgement as the faculty of a comprehensive teleological world view (see Kant, I.).

Tanabe's attention then turned to dialectics. He held Hegelian dialectics to be just the logical method that could give concrete shape to a teleological comprehension of the world (see Hegel, G.W.F.). According to him, the formula 'thesis-antithesis- synthesis' represents the process through which teleologically driven objects gradually realize themselves. The final aim of the process indicates the teleological perfection of the world. To this he applied the term 'absolute nothingness', chiefly on the grounds that this term describes the activity of dialectical negation. Thus 'absolute nothingness' took on an active meaning in his understanding, where it had been originally used by Nishida in a meditative, introspective way. Tanabe described his steps to a thorough acceptance of dialectics in his work *Hegerutetsugaku to Benshoho (Hegel's Philosophy and Dialectics)* in 1932.

In 1934 Tanabe introduced a new notion, 'the logic of species', the novelty of which earned him a reputation as an original thinker. As he understood it, the triad of genus-species-individual was simply a variant of the triad of propositional logic, universal- particular-singular, as it is applied to human existence. In his logic of species, therefore, Tanabe was concerned with giving a new priority to the particular in human experience.

In its methodological structure, the logic of species was an attempt to apply dialectics to the real world. The individual is determined decisively by the species; conversely, the former seeks of its own free will to determine the latter. The two face each other as opposites in tension. Each is irrational and selfish in that each immediately affirms its own being, but through mutual interpenetration they can be purified of their irrational qualities and raised to the level of rationality. The resultant synthesis is the genus. Thus the individual and the species are synthesized in the genus; that is, a particular society is 'generalized'. Tanabe lumped together these results in the notion of the 'state'. The state as the realization of the universal had an important position in his thought. It was only natural for him to take the next steps towards a theory of the state, identifying the dialectical process with the perpetual self-renewal of the polity. Among his articles during this period, great importance can be attached to the one entitled 'Kokkatekisonzai no Ronri' (The Logic of National Existence), which appeared in 1939.

Tanabe's theory of the state was in its basic structure an admittedly rationalistic one. Pivoting on a dynamic in which ethnic unity is 'generalized' into rationality through the activity of free individuals, it offered the ideal of the nation-state. It was Tanabe's intention to give the Japanese state a rational, democratic basis; in this way, he set himself to check the irrationalistic, warlike modes of thinking. Obviously his efforts failed, chiefly due to the confusion of the ideal state with the real Japanese state. Tanabe watched as his own words were twisted by the nationalists into a justification for military aggression and, with pangs of conscience at this turn of events, his thinking came to a standstill.

After December 1941, Tanabe was no longer able to publish any philosophical work. It was only at the end of the war, with *Zangedo toshiteno Tetsugaku (Philosophy as Metanoetics)*, that his struggles bore fruit and released his thinking from its predicament. This work is often characterized as Tanabe's 'Confessions', for he described in it his experience of personal salvation. In despair over his philosophy up to that time, he had abandoned all attempts at realization under his own power and repented his errors. At this point he came to the awareness of returning to life by the grace of a transcendent Other-power. Since the new form of his philosophy was based on this experience, he named it 'metanoetics'. Thereafter he concentrated on religious thinking.

Tanabe then tried to reorganize his logic of species in light of the experience of metanoetic salvation. The particular society is 'generalized' through human action deriving ultimately from Other-power. This time, the resulting society is called not the state, but the religious community in its universal character. From here, Tanabe developed his interest in sociological research into the history of religions. In his notable work *Kirisutokyo no Bensho (Dialectic of Christianity)*, published in 1948, he explicated the development of primitive Christianity as typical of the way in which a religious community is founded. In his last years, he theorized on the bodhisattva-way of Mahāyāna Buddhism from the perspective of sociological philosophy of religion.

See also: Japanese philosophy; Kyoto School; Logic in Japan

HIMI KIYOSHI

List of works

Tanabe Hajime (1910-61) *Tanabe Hajime zenshū(The Complete Works of Tanabe Hajime)*, ed. Y. Takeuchi *et al.*, Tokyo: Chikuma Shobo, 1963-4, 15 vols.(Edition of Tanabe's complete works.)

Tanabe Hajime (1924) *Kanto no Mokutekiron (Kant's Teleology)*, in *Tanabe Hajime zenshū(The Complete Works of Tanabe Hajime)*, Tokyo: Chikuma Shobo, 1963-4.(On Kant's reflective judgment.)

Tanabe Hajime (1932) *Hegerutetsugaku to Benshoho (Hegel's Philosophy and Dialectics)*, in *Tanabe Hajime zenshū(The Complete Works of Tanabe Hajime)*, Tokyo: Chikuma Shobo, 1963-4.(Tanabe's view of Hegelian dialectics.)

Tanabe Hajime (1939) 'Kokkatekisonzai no Ronri' (The Logic of National Existence), in *Tanabe Hajime zenshū(The Complete Works of Tanabe Hajime)*, Tokyo: Chikuma Shobo, 1963-4.(Identifies Tanabe's dialectical process with a theory of the state.)

Tanabe Hajime (1946) *Zangedo toshiteno Tetsugaku (Philosophy as Metanoetics)*, trans. Y. Takeuchi and J.W. Heisig, Los Angeles, CA: University of California Press, 1986.(Tanabe's 'Confessions', as the start of his late philosophy.)

Tanabe Hajime (1947) 'The Logic of the Species as Dialectics', trans. D.A. Dilworth, *Monumenta Nipponica* 1969, 24 (3): 273-88.(An attempt to reorganize the logic of species in the light of the metanoetic experience.)

Tanabe Hajime (1948) *Kirisutokyo no Bensho (Dialectic of Christianity)*, in *Tanabe Hajime zenshū(The Complete Works of Tanabe Hajime)*, Tokyo: Chikuma Shobo, 1963-4.(Tanabe's study of the development of Christianity.)

Tanabe Hajime (1959) 'Memento Mori', trans. V.H. Viglielmo, *Philosophical Studies of Japan* I, Compiled by the National Commission for UNESCO: 1-12.(Summary of Tanabe's meditation on death in his late days.)

References and further reading

Himi Kiyoshi (1990) *Tanabetetsugaku Kenkyu - Shukyotetsugaku no Kanten kara (Studies in the Thought of Tanabe: A Perspective from the Philosophy of Religion)*, Tokyo: Hokuju Shuppan.(Pursuit of the development of Tanabe's thought, laying great emphasis on his late religious philosophy.)

Laube, J. (1984) *Dialektik der absoluten Vermittlung (Dialectic of the Absolute Mediation)*, Freiburg: Verlag Herder.(Critical researches on Tanabe's dialectical philosophy.)

Unno Taitetsu and Heisig, J.W. (eds) (1990) *The Religious Philosophy of Tanabe Hajime: The Metanoetic Imperative*, Berkeley, CA: Asian Humanities Press.(Record of the International Symposium on Tanabe's Metanoetics held in Massachusetts, 1989.)

Tarski, Alfred (1901-83)

Alfred Tarski was a Polish mathematician and logician. He worked in metamathematics and semantics, set theory, algebra and the foundations of geometry. Some of his logical works, in particular his definition of truth, were also significant contributions to philosophy. He was a successful teacher and a master of writing simply and with great clarity about complicated matters.

Tarski was born Alfred Tajtelbaum in Warsaw. (The family name was changed in 1924.) Between 1918 and 1924 Tarski studied at the University of Warsaw where he received his doctorate in mathematics under the direction of S. Leśniewski. In 1926 he was appointed as a lecturer. In 1939 Tarski set out for a lecture tour of the USA and was prevented from returning to Poland by the outbreak of the Second World War. He then briefly held positions at Harvard University, the City College of New York and the Institute for Advanced Study at Princeton. In 1942 he was appointed to the mathematics department of the University of California at Berkeley, where he remained until his retirement in 1968.

Tarski's best known logical achievements, being also of philosophical significance, are certainly his definitions of satisfaction and truth and his work on logical consequence (see Tarski's definition of truth; Consequence, conceptions of). All Tarski's logical investigations were determined by his interests in the methodology of the deductive sciences, which he termed metamathematics. In fact, general metamathematics was one of his chief areas of activity. He distinguished three main types of deductive systems and studied their metamathematical properties - in particular, their axiomatizability, the independence of their axioms, and their consistency and completeness. Gradually, his metamathematical investigations took on an algebraic character, as reflected in his increasing emphasis on semantics. This approach led to the algebraic treatment of logic and the development of model theory, of which he was a founder and the author of various fundamental results.

Tarski's interest in model theory was connected to his studies on decidability. Here too his approach was fundamentally semantical and connected to matters of definability in theories. The key notion was that of interpretability, and Tarski developed general methods of proving (essential) undecidability of various theories (in particular, various algebraic and geometrical theories) by considering their interpretability in certain other theories. Another method used by Tarski in connection with the study of definability was that of quantifier elimination. It was also applied by Tarski to prove decidability of various theories (for example, of the theory of real closed fields).

Tarski used logical methods to study the foundations of various mathematical theories, in particular of geometry. Being interested mainly in Euclidean and general affine geometry he proposed various axiom systems for them and studied their metamathematical properties. Particularly prominent in Tarski's geometrical work was the problem of determining and selecting primitive notions. Connected with his work in geometry and set theory is the Banach-Tarski theorem, which is a remarkable consequence of the axiom of choice. It asserts that any sphere can be decomposed into a finite number of subsets which can be reassembled by rigid motions to form two spheres each having the same volume as the original. Tarski also proved that several statements of the arithmetic of the infinite cardinals are equivalent to the axiom of choice.

Tarski left few philosophical papers. Nevertheless, various parts of his logical work and, in particular, those concerned chiefly with the methodology of the deductive sciences and the nature of truth and consequence, are of considerable philosophical interest. He described himself as 'a mathematician (as well as a logician, perhaps a philosopher of a sort)'. It is difficult to characterize his philosophical views. He did, however, indicate that he believed logical and mathematical research ought not to be restricted by any general philosophical viewpoint.

ROMAN MURAWSKI

List of works

Tarski's bibliography comprises 109 papers, 94 abstracts and 20 monographs (some of them translated into several languages). Listed here are some of his works with the most importance for logic and philosophy.

Tarski, A. (1981) *The Collected Works of Alfred Tarski*, ed. S.R. Givant and R.N. McKenzie, Berkeley, CA: University of California Press, 4 vols.(Includes all Tarski's published works.)

Tarski, A. and Banach, S. (1924) ‘Sur la décomposition des ensembles de points en parties respectivement congruentes’ (On the Decomposition of Point-Sets into Mutually Congruent Parts), *Fundamenta Mathematicae* 6: 244-77.(Includes the Banach-Tarski theorem on the decomposition of the sphere.)

Tarski, A. (1930) ‘Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften I’, *Monatshefte für Mathematik und Physik* 37: 361-404; trans. J.H. Woodger (1956), ‘Fundamental Concepts of the Methodology of Deductive Sciences’, in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 60-109.(Introduces and investigates fundamental notions of the methodology of deductive sciences.)

Tarski, A. (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956), ‘The Concept of Truth in Formalized Languages’, in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(One of the most fundamental of Tarski’s works, including his famous definition of truth.)

Tarski, A. (1935) *O logice matematycznej i metodzie dedukcyjnej*, Lwów; trans. *Introduction to Logic and to the Methodology of Deductive Sciences*, Oxford and New York: Oxford University Press, 1941.(This book gives a general introduction, and shows how logic can be applied in the construction of mathematical theories.)

Tarski, A. (1944) ‘The Semantic Conception of Truth and the Foundations of Semantics’, *Philosophy and Phenomenological Research* 4: 341-75; repr. in L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University of Illinois Press, 1952.(Tarski’s definition of satisfaction and truth.)

Tarski, A. (1948) *A Decision Method for Elementary Algebra and Geometry*, prepared for publication by J.C.C. McKinsey, Santa Monica, CA: Rand.(One of the most important of Tarski’s results in decidability theory, namely the proof of decidability of the theory of reals.)

Tarski, A. (1949) *Cardinal Algebras*, Oxford and New York: Oxford University Press.(A monograph devoted to the algebra of cardinal numbers.)

Tarski, A., Mostowski, A. and Robinson, R.M. (1953) *Undecidable Theories*, Amsterdam: North Holland. (Includes three papers devoted to general methods of proving undecidability of theories, of various systems of arithmetic and of group theory.)

Tarski, A., Mostowski, A. and Robinson, R.M. (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press; repr. and ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(Includes English translations of Tarski’s main logical papers from 1923 to 1938.)

Tarski, A., Mostowski, A. and Robinson, R.M. (1967) *The Completeness of Elementary Algebra and Geometry*, Paris: Institut Blaise Pascal.(The proof of the decidability of the theory of reals and of geometry.)

Tarski, A., Henkin, L. and Monk, J.D. (1971, 1985) *Cylindric Algebras*, Amsterdam: North Holland, 2 vols.(A monograph devoted to cylindric algebras.)

Tarski, A., Doner, J. and Mostowski, A. (1978) ‘The Elementary Theory of Well-Ordering - A Metamathematical Study’, in A. MacIntyre *et al*. (eds) *Logic Colloquium 77*, Amsterdam: North Holland, 1-54. (The proof of the decidability of the elementary theory of well-ordering.)

Tarski, A., Schwabhäuser, W. and Szmielew, W. (1983) *Metamathematische Methoden in der Geometrie*, Berlin and New York: Springer.(The development and metamathematical investigation of absolute and Euclidean geometry on the basis of a system of axioms introduced by Tarski.)

Tarski, A. and Givant, S. (1987) *A Formalization of Set Theory Without Variables*, Providence, RI: American Mathematical Society.(A demonstration that set theory and number theory can be developed within the framework of a new, different and simple equational formalism, closely related to the formalism of the theory of relational algebras.)

References and further reading

Blok, W.J. and Pigozzi, D. (1988) ‘Alfred Tarski’s Work on General Metamathematics’, *Journal of Symbolic Logic* 53: 36-50.(A discussion of Tarski’s work on the methodology of the deductive sciences.)

Doner, J. and Hodges, W. (1988) ‘Alfred Tarski and Decidable Theories’, *Journal of Symbolic Logic* 53: 20-35. (Discusses Tarski’s work on the decidability of theories.)

Dries, L. van den (1988) ‘Alfred Tarski’s Elimination Theory for Real Closed Fields’, *Journal of Symbolic Logic* 53: 7-19.(Discusses Tarski’s theorem on quantifier elimination for the theory of reals, and its influence on mathematics and logic.)

Etchemendy, J. (1988) ‘Tarski on Truth and Logical Consequence’, *Journal of Symbolic Logic* 53: 51- 79. (Discusses Tarski’s writings on the concept of truth and logical consequence.)

Givant, S. (1986) ‘Bibliography of Alfred Tarski’, *Journal of Symbolic Logic* 51: 913-41.(With a complete

bibliography, including information on translations.)
Jønsson, B. (1986) 'The Contributions of Alfred Tarski to General Algebra', *Journal of Symbolic Logic* 51: 883-9. (A discussion of Tarski's writings on general algebra.)
Levy, A. (1988) 'Alfred Tarski's Work in Set Theory', *Journal of Symbolic Logic* 53: 2-6.(Considers Tarski's contribution to set theory.)
Monk, J.D. (1986) 'The Contribution of Alfred Tarski to Algebraic Logic', *Journal of Symbolic Logic* 51: 899-906.(Studies the meaning of Tarski's writings on algebraic logic.)
McNulty, G.F. (1986) 'Alfred Tarski and Undecidable Theories', *Journal of Symbolic Logic* 51: 890-8.(Discusses Tarski's methods of proving undecidability of theories, as well as their applications.)
Suppes, P. (1988) 'Philosophical Implications of Tarski's Work', *Journal of Symbolic Logic* 53: 80-91.(Discusses the implications for philosophy of Tarski's work.)
Szczerba, L.W. (1986) 'Tarski and Geometry', *Journal of Symbolic Logic* 51: 907-12.(Considers Tarski's work in the field of foundations of geometry.)
Vaught, R.L. (1986) 'Alfred Tarski's Work in Model Theory', *Journal of Symbolic Logic* 51: 869-82.(Studies Tarski's contribution to the development of model theory.)

Tarski's definition of truth

Alfred Tarski's definition of truth is unlike any that philosophers have given in their long struggle to understand the concept of truth. Tarski's definition is more clear and precise than any previous definition, but it is also unusual in character and more restricted in scope. Tarski does not provide a general definition of truth. He provides instead a method of constructing, for a range of formalized languages L, definitions of the notions 'true sentence of L'. A remarkable feature of Tarski's work on truth is his 'Criterion T', which lays down a general condition that any definition of 'true sentence of L' must satisfy. Tarski's ideas have exercised an enormous influence in philosophy. They have played an important role in the formulation and defence of a range of views in logic, semantics and metaphysics.

1 Formal correctness and material adequacy

Any definition that aims to elucidate some aspect of a concept must satisfy two sets of requirements: the definition must be, in Tarski's terminology, 'formally correct' and 'materially adequate'. Under formal correctness fall requirements concerning the form of the definition and the vocabulary that the definition may use. Concerning these requirements, let us note only that a definition of 'true' of the form

x is true iff *x*

is bound to be formally correct provided that the right-hand side of the equivalence contains only 'admissible' vocabulary, that is, vocabulary that may legitimately be used in defining 'true'. For example, the definition

x is true iff *x* corresponds to reality

fails to satisfy the requirement of formal correctness, because the terms 'corresponds' and 'reality' are much too unclear to be considered admissible. On the other hand, no doubts should arise about the formal correctness of the definition

x is true iff *x* is identical to *x*.

The definition is, of course, unsatisfactory. But that is because it is materially inadequate: it fails to conform to our ordinary notion of truth.

The most remarkable feature of Tarski's work on truth (1936, 1944) is his formulation, known as 'Criterion T', of the requirement of material adequacy. It is this Criterion that is responsible for the special characteristics, both positive and negative, of Tarski's definition. On the positive side, it enables Tarski to give a precise definition of truth and to prove its material adequacy. On the negative side, it is the source of the unusual and, from the philosophical point of view, strange character of Tarski's definition.

Criterion T lays down two conditions that a materially adequate definition of truth must meet. The second, and less important, condition is that the definition should imply that only sentences are true. Tarski imposes this condition because he takes truth to be a predicate of sentences. The first, and more important, condition is that the definition should imply all the 'T-biconditionals', that is, all sentences of the form

(T) '' is a true sentence (of English) iff .

Thus, Criterion T requires that a definition of truth (for English) should imply, for example, the famous T-biconditional,

'Snow is white' is a true sentence (of English) iff snow is white.

Note that any definition that implies this biconditional is bound to define a concept that agrees with truth over the sentence 'Snow is white': the sentence will fall under the concept if and only if it falls under truth. Hence, any definition that implies *all* the T-biconditionals is bound to define a concept *C* that agrees with truth over all the sentences: the concept defined is bound to be coextensive with truth (assuming that truth and *C* agree over non-sentences). Furthermore, since this argument uses no special contingent assumptions, the claim of

coextensiveness holds irrespective of the contingent facts. In one sense of 'intension', then, the concept defined is bound to have the same intension as truth.

The equivalence of the two sides of the T-biconditionals has been noted by many philosophers and logicians, including Aristotle, Gottlob Frege and Frank Ramsey. Tarski is the first, however, to have used the equivalence to formulate a material adequacy condition on a definition of truth. The simplicity of Tarski's formulation should not obscure the important and surprising claim that it contains. Definitions of virtually all of our concepts can be assessed for material adequacy only in a piecemeal way, namely, by comparing their consequences with the ordinary uses of the concepts defined. No general conditions of material adequacy can be laid down. One would have expected the same of truth. But Criterion T puts forward a general and a priori condition of material adequacy on a definition of truth - a condition that, surprisingly, is highly plausible. Criterion T embodies Tarski's key philosophical claim. If it is granted then there can be no denying that it marks an important philosophical advance.

The formulation of Criterion T given above, it must be acknowledged, is unsatisfactory in several respects. First, it is formulated only for a definition of truth *for* English *in* English. What is wanted - and what Tarski's works provide - is a more general criterion that also applies to other languages. Second, the criterion as formulated is plausible only for sentences that are free of ambiguity, indexicals, tenses and other context-dependent elements. Tarski himself never attempted to overcome this limitation - his primary interest was in formalized languages for which these problems do not arise - but other philosophers have attempted to do so. Third, and most important, the criterion as formulated requires the definition of truth to be inconsistent. This inconsistency is a consequence of the liar paradox.

2 The liar paradox

Suppose, following Saul Kripke, that we introduce a name 'Jack' into English with the understanding that it names the sentence 'Jack is not a true sentence (of English)'. Suppressing the parenthetical relativization, the T-biconditional for Jack is

'Jack is not a true sentence' is a true sentence iff Jack is not a true sentence.

Our stipulation ensures that

Jack = 'Jack is not a true sentence'.

Hence, by the principle of substitutivity of identicals, the T-biconditional for Jack implies

Jack is a true sentence iff Jack is not a true sentence.

This in turn yields a contradiction by the rules of propositional logic. Note that the difficulty cannot be overcome by stipulating away names such as 'Jack'. Self-referential 'liar' sentences, which yield an inconsistency, can be constructed without the aid of such names.

Tarski makes constructive use of the argument of the liar paradox. He uses it to prove (a version of) the indefinability theorem named after him: arithmetical truth is not arithmetical. More precisely, no formula of Peano arithmetic (PA) is true precisely of (the codes of) the true sentences of PA (see Gödel's theorems). The theorem has a more general version: no classical language L_1 that is rich in syntactic resources can contain its own notion of truth. L_1 can be extended to a new language L_2 by the addition of a predicate that represents 'true in L_1'. But, by Tarski's theorem, L_2 will not contain its own truth predicate. As before, L_2 can be extended to yet another language L_3; this will contain a truth predicate for L_2, but again not for L_3. By repeating this procedure, we obtain a Tarskian hierarchy of languages, $L_1, L_2, \ldots, L_n, \ldots$, in which truth for a language of level *n* is contained in the language at level $n + 1$. 'Tarski's indefinability theorem' motivates a way of getting around the difficulty that the paradoxes pose for Criterion T: (1) Draw a distinction between the language *in* which the definition of truth is given (the 'metalanguage') and the language *for* which the definition is given (the 'object language'); and (2) require that the object language not contain its own notion of truth. This manoeuvre ensures that no liar sentences occur in the object language. Consequently, the threat of inconsistency is removed from the T-biconditionals, and we regain Criterion T as a plausible material adequacy condition.

The manoeuvre allows us to proceed with the definition of truth for certain kinds of languages. But it does not help

us to understand the behaviour of truth in languages (such as English) that contain paradoxical sentences. Tarski's response to this problem was an impatient one: he dismissed all such languages as inconsistent. In the past twenty-five years or so, however, Tarskian ideas have been used to develop a better, and certainly less impatient, response to the problem. The response that is most Tarskian in spirit is the 'hierarchy theory' of Charles Parsons and Tyler Burge. This theory sees our ordinary notion of truth as implicitly hierarchical. Each occurrence of 'true', according to the theory, corresponds to a truth predicate at a certain level in Tarski's hierarchy - with different occurrences corresponding to predicates at possibly different levels. So, for example, in the sentence

'''Snow is white'' is true' is true.

the first occurrence of 'true' might correspond to the truth predicate at the first level of the hierarchy, and the second occurrence to that at the second level. The substantial burden on a hierarchy theory is to give a systematic explanation of the levels of the occurrences of 'true'. There are other approaches to the paradoxes that build on Tarski's ideas. The principal ones are the 'fixed-point theory' (see Semantic paradoxes and theories of truth §4) and the 'revision theory' (see Definition §4).

Tarski showed how to construct definitions of truth that meet Criterion T. The next section contains some examples of Tarskian truth definitions. But before turning to these, let us note some general features of Tarski's method. First, this method does not yield a definition of the relational notion 'true in *L*', with variable *L*; its primary application is to the definition of 'true in *L*' for particular *L*. Second, the method enables us to give truth definitions only for formalized languages, not for natural languages. Third, it yields a definition in a metalanguage *ML* that is distinct from the object language *L*. If *ML* is a classical language then the distinction is forced by Tarski's indefinability theorem.

3 Examples of Tarskian truth definitions

This section gives three examples of Tarskian truth definitions. Two of these are very simple; only the third begins to approach the complexity of Tarski's actual definition.

Example A. Suppose that *L* is a fragment of English that contains only a finite number of sentences (say, a million). Suppose that the sentences of *L* contain no occurrence of 'true' or of any other term whose use is inadmissible in a definition of truth. Further, let 'S_1', 'S_2', . . . , 'S_{10^6}' be abbreviations of the sentences of *L*. A definition of 'true sentence of *L*' may now be given as follows:

x is a true sentence of *L* iff S_1 and *x* is 'S_1', or S_2 and *x* is 'S_2', or…, or S_{10^6} and *x* is 'S_{10^6}'.

The definition is evidently formally correct. Apart from the vocabulary of $S_1, S_2, \ldots, S_{10^6}$, which is stipulated to be admissible, it uses only the logical notions 'or' and 'is' (identity) and the quotation names for the sentences of *L*. The definition is also materially adequate since, under the assumption of such simple syntactic facts as 'S_1' is not 'S_2', it implies all instances of the form (T).

Example B. Let us now construct a definition of truth for a language L' that has an infinite number of sentences. Suppose that L' has two sentential connectives, '$\sim$' (negation: 'it is not the case that') and 'Čamp;' (conjunction: 'and'). Suppose further that L' has a finite number of atomic sentences, that is, sentences that are not built using the connectives '$\sim$' and 'Čamp;'. Now truth for atomic sentences can be defined as in example A. Over compound sentences, truth can be defined recursively: the negation of a sentence *A* is true iff *A* is not true; a conjunction is true iff both its conjuncts are true. Putting these elements together we can construct an explicit definition of 'true sentence of L'' in the following way. First let us define an intermediary notion, that of a 'closed' set; this will enable us to give a more compact definition of truth. A set *Z* is 'closed' iff (1) *Z* contains all true, and only true, atomic sentences of L'; (2) *Z* contains the negation of a sentence *A* iff it does not contain *A*; and (3) *Z* contains the conjunction of sentences *A* and *B* iff it contains the conjuncts *A* and *B*.

x is a true sentence of L' iff *x* belongs to every closed set.

Assuming that the atomic sentences of L' do not contain any terms that are inadmissible, our definition is formally correct. Further, it is materially adequate, since it implies all the T-biconditionals of L'. Note that the implication requires some syntactic facts about L' and some low-level set-theoretic facts.

Example C. Finally, let us see how to construct a definition of truth for a quantificational language L''. Such a language has an infinite number of sentences, so the definition of truth cannot proceed in the list-like way of example A; it relies instead on a recursion. However, in these languages compound sentences may have immediate components that are not themselves sentences. For example, the quantificational compound '$(\forall v)Red(v)$' ('Everything is red') has as its immediate component '$Red(v)$', which is not a sentence. Consequently, in these languages, the truth of compounds cannot in general be explained recursively in terms of the truth of their components. To solve the problem Tarski first defines recursively a more general notion, namely, 'satisfaction', and then he defines truth using satisfaction.

Satisfaction is a relation that holds between 'assignments' (of values to variables) and formulas. We can think of assignments as functions from variables into objects. So, if *s* is an assignment and *v* is a variable then $s(v)$ is the value *s* assigns to *v*. Intuitively, an assignment *s* satisfies a formula *A* iff *A* is true when the free variables in it are interpreted as naming the values that *s* assigns them. The notion of satisfaction is easy to define for atomic formulas: any assignment that assigns a red object to the variable *v* satisfies the formula '$Red(v)$'. Similarly, any assignment *s* such that $s(v)$ is larger than $s(w)$ satisfies the formula '$Larger(v, w)$'. The clauses for *n*-ary predications are analogous. Satisfaction for complex formulas is defined recursively in terms of the satisfaction of the components:

An assignment *s* satisfies the negation of a formula *A* iff *s* does not satisfy *A*.
An assignment *s* satisfies the conjunction of *B* and *C* iff *s* satisfies *B* and *s* satisfies *C*.
An assignment *s* satisfies the universal quantification of *A* with respect to the variable *v* iff *A* is satisfied by all assignments s' such that for all variables v' distinct from *v*, $s'(v') = s(v')$.

This recursive explanation can be converted to an explicit definition of satisfaction (see example B). It is easy to show that the satisfaction of a formula *A* by an assignment *s* depends only on the values that *s* assigns to the free variables of *A*. Consequently, a sentence is satisfied either by all assignments or by none. This motivates the following definition of truth:

x is a true sentence of L'' iff *x* is a sentence of L'' that is satisfied by all assignments.

The reader will find a more leisurely presentation of this example in W.V. Quine's *Philosophy of Logic* (1970).

4 Philosophical influence

Tarski's avowed aim in constructing a definition of truth was to lay down the foundations of a scientific semantics, and it is here that his work has been most influential. Tarski's work provides the foundations for a semantic study of logic - a study that has been vigorously pursued in the twentieth century and to which Tarski made important contributions. Two key notions in this study - 'model' and 'truth in a model' - are generalizations of the notions of 'assignment' and 'satisfaction' introduced in example C. An assignment specifies the values of the variables; a model specifies the domain of quantification and the 'values' of all the nonlogical constants in the language. Furthermore, just as satisfaction tells us what formulas are true under what assignments, similarly 'truth in a model' tells us what formulas are true when the quantifiers and the nonlogical constants are interpreted as specified in the model. The notions 'model' and 'truth in a model' enable a semantic explanation of the various logical notions. For example, an argument is 'valid' if and only if its conclusion is true in all models in which all the premises are true. The semantic analysis of logical concepts has yielded valuable insights into logic (see Model theory).

Tarski's definition of truth is also of importance in natural-language semantics. Here it has inspired two different kinds of approaches: the Davidsonian approach and the model-theoretic approach. The Davidsonian approach takes the goal of a semantic theory for a language *L* to be the construction of a theory of truth for it: a finite theory that implies all the T-biconditionals of *L*. (This is only a rough characterization; see Meaning and truth.) Davidson has argued that such a theory must proceed recursively, explaining semantic properties of compounds in terms of the semantic properties of the components. As a result, the theory will reveal the logical structure of language. The model-theoretic approach, developed by Richard Montague and others, takes the goal of the semantic theory to be the definition of the notion 'truth in a model'. This requires the characterization of the notion of model and it, in turn, requires an analysis of the logical structure of language. Davidson's work has been highly influential in

philosophy, but less so in linguistics. The model-theoretic approach has, however, gained many adherents in linguistics.

Tarski gave his definition in a philosophical climate that was sceptical of the notion of truth. Some logical positivists suggested, for example, that the notion was unanalysable in empirical terms and, hence, unacceptable in a scientific philosophy. Tarski's definition provided a decisive refutation of their suggestion. Let *L* be the positivists' language - purified of all elements that the positivists judge to be unempirical and meaningless. The notion 'true in *L*', Tarski showed, can be rigorously defined using only resources that the positivists would accept: the definition uses only some syntactic and logical concepts and those that are expressed in *L*. Tarski's results were accepted by the positivists and significantly influenced their philosophical work (especially that of Rudolf Carnap).

Tarski's definition has also been influential in more recent philosophy. Some philosophers (for example, Hartry Field (1972) and Michael Devitt) have sought to construct a correspondence theory of truth on its basis. These philosophers accept the recursive parts of Tarski's definition. But they argue that the definition does not provide a substantial account of the semantic properties of the simplest elements of the language. It does not explain, for example, when a name denotes an object or when an object falls under a predicate. To obtain a fully fledged correspondence theory, they suggest, Tarski's recursive explanation of truth should be supplemented with a substantial account of the semantics of the simples. Some philosophers have seen in this kind of theory a way of vindicating realism. But others have questioned both the feasibility of the theory and its links to the realism issue.

It is ironic that while Tarski put forward his definition to show that the concept of truth is useful for certain logical and philosophical purposes, his ideas have also been used in defence of 'deflationism' - the view that the concept has no substantive role to play in logic and philosophy. The thought here is that something like Tarski's definition explains completely the meaning of 'true', and that this makes truth a philosophically lightweight concept. Hence, it is claimed, the concept can play no substantive role in, for example, the explanation of validity or the realism debate. Consider, for instance, the explanation of validity - say, of the argument *P*, therefore *Q* - in terms of truth: the argument is valid iff it is impossible that *P* be true but *Q* false. The T-biconditionals, the deflationists argue, reduce this explanation to the nearly vacuous claim that the argument is valid iff it is impossible that *P* and not *Q*. Neither the explanation nor the concept of truth seem to be doing any substantive work.

The role of the concept of truth in philosophy continues to be a topic of intense debate. In this debate, Tarski's definition and Tarski's ideas occupy centre stage.

See also: Consequence, conceptions of §5; Logical and mathematical terms, glossary of; Semantics

ANIL GUPTA

References and further reading

Etchemendy, J. (1988) 'Tarski on Truth and Logical Consequence', *Journal of Symbolic Logic* 53: 51- 79. (Nontechnical; makes useful critical points.)

Field, H. (1972) 'Tarski's Theory of Truth', *Journal of Philosophy* 69: 347-75; repr. in M. Platts (ed.) *Reference, Truth and Reality*, London: Routledge & Kegan Paul, 1980.(Referred to in §4; argues for a transformation of Tarski's definition into a correspondence theory of truth.)

Gupta, A. and Belnap, N.D. (1993) *The Revision Theory of Truth*, Cambridge, MA: MIT Press. (Chapters 1 and 4 are nontechnical and discuss Criterion T and the liar paradox; the rest of the book is quite technical.)

Horwich, P. (1990) *Truth*, Oxford: Blackwell.(A clear defence of deflationism.)

Quine, W.V. (1970) *Philosophy of Logic*, Englewood Cliffs, NJ: Prentice-Hall.(Referred to in §3; presents a Tarskian truth definition in detail.)

Tarski, A. (1936) 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica* 1: 261-405; trans. J.H. Woodger, 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Oxford: Clarendon Press, 1956; ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(*The* classic paper on the subject; the introduction and the first section of the paper are nontechnical and are highly recommended.)

Tarski, A. (1944) 'The Semantic Conception of Truth and the Foundations of Semantics', *Philosophy and Phenomenological Research* 4: 341-76; repr. in L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University of Illinois Press, 1952.(Nontechnical; includes valuable philosophical material.)

Tarski, A. (1969) ‘Truth and Proof’, *Scientific American* 220 (6): 63-77; repr. in R.I.G. Hughes (ed.) *A Philosophical Companion to First-Order Logic*, Indianapolis, IN: Hackett Publishing Company, 1993. (A particularly good popular exposition of the main ideas.)

Tauler, John (*c.*1300-1361)

Tauler was a Dominican preacher and mystic, the author of seventy-nine vernacular sermons which presented the Neoplatonic speculative mysticism of his teacher Eckhart in more personal and concrete terms. Tauler greatly influenced late medieval mendicant spirituality, Luther, and the German pietists and romantics. His primary aims were to inculcate the spiritual virtues and devotion to God.

Tauler, his friend Henry Suso and Meister Eckhart are the central figures in the Rhineland school of mysticism. Tauler enrolled in the Dominican *studium generale* in Cologne in 1324, together with Suso, but little is known of a life devoted to preaching and spiritual direction in the Dominican nunneries and Beguines around Strasbourg and Basel. Among Tauler's spiritual contacts were Suso, who was like Tauler also a leader of and moderating influence on the Friends of God who were inclined to antinomian spirituality, the influential mystical Dominican nun Margareta Ebner and the famous Rhenish mystic John Ruusbroec.

The social situation during Tauler's maturity was chaotic. The Strasbourg Dominicans were exiled in Basel (1339-43) during the time the anti-papal civic authorities supported Emperor Louis. The ensuing civil war and the Black Death strained conventional morality, creating a challenge for a preacher calling lay people to the spiritual life. These conditions may have contributed to the ferocity and intense gloom of some of the sermons. In any case, Tauler's sermons are more personal and emotional than Eckhart's but generally more moderate and balanced than Suso's. At his death Tauler had published nothing; the sermons were written down in the Middle High German vernacular by listeners who were most often Dominican nuns.

Tauler's sermons combine biblical commentary with German mystical themes, supplemented by distinctions drawn from scholastic thinkers, especially Peter Lombard and Thomas Aquinas. They orient pastoral concern towards spiritual transformation, and combine various strands of medieval Christian mysticism: practical monastic spirituality, Dominican intellectualist mysticism, and beatitude and divinisation. Philosophically, Tauler was most influenced by the Neoplatonised Aristotelianism of Albert the Great and the stronger Neoplatonism of Pseudo-dionysius and Meister Eckhart. Notably absent from the sermons is the love imagery prominent in medieval exponents of *Song of Songs* mysticism. Tauler is deeply versed in Eckhart's teachings; thus, he regularly employs the themes and concepts of Dominican mystical psychology, though his language is usually more concrete. Fundamental are the notions of 'the ground', 'summit' or 'spark' of the soul, the mysterious essence of human nature that can touch the ineffable divine nature, the 'imageless Image' which is the Trinity. Paradoxically, however, the soul returns to God by returning to itself: 'This spark flies up to the summits where its true place is, even beyond this world, where intelligence cannot follow it, for it does not rest until it returns to that centre from which it originates and where it used to be in its uncreated state' (Sermons 64, 2). Clearly, Tauler suscribes to the 'breakthrough' (*Durchbruch*) of the soul into eternity, but unlike Eckhart, he favours an account of divinization consistent with the orthodox distinction between God and creature. This and other departures from his teacher's extreme views should be understood in part as precautions against the ecclesiastical pressures increased by Eckhart's condemnation in 1328.

Negative theology becomes a supple psychological instrument in Tauler's hands (see Negative theology). Instead of emphasising Eckhart's heterodox notion of the soul's birth as the Son, Tauler counsels humility, self-denial and detachment (*Gelassenheit*) - also Eckhartian themes. Meditative concentration on the humanity of Christ entails for Tauler the negation of overly abstract theological concepts and the abandonment of free will, external goods and attachments, as in Matthew 19: 29: 'whoever leaves father and mother and all possessions for my sake shall receive a hundredfold and eternal life'. Tauler's own words serve well as the motto of this more homely, devotional spirituality: 'The lower the descent, the higher the ascent'. Like Suso, Tauler is more inclined than Eckhart to explain human error and sin and their overcoming through ascetic exercises and the sacraments. Unlike his friend, however, Tauler does not speak explicitly of his own spiritual experiences. Paradigmatic stages of the spiritual life are explored via scriptural exegesis. Instead of Suso's imaginative flights Tauler employs vivid imagery drawn from common life, crafts and popular examples to convey abstract points. His overall aim is to strike a balance between the aridity of scholastic theology and the excesses of pseudo-mystical enthusiasm, with its widespread dangers of mistaking inner spiritual freedom with libertinism.

See also: Meister Eckhart; Mysticism, history of; Neoplatonism; Platonism, medieval

JOHN BUSSANICH

List of works

Tauler, John (*c.*1300-61) Sermons, ed.F. Vetter, *Die Predigten Taulers*, Frankfurt: Weidmann, 1910; 2nd edn, 1968(A good English translation of selected sermons is E. Colledge and Sister M. Jane (ed. and trans.) *Spiritual Conferences by John Tauler*, Rockford, IL: Tan Books, 1978. *Johannes Tauler: Sermons*, trans. M. Shrady, New York: Paulist Press, 1985, is a fine collection of twenty-three sermons in a good English translation, accompanied by a useful preface by A. Haas and an excellent introduction by J. Schmidt.)

References and further reading

de Gandillac, M. (1956) *Valeur du temps dans la pédagogie spirituelle de Jean Tauler*, Montreal, Que.: Institut d'études medievales.(A fine assessment of Tauler's spirituality.)

Haas, A. (1971) *Nim din selbes war. Studien zur Lehre von der Selbsterkenntnis bei Meister Eckhart, Johannes Tauler und Heinrich Seuse*, Freiburg: Universtätsverlag. (Excellent account of the theme of self-knowledge in the Rhineland mystics.)

Ozment, S.E. (1969) *Homo Spiritualis: A Comparative Study of the Anthropology of Johannes Tauler, Jean Gerson and Martin Luther (1509-16)*, Leiden: Brill.(Valuable exploration of Tauler's mysticism and its influence.)

Taxonomy

The fundamental elements of any classification are its theoretical commitments, basic units and the criteria for ordering these basic units into a classification. Two fundamentally different sorts of classification are those that reflect structural organization and those that are systematically related to historical development.

In biological classification, evolution supplies the theoretical orientation. The goal is to make the basic units of classification (taxonomic species) identical to the basic units of biological evolution (evolutionary species). The principle of order is supplied by phylogeny. Species splitting successively through time produce a phylogenetic tree. The primary goal of taxonomy since Darwin has been to reflect these successive splittings in a hierarchical classification made up of species, genera, families, and so on.

The major point of contention in taxonomy is epistemological. A recurrent complaint against classifications that attempt to reflect phylogeny is that phylogeny cannot be 'known' with certainty sufficient to warrant using it as the object of classification. Instead, small but persistent groups of taxonomists have insisted that classifications be more 'operational'. Instead of attempting to reflect something as difficult to infer as phylogeny, advocates of this position contend that systematists should stick more closely to observational reality.

1 The goals of taxonomy

Any set of entities can be classified in indefinitely many ways. Books can be classified according to author, title, subject matter, and so on. Although these various classifications can be integrated into a single reference system so that any book can be retrieved on a variety of different counts, books as physical objects can be arrayed on library shelves in only one order. Even here, contingent problems arise. For example, very large books may have to be shelved out of order, and expensive first editions may be locked away in separate collections.

More stringent requirements apply to scientific classifications. The ultimate goal for scientific classifications is to group entities so that these classes function in, or facilitate the formation of, scientific laws (see Laws, natural §1). Aristotle divided motion into super- and sub-lunar as well as forced and natural (see Mechanics, Aristotelian). The primary justification of his classification was the system of laws that he was able to generate using it. When Newton introduced his quite different system, his classification replaced Aristotle's because Newton's system of laws was more powerful, accurate and inclusive. In general, systems of scientific classification are intimately connected to scientific theories and cannot be evaluated independently of them. Different sorts of theories require different classifications.

One major difference is between structural and historical classifications. The periodic table of physical elements is structural. The elements are individuated and ordered linearly according to their atomic number. Hydrogen comes first, then helium, lithium, and so on. These elements in turn can be arranged hierarchically as metals, rare earths, and so on. Some of these arrangements are perfectly nested; others are not. A more contemporary classification would include reference to subatomic particles and their relations (see Chemistry, philosophical aspects of §4). In general, structural hierarchies do not include very many levels. They are not very deep. Although cosmology is a legitimate area of physics, no one has suggested a historical classification of the physical elements; for example, classifying them in the order in which they appeared after the Big Bang.

Historical classifications are more prominent in other areas of scientific investigation. For example, linguists produce both structural and historical classifications of languages. Phonemes are the basic units of the structural classifications. Historical linguists also reconstruct the history of languages, indicating which evolved from which and how various sounds changed through time. One sign that a classification is historical rather than structural is the number of levels it exhibits. Historical classifications tend to be much deeper than structural classifications.

The logical structure of structural classifications is quite intuitive. Geometric figures can be divided into those that have three, four, five, and more sides. Three-sided figures in turn can be divided into triangles that have three, two and no sides of equal length. The logical structure of historical classifications also appears to be intuitively clear. It is anything but. Appearances notwithstanding, historical relations, especially those that form genealogical trees, are not easily represented in hierarchical classifications. Because these issues have been worked out most extensively with respect to classifications of living creatures, and because biological classifications are among the

most familiar, the rest of this discussion concentrates on biological classifications.

2 The basic units of classification

The organisms with which people are most familiar seem to fall naturally into discrete kinds. No one has any trouble telling apart cats, dogs, elephants and whales. However, this ease stems from limitation of perspective. If a particular species is followed across its range, problems frequently arise as the traits that characterize the species change geographically. Populations of a species in one locale frequently differ statistically from populations at other locales. If biological species are traced through successive geological strata, the same sort of variation occurs. In addition, not all organisms seem to form species of the same sort. For example, organisms that reproduce sexually may well form species that are significantly different from the species formed by organisms that reproduce asexually. Plant species may differ systematically from animal or protist species. For example, hybrid species are rare among animals but common among plants.

Most systematists are willing to put up with these difficulties in order to produce classifications that reflect the evolutionary process. Others, however, have suggested subdividing organisms into basic units of the sort that can be recognized with greater ease and reliability. For example, the basic units of classification might reflect some one measure of overall similarity. One of the problems with this suggestion is deciding on the measure. Another is that, no matter the measure, many of the resulting ‘species’ are exceedingly nonstandard; for example, males and females of the same evolutionary species may be placed in separate ‘species’. Still other systematists have proposed to define species so that males, females and their immediate descendants are included in the same species, but thereafter all that is required is one or more diagnostic characters. Species can be defined in numerous different ways for different purposes, but systematists must settle on a single way for their classifications (see Species §1).

3 Homology and monophyly

For centuries systematists attempted to produce structural classifications of living creatures. At times, they thought that they had found criteria that promised to group organisms into something like the periodic table, but insuperable problems always arose. Living creatures seemed to embody two different kinds of similarity. As much as sharks, boney fishes and whales look alike on the basis of some characteristics, they are fundamentally different on other sets. With the introduction of Darwin’s theory of evolution, systematists realized why they were having such a difficult time. Biological kinds (taxa) are characterized by both structural and historical relations, and no one classification can represent both (see Darwin, C.R.; Evolution, theory of).

Nothing requires systematists to accept descent as the relation that they represent in their classifications. If some other principle of order had proven feasible and productive, it might have been adopted. But once systematists accepted the task of constructing classifications that reflect genealogy, they had to rework their basic conceptual system from the ground up, redefining terms that had no genealogical component so that they did. For example, the terms ‘homology’ and ‘analogy’ originally implied nothing about descent but concerned differences in structural patterns. Evolutionists redefined these terms to indicate descent. Apes, fish and amphibians all have eyes, and these eyes are homologies because they evolved as eyes in a common ancestor. They are vertebrate eyes. Cephelapods also have eyes, but these ‘eyes’ cannot count as the same character as vertebrate eyes because they evolved independently. Similarly, snakes belong to Tetrapoda even though they have lost their legs. Having lost a character is importantly different from never having possessed it in the first place.

According to evolutionary systematists, all taxa must be monophyletic. However, some systematists are satisfied with a weak notion of monophyly, while others insist on a much stronger notion. On the weak definition, all species that make up a monophyletic taxon must arise from a single immediately ancestral stem species. Willi Hennig and his intellectual descendants (termed ‘cladists’) are convinced that a much stronger notion of monophyly is required. Not only must all species included in a monophyletic taxon be descended from a single stem species, but also all the species descended from a single stem species must be included in a single higher taxon. One problem with this stronger definition of ‘monophyly’ is that it obliterates numerous time-honoured taxa. For example, Reptilia has been part of biological classifications for centuries. However, Reptilia is not monophyletic in this stronger sense because some organisms that share an immediate common ancestor with reptiles are nevertheless not included in Reptilia, for example, birds.

4 Taxonomic hierarchies

Because phylogenetic trees consist of successive branches and hierarchical classifications consist of successively less inclusive taxa, systematists were led for a long time into thinking that the relationship between the two systems is quite straightforward. It is not. The source of this mistake can be seen in Figure 1(A), which represents a typical phylogeny. The intuitive classification of this phylogeny is presented in Figure 1(B). A moment's reflection should reveal that something is wrong with this classification. The phylogenetic tree is made up of five species, while the classification lists only two species. The taxon listed as Tribe *a* is not a Tribe but a species.

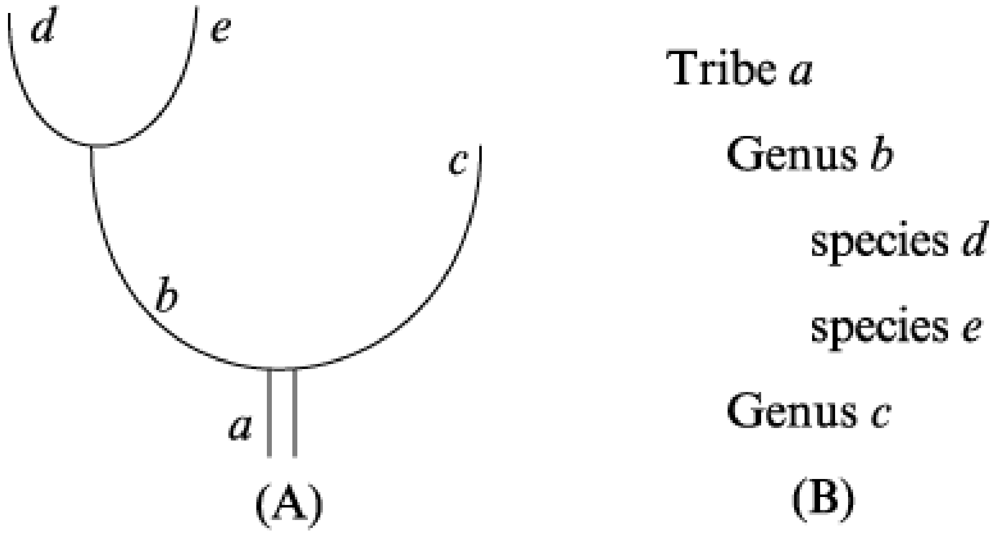

Figure 1. A representative phylogeny (A) and an intuitive but mistaken way of classifying it (B)

A more appropriate classification of the phylogeny depicted in Figure 1(A) is presented in Figure 2(A). The way in which this classification groups species into higher taxa is represented in Figure 2(B). In this classification, the five species are at least listed as species. However, problems remain. All the taxa in this classification are monophyletic in the weak sense. Each stems from a single ancestral species. But not all of these taxa are monophyletic in the stronger sense. The source of the problem is, surprisingly, the presence of common ancestors. No collection of species that includes one or more species that are ancestral to the others can be classified monophyletically in the strong sense. If species *a* in Figure 1(A) is classified with species *b*, then species *c* is left out. Conversely, if species *a* is classified with species *c*, then species *b* is left out. As a result, no classification that includes common ancestors can be monophyletic in the strong sense.

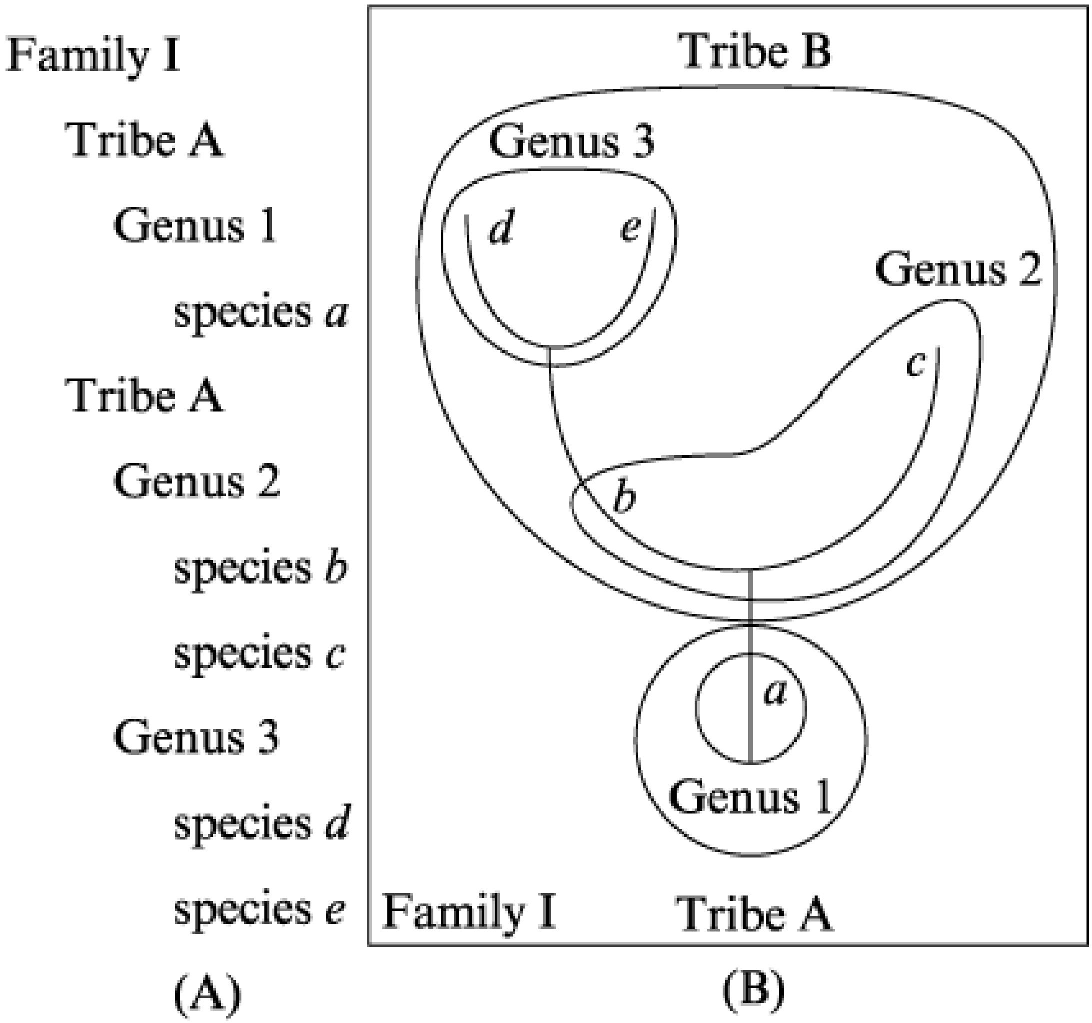

Figure 2. Classification (A) of a phylogeny and that same classification indicated on that phylogeny (B)

Parallel problems arise at the level of characters. If the list of species to be classified happens to include a common ancestor, then no single set of nested characters can be found to define the relevant higher taxa. Some characters will imply that species *a* should be classed with species *b*; other characters will imply that species *a* should be classed with species *c*. Every time an ancestral species is included in a taxonomic study, the result will be conflicting character distributions. In fact, one hint that a set of species under investigation includes common ancestors is continued conflict in character distributions as the study progresses.

The source of the preceding problem is not nature but the structure of the Linnaean hierarchy. The Linnaean hierarchy was developed for structural classifications (see Linnaeus, C. von). When the goal of systematics was changed to have classifications reflect genealogy, no one noticed the deep problems posed by any attempt to translate a historically connected tree into nested sets of taxa. The traditional Linnaean hierarchy is not up to the task. It cannot represent splitting. Nor can it represent the symmetrical relation - merger. Species of hybrid origin are no easier to represent than are common ancestors.

5 Epistemology and the goals of classification

Throughout the history of biological systematics, conflicts have arisen over the kind and degree of certainty that is required in science. At one extreme, some systematists are satisfied to produce classifications based on a host of tacit evaluations and patchy evidence. For them, classification is mainly an art. At the other extreme, are those systematists who insist that intuitive guesses must be replaced by strict quantitative techniques. These systematists

also view scientific theories with suspicion. Two sorts of 'theory' have come under special scrutiny - phylogeny reconstruction and evolutionary theory itself. Systematists have rediscovered a problem long familiar to philosophers. How can one know that a particular chunk of metal is gold unless one knows what gold is, and how can one know what gold is without inspecting some samples of gold? But if one does not know what gold is, one cannot decide what to inspect….

The appropriate response to the preceding question is not all that satisfying. On the basis of very little knowledge mixed with lots of error, one develops a conception of gold. On the basis of this imperfect conception, additional investigations are performed, and these investigations uncover past errors, which increase our ability to discern gold, and so on. On the basis of sketchy data and an imperfect understanding of how evolution occurs, one attempts to reconstruct phylogeny. From these imperfect reconstructions, evolutionary theory is improved, which permits better phylogenetic reconstructions, and so on (see Scientific method §2).

Although character covariations are the primary evidence in phylogeny reconstruction, they are not the sole evidence. Two sessile species that have equally sessile propagules cannot possibly be closely related phylogenetically if the one lives in Mexico and the other in Ethiopia. The goal of making scientific method as simple, straightforward and reliable as possible is justified, but such methodological injunctions should not be made so rigorous that the progress of science is stopped in its tracks. In the last analysis, classifications succeed or fail on the basis of the use that scientists can make of them.

See also: Theories, scientific

DAVID L. HULL

References and further reading

Ereshefsky, M. (ed.) (1992) *The Units of Evolution: Essays on the Nature of Species*, Cambridge, MA: MIT Press. (A collection of classic papers on the species' problem.)

Mayr, E. and Ashlock, P.D. (1991) *Principles of Systematic Zoology*, New York: McGraw-Hill, 2nd edn.(A discussion of the principles of traditional evolutionary taxonomy.)

Ridley, M. (1986) *Evolution and Classification: The Reformation of Cladism*, London and New York: Longman. (A philosophical critique of cladistic analysis.)

Sneath, P.H.A. (1995) 'Thirty Years of Numerical Taxonomy', *Systematic Biology* 44: 281-98.(An evaluation of numerical taxonomy by one of its founders.)

Sober, E. (1988) *Reconstructing the Past: Parsimony, Evolution, Inference*, Cambridge, MA: MIT Press.(A philosophical evaluation of cladistic methods for reconstructing the past.)

Sokal, R.R. (1986) 'Phenetic Taxonomy: Theory and Methods', *Annual Review of Ecology and Systematics* 17: 423-42.(A summary of the principles of phenetic taxonomy by one of its founders.)

Wiley, E.O., Siegel-Causey, D., Brooks, D.R. and Funk, V.A. (1991) *The Complete Cladist: A Primer of Phylogenetic Systematics*, Lawrence, KS: Special Publication, Museum of Natural History.(A discussion of the principles of cladistic analysis.)

Taylor, Charles (1931-)

Among the most influential of late twentieth-century philosophers, Taylor has written on human agency, identity and the self; language; the limits of epistemology; interpretation and explanation in social science; ethics; and democratic politics. His work is distinctive because of his innovative treatments of long-standing philosophical problems, especially those deriving from applications of Enlightenment epistemology to theories of language, the self and political action, and his unusually thorough integration of 'analytic' and 'continental' philosophical concerns and approaches.

Taylor's work is shaped by the view that adequate understanding of philosophical arguments requires an appreciation of their origins, changing contexts and transformed meanings. Thus it often takes the form of historical reconstructions that seek to identify the paths by which particular theories and languages of understanding or evaluation have been developed. This reflects both Taylor's sustained engagement with Hegel's philosophy and his resistance to epistemological dichotomies such as 'truth' and 'falsehood' in favour of a notion of 'epistemic gain' influenced by H.G. Gadamer.

1 Explanation and interpretation

In his most prominent early publications, Taylor addressed the status of explanation in psychology and the social sciences. *The Explanation of Behaviour* (1967) challenged the adequacy of behaviourism in psychology, principally by arguing that accounting for intentional action entails a teleological understanding of the ends of action that cannot be achieved within purely causal theories. 'Interpretation and the Sciences of Man' (1971) extended this argument to politics and social analysis, showing that attempts at explanation in terms of external approaches to 'brute facts' not only fail to satisfy those who seek more meaningful understandings of human agency, but are also incoherent or incomplete on their own terms. These studies are all parts of a more general argument against the epistemology that both inspired and drew strength from the scientific revolutions of the seventeenth century and after. Taylor holds that this epistemological tradition both relied on an atomistic account of putatively undifferentiated nature (including human nature) and erected a perniciously sharp distinction between knower and object of knowledge, with the result that it drastically distorted and narrowed the scope of understanding of human life (see Behaviourism, methodological and scientific §1).

2 Hegel

The same issues motivate much of Taylor's engagement with Hegel. Hegel too challenged the reigning epistemological tradition and especially its atomistic inattention to the necessary relatedness of all subjects and objects, and to the internal differentiation of both subjects and objects. He argued that any adequate account of the human subject must rely on an understanding of persons as existing only in interaction, as becoming individuals only through participation in an intersubjective reality. Taylor argues that this richer account of the person requires an understanding of language as not merely neutrally picking out objects of attention or reflecting pre-given inner states, but as helping to constitute phenomena and our understanding of external phenomena. Following Johann Herder (see Herder, J. §2), Taylor places these positions on bases significantly different from Hegel's.

This is necessary because Taylor holds that Hegel ultimately failed to achieve the rational certainty about the absolute that he sought. Hegel's arguments reveal an interpretive vision of power and insight, but not a system of determinate necessity. Building then on the critical foundation he shares in large part with Hegel, but rejecting the more extreme claims of Hegel's *Logic* and related elements of his substantive philosophical anthropology and social theory, Taylor has sought to advance an understanding of the nature and activities of and relations among human subjects and of the kind of science that can grasp these subjects, their relations and activities. This entails moving 'beyond epistemology', but not following Hegel in the attempt to ground all argument in ontology.

3 Language

Dualist epistemology is predicated on a rigid separation of subject and object that makes us unable to grasp distinctive features of human life and activity as distinct from the behaviour of physical objects and natural systems. Taylor moves beyond this by developing 'expressivist' or 'constitutive' theories of language, inaugurated by Herder and important to the Romantic tradition, but often left inadequately grounded in an appeal to immediate

self-knowledge. These theories help Taylor to show the human agent to be understandable only as a participant in a linguistic community. Correspondingly, language itself cannot be understood entirely as a matter of reference and predication; instead, Taylor suggests that linguistic activity also involves constructing objects by making identifications of significance that cannot adequately be rendered in truth-conditional form, and that agents are never in a purely external relationship to language, nor indeed to the rest of their worlds in so far as these are constituted through language.

4 Strong evaluation and the self

In a similar fashion, Taylor argues that human agents necessarily engage in different orders of cognition and evaluation. We engage in practical reason always against a background of ‘strong evaluations’. These are simultaneously intellectual and moral commitments that constitute us as knowers and judgers, and that make possible our more specific and immediate knowledge and judgments. Such commitments may vary, but are necessary to the constitution of the self.

Modern moral philosophy has tended to focus on what it is right to do rather than on what it is good to be, and on defining the content of obligation rather than the nature of the good life. In this, it reflects the reliance on the notion of a disembodied, decontextualized and disengaged subject pioneered by post-seventeenth-century science and the epistemology to which it helped give rise. Accordingly, it is a philosophical priority to reconstruct both the dominant modern understandings of the self and alternative interpretations of human agents. Taylor takes on this task in *Sources of the Self* (1989). A key feature of human agency, he shows, is that it *is* constituted only within frameworks of strong evaluation - whether these are traditional notions of the primacy of honour, Platonic accounts of the virtues of reason and self-mastery, modern understandings of the expressive power of inner selves or the virtues of counting everyone’s interests equally. The historical story of the changing character of the modern self is thus inextricably an account of the transformation of more capacities, because these are rooted in changing constructions of agency. Changes in the idea of self, moreover, were often driven directly by attempts to resolve moral or religious problems, though their long-term results were sometimes to undermine the theological or other commitments that give rise to the new conceptions. A crucial moment in this process was the transformation in evaluation of ordinary life, the movement of the world of work and family from the margins of morality to the centre of the modern agent’s moral commitments. This helped to make possible new positive understandings of the self as a physical, including sexual being, and contributed both to utilitarianism, with its reckonings of all manner of satisfactions without reference to the hierarchy that had previously denigrated those of ordinary life, and to Romanticism, with its understanding of the primacy of individual expression (see Subject, postmodern critique of the §1).

5 Communitarianism and multiculturalism

Although these modern transformations of the self lead to new capacities for individuation and fulfilment in interpersonal relationships, these new capacities give rise to new ethical and political challenges. Increased attention to intersubjective community is therefore important not only for philosophical accuracy but for moral life and personal satisfaction. Taylor is thus part of the diverse philosophical and political movement termed ‘communitarianism’. In addition to his enduring emphases on the constitutive role of language and the intersubjective nature of agency, in his work of the 1990s Taylor focused attention on the sociological dimensions of community, including especially the nature of nationalism and the irreducibly social nature of some human goods.

Drawing on his analysis of the modern self, Taylor shows in *The Ethics of Authenticity* (1991) how the search for authentic self-fulfilment can become incoherent and self-defeating when it is tied to atomistic individualism, the overvaluation of instrumental reason and an alienation from public life. At the same time, he argues against pessimism, suggesting that the other elements of our philosophical and cultural traditions give us resources for confronting our current crisis - including the recognition that our wants are necessarily qualitatively distinguishable (so that, among other things, we can want to have better wants), that our individuality is grounded in sociality (so that we can conceive of freedom in ways other than absence of external constraint), and that frameworks of strong evaluation are inescapable (so that the attribution of significance is not simply a matter of immediate subjective choice).

Among the most important themes in this work is a renewed link to Hegel. In many different versions of the fragmentation of political life, Taylor sees a common theme of competing demands for recognition of the legitimacy or value of different identities. This ‘politics of recognition’, appearing in nationalism, ethnic politics, feminism and multiculturalism, is an outgrowth of the modern valuation of self and ordinary life. Claims often assert the rightness and value of differences among people, in contradiction to earlier politics that stressed universal dignity by recommending blindness to differences. Many are incoherent, however, in demanding a recognition of equal worth that can only be met by a ‘soft relativism’, since it is demanded in advance of genuine evaluative engagement. There is no resolution to this dilemma in pure individualistic liberalism because of its homogenizing conception of the person and consequent incapacity to provide a sense of significant differentiation so that partial communities can be centres of value within larger politics in ways that connect members to the whole. A presumption of mutual respect is a useful beginning, but also a ‘mere ought’, unless linked to a notion of the self as, first, necessarily socially engaged rather than merely observing from an external vantage point; second, limited in its capacity for understanding by the very cultural frameworks that make its individuality and understanding possible; and, third, open to change through communicative interaction. Such a notion of the self fits with the aspiration to combine full moral autonomy with the recovery of community both expressive of the common life of its members and constitutive of their individuality (see Community and communitarianism §§2-3).

CRAIG CALHOUN

List of works

Taylor, C. (1967) *The Explanation of Behaviour*, London: Routledge.(A critique of psychological behaviourism, emphasizing the limits of purely causal theories and the need for teleological accounts of the ends of action.)

Taylor, C.(1971) ‘Interpretation and the Sciences of Man’, *The Review of Metaphysics* 25 (1): 3-51.(A classic essay on the inevitability of interpretation to the human sciences, and the consequent impossibility of pure objectivist behaviourism.)

Taylor, C.(1975) *Hegel*, Cambridge: Cambridge University Press.(A sustained inquiry into Hegel’s philosophy, establishing the unity of his thought, the embeddedness of his project in the problems of its historical epoch, the reasons for the ultimate failure of that project, and the insights it none the less offers for philosophy and social thought.)

Taylor, C.(1977) *Hegel and Modern Society*, Cambridge: Cambridge University Press.(A consideration of Hegel’s contribution to grasping the distinct nature of modern society and its enduring problems.)

Taylor, C. (1985a) *Philosphical Papers, vol. 1: Human Agency and Language*, Cambridge: Cambridge University Press.(Taylor’s major early papers addressing themes of agency, language and the intersubjective constitution of the person.)

Taylor, C. (1985b) *Philosophical Papers, vol. 2: Philosophy and the Human Sciences*, Cambridge: Cambridge University Press.(Taylor’s major early papers on the social and behavioural sciences, including Taylor (1971).)

Taylor, C. (1989) *Sources of the Self*, Cambridge, MA: Harvard University Press.(Taylor’s investigation of the genesis of modern Western constructions of the self, emphasizing that agency is only constituted within frameworks of ‘strong evaluation’, and studying how these change with attempts to resolve moral and religious problems.)

Taylor, C. (1991) *The Malaises of Modernity*, Toronto, Ont.: Anansi; also publ. as *The Ethics of Authenticity*, Cambridge, MA: Harvard University Press.(A series of lectures examining the challenges posed to the modern era by tying the search for self-fulfilment to atomistic individualism, an overvaluation of instrumental reason, and alienation from public life, and arguing for the inadequacy of ‘soft relativism’ when faced with significant cultural and interpersonal differences.)

Taylor, C.(1995) *Philosophical Arguments*, Cambridge, MA: Harvard University Press.(A collection of Taylor’s major recent papers on themes ranging from moral obligation to public goods and the self to nationalism, including the widely debated ‘The Politics of Recognition’.)

References and further reading

Calhoun, C. (1991) ‘Morality, Identity and Historical Explanation: Charles Taylor on the Sources of the Self’,*Sociological Theory* 9 (2): 232-63.(An analysis of Taylor’s account of transitions in self and moral reasoning.)

Gutmann, A. (ed.) (1994) *Multiculturalism and the Politics of Recognition*, Princeton, NJ: Princeton University Press.(A collection of responses to one of Taylor’s key arguments by Anthony Appiah, Jürgen Habermas,

Michael Walzer and others).
Phillips, D. (1993) *Looking Backward: A Critical Appraisal of Communitarian Thought*, Princeton, NJ: Princeton University Press.(An unsympathetic but wide-ranging critique of Taylor and other communitarians in favour of a version of liberalism.)
Tully, J. (ed.) (1994) *Philosophy in an Age of Pluralism*, Cambridge: Cambridge University Press.(A collection of papers, mainly concerning *The Sources of the Self*, with a reply by Taylor.)

Taylor, Harriet (1807-58)

Harriet Taylor was John Stuart Mill's intellectual collaborator and great love. Married to John Taylor in 1826, Harriet met Mill in 1830 and they began a brazenly unconventional intimacy which lasted throughout her life. Her thoughts on poetry, equality, liberty and individuality shaped Mill's work on these topics.

Harriet Taylor, *née* Hardy, was born and raised in Walworth, England, where she married John Taylor in 1826. She was introduced to John Stuart Mill in 1830. At the time, she was a restless young wife and mother, deeply involved in a coterie of unitarian radicals whose liberal politics, feminism, and literary and philosophical activities suited her exactly. She had hoped to become a writer, but put aside this aspiration in order to become Mill's intellectual partner. Her amiable and generous husband acquiesced in the peculiar arrangement for her sake. In 1851, two years after Taylor's death, Taylor married Mill, and the couple lived and worked together until her death in 1858. In 1859, Mill began publishing the work of their marital collaboration, extravagantly dedicating everything to her.

Their first collaborative work was the chapter on the labouring classes in Mill's *Principles of Political Economy* (which, published in 1848, prompted an overblown dedication from Mill, discreetly pasted into a few gift copies). Taylor also directed his work on the writings later published as *Three Essays on Religion* (1874). In 1851, she wrote 'The Enfranchisement of Women', which argued for full equality between the sexes with special emphasis on educational reform and equal access to employment. Taylor's greatest claim to lasting importance stems from her role in the writing of *On Liberty* (1859), which Mill later called 'more directly and literally our joint production than anything else which bears my name' (Mill 1873: 249).

Unfortunately, Mill's sober record of Taylor's contributions has been rendered suspect by the hallucinatory lavishness of his praise of her powers. The questions surrounding their collaboration, however, do not concern Taylor's personal attributes, but the extent and nature of her contribution to Mill's work. Her own writing does not settle those questions definitively.

Most of the relevant letters, essays and notes in Taylor's solitary corpus are allusive supplements to ongoing dialogue with Mill. Some of this writing supports Mill's account of her work. Taylor's paragraphs on marriage (*c*.1832, Hayek 1951) are echoed in 'The Subjection of Women' (1869) and the distinctive themes in *On Liberty* first appear in her essay on toleration (*c*.1832, Hayek 1951). There she treats social conformity as a great evil which breeds injustice, stifles individuality and discourages character development. She charges that the indolent acceptance of popular opinion leaves people with baseless conviction and neither the patience nor skills needed to acquire genuine knowledge. She complains that independent thinkers must contend with ignorant majorities. Her arguments are weak, but her whole mode of thought on these topics informs in *On Liberty*.

Mill's corpus sometimes seems to split irreconcilably in two, and Taylor's influence may have caused the apparent fissure. Perhaps the author of in *On Liberty*, 'The Subjection of Women' and various writings on art and society was the offspring of their intimacy, conceived and nurtured in their conversations.

See also: Feminist political philosophy §2

CANDACE VOGLER

List of works

Hayek, F.A. (ed.) (1951) *John Stuart Mill and Harriet Taylor: Their Friendship and Subsequent Marriage*, Chicago, IL: University of Chicago Press.(The standard collection of their correspondence, which includes Harriet's poems and early essays on marriage and toleration, sometimes catalogued with 'J.S. Mill' listed as author.)

Rossi, A. (ed.) (1970) *Essays on Sex Equality*, Chicago, IL: University of Chicago Press.(Contains Taylor's essays on marriage, 'The Enfranchisement of Women' (1851) and 'The Subjection of Women' (1869) with an excellent critical preface by Alice Rossi.)

References and further reading

Bain, A. (1882) *John Stuart Mill: A Criticism with Personal Recollections*, London: Longman, Green.(In the final

chapter, ‘Character and Influence’, Bain takes up the question of Mill’s relationship with Harriet, notes ‘Mill’s hallucination about his wife’s genius’, but supports Mill’s detailed, precise account of Harriet’s contributions.)

Himmelfarb, G. (1974) *On Liberty and Liberalism: The Case of John Stuart Mill*, New York: Alfred A. Knopf. (A lively, detailed examination of Mill’s conflicting accounts of liberty, hostile to Harriet Taylor and supportive of Mill’s claim that *On Liberty* was a joint production.)

Mill, J.S. (1848) *Principles of Political Economy*, in *Collected Works of John Stuart Mill*, Toronto, Ont.: University of Toronto Press, vols 2-3.(Mill’s work on empirical and theoretical economics as a social science.)

Mill, J.S. (1859) *On Liberty*, in *Collected Works of John Stuart Mill*, Toronto, Ont.: University of Toronto Press, vol. 18: 213-310.(Mill’s powerful defence of extensive civil liberties, probably co-authored by Harriet Taylor.)

Mill, J.S. (1873) *Autobiography*, in *Collected Works of John Stuart Mill*, Toronto, Ont.: University of Toronto Press, vol. 1: 1-290.(Mill’s oddly impersonal account of his own life, containing the sober record of Taylor’s contribution to his work; the same volume of the *Collected Works* contains literary essays which betray her influence.)

Mill, J.S. (1874) *Three Essays on Religion*, in *Collected Works of John Stuart Mill*, Toronto, Ont.: University of Toronto Press, vol. 10: 369-489.(Essays on religion and nature for which Taylor set topics and about which they corresponded.)

Packe, M. (1954) *The Life of John Stuart Mill*, New York: Macmillan.(An excellent biography with a superb account of Harriet Taylor’s relationship with Mill.)

Technology and ethics

Only within the modern period have philosophers made a direct and sustained study of ethics and technology. Their work follows two philosophical traditions, each marked by distinct styles: the Continental or phenomenological tradition, and the Anglo-American or analytical tradition.

Hans Jonas (1979) articulated one of the basic premises of Continental approaches when he argued for technology as a special subject of ethics: because technology has fundamentally transformed the human condition, generating problems of global magnitude extending into the indefinite future, it calls for a new approach to ethics. Jonas' basic premise is expressed variously in the works of Karl Marx, Max Scheler, José Ortega y Gasset, Martin Heidegger and others.

Work within the Anglo-American tradition tends not to deal with technology as a whole but to be organized around particular technologies, such as computing, engineering, and medical and biological sciences. It draws on concepts and principles of traditional ethical theory at least as a starting point for analyses. Although each of the technologies has a unique set of problems, certain themes, such as responsibility, risk, equity and autonomy, are common to almost all.

Social scientists have also raised important issues for the field of ethics and technology. Their work has yielded two dominant schools of thought: technological determinism and social constructivism.

1 Attitudes to technology

The history of philosophy contains two broad moral appraisals of technology. The first, implicitly present in most normative theories prior to the Renaissance, is that technological change is socially destabilizing and therefore should be delimited carefully. Without social stability, it was argued, people die even under conditions where nature is abundant, and technological change easily undermines such stability. The second, characteristic of modern ethics, is that technological change is inherently beneficial because it enhances human welfare and autonomy. Here the argument is that people suffer more from the elements than from other human beings, and that they should therefore work together to conquer nature through technological progress. These two broad ethical views further reflect opposed ideals of human life: on the one hand, an ideal of social community subordinate to nature; on the other, human autonomy and freedom from natural constraint (see Technology, philosophy of).

Only within the modern period have questions of ethics and technology been examined in detail, with two approaches emerging. One grows out of the Continental or phenomenological tradition in philosophy, the other out of the Anglo-American analytical tradition (see Analytical philosophy; Phenomenological movement). Continental philosophers seek to evaluate technology as a whole while Anglo-American philosophers are oriented towards piecemeal assessments of particular technologies. Continental philosophers also commonly argue that traditional ethical theories are inadequate to the moral issues presented by modern technology, whereas Anglo-American philosophers have been more comfortable adapting existing utilitarian or deontological ethical frameworks (see Deontological ethics; Utilitarianism).

2 Continental discussions

Within Continental discussions, Hans Jonas has made the most sustained argument for technology as a special subject of ethics. For Jonas, ethics in premodern times could reasonably allow technology to remain in the background because technology itself had no high moral purpose. Unlike politics or religion, for instance, technology was treated as a marginal aspect of human life, one limited in both power and effect. By contrast, during the modern period technology entered the foreground of human experience. According to Francis Bacon (1620), for instance, the inventions of printing, gunpowder and the compass have done more to benefit humanity than all politics and religion. Jonas believes that because of the Baconian evaluation of technology, 'modern technology has introduced actions of such novel scale, objects, and consequences that the framework of former ethics can no longer contain them' ([1979] 1984: 6). According to Jonas, 'no previous ethics had to consider the global condition of human life and the far-off future, even existence of the race. These now being an issue demands… a new conception of duties and rights, for which previous ethics… provided not even the principles' ([1979] 1984: 8).

The exact way in which technology engenders an epochal transformation of the human condition is, however, open to dispute within Continental discussions. Philosophers as diverse as Karl Marx, Max Scheler, José Ortega y Gasset, and Martin Heidegger all claim that modern technology has transformed the human condition and undermined traditional moralities. But about the particular ways in which technology has altered the human lifeworld there are major disagreements. With regard to appropriate moral responses there are still further disputes.

According to Marx (1867), for instance, 'the modern science of technology' undermines traditional skills and the satisfactions of craft production, placing workers under the control of large-scale, capitalist-owned factories in which labour functions have become equal and interchangeable. The new system of production destroys a traditional social ecology in which the 'species essence' of making things benefited all social classes. Under capitalism material production unequally benefits the upper classes. This disequilibrium can be corrected only by means of a social revolution in ownership of the new technologies. The Marxist ethical assessment of industrial technology thus points up the inadequacy of the modern economic order as a means for the social control of technology, because of the way participation in that order has been restructured by technological change (see Marx, K. §4).

According to Scheler (1915), the technological transformation of the lifeworld is more than an economic phenomenon; it is also the rise and dominance of a new 'ethos of industrialism' even among technical workers themselves. Such an ethos exalts utility and instrumental values such as efficiency over vital and organic ones such as love. This is an axiological disorder that calls for cultural reformation.

For Ortega (1939), however, inherent within the modern science of technology and the ethos of industrialization there arises a moral problem that cannot be addressed by means of either Marx's social revolution or Scheler's cultural reform. Modern scientific technology, in contrast with traditional crafts, radically increases what can be done without any corresponding enhancement of ideals about what should be done. In Ortega's own formulation of the issue: in the pre-modern period, human beings acquired only very specific technical abilities, tightly coupled to particular uses, such as pottery for making pots. They never possessed any generalized technological powers. Now, however, human beings do possess technology in general without any clear idea of particular uses. To address this problem Ortega suggested a need to cultivate what he called 'techniques of the soul' such as yoga.

Undoubtedly the most influential European philosopher to address the issue of ethics and technology, even though he explicitly rejects the discipline of ethics as such, is Heidegger. Heidegger undercuts the distinction between science and technology, and argues that modern scientific technology or technoscience is not so much an ethos as a form of truth. This truth or knowledge reduces the world to *Bestand* or resources available for manipulation by a world-configuring, nihilistic destiny he calls *Gestell*. Heidegger seems at once to make ethical reflection more necessary than ever before and to destroy its very possibility (see Heidegger, M. §6).

One of Heidegger's less controversial theses is the notion that science and technology are interpenetrating practices: the science of nuclear physics is as much the applied technology of cyclotrons and reactors as the technology of nuclear engineering is applied nuclear physics. To the extent that this is the case, the ethics of science tends to merge with the ethics of technology (see Heideggerian philosophy of science).

3 Anglo-American discussions

In contrast to the Continental approach, philosophers following an Anglo-American analytical tradition organize ethical discussions around particular technologies. For example, biomedical ethics includes the study of ethical implications of the use and development of advanced medical technologies (see Bioethics; Medical ethics); information technology ethics (also known as computer ethics) examines social and ethical ramifications of computers and high-speed digital networks (see Information technology and ethics); engineering ethics studies the professional responsibilities of engineers (see Engineering and ethics); and environmental ethics evaluates the effects of various technologies on the natural environment (see Environmental ethics). Work in these areas applies concepts and modes of analysis drawn from analytical moral philosophy and political theory even though, at times, the special problems of new technologies demand an extension of these concepts and methodology beyond their traditional usage.

Each sub-area of ethics and technology is marked by a distinctive community of discourse and unique set of issues. Nevertheless, a number of common themes cut across these specialized domains. Prominent among them are: equity or justice; the problem of risk; responsibility for technology; and the effect of technology on liberty and autonomy.

Historically, issues of social justice, related to the distribution of technological goods and services, were first to become a focus of moral concern. This set of issues played a major role in the nineteenth-century rise of the labour movement, socialism, and the state regulation of technology. More recently, equity issues have re-emerged in acute form in relation to biomedicine, environmental pollution and computers. On what basis should scarce medical resources such as donated or artificial organs be allocated? To what extent does protection of an environmental commons legitimately limit private ownership? How should access to information technologies be facilitated under democratic capitalist structures?

A different dimension of the problem of just distribution of technological benefits is the problem of the equitable distribution of technological costs and risks. The general problem of risk due to technology, however, goes beyond equity concerns, and is part of the more general effort of technology assessment. Technologies often have unintended consequences which, if known in advance, might have altered decisions about their adoption. These consequences may include an increase in risk. Two scholars who have especially influenced work in this area are David Collingridge (1980) and Kristin Shrader-Frechette (1991). Collingridge is well-known for his dilemma of technological (and risk) assessment: in the early development of a technology, when it is relatively easy to control its direction, we inevitably lack the knowledge to exercise reasonable control; yet by the time we have more experience and, along with it, a better understanding of the risks, control has become difficult, if not impossible. Shrader-Frechette argues that persons should not be subjected to technological risk until they have clearly understood the risk and have granted their consent without being unduly constrained by economic or other external pressures. She contends that the concept of free and informed consent as applied in the field of medicine is applicable to technology in general and ought to be a part of what guides morally grounded public policy (see Risk; Risk assessment).

The issue of responsibility for the effects of technology involves two dominant lines of inquiry. One is concerned with the special responsibilities of technical professionals (see Professional ethics). Paul Durbin (1992), for instance, argues that engineers have an obligation to go outside their technical communities to lobby public policy, as when physicists during the 1950s and early 1960s lobbied for a worldwide ban on the atmospheric testing of nuclear weapons or when computer scientists in the United States opposed funding of the Strategic Defense Initiative (the ‘Star Wars’ project) in the 1980s. The second line of inquiry focuses on questions of blameworthiness and accountability when technological innovations or products cause individual or societal harms (see Responsibility; Responsibilities of scientists and intellectuals §3).

Finally, invention and utilization of advanced technologies may be evaluated in terms of liberty and autonomy. In biomedicine, for example, in the name of liberty and autonomy, there have been numerous efforts to work out the exact parameters of free and informed consent, and then to propose ways to institutionalize them (see Consent). In the area of information technology, liberty and autonomy are a key to explaining normative dimensions of privacy as well as prescribing the extent of freedom of speech over the digital networks (see Privacy §2). Langdon Winner (1986) has discussed, in more general terms, the relationship of technology with control and autonomy by looking at ways in which the technological design process and large technical systems, such as roads, restrict human agency, affect political and personal autonomy and delimit the exercise of democratic citizenship.

4 Other approaches to ethics and technology

Political, economic, and social science studies of technology also often have ethical dimensions. For example, political science studies of the governmental regulation of technology and of law-technology relations have obvious implications for ethics. Not only public policy efforts at technology assessment, but attempts to develop scientific techniques of policy analysis ultimately reflect and influence ethical decision-making. The same is true for efforts to manage more effectively social investments in innovation at both the state and corporate levels, and even for welfare economics.

The ethical implications of social science approaches to technology centre around the issue of technological

determinism, that is, the idea that technology is the primary determinant of social life, developing in an almost independent or autonomous manner. During the 1950s and 1960s, critics often opposed popular optimism about the unqualified benefits of technological progress with theories of technological determinism. Ironically, these theories almost immediately galvanized moral protest against technology. This protest was subsequently provided with theoretical justification by new studies that developed a social constructivist understanding of technological change, in which numerous agents and influences could be seen to exercise a series of micro-influences often invisible at the macro level (see Technology, philosophy of §4). Determinist and social constructivist interpretations, although apparently opposed, have quite similar moral implications. Both effectively highlight the continuing need for ethical examination of technology in a highly technological world.

CARL MITCHAM
HELEN NISSENBAUM

References and further reading

Bacon, F. (1620) *Novum organum (The New Method)*, trans. P. Urbach and J. Gibson, La Salle, IL: Open Court, 1994, esp. book I, ch. 129.(Argues that the goal of science is not simply knowledge but technological power.)

Baier, K. and Rescher, N. (eds) (1969) *Values and the Future: The Impact of Technological Change on American Values*, New York: Free Press.(An empirical and analytic philosophical study.)

Bijker, W.E., Hughes, T.P. and Pinch, T.J. (eds) (1987) *The Social Construction of Technological Systems*, Cambridge, MA: MIT Press.(A standard presentation of the social constructivist theory of technology.)

Borgmann, A. (1984) *Technology and the Character of Contemporary Life: A Philosophical Inquiry*, Chicago, IL: University of Chicago Press.(A comprehensive ethics and politics of technology.)

Collingridge, D. (1980) *The Social Control of Technology*, New York: St Martin’s Press.(Outlines the dilemma in assessing technology: by the time we know enough about a technology to want to control it, the technology has often developed so much social momentum that it is almost impossible to do so.)

Dahl, R. (1985) *Controlling Nuclear Weapons: Democracy Versus Guardianship*, Syracuse, NY: Syracuse University Press.(This contrast between two political approaches to the control of a particular technology has general implications.)

Durbin, P.T. (1992) *Social Responsibility in Science, Technology, and Medicine*, Bethlehem, PA: Lehigh University Press.(On obligations of engineers and others.)

Ellul, J. (1954) *La Technique ou L’Enjeu du Siècle*, trans. J. Wilkinson, *The Technological Society*, New York: Alfred A. Knopf, 1964.(The classic presentation of technological determinism.)

Feenberg, A. (1991) *Critical Theory of Technology*, New York: Oxford University Press.(Extends Frankfurt School Neo-Marxist critical theory to include issues of the social critical response to technology.)

Feenberg, A. (1995) *Alternative Modernity: The Technical Turn in Philosophy and Social Theory*, Berkeley, CA: University of California Press.(Because technology is a kind of legislation it calls for democratic control, and there are alternatives to technological modernism and postmodern abdication to responsibility. This book gives concrete examples of alternative socio-cultural forms of technology.)

Foucault, M. (1988) *Technologies of the Self*, ed. L.H. Martin, H. Gutman and P.H. Hutton, London: Tavistock. (Complements Ortega’s reference to "techniques of the soul".)

Goodpaster, K. and Sayre, K. (eds) (1979) *Ethics and the Problems of the 21st Century*, South Bend, IN: University of Notre Dame Press.(Represents the Anglo-American analytic approach.)

Guardini, R. (1926) *Letters from Lake Como: Explorations in Technology and the Human Race*, trans. G.W. Bromiley, Grand Rapids, MI: Eerdmans, 1994.(A lucid presentation of one Continental attitude.)

Habermas, J. (1970) *Toward a Rational Society: Student Protest, Science, and Politics*, trans. J.J. Shapiro, Boston, MA: Beacon Press.(Translated collection of essays from German texts published in 1968 and 1969. Provides a qualified defence of modern optimism.)

Heidegger, M. (1954) ‘Die Frage nach der Technik’, in *Vorträge und Aufsätze*, Pfullingen: Neske; trans. W. Lovitt, ‘The Question Concerning Technology’, in *The Question Concerning Technology and Other Essays*, New York: Harper & Row, 1977.(Argues that technology is not just a neutral or instrumental means but a way of revealing the world that influences the whole life of all who are involved with it.)

Hottois, G. (1984) *Pour une ethique dan un univers technicien*, Brussels: Éditions de l’Université de Bruxelles. (Moves the Continental approach towards specific problems.)

Illich, I. (1973) *Tools for Conviviality*, New York: Pantheon.(Representative of the moral protest against determinism.)

Jasanoff, S. (1990) *The Fifth Branch: Science Advisers as Policymakers*, Cambridge, MA: Harvard University

Press.(Examines how science and technology are regulated in an advanced industrial state.)
Jaspers, K. (1931) *Die geistige Situation der Zeit*, trans. E. Paul and C. Paul, *Man in the Modern Age*, London: Routledge, 1933.(Representative of the classical Continental attitude.)
Jonas, H. (1979) *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*, trans. H. Jonas and D. Herr, Chicago, IL: University of Chicago Press, 1984.(Enlargements of human power through technology carry with them expansions of human moral responsibility.)
Marcuse, H. (1964) *One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society*, Boston, MA: Beacon Press.(Important Marxist presentation of technological determinism and moral protest.)
Marx, K. (1867) *Das Kapital: Kritik der Politischen Ökonomie*, trans. D. Fernbach, *Capital: A Critique of Political Economy*, New York: Penguin, 3 vols, 1992, book I, ch. 13.(Critique of technology under capitalism. Includes an analysis of technology as a new force that, along with capitalist ownership, has transformed economic life.)
Mitcham, C. (1989) *Ethics and Technology: Research in Philosophy and Technology*, vol. 9, Greenwich, CT: JAI Press.(A collection of original papers representative of both Continental and Anglo-American discussions.)
Mitcham, C. (ed.) (1994) 'Ethics in Science, Technology, and Medicine', in C. Mitcham and W.F. Williams (eds) *The Reader's Adviser*, vol. 5, *The Best in Science, Technology, and Medicine*, New York: Bowker, 105-46.(An extended annotated bibliography.)
Mumford, L. (1967, 1970) *The Myth of the Machine*, vol. 1, *Technics and Human Development*, vol. 2, *The Pentagon of Power*, New York: Harcourt Brace.(Argues that the long historical development of power-centred modern technology has had a largely detrimental impact of human ways of life.)
Nelkin, D. (ed.) (1984) *Controversy: Politics of Technical Decisions*, Newbury Park, CA: Sage; 3rd edn, 1992. (Empirical case studies.)
Ortega y Gasset, José (1939) *Meditación de la técnica*, Buenos Aires: Escalpe; repr. in *Obras Completas*, vol. 5, Madrid: Revista de Occidente, 2nd edn, 1950/1; trans. H. Weyl and E. Williams, 'Thoughts on Technology', in C. Mitcham and R. Mackey (eds) *Philosophy and Technology: Readings in the Philosophical Problems of Technology*, New York: Free Press, 1972.(On modernity and technology. Argues that human beings are essentially technological and that history is transformed by changes in technology.)
Ropohl, G. (ed.) (1987) *Technik und Ethik*, Stuttgart: Reclam.(An introduction to engineering ethics in Germany.)
Rosenberg, N. (1982) *Inside the Black Box: Technology and Economics*, New York: Cambridge University Press. (On the interactions between technology and the eceonomy.)
Russell, B.H. and Pelto, P. (eds) (1972) *Technology and Social Change*, New York: Macmillan; 2nd edn, Prospect Heights, IL: Waveland Press, 1987.(A collection of anthropological studies.)
Sclove, R. (1995) *Democracy and Technology*, New York: Guilford Press.(Develops nine criteria for technology that would enhance democratic control. A positive, solutions-oriented book.)
Scheler, M. (1915) *Das Ressentiment im Aufbau der Moralen*, trans. W.W. Holdheim, *Ressentiment*, Milwaukee, WI: Marquette University Press, 1994, ch. 5.(Discusses 'ethos of industrialism', suggesting that the modern world is characterized by an egalitarian resentment of aristocratic achievement that also supports technology.)
Sen, A. (1987) *On Ethics and Economics*, Oxford: Blackwell.(Critical of a welfare economic analyses of technological development that ignore equity issues.)
Shrader-Frechette, K.S. (1991) *Risk and Rationality: Philosophical Foundations for Populist Reforms*, Berkeley, CA: University of California Press.(The subjection of human beings to technological risks is morally legitimate only to the extent people have given their free and informed consent.)
Tiles, M. and Oberdiek, H. (1995) *Living in a Technological Culture: Human Tools and Human Values*, London: Routledge.(Broad overview from an Anglo-American perspective.)
Unger, S.H. (1982) *Controlling Technology: Ethics and the Responsible Engineer*, New York: Addison Wesley; 2nd edn, 1994.(A socially critical engineer's view of responsibility in the technical professions.)
Westra, L. and Shader-Frechette, K. (eds) (1996) *Ethics and Technology*, LaJolla, CA: Jones and Bartlett. (Useful general collection.)
Winner, L. (1986) *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, Chicago, IL: University of Chicago Press.(Technologies are like political constitutions in that they set up ways of life, and as such ought to be created on a democratic basis.)

Technology, philosophy of

The philosophy of technology deals with the nature of technology and its effects on human life and society. The increasing influence of modern technology on human existence has triggered a growing interest in a philosophical analysis of technology. Nevertheless, the philosophy of technology as a coherent field of research does not yet exist. The subject covers studies from almost every branch of thinking in philosophy and deals with a great variety of topics because of a lack of consensus about the primary meaning of the term 'technology', which may, among others, refer to a collection of artifacts, a form of human action, a form of knowledge or a social process.

Among the most fundamental issues are two demarcation problems directly related to the definition of technology. The first concerns the distinction between technological (artificial) and natural objects. It involves the relation between man, nature and culture. The second pertains to the distinction between science and technology as types of knowledge. The science-technology relationship has become of central importance because of the widespread assumption that the distinguishing feature of modern technology, as compared to traditional forms of technology, is that it is science-based. Another much discussed issue is the autonomy of technology. It deals with the question of whether technology follows its own inevitable course of development, irrespective of its social, political, economic and cultural context.

1 Philosophy and technology

The following enumeration of approaches is intended to convey an impression of the strongly fragmented field of study that goes by the name of philosophy of technology.

What is (the essence of) technology? In philosophical-anthropological studies, the starting point for answering this question is the human being and its place in and relation to nature. The human being is considered to be a defective animal that is dependent on technology for its survival; technology becomes the substitute for biological shortcomings and is therefore determined to a large degree by the nature of these shortcomings. For Heidegger (1977) such a characterization of the essence of technology is not sufficient; an answer at the metaphysical level is also needed. For him the essence of technology is not that it is a means to some end: technology brings to the fore what was hidden and does not by itself present itself (see Heidegger, M. §6; Heideggerian philosophy of science). In Dessauer's (1927) metaphysical approach, invention is the essence of technology and the ontological conditions that make invention possible are explored.

The social philosophy of technology focuses on the relations between (specific forms of) technology and social, economic and political structures. It analyses technological development as a social process and addresses the problem of how to control its development. One of the key problems in this field is whether technological development is primarily determined by its context (social shaping of technology), or whether technology determines the social context including its systems of norms and values (a position often attributed to Marx). In the latter case, the idea of a technocratic society emerges in which technological rationality imposes itself on all domains of social life (technology as an ideology (Habermas 1968); see Habermas, J. §1).

Ethical studies take a prominent place in the philosophy of technology. New technological possibilities for human intervention create new moral problems. Do these require new ethical principles? Arguments in favour are based mostly on the idea that modern, science-based technology is essentially different from earlier forms of technology (the crafts), and that its impact on man and nature is of a different order (for example, the consequences of applying modern technology are no longer limited in space and time (Jonas 1984)). Another issue in this field concerns the claim that technology itself, as a system of means, is ethically neutral. Arguments against the neutrality thesis attempt to show that the conception of technology as a mere system of means is inadequate, because its impact on human life stretches much further: it replaces the natural with an artificial environment (see Technology and ethics).

The so-called 'analytical philosophy of technology' (Rapp 1974) shows a strong focus on epistemological and methodological problems of technology, particularly of the engineering sciences. These problems have long been neglected, because of the widespread idea that technology is applied science. A specific feature of this approach is that it takes into account the various types and stages of technology and its concrete historical form of appearance.

Key topics are the nature of technological knowledge, the nature of engineering design, design methodology and the relation between science and technology. Its main disciplinary sources are, besides philosophy, technology itself and the history of technology.

2 Technology and artifacts

The usual conception of technology is that it is the transformation or manipulation of nature (the existing physical (material) and biological environments) to satisfy human needs and goals. Technology is thus conceived to be a specific form of purposeful (teleological) action, that may result in a 'technological artefact': a human-made object or state of affairs that fulfils a utilitarian or practical function. The transformation of nature may or may not itself be mediated by artifacts, which are then called tools (see Teleology).

This conception of technology raises many questions. On the one hand, it appears too restrictive, for it does not fit certain domains which are considered to belong to modern technology, such as software engineering which deals with the transformation of something immaterial (information). On the other hand, it is too broad, since it makes any object or state of affairs which satisfies a practical need, and is the result of intentional human intervention in nature, a technological artefact (for example, a wild tree planted deliberately at a certain place to provide shadow, or an organism with a slightly modified genetic structure).

The demarcation problem - what kind of action constitutes a technological action and what kind of objects or states of affairs are technological artifacts? - remains an open issue. Not only is the distinction between the technological and the artificial problematic, but also that between the artificial and the natural. The latter raises fundamental philosophical issues about the relation between the human race and nature. The distinction makes sense only if the human race is considered in some respect not to be part of nature. As an integral part of nature (and as a result of natural evolution), a human being cannot interfere with nature. The distinction between the natural and the artificial is commonly taken to be identical to the distinction between the spontaneous and the intentional; these notions themselves, however, raise all kinds of philosophical problems (see Environmental ethics §4).

Similar questions arise when technological artifacts are characterized as objects that perform practical functions on the basis of *human designs*. In technology a design is taken to be a pattern or scheme that describes the structure and mode of operation of a system and shows how a given practical aim or function may be realized. The notion of design stresses the inherently intentional/teleological nature of technological artifacts. The distinctive feature of technological artifacts as compared to objects from nature remains problematic, however. Is the difference primarily a genetic one (produced by humans or by nature), or is it more fundamental in the sense that the attribution of a design to objects of nature is meaningless (as in the modern scientific conception of nature)? These questions inevitably lead to issues in the philosophy of nature.

3 Technology as knowledge

For Aristotle science and technology belonged to two different spheres of human experience (contemplation versus productive action) and constituted two different forms of knowledge (theoretical versus practical knowledge). Scientific knowledge was, moreover, not inherently relevant for solving technological problems (see Aristotle §6). Modern technology and science, however, have merged to such a degree that even the demarcation between them has become problematic. Modern technology is science-based (and modern science, technology-based) and alongside the traditional natural sciences engineering sciences have established themselves. The so-called 'scientification of technology' is generally considered to be the characteristic feature of modern technology that is directly related to its prominent role in society. This has directed attention to the problem of the relation between science and technology and how science has altered the nature of technology. From a cognitive point of view, the nature of technological knowledge and its relation to scientific knowledge is at issue.

One of the most influential models for the science-technology relationship has been the technology-is-applied-science model. It stresses that technology, in contrast to the traditional crafts, is as theory-laden as science and that it applies scientific theories to systems which are of practical use. It considers technological knowledge to be a derivative kind of scientific knowledge. This model has been criticized severely. It is historically inadequate because it makes technological progress wholly dependent on scientific development. From a cognitive point of view it is highly problematic because it assumes that there is a logically deductive path

from scientific knowledge (theories) to technological designs.

A generally accepted alternative for the technology-is-applied-science model is, however, still missing. A strong case could be made for considering technological knowledge, with respect to scientific knowledge, to be a form of knowledge *sui generis*, which deals with the design and production of artifacts. One of the considerations in favour of such a view concerns the fact that the criteria for evaluating cognitive claims are fundamentally different in the two domains.

In its most basic form technological knowledge is prescriptive; it consists of procedures (rules) which describe actions that have to be performed in order to achieve practical ends. Often, an adequate performance of these actions requires practical skills; these also constitute a form of technological knowledge, which, however, cannot be expressed in terms of prescriptions in language. The first criterion for evaluating a procedure that purports to solve a technological problem is its *effectiveness*: does the procedure bring about the desired state of affairs? The second criterion for assessing a solution is its degree of *efficiency*: is it possible to bring about the desired state of affairs in a better way, that is, with less effort or costs. (In general, measures of efficiency are based on quantitative comparisons of input and output of technological systems.) In modern technology, the notion of efficiency plays a dominant role, and the evaluation of the efficiency of artifacts is often carried out with theoretical means; at this point scientific theories about the operative principles of an artefact turn out to be of great value.

While efficiency, effectiveness, and other criteria like durability, costs, manufacturability, safety and utility, are key notions in the structure of thinking in technology, they play almost no role in science, because scientific knowledge is primarily descriptive and explanatory. Among its principal evaluative criteria we find truth, empirical adequacy and explanatory power. This difference in basic concepts for interpreting and evaluating knowledge claims strongly supports the idea that two different forms of knowledge (and of rationality) are involved in science and technology. Further insight into the characteristic features of technological knowledge, that is, in the epistemology of technology, has to come from detailed case studies (Vincenti 1990).

4 The dynamics of technological change

A recurrent theme in the philosophy of technology is the autonomy of technological development relative to its social embedding. It concerns the dynamics of technological change. This is an extremely complex topic since technological innovation, in its various stages from design, development, production and diffusion, is influenced by very heterogeneous factors (cognitive, economic, social, political, military, geographic, cultural, and so on). Views on the dynamics of technological change tend to show strong reductionistic tenets by taking one of these factors as the prime mover of technological change. Well-known illustrations are ‘technological determinism’ and ‘social constructivist interpretations of technology’. According to technological determinism technology itself is the prime mover; this view maintains that technology follows its own intrinsic course of evolution to which society has to adapt; it is self-determinative with increase of efficiency as one of its main dynamical principles. So there is no room for alternative forms of technology. Social constructivist interpretations of technology, on the other hand, claim that technology is to a large degree, or even completely, socially determined (see Constructivism).

In order to uncover some of the basic assumptions underlying these ideas we will briefly examine the shaping of artifacts in the design phase. The final form of an artifact is schematically determined by two kinds of constraint. On the one hand, there is the list of specifications which describes all kinds of requirements the artefact should fulfil. This heterogeneous list may contain, among others, constraints derived from the primary technological function of the artefact, the conditions under which this function has to be performed in practice (for instance, safety regulations), the conditions under which the artifact has to be produced (for example, mass production), its price, standards and norms. The defining feature of this list is that it is the outcome of a process of negotiation between all parties with some interest in the artefact; it contains social or contextual constraints, which are imposed by convention. On the other hand, there is a list of technological constraints, constraints due to what is, as a matter of fact, physically and technically possible (this list may, of course, change over time). The desired artefact should satisfy both lists of constraints, but because of conflicts this is often not possible in practice. Such conflicts can be resolved, again schematically, by adapting the list of specifications or by creating new technological possibilities.

According to this conception of the design process, the driving force behind technological change is the tension

between two lists of constraints which are different in nature (one describing what is desirable, the other what is possible) and in principle independent of each other. Defenders of the autonomy of technology reject this independence; they assume that the list of specifications is determined primarily by technological constraints ('what can be done, will be done'). The decisions leading up to the list of specifications are supposed to be dictated by technological imperatives (technological rationality). Social constructivists, on their part, tend to deny any difference in nature between the two types of constraint: they assume that physical and technological constraints are socially constructed in ways similar to the constraints contained in the list of specifications.

These assumptions may be criticized on philosophical grounds, but much more important for the assessment of these views is that the issue of the autonomy of technology is not to be settled by philosophical arguments alone. Models for the dynamics of technological change (with their underlying philosophical assumptions) must prove their value in empirical research, particularly in the history of technology. Generally speaking, a 'historical-empirical turn' in the philosophy of technology might prove to be fruitful for developing more adequate conceptual schemes for interpreting technology and its development.

See also: Functional explanation; Gender and science §5; Marxist philosophy of science; Risk assessment

PETER KROES

References and further reading

The following list contains works on philosophy of technology at a general level and at a more advanced level.

Bijker, W.E., Hughes, T.P. and Pinch, T.J. (eds) (1989) *The Social Construction of Technological Systems*, Cambridge, MA: MIT Press.(Contains contributions which interpret technology from a social constructivist point of view. See §4.)

Dessauer, F. (1927) *Philosophie der Technik (Philosophy of Technology)*, Bonn: Cohen.(Metaphysical analysis of invention as the core activity of technology. See §1.)

Habermas, J. (1968) *Technik und Wissenschaft als 'Ideologie' (Technology and Science as 'Ideology')*, Frankfurt: Suhrkamp.(Analysis of the ideological character of technology and science. See §1.)

Heidegger, M. (1977) *The Question Concerning Technology and other Essays*, trans. W. Lovitt, New York: Harper & Row.(Generally considered to be the most outstanding metaphysical analysis of technology. See §1.)

Jonas, H. (1984) *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*, Chicago, IL: University of Chicago Press.(Argues that modern technology requires a new ethics based on the notion of responsibility. See §1.)

Philosophy and Technology (1983-93) Society for Philosophy of Technology, 10 vols.(Official publication of the society for philosophy of technology; vols 6 and 7 discuss various interpretations and approaches in the philosophy of technology.)

Rapp, F. (ed.) (1974) *Contributions to a Philosophy of Technology*, Dordrecht: Reidel.(This anthology marks the beginning of the analytical philosophy of technology and addresses mainly methodological and epistemological issues. See §1.)

Research in Philosophy and Technology (1978-93) Greenwich, CT: JAI Press, vols 1-13.(For a long time the only regular publication in the field; its content illustrates the diversity of approaches and of topics addressed in the philosophy of technology. Several volumes contain extensive bibliographies on various topics in the philosophy of technology.)

Staudenmaier, J.M. (1985) *Technology's Storytellers*, Cambridge, MA: MIT Press.(Contains an extensive discussion of attempts by historians and philosophers of science and technology to analyse the relationship between science and technology and of the debate over technological determinism. See §5.)

Vincenti, W.G. (1990) *What Engineers Know and How They Know It: Analytical Studies from Aeronautical History*, Baltimore, MD: Johns Hopkins University Press.(Contains detailed studies of the nature of technological knowledge and its use and generation by engineers. See §3.)

Teilhard de Chardin, Pierre (1881-1955)

Pierre Teilhard de Chardin taught that the evolutionary process is governed by a 'law of complexification' which dictates that inorganic matter will reach ever more complex forms, resulting in inorganic matter being followed by organic matter and organic matter being followed by conscious life forms. Viewed by observers, humans are material systems within a larger physical system. Viewed introspectively, a human being is a self-conscious creature possessed of freedom and rationality, with the capacity for action and inquiry. Each element in the world has some form of this dual 'exterior' aspect and 'interior' aspect, though consciousness arises only late in the evolutionary history. Teilhard de Chardin saw neither reason to doubt that matter can give rise to mind, nor any basis for reducing mind to matter. The prospects for humanity are gratifying, as evolution, following the law of complexification with the cooperation of human choice, moves to an Omega point at which Christ's fullness will include as his 'body' a unified humanity that is at peace.

Scientific critics of Teilhard de Chardin's theory have charged that his optimism involves extrapolation far beyond what the present evidence warrants. Theological critics have argued that he does not sufficiently consider the degree of evil in the world; optimism can only be justified if we assume that evil can be redeemed by transcendent divine action, because immanent evolutionary processes may not suffice.

1 Life

Pierre Teilhard de Chardin, the fourth of eleven children, received his secondary education at the Collège Notre-Dame at Mongre, and, at eighteen, entered the Society of Jesus. He read Bergson's writings, and an association with Eduard Le Roy, a Bergsonian scholar, deepened the Bergsonian influence on Teilhard de Chardin's thought. While his views often sparked keen dissent within the Church, he remained a Jesuit and a loyal Catholic. Repeated requests to the Church for permission to publish his best-known work, *Le Phénomène humaine* (1953), were refused, and it was published only posthumously. He sought a way to unite his 'love for God' and his 'love for the world'. The product was not merely a view which sanctions the dignity of scientific endeavour, but a complex worldview in which cosmic evolution is the process by which God brings into being a 'fullness of Christ' that includes a morally and spiritually mature humanity and a fully developed natural environment.

Teilhard de Chardin taught in the Jesuit College in Cairo (1906-8), served in the First World War, and taught at the Catholic Institute in Paris; when he was fired from this post, he travelled to China. He was a distinguished geologist and palaeontologist, especially concerned with human prehistory, and in China participated in the palaeontological researches surrounding the discovery of Peking Man. He was associated with many expeditions and became director of the research office of the Centre National de la Recherche Scientifique in Paris. After the Second World War, he was offered a chair at the Collège de France, but instead went to New York, where he was made permanent assistant at the Wenner-Gren Foundation for Anthropological Research. He was an Honorary Fellow of the Royal Anthropological Society of Great Britain, and an honorary member of the Academy of Sciences of New York, as well as a member of the French Academy of Sciences and the American Association for Geology.

2 Priest and scientist

It was Teilhard de Chardin's goal to find for himself, and then share with others, a worldview that unified his theological beliefs as a Catholic and professional beliefs as a scientist. He thus developed a synthesis of Catholic theology and evolutionary theory which he claimed was satisfactorily founded in evidence and reason. In it, he rightly remarked, the traditional juridical and political terms in which much of Catholic thought had come to be classically expressed were replaced by terms derived from the natural sciences, although he extended them beyond their native home in scientific theory. It would seem that he thought these terms received much less stretching at his hands than many of his readers, both scientists and theologians, suppose.

It seems clear as well that while Teilhard de Chardin had apologetic interests - he wanted to argue for his faith - and communication interests - he wanted his faith to be intelligible to modern readers - he had a more basic motivation. He thought that his newly crafted worldview was a natural and appropriate convergence of the picture of the world presented descriptively by evolutionary science, and the doctrines of an orthodox (if nontraditionally

stated) Catholicism; his main reason for presenting his worldview was that he thought it was true.

Teilhard de Chardin described his intellectual project as providing a phenomenology of the physical universe - a description of the universe that arises from and is justified by what the natural sciences have discovered the world to be. The resulting account of things has a surprising consequence. Included in his notion of a phenomenology is that it should say what the functions of the phenomena are and give an account of the purposiveness that phenomena actually exhibit. His account of the function and purpose of the evolutionary process is, of course, controversial. He contends, for example, that since the evolutionary process has plainly produced human beings, and we see no new species on the horizon to replace them, it is in humanity that we find the best clue to understanding the process. This in effect involves him in thinking in terms of final as well as efficient causes (to put the matter in traditional language), and in viewing this as part of what is properly included in a phenomenology that is scientifically based.

Teilhard de Chardin viewed the universe in both theistic and evolutionary terms; the world owes its existence and structure to God, and God is immanent in it. The universe is an immense organic whole which is driven by its inner energies toward greater and greater order and complexity. His notion of evolution is cosmic, not simply biological; within the framework of his evolutionary theory there appear not only mutating forms of life but also inorganic matter and, ultimately, social and cultural systems.

3 Optimism and evolution

Much has been written about the medieval view of an orderly universe, quite limited in scope, created for human habitation, and centred on our earth, whose plants and animals served human ends. Earth was a rather manageable habitation which served its resident landlords rather well as a temporary home until their real home was reached in the afterlife. Then astronomers revealed the vastness of the universe, physicists discovered that the same laws applied beyond the moon as applied below it and that the universe is a vast machine, biologists found a minuscule universe of a vastness comparable to the macrocosm, Darwin robbed humanity of its uniqueness, and Freud stole from humanity its fundamental rationality. The result is a universe in which human beings live a precarious unfree existence in a natural environment that is at best neutral, while beset with inner forces that threaten what sanity they possess as they wait for entropy to wind things down.

Without considering how accurate all this is, either as regards medieval or modern thought, perhaps the broad strokes of the preceding paragraph give some sense of what Teilhard de Chardin was fighting. He insisted that the very evolutionary theory that played a significant role in questioning the uniqueness and hence special worth of human beings, when properly understood and articulated, restored that very uniqueness and worth. He denied that the universe was machinelike, static, deterministic, neutral or unfriendly to human existence, or doomed to an inelegant ending. He held that humanity is an entirely natural denizen of the world without thinking that the existence of human beings is to be naturalistically explained - not, at any rate, if 'being naturalistically explained' denies either human rational and volitional transcendence of determining conditions or theological explanation of the evolutionary process.

Teilhard de Chardin suggested that the earth's history has three phases. First, there is the cooling of the earth's crust and the presence of inorganic matter. In the second stage, organic matter arises as life emerges and diversifies. The third stage begins with the appearance of mind. Geosphere is succeeded by biosphere, biosphere is followed by noosphere. While the first stage leads naturally to the second, and the second to the third, both later stages none the less contain something qualitatively new. In the third stage, there is a qualitative shift in the nature of the evolutionary process itself as the products of evolution become capable of contributing to, and potentially directing, the future of the process.

Humanity, Teilhard de Chardin held, is 'separated [from the other elements of the evolutionary process] by a chasm - or a threshold - which it cannot cross. Because we are reflective we are not only different but quite other. It is not a matter of change of degree, but a change of nature, resulting from a change of state' ([1953] 1959: 165-6). Thus the evolutionary process exhibits discontinuity within continuity in an ascent from less to more organized forms of matter. Material synthesis and complexity are viewed as one aspect of the same stuff that, in its highly complex versions, has intellectual and spiritual capacity as another aspect.

4 Beginning and goal

Teilhard de Chardin saw human beings as both the end or goal of the evolutionary process and as a new evolutionary beginning, not to be replaced by a new species. Evolution now and in future is a process under new rules; humans act freely in the context of natural laws, rather than one nonintelligent event following another. Without supposing that detailed projections are possible, he took the future to be the key to the past and believed that science justifies an optimistic forecast of the future of humanity. It is likely that this optimism, grounded at least in intent in a rationally developed worldview, is part of the explanation for the enthusiasm Teilhard de Chardin's views often elicit.

Currently, humanity experiences diverse cultures and forms of intellectual life; it is an often warring, significantly immature species. None the less, humanity's capacity for self-conscious thought and action provides a new layer or sphere on the surface of the earth - a 'noosphere' that is the unique environment of human beings, shared not even with the highest nonhuman animals. This sphere too is subject to the law of complexification, and thus there will be a cultural unification in which humans learn to live in love and peace. When this 'Omega point' is reached, a united humanity will serve as the body of Christ, and Christ will have achieved the fullness that it has been the point of evolution to produce.

It follows from Teilhard de Chardin's views that it is important to understand the world in terms of theistic cosmic evolution. While there are powerful tendencies to 'complexification' in nature, once human beings arrive on the scene there is in principle the possibility of their acting sufficiently unwisely or wickedly as to postpone, diminish or even prevent the arrival of the Omega point. After all, according to Teilhard de Chardin's view (though this is certainly not unique to his perspective), the course of evolution is now significantly in human hands; none the less, it is plainly his view that optimism is justified concerning the prospects of the Omega point actually being reached. The later chapters of the book of evolution seem sure to be happy ones.

For Teilhard de Chardin, Christ himself is primarily the Cosmic Goal of the evolutionary process ('Christ' here is intended to refer to both the Second Person of the Christian trinity and his 'body' of collective mature humanity). The redemption that Christ provides according to the Christian message is seen in cosmic as well as individual terms, and for Teilhard de Chardin, these cosmic terms are seen along the lines of a gradual and predictable change for the better, and are conceived more in terms of immanence than transcendence.

5 Criticisms

Teilhard de Chardin's scientific critics, even those who agree that the course of evolution is now partly under human management, find in rising population, environmental pollution and decreasing resources significant grounds for being dubious about evolutionary optimism. They worry that entropy in the long run will end human life and that increasing specialization will diminish the quality of life. Most fundamentally, they find in Teilhard de Chardin's perspective an extrapolation that goes far beyond anything they discover in actual science, and are unconvinced that anyone who follows his lead will thereby be walking in the overall direction (if any) suggested by scientific conclusions. Thus they see little or no justification in actual scientific theory for Teilhard de Chardin's Christian cosmic evolutionism.

Theological critics (for example, Smulders 1967 and Thielicke 1964) have worried that the 'in principle' possibility of human evil, manifested individually and institutionally, might not be too deep for real confidence in any immanent process leading to a cosmic kingdom of God. They view Teilhard de Chardin as taking altogether insufficient account of the depth of evil in the human heart and in the institutions that we have created. It seems to them all too likely that the outcome will be at least no better than what we currently experience unless there is special divine intervention of a sort not really in the spirit of his theory.

Teilhard de Chardin sees sin as an inevitable feature of the evolutionary process, which is to be overcome as the law of complexification works in concert with right human choices and actions. This seems to his critics to involve a sort of self-healing theory of the universe in which they have no great confidence. Much of the Christian tradition has insisted on a significantly greater role for transcendent divine action than seems evident in, or consistent with, Teilhard de Chardin's conception of things. For most of that tradition, the Fall is neither a fall 'up' towards maturity, nor an inevitable part of a process that leads upward. It has thought in terms of people, and often also of nature, as needing regeneration, not merely time to develop. Its optimism rests more on faith and hope in

the activities of a transcendent Saviour than on faith in the progressive unfolding of immanent principles built by the Creator into the creation. Thus Teilhard de Chardin's Christian critics stress that the problem that Christ came to deal with was our sin, which includes not only individual wrong actions knowingly performed, but a commitment to a style of life centred on our own apparent advantage in disregard to the will of God and the good of others. Christ is thus the Saviour who dies for our sins and rises again for our justification, the Judge who condemns both sin and the unrepentant sinner who clings to sinfulness and rejects forgiveness and grace, and the Lord of all Creation. The Christian critics argue that Teilhard de Chardin subordinates the roles of Saviour and Judge to that of Lord of Nature in a way that trims the sails of Christian doctrine to fit the masts of evolutionary optimism.

At times asserting that his views arose naturally and with considerable evidential support from the data of science, Teilhard de Chardin at other times emphasized the tentative and partial nature of his perspective. At the least, he coherently placed a massive amount of scientific data within a theological perspective that most of his colleagues had dismissed or ignored, and (perhaps not altogether intentionally) illustrated the fact that scientific data, as philosophers of science say, 'underdetermine' the world views based on them.

See also: Religion and science §§4-5

KEITH E. YANDELL

List of works

Teilhard de Chardin, P. (1955) *Le Phénomène humaine*, Paris: Éditions de Seuil; trans. B. Wall, *The Phenomenon of Man*, London: Collins, 1959.(This is Teilhard de Chardin's *magnum opus*, in which he endeavours to unify the data of science and the data of the Christian revelation.)

Teilhard de Chardin, P. (1956) *L'Apparition de l'homme*, Paris: Éditions de Seuil; trans. J.M. Cohen, *The Appearance of Man*, London: Collins, 1965.(Teilhard de Chardin here offers his views on the evolutionary rise of humanity.)

Teilhard de Chardin, P. (1959) *L'Avenir de l'homme*, Paris: Éditions de Seuil; trans. N. Denny, *The Future of Man*, London: Collins, 1964.(Teilhard de Chardin's vision of the future of humanity and the world.)

References and further reading

Cuenot, C. (1965) *Teilhard de Chardin*, London and Baltimore, MD: Burns, Oates & Helicon.(A substantial biography.)

Lubac, H. de (1962) *La Pensée religieuse du Père Teilhard de Chardin*, Paris: Éditions Montaigne; trans. R. Hague, *The Religion of Teilhard de Chardin*, London: Collins, 1967. (Includes an appendix of letters by Teilhard de Chardin.)

Smulders, P. (1967) *The Design of Teilhard de Chardin*, Westminster, MD: Newman Press.(Examines the theological implications of Teilhard de Chardin's views.)

Thielicke, H. (1964) *Mensch Sein - Mensch Werden*, Munich: R. Piper Verlag; *Being Human… Becoming Human*, Garden City, NY: Doubleday, 1976.(Contains a theological critique of Teilhard de Chardin, and reviews a scientific critique.)

Wilders, N.M. (1963) *Teilhard de Chardin: Eeninleiding in zijn denken*, Antwerp and Amsterdam: N.V. Standaard-Boekhandel Uitgeversmaatschappij; trans. H. Hoskins, *An Introduction to Teilhard de Chardin*, New York: Harper & Row, 1963.(An accessible, sympathetic discussion of Teilhard de Chardin's perspective.)

Wilders, N.M. (1966) *Teilhard de Chardin and the Mystery of Christ*, London: Collins. (Endeavours to exhibit Teilhard's thought as a single system.)

Tel Quel School

Tel Quel was a review published in Paris from 1960 to 1982. Under the direction of Philippe Sollers, it became a key source of avant-garde work in literature and critical theory. Concerned with the relations between art and politics, the Tel Quel *group drew on semiotics, psychoanalysis and Marxism as the bases for an overall theory that would establish writing -* écriture *- as having its own specific and necessary revolutionary force. Influential in its emphasis on literary practices seen as breaking with the given social ordering of 'reality' and 'subject' (the 'limit-texts' of writers such as Sade or Artaud), the review emphasized textuality, the condition of all fields of knowledge as textual productions. Less a coherent school of thought than a site of shifting theoretical-political interventions and new explorations in writing, Tel Quel was at its most powerful in the late 1960s and early 1970s.*

1 History

Founded in 1960 by a handful of young writers under the auspices of the French publishing house *Editions du Seuil*, *Tel Quel* was a review appearing quarterly until 1982. Purely literary concerns quickly gave way to attempts to define a specific role for literature as a factor in social change and to explore possible alliances with existing political formations. In the late 1960s, while elaborating a revolutionary cultural theory of writing and its practice, the review drew close to the French Communist Party (*PCF*), espousing many of its positions. In 1969 there was a shift to criticism of the Party's 'revisionism', with *Tel Quel* asserting a Marxism-Leninism of which Maoism and the Chinese Cultural Revolution were regarded as the authentic contemporary manifestation. The final break with the *PCF* came in 1971, with some members of the review's board forming the *Mouvement de Juin 1971* which briefly issued *Tel Quel Informations*, a bulletin devoted to violent criticism of the *PCF* and unbridled celebration of the People's Republic. A *Tel Quel* delegation visited China in 1974 but the visit coincided with, and contributed to, a disaffection with Communism that led in 1976 to dissociation from the Chinese model and the 'Marxist church'. Communism and the 'myths' of the left now became the object of the review's combat, with dissidence valued as the epitome of literature's purpose. *Tel Quel* ended in 1982, only to be resurrected as *L'Infini*.

The core members of the editorial board were Philippe Sollers, Marcelin Pleynet and Julia Kristeva. Jacques Derrida was a decisive contributor in the 1960s and Roland Barthes a major one throughout. A series of books - the '*Collection Tel Quel*' - was linked with the review, which also instigated various cultural events.

2 Theory

Tel Quel's initial orientation was determinedly literary, in opposition to Sartrean existentialism and the demand for *littérature engagée* (literature as an instrument of social commitment). Stating its aim to be a clear understanding of 'the powers of writing', the review began with a vaguely romantic declaration of literature's value and a determination to free it from moral and political imperatives: literature was to be taken on its own necessary terms as the mode of discovery of the world as sensory whole. Surrealism was acknowledged as a significant historical precedent but also as a movement in relation to which the review sought to make its own distinct advance.

Consideration of writing's powers and of the nature of its realization of the world 'as is' (the sense of the review's title) was undertaken both through an analysis of literary forms and through an engagement with bodies of writing that questioned given conventions. An immediate reference was to the critique of the traditional terms of the novel represented by the *nouveau roman* and notably to the work of Alain Robbe-Grillet, itself critical of Sartre's instrumentalization of literature and self-reflexively involved in the showing up of conventional fictional structures. The nouveau roman, however, was soon repudiated as complicit with old assumptions, as once more referring literature to the representation of a reality projected as external to its writing, whether physical ('realism') or mental ('psychologism'). Hence the contrary importance of 'limit-texts' that disturb the homogeneity of reality in its accepted forms, subverting representational coherence, subject unity and the notion of a truth outside linguistic constructions: Sade and Artaud, Lautréamont and Bataille were read as inscribing within the Western tradition a permanent refusal of the comforts of any such 'truth'. The force of literature was defined as exactly that refusal, its practice as a radical knowledge of the tissue of fictions in which we and our world are 'made up'.

To take literature in this way was to break with ideas of it as the expression of universal ('human') meanings and values, and to assign to it a fundamental role for change. Literature and politics were seen as interdependent, not in

terms of the former's subordination to the latter (submission to imperatives again), but in as much as literature - in *Tel Quel*'s avant-garde version of its excessive practice - was understood as a condition of political transformation: no economic and social revolution without symbolic revolution. Surrealism's weakness was identified precisely as its failure scientifically to make the link between revolution and writing - it lacked the textual theory that from the mid-1960s to the mid-1970s became *Tel Quel*'s purpose to provide.

The beginnings of the theory were in structuralism and semiotics: the analysis of the underlying structures of cultural phenomena grasped as systems of signs seemed to offer a possible science of literature rid of any idea of the latter as externally derived (works as originating in a reality they represent or an author they express). Confronted, however, with the experience of writing realized in the limit-texts, as in the *Tel Quel* group's own novels and poetry, structuralism appeared as based on a static conception of structure which, while effectively serving to remove the subject as consciousness from any central and founding position, ended in an ahistorical 'objectivity'. If everything is only signs, structuralism's metalinguistic certainty of structures to be perceived through the agency of its analyses is itself in question. The recognition semiotics should provoke is that there can be no mastery of language since language includes - is constitutive of - the subject assuming such mastery. A science of literature must work with just this recognition which is, indeed, literature's own troubling power and so its challenge to any science or knowledge classically conceived (in terms of a closure of textual production, of the elimination precisely of all 'literariness' of language).

Crucial here was the work of Jacques Derrida with its development of an idea of writing - tracing of differences - as the structuring condition of presence (see Derrida, J. §2). There is nothing without writing, no 'outside-text' in the sense of some instance to which a text could be referred that would not itself be textual. Derrida's critique of the Western conception of the sign as underpinning a continual denial of writing (the signified received as having an identity distinct from the signifier) could be given an explicit political thrust when carried over to the understanding of ideologies as systems masking textuality and the production of posited 'grounds' of meaning. The struggle against what *Tel Quel* called 'the hypostasized result of a genesis effaced' and the retrieval of the materiality of the signifying process determined the ambition of the group's textual theory: 'a kind of permanent writing' that would break down the discursive orders of social intelligibility and accede to 'the figured truth of the world'. The *semanalysis* of J. Kristeva reformulated semiotics in these terms, pushing it beyond the study of sign systems into that of signifying practices, demanding an attention each time to the production of meaning and subject involved.

Tel Quel was thus characterized by a militant critique of representation for its containment of textual productivity but this critique was characterized in turn by the resolve not to be caught thereby in the position of accounting language the only reality. Against such idealism, *Tel Quel* proposed a 'semantic materialism': a text - any representation, discourse, piece of literature, philosophy or whatever - is never derived or justified from a point outside of textuality *and* there is a heterogeneous material reality inscribed in the subject's production in and through textuality that is at once social and psycho-sexual. The review thus took as its necessary task the articulation of historical materialism (using the work of Louis Althusser) and psychoanalysis (using that of Jacques Lacan) within a general theory of this textual process of the production together of subject-history-desire.

By 1977, belated disillusionment with existing Communist societies had shifted the review into distrust of any political vision of the world and the conviction of the impossibility of revolution. With Freud's pessimism concerning the human psyche defeating Marx's optimism regarding social transformation, the limit-texts were again important, read as engaging the impasses of sexuality and subjectivity, as exploding the dangerously deluded Enlightenment ideas of reason and progress to which political projects like Marxism refer. To these texts were now added the literary testimonies of the modern dissidents (Alexander Solzhenitsyn in particular). *Tel Quel*'s conception of writing was discharged of political terms and returned to an insistence on literature's value in itself as the fundamental site of freedom, dissidently marginal to all establishments of meaning.

Throughout the changes and reversals in the review's positions, what remained constant was the will to propose literature as a vital and vitally specific force. *Tel Quel*'s significance lies in its extreme theoretical and practical exploration of the political possibilities and contradictions of such a proposition.

See also: Semiotics; Structuralism

STEPHEN HEATH

References and further reading

Forrest, P. (1995) *Histoire de Tel Quel*, Paris: Editions du Seuil.(A detailed, sympathetic account of *Tel Quel*'s intellectual history.)

Kauppi, N. (1990) *The Making of an Avant-Garde: Tel Quel*, trans. A.R Epstein, Berlin and New York: Mouton de Gruyter, 1994.(A critical examination of the socio-cultural conditions of the review's appearance, development and importance.)

Kristeva, J. (1969) Σημειωτικη*: Recherches pour une sémanalyse*, Paris: Editions du Seuil, Collection *Tel Quel*. (The collection of essays in which Kristeva first defined and elaborated semanalysis.)

Sollers, P. (1968) *Logiques*, Paris: Editions du Seuil, Collection *Tel Quel*.(Readings of limit-texts as the approach to a theory of writing.)

Sollers, P. and Hayman, D. (1981) *Visionà New York*, Paris: Grasset.(In a series of interviews Sollers describes and explains the evolution of *Tel Quel* in the 1960s and 1970s.)

Suleiman, S. R. (1989) 'As is', in D. Hollier (ed.) *A New History of French Literature*, Cambridge, MA: Harvard University Press.(A succinct presentation of the review and its context.)

Tel Quel (1962-82) Paris: Editions du Seuil.(The original quarterly review.)

Tel Quel (1968) *Théorie d'ensemble*, Paris: Editions du Seuil, Collection *Tel Quel*. (The fundamental collective statement, in a series of essays by members of the board, of the review's theoretical project at the moment of its greatest influence.)

Teleological ethics

The Greek telos *means final purpose; a teleological ethical theory explains and justifies ethical values by reference to some final purpose or good. Two different types of ethical theory have been called teleological, however. Ancient Greek theories are 'teleological' because they identify virtue with the perfection of human nature. Modern utilitarianism is 'teleological' because it defines right conduct as that which promotes the best consequences.*

Twentieth-century philosophers have distinguished ethical theories as 'teleological' or 'deontological', placing classical Greek and Roman theories alongside utilitarianism in the first category, and Kant's ethical theory together with rational intuitionism in the second. As the distinction is usually drawn, a teleologist explains the rightness or virtue of action in terms of the good realized by means of it, while a deontologist treats rightness as an independent value, and argues that we may be required to do what is right at the expense of realizing the good (see Good, theories of the; Right and good).

This distinction, however, obscures the difference between two different types of theories that have been called 'teleological'. Classical philosophers like Plato and Aristotle, and modern Idealists like Bradley and Green, are 'teleologists' because they believe that human nature is fully realized or perfected through the practice of virtue (see Perfectionism). Virtue, according to these philosophers, is not a means to human perfection: instead, virtue is itself the perfection of human nature - in the way that intelligence is the perfection of thought, or beauty of appearance. In these theories there is no general distinction between 'the right' and 'the good': as expressions of the virtues, virtuous actions are good in themselves.

Modern 'teleologists', by contrast, assign virtuous or right action only an instrumental value (see Consequentialism). In this respect, modern 'deontologists', who see the rightness or virtue of an action as a kind of value that is intrinsic to it, are closer to the classical teleologists (see Deontological ethics). What distinguishes modern deontologists from classical thinkers is that they recognize forms of 'natural' or nonethical goodness which are independent of, and can conflict with, ethical values. For instance, many modern deontologists identify happiness with nonethical goods like pleasure, the pursuit of which can conflict with the right (see Happiness). By contrast, Plato and Aristotle argued that *eudaimonia*, happiness or flourishing, essentially involves virtue (see Eudaimonia). In treating happiness as an independent or nonethical value, modern deontologists resemble the modern 'teleologists', who believe that right action is simply that which promotes an independently identifiable good.

Thus although both classical and modern teleologists believe that the good is realized through virtuous action, it is in different senses. Classical teleologists argue that virtue is identical with the best state of a human being, while modern ones argue that virtue promotes an independent, nonethical good. Philosophers in the late twentieth century have begun to mark this distinction by referring to those who believe that rightness consists in the production of nonethical good as 'consequentialists' rather than 'teleologists'.

See also: Intuitionism in ethics; Kantian ethics; Stoicism; Utilitarianism; Values

CHRISTINE M. KORSGAARD

References and further reading

Frankena, W. (1963) *Ethics*, Foundations of Philosophy Series, Englewood Cliffs, NJ: Prentice Hall.(A very influential introduction to ethics, in which the teleological/deontological distinction is made, and various theories categorized, very much as they are in the opening paragraphs of this entry.)

Herman, B. (1993) 'Leaving Deontology Behind', in *The Practice of Moral Judgment*, Cambridge, MA: Harvard University Press.(An attack on the idea that Kant is a deontologist, as well as on the idea of deontology, and with it, the deontological/teleological distinction.)

Korsgaard, C.M. (1996) 'From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action', in S. Engstrom and J. Whiting (eds) *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, New York: Cambridge University Press.(A comparison of Kant's and Aristotle's views on the value of actions, defending the claim that Kant inherits the classical view that there is a special kind of value intrinsic to

actions.)

Mill, J.S. (1861) *Utilitarianism*, ed. G. Sher, Indianapolis, IN: Hackett Publishing Company, 1979.(Mill claimed an alliance between utilitarians and ancient Greek moralists, and he opposed this approach to that of the intuitionists and Kant. His influence helped to popularize this way of categorizing theories.)

Muirhead, J.H. (1932) *Rule and End in Morals*, Oxford: Oxford University Press.(Muirhead was the first to divide ethical theories into the categories ‘teleological’ and ‘deontological’. In this book he discusses the debate between the two types of theories, and urges the one-sidedness of both positions.)

Teleology

Teleology is the study of purposes, goals, ends and functions. Intrinsic or immanent teleology is concerned with cases of aiming or striving towards goals; extrinsic teleology covers cases where an object, event or characteristic serves a function for something.

Teleological explanations attempt to explain X by saying that X exists or occurs for the sake of Y. Since the question 'For what purpose… ?' may be construed either intrinsically or extrinsically, such explanations split into two broad types: those that cite goals of an agent, and those that cite functions.

The history of Western philosophy and science has been characterized by major debates about the logic, legitimacy and proper domains of these types of explanation. They still raise problems in contemporary biology and psychology. The modern debates have progressed considerably from the earlier ones, although continuities do exist.

1 Goals

Aristotle's views about the domain of striving that is present in nature were challenged during the Renaissance. Aristotle held that goals were 'final causes', that inanimate things seek natural places or states which are proper to their kind, and that growth and development in living things is directed towards the attainment of maturity.

The term 'final cause' misled some commentators, who assumed that a final cause is an efficient cause that comes after its effect. This could not have been Aristotle's view. Aristotle believed that trees grow leaves in order to protect their fruit (*Physics:* 199a 26-9), but he recognized that the fruit is not always successfully protected. If birds eat the fruit, and hence the final cause fails to come about, this fact in no way undermines the teleological explanation. If he had meant that fruit-protection was an efficient cause working backwards in time, the failure to come about would undercut the efficient causal explanation. A final cause, for Aristotle, is a 'that for the sake of which' (see Aristotle §9).

Still, there were other objections to his explaining the movements of inorganic bodies by ascribing goals to them. Francis Bacon, Galileo and Newton eschewed such explanations on the grounds that they were entirely speculative and otiose. First, the alleged striving could not be identified independently of the changes that actually occurred to the body, nor could its goal be identified - the hypothesis was untestable. Second, the hypothesis was unnecessary, at least in the fields of mechanics and dynamics, since a complete explanation could be provided in terms of antecedent causes and the laws of motion. As Bacon (1623 III: ch. 5) put it, 'Inquiry into final causes is sterile, and, like a virgin consecrated to God, produces nothing'.

Modern science does not sanction the ascription of goals or strivings to inanimate objects, except possibly to such artefacts as guided missiles, autonomous robots, and mechanical searching devices, which were first designed in the 1940s. But here the goal-talk is perhaps only 'as if' (Woodfield 1976: ch. 11).

In the life sciences, however, it is accepted that human beings and animals *do* strive after goals. The conviction that we are intentional agents is central to our self-image and to society. Moreover, animal and human goals can be identified in advance of the behaviour that they explain. The hypothesis that an animal is striving for food, for example, can be tested experimentally. So it is not true that goal-explanations are *in principle* vacuous.

Whether plants strive is, perhaps, unclear. Although we speak as if they did ('The flower turned in order to face the sun' and so on), such locutions may be merely a hangover from an Aristotelian tradition. They survive because we find them picturesque or convenient. In late-twentieth-century biology, vitalistic teleological theories of growth and development are not thought to be respectable, even though such processes cannot quite be explained in wholly physico-chemical terms (see Vitalism).

Are goal-explanations restricted, then, to the domain of intentional behaviour performed by intelligent organisms? The central case is surely that of the animate agent, conscious of what it wants, sensitive to information about its environment and able to represent alternative plans to itself. If goals always involve desires, beliefs and other mental states, then intrinsic teleological explanations are a species of mentalistic explanation. The main problem that arises next is to provide a satisfactory account of intentionality (see Intentionality).

Several philosophers in the twentieth century have tried to make room for a distinctive form of goal-explanation which is not necessarily mentalistic (Braithwaite 1953; Nagel 1961; Taylor 1964; Wright 1976). Such theories may be viewed as broadly Aristotelian in the sense that they locate striving in activity which exhibits a distinctive pattern or causal structure.

2 Functions

Aristotle maintained that if an item *X* is a part of a system *S* in which *X* performs a characteristic activity that benefits *S*, then *X* exists and acts for the sake of *S*. *S* might be a living organism or something bigger, such as a bee-colony, an ecosystem or even the world as a whole. The fact that *X* serves a function for *S* is supposed to explain why *X* is present in *S*. Aristotle's doctrine was naturalistic in the sense that it did not postulate a supernatural designer, but it was not wholly naturalistic, since it employed the notion of *benefiting*. The main problem with extrinsic teleological explanations in biology is to see precisely how they work.

Even supposing that it is a 'fact' that *X* does good to a bigger system *S*, that fact alone is insufficient to account for *X*'s existence. Some additional premise or principle seems to be required. For example, if nature had been designed and created by a benevolent and omnipotent God, the existence of *X* in *S* would be explicable in terms of God's wishes, beliefs and creative acts. This form of functional explanation has a familiar logic: we use such explanations when giving the reasoning that leads human beings to design and produce useful artefacts. By supplementing the explanation in this way, we present extrinsic teleology as being derivative upon the intrinsic teleology of the designer (see God, arguments for the existence of §4). This solution is unsatisfactory, however, since neither biologists nor laypeople feel that the validity of 'natural function' explanations is dependent upon any theological assumptions. Either the explanations have some other form, then, or they are not genuine explanations at all.

In 1859, Darwin's theory of evolution by natural selection showed how harmonious systems could have arisen naturally, without the need for a designer. Darwin's theory explains the existence of *X* in *S* as the outcome of a gradual process. Ancestors of *S* who possessed parts similar to *X* survived and reproduced more successfully than their relatives who lacked parts similar to *X*, and these ancestors reproduced true to form (their offspring had organs like *X*, including *S* who has *X*).

Darwin made the designer-hypothesis redundant. Was his theory anti-teleological? Darwin took it as a datum that biological parts and characteristics which have survived the selection process are normally useful for their owners; his theory asserted that they persisted *because* they were useful. This looks like a vindication of Aristotelian extrinsic teleology. Upon further reflection, however, it hardly amounts to a ringing endorsement since Darwin's theory can be stated without employing the term 'function' or any teleological language at all (see Darwin, C.R.; Evolution, theory of).

Contemporary philosophers separate into two camps on the question of the logic of functional explanations in biology. The 'naturalistic revisionists' *redefine* the concept of function in terms of the causal-historical Darwinian selection hypothesis. They keep the old teleological language but sanitize it (Wright 1976; Millikan 1984).The other camp consists of 'semantic conservationists' who maintain that natural functionality cannot be *defined* in terms of Darwin's theory (Woodfield 1976). On the latter view, talk of natural functions is committed to assumptions about benefit and harmony and goodness that are extraneous to science and probably perspective-dependent. This view implies that functional explanations are still potentially problematic.

It is possible to maintain that functional explanations of undesigned phenomena have some scientific merit even if they are not wholly objective. Kant (1790) argues that the attribution of natural functions to organs is heuristic: it helps to systematize our knowledge of organisms and generates further 'how?' questions. Kant's sophisticated defence does not license the unbridled attribution of good consequences to everything in nature - a tendency ridiculed earlier by Voltaire (1758) (see Kant, I. §13).

In the 1980s, naturalistic revisionists, notably Millikan (1984), began to exploit the Darwinian account of functionality as a tool for solving the problem of intentionality. The key insight is that desires, intentions and other mental states with intentional contents can themselves be seen as biologically adaptive states or as the products of mechanisms that are adaptive. The hope is to provide a naturalistic reduction of intentionality. This ambitious

research-programme would make extrinsic teleology more fundamental than intrinsic teleology.

After more than two millennia of debates since Aristotle, teleology continues to provoke lively controversy among analytic philosophers.

See also: Functional explanation

ANDREW WOODFIELD

References and further reading

Aristotle (*c.* mid 4th century BC) *Physica*, trans. R.P. Hardie and R.K. Gaye in *The Works of Aristotle*, ed. W.D. Ross and J.A. Smith, vol. 2, Oxford: Oxford University Press, 1930.(The classic account of teleology.)

Bacon, F. (1623) *De Augmentis Scientiarum*, in *Works of Sir Francis Bacon*, vol. 7, London: W. Baynes & Son, 1824.(Renaissance classic, emphasizing observation and induction as the royal route to knowledge.)

Braithwaite, R.B. (1953) *Scientific Explanation*, Cambridge: Cambridge University Press.(Textbook in analytic philosophy of science, still rewarding after nearly half a century.)

Darwin, C. (1859) *The Origin of Species*, London: John Murray.(Classic. The source of modern evolutionary theory.)

Kant, I. (1790) *Critique of Teleological Judgement*, trans. J.C. Meredith, Oxford: Oxford University Press, 1928.(Complement to Kant's Critiques of Pure Reason and Practical Reason.)

Millikan, R. (1984) *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.(Original defence of a biological theory of representation.)

Nagel, E. (1961) *The Structure of Science*, London: Routledge & Kegan Paul. (Logical empiricist textbook in the philosophy of science.)

Taylor, C. (1964) *The Explanation of Behaviour*, London: Routledge & Kegan Paul. (Defines a kind of non-mentalistic explanation that has a distinctive teleological form.)

Voltaire, F.-M. de (1758) *Candide*, trans. J. Butt, Harmondsworth: Penguin, 1947. (Satire directed against unthinking optimism.)

Woodfield, A. (1976) *Teleology*, Cambridge: Cambridge University Press. (Monograph defending a unified analysis of teleological language.)

Wright, L. (1976) *Teleological Explanations*, Berkeley, CA: University of California Press.(Monograph arguing that such explanations are a special kind of causal explanation.)

Telesio, Bernardino (1509-88)

Bernardino Telesio was a philosopher from southern Italy. He was one of the Renaissance philosophers who developed a new philosophy of nature: his most important book was called De rerum natura iuxta propria principia (On the Nature of Things According to Their Own Principles). Telesio approached natural philosophy empirically, and regarded it as a separate field of study from theology and metaphysics. He believed that all natural beings were animate; and, by arguing that the two general principles of heat and cold affected the whole universe, he resisted the Aristotelian division between the corruptible earth and the eternal heavens. However, despite his apparent anti-Aristotelianism and his sympathy for the Presocratics, Telesio owed much to Aristotle, and tried to transform rather than destroy Aristotle's work. Telesio became the head of a Calabrian school, and was influential and widely discussed in his own time. Francesco Patrizi criticized him, but with respect; Tommaso Campanella followed him in his early works; and Thomas Hobbes drew inspiration from him.

1 Life

Not much is known about the circumstances of Telesio's life, the 'man of one book' who was entirely devoted to the development of a new philosophy of nature and who died when he had completed his task. He was born at Cosenza in southern Italy, the first of eight children in a poor but noble family. At an early age he followed his uncle and teacher Antonio Telesio to the north, where his uncle taught Latin and Greek (first in Milan and then in Venice). He attended the University of Padua, which was well known throughout Europe, studying mathematics with Federico Delfino and philosophy with the Aristotelian Vincenzo Maggi (or Madius). He probably received his doctorate in 1535, after which he returned to his native province and retired to a monastery to work on his own philosophical projects. In 1552 he married a noble widow, who died eight years later. In 1565 Pope Pius IV offered Telesio the archbishopric of Cosenza, but he refused, and in 1576 he also refused Pope Gregory XIII's invitation to teach his philosophy in Rome. Despite his reluctance to enter public life, from 1547 on he made it known that he was working on a philosophy which would destroy the tyranny of Aristotle. In 1563 he travelled to Brescia, to discuss his ideas with his old friend and teacher, Maggi. With Maggi's approval, he published the first version of his natural philosophy, *De natura iuxta propria principia (On Nature According to Its Own Principles)* in 1565. Five years later, he published a revised second version with the definitive title *De rerum natura iuxta propria principia (On the Nature of Things According to their Own Principles)*, as well as three minor treatises, containing additional material. These works met with the critical approval of Francesco Patrizi, who was to develop his own views in his *Nova de universis philosophia (New Philosophy of Universes)* (1591). In 1586 Telesio published the final version of his treatise, now enlarged from two books to nine. He founded the *Accademia Cosentina* in Cosenza in order to promote the study of nature according to Telesio's principles, and it was in Cosenza that he died in 1588.

2 Approach to nature

Telesio's initial claim is of fundamental importance for his new philosophy of nature. He said that, in contrast to his predecessors, he would not rely on abstract principles invented by human reason, but would content himself with those concrete aspects of nature which are given in sense perception, and with what can be inferred from their similitudes. He did not wish to make the false claim that he was the first to deal with nature in an empirical manner, nor, contrary to Patrizi's accusation, did he pretend to be able to philosophize on the basis of mere sense perception. As he makes clear in the ninth and tenth chapters of the first book, he meant that he would refuse to explain what is perceived in nature in terms of what lies beyond the world of sensible experience, and hence is properly a metaphysical concern. Thus from the very beginning he postulated that natural philosophy should be developed as an autonomous science, independent of metaphysics and theology, and that nature itself should be considered as an autonomous structure, rationally designed to preserve itself eternally according to its own principles (*iuxta propria principia*).

3 The principles and structure of the universe

When Telesio began to determine the first principles of nature itself, he simply said that it is obvious to the senses that heat and cold are the most active general principles which constitute all natural beings, and offered no other justifications for this opinion. Patrizi and many scholars after him objected that these were the principles of

Parmenides, so Telesio was not original, nor were his principles based on sense perception. However, one should not forget that for Aristotle natural beings (in so far as they are sensible bodies) are constituted by the most basic sensible, that is tangible, qualities (hot, cold, wet and dry). Of these primary qualities, heat and cold are defined as the active ones. Thus in the basically Aristotelian world of sixteenth-century thought, Telesio's identification of the first sensible active principles with heat and cold seemed quite natural.

Telesio's originality lies in his use of the two principles. He did not regard them as being primary qualities or forms of prime matter, but as incorporeal substances which are received by the third passive principle, the corporeal mass (*moles*) on which they operate through condensation and rarefaction. Heat, since it is white, bright, extremely rarefied and in rapid continuous motion, is concentrated in the sun and the heavens. Cold, on the other hand, since it is black, dark, dense and totally unmoved, is concentrated in the earth. Since both are endowed with the capacity to regenerate and multiply themselves, to spread out in all directions, and to occupy and transform the whole mass according to their own nature, they fight each other. In so doing, they constitute the whole variety of natural beings through a continuous process of self-organization.

This general model of the universe has some interesting consequences from the perspective of the modern concept of nature. By locating the centre of one of the active principles, heat, in the sun and the heavens, Telesio identifies elementary heat, present in fire, with heavenly warmth, operating through the sun. In the Aristotelian tradition these were regarded as being different, and nature, which for centuries had been divided into the corruptible earthly sphere and the eternal heavenly natures, was thus unified by Telesio. Moreover, since it would obviously be difficult to derive space and time from his principles of nature, Telesio taught that time was not identical with the motion of bodies, and that space was not identical with the place (*locus*) formed by the surface of bodies. Rather, both are absolute entities, presupposed by the operations of the active principles. As a result, there is no doubt that a vacuum is possible, even though bodies, thanks to a certain delight in contact with one another, try to avoid it (see Cosmology §1).

4 The soul

When the active principles operate, they do so in order to overthrow their opposite and to preserve themselves. In order to be able to do this, they need the capacity to perceive both favourable and unfavourable operations, and to discriminate between them. Thus sense perception is as basic as the active principles themselves, for without it the whole process of interaction and self-organization would not work. If sense perception is linked to having a soul, as in the Aristotelian tradition, the active principles must be animate, and as a result all natural beings that are constituted by these principles and that are active in self-preservation must also be animate.

Thanks to the general animation of natural beings in so far as they are natural, even the lowest natural entities are raised to the dignity of sense perception and discrimination; and indeed the whole cosmos seems to be regarded as a living being. The capacities of sense perception and discrimination come to be seen as the most basic and most common natural qualities, which provide the general conditions for any natural process as such, even the most simple and mechanical ones such as rarefaction and condensation through heat and cold. However, when we arrive at more sophisticated organizations in the hierarchy of natural beings, such as those of plants, animals and human beings, it is only natural that their powers of perception and discrimination are also more sophisticated. Sense perception, cognition and judgment as found in human beings differ from the faculties of the souls of other natural beings only in their higher perfection, and it is tempting to suggest that for Telesio the very development of a philosophy of nature is just the natural operation of the most perfect natural being in its striving for self-preservation (see Soul, nature and immortality of the).

5 The immortal intellect of human beings

The soul which enables human beings, the most perfect of natural beings, to fulfil their specific operations, is diffused through their bodies in the form of spirits (*spiritus*) or highly rarefied matter, and since human spirits differ from animal spirits only in the degree of rarefaction, the perfection of human beings is based on this difference alone. As a result it seems puzzling that in his eighth book, Telesio introduces a second type of intellect which is immaterial and immortal. This intellect is created by God and infused into human beings so that they can contemplate and know divine and eternal objects in addition to sense objects, and so that they can strive for the preservation of their supernatural, incorporeal self. It has been supposed that this divine intellect, which seems

inconsistent with Telesio's pure natural philosophy, was merely a prudent sign of submission to the church, or else a proof that Telesio was conscious of the indispensability of metaphysics. However, before the general separation of physics and metaphysics developed in sixteenth-century Aristotelianism, the autonomy of natural philosophy was not thought to be affected by the admission of such metaphysical notions as an intellect that enters the body from outside (the so-called *nous thyrathen*) or of a second end for human beings (as discussed in Aristotelian ethics) beyond the purely natural one of self-preservation. Telesio may well have placed himself in this tradition.

See also: Aristotelianism, Renaissance

ECKHARD KESSLER

List of works

Telesio, B. (1565) *De natura iuxta propria principia liber primus et secundus (On Nature According to Its Own Principles)*, Rome: Antonius Bladus Impressor Cameralis.(The first version of Telesio's natural philosophy.)

Telesio, B. (1570) *De rerum natura iuxta propria principia liber primus et secundus, denuo editi (On the Nature of Things According to their Own Principles)*, Naples: Josephus Cacchius; repr. Naples: Istituto Suor Orsola Benincasa, 1989.(The revised second edition of Telesio's work on natural philosophy, in two books. The 1989 reprint includes Telesio's own notes.)

Telesio, B. (1586) *De rerum natura iuxta propria principia libri IX (On the Nature of Things According to their Own Principles)*, Naples: Horatius Salvianus; repr. Hildesheim: Olms, 1971; ed. with Italian trans. L. De Franco, vols 1-2, Cosenza: Casa del libro, vol. 3, Florence: La Nuova Italia, 1965-76.(The final version of Telesio's work on natural philosophy, enlarged to nine books from the original two. The 1971 reprint has a preface by C. Vasoli.)

Telesio, B. (1590) *Varii de naturalibus rebus libelli (Diverse Short Works on Natural Things)*, ed. A. Persius, Venice: Felix Valgrisius; repr. Hildesheim: Olms, 1971; ed. L. De Franco, Florence: La Nuova Italia, 1981. (Several treatises, written by Telesio while he was preparing his natural philosophy. The 1971 reprint has a preface by C. Vasoli; De Franco's 1981 edition includes Patrizi's objections to Telesio's philosophy on pages 463-74.)

References and further reading

De Franco, L. (1989) *Bernardino Telesio. La vita e l'opera (His Life and Works)*, Cosenza: Edizioni Periferia. (Collection of essays on the life and work of Telesio by one of the most competent scholars in the field.)

De Franco, L. (ed.) (1991) *Atti del convegno internazionale su Bernardino Telesio-Cosenza 12-13 Maggio 1989 (Proceedings of the International Convention on Bernardino Telesio…)*, Cosenza: Edizioni Periferia.(Various contributions on the life, works, and fortunes of Telesio.)

Kessler, E. (1992) 'Selbstorganisation in der Naturphilosophie der Renaissance' (Self-organization in Renaissance Natural Philosophy), *Selbstorganisation. Jahrbuch für Komplexität in den Natur-, Sozial- und Geisteswissenschaften* 3: 15-29.(Analysis of Telesio's universe as a self-organized system.)

Kristeller, P.O. (1964) 'Telesio', in *Eight Philosophers of the Italian Renaissance*, Stanford, CA: Stanford University Press, 91-109.(Short but reliable overview.)

Schuhmann, K. (1988) 'Hobbes and Telesio', *Hobbes Studies* 1: 109-33.(Shows that Hobbes was inspired by Telesio.)

Van Deusen, N.C. (1932) *Telesio: The first of the moderns*, New York: Columbia University Press.(On the life, cosmology, psychology, theology and fortunes of Telesio.)

Temple, William (1881-1944)

William Temple was concerned to unite personal Christian religion and social action, finding in the doctrine of the incarnation of God in Jesus Christ the basis for human dignity. He held that while the universe long existed without finite minds, and finite thought has its origin in the functioning of physical organisms, thought transcends its origins, creating ideas not occasioned by or referring to its physical environment, and purposively affecting that environment. Our capacities to seek truth, appreciate beauty and recognize duty are best explained by a purposive creative Mind using physical creation to bring other minds into existence. Created minds continually depend for their existence on God's continuing to sustain them.

Born in 1881, the son of an Archbishop of Canterbury, William Temple entered the Church of England and rose to the positions of Bishop of Manchester and then Archbishop of York before himself becoming Archbishop of Canterbury in 1942, at the height of the Second World War. He died in 1944.

An Anglican Platonist, Temple was concerned with overcoming the problems of evil and embracing good within a Christian context, and sought to reconcile the existence of evil and the existence of God. While it is the nature of evil to be against the will of God, monotheism, in order to retain plausibility, must be able not merely to show that the propositions 'God exists' and 'There is evil' are logically consistent, but to explain how God can use evil to further a providential plan. A basic theme in Temple's *Nature, Man, and God* (1934) is that what is wanted is some ground for the belief that the occurrence of evil is an element in the total good. Temple endeavoured to provide this by arguing that human beings are so created as to act in terms of what they believe their good to be. Creatures such as ourselves, in whom thought is rooted in a physiological organism, naturally tend to find their apparent good in what brings gratification, comfort and security. While not strictly necessary, that such creatures should sin is 'too probable not to happen'. Thus if thought is to arise from material conditions, it is natural (that is, easily explicable and to be expected) that those in whom it arises should have a bent towards self-concern and selfishness. This leads us to identify as our good conditions that are not our true good; we tend to a self-concern unwise even as regards our own interests. From this oversized self-concern arise individual and social evils.

Perhaps, Temple suggested, the conditions that involve the suffering of nonhuman animals are those which in humans give opportunity for forbearance, fortitude, courage, compassion and the like, presaging human moral life in the way that nonhuman animal consciousness prepares the way for human thought. In any case, since our actual good includes our coming to recognize a worth in our neighbours equal to that which we find in ourselves, our bent to selfishness precludes our identifying and successfully embracing our real good. We are unable on our own to correct this situation. While the successes of our searches for truth, beauty and goodness may not be ignored or repudiated, we can embrace our own good only by help from resources beyond our own. We find that self-centredness is sometimes temporarily and partially overcome by unselfish love.

If it is to be sufficiently conquered for us firmly to identify and seek our genuine good, we must ourselves be objects of an unselfish love that enables us to make others the object of an unselfish love of our own. Philosophical reflection and our knowledge of ourselves can make clear what is needed, but cannot say whether it is available to us. Temple endeavoured, then, to develop a defensible monotheism which would explain human thought and leave as little of the puzzle of human existence philosophically unilluminated as possible. He contended that when this is done, there remains a missing piece whose shape exactly corresponds to that of the Christian gospel, which proclaims that God became incarnate in Jesus Christ, whose exemplary life, atoning death, and resurrection from the dead for our justification manifested a costly unselfish love on the part of a God who can empower us to achieve our actual good. Temple claimed that this position is not refuted by the existence of evil, quoting as expressing his own position a passage from Bosanquet's *The Principle of Individuality and Value*:

> For a Christianity which has the courage of its opinions the idea of victory involves the idea of the Fall, and the answer [to the question of whether a world with both sin and divine atonement is better than one without sin] would be that the scheme of salvation, involving finiteness and sin, seem essential to the nature of God and the perfection of the universe.

(Bosanquet 1912: 425)

Temple was important in Anglican Christianity's effort to relate its worldview to increasing scientific knowledge and contemporary social problems without abandoning its distinctive traditional content.

See also: Evil, problem of

KEITH E. YANDELL

List of works

Temple, W. (1911) *The Nature of Personality*, London: Macmillan.(An account of personality developed in terms of Trinitarian doctrine.)

Temple, W. (1917) *Mens Creatrix (The Creative Mind)*, London: Macmillan.(Philosophical companion to *Christus Veritas* that argues that a correct philosophical perspective leads to the doctrine of the incarnation of God in Jesus Christ.)

Temple, W. (1923) *The Universality of Christ*, London: SCM Press.(Discussion of criteria for, and of Christianity as, a universal religion.)

Temple, W. (1924) *Christus Veritas* (The True Christ), London: Macmillan.(Theological companion to *Mens Creatrix* that finds the clue to the meaning of history in the doctrine of the Incarnation.)

Temple, W. (1934) *Nature, Man, and God*, London: Macmillan.(Gifford Lectures that present a monotheistic and Christian understanding of the development of human beings and the fact of evil.)

References and further reading

Bosanquet, B. (1912) *The Principle of Individuality and Value: The Gifford Lectures for 1911*, London: Macmillan.(Outlines Bosanquet's idealist metaphysics.)

Elliot-Binns, J.E. (1956) *English Thought 1860-1900: The Theological Aspect*, London: Longmans, Green & Co. (Discusses the historical, philosophical, theological and social setting of English theology in the indicated period.)

Horton, W.M. *Contemporary English Theology*, London: SCM Press.(Treats pre- and post-First-World-War English theology from a perceptive American perspective.)

Iremonger, F. (1948) *William Temple, Archbishop of Canterbury*, London: Oxford University Press.(The standard biography.)

Langford, T. (1969) *In Search of Foundations: English Theology 1900-1920*, Nashville, TN: Abington Press. (Presentation of English theological thought in its cultural and philosophical context; good bibliography.)

Mozley, J.K. (1951) *Some Tendencies in British Theology*, London: SPCK. (Discussion of British theology from *Lux mundi* (1889) to the thought of Oman.)

Thomas, O.C. (1961) *William Temple's Philosophy of Religion*, London: SPCK.(General philosophical assessment of Temple's views.)

Tennant, Frederick Robert (1866-1957)

A prolific writer on religion and philosophical theology, Tennant produced book-length studies of topics as diverse as the philosophy of science and the origin of sin. He is best known by philosophers for his two-volume Philosophical Theology (1928, 1930), which offered the most sophisticated version then available of the argument from design for the existence of God. Tennant's philosophical legacy consists primarily in the influence his methodology has exerted on later philosophical theologians.

F.R. Tennant received his graduate education at Cambridge University and spent his academic career there, as a fellow of Trinity College and, eventually, as a university lecturer in philosophy of religion. His specifically theological works include three studies of the concepts of sin and the Fall. His philosophical works include monographs on the concept of miracles and the nature of belief.

Tennant's most influential contribution to philosophy, the richly detailed *Philosophical Theology* (1928, 1930), employs his thinking on topics in the philosophy of empirical science to produce an intricate version of the argument from design for the existence of God (see God, arguments for the existence of §§4-5). Recognizing the force of Hume's attack on traditional forms of the argument (see Hume, D. §6), Tennant proposes a deliberately weakened conclusion: 'The attributes to be ascribed to God will be such as empirical facts and their sufficient explanation indicate or require' (1930: 78), a list of attributes that, he concedes, may fall well short of the elaborate list theists have traditionally ascribed to God.

Like previous versions of the design argument, Tennant's version claims only to render its conclusion probable, where the probability in question is inductive, rather than statistical or logical. Inductive probability, the sort of probability that empirical evidence is supposed to lend to a scientific hypothesis, forms the basis, according to Tennant, of empirical reasoning in both science and everyday life. The 'multitude of interwoven adaptations by which the world is constituted a theatre of life, intelligence, and morality', he argues, makes inductively probable the existence of a purposive, intelligent creator (1930: 121).

Tennant concedes that naturalistic accounts such as evolutionary theory may explain each of the individual adaptations he cites, but he insists that in this case the whole exceeds the sum of its parts: naturalism can explain each adaptation but not their totality. Here Tennant anticipates the kind of 'cumulative case' arguments for theism offered by Basil Mitchell (1973) and others: while no single piece of evidence rationally compels belief in God, the pieces are said to reinforce one another in such a way that theism becomes inductively probable. Critics have insisted on focusing on the cogency of each piece of theistic evidence - reminding us that, in the end, ten leaky buckets hold no more water than one.

Tennant's design argument has encountered other, more specific objections. Some critics, such as John Hick (1966) and D.H. Mellor (1968), have objected to Tennant's particular use of probability theory and have challenged the relevance of any kind of probabilistic reasoning to theistic belief. They give Tennant credit for recognizing that statistical and logical probability apply only dubiously to the question of how likely it is that the entire universe arose by chance. But they charge him with appealing to a notion of inductive probability that he fails to explain rigorously and with using even inductive probability in a fallacious way. Others, including D.L. Scudder (1940), have faulted Tennant for failing to treat religious experiences either as a fundamental basis for religious belief or as further data in support of the theistic hypothesis.

These objections have since been addressed by a number of philosophers who, because of Tennant's apparent influence on them, may be counted among his successors, including Mitchell, George Schlesinger and especially Richard Swinburne. Like Tennant, they have treated theism as in some sense a scientific hypothesis to be tested by the usual inductive methods. Swinburne's contributions to philosophical theology have sought to apply more sophisticated versions of probability theory to the question of God's existence, a methodological improvement on Tennant's work but squarely in the same spirit. Swinburne has also tried to meet another objection to Tennant's argument by making the widespread existence of religious experiences an explicit part of the case for theism (see Religious experience §2). Although cited less frequently with each generation, Tennant's writings stand at the head of a steady stream of probabilistic arguments from design.

STEPHEN MAITZEN

List of works

Tennant, F.R. (1902) *The Origin and Propagation of Sin*, Cambridge: Cambridge University Press.(Four very clear lectures on original sin, with a long preface responding to criticism.)

Tennant, F.R. (1903) *The Sources of the Doctrines of the Fall and Original Sin*, Cambridge: Cambridge University Press.(A companion to the previous volume, containing a thorough survey of the historical origins of the doctrine.)

Tennant, F.R. (1912) *The Concept of Sin*, Cambridge: Cambridge University Press.(A clear and precise conceptual clarification of the connotation of 'sin'.)

Tennant, F.R. (1925) *Miracle and Its Philosophical Presuppositions*, Cambridge: Cambridge University Press. (Three very short and lucid lectures on miracle, law and supernatural causation, credibility and actuality.)

Tennant, F.R. (1928, 1930) *Philosophical Theology*, Cambridge: Cambridge University Press, 2 vols.(Immensely detailed and fairly technical but still readable philosophy of self and soul, of religious experience and knowledge, and the idea of God.)

Tennant, F.R. (1932) *Philosophy of the Sciences*, Cambridge: Cambridge University Press.(Another short, lucid lecture series, placing science and theology with respect to the 'psychology of knowledge'.)

Tennant, F.R. (1943) *The Nature of Belief*, London: Centenary Press.(A slightly longer and more complex work on the nature and justification of faith.)

References and further reading

Hick, J. (1966) *Faith and Knowledge*, Ithaca, NY: Cornell University Press, 2nd edn. (Questions the applicability of probability theory to religious belief.)

Mellor, D.H. (1968) 'God and Probability', *Religious Studies* 5: 223-34.(Criticizes Tennant's use of probability theory and questions the applicability of probabilistic reasoning to religious belief.)

Mitchell, B. (1973) *The Justification of Religious Belief*, London: Macmillan. (Explores the prospects for a probabilistic, 'cumulative case' argument for theism.)

Schlesinger, G. (1977) *Religion and Scientific Method*, Dordrecht: Reidel. (Assimilates reasoning about God to scientific reasoning; useful discussions of confirmation theory.)

Scudder, D.L. (1940) *Tennant's Philosophical Theology*, New Haven, CT: Yale University Press.(Detailed critique of Tennant's major work, focusing on Tennant's allegedly inadequate treatment of religious experiences.)

Swinburne, R. (1979) *The Existence of God*, Oxford: Clarendon Press.(Detailed and influential argument, in the spirit of Tennant, for the conclusion that, more probably than not, God exists.)

Tense and temporal logic

A special kind of logic is needed to represent the valid kinds of arguments involving tensed sentences. The first significant presentation of a tense logic appeared in Prior (1957). Sentential tense logic, in its simplest form, adds to classical sentential logic two tense operators, P and F. The basic idea is to analyse past and future tenses in terms of prefixes 'It was true that' and 'It will be true that', attached to present-tensed sentences. (Present-tensed sentences do not need present tense operators, since 'It is true that Jane is walking' is equivalent to 'Jane is walking'.) Translating the symbols into English is merely a preliminary to a semantics for tense logic; we may translate 'P' as 'it was true that' but we still have the question of the meaning of 'it was true that'. There are at least two versions of the tensed theory of time - the minimalist version and the maximalist version - that can be used for the interpretation of the tense logic symbols.

The minimalist version implies that there are no past or future particulars, and thus no things or events that have properties of pastness or futurity. What exists are the things, with their properties and relations, that can be mentioned in certain present-tensed sentences. If 'Jane is walking' is true, then there is a thing, Jane, which possesses the property of walking. 'Socrates was discoursing', even if true, does not contain a name that refers to a past thing, Socrates, since there are no past things. The ontological commitments of past and future tensed sentences are merely to propositions, which are sentence-like abstract objects that are the meanings or senses of sentences. 'Socrates was discoursing' merely commits us to the proposition expressed by the sentence 'It was true that Socrates is discoursing'.

The maximalist tensed theory of time, by contrast, implies that there are past, present and future things and events; that past items possess the property of pastness, present items possess the property of presentness, and future items possess the property of being future. 'Socrates was discoursing' involves a reference to a past thing, Socrates, and implies that the event of Socrates discoursing has the property of being past.

1 *De dicto* and *de re* occurrences of tenses

If a tense occurs *de dicto*, then the tensed expression ascribes a temporal property to a 'dictum' (proposition). 'It will be true that the sun is exploding' involves a *de dicto* occurrence of 'will be': the property 'having future truth' is ascribed to the proposition expressed by 'The sun is exploding'. If a tense occurs *de re*, then the tensed expression ascribes a property to a thing or event. The sentence 'The sun will explode' involves a *de re* use of 'will' since the tensed expression 'will explode' ascribes the temporal property 'exploding in the future' to the sun (see De re/de dicto).

The tense logic favoured by minimalists is based on the idea that all occurrences of past and future tenses are *de dicto* (even though some might appear to be *de re*) and therefore can be represented by sentential operators, such as 'it was true that'. According to this theory, the sentence 'The sun will explode' only appears to ascribe a temporal property to the sun; in reality it ascribes a temporal property to a proposition. 'The sun will explode' means the same as 'It will be true that the sun is exploding'.

2 Quantified tense logic

The *de re*/*de dicto* distinction also appears in the different approaches of the maximalist and minimalist theorists to the syntax of quantified tense logic. Quantified tense logic may be formulated in at least two ways. One way is based on the idea that quantifiers ('any', 'some') occur only within the scope of tense operators. In the sentence 'Some thing will fly to Mars', the quantifier 'some' does not occur within the scope of a tense operator. But it does in 'It will be true that some thing flies to Mars'.

The minimalist theory implies that quantification over past or future particulars should occur only in the scope of tense operators. Consider the sentence 'Some person did exist who wrote *The Republic*'. The minimalist will take this to express the same proposition as 'It was true that some person is writing *The Republic*', which has the symbolic form $P(\exists x)Gx$. The quantified expression 'some person' occurs within the scope of 'it was true that'; this implies that 'some person' does not now refer to anything (there are no past people for it to refer to), but used to refer to something; to a person that existed at the time the proposition was true.

The minimalist assumption that temporal properties should be ascribed only to propositions also appears in the treatment of quantification over present particulars. The sentence ‘Some person is presently writing a book’ is not interpreted as ascribing the temporal property ‘presently writing a book’ to a person. Rather, it is taken to mean ‘It is true that some person is writing a book’, which would involve a present tense operator. This sentence ascribes the property of being presently true to the proposition ‘Some person is writing a book’. This minimalist interpretation of quantification is exemplified in A.N. Prior’s remark in *Past, Present and Future*: ‘a quantifier preceding any such operator [any tense operator] is naturally taken to be governed by the ‘‘It is the case that-’’, which is prefixable to anything we say, and therefore to range over what *now* exists’ (1967: 144; emphasis added).

The maximalist tensed theory of time, by contrast, allows quantification over past, present or future particulars to occur outside the scope of tense operators. The expression ‘some person’ in the sentence ‘Some person existed who wrote *The Republic*’ quantifies over all past, present and future persons. This reflects the assumption that some people exemplify pastness, some, presentness and some, futurity.

The maximalist theory implies a *de re* use of the symbols *P* and *F*. Traditional tense logic (following Prior) uses *P* and *F* only in a *de dicto* manner, to mean ‘it was true that’ or ‘it will be true that’. But in an expression of the form $(\exists x)Px$, ‘*P*’ does not mean ‘it was true that’ but stands for a property of *x*; the property of pastness. The expression translates as ‘Something is past’. Since ‘*P*’ here is not used as a sentential operator, but as a predicate, this should be marked in our symbolic notation. We may put an asterisk after *P* to mark the use of this symbol as a predicate. Thus we should say $(\exists x)P^{*}x$.

The symbol *P* is also used in a *de re* manner in $(\exists x)[PG]x$. Here the symbol *P* operates on the predicate *G* to form a more complex predicate. This translates as ‘Something exemplifies past *G*-ness’ or, equivalently, ‘Something’s exemplification of *G* has pastness’. To mark this distinctive use of *P* as a predicate operator, we can underline *P*, so that we say $(\exists x)[\underline{P}G]x$.

According to the maximalist tensed theory of time, an adequate quantified tense logic must include *P* (a sentential operator), P^{*} (a predicate) *and* $\underline{P}$ (a predicate operator). Given these symbolic distinctions, maximalists can allow that *P* is used only *de dicto*.

The distinction between *de dicto* and *de re* occurrences of tense logic symbols is partly analogous to the same distinction in modal logic. The box is used *de dicto* in $\Box A$, which means ‘It is necessarily true that *A*’, where *A* is some proposition. The box is used *de re* in $(\exists x)\Box Fx$, which means that some thing is necessarily *F*.

3 A paradox for the minimalist theory

The minimalist tensed theory of time is false if there is a true sentence which a minimalist interpretation turns into a logical contradiction. The maximalist may claim there are such sentences. For example, the maximalist may claim that 10 billion years ago, before any intelligent organisms evolved, there was no language and thus no truths or falsehoods. Consider the sentence ‘It is logically possible that there was a time *t* at which there were no truths’. If ‘there was’ and ‘there were’ are *de dicto* occurrences of tense, then this sentence says the same as ‘It is logically possible that it was true at time *t* that there are no truths’, which is a logical contradiction. The sentence must instead be taken to mean that it is logically possible that some time *t* exemplifies both pastness and having contained no truths.

The minimalist may counter that Platonic realism is true and thus that it is necessary that there be truths at all times. The maximalist may respond that the truth of Platonic realism and the falsity of nominalism are determined by metaphysics, not by formal logic. ‘Platonic realism is true’ is not a logical theorem and ‘There were no truths 10 billion years ago’ is not a logical contradiction. Thus a formal logic that implies these sentences are logical truths or falsehoods is not in fact valid.

The minimalist may have a second response, namely that the sentence ‘It is logically possible that it was true at time *t* that there are no truths’ does not imply the existence of a truth at time *t*. The tense operator ‘it was true at time *t* that’ does not imply that the proposition ‘There are no truths’ *existed* at time *t*. Rather, it implies merely that the time at which the (present tense) proposition should be evaluated for truth or falsity is the past time *t*, and that the proposition possesses truth at this time. The maximalist will say that this response depends on the intelligibility of the thesis that the proposition ‘There are no truths’ possesses at time *t* the property of being true and yet does

not exist at this time. How can a nonexistent (at time *t*) proposition possess a property (at time *t*)? It is debatable whether the maximalist can derive a contradiction from the minimalist theory that some propositions do not exist at some of the times at which they exemplify the property of being true.

A.N. Prior, who presented the first significant tense logic (1957), defends a minimalist tensed theory of time (1967) and is the most influential opponent of the maximalist theory and *de re* occurrences of tenses. Prior's critical stance is analogous to Quine's position regarding quantified modal logic. Prior found problematic *de re* occurrences of 'was' and 'will be', objecting to the realism about properties of pastness, futurity and presentness associated with constructions of the form $(\exists x)P^*x$ (something exemplifies 'being past'). (Similarly, Quine found problematic *de re* occurrences of 'must be', and objected to the realism about essential properties associated with constructions of the form $(\exists x)\Box Fx$ (something exemplifies 'being necessarily *F*').)

In Prior's case, the rejection of the realist position is in part due to his categorization of pastness, presentness and futurity as *activities*. Prior suggests that these temporal properties of events (if there were such properties) belong in the same ontological category as activities done by things. He denies that there are *events* 'momentarily doing something called "being present" and then doing something else called "being past" for much longer' (1967: 18). Prior may be interpreted as suggesting that it is a mistake to categorize presentness and pastness along with such activities as running or talking, and then say that presentness and pastness (if there are such properties) are activities performed not by things, but by events.

The realist may grant this point, but argue that presentness and pastness need not be categorized as activities performed by events. 'Events possess temporal properties of being present or being past or being future' does not entail 'Presentness, pastness and futurity are activities done by events'. These temporal properties, like such properties as existence or self-identity, are unusual kinds of properties and should be placed in an ontological category by themselves.

4 Do *de dicto* occurrences imply *de re* occurrences?

The idea that all occurrences of tenses are *de dicto* leads to a minimalist semantic interpretation of such present-tensed sentences as 'The moon is bright'. According to this interpretation, this sentence does not report that an event, the moon's exemplification of brightness, has presentness. There is no first-order property of presentness and no event; there is only the thing, the moon, exemplifying brightness. The maximalist may argue that problems with this interpretation arise as soon as we ask *when* the moon exemplifies brightness. If 'exemplify' is tenseless, then there is no answer, and the analysis is faulty, since the present tense sentence 'The moon is bright' implies the moon is now exemplifying brightness. But if 'exemplify' is present-tensed, then we must ask what the semantic content of the present-tensed aspect of this verb is. The present tense must have some semantic content or meaning, since it conveys temporal information and is not a meaningless syntactic device such as 'It' in 'It is raining'. Further, to what ontological category does this temporal information belong?

According to the minimalist *de dicto* theory of tenses, the semantic category of the present tense is 'sentential operator' (the present tense has the same semantic content as 'it is true that') and the relevant ontological category is 'second-order property'; the present tense ascribes the property of 'being presently true' to a proposition. This theory implies that if the sentence 'The moon is bright' is true, we are ontologically committed to (1) the proposition, 'The moon is bright', (2) the proposition's exemplification of the property, 'being presently true', (3) the thing, the moon, and (4) this thing's exemplifying the property of brightness. The maximalist may object that 'exemplifying' in (4) is present-tensed, reflecting the fact that the thing presently exemplifies a property of brightness. According to the maximalist, the thing's *present* exemplification of brightness (rather than a timeless, past or future exemplification of brightness) is the ground for the proposition's being presently true.

5 Past and future existents

According to the minimalist theory, the ontological commitments of the sentence 'Plato had been alive' are merely the proposition 'Plato is alive' and the proposition's property of having been true. The maximalist may grant that these are the only ontological commitments if we accept the coherentist or eliminativist theory of truth. But the maximalist accepts the correspondence theory of truth and argues that this theory requires further ontological commitments. 'Plato had been alive' commits us to a past thing, namely, Plato, and a past event or state, namely,

Plato's being alive. 'Correspondence' is a symmetrical property. (It is a dyadic property and thus a relation.) If the proposition 'Plato is alive' had been true in the past, then this proposition exemplifies past correspondence to Plato's being alive. Since correspondence is a symmetrical relation, this implies that some event, Plato's being alive, exemplifies past correspondence to the proposition 'Plato is alive'. The reason this event exemplifies past correspondence is that the event exemplifies pastness. Since Plato is a part of this past event, Plato exemplifies pastness. It follows, the maximalist concludes, that some events and things are past.

This maximalist conclusion implies that the temporal property 'pastness' is *de re*, since it is a property of a thing, Plato, and of an event, Plato's being alive. There was something, Plato, over which we can now quantify; it is the case that $(\exists x)P^{*}x$. The same results hold for propositions that will correspond to something exemplifying some property.

One concern that minimalists have about the maximalist interpretation of tense logic pertains to the maximalists' thesis that past or future things or events *presently* possess properties. Consider the formula $(\exists x)P^{*}x$. Does not $(\exists x)$ in this formula convey that there presently exists an *x*? And does not $P^{*}x$ mean that this *x* is not presently existing but is past? If so (the minimalist may argue), the symbolic representation of *de re* occurrences of tenses turns true sentences into contradictions.

But the maximalist may respond that $(\exists x)$ has at least three different meanings in an adequate quantified tense logic, corresponding to three different senses of 'exists'. If 'exists' is present-tensed, it can have one of two senses. In one sense, it is logically equivalent to 'has the property of presentness', so that 'Mount Everest exists' is logically equivalent to 'Mount Everest has the property of presentness'. In the second sense, it can be equivalent to 'presently possesses some property'. Something that is past or future exists in this present-tensed sense, since if something is now past, it presently possesses the property of pastness. If 'exists' is tenseless, it is in effect disjunctively tensed; that is, it is equivalent to 'has existed, exists or will exist', where the middle 'exists' is equivalent to 'has the property of presentness'. If we say 'Socrates exists', we mean that Socrates either has existed or exists (presently) or will exist.

The maximalist argues that this threefold distinction shows that there is no logical difficulty in the idea that some past or future item presently possesses a property. The maximalist may note that 'Some past thing presently has the polyadic property of being remembered' is not a logical contradiction. Nor can it be turned into a logical contradiction by substitution of synonyms. Indeed, the maximalist continues, an analysis of the concept expressed by 'remembering' implies that the minimalist theory is false. Remembering is not a relation between a present person and a present image. If it were such a relation, it would follow that what is being remembered is in every case an image that exists simultaneously with the act of remembering. But this is an analytic contradiction, since it is a part of the logic of 'remembers' that what is remembered exists earlier than the act of remembering. Remembering a thing is a three-termed relation: a person remembers a past thing by forming a present image of that past thing. The 'of' expresses the relation in which the present image stands to the past thing.

The maximalist may elaborate that there will appear to be a difficulty about non-present items presently possessing properties if one is mistaken about the sorts of properties a past or future item can possess. Since Plato does not now exist, Plato cannot now possess the properties of walking, breathing, writing, thinking and so on. These properties can only be possessed by Plato when he is present. But Plato now possesses the properties of being past, of being referred to by the name 'Plato', of being thought about, of being earlier in time than Descartes, and so on. Further, Plato now possesses the properties of having walked, having written *The Republic*, having been alive, having been human, and so on, all of which can be expressed in a symbolic notation of the form $(\exists x)[\underline{P}G]x$, where $(\exists x)$ can be taken as tenseless or as present-tensed in the second sense distinguished by the maximalist.

See also: Adverbs; Continuants §2; Demonstratives and indexicals §1; Intensional logics; Logical and mathematical terms, glossary of; Modal logic; Modal operators; Time

QUENTIN SMITH

References and further reading

Burgess, J. (1984) 'Basic Tense Logic', in D. Gabbay and F. Guenthner, *Handbook of Philosophical Logic*, vol. 2, Dordrecht: Reidel.(A technical account of sentential tense logic.)

Cocchiarella, N. (1984) 'Philosophical Perspectives on Quantification in Tense and Modal Logic', in D. Gabbay

and F. Guenthner, *Handbook of Philosophical Logic*, vol. 2, Dordrecht: Reidel.(A presentation, with technical detail, of issues in quantified tense logic.)

Craig, W.L. (1996a) ‘Tense and the New B-Theory of Language’, *Philosophy* 71: 5-26.(Presents a defence of the tensed theory of time against the new tenseless theory (the B-theory) of time.)

Craig, W.L. (1996b) ‘The New B-Theory’s *tu quoque* Argument’, *Synthese* 105: 1-21.(Shows how recent work on indexicals, such as Kaplan’s and Perry’s, supports the tensed theory of time.)

Forbes, G. (1984) *The Metaphysics of Modality*, Oxford: Clarendon Press.(Includes a clear discussion of tense logic and its relation to modal logic.)

Gale, R. (1968) *The Language of Time*, New York: Humanities Press.(An influential book that includes a classic defence of the maximalist tensed theory of time.)

Haack, S. (1978) *Philosophy of Logics*, Cambridge: Cambridge University Press.(Chapter 9 of this book includes one of the most accessible introductions to tense logic.)

Kamp, J.A.W. (1971) ‘Formal Properties of "Now"’, *Theoria* 37: 227-73.(An influential technical account.)

Le Poidevin, R. (1991) *Change, Cause and Contradiction: A Defence of the Tenseless Theory of Time*, New York: St Martin’s Press.(Includes an accessible defence of the thesis that the tensed theory of time is false.)

Le Poidevin, R. (ed.) (1998) *Question of Time and Tense*, Oxford: Clarendon Press.(A collection of essays by Le Poidevin, Craig, Nerlich, Oaklander, Dyke, Smith and others about the tensed theory of time.)

Oaklander, L.N. (1984) *Temporal Relations and Temporal Becoming*, Lantham, MD: University of America Press.(Presents a sustained criticism of the tensed theory of time.)

Oaklander, L.N. and Smith, Q. (eds) (1994) *The New Theory of Time*, New Haven, CT: Yale University Press.(Essays by Kaplan, Oaklander, Mellor, Schlesinger, C. Williams, Smith and others on the tensed theory of time versus the new tenseless theory of time.)

Prior, A.N. (1957) *Time and Modality*, Oxford: Clarendon Press.(This classic is the primary source for modern tense logic.)

Prior, A.N. (1967) *Past, Present and Future*, Oxford: Clarendon Press.(A collection of papers on time and tense logic, some of which articulate and defend the minimalist tensed theory of time.)

Prior, A.N. (1968) *Papers on Time and Tense*, Oxford: Clarendon Press.(Papers relevant to tense logic and the minimalist tensed theory of time.)

Smith, Q. (1993) *Language and Time*, New York: Oxford University Press.(An explication and defence of the maximalist tensed theory of time, and criticism of the minimalist tensed theory and the tenseless theory of time.)

Smith, Q and Oaklander, L.N. (1995) *Time, Change and Freedom*, London: Routledge.(An introductory discussion of the tensed and tenseless theories of time and their relation to other topics in metaphysics.)

Tertullian, Quintus Septimus Florens (*c.* AD 160-*c.* AD 220)

Tertullian was the first Christian theological author to write in Latin, and is responsible for initiating the Latin vocabulary of Christian theology, including such important terms as persona *(person) and* substantia *(substance). His early works, including the Apologeticum, refute pagan misconceptions about Christianity and argue on philosophical and juridical grounds for religious freedom. His later theological treatises, such as De anima (On the Soul) and Adversus Marcionem (Against Marcion) reflect Tertullian's adherence, in about* AD *205-6, to Montanism, a Christian sect which emphasized asceticism, apocalypticism and prophecy. These lengthy works represent an effort to oppose those forms of Christianity that sought to ally themselves with Platonism, such as Gnosticism. After these defences of apocalyptic Christianity, Tertullian fades from historical view around* AD *220, leaving a legacy of charismatic truculence.*

Tertullian was born about AD 160 in Carthage. His family was pagan and his early education included rhetoric and law, together with some philosophy. Tertullian practised law at Rome until his conversion to Christianity around AD 193, after which he returned to Africa where he devoted himself to propagating his new faith and producing apologetical literature on its behalf.

Tertullian's lasting impact on the history of Western thought is a result of the brilliance of his rhetoric. His writings are highly individualistic: intense, pungent and paradoxical. They are full of striking aphorisms that have become famous in Western culture. Of the relationship of Christianity to reason and philosophy, he maintained 'What does Athens have to do with Jerusalem? What agreement is there between the Academy and the Church, between heretics and Christians?' (*De praescriptione haereticorum* 7) and of the incarnation: 'it is certain because impossible' (*De carne Christi* 5).

Tertullian was strongly opposed to the transcendentalism of earlier Christian apologists and Platonists (Justin Martyr, Theophilus, Irenaeus) and of the Gnostics (Marcion, Hermogenes, Valentinus). His ontology is Stoic in its origins. His major philosophical treatise is *De anima (On the Soul)*, an attempt to disprove the Platonic conception of a naturally immortal soul. For Tertullian, the soul is corporeal, a product of a 'soul-seed' transmitted at conception. According to Tertullian, the soul is a free material being whose continued existence depends upon its redemption by Christ and whose spiritual hope rests in its resurrection (see Soul, nature and immortality of the). By adopting this materialist psychology, Tertullian sought to associate Christianity with opponents of psychic pre-existence and transmigration. This was an act of deliberate philosophical positioning, setting the nascent Christian movement into alliance with Stoic and medical writers whose outlook on theology was materialistic.

Yet it is difficult to assess Tertullian's theological materialism. He insists that everything which exists is corporeal (*De carne Christi* 11) and that God is corporeal, albeit a 'spiritual' body (*Adversus Praxean* 7). This conflation of *spiritus* and *corpus* in God signals his unique character, since God alone is a perfect corporeal being, not produced by any other. However, God remains a substance distinct from other material entities, which are his products. His existence is known through them (*De resurrectione* 2-3). God is, however, understood by Tertullian to be at the extremity or outer surface of the cosmos, so that he is literally greater than all other things (*Adversus Praxean* 16). Thus, Tertullian draws on the Stoics for his attenuated materialist theology, but his precludes the Stoic tendency to assimilate God to the world. In insisting that God is *corpus*, Tertullian is unique among Patristic authors. His early treatise on Gnosticism, *Adversus Hermogenem (Against Hermogenes)*, attacks the common ancient theory that matter is eternal, and with it the 'demiurgic' view of God as the fashioner of order from disordered matter. Tertullian maintains instead that God, while himself a material spirit, is free from any constraint imposed by pre-existent matter, creating *ex nihilo* by his free will.

Finally, Tertullian contributed importantly to the Latin Christian definition of God as a 'trinity'; he is the first Latin author to employ the term *trinitas* in a technical sense (see Trinity). His formulation 'one substance, three persons' (*una substantia, tres personae*) became fixed in subsequent Latin theology. It was meant to capture the original unity of his material God, who is conceived as 'shooting forth' the Logos and the Spirit in a temporal process. The Logos is understood to exist before this act of generation within the Father's mind as a *ratio*, but to emerge subsequently as a distinct *persona* with the beginning of the cosmos. While Nicene orthodoxy would later reject these materialist and temporal implications, his innovative terminology would be adopted by Christian

orthodoxy.

Tertullian is thus an outrider in early Christian thought, a radical opponent of its developing Platonic mainstream. As such, he helps to show how different Christian theology might have been had its scriptural deposit been read through a materialist philosophical prism.

See also: Gnosticism; God, concepts of; Neoplatonism; Patristic philosophy; Stoicism; Trinity

JOHN PETER KENNEY

List of works

Tertullian, Quintus Septimus Florens (*c.* AD 197-*c.* AD 220) *Opera (Works)*, ed. J.G.P. Borleffs, E. Dekkers *et al.*, *Quinti Septimi Tertulliani Opera*, The Hague: Brepols, 1953-4, 2 vols; Corpus Scriptorum Ecclesiasticorum Latinorum vol. 20, ed. A. Reifferscheid and G. Wissowa, Vienna: F. Tempsky, 1890; vol. 47, ed. A. Kroymann, Vienna: F. Tempsky, 1906; vol. 69, ed. H. Hoppe, Vienna: F. Tempsky, 1939; vol. 70, ed. A. Kroymann, Vienna: F. Tempsky, 1942.

Tertullian, Quintus Septimus Florens (AD 197) *De testimonio animae (The Testimony of the Soul)*, Corpus Scriptorum Ecclesiasticorum Latinorum vol. 20, ed. A. Reifferscheid and G. Wissowa, Vienna: F. Tempsky, 1890.(A short treatise on the soul, closely related to the *Apologeticum*.)

Tertullian, Quintus Septimus Florens (*c.* AD 197) *Apologeticum*, Corpus Scriptorum Ecclesiasticorum Latinorum vol. 69, ed. H. Hoppe, Vienna: F. Tempsky, 1939.(An important legal and cultural defence of Christianity.)

Tertullian, Quintus Septimus Florens (before *c.* AD 200) *De praescriptione haereticorum (The Prescription of Heretics),* Corpus Scriptorum Ecclesiasticorum Latinorum vol. 70, ed. A. Kroymann, Vienna: F. Tempsky, 1942.(A debate between orthodox Christianity and its heretical opponents, written during Tertullian's Catholic phase.)

Tertullian, Quintus Septimus Florens (after AD 200) *Adversus Hermogenem (Against Hermogenes)*, Corpus Scriptorum Ecclesiasticorum Latinorum vol. 47, ed. A. Kroymann, Vienna: F. Tempsky, 1906.(An attack on the Gnostic Hermogenes of Carthage.)

Tertullian, Quintus Septimus Florens (*c.* AD 202-10) *Adversus Valentinianos (Against the Valentinians)*, Corpus Scriptorum Ecclesiasticorum Latinorum vol. 47, ed. A. Kroymann, Vienna: F. Tempsky, 1906.(On the tenets of the Valentinians, a Gnostic group. A good source for the developing polemical traditions within Christianity.)

Tertullian, Quintus Septimus Florens (*c.* AD 205) *De carne Christi (On the Flesh of Christ)*, Corpus Scriptorum Ecclesiasticorum Latinorum vol. 70, ed. A. Kroymann, Vienna: F. Tempsky, 1942.(An initial treatment of Christ's bodily resurrection, closely related to *De resurrectione carnis*.)

Tertullian, Quintus Septimus Florens (*c.* AD 205-7) *De resurrectione carnis (The Resurrection of the Flesh)*, Corpus Scriptorum Ecclesiasticorum Latinorum vol. 47, ed. A. Kroymann, Vienna: F. Tempsky, 1906; ed. and trans. E. Evans, *Treatise on the Resurrection*, London: SPCK, 1960.(A defense of the Christian concept of the resurrection of the body.)

Tertullian, Quintus Septimus Florens (*c.* AD 207-12) *Adversus Marcionem (Against Marcion)*, Corpus Scriptorum Ecclesiasticorum Latinorum vol. 47, ed. A. Kroymann, Vienna: F. Tempsky, 1906; ed. and trans E. Evans, *Adversus Marcionem*, Oxford: Clarendon Press, 1972, 2 vols.(A lengthy refutation of the Gnostic heresy of Marcion.)

Tertullian, Quintus Septimus Florens (*c.* AD 210-13) *De anima (On the Soul)*, Corpus Scriptorum Ecclesiasticorum Latinorum vol. 20, ed. A. Reifferscheid and G. Wissowa, Vienna: F. Tempsky, 1890.(One of Tertullian's longest and most important philosophical treatises. Written during Tertullian's later Montanist period.)

Tertullian, Quintus Septimus Florens (*c.* AD 213) *Adversus Praxean (Against Praxeas)*, Corpus Scriptorum Ecclesiasticorum Latinorum vol. 47, ed. A. Kroymann, Vienna: F. Tempsky, 1906; ed. and trans. E. Evans, *Treatise Against Praxeas*, London: SPCK, 1948.(Directed against Praxeas, the work articulates Tertullian's mature Trinitarian theology.)

References and further reading

Koch, H. (1934) 'Tertullianus', in A. Pauly and G. Wissowa (eds) *Real-Encyclopädie der classischen Altertumswissenschaft*, cols. 822-44.(An old but still significant study.)

Sider, R.D. (1971) *Ancient Rhetoric and the Art of Tertullian*, London: Oxford University Press.(Tertullian and ancient rhetorical traditions.)

Spanneut, M. (1957) *Le stoïcisme des pères de l'église (The Stoicism of the Church Fathers)*, Paris: Éditions du

Seuil.(Study locating Tertullian in the history of ancient thought.)
Spanneut, M. (1969) *Tertullien et les premiers moralistes africains (Tertullian and the First African Moralists)*, Gembloux: J. Duculot.(Study locating Tertullian in the history of ancient thought.)

Testimony

Philosophical treatment of the problems posed by the concept of knowledge has been curiously blind to the role played by testimony in the accumulation and validation of knowledge or, for that matter, justified belief. This is all the more surprising, given that an enormous amount of what any individual can plausibly claim to know, whether in everyday affairs or in theoretical pursuits, is dependent in various ways upon what others have to say. The idea that someone can only really attain knowledge if they get it entirely by the use of their own resources provides a seductive ideal of autonomous knowledge that may help explain the way epistemologists have averted their gaze from the topic of testimony. But, unless they are prepared to limit the scope of knowledge dramatically, theorists who support this individualist ideal of autonomy need to explain how our wide-ranging reliance upon what we are told is consistent with it. Characteristically, those who consider the matter acknowledge the reliance, but seek to show that the individual cognizer can 'justify' dependence upon testimony by sole resort to the individual's resources of observation, memory and inference. Testimony is thus viewed as a second-order source of knowledge. But this reductionist project is subject to major difficulties, as can be seen in David Hume's version. It has problems with the way the proposed justification is structured, with its assumptions about language and with the way the individual's epistemic resources are already enmeshed with testimony. The success or failure of the reductionist project has significant implications for other areas of inquiry.

1 The epistemological background

Philosophy in the Western tradition, and beyond, has long enshrined a particular picture of the starting point and the task of epistemology. The task has principally been seen as trying to understand what sort of thing or goal knowledge is and thereby to distinguish it from counterfeit, inferior or merely different items such as superstition, opinion or belief. In pursuit of this, it became necessary to further contrast or identify knowledge with such notions as justified or reliable belief. Often the task was propelled by worries about (or enthusiasm for) scepticism, either total or partial. But equally, philosophers uninterested in the sceptical problem have attempted to distinguish knowledge from various pretenders to its title and to chart the interrelations of the various areas, types or sources of knowledge (see Epistemology, history of; Ryle, G.).

In all of this thinking about the nature and value of knowledge, however, the picture has persisted (except in the speculations of a handful of thinkers) of the subject doing the investigation as initially positioned in a state of cognitive isolation. This picture has endured in spite of the fact that the data to be accommodated and explained seem ill at ease with it. Such data include the fact that these cognizers unhesitatingly claim to know many things that surely transcend the power of the individual resources they bring to the epistemic task. To be sure, individual cognizers know about the existence and disposition of middle-sized objects and their surroundings in the more or less immediate vicinity, but they also know facts well beyond their immediate ken, such as numerous everyday facts of geography and history. They also know how conceptually to identify and linguistically to label these objects and surroundings in terms intelligible to others. Beyond this array of relatively ordinary convictions, there are a host of more recondite things many such individuals know - such as the population of the USA, the day-to-day temperature of the environment, the non-existence of a largest prime number, the military practices of ancient peoples, the chemical composition of water or of various medicines, certain facts about the workings of engines and about the workings of institutions.

It is, I think, perfectly clear that the isolated individual imagined as the hero of so much epistemological writing is not equal to the task of amassing such knowledge from purely individual resources of perception, experiment, memory and inference. Some theorists, it is true, heroically deny that anything that involves accepting the word of others can really deserve the name of knowledge. But they have either failed to appreciate the enormity of this denial, in terms of the massive amount and quality of information thus excluded and its often intimate connection with much else that is included, or they have so construed knowledge as to make its attainment a rare privilege for the elite (as in Plato who banishes not only testimony but also perception from the realm of knowledge, and then imagines that a purified 'science' will emerge from contemplation of the unchanging Forms). Short of such heroic, but either myopic or puritan, responses we have no option but to recognize the need for a pooling of resources so that knowledge requires some reliance upon one's fellow cognizers. But the individualist picture dies hard. So the idea emerges that the necessary reliance upon the word of others in the enterprise of knowledge is itself something

for which the isolated individual discovers the need and provides (from individual resources alone) the justification.

2 Autonomy and reduction

This individualist idea is essentially reductionist, in the sense that it attempts to show that, at the significant core of the epistemological enterprise, we have no need of any basic or primitive reliance upon the word of others. At this level, we need only the individual resources of perception, memory and inference. We rely extensively on the reports of others, of course, in the final development of knowledge, but this reliance is itself established by and dependent upon the primary operations of those individual resources. The idea that this must be so has great appeal because of the influence of a certain ideal of intellectual 'autonomy'. This ideal is well captured at the beginning of modern philosophy in Descartes' philosophy. There we find an outlook on knowledge that remains immensely influential today even among those who reject, and even denounce, most of Descartes' specific conclusions and many of his assumptions. Descartes makes it clear that the task he sets himself is precisely that of finding out whether it is possible to have fully autonomous knowledge, knowledge that you have rationally guaranteed entirely by and for yourself. As he puts it in the *Discourse on Method*:

> So, too, I reflected that we were all children before being men and had to be governed for some time by our appetites and our teachers, which were often opposed to each other and neither of which, perhaps, always gave us the best advice; hence I thought it virtually impossible that our judgments should be as unclouded and firm as they would have been if we had had the full use of our reason from the moment of our birth, *and if we had always been guided by it alone*.
>
> (Descartes [1637] 1985: 117, emphasis added)

Although Descartes then tries to establish how he is entitled to trust himself in the proper exercise of his own isolated cognitive powers, he never addresses the question of how solitary cognizers can employ their reason to validate in an 'unclouded' fashion their inevitable reliance upon others. Later theorists, however, occasionally try to deal with this problem. From an ideal of entirely 'autonomous' knowledge there is born the correlative portrait of the autonomous knower who extends and enriches knowledge by relying upon others only by licensing their word to be reliable by unaided epistemic efforts. Different theorists have different stories to tell about how the autonomous knower brings off this remarkable feat. One of the most vivid presentations of the picture, and one of the most influential accounts of how the feat is to be done, may be found in David Hume, though there are also different attempts in a similar reductive spirit by other philosophers such as F.H. Bradley and Bertrand Russell.

3 Hume and the reductionist project

Hume's idea is that the 'autonomous knower' does indeed rely, and rely quite extensively, on the testimony of others, but does so because of constant conjunctions discovered by personal checking on their reliability in a testimony-free way. So Hume says of reliance upon testimony:

> our assurance in any argument of this kind is derived from no other principle than our observation of the veracity of human testimony, and of the usual conformity of facts to the reports of witnesses. It being a general maxim, that no objects have any discoverable connexion together, and that all the inferences, which we can draw from one to another, are founded merely on our experience of their constant and regular conjunction; it is evident that we ought not to make an exception to this maxim in favour of human testimony, whose connexion with any other event seems, in itself, as little necessary as any other.
>
> (Hume [1748] 1975: 111)

The idea turns out to be methodologically ambiguous: do the autonomous knowers accept testimony in an entirely piecemeal fashion because they have independently ascertained the reliability of this or that particular witness, or do they proceed by determining the reliability of the various classes of cognitive agents to which witnesses belong, or is it merely that they know they can do one or other of these checks if needs be? Given the scope and depth of our dependence on what others tell us, the first suggestion opens up our cognitive landscape to interminable investigative travelling that would make a nonsense of inquiry itself. The third asserts a mere possibility and hardly seems a move in the justificatory game. The second, which is more plausibly Hume's idea, reduces the labour somewhat, but still requires an immense amount of checking, and more significantly owes us a non

question-begging account of what counts as a class of witnesses. This is much harder to provide than it seems.

When Hume discusses the task of justifying reliance upon testimony, he talks at one point of its being ‘founded on past experience’ and varying in probity ‘according as the conjunction between any particular kind of report and any kind of object has been found to be constant or variable’ (Hume [1748] 1975: 112). But it is unclear whether ‘kinds of report’ are to be individuated by reference to type of speaker or type of content. If the former, then we are really checking upon expertise: we are seeking to find whether there is a high correlation between expert reports and the sorts of situations that experts report upon. But, even if we allow a broad interpretation of expertise (as we must if we are to include enough testimony for the project’s ambitions), we cannot detect expertise as we might detect colour or smell. That someone is an expert on the geography of Southeast Asia is either itself known in ways that rely directly or indirectly on the testimony of others, or it must be determined by observing personally some very high correlation between their reports and the state of the world. If the former, then we are no further advanced on the reductionist programme since the same problem arises over and over again; if the latter, then not only does the notion of an expert no longer provide us with a specification of a kind of report, but we would usually have to be experts ourselves to be in a position to determine whether the particular correlations obtain. On the other hand, if we interpret ‘kinds of report’ as ‘reports of kinds of situation’ we face other intractable problems. For instance, a problem arises immediately about the degree of specificity to be attached to the classification of kinds of situation. For the purpose of checking do we classify the utterance ‘This is a car that is in perfect working order’ as an existence report, a motor vehicle report, a mechanical report or an evaluation? There seems no intrinsic reason to proceed one way or another here, and yet it seems to be of the first importance how we go.

4 Vulnerable assumptions

Another problem for any reductionist project concerns the fact that the plausibility of supposing that testimony-free verifications of reliability are usually available seems covertly to depend upon accepting testimony-laden checks, observations and falsifications. We have seen this problem arise in the particular case of checking for expertise, but its significance is wider than that since all sorts of apparently personal verifications will turn out to be dependent upon unchecked testimony. If I become suspicious of a job applicant’s claim to have a Ph.D., for instance, I can seek to check personally on the claim (in the sense of conducting the investigation on my own behalf) but this endeavour will invariably involve accepting the word of some people or institutions that the degree was or was not awarded. Moreover, the reliability of these people or institutions will often be either taken for granted or accepted on the basis of reputation, which is itself a complex form of hearsay, as are the documents and certificates the institutions may offer as evidence. Such often-ignored facts account for the way Hume and others persistently and unconsciously conflate individual and social readings of the term ‘observation’, a conflation that makes Hume’s project seem more manageable than it really is. When he refers to the need to base ‘our’ (that is, the sum of individual cognizers’) reliance upon testimony on ‘no other principle than our observation of the veracity of human testimony’ ([1748] 1975: 111), Hume is sliding between the two senses of ‘observation’. ‘We’ can observe the veracity of human testimony only if we already trust each other’s testimony sufficiently for the project to be no longer that of the autonomous knower.

Hume’s problems are symptomatic of the fact that the reductionist picture in which the autonomous knower figures so centrally is flawed at heart since the existence of a common language in which reports are made and accepted or rejected already carries with it a commitment to some degree of unmediated acceptance of testimony. The reductionist strategy necessarily supposes that we can fully understand what we are told in a public language and then independently investigate its truth or correctness, but the connections between linguistic comprehension and truth (or at least correct assertibility) are much closer than this picture allows. The reductionist project must concede that the autonomous knower might consistently understand correctly all that they were told but none the less find every report mistaken or deceitful. But any such ‘discovery’ would rather serve to refute the initial supposition that the reports had been correctly understood. Public languages are, after all, learnt and their use is subject to correction, so reports on usage and on some facts embodying that usage must have a degree of success that the reductionist ignores. Fixity of meaning depends upon some degree of successful reporting, though just what degree is a further question (Coady 1992: Ch. 9).

5 Further issues

Foundationalist theories of knowledge have recently been subjected to much scorn and criticism, a good deal of it

deserved, but one thing that seems right about the foundationalist emphasis is the idea that individual perception, memory and inference deserve a particular respect as sources of information in a way that speculation, guessing and suggestion (whatever their other cognitive value) do not (see Foundationalism). The challenge posed by testimony is that it seems to deserve a similar respect to perception, memory and inference, indeed, the respect we owe them may require us to accord it parallel honour. This is because testimony puts us in touch with the perceptions, memories and inferences of others, and a certain coherence between individual and communal resources underpins the individual's reliance on their own epistemic skills and achievements. It is precisely this status that the ideal of autonomous knowledge rejects.

Abandonment of the reductionist project does not mean that our general attitude to testimony should be one of gullibility. A comparison with memory makes the situation clearer: we cannot reduce our reliance upon memory to inferences from present perceptions but we also learn not to trust our memories blindly in all circumstances. The question of *how* trustworthy memories or testimonies are cannot be addressed by initially allowing that they might not be trustworthy at all, but this still allows that there are contexts and circumstances in which we should be wary of them. It also allows that there is room for serious investigation of the reliability of the testimonies of certain groups whose dependability is often questioned in legal and other contexts (for example, very young children, the seriously mentally ill, parties to an issue for adjudication), and the more general question of how reliable testimony might be in any given culture or context.

Throughout the above I have, in accordance with philosophical tradition, used 'testimony' as a broad term for certain sorts of telling, but it is an interesting and difficult task to give a serviceable definition of what these sorts are. The term is most at home in legal settings, but clearly a wider notion than that is required. One strategy is to define it as a type of speech act, but to allow extensions from the speech-act model for such things as maps, signposts and documentary records. There remain problems about whether to include competency and sincerity conditions in the definition and whether there is a sense of 'evidence' in which testimony by definition falls into the category of evidence. Some would see the former move as too demanding and the latter as question-begging.

Finally, the neglect of the topic by philosophers has left many important questions of law, history, psychology and religion to be debated without the benefit of concentrated philosophical input. These include such matters as the status of children's testimony at law, the problems posed by expert evidence, psychological claims about the unreliability of eye-witness identifications, and certain aspects of historical methodology, including the 'higher' biblical criticism.

See also: Social epistemology; Testimony in Indian philosophy

C.A.J. COADY

References and further reading

Anscombe, G.E.M. (1981) 'Hume and Julius Caesar', in *The Collected Philosophical Papers of G.E.M. Anscombe*, vol. 1, Oxford: Blackwell.(Strong criticism of Hume's views on the transmission of historical knowledge.)

Aquinas, T. (1257-9) *Commentary on Boethius's De Trinitate*, Toronto, Ont.: Pontifical Institute of Medieval Studies, 1958.(Aquinas has interesting reflections on testimony as a kind of natural faith, occupying a position midway between knowledge and opinion.)

Augustine (386-429) *The Works of Aurelius Augustinus*, ed. M. Dods, Edinburgh: T.&T. Clark, 1871-6.(Of all the philosophers of the ancient world, Augustine has the surest appreciation of the extent of our dependence upon the word of others and makes an interesting attempt to fit it into a Platonist epistemological framework. See especially, *De Trinitate*, *De Magistro*, *De Utilitate Credendi*, *Retractiones*.)

Bradley, F.H. (1935a) 'The Presuppositions of Critical History', in *Collected Essays*, vol. 1, Oxford: Oxford University Press.(Bradley's first publication, this is an essay on historical method which supports the new German school of historical criticism of the Bible, and seeks to undermine the credibility of historical testimony to what is 'non-analogous' to present experience.)

Bradley, F.H. (1935b) 'The Evidences of Spiritualism', in *Collected Essays*, vol. 2, Oxford: Oxford University Press.(After Hume, Bradley is one of the most significant writers on testimony. Here he tries to accommodate the importance of testimony by giving a reductive treatment of its epistemic significance, and attempting to show the limits of its value to 'critical' history. His attack on spiritualism extends his treatment in interesting ways.)

Burnyeat, M.F. (1980) ‘Socrates and the Jury: Paradoxes in Plato’s Distinction between Knowledge and True Belief’, *Proceedings of the Aristotelian Society* supplementary vol. 54: 173-91.(An important discussion of Plato’s epistemology, which includes an analysis of Plato’s approach to epistemology.)

Burnyeat, M.F. (1987) ‘Wittgenstein and Augustine De Magistro’, *Proceedings of the Aristotelian Society* supplementary vol. 61: 1-24.(Augustine’s views on teaching are explored and compared to Wittgenstein’s epistemological outlook. The discussion illuminates the significance of testimony for the theory of knowledge.)

Coady, C.A.J. (1992) *Testimony: A Philosophical Study*, Oxford: Oxford University Press.(Extended treatment of the topic that defends a non-reductionist approach to the epistemology of testimony.)

Coady, C.A.J. (1994) ‘Speaking of Ghosts’, in F.F. Schmitt (ed.) *Socializing Epistemology*, Lanham, MD and London: Rowman & Littlefield.(Discusses what seems to be a challenge to some of the methodology and conclusions of Coady’s 1992 work.)

Descartes, R. (1637) *Discourse on Method*, in *The Philosophical Works of Descartes*, vol. 1, trans. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge, Cambridge University Press, 1985.(Formative work on the theory of knowledge by one of the great figures of modern philosophy. Does not discuss testimony, but his outlook embodies the individualism that significantly influenced later theorists.)

Dummett, M.A.E. (1994) ‘Memory and Testimony’, in B.K. Matilal and A. Chakrabarti (eds) *Knowing from Words*, Dordrecht: Kluwer.(Exploration of the deep links between the epistemological status of testimony and of memory. Argues for an anti-reductionist stance on testimony.)

Fricker, E. (1987) ‘The Epistemology of Testimony’, *Proceedings of the Aristotelian Society* supplementary vol. 61: 57-83.(Argues for the superiority of justificationist over reliabilist accounts of the status of testimony, and treats testimony as a ‘secondary’ epistemic link compared with such ‘primary’ links as seeing.)

Fricker, E. (1995) ‘Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony: Critical Notice of C.A.J. Coady: *Testimony: A Philosophical Study*’, *Mind*, 104: 393-411.(Seeks to distinguish the merits of different kinds of reductionism and opposition to them in this area.)

Hardwig, J. (1985) ‘Epistemic Dependence’, *Journal of Philosophy* 82 (7): 335-49. (Argues that appeal to epistemic authority is an essential ingredient in much knowledge and that, because of this, rationality sometimes consists in refusing to think for oneself.)

Hardwig, J. (1991) ‘The Role of Trust in Knowledge’, *Journal of Philosophy* 88 (12): 693-708.(Emphasizes the significance of trust for the project of knowledge, and the related importance of ethics for epistemology.)

Hume, D. (1748) *An Enquiry Concerning Human Understanding*, in *Enquiries*, ed. P.H. Nidditch and L.A. Selby-Bigge, Oxford: Oxford University Press, 1975.(Although Hume does not have an extended treatment of testimony, his discussions of its epistemic status in the essay ‘On Miracles’ represent an acknowledgement of its practical importance combined with an influential attempt to make it theoretically derivative from more fundamental sources of information.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Oxford University Press, 1975.(Locke discusses problems about the assessment of particular testimonies and the apparent decay of the reliability of testimony through transmission chains in book 4, chapter 16.)

Mackie, J.L. (1969-70) ‘The Possibility of Innate Knowledge’, *Proceedings of the Aristotelian Society* 70: 245-57.(Discusses the problem of testimony and coins the phrase ‘autonomous knowledge’ to describe the ideal driving the reductionist about testimony.)

Matilal, B.K. and Chakrabarti, A. (1994) *Knowing from Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, Dordrecht: Kluwer.(Collection of papers discussing issues about testimony from Western and Indian perspectives. In addition to the essays by Dummett and Strawson listed separately here, there are, among others, chapters by Coady, Fricker, Lehrer, Sosa and Welbourne.)

Plato (4th century BC) *Theaetetus* and *Meno*, in *The Dialogues of Plato*, trans. and intro. B. Jowett, Oxford, Oxford University Press, 1953.(*Theaetetus* 201 and *Meno* 97a-b insist that testimony would not provide knowledge even if perception did.)

Popper, K. (1968) *Conjectures and Refutations: the Growth of Scientific Knowledge*, 2nd. edn, New York: Harper & Row.(Notable for viewing our reliance upon testimony as a reason for being opposed to foundationalism.)

Price, H.H. (1969) *Belief*, Lecture 5, New York: Humanities Press.(Interesting modern discussion of testimony that is fully sensitive to many of the central problems.)

Reid, T. (1983) *Philosophical Works*, with notes by Sir W. Hamilton, ed. H. Bracken, Hildesheim: George Olms Verlag.(Reid treats our regard for testimony as epistemologically basic because it stems from first principles of common sense.)

Ross, J.F. (1975) ‘Testimonial Evidence’, in *Analysis and Metaphysics: Essays in Honor of R.M. Chisholm*, ed. K. Lehrer, Dordrecht: Reidel.(Perceptive treatment, in the mode of Chisholm, of what it is to know on the basis of

testimony.)
Russell, B. (1948) *Human Knowledge: Its Scope and Limits*, New York: Allen & Unwin.(Here and elsewhere Russell shows awareness of the complexities of our dependence upon testimony, and he offers an analogical argument to deal with them.)
Schmitt, F.F. (1987) 'Justification, Sociality and Autonomy', *Synthèse* 73: 43-86. (Argues against the individualism implicit in traditional epistemologies and criticizes their treatment of testimony. Rejects the 'extreme' theory of Locke and the 'milder' theory of Hume, and explores an alternative view of intellectual autonomy to that inherent in the tradition.)
Schmitt, F.F. (1994) 'Socializing Epistemology: An Introduction through Two Sample Issues', in *Socializing Epistemology: the Social Dimensions of Knowledge*, Lanham, MD and London: Rowman & Littlefield.(Schmitt uses testimony as one example of the need for a new, social approach to epistemology.)
Strawson, P.F. (1994) 'Knowing from Words', in B.K. Matilal and A. Chakrabarti (eds) *Knowing from Words,* Dordrecht: Kluwer.(A brief critique of reductionist approaches to testimony.)
Welbourne, M. (1986) *The Community of Knowledge*, Aberdeen: Aberdeen University Press.(Short but interesting book that argues for a conception of knowledge as 'commonable' and is strongly anti-reductionist about testimony. Also argues that knowledge is conceptually prior to belief and cannot be defined in terms of it.)

Testimony in Indian philosophy

A prominent topic in Indian epistemology is śābdapramāṇa, *knowledge derived from linguistic utterance or testimony. The classical material is extensive and varied, initially concerned with providing grounds for accepting the wisdom of* śruti *or 'the heard word', that is, the canonical scriptures. The Buddhists, however, saw no need for* śābdajñāna *(information gained through words) as an independent source of knowledge, because any utterance (including the Buddha's) that has not been tested in one's own experience cannot be relied upon; and in any case, the operation of such knowledge can be accounted for in terms of inference and perception.*

The Nyāya, following the Mīmāṃsā, developed sophisticated analyses and a spirited defence of the viability and autonomy of testimony. The problem is recast thus: is śābdapramāṇa *linguistic knowledge* eo ipso, *or does verbal understanding amount to knowledge only when certain specifiable conditions, in addition to the generating conditions, are satisfied? The more usual answer is that where the speaker is reliable and sincere, and there is no evidence to the contrary, the generating semantic and phenomenological conditions suffice to deliver valid knowledge. If doubt arises, then other resources can be utilized for checking the truth or falsity of the understanding, or the reliability of the author (or nonpersonal source), and for overcoming the defects.*

1 The problem of *śābda*

Ever since some traditional philosophers made the intriguing claim that the Vedas (canonical Brahmanical texts) are an inviolable authority on all manner of things, debate has raged in Indian thought over whether *śābda*, or information derived from words, can rightfully be accepted as an autonomous mode of knowing. In other words, how can hearing a sentence under appropriate conditions serve as a means of acquiring knowledge? Philosophers in the tradition recognized that a theory is needed that gives an account of both the instrumental conditions by which words or linguistic utterances convey appropriate understanding and the epistemological grounds for adjudicating the validity of the understanding or cognition so derived. Indian schools of philosophy have responded in differing ways to this challenge.

In the general accounts of knowledge developed in India, a properly accredited source of knowledge (*pramāṇa*) is thought to be fundamental to the production of knowledge (see Knowledge, Indian views of). What could such a source be in the case of *śābdapramāṇa*? Obviously, but nontrivially, words themselves. This response indicates a remarkable recognition of the deep connection between words and knowledge, beyond the usual preoccupations with words and objects, names and things, and even knowing words. It does not take much argument to point out that a vast body of knowledge and beliefs is derived from sources other than perception and inference, namely, from the words of others, notably through hearing (direct conversation, rumour, radio and television) or reading (books, newspapers, documents, lexicons, inscriptions), not to mention the role of testimony under oath ('nothing but the truth') submitted in legal proceedings and courts of law. It should be noted that this account is not committed to suggesting that words exhaust or constitute the sole set of conditions that deliver the knowledge, or that the delivery is unproblematic in every instance. A number of factors are involved, and words arguably happen to be the primary causal antecedent, without which the delivery would flounder. That much is uncontroversial. The extent of the input from or reliance on other *pramāṇas*, including further testimony, remains to be established; much depends on the nature of the individual inquiry and the strength of conviction expected. Seen in this light, the question of the autonomy of testimony takes on a somewhat different significance, as does that of the reductivity of *śābdapramāṇa*. We shall return to this point.

2 Sentential meaning

The primary causal antecedent of knowledge through testimony is not just any string of words but a sentence (*vākya*). (This is also the basic unit of linguistic understanding for grammarians.) A sentence is made up of words, and sentential meaning is made up of word-meanings. A word (for example, 'cow') signifies either a universal (cowness) or a particular (dewlap-thing) possessing that universal, or it denotes a general class (cows), or perhaps even names an entity (Gokula). There is much discussion as to whether it is words or their meanings, whether singly or together, or the indivisible sentential structure, that convey the meaning of a sentence (*vākyārtha*) to a hearer. And the corollary issue of whether words achieve their signification individually, or by piecemeal compounding, or only by being related to other parts in the whole context of a sentence, is equally rehearsed, with

various compromise positions. However, the main issue boils down to this: what does a sentence succeed in conveying, under appropriate conditions? It is clearly the (sentential) meaning that is grasped. But 'meaning' here is not exactly what linguists and grammarians are usually concerned with; it is more than the structure of the cognitive state constructed from the grasped words. *Vākyārtha* is literally the object out there, a compound object whose components are the component word-meanings. For example, grasping the meaning of the expression 'The lotus (is) blue', regardless of whether the object spoken of is present or not, amounts to cognizing a substratum, possessing lotusness, characterized by the quality blueness (the substratum that registers in awareness is represented by an oblique 'this': the awareness that there is something, *X*, qualified by blueness). Let us for now call this special cognition *śābdabodha* (standing in for *śābdajñāna* - awareness of the linguistic components or structure - and *anvayabodha* - comprehension of sense, or linguistic understanding), and render this as 'linguistic knowing'.

But what exactly is the meaning in respect of - reference or sense, or some other kind of designation? Some writers have argued that Indian philosophy does not have a clear notion of sense (as an abstracted conceptual representation or phenomenological 'thought') either because the sense-reference distinction is altogether lacking, or because some theory of reference has been dominant and sense has been subsumed as merely a mode of presentation consistent with the theory of reference (Mohanty 1992). Thus it falls short of allowing a sentence to achieve signification of any kind other than a direct denotation of the objective state that is its referent; this is construed as a complex relational object (in Nyāya) or related designation (in Prābhākara Mīmāṃsā). K.C. Bhattacharyya (1956) put it succinctly thus: 'The word directly refers to the thing, expresses the thing, touches it in a sense… The sentence at once refers to an objective relation.' (Clearly it is neither belief nor proposition that is of concern here, for the epistemic content of the word-derived cognition is more primitive than presupposed by either of these two pervasive categories of analysis.)

What is grasped in the cognitive state - that is, in *śābdabodha* under appropriate conditions - is a structure with various qualifying items that bear direct reference to objects thereby known. It follows that if in the process of grasping an expression some object fails to be clearly registered in the cognitive state (*anubhava*) or conflicts with another item in the structure, then a *śābdabodha* properly speaking cannot arise. The epistemic significance of this rider cannot be overstated, as it is connected with the fulfilment of the often-mentioned appropriate or specified conditions (*kāraṇas*). These conditions mark the uniqueness and therefore autonomy of *śābdabodha* in a way that is not available to perception and inference.

3 The conditions of understanding

Just as a sentence is not just any string of words, so sentential meaning is not conveyed unless certain conditions obtain. These conditions pertain to the relation, connection, integration and presentation of the words and their respective meanings or senses, and provide checks against vitiating factors, thus ensuring an unhindered and 'valid' grasp of sentential meaning. There are four conditions, whose relative degrees of importance depend on the school one follows. The first is contiguity (*āsatti*), which requires that words should follow each other in contiguous or proximate relation, for too large a spatial or temporal distance between words can be a linguistic hazard. The second condition is expectancy (*ākāōkṣā*). When a word is uttered it arouses the expectation of another word that will appropriately complete the syntactical relation. When a speaker utters the words 'the door' the hearer expects another word (or series of words), such as 'close' or 'paint', to connect a pending meaning so as to complete the *śābdabodha* set in motion. The third condition is intentionality (*tātparya*), which refers to the speaker's intention or the intended purport of the expression. This is needed especially to disambiguate petrified meanings where the speaker desires a particular meaning to be conveyed, and also to carry the meaning of an expression whose speaker is either unavailable or nonexistent beyond the didactic (or literal) meaning (as in the case of the allegedly authorless scriptures and impersonal ordinances). The fourth and probably most controversial condition is semantic compatibility (*yogyatā*).

Yogyatā or semantic 'fitness' is related to *ākāōkṣā*, as what is to follow the anticipation aroused should be a meaning which appropriately and not simply contingently completes the objective reference. In other words, it would be a source of frustration if after hearing someone utter the words 'the door', the hearer were to be presented with the words 'fuels the sun's fire'. With respect to whole utterances, the condition of *yogyatā* would rule out *śābdabodha* from expressions such as 'He wets with fire', since 'wetting' and 'fire' are mutually

incompatible. On the strict Nyāya interpretation, there is no possibility of linguistic knowing arising here because there is incompatibility between the items being conjoined. The fitness expected is as much between the things signified (related in the reference) as it is with the signifying terms. With one part of the reference left as it were 'unsaturated' the truth-value presented is incomplete. On this interpretation, a sentence '*a* is *F* is characterized by fitness (*yogyatā*) only if *a* is *F*. It follows that if a sentence is false, it cannot generate *śābdabodha*. Only true sentences, on this interpretation, are fit and so can generate *śābdabodha*. Not only contradictory sentences such as 'The fireless hill has fire', but also syntactically consistent but semantically absurd sentences such as 'Idleness is green' and even contingently false sentences such as 'It is raining' (when it in fact is not) - all these, strictly speaking, cannot generate *śābdabodha*, for there are no objective relational referents that answer to the disparate component-meanings.

The virtue of this argument is that it restricts the scope of *śābdapramāṇa* or testimony by stipulating a criterion that raises the epistemic value of the word-derived cognition so that not every expression can present itself as a fit candidate. And this criterion is built into the linguistic operation itself. But this also implies that if a sentence is not true, it lacks fitness and will fail to generate linguistic knowing. Two questions arise: (1) Does this mean that in the case of a false sentence no linguistic understanding at all arises? (2) Is the truth of a sentence always established or given a priori when the sentence is being understood? With respect to (1), it is urged that even with a patently false sentence (such as 'The sun is a ball of golden jelly') one does understand something from it; one at least knows that it is false. Why should this be any different from erroneous perceptions, which are still admitted to be 'perceptions' even though hopelessly false? Why can there not simply be 'false linguistic cognitions'? But even to know that a sentence is false, one has to understand something about the complex content of the sentence, in a way not dissimilar to understanding the complex relational object of a true sentence.

This recognition of a further, if weaker, cognitive process requires that we make a distinction between knowing and understanding. But the line of analysis followed so far has almost precluded the admission of such a distinction. This is due in large measure to the overwhelming influence of the Nyāya approach to the whole problem of language, which, with its unstinted commitment to a naïve realist ontology, has tended to eclipse more sanguine and less rigid analyses which not only admit of such a distinction but consider the function of understanding to be indispensable and necessary to the process. Prābhākara Mīmāṃsā, the Vedānta, the Jaina and the grammarians would side with the latter position. It is generally overlooked that Mīmāṃsābegins its theory of meaning with the thesis of language being eternal (*autpattika*), which, like Saussurean semiology, posits the inseparability of signifier (sound-form) and signified (mental/conceptual image) in a word (see MīmāṃsāČsect;3; Saussure, F. de §2). The composite designation derived from the component meanings need refer to no object(s) in the world as such, or can be simply a mode of presentation, and so on. Often, in the transmission of teachings, the pupil may acquire very directly images or complex ideas which have no objective correlate answering to the 'related designation' (see §2), which would be therefore better rendered as 'sense'. The inquiry into duty (*dharma*), the highest good and supplementary ethical notions, injunctional precepts, performatives (involving mantras), various illocutionary and perlocutionary speech-acts, poetic and mimetic evocations, and so on, involve an endeavour to generate verbal associations of this sort; the question of their truth and falsity, and of their being knowledge, can hardly be said to arise before their full significance and purport (teleological, soteriological or transcendental) is assimilated into the larger body of one's knowledge or beliefs. If the distinction is admitted, and understanding is permitted before the question of knowing is raised, then the thesis that identifies or collapses semantic fitness with truth-value is effectively shattered. This takes us to the second question.

4 Truth and falsity

Do all sentences come ready-stamped with the mark of truth or falsity? The idea that they might is not only counterintuitive, but an opponent might well urge that one hears a sentence and then, for the sake of conjecture and adventure, proceeds to determine whether it is true or false. From the hearer's subjective position, there is no other sense in which one can speak of the sentence being true or false moments prior to it being heard, without begging the question. (The speaker being aware or believing that the sentence uttered is false is of no relevance as this information is not available to the hearer, and besides, the sentence may well turn out to be true.)

Furthermore, a hearer may defer to the reliability of a speaker, trusting that the speaker is and has always been sincere and reliable. There may be no reason to doubt the reliability of the speaker and the hearer may take the

speaker's authority to be a reason for attending to the utterance in the first place. And suppose that the hearer wishes to know about certain things to which it is not possible to have access through perception and inference. Would it be an act of epistemological violence if the hearer decided to accept what they heard from the testifier, but perhaps subsequently made detailed investigations to either prove or disprove what they thus came to 'know'? When no further evidence either way is forthcoming - notably from perception and inference - one either reviews one's reliance on the authority itself, or one turns to logical and metaphysical arguments, as philosophers frequently do, to assuage growing doubts about the truth of one's forebears' utterances.

Finally, on the question of the non-reductiveness of *śābdapramāṇa*, some general points only can be noted here. The first issue is whether the requisite linguistic conditions are sufficient, in and of themselves, to mediate and deliver the kind of knowledge claimed for language. We observed earlier a resistance to collapsing semantic fitness into the requirement for truth so that it is not always possible to question whether what one understands is true or not. Even the scholastic Prābhākara Mīmāṃsā, for whom all cognitions arising from the appropriate use of the accepted *pramāṇas* are intrinsically true - that is, truth is given by virtue of the cognition being a knowing-state - will concede that errors do creep in, that one could be mistaken about what one takes to be true, especially with regard to nonperceptibles, and that such falsified knowledge has to be set aside.

Ironically for Nyāya, it is a general principle that truth is never self-evidentially given, for whether anything is a source of knowledge is a matter to be established *ab extra* (*parataḥ*), by certain 'marks of excellence' of evidence, and further corroboration that may involve pragmatic testing. One may even attempt to falsify a cognition to assess its tenacity. But if this is extended to *śābdabodha*, then it could be argued that testimony comes to depend on other resources for its validation and therefore loses its claim to be *sui generis*. Paradoxically, if such resources are not available in the case of nonperceptibles, then the risk of allowing certain linguistic understandings to stand without validation could become onerous; and deferring to further testimonies or to consensus on their reliability would only serve to perpetuate the process.

See also: Language, Indian theories of; Meaning, Indian theories of

PURUSHOTTAMA BILIMORIA

References and further readings

Bhattacharyya, K.C. (1956) *Studies in Philosophy*, ed. G. Bhattacharyya, vol. 1, ch. 2, Calcutta: Progress Publishers; repr. Delhi: Motilal Banarsidass, 1983, page 83-.(Bhattacharyya pioneered a uniquely Indian genre of Neo-Kantian realism and phenomenological style in the 1920s, working out of Calcutta, where he had no contact with philosophers in the West.)

Bilimoria, P. (1988) *Śābdapramāṇa: Word and Knowledge, A Doctrine in Nyāya-MīmāṃsāPhilosophy*, Dordrecht: Reidel/Kluwer.(The first comprehensive treatment and analytical exposition in English of testimony in Indian philosophy. Based on classical material, it makes relevant connections with Western theories of knowledge and philosophy of language.)

Bilimoria, P. (1993) '*Pramāṇa* Epistemology: Some Recent Developments', in G. Floistad (ed.) *Contemporary Philosophy: A New Survey*, Asian Philosophy 7, Dordrecht: Kluwer.(This essay surveys current developments in Indian epistemology pertaining to debates over language, knowledge and truth.)

Bilimoria, P. (1995) 'Authorless Voice, Tradition and Authority in Mīmāṃsā', *Nagoya Studies in Indian Culture and Buddhism* 16: 137-60.(The paper analyses the locution of 'authorless texts' in Mīmāṃsāexegetical hermeneutics, via the framework of Saussurean semiology, which is also criticized here.)

Hayes, R. (1988) *Dignāga on the Interpretation of Signs*, Dordrecht: Reidel/Kluwer.(Hayes presents an elegant treatment of Dignāga's theory of signs and language, critically reviewing non-Buddhist ideas on semiology and interpretation.)

Matilal, B.K. (1990) *The Word and the World: India's Contribution to the Study of Language*, Oxford: Oxford University Press.(The last book to be published before the author succumbed to cancer in Oxford is an eclectic set of reflections on the problem of language and knowledge across Indian thought, from Nyāya-Bhartṛhari holism to Derridean deconstructionism.)

Matilal, B.K. and Chakrabarti, A. (eds) (1994) *Knowing from Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, Dordrecht: Kluwer.(This is a useful collection of essays from philosophers in India and the West (including Strawson and Coady) on the variety of ways in which the question of testimony, its viability in terms of linguistic and epistemological process-content, can be approached in both the 'pure' and 'comparative' understandings.)

Matilal, B.K. and Evans, R. (eds) (1986) *Buddhist Logic and Epistemology*, Dordrecht: Reidel/Kluwer.(This is a useful collection of essays on the development of logic and its relations with the problem of knowing in Buddhist philosophy. The more difficult and technical areas of Nāgārjunian dialectic also receive lucid treatment here.)

Mohanty, J.N. (1992) *Reason and Tradition in Indian Thought: An Essay on the Nature of Indian Philosophical Thinking*, Oxford: Clarendon Press.(This is about the most authoritative philosophical work to date on reason, rationality and responses to the question of tradition in Indian thought, and is written in a highly accessible style combining analytical and phenomenological idiom.)

Mohanty, J.N. (1993) *Essays in Indian Philosophy Traditional and Modern*, ed. with intro. by P. Bilimoria, New Delhi/York: Oxford University Press; repr. 1995.(A collection of rare essays spanning a thirty-year period. Issues in metaphysics and social philosophy are covered, and the comparative standing of key doctrines, including testimony, in Indian and European philosophy is examined.)

Siderits, M. (1991) *Indian Philosophy of Language: Studies in Selected Issues*, Dordrecht: Kluwer.(A short, highly technical work treating the key issues in Indian philosophy of language. It picks up on recent linguistic developments and major paradigm debates in the analysis of language in modern philosophy, but focuses for its responses on Buddhist critiques of Nyāya and Mīmāṃsātheories.)

Staal, F. (1989) *Rules without Meaning: Rituals, Mantras and the Human Sciences*, New York: Peter Lang.(Staal explores a thesis of meaning which goes beyond semantic and purely linguistic components by shifting the analysis to the ritual structures and perfomative context of utterances (such as *mantras*) in highly sophisticated texts exemplified in the Vedas.)

Tetens, Johann Nicolaus (1736-1807)

Tetens was a German philosopher, mathematician and physicist, with a second career as a Danish government official, who was active in Northern Germany and Denmark during the second half of the eighteenth century. Together with Johann Heinrich Lambert and Moses Mendelssohn, Tetens forms the transition from the German school philosophy of Leibniz, Wolff and Crusius to the new, critical philosophy of Kant. Tetens' philosophical work reflects the combined influence of contemporary German, British and French philosophical currents. His main contribution to philosophy is a detailed descriptive account of the principal operations of the human mind that combines psychological, epistemological and metaphysical considerations. While showing a strong empiricist leaning, Tetens rejected the associationist and materialist accounts of the mind, favoured in Britain and France, and insisted on the active, spontaneous role of the mind in the formation and processing of mental contents.

1 Life

Tetens was born in Tetenbüll (Schleswig, Northern Germany) - or, on another account, in 1738 in Tönning (Schleswig) - and died in Copenhagen. From 1755 until 1758 he studied mathematics, physics and philosophy at the universities of Copenhagen and Rostock. In 1759 he earned the degree of *magister* and began teaching at Rostock, where he defended his dissertation in 1760. From 1760 until 1776 he taught physics and philosophy at the newly founded Bützow Academy. In 1765 he also assumed the directorship of the local *Gymnasium*. From 1776 until 1789 he held a professorship at the University of Kiel, first in philosophy, and later also in mathematics. In 1789 he entered public service and had a distinguished career in the Danish Ministry of Finance.

Tetens' main philosophical works were written in the 1770s under the influence of Leibniz's posthumously published *Nouveaux Essais* (1765) and Kant's pre-critical writings. Tetens also made contributions to applied mathematics, meteorology, pedagogy and environmental engineering. With his wide range of research activities and his dual career in academia and government, Tetens was one of the few truly universal scholars in the tradition of Leibniz. His philosophical work places him among the very best in German philosophy between Leibniz and Kant. In several respects he is a direct precursor of Kant's critical philosophy.

2 Metaphysics

Tetens' philosophical work is informed by a sweeping knowledge of the main philosophical movements of his time. He was well acquainted with the Leibnizian-Wolffian school of philosophy and its pietist critics, was conversant with the Scottish Common Sense philosophers, knew the French materialist and sensualist philosophers, and was one of the few philosophers in Germany to read Locke and Hume before their works became available in translation (see Wolff, C.; Pietism; Common Sense School). In his eclectic stance towards those philosophical currents, Tetens resembled the position of the Berlin Academy and German *Popularphilosophie* in general. Yet unlike the main representatives of that movement, he held on to the German school-philosophical project of a 'first philosophy' or metaphysics. Like other proponents of metaphysics in eighteenth-century Germany, Tetens responded to the sceptical and materialist challenges to metaphysics with an investigation into the very possibility of metaphysical knowledge. In Tetens this methodological and epistemological reorientention of metaphysics has a decidedly psychological character; metaphysical concepts and the principles into which they enter are traced to their origin in the general operations of the human mind.

Tetens' guiding interest in the reform of metaphysics shows most clearly in his first significant work, *Über die allgemeine speculativische Philosophie (On General Speculative Philosophy)* of 1775. He argues for a truly general and comprehensive foundational science that is indifferent to the specifics of the two principal object domains of metaphysics - spiritual and corporeal reality. The science in question would not have specific objects as its domain but only the most general concepts and principles that pertain to being as such. Tetens' term for this discipline, that transcends all knowledge of specific objects, is 'transcendent philosophy'. According to Tetens, previous metaphysics suffers from a confusion of what is genuinely transcendent with the features of either corporeal or spiritual reality.

Tetens' chief concern is to ensure that the transcendent concepts and principles are not merely figments of the mind but have reality and pertain to objects outside the mind entertaining them. According to Tetens, the

validation of such metaphysical claims proceeds by appealing to psychological facts to be discovered through observation of the workings of one's own mind. Transcendent principles, such as the Principle of Sufficient Reason, are validated by the feeling of necessity that accompanies consideration of them. The reality of general concepts, including such transcendent concepts as causality and substance, is established by showing their material basis in sensations. For Tetens, there is no sceptical wedge to be inserted between our best evidence about the world and the way the world really is. Psychological necessity carries logical and ontological weight. He seeks to respond to Hume's doubt about the reality of causal connection by making a distinction between contingent mental habits and necessary manners of thinking. Yet while assuming the reality of universal forms of thinking, Tetens does not go on to draw ontological consequences from this fact, in the manner of Kant's distinction between things in themselves and appearances. Tetens came to see that before transcendent philosophy can be applied to objects outside the understanding, its concepts and principles have to be derived through a comprehensive philosophical account of human understanding.

3 Philosophical psychology

Tetens undertook such a detailed examination of the human mind in his main philosophical work, the *Philosophische Versuche über die menschliche Natur und ihre Entwickelung (Philosophical Essays on Human Nature and its Development)*, which was published in two volumes in 1777. In the first ten of the work's fourteen essays, Tetens follows Locke's descriptive method and provides a detailed classification of the mental powers and their modes of operation. The emphasis is on the power of cognition. Tetens here exceeds the confines of contemporary empirical psychology through the ontological and epistemological scope of his psychological considerations. The last four essays take up topics in rational psychology by discussing the nature of the soul with regard to its basic power, its freedom, its relation to the body and its perfectibility. The work often digresses and the relation between the various classifications of mental powers offered is not always clear. Yet the *Philosophische Versuche* contains a number of original contributions to the study of the human mind and constitutes a significant step in the prehistory of scientific psychology and cognitive science.

Tetens distinguishes three principal forces that make up the human power of cognition: feeling, representing and thinking. Feeling is the initial, merely passive stage of cognition. It consists in the mind's modificability or receptivity to sensations, but also includes the mind's passive ability of noticing its own states. Representing involves both phantasy, responsible for the associative grouping of representations, and imagination, through which new representations are formed. Thinking consists in the detection of relations among the mental contents presented through feeling and representation. The relations are articulated through concepts and judgments. Tetens provides a psychological derivation of the general relational concepts, such as identity and diversity or simultaneity and succession. He considers those general concepts and the judgments into which they enter sufficiently validated through the psychological necessity with which they present themselves in the mind.

Tetens stresses the active, spontaneous nature of representing and thinking. Even perception is said to involve mental activity beyond the reception of sensation. He considers whether there might be a fundamental principle underlying the three basic powers of cognition, but concedes the hypothetical nature of his speculations on a primitive mental force that would be both feeling and active.

When surveying the overall organization of the human mind with an eye to its most basic powers, Tetens separates the active cognitive powers from passive feeling. Representing and thinking now make up the understanding, while feeling is taken in a broader sense, encompassing the mind's affective response to its modifications through the sensations of pleasure and pain. To understanding and feeling Tetens adds volition as the third basic mental power, thus providing the immediate precedent for Kant's threefold division of the human mind into the faculty of cognition, the feeling of pleasure and displeasure, and the faculty of desire.

Tetens' inquiry into the basic force of the human mind as a whole leads him to broader considerations regarding the basic character of humanity, which he identifies as extraordinary modificability and disposition to spontaneous activity. This double characterization combines the recognition that physical and mental development is highly susceptible to external circumstances with the recognition that human mental life develops on the basis of internal forces.

While Tetens' contributions to the reform of metaphysics were soon overshadowed by Kant's *Critique of Pure*

Reason, the psychological approach to philosophical issues practised by Tetens found a number of successors in post-Kantian philosophy from Fries and Herbart to Laas and Brentano. In fact, Kant's distinction between a psychological, descriptive, and a nonpsychological, justificatory, dimension of knowledge, which had been developed mainly in response to Tetens, did not achieve canonical status until the neo-Kantianism of the late nineteenth century and the advent of anti-psychologism in logic - only to be questioned again in twentieth-century efforts to naturalize epistemology and ontology (see Neo-Kantianism).

See also: Lambert, J.H.; Mendelssohn, M.; Transcendental arguments

GÜNTER ZÖLLER

List of works

Tetens, J.N. (1775) *Über die allgemeine speculativische Philosophie (On General Speculative Philosophy)*, Bützow and Wismar; ed. W. Uebele, Berlin: Reuther & Reichard, 1913.(The Uebele edition also reprints the *Philosophische Versuche*.)

Tetens, J.N. (1777) *Philosophische Versuche über die menschliche Natur und ihre Entwickelung (Philosophical Essays on Human Nature and its Development)*, Leipzig, 2 vols; repr. in *Die philosophischen Werke*, vols 1-2, Hildesheim: Olms, 1979; vol. 1 repr. in W. Uebele (ed.), Berlin: Reuther & Reichard, 1913.(The Uebele edition also reprints *Über die allgemeine speculativische Philosophie*.)

Tetens, J.N. (1971) *Sprachphilosophische Versuche (Experiments in Philosophy of Language)*, ed. H. Pfannkuch, Hamburg: Felix Meiner.(Introduction by E. Heintel. This collection contains *Über die Grundsätze und den Nutzen der Etymologie (On the Principles and Advantages of Etymology*, 1765-6); *Über den Ursprung der Sprachen und der Schrift (On the Origin of Languages and Writing*, 1772); and selections from Essay 11 of the *Philosophische Versuche*. Includes a bibliography.)

References and further reading

Barnouw, J. (1979) 'The Philosophical Achievement and Historical Significance of Johann Nicolaus Tetens', *Studies in Eighteenth Century Culture* 9: 301-35.(De-emphasizes Tetens' relation to Kant.)

Baumgarten, H.-U. (1992) *Kant und Tetens. Untersuchungen zum Problem von Vorstellung und Gegenstand (Kant and Tetens. Investigations into the Problem of Representation and Object)*, Stuttgart: Metzler & Poeschel.(Emphasizes Tetens' relation to Kant.)

Beck, L.W. (1969) *Early German Philosophy: Kant and his Predecessors*, Cambridge, MA: Harvard University Press, 412-25.(Contrasts Tetens with Kant.)

Ciafardone, R. (1983) *Johann Nicolaus Tetens. Saggi filosofici e scritti minori*, L'Aquila.(A comprehensive monograph on Tetens.)

Dessoir, M. (1892) 'Des Nic. Tetens Stellung in der Geschichte der Philosophie' (On Tetens' Place in the History of Philosophy), *Vierteljahresschrift für wissenschaftliche Philosophie* 16: 335-68.(A concise account of Tetens' position in the history of philosophy.)

Gelder, B. de (1975) 'Kant en Tetens', *Tijdschrift voor filosofi* 37: 226-60.(A comparison of Kant and Tetens.)

Hauser, C. (1994) *Selbstbewußtsein und personale Identität. Positionen und Aporien ihrer vorkantischen Geschichte. Locke, Leibniz, Hume und Tetens (Self-consciousness and Personal Identity. Positions and Aporia of their Pre-Kantian History. Locke, Leibniz, Hume and Tetens)*, Stuttgart-Bad Cannstatt: Frommann-Holzboog, 124-51.(An account of Tetens' contributions to eighteenth-century accounts of self-consciousness and personal identity; includes bibliography.)

Kuehn, M. (1987) *Scottish Common-sense in Germany*, Kingston, NY, and Montreal: McGill University Press, 119-40.(An account of Tetens' relation to Thomas Reid, James Oswald and James Beattie.)

Störring, G. (1901) *Die Erkenntnistheorie von Tetens (Tetens' Epistemology)*, Leipzig.(An account of Tetens' theory of knowledge.)

Uebele, W. (1911) *Johann Nicolaus Tetens nach seiner Gesamtentwicklung betrachtet, mit besonderer Berücksichtigung des Verhältnisses zu Kant (Johann Nicolaus Tetens and his Development Considered, with Special Consideration of the Relationship to Kant)*, Berlin: Reuther & Reichard.(Emphasises the relationship between the philosophy of Tetens and that of Kant.)

Thales (*fl. c.*585 BC)

Known as the first Greek philosopher, Thales initiated a way of understanding the world that was based on reason and nature rather than tradition and mythology. He held that water is in some sense the basic material, that all things are full of gods and (purportedly) that all things possess soul. He predicted an eclipse of the sun and was considered the founder of Greek astronomy and mathematics.

1 Life and legacy

The first Presocratic philosopher and reputed founder of Western philosophy and science, Thales lived in the early sixth century BC in Miletus, a prosperous Ionian Greek city on the seaboard of Asia Minor. At Miletus he was followed in turn by Anaximander and Anaximenes, the other members of the so-called Milesian School. These two refined his rational approach to the study of nature and extended its application, leaving an intellectual legacy that was decisive for later Presocratic philosophers and, through them, for Plato, Aristotle and the Hellenistic philosophical schools, and consequently for the rest of the Western cultural tradition.

It is doubtful whether Thales wrote anything. Although ancient sources ascribe several works to him, no writing of his was available to Aristotle. Our knowledge depends entirely on the reports of others, many of which are inaccurate and even fanciful, thanks to the legendary status Thales achieved. Thales was the only Presocratic philosopher numbered among the Seven Sages, and this honour is probably due to his reputation for practical advice rather than for his scientific and philosophical work. He is said to have told the Ionian cities to form a political union with a common, centrally located governing council, the better to resist Persian aggression (Herodotus, I 170, A4) and to have advised the Lydian king Croesus how to ford the Halys River (I 75, A6). He was most famous for predicting a solar eclipse that took place on 28 May 585 BC (I 74, A5).

2 Philosophy

Thales is widely believed to have held that all things are made of water. This interpretation goes back at least to Aristotle (*Metaphysics* I 3, A12), who attributes it to Thales in summarizing the views of his predecessors which he regards as pertinent to his own investigation of *aitiai* (causes or explanatory factors). Thales, like ‘others who first pursued philosophy’, held that ‘the only principles (*archai*) of all things are principles in the form of matter’. He chose water as the unique material principle, in the sense that everything originally evolves from water and ultimately perishes into water and, more strongly, that all things are composed of water, which persists in them although its attributes change. Thus Thales is the first Ionian material monist and the founder of physical science (see Archē).

How accurately Aristotle’s interpretation represents Thales’ views is open to question, for it dresses Thales in Aristotelian language and concepts and presupposes that he was interested in the very problems that excited Aristotle two centuries later. Furthermore, it is possible that Aristotle’s account (‘Thales, the originator of this kind of philosophy, says it is water; this is why he declared that the earth rests on water’ (*Metaphysics* I 3, A12)) merely records Aristotle’s own inference that Thales held water to be the material principle on the basis of the view, which he confidently attributes to Thales, that the earth rests on water. This possibility, together with the weakness of the inference, makes it less likely that Thales was a material monist in Aristotle’s sense.

On an alternative theory, also based on ancient evidence, Thales borrowed the belief that the world floats on water from Egyptian and/or Babylonian creation myths such as the *Enuma Elish* (see Egyptian philosophy, influence on Ancient Greek thought §1). This view is not implausible since Thales reportedly visited Egypt and Miletus had commercial connections with Egypt and the Near East. According to these creation myths, the earth or land arose in the middle of primeval waters through the agency of one or more gods. Thales may thus have borrowed also the idea that water is the origin of the world, conceiving of it not as an Aristotelian material principle, but as the ultimate ancestor of all things. It may be significant that *archē*, Aristotle’s word for ‘principle’, also means ‘origin’ or ‘beginning’.

Even on this account Thales did more than borrow; he took a decisive step away from the traditional and mythological towards the philosophical and scientific by demythologizing the views he adopted. Not that he was an atheist; he and his successors believed strongly in the divine and its importance in the universe. However, they

abandoned the traditional Greek tendency to explain events through the Olympian gods. Where tradition had attributed earthquakes to Poseidon, Thales explained them in terms of the motion of the water that supports the earth (Seneca, *Natural Questions* III 14, A15). This way of thinking is more apparent in the theories of later Presocratics than in what we know of Thales himself, but their agreement on this approach was doubtless one of Aristotle's reasons for calling Thales the originator of materialism and for distinguishing him from purveyors of mythological accounts of creation.

Thales' views on divinity and on soul are attested as follows. 'He said that the magnet possesses soul (*psychē*) because it moves iron' (Aristotle, *On the Soul* 405a20, A22) (see Psychē). 'Some say that soul permeates the universe. This may have been Thales' reason for thinking that all things are full of gods' (*On the Soul* 411a7, A22). These reports have led some to describe Thales' world view as panpsychism and pantheism, but such labels by themselves contribute little to understanding a philosopher's thought. In the first place, the notion of soul needs clarification. Briefly and dogmatically, on the Greek view of soul which Aristotle develops and for which he finds evidence in Thales, the presence of soul is what makes something alive; non-living things have no soul, and dead things no longer have one. Signs of life include the capacities to move and change, and to cause other things to move and change. In Thales' case we have three claims: (1) the magnet moves iron, therefore it possesses soul; (2) soul permeates the universe; and (3) all things are full of gods. Only (1) and (3) are explicitly attributed to Thales; (2) is conjectured to be his reason for holding (3), and no connection is asserted to hold between (1) and (3). Now (3) does not follow from (2) without further premises, including one that amounts to: (4) wherever there is soul, there are gods. If Thales held this view (a hypothesis that is historically plausible, if unprovable), then (1) becomes relevant, since it establishes that soul occurs in more things than people suppose, and hence provides support for (2). Even so, the argument falls short. Even granted that anything that makes other things move is alive, we need not suppose that everything whatsoever is alive. Still, this could be how Thales reasoned, in view of the Presocratic tendency to make generalizations on the basis of slender evidence.

This reconstruction of Thales' reasoning is tentative and optimistic, for (1) and (3) are the only views ascribed to him, and he could have held them independently of one another. However, his views on the role of water in the universe support the connection of ideas suggested above. Briefly and dogmatically again, the Greeks commonly held that anything was divine that was immortal and had the power to affect things independently of human will. On the second account of water presented above, water could be considered as divine and (for reasons given above) also as possessing soul. Things made out of water, or ultimately descended from water - that is, *all* things - thus have a claim to possessing soul and being divine, although these properties are more or less obvious in different things (just as some things are wetter than others).

3 Mathematics and astronomy

As Aristotle made Thales the originator of philosophy, so Aristotle's student Eudemus, who wrote histories of geometry and astronomy, made him the first Greek mathematician and astronomer. Here too the nature of his originality and contributions is uncertain. Reportedly he brought geometry to Greece from Egypt, where it was used for practical purposes and 'attacked some [theorems] more generally and others more perceptually' (Proclus, *On the First Book of Euclid's Elements* 65.3, A11). He was the first to prove that a circle is bisected by its diameter (157.10, A20), and to know that triangles having one side and the two adjacent angles equal are congruent, 'since he must have used this theorem in his method for demonstrating the distance of ships at sea' (352.14, A20). He is also credited with discovering several other theorems. The evidence supports different interpretations, from the extreme position that Thales stated and proved certain theorems exactly as they appear in Euclid, to the other extreme that he was no mathematician at all, but that Eudemus attributed certain results to him in order to give geometry as noble an ancestry as philosophy. An intermediate view is preferable; that he made certain discoveries about spatial relations and used them to solve practical problems, but did not have a general conception of mathematical proof, let alone of a Euclidean axiomatization of geometry, and probably did not even state the propositions generally. He is explicitly credited with proving only the claim about the circle - a proposition which Euclid does not prove, but incorporates into his definition of the diameter of a circle (*Elements* I def. 17). He may have proved this result by 'superposition' - drawing a circle, cutting it along a diameter, and seeing that one piece fits exactly over the other. This is not a rigorous demonstration, but since it shows that one fact follows from another and holds generally of all circles, not only the one used in the proof, it contains the essence of the conception of mathematical proof, and would represent a decisive advance from the

problem-solving and rule-oriented approach of the Egyptians and Babylonians.

In astronomy he was credited with many discoveries, but in the end his fame as an astronomer rests on his prediction of the eclipse. However, even this cannot be accepted straightforwardly. For Thales certainly lacked the knowledge needed to predict eclipses by modern methods. At most he reported a Babylonian prediction: the Babylonians had methods of predicting when eclipses might occur, and there is nothing improbable about Thales' having come into contact with such knowledge.

The central problem in assessing Thales' contribution concerns his originality and his borrowing from others: how much was there of each, and what relative value are we to assign them? The evidence is unanimous in singling him out as a great innovator, and it would be rash to deny him all claims to importance, even if much of what he professed was due to others, especially non-Greeks. Later Ionian philosophers admired him and recognized him as their predecessor, and their ideas and methods are plausibly interpreted in terms of a historical sequence that originated with his evident curiosity about the world, his polymathy and his rational approach.

See also: Doxography; Hesiod

RICHARD McKIRAHAN

References and further reading

Classen, C.J. (1965) 'Thales', in A. Pauly, G. Wissowa and W. Kroll (eds) *Realencyclopädie der Altertumswissenschaft*, Stuttgart: Druckenmueller, suppl. vol. 10, cols 930-47.(Detailed and exhaustive discussion of the evidence; German text.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols. (Volume 1 pages 45-72 offers a thorough discussion of Thales; good bibliography.)

Heath, T.L. (1921) *A History of Greek Mathematics*, Oxford: Oxford University Press.(Still the best general account of Greek mathematics; volume 1 pages 128-39 covers Thales.)

Kahn, C.H. (1960) *Anaximander and the Origins of Greek Cosmology*, New York: Columbia University Press; repr. Indianapolis, IN: Hackett, 1994.(The best book on the Milesians.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1982) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(Chapter 2 offers a general treatment of Thales based on texts which are given in the original and in English translation.)

McKirahan, R.D. (1994) *Philosophy before Socrates*, Indianapolis, IN: Hackett.(Translation and discussion of source materials; chapter 4 covers Thales.)

Thales (*fl. c.*585 BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 1, 67- 81.(The standard collection of the ancient sources in the original languages; includes Greek texts with translations in German.)

The constructible universe

the 'universe' of constructible sets was introduced by Kurt Gödel in order to prove the consistency of the axiom of choice (AC) and the continuum hypothesis (CH) with the basic (ZF) axioms of set theory. The hypothesis that all sets are constructible is the axiom of constructibility *(V = L). Gödel showed that if ZF is consistent, then ZF + V = L is consistent, and that AC and CH are provable in ZF + V = L.*

1 The universe of constructible sets

The 'universe' of constructible sets was the first example of an inner model (see Set theory, especially §7, the notation and terminology from which will be used throughout). The 'universe' of all sets consists of a hierarchy of (transitive) levels $V(\alpha)$, where:

$$V(0) = \emptyset V(\rho+1) = V(\rho) \cup \wp(V(\rho)) \text{ at limits } V(\tau) = \cup\{V(\sigma) \mid \sigma < \tau\}$$

The universe of constructible sets has two equivalent definitions. According to the first definition, the constructible universe consists of a hierarchy of (transitive) levels $L(\rho) \subseteq V(\rho)$ where:

$$L(0) = \emptyset L(\rho+1) = L(\rho) \cup \lambda(L(\rho)) \text{ at limits } L(\tau) = \cup\{L(\sigma) \mid \sigma < \tau\}$$

where the elements of $\lambda(X)$ are those subsets of *X* that are obtainable from elements of $X \cup \{X\}$ by a few simple set-theoretic operations:

$G_1(u,v) = \{u,v\}$
$G_2(u,v) = u - v$
$G_3(u,v) = u \oplus v$
$G_4(u) = dom(u)$
$G_5(u) = \in_u = \{(w,v) \in v \in u\}$
$G_6(u) = \{((a,b),c) \mid ((b,c),a) \in u\}$
$G_7(u) = \{((a,b),c) \mid ((c,b),a) \in u\}$
$G_8(u) = \{((a,b),c) \mid ((a,c),b) \in u\}$

According to the second definition, the constructible universe consists of a hierarchy of (transitive) levels $M(\rho) \subseteq V(\rho)$ where:

$$M(0) = \emptyset M(\rho+1) = M(\rho) \cup \mu(M(\rho)) \text{ at limits } M(\tau) = \cup\{M(\sigma) \mid \sigma < \tau\}$$

where the elements of $\mu(X)$ are those subsets *Y* of *X* that are *definable* over *X*, in the sense that for some formula $\Phi(u,v,w,\ldots)$ and some $y,z,\ldots \in X, Y = \{x \in X \mid \Phi^X(x,y,z,\ldots)\}$. The highly technical proof of the equivalence of the two definitions involves correlating the set operations $G_1,\ldots,G_8$ with certain syntactic operations by which a complicated formula Φ can be built up from simpler formulas. A set *x* is *constructible* if $x \in L(\rho)$ for some ordinal ρ, or equivalently if $x \in M(\rho)$ for some ρ.

The condition '*x* is constructible' is expressible by a formula $\Xi(x)$ of the symbolic language of set theory. Gödel showed that if Ψ is an axiom of ZF+V=L, then Ψ^Ξ is provable in ZF, where Ψ^Ξ is obtained from Ψ by restricting all quantifiers 'for all *x*' and 'for some *x*' to read 'for all *x* such that $\Xi(x)$' and 'for some *x* such that $\Xi(x)$', in other words, by restricting them to read 'for all constructible *x*' and 'for some constructible *x*'. The highly technical proof involves heavy use of 'absoluteness' considerations and of various necessary and sufficient conditions for being a model of this or that axiom of set theory (see Set theory §6).

In handling the axioms of ZF, it is generally most convenient to use the second definition of constructibility. For instance, to handle the axiom of union, one must show that if *x* is constructible, then $\cup x$ is constructible: If $x \in M(\rho)$ for some ρ, then $\cup x \in M(\sigma)$ for some σ. To do this, one observes that the formula $\Phi(x,z)$ or '$z \in y$ for some '$y \in x$', which expresses the condition '$z \in \cup x$', is limited and therefore absolute for transitive sets such as the levels $M(\rho)$, so that $\cup x = \{z \mid \; Phi(z,x)\} = \{z \in M(\rho) \mid \Phi^{M(\rho)}(z,x)\} \in \mu(M(\rho)) \subseteq M(\rho+1)$. The separation axiom of ZF requires special handling. To handle it, one must show that for any formula Φ, if *x* is constructible, then $\{y \in x \mid \Phi^{Xi}(y)\}$ is constructible. Now if $x \in M(\rho)$ one has $\{y \in x \mid \Phi^{M(\rho)}(y)\} \in M(\rho+1)$, so it would suffice to show that for any formula Φ there are arbitrarily large

ordinals ρ such that for any $y \in M(\rho)$, $\Phi^{Xi}(y)$ holds if and only if $Phi^{M(\rho)}(y)$ holds. This result can be established, and indeed it is a fairly straightforward variant of the reflection principle (see Set theory §7).

In addition to handling the axioms of ZF, the axiom of constructibility must also be handled, for which purpose it is most convenient to use the first definition of constructibility. The axiom of constructibility asserts that $\Xi(x)$ holds for all *x*. What one must prove is the assertion that results when all quantifiers in this axiom are restricted to sets *x* such that $\Xi(x)$: What one must prove is the assertion that $\Xi^{\Xi}(x)$ holds for all *x* such that $\Xi(x)$, where Ξ^{Ξ} is the result of taking the rather complicated formula Ξ expressing constructibility and restricting all its quantifiers to constructible sets. Considerations of absoluteness are crucial. One successively establishes the absoluteness of the formulas expressing the conditions '$w = G_i(u, v)$', '$Y = \lambda(X)$', '$X = L(\rho)$'.

2 The axiom of constructibility

By basic ordinal arithmetic (see Set theory §4), if there is a well-order of a set *X* of order type σ, then there is a well-order of $X \otimes X = \{(u, v) \mid u, v \in X\}$ of order type σ^2. This basic fact, together with the fact that each element of $\lambda(X)$ is of the form $G_i(u, v)$ for some *i* and some $(u, v) \in X$, can be used to show that if there is a well-order of a transitive set *X*, then there is a well-order of the transitive set $X \cup \lambda(X)$ in which the elements of *X* come before those of $\lambda(X) - X$. In particular, if there is a well-order of $L(\rho)$, then there is a well-order of $L(\rho + 1)$ in which the elements of $L(\rho)$ come before the elements of $L(\rho + 1) - L(\rho)$. This fact can be used to prove by transfinite induction that there is a well-order of $L(\rho)$ for every ρ. It follows that for any set *x* that belongs to some $L(\rho)$ there is a well-order of *x*, namely the well-order of $L(\rho)$ restricted to *x*. It follows that if the axiom of constructibility V = L holds, so that every set belongs to some $L(\rho)$, then for every set *x* there is a well-order of *x*, so that the well-ordering principle WO holds. But WO is a well-known equivalent of the axiom of choice AC. Thus V = L implies AC.

By basic cardinal arithmetic (see Set theory §3, notation and terminology from which is used throughout), if *X* is countable, then so is $X \oplus X$. This basic fact, together with the fact that each element of $\lambda(X)$ is of the form $G_i(u, v)$ for some *i* and some $(u, v) \in X$, can be used to show that if a transitive set *X* is countable, then so is $X \cup \lambda(X)$. In particular, if $L(\rho)$ is countable, then so is $L(\rho + 1)$. This fact can be used to prove by transfinite induction that $L(\rho)$ is countable for every countable ρ. It follows that the set of those sets *A* of natural numbers such that $A \in L(\rho)$ for some countable ρ cannot have cardinal greater than $\aleph_1$. Gödel proved that if *A* is a set of natural numbers is constructible, so that $A \in L(\sigma)$ for some σ, then in fact $A \in L(\rho)$ for some *countable* ρ. His proof combined heavy use of absoluteness considerations with a celebrated result of mathematical logic (see Löwenheim-Skolem theorems and nonstandard models). It follows that the set of constructible sets of natural numbers cannot have cardinal greater than $\aleph_1$.It follows that if the axiom of constructibility V = L holds, so that every set (of natural numbers or of anything else) is constructible, then the set $\wp(\omega)$ of *all* sets of natural numbers must have cardinal exactly $\aleph_1$ so that the continuum hypothesis CH holds. A generalization establishes that V = L in fact implies the generalized continuum hypothesis GCH.

V = L also has implications for descriptive set theory (see Set theory §5): in descriptive set theory, well-orders and AC are used both to produce sets with desirable properties (uniformizability) and to produce sets with undesirable properties (non-measurability). Assuming V = L, the well-order of the real numbers $\mathbf{R}$ obtained as above, when considered as a subset of the plane $\mathbf{R} \otimes \mathbf{R}$, turns out to be not too complicated (a $\sum_2^1$ set). It follows that $\sum_p^1$-uniformization for all *p*, and hence projective uniformization PU, can be proved assuming V = L; but $\sum_2^1$-measurability, and hence projective measurability PM, can be refuted.

V = L also has implications for combinatorial set theory (for which see Set theory §4, notation and terminology from which will be used). One such consequence of V = L is the principle $\diamond$, asserting that there is a function *S* from Ω to $\wp(\Omega)$ with $S(\rho) \subseteq \rho$ for all $\rho < \Omega$ such that for any $X \subseteq \Omega$, $\rho < \Omega \mid S(\rho) = X \cap \rho$ is stationary. While the significance of this principle is hardly obvious on first reading, it has an important consequence, namely ~SH, the negation of Suslin's hypothesis SH about the order type θ of the usual order on the real numbers. Much more complicated combinatorial principles, implying or implying the negation of various historic conjectures have been derived from V = L through a minute analysis of the fine structure of the constructible universe by Ronald Jensen.

While the addition of certain hypotheses to ZFC as new axioms has sometimes been advocated (see Set theory §6), the addition of V = L to ZF as a new axiom has seldom been so proposed (and was not so proposed by Gödel). This is because V = L seems contrary to the intuition that the universe of all sets is as 'wide' as possible, containing wholly arbitrary collections, and because its consequences in descriptive set theory are not all desirable or attractive. Indeed, strong large cardinal and determinacy axioms, which imply both PU and PM, and which *have* often been often advocated, are incompatible with V = L. But even in investigating these other axioms, the method of 'inner models', of which the constructible model is the prototype, has been widely useful, as have variants of the constructible universe itself. (For instance, the variant $L\,[\mathbf{R}]$, which starts with $L(0) = \mathbf{R}$ rather than $L(0) = \emptyset$, plays a role in such investigations: The proof that SC implies PD proceeds by proving that SC implies that AD holds in $L\,[\mathbf{R}]$, and that AD holding in $L\,[\mathbf{R}]$ implies PD.)

See also: Axiom of choice; Continuum hypothesis

JOHN P. BURGESS

References and further reading

Devlin, K.J. (1977) 'Constructibility', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 453-89.(An authoritative survey covering both Gödel's and Jensen's work.)

Gödel, K. (1940) *The Consistency of the Continuum Hypothesis*, Princeton, NJ: Princeton University Press, 1994. (The original source.)

Themistius (*c.* ad 317-*c.* AD 388)

As a pagan philosopher and adviser to Christian Roman emperors, Themistius aimed at making the celebrated writings of his heroes Plato and Aristotle more accessible through explanatory paraphrase. An apostle of cultural Hellenism to his contemporaries, in the Middle Ages he was widely known as an important epitomizer of Aristotle.

Themistius was a philosopher, orator, politician and imperial adviser. In Constantinople he administered a school from AD 345 to 355, where he lectured to large audiences on classical philosophy, especially that of Aristotle. Unlike most Aristotelian commentators in late antiquity, Themistius combined theoretical and practical aspects of Hellenistic thought and presented them in a rhetorically effective manner. After his official appointments in 355 to a chair of philosophy and to the new Senate at Constantinople, he served Christian emperors from Constantius II to Theodosius I, eventually attaining the high rank of City Prefect in 384. His thirty-four surviving speeches celebrate imperial occasions and, more generally, articulate an influential ideology of monarchy and ecumenical empire, where foreigners with their different religions would be welcome. As a champion of the new secular Hellenism, Themistius argued that an emperor should display the qualities of a philosopher-king, that is, be both an Alexander and a Socrates. His advocacy of religious toleration is remarkable for a fourth-century pagan intellectual; he considered Christianity valid as one of the ancient cults ordained by the highest god.

The popularizing method of teaching employed by Themistius is evident in his philosophical writings. Instead of the traditional detailed textual commentary for advanced students, Themistius composed paraphrases which were designed as explanatory recapitulations to be read in conjunction with the original texts. This type of exegesis, particularly of Aristotelian works, became extremely popular throughout the medieval period in western Europe, Islamic countries and Byzantium. Of the five surviving paraphrases of Aristotelian works, three (on the *Posterior Analytics*, on *On the Soul* and on the *Physics*) survive in Greek. Those on *On the Heavens* and *Metaphysics* XII survive in medieval Hebrew translations from Arabic. Perhaps the most noteworthy of the lost paraphrases is one on the *Topics*, which influenced Boethius and early medieval logical theory (see Logic, medieval). Themistius' commentaries on Plato do not survive.

Certainly the most significant and influential of the surviving works is the paraphrase of the *On the Soul*. Themistius' interpretation of Aristotle's theory of the intellect became available to thirteenth-century philosophers through Averroes' commentaries on Aristotle (see Ibn Rushd) and William of Moerbeke's Latin translation. It had considerable impact on Thomas Aquinas and Siger of Brabant, the most important representative of Latin Averroism (see Averroism). New translations appeared in the fifteenth and sixteenth centuries.

Inasmuch as Themistius' epitomizing writings closely follow their Aristotelian originals, attention should be focused on his original interpretation of Aristotle's notoriously obscure account of the agent intellect (*nous poêtikos*) in *De Anima* III, 5. Themistius adds a substantial excursus on this difficult chapter, which analyses it in tandem with other Aristotelian texts and passages from Plato and Theophrastus. Aristotle spoke of two intellects, one actual and productive and one potential and receptive. The agent intellect is 'separable and unaffected and unmixed, being essentially actuality…. And when separated it is just what it is, and this alone is immortal and eternal.' Alexander of Aphrodisias, the last great Aristotelian commentator before Themistius (more than a century earlier), boldly asserted that the agent intellect was identical with Aristotle's supreme principle and cause of the universe, the unmoved mover which is a self-thinking intellect. Themistius rejected Alexander's equation on the grounds that Aristotle located the agent intellect 'in' the soul. He suggested instead that the productive intellect might be a secondary god, a suprahuman entity which is the cause activating each human being's 'possible' intellect. Thus, as the cause of thinking Themistius identified it with the Platonic form of the good (see Platonism, Early and Middle). The agent intellect, which Aristotle compared to light, is present like a ray to each mind when it thinks. The highest part of the human self is therefore a compound formed of the transcendent agent intellect and the individual's potential intellect. Unlike Aristotle and Alexander, Themistius claimed that the possible intellect when joined to the productive intellect constitutes the immortal aspect of human nature. Some scholars argue that human access to the transcendent aspect of the intellect is evidence of the influence of the noetic theory of Plotinus. Others emphasize Themistius' straightforward restatements of Aristotelian views and his preference for Aristotle's less idealistic ethical and political positions as compared to those of Plato and Plotinus as evidence that

his views are fundamentally Peripatetic.

See also: Aristotelianism, medieval; Aristotle Commentators; Platonism, Early and Middle

JOHN BUSSANICH

List of works

Themistius (AD 345-55) *Analyticorum posteriorum paraphrasis (Paraphrase on Posterior Analytics)*, ed. M. Wallies, *Commentaria in Aristotelem Graeca*, vol. 5, part 1, Berlin: G. Reimeri, 1900.(Aristotelian commentary.)

Themistius (AD 345-55) *In libros Aristoteles De anima paraphrasis (Paraphrase on De anima)*, ed. R. Heinze, *Commentaria in Aristotelem Graeca*, vol. 5, part 3, Berlin: G. Reimeri, 1899; trans. R.B. Todd in F.M. Schroeder and R.B. Todd (eds) *Two Greek Aristotelian Commentators on the Intellect: The De Intellectu Attributed to Alexander of Aphrodisias and Themistius's Paraphrase of Aristotle De Anima, 3.4-8*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1990.(An edition of the medieval Latin translation by William of Moerbeke is in G. Verbeke (ed.) *Commentaire sur le traite De l'ame d'Aristote traduction de Guillaume de Moerbeke*, Louvain: Publications universitaires de Louvain, 1957; this edition also includes an essay by Verbeke on Thomas Aquinas' use of Themistius' commentary on Aristotle.)

Themistius (AD 345-55) *In Aristotelis Physica paraphrasis (Paraphrase on the Physics)*, ed. H. Schenkl, *Commentaria in Aristotelem Graeca*, vol. 5, part 2, Berlin: G. Reimeri, 1900.(Aristotelian commentary.)

Themistius (AD 345-55) *In libros Aristotelis De caelo paraphrasis hebraice et latine (Paraphrase on On the Heavens in Hebrew and Latin)*, ed. S. Landauer, *Commentaria in Aristotelem Graeca*, vol. 5, part 4, Berlin: G. Reimeri, 1902.(Aristotelian commentary.)

Themistius (AD 345-55) *In Metaphysicorum librum a paraphrasis hebraice et latine (Paraphrase on the Metaphysics in Hebrew and Latin)*, ed. S. Landauer, *Commentaria in Aristotelem Graeca*, vol. 5, part 5, Berlin: G. Reimeri, 1903.(Aristotelian commentary.)

Themistius (after AD 355) *Orationes*, ed. H. Schenkl and G. Downey, Leipzig: Teubner, 1965-74.

References and further reading

Blumenthal, H. (1990) 'Themistius, the last Peripatetic commentator on Aristotle?', in R. Sorabji (ed.) *Aristotle Transformed*, London: Duckworth, 113-23.(Argues that Themistius is straightforwardly Aristotelian in his commentaries on Aristotle, without the Neoplatonic commitments of other commentators.)

Schroeder, F.M. and Todd, R.B. (1990) *Two Greek Aristotelian Commentators on the Intellect: The De Intellectu Attributed to Alexander of Aphrodisias and Themistius' Paraphrase of Aristotle De Anima, 3.4-8*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(Introduction, translation and commentary. In his introduction to Themistius' Paraphrase, pages 31-41, and in the commentary accompanying the translation, Todd outlines Themistius' thought and argues against Blumenthal and in favor of considerable Neoplatonic influence on Themistius.)

Theological virtues

The three theological virtues of faith, hope and love, referred to frequently by the apostle Paul in his letters, play an indispensable role in Christian theorizing about a person's duties with respect to God. Thomas Aquinas is responsible for the most thorough and influential philosophical theory of the theological virtues. According to him, faith, hope and love are virtues because they are dispositions whose possession enables a person to act well to achieve a good thing - in this case, the ultimate good of salvation and beatitude. Without them, people would have neither the awareness of nor the will to strive for salvation. Despite the fact that they are infused in persons by God's grace, one can wilfully and culpably fail to let them develop.

Faith for Aquinas is the voluntary assent to propositions about God that cannot be known by the evidence available to the natural capacities of humans. Other theologians, such as Martin Luther and Søren Kierkegaard, deny the assumption that faith is primarily cognitive or propositional in nature, insisting instead that it is trust in God. Kierkegaard even challenges the presupposition that faith is logically continuous with natural knowledge. There has been much debate in the second half of the twentieth century as to whether it is ever rationally permissible to believe something on the basis of insufficient evidence.

According to Aquinas, hope for one's salvation requires that one already have faith. Hope requires that one remain steadfast in the face of despair on the one hand and presumption on the other. Aquinas models the virtue of love on one strand of Aristotle's notion of friendship. Love of God entails desiring the good that God has to offer, seeking to advance God's goals, and communicating one's love to God. Love for others follows from the realization that they are also created with good natures by God.

1 Aquinas on the theological virtues in general

For Aquinas (§13), a human virtue is a disposition whose operation enables its possessor to act well to achieve an end that is good for human beings. Following Aristotle, Aquinas distinguishes between intellectual and moral virtues. The intellectual virtues pertain to the human intellect in so far as it is capable of acquiring theoretical or scientific knowledge. The moral virtues help to govern human actions and emotions. The division of human virtues into intellectual and moral would be exhaustive if every end good for humans were natural. But Aquinas thinks that the ultimate end for humans is supernatural, the beatific vision of God (see Heaven §3). Because this end is beyond the natural order of things and because the intellectual and moral virtues are fit only for the achievement of natural ends, acts directed towards achieving the supernatural end must issue from dispositions that go beyond the natural capacities of humans, namely, the theological virtues. The theological virtues are unlike the moral virtues in two important respects: (1) the moral virtues are acquired by practice - a person becomes courageous, for example, by acting as the courageous person would act; and (2) with regard to the emotions, choices and actions appropriate to a particular moral virtue, the possessor of that virtue displays a state that is a mean, neither deficient nor excessive. Thus a courageous person's emotions, choices and actions will be neither cowardly (the deficient state) nor rash (the excessive state). In contrast to point (1), the theological virtues are 'infused', instilled in a person by an act of God's grace. The infusion need not lead to success: a person can freely and culpably fail to respond to and develop the theological virtues. Aquinas claims that even though faith, hope and love are infused simultaneously, faith and hope can develop without the development of love; we then have 'unformed' or 'dead' faith. Love, however, cannot develop without the development of faith and hope, even though the blessed in heaven retain love but no longer have any need for faith or hope. In contrast to point (2), even though the theological virtues can be enhanced or diminished, they do not conform to the doctrine of the mean in the way in which the moral virtues do: one cannot believe in, hope for or love God excessively.

2 Aquinas on faith

The theological virtues, on Aquinas' understanding of them, are complexes whose analysis reveals both cognitive and volitional components. This interplay between intellect and will makes Aquinas' account of faith distinctive and subject to criticism. The primary object of faith (*fides*) is God, in so far as he can be grasped by human intellects. Human intellects operate by considering propositions. Some true propositions about God can be known by the intellect functioning within its natural capacities. For example, Aquinas thinks that there are sound arguments, appealing only to the testimony of our senses and obvious metaphysical principles, that enable us to

know the truth of the proposition that God exists. Other propositions about God and God's relation to the world cannot be ascertained by reason and experience alone. Only divine revelation can convey such truths as that the world has not existed forever but was created by God, and that God is triune in nature.

Aquinas reserves the term 'understanding' (*intellectus*) for the intellectual virtue that enables one to grasp self-evident truths. 'Knowledge' (*scientia*) can refer either to the intellectual virtue that enables one to make correct inferences, or to the cognitive state one is in when one grasps a truth by understanding or by inference, or to the body of propositions thereby known. In contrast to the proposition that triangles have three sides, the mere understanding of which compels the intellect's assent, or the proposition that the sum of the interior angles of a triangle is equal to two right angles, which follows deductively from other propositions that are obviously true, the proposition that God is triune in nature is neither self-evident nor a demonstrable consequence of self-evident truths. Yet belief in this and similar supernatural doctrines of the Christian faith is necessary for human salvation. Faith is not identical to understanding or knowledge, because the propositions constitutive of faith are not rationally compelling. An act of will is necessary to bring the intellect to assent to the propositions of faith. The resultant mental state, belief, is not identical with any of the other cognitive states induced by the will, such as opinion, suspicion or doubt, because the latter involve, to varying degrees, an element of psychological uncertainty, whereas belief involves certainty. Aquinas appears to endorse the notion that faithful belief lies between knowledge and opinion. Belief is more like opinion and less like knowledge in so far as the evidence is insufficient to warrant psychological certainty. But because faithful belief nevertheless has no taint of psychological uncertainty, it is in that respect more like knowledge than opinion.

Aquinas believes that the possession of any virtue must be meritorious. The meritoriousness of 'formed' or 'living' faith resides in the facts that the belief is the voluntary assent to revealed divine authority, and that the assent is directed by love for the ultimate good, God. If either the voluntariness or the love for the ultimate good were lacking, faith would not be a virtue.

Aquinas' position, then, is that faith is a voluntarily induced belief whose content is a set of propositions revealed authoritatively by God and typically codified in a confessional creed. The disposition to believe is infused by God. Faith is not knowledge because the intellect is not able by natural means to acquire knowledge of the propositions to which the believer assents. Unlike other voluntarily induced cognitive states, faith entails psychological certainty. Living faith is a virtue because the relevant propositional assent is voluntary and motivated by love for the object of faith, God. Some criticisms of Aquinas' position appear to be compatible with modification and redeployment (see §3 below). Other criticisms are more radical, advocating the abandonment of Aquinas' position (see §4 below).

3 Difficulties with Aquinas' view

Aquinas insists that the virtue of faith cannot be separated from the propositions that give faith its content. Since the propositional content is revealed by an essentially truthful God, one cannot have faith in a proposition that is false. (Thus faith must be distinguished from the ordinary cognitive state of belief.) Moreover, one has faith, according to Aquinas, only if one assents to all the propositions contained in the confessional creeds. To reject some part of the creeds is to fail to regard the creeds themselves as underwritten by divine authority. Thus Aquinas' view entails that all non-Christians lack faith. It also entails that a Christian heretic, who might deny some article of faith while accepting all the others, lacks faith. If one seeks an analysis of faith that is neutral among different religions, one will either have to modify Aquinas' view or look elsewhere.

Aquinas' compartmentalization of various cognitive attitudes contributes to two difficulties for his view, one involving the difference between knowledge and belief, another involving the difference between faith and doubt. On the biblical authority of the Epistle of James, Aquinas thinks that there are demons or malignant spirits who believe various propositions about God. If the virtue of faith is simply a matter of believing those propositions, then the demons would appear to have faith, contrary to their diabolical natures. Aquinas attempts to deny merit for the demons' cognitive state by attributing to them intellectual powers superior to those of humans, appearing to suggest that the evidence is so clear for the demons as to make their assent involuntary. If Aquinas accepts this position, then he is forced to concede that the demons do not *believe*; they *know*. Since Aquinas regards knowledge and belief as two separate and incompatible cognitive states, he would then have to gainsay James, whom he acknowledges to be one source of revelation. On the other hand, if the demons' belief is voluntary, then Aquinas'

account of faith would seem to confer merit on their belief.

Because faith as a virtue is a disposition that entails psychological certainty, and doubt entails lack of certainty, it is impossible for Aquinas' believer to harbour any doubts about the content of faith. Aquinas' account thus has the consequence that believers who struggle to hold on to their religious convictions in the face of lingering doubts simply do not have faith at all. In the light of this, one might seek to deny Aquinas' requirement of psychological certainty.

Aquinas' view presupposes that, to some extent, what people believe is within their voluntary control. Aquinas does not believe that all beliefs are voluntary: the verdicts of understanding and knowledge, in his technical senses of those terms, are not subject to the will. One might wonder, however, whether any beliefs can be voluntarily induced, and, in particular, whether religious beliefs can be adopted as a matter of conscious decision, as Aquinas' doctrine requires. It is unclear whether Aquinas thinks that someone infused with faith must be able to come *directly* to will to believe, or that infused faith allows for the *indirect* development of belief by the will. Consider this analogy. A person might realize that they would be happier if they did not believe that others were conspiring against them, yet be unable to eradicate the belief simply by willing it away. Nevertheless, they might be able to take small steps, such as associating cautiously but more freely with others, which would have the indirect and cumulative effect of weakening and eventually dislodging the paranoid belief. Similarly, a person might be able to embrace religious belief by associating with believers and emulating their practices. Even if no one can directly will to believe, it may be enough for an account of faith like Aquinas' that one be able to nourish infused faith by such indirect means.

4 Alternative conceptions of faith

Martin Luther's articulation of faith has been traditionally interpreted as a rejection of any view that maintains that faith is a cognitive virtue (see Luther, M.). It would seem that if faith is a virtue, then its possessors have the power to make progress towards achieving salvation on their own. But Luther insists that personal salvation can only be attained through the recognition that one's sinful state makes one incapable of living righteously. Nor is faith primarily cognitive. A person of faith believes that various propositions about God are true, but those beliefs are subordinate to trust in God, which for Luther is the core of faith. To trust in God is to resign oneself to the will of God, a resignation that is precipitated by awareness of one's own helplessness and lack of merit. But even the resignation is not under one's voluntary control: Luther views it as a gift from God. Luther's insistence on 'justification by faith alone, not by works' epitomizes his view that faith is the only path to salvation (see Justification, religious §4).

Some observers have argued that the differences between Aquinas' and Luther's accounts of faith are not as great as they may first appear. Luther claims that love is a necessary consequence of faith, whereas Aquinas regards living faith as faith informed by love. For both accounts, then, genuine faith entails love of God and neighbour. Even so, Aquinas and Luther appear to differ about the status of hope (see §5 below).

The implications of treating faith as absolute trust in God are pursued further in the writings of Søren Kierkegaard (§§4-5). Aquinas' view allows faith's cognitive content to extend beyond the content accessible to natural reason while remaining compatible with it. Many of Kierkegaard's remarks can be interpreted as an attack on such a view, on the grounds that it underestimates the radical 'leap' that religious faith requires. To conceive of faith as somehow continuous with natural knowledge is to succumb to the temptation to base faith on evidence, whether that evidence be historical (such as the testimony of miraculous events) or scientific (such as the evidence that the world is only finitely old). But the evidence will always be less than conclusive, whereas the passion of faith is infinite. Kierkegaard thinks that as long as one is guided by canons of objectivity, under conditions of incomplete or ambiguous evidence one will endlessly defer forming a religious conviction, refusing to take the 'risk'. In fact, Kierkegaard thinks that from the point of view of natural objectivity, faith can only appear absurd. Abraham's complete trust in God, manifested by his unhesitating willingness to sacrifice Isaac at God's command, is for Kierkegaard an admirable example of faith that shows how alien faith is from objective inquiry. Although the differences between Aquinas' and Kierkegaard's views are pronounced, it is not clear whether those differences preclude Kierkegaard from regarding faith as a virtue.

W.K. Clifford provides the classic statement of the opinion that faith is a vice and not a virtue. According to

Clifford, 'It is wrong always, everywhere, and for any one, to believe anything upon insufficient evidence' (1877: 346). A person who subscribes to any belief, religious or otherwise, on the basis of insufficient evidence thereby exhibits the vice of credulity, which Clifford regards as both an intellectual and a moral flaw. It follows from Clifford's position that a religious believer would be free from credulity only if the content of the belief were backed by evidence, available to and entertained by the believer, that is sufficient to warrant the belief. Many contemporary philosophers have denied Clifford's position, arguing that there need be no flaw in believing some propositions more strongly than the strength of the evidence would warrant. Even if this opinion can be sustained, it would take further argument to show that religious propositions can fit into the favoured class of beliefs.

Finally, some philosophers challenge the assumption, common to thinkers as diverse as Aquinas and Kierkegaard, that faith is necessarily distinct from knowledge. The challenge takes form by maintaining that Aquinas' conception of knowledge is overly restrictive, confining knowledge to the systematic rigour exhibited in a treatise on geometry or an axiomatization of particle mechanics. Epistemological reliabilism, in contrast, claims that a true belief is a specimen of knowledge if the belief was produced in the believer in a reliable way - for example, a perceptual belief brought about by means of a perceptual faculty functioning properly in an environment in which it is fit to function. Reliabilists typically insist that it is not necessary, in order for a true belief to count as knowledge, that the believer be in a position to justify the belief. A reliabilist who agreed with Aquinas that faith is divinely infused could regard divine infusion itself as a reliable process. If so, then contrary to what Aquinas claims, the believer's infused faith is a kind of knowledge. A proponent of such a view could still agree on Aquinas' own terms that faith is a virtue: that faith turns out to be knowledge is compatible with its being both voluntary and motivated by love of God.

5 Hope and love

Like faith, hope as a theological virtue has a specific propositional content, namely, that one will survive death and final judgment to enjoy the beatific vision of God. Augustine, in *Faith, Hope and Charity*, claimed that hope (*spes*) is directed only towards good things to be had in the future by the person who hopes. This seems to preclude the possibility that one could hope for the salvation of others. If one can hope only for one's own salvation, then it is not obvious why hope should count as a virtue, conferring moral credit on its possessor.

Aquinas' opinion is that unlike wishful thinking, hope is arduous. First, hope presupposes faith: those who have no beliefs about God's plans for humankind can have no hope that they will be saved by God. Second, even though the theological virtues do not conform to the doctrine of the mean, Aquinas thinks that hope does steer a course between two opposite vices, despair and presumption. Those who despair do not have too little hope; they have utterly abandoned hope, believing that they are too sinful to be an object of God's love. Similarly, presumption is not too much hope; it is rather a rejection of hope as unnecessary, on the belief either that one can attain salvation without God's aid or that God is so merciful that his salvation will extend to those who are unrepentant.

Aquinas' conception of hope is augmented by his remarks on love. He uses the term *amor* to characterize the ordinary notion of love as a passion, reserving the term *caritas* for the kind of love that is a theological virtue. Aquinas appeals to Aristotle's remarks on friendship in the *Nicomachean Ethics* (Books VIII and IX) in order to explicate the notion of *caritas* (see Aristotle §25). The sort of friendship (*amicitia*) that constitutes *caritas* involves three components. First, one must desire the good that the loved thing or person is in a position to provide. Although necessary, such a desire is not sufficient to distinguish *caritas* from *amor*, for the desire can be directed towards animals and inanimate objects, with whom one cannot enter into a relation of friendship. Second, one must wish well for the object of one's desire, seeking to further that person's own interests. Aquinas believes that one cannot literally have this type of desire towards animals, for it presupposes in the object of *caritas* something that animals do not have, namely, freedom of will concerning the disposal of interests. The satisfaction of the first two components is compatible with one's love remaining hidden or unrequited, and is therefore insufficient for an analysis of *caritas* as friendship. The third requisite component is that *caritas* be shared and communicated among those who participate in it. The first component, supplemented by the principle that one's love should be proportionate to the goodness of the object loved, undergirds Aquinas' claim that *caritas* requires that one love God more than anything else, including oneself, because God is the supreme good and source of all other goods. The second component allows Aquinas to extend *caritas* to other persons, including one's enemies, inasmuch as all persons are created with good natures. The second component thus explains how one can hope for the salvation

of others: if one is united to others by *caritas,* one adopts their ultimate best interests as one's own.

Aquinas' remarks on hope and his depiction of love as involving friendship between God and person reflect a guardedly optimistic conception of the human situation with respect to God. This conception appears to be irreconcilable with views expressed by Luther in his *Lectures on Romans* (1515-16). 'Pure love' of God, which Luther regards as the highest state a person can attain in this life, entails hatred of oneself, based on an awareness of one's inability to avoid sin and of God's detestation of all sinners. Luther's notion of pure love implies that even if hope is a virtue, it is not necessary for salvation and that despair need not be a vice. Moreover, it is hard to see how love of others can be a virtue for one who has pure love of God: if one hates oneself and realizes that all other are equally detestable, then one should hate them also.

See also: Faith; Grace; Religion and epistemology; Virtue ethics; Virtues and vices

WILLIAM E. MANN

References and further reading

Adams, R.M. (1987) *The Virtue of Faith and Other Essays in Philosophical Theology*, New York: Oxford University Press.(The title essay provides some reason for thinking that faith is better than knowledge. Two other essays discuss Kierkegaard's views.)

Aquinas, T. (*c*.1265-73) *Summa theologiae*, trans. Fathers of the English Dominican Province, London: Burns, Oates & Washbourne, 2nd edn, 1920-9; repr. Westminster, MD: Christian Classics, 1981.(The writing style requires getting used to. Virtues and vices in general are discussed in questions 49-89 of the first part of the second part; the theological virtues and vices in questions 1-46 of the second part of the second part.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985.(Perhaps the most influential book in the history of ethics, its thought is evident in the massive second part of Aquinas' *Summa theologiae*.)

Augustine (*c*.423) *Faith, Hope and Charity*, trans. and with introduction by B.M. Peebles, The Fathers of the Church, vol. 2, Washington, DC: Catholic University of America Press, 1947, 355-472.(Often referred to as the *Enchiridion*, this is perhaps the first sustained treatise on the theological virtues.)

Clifford, W.K. (1877) 'The Ethics of Belief', in L. Stephen and F. Pollock (eds) *Lectures and Essays by the Late William Kingdon Clifford, F.R.S.*, London: Macmillan, 2nd edn, 1886, 339-63.(Clifford's classic attack on belief based on insufficient evidence.)

James, W. (1896) 'The Will to Believe', in *The Will to Believe and Other Essays in Popular Philosophy*, The Works of Willam James, vol. 6, ed. F.H. Burkhardt, F. Bowers and I.K. Skrupskelis, Cambridge, MA: Harvard University Press, 1979, 13-33.(A famous rejoinder to Clifford's position.)

Kenny, A. (1992) *What is Faith?*, Oxford: Oxford University Press.(Contains a series of lectures on faith and reason, endorsing parts of Plantinga's case against views like Clifford's.)

Kierkegaard, S. (1843) *Fear and Trembling*, trans. and with introduction by A. Hannay, Harmondsworth: Penguin, 1985.(Kierkegaard's meditation on Abraham's faith.)

Kierkegaard, S. (1846) *Concluding Unscientific Postscript*, trans. D.F. Swenson and W. Lowrie, Princeton, NJ: Princeton University Press, 1941.(Part II, chapter 2 contains Kierkegaard's comments on the discontinuity between faith and objectivity.)

Luther, M. (1515-16) *Lectures on Romans*, trans. J.A.O. Preus, *Luther's Works*, vol. 25, St Louis, MO: Concordia, 1972.(Chapters 8 and 9 contain Luther's comments on pure love.)

Luther, M. (1520) 'The Freedom of a Christian', trans. W.A. Lambert, revised H.J. Grimm, *Luther's Works*, vol. 31, Philadelphia, PA: Muhlenberg, 1957, 343-77.(Luther on justification by faith alone.)

Mann, W.E. (1993) 'Hope', in E. Stump (ed.) *Reasoned Faith*, Ithaca, NY: Cornell University Press, 251-80.(A discussion of Luther's notion of pure love and Aquinas' remarks on hope.)

Penelhum, T. (1977) 'The Analysis of Faith in St Thomas Aquinas', *Religious Studies* 13 (2): 133-54.(A careful discussion of Aquinas' views that develops the criticisms about demonic belief and the incompatibility of faith with doubt.)

Pieper, J. (1986) *On Hope*, trans. M.F. McCarthy, San Francisco, CA: Ignatius Press. (An elementary presentation of Aquinas' views.)

Plantinga, A. (1982) 'Rationality and Religious Belief', in S.M. Cahn and D. Schatz (eds) *Contemporary Philosophy of Religion*, New York: Oxford University Press, 255-77.(An influential critique of Clifford that is hospitable to reliabilism.)

Swinburne, R. (1981) *Faith and Reason*, Oxford: Clarendon Press.(Chapter 4 argues that the differences between

Aquinas' and Luther's conceptions are not as great as they seem.)

Theology, political

The concept of political theology was the subject of important controversies in European, and especially German, philosophy, social science and jurisprudence in the twentieth century. After the First World War, a debate took place between the jurist Carl Schmitt, an influential right-wing critic of parliamentary democracy in the Weimar Republic, and the theologian Erik Peterson. Another debate was occasioned by a new, leftist political interpretation of biblical texts in the years after 1960. In that context, 'political theology' designates philosophical positions influenced by neo-Marxist philosophies, such as the critical theory of the Frankfurt School. The earlier controversy between Schmitt and Peterson played a role in this later debate as well. Johann Baptist Metz is probably the most important representative of political theology in the later controversy.

In his writings on the problem of the legitimacy of the modern state and its constitution, the jurist Carl Schmitt focused on the question of sovereignty. In his view, the sovereignty of the modern state is derived from an earlier theological tradition centred on God's existence and nature. The modern theory of the state, Schmitt held, had shifted the idea of God's omnipotence, which theologians understood as determined only by God's goodness, to the concept of the sovereignty of the people. Rousseau (§3), for example, had described 'the general will' as the expression of the absolute will of a state's citizens, which could never be wrong or evil. The problem is that when reference to transcendence is abandoned, the notion of the state that remains is despotic, as, for example, in Hobbes' *Leviathan* (see Hobbes, T. §7). To describe the fundamental condition of modern politics, Schmitt introduced the notion of a situation of exemption. In a situation of exemption, radical political movements and charismatic leaders are allowed to annul laws simply by a decision to do so. Schmitt articulated a principle which helped to support both left- and right-wing political messianism: 'Authority, not truth, makes law'.

The most important opposition to Schmitt's account of the political theology of the modern state did not arise either from classical theories of natural law, in the Aristotelian or Thomistic traditions, or from a Kantian theory of law as based on arguments supported by practical reason. Instead it came from a theologian, Erik Peterson, who had converted to Catholicism from Protestantism. Peterson emphatically rejected Schmitt's understanding of the voluntary character of law as theological. He based his argument against Schmitt's position on historical as well as philosophical grounds. He pointed to the decision of the early Church to reject Arianism and to accept the doctrine of the Trinity. One of the consequences of this position was set out by Augustine, Peterson thought. In his *De civitate Dei (City of God)*, Augustine rejects a version of political theology which he thinks he finds in Eusebius of Caesarea; instead, Augustine maintains that there is and must be a separation between the City of God, on the one hand, and the earthly city, or state, on the other. In Peterson's view, this Christian position effectively undermines any possibility of a political theology of the sort Schmitt propounded.

After the Second World War, the Catholic theologian Johann Baptist Metz established a new version of political theology. In the 1960s, and especially during the time of Vatican II (1962-5), the Catholic Church was engaged in a process of reform and was searching for a renewed dialogue with the society in which it found itself. Metz's political theology was part of this larger trend. Metz took a positive view of the Enlightenment and modern freedom of thought, and saw his position as a legitimate consequence (supported by biblical studies) of modern anthropology. Like his teacher Karl Rahner, Metz was concerned to bring the work of Thomas Aquinas into contemporary philosophical discussion. In his earlier writings, Metz also followed Rahner in trying to integrate Kant's transcendental philosophy and German idealism in general with the philosophical presuppositions of systematic theology. These efforts significantly transformed the Neo-Thomist account of human nature.

Metz's political theology was also based on a critical understanding of Kant's account of human freedom and Marx's concept of history. To some extent, it was influenced also by the views of Horkheimer and Adorno, who were critical of the Enlightenment and of modern theories of rationality, and by Benjamin's view of history as 'one single catastrophe where we perceive a chain of events' (1950, Thesis IX). So understood, political theology constitutes an important theological theory of modern times. At its heart, there is the concept of an eschatological reservation with regard to all stages of progress and emancipation in history. This feature of Metz's political theology allows it to avoid the sort of mistakes for which Peterson reproached Schmitt.

See also: Religion and political philosophy

MATTHIAS LUTZ-BACHMANN

References and further reading

Anglet, K. (1995) *Messianität und Geschichte. Walter Benjamins Konstruktion der historischen Dialektik und deren Aufhebung ins Eschatologische durch Erik Peterson (Messianism and History. Walter Benjamin's Construction of the Historical Dialectic and their Shift into the Eschatological through the Work of Erik Peterson)*, Berlin: Akademie.(An important study of the relation between political theology, Critical Theory and messianism.)

Augustine (413-27) *De civitate Dei (The City of God)*, Corpus Christianorum, Series Latina, 47-8, Turnhout: Brepols, 1955-.(One of the most influential Christian views of history and politics, criticizing the ancient pagan identification of state and religion.)

Bendersky, J.W. (1983) *Carl Schmitt, Theorist for the Reich*, Princeton, NJ: Princeton University Press.(General introduction with bibliography.)

Benjamin, W. (1950) 'Über den Begriff der Geschichte' (On the Concept of History), *Neue Rundschau* 61; trans. H. Zohn, New York: Schocken Books, 1969.(Although first published in 1950, this was written in 1940. This is the most important text of Benjamin, containing his programmatic intentions.)

Habermas, J. (1997) 'Israel oder Athen: Wem gehört die anamnestische Vernunft?' (Israel or Athens: To Whom does Anamnestic Reason Belong?), in *Vom sinnlichen Eindruck zum symbolischen Ausdruck (From Sensible Impression to Symbolic Expression)*, Frankfurt: Suhrkamp.(Philosophically, one of Habermas' most important studies defining the relation between his version of Critical Theory and political theology.)

Horkheimer, M. and Adorno, T. (1947) *Dialektik der Aufklärung (Dialectic of Enlightenment)*, trans. J. Cumming, New York: Continuum, 1994.(The programmatic study of earlier Critical Theory on the internal dialectic of the process of Western civilization.)

Lutz-Bachmann, M. (1980) 'Freiheitsgeschichte und Theologie' (Liberation History and Theology), in *Philosophische Rundschau* 27.(Introduction to the political theology of Metz and others.)

Maier, H. (1970) *Kritik der politischen Theologie*, Einsiedeln: Johannes.(A criticism of Metz inspired by the work of Erik Peterson.)

Metz, J.B. (1968) *Zur Theologie der Welt (On the Theology of the World)*, Mainz/Munich: Grünewald.(Collected articles introducing the new concept of political theology.)

Metz, J.B. (1977) *Glaube in Geschichte und Gesellschaft, Studien zu einer praktischen Fundamentaltheologie (Faith in History and Society, Studies Towards a Practical Fundamental Theology)*, Mainz: Grünewald. (Collected articles deepening the new concept of political theology.)

Metz, J.B. (1980) *Jenseits bürgerlicher Religion, Reden über die Zukunft des Christentums (Beyond Bourgeois Religion, Lectures on the Future of Christianity)*, Mainz: Grünewald.(A collection of lectures and homilies, easy to understand.)

Nichtweiß, B. (1992) *Erik Peterson. Neue Sicht auf Leben und Werk (Erik Peterson. A New View of his Life and Work)*, Freiburg: Herder.(An informative introduction to Peterson's work.)

Peterson, E. (1935) *Der Monotheismus als politisches Problem, ein Beitrag zur Geschichte der politischen Theologie im Imperium Romanum (Monotheism as a Political Problem, a Contribution to the History of Political Theology in the Roman Empire)*, Leipzig: Jakob Hegner.(An important critique of Carl Schmitt's concept of political theology.)

Peterson, E. (1994) *Ausgewählte Schriften (Selected Writings)*, vol. 1, *Theologische Traktate*, ed. B. Nichtweiß, Würzburg: Echter.(First volume of the new edition of Peterson's works.)

Schmitt, C. (1922) *Politische Theologie*, Munich and Leipzig: Duncker & Humblot. (The important work in which Schmitt introduces his theory of the religious origin of the modern state.)

Schmitt, C. (1970) *Politische Theologie II*, Berlin: Duncker & Humblot.(An apology in which Schmitt defends his concept of political theology against his critics.)

Sölle, D. (1982) *Politische Theologie*, Stuttgart: Kreuz.(A Protestant version of modern political theology in Germany.)

Theology, Rabbinic

The Talmud, a shelf of folio volumes built up out of the expansive reflections of generations of scholar/thinkers whose discourse formed a commentary or complement (Gemara) to the ancient legal code of the Mishnah, encapsulates rabbinic sayings and discussions dating from before the first century to around 600. The monotheistic idea of God affords a key perspective on the Talmud's variegated themes: God's uniqueness and ultimacy preclude any easy direct commerce between the human and the divine. Literal contact may be endlessly deferred; yet God remains ever present and ever active in human life. Rabbinic thought, textually represented in the Talmud and in the institutions it fosters, seeks to mediate God's 'absent presence' to a community of believers in a way that renders manifest the penetration of divine concern into every cranny of human consciousness without compromising God's transcendence or explanatory uniqueness.

1 The rabbinic framework

The location of the believer *vis-à-vis* the God of Israel is elegantly captured in the Talmudic dictum: 'Everything is in the hands of Heaven except the fear of Heaven' (Babylonian Talmud (B.T.) Berakhoth 33b; Megillah 25a; Niddah 16b). Everything falls under God's sway, but only human beings can decide whether or not to hold God in awe. Whatever occurs in the world comes under the purview of the monotheistic God - except the affirmation of his role as God. The concatenated system of beliefs and practices that constitute Judaism works only through such acknowledgement of God as creator and ruler of the universe.

The unbridgeable conceptual gap between God and ourselves presupposes a critical epistemology readily capable of registering the inadequacy of our knowledge claims about God. It delineates the outermost limits of defensible human utterance and locates God beyond those limits. The pattern of argument parallels that of philosophical scepticism in urging that whole species of statements exceed the bounds of coherence. But where scepticism introjects its critique, negative theology projects it outward, onto God (see Negative theology).

Since all descriptions of God here are irresolvably metaphoric, it is only human accetance of God's centrality that sets the religious life in motion. All discourse about God is ultimately metaphoric. Its 'cash value', the only 'hard currency' for which invocations of God can be exchanged in the end, is the act of human consent that establishes God's sovereignty in our lives, acceptance, as the Talmudic rabbis call it, of the yoke of the kingdom of Heaven (God, concepts of).

For the rabbis the supreme mitzvah or divine commandment is the study of Torah - both the written law of Scripture and the oral law of the rabbinic canon itself (B.T. Shabbat 127b). The centrality of the monotheistic idea helps to explain this privileging of the intellectual in a system that is, by and large, practical, even pragmatic, in orientation: The rabbis find spiritual elevation, even rapture, in the intellectual cultivation linked with the study of Torah, its norms and intricacies. Having no tangible object of religious devotion, we pursue the Infinite through unending occupation with God's word, will and law, ceaselessly testing the limits of our rational faculties, trying again and again to transcend literalism and probe the nexus between symbol and object that is the genesis of metaphor. The outward and upward movement towards a grasp of the divine thus converges with an inward and downward movement, towards self-reflection, integrating the lessons of the limitations of reason.

2 Rabbinic norms

Rabbinic ideas of the good life reflect the metaphysical priorities evoked by the monotheistic idea. The rabbis look with sharp disapproval on a life of asceticism (see Asceticism). One who merely refrains from drinking wine, they observe, is deemed a sinner against the self (B.T. Nedarim 10a) - how much more so those who refrain from marriage! The Sages (as the ancient rabbis are often called) warmly echo Isaiah's words about the earth, that God 'created it not a waste, He formed it to be inhabited' (Isaiah 45: 18; B.T. Yevamot 62a): divine transcendence does not empty the world of value, since the world is God's work; but it does mean that our guidelines to human perfectibility must emerge from a sensibility informed by an enlightened understanding of our concerted interactions with one another and with our natural environment. Such guidelines are not simply vouchsafed to us - not even by revelation. Just as monotheism continually thrusts us back upon our own intellectual resources theologically, so too does it prod us persistently to refocus on the ways in which our existence in the world can be

made better by greater attentiveness to the norms and limitations of the human condition. Monotheism does not urge us to abandon these norms or transcend these limitations by some lurch into an Ethical Beyond; it does not countenance notions of a transmoral imperative or teleological suspension of the ethical.

Six times in the Talmud (for example, B.T. Shabbat 63a) the prominent rabbinic Sage Samuel emphasizes that only Israel's subjection to the nations distinguishes historical time from the Messianic age. The remark is highly evocative of the rabbinic perspective: since it indicates that human success in reaching the consummation of history has no overtly miraculous emblem by which it is to be judged: monotheism does not give us the mind of God to read. There are human initiatives and reconsiderations but no end-stage within sight that can be declared to have God's full sanction, no Hegelian or Marxian final stage of history, until the parameters of history themselves have been transcended (see History, philosophy of).

We approach God not by theorizing but by continual mobilization of our energies and will, orienting ourselves towards what we can grasp of our canonical texts and thereby, in the course of time, improving human lives and the world. But the Judaic stress on fulfilling God's commandments goes hand in hand with affirmation of what we might call partial redemption. Judaism is deeply suspicious of human claims to have attained complete redemption. The inaccessibility of God redirects our energies towards this world - and so towards events whose outcomes, inevitably ambiguous and partial, spur us to further action, *since they are never validated supernaturally as the sufficient consummation of human effort and achievement*.

3 Action and intention

A recurring controversy in the Talmud (for example, B.T. Rosh Hashanah 28b) addresses the question whether religious obligations need to be fulfilled with deliberate intention. That such an issue could even be disputed is highly illuminating. The rabbis view with equanimity the prospect of assigning primacy to actions whose motives, in an underlying network of thoughts and feelings, are not perfectly defined. Two reflections on the monotheistic idea shed light on this rabbinic stance. First, as we have seen, the conceptual non-negotiability of the monotheistic God makes all human formulations about Him rebound into metaphoric approximations whose major point of intersection with our reality is pragmatic. If what we are comparing - and what we are comparing it to - can never be fixed as divine unreservedly, then our frame of reference in assigning content to our statements about God remains irredeemably human. For example, God is conceived by us relationally - that is, through his actions - as creator of the world, not as he is in himself. Human beings, correspondingly, are invisible to ourselves and others, and are known, if we can be known at all, only through what we do, not through what we think or feel. Second, since action is as metaphoric in relation to God as are the categories of thought and feeling (its primacy in a monotheistic vocabulary merely reflecting the impossibility of our penetrating God's interiority - God's thoughts and feelings, as it were) monotheistic 'action', when viewed as a registering of human possibilities, would be, correspondingly, much more amorphous than our traditional concepts of action would suggest. It might invoke no more specific a reality than that of power - where everything that we utter, think, feel, or do constitutes an assertion of power, but where the familiar demarcations of thought and action, expression and intention become blurred. It is perhaps for this reason that the rabbis can at times seem rather cavalier about thoughts and feelings in the performance of *mitzvot* - not because they denigrate human intentions but because intentions themselves have been absorbed in a broader ontology. Some rabbinic authorities (for example, Abraham Danzig, Hayyei Adam 68: 9), consider the mere engagement in the performance of a mitzvah a sufficient registering of intention; they see no further need for an explicit or separate mental articulation of intention.

4 Progress and retrospection

Two famous themes of rabbinic exegesis seem to point in conflicting directions. The first is the notion of 'It is not in heaven' (Deuteronomy 30: 12) as expounded in the Babylonian Talmudic tractate Baba Metzia 59b, where the biblical dictum about the accessibility of the law is taken to assign to human inquirers the power and responsibility to interpret (and elaborate) it. The other is the idea that 'Even what a conscientious student is destined to say before his teacher was stated to Moses at Sinai' (Midrash Vayikrah Rabbah 22: 1). The first statement conjures up images of irreversible forward movement, in which we are thrust upon our own resources, however paltry, to structure and organize a *halakhic* (Jewishly mandated) life. The second suggests preformation rather than evolution and an inevitable retrospection or even backward movement.

One strategy for reconciling these two texts is to say that the principle of 'It is not in heaven' does not ordain a strictly linear and irreversible development of Jewish law, but something more dialectical and more nearly akin to Hans-Georg Gadamer's hermeneutics. Gadamer argues for the contemporaneity of all texts, their essential incompleteness, subject to the fusion of horizons between an author and all subsequent generations of readers. Yosef Hayim Yerushalmi in his book *Zakhor* (1982) similarly argues eloquently for an unbridgeable tension between the ever-dynamic practice of Torah and the more positive, diachronic demands of ordinary historical writing. For historical writing, the narrative sequence moves irreversibly from past to present. The uniqueness and separateness of the past are preserved forever. But, in the perspective of the Sages, the movement, at least as momentously, is from future to past - so that time forms a seamless web with no sharp chronological divisions. Later interpretations fill in the content of the original teaching, retroactively. The past remains present as a reservoir of legitimation for the expansion of meaning; the future constitutes an inevitable resource for the extension of meaning and reference. The hermeneutical principle of the underdetermination of meaning by text becomes a cardinal rabbinic tool for balancing the authority of precedent and the demands of crisis and creativity (see Hermeneutics, biblical).

5 The underdetermination of meaning by text

As the few examples cited here clearly suggest, biblical and earlier rabbinic texts are often quoted out of context in the Talmud, stripped of their original significations and outfitted with wholly new connotations germane to matters at hand. By such unremitting recontextualizing, the Talmud makes texts speak to ever new issues. What is the significance of this pervasive Talmudic device based on the underdetermination of meaning by text?

From a purely practical standpoint, recognition of the underdetermination is a key factor in rabbinic innovation, allowing the application to unanticipated circumstances of ideas, methods and principles well established in Jewish law.

Hermeneutically as well, the recognition of underdetermination introduces an element of complexity into the idea and the project of historical continuity and faithfulness. A good part of what the Talmudic Sages were doing was consciously to subvert a slavishly historical approach to biblical texts and rabbinic precedents. The rabbis work and play with the texts, almost endlessly reinventing them. The Talmud thus contains an ineradicably storytelling element, even in the most austerely legal discussions. Thus, paradoxically, to give a historically faithful account of a Talmudic discussion or legal analysis, one must somehow accommodate historiography to the Talmud's own resolute anti-historicism.

A jarring theological issue emerges from the centrality of the underdetermination principle in Talmudic exegesis. For this principle points towards scepticism and seems to destabilize all interpretation. Once scepticism is unleashed with regard to the textual sources of Jewish tradition, how can it be contained with regard to the norms that those texts enunciate? The tension between the demand to conserve the Torah's norms and the pressure to decontextualize them (whether as expressed in the Torah text or in its canonical offshoots among the codes), so as to make the underlying norms relevant in transformed contexts - temporally extending the normative reach of the Torah in the broadest sense by way of skeptical strategies of underdetermination - has profound implications for the project of the rabbis as founders and exponents, and, in a sense, reinventors, of Jewish tradition.

The same underdetermination illuminates the priority assigned to speech over writing in Talmudic discourse. It was only by reference to the pressure of persecution that reduction to writing was deemed permissible (B.T. Berakhot 54a). Even then, apparently, it was expected that the writing should resemble speech as closely as possible, by remaining dialectical in structure and even in diction. Texts can engender an illusion of fixity and finality that orality continually belies. Only speech, it seems, is deemed in good faith; writing is always in bad faith. The fact of persecution and the risk of the utter loss of collective memory permits recourse to writing. But dialectical writing, with its tentativeness and recursiveness, its continually interrogatory aspects, presents itself as the mere record of living speech.

A contrast is often drawn between the Talmud's legalistic mode of argument and the methods familiar to philosophers. The Talmud is practically oriented, we are told, where philosophy is theoretical. But in the light of our present analysis, that dichotomy loses its force. The vast canon of Talmudic-legalistic argument itself insinuates a monumental philosophical project.

Articulation simultaneously constitutes erasure, effacing the very possibilities that have just been spoken. What each speaker says can always be decomposed and reconstituted along lines other than those in which the words seemed officially to be pointing. Sentences continually posit their silent others, as dialectical alternatives, and as presuppositions and implications that can always be regrouped and rearranged to yield new configurations of meaning and unsuspected, but possibly hinted, constellations of referents. The whole notion of the primacy of intention is thus reverently mocked, as dicta are made precedents or pretexts for understandings far removed from their literal or naive or historical imputations. Talmudic argument, viewed from this perspective, constitutes one of the most imposing rhetorical edifices ever erected in defence of the notion of the 'floating statement' and the endless process of de-contextualization and re-contextualization.

The theological backdrop for all Talmudic argument is the fabric formed by the tenets of monotheism. The very utterance of the word 'God' initiates a process of endless displacement that finds no resting place. All we can ever do towards assigning a content or pinpointing a sense for this word is to say, without letup, that God is not literally to be construed in this way or that, not to be found in a humanly cognizable sense here or there or elsewhere. Talmudic discourse itself mimics this elusiveness; its stability and steadiness of focus through the endless process of deconstruction and reconstruction refuses to come to rest in some merely fixed and positive content. The deliteralizing of positive ascriptions to God of particular attributes is translated in the Talmud - with its continually rendering fluid of the sense and reference of human utterances - into an implicit acknowledgement of the ultimate ungroundedness of all mere acts of language. In this negative sense, God for the Talmudic speakers and editors becomes the paradigm case of the humanly sayable.

See also: Halakhah; Islamic theology; Karaism; Midrash

ARYEH BOTWINICK

References and further reading

Botwinick, A. (1995) 'Underdetermination of Meaning by the Talmudic Text', in D.H. Frank (ed.) *Commandment and Community: New Essays in Jewish Legal and Political Philosophy*, New York: State University of New York Press.(Expands on the theme of underdetermination of meaning by text in the rabbinic sources.)

Handelman, S.A. (1982) *The Slayers of Moses*, New York: State University of New York Press.(Invokes post-structuralist and postmodernist methods and perspectives for analysing rabbinic texts.)

Heschel, A.J. (1962, 1965, 1990) *Torah Min H'Shamayím B'Asplakaria shel ha-Dorot (Theology of Ancient Judaism)*, vol. 1, *Darkei Machshavah B'Tekufat ha-Tana'im (Pathways of Thought in the Age of the Tana'im)*, vol. 2, *Torah M'Sinai V'Torah Min ha-Shamayim (Torah from Sinai and Torah from Heaven)*, London: Soncino Press; vol. 3, *Elu V'Elu Divrei Elokim chaim (Both these Words and Those are the Words of the Living God)*, Jerusalem: Jewish Theological Seminary.(Heschel builds a typological contrast between the approaches to the Torah of Rabbi Akiva and Rabbi Ishmael and their schools. He shows how the unbridgeable conceptual distance between God and human beings is compatible with normal logical-linguistic protocols in deciphering the Torah text.)

Kaddushin, M. (1972) *The Rabbinic Mind*, New York: Bloch, 3rd edn.(Focuses on 'normal mysticism' - the rabbinic theological approach that domesticates divine distance and integrates it into a fabric of daily living.)

Neusner, J. (1988) *Torah: From Scroll to Symbol in Formative Judaism*, Atlanta, GA: Scholars Press.(Shows how the Torah as finite text, a simple scroll, is transformed into an all-encompassing context that defines a way of life for individuals and for the people of Israel.)

Urbach, E.E. (1969) *Chazal: Pirkei Emunot V'Deot*, Jerusalem: The Magnes Press; trans. I. Abrahams, *The Sages - Their Concepts and Beliefs*, Jerusalem: The Magnes Press, 1975, 2 vols.(A classic reconstruction of the fundamental normative and descriptive categories of rabbinic Judaism.)

Yerushalmi, Y.H. (1982) *Zakhor: Jewish History and Jewish Memory*, Seattle, WA: University of Washington Press.

Theophrastus (*c*.372-*c*.287 BC)

Theophrastus, the pupil and successor of Aristotle, shared all the latter's interests, and produced a large number of works on the same topics. Some, like the extant botanical works, went far beyond Aristotle, and Theophrastus is known as the Father of Botany; others amplified and criticized what Aristotle had done. The short Metaphysics, also extant, raises many questions about the nature and the possibility of metaphysics, but most of his work on logic, science, psychology, ethics, politics and religion survives only in fragments, some material coming from the Arabs, and some only from medieval Latin sources. His developments of modal logic and various forms of the syllogism were regarded as important, and his amplification of Aristotle's account of the human intellect was studied in the Middle Ages in the West. His little Characters, an entertaining set of sketches of human peculiarities, has had considerable influence on later literature, and his surveys of earlier opinions, of which his On the Senses survives, influenced later doxographers. He was an older contemporary of Zeno, the founder of Stoicism, but his influence on Stoicism remains uncertain, and we also know little of his relationship with Epicurus.

1 Life and writings

Theophrastus was born in Eresos on the island of Lesbos, but went to Athens for higher education and became the pupil and then the constant companion of Aristotle, whom he succeeded as head of his school, the Peripatos, in 323 BC. He probably shared Aristotle's sojourns from 348 onwards in Assos, Lesbos and Macedonia, where Aristotle taught the future Alexander the Great, and returned with him in 335 to Athens, where most of the rest of his life was spent. He was succeeded by his pupil Strato (died *c*.269 BC). He shared all Aristotle's interests, and wrote widely in a very careful style. Diogenes Laertius lists over 200 of his works. His two botanical works, known popularly as the *History of Plants* and the *Causes of Plants*, have survived, as well as his *Characters*, an entertaining and widely influential work, whose original purpose is uncertain: it seems to have influenced the playwright Menander, himself a pupil of Theophrastus, and may have rhetorical, ethical or theoretical purposes. It enjoyed considerable vogue in the seventeenth and eighteenth centuries, particularly in France. There survive also the short *Metaphysics*, and a number of treatises on scientific matters, but much more survives only in fragments. It is likely that he and his fellow Peripatetic Eudemus worked out their logical theories while Aristotle was alive, but other material presupposes the existence of Aristotle's works as we have them now, which suggests that it was produced after his death. A large part of it was concerned with questioning, widening and improving the works of Aristotle, and Theophrastus himself probably played a part in editing Aristotle's lectures into the form in which we now have them. It is also likely that through his pupil Demetrius of Phalerum he had some influence on the direction taken by the Library and Museum of Alexandria. His contemporary Epicurus wrote a work addressed to him (fr. 280), but we know little else about his relations with either Epicurus or the Stoics.

2 Logic and grammar

Theophrastus made a number of contributions to logic (see Logic, ancient §4). He added a few footnotes to Aristotle's system of categorical propositions, such as a new account of the relations between the second and third figures of the syllogism and the first, and a study of the basis on which the premises of a syllogism are themselves to be accepted (fr. 94). The invention of the fourth figure of the categorical syllogism has been ascribed to him, mistakenly, but he did add a number of additional moods to those recognized by Aristotle (frs 91-6). He went much further in attacking Aristotle's unsatisfactory system of modal logic, working closely with his contemporary, Eudemus. For this, evidence from Arabic sources goes beyond what we have from the Greeks (frs 98-109). Aristotle held that in a modal syllogism, which uses the notions of 'necessary', 'actual' or 'assertoric', and 'possible' in its premises and conclusions, one of which might for example be 'All men are possibly good', the mode of the conclusion follows that of the major premise. But his pupils saw flaws in that position and suggested that the conclusion always follows the mode of the weaker premise: that is, if one premise is necessary and the other assertoric then the conclusion is always assertoric, whichever of the premises is. This also involved revised accounts of the meaning of terms such as 'necessary' and 'possible', and of the conversion of propositions. 'Some *A* is *B*' implies 'Some *B* is *A*', but if *A* possibly belongs to no *B*, does *B* possibly belong to no *A*? Aristotle argued against this, but Theophrastus and Eudemus produced a proof that it does.

Theophrastus also developed some remarks of Aristotle into a theory of prosleptic syllogisms, such as ‘*A* is said of all of that of all of which *B* is said’ (fr. 110), and with Eudemus worked out a theory of hypothetical syllogisms (frs 111-13), such as ‘If virtue is knowledge, it is teachable. But it is knowledge. Therefore it is teachable’, the relation of which to Stoic logic is controversial. Finally, he revised Aristotle’s system of topics, and gave some definitions of key terms such as ‘topic’.

He also made advances in the theory of grammar, which Plato and Aristotle had begun by recognizing nouns and verbs. Theophrastus added ‘joints and ligaments’, which probably included conjunctions, prepositions and articles, and he also paid attention to aspects of style (fr. 683).

3 Psychology

For Theophrastus’ psychology we have evidence from Themistius (fourth century AD) and Priscian of Lydia (sixth century AD) which provides a considerable proportion of Theophrastus’ *On the Soul*, a work devoted mainly to a critique of Aristotle’s own *On the Soul*. Curiously, Priscian used Theophrastus as a peg on which to hang his own Neoplatonic account of the same subject matter, sensation, imagination and thought. Theophrastus laid the ground for future work by bringing material from Aristotle’s *Generation of Animals* into discussion about the origin of soul and intellect in the embryo. In the field of sensation he worried over Aristotle’s claim that in sensing the sense organ becomes like its object (see Aristotle §18), for it could hardly become like it in colour or taste, for example, so he fastened on Aristotle’s references to form and *logos* as explanations of this process, and considered detailed aspects of sensation, such as light, colour, reflections and the transparent. Where Aristotle is obscure he can be obscure himself, as in the section on why there are only five senses. He also found difficulties in Aristotle’s condensed account of intellect (see Aristotle §19; Nous), asking how intellect differed from matter, for both seemed to have no characteristics in themselves; and how intellect and its objects, the thinker and its thoughts, could causally affect one another. He seems to have distinguished intellect from matter by saying that while matter receives individual forms intellect receives universal forms (fr. 308). He insisted again and again on the importance of using language appropriately, to avoid being misled, accepting that we have sometimes to transfer words appropriate primarily to one subject matter to another subject matter; this can be done successfully provided we are aware of what we are doing.

In his epistemology, if Sextus Empiricus (second century AD) is correct, he provided an empiricist theory of knowledge (fr. 301A) similar to that given by Aristotle at the end of his *Posterior Analytics*. He stressed the importance of self-evidence as a criterion of truth, and claimed that both sensation and thought are foundations of knowledge.

4 Metaphysics

Theophrastus’ *Metaphysics*, now generally agreed to be a complete work in itself, is a critique not only of the views of Aristotle but also of those of Plato and other philosophers. In particular it deals with themes in *Metaphysics*, Aristotle’s ‘theological’ work, and it has been argued plausibly that it was an early work of Theophrastus. Its original title is unknown: ‘First Philosophy’ would be appropriate. It has been seen as the classic formulation of a whole complex of problems. First he tackles some of the assumptions made by Aristotle, starting with the claim that there exist ‘first things’, known by reason. These must be either in mathematical objects, or in something prior to them; they are the cause of motion, but are themselves unmoved; as objects of desire they cause the rotation of the heavens. In all this he is following Aristotle, but he then asks how many they are: if there is only one, why do the heavenly bodies move differently, but if there are more than one, how is their influence harmonized? And why does love of the unmoved cause movement as an imitation? Further, if the heavenly bodies have desire, they must have soul, and the movement of soul, which is thought, is better than rotary movement.

After a criticism of Plato and his successor, Speusippus (died 339 BC), for not carrying through their accounts of the nature of the world into a complete system, he suggests that perhaps metaphysics is only concerned with first principles. But then what are these principles? He appears to consider two kinds: the ultimate entities out of which things are made and the general laws by which all things are governed.

Turning to details, he considers various basic notions, form and matter as conceived by Aristotle, good and evil, being, and how various kinds of being may be known. He wonders why there is so much evil in the world. On a

lower level he considers cases which seem to lack final causes, such as floods and male breasts. These raise difficulties for Aristotle's claim that nothing happens without a purpose. Theophrastus suggests that perhaps these anomalies result automatically from the rotation of the heavens, or perhaps there is a limit to purposiveness and the desire for what is good. Recently discovered Arabic texts show him arguing that god is responsible only for order in the universe, and that things like thunderbolts are not sent by god.

It is possible that the later lack of interest in metaphysics in the Peripatos is due to Theophrastus' devastating criticism. Certainly his successor, Strato, was primarily interested in science.

5 Theology and religion

For Theophrastus' theological views, apart from what is mentioned above, we are dependent on later writers. Cicero (106-43 BC) says that he identified god with heaven and with mind or spirit, but his evidence is superficial. Albert the Great (thirteenth century) says that Theophrastus argued that god had no will, giving five detailed arguments (fr. 259) which may however owe much to Arabic developments, and that the relationship between god and the intellect is like light. Denis the Carthusian (1402-71) suggests that he was the founder of negative theology (fr. 260), saying that we know better what god is not than what he is. It also seems that he wrote about *daimons*, the beings supposed to exist on a level between gods and mortals (see Platonism, Early and Middle §7). Porphyry (mid 3rd-early 4th century) gives us long excerpts from a work which may be his *On Piety*, which are largely on what types of sacrifice are appropriate to the gods, and advocate vegetarianism (frs 584-5).

6 Ethics

Much of the evidence for his ethics is trivial, but Porphyry gives us sections from his attack on meat eating, probably slanted to suit Porphyry's own interests, and Jerome quotes from his very funny attack on marriage, in order to support Christian views on chastity (fr. 486). We also have Theophrastus' doctrine of the natural kinship of all human beings and, more, of human beings and animals, because of their physiological and psychological similarities - sometimes known as the doctrine of *oikeiotēs* (frs 531, 584), though it is still debated whether it is related to the *oikeiōsis* of the Stoics (see Stoicism §14); also his attack on Plato's account of false pleasure in the *Philebus* (fr. 556), where he stresses the importance of language. There is also evidence that he had new ideas about emotions (frs 438-48), in particular using the notion of 'the more and less', that is, differences in degree, to relate similar emotions such as anger and rage. He attained some notoriety among the Romans for holding a less austere view of how virtue and happiness are related than did Aristotle (and after him the Stoics), saying that a happy life does not lie in virtue alone (fr. 497). His three books on friendship had considerable influence, but little now survives (frs 532-46).

7 Science

The line between philosophy and early science is difficult to draw, and his work in botany and biology contains both empirical work based on observation and attempts at establishing theoretical foundations. The opening of his *Physics* gives an account of the fundamentals of natural science, that natural bodies are composite, and have elements and causes and principles, and he goes on to place (criticizing in detail Aristotle's odd account; frs 146-9), time (relating it, like Aristotle, to motion; frs 150-1), motion and change. His remarks on place, at least adumbrating a relativist view of space, have been given great significance by some scholars, but it is more likely that as usual he was raising more problems than he answered. Large sections of a work in which he stated and refuted in detail several arguments of his predecessors against the eternity of the universe are preserved by Philo of Alexandria (frs 184-5). He extended Aristotle's work in zoology with studies of the nature and habits of many creatures, and focused in particular on peculiar phenomena such as swarming and hibernation, trying to show that these had a natural explanation. In botany he applied to plants the principle of classification originally developed by Plato, but further developed by Aristotle. The opening of the *History of Plants*, or better, *Inquiry into Plants*, attempts to list the parts of plants, stem, root, leaf, flower and so on, and later there is a division of plants into trees, shrubs, undershrubs and herbs. In *On the Causes of Plants*, or better, *Plant Explanations*, he covers aspects of the growth of plants, such as methods of reproduction and climatic effects. He built on a mass of information gathered by others, including herbalists and agriculturalists, but also by his own observation, and provided a theoretical framework which lasted for centuries.

In addition there are several short treatises including *On Fire*, *On Stones*, *On Tiredness* and *On Sweating*, and a number of scientific works attributed to Aristotle but not thought to be by him which have by some been attributed to Theophrastus. These include *On Colours*, parts of the *History of Animals* VIII and IX, and parts of the *Mirabilia*.

8 Other work

Theophrastus has also been seen as an influential doxographer and the source of other later doxographies (see Doxography): his *On the Senses* has survived with its treatment of the views of the Presocratics and Plato about sensation, and we have evidence for similar work in physics (frs 224-45).

He followed Aristotle in working on politics (frs 589-665) and rhetoric (frs 666-713), particularly on style (frs 684-704), and wrote some difficult theoretical material on music (frs 714-26).

PAMELA M. HUBY

List of works

Theophrastus (*c*.372-287 BC) Fragments, in W.W. Fortenbaugh, P.M. Huby, R.W. Sharples and D. Gutas (eds) *Theophrastus of Eresus, Sources for his Life, Writings, Thought and Influence*, Leiden: Brill, 1992.(A collection of texts and translations relating to Theophrastus until the fourteenth century, including evidence for his work in logic, science, psychology, theology, religion, ethics, political theory and rhetoric; volume 5, *Sources on Biology*, ed. R.W. Sharples appeared in 1995.)

Theophrastus (*c*.372-287 BC) *Metaphysics*, trans. W.D. Ross and F.H. Fobes, Oxford: Clarendon Press, 1929; repr. Hildesheim: Olms, 1967; trans. M. van Raalte, Leiden: Brill, 1993.(Greek and English text, with notes.)

Theophrastus (*c*.372-287 BC) *Characters*, ed. and trans. J. Rusten Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 2nd edn, 1992.(Sketches of character weaknesses; parallel Greek text and English translations)

Theophrastus (*c*.372-287 BC) *Inquiry into Plants*, ed. and trans. A. Holt Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 2 vols 1916-26.(Botany; parallel Greek text and English translations.)

Theophrastus (*c*.372-287 BC) *Causes of Plants*, ed. and trans. B. Einarson and G. Link Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 3 vols, 1976-90.(Botany; parallel Greek texts and English translations.)

Theophrastus (*c*.372-287 BC) *On the Senses*, trans. in G.M. Stratton, *Theophrastus and the Greek Physiological Psychology before Aristotle*, London: Allen & Unwin, 1917; repr. Amsterdam: Grüner, 1964.(A study of earlier theories of perception. Greek text, translation and notes by an eminent cognitive psychologist)

References and further reading

Rutgers University Studies in Classical Humanities (1983-) New Brunswick, NJ, and London: Transaction Publishers.(Includes several volumes of conference proceedings relating to Theophrastus, especially volume 2, *Theophrastus of Eresus: On His Life and Work*; volume 3, *Theophrastean Studies: On Natural Science, Physics and Metaphysics, Ethics, Religion and Rhetoric*; and volume 5, *Theophrastus: His Psychological, Doxographical and Scientific Writings*.)

Huby, P.M. (1985) 'Theophrastus in the Aristotelian Corpus, with Particular Reference to the Biological Problems', in A. Gotthelf (ed.) *Aristotle on Nature and Living Things*, Bristol: Bristol Classical Press.(Surveys suggestions about works attributed to Aristotle which might have been written by Theophrastus.)

Theoretical (epistemic) virtues

When two competing theories or hypotheses explain or accommodate just the same data (and both are unrefuted), which should be preferred? According to a classical, purely formal confirmation theory, neither - each is confirmed to the same degree, and so the two hypotheses are precisely equal in epistemic status, warrant or credibility. Yet in real life, one of the two may be preferred very strongly, for any of a number of pragmatic reasons: it may be simpler, more readily testable, more fruitful or less at odds with what we already believe. The philosophical question is whether such pragmatic virtues are of no specifically epistemic, truth-conducing value, or are instead genuine reasons for accepting a theory as more likely to be true than is a competitor that lacks them.

1 Some of the virtues

The preference for simplicity in particular is illustrated by the example of experimental scientists' practice in curve-fitting on graphs: given a set of data points that lie along a straight line, any scientist will draw a straight line through them rather than any more complicated curve, and leave it that way unless further, refuting data should come in. This compelling smoothness of the linear hypothesis is a virtue of some sort, one that is not shared by the hypotheses expressed by the countlessly many more complex curves that pass through the very same data points and hence are just as well confirmed in the classical sense as is the linear theory. (The immediacy and vividness of the curve-fitting example should not be allowed to suggest that 'simplicity' is easily measured, or even that it can be given a clear general characterization - Foster and Martin 1966; Sober 1975 - see Simplicity in scientific theories)

Other pragmatic virtues having simplicity's nonclassical and doubtful epistemic character include: (1) *Testability* - other things being equal, a hypothesis H_1 will be preferred to a competitor H_2 if H_1 has more readily testable implications; (2) *Fertility* - H_1 will be preferred to H_2 if H_1 is more fruitful in suggesting further related hypotheses, or parallel hypotheses in other areas; (3) *Neatness* - H_1 will be preferred to H_2 if H_1 leaves fewer messy unanswered questions behind, and especially if H_1 does not itself raise such questions; (4) *Conservativeness* - H_1 will be preferred to H_2 if H_1 fits better with what we already believe. (If this sounds dogmatic, notice that, inescapably, we never even consider competing hypotheses that would strike us as grossly implausible; for example, no detective considers the hypothesis that the crime was committed by invisible Venusian invaders.) Also, often listed as a pragmatic virtue is 'generality' or *explanatory power*, the property of comprehending more various data than one's competitors; but we have stipulated for the sake of argument that our competing theories comprehend just the same data, and in any case it seems that power is just a wider, more global simplicity.

Every pragmatic virtue is a matter of degree. And our preference for any one of them always comes qualified by 'other things being equal', for clearly they can conflict among themselves. Perhaps the most obvious tension holds between simplicity and conservativeness, since often simplification is gained only through bold overthrowing of previously accepted theory, as in the case of Copernican versus Ptolemaic astronomy. But because of the complexity of any detailed real-world case study, there is no generally accepted policy for weighing the various degrees of the various virtues against each other in any particular inquiry. The lack of such a policy, understandably, has led some epistemologists to scepticism or relativism concerning theory choice based on pragmatic virtues. But there is also a more fundamental issue, outlined in section two below.

2 The spartan argument

It seems we have considerations of two sorts: data or hard evidence bearing a quasi-formal probabilifying or confirming relation to each of the competing hypotheses H_1 and H_2, and the pragmatic virtues attaching differentially to H_1 and H_2. The former is expressed commonly by words like 'likely', 'probable' and 'confirms' in our narrow sense; the model for it is formal inductive logic or confirmation theory based on probability theory, conceived by the Logical Positivists as a strict analogue to deductive logic (see Confirmation theory). Let us call this 'narrow confirmation'. Following mainstream epistemology, let us also use the words 'justified' and 'justification' respectively to mark out the class of beliefs it is epistemically rational to hold, and the relation

between those beliefs and the evidence in virtue of which holding them is rational. The question then arises, ‘is justification exhausted by narrow confirmation?’

The affirmative is emphatically urged by van Fraassen (1980). On that spartan view, the pragmatic virtues are merely pragmatic, a matter of what is utile or convenient and (in Ian Hacking’s phrase) makes our minds feel good; the virtues contribute not at all to justification. It is the spartan view that drives van Fraassen’s and others’ radical scepticism about the unobservable. Unobservable entities are posited or hypothesized at best, so if hypothesis-choice requires appeal to pragmatic virtues and the virtues do not ever make for justification, no belief in unobservables can be justified. It is also the spartan view that drives evil-demon scepticism about the external world: by hypothesis, the evil-demon theory makes exactly the same observational predictions as does the realist external-world theory, so both are equally probable or well-confirmed on our evidence, so we have no reason to believe the external-world theory to the exclusion of the other. Further, the spartan view drives Nelson Goodman’s famous ‘Grue Paradox’, for the problem is that for every well-confirmed hypothesis, say that the next emerald to be observed after a certain future time *t* will be green, there are innumerably many truth-functionally mocked-up competitors that are equally well confirmed, say that the next emerald will be ‘grue’ (either examined before *t* and green, or not examined before *t* and blue) (see Induction, epistemic issues in; Knowledge, concept of §9).

Since the spartan position gives rise to such sceptical quandaries, why not abandon it forthwith and admit the pragmatic virtues’ justificatory power? But one must not simply presume the falsity of scepticism in the face of arguments that to many philosophers have seemed unanswerable. Also, there is a very attractive direct argument against taking the virtues as justificatory.

A first formulation of the argument is this: the virtues’ pragmatic ‘utility’ is really a mixture of corner-cutting convenience - a form of intellectual laziness - and merely aesthetic appeal. We prefer the smoothest hypothesis in curve-fitting because the smooth curve is easier to draw and looks prettier. But why should anyone think that convenience and prettiness count in any way towards truth? Few philosophers would share the Grecian Urn’s motto that beauty and truth are one.

The spartan argument is most compelling for the case of conservativeness. That a hypothesis fits comfortably with what we already believe makes that hypothesis pleasant and attractive to us, but hardly justifies it; for to think it does is to assume that what we already believe is justified merely by the fact of our believing it, and that idea strikes most philosophers as blatantly false - however, see Sklar (1975) and Lycan (1988). The spartan argument is memorably embellished by van Fraassen:

> Judgements of simplicity and explanatory power are the intuitive and natural vehicle for expressing our epistemic appraisal. What can an empiricist make of these… virtues which go so clearly beyond the ones he considers pre-eminent?
>
> There are specifically human concerns, a function of our interests and pleasures, which make some theories more valuable and appealing to us than others. Values of this sort, however… cannot rationally guide our epistemic attitudes and decisions. For example, if it matters more to us to have one sort of question answered rather than another, that is no reason to think that a theory which answers more of the first sort of question is more likely to be true.
>
> (van Fraassen 1980: 87)

Truth is a relation between a theory or hypothesis and the world. But the pragmatic virtues are relations between theories and our human minds, to which relations the world seems irrelevant. The virtues have to do with the roles that hypotheses play in our private cognitive economies, not with anything external to us. Making our minds feel good is hardly a warrant of truth.

Notice that, if correct, the spartan position undermines an entire theory of epistemic justification, that which in this century originated with Quine (1960) and with Sellars (1963, 1973). It is the view that what justifies some (or all) beliefs is those beliefs’ felicitous explanatory relations to others, and since the goodness of an explanation is measured almost entirely by its pragmatic virtues, they are indispensable to the explanationist theory of justification (see Inference to the best explanation).

3 Replies to the sceptic

Two replies may be made to the embellished spartan argument. First, it falsely assimilates the pragmatic virtues to crassly self-seeking 'reasons' for believing things (as in Pascal's Wager, or a case in which we are offered money if we can get ourselves to believe that Mrs Thatcher was Britain's greatest leader ever). The virtues are at least genuinely *cognitive* and, in one important sense, epistemic values. It is fairly easy to see that truth cannot be the only epistemic value. For suppose it were - if the idea, like Descartes', is merely to avoid falsehood, then we could reach our ultimate epistemic goal simply by confining our assent to tautologies; if instead the idea is to believe *all* truths, the goal would be radically unreachable. Fully realizing those things, the truth-centred epistemologist usually alludes to a 'favourable balance of truth over error'. But 'favourable' as regards what? Some further value or interest must be consulted to judge what is 'favourable'. Also, more specifically, it is hardly unreasonable to suppose that beliefs are *for* something, and that cognition has a function. Truth is often called 'the goal of' cognition, but we have just seen that truth cannot be the only goal of cognition; there must at least be something in the way of informativeness, however it might be measured. Since belief is a guide to action, a belief's other pragmatic virtues may also contribute to its overall cognitive goodness. And now the burden is on the spartan to say why justification is constituted by only a subcomponent of overall cognitive goodness, in the face of the sceptical quandaries expounded in §2.

Second, we should advert to the epistemology of epistemology. Epistemology is a study of norms, and (so the present reply goes) the epistemology appropriate to a normative subject is that of 'reflective equilibrium' as first proposed by Goodman (1955) and developed by Rawls (1971). Roughly, we begin with our instinctive normative intuitions and build an accordingly normative theory to systematize them. Mutual adjustment occurs until what Rawls calls 'narrow' reflective equilibrium is reached; then factual knowledge and perhaps also other norms are admitted to the equation, resulting in further adjustment and eventual 'wide' reflective equilibrium. In epistemology, as in ethical theory and for that matter in deductive logic, the intuitions in question are normative to begin with.

On the present view, epistemology starts with the attempt at narrow reflective equilibrium. The move to wide equilibrium will involve attending to empirical cognitive science. But each is likely to respect the pragmatic virtues. For our pragmatic preferences are not merely preferences, but normative practices: we instruct our science students in the techniques of curve-fitting, epicycle elimination and the like, and science would be in a bad way if we did not.

One may or may not be persuaded by the foregoing two rebuttals of spartan scepticism. If not, and if one does not want simply to acquiesce in a radical scepticism about all of nature, one has the task of showing how a classical confirmation theory can overcome the sceptical quandaries without at least tacit appeal to the pragmatic virtues. If one is persuaded, there are three paths: to modify classical confirmation theory in order to respect the virtues; to relegate classical confirmation theory to a confined role in inquiry, granting that justification far outruns narrow confirmation; or to abandon confirmation theory entirely as a bad job.

Roughly the first path has been taken by Clark Glymour (1980). The idea is to broaden the notion of 'confirmation', so that H_1 may be counted as 'better confirmed' than H_2 even though H_1 and H_2 are still equally probabilified by the evidence base. If, however, one wants to demote or abandon confirmation theory entirely, one faces the daunting task of building a systematic account of the pragmatic virtues, their measurement and their comparative interaction - to date, no theorist has taken more than a step or two in that direction.

See also: Justification, epistemic; Scepticism

WILLIAM G. LYCAN

References and further reading

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Oxford University Press. (Attacks the pragmatic virtues conceived as justificatory.)

Foster, M.H. and Martin, M.L. (1966) *Probability, Confirmation, and Simplicity*, New York: Odyssey Press. (Contains some classic papers on simplicity. R. Ackermann's 'Inductive Simplicity' in particular expresses pessimism as regards the taming of that notion.)

Fraassen, B. van (1980) *The Scientific Image*, Oxford: Oxford University Press. (Attacks the pragmatic virtues conceived as justificatory.)

Glymour, C. (1980) *Theory and Evidence*, Princeton, NJ: Princeton University Press. (Tries to broaden narrow confirmation in the direction of the pragmatic virtues.)
Goodman, N. (1955) 'The New Riddle of Induction', in *Fact, Fiction, and Forecast*, Cambridge, MA: Harvard University Press.(Proposes the method of reflective equilibrium, and expounds the Grue Paradox.)
Hacking, I. (1982) 'Experimentation and Scientific Realism', *Philosophical Topics* 13 (1): 71-88.(Attacks the pragmatic virtues conceived as justificatory.)
Lycan, W.G. (1988) *Judgement and Justification*, Cambridge: Cambridge University Press.(Part 2 defends an explanationist epistemology; chapter 7 defends the pragmatic virtues as justificatory, and chapter 8 makes the case for conservativeness in particular.)
Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Seminal work in which Quine brings his own brand of empiricism to bear on epistemic issues.)
Quine, W.V. and Ullian, J.S. (1970) *The Web of Belief*, New York: Random House, 2nd edn, 1978.(Chapter 5 discusses and defends some representative pragmatic virtues.)
Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press. (Develops the theory of reflective equilibrium.)
Sellars, W. (1963) 'Some Reflections on Language Games', in *Science, Perception and Reality*, London: Routledge & Kegan Paul.(The locus classicus of Sellars' explanationist epistemology.)
Sellars, W. (1973) 'Givenness and Explanatory Coherence', *Journal of Philosophy* 70 (18): 612-24.(Further development of Sellars' view.)
Sklar, L. (1975) 'Methodological Conservatism', *Philosophical Review* 84 (3): 374-400.(A qualified defence of conservativeness as justificatory.)
Sober, E. (1975) *Simplicity*, Oxford: Oxford University Press.(Standard work on the subject.)
Thagard, P. (1978) 'The Best Explanation: Criteria for Theory Choice', *Journal of Philosophy* 75 (2): 76-92. (Discusses and defends three of the pragmatic virtues.)

Theories, scientific

The term 'theory' is used variously in science to refer to an unproven hunch, a scientific field (as in 'electromagnetic theory'), and a conceptual device for systematically characterizing the state-transition behaviour of systems. Philosophers of science have tended to view the latter as the most fundamental, and most analyses of theories focus on it.

The Einsteinian revolution involved the rejection of the chemical ether on experimental grounds. It thus prompted philosophers and scientists to examine closely the nature of scientific theories and their connections to observation. Many sought normative analyses that precluded the introduction of 'fictitious' theoretical entities such as the ether. Such analyses amounted to criteria for demarcating scientific or cognitively significant claims from unscientific or metaphysical claims.

Logical positivism sought to develop an ideal language for science that would guarantee cognitive significance. The language was symbolic logic with the nonlogical vocabulary bifurcated into observational and theoretical subvocabularies. Observation terms directly designated observable entities and attributes, and the truth of statements using them was unproblematic. To prevent postulation of fictitious unobservable entities, theoretical terms were allowed only in the context of a theory which guaranteed the cognitive significance of theoretical assertions. Theories were required to contain correspondence rules that interpret theoretical terms by coordinating them in some way with observational conditions.

In the 1960s this 'received view' was attacked on grounds that the observational-theoretical distinction was untenable; that the correspondence rules were a heterogeneous confusion of meaning relationships, experimental design, measurement and causal relationships; that the notion of partial interpretation associated with more liberal requirements on correspondence rules was incoherent; that theories are not axiomatic systems; that symbolic logic is an inappropriate formalism; and that theories are not linguistic entities.

Alternative analyses of theories were suggested - construing theories as answers to scientific problems or as paradigms or conceptual frameworks. Gradually analyses that construe theories as extra-linguistic set-theoretic structures came to dominate post-positivistic thought. The semantic conception *identifies theories with abstract theory structures like configurated state spaces that stand in mapping relations to phenomena and are the referents of linguistic theory formulations. Depending on the sort of mapping relationship required for theoretical adequacy, realist, quasi-realist or antirealist versions are obtained. Correspondence rules are avoided and some versions eschew observational-nonobservational distinctions altogether. Development of the semantic conception has tended to focus on the mediation of theories and phenomena via observation or experiment, the relations between models and theories, confirmation of theories, their ontological commitments, and semantic relations between theories, phenomena and linguistic formulations. The* structuralist *approach also analyses theories set-theoretically as comprised of a theory structure and a set of intended applications, but is neopositivistic in spirit and in its reliance on a relativized theoretical-nontheoretical term distinction. It has been used to explore theoreticity, the dynamics of theories as they undergo development, and incommensurability notions.*

One's analysis of theories tends to influence strongly the position one takes on issues such as such as observation, confirmation and testing, and realism versus instrumentalism versus antirealism.

1 Background

Part of the mathematically specified content of special relativity cannot be reduced to observations. Logical positivism emerged out of attempts to reconcile that feature with Mach's doctrine that observations exhaust the empirical content of a theory. Exploiting recent developments in logic and the foundations of mathematics such as *Principia mathematica*, the logical positivists concluded that mathematics was tautological and so mathematical portions of scientific theories did not add to their empirical content - a reconciliation called into question by their own later work on pure versus applied geometries.

Joining with this effort was a strongly Neo-Kantian approach to epistemology and metaphysics. Carnap 's *Aufbau* (1928) essentially reworked Kant's 'transcendental analytic': the unique categories and their principles of the metaphysical deduction were replaced by the notion of a nonunique constructional system (*Konstitutionstheorie*)

and the transcendental ego, by a physical person constructed out of subjective experiences. Carnap maintained that this particular constitutional scheme was not unique, and thus that realism, idealism, and his favoured constructional scheme were just different ways of speaking. Insofar as such schemes were valid they were in agreement. The disadvantage of realist or idealist schemes was that they contained language that allowed metaphysical assertions which were not cognitively significant. The advantage of a logically ideal scheme was that only what was cognitively significant could be asserted. Therefore anything legitimate in any other scheme could be translated into a logically ideal scheme. Philosophical analysis became canonical formulation in the ideal syntactical language.

2 Syntactical approaches: the 'received view'

Syntactical approaches analyse theories as *linguistic entities* which are individuated by their syntactic characteristics. Thus alterations in the syntax of a theory (say, by changing the axioms or the language) result in a different theory.

Whitehead and Russell's *Principia mathematica* did not clearly distinguish syntax from semantics, assuming syntactical axioms had a given standard interpretation, and so early syntactical investigations were of interpreted sentences. Thus when positivists syntactically distinguished the observational and theoretical components of a scientific language, the observational portion was presumed to be interpreted; however, the problem of fictitious theoretical entities rendered problematic whether theoretical assertions genuinely referred.

Following constructions in the *Aufbau*, Rudolf Carnap initially construed the observational sublanguage as a sense datum language. Later he switched to the more realistic 'thing' language. Observational assertions were statements about directly observable entities and their attributes. Carnap never attempted rigorous delineation of the observational-theoretical distinction because he viewed it as somewhat arbitrary and probably just a matter of language choice. In the terminology of his 'Empiricism, Semantics, and Ontology' (1950), choice of a language is an external question, but truth is an internal question that can only be raised in the context of a specific language system. Thus demarcation of the observation language from the theoretical language is an external question.

A logically ideal scientific language trifurcated its vocabulary into the logical, *observational* V_O and *theoretical* V_T. The observational sub-language consisted in sentences using just V_O augmented by the logical, and the theoretical sublanguage utilized V_T and logical terms. *Mixed sentences* contained both observational and theoretical terms. Observation language sentences were fully interpreted, whereas theoretical assertions depended on the theory for their interpretation. The laws, *T*, (axioms in the theoretical language) *implicitly defined* their terms by imposing relationships that restricted potential interpretations. Mixed language *correspondence rules C* infused observational meaning into the theoretical terms, so that the interpretation of theoretical assertions was a function of implicit definition and infused operational meaning. The theory consisted of the conjunction *TC*.

Since implicit definition restricted but did not fix referents for theoretical terms, the empirical referents of theoretical assertions depended essentially on the standard interpretation of V_O mediated via *C*. Initially the *C* were *explicit definitions* that identified the content of theoretical assertions with complex observable conditions. Existence of alternative procedures for defining the 'same' concept and problems in capturing the logic of dispositional terms using symbolic logic led to the more liberal requirement that the *C* be *reduction sentences* partially defining theoretical terms in the context of a particular experimental set-up as characteristic outcomes to stimuli. Difficulties in giving observational interpretations to theoretical terms such as the ψ function in quantum mechanics and real valued measures led to allowing *C* to be *interpretive systems* that do not individually coordinate V_T terms with observable conditions, instead requiring that their inclusion in a theory make a difference in the theory's observable predictions.

Figure 1 will help clarify the differences in these proposals. When the *C* were required to be explicit definitions, theoretical assertions were guaranteed to refer to observable reality within *TC*'s scope. Once reduction sentences or interpretative systems were allowed, V_T terms could refer to observables or to nonobservables. That is to say that the models of *TC* included both. A general problem was that the Löwenheim-Skolem theorem implied that the models of *TC* included both intended models and wildly unintended ones (see Model theory). Unintended models are the source of potential counterexamples. Blocking them more concerns eliminating artifacts of the syntactical approach than it does dealing with substantive content of the analysis. Positivistic syntactical analyses of theories,

confirmation and explanation persistently were plagued by problems of unintended models.

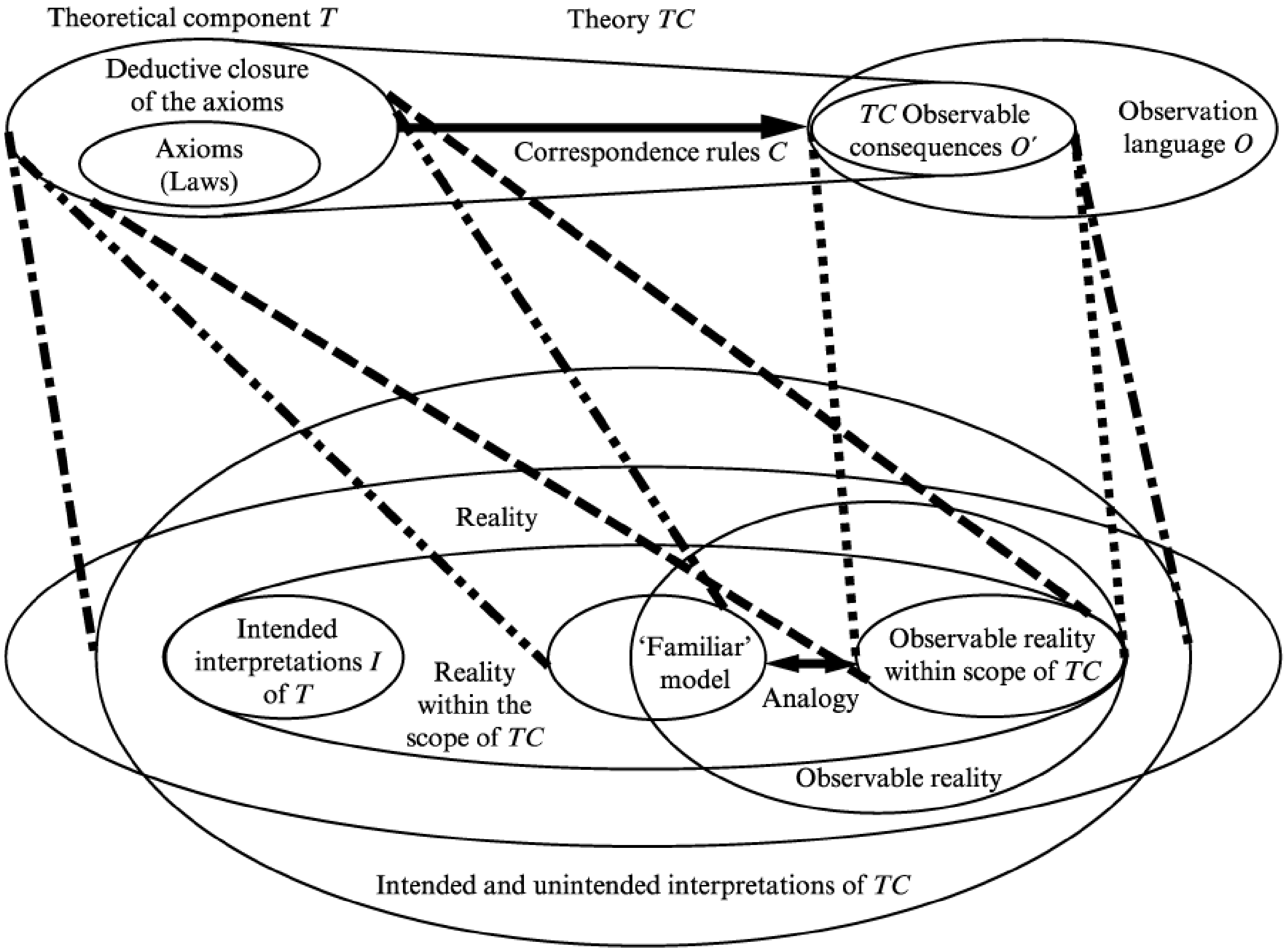

Figure 1. Semantic properties of the received view on theories under explicit and partial definition versions and when augmented by 'familiar' analogical models.

Key:consequences O' to pick out observable reality within the scope of *TC*.
TC and the interpretation of O' cause the theory component *T* to pick out observable consequences redundent of those O' pick out.
C are reduction sentences or interpretive systems that only partially define theoretical terms in *T*, the allowed interpretations of *T* include not only the intended interpretations of *T* but many wildly unintended and unrealistic interpretations as well. Further the choice of the correspondence rules *C* that partially define *T* is alleged by Campbell and Hesse to be highly irrational.
T must admit of a second interpretation in terms of a 'familiar' model that stands in an analogical relationship to the observable portion of reality within the intended scope of theory *TC*. This analogy rationally suggests plausible candidates for correspondence rules *C*, but does nothing to solve the problem of wildly unintended interpretations for *T*.

3 Instrumentalism versus realism

The arguments that prompted Carnap and C.G. Hempel to move from explicit definitions to reduction sentences or interpretive systems as preferred correspondence rules *C* are persuasive only within a realistic perspective. Those who resisted realism turned to instrumentalism to accommodate the undeniable benefits of theory not reducible to observation. Ramsey proposed that theoretical content be construed as a second-order existential quantification of theoretical laws which entailed the observable consequences of the theory. This, of course, only asserts that *some* system within the intended and unintended interpretations of Figure 1 yields observational consequences O' and thus does little to avoid realism, other than being noncommittal.

Craig's theorem asserts that whenever we bifurcate the nonlogical vocabulary of a theory *T* into V_T and V_O there is a new theory T_O using only the sub-vocabulary V_O that has precisely the T_O theorems. Thus theoretical assertions are alleged to be eliminable in science. Unfortunately T_O has an infinite axiom set whereas positivists construe theories as being finitely axiomatizable, and so this theorem's relevance is suspect. With the gradual liberalization of *C*-rules the received view became increasingly realistic. Nevertheless there is an artificiality to the realistic versus instrumentalistic debates since they erroneously assume that either all the terms in a theory are interpreted realistically or none are.

4 Demise of the received view

Logical positivists were concerned to accommodate real science and whenever they found incompatibilities with uncontroversial scientific achievement they modified their analyses accordingly. In practice this meant that they adjusted their analyses to accommodate the best physics had to offer, but deployed them normatively in evaluation of the social sciences. Positivists, especially Carnap and Hempel, were paragons of intellectual integrity, relentlessly subjecting their own analysis to scrutiny and revelling in the discovery of defects that required modifications.

Such internal critique did not challenge the presuppositions of their enterprise, but a series of external critiques in the 1950s and 1960s did. Most fundamental were the attacks on the observational-theoretical distinction by Achinstein and Hilary Putnam who charged that the distinction could not be drawn in a satisfactory manner. Later Suppe (1989) showed that these arguments established only that such a distinction could not be drawn on the basis of natural language usage and that, if artificially drawn, did not mark an epistemologically significant distinction. Achinstein and Putnam also charged that the notion of 'partial interpretation' associated with more liberal *C*-rules was incoherent; later Suppe (1989) gave a model-theoretic construal compatible with positivism, once it was realized that not all interpretation had to be via the observation language.

Putnam and Achinstein's attacks prevailed and were joined by new historically based accounts of theorizing that semi-explicitly rejected the received view. Hanson argued that observations were theory-laden - from which it followed that the observational-theoretical distinction was untenable. P.K. Feyerabend noted that observations and theories use the same descriptive vocabulary.

Thomas S. Kuhn's *Structure of Scientific Revolutions* is the classic reworking of positivism. He rejected the positivistic observational-theoretical distinction, instead embracing a version of N.R. Hanson's claim that observation is theory-laden. He denied that the connections between theory and phenomena can be mediated by any *explicit* set of correspondence rules. Instead exemplary applications are studied as one becomes a scientist and subsequent applications are modelled on these exemplars, and both phenomena and theory use the same vocabulary. His account of how, in revolutionary episodes, science switches 'paradigms' or organizing frameworks closely resembles Carnap's views on the arbitrariness of choosing a particular language or conceptual framework. In particular, both agree that the choice cannot be assessed on the basis of truth determination.

Schaffner (1969) charged that positivist *C*-rules conflated meaning relations, experimental procedures and causal relations. Thus they lumped together components that were epistemically distinct. Suppes (1962) argued that the *C*-rules account failed to reveal what was epistemologically central to the application of theories to phenomena. Suppe (1977) argued that the inclusion of *C*-rules as proper parts of theories resulted in improper individuation criteria for theories since the received view required that, contra actual scientific practice, a change in experimental procedure was a change in theory. The relevance of formalist approaches was challenged by Toulmin and others including, surprisingly, Hempel (in Suppe 1977).

Increasingly positivists had become sidetracked from the development of their substantive ideas by technical problems that were mere artifacts of their strategy of modelling science via predicate calculus axiomatizations. Suppes (1967) contrasted their 'intrinsic formulation' approach to the 'extrinsic' approach wherein one directly designates an intended family of models. Although formal semantics is incapable of designating precisely the intended set of models, our ordinary linguistic resources suffice. In practice we do pick out just intended models (else there would not be a problem of unintended models), and so we can proceed directly to the specification of these models without recourse to syntactic axiomatization. What is surprising about this line of criticism is that Carnap and others were seminal developers of formal semantics, yet that work had little impact on their approach

to formal analysis of theories which, to the end, was aggressively syntactical.

Under the weight of these objections the consensus emerged that the received view was fundamentally untenable.

5 The semantic conception

Questions such as whether Schrödinger's wave mechanics and Heisenberg's matrix mechanics are the same theory were resolved when von Neumann showed that they both admitted of the same canonical Hilbert space representations. That and a related paper construing theories as configurated state spaces mapped onto observation spaces inspired the development of the semantic conception. The basic idea is that a theory is identified by a suitably connected family of models.

Everet Beth used set-theoretic techniques to analyse several specific theories, speculated about the potential for semantic analyses, and pioneered the semantics of amplified usage wherein the same formalism can refer variously to different set-theoretic structures.

In the early 1950s Patrick Suppes, J.C.C. McKinsey and others investigated foundational issues in physics via set-theoretic axiomatizations. Generalizing from these studies Suppes argued that theories were extra-linguistic set-theoretic structures and offered a general structural analysis. One argument for this semantic approach was that extrinsic presentation of theories avoided the problem of blocking unintended interpretations that had so diverted positivism from its philosophical agenda. Suppes (1962) explored theory structures associated with phenomena in experimental circumstances (see Measurement, theory of). Drawing on his mathematical learning theory laboratory experience, he argued that the connection between theory and experiment is mediated via a hierarchy of models including the model of the experiment, models of data, experimental design, and *ceteris paribus* conditions.

Drawing on his jet-engine flight-test research experiences, Suppe (1989) argued that correspondence rule accounts bore little resemblance to how theories attached to phenomena or how theories were tested. Analysing theories as essentially describing state transitions in a 'state space', he argued that propounding a theory consisted in asserting a suitable mapping relation between the configurated state space and systems within the theory's scope. Because theories ignored all but a selected finite number of variables, the relation had to be counterfactual wherein the theory structure specified how systems within the theory's scope *would* behave *were* they isolated from influences not showing up as variables in the theory.

The counterfactual interpretation of theories has implications for how theories are confirmed by experiment. Characteristic of experimental design is control of extraneous variables so that only variables of the theory are influential. Suppe viewed the logic of confirmation as turning essentially on issues of experimental control, and offered a rigorous epistemology where the adequacy of theories under controlled circumstances licensed the generalized conclusion that the theory was counterfactually true of *all* systems within its scope. On Suppe's epistemology, observational knowledge of one carefully crafted experimental instance is sufficient to establish a scientific theory non-inductively.

Whereas Suppes and Suppe have concentrated on the experimental mediation of theory to experiment, van Fraassen's version of the semantic conception focuses on ontological issues. His investigations have focused on the semantic relations between theories, their formulations, and reality. His starting point is the supposition that not everything referred to in a theory engenders ontological commitment. The problem is to give full semantic interpretations of theoretical language without excessive ontological commitments. His solution is the theory of *semi-interpreted languages* wherein languages are interpreted as referring to 'logical spaces' which provide a full semantic interpretation for the language of theorizing. However ontological commitments are left unconstrained, being a matter of which points in the logical space one wishes to ontologically commit to. Such commitments are made via individual mapping functions (called *Loc* functions) from the real-world objects to points in logical space. For van Fraassen scientific theories are just suitably configurated logical spaces.

Figure 2 illustrates the main features of the semantic conception and relations between prominent versions. Detailed examinations of biological theory using the semantic conception have been made by Lloyd (1988) and Thompson (1989), and of quantum theory by van Fraassen (1991).

In contrasting semantic and syntactical approaches, some people refer to the latter as 'the statement view'. That is misleading since some philosophers view statements as linguistic entities having a syntax, and others do not. When

theories are characterized as collections of statements construed extra-linguistically, one is presenting a semantic analysis. Proponents of the semantic conception view such statements as structurally too problematic or impoverished to shed much illumination on the structure of theories.

6 Realism versus antirealism

Since theories are extra-linguistic entities, realism-instrumentalism disputes have to be recast. On the semantic conception theory, structures specify state-transitions where states are *n*-tuples of simultaneous values for the theory's variables or, in theories like quantum mechanics and natural selection, probability distributions over such values. Realism debates concern the nature of the mapping relationship between theory structure and the world (see Figure 2).

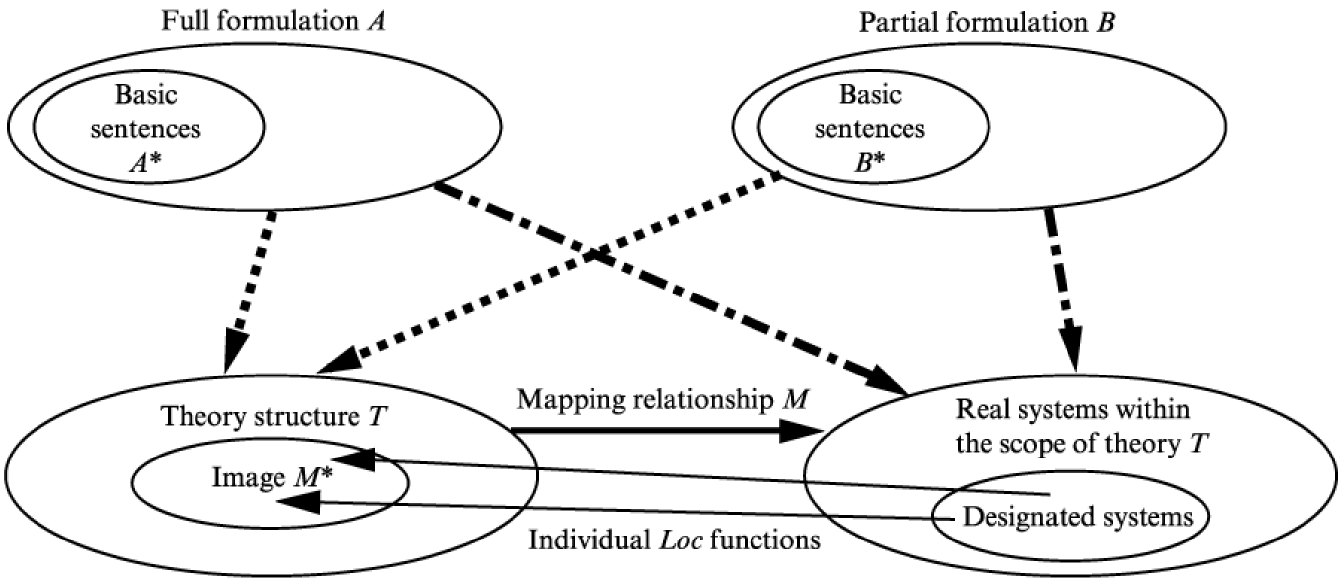

Figure 2. The semantic conceptions of theories.

Here theory structure *T* consists of precisely the intended interpretations of formation *A*. Note that no observational-theoretical interpretation of the theory formulation languages *A* or *B* are presupposed nor are there any correspondence rules as proper parts of the theory. Theory-structure *T* provides a full semantic interpretation the formulation languages *A* and *B* which are also interpreted real systems within *T*'s scope. Theories are asserted to stand in some mapping relationship *M* to real systems within the scope of the theory. On a realistic version, *M* would be a homomorphism; on Suppe's quasi-realistic version, *M* would be a counterfactual relationship specifying how the real systems would behave were they isolated from influence by variables not in *T*; on van Fraassen's constructive empiricism the mapping *M* is between a designated subset of the real systems and its image *M** under individual *Loc* functions. When *M** is contained in *T* then *T* is said to be empirically adequate. The *Loc* functions specify the ontological commitments one makes in asserting the theory *T*. On Suppes' version, the mapping relation *M* is mediated by a hierarchy of models including models of the experiment and models of data. Experimental design, instrumentation and so on, are not proper parts of theories but are used to determine whether mapping *M* does hold.

Realism maintains that theories give literally true accounts of reality. That means that one is ontologically commited to all the state variables. A theory is empirically true just in case the allowed state-transitions in the theory structure are identical to those which could occur in the actual world.

Van Fraassen's *antirealism* does not commit ontologically to all state variables. On his scheme a realist will have *Loc* functions onto every state variable, whereas he only countenances *Loc* functions from observables. Let *W* be that portion of reality to which one attaches *Loc* functions. Then according to van Fraassen a theory is *empirically adequate* precisely if the image *M** of *W* is among the models comprising the theory. Although van Fraassen resorts to an observational-nonobservational distinction to motivate his analysis, it works for any vocabulary bifurcation. The result is a general account of how language can have full semantic interpretations without generating ontological commitments to every semantically utilized entity. Van Fraassen is particularly concerned

about the fact that at various junctures science is irreducibly modal, yet it seems excessive to commit ontologically to unreal possible worlds. Semi-interpreted languages provide an escape from such ontological commitment. Within the range of ontological commitment empirical adequacy is equivalent to empirical truth. His antirealism thus is just realism attenuated to the range of one's ontological commitment.

Suppe questions whether any theories meet realistic truth or antirealistic empirical adequacy requirements since the scope of a theory includes systems influenced by variables not taken into account by the theory. Real world state-transitions often diverge from those specified in theory structures - thus making the theory false or empirically inadequate. In actual practice we only test theories against systems isolated under experimental control. Suppe's *quasi-realism* consists in ontological commitment to all variables that can be detected, together with the claim that empirically true theories provided counterfactual characterizations of how systems would behave were they isolated from influences not taken into explicit account by the theory. One could develop quasi-realisms with van Fraassen-like restricted ontological commitments.

Van Fraassen's semi-interpreted languages provide a general framework for evading excessive metaphysical commitment. (Thus, Suppe who is far more realistic, utilizes van Fraassen's semi-interpreted languages in his treatment of the relations between theory structure and linguistic formations, and in the analysis of empirical probabilities and causal modalities.) It accommodates, but does not require, commitment to van Fraassen's *antirealism*.

Although van Fraassen's enterprise is antimetaphysical in a manner not unlike Carnap's 'Empiricism, Semantics, and Ontology', neither he nor the other developers intend the semantic conception to demarcate the scientific from the unscientific or the real from the metaphysical. Indeed, semi-interpreted languages show that issues of ontological commitment are independent of the structure of theories, though not of one's epistemology.

7 Models and theories

For positivists theoretical terms have no empirical referents until they receive infused meaning from correspondence rules. How does one rationally choose reduction sentences or interpretive systems if theoretical terms have no independent meaning? N.R. Campbell argued that the choice was irrational unless there were some interpretation of the theory's axioms to guide you. He proposed that theoretical terms should be given an independent semantic interpretation in terms of concrete familiar models that are analogous to systems in the observable portion of the theory's scope (see Figure 1). Correspondence rules that mirror the analogy then would be tried. Later followers like Harré and Hesse argued that such models were what made theories explanatory.

To accommodate theories like quantum mechanics, which does not admit of concrete familiar models, they allowed that the mathematical formalisms (presumably the structures not the equations) themselves could be the models. The proposal did nothing to solve the problem of unintended interpretations of the theory. Little illumination as to the nature of models resulted beyond the notion that they were analogical and hence akin to metaphors.

Suppes noted that the word 'model' variously refers to metamathematical models (set-theoretic structures used to interpret axiom systems), scientific theories, and physical or scientific models. He argued that the latter two sorts can be analysed in terms of metamathematical models. The semantic conception does so for scientific theories.

Although the philosophical literature on scientific models has focused on familiar concrete models used heuristically or along Campbell's lines, models are often vehicles of scientific knowledge, not mere analogues or metaphors. Data typically are presented as a *model of the data* into which the raw data have been embedded and structurally enhanced. Typically the model is some sort of configurated hyperspace. *Simulation modelling* is increasingly an alternative source of experimental data. Simulation and other *dynamic models* are state-transition systems used to probe or model real-world systems. Thus both theories and models consist of configurated mathematical spaces or structures standing in mapping relations to other systems, and so the structure of various scientific models can be studied along the lines of the semantic conception. Fair progress analysing dynamic and simulation models has been made by Burks and Suppe. As other sorts of scientific models are studied and given semantic analyses, we should come to the point where systematic comparative investigations of the scientific models and scientific theories can be given.

8 Dynamics of theories

One reason for rejecting the positivists' analysis was that theories were improperly individuated since experimental procedures were incorporated into correspondence rules. The semantic conception also is inadequate if it cannot properly individuate theories.

Theories do undergo development. Lakatos construed this as a progression of static theories within a research programme. Others maintain that theories are dynamic entities that expand their scope and undergo development. Developing a dynamic account where partial or sub-models individuate theories is central to Sneed's *structuralist* analysis which began as an application of Suppes's set-theoretic techniques to the problem of theoretical terms. Sneed noticed that determination of mass and force function values when applying classical particle mechanics must invariably utilize classical particle mechanics itself and claimed this is the characteristic mark of theoretical functions or terms. Theoretical functions threaten to make the predictive testing of a theory viciously circular unless their contribution to the empirical content of a theory is restricted.

Sneed's solution is that the empirical content of a theory varies from application to application, and so the portion of the theory used in an application to calculate theory functions is not part of the empirical content of *that* application of the theory. Yet theories have a unified content that cannot be accounted for by a conjunction of statements. More precisely: a theory consists of a theory structure *T* and a set *I* of intended applications that are physical systems. Let M_p be the possible physical systems, and let M_{pp} be the partial possible systems obtained by function *r* that strips the M_p of their theoretical functions. The theory's empirical content is expressed by the statement that there exists a class of extensions *E* of the M_{pp} obtained via the addition of suitably constrained theoretical functions, such that these extensions *E* all are models in *T*. The analysis here is a set-theoretic modification of the Ramsey sentence (§3).

It follows that theories cannot be individuated on the basis of theory structure *T*. Sneed generalizes the notion of the empirical content of a theory to a set-theoretic analysis of the *core* $T_C = \langle M_p, M_{pp}, r, M, C \rangle$ of the theory structure *T*, where *M* is the law of the theory and *C* is a set of constraints. Although *I* is a part of the theory $\langle T, I \rangle$, individuation of the theory only depends on a core I' of paradigmatic intended applications. Theories are individuated on the basis of T_C and I'.

Sneed's analysis of the dynamics and individuation of theories is driven by his formulation of the problem of theoretical terms, which depends crucially on the supposition that the testing of a theory requires a prediction and an independent determination of whether the prediction is correct. If, as a growing number of philosophers and historians maintain, prediction is not central to testing or confirmation, Sneed's analysis is unlikely to be compelling. More generally, although set-theoretic apparatus makes this a semantic approach, its perspective is neo-positivistic, being wedded to a hypothetico-deductive view of testing, concern over distinguishing theoretical from nontheoretical terms and modified Ramsey sentences. In this respect it joins a number of other neo-positivistic efforts that attempt to exploit model-theoretic notions. Subsequent development of the structuralist programme has focused on applying Sneed's perspective to Kuhn's views on science and incommensurability controversies (see Incommensurability).

Despite such reservations over the structuralist programme, the idea of individuating dynamically evolving theories on the basis of a structural core or partial models is a promising one. DaCosta and French (1990) present an analysis of partial models that is not tied to neo-positivistic ideas, but they do not apply it to theory individuation issues. Another untried approach would be to analyse theory structures as adaptive systems capable of reorganizing. Before expending much effort here, better data on how developing theories are individuated should be obtained.

See also: Experiment; Idealizations; Logical positivism; Models; Observation; Operationalism; Relativity theory, philosophical significance of; Scientific realism and antirealism

FREDERICK SUPPE

References and further reading

Achinstein, P. (1968) *Concepts of Science: A Philosophical Analysis*, Baltimore, MD: Johns Hopkins University Press.(Contains reprintings of his objections to positivistic operational-theoretical distinction and partial

interpretation notion.)
Beth, E. (1949) ‘Towards an Up-to-Date Philosophy of the Natural Sciences’, *Methodos* 1: 178-85.(Articulates his vision of a semantic conception of theories developed in his *Natuursphilosophie*, Gorinchem: Noorduyn, 1948, that inspired van Fraassen’s version of the semantic conception.)
Beth, E. (1963) ‘Carnap’s Views on the Advantages of Constructed Systems Over Natural Languages in the Philosophy of Science’, in P. Schilpp (ed.) *The Philosophy of Rudolf Carnap*, La Salle, IL: Open Court, 469-502.(Develops his amplified usage semantic doctrines that influenced Suppe’s version of the semantic conception.)
Birkhoff, G. and Neumann, J. von (1936) ‘The Logic of Quantum Mechanics’, *Annals of Mathematics* 37: 823-43; repr. in J. von Neumann, *Collected Works*, New York: Pergamon Press, 1962, vol. 4, 1-5-125.(The fundamental paper that influenced the development of the semantic conception of theories.)
Bromberger, S. (1963) ‘A theory about the Theory of Theory and about the Theory of Theories’, in B. Baumrin (ed.) *Philosophy of Science: The Delaware Seminar*, New York: Interscience, vol. II, 79-106.(Proposes an erotetic logic approach to theories and distinguishes different senses of ‘theory’.)
Burks, A.W. (1975) ‘Models of Deterministic Systems’, *Mathematical Systems Theory* 8: 295-308.(Detailed technical analysis of modelling deterministic and indeterministic state-transition systems using the mathematics of sequence generators.)
Campbell, N. (1920) *Physics: The Elements*, Cambridge: Cambridge University Press; repr. as *Foundations of Science*, New York: Dover, 1957.(Positivistic analysis arguing that theories must contain analogical models.)
Carnap, R. (1928) *Der Logische Aufbau der Welt*, Berlin; 2nd edn, Berlin: Felix Meiner, 1961; trans. R. George, *The Logical Structure of The World*, Berkeley, CA: University of California Press, 1969.(Carnap’s reworking of the transcendental analytic of Kant’s first *Critique* which lays the basis for his later treatment of theories. The translation is of the 2nd edition.)
Carnap, R. (1936-7) ‘Testability and Meaning’, *Philosophy of Science* 3: 420-66, 4: 1-40.(Classic rejection of explicit definition correspondence rules in favour of reduction sentences.)
Carnap, R. (1950) ‘Empiricism, Semantics, and Ontology’, *Revue internationale de Philosophie* 11, 208-28; repr. in *Meaning and Necessity*, enlarged edn, Chicago, IL: University of Chicago Press, 1956.(Carnap’s later semantic reworking of his *Aufbau*.)
Carnap, R. (1956) ‘Methodological Character of Theoretical Concepts’, in H. Feigl and M. Scriven (eds) *Minnesota Studies in the Philosophy of Science*, Minneapolis, MN: University of Minnesota Press, vol. 1, 33-76.(Carnap’s most comprehensive received-view treatment of the structure of theories.)
Carnap, R. (1963) ‘Replies and Systematic Expositions’, in P. Schilpp (ed.) *The Philosophy of Rudolf Carnap*, La Salle, IL: Open Court, 859-1013.(Contains responses to Hempel’s version of the positivistic received view on theories.)
Carnap, R. (1966) *Philosophical Foundations of Physics*, New York: Basic Books.(Contains Carnap’s most sustained discussion of the observational-theoretical distinction.)
Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Clarendon Press. (Criticizes realistic construal of theories and discusses how models attach to phenomena.)
Churchland, P. and Hooker, C. (eds) (1985) *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. van Fraassen*, Chicago, IL: University of Chicago Press.(A volume of criticisms of van Fraassen’s constructive empiricism account of theory testing on his version of the semantic conception, with replies by van Fraassen.)
Coffa, A. (1991) *The Semantic Tradition from Kant to Carnap: To the Vienna Station*, ed. L. Wessels, Cambridge: Cambridge University Press.(Authoritative historical account of the positivistic programme of language reform and its precursors.)
Craig, W. (1953) ’On Axiomatizability with a System’, *Journal of Symbolic Logic* 18: 30-2.(Theorem that some sought to use to defend instrumentalism.)
DaCosta, N.C. and French, S. (1990) ‘The Model-Theoretic Approach in the Philosophy of Science’, *Philosophy of Science* 57 (2): 248-65.(A partial models attempt to deal with the dynamics of theory development on the semantic conception.)
Dalla Chiara Scabia, M. and Toraldo di Franca, G. (1973) ‘A Logical Analysis of Physical Theories’, *Revisto del Nuovo Cimento*, 2nd series, 3: 1-20.(Neo-positivistic semantic analysis of theories.)
Feyerabend, P. (1958) ‘An Attempt at a Realistic Interpretation of Experience’, *Proceedings of the Aristotelian Society*, new series, 58: 143-70.(Analysis of scientific observation wherein theory and observation use the same vocabulary.)
Fraassen, B.C. van (1967) ‘Meaning Relations Among Predicates’, *Nous* 1: 161-80.(Fundamental paper on the

theory of semi-interpreted languages.)
Fraassen, B.C. van (1969) 'Meaning Relations and Modalities', *Nous* 3: 155-68.(Extends theory of semi-interpreted languages to include modal operators.)
Fraassen, B.C. van (1970) 'On the Extension of Beth's Semantics of Physical Theories', *Philosophy of Science* 37: 325-39.(Develops a semantic conception of theories based upon his theory of semi-interpreted languages.)
Fraassen, B.C. van (1979) 'Foundations of Probability: A Modal Frequency Interpretation', in G. Toraldo di Francia (ed.) *Problems in the Foundations of Physics*, Amsterdam: North Holland, 343-94.(Extends the theory of semi-interpreted languages to probabilistic and relative frequency statements.)
Fraassen, B.C. van (1980) *The Scientific Image*, New York: Oxford University Press.(Develops the associated constructive empiricist epistemology for his antirealist version of the semantic conception.)
Fraassen, B.C. van (1989) *Laws and Symmetry*, Oxford: Oxford University Press.(Contains his most systematic articulation of semantic approaches to the analysis of science.)
Fraassen, B.C. van (1991) *Quantum Mechanics: An Empiricist View*, New York: Oxford University Press. (Extends theory of semi-interpreted languages to modal probabilistic systems by presenting a semantic conception analysis of quantum mechanics.)
Giere, R. (1979) *Understanding Scientific Reasoning*, New York: Holt.(Elementary textbook that presents a clear nontechnical discussion of the semantic conception of theories.)
Giere, R. (1988) *Explaining Science: A Cognitive Approach*, Chicago, IL: University of Chicago Press.(Defends a realistic version of the semantic conception of theories.)
Hanson, N.R. (1958) *Patterns of Discovery*, Cambridge: Cambridge University Press.(Classic work arguing that observation is theory-laden.)
Harré, R. (1983) *An Introduction to the Logic of the Sciences*, New York: St Martin's Press, 2nd edn.(Contains a comprehensive presentation of the analogical view of models and their necessity within theories.)
Hempel, C.G. (1952) *Fundamentals of Concept Formation in Empirical Science*, Chicago, IL: University of Chicago Press.(Argues that some legitimate theoretical concepts cannot be defined by reduction sentences and introduces interpretive sentences as the general form of correspondence rules for the positivistic received-view account.)
Hempel, C.G. (1963) 'Implications of Carnap's Work for the Philosophy of Science', in P. Schilpp (ed.) *The Philosophy of Rudolf Carnap*, La Salle, IL: Open Court, 685-710.(Hempel's critical discussion of Carnap's version of the positivistic received view analysis of theories.)
Hempel, C.G. (1965) *Aspects of Scientific Explanation and Other Essays in Philosophy of Science*, New York: Free Press.(Reprints a number of his papers, including ones on cognitive significance, theories and instrumentalism versus realism.)
Henkin, L., Suppes, P. and Tarski, A. (eds) (1959) *The Axiomatic Method with Special Reference to Geometry and Physics*, Amsterdam: North Holland.(Presents a number of attempts to axiomatize theories using semantic methods.)
Hesse, M. (1966) *Models and Analogies in Science*, Notre Dame, IN: University of Notre Dame Press.(Modern classic defence of the necessity of analogical models as components of theories.)
Hooker, C. (1975) 'On Global Theories', *Philosophy of Science* 42: 152-79.(Appendix presents a detailed cloud chamber example of the complexities of applying theory to phenomena.)
Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press; rev. edn 1970.(The classic reform of positivist doctrines that rejects the observational-theoretical distinction and correspondence rules, while maintaining that basic choice of theory cannot be made on the basis of testing.)
Lakatos, I. (1970) 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (eds) *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press, 149-86. (Presents an approach to dealing with the dynamics of theorizing via a 'problem shift' sequence of successor theories.)
Lambert, K. and Brittan, G.G., Jr (1987) *An Introduction to the Philosophy of Science*, Atascadero, CA: Ridgeview Publishing Co., 3rd edn.(Contains excellent elementary presentation of positivistic and semantic analyses of theories.)
Laymon, R. (1985) 'Idealization and the Testing of Theories by Experimentation', in P. Achinstein and O. Hannaway (eds) *Observation, Experiment, and Hypothesis in Modern Physical Science*, Cambridge, MA: MIT Press, 147-74.(Focuses on the role of idealization in theories and implications for theory testing.)
Leplin, J. (ed.) (1984) *Scientific Realism*, Berkeley, CA: University of California Press.(Anthology containing articles representing many main stances on scientific realism versus antirealism.)
Lloyd, E. (1988) *The Structure and Confirmation of Evolutionary Theory*, Westport, CT: Greenwood Press.

(Develops a semantic conception account of evolutionary theory.)
Moulines, C.U. (1975) 'A Logical Reconstruction of Simple Equilibrium Thermodynamics', *Erkenntnis* 9: 101-30.(An application of Sneed's structuralist analysis of theories to thermodynamics.)
Moulines, C.U. (1976) 'Approximate Application of Empirical Theories: A General Explication', *Erkenntnis* 10: 201-27.(Concerns the application of theories to phenomena a under Sneed's structuralist account.)
McKinsey, J.C.C., Sugar, A. and Suppes, P. (1953) 'Axiomatic Foundations of Classical Particle Mechanics', *Journal of Rational Mechanics and Analysis*, 2: 253-72.(Semantic axiomatization of a theory via Suppes' set-theoretic predicates version of the semantic conception of theories.)
McKinsey, J.C.C. and Suppes, P. (1953) 'Transformations of Systems of Classical Particle Mechanics', *Journal of Rational Mechanics and Analysis* 2: 273-89.(Semantic axiomatization of a theory via Suppes' set-theoretic predicates version of the semantic conception of theories.)
Neumann, J. von (1932) *Mathematische Gründlagen der Quantenmechanik*, Berlin: Springer; trans. *Mathematical Foundations of Quantum Mechanics*, Princeton, NJ: Princeton University Press, 1955.(Classical demonstration of the power of semantic approaches by showing the equivalence of wave and matrix mechanics using a canonical Hilbert space model.)
Prezełecki, M., Szaniawski, K. and Wójcicki, R. (eds) (1976) *Formal Methods in the Methodology of Science*, Wrocław: Ossolineum.(Contains a number of semantic approaches to the analysis of theories, most of them being neo-positivistic.)
Putnam, H. (1962) 'What Theories Are Not', in E. Nagel, P. Suppes, and A. Tarski (eds) *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, Stanford, CA: Stanford University Press, 240-51.(Putnam's classic attacks on the positivistic observational/theoretical distinction and partial interpretation notion; stresses problems of unintended interpretations.)
Ramsey, F.P. (1931) 'Theories', in R.B. Braithwaite (ed.) *The Foundations of Mathematics*, London: Kegan-Paul, 212-36; repr. Paterson, NJ: Littlefield, Adams & Co., 1960.(Introduces a canonical received view formulation for theories that others attempted to transform into method for eliminating theoretical terms used in defence of instrumentalism. Sneed's structuralist account of theories and theoricity utilizes a modification of Ramsey's method.)
Schaffner, K. (1969) 'Correspondence Rules', *Philosophy of Science* 36: 280-90.(Argues that positivistic received-view correspondence rules jumble together meaning, causal and experimental relationships that are epistemologically distinct.)
Schaffner, K. (1993) *Discovery and Explanation in Biology and Medicine*, Chicago, IL: University of Chicago Press.(Explores modifications of the semantic conception in the analysis of theories in molecular biology.)
Shapere, D. (1969) 'Notes Toward a Post-Positivistic Interpretation of Science', in P. Achinstein and S. Barker (eds) *The Legacy of Logical Positivism*, Baltimore, MD: Johns Hopkins University Press, 115-60.(Criticizes the assumption that all terms must be given realistic or all must be given instrumentalistic interpretations.)
Sneed, J. (1971) *The Logical Structure of Mathematical Physics*, Dordrecht: Reidel.(Basic work presenting the structuralist analysis of theories.)
Sneed, J. (1976) 'Philosophical Problems in the Empirical Science of Science: A Formal Approach', *Erkenntnis* 10: 115-46.(Further development of Sneed's structuralist account of theories.)
Stegmüller, W. (1976) *The Structure and Dynamics of Theories*, New York: Springer.(Attempts, with partial success, to make Sneed's structuralist analysis more accessible.)
Suppe, F. (1977) *The Structure of Scientific Theories*, Urbana, IL: University of Illinois Press, 2nd edn. (Introduction and Afterword provide a comprehensive account of the development of the positivistic received view, criticisms of it (including individuation arguments) that led to its eventual demise, and post-positivistic attempts to understand scientific theorizing including the semantic conception. The volume also contains a number of important papers on theories, including Hempel's objections to Suppes' like formalization.)
Suppe, F. (1989) *The Semantic Conception of Theories and Scientific Realism*, Urbana, IL: University of Illinois Press.(Comprehensive detailed development of a quasi-realistic version of the semantic conception of theories; also contains a history of the origins of the semantic conception, reprintings of his rejoinders to Achinstein and Putnam on positivistic partial interpretation and observational-theoretical doctrines, and a response to Schaffner's attempts to modify the semantic conception to deal with molecular biology theories.)
Suppe, F. (forthcoming) *Facts, Theories, and Scientific Observation*, vol. 1, *A Posteriori Knowledge and Truth*, vol. 2, *Scientific Knowledge*, Urbana, IL: University of Illinois Press.(Detailed development of the epistemology of observation and theory testing for the quasi-realistic version of the semantic conception. Also contains a detailed formal analysis of simulation modelling.)
Suppes, P. (1957) *Introduction to Logic*, New York: Van Nostrand.(Part II articulates the linguistic versus

set-theoretic entity distinction and presents Suppes' set-theoretic predicate axiomatization approach to the semantic conception of theories. Technical but accessible.)

Suppes, P. (1961) 'A Comparison of the Meaning and Use of Models in Mathematics and the Empirical Sciences', in H. Freudenthal (ed.) *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*, Dordrecht: Reidel, 163-77.(Distinguishes theory, metamathematical and scientific model senses of 'model' and argues all can be analysed metamathematically.)

Suppes, P. (1962) 'Models of Data', in E. Nagel, P. Suppes and A. Tarski (eds) *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, Stanford, CA: Stanford University Press, 252-61.(Model theoretic analysis of how theories are fitted experimentally to phenomena under his version of the semantic conception of theories.)

Suppes, P. (1967) 'What is a Scientific Theory?', in S. Morgenbesser (ed.) *Philosophy of Science Today*, New York: Basic Books, 55-67.(Contains Suppes' articulation of extrinsic and intrinsic approaches to specifying theories, arguing for the superiority of extrinsic semantic approaches.)

Suppes, P. (1968) 'The Desirability of Formalization in Science', *Journal of Philosophy* 65, 651-4.(Addresses criticism of formal approaches to the analysis of science, including his vision of the semantic conception.)

Thompson, P. (1989) *The Structure of Biological Theories*, New York: State University of New York Press.(A semantic conception analysis of biological theories.)

Toulmin, S. (1953) *Philosophy of Science*, London: Hutchinson.(Argues against formalized analyses of theories and maintains that theories do not contain their scopes.)

Whitehead, A.N. and Russell, B. (1910-3) *Principia mathematica*, Cambridge: Cambridge University Press. (Syntactical axiomatization of mathematics that convinces positivists that mathematics is tautological, hence does not add any empirical content to scientific theories.)

Theory and observation in social sciences

The concept of observation has received relatively little systematic attention in the social sciences, with the important exceptions of social psychology, social anthropology and some areas of sociological methodology such as 'participant observation'. In a broader sense, however, concern with the relation between theory and 'reality', 'data', 'empirical research' and so on, has been a pervasive theme in the philosophy of social science and in the methodological self-reflection of the individual social sciences.

The social sciences differ very substantially in the extent and character of their metatheoretical preoccupations. Some are still dominated by what sometimes used to be called the 'standard view' in Anglo-American philosophy of science derived from the Vienna Circle's logical empiricism and consolidated by Karl Popper in the middle decades of the twentieth century (see Positivism in the social sciences). In this conception, theories consist essentially of law-like statements verified or, in Popper's more refined version, falsified, in a fairly direct confrontation with 'the facts'.

Both elements of this relation came under fire from the 1960s onwards. The traditional view of theory was attacked from three directions. Philosophers of science such as Mary Hesse, Rom Harré, Norwood Russell Hanson and Michael Scriven questioned the deductivist model of scientific theory. Historians and sociologists of science, building on the pioneering work of Thomas Kuhn (1962), himself inspired by a much earlier monograph by the neurologist Ludvik Fleck (1935), noted that scientists were much more collectivistic and conservative in their theoretical affiliations than Popper's model suggested (see Sociology of knowledge; Kuhn, T.). And social scientists, beginning with historians and philosophers of history, pointed out that explanations by reference to general laws had little application in the social world (see Explanation in history and social science §§1-2).

At the 'factual' or observational end of the relation, a convergent set of criticisms stressed the theory-dependence of observation statements in science as a whole and the additional complexities of 'understanding' social phenomena. Peter Winch (1958) developed themes of Wittgenstein's later work in directions which paralleled the traditional concerns of 'continental' hermeneutic theory. Here 'theory' builds on, and generally remains close to, the informal or common-sense understanding of a society of the kind possessed by its more reflective members. A variety of approaches, often grouped under the heading of 'social constructionism' and inspired directly or indirectly by Berger and Luckmann (1966), pursue their claim that what we call social reality is essentially the product of our own conceptualization. 'Critical' hermeneutics, combining hermeneutics with the Frankfurt School's critical theory, emphasizes, however, that these understandings can be systematically distorted by unequal power relations and the ideologies which sustain them (see Critical theory).

All this occurred against the background of the increasing polarization of the social sciences in the late 1960s and early 1970s, a marked revival of interest in the classics of nineteenth- and early twentieth-century sociology and, to a lesser extent the other social sciences, and more broadly, what has been called 'the return of grand theory' in the human sciences (Skinner 1985). The diversity of old and new perspectives led to something of a free-for-all in the more theoretically reflective social sciences. It also suggested however that earlier conceptions of the relation between theory and observational data were perhaps more relevant and productive than those found in the standard view. In both their substantive theorizing and in their metatheoretical reflections, Marx, Weber, Durkheim and the other classics, seemed less like remote ancestors of a social scientific practice which had since become 'scientific', and more like senior contemporaries.

As a result, English-language social science re-established connections with more theoretical traditions in continental Europe, and also with the philosophy of science (see Social science, contemporary philosophy of §4). In particular, the realist philosophies of science which became influential in Britain and elsewhere in the 1970s were inspired largely by a rejection of positivistic social science (see Scientific realism and social science §§1-2). But whereas the interpretive or hermeneutic traditions and those of critical theory had focused mainly on the inappropriateness of positivism as a theory of social science, realists offered an explicit alternative to positivist and conventionalist theories of science as a whole, thus leaving the way open, as was noted by Russell Keat (1971), to a restatement of naturalism, the thesis that the natural and the social sciences were alike concerned with the identification of structures and explanatory mechanisms. Realists accept that observations are theory-laden but

reject conventionalist conclusions, asserting instead that science is an open-ended, developing and fallible attempt to capture in our descriptions, models and theories the underlying structures of social reality. With this view, theory-choice cannot simply be based on observational adequacy or predictive success, but on a more complex and contestable assessment of explanatory power, involving bolder conjectures about the stratification of nature than one finds in the standard view, even among those like Popper, who defended a version of realism.

See also: Hermeneutics

WILLIAM OUTHWAITE

References and further reading

Berger, P. and Luckmann, T. (1966) *The Social Construction of Reality: a Treatise in the Sociology of Knowledge*, Harmondsworth: Penguin.(Despite the subtitle this book is better understood as a phenomenological theory of society.)

Bhaskar, R. (1979) *The Possibility of Naturalism*, Brighton: Harvester Press.(Brilliant extension of the model developed in his *A Realist Theory of Science* to the social sciences.)

Fleck, L. (1935) *Genesis and Development of a Scientific Fact*, Chicago, IL and London: University of Chicago Press, 1979.(This classic monograph on theories of syphilis and the development of the Wassermann test explores the nature of scientific communities and their ‘styles of thought’.)

Halfpenny, P. (1982) *Positivism and Sociology*, London: Allen & Unwin. (Outstanding brief discussion of the multifarious relations between the two.)

Keat, R. (1971) ‘Social scientific knowledge and the problem of naturalism’, *Journal for the Theory of Social Behaviour* 1 (1): 3-17.(Classic early statement of the implications of scientific realism for the philosophy of social science.)

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(Kuhn’s account of the history of science transformed the subject, and also had a substantial impact on the philosophy of science.)

Outhwaite, W. (1987a) ‘Laws and explanations in sociology’, in R.J. Anderson, J.A. Hughes and W.W. Sharrock (eds), *Classic Disputes in Sociology*, London: Allen & Unwin, 157-83.(Critique of the covering-law model of explanation and its application to the social sciences.)

Outhwaite, W. (1987b) *New Philosophies of Social Science: Realism Hermeneutics and Critical Theory*, London: Macmillan.(Argues for a convergence between the three.)

Outhwaite, W. (1988) ‘Theory’, in M. Haralambos (ed.), *Developments in Sociology*, Ormskirk: Causeway Press, vol. 5, 165-86.(Survey of recent trends.)

Skinner, Q. (ed.) (1985) *The Return of Grand Theory in the Human Sciences*, Cambridge: Cambridge University Press.(Documents the revival of ‘theory’ with essays on a number of influential recent philosophers and social and political thinkers.)

Winch, P. (1958) *The Idea of a Social Science and its Relation to Philosophy*, London: Routledge, 2nd edn, 1976. (Classic application of Wittgensteinian philosophy to the understanding of societies, thus demonstrating links back to German idealist philosophy.)

Theory and practice

Questions concerning the relation of 'theory' to 'practice' include whether there is a role for theory in the practical realm of ethics and politics; if so, how it can guide or provide justificatory reasons for practice; how reference to ethical practices might enter into the justification of theory; and whether theory can play a role in the critical appraisal of social practice. In responding to these issues, different conceptions of theory and practice need to be distinguished. Justifiable scepticism about ambitious claims for ethical theory need not rule out a more modest role for theoretical reflection on practice.

1 Theory and practice

The concepts of both 'theory' and 'practice' are used in a variety of ways. The term 'theory' is sometimes a shorthand for 'ethical theory' (see Morality and ethics §1). Minimally it can refer to any systematic framework of general ethical principles. More elaborately it refers to those systematizations which specify some basic set of principles from which more detailed moral injunctions that guide practice can be derived or with reference to which they can be tested: Kantian and utilitarian theories provide standard examples (see Kantian ethics; Utilitarianism). Other properties are commonly attributed to 'ethical theory' in this style: that it specifies principles that are 'abstract', that it be justifiable from an impartial perspective, that it make a claim on any rational agent (see Impartiality; Universalism in ethics §2). In addition to this specific use, the concept of theory is also used more broadly to include social and political thought which has both explanatory and normative components: Marxian and Austrian economic theories are standard examples (see Economics, philosophy of §§1, 3). The concept of practice is similarly used in a variety of ways. It can refer to specific actions (such as the act of giving this drug to this patient); to a kind of act (the giving of drugs, for instance); to a group of systematically related activities pursued for some common end (such as the practice of medicine); or, more broadly still, to a set of social institutions (for example, political and economic arrangements with different distributions of rights and goods). Finally, the concept of 'moral practices' is often used in contrast to 'ethical theory' to refer to the embodiment of ethical life in the specific responses and institutions of particular communities.

2 Principle and action

How are ethical principles which appear as general statements in a theory to specify or guide particular courses of action? One possible answer, the algorithmic, is that knowledge of the principle determines all the cases that fall under it. The principle specifies a finite list of properties that can be mechanically ascertained, so that for any situation it specifies unequivocally the act to be performed. Denials of this algorithmic approach standardly invoke the need for judgment (see Universalism in ethics §3). Rules cannot specify in advance the conditions under which they apply for they require an act of judgment about particulars which cannot be itself rule-determined. This role of judgment can be understood in a strong or weak sense. In the weak sense it is confined to seeing that a particular case falls under the rule. The principle is not itself infected by the judgment. On the strong position, the act of judgment, ascertaining whether an act falls under the rule, is akin to an act of interpretation which infects our understanding of the principle itself. The application extends our understanding of the principle. Defences of such positions often invoke Wittgensteinian considerations about rule-following: every application is a new interpretation (see Wittgenstein, L. §10). Consider, for example, the question of whether an act of 'date rape' should come under the principle prohibiting rape. On the algorithmic position, there should be general rules which will determine for us whether the act falls under the principle. On the weak judgmental position, a concept of rape is given and judgment is applied to ascertain whether this particular case falls under it. On the strong position, what the case does is to raise the question of how we are to understand the general principle against rape: the concept of rape itself is extended by this (and every other) new case.

The positions outlined put increasing emphasis on the importance of 'particular' as against 'universal' principles in ethical deliberation. This emphasis on the particular often takes the form of general scepticism about the very idea of ethical theory as a set of general principles that guide practice. The strong particularist rejects the whole picture of moral theory as a set of systematic general principles from which particular moral injunctions can be derived (see Moral particularism). Ethics is not codifiable. It is a matter of having the virtues, those dispositions of character that enable one to see and respond to a situation in a certain way, which cannot be settled in advance by

the application of principles (McDowell 1979) (see Virtue ethics; Virtues and vices §3). Within the conservative tradition of politics the idea of ethical theory is taken to be a rationalist misunderstanding of the nature of moral knowledge: a theoretical approach involves a confusion of technical knowledge, which can be articulated in a set of rules and reflected upon, with practical knowledge, which is unreflective, exists only in use, and cannot be formulated in rules (Oakeshott 1962) (see Conservatism §5). Scepticism about theory as a guide to practice is often allied with scepticism about the possibility of ethical theories providing justificatory reasons for a practice. There is no perspective external to ethical practices from which justification is possible. Only from within an ethical practice do we know how to go on. Correspondingly, appeals to principles of reason such as consistency or to formal features of moral language are too thin to justify any system of ethical belief (see Moral scepticism §2).

This sceptical approach to moral and political theory may have force against a particular ambitious model of theory according to which we can have a moral theory that takes a standpoint entirely external to any particular moral practice, makes a claim on any rational agent irrespective of time and place, and offers a decision procedure for settling any practical problem by the mechanical application of a system of universal principles. Criticism of ambitious theory can be granted. However, the sceptical arguments do not rule out theory in a more modest sense of systematic reflection on our ethical beliefs, attitudes and practices, nor do they show that such reflection cannot give us reasons to change and even abandon some existing practices. An antitheoretical position that rules out even modest ethical theory renders both particular acts and social practices opaque to criticism and public judgment. Reflective equilibrium in the Rawlsian sense in which theory starts from considered ethical judgments in everyday practices, subsequently revising some in the attempt to render judgments coherent, does not require us to reason from a point of view external to our practices (Rawls 1971). Neither do the forms of radical critical social theory developed within the Marxian tradition.

3 Social theory and political practice

Critical social theory within the Marxian tradition especially as systematized in the Frankfurt School (Fay 1985) proceeds through immanent critique of social practices from the inside through theoretical reflection (see Frankfurt school; Critical theory). The account runs roughly as follows. Social practices, in the widest sense of the term, are constituted and sustained by particular pre-theoretical self-understandings and norms. For example, market contract is constituted by self-understandings of the participants as independent agents with property rights which are exchanged according to certain norms; and it is sustained by the assumption that these norms are legitimate (see Ideology). Theoretical reflection on social practices will in part involve making explicit these assumptions. At the same time it may make assertions that conflict with the self-understandings of actors by claiming that they are either false or incoherent. Typical is Marx's claim that the market contract between wage worker and the owner of capital is not the free exchange between independent agents that the actors within it conceive it to be (see Marx, K. §9). The conflicts between the self-understandings of actors and the claims of social theory form the core of critical social theory, which takes as its aim the liberation of agents from those self-understandings through the changing of those practices that depend on them (Marx 1843). Within the early philosophical work of Marx this is conjoined with the more radical claim that, since many theoretical errors are the articulation of the false understandings that constitute or sustain particular practices, overcoming the theoretical errors requires changes to those practices (Marx 1845).

The extent to which critical theory can provide a purely internal criticism of any social practice is open to doubt. It presupposes contested Enlightenment norms of consistency and transparency of self-understanding. Moreover, those norms of themselves may be too thin to sustain a critique of social practices without the support of others. Hence a more general reflection on the goods of human life may be a presupposition of critical theory.

See also: Applied ethics; Casuistry; Examples in ethics; Ideals

JOHN O'NEILL

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985, esp. bks VI and X.(The classic account of the role of practical judgment about particulars in ethical deliberation.)

Fay, B. (1985) *Critical Social Science*, Oxford: Blackwell.(An accessible statement and assessment of the central

claims of critical social theory.)
Kant, I. (1793) ‘On the Common Saying: ‘‘This May be True in Theory, but it does not Apply in Practice"’, in H. Reiss (ed.) *Political Writings*, Cambridge: Cambridge University Press, 1991.(An accessible account of Kant’s account of the relation of theory and practice.)
McDowell, J. (1979) ‘Virtue and Reason’, *Monis*t 62: 331-50.(An influential statement of a particularism in ethics.)
Marx, K. (1843) *Zur Kritik der Hegelschen Rechtsphilosophie*, trans. R. Livingstone and G. Benton, *Critique of Hegel’s Philosophy of Right*, in L. Colletti (ed.) *Early Writings*, Harmondsworth: Penguin, 1975, 57-198.(An early and influential statement of the aims of critical social theory.)
Marx, K. (1845) ‘Theses on Feuerbach’, trans. R. Livingston and G. Benton, in L. Coletti (ed.) *Early Writings*, Harmondsworth: Penguin, 1975, 421-3.(The classic text on theory and practice in the Marxian tradition.)
Oakeshott, M. (1962) *Rationalism in Politics*, London: Methuen.(A clearly stated conservative defence of the primacy of practical knowledge in political and social life.)
Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press, ch. 1, sect. 9.(Contains the basic account of the methodology of ‘reflective equilibrium’.)
Sidgwick, H. (1895) ‘Theory and Practice’, *Mind* (new series) 4: 370-75.(A pithy account of the relation of theory to practice from a classical utilitarian theorist.)
Williams, B. (1985) *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press and London: Fontana.(An influential criticism of the ambitions of ethical theory.)

Theory of types

The theory of types was first described by Bertrand Russell in 1908. He was seeking a logical theory that could serve as a framework for mathematics, and, in particular, a theory that would avoid the so-called ‘vicious-circle’ antinomies, such as his own paradox of the property of those properties that are not properties of themselves - or, similarly, of the class of those classes that are not members of themselves. Such paradoxes can be thought of as resulting when logical distinctions are not made between different types of entities, and, in particular, between different types of properties and relations that might be predicated of entities, such as the distinction between concrete objects and their properties, and the properties of those properties, and so on. In ‘ramified’ type theory, the hierarchy of properties and relations is, as it were, two-dimensional, where properties and relations are distinguished first by their order, and then by their level within each order. In ‘simple’ type theory properties and relations are distinguished only by their orders.

1 Ramified type theory

The theory of (ramified) types was first described by Bertrand Russell (1908) as a way of avoiding vicious-circle antinomies, including the one discovered by Russell himself regarding the property that applies to all and only those properties that do not apply to themselves (see Russell, B.A.W. §§7-8; Paradoxes of set and property §5). A property is said to be ‘impredicable’ if, and only if, it is not predicable of itself; thus, the property of being impredicable is impredicable if, and only if, it is not impredicable.

$$\mathrm{Impred}(\mathrm{Impred}) \equiv \neg\mathrm{Impred}(\mathrm{Impred})$$

From this the contradiction follows that the property of being impredicable both is and is not impredicable. The theory of ramified types for propositional functions was developed in *Principia Mathematica* (1910-13). For Russell, such a theory was the framework of a ‘logically perfect language’, by which he meant a language that would show at a glance the logical structure of the facts that can be described by means of it, and in particular a language ‘in which everything that we might wish to say in the way of propositions that are intelligible to us, could be said, and in which, further, structure would always be made explicit’ (1959: 165). All that would be needed to add to such a theory is a vocabulary of descriptive constants that correspond to the meaningful words and phrases of the natural sciences or of natural language. The basic laws of a natural science would then be listed as nonlogical postulates of that science, and (as later proposed by Rudolf Carnap) the analytic connections between the words and phrases of natural language would be listed as meaning postulates. The constants and postulates of pure mathematics do not also need to be added, according to Russell, because they are all definable and provable in purely logical terms within the framework of type theory itself - a view known as (Russell’s form of) logicism (see Logicism §1).

The way type theory avoids vicious-circle antinomies is by imposing logico-grammatical constraints on the logic of predicates. The basic idea is that predicates (and the propositional functions they stand for) are to be typed in a hierarchical manner so that predicates of a given type (and their corresponding propositional functions) could never meaningfully be applied to predicates (propositional functions) of the same or higher type. (Thus both $\mathrm{Impred}(\mathrm{Impred})$ and $\neg\mathrm{Impred}(\mathrm{Impred})$ are meaningless in type theory.) In ‘ramified’ type theory, the hierarchy is, as it were, two-dimensional, having a ‘vertical’ dimension consisting of a hierarchy of orders and, within each order, a ‘horizontal’ dimension consisting of a hierarchy of levels. That is, properties are distinguished in ramified type theory as being both of a given order (for example, first-order, second-order and so on) and of a specific level within that order. Thus, for example, where ‘being shrewd’, ‘being courageous’, ‘being insightful about enemy positions’ and so on are examples of first-order/first-level properties that Napoleon and other great generals shared (that is, properties of r-type (i)/1, as defined below, assuming Napoleon and other generals to be of r-type i), Napoleon’s having *all of the (first-order/first-level) properties of a great general* - that is, all of the (first-order/first-level) properties that make up what it is to be a great general - amounts to Napoleon having a first-order property of level two (that is, a property of r-type (i)/2, as defined below). ‘Simple’ type theory has only the ‘vertical’ hierarchy.

The description of the ramified hierarchy given by Russell was informal and relied on a use/mention confusion (see Use/mention distinction and quotation §3) between propositional functions (properties or relations) and the

open formulas with 'apparent' (that is, 'bound' as opposed to 'free') variables taken to represent those functions. The ramified theory was not given a precise, formal description until Alonzo Church's paper of 1976, which, in accordance with contemporary terminology, uses the notion of 'order' to represent the 'vertical' hierarchy (instead of Russell's informal notion of order corresponding to the 'horizontal' hierarchy), and which uses the notion of 'level' to represent the 'horizontal' hierarchy. Church's definition of ramified types ('r-types') is as follows:

(1) There is an r-type *i* to which all and only individuals belong, and whose order is stipulated to be 0.
(2) If *m* is a non-negative integer and *n* is a positive integer and $\beta_1, \ldots, \beta_m$ are any given r-types, then there is an r-type $(\beta_1, \ldots, \beta_m)/n$ to which belong all and only *m*-ary propositional functions of level *n* and with arguments of r-types $\leq \beta_1, \ldots, \beta_m$, respectively; and the order of such a function is $N + n$, where *N* is the greatest of the orders corresponding to the types $\beta_1, \ldots, \beta_m$ (and $N = 0$ if $m = 0$).

Here, by an 'individual', that is, an entity of r-type *i*, Russell meant a concrete object (in particular, an event) as compared to the propositional functions (abstract properties and relations) that make up the entities of all other r-types. (It is not essential for all logical purposes that the entities of the basic type *i* be concrete, however.) The notion of the level of a propositional function ϕ of r-type $(\beta_1, \ldots, \beta_m)/n$ represents what Russell intended in speaking informally of the 'apparent' (bound) variables occurring in ϕ. If *N*, for example, is the greatest of the orders of $\beta_1, \ldots, \beta_m$, and *k* is the greatest of the orders of the 'apparent' variables occurring in ϕ (construed as a formula of r-type $(\beta_1, \ldots, \beta_m)/n$), then $n = 1$ if $k \leq N$, and $n = k + 1$ if $N < k$. According to Russell, a propositional function (or formula for such) was said to be 'predicative' if 'it is of the lowest order compatible with its having the arguments it has' (1910-13: vol. 1, 53); that is, in terms of the notion of level, ϕ is predicative if, and only if, $n = 1$.

In the simple theory of types the distinction between levels is dropped and only the distinction between orders is retained, so that in effect all propositional functions of the simple theory are predicative - although the comprehension principle of simple type theory, unlike that of ramified type theory, allows that such functions can be specified 'impredicatively', that is, in terms of a totality to which they belong.

2 The axiom of reducibility

It was Russell's view that all of traditional mathematics about numbers as abstract objects could be represented in terms of the predicative propositional functions of his theory of ramified types. He could not justify such a view, however, without assuming an 'axiom (schema) of reducibility', which stipulates that every propositional function of whatever r-type is coextensive with a predicative propositional function (having arguments of the same r-type). In symbols:

$$(\forall\phi)(\exists\psi)(x_1)\ldots(x_m)[\phi(x_1, \ldots, x_m) \\ \equiv \psi!(x_1, \ldots, x_m)],$$

where 'ϕ' and 'ψ' are variables for propositional functions of r-types $(\beta_1, \ldots, \beta_m)/n$ and $(\beta_1, \ldots, \beta_m)/1$, respectively, and '$x_1$', …, '$x_m$' are variables of r-type $\beta_1, \ldots, \beta_m$, respectively. (Russell used '$\psi!$' to indicate that the propositional function represented by 'ψ' is predicative.) This axiom was criticized by a number of logicians (including L. Chwistek 1922; F. Ramsey 1931; W.V. Quine 1936) who argued that, on the assumption that type theory is an extensional system in which propositional functions of the same r-type are identical if they are coextensive (that is, true of all the same objects), the effect of the axiom of reducibility was to nullify the distinction between the different levels of propositional functions - in which case ramified type theory amounted to nothing more than a notationally complex version of simple type theory.

The extensionality assumption is dubious, however, if propositional functions are interpreted intensionally as properties and relations (as Russell originally intended). The point of the axiom of reducibility was not to identify properties and relations with their extensions, but to represent their extensions within the wider intensional framework, which is why Russell also called it 'the axiom of classes' (1910-13: vol. 1, 167). The effect of the axiom is that so-called 'impredicative' specifications of propositional functions - that is, specifications of functions in terms of a totality to which they belong - are 'annulled in extensional but not in intensional matters' (Church 1976: 758).

Ramsey emphasized a distinction between so-called logical and semantic antinomies, both of which the theory of

ramified types was designed to avoid. Russell's paradox, for example, is one of the logical antinomies, as is Georg Cantor's paradox that there are classes (such as the universal class) larger than all other classes, even though, by a theorem of Cantor's, there is a class larger than any given class. These are paradoxes that, without some restrictions such as those imposed in a theory of types, simple or ramified, might occur in any system rich enough to represent properties as objects having properties, or classes as members of classes. The semantic antinomies, on the other hand, assume special semantic notions, such as truth (as in the liar paradox of a sentence that says of itself that it is false, which if true is false, and if false is true), or definability (as in G.G. Berry's paradox of 'the least integer not nameable in fewer than nineteen syllables', which is itself a name with just eighteen syllables), and other related notions of semantics. Ramsey (1931) thought of the semantic paradoxes as 'epistemological' in nature. (See Paradoxes of set and property.)

3 The simple v. the ramified theory of types

The simple theory of types suffices to avoid the logical paradoxes, but not the semantic ones, unless a distinction is made (as first proposed by A. Tarski) between the object language of simple type theory and a metalanguage in which the different semantic notions for the expressions of simple type theory can be defined or developed axiomatically (see Tarski's definition of truth). The ramified theory of types avoids the semantic paradoxes by incorporating an additional hierarchy of levels that (combined with the hierarchy of orders) implicitly contains a hierarchy of languages, and that in effect amounts to a special form of the metalanguage/object language distinction by which the semantic (and related intensional) paradoxes can be avoided (as noted by Church 1976: §4). Adding the axiom of reducibility does not nullify the way ramified type theory resolves or avoids the semantic (and related intensional) paradoxes. (See Church (1976: §5) and Myhill (1979) for justification of this claim.)

Different versions of the simple theory of types have been formulated by L. Chwistek (1922), F. Ramsey (1931), R. Carnap (1929), K. Gödel (1931), and A. Church (1940). The simple theory is not committed to an extensional logic, although it is often formulated as such. R.M. Montague (1974), for example, extended Church's simple theory of types by incorporating into it operators for representing the intensions and extensions of expressions, and then used the resulting system as an intensional logic by which to provide an interpretation of a variety of intensional contexts of natural language (along lines originally described by Frege). Church (1974) has also formulated an intensional version of the simple theory by adding to it a formula operator for the Russellian notion of propositional identity. But in 1984, Church also showed that, given Russell's notion of a proposition, and hence of propositional identity, Russell's antinomy of propositions (Russell 1903: §500) can be reconstructed in simple type theory, but not in ramified type theory. According to this antinomy, there cannot be more classes of propositions than there are propositions - because every non-empty class of propositions can be associated with its 'logical product', which, by Russell's notion of propositional identity, is unique for that class; and yet, by Cantor's theorem, there must be more classes of propositions than there are propositions (see Cantor's theorem).

See also: Church, A. §3; Intensional logic; Logical and mathematical terms, glossary of

NINO B. COCCHIARELLA

References and further reading

Carnap, R. (1929) *Abriss der Logistik, mit besonderer Berücksichtigung der Relationstheorie und ihrer Anwendungen (Outline of Logistic, with Special Consideration of the Theory of Relations and their Applications)*, Vienna: Springer.(This work includes an early formulation of simple type theory and its applications as a logical framework for the axiomatic method. Examples of axiomatic theories, such as set theory, Peano arithmetic, projective geometry and space-time topology are given within the simple type theory formulated. The book was later expanded and developed as Carnap (1954).)

Carnap, R. (1954) *Einführung in die symbolische Logik, mit besonderer Berücksichtigung ihrer Anwendungen*, Vienna: Springer; trans. *Introduction to Symbolic Logic and its Applications*, New York: Dover, 1958.(A simple, clear description of Carnap's version of simple type theory can be found in this book, which also includes many important applications of type theory to the construction of the different kinds of numbers, the foundations of mathematics, spacetime physics, anthropological kinship theories, and axiomatic biology.)

Church, A. (1940) 'A Formulation of the Simple Theory of Types', *Journal of Symbolic Logic* 5: 56-68.(Elegant, but somewhat difficult for students with only an introductory understanding of logic.)

Church, A. (1974) 'Russellian Simple Type Theory', *Proceedings and Addresses of the American Philosophical*

Association 47: 21-33.(This paper includes a formulation of simple type theory with a primitive operator for propositional identity.)

Church, A. (1976) ‘Comparison of Russell’s Resolution of the Semantical Antinomies with That of Tarski’, *Journal of Symbolic Logic* 41: 747-60.(An intermediate level understanding of logic should suffice to read this paper, which includes some important observations on how ramified, but not simple, type theory can avoid certain semantic and intensional paradoxes.)

Church, A. (1984) ‘Russell’s Theory of Identity of Propositions’, *Philosophia Naturalis* 21 (2-4): 513-22. (Reviews Russell’s antinomy about propositions (Russell 1903: §500). Church shows that this antinomy cannot be resolved in simple type theory, given Russell’s view of propositional identity, but it can be resolved in ramified type theory.)

Chwistek, L. (1922) ‘Über die Antinomien der Prinzipien der Mathematik’ (On the Antinomies of the Principles of Mathematics), *Mathematische Zeitschrift* 14: 236-43.(Now read mostly for historical interest.)

Cocchiarella, N.B. (1987) *Logical Studies in Early Analytic Philosophy*, Columbus, OH: Ohio State University Press.(Chapter 1 includes a detailed description of the theory of logical types in the development of Russell’s views from his 1903 analysis of his paradox to the 1910-13 formulation of his ramified theory. Chapter 2 compares and reconstructs without paradox both Russell’s and Frege’s early versions of logicism. Chapter 5 explains how Russell’s 1925 modifications to his ramified theory commit him to a system too weak for his earlier logicism.)

Copi, I. (1971) *The Theory of Logical Types*, London: Routledge & Kegan Paul.(A good, elementary introduction.)

Gödel, K. (1931) ‘Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I’, *Monatshefte für Mathematik und Physik* 38: 173-98; trans. ‘On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 592-617.(A classic, including Gödel’s famous incompleteness theorems as applied to a type-theoretical development of arithmetic.)

Montague, R. (1974) *Formal Philosophy: Selected Papers of Richard Montague*, ed. R.H. Thomason, New Haven, CT: Yale University Press.(See especially the essays ‘Universal Grammar’ and ‘The Proper Treatment of Quantification in Ordinary English’, for Montague’s development of Church’s simple type theory as an intensional logic.)

Myhill, J. (1979) ‘A Refutation of an Unjustified Attack on the Axiom of Reducibility’, in G.W. Roberts (ed.) *Bertrand Russell Memorial Volume*, London: Allen & Unwin.(Readable with an intermediate level of understanding of logic; explains how the addition of the axiom of reducibility to ramified type theory results in a system essentially different from simple type theory.)

Quine, W.V. (1936) ‘On the Axiom of Reducibility’, *Mind* 45: 498-500.(One of the earliest arguments that the addition of the axiom of reducibility to ramified type theory amounts to a system only notationally different from simple type theory.)

Ramsey, F.P. (1931) ‘The Foundations of Mathematics’, in *The Foundations of Mathematics and Other Logical Essays*, ed. R.B. Braithwaite, London: Routledge & Kegan Paul.(Ramsey’s version of simple type theory, and his arguments why, as a foundation for mathematics, the simple theory is to be preferred to the ramified theory.)

Russell, B.A.W. (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992, esp. ch. 10.(Russell’s first extended discussion of his paradox, along with some proposals for avoiding it. His first rough ideas about a theory of types are described in appendix B, which also contains an account, in §500, of his antinomy about propositions.)

Russell, B.A.W. (1908) ‘Mathematical Logic as Based on the Theory of Types’, *American Journal Of Mathematics* 30: 222-62; repr. in *Logic and Knowledge, Essays 1901-1950*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 59-102; repr. London: Routledge, 1992.(Accessible to readers with an intermediate-level understanding of logic.)

Russell, B.A.W. (1959) *My Philosophical Development*, London: Allen & Unwin; repr. London: Routledge, 1993. (An autobiographical account of the development of Russell’s philosophical and logical views.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(Development of the theory of ramified types.)

Theosophy

Etymologically, 'theosophy' means wisdom concerning God or divine things, from the Greek 'theos' (God) and 'sophia' (wisdom). Seventeenth-century philosophers and speculative mystics used 'theosophy' to refer to a knowledge of nature based on mystical, symbolical or intuitive knowledge of the divine nature and its manifestations. It referred also to an analogical knowledge of God's nature obtained by deciphering correspondences between the macrocosm and God.

In the late nineteenth century, 'theosophy' became associated with the doctrines of Helena Petrovna Blavatsky, the founder of the popular Theosophical Society. She drew on Buddhist and Hindu philosophy and fragments from the Western esoteric tradition, especially Neoplatonism. She espoused an absolutist metaphysics in which there is a single, ultimate, eternal principle which remains unchanged and undiminished, despite manifesting itself partially in the periodic emanation and reabsorption of universes. Her cosmology included a spiritual account of the evolution of material bodies, which serve as the necessary vehicles by which individuals gradually perfect themselves through cyclic rebirth.

1 Traditional theosophy and Western esotericism

Prior to the influence of the Theosophical Society, the term 'theosophy' was used to refer to any intellectual system of pagan, Jewish or Christian origin which purported to describe the secret constitution of the divine nature. Typically, this wisdom or 'gnosis' was used to achieve an understanding of the mystical relationships among the universe, humanity and God unavailable to either the empirical sciences or rational theology. Western theosophists saw themselves as heirs to a tradition of mystical and occult knowledge with roots in Pythagoreanism, Platonism, Neoplatonism, Gnosticism, Hermetism, the Jewish Kabbalah, medieval Christian mysticism and Renaissance nature philosophy.

According to Antoine Faivre (1994), the term 'theosophy' was used occasionally as a synonym for 'theology' even by some medieval philosophers, for example by Pseudo-Dionysius and Johannes Scottus Eriugena. It was used more frequently during the Renaissance, but acquired its distinctive meaning only at the beginning of the seventeenth century. At that time, alchemy, Alexandrian Hermetism and Jewish mysticism (the Kabbalah) were influencing the philosophy of nature and Rosicrucianism was gaining adherents. The latter was a loosely knit movement associated with members of the fraternal Society of the Rose Cross. Rosicrucians followed the teachings of the legendary Christian Rosencreutz, who preached the possibility of an initiatory knowledge of a supersensible reality. In this context, some German authors began to use 'theosophy' to describe the knowledge of God's nature obtained by esoteric and mystical interpretation of sacred texts. Jakob Boehme, for example, the German speculative mystic, described himself as a theosophist.

During the seventeenth century, 'theosophy' was closely associated with 'pansophy', a term used by Rosicrucians and Paracelsians to refer to a knowledge of divine things obtained by deciphering the 'signatures' found in the 'book of nature' (see Paracelsus §2). For seventeenth-century theosophy, the wisdom concerning the divine nature was understood first and a gnostic comprehension of nature was derived from it, whereas for pansophy the order was reversed - decoding the 'signatures' came first and God was known by understanding the correspondences between nature and God. But from the eighteenth century onward, 'theosophy' was used to refer to hermeneutic speculations in either direction.

'Theosophy' entered the world of academic discourse largely through Jakob Brucker's chapter on theosophy in his *Historia critica philosophiae (Critical History of Philosophy)* (1741), a work which became a standard reference for the history of philosophy in the eighteenth century. According to Faivre (1994), Denis Diderot's *Encyclopédie* article on theosophers used long passages from Brucker's chapter without citation. Diderot's article introduced 'théosophie' into the French language and served as the definitive exposition of the term during the Enlightenment.

Christian theosophy was unified more by its hermeneutical method and its goal of achieving a transformative experience of divine revelation than by the specific claims made by various theosophists about nature and God's nature. Christian theosophists interpreted divine revelation by symbolically re-experiencing the events described in

a sacred narrative. They relied on their imaginative responses to the text, confident that the imagination could be used as an organ of perception when divinely inspired and interiorly illuminated. Analogical principles that related the created world to God provided both an intellectual framework for, and a corrective to, the enthusiastic imagination. For example, Christian theosophists used this hermeneutical method to explain the existence of evil and the alienated spiritual condition of humanity in terms of the constitution of the Godhead.

Important figures in the Christian theosophical tradition are Meister Eckhart, Nicholas of Cusa (§2), Agrippa von Nettesheim (§§1-2, 5), Robert Fludd, Angelus Silesius, Johann Georg Gichtel, Emanuel Swedenborg, Friedrich Christoph Oetinger, Louis-Claude de Saint-Martin, F.W.J. von Schelling (§3) (especially in his later work - for example, his *Of Human Freedom*) and Franz von Baader.

In the late nineteenth century, the Theosophical Society introduced the term 'theosophy' into popular culture and widened its scope by including the theosophical doctrines of Oriental religion and philosophy.

2 Theosophy and the Theosophical Society

Helena Petrovna Blavatsky (1831-91) and the American journalist and lawyer Henry Steel Olcott founded the Theosophical Society in 1875 in New York City. Its mission was defined by three objects: first, to form a nucleus of the universal brotherhood of humanity, without distinction of race, creed, sex, caste or colour; second, to encourage the study of comparative religion, philosophy and science; and third, to investigate unexplained laws of nature and the powers latent in human beings. Blavatsky and Olcott moved to India in 1879, and in 1882 they established the Theosophical Society's headquarters at Adyar. In 1896, the American branch broke away, establishing the Theosophical Society of America.

Blavatsky's message appealed to nineteenth-century Americans because it harmonized science and religious faith, and delineated a worldview in which salvation was largely up to the individual. The publication of Darwin's *The Origin of Species* in 1859 contributed to a growing awareness that the findings of science were undermining belief in biblical texts as divine revelation. The apparent conflict between science and religion prompted some believers to look for a scientific basis for religious faith. The American ideal of self-improvement led some to question the moral acceptability of the Christian doctrines of vicarious atonement and eternal damnation. Blavatsky responded to these concerns by spiritualizing evolutionary theory and emphasizing self-reliance in the quest for moral and spiritual perfection. The Theosophical Society's influence was not confined to America, however: the Eastern ideas it popularized inspired many European artists at the beginning of the twentieth century, and its members were even influential in the revival of Buddhism in Ceylon (now Sri Lanka).

Blavatsky's theosophy has several distinctive features, some in common with the traditional theosophy of Western esotericism, and others in common with Oriental esotericism:

(1) It has an absolutist metaphysics which tends towards monism and idealism. There is an unknowable, impersonal, eternal, unitary First Principle which is the transcendent ground and source of everything (see Brahman). The doctrine is monistic because the absolute is the only self-sufficient substance, and ultimately all individual minds are substances in virtue of being rooted in the absolute. They are one with the absolute and with each other. The doctrine is idealistic because intelligence and consciousness are the primary manifestations of the absolute, although these attributes cannot properly be affirmed of it, according to theosophy's negative theology.
(2) Theosophy has an emanationist cosmology: the absolute's perfection remains undiminished and unchanged, though it manifests periodically as a lesser reality, the Logos or the manifest God, from whom numberless universes gradually emanate and dissolve in a rhythmic process (see Creation and conservation, religious doctrine of §3). The absolute does not create the universe; both creation from nothing and from a pre-existing matter are ruled out. Rather, the Logos periodically differentiates itself from the absolute as an expression of the absolute's inner nature.
(3) Theosophy has a hierarchical account of reality: there are seven degrees of being which manifest different aspects of the Logos, and for each 'plane' of being there is a grade of matter appropriate for manifesting intelligence and consciousness at that level. Ultimately, however, both matter and intelligence are manifestations of the Logos, which in turn is a manifestation of the First Principle.
(4) Theosophy espouses a teleological and idealistic evolutionary theory. It is idealistic because it claims that the evolution of diversified physical forms is preceded by an involution of consciousness into matter. General

species-specific characteristics first exist in the Logos, and through the process of physical evolution become more fully expressed by the functioning and capacities of biological organisms. In contrast, materialistic Darwinian theory claims that these characteristics are the result of entirely natural processes, that is, random genetic mutation and natural selection. Theosophical evolutionary theory is teleological because the evolutionary process aims at developing enlightened individuals who are conscious of the full range of their multidimensional existence and are dedicated to the project of leading others to spiritual truth. All individuals, from solar systems to human beings, gradually descend from the highest degree of being to the lowest and then ascend again. Human beings have a sevenfold constitution which reflects the seven degrees of being, and each principle of the human microcosm enables the individual to participate in the corresponding macrocosmic principle.

(5) The evolutionary process requires that human beings actualize and balance their gnostic, affective and volitional capacities. These capacities include direct insight into the transcendent reality, imaginative and mythic grasp of truths about the correspondences between the macrocosm and microcosm, unbiased reasoning which corrects intuition and unmasks self-deception, and compassionate concern for all living things. The awakening of compassion leads to a forceful commitment to the project of reshaping one's character and overcoming all obstacles to enlightenment, first in oneself and then in others.

(6) Theosophy claims that the universe works in accordance with ethical principles which guarantee that morality coincides with self-interest; in the long term, justice prevails. The 'law of karma' holds individuals accountable for their thoughts and intentional actions via a natural causal process. Both justice and evolutionary teleology require a long series of rebirths, in which individuals slowly perfect themselves by experiencing the painful consequences of their ignorance and misguided actions (see Reincarnation).

(7) The apparent conflict between the teachings of the various historical religions can be harmonized by understanding their esoteric and mystical core. In some cases, development or correction must be supplied from the unadulterated primeval tradition, allegedly preserved by a secret brotherhood of adepts and sages.

(8) There is an 'occult science' comprised of natural laws unknown to modern science.

See also: Gnosticism; Hermetism; Kabbalah; Mysticism, history of; Neoplatonism; Pythagoreanism; Steiner, R.

MICHAEL B. WAKOFF

References and further reading

Blavatsky, H.P. (1877) *Isis Unveiled: A Master-Key to the Mysteries of Ancient and Modern Science and Theology*, New York: J.W. Bouton; repr. Pasadena, CA: Theosophical Press, 1972.(Syncretic, wide-ranging and fragmentary discussions of Blavatsky's thesis that there is a secret wisdom-tradition which harmonizes spiritual truths and scientific theory.)

Blavatsky, H.P. (1888) *The Secret Doctrine: The Synthesis of Science, Religion, and Philosophy*, ed. with intro. by B. de Zirkoff, London: Theosophical Publishing Company; repr. Adyar: Theosophical Publishing House, 1978.(Volume 1, 'Cosmogenesis', describes an emanationist cosmology and Volume 2, 'Anthropogenesis', describes the history of the earth and the evolution of successive human 'root races'. Difficult reading, but the 'Proem' summarizes some key points. The 1978 edition has a useful index.)

Blavatsky, H.P. (1889) *The Voice of the Silence, Being Chosen Fragments from the 'Book of the Golden Precepts'*, London: Theosophical Publishing Company; repr. Wheaton, IL: Theosophical Publishing House, 1992.(Very readable - a devotional manual about the theosophical path to enlightenment.)

Blavatsky, H.P. (1889) *The Key to Theosophy*, London: Theosophical Publishing Company; repr. London: Theosophical Publishing House, 1968.(An accessible introduction to the tenets of the Theosophical Society.)

Brucker, J. (1741) *Historia critica philosophiae (A Critical History of Philosophy)*, Leipzig.(No English translation available.)

Brunton, P. (1943) *The Wisdom of the Overself*, New York: E.P. Dutton.(A clear, jargon-free exposition of the central teachings of Oriental theosophy written for those who aspire to spiritual wisdom.)

Brunton, P. (1984-8) *The Notebooks of Paul Brunton*, Burdett, NY: Larson, 16 vols.(A posthumously published selection from Brunton's notebooks, arranged into sixteen categories. Volumes 13-16 are especially useful for those interested in both the theory and practice of Oriental theosophy.)

Campbell, B. (1980) *Ancient Wisdom Revived: A History of the Theosophical Movement*, Berkeley, CA: University of California Press.(Accessible introduction to the history, tenets and cultural influences of Blavatsky's theosophical movement.)

Diderot, D. (1779-82) *Encyclopédie*, in J. Assézat and M. Tourneux (eds) *Oeuvres complètes de Diderot*, Paris,

1875-7.(Volume 17 contains the article on theosophers, but no English translation of it is available.)

Ellwood, R. (1986) *Theosophy, a Modern Expression of the Wisdom of the Ages*, Wheaton, IL: Quest Books.(A sympathetic discussion of Blavatsky's theosophy by a scholar of religious studies.)

Faivre, A. (1994) *Access to Western Esotericism*, Albany, NY: State University of New York Press.(Contains an expansion of the material of §1 of this entry on the history of the term 'theosophy', a history of Western esoteric currents, historical studies of several Christian theosophists, and an excellent bibliographical guide to research, including recent European scholarship.)

Faivre, A. and Needleman, J. (eds.) (1992) *Modern Esoteric Spirituality*, New York: Crossroad.(Studies of esoteric traditions and writers in the modern period, including Paracelsus, Rosicrucianism, Jakob Boehme and the Theosophical Society, introduced by an article on the ancient and medieval sources of modern esoteric movements.)

Godwin, J. (1989) *The Beginnings of Theosophy in France*, London: Theosophical History Centre.(A scholarly historical work.)

Godwin, J. (1994) *The Theosophical Enlightenment*, Albany, NY: State University of New York Press.(Historical study of theosophy and occultism in England from the seventeenth century to the early twentieth century which argues that theosophy is indebted to both the French Enlightenment and the Hermetic tradition.)

Gomes, M. (1994) *Theosophy in the Nineteenth Century: An Annotated Bibliography*, New York: Garland Publishing.(Comprehensive bibliography of nineteenth-century works by and about H.P. Blavatsky, the Theosophical Society and other theosophists.)

Nasr, S.H. (1989) *Knowledge and the Sacred*, Albany, NY: State University of New York Press.(Based on Nasr's 1981 Gifford Lectures on esotericism and tradition. In the context of arguing that knowledge must be resacralized, he presents a historical survey of Western theosophical currents from ancient Greek thinkers to the present.)

Peuckert, W.E. (1936) *Pansophie. Ein Versuch zur Geschichte der weissen und schwarzen Magie (Pansophy: An Essay in the History of White and Black Magic)*, Stuttgart: Kohlhammer.(A study of Christian theosophy and pansophy, not available in English.)

Schelling, F.W.J. (1809) *Philosophische Untersuchungen über das Wesen der menschlichen Freiheit und die damit zusammenhängenden Gegenstände (Philosophical Inquiries into the Nature of Human Freedom and Connected Topics)*, trans. J. Gutmann, *Schelling: Of Human Freedom*, Chicago, IL: Open Court, 1936. (The translator's introduction discusses Schelling's views about mysticism and theosophy in relation to philosophy.)

Weeks, A. (1993) *German Mysticism from Hildegard of Bingen to Ludwig Wittgenstein: A Literary and Intellectual History*, Albany, NY: State University of New York Press.(Accessible intellectual history of German mysticism and theosophy.)

Thermodynamics

Thermodynamics began as the science that elucidated the law-like order present in the behaviour of heat and in its transformations to and from mechanical work. It became of interest to philosophers of science when the nature of heat was discovered to be that of the hidden energy of motion of the microscopic constituents of matter.

Attempts at accounting for the phenomenological laws of heat that make up thermodynamics on the basis of the so-called kinetic theory of heat gave rise to the first fundamental introduction into physics of probabilistic concepts and of probabilistic explanation. This led to so-called statistical mechanics.

Some of the issues of thermodynamics with importance to philosophers are: the meaning of the probabilistic claims made in statistical mechanics; the nature of the probabilistic explanations it proffers for the observed macroscopic phenomena; the structure of the alleged reduction of thermodynamic theory to the theory of the dynamics of the underlying microscopic constituents of matter; the place of cosmological posits in explaining the behaviour of local systems; and the alleged reducibility of our very notion of the asymmetry of time to thermodynamic asymmetries of systems in time.

1 Historical sketch

One of the great discoveries of phenomenological thermodynamics was the joint conservation of the totality of overt mechanical energy and internal heat, a law confirmed by Joule and called the first law of thermodynamics. Another was the realization that the ability to transform the internal heat of a system into usable mechanical work required a process in which heat was introduced into the system at a higher temperature and recovered at the end of the process only at a lower temperature where it was less available for conversion into work. These insights began with the investigations of Carnot, and were brought to fruition by Clausius with his introduction of the concept of entropy and the postulation of the second law of thermodynamics.

The idea that heat was merely energy of motion of microscopic constituents of matter was already suggested by Francis Bacon and Bernoulli. These ideas were resurrected by Herepath and Waterston, but ignored by the scientific community. Finally, when once more discovered by Krönig and developed by Clausius, this theory became the dominant theory of heat.

The full-blown kinetic theory receives its greatest impetus from Maxwell's ingenious derivation of the distribution of velocities of the molecules of a system which has obtained equilibrium, and from the derivation of a law governing the path followed by a system from an initial non-equilibrium condition to equilibrium by J.C. Maxwell and by Boltzmann. Boltzmann also provided a proof that any system not in equilibrium must move to equilibrium and then stay there, providing an underpinning at the kinetic level for the behaviour demanded by the second law.

The kinetic theory of heat was criticized by such positivistically minded philosophers as Duhem and Mach for its introduction of unobservables and its insistence on mechanistic explanations (see Duhem, P.M.M. §2; Mach, E.). More important were the specific objections made to the theory of approach to equilibrium. The kinetic theory predicts inevitable approach to equilibrium in the future time direction. But the underlying dynamics of the molecules is time-reversal invariant. Furthermore, the dynamical theory predicts the 'almost recurrence' to its initial state by the system, a recurrence that is incompatible with a monotonic approach to stationary equilibrium.

In response to these observations Maxwell and Boltzmann introduced probabilistic considerations into the theory. Boltzmann characterized equilibrium as the 'most probable' state of a system. He also argued that his equation gave the probable evolution of a system, not its inevitable evolution. Even this interpretation of Boltzmann's equation is subject to objections, and Boltzmann finally moved to a time-symmetric interpretation of his theory: almost all the time systems are at equilibrium. Occasionally they fluctuate from it. When a system is in non-equilibrium it is usually closer to equilibrium in both future and past time, as described by Boltzmann's equation.

But why, then, do we live in a non-equilibrium world? Boltzmann suggested that our region of the universe is a local fluctuation from equilibrium, the universe as a whole being at that most probable state. The consistency of the Boltzmann account was elucidated in an important survey piece by Ehrenfest and Ehrenfest (1959).

2 Equilibrium theory

A standard method for calculating the values of quantities at equilibrium was invented by Maxwell and Boltzmann, and generalized by Gibbs. A probability distribution is placed over the set of microscopic conditions of the system compatible with its macroscopic constraints (such as its fixed total energy and confinement to a box of specific volume). Macroscopic quantities at equilibrium are then identified with averages, calculated using the probability distribution, of the values of functions of the possible microscopic conditions.

But what justifies choosing a specific probability distribution? And with what physical proportion or frequency in the world is it to be associated?

One proposal (by Jaynes (1983)) takes it that the distribution is justified by standard principles of inductive reasoning and that the probability in question is to receive a subjective interpretation. More common is the proposal, suggested by Maxwell and Boltzmann, that the probabilities in question are representative of proportions of time, in the limit as time goes to infinity, in which a single system has microscopic conditions of a given kind (see Probability, interpretation of §§3, 5).

The attempt to justify this latter claim led Maxwell and Boltzmann to the ergodic hypothesis, the posit that over infinite time a system's microscopic condition would take on every possible value. This posit leads one to the conclusion that the standard probability distribution is the unique such distribution stationary in time, and that its probabilities do indeed correspond to time proportions in the infinite time limit. But the hypothesis is provably false.

The ergodic hypothesis was later replaced by the ergodic theorem, a theorem that holds for specifiable dynamical systems and from which weakened versions of the consequences of the hypothesis can be derived. The central question for the philosopher is the degree to which the results of ergodic theory provide statistical explanations of the equilibrium nature of systems. Two crucial problems are the facts that the results require idealizations concerning systems often not met by realistic systems (see Idealizations), and that the results connect probabilities with proportions only in an infinite time limit.

3 Non-equilibrium theory

Equilibrium theory cannot explain approach to equilibrium observed experimentally in the world. For this, one needs a statistical mechanical derivation of the appropriate 'kinetic equation', the equation generalizing that found by Boltzmann for gases of low density.

The standard derivations of the kinetic equations invoke re-randomizing posits, the descendants of Boltzmann's hypothesis with regard to collision numbers and of the hypothesis of molecular chaos of the Ehrenfests. The problem is, however, that once an initial probability distribution is specified for a system, its dynamic evolution is fixed by the underlying deterministic dynamical laws governing the micro-constituents. Therefore the consistency of a re-randomization posit is always in doubt. Much of orthodox non-equilibrium theory consists in getting rid of such dubious posits.

Two important approaches to non-equilibrium theory are those of 'mixing' and of the 'rigorous derivation of the Boltzmann equation'. The former approach generalizes the results of the ergodic theorem to rationalize a model presented by Gibbs. In this approach one tries to show that for appropriately idealized systems, at least in the infinite time limit, any initial probability distribution assigned to a system and associated with its initial non-equilibrium condition will approach the probability distribution associated with equilibrium for the system in a 'coarse-grained' sense. One uses only the dynamical laws and imposes no re-randomizing. The kinetic equation is taken as having its solution curve give the sequence of 'most probable states' of a collection of systems all started in a common non-equilibrium condition, but having different initial micro-states.

The latter approach tries to show that in a different idealized circumstance (the Boltzmann-Grad limit) the kinetic equation solution can be correctly asserted to show, at least for short time intervals, the overwhelmingly probable evolution of members of a collection of systems described by a special initial probability distribution.

Important questions arise as to the nature and role of the idealizations used and the kind of probabilistic explanation of the experimentally determined nature of approach to equilibrium offered by these alternative

accounts. Especially important is the observation that they offer accounts of the physical processes that are conceptually at odds with one another, although they are not contradictory to one another since they posit quite different idealizations.

4 Asymmetry in time

A core issue of non-equilibrium theory remains the attempt to find an appropriate physical explanation for the fact that macroscopic systems show asymmetry in time in their thermodynamic features, whereas the underlying laws governing the micro-constituents are taken to be time-symmetric.

A number of unorthodox approaches to the problem exist. In some the underlying dynamical laws have their time symmetry denied. In others the origin of time asymmetry of systems, systems usually taken to be energetically isolated from the world, is accounted for in terms of their ineliminable residual causal interaction with their external environments. A subjectivist approach to the second law exists as well.

In the orthodox accounts a crucial role is played by the initial probability distribution over micro-states assigned to a non-equilibrium system. In the ‘rigorous derivation’ approach, this initial probability assignment is necessary to derive the asymmetric Boltzmann behaviour. In the ‘mixing’ approach it is noted that mixing and its generalizations are all time-symmetric notions. But whereas a non-equilibrium system will coarsely approach equilibrium in both the infinite future and the infinite past, it is possible to specify an initial probability distribution that will behave asymmetrically in time for finite time intervals. Another approach, due to Prigogine (1980), focuses on singular initial probability distributions to break the time symmetry.

In order to account for the asymmetry in time of the behaviour of individual systems, resort is often made to the behaviour in time of the cosmos as a whole. In the new cosmological models one thinks of the overall universe not as in equilibrium, as Boltzmann speculated it was, but as in expansion from an initial Big Bang singularity and showing entropic increase.

Why does the entropy of the universe increase in time? Some accounts (Gold) tie the entropy increase to the expansion of the universe. The more common view is, however, that expansion by itself will not explain the entropic asymmetry of the cosmos. Rather one must assume an initial low-entropy state for the universe, not paralleled by a low-entropy final state. One important version of this approach (Penrose) takes the initial low entropy to be found in the ‘smooth’ initial spacetime structure (see Cosmology §3). The matter of the world starts thermalized and then clumps into hot stars living in cold space, the decrease in entropy of the matter being paid for by the increased ‘clumpiness’ of spacetime, a change that corresponds to an increase of entropy overall.

Many profound philosophical questions arise when one speculates about what kind of an answer could be given to such questions as: ‘why is the initial entropy of the universe so surprisingly low?’ Here philosophical discussion is reminiscent of older arguments about the origin of order in the universe and teleological arguments for the existence of a deity.

The purpose of invoking the entropic asymmetry of the cosmos as a whole was to give a physical account of the ground of the second law of thermodynamics. But cosmic entropy increase can perform this task only if it can be shown that it can account for the parallel increase in entropy in time of the individual systems with which the second law deals. Reichenbach calls these individual systems, energetically isolated from the overall cosmos only for finite time intervals, ‘branch systems’ (Reichenbach 1956).

One could simply posit that branch systems show an entropy asymmetry in time parallel to one another and parallel in time to the entropic asymmetry of the cosmic system. Several proposals exist to do away with this posit, but all of them, including Reichenbach’s own (positing that the collection of branch systems constitutes what he called a ‘lattice of mixture’), ultimately introduce parallelism of the direction of entropy increase for the branch systems with each other and with the main system as a posit. The physical origin of the temporal asymmetry of systems remains a very controversial subject.

5 The direction of time

Suppose we accept it as a fact that branch systems do indeed show parallel increase of entropy in the same direction of time. The question remains, ‘Why do systems have their entropy increase towards the future and not

towards the past?' It was Boltzmann who first suggested the answer that our very conception of the past-future distinction in time is one that is itself founded upon the asymmetry in time of the entropy of systems. What counts as the future direction of time, he claimed, is just the direction of time in which systems have their entropy increasing.

In order to evaluate this claim one needs to know what it really amounts to. It is sometimes suggested that the claim amounts to the assertion that we determine which direction of time is the future by examining the entropic states of systems. Thus we can tell if a film has been put in a projector in the correct direction only by consulting the entropic features of the states it pictures. But such an understanding of Boltzmann's claim founders on the fact that we have direct, non-inferential knowledge for many pairs of states as to which member of the pair is later than the other.

A more plausible reading of the Boltzmann claim takes it that entropic facts ground our intuitive means for determining time order in much the same way that facts about gravity ground our distinction between the directions in space we take as upward and downward. From this point of view the entropic asymmetry of the world offers a full explanatory account of all the other general asymmetries of the world we intuitively take to hold. These intuitive asymmetries include our differential epistemic access to past and future (we remember the past and have records of it, but not the future); the direction of causation from past to future; our differential concern with past and future; and the alleged 'fixity' of the past but openness of the future. Although interesting arguments have been given to back up this claim, it remains an open one (see Causation; Time).

LAWRENCE SKLAR

References and Further Reading

Arnold, V. and Avez, A. (1968) *Ergodic Problems of Classical Mechanics*, New York: Benjamin.(Ergodicity and its generalizations treated with mathematical sophistication.)

Batterman, R. (1990) 'Irreversibility and Statistical Mechanics: A New Approach?', *Philosophy of Science* 57 (3): 395-419.(Discussion of Krylov's approach to non-equilibrium.)

Batterman, R. (1991) 'Randomness and Probability in Dynamical Theories: On the Proposals of the Prigogine School', *Philosophy of Science* 58 (2): 241-63.(Discussion of Prigogine theory of non-equilibrium.)

Brush, S. (ed.) (1965) *Kinetic Theory*, Oxford: Pergamon Press.(Collection of important original papers in kinetic theory and statistical mechanics.)

Brush, S. (1983) *Statistical Physics and the Atomic Theory of Matter*, Princeton, NJ: Princeton University Press. (Outline of foundational issues.)

Buchdahl, H. (1966) *The Concepts of Classical Thermodynamics*, Cambridge: Cambridge University Press. (Thermodynamics from foundational point of view.)

Davies, P. (1974) *The Physics of Time Asymmetry*, Berkeley, CA: University of California Press.(Outline of physical aspects of asymmetry in time.)

Earman, J. (1974) 'An Attempt to Add a Little Direction to "The Problem of the Direction of Time"' *Philosophy of Science* 41 (1): 15-47.(Philosophical discussion of entropy asymmetry and the direction of time.)

Ehrenfest, P. and Ehrenfest, T. (1959) *The Conceptual Foundations of the Statistical Approach in Mechanics*, Ithaca, NY: Cornell University Press.(Important foundational review, originally published in 1910.)

Horwich, P. (1987) *Asymmetries in Time*, Cambridge, MA: MIT Press.(Philosophical discussion of asymmetries in time and the direction of time problem.)

Jancel, R. (1969) *Foundations of Classical and Quantum Statistical Mechanics*, Oxford: Pergamon Press.(Physics of statistical mechanics from a foundational point of view.)

Jaynes, E. (1983) *Papers on Probability, Statistics and Statistical Physics*, Dordrecht: Reidel.(Subjectivist approach to statistical mechanics.)

Katz, A. (1967) *Principles of Statistical Mechanics*, San Francisco: W.H. Freeman.(Subjectivist statistical mechanics.)

Krylov, N. (1979) *Works on the Foundations of Statistical Physics*, Princeton, NJ: Princeton University Press. (Foundational aspects of statistical physics.)

Lewis, D. (1986) 'Counterfactual Dependence and Time's Arrow', in D. Lewis, *Philosophical Papers*, vol. II, Oxford: Oxford University Press.(Intuitive asymmetries and the direction of time.)

Prigogine, I. (1980) *From Being to Becoming*, San Francisco: W.H. Freeman.(Nontechnical discussion of foundational issues.)

Prigogine, I. (1984) *Order Out of Chaos*, New York: Benjamin.(Introduction to foundational problems including the perspective of the author's school.)

Reichenbach, H. (1956) *The Direction of Time*, Berkeley, CA: University of California Press.(Foundations of statistical physics with emphasis on the direction of time problem.)

Sklar, L. (1992) *Philosophy of Physics*, Boulder, CO: Westview.(Elementary survey of the issues in chapter 3.)

Sklar, L. (1993) *Physics and Chance*, Cambridge: Cambridge University Press. (Comprehensive survey of the physical and philosophical issues.)

Tolman, R. (1938) *The Principles of Statistical Mechanics*, Oxford: Oxford University Press.(Clear foundational treatment of statistical mechanics at the time of its publication.)

Zeh, H.-D. (1989) *The Direction of Time*, Berlin: Springer.(Survey of time asymmetries in physics.)

Thielicke, Helmut (1908-86)

Helmut Thielicke presented a systematic Christian theology with particular reference to concrete issues and situations within the world. He held that statements about God are inevitably conditioned by the fact that God reveals himself to and through humanity; they thus always have an anthropological aspect. Similarly, while theology arises from transcendent revelation, it necessarily belongs to a particular historical context. Thielicke maintained that the notion of a 'perennial theology' is therefore mistaken. This raises the question as to whether it is a perennial truth that there is no perennial theology, and, if so, how Thielicke could, on his grounds, know this.

Born in 1908, Helmut Thielicke was a Lutheran theologian whose preaching was silenced by the Nazis in 1940. He was a professor at Heidelberg, Tübingen and, from 1954, Hamburg. He died in 1986.

Thielicke expounded a Lutheran (but also typically Christian) doctrine of justification, the doctrine of forgiveness of sins through the death and resurrection of Christ. He aimed to decline 'the doctrine of justification through all the case forms in which it appears within the grammar of our existence' ([1959] 1969: xiv). As the Latin professor teaches by putting nouns through their declensions, Thielicke presented Christian theology by weaving a Christian perspective on politics, society, economics, sexuality, law and art. He used concrete issues (for example, truthfulness at the sick bed) as models: the doctrines and principles of Christian ethics are explained in terms of specific situations and described in detail. He wrote:

> Since God discloses himself in man and seeks to be his God, any statement about God is also a statement about his relation to man. To that degree man is there in every theological utterance and thus gives it an anthropological reference… [this] book is about the relation between God and man and not about a purely transcendent God taken theistically in and for himself. I understand primarily by this relation an ontic rather than a noetic relation. Man 'is' his relation to God. He is the one created by him, in flight from him, visited by him and justified.
>
> ([1968] 1974: 15)

That each human person is created and constantly sustained by God, and that each person's good lies in trust and obedience to God, structures Thielicke's ethic.

Thielicke saw it as the nature of theology to respond to challenges. In this lies its potential for growth and continued relevance. He sought a worldview that was uncompromisingly Christian, preachable in and to the Church, and plainly of relevance to moral problems. He interacted with the best of contemporary thought without sacrificing the distinctiveness of the Christian gospel.

In *Der evangelische Glaube* (*The Evangelical Faith*), he wrote that:

> [God's] Word is historical not merely in the sense of being grounded in history [that is, in part concerning the religious significance of historical events in Israel's history and the life of Jesus] but also as it is addressed to historical situation. Both the authors and the recipients of verbal messages are subject to the process of history. The message, then, cannot be detached at either point. If an effort is made in this direction, there arises the false notion of perennial theology characterized by an abstract conceptual system. Scholasticism and seventeenth century orthodoxy are classical examples.
>
> ([1968] 1974: 23)

Theology necessarily arises from a transcendent revelation conditioned by the cultural context of, and the challenges peculiar to, a particular time. It exists in an inescapable tension, requiring faithfulness to its status as revelation and sensitivity to the historically conditioned nature of its proclamations. In effect, Thielicke wanted to retain what he saw as the best in his evangelical colleagues and his neo-orthodox predecessors (such as Karl Barth and Emil Brunner).

There remains this question: can a theology that allows only the sorts of statements about God that Thielicke permitted have any content, let alone respond to contemporary challenges? If there are no statements concerning

God save those that describe God's relations to humans, then those relational statements themselves do not inform us about God's nature. But then it is hard to see what the relational statements themselves mean. If 'God loves us' does not entail 'God is by nature loving', or anything else about God's nature, it is unclear just what 'loves' means.

In addition, Thielicke's statement that theology must rest on a transcendent revelation only expressible in the terms of some particular culture is itself expressed in the terms of Thielicke's culture. Yet it is supposed to be categorically and universally true, and hence it is supposed to escape being conditioned by Thielicke's culture. But if that statement can somehow, or in some degree, escape such conditioning, is it impossible that theological claims should somehow, or in some degree, do so as well?

KEITH E. YANDELL

List of works

Thielicke, H. (1958, 1959, 1964) *Theologische Ethik*, Tübingen: Mohr (Paul Siebeck), 3 vols; ed. W.H. Lazareth, *Theological Ethics*, Philadelphia, PA: Fortress Press, vols 1-2, 1966, 1969, and Grand Rapids, MI: Baker Book House, vol. 3, 1975. (The English translation is an abridgement of the German work.)

Thielicke, H. (1964) *Mensch Sein - Mensch Werden*, Munich: R. Piper Verlag; *Being Human… Becoming Human*, Garden City, NY: Doubleday, 1976.(Presents Thielicke's anthropology in the context of dialogue with competing perspectives.)

Thielicke, H. (1968, 1973, 1978) *Der evangelische Glaube*, Tübingen: Mohr (Paul Siebeck), 3 vols; trans. and ed. G.W. Bromiley, *The Evangelical Faith*, Grand Rapids, MI: Eerdmans, vols 1-2, 1974, 1977, and Edinburgh: T. & T. Clark, vol. 3, 1982.(Relates Christian theology to modern forms of thought.)

Thielicke, H. (1988) *Glauben und Denken in der Neuzeit*, Tübingen: Mohr (Paul Siebeck); trans. G.W. Bromiley, *Modern Faith and Thought*, Grand Rapids, MI: Eerdmans, 1990.(Thielicke surveys and assesses contemporary theological movements.)

References and further reading

McGrath, A.E. (1994) *Christian Theology: An Introduction*, Oxford: Blackwell. (Accessible introduction to many of the issues that concerned Thielicke.)

Nygren, A. (1972) *Meaning and Method: Prolegomena to a Scientific Philosophy of Religion and a Scientific Theology*, trans. P.S. Watson, London: Epworth Press.(Authorized translation of an unpublished Swedish manuscript that presents an alternative approach to Thielicke's.)

Thierry of Chartres (*fl. c.*1130-50)

Thierry of Chartres, who taught at Paris and Chartres in the mid-twelfth century, was a polymath and a Platonist. The Heptateuchon, a large and ambitious collection of texts for teaching the liberal arts, testifies to the range of his interests from grammar, logic and rhetoric to mathematics and astronomy; they also stretched to theology. To Thierry is attributed an explanation of the account of creation in Genesis, after God's initial action, in physical terms. He also used arithmetical analogies to illustrate the Trinity and, drawing on a variety of Platonic and Neoplatonic sources, analysed the relationship between God and his creation.

Thierry of Chartres was born in Brittany (he usually called himself Theodoricus Brito, Thierry the Breton). In the 1140s he was chancellor of Chartres Cathedral, before he retired in the 1150s to a monastery. He taught, almost certainly in both Paris and Chartres, from the 1130s or earlier. Among his pupils were John of Salisbury, the grammarian Peter Helias and the scientist and translator Hermann of Carinthia, who described him as 'the soul of Plato restored to mankind from heaven'. Unfortunately, the records of his wide range of teaching are scant. His commentaries (*c.*1130-9) on the two main textbooks of rhetoric, Cicero's *On Invention* and the pseudo-Ciceronian *Rhetorica ad Herennium* survive. His logical writings have been lost, although Thierry is known to have been one of the earliest writers to teach Aristotle's *On Sophistical Refutations* (before 1132) and perhaps also the *Prior Analytics*. He also taught grammar, and the *Heptateuchon* (which includes extracts from Euclid and al-Khwarizmi's astronomical tables) shows his lively interest in the various branches of mathematics. Thierry's modern reputation as a philosopher depends mainly, however, on a short treatise *De sex dierum operibus (On the Work of the Six Days)* and on various commentaries and glosses to Boethius' theological treatises (*Opuscula sacra*). Yet these too pose considerable problems of integrity and authorship.

It was Thierry's pupil Clarembald of Arras who gave *De sex dierum operibus* its title and attributed it to Thierry. He added to it a further collection of remarks on Genesis, taken from Thierry's lectures. Sir Richard Southern's suggestion (Southern 1970) that the treatise itself (which circulated separately in other manuscripts) was also reworked by Clarembald is implausible; but the work is certainly unfinished in all surviving copies. As it stands, it divides into two parts. The first offers what is called a 'natural and literal' exposition of the first five verses of Genesis. Here Thierry is following a tradition going back to Augustine and the Greek Fathers of using scientific material to gloss and expand the biblical account of creation. However, he is unusually rigorous in restricting himself to naturalistic explanation, drawing on Plato's *Timaeus* and Calcidius' commentary and using a number of ideas he may have taken from William of Conches. The second section is a long (and unfinished) digression on the nature of God and his triunity. Thierry promises to use four methods of mathematical reasoning to lead us to knowledge of the creator, arithmetical, musical, geometrical and astronomical; but only the arithmetical 'proofs' are preserved. God, who is the 'form of being' for individual things, is unity. The unities from which numbers are made are participations in this divine unity. The Son is related to the Father as 'equality' to unity and, when the treatise breaks off, Thierry is about to discuss the Holy Spirit as the link (*connexio*) between unity and equality.

To Clarembald also is owed the identification of Thierry's most important philosophical work. From his remarks, it is clear that Clarembald based his own commentary on Boethius' *De trinitate* (On the Trinity) mainly on Thierry; and comparison shows that a commentary known as *Librum hunc* (written late 1140s; incipit 'Inchoantibus librum hunc', edited as the *Commentum*) is, or closely resembles, Clarembald's source. *Librum hunc*, therefore, is certainly based on Thierry's teaching, but exactly how accurately remains an open question. Two other commentaries on the *De trinitate* related to *Librum hunc* have been argued by his editor to be later works of Thierry: the *Lectiones (Lectures: incipit 'Intentio auctoris')* and the *Glosa* (Gloss: incipit 'Aggreditur propositum'). But do the doctrines they contain, often at variance with those of *Librum hunc*, all go back to the same man?

Besides developing a numerological account of the Trinity similar to that in Thierry's treatise on the creation, *Librum hunc* shows a concern to develop Boethius' somewhat disparate remarks on matter and form into a theory about God's relation to his creation. God is said to be the only true form. By contact with matter this form is determined into forms of different kinds (humanity, stone-ness and so on). These forms cannot exist apart from the individuals they inform. Were there to be, for instance, no human beings, the form humanity would lose its specific

identity and return to the simplicity of the one true form, God. Both the *Lectiones* and the *Glosa* share this view of God as the single true form, but they also admit a multiplicity of pure forms as ideas in the mind of God; and the *Glosa* elaborates this plan in terms of act and potentiality. Peter Dronke has made a bold attempt to find a unity of thought in these various works by or connected with Thierry. He argues (looking to the *Lectiones* and the *Glosa*) that Thierry distinguished four modes of envisaging the universe: as 'absolute necessity' or the 'form of forms'; 'necessity of make-up' (linked to Ideas, natural law and the World Soul); absolute possibility (primordial matter); and determinate possibility (contingent reality). By using this diversity of perspectives, Thierry is able to reconcile 'an extreme Platonism - in which forms and names exist indissolubly in the mind of God, and in which "names essentiate things" - with a far reaching naturalism' (Dronke 1988: 384).

See also: Chartres, School of; Clarembald of Arras

JOHN MARENBON

List of works

Thierry of Chartres (1130-39) Commentary on Cicero's *On Invention*, ed. K.M. Fredborg, *The Latin Rhetorical Commentaries by Thierry of Chartres*, Studies and Texts 84, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1988.(Thierry's Ciceronian commentary.)

Thierry of Chartres (1130-39) Commentary on *Rhetorica ad Herennium*, ed. K.M. Fredborg, *The Latin Rhetorical Commentaries by Thierry of Chartres*, Studies and Texts 84, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1988.(Shows Thierry's work as a teacher of rhetoric.)

Thierry of Chartres (*c*.1150) *Heptateuchon*.(The original manuscripts of this work from the Chartres library have been destroyed, but a microfilm exists.)

Thierry of Chartres (*c*.1150) *De sex dierum operibus (On the Work of Six Days)*, ed. N.M. Häring, *Commentaries on Boethius by Thierry of Chartres and his School*, Studies and Texts 20, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1971, 557-75.(Naturalistic account of the stages of creation.)

Thierry of Chartres (*c*.1150) *Librum hunc (Commentum)*, ed. N.M. Häring, *Commentaries on Boethius by Thierry of Chartres and his School*, Studies and Texts 20, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1971, 57-121.(Based on Thierry's teaching, but not certainly his work.)

Thierry of Chartres (*c*.1150) *Lectiones (Lectures)*, ed. N.M. Häring, *Commentaries on Boethius by Thierry of Chartres and his School*, Studies and Texts 20, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1971, 125-229.(Based on Thierry's teaching, but not certainly his work.)

Thierry of Chartres (*c*.1150) *Glosa (Gloss)*, ed. N.M. Häring, *Commentaries on Boethius by Thierry of Chartres and his School*, Studies and Texts 20, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1971, 259-300. (Based on Thierry's teaching, but not certainly his work.)

References and further reading

Dronke, P. (1988) 'Thierry of Chartres', in P. Dronke (ed.) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press, 358-85.(A bold reconstruction of Thierry's underlying ideas.)

Gersh, S. (1982) 'Platonism - Neoplatonism - Aristotelianism: a Twelfth-Century Metaphysical System and its Sources', in R.L. Benson and G. Constable (eds) *Renaissance and Renewal in the Twelfth Century*, Oxford: Clarendon Press, 512-34. (The *Librum hunc* and its Neoplatonic sources.)

Häring, N.M. (1965) *Life and Works of Clarembald of Arras*, Studies and Texts 10, Toronto: Pontifical Institute of Mediaeval Studies.(Arguments for attributing *Librum hunc* and related commentaries to Thierry.)

Jeauneau, E. (1973) *Lectio philosophorum*, Amsterdam: Hakkert.(Includes essays on the *Heptateuchon* and on Thierry's mathematical discussions of the Trinity.)

Southern, R.W. (1970) 'Humanism and the school of Chartres', *Medieval Humanism and Other Studies*, Oxford: Blackwell, 61-85.(The canon of Thierry's works and its relation to twelfth-century Platonism.)

Thomas of York (*fl. c.*1255)

A philosopher of remarkably wide reading in the works of Western and non-Western thinkers, Thomas attempted to assemble, and to some extent to synthesize, the views he had encountered. Whatever the eventual judgement on Thomas' thought, the Sapientiale (Wisdom), a lengthy metaphysical work, certainly reflects the intellectual ferment brought about in the thirteenth century by the influx of new works by Aristotle and by Jewish and Islamic philosophers and commentators.

A member of the Franciscan order from at least 1245, Thomas of York incepted as master of theology at Oxford in 1253. He was lecturer to the Oxford Franciscans until 1256, and to the Cambridge convent in 1256-7. He wrote a work entitled the *Manus quae contra omnipotentem (The Hand That is Raised Against the Omnipotent)*, a polemic work defending voluntary poverty and mendicancy, but his most important philosophical work is the *Sapientiale (Wisdom)*.

The *Sapientiale* is a summa - or it might be more properly called a compendium - of arguments addressing what Thomas labels as 'general' and 'special' metaphysics. The first five books discuss God, creation, being and causation, substance and accident, and attributes of being such as truth and falsity, while the remaining section on special metaphysics discusses the soul. The second section is incomplete: it breaks off unfinished, omitting projected discussions of the inanimate world and of plants and animals. Prominent throughout is Thomas' concern to include extensive and accurate citations from a very wide range of sources (including writers of antiquity, Jewish and Islamic philosophers and commentators, patristic writers and Thomas's predecessors at Oxford and Paris), and where possible to adjudicate between or reconcile opposed views. In addition to the difficulty of fixing Thomas' views amid the remarkable breadth and extent of his citations - unmatched in any later Franciscan philosophical work - there is the problem that an edition of the *Sapientiale* is not yet available.

In his general metaphysics, Thomas develops an elaborate explanation of the natures of matter and form. While the complexity of his scheme may in part be attributed to his efforts to accommodate both Aristotelian and Neoplatonic systems, its intrinsic interest has yet to be evaluated. Thomas equates form and matter with act and potency, and identifies these as the two principles of created being. Not willing to admit a third principle, he classes privation as an attribute of matter. However, since not all matter is subject to privation, he must distinguish three varieties of matter: 1) matter subject to privation, and hence to generation and corruption; 2) matter having dimension and location, but not subject to privation (such as heavenly bodies); and 3) 'spiritual matter', a potentiality capable of taking on form but neither subject to privation nor having any particular location or dimension. Thomas also distinguishes a 'potential' variety of form. He attempts to reconcile those who insist that common forms or natures (such as 'man' or 'white') exist independently, separate from their realization in matter and outside the minds of those who understand them, with their opponents who deny any such separate existence to forms. He does so by proposing that common forms do exist outside the mind, for otherwise demonstrative knowledge would be impossible, but that their external existence is merely potential, awaiting realization in matter.

Thomas' special metaphysics focuses on the human soul but does not, as might be expected, discuss philosophical psychology. Instead, it is largely devoted to the metaphysics of the rational soul, arguing that the soul is not a form but rather is compounded from form and spiritual matter and related to the body as a pilot to a ship. (Thomas does not offer any solution to the potential difficulty that creatures with souls would then seem to have more than one substantial form.) Considerable attention has nonetheless been devoted to Thomas's philosophy of mind. However, granted that Thomas affirms earlier in the *Sapientiale* that some human knowledge derives 'from something higher and not from sense', and that he notes the incompleteness of Aristotle's account of the soul - the only one, he says, to assert that all knowledge stems from the senses - Thomas nonetheless devotes the bulk of his attention to explaining how the intellect apprehends though a process of abstraction from the data of sensory perception. His acceptance that some human knowledge derives from divine illumination is standard rather than controversial: Aquinas, whom it has been suggested he opposed on this issue, also granted it a restricted role.

See also: Aristotelianism, medieval; Matter

FIONA SOMERSET

List of works

Thomas of York (*c.*1256) *Manus quae contra omnipotentem (The Hand That is Raised Against the Omnipotent)*, ed. M. Bierbaum, *Franziskanische Studien* Beiheft 2, 37-168.(A polemic work defending voluntary poverty and mendicancy.)

Thomas of York (*c.*1250-60) *Sapientiale (Wisdom)*.(No complete edition of this work is available. Portions edited in four University of Toronto dissertations are not yet available on microfilm. Book II, chapters 4, 5 and 6 appear in E. Longpré, 'Thomas d'York et Matthieu d'Aquasparta', *Archives d'histoire doctrinale et littéraire du Moyen Age* 1, 1926: 268-308.)

References and further reading

Emden, A.B. (1957-9) *A Biographical Register of the University of Oxford to a. d. 1500*, 3 vols, Oxford: Clarendon Press.(Gives all known details of the lives and works of Oxford scholars to 1500.)

Gilson, E. (1955) *History of Christian Philosophy in the Middle Ages*, London: Sheed & Ward, 666 n.51.(A summary of Treserra 1929, in English.)

Grabmann, M. (1913) 'Die Metaphysik des Thomas von York' (The Metaphysics of Thomas of York), Beiträge zur Geschichte der Philosophie (und Theologie) des Mittelalters Suppbd. I, Münster: Aschendorff, 181-93. (Includes an overview briefly summarizing the contents of each book of the *Sapientiale*.)

Longpré, E. (1926) 'F. Thomas d'York, O.F.M.', *Archivum Franciscanum Historium* 19: 875-930. (Includes a transcription of the table of contents to the *Sapientiale*.)

Reilly, J.P., Jr (1953) 'Thomas of York on the Efficacy of Secondary Causes', *Mediaeval Studies* 15: 225-32. (Explains Thomas' theory of secondary causation and contrasts it with his theory of knowledge. The reader should note that Thomas discusses secondary causation in animal and human reproduction rather than in activity more generally.)

Sharp, D.E. (1930) *Franciscan Philosophy at Oxford in the Thirteenth Century*, London: Oxford University Press, 49-112.(The only detailed treatment of the whole of the *Sapientiale* available in English. Unfortunately the discussion is organized under rather idiosyncratic categories applied in turn to a series of philosophers, but some parts are still useful. A full list of works cited by Thomas appears on pages 53-5.)

Treserra, F. (1929) 'De Doctrinis Metaphysicis Fratris Thomae de Eboraco OFM' (The Metaphysical Teachings of Thomas of York OFM), *Analecta Sacra Tarraconensis* 5: 33-102.(An exposition of Book III of the *Sapientiale*.)

Thomas White (1593-1676)

Thomas White's reputation has suffered unmerited decline since he was described by John Evelyn in 1651 as 'a learned priest and famous philosopher'. His works embrace theology, metaphysics, natural philosophy and political theory. The leader of a minority faction of English Catholics, known after his alias as 'Blackloists', White's overall intellectual position is determinedly antisceptical, characterized by a certainty-seeking synthesis of old and new. The traditional Aristotelianism of his own education is blended with aspects of the 'new philosophy' which he encountered in the 1640s; and in this respect White stands as an important representative of the intellectually turbulent times in which he lived.

As an English Catholic, White was educated in continental Europe. He was particularly associated with the English College at Douai, where he was ordained as a priest in 1617, and at which for some years he taught scholastic philosophy, or 'rigid Thomistry'. His ideas were extended during periods in Rome from 1625-9, in Paris in the 1640s and in London after 1655. His intellectual contacts included members of the Mersenne circle and of the embryonic Royal Society (see Mersenne, M.). White's thought is remarkable for embracing such varied influences within an essentially Aristotelian framework.

White's antisceptical metaphysics is most clearly outlined in his debate with Joseph Glanvill. Responding to the fashionably Pyrrhonian *Vanity of Dogmatizing* (1661), White argues in his *Sciri, sive sceptices* (1663) that a sceptical denial of the possibility of certain knowledge not only insults past and present philosophers, but also threatens the very basis of morality. Seeking some fundamental starting-point, and specifically rejecting the validity of Descartes' introspective quest for an infallible generator of knowledge, White resorts to an essentially Aristotelian, logical procedure. He thus assumes the existence of some unquestionable 'identical propositions', from which certain knowledge can be generated: in its simplest form, what is, is, and 'whilst it is, it cannot not-be'. This invincible truth provides at least one 'first step' against the sceptics.

As a second step, 'self-known propositions' are cited. For example, a number must be odd or even, on the assumption that 'even and not-even are all'; or a man must be an animal, where 'rational animal' is subsumed within the category 'animal'. From such propositions, it is then possible to proceed by syllogistic reasoning: thus, we may start with the self-known propositions that 'every man is an animal', and that 'every animal is a living creature', and from these we may go on justifiably to conclude that what is a 'man' is a 'living creature'. Though obviously limited in scope, such examples served to demonstrate for White that at least some certain knowledge was in principle possible; and on this solid foundation, more could then be built.

The defeat of scepticism was particularly important in a theological (and related moral) context. Though denying some important Catholic orthodoxies, including the widespread belief in papal authority and conventional notions of purgatory, White retained an emphasis on 'tradition' as the rule of faith, and additionally synthesized theology with his own brand of natural philosophy, to produce a so-called 'philosophical divinity'. Catholic tradition was to be seen as perfectly compatible with a natural philosophy grounded on Aristotelian first-principles and incorporating important elements of the new science. White's approach here is well illustrated in the Theological Appendix to his *Institutionum Peripateticarum* (1646). Here the seemingly miraculous story of the Flood is shown to be in perfect conformity with the rational and empirical methods of the new science; and the two approaches are to be seen as mutually-supportive in their stand against scepticism.

White's natural philosophy is best studied in *De Mundo* (1642) and *Institutionum Peripateticarum* (1646). It is, again, particularly interesting for its synthesizing of old and new philosophies. This is evident in both cosmology and physics. With some ambiguities, Copernican cosmology can be shown to be consistent with Aristotelian philosophy and Christian theology. Thus while heliocentrism is nominally accepted, the earth is still described as 'situated in the very middle of the universe': this apparent anomaly can allegedly be justified by identifying the 'middle' not with the centre of the terrestrial globe but with the whole circumference of the earth's orbit; and that ploy is legitimated by reference to the magnitude of the universe, in terms of which that circumference is itself a mere 'point'. The earth's continuing motor (still required by an Aristotelian) is found in the wind acting on contiguous sea; and the air surrounding the earth is also assumed to move, circumventing another traditional scholastic argument. Finally, Christian scripturally-based insistence on an immobile earth can be appeased by an

appeal to the relativity of motion, in terms of which 'both statements are made with absolute correctness and accuracy: that the earth has been continually moved with regard to the sun, and the sun with respect to the earth'.

White's ambivalence in cosmology is paralleled in his physics: retaining such Aristotelian premises as the impossibility of a vacuum, motion's need of a mover, and the infinite divisibility of a continuum, he subscribes none the less to a form of atomic theory, by conflating newly-revived Epicurean 'atoms' with traditional Aristotelian 'minima'. In this White may be compared to his fellow-Blackloist friend Kenelm Digby.

Blackloist political philosophy is outlined in White's *Grounds of Obedience and Government* (1655). Aiming for toleration of Catholic worship, and from the starting-point of 'salus populi', White argues essentially for a contract theory, whereby the king's power is held in trust for the people. Revolution may therefore be justified; and in practical terms Catholics should accept the *de facto* rule of Cromwell. White's outspokenness proved his downfall: following the restoration of the monarchy he was repudiated by Protestants and fellow-Catholics alike, and was all but lost to history.

See also: Aristotelianism in the 17th century; Digby, K.; Sergeant, J.

BEVERLEY SOUTHGATE

List of main works

White, T. (1642) *De Mundo Dialogi Tres (Three dialogues concerning the world)*, Paris.(White's first comprehensive treatise on natural philosophy, which established his reputation amongst contemporaries. The work was criticized by Hobbes (see below).

White, T. (1646) *Institutionum Peripateticarum… pars theorica*, Lyons; trans. as *Peripateticall Institutions*, London, 1656.(Cosmological treatise apparently influenced by Kenelm Digby. Scientific matters are subordinated to religious and moral concerns, especially in a Theological Appendix.)

White, T. (1653) *Villicationis suae de medio animarum statu…*, Paris; trans. as *The Middle State of Souls*, London, 1659.(Notorious theological work, which includes denial of purgatory. Condemned both by Roman authorities and by the English Parliament.)

White, T. (1655) *The Grounds of Obedience and Government*, London; repr. facsimile Farnborough, Gregg International Publishers Ltd, 1968.(White's much condemned, allegedly pro-Cromwellian, political treatise.)

White, T. (1657) *Euclides Physicus*, London.(Mathematically expressed work of physics, claimed to have influenced Leibniz.)

White, T. (1659) *Controversy-Logicke or the method to come to truth in debates of religion*, Paris.(An apologia for Catholicism, in response to the Protestant John Biddle.)

White, T. (1660) *Religion and Reason*, Paris.(A theological work aiming to show the compatibility of science and reason with Christianity.)

White, T. (1663) *Sciri, sive sceptices et scepticorum a jure disputationis exclusio*, London; trans. as *An Exclusion of Scepticks from all Title to Dispute*, London, 1665.(White's anti-sceptical response to Joseph Glanvill's *Vanity of Dogmatizing*.)

References and further reading

Henry, J. (1982) 'Atomism and Eschatology: Catholicism and natural philosophy in the inter-regnum', *British Journal for the History of Science* 15: 211-39.(Argues for importance of 'Blackloists' in early development of 'mechanical philosophy'.)

Hobbes, T. (*c*.1642-3) *Thomas Hobbes: Thomas White's De Mundo Examined*, ed. H.W. Jones, London: Bradford University Press/Crosby Lockwood Staples, 1976. (Translation of Hobbes' critique of White's cosmology.)

Jones, G.H. (1975) 'Leibniz' Cosmology and Thomas White's *Euclides Physicus*', *Archives Internationales d'Histoire des Sciences* 25: 277-303.(Discusses White's originality, and claims his influence on Leibniz.)

Southgate, B.C. (1993) *'Covetous of Truth': The Life and Work of Thomas White, 1593-1676*, Dordrecht: Kluwer Academic Publishers.(Includes bibliography.)

Tavard, G.H. (1978) *The Seventeenth Century Tradition: A Study in Recusant Thought*, Leiden: E.J. Brill. (Includes chapter on White's theology.)

Thomas à Kempis (1379/80-1471)

*Thomas Hemerken was born in Kempen, Germany. He spent his life in foundations of the Modern Devotion (*Devotio Moderna*), a spiritual movement of the late fourteenth and fifteenth centuries that originated in the Low Countries and spread throughout northern Europe. In 1406 he entered the monastery of Mount Saint Agnes in Windesheim (St Agnietenberg, the Netherlands), the origin and centre of a reformed congregation of Augustinian Canons Regular, which disseminated the Modern Devotion in the Low Countries and Germany. Thomas was ordained a priest in 1413, and was the novice master in the monastery for many years. He is generally recognized as the author of De imitatione Christi (The Imitation of Christ), perhaps the most popular work on the spiritual life ever written.*

Thomas wrote a number of spiritual and ascetic treatises, meditations on the life of Christ, lives of the leading figures of the Modern Devotion and a chronicle of his monastery. Most of these writings can be related to his duties as novice master. He is most famous as the author of *De imitatione Christi (The Imitation of Christ)*. His authorship of this work, however, has always been disputed. His claim rests mainly on the tradition of his religious order. *De imitatione Christi* comprises four treatises, composed separately. Some early copies (1420s) of the treatises attribute them to Thomas, but most are anonymous. Thomas copied the treatises more than once; one copy made by him, containing the four treatises and completed in 1441, bears corrections indicating that he worked on the manuscript for many years. The manuscript suggests that if Thomas was not the original composer of all the texts, he was their compiler and editor. However, none of the evidence is conclusive. Other evidence points to a Carthusian monk as the author, possibly as early as the end of the fourteenth century; Carthusians were closely associated with early foundations of the Modern Devotion, and through them the core texts of *De imitatione Christi* may have circulated among the Devout (that is, the followers of the Modern Devotion).

The style of *De imitatione* complicates attribution. The Devout were accustomed to collect texts of spiritual authorities, ordering them under general topics for their meditation and personal use. The books of *De imitatione* are related to such collections, weaving together ‘sentences’, or moral and psychological maxims, into a continuous discourse. The ‘sentences’ of *De imitatione* are drawn largely from the Scriptures, but also from anonymous teachers and a few named authorities. Doubtless the author crafted his own ‘sentences’ as well. Such a text readily became the common property of those in the monastic life; perhaps this is why *De imitatione* was later ascribed to so many authors.

Similarly, the treatises of *De imitatione* express the common themes of the Modern Devotion. They develop the dialectic between the world and the kingdom of God, which is within. They contrast the vanity of the world, the emptiness of sensible goods, false happiness, self-love and the pride of the intellect with the inner life of the kingdom of God, which requires purification of the senses and mind, abandonment to God, obedience, patience, humility and the practice of divine charity, which directs every act to God. These virtues yield interior peace (the ordering and calming of desires and passions), interior freedom (liberation from concupiscence and regard for human esteem), and simplicity of spirit. Christ is the exemplar of the soul’s dispositions; only by participation in, and responsiveness to, his divine grace can the disorder of sinful human nature be repaired and the kingdom of God within be attained. *De imitatione* seeks a practical reform of Christian life, in principle accessible to all. It avoids complex questions concerning mystical contemplation.

Although in its time the devotion of *De imitatione* was called ‘new’, it in fact echoes an older Christian wisdom, the chief sources of which are Augustine, Gregory the Great, John Cassian and Bernard of Clairvaux (see Patristic philosophy). Devout authors were sceptical of theological speculation which, they thought, induced vanity and was a diversion from the evangelical life. Some historians nevertheless see the Modern Devotion as a coefficient of the religious attitude underlying the scholastic *via moderna*. This interpretation is not wholly convincing. The ‘philosophic’ roots of *De imitatione Christi* lie elsewhere, as its ‘sententious’ and ‘gathering’ style suggests. The letters of Seneca, which treat self-knowledge, recollection of the memory, the ordering of the passions, and the necessity of solitude, were favourite readings among the Devout and the earlier monastic writers who inspired their spiritual teaching. The author of *De imitatione* adapts Stoic psychological teaching to the narrative of Christ’s life and death and to the Christian doctrines of sin, grace and redemption (see Stoicism).

See also: Grace; Mysticism, history of

KENT EMERY, JR

List of works

Thomas à Kempis (1379/80-1471) *Opera omnia (Complete Works)*, ed. M.J. Pohl, *Thomae Hemerken a Kempis Opera omnia*, 7 vols, Freiburg: Herder, 1902-22.(Reprinted Hildesheim, 1985. This is the standard critical edition of Thomas' works; *De imitatione Christi* is in vol. 2, 3-263.)

Thomas à Kempis (1379/80-1471) *De imitatione Christi (The Imitation of Christ)*, ed. L.M.J. Delaisse, *Le manuscrit autographe de Thomas à Kempis et 'L'Imitation de Jésus-Christ': Examen archéologique et édition diplomatique de Bruxellensis 5855-61*, Les Publications de Scriptorium 2, Paris: Erasme, 1956; ed. T. Lupi, *De imitatione Christi libri quatuor*, Vatican City: Libreria Editrice Vaticana, 1982.(Delaisse includes a detailed study of Thomas' 1441 copy of the treatises of *De imitatione*, with a diplomatic edition of the manuscript's Latin text; Lupi is an edition of the Latin text by one of the leading scholars of its manuscript transmission, who disputes the authorship of Thomas à Kempis. There are a number of English translations, including those of W.C. Creasy, Macon, GA: Mercer University Press, 1989 (a new reading of the 1441 Latin autograph manuscript); E. Daplyn, London: Marshall, Morgan & Scott, 1979 (from the full text of the autograph manuscript of 1441); G.F. Maine, London: Collins, 1957; and L. Sherley-Price, Hammondsworth: Penguin, 1952.)

References and further reading

Ampe, A. (1973) *L'Imitation de Jésus-Christ et son auteur: Reflexions critiques (The Imitation of Jesus Christ and its Author: Critical Reflections)*, Rome: Edizioni di storia e letteratura.(A detailed study of the question of authorship, which concludes that none of the previous arguments for one author or another is decisive.)

Ampe, A. and Spaapen, B. (1971) 'Imitation de Jésus-Christ: I. Le livre et auteur. II. Doctrine' (The Imitation of Jesus Christ: I. The Book and its Author. II. Doctrine), *Dictionnaire de spiritualité ascétique et mystique, histoire et doctrine*, Paris: Beauchesne, vol. 7, 2338-68.(Ampe here outlines the evidence concerning the authorship of *De imitatione*, and Spaapen studies the work's doctrine. The article contains an ample bibliography.)

Colish, M.L. (1985) *The Stoic Tradition from Antiquity to the Early Middle Ages*, Studies in the History of Christian Thought 34-5, Leiden: Brill, 2 vols.(The second volume contains surveys of the Stoic influences on Latin fathers and early monastic writers, including the major sources of the Modern Devotion: Jerome, Augustine, Cassian and Gregory the Great.)

Geurts, A.J. *et al.* (eds) (1985) *Moderne Devotie Figuren en Facetten: Tentoonstelling ter herdenking van het sterfjaar van Geert Grote 1384-1984, Catalogus (Figures and Facets of the Modern Devotion: An Exhibition on the Anniversary of the Death of Geert Grote, 1384-1984)*, Nijmegen: Katholieke Universiteit.(This exhibition catalogue is the best scholarly work on the different genres of of literature produced by the Devout. For 'collection books' and *De imitatione*, see pp 152-67, 204-11.)

Post, R.R. (1968) *The Modern Devotion: Confrontation with Reformation and Humanism*, Leiden: Brill.(The most authoritative historical study of the movement, disputing the notion that it was connected with 'humanism' and was a 'precursor' of the Protestant Reformation.)

Van Dijk, R.T.M. (1990) 'Thomas Hemerken à Kempis', *Dictionnaire de spiritualité ascétique et mystique, histoire et doctrine*, Paris: Beauchesne, vol. 15, 817-26.(An introduction to the life, works and doctrine of Thomas à Kempis; the bibliography updates the article of Ampe and Spaapen.)

Van Engen, J. (1988) *Devotio Moderna: Basic Writings*, New York: Paulist Press. (Translations of typical writings by the Devout, with an excellent introduction to the spirituality of the movement. The notes contain extensive bibliography.)

Thomasius (Thomas), Christian (1655-1728)

Christian Thomasius' stature as the 'founder' of the German Enlightenment has been the source of much debate. His many essays dealing with issues in moral enlightenment and law reform (bigamy, witchcraft, torture, heresy, adultery, the use of the vernacular and so on) certainly single him out from other seventeenth-century writers. He was the public philosopher par excellence, a suitable match for August Hermann Francke, the great public theologian. Both men spent most of their career in Halle (in Brandenburg), and it was there that Francke institutionalized pietism, just as Thomasius propagated secular natural law theory. Despite many tensions, pietism and modern natural law thereby fused into a social duty-ethics that was of the greatest importance in shaping the modern Prussian state. The basis for natural law was God's will and it was the attempt to follow this law that made humanity a moral species. Since humankind could not have any certain knowledge of the content *of God's law, the natural powers of the mind would have to be relied upon, and Thomasius' thought was an investigation into the nature and social effect of these powers. His best-known result was a series of linked divisions between law and morality, between public and private spheres, between external and internal obligation, and between action and intention.*

1 Life

Christian Thomasius was born in Leipzig, Saxony, the son of the jurist and philosopher Jacob Thomasius. He was educated at the University of Leipzig by Valentin Alberti, one of the most prominent critics of Pufendorf, and by his own father who some years previously had taught Leibniz. Thomasius took his law degree at the University of Frankfurt an der Oder in 1679. He spent his early career in Leipzig as an advocate, private lecturer and founder of a controversial periodical, *Monatsgespräche* (Monthly conversations), the first intellectual monthly written in German. He made it his cause to criticize Aristotelianism, orthodox Lutheranism and Roman jurisprudence, and his public doubts about the divine right of kingship eventually made it prudent for him to leave Saxony in 1690. His alternative programme of German history, natural law and German jurisprudence was, however, welcome in neighbouring Brandenburg, where Thomasius was instrumental in the foundation of the University of Halle in 1694. In 1710 he became Direktor (President) for life of this institution which had a formative influence on German Enlightenment universities. By the time of his death here in 1728, Thomasius had established himself as a towering cultural figure, with a public philosophy promoting polite sociability.

2 Theoretical philosophy

From a philosophical point of view, Thomasius' huge *oeuvre* divides into three periods. In the 1680s he worked to complete Pufendorf's project of separating natural law from theology, culminating in 1688 with *Institutiones jurisprudentiæ divinæ (Institutions of divine jurisprudence)*. This was accompanied by *An Introduction to Court Philosophy* - that is to say, a theory of the centralist political culture that was necessary to maintain a state based upon the law of reason alone. Thomasius' reaction against orthodoxy broadened to a comprehensive philosophical outlook, resulting in works on reason and on moral theory in the 1690s. Strongly influenced by pietism, Thomasius pursued his doubts about the sufficiency of reason in the conduct of life, leading him to a spiritualist metaphysics in the tradition of fashionable mysticism. This prepared the way for a change after 1700 to a deterministic theory of the will and a scheme of moral, legal and political regulation based on governance of the passions.

Thomasius owed much to the long tradition of eclecticism in philosophy, but he can hardly be called an eclectic in a strict sense. His debt to eclecticism was largely methodological, not a substantive philosophical doctrine. And while he argued against the dominance of Aristotelianism and Cartesianism and other schools, he did not, in the manner of eclecticism, see his philosophy as inherently an argument for the freedom of the individual spirit from the tyranny of all 'sects'.

His epistemology was a basic empiricism inspired by Aristotle and, subsequently, reinforced by his reading of Locke's *Essay* (see Empiricism). But although he maintained that all knowledge has a basis in experience, Thomasius retained the notion of ideas of reason that are merely *awoken* through experience. These have a universality and certainty missing from the merely probable knowledge that is formed inductively from experience. It never becomes clear whether the ideas of reason are a remnant of humanity's pristine nature before the Fall. In that condition, Thomasius suggests, each person, from birth, was able immediately to apprehend the

very nature of things and actions. After the Fall, experience, including the sciences, has to make up for this loss of insight into the essence of life and the world by piecing together such appearances as humanity can manage. Any success in this endeavour presupposes a systematic critique of all prejudices and this is in effect an implementation of systematic doubt - not ‘sceptical’ Cartesian doubt which is contradictory in denying the possibility of truth and falsehood, but ‘dogmatic’ doubt premised on such a possibility.

The world is, for Thomasius, ultimately spiritual; each thing is a living being with a purpose. But humans stand out as the only beings not entirely determined by their purpose but prone to error through the influence of other aims. He develops a theory of the passions to account for these pervasive aberrations, while his moral theory concerns the moral rules and institutions that allow us to live in pursuit of moral aims that, ultimately, are beyond most of us in this life. The great change in Thomasius is that in the earlier work, passions and morals are vying for control of a free will; but from his 1705 work *Fundamenta juris naturae et gentium ex sensu communi deducta (Foundations of the Law of Nature and of Nations, deduced from Common Sense)* onwards, morality is an epiphenomenon in a deterministic world.

3 Practical philosophy

In practical philosophy, Thomasius’ starting-point was Samuel von Pufendorf and, in a sense, this remained the case across the otherwise dramatic developments of his ideas. The Pufendorfian-Thomasian theory is profoundly voluntarist in the sense that it sees all moral values - in fact, all values - as directly or indirectly dependent upon acts of will. No thing (such as a creature), no state of affairs (such as a relationship between creatures), and no event (such as an act of one creature towards another) has any *inherent* value. Human acts of will are in themselves nothing but natural events, and only assume a moral aspect through their relationship to the law of nature; the law of nature is only a moral law because it is God’s will.

Much intellectual energy was spent during Thomasius’ generation on the obvious problem that this line of argument simply transforms the question of the foundation of morals into the question of the goodness of God, and Thomasius’ early work is first of all a contribution to this debate. His answer was that it is impious to raise the question of God’s goodness, for it assumes that there is enough in common between God and humans to make sense of such a question; but we can know nothing whatsoever by natural means about God - except the bare fact that we are obliged to obey his will. Even if we could know God’s goodness, this would not imply *what* we should do, if it were not for the fact that God, in addition to being good, also would will us to do what his goodness implies. We here meet the common assumption that obligation arises only from a law imposed by a superior. In sum, we know that we are obliged to obey God’s will but also that we do not have a sufficient notion of God from which we can derive what our actual duties are. As a consequence, we have to turn to that which God’s will has left us with, namely our reason. In this way Thomasius effectively excludes theology from any role in the theory of morals, including jurisprudence.

Central to the concept of reason upon which Thomasius relies is the idea of a dialogue or conversation. Reasoning consists of the manipulation of signs, and we learn the use of signs only in dealing with other people. So, without some minimum of social living, there could be no language and no ability to reason; these functions are interdependent. They are also distinctive for humanity, and they are the means by which we alone can live under the guidance of natural law. It is through reason that we can reflect upon the fact that we would not have reason if we were not social, and that we could not be social if others were not reasoning likewise. Thomasius takes these undeniable facts as the best indication we have of what God’s will for us is, namely that we should always act so as to benefit the whole of humanity considered as rational and social. This *utilitas totius humani generis* is again to be understood as *pax* or *vita tranquilla*, peace or the quiet life with others, and this can also be characterized as the essence of temporal happiness, *beatitudo*. This rational sociality (*socialitas* or *Geselligkeit*) is the basic natural law and thus the foundation for all society (*societas* or *Gesellschaft*).

In the 1688 work, the *Institutiones jurisprudentiae divinae*, Thomasius thus identified morality with a broad concept of obligation to natural law or, in other words, justice. He never indicated any room for other areas of morality nor tried to differentiate the concept of natural law to make room for such areas; in fact, he simply followed Pufendorf’s traditional division according to the objects of the duties imposed by natural law, namely God, other people and oneself. However, Thomasius was certainly at work on this problem. Just four years after the *Institutiones*, he published a different account of morals in *Einleitung zur Sitten Lehre*. Morality is now seen as

based on love rather than sociality, but the nature of love is that it cannot be enforced and this is the distinguishing characteristic of morality as opposed to justice and law. The acts that comprise justice - giving each their due, abstaining from injury and so on - may be done out of neighbourly love, but they may also be enforced, in which case they move from love to justice, from morality to law.

In the *Einleitung* Thomasius pays little attention to justice and the foundation of law; his concern is to develop a theory of morality as rational love. This is a topic which we cannot pursue here, just as we have to ignore Thomasius' prolonged despair at making sense of the way in which reason might be said to regulate the love of humanity as the basis for the moral life. He could not see how a free will could escape the storms of passion and make any moral progress merely by rational means; the only way was to lose itself in mystical identification with God.

Thomasius escaped this pietistic replacement of morality by faith through a rejection of the idea of free will. Instead, he regarded the will as determined by the good, and developed a theory of the varieties of goods in life that we may cultivate as the objects of will. This was the basis for his attempt in the late work, especially the *Fundamenta* (1705), to put together a theory that could accommodate both the theory of justice and law, and the theory of morality, and to account for their difference.

The starting point is an identification of the good with peace or quietness of life that can be either internal - peace of mind - or external - security of action. Further, the external peace can either be a purely negative matter of being left in peace, or it can be a matter of having one's welfare actively secured or promoted. This is the basis for Thomasius' famous distinction between *honestum, justum* and *decorum*. Actions that contribute to or are in accordance with the inner peace or balance of a person make up the *honestum*. This is the sphere of morality proper, for such actions are obligatory as part of the moral nature God has prescribed for humanity, but they are only obligatory internally, in conscience, and cannot be enforced - like acts of love, they would lose their specific moral character if enforced. Actions that simply avoid breaking the external peace make up the *justum*; they carry external obligation or obligation to action, and they are enforceable and thus suitable objects of positive legislation. Actions that actively promote the external quiet of life, finally, make up the *decorum*. They are externally obligatory but not enforceable; they thus provide a middle way between the *honestum* and the *justum*, between morality and law, and they are characterized as prudence, or politics.

4 Conclusion

Thomasius' threefold division of morals made available much clearer conceptual and linguistic means than previously for distinguishing between law and morality, and thus between public and private spheres. It also showed just how easily voluntaristic natural law could lose most of its meaning *as* natural law. The gain and the loss went hand in hand as follows: once the basic obligation to morality - or natural law - is reduced to mere obedience to the supreme divine will, all content of this law has to be derived from temporal sources, that is, from human acts of will. But the investigation of human voluntary activity invites empirical methods, and much of Thomasius' progress consisted of his erratic acceptance of this invitation, resulting in a combination of a kind of philosophical anthropology and moral, especially legal, history based upon his three spheres of *honestum*, *justum* and *decorum*. Morals and law are thus increasingly seen as historical or cultural - in the broad sense, conventional - phenomena, and it is not evident that any objectively or universally valid *natural* law is left on the basis of which they could be criticized and changed. At the same time, of course, his linked divisions between law and morality, between public and private, between external and internal obligation, between action and intention, were tools to be taken up and refined by Kant in a way that was far from Thomasius' own troubled Lutheran horizon (see Kant, I. §§9-10).

Of Thomasius' many students and disciples, the most prominent include the theologian Johann Franz Budde (Buddeus), the philosopher Nicolaus Hieronymus Gundling and the legal theorist Johann Peter von Ludewig. He was also influential in the new University of Göttingen, but more in jurisprudence than in philosophy, and especially on J.J. Schmauß and Gottfried Achenwall, whose textbook was used by Kant. But despite their numbers, the Thomasius students never formed a 'school' as did followers of Wolff, partly because the master was difficult to get on with, partly because the Wolff-gang swept the field.

See also: Enlightenment, continental; Natural law

KNUD HAAKONSSEN

List of works

Thomasius' works are not available in English translation.

Thomasius, C. (1995-) *Ausgewählte Werke*, Werner Schneiders, Hildesheim: Olms Verlag.(In progress - when complete, this comprehensive edition will contain all works listed below.)

Thomasius, C. (1688) *Institutiones jurisprudentiæ divinæ libri tres (Institutions of divine jurisprudence in three books)*; repr. 7th edn (1730) Aalen: Scientia Verlag, 1963; German trans., *Drei Bücher der Göttlichen Rechtsgelahrheit*, 1709.(Thomasius' most Pufendorfian work.)

Thomasius, C. (1688) *Introductio ad philosophiam aulicam (Introduction to Court Philosophy)*; German trans., *Einleitung zur Hof-Philosophie*, 1710.(Politico-pedagogical application of the theory of reason.)

Thomasius, C. (1691a) *Einleitung zu der Vernunfft-Lehre (Introduction to the theory of reason)*, repr. Hildesheim: Olms Verlag, 1968. (This and the following are his main epistemological works.)

Thomasius, C. (1691b) *Aushbung der Vernunfft-Lehre (Application of the theory of reason)*, repr. Hildesheim: Olms Verlag, 1968.(Companion volume to the preceding work.)

Thomasius, C. (1692) *Einleitung zu der Sitten-Lehre (Introduction to moral philosophy)*, repr. Hildesheim: Olms Verlag, 1968.(Morality based on love rather than sociality - compare this with the *Institutiones*.)

Thomasius, C. (1696) *Aushbung der Sitten-Lehre (Application of moral theory)*, repr. Hildesheim: Olms Verlag, 1968.(Under the influence of pietism, the sufficiency of reason is in doubt.)

Thomasius, C. (1699) *Summarischer Entwurff derer Grundlehren, die einem Studioso Juris zu wissen… nöthig (Sketch of basic teaching, necessary knowledge for a student of law)*, repr. Aalen: Scientia Verlag, 1979. (Useful jurisprudential compendium.)

Thomasius, C. (1699) *Versuch vom Wesen des Geistes (Essay on the Nature of Spirit)*. (Pietistic mysticism.)

Thomasius, C. (1705) *Fundamenta juris naturae et gentium ex sensu communi deducta (Foundations of the Law of Nature and Nations, deduced from Common Sense)*; repr. 4th edn (1718) Aalen: Scientia Verlag, 1970; German trans., *Grundlehren des Natur- und Völkerrechts*, 1709.(Re-writing of the practical philosophy on a deterministic basis; development of the influential theory of *honestum*, *justum* and *decorum* - morality, law and prudence.)

Thomasius, C. (1719) *Paulo plenior, Historia juris naturalis*, repr. Stuttgart-Bad Cannstatt: Frommann-holsboog, 1972.(Brief history of natural jurisprudence.)

References and further reading

Barnard, F.M. (1965) 'Christian Thomasius: Enlightenment and Bureaucracy', *American Political Science Review* 59: 430-8.(This and the following two works by Barnard are general and accessible studies of Thomasius.)

Barnard, F.M. (1971) 'The "Practical Philosophy" of Christian Thomasius', *Journal of the History of Ideas* 32: 221-46.(General overview of Thomasius' practical philosophy.)

Barnard, F.M. (1983) 'Self-Direction: Thomasius, Kant and Herder', *Political Theory* 11: 343-68.

Barnard, F.M. (1988) 'Fraternity and Citizenship: Two Ethics of Mutuality in Christian Thomasius', *Review of Politics*, 50: 582-602.

Battaglia, F. (1936) *Christiano Thomasio. Filosofo e giurista*, Bologna: Editrice CLUEB, repr. 1982.(The main scholarly study prior to Schneiders' 1971 work.)

Dreitzel, H. (1991) 'Zur Entwicklung und Eigenart der "Eklektischen Philosophie"', *Zeitschrift für Historische Forschung* 18: 281-343.(Thorough study of the eclectic tradition which questions Thomasius' membership of the tradition.)

Haakonssen, K. (1996) *Natural Law and Moral Philosophy: From Grotius to the Scottish Enlightenment*, Cambridge: Cambridge University Press.(Comprehensive study of early modern natural law and its influences.)

Haakonssen, K. (forthcoming) 'Natural Law and the German Tradition', in M. Goldie and R. Wokler (eds) *Cambridge History of Eighteenth-Century Political Thought*, Cambridge: Cambridge University Press. (Thomasius in the context of German natural law theory from Pufendorf to Kant.)

Hammerstein, N. (1972) *Jus und Historie. Ein Beitrag zur Geschichte des historischen Denkens an deutschen Universitäten im späten 17. und im 18. Jahrhundert (Law and history. A contribution to the development of historical thought at German universities in the late seventeenth and eighteenth centuries)*, Göttingen: Vandenhoeck & Ruprecht.(Important, comprehensive study of Thomasius' intellectual and university milieu.)

Hinrichs, C. (1971) *Preußentum und Pietismus. Der Pietismus in Brandenburg-Preußen als religiös-soziale*

Reformbewegung (Prussianism and pietism. Pietism in Brandenburg-Prussia as a socio-religious reform movement), Göttingen: Vandenhoeck & Ruprecht. (Fundamental study of pietism with much attention to Thomasius.)

Krieger, L. (1957) *The German Idea of Freedom. History of a Political Tradition*, Chicago, IL and London: University of Chicago Press, 59-66.(Comprehensive study of Pufendorf and his influence.)

Lieberwirth, R. (1955) *Christian Thomasius. Sein wissenschaftliches Lebenswerk. Eine Bibliographie (Christian Thomasius. Bibliography of a scholarly life)*, Thomasiana. Arbeiten aus dem Institut für Staats- und Rechtsgeschichte bei der Martin-Luther-Universität Halle-Wittenberg. Heft 2. Weimar.(Standard bibliography.)

Rüping, H. (1968) *Die Naturrechtslehre des Christian Thomasius und ihre Fortbildung in der Thomasius-Schule (The natural-law theory of Christian Thomasius and its development in the Thomasian school)*, Bonner Rechtswissenschaftliche Abhandlungen, Bd. 81. Bonn.(Comprehensive study of Thomasius' jurisprudence with emphasis on the later work.)

Schmidt, W. (1995) *Ein vergessener Rebell. Leben und Wirken des Christian Thomasius (A forgotten rebel. The life and work of Christian Thomasius)*, Munich: Diedrichs.(General biography.)

Schneiders, W. (1971) *Naturrecht und Liebesethik. Zur Geschichte der praktischen Philosophie im Hinblick auf Christian Thomasius (Natural law and the love of ethics. Christian Thomasius and the history of practical philosophy)*, Hildesheim and New York: Olms Verlag.(The most important philosophical analysis of Thomasius.)

Schneiders, W. (1983) *Aufklärung und Vorurteilskritik. Studien zur Vorurtheilstheorie*, Stuttgart-Bad Cannstatt: frommann-holzboog.(Work of great scholarship which includes a valuable discussion of Thomasius' theory of prejudice.)

Schneiders, W. (ed.) (1989) *Christian Thomasius, 1655-1728. Interpretationen zu Werk und Wirkung. Mit einer Bibliographie der neueren Thomasius-Literatur (Christian Thomasius, 1655-1728. Interpretations of his work and influence. With a bibliography of modern scholarship)*, Hamburg: Meiner Verlag.(Valuable collection of essays covering most aspects of Thomasius' thought. Excellent bibliography of secondary literature.)

Spaeting, R. (1971) 'On Christian Thomasius and his Alleged Offspring, the German Enlightenment', *Lessing Yearbook* 3: 194-213.(Useful discussion of the connection between Thomasius and the later Enlightenment.)

Thomism

Deriving from Thomas Aquinas in the thirteenth century, Thomism is a body of philosophical and theological ideas that seeks to articulate the intellectual content of Catholic Christianity. In its nineteenth and twentieth-century revivals Thomism has often characterized itself as the 'perennial philosophy'. This description has several aspects: first, the suggestion that there is a set of central and enduring philosophical questions about reality, knowledge and value; second, that Thomism offers an ever-relevant set of answers to these; and third, that these answers constitute an integrated philosophical system.

In its general orientation Thomism is indeed preoccupied with an ancient philosophical agenda and does claim to offer a comprehensive, non-sceptical and realist response based on a synthesis of Greek thought - in particular that of Aristotle - and Judaeo-Christian religious doctrines. However, in their concern to emphasize the continuity of their tradition, Thomists have sometimes overlooked the extent to which it is reinterpretative of its earlier phases. The period from the original writings of Thomas Aquinas to late twentieth-century neo-scholastic and 'analytical' Thomism covers eight centuries and a stretch of intellectual history more varied in its composition than any other comparable period.

Not only have some self-proclaimed Thomists held positions with which Aquinas would probably have taken issue, some have advanced claims that he would not have been able to understand. Examples of the first are found in Neo-Kantian treatments of epistemology and ethics favoured by some twentieth-century Thomists. Examples of the second include attempts to reconcile Aquinas' philosophy of nature with modern physics, and his informal Aristotelian logic with quantified predicate calculus and possible world semantics.

The term 'Thomism' is sometimes used narrowly to refer to the thought of Aquinas, and to its interpretation and elaboration by sixteenth- and seventeenth-century commentators such as Cajetan, Sylvester of Ferrara, Domingo Bañez and John of St Thomas. At other times it is employed in connection with any view that takes its central ideas from Aquinas but which may depart from other of his doctrines, or which combines his ideas with those of other philosophers and philosophies. Prominent examples of Thomists in this wider sense include Francisco Suárez (1548- 1617) who also drew on the epistemology and metaphysics of another great medieval thinker Duns Scotus; and, more recently, Joseph Marechal (1878-1944) whose 'Transcendental Thomism' accepted as its starting point the Kantian assumption that experience is of phenomena and not of reality as it is in itself. An example drawn from the ranks of contemporary analytical philosophers is Peter Geach who draws in equal measure from Aquinas, Frege and Wittgenstein.

In the twentieth century there have been two major proponents of the philosophy of Aquinas, namely Jacques Maritain and Etienne Gilson, both of whom contributed significantly to the development of Neo-Thomism in North America. Interestingly, both men were French, neither had been trained in a Thomistic tradition and both were drawn into philosophy by attending lectures by Henri Bergson at the Collège de France in Paris. The Neo-Thomism they inspired declined following the Second Vatican Council (1962-5) as Catholics looked to other philosophical movements, including existentialism and phenomenology, or away from philosophy altogether. Today Thomists tend to be close followers and interpreters of the writings of Aquinas, but there is also a growing interest among mainstream English-language philosophers in some of his central ideas. While not a movement, this approach has been described as 'analytical' Thomism.

1 Aquinas and the first Thomism: 13th-15th centuries

Aquinas was born into a religious culture in which the dominant style of thought was a form of Neoplatonist Catholic theology. The main source of this was Augustine of Hippo, mediated via later Latin thinkers such as Anselm. Early in his life, however, Aquinas fell under the influence of the recently founded Dominican Order of which he became a member. Under the direction of Albert the Great, the foremost of the Order's early masters, he developed an intense interest in the more naturalistic philosophy of Aristotle whom he came to refer to as 'the philosopher'. Works of Aristotle were being translated into Latin for the first time during Aquinas' life, having been rediscovered through contact with the Arab world where they had been preserved.

To Albert and Aquinas, Aristotle offered a more promising resource for the articulation of Christian doctrine than

did the Augustinian Platonism current in the cathedral schools and universities. However this new synthesis met with considerable opposition since it seemed to be at odds with orthodoxy. In fact, Aquinas found himself in dispute with two groups. To one side were the Augustinians represented by the secular teachers and the Franciscans; to the other were extreme Aristotelian naturalists who held doctrines that are indeed difficult to reconcile with Christian orthodoxy. Aquinas sought to tread a middle path, directing writings against each group in turn: *De aeternitate mundi contra murmurantes* (On the eternity of the world) (1271) against Bonaventure and other Augustinian Franciscans, and *De unitate intellectus* (On the Unity of the Intellect) (1270) against Siger of Brabant and other Latin Averroists (named after Averroes (Ibn Rushd) the Arab philosopher and interpreter of Aristotle - see Averroism).

Although Aquinas' Christian Aristoteliansm was later to be judged the 'most perfect' reconciliation of philosophy and faith, and was made, in effect, the official system of thought of Roman Catholicism, it was initially attacked and subjected to ecclesiastical denunciation. In 1270 Bishop Tempier of Paris condemned a number of propositions associated with Aristotelianism. None directly attributable to Albert or Aquinas was included; but in 1277, three years after the death of Aquinas, Tempier issued a further condemnation and this time Thomistic claims (about the soul and the total nonmateriality of spiritual substances) were specified, although Aquinas was not named. In the same year Robert Kilwardby, Bishop of Oxford (and, like Aquinas, a Dominican) issued a similar condemnation; the following month the Pope endorsed Tempier's decree. Two years later William de la Mare, a Franciscan, produced a work 'correcting' the error of Aquinas' ways (*Correctorium fratris Thomae*).

The general Dominican response was to defend their master against these attacks from within and without. In 1278 the Dominican general chapter appointed a committee to investigate English Dominican disloyalty, and in the meantime set about promoting the cause of Thomas as a thinker and as a saint. In 1282 William of Macclesfield responded to de la Mare countering the charges with his *Correctorium corruptorii 'Quaestione'*. Around this time John of Paris responded similarly; by 1286 study of Aquinas was made compulsory by the Parisian Dominicans and this was repeated elsewhere: in Saragossa (1309) in London (1314) and in Bologna (1315). Defence gave way to counter-attack and on 18 July 1323, within fifty years of his death, Aquinas was declared a saint by Pope John XXII. Two years later Bishop Bourret of Paris revoked Tempier's condemnation.

Ecclesiastical approval removed one obstacle to acceptance of Thomistic thought, and by stages his ideas spread and gained influence. Apart from the intrinsic merit of those ideas, an important factor in this development was the increasing number of colleges and universities. Each approved place of study (*studium generale*) had houses belonging to the main teaching orders, and by this means, through the multiplication of copies of Thomistic texts and by translation into other languages (German and Greek), the Dominicans ensured that their master's voice could be heard throughout Europe.

Early in the fifteenth century Aquinas found a powerful follower in John Capreolus. In the late scholastic period, the ideas of Aquinas had to compete with those of two other medieval figures, namely Duns Scotus and William of Ockham. Capreolus challenged various views of Scotus and others to such good effect that he earned the title 'foremost Thomist' (*princeps Thomistarum*). More common than dialectical defences, however, were informed commentaries on the works of Aquinas, in particular the *Summa theologiae* (1266-73). These were important in transmitting Thomist doctrines, yet in themselves they did little to combat the rising tide of Ockhamist nominalism. Also, and somewhat unjustly, they were in part associated with the formalistic scholasticism against which the Renaissance humanists rebelled.

2 The Second Thomism: 16th-18th centuries

The sixteenth century was perhaps the most troubled in the history of post-medieval Christendom. The Reformation divided Europe into Catholic and Protestant states. It would be natural - but incorrect - to suppose that Aquinas was read only by Catholics. In England, the Anglican Richard Hooker, and in Holland the Calvinist Hugo Grotius were both influenced by studying his work. However, it was within Catholic countries and regions that the next phase of Thomism developed. In particular, Spain and Italy gave rise to new theologically and metaphysically oriented presentations of Thomas' thought. In England the Reformation was more a matter of politics than theology; elsewhere (including Scotland) the reformers drove their axe to the roots of Catholic belief. In response, the Church of Rome set about renewing its intellectual resources. Thus was born the Counter-Reformation.

The Council of Trent (1545-63) sought to systematize Catholic doctrine and led to the production of a definitive Catechism in 1566 (*Catechismus romanus*) on which the thought of Aquinas had a major influence. (Contrary to an often-repeated tale, however, the *Summa theologiae* was not placed on the altar alongside the Bible during meetings of the Council.) Trent encouraged the study of philosophy and theology in all Catholic colleges, seminaries and universities. This created a need for appropriate textbooks and that was met with a new style of manual setting out Thomistic thought. A further response to the reformers was the development of new religious orders. The most famous of these was the 'Jesuits', the Society of Jesus founded by Ignatius of Loyola. Ignatius explicitly encouraged the study of Aquinas and Aristotle, and expressed the hope that interpretations of their ideas adapted to the needs of the time would be forthcoming.

In saying this, Ignatius was drawing upon an earlier pre-Tridentine tradition of teaching and commenting on Aquinas. The major figures associated with this tradition were Dominicans. In the first decade of the sixteenth century Peter Crockaert, a Belgian working in Paris, had substituted the *Summa theologiae* for the *Sentences* of Peter Lombard which had previously been the standard text for theological instruction. Likewise in Italy Thomas de Vio Cajetan was lecturing on the *Summa* and producing a major commentary which was later to be published in an edition of the complete works of Aquinas. In Spain Francisco de Vitoria also used the *Summa* as a basis of theological education and he was followed in this by his disciple Domingo de Soto.

Early Jesuit Thomists included students of de Soto. But the full Jesuit appropriation of Aquinas came later with Luis de Molina and, most notably of all, with Francisco Suárez. The latter's great work was the 1597 *Disputationes Metaphysicae (Metaphysical Disputations)*. Starting from a recognized need to produce theology adequate to meet that of the reformers, Suárez came to the conclusion that it was not appropriate simply to invoke the philosophy of Aristotle; rather fundamental issues needed to be addressed afresh. The result was a mix of Thomistic and non-Thomistic (largely Scotist) metaphysics. In fact, Suárez anticipates much of the thinking about essence, existence, identity and modality that emerged as a result of the equally focused attention in a century of analytical metaphysicians.

The Dominicans meanwhile had stayed closer to the detail of Aquinas' philosophy, in part because they were still operating with the idea that their first duty was one of loyalty to a brother who had long been misrepresented and maligned. The need to evangelize and to educate added to the tradition of exact commentary a series of course-book texts, the most famous of which (still in use in Catholic institutions into the twentieth century) were the philosophical and theological courses (*Cursus philosophicus* and *Cursus theologicus*) of Jean Poinsot, better known as John of St Thomas.

In 1568 Aquinas was named a 'Doctor of the Church'. Taken from the Latin for teacher (*docens*), the first official use of the title occurred in the thirteenth century when eight saints (including Augustine) were so honoured. It is a measure of the Church's estimate that Aquinas was named its ninth *docens*, known honorifically as the 'Angelic Doctor'.

The sixteenth century ended with two main schools of Thomism. The first had its strongest base in Italy and was associated with Dominican exegetical interpretation. The second was rooted in Spain and centred around the Jesuit appropriation of Aquinas. The first was more narrowly Thomist, being closer to the historical doctrines of St Thomas, and found its natural expression in the form of close textual commentaries. The second was more broadly Thomistic and gave rise to treatises on particular philosophical themes such as existence and essence.

Inevitably there was competition between these traditions; but the most heated conflict was not methodological or interpretative but doctrinal. This occurred from around 1590 to 1610, and continued intermittently thereafter. The subject was the nature of grace, free-will and divine foreknowledge. On the one side, the Jesuit Molina argued in his 1588 *Liberi arbitrii cum gratiae donis (A Reconciliation of Free Choice with the Gifts of Grace, Divine Foreknowledge, Providence, Predestination and Reprobation)* that God's total omniscience is compatible with human liberty, because the former includes 'middle knowledge' (*scientia media*) by which God knows what each person would freely do in every possible circumstance of choice, and distributes grace accordingly. In opposition to what is certainly not a position advanced or accepted by Aquinas, the Dominicans, of whom the most prominent was Dominic Bañez, contended that God knows who will be saved and who will be damned because he has distributed fully effective grace to some but not all. The Jesuits accused the Dominicans of embracing Calvinist

predestinarianism, while the Dominicans charged the Jesuits with Pelagianism (see Pelagianism; Predestination).

This 'heresy calling' led to papal attempts to tame the debate, if not to end it, although without much success. In the meantime the intellectual energies of Thomists had been largely distracted from the important task of developing the general system so as to take account of the rise of modern science and the new philosophies of rationalism and empiricism. The trial of Galileo and the replies to Descartes show the Thomists to have fallen behind the times. Indeed, so far as the intellectual aspect is concerned, it was their ill-preparedness to engage modern thought, rather than intrinsic weaknesses within Thomism, that led to the marginalization of the tradition in the seventeenth and eighteenth centuries.

The fate of Catholic Thomism in the eighteenth century was also a reflection of the general circumstance of the Church. As in the Reformation, the period featured many social and political disruptions and much ecclesiastical infighting. In 1772 the Jesuits were suppressed on the order of the Pope and, in the next decade, Catholicism itself was battered by the French Revolution and by the rise of secularism. The new political thinking was antitheocratic, anticlerical, broadly democratic and at best deistic, although often atheistic (see Deism). It is hardly surprising, therefore, that a system of thought born out of medieval Catholicism did not flourish in these circumstances. The Dominicans continued to produce critical editions of Aquinas' writings but it is doubtful whether they were read outside the Dominican order. Even in Rome ecclesiastics had lost interest in Thomism.

3 Neo-Thomism: 19th and 20th centuries

As in the past, however, a process of revitalisation led in due course to a renaissance. Following the period of the French Revolution, Catholic thought in France, Belgium and Italy tended to be divided between two movements: a quasi-fideistic movement that emphasized the centrality of faith and sought to deal with the threat from rationalism by side-stepping it; and a philosophical approach which maintained, along lines first suggested by Christian Neoplatonists, and later developed by Cartesians such as Malebranche, that the intellect directly intuits God in all its acts of knowledge. These two approaches came to be known as 'traditionalism' and 'ontologism'; their main proponents being Lammenais (1782-1854) and de Maistre (1752-1821), and Gioberti (1801-52) and Rosmini-Serbati, respectively.

Elsewhere in Italy (particularly in Naples) and in Spain, the Dominicans maintained their loyalty to Aquinas. The Italian Dominican Tommaso Zigliara (1833-93) found favour with Bishop Pecci of Perugia (later to become Pope Leo III), and in 1873 was appointed Regent of Studies in the Dominican College in Rome where he was joined by Alberto Lepidi (1838-1922). Both men were critical of the traditionalists and the ontologists; and through their writings, teachings and administration they did much to encourage members of their order in Italy and France to develop Neo-Thomistic responses to these movements, as well as to the empiricism and rationalism to which they had been reactions. The revival of Thomism was much encouraged by the papacy. In 1846 Pope Pius IX argued that reason and faith are compatible and that lapses into fideism and intellectual intuitionism are both to be avoided. Versions of traditionalism and ontologism were condemned in 1855 and 1866, respectively, and a return to the scholastic approaches was openly favoured.

An important figure in this revival was Joseph Kleutgen (1811-83) a German Jesuit. In a five-volume work entitled *Die Theologie der Vorzeit*, Kleutgen identified the weaknesses in Catholic intellectual responses to modern thought. He argued that only Aristotelian metaphysics could provide a sure foundation for Catholic theology. In a second work, *Die Philosophie der Vorzeit*, he expounded his own version of neo-Aristotelianism, rejecting the epistemological method of Descartes in favour of cognitive realism, attacking Cartesian mind-body dualism. Like Aquinas, Kleutgen and his colleagues, such as Matteo Liberatore (1810-92), affirmed the unity of the human person as a psychophysical substance. Also while upholding the epistemological primacy of experience they maintained the possibility of establishing, by abstract reflection, various necessary truths about reality - principally that it is the creation of God. Although the thought of Aquinas featured here, this movement was more generally a revival of scholasticism rather than of Thomism as such. In fact neither Kleutgen nor Liberatore were Thomists in the narrow sense.

In 1878 Gioacchino Pecci was elected pope and was crowned Leo XIII. He had long supported the revivalist movement, and the following year he published the famous encyclical *Aeterni Patris* in which Aquinas is commended as providing the surest intellectual foundation for, and articulation of Catholic doctrine. Kleutgen is

reputed to have contributed to the draft of the encyclical and certainly his scholastic stance was vindicated by it. Leo also appointed neo-scholastics to important posts in Rome. Once again, then, through the edict of a pope, Thomism became the orthodox system of thought for Roman Catholicism.

Neo-Thomism looked in two directions. On the one hand it was called into being to address contemporary philosophical issues; on the other hand it drew its inspiration from the distant past. These two orientations gave rise to two strands, one ‘problematic’ the other historical. An unquestionable beneficiary of the revival was the history of medieval philosophy. At the University of Louvain in Belgium, where once ontologism had been enthroned, a new school of scholastic scholarship developed to which important contributions were made by Cardinal Mercier (1851-1926), Maurice de Wulf (1867-1947) and Martin Grabmann (1875-1959).

The problematic strand was first developed in response to the challenges of empiricism and Kantian idealism. Traditional Thomism assumed that the mind was in direct engagement with reality through experience. According to Aquinas both perception and intellection involve the reception of the forms of external substances. After Descartes, Locke, Hume and Kant, however, this view seemed difficult to maintain - indeed some found it difficult even to make sense of. The new orthodoxy was that the starting point of all philosophy is consciousness and the appearances it offers. Somehow from this we need to argue to the existence of something independent - external reality. Like others awed by Kantianism, several Neo-Thomists maintained that no philosophy could be credible that did not accept the new starting point of immanent consciousness. Influenced by Maurice Blondel (1861-1947) and Pierre Rousselot (1878-1915), the philosopher Joseph Marechal (1878-1944), and the theologians Karl Rahner and Bernard Lonergan - all of whom, like Rousselot, were Jesuits - tried to show that it was possible to combine Kant’s critical philosophy with the transcendental realism and theism of Aquinas. The result, known as ‘Transcendental Thomism’, though widespread in its influence among theologians, was never taken very seriously by philosophers.

Another example of the attempt to synthesize Thomism with a modern philosophy is represented by the ‘Lublin school’ founded with the establishment (in 1946) of the Catholic University of Lublin in Poland, the only independent university during the period of the Soviet bloc. Here the sources were several: Thomism was represented by French interpreters; realist phenomenology was advocated by Roman Ingarden who had been a student of Husserl; and logic and philosophy of science in the technical style was favoured by the Lwów-Warsaw school (see Poland, philosophy in §3). The leading figures were Kamiński and Krąpiec. The most famous member of the school, however, was Karol Wojtyla (1920-) who drew on the value theory of Max Scheler, Husserlian phenomenology and the anthropology of Aquinas, to devise a form of Thomist personalism articulated in his work *The Acting Person*. Better known as John Paul II, Wojtyla was not the first, nor probably the last pope to favour Thomism. His time at Lublin was short, however, being promoted to the see of Cracow in 1958, and elected Pope in 1978. Other members of the movement moved on, were diverted into administration, or died (Kamiński in 1986). The Lublin school is now moribund (see Poland, philosophy in §4).

Two of the most important twentieth-century Neo-Thomists were French laymen, Jacques Maritain and Etienne Gilson. Both were critics of transcendental Thomism and both had enormous influence in North America as well as in Europe. Maritain was raised in a comfortable, politically liberal, Protestant family. Despairing of the materialism and secularism characteristic of Paris university and intellectual life, he and his wife made a suicide pact but revoked this after attending lectures by Henri Bergson. While Bergson’s ‘vitalist’ philosophy lifted their despair, it was not until they converted to Roman Catholicism in 1906, and thereafter discovered the philosophy of Aquinas, that the Maritains felt they had found a wholly adequate worldview combining humanism with transcendence.

Maritain lectured and published very widely in almost all areas of philosophy, and was a dominant influence in Catholic thought from the post-First World War period to the Second Vatican Council (1962-5). He approached the thought of Aquinas somewhat ahistorically rather than as medieval revivalist, and derived from it a realist metaphysics, epistemology and value theory. In his most important work of speculative philosophy, *The Degrees of Knowledge* (1959), he argues, following Aquinas and John of St Thomas, that concepts or ideas are principles of thought but not (save in reflection) the objects of cognition. Likewise, he insists that the natural order has an objective metaphysical structure of essential kinds and that these are the proper concern of science. His theory of value is likewise keyed to external realities, but value is seen as directed towards participation in the life of God.

In his social and political philosophy Maritain emphasized the irreducibility of community and of the common good. These notions featured prominently in Catholic social teaching throughout the twentieth century and Maritain is often looked to as a source of inspiration by followers of this tradition (see Maritain, J.).

Like Maritain, Gilson was taught by Bergson, but his own interests lay principally in the history of ideas, and in particular in the relationship between modern philosophy as represented by Descartes and pre-modern scholasticism. From 1921 he held several Parisian posts in medieval philosophy at the Sorbonne, the École Pratique des Hautes Études, and at the Collège de France. His influence on the Neo-Thomist revival was as much through his teaching and academic leadership as through his writings. He lectured in North America, and co-founded the Pontifical Institute of Medieval Studies in Toronto. This has been one of the major centres of medieval scholarship and has helped to shape the interpretation of Aquinas by insisting on the need to understand his ideas in their historical context.

Gilson deplored the subjective turn introduced by the modern doctrines of mental images and ideas. However, he believed that any attempt to invoke medieval thinking in opposition to modern philosophy must be mindful of the variety of views held during the Middle Ages and, even more importantly, of the fact that they were developed primarily in theological contexts. Whereas Maritain presented Thomism as if it were a set of timeless abstract ideas, Gilson distinguished between the teachings of Aquinas and those of late commentators who sometimes imported their own views or who sought to synthesize Thomism with approaches current in their own day. Similarly he argued that while Aquinas drew heavily on the work of Aristotle he often used Aristotelian notions for different purposes, generally to defend Christian theology, and added ideas of his own; the most important of these being the claim that God is necessary existence and the source of the existence of contingent being.

Gilson's contextualist approach has been most widely followed among the Catholic historians of the medieval period, but its indirect influence is also apparent in the philosophical account of reason as immanent within traditions of enquiry advanced by Alasdair MacIntyre. A convert, like so many other English-speaking Catholic philosophers (including Anscombe, Dummett and Geach), MacIntyre's understanding and use of Aquinas has been shaped not by a Thomistic education but by personal study and in response to the views of others. For some years he taught at the University of Notre Dame, Indiana, where the Medieval Institute and the Jacques Maritain Center each engage scholastic thought. The second of these has been directed by Ralph McInerny, one of the leading representatives of Neo-Thomism in North America. McInerny has published extensively on Aquinas, increasingly in a popularizing vein. He was also for many years editor of the *New Scholasticism* (now the *American Catholic Philosophical Quarterly*). Together with the *Modern Schoolman*, the *Thomist* and to a lesser extent the *Review of Metaphysics* (edited by another senior Neo-Thomist, Jude Dougherty of the Catholic University of America), this journal and its successor have been the main fora for the presentation of essays in neo-scholastic historical and problematic traditions.

4 Aquinas and Thomism: future prospects

Contemporary historical scholarship in Thomistic philosophy is of a high standard. In the Anglo-Saxon world this is a consequence of the particular efforts of Gilson and his followers in Toronto, Notre Dame and elsewhere. Aquinas and other medieval and scholastic figures have also benefited from a general rise of interest in the history of philosophy, and by no means are all who now study Christian medieval thought themselves Catholics or even theists. At the same time it is natural that those who avow a Christian worldview should look with intense interest at the work of the individual who is beyond question the greatest Christian philosopher-theologian.

Future trends are difficult to predict in any detail, but if there is a future, and if it resembles the past even in broad outline, then the tide of interest in the thought of Aquinas will rise and fall as before. To some extent this will reflect the intellectual condition of the Roman Catholic Church and that of the colleges, seminaries and universities established to serve it. As was noted, however, Aquinas and Thomism are not the preserve of Catholics only. Indeed, there is a growing interest among philosophers trained wholly or partly in analytical philosophy. Some younger writers, drawing on the example of the British philosophers Elizabeth Anscombe, Peter Geach and Anthony Kenny, are using Thomistic resources to deal with contemporary philosophical problems. Others in the English-speaking world have turned to Aquinas as an important figure in the history of philosophy to be studied as one might any other thinker from the past. This second (mainly US) group is represented principally by scholars trained at Cornell under Norman Kretzmann and at Notre Dame under Ralph McInerny. As throughout previous

centuries, interest in Aquinas will continue along two broad and connecting paths: one followed by those who wish to give an accurate representation of his thought; the other taken by those who wish to mine it as a source of interesting ideas.

See also: Garrigou-Lagrange, R.; Religion, philosophy of

JOHN J. HALDANE

References and further reading

Works cited in the text not listed here can be found in the respective biographical entries of the particular author.The following English-language journals continue to publish work in the broadly Thomistic tradition: the *American Catholic Philosophical Quarterly* edited from the University of Dallas, TX (formerly the *New Scholasticism*, edited from the University of Notre Dame); the *Modern Schoolman* (Quarterly) edited from the University of St Louis, MO; the *Review of Metaphysics* (Quarterly) edited from the Catholic University of America in Washington, DC; and the *Thomist* (Quarterly) edited from the Dominican house of studies in Washington, DC. Of these, the *ACPQ* and the *Review of Metaphysics* are the most pluralistic, and the *Thomist* is the most traditional.

Bourke, V.J. (1945) 'Thomistic Bibliography 1920-1940', *The Modern Schoolman*, supplementary vol. 21.(The first English-language bibliography of modern neo-Thomism.)

Brezik, V.B. (1981) *One Hundred Years of Thomism*, Houston, TX: Center for Thomistic Studies.(Collection of essays on the past, present and future of modern Thomism with the text of *Aeterni Patris* 'Encyclical Letter on the Restoration of Christian Philosophy' by Pope Leo XIII.)

Chesterton, G.K. (1933) *St Thomas Aquinas*, London: Hodder & Stoughton. (Although unscholarly and somewhat idiosyncratic, this shows a strong intuitive grasp of the main realist thrust of Thomism.)

Dougherty, J. (1995) 'Analysis in Search of Wisdom', ed. T. Druart, *Reason in History, Proceedings of the American Catholic Philosophical Association* 68.(Reflections by a senior contemporary neo-Thomist on the current condition of philosophy, delivered on the occasion of his receiving the Aquinas Medal from the ACPA in 1994.)

Fitzpatrick, P.J. (1982) 'Neoscholasticism', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(The best single-volume reference work on the philosophy of the high-middle ages. It contains over forty essays by as many authors on a wide range of subjects. Also included are biographies of medieval figures, an extensive bibliography, and name and subject indexes.)

Gilson, E. (1936) *The Spirit of Medieval Philosophy*, New York: Scribner.(The text of Gilson's 1931-2 Gifford Lectures, in which he argues for the importance of seeing medieval thinkers as religious, while also recognizing the very broad range of views held by them.)

Gilson, E. (1937) *The Unity of Philosophical Experience*, New York: Scribner.(A short study, much admired within and beyond neo-Thomist circles for its inspirational view of the nature of philosophical enquiry.)

Gilson, E. (1952) *Being and Some Philosophers*, Toronto, Ont.: Pontifical Institute of Medieval Studies.(An advanced work in which Gilson explores the ways in which the notion of being features in the philosophy of Aquinas' predecessors and argues for the originality of Aquinas' own contribution.)

Gilson, E. (1964) *The Spirit of Thomism*, New York: P.J. Kennedy & Sons.(A short account of the nature and condition of neo-Thomism in relation to the religious philosophy of Aquinas. Useful on the question of how basic Thomist principles should be applied in challenging modern philosophies.)

Haldane, J. (1994) 'MacIntyre's Thomist Revival: What Next?' in J. Horton and S. Mendus (eds) *After MacIntyre*, Cambridge: Polity Press.(A critical examination of MacIntyre's claims concerning Thomism - as developed in MacIntyre's *Whose Justice? Which Rationality?* and *Three Rival Versions of Moral Inquiry*.)

Haldane, J. (1997) 'Analytical Thomism', *The Monist* 80 (4).(Collection of new articles by philosophers in the analytical tradition exploring aspects of Aquinas' thought as it bears on central philosophical problems.)

Haldane, J. (1998) 'What Future has Catholic Philosophy?', ed. T. Druart, *Virtues and Virtue Theories, Proceedings of the American Catholic Philosophical Association* 71. (An examination of the current state of philosophy as practised by Catholic thinkers. It argues that neo-Thomism has suffered from isolation from the mainstream analytical tradition of English-language philosophy.)

Hudson, D.W. and Moran, D.W. (1992) *The Future of Thomism*, Mishawaka, IN: American Maritain Association.(Collection of essays presenting a wide range of sometimes conflicting interpretations of the meaning of Thomism and its place within contemporary philosophy.)

Ingardia, R. (1993) *Thomas Aquinas: International Bibliography 1977-1990*, Bowling Green, OH: Philosophy

Documentation Center.(The most up-to-date guide to publications on Aquinas and Thomism in English, French, German, Italian and Spanish.)

Kennedy, L.A. (1987) *A Catalogue of Thomists, 1270-1900*, Notre Dame, IN: University of Notre Dame Press.(A scholarly listing without commentary of 2034 Thomists: from Richard Knapwell (of England) in the thirteenth century to Walsch (of unknown country) at the end of the nineteenth.)

Kleutgen, J. (1850-3) *Die Philosophie der Vorzeit Verteidigt (A Defence of the Philosophy of the Past)*, Munich: Theissings'sche Buchhandlung, 5 vols.(Influential work in the nineteenth-century revival of Thomism which argues that Western philosophy suffered a radical break in the modern period and thus can be divided into two phases: from Socrates to the medievals and the late scholastics, and from Descartes onwards.)

Knasas, J.F.X. (1994) *Thomistic Papers VI*, Houston, TX: Center for Thomistic Studies.(Set of essays by many of the same authors as in Brezik's *One Hundred Years of Thomism*, discussing the claims of McCool's *From Unity to Pluralism*.)

MacIntyre, A. (1988) *Whose Justice? Which Rationality?*, London: Duckworth.(An important statement of MacIntyre's view that all reason is immanent within traditions of inquiry.)

MacIntyre, A. (1990) *Three Rival Versions of Moral Inquiry*, London: Duckworth.(A revised version of his 1988 Gifford Lectures, this is a further development of MacIntyre's account of reason, this time applied in relation to three contrasting perspectives, one of which - Thomism - is judged to be superior.)

Maritain, J. (1930) *Art and Scholasticism*, London: Sheed & Ward.(An influential and once widely read account of the nature of art and beauty conceived of from a Thomistic perspective.)

Maritain, J. (1948) *A Preface to Metaphysics*, New York: Sheed & Ward.(A useful starting point for Maritain's Thomism, this short book consists of seven lectures and is something of a prospectus for the metaphysics he favoured.)

Maritain, J. (1959) *The Degrees of Knowledge*, New York: Charles Scribner.(The definitive presentation of Maritain's epistemology in which he draws upon the interpretation of John of St Thomas to argue for realism. An important work, though hard going for the uninitiated.)

McCool, G. (1989) *Nineteenth Century Scholasticism: The Search for a Unitary Method*, New York: Fordham University Press.(An examination of the development of Catholic theology and the impact on this of the revival of Thomism.)

McCool, G. (1992) *From Unity to Pluralism: The Internal Evolution of Thomism*, New York: Fordham University Press.(A study of the debate between traditional and largely Jesuit neo-Kantian Thomists over the question of whether there can be a diversity of philosophical approaches.)

McCool, G. (1994) *The Neo-Thomists*, Milwaukee, WI: Marquette University Press.(The best available survey of Thomism, concentrating mainly on the nineteenth and twentieth centuries).

McInerny, R. (1968) *Thomism in an Age of Renewal*, Notre Dame, IN: University of Notre Dame Press.(A study of central features of Thomistic thought in relation to developments in Catholic theology following the Second Vatican Council.)

McInerny, R. (1982) *St Thomas Aquinas*, Notre Dame, IN: University of Notre Dame Press. (A short but clear and systematic account of Aquinas, indicating his reliance on the thought of his predecessors, in particular Aristotle and Boethius.)

Miethe, T.L. and Bourke, V.J. (1981) *Thomistic Bibliography 1940-1978*, London: Greenwood Press.(A further useful bibliography, this spanning the gap between the periods covered by Bourke and Ingardia.)

Phillips, R.P. (1934) *Modern Thomistic Philosophy, Volumes I and II*, London: Burns Oates & Washbourne. (Example of the seminary-manual style of systematic presentation of Thomistic philosophy - in this case philosophy of nature and metaphysics).

Roensch, F.J. (1964) *The Early Thomistic School*, Dubuque, IA: Priory Press.(A scholarly study of the range of issues and opinions disputed among medieval and neo-scholarly Thomists.)

Trentman, J.A. (1982) 'Scholasticism in the Seventeenth Century', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(Good survey - see Fitzpatrick (1982) above.)

William de la Mare (1279) *Correctorium Corruptorii*, in P. Glorieux (ed) *Les premières polemiques thomistes (The First Thomistic Polemics)*, Kain: Bibliothèque thomiste, 1927.(A scholarly collection of various writings by de la Mare advancing Augustinian positions in opposition to neo-Aristotelian ones.)

Wojtyla, K. (Pope John Paul II) (1979) *The Acting Person*, trans. A. Potocki, Dordrecht: Reidel.(An advanced presentation of Lublin-style, Thomistic philosophical anthropology. Drawing on Aquinas and phenomenology, Wojtyla emphasizes the unity of the agent.)

Thoreau, Henry David (1817-62)

Thoreau was one of the founders of the new literature that emerged within the fledgling culture of the United States in the middle decades of the nineteenth century. He inherited an education in the classics and in the transcendentalism of his older friend and teacher Ralph Waldo Emerson. Thoreau forged a means of writing which was dedicated to recording particular events in all their transience but capable of rendering graphic the permanent laws of nature and conscience. His incorporation of both confidence and self-questioning into the texture of his writing forms the ground of his standpoint as an observer of human lives and other natural histories.

Thoreau's relation to philosophy goes beyond his inheritance from Plato, Kant, Emerson and Eastern thought. Above all, his quest for philosophy is evident in the ways his writing seeks its own foundations. It is in the act of writing that Thoreau locates the perspectives within which to give an account of the humanness of a life. His project is to report sincerely and unselfconsciously a life of passion and simplicity, using himself as a representative of basic human needs and projects. Influenced by Plato's Republic, *Thoreau gives an account of some basic human needs, such as food, shelter and society. But also, like Plato, he shows that the particular institutions by which human needs are met are very far from being necessary. Tracing the relationship between need and necessity is one of the primary goals of Thoreau's work.*

1 Thoreau's vocation

During two years spent at Walden Pond, a small lake near his birthplace of Concord, Massachusetts, Thoreau began the work that culminated in the book called *Walden* (1854). In that same period of time, he spent the night in jail, and this experience is the point of departure for his essay 'Resistance to Civil Government' (1849), better known under its later title 'Civil Disobedience'. Despite the efforts of scholars like Perry Miller to promote the accomplishment of Thoreau's fourteen-volume *Journal* (1906), it is for *Walden* and 'Civil Disobedience' that Thoreau is best known. As literature, *Walden* stands with the work of Melville, Dickinson and Whitman, and is one of the founding documents of the transcendentalist tradition in American philosophy.

Thoreau possessed unrivalled powers of observation, the desire for a life of independence, and a vocation as a writer that dominated his everyday existence. His powers of observation were devoted to providing an account of the natural and human events that occurred around him. He sought both to find the order in these events and to disrupt the received categories of perception, understanding and narration. Long before twentieth-century physics, he was questioning the standpoint of the observer of nature, as well as the standpoint of the observer of human affairs. He sought the laws by which we could calculate the path of the apparently irregular occurrences of observers, along with the regularities in their observations.

Thoreau's requirement that we seek solitude and independence was not a recommendation of total isolation. He focused on solitude as a condition in which human beings learn how to hear what there is to hear. He thought that the modern conditions of cities and societies place us generally in a false proximity to other human beings. This falseness makes it difficult to take any steps towards a more genuine closeness or communication. Withdrawing to a place like *Walden* is not the same as acquiring freedom from the false sense of companionship and society. Taking that step, however, is meant to be an acknowledgement of the difficulties involved in achieving the proper distance from human affairs. Physical distance and isolation were symbols for the kind of mental perspective he sought.

Thoreau's work shows the way he absorbed and re-worked his reading of classical philosophers, especially Plato. He was one of the first thinkers in the West to avail himself of Eastern works: the *Bhagavad Gītā* and the *Laws of Manu* figure pervasively in *Walden* (see Duty and virtue, Indian conceptions of). Thoreau does not distinguish greatly between Eastern and Western traditions; rather he is concerned to distinguish between 'easy reading' and what he calls the 'heroic books' (Thoreau [1854] 1963 ch. III: 75). The latter require our best efforts of reading, while the former suits the cheapness of contemporary literature and serves to dissipate its readers' faculties and to confirm their shallowness.

What Thoreau characterizes as 'heroic' is connected to what he sees as the need for an epistemology of human possibility. He offers us an investigation of the relation between the literary depiction of mental liberation and the

natural conditions of practical freedom. The ways in which we read, as much as the newspapers and novels that constitute the bulk of our reading, are indications of our habits of attention in every sphere of human practice. Thoreau finds the dominant characteristic of these various enterprises to be a self-dissipation that undermines our ability to concentrate our powers. Reading in a 'high sense' is a primary path on which we may recover our concentration and discernment. Reading thus becomes a way to discover the conditions under which we work and act.

2 Emerson and transcendentalism

Early in his life, Thoreau experienced the power of the work and capacity for friendship of Ralph Waldo Emerson. Paradoxically, it was Emerson's entry into life that enabled Thoreau to discover his originality. Thoreau's reputation as a thinker has suffered doubly from this situation: he is almost invariably regarded not only as a transcendentalist, but as one whose basic ideas were derived from Emerson. And Emerson's own writing is consistently characterized as lacking the rigour and methodical development of philosophical thought.

The transcendentalist movement took its name from Kant's notion of the transcendental employment of ideas. Rather than seeking to go beyond human experience, transcendental knowledge seeks the conditions under which our current human knowledge and the experience of ordinary objects is possible. Transcendentalists interpreted this Kantian problematic as requiring a reciprocal relation between the quality of daily experience and the possibility of achieving a knowledge of the conditions of that experience. Taking the state of mind in which a thing or event gets perceived to be as critical as our knowledge of the manner in which the thing or event is caused, transcendentalism combined an idealist vision of the world as created or coloured by the human spirit, with a Romantic characterization of human beings as losing interest in the world. This loss of interest is taken not merely as a mood but as an increasingly fundamental and characteristic relationship between the individual and the world. The sense of boredom and melancholy concerning the world was understood as affecting the very perception of the world. A renewal of interest in the world was regarded as necessary to the renewal both of knowledge and of the powers of creativity.

Thoreau's transcendentalism provides a critical piece of the background to his justification of resistance to unjust government. 'Civil Disobedience' is both a statement of principle and an account of the actions which led to his spending a night in jail for refusing to pay a poll tax. The work was intended as both an intellectual example and a moral provocation. Thoreau exhibits the connections of the call to conscience with his own call to a less economically dependent mode of life. As Plato's philosopher must go back into the cave, so Thoreau declares that some of his 'generative energy' must be devoted to abolishing the conditions that prevent other individuals from responding to fundamental human possibilities. His appeal to his fellow citizens to resist the tyrannies of slavery and imperialism are couched in terms that demonstrate his hopes for a national awakening and moral regeneration.

3 Cavell's reading of Thoreau

In *The Senses of Walden* (1972), the American philosopher Stanley Cavell depicts the unwillingness of readers to grant the powers of philosophy or consecutive thought to Thoreau's words. He characterizes this blankness towards Thoreau as an estrangement from a region of our own powers of thought. Cavell's project has spiralled back to a reconsideration of Emerson's work, and he propounds a vision of perfectionism in Thoreau and Emerson. This interpretation of a dimension of moral thought at least as old as Socrates' 'care of the soul' requires the shedding of false possibilities and false necessities. Only by giving up an apparent wholeness of the self can we take a step towards an integrity that is internal to the self's own constitution. Emerson and Thoreau thus stand at the core of a tradition of American thinking that has been neglected and even suppressed.

In characterizing our distance from Thoreau's writing as a form of distance from our own capacities for self-reflection, Cavell's work provides the grounds for considering Thoreau's relation to philosophy. Among the central claims of Cavell's investigation are these: (1) Thoreau's penetration of the world is made possible by his acceptance of our necessities and by the capacity to convert necessity into forms of habitation. Specifically, (2) human labour is not only one of the necessities of our lives but one of the forms in which we make contact with the necessities of the world. All forms of labour are thus expressive of one another, not only in the debased terms of economic value but as different modes of knowledge and of contact with reality (see Work, philosophy of). (3) Cavell locates in *Walden* the apparent paradox according to which we must let a book provide the terms that teach

us how to read the book. The immediate methodological step is that we are to relinquish our inherited images and ways of reading. (4) Cavell discerns in Thoreau the intention to compose a nation's scripture, overturning not only personal deadness to our words but a national faithlessness to its founding vision. (5) There is a corresponding project to redeem the economic terms of value and expression, and this project in turn releases a more humanly adequate conception of labour and value. (6) Cavell discovers in Thoreau a quasi-Kantian vision of the necessities of life in nature as creating the medium within which a life of freedom is found to be liveable. Thoreau's task is not so much to learn independence by living in the woods as to find out what could make his life in the woods a form of independence. From the beginning of his sojourn there, he demonstrates his awareness that he must be prepared to leave it.

4 Thoreau as inheritor of philosophy

Thoreau's work often invokes the figure of the philosopher, as when he writes that 'To a philosopher all news, as it is called, is gossip' (Thoreau [1854] 1963 ch. II: 70). Thoreau thought of newspapers as purveying what was essentially gossip, and urged philosophers to find ways of getting beyond the superficial claims of merely current events. At the same time, Thoreau characterizes gossip as one of the primary modes in which the members of modern society communicate messages to each other. What makes it gossip is not the content of what is said but the uprooted disposition of the speakers from the ground of their real concern. Thoreau understood gossip as disguising the sources of our concern and the unsettled quality of our communications.

The sense of the social as precisely constituted by a certain kind of talk is most immediately related to Emerson's depiction of the whole 'cry of voices' standing against the solitary voice of the self-reliant. Thoreau's sense of social communication also bears comparison to the fifth chapter of J.S. Mill's *On Liberty* and to the places in *Being and Time* where Heidegger attempts to specify the conditions of 'mere talk' (*Gerede*) as constitutive of our being with one another. The degeneracy of our talk and speech is linked by Thoreau ([1854] 1963 ch. XIII: 186) to the distance between our 'parlours', or places of speech, and the 'kitchens' where our the object of our needs are most immediately transformed into the possibility of satisfaction. The proximity of human words to the concerns from which the words arise is part of the accomplishment of the truer speech and deeper knowledge which is the province of the humanities.

Thoreau sees his own capacity for vision as deriving from the same sources as other philosophers, in other places and other epochs. He explicitly mentions sharing certain perceptions with the 'oldest Egyptian or Hindoo philosopher' ([1854] 1963 ch. III: 74). This mingling of identities in our capacity to apprehend our condition is not mystical. It is a precise declaration of the impersonality of that part of the mind which is capable of spectatorship and criticism. He declares his sense that:

> To be a philosopher is not just to have subtle thoughts but so to love wisdom as to live according to its dictates, a life of simplicity, magnanimity, and trust. It is to solve some of the problems of life, not only theoretically, but practically.
>
> (Thoreau [1854] 1963 ch. I: 10)

Many readers have taken such remarks to imply that Thoreau is exhorting us to bring our lives into accord not only with simplicity but with a relatively simple doctrine of life. Richardson's 1986 biography simply omits the word 'just', as if to insist that a philosopher should dispense with subtle thoughts. Thoreau associates philosophy with its classical task of searching out and heeding the dictates of a practical wisdom. This association is a valuable clue to Thoreau's conception of a philosophical life. Thoreau's search for what he called the writer's 'equable life' is also a search for the 'occasions' of writing ([1854] 1963 ch. III: 76). These occasions provide simultaneously a chance for immediate perception and a chance to secure the vantage points of his observations. The reader must remember that the search for these occasions of simplicity is not itself very simple. Thoreau scrupulously records the failures as well as the successes in his search for the conditions of writing and observation. In these moments of reflection and self-criticism, we find Thoreau's deepest engagements with the issues of philosophy. One train of Thoreau's thought about simplicity leads to the analogy that:

> Our life is like a German confederacy, made up of petty states, with its boundary forever fluctuating, so that even a German cannot tell you how it is bounded at any moment.
>
> (Thoreau [1854] 1963 ch. II: 68)

The object of this little gibe is the effort of post-Kantian German philosophers to chart in systematic prose the limits of the mind and the limits of its freedom. As *Walden* unfolds, we learn that 'simplicity' is only one face of the answer. Simplicity is as little or as much of a solution to the problem of the soul as it is in Plato's *Phaedo*. It is the smallness as well as the transience of the states composing our minds that Thoreau wants us to overcome.

Thoreau depicts the fluctuation of the boundaries of our life as containing a philosophically positive possibility. In the conclusion of *Walden* he speaks of passing beyond an 'invisible boundary'. He goes on to say that 'in view of the future or possible' we should learn to live 'quite laxly and undefined in front' ([1854] 1963 ch. XVIII: 245). He was not content to arrive at a simple life, if that meant establishing rigid defences against the complications of society. He also suggested that the transformations of an individual's development could not be bounded by a single experience, or a single theory of such an experience. The primary example in his culture was the Christian experience of conversion and its role in the doctrine of atonement.

Thoreau's reflections on language attempt to characterize the volatile presence of words at the core of all experience. The transience of our spoken words is analogous to the disappearance of significance from a written text. Both can be instructive to those who know how to listen and to read. An appropriate mode of reading points inwards to the possibility of locating our freedom in a world of natural law. For Thoreau the act of reading thus takes on some of the functions of judgment in Kant's *Critique of Aesthetic Judgment*. As judgments of the beautiful are meant to apprise us of 'traces' of harmony between nature and the functioning of our practical reason, so the act of reading teaches us how to understand and to appropriate what is rightfully ours. For Thoreau, reading is the path on which we learn to conceive the connection between the accidents of nature and the necessities of human freedom.

See also: American philosophy in the eighteenth and nineteenth centuries

TIMOTHY GOULD

List of works

Thoreau, H.D. (1906) *The Writings of Henry David Thoreau*, Boston, MA: Houghton Mifflin, 20 vols; vols. 7-20, ed. B.Torrey and F. Allen repr. as *The Journal of Henry David Thoreau*, ed. W. Harding, Salt Lake City, UT: Gibbs & Smith, 1984.(Volumes 1-5 were unedited reprints from earlier editions of his works; volume 6, 'Familiar Letters', was edited by F.B. Sanborn. Volumes 7-20 were the first complete edition of the 'Journal'.)

Thoreau, H.D. (1981-) *The Writings of Henry David Thoreau*, ed. E. Witherall, Princeton, NJ: Princeton University Press.(Now the standard scholarly edition.)

Thoreau, H.D. (1849a) *A Week on the Concord and Merrimack Rivers*, Boston, MA: James.(Structured as a journey, it contains early versions of Thoreau's preoccupations with nature, friendship, memory and the act of writing.)

Thoreau, H.D. (1849b) 'Resistance to Civil Government', *Aesthetic Papers*, ed. E. Peabody, Boston, MA: G.P. Putnam, 189-211.(Later re-titled 'Civil Disobedience' and often reprinted.)

Thoreau, H.D. (1854) *Walden; or Life in the Woods*, Boston, MA: Ticknor & Fields; repr. as *The Variorum Walden*, ed. W. Harding, New York: Washington Square Press, 1963. (The subtitle was omitted after the first edition. The 1963 annotated edition contains useful accounts of sources, references and background.)

Thoreau, H.D. (1864) *Maine Woods*, Boston, MA: Ticknor & Fields.(Contains an account of climbing Mount Ktadin, a late masterpiece of Thoreau's writing.)

Thoreau, H.D. (1890) *Anti-Slavery and Reform Papers*, London: Swan Sonnenschein.(Contains 'Civil Disobedience', 'A Plea for Captain John Brown' and other political writings.)

Thoreau, H.D. (1981) *Walden and other Writings*, ed. W. Howarth, New York: Random House.(Good selection of writings including *Walden*, 'Walking', 'Civil Disobedience' and 'A Plea for Captain John Brown'.)

References and further reading

Cameron, S. (1985) *Writing Nature: Henry Thoreau's Journal*, New York: Oxford University Press.(Argues for the priority of Thoreau's Journals as the main embodiment of his literary enterprise.)

Cavell, S. (1972) *The Senses of Walden*, Chicago, IL: University of Chicago Press; expanded edn, 1980.(An exploration of Thoreau's relation to Romanticism, Emerson, Kant and modern philosophy.)

Emerson, R.W. (1883) 'Thoreau'; repr. in *The Complete Works of Ralph Waldo Emerson*, vol. 10, ed. E.W. Emerson, Boston, MA: Houghton Mifflin, 1903-4.(A revised version of Emerson's memorial address, delivered

at Thoreau's funeral, containing a pertinent introduction to the man and his work.)

Gould, T. (1992) 'Henry David Thoreau', *Encyclopedia of Ethics*, New York and London: Garland Press.(Outline of Thoreau's contribution to transcendentalism, moral perfectionism and ethical action.)

Harding, W. (1966) *The Days of Henry Thoreau*, New York: Knopf.(Reconstructs the way Thoreau spent his days from the writings themselves.)

Heidegger, M. (1927) *Being and Time*, trans. J. Macquarrie and E. Robinson, New York: Harper & Row, 1962. (Heidegger's first masterwork, containing diagnoses of inauthenticity and idle speech as intrinsic to the social constitution of the human being.)

Kant, I. (1790) *Critique of Judgment*, trans. W. Pluhar, Indianapolis, IN: Hackett Publishing Company, 1987. (Kant's third critique cast an immense shadow on the thinkers of New England, in part through the influence of Goethe and Coleridge. See especially §42, pages 165-70.)

Mill, J.S. (1859) *On Liberty*, in *The Philosophy of John Stuart Mill*, ed. M. Cohen, The Modern Library Edition, New York: Random House, 1961.(See pages 248-71. In addition, this collection also contains Mill's writing on poetry, originality and Coleridge as opposing conformity and custom which forms a rare contemporary point of comparison with the work of Thoreau and Emerson.)

Miller, P. (1958) *Consciousness at Concord: The Text of Thoreau's Hitherto Lost Journal (1840-1841)*, Boston, MA: Houghton Mifflin.(A polemical but still useful confrontation with Thoreau, by one of the most important historians of American literature.)

Myerson, J. (1988) *Critical Essays on Henry David Thoreau's Walden*, Boston, MA: G.K. Hall.(Contains a selection of nineteenth-century reviews, including George Eliot, Horace Greely and others, and twentieth-century essays.)

Packer, B.L. (1995) *The Transcendentalists*, in *The Cambridge History of American Literature*, ed. S. Bercovitch, Cambridge: Cambridge University Press, vol. 2, 329-604.(A consideration of Thoreau's literary projects in relation to transcendentalism and to the social and political conditions of the age.)

Paul, S. (1958) *The Shores of America: Thoreau's Inward Exploration*, Urbana, IL: University of Illinois Press.(A major critical statement about Thoreau's enterprise as a writer.)

Richardson, R., Jr. (1986) *Henry Thoreau: A Life of the Mind*, Los Angeles, CA: University of California Press. (Intellectual biography which maps Thoreau's background of reading and some central themes of his writing against the major events of his life.)

Sattelmeyer, R. (1988) *Thoreau's Reading: A Study in Intellectual History, with Bibliographical Catalogue*, Princeton, NJ: Princeton University Press.(Extensive background, with chronological details of Thoreau's reading.)

Thought experiments

Thought experiments are strange: they have the power to present surprising results and can profoundly change the way we view the world, all without requiring us to examine the world in the way that ordinary scientific experiments do. Philosophers who view all hypothetical reasoning as a form of thought experimentation regard the method as being as old as philosophy itself. Others maintain that truly informative thought experiments are found only in mathematics and the natural sciences. These emerged in the seventeenth century when the new experimental science of Bacon, Boyle, Galileo, Newton and others forced a distinction between the passive observation of Aristotelian mental narratives and the active interventions of real-world experiment. The new science gave rise to a philosophical puzzle: how can mere thought be so informative about the world? Rationalists argue that thought experiments are exercises in which thought apprehends laws of nature and mathematical truths directly. Empiricists argue that thought experiments are not exercises of 'mere thought' because they actually rely upon hidden empirical information - otherwise they would not count as experiments at all. More recently it has been argued that thought experiments are not mysterious because they are constructed arguments that are embedded in the world so as to combine logical and conceptual analysis with relevant features of the world.

1 Features of thought experiments

As literary entities, thought experiments are more accessible, repeatable and persuasive than any real-world experiment. Thought experiments have other features, which we illustrate with an experiment by Simon Stevin. This shows that the force needed to balance a ball on an inclined plane varies inversely with the length of the plane. We know that a ball on a smooth horizontal plane is supported by the plane: no force is needed to keep it stationary. A ball against a vertical plane is not supported by the plane at all, so a force equal and opposite to its weight is needed to keep it stationary. What force is needed for the intermediate cases where the plane is inclined?

Can this question be answered just by thinking? First imagine a triangular prism positioned with its base horizontal some distance above a table. Now imagine a necklace consisting of fourteen identical, perfectly spherical-balls draped over the prism so that four rest on the left-hand longer incline and two rest on the shorter incline. The remainder of the chain forms a loop below the prism. The balls are connected at equal distances in a way that allows them to roll freely. Will the necklace move? To imagine that it moves is to suppose that perpetual motion is possible. Observe, instead, that the necklace is in equilibrium. The third step in the experiment is to imagine cutting the necklace simultaneously at two points just beneath the prism (or more easily, perhaps, at a single point where it is closest to the table). What happens to the six balls remaining on the prism? Our knowledge of this situation suggests that they will remain in equilibrium: four balls on the left balance the two on the steeper incline on the right. This is an imaginary result. But we do not need to make a real experiment because to suppose any other outcome is to deny one or more conditions of the experiment. Now the length of each plane is given by the number of balls resting on it (here, $4:2$), so the force needed to balance any ball varies inversely with the length of the plane.

This illustrates nine features shared by most thought experiments in science and by many in philosophy:

(i) Method of abstraction: the experiment presupposes a method (determine the limiting cases (force = 0, force = weight of the ball)) then constructs an idealized intermediate case in which only two outcomes are possible.
(ii) There is a situated procedure: this is conveyed as a narrative describing a sequence of operations on simple objects, which takes place in an idealized situation (Stevin's set-up is frictionless).
(iii) Causality is important: operations are connected to outcomes by causal principles and causality is governed by simple principles such as symmetry or conservation.
(iv) Intervention is important but unproblematic: the reader/observer is involved, but is perfectly competent (the cuts in the necklace must be made perfectly simultaneously, so they are).
(v) Interpretation is constrained: the scenario is interpreted according to agreed principles (impossibility of perpetual motion).
(vi) Results are unambiguous and are usually limited to just two possibilities (the balls move, or they do not). Binary outcomes are a feature through which thought experiments expose the tension between two disparate ways of thinking about the situation.
(vii) Implications are clear: the meaning of the result is unequivocal (lack of motion implies the proportionality

described above). We describe other examples below in which thought experiments generate a paradox or contradiction whose resolution requires a profound change of view.
(viii) Implications are compelling: however surprising the resolution may be, the intellectual experience engendered by a thought experiment is compelling; whether the experiment could be performed in the world seems as irrelevant as whether the story-line of a good joke is actually true.
(ix) Thus, a thought experiment always works and is easy to replicate. In this respect it differs from real experiments (excepting text-book idealizations).

A thought experiment is an idealization which transcends the particularity and the accidents of worldly human activities in order to achieve the generality and rigour of a demonstrative procedure. We argue in §4 that this effectiveness owes as much to knowledge brought to bear by a reader of the narrative as it does to the ingenuity and knowledge of the experimentalist who authored it. We note in §3 that thought experiments in philosophy are typically less constrained by situational knowledge and are open to counter-interpretation; that is, they may lack features (ii), (iii) and (v) and, therefore, (vii) as well.

2 Thought experiments in science

In Aristotelian and medieval science the narration of an experiment usually supplanted actual performance. The new experimental narratives of Francis Bacon, Boyle, Galileo Galilei, and Newton discredited these speculative narratives. By the end of the seventeenth century actual performance was necessary to establish a fact. Yet thought experiments are advanced as fictions. How could they assume such an important role in the new science? A Platonist explanation advanced by Koyré is that thought experiments became central to science because they use idealizations necessary for mathematical descriptions of nature. Since nature is ultimately a rational mathematical system it is to be expected that such experiments deliver truths about reality. They work because they appeal directly to intuitive knowledge rather than empirical experience. Philosophers go on to argue that the informativeness of thought experiments is evidence for the Platonist view that human knowledge is of transcendent realities. Einstein, a master of thought experiments, is often enlisted in support of this rationalist view.

The power of thought experiments as seemingly independent of publicly repeatable observations led a few empiricist philosophers, such as Pierre Duhem (1954) and Carl Hempel (1965), to deny that they constitute a legitimate method of scientific investigation: at best they have an heuristic role as guides to discovery. Another empiricist interpretation is that all thought experiments are deductive arguments whose empirical content is hidden as suppressed premises. In a response to Koyré’s Platonist interpretation, Kuhn (1964) argued that thought experiments can work effectively only in simulated worlds to which experimenters successfully apply their own real-world knowledge and experience. We develop this important insight further in §4. It is true that most thought experiments can be reconstructed as deductions; however, this does not show that thought experimentation is deductive. Some philosophers argue that the content of an idealization may be greater than can be expressed by a set of propositions. Thought experiments are narratives from which readers construct non-propositional mental models. The persuasiveness of such experiments does not necessarily depend on logical structure. Logic is only used afterwards, to validate conclusions reached by thinking with models.

The view that thought experiments are no more than deductions can be refuted by an example, one of Galileo’s best-known objections to the Aristotelian doctrine of motion applied to falling bodies. According to Aristotle (and to untutored common sense) heavy bodies fall more quickly than light ones. Imagine that two stones, L and H, are falling. Grant the common-sense belief that H falls more quickly than L. Now imagine that H becomes attached to L. Since light objects fall more slowly, L will retard the fall of H. Yet $L + H$ weigh more than H, so the conjoined stones must fall more quickly than H. This experiment displays all the features listed in §1. There is a narrative in which a sequence of events and interventions can cause two contradictory outcomes. The experimenter must choose. In a truly efficacious thought experiment the procedures cannot be faulted (as they routinely are in real experiments). A contradiction or paradox therefore exposes a fallacious assumption rather than incompetent practices. Here the only candidate is the postulate that velocity is proportional to weight. The problem is resolved if velocity is not affected by weight:

$$V_H = V_L = V_{H+L}\ .$$

This means that the accelerative force is the only relevant factor. Galileo’s narrative sequence may be

reconstructed as a deductive argument. However the crucial move - assenting to the proposition that ' $V_H = V_L = V_{H+L}$ ' - cannot be *deduced* from what has gone before. It is an insight, a new way of perceiving the whole sequence of events which also creates new experimental possibilities.

After a crucial insight has been accepted, another kind of thought experiment is possible, in which material impediments such as air resistance are removed (see Idealizations). These idealizations sometimes become real experiments as technology develops. Consider an imaginary experiment in Newton's *System of the World* in which the altitude of the point from which a cannon ball is fired becomes higher and higher so that the projectile eventually goes into orbit. Incremental steps take us from plausible, familiar circumstances and consequences to one that is impossible and implausible. Newton's purpose was to demonstrate the counter-intuitive proposition that the force that pulls a projectile back to the ground is the same force that holds the moon in orbit. Here conceivability suggests the physical possibility of artificial satellites. Viewed as a deductive argument, it presents directly the impossible consequence of applying the laws of gravitational and centripetal force to projectiles. We can reconstruct this as a deductive argument, but as a self-contained argument this is no longer an experiment.

3 Thought experiments in philosophy

Thought experiments are used widely in philosophy. Some philosophers define thought experiments to include any argument that invokes a hypothetical situation (as, for example, moral dilemmas raised by a pressing need to choose between two equally unjustifiable actions). Others hold that such arguments only appear to be experiments. Plato argued Socrates' thesis that all knowledge is recollection by means of a thought experiment in the *Meno* in which Socrates responds to a dilemma about inquiry (see Plato §11). Someone posing a question either knows the answer (so has no need to ask) or has no knowledge of the answer (so would not recognize the answer). The questioning is therefore pointless. Meno's uneducated slave boy is asked to construct a square double the size of a given square. Does his success prove that the boy remembers mathematical knowledge that he had not realized that he possessed? Not necessarily. An alternative interpretation is that in the course of eliciting an answer, Socrates' questioning educates the boy, enabling him to solve the problem. This illustrates the point made in §1 that scientific thought experiments are less open to interpretation than are philosophical ones.

Some philosophical experiments are imaginary excursions into intriguing but impossible worlds. Hobbes' refutation of the immateriality of mind involves imagining oneself shrinking until one can wander amongst the smallest corpuscular components of the brain. Could such an observer see thoughts? Not if thoughts are due to the interactions of corpuscles. The experiment works by diverting our attention from the fact that to enter this corpuscular world is to assent implicitly to a materialist metaphysics. One can always deny the presumption on which the possibility of making the experiment depends. If, for example, we re-interpret Plato's narrative about Meno's slave as an example of instruction, then it ceases to be a demonstration that knowledge is the recollection of eternal, transcendent truths. Because of this interpretive flexibility about central features of the narrative, philosophical thought experiments are rarely as decisive as their scientific counterparts.

The argument of Searle's Chinese Room experiment is that semantic meaning is not reducible to physical symbol-manipulation. This develops a version of Turing's behavioural test of the indistinguishability of the linguistic competence of machines and humans. Searle aims to undermine Turing's criterion of equivalence by granting to a Chinese-speaking room the capacity to do whatever is necessary to pass Turing's test (that is, solve arithmetical problems, converse about preferences, tell jokes and so on). The experiment involves examining the inner workings, especially the procedures whereby an operator inside applies rules of syntax to interpret and respond to questions. Since the operator can do this without understanding any Chinese pictogram, Searle concludes that symbol manipulation does not involve semantic understanding of this or any other language (see Chinese room argument). Are Hobbes and Searle describing experiments? Or do their narratives have only the appearance of being experimental? These exercises expound the consequences of a metaphysical position. It is difficult to see that they test it in the way that Galileo's experiment (see §2) tests Aristotle's theory of motion.

4 Further issues

A thought experiment constructed by an ill-informed experimenter may persuade, but it will also mislead if it is not constrained in the way that thought experiments in mathematics and physics are. This is why similarities between armchair experiments and scientific ones should not be pressed too far. The constraints are illustrated by

an armchair refutation of Einstein's principle that no signal can travel faster than light. Imagine that a rigid rod is extended upwards from the surface of the rotating earth. The velocity of its tip will eventually exceed the speed of light. However this logical necessity has no physical relevance. Physicists deny the possibility of such an experiment, pointing to the fact that, as material from the earth is used to extend the rod, the velocity of rotation must decrease. The experiment indicates ignorance of the relevant physical laws. A further example illustrates the importance of knowing how to apply such knowledge consistently. What happens to a very small hole bored through a large metal ring when the latter is heated: does the hole get larger or smaller? A would-be experimenter needs to know both that heating causes expansion and how to apply this fact to the ring. What is needed is a combination of empirical knowledge and the ability to reason with it. This fact opens the way for an answer to our opening question about the informativeness of imaginary experiments.

Can the mysterious efficacy of thought experiments be explained? We approach this by addressing two further questions:

(1) What distinguishes this kind of mental activity as experimentation rather than hypothetical reasoning? If a thought experiment is a purely mental affair then it cannot be an experiment at all. If it is not a purely mental affair then there is no paradox. Naturalism argues that personal participation - represented by items (ii)-(iv) in §1 - is what makes a thought experiment more than a hypothetical argument. (This is why we perform experiments but do not 'perform' an argument). A thought experiment is effective if it is both logically well-formed and empirically informed about relevant features of the world it purports to investigate. It expresses the minimum set of conditions necessary to sustain the appearance of an argument of some generality, whilst giving the appearance of the situated-ness that real experiments actually have.

(2) Does an informed thought experiment express tacit knowledge or intuitively apprehended principles? Empiricists have argued that thought experiments articulate and apply intuitions formed from human experience. Ernst Mach appealed to Darwin's principle of natural selection to guarantee that these intuitions adequately represent the world: natural selection favours those minds which successfully imitate causal patterns in nature. This explanation embeds experiment in the world but it does not go far enough. Naturalistic explanations emphasize cognitive capacities and culturally transmitted knowledge (see Naturalized philosophy of science). Such capacities and knowledge enable abstraction. The degree of abstraction possible in a thought experiment depends on how much both its author and its readers have participated in the culture of the experiment. In this respect thought experiments have much in common with jokes. Both are sparse, carefully crafted, narratives which include only essential details. There is a punch-line requiring an insight which changes our understanding of the story. In both cases we see the point without its being articulated as an argument. Yet philosophers do not appeal to intuition or sense experience to explain how people understand jokes. Thought experiments are powerful because they appeal to lived experience of a world which their narratives reflect, selectively, back at us.

See also: Empiricism; Experiment; Rationalism; Scientific method

DAVID C. GOODING

References and further reading

Brown, J.R. (1991) *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, London: Routledge.(An accessible account of influential thought experiments in physics and mathematics; defends the rationalist interpretation advanced by Koyré, described in §2.)

Duhem, P. (1954) *The Aim and Structure of Physical Theory*, Princeton, NJ: Princeton University Press. (Originally published in the *Revue de Philosophie* 4, 1904; 5, 1904; and 6, 1905. Duhem's rejection of thought experiments for scientific investigation.)

Gooding, D. (1993) 'What Is *Experimental* about Thought Experiments?', in D. Hull, M. Forbes and K. Okruhlik (eds) *PSA 1992*, East Lansing, MI: Philosophy of Science Association, vol. 2, 280-90.(Develops a naturalistic resolution of the paradox generated by approaching thought experiments from either a rationalist or an empiricist standpoint.)

Hempel, C.G. (1965) *Aspects of Scientific Explanation*, London: Macmillan, and New York: Free Press. (Reproduces some fifteen previously published essays, with commentary, blending and revision. The title essay is original. Hempel, like Duhem, rejects the legitimacy of thought experiments in scientific investigation.)

Horowitz, T. and Massey, G.J. (eds) (1991) *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman & Littlefield.(Essays in the analytical tradition illustrating the variety of thought experiments in

philosophy, logic, mathematics and various sciences.)
Kuhn, T.S. (1964) 'A Function for Thought Experiments', in *L' aventure de la science, Mélanges Alexandre Koyré*, Paris: Hermann, vol. 2, 307-34, 2 vols; repr. in T.S. Kuhn *The Essential Tension*, Chicago, IL: University of Chicago Press, 1977, 240-65.(This influential essay argues that thought experiments bring about major changes of viewpoint by using tensions between conceptual schemes to expose anomalies; that is, inconsistencies in the way that one of the schemes relates to the world.)
Sorensen, R.A. (1992) *Thought Experiments*, Oxford: Oxford University Press.(An accessible comprehensive study, containing many examples, which develops an empiricist explanation of the efficacy of thought experiments as limiting cases of ordinary experiments.)
Tiles, J.E. (1993) 'Experiment as Intervention', *British Journal for the Philosophy of Science* 44: 463-75. (Develops the argument in §2 that the new science of Bacon and others forced a distinction between the passive observation of mental narratives and the active interventions of experimental ones, giving rise to the distinction between thought experiments and real-world experiments).

Thrasymachus (late 5th century BC)

Thrasymachus, a Greek Sophist and orator, is known principally for his role in book 1 of Plato's Republic, in which he argues that justice is simply a social institution created by rulers to further their own interests. It is intended solely for the subjects; the rulers themselves need not practise it. Since justice thus consists in promoting another's advantage rather than one's own, injustice is far more profitable. Apart from issues of internal coherence, his claims raise many questions. What, for example, are our true interests? And what are the actual and ideal operations of power?

Very little is known of Thrasymachus' life. He was a teacher of rhetoric, of an apparently aggressive disposition, although Plato's portrait may be distorted by anti-Sophistic bias (see Sophists). In the *Republic* (see Plato §14) he makes two main speeches, and it is debated whether their claims are compatible. In the first he declares that what we call justice is simply the interest of the stronger: in every city the stronger wield political power and make the laws in their own interest, calling obedience to these laws 'justice'. Whether the government is tyrannical, aristocratic or democratic, the institution of justice is merely a cover for *force majeure*. This legalistic conception of justice, however, is modified when Thrasymachus adds that obedience to the laws will only constitute justice when the laws really are in the interests of the rulers, since the rulers may be mistaken about where their advantage lies.

In his second speech Thrasymachus extends the discussion to include private as well as political relations. What we term injustice is simply the pursuit of self-interest, and pays the individual far better than justice. He boldly suggests that injustice is positively a virtue (*aretē*), the virtue of common sense. The unjust are prudent and admirable; the just are naïve fools. Indeed, we partially admit this: although we criticize and punish petty wrongdoers, when injustice is practised on a sufficiently grand scale by, say, a tyrant, we merely call the tyrant fortunate. We would all be tyrants if we could.

These two positions may appear at odds. If justice is simply the interest of the stronger, then surely the tyrant who supremely promotes his self-interest is supremely just? Yet Thrasymachus clearly states that the tyrant is supremely *un*just. Some have argued that he is operating with two notions of justice, conventional and natural (see Callicles), but a simpler reading suggests that a conventional description is retained throughout. Injustice is ruthlessly promoting your own desires at the expense of others; justice is giving others their due. In theory each can be practised by both rulers and ruled (as Thrasymachus admits in his second speech); in practice, however, opportunities for injustice will chiefly be available to the rulers, whereas the ruled will generally have to accept the imposition of justice. The initial statement that justice is the interest of the stronger is therefore not to be taken as a complete definition, but as a description of how justice operates in the world of *realpolitik*.

Thrasymachus does, however, complicate his position by admitting that the practice of injustice may sometimes be an unsuccessful means of satisfying one's desires (one may get caught and imprisoned), and this leaves open the possibility that in some circumstances justice would be the more prudent course of action. His implicit ideal is the person who successfully promotes their own interests. This will often involve acting unjustly, but not necessarily always - a key consideration will be one's political status. Underlying the ideal is a notion of human flourishing seen in terms of material wealth, autonomy and power.

Thrasymachus' position thus goes deeper than its superficial cynicism might suggest. But does his ideal carry with it any notion of moral obligation? The answer depends on how one interprets (1) his use of the word 'virtue', and (2) his claim that everything normally predicated of justice should in fact be predicated of injustice. Whether he is prescribing or simply commending his ideal, however, the practical implications remain unclear. If the rulers fare so much better than the ruled, is he advocating that the ruled seize power? Or that they embark on a programme of mass disobedience? Or is he just recommending opportunism? Whatever the answer, it is plain that his doctrines could be used to endorse all these possibilities, plus tyranny and imperialism, as Thucydides shows. It is because of this that Plato takes his challenge so seriously in the *Republic*.

It is also debated whether Thrasymachus actually endorses the doctrines he voices. The ambiguous tone of his responses to Socrates precludes certainty, but it is possible (and may be supported by a separate fragment) that his

attitude is that of the embittered and disillusioned moralist, ironically praising injustice because justice no longer pays. His intention may be simply to expose current hypocrisies, rather than to applaud their manipulation.

See also: Antiphon; Physis and nomos

ANGELA HOBBS

References and further reading

Chappell, T.D.J. (1993) 'The Virtues of Thrasymachus', *Phronesis* 38: 1-17.(A crisp and stimulating discussion.)

Guthrie, W.K.C. (1969) *A History of Greek Philosophy*, vol. 3, Cambridge: Cambridge University Press; part of vol.3 repr. as *The Sophists*, Cambridge: Cambridge University Press, 1971.(Thoughtful discussions of Thrasymachus, especially pages 294-8.)

Kerferd, G.B. (1947-8) 'The Doctrine of Thrasymachus in Plato's Republic', *Durham University Journal* 40: 19-27; repr. in C.J. Classen (ed.) *Sophistik*, Darmstadt: Wissenschaftliche Buchgesellschaft, 1976, 545-63.(A classic study; lucid and exhilarating.)

— (1981) *The Sophistic Movement*, Cambridge: Cambridge University Press.(The best introductory handbook to the Sophists in general; for Thrasymachus see especially pages 120-3.)

Plato (*c*.380-367 BC) *Republic*, trans. H.D.P. Lee, Harmondsworth: Penguin, 1955; repr. 1974.(Book I is our main source on Thrasymachus.)

Sprague, R.K. (ed.) (1972) *The Older Sophists*, Columbia, SC: University of South Carolina Press, 86-93.(Full English translation of the fragments and testimonia from Diels and Kranz (1952).)

Thrasymachus (late 5th century BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 2, 319-26.(The standard collection of the ancient sources; includes Greek texts with translations in German.)

Thucydides (*fl. c.*400 BC)

A Greek historian with philosophical interests, Thucydides wrote about the Peloponnesian War between Athens and Sparta (431-404 BC). He elaborates on the decisions of war in brilliantly reconstructed debates and speeches, reflecting his training under various Sophists. Many of these speeches take for granted that people care less for justice than for their own narrow interests. This dark view of human nature influenced Hobbes, while the style of the debates and speeches has had an enduring effect on public rhetoric. His account of Athenian democracy in action is cautionary, and his conservative political views anticipated Aristotle's in some respects.

A wealthy Athenian, Thucydides served as general in 424 BC, but failed to prevent the Spartan capture of an important outpost and was punished by exile. He did not return to Athens until after its defeat in 404. Beginning work on his *History of the Peloponnesian War* during this exile, he brought his account up to the year 411, at which point it breaks off. The dates of his birth and death are not known, but it is likely that he was born between 460 and 455 and died early in the fourth century BC. He is said to have studied with the sophist Antiphon and the natural philosopher Anaxagoras, and it seems likely that he also learned style and argument from other Sophists such as Protagoras (see Sophists). His *History* is unusual in that nearly one fourth of the text consists of speeches and debates. These are all composed in Thucydides' style and display methods of argument taught by Sophists. Pericles' funeral oration is especially well known, with its reflections on the culture and constitution of Athens. Thucydides states his own views sparingly, 'having', as Hobbes puts it, 'so clearly set before men's eyes the ways and events of good and evil counsels, that the narration itself doth secretly instruct the reader' (Schlatter 1975: 18).

On human motivation, Thucydides' views are no secret. Fear, ambition and avarice dominate his explanations of events. Expansion of the Athenian empire is due to these factors; this expansion put fear into the Spartans which drove them to start the war. Thucydides' description of the plague at Athens in 430 BC shows how badly he believes people behave when they no longer have any reason to fear the laws. He sounds a similar theme in his account of a civil war on the island of Corcyra between the parties of 'the many' and the 'few', that is, the democrats and the oligarchs. These combatants were led by avarice and ambition to commit violent acts that they justified by a reversal in their use of moral language; the results included all sorts of betrayal, revenge and the slaughter of prisoners. Human nature, Thucydides believed, will overcome the effects of law whenever a breakdown of authority gives it the opportunity.

The same principle is at work in international affairs. Neither Athens nor Sparta considers justice when deciding how to treat its neighbour states. When the Athenians decide not to slaughter the people of Mytilene, whose rebellion they have just put down, they do so from a calculation of their own advantage. Justice, the Athenians believe, is relevant only when both parties are equally subject to authority. The Athenians do in each case what they deem necessary for the survival of their empire, and go so far as to conquer the harmless island of Melos and execute its male population for this reason. Although Sparta claims to support justice, it too destroys a city and its people out of self-interest.

Thucydides' dark view of human nature explains the value he places on maintaining a stable system of law and fending off radical democracy, which, in his view, leads to lawlessness. Thucydides praises Pericles for ruling like a monarch in Athens when it was a democracy 'in name only'. Much later, when Athens blundered into war with Syracuse, it was defeated by the conservative leadership of Hermocrates, which Thucydides contrasts with the foolish partisanship of the democrat Athenagoras. Conservative democracy was the strength of Syracuse, but democracy after Pericles did not serve Athens well. The best sort of constitution, according to Thucydides, is neither democracy nor oligarchy but 'a moderate blending of the interests of the few and many', in which neither party is allowed to make gains at the expense of the other.

The methods of argument displayed in the *History* are as important as the views expressed. Historiographers are interested in the manner in which Thucydides constructs an account of prehistoric Greece using myth, poetry, archaeology, local traditions and analogies with primitive peoples of his own time. Here and in the speeches he makes use of a type of inference known as *eikos* ('probable' or 'reasonable expectation'). Thucydides' speakers often argue in favour of actions they believe it is necessary (*anankē*) for them to take in view of the outcome they think it is reasonable to expect (*eikos*). Such a speaker is frequently contradicted either by the outcome (which is

often unexpected) or by another speaker making an equally powerful argument for the other side. Reasoning on such matters is what philosophers now call *defeasible*; it holds only for normal conditions and is defeated by unexpected abnormalities. Alcibiades, for example, predicts that when the Athenian army reaches Syracuse the people there will be divided, owing to their complex cultural heritage, and so more easily defeated. Such a strategy worked for Athens in cities where civil war was the norm. Syracuse, however, was exceptionally stable owing to the conservative leadership of its democracy; the city did not divide and it was not conquered. Unusual circumstances defeat arguments based on reasonable expectations, and this pattern is illustrated throughout the *History*. Thucydides' opposed pairs of speeches, as well as his frequent oppositions of speech and event, illustrate not only his theory of motivation but also kinds of reasoning, good and bad, that were current in the teaching of the sophists.

See also: Physis and nomos

PAUL WOODRUFF

List of works

Thucydides (*fl*. *c*.400 BC) *History of the Peloponnesian War*, trans. R. Crawley, London, 1876, revised by R. Feetham, Chicago, IL, 1910; trans. B. Jowett, *Thucydides, History of the Peloponnesian War*, 1881, revised by S. Hornblower, 3rd edn, 1995.(Crawley is the most admired and widely used translation; the Jowett translation reflects recent scholarship.)

References and further reading

Connor, W.R. (1984) *Thucydides*, Princeton, NJ: Princeton University Press.(The most thorough modern treatment.)

Farrar, C. (1988) *The Origins of Democratic Thinking*, Cambridge: Cambridge University Press.(Substantial discussion of Thucydides' importance to political thought.)

Hornblower, S. (1988) *Thucydides*, London: Duckworth.(Contains a useful discussion of Thucydides' views.)

Orwin, C. (1994) *The Humanity of Thucydides*, Princeton, NJ: Princeton University Press.(A discussion of Thucydides as a political thinker.)

Ostwald, M. (1988) *ANANKE in Thucydides*, American Classical Studies 18, Atlanta, GA: Scholars Press.(A brilliant study of Thucydides' views and method.)

Schlatter, R. (1975) *Hobbes' Thucydides*, New Brunswick, NJ: Rutgers University Press.(Hobbes' famous translation of 1629, still the most vigorous in English.)

Ti and yong

Ti *and* yong *(literally 'body' or 'substance' and 'use' or 'function') are technical terms in Chinese philosophy.* Ti *often is used to denote the essence or fundamental nature of a given thing, for example, 'the substance/true essence of the Way'. As a verb, it can also mean to 'embody' or 'instantiate' a given characteristic or virtue, for example 'to embody/fully realize humanity'. A thing's* yong *is its characteristic activity in accordance with its nature.*

Neo-Confucians made extensive use of the idea that the universe was - like the parts of a single individual - 'one body'. Those who realize this, both in the sense of cognitively assenting to this proposition as true as well as appreciating its moral implications, will feel an injury to another (in some cases even to inanimate objects) as an injury to their own bodies. The notion of 'one body' was also used in a related but distinct sense to describe phenomena which, while logically distinguishable, cannot actually exist apart from one another: in Chapter 15 of the *Platform Sutra*, a lighted lamp and its light are said to be 'one body' (see Platform Sutra). This recalls Aristotle's distinction between phenomena which are by definition distinct but by nature inseparable, such as concave and convex in the circumference of a circle.

The first occurrence of *ti* and *yong* as a conceptual pair occurs in Wang Bi's commentary on the thirty-eighth chapter of the *Daodejing*, where he uses it to discuss the *ti* (essence) of things and their *yong* (characteristic functions). Here, and in other passages, the *yong* of a thing is not simply what a given thing happens to be doing; there is a normative aspect to the notion. It is thus not wholly unlike Aristotle's use of the term *ergon* (see Aristotle §22). It is what that thing will do according to its particular *ti* (essence) when allowed to operate freely in an environment conducive to its natural functioning. Thus, in the passage from the *Platform Sutra* quoted above, the *yong* of a lighted lamp is to emit light; or as Chen Chun says, 'That which is one whole within is the substance; its response when stimulated is the function' (Graham 1992: 40).

The pair *ti* and *yong* are used extensively by neo-Confucian thinkers in the normative sense mentioned above. Thus, in his commentary on *Analects* 1.2, Zhu Xi says, 'Humanity is human nature, filial piety and brotherly respect are its *yong* "function"'. On *Analects* 1.12, he cites an earlier commentator who says, 'The *ti* "substance" of all rituals rests in reverence but in their *yong* "application" harmony is valued'.

Later Confucian thinkers tried, unsuccessfully, to deploy the two concepts to effect a synthesis or at least a symbiosis of Chinese and Western culture. The first and best-known example of such attempts was that of Zhang Zhidong (1837-1909), who coined the slogan, 'Chinese learning for *ti* "substance" and Western learning for *yong* "utility"' (Levenson 1965: 60): traditional Chinese culture was to serve as the fundamental basis for a modern Chinese society that adopted Western technology and ideas, but only for their efficacy. However, since the *ti* and *yong* of any given thing are inextricably related to one another, there was no hope of ever separating them in the way Zhang suggested.

See also: Daodejing; Neo-Confucian philosophy; Platform Sutra

PHILIP J. IVANHOE

References and further reading

Graham, A.C. (1992) *Two Chinese Philosophers*, LaSalle, IL: Open Court. (Contains very helpful discussions and illustrations of both terms; see the index under *t'i* and *yung* (substance and function).)

Levenson, J.R. (1965) *Confucianism and Its Modern Fate*, Berkeley, CA: University of California Press. (Contains an excellent discussion of late nineteenth and early twentieth century Chinese attempts to apply the notions of *ti* and *yong* to meet the challenges posed by Western imperialism; see in particular Chapter 4, pages 59-78.)

Tian

Tian*, conventionally translated as 'Heaven', is both what our world is and how it is. The myriad things are not the creatures of* tian *or disciplined by a* tian *which stands independent of what is ordered; rather, they are constitutive of it.* Tian *is both creator and the field of creatures. There is no apparent distinction between the order itself and what orders it. This absence of superordination is a condition made familiar in related notions of the Daoist* dao *and the Buddhist* dharma*, which also refer to concrete phenomena and the order that obtains among them. On this basis,* tian *can be described as an inhering, emergent order negotiated out of the dispositioning of the particulars that are constitutive of it. In the human world,* tian *is the experience of meaningful context felt differently by each person in the fellowship of family and community.*

The conventional translation of *tian* as 'Heaven' (with a capital 'H') inappropriately conjures up the notion of transcendence, concealing precisely those aspects of the term most essential to a correct appreciation of its meaning. We can, however, make several observations that reinstate aspects of *tian* that tend to be concealed by this translation. First, the association between *tian* and the sky encourages proper notice of the profound temporality and historicity that attends this idea, frustrating any analogy one might want to find between '*tian*' and 'sky' on the one hand and 'Heaven' and 'the heavens' on the other. *Tian* is inextricably linked to the pervasive processes of change, and is understood often as an abbreviation of *tiandi* (heaven and earth), or 'the autogenerative and self-sustaining world'. The Judaeo-Christian God, often referred to metonymically as 'Heaven', *creates* the world, but classical Chinese *tian is* the world.

Tian further is not only 'the sky', but an articulated and patterned sky. *Tian* is thus defined as the 'day' and the 'skies' under which culture accumulates, rather than as some more disjunctive atemporal and aspatial 'Other', some ontologically different order of Being. Significantly, there is a continuity between the articulation of nature generally, and the inscription of human culture. The nature-nurture dualism familiar in Greek-based culture is not operative; instead, the natural world and human culture are both vigorous and continuous.

A corollary to this notion of an invigorated world is the absence of any final boundary between the sentient and insentient, animate and inanimate, living and lifeless. Since spirituality and life go hand in hand, spirituality, like life, pervades all things.

To say that spirituality and life are all-pervasive has two immediate implications. First, there is no value-neutral nature which can stand as encouragement for the pathetic fallacy. Humanity, instead of constituting a privileged link along a Chain of Being which discriminates on the basis of given ontological disparities, achieves its hierarchical cosmic role through its cultivation of a complex spirituality, where this spirituality is understood in terms of extension, influence, and inclusion. In fact, as is evident from the connotations of the character *shen*, conventionally translated as both 'human spirituality' and 'divinity', and also from entrenched cultural practices such as ancestor worship and the worship of cultural heroes such as Confucius, gods in the Chinese tradition are generally 'extended human beings' (see Confucius). *Tian* itself is the aggregate spirituality generated by a continuous culture.

Tian is 'self-so-ing' (*ziran*). There is no distinction between nature and its power of organization and generation. There is nothing antecedent to it; there is no beginning to it or end of it. *Tian* is also anthropomorphic, suggesting its intimate relationship with the process of euhemerization that grounds Chinese ancestor worship. It is probably this common foundation in ancestor worship that allowed for the conflation of the Shang dynasty's *di* with the notion of *tian* imported with the Zhou tribes. There seems to be sufficient reason to assume that *tian*, like all Chinese gods, is by and large constituted by dead people. Culturally significant human beings - persons such as the Duke of Zhou and Confucius - ascend to become *tian*, and *tian* is itself made determinate in their persons. This is the meaning of the familiar notion, *tianren heyi* (the continuity of *tian* and humanity). Not only does *tian* entail anthropomorphism - gods are man-shaped - it also entails a 'theomorphism' - exemplary persons are god-shaped. Worthiness in the human world defines *tian*. As a narrative constituted by cultural heroes, *tian* thus is genealogical and biographical.

Tian is not only culturally specific, it is also geographical. The discovery of a new and sophisticated culture would

anticipate the discovery of a *tian* representative of that culture. Just as there are many skies, one would expect other cultural traditions to have *tian* of their own.

Finally, *tian* does not speak but communicates effectively, although not always clearly, through oracles, perturbations in the climate and alterations in the natural conditions of the human world. *Tian* participates in a discourse shared by the human community: at least, by the most worthy among them. Given the interrelatedness and interdependency of the orders defining the Chinese world, what affects one affects all. A failure of order in the human world will automatically be reflected in the natural environment. Although *tian* is not a 'personal' deity responsive to individual needs as in the Judeo-Christian world view, as aggregate ancestor it would seem that *tian* functions impartially on behalf of its progeny to maximize the possibilities of emergent harmony at all levels.

See also: Anaxagoras §3; Chinese philosophy; Confucian philosophy, Chinese; Creation and conservation, religious doctrine of; Dao; De; Daoist philosophy; God, concepts of

DAVID L. HALL
ROGER T. AMES

References and further reading

Berthrong, J. (1994) *All Under Heaven: Transforming Paradigms in Confucian-Christian Dialogue*, Albany, NY: State University of New York Press. (A study of the critical issues which arise from the confrontation between modern interpretations of Confucian philosophy and Christian theology.)

Hall, D.L. and Ames, R.T. (1997) *Thinking From the Han: Self, Truth, and Transcendence in Chinese and Western Culture*, Albany, NY: State University of New York Press. (An exploration of several central philosophical issues that illustrate the difference between Chinese and Western assumptions.)

Tibetan philosophy

*Tibetan philosophy - if we can make a rough separation between what is predominantly argument-oriented and analytical and what is more a question of ritual, devotion or vision - is best characterized as a form of scholasticism. It exhibits marked parallels with philosophy in Western medieval contexts, including a heavy emphasis on logic, philosophy of language and metaphysics, all in the service of exegesis of religious doctrine found in root texts. Just as in Western scholasticism, there is a reliance upon scripture, but within that traditional context there is also ample room for rational analysis and synthesis of potentially disparate doctrines, as well as a considerable quantity of argumentation which is a type of 'fine tuning' of Indian issues. Tibetan thinkers explored matters which are often of genuine importance in our understanding of Indian texts. In particular, in Mādhyamika Buddhist philosophy we find an important synthesis of Indian Yogācāra ideas with a relatively natural interpretation of key ideas in the literature on the Buddha-nature (*tathāgatagarbha*); we also find important debates on the nature of the two truths, the status of means of valid cognition (*pramāṇas*), and on questions of philosophical method, such as the possibility or impossibility of Mādhyamikas holding theses and themselves defending positions. Beginning with the Great Debate of bSam-yas (Samyay) in the latter part of the eighth century, we find constantly recurring reflection on questions concerning the nature of spiritual realizations and the role of conceptual and analytic thought in leading to such insights. In the logico-epistemological literature, the hotly debated issues generally centre around the problem of universals, the Indian Buddhist philosophy of language and the theory of the triply characterized logical reason (*trirūpahetu*). In addition, the Tibetans developed an elaborate logic of debate, an indigenous system containing many original elements unknown in or even alien to Indian Buddhist logic.*

1 1 Issues in Mādhyamika philosophy

Tibetan Buddhist thinkers, being generally adherents of some form of the Mādhyamika, had to provide answers to the question as to what was the correct and highest interpretation of Nāgārjuna's thought and what were the erroneous interpretations and traps into which thinkers might fall when dealing with this difficult and often maddeningly ambiguous Indian philosopher (see Nāgārjuna; Buddhism, Mādhyamika: India and Tibet). Thus, for example, the Sa-skya-pa (Sagyaba) thinker Go ram pa bSod nams seng ge (Goramba Sönam sengge, 1429-89), in a text called *lTa ba'i shan'byed (Differentiation of the Views)*, polemically described three types of Mādhyamika philosophy in Tibet: (1) those which took the extreme of permanence to be the middle way; (2) those which took the extreme of annihilation to be the middle way; (3) the Mādhyamika free from extremes. The first two views are those of mistaken interpreters, while the latter is supposedly Go ram pa's own. The dGe-lugs-pa (Gelukba) thinker 'Jam dbyangs bzhad pa (Jamyang shayba, 1648-1722), in his *Grub mtha' chen mo (Great Systems of Tenets)*, speaks of three sorts of erroneous view: (1) the theory of the Chinese monk Hva-shang and certain bKa'-brgyud-pa (Gagyuba) philosophers that nothing exists (*ci yang med*); (2) the voidness-of-what-is-other (*gzhan stong*) theory of the Jo-nang-pa (Jonangba); (3) the view that the Mādhyamika does not accept means of valid cognition (*pramāṇas*), does not accept anything himself on either of the two levels of truth and does not have a system of his own (*rang lugs*).

To take the first view mentioned by Go ram pa (the second one mentioned by 'Jam dbyangs bzhad pa), much of Tibetan thought was indeed strongly influenced by an indigenous version of the Mādhyamika which attempted to integrate Nāgārjuna's thought with Yogācāra and with the principal ideas in Indian texts such as the *Ratnagotravibhāga (Differentiation of the Lineage of the [Three] Jewels)*, an early fifth-century text which notoriously speaks of a permanent (*nitya*), stable (*dhruva*) and eternal (*śāśvata*) Buddha-nature present in sentient beings. This Tibetan synthesis was initially put forward by the Jo-nang-pa school, founded by Dol bu pa Shes rab rgyal mtshan (Dolbuba Shayrap gyeltsen, 1292-1361); for reasons that will become clear, the Jo-nang-pas and their successors came to be known as gZhan-stong-pas (Shendongbas), 'those who accept voidness-of-what-is-other'. The gZhan-stong-pa position had adherents in three of the four major traditions of Tibet - namely the Sa-skya, the bKa'-brgyud and the rNying-ma (Nyingma) traditions - and was championed by such major Tibetan thinkers as gSer mdog Paṇ chen Śākya mchog ldan (Serdok Paṇchen Shākya chokden, 1428-1507), Tāranātha (b. 1575) and many others (see Mi bskyod rdo rje); in the nineteenth century, with the 'rediscovery' of the Jo-nang-pa texts, it became the philosophical underpinning of the 'nonsectarian movement' (*ris med*), founded by the syncretic thinker 'Jam mgon kong sprul (Jamgön gongtrul, 1813-99). Only the

dGe-lugs-pa school rejected it completely and unanimously, the Jo-nang-pas being for them a sort of *bête noire* both politically and philosophically, against which Tsong kha pa Blo bzang grags pa (Dzongkaba Losang dragba, 1357-1419) and his school repeatedly directed their ire.

The Jo-nang-pas and other gZhan-stong-pas relied on an adaptation of the Yogācāra theory of voidness that one finds in the *Tattvārtha* chapter of Asaōga's *Bodhisattvabhūmi*. This they combined with the three-nature theory found in the *Saṃdhinirmocanasūtra(Discourse on the Interpretation of the [Buddha's] Thought*; possibly second-fourth centuries AD), the theory that things have falsely imagined natures (*parikalpitasvabhāva*) invented by thought and language, causally dependent natures (*paratantrasvabhāva*) and an absolute or a perfected nature (*parinispannasvabhāva*) (see Buddhism, Yogācāra school of). The combination of Yogācāra with the *Ratnagotravibhāga* yields the key Tibetan distinction between 'void-of-oneself' (*rang stong*) and 'void-of-what-is-other' (*gzhan stong*) that has inspired enormous debates in Tibetan Mādhyamika up to the present day. In brief, the fundamental gZhan-stong-pa ideas go like this for a Jo-nang-pa: the Absolute, *parinispannasvabhāva*, whose existence enables us to avoid the nihilistic view that everything is just a complete illusion, is only void of the imagined and dependent natures: it is void of what is other than it, but is not void of itself. The imagined and dependent natures, on the other hand, are nonexistent and are void of themselves. This key stance is then fleshed out to include most of the other major currents in Indian Mādhyamika, Yogācāra, Tathāgatagarbha and even Tantric systems: for example, Nāgārjuna's arguments in the *Mūlamadhyamakakārikā* show only that conventional truths (*saṃvṛtisatya*) are void of themselves; the Absolute is an existent, truly established gnosis (*ye shes*); as in Yogācāra thought, this gnosis admits of no distinction between subject (*grāhaka*) and object (*grāhya*) and is suchness (*tathatā*) and the *bhūtakoṭi* ('limit of the real'); it is identifiable with the Buddha-nature spoken of in the *Ratnagotravibhāga*; Tantric principles such as the union of voidness and bliss (*bde stong zung 'jug*) are said to be inexplicable without this version of voidness.

Not surprisingly, the Jo-nang-pas were often criticized, especially by the dGe-lugs-pas, but also by Sa-skya-pas such as Go ram pa, as reifying the Absolute and thus transforming Buddhism into a substantialist philosophy. This is what was behind Go ram pa's charge that this school 'took the extreme of permanence as the Middle Way'. (The Mongolian dGe-lugs-pa writer Thu'u bkwan Chos kyi nyi ma (Tugen Chögyi nyima, 1737-1802) went so far as to say that the Jo-nang-pas were in effect like partisans of a type of Brahmanism.) The Jo-nang-pas thus supposedly went badly astray from Indian Mādhyamika by adopting positive descriptions which hypostasized a permanent Absolute, although, in all fairness, it has to be said that this criticism largely depends on which Indian texts one emphasizes and what literature one takes as authoritative. It can be intelligently argued in defence of the Jo-nang-pas that there were Indian Mādhyamika texts, like the hymns attributed to Nāgārjuna, which did exhibit a positive, cataphatic approach not far from that of the *Ratnagotravibhāga*, and that Indian Mādhyamika did not consist exclusively in the negative apophatic dialectic or the insistence upon dependent origination (*pratītyasamutpāda*) that one finds in Nāgārjuna's *Mūlamadhyamakakārikā*.

2 2 Subschools of Mādhyamika

Tsong kha pa and his followers, the dGa'-ldan-pa (Gandenba) and then the dGe-lugs-pa school, reject this all-encompassing syncretism, arguing that Yogācāra and Mādhyamika are quite distinct schools, and that Tantra and the Tathāgatagarbha theories are to be interpreted along the lines of Nāgārjuna, and especially along the lines of the branch of Mādhyamika founded by Candrakīrti, the so-called Čast;Prāsaōgika or Thal-'gyur-ba (Talgyurwa) school. Whereas the Čast;Svātantrika or Rang-rgyud-pa (Ranggyuba) philosophy of Indian Mādhyamika, attributed by Tibetans to Śāntarakṣita, Bhāvaviveka *et al.*, had been dominant in the earlier diffusion of Buddhism in Tibet (starting in the eighth century) and had been the philosophy of the first traditions in the beginning of the second diffusion (from the eleventh century on), Tsong kha pa came squarely down on the side of the Čast;Prāsaōgika Mādhyamika, a school which owes much of its influence in Tibet to Pa tshab Nyi ma grags (Batsap Nyimadrak, b. 1054/55), who was the translator of Candrakīrti's *Madhyamakāvatāra (Introduction to the Middle Way)* and other works. Note that the terms Čast;Prāsaōgika (*thal 'gyur ba*) and Čast;Svātantrika (*rang rgyud pa*), although frequently used in the modern secondary literature, are not clearly attested in Indian texts. This terminology to distinguish between Mādhyamika schools is Tibetan in origin and is especially found in a genre of literature which had great importance in Tibet, the hierarchical presentation of the 'systems of tenets' (*siddhānta*; *grub mtha'*) of the four Indian Buddhist schools and their various subschools.

There were numerous theories among Tibetans of the eleventh to fifteenth centuries about what the difference between Čast;Svātantrika and Čast;Prāsaōgika actually came down to - indeed, the matter took on such importance that the question of how to construe the difference became one of the key issues in Tsong kha pa’s thought, just as it was a key issue for his predecessors and for many of his successors. Some Tibetans had conceived of the difference between Čast;Svātantrika and Čast;Prāsaōgika solely in terms of the logical forms which the two schools employed (namely, objectively valid reasons versus mere *reductio ad absurdum*); others focused on the fact that Čast;Prāsaōgika rejected means of valid cognition (*pramāṇas*) while Čast;Svātantrika did not, or on the fact that Čast;Prāsaōgikas did not have any theses (*pakṣa*) whatsoever of their own, not even on conventional matters. (Compare ’Jam dbyangs bzhad pa’s third type of erroneous Mādhyamika.) Reacting to the views of these opponents, Tsong kha pa formulated the difference as turning, above all, on a difference in ontology, in particular on the acceptance or rejection of things being in some way established with/through their own intrinsic natures (*rang bzhin*; *svabhāva*). *Svabhāva*, a key notion that Nāgārjuna and his commentators argue against, figures throughout the Tibetan presentations of the systems of Buddhist schools in the *grub mtha’* literature, but with each school adopting an ever subtler position on the question - it is only, according to Tsong kha pa, the Čast;Prāsaōgika branch of the Mādhyamika school which rejects it completely and unconditionally; the Čast;Svātantrikas still hold on to a kind of rarefied version of *svabhāva*, albeit only on the level of conventional truth. The basic argumentation for this attribution turns on the Čast;Svātantrikas’ own construal of the Yogācāra three-nature theory.

The question as to precisely what a Mādhyamika denies became a major issue in Tibet. A dominant trend in Tibetan Mādhyamika thought was that promulgated by Pa tshab’s disciple Zhang Thang sag pa (Shang tangsakba), Go ram pa bSod nams seng ge and others, a position which came to be known as ‘the Mādhyamika free from extremes’ (*mtha’ bral dbu ma*) and which was often depicted as ‘the view of neither being nor nonbeing’ (*yod min med min gyi lta ba*). This view advocated using Mādhyamika-style arguments to show the incoherence of any and all things that one might analyse, with the result that no predication of being, nonbeing, both being and nonbeing, or neither being nor nonbeing is rationally justified. We thus have, on this interpretation of the Mādhyamika, a literal and unqualified negation of all members of the ‘tetralemma’ (*catuṣkoṭi*), with the purpose of thoroughly overcoming the possibility of any conceptualizations and discursive thought. This ‘Mādhyamika free from extremes’, it should be remarked, is a pretty fair characterization of the usual Indian position, and corresponds quite well to the way in which certain important Western authors would characterize Indian Mādhyamika. (In this sense, S. Matsumoto (1990) is certainly right to stress just how remarkable and unique Tsong kha pa was in the history of Indo-Tibetan Mādhyamika in *opposing* the fundamental position of the ‘Mādhyamika free from extremes’.)

The question does, however, arise as to whether this version of the fourfold negation leads to a deviant logic, as negating nonbeing would seem to imply being, and negating the last lemma would seem to imply being or nonbeing. Tsong kha pa and his school avoid these types of unwelcome consequences by adding the modal qualifiers ‘truly’ (*bden par*) and ‘conventionally’ (*tha snyad du*) to the tetralemma, so that, for example, the first two negations end up being that things are not truly existent and not conventionally nonexistent. The tactic obviously avoids violations of logical laws. Equally, it allows the Mādhyamikas to make a clear separation between the conventional and the ultimate, so that they can argue that a thing like a pot exists (conventionally) and is established by means of valid cognition. We can thus easily hold the thesis that the pot exists, but where the Mādhyamika philosophy comes in is to show that it does not *truly* exist. Indeed, the advocates of the ‘Mādhyamika free from extremes’, as well as the gZhan-stong-pas, do have serious problems in accounting for conventional truths: to say, for example, that such truths merely exist for mistaken minds (*blo ’khrul ba’i ngor yod pa*) and are not established by means of valid cognition (*pramāṇas*) comes perilously close to just falling into the trap of subjectivism, where one says that things exist or have such-and-such properties when one believes that they do. (Quite a bit more than *that* is needed to explain how we can discover new things and how we can show that things which many people believe in at some particular time do not exist and have never existed even conventionally.) The issue of the acceptance or rejection of *pramāṇas* establishing conventional truths is developed at length from the fifteenth century onwards, starting with a critique of Tsong kha pa on this and other points by the Sa-skya-pa thinker sTag tshang lōtsa ba (Daktsang lodzawa, b. 1405).

It should be stressed that the logical simplicity and lucidity of Tsong kha pa’s version of Mādhyamika comes at a

high exegetical and philosophical price: not only is he constrained to add words everywhere to Indian texts, but, what is perhaps more serious, he can well be accused of inventing a factitious notion, ‘true existence’, which he then refutes as incoherent and contradictory, leaving the ordinary world intact and even reified. This was a common criticism of Tsong kha pa. Arguably, as his critics maintained (be they advocates of the ‘Mādhyamika free from extremes’ or gZhan-stong-pas), Nāgārjuna and his commentators were showing the utter incoherence of ordinary things themselves, and were leaving nothing whatsoever unassailed and intact among these conventional entities.

3 3 The role of conceptual understanding in spiritual realization

Tibetan thought on the nature of meditation and spiritual realization seems to be divided into two camps, both stemming (in very complex ways) from the Great Debate of bSam-yas (Samyay) in the latter part of the eighth century. This debate was a (year-long?) series of exchanges which are reputed to have opposed the ‘Gradualist’ (*rim gyis pa*) Indian Buddhism represented by Kamalaśīla and others to the Chinese ‘Subitists/Simultaneists’ (*cig car ba*), who were represented by Hva-shang Mahāyāna, a monk from Dunhuang holding Chan positions. The historical details of the Great Debate (which is with few exceptions depicted in Tibetan sources as having been won by the Gradualists) cannot be developed here, except to say that it now seems clear that several ‘heretical’ positions of Chinese, Tibetan and even Indian origin were under attack from the Kamalaśīla side.

In fact, as D. Seyfort Ruegg (1992) has insightfully remarked, the debate became a partly dehistoricized topos in the Tibetan representation of the history of their own philosophy. The bSam-yas debate and its traces were perceived to recur constantly in the form of an opposition between two models of the Buddhist path, namely mediacy versus immediacy - the path taken as a progression, where philosophical analysis and various moral practices lead step by step to awakening versus the path which is not a series of intermediate steps, but is fundamentally identical with the fruit to be attained. This latter ‘path’ is one where realization comes from a mode of awareness that is characterized by no mentation (*amanasikāra*; *ci yang yid la mi byed pa*), an awareness that consists in seeing or coming face to face (*ngo ’phrod*) with the nature of the mind (*sems nyid*) without reliance upon any concepts whatsoever - this realization is all we need to become enlightened at once. To take an important example of the recurrence of this debate, the thirteenth century’s greatest thinker, Sa skya Paṇḍita (Sagya Paṇḍita, 1182-1251), makes a rapprochement between the Hva-shang side and the Tibetan tradition of Great Perfection (*rdzogs chen*) by criticizing the ‘Self-sufficient White Remedy’ (*dkar po chig thub*) doctrine of the bKa’-brgyud-pa scholars sGam-po-pa (Gamboba, 1079-1153) and Zhang-Tshal-pa (Shangtselba, 1123-93) as being a revival of the ‘Great Perfection of the Chinese tradition’ (*rgya nag lugs kyi rdzogs chen*). (Compare ’Jam dbyangs bzhad pa’s first type of erroneous Mādhyamika.) Another example: for Tsong kha pa, Hva-shang becomes an exponent of a mere blank-mindedness, a kind of stupefaction where one supposedly rests virtually unconscious. This Hva-shang-style blank-mindedness is then taken to be what will ensue if, like certain ‘Mādhyamikas free from extremes’ who deny ‘too much’, one rejects all entities and all predication whatsoever and hence fails to make the proper distinction between so-called truly established things and conventional things.

What was the issue? Did the Hva-shang position lead to nothing but voluntary stupefaction, as Tsong kha pa and even some contemporary Western writers depict it? To use Ruegg’s phrase, the Great Debate involved ‘lattices’ of related ideas, and the lattice of ideas associated with the partly dehistoricized version of the Hva-shang position involves, very noticeably, the idea of an innate nature which is in some sense already perfect and actual. Indeed, in a twelfth-century account of the Great Debate, the *Chos ’byung me tog snying po*, and in the sixteenth-century historical work the *mKhas pa’i dga’ ston (Feast for Scholars*; Minzu chubanshe edition, 1986: 389; see also Ruegg 1992: 73, 86, note 164), one of the ways in which the issue is formulated is whether sentient beings are ‘Buddhas from the beginning’ (*dang po nas sangs rgyas*) - Hva-shang’s side supposedly advocated a kind of innate version of the Buddha-nature (*tathāgatagarbha*) existing fully *ab initio*, not just as a germ or potentiality for enlightenment, but as already accomplished. (Compare, for example, the view found in the nine sorts of Great Perfection teachings described in an early work, the *bSam gtan mig sgron (Lamp for the Eye of Meditative Absorption)* of gNubs Sangs rgyas ye shes (Nup Sanggyay yayshay, tenth century): ‘The nature of the Buddha, sentient beings and their objects is to be without exception enlightened in the great state of the spontaneous *dharmatā*, which is without beginning nor end’ (1974: 320; see also Karmay 1988: 114).) It is not surprising, if one holds that there is a type of innate perfected nature, that the path should consist essentially in stopping thought and seeing this nature. It is also probably not very surprising that focusing on good actions along the path should

seem to Subitists such as Hva-shang and others a significant distraction. Non-mentation would not be a stupefaction, where the mind is reduced to virtual insentience, but would be more like a re-cognition or ‘anagnosis’ (to use the phrase of Giuseppe Tucci (1980: 13)) of what was already there.

Kamalaśīla’s side and its later followers of course panned this Buddhahood *ab initio* as an utter absurdity. During the Debate, for example, Ye-shes dbang-po (Yayshay wangbo) supposedly argued: ‘If you accede simultaneously [to enlightenment] then why are you still doing anything; if you are a *buddha* from the beginning (*dang po nas sangs rgyas*), what then is wrong?’ (*mKhas pa’i dga’ ston*, 1986: 389). And elsewhere on the same page in the *mKhas pa’i dga’ ston*, he is represented as criticizing the idea of ‘being enlightened without having done anything’ (*ci yang ma byas par ’tshang rgya ba*). All this is a kind of Gradualist common sense, but it is also somewhat unimaginative, especially so in that a number of clear-headed Tibetan and Chan philosophers devoted a lot of deep thought to this difficult subject and seem to have been quite aware of the gross absurdities into which one should not fall.

Finally, it is hard to agree with Samten Karmay (1988: 87-8) that the Subitists and Gradualists both accepted an identical version of the ‘spiritual basis’ (*gzhi*), and that therefore the debate turned only on questions of method. Karmay argues this because of ‘correspondences’ that he sees between various key terms on each side, although these correspondences are far from clear; even when the same word is used on the Gradualist and Subitist sides (for example, *tathāgatagarbha*), it is unlikely that the concept is being taken in altogether the same way. In short, the idea of some sort of innate perfected nature is a crucial issue in the Great Debate; we are not dealing just with methods. Indeed, if there is only method involved, and no question of Buddhahood *ab initio* being re-cognized, the general Subitist position becomes an almost grotesquely unintelligent method, and the criticisms by Kamalaśīla, Tsong kha pa and others about meditation without analysis yielding a rather useless mental *tabula rasa* become difficult to avoid.

4 4 Epistemology, logic and philosophy of language

The major developments by Tibetans in this area are to be found in the interaction of two currents of interpretation of Dignāga and Dharmakīrti’s thought (see Dignāga; Dharmakīrti). The first current has its origin largely in gSang-phu Ne’u-thog monastery (Sangpu Nay-utok, founded in 1073), in the works of rNgog lōtsa ba Blo ldan shes rab (Ngok lodzawa Loden shayrap, 1059-1109), Phya pa Chos kyi seng ge (Chaba Chögyi sengge, 1109-69) and their disciples. Subsequently, Phya pa’s *Tshad ma’i bsdus pa (Summaries of Epistemology and Logic)* became the groundwork for the ‘Collected Topics’, or *bsdus grwa* (düra) literature, which furnished so much of the epistemology of the dGe-lugs school. From what we can glean of Phya pa’s thought from other works in his tradition and the works of his adversaries - his own works have not survived - he combined a creative intellect with what was probably a quite lacunary knowledge of the main texts of the Indian Buddhist epistemological school: the result is not unlike what we find in certain early Chinese Buddhists, namely numerous insightful interpretations and new ideas, and, at the same time, many notions which were no doubt invented and had no actual basis in Indian texts.

The gSang-phu positions were strongly criticized by Sa skya Paṇḍita, the Tibetan who probably came closest to faithfully representing the positions of later Indian epistemology and logic. In his *Rigs gter (The Treasure of Reasoning)*, we find a repeated trenchant critique of ‘Tibetans’ or ‘Tibetans who pride themselves on being logicians’ - polemical shorthand for the gSang-phu philosophers. None the less, gSang-phu thought had a persistent appeal due to what must have been its extremely seductive appearance of subtlety and rigour - indeed, as D. Jackson (1987: 137-8) points out, Śākya mchog ldan reported that in his time, in the fifteenth century, people often felt that the ‘Summaries’ were subtle and proven correct, while the *Rigs gter* was ‘extremely rough’.

Central to the gSang-phu-ba versus Rigs-gter-ba debates was the problem of universals (see Universals, Indian theories of §3). The gSang-phu side (including the dGe-lugs) maintained a type of realism where at least some universals were thought to be real entities (*dngos po*), while Sa skya Paṇḍita himself and many of his Rigs-gter-ba followers, such as Go ram pa and Śākya mchog ldan, argued for the more usual Dharmakīrtian *apoha* theory, maintaining the unreality of all universals, indeed of all objects of conceptual thought. (The *apoha* theory of language is the Buddhist explanation of the unreality of all concepts, treating a concept of *A* as a mind-invented exclusion of non-*A* (see Nominalism, Buddhist doctrine of).) Indian Buddhism’s philosophy of language was thus considerably reinterpreted by the Phya-pa school, especially by the dGe-lugs-pa. The latter would find textual

sources in later writers like Śaōkaranandana, but while certain Indian Buddhists (for example, Dharmottara, Mokṣākaragupta and perhaps Śaōkaranandana) were probably tending towards some form of realism on the question of universals, it is very difficult to see how the Tibetans' rationale for realism had much, if anything, to do with the reasonings of these Indian Buddhists. The Phya-pa school's interpretation of the Dignāgan-Dharmakīrtian version of concepts-*qua*-exclusion (*ldog pa*; *vyāvṛtti*), which they also applied to universals, logical reasons and all other mind-invented constructs, was to say that 'exclusion itself' or 'universal itself' (*spyi kho rang*) is unreal, but not all exclusions, universals, and so on, are unreal. The point probably partly turns on certain features of the Tibetan language which make this type of explanation possible. This and closely related moves, which the dGe-lugs-pas would maintain to be at the heart of 'the difficult point of the *apoha* theory' (*gzhan sel dka' gnad*), had already been criticized by Sa skya Paṇḍita himself and would be devastatingly attacked by Śākya mchog ldan.

There were also numerous Tibetan developments in epistemology which genuinely contributed to clarifying key Dignāgan and Dharmakīrtian issues: for example, the Tibetan debates on the theory of truth in Indian Buddhism, on the notion of nonperception (*anupalabdhi*), and on the use of scriptural arguments, and Tibetan positions on the question of how to interpret Dharmakīrti's proofs of the Buddha's authority. We find gSang-phu-bas and Rigs-gter-bas offering (conflicting) explanations of Dignāga's *Hetucakra (The Wheel of Logical Reasons)*, in particular of the so-called 'inconclusive reason which is too exclusive' (*asādhāraṇānaikāntikahetu*), which is by far the most difficult fallacy to understand in Dignāga's system of nine reasons. We also find the gSang-phu traditions, including the dGe-lugs, accepting many of the key elements of Ratnākaraśānti's idea of 'intrinsic entailment' (*antarvyāpti*) as presented in his *Antarvyāptisamarthana (The Justification of Intrinsic Entailment)*, even though these same thinkers still did not go so far as to accept the Antarvyāptivāda position that examples are dispensable in argumentation, and are only used for dullards (*rmongs pa*). There were actually very few Tibetans who considered themselves real Antarvyāptivādins, the major possible exception being the Sa-skya-pa thinker Nya dbon Kun dga' dpal (Nyawön Güngapel, *c*.1300-80), who is said to have espoused this view in his now lost commentary on Dharmakīrti's *Pramāṇavārttika (Commentary on Valid Cognition)*.

The Phya-pa tradition, largely independently, developed a remarkably sophisticated type of debating logic, with consequences (*thal 'gyur*; *prasaōga*) replacing the usual triply-characterized reasons (*trirūpahetu*) and inferences-for-others (*parārthānumāna*) of Indian Buddhist argumentation. This logic, which one finds elaborately developed in *bsdus grwa* texts, has a unique type of quantification and employs variables. The Dharmakīrtian idea of the general entailment (*vyāpti*, literally 'pervasion') between the reason and what is to be proved had presented enormous philosophical problems in Indian logic, as it was very difficult to say precisely how one was to establish or know for certain when there was *vyāpti* - *vyāpti* had to be assured by ascertaining a necessary connection (*sambandha*) between terms (see Inference, Indian theories of §§5-6). In this debating logic, however, the epistemological questions as to how we know that there really is *vyāpti*, or that there can be no counterexamples, are of curiously minor importance; they are, in any case, separated from the purely logical matter of the truth conditions for a universally quantified statement, that is, absence of counterexamples. This is a type of progress, freeing the logical and formal matters around *vyāpti* from the probably unsolvable epistemological issues which preoccupied the Indians. Interestingly enough, the Phya-pa school and its successors, in their classifications of various types of consequences as having twelve different types of *vyāpti*, went into considerable detail on the logical connections between statements in a way which was quite unknown in India. Moreover, we see here and elsewhere a marked turn towards a formal perspective: saying, for example, that such-and-such a consequence-statement will satisfy a certain kind of *vyāpti* if it satisfies another kind is indeed an abstraction away from content to matters of form. It is, however, difficult to argue that the rules and approach of this logic represent a propositional or term logic of the sort developed by Stoic or Aristotelian philosophers, although it is certainly not opposed to classical theorems.

See also: Buddhist concept of emptiness; Buddhist philosophy, Indian; Buddhist philosophy, Chinese; mKhas grub dge legs dpal bzang po; rGyal tshab dar ma rin chen

TOM J.F. TILLEMANS

References and further reading

Cabezón, J. (1994) *Buddhism and Language: A Study of Indo-Tibetan Scholasticism*, Albany, NY: State University of New York Press.(A relatively nontechnical introduction to epistemological issues.)

dPa' bo gtsug lag phreng ba (16th century) *mKhas pa'i dga' ston (Feast for Scholars)*, ed. rDo rje rGyal po, Beijing: Minzu chubanshe, 1986.(Cited in §3; the full title is *Dam pa'i chos kyi 'khor lo bsgyur ba rnams kyi byung ba gsal bar byed pa mkhas pa'i dga' ston*.)

Dreyfus, G. (1997) *Recognizing Reality. Dharmakīrti's Philosophy and its Tibetan Interpretations*, Albany, NY: State University of New York Press.(Detailed discussion of Sa-skya-pa versus dGe-lugs-pa debates on epistemology.)

gNubs chen Sangs rgyas ye shes (10th century) *bSam gtan mig sgron (Lamp for the Eye of Meditative Absorption)*, Smanrtsis Sherig Spendzod Series 74, Leh, India: N. Topgyal, 1974.(Cited in §3; excerpts are discussed in Karmay 1988. The full title is *sGom gyi gnad gsal bar phye ba bsam gtan mig sgron*.)

Gomez, L.O. (1983) 'Indian Materials on the Doctrine of Sudden Enlightenment', in W. Lai and L.R. Lancaster (eds) *Early Ch'an in China and Tibet*, Berkeley, CA: University of California Press, 393-434.(The bSam-yas debate and the role of Indian 'Subitists' such as Vimalamitra.)

Jackson, D.P. (1987) *The Entrance Gate for the Wise*, Vienna: Arbeitskreis für Tibetische und Buddhistische Studien.(Detailed information on Sa skya Paṇḍita and his tradition.)

Jackson, D.P. (1994) *Enlightenment by a Single Means: Tibetan Controversies on The 'Self-Sufficient White Remedy' (dkar po chig thub)*, Vienna: Österreichischen Akademie der Wissenschaften.(Expands upon the discussion in §3 about the continuation of themes from the Great Debate into certain bKa'-brgyud-pa schools.)

Karmay, Samten Gyaltsen (1988) *The Great Perfection: A Philosophical and Meditative Training of Tibetan Buddhism*, Leiden: Brill.(A reference work on the history and principal doctrines of the Great Perfection school and on issues concerning the Subitist school in the Great Debate.)

Kuijp, L. van der (1983) *Contributions to the Development of Tibetan Buddhist Epistemology from the Eleventh to the Thirteenth Century*, Wiesbaden: Franz Steiner. (A reference work on the subject.)

Matsumoto, S. (1990) 'The Mādhyamika Philosophy of Tsong kha pa', *Memoirs of the Research Department of the Toyo Bunko* 48: 17-47.(A vigorous argument for the importance and uniqueness of Tsong kha pa in Indo-Tibetan Mādhyamika thought.)

Mimaki, K. (1982) *Blo gsal grub mtha': Chapitres IX (Vaibhāṣika) et XI (Yogācāra) édités et Chapitre XII (Mādhyamika) édité et traduit*, Kyoto: Zinbun Kagaku Kenkyusyo.(A reference work on Tibetan philosophies and the grub mtha' literature.)

Onoda, S. (1992) *Monastic Debate in Tibet: A Study on the History and Structures of bsDus grwa logic*, Vienna: Arbeitskreis für Tibetische und Buddhistische Studien.(A reference work on the Phya-pa tradition and its successors. See page 98- on the twelve sorts of *vyāpti* mentioned in §4.)

Ruegg, D. Seyfort (1963) 'The Jo naō pas: A School of Buddhist Ontologists according to the Grub mtha' śel gyi me loō', *Journal of the American Oriental Society* 83: 73-91.(Expands upon §1; discusses the Jo-nang-pa position and its critique by the dGe-lugs-pa.)

Ruegg, D. Seyfort (1992) *Buddha-nature, Mind and the Problem of Gradualism in a Comparative Perspective: On the Transmission and Reception of Buddhism in India and Tibet*, London: School of Oriental and African Studies.(An important study of the bSam-yas debate and its philosophical issues.)

Sa skya Paṇḍita Kun dga' rgyal mtshan (1182-1251) *Rigs gter (Treasure of Reasoning)*, in *Sa skya pa'i bKa' 'bum, The Complete Works of the Great Masters of the Sa skya Sect of Tibetan Buddhism*, compiled by bSod nams rgya mtsho, vol. 5, Tokyo: The Toyo Bunko, 1968.(The full title is *Tshad ma rigs pa'i gter*. The text is being edited and translated into Japanese in a series of monographs by Y. Fukuda *et al*. as *Chibetto ronrigakukenkyū(Studies in Tibetan Logic)*, Tokyo: The Toyo Bunko.)

Tillemans, T. (1989) 'Formal and Semantic Aspects of Tibetan Buddhist Debate Logic', *Journal of Indian Philosophy* 17: 265-97.(Expands upon §2 and §4 of this entry.)

Tillemans, T. (1990) 'On Sapakṣa', *Journal of Indian Philosophy* 18: 53-79.(Expands upon §4 of this entry.)

Tillemans, T. (1995) 'On the So-called Difficult Point of the *Apoha* Theory', *Asiatische Studien-Études Asiatiques* 49(4): 853-89.(Expands upon §4 of this entry.)

Tucci, G. (1980) *The Religions of Tibet*, trans. G. Samuel, London: Routledge. (Tucci's work remains one of the best introductions to Tibetan religions.)

Williams, P. (1992) 'Non-Conceptuality, Critical Reasoning and Religious Experience: Some Tibetan Buddhist Discussions', in M. McGhee (ed.) *Philosophy, Religion, and the Spiritual Life*, Cambridge: Cambridge University Press, 189-210.(A strong philosophical defence of the Gradualist position and of the dGe-lugs critique of Hva-shang's 'blank-minded' meditation.)

Yoshimizu, C. (1993) 'The Madhyamaka Theories Regarded as False by the Dge lugs pas', *Wiener Zeitschrift für die Kunde Südasiens* 37: 201-27.(On some of the historical issues discussed in §§1-2.)

Tillich, Paul (1886-1965)

Tillich was one of the most influential Christian theologians of the twentieth century. Notable for his effort to translate the language of the Western religious tradition into terms comprehensible to modernity, he drew upon various secular philosophies, including Marxism, existentialism and psychoanalysis, as well as literature and the arts. In his view, these contemporary secular expressions contain the questions which theology must address. He was sometimes criticized for losing one or another aspect of Christian orthodoxy, but more often praised for making it possible to be both Christian and modern. He fled Germany in 1933, in the early days of Nazism. As an expatriated German who became an American citizen, Tillich came to understand his life as one standing 'on the boundary'. He saw himself as an interpreter, occupying the boundary between the Old World and the New, between philosophy and theology, between religious orthodoxy and humanistic secularity, and between university and church.

Tillich was an intellectual who achieved widespread popular acclaim. Even though his lectures and publications were strewn with allusions to obscure thinkers, he gained a substantial following and was frequently quoted in the popular press. His courses were immensely popular, and his sermons - delivered mostly in college chapels - met with great public approbation.

Two of Tillich's themes stand out as most influential. First was his advocacy of a broadened category of the religious. By defining religion as a person's 'ultimate concern', he was able to maintain that virtually everyone has some religious commitment. Through this conceptualization it became possible to view such twentieth-century ideologies as Nazism and Communism - as well as Americanism - as in significant respects religious perspectives. This broadened definition of religion gained wide acceptance, with sociopolitical and even judicial implications. (The US Supreme Court's definition of conscientious objection was influenced by Tillich's formulation.) Second, Tillich's persistent claim that all language about God is symbolic had great impact. Objecting to views of religious language as merely *symbolic, he contended that efforts to be literal in one's talk of God are seriously deficient. By the same token, he argued that the mythic quality of religious narratives cannot be removed without detriment. Finding much American religion strongly literalistic, Tillich persistently argued the contrary view.*

1 Life

Paul Tillich was born in the village of Starzeddel in eastern Germany (now in Poland), the son of a minister and high official of the Evangelical Church of Prussia. After studying philosophy and theology, receiving advanced degrees at Breslau (1910) and Halle (1912), he entered the ministry and became a chaplain in the German army at the outbreak of World War I. At the conclusion of the war - which was at once traumatic and transformative for him - he began an academic career, becoming a professor of philosophy and theology, first at Berlin and subsequently at Marburg, Dresden and Frankfurt. In the late 1920s he participated in a religio-political movement called 'religious socialism', which attempted to offer an ideological alternative to traditionalistic and pro-capitalist views on the one hand and fascist and communist views on the other. Tillich's *The Socialist Decision* (1933) derives from this period and is thought to be one of the first theological essays to analyse the dangers of the Nazi movement. During this period he also argued that German Christians ought to align themselves with the Jews of Germany rather than succumbing to the prevailing anti-Semitism. With Hitler's accession to power in 1933, Tillich was dismissed from his position at the University of Frankfurt; he was among the first to be removed from his state university appointment. Soon thereafter he sought refuge in the USA.

Gradually mastering the English language, Tillich spent his American career as a professor at three influential educational institutions: Union Theological Seminary at Columbia University (1933-55), Harvard University (1955-62) and the University of Chicago (1962-5). Although his involvements were therefore predominantly academic rather than churchly or political, he became widely influential as an interpreter of American culture and its religious dimension (termed by Tillich 'the depth dimension'). The English-speaking public became familiar with his earlier cultural analyses, which were frequently informed by Marxian categories, through the translated essays collected in *The Protestant Era* (1948); during his years in the USA, he produced the essays found in *Theology of Culture* (1959). Tillich's increasing engagement with psychology, principally Freudian, and his increasing use of existentialist categories were reflected in his widely read work *The Courage To Be* (1952), while

his distinctive combination of ontology and ethics was displayed in *Love, Power and Justice* (1954). The most significant of his English language publications was the three-volume *Systematic Theology* (1951, 1957, 1963), the first volume of which contains many of his most penetrating contributions to philosophy and theology.

In 1960, Tillich journeyed to Japan, where he engaged in dialogues with Zen masters and other religious leaders. The trip represented the culmination of a shift in his interests from dialogue with Western secularity to dialogue with non-Western religions, a shift reflected in the slender but important volume *Christianity and the Encounter of the World Religions* (1963), and in his last public lecture, ‘The Significance of the History of Religions for the Systematic Theologian’, found in the posthumous collection of essays *The Future of Religions* (1966).

2 Epistemology and ontology

Like his nineteenth-century mentors, the German idealist philosophers Fichte, Schelling and Hegel, Tillich sought to discover a new path to ontology. Following the lead of his contemporary, the existentialist philosopher Martin Heidegger (§2), he concluded that ‘the way to ontology passes through the doctrine of man’. Reaffirming the Renaissance notion of the human as microcosmic, Tillich asserted that the structure of being can be discovered in and through us in the polarity of the ‘self-world correlation’. The structures of finitude (as well as its destructive tensions) can be analysed philosophically, but finitude is not self-sufficient. Finitude has to ask about the possibility of a self-caused ground of being, beyond finitude. Such a ground can be conceived as ultimate reality or the unconditioned, but proofs for its existence fail. Indeed, argued Tillich, existence belongs to finitude and suggests limitation. If the unconditioned is real, it lies beyond the self-world correlation and is not a being which exists as an object for us as subjects. Nevertheless, ontology shares the existential anguish: it longs for an ultimate in being and meaning.

3 Revelation and God

Tillich followed Aquinas in asserting the equivalence of the ground of being and God. The modern ontological quest replaces static with dynamic categories; power and life replace being and substance as the fundamental ontological concepts. What is sought is an unambiguous power of being, an eternal life. Revelation - the ecstasy of reason - fulfils this quest. In Tillich’s view, the world’s religions are based upon revelation. All revelatory religious language about God, however, is symbolic; religious terminology, always experiential and drawn from finitude, can only point beyond the finite towards the transcendent reality (see Religious language §3). Taken literally, it becomes idolatrous. ‘God’ is a symbol for the ultimate reality. Primary among the symbols constituting the theistic God is the personal symbol. For Tillich, this is an essential symbol because it authenticates the moral relationship of divinity and humanity; moral demand suggests that human beings are separated from what they essentially are and ought to be (thus God is judge). But this symbol must be balanced by a participatory symbol: the unity of spiritual presence (thus God is reconciling love). Morality and mystical experience - interpreted broadly as experiential aspects of individualization and participation - are for Tillich the two principal avenues to knowledge of God. Notable here is the way Tillich’s approach to God is compatible with both fatherly and motherly symbols.

4 Anxiety and faith

In his analysis of human existence Tillich drew much from Kierkegaard and other existentialists, though describing Tillich as an existentialist theologian is misleadingly one-sided (see Existentialist theology §3). Like his one-time colleagues in the Frankfurt Institute for Social Research, he drew upon a wide range of psychological, sociological and philosophical analyses of the human condition, tending, however, to use as a key coordinating concept the Hegelian/existentialist theme of estrangement. In our finite freedom we find ourselves estranged from the power of being; finitude is invariably experienced in us as anxiety. Hypothetically, any one of us could obey the injunction of Jesus to ‘be not anxious’; in fact, however, without the sustenance of some faith we fall into destructive despair.

Religious faith as interpreted by Tillich is a virtually universal phenomenon, which he analysed both as gift (or passive mystical experience) and as risky moral choice (see Faith §2). Religion, first and foremost, is the experience of being grasped by the power of an ultimate concern. All objects of such concern possess not only ultimacy or holiness but also concreteness. To the extent that they involve being grasped by ultimate reality, all faiths offer a kind of compelling self-authentication; but true faith can nevertheless be distinguished from

idolatrous faith by the extent to which the concrete symbolic object of faith is perceived as pointing beyond itself to what is truly ultimate in being and meaning. These self-authenticating aspects of faith are most fully explored in Tillich's book *Dynamics of Faith* (1957), where he shows evidence of his Lutheran heritage.

The more active, willing aspect of faith is the courage of self-affirmation and the affirmation of meaning in the face of threats to our finitude. Tillich identified the principal threats as death, guilt and meaninglessness, all of which may be subsumed under the threat of nonbeing. Both the moral/personal God of theism and the all-encompassing God of mysticism have been offered as sources of the courage to be. In the face of doubts about both of these symbols, Tillich in his most radical arguments, at the end of *The Courage To Be*, proposed absolute faith. Absolute faith may be described as affirmation of the meaningfulness of individual finite existence without the symbolic moral endorsement of the theistic God, but also without a sense of mystical fusion with the divine. Following Nietzsche, Tillich appears to arrive here at a point of the death of concrete symbols for the ultimate and a quest for new symbols.

5 The doctrine of Christ

For Tillich, Christianity's primary symbol is the sacrificial life of Jesus, or the New Testament picture of Jesus as the Christ. Jesus' life, as remembered by the early church, was a pivotal point of history, a central manifestation of Spiritual Presence. Exemplifying fully the New Being - a state beyond estrangement available fragmentarily to the believer - Jesus at the same time consistently pointed beyond himself to the ultimate reality. Consequently, his life and death became the prototypical religious symbol, in relation to which other symbols can be judged.

In this approach to religious symbols, Tillich combined what he called 'Catholic substance' and 'Protestant principle'. Symbols for God are manifold; some are more powerful than others, and around ones of uncommon power, institutions and traditions gather. These symbols and their institutional embodiments grow (and they may die) in the course of historical contingencies. Symbolic content is religious or Catholic substance in the Tillichian lexicon; without it, traditions become sterile and spiritually empty. But claims are sometimes made for religious symbols that they are, or possess significant aspects of, the reality symbolized. Then they become idolatrous and even demonic and must be criticized. Such claims fail to recognize the religious principle that only God is God; only the ultimate reality can claim ultimacy. Tillich called this the prophetic or Protestant principle. This principle of criticism must be applied against all concrete religions, including Christianity itself.

In his later lectures and publications, Tillich advocated dialogue with other religions. This was not with the intent of achieving a superficial eclectic synthesis, but rather with the hope of finding convergences by plumbing the depths of each concrete tradition.

6 Ethics and social philosophy

Tillich linked moral demand and conscience with separation and individualization; he linked love and 'transmoral' conscience (the conscience that accepts forgiveness) with reunion and participation. Indeed, he insisted that love in all its forms is the drive towards the reunion of the separated. He explicitly rejected the sharp opposition of desiring love (*eros*) and self-forgetting love (*agapē*) postulated by some theologians (see Nygren, A.). Distinctions should be made, but within a broader unity; self-forgetting love seeks the good of the other in God, but is no less than desiring love the drive towards reunion. Justice, too, is encompassed in this schematism: it preserves that which is to be reunited. Tillich's high evaluation of desiring love is unique among major twentieth-century theologians. His sensitive reflections on the intricate interweavings of body and spirit will endure as permanent contributions even if he 'failed to reconcile the power of eros with the demands of justice in his private life' (Irwin 1991: 118).

If Tillich's approach to ethics appears much indebted to German idealism, so also does his social philosophy. Early cultures are thought to be characterized by a high degree of communalism, rooted in religion (in the broad sense), and are described by Tillich as archaic 'theonomies'. Modernity brings with it trends towards autonomy, with concomitant rationalization. But modern cultures tend to become increasingly sterile and empty of meaning-giving linkages. Authoritarian heteronomies arise to counteract these threats. Only in a culture in touch with its own religious depth, yet prophetically self-critical - a new theonomy - can autonomous individualization be preserved along with meaningful participation in community. In his later work, *Systematic Theology* (Volume 3) and

Christianity and the Encounter of the World Religions, Tillich interpreted the theme of the kingdom of God as the key Christian symbol for the goal of history. This personalistic, social symbol captures the affirmation of the essential goodness of finite existence, along with a critique of the injustices of the presently existing kingdoms of this world. Although Tillich backed away from explicit commitment to socialism in his American years, he remained a steadfast critic of capitalism and the secular culture of self-sufficient finitude. He was similarly steadfast in his criticism of all forms of utopianism, whether religious or secular; nevertheless, he sought to rehabilitate the spirit of utopia, in the process suggesting a distinction relevant to liberation theology.

GUYTON B. HAMMOND

List of works

Tillich, P. (1959-75) *Gesammelte Werke (Collected Works)*, ed. R. Albrecht, Stuttgart: Evangelisches Verlagswerk, 14 vols.(Bibliography in volume 14; the second edition of volume 14 (Berlin and New York: de Gruyter, 1990) contains an updated bibliography.)

Tillich, P. (1933) *Die Sozialistische Entscheidung*, Potsdam: Protte; trans. F. Sherman, *The Socialist Decision*, New York: Harper & Row, 1977.(An analysis of Germany in the early 1930s and a defence of religious socialism as an alternative to national socialism (Nazism).)

Tillich, P. (1948) *The Protestant Era*, trans. and ed. J.L. Adams, Chicago, IL: University of Chicago Press.(A translation of a number of Tillich's German-language articles.)

Tillich, P. (1951, 1957, 1963) *Systematic Theology*, vols 1, 2 and 3, Chicago, IL: University of Chicago Press. (Tillich's fullest theological statement, employing the method of correlation.)

Tillich, P. (1952) *The Courage to Be*, New Haven, CT: Yale University Press. (Existentialist analysis of anxiety and faith.)

Tillich, P. (1954) *Love, Power and Justice*, New York and London: Oxford University Press.(Tillich's distinctive combination of ontology and ethics.)

Tillich, P. (1957) *Dynamics of Faith*, New York: Harper & Row.(Analysis of what faith is and is not; valuable as an introduction to Tillich's thought.)

Tillich, P. (1959) *Theology of Culture*, ed. R.C. Kimball, New York and London: Oxford University Press. (Important essays from Tillich's later period, including discussions of existentialism and psychoanalysis.)

Tillich, P. (1963) *Christianity and the Encounter of the World Religions*, New York: Columbia University Press. (Tillich in dialogue with non-Western religions, especially Buddhism.)

Tillich, P. (1987-93) *Main Works/Hauptwerke*, gen. ed. C.H. Ratschow, Berlin and New York: Evangelisches Verlagswerk and de Gruyter, 6 vols.(A simultaneously published critical edition of the collected works of Tillich in the original language (German or English).)

References and further reading

Adams, J.L. (1965) *Paul Tillich's Philosophy of Culture, Science, and Religion*, New York: Harper & Row.(An authoritative interpretation by a close associate, covering Tillich's works written before 1945.)

Adams, J.L., Pauck, W. and Shinn, R.L. (eds) (1985) *The Thought of Paul Tillich*, San Francisco, CA: Harper & Row.(Essays on different aspects of Tillich's thought by recognized scholars.)

Carey, J.J. (ed.) (1984) *Theonomy and Autonomy: Studies in Paul Tillich's Engagement with Modern Culture*, Macon, GA: Mercer University Press.(A collection of essays sponsored by the North American Paul Tillich Society.)

Gilkey, L. (1990) *Gilkey on Tillich*, New York: Crossroad Publishing.(An appraisal of Tillich, including personal recollections, by a younger contemporary on the faculty at the University of Chicago, now a major theologian.)

Hopper, D. (1967) *Tillich: A Theological Portrait*, Philadelphia, PA: J.P. Lippincott. (A summary view of the man and his work for the non-specialist.)

Irwin, A.C. (1991) *Eros Toward the World: Paul Tillich and the Theology of the Erotic*, Minneapolis, MN: Fortress.(Traces the theme of *eros* in Tillich's writings and brings Tillich into dialogue with feminist theology.)

Kegley, C.W. (ed.) (1952) *The Theology of Paul Tillich*, New York: Pilgrim; revised edn, 1982.(Contains a collection of interpretive essays, autobiographical reflections and a reply to the essays by Tillich, and a complete bibliography of Tillich's writings.)

O'Meara, T.A. and Weisser, C.D. (eds) (1964) *Paul Tillich in Catholic Thought*, Dubuque, IA: Priory.(Essays on the theme, with an afterword by Paul Tillich.)

Pasewark, K.A. (1993) *A Theology of Power: Being Beyond Domination*, Minneapolis, MN: Fortress Press, 236-336.(Detailed analysis of Tillich's approach to 'power', exploring its ontological and ethical dimensions.)

Pauck, W. and Pauck, M. (1976) *Paul Tillich: His Life and Thought*, vol. 1, *Life*, New York: Harper & Row.(The standard biography of Tillich; volume 2 was never published.)

Rowe, W.L. (1968) *Religious Symbols and God: A Philosophical Study of Tillich's Theology*, Chicago, IL: University of Chicago Press.(A representative philosophical study of Tillich focusing on the doctrine of God and the theory of religious symbols.)

Scharlemann, R.P. (1969) *Reflection and Doubt in the Thought of Paul Tillich*, New Haven, CT: Yale University Press.(A closely reasoned exploration of theological themes such as correlation and paradox in Tillich's work.)

Stumme, J.R. (1978) *Socialism in Theological Perspective: A Study of Paul Tillich, 1918-1933*, Missoula, MT: Scholars Press.(A careful study of Tillich's religious socialism, with primary and secondary bibliography.)

Tillich, H. (1973) *From Time to Time*, New York: Stein & Day.(Memoirs and reflections by Tillich's wife, Hannah.)

Time

Time is the single most pervasive component of our experience and the most fundamental concept in our physical theories. For these reasons time has received intensive attention from philosophy. Reflection on our ordinary-tensed language of time has led many to posit a relation of metaphysical importance between time and existence. Closely connected with such intuitions are claims to the effect that time is unlike space, and in deep and important ways.

The development of physical theories from Newtonian dynamics through relativistic theories, statistical mechanics, and quantum mechanics has had a profound effect on philosophical views about time. Relativity threatens the notion of a universal, global present, and with it the alleged connections of time to existence. The connection between temporal order and causal order in relativity theories, and between the asymmetry of time and entropic asymmetry in statistical mechanics, suggest various 'reductive' accounts of temporal phenomena.

Finally, the radical differences between time as it appears in our physical theory and time as it appears in our immediate experience, show important and difficult problems concerning the relation of the time of 'theory' to the time of 'our immediate awareness'.

1 Time and existence

There is much philosophical debate about the degree to which the temporality of the world is like or unlike its spatiality. All agree that space and time differ in their dimensionality, and all agree, even in a relativistic context, that the temporality of the world is not a spatial dimension. Spatial and temporal aspects of the world are, for example, differentially interconnected with causal features.

One aspect that distinguishes time from space is the intuitive asymmetry of the former, unmatched by any related feature of the latter. While there is debate about what this asymmetry consists in, and how it is connected to other features of the world such as entropic asymmetry, there is not much question that it is uniquely characteristic of the temporal structure of the world (see Thermodynamics §§4-5).

But that distinction alone is often thought insufficient to capture the special nature of time. Expressions such as the 'flux' or 'flow' of time are often invoked metaphorically to capture what is intended, as in Bergson's treatment of temporality (see Bergson, H.-L. §2). McTaggart (1934) distinguished between those aspects of time that could be characterized using timeless temporal relations (such as one event being after another), called by him the B-series, and those aspects of time that were expressed in tensed discourse, called by him the A-series. It was McTaggart's contention that only the A-series captured the essence of temporality, but that since this essence was self-contradictory, time was 'unreal' (see McTaggart, J.M.E. §2).

The core of such claims lies, perhaps, in the intuition expressed by St Augustine that while being elsewhere than here hardly reduced the reality of an object, that which was not presently existing had no genuine existence at all (see Augustine §6). Other versions of this line take it that only past and present have 'determinate reality', the future, not yet having come into being, having no genuine determinate reality at all.

An important response to such ideas argues that there is no ultimate distinction in 'reality' between the present as contrasted with the past and future or between the present and past as contrasted with the future. This approach admits that past and future do not 'exist', of course, if 'exist' is being used in its tensed form, but finds in that only an interesting feature of natural language (having tensed-verb forms but no parallel structure for spatial location) and not anything of metaphysical importance.

McTaggart's claim that tensed expressions cannot be translated into tenseless ones, since the former have truth-values that vary with time and the latter do not, is admitted, but the proponents of the view that nothing metaphysical is represented by tense take it that this shows only that tensed discourse has an ineliminable 'token reflexive' or 'indexical' nature. To say '*x* occurred' is to say '*x* occurred before *now*'. 'Now' is a term whose reference varies with its use, since it refers to any moment of time at which it is uttered. Only this implicit indexicality, it is claimed, distinguishes tensed from untensed discourse, not some ability of the former but not the latter, to capture the essential metaphysical nature of time.

Those who reject this line offer several ways of trying to express what they think is missing in its account. Tense logics and models of splitting worlds (with the past as unique and the future as still offering manifold possibilities) are one approach. Others note the analogy with modal logic where the actual world is (in most accounts) metaphysically distinguished in its reality from the other possible worlds, much as the present is, allegedly, distinguished in its reality from other temporal moments of past and future (see Tense and temporal logic §5). The analogy goes only so far, however, since each moment of time is, when it is the present, the moment of reality. Formulating a full-fledged metaphysical perspectivalism that would capture the essence of the metaphysical claims of those who think time is specially connected with existence has proven a difficult task.

2 Relationism and its problems

Just as in philosophical explorations of space, the idea that time can be considered as nothing but a structured family of relations of material events is an important one (see Space §3). The doctrine is, perhaps, already implicit in Aristotle's description of time as 'the measure of motion' and becomes explicit in Leibniz's full-fledged relationism for space and time (see Leibniz, G.W. §11).

One question that immediately arises in a relationist perspective is whether there could be passage of time without change (see Change §2). One response to this is to introduce modality into the discourse. Just as empty space can be thought of as given by unactualized possibilities of spatial relations, so one might think of time intervals without actual change as given by possible but not actual changes. Important insights are contained in Shoemaker's notion of limits to situations themselves relationally acceptable (such as the limit of a sequence of ever-larger regions of the universe that abide without change) (Shoemaker 1969).

Newton introduces novel elements into the philosophical discussion in his insistence on 'absolute time' to accompany space as reference frame for absolute inertial motions (see Newton, I. §3). Many relationists like to think of the time interval as being arbitrarily specified by any periodic process. Newton emphasizes the distinction between the ideal process and any actual one. More importantly his theory, with its notion of distinguished absolutely uniform motions, presupposes that uniformity of time interval is fixed by physical processes other than arbitrarily chosen periodic processes. Without an absolute notion of equality of time interval the absolute notion of uniformity of motion is incoherent.

Within contemporary relativistic theories the issue of the adequacy of a relationist account of time becomes absorbed into the more general issues concerning the opposition between relationistically and substantivally interpreted spacetime theories (see Relativity theory, philosophical significance of §§5-7).

3 Time in relativistic theories

The introduction of the novel conceptualization of space and time forced upon us by the theories of special and general relativity once again illustrates the crucial interdependence of philosophical and physicist thought in this area.

In the special theory of relativity, the notion of simultaneity for events spatially separated from one another, and with it the time interval between any two separated non-simultaneous events, becomes relative to a chosen inertial reference frame (see Relativity theory, philosophical significance of §2). There is no longer a global 'now' that selects one and the same set of events for all observers, no matter what their state of motion relative to one another. Such a relativization of the notion of time is not to be confused with the relationist conception of time described above. With the advent of special relativity one has the beginning of the radical divergence between our intuitive notions of time and those posited by physical theories.

It is often claimed that within the relativistic picture the doctrines that connect existence with temporality (as discussed above) are no longer viable. If what is present and what is past can vary from observer to observer, then how can one maintain that past and future (or, alternatively, future alone) have no genuine reality? A frequent claim is that relativity forces upon us the view of the timelessness of existence contrasted with the 'essentially tensed nature of existence' theories.

As is so often the case, however, such an inference from physics to philosophy is premature. One could, for example, relativize one's notion of existence, thus claiming that past and future are unreal, but that events unreal

for one observer can be real for another observer itself real to the first. Alternatively, one could utilize only invariant spacetime features in the account. In one version of this approach the past remains real at a spacetime point-event, in the sense of the past light-cone at the point-event constituting the real (see Relativity theory, philosophical significance of §1). In another version reality is collapsed even more radically than in the Augustinian view, so that what is real at a place-time is only that unextended point place-time and its features.

General relativity, with its dynamic spacetime, makes matters even more peculiar. Models of the world in this theory exist with closed timelike curves. In such a world an event can be (globally) in its own past and future (see Time travel). Such possibilities led Kurt Gödel (1949) to assert that time was 'ideal' and that the '*t*' parameter of physics did not stand for time at all.

4 Time and causation

Various claims have been made to the effect that time 'is defined by' or 'reduces to' some other feature of the world. Causal theories date back to Leibniz, who pointed out that simultaneity for events could be characterized by their not being causally connectable. The breakdown of this association in the special theory of relativity is often taken as the key factor in claims to the effect that distant simultaneity is a matter of convention.

The issue of the extent to which temporal notions are, or are not, causally definable in relativistic theories is a complex one. In the special theory there is a relation framed in terms of causal connectability that is coextensive with simultaneity. In general relativity causal definitions of temporal metric and topological notions utilizing causal connectability alone fail in general. But one can characterize topological spatiotemporal notions in terms of the continuity of causal (timelike) paths. This suggests a reduction of the spatiotemporal notions not to causal notions*per se*, but to the epistemically accessible topological features of the spacetime, those directly available to an observer (see Relativity theory, philosophical significance of §3).

Both the physical connections of spatial and temporal (and spatiotemporal) features to causal features of the world, and the philosophical understanding of what these might tell us about 'what time is' in the world, remain controversial issues (see Causation).

5 Time in experience and time in nature

Temporality enters our conceptual framework both as a descriptive component of our immediate experience and as a component of our theoretical description of the physical world. But how are these aspects of time related to one another? This would be a problem for philosophy even if we took the time of nature to have the features we think of time having prescientifically. Once we have been told by science that time in nature is radically unlike anything we encounter in our immediate experience, as we seem to be told in relativistic theories and as we are likely to be even more radically informed when a fully quantum mechanical account of spacetime is forthcoming, the problem of relating experienced time to the time of theory becomes even more pressing.

One tradition starts from the time of our immediate experience and suggests that the time of nature is in some way or another a construct dependent on the time of conscious awareness. A seminal version of this approach is the treatment of time in Kant (1781/1787). Time is, along with space, one of the structuring principles of all experience, called by Kant, a pure intuition. As such it is a feature of the phenomenal world and not of 'things-in-themselves'. Both outer experience of physical objects and inner experience of psychological states are framed in the format of the temporal intuitive structure. Behind the temporal psychological self is the transcendental self, that which unifies all our experience with its implicit 'I think'. Non-temporal, the transcendental self is the ground of the temporal structuring of physical and psychological experience (see Kant, I. §§5-7).

The idealist tradition is continued in Husserl's phenomenological account of the ground and nature of our experience of time (1928). Here emphasis is placed on what the necessary features of such experience must be like in order that we experience things as past, present and future, and have the sense that we do of time as passage. In Heidegger's pragmatist-phenomenology, time as it functions in human activity is the ground of time altogether (see Heidegger, M. §§1-2). One begins with time as it appears in our experienced world of the fixed past, the present of action and the future of projected intentions. The time of nature and science, the time of the 'present-at-hand', is only derivative from the primordial time of experience as decision and action. In these

(external, transcendental) idealist accounts of time, the time of past, present and future is taken as central.

Contrasted with such approaches is that of the physicalist-naturalist, to whom the time of the physical world is basic and the time of human experience supervenient at best. In these accounts it is the timeless temporal relations of the B-series that are usually taken as fundamental. Time in experience is usually thought of in something in the vein of a secondary quality.

But even such naturalistic accounts have their problematic aspects. Any proposal to make time as experienced a mere secondary quality (such as that of Gödel noted above), and that takes the '*t*' of physics to be a parameter disconnected from time as we immediately encounter it, leaves little ground for understanding how physical theories can receive a realistic (as opposed to instrumentalistic) interpretation, or for theoretical realism in general (see Scientific realism and antirealism §1).

See also: Continuants §2

LAWRENCE SKLAR

References and further reading

Broad, C. (1938) *Examination of McTaggart's Philosophy*, Cambridge: Cambridge University Press.(Tensed versus tenseless theories of time.)

Capek, M. (1961) *The Philosophical Impact of Contemporary Physics*, Princeton, NJ: Van Nostrand.(Relativity and the relation of time to existence.)

Davies, P. (1974) *The Physics of Time Asymmetry*, Berkeley, CA: University of California Press.(Time asymmetry and the asymmetry of physical processes in time.)

Earman, J. (1974) 'An Attempt to Add a Little Direction to "The Problem of the Direction of Time"', *Philosophy of Science* 41 (1): 15-47.

Flood, R. and Lockwood, M. (1986) *The Nature of Time*, Oxford: Blackwell.(Essays on time in physics and in philosophy.)

Gödel, K. (1949) 'A Remark About the Relationship Between Relativity Theory and Idealistic Philosophy', in P. Schillp (ed.) *Albert Einstein: Philosopher-Scientist*, New York: Tudor.(General relativity, closed timelike lines and the distinction between the time of physics and the time of experience.)

Grünbaum, A. (1973) *Philosophical Problems of Space and Time*, Dordrecht: Reidel, 2nd edn.(Time in physics and philosophy.)

Horwich, P. (1987) *Asymmetries in Time*, Cambridge, MA: MIT Press.(Time asymmetry and the asymmetry of physical processes in time.)

Husserl, E. (1928) *Vorlesungen zur Phänomenologie des inneren Zeitbewußtseins*; trans. J.B. Brough, *On the Phenomenology of the Consciousness of Internal Time (1893-1917)*, The Hague: Nijhoff, 1991.(Husserl's main texts on time and time-consciousness. Phenomenological treatment of time awareness.)

Kant, I. (1781/1787) *The Critique of Pure Reason*; trans. N. Kemp Smith, London: Macmillan, 1929.(Kant's theory of time as pure intuition.)

Kroes, P. (1985) *Time: Its Structure and Role in Physical Theories*, Dordrecht: Reidel.(Time and time asymmetry in physics.)

Landsberg, P. (1983) *The Enigma of Time*, Bristol: Adam Hilger.(Readings on time in physics.)

Le Poidevin, R. and MacBeath, M. (1993) *The Philosophy of Time*, Oxford: Oxford University Press.(Collection of philosophical papers on time.)

McInerney, P. (1991) *Time and Experience*, Philadelphia, PA: Temple University Press.(Survey of phenomenological approaches to time in experience.)

McTaggart, J. (1934) *Philosophical Studies*, ed. S.V. Keeling, London: Arnold, and New York: Longmans. (McTaggart on tense and the unreality of time.)

Mellor, H. (1981) *Real Time*, Cambridge: Cambridge University Press.(Tense and indexicals.)

Newton-Smith, W. (1980) *The Structure of Time*, London: Routledge & Kegan Paul. (Philosophical issues about time.)

Prior, A. (1967) *Past, Present and Future*, Oxford: Oxford University Press.(Time, tense, reality and tense logic.)

Reichenbach, H. (1928) *Philosophie der Raum-Zeit-Lehre*, Berlin and Leipzig: de Gruyter; trans. M. Reichenbach, *The Philosophy of Space and Time*, New York: Dover, 1957.(Time in the spacetime of relativity theories.)

Reichenbach, H. (1956) *The Direction of Time*, Berkeley, CA: University of California Press.(Asymmetry of physics in time and the direction of time problem.)
Schlesinger, G. (1980) *Aspects of Time,* Indianapolis, IN: Hackett Publishing Company.(Time and tense and other philosophical issues.)
Sherover, C. (1975) *The Human Experience of Time*, New York: New York University Press.(Historical survey of philosophical views on the place of time in human experience.)
Shoemaker, S. (1969) 'Time Without Change', *Journal of Philosophy* 66: 363-81.(Intelligibility of changeless time.)
Sklar, L. (1993) *Physics and Chance*, Cambridge: Cambridge University Press. (Asymmetry of systems in time and the direction of time.)
Sklar, L. (1985) *Philosophy and Spacetime Physics*, Berkeley, CA: University of California Press.(Papers on time in physics and philosophy.)
Sklar, L. (1974) *Space, Time and Spacetime*, Berkeley, CA: University of California Press.(Time in relativity and the direction of time.)
Smart, J. (1964) *Problems of Space and Time*, New York: Macmillan.(Readings on time in philosophy and in physics.)
Stein, H. (1991) 'On Relativity Theory and Openness of the Future', *Philosophy of Science* 58 (2): 147-67.(Time and reality in a relativistic context.)
Yourgrau, P. (1991) *The Disappearance of Time*, Cambridge: Cambridge University Press.(Time in experience and in relativity, emphasizing Gödel's views.)
Zeh, H. (1989) *The Direction of Time*, Berlin: Springer.(Physical asymmetries in time.)

Time travel

The prospect of a machine in which one could be transported through time is no longer mere fantasy, having become in this century the subject of serious scientific and philosophical debate. From Einstein's special theory of relativity we have learned that a form of time travel into the future may be accomplished by moving quickly, and therefore ageing slowly (exploiting the time dilation effect). And in 1949 Kurt Gödel announced his discovery of (general relativistic) spacetimes whose global curvature allows voyages into the past as well. Since then the study of time travel has had three main strands. First, there has been research by theoretical physicists into the character and plausibility of structures, beyond those found by Gödel, that could engender closed timelike lines and closed causal chains. These phenomena include rotating universes, black holes, traversable wormholes and infinite cosmic strings (Earman 1995). Second, there has been concern with the semantic issue of whether the terms 'cause', 'time' and 'travel' are applicable, strictly speaking, to such bizarre models, given how different they are from the contexts in which those terms are normally employed (Yourgrau 1993). However, one may be sceptical about the significance of this issue, since the questions of primary interest - focused on the nature and reality of the Gödel-style models - seem independent of whether their description requires a shift in the meanings of those words. And, third, there has been considerable discussion within both physics and philosophy of various alleged paradoxes of time travel, and of their power to preclude the spacetime models in which time travel could occur.

1 Paradoxes of time travel

The simplest (and weakest) of the alleged paradoxes is the argument that time travel would permit 'changing the past', which would be a patent absurdity. Suppose, for example, you were *not* at the Battle of Hastings. If time travel were possible, it would seem that there would be nothing to stop you from going to the Battle, thus bringing about a contradiction with what was the case, namely, with your not having been there. However, this apparent problem is removed by recognizing that time travel would merely allow one to *influence* the past - to help make it what it was - but not to *change* it. Thus, if you were at the Battle, this fact would be a consequence of your present decision to travel back in time; and if you were not there, this would be because you now freely choose not to go, or fail to arrive for some other reason. Either way there is nothing that deserves to be called 'changing the past' and nothing contradictory.

But there are related problems, not so easily dismissed: most notoriously that time travel would appear to open the door to 'killing one's infant self'; or, more generally, to going back and doing something that one has strong present evidence was not done; or, even more generally, to producing a self-defeating causal chain - a closed string of events in which the occurrence of E at some particular spacetime point has a looping sequence of effects eventuating in the absence of E at that very point. It might be argued that since: one, such chains are evidently self-contradictory, and since, two, their possibility is entailed by the possibility of time travel, it follows that time travel is impossible. But in this formulation the allegation of paradox is unpersuasive, since premise two may well be denied. For it need not be conceded that in countenancing the prospect of time travel, hence of *some* closed causal chains, we necessarily become committed, in addition, to the possibility of self-defeating logically impossible chains (Lewis 1986).

An alternative formulation of the puzzle raised by these examples focuses on certain physical anomalies that would be associated with time travel, given the logical inadmissibility of self-defeating chains. For example, repeated attempts to commit autoinfanticide would have to be constantly yet inexplicably frustrated by unfavourable circumstances (such as memory lapses, gun jammings, brilliant surgeons on hand, and so on), which would add up to a peculiarly extreme coincidence; but we have empirical reason to believe that such a coincidence would never occur in our world; so we can infer that regular time travel will not take place. Not that such a thing is logically or conceptually impossible, but rather that what we see of our world (in particular, that massive coincidences do not happen) provides evidence against its occurrence (Horwich 1989).

2 Implications for the character of spacetime

Could some such argument show, in addition, that spacetime does not have the sort of structure (containing closed timelike lines) that is needed for time travel into the past? Not according to Gödel. He reasoned that the

mass-energy required for such trips would be so immense (Malament 1984) that they could never become technologically feasible; and so we should not think that if we were really in a Gödelian spacetime, we ought to be encountering the kind of anomaly (massive coincidences) that we do not in fact observe.

But this reasoning has been questioned by De (1969) who considers electromagnetic energy sources that Gödel did not take into account. Moreover one might suspect that a version of the autoinfanticide argument could be developed that does not depend upon things being transported back into the past. For it would seem that the existence of closed timelike lines, and the requirement that the causal chains located on them not be self-defeating, will together imply certain constraints on the distribution of matter (on the constituents of the chains) - constraints which we might observe to be violated. The idea, again very crudely, is that of all the otherwise possible 'initial conditions', only a restricted subset would, given our dynamical laws, be capable of 'causing themselves', and thereby avoiding the production of self-defeating chains; but insofar as we observe no such restriction, it would follow that our world could not contain closed causal chains.

3 Open questions

Although suggestive, this line of thought stands in need of considerable development in various directions. For it is as yet unclear what exactly the constraints implied by closed timelike curves would be, whether their severity is or is not substantially increased by the supposition of the special type of closed causal chains that embody human voyages into the past, and whether we do or do not have reason to believe that the constraints are violated. These are hard questions even to formulate precisely, let alone answer. Thus the topic of time travel promises to remain puzzling and vibrant for some time to come.

See also: Relativity theory, philosophical significance of; Spacetime; Time

PAUL HORWICH

References and further reading

De, U.K. (1969) 'Paths in Universes Having Closed Time-Like Lines', *Journal of Physics A*, series 2, 2: 427-32. (See §3.)

Earman, J. (1995) 'Recent Work on Time Travel', S. Savitt (ed.), in *Time's Arrow Today*, Cambridge: Cambridge University Press.(Includes a comprehensive, critical review of the physics literature. See §1.)

Gödel, K. (1949a) 'An Example of a New Type of Cosmological Solution to Einstein's Field Equations of Gravitation', *Reviews of Modern Physics* 21, 447-50.(A technical paper providing solutions to the field equations of general relativity that characterize spacetimes with closed timelike lines.)

Gödel, K. (1949b) 'A Remark about the Relationship between Relativity Theory and Idealistic Philosophy', in P.A. Schilpp (ed.) *Albert Einstein: Philosopher-Scientist*, La Salle, IL: Open Court, 1988, 557-62.(An informal discussion of the implications of Gödel's solutions, especially of their import for time travel and for the 'moving-Now' conception of time. See §2.)

Horwich, P. (1989) *Asymmetries in Time*, Cambridge, MA: MIT Press.(Discusses the alleged paradoxes in the context of a general treatment of time-asymmetric phenomena, and argues that the autoinfanticide considerations provide empirical evidence that time travel to our local past will never be feasible. See §2.)

Lewis, D. (1986) 'The Paradoxes of Time Travel', *Philosophical Papers*, Oxford: Oxford University Press, vol. 2, 67-80.(Defuses various alleged paradoxes and argues that time travel is a conceptual possibility. See §1.)

Malament, D. (1984) '"Time Travel" in the Gödel Universe', in D. Hull, M. Forbes and R.M. Burian (eds) *PSA 1984*, East Lansing, MI: Philosophy of Science Association, vol. 2, 91-100.(Calculates the mass-energy required for trips into the local past. See §2.)

Yourgrau, P. (1993) *The Disappearance of Time*, Cambridge: Cambridge University Press.(Argues that the import of Gödel's results is the non-existence of time rather than the possibility of time travel.)

Timon (*c*.315-*c*.225 BC)

Timon was a Greek philosopher-poet. The formative influence on his life was his meeting with Pyrrho, who was later hailed as the founder of Scepticism. He devoted his literary talents to eulogizing Pyrrho, and his satirical vigour to criticizing other philosophers. He, more than anyone else, carved the image of Pyrrho into what was to become its traditional form and placed it on its pedestal. In this, Timon seems to have been not only a fervent propagandist but also a major philosophical figure, exerting a decisive influence on the history and the very definition of neo-Pyrrhonian scepticism.

Timon was born in Phlius around 315 BC, dying in Athens around 225 BC (this chronology revises the traditional dating forwards by a decade, lengthening Timon's potential acquaintance with the philosophical debates of Hellenistic Athens). What evidence we have portrays him as cultivated, spiritual, a *bon vivant*, a caustic wit, absent-minded and unselfish. According to Diogenes Laertius (IX 109-16) he had a colourful life, full of ups and downs. Orphaned early, he first became a dancer, then lived at Megara, where he worked with the philosopher Stilpo (see Megarian School). His meeting with Pyrrho transformed his life and thought: he moved with his family to live close to Pyrrho, where he remained for twenty years. After Pyrrho's death, material problems flung him into an itinerant life, but his manifold talents in the end won him wealth and renown. He ended his days in Athens. There the Platonic Academy had just passed into the hands of Arcesilaus, and it was by contrast with the type of scepticism which Arcesilaus represented that Timon set out to define Pyrrhonian scepticism (see Academy).

Timon was quite unlike Pyrrho in that he was a prolific, versatile and talented writer. According to Diogenes Laertius (IX 110), 'the leisure which he snatched from philosophy he devoted to composing poetry.' He must have snatched a lot of it, because he wrote epics, dozens of tragedies and comedies, his *Silloi*, and prose works totalling 20,000 lines. The *Silloi* (that is 'Mockeries' - the original meaning of the word is disputed) are the least unknown of his works. Of these we possess sixty-six fragments (some very short) and a synopsis. Taking up the literary genre of parody (mainly of Homer) already made familiar by Xenophanes and the Cynic Crates of Thebes (see Cynics §3), Timon turned the mockery onto dogmatic philosophers. In the last two of the poem's three books, he portrayed himself - probably under the fictional guise of a descent to Hades - in conversation with Xenophanes, first about early philosophers, then about recent ones. The surviving fragments confirm that the style was parodic, and show the precision with which he aimed his barbs. More than thirty philosophers are mentioned, from the Presocratics down to Timon's contemporaries, each spiced with jokes of greater or lesser piquancy (and greater or lesser intelligibility to us). Only Pyrrho is unreservedly eulogized. Xenophanes, owing to his sceptical tendencies (see Xenophanes §5), occupies a privileged place: he is an interlocutor and guide, and perhaps intercedes on behalf of some philosophers who are treated with relative leniency (the Eleatics, Democritus, Protagoras). Even so Timon, not without malice, makes him say that in his old age he ended up himself falling back into dogmatism. This gallery of philosophers and the accompanying comments certainly played a leading role in the process by which Pyrrhonism defined itself (see Pyrrhonism), establishing its own pedigree and locating itself relatively to those philosophies with which it had affinities and differences.

Beyond the primarily critical *Silloi*, Timon wrote another philosophical poem, the *Indalmoi* ('Images'), apparently the constructive part of his programme for exalting, if not deifying, the moral figure of Pyrrho. His prose works are even less well known. The most important seems to have been the *Python*, where Timon's meeting with Pyrrho was ceremonially depicted, and where the pupil, in a restrained tone, expounded the main outlines of his master's work. This last work is probably the source of a summary of Pyrrho which Aristocles attributes to Timon, and which is undoubtedly the most important text in the entire doxography of Pyrrho (§3). Treatises by Timon titled *On Sensations* and *Against the Physicist* are also cited: remains of these works show that this master of mockery could discuss technical philosophical questions with the utmost seriousness.

Timon's place in the history of Pyrrhonian scepticism gave rise to disagreements in antiquity (Diogenes Laertius, IX 115). According to some, the 'succession' of teachers and pupils continued without interruption from Timon down to the Neo-Pyrrhonism we know from the writings of Sextus Empiricus (*c*. AD 200). Others, more plausibly, held that the line was broken after Timon and renewed later. It would be especially interesting to know whether Timon himself was merely Pyrrho's 'spokesman' (Sextus' description of him in *Against the Professors* I 53),

however effective and brilliant, or whether he left his own mark on the image he projected of his master's thought. Since Pyrrho wrote nothing, and is known through testimonies which depend, directly or indirectly, mainly on Timon, the answer can only be conjectural. In the summary reported by Aristocles, some details seem to imply that Pyrrho's own aim was exclusively or primarily ethical, and that it was Timon who imposed on him a sort of epistemological distortion. That would make Timon in effect the first of the Neo-Pyrrhonians. Even if one simply reflects on Timon's location (in both time and space), interests and personality, it will seem most unlikely that he was a mere clone of Pyrrho.

JACQUES BRUNSCHWIG

References and further reading

Brochard, V. (1887) *Les Sceptiques grecs*, Paris: Alcan; repr. 1923.(A classic which has aged well.)

Brunschwig, J. (1994a) 'Once Again on Eusebius on Aristocles on Timon on Pyrrho', in *Papers in Hellenistic Philosophy*, Cambridge: Cambridge University Press, 190-211.(Defends the hypothesis that Timon added the epistemological dimension to an essentially moral philosophy taught by Pyrrho.)

Brunschwig, J. (1994b) 'The Title of Timon's Indalmoi: From Odysseus to Pyrrho', in *Papers in Hellenistic Philosophy*, ed. J. Brunschwig, Cambridge: Cambridge University Press, 212-23.(Analysis of Timon's use of the text of Homer, based on one specific example.)

Diogenes Laertius (*c*. early 3rd century AD) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols.(See book IX 109-16 for Timon's life, character, works and pupils.)

Hankinson, R.J. (1995) *The Sceptics*, London and New York: Routledge.(The best modern book-length study.)

Lloyd-Jones, H. and Parsons, P. (1983) *Supplementum Hellenisticum*, Berlin and New York: de Gruyter.(Now the standard reference work for Timon's poetry.)

Long, A.A. (1978) 'Timon of Phlius: Pyrrhonist and Satirist', *Proceedings of the Cambridge Philological Society* 204: 68-91.(Pioneering work, which dramatically advanced the study of Timon.)

Long, A.A. and Sedley, D.N. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press, 2 vols.(Includes Timon's main philosophical fragments, with English translation and brief but excellent commentary.)

Pyrrho (*c*.365-*c*.275 BC) Testimonia, ed. F. Decleva Caizzi, *Pirrone - Testimonianze*, Naples: Bibliopolis, 1981. (Standard edition of the testimonies on Pyrrho; includes Italian translation and valuable commentary.)

Timon (*c*.315-*c*.225 BC) *Silloi*, ed. M. Di Marco, *Timone di Fliunte: Silli - Introduzione, edizione critica, traduzione e commento*, Rome: Edizioni dell'Ateneo, 1989.(Authoritative scholarly edition of the *Silloi*, with in-depth literary and philosophical commentary and an extensive bibliography.)

Tindal, Matthew (1657-1733)

Matthew Tindal was one of the last and most learned exponents of English deism. His most famous work is Christianity as Old as the Creation (1730), a comprehensive apology for natural religion. In it, he argued that God's law is imprinted on the nature of all things, including the human soul, and is accessible to reason. Revealed religion merely restates this universal law - the will of God - in a different form. Religion enables us to act in accordance with this natural order, and its end is happiness. However, Tindal was scathingly critical of the clergy, and cast doubt on the reliability of the Bible. Although Tindal's work was severely criticized by William Law, it exerted a considerable influence on the English and Continental Enlightenment.

1 Life and work

Born in Bere Ferrers in Devonshire, Matthew Tindal was brought up in the High-Church tradition of his father, the local parish priest. In 1673, he matriculated at Lincoln College at Oxford, where he was tutored by George Hickes, a High-Churchman. He later moved to Exeter College. He was elected a Fellow of All Souls in 1678 and was made a Doctor of Civil Law in 1685.

Tindal's England was deeply divided along political and religious lines. Theological controversies between Catholics and Protestants were further complicated by the self-confidence of reason, the claims of natural philosophy, and a rising interest in textual criticism. Like John Chillingworth and Pierre Bayle, Tindal went through a short Roman Catholic phase. But craving for freedom of thought, he returned to Protestantism in early 1688, eventually disencumbered of High-Churchism. At the time of the Glorious Revolution, Tindal rallied to William III, championed Whig opinions and discovered affinities with Low-Churchism. His early publications - *An Essay Concerning Obedience to Supreme Powers* (1693) and *An Essay Concerning the Laws of Nations, and the Rights of Soveraigns* (1694) - were concerned with legal matters. An unmitigated assault on episcopacy, *The Rights of the Christian Church Asserted*, caused a considerable stir in 1706 and established his reputation as a freethinker. In spite of a *Defence* (1707-8), the book was condemned by Parliament in 1710 and burned by the hangman. By then his offensive against the High-Church party had gathered pace with *A New Catechism, with Dr. Hickes's thirty nine Articles* (1710).

The first volume of *Christianity as Old as the Creation or, the Gospel, a Republication of the Religion of Nature* came out in 1730, ending two decades of quiet. Inspired by Cicero's *On the Nature of the Gods* and *On Divination* (see Cicero §3), Tindal's dialogue on religion, though repetitive and crammed with innumerable excerpts from Latitudinarian and rationalist divines, is typically in both argument and style the work of a mature and aged scholar. The second volume, though completed, was never published. The executor of Tindal's will was delinquent and the manuscript vanished. The first volume, Tindal's masterpiece, was a synthesis of the deist argument developed over a century by Herbert of Cherbury (§§3-4), Charles Blount, John Toland, Lord Shaftesbury (§2), Anthony Collins (§1), Thomas Chubb and Thomas Woolston (see Deism). After Tindal's death in 1733, it was hailed as the Bible of the deists, and in 1741 it was translated into German. It spurred the Enlightenment; Reimarus, Lessing and Kant (§14) are indebted to Tindal for their understanding of natural and revealed religion. In France, the *philosophes* revelled in Tindal, whom Voltaire (§2) praised as the most intrepid defender of natural religion. Jacques-André Naigeon was, however, not permitted to publish an abridged translation in French before 1770.

2 The religion of nature

It was the ambition of deism to define a creed, minimalist and natural, that might provide a common basis for belief, upon which theologians and natural philosophers would agree regardless of their denomination. In promoting the religion of nature on the basis of 'a Law of Nature, or Reason… a Law which is common, or natural, to all rational creatures' ([1730] 1995: 7), Tindal, a rationalist in principle, pursued a threefold objective: the reconciliation of Christian revelation with human rationality, the settlement of religious conflicts, and the establishment of religious freedom and toleration. Beyond its theological and cultural aspects, the construction of a philosophical religion responded to a social and political need of the time.

In order to solve problems that the Latitudinarians (who advocated an alliance of Christian theology with reason)

were among the first to perceive, Tindal postulated the innateness of religion. A law is imprinted on the human soul by virtue of creation and can be known by reason regardless of historical and cultural circumstances. This law is the will of the supreme legislator, the creator of the universe. Borrowing a key ontological principle from Samuel Clarke (§1), whose Boyle Lectures were emblematic of a rationalist age, Tindal enunciated the view that God's will was entrenched in 'the Nature and Reason of Things' ([1730] 1995: 56), in the immutable relations governing nature. The possibility of religion in general, and natural religion in particular, is established on a rationalist ontology.

But if true religion is a priori, natural and universal, should Christian belief not be dispensed with? The anthropomorphisms propagated by the Judaeo-Christian scriptures were as revolting to Tindal as the arbitrariness of a God who selected a miserable and ignorant tribe and demanded that human beings believe in magical phenomena and acknowledge absurd doctrines in order to receive the benefits of salvation. None the less, since Tindal aimed at bringing the Judaeo-Christian God and the God of deism into harmony, he espoused a hypothesis which has since become famous, namely that the a posteriori, revealed or historical religion was 'a Republication of the Law of Nature', 'a Restoration' or 'a Transcript of the religion of Nature' ([1730] 1995: 68, 176, 334; the first of these was originally a quote from Thomas Sherlock). This hypothesis had already gained some currency in seventeenth-century England owing to the sermons of Archbishop John Tillotson (1630-94) and Bishop Thomas Sherlock. Natural and revealed religion were 'like two Tallies… [which] exactly answer one another' ([1730] 1995: 51). It follows from this that Christianity is not a new religion, but, as the title of Tindal's book audaciously claimed, as old as creation.

Tindal's definition of true religion as 'plain, simple, and natural' ([1730] 1995: 241) is followed by six corollaries: (1) The sufficiency of natural religion. Natural religion is perfect: it originates from 'the Author of all Perfection' ([1730] 1995: 255); it is accessible universally; it discloses the divine will exhaustively; it allows for a reasonable worship; and it grounds revealed religion. (2) The equation of morality with nature. The law of nature expresses the will of God. All natural duties are summarized in the commandment to love God and our neighbour. (3) The equation of religion with ethics. Both Confucius and Jesus taught truths that encapsulated the substance of natural religion. Whereas morality is 'acting according to the Reason and Nature of Things considered in themselves', religion is 'acting according to the same Reason of Things consider'd as the Will of God' ([1730] 1995: 270). (4) Eudaimonia constitutes the purpose of religion. 'The ultimate End of all God's Laws, and consequently, of all Religion, is human happiness' ([1730] 1995: 90). (5) The difference between natural and revealed religion is reduced to a question of form. 'Natural and Revealed Religion only differ in the Manner of their being delivered' ([1730] 1995: 114). They are identical as regards the source of their being, namely the divine will, but distinct as regards the way in which they are known. Natural religion is known rationally, revealed religion historically. (6) The distinction between true religion and superstition. True religion deals with ends and prescribes means. What is concerned with matters irrelevant to morality is called superstition.

3 Polemical deism and its opponents

In the wake of Collins' *Priestcraft in Perfection* (1709), Tindal launched a veiled attack against the clergy, the Bible and the external evidences of Christian belief. He stigmatized the power of priestcraft and the evils of religious enthusiasm by ridiculing the customs, rites and mores of heathens and Jews. The clergy is charged with bigotry, fraud and cruelty. Self-serving Christian priests teach absurdities, impose arbitrary duties and foster superstition. Subjecting the Judaeo-Christian scriptures to the critique of reason, Tindal drew attention to a collection of scriptural problems that the nascent biblical criticism was beginning to exhume (see Hermeneutics, biblical §2). Not only did he cast doubt on the authority and reliability of the Bible, but he also disparaged the allegorical and typological interpretations of the Anglican tradition. The refutation of the external proofs of revealed religion had been a main feature of deism since John Toland. Seeking to eliminate from Christianity teachings that were either against or beyond reason, Tindal listed miracles among those things that can only nurture superstition. 'If you look no further than the Christian World, you will find, that Ignorance, and the Belief of daily Miracles go hand in hand; and that there's nothing too absurd for the Peoples Belief' ([1730] 1995: 192). Tindal's offensive against Christian orthodoxy anticipated d'Holbach's *Christianity Unveiled* (1756), which provided Feuerbach's and Marx's critiques of religion with their main arguments (see Feuerbach, L.A. §2; Marx, K. §3).

Tindal's natural and ahistorical rationalism resulted in a sophisticated moral elitism which cut right across doctrinal and ecclesiastical boundaries, but at the huge cost of an unsophisticated anticlericalism, antibiblicism and anti-Judaism. There were not many replies to his polemical book. In 1731, Daniel Waterland unsuccessfully tried to vindicate the Old Testament picture of God in an anonymous *Scripture Vindicated*, while William Law more successfully exculpated the God of orthodoxy from the charge of arbitrariness in *The Case of Religion and Reason, or Natural Religion, Fairly and Fully Stated* (1731). Law further blamed Tindal for failing to see in human beings the importance of both irrationality and evil.

See also: Natural theology

JEAN-LOUP SEBAN

List of works

Tindal, M. (1693) *An Essay Concerning Obedience to Supreme Powers*, London; microfilm, Ann Arbor, MI: University Microfilms, 'Early English Books 1641-1700', 1969.(A combination of reflection on the contemporary political scene and discussion of the general duties of citizenship in times of crisis; the microfilm version is of a 1709 reprint that also includes Tindal 1694a and 1697, and another essay.)

Tindal, M. (1694a) *An Essay Concerning the Laws of Nations, and the Rights of Soveraigns*, London; microfilm, Ann Arbor, MI: University Microfilms, 'Early English Books 1641-1700', 1969.(Tindal indulges in a controversy sparked by naval policy.)

Tindal, M. (1694b) *A Letter to the Reverend Clergy of Both Universities, concerning the Trinity and the Athanasian Creed*, London; microfilm, Ann Arbor, MI: University Microfilms, 'Early English Books 1641-1700', 1974.(Tindal's first major religious controversy.)

Tindal, M. (1697) *An Essay concerning the Power of the Magistrate, and the Rights of Mankind in Matters of Religion*, London; microfilm, Ann Arbor, MI: University Microfilms, 'Early English Books 1641-1700', 1974. (Argues, among other things, for the political rights of dissenters.)

Tindal, M. (1706) *The Rights of the Christian Church Asserted*, London, Part I (no more published); microfilm, Woodbridge, CT: Research Publications, 'The Eighteenth Century', 1991.('… against the Romish, and all other priests, who claim an independent power over it' - Tindal's first major work, the source of heated controversy.)

Tindal, M. (1707-8) *A Defence of the Rights of the Christian Church*, London, 2 parts. (Tindal's two vain attempts to defend his 1706 work; republished together in 1709.)

Tindal, M. (1710) *A New Catechism, with Dr. Hickes's thirty nine Articles*, London; microfilm, Woodbridge, CT: Research Publications, 'The Eighteenth Century', 1986. (Against the High-Church party in the Church of England, and Tindal's old Oxford tutor, Hickes.)

Tindal, M. (1718) *An Account of a Manuscript, entitul'd, Destruction the Certain Consequence of Division*, London; microfilm, Woodbridge, CT: Research Publications, 'The Eighteenth Century', 1986.(Turning to a more positive tack, Tindal makes a plea for the political unity of all Protestants.)

Tindal, M. (1730) *Christianity as Old as the Creation or, the Gospel, a Republication of the Religion of Nature*, London, vol. 1 (no more published); facs. edn ed. J.V. Price, London: Thoemmes, 1995.(Tindal's most famous work, which provoked some 150 published replies in the years following.)

References and further reading

Collins, A. (1709) *Priestcraft in Perfection*, London; microfilm, Woodbridge, CT: Research Publications, 'The Eighteenth Century', 1986.(Arguing against the power of priests, via a criticism of the Articles of the Church of England, this essay paved the way for Tindal's own attacks.)

Conybeare, J. (1732) *A Defence of Reveal'd Religion*, Dublin.(One of the most sympathetic and insightful replies to Tindal 1730.)

Gawlick, G. (1967) 'Einleitung', in *Christianity as Old as the Creation*, Stuttgart: Cannstatt, facs. edn.(A good discussion of the famous book, with an extensive bibliography.)

Law, W. (1731) *The Case of Religion and Reason, or Natural Religion, Fairly and Fully Stated*, London; microfilm, Woodbridge, CT: Research Publications, 'The Eighteenth Century', 1988.(The most successful attack on Tindal and on deism in general, showing that Tindal's arguments hit natural religion as much as revealed.)

Rupp, E.G. (1986) *Religion in England 1688-1791*, Oxford: Clarendon Press. (Contains a brief but clear description of the deist controversies.)

Stephen, L. (1991) *History of English Thought in the Eighteenth Century*, Bristol: Thoemmes, 2 vols.(This is a facsimile of the 1902 third edition; it is dated, but expansive and readable.)
Waterland, D. (1731-2) *Scripture Vindicated*, London, 3 parts; microfilm, Woodbridge, CT: Research Publications, 'The Eighteenth Century', 1986.(One of the many rather hasty replies to Tindal, itself attacked by moderates for its strict literalism in interpreting the Old Testament.)

Tocqueville, Alexis de (1805-59)

Tocqueville once observed that his temperament was the 'least philosophical' imaginable. He meant that his mind was governed by passionate commitment, a determination to defend civil and political liberty against threats resulting from social levelling and the growth of state power. Thus, Tocqueville's most famous work, De la démocratie en Amérique (Democracy in America) (1835, 1840), did not spring from detached curiosity about US institutions. It was rather an attempt to draw lessons from US society and government which could be used to reform French institutions. His belief in local autonomy - he called the New England township a 'school for citizens' - led him to develop a distinctive conception of liberty that combined elements of ancient citizenship and modern autonomy. That conception also shaped his own political career and later writings. In L'Ancien régime et la révolution (The Old Regime and the French Revolution) (1856) Tocqueville traced bitter class conflicts in France to the destruction of local autonomy long before 1789.

Tocqueville was the child of aristocratic parents who had nearly died on the scaffold in the Reign of Terror. His childhood was overshadowed by stories of relations who did not escape, including his famous grandfather, Lamoignon de Malesherbes - the pre-Revolutionary reformer and critic of royal despotism who became the young Tocqueville's hero. The sense of danger surrounding an aristocratic family after the Revolution left a profound mark on the boy. But that sense was strengthened by political developments in France during the Restoration (1815-30). The advent of a reactionary ultra-royalist government in the 1820s gradually convinced Tocqueville that it was not only futile but unjust to resist social change. He learned from a group of Restoration liberals called the *Doctrinaires* to consider the long-term movement from an aristocratic to a democratic social order - the 'Democratic Revolution' - as 'irresistible' and 'providential'. But that is not all he learned from the *Doctrinaires*. After completing his legal studies and becoming a junior magistrate, Tocqueville attended lectures by the historian François Guizot which deepened his understanding of two potentially dangerous consequences of the democratic social revolution in France, what the older Royer-Collard had called 'atomization' and 'centralization'.

These concepts shaped Tocqueville's intellectual development after 1828. By 'atomization' he understood the threat to association which resulted from the destruction of a corporate society, leaving a society in which individuals are equal but isolated and equally weak. The beneficiary is the state with its claim to a sovereign right which can be used to justify an indefinite concentration of power at the centre, controlling the periphery of society through a bureaucratic machine. Restoration liberals had begun to invoke a 'new federalism' to meet this threat and establish a balance between central power and local autonomy. Guizot, in particular, had criticized the concept of sovereignty passed down from Bodin and Hobbes, with its stipulation of a monopoly of authority in one agency (see Sovereignty).

Tocqueville evidently found the new liberal prescription tantalizing but vague - for would not local autonomy lead to the dissolution of the state through the emergence of competing 'sovereigns'? Such questions probably helped to direct his attention to the USA. Restoration liberals had laid emphasis on the role of class conflict between the aristocracy and middle classes in explaining why power had been centralized in France under the *ancien régime*. Tocqueville began to wonder if US political institutions might have a didactic potential because they had developed in a society which had been democratic from the outset.

The journey Tocqueville and his friend Beaumont made through America in the years 1831-2 confirmed this suspicion. US federalism embodied a balance between central power and local autonomy that enabled Tocqueville to complete the critique of the concept of sovereignty by demonstrating that political unity did not entail administrative centralization. But observation of the vibrant life of townships in New England led him beyond his initial concern with dispersing power into a new conception of liberty - what he was to call 'free *moeurs*', or the attitudes and habits of a free people. In that way Tocqueville stepped outside the confines of late eighteenth- and early nineteenth-century liberalism, with its emphasis on constitutionalism and legal rights. Instead, social attitudes became the focus of his concern. But while seeking to analyse and defend the civic spirit or citizenship which local participation fostered in the USA, Tocqueville also sought to avoid the drawbacks of a concept like Rousseau's general will (see General will).

In his *Du contrat social (The Social Contract)* (1762) Rousseau had defined active citizenship and concern for the

general welfare into a concept of real or moral freedom (the general will) which might conflict with the *de facto* wishes of individuals (see Rousseau, J.-J. §3). At the outset of the Restoration, Benjamin Constant (§1) had argued that Rousseau's defence of citizenship or ancient liberty had sacrificed the most important modern meaning of liberty; that is, individual autonomy protected by rights. Restoration liberals such as Constant charged Rousseau with creating a concept which could serve as a weapon of tyranny. They had in mind especially appeals to the 'will of the people' which had been used during the Reign of Terror to justify the sacrifice of individual liberty. Consequently they looked upon arguments for political participation with suspicion, if not with hostility. Tocqueville, after his journey through America, stood out against that trend. What he had observed about local autonomy persuaded him that it had a crucial moralizing potential. What is striking about the argument of *De la démocratie en Amérique (Democracy in America)* (1835, 1840) is that instead of tying political participation to 'correct' political decisions by definition, Tocqueville put forward a concept of free *moeurs* - the attitudes and habits of a people by self-reliance and the habit of association - which makes the moralizing process probable rather than necessary. In that way Tocqueville 'rescued' the value of citizenship by reconciling it with individual autonomy or rights in an unambiguously liberal doctrine of citizenship (see Citizenship §2).

The success of *Democracy in America* was stunning. It helped to launch Tocqueville into a political career. Yet the 1840s turned him into a sad and frustrated critic of Louis Philippe's July Monarchy, with its cynical manipulation of the limited franchise. Tocqueville's greatest political prominence came after the revolution of 1848, when he served briefly as French foreign minister. But it was a short-lived success. For Louis Napoleon's *coup d'état* (1859) and the Second Empire obliged Tocqueville to withdraw into a kind of internal exile. He would not serve a tyrant. Instead, he devoted his final years to showing how the bureaucratic French state and class hatreds had developed together in a vicious alliance. The structure of argument in *L'Ancien régime et la révolution (The Old Regime and the French Revolution)* (1856) remained essentially the same as that of *Democracy in America*. Tocqueville's insistence on the importance of decentralization and political participation made him a liberal moralist of a new kind. In his view, liberalism could not safely confine itself to an interest in market relations or legal structures. Fostering the virile attitudes of a free people was all-important. Sad but unswerving in his opposition to tyranny, he died in 1859.

See also: Liberalism

L.A. SIEDENTOP

List of works

Tocqueville, A. de (1835, 1840) *De la démocratie en Amérique (Democracy in America)*, trans G. Lawrence, ed. J.P. Mayer and M. Lerner, London: Fontana, 1966.

Tocqueville, A. de (1856) *L'Ancien régime et la révolution (The Old Regime and the French Revolution)*, trans. S. Gilbert, Garden City, NY: Doubleday, 1955.

References and further reading

Jardin, A. (1988) *Tocqueville*, London: Halban.

Kelly, G.A. (1992) *The Humane Comedy: Constant, Tocqueville and French Liberalism*, Cambridge: Cambridge University Press.(Uneven but imaginative.)

Lamberti, J.-C. (1983) *Tocqueville et les deux Démocraties (Tocqueville and the Two Democracies)*, Paris: Presses Universitaires de France.(An interesting study.)

Lively, J. (1965) *The Social and Political Thought of Alexis de Tocqueville*, Oxford: Clarendon Press.(A somewhat unhistorical, topical approach to Tocqueville's thought.)

Mill, J.S. (1840) 'M. de Tocqueville on *Democracy in America*', in *Collected Works of John Stuart Mill*, vol. 17, Toronto, Ont.: University of Toronto Press, 1977, 47-90 and 153-204.(This remains a penetrating analysis.)

Pierson, G. (1938) *Tocqueville and Beaumont in America*, New York: Oxford University Press.(A magisterial account of the journey through America.)

Rousseau, J.-J. (1762) *Du contrat social*, in *The Social Contract, with Geneva Manuscript and Political Economy*, ed. R.D. Masters, New York: St Martin's Press, 1978.

Schleifer, J. (1980) *The Making of Tocqueville's Democracy in America*, Chapel Hill, NC: University of North Carolina Press.

Siedentop, L. (1979) 'Two Liberal Traditions', in *The Idea of Freedom*, ed. A. Ryan, Oxford: Oxford University Press.(A contrast between English and French liberalism.)

Siedentop, L. (1994) *Tocqueville*, Oxford: Oxford University Press.(An intellectual biography and survey of his thought.)
Zeitlin, I. (1971) *Liberty, Equality and Revolution in Alexis de Tocqueville*, Boston, MA: Little, Brown.(Useful.)

Todorov, Tzvetan (1936-)

Tzvetan Todorov is a Bulgarian thinker and literary theorist who lives and works in France. His thought is 'structuralist' in that it seeks stable abstract principles explanatory of, but not directly given in, phenomena of literature and history. His thought is not only a milestone in the history of structuralism but is 'negatively' important to the development of post-structuralist thought: the rules and oppositions conveyed by structuralist typologies such as those of Todorov constitute much of what thinkers such as Jacques Derrida seek to 'deconstruct'.

Tzvetan Todorov was born and received his early education in Sofia, Bulgaria. At the age of 24, he moved to France, where he is (1996) Director of Research at the Centre National de Recherches Scientifiques in Paris. He has also taught at the University of Paris and at several American universities, among them Columbia and Yale. In addition to broader, more 'philosophical' works, he has published important studies on Rousseau, Goethe and Bakhtin. He has also edited a number of anthologies, including the *Dictionnaire encyclopédique des sciences du langage* (1972).

Todorov's earliest works (1968, 1971, 1978b) concern the study of literature, which he divides into two parts: 'poetics', which seeks the abstract structures governing literary genres, and 'interpretation', which concerns the meaning of individual texts. Poetics, for its part, is 'objective' and 'scientific'; its central concept, that of genre, plays a role in it akin to that of species in biology. 'Interpretation' is more subjective: while some interpretations can be wrong, none is ever the only correct one.

Todorov's poetics explores three basic aspects of the literary text: semantic, syntactic and verbal. Unlike traditional semantic units (for example, sentences), the literary work is 'dialogical' or 'intertextual': it refers, not merely to non-literary states of affairs, but also to prior and posterior texts (including earlier and later stages of itself). As fictional, or 'feigned' speech, it claims, not truth, but verisimilitude: accordance with the rules of its genre and with the expectations of its readers, or 'public opinion'. Following Russian formalists such as Mikhail Bakhtin, Vladimir Propp and others (see Bakhtin, M.; Russian literary Formalism), Todorov argues that the 'syntactic' order of the literary text is not grammatical, but narrative: the events narrated follow each other in a logical order ('mythological' narrative) or contribute to the manifestation of a single idea ('ideological' narrative). Its 'verbal' aspect is the way in which it enables the reader to construct a fictional world and includes the text's use of time, perspective and voice. None of these aspects, of course, is limited to what we call 'literature': all are widespread in many types of discourse. Hence no structural definition of literature is possible (1978b: 11). The proper object of literary theory remains the genre.

But a literary genre differs crucially from a biological species:

> The impact of individual organisms on the evolution of the species is so slow that we can discount it in practice. …[But] we grant a text the right to figure in a history of literature or of science only insofar as it produces a change in our notion of one activity or the other.
>
> (1970: 6)

Hence, poetics by itself is incomplete, and must be supplemented by the interpretive investigation of individual works. Todorov's most extended pursuit of such investigation can be found in *The Fantastic* (1970). His main historical account of it is in *Theories of the Symbol* (1977), which investigates the development of the modern or 'romantic' conception of the symbol, in which the work of art no longer 'imitates' reality but functions according to its own inner logic (thus making poetics itself possible). Finally, *Symbolism and Interpretation* (1978a) undertakes a general typology of interpretive strategies.

The rules which a work of literature obeys - those of its genre and of its own intertextuality - are not thought up by its author: they are beyond authorial control. Such a text is thus governed by elements which may be called foreign to it. If allowing foreign elements to dominate one's own thought and work is 'primitive', then as early as *Theories of the Symbol* Todorov is engaged in 'identifying the supposedly primitive mechanisms in our own thinking' (1977: 223). Structuralist poetics is thus from the start an enquiry into the otherness always resident in our own thought. By the time of *Literature and its Theorists* (1984), the theme of alterity has become 'the focal point of

[Todorov's] work' (1984: 3). Together with this new emphasis on alterity goes a shift in interest from literature to history.

The most important document in this shift is Todorov's *The Conquest of America* (1982). In this classic work, Todorov undertakes to examine the ways in which the European conquerors of America thought about its pre-Columbian inhabitants. The book is not, then, a 'theory' of anything. Rather, it recounts an 'exemplary narrative' with ethical meaning (1982: 4). The aim is not merely to understand the practices of the *conquistadors*, but to change those aspects of contemporary culture which exhibit their vestiges: 'to become aware of the relativity (hence the arbitrariness) of any feature of our culture is already to shift it a little' (1982: 254).

But the book retains a 'typological perspective' (1982: 252); the typology in question is developed in *On Human Diversity: Nationalism, Racism, and Exoticism* (1989). There are, it argues, three main ways to handle the empirical fact of human diversity. One is ethnocentrism, which simply generalizes characteristics of one's own society into 'universal' values. A second, more respectable today, is 'scientism': the scientific examination of human beings will teach us what they should pursue and avoid. Both are dangerous: ethnocentrism inevitably arrives at the result that we should rule others, while scientism arrives at the rule of scientists. Apart from the 'indefensible' approach of complete relativism, which falls to a version of the sceptical paradox (1989: 389) (see Relativism), only 'critical humanism' is left. This sees liberty - the capacity to reject any determination of one's nature - as the central feature of humanity. The universalism to which it aspires is always open to revision, but - like truth itself - always there as a goal.

See also: Structuralism in literary theory

FRANÇOISE LIONNET

List of works

Todorov, T. (1968; rev. edn 1973) *Qu'est-ce que le structuralisme: Poétique*, Paris: Éditions du Seuil, rev. edn 1973; trans. R. Howard, *Introduction to Poetics*, Minneapolis, MN: University of Minnesota Press, 1981.(An introductory handbook to structuralism, containing a key statement of its basic principles.)

Todorov, T. (1970) *Introduction à la littérature fantastique*, Paris: Éditions du Seuil; trans. R. Howard, *The Fantastic: A Structural Approach to a Literary Genre*, Cleveland, OH: Case Western Reserve University Press, 1973.(Pursues the structural principles of the literary distinction between what strikes the reader as strange, but natural, and what is truly marvellous or 'fantastic'.)

Todorov, T. (1971) *Poétique de la prose*, Paris: Éditions du Seuil; trans. R. Howard, *The Poetics of Prose*, Ithaca, NY: Cornell University Press, 1977.(One of Todorov's most influential works; at once an introduction to and a critical discussion of structuralist literary criticism, it also applies structuralism to texts from a variety of languages and periods.)

Todorov, T. (1972) *Dictionnaire encyclopédique des sciences du langage*, Paris: éditions du Seuil.

Todorov, T. (1977) *Théorie du symbole*, Paris: Éditions du Seuil; trans. C. Porter, *Theories of the Symbol*, Ithaca, NY: Cornell University Press, 1982.(A history of semiotics through one of its key concepts - that of the symbol - this work constitutes a major irruption of history into the supposedly atemporal features uncovered by structuralist analysis.)

Todorov, T. (1978a) *Symbolisme et interpretation* , Paris: Éditions du Seuil; trans. C. Porter, *Symbolism and Interpretation*, Ithaca, NY: Cornell University Press, 1982.(A general typology of interpretive strategies that can be taken with respect to 'symbols'.)

Todorov, T. (1978b) *Les Genres du discours*, Paris: Éditions du Seuil; trans. C. Porter, *Genres in Discourse*, Ithaca, NY: Cornell University Press, 1990.(A discussion of the types of, and principles underlying, literary genres.)

Todorov, T. (1981) *Mikhail Bakhtin: le principe dialogique*, Paris: Éditions du Seuil; trans. W. Godzich, *Mikhail Bakhtin: The Dialogical Principle*, Minneapolis, MN: University of Minnesota Press, 1984.(An extremely influential systematic account of Bakhtin, showing how the polysemy of his writings is meant to engage the reader in a dialogue in which the reader must decipher (or not decipher) the various ambiguities in Bakhtin's works.)

Todorov, T. (1982) *La Conquête de l'Amérique*, Paris: Éditions du Seuil; trans. R. Howard, *The Conquest of America*, New York: Harper & Row, 1984.(One of Todorov's most influential works; examines Europe's encounter with the Other (in this case, with the inhabitants of what came to be called Latin America) via an examination of the mind of the conquistadors as evidenced by contemporary accounts.)

Todorov, T. (1984) *Critique de la critique*, Paris: Éditions du Seuil; trans. C. Porter, *Literature and its Theorists*, Ithaca, NY: Cornell University Press, 1987.(A personal and critical history of major developments in twentieth-century literary criticism.)

Todorov, T. (1989) *Nous et les autres*, Paris: Éditions du Seuil; trans. C. Porter, *On Human Diversity: Nationalism, Racism, and Exoticism in French Thought*, Cambridge, MA: Harvard University Press, 1993.(A critical and typological account of ways of handling human diversity; widely recognized as contributing to the French debates on decolonization.)

Todorov, T. (1991a) *Face à l'extrème*, Paris: Éditions du Seuil; trans. A. Denner and A. Pollak, *Facing the Extreme*, New York: Metropolitan Books, 1996.

Todorov, T. (1991b) *Les Morales de l'histoire*, Paris: Grasset & Fasquelle; trans. A. Waters, *The Morals of History*, Minneapolis, MN: University of Minnesota Press, 1995.(A discussion of the types and structures of cross-cultural encounter.)

Toland, John (1670-1722)

Deist, freethinker and political republican, the Irishman John Toland's reputation is closely associated with the radical attack on Christian metaphysics and institutions in the Augustan period. His philosophical achievement was to turn the more erudite thinking of Spinoza, Hobbes and Locke into a popular polemic against the shibboleths of orthodox religious belief. In Christianity Not Mysterious (1696), burnt in Dublin by Parliamentary command in 1697, he exploited and extended the epistemology of Locke's Essay Concerning Human Understanding into a revision of Christian descriptions of the relationship between faith and knowledge, and a consequent defence of liberty of thought and belief.

Toland's contribution to philosophical developments can be traced to his scholarly training in the universities of Glasgow, Edinburgh, Leiden and Oxford in the 1690s. Skilful in languages and criticism, Toland forged his contemporary reputation as a figure of learning in the Republic of Letters. Associated at different times with powerful men such as John Locke, Anthony Ashley Cooper, the third Earl of Shaftesbury and Robert Harley in English society, and Prince Eugene of Savoy and Sophia of Hanover on the continent, Toland was a conduit for the diffusion of radical ideas across Europe. Controversial since his university days, he repeatedly found himself the subject of orthodox condemnation and hostility, whether publishing studies of apocryphal literature or defences of his political philosophy. The novelty of Toland's approach was to bring together and radicalize the achievements of a number of earlier philosophical positions.

As a pamphleteer and philosopher, Toland combined political radicalism with his heterodoxy: during the 1690s he was single-handedly responsible for publishing standard editions of the republican figures of the 1640s and 1650s, adapting the political philosophy of such commonwealthsmen as James Harrington to the exigencies of a monarchical constitution. Crucially, as his editorial work on the life and work of John Milton indicates, Toland made textual criticism (again adapted from the hermeneutical advances of Spinoza and the French Biblical critic Richard Simon) a tool of political and philosophical reform. Coiner of the neologism 'pantheist', Toland's private metaphysics (a combination of a Ciceronian neo-Stoicism, materialism and Spinozist ethics), disseminated in the Latin work *Pantheisticon* (1720), was a profound influence on the heterodoxy of the later French 'High' Enlightenment.

Christianity Not Mysterious (1696) was Toland's first major work. Premised on a combination of the epistemologies of Locke and Hobbes which emphasized the role of reason in the construction of knowledge, Toland argued that no Christian doctrine could be mysterious. Insisting that all theologies must be rational in order that every human being might understand them, Toland undercut the traditional Christian assertion that some doctrines (most obviously the Trinity for the Church of England) might be above the capacity of human rationality. Executing a historical examination of the meaning of the word 'mystery', Toland suggested that mysteries might be 'hidden' from human understanding rather than 'incomprehensible' to rational capacities. Toland's point, much vilified by Churchmen, was that no belief could legitimately be imposed by either Church or State on any individual. As Toland made clear in this and later works, the corrosion of the authority of clerical imposition in matters of interpretation was as much a political as a religious action.

Toland made the link between religion and politics in his series of canonical editions of commonwealth political philosophy in the 1690s. The works of Algernon Sidney, James Harrington, John Milton and Edmund Ludlow were adapted to the demands of eighteenth-century political culture in Toland's editions. Re-writing the millenarian language of Ludlow, or the puritanical idiom of John Milton, to the more civil demands of political discourse in the late seventeenth century was a skilled and long-lived achievement: Toland's editions remained canonical into the nineteenth century. In changing republican political philosophy from a programme of institutional reform into a language more suitable for a monarchical constitution, Toland placed the development of a civic theology and religion at the heart of his revision. 'Commonwealthsmen' pursued not the execution of all monarchs, but the imperatives of civic virtue. Protecting the safety of the people and advancing the cause of civic sociability were the central planks of this political philosophy. Importantly, making the connection between religion and politics, Toland, again adapting Harringtonian suggestions, insisted that public religion was the most effective way of establishing civic virtue.

As a corollary to his insistence that religion must be tuned to establish the harmony of civic communities, Toland embarked on a lifelong polemic against the corruptions and deviancies of priestcraft. His strategy was contrived as a deliberately public attempt to undercut the apostolic claims of clerical orthodoxy by establishing all religion as the manifestation of historical circumstance. Writing with a deliberately accentuated emphasis on his learning and skill, appropriating the scholarship of more orthodox authors, Toland, in works such as *Letters to Serena* (1704), suggested that key Christian doctrines such as the immortality of the soul were derived from 'heathen' origins. Similar essays on 'prejudice', 'idolatry' and the primitive church, stressed the continuities of all false religions: priests had ever corrupted the virtuous injunctions of natural religion. Pre-empting the naturalistic accounts of later thinkers such as David Hume and the *philosophes* of the Enlightenment, Toland established - importantly for a non-learned audience - that the ceremonial, doctrinal and institutional elements of all religions were culturally specific (see Enlightenment, Continental).

Drawing from the relativism implicit in such historical perspectives, Toland insisted that since there were no clear divine imperatives in favour of any one religious economy, all religions were equal. Concepts of religious truth were thus displaced for Toland by assessments of religious function. Good religion was focused on facilitating the values of civic community. Corrupt religion was a system calculated for the false private advantage of priests and kings. The high point of such arguments was to be found in Toland's *Nazarenus. Or Jewish, Gentile and Mahometan Christianity* (1718), a work first composed as a private, clandestine manuscript for the purview of Prince Eugene of Savoy. In the published work, using the device of a commentary on a suppositious gnostic work, the Gospel of Barnabas, Toland surveyed the historical relationships of Judaism, Christianity and Islam, concluding that as historical manifestations of specific cultural contexts the three religions were of equal value. Such an emphasis on locating the claims of religious truth in historical context has led some commentators to suggest that in Toland's works, and in the scholarly response to them on the continent, lie the origins of a modern *Textsgeschichte* identified in the Tübingen school.

J.A.I. CHAMPION

List of works

Toland, J. (1696) *Christianity not Mysterious*, London: Samuel Buckley.(The work was condemned and prosecuted in London and Dublin. There are various modern reprints.)

Toland, J. (1698) *The Life of John Milton*, Amsterdam; repr. in H. Darbishire (ed.) *The Early Lives of Milton*, London: Constable, 1932.(Controversial edition and defence of the republican poet's life and work.)

Toland, J. (1699) *Amyntor*, London: The Booksellers of London.(Critical defence of his edition of Milton; exploits textual criticism of Scripture to undercut clerical authority.)

Toland, J. (1700) *The Oceana of James Harrington and his other works*, London: The Booksellers of London. (Critical edition, much reprinted in the eighteenth century.)

Toland, J. (1704) *Letters to Serena*, London: Bernard Lintot; modern edn, ed. G. Gawlick, Stuttgart: Friedrich Frommann, 1964.(Five essays on idolatry, prejudice and materialist metaphysics; sections were later translated into French by the atheistic Baron D'Holbach.)

Toland, J. (1705) *Socinianism Truly Stated*, London.(A very rare pamphlet that contains the first use of the word 'Pantheist'.)

Toland, J. (1709) *Adeisdaemon, sive Titus Livius a superstitione vindicatus (The unsuperstitious man. A vindication of Titus Livy from the charge of superstition)*, The Hague: Thomas Johnson.(Radical historical reading of the life of Moses as a political legislator: poses hermeneutical doubts about the status of revelation.)

Toland, J. (1714) *Reasons for Naturalising the Jews in Great Britain and Ireland*, London: J. Roberts; modern edn, ed. H. Mainusch, Stuttgart: W. Kohlhammer, 1965.(Radical defence of toleration for Jews.)

Toland, J. (1718) *Nazarenus, or Jewish, Gentile, and Mahometan Christianity*, London: J. Roberts; critical edn, ed. J.A.I. Champion, Oxford: Voltaire Foundation, 1998. (Controversial re-reading of the historical relationships between the three great religions.)

Toland, J. (1720) *Pantheisticon*, Cosmopoli (London).(Translated into English in 1751; a parody of Christianity and vehicle for Toland's pantheist metaphysics.)

Toland, J. (1726) *A Collection of several pieces of John Toland*, London: J. Peele, 2 vols.(The posthumous works and unpublished correspondence.)

References and further reading

Brykman, G. (ed.) (1995) 'John Toland (1670-1722) et la crise de conscience europééne' (John Toland and the crisis of the European mind), in *Revue de synthèse*, Paris: Albin Michel. (Collection of essays that covers the diverse contexts of Toland's thought.)

Carabelli, G. (1975) *Tolandiana. Materiali bibliografici per lo studio dell'opera e della fortuna di John Toland (1670-1722) (A bibliography for the study of the life and works of John Toland)*, Florence: La Nuova Italia. (Massive bibliographical study of Toland: indispensable for context and reception.)

Jacob, M. (1981) *The Radical Enlightenment*, London: Allen & Unwin.(Locates Toland's thought within the radical context of republican freemasonry.)

Sullivan, R.E. (1982) *John Toland and the Deist Controversy*, Cambridge MA: Harvard University Press.(The best intellectual biography of Toland, which locates him within the broader intellectual and religious traditions.)

Toleration

Toleration emerged as an important idea in the seventeenth century, receiving its fullest defence in John Locke's A Letter Concerning Toleration (1689). Initially developed in the context of attempts to restore peace in a Europe convulsed by religious conflicts, in the nineteenth and twentieth centuries it came to be extended to the accommodation of disputes about racial, sexual and social differences. Toleration is widely thought to be an essential element of a free society, especially one marked by moral and cultural pluralism, and it figures particularly prominently in the political theory of liberalism.

The paradigm example of toleration is the deliberate decision to refrain from prohibiting, hindering or otherwise coercively interfering with conduct of which one disapproves, although one has the power to do so. The principal components of the concept of toleration are: a tolerating subject and a tolerated subject (either may be an individual, group, organization or institution); an action, belief or practice which is the object of toleration; a negative attitude (dislike or moral disapproval) on the part of tolerator toward the object of toleration; and a significant degree of restraint in acting against it.

Philosophical arguments have mostly concerned: the range of toleration (what things should or should not be tolerated?); the degree of restraint required by toleration (what forms of opposition are consistent with toleration?); and, most importantly, the justification of toleration (why should some things be tolerated?).

1 Historical development

While intimations of the idea of toleration have been discerned in aspects of Ancient Greek, Roman and medieval thought, it was in sixteenth-century Europe that the concept began to emerge in a distinct and recognizable form (see, for example, the work of Jean Bodin); and in the seventeenth century that it became a crucial component in theorizing about how civil peace could be restored and maintained in a Europe riven by wars of religion. Althusius, Spinoza, Milton and Bayle were all important defenders of some conception of toleration. However, it is in John Locke's *A Letter Concerning Toleration*, first published anonymously in Latin in 1689, that the formulation and defence of the idea of toleration received its fullest statement.

Locke's arguments were addressed exclusively to the religious disputes of his time. Many of his arguments were based on religious or scriptural considerations and there was little in them that was entirely new. What Locke did, however, was to organize existing arguments into a coherent, powerful and principled defence of toleration. In particular, he emphasized the irrationality of religious intolerance, arguing that the attempt to coerce people into believing what they could not conscientiously assent to was both futile and destructive of the civil order. In Locke's view the prime function of government was to maintain civil order and public security: it was not the business of the government to save souls, for which it had neither the appropriate means nor the requisite expertise. However, where religious belief, or its lack, posed a genuine threat to civil order - for example, in the case of Catholics who owed their allegiance to a foreign power, or that of atheists who, in the absence of an effective threat of divine retribution, could not be trusted to keep their promises - intolerance was in principle entirely proper, although whether in fact it should be employed was a matter of prudential judgment (see Locke, J. §7).

Locke's argument for toleration was modest in its scope and cautious in its application. However, its central contention about the proper limits of government intervention in the lives of its citizens contained the seeds of a more expansive and positive conception of toleration. Both the ideas of the Enlightenment and the theory and practice of the US Constitution provided significant further support for toleration, especially for the toleration of heretical or unpopular ideas and opinions. Voltaire was a particularly influential publicist, being attributed with perhaps the most famous remark in defence of toleration: 'I disapprove of what you say, but I will defend to the death your right to say it'.

It is in J.S. Mill's *On Liberty* (1859), probably the most sustained and eloquent plea for toleration yet written, that toleration is wholeheartedly celebrated as part of a general defence of the value of liberty. Essentially, Mill argued for toleration on two grounds. First, he contended that it was a necessary condition for social improvement, intellectual progress and the growth of knowledge. Second, he claimed that it was required for the moral and

spiritual development of the individual. Toleration, therefore, is an essential ingredient of a good society, being both a means to social progress and a prerequisite for the development of individuality (see Mill, J.S. §12). Mill's arguments have been enormously influential for modern liberal theorists such as Ronald Dworkin, H.L.A. Hart, Joseph Raz and, more ambiguously, John Rawls.

Arguments for toleration have been important in removing religious disabilities, easing discrimination against minority groups and promoting a more permissive attitude towards differing life styles. Modern arguments about toleration invariably presume that some substantial element of toleration is essential to a free and democratic society and disagreement generally concerns where, and on what grounds, the limits of toleration should be drawn.

This is not to deny that any robust idea of toleration continues to have, as it always has, its opponents. The case against toleration has usually centred on the claim that it effectively condones immorality or undesirable behaviour; and that in doing so it undermines personal virtue and social cohesiveness. In relation to modern Western society, opponents claim that we have too much toleration, not too little.

2 The concept of toleration

Toleration is sometimes used loosely to mean freedom or liberty but is more accurately understood as a narrower concept. It is concerned specifically with permitting the expression of beliefs, actions or practices which the tolerator would prefer not to exist. Three components of the core concept of toleration can be identified. First, there must be some conduct about which the tolerator has a negative attitude. Second, this negative attitude must not be acted upon in ways which coercively interfere with others acting in the disapproved manner. Third, this refusal to interfere coercively must be more than mere acquiescence or resignation. Toleration is particularly important and problematic when it involves a principled refusal to prohibit conduct believed to be wrong. This gives rise to the so-called 'paradox of toleration' according to which toleration requires that it is right to permit that which is wrong.

Normally, one cannot be tolerant of something if one is either indifferent to it or approves of it. Acting tolerantly characteristically implies some sort of negative attitude or judgment on the part of the tolerator - ranging from mere dislike to intense moral disapproval - towards that which is tolerated. However, a tolerant disposition may also relate to the extent of the negative judgments. For example, to say that a person is tolerant may be to claim that they do not possess an inappropriate range or unjustified number of negative attitudes to the conduct of other people.

To describe a person or group as acting tolerantly standardly implies that they have the power to prohibit or interfere with that which they are tolerating, but choose not to do so. However, not all cases of such restraint are instances of toleration: only where the restraint is motivated by an appropriate reason is the response a distinctively tolerant one. There is some unavoidable vagueness about how restrictive 'interference' with the disapproved conduct can be while remaining consistent with toleration. At one extreme, in normal circumstances, the mere attempt to dissuade a person by reasonable argument from engaging in the disapproved conduct is entirely consistent with acting tolerantly. At the other extreme, the use of physical coercion to prevent an agent engaging in the disapproved conduct is utterly inconsistent with toleration. The key variable is the degree of coerciveness of the interference. The extent to which the imposition of costs which fall short of prohibition or prevention, such as informal social ostracism, higher taxation of disapproved conduct and other forms of disincentive, is compatible with toleration is partly a matter of circumstance and intent, and often involves judgments which allow for reasonable disagreement. In general terms, the more such disincentives incline towards coercion the less consistent they are with toleration. The paradigm of intolerance is the deliberate attempt to eliminate disapproved conduct by coercive means; in its most extreme and ruthless form, persecution. Toleration is therefore a matter of degree (see Coercion).

The moral ideal of toleration can be seen as a mean between intolerance, or the refusal to permit that which should be permitted, and indulgence or laxity, the permitting of that which should not be permitted, without lapsing into indifference, the refusal to judge that which should be judged. Much dispute about toleration has concerned its proper scope and limits. This, in turn, is usually related to differing accounts of the moral grounds or justification of toleration.

3 The justification of toleration

Many arguments for a policy of toleration with respect to particular beliefs or actions are largely prudential or pragmatic. They are couched in terms of the costs (social, economic or political) of intolerance for the intolerant themselves. Such arguments have been, and continue to be, of enormous practical importance, especially in the context of violent religious or political disputes. They are often effective where appeals to more high-minded concerns are not. Nor can such prudential considerations be separated entirely from moral arguments for toleration since they commonly invoke values such as peace, order, security and the avoidance of suffering. However, they do not involve any deeply principled defence of toleration.

The limitations of such arguments are evident in assessing utilitarian justifications of toleration. Utilitarians typically argue for toleration in terms of the maximization of happiness, welfare, preference satisfaction or whatever interpretation is given to utility. The limits of toleration should be set so as to obtain the optimum balance of benefits over harms. As with all utilitarian arguments, however, this defence of toleration is unavoidably conditional upon the truth of many controversial and disputable historically and culturally variable empirical claims. Thus utilitarianism offers a rather precarious and uncertain defence of toleration, heavily dependent upon contingent social circumstances. Moreover, in view of the difficulty of establishing the truth of many of the empirical claims, it is perhaps not surprising that while some utilitarians, such as J.S. Mill, have been uncompromising advocates of toleration, others, such as Mill's contemporary J.F. Stephen, have been considerably less enthusiastic in their support. Indeed, utilitarians are likely to have most difficulty justifying toleration precisely where it may be most needed; that is, in a society with an intensely intolerant majority (see Utilitarianism).

Another argument for toleration has derived from moral or religious scepticism (see Moral scepticism). On this view, since we can have no guarantee that we possess moral or religious truth we are not justified in imposing on others our own beliefs about these matters. Toleration, therefore, is the only proper response to our fallibility or to the absence of any discernible truth with respect to moral and religious beliefs. Undoubtedly the growth of religious scepticism was an important source of increased religious toleration but the affinity between toleration and scepticism is not as straightforward as it might appear. It has also been argued that if we can never have knowledge of the truth in moral and religious matters, or if there is no truth about such matters to be discovered, then there is nothing wrong with enforcing uniformity of belief. If allowing freedom of belief promotes disunity and creates social tensions, and suppression of beliefs involves no loss of truth, why not seek to rigorously control such diversity?

What this discussion so far suggests is that neither pragmatic nor utilitarian nor sceptical arguments can provide a very secure basis for a principle of toleration. Such a principle must be underpinned by a moral theory which provides stronger support. Much recent moral and political philosophy has been concerned to articulate a defence of toleration in terms of principles or values such as autonomy, impartiality and respect for persons. What these arguments typically seek to show is that toleration is an essential element of any society in which each individual will have the possibility of living an autonomous life; or a necessary implication of the requirement that the state should be impartial between different conceptions of the good; or integral to the idea that every person is an end in themselves and therefore deserving of respect.

All of these arguments are complex and give rise to controversy, and each has been developed in interestingly different ways. In very broad and general terms, however, they are all informed by a particular conception of the person and a view about the proper limits of coercive interference, especially by the state, in people's lives. Persons are conceived as choosers, capable of forming and pursuing their own conception of the good. Each person should be able to exercise significant control over their own life, to make and act on their own choices and decisions, without being subject to coercive interference by other people or the state. The responsibility of government is to secure the conditions under which this is possible, though it is a matter of dispute exactly what this will entail. There is, however, widespread agreement among proponents of these arguments for toleration that it is not the government's role to prohibit, or even on some views to promote, any specific conceptions of the good (see Neutrality, political §2).

In modern Western societies, disagreements about the scope and limits of toleration have centred around race, religion, gender, sexual practices and political affiliation. In particular, dispute has surrounded the permissibility of

‘hate’ speech, pornography and, in the UK at least, blasphemy. Many of these disputes have been further complicated by the development of multicultural societies and the need to arrive at mutually acceptable terms of accommodation for groups with deeply opposed conceptions of what constitutes a valuable life. What is particularly noteworthy is that attacks on toleration in these areas have frequently come not from the traditional opponents of toleration but from groups such as feminists, socialists and even some liberals. While the question of the extent to which the intolerant should themselves be tolerated has always proved difficult, many of these more recent disputes go well beyond this familiar problem, calling into question some of the most basic tenets of typical liberal accounts of toleration.

See also: Freedom of speech; Law, limits of; Liberalism; Multiculturalism

JOHN HORTON

References and further reading

Bollinger, L. (1986) *The Tolerant Society: Freedom of Speech and Extremist Speech in America*, New York: Oxford University Press.(A thorough exploration of the justification and limits of freedom of speech in the context of the USA.)

Budziszewski, J. (1992) *True Tolerance: Liberalism and the Necessity of Judgement*, New Brunswick, NJ: Transaction Publishers.(A critique of tolerance as ethical neutrality.)

Gray, J. and Smith, G.W. (eds) (1991) *J.S. Mill: On Liberty in Focus*, London: Routledge.(Mill’s *On Liberty* along with the most important modern discussions of it.)

Heyd, D. (ed.) (1996) *Toleration: An Elusive Virtue*, Princeton, NJ: Princeton University Press.(A collection of articles on the meaning, scope and justification of toleration in the modern world.)

Horton, J. (ed.) (1993) *Liberalism, Multiculturalism and Toleration*, London: Macmillan.(Essays on multiculturalism and toleration with particular reference to the ‘Rushdie Affair’.)

Horton, J. and Mendus, S. (eds) (1991) *John Locke: A Letter Concerning Toleration in Focus*, London: Routledge.(Popple’s translation of Locke’s *Letter* with historical and philosophical discussions.)

Horton, J. and Nicholson, P.P. (eds) (1992) *Toleration: Philosophy and Practice*, Aldershot: Avebury.(A diverse collection of articles on various aspects of toleration.)

Kamen, H. (1967) *The Rise of Toleration*, London: Weidenfeld & Nicolson.(An introductory historical account of the development of toleration.)

King, P. (1976) *Toleration*, London: Allen & Unwin.(A comprehensive analysis of the concept of toleration.)

Locke, J. (1689) *A Letter Concerning Toleration*, ed. R. Kilbansky, trans. J.W. Gough, Oxford: Clarendon Press, 1968.(An accurate modern translation of Locke’s letter.)

Mendus, S. (1988) *Justifying Toleration. Conceptual and Historical Perspectives*, Cambridge: Cambridge University Press.(An excellent collection of historical and analytic discussions.)

Mendus, S. (ed.) (1989) *Toleration and the Limits of Liberalism*, London: Macmillan.(A probing assessment of liberal arguments for toleration.)

Mill, J.S. (1859) *On Liberty*, in *Collected Works of John Stuart Mill*, ed. J.M. Robson, London: Routledge, 1991, vol. 18, 213-310.(A classic liberal defence of toleration.)

Newman, J. (1982) *Foundations of Religious Tolerance*, Toronto, Ont.: University of Toronto Press. (An analysis and defence of religious tolerance.)

Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.(The latest statement of the most influential contemporary liberal defence of toleration.)

Richards, D.J. (1986) *Toleration and the Constitution*, New York: Oxford University Press.(A magisterial discussion of the US Constitution and its interpretation of toleration.)

Tinder, G. (1975) *Tolerance: Toward a New Civility*, Amherst, MA: University of Massachusetts Press.(Develops an account of toleration on grounds which challenge some of the standard liberal argument in its defence as ethical neutrality.)

Wolff, R.P., Marcuse, H. and Moore, B. (1969) *A Critique of Pure Tolerance*, London: Jonathan Cape.(A celebrated series of radical critiques of ‘liberal’ tolerance.)

Toletus, Franciscus (1533-96)

Toletus had an independent, somewhat eclectic, but fundamentally Thomistic outlook. In philosophy his most important works were his commentaries on Aristotle in the areas of logic and natural philosophy. In these commentaries he drew upon the whole previous scholastic tradition to raise and answer questions which were debated in his time and later. In theology he commented upon the greater part of Aquinas' Summa theologiae. Here again he drew upon scholastic philosophers to raise and discuss a wide variety of metaphysical, epistemological and ethical topics. Far from being a slavish follower of Aquinas or Aristotle, he expressed his respectful disagreement with them wherever reason compelled it.

1 Life

Franciscus Toletus (Francisco de Toledo) was born at Cordoba in Spain. After receiving a master of arts at Valencia, he studied theology at Salamanca under the famous Domingo de Soto. Ordained a priest in 1556, Toletus was already teaching philosophy at Salamanca when he entered the Society of Jesus (the Jesuits) in 1558. In the following year he was sent to Rome where he taught philosophy and then theology, bringing with him the Thomistic outlook emphasized at Salamanca by Francisco de Vitoria and his disciple, Soto. Toletus went on diplomatic missions to Germany, Poland, the Low Countries and France for the Holy See, and in 1593 became the first Jesuit cardinal. He died in Rome.

2 Commentary on Aquinas

His partial commentary on the *Summa theologiae* of Thomas Aquinas (§6) was left in autograph at the time of his death and was published in four volumes only in the nineteenth century. It addresses issues about most of the first sixty-four questions of the First Part (Ia), the first eighty-eight questions of the second part of the Second Part (IIaIIae) and the ninety questions of the Third Part (IIIa) of the *Summa*. Written in a clear and lively style, it is not simply an exposition of the text of Aquinas. Toletus is interested in the whole of scholasticism, especially in Duns Scotus, Durandus of St Pourçain, William of Ockham, Cajetan, Lychetus and others in relation to Aquinas. In discussing their concerns, he goes far beyond the text of the *Summa*. For example, in a First Part (*In Summam Theologiae* q.2 a.1) discussion of self-evident propositions he explores the historical issue of the self-denying statement, 'There is no truth.' In the same part (*In Summam Theologiae* q.2 a.3), most likely reflecting Cajetan, he says that taken by itself each of Aquinas' famous five ways proves the existence of God only imperfectly, but taken together they demonstrate a collection of attributes which can belong only to God. In a Second Part (*In Summam Theologiae* q.57 a.3) discussion Toletus separated the 'law of nations' (*jus gentium*) from the natural law and thus set the stage for later Jesuits such as Suárez (§4) to do the same. Again in the Second Part, he expanded Aquinas' treatment of war in various ways. For example (*In Summam Theologiae* q.34 a.1 dub.4), he distinguished three classes of soldiers (those subject to some prince, those not subject who have accepted salaries to fight in time of war, those not subject and not salaried) and discussed their respective obligations to consider the justice of any war in which they may take part, as well as the obligations to restitution they may incur.

3 Logic

The first of Toletus' expressly philosophical works was his *Introductio in dialecticam Aristotelis (Introduction to the Dialectic of Aristotle)* which appeared at Rome in 1560. Reflecting medieval as well as Aristotelian concerns, through five books it dealt with the definition and division of dialectics; with terms (nouns and verbs) and propositions; the supposition, the ampliation, the restriction and so on, of terms; the opposition, equipollence and conversion of propositions; modal propositions and exponibles; reasoning and syllogisms (both proper and fallacious). In 1572, Toletus published at Rome a commentary together with questions on Aristotle's logic. Through seven parts, he treated questions on logic in general, on the five universals of Porphyry, as well as on the *Categories*, *Perihermeneias* and *Posterior Analytics* of Aristotle. His procedure, as he laid it out, was first to explain a question, and then to give those opinions 'which beyond the fact that they are the views of eminent men seem able to offer some reward to their hearer' ('Lectori', *Introductio in dialecticam*). Third, he showed what should be held as more certain and, finally, he refuted opposing arguments (see Logic, Renaissance).

4 Physics

In 1573, Toletus published a commentary with questions on the eight books of Aristotle's *Physics*. In this work, disagreeing with Aquinas and a majority of scholastics, he sided with Bonaventure and Henry of Ghent to deny the possibility of an eternal creation. In support of his position, he offered twelve arguments, of which the first seven he said were demonstrative and the last five less than that. His first argument which went back beyond Bonaventure, at least to al-Ghazali, was to the effect that if the world had existed from all eternity there would have been by now an infinity of human beings. But since every human being has an immortal soul, there would at the present time exist an actual infinite number of such souls. But an actual infinite number is an impossibility. His sixth argument asks, if the world had been created from all eternity, where would the sun have been created? Would it have been created in the east or in the west, in the north or in the south? It could not have been created simultaneously in all these places. Neither could it have been created in one of these places to the exclusion of the others. Otherwise, we could obviously mark a first *terminus a quo* for the motion of the sun and thus a first moment of time. Hence the world could not have been created from eternity. His ninth argument is that if the world were eternal, by now (judging from the multiplication of the human race from Adam) there would be an actually infinite number of human beings on this earth. Again, apart from the obvious fact that this is not the case, such an actual infinite number would be impossible. His tenth argument is that if the world were eternal, the hours, days, months and years before now would all be infinite. And inasmuch as one infinity cannot be greater than another, all of these would be equal. But this would amount to saying that an hour would be equal to a day, to a year, or to a century - which is patently absurd.

It is difficult to reconcile Toletus' position here with the one he took on an actual infinite number when commenting on the First Part (Ia) of the *Summa theologiae*. In his *In Summam Theologiae* (q.7 a.4), Toletus treats the matter 'theologically', siding with Avicenna (see Ibn Sina §4), al-Ghazali (§2), and also William of Ockham, Gregory of Rimini and Gabriel Biel (§4), to oppose Aquinas and say that most probably such a number can be realized. In support of this he claims that God can create more and more human beings without limit and that such an unlimited multitude can actually exist together from eternity in the divine knowledge. It is further probable, he thinks, that God, by his absolute power, could create such an unlimited number all at once. In an *ad hominem* argument aimed at Aquinas, Toletus says that anyone who accepts the possibility of an eternal world should admit the possibility of an actual infinite number of immortal human souls. At the same time, consistent with his commentary on the *Physics*, Toletus does allow some probability to the opposite view that an actual infinite number is impossible. Linked with this he says that 'because I do not understand, I prefer to grant to, rather than take away from, the divine power what is a mark of perfection and what I do not understand' (*In Summam Theologiae* q.7 a.4).

5 Other works

Two other philosophical works published by Toletus at Venice in 1575 were in the form of commentaries with questions, on Aristotle's *On the Soul* and *On Generation and Corruption*. In the former work (which was reprinted twenty-two times by 1625), he allowed (in opposition to Aquinas) a direct intellectual cognition of a singular material thing. And although he thinks it 'more probable' that an agent intellect is necessary, he regards it as 'probable' that there is no agent intellect or that such an intellect differs only aspectually (*ratione*) from the possible intellect. Affirming the immortality of the soul, he held it as probable (again in opposition to Aquinas) that although a phantasm was necessary for the inception of intellectual knowledge, once such knowledge was acquired the intellect could dispense with phantasms, and thus could function after death in separation from the body. He also maintained that the intelligible species is not an efficient cause of cognition, but is at best only a condition required for the causality of the possible intellect. Again, apparently reflecting Duns Scotus, Toletus sharply distinguished the apprehension of a propositional synthesis from the judgment which either assents to or dissents from it. This last was to be a much debated item in seventeenth-century scholasticism.

See also: Aquinas, T.; Aristotelianism in the 17th century §§2-4; Aristotelianism, medieval; Aristotelianism, Renaissance; Language, Renaissance philosophy of; Logic, Renaissance

JOHN P. DOYLE

List of works

Toletus, F. (1533-96) *Opera omnia philosophica (Complete Philosophical Works)*, Cologne, 1615; repr. Hildesheim: Olms, 1985.(Contains Toletus' logical works from 1560 and 1572, plus his physical and

psychological works from 1575.)

Toletus, F. (1560) *Introductio in dialecticam Aristotelis (Introduction to the Dialectic of Aristotle)*, in F. Toletus, *Opera omnia philosophica*, vol. 1, Hildesheim: Olms, 1985.(An Aristotelian introduction to the logic of terms, propositions and argumentation.)

Toletus, F. (*c*.1563) *In Summam Theologiae S. Thomae Aquinatis Enarratio (An Exposition of Aquinas' Summa theologiae)*, Rome, 1869, 4 vols.(Toletus' partial exposition of the *Summa*, which remained an unpublished manuscript at Toletus' death.)

References and further reading

Ashworth, E.J. (1969) 'The Doctrine of Supposition in the Sixteenth and Seventeenth Centuries', *Archiv für Geschichte der Philosophie* 51: 260-85.(A specific doctrine of Toletus is treated within a wider context.)

Ashworth, E.J. (1974) *Language and Logic in the Post-Medieval Period*, Dordrecht and Boston, MA: Reidel.(A general study which includes discussion of Toletus.)

Charron, W.C. and Doyle, J.P. (1993) 'On the Self-Refuting Statement "There is no Truth": A Medieval Treatment', *Vivarium* 31: 24l-66.(Contains an exposition of Toletus' summation of and judgment on a long-standing scholastic discussion.)

Lohr, C.H. (1987) *Latin Aristotle Commentaries. II Renaissance Authors*, Florence: Olschki, 458-61.(Contains bibliographical information on Toletus.)

Risse, W. (1964) *Die Logik der Neuzeit. I Band. 1500-1640 (The Logic of the Modern Period)*, Stuttgart and Bad Cannstatt: Frommann, vol. 1, 382-5.(Toletus' logic treated within its historical context.)

Solana, M. (1940) *Historia de la filosofía española. Época del renacimiento (siglo XVI) (History of Spanish Philosophy. The Time of the Renaissance (16th Century))*, Madrid: Asociación Española para el Progreso de las Ciencias, vol. 3, 311-37.(An introductory overview of Toletus.)

Tolstoi, Count Lev Nikolaevich (1828-1910)

Tolstoi expressed philosophical ideas in his novels Voina i mir (War and Peace) (1865-9) and Anna Karenina (1875-7), which are often regarded as the summit of realism, as well as in shorter fictional works, such as Smert' Ivana Il'icha (The Death of Ivan Il'ich) (1886), often praised as the finest novella in European literature. In addition, he wrote numerous essays and tracts on religious, moral, social, educational and aesthetic topics, most notably 'Chto takoe iskusstvo?' ('What Is Art?') *(1898), Tsarstvo Bozhie vnutri vas (The Kingdom of God Is Within You) (1893) and his autobiographical meditation 'Ispoved'' (A Confession) (1884).*

Tolstoi apparently used his essays, letters and diaries to explore ideas by stating them in their most extreme form, while his fiction developed them with much greater subtlety. Critics have discerned a sharp break in his work: an earlier period, in which he produced the two great novels, is dominated by deep scepticism; and a later period following the existential trauma and subsequent conversion experience described in 'Ispoved''. Tolstoi stressed the radical contingency of events, valued practical over theoretical reasoning, and satirized any and all overarching systems. After 1880, he assumed the role of a prophet, claiming to have found the true meaning of Christianity. He 'edited' the Gospels by keeping only those passages containing the essence of Christ's teaching and dismissed the rest as so many layers of falsification imposed by ecclesiastics. Tolstoi preached pacifism, anarchism, vegetarianism, passive resistance to evil (a doctrine that influenced Gandhi), a radical asceticism that would have banned sex even within marriage, and a theory of art that rejected most classic authors, including the plays of Shakespeare and Tolstoi's own earlier novels.

1 Life

Born into a prominent aristocratic family, Tolstoi studied oriental languages and law at Kazan University, where he failed to earn a degree. He then tried to reform his estate according to philosophical principles, a topic which recurs in *Anna Karenina*, whose hero, Konstantin Levin, meditates on why such efforts almost invariably fail. A compulsive diarist, the young Tolstoi repeatedly formulated complex rules for behaviour, which he never followed, then devised rules for insuring obedience to the earlier rules, and at last reflected on the inadequacy of all rules or systems for an understanding of human behaviour, another key theme of his fiction. Enlisting in the army during the Crimean War, he wrote military stories (the Sevastopol sketches, 1855-6) concentrating on the difference between standard narratives, whether official, literary or historical, and the realities of combat. He later recalled that immediately after a battle, soldiers would narrate their experiences in a confused way and describe events as rife with contingency; but a month later, after official accounts had described a smooth series of military engagements in conventional language, the same soldiers would 'remember' events according to the received narratives. Such experiences led Tolstoi to a deep suspicion of narrative neatness, which rules out the contingency actually governing events, a theme central to *War and Peace*.

After the Crimean War, Tolstoi was hailed as a major writer but, always suspicious of the intelligentsia he deemed self-congratulatory, he joined no intellectual camp. In 1862 he married Sofia Andreevna Bers, and devoted the ensuing years to his marriage (he eventually had thirteen children) and the writing of *War and Peace*. In the 1870s he wrote *Anna Karenina* and then experienced the most significant of his many psychological crises, leading to the adoption of his own reformulated Christianity. His new ideas led to estrangement from his wife, and their marriage from this point on deteriorated and became complicated; it has been the subject of numerous, vaguely voyeuristic biographical treatments.

In addition to theological tracts and many essays, Tolstoi wrote endless diaries, which were evidently designed to be published and so resemble stage whispers. Regarded around the world as a sage, he corresponded with countless writers and thinkers, and his estate, Iasnaia Poliana, became a place of pilgrimage. His best fiction of the period deals with sexuality (in *Kreitserova sonata (The Kreutzer Sonata)* (1891)), death (*The Death of Ivan Il'ich*), and the complex psychology of sainthood (*Father Sergei* (1898)), but he also wrote one long novel, *Voskresenie (Resurrection)* (1899), usually regarded as a heavy-handed, didactic failure. Always perplexed by the contradiction between his wealthy life and the simple existence he preached, he fled Iasnaia Poliana and soon died in the course of his pilgrimage.

2 *War and Peace*

In his most lasting work, Tolstoi was deeply impressed by all those aspects of life that elude systematization or theoretical knowledge. He loved to formulate negative laws, dicta about what cannot be known or found in an unmixed state. ’Pure and absolute sorrow is as impossible as pure and absolute joy’, he writes in *War and Peace*, a sentence that reflects in its very style Tolstoian habits of thought.

In Tolstoi’s view, it is not primarily grand incidents, broad laws or dramatic decisions that shape individual or historical life, but the sum total of ordinary events. Life is made by ‘tiny, tiny alterations’ of action and consciousness, changes so small that they usually pass unnoticed, even though they are right before our eyes. What really counts is hidden in plain view, a sentiment that seems to have influenced Wittgenstein, who was an admirer of Tolstoi. At each small moment of existence, contingency reigns, not in the sense that anything can happen, but in the sense that more than one thing can happen. Time branches not only at critical moments, but always, and so our smallest decisions have moral value. Apparently inconsequential events may turn out to have great impact over time. Our decisions at critical moments are shaped by the climate of our minds, which in turn depends on the sum total of our smallest thoughts at odd moments.

No algorithm can reduce these small events to laws. With withering irony, Tolstoi, who was superb at logical analysis, detects the fallacies in attempts to construct systems that aspire to eliminate contingency. Theoreticians often assume that behind the apparent chaos of the social world reigns a hidden order, which may be expressed in a few simple laws; but Tolstoi insists that all such reasoning manages to rule out inconvenient facts and at crucial moments presumes what it hopes to prove. One example he cites for this is the Hegelian assumption that certain events not fitting a central story are ‘nonhistorical’. In *War and Peace*, he refers to reasoning by ‘stencil work’. Tolstoi suggests that the opposite assumption is closer to lived experience: the social world is fundamentally messy and order is always the result of work. It is regularity, not contingency, that requires an explanation.

In *War and Peace*, Tolstoi develops these insights in the philosophical reflections of his characters, in the events of their lives and in embedded essays, expressed not through a narrator but directly through the voice of the author. In this work, battle becomes a metaphor for history, and the radical uncertainty of combat, which only fools think can be reduced to a science of warfare, suggests the radical unpredictability of history. Tolstoi utterly rejects the possibility of laws of history and in several writings expressed particular animus to the idea of an underlying law of progress. One of the novel’s heroes, Prince Andrei, begins the novel believing in heroism, which he conceives as the ability to grasp the laws of battle and to act decisively upon them in dangerous moments. He eventually learns that battles cannot be shaped by strategy but are decided instead by ‘a hundred million diverse chances, which will be decided on the instant by whether we run or they run, whether this man or that man is killed’ - decided, that is, by events that are in principle unpredictable and chancy. The emphasis on what happens ‘on the instant’ reflects Tolstoi’s belief in the crucial importance of presentness and on the futility of assuming that only timeless principles represent the highest form of thinking. Tolstoi believed deeply in what Aristotle called *phronesis*, or practical reasoning (see Aristotle §23).

‘What science can there be’, Prince Andrei asks, ‘in a matter in which, as in every practical matter, nothing can be determined and everything depends on innumerable conditions, the significance of which becomes manifest at a particular moment, and no one can tell when that moment will come?’ Thus, the most effective soldier in the novel, Nikolai Rostov, cares nothing for theories but has developed the practical ability to grasp the shifting situations of each moment and, on the basis of unformalizable experience, to take advantage of unique and short-lived opportunities. The best generals in the novel, such as the Russian commander Kutuzov, know that they can be effective not by planning battles but by inspiring soldiers with the confidence to act intelligently. Kutuzov, in fact, falls asleep during councils of war because effective action depends on alertness. He explains that the best preparation for a battle, or any especially unpredictable and rapidly changing situation, is ‘a good night’s sleep’.

For a variety of reasons, historical narratives smooth out the contingencies of battles and of history. They therefore misrepresent the course of events, and, when taken as empirical knowledge on which to base future actions, lead to ill-conceived plans and to overconfidence in planning itself. In large part, the regularity of historical narratives reflects the essentially aesthetic need for a good story. Genres of historical writing differ, but each reflects a particular kind of plotting, all of which understate the radical uncertainty of the world precisely by telling a neat story. Plot contains implicit philosophy, and Tolstoi relentlessly teases out the hidden and fallacious assumptions behind various narrative forms. For reasons of mental economy, memory also works by filtering out the

contingent. Indeed, the fallacies of memory and narrative are present in initial acts of perception. In his novels, Tolstoi demonstrates these truths with portraits of events, their perception, their later appearance in memory and their encapsulation in the narratives characters and real historians construct.

In a series of draft introductions to *War and Peace*, Tolstoi evokes the image of a man viewing a distant hill on which only trees are visible. The man may conclude that the region in question contains nothing but trees. This conclusion would be a trick of perception, because smaller objects are simply not visible at a distance. In the same way, historians focus on unusual and big events because they are the most likely to be recorded. Part of the strategy of *War and Peace* is to contrast the picture of historians with a recreation of what events must have been like. We also commit this fallacy of perception in our individual lives, because memory does not preserve, or we cannot easily recall, small events that do not fit a pattern or story, but which may have been most effective and may even have constituted the essential quality of our lives. *War and Peace* is largely about the need to reverse some mental habits.

3 *Anna Karenina*

In both *War and Peace* and *Anna Karenina*, Tolstoi develops a theory of psychology. Perhaps indebted to Locke through the medium of Laurence Sterne's novel *Tristram Shandy*, Tolstoi conceives the mind as an agglomeration of mental habits, each the result of small actions taken many times. These habits do not cohere into a whole; wholeness, like moral integrity, is a project we undertake but never complete. At any given moment, the mind has many thoughts and sensations, too many for effective action if we did not have the capacity to focus our attention on one or a few of them. Numerous tiny alterations are taking place on the periphery of consciousness, much as small events are always taking place in history, and these have their effect. But we rarely notice them, precisely because by definition one cannot focus one's attention on what escapes one's attention. One can only glimpse them at moments when attention shifts, and Tolstoi illustrates this process in detail.

Tolstoi is also famous for exploring in meticulous detail the relation of the body to the mind. At the periphery of attention characters are dimly aware of bodily sensations but are usually unaware of how they affect the direction of their thoughts. Or a character may be smiling out of habit or because of a thought experienced some time earlier, with the smile simply forgotten on his face. Anyone trying to guess his current thoughts by that smile would be mistaken because the mind is never whole and because at any given moment is temporally layered. We are all palimpsests.

In both *War and Peace* and *Anna Karenina*, Tolstoi develops an approach to ethics that might be called casuistical in the root sense of the word. Theoretical reasoning, proceeding from general principles to the particular situation, tends to oversimplify events. We do better by educating an ethical sensibility through sensitive attention to particular cases as they arise throughout life and then, at each moment, trusting a well-developed ethical sensibility (see Moral particularism).

Anna Karenina applies these general ideas to social life and to the problem of reform. As *War and Peace* rejects the romanticized idea of great heroes directing history, *Anna Karenina* may be viewed as a polemic against the assumption that romantic love (the sort to be found in *Romeo and Juliet* and in countless works of popular culture) is the only kind. The tragedy of Anna derives from her acceptance of this myth. The wiser characters, such as Kitty, understand that love may also be prosaic and that a successful marriage must be based precisely on prosaic love, which cultivates not grand gestures but small acts of intimacy that do not make a compelling story. The novel's famous first sentence - 'All happy families resemble each other; each unhappy family is unhappy in its own way' - suggests that unhappy families each have a unique story, whereas happy family lives are composed of prosaic incidents that do not make a compelling story. Plot is an index of error and of a life lived badly.

The novel's hero, Levin, hopes to reform Russian agriculture by introducing Western machinery and applying the conclusions of social science. He discovers that successful reform never proceeds down from universal principles forced on a reluctant population, but by reasoning up from local conditions, which are always particular, and adopting measures that improve upon what already exists. We sense here Tolstoi's deep conservatism and scepticism about the abstract theorizing of intellectuals.

4 Aesthetics and late writings

For Tolstoi before his conversion, realistic novels were the supreme art form because they provide, more than any other literary or non-literary writing, a rich sense of the particularities of ordinary existence. Nevertheless, novels also mislead because they, like other forms of narrative art, offer too neat a picture. Precisely because the author knows the story in advance and plans everything to fit, novels misrepresent our temporal experience by letting a pattern of the whole dictate what happens and by ruling out contingency. Tolstoi experimented with various ways of avoiding this error, and so wrote *War and Peace* (and to a great extent *Anna Karenina*) without an advance plan, letting each serialized part develop potentials in earlier parts but with no overall end in view. In fact, *War and Peace* never really ends, it just breaks off.

In *Anna Karenina*, Tolstoi developed a general theory of art on which he expanded in *What Is Art?* and other writings. The novel's true artist, Mikhailov, paints not by applying techniques or by adhering to the tenets of any school, but by learning to see and record the particularities of experience. He is always observing small events of life and remembering them. Because everyone's experience is unique, he can be certain that his painting will convey an aspect of experience no one else has conveyed.

In *What Is Art?*, Tolstoi first offers a showcase demolition of all the received schools of aesthetics and then presents his own view. Like Mikhailov, the true artist conveys an experience so sensitively that he manages to 'infect' his audience with the feelings he himself had. But many so-called artists work quite differently, by applying a set of abstract techniques copied from other artists. Though gaudily 'interesting', their work is not true but 'counterfeit' art because it results from no particular and unique experience. True art may be divided into good and bad on moral grounds. A bad but true artist successfully infects his readers with pernicious moral feelings (see Art and morality §3). Tolstoi rejects most of the canon of Western art as either immoral (Maupassant) or counterfeit (Wagner) or both (Shakespeare). His favourite example of good, true art was the story of Joseph in the Bible.

By goodness Tolstoi came to mean the tenets of Tolstoian Christianity. He rejected all miracles, all sacraments and the divinity of Christ. Relying on the Sermon on the Mount (Matthew: 6-7), he took Christ's most important commandment to be 'non-resistance to evil', from which he derived a rejection of any institution based on force and so arrived at anarchism. He interpreted social and cultural institutions as so many curtains for concealing from ourselves our mortality (the theme of *The Death of Ivan Il'ich*) and for creating intermediaries between the practitioners of violence and ourselves, who benefit from it. He went so far as to argue that crime would cease if prisons and police were abolished. In these writings, Tolstoi seems far from the radical sceptic of his earlier years - and from occasional works of his old age, such as his novella *Father Sergei*, which reflects with evident self-irony on the distortions induced by aspirations to sainthood. Admirers of Tolstoi's thought are divided between those who think of him as the prophet of non-resistance and those who prefer his sense of the futility of all-embracing systems.

See also: Gandhi, M.K.; War and peace, philosophy of §4

GARY SAUL MORSON

List of works

Tolstoi [Tolstoy], L. (1929-58) *Polnoe sobranie sochinenii*, Moscow: Khudozhestvennaia literatura, 90 vols.(The authoritative edition of Tolstoi's works in Russian.)

Tolstoi [Tolstoy], L. (1967) *Great Short Works* trans. L. and A. Maude, New York: Harper & Row.(Contains *The Death of Ivan Il'ich* (1886), *The Kreutzer Sonata* (1891), *Father Sergei* (1898) and other works.)

Tolstoi [Tolstoy], L. (1865-9) *Voina i mir*, trans. A. Dunnigan, *War and Peace*, New York: Signet, 1968. (Discussed in detail in §2.)

Tolstoi [Tolstoy], L. (1875-7) *Anna Karenina*, trans. C. Garnett, ed. L. Kent and N. Berberova, New York: Modern Library, 1965.(Discussed in detail in §3.)

Tolstoi [Tolstoy], L. (1893) *Tsarstvo Bozhie vnutri vas*, trans. L. Weiner, *The Kingdom of God Is Within You: Or, Christianity Not As A Mystical Teaching But As A New Concept of Life*, New York: Noonday, 1961.

Tolstoi [Tolstoy], L. (1899) *Voskresenie*, trans. L. Maude, *Resurrection*, Oxford: Oxford University Press, 1994.

Tolstoi [Tolstoy], L. (1971) *A Confession*, *The Gospel in Brief and What I Believe*, trans. A. Maude, London: Oxford University Press.(From the original Russian texts of 'Ispoved'' (1884), 'Soedinenie i perevod chetyrekh evangelii' (1881) and 'V chem moia vera' (1884) respectively.)

Tolstoi [Tolstoy], L. (1961) *Recollections and Essays*, trans. A. Maude, London: Oxford University Press.

(Anthology containing ‘Why Do Men Stupefy Themselves?’ (1890), in which Tolstoi develops his theme of ‘tiny alterations’; his main critique of Shakespeare, ‘Shakespeare and the Drama’ (1906); ‘Non-Acting’ (1893); and the Gandhi letters (1910).)

Tolstoi [Tolstoy], L. (1969) *What Is Art? and Essays on Art*, trans. A. Maude, London: Oxford University Press. (Anthology containing *What Is Art?* (1898); ‘Introduction to the Works of Guy de Maupassant’ (1894); Tolstoi’s ‘Afterword’ (1905) to Chekhov’s story ’Darling’; and other essays on art.)

References and further reading

Berlin, I. (1970) *The Hedgehog and the Fox: An Essay on Tolstoy’s View of History*, New York: Simon & Schuster.(Views Tolstoi as torn between scepticism and dogmatism.)

Eikhenbaum, B. (1930) *Tolstoi in the Sixties*, trans. D. White, Ann Arbor, MI: Ardis, 1982. (By a leading Russian formalist, this book also examines Tolstoi’s relation to the thinkers of his time.)

Feuer, K.B. (1996) *Tolstoy and the Genesis of ’War and Peace’*, ed. R. Miller and D. Orwin, Ithaca, NY: Cornell University Press.(Considers the creation of the novel in relation to Tolstoi’s evolving thought about society.)

Gustafson, R.F. (1986) *Leo Tolstoy, Resident and Stranger: A Study in Fiction and Theology*, Princeton, NJ: Princeton University Press. (Outlines Tolstoi’s theological views and reads Tolstoi’s late ideas into his early fiction.)

Morson, G.S. (1987) *Hidden in Plain View: Narrative and Creative Potentials in ’War and Peace’*, Stanford, CA: Stanford University Press.(Discusses Tolstoi’s views on history, historiography and psychology in relation to the forms and themes of *War and Peace*.)

Tominaga Nakamoto (1715-46)

Tominaga Nakamoto was a leading representative of what some scholars have called the eighteenth-century 'enlightenment' movement in Tokugawa thought. Nakamoto's philological critiques of the historical development of Buddhist, Confucian and Shintōdoctrines are noteworthy for their modern, empiricist tendencies. His advocacy of makoto no michi, *or 'the True Way', a quotidian ethics advocating practical morality, gained no real following during Nakamoto's brief life.*

Nakamoto was the third son of Tominaga Hōshun, an Ōsaka *shōyu* (soya sauce) merchant who cofounded the Kaitokudōacademy. The neo-Confucian scholar Miyake Sekian (1665-1730), directed Nakamoto's education at the Kaitokudō. However, early on Nakamoto studied *kogaku*, or 'Ancient Learning', the philologically-oriented Confucianism of Itō Jinsai, OgyūSorai and Itō Tōgai. Those *kogaku* scholars had criticized neo-Confucians for misconstruing the original Confucian teachings. The latter, they generally agreed, conveyed a sagely wisdom worthy of faithful obedience.

Nakamoto surpassed contemporary *kogaku* schools by claiming that there was nothing sacrosanct about ancient Confucianism: it was one of several schools which appeared during an age of profuse philosophical activity in ancient China. After expounding this view in his *Setsuhei (Philosophical Obscurantism)*, Nakamoto was expelled from the Kaitokudō. The *Setsuhei* is no longer extant, suggesting that contemporaries saw it as a pariah work. Its ideas were recalled, however, in Nakamoto's *Okina no fumi (An Old Man's Notes)*.

In his *Shutsujō gogo (Buddha's Post-Enlightenment Teachings)*, Nakamoto similarly criticized the history of Buddhism, claiming that Mahāyāna teachings were later accretions, not legitimate accounts of the historical Buddha's sermons as Mahāyāna Buddhists claimed them to be. *Shutsujō gogo* explained the cardinal principle of Nakamoto's critical philology, *kajō*, or 'transformative accretion'. It asserted that throughout the Buddhist tradition new ideas had been attributed (*ka*) to the Buddha, effectively and illegitimately transforming (*jō*) his message time and again (see Buddhist philosophy, Japanese).

Nakamoto cautioned that language must be understood relative to three factors: the person using it, the historical period in which it is used, and the context of its use (see Language, philosophy of). Nakamoto also distinguished five contexts of meaning: (1) later explication, (2) the original teaching, (3) implicit significance, (4) heated debate and (5) rebuttal. Through hermeneutics based on these contexts, one could come to understand the history of philosophical thought.

In his last work, *Okina no fumi*, Nakamoto criticized Shintō, arguing that it had been invented by medieval Japanese to compete with Confucianism and Buddhism (see Shintō). Nakamoto faulted Shintō for its excessive secrecy, which contradicted his notion of *makoto no michi* (the True Way). While allowing that elements of Confucianism, Buddhism and Shintō were intrinsic to *makoto no michi*, Nakamoto suggested, contrary to his principle of *kajō*, that his 'True Way' was not just a later accretion.

Nakamoto's writings never circulated widely. Prior to the efforts of Professor Naitō Konan at Kyoto University in the 1920s, *Okina no fumi* was an obscure work even among specialists in Tokugawa thought. Nakamoto's *Shutsujō gogo* did, however, influence *kokugaku* (National Learning) scholars such as Motoori Norinaga and Hirata Atsutane in their rejection of Buddhism.

See also: Buddhist philosophy, Japanese; Confucian philosophy, Japanese; Shintō

JOHN ALLEN TUCKER

List of works

Tominaga Nakamoto (1745) *Shutsujōgogo (Buddha's Post-Enlightenment Teachings)*, in Mizuta Norihisa and Arisaka Takamichi (eds) *Tominaga Nakamoto/Yamagata Bantō*, Nihon shisōtaikei vol. 43, Tokyo: Iwanami shoten, 1975.(Edition of *Shutsujōgogo* in Sino-Japanese along with a classical Japanese translation and copious annotations.)

Tominaga Nakamoto (1746) *Okina no fumi (An Old Man's Notes)*, in Nakamoto Yukihiko (ed.), *AndōShōeki/Tominaga Nakamoto/Miura Baien/Ishida Baigan/Ninomiya Sontoku/KaihōSeiryō*, Nihon no shisō

vol. 18, Tokyo: Chikuma shobō, 1971.(Includes *Okina no fumi* in classical Japanese along with a modern translation and notes.)

References and further reading

Minamoto Ryōen (1973) *Tokugawa shisōshōshi*, Tokyo: Chūōkōronsha.(Focuses on Nakamoto's critiques of Confucianism, Buddhism and Shintō, arguing that Nakamoto's ideas relativized traditional currents of Japanese thought.)

Najita Tetsuo (1987) *Visions of Virtue in Tokugawa Japan: The KaitokudōMerchant Academy of Osaka*, Chicago, IL: University of Chicago Press.(Examines Nakamoto in the context of the Kaitokudōacademy, the *kogaku* movement and *kokugaku* thought.)

Tsunoda Ryusaku, de Bary, W.T. and Keene, D. (eds) (1964) *Sources of Japanese Tradition*, vol. 1, New York: Columbia University Press.(Discusses Nakamoto's 'historical relativism' as part of Japan's 'Eighteenth-Century Rationalism'. Includes translated excerpts of *Okina no fumi*.)

Tonghak

Tonghak is an indigenous religion in Korea. Founded by Ch'oe Cheu (1824-64), it presently flourishes under the new name of Ch'ôndogyo. An eclectic religion, Tonghak borrowed from Confucianism, Buddhism, Daoism, Christianity, shamanism and other folk beliefs. Its central tenet is founded on the concept of In nae ch'ôn *(Man is God). God is not a supernatural God who exists outside or beyond man, but is an immanent God who is present within every man.*

Disturbed by the unsettling political, social and economic conditions, and frustrated by his own unhappy family circumstances, the founder of Tonghak, Ch'oe Cheu (also known by his honorific name Suun), embarked on a long and tortuous search by means of meditation and study for a new way for the salvation of his fellow men. Suddenly, on the fifth day of the fourth month of 1860, he fell shivering into a trance, during which he heard a voice from Heaven. For a year thereafter, he experienced similar encounters from time to time. In 1861 he started to proselytize, based on the revelations he received. He named his teaching Tonghak (Eastern Learning) as a counter to Catholicism (known then as Sôhak, or Western Learning), which had made significant inroads in Korea. In 1864, he was executed by the provincial government on the charge of spreading false teaching to confuse the people.

The basic doctrine as taught by Ch'oe Cheu is embodied in the *Tonggyông taejôn (Bible of the Tonghak Doctrine)* and *Yongdam yusa (Hymns from the Dragon Pool)*, and was further clarified and refined by the leaders who succeeded him. From the moment of conception, according to Ch'oe Cheu, man is endowed with the mind of God and God is present within every man ever after. (In referring to God, Ch'oe Cheu used both the Sino-Korean word *Ch'ônju (Lord of Heaven)* and the native Korean word *Hanulim* (God) interchangeably.) Thus, God said: 'My mind is your mind (*Osim chûk yôsim*).' Accordingly, Tonghak emphasizes the oneness of God and man, in which man is a manifestation of God. Though God is omniscient and omnipotent, he can realize his will only through men. Hence there is room for evolution in this world (see God, concepts of).

Man must work, according to Ch'oe Cheu, to 'serve God (*si ch'ônju*)' by maintaining the upright mind and proper deportment (*susim chônggi*) in addition to the 'humanity, righteousness, propriety and wisdom that the former sage taught'. The essence of the Tonghak teaching, he said, lies in the three words: sincerity (*sông*), reverence (*kyông*) and faith (*sin*) (see Xin). But whereas in Confucianism, only the educated are capable of becoming 'superior men', in Tonghak all men without regard to their educational or social background can attain that status (see Confucian philosophy, Chinese).

Concomitant with the idea that Man is God is the concept emphasized by Ch'oe Sihyông (1829-98), the second leader of Tonghak, who maintained that because man is God, 'every man should be treated as a God (*sain yôch'ôn*)'. There can be no social distinction based on birth, education or any other factor. Ch'oe Sihyông said: 'Because man is God, all men are equal and there can be no discrimination. Any artificial distinction of men between high and low violates God's will.' Another concept that underlies Tonghak is the idea that 'all men evolve toward oneness' (*tonggwi ilch'e*). When people realize the truth that man and God are one, all will turn to this truth and evolve towards becoming one with God. The Tonghak teaching also contains a nationalistic element, in so far as Ch'oe Cheu called for 'protecting the nation and securing peace for the people' (*poguk anmin*) against the potential threat from Japan and the Western nations.

In addition to such theological and scholarly principles, Tonghak appealed to the uneducated masses by promising salvation via mystical experience (see Mysticism, nature of). It promised that those who practised the true faith would have their diseases cured, and that the ritual incantations of certain sacred formulas and the possession of a prescribed talisman would give them supernatural power to ward off evil and sickness. It also assured faithful followers freedom from worldly oppression and suffering. The Tonghak founder then urged his followers to work for 'the salvation of all people' (*Kwangje ch'angsaeng*).

Attracted by its humanitarian and egalitarian teaching, many people at the grassroots level accepted Tonghak, especially peasants. In spite of harassment and persecution by the government, its strength grew rapidly and it became a powerful social and political force in Korea. In 1894, Tonghak followers played a central role in leading

the great insurrection known as the Tonghak rebellion that almost brought down the Chosôn dynasty and touched off the first Sino-Japanese war. In 1905, under the leadership of Son Pyônghûi, Tonghak was reorganized, adopting the new name Ch'ôndogyo (Religion of the Heavenly Way). Throughout the Japanese colonial period, Ch'ôndogyo remained strongly nationalistic, playing a leading role in the March First movement in 1919. In addition to being a religious organization, Ch'ôndogyo has played an active role in promoting Korea's national, social and cultural causes. Today it has about 600,000 followers.

See also: Confucian philosophy, Korean; Religion, philosophy of

YÔNG-HO CH'OE

References and further reading

Ch'ôndogyo paengnyôn yaksa (Abridged One Hundred Year History of Ch'ôndogyo) (1981), Seoul: Mirae munhwasa.(An official history compiled by a special committee of the church, to be published in two volumes.)

Hong Chang-hwa (1990) *Ch'ôndogyo kyori wa sasang (Doctrine and Thought of Ch'ôndogyo)*, Seoul: Ch'ôndogyo Chungang Chongbu Ch'ulp'anbu.(An explanation of the basic doctrine of Ch'ôndogyo in simple Korean.)

Kim Yong Choon (1972) 'The Ch'ôndogyo Concept of the Origin of Man', *Philosophy East and West* 22: 373-84. (Examines the original and unique nature of its concept with a contrasting Judaeo-Christian perspective.)

Kim Yong Choon (1973) 'The Ch'ôndogyo Concept of the Nature of Man', *International Philosophical Quarterly* 13: 209-28.(Analyses the basic concept of 'Man is God'.)

O Chiyông (1940) *Tonghak sa (History of Tonghak)*, Seoul: Yôngch'ang Sôgwan.(An authoritative history by an active participant.)

Shin, S.S. (1978) 'The Tonghak Movement: From Enlightenment to Revolution', *Korean Studies Forum* 5: 1-79. (A good study of how the obscure and apolitical religion of Tonghak was transformed to play a leading role in the nationwide uprising of 1894.)

Shin, S.S. (1979) 'Tonghak Thought: The Roots of Revolution', *Korea Journal* 19 (9): 11-20.(A scholarly study of the basic doctrine of Tonghak thought tracing the formulation of its founder's teachings.)

Weems, B.B. (1964) *Reform, Rebellion and the Heavenly Way*, Tucson, AZ: University of Arizona Press.(A still useful English history of Tonghak from its beginnings until 1950.)

Yi Ton-hwa (1933) *Ch'ôndogyo ch'anggôn sa (History of Founding of Ch'ôndogyo)*, Seoul: Ch'ôndogyo Chungang Chongriwôn.(An instructive account of Ch'ôndogyo's early history by one of its active members).

Totalitarianism

A term adopted in the 1920s by the Italian philosopher Giovanni Gentile to describe the ideal fascist state, 'totalitarianism' quickly acquired negative connotations as it was applied to the regimes of Hitler in Germany and Stalin in the USSR. Within political science it has generally been used to refer to a distinctively modern form of dictatorship based not only on terror but also on mass support mobilized behind an ideology prescribing radical social change. Controversially, the specific content of the ideology is considered less significant than the regime's determination to form the minds of the population through control of all communications.

Totalitarianism has attracted the attention of philosophers as well as political scientists because a number of classic philosophical systems have been suspected of harbouring totalitarian aspirations, and also because the model of total power exercised through discourse has been used by critical theorists to mount an attack on modernity in general.

1 The fascist doctrine of totalitarianism

The concept of totalitarianism began its career in Italy in the 1920s as an ideal formulated for Mussolini's Fascist regime by Giovanni Gentile (§3) (see Fascism). Gentile, whose idealist philosophy was deeply indebted to Hegel, argued that the state is an all-embracing ethical reality, that state and individual are inseparable and that only in and through the state can individuals realize their true freedom. Fascism, to Gentile, was essentially a spirit that ought to manifest itself in every aspect of life. In 'The Doctrine of Fascism', published under Mussolini's name in 1932, he declared: 'For the Fascist all is in the state and nothing human or spiritual exists, or much less has value, outside the state. In this sense, Fascism is totalitarian' (Gregor 1969: 223). For a while the connotations of the term were favourable, sufficiently so for it to be used approvingly by the Marxist thinker Antonio Gramsci to describe the Communist Party. However, it took on different senses upon being borrowed to describe the more formidable dictatorships of Hitler and Stalin, and was developed into an ideal-type within political science.

2 Ideological domination

According to the classic theories put forward after the Second World War by Hannah Arendt in *The Origins of Totalitarianism* (1951) and by Friedrich and Brzezinski in *Totalitarian Dictatorship and Autocracy* (1967), 'totalitarianism' means a distinctively modern type of dictatorship of which Nazism and Stalinism are the most prominent examples. Power within such a state rests not only upon the control exercised by a single party over all aspects of social organization, nor even merely upon the terror wielded by the secret police. An essential aspect of totalitarian rule is ideological domination, guided by an official ideology aimed at radical transformation. Whether this ideology appears to be right-wing or left-wing is beside the point. Differences of content between the racist ideology of Nazism and Stalin's Marxism-Leninism seemed to the theorists of totalitarianism to be less significant than the structural similarities between two regimes that attempted to mould the thoughts of their subjects into conformity with official ideology. Criticism of the concept of totalitarianism has focused particularly upon this central claim that regimes whose ideologies appear to be diametrically opposite are in fact essentially similar.

Friedrich and Brzezinski regard all ideological one-party regimes as totalitarian, whereas Arendt's use of the concept is more complex and is restricted to a few particularly extreme cases. For Arendt, the most characteristic totalitarian institution is the concentration camp, which exemplifies the pursuit of total power by dehumanizing its victims to the point where the human attributes of plurality and spontaneity have been destroyed. Totalitarian ideology is the intellectual expression of this, since it allows no room for individual thought. The key feature of these regimes for Arendt was not simply the adoption of an ideological principle (such as the claim that class struggle or race war is the key to history), but rather the relentless working out of that principle to its logical conclusion in the extermination of whole peoples or classes deemed to be 'objective enemies'. At the heart of totalitarianism, according to Arendt, was the blind pursuit of logical reasoning unrestrained by common sense. In her view, atomized individuals in modern mass societies are particularly vulnerable to the appeal of this kind of ideological reasoning.

3 Philosophy as a source of totalitarianism

If there is an affinity between totalitarianism and logic, if totalitarian regimes are distinctively ideological

dictatorships, and if the concept itself began life as a philosophical ideal, does philosophy itself (or at any rate certain philosophers) bear some responsibility for the horrors of the twentieth century? This claim was advanced most notoriously by Karl Popper (§4), who argued in *The Open Society and Its Enemies* (1945) that Plato in the ancient world and Hegel and Marx more recently had harboured the seeds of totalitarianism, opposing the ‘open society’ in which critical thinking, freedom and progress can flourish. Popper concentrated his attack particularly on ‘historicism’, or the claim that there are laws of history which can be discovered, enabling one to predict the future (see Historicism §2).

As the most notable philosopher of history, Hegel (§8) has been especially liable to be accused of being a progenitor of totalitarianism, particularly since his influence can be clearly traced both in Marxism-Leninism and in Gentile’s fascism (although not in Nazism). Students of Hegel maintain that this imputation is unjust, and even that Hegel should be seen as a liberal. It should be noted, however, that Gentile was also a liberal who considered that since true freedom is to be found through the ethical state, fascism was the consummation of liberalism. Isaiah Berlin’s attack in ‘Two Concepts of Liberty’ (1969) on ‘positive’ concepts of freedom was largely directed at the use made by fascists and Marxists of Hegelian ideas (see Berlin, I.).

The notion of any *direct* connection between totalitarian atrocities and philosophic indiscretions was dismissed by Hannah Arendt, who argued that the advent of totalitarianism owed more to practices developed by imperialist regimes than to abstract theories. Nevertheless she maintained that philosophers from Plato to Marx and Heidegger (a prominent supporter of Nazism) had facilitated totalitarianism because they had failed to appreciate human plurality. She suggested that this failure to notice the political implications of plural actions and multiple points of view might be an occupational hazard of philosophic solitude.

Another political thinker prompted by bitter experience to criticize intellectual hubris was Albert Camus, who argued in *The Rebel* (1951) that totalitarianism was the ultimate consequence of humanism gone sour, of a metaphysical rebellion that had culminated in the attempt to deify man.

4 Totalitarianism and the critique of modernity

Although the notion of totalitarianism has been applied primarily by theorists living in comparatively ‘open’ societies to regimes that have been manifestly terroristic, this model of domination in and through ideas and discourse has also been turned into a stick with which to beat liberal societies themselves. In *One-Dimensional Man* (1964), Herbert Marcuse argued that the apparent freedom of individuals in modern capitalist societies was illusory, because the desires and choices of consumers were moulded by cultural forces that served the purposes of the system of production while failing to satisfy the real needs of the population. ‘Under the rule of a repressive whole, liberty can be made into a powerful instrument of domination’ (Marcuse 1964: 21). From this point of view, the fact that people in the USA could be controlled without the use of terror made the USA even more effectively totalitarian than more overtly repressive regimes, and made liberation less likely.

Marcuse’s Frankfurt School colleagues Max Horkheimer and Theodor Adorno had already discovered totalitarianism deeply embedded in modern civilization. In *Dialectic of Enlightenment* (1973), they claimed that the project of liberating mankind by developing instrumental reason, science and technology had given rise first to the domination of nature and then (by a dialectical process) to the enslavement of mankind. On this deeply pessimistic view, totalitarianism is the direct outcome of enlightenment (see Critical theory §5). A generation later Michel Foucault, focusing on the development during the Age of Reason of prisons and other disciplinary institutions, also discovered totalitarianism at the heart of liberal modernity, and set out to unmask the power inherent in the production of truth and the discipline exercised through scientific disciplines (Foucault 1977). As in the cases of Horkheimer and Adorno and of Marcuse, power is diffused throughout Foucault’s ‘disciplinary society’ rather than being concentrated in the hands of the dictator. Even in Arendt’s version, however, the apparently all-powerful dictators have made themselves slaves of suprahuman forces. Common to many theories of totalitarianism is a sense that the imprisoning structures and destructive processes have been created by human beings but have escaped from human control. This nightmare forms a sharp contrast to the original philosophical vision of a totally intelligible and totally harmonious society.

MARGARET CANOVAN

References and further reading

Arendt, H. (1951) *The Origins of Totalitarianism*, New York: Harcourt Brace.(A classic text of the theory of totalitarianism.)
Berlin, I. (1969) 'Two Concepts of Liberty', in *Four Essays on Liberty*, Oxford: Oxford University Press.(Finds totalitarian implications in 'positive' concepts of liberty.)
Camus, A. (1951) *The Rebel*, trans. A. Bower, London: Hamish Hamilton, 1953.(Traces totalitarianism to intellectual hubris.)
Foucault, M. (1977) *Discipline and Punish*, London: Allen Lane.(Study of the development of prisons and organizations for social control, and of the growth of disciplinary society as a whole.)
Friedrich, C.J. and Brzezinski, Z. (1967) *Totalitarian Dictatorship and Autocracy*, New York: Praeger.(The classic text of the theory of totalitarianism as an ideal type in political science.)
Gregor, A.J. (1969) *The Ideology of Fascism: The Rationale of Totalitarianism*, New York: Free Press.(Includes an account of the ideal of totalitarianism.)
Harris, H.S. (1960) *The Social Philosophy of Giovanni Gentile*, Urbana, IL: University of Illinois Press.(Includes a comprehensive account of Gentile's philosophy of fascism.)
Horkheimer, M. and Adorno, T. (1973) *Dialectic of Enlightenment*, London: Allen Lane.(Argues that the roots of totalitarianism are to be found in instrumental reason.)
Marcuse, H. (1964) *One-Dimensional Man*, London: Routledge & Kegan Paul.(Depicts advanced capitalist societies as totalitarian.)
Popper, K. (1945) *The Open Society and Its Enemies*, London: Routledge, 2 vols.(A denunciation of Plato, Hegel and Marx as proto-totalitarians.)

Tradition and traditionalism

Tradition is that body of practice and belief which is socially transmitted from the past. It is regarded as having authority in the present simply because it comes from the past, and encapsulates the wisdom and experience of the past. For some, the very idea of tradition is anathema. It is characteristic of modernity to reject the authority of the past in favour of the present deployment of reason, unencumbered by tradition or prejudice. While prior to the seventeenth century tradition was largely unquestioned as a source of insight, and in need of no defence, since the Enlightenment the notion of tradition has been defended by traditionalists such as Burke and, more recently, Hayek. Upon inspection, however, traditionalism, if not indefensibly irrational, turns out to be a demonstration of the overlooked rationality contained within traditions. Traditions often turn out upon inspection to be not so much irrational as subtle and flexible deployments of reason in particular spheres.

1 Antitraditionalism

According to Karl Popper, a prevalent modern attitude to the past holds that 'I am not interested in tradition. I want to judge everything on its own merits… quite independently of any tradition… with my own brain, and not with the brains of other people who lived long ago' ([1949] 1963: 120-1).

Have there ever been people who thought in this way? In *The Frogs* Aristophanes depicts Euripides as wanting to teach his audiences to use their own brains, unfettered by meaningless traditions, and uncowed by authoritarian windbags who simply spout outdated myths. But we should remember that Aristophanes was hostile to Euripides. Furthermore, Euripides' plays certainly exploit old myths, and not always with new intentions.

In the eighteenth century, Diderot insisted that a true philosopher would 'trample underfoot prejudice, tradition, venerability' and admit nothing 'save on the testimony of his own reason and experience', possibly oblivious of the fact that that doctrine itself had by 1750 become well embedded in the philosophical tradition to which he belonged (Wilson 1972: 237). Critics of the French Revolution, such as Burke and de Maistre certainly saw the revolutionaries as attempting to break with tradition, but again, in view of their constant invocations of ancient Roman and republican virtue, it is doubtful that the revolutionaries conceived themselves as operating independently of any tradition. In the twentieth century, Keynes, along with many others, saw traditional morality as a crippling burden. Every decade or so artistic avant-gardists have called for the need to create anew, 'burning one's path behind one' (as the Russian constructivist Malevich put it). In retrospect, of course, the reality is that antitraditionalism is always selectively antitraditional, and that, as Popper says, whether we like it or not, we always stand on the shoulders of at least some of our predecessors.

2 Burkean traditionalism

Nevertheless since the end of the eighteenth century, in face of explicit verbal attacks on traditions, attacks on respect for tradition and successful assaults on parts of actual traditions, a line of thinking has developed which defends the virtues of tradition against its critics. In this explicit defence we encounter what might be called traditionalism, an attitude of mind intended to curb the pretensions of present reason to criticize traditional beliefs, institutions and practices.

One of the first and certainly one of the most famous expositions of traditionalism is that of Burke.

> We are afraid to put men to live and trade each on his own private stock of reason; because we suggest that this stock in each man is small, and that the individuals would do better to avail themselves of the general bank and capital of nations and of ages.
>
> ([1790] 1967: 84)

Burke argues that old prejudices should be cherished both because they are prejudices, and because they are old. Prejudices are easily applied in emergencies, and they can also become habitual, rendering, as he says, a man's virtue his habit. He also argues that if they are examined sympathetically, they will generally be found to contain 'latent wisdom', encapsulating the wisdom of the ages and of much experience, which would be likely to elude us if we relied solely on our present stock of reason (see Burke, E.).

In Burke's approach to tradition there is not a little hint of Adam Smith's invisible hand, transposed to the moral and political sphere; that is, reason as well as economic equilibrium can emerge through the actions and decision of individuals who have no conception of the greater order to which they contribute (see Smith, A.). By a process resembling natural selection, fruitful lines of behaviour will be rewarded and reinforced and eventually embedded in society as a whole, although the individuals concerned may be quite unaware of this aspect of their decisions, and be taking them for reasons quite other than the evolutionary advantage they bring to the society to which they belong.

3 Hayek's evolutionary account of tradition

In the twentieth century the conception of a tradition in terms of spontaneously developing orders has been associated particularly with the work of Hayek (§3). Hayek points out that societies with successful economies have reproduced and spread themselves more than those with unsuccessful economies. Along with economic success the rules of behaviour which underlie or accompany success have also been reproduced. In particular, in Hayek's view, the institutions of the family and private property have not just been an irrelevant accompaniment to economic success; they have been central to the market order. But in market societies, these institutions have usually been promulgated and defended for religious reasons, and as the societies have prospered, so has their religion. Conversely, societies with belief systems prejudicial to the market order have tended to die out, their belief systems dying with them (see Hayek 188).

While Hayek's conception of tradition might help throw light on unsuspected links between a society's historical success and apparently unrelated aspects of its beliefs and practices, it is less clear how far what he says can amount to a defence of the traditional as such. He and Burke can certainly be read as counselling us against premature attempts to overthrow traditions, warning us to examine the unsuspected consequences of so doing in greater depth than would a Diderot or a Keynes. It is also true that any explicit judgment we make will always be made against a background of unspoken and largely traditionally based agreement, again to a greater extent than many rationalists might suspect. But, at the end of the day, when confronted by any explicit controversy, we still must make a decision, and the mere fact that something has always been done is a less than compelling reason for continuing to do it. For example, we need to know whether its always having been done has had good or bad consequences. The simple fact that a particular religion's morality is conducive to the market order is not, in itself, a reason for a sceptic to accept that religion, even if they were concerned to promote or promulgate the market. The sceptic would have to accept or reject the religion in the light of its credibility, and not in the light of its supposed contribution to social wellbeing. A sceptic may even lament their inability to go along with a traditional belief or practice precisely because they are aware of its incidental benefits.

Hayek is particularly critical of modern thinkers who find market arrangements unjust, and who believe that they can improve on them by use of their own reason. He argues that experience of the comparative success of planned and unplanned economies shows that tradition is in some respects superior to conscious human reason. In reality, it demonstrates no such thing. What experience shows is that adherence to some traditions is sometimes better than attempting to improve or replace them. And if, following Hayek, we decide to defend and develop the traditions central to the market order, this will not be because of our love of tradition as such, but because we have reasoned that experience has shown one particular tradition has superiority over others. The fact that tradition suggests limits to the scope of rational planning and advances strategies for encouraging unplanned economic activity does not mean that a decision in its favour is either irrational or purely tradition-based. So while a prejudice in favour of tradition need not be irrational, particularly if associated with a Burkean attempt to reveal its latent wisdom, this hardly amounts to a full-blooded commitment to tradition as such. Traditionalism, then, is itself either irrational, or a rather sophisticated form of rationalism, a prejudice in favour of tradition, but for good reason in so far as the traditions in question can be shown to be reasonable.

4 The flexibility of traditions

Flexibility as a characteristic of traditions would not necessarily prove unacceptable to traditionalists, who usually emphasize such a flexible nature, the way in which traditions respond to circumstances of various sorts. One of the circumstances to which traditions respond is the reasoning and discussion their adherents engage in for both internal and external reasons. There can be no denying that a long-lasting tradition of belief, such as Roman Catholicism, has undergone many developments and changes of emphasis over the centuries, so much so that one

might be tempted to wonder just how much there is in common between a twentieth-century bishop and one of the Apostles. It was this question among others which Cardinal Newman addressed in his 1845 *Essay on the Development of Christian Doctrine* (see Newman, J.H.). Although specifically dealing with theology and the Church, Newman's essay is a profound analysis of the continuity through discontinuity present in any long-lasting tradition, with implications for any field of human endeavour which manifests creative interplay between inherited tradition, rational reflection and the wider social circumstances in which it is located. Doubtless many fields do manifest such creative interplay, although it is noteworthy that visual art has largely turned its back on its past achievements. Newman shows that the success of a tradition is related to its ability to assimilate new data, while conserving its past principles and achievements, and also to its ability to develop complex sequences of thought and practice while anticipating future development. He brings to the study of tradition a subtlety and a comparative perspective often lacking in the blanket statements of self-professed traditionalists and antitraditionalists alike.

See also: Conservatism

ANTHONY O'HEAR

References and further reading

Burke, E. (1790) *Reflections on the Revolution in France*, London, Dent, 1967.(The classic statement of the argument for tradition.)

Hayek, F.A. (1988) *The Fatal Conceit*, London, Routledge.(The source for the discussion of Hayek in §3. Gives an evolutionary account of the importance of apparently unthinking traditions.)

Newman, J.H. (1845) *Essay on the Development of Christian Doctrine*, London: Longmans Green, 1890. (Analyses the continuity in discontinuity in a long-lasting tradition. Referred to in §4.)

O'Hear, A. (1992) 'Criticism and Tradition in Popper, Oakeshott and Hayek', *Journal of Applied Philosophy* 9 (1): 65-75.(Explores the interplay of tradition and reason in the three cited authors.)

Popper, K.R. (1949) 'Towards a Rational Theory of Tradition', in *Conjectures and Reflections*, London: Routledge & Kegan Paul, 1963, ch. 4.(The work referred to in §1. What its title suggests: an attempt to defend traditions against its rationalistic opponents and its conservative friends.)

Shils, E. (1981) *Tradition*, London: Faber & Faber.(An analytical treatment of tradition, defending it against those who regard it simply as a barrier to progress.)

Taylor, C. (1990) *Sources of the Self*, Cambridge: Cambridge University Press.(Explores the development of rationalism in the modern world.)

Wilson, A. (1972) *Diderot*, Oxford: Oxford University Press.(Diderot gives voice to the classic Enlightenment hostility to tradition.)

Tragedy

Tragedy is primarily a type of drama, though non-dramatic poetry ('lyric tragedy') and some novels (for example, Moby Dick) have laid claim to the description. As a genre, it began in ancient Greece and forms a part of the western European tradition. Historically, it has carried prestige for playwrights and actors because it dealt with persons, generally men, of 'high' or noble birth, who, by virtue of their stature, represented the most profound sufferings and conflicts of humanity, both morally and metaphysically. The history of the genre is part of the history of how art and culture reflect views about class and gender.

Tragic theory has concentrated primarily on how to define the genre. A persistent feature is the tragic hero, who begins by occupying a position of power or nobility, but comes to a catastrophic end through some action of his own. According to the Aristotelian tradition, the audience is supposed to experience pity and fear in response to the sufferings of the tragic hero, and perhaps pleasure from its cathartic effects. Hegel initiated a paradigm shift in tragic theory in proposing that tragic plots essentially involve conflicts of duty rather than suffering.

Greek tragedy and Shakespearean tragedy provide two different exemplars of the genre. The tradition inspired by Greek tragedy emphasized a rigidly defined genre of dramatic poetry; French neoclassic tragedy is part of this tradition. Shakespearean tragedy, on the other hand, is written partly in prose, and includes comic elements and characters who are not nobly born. Lessing and Ibsen also resisted restraints imposed on the genre in terms of its representation of social class and gender in favour of drama that was more realistic and relevant to a bourgeois audience. Twentieth-century criticism has questioned the viability of the genre for modern times.

1 Greek tragedy to Plato

Greek tragedy sprang from dithyrambs (choral songs to Dionysus); when dialogue was included in these, tragedy developed as a distinct genre. In about 534 BC the poet Thespis is said to have introduced dialogue between the chorus (or its representative) and a choral leader, and in about 500 BC Aeschylus introduced a second actor, thereby making dialogue independent of the chorus. Either could be seen as the originator of tragedy. Tragedy was considered to be a kind of poetry; poets, and hence tragedians, were regarded as teachers of morality and religion, and tragedies were performed at annual festivals rich with civic and religious significance. The classic tragedies of Aeschylus, Sophocles and Euripides, all written in the fifth century BC, dramatize well-known myths involving important and powerful families, sometimes in altered forms. There were no female tragedians and all roles were performed by men, including those of female characters. Tragedy arose out of ritual, and the exclusively male participation in ritual reinforced a gender-linked social hierarchy.

Gorgias of Leontini, Sicily, held that tragedy (and other poetry) produced pleasure through deception. He also recognized the two main emotions generally associated with tragedy - pity and fear - though it remains unclear why precisely these two were singled out as characteristic of the genre. Plato argued that poetry was imitation, not necessarily deception. Nevertheless, he argued that writing, performing and witnessing performances of poetry could have undesirable, damaging effects on our ability to reason. In his early dialogue *Ion*, he argued that one writes good poetry and performs (recites) it well by 'divine inspiration', not rational understanding. He also argued that our emotional responses to poetry are not rational.

Plato refused to allow imitative or mimetic poetry (as opposed to narrative poetry) into the Republic, his ideal state. One reason was that poetry often portrays things that are not true and describes gods doing things that are not good models for behaviour (*Republic* 387b *et passim*). Another was that most people enjoy giving vent to emotions; we do not enjoy representations of intelligent and temperate dispositions (*Republic* 604e), but what we do enjoy vicariously tends to become part of ourselves (*Republic* 606b). The audience's vicarious pleasure in experiencing appetites (for example, sex) and pains and pleasures 'of the soul' (anger) strengthens those appetites when they ought to be weakened and brought under the control of reason (*Republic* 606d).

2 From Aristotle to Rome

Aristotle's *Poetics* is probably the single most influential work on tragedy ever written (see Aristotle). In contrast to Plato, Aristotle argued that writing tragedy required an understanding of the functions of tragedy and of the principles that ensure that those functions are fulfilled. Writing good tragedy thus develops reason. He also argued

that some pleasure gained from tragedy derives from learning about the things imitated, so that responses to it are cognitively rewarding.

Aristotle's famous definition of tragedy at the beginning of chapter six of the *Poetics* differentiates it from other types of imitation (painting, music, dithyramb, epic, comedy): 'Tragedy is an imitation of an action that is serious, complete, and possessing magnitude; in embellished language; …in the mode of action and not narrated; and effecting through pity and fear the catharsis of such emotions' (1449b24-28). Aristotle also held that tragedy has an end or purpose (*telos*), but there is controversy about what this is. Is its purpose to produce pity and fear, a catharsis, or pleasure (through imitation, pity and fear, and/or catharsis), or is it possibly to construct a plot of the kind which produces the appropriate effect? Aristotle clearly held that only certain kinds of plots will produce the emotions appropriate to tragedy. They are complex plots, involving recognition (of the tragic implications of the action) and reversal (of fortune). Tragic characters must be presented as better people than most and as agents actively affecting the course of action through their own choices, but operating in a world beyond their control.

The suffering that elicits pity and fear comes about through the actions of an agent who makes some kind of mistake or error in judgment (*hamartia*). Even very good people can make mistakes when subjected to forces beyond their control: we thus pity them (since pity is the appropriate response to undeserved misfortune) and fear for them and for ourselves (since if it can happen to people better than ourselves it can certainly happen to us). Aristotle allowed that tragedies could even have happy endings, on the grounds that the *threat* of suffering is capable of generating the 'tragic emotions'.

Historically, catharsis has loomed large in discussions of Aristotle's conception of tragedy, though the passage quoted above contains the *Poetics*' only occurrence of *katharsis* in the relevant sense, and gives no explanation of what it means. *Katharsis* has in fact been translated as both a purging and a clarification. The 'homeopathic' account, in which emotional responses to tragedy purge us of their harmful effects, is probably the most popular. Other accounts of catharsis hold that emotions are rendered pure and clear, so that knowledge of them as psychological states is enhanced; emotions are clarified in the sense that we feel them in response to their proper objects, that is, we pity and fear the right kinds of things; and emotional responses clarify our understanding of the structure of the plot (see Katharsis).

Tragedies were written during Roman times, notably by the Roman Stoic politician and philosopher Seneca. He emphasized the nobility of suffering and gave little attention to action. His works were filled with rhetorical conceits, and are now generally held to be affected, sentimental and bombastic. Horace's *Ars Poetica* proposed that poetry (including tragedy) is *utile dulce*, 'delightful instruction': poetry is seen here as a form of rhetoric that is styled to give pleasure and moral instruction.

3 Italy and France

Seneca's influence was significant, especially in fifteenth- and sixteenth-century Italy and in seventeenth-century France. In Italy, the sixteenth century saw the development of 'true' tragedy, with the playwright Torquato Tasso emerging as the major figure. Though animated by religious ideals, he tried to adhere to the prescriptions of Aristotle and Horace. French tragedy of this period consisted of a great deal of declamation and little action, and is widely regarded as an insignificant precursor to the neoclassical drama of Corneille and Racine, which also showed Senecan influence.

From the time of its first Latin translation in 1498, Aristotle's *Poetics* had substantial influence. Inspired by Aristotle, Bernardino Daniello formulated the Doctrine of Fixed Forms (1536) and Lodovico Castelvetro invented the three unities of time, place and action (1576): the action must take place during a single day, in a single place, and there should be only one main plot with no subplots. Castelvetro also argued that pleasure is tragedy's proper end; if instruction were its end, it would be a utilitarian art. Other rules acquiring more limited acceptance were that the ending should be unhappy, the performance time should be the same as the time of the action, there should be five acts, and no death or violence should take place on stage. Culminating with Nicholas Boileau-Despréaux (*L'art poëtique*, 1674), French theorists entrenched the rules developed by sixteenth-century Italians as absolute and unchanging standards for the production and evaluation of tragedy, though exactly which rules and how absolute their status became issues of continuing controversy.

Corneille and Racine were rivals for the title of pre-eminent dramatist of their day. Racine was best known for his

literary craft, but Corneille, whose most notable tragedy was *Le Cid*, also wrote essays (*Discourses*). Corneilleargued that we should recognize that many tragedies are effective even though they do not obey strict Aristotelian rules. However, it was still the function of poets to please according to the rules of their art, and their art included more than the three unities: for example, the tragic hero must be noble, plays were to be entirely in verse, there were to be no representations of violence on stage and no more than three speaking characters on stage at once. There was, moreover, more concern with general character types than the psychology of individuals. Plays were still declamatory in style, and extolled the virtues of reason or will in the face of the temptation to be ruled by passion. They were thus part of the Cartesian rationalist spirit, following up Plato’s praise of reason and critiques of pleasure and emotion.

4 England

In *The Monk’s Tale*, Geoffrey Chaucer provided a famous definition of tragedy as a (narrative) story of one who falls from prosperity into misery (teaching us not to trust our temporary good fortune). Sir Philip Sidney (*An Apology for Poetry*, 1598) argued that emotional responses to tragedy are morally useful because they move us to want to know and do what is good. He held that our admiration and commiseration teach us about the uncertainties of this world, but that fear is felt only by royal spectators!

The tragedies of William Shakespeare are now viewed as hallmarks of the genre. In a major break from both Greek and Roman tragedy and the morality plays of the Middle Ages, Shakespeare’s plays explore the psychology and actions of individuals, rather than general character types. Writing and acting were still male preserves, and women did not act female roles in England until the Restoration.

One topic of recurrent speculation among English theorists is the so-called paradox of tragedy: how and why do we enjoy being moved by scenes of pity, fear, suffering, and distress? Thomas Hobbes, true to his psychological egoism, suggested that the pleasure is of the thank-God-it’s-not-me sort: the suffering of the tragic hero makes spectators appreciate their own relative security and comfort. Thomas Rymer (1693), who developed the concept of ‘poetic justice’, which requires that the plot must provide a clear moral lesson, held that pleasure arises from the moral appreciation of seeing poetic justice take place.

David Hume’s essay ‘Of Tragedy’ (1757) attempts to explain ‘the unaccountable pleasure…[we] receive from sorrow, terror, anxiety and other passions, that are in themselves disagreeable and uneasy’. He claims that there may be some value in the observation of L’Abbé Du Bos that simply being dislodged from listlessness is itself pleasurable, and in Fontenelle’s idea that the sorrows of the theatre are softened by our recognizing ‘it is nothing but a fiction’. He adds that the eloquence of the poet pleases us and that all imitations please in themselves, so that these new passions predominate over and ‘convert’ the unpleasant ones. Hume’s view was not moralistic, though he asserted that conflicts between our own morality and the morality expressed in the play, and the sight of horrific violence (such as being ‘besmeared all over with mingled blood and gore’), can prevent us from being pleased by the play (see Hume, D.).

5 Germany

In the mid-eighteenth century, Johann Christoph Gottsched led a movement aiming to bring a strict and pedantic version of neoclassic French tragic theory and practice, along the lines of Boileau, to Germany. Tragedy carried literary prestige, but the forms prescribed by the rules were increasingly felt to be forced and unnatural, the elevated diction to be stilted, and the practice of *éloignement* - the distancing of characters and their actions from the audience in status, time and place - to be remote from most people’s experience. Gotthold Ephraim Lessing gradually acknowledged the need to change the forms. His *Miss Sara Sampson* has been called the first really ‘modern’ tragedy; it is written in prose, with fictional (rather than historical) characters from the lower fringe of the nobility, and has a contemporary setting. His *Hamburg Dramaturgy* (1767-8)articulated later ideas about tragedy, notably that the hero should not be admired but pitied, and that tragedy should enhance the audience’s moral awareness through this pity.

Friedrich Schiller’s views (see Schiller, J.C.F. §2) on tragedy grew out of Kantian metaphysics. The ‘play impulse’ unifies the phenomenal world (of determinism) with the noumenal world (of free choice). The tragic hero is not simply determined by physical forces, but challenged by suffering to choose freely what is morally right.

Schiller thought that neoclassic French tragedy erred by overemphasizing reason: to be truly effective tragedy must strike a balance between reason and passion. Both Schiller and Goethe wrote so-called ‘bourgeois tragedies’, spurning the requirement that tragedy should portray persons of noble birth.

A precursor to Hegel, August Schlegel saw as essential to tragedy the conflict between our aspirations to know both the (noumenal) infinite and our own (phenomenal) finite existence. Our pleasure derives from the affirmation of human striving for something beyond ourselves, even though we suffer in the effort. Goethe’s *Faust* could be considered tragedy on this account. Hegel himself transformed tragic theory by focusing on conflict and by de-emphasizing the role of suffering, emotion, pleasure and purgation. Conflicts arise between parents and children, or between the family and the state or ruler, because of one-sidedness: each party ignores the rights of the other. Tragedies can end with a peaceful reconciliation, at which the tragic hero submits to a course of action that he has previously fundamentally opposed. In these resolutions, what is justified on each side is preserved. Among his examples are the *Eumenides*, where claims of contending powers are adjusted, and *Oedipus Coloneus*, where Oedipus accepts as just the exile which is his fate. Thus, Hegel’s dialectical notion of tragedy looks to the (rational) reconciliation of conflicting ethical claims, in contrast with the ancient concept which saw people as victims of unpredictable, irrational forces.

Arthur Schopenhauer (§§4-6) borrowed from Buddhist philosophy the idea of nirvana - the state of release from the compulsions of human will and desire. In Schopenhauer’s writings, will is a blind, irrational cosmic force and its unceasing demands are responsible for all the suffering in the world. The essence of tragedy is suffering, and not action and conflict, as Hegel proposed. The greatest tragedies show the resignation and surrender of the will, including the will to live. It is not replaced by a will to die, he adds, but by disinterested pleasure in the momentary release from ‘the penal servitude of willing’.

Friedrich Nietzsche’s *The Birth of Tragedy* (1872) portrayed two forces at work in human nature. The Apollonian force is calm, orderly, structured, harmonious, rational. It exemplifies the glory of the individual and the wisdom of the ancient Greek maxim, ‘Know thyself.’ The Dionysian force is, by contrast, intoxicated, irrational, aggressive, disorderly, destructive, wilful. The tragic hero represents the Apollonian side, and the chorus - where the individual disappears - the Dionysian. The hero is annihilated, but in this destruction of the individual there exists an affirmation of (Dionysian) life in a cosmic sense. Nietzsche thus rejected Schopenhauer’s resignation to fate and instead championed the joyful identification with nature, no longer limited by concerns for self (see Nietzsche, F.).

6 The twentieth century

Henrik Ibsen gave his 1879 play *A Doll’s House* the subtitle ‘A Modern Tragedy’, though the heroine Nora embodies a (Hegelian) conflict of respect for authority and being true to one’s own feelings. But the play is modern in its middle class, domestic, contemporary setting, its use of prose and its recognition of the rights of women. That women could be tragic figures, representing humankind and the conflicts that plague us all, was an extraordinary advance.

A.C. Bradley defended a fundamentally Hegelian analysis of tragedy in ‘Hegel’s Theory of Tragedy’ (1909). He agreed with Hegel’s emphasis on conflict rather than suffering, but argued that some conflicts depend not only on the character of the agents but on fate, and result from warring between good and evil, rather than just conflicts between different goods.

Miguel de Unamuno (1913) proposed that tragedy was a way of looking at life, and many writings of the twentieth century emphasize a ‘tragic vision’ or ‘tragic sense’ rather than its essential properties as a literary genre. For example, Joseph Wood Krutch’s impassioned essay ‘The Tragic Fallacy’ (1929) held it a fallacy to attribute nobility to actions themselves, as an objective property of certain actions. Rather, tragedy represents human actions *as* noble. Tragedy satisfies the ‘universally human desire to find in the world some justice, some meaning, or, at the very least, some recognizable order’. He claims that Ibsen presented life as trivial and meaningless, and hence that his work lacked the tragic spirit that expressed the value of human life, rather than despair.

Several forces emerged in the twentieth century to question the viability and relevance of tragedy as a genre. Ibsen’s ‘democratization’ of tragedy (with its own precursors in German ‘bourgeois tragedy’) rejected rigid class structures and gender roles. Marxist and feminist theories undermine the concept of a tragic hero as an independent

agent responsible for the course of the 'tragic action'. Bertolt Brecht, himself a Marxist, encouraged a stylized form of acting which produces an intellectualized rather than emotional response. This is reminiscent of early French neoclassic tragic acting, but to an opposite end: whereas the French reaffirmed the power of an individual's rationality to guide their fate, Brecht saw the individual as alienated from the social forces that determine the course of history. The Absurdists (Ionesco, Genet, Beckett, Pinter) depicted the world as meaningless and refused to represent human actions, in Krutch's words, *as* noble. In general, the side of tragedy consisting of a positive affirmation of human life, of value, of the power and nobility of the individual, has been subdued. But forces working in the late twentieth century for the empowerment of and respect for difference among traditionally marginalized and disempowered peoples - women, people of colour, and cultural minorities of all types - create an environment in which a new set of values could provide a foundation for the affirmation of human worth which tragedy entails.

See also: Comedy §§3-4; Emotion in response to art §5; Hegel, G.W.F. §8; Lessing, G.E.; Mimēsis; Poetry

SUSAN L. FEAGIN

References and further reading

Aristotle (*c*. mid 4th century BC) *Poetics*, trans., with intro. and notes J. Hutton, New York: W.W. Norton, 1982. (Highly influential account of what could be called a 'science' of writing and understanding tragedy, describing its function and how it should be structured to fulfil that function. Established the paradigm of tragedy involving a tragic hero; pity, fear, and suffering; and its effect as catharsis.)

Bradley, A.C. (1909) 'Hegel's Theory of Tragedy', in *Oxford Lectures on Poetry*, London: Macmillan.(Lucid discussion of Hegel's views, extending them to cover cases of conflict between good and evil, not just a reconciliation of various goods.)

Hegel, G.W.F. (1835-8) *The Philosophy of Fine Art*, trans. and intro. F.P.B. Osmaston, London: G. Bell & Sons, 1920, 4 vols.(A dense and difficult work in which Hegel integrates theories of tragedy and poetry with his larger metaphysical scheme. He establishes a new paradigm of tragic action as conflict and its resolution, rather than pity, fear and suffering.)

Hume, D. (1757) 'Of Tragedy', in J.W. Lenz (ed.) *Of the Standard of Taste and Other Essays*, Indianapolis, IN: Bobbs-Merrill, 1965.(A delightful essay attempting to explain why we enjoy tragic drama.)

Krutch, J.W. (1929) 'The Tragic Fallacy', in *The Modern Temper*, New York: Harcourt Brace.(Passionate - perhaps maddening, but certainly stirring - defence of the nobility of the desire to find meaning in the world. Condemns plays that end in despair, where the pursuit of the good is depicted as not important enough to justify the suffering.)

Lessing, G.E. (1767-8) *Hamburg Dramaturgy*, trans. H. Zimmern, New York: Dover, 1962.(A series of essays that turns away from neoclassicism, proposing instead, for example, that the language of tragedy should be more natural and the psychology of its characters more realistic.)

Nietzsche, F. (1872) 'The Birth of Tragedy', in W. Kaufmann (ed. and trans.) *The Birth of Tragedy and the Case of Wagner*, New York: Random House, 1967.(A dramatic and challenging account of the Apollonian and Dionysian as psychological forces at work in tragedy.)

Plato (*c*.395-387 BC) *Ion*, trans. L. Cooper, Ithaca, NY: Cornell University Press, 1938.(Short, accessible dialogue arguing that the abilities to write, appreciate and dramatically recite poetry are not developed through reason.)

Plato (*c*.380-367 BC) *Republic*, III and X, trans. B. Jowett, *The Dialogues of Plato*, 3rd edn, London: Oxford University Press, 1892.(A complex dialogue presenting, *inter alia*, moral, epistemic and metaphysical reasons for why tragedy and mimetic poetry do not have a useful role to play in the ideal state.)

Rymer, T. (1693) *A Short View of Tragedy*, London: Richard Baldwin. (A dated but at times amusing critique of tragedies through the centuries, arguing that they should serve the purpose of edifying the public about justice and virtue.)

Schopenhauer, A. (1883) *The World as Will and Idea*, vol. 3, trans. R.B. Haldane and J. Kemp, New York: Charles Scribner's Sons.(A clear and accessible discussion of how tragedy shows us that suffering results from the drive to satisfy our will, and pleasure results from surrendering the will to live.)

Steiner, G. (1961) *The Death of Tragedy*, New York: Alfred A. Knopf.(Proposes that tragedy arises in cultures that see indomitable and implacable forces as limiting and thwarting human power and reason, leading to suffering and destruction.)

Unamuno, M. (1913) *The Tragic Sense of Life*, trans. J.E. Crawford Flitch, New York: Dover, 1954.(A passionate exploration of themes we now identify as existentialist, proposing that human tension and conflict, sorrow and anguish are 'solved', paradoxically, only by a longing for eternal life.)

Williams, R. (1966) *Modern Tragedy*, London: Chatto & Windus.(A poet and critic, he argues that our understanding of the history of tragedy and tragic theory is coloured by our own perspectives, and this helps us to resolve the conflict between what is accepted as tradition and our own ordinary notion of a tragic event.)

Transcendental arguments

Transcendental arguments seek to answer scepticism by showing that the things doubted by a sceptic are in fact preconditions for the scepticism to make sense. Hence the scepticism is either meaningless or false. A transcendental argument works by finding the preconditions of meaningful thought or judgment. For example, scepticism about other minds suggests that only the thinker themselves might have sensations. A transcendental argument which answered this scepticism would show that a precondition for thinking oneself to have sensations is that others do so as well. Expressing the scepticism involves thinking oneself to have sensations; and the argument shows that if this thought is expressible, then it is also false.

Arguments with such powerful consequences have, unsurprisingly, been much criticized. One criticism is that it is not possible to discover the necessary conditions of judgment. Another is that transcendental arguments can only show us how we have to think, whereas defeating scepticism involves showing instead how things really are.

1 Nature of argument

The name ‘transcendental argument’ originally comes from Kant, who called the most difficult part of his most difficult work a ‘transcendental deduction’ (see Kant, I. §6). In the twentieth century, the name has been applied by both proponents and critics to a loosely similar set of arguments, sometimes directly inspired by Kant, sometimes not. The central figure in re-launching such arguments was P.F. Strawson in his work on metaphysics, *Individuals* (1959), and his reconstruction of Kant, *The Bounds of Sense* (1966) (see Strawson, P.F. §8). Subsequently the work of Davidson, Putnam and Searle on the relations between language and the world has been called ‘transcendental argument’ both by themselves and by others; indeed Richard Rorty calls Davidson’s argument a ‘ transcendental argument to end all transcendental arguments’ (Bieri, Horstmann and Krüger 1979: 78).

One example of a transcendental argument is Strawson’s attack on scepticism about other minds. He maintains that the ability to attribute mental states to others is a precondition for attributing mental states to oneself. Therefore, if the scepticism is stated in its standard form of wondering whether anyone apart from the thinker has thoughts or feelings, it is also answerable. For the argument shows that a precondition for any such wondering is that we also attribute such states to others. Hence, if scepticism makes sense at all, it is mistaken; the sceptic is reduced to either error or silence.

This particular example is a kind of private language argument (see Private language argument). Wittgenstein has retrospectively been seen as having made these kinds or arguments, as more obviously and recently has Donald Davidson,. The closest analogy to them in Kant is not the *Transcendental Deduction* but, rather, the *Refutation of Idealism*, where Kant argued that knowledge of outer states is a precondition for knowledge of inner states, and hence claimed that ‘the game played by idealism has been turned against itself’ (1781/1787: B276). As in Strawson, the way that the sceptical position (here idealism) enters the game means that it can be defeated.

This briefly sketched family of arguments illustrates several things about modern transcendental arguments. They are arguments about the preconditions of thought or judgment. They start with a supposition about our thoughts, such as that we have thoughts of some particular kind. A necessary condition for having such thoughts is then derived, followed by a necessary condition for this necessary condition, and so on. Assuming that the first assumption is correct, all its necessary conditions will then have been found also to apply.

2 Problems

As an answer to scepticism, a transcendental argument is only as strong as its initial assumption. In the example in §1 above, the assumption is that we think ourselves to have mental states. However, a sceptic could try to meet this by claiming that we might be mistaken about the content of our thoughts. Similarly Putnam’s transcendental argument that we would not even be able to think that we were seeing an apple unless this kind of state was typically caused by real apples (and hence that external-world scepticism must be false) could be met by doubting whether I correctly identify my thoughts as being of the apple variety.

If, however, the initial assumption is not about the contents of thoughts but is just that there is thought (or

language), then the sceptical move can be met. For this is something which no one, and hence no sceptic, can properly think that they doubt. Hence the most robust transcendental arguments concern themselves merely with the preconditions of there being thought or language at all. Examples are (again) the Private Language Argument, or Davidson's argument that there is only one conceptual scheme. The next possible point of weakness after the initial assumption is the discovery of its necessary conditions. Obviously this must not have a purely observational basis. The argument that people could not think without brains - we think, hence we have brains, hence the external world exists - would not be an effective answer to scepticism. So the arguments have to be (at least relatively) a priori.

Some people have therefore held that transcendental arguments may only draw out the purely conceptual, or analytic, consequences of their initial assumptions. This gives them better security against scepticism but makes their chances of reaching interesting conclusions rather slim. Others have held, rather more ambitiously, that the connections do not need to be narrowly analytic: transcendental arguments show the necessary conditions for having thoughts. Therefore, if the having of one kind of thought is a necessary condition for having another kind of thought, this will be good enough, even if the contents of the thoughts themselves are not analytically connected.

3 Further criticisms

In transcendental arguments, necessary conditions are often demonstrated by discovering the unique conditions which enable some kind of thought to be had. Yet it has been claimed, most prominently by Stefan Körner (1969), that the uniqueness of a conceptual structure cannot be established. Uniqueness, he says, could only be demonstrated by the elimination of all possible rivals. Yet although we may eliminate all rivals that occur to us, there may always be a possibility that we might have overlooked or might not be able to envisage. Other people have similarly claimed that, at best, we can only describe how we currently think, or the limits of currently imaginable alternatives. However, neither of these can show how we must think; they cannot demonstrate the necessary conditions of thought.

Another criticism is that, even if it can be shown what we must think, this will not defeat scepticism. Defeating scepticism demands a demonstration that we are justified in thinking that our beliefs show how things actually are. However, showing that we must think something does not show that the thought is correct - a necessary illusion is still an illusion. The force of this criticism depends on the kinds of necessary conditions derived in transcendental arguments. If some matter of fact is shown to be a necessary condition for having thoughts, then it does not apply. However, in modern arguments, the necessary conditions derived for having one kind of thought are frequently other kinds of thought. For example in the Strawson argument discussed above in §1, it is belief in other people's pains that is the required presupposition. It can be objected that this is insufficient to answer scepticism (in this case about other minds) since the crucial point is not whether we have to believe that others have pains, but, rather, whether this belief is correct.

A version of this objection is the one which Stroud mounts against Strawson and others in his highly influential paper about transcendental arguments (1968). This is his claim that transcendental arguments need a version of the Verification Principle in order to work (see Meaning and verification). However, Stroud continues, if we are entitled to presuppose the Verification Principle, then transcendental arguments are not needed, since the Verification Principle by itself will do the anti-sceptical work. The principle shows that it makes no sense to suppose that our best-verified beliefs might not be true. Analogously, it can show that it makes no sense to suppose that beliefs that we must have are not true. However it is the verificationist assumption which is doing the deadly anti-sceptical work, not the transcendental argument which gets us from belief to necessary belief. Therefore the route through necessary belief is superfluous; transcendental arguments are either redundant or invalid.

Alternatively, the transition from what we (have to) believe to how things actually are can be made by shrinking the distance between them. A supposition of idealism would do this; so that in the end we do not distinguish between the structure of thought and an independent, real world waiting to be described. In Kant, some kind of idealism was required in order to make the transcendental arguments work, and it has been suggested by Bernard Williams (1973) and others that this also applies to modern users of transcendental arguments. However, some modern users would not take this to be a criticism (see Idealism; Realism and antirealism).

See also: Scepticism

ROSS HARRISON

References and further reading

Bieri, P., Horstmann, R.-P. and Krüger, L. (1979) *Transcendental Arguments and Science*, Dordrecht: Reidel. (Contains papers by Bennett, Rorty, Stroud and others on transcendental arguments.)

Brueckner, A. (1983) 'Transcendental Arguments I', *Nous* 17 (4): 551-75.(Pushes Stroud's verificationist objection against modern attempts at making transcendental arguments.)

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press.(Rather elliptical, at times technical, but very stimulating. Essays 11 and 13 contain examples of transcendental arguments.)

Davidson, D. (1991) 'Three varieties of knowledge', in A. Phillips Griffiths (ed.) *A.J. Ayer Memorial Essays*, Cambridge: Cambridge University Press.(Succinct summary of his transcendental arguments about other minds and external world; easy to follow.)

Grayling, A.C. (1985) *The Refutation of Scepticism*, London: Duckworth.(Puts reconstructed Strawson and reconstructed Davidson together to build a transcendental argument which he thinks survives all the standard objections.)

Harrison, R. (1982) 'Transcendental arguments and idealism', in G. Vesey (ed.) *Idealism, Past and Present*, Cambridge: Cambridge University Press.(Replies to the criticism made above in §3 that transcendental arguments involve idealism.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1929.(Contains the passages referred to in §1.)

Körner, S. (1969) *Fundamental Questions in Philosophy*, Harmondsworth: Allen Lane, The Penguin Press. (Chapter 12 contains the objection mentioned in §2 above; moderately difficult work.)

Peacocke, C. (1989) *Transcendental Arguments in the Theory of Content*, Oxford: Oxford University Press. (Exhibits his work on mental content as a transcendental argument; at times over-condensed.)

Putnam, H. (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press.(Chapter 1 contains the striking argument referred to above in §2; moderate difficulty, but not technical.)

Schaper, E. (1972) 'Arguing transcendentally', *Kant-Studien* 63 (1): 101-16.(Direct reply to Körner.)

Schaper, E. and Vossenkuhl, W. (1989) *Reading Kant*, Oxford: Blackwell.(In spite of its title, this collection contains papers on the modern use of transcendental arguments against scepticism as well as a good bibliography.)

Searle, J. (1995) *The Construction of Social Reality*, Harmondsworth: Allen Lane, The Penguin Press.(The argument mentioned above in §1 appears in chapter 8; easy work.)

Strawson, P.F. (1959) *Individuals*, London: Methuen.(Mentioned in §1 above; influential transcendental arguments in first three chapters. Very solid argument, but no technicality.)

Strawson, P.F. (1966) *The Bounds of Sense*, London: Methuen.(Mentioned above in §1; a reconstruction of Kantian arguments.)

Stroud, B. (1968) 'Transcendental Arguments', *The Journal of Philosophy* 65 (9): 241-56.(Mentioned in §3 above; frequently cited article which is not particularly difficult.)

Williams, B. (1973) *Problems of the Self*, Cambridge: Cambridge University Press. (The argument referred to in §3 above is contained in paper number 8; a fairly difficult work.)

Translators

Translators played a crucial role in the history of medieval philosophy. Since multilingualism was generally restricted to places in which a direct contact between different languages was possible, such as Byzantium, the Near East, southern Italy or Spain, the dissemination of knowledge into foreign cultures was mainly brought about by means of translation. In this conversion process various kinds of writings were involved, including the Bible, the Qur'an and liturgical and hagiographic works as well as literary and historiographic texts.

1 Early Greek-Latin translations

The tradition of Greek-Latin translations of philosophical and scientific texts goes back to Cicero, whose philosophical works contain some translated fragments (for example, of Plato's *Timaeus*), but consist for the greater part of free adaptations of some contemporary Greek models. The practice of both translation and paraphrase was continued by early Christian writers such as Marius Victorinus and Ambrose (see Patristic philosophy). As for Plato's *Timaeus*, the greater part of it was known to medieval thinkers through the Latin translation and commentary of Calcidius in the fourth/fifth century.

The famous debate over translation *ad verbum* (according to the verbal expression) and *ad sensum* (according to the meaning) also originated in Roman times. Jerome in the fourth century, following Cicero, was a representative of the latter method but defended literal translation when a highly authoritative text such as the Bible was at issue. Boethius in the sixth century adopted the same position with respect to the works of a renowned philosopher such as Aristotle: he translated 'word for word' the *Categories*, *De interpretatione*, *Prior Analytics*, *Topics* and *Sophistici Elenchi*, as well as Porphyry's *Isagoge* (see Porphyry). Boethius' translation strategy was followed in the Carolingian Renaissance by Johannes Scottus Eriugena, who made the Neoplatonism of the Greek Fathers accessible to Latin readers (see Carolingian Renaissance).

2 The science of the Arabs

The twelfth century was an age of revival in European science and philosophy. This cultural phenomenon was, to a large extent, a consequence of the appearance of Latin translations of a number of Greek works, and also of some writings of the Arabs in which Greek science had been incorporated and developed. The initiator of this movement was Constantine the African, a monk of Monte Cassino in the second half of the eleventh century. He was a contemporary of Alfanus, archbishop of Salerno, who translated from Greek (under the title *Premnon physicon*) the treatise *On the Nature of Man* by Nemesius of Emesa. Constantine wrote several Latin versions of Arabic works, or of Arabic translations from Greek, on medicine, the most influential of which was the *Pantegni*, an adaptation of Haly Abbas' *Kitab al-malaki*.

An intriguing personality is Adelard of Bath, active in the first half of the twelfth century; his translations include the *Elements* of Euclid, the *Astronomical Tables* of al-Khwarizmi and the *Shorter Introduction to Astronomy* by Abu Ma'shar, and yet he does not seem to have had a profound knowledge of Arabic. Probably these versions were made with the collaboration of Adelard's teacher, the Spanish Jew Petrus Alfonsi. It was actually in Spain that most Arabic-Latin translations originated in the course of the twelfth century. Important translators were John of Seville and Hermann of Carinthia, both of whom produced Latin versions of Abu Ma'shar's *Greater Introduction* (in 1135 and 1140), and Hugh of Santalla, who shared with his contemporaries a strong interest in astronomy and astrology.

3 The *Aristoteles Latinus*

In the second half of the twelfth century, the flourishing centre of Arabic-Latin translations was Toledo. It was here that Dominicus Gundissalinus, with the help of the Jewish scholar Avendauth, translated into Latin some parts of Avicenna's *Kitab al-shifa'* (see Ibn Sina). Gerard of Cremona, probably the most prolific of all medieval translators, also worked in Toledo. Apart from his Latin translations of scientific works in almost every field, Gerard made a substantial contribution to the *Aristoteles Latinus*, the medieval Latin version of Aristotle's works: *Posterior Analytics*, *Physics*, *On the Heavens*, *On Generation and Corruption*, *Meteora* I-III and the pseudo-Aristotelian *Liber de causis* (see Aristotelianism, medieval; Liber de causis). Meanwhile, probably before 1150, the first two of these works as well as *On the Heavens*, some of the *Parva naturalia* and *Metaphysics*, had

been translated directly from Greek by James of Venice. Also known as 'Iacobus Veneticus Grecus', this translator probably lived in Byzantium for at least some time, like his contemporary Burgundio of Pisa, who produced a new version of Nemesius' *On the Nature of Man* and of several Galenic treatises (see Nemesius; Galen). In addition, Burgundio was probably responsible for the oldest Greek-Latin translations of Aristotle's *On Generation and Corruption* and *Nicomachean Ethics*, while the fourth book of the *Meteora* had been translated from Greek by Henricus Aristippus, archdeacon of Catania, before 1162. To Henricus we owe also a Latin version of Plato's dialogues *Phaedo* and *Meno*.

Many twelfth-century Greek-Latin translations that have come down to us are still anonymous. Sometimes it is possible, on the basis of certain similarities of style and terminology, to ascribe two or more of these versions to one and the same unknown scholar. Thus both the so-called 'Fragmentum Vaticanum' of the *Physics* and the 'Metaphysica Media' are probably the work of one person. Likewise, the anonymous Greek-Latin translation of Euclid's *Elements* seems to go back to the translator of the oldest Latin version of Ptolemy's *Almagest*, translated from Greek in Sicily about 1160 and so preceding Gerard of Cremona's Arabic-Latin rendering by more than ten years.

Around the turn of the century, Alfred of Sareshel, who may have learned Arabic in Spain, translated the pseudo-Aristotelian *De plantis* and an extract from Avicenna's *Meteora*. One of the most influential medieval translators was Michael Scot, active in the early thirteenth century, but the real extent of his work is not well known; attested are his versions of al-Bitruji's *On the Sphere* (made at Toledo with the help of 'Abuteus levita'), Aristotle's *De animalibus* in nineteen books (probably executed at the same place), and Avicenna's *De animalibus* (which must have been done in Italy, since it is dedicated to Emperor Frederick II). We cannot be as certain of his part in the translation of the huge Averroistic corpus. He certainly or probably translated into Latin Averroes' 'great' (that is, long) commentaries on the *Physics*, *On the Heavens*, *On the Soul* and *Metaphysics*, and these versions were the starting point for intensive study of those Aristotelian treatises (see Ibn Rushd). Averroes' 'middle' commentaries on the logical works were translated about the same time by William of Luna, while those on the *Nicomachean Ethics*, *Rhetoric* and *Poetics* were translated in this period by Hermann the German.

In the field of Greek-Latin translations, it is important to mention Robert Grosseteste for his rendering of the *Nicomachean Ethics* with commentaries by various authors, and of a fragment of the *De caelo* with Simplicius' commentary. In the second half of the thirteenth century, a number of pseudo-Aristotelian works were translated by Bartholomew of Messina at the court of Manfred, King of Sicily, while the corpus of Greek-Latin translations of Aristotle's genuine works was revised and completed by William of Moerbeke. This prolific translator, who worked in Greece and at the papal court, was responsible for the first complete Greek-Latin versions of the *On the Heavens*, *Meteora*, *De animalibus*, *Metaphysics*, *Politics* and *Poetics*. Moreover, he produced an impressive series of Latin versions of Greek commentaries on Aristotle, including *In Meteora* and *In de sensu* by Alexander of Aphrodisias, *In de interpretatione* by Ammonius, *In de anima* by Philoponus, *In categorias* and *In de caelo* by Simplicius, and *In de anima* by Themistius. In addition, William translated Proclus' *Elements of Theology* (see Proclus), *Tria opuscula*, *In Parmenidem* and *In Timaeum*, Alexander's *De fato*, several treatises of Archimedes and Eutocius, Heron's *Catoptrica*, Ptolemy's *De analemmate* and *Quadripartitum* and Galen's *De virtute alimentorum*. The Greek-Latin Galenic corpus was added to by Peter of Abano around the turn of the century and, particularly, by Nicholas of Reggio in the early fourteenth century.

See also: Aristotelianism, medieval; Aristotle Commentators; Averroism; Boethius, A.M.S.; Carolingian renaissance; Chartres, School of; Encyclopedists; Gerard of Cremona; Grosseteste, R.; Islamic philosophy: transmission into Western Europe; Patristic philosophy; Platonism, early and middle; Platonism, medieval

JOZEF BRAMS

References and further reading

d'Alverny, M.-T. (1982) 'Translations and Translators', in R.L. Benson, G. Constable and C.D. Lanham (eds) *Renaissance and Renewal in the Twelfth Century*, Cambridge, MA: Harvard University Press, 421-62.(Updates Haskins' work.)

Contamine, G. (ed.) (1989) *Traduction et traducteurs au moyen âge (Translation and Translators in the Middle Ages)*, Paris: Éditions du CNRS.(Various kinds of translations are considered from divergent points of view, such as differences between languages, working conditions of translators and centres of translation.)

Dod, B.G. (1982) ‘Aristoteles Latinus’, in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 45-79.(Includes useful table of medieval Latin translations of Aristotle’s works and of Greek and Arabic commentaries.)

Hamesse, J. and Fattori, M. (eds) (1990) *Rencontres de cultures dans la philosophie médiévale. Traductions et traducteurs de l’antiquité tardive au XIVe siècle (The Meeting of Cultures in Medieval Philosophy: Translations and Translators from Late Antiquity through the Fourteenth Century)*, Louvain: Université Catholique de Louvain, and Cassino: Università degli Studi di Cassino.(Concentrates on the methodological problems of medieval translations of philosophical and scientific texts.)

Haskins, C.H. (1924) *Studies in the History of Mediaeval Science*, Cambridge, MA: Harvard University Press. (Despite its age, still valuable as a source of information about translators of the twelfth and early thirteenth centuries.)

Lindberg, D.C. (1978) ‘The Transmission of Greek and Arabic Learning to the West’, in *Science in the Middle Ages*, Chicago, IL: University of Chicago Press, 52-90.(A handy survey of the data concerning the translators.)

Minio-Paluello, L. (1972) *Opuscula - The Latin Aristotle*, Amsterdam: Hakkert.(Contains almost all the articles published before 1969 by the famous specialist in medieval Latin translations of Aristotle’s works.)

Trinity

The doctrine of the Holy Trinity is a central and essential element of Christian theology. The part of the doctrine that is of special concern in the present entry may be stated in these words: the Father, the Son and the Holy Spirit are each God; they are distinct from one another; and yet (in the words of the Athanasian Creed), 'they are not three Gods, but there is one God'. This is not to be explained by saying that 'the Father', 'the Son' and 'the Holy Spirit' are three names that are applied to the one God in various circumstances; nor is it to be explained by saying that the Father, the Son, and the Holy Spirit are parts or aspects of God (like the leaves of a shamrock or the faces of a cube). In the words of St Augustine:

> *Thus there are the Father, the Son, and the Holy Spirit, and each is God and at the same time all are one God; and each of them is a full substance, and at the same time all are one substance. The Father is neither the Son nor the Holy Spirit; the Son is neither the Father nor the Holy Spirit; the Holy Spirit is neither the Father nor the Son. But the Father is the Father uniquely; the Son is the Son uniquely; and the Holy Spirit is the Holy Spirit uniquely.*
>
> *(De doctrina christiana I, 5, 5)*

The doctrine of the Trinity seems on the face of it to be logically incoherent. It seems to imply that identity is not transitive - for the Father is identical with God, the Son is identical with God, and the Father is not identical with the Son. There have been two recent attempts by philosophers to defend the logical coherency of the doctrine. Richard Swinburne has suggested that the Father, the Son and the Holy Spirit be thought of as numerically distinct Gods, and he has argued that, properly understood, this suggestion is consistent with historical orthodoxy. Peter Geach and various others have suggested that a coherent statement of the doctrine is possible on the assumption that identity is 'always relative to a sortal term'. Swinburne's formulation of the doctrine of the Trinity is certainly free from logical incoherency, but it is debatable whether it is consistent with historical orthodoxy. As to 'relative identity' formulations of the doctrine, not all philosophers would agree that the idea that identity is always relative to a sortal term is even intelligible.

1 The logical problem of the Trinity

The words 'the Trinity' are the English equivalent of the Latin word *Trinitas*, which was coined by the early Christian writer Tertullian. The word, which, etymologically, means something like 'the tripleness', is used to refer collectively to the Father, the Son and the Holy Spirit. (Tertullian also originated the use of the word 'person' (*persona*) as a common noun that applies to the Father, the Son and the Holy Spirit. Outside theology, the Latin word means a mask of the sort worn by characters in a classical drama, and, by extension, a *dramatis persona*, a character in a drama. What Tertullian's application of this word to the Father, the Son and the Holy Spirit was intended to suggest is disputed.) Theologians writing in Latin have generally said that, although God is a single *substantia*, there are in God three *personae*. Theologians writing in Greek have generally said that, although God is a single *ousia*, there are in God three *hypostases*. These two pairs of terms have caused some confusion, owing to the fact that *substantia* and *hypostasis* have the same literal or etymological meaning: 'that which stands under'.

The purpose of this entry is neither theological nor historical. Its purpose is rather to discuss the philosophical difficulties presented by the 'developed' doctrine (as it is to be found in the Athanasian Creed, of around AD 500). These difficulties are mainly logical. They are well stated in an anonymous seventeenth-century work that has been ascribed to the Socinian John Biddle:

> You may add yet more absurdly, that there are three persons who are *severally and each of them true God*, and yet there is but one God: this is *an Error* in counting or numbering; which, when stood in, is of all others the most brute and inexcusable, and not to discern it is not to be a Man.
>
> (quoted in Hodgson 1940)

The author of this passage is, essentially, charging that the doctrine of the Trinity implies a violation of the principle of the transitivity of identity, for it implies that the Father is identical with God, God is identical with the Son, and the Father is not identical with the Son. (For a full development of this charge, see Cartwright 1987.) The

central problem that faces the doctrine of the Trinity is this: how can the doctrine be stated in a way that is orthodox, clear and does not violate the principle of the transitivity of identity?

The doctrine of the Trinity is one of the Christian mysteries, which means that it cannot be seen to be true, or even to be possible, by the use of unaided human reason. This does not mean, however, that human beings, employing only their unaided reason, cannot usefully discuss the question whether the doctrine is formally self-contradictory. (If it could be demonstrated that the doctrine of the Trinity was formally self-contradictory, that would, of course, show that it was impossible; but the converse entailment does not hold.) The task undertaken in this entry does not, therefore, rest on a failure to appreciate the fact that the doctrine is held by those who accept it to be a mystery.

This entry will consider two recent attempts to avoid the conflict with Leibniz's Law that the doctrine of the Trinity seems to face (see Identity of indiscernibles §1). One proceeds by affirming that the Father, the Son and the Holy Spirit are numerically distinct from one another, and attempting to show that this thesis is consistent with historical orthodoxy. The other proceeds by denying the ultimate reality of numerical identity - and thus by denying that Leibniz's Law has anything to apply to. The first risks falling into tritheism, the heresy that there are three Gods. The second risks incoherence if not outright unintelligibility.

2 Swinburne's theory

Richard Swinburne (1988) has argued for a Trinitarian theology according to which the Father, the Son and the Holy Spirit are numerically distinct from one another and each of them is a God - each is a necessarily existent, omnipotent, omniscient, perfectly good being who is the creator of whatever world there may be, and who has each of these attributes essentially. Swinburne's theology, moreover, represents the Father as the creator of the Son. He does not, however, freely choose to create the Son, as he freely chooses to create a physical world. He is, rather, constrained by his own nature - by his perfect goodness - to create the Son (that is, he is constrained to create that very being, as opposed to being constrained to create some being or other who has certain properties that in actuality belong to the Son). 'There being a God and there being no physical world' and 'There being a God and there being a physical world that is "very good"' are morally or ethically indifferent states of affairs, and a God's perfect goodness does not, therefore, constrain him to prefer either to the other: which of these states of affairs obtains is a matter of the exercise of divine free will. But the two states of affairs 'There being only one God' and 'There being more than one God' are not morally or ethically indifferent; the second is better than the first, and the Father is, therefore, constrained by his own perfect goodness to prefer the latter. He therefore creates - eternally, of course: not at some point in time - the Son. Although Swinburne does not explicitly say this, it would appear that the individual essence of the Son must be supposed to include the property 'being created by the Father if any divine being is created by the Father'; if this were not the case, there would be no ontological ground for the fact that the Father creates the Son and not some other divine being. The Son is therefore a necessary being: he exists in all possible worlds, for the Father exists in all possible worlds, and, in every world in which he exists, he is constrained by his essential nature to create the Son. The necessity of the Father and the necessity of the Son may, in consequence, be contrasted by using a pair of phrases that Aquinas used in respect of a different kind of necessity (imperishability): the Father has his necessity of himself, but the Son receives his necessity from another.

The state of affairs 'There being more than one God' is better than the state of affairs 'There being only one God' because it is better that there should be a plurality of Gods who form a community of love than that there should be a solitary God. Swinburne argues, moreover, that it is better for a divine community of love to contain more than two Gods than to contain only two, for it is good for two beings to cooperate to benefit a third, and such cooperation could not exist within the divine nature if there were only two Gods. Hence, the Father and the Son are constrained by their moral perfection to cooperate to create a third God, the God called the Holy Spirit. There is, however, no good that requires the existence of more than three Gods, and the 'process' of the successive creation of Gods stops at three. (The ontological priority of the Father, Swinburne argues, gives him an authority over the Son and the Spirit, with the consequence that - of necessity - they conform their wills to his in matters about which a solitary God would have a free choice. The wills of the three Gods, therefore, can never be in conflict.)

Can Swinburne plausibly contend that his account of the Trinity is orthodox? There would seem to be two points on which Swinburne might be charged with unorthodoxy. There is, first, the fact that both the Creeds of the Church and every Trinitarian theologian whose writings have not been condemned have insisted that (as the Nicene Creed puts it) the Son is 'begotten, not made' (*genitus, non factus*). And, historical orthodoxy insists,

although the word ‘begotten’ is not used of the Holy Spirit, he too is ‘not made’. Second, one might well ask Swinburne why he should not be called a tritheist: after all, he says that there are three Gods, and tritheism is the thesis that there are three Gods. As to the first point, Swinburne contends that in the vocabulary of traditional theology, ‘create’ (*creare*) and ‘make’ (*facere*) have been used to express relations that God bears to finite, contingent creatures, and that traditional theologians would have objected to the words ‘*Pater filium creavit*’ only because they would have understood those words to imply that the Son was a finite, contingent being. If, however, the word ‘create’ is used in the very abstract sense of ‘eternally bring about the existence of’ - there being no implication that the being whose existence is brought about be contingent or finite - nothing contrary to historical orthodoxy is implied by ‘The Father created the Son’. On the second point, the charge of tritheism, Swinburne has chosen his words very carefully:

> A substance is not unnaturally understood as an individual thing which does not have parts capable of independent existence. Now the three persons are such that of logical necessity none can exist without the other…. They are therefore not unnaturally said to form one ‘first substance,’ and we may follow a natural tradition in calling that substance ‘God’.
>
> (1988: 236)

The sense of this passage seems to be this: the Father, the Son and the Holy Spirit are parts (albeit parts that are not ‘capable of independent existence’) of a composite being, and it is therefore natural to apply the name ‘God’ (derived from the general term ‘a God’, whose extension is the three divine parts of the composite being) to this composite being. If this is a correct interpretation of this passage, it seems unlikely that St Augustine or the framers of the Athanasian Creed would agree that Swinburne’s theory adequately captured the sense in which it is true of the Father, the Son and the Holy Spirit that ‘all are one God; and each of them is a full substance, and at the same time all are one substance’ (Augustine, *De doctrina christiana* I, 5, 5).

Whether or not Swinburne’s theory of the Trinity can plausibly be identified with the historical doctrine of the Trinity, it is clear that it faces none of the logical difficulties that the historical doctrine seems to face, for there are, according to Swinburne, three metaphysically simple beings to which the general term ‘a God’ applies, and one composite being to which the name ‘God’ applies. None of these four beings (of course) is numerically identical with any of the others, and each has - as their non-identity allows - properties that the others lack. Swinburne’s purpose was not simply to solve the logical problems that the historical doctrine seems to face, but to provide and argue for the truth of a comprehensive account of the ‘internal structure’ of the Trinity.

The other recent attempt to solve the logical problems raised by the doctrine of the Trinity is that and no more; the philosophers who have contributed to this attempt have been concerned only to show that the doctrine can be stated without internal logical contradiction, and they have said very little of an ontological nature about the Trinity.

3 Relative identity

The originator of this approach to the logical problems raised by the doctrine of the Trinity is Peter Geach (1977; Geach and Anscombe 1963), who has developed a theory according to which ‘identity is always relative to a sortal term’, which he has applied to the problems of counting and predication that confront the doctrine of the Trinity. Geach’s work has been continued by Martinich (1978) and van Inwagen (1988). The exposition that follows is a composite of things said by these three authors.

The ‘theory of the relativity of identity’ proceeds from the axiom that there is no such relation as numerical identity *simpliciter*: there is rather an indefinite number of relations expressed by phrases of the form ‘is the same *N* as’, where *N* represents the place of a count-noun. There are, for example, such relations as ‘is the same horse as’ and ‘is the same apple as’, but there is, strictly speaking, no such relation as ‘is the same as *simpliciter*’ or ‘is numerically identical with’. Identity *simpliciter* (expressed below by ‘Čequals;’) is defined by two characteristics: it is *universally reflexive* (everything bears identity *simpliciter* to itself) and it *forces absolute indiscernibility* (this characteristic is embodied in Leibniz’s Law or the principle of the indiscernibility of identicals: if $x = y$, then anything whatever that is true of *x* is also true of *y*). Relative-identity relations, however, are not in general universally reflexive. (Socrates is not the same horse as Socrates because Socrates is not the same horse as anything; that is to say, Socrates is not a horse.) Relative-identity relations, moreover, cannot be assumed to force

absolute indiscernibility - although any given such relation may have this feature. If it were assumed that every relative-identity relation forced absolute indiscernibility, then the logic of relative identities would simply be a fragment of the standard logic of identity *simpliciter*, and anything that could be said by using relative-identity predicates could be said equally well without them. (If every relative-identity relation forced absolute indiscernibility, then '*x* is the same *N* as *y*' could always be replaced by '*x* is an *N* and $x = y$'.)

The logic of relative identities is easily described. Its language is that of first-order predicate logic (without 'Čequals;' and the description operator, and without singular terms), its two-place predicates being partitioned into two classes (somehow visibly differentiated), the 'ordinary' two-place predicates, and the 'relative-identity' predicates. Its rules of inference are those of ordinary predicate logic, plus two rules that state, in effect, that relative-identity predicates express symmetrical and transitive relations. Relative-identity logic must do without anything corresponding to Leibniz's Law, for the reason outlined above. It must also do without singular terms. This is because a singular term is supposed to denote exactly one object (if it does not fail of denotation), and the concept of a singular term therefore involves the notion of identity *simpliciter*. (If *a* denotes *x* and also denotes *y*, it follows that $x = y$.) If, however, relative-identity logic is to have any power to represent ordinary, informal reasoning, its users must be able to employ some substitute for singular terms. This can be done through the use of an adaptation of Russell's Theory of Descriptions. For example, 'The present pope is bald' could be read as 'There is an *x* such that [*x* is at present a pope, and, for any *y* (if *y* is at present a pope, then *y* is the same man as *x*), and *x* is bald].' There is, of course, nothing special about the word 'man' that dictated its use in this sentence; we might as well have used 'person' or 'animal' or any of indefinitely many other count-nouns that would apply to anyone who was a pope. The sentence obtained by substituting 'person' in the above sentence is not equivalent in relative-identity logic to that sentence; to deduce either from the other, one would need a premise not endorsed by relative-identity logic. For example: 'For any *x* and for any *y*, if *x* is a man (that is, if *x* is the same man as something) and if *y* is a man, then *x* is the same person as *y* if and only if *x* is the same man as *y*.' No doubt most people would say that this proposition was *true*, but it is of the essence of the theory of the relativity of identities not to regard such propositions as truths of logic.

The customary term for 'what there is one of' in the Trinity is 'substance'. (But Geach and Martinich use 'God' for 'what there is one of' in the Trinity, and van Inwagen uses 'being'.) The customary terms for 'what there are three of' in the Trinity are 'person' and 'hypostasis'. (The relation between the meaning of 'person' in Trinitarian theology and 'person' in ordinary speech is a matter of dispute.)

All of the propositions of Trinitarian theology that raise logical problems can be represented using two relative-identity predicates ('is the same substance as' and 'is the same person as'), a predicate that expresses the divine nature ('is a God' or 'is divine'), and some predicates that express the relations that individuate the Father, the Son and the Holy Spirit. (The three persons or hypostases have traditionally been held to be individuated by the relations they bear to one another: the Father *begets* the Son; the Holy Spirit *proceeds from* the Father *and* - or *through* - the Son.) For example, the proposition that there are three divine persons can be expressed as 'There exist *x*, *y* and *z*, all of which are divine and are such that none of them is the same person as the others, and such that anything divine is the same person as one of them.' The proposition that there is one God (one divine substance) can be expressed as 'Something is divine and anything divine is the same substance as it.' These two sentences are consistent in relative-identity logic. The proposition that God is omnipotent can be expressed as 'Something is divine and anything divine is the same substance as it and it is omnipotent.' 'Reference' to the Father, the Son and the Holy Spirit can be accomplished by a device similar to the one that was used to 'refer' to God in the preceding sentence; in applying this device, use must be made of the predicates that express the relations that individuate the Father, the Son and the Spirit. Van Inwagen has shown (by constructing a model in which the interpretations of these sentences are true and in which 'is the same person as' and 'is the same substance as' express symmetrical and transitive relations) that the formal analogues of the whole set of logically problematic sentences endorsed by the doctrine of the Trinity are consistent in relative-identity logic. One striking consequence of this result is that the formal analogues of the sentences 'The Father is the same person as God', 'God is the same person as the Son' and 'The Father is not the same person as the Son' are consistent - and this despite the fact that 'is the same person as' expresses a transitive relation. (Needless to say, the formal sentences do not have the logical forms suggested by the English sentences they are held to translate.)

The main problem facing this account of the 'logic' of the Trinity would seem to be whether it is intelligible. Is it,

in the final analysis, intelligible to suppose, for some *x* and for some *y* - where *x* and *y* are both substances and both persons - that *x* is the same substance as *y*, but not the same person as *y*? Alleged non-theological cases in which *x* is the same *N* as *y*, but not the same *M* (the statue is the same lump of clay as the vase, but not the same artefact; Dr Jekyll and Mr Hyde were the same man but not the same person; James I of England and James VI of Scotland were the same man but not the same monarch) are all susceptible of lucid and plausible philosophical analyses that do not presuppose that 'identity is always relative to a sortal term'.

See also: Identity; Incarnation and Christology; Substance

PETER VAN INWAGEN

References and further reading

Augustine (396-426) *De doctrina christiana* (On Christian Doctrine), trans. E. Hill, *Teaching Christianity*, Hyde Park, NY: New City Press, 1996.(A good modern translation.)

Brown, D. (1985) *The Divine Trinity*, London: Duckworth.(A widely discussed book on the topic.)

Cartwright, R. (1987) 'On the Logical Problem of the Trinity', in *Philosophical Essays*, Cambridge, MA and London: MIT Press.(A sustained, rigorous attempt to show that the logical problems faced by the doctrine of the Trinity are insoluble by any means that have so far been proposed.)

Geach, P. (1977) *The Virtues*, Cambridge: Cambridge University Press.(This and the following item are Geach's most important statements of his attempt at a solution to the logical problems presented by Trinitarian theology; see especially pages 72-81.)

Geach, P. and Anscombe, G.E.M. (1963) *Three Philosophers*, Oxford: Blackwell.(See previous item; see especially pages 118-20.)

Hill, E. (1985) *The Mystery of the Trinity*, London: Geoffrey Chapman.(A very useful exposition of Augustine and Aquinas on the Trinity.)

Hodgson, L. (1940) *The Doctrine of the Trinity*, New York: Scribner.(A classic work.)

McGrath, A.E. (1994) *Christian Theology: An Introduction*, Oxford: Blackwell. (Recommended for readers with no background in theology or church history. Clear and reliable.)

Martinich, A.P. (1978) 'Identity and Trinity', *Journal of Religion* 58: 169-81.(An attempt at a Geach-style solution to the logical problems of the Trinity. More systematic than Geach.)

Swinburne, R. (1988) 'Could There Be More Than One God?', *Faith and Philosophy* 5: 225-41.(Swinburne's account of the Trinity.)

Tertullian, Q.S.F. (*c*.213) *Adversus Praxean (Against Praxeas)*, trans. A. Souter, London: SPCK, 1920.(Contains Tertullian's treatment of the Trinity; this is the source of the technical terminology used in Latin Christian discussions.)

Van Inwagen, P. (1988) 'And Yet They Are Not Three Gods but One God', in T.V. Morris (ed.) *Philosophy and the Christian Faith*, South Bend, IN: University of Notre Dame Press.(An attempt at a Geach-style solution to the logical problems of the Trinity. A broader range of problems than those considered by Geach and Martinich is addressed. This essay may be consulted for further references.)

Troeltsch, Ernst Peter Wilhelm (1865-1923)

Ernst Troeltsch was a theologian, sociological historian, and philosopher of religion and history. He aimed to reconcile theology with modern scientific culture by grounding his philosophy of religion on historical analysis, and is regarded as the systematician of the 'history of religion school'. He is famous for his critical appraisal of the Protestant Reformation, which, he argued, had retarded the development of modern culture.

A native of Haunstetten, Bavaria, Troeltsch studied theology first at Erlangen, where the philosopher Gustav Class strove to reconcile idealism with empiricism and historicism, then at Berlin, where he attended Julius Kaftan's and Heinrich von Treitschke's lectures. Finally he went to Göttingen, where he became one of Albrecht Ritschl's (1822-89) last pupils. There he met progressive theologians who later formed the 'history of religion school', submitted his dissertation, and started work as an assistant lecturer. In 1894 he assumed the chair of systematic theology at Heidelberg, which he left in 1915 to take up a personal chair of philosophy at Berlin. The collapse of the German Empire lured him into politics and, as a committed democrat, he took an active yet unrewarding part in establishing the Weimar Republic. Politics clouded the close of his otherwise successful career.

As a theologian, Troeltsch searched for an auspicious compromise with modern secular culture. He was an apologist in disguise, who sought to moderate the shattering effect of modern science on orthodox belief. In a positivistic and historically conscious age, Christian theologians had to establish their discipline on a scientifically unassailable foundation. Neither supernatural revelation nor rational metaphysics but only history could fulfil such a function (Troeltsch 1902). A philosophy of religion drawn from analysis of the history and development of religious consciousness would, in Troeltsch's opinion, strengthen the standing of the science of religion among other sciences. In addition, it would assist the Christian strategy against materialism, naturalism, scepticism, aestheticism and pantheism. Besides, Troeltsch shared Wilhelm Herrmann's assumption that ethics provided the framework for the study of religion. An heir to both Schleiermacher's religious consciousness and Hegel's concept of development, Troeltsch unfolded an idealism of historical development, which aimed at bridging the gulf between 'scientific' (historical) and 'dogmatic' (ecclesiastical) theology. He therefore ascribed to theology a threefold task: the justification of the supremacy of Christianity on a historical and comparative religion basis; the specification of the essence of Christianity; the exposition of the Christian understanding of God, the world, the soul and redemption.

As a historian, Troeltsch interpreted Protestant and Roman Catholic Christianity, as well as contemporary intellectual culture, from a philosophical-sociological perspective (Troeltsch 1906a, 1915). He emphasized the cultural value of religion and derived most standards of truth and value from an analysis of history. In Dilthey's wake, he also reflected on the relation of historiography and relativism (Troeltsch 1922). As a philosopher, Troeltsch upheld a post-Kantian, eclectic idealism. Initially close to Hermann Lotze and Wilhelm Dilthey, he later shifted to the School of Baden (Windelband, Rickert). However, he opposed the neo-Kantian distinction between knowledge and faith then current, and refused to ground religious belief in value-judgments alone. In *Das Historische in Kants Religionsphilosophie* (The Historical in Kant's Philosophy of Religion, 1904), Troeltsch strove to restore a more authentic Kant, in whom he found evidences of the beginning of a modern historical consciousness.

Die Bedeutung des Protestantismus für die Entstehung der modernen Welt (*Protestantism and Progress*, 1906b) argues that early Protestantism marked a retreat to medieval theocratic civilization. By holding to the principle of authority and enforcing inward asceticism, the Protestant Reformation merely modified medieval Catholicism and thereby delayed the development of modern culture. It was the Enlightenment that fulfilled the Renaissance and produced the modern world. With regard to sociology, Troeltsch closely followed Max Weber, though he disagreed with him on some issues. His masterpiece, *Die Soziallehren der christlichen Kirchen* (*The Social Teaching of the Christian Churches*, 1912), considers whether the development of religious beliefs and movements was conditioned by external factors and whether, in turn, religion has influenced society and culture. From a detailed study of Christian social history, Troeltsch derived three types of sociological self-formation of the Christian idea: the Church, the sect and the mystic. It was Troeltsch's conclusion that the permanent quest for, and opposition to, a cultural compromise wove the fabric of the Christian ethos.

JEAN-LOUP SEBAN

List of works

Troeltsch, E. (1912-25) *Gesammelte Schriften (Collected Writings)*, Tübingen: Mohr. (Mainly published during Troeltsch's lifetime, this collection is not exhaustive.)

Troeltsch, E. (1895-6) 'Die Selbständigkeit der Religion' (The Autonomy of Religion), in *Zeitschrift für Theologie und Kirche* 5-6.(An over-ambitious set of addresses on the origins and development of religion in the individual consciousness and in history.)

Troeltsch, E. (1900) 'Über historische und dogmatische Methode der Theologie', in G. Sauter (ed.) *Theologie als Wissenschaft (Theology as Science)*, Munich: Chr. Kaiser Verlag, 1971; trans. W.F. Bense, 'Historical and Dogmatic Method in Theology', in J.L. Adams and W.F. Bense (eds) *Religion in History*, Edinburgh: T. & T. Clark. (Responding to criticisms of his work by F. Niebegall, Troeltsch presents his version of historical method.)

Troeltsch, E. (1902) *Die Absolutheit des Christentums und die Religionsgeschichte*, Munich/Hamburg: Siebensten Taschenbuch Verlag, 1970 (repr. of 3rd, 1929, edn); trans. D. Reid, *The Absoluteness of Christianity and the History of Religions*, London: SCM Press, 1972.(A famous and much reprinted work, in which Troeltsch argues that judgments of the value of Christianity can only emerge from study of its historical development.)

Troeltsch, E. (1904) *Das Historische in Kants Religionsphilosophie (The Historical in Kant's Philosophy of Religion)*, Berlin: Reuther & Reichard.(Both an appropriation of Kantian themes in the philosophy of religion and an historical discussion of Enlightenment thought.)

Troeltsch, E. (1905) *Psychologie und Erkenntnistheorie in der Religionswissenschaft (Psychology and Epistemology in the Scientific Study of Religion)*, Tübingen: Mohr. (Lecture given at the World Congress of Arts and Sciences in the United States, which spread Troeltsch's fame and included his first discussion of the religious a priori.)

Troeltsch, E. (1906a) 'Protestantisches Christentum und Kirche in der Neuzeit' (Protestant Christianity and Church in the Modern Era), in P. Hinneberg (ed.) *Die christliche Religion mit Einschluss der israelitisch-jüdischen Religion (The Christian Religion with the Inclusion of the Israelite-Jewish Religion)*, Berlin/Leipzig: B.G. Teubner.(A contribution towards a history of the European mind, focusing on the Reformation and later developments springing from it, all in cultural context.)

Troeltsch, E. (1906b) *Die Bedeutung des Protestantismus für die Entstehung der modernen Welt (The Significance of Protestantism for the Formation of the Modern World)*, Munich/Berlin: Oldenbourg; trans. W. Montgomery, *Protestantism and Progress: A Historical Study of the Relation of Protestantism to the Modern World*, Boston, MA: Beacon Press, 1966.(A controversial work, claiming that early Protestantism was a retrogade step in the development of modernity.)

Troeltsch, E. (1909) 'Zur Frage des religiösen a priori', *Gesammelte Schriften* II, Tübingen: Mohr; trans. W.F. Bense, 'On the Question of the Religious A Priori', in J.L. Adams and W.F. Bense (eds) *Religion in History*, Edinburgh: T. & T. Clark.(Clarifies Troeltsch's views on the religious a priori, in response to criticisms by P. Speiss.)

Troeltsch, E. (1912) *Die Soziallehren der christlichen Kirchen*, *Gesammelte Schriften* I, Tübingen: Mohr; trans. O. Wyon, *The Social Teaching of the Christian Churches*, London: Allen & Unwin, 1949.(Troeltsch's favourite of his own works, in which he develops his famous typology of social forms of Christian institution.)

Troeltsch, E. (1913a) 'Prinzip, religiöses', in F.M. Schiele and L. Zscharnack (eds) *Die Religion in Geschichte und Gegenwart (Religion in History and the Present)*, Tübingen: Mohr, 5 vols, 1909-14; trans. R.A. Wilson, 'Religious Principle', in J. Pelikan (ed.) *Twentieth Century Theology in the Making*, London: Fontana, vol. 2, 1970.(A good, short introduction to Troeltsch's combination of historical method and philosophy of religion; there are several more articles by Troeltsch in the Schiele and Zscharnack volumes.)

Troeltsch, E. (1913b) *Zur religiösen Lage, Religionsphilosophie und Ethik (On the Religious Situation, Philosophy of Religion, and Ethics)*, *Gesammelte Schriften* II, Tübingen: Mohr; excerpts trans. in J.L. Adams and W.F. Bense (eds) *Religion in History*, Edinburgh: T. & T. Clark.(Collected essays in the philosophy of religion.)

Troeltsch, E. (1915) *Augustin, die christliche Antike und das Mittelalter im Anschluß an die Schrift 'De Civitate Dei' (Augustine, Christian Antiquity and the Middle Ages, with Reference to 'The City of God')*, Aalen: Scientia Verlag, new edn, 1963.(Another philosophical and sociological investigation of the development of Christianity, this time of an earlier period.)

Troeltsch, E. (1922) *Der Historismus und seine Probleme: Das logische Problem der Geschichtsphilosophie*

(Historicism and its Problems: the Logical Problem of the Philosophy of History), *Gesammelte Schriften* III, Tübingen: Mohr.(This was to be the first of two volumes, but the second was never finished; it presents Troeltsch's mature philosophy of history at length.)

Troeltsch, E. (1925) *Aufsätze zur Geistesgeschichte und Religionssoziologie (Essays on Intellectual History and the Sociology of Religion)*, ed. H. Baron, *Gesammelte Schriften* IV, Tübingen: Mohr.(The closest we have to the unfinished second volume of *Der Historismus und seine Probleme*, containing various short essays.)

Troeltsch, E. (1977) *Ernst Troeltsch: Writings on Theology and Religion*, ed. R. Morgan and M. Pye, London: Duckworth.(Short essays on theology in the nineteenth century, the essence of Christianity, and the historical Jesus.)

References and further reading

Clayton, J.P. (ed.) (1976) *Ernst Troeltsch and the Future of Theology*, Cambridge: Cambridge University Press. (The best collection of essays around; includes a bibliography of secondary sources and of English translations.)

Coakley, S. (1988) *Christ without Absolutes: A Study of the Christology of Ernst Troeltsch*, Oxford: Clarendon Press.(Explores how Troeltsch treated Christian claims about Jesus within his pluralist, relativist framework.)

Drescher, H.-G. (1991) *Ernst Troeltsch. Leben und Werk*, Göttingen: Vandenhoeck & Ruprecht; trans. J. Bowden, *Ernst Troeltsch, His Life and Work*, London: SCM Press, 1992.(The most detailed intellectual biography available.)

Graf, F.W. (1982) *Ernst Troeltsch Bibliographie*, Tübingen: Mohr, 4 vols.(Contains brief, mainly bibliographical comments to many entries.)

Niebuhr, H.R. (1960) 'Introduction', in E. Troeltsch, *The Social Teaching of the Christian Churches*, New York: Harper.(A prominent American theologian, who was influenced by a curious mix of Karl Barth and Ernst Troeltsch, pays tribute to the latter.)

Rubanowice, R.J. (1982) *Crisis in Consciousness: The Thought of Ernst Troeltsch*, Tallahassee, FL: University Press of Florida.(Describes Troeltsch as a representative of an early twentieth-century crisis, in which ideas of cultured civilization became problematized.)

Yamin, G.S., Jr (1993) *In the Absence of Fantasia: Troeltsch's Relation to Hegel*, Gainesville, FL: University Press of Florida.(Although Troeltsch criticized Hegel, the latter had a profound effect on him.)

Trotsky, Leon (1879-1940)

Trotsky's chief claim to attention is as the leader of the Russian Revolution who opposed the consolidation of the Stalin regime in the Soviet Union and sought to dissociate the classical Marxist tradition from that regime and its official ideology. In doing so, however, he developed a version of Marxism which sought to give proper place to the 'subjective factor' in history, and at the same time to integrate Marx's social theory into a broader, dialectical theory of nature.

Trotsky's life belongs chiefly to the principal drama of the twentieth century - the Russian Revolution of October 1917 and the emergence, in its aftermath, of the Stalinist regime which survived until the collapse of the Soviet Union in 1991. Born in Yankova, Lev Davidovich Bronstein assumed the name Leon Trotsky while escaping exile in Siberia in 1902. A revolutionary activist from his teens, he presided over the St Petersburg Soviet of Workers' Deputies during the 1905 Revolution, and held the same position when, as a leader of the Bolshevik Party, he helped organize the uprising of October 1917. As War Commissar, Trotsky played a pivotal role in securing a Bolshevik victory in the post-revolutionary Civil War. Defeated in the factional struggles which led to the dominance of Josef Stalin in Party and state, he was exiled in 1929. Trotsky spent his remaining years desperately seeking to rally socialist opposition to the Stalin regime. Isolated and persecuted, he was murdered by a Soviet agent.

Trotsky's voluminous writings reflect this life. They represent an attempt to preserve and to extend the classical Marxist tradition in opposition to Stalinism and its official ideology of Marxism-Leninism. This stance distinguishes Trotsky from Western Marxists such as Adorno, Marcuse and other members of the Frankfurt School, who kept their distance equally from the Soviet regime and from classical Marxism, which they tended to regard as excessively and reductively materialist (see Frankfurt School; Critical theory).

Trotsky was, by contrast, an unapologetically orthodox Marxist. Thus he sought to provide an explanation of Stalinism which took as its starting point the economic circumstances of the Bolshevik regime after 1917. It was the isolation of the Revolution and the consequent prevalence of scarcity in a backward, predominantly peasant country that accounted for the rise of Stalin and the *nomenklatura*. At the same time, however, he vehemently denied that Marxism was a version of economic determinism. His theory of permanent revolution, first developed after the 1905 Revolution, sought to show how processes of 'uneven and combined development' on a world scale made possible socialist revolutions even in underdeveloped countries such as Tsarist Russia. Similarly, Trotsky's political and historical writings constantly stress that the outcomes of social and political crises are not predetermined by the economic contradictions that cause them, but depend on the 'subjective factor' - human agency in the shape of social classes, political organizations and even individual leaders.

A similar combination of orthodoxy and flexibility characterizes Trotsky's strictly philosophical writings, which are devoted to a defence of the proposition that Marxism possesses a dialectical method which sets it apart from other social theories. Marx inherited from Hegel the idea that contradictions exist in reality and developed it into a theory of social contradictions - for example, the structural conflict between capitalist relations of production and the development of the productive forces. Engels then extended this, most notably in *Dialectics of Nature* (1925), into a general theory of nature, in which certain dialectical laws govern both physical and social worlds.

'Dialectical materialism' thus understood became a basic element of Marxist orthodoxy (see Dialectical materialism). During the 1930s, however, some of Trotsky's US followers, influenced by pragmatism, sought to separate Marxist social theory from this broader metaphysical doctrine. Trotsky responded vigorously, arguing that abandonment of the dialectic would lead sooner or later to the rejection of Marxism itself. The dialectical method was indispensable for scientific inquiry. Following Hegel, Trotsky argued that formal logic is unable to comprehend any process of change:

> It wishes to content itself with motionless imprints of a reality which consists of eternal motion. Dialectical thinking gives to concepts, by means of closer approximations, corrections, concretizations, a richness of content and flexibility; I would even say a succulence which to a certain extent brings them close to living phenomena.

([1942] 1973: 50)

The more scientific discoveries reveal nature itself to be undergoing processes of historical transformation, the more important the dialectical method becomes: thus Darwin was an ‘unconscious dialectician’ ([1942] 1973: 51-4). But this line of argument poses the following dilemma. If Darwin could get as far as he did ‘unconsciously’, what need of a ‘conscious’ dialectic? If, on the other hand, research is held to depend on dialectical thinking, does this not license the kind of ideological terror suffered by Soviet biologists during the Lysenko era?

Trotsky, however, loosened Engels’ conception of universal ‘laws of the dialectic’. Only ‘the conversion of quantity into quality’ - the concept of qualitative transformation - is ‘the fundamental law of dialectics’ because ‘the entire universe’ is ‘a product of formation and transformation’ (1986: 88-9). One such ‘transition from quantity to quality’ is ‘"the autonomy" of psychological phenomena’: ‘the psyche, arising from matter, is "freed" from the determinism of matter, so that it can - by its own laws - influence matter’. Trotsky compared this to the interaction between base and superstructure - ‘politics grows out of economics in order for it in turn to influence the base’ - but admitted that the relationship between mind and body is ‘incomparably more puzzling’ (1986: 106-7). Nature is thus conceived as stratified - divided into distinct domains each governed by its own laws - yet forming an integrated and historically evolving totality. The dialectic is not a substitute for empirical research into these different domains, but ‘it gives investigative thought elasticity, helps it to cope with ossified prejudices, arms it with invaluable analogies, and educates it in a spirit of daring, grounded in circumspection’ (1986: 92). Although even many Marxists regard the idea of a dialectic of nature as an embarrassment, Trotsky would no doubt argue that it has been vindicated now that cosmologists tell us of a universe formed through a series of phase transitions producing radical changes in physical structures.

See also: Marxist philosophy, Russian and Soviet

ALEX CALLINICOS

List of works

Trotsky, L. (1909) *1905*, trans. A. Bostock, Harmondsworth: Penguin, 1973.(An account of the Russian Revolution of 1905 containing a particularly good statement of Trotsky’s theory of permanent revolution.)

Trotsky, L. (1930) *Moia zizhn’: opyt avtobiografii*, trans. as *My Life*, Harmondsworth: Penguin, 1975.(A polemical autobiography; the opening chapters, on Trotsky’s early years, are of exceptional literary quality.)

Trotsky, L. (1931-2) *Istoriia russkoi revoliutsii*, trans. M. Eastman, *The History of the Russian Revolution*, London: Pluto Press, 1977.(A historiographic masterpiece, and Trotsky’s most powerful statement of his conception of the role of human agency in the historical process.)

Trotsky, L. (1936) *The Revolution Betrayed*, trans. M. Eastman, New York: Pathfinder, 1972.(Trotsky’s definitive account of the nature and origins of Stalinism.)

Trotsky, L. (1942) *In Defence of Marxism*, New York: Pathfinder, 1973.(Trotsky’s last thoughts on Stalinism and the future of Marxism; also contains his most celebrated defence of the dialectic.)

Trotsky, L. (1969) *The Permanent Revolution and Results and Prospects*, trans. J.G. Wright and B. Pearce, New York: Pathfinder.(The two main statements of the theory of permanent revolution.)

Trotsky, L. (1986) *Notebooks 1933-1935*, trans. P. Pomper, New York: Columbia University Press.(A previously unpublished manuscript containing Trotsky’s most sustained discussion of the dialectic.)

Trotsky, L., Dewey, J. and Novack, G. (1978) *Their Morals and Ours*, New York: Pathfinder.(A debate on revolutionary morality in which Dewey makes some shrewd points against Trotsky.)

Further reading and references

Anderson, P. (1984) ‘Trotsky’s Interpretation of Stalinism’, in T. Ali (ed.) *The Stalinist Legacy*, Harmondsworth: Penguin, 118-28.(A sympathetic but critical assessment of *The Revolution Betrayed*.)

Beilharz, P. (1987) *Trotsky, Trotskyism and the Transition to Socialism*, Beckenham: Croom Helm.(Trotsky interpreted as a fatalist and evolutionist Marxist.)

Callinicos, A. (1990) *Trotskyism*, Milton Keynes: Open University Press.(A brief critical survey of Trotsky’s thought and the traditions stemming from it.)

Cliff, T. (1955) *State Capitalism in Russia*, London: Bookmarks, 1988.(The most important single attempt to develop Trotsky’s analysis of Stalinism.)

Deutscher, I. (1954-63) *The Prophet Armed: Trotsky 1879-1921*, *The Prophet Unarmed: Trotsky 1921-29* and *The Prophet Outcast: Trotsky 1929-40*, Oxford: Oxford University Press, 3 vols, 1970.(An outstanding biography.)

Engels, F. (1925) *Dialectics of Nature*, Moscow: Progress, 1972.(The main source for the view that both physical and social reality is governed by certain universal, dialectical laws.)

Knei-Paz, B. (1978) *The Social and Political Thought of Leon Trotsky*, Oxford: Clarendon Press.(A systematic critical treatment of the theory of permanent revolution which relates it to more recent social-scientific literature on development and modernization.)

Levins, R. and Lewontin, R. (1985) *The Dialectical Biologist*, Cambridge, MA: Harvard University Press.(Two working scientists seek to show the relevance of dialectical concepts to contemporary biology.)

Löwy, M. (1981) *The Politics of Combined and Uneven Development*, London: Verso.(A historical and theoretical survey of the idea of permanent revolution.)

Novack, G. (1972) *Understanding History*, New York: Pathfinder.(Includes an influential discussion of the concept of uneven and combined development.)

Rees, J. (1990) ‘Trotsky and the Dialectic of History’, *International Socialism* 47: 113-35.(A discussion of Trotsky’s *Notebooks*.)

Trust

Most people writing on trust accept the following claims: trust involves risk; trusters do not constantly monitor those they trust; trust enhances the effectiveness of agency; and trust and distrust are self-confirming. Three further claims are widely accepted: trust and distrust are contraries but not contradictories; trust cannot be willed; and trust has noninstrumental value. Accounts of trust divide into three families: risk-assessment accounts, which are indifferent to the reasons why one trusts; will-based accounts, which stress the importance of the motives of those who are trusted; and affective attitude accounts, which claim that trust is a feeling as well as a judgment and a disposition to act. One of the central questions concerns when trust is justified, and, in particular, whether justified trusting can outstrip evidence for the belief that the person trusted is trustworthy. If trust can leap ahead of evidence of trustworthiness, then trust poses a problem for evidentialism, or the view that one should never believe anything without sufficient evidence. Further central questions include whether trusting is a virtue and trustworthiness morally required, while a final set of questions concerns the role of trust in politics and the connection between interpersonal trust and trust in institutions.

1 Common observations about trust

Trust is part of love and friendship, and vital for any cooperative task. Trust's pervasiveness can render it almost invisible; we notice how much we trust only when trust has been betrayed.

Any account of trust should aim to be compatible with and seek to explain a number of commonly accepted observations. There is a core of four such observations which are generally accepted by people writing about trust and a further three which, while not universally accepted, are frequently noted. The four core observations about trust are as follows.

(1) Trust involves risk. A person who trusts another is willing both to run the risk of letting them near things that they care about, and to rely on their words and promised actions. When trust is betrayed the truster is usually left worse off than had they refrained from trusting. (2) Those who trust are willing to forgo an immediate accounting of how or even whether the one trusted has responded to trust with trustworthiness, and to allow the one trusted some discretion as to how to fulfil that trust. While contractual relations might involve a degree of trust, noncontractual trust relations are distinguished from contractual arrangements by their relative open-endedness. When there is trust, there is no need to specify exactly what must be done to fulfil it. The flexibility of trust relations in part explains trust's usefulness: contingency and ignorance often make us unable to specify what we are counting on the other to do; trust removes the need to do so. (3) Trust enhances the effectiveness of agency: it enables us to attempt cooperative tasks that could not rationally be attempted without trust, and, as Luhmann (1973) points out, it lets us inhabit a simpler and less threatening world, a world in which we need not plan for every contingency. (4) Trust and distrust are self-confirming, distrust even more strongly so than trust. When there is distrust, cooperative relations will be minimized and there will be little chance of getting evidence that the person is after all trustworthy. Trust and distrust also influence how we interpret the actions and behaviour of others: when there is room for a more or less favourable interpretation, we will adopt the interpretation which most fits with our view of their trustworthiness. In addition, sometimes displaying trust or distrust itself affects the behaviour of the person in question and tends to make them either trustworthy or untrustworthy.

Three additional claims enjoy wide agreement. (5) Trust and distrust are contraries but not contradictories. Just because one does not trust, it does not follow that one thereby distrusts, for there can be a neutral stance of neither trusting nor distrusting. (6) Trust cannot be willed. One cannot will oneself to trust in the acknowledged presence of substantial reasons to distrust, though one may be able to make oneself act as if one trusted. (7) The value of trust is not exhausted by its instrumental role in making possible cooperative relations. Trust, when well placed, is valuable in itself and is a constitutive part of other things valuable in themselves, such as love and friendship (see Friendship §1).

2 Accounts of trust

Accounts of trust divide into three families, depending on the degrees to which they distinguish trust from reliance, and to which they emphasize affect or feeling.

The first family of accounts view trust as a form of risk assessment (see Risk assessment). Such accounts tend to begin by considering trust in institutions and then extend the results of that reflection to cover interpersonal trust. Gambetta offers the following definition of trust:

> trust (or, symmetrically, distrust) is a certain level of subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both *before* he can monitor such action (or independently of his capacity ever to be able to monitor it) *and* in a context in which it affects *his own* action.
> (1988: 217; original emphasis)

The definition emphasizes the way trusting makes one vulnerable, but it does not distinguish among the reasons why one might assign a sufficiently high degree of probability to agents' performing an action. One might think agents will perform that action out of fear or stupidity, or because it coincides with their own self-interest, or because they have, and wish to display, goodwill towards those who are counting on them. Gambetta includes all these alike in his definition. Proponents of such accounts are thus unlikely to accept that trust is anything other than instrumentally good, for if trust can rest on any assessment of motivation it is hard to see how it could be valuable in itself. Several writers hold that accounts which ignore motivation confuse trust with mere reliance.

Trust is distinguished from reliance in two ways. One can rely on something, or someone, simply because one has no choice but to do so: the bridge, rickety and unsafe though it might be, is the only way to escape the oncoming fire and so one is forced to rely on it and hope for the best. When one trusts, in contrast, one has the confident expectation that the one trusted will respond favourably. According to Thomas Hobbes, '[t]rust is a passion proceeding from the belief of him whom we expect or hope for good, so free from doubt that upon the same we pursue no other way' (1650).

Trust is thus distinguished from hope by confidence of expectations. But confident expectations can be grounded in all manner of psychological features. Most who clearly distinguish trust from reliance suggest that the difference is that in trust one relies on goodwill rather than on some other aspect of a person's psychology. Thus, the second family of accounts of trust, will-based accounts, tend to begin by considering interpersonal trust. Annette Baier's article, 'Trust and Anti-trust', offers an entrusting model. On this model, trusting is a three-place relation: A trusts B with valued thing C (1986: 236); and trusting is distinguished from mere reliance in so far as trust is always reliance on the goodwill (and competence) of another. We value another's goodwill towards us in itself, not merely because of its usefulness; therefore, on will-based accounts, the value of well-placed trust is not exclusively instrumental.

The first two families of accounts of trust take trust to involve both cognitive and conative aspects, both judgments about the likely actions of another and readiness to undertake risky actions on the basis of those judgments. The third family of accounts, affective attitude accounts, adds an affective aspect: trust is a feeling as well as a judgment and a disposition to act. '[Trust] has a special "feel", most easily acknowledged when it is missed, say, when one moves from a friendly, "safe" neighbourhood to a tense, insecure one' (Baier 1992: 112-3).

3 When trust is justified

If our trust in strangers can be justified, then it seems that we sometimes trust without good evidence for the belief that the one trusted is trustworthy. Thus trust, like faith, might seem to conflict with evidentialism, or the view that we should never believe anything without sufficient evidence (see Faith). In response, we might claim that in certain social climates we have sufficient evidence that strangers will be trustworthy and that outside such climates our trust in strangers is unjustified. In this way we reconcile trust with evidentialism by requiring trust to meet the normal justification standards for belief. Alternatively, we might, following some writers on faith, reject evidentialism and claim that trust does not require the same evidence as other beliefs but can still be justified. Finally, we might stress the affective aspect of trust and claim that since evidentialism governs the adoption of beliefs and not the adoption of affective attitudes, trust falls outside the scope of the evidentialist requirement.

It might be thought that trust should be our default stance: if we distrust without solid evidence of untrustworthiness our willingness to assume ill will or unreliability shows a lack of respect for the person distrusted (see Respect for persons). At least initially, everyone should be given the benefit of the doubt. On this view, willingness to trust is a virtue (see Virtues and vices §§2-3). Similarly, it might be thought that we are

required to respond to trusting with trustworthiness, and so when trust is betrayed the moral fault always lies with the one who responded to trust with untrustworthiness. Both of these claims ignore how trust and trustworthiness can be used to foster exploitation, to hide abuse and to promote cooperation in the service of unworthy ends. In this respect, trust and trustworthiness are like loyalty. Loyalty can promote bad causes as well as good ones. Sometimes, one is required to respond to trust with selective untrustworthiness, and to demands for loyalty with judicious disloyalty. Just because well-placed trust has noninstrumental value, it need not follow that poorly placed trust has such value.

What conditions are likely to give rise to trust and how is trust to be maintained and extended? Good's contribution to the Gambetta volume shows that some insight into these questions can be gained from game theory and from laboratory studies of cooperation (see Decision and game theory §§3, 6, 8). Trustworthy behaviour is more likely to be shown by those who are engaged in ongoing cooperative exchanges since, by being untrustworthy, they risk losing the benefits of all future cooperative exchanges. Likewise, communication facilitates trust since the more we know about persons, the more certain we can be about how they will act. Trust is more easily extended in incremental steps because we do not deliberate extensively about small extensions to previous policy; further, trust is more easily developed through small initial risks. These observations suggest strategies for those who would initiate and develop trusting relations. Baier (1992) offers a sensitive discussion of the functional virtues needed to maintain trust. Chief among them are tact and willingness to forgive breaches in trust, though judgment is indispensable in working out which breaches should be forgiven, and which not.

4 Trust and social and political institutions

Social and political institutions (such as licensing boards, marketplaces and schools) can make possible and strengthen interpersonal trust, by removing incentives to untrustworthiness and so creating a confluence of motives for trustworthiness, by helping to create stable expectations and by providing a framework for understanding communication and interpreting behaviour.

If social and political institutions can underwrite interpersonal trust, could it be possible to replace interpersonal trust with trust in well-designed institutions? Trusting in institutions rather than in individuals would eliminate the need for information about particular persons filling institutional roles. We could count on whoever happens to fill a well-designed social role. In large impersonal societies, where we must frequently interact with persons unknown to us, the advantages of shifting from interpersonal to institutional trust are obvious. The contrast between institutional and interpersonal trust is not hard and fast, and a particular instance of trust, say trust in one's physician, could involve elements of both. That the two sorts of trust can overlap suggests we need not think it possible to replace interpersonal trust with institutional trust to think it worth heading in such a direction.

One reason for being suspicious of the claim that we should trust in institutions is the concern that doing so requires that we forswear vigilance and forgo checking. Consider the claim, advanced by John Locke (1690), that the relation between citizens and government is one of trust. If trust is generally appropriate towards parents and always appropriate towards God, it might be thought just plain foolish towards governments. This objection is based, however, on a misunderstanding of the way theorists have thought that the notion of trust enters into the relation between citizens and government. According to Locke, citizens entrust the government with coercive power to act on their behalf for their benefit and for the protection of their rights. When governments fail to fulfil their trust, they lose legitimate authority (see Locke, J. §10). The people also retain the power to judge when a breach of trust has occurred:

> for who shall be *Judge* whether his Trustee or Deputy acts well, and according to the Trust reposed in him, but he who deputes him, and must, by having deputed him, have still a Power to discard him, when he fails in his Trust?
>
> (1690: 445; original emphasis)

That the relation between government and governed is one of trust is thus compatible with holding the government accountable, provided that we understand the relation in question to be one of entrusting. However, for societies to flourish they need political institutions that merit trust and allow citizens to be confident that power is safely entrusted to them.

How significant one finds trust to be in public life will depend in part on the account of trust that one adopts. Thus, while risk-assessment accounts encounter problems thinking about interpersonal trust, will-based accounts are less readily generalized from interpersonal relations to social and political institutions.

See also: Hope; Professional ethics; Promising; Reciprocity; Solidarity; Truthfulness; Vulnerability and finitude; Xin (trustworthiness)

KAREN JONES

References and further reading

Baier, A. (1986) 'Trust and Anti-trust', *Ethics* 96: 231-60. (The best introduction to the topic and the single most influential article in the area.)

Baier, A. (1992) 'The Pathologies of Trust' and 'Appropriate Trust', *Tanner Lectures on Human Values*, vol. 13, Salt Lake City, UT: University of Utah Press.(Delivered originally at Princeton University in 1991, these lectures develop the account given in 'Trust and Anti-trust'.)

Baker, J. (1987) 'Trust and Rationality', *Pacific Philosophical Quarterly* 68: 1-13.(A useful introduction to the problems of trust and evidentialism.)

Becker, L. (1996) 'Trust as Noncognitive Security About Motives', *Ethics* 107: 43-61.(Emphasizes the noncognitive aspects of trust.)

Gambetta, D. (1988) 'Can We Trust Trust?' in D. Gambetta (ed.) *Trust: Making and Breaking Cooperative Relations*, New York: Blackwell, 1988, 213-37.(Addresses the question of whether and when it is rational to trust.).

Gambetta, D. (ed.) (1988) *Trust: Making and Breaking Cooperative Relations*, New York: Blackwell.(A collection of essays most of which provide risk-assessment accounts of trust. Contains articles by Patrick Bateson, Partha Dasgupta, John Dunn, Ernest Gellner, Keith Hart, Geoffrey Hawthorn, Edward Lorenz, Niklas Luhmann, Anthony Pagden and Bernard Williams. Dunn's contribution, 'Trust and Political Agency', is an especially useful discussion of the role of trust in politics.)

Good, D. (1988) 'Individuals, Interpersonal Relations, and Trust', in D. Gambetta (ed.) Trust: Making and Breaking Cooperative Relations, New York: Blackwell, 1988, 31-48.(Applies game theory to the question of when to trust.)

Govier, T. (1992) 'Trust, Distrust, and Feminist Theory', *Hypatia* 7: 16-33. (A survey of feminist reflections on trust.)

Hardin, R. (1991) 'Trusting Persons, Trusting Institutions', in R. Zeckhauser (ed.) *Strategy and Choice*, Cambridge, MA: MIT Press, 185-209. (Presents a critique of will-based and affective attitude accounts of trust.)

Hobbes, T. (1650) 'Human Nature', part I of *The Elements of Law, Natural and Politic*, Oxford: Oxford University Press, 1994, chapter IX, section 9.(Contains a brief treatment of trust in the context of a discussion of the passions.)

Locke, J. (1690) *Two Treatises of Government*, ed. and with notes by P. Laslett, Cambridge: Cambridge University Press, 1960.(Argues that the relation between citizens and government is one of trust. The editor's introduction to this critical edition deals with Locke's use of the concept of trust.)

Luhmann, N. (1973) 'Trust: A Mechanism for the Reduction of Social Complexity', trans. H. Davis, J. Raffman and K. Rooney, in T. Burns and G. Poggi (eds) Trust and Power, Chichester, NY: Wiley, 1979, 4-103.(A complex and subtle discussion of trust from a sociological perspective. Quite difficult.)

Thomas, L. (1990) 'Trust, Affirmation, and Moral Character: A Critique of Kantian Morality', in O. Flanagan and A. Rorty (eds) Identity, Character and Morality, Cambridge, MA: MIT Press, 1990, 235-57.(Offers a detailed will-based account of trust and argues that trust has noninstrumental value.)

Truth, coherence theory of

The term 'coherence' in the phrase 'coherence theory of truth' has never been very precisely defined. The most we can say by way of a general definition is that a set of two or more beliefs are said to cohere if they 'fit' together or 'agree' with one another. Typically, then, a coherence theory of truth would claim that the beliefs of a given individual are true to the extent that the set of all their beliefs is coherent. Such theories, thus, make truth a matter of a truth bearer's relations to other truth bearers rather than its relations to reality. This latter implication is the chief hindrance to plausibility faced by coherence theories, and most coherence theorists try to escape the problem by denying that there is any extra-mental reality.

1 Essential elements of coherence theories

John Locke voiced what might be the earliest statement of a coherence theory of truth when he said in *An Essay Concerning Human Understanding* that ideas are true if they agree with each other. But a few pages later he reverted to a correspondence theory by declaring that there were 'chimerical' truths and real truths, and what distinguishes the latter from the former is that real truths agree with reality. When the character Philonous, representing George Berkeley in the latter's *Three Dialogues*, is asked how his idealist doctrine can distinguish fact from fiction, reality from illusion, and waking life from dreaming, he replies that fictional ideas, illusions, and dreams are known to be such because they are disconnected from the great mass of the rest of our ideas. This can be taken as an endorsement of what would today be called a coherence theory of truth. Moreover, coherence theories were in their golden age during the nineteenth century, the heyday of ontological idealism in Western philosophy. It is not surprising that coherence theories should cohabit with anti-realist ontology. If reality itself is just a system of mental entities - thoughts - then not even the true thoughts can correspond to some *other* reality, since there is no other. So, it would be natural to suppose that the truth of a true thought must consist in its relations to other thoughts (see Truth, correspondence theory of §2).

Nineteenth-century coherence theories provided a more detailed definition of coherence: a set of beliefs coheres if and only if each member of the set is consistent with any subset of the others, and if each is implied (inductively, if not deductively) by all of the others taken as premises or, according to some coherence theories, each is implied by each of the others individually.

Unfortunately, the distinction between theories of justification and theories of truth was not deployed in the nineteenth century. As a result, there are few individuals to whom one can unhesitatingly attribute a coherence theory of truth, distinct from a coherence theory of justification. One to whom we can make such an attribution is Brand Blanshard: ' coherence is the sole criterion of truth. We have now to face the question whether it also gives us the nature of truth… one may reject coherence as the definition of truth while accepting it as the test' (Blanshard 1941: 260). *Pure* truth, he says, is a fully coherent set of beliefs, and ' fully coherent knowledge would be knowledge which in every judgment entailed, and was entailed by, the rest of the system'. But even this understates the entailment relations between the members of the system: ' every proposition would be entailed by the others jointly and even singly'. Since an inconsistent subset of beliefs would trivially entail anything and everything, and since presumably Blanshard does not want that kind of entailment to be a truth-making relation, there is an implied condition of consistency on any system of beliefs before it will count as true. However, Blanshard does not want to say that two such systems, each coherent within itself but inconsistent with the other, are *both* true. A true system of beliefs would be one 'in which everything real and possible is coherently included'. So, a purely true system would be one that gives us a complete picture of the world. This theory can be expressed in the following formula:

> For each belief, *b*, *b* is purely true if and only if *b* is a member of a consistent set of beliefs that among them give a complete picture of the world and each of which entails each of the others.

But pure truth has never been attained, so Blanshard, typically for a coherence theorist, proposes that truth comes in degrees: ' A given judgment is true in the *degree* to which its content could maintain itself in the light of a completed system of knowledge, false in the *degree* to which its appearance there would require transformation' (Blanshard 1941: 304). This is captured in the formula:

For each belief, *b*, *b* is true to degree *n* if and only if *n* per cent of the content of *b* would be present in a purely true system of beliefs.

2 Objections to coherence theories

The most common objection to coherence theories is that the conditions they place on truth are too weak. Traditionally, this objection is expressed with the quip that on a coherence theory a well-written novel would be true, literally not figuratively. Taken at face value, such an objection is only telling against those theories which require mere mutual consistency for coherence. But there are no such coherence theories. Blanshard's, as we have seen, requires mutual entailment among the beliefs in a set if the set is to count as coherent. Nevertheless, it is certainly possible to construct two sets of consistent, mutually entailing propositions which contradict each other and which, thus, cannot both be true. Blanshard of course, following Bradley (1914), tries to eliminate this possibility by insisting that final truth is found only in a set which is not only coherent but also *complete*. Yet even this might not be a strong enough condition on truth. Could there not be two complete pictures of the universe, inconsistent with each other, but each internally coherent? Since there is not even one such complete picture, an answer to that question must remain uncertain; but it is worth noting that there are large domains of phenomena for which humankind possesses two (or more) competing theories, each of which seems to have approximately equal degrees of internal coherence. There are for example the geometries of Euclid and Riemann; and one thinks also of Western medicine, on the one hand, and the traditional medicine of China, on the other.

Since the concepts of entailment and consistency are usually defined in terms of truth, a theory which, like coherence theories, defines the latter concept in terms of entailment or consistency is uninformatively circular. One might try to eliminate the circularity by defining entailment and consistency in terms of a certain set of rules of derivation; for example '*p* entails *q* if and only if *q* can be derived from *p* via one or more of the rules in *R*', where *R* is a set of rules of inference whose members refer only to the syntactic structure of propositions, not to their truth values. One such rule might be 'from propositions of the form *p* and "if *p*, then *q*", one may derive the proposition *q*'. But suppose *R* contains 'from a proposition *p*, one may derive the proposition "not-*p*"'. Then no one would take seriously a theory claiming that a set of propositions that are mutually 'entailing' via such a set of rules is, for that reason, true. Hence, defining entailment and consistency proof-theoretically requires that the set of rules in question be restricted to those which are *truth*-preserving; and so the circularity problem arises again.

Defining the truth-value of a truth bearer in terms of its relations with other truth bearers instead of its relations with the world is initially considered implausible by most, because doing so seems to make it logically possible for 'Kendall is red-headed' to be true even if she is not in fact red-headed, and for 'Emory is handsome' to be false even if he is, in fact, handsome. The only way to make such situations impossible is to embrace an anti-realist notion of what it is to be a fact. For example, if we define a fact to be an idea in the mind of God and declare that God's ideas are a complete and perfectly coherent set, then facts and truths are not disjoint. But since the plausibility of the coherence theory depends on an anti-realist ontology, the difficulties facing such ontologies become difficulties for the coherence theory (see Idealism; Realism and antirealism).

3 Scepticism and coherence theories

Given the difficulties facing coherence theories of truth, one might wonder why some have found them attractive. There are remarkably few positive arguments for such theories in the literature, and those which can be found are nearly all variations on what might be called the argument from scepticism - that it is very hard (some would say it is impossible) ever to be justified in believing that this or that is a mind-independent fact. Our subconscious minds so thoroughly filter and modify observational input that what we take to be facts have no resemblance to any mind-independent facts. So, if we make the existence of a mind-independent fact a necessary condition for the truth of a belief (or sentence, or whatever) we would never be justified in thinking that *any* belief is true. Thus, such a theory of truth would entail scepticism. Given that scepticism is wrong, so too must be any such theory of truth.

But begging the question is no more respectable when directed against scepticism than when directed against any other philosophical proposal. A sceptical thesis such as 'No belief is justified as true' (which does not say 'I am justified in believing that no belief is justified as true') is not self-contradictory and thus might well be correct. That pill will be less bitter if we remember that it need not imply that no belief is any more rational than any other;

for the rationality of beliefs could be defined in terms of their possession of values, such as explanatory power, *other than* the value of being justified as true.

A second problem with the argument from scepticism for coherence theories of truth is that the latter theories do not really solve the problem of scepticism anyway. Coherence theorists of truth, in effect if not in intention, try to deal with scepticism by redefining truth so as to make it more attainable. They define it in such a way that it becomes easy to have beliefs justified as probably true. But no satisfying rebuttal to scepticism can come from this strategy, for the problem of scepticism is the concern that our beliefs may not be justified as accurate reflections of a mind-independent world. The manoeuvre under consideration tells us in the large print that we have adequate justification for most of our beliefs about the world; but, simultaneously, the small print tells us that this 'world' we believe in consists of mental constructs and may not even vaguely resemble the mind-independent world. Of course, a coherence theorist of truth believes we are mistaken in thinking there is any mind-independent world in the first place. But to say this is to surrender to scepticism, not to refute it. Davidson makes the same point thus: '[Such theories] are sceptical in the way idealism or phenomenalism are sceptical; they are sceptical not because they make reality unknowable, but because they reduce reality to so much less than we believe there is' (Davidson 1990: 298).

See also: Meaning and truth; Truth, deflationary theories of; Truth, pragmatic theory of

RICHARD L. KIRKHAM

References and further reading

Berkeley, G. (1713) *Three Dialogues between Hylas and Philonous*, Peru, IL: Open Court, 1969.(A readable and entertaining summary of the author's philosophy.)

Blanshard, B. (1941) *The Nature of Thought*, New York: Macmillan, vol. 2.(The author's coherence theory of truth is presented on pages 260-308.)

Bradley, F.H. (1914) *Essays on Truth and Reality*, Oxford: Clarendon Press.(A coherence theory which unfortunately mixes together issues of justification with issues of truth.)

Davidson, D. (1990) 'The Structure and Content of Truth', *Journal of Philosophy* 87: 279-328.(A summary of the author's mature views on truth and semantics.)

Kirkham, R.L. (1992) *Theories of Truth: A Critical Introduction*, Cambridge, MA: MIT Press.(An introduction to theories of truth. Sects 3.5-3.6 expand on the issues in this entry.)

Locke, J. (1690) *An Essay Concerning Human Understanding*, New York: E.P. Dutton, 1976.(Book IV, Chapter V contains the author's theory of truth.)

Truth, correspondence theory of

The two oldest theories of truth in Western philosophy, those of Plato and Aristotle, are both correspondence theories. And if the non-philosopher can be said to subscribe to a theory of truth, it would most likely be to a correspondence theory; so called because such theories are often summed up with the slogans 'truth is correspondence with the facts' or 'truth is agreement with reality'. Aristotle puts it thus: 'to say that [either] that which is is not *or that which is not* is, *is a falsehood; and to say that that which is is and that which is not is not, is true'. In epistemology, such theories offer an analysis of that at which, supposedly, investigation aims: truth. But correspondence theories are also now thought to play important roles in philosophical semantics and in the physicalist programme, which is the task of reducing all non-physical concepts to the concepts of logic, mathematics, and physics.*

1 Two kinds of correspondence theory

Correspondence theories come in two varieties - correspondence as correlation, and correspondence as congruence. The former variety asserts that every truth bearer (proposition, sentence, belief, and so on) is correlated to a possible fact. If the possible fact to which a given truth bearer is correlated actually obtains, the truth bearer is true; otherwise it is false. What the correspondence-as-correlation theory does not claim is that the truth bearer depicts, or is structurally isomorphic with, the possible fact to which it is correlated. Rather, a truth bearer as a whole is correlated to a possible fact as a whole. A correspondence-as-congruence theory does claim that truth bearers and the possible facts to which they correspond have parallel structures. Representing this second school is Bertrand Russell, who affirms that Othello's belief that Desdemona loves Cassio is a complex relation between Othello (the subject), Desdemona (an object-term), Cassio (another object-term) and loving (the object-relation) (Russell 1912). Truth requires a congruence between this four-term relation and a second, three-term relation called 'a fact' which has Desdemona, loving, and Cassio (in that order) as its terms. If such a three-term relation exists in reality, then Othello's belief is true. If there is no such fact, the belief is false (see Russell, B. §10).

J.L. Austin has offered a correspondence-as-correlation theory. Truth is considered as a single, four-term relation between a statement, a sentence, a state of affairs (that is, a possible fact), and a *type* of state of affairs. For Austin, a statement is the information conveyed by a declarative sentence. So a sentence is the medium in which a statement is made. And the meaning of statements is a matter of two kinds of conventions that have evolved in our language. First, there are descriptive conventions correlating *sentences* with types of states of affairs. Second, there are demonstrative conventions correlating *statements* to particular states of affairs. Thus, ' A statement is said to be true when the historic [that is, particular] state of affairs to which it is correlated by the demonstrative conventions (the one to which it "refers") is of a type which the sentence used in making it is correlated by the descriptive conventions' (Austin [1950] 1970: 121-2). Hence, 'there is no need whatsoever for the words used in making a true statement to "mirror" in any way, however indirect, any feature whatsoever of the situation or event'. The correspondence between the truth bearer and the world is '*absolutely and purely* conventional'. For example, 'the cat is on the mat' would in an ordinary context refer to the present state of affairs in which speaker and hearer find themselves, along with a cat and a nearby mat. Therefore, if a cat is on a mat in the state of affairs in which the speaker is located, the statement is true because the present state of affairs is of just the type described by the sentence.

In favour of the correspondence-as-correlation view is the fact that some sentences (for example, the Latin word 'sum', which translates as 'I am') are atomic meaning units which have no internal semantic structure and thus could correlate with possible facts only whole-for-whole.

2 The common essence of correspondence theories

If we strip out of Russell's theory of truth the assumption that beliefs are four-termed relations, and strip out of Austin's theory his assumptions about meaning being a matter of conventions, we can then encapsulate the two stripped-down theories of truth in these two formulas:

Russell: A belief is true if and only if it is a belief that an object x bears relation R to another object y, and x does

bear relation *R* to *y*.

Austin: A statement is true if and only if it expresses a state of affairs, and that state of affairs obtains.

If we abstract the common elements of these two theories, we get the following thesis which can be taken as expressing the common denominator of all correspondence theories of truth:

Correspondence: A truth bearer is true if and only if it corresponds to a state of affairs and that state of affairs obtains.

But it must be emphasized that the 'corresponds to' in this thesis does not express some particular relation. Rather it should be thought of as a placeholder which any given correspondence theory would replace with some particular and familiar relation. Which relation a particular correspondence theorist would cite will vary, depending on their choice of truth bearer. For beliefs, the relation would be 'is a belief that'. For linguistic entities such as sentences, 'says that' or 'means that' or 'expresses' would be more appropriate choices. If a proposition is the bearer in question, then the appropriate relation would depend on how one defines propositions. (The differences between the formula expressing Russell's theory and the schema are attributable to Russell's conviction that the belief must be isomorphic to the corresponding fact.)

Note that the 'obtains' in the correspondence schema does not require that the state of affairs should obtain mind-independently. It is perfectly possible to hold that truth consists in correspondence with facts and to hold also that facts are mind-dependent entities. McTaggart (1921) endorsed such a correspondence theory, and Kant (1781) and Wilfred Sellars (1963) can be read this way too (see Kant, I.; McTaggart, J.M.E.; Sellars, W.). Thus, while most correspondence theorists have been ontological realists, the common belief that a correspondence theory is committed to realism is erroneous.

3 Objections to correspondence theories

Traditionally, three kinds of criticisms have been levelled at correspondence theories. The first concerns whatever it is the theory identifies as the truth bearer (beliefs, propositions, statements and so on): some allege that such things cannot, for various reasons, be truth bearers. Second, some allege, about whatever (facts, situations, states of affairs and so on) is identified by the theory as the correspondent of the truth bearer, that such things cannot, for various reasons, serve as the correspondents. Third, there are objections to the alleged relation between truth bearers and reality on the grounds that there is no such relation or that its nature has not been clearly explained by the theory.

The first sort of issue arises with any kind of truth theory, so nothing more will be said of it here. The third sort of objection is rarely legitimate because it aims at the sort of slogans with which correspondence theorists sum up their theories (for example, 'truth is a correspondence with the facts') and is thus irrelevant to the theories themselves, such as that of Russell and Austin, in which no such special relation makes any appearance. A closely related objection is that correspondence theories use concepts like 'belief' or 'means that' which themselves have no undisputed philosophical analysis. But such an accusation is of dubious significance precisely because it is difficult to imagine any philosophical theory against which the same objection could not be made: philosophically important topics tend to link with one another, so such an objection implicitly assumes that a theory does not solve any problem unless it solves every problem.

A correspondence theorist's description of the connection between truth bearers and facts is largely dictated by the choice of truth bearer anyway. And the latter choice, in turn, is dictated by the needs of the broader philosophical programme which has 'placed an order' for a theory of truth. Traditional epistemology is concerned with the comparative evaluation of competing theories about how our beliefs can be justified as likely to be true. But such evaluations would be impossible without first having a notion of what truth, specifically, what true *belief*, is. On the other hand, the Davidson programme aims to put a theory of truth to work in semantics and, thus, requires a theory of true *sentences* (see Davidson, D. §4).

Many have rejected the very notion of a fact as some non-linguistic entity existing in the world; facts are really just reifications of true sentences. The evidence for the claim that 'fact' is just another name for 'true sentence' is supposed to be that we cannot individuate and identify any particular fact save by using the very same words that we use to individuate and identify its corresponding sentence. There are, however, good reasons for resisting this

line of thought. First, it should be no surprise that we cannot specify a given fact save by means of the sentence to which the fact corresponds, because it could not possibly be otherwise. Suppose our language contained a series of nouns (‘Fact1’, ‘Fact2’ and so on) each referring uniquely to a different fact. We could then use these nouns to identify particular facts, but we could also use these terms to make statements. Any expression that could be used to identify facts could be used to make statements. Second, facts can enter into causal relations in a way that true sentences cannot: the fact that the war was lost caused the government to fall, but the true sentence ‘the war was lost’ cannot cause the government to fall. Third, one of the constituents of the fact that the war was lost is a certain war, but no war (distinct from the word ‘war’) can be a constituent of the true sentence ‘the war was lost’.

Some who would accept the existence of atomic facts would still object to the correspondence theory on the grounds that there are no disjunctive, conditional or negative facts; hence, there are no facts to be the correspondents of true disjunctions, conditionals, or negations. In response to this, a correspondence theorist would rightly note that, on the ordinary sense of fact, it is perfectly correct to say ‘It is a fact that either she gets here on time or I spend the night under a bridge’ or ‘It is a fact that if the donated liver is late, then the patient will die’ or ‘It is a fact that I am not going to make it’. Moreover, a correspondence theory could avoid commitment to conditional, negative or disjunctive facts by giving a recursive analysis of the truth of non-atomic truth bearers.

See also: Meaning and truth; Realism and antirealism; Truth, coherence theory of; Truth, deflationary theories of; Truth, pragmatic theory of

RICHARD L. KIRKHAM

References and further reading

Aristotle (*c.* mid 4th century BC) *Metaphysics*, in R. McKeon (ed.) *Basic Works of Aristotle*, New York: Random House, 1941.(At 1011b26, the author states the earliest correspondence-as-correlation theory of truth in Western philosophy.)

Austin, J.L. (1950) ‘Truth’, in J.O. Urmson and G.J. Warnock (eds) *Philosophical Papers*, Oxford: Oxford University Press, 1970.(A modern correspondence-as-correlation theory, originally published in *Proceedings of the Aristotelian Society* supplement to vol. 24: 111- 28.)

Bealer, G. (1982) *Quality and Concept*, Oxford: Clarendon Press.(Pages 11 and 186-203 offer a contemporary correspondence-as-congruence theory that identifies the fundamental elements of thoughts with the fundamental elements of states of affairs.)

Kant, I. (1781) *The Critique of Pure Reason*, trans. N. Kemp Smith, New York: St Martin’s Press, 1929.(The author’s words at A58 suggest a correspondence theory, which in light of the rest of his views could not be taken as implying ontological realism.)

Kirkham, R.L. (1992) *Theories of Truth: A Critical Introduction*, Cambridge, MA: MIT Press.(An introduction to theories of truth. Chapter 4 expands on the issues in this entry.)

McTaggart, J. McT.E. (1921) *The Nature of Existence*, vol. 1. Cambridge: Cambridge University Press.(Pages 10-37 present the author’s correspondence theory in the context of his ontological anti-realism.)

Pitcher, G. (ed.) (1964) *Truth*, Englewood Cliffs, NJ: Prentice Hall.(The only anthology of theories of truth in print. Somewhat dated, but still valuable.)

Plato (*c.*366-360 BC) *Sophist*, in E. Hamilton and H. Cairns (eds) *Plato: The Collected Dialogues*, Princeton, NJ: Princeton University Press, 1961.(At 262E-263D the Stranger presents the Western world’s first theory of truth.)

Russell, B. (1912) *The Problems of Philosophy*, Oxford: Oxford University Press. (The chapter on ‘Truth’ is the prototypical modern correspondence-as-congruence theory.)

Sellars, W. (1963) *Science, Perception and Reality*, New York: Humanities Press. (Another modern example of an ontological anti-realist endorsing a correspondence theory.)

Truth, deflationary theories of

So-called deflationary theories of truth, of which the best known are the redundancy, performative and prosentential theories, are really theories of truth ascriptions. *This is because they are not theories of what truth* is*; rather, they are theories of what we are saying when we make utterances like '"Routledge editors are fine folks" is true'. The surface grammar of such utterances suggests that we use them to predicate a property, truth, of sentences or propositions; but the several deflationary theories all deny this. Indeed, they all endorse the Deflationary Thesis that there is no such property as truth and thus there is no need for, or sense to, a theory of truth distinct from a theory of truth ascriptions. Thus, for deflationists, the classical theories of truth, such as correspondence, coherence and pragmatic, are not wrong. They are something worse: they are wrong-headed from the start, for they are attempting to analyse something which simply* is not there.

1 Problems with the Deflationary Thesis

What *are* we saying when we make utterances that *appear* to be ascribing truth to some truth bearer? According to the redundancy theory, whose invention is usually credited to F.P. Ramsey, we are saying nothing more or less than what is said in the statement to which we appear to be ascribing truth: '"Routledge editors are fine folks" is true' is synonymous with 'Routledge editors are fine folks' (see Ramsey, F.P. §3). The redundancy theory, then, endorses the view that anything we can say or do with the predicate 'is true' can be said or done perfectly well without that predicate. The latter contention, which I shall call the Gratuity Thesis, is shared by the performative and prosentential theories of truth ascriptions discussed below, although these offer different accounts of what we are saying or doing when we ascribe truth. By itself, the Gratuity Thesis would be of importance only to linguists. Philosophical interest arises because redundancy theorists, and their performative and prosentential counterparts, make a breathtaking and rather cloudy inferential leap from the Gratuity Thesis to the Deflationary Thesis.

The description of the Deflationary Thesis given above seems to presuppose Platonism: that there are universal properties, and most predicates name properties. We can make it neutral with regard to Platonist-Nominalist debates by rewriting it thus: '"is true" is not a genuine predicate'.

It would be absurd to suppose that most predicates are not genuine predicates, so it would be equally absurd to take as a methodological principle that we should assume a predicate is not genuine until it is proven to be. Strangely, however, deflationists have adopted that principle, and anti-deflationists have let them get away with it. Indeed, it was not until the late 1970s and 1980s that the slow decline of ordinary language philosophy brought either contender to an awareness that linguistic analysis alone cannot settle the question (see Ordinary language philosophy §1). If indeed we do not need to postulate a property of truth (whether Platonically or Nominalistically construed) to explain what we are doing with our truth-ascribing utterances, then that is a point in favour of the Deflationary Thesis. But it is not decisive, for there may well be non-linguistic programmes which cannot be carried out without supposing that truth is some kind of property. Anti-deflationists have suggested, for example, that without such a postulation, we cannot explain why our scientific theories are so successful. To this, deflationists have responded that there is really nothing here that needs explanation: we have the theories we do just *because* they are successful. But this will not do. The fact that we would not have automobiles if they did not succeed in getting us from place to place does not preclude the need for an explanation of why they work, which would presumably be given in terms of the properties of automobiles and the forces at work in them.

Explaining the success of science is just a special case of the traditional epistemological programme of discovering the correct theory of justification. The task of the latter project is to refute scepticism by finding some property, called a 'mark' of truth, which correlates, perhaps imperfectly, with truth, and whose possession or non-possession by a given proposition is reasonably easy to detect. Coherence theories of justification (not truth) propose 'coherence with other propositions' as the mark. Foundationalists propose that it is 'self-evident (perhaps defeasibly), or inferred from self-evident premises'. One cannot judge whether a proposed mark does indeed correlate with truth unless one has some notion of what truth *is,* so this programme must postulate that truth is some kind of property. Those few deflationists who have considered this proposal respond by insisting, in effect, that we are entitled just to beg the question against scepticism, and hence there is no need for traditional epistemology in the first place. Begging the question against scepticism has been popular in contemporary

philosophy, but anyone tempted to think that this is philosophically respectable should consider the following analogy: to prove the soundness and completeness theorems for classical logic is, in effect, to prove that the results of the logic's rules of inference correlate (perfectly) with truth (as defined in the logic's model theory). How seriously would we take one who claimed that such proofs (and hence the model theory itself) are unneeded because we are entitled to assume that the inference rules of classical logic are correct? (And how would 'correct' as used here be defined without postulating a property of truth?) Even mathematical intuitionism does not want to say that there is no such thing as truth. It simply defines truth in terms of justification, that is, proof.

2 Problems with the Gratuity Thesis

The intense interest in the Deflationary Thesis in the later decades of the twentieth century should not give the impression that the Gratuity Thesis is itself beyond doubt. All existing theories of truth ascriptions face problems. In particular, at least two of the difficulties confronting the redundancy theory are worth mentioning. First, what are we to make of blind truth ascriptions such as 'What the editor says is true', which do not explicitly identify the proposition to which truth is being ascribed? Given that the speaker (and for that matter the audience) might not even know what proposition the editor has asserted, the idea that this truth ascription is just an alternative way of asserting that same proposition loses some plausibility. The two best redundancy theories, those of C.J.F. Williams (1976) and Paul Horwich (1990), handle this problem quite cleverly. But neither of them deals adequately with the fact that, as Horwich concedes, the redundancy theory cannot actually be stated. It can be *described* as the conjunction of all clauses of the form: '"*p* is true" is synonymous with "*p*"'. Obviously, one cannot state an infinitely long conjunction. Nor can one simply append the quantifier 'For all *p*' to the front of this schema. Such a quantifier cannot be objectual, on the one hand, because the first appearance of the '*p*' in the schema stands in for a noun, while the second appearance stands in for a grammatically complete clause. On the other hand, such a quantifier cannot be substitutional, for then the redundancy theory would be saying 'all substitution instances of the schema are true' and thus the theory would be parasitic on some other antecedent notion of truth.

P. F. Strawson's (1950) performative theory of truth (ascriptions) denies that apparent truth ascriptions say *anything* in any familiar sense of 'say' (see Strawson, P. §4). Such utterances are more *doings* than *sayings*. They are gestures of agreement, much like nodding one's head. Hence, to utter '"Routledge editors are fine folks" is true' is to signal agreement with the notion that Routledge editors are fine folks. Utterances which do something rather than say something are called illocutionary or performative utterances (see Performatives).

Strawson concedes that certain conditions in the non-linguistic world must obtain before one may properly signal one's agreement, where the force of 'properly' is not 'appropriate from the standpoint of honesty or accuracy' but 'appropriate from the standpoint of correct usage'. For example, one ought not utter the words '"The dam has collapsed" is true' unless one believes it is a fact that the dam has collapsed. But if uttering an apparent truth ascription is just to signal agreement, why would the facts of the matter have any relevance? If 'is true' in no way asserts that these conditions are fulfilled, then how and why would these conditions be conditions for uttering 'is true'? One can after all *signal* agreement, dishonestly, even when one does not in fact agree. A second reason for doubting that the performative theory can establish the Gratuity Thesis is that, as H. Price (1988) has pointed out, the theory makes it a mystery why 'is true' is only applicable to indicative sentences. Why can we not use it to endorse the appropriateness of some question (for example, '"Is snow white?" is true') just as we can use it to endorse someone's assertion? And why can we not say '"Shut the door!" is true' to express our endorsement of the importance of obeying this command? It seems that any answer to these questions would have to concede that 'is true', 'is a good question', and 'ought to be obeyed' *mean* different things and this would imply that each of them *says* something.

Besides pronouns, English also contains pro-verbs, such as the 'did' in 'Mary ran quickly, so Bill did too', and pro-adjectives such as 'such' in 'The happy man was no longer such'; and pro-adverbs such as the 'so' in 'She twitched violently and, while so twitching, expired'. D. Grover, J. Camp and N. Belnap (1975), inventors of the prosentential theory of truth (ascriptions), suggest that there are also prosentences. 'So' is used this way in 'I do not believe Rachel is sick, but if so, she should stay home'. More to the immediate point, they contend that the phrases 'it is true' and 'that is true', despite their subject-predicate structure, are really one-word prosentences. And they claim that every locution in which 'is true' appears can be replaced by a locution that uses one of these prosentences. Thus, they endorse a modified version of the Gratuity Thesis: 'is true', as a separable predicate, is a

gratuitous feature of the language. For example, ‘Everything John says is true’ is ‘For every proposition, if John says that *it is true*, then *it is true*’, where neither the ‘it’ nor the ‘is true’ have any separate meaning whatsoever.

The most serious objection to the prosentential theory concerns modified uses of ‘is true’ such as ‘is not true’ or ‘will be true’. The prosentential theory postulates that the deep structure of English contains a number of sentential operators including ‘It-will-be-true-that’ and ‘It-is-not-true-that’, and that all modified uses of ‘is true’ involve one or the other of the two prosentences within the scope of one of these operators. Thus, the deep structure of ‘Everything John says will be true’ is ‘For all propositions, if John says that it is true, then it-will-be-true-that it is true’, where, again, the ‘it is true’ is a prosentence. To analyse apparent predications of falsehood, they postulate the operator ‘It-is-false-that’. Thus, the deep structure of ‘That is false’ is ‘It-is-false-that that is true’. But, then, how do we account for the ‘true’ that appears in the sentential operators? The operators cannot themselves be given a prosentential analysis: if they could, there would have been no need to postulate them in the first place. The use of hyphens to connect the words of the operators tempts us to think of them each as a single word, in which case the ‘true’ in them has no independent meaning. But it is not easy to see how we could explain the meanings of these one-word operators without using a substantive notion of truth, and even if we could, our explanations of such one-word operators would miss out entirely on the common element in the meanings of, say, ‘it-will-be-true-that’ and ‘it-was-true-that’.

See also: Doōgen; Justification, epistemic; Meaning and truth; Truth, coherence theory of; Truth, correspondence theory of; Truth, pragmatic theory of

RICHARD L. KIRKHAM

References and further reading

Field, H. (1986) ‘The Deflationary Conception of Truth’, in G. MacDonald and C. Wright (eds) *Fact, Science and Morality*, Oxford: Blackwell.(An important attack on the Deflationary Thesis. Not for beginners.)

Geyer, D.L. (1917) ‘The Relation of Truth to Tests’, *Journal of Philosophy, Psychology and Scientific Methods* 13: 626-33.(Contains the earliest statement of what today would be called a redundancy theory.)

Grover, D.L., Camp, J.L., Jr and Belnap, N.D., Jr (1975) ‘A Prosentential Theory of Truth’, *Philosophical Studies* 27: 73-125.(Presentation of a prosentential theory of truth by its inventors.)

Horwich, P. (1990) *Truth*, Oxford: Blackwell.(Along with Williams - see below - one of the most thorough and best-argued defences of the redundancy theory. Horwich, however, rejects the Gratuity Thesis and endorses a milder than usual version of the Deflationary Thesis.)

Kirkham, R.L. (1992) *Theories of Truth: A Critical Introduction*, Cambridge, MA: MIT Press.(An introduction to theories of truth. Chapter 10 expands on the issues in this entry.)

Körner, S. (1955) ‘Truth as a Predicate’, *Analysis* 15: 106-9.(A potent, early attack on what is now called deflationism before the latter became popular.)

Price, H. (1988) *Facts and the Function of Truth*, Oxford: Blackwell.(Presents a Darwinian theory of truth ascriptions.)

Ramsey, F.P. (1927) ‘Facts and Propositions’, *Proceedings of the Aristotelian Society* supplement to 7: 153-70. (The first statement of a redundancy theory to have a widespread impact.)

Strawson, P.F. (1950) ‘Truth’, *Proceedings of the Aristotelian Society* supplement to 24: 129-56.(The classic statement of the performative theory of truth ascriptions.)

Williams, C.J.F. (1976) *What is Truth?*, Cambridge: Cambridge University Press. (Along with Horwich - see above - one of the most thorough and best-argued defences of what could fairly be called a redundancy theory.)

Truth, pragmatic theory of

Two distinctly different kinds of theories parade under the banner of the 'pragmatic theory of truth'. First, there is the consensus theory of C.S. Peirce, according to which a true proposition is one which would be endorsed unanimously by all persons who had had sufficient relevant experiences to judge it. Second, there is the instrumentalist theory associated with William James, John Dewey, and F.C.S. Schiller, according to which a proposition counts as true if and only if behaviour based on a belief in the proposition leads, in the long run and all things considered, to beneficial results for the believers. (Peirce renamed his theory 'pragmaticism' when his original term 'pragmatism' was appropriated by the instrumentalists.) Unless they are married to some form of ontological anti-realism, which they usually are, both theories imply that the facts of the matter are not relevant to the truth-value of the proposition.

1 Pragmaticism

C.S. Peirce believed that any two minds investigating a given question would tend eventually to arrive at the same answer, even if they used different methods and different pools of evidence: ' Let any human being have enough information and exert enough thought upon any question, and the result will be that he will arrive at a certain definite conclusion, which is the same that any other mind will reach' (Peirce 1931-58 (7): 319). Moreover, this one answer that all would reach is, by definition, the true answer: ' The opinion which is fated to be ultimately agreed to by all who investigate is what we mean by truth' (Peirce 1931-58 (5): 407). Indeed, in principle, consensus embodies the truth *no matter what method was used to bring about the consensus*. ' If a general belief…can in any way be produced, though it be by the faggot and the rack, to talk of error in such belief is utterly absurd' (Peirce 1931-58 (8): 16). Although Peirce thought that in the long run the only method which could produce and sustain a consensus agreement is what he called 'abduction' (what is now called 'inference to the best explanation'), it is important to remember that he did not think propositions which would be universally accepted are true because they were arrived at by abduction; rather, he thought they are true just because they would be universally accepted (see Peirce, C.S. §3).

Whence Peirce's confidence that investigators would move towards a common conclusion? Ultimately, our evidence takes the form of perceptions, and these perceptions are controlled by a single fixed reality which is public to all. Since there is just one objective reality and it is driving all of us to beliefs that accurately reflect it, we are driven to agree with one another. So, in the long run, the only propositions with which everyone would agree are those that accurately reflect reality. Hence, 'is true' is equivalent to 'accurately reflects reality'. It might seem odd, then, that Peirce turns his attention away from this equivalence and focuses instead on what would otherwise seem to be an incidental equivalence between 'is true' and 'would eventually be agreed to by everyone with sufficient relevant experiences'. But for Peirce it is the former relation which is the trivial one because reality, he thought, is just a construct of the community of human minds. Specifically, what is real is just whatever we would come to agree is real: ' the real is the idea in which the community ultimately settles down' (Peirce 1931-58 (6): 610) and ' everything, therefore which will be thought to exist in the final opinion is real, and nothing else' (Peirce 1931-58 (8): 12). Peirce called this his 'social theory of reality'.

But notice now how this ontological doctrine undermines Peirce's own explanation for why all who investigate a given question would ultimately come to agreement: on a realist ontology, the notion of reality controlling our perceptions is based on common sense; but on the social theory of reality, it is an idea in the minds of those who have already reached the final opinion which is causing those who have not reached it to have certain perceptions. Indeed, matters are even stranger than this; for the very perceptions that caused those who have reached the final conclusion to reach it were forced on them, *in a reverse-chronological direction*, by the final conclusion which, at the time they had the perceptions, they had not reached. Therefore, some of the perceptions you and I are having right now are forced on us by an idea which, if we have it at all, we will only have at some future time. Peirce himself was aware of this rather fantastic implication of his views and attempted to defend it:

At first sight it seems no doubt a paradoxical statement that, ' The object of final belief which exists only in consequence of the belief, should itself produce the belief';…there is:

> nothing extraordinary…in saying that the existence of external realities depends upon the fact, that opinion will

> finally settle in the belief in them. And yet that these realities existed before the belief took rise, and were even the cause of that belief, just as the force of gravity is the cause of the falling of the inkstand - although the force of gravity consists merely in the fact that the inkstand and other objects will fall.
>
> (Peirce 1931-58 (7): 340-4)

But this will not do. On the analysis of causation which Peirce is assuming, the relation between gravity and the tendency of things to fall is one of identity, not of mutual causation. Moreover, even if we allow that gravity is in some sense a consequence of the fact that things tend to fall, neither gravity nor the tendency are events in time. Hence, to assert that the former is both a cause and a consequence of the latter is not to assert the possibility of reverse-chronological causation. But coming to believe the final conclusion, and the occurrence of the perceptions that bring about that belief, are both events in time, and they come at different times. Hence, to assert that the chronologically earlier of these is caused by the chronologically later is to assert something not at all analogous to any causal relations involving the inkstand.

The problems in Peirce’s account go even deeper. Although Peirce sometimes speaks as if the final conclusion is fated or destined, on those occasions on which he self-consciously considers whether this conclusion will ever actually be reached, he is much more cautious: ‘ We cannot be quite sure the community will ever settle down to an unalterable conclusion upon any given question…nor can we rationally presume any overwhelming *consensus* of opinion will be reached upon every question’ (Peirce 1931-58 (6): 610).

> I do not say that it is infallibly true that there is any belief to which a person would come if he were to carry his inquiries far enough. I only say that that alone is what I call Truth. I cannot infallibly know that there *is* any truth.
>
> (Peirce 1966: 398)

So the causal action of the final conclusion on our present actual perceptions is not only reverse-chronological, it is also action from within a hypothetical domain to the actual domain.

2 Instrumentalism

William James always claimed to accept the definition of truth embodied in correspondence theories of truth, namely that a true belief or statement is one that ‘agrees with reality’ (see James, W. §5). But, for James, the reality to which true ideas must agree is mind-dependent: ‘ By "reality" humanism [one of James’s names for his philosophy] means nothing more than the other conceptual or perceptual experiences with which a given present experience may find itself in point of fact mixed up’. And, ‘ we are not required to seek [truth] in a relation of experience as such to anything beyond itself’. And, ‘ reality is an accumulation of our own intellectual inventions’. These inventions include ‘ the notions of one Time and of one Space as single continuous receptacles; the distinction between thoughts and things, matter and mind; between permanent subjects and changing attributes; the conception of classes with sub-classes within them; the separation of fortuitous from regularly caused connexions’. Unlike Kant, however, James does not think that these constructs are built into our minds. Rather, these constructs are inventions of our ancestors. They *made* the world this way, by so conceiving of it. Why did they choose to structure the world with these features and not some other features? James’s answer is that they found it more *useful* to organize the world in this manner. The last quotation continues:

> surely all these were once definite conquests made… by our ancestors in their attempts to get the chaos of their crude individual experiences into a more shareable and manageable shape. They proved of such sovereign use as *denkmittel* [instruments of thought] that they are now a part of the very structure of our mind.
>
> (James 1909: 42)

James also wants to give the word ‘agree’, in the phrase ‘agrees with reality’, a different sense from the typical correspondence theorist. Given that reality is just useful mental constructs of the collection of past and present minds, a belief agrees with reality by proving useful to those who believe it. The examples James offers suggest that useful beliefs are those which: (1) enable us to manipulate the objects of the world; (2) allow us to communicate successfully with our fellows; (3) provide good explanations for other occurrences; and (4) lead to accurate predictions.

It should be kept in mind, if for no other reason than to forestall overly facile counterexamples, that James identifies truth with beliefs that are useful over the long run and all things considered:

> ‘"*The true*",*to put it very briefly, is only the expedient in the way of our thinking, just as "the right" is only the expedient in the way of our behaving.* Expedient in almost any fashion; and expedient in the long run and on the whole of course; for what meets expediently all the experience in sight will not necessarily meet all farther experiences equally satisfactorily’
>
> (James 1907: 106)

Still, opponents of instrumentalism would insist, it might be useful throughout a person’s life for them to believe that they are better at their job than anyone else (because, for example, the increased confidence it gives them pays huge dividends.) And this can be the case even if they are not in fact better at their job than anyone else. James, however, denied that there can be any cases in which the truth and the facts are disjoint, and the reason for his denial lies in his ontology: if the facts are themselves just mental constructs which have proved useful, then there cannot be a case of a useful belief that does not agree with the facts.

So, both Peirce and James can happily accept what have been called ‘T-sentences’ - sentences of the form ‘"p is true" if and only if p’. But for both of them, this coordination of truth and reality is itself an incidental side effect which distracts from, rather than reveals, the essential nature of truth.

See also: Meaning and truth; Pragmatism; Realism and antirealism; Schiller, F.C.S.; Truth, coherence theory of; Truth, correspondence theory of; Truth, deflationary theories of

RICHARD L. KIRKHAM

References and further reading

James, W. (1907) *Pragmatism*, Cambridge, MA: Harvard University Press, 1975.(A superbly produced critical edition of the author’s most famous work, in which he attempts to provide a survey of his philosophical views.)

James, W. (1909) *The Meaning of Truth*, Cambridge, MA: Harvard University Press, 1975.(A collection of essays on truth, mainly written in the last few years before the author’s death, in an excellent critical edition.)

Kirkham, R.L. (1992) *Theories of Truth: A Critical Introduction*, Cambridge, MA: MIT Press.(An introduction to theories of truth. Sects 3.2-3.4 expand on the issues in this entry.)

Peirce, C.S. (1931-58) *Collected Papers of Charles Sanders Peirce*, Cambridge, MA: Harvard University Press, 8 vols, C. Hartshorne and P. Weiss (eds) vols 1-6, A.W. Burks (ed.) vols 7-8.(The author’s views on truth are scattered throughout his works, so the reader will have to look under ‘truth’ in the index of this collection. References in this entry to this work have been by volume and section.)

Peirce, C.S. (1966) ‘Letters to Lady Welby’, in P.P. Wiener (ed.) *Charles S. Peirce: Selected Writings*, New York: Dover.(An interesting selection of Peirce’s writings.)

Truthfulness

Humans are the only species capable of speech and thus of lies. Choices regarding truthfulness and deceit are woven into all that they say and do. From childhood on, everyone knows the experience of being deceived and of deceiving others, of doubting someone's word and of being thought a liar. Throughout life, no moral choice is more common than that of whether to speak truthfully, equivocate, or lie - whether to flatter, get out of trouble, retaliate, or gain some advantage.

All societies, as well as all major moral, religious and legal traditions have condemned forms of deceit such as bearing false witness; but many have also held that deceit can be excusable or even mandated under certain circumstances, as, for instance, to deflect enemies in war or criminals bent on doing violence to innocent victims. Opinions diverge about such cases, however, as well as about many common choices about truthfulness and deceit. How open should spouses be to one another about adultery, for example, or physicians to dying patients? These are quandaries familiar since antiquity. Others, such as those involving the backdating of computerized documents, false claims on résumés in applying for work, or misrepresenting one's HIV-positive status to sexual partners, present themselves in new garb.

Hard choices involving truthfulness and lying inevitably raise certain underlying questions. How should truthfulness be defined? Is lying ever morally justified, and if so under what conditions? How should one deal with borderline cases between truthfulness and clear-cut falsehood, and between more and less egregious forms of deceit? And how do attitudes towards truthfulness relate to personal integrity and character? The rich philosophical debate of these issues has focused on issues of definition, justification, and line-drawing, and on their relevance to practical moral choice.

1 Truth, truthfulness, falsity and deceit

By what criteria can we distinguish between truthfulness and deceit? Should all deceptive messages, whether intended to deceive or not, whether verbal or nonverbal, count as lies so long as they end up misleading recipients? Such an all-encompassing view is unpersuasive: it is hard to claim that erroneous weather forecasts, for example, or promises people make in good faith but are later unable to honour, constitute lies.

If one limits the category of lies to intentionally false speech, however, problems arise about how to characterize the use of metaphor and other figurative speech, rhetorical flourishes, and works of fiction (see Metaphor; Fictional entities; Art and truth §2). Some thinkers have been known to hold that the very purpose of all art, including fiction, is to deceive: they may defend the practice, as Oscar Wilde does in 'The Decay of Lying' (1892), suggesting that, 'the final revelation is that Lying, the telling of beautiful untrue things, is the proper aim of Art'; or deplore it, as does Socrates in Plato's *Republic*, when he argues that poets mislead their audiences in speaking, for instance, of the gods as undignified or immoral. Many people, rejecting the view that fiction must involve lying, prefer to limit the category of lies to spoken communication intended to mislead. They take what Samuel Coleridge called 'the willing suspension of disbelief' (1817: 6) to operate in such a way that fiction does not count as deceitful, and may regard persons who cannot distinguish between lies and fiction as obtuse. Thus John Stuart Mill pointed to Jeremy Bentham's aphorism 'All poetry is misrepresentation' as revealing his literalness and his impoverished understanding of human nature (1838: 50-1).

The failure to sort out such disagreements about the scope of definitions generates needless confusion in debates about truthfulness and deceit. In turn, this failure risks perpetuating a conceptual muddle with regard to the crucial differences between the *moral* domain of intended truthfulness and deception and the *epistemological* domain of truth and falsity. The moral question of whether you are lying or not is not *settled* by establishing the truth or falsity of what you say. In order to settle this question, we must know whether you intend your statement to mislead those to whom you address yourself. The two domains often overlap; but truth and truthfulness are not identical, any more than falsehood and falsity (see Aquinas, *Summa theologiae*). Confusing the domains of 'truth' and 'truthfulness' has especially problematic consequences for persons whose scepticism about all epistemological claims concerning truth bleeds into their attitudes with respect to the moral questions about deceit and truthfulness. They may leap from doubting that we can ever know the full truth about anything to doubting that there are differences between aiming to lie or to speak truthfully, and in turn to regarding all communication

as equally unreliable. Iris Murdoch, in *Metaphysics as a Guide to Morals* (1992: 198), discusses how structuralists make this leap with regard to history, thus blotting out 'the conception of seeking carefully for some truthful conception of the past'. 'Fabulation' means providing accounts in the form of fables, as did Ovid, or, more generally, to concoct and fabricate. 'Confabulation' has come into common parlance as a psychiatric term so recently that it has yet to be recorded in a major dictionary. The term, which used to carry the meaning of persons simply coming together to talk or chat, is now used to refer to the stories told by brain-damaged persons suffering from Alzheimer's disease and a variety of other psychiatric and neurological conditions. They may spin false tales about their lives with great aplomb and in utter confidence that they are correct. They cannot, therefore, be thought of as engaging in lying or any form of deceit; at the same time, because their statements depart so clearly from the truth, it is difficult to speak of truthfulness in characterizing their stories.

Such cases show that the moral dimensions of choices concerning truthfulness and deceit are not exhausted by referring to the intentions of those who make statements. There is a large category of statements where deceit is not intended but where truthful communication is far from being achieved. In exploring this category, it is important to take into account all that can help to distort communication quite apart from an intention to deceive. When people convey false information in the belief that it is true, they may be tired, mistaken, uninformed, inarticulate, intoxicated, or duped by others; but so long as they do not intend to mislead anyone, they are not acting in a manner that is in any way deceitful. Their statements may be false, but they have not knowingly uttered falsehoods. If the information is conveyed through intermediaries, as through gossip or via the media, further distortion from such causes is likely to ensue. At the receiving end of such information, likewise, similar factors and others such as deafness may operate so that people end up deceived through no fault of the person who originated the message or those who passed it along.

Allowing for these qualifications, a truthful person is someone who aims consistently to speak so as not to mislead others. A scrupulously truthful person is one who takes special care to try to counteract factors known to skew or distort communication. The trait of 'truthfulness', defined as the disposition to tell the truth, is widely regarded as a virtue. Truthfulness can take either the baseline form of avoiding all lies or the wider form of being as forthcoming and sincere as possible at all times. The fourth Buddhist Precept of Right Action spans this spectrum, in enjoining followers of Buddha 'not to lie but to practice sincerity and honesty'. And Michel de Montaigne, who took a firm stance against lying and all forms of deceit, put the difference as follows: 'We must not always say everything, for that would be folly; but what we say must be what we think; otherwise it is wickedness' (1533-92: 401).

The disposition to be truthful is generally held to be both admirable in its own right and central to the larger cluster of traits called honesty. 'Honesty' characterizes a person of integrity and trustworthiness who avoids not only lies and intentional deception, but also theft, cheating, plagiarism and other forms of duplicity and betrayal. Conversely, 'dishonesty' refers to the disposition to lie, deceive, steal, cheat and defraud more generally. 'Veracity' is often used synonymously with 'truthfulness', in the above sense, but carries an additional stress on the accuracy of what is stated.

Just as false statements need not be deceitful, moreover, so true statements can sometimes be meant to deceive. An isolated true statement told by someone known to be a habitual liar can be intended to mislead listeners, and succeed in doing so, as much as lies by persons thought honest. Persons adept at producing 'information overload', moreover, as in lengthy, confusing sales agreements or insurance documents, can aim to deceive others without actual lies.

Withholding part of the information needed by interlocutors, finally, can be as deceptive as any lie. Sellers of medications, toys or automobiles who purposely omit information regarding dangerous defects or other risks attendant on using their products, engage in such deceit. So do investigators who enrol patients in studies of experimental drugs without asking for their fully informed consent. Persons wishing to engage in such communications while at the same time maintaining that they adhere strictly to the truth have often had recourse to the concept of a 'mental reservation'. It involves speaking only a partial but highly misleading truth with the intent to deceive, while adding in one's mind the missing words that would render the statement non-deceptive: as, for instance, when a thief responds to an inquiry about a theft the previous week by claiming that he had not stolen anything, adding silently 'today'.

2 Justification

From the Bible's Ten Commandments to the Buddhist Five Precepts for Right Action, from the Five Jaina Great Vows to the Hindu prescriptions in the *Bhagavadgītā* (see Bok 1996), from the maxims of Confucius to the dictates of the Roman Stoics, false speech, along with the resort to violence, has been rejected consistently. However much these traditions differ over questions of religious belief, asceticism or sexual conduct, they speak in unison in condemning violence and lies, the two ways by which human beings deliberately bring about injury to one another. Because no society or human relationship could survive without at least a degree of truthfulness in communication, all communities have stressed truthfulness as a trait to be fostered, at least with respect to intimates, friends and colleagues. Even the devils themselves, as Samuel Johnson said, quoting Sir Thomas Brown, do not lie to one another: truth is as necessary to their society as to all others (see Hindu philosophy; Jaina philosophy; Confucian philosophy, Chinese; Stoicism).

Even so, the world's traditions rest their preference for truthfulness and their condemnations of deceit on widely differing forms of justification. They may claim a variety of religious reasons for condemning lying, argue that it goes against the natural order, hold that it jars with our 'moral sense', or invoke other reasons. Further differences arise within each tradition with respect to whether or not exceptions are permitted under certain circumstances, and, if so, which ones and on what grounds. Is it excusable to lie, for example, where truthfulness might endanger innocent persons, precipitate violence, or cause despair? Or possibly even justifiable?

While a degree of truthfulness has been recognized as necessary everywhere, the pull of lying has been ever-present. It offers the simplest and most tempting way of trying to generate false beliefs among listeners and thus of achieving power over them or eluding their power over oneself. But because truthfulness is uniformly seen as the norm and lies as morally problematic, the burden of proof falls on those who would defend particular categories of departures from truthfulness.

Two positions have been taken within most traditions on this score. The first, absolutist, position was held by, among others, Augustine and Immanuel Kant, who took there to be no sufficient justification for lying. According to Augustine, lies can be ranked according to their severity, with lies about God the most severe and lies by a virgin to guard against violation the most excusable. Such excusable lies, while not mortal sins unlike the former, are nevertheless forbidden by God and should be avoided. Kant (1788) also rejected all lies as unjustifiable and recognized, under the heading of 'casuistical questions', cases of seemingly more excusable lies; but he refused to rank them as Augustine had done. He insisted, instead, that truthfulness is an unconditional duty which holds in all circumstances. All persons who lie, he claimed, thereby repudiate their own human dignity and contribute to undermining the precarious trust on which human society is based.

The second position is shared by many thinkers who set forth certain types of lies as excusable or even justifiable. Some of them argue that lies to certain kinds of people - children, the dying, the mentally ill, or enemies in war - should not count as lies properly speaking, and are justifiable whenever told to deceive these persons for their own good or in self-defence. There is no need for complex moral reasoning in such cases, these writers claim. Thus Hugo Grotius (1625) held that falsehoods told to children or to the insane show no disrespect for their liberty of judgment, since they have no such liberty in the first place. Others argue that lies *by* certain persons are justified in the exercise of their professions, as in the law or policing or medicine or, as urged by Machiavelli in *Il principe (The Prince)* (1532), governing (see Machiavelli, N. §2). While they agree that high standards of honesty should be encouraged in most walks of life, these advocates maintain that there are special circumstances and pressures within their own profession that necessitate forgoing candour at times. Important goals such as the interests of one's nation or one's clients, the public welfare, the advancement of knowledge, or the uncovering of corruption, may be at stake, and may outweigh a strict commitment to veracity.

John Stuart Mill argued, on utilitarian grounds, that it is possible so to estimate the consequences of particular lies as to sort out those few lies that would be justifiable (see Utilitarianism; Consequentialism). Henry Sidgwick (1874) likewise concluded on such grounds that certain lies are justifiable - among them those told to invalids where the truth might prove too great a shock to them. W.D. Ross (1930) held that, while there is a *prima facie* duty of fidelity, which includes that of veracity, it can conflict with other *prima facie* duties such as that of 'not injuring others'. It is worth noting, however, that Mill and Sidgwick both took seriously what Kant also emphasized but Machiavelli did not: the central importance of the virtue of veracity for purposes of societal trust.

3 Line-drawing

Problems of line-drawing will always arise with respect to the interlocking distinctions having to do with truth, truthfulness, falsity, and deceit, both when it comes to definitions and forms of justification or excuse. In considering definitions, first of all, just where should the boundaries be between intentional and unintentional deception? Or between jokes understood as such by all listeners and tricks meant to mislead them? Or between mild exaggeration and outright lies? Are so-called white lies even lies in the first place? When do errors due to poor memory shade into confabulation? When does the transmission of false information - say by reporters covering public officials - shade into complicity in deceit? How easy is it even for historians to tell when they are engaged in conscious rewriting of history, let alone their readers? And in evaluating possible justifications for accepting or ruling out lying, by what means might a Kantian best distinguish intentionally and unintentionally deceptive statements? How might Grotius compare what degree of truth is owed to preadolescents? How do Ross and others weigh conflicting *prima facie* duties in marginal cases? And how should utilitarians weigh the gains or losses that will flow from comparatively similar forms of deceit?

From the earliest texts onward, the moral perplexities in trying to sort out and compare forms of deceit have been set forth: as in the account of Adam and Eve and the serpent in Genesis, the kiss of Judas in the New Testament, the false protestations of lifelong truthfulness in the Egyptian *Book of the Dead*, or the trickery of Ulysses in Homer's *Odyssey* and *Iliad*. And thinkers in the great traditions of casuistry and practical moral inquiry - among them Confucians, Stoics and commentators on Christian, Jewish, Buddhist, Hindu and Islamic ethics - have worked to weigh the difficult 'cases of conscience' in which the duty not to lie appears to conflict with such duties as that of keeping promises of secrecy or saving intended victims from murderers.

Drawing lines with respect to truthfulness and deceit often turns out differently, however, depending on whose choices are being considered. Truthfulness is a trait that people tend to prize more in others than regard as fully called for when it comes to their own choices. When they consider themselves as being on the receiving end of other people's lies, they are more suspicious of the underlying motives than when they weigh possible lies of their own. Each of these perspectives on truthfulness and deceit - that of the deceived and that of those considering engaging in deception - is incomplete, biased, and undiscriminating by itself. The perspective of those who see themselves as at the receiving end of lies is, understandably, suspicious in the extreme. As a result, they are more likely to blur the distinction between errors and lies, between the honest and the dishonest, between intended and unintended injuries. By contrast, the perspective of those who wonder whether or not to lie blurs and confuses the harm that lying does not only to those lied to but also to those who do the lying, and, most insidiously and indelibly, to trust. To persons adopting this latter perspective, lies may seem far more excusable or innocuous and the risks of discovery or of slipping into broader practices of deceit less worrisome.

It is only by examining one's choices not only as subject and agent but as recipient, sometimes victim, that a fuller perception of truthfulness and lying becomes possible, and of ways to counteract the biases present in each of the perspectives. These biases operate not only to skew line-drawing when it comes to defining truthfulness and lying, but also with respect to justifying or excusing particular choices. In this way, those considering lies of their own may give themselves far more latitude than they grant to others in declaring certain forms of duplicity such as white lies nondeceptive; and greater latitude, too, when it comes to excusing or justifying what they acknowledge as lies.

Both Kant, who characterized lying as a repudiation of one's own human dignity, and Mill, who wrote in *Utilitarianism* (1861) of the need to cultivate in oneself a 'sensitive feeling on the subject of veracity', addressed, albeit from different points of view, the importance of taking truthfulness seriously in one's own life. The role that one assigns to truthfulness, they both thought, should be central in considering what kind of person one wants to be - how one wishes to treat not only other people, but oneself. Everyone can make mistakes when it comes to truthfulness and deception and be swayed by poorly understood forces and temptations; it is another matter altogether to choose to be someone who deals with others through deceit. Such a choice is made easier by neglecting to take these forces and temptations into adequate account. Though Mill would have insisted on certain rare exceptions to what he characterized as the 'sacred rule' of veracity, he might otherwise have concurred in Kant's linking internal and external truthfulness, in the last of his books published in his lifetime, *Anthropology from a Pragmatic Point of View* (1798: 207): 'Briefly, as the highest maxim, uninhibited internal truthfulness

toward oneself, as well as in the behavior toward everyone else, is the only proof of a person's having character.'

See also: Self-deception; Self-deception, ethics of; Trust; Virtues and vices

SISSELA BOK

References and further reading

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, trans. Fathers of the English Dominican Province, Westminster, MD: Christian Classics, 1948, IIaIIae.109-113.(Authoritative statement of the Roman Catholic Natural Law position, building on Augustine.)

Arendt, H. (1970) 'Truth and Politics', in P. Laslett, and W.G. Runciman (eds) *Philosophy, Politics, and Society*, New York: Barnes and Noble, 1967.(An analysis of the politics of lying and truthfulness in the context of the Vietnam War.)

Augustine (395) *De mendacio (Of Lying)*, ed. R.J. Deferrari, in *Fathers of the Church: A New Translation*, vol. 16, New York: Catholic University of America Press, 1952, 45-110.(A subtle analysis of issues related to truthfulness and the sin of lying.)

Augustine (420) *Contra mendacium (Against Lying)*, ed. R.J. Deferrari, in *Fathers of the Church: A New Translation*, vol. 16, New York: Catholic University of America Press, 1952, 111-79.(The second of Augustine's treatises on lying and truthfulness.)

Bok, S. (1978) *Lying: Moral Choices in Public and Private Life*, New York: Pantheon.(A contemporary treatment of the ethics of lying and truthfulness and of the traditions of debate on these subjects. Appendix has selections from Augustine, Thomas Aquinas, Grotius, Kant, Sidgwick, Ross and others.)

Bok, S. (1996) *Common Values*, Columbia, MO: University of Missouri Press.(Includes discussion of the Buddhist Five Precepts for Right Action, the Five Jaina Great Vows, and the Hindu prescriptions in the Bhagavadgītā.)

Coleridge, S.T. (1817) *Biographia Literaria*, in J. Engels and W. Jackson Bate (eds) *Biographia Literaria*, vol. 2, Princeton, NJ: Princeton University Press, 1984.(The finest text of philosophical and literary criticism of the romantic period.)

Grotius, H. (1625) *De iure belli ac pacis (On the Law of War and Peace)*, trans. F. Kelsey, Indianapolis, IN: Bobbs-Merrill, 1925, bk III, ch. 1.(Treats of lies in the context of war and peace.)

Kant, I. (1788) 'On a Supposed Right to Lie from Benevolent Motives', in *The Critique of Pure Practical Writings in Moral Philosophy*, trans. L.W. Beck, Chicago, IL: University of Chicago Press, 1949, 346-50.(Kant's much-discussed statement of an absolutist position on lying.)

Kant, I. (1798) *Anthropology from a Pragmatic Point of View*, trans. V.L. Dowdell, Carbondale, IL: Southern Illinois University Press, 1996, 207.(Kant's last published book, containing the summation of decades of teaching.)

Machiavelli, N. (1532) *Il principe*, trans. L. Ricci, *The Prince*, in *The Prince and The Discourses*, New York: Modern Library, 1950.(Machiavelli's advice to princes on politics and morality, with indications regarding when, how, and why to lie most advantageously.)

Mill, J.S. (1838) *Bentham*, in M. Cohen (ed.) *The Philosophy of J.S. Mill*, New York: Modern Library, 1961, 50-1.(Mill's evaluation of Bentham's contributions.)

Mill, J.S. (1861) *Utilitarianism*, in M. Cohen (ed.) *The Philosophy of J.S. Mill*, New York: Modern Library, 1961, 348.(Mill's statement of his position on ethics in general, including his views on lying.)

Montaigne, M. de (1533-92) 'Des Menteurs', trans. D.M. Frame, 'On Liars', in *The Complete Works of Montaigne*, Stanford, CA: Stanford University Press, 1957, bk I, ch. 9.(One of the most profound meditations on lying in the literature.)

Murdoch, I. (1992) *Metaphysics as a Guide to Morals*, London: Chatto & Windus.(Includes discussion of how structuralists, with regard to history, make the leap from doubting that we can ever know the full truth about anything to doubting that there are differences between aiming to lie or to speak truthfully, and in turn to regarding all communication as equally unreliable.)

Plato (*c*.380-367 BC) *Republic*, trans. A. Bloom, New York: Basic Books, 1968, bk III.(Plato's rejection of theatre and poetry that corrupt and mislead.)

Radhakrishnan, S. and Moore, C.A. (eds) (1957) *A Sourcebook in Indian Philosophy*, Princeton, NJ: Princeton University Press, 1957.(Index listings refer to Buddhist, Hindu, and Jain texts and to explorations of truthfulness and deceit.)

Ross, W.D. (1930) *The Right and the Good*, Oxford: Clarendon Press.(Contains Ross' specifications of the duty not to tell lies as a *prima facie* duty of fidelity.)

Sidgwick, H. (1874) 'Classification of Duties - Veracity', in *The Methods of Ethics*, London: Macmillan; 7th edn,

1907.(Sidgwick's uncharacteristically casual discussion of the duty of veracity and of exceptions to this duty.)
Wilde, O. (1892) 'The Decay of Lying', in V. Holland (ed.) *The Complete Works of Oscar Wilde*, New York: Harper & Row, 1989, 970-92.(A defence of lying on aesthetic grounds.)

Tschirnhaus, Ehrenfried Walther von (1651-1708)

The natural philosopher E.W. von Tschirnhaus emphasized bodily and mental health, was a friend of Spinoza and correspondent of Leibniz. He perfected the construction of concave mirrors (used to generate extremely high temperatures) and was probably the first European to produce porcelain. Hoping to make scientific progress more predictable, Tschirnhaus devised a method of inquiry orientated to mathematics and experimentation.

Ehrenfried Walther (Walter) von Tschirnhaus was born in Kieslingswalde, east of Dresden, on the family's estate. After studying medicine and science at the University of Leiden, he went on extended educational travels. In 1682, he was elected the first German member of the Académie Royale des Sciences in Paris for his chemical experiments and innovations in making burning mirrors. Upon returning to Kieslingswalde, Tschirnhaus built a laboratory and tried to unlock the secrets of porcelain production. He died shortly after the decisive experimental breakthrough. Ludwig Böttger, his assistant and co-worker, passed himself off successfully as the sole inventor of European porcelain after Tschirnhaus' death.

The general theme of Tschirnhaus' writings is the achievement of bodily and mental health. Mental health consists in a state of wisdom, tranquillity and happiness. This state results from the striving for self-perfection through scholarly study, intellectual creativity and scientific discovery.

The *Medicina Corporis (Medicine for the body)* (1686) was Tschirnhaus' first book, comprising twelve rules for a physically healthy life. His second and main work was the *Medicina Mentis (Medicine for the mind)* (1687). It contains a methodology for a philosophy of nature based on experimentation and deduction. The oblique but nonetheless obvious references to Spinoza in the *Medicina Mentis* prompted Christian Thomasius to charge Tschirnhaus with heresy and Spinozism, to which the accused replied with *Eilfertiges Bedencken* (1688). Tschirnhaus' later treatises are devoted to pedagogy: the *Gründliche Anleitung zu nützlichen Wissenschafften (The select primer of useful sciences)* (1700), on the scientific education of young people, and *Dreissig Hinweise für einen guten Hofmeister (Thirty suggestions for a good steward)* (1704, published posthumously 1727), on the general education of gentlemen and court officials. He also published a variety of mathematical papers in the *Acta Eruditorum*, founded in Leipzig in 1686. However, the errors these contained did more harm than good to Tschirnhaus' reputation.

Since mathematics is the most successful of all sciences, Tschirnhaus argued, like Descartes, that the scientific investigation of nature ought to follow the example of mathematics and adopt its deductive methods. Although mathematics is abstract, dealing solely with notional entities, it can and should be applied to the study of material entities occurring in nature. Physics is most successful when quantitative. The study of nature ought to start with the classification of evident truths about the object under investigation. Empirical features of the object should be identified by experiments that are to be performed systematically, according to rules. Ideally, a complete list of all relevant data will then be generated, containing the basic notions (*primi conceptus* or definitions) of the object under investigation. Individual definitions yield axioms, combined definitions yield theorems, and the statements so deduced constitute the exhaustive knowledge of the object.

Tschirnhaus has often been misunderstood. Christian Wolff, who claimed to have been influenced by him, took this methodology of natural philosophy as an apology for the so-called geometric method of metaphysics. Wolff held that philosophy can imitate mathematics in style without employing it in content, and aprioristic non-quantitative deductions are permissible without experimental confirmation. Twentieth-century Marxists regarded Tschirnhaus as a spiritual forerunner, whose alleged atheism was illustrated by his Spinozist sympathies and whose putative social awareness was indicated by his interest in practical questions of engineering and manufacturing.

See also: Enlightenment, Continental

MARTIN SCHÖNFELD

List of works

Tschirnhaus, E.W. von (1686) *Medicina Corporis, seu Cogitationes admodum probabiles de conservanda*

Sanitate (Medicine for the body, or, quite plausible thoughts on the conservation of health), Amsterdam: Albertus Magnus & Johannes Rieuwerts Jr; 2nd edn, Leipzig: Fritsch, 1695; 3rd edn, Leipzig: A. Martin, 1733; Dutch trans., Amsterdam, 1687; German trans., Frankfurt & Leipzig, 1688. (A physician's text, containing dietary recommendations and advice for a healthy lifestyle.)

Tschirnhaus, E.W. von (1687) *Medicina Mentis, sive Tentamen genuinae Logicae, in qua disseritur de Methodo detegendi incognitas veritates (Medicine for the mind, or an essay with genuine logic, investigating the method for the discovery of unknown truths)*, Amsterdam: Albertus Magnus & Johannes Rieuwerts Jr; Dutch trans., Amsterdam, 1687; 2nd revised edn, *Medicina Mentis, sive Artis inveniendi Praecepta generalia (Medicine for the mind, or, general precepts for the art of invention)* Leipzig: Fritsch, 1695; repr. Hildesheim: G. Olms, 1964; German trans. J. Haussleiter, in *Acta Historica Leopoldina 1963/1*, Leipzig: Barth; 3rd edn, Leipzig: A. Martin, 1733.(Tschirnhaus' best-known work; introductory reflections on the relation of knowledge and happiness are followed by a detailed methodology for scientific discovery that involves both rational and empirical principles.)

Tschirnhaus, E.W. von (1688) *Eilfertiges Bedencken wieder die Objectiones, so im Menso Martio Schertz und Ernsthafftiger Gedancken über den Tractat Medicinae Mentis enthalten* (Zealous Misgivings about the Objections to the Tract Medicinae Mentis in the March Issue of the Journal Schertz und Ernsthafftiger Gedancken), Halle: Salfelden. (A short pamphlet directed against Christian Thomasius' critique of the *Medicina Mentis*.)

Tschirnhaus, E.W. von (1700) *Gründliche Anleitung zu nützlichen Wissenschafften absonderlich Zu der Mathesi und Physica Wie sie anitzo von den Gelehrtesten abgehandelt werden (The select primer of useful sciences, in particular of mathematics and physics, as they are treated by truly competent researchers)*, Halle; 2nd edn, Frankfurt and Leipzig: Ritscheln, 1708; 3rd edn, 1712; 4th edn, 1729; repr. Stuttgart and Bad Cannstatt: Frommann-Holzboog, 1967.(A brief sequel to the *Medicina Mentis* for students, containing methods for the efficient and systematic acquisition of knowledge.)

Tschirnhaus, E.W. von (1704) *Dreissig Hinweise für einen guten Hofmeister (Thirty suggestions for a good steward)*, in W.B. von Tschirnhaus (ed.) *Getreuer Hofmeister auf Academien und Reisen (The Loyal Steward at Academies and on Travels)*, Hanover: Nicolaus Förster, 1727.(A short pedagogical tract concerning manners and learning.)

References and further reading

Van Peursen, C.A. (1993) 'E.W. von Tschirnhaus and the *Ars Inveniendi*', *Journal of the History of Ideas* 54: 395-410.(A detailed exposition of Tschirnhaus' method for philosophy.)

Winter, E. (1959) 'E.W. v. Tschirnhaus (1651-1708). Ein Leben im Dienste des Akademiegedanken' (E.W. v. Tschirnhaus (1651-1708). A Life Devoted to the Project of an Academy), *Sitzungsberichte der deutschen Akademie der Wissenschaften zu Berlin* 1959/1, Berlin: Akademieverlag.(A concise account of Tschirnhaus' scientific activities.)

Winter, E. (ed.) (1960) *E.W. von Tschirnhaus und die Frühaufklärung in Mittel- und Osteuropa (E.W. von Tschirnhaus and the Early Enlightenment in Central and Eastern Europe)*, Berlin: Akademieverlag.(A collection of rather ideological papers by East German, Czech, Polish and Soviet authors on the political and economic importance of Tschirnhaus' work.)

Winter, E. (1977) 'Der Freund B. Spinozas E.W. v. Tschirnhaus. Die Einheit von Theorie und Praxis' (E.W. v. Tschirnhaus as Friend of B. Spinoza: The Unity of Theory and Practice), *Akademie der Wissenschaften der DDR: Sitzungsbericht G 1977/7*, Berlin: Akademieverlag.(A study of the relation between Tschirnhaus and Spinoza.)

Winter, E. (1979) 'Ein Ketzerspiegel: Ehrenfried Walter von Tschirnhaus (1651-1708)' (Example of a Heretic: Ehrenfried Walter von Tschirnhaus (1651-1708)), *Ketzerschicksale. Christliche Denker aus 9 Jahrhunderten*, Berlin: Union Verlag, 164-83.(A Marxist characterization of Tschirnhaus as a revolutionary heretic.)

Wollgast, S. (1988) *Ehrenfried Walther von Tschirnhaus und die deutsche Frühaufklärung (Ehrenfried Walther von Tschirnhaus and the Early German Enlightenment)*, Berlin: Akademieverlag.(Tschirnhaus' role in the early German Enlightenment, as interpreted through the Marxist categories of historical materialism.)

Wurtz, J.-P. (1988) 'Über einige offene oder strittige, die *Medicina Mentis* von Tschirnhaus betreffene Fragen' (On some Open or Disputed Questions concerning Tschirnhaus' Medicina Mentis), *Studia Leibnitiana* 20: 191-211.(A scholarly analysis of Tschirnhaus' response to Thomasius' charge of Spinozism, the role of experimental confirmation in Tschirnhaus' methodology, and other issues.)

Zaunick, R. (1963) 'Einführung: Ehrenfried Walther von Tschirnhaus in seinem Werden und Wirken' (Introduction: The Development and Influence of Ehrenfried Walther von Tschirnhaus), introduction to E.W.

von Tschirnhaus, *Medicina mentis sive artis inveniendi praecepta generalia*, ed. and German trans. J. Hausleiter, Leipzig: Barth, 5-28.(Probably the most useful survey of Tschirnhaus' philosophical, scientific and technical work.)

Tsong kha pa Blo bzang grags pa (1357-1419)

Tsong kha pa Blo bzang grags pa (Dzongkaba Losang dragba), the founder of the dGa'-ldan-pa (Gandenba) school of Tibetan Buddhism, was born in Tsong-kha, in the extreme northeastern region of Tibet. He is often depicted as a type of reformer, putting great emphasis on moral precepts and interpreting Tantra in a way which would not create any conflict with the traditional Mahāyāna doctrines found in the sūtras *and treatises. He was also an eclectic, drawing upon and synthesizing numerous different currents of Indian Buddhism - for example, he put forth a version of Čast;Prāsaōgika-Mādhyamika which was inextricably bound up with the logical tradition of Dignāga and Dharmakīrti. On the Tibetan side, one of his major philosophical debts was undoubtedly to the gSang-phu (Sangpu) traditions stemming from the highly original thinker Phya pa Chos kyi seng ge (Chaba Chögyi sengge, 1109-69). Finally, his dGa'-ldan-pa school subsequently became the dGe-lugs-pa (Gelukba), a predominantly monastic tradition which in time became the dominant current of Buddhism in Tibet. Tsong kha pa thus had, in addition to his philosophical influence, a long-term impact on the Tibetan political situation, contributing to the transfer of power from the southern provinces to the Lhasa region and laying the groundwork for the peculiarly Tibetan synthesis of religion and political power which was to be embodied in the institution of the Dalai Lamas.*

1 Life and works

The education of Tsong kha pa Blo bzang grags pa (Dzongkaba Losang dragba) was characterized by a series of stays at various monasteries, where he studied the principal philosophical treatises of Indian Buddhism as well as the root texts and commentaries of Tantra. His scholastic knowledge was regularly tested and refined through debate, but equally he engaged in long periods of analytic meditation and in meditative retreats devoted essentially to Tantric practices and rituals. Among the figures who played a role in Tsong kha pa's intellectual development, two are particularly prominent: Red mda' ba gZhon nu blo gros (Rendawa Shönnu lodrö, 1349-1412), who taught him, *inter alia*, Prajñāpāramitāphilosophy, Abhidharma and Mādhyamika; and a certain dBu ma pa (Umaba), who seems to have been, for some time at least, the intermediary for Tsong kha pa to question the bodhisattva Mañjuśrī and thus to come to definite convictions on various matters, including the superiority of Čast;Prāsaōgika-Mādhyamika to all other views. While Red mda' ba is relatively well known to us, and some of his works are still extant, Lama dBu ma pa remains an obscure figure.

Tsong kha pa's *Collected Works* span eighteen Tibetan volumes (from volume *ka* to *tsha*), with about nine volumes (*ga* to *na*) devoted to Tantra, and the other volumes devoted to Mādhyamika, monastic discipline, Prajñāpāramitātopics, rituals, and miscellaneous works and letters. His most famous work is probably the voluminous *Lam rim chen mo (The Great Work on the Stages of the Path*, volume *pa*), a summa of Buddhist thought which was written as an elaborate and loose commentary on Atīśa's *Bodhipathapradīpa (The Lamp for the Path to Enlightenment)*. Equally extremely important is *Drang nges legs bshad snying po (The Essential Correct Explanation of Intentional and Definitive Meaning*, volume *pha*), a later work where we find Tsong kha pa's position on the classification of scriptures as being of definitive (*nītārtha*) or intentional (*neyārtha*) meaning; the work gives his stance on his use of Indian sources and was written largely as a response to Jo-nang-pa (Jonangba) hermeneutics. His major Mādhyamika works are his commentaries on Candrakīrti's *Madhyamakāvatāra (Introduction to the Middle Way*, volume *ma*) and on Nāgārjuna's *Mūlamadhyamakakārikā(The Fundamental Treatise on the Middle Way*, volume *ba*). In Tantric exegesis, besides several commentaries on the *Guhyasamājatantra*, *Cakrasaṃvara*, *Pañcakrama (The Five Steps)*, and so forth, he is well known for his great compendium, the *sNgags rim chen mo (The Great Work on the Stages of Tantra*, volume *ga*). Although Tsong kha pa lectured on logic and epistemology (*tshad ma*), curiously enough he did not write any major works on these subjects; the few texts in his *Collected Works* concerning *tshad ma* are (with perhaps one exception) notes of his teachings taken down by his two closest disciples, rGyal tshab Dar ma rin chen (Gyeltsap Darma rinchen, 1364-1432) and mKhas grub dGe legs dpal bzang po (Kaydrup Gelek belsangbo, 1385-1438). It is, none the less, abundantly clear that his positions on the soteriological import of Dharmakīrti's *Pramāṇavārttika (Commentary on Valid Cognition)*, on the nominalist theory of language and on the importance of Dharmakīrtian logic generally (even for a Mādhyamika), constitute some of the most fundamental elements of his thought (see Dharmakīrti).

2 Elements of Tsong kha pa's Mādhyamika

Tsong kha pa advocated what might be called a ‘minimalist’ Mādhyamika, with no place for the type of Absolute that one finds in the synthesis of Mādhyamika and Buddha-nature doctrines developed by rival Tibetan Mādhyamika schools, notably those of a gZhan-stong-pa (Shendongba) orientation. For him, contrary to gZhan-stong schools, Mādhyamika’s analysis of the inconsistency of phenomena was not designed to prove the existence of an Absolute free of phenomenal qualities, but rather to lead to an understanding of the principle of dependent origination (*pratītyasamutpāda*), the thoroughly relative nature of the phenomena themselves. His *brTen ’grel bstod pa (Praise of Dependent Origination*, volume *kha*), a short text in verse, is often cited as an eloquent statement of this theme.

Undoubtedly, one of the major issues of Mādhyamika philosophy for Tsong kha pa is the question of precisely what and how much Mādhyamikas should deny if they are to avoid reification of entities, yet preserve conventional truth. How one stands on this matter - in particular, how one ‘recognizes the object to be negated’ (*dgag bya ngos ’dzin*) - is argued to have very wide-ranging consequences: it will, *inter alia*, significantly affect the exact formulations of the logical arguments which the Mādhyamika uses, and will even determine the importance that one attaches to analysis and conceptual thought in the quest for spiritual realization. Other important aspects of Tsong kha pa’s Mādhyamika philosophy can be summarized in a series of recurring concepts known as the ‘eight difficult points’. These form the subject matter of a short work, the *dKa’ gnad brgyad kyi zin bris (The Résumé of [Tsong kha pa’s teaching on] the Eight Difficult Points)*. In the opening section he says:

> Concerning the [ontological] bases, there are the following [three points]: (1) the conventional nonacceptance of particulars and of (2) the storehouse consciousness, and (3) the acceptance of external objects. Concerning the path, there are the following [four points]: (4) the nonacceptance of autonomous reasonings as being means for understanding reality and (5) the nonacceptance of self-awareness; (6) the way in which the two obscurations exist; (7) how it is accepted that the Buddha’s disciples and those who become awakened without a Buddha’s help realize that things are without any own-nature. Concerning the result, there is: (8) the way in which the *buddhas* know [conventional] things in their full extent. Thus, there are four accepted theses and four unaccepted theses.
>
> (Volume *ba*-, 1b-2a; translation by T.J.F. Tillemans)

Some of these points are arguably quite faithful to the Indian positions, but others - such as (2), the rejection of the usual Buddhist ideas about latent karmic tendencies being transmitted from life to life via some type of medium like the ‘storehouse consciousness’ in favour of a mere continuum of causally efficient cessations with no medium of transmission whatsoever - are extraordinarily creative developments with fairly little or no grounding in Indian sources (see Buddhism, Yogācāra school of §§5-8). Almost needless to add, they were also very controversial in Tibet. Indeed, Tsong kha pa was savagely mocked by one of his opponents, Go ram pa bSod nams seng ge (Goramba Sönam Sennge, 1429-89), for formulating his Mādhyamika with reliance upon deceptive ‘visions’ of Mañjuśrī, coming to him second-hand via dBu ma pa, rather than upon actual Indian texts.

Although he was a staunch follower of Nāgārjuna and Candrakīrti, Tsong kha pa remained uncompromisingly committed to having philosophical theses and doctrinal positions, and in this respect too he was significantly different from most other Tibetan Mādhyamika philosophers. A singular feature of his approach was that while Čast;Prāsaōgika was the school that he professed, he remained uncomfortable with the usual Indo-Tibetan version of *prasaōga*-method, namely a systematic *reductio ad absurdum* which was required only to refute adversaries by pointing out internal inconsistencies in their views, and did not have to establish true propositions. Repeatedly, Tsong kha pa argued that it was not possible to lead an adversary to a correct view on the basis of arguments which were not supported by some means of valid cognition (*pramāṇa*): as a result, the ‘adversary-established reason’ (*gzhan grags kyi gtan tshig*) could not be established *only* for the adversary; it also had to be, in some way, actually valid and established for the Mādhyamika proponent. Not surprisingly, he was also attacked by critics on this. Go ram pa, in his *lTa ba’i shan ’byed (Differentiation of the Views*, 31b-32a), accused Tsong kha pa of transforming Čast;Prāsaōgika methods into a logic like that of the Čast;Svātantrikas, who maintained that reasoning had to follow the strictures of Dignāga and Dharmakīrti, that is, that the terms had to be established for both parties in the debate.

3 Tsong kha pa on Tantra

Tsong kha pa’s Tantric system was mainly based on the textual cycles of the divinities Guhyasamāja, Saṃvara and Vajrabhairava. Though the majority of his writings were devoted to Guhyasamāja and Saṃvara, Vajrabhairava did have a marked importance for him. Vajrabhairava was Tsong kha pa’s tutelary deity (*yi dam*) from the time of his ordination as a novice monk and represented the fierce aspect of Mañjuśri, who, as we have seen, played such a recurring role in Tsong kha pa’s choices of philosophical orientation. As for the Guhyasamāja cycle, which was without a doubt the main Tantric cycle to which he devoted his attention, Tsong kha pa relied here upon the Indian exegetical system known as the ‘Ārya tradition’, believing (as did most other Tibetan scholars) that the authors of the fundamental treatises of this tradition, namely Nāgārjuna, Āryadeva and Candrakīrti, were identical with the famous Mādhyamika philosophers of the same name. In choosing the Ārya tradition, then, Tsong kha pa seems to have sought to reinforce his basic methodological standpoint that both Tantric and non-Tantric Buddhism had to be understood according to the Čast;Prāsaōgika philosophy of Candrakīrti.

In many respects, however, Tsong kha pa’s Tantric system owes much to syntheses with cycles other than Guhyasamāja. The instructions on the ‘veins’ (*nāḍī*), ‘energy centres’ (*cakra*) and ‘inner heat’ (*caṇḍālī*), all of which play an important role in Tsong kha pa’s views on Tantra, are, notably, virtually absent in the fundamental Indian texts of the Ārya tradition on Guhyasamāja. Rather, we must look to the *Nāro chos drug (Six Teachings of Nāropa)*, a later Indian synthesis of diverse Tantric practices, as having influenced Tsong kha pa in this regard. Indeed, it was the *Pañcakramasaṃgrahaprakāśa (Illumination of the Summary of the Five Steps)*, a short treatise attributed to Nāropa (956-1040) combining the Six Teachings with the ‘Five Steps’ (*pañcakrama*) of the Ārya tradition, which provided Tsong kha pa with the basic ideas of his Tantric system.

See also: Buddhism, Mādhyamika: India and Tibet; Mi bskyod rdo rje; Sa skya Paṇḍita; Tibetan philosophy; Wônch’ūk

TOM J.F. TILLEMANS

List of works

Tsong kha pa (1357-1419) *Collected Works*, Gedan sungrab mi nyam gyunphel Series 79-105, Delhi: Ngawang Gelek Demo, 1975-9, 27 vols.(The volumes are ordered following the letters of the Tibetan alphabet. For an enumeration and brief description of the individual works in this edition, see the section on Tsong kha pa in Yoshimizu 1989.)

Tsong kha pa (1357-1419) *Drang nges legs bshad snying po*, trans. R. Thurman, *Tsong kha pa’s Speech of Gold in the Essence of True Eloquence*, Princeton, NJ: Princeton University Press, 1984.(A translation usable with considerable reservations.)

Tsong kha pa (1357-1419) *Yid dang kun gzhi dka’ ba’i gnas*, trans. G. Sparham, *Ocean of Eloquence: Tsong kha pa’s Commentary on the Yogācāra Doctrine of Mind*, Albany, NY: State University of New York Press, 1993.(An excellent translation of Tsong kha pa’s text on the mental faculty and the storehouse consciousness.)

References and further reading

Go ram pa bSod nams seng ge (1429-89) *lTa ba’i shan ’byed (Differentiation of the Views)*, in *Sa skya pa’i bKa’ ’bum, The Complete Works of the Great Masters of the Sa skya Sect of Tibetan Buddhism*, compiled by bSod nams rgya mtsho, vol. 13, Tokyo: The Toyo Bunko, 1969.(The full title of this work is *lTa ba’i shan ’byed theg mchog gnad kyi zla zer*.)

Kaschewsky, R. (1971) *Das Leben des Lamaistischen Heiligen Tsongkhapa Blo bzang grags pa (1357-1419) (The Life of the Lamaist Holy Man Tsong kha pa Blo bzang grags pa (1357-1419))*, Wiesbaden: Harrasowitz.(An account of Tsong kha pa’s life based on an eighteenth-century Tibetan biography.)

Napper, E. (1989) *Dependent-Arising and Emptiness*, London: Wisdom Publications, 1989.(An introduction to Tsong kha pa’s Mādhyamika with a translation of parts of the *Lam rim chen mo*.)

Tauscher, H. (1995) *Die Lehre von den zwei Wirklichkeiten in Tsoō kha pas Madhyamaka-Werken (The Two-Truth Theory in Tsong kha pa’s Madhyamaka Works)*, Vienna: Arbeitskreis für Tibetische und Buddhistische Studien.(A thorough study of an important theme in Tsong kha pa’s philosophy.)

Tillemans, T. (1992) ‘Tsong kha pa *et al*. on the Bhāvaviveka-Candrakīrti Debate’, in Shōren Ihara and ZuihōYamaguchi (eds) *Tibetan Studies*, Narita: Naritasan Shinshoji, vol. 1, 316-26.(Tsong kha pa and other dGe-lugs-pas on the debate in the first chapter of Candrakīrti’s *Prasannapadā*.)

Tomabechi, T. (1992). ‘Nāropa no Pañcakrama chūto sono Tibetto Bukkyōni okeru Ichi - Tsoō kha pa no Tachiba o Chūshin ni’ (Nāropa’s Commentary on the Pañcakrama, and its Place in Tibetan Buddhism - On Tsong kha

pa's Standpoint), *Report of the Japanese Association for Tibetan Studies* 38: 1-9.(An investigation of Tsong kha pa's synthesis of the *Pañcakrama* of Tantric Nāgārjuna and the Six Teachings of Nāropa. We gratefully acknowledge Tomabechi's assistance in the preparation of section §3 of the present entry.)

Yoshimizu, C. (1989) *Descriptive Catalogue of the Naritasan Institute Collection of Tibetan Works*, vol. 1, Narita: Naritasan shinshoji.(Includes a list, with brief descriptions, of the works in Tsong kha pa's *Collected Works*.)

Tucker, Abraham (1705-74)

Like many of his eighteenth-century British contemporaries, Abraham Tucker was an empiricist follower of John Locke. Tucker held that the mind begins as a blank slate and remains nothing more than a passive receptacle for 'trains' of ideas with 'a motion of their own'. In his moral philosophy Tucker proposed that the motive of all our actions is the prospect of our own satisfaction, and that the maximization of everyone's satisfaction is the ultimate moral good. (The latter view became a central tenet of the utilitarians who followed him.) According to Tucker, God ensures that our self-interested motivation will be congruent with morality, for God has arranged that we will be rewarded for good and punished for evil - either in this world or in the next. Among those most influenced by his work was the utilitarian and philosophical theologian William Paley.

1 Life

Abraham Tucker was born in London to wealthy parents in 1705. His father died during Abraham's infancy, and an uncle, Isaac Tillard, became his guardian. Tucker studied law at Oxford, but never practised. At 22, he purchased Betchworth Castle; at 31, he married Dorothy Barker, with whom he had two daughters.

Tucker's most significant philosophical work is *The Light of Nature Pursued*, published under the pseudonym Edward Search. A preliminary version of this work appeared in 1763 under the title *Freewill, Foreknowledge, and Fate*, but the full seven volumes of *The Light of Nature Pursued* were not published until 1777, three years after Tucker's death.

2 Associationist psychology

Like several other philosophers in the British empiricist tradition (notably David Hume), Tucker sought to explain the workings of the mind in terms of the principles of association that cause one idea to follow upon another. Among the associationists, Tucker makes explicit what some others seem to take as an unstated assumption: that if the flow of our ideas is entirely determined by certain principles of association, then the understanding has no role to play save that of a passive receptacle for ideas that flow through it according to laws of their own.

Our ideas combine with one another in either of two ways - by composition, in which several ideas 'melt together' to form a complex idea, and by association proper, in which ideas 'appear in couples strongly adhering to each other, but not blended' ([1777] 1977, vol. 1: 221). Ideas that are related through association can link together severally to form 'trains', in which a first idea couples with a second, which couples with a third, and so on.

After ideas become associated in a 'train', some of them may drop out, so that the earlier elements of the train come to associate directly with the later ones by means of 'translation'. When we first see a proof of a mathematical theorem, the theorem, the steps of the proof, and the certainty that we attach to each of these steps will associate with one another to form a train. Yet later we may recall only the theorem itself and the certainty that we once associated with the proof.

Tucker takes this process of translation to be the explanation for a variety of psychological phenomena, including (for example) our sympathy with the moods of others. As children, we soon learn to associate our own happiness with the actions of others, and to associate their actions with their moods - they may play with us when they are happy or ignore us when they are sad. Eventually, we come simply to associate our own pleasure or displeasure with their happy or unhappy moods - without any intervening link of action.

3 Moral and religious philosophy

Tucker says of the will (the active element of the mind) that 'all her motions depend upon motives, thrown upon her from external objects, or conveyed by the channels of experience, education, and example, or procured by her own cares and industry, whereto she was instigated by former motives' ([1777] 1977, vol. 4: 302). The 'active ingredient' in all our motives, however we might describe their details, is the prospect of our own satisfaction. (Tucker understood 'satisfaction' to be a more general term than pleasure, combining the presence of any desirable feelings with the avoidance of any undesirable feelings.) And 'the summum bonum… happiness… is the aggregate of satisfactions' ([1777] 1977, vol. 2: 233). Thus we are absolutely determined to seek our own happiness. But

Tucker's is a soft determinism that is compatible with free will, since 'freewill needs no compulsive force to keep her steady, for she communicates, by antecedent and external causes giving birth to her motives, with the fountain [God's design] whence all the other streams derive' ([1777] 1977, vol. 4: 303).

According to Tucker, 'rectitude hath not a substantiality or distinct essence of its own, but subsists in the relation to happiness, those actions being right which upon every occasion tend most effectually to happiness' ([1777] 1977, vol. 7: 151). And in this context, Tucker understands happiness to encompass the aggregate of satisfaction for all God's creatures. Thus he anticipates the later utilitarians William Paley and Jeremy Bentham not only in asserting that the greatest total happiness is the greatest good, but also in proposing a consequentialist morality in which the good is prior to the right (see Utilitarianism).

So we are absolutely determined to pursue our own happiness, yet moral rectitude lies in the maximal happiness of all. But according to Tucker there is no difficulty here: our self-interested motivation and moral rectitude are congruent; for God has arranged that each of us shall achieve our greatest happiness through the greatest happiness of all.

In order to establish the existence of God and gain insight into God's attributes, Tucker employs an extended argument that combines elements of the first-cause argument with elements of the argument from design (see God, Arguments for the Existence of §§1, 4-5). (At one point in this argument Tucker makes use of the clock and clockmaker analogy that William Paley would later popularize.) Tucker's God is omnipresent (existing necessarily, thus existing everywhere and for all time), omnipotent (powerful enough to be the cause of everything), and omniscient (intelligent enough to plan a universe).

Among the attributes of the creator which are most important to Tucker's moral philosophy are two manifestations of God's goodness: God's providence and God's equity. Tucker proposes that the workings of nature give evidence of the providence of a creator who aims at the happiness of everyone. Tucker also argues that because of 'the wants and weaknesses of human nature' ([1777] 1977, vol. 3: 278) we humans are partial to those from whose good opinion we think we may benefit, while God wants no benefit that any human being can provide. So God must be impartial.

Thus the creator aims at our happiness, and the creator also plays no favourites - each of us will receive an equal measure of happiness over eternity. (Tucker rejects the notion of eternal punishment for the wicked.) For that reason, if we seek to advance our own satisfaction at the expense of others, we know that any short-term advantage we may gain will ultimately be lost - either in this world or in the next. And if we choose to sacrifice our own satisfaction to promote the general good, we know that God will compensate us fully in the end.

Now it may seem that on Tucker's account, we will come out no better if we do good than if we do evil. It may not pay to do wrong, but does it pay to do right? Tucker argues that when we promote the general good, we increase the store of happiness that is available to us all, so that every benevolent act will enhance the 'profit to the common stock in partnership' ([1777] 1977, vol. 4: 502). Thus 'whoever adds to the happiness of another, adds thereby to his own' ([1777] 1977, vol. 4: 500).

4 Influences

Throughout *The Light of Nature Pursued* Tucker repeatedly acknowledges his debts to John Locke, but his other influences are hard to trace. He often mentions David Hartley (whose associationist psychology probably inspired Tucker's own) and George Berkeley, but both of these thinkers are usually cited in criticism. Tucker undoubtedly owes a debt to John Gay, but never refers to him by name; perhaps Gay's influence on Tucker came chiefly through Hartley. And Tucker may have been completely ignorant of his most famous contemporary, David Hume. Even where the two thinkers' ideas seem most similar, as in their belief in soft determinism and their associationist accounts of sympathy, they diverge in their arguments and in the details of their views.

Tucker's influence on the utilitarians who followed him, particularly William Paley, is powerful and direct. In the preface to his *Principles of Moral and Political Philosophy*, Paley says:

> There is, however, one work to which I owe so much that it would be ungrateful not to confess the obligation. I mean the writings of the late Abraham Tucker…. I have found in this writer more original thinking and observation, upon the several subjects that he has taken in hand, than in any other, not to say, than in all others

put together.

(1785: xiii)

See also: Hume, D.; Locke, J.

T. McNAIR

List of works

Tucker, A. (1755) *The country gentleman's advice to his son, on his coming of age, in the year 1755, with regard to his political conduct. Shewing, amongst other things, the folly and pernicious consequences of all party clubs*, London: Owen.(A short tract advising against affiliation with political parties.)

Tucker, A. (1763) *Freewill, Foreknowledge, and Fate*, London: Dodsley.(A preliminary version of what would later become *The Light of Nature Pursued*. Published under the pseudonym Edward Search.)

Tucker, A. (1763) *Man in Quest of Himself: or A Defence of the Individuality of the Human Mind, or Self*, London: Dodsley; repr. in S. Parr (ed.) *Metaphysical Tracts by English Philosophers of the Eighteenth Century*, London: Lumley, 1837, 171-210.(A response to some criticisms of *Freewill, Foreknowledge, and Fate*. Originally published under the pseudonym Cuthbert Comment.)

Tucker, A. (1773) *Vocal Sounds*, Menston: Scholar Press, 1969.(A curious little volume in which Tucker proposes a reformed English alphabet.)

Tucker, A. (1777) *The Light of Nature Pursued*, 7 vols; 2nd edn, 1805; repr. London: Garland, 1977.(Tucker's major work, written in a discursive and painstaking style. Includes a short biography of the author. The first edition was published under the pseudonym Edward Search.)

References and further reading

Albee, E. (1902) *A History of English Utilitarianism*, Bristol: Thoemmes, 1990, 130-64.(A clear summary of Tucker's thought and its place in the context of British moral philosophy.)

Paley, W. (1785) *Principles of Moral and Political Philosophy*, London: Faulder, xiii.(Source of the quotation that closes this entry.)

Stephen, L. (1876) *History of English Thought in the Eighteenth Century*, New York: Harcourt Brace, 1962, vol. 2, 92-104.(A short but somewhat abstruse account of Tucker's thought, including his religious philosophy.)

Turing machines

Turing machines are abstract computing devices, named after Alan Mathison Turing. A Turing machine operates on a potentially infinite tape uniformly divided into squares, and is capable of entering only a finite number of distinct internal configurations. Each square may contain a symbol from a finite alphabet. The machine can scan one square at a time and perform, depending on the content of the scanned square and its own internal configuration, one of the following operations: print or erase a symbol on the scanned square or move on to scan either one of the immediately adjacent squares. These elementary operations are possibly accompanied by a change of internal configuration. Turing argued that the class of functions calculable by means of an algorithmic procedure (a mechanical, stepwise, deterministic procedure) is to be identified with the class of functions computable by Turing machines. The epistemological significance of Turing machines and related mathematical results hinges upon this identification, which later became known as Turing's thesis; an equivalent claim, Church's thesis, had been advanced independently by Alonzo Church. Most crucially, mathematical results stating that certain functions cannot be computed by any Turing machine are interpreted, by Turing's thesis, as establishing absolute limitations of computing agents.

1 Human computors

In his seminal article of 1937, 'On Computable Numbers, With an Application to the *Entscheidungs-problem* [decision problem]', Alan Mathison Turing argued that a function is calculable by some algorithmic procedure if and only if it is computable by a Turing machine. The main claim in Turing's argument is that the behaviour of a human being carrying out an algorithmic procedure (adopting the terminology of Gandy (1988), a 'computor') is constrained by limitations of human memory and sensory capacities. It is appropriate to review briefly these restrictive conditions, for they are reflected in the mathematical definition of a Turing machine. (For a careful analysis of Turing's argument, see Sieg (1994); an analogue of Turing's argument for parallel computations by digital computers is given by Gandy (1980).)

Turing considered a 'one-dimensional' computing space formed by a potentially infinite tape uniformly divided into squares, in each of which at most one symbol may be written. What are the possible processes which can be carried out by a computor working on this computing space?

(1) The computor can write on the tape symbols from a fixed finite list (for example, letters of the alphabet). This restriction is argued for on the ground that if one allowed an infinity of symbols, then one would have symbols differing to an arbitrarily small extent (which, presumably, could not be unambiguously recognized or written down by the computor in view of sensory and memory limitations).
(2) The computor can observe squares on the tape. Limitations of human visual apparatus seem sufficient to justify the existence of a fixed finite bound *B* on the number of contiguous squares observable at any one moment.
(3) The computor can remember earlier stages of the computation and bring this information to bear on future behaviour. This memory consists in what Turing calls the computor's 'states of mind'. There is a fixed finite bound on the number of 'states of mind' which need be taken into account in a mechanical calculation. If one admitted the possibility of using infinitely many states of mind, Turing observes, some of them would be arbitrarily close and would be confused by the computor.
(4) The elementary actions the computor can perform consist of a change of either (a) observed squares or (b) the contents of some squares, possibly accompanied by a change of 'state of mind'. Any other operation of the computor can be viewed as a composition of these elementary actions. Furthermore, Turing assumes that (c) each of the newly observed squares must be within a bounded distance *L* of an immediately previously observed square, and that nothing will be lost by supposing that only the contents of observed squares can be changed.
(5) Which elementary action the computor is to undertake at a given moment depends only on what the computor currently sees and remembers. Thus, the observed squares and the 'state of mind' at any moment uniquely determine the action to be undertaken.

These restrictive conditions ensure that the computor's behaviour is fixed by a finite list of instructions: there are only finitely many different types of observable configurations of symbols (by (1) and (2)) and finitely many 'states of mind' (by (3)); to each pair formed by a 'state of mind' and type of observable configuration is uniquely

associated (by (5)) an elementary action (fixed by (4)).

2 (Universal) Turing machines

Turing claims that the behaviour of a computor satisfying the restrictive conditions above can be faithfully simulated by a machine organized as follows. The internal configurations $q_1, \ldots, q_n$ of the machine correspond to the states of mind $m_1, \ldots, m_n$ of the computor. The machine scans *B* squares corresponding to the *B* squares observed by the computor. In any of its actions, the machine can change a symbol on one of the scanned squares or shift attention to observe another configuration of symbols in such a way that condition (4)(c) is satisfied. The action undertaken at any moment by the machine, and its next internal configuration, are uniquely determined by the scanned symbols and the internal configuration at that moment. In turn, one can show that any computation of this machine can be carried out by a machine *M* endowed with a more elementary repertoire of actions, which provides an intuitive visualization of a Turing machine. This machine has a fixed finite number of internal configurations; it is capable of scanning only one square at a time (that is, $B = 1$); it may print or erase a symbol on the scanned square or else move on to scan either of the immediately adjacent squares (that is, $L = 1$).

To the machine *M* one can associate a function $f(x_1, \ldots, x_n)$ of *n* non-negative integers in the following way. Given an *n*-tuple $a_1, \ldots, a_n$, since the machine can print at least one symbol, say 1, the input $\langle a_1, \ldots, a_n \rangle$ can be represented on an otherwise blank tape by a sequence of $a_1 + 1$ consecutive 1's followed by a blank square followed by a sequence of $a_2 + 1$ consecutive 1's followed by a blank square… followed by a sequence of $a_n + 1$ consecutive 1's. The machine is to be started in a specific internal configuration, q_1, scanning the leftmost square containing a 1. The output number (if any), which is to be identified with $f(a_1, \ldots, a_n)$, is given by the number of 1's written on the tape to the right of the scanned square when (and if) the machine halts. These stipulations enable one to associate to *M* an *n*-ary *partial* (rather than total) function (that is, a function which may be undefined for some *n*-tuple of non-negative integers), because for some input the machine may fail to provide an output by not halting.

The initial stage of computation of *M* on the input $\langle a_1, \ldots, a_n \rangle$ has been described by giving (1) the tape content, (2) the machine's internal configuration, and (3) the scanned square. Each subsequent stage of the computation can be described by providing this information, which is called an 'instantaneous description' of the machine. Any computation which produces a final output can be described by a finite sequence of instantaneous descriptions.

The essential features of this type of machine can be captured by precise mathematical definitions. Let us consider the list of symbols $S_0, S_1, S_2, \ldots,$ $q_1, q_2, q_3, \ldots,$ *R* and *L*, and sequences of symbols from this list having one of the following forms:

(1) $q_i S_j S_k q_l$
(2) $q_i S_j R q_l$
(3) $q_i S_j L q_l$.

Sequences of this form are called 'quadruples'. Turing machines will be now defined (following Davis 1958) entirely in terms of quadruples, which can be interpreted as specifying the next action of a Turing machine when its internal configuration is q_i and the symbol it is scanning is S_j. Quadruple (1) indicates that S_j is to be replaced by S_k and the machine is to enter internal configuration q_l; quadruples (2) and (3) indicate that the machine is to shift one square to the right or to the left, respectively, and then enter internal configuration q_l. We stipulate that the symbol S_0 will serve as a blank; thus, replacing symbol S_j by S_0 amounts to erasing it.

Now a Turing machine is a finite, non-empty set of quadruples that contains no two quadruples with the same first two symbols. The latter condition guarantees the deterministic character of the machine operation, since no more than one instruction will be applicable at any given moment.

This definition identifies a Turing machine with a finite table of instructions in a certain normal form. Equally rigorous definitions can be given for the notions of instantaneous description and computation, as well as for the Turing computable functions, that is, the class of partial functions that Turing machines can compute (see Davis 1958: ch. 1). A basic result of computability theory is the extensional equivalence between this class of functions and the partial recursive functions. Analogous results have been obtained for classes of functions defined in terms of other algorithmic processes (Markov algorithms, lambda-calculi, Post's production systems, Kleene's

equational calculus, and so on; see Church's thesis; Computability theory).

One can define an algorithmic one-one correspondence of the members of the class of Turing machines with the set of all non-negative integers. This fact entails that there is an algorithm (and, by Turing's thesis, a Turing machine) enabling one to retrieve, given any integer *x*, the Turing machine indexed by *x* and, conversely, given any Turing machine, the integer associated with it (see Rogers 1958). This mapping yields an enumeration $M_0, M_1, M_2, \ldots$ of the class of Turing machines. Correspondingly, we shall indicate by f_x the partial function computable by M_x.

Under any such enumeration, there is a number *u* such that for any numbers *x* and *y*, the machine M_u computes the following binary function:

$$f_u(x, y) = \begin{cases} f_x(y) & \text{if } f_x(y) \text{ is defined} \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Thus, the machine M_u can be employed to compute any unary Turing computable function. Furthermore, M_u can also compute the *n*-ary Turing computable functions, provided one codes any *n*-tuple $\langle y_1, \ldots, y_n \rangle$ of input numbers by means of a non-negative integer. Again, this coding can be given by means of Turing computable functions (see Davis 1958: ch. 4). In view of these properties, M_u is called a universal Turing machine. A universal Turing machine is explicitly described in Turing (1937).

3 The horizon of computability (and beyond)

There are number-theoretic functions which cannot be computed by any Turing machine. The mere existence of such functions follows from the fact that the class of Turing computable functions is denumerable (it can be put into a one-one correspondence with the non-negative integers) whereas, by Cantor's theorem (see Cantor, G.), there are indenumerably many number-theoretic functions.

Specific examples of non-computable functions were provided by Turing in his 1937 paper. Consider the so-called 'halting problem': is there a Turing machine capable of deciding for arbitrary *x* and *y* whether the Turing machine M_x will eventually halt after being started on input *y*? In other words, is the total number-theoretic function

$$h(x, y) = \begin{cases} 1 & \text{if } f_x(y) \text{ is defined} \\ 0 & \text{otherwise} \end{cases}$$

a Turing computable function? The answer to this question is negative: the function $h(x, y)$ is not Turing computable, and the halting problem is said to be unsolvable. Turing used the unsolvability of the halting problem to establish a negative solution to Hilbert's *Entscheidungsproblem* [decision problem] (see Church's theorem and the decision problem; Hilbert's programme and formalism).

The unsolvability of the halting problem implies that the behaviour of some Turing machines cannot be completely predicted by any Turing machine. An example of such 'unpredictable' machines is just the universal Turing machine M_u. The existence of a Turing machine capable of determining, for every *x* and *y*, whether the machine M_u will eventually halt when started on input $\langle x, y \rangle$ conflicts with the negative solution to the halting problem.

The import of this and other 'negative' results concerning Turing machines crucially hinges upon Turing's thesis. Indeed, if a function cannot be computed by a Turing machine then, by Turing's thesis, one can infer that there is no algorithm enabling one to calculate the values of this function. In particular, no algorithmic procedure can enable one to answer, for each *x* and *y*, all questions of the form 'Does M_x halt on *y*?' or, equivalently, 'Does *x* belong to the set K_y formed by the numbers *z* such that $f_z(y)$ is defined?'.

Suppose we are provided with some unspecified means - an oracle, as it were - for solving problems of the form 'Does *x* belong to *X*?', where *X* is a set of non-negative integers. One may then define a new class of machines, which one may call *O*-machines, suitably extending the definition of Turing machine so as to allow for quadruples of the form

(4) $q_i S_j q_k q_l$,

and associating to each *O*-machine a specific set *X* of non-negative integers. An *O*-machine in internal configuration q_i which is scanning symbol S_j may be interpreted as asking the oracle: 'Does the number *x* of occurrences of 1 on my tape at this moment belong to *X*?' If the answer is 'yes' then the *O*-machine changes its internal configuration to q_k; if the answer is 'no' then the *O*-machine changes its internal configuration to q_l (see Davis 1958: ch. 1). The function computable by an *O*-machine is said to be computable in the set *X* associated with that machine.

A class of functions computable in a set *X* may include functions that are not Turing computable. Furthermore, given a set *X*, one can prove the existence of functions that are not computable in *X*, that is, functions which cannot be computed by *O*-machines having access to an oracle capable of deciding membership of the set *X*. Results of this kind are central to the theory of the degrees of unsolvability, which introduces a hierarchy of classes of unsolvable problems.

One may also distinguish between classes of solvable problems, in terms of the amount of resources a computing machine may need for their solution. Among the various, equivalent models that are available, the model of computation commonly chosen to estimate these resources is that of the Turing machine (see Complexity, computational).

The notion of (universal) Turing machine also plays a prominent role in philosophical discussions of the mind-machine problem. Theoretical limitations of (universal) Turing machines have been appealed to in arguments purporting to show that the human mind is not subject to analogous limitations. Turing (1950) employed this notion in order to circumscribe the class of devices that are admitted to compete with human beings in his test for intelligent behaviour. More recently, (universal) Turing machines have been used to isolate distinguishing features of the notion of computational explanation in cognitive science, as well as heuristic ideas underlying research programmes in artificial intelligence (see, for example, Newell and Simon 1976).

See also: Logical and mathematical terms, gossary of

GUGLIELMO TAMBURRINI

References and further reading

Davis, M. (1958) *Computability and Unsolvability*, New York: McGraw-Hill; repr. New York: Dover, 1982. (Classic introduction to Turing machines and the theory of computability. The first part of this book is accessible to readers with no special mathematical training.)

Davis, M. (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Hewlett, NY: Raven Press.(An anthology of fundamental papers on undecidability and unsolvability, which includes Turing (1937).)

Gandy, R. (1980) 'Church's Thesis and Principles for Mechanisms', in J. Barwise, H.J. Keisler and K. Kunen (eds) *The Kleene Symposium*, Amsterdam: North Holland, 123-45.(Presents an analogue of Turing's argument for parallel computations by digital computers.)

Gandy, R. (1988) 'The Confluence of Ideas in 1936', in R. Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press, 55-112.(An informative discussion of the conceptual background of Turing's analysis.)

Herken, R. (ed.) (1988) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press. (See especially Kleene's and Gandy's articles. Feferman's article reviews the ideas leading to the notion of oracle machine.)

Kleene, S.C. (1988) 'Turing's Analysis of Computability, and Major Applications of It', in R. Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press, 17-54. (Includes a terse outline of Gödel's incompleteness theorems, highlighting their relation to Turing machines in a form accessible to non-specialists.)

Newell, A. and Simon, H. (1976) 'Computer Science as Empirical Enquiry: Symbols and Search', *Communications of the ACM* 19: 113-26.(Formulates broad heuristic hypotheses in artificial intelligence and cognitive science by crucial appeal to universal computing machines.)

Rogers, H. (1958) 'Gödel Numbers of Partial Recursive Functions', *Journal of Symbolic Logic* 23: 331-41.(A technical analysis of the numerical codings referred to in §2.)

Sieg, W. (1994) 'Mechanical Procedures and Mathematical Experience', in A. George (ed.) *Mathematics and Mind*, Oxford: Oxford University Press, 71-117.(Includes a careful analysis of Turing's argument.)

Turing, A.M. (1937) 'On Computable Numbers, With an Application to the *Entscheidungsproblem* [decision problem]', *Proceedings of the London Mathematical Society*, series 2, 42: 230-65; repr. in M. Davis (ed.) *The Undecidable*, Hewlett, NY: Raven Press, 1965.(Turing's seminal article in the theory of computability: §9 presents the major argument for Turing's thesis.)

Turing, A.M. (1950) 'Computing Machinery and Intelligence', *Mind* 59: 433-60; repr. in D.C. Ince (ed.) *Collected Works of A.M. Turing*, vol. 3, *Mechanical Intelligence*, Amsterdam: North Holland, 1992. (An influential article on the mind-machine problem, presenting Turing's reflections on (universal) computing machines and intelligent behaviour.)

Turing reducibility and Turing degrees

A reducibility is a relation of comparative computational complexity (which can be made precise in various non-equivalent ways) between mathematical objects of appropriate sorts. Much of recursion theory concerns such relations, initially between sets of natural numbers (in so-called classical recursion theory), but later between sets of other sorts (in so-called generalized recursion theory). This article considers only the classical setting. Also Turing first defined such a relation, now called Turing- (or just T-) reducibility; probably most logicians regard it as the most important such relation. Turing- (or T-) degrees are the units of computational complexity when comparative complexity is taken to be T-reducibility.

1 Usage and notation

ω = the set of all natural numbers; a real is a subset of ω. For sets A, B: $A \cup B = \{x : x \in A \text{ or } x \in B\}$; $A \times B = \{< x, y >: x \in X, y \in Y\}$; for $n \in \omega$, $n \neq 0$: $A^n = A \times \ldots \times A$ (taking *A* *n*-times). For a function φ: $\mathrm{dom}(\varphi)$ = the domain of φ; φ is a particle function on ω^n iff $\mathrm{dom}(\varphi) \subseteq \omega^{\mathrm{n}}$; such a function is total iff its domain is ω^n. In this article a partial function will be partial on ω^n for some $n \in \omega$; φ is total iff $\mathrm{dom}(\varphi) = \omega^n$. (So a total function on ω^n is also partial - an unfortunate, but standard, usage.) For a real *X*, χ_X, is *X*'s characteristic function: $\chi_X(x) = 1$ if $x \in X$; $\chi_X(x) = 0$ if $x \notin X$.

2 Informal characterization of T-reducibility

For reals *X* and *Y*: *X* is reducible to *Y* iff *Y* is at least as computationally complex (in a sense whose specification determines the reducibility relation) as *X* is, that is, given any $n \in \omega$ we could answer the question '$n \in x$?' by means that are in some sense computational except (perhaps) for using (subject to various possible constraints) answers to membership-questions about *Y*. For T-reducibility, we require computation to be effective and impose no restraints. A relativized algorithm consists of instructions like an 'absolute' algorithm (see Computability theory) except that they may tell one to 'ask a question of an oracle', one of the form 'what about *y*?', where *y* can be any natural number given to the oracle as a numeral; running the algorithm relative to *Y*, in a finite time the oracle says 'yes' if $y \in Y$, and 'no' otherwise. Suppose *A* is such a relativized algorithm, and we have an oracle for the real *Y* and infinite amount of time. Were we to apply *A* to a given input, we would grind along following *A* as we would an 'absolute' algorithm, except that from time to time *A* might specify a natural number (in some sort of numeral form) for us to ask the oracle about; the oracle would answer; we would grind on following *A*, exactly what *A* requires us to do being 'mechanically' determined by the answers the oracle gave to this and previous queries; eventually *A* might tell us to stop, delivering an output; *A* also might leave us computing forever. The oracle is oracular in that we do not trouble ourselves about how it comes up with its answers; going from a query of the oracle to its answer is a non-effective step in an otherwise effective computation. Now we will introduce some 'informal' definitions.

(I) A partial function φ on ω is computed by an algorithm *A* relative to a real *Y* iff for any $x \in \omega$, were we to run *A* on input *x* using an oracle for *Y*: if $x \in \mathrm{dom}(\varphi)$ the computation would halt giving the value $\varphi(x)$; if $x \notin \mathrm{dom}(\varphi)$ the computation would not halt. φ is computable relative to *Y* iff φ is computed by some such *A* relative to *Y*. *X* is decidable relative to *Y* iff χ_X is computable relative to *y* (that is, for some *A* for any $\chi \in \omega$, were we to run *A* on input *x* using an oracle for *Y*, the computation would eventually halt with a correct answer (1 = 'yes', 0 = 'no') to the question '$x \in X$?'. *X* is recursively enumerable relative to *Y* iff *X* is the range of a function that is total on ω and computable relative to *Y*.

3 Two rigorous definitions

One way to replace these informal definitions by rigorous definitions uses the notion of an oracle machine (that is, a relativized Turing machine (see Turing machines)), credited by Rogers (1967) to M. Davis. Fix a finite set ALPH (an alphabet) whose members we will call symbols, say with $0 (= \mathrm{blank}), 1, 2 \in \mathrm{ALPH}$; without mathematical loss, we could suppose that $\mathrm{ALPH} = \{0, 1, 2\}$. Recall the three usual rule-schemata defining being a Turing-machine table (for a single tape infinite in both directions and with '0' meaning 'blank') relative to ALPH and to a set STATE of internal states. Add to them this fourth schema (Query): for *y* = the number of occurrences of 1 on the tape at the moment, if in state *q* then: if the oracle says 'yes' to *y* then change to state q'; if it says 'no'

change the state q''. (By using occurrences of 2 rather than 1, the device can record and preserve results of previous steps while 'asking the oracle a question', insulating this record from the question asked.)

Pick $L, R \notin \mathrm{ALPH}, \mathrm{ACT} = \mathrm{ALPH} \cup \{L, R\}$. Represent any Turing-machine rule by a quadruple of the form $< q, s, a, q' >$ for $q \in \mathrm{STATE}, \mathrm{s} \in \mathrm{ALPH}, \mathrm{a} \in \mathrm{ACT}, \mathrm{q}' \in \mathrm{STATE}$, with this understanding: if $a \in \mathrm{ALPH}$ the instruction says 'Erase the cell and write *a*'; *L* and *R* represent 'Move the tape one cell to the left (that is, the head one cell to the right)' and 'Move the tape one cell to the right', respectively. Any rule of form (Query) can be represented by a quadruple $< q, s, a, q'q'' >$for $q, q', q'' \in \mathrm{STATE}, \mathrm{s} \in \mathrm{ALPH}$.

Let an oracle-machine table on $<< \mathrm{ALPH}, \mathrm{STATE} >$ be a function from $\mathrm{STATE} \times \mathrm{ALPH}$ into $(\mathrm{ACT} \cup \mathrm{STATE}) \times \mathrm{STATE}$. Identifying a function *f* with the set of ordered pairs of the form $< x, f(x) >$, and identifying an ordered pair of the form $<< a, b >, < cd >>$ with $< a, b, c, d >$, such a table is a set of quadruples coding instructions. For mathematical purposes, an oracle machine is an ordered pair $M =< T, q_0 >$ for *T* an oracle-machine table and $q_0 \in \mathrm{STATE}$; q_0 is *M*'s initial state. Given a tape with cells in any condition, a particular cell on that tape, and a real *Y*, if each query of the oracle is true to *Y* (that is, if the device is in state *q* reading *s*, *y* is as above, and $T(< q, s >) =< q', q'' >$, then the machine goes into state q' if $y \in Y$, and goes into state q'' otherwise), an oracle machine *M* carries out a uniquely determined succession of operations starting in state q_0 with its head on that chosen cell, one which may or may not terminate in a finite number of steps; this is its computation (also known as its run) relative to *Y*. The definition of an oracle machine does not involve a particular value for '*Y*': a given oracle machine *M* with a given input can run with any real *Y*, perhaps producing different computations for different reals. Also , talk of tapes and devices is heuristic, replaceable by an austere but tedious definition of the computation *M* carries out input *x* using *Y*.

(II) We associate with each such *M* and *Y* a partial function φ on ω as follows: represent a natural number *x* as input by string of $x + 1$ consecutive occurrences of 1 on the tape with all other cells blank; adding 1 distinguishes input 0 from starting with a blank tape. Start *M* in state q_0 reading the leftmost cell containing a 1; as output take the total number of occurrences of 1 appearing anywhere on the tape when and if the machine stops; thus $\varphi(x) =$ that total number of 1s; $x \in \mathrm{dom}(\varphi)$ iff *M* eventually halts when run with input *x*; if *M* does not halt, $x \notin \mathrm{dom}(\varphi)$ and $\varphi(x)$ does not exist (as mathematicians say, ' $\varphi(x)$ is undefined'). This φ is the partial function computed by *M* using *Y*. φ is T-computable from *Y* iff φ is computed by some oracle machine using *Y*. *X* is T-reducible to *Y* (in symbols $X \leq_T Y$) iff χ_X is T-computable from *Y*.

For $\mathrm{ALPH} = \{0, 1, 2\}$, represent the possible 'actions' by natural numbers: $i \leq 2$ for 'Erase the cell and write *i*'; 3 for *L*; 4 for *R*; represent the internal states as natural numbers greater than 4. Represent (or better, define) a machine-table as a function from a finite subset of $\omega \times \{0, 1, 2\}$ into a finite subset of $\omega \times \omega$; such a function can then be represented by a single natural number. An oracle machine *M* can be coded by a single natural number, *M*'s Gödel-number (see Gödel's theorems). For *e* = *M*'s Gödel-number, let φ^e_Y (alternative notation: $\{e\}^Y$) be the partial function *M* computes using *Y*.

(III) Turing himself did not use the notion of an oracle machine in his original definition of T-reducibility; rather he defined the class of partial functions (each on ω^n for any $n \in \omega$) recursive in a real *Y* (also known as the class of functions partial recursive in *Y*) to be the least class $\underline{C}$ meeting these conditions: (1) all initial functions belong to $\underline{C}$; these are successor, the function mapping each ω^n into $\{0\}$, and all projections (that is, functions of the form $f(x_1, \ldots, x_n) = x_i$ for some *i*, $1 \leq i \leq n$); (2) $\chi_Y \in \underline{C}$; (3) $\underline{C}$ is closed under composition, primitive recursion and minimization. A real *X* is recursive in *Y* iff χ_X is.

It is tedious but not hard to prove that this class is exactly the class of partial functions T-computable from *Y*. Thus *X* is recursive in *Y* iff *X* is T-reducible to *Y*. In fact, each of the various ways of defining being a recursive function extends naturally to a definition of being a function recursive in an arbitrary real *Y*. Some regard the coextensiveness of (II), (III), and those other definitions, as evidence that they all rigorously define what (I) informally defines, that is, as evidence for the relativized version of Church's thesis: a function is computable relative to *Y* iff it is T-computable relative to *Y* (see Church's thesis).

4 More definitions and facts

Let $X <_T Y$ iff $X \leq_T Y$ and not $Y \leq_T X$; *X* and *Y* are T-equivalent (in symbols $X \equiv_T Y$) iff $X \leq_T Y$ and

$Y \leq_T X$. As the notation suggests: $<_T$ is a reflexive transitive relation; $\equiv_T$ is an equivalence relation. We could prove these facts from the mathematical definition of (II) or (III); but it is easier to work informally.

Sample proof Let A and B be algorithms computing χ_X and χ_Y relative to *Y* and *Z* respectively. Consider this algorithm: given any *x*, run *A* on *X*; whenever A tells us to ask the oracle for *Y* about a *y*, run B on *y*, using the oracle for *Z*; eventually such a run of B halts; if it yields 1 or 0 take that as a 'yes' or 'no' respectively from an oracle for *Y*. A halts on an input if responses to its queries of an oracle are true to *Y*; a run of B using an oracle for *Z* does deliver the truth about *Y* (since it computes χ_Y); so this process halts, yielding $\chi_X(x)$, and it uses an oracle for Z (rather than *Y*, unless $Y = Z$). Thus $\leq T$ is transitive. (One of recursion theory's advantages is that we can almost always work in this informal way.)

For a real *X* let $\deg(x) = \{Y{:}\, X \equiv_T Y\}$. Let *a* be a T-degree iff $a = \deg(\mathrm{x})$ for some real *X*; $\mathcal{D}$ = the set of T-degrees. For $a, b \in \mathcal{D}$ let $a \leq b$ iff for some $X \in a, Y \in b, X \leq_T Y$. Note that if $a \leq b$ then for any $X \in a, Y \in b{:}\, X \leq_T Y$. The relation $\leq$ is partial ordering, that is, it is reflexive , transitive, and antisymmetric (that is, if $a \leq b$ and $b \leq a$ then $a = b$). Let 0 = the set of recursive reals. For *X*, $Y \in 0{:}\, X \equiv_T Y$; for any real *Z* if $Z \leq_T X$ then $Z \in 0$; thus $0 \in \mathcal{D}$. T-degrees are sometimes called 'degrees of unsolvability', rather misleading since one of them, 0, represents solvability (= recursiveness). For any real *Y* and any $X \in 0, X \leq_T Y$; thus 0 is the least T-degree under this (the standard) ordering. Study of the algebraic properties of the structure $<< \mathcal{D}, \leq>$ is a major component of recursion theory. One basic fact is that the partial ordering $\leq$ is very partial; for $a, b \in \mathcal{D}$, *a* and *b* are incomparable iff neither $a \leq b$ nor $b \leq a$; each T-degree is incomparable with continuum-many other T-degrees - as many as there are reals.

Let a degree be recursively enumberable (hereafter r.e.) iff it contains an r.e. real. Recursive reals are r.e.; so 0 is r.e. Let $K^X = \{x{:}\, \varphi_x^X(x) \text{ is defined}\}$. In 1944 E. Post gave the first example of a non-recursive r.e. set: for $K = K^{\{\}}$ for {} = the empty set; so deg(*K*) is another r.e. degree. He went on to pose 'Post's problem': are there other r.e. T-degrees? In 1956 R. Friedberg and A. Muchnik independently proved that there are others; some consider this the first deep result in degree theory. (See Friedberg 1957 or any of the textbooks referred to below.) The study of r.e. degrees is a flourishing corner of recursion theory.

The jump function on the reals assigns each real *X* to K^X. The effective version of Cantor's diagonal argument for the unaccountability of the set of reals (see Cantor's theorem) shows that $K^X \leq_T X$; since $X \leq_T K^X, X \leq_T K^X$. If $X \equiv_T Y$ then $K^X \equiv_T K^Y$; so it makes sense to define the jump function on the T-degrees: the jump of *a* (usually written a') is the T-degree of K^X for some (thus any) $X \in a$. Thus $a < a'$, showing the jump for degrees is a natural uniform way of transforming a T-degree into a more complex one. Obviously jumping iterates any finite number of times, yielding degrees of increasingly complex sets; in fact, it iterates through the so-called constructibly countable ordinals (see Hodes 1980), a matter closely connected to Gödel's constructible hierarchy (see The constructible universe).

A sentence in a formal language can be effectively coded as a natural number by 'Gödel numbering' (seeGödel's theorems); so a set of sentences, for example, a theory (in the logician's sense (see Model theory)), can be coded as a real; so its complexity can be identified with the T-degree of that real. Examples are: the theory of a dense linear ordering without endpoints has degree 0; the set of validities of first-order logic is r.e. and so has degree $0'$; ditto for the set of theorems of any axiomatizable theory in a countable language; the set of truths of first-order arithmetic has degree = 0^ω = the result of jumping 0^ω-times.

S. Simpson (1977) proved that the first-order theory of $< \mathcal{D}, \leq,' >$ (that is, the set of true first-order sentences based on a two-place predicate for $\leq$, a one-place function-constant for the jump, and 'first-order logic' with variables ranging over $\mathcal{D}$) was very complicated - of the same T-degree as the set of truths of second-order arithmetic. In 1990 B. Cooper announced a proof (still unpublished at the end of 1993) that the jump-function on $\mathcal{D}$ is first-order definable over $< \mathcal{D}, \leq>$, that is, there is a formula $F(x, y)$ with exactly free variables *x* and *y*, based on a two place predicate for $\leq$ and 'predicate logic', so that with quantifiers ranging over $\mathcal{D}$: for any $a, b \in \mathcal{D}$ $F(a, b)$ is true iff $b = a'$. With Simpson's theorem, this showed that the first-order theory of $< \mathcal{D}, \leq>$ is as complex as the truth-set of second-order arithmetic. Analysis of T-degree of the truth-set have been carried out for many other kinds of structures, including various pieces of $< \mathcal{D}, \leq>$.

See also: Logical and mathematical terms, glossary of

References and further reading

Friedberg, R. (1957) ‘Two recursively enumberable sets of incomparable degrees of unsolvability’, *Proceedings of the National Academy of Science* 43: 236-8.(One of the first two answers to Post’s problem.)

Hodes, H. (1980) ‘Jumping through the transfinite’, *Journal of Symbolic Logic* 45: 204-20. (On iterating the jump through the constructibly countable ordinals.)

Griffor, E. ed. (forthcoming) *Handbook of Recursion Theory*, Amsterdam: North-Holland Publishing.(Survey articles that will be up-to-date as of publication.)

Lerman, M. (1983) *The Degrees of Unsolvability*, Berlin: Springer-Verlag.(For advanced students.)

Nerode, A., and Shore, R. (1980) ‘Second-order logic and first-order theories of reducibility orderings’ in Barwise *et al.* (eds) *Kleene Symposium*, Amsterdam: North-Holland Publishing, 181-200.(A proof of Simpson’s theorem, one easier that the original proof.)

Odifreddi, P. (1989) *Classical Recursion Theory,* Amsterdam: North-Holland Publishing, Elsevier Science Publishers.(A very good introductory text, emphasising breadth, up-to-date as of 1989; the author plans to extend this with at least one further volume.)

Rogers, H. (1967) *The Theory of Recursive Functions and Effective Computability*, New York: McGraw Hill Inc. (The original text on recursion theory, and excellent complement to Odifreddi’s text, with very good exercises.)

Shoenfield, J. (1971) *Degrees of Unsolvability,* Amsterdam: North-Holland Publishing, and New York: American Elsevier.(Lecture notes, not comprehensive, but very readable.)

Simpson, S. (1977) ‘First-order theory of the degrees of recursive unsolvability’, *Annals of Mathematics* 105: 212-138.(The proof in Nerode and Shore (1980) is easier.)

Soare, R. (1987) *Recursively Enumberable Sets and Degrees*, Berlin: Springer-Verlag.(For advanced students.)

Turing, Alan Mathison (1912-54)

Alan Turing was a mathematical logician who made fundamental contributions to the theory of computation. He developed the concept of an abstract computing device (a 'Turing machine') which precisely characterizes the concept of computation, and provided the basis for the practical development of electronic digital computers beginning in the 1940s. He demonstrated both the scope and limitations of computation, proving that some mathematical functions are not computable in principle by such machines.

Turing believed that human behaviour might be understood in terms of computation, and his views inspired contemporary computational theories of mind. He proposed a comparative test for machine intelligence, the 'Turing test', in which a human interrogator tries to distinguish a computer from a human by interacting with them only over a teletypewriter. Although the validity of the Turing test is controversial, the test and modifications of it remain influential measures for evaluating artificial intelligence.

1 Life

Turing was born in London. He was educated at King's College, Cambridge, where he began work in mathematical logic, soon writing his most brilliant article 'On Computable Numbers with an Application to the *Entscheidungsproblem*'. The German '*Entscheidungsproblem*' refers to David Hilbert's question whether mathematics is *decidable*, that is, whether there is for any mathematical assertion a definite procedure which will determine whether or not the assertion is true. Turing's work, along with contemporary work of Kurt Gödel and Alonzo Church, demonstrated the impossibility of such a general decision procedure. Moreover, his work on computable numbers and universal computing machines provided the conceptual foundation for the development of the modern computer.

During the Second World War Turing helped to design special computing equipment to decipher the German 'enigma' codes. Afterwards he led efforts to design some of the earliest computers, including the Automatic Computing Engine (ACE) in 1945 and the Manchester Automatic Digital Machine (MADAM) in 1948. Turing was prosecuted for his homosexuality in 1952. He died at the peak of his intellectual powers, an apparent suicide.

2 Turing machines and minds

Turing machines are highly idealized 'machines' (properly regarded as abstract objects that may or may not be realized in any physical material) that consist of a potentially infinite tape, divided into individual cells, on which the machine can read and write discrete symbols, and a set of state descriptions which specify what the machine is to do when, in a particular state, it reads a particular symbol: it can leave it alone, erase it, write a new symbol in its place, move one cell to the left or right, and enter a new or the same state (see Turing machines). Turing showed how such machines could compute many ordinary arithmetical functions, and conjectured that anything that could be intuitively regarded as computable at all could be computed by such a machine (what has become known as the 'Church-Turing thesis'; see Church's thesis).

A particularly important Turing machine is the universal Turing machine, which serves as the prototype of the standard 'computer': when its tape is supplied with a coded description of any Turing machine (or 'program') and some input for that machine, it computes precisely what *that* machine *would* compute when supplied with that input.

The concept of a Turing machine furnishes a theoretical foundation for much work in cognitive science and artificial intelligence, where the driving idea is that cognitive processes, and maybe all mental processes, are ultimately computational (see Artificial intelligence; Mind, computational theories of). Processes of perception, reasoning, and decision making, for example, are to be explained in terms of computations operating on the input to sense organs. A version of this computational approach to the mind is called 'Turing machine functionalism' (see Block 1980).

One of the philosophically most important consequences of Turing's proposal is that it provides a method of explaining overall mental activity in simple mechanical terms. It thereby meets a recurring worry about psychological theories; that they vacuously postulate an intelligent 'homunculus' to explain intelligent activity.

Rather than resorting to a master homunculus to explain the mind, a hierarchy of computational systems is postulated, each of which is made up of simpler computational systems, until finally there is nothing left but simple mechanical components whose operations can be understood purely in terms of physical causation.

3 The Turing test

Turing not only provided a rigorous account of computation in terms of Turing Machines, but also proposed a classic test for examining machine mentality. His test, the Turing test, is designed to examine the ability of computers to imitate humans. In the Turing test a human interrogator or 'judge' puts questions to two unseen subjects - one a computer and the other a human. The task of the interrogator is to determine, based on teletypewriter communication with the two unseen subjects, which is the computer and which is the human. The goal of the computer is to imitate a human, including the imitation of typical human strengths and weaknesses. To use Turing's own example, when the computer subject is asked to add 34,957 and 70,764, it does not respond immediately but waits about 30 seconds and then responds with 105,621 rather than the correct answer 105,721. Although Turing does not specify, it is reasonable to assume that the interrogator is allowed to ask many questions over an extended period of time. The computer passes the test if and only if the interrogator cannot distinguish (better than chance) between the computer and human responses.

The Turing test has a number of virtues as a scientific test. First, the test is conducted with blind impartiality. The interrogator in the test cannot judge the computer to be non-intelligent because of its appearance alone. The sophistication of the behaviour is the crucial variable, and linguistic behaviour in the test can be evaluated without subtle bias due to appearance (see Behaviourism, analytic). Second, the test is reproducible. If the computer really has intelligence, then it has a cluster of abilities that can be demonstrated over and over again.

The Turing test is, however, open to many criticisms.

(1) The *action* objection: the Turing test examines only linguistic behaviour, but intelligence and thinking must be evaluated in terms of bodily actions within real contexts, not merely through linguistic behaviour over a teletypewriter. A possible reply: actions of any kind can be *discussed* within the Turing test, and that is sufficient to establish intelligence. After all, a human who is largely paralysed but who has adequate linguistic capabilities can still demonstrate intelligence.

(2) The *jukebox* objection: because the test is finite, pre-selected answers could be entered for every possible question (see Block 1981), but surely merely accessing such a list should not be intelligence. A possible reply: this is not a practical option as the number of answers needed is astronomically high. In any remotely realistic test, the proper response at any given time depends on common knowledge of the real world which is constantly and unpredictably changing.

(3) The *syntax* objection: computers operate syntactically, not semantically, and can never understand what they produce even if the result passes the Turing test (see Searle 1980; Chinese room argument). A possible reply: although the computer's operations are defined purely syntactically (in terms of what the cursor reads and writes), this does not preclude the possibility that they possess causally efficacious semantic properties. A computer as a whole functioning system may possess semantics and hence be capable of understanding (see Language of thought).

(4) The *Barnum* objection: there have been *restricted* Turing tests (subject to limits of time, topic and style of question) that have been actually staged, and which a number of computer programs have passed (see Shieber 1994 for discussion). However, as Block (1981) points out, they only show what P.T. Barnum claimed: people are easy to fool. A possible reply: the restrictions on these tests have been too severe; it is very implausible that a computer that passed a fully *unrestricted* test, in which any topic could be discussed in any way for substantial amounts of time, could really lack intelligence.

The Turing test is often understood as providing necessary and sufficient conditions, an 'operational' definition, for computer thinking. Turing, himself, did not insist that the test be so interpreted. Indeed, on the contrary, Turing believed that a machine might fail the test, not due to a lack of intelligence, but due to a lack of the ability to imitate a human. The computer might add too quickly and give itself away. A weaker, inductive interpretation of the Turing test is more defensible. On the inductive interpretation the Turing test is not expected to provide

deductively conclusive evidence, but merely *scientific inductive* evidence for the hypothesis that computers have intelligence similar to humans. How good the evidence is depends on how well the computer does in the test. If a computer actually passed a full Turing Test, then the proponent of the Turing test would claim that good inductive evidence would exist to infer that the computer has intelligence similar to our own.

JAMES H. MOOR

List of works

Turing, A.M. (1992) *Collected Works of A.M. Turing*, ed. D.C. Ince, Amsterdam: Elsevier.(The standard edition of Turing's works.)

Turing, A.M. (1964) 'Computing, Machinery and Intelligence', repr. in A.R. Anderson (ed.) *Minds and Machines*, Englewood Cliffs, NJ: Prentice Hall.(Turing's proposal of the 'Turing test'.)

References and further reading

Block, N. (ed.) (1980) *Readings in Philosophy of Psychology*, vol. 1, Cambridge, MA: Harvard University Press. (Part 3 has an introduction and selection of articles discussing Turing machine functionalism.)

Block, N. (1981) 'Psychologism and Behaviorism', *Philosophical Review* 90: 5-43.(Develops the *jukebox* - or 'blockhead' - counterexample to the Turing test, claiming it could be passed by a machine consisting of a vast look-up table.)

Block, N. (1990) 'The Computer Model of the Mind', in D. Osherson and E. Smith (eds) *An Invitation to Cognitive Science*, vol. 3, *Thinking*, Cambridge, MA: MIT Press.(A useful introduction to the topic, and some problems both with the model and with the Turing test.)

Boden, M. (ed.) (1990) *The Philosophy of Artificial Intelligence*, Oxford: Oxford University Press.(Selection of articles about the Turing test and the computer model of the mind.)

Hodges, A. (1983) *Alan Turing: The Enigma*, New York: Simon & Schuster.(Excellent biography.)

Searle, J. (1980) 'Minds, Brains, and Programs', *Behavioral and Brain Sciences* 3: 417-24.(Gives the well-known Chinese room argument to criticize the Turing test.)

Shieber, S. (1994) 'Lessons from a Restricted Turing Test', *Communications of the ACM* 37: 70-8.(A discussion of some actual stagings of highly restricted Turing tests, which some computers passed.)

Turnbull, George (1698-1748)

George Turnbull was an early champion of the use of empirical methods in the moral sciences. Involved in contemporary religious debate, he favoured religious toleration and the use of rational argument in defence of Christian belief. He also made contributions to educational theory and practice.

Born in Alloa, Scotland, on 11 July 1698, George Turnbull studied at the University of Edinburgh, where he was an early member of the Rankenian Club. Like his fellow Rankenians, Turnbull was interested in contemporary religious debates, and in 1718 he tried to initiate a correspondence with the deist John Toland (1670-1722) (see Deism; Toland, J.). During the late 1710s he also composed a manuscript treatise on civil religion which remained unpublished because of the controversial nature of his argument in favour of religious toleration. After graduating with his MA from Edinburgh in April 1721, he became a regent at Marischal College, Aberdeen, and taught there until he resigned in 1727.

Although Turnbull had a chequered career in Aberdeen, he played a major part in the transformation of Marischal's philosophy curriculum. He helped to popularize the writings of the Third Earl of Shaftesbury, and he joined with other regents to promote the study of natural jurisprudence and history. Along with some of his colleagues, Turnbull expounded the political views of Old Whigs like Lord Molesworth (1656-1725), with whom he corresponded about education and religion. He also mobilized Newtonian natural philosophy in the service of religion, and developed a rational form of Christianity designed to counter the attacks of deists and atheists. But perhaps his most important contribution was his proposed methodological reformation of moral philosophy. Although he may have been indebted to his teachers at Edinburgh for the initial idea, Turnbull was the first Scottish moralist to advocate in print the use of the experimental method to investigate moral questions, and his advocacy of a scientistic approach to morals left a permanent mark on eighteenth-century Scottish philosophy, especially through the influence of his pupil Thomas Reid.

After such an auspicious start, Turnbull's subsequent career was something of a prolonged anticlimax. Following his departure from Aberdeen, he worked as a tutor to young aristocrats on the Grand Tour until he settled in London in the mid-1730s. During his travels he published two tracts on the rational basis of Christian belief (1731, 1732), and he addressed this topic again in 1740 in his most substantial publication, *The Principles of Moral Philosophy*. Turnbull here reiterated his call for the use of the experimental method in the moral sciences, and elaborated upon this theme in a dissertation appended to his translation of Heineccius' work on natural law (1741). 1740 also saw the appearance of his *Treatise on Ancient Painting*, wherein he expanded upon the basic elements of Shaftesbury's aesthetics in the context of a historical review of the development of painting and sculpture among the ancients. In addition, the *Treatise* considered the role of the fine arts in education, and issues of pedagogy were the primary focus of his last significant book, the *Observations upon Liberal Education* (1742), which despite its derivativeness makes a persuasive case for the structural unity of all branches of human learning. Turnbull's feverish literary activity points to his continuing search for settled employment and, having been ordained in 1739, he sought preferment in the Anglican Church. He was eventually rewarded with a minor position in Ireland, and he died in relative obscurity in the Hague on 31 January 1748.

See also: Common Sense School; Enlightenment, Scottish; Oswald, J.; Tindal, M.

PAUL WOOD

List of works

Turnbull, G. (1723) *De scientiae naturalis cum philosophia morali conjunctione*, Aberdeen: Nicol.(The text of this graduation thesis argues for the close connection of natural science and moral philosophy.)

Turnbull, G. (1726) *De pulcherrima mundi cum materialis tum rationalis constitutione (On the Beauty of the World in its Material and Rational Constitution)*, Aberdeen: Nicol.(Turnbull's formulation of the argument from design in response to the deists' critique of Christianity.)

Turnbull, G. (1731) *A Philosophical Enquiry concerning the Connexion betwixt the Doctrines and Miracles of Jesus Christ*, London: Willock; rev. edns 1732, 1739.(Ostensibly written in 1726, this pamphlet argues that Christ's miracles provided 'experimental' proofs for the central doctrines of Christianity.)

Turnbull, G. (1732) *Christianity neither False nor Useless, tho' not as Old as the Creation*, London: Willock.(A

defence of Samuel Clarke’s Boyle Lectures against the criticisms advanced in Matthew Tindal’s *Christianity as Old as the Creation* (1730).)

Turnbull, G. (1740a) *The Principles of Moral Philosophy*, London: Noon, 2 vols; repr. Hildesheim: Olms, 1976. (Based on Turnbull’s moral philosophy lectures at Marischal College, and incorporating ideas taken from Shaftesbury, Hutcheson, Butler, Berkeley, Pope, Newton and John Clark among others.)

Turnbull, G. (1740b) *A Treatise on Ancient Painting*, London: Millar; repr., with intro. by V. Bevilacqua, Munich: Fink, 1971.(The reprint does not contain the lavish plates of the original, which Turnbull also published separately in 1741.)

Turnbull, G. (1740c) *An Impartial Enquiry into the Moral Character of Jesus Christ; Wherein he is Considered as a Philosopher. In a Letter to a Friend*, London: Roberts; 2nd edn, 1743.(Published anonymously, this pamphlet argues that Christ was a greater philosopher than Socrates, and that the purity of his morals demonstrates the truth of the essentials of Christianity.)

Turnbull, G. (trans.) (1741) *A Methodical System of Universal Law*, by J.G. Heineccius, London: Noon, 2 vols. (Includes Turnbull’s important dissertation ‘A Discourse upon the Nature and Origine of Moral and Civil Laws’.)

Turnbull, G. (1742) *Observations upon Liberal Education*, London: Millar.(Turnbull’s most comprehensive statement of his pedagogical ideals, which blends together themes from the writings of Locke, Shaftesbury, Molesworth, Rollin, Cicero and Plato.)

References and further reading

Gibson-Wood, C. (1987) ‘Painting as Philosophy: George Turnbull’s *Treatise on Ancient Painting*’, in J.J. Carter and J.H. Pittock (eds) *Aberdeen and the Enlightenment*, Aberdeen: Aberdeen University Press, 189-98.(Draws out the subversive implications of Turnbull’s view of the didactic functions of the fine arts for neoclassical art theory.)

Mackinnon, K.A.B. (1987) ‘George Turnbull’s Common Sense Jurisprudence’, in J.J. Carter and J.H. Pittock (eds) *Aberdeen and the Enlightenment*, Aberdeen: Aberdeen University Press, 104-10.(A brief survey of the main themes of Turnbull’s science of morals.)

Norton, D.F. (1975) ‘George Turnbull and the Furniture of the Mind’, *Journal of the History of Ideas* 35: 701-16. (The only substantial treatment of Turnbull’s epistemology.)

Stewart, M.A. (1987) ‘George Turnbull and Educational Reform’, in J.J. Carter and J.H. Pittock (eds) *Aberdeen and the Enlightenment*, Aberdeen: Aberdeen University Press, 95-103.(A careful reconstruction of Turnbull’s career.)

Wood, P.B. (1993) *The Aberdeen Enlightenment: The Arts Curriculum in the Eighteenth Century*, Aberdeen: Aberdeen University Press.(Discusses Turnbull’s teaching career at Marischal College and his impact on the Aberdeen philosophical tradition.)

Twardowski, Kazimierz (1866-1938)

Twardowski, one of the most distinguished of Brentano's students, became famous for his distinction between the content and object of presentations. Twardowski, after his appointment as a professor of philosophy at the University of Lwów (Lvov), considerably limited his own philosophical research for the sake of teaching activities. He set himself an ambitious task: to create a scientific philosophy in Poland. Twardowski fully realized his aim, giving the first step towards the so-called Lwów-Warsaw School, a group of philosophers working in analytic philosophy - in particular, logic, philosophy of science, and philosophy of language. In spite of his concentration on teaching, Twardowski also made remarkable contributions to philosophy after coming to Lwów.

1 Life and teaching activity

Kazimierz Twardowski was born in Vienna on 10 October 1866. He was educated in a famous Viennese secondary school, *Theresianum*. In 1885-9 he studied philosophy at Vienna University, mainly under Franz Brentano. He obtained his Ph.D. in 1892 for a thesis on the relation between ideas and perception in Descartes. Two years later Twardowski obtained his *Habilitation* on the basis of a dissertation on the content and object of presentations. Having *venia legendi*, Twardowski started lecturing at the University of Vienna as *Privatdozent*. In 1895 he was appointed Professor of Philosophy at the University of Lwów. Twardowski came to Lwów with the definite aim of establishing a scientific philosophical school in Poland. Indeed he subordinated all his academic activities to the realization of this goal, which he regarded as an important national duty; when Twardowski came to Lwów, Poland was still partitioned between Germany, Russia and the Austro-Hungarian Empire. Twardowski spent the rest of his life in Lwów, where he died on 11 February 1938.

The results of Twardowski's activities as a teacher were enormous. Almost all important Polish philosophers of the first half of the twentieth century were his students, in particular Kazimierz Ajdukiewicz, Tadeusz Kotarbiński, Stanisław Leśniewski, Jan Łukasiewicz and, briefly, Roman Ingarden. Twardowski's teaching activity gave rise to the Lwów-Warsaw School, the most important school in the history of Polish philosophy.

2 The conception of philosophy

Following Brentano, Twardowski embraced a conception of philosophy as a science based on descriptive psychology. For Twardowski, scientific philosophy should avoid hopeless general problems, like the question whether the world is material or spiritual, and concentrate on particular problems which might be properly analysed. Twardowski regarded most traditional philosophical questions, particularly in metaphysics, as matters of faith and worldview rather than proper science. He strongly insisted that philosophy must begin with careful and precise clarifications, because otherwise it is simply impossible to understand what is going on in philosophical questions and discussions. Thus, a clear statement of a given question is the first duty of philosophers. Twardowski's general metaphilosophical claims decisively favoured some philosophical fields, like logic, semantics (in the broad sense: that is, including syntax, semantics as a theory of reference, and pragmatics), epistemology, and philosophy of science. This is why his students undertook their basic research in those areas of philosophy.

3 The content-object distinction

Twardowski adopted the main theses of Brentano's descriptive psychology, in particular his conception of the psychic as essentially intentional and his distinction between mental acts and their objects. Twardowski's main contribution to the Brentanist philosophical tradition consists in his having added to the act-object distinction a third ingredient: the content. In his *Habilitation* dissertation Twardowski gave a detailed analysis of the content-object distinction in the case of presentations (see Intentionality).

Twardowski offers three direct arguments for introducing contents of presentations: (a) contents and objects cannot be identical, because true negative judgments which deny the existence of objects are based on presentations; their contents exist, their objects do not; (b) contents have properties which objects lack, and conversely objects have properties which contents lack; for example, redness as a property of an object cannot be attributed to the content of this object; (c) there are non-equivalent contents of presentations which refer to the same objects; for example, 'the city located at the site of Roman Juvavum', and 'the birthplace of Mozart'.

Twardowski observed that there is a close analogy between psychology and semantics. In particular, names and presentations are parallel, because names (a) make known acts of presentation which occur in the speaker, (b) arouse mental contents in the listener (for Twardowski, mental contents are meanings of names), (c) designate objects which are presented by meanings of names. This parallelism played a significant heuristic role in Twardowski's distinction. However, perhaps its more important function was that it led Twardowski to employ semantic analysis in descriptive psychology. Thus, Twardowski noted that when we use the expression 'a presented object', the adjective 'presented' may function as a determiner or as a modifier. This ambiguity is responsible for a confusion between contents and objects, in particular for the idealistic thesis that contents are genuine objects of our presentations. This observation led Twardowski to the realistic account of intentionality: our mental acts are usually directed to ordinary objects.

Twardowski extended his psychological and semantic analysis by important ontological considerations. Contrary to Bolzano, he defended the thesis that there are no objectless presentations. He also outlined a mereological theory of objects of presentations.

Twardowski's *Habilitation* dissertation had considerable historical significance. It influenced Meinong, in his general theory of objects, and Husserl's theory of intentionality. However, Husserl criticized Twardowski for psychologism and unclarities in the nature of content. Although the objection of psychologism was fully justified and recognized by Twardowski himself, Husserl's accusation that Twardowski fell into the picture theory of content is a misunderstanding.

4 Other contributions to philosophy

As already noted, Twardowski agreed that his account in his *Habilitation* dissertation was too psychologistic. Later he proposed an important distinction between actions and products. Twardowski begins by pointing out that there are characteristic pairs in which the first word is a verb (for example, sing, command) or a gerund (singing, commanding) and the second is a related noun (song, command). The first word refers to an action and the second to the product of that action. Products may be non-durable, for example the song which terminates when singing stops, or durable, for example the painting as a product of painting. Twardowski was mainly interested in durable psychophysical products. For Twardowski, such products express mental states and become their signs. In this sense, expressed mental states may be considered as meanings of the given signs. This scheme enables us to draw a distinction between logic and psychology. Psychology is primarily interested in actions, but logic in their durable products and their formal connections. Moreover, there is an obvious linkage between both fields, because meanings are generated by actions.

Twardowski strongly defended the view that there are no relative truths: if something is true, it is absolutely true. He argued that typical arguments for the existence of relative truths are based on simple confusions. For example, if one says that 'today it is raining' is an example of a relative truth, one forgets that this phrase expresses only an incomplete proposition. Once it is made complete (by indicating spatiotemporal coordinates), it becomes absolutely true. Also empirical hypotheses do not provide examples of relative truths. Twardowski points out that if we accept hypotheses as only probable, this means only that we do not know their actual logical values. If it turns out that a given hypothesis is at variance with fact, we consider it as false from the very beginning, and not as a proposition which changes its logical value.

See also: Poland, philosophy in

JAN WOLEŃSKI

List of works

Twardowski, K. (1892) *Idee und Perzeption: Eine erkenntnistheoretische Untersuchung aus Descartes*, Vienna: Konegen.(Twardowski's Ph.D. dissertation.)

Twardowski, K. (1894) *Zur Lehre vom Inhalt und Gegestand der Vorstellungen: Eine psychologische Untersuchung*, Vienna: Hölder; 2nd edn, Munich: Philosophia-Verlag, 1982; trans. H. Grossmann, *On the Concept and Object of Presentations*, The Hague: Nijhoff, 1976.(Twardowski's *Habilitation* dissertation.)

Twardowski, K. (1998) *Selected Papers*, eds J. Brandl and J. Wolenski, Amsterdam: Rodopi.(This collection contains Twardowski's most important papers.)

References and further reading

Cavallin, J. (1990) *Content and Object: Husserl, Twardowski and Psychologism*, Stockholm: Department of Philosophy, University of Stockholm.(A detailed presentation of Twardowski's content-object distinction and Husserl's criticism of it; the book contains valuable bibliographical information.)

Smith, B. (1989) 'Kasimir Twardowski: An Essay on the Borderlines of Ontology, Psychology and Logic', in K. Szaniawski (ed.) *The Vienna Circle and the Lvov-Warsaw School*, Dordrecht: Kluwer, 312-73.(An extensive treatment of Twardowski's ontological views.)

Woleński, J. (1989) *Logic and Philosophy in the Lvov-Warsaw School*, Dordrecht: Kluwer.(Chapter II is devoted to Twardowski.)

Type/token distinction

The type/token distinction is related to that between universals and particulars. C.S. Peirce introduced the terms 'type' and 'token', and illustrated the distinction by pointing to two senses of 'word': in one, there is only one word 'the' in the English language; in the other, there are numerous words 'the' on the physical page you are now looking at. The latter are spatiotemporal objects composed of ink; they are said to be word tokens of the former, which is said to be the word type and is abstract. Phonemes, letters and sentences also come in types and tokens.

The distinction between 'type' and 'token', or something very like it, seems to be applicable beyond language to, for example, Beethoven's *Fifth Symphony* and performances of it, the grizzly bear and specimens of it, the Kentucky Derby and runnings of it, and the bubonic plague and outbreaks of it. The type/token distinction is important to linguistics, logic, aesthetics and philosophy of science. In philosophy of mind it is critical to distinguish types of events/states/processes from tokens of them, because those who identify mental events with physical events agree that every mental event token is a physical event token, but divide over whether mental event types, for example, pain, are identical to physical event types, for example, C-fibre stimulation (see Mind, identity theory of). Just what types are, whether they exist, whether they are 'present in' their tokens and whether all types have tokens are matters of controversy.

To see that we often refer to types, consider the grizzly bear. At one time its US range was most of the west and it numbered 10,000 in California alone. At the end of the twentieth century its range is Montana, Wyoming and Idaho, and it numbers fewer than 1,000. But no particular flesh and blood bear numbers 1,000 or once had a range comprising most of the western USA; if anything, it is a type which does. Similarly, Old Glory had twenty-nine stars in 1846 but fifty in 1996, whether or not any particular American flag underwent such a transformation; Old Glory is also a type.

This last example points to a related notion, that of an 'occurrence'. The stars on Old Glory number fifty; but fifty star *types* or fifty star *tokens*? In fact it cannot be either: not star types, because all the stars on the flag are of the same (five-pointed) type; nor star tokens, because tokens are concrete and the flag in question is abstract and cannot contain concrete parts. Old Glory contains fifty occurrences of the (five-pointed) star. Similarly, the letter '*x*' (the very same letter type) occurs three times in the formula (type) ' $(\forall x)(Fx \rightarrow Gx)$'. Thus the notion of an occurrence of '*x*' must not be confused with the notion of a token of '*x*' (although it often is so confused). The notion of 'an occurrence of *x* in *y*' involves not only *x* and *y*, but also how *x* is situated in *y*. If we think of a formula as a sequence, then the air of mystery over how the same identical thing can occur twice vanishes. Even concrete particulars can occur more than once (for example, the same person occurs twice in the sequence of New Jersey million dollar lottery winners).

Tokens are particulars. Are types universals? They are if having instances makes something a universal. But in many ordinary and theoretical contexts, terms which refer to types function not as predicates but as singular terms, in sentences that permit existential generalization. ('The grizzly bear ranged over most of the west' entails 'Something ranged over most of the west', for example.) As values of bound first-order variables, they meet a Quinean necessary condition for existence, and thus may also be viewed as particulars (see Ontological commitment §1).

Do types exist? If types are universals, then the debate over whether they exist goes back at least to Plato. Still, as the preceding paragraph noted, types seem to have an existential advantage over more traditional universals. (Quine held that expression types exist, though not more traditional universals (1987).) Nominalists who deny that there are any abstract objects argue that this is an illusion; that talk about types is just shorthand for talk about tokens. The matter cannot be resolved here, but it may be useful to examine one popular nominalistic account, namely, that every reference to a type (for example, 'The grizzly bear is ferocious') can be replaced without change of information by quantification over all tokens (for example, 'Every grizzly bear is ferocious'). One problem for this account is the existence of numerous counterexamples. 'The grizzly bear ranged over most of the west', for example, cannot be analysed as 'Every grizzly bear ranged over most of the west'. (Perhaps it can be analysed in some other 'type-free' way, but the point is that the nominalistic account under consideration cannot be the systematic one needed to eliminate all the many references to types which occur in our ordinary and

scientific talk.)

A second problem is created by the implication that all types have tokens. Peirce (1931-58) claimed they do, but many thinkers since have denied this. Chomsky (1957), for example, claimed that there are sentences that never have been or will be instantiated - infinitely many. A third problem arises when we try to find a replacement for even a simple true sentence about, say, the word ‘one’, given the myriad senses it has, and the myriad forms its tokens may take: printed in ink in any number of fonts and handwritings, raised in Braille, incised in marble, existing briefly as so many pixels of light on a computer screen, electronic strings of dots and dashes, smoke signals, hand signals, individual air disturbances produced by human vocal cords, electromagnetic pulses on phone lines, pronounced (and mispronounced) in countless ways in countless accents. Even an (appropriately surrounded) empty space can be a token of ‘one’. About the only thing all and only tokens of ‘one’ clearly have in common is being tokens of ‘one’ - the type, that is; if that word (type) does not exist, then they would seem to have nothing in common and hence no appropriate quantification could even be formulated.

See also: Abstract objects; Logical and mathematical terms, glossary ofUniversals

LINDA WETZEL

References and further reading

Block, N. (ed.) (1980) *Readings in Philosophy of Psychology*, Cambridge, MA: Harvard University Press, vol.1. (Includes numerous important articles in philosophy of mind pertinent to the debate between token materialists and type materialists.)

Bromberger, S. (1992) ‘Types and Tokens in Linguistics’, in *On What We Know We Don’t Know*, Chicago, IL: University of Chicago Press.(A fine attempt to explain the relationship between types and tokens, in order to explain how linguistics is an empirical science although about abstract types.)

Chomsky, N. (1957) *Syntactic Structures*, The Hague: Mouton.(The classic attack on structuralism which started the Chomskian revolution in linguistics.)

Goodman, N. (1951) *The Structure of Appearance*, Dordrecht: Reidel, 3rd edn, 1977. (Although not a book about types, Goodman’s attempt to do without them produces many instructive passages.)

Goodman, N. and Quine, W.V. (1947) ‘Steps Toward a Constructive Nominalism’, *Journal of Symbolic Logic* 12: 105-22; repr. in *Problems and Projects*, Indianapolis, IN: Bobbs-Merrill, 1972.(A nominalist attempt at proof theory without expression types.)

Hale, R.J. (1987) *Abstract Objects*, Oxford and New York: Blackwell.(An excellent discussion about abstract objects generally, including types, and our reasons for being justified in concluding they exist.)

Hutton, C. (1990) *Abstraction and Instance: The Type-Token Relation in Linguistic Theory*, Oxford: Pergamon. (By means of an impressive survey of the relevant literature, Hutton argues that linguistic theories rely uncritically on the type/token distinction (and on related notions such as that of *langue-parole* and competence-performance) and that all such accounts are problematic.)

Katz, J.J. (1981) *Language and Other Abstract Objects*, Totowa, NJ: Rowman & Littlefield.(A robust defence of languages and sentences as abstract objects.)

Peirce, C.S. (1931-58) *Collected Papers of Charles Sanders Peirce*, ed. C. Hartshorne and P. Weiss, Cambridge, MA: Harvard University Press, vol. 4, para. 537; vol. 8, para. 334; vol. 2, para. 243-6.(The eight-volume collection of Peirce’s papers.)

Quine, W.V. (1987) *Quiddities: An Intermittently Philosophical Dictionary*, Cambridge, MA: Harvard University Press, 216-19.(This is Quine distilled down to eighty-three alphabetically arranged entries, including a two-page entry on ‘Type versus Token’.)

Wetzel, L. (1993) ‘What Are Occurrences of Expressions?’, *Journal of Philosophical Logic* 22: 215-20.(Provides an answer to the title question.)

Wollheim, R. (1968) *Art and Its Objects*, New York: Harper & Row.(Argues that many works of art cannot be physical objects/events or classes thereof, but are types that none the less are ‘present in’ their tokens.)

Wolterstorff, N. (1975) ‘Toward an Ontology of Art Works’, *Noûs* 9: 115-42. (Wolterstorff argues that many works of art are types, or ‘kinds’, and proposes a suitable analysis of the notion of a kind.)

Udayana (11th century)

*Perhaps the most important philosopher of the Nyāya school, Udayana authored several works in the eleventh century which brought to a close the long-standing debate between Nyāya and Buddhist philosophers. The realist Nyāya philosophers had argued for the existence of an enduring self (*ātman*), a thesis denied by their Buddhist opponents. Such was the importance of this disagreement that it pervaded all other areas of philosophical contention between them. In the Ātmatattvaviveka (On the Discrimination of the Reality of the Self), Udayana systematically clarified the connections between the* ātman *debate and many other areas of philosophical dispute, with the result that, in defending* ātman, *he also produced a masterly defence of Nyāya realism. Udayana is also credited with giving the definitive defence of theism in the Nyāyakusumāñjali (A Handful of Nyāya-Tree Flowers).*

Tradition has it that Udayana came from Mithila in north-east India, a centre of great learning, especially for the Nyāya school. His life is shrouded in mythical tales of his brilliance. There is no question, however, about his pre-eminent place in the evolution of Nyāya thought. Udayana was the last of the Nyāya philosophers to enjoy the stimulus of a powerful Buddhist opposition, for it was shortly after his death that Buddhism began to disappear from India. After Buddhism's demise, Nyāya thought became characterized by increasing subtlety of argument and a greater preoccupation with method than with content. Although Navya-Nyāya (New Nyāya), as the school in its later period is called, is usually said to date from the fourteenth century, the subtlety of Udayana's work is such that some critics regard him as its real source. Certainly Udayana's style and rigour look forward to Navya-Nyāya, but at the same time his work brings to fruition the arguments that traditional Nyāya philosophers had developed in their efforts to defeat the Buddhists. While no one would want to credit the demise of Buddhism in India to Udayana's polemics, the quality of the philosophical exchange between him and his Buddhist opponent, the monk Jñānaśrīmitra, represents one of the greatest achievements of Indian philosophy.

In the *Ātmatattvaviveka (On the Discrimination of the Reality of the Self)*, Udayana states that there are four main theses which the Buddhists give to refute the existence of *ātman* (Self): (1) all existent things are momentary; (2) there are no objects external to our judgments; (3) there is no difference between an object and its qualities; and (4) the *ātman* is not perceived. Apart from the last thesis, it can be seen that the Buddhist denial of *ātman* is a consequence of a more general argument about the nature of reality (see Momentariness, Buddhist doctrine of). Udayana concentrates his arguments against these general theses rather than arguing about *ātman* in particular, and it is this that gives the work its philosophical breadth and interest.

For example, the thesis of universal momentariness not only rules out the possibility of an enduring *ātman*, but is also the antithesis of Nyāya's realist view of the world and becomes a denial of the enduring existence of everything, from the *ātman* to the kitchen pot. Thus Udayana denies any concomitance between existence and momentariness and attacks the related notion of causal efficacy. It is only when causal efficacy is understood to mean the immediate production of an effect that the concomitance between existence and momentariness is established. Jñānaśrīmitra states that an enduring entity is incapable of causal efficacy and so lacks existence. The relationship between a dormant seed in the granary and a sprouting seed in the field is the locus for this debate. For the Buddhist argument to succeed, the nature of causal efficacy must be such that it creates a difference in identity between the nonsprouting granary seed and the sprouting field seed. Udayana argues for an interpretation of causal efficacy which allows for one and the same object to be causally efficacious at one time and not at another, thus allowing for the continuity of the granary seed with the field seed. He argues that to be causally potent does not mean that a cause must produce its effect immediately, but rather that it has the potential to do so given the presence of the right accessories - in the case of the seed, such things as sunshine and water. Udayana notes that we call seeds in the granary 'seeds' precisely because of their potential to produce sprouts. He charges that the Buddhists are not even entitled to call the seeds in the granary 'seeds' if, as they claim, those seeds are not responsible for the final production of sprouts. According to Udayana, the capacity or disposition of a seed to produce a sprout ultimately depends on the universal 'seedness' inhering in each individual seed. The presence of this universal along with the necessary accessories allows a nonproductive seed to become productive. The Buddhist position represents a lapse into chaos, where sprouts can be produced without seeds, and seeds are deprived of their very nature. Udayana's detailed critique of the Buddhist understanding of causality was accepted as definitive by later Nyāya philosophers (see Causation, Indian theories of §§5-6).

Although real universals were essential to Udayana’s attack on the Buddhist account of causality, and although he was a severe critic of Buddhist nominalism, he did recognize that not all general terms represent universals that have an independent reality (see Nominalism, Buddhist doctrine of §§1-2; Universals, Indian theories of §§2-3). Uddyotakara had distingushed real universals, such as ‘cowness’, from those constructed by us, such as the property of being a cook. In the *Kiraṇāvalī(Garland of Rays)*, Udayana contributed to the debate by giving a list of formal conditions which distinguish real universals (*jāti*) from those that are convenient devices of our own construction (*upādhi*). Some contemporary philosophers of language have noted an affinity between Udayana’s *jāti* and what are now called ‘natural kinds’.

Despite Udayana’s obvious flair for rigorous philosophical debate, he did not abandon the traditional soteriological concerns of philosophy. In the *Nyāyakusumāñjali (A Handful of Nyāya-Tree Flowers)*, he defends the theism of his school by developing a rational theology, using arguments that would be familiar to Western theologians. We infer the existence of God, for example, from our experience of order in the world, just as we infer the existence of a potter from the presence of a pot. For Udayana, the realism of the Nyāya school provided the most coherent support for *ātman* and, in a remarkable passage in the *Ātmatattvaviveka*, he places Nyāya at the pinnacle of philosophical thought. In accordance with the Nyāya system, he emphasizes that liberation cannot just be a matter of transcendental experience, but must incorporate correct philosophical knowledge.

See also: Nyāya-Vaiśeṣika

JOY LAINE

List of works

A great deal of Udayana’s work remains to be translated from the Sanskrit. His works are listed below in chronological order. These are the Sanskrit texts unless indicated otherwise.

Udayana (11th century) *Lakṣaṇāvalī(The Garland of Definitions)*, ed.G. Jha, Darbhanga: Mithila Institute Series 13, 1963; trans. M. Tachikawa, in *The Structure of the World in Udayana's Realism*, Dordrecht: Reidel, 1981. (A manuscript on the categories of the Vaiśeṣika school.)

Udayana (11th century) *Lakṣaṇamāla (The Wreath of Definitions)*, ed. G. Jha, Darbhanga: Mithila Institute Series 14, 1964.(A manual on the categories of the Nyāya school.)

Udayana (11th century) *Ātmatattvaviveka (On the Discrimination of the Reality of the Self)*, ed. D. Sastri, Varanasi: Chowkhamba, 1940.(A comprehensive survey of the Buddhist arguments against the existence of the self or *ātman*. In countering these arguments, Udayana also gives a defence of the realism of the Nyāya school.)

Udayana (11th century) *Nyāyapariśiṣṭa (The Nyāya Appendix)*, ed. N.C. Vedantatirtha, Calcutta: Calcutta Sanskrit Series, 1938.(A commentary on the fifth chapter of the *Nyāyasūtra*, the foundational text of the Nyāya school.)

Udayana (11th century) *Nyāyavārttika-tātparyaṭīkā-pariśuddhi (An Exoneration of the Nyāyavārttika-tātparyaṭīkā)*, ed. L.S. Dravid and V.P. Dvivedin, Calcutta: Bibliotheca Indica, 1911; trans. in part in Jha, G., *The Nyāya-Sūtras of Gautama with the Bhāṣya of Vātsyāyana and the Vārtika of Uddyotakara*, Delhi: Motilal Banarsidass, 1984. (A partially extant work which continues with the tradition of building on previous subcommentaries on the *Nyāyasūtra*, in this case Vācaspati Miśra’s commentary, the *Nyāyavārttika-tātparyaṭīkā*, which is itself a commentary on Uddyotakara’s *Nyāyavārttika*.)

Udayana (11th century) *Nyāyakusumāñjali (A Handful of Nyāya-Tree Flowers)*, ed. L.S. Dravid and N.C. Vedantatirtha, Calcutta: Kashi Sanskrit Series, 1912; Books 1 and 2 trans. Ravitirtha, *Adyar Library Bulletin* 5 (1941), 6 (1942), 7 (1943), 10 (1946).(Considered by some to be Udayana’s greatest work, in which he gives a definitive defence of the Nyāya conception of God.)

Udayana (11th century) *Kiraṇāvalī(The Garland of Rays)*, ed. Sivachandra Sarvabhauma and N.C. Vedantatirtha, Calcutta: Bibliotheca Indica, 1956.(A commentary on Praśastapāda’s *Padārthadharmasaōgraha*, an early, foundational work on the Vaiśeṣika system. Portions of the *Kiraṇāvalī* are translated in Tachikawa 1981.)

References and further reading

Dravid, R. (1971) *The Problem of Universals in Indian Philosophy*, Varanasi: Motilal Banarsidass.(Chapter 11 of this survey focuses on Udayana’s arguments against theBuddhist theory of meaning as presented by Jñānaśrīmitra and Ratnakīrti.)

Matilal, B.K. (1977) *A History of Indian Literature*, vol. 6, *Scientific and Technical Literature*, Wiesbaden: Otto

Harrassowitz.(Part 3, fasc. 2 is about Nyāya-Vaiśeṣika, and contains a translation of a small portion of the *Ātmatattvaviveka*.)

Matilal, B.K. (1986) *Perception: An Essay on Classical Indian Theories of Knowledge*, Oxford: Clarendon Press. (In this comprehensive work describing the epistemological differences between the Nyāya realists and their Buddhist opponents there are frequent references to Udayana's contributions.)

Phillips, S.H. (1995) *Classical Indian Metaphysics*, La Salle, IL: Open Court.(Although the focus of this book is on the debate between the New Nyāya school and the Advaitin philosopher Śrīharṣa in the period immediately after Udayana, it does contain frequent references to him, and one chapter is devoted to the views of traditional Nyāya philosophers up to and including Udayana.)

Potter, K.H. (ed.) (1977) *Encyclopedia of Indian Philosophies*, vol. 2, *Nyāya Vaiśeṣika*, Delhi: Motilal Banarsidass.(A comprehensive introduction to the Nyāya-Vaiśeṣika system, containing detailed summaries of all known Nyāya-Vaiśeṣika works.)

Sastri, D.N. (1964) *The Philosophy of Nyāya-Vaiśeṣika and its Conflict with the Buddhist Dignāga School*, Delhi: Bharatiya Vidya Prakashan.(A detailed account of the defence of Nyāya realism in the face of Buddhist opposition. Although the focus is on philosophers of an earlier period, there are frequent references to Udayana and his role in this philosophical debate.)

Sharma, D. (1969) *The Differentiation Theory of Meaning in Indian Logic*, The Hague: Mouton.(Sharma introduces his translation of Ratnakīrti's *Apohasiddhi* by describing the differences separating Nyāya and Buddhist theories of meaning.)

Tachikawa, M. (1981) *The Structure of the World in Udayana's Realism*, Dordrecht: Reidel.(Contains translations of the *Lakṣaṇāvalī* and the *Kiraṇāvalī*.)

Uddyotakara (6th century)

Uddyotakara, a philosopher of the Nyāya school, wrote the Nyāyavārttika, a lengthy commentary on the Nyāyasūtra. His most urgent task was to re-establish the authority of the Nyāya school in the face of extensive criticism from the great Buddhist logician Dignāga. Dignāga had been particularly critical of the logical work of Vātsyāyana, Uddyotakara's predecessor. In response, Uddyotakara incorporated Dignāga's logical work into the Nyāya school, and added his own interpretation. He was less receptive to Dignāga's other views, especially his account of perception and its relation to language.

Uddyotakara ('The Enlightener') was sometimes referred to as a Pāśupata teacher, indicating membership of a particular sect of devotees of Śiva. It is not therefore surprising to find in him a strong proponent of theism. Certainly since Vātsyāyana, Nyāya contained an element of theism inasmuch as God was recognized as a substantial Self distinct from other selves. Uddyotakara, however, gives a much fuller articulation of the Nyāya conception of God as creator, as omnipotent and as being beyond *dharma*. His arguments in support of such a God became an important aspect of the Nyāya tradition. Udayana's classic defence of theism, in his *Nyāyakusumāñjali (A Handful of Nyāya-Tree Flowers),* was the culmination of Uddyotakara's efforts in this vein.

Uddyotakara's importance for the Nyāya school, however, lies in his response to Dignāga. A skilled logician himself, he fully appreciated Dignāga's work. Dignāga had grasped the formal character of inference and had formulated three conditions under which the reason given for a thesis would result in a correct inference. Uddyotakara's acceptance of his opponent's analysis of logical inference was mitigated by his criticisms of its details. He demonstrated that Dignāga's three conditions actually mask sixteen possible ways in which the reason could be combined with the other members of the inference. Thus a more precise analysis would be needed to demarcate correct and incorrect inferences (see Inference, Indian theories of §§5-6).

In other areas Uddyotakara asserted his differences with Dignāga more vehemently. Unlike Nāgārjuna, Vātsyāyana's opponent, Dignāga offered a clearly articulated metaphysics in opposition to Nyāya realism. In his system there is a radical break between language and the world. The world is made up of unique momentary particulars (*svalakṣaṇas*) whose fleeting existence can be grasped only by perception. The judgments that we make about the world and express in language create a set of fictional categories, such as physical objects and universals, which are a result of our mental construction. Uddyotakara opposed this view by reaffirming and expanding Vātsyāyana's arguments for the perceptual basis of physical objects (see Vātsyāyana). He also initiated the Nyāya opposition to the Buddhist theory of meaning, known as *apohavāda*, which originated with Dignāga. Dignāga had avoided all commitment to any real basis for the terms of our language, such as universals, by proposing that words derive their meaning and usefulness on the basis of a common exclusion, rather than any set of properties held in common.

Since the *Nyāyasūtra* had given a definition of perception as being that which is not expressible, it might seem that Dignāga's separation of perception from language was in accord with them. Uddyotakara's commentary reaffirms Vātsyāyana's interpretation of this *sūtra* (1.1.4). Like Vātsyāyana, he states that the purpose of defining perception in this way is to stress that, since perception comes about as a result of contact between object and sense organ, no linguistic skills are necessary for its occurrence. This does not mean, however, that when we acquire those skills and begin to make judgments about our perceptions we are superimposing mental constructs, which have no basis in reality.

Hence for Uddyotakara physical objects are perceptual 'firsts' and he produces a dazzling array of counterarguments to defend Nyāya realism against Dignāga's offensive. As with Vātsyāyana, the recurring theme is that without the acceptance of a whole distinct from its component parts, we cannot explain how we arrive at the notion of an object such as a tree (see Sense perception, Indian views of §5). An additional argument which Uddyotakara uses concerns the relationship between vision and touch. For the Buddhist, there is no single entity apprehended by both vision and touch, so that we do not see the very same jar that we touch. Our conception of a jar arises from grouping a particular set of perceptual qualities together. Uddyotakara therefore argues from the premise that we do in fact have a cognition of a jar, and that only the Nyāya view of a substance as separate from

its qualities can explain this cognition that we all have. Udayana, in the *Ātmatattvaviveka (On the Discrimination of the Reality of the Self)*, was later to argue this same point in meticulous detail. Uddyotakara also addresses the difficult question of exactly how the whole is related to its parts. In answer to the question whether the whole resides in each of its parts partially or in its entirety, Uddyotakara replies that neither alternative is applicable. He resists all attempts to pull the whole apart and compares it to other entities, such as numbers, that are indivisible yet can be present in several things.

Uddyotakara employs a very similar argument when explaining the relationship between a universal (such as ‘cowness’) and the individual objects in which the universal is manifested (in this case, individual cows). As in the above argument, Uddyotakara deflects the question as to whether the universal is present in the individual in its entirety or in part. Universals are simple entities not subject to partition in this way, and they are related to the appropriate individuals by the relation of ‘inherence’. Despite Uddyotakara’s efforts, there is no doubt that the notion of real universals remains problematic for the Nyāya system. It might seem puzzling that the Naiyāyikas clung to it, especially given Uddyotakara’s recognition that in many instances we can explain our use of a common noun without invoking real universals. It would be a mistake, Uddyotakara tells us, to conjure up the notion of ‘cookness’ to explain how we label a variety of individuals as cooks. People are cooks in virtue of the fact that they all perform the common action of cooking, not in virtue of some universal inhering in them. Uddyotakara demonstrates a sophisticated understanding of language in his recognition that no uniform account can be given of how general terms function in a language. Many contemporary philosophers of language would agree. The role of real universals, such as ‘cowness’, was crucial, however, in the Nyāya account of causality (see Udayana) and in their attack on the Buddhist negative theory of meaning (*apohavāda*).

See also: Nominalism, Buddhist doctrine of; Nyāya-Vaiśeṣika; Universals, Indian theories of §§2-3

JOY LAINE

List of works

Uddyotakara (6th century) *Nyāyavārttika*, in G. Jha (trans.) *The Nyāya-Sūtras of Gautama with the Bhāṣya of Vātsyāyana and the Vārttika of Uddyotakara*, Delhi: Motilal Banarsidass, 4 vols, 1984.(This four-volume set gives a translation of Uddyotakara’s *vārttika*, which can be read alongside the *sūtra* and the *bhaṣya*.)

References and further reading

Matilal, B.K. (1971) *Epistemology, Logic and Grammar in Indian Philosophical Analysis*, The Hague: Mouton.(A detailed account of the differences between the Nyāya and Buddhist schools, with particular references to theories of perception and language. Contains frequent references to Uddyotakara.)

Matilal, B.K. (1977) *A History of Indian Literature*, vol. 6, *Scientific and Technical Literature*, Wiesbaden: Otto Harrassowitz. (Part 3, fasc. 2 deals with Nyāya-Vaiśeṣika.)

Nyāyasūtra (*c*.400 AD) trans. G. Jha, *The Nyāya-Sūtras of Gautama with the Bhāṣya of Vātsyāyana and the Vārttika of Uddyotakara*, Delhi: Motilal Banarsidass, 4 vols, 1984.(Complete translation with detailed notes.)

Phillips, S.H. (1995) *Classical Indian Metaphysics*, La Salle, IL: Open Court.(Although this study of Indian metaphysics focuses on the philosophers of a later period of the Nyāya school, there is a chapter specifically about the earlier Nyāya philosophers, including Uddyotakara.)

Potter, K.H. (ed.) (1977) *Encyclopedia of Indian Philosophies*, vol. 2, *Nyāya Vaiśeṣika*, Delhi: Motilal Banarsidass.(A comprehensive introduction to the Nyāya-Vaiśeṣika system, containing detailed summaries of all known Nyāya-Vaiśeṣika works.)

Sastri, D.N. (1964) *The Philosophy of Nyāya-Vaiśeṣika and its Conflict with the Buddhist Dignāga School*, Delhi: Bharatiya Vidya Prakashan.(A detailed account of the defence of Nyāya realism in the face of Buddhist opposition. There are frequent references to Uddyotakara and his role in this long philosophical dispute.)

Ûisang (605-702)

Ûisang was the founder of the Korean branch of the Flower Garland (Hwaôm; in Chinese, Huayan) school of East Asian Buddhism which emerged as the main scholastic tradition within Korean Buddhism, thanks in large measure to Ûisang himself. His works emphasize the unimpeded interpenetration that Hwaôm posits to pertain between all phenomena in the universe.

Ûisang was ordained in the Silla kingdom of Korea, but soon decided to make a pilgrimage to the mecca of the Chinese mainland to study with Chinese teachers. After one unsuccessful attempt in 650, when he was repatriated to Silla on suspicion of espionage, Ûisang eventually arrived in Tang China in 661, where he became one of the principal disciples of the second patriarch of the Chinese Huayan school, Zhiyan (602-68). After his teacher's death he emerged as one of the two leaders of Zhiyan's congregation, along with Fazang (643-712). In 670, however, Ûisang learned of an imminent Chinese invasion of Korea and promptly returned to Korea to warn his king of the danger. The invasion forestalled, Ûisang was rewarded with munificent royal support and his Flower Garland School dominated Korean Buddhist scholasticism from that point onward.

Ûisang's works, though few in number, were extremely influential throughout East Asia and contributed much to establishing the doctrinal foundations for a distinctively Korean tradition of Buddhism. His vision of Buddhism is laid out in summary form in his principal work, the *Hwaôm ilsûng pôpkyedo (Diagram of the Flower Garland One-Vehicle Realm-of-Reality)*, a verse of 210 Chinese characters with accompanying autocommentary, written in 668 while he was still in China. The chart is a novel attempt to represent diagrammatically all the teachings of the *Avataṃsakasūtra (Flower Garland Scripture)*. Ûisang's poem is arranged in a wavelike form, the 'Ocean seal diagram' (*sagaramudrā-maṇḍala*), which represents the Flower Garland teaching of the 'six marks': universality and particularity, identity and difference, and integration and disintegration. The totality of the diagram represents the marks of universality, identity and integration, while its sinuous curves represent particularity, difference and disintegration. The one continuous line of the chart, upon which each individual logograph of the poem is superimposed, demonstrates the cardinal Flower Garland doctrine of the unimpeded interpenetration between all independent phenomena in the universe. The meanderings of the line of verse around the diagram's fifty-four corners symbolize the pilgrimage of the lad Sudhana to spiritual teachers, as recounted in the lengthy 'Entering the Realm of Reality' (*Gaṇḍavyūha*) chapter of the *Avataṃsakasūtra*. The fact that the snaking line of verse begins and ends at the same spot on the diagram illustrates the Hwaôm doctrine of soteriological interpenetration: the initial arousing of the thought of enlightenment (*bodhicittotpāda*), the inception of spiritual practice, is the same as complete, perfect enlightenment (*anuttarasamyaksaṃbodhi*), the consummation of that practice. Ûisang's autocommentary helps to elaborate the doctrinal implications of the diagram's visual form, creating a uniquely East Asian type of philosophical document.

See also: Buddhist philosophy, Chinese; Buddhist philosophy, Korean; Fazang

ROBERT E. BUSWELL, JR

List of works

Ûisang (625-702) Collected works, in *Han'guk Pulgyo chônsô (Collected Works of Korean Buddhism)*, ed. Han'guk Pulgyo chônsô p'yônch'an wiwônhoe, Seoul: Tongguk University Press, vol. 2, 1-9.(The standard Korean edition of Ûisang's works.)

Ûisang (668) *Hwaôm ilsûng pôpkyedo (Diagram of the Flower Garland One-Vehicle Realm-of-Reality)*, trans. S. Odin in *Process Metaphysics and Hua-yen Buddhism: A Critical Study of Cumulative Penetration vs. Interpenetration*, Albany, NY: State University of New York Press, 1982.(The diagram, along with its autocommentary, is translated in the appendix, though without annotation.)

References and further reading

Odin, S. (1982) *Process Metaphysics and Hua-yen Buddhism: A Critical Study of Cumulative Penetration vs. Interpenetration*, Albany, NY: State University of New York Press.(The expository section includes a comparative treatment of Ûisang and Whitehead.)

Ulrich of Strasbourg (*c.*1220/5-1277)

A Dominican theologian and philosopher and a student of Albert The Great, Ulrich was well known for a widely studied summa theologiae, *De summo bono (On the Supreme Good), which represents an advance over previous summae in plan and organization. Ulrich provides a rich synthesis of Christian Neoplatonic theology and mysticism by systematizing the Aristotelianized Neoplatonic philosophical theologies of Albert the Great, Pseudo-Dionysius' De divinis nominibus and the Liber de causis. He exercised a notable influence on the Rhineland mystics.*

Ulrich was a student of Albert the Great at Cologne between 1248 and 1254; his fellow students included Thomas Aquinas and Hugh of Strasbourg. Ulrich was both a popular lecturer in theology at Strasbourg and, during the last five years of his life, the influential provincial of the German province. Ulrich was an intimate friend of Albert and his regular walking companion. Their long, warm friendship is evident in extant letters.

During the thirteenth century, Albert was the great figure in the Dominican school at Cologne from which his scientific interests and philosophical theology were transmitted through his immediate students Ulrich and Hugh to later important figures such as Dietrich of Freiberg and to the Rhineland mystics Meister Eckhart, John Tauler and Henry Suso. The dominant philosophical influences in the school were Aristotle, Proclus, Avicenna (see Ibn Sina), al-Farabi, Augustine, the *Liber de causis* (see Liber de causis) and Pseudo-Dionysius the Areopagite. Ulrich is deeply indebted to Albert's complex synthesis of the Augustinian theology of William of Auxerre and William of Auvergne, which was enriched with Aristotle's precise terminology, and the Neoplatonic ontology of Proclus and Pseudo-Dionysius.

Ulrich's works included commentaries on Aristotle's *Meteorology* and on Peter Lombard's *Sentences*. His only extant work is the summa *De summo bono (On the Supreme Good)*, composed between 1262 and 1272. Probably completed in eight books, the surviving manuscripts end at Book 6 tractate 5 and fill more than 600 folios; publication of the text is still in progress. The importance of the treatise is evident from the large number of manuscripts circulating in the thirteenth and fourteenth centuries. Ulrich was widely considered one of the most important Dominican thinkers after Albert and Thomas.

The treatise is organized according to an architectonic scheme derived from Peter Lombard. Book 1 concerns theology as science of the highest good; Book 2 concerns the being of the highest good; Book 3 concerns the Trinity; Book 4 concerns the Father and creation; Book 5 concerns Christ; Book 6 (200 folios) concerns the Holy Spirit and its ethical workings in human psychology; Book 7 concerns the sacraments; and Book 8 concerns beatitude and the attainment of the highest good (see Good, theories of the §1). The central themes of Ulrich's philosophical theology are presented in Books 1, 2 and 4.

The pervasive influence on Ulrich of the Christian Neoplatonist Pseudo-Dionysius is evident in the first two books, which comprise a loose commentary on the latter's *De divinis nominibus (On the Divine Names)*. In Book 1 tractate 1, the Neoplatonic metaphysics of light lays the foundation for an account of how the mind can know God. We know that God exists and what God is by means of the light conveyed by the angelic intelligences from the divine nature itself, both of which illuminate the agent intellect, the natural image of God in us. The theory of illumination endows the mind with the capacity to understand God in three ways: first, the negation of finite and imperfect properties generates infinite perfections which are worthy of the divine nature (symbolic theology); second, God is identified as the cause and source of all perfections (positive theology); and third, all properties are transcended in God (mystical theology) (see God, concepts of).

Book 1 tractate 2 argues that theology must be a supernatural science since even our natural knowledge of God's effects is not based on the dark and defective human intellect. It is faith, therefore, which presents God as the object of this science, and theology which grasps scientifically what is believed: God in himself and in his attributes, and God as both the supreme cause or first principle of creation and the supreme end or salvational goal of all. With greater precision and clarity than many scholastics, Ulrich articulates first principles for theological inquiry. In an Aristotelian manner, he argues that the four absolute first principles of theology, from which all other truths can be known, are immediately self-evident, again by means of intellectual illumination: (1) God is the

highest truth and the source of all truth; (2) this primary truth is infallible and warrants unconditional faith; (3) God's spokespersons must be believed; and (4) since it is inspired by God, Scripture must be true. It is only on the basis of the certain knowledge of these rules that the articles of faith, which are not evident in themselves, can be established. Book 2 expounds the divine names, divine substance and the transcendentals.

Book 4 tractate 1 examines the First Principle and its relations with created beings, and offers a compendium of Ulrich's doctrines in metaphysics, natural theology and cosmogony. His metaphysics of being derives from the *Liber de causis*: that 'being is the first among created things'. This Neoplatonic formulation holds that being is the first expression of or 'outflow' (*emanatio*) from God; it is the first form and basis of all other forms. Since being is universal it is indefinable in itself; but since being is mixed with non-being, it differs from the pure absolute being of God (see Being). Ulrich stresses that God must be distinguished from this universal being in order to avoid pantheism. The two realities also differ fundamentally as regards unity and relations. Universal being is not absolutely simple since it stands in real relations to non-being and God, for example, whereas God's pure unity is not abrogated by the relations of created beings to him. The rest of Book 4 comprises an extensive discussion of the Aristotelian causes and categories (tractate 2) and of transcendent substances and their mode of knowledge (tractate 3).

Ulrich's neglected treatise is a rich synthesis of the Christianized version of Neoplatonic themes scattered throughout the voluminous writings of Albert the Great. These include the identity of being and form, the hierarchy of intellects (Avicenna) combined with the doctrine of angels (Pseudo-Dionysius), the emanation of beings and forms from God via the process of universal illumination, and the confluence of being and value in the supreme principle.

See also: Albert the Great; Liber de causis; Pseudo-Dionysius

JOHN BUSSANICH

List of works

Ulrich of Strasbourg (1262-72) *De summo bono (On the Supreme Good)*, Corpus philosophorum Teutonicum Medii Aevi, Book 1 ed. A. de Libera and B. Mojsisch, Book 2 tractates 1-4 ed. A. de Libera, Book 4 tractates 1, 2 and 7 ed. S. Pieperhoff, Hamburg: F. Meiner, 1987-9.(Earlier partial editions include Book 1, ed. J. Daguillon, *La 'Summa de bono'. Livre I*, Paris: Vrin, 1930; Book 4, tractate 1 in F.J. Lescoe, *God as First Principle in Ulrich of Strasbourg*, New York: Alba House, 1979, 145-242.)

Ulrich of Strasbourg (1272-7) Letters, in H. Finke (ed.) *Ungedruckte Dominikanerbriefe des 13. Jahrhunderts*, Paderborn: F. Schoningh, 1891, 78-104.(Edition of collected letters.)

References and further reading

Backes, I. (1975) *Die Christologie, Soteriologie und Mariologie des Ulrich von Strassburg: ein Beitrag zur Geistegeschichte des 13. Jahrhunderts (The Christology, Soteriology and Mariology of Ulrich of Strasbourg: A Contribution to Thirteenth-Century Intellectual History)*, Trier: Paulinus-Verlag.(The most extensive study of central topics in Ulrich's theology.)

Grabmann, M. (1926) 'Studien über Ulrich von Strassburg' (Studies on Ulrich of Strasbourg), in M. Grabmann, *Mittelalterliches Geistesleben: Abhandlungen zur Geschichte der Scholastik und Mystik*, Munich: Max Hueber, vol. 1, 147-221.(A comprehensive survey of Ulrich's life, the manuscripts of his treatise and a summary of its contents.)

Lescoe, F.J. (1979) *God as First Principle in Ulrich of Strasbourg*, New York: Alba House.(As well as the edited but untranslated text of Book 4 tractate 1, Lescoe provides a comprehensive discussion of the authenticity and contents of the entire treatise, a detailed study of the central metaphysical themes in Ulrich's thought, and an extensive bibliography which is especially useful for information on the many bits and pieces of Ulrich's treatise that have been edited and studied in unpublished doctoral dissertations.)

Putnam, C. (1961) 'Ulrich of Strasbourg and the Aristotelian Causes', in J.K. Ryan (ed.) *Studies in Philosophy and the History of Philosophy* vol. I, 139-59.(Useful account of Ulrich's adaptation of Aristotelian causal theory within a Neoplatonic ontology.)

Théry, G. (1922) 'Originalité du plan de la Summa de Bono d'Ulrich de Strasbourg', *Revue Thomiste* 27: 376-97. (Demonstrates that Books 1-2 of the treatise comment on Pseudo-Dionysius' *De divinis nominibus*.)

Unamuno y Jugo, Miguel de (1864-1936)

The Spanish philosopher-poet Miguel de Unamuno upheld a heterodoxical Catholicism, resembling much nineteenth-century Liberal Protestantism, which viewed reason and faith as antagonistic. By 'reason', he understood scientific induction and deduction; by 'faith', a sentiment varying with his readings and personal experiences. Adolescent scepticism led him to reconcile science with religion by grafting Spencer's positivism onto various German idealisms, but a family tragedy brought this period of experimentation to an abrupt end. Obsessed with mortality, Unamuno achieved philosophical maturity with a blend of Liberal Protestant theology and the philosophies of James and Kierkegaard in his conception of the 'tragic sense of life' - the theme of his essays, novels, dramas, poetry and journalism. He acquired deep and intense insights into the quest for immortality. Unamuno was a professional in neither philosophy nor theology.

1 Life

Like many Spanish intellectual reformers of the late nineteenth and early twentieth centuries, Miguel de Unamuno sought to save Spain with rationalized religiousness. He dreamed of becoming the Spanish Luther. Brought up as a strict Catholic in the liberal Basque city of Bilbao, he lost his faith in Madrid of 1880, a city awash with many different philosophical currents, and thenceforth struggled to reconcile reason with religion and to stimulate Spaniards to follow his lead. To those ends he used his Chair in Greek Philology, obtained in 1891, and his controversial rectorship at the University of Salamanca (1901-14; 1930-6).

2 Reason and faith

Unamuno transposed problems of logic to the anthropological problem of life after death. Knowing that Hegel's dialectic somehow equated being and nonbeing, the sceptical young Unamuno of 1880 dreaded nonbeing after death. Hegel's logic starts from the datum of being (the thesis), whose indeterminacy is nonbeing (the antithesis). Thesis and antithesis rise to equality on a higher plane of discourse, the synthesis, termed by Hegel 'becoming'. By analogy, might not the negation of human being (death) imply salvation in a (higher) life eternal? Such a hope appeared in Unamuno's diaries and verses. In the 1880s he avidly read Herbert Spencer, who saw self-conscious individual existence as an incontrovertible truth, and in 1886 Unamuno began elaborating his own *'Filosofía lógica' ('Philosophy of Logic')* to unite Spencer with Hegel. Beginning with the undeniable fact of individual existence, Unamuno affirmed with Spencer the relativism of every relationship between subject and object (including mental representations of God and the soul). Unamuno called these representations 'images' in Hegel's sense of propositions or forms transcending common sense. Had the youthful Unamuno been bolder, he might have asserted endless individual existence as a matter of faith, defying common sense, yet demanded by a higher, rational dialectic; but he stopped himself making such a leap of faith.

Later he grafted Spencer onto a pre-existing system combining religion and science: Krausism, imported from Germany in 1844 (see Krause, K.C.F.). A leading Spanish Krausist, Unamuno's teacher Francisco Giner de los Ríos reinterpreted Krause to aid in the moral and intellectual reform of Spain. Younger Krausists sometimes adopted positivistic methods to buttress idealistic speculations, and styled themselves 'Krausopositivists'. Accordingly, Unamuno used Spencer to prove Krausist theories of human history in his influential 'Krausopositivistic' series of essays, En torno al casticismo (On Authentic Tradition) (1895). Here humanity is said to harmonize contraries such as reason and faith in a subrational collective heritage, the 'inner history'. However, Castilian intransigence hinders such syntheses and accounts for Spanish decadence. Unamuno's 1895 work helped writers of his literary generation of 1898 to explain Spain's defeat in her humiliatingly brief war with the United States.

Nevertheless, a personal crisis caused Unamuno to sweep aside previous intellectual accomplishments. In 1897, his son's death aroused in him nightmares of the abyss. Renouncing intellectualism, Unamuno sought to reconquer his childhood faith. He oscillated between retreating to orthodox Catholicism, converting to Liberal Protestantism, and yielding to scepticism. At the time, he was studying the prestigious Neo-Kantian school of Liberal Protestantism known as Ritschlianism, which he found compatible with Krausism. Aiming for an original philosophy of religion, he immersed himself in the history of Christian dogma as interpreted by the Ritschlian theologian Adolf von Harnack. Ritschl and Harnack held that Protestants, through justification by faith, receive

forgiveness for their sins and are consequently guaranteed salvation; while Catholics harbour doubts about salvation because of their indecisive theology. The troubled Unamuno uncritically accepted this Protestant opinion: hence the resemblances of his heterodoxy to conservative Ritschlianism. Yet in 1904 he valiantly tried to turn alleged Catholic uncertainty about salvation to his advantage, presenting it as a dynamic heritage he shared with all Spaniards. His concurrent readings in William James suggested the universality of the yearning for immortality.

Unamuno resolved to inculcate in his readers anxiety for their own immortality as a way to save Spain from dogmatic inertia. Spain's national Bible could be *Don Quixote*, whose hero equated the act of believing (*creer*) with creating (*crear*) its object: belief in a saving God makes the Saviour real. Just as Ritschl desired a nation of believers in forgiveness of sin, so Unamuno sought a nation of doubting believers in salvation. He modified Protestant thinkers such as Kierkegaard and Ritschl by substituting their concern about redemption with his own desire for immortality (see Ritschl 1883).

3 Chief works

Unamuno's main philosophical work, *Del sentimiento trágico dela vida en los hombres y en los pueblos (The Tragic Sense of Life)* (1911-12), contains a philosophy of religion more systematic than its doctrines allowed him to admit. Every flesh-and-blood human, Unamuno finds, desires immortality, but this desire clashes with sceptical scientific reason. If faith wins out, then God exists for human salvation; consists of Will and Personality, rather than of Reason; as Personality, is both finite and infinite, defying discursive reason; and is multifaceted and merciful enough to save every creature. Therefore the end of human history is the 'apocatastasis', the union of everything in God and the spiritualization of the universe. It is the human's moral duty to make the 'apocatastasis' a practical norm of everyday living and to deserve immortality. Structurally, Unamuno's system of ideas follows Ritschl's, which also begins with reflections on the conservation of personality, presents God as Personality and Love, and ends with an ethics of dominion over the world with devotion to vocation. Yet Unamuno substitutes considerations about life after death for Ritschl's concern with redemption here and now.

While *The Tragic Sense of Life* is Unamuno's most respected essayistic work, *La agonía del cristianismo (The Agony of Christianity)* (1931) is his most terse and anguished. Written in 1925 while Unamuno was in exile, this booklet derives the word 'agony' from the Greek word for struggle. Christianity entails struggle between universality and personalism; between truth, social and collective, and revelation, personal by essence; between Jewish belief in resurrection of the flesh and Hellenic credence in immortality of the soul; between monkish nihilism and the immortalizing desire for parenthood; between Catholicism and Protestantism, opposites which attract. Unamuno himself 'agonizes' while writing, hoping to serve Spain best by polemicizing from afar against her dictator Primo de Rivera.

The tragic sense of life generated Unamuno's best philosophical novels, *Niebla (Mist)* (1914), *Abel Sánchez* (1917) and *San Manuel Bueno, mártir (Saint Emmanuel the Good, Martyr)* (1933). *Mist* protests against the world-plan hypothesized by William James, with God and humans unfairly competing as if in a game of chess with an invincible champion. Unamuno yearned for a relationship of saving love between deity and humankind. His protagonist Augusto Pérez, confronting his Author-God Unamuno (a caricature of the flesh-and-blood author), anguishes for independence of his tyrannical creator. *Abel Sánchez* lodges a more specific protest against the world order, seen this time as a hospital for incurables afflicted by spiritual diseases, such as hate. Unamuno's modernization of the tale of Cain and Abel recounts Cain's experience from two tragically clashing viewpoints - the orthodox and the modern. Cain's malice is comprehensible both from the traditional standpoint, as envy, a sin, a disease, and, from the perspective of Unamuno's correspondent Croce, as the fact of human passion, capable of aggrandizing the individual.

In a subtler novelistic experiment, *Saint Emmanuel the Good, Martyr*, a beloved rural priest preaches immortality, in which he secretly does not believe. Greenfield (1979) defines him as a suicidal drive with a life-promoting direction. Ethically, his selflessness towards his parishioners daily helps spiritualize the universe; psychologically, his self-destructive tendency clashes with his desire for parenthood, which is equated in *The Agony of Christianity* to a yearning for immortality; existentially, he corresponds to Kierkegaard's individual feeling himself nothing before God, not because of sinfulness as in the case of the Danish philosopher himself, but because of doubt in immortality. In conclusion, though neither a professional philosopher like the Kierkegaardian phenomenologist

Heidegger, nor a professional theologian like the post-Ritschlian Barth, Unamuno expressed the tragedy of human mortality with multifaceted intensity.

See also: Spain, philosophy in

NELSON R. ORRINGER

List of works

Unamuno, M. de (1886) *'Filosofía lógica' (Philosophy of Logic)*, unpublished manuscript, Casa-Museo Unamuno (Unamuno Homestead Museum), Salamanca.(Early work, influenced by Hegel, superseded by *Del sentimiento*….)

Unamuno, M. de (1895) En torno al casticismo (On Authentic Tradition), *La Revista Moderna* 7 (74-8); Madrid: Biblioteca Moderna de Ciencias Sociales, 1902; 1 essay of 5, 'El espíritu castellano', trans. J.F. Crawford-Flitch, in *Essays and Soliloquies*, New York, 1925.(Series of essays on the concept of *casticismo* and Spanish national identity, considered as important, if idiosyncratic, statement of the principles of the group of writers and thinkers known as the 'Generation of '98'.)

Unamuno, M. de (1911-12) *Del sentimiento trágico de la vida en los hombres y en los pueblos, La España Moderna*, Madrid; Madrid: Renacimiento, 1913; trans. J.E. Crawford-Flitch, *Tragic Sense of Life*, London, 1921.(His major philosophical work, influenced by Nietzsche, pre-empting much existential thought and theologically unorthodox for Spain at the time; argues for the capacity of feeling rather than reason as the motivating force behind human creativity.)

Unamuno, M. de (1914) *Niebla (Nivola)*, Madrid: Renacimiento; trans. W. Fite, *Mist: A Tragi-Comic Novel*, New York, 1928.(The first modernist novel in Spain, constituting a playful questioning and overturning of the hierarchical author-character and author-reader relationship, in the form of a transposition onto aesthetics of his theological and philosophical thought.)

Unamuno, M. de (1917) *Abel Sánchez. Una historia de pasión*, Madrid: Renacimiento; trans. A. Kerrigan, *Ábel Sánchez and Other Stories*, Chicago, IL, 1956.(Mid-period novel, reworking of the Cain and Abel story.)

Unamuno, M. de (1920) *Tres novelas ejemplares y un prólogo*, Madrid: Espasa-Calpe; trans. Á. Flores, *Three Exemplary Novels*, New York: Grove Press, 1956.(Novellas: the prologue constitutes a major work of literary criticism; elaborates the idea of reading as creative play or *re-creación*, and marks the beginning of his interest in the status of women within society.)

Unamuno, M. de (1921) *La tía Tula*, Madrid: Espasa-Calpe.(A novel which attempts to reconcile and comment on the contradictory images of women within Christianity; contains some proto-feminist statements by the main character, Gertrudis.)

Unamuno, M. de (1925) *La agonía del cristianismo*, Madrid: Compañía Ibero Americana de Publicaciones; trans. J. Cassou, *L'Agonie du christianisme*, Paris, 1931; trans. A. Kerrigan, *The Agony of Christianity* and *Essays on Faith*, Princeton, NJ, 1974.(Written in exile, Unamuno's last major essayistic statement on theology, still playing on the concept of anguished suffering in the face of *la nada* (*le néant*) as humanity's source of creativity.)

Unamuno, M. de (1933) *San Manuel Bueno, mártir y tres historias más*, Madrid: Espasa-Calpe; trans. A. Kerrigan, *Saint Emmanuel the Good, Martyr, in Abel Sánchez and Other Stories*, Chicago, IL, 1956; repr. in *The Existential Imagination*, ed. F.R. Karl and L. Hamalian, Greenwich, CT, 1963.(Short novel on the paradoxes of faith and doubt, knowledge and innocence; the last work in what Unamuno considered to be a trilogy consisting of *Del sentimiento*…, *La agonía*…and *San Manuel*….)

Unamuno, M. de (1959-64) *Obras completas (Complete Works)*, Madrid: Vergara Editorial, 16 vols; repr. Madrid: Escelicer, 1966, 9 vols.(Standard complete works, but neither complete nor carefully edited.)

References and further reading

Fernández, P.H. (1976) *Bibliografía crítica de Miguel de Unamuno*, Madrid: J. Porrúa.(The most complete bibliography on Unamuno criticism, ordered by date of publication.)

Greenfield, S. (1979) 'La iglesia terrestre de San Manuel Bueno', *Cuadernos Hispanoamericanos* 348: 609-20.(A view of Manuel Bueno, Unamuno's best-known fictional character, as a suicidal drive which takes a socially productive direction.)

Jiménez García, A. (1985) *El krausismo y la Institución Libre de Enseñanza*, Madrid: Cincel.(A concise history of Spanish Krausism.)

Nozick, M. (1971) *Miguel de Unamuno*, New York: Twayne.(The deepest, most concise and comprehensive treatment of Unamuno's philosophy.)

Orringer, N.R. (1986-7) ‘Martin Nozick’s Unamuno: A Fountainhead of Future Discoveries’, *Siglo XX/20th Century* 4 (1-2): 30-43.(Survey of convincing philosophical criticism of all Unamuno’s works, with suggestions for future research.)

Orringer, N.R. (1985) ‘Unamuno y los protestantes liberales’, in *Sobre las fuentes de ‘Del sentimiento trágico de la vida’*, Madrid: Gredos.(Contributor’s analysis of Liberal Protestantism within *Tragic Sense of Life*.)

Ritschl, A. (1883) *Die christliche Lehre der Rechtfertigung und Versöhnung*, Bonn: Adolph Marcus, 2nd edn.(The Liberal-Protestant theological treatise from which Unamuno took the structure of ideas he used in *‘Tragic Sense of Life’*.)

Valdés, M.J. and Elena, M. (1973) *An Unamuno source book: A Catalogue of Readings and Acquisitions with an Introductory Essay on Unamuno’s Dialectical Enquiry*, Toronto, Ont.: University of Toronto Press. (Indispensable list of holdings and annotations in Unamuno’s personal library at the Casa-Museo Unamuno, Salamanca.)

Unconscious mental states

Unconscious phenomena are those mental phenomena which their possessor cannot introspect, not only at the moment at which the phenomenon occurs, but even when prompted ('Do you think/want/…?'). There are abundant allusions to many kinds of unconscious phenomena from classical times to Freud. Most notably, Plato in his Meno defended a doctrine of anamnesis *according to which a priori knowledge of, for example, geometry is 'recollected' from a previous life. But the notion of a rich, unconscious mental life really takes hold in nineteenth-century writers, such as Herder, Hegel, Helmholtz and Schopenhauer. It is partly out of this latter tradition that Freud's famous postulations of unconscious, 'repressed' desires and memories emerged.*

Partly in reaction to the excesses of introspection and partly because of the rise of computational models of mental processes, twentieth-century psychology has often been tempted by Lashley's view that 'no activity of mind is ever conscious' (1956). A wide range of recent experiments do suggest that people can be unaware of a multitude of sensory cognitive factors (for example, pupillary dilation, cognitive dissonance, subliminal cues to problem-solving) that demonstrably affect their behaviour. And Weiskrantz has documented cases of 'blindsight' in which patients with damage to their visual cortex can be shown to be sensitive to visual material they sincerely claim they cannot see.

The most controversial cases of unconscious phenomena are those which the agent could not possibly introspect, even in principle. Chomsky ascribes unconscious knowledge of quite abstract principles of grammar to adults and even newborn children that only a linguist could infer.

Many philosophers have found these claims about the unconscious unconvincing, even incoherent. However, they need to show how the evidence cited above could be otherwise explained, and why appeals to the unconscious have seemed so perfectly intelligible throughout history.

1 What counts?

Some care is needed to define exactly what counts as unconscious. Not just any phenomenon that fails to be conscious is *unconscious* in the intended sense. Thus, planets do not unconsciously follow elliptical orbits; nor does one's stomach unconsciously follow rules for digesting food; nor (to take a fallacy from popular biology) do we or our genes unconsciously seek replication when we act in ways that may in the past have had that consequence. We might call these phenomena simply '*non*-conscious', distinguishing as *un*conscious only those *mental* phenomena of which the possessor could not be made aware at the time of their occurrence, even if prompted. Thus, planets, stomachs and genes are non-conscious, having no mental lives at all.

But this definition is not quite enough. Suppose it were true that someone, Seymour, unconsciously feared castration; and suppose that Seymour read in a book that all men unconsciously fear castration, and concluded *merely* from this that he, too, must have such a fear. Presumably his unconscious fear would not *thereby* have been rendered conscious. The fear remains 'unconscious' so long as he is not conscious of it in the *right*, 'introspective' sort of way. Specifying this right way may be none too easy; but this is a problem that arises with respect to other mental phenomena (for example, the relation of intentions to the intended act).

Whyte (1978) has usefully compiled allusions to unconscious processes from classical times to Freud (not always distinguishing, however, between unconscious and non-conscious). Not only did Plato posit unconscious memories (see Plato §11), but the Greek physiologist, Galen, postulated unconscious inferences in perception and Plotinus, 'feelings without awareness'. Aquinas inferred the existence of unconscious 'processes of the soul' from the fact that he 'did not observe [his] soul apart from its acts'. Leibniz argued that ordinary perceptions must be the result of innumerable smaller perceptions occurring below a quantitative threshold (see Leibniz, G.W. §8). Unconscious thoughts and desires were a common theme among eighteenth- and nineteenth-century romantic poets and philosophers, such as Herder, Novalis, Richter, Wordsworth, Hegel and Schopenhauer; and, anticipating contemporary computational models of mind, the great nineteenth-century psychologist, Helmholtz, developed a fairly elaborate theory of unconscious inferential processing (see Helmholtz, H. von §2) Much of this served as the background for perhaps the most famous advocate of the unconscious, Freud.

2 Freud

Sigmund Freud postulated unconscious mental states to explain a wide variety of behaviours. His earliest evidence came from experiments involving hypnotic suggestion, experiments which, though fascinating, remain still quite controversial (see Hilgard 1977). Freud's primary argument is that parapraxes (for example, slips of the tongue), dreams, hysterical and neurotic symptoms, even 'ideas that come into our head we do not know from where' remain 'disconnected and unintelligible' if we insist that all mental states are conscious, whereas they 'fall into a demonstrable connection if we interpolate between them the unconscious acts we [psychoanalysts] have inferred' (1915: 168). Moreover, he argues, if it turns out that this interpolation enables us to 'exert an effective influence upon the course of conscious processes, this success will have given us an incontrovertible proof of the existence of what we have assumed' (1915: 168).

'Incontrovertible' seems a little strong. Certainly there *could be* plausible theories of 'the ideas that come into our head', and perhaps even of dreams and neuroses, that do not postulate unconscious mental states: especially the random flow of ideas in both dreams and daily life could have a purely physical explanation. Of course, we can imagine a standard mentalistic explanation in terms of beliefs and motives. But it is not enough to imagine the pattern; one has to provide some reason to think the pattern is genuine. For example, one has to find some critical evidence that could not be explained by any independently warranted purely physiological hypothesis. Of course, these latter explanations may not make a person's behaviour 'intelligible', if what that means is that they should be seen as part of some kind of 'rational' or 'symbolic' pattern. There is no doubt that Freud's explanations often possess an extraordinary charm in this way (Wittgenstein 1966). But then we need an argument that the charming pattern Freud seeks actually corresponds to any real processes in a person's mind. Perhaps there is no true *causal* explanation of dreams or neuroses that makes them 'intelligible' in *that* particular way.

Freud's further argument from therapeutic success is, moreover, a notorious fallacy: the history of religion and psychotherapy is full of success stories involving theories that there is no serious reason to believe. Many such effects may well be due to suggestion or well-established 'placebo' effects, whereby someone's belief merely that they have been cared for at all can actually improve their condition (Grünbaum 1984). None of this implies that Freud's postulations are *false*. The point is that the truth in this domain may be harder to ascertain than Freud recognizes.

What Freud needs, and sometimes does provide, are regularities that could not be explained except by adverting to the specific contentful states he posits. He and Breuer began to do this when they noted that hysterical patients were often 'paralysed' not at places at which normal paralyses occurred, but, rather, at places that were *popularly and mistakenly believed* to be subject to paralysis (for example, 'glove' anaesthesia). Here the only plausible explanation of such curious paralyses involves appeal to the *content* of those popular beliefs, and this it does whether or not the believer is aware of that content.

Freud's actual account of the unconscious (in what he sometimes calls 'the truly dynamic sense', and which he comes to call the 'id') is actually quite special. He thinks that unconscious or 'primary' processes are intellectually quite primitive: they have no representation of negation, degrees of certainty, time or an independent reality, and are only subject to 'the pleasure principle'. Interestingly, he denies the existence of unconscious *affect* (Freud 1933). Nevertheless, he claims, 'every psychical act begins as an unconscious one, and…may remain so or go on developing into consciousness according as it meets resistance or not' (1915: 55). It appears that the only reason he can see for material not becoming conscious is that it has been in this way censored or 'repressed'. It seems not to have occurred to him that there could be other reasons for material to remain unconscious: for example, that there may not be the right kind of causal or 'informational' path between the material and mechanisms of introspection (see Psychoanalysis, methodological issues in).

3 Recent experiments

This latter possibility emerges in contemporary cognitive science. In an influential article, Nisbett and Wilson (1977) reviewed a wide range of experiments in which people can be shown to be sensitive to a variety of what would appear to be conceptualized contents, but without any awareness of them. Instead of noticing these factors, the subjects frequently 'introspect' material that can be independently shown to be irrelevant to the causation of their behaviour. Indeed, even when explicitly asked about the relevant material, they will deny that it played any

role. For example, Latané and Darley (1970) showed that subjects are less likely to aid strangers in distress the greater the number of bystanders in the situation, but with no awareness of this effect. And Nisbett and Wilson (1977) asked fifty-two subjects to choose between what were (unknown to them) four *identical* pairs of nylon stockings presented in a left-to-right array: 80 per cent of them tended to choose the right-most pair; but, when asked, virtually all subjects denied that position had any effect. Similar results obtain for cognitive dissonance, subliminal clues in problem solving and semantic disambiguation, and subtle perceptual cues, such as the dilation of pupils in a face (Hess 1975).

One of the most philosophically interesting phenomena supporting the evidence for unconscious mental states is that of 'blindsight'. This occurs in patients who have suffered damage to their visual cortex and who claim not to be able to see various stimulus materials that have been presented to them; nevertheless, they can be shown to score well above chance when they are forced to 'guess' at what they may be seeing (Weiskrantz 1986).

What exactly such data show about the *extent* of unconscious processes is still a matter of considerable controversy. Velmans (1991), supporting Lashley (1956), argues that all cognitive processes can occur unconsciously. In an important study, Ericsson and Simon (1984) try to sort out the different informational paths involved in different mental processes such as perception, long- and short-term memory retrieval, and in a person's surmises about the causation of their behaviour, arguing that some of these paths do and some do not provide a basis for genuine introspection.

4 Chomsky

All the examples considered so far are examples of cognitive contents that people would at least ordinarily *understand*. However, more theoretical cognitive scientists, influenced especially by the linguist, Noam Chomsky, often ascribe quite technical contents even to newborn children. Chomsky (1986) argues, for example, that the best explanation of the rapidity with which children acquire natural language is that they have innate knowledge of general principles of what he calls 'universal grammar' (see Language, innateness of). This includes such principles, for example, as that a pronoun cannot be used to refer to something picked out by a noun in a clause subordinate to the clause in which it occurs. For example, in sentence (1), 'he' cannot be knowingly used to refer to John, although it can in (2):

(1) He hoped John would win.
(2) John hoped he would win.

The full statement of the rules for pronouns is actually quite technical, involving notions such as 'c-command' and 'phrasal head' that require a good deal of theoretical study to understand. If a newborn child is said to 'know' such a principle, it would patently be a piece of unconscious (or 'tacit') knowledge, at least in the sense that the child is entirely unable to express it, even when prompted. Indeed, there is no *introspective* route even in an adult linguist from the unconsciously known principle to any awareness of it.

Chomsky's research has stimulated an entire movement not only in linguistics, but in many other parts of psychology (for example, vision, reasoning, cognitive development), where knowledge of similarly abstract principles about material objects, biological kinds and folk physics are also attributed to adults and young children alike. In a different vein, Polanyi (1967) has stressed the role of 'tacit' knowledge in guiding creativity and scientific research (see Knowledge, tacit; Nativism).

There can be no doubt that much of the evidence adduced by Chomsky and these other psychologists is impressive and cries out for explanation, which is not easy to provide. But many philosophers doubt that the explanation needs actually to claim that people *know*, *believe* or, in Chomsky's phrase, even 'cognize' (1986: 265) the relevant principles. It may be enough that they simply *obey* the principles in the way that their bodies obey the laws of gravity, without *following* them, indeed, without them being represented or otherwise part of their *mental* lives. They may explain people's knowing *how* to do certain things, without requiring that they know *that* certain principles obtain (see Ryle 1949; Devitt and Sterelny 1989).

Much here depends upon being clearer than anyone is about what is required for ascribing any mental content at all. Quite apart from issues of awareness, Davidson (1980) claims that content ascription only makes sense when many contents are ascribed as some kind of interacting rational system. If so, then many of Freud's and

Chomsky's claims would appear to be in trouble: as Stich (1978) points out, if Chomsky were correct, someone could know innately some principle *P*, be convinced as an adult that *P* entailed *Q*, but not be the least disposed to infer *Q*. But then, on a Davidsonian view, that should be a reason not to ascribe *P* to that person in the first place.

Stich proposes thinking of Chomsky's grammatical principles as contents occurring at a different, 'sub-doxastic' level of psychology, a suggestion taken up systematically by Fodor (1983) in his postulation of 'informationally encapsulated modules' of the mind (see Modularity of mind). But the question persists: why think that the principles comprise genuine *mental contents* inside a module? Peacocke (1989) suggests that it is enough that a system be 'sensitive to the information' expressed by the principle. But this seems too broad: for example, if we are to take talk of 'genetic information' seriously, DNA would seem to have contentful attitudes about the structure of amino acids.

5 Philosophical qualms

A number of philosophers have objected to the postulation of unconscious mental states. McIntyre (1958) and Searle (1992) have insisted that it is some sort of conceptual truth that mental states involve at least a disposition to have certain experiences. Searle, for example, defends what he calls the 'connection principle', according to which 'we understand the notion of an unconscious mental state only as…the *sort of thing* that…could be…conscious' (1992: 155-6). Now, of course, *any* contents - from the most abstruse Chomskean principles to the most sordid of Freudian analyses - are, for almost anyone, 'possible contents of consciousness': most people can be brought at least to *understand the words*. Reiterating the point we made early on, what Searle presumably means is that mental contents have to be *accessible* to their possessor in something like the normal introspective way, whatever way that is. But once one reflects on how very special that way may be, it is puzzling why the existence of a mental state should depend upon it.

Strawson goes further and claims that 'there are, strictly speaking, no dispositional non-experiential mental phenomena' (1994: 167), comparing a non-experiential brain to a CD recording of a Beethoven quartet that is not currently being played, and which could be used either to produce a genuine acoustical event on a CD player or a light show.

What tells against all this scepticism, however, is the explanatory usefulness of unconscious intentional states having genuine causal powers. Freud's and contemporary cognitive postulations of unconscious states are interesting because, if true, they would explain certain *content-sensitive* behaviours, such as parapraxes, neuroses, and differential willingness to come to the aid of strangers; Chomsky's, because, if true, they would explain language acquisition. Weiskrantz's postulation of'blindsight' explains his patients' ability to perform better than chance at picking out stimuli of which they are unaware. The crucial difference between Strawson's example of the unplayed CD and the unconscious brain is that, whereas the *musical* properties of the unplayed CD are, indeed, at that time causally inert, the intentional content properties of these unconscious mental states are not. If Freud, Chomsky, Nisbett and Wilson or Weiskrantz are correct, they are responsible for a variety of observable phenomena. Someone sceptical of unconscious states needs to show how all the phenomena these figures cite can be otherwise explained.

Searle does provide a further argument against entirely unconscious states, claiming that no postulation of an attitude lacking consciousness can account for 'aspectual shape' (1992: 156-61, 169-72), or what is better known as the phenomena of (hyper-)intensionality, whereby minds are able to distinguish among even necessarily coextensive concepts, such as [water] and [H_2O]. Searle believes that only a conscious mind can possess such an ability (see Bealer 1984 for a similar, more formidable argument).

A full reply to this point would require setting out the various proposals of a great number of writers regarding how the problems of both intentionality and (hyper-)intensionality are to be solved. Suffice it to say that there are a number of possibilities to be considered in detail. For example, many think that a computational theory of mind involving a language of thought, endowed with both narrow and wide contents, will suffice (see Language of thought; Content: wide and narrow). Many different sentences with different causal roles within the mind - for example, 'Water is wet', 'H_2O is wet' - may express the same *wide* content, and so constitute different 'aspectual shapes' of that content. Searle would in turn reject such an account, invoking his 'Chinese room argument'; however, the soundness of this argument is widely disputed (see Chinese room argument).

In any case, the fact that so many people, from antiquity to the present, find the postulation of unconscious phenomena both intelligible and illuminating provides an at least *prima facie* case that they are not incoherent in the ways these philosophers claim. Indeed, until better explanations of the various phenomena we have discussed are presented, that postulation needs to be taken seriously.

See also: Consciousness; Introspection, epistemology of; Introspection, psychology of; Psychoanalysis, methodological issues in

GEORGES REY

References and further reading

Bealer, G. (1984) ‘Mind and Anti-Mind’, *Midwest Studies in Philosophy* 9: 283-328.(A rich presentation of an argument, derived from Quine, that the phenomenon of hyper-intensionality requires first-person conscious experience.)

Brown, D. (1995) ‘Pseudo Memories: The Standard of Science and the Standard of Care in Trauma Treatment’, *American Journal of Clinical Hypnosis* 37: 1-24.(Recent experimental work on ‘repressed memory’.)

Chomsky, N. (1986) *Knowledge of Language*, New York: Praeger.(A recent, relatively non-technical presentation of his views about innate knowledge of grammar.)

Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Oxford University Press.(Development of a view of the ascription of mental states as involving making rational sense of an agent overall.)

Davies, M. (1989) ‘Tacit Knowledge and Sub-Doxastic States’, in A. George (ed.) *Reflections on Chomsky*, Oxford: Blackwell.(Subtle discussion of whether ‘modules’ for language and perception can be full mental states.)

Davies, M. (forthcoming) *Tacit Knowledge*, Oxford: Blackwell.(Brings together the author’s developments of his views about the kind of tacit knowledge Chomsky ascribes to language users.)

Devitt, M. and Sterelny, K. (1989) ‘Linguistics: What’s Wrong with "the Right View"’, in J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 3, *Philosophy of Mind and Action Theory*, Atascadero, CA: Ridgeview.(A sceptical response to Chomsky’s claims about the psychological reality of grammars.)

Ericsson, K. and Simon, H. (1984) *Protocol Analysis: Verbal Reports as Data*, Cambridge, MA: MIT Press, 2nd edn, 1993.(A detailed development of a theory of reliable introspective processes, and the ‘protocols’ to which they give rise.)

Evans, G. (1981) ‘Semantic Theory and Tacit Knowledge’, in S. Holtzman and C. Leich (eds) *Wittgenstein: To Follow a Rule*, Oxford: Routledge & Kegan Paul.(Subtle discussion of whether a Chomskean appeal to tacit knowledge of rules can explain people’s grasp of the *semantics*, as well as of the syntax, of their language.)

Fisher, J. (1974) ‘Knowledge of Rules’, *Review of Metaphysics* 28: 237-60.(Useful discussion distinguishing knowledge *of* a rule of grammar from knowledge *about* such a rule.)

Fodor, J. (1983) *The Modularity of Mind*, Cambridge, MA: MIT Press.(Postulation of ‘informationally encapsulated modules’ of the mind.)

Freud, S. (1912) ‘A Note on the Unconscious in Psychoanalysis’, in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 12, trans. and ed. J. Strachey *et al.*, London: Hogarth Press, 1958.(Short, early anticipation of the 1915 essay.)

Freud, S. (1915) ‘The Unconscious’, in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 14, trans. and ed. J. Strachey *et al.*, London: Hogarth Press, 1958.(Freud’s most extended, explicit treatment of the topic.)

Freud, S. (1933) ‘New Introductory Lectures on Psychoanalysis’, in *The Complete Introductory Lectures on Psychoanalysis*, trans. and ed. J. Strachey, New York: Norton, 1966.(Freud’s best summaries of his own theories.)

George, A. (ed.) (1989) *Reflections on Chomsky*, Oxford: Blackwell.(Excellent collection of a number of recent articles listed here on Chomsky’s claims about tacit knowledge.)

Grünbaum, A. (1984) *The Foundations of Psychoanalysis: A Philosophical Critique*, Berkeley, CA: University of California Press.(A standard philosophical critique of Freudian theories.)

Hess, R. (1975) ‘The Role of Pupil Size in Communication’, *Scientific American* 233 (5): 110-18.(Experiment in which males’ preferences for female faces were shown to depend upon the size of the pupil in the depicted face, without the males noticing this fact.)

Hilgard, E. (1977) *Divided Consciousness: Multiple Controls in Human Thought and Action*, New York: Wiley & Sons.(Interesting contemporary research on hypnosis and suggestion.)

Lashley, K. (1956) ‘Cerebral Organization and Behavior’, in H. Solomon, S. Cobb and W. Penfield (eds) *The*

Brain and Human Behavior, Baltimore, MD: Williams and Wilkens Press.(Classic statement of the widespread contemporary view that most - Lashley claims 'all' - important mental processes are unconscious.)

Latané, B. and Darley, J. (1970) *The Unconscious Bystander: Why Doesn't He Help Us?*, New York: Appleton-Century-Crofts.(Experiment in which a person's readiness to come to the aid of someone in distress was reduced if there were more bystanders, without the person being aware of this fact.)

Loftus, E., Polonsky, S. and Fullilove, M.T. (1994) 'Memories of Childhood Sexual Abuse: Remembering and Repressing', *Psychology of Women Quarterly* 18: 67-84.(Recent experimental studies of 'repressed memory' of sexual abuse.)

Lyons, W. (1986) *The Disappearance of Introspection*, Cambridge, MA: MIT Press.(A discussion of the decline of introspection in our understanding of the mind.)

McIntyre, A. (1958) *The Unconscious*, Oxford: Routledge & Kegan Paul.(Some philosophical criticism of Freudian appeals to the unconscious, from days prior to cognitive science.)

Nisbett, R. and Wilson, T. (1977) 'On Telling More Than We Can Know: Verbal Reports on Mental Processes', *Psychological Review* 84 (3): 231-59.(Influential and fascinating review of a wide range of experiments showing people to be sensitive to phenomena of which they are unaware.)

Peacocke, C. (1989) 'When Is a Grammar Psychologically Real?', in A. George (ed.) *Reflections on Chomsky*, Oxford: Blackwell.(Argues that psychological reality turns on a system's sensitivity to information - irrespective of whether it is conscious.)

Pinker, S. (1994) *The Language Instinct: How the Mind Creates Language*, New York: Harper & Row.(A superb introduction to the revolution caused by Chomsky's postulation of innate, tacit knowledge of grammar.)

Polanyi, M. (1967) *The Tacit Dimension*, Oxford: Routledge & Kegan Paul.(Discussion of the role of unconscious knowledge in creative thought.)

Rey, G. (1997) *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Oxford: Blackwell.(Sets out in detail the structure of contemporary computational/representational theories of mind, arguing that their success does not depend upon the postulated cognitive states' being conscious.)

Ryle, G. (1949) *The Concept of Mind*, London: Hutcheson.(Historical source of the suggestion that much unconscious knowledge may be merely 'know how', not 'knowledge that'.)

Searle, J. (1992) *The Rediscovery of the Mind*, Cambridge, MA: MIT Press.(A vigorous denial of the possibility of unconscious mental phenomena.)

Stich, S. (1978) 'Beliefs and Sub-Doxastic States', *Philosophy of Science* 45: 499-518.(Proposes thinking of Chomsky's grammatical principles as contents occurring at a different, 'sub-doxastic' level of psychology.)

Strawson, G. (1994) *Mental Reality*, Cambridge, MA: MIT Press.(An effort to argue against unconscious mental phenomena on intuitive philosophical grounds.)

Velmans, M. (1991) 'Is Human Information Processing Conscious?', *Behavioral and Brain Sciences* 14 (4): 691-2.(Argues that consciousness is inessential to all cognitive processing.)

Weiskrantz, L. (1986) *Blindsight: A Case Study and Implications*, Oxford: Oxford University Press.(The major study of the surprising phenomenon of blindsight.)

Whyte, L. (1978) *The Unconscious Before Freud*, New York: St Martin's Press.(A not particularly philosophical, but very useful, survey of claims about the unconscious from classical Greece to Freud.)

Wittgenstein, L. (1966) *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*, Berkeley, CA: University of California Press.(Early philosophical criticism of Freud.)

Underdetermination

The term underdetermination refers to a broad family of arguments about the relations between theory and evidence. All share the conclusion that evidence is more or less impotent to guide choice between rival theories or hypotheses. In one or other of its guises, underdetermination has probably been the most potent and most pervasive idea driving twentieth-century forms of scepticism and epistemological relativism. It figures prominently in the writing of diverse influential philosophers. It is a complex family of doctrines, each with a different argumentative structure. Most, however, suppose that only the logical consequences of a hypothesis are relevant to its empirical support. This supposition can be challenged.

1 Relevant terminology

The thesis of underdetermination involves a set of specific claims about the relations between bodies of evidence or statements about that evidence, on the one hand, and theories, on the other. All forms of that thesis insist that multiple, mutually incompatible, theories can enjoy the same relation to any given body of evidence. Beyond that point of agreement, however, there are more and less ambitious versions of underdetermination. A few terminological preliminaries are in order before we proceed to examine some of them. To facilitate comparison of the various versions of underdetermination, I shall use this vocabulary: *e* will refer to whatever evidence is in hand; *c* will refer to relevant rules or criteria for choosing between hypotheses; and I shall further suppose that we are confronted by a single hypothesis *h* or, more commonly, by a set of rival hypotheses, $h_1, h_2, \ldots,_n$ (some of which may not have been explicitly formulated).

Now, what every version of the thesis of underdetermination shares in common is this claim: that the relevant criteria, *c*, and available evidence, *e*, will provide no rational grounds for preferring h_i to some (and perhaps not to any) of the rivals to h_i. Various versions of the thesis of underdetermination explore its plausibility under various interpretations of *c*, the criteria of appraisal.

2 Deductive and ampliative underdetermination

Hume argued convincingly that one cannot derive a genuinely universal statement from any finite set of its positive instances (see Hume, D. §2). This is surely correct, but what does it entail for the theory of knowledge? One thing it certainly shows is that if the only criteria of theory choice we allow are those of deductive logic, then the rational acceptance of a theory is always underdetermined by any conceivable evidence. Karl Popper explicitly drew this moral fromHume’s work (see Popper, K.R. §2). It explains why Popper resisted the idea that positive belief in a theory could ever be rational. (Belief for Popper, as for Hume, was a matter of ‘animal instinct’.)

Most philosophers refuse to follow Popper in supposing that the rules of deductive logic exhaust the epistemic and methodological resources available for the appraisal of hypotheses. They believe that there are various principles of ampliative inference that come into play in the evaluation of hypotheses. The fact that deductive logic underdetermines theory choice leaves open, they insist, the possibility that these other nondeductive principles may guide theory choice unambiguously.

One plausible, if simplistic, ampliative rule of theory evaluation is what Carl Hempel called the Nicod criterion (1965). According to this idea, if some *h* entails a particular observation report, then the truth of that observation report counts as evidence for *h*; if the observation report is false, then it counts as evidence against *h*. About a century ago, Henri Poincaré and others stressed the significance of the fact that, through any finite number of points, indefinitely many curves could be drawn. This made clear that indefinitely many alternative hypotheses were compatible with (indeed, deductively entailed by) any given body of data. If one supposes, with the Nicod criterion, that the potential evidence for *h* consists simply in its observational consequences, then it is clear that indefinitely many different and conflicting hypotheses will share the *same* evidential instances. If we take the Nicod rule as a criterion for hypothesis evaluation, it follows that, whenever we are confronted with an *h* enjoying apparent evidential support (that is, having many true positive or entailed instances), we know that there will be indefinitely many incompatible hypotheses possessing those same instances as ‘evidence’ - even if we cannot formulate them. In sum, theory choice is underdetermined by any methodology which countenances all and only the known positive (or otherwise entailed) instances of a hypothesis as confirmatory.

Nelson Goodman’s new riddles of induction provided concrete illustrations of the point (1965). Consider a hypothesis like ‘all emeralds are green’ and a contrived rival like ‘all emeralds are green before the year 2000 AD and blue thereafter’. If our observations consist entirely of green-appearing emeralds examined before 2000, then both hypotheses can claim all the same positive instances. Prior to the year 2000, the evidence, however extensive, appears impotent to guide choice between them (see Goodman, N. §3; Induction, epistemic issues in).

It is important to note that, even if hypothesis choice is underdetermined in this specific sense, that does not entail that *every* rival to a given *h* will fare as well as *h* does. It means, rather, that *there will be* rivals which will fare as well.

3 Holistic underdetermination (the ‘D-thesis’)

Arguments of a very different sort undergird an even more global form of underdetermination. During the 1950s, W.V. Quine revived arguments made earlier in the century by Pierre Duhem (1906) about the inconclusiveness of falsification (see Duhem, P.M.M.; Quine, W.V. §3). Specifically, the Duhem thesis (D-thesis) holds that hypotheses never approach the tribunal of experience singly but only as parts of much larger webs of belief. This must be so, since auxiliary hypotheses are typically needed even to determine what a given hypothesis predicts about experience. Suppose, asks Quine, that the result predicted by a web of such hypotheses turns out to be mistaken? In that case, he maintains, all we have learned is that we have made a mistake somewhere in our body of beliefs but that blame cannot be further localized. Moreover, he insists that we can hang on to any particular hypothesis we like ‘come what may’, since we can always blame any empirical failures of the web on other strands. More precisely, Quine holds that *any* given *h* can be maintained in the face of any apparently negative evidence because (he thinks) there will always be some auxiliary that can be legitimately invoked which will bring *h* into line with whatever we observe. Any *h* can, with suitable changes elsewhere in our beliefs, be made to explain any evidence. Although Quine and others have repeatedly asserted this thesis to be true, it has never been proven that every apparently refuted hypothesis can be saved in the face of any evidence by invoking a nontrivial auxiliary hypothesis.

If, however, this is correct, then we have a much more global sort of underdetermination to contend with. Whereas previous versions of the underdetermination thesis argued that, for any apparently successful *h*, there will be at least one rival which exhibits the same success, the holistic version of the underdeterminationist thesis seems to say that, given any apparently successful *h*, *all* of its rivals can be made to fit the same evidence that supports *h*, whatever *e* is. This view appears to entail a thesis of cognitive egalitarianism, namely, that all rival hypotheses can - with sufficient ingenuity - be made to appear equally well supported by whatever evidence is in hand.

4 Undermining underdetermination

What Quine’s version of the underdetermination thesis takes for granted, as do virtually all others, is the idea that a hypothesis’ entailed instances are coextensive with its epistemically supporting instances. He is committed to this since he countenances two hypotheses as equally well supported by *e* provided that both hypotheses entail *e*. This idea, the Nicod criterion, undergirds the paradoxes of confirmation, the Goodman paradoxes of induction and Quinean underdetermination.

Those who are less taken with the underdetermination thesis believe that its identification of entailed consequences with evidentially supporting instances rests on a misunderstanding of how empirical support works. Specifically, it is easy to show both (a) that many entailed instances of an hypothesis lend it no empirical support and (b) that much of the support enjoyed by many hypotheses comes from outside the set of its entailed consequences. As an example of (a), consider the fact that data used to fix the parameters of a hypothesis, by virtue of being built into *h* itself, cannot then be used as reasons for accepting *h*, even though (trivially) *h* entails them. As regards (b), it is virtually a truism that hypotheses can derive evidential support from instances not entailed by those hypotheses. For instance, the information that the last 10,000 crows have been black is evidence for the hypothesis that the next crow I observe will be black, even though that latter hypothesis entails nothing about the 10,000 crows. Equally, probabilistic statements (for example, ‘most humans die before the age of 100’) typically do not entail any specific claim about observed relative frequencies, even though those same relative frequencies can both confirm and disconfirm statistical hypotheses. Even with respect to a nonprobabilistic hypothesis, one can often derive empirical support from evidence which *h* does not entail but which is entailed by yet more general

hypotheses that in turn entail *h*. Thus, after the Newtonian synthesis, evidence about bodies falling on the face of the earth supports Kepler's laws, even though Kepler's laws themselves entail nothing whatever about cases of terrestrial free fall.

In sum, the supporting instances for a hypothesis are not coextensive with its entailed instances nor are its undermining instances exclusively those denied by *h*. Where does this leave the thesis of underdetermination? Because all the advocates of underdetermination have operated with this mistaken notion of empirical support - which assimilates support to being a logical consequence of a hypothesis - we do not presently know whether choice of theory is underdetermined by the evidence. While we do know that rival theories may well have the same known empirical consequences, this knowledge seems to be irrelevant to questions of empirical support, since there are ample grounds for denying that empirical consequences and supporting instances amount to the same thing.

It should be noted, finally, that virtually all formulations of the thesis of underdetermination presuppose that, in judging whether the available evidence supports a given hypothesis, we must assure ourselves that there is no *possible* hypothesis which could be equally well supported. The idea that extant hypotheses are being compared (at least implicitly) with all possible rival hypotheses is probably a hyperbolic view of the theory choice situation. Scientists repeatedly point out that the problem generally facing them is finding even one hypothesis that will fit with the available evidence. The philosopher's penchant for supposing that the function of evidence is to enable one to choose, not merely between extant hypotheses, but between all possible hypotheses is a conceit that could profitably be dispensed with.

See also: Confirmation Theory; Crucial experiments; Inductive Inference; Scientific method

LARRY LAUDAN

References and further reading

Duhem, P. (1906) *La théorie physique. Son objet et sa structure*, Paris: Chevalier et Rivière; trans. P.P. Wiener, *The Aim and Structure of Physical Theory*, Princeton, NJ: Princeton University Press, 1954.(One of the earliest attempts to explore the implications of underdetermination for the epistemology of science.)

Glymour, C. (1980) *Theory and Evidence*, Princeton, NJ: Princeton University Press. (A detailed account of the evidence relation in science.)

Goodman, N. (1965) *Fact, Fiction and Forecast*, Indianapolis, IN: Bobbs-Merrill.(A nominalist's formulation of new challenges to the idea of instance confirmation, discussed in §2.)

Grünbaum, A. (1974) *Philosophical Problems of Space and Time*, Dordrecht: Reidel. (The standard critique of Duhem and Quine's accounts of underdetermination.)

Hempel, C.G. (1965) *Aspects of Explanation*, New York: Free Press.(The principal formulation and critique of the qualitative theory of confirmation.)

Laudan, L. and Leplin, J. (1991) 'Empirical Equivalence and Underdetermination', *Journal of Philosophy* 88: 449-72.(Elaboration of many of the themes discussed here.)

Quine, W.V. (1953) *From a Logical Point of View*, Cambridge, MA: Harvard University Press.(The *locus classicus* for the 'Duhem-Quine thesis', discussed in §3.)

Unity of science

How should our scientific knowledge be organized? Is scientific knowledge unified and, if so, does it mirror a unity of the world as a whole? Or is it merely a matter of simplicity and economy of thought? Either way, what sort of unity is it? If the world can be decomposed into elementary constituents, must our knowledge be in some way reducible to, or even replaced by, the concepts and theories describing such constituents? Can economics be reduced to microphysics, as Einstein claimed? Can sociology be derived from molecular genetics? Might the sciences be unified in the sense of all following the same method, whether or not they are all ultimately reducible to physics? Considerations of the unity problem begin at least with Greek cosmology and the question of the one and the many. In the late twentieth century the increasing tendency is to argue for the disunity of science and to deny reducibility to physics.

What kinds of integration are manifested, or sought after, in the claims and practices of the different fields of inquiry we call ‘sciences’? This question should be carefully distinguished from any of the different specific theses addressing it and yet it should be stressed as the linking thread of a time-honoured philosophical debate. The question belongs to a tradition of thought that can be traced back to Presocratic Greek cosmology, in particular to the preoccupation with the question of the one and the many. In what senses are the world, and thereby our knowledge of it, one? A number of representations of the world emerged in terms of its decomposition into a few simple constituents: Parmenides’ static substance, Heraclitus’ flux of becoming, Empedocles’ four elements, Democritus’ atoms, or Pythagoras’ numbers, Plato’s forms, Aristotle’s categories.

With the advent and expansion of Christian monotheism, the organization of knowledge reflected the idea of a world governed by the laws dictated by God, creator and legislator. Encyclopaedic efforts such as the Etymologies, compiled in the sixth century by Isidore, Bishop of Seville, the works of Ramon Llull in the Middle Ages and of Petrus Ramus in the Renaissance emerged from this tradition. Llull introduced tree-diagrams and forest-encyclopedias organizing different disciplines (including law, medicine, theology and logic). He also introduced more abstract diagrams - not unlike some found in Kabbalistic and esoteric traditions - in an attempt to encode combinatorially the knowledge of God’s creation in a universal language of symbols. Ramus introduced diagrams representing dichotomies and gave prominence to the view that the starting point of all philosophy is the classification of the arts and sciences.

The emergence of a distinctive tradition of scientific thought addressed the question of unity through science’s designation of a privileged method, set of concepts and language. Francis Bacon held that a unity of the sciences emerges out of our classifications of material facts in the form of a pyramid with different levels of generalities. Galileo Galilei proclaimed that the Book of Nature was written by God in the language of mathematical symbols and geometrical truths: the story of Nature’s laws is told in that language in terms of a reduced set of primary qualities - extension, quantity of matter and motion.

Descartes and Leibniz gave this tradition a rationalist twist centred on the powers of human reason. Like Llull’s, their conception of unity is combinatorial, that is, determined by rules of analysis and combination. According to Descartes the science of geometry, with its demonstrative reasoning from the simplest and clearest thoughts, constitutes the paradigm for the ideal of unification of all knowledge (see Descartes, R. §§2-3). Descartes symbolized this unity again using the image of a tree. Unlike in the manual arts, there should be no division of labour in the sciences; they ought not to be studied separately according to their subject matter. Leibniz proposed a ‘general science’ in the form of a Demonstrative Encyclopaedia. This would be based on a ‘catalogue of simple thoughts’ and an algebraic language of symbols, ‘characteristica universalis’, which would render all knowledge demonstrative and allow disputes to be resolved by precise calculation (see Leibniz, G.W. §10).

The notion of unity of science, along with that of the universality of rationality, was given its strongest promotion during the European Enlightenment. The most important expression of the encyclopaedic tradition came in the mid-eighteenth century from Diderot and D’Alembert, editors of the *Encyclopédie* (1751-72). Diderot stressed in his own entry, ‘Encyclopaedia’, that ‘the word Encyclopaedia signifies the unification of the sciences’. The function of the encyclopedia was to exhibit the unity of human knowledge. Diderot and D’Alembert, in contrast with Leibniz, made classification by subject primary, and introduced cross-references by way of logical

connections.

For Kant the unity of science is not the reflection of a unity found in nature; rather, it has its foundations in the allegedly unifying nature of reason itself (see Kant, I. §7). This, however, lacks the Leibnizian form of a deductive system and combinatorial framework. Instead, Kant adopts the image of an organic architecture. Each science is a self-subsistent whole and separate buildings connect to other sciences by passages. He also held the view that biology cannot be reduced to mechanics.

Kant's ideas set the frame of reference for discussions of unification in German thought throughout the nineteenth century. He gave philosophical currency to the notion of world-picture (*Weltbild*), establishing unity of science as an intellectual ideal. Influenced by this tradition, which culminated with the work of Albert Einstein, the physicists Max Planck and Ernst Mach engaged in a heated debate about the precise character of the unified scientific world-picture. Mach's, more influential, view was phenomenological and Darwinian: the unification of knowledge took the form of an analysis of ideas into elementary sensations and was ultimately a matter of adaptive economy of thought. In physics the 'electromagnetic world-picture' and Planck's thermodynamical worldview constituted some of the alternatives to a long-standing mechanistic view that since Newton had affected biology as well as most branches of physics.

Kant also distinguished between several types of judgement that, in turn, characterized different intellectual disciplines. In this way he set the basis for the famous distinction between the natural sciences (*Naturwissenschaften*) and the cultural or social sciences (*Geisteswissenschaften*) introduced by Wilhelm Dilthey. Followers of this distinction, such as Max Weber and Heinrich Rickert (for Rickert, see Neo-Kantianism §7), claimed that the difference in subject matter between the two kinds of science forced a distinctive difference between their respective methods - this despite the fact that since Hume, Comte and J.S. Mill, the social sciences had relied on analogies with the natural sciences, from Newtonian and statistical mechanics and also from biology.

The logical positivists, notably the members of the Vienna Circle, introduced, under the banner of 'unity of science without metaphysics', a model of demarcation between science and metaphysics based on the unity of method and language that included all the sciences, natural and social (see Demarcation problem; Logical positivism; Vienna Circle). Notice that a common method does not imply a more substantive unity of content involving theories and their concepts. In the stronger model especially recommended by Rudolf Carnap, however, unity was affirmed because of the associated value of simplicity, and was characterized by axiomatic structures, reductive logical connections between concepts, and laws of the different sciences at different levels, with physics as the fundamental basis, and the possibility of empirical grounding and testability. Carnap was influenced by the empiricist tradition (especially Russell and Mach) and the ideals of simplicity and reductive logical analysis in the early works of Russell and Wittgenstein (see Analytical philosophy §2). Otto Neurath, by contrast, favoured a more realistic and less reductive model of unity which emphasized a unity of language and the local exchanges of scientific tools, although it was not constrained by Carnap's ideals of conceptual precision, deductive systematicity and logical rigour (Neurath spoke of a 'boat', a 'mosaic', an 'orchestration', a 'universal jargon'). In 1934 Neurath spearheaded a movement for Unity of Science that encouraged international cooperation among scientists and launched the project of an *International Encyclopedia of Unified Science*. For both Carnap and Neurath the ideal of unified science against metaphysics had deep social and political significance.

Two new ideas by logical positivists in the USA after the Second World War again placed the question of unity of science at the core of philosophy of science: Carl Hempel's deductive-nomological model of explanation and Ernest Nagel's (1961) model of reduction (see Explanation; Reduction, problems of §§2-3). Hempel's model characterizes the scientific explanation of events as a subsumption under an empirically testable generalization. In the 1950s, when positivism was extending to the social sciences, the model was offered as a criterion of demarcation. Explanations in the historical sciences too must fit the model if they are to count as scientific (see Positivism in the social sciences §2). The applicability of Hempel's model was soon contested - notably by W. Dray - and this opened a debate about the nature of the historical sciences that remains unresolved. In the process, some have claimed as historical natural sciences the like of geology and biology. It has been argued that Hempel's model, especially the requirement of empirically testable strict universal laws, is satisfied neither in the physical sciences nor the historical sciences, including biology (Cartwright 1983; Ehreshefsky 1992).

Since Nagel's influential model of reduction by derivation, most discussions of unity of science have been cast in

terms of reductions between concepts and between theories. The hierarchy of levels of reduction was set by the levels of aggregation of entities all the way down to atomic particles, thus rendering microphysics the fundamental science (Oppenheim and Putnam 1958). Criticism for the rejection of Nagel's model and its emendations has occupied the last three decades of philosophical discussions of the unity of physics, and especially in psychology and biology. In the latter the availability of laws and theories and the feasibility of global reductions have been contested (see Reductionism in the philosophy of mind). The possibility and the necessity of reductions in a number of forms have been discussed in connection with issues of explanation, realism and scientific progress (see Naturalized philosophy of science §1). Debates on unification and reduction have taken place also within the sciences as renewed reductionistic ideals have been revived in physics, driven by the success of the unified theories of fundamental forces in the 1970s, and in biology - especially sociobiology - driven by successes in molecular genetics (see Molecular biology; Sociobiology).

In anti-reductionist quarters, models of unification without reduction have appeared: for instance, the notion of 'interfield theories' (Darden and Maull 1977) and, less theory-oriented, the emphasis on the plurality of kinds of unity (Hacking 1996) and the idea of science as a historical process, manifesting a historical unity with a Darwinian-like pattern of evolution (Hull 1988). A more radical departure is the more recent criticism of the methodological values of reductionism and unification in science - also as embedded in society and affecting it. This view argues for the replacement of the emphasis on global unity - including unity of method - by the emphasis on disunity, epistemological and ontological pluralism and local cooperation. The suggestion is that the notion of a science should be understood, now following the later Wittgenstein, as at best a family resemblance concept (Dupré 1993).

JORDI CAT

References and further reading

D'Alembert, J. and Diderot, D. (eds) (1751-72) *Encyclopédie, ou dictionnaire raisonné des sciences, des arts at des metiers*, Paris: Plon.(This is the classic paradigm of encyclopedic work, with especial significance in the context of the culture of the enlightenment.)

Carnap, R., Morris, C. and Neurath, O. (eds) (1938-70) *International Encyclopedia of Unified Science*, vols 1 and 2, Foundations of the Unity of Science, Chicago, IL: University of Chicago Press.(The *Encyclopedia* was planned as a series of monographs applying the philosophy of logical empiricism to the different sciences and as a way of showing and promoting cross-bearings. Publication ended in 1970 and, ironically, it contains as one of the later pieces Kuhn's *The Structure of Scientific Revolutions*.)

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Oxford University Press.(Criticizes the application of Hempel's deductive-nomological model of explanation in physics.)

Cohen, I.B. (1995) *Interactions*, Cambridge, MA: MIT Press.(Historical examination of the different mechanistic and organistic analogies and metaphors used in the development of the social sciences.)

Darden, L. and Maull, N. (1977) 'Interfield Theories', *Philosophy of Science* 44: 43-64.(Against standard models of theory-reduction, it proposes a model of unification, especially in biology, in which theories do not identify fields but rather bring them together.)

Dupré, J. (1993) *The Disorder of Things. Metaphysical Foundations of the Disunity of Science*, Cambridge, MA: Harvard University Press.(Powerful criticism of reductionism and essentialism, as well as other formulations of the unity of science; a *fin-de-siècle* classic.)

Ehreshefsky, M. (1992) 'The Historical Nature of Evolutionary Theory', in M.H. Nitecki and D.V. Nitecki (eds) *History and Evolution*, New York: State University of New York Press. (Claims that neither the physical nor the biological sciences can be adequately brought together under Hempel's model of explanation).

Galison, P. and Stump, D. (eds) (1996) *The Disunity of Science. Boundaries, Contexts and Power*, Stanford, CA: Stanford University Press.(A very interesting collection of articles with an emphasis on disunity and cultural contexts.)

Hacking, I. (1996) 'The Disunities of Science', in P. Galison and D. Stump (eds) *The Disunity of Science. Boundaries, Contexts and Power*, Stanford, CA: Stanford University Press, 1996.(Points to the number of different possible kinds of unities).

Hull, D. (1988) *Science as Progress*, Chicago, IL: University of Chicago Press.(An attempt to apply evolutionary ideas to the dynamics of scientific practice.)

McRae, R. (1961) *The Problem of the Unity of the Sciences: Bacon to Kant*, Toronto, Ont.: University of Toronto Press.(The only monograph on the issue in Bacon, Descartes, Leibniz and Kant; very general; it omits Hume

and it does not address the relation between natural and social sciences.)

Mayr, E. (1982) *The Growth of Biological Thought*, Cambridge, MA: Harvard University Press.(A classic in philosophy of biology that criticizes physicists' reductionism, especially Weinberg's, and defends the concept of emergence in biology.)

Nagel, E. (1961) *The Structure of Science*, London: Routledge.(Chapter 11 presents the author's account of reduction.)

Oppenheim, P. and Putnam, H. (1958) 'The Unity of Science as a Working Hypothesis', in H. Feigl *et al.* (eds) *Minnesota Studies in the Philosophy of Science*, vol. 2, Minneapolis, MN: Minnesota University Press.(A standard formulation of the metaphysics of reductionism as a hierarchy of objects in the scientific domain.)

Primas, H. (1983) *Chemistry, Quantum Mechanics and Reductionism*, Berlin: Springer.(A criticism of the straightforward reducibility of chemistry to quantum mechanics.)

Schaffner, K. (1993) *Discovery and Explanation in Biology and Medicine*, Chicago, IL: University of Chicago Press.(Presents a survey of the literature and a discussion of the applicability of different notions of reductionism in the biomedical sciences, especially his own; it includes a very helpful bibliography with essential primary sources.)

Universal language

Most often associated with attempts to establish an international language such as Esperanto, the idea of a universal language is rooted in the biblical claim of an original language common to all human beings. The idea received its most thorough investigation during the seventeenth century. Drawing on the example of Chinese characters, early schemes involved a system of written signs that would allow communication between speakers of different languages. Later thinkers argued for the importance of an ideal 'philosophical language' in which the structure of signs exactly mirrored the structure of reality. While such projects fell short of their authors' expectations, their influence can be discerned in the formalisms of modern logic and science.

1 Adamic language

The earliest source for the idea of a universal language is the biblical story of an original language used by Adam to name the different species of animals created by God (*Genesis* 2: 19-20). In both the Jewish and Christian traditions, it was commonly assumed that this 'Adamic' language expressed Adam's perfect knowledge of things prior to the Fall: each name of a creature exactly conveyed its essence. It was natural to suppose a corruption of this language at the Fall; however, the crucial event in the loss of a universal language is told in the story of Babel, when 'the Lord confused the language of all the earth' (*Genesis* 11: 9). Thereafter, human beings were condemned to speak a multitude of languages and to suffer the pain of mutual incomprehension.

This biblical background remained central to conceptions of a universal language well into the seventeenth century. Numerous attempts were made to identify and recover the original language of Adam. Among Jewish and Christian thinkers alike, Hebrew retained the greatest claim to this honour, although other candidates were proposed (including Latin, Chinese, Dutch and Swedish). The attempt to recover the Adamic language, and the knowledge implicit in it, was significantly influenced by the Jewish kabbalah, which assigned a mystical significance to the letters of the Hebrew alphabet and prescribed techniques for their interpretation and manipulation (see Kabbalah). Another important source was the thirteenth-century theologian Ramon Llull, who stressed a method of achieving universal knowledge through combinations of letters signifying fundamental categories of reality (see Llull, R. §2).

2 Artificial schemes

During the seventeenth century, there occurred an explosion of interest in universal language schemes, conceived for the first time as systems of artificial signs or 'characters' constructed by human beings as a means of overcoming the limitations of natural languages. The causes of this newfound fascination with universal languages are complex. Clearly, a role must be assigned to the growing importance of vernacular languages and the decline of Latin as a shared medium of commerce, scholarship and diplomacy. An increase in millenarian religious sentiment, particularly in England, renewed the call for a recovery of the Adamic language. Finally, and most importantly, there was a growing recognition of the significance of language as a factor in the acquisition of scientific knowledge.

One of the earliest and most influential statements on the topic was Francis Bacon's reference in *The Advancement of Learning* (1605 II: ch. 16) to languages such as Chinese, whose users 'write in characters real, which express neither letters nor words in gross, but things or notions'. To many seventeenth-century thinkers, Chinese provided a model of what a universal language might be: a single set of characters that could be pronounced differently in different languages, but which when written would offer a shared basis for understanding. Implicit in Bacon's remark, however, was a further idea that pointed toward the scientific value of such a language. In his *New Organon* (1620 I: aph. 59), Bacon complained of how words commonly obscure 'the true divisions of nature'. It was a short step from this to the idea that these 'true divisions' might be better represented in a language composed of 'real characters', which directly expressed 'things or notions' (see Bacon, F. §§4-5).

Most early proponents of artificial language schemes stressed the practical value of their inventions as instruments of communication. Their works were primarily attempts to devise a system of writing, modelled variously on Chinese characters, cryptographic codes or shorthand notation, whereby synonymous words in different languages would be represented by a common sign. It was not long, however, before the more ambitious idea of a

'philosophical language' took hold. In a 1629 letter to Marin Mersenne, Descartes had already expressed scepticism concerning the usefulness of artificial languages of the first sort. However, he went on to postulate another kind of language in which ideas would be represented so clearly that errors of judgment would be 'almost impossible'. To realize such a language, all of our thoughts would first have to be given a proper order 'like the natural order of the numbers'; and this presupposes the 'true philosophy', by which the analysis and ordering of thoughts would be carried out. Although Descartes pursues the plan no further, he is optimistic that 'such a language is possible and that the knowledge on which it depends can be discovered'.

A philosophical language of the sort envisioned by Descartes is described in Mersenne's *Harmonie Universelle* (1636) and was pursued in many later works, culminating in George Dalgarno's *Ars Signorum* (1661) and John Wilkins' *An Essay Towards a Real Character and a Philosophical Language* (1668). Such schemes typically consisted of two parts: a system of categories summarizing the 'true divisions' of nature, and a set of characters suitable for representing these categories and the elements within them. Although informed by the discoveries of seventeenth-century science, the first part of the scheme was strongly indebted to the systems of categories propounded by Aristotle and medieval philosophers. In his *Essay*, Wilkins begins with forty genera (classified as transcendentals, substances, quantities, qualities, actions and relations), each of which he subdivides into its 'proper differences and species'. He then proposes two ways of representing the composition of concepts from their respective genus, difference and species: (1) a real character, or system of ideographic signs formed from combinations of vertical and horizontal lines; (2) a speakable philosophical language, consisting of novel combinations of syllables, consonants and vowels. In both cases, additional signs must be added to play the role of particles underwriting the grammatical structure of the language.

3 Leibniz

The idea of a universal language received its final significant development in the seventeenth century in the writings of Gottfried Wilhelm Leibniz. Although indebted to the efforts of Dalgarno and Wilkins, Leibniz criticizes these authors for their reliance on what he sees as an arbitrary number of basic categories. Harking back to Descartes, Leibniz locates the key to a philosophical language in an analysis by which every concept would be broken down into its simplest elements, the 'alphabet of human thoughts'. With these simples identified, and appropriate characters assigned to each, we would, in effect, possess the Adamic language: one in which the structure of every character perfectly mirrors the structure of reality.

Leibniz's plans for a universal language are among the most ambitious devised; in practice, however, they met with limited success. Early on, he recognized that an analysis of concepts into ultimate simples lies beyond the power of the human mind. What we are left with, then, are classificatory systems, like that of Wilkins, which represent our best attempts to order reality as we conceive it. Leibniz's greatest achievement in the area lies in his *specieuse générale* or 'general science of forms', a collection of calculi designed to support the formalization of reasoning in every branch of knowledge. Anticipating twentieth-century developments in logic and methodology, Leibniz's studies represent one of the enduring legacies of the universal language movement (see Leibniz, G.W. §10).

See also: Formal languages and systems

DONALD RUTHERFORD

References and further reading

Bacon, F. (1605) *The Advancement of Learning*, ed. G.W. Kitchin, London: Dent, 1976.(Bacon's discussion of 'characters real' occurs in book 2, chapter 16.)

Bacon, F. (1620) *The New Organon*, ed. F.H. Anderson, New York: Macmillan, 1960. (Bacon criticizes words as 'idols of the marketplace' in book I, aphorisms 43, 59-60.)

Dalgarno, G. (1834) *The Works of George Dalgarno*, Edinburgh; repr. New York: AMS Press, 1971.(Contains all of Dalgarno's writings on artificial language schemes, including his 1661 *Ars Signorum*.)

Descartes, R. (1991) *The Philosophical Writings of Descartes*, vol. 3 (The Correspondence), trans. J. Cottingham, R. Stoothoof, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press.(Descartes' letter to Mersenne is dated 20 November 1629.)

Katz, D.S. (1982) *Philo-Semitism and the Readmission of the Jews to England (1603-1655)*, Oxford: Clarendon Press.(Chapter 2 argues persuasively for the importance of the idea of an Adamic language in

seventeenth-century England.)

Knowlson, J. (1975) *Universal Language Schemes in England and France 1600-1800*, Toronto, Ont.: University of Toronto Press.(The best survey of the topic; contains a valuable checklist of universal language schemes in the seventeenth and eighteenth centuries.)

Leibniz, G.W. (1966) *Logical Papers*, trans. and ed. G.H.R. Parkinson, Oxford: Clarendon Press.(The fullest collection in English of Leibniz's writings on logic and language; includes selections from his 1666 essay *On the Art of Combinations*, in which appears his earliest plan for a universal language.)

Mersenne, M. (1636) *Harmonie Universelle*, Paris; repr. Paris: Centre National de la Recherche Scientifique, 1986.(Book I, proposition 24 contains Mersenne's proposal for a 'perfect language', in which mathematical and scientific truths would be expressed by combinations of musical tones.)

Rutherford, D. (1995) 'Philosophy and Language', in N. Jolley (ed.) *The Cambridge Companion to Leibniz*, New York: Cambridge University Press.(A comprehensive review of Leibniz's interests in language, including the idea of a universal language.)

Salmon, V. (1988) *The Study of Language in 17th-Century England*, Amsterdam: John Benjamins.(A collection of ground-breaking essays on the projects of Dalgarno, Wilkins and their contemporaries; includes a bibliography of recent work in the area.)

Slaughter, M.M. (1982) *Universal Languages and Scientific Taxonomy in the Seventeenth Century*, Cambridge: Cambridge University Press.(An important study of the relationship between seventeenth-century scientific thinking and the idea of a universal language.)

Wilkins, J. (1668) *An Essay towards a Real Character, and a Philosophical Language*, London; repr. in the series *English Linguistics 1500-1800*, number 119, Menston: Scolar Press, 1968.(Wilkins' attempt at constructing a universal language, discussed in §2 above.)

Universalism in ethics

The claim that ethical standards or principles are universal is an ancient commonplace of many ethical traditions and of contemporary political life, particularly in appeals to universal human rights. Yet it remains controversial. There are many sources of controversy. Universalism in ethics may be identified with claims about the form, scope or content of ethical principles, or with the very idea that ethical judgment appeals to principles, rather than to particular cases. Or it may be identified with various claims to identify a single fundamental universal principle, from which all other ethical principles and judgments derive. These disagreements can be clarified, and perhaps in part resolved, by distinguishing a number of different conceptions of universalism in ethics.

1 Form and scope: principles for everybody

One distinctive understanding of universalism in ethics is that ethical principles are principles *for everybody*. They prescribe obligations for everybody, define rights for everybody, list virtues for everybody. The most minimal version of ethical universalism is a claim about the *form* of ethical principles or standards. It is the claim that ethical principles hold for all and not merely for some, that is, for everybody without exception.

Those who hold that ethical principles are universal in form often disagree about their *scope*, that is to say about which beings comprise ‘everybody’. Plato’s character Meno tells Socrates that there are quite different virtues for men and women, for boys and girls, for old men and slaves (*Meno* 71e). On the other hand, Cicero famously asserted that ‘there will not be different laws at Rome and at Athens, now and in the future, but one eternal and unchangeable law for all nations and all times’ (*De Republica* III, 33) (see Cicero, M.T. §2); and St Paul proclaimed that ‘there is neither Jew nor Greek, there is neither bond nor free, there is neither male nor female: for ye are all one in Christ Jesus’ (Galatians 3:28). One very influential understanding of universalism in ethics, shared by many religions, by the natural law and liberal traditions, and by many others, is the contention that ethical principles are universal in form and cosmopolitan in scope, in that they hold for all humans.

However, any cosmopolitan view of the scope of ethical principles must note that living up to obligations, virtues and even some rights is impossible for human beings who lack mature capacities for action. Neither infants nor small children, neither the retarded nor the senile, can be held accountable for carrying out obligations, living virtuously or exercising certain rights, such as political rights. Yet humans who lack these capacities might have other rights, for example, to care and protection. Those who can suffer but not act can be moral patients but not moral agents, possessing some rights, but not the full range of obligations, virtues or rights. The scope of different sorts of principles of universal form will evidently have to vary (see Moral agents; Responsibility).

Many think that the scope of some ethical principles is more-than-cosmopolitan. Jeremy Bentham famously declared that the criterion of moral standing was ‘not, Can they *reason*? nor, Can they *talk*? but Can they *suffer*?’ (1789: 412, footnote; original emphasis); Hindus and Buddhists too extend moral concern beyond humankind; some environmentalists extend concern not only to nonhuman animals, but to plants, even to species and habitats (see Animals and ethics; Bentham, J. §2; Duty and virtue, Indian conceptions of; Environmental ethics; Moral standing §2-3).

Other advocates of principles of universal form join Meno in holding that their scope is less-than-cosmopolitan. For example, communitarians (who sometimes describe themselves as *rejecting* universal principles) advocate principles of universal form whose scope is restricted to particular communities (see Community and communitarianism); some virtue ethicists hold similar views (see Virtue ethics §5).

2 Content: formal principles or uniform requirements?

A second conception of universalism in ethics emphasizes the content as well as the form and scope of principles. Principles which hold for everybody will prescribe or recommend the same for everybody (same obligations, same rights, same virtues and so on). Advocates of universal principles see this as a merit: they see equality of requirement and entitlement as ethically important (see Equality §3). For example, discussions of universal human rights emphasize not only that all humans have rights, but that they all have the same rights.

Two objections are commonly raised. The first is that principles which prescribe the same for all will be abstract

and general, so provide too little guidance. The second is that they will be too demanding and specific, prescribing with senseless and heartless uniformity for differing cases and situations. On this account, universal principles are either too formal and minimal or else too uniformly demanding. Evidently the two criticisms cannot both be true of one and the same universal principle. If a principle is so abstract that it provides no practical guidance, then it will not prescribe rigid uniformity of action; conversely, if it prescribes with rigid uniformity it will not fail to guide action.

The charge that ethical principles which prescribe the same for all abstract from differences between cases is true, but not damaging. No principle of action - whether of universal or non-universal form, whether of cosmopolitan or lesser scope - can prescribe with total specificity; even very explicit principles abstract from many circumstances. It follows that principles of action can always be satisfied in varied ways. A principle such as ‘Tell the truth’ does not prescribe what we must say to whom or when; a principle such as ‘Pay your debts’ does not determine the means or manner of repayment. Principles of action, including ethical principles, *constrain* action or entitlements, rather than picking out a single, wholly determinate line of action. Abstract principles can therefore guide action yet allow for flexible interpretation or application that takes account of differences between cases. So an ethics of universal principles can readily avoid both barren formalism and doctrinaire rigorism.

3 Universalism and particularism: principles and judgment

Since universal ethical principles are always to some extent abstract or indeterminate they must be supplemented by *judgment* in selecting among possible implementations. This point is recognized, indeed stressed, by advocates of universal principles. Kant (1781/1787), for example, insisted that there cannot be complete rules for judgment, and that principles cannot entail their determinate applications, which require judgment (see Kant, I. §12).

The serious disagreement lies not between those who think that ethics needs *only* principles and those who think it needs *only* judgment of particular cases, but between those who think principles and judgment are both needed and those who believe that judgment alone will be enough. The deepest opposition to any sort of ethical universalism comes from ethical particularists who hold that unmediated apprehension of particular cases can guide ethical life. Ethical particularists seek to anchor ethical judgment in perception of and responsiveness to the particular, in attentiveness to the case at hand, in the salience of the personal relationship and its claims (see Friendship; Impartiality §4). They usually hold that ethical life revolves around character and virtue. The most radical cast doubt on the very conception of following a rule or principle; the less radical cite the issues of §2 as evidence that an ethics of rules and duties is inadequate (see Moral particularism).

Both ethical particularists (who appeal only to judgment) and universalists (who argue that judgment is used in combination with principles) have found it difficult to explain how judgment works. Some particularists describe it as analogous to perceiving or attending or to the exercise of a craft skill. Some universalists see judgment as the skill of identifying acts that fall within the constraints set by a plurality of principles (see Moral judgment).

4 Fundamental principles: ‘golden rules’

Other conceptions of universalism in ethics combine views of the form, scope and sameness of content of principles with ambitious claims that a single fundamental universal principle provides the basis for all derivative ethical principles and ultimately for ethical judgment of particular cases.

Often the proposed fundamental principle is a version of a ‘golden rule’. Variously formulated golden rules are found in Hindu and Confucian sacred texts, and in many other traditions, including natural law and popular ethical debate. One well known golden rule is Christian with Jewish antecedents: ‘Do unto others as you would that they should do unto you’ (for specific formulations see Matthew 7:12, Luke 6:31; for antecedents Tobias 4:15). Others are prohibitions rather than injunctions, such as ‘Do not do unto others what you would not have them do unto you’.

These would-be foundational principles have been criticized for linking ethics too closely to agents’ desires or consent. Why should willingness to be on the receiving end of like action make it permissible? If masochists are willing to suffer others’ sadism, would that make sadism right? More generally, can acceptance of being on the receiving end of like action legitimate anything?

This problem can be overcome only by building additional constraints or complexities into the idea of considering what one would desire or consent to when putting oneself into another's shoes. This has been attempted in various principles, which are first cousin to golden rules, that have been influential in secular work in ethics. Most famously J.S. Mill asserted, in *Utilitarianism* (1861), that 'in the golden rule of Jesus of Nazareth, we read the complete spirit of the ethics of utility'. The link Mill draws between utilitarianism and golden rules arises only if agents consider not what they as individuals would want if on the receiving end, but what they if taking account of all others' desires would want if on the receiving end. Only then can a golden rule reflect everybody's desires, and so be thought of as aiming at the greatest happiness (see Utilitarianism; Mill, J.S. §10).

Other approaches of this sort have recently been advocated by P. Singer (1972), R.M. Hare (1975) and A. Gewirth (1987), each of whom recognizes affinities as well as differences between his proposal for the foundations of ethics and traditional golden rules. For example, Gewirth suggests that a rational golden rule would read 'Do unto others as you have a right that they do unto you', while Hare advocates a universal prescriptivism by which the fundamental criterion for ethical judgment is that agents be willing to universalize their judgments, that is extend them to all situations identical in their universal properties (see Prescriptivism; Hare, R.M. §§1-2). There has been much discussion of the plausibility of these proposals, which generally reject the emphasis traditional golden rules give to what one would have if the particular victim reciprocated, and introduce some reference to what one would want if one's own principle were to be universally adopted or if one's desires took account of others' desires. These writers advocate a strong form of ethical universalism: not merely do they defend a single fundamental ethical principle, but they insist that it refer to the desires that all hold, or ought if rational to hold.

5 Fundamental principles: Kantian universalizability

An alternative conception of universalism in ethics rejects golden rules and seeks to anchor all ethical justification in a more formal fundamental universal principle, which does not refer to desires or consent to fix the content of ethics. The most famous and most ambitious attempt to go further is Kant's 'categorical imperative', of which the best known version runs: 'Act only on that maxim through which you can at the same time will that it should become a universal law' ([1785] 1903: 421). Kant claims to show that 'all imperatives of duty can be derived from this one imperative as their principle' (421). He insists that in such derivations no reference be made either to anyone's happiness or desires, consent or agreement, and that the categorical imperative is not a version of a golden rule (which he dismisses as trivial, 430, footnote). Kant's views have been influential: a German scholar recently commented that 'Kant succeeded with his objection almost in invalidating the golden rule and disqualifying it from future discussion in ethics' (Reiner 1983: 274).

English language philosophy has been less convinced that Kant undermined golden rule approaches. J.S. Mill was neither the first nor the last to think that Kant's claim to derive all principles of duty from the categorical imperative was complete nonsense. He wrote of Kant

> when he begins to deduce from this precept any of the actual duties of morality, he fails, almost grotesquely, to show that there would be any contradiction, any logical (not to say physical) impossibility, in the adoption by all rational beings of the most outrageously immoral rules of conduct. All he shows is that the *consequences* of their universal adoption would be such as no one would choose to incur.
>
> (1861: 207; original emphasis)

There has been widespread scepticism about Kant's supposed claim to show that 'immoral rules of conduct' are self-contradictory. However, he in fact makes the more circumspect *modal* claim that we should not act on principles which we *cannot* simultaneously 'will as universal laws'. An example of such a principle is that of false promising. Kant holds that false promisers who try (incoherently) to will false promising as a universal law thereby will the destruction of the very trust on which their own attempts to promise falsely must rely. Hence when we try to act on such principles Kant holds that

> we in fact do not will that our maxim (principle) should become a universal law - since this is impossible for us - but rather that its opposite should remain a law universally: we only take the liberty of making an *exception* to it for ourselves (or even just for this once).
>
> ([1785] 1903: 424; original emphasis)

In 'deriving' an 'actual principle of duty' from the categorical imperative, Kant takes it that agents not only seek principles of universal form and cosmopolitan scope which prescribe the same for all, but shun any principles which cannot be 'willed for all'. Kantian justifications of such principles, unlike golden rule justifications, do not appeal to either the desires, the happiness or the acceptance of those on the receiving end, nor indeed to actual or hypothetical desires of any or of all agents. The distinctive modal character of Kantian universalizability is its appeal to what *can be willed for all* (rather than to what *actually is* or *hypothetically would be willed by all*). It remains a matter of considerable controversy whether a strictly Kantian approach can be used to construct an account of specific principles of duty, virtue or entitlement, or whether it is indeed too formal and minimal to sustain these derivations.

See also: Critical Theory; Intuitionism in ethics; Theological Virtues

ONORA O'NEILL

References and further reading

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, revised F. Rosen, Oxford: Clarendon Press, 1996.(Classic statement of utilitarianism.)

Cicero, M.T. (54-51 BC) *De Republica*, trans. M. Grant, *On Government*, Harmondsworth: Penguin, 1993, books III, V and VI.(Early form of universalism about scope of moral principles.)

Gewirth, A. (1987) 'The Golden Rule Rationalized', *Midwest Studies in Philosophy* 3: 133-44.(Modern form of universalism, with affinities to golden rules.)

Gewirth, A. (1988) 'Ethical Universalism and Particularism', *Journal of Philosophy* 85: 283-301.(A universalist approach to ethical judgment.)

Hare, R.M. (1963) *Freedom and Reason*, Oxford: Clarendon Press.(Basic source on universal prescriptivism.)

Hare, R.M. (1975) 'Abortion and the Golden Rule', *Philosophy and Public Affairs* 3: 201-22.(Application of universal prescriptivism to problem of abortion; includes discussion of golden rules.)

Herman, B. (1993) *The Practice of Moral Judgment*, Cambridge, MA: Harvard University Press.(Insightful discussions of difficulties raised about Kantian ethics.)

Kant, I. (1781/1787) *Kritik der Reinen Vernunft*, trans. N. Kemp Smith, *Critique of Pure Reason*, London: Macmillan, 1973.(*Locus classicus* for Kant's insistence that there cannot be complete rules for judgment.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, in *Kants gesammelte Schriften*, ed. Königlichen Preußischen Akademie der Wissenschaften, Berlin: Reimer, vol. 4, 1903; trans. H.J. Paton, *Groundwork of the Metaphysics of Morals* (originally *The Moral Law*), London: Hutchinson, 1948; repr. New York: Harper & Row, 1964.(References made to this work in the entry give the page number from the 1903 Berlin Akademie volume; these page numbers are included in the Paton translation. Classic, short, if difficult, exposition of Kant's ethics.)

McDowell, J. (1979) 'Virtue and Reason', *Monist* 62 (3): 331-50; revised version repr. as 'Non-cognitivism and Rule Following', in S. Holtzman and C. Leach (eds) *Wittgenstein: To Follow a Rule*, London: Routledge & Kegan Paul, 1981. 141-62.(An influential statement of a radical particularist position, which questions the very possibility of following rules or principles.)

Mill, J.S. (1861) *Utilitarianism*, in J.M. Robson (ed.) *Collected Works of John Stuart Mill*, vol. 10, *Essays on Ethics, Religion and Society*, Toronto: University of Toronto Press, 1969.(Account of utilitarianism expanding on Bentham.)

O'Neill, O. (1987) 'Abstraction, Idealization and Ideology in Ethics', in J.D.G. Evans (ed.) *Moral Philosophy and Contemporary Problems*, Cambridge: Cambridge University Press.(What is abstraction? Is it avoidable? Is it harmful? In asking these questions, this essay is particularly relevant to §§2 and 3.)

O'Neill, O. (1989) *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge: Cambridge University Press.(Papers on patterns of universalist ethical reasoning that use Kant's categorical imperative.)

O'Neill, O. (1991) 'Kantian Ethics', in P. Singer (ed.) *A Companion to Ethics*, Oxford: Blackwell, 175-85. (Overview of Kant's position and some well-known criticisms).

Plato (*c*.386-380 BC) *Meno*, in *Protagoras and Meno*, trans. W.K.C. Guthrie, Harmondsworth: Penguin, 1956. (Discusses the nature of virtue.)

Potter, N. and Timmons, M. (eds) (1985) *Morality and Universality: Essays on Ethical Universalizability*, Dordrecht: Reidel.(Useful papers on different conceptions of universality in ethics; large bibliography.)

Reiner, H. (1983) 'The Golden Rule and the Natural Law', in *Duty and Inclination: The Fundamentals of Morality Discussed and Redefined with Special Regard to Kant and Schiller*, trans. M. Santos, The Hague: Nijhoff, 271-93.(Historical overview; useful references especially to German literature.)

Singer, P. (1972) ‘Famine, Affluence and Morality’, *Philosophy and Public Affairs* 1: 229-43.(Well-known utilitarian argument to show that beneficence should have cosmopolitan scope: the affluent should help the hungry however far away they may be.)

Wattles, J. (1996) *The Golden Rule,* Oxford: Oxford University Press.(Examines the principle ‘Do to others as you want others to do to you’ in contexts of psychology, philosophy and religion, from Confucius, Hillel and Jesus to R.M. Hare and Paul Ricoeur.)

Wiggins, D. (1980) ‘Deliberation and Practical Reason’, in *Needs, Values, Truth: Essays in the Philosophy of Value,* Aristotelian Society Series, vol. 6, Oxford: Blackwell, 1987; revised edn, 1991.(A particularist reading of Aristotle on judgment.)

Williams, B. (1985) *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press and London: Fontana. (Particularist criticism of aspects of ethical universalism.)

Universals

In metaphysics, the term 'universals' is applied to things of two sorts: properties (such as redness or roundness), and relations (such as kinship relations like sisterhood, or the causal relation, or spatial and temporal relations). Universals are to be understood by contrast with particulars. Few universals, if any, are truly 'universal' in the sense that they are shared by all individuals - a universal is characteristically the sort of thing which some individuals may have in common, and others may lack.

Universals have been conceived to be things which enable us intellectually to grasp a permanent, underlying order behind the changing flux of experience. Some of the gods of ancient mythologies correspond roughly to various important underlying universals - social relations for instance, as for example if Hera is said to be the goddess of Marriage and Ares (or Mars) is said to be the god of War. Many traditions, East and West, have dealt with the underlying problem which generates theories of universals; nevertheless the term 'universals' is closely tied to the Western tradition, and the agenda has been set largely by the work of Plato and Aristotle.

The term often used in connection with Plato is not 'universals' but 'Forms' (or 'Ideas', used in the sense of ideals rather than of thoughts), the term 'universals' echoing Aristotle more than Plato. Other terms cognate with universals include not only properties and relations, but also qualities, attributes, characteristics, essences and accidents (in the sense of qualities which a thing has not of necessity but only by accident), species and genus, and natural kinds.

Various arguments have been advanced to establish the existence of universals, the most memorable of which is the 'one over many' argument. There are also various arguments against the existence of universals. There are, for instance, various vicious regress arguments which derive from Aristotle's so-called 'third man argument' against Plato. Another family of arguments trades on what is called Ockham's razor: it is argued that we can say anything we need to say, and explain everything we need to explain, without appeal to universals; and if we can, and if we are rational, then we should. Those who believe in universals are called Realists, those who do not are called Nominalists.

1 Sources in ancient mathematics and biology

Plato looked to mathematics as a model to find ideal 'forms' which can be grasped by the intellect and which we find to be imperfectly reflected in the world of the senses. Moral and political ideals too, Plato thought, are reflected only very imperfectly in the world of appearances. Aristotle's conception of universals was tailored to fit not mathematics but biology. Individual animals and plants fall into natural kinds, or species, such as pigs or cabbages. Various different species, in turn, fall under a genus.

Universals impose a taxonomy on the plurality of different individuals in the world. Regularities in the world can then be understood by appeal to the universals, or species, under which individuals fall, explaining why pigs never give birth to kittens, for instance, and in general why each living thing generates others of its kind.

Plato conceived of universals as transcendent beings, *ante rem* in Latin ('before things'): the existence of universals does not depend on the existence of individuals which instantiated them. This is a natural thought if your model of universals lies in mathematics: geometrical truths about circles, for instance, do not depend on the existence of any individuals which really are perfectly circular. Aristotle, in contrast, held a theory of universals as immanent beings, *in rebus* ('in things'): there can be no universals unless there are individuals in which those universals are instantiated. This is a natural thought if your model of universals lies in biology: a species cannot exist, for instance, if there are no animals of that species. Thus, one of the key distinctions between Plato's transcendent and Aristotle's immanent realism is that the Platonist allows, and Aristotle does not allow, the existence of uninstantiated universals (see Aristotle §15; Plato).

2 Samenesses and differences

When a property is shared by two individuals, there is something which is in or is had by both. But it is in a quite distinctive sense that one universal can be 'in' two distinct individuals. An individual person may be 'in' two places at once if, for instance, their hand is in the cookie jar and their foot is in the bath. But a universal is 'in'

distinct individuals in a way which does not mean that there is one part of the universal in one thing and a distinct part of it in another. Thus, a universal is said to be the sort of thing which can be wholly present in distinct individuals at the same time: a person cannot be wholly present in two places at once, but justice can.

Some draw a distinction between certain special properties and relations which qualify for the label ‘universals’, and other properties and relations which do not. It is suggested that, whenever something is true of an individual (whenever a description can truly be predicated of an individual), then there is always a ‘property’ which that individual may be said to have. On this view, a ‘property’ is just a shadow of a predicate, whereas a genuine universal is something more. A genuine universal has to be something which is literally identical in each of its instances. Alternatively, the sorts of ‘properties’ which are just shadows of predicates are sometimes construed as set-theoretical constructions of various sorts, as for instance if we say that the ‘property’ of redness is the set of actual red things, or of actual and possible red things. In this spirit it is now standard practice in mathematics to use the term ‘relations’ to refer just to any set of ordered pairs. Set-theoretical constructions are not, however, universals - or at least they are not to be confused with the universals which are the subject matter of traditional debates.

3 Arguments for and against

Various arguments have been advanced to establish the existence of universals, the most memorable of which is the ‘one over many’ argument. Although it is memorable, there is little consensus on just how this argument works. Very roughly, it begins with an appeal to the manifest fact of recurrence, the fact that, as it says in the biblical text of *Ecclesiastes* (1: 9), ‘What has been is what will be, and what has been done is what will be done; and there is nothing new under the sun’. There are many things, and yet they are all in some sense just the very same things over and over again. From this manifest fact of recurrence, the argument purports to derive the conclusion that there are universals as well as particulars.

There are also various arguments against the existence of universals. One family of such arguments derives from Aristotle’s so-called ‘third man argument’ and is designed to demonstrate that Plato’s Theory of Forms entails an unacceptable infinite regress. Roughly, Plato’s problem is that he needs some relation to hold between the Form of Man and individual men before this Form can help to explain what it is that individual men have in common. So the theory would seem to call into being another Form, a third man, which is what the Form of Man has in common with individual men. This leads to an infinite regress, hence Plato’s Theory of Forms is unacceptable. Of course, Aristotle had only intended to demonstrate the nonexistence of Plato’s Forms, not of universals in general; but enemies of universals frequently advance related infinite-regress arguments against the existence of universals of any kind. Whatever you call the instantiation relation between particulars and universals, if you think of it as another universal then you are off on a regress, and this seems to count against any theory of universals.

Another argument against the existence of universals trades on what is called ‘Ockham’s razor’ - the principle that you should not postulate more entities when everything you want to explain can be explained with fewer (see William of Ockham §2). It is sometimes argued that everything you can explain with universals can be explained just as well without them. Things which superficially seem to refer to universals can, it is maintained, generally be rephrased in ways which make no apparent reference to universals - reference to universals can be paraphrased away. If we can do without universals, then obviously we should; when you supplement this Ockhamist argument with allusions to the interminable and unresolvable internecine conflicts among Realists over numerous details, you have an even stronger case against the existence of universals.

4 Nominalism and Realism

During the Middle Ages in Europe, universals played a focal role in the intellectual economy: many issues revolved around what became known as the problem of universals. Famously, a commentary by Boethius on Porphyry’s *Isagoge*, which in turn was intended as an introduction to Aristotle’s Categories, set very crisply but vividly and tantalizingly what came to be taken as a compulsory question in the Medieval pursuit of learning: whether genera and species are substances or are set in the mind alone; whether they are corporeal or incorporeal substances; and whether they are separate from the things perceived by the senses or set in them (Boethius *c*.510; Spade 1994). The initial problem for many was not one of deciding whether there are any universals, but of choosing between Plato and Aristotle and then fine-tuning further details.

Later in the Middle Ages, however, a growing number of philosophers and theologians became more and more impressed by arguments against the existence of universals. They began to adopt the position called 'Nominalism' which was opposed to all the various forms of Platonic or Aristotelian Realism. According to Nominalists like Abelard and Ockham, the only thing which distinct individuals share in common is a common name, a label which we choose to apply to each of those individuals and not to others.

Nominalistic claims were echoed by many of the champions of the modern sciences as they emerged at the end of the Middle Ages. It was standardly said to be granted on all hands that all existing things are merely particular. Being assumed as granted on all hands, it was not up for debate, and so the problem of universals, explicitly so described, settled into the shadowy background of scientific and philosophical discussion. For example, an archaeologist of ideas might argue that, in Kant, the problem of universals is really alive and working very hard in the background, playing a role in discussions on almost every topic that arises. Nonetheless the problem of universals, under that name or any clear equivalent, is not featured on Kant's explicit agenda. Kant speaks of intuitions and concepts in ways which have some relation to the old problem of particulars and universals, but more has shifted than just the labels. Hence the problem of universals has received little attention across a great span of philosophical history, right through to twentieth-century philosophy in France and Germany (see Nominalism §2; Realism and antirealism).

5 Frege exhumes universals

In the twentieth century, the problem of universals has re-emerged under its familiar name, accompanied by more or less the same guiding illustrations used by Plato and Aristotle. This rebirth has occurred in the tradition of analytic philosophy, notably in the work of Frege, Russell, Wittgenstein, Quine and Armstrong.

A new twist to the theory of universals can be traced to groundbreaking work by Frege on the nature of natural numbers in his *Grundlagen der Arithmetik (The Foundations of Arithmetic)* (1884). As for Plato, so too for Frege, Russell and others in recent times, advances in mathematics have been the source of a philosophical focus on the problem of universals. Frege's analysis of natural numbers (1, 2, 3,…) proceeded in three very different stages (see Frege, G. §9).

In the first stage of his analysis of numbers, Frege introduced the idea that numbering individuals essentially involves not the attribution of properties to individuals but, rather, the attribution of properties to properties. To illustrate: when asked 'How many are on the table?', Frege notes that there will be many different possible answers, as for instance (1) 'Two packs of playing cards' or (2) '104 playing cards'. The metaphysical truth-makers identified by Frege for these two sample answers are (1) that the property of being a pack of playing cards on the table is a property which has the property of having two instances, and (2) that the property of being a playing card on the table is a property which has the property of having one hundred and four instances. In general, natural numbers number individuals only via the intermediary of contributing to second-order properties, or properties of properties, namely properties of the form 'having *n* instances'. Like Kant, Frege speaks of concepts (*Begriffe*) rather than of 'universals'. Yet Frege's concepts are definitively not private mental episodes, but are thoroughly mind-independent, more like Plato's Forms than Aristotelian universals.

In the second stage of his analysis of numbers, Frege gives a very new twist to the theory of universals. He argues that the nature of universals, or concepts, is such as to make it impossible in principle ever to refer to a universal by any name or description. Thus for instance, in saying 'Socrates is wise', the universal which is instantiated by Socrates is something which is expressed by the whole arrangement of symbols into which the name 'Socrates' is embedded to yield the sentence 'Socrates is wise'. Suppose you were to try to name this universal by the name 'wisdom'. Then, compare 'Socrates is wise' with the concatenation of names - 'Socrates wisdom'. The mere name 'wisdom' clearly leaves out something which was present in the attribution of wisdom to Socrates. Hence a universal cannot be referred to by a name.

Thus, a property can only be expressed by a predicate, never by a name or by any logical device which refers to individuals. Indeed, if we wish to attribute existence to universals, we cannot do so by the use of the same sort of device (the first-order quantifier) which is used to attribute existence to individuals. Thus, for instance, from 'Socrates is wise' we may infer 'There exists something which is wise', and 'There exists something which is Socrates':

$(\exists x)(\text{wise}(x))$, and
$(\exists x)(x = \text{Socrates})$.

Yet we may not infer that ‘There exists something which Socrates possesses’, or that ‘There exists something which is wisdom’:

$(\exists x)(\text{has}(\text{Socrates}, x))$, or
$(\exists x)(x = \text{wisdom})$.

Frege does, however, allow us to attribute existence to universals, using logical devices called higher-order quantifiers, which he introduced in his *Begriffsschrift* (1879). That is, we can infer from ‘Socrates is wise’ to ‘There is somehow such that: Socrates is that-how’:

(E*f*)(*f*(Socrates)).

But although there is somehow that Socrates is, this does not entail that there is anything which is the somehow that Socrates is: universals (concepts) can only have second-order existence, not first-order existence.

For Frege, numbering things essentially involved attribution of properties to properties. So the sorts of things being attributed are not the sorts of things which can be named. Yet, Frege argued, numbers can be named - numbers are abstract individuals, he says, objects not concepts. Hence the third stage of Frege’s analysis of numbers consists in the attempt to find individuals - objects - which could be identified with the numbers. It was this stage of the analysis which resulted in the emergence of modern set theory. For every property, Frege argued, there is a corresponding individual: the extension of that universal, the set of all the things (or all the actual and possible things) which instantiate that universal. Thus, for instance, corresponding to the property of being a property which has two instances, there will be a set of sets which have two members. Modern mathematics has selected different candidates for identification with the natural numbers, but it has followed Frege hook, line and sinker with respect to the broad strategy of identifying numbers, and functions and relations, with sets.

Frege’s legacy has significantly changed the agenda for any theory of universals which, like Plato’s, aspires to do justice to mathematics. It leaves three courses open for exploration. One course is that charted by Quine (1953, 1960), of allowing the existence of sets but not of any other nameable things which might be called universals, nor of any of Frege’s higher-order, unnameable universals. Another course is that of allowing the existence of nameable things other than sets: this is a course charted, for example, by Armstrong (1978). A third course allows also the irreducible significance of higher-order quantification (Boolos 1975; Bigelow and Pargetter 1990).

See also: Abstract objects; Particulars; Universals, Indian theories of

JOHN C. BIGELOW

References and further reading

Armstrong, D.M. (1978) *Universals and Scientific Realism*, London: Cambridge University Press, 2 vols.(A groundbreaking resuscitation of a broadly Aristotelian realism about universals as extra ‘objects’ in Frege’s sense.)

Bigelow, J. and Pargetter, R. (1990) *Science and Necessity*, Cambridge: Cambridge University Press.(A defence of both a broadly Platonic realism about universals as extra ‘objects’ in Frege’s sense, and of a realist construal of Fregean higher-order quantification.)

Boethius, A.M.S. (*c*.510) *In Isagogen Porphyrii Commenta*, in S. Brandt (ed.) *Corpus Scriptorum Ecclesiasticorum Latinorum*, vol. 48, Vienna: Tempsky, 1906. (Latin source for the influential text on genera and species mentioned in §4 above. Translated in Spade 1994.)

Boolos, G. (1975) ‘On Second-Order Logic’, *Journal of Philosophy* 72: 509-27.(A very impressive explanation of the virtues of higher-order logic, relevant to Frege’s fundamental distinction between ‘objects’ and ‘concepts’.)

Dooley, W.E. (1989) *Alexander of Aphrodisias: On Aristotle, Metaphysics 1*, London: Duckworth; Ithaca, NY: Cornell University Press.(Chapter 9 is especially relevant. Rich material on Plato and Aristotle, especially relevant to the third man argument outlined in §3.)

Frege, G. (1879) *Begriffsschrift*, Halle: Louis Nebert; trans. J. van Heijenoort, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967.(Difficult, but the only really epoch-making event in logic since Aristotle.)

Frege, G. (1884) *Die Grundlagen der Arithmetik*, Breslav: W. Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A logico-mathematical enquiry into the concept of number*, 2nd revised edn, Oxford: Blackwell, 1959.(An accessible, informal and exciting introduction to Frege's epoch-making theory of natural numbers.)

Gödel, K. (1944) 'Russell's Mathematical Logic', in P.A. Schilpp (ed.) *The Philosophy of Bertrand Russell*, Cambridge: Cambridge University Press.(Rich, deep and mysterious reflections on the post-Frege revolution in mathematics.)

Lewis, D. (1983) 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* 61: 343-77.(Clear exposition of a menu of options open to us in the post-Frege era.)

Quine, W.V. (1953) *From a Logical Point of View*, Cambridge, MA: Harvard University Press; 2nd edn, New York: Harper & Row, 1961.(Contains a classic, accessible and entertaining, sceptical discussion of universals in 'On What There Is', and a hard but important piece on the relation between universals and sets in 'Logic and the Reification of Universals'.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(In addition to notorious scepticism about semantics, this classic gives a landmark presentation of a sparse ontology containing only particulars and sets.)

Russell, B. (1903) *The Principles of Mathematics*, New York: Norton.(Exciting and accessible book assimilating Frege's revolution in logic, turning universals into sets and displaying some of the reasons why mathematical Platonists now believe in sets rather than in universals proper.)

Russell, B. (1912) *The Problems of Philosophy*, London: Clarendon Press.(An accessible classic, which, in addition to epistemological themes, also presents a classic, broadly Platonist vision of universals.)

Spade, P.V. (trans. and ed.) (1994) *Five Texts on the Medieval Problem of Universals*, Indianapolis, IN: Hackett. (Great works from the heyday of the problem of universals, including those mentioned at the beginning of §4 above; requires no formal logic, but both historically and conceptually difficult.)

Universals, Indian theories of

Indian philosophers postulated universals for two principal reasons: to serve as the 'eternal' meanings of words, upon which the eternality of language - in particular, the Hindu scriptures, the Veda - is based, and to account for why we conceive of things as being of certain types. However, universals were seen as problematic in various ways. How can something exist simultaneously in numerous individuals without being divided into parts? How can a universal, which is supposed to be eternal, continue to exist if all its substrata are destroyed? In what sense can a universal be said to 'exist' at all? Is a universal distinct from or identical with the individuals in which it inheres? In light of such difficulties, it is not surprising that certain other Indian philosophers - specifically Buddhist philosophers, who did not accept the doctrine of the eternality of the Veda - rejected universals and took up a nominalist stance. They held that general terms refer to mentally constructed 'exclusion classes', apohas. *The use of the term 'cow', for example, is grounded not on some positive entity common to all cows but on the idea of the class of things that are different from all things that are not cows. This proposal, which originated with Dignāga in the sixth century* AD*, was debated vigorously until the eleventh century.*

1 Universals as the meanings of words

The question of universals first arose in Indian philosophy with regard to language. In his commentary on Pāṇini's grammar of classical Sanskrit, Patañjali (second century BC) considers the thesis upon which the Hindu belief in the eternality of the Veda is grounded, that word, meaning and the connection between the two are 'eternal'. What would the meaning of a word have to be in order for it to be eternal? Presumably, not the individual thing or 'substance' (*dravya*), which is perishable, but the 'shape' or 'configuration' (*ākṛti*), which is common to many individuals. Although the *ākṛti* may be destroyed with one individual, it will continue to exist in others ((*Vyākarana*-) *Mahābhāṣya*, 1.1.1, 1962, vol. 1: 7). Here, *ākṛti* is not yet a universal in the sense of an abstract entity that makes various individuals members of a single class, but the visible form by which the members of a class are recognized. Patañjali also notes, however, that the 'substance', in the sense of the stuff out of which a thing is made, can also be considered eternal: the shape of a lump of gold may be altered in various ways, but the gold itself will remain the same. Thus, he settles on neither *ākṛti* nor *dravya* as the exclusive meaning of a word. Depending on the context, one or the other emerges as the meaning. In particular, when Pāṇini says (*Aṣṭādhyāyī* 1.2.58) that nouns may optionally take plural endings when they refer to a class - that is, one may say either 'the Brahman' or 'Brahmans' when referring to the class of Brahmans - he implies that the meaning of the word is the *ākṛti*. On the other hand, when he says (*Aṣṭādhyāyī* 1.2.64) that the plural form of a word is really an abbreviation for the same word iterated several times, he seems to have the *dravya* - even in the sense of the concrete individual - in mind (*Mahābhāṣya* 1.1.1, 1962, vol. 1: 6).

Patañjali records the opinions of other ancient grammarians (*Mahābhāṣya* 1.2.64, vol. 1: 242-). Vājapyāyana held that a word must express the *ākṛti*, because the rules of *dharma* apply universally and not just to specific things. When the law forbids the killing of a Brahman, it means that one should not kill *any* Brahman, not just a particular Brahman! Moreover, words evoke general notions. When someone says 'cow', we do not think of a particular white or black cow; we just think 'cow' in general. Vyāḍi, on the other hand, pointed out that actions can be carried out only with reference to individuals. An injunction that one should sacrifice a certain animal means that one should sacrifice an individual animal of a certain type; one could hardly carry out the action with regard to the entire species! Also, it would seem that nouns could have number and gender only if they referred to individuals. Finally, Vyāḍi wonders how the *ākṛti* can be fully present in many individuals simultaneously. Vājapyāyana responds that it is possible in the same way that it is possible for the sun to shine in many places or for Indra to be simultaneously present at many sacrifices.

This same debate is played out in the *Mīmāṃsāsūtra* (1.3.30-5), where the issue is decided in favour of the *ākṛti*, which is now explicitly declared by the commentators to be a universal (*sāmānya*). It is still recognized, however, that in certain contexts, such as sacrificial injunctions, words indirectly refer to individuals; for universals have individuals as their substrata (*Mīmāṃsāsūtrabhāṣya*, 1981, vol. 2: 266-7). The teaching of the influential Bhāṭṭa school of Mīmāṃsāwas that the universal is the primary meaning of a word, while the individual is the secondary or metaphorical meaning to which one resorts when an expression cannot be construed literally (for example, in a sacrificial context) (see MīmāṃsāČsect;Čsect;2-3). In the *Nyāyasūtra* all three items - individual (*vyakti*),

universal (*jāti*) and configuration (*ākṛti*) - are declared to be jointly the meaning of a word (2.2.66). The individual alone cannot be the meaning, because we never apprehend a mere individual, an individual which is not of a certain kind; the configuration alone cannot be the meaning, because things having the configuration of, say, a cow are not necessarily cows (for example, images of cows); and the universal by itself cannot be the meaning, because one does not apprehend the universal without perceiving an individual having a certain configuration. Thus, all three together comprise the meaning of a word, one becoming dominant, the others subordinate, in different circumstances. When one, for example, intends the distinctness of the referent from other things - as in the sentence 'Wash *that* cow' - then the *vyakti* becomes predominant; when one intends to include many things together - as in the sentence 'One should not kill a cow' - the *jāti*, or universal, becomes predominant.

2 The Vaiśeṣika theory

In the literature of the Hindu Vaiśeṣika school of natural philosophy the question of universals is approached from the standpoint of ontology. The *Vaiśeṣikasūtra* lists universals (*sāmānya*) as one of the six principal categories of things (1.1.4) (see Nyāya-Vaiśeṣika §§4-5). Their existence is known from the fact that we have common notions and terms; there must be something identical existing in different individuals that causes us to conceive of them in the same way and refer to them by the same word. Universals can be arranged in hierarchical series: animality is higher than cowness, while substantiality is higher than animality. The highest universal of all, which stands at the pinnacle of every such series and falls under no other universal, is 'existence' (*sattā*). Universals below existence can be construed either as functions of generality or particularity (*Vaiśeṣikasūtra* 1.2.3). Cowness causes us to group together all cows in one class, but also separates them from other animals. Moreover, cowness functions as a particularity with respect to universals of wider scope, such as animality: it distinguishes a certain type of animal within the genus of animals.

All cows, all animals - in general, all substances, qualities and actions - are considered 'existents' (*sat*) due to the fact that the universal 'existence' resides in them (*Vaiśeṣikasūtra* 1.2.7). Universals themselves, however, are not existents. While it can be said of a universal that it 'is' (*asti*) in so far as it possesses intrinsic nature (*svarūpa*), and while it can be known and be named, it does not exist in the same concrete sense as substances, qualities and actions: it is 'not expressible by the word *artha* ['thing']' (*Padārthadharmasaōgraha*, 1971: 49; *Vaiśeṣikasūtra* 8.2.3). In a way similar to Aristotle's account, then, 'being' does not apply univocally to all categories in Vaiśeṣika (though it should be noted that Aristotle does not consider 'universal' a category; thus, it can also be said that Vaiśeṣika differs from Aristotle in considering 'being' to apply univocally to substances, qualities and actions) (see Aristotle §7). Nor is there a universal 'universalness' that inheres in all universals; for that would lead to a 'third-man' type of regress.

Qualities are not universals in Vaiśeṣika thought. The red that occurs in a particular object is itself a particular. There is, however, a universal redness that is common to all particular occurrences of red. This contrasts with the view found in grammatical and Mīmāṃsāliterature that qualities are repeatable types: it is the *same* quality red that occurs in all things red.

It is evident that the universals of Vaiśeṣika and Nyāya - the Hindu school devoted primarily to logic and epistemology, which adopted the metaphysics of Vaiśeṣika - resemble Aristotelian universals more than Platonic forms (see Universals §1). They are neither substances nor causes of substances; they are neither essences nor responsible for the natures of things; nor are they ideal entities - a point that distinguishes them from both Platonic and Aristotelian universals. Rather, a universal simply endows an individual thing, whose nature derives from other causes, with the capacity to give rise to a certain notion such as 'cow' or 'pot', a capacity that it shares with other individuals. Moreover, although universals are quite distinct from the individuals in which they inhere - for they are eternal and 'common', while individuals are perishable and particular - they are never cognized, nor in any clear sense exist apart from them.

Other thinkers, however, conceived of universals as being more intimately related to individual substances. The grammarian Bhartṛhari (fifth century AD) held universals to be causally involved in the formation of substances; they activate the material causes that bring about the substances in which they inhere. He noted this phenomenon in the production of artefacts, much as Aristotle did: the artisan is guided by the universal (for Aristotle, read 'substantial form') in producing an object of a certain type, which will in turn exhibit the universal (*Vākyapadīya* 3.25-7). Ultimately for Bhartṛhari, who was a monist (Advaitin), both universals and individual substances are

creative potencies (*śakti*) within Brahman, different ways in which it limits itself as it becomes the empirical world. Universal and individual, that is, are merely different aspects of the finite self-expressions of Brahman, hence they are identical (*Vākyapadīya* 3.22-3, 32-3; 1.1-4). This metaphysical doctrine can be seen as a reflection of the linguistic one that words refer to both universals and individuals.

The Mīmāṃsāphilosopher Kumārila (seventh century AD) also held that the universal and the individual form a unity; for we do not cognize a universal as externally related to the individual. We do not cognize the universal 'cowness' as existing *in* the cow, rather we cognize the individual *as being* a cow (*Ślokavārttika*, 'Pratyakṣa' 141-2). The fact that individual and universal never exist apart from each other also bespeaks their identity. Furthermore, Kumārila held that one cannot make sense of the relation of 'inherence' (*samavāya*), by which Nyāya and Vaiśeṣika philosophers believed the universal and individual to be related. If one asks how *samavāya* itself is related to each of the terms it relates together, a regress ensues. Thus, according to Kumārila, who was a realist and a pluralist, the world is populated by numerous concrete entities that are both universal and particular in nature (*Ślokavārttika*, 'Ākṛti' 5-6, 59-65). A similar view was expressed by Jaina philosophers: universal and particular are neither absolutely different nor absolutely identical.

Jainas also held the distinctive doctrine of two types of universals, horizontal (*tīryaksāmānya*) and vertical (*ūrdhvatāsāmānya*). The former account for the various types of things that exist at the same time - cows, horses, and so on; the latter account for the identity of individual substances through time, that is, they enable us to recognize the various stages of a substance as the same thing.

One of the peculiarities of the Indian discussion of universals is that all those who believed in the reality of universals held them to be objects of perception; indeed, to be real and to be an object of perception were nearly the same in classical Indian philosophy. To say that universals are objects of thought would have been to say that they are mental constructs and therefore unreal. The latter view was in fact held by Buddhist philosophers (see §3). A theory of universals as real entities cognized by a process of mental abstraction was never achieved. Nyāya, Vaiśeṣika and Mīmāṃsāphilosophers argued that universals are objects of perception because their cognition occurs simultaneously with sense organ-object contact. Even when encountering a thing for the first time - say, a camel - one is able to *see* that it is an entity of a certain type, different from other kinds of entities one knows. One never perceives something that is completely particular in nature and devoid of any general character.

3 Controversies

In the face of the Buddhist denial of the reality of universals, Hindu and Jaina philosophers were compelled to develop more rigorous arguments for their existence. The most common argument was simply that distinct cognitions must have distinct causes. We have, besides the cognition of something as an individual, a cognition of it as a member of a certain class. There must be some real basis for the latter, in the same way that, when we cognize something as blue, there is some real basis of the cognition other than the individual substance itself, namely the quality blue (*Nyāyavārttika* 2.2.64, 1985: 667). Nor can the cognition of an individual as possessing a general character be based on a similarity between it and other individuals. Cows are in certain ways similar to each other, but in important respects also different; and almost to the same extent that there is a similarity between cows, there is a similarity between cows and horses, yet we do not call a horse a cow! Moreover, similarity would merely give rise to the notion that various things are 'like' each other, not that they are the same thing. If universals were not real, it was also argued, inference and reasoning, which are based on the observation of general traits, would not be valid means of knowledge. The Nyāya philosopher Udayana (eleventh century) contended that universals must be postulated to explain the necessity and uniformity of causal relations, such as the fact that not just this bit of milk will now produce yoghurt, but any portion at any time.

One problem that universal theorists had to face was whether a universal exists wholly in an individual or only partially. If wholly, then it would only be in that individual and could not also exist in others. If only a part of the universal existed in each individual, then it would not be one thing common to all (compare Plato's *Parmenides* 131a-e). The standard response to this dilemma was simply to stress that the property of existing completely in each of a class of individuals is one of the things that distinguishes universals from other categories. As the Nyāya philosopher Uddyotakara (sixth century) puts it: a universal is not a whole comprised of parts, nor a collection of many things, thus it is not really proper to put the question about whether it exists 'wholly' or 'partially' in something (*Nyāyavārttika* 2.2.64, 1985: 669).

What happens to a universal if all its individual substrata perish? It cannot cease to exist, for it is eternal; yet it is never perceived independently of its substrata either. Nyāya-Vaiśeṣika philosophers held that during periods of cosmic dissolution (*pralaya*), when there are no individuals at all, universals continue to exist in an unmanifest state - in so far as they can be said to 'exist' at all. Otherwise, some (for example, Jayantabhaṭṭa) held that when all the individuals of a particular universal are destroyed, universals continue to inhere in other individuals. That is to say, cowness (along with every other universal) is present in all things, even non-cows, and so will still be located in individuals even when cows have died out; but it is *perceived* only in cows (*Nyāyamañjarī*, 1983, vol. 2: 36-7). In such doctrines one sees a tendency to favour the notion that universals exist only *in rebus*. There were even philosophers who held that during *pralaya* universals continue to exist by virtue of inhering in individuals in other universes! Perhaps only Bhartṛhari would have been sympathetic to the Western medieval view that universals are *ante res*, in so far as they are eternal potencies within Brahman.

Finally, there is the matter of *upādhis*, general traits that are not universals. Nyāya-Vaiśeṣika philosophers wanted to deny that such terms as 'cookness' or 'servantness' refer to real, objective universals. Being a cook or a servant is not natural to a person; one becomes a cook or servant only when one comes to perform the actions appropriate to a cook or servant. Universals, however, are supposed to inhere permanently in their substrata from the moment of origin. Similarly, several men carrying umbrellas can be grouped together conceptually, but that will not be due to the presence of a universal in them. Thus, a distinction was made between general terms that are based on real universals and general terms that are based on inauthentic, nominal properties. Buddhists used this point to their own advantage: if some general terms are based on inauthentic, nominal properties, why not all?

The Buddhist critique of universals was initiated by Dignāga (sixth century) and continued by Dharmakīrti (seventh century), Śāntarakṣita (eighth century) and other thinkers until the eleventh century. An often-quoted verse of Dharmakīrti charges that the concept of universals is beset with a plethora of difficulties: when an individual arises in a particular place, the universal cannot move to where it is and unite with it, for a universal is without motion; nor could it have been present there all along, for then the empty space would have been cognized as falling under the universal. It is also absurd to think that a universal could join with an individual after it has arisen, since the individual would exist for a moment without any universal inhering in it (*Pramāṇavārttika*, 'Svārthānumāna' 152-3). If various things were identical in nature through the presence of a universal, they would not really be diverse; if a concrete thing were both universal and particular in nature, how would it be unified (*Pramāṇavārttika*, 'Pratyakṣa' 41; 'Svārthānumāna' 176-9)? Less polemically, Dharmakīrti points out that the fact that various things give rise to the same idea does not require the postulation of something that they all have in common; for we observe that a variety of herbs that have nothing in common have the same effect of alleviating fever (*Pramāṇavārttika*, 'Svārthānumāna' 73-4).

In reality, however, various individuals never give rise to the same *positive* idea, but only to a conception that is completely negative in content. That is to say, a particular cow may be seen as being different from all goats, all horses, all pots, all vegetables, and so on; another particular cow may be seen as being different from the same things. In so far as both cows are conceived of as being different from the same things, they are thought of as being of the same type; in themselves, however, they are quite distinct. This is the notorious *apohavāda*, 'the doctrine of exclusion', which asserts that general terms refer only to mentally constructed exclusion classes, *apohas*. The main thrust of the theory is that there is nothing at all positive to being an entity of a certain kind. Only individuals, which are inconceivable in nature, yet grasped immediately in perception (which is by definition devoid of all conceptualization), are real. The *apoha* doctrine was used by the Buddhists to question the reality of a wide range of phenomena that appear to be conceptually based, including substances, words and meanings (see Nominalism, Buddhist doctrine of).

The *apohavāda* evoked a spirited response on the part of realist thinkers, with numerous texts being devoted to the debate. The most serious objection to it, originally developed by Kumārila, is that it is unclear whether one can even recognize that a particular thing is *different* from other things without seeing positively that it and other things are of certain types. And if in attempting to identify a certain thing one's recognition of the 'other things' from which it differs were based on *apoha*, then one would have to be able to see that they exclude, among other things, the thing one is trying to identify, in which case one would already have to be able to identify something as being of a certain type in order to see that it is of a certain type (*Ślokavārttika*, 'Apohavāda' 2-, 65-6). Moreover, the *apohavāda* renders all words nearly synonymous: the only difference in meaning between the words 'cow' and

'horse' is that the former excludes horses while the latter excludes cows; otherwise, the vast number of the things each word excludes is identical (*ibid*. 53-4). In general, the realist arguments against the *apohavāda* raise an issue with which even modern nominalists must reckon, namely, while it may be possible to construct theories that make no reference to abstract entities, can one really understand the locutions of such theories without tacitly appealing to them?

The heated debate over universals was no mere intellectual exercise for Indian philosophers. The defenders of the reality of universals were concerned to preserve, at least up to a point, the validity of the distinctions of natural kinds upon which the prescriptions of scripture are based. The opponents of universals, on the other hand, sought ultimately to deny the reality of the world of ordinary experience and encourage the cultivation of a mystical awareness of an ineffable Absolute.

See also: Language, Indian theories of; Meaning, Indian theories of; Ontology in Indian philosophy

JOHN A. TABER

References and further reading

Bhartṛhari (5th century) *Vākyapadīya*, trans. K.A. Subramania, Iyer, Poona: Deccan College, 1965-74, 4 vols. (The main treatise of the grammarian school of Hindu thought.)

Dharmakīrti (7th century) *Pramāṇavārttika*, ed. Swami Dwarikadas Shastri, Varanasi: Bauddha Bharati, 1984. (An important text of the logical-epistemological school of Yogācāra Buddhism.)

Dravid, R.R. (1972) *The Problem of Universals in Indian Philosophy*, Delhi: Motilal Banarsidass.(A comprehensive treatment of the views of all schools.)

Frauwallner, E. (1982) 'Beiträge zur Apohalehre' (Contributions to the Theory of Apoha), in *Kleine Schriften (Smaller Writings)*, Wiesbaden: Franz Steiner Verlag. (Seminal articles on Buddhist nominalism.)

Halbfass, W. (1992) *On Being and What There Is: Classical Vaiśeṣika and the History of Indian Ontology*, Albany, NY: State University of New York Press.(A penetrating study of Vaiśeṣika metaphysics.)

Jayantabhaṭṭa (9th century) *Nyāyamañjarī*, ed. K.S. Varadacharya, Mysore: Oriental Research Institute, 1983.(A major Nyāya text of the late ninth century.)

Kumārila (7th century) *Ślokavārttika*, ed. Swami Dwarikadas Shastri, Varanasi: Tara Publications, 1978; trans. G. Jha, Delhi: Sri Satguru Publications, 1983.(One of two principal subcommentaries on Śabarasvāmin's *Mīmāṃsāsūtrabhāṣya*.)

Matilal, B.K. (1986) *Perception*, Oxford: Clarendon Press.(Treats a wide range of issues in Indian metaphysics and epistemology; chapter 12 is devoted to universals.)

Patañjali (2nd century BC) *Vyākaraṇamahābhāṣya*, ed. F. Kielhorn, Poona: Bhandarkar Oriental Research Institute, 1962.(The principal commentary on Pāṇini's grammar.)

Praśastapāda (5th century) *Praśastapādabhāṣya* (*Padārthadharmasaōgraha*), ed. Bhāgīratha Prasāda Tripāthī, Varanasi: Sampurnand Sanskrit Vishvavidyalaya, 1971; trans. G. Jha, Benares: E.J. Lazarus, 1916.(The most important commentary on Pāṇini's grammar.)

Śabarasvāmin *Mīmāṃsāsūtrabhāṣya*, in Gaṇeṣaśāstri Jośī (ed.) *Mīmāṃsādarśanam*, Varanasi: Anandaśrama Sanskrit Series, 1981; trans. G. Jha, *Śābara-Bhāṣya*, The Gaekwad's Oriental Series, 66, 70, 73, Baroda: Oriental Institute, 1933-6.(The oldest extant commentary on the *Mīmāṃsāsūtra*; includes the text of the latter.)

Uddyotakara (6th century) *Nyāyavārttika*, in Taranatha Nyaya-Tarkatirtha (ed.) *Nyāyadarśanam*, Delhi: Munshiram Manoharlal, 1985; trans. G. Jha, *The Nyāya-Sūtras of Gautama with the Bhāśya of Vātsyāyana and the Vārttika of Uddyotakara*, Delhi: Motilal Banarsidass, 1984, 4 vols.(An important subcommentary on Vātsyāyana's commentary on the *Nyāyasūtra*; both Sanskrit edition and translation include the text of the *Nyāyasūtra*.)

Vaiśeṣikasūtra (1st century AD?), trans. Nandalal Sinha, *The Vaiśeṣika Sūtras of Kaṇāda with the Commentary of Śaōkara Miśra*, Allahabad: Sacred Books of the Hindus, 1910-11.(The foundational text of the Vaiśeṣika school.)

Use/mention distinction and quotation

Speakers 'use' the expressions they utter and 'mention' the individuals they talk about. Connected with the roles of used expressions and mentioned individuals is a way of uniting them and a characteristic mistake involving them. Usually the expression used in an utterance will not be the same as the individual mentioned, but the two can be made to converge. The means is quotation. Quotation is a special usage in which an expression is used to mention itself. A failure to distinguish between the roles of used expressions and mentioned individuals can lead to mistakes. Such mistakes are called use/mention confusions. In themselves use/mention confusions are a minor linguistic faux pas, *but under unfavourable conditions, they have the potential to cause greater problems.*

1 Use and mention

Though philosophers since Plato have distinguished between language and the world it is used to describe, the terms 'use' and 'mention' were first applied to these categories by the philosopher and logician Willard Quine in 1940. Quine's choice of the terms was intended to encourage associations with their ordinary meanings. To *use* an expression linguistically is to use it as a tool; to employ it towards a linguistic end. The ends are performances of speech acts, such as making claims, asking questions and giving responses. In contrast, to *mention* an individual is to say something about it; to make it the topic of conversation between speaker and listener. An individual mentioned in a speech act is one whose characteristics help determine how the act is to be evaluated: whether a claim is true, a question pertinent or a response correct depends on the individuals mentioned in the act and what is said about them.

In every communicative setting, some expression is used and, in most, some individual is mentioned. For example, a speaker uttering (1) below succeeds in making a claim about Tucson, the capital city of the state of Arizona:

(1) Tucson is dry.

By using the words 'Tucson', 'is' and 'dry' in the appropriate order, the city is mentioned. To say that Tucson is mentioned is to say that it is Tucson's weather, rather than that of Rangoon or Vienna, which is relevant to the truth of the claim. Greetings, such as 'Good Morning' or 'Hello', are examples of speech acts in which no individual is mentioned; their success depends only on the circumstances and the speaker's intentions.

While thinking about the categories of expressions used and individuals mentioned, one should bear in mind that the distinction divides the roles individuals can play rather than the kinds of individuals that can play these roles. That the distinction does not separate individuals by kinds is shown by the fact that linguistic expressions can play both parts: expressions can be mentioned as well as used. By uttering (2), for example, a speaker manages to mention the name 'Tucson' and, in fact, to say something true about it:

(2) The name of the capital city of Arizona contains six letters.

(1) should be compared with (2). The former uses the name of the city to mention the city, while the latter uses a description of the name to mention the name itself.

A detailed understanding of use and mention requires that one draw a distinction between the 'type' of an expression and 'tokens' of that type (see Type/token distinction). Expression tokens are concrete objects, individual marks or sounds made of raw materials: ink, perhaps; or sound waves. Types are abstract individuals and, as such, not made of anything. There is only one word 'Tucson' in the English language (so 'Tucson' is a type), of which there are several tokens in this entry.

The type/token distinction brings with it possibilities for what is used and mentioned. Both tokens and types can be used. To use a token, one has to utter it with the intention of saying something. Using a type requires the same intention, but also requires that the token have the right shape or sound. Tokens and types can also be mentioned. (2) mentions the type and (3) the token used in (1):

(3) The first word token in the first example sentence of the entry on 'Use/mention distinction and quotation' in the *Routledge Encyclopedia of Philosophy* has six letters.

2 Standard uses of quotation and quasi-quotation

Expressions of each grammatical category have what may be called standard uses. The standard use of noun phrases is to mention: 'Tucson', 'The name of the capital city of Arizona' and 'The first word token…*Philosophy*' from (3) above are all used in standard fashion. Expressions of most other categories have standard uses which do not involve mentioning. Articles and quantifiers ('every', 'some', 'all', 'most') do not themselves mention, but rather help determine the range of individuals that are mentioned. Conjunctions govern the manner in which the meanings of compound expressions depend on the meanings of their constituents. According to some theories, adjectives, adverbs and verbs mention relations, properties or events. According to other theories, they do not mention at all.

Along with their standard uses, expressions also have nonstandard uses. Chief among these is their use in quotation. Quotation is a way of using an expression to mention itself. It is the most general and efficient way of talking about expressions. Instead of the long-winded description in (2), a speaker wanting to mention 'Tucson' can simply quote it:

(4) 'Tucson' has six letters.

From the notational point of view, quotation is trivial. Its basic rule is simple: to quote an expression, enclose a token of it in quotation marks - the purpose of the marks being to indicate the nature of the usage. Or it can be simpler yet; one can leave out the marks when it is clear from the context that one wants to talk about an expression rather than use it in the standard way. (Passages which are indented or examples marked by numbers - as in this entry - are quoted using this method.) Quotation is generally used to mention expression types, unless there is an explicit qualification to the contrary.

Quotation is a leveller. No matter what they may standardly be used to mention, when quoted, all nouns mention expressions. It is instructive to compare (1) and (4) on this count to see how quotation can bring about a change in subject matter: (1) is about a large city, while (4), which differs only in the addition of a set of quotation marks, is about a medium-sized word. The effect of quotation on other categories of expressions is even more significant. It reduces the whole variety of standard uses to one. No matter what their standard use may be, when quoted, expressions of all categories become mentioning expressions. This effect can be seen in the example below:

(5) 'A' is an article and 'and' is a conjunction.

In this sentence, the first token of the word 'a' and the second token of the word 'and' are used to mention, while the second token of the word 'a', the first of 'and' and the only one of 'an' operate in standard fashion.

Related to regular quotation is a notation invented by Quine (1940), called 'quasi-quotation'. Quine, who has done more to encourage the correct use of quotation marks than any other writer, created quasi-quotation to help in capturing linguistic generalizations. Quasi-quotation treats an expression in two ways. Towards one part of the expression ('Čamp;' in the example below) it behaves as regular quotation. The other part or parts (' α' and ' β' in the example below) it treats as indeterminate elements that can mention any expression within a specific category. The marks of quasi-quotation are called corner quotes (' $\ulcorner$ ' on the left and ' $\urcorner$ ' on the right). A full quasi-quotation consisting of the expression enclosed by two corners is itself an indeterminate that can mention any expression consisting of the quoted part and any expression mentioned by the indeterminate part. To take an example from logic, let ' α' and ' β' be indeterminates mentioning sentences and 'Čamp;' be the conjunction symbol. A rule giving the extensional semantics for conjunction can be stated using quasi-quotation:

(6) $\ulcorner\alpha \,\&\, \beta\urcorner$ is true iff α is true and β is true,

where ' $\ulcorner\alpha \,\&\, \beta\urcorner$ ' mentions an expression that is a conjunction of the sentences mentioned by ' α' and ' β'. If α is 'Cicero died' and β is 'Caesar spoke', then $\ulcorner\alpha \,\&\, \beta\urcorner$ is the result of writing α followed by 'Čamp;' followed by β, namely 'Cicero died and Caesar spoke'.

3 Use/mention confusions

The distinction between use and mention can turn slippery. When it does and one loses hold of it, use/mention

confusions can occur. The typical result of such a confusion is that a speaker ends up making a claim about nonlinguistic individuals that is appropriate only to expressions or making a claim about expressions that is only appropriate to nonlinguistic individuals. Where the distinction between use and mention turns most slippery is on the subject of abstract individuals. Numbers and expressions used to mention them can be especially hard to keep apart. The category of numbers includes whole numbers such as 1 and 3, ratios such as $\frac{1}{2}$ and $\frac{4}{3}$ and irrational numbers such as π. The expressions that mention them include the numerals '1' and '3', the fractions '$\frac{1}{2}$' and '$\frac{4}{3}$' and the Greek letter 'π'. Classroom presentations of ratios and fractions commonly involve use/mention confusions. Students are often told something like the following:

Fractions consist of one number divided by another. The number to be divided, called the 'numerator', is placed on the top, while the number to divide by, called the 'denominator', is placed on the bottom, with a line between the two. So, 2 is the numerator and 3 the denominator of $\frac{2}{3}$.

However, the clever student may wonder how this could be. Since $\frac{2}{3}$ is the same quotient as $\frac{4}{6}$, the above remarks suggest that they ought to be the same fraction. Therefore 3 ought to be the denominator of both (on the grounds that the denominators of identicals are identical), but it is not. This puzzle exposes the use/mention confusion. The teacher has confused fractions with ratios. '$\frac{2}{3}$' is a fraction; $\frac{2}{3}$ a ratio. Only fractions have numerators and denominators, while it is the ratios mentioned by fractions that may be identical. The numeral '3' is the denominator of the fraction '$\frac{2}{3}$', but it is the *ratios* $\frac{2}{3}$ and $\frac{4}{6}$ which are identical. Because '$\frac{2}{3}$' is not the same fraction as '$\frac{4}{6}$', they need not, and indeed do not, have the same denominator.

The need to achieve clarity and avoid confusions such as the one above is the main reason why it is important to maintain a clear separation between use and mention. As far as errors go, the mix-up with fractions and ratios is probably mid-level. Some use/mention confusions are of less consequence. Others pose greater problems (see Modal operators §1). Where subtlety and precision are at a premium, as they are in logic and linguistics, use/mention confusions can severely impair the intelligibility of a discussion.

See also: De re/de dicto; Logical and mathematical terms, glossary of

COREY WASHINGTON

References and further reading

Davidson, D. (1984) 'Quotation', in *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1991, 79-92.(Davidson puts forward the view that quotation marks are demonstratives, an alternative to the view advanced in §2.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. 'On Sense and Reference', in A.P. Martinich (ed.) *The Philosophy of Language*, New York: Oxford University Press, 3rd edn, 1996.(Frege's discussion in this essay contains what may be the first application of the use/mention distinction in the context of a philosophical argument. Frege also introduces the view, stated in §2, that quotation marks are context markers and that a quoted word is used to mention itself.)

Garcia-Carpentera, M. (1994) 'Ostensive Signs: Against the Identity Theory of Quotation', *Journal of Philosophy* 91: 253-64.(Endorses Davidson's theory and defends it against criticisms.)

Geach, P. (1957) *Mental Acts*, London: Routledge & Kegan Paul, 79-82.(Clearest statement of the view, first formulated by Tarski, that quotations can be seen as descriptions.)

Quine, W.V. (1940) *Mathematical Logic*, New York: Norton; repr. Cambridge, MA: Harvard University Press, 1951, 23-37.(A careful discussion of the use/mention distinction, quotation and quasi-quotation.)

Russell, B.A.W. (1919) 'Descriptions', in *Introduction to Mathematical Philosophy*, London: Allen & Unwin, 167-80; repr. in A.P. Martinich (ed.) *The Philosophy of Language*, New York: Oxford University Press, 1985, 213-19; 3rd edn, 1996.(Contains numerous examples of use/mention confusions.)

Tarski, A. (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956), 'On the Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(The classic work on the view that quotations are names.)

Washington, C. (1992) 'The Identity Theory of Quotation', *Journal of Philosophy* 91: 582-605.(Gives a detailed

description and defence of Frege's view.)

Utilitarianism

Utilitarianism is a theory about rightness, according to which the only good thing is welfare (wellbeing or 'utility'). Welfare should, in some way, be maximized, and agents are to be neutral between their own welfare, and that of other people and of other sentient beings.

The roots of utilitarianism lie in ancient thought. Traditionally, welfare has been seen as the greatest balance of pleasure over pain, a view discussed in Plato. The notion of impartiality also has its roots in Plato, as well as in Stoicism and Christianity. In the modern period, utilitarianism grew out of the Enlightenment, its two major proponents being Jeremy Bentham and John Stuart Mill.

Hedonists, believing that pleasure is the good, have long been criticized for sensualism, a charge Mill attempted to answer with a distinction between higher and lower pleasures. He contended that welfare consists in the experiencing of pleasurable mental states, suggesting, in contrast to Bentham, that the quality, not simply the amount, of a pleasure is what matters. Others have doubted this conception, and developed desire accounts, according to which welfare lies in the satisfaction of desire. Ideal theorists suggest that certain things are just good or bad for people, independently of pleasure and desire.

Utilitarianism has usually focused on actions. The most common form is act-utilitarianism, according to which what makes an action right is its maximizing total or average utility. Some, however, have argued that constantly attempting to put utilitarianism into practice could be self-defeating, in that utility would not be maximized by so doing. Many utilitarians have therefore advocated non-utilitarian decision procedures, often based on common sense morality. Some have felt the appeal of common sense moral principles in themselves, and sought to reconcile utilitarianism with them. According to rule-utilitarianism, the right action is that which is consistent with those rules which would maximize utility if all accepted them.

There have been many arguments for utilitarianism, the most common being an appeal to reflective belief or 'intuition'. One of the most interesting is Henry Sidgwick's argument, which is ultimately intuitionist, and results from sustained reflection on common sense morality. The most famous argument is Mill's 'proof'. In recent times, R.M. Hare has offered a logical argument for utilitarianism.

The main problems for utilitarianism emerge out of its conflict with common sense morality, in particular justice, and its impartial conception of practical reasoning.

1 Introduction and history

Defining utilitarianism is difficult, partly because of its many variations and complexities, but also because the utilitarian tradition has always seen itself as a broad church. But before offering a history, we must supply a working definition. First, utilitarianism is, usually, a version of *welfarism*, the view that the only good is welfare (see Welfare). Second, it assumes that we can compare welfare across different people's lives (see Economics and ethics). Third, it is a version of consequentialism (see Consequentialism). Consequentialists advocate the impartial maximization of certain values, which might include, say, equality. Utilitarianism is welfarist consequentialism, in its classical form, for instance, requiring that any action produce the greatest happiness (see Happiness).

The concern with welfare, its measurement and its maximization is found early, in Plato's *Protagoras*. In the process of attempting to prove that all virtues are one, Socrates advocates *hedonism*, the welfarist view that only pleasurable states of mind are valuable, and that they are valuable solely because of their pleasurableness (see Plato §9; Socrates §24; Hedonism).

The debate in the *Protagoras* is just one example of the many discussions of welfare in ancient ethics (see Eudaimonia). Some have seen Greek ethics as primarily egoistic, addressing the question of what each individual should do to further their own welfare (see Egoism and altruism §4). Utilitarianism, however, is impartial.

The Stoics, who followed Plato and Aristotle, began to develop a notion of impartiality according to which self-concern extended rationally to others, and eventually to the whole world (see Stoicism §18). This doctrine, allied to Christian conceptions of self-sacrifice, and conceptions of rationality with roots in Plato which emphasize the objective supra-individual point of view, could plausibly be said to be the source of utilitarian impartiality (see

Impartiality).

In the modern period, the history of utilitarianism takes up again during the Enlightenment. The idea of impartial maximization is found in the work of the eighteenth-century Scottish philosopher Francis Hutcheson (1755) (see Hutcheson, F. §2). The work of his contemporary, David Hume (1751), also stressed the importance to ethics of the notion of 'utility' (see Hume, D. §4.1). A little later, the so-called 'theological utilitarians', Joseph Priestley (1768) and William Paley (1785), argued that God requires us to promote the greatest happiness (see Priestley, J.; Paley, W.). Meanwhile, in France, Claude Helvétius (1758) advocated utilitarianism as a political theory, according to which the task of governments is to produce happiness for the people. He influenced one of the most extreme of all utilitarians, William Godwin (1793) (see Helvétius, C.; Godwin, W.).

It was Jeremy Bentham, however, who did most to systematize utilitarianism. Bentham's disciple, J.S. Mill, was the next great utilitarian, and he was followed by Henry Sidgwick. G.E. Moore (1903) distanced himself from Mill's hedonism, and offered an influential 'ideal' account of the good. One of the most important recent versions of utilitarianism is that of R.M. Hare (see Moore, G.E.; Hare, R.M.).

2 Conceptions of utility

Before you can maximize utility, you need to know what utility is. It is essential to note that the plausibility of utilitarianism as a theory of right action does not depend on any particular conception of welfare. An account of the good for a person is different from an account of right action (see Right and good).

Utilitarians have held many different views of utility. The 'classical' utilitarians - primarily Bentham (1789) and Mill (1861) - were hedonists. There are many objections to hedonism. What about masochists, for example, who seem to find pain desirable? Well, perhaps pain can be pleasurable. But is there really something common - pleasure - to all the experiences that go to make up a happy life? And would it be rational to plug oneself into a machine that gave one vast numbers of pleasurable sensations? Here there may be a move towards the more eclectic view of Sidgwick (1874), that utility consists in desirable consciousness of any kind. Some philosophers, however, such as Nietzsche (1888), have suggested that a life of mere enjoyment is inauthentic.

Hedonists have been criticized for sensualism for millennia J.S. Mill sought to answer the charge, suggesting that hedonists do not have to accept that all pleasurable experiences - drinking lemonade and reading Wordsworth - are on a par, to be valued only according to the *amount* of pleasure they contain. Bentham and others had suggested that the value of a pleasure depends mainly on its intensity and its duration, but Mill insisted that the *quality* of a pleasure - its nature - also influences its pleasurableness and hence its value. But why must the effect on value of the nature of an experience be filtered through pleasurableness? Why cannot its nature by itself add value?

Perhaps the most serious objection to any theory that welfare consists in mental states is the so-called 'experience machine'. This machine is better than the pleasure machine, and can give you the most desirable experiences you can imagine. Would it be best for you to be wired up to it throughout your life? Note that this is not the question whether it would be *right* to arrange for yourself to be wired up, leaving all your obligations in the real world unfulfilled. Even a utilitarian can argue that that would be immoral.

Some people think it makes sense to plug in, others that it would be a kind of death. If you are one of the latter, then you might consider moving to a desire theory of utility, according to which what makes life good for you is your desires' being maximally fulfilled. On the experience machine, many of your desires will remain unfulfilled. You want not just the experience of, say, bringing about world peace, but actually to bring it about. Desire theories have come to dominate contemporary thought because of economists' liking for the notion of 'revealed preferences' (see Rationality, practical). Pleasures and pains are hard to get at or measure, whereas people's preferences can be stated, and inferred objectively from their behaviour.

A simple desire theory fails immediately. I desire the glass of liquid, thinking it to be whisky. In fact it is poison, so satisfying my desire will not make me better off. What desire theorists should say here is that it is the satisfaction of intrinsic desires which counts for wellbeing. My intrinsic desire is for pleasure, the desire for the drink being merely derived.

The usual strategy adopted by desire theorists is to build constraints into the theory in response to such counterexamples: what makes me better off is not the fulfilment of my desires, but of my informed desires.

But why do desire theorists so respond to such counterexamples? It is probably because they already have a view of utility which guides them in the construction of their theories. This means that desire theories are themselves idle, which is to be expected once we realize that the fulfilment of a desire is in itself neither good nor bad for a person. What matters is whether what the person desires, and gets, is good or bad.

For reasons such as this, there is now a return to ancient *ideal* theories of utility, according to which certain things are good or bad for beings, independently in at least some cases of whether they are desired or whether they give rise to pleasurable experiences (see Perfectionism). Another interesting ancient view which has recently been revived is that certain nonhedonistic goods are valuable, but only when they are combined with pleasure or desire-fulfilment (see Plato, *Philebus* 21a-22b). The nonhedonistic goods suggested include knowledge and friendship. Questions to ask of the ideal theorist include the following. What will go on your list of goods? How do you decide? How are the various items to be balanced?

3 Types of utilitarianism

Theories of right and wrong have to be about something, that is, have to have a focus. Usually, at least in recent centuries, they have focused on actions, attempting to answer the questions, ‘Which actions are right?’, and, ‘What makes those actions right?’. The ancients also asked these questions, but were concerned also to focus on lives, characters, dispositions and virtues. Nearly all forms of utilitarianism have focused on actions, but in recent decades there has been some interest in utilitarianism as applied to motives, virtues and lives as a whole.

Utilitarianism is a form of consequentialism. But it is important to note that, since utilitarians can attach instrinsic moral importance to acts (especially, of course, the act of maximizing itself), there are problems in attempting to capture the nature of utilitarianism using the act/consequence distinction. A recent alternative has been to employ the ‘agent-neutral’/‘agent-relative’ distinction. Agent-neutral theories give every agent the same aim (for example, that utility be maximized), whereas agent-relative theories give agents different aims (say, that *your* children be looked after). Logically, however, there is nothing to prevent a utilitarian’s insisting that *your* aim should be that *you* maximize utility. Though this theory would be practically equivalent to an agent-neutral theory, its possibility suggests there may be problems with attempting to use the agent-neutral/agent-relative distinction to capture the essence of utilitarianism.

What clearly distinguishes utilitarianism from other moral theories is what it requires and why, so we should now turn to that. The commonest, and most straightforward, version of utilitarianism is *act-utilitarianism*, according to which the criterion of an action’s rightness is that it maximize utility.

Act-utilitarians might offer two accounts of rightness. The objectively right action would be that which actually does maximize utility, while the subjectively right action would be that which maximizes expected utility. Agents would usually be blamed for not doing what was subjectively right.

Another distinction is between *total* and *average* forms. According to the total view, the right act is the one that produces the largest overall total of utility. The average view says that the right action is that which maximizes the average level of utility in a population. The theories are inconsistent only in cases in which the size of a population is under consideration. The most common such case occurs when one is thinking of having a child. Here, the average view has the absurd conclusion that I should not have a child, even if its life will be wonderful and there will be no detrimental effects from its existence, if its welfare will be lower than the existing average.

But the total view also runs into problems, most famously with Derek Parfit’s ‘repugnant conclusion’ (1984), which commits the total view to the notion that if a population of people with lives barely worth living is large enough it is preferable to a smaller population with very good lives. One way out of this problem is to adopt a *person-affecting* version of utilitarianism, which restricts itself in scope to existing people. But there are problems with this view (see Parfit 1984: ch. 18). Recently, certain writers have suggested that one way to avoid the ‘repugnant conclusion’ would be to argue that there are *discontinuities* in value, such that once welfare drops below a certain level the loss cannot be compensated for by quantity. There is a link here with Mill’s view of the relation of higher pleasures to lower.

Imagine being an act-utilitarian, brought up in an entirely act-utilitarian society. You will have to spend much time calculating the utility values of the various actions open to you. You are quite likely to make mistakes, and, being

human, to cook the books in your own favour.

For these reasons, most act-utilitarians have argued that we should not attempt to put act-utilitarianism into practice wholesale, but stick by a lot of common sense morality (see Common-sense ethics). It will save a lot of valuable time, is based on long experience, and will keep us on the straight and narrow. Act-utilitarians who recommend sole and constant application of their theory as well as those who recommend that we never consult the theory and use common sense morality can both be called *single-level* theorists, since moral thinking will be carried on only at one level. Most utilitarians have adopted a *two-level* theory, according to which we consult utilitarianism only sometimes - in particular when the principles of ordinary morality conflict with one another.

The main problem with two-level views is their psychology. If I really accept utilitarianism, how can I abide by a common sense morality I know to be a fiction? And if I really do take that common sense morality seriously, how can I just forget it when I am supposed to think as a utilitarian? The two-level response here must be that this is indeed a messy compromise, but one made to deal with a messy reality.

Act-utilitarianism is an extremely demanding theory, since it requires you to be entirely impartial between your own interests, the interests of those you love, and the interests of all. The usual example offered is famine relief. By giving up all your time, money and energy to famine relief, you will save many lives and prevent much suffering. Utilitarians often claim at this point that there are limits to human capabilities, and utilitarianism requires us only to do what we can. But the sense of ‘can’ here is quite obscure, since in any ordinary sense I can give up my job and spend my life campaigning for Oxfam.

The demandingness objection seems particularly serious when taken in the context of widespread non-compliance with the demands of act-utilitarian morality. Most people do little or nothing for the developing world, and this is why the moral demands on me are so great. An argument such as this has been used to advocate *rule-utilitarianism*, according to which the right action is that which is in accord with that set of rules which, if generally or universally accepted, would maximize utility. (The version of the theory which speaks of the rules that are *obeyed* is likely to collapse into act-utilitarianism; see Lyons 1965.)

Unlike act-utilitarianism, which is a *direct* theory in that the rightness and wrongness of acts depends directly on whether they fit with the maximizing principle, rule-utilitarianism is an *indirect* theory, since rightness and wrongness depend on rules, the justification for which itself rests on the utilitarian principle.

The demandingness of act-utilitarianism has not been the main reason for adopting rule-utilitarianism. Rather, the latter theory has been thought to provide support for common sense moral principles, such as those speaking against killing or lying, which appear plausible in their own right.

Rule-utilitarianism has not received as much attention as act-utilitarianism, partly because it detaches itself from the attractiveness of maximization. According to rule-utilitarianism there may be times when the right action is to bring about less than the best possible world (such as when others are not complying). But if maximization is reasonable at the level of rules, why does it not apply straightforwardly to acts?

4 Arguments for utilitarianism

The most famous argument for utilitarianism is John Stuart Mill’s ‘proof’ (1861). This has three stages:

1 Happiness is desirable.
2 The general happiness is desirable.
3 Nothing other than happiness is desirable.

Each stage has been subjected to much criticism, especially the first. Mill was an empiricist, who believed that matters of fact could be decided by appeal to the senses (see Empiricism). In his proof, he attempted to ground evaluative claims on an analogous appeal to desires, making unfortunate rhetorical use of ‘visible’ and ‘desirable’. The first stage suggests to the reader that if they consult their own desires, they will see that they find happiness desirable.

The second stage is little more than assertion, since Mill did not see the vastness of the difference between egoistic and universalistic hedonism (utilitarianism). In an important footnote (1861: ch. 5, para. 36), we see the

assumption that lies behind the proof: the more happiness one can promote by a certain action, the stronger the reason to perform it. Egoists will deny this, but it does put the ball back in their court.

The final stage again rests on introspection, the claim being that we desire, ultimately, only pleasurable states. Thus even a desire for virtue can be seen as a desire for happiness, since what we desire is the pleasure of acting virtuously or contemplating our virtue. One suspects that introspection by Mill's opponents would have had different results.

Perhaps the most common form of utilitarianism, as of any other moral theory, is, in a weak sense, intuitionist (see Intuitionism in ethics). To many, utilitarianism has just seemed, taken by itself, reasonable - so reasonable, indeed, that any attempt to prove it would probably rest on premises less secure than the conclusion. This view was expressed most powerfully by Henry Sidgwick (1874). Sidgwick supported his argument with a painstaking analysis of common sense morality. Sidgwick also believed that egoism was supported by intuition, so that practical reason was ultimately divided (see Egoism and altruism §§1, 3).

In the twentieth century, R.M. Hare wished to avoid appeal to moral intuition, which he saw as irrational. According to Hare (1981), if we are going to answer a moral question such as, 'What ought I to do?', we should first understand the logic of the words we are using. In the case of 'ought', we shall find that it has two properties: prescriptivity (it is action-guiding) and universalizability (I should be ready to assent to any moral judgment I make when it is applied to situations similar to the present one in their universal properties) (see Prescriptivism). Hare argues that putting yourself in another's position properly - 'universalizing' - involves taking on board their preferences. Once this has been done, the only rational strategy is to maximize overall preference-satisfaction, which is equivalent to utilitarianism.

Hare's moral theory is one of the most sophisticated since Kant's, and he does indeed claim to incorporate elements of Kantianism into his theory (see Kantian ethics). Objectors have claimed, however, that, rather like Kant himself, Hare introduces 'intuitions' (that is, beliefs about morality or rationality) through the back door. For example, the logic of the word 'ought' may be said not to involve a commitment to the rationality of maximization even in one's own case.

5 Problems for utilitarianism

There are many technical problems with the various forms of utilitarianism. How are pleasure and pain to be measured? Which desires are to count? Is knowledge a good in itself? Should we take into account actual or probable effects on happiness? How do we characterize the possible world which is to guide us in our selection of rules? These are problems for the theorists themselves, and there has been a great deal said in attempts to resolve them.

More foundational, however, is a set of problems for any kind of utilitarian theory, emerging out of utilitarianism's peculiarly strict conception of impartiality. A famous utilitarian tag, from Bentham, is, 'Everybody to count for one, nobody for more than one'. This, however, as Mill implies (1861: ch. 5, para. 36), is slightly misleading. In a sense, according to utilitarianism, no one matters; all that matters is the level of utility. What are counted equally are not persons but pleasures or utilities.

This conception of impartiality has made it easy for opponents of utilitarianism to dream up examples in which utilitarianism seems to require something appalling. A famous such example requires a utilitarian sheriff to hang an innocent man, so as to prevent a riot and bring about the greatest overall happiness possible in the circumstances (see Crime and punishment §2).

Utilitarians can here respond that, in practice, they believe that people should abide by common sense morality, that people should accept practical principles of rights for utilitarian reasons (see §2). But this misses the serious point in many of these objections: that it matters not just how much utility there is, but how it is shared around. Imagine, for example, a case in which you can give a bundle of resources either to someone who is well-off and rich through no fault of their own, or to someone who is poor through no fault of their own. If the utility of giving the bundle to the rich person is only slightly higher than that of giving it to the poor person, utilitarianism dictates giving it to the rich person. But many (including some consequentialists) would argue that it is reasonable to give some priority to the worse-off.

These are problems at the level of the social distribution of utility. But difficulties arise also because of the fact that human agents each have their own lives to live, and engage in their practical reasoning from their own personal point of view rather than from the imaginary point of view of an 'impartial spectator'. These problems have been stated influentially in recent years by Bernard Williams (Smart and Williams 1973), who puts them under the heading of what he calls 'integrity' (see Williams, B.A.O. §4).

In a famous example, Williams asks us to imagine the case of Jim, who is travelling in a South American jungle. He comes across a military firing squad, about to shoot twenty Indians from a nearby village where some insurrection has occurred. The captain in charge offers Jim a guest's privilege. Either Jim can choose to shoot one of the Indians himself, and the others will go free, or all twenty will be shot by the firing squad.

Williams' point here is not that utilitarianism gives the wrong answer; indeed he himself thinks that Jim should shoot. Rather, it is that utilitarianism reaches its answer too quickly, and cannot account for many of the thoughts we know that we should have ourselves in Jim's situation, such as, 'It is I who will be the killer'. Practical reasoning is not concerned only with arranging things so that the greatest utility is produced. Rather it matters to each agent what role they will be playing in the situation, and where the goods and bads occur. This point emerges even more starkly if we imagine a variation on the story about Jim, in which the captain asks Jim to commit suicide so as to set an example of courage and nobility to the local populace, on the condition that if he does so the twenty Indians will go free. The utility calculations are as clear, perhaps clearer, than in the original story. But it is only reasonable that Jim in this story should think it relevant that it is he who is going to die. To any individual, it matters not only how much happiness there is in the world, but who gets it.

See also: Animals and ethics; Deontological ethics; Good, theories of the; Help and beneficence §2; Teleological ethics

ROGER CRISP
TIM CHAPPELL

References and further reading

Adams, R.M. (1976) 'Motive utilitarianism', *Journal of Philosophy* 73: 467-81.(Utilitarianism with focus on motives.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. J.H. Burns and H.L.A. Hart, revised F. Rosen, Oxford: Clarendon Press, 1996.(Highly influential statement of hedonistic act-utilitarianism.)

Brandt, R.B. (1985) *A Theory of the Good and the Right*, Oxford: Clarendon Press.(Defends hedonistic rule-utilitarianism.)

Brink, D. (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.(Defends two-level ideal utilitarianism.)

Broome, J. (1991) *Weighing Goods*, Oxford: Blackwell, ch. 1.(Contains discussion of structure of moral theories. Moderately difficult.)

Crisp, R. (1992) 'Utilitarianism and the life of virtue', *Philosophical Quarterly* 42: 139-60.(Discussion of two-level utilitarianism with focus on lives.)

Frey, R. (ed.) (1985) *Utility and Rights*, Oxford: Blackwell.(Collection devoted to discussion of rights and utilitarianism.)

Glover, J. (ed.) (1990) *Utilitarianism and Its Critics*, New York: Macmillan.(Contains essential readings, with helpful guidance by the editor. Good starting place; helpfully structured bibliography.)

Godwin, W. (1793) *Enquiry Concerning Political Justice*, ed. K. Codell, Oxford: Clarendon Press, 1971.(Radical anarchist utilitarianism.)

Goodin, R. (1991) 'Utility and the good', in P. Singer (ed.) *A Companion to Ethics*, Oxford: Blackwell, 241-8.(Introductory discussion of welfare.)

Griffin, J. (1982) 'Modern utilitarianism', *Revue Internationale de Philosophie* 141: 331-75.(Good starting place for study of recent work on utilitarianism.)

Griffin, J. (1986) *Well-Being: Its Meaning, Measurement, and Moral Importance*, Oxford: Clarendon Press.(Key discussion of different conceptions of utility. Defends ideal account with preferences entering at level of explanation of value.)

Hare, R.M. (1981) *Moral Thinking: Its Methods, Levels and Point*, Oxford: Clarendon Press.(Sophisticated modern two-level act-utilitarian, influenced by Aristotle, Kant and Mill.)

Helvétius, C. (1758) *De l'esprit (On the Mind)*, Paris: Arthème Fayard, 1988.(Enlightenment utilitarian political theorist.)

Hooker, B. (1990) 'Rule consequentialism', *Mind* 91: 67-77.(Contains defence of rule-consequentialism - also applicable to rule-utilitarianism - on grounds of undemandingness and fairness.)

Hume, D. (1751) *An Enquiry Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 3rd edn, 1975.(See esp. section 5, 'Why utility pleases', on the usefulness of the social virtues.)

Hutcheson, F. (1755) *A System of Moral Philosophy*, in *Collected Works*, vols 5-6, Hildesheim: Georg Olms, 1969.(Important Enlightenment work, anticipating Bentham and Mill.)

Kagan, S. (1989) *The Limits of Morality*, Oxford: Clarendon Press.(Defends demandingness of utilitarianism. Moderately difficult.)

Long, A.A. and Sedley, D. (1987) *The Hellenistic Philosophers*, vol. 1, Cambridge: Cambridge University Press, sections 56-67.(Contains excerpts from and discussions of Stoic ethics.)

Lyons, D. (1965) *Forms and Limits of Utilitarianism*, Oxford: Clarendon Press.(Classic discussion of act- and rule-utilitarianism.)

McCloskey, H.J. (1957) , *Philosophical Review* 66: 466-85.(Original version of the 'sheriff' case, in which a utilitarian sheriff is required to hang an innocent man, so as to prevent a riot and bring about the greatest overall happiness possible in the circumstances.)

McNaughton, D. and Rawling, P. (1995) 'Value and agent-relative reasons', *Utilitas* 7: 31-47.(Sophisticated attempt to capture consequentialist/nonconsequentialist distinction using deontic logic. Difficult.)

Mill, J.S. (1861) *Utilitarianism*, ed. R. Crisp, Oxford: Clarendon Press, 1998.(One of the most important and widely studied works in moral philosophy. Contains argument that pleasures can be seen as higher and lower.)

Miller, H.B. and Williams, W.H. (eds) (1982) *The Limits of Utilitarianism*, Minneapolis, MN: University of Minnesota Press.(Interesting and wide-ranging collection. Contains bibliography of works on utilitarianism from 1930 to 1980.)

Moore, G.E. (1903) *Principia Ethica*, ed. T. Baldwin, Cambridge: Cambridge University Press, revised edn, 1993.(Includes critique of Mill, and development of ideal account of the good.)

Nietzsche, F. (1888) *Die Götzen-Dämmerung*, trans. A.M. Ludovici, *The Twilight of the Idols*, London: Allen & Unwin, 1911, 94-6. (Criticizes the life consisting solely of happiness or pleasure.)

Nozick, R. (1984) *Anarchy, State, and Utopia*, Oxford: Blackwell.(Contains the 'experience machine' objection and defence of rights as constraints on maximization.)

Nussbaum, M. and Sen, A. (eds) (1993) *The Quality of Life*, Oxford: Clarendon Press.(Wide-ranging collection on welfare.)

Paley, W. (1785) *Principles of Moral and Political Philosophy*, New York: Garland, 1978.(Theological utilitarian, whose great success is said to have spurred Bentham to publish his *Introduction to the Principles*.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press.(Central modern work. Part IV contains discussion of population problems.)

Pettit, P. (1991) 'Consequentialism', in P. Singer (ed.) *A Companion to Ethics*, Oxford: Blackwell, 230-40.(Introductory discussion of consequentialism. Has brief bibliography.)

Plato (*c*.386-380 BC) *Protagoras*, trans. with notes by C.C.W. Taylor, Oxford: Clarendon Press, 2nd edn, 1991.(Develops early account of hedonism.)

Plato (*c*.360-347 BC) *Philebus*, trans. J.C.B. Gosling, *Plato: Philebus*, Oxford: Clarendon Press, 1975.(A discussion of the relative merits of pleasure and wisdom in the good life.)

Priestley, J. (1768) *Essay on the First Principles of Government; and on the Nature of Political, Civil, and Religious Liberty*, London: Dodsley, Cadell, Johnson; repr. in P. Miller (ed.) *Political Writings*, Cambridge: Cambridge University Press, 1993.(Influential theological utilitarian.)

Quinton, A. (1973) *Utilitarian Ethics*, London: Macmillan.(Historical introduction. Includes chapters on Bentham and Mill.)

Railton, P. (1984) 'Alienation, consequentialism, and the demands of morality', *Philosophy and Public Affairs* 13: 134-71.(On the everyday moral thinking recommended by utilitarianism.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(Significant defence of anti-utilitarian view of justice, based on 'separateness of persons'. Moderately difficult.)

Scarre, G. (1996) *Utilitarianism*, London: Routledge.(Useful introduction, including history. Contains bibliography.)

Scheffler, S. (ed.) (1988) *Consequentialism and Its Critics*, Oxford: Oxford University Press.(Collection of important recent critiques, with helpful introduction.)

Sen, A. (1981) 'Plural utility', *Proceedings of the Aristotelian Society* 81: 193-215.(Defends pluralistic conception of utility.)

Sen, A. and Williams, B. (eds) (1982) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press. (Useful collection. Introduction provides general survey, including discussion of desire theory and rights. Contains bibliography. See essays by R.M. Hare for introduction to his views, and by T.M. Scanlon for contractualist discussion of utilitarianism.)

Sidgwick, H. (1874) *The Methods of Ethics*, London: Macmillan; 7th edn, 1907, books III and IV.(Intuitionist defence of utilitarianism embedded within perceptive discussion of common sense morality.)

Smart, J.J. and Williams, B. (1973) *Utilitarianism For and Against*, Cambridge: Cambridge University Press. (Smart discusses the 'experience machine', Williams integrity.)

Sumner, L.W. (1981) *Abortion and Moral Theory*, Princeton, NJ: Princeton University Press.(Utilitarian discussion of abortion.)

Williams, B. (1976) 'Persons, Character and Morality', in A.O. Rorty (ed.) *The Identities of Persons*, Berkeley, CA: University of California Press, 197-216; repr. in *Moral Luck: Philosophical Papers 1973-80*, Cambridge: Cambridge University Press, 1981, 1-19.(Critique of impartialist ethics.)

Williams, B. (1985) *Ethics and the Limits of Philosophy*, London: Fontana, ch. 6.(Contains sophisticated critique of split-level theories. Moderately difficult.)

Utopianism

Utopianism is the general label for a number of different ways of dreaming or thinking about, describing or attempting to create a better society. Utopianism is derived from the word utopia, coined by Thomas More. In his book Utopia *(1516) More described a society significantly better than England as it existed at the time, and the word* utopia *(good place) has come to mean a description of a fictional place, usually a society, that is better than the society in which the author lives and which functions as a criticism of the author's society. In some cases it is intended as a direction to be followed in social reform, or even, in a few instances, as a possible goal to be achieved.*

The concept of utopianism clearly reflects its origins. In Utopia *More presented a fictional debate over the nature of his creation. Was it fictional or real? Was the obvious satire aimed primarily at contemporary England or was it also aimed at the society described in the book? More important for later developments, was it naïvely unrealistic or did it present a social vision that, whether achievable or not, could serve as a goal to be aimed at? Most of what we now call utopianism derives from the last question. In the nineteenth century Robert Owen in England and Charles Fourier, Henri Saint-Simon and Étienne Cabet in France, collectively known as the utopian socialists, popularized the possibility of creating a better future through the establishment of small, experimental communities. Karl Marx, Friedrich Engels and others argued that such an approach was incapable of solving the problems of industrial society and the label 'utopian' came to mean unrealistic and naïve. Later theorists, both opposed to and supportive of utopianism, debated the desirability of depicting a better society as a way of achieving significant social change. In particular, Christian religious thinkers have been deeply divided over utopianism. Is the act of envisaging a better life on earth heretical, or is it a normal part of Christian thinking?*

Since the collapse of communism in eastern Europe and the former Soviet Union, a number of theorists have argued that utopianism has come to an end. It has not; utopias are still being written and intentional communities founded, hoping that a better life is possible.

1 Background

Utopianism has existed at least as long as written history. Descriptions of better societies are found on a Sumerian clay tablet, in the Old Testament and among early Greek writers. There is considerable debate over whether or not it is a universal human phenomenon, but the only culture that seems to be lacking an early utopian tradition is Japan. The term that is used to characterize the phenomenon first occurs in the title of Thomas More's book *Libellus vere aureus nec minus salutaris quam festivus de optimo reip[ublicae] statu, deq[ue] noua Insula Vtopia* (1516), now known simply as *Utopia*. The word was based on the root *topos*, meaning *place*, and the prefix *u* or *ou*, meaning *no* or *not*. Thus, strictly speaking, utopia refers simply to a nonexistent place, but More punned on an alternative prefix, *eu* or *good*. As a result the word *utopia* came to refer to a nonexistent good place. Even though some scholars prefer to retain the distinction between *utopia* and *eutopia*, common usage conflates them (Sargent 1994).

The word *utopia* was rapidly added to most European languages and was in fairly common use by the end of the sixteenth century, being employed as part of a book title as early as 1520. The utopia as a literary genre developed quickly (Hölscher 1996); more than twenty utopias were published in the sixteenth century after More's*Utopia* (Sargent 1988), and the first study of this new phenomenon was published in 1704 (Ahlefeld 1704; see Widdicombe 1992 for a discussion of early studies).

Utopia is a small but complex book in that it presages much of the later debate over the importance of utopianism. Within *Utopia* More presented arguments for and against his invention, using arguments that have a long history and still continue to be used today.

Discussion of utopianism and its role in political philosophy surfaces as early as Aristotle, who in *Politics* criticized Plato and other projectors of better societies. Aristotle (§§27-8) argued that depicting an ideal society is the wrong approach to political theorizing. He suggested that both the criticism of contemporary societies and the development of proposals for bettering them need to be based on a detailed, realistic understanding of how human beings behave and the ways that political systems are both shaped by, and can shape, that behaviour. Plato can be

interpreted as doing just that, particularly in his *Laws*, which contrary to popular opinion is closer to later utopian thinking than his *Republic*. Aristotle's position, however, still forms one of the starting points for much contemporary criticism of utopianism. Today both Marxists and liberals use utopianism as a pejorative term, meaning 'unrealistic' or 'against human nature' or simply 'too idealistic'. Others have argued that utopianism is essential to any positive social change and may even be part of what makes us human (see Human nature §1).

In the late 1920s Karl Mannheim, primarily known as the theorist of the sociology of knowledge, used the concepts utopia and ideology as the two basic mental constructs that distort political discourse. Mannheim hoped to rid the words utopia and ideology of their negative meanings and reconstitute them as means of understanding the ways that certain social groups perceive the world around them. He first developed these ideas in his *Ideology and Utopia* (1929). In 1935, in an influential encyclopedia article, he defined the two concepts succinctly:

> The term utopian, as here used, may be applied to any process of thought which receives its impetus not from the direct source of reality but from concepts, such as symbols, fantasies, dreams, ideas and the like, which in the most comprehensive sense of that term are non-existent. Viewed from the standpoint of sociology, such mental constructs may in general assume two forms: They are ideological if they serve the purpose of glossing over or stabilizing the existing social reality; utopian if they inspire collective activity which aims to change such reality to conform with their goals, which transcend reality.
>
> (Mannheim 1935)

Mannheim's definitions influenced the use of the words, although not in the way he intended. His definition reinforced the old message that to be utopian meant to be naïve and out of touch with the real world. Mannheim scholars now consider *Ideology and Utopia* to be a minor part of his work, but his definitions are still used, not always accurately, by scholars of both ideology and utopianism (see Ideology §1).

2 Utopian socialism

Historically the most important critics of utopianism after Aristotle were Karl Marx and Friedrich Engels. In his *Die Entwicklung des Sozialismus von der Utopie zur Wissenschaft (Socialism: Utopian and Scientific)*, published in 1882, Engels coined the phrase 'utopian socialist' to describe Owen, Fourier, Saint-Simon and, in some editions, Cabet, whose idealistic proposals he regarded as distinctly inferior to Marx's 'scientific socialism'. Among these writers only Cabet wrote a work that a narrow definition would call a utopia - *Voyage en Icarie* (1840). Owen and Fourier described social systems based on relatively small communities, now generally called 'intentional communities' but in the past frequently called 'utopian communities'. Owen and Cabet actually participated in the development of such communities - Owen in the UK and the USA and Cabet in the USA - while followers of Fourier and Saint-Simon built them, mostly in France and the USA. Such communities are now generally thought of as examples of utopianism, and there are currently thousands of them based in at least twenty-one countries, the best-known example being the kibbutz movement in Israel.

Engels' primary target was not the building of small communities; in fact at one time he actually praised such community construction, using the example of the Rappites in Pennsylvania. What bothered the Marxists was that the utopian socialists made what Marx and his followers thought were unwarranted assumptions about human nature and about their own ability to predict and construct a better future. For all of his desire to overthrow the currently existing socio-economic system, Marx was adamant that a future constructed by people freed from domination could not be predicted. Marxists also argued that the utopian socialists were extremely naïve about the problems of constructing such a future society and the opposition that would be brought against any attempt to do so. Marxists contended that the construction of small communities was the wrong approach because they would never be allowed to succeed or, if they were, it would be because they were no threat to those in power. None of the communities built by the utopian socialists lasted very long (the seven Icarian communities founded by the followers of Cabet had a combined life span of fifty years), and none ever posed a threat. In contrast, nearing the end of the twentieth century, certain religious intentional communities have been in existence for nearly two hundred years, while fifty years is no longer uncommon and twenty-five years is almost the norm, even for secular communities; however, as before, these are not in any real sense threats to the existing order in any of the countries in which they exist (see Communism).

3 Utopian social theory

While the main line of criticism of utopianism initially developed in the context of utopian socialism and the small community, it was not that aspect of utopianism that became the focus for either its attackers or defenders, and Marxists have appeared in both ranks. Both Marxists and anti-Marxists argue that utopian thinking is naïve and unrealistic and makes assumptions about 'human nature' that will not withstand scrutiny. Marxists argue that 'human nature' is socially constructed and varies from society to society and time to time. Conservative, mostly Christian, opponents argue that sinful human nature is incapable of the improvements that they believe utopianism requires (Molnar 1967). It is true that many utopians have assumed that human nature is fundamentally 'good', usually meaning cooperative, but that contemporary society distorts that goodness; a new, better society would allow it to be expressed. However, such thinking is by no means the norm among utopians.

The modern debate around utopianism developed historically from its rejection. At the turn of the century, Edward Bellamy's novel, *Looking Backward 2000-1887* (1888), depicting a centralized socialist (he called it nationalist) utopia became the best-selling novel published up to that time and remained so until well into the twentieth century. Nationalist societies were founded in countries around the world. Bellamy effectively rekindled interest in utopianism after Engels had destroyed it. Since then, although deeply disputed, utopianism has never been ignored, even though for one fairly lengthy period it was the anti-utopians that kept it visible.

The anti-utopians generally regard utopias as blueprints of a desired perfect future society that must be achieved in all of its details. Since perfection is beyond human capabilities, the utopians, it is believed, will have to resort to force and violence to achieve utopia. Therefore, utopias are a first step that almost inevitably leads to totalitarianism (see Totalitarianism). Karl Popper is the best-known exponent of this position. In *The Open Society and Its Enemies*, he wrote: 'the Utopian approach can be saved only by the Platonic belief in one absolute and unchanging ideal, together with two further assumptions, namely (a) that there are rational methods to determine once and for all what this ideal is, and (b) what the best means of its realization are' ([1945] 1957, vol. 1: 161).

Utopians deny that utopias are the problem; the problem is with the believers. They argue that utopias are not blueprints of some future perfect society but act both as a critique of contemporary society and as a generalized goal rather than something to be realized in all its details. Utopias are seen by their defenders as expressions of the highest human aspirations.

This position is put most forcefully by the Dutch sociologist F.L. Polak in his *The Image of the Future*. He argued that our images of the future (whether utopias or dystopias) affect the actual future: 'We will view human society and culture as being magnetically *pulled* towards a future fulfilment of their own preceding and prevailing, idealistic images of the future, as well as being *pushed* from behind by their own realistic past' (1961, vol. 1: 15; original emphasis). Polak goes on to contend that 'if Western man now stops thinking and dreaming the materials of new images of the future and attempts to shut himself up in the present, out of longing for security and for fear of the future, his civilization will come to an end. He has no choice but to dream or to die, condemning the whole of Western society to die with him' (vol. 1: 53).

Ernst Bloch (§3) has become the most influential utopian thinker of the twentieth century. Bloch has become important for utopians because he argues that there always exists unrealized potential for positive social change. In other words, there are always utopian possibilities in any given situation. For scholars of utopianism, Bloch is significant today because in his most important work, *Das Prinzip Hoffnung* (*The Principle of Hope*, 1959), he demonstrates the long history of utopianism and its important contribution to social improvement.

For Bloch utopia is constantly on the horizon, not yet achieved, but open to the possibility of being achieved. It is these utopian aspirations and the hope that they can be realized that make human life bearable and a better life possible. For Bloch, human beings are active participants in creating that better future; we must think and act experimentally, driven by the utopian visions that Bloch found in art, literature, music, and social theory. In this he separates himself from the more deterministic versions of Marxism while at the same time seeing Marx as one of the great utopians of the past.

Utopianism has been a particular problem for Christian theorists, with some seeing any expression of utopianism as heretical (Molnar 1967) and others seeing it as an essential aspect of Christianity (Tillich 1971).

4 The end of utopia

Recent discussion of utopianism has focused on the so-called 'end of utopia'. Proponents of this have tended to identify utopian aspirations with the communist regimes of the former Soviet Union and eastern Europe. Thus, they saw in the collapse of communism a vindication of the futility of utopianism in the face of the obdurate material of human nature. The end of utopia was predicated on the same mis-identification of utopia with totalitarianism found earlier in Popper.

Just as with the 'end of ideology' debate of the 1950s - which was based on almost precisely the same arguments - utopia has not ended. In fact, it has been argued that the collapse of communism (a utopia that became a dystopia) was based on the development of a new and powerful utopia within communist countries, a utopia of freedom and materialism. Generalized, the position is that a new utopia is always needed to overcome the old dystopia. This point was made by the late Viktoriia Chalikova in *Utopiia rozhdaetsia iz utopii: cesse raznykh let (Utopia is Born by Utopia*, 1992).

Utopianism remains alive and well in all three of its forms. A constant stream of utopian novels is published in a wide variety of languages. The utopias found in these novels are more complex and self-reflective than in the earlier tradition, less certain that utopia can be both created and maintained, more aware both of the powerful forces marshalled against social betterment and the dangers inherent in the utopian enterprise, dangers pointed out by the anti-utopians. This has led one critic, Tom Moylan, to argue in his *Demand the Impossible* (1986) that a new type of utopia, the 'critical utopia', has been created. Moreover, since the beginning of the 1970s feminist novelists and theorists have produced a steady and continuing stream of feminist utopias.

There are more long-lived and well-established intentional communities currently in existence than at any time in the past, including the 1960s, when there was a great upsurge of community building. While generally little known, they are found throughout the world, with the greatest numbers in Australia, Europe and North America, and a number of ecologically based communities have recently been founded in the former Soviet Union. These communities express the desire of thousands of people to live a better life now.

Utopian theorists are also publishing more than ever before, arguing that social dreaming is both essential and potentially dangerous. Alongside Bloch and Polak, such theorists see the need to have utopian aspirations, but they are also intensely aware that there are always people ready to impose their utopia on others. None the less, the utopians argue, this does not in any way diminish the need to dream of a better life for all.

See also: Saint-simon, Comte de

LYMAN TOWER SARGENT

References and further reading

Ahlefeld, H. ab (1704) *Disputatio philosophica de fictis rebuspublicis. Quam divina favente gratia, praeside Georgio Paschio, Artis rationis, Philos. In Auditorio Majori publice defendet Henricus ab Ahlefeld, Esques Holsatus (A Philosophical Argument Concerning Fictional Republics),* Kiloni: Typis Bartholdi Reutheri; trans. W. Biesterfeld, 'Ein früher Beitrag zu Begriff und Geschichte der Utopie. Heinrich von Ahlefeldts *Disputatio philosophica de fictis rebuspublicis*' (An Early Contribution to the Concept and History of Utopia), *Archiv für Begriffsgeschichte* 16 (1): 28-47, 1972.(Earliest study of utopianism.)

Bellamy, E. (1888) *Looking Backward: 2000-1887,* ed. C. Tichi, Harmondsworth: Penguin, 1982.(The most popular utopian novel ever published.)

Bloch, E. (1959) *Das Prinzip Hoffnung*, Frankfurt: Suhrkampf Verlag; trans. N. Plaice, S. Plaice and P. Knight, *The Principle of Hope,* Oxford: Blackwell, 1986, 3 vols.(Bloch's most important work.)

Buber, M. (1949) *Paths in Utopia*, trans. R.F.C. Hull, Boston, MA: Beacon Press.(Early work arguing for the importance of utopian goals for social change.)

Cabet, E. (1840) *Voyage en Icarie,* Paris: Bureau au Populaire.(Utopian novel.)

Chalikova, V.A. (1992) *Utopiia rozhdaetsia iz utopii: cesse raznykh let (Utopia is Born by Utopia)*, London: Overseas Publications Interchange.(Argues that utopias are essential for overcoming regimes that have become dystopic.)

Dahrendorf, R. (1958) 'Out of Utopia: Toward a Reorientation of Sociological Analysis', *American Journal of Sociology* 64 (September): 115-27.(Anti-utopian.)

Engels, F. (1882) *Die Entwicklung des Sozialismus von der Utopie zur Wissenschaft*, Zurich: Schweizerische Genossenschaftsdruckerei; trans. E. Aveling (1892) *Socialism: Utopian and Scientific*, New York: International

Publishers, 1935.(Classic Marxist statement on utopian socialism.)

Goodwin, B. (1980) ‘Utopia Defended Against the Liberals’, *Political Studies* 28 (3): 384-400.(Argument for the importance of utopianism.)

Hölscher, L. (1996) ‘Utopie’, *Utopian Studies* 7 (2): 1-65.(A history of the concept of utopia.)

Kateb, G. (1963) *Utopia and Its Enemies*, New York: Free Press.(Early study of utopian and anti-utopian thinkers.)

Levitas, R. (1990) *The Concept of Utopia*, Hemel Hempstead: Philip Allan, and Syracuse, NY: Syracuse University Press.(Best consideration of utopianism in recent social theory.)

Mannheim, K. (1929) *Ideology and Utopia: An Introduction to the Sociology of Knowledge*, trans. L. Wirth and E. Shils, London: Routledge, 1936; new edn, 1991.(Classic treatise on the sociology of knowledge.)

Mannheim, K. (1935) ‘Utopia’, in *Encyclopedia of the Social Sciences*, New York: Macmillan, vol. 15, 200-3. (Definition of the concept from Mannheim’s viewpoint.)

Manuel, F.E. (ed.) (1967) *Utopias and Utopian Thought*, Boston, MA: Beacon Press.(Collection of essays on aspects of utopianism.)

Manuel, F.E. and Fritzie, P.M. (1979) *Utopian Thought in the Western World*, Cambridge, MA: Belknap Press. (Best general history of utopian thought.)

Minerva, N. (ed.) (1992) *Per una definizione dell’utopia: Metodologie e discipline a confronto. Atti del Convegno Internazionale di Bagni di Lucca 12-14 settembre 1990 (On the Definition of Utopia)*, Ravenna: Longo Editore. (Excellent collection on various aspects of utopian studies. Includes essays in English and French as well as Italian.)

Molnar, T. (1967) *Utopia: The Perennial Heresy*, New York: Sheed & Ward.(Anti-utopian.)

More, T. (1516) *Utopia*, ed. G.M. Logan and R.M. Adams, Cambridge: Cambridge University Press, 1989.(The origin of the word utopia and an early statement of communism and religious toleration.)

Moylan, T. (1986) *Demand the Impossible: Science Fiction and the Utopian Imagination*, London: Methuen. (Exposition of the ‘critical utopia’.)

Polak, F.L. (1961) *The Image of the Future; Enlightening the Past, Orientating the Present, Forecasting the Future*, New York: Oceana Publications, 2 vols.(Strongest pro-utopian argument.)

Popper, K.R. (1945) *The Open Society and Its Enemies*, London: Routledge & Kegan Paul, 3rd edn, 1957, 2 vols. (Strongest anti-utopian argument.)

Popper, K.R. (1948) ‘Utopia and Violence’, *Hibbert Journal* 46 (January): 109-16.(Anti-utopian.)

Saage, R. (ed.) (1992) *Hat die politische Utopie eine Zukunft? (Does Political Utopia Have a Future?)*, Darmstadt: Wissenschaftliche Buchgesellschaft.(Essays on the ‘end of utopia’.)

Sargent, L.T. (1982) ‘Authority and Utopia: Utopianism in Political Thought’, *Polity* 14 (4): 565-84.(Argument for the importance of utopianism.)

Sargent, L.T. (1988) *British and American Utopian Literature 1516-1985: An Annotated, Chronological Bibliography*, New York: Garland.(The most comprehensive bibliography of materials in English.)

Sargent, L.T. (1994) ‘The Three Faces of Utopianism Revisited’, *Utopian Studies* 5 (2): 1-37.(General consideration of various aspects of utopianism.)

Tillich, P. (1971) ‘The Political Meaning of Utopia’, in *Political Expectation*, trans. W.J. Crout, W. Bense and J.L. Adams, New York: Harper & Row, 125-80.(Strong pro-utopian argument.)

Utopian Studies (1990-).(Best journal in the field.)

Widdicombe, R.T. (1992) ‘Early Histories of Utopian Thought (to 1950)’, *Utopian Studies* 3 (1): 1-38.(Survey of early research on utopianism.)

Vagueness

It seems obvious that there are vague ways of speaking and vague ways of thinking - saying that the weather is hot, for example. Common sense also has it that there is vagueness in the external world (although this is not the usual view in philosophy). Intuitively, clouds, for example, do not have sharp spatiotemporal boundaries. But the thesis that vagueness is real has spawned a number of deeply perplexing paradoxes and problems. There is no general agreement among philosophers about how to understand vagueness.

1 Vagueness in language and the world

Many linguistic terms are vague. But of what exactly does their vagueness consist? Consider, for example, the term 'bald'. This term has borderline cases of application. There are people to whom the term, as it is ordinarily used, neither clearly applies nor clearly fails to apply. In this respect, the term 'bald' is different from the term 'square root', say. There are no borderline square roots, nor could there be. Let us say that a general term (a word or phrase which potentially applies to many things) is intensionally vague (that is, vague with respect to its meaning) only if its meaning permits possible borderline cases of application. Then 'bald' is intensionally vague.

Vagueness can also be found in singular terms (words or phrases that potentially apply to single things). Consider, for example, 'the friend of Amy'. Suppose Amy has a love-hate relationship with Jane, without Amy clearly having any other friends. Then, it is indeterminate whether the singular term, 'the friend of Amy', designates Jane. So, 'the friend of Amy' has a possible borderline case of designation. And this indicates that it is intensionally vague.

It is plausible to suppose that intensional vagueness in general terms is more basic than intensional vagueness in singular terms, since uncontroversial examples of vague singular terms either contain vague general terms (non-relational or relational) or are the abstract singular counterparts to those terms. In the case of 'the friend of Amy', the term 'friend of' is intensionally vague.

The vagueness so far discussed does not require vagueness in the world. But worldly vagueness can be treated in a similar way. Suppose, as we normally do, that there really is a nonlinguistic property of being red, a property expressed by the predicate 'is red'. This predicate is vague and so is the property. Just as the former has borderline cases of application, so too the latter has borderline instances (objects which are neither clearly red nor clearly not red). Moreover, even if the actual world had not contained borderline red objects, still, intuitively, that would not have shown that the property, redness, is not vague. What the vagueness of properties seems to require is that there be possible borderline instances. Consider next concrete objects, for example, Mount Everest. Some molecules are definitely inside Everest and some are definitely outside. But intuitively, some have a borderline status: there is no determinate fact of the matter about whether they are inside or outside. Everest, then, has borderline spatiotemporal parts. In this way, it is like a cloud. So Everest is a vague concrete object.

There is another way in which, according to some philosophers, objects in the world can be vague. If, for a given object, o, there is an object, o', such that it is indefinite whether o is identical with o', then o, by virtue of its entering into an indefinite identity relation, is a vague object. Here is one possible example (due to D. Parfit, cited in Broome 1984). There is a club which has a clubhouse, a membership list and a set of rules. This club is never formally disbanded, but through time its members meet less and less frequently and the clubhouse becomes run down. There are no meetings for several years. Twelve years later, however, a few of the original members get together with some new people and start to meet once more in the same building (now redecorated). The club they belong to at this later date has the same name as the earlier one.

It has been held that the claim that the first club is identical with the second one is indefinite. Moreover, this indefiniteness, it has been suggested, is not an epistemic matter, since there is no further information which would settle the issue. On this view, each club has a vague identity and is thereby a vague object.

There is at least one respect in which the proposed characterizations of vagueness are incomplete. Consider the following sequence of conditionals, each of which includes the vague term 'tall':

(1a) If a man whose height is 7 feet is tall, then a man whose height is 6 feet 11 and $\frac{99}{100}$ inches is tall.

(1b) If a man whose height is 6 feet 11 and $\frac{99}{100}$ inches is tall, then a man whose height is 6 feet 11 and $\frac{98}{100}$ inches is tall.

⋮

(1*n*) If a man whose height is 5 feet and $\frac{1}{100}$ inches is tall, then a man whose height is 5 feet is tall.

Intuitively, there is no sharp dividing line between the true conditionals in this sequence and those that have some other value. This fact is not captured in the earlier account of intensional vagueness. But it seems to be part and parcel of our ordinary conception of the vagueness of terms such as 'tall' or 'bald'. Their vagueness is robust or resilient. And a corresponding point can be made about vague concrete objects. For example, it does not seem to be true that there is a sharp dividing line between the molecules that are inside Everest and those that have a borderline status.

2 Paradoxes and problems

The oldest puzzle of vagueness, which allegedly derives from Eubulides, is the paradox of the bald man. It goes as follows:

(2) A man with no hairs on his head is bald.
(3) For any number, *n*, if a man with *n* hairs on his head is bald then a man with $n+1$ hairs on his head is bald.
(4) Therefore, a man with ten thousand hairs on his head is bald.

The conclusion is derived from the premises via ten thousand applications of the classical logical rules of *modus ponens* and universal instantiation, so this argument is valid. Now, premise (2) is certainly true and the conclusion, (4), is certainly false, so it is inferred that premise (3) is false. Therefore, there is an *n* such that a man with *n* hairs on his head is bald and a man with $n+1$ is not. Therefore, the term 'bald' is precise, contrary to appearances. And what is true in this one case is true by parallel reasoning for any general term which is ordinarily classified as vague.

It is worth observing that the paradox of the bald man does not rely essentially on the use of a universal generalization. The same result can be generated by means of a sequence of (ten thousand) conditionals such as in (1a)-(1n), with (2) as a starting point. Since the conclusion, (4), is false, and (2) is true, then at least one of the conditionals must be false. So, given that (2) is true and that each conditional is either true or false, there must be an adjacent pair of conditionals such that the first is true and the second false. This runs directly counter to the idea that 'bald' is vague.

If the paradox of the bald man really does demonstrate that the term 'bald' is not vague, contrary to what we all ordinarily believe, then 'bald' cannot express a vague property. So, the property of being bald cannot itself be vague. Since the paradox of the bald man can be restated so as to apply to any vague general term, one overall conclusion which has been reached is that no general terms or properties are vague. This conclusion, of course, not only threatens the thesis that there are vague abstract objects but also challenges the idea that there is any such thing as vagueness in language at all, however it is understood.

There are also sorites arguments which attack the concrete objects of common sense, on the assumption that these objects have vague boundaries. These arguments, like the paradox of the bald man, can be stated without the use of a universal generalization.

There are other problems for the view that individual concrete objects are vague. One of these is directed against vague objects, conceived of as objects having indeterminate identities (Evans 1978). Suppose that '*a*' and '*b*' are (precise) singular terms for vague objects and that '$a = b$' is indefinite in truth-value. Then, if we let '∇' symbolize 'indefinitely', the following is true:

(5) $\nabla(a = b)$

(5) ascribes to *b* the property of being indefinitely identical with *a*. Now surely we have

(6) $\sim\nabla(a = a)$

and hence that *a* lacks the property of being indefinitely identical with a. By the principle that if object *o* and object *o′* differ in a property, then they are not identical (the contrapositive of Leibniz's Law; see Identity), it follows that *a* is not identical with *b*. Since, on the standard understanding of 'indefinitely', it cannot be the case both that it is indefinite whether *a* is identical with *b* and that *a* is not identical with *b*, there cannot be vague identities.

The deepest problems of vagueness, in my view, are the sorites paradoxes. And these arise whether or not there is vagueness in the nonlinguistic world.

3 Theories

It has been suggested that a proper understanding of vagueness requires the admission that truth and set membership come in degrees, rather than being 'all or nothing' (see Zadeh 1965, Goguen 1969). On this approach, real numbers in the interval from 0 to 1 are typically taken to be truth-values, with 1 being fully fledged truth and 0 being fully fledged falsity. The same numbers are assigned to the degrees to which objects belong to sets. So, for example, if Herbert is clearly bald, then the sentence

(7) Herbert is bald

is assigned the value 1, and Herbert is taken to belong to the set of bald men to degree 1. But if Herbert is a borderline case then (7) is assigned some number less than 1, say, 0.6, and the degree of Herbert's membership in the set of bald men is now taken to be 0.6 too. In general, a singular sentence '*Fa*' is treated as having the truth-value *n*, where $0 \leq n \leq 1$, if and only if the referent of '*a*' belongs to the set of *F*s to degree *n* (see Fuzzy logic).

A consequence of this approach is that vague predicates do not sharply divide the world into those things to which they apply and those things to which they do not. Rather, vague predicates apply to objects to varying degrees. As a man loses hair he becomes more and more bald; so the predicate 'is bald' applies to him more and more and the assertion that he is bald increases gradually in its degree of truth.

One serious objection to this view is that it really replaces vagueness with the most refined and incredible precision. Set membership, as viewed by the degrees-of-truth theorist, comes in precise degrees, as does predicate application and truth. The result is a commitment to precise dividing lines that is not only unbelievable but also thoroughly contrary to how we think of everyday vagueness (as noted in §1).

It is often supposed by philosophers that vagueness resides in language. One popular view here is that vagueness itself is a matter of semantic indecision (for example, Lewis 1986). On this view, all external objects, properties and relations are precise. Each vague linguistic term has a meaning which can be made precise in a whole range of different permissible ways. For example, 'over 70 years of age' is one acceptable way of making precise the meaning of the term 'old'. But other equally good ways are 'over 72' and 'over 74'. The rules that govern the use of the term 'old' are not specific enough for us to be able to choose non-arbitrarily some one precise term as capturing its meaning. When the term 'old' was first introduced, no decision was made to link it exclusively with one particular precise property. Instead it was applied widely in a range of cases, thereby acquiring a vague meaning. And what is true here for 'old' is true *mutatis mutandis* for all other vague terms.

Now if it is indeed true that there are no vague properties expressed by vague predicates, but instead ranges of precise properties, then, under some ways of making precise a vague predicate, it will definitely apply to a given object (in virtue of the object having the appropriate precise property) and, under other ways, it will not. Suppose that Alfred is 67. If 'old' is sharpened to mean 'over 70' then Alfred is not old: he lacks the precise property of being over 70. But if 'old' is sharpened to mean 'over 65', then Alfred is old - this time he has the right precise property. Is it really true that Alfred is old, then? If these two ways of sharpening the term 'old' are equally acceptable - if they both express precise properties falling in the range associated with 'old' - then the natural answer is that it is neither true nor false that Alfred is old. He is a borderline case. So, there really is no definite age at which Alfred becomes old: the boundary between being old and not being old is not sharp. By contrast, if Alfred is 97 years of age, then, under any acceptable way of making 'old' precise, he will count as old. So here it is unquestionably true that Alfred is old. And if he is 21, then, no matter which acceptable precisification we choose for 'old', he will not count as old. The claim that Alfred is old is false.

The thesis that vagueness is semantic indecision thus leads straightforwardly to the following claims. Vague sentences have three possible truth-values: true, false and ‘indefinite’ (neither true nor false). A vague sentence is to be counted as true if it comes out true under all acceptable ways of making precise its component vague terms; as false if it comes out false in every such case; and as neither true nor false if it comes out true under some ways and false under others. These claims entail that the law of excluded middle (LEM) remains true for vague sentences. So, the semantic indecision theory of vagueness provides us with a classical conception of reality (all that there is in the world is precise), a safe haven for vagueness in the linguistic realm and a conservative attitude towards the retention of LEM.

There are grave difficulties, however. In particular, it is far from clear that the semantic indecision approach fares any better than the degrees-of-truth view with respect to sorites paradoxes and robust or resilient vagueness. Another objection concerns LEM. The objection is that

(8) Either Herbert is bald or Herbert is not bald,

to take one instance of LEM, should not be counted as true if Herbert is a borderline bald man. For if (8) is true then either Herbert is bald or he is not bald. If this is the case then the question, ‘Well, which is he then?’, must surely have an answer; that is, if (8) is true then precisely one of the disjuncts in (8) must be true. So ‘Herbert is bald’ must be either true or false. This runs contrary to the assumption that Herbert is a borderline bald man. So, (8) is not true.

One common feature of both the standard degrees-of-truth approach and the semantic indecision theory is that they employ precise metalanguages in stating the truth-conditions for vague object language sentences (a metalanguage is used to talk about the object language). It is this feature which, in my view, dooms them to failure. A central aspect of the resilient vagueness of ordinary terms is that, in sorites sequences, there simply is no determinate fact of the matter about the transition from true to some other value. Any attempt to state truth-conditions for vague discourse in precise language will inevitably fall foul of this fact.

Another possible approach, then, is to start out using a vague metalanguage, which mirrors in its vagueness the object language. One proposal along these lines is based on the three truth-values, true, false and indefinite. Just as sentences in the object language can be indefinite, so too can sentences in the metalanguage (and indeed in all the higher metalanguages). If this approach is to avoid the sorites paradoxes, it is essential that it not be true that every object language sentence is true or false or indefinite. For this would create sharp dividing lines. But neither can it be false that every such sentence is true or false or indefinite. For this would require further truth-values. Instead, it has been argued that the above generalization about object language sentences must be indefinite. This permits us to hold that the claim that there is a last true sentence, followed by a first indefinite one, in a sorites sequence is indefinite. And if this is the case then sorites arguments cannot be sound (see Tye 1990).

There are many philosophers who would reject all three of the alternative proposals sketched so far. The fourth and final approach I shall mention is the epistemic view (Sorensen 1988, Williamson 1994). According to this position, vagueness is a kind of ignorance. The nonlinguistic world is precise. Standard logic holds even for vague discourse. Every molecule is either inside or outside Everest, for example. We just do not know where the boundaries lie. Likewise, there is always a single hair, the addition of which would turn a bald man into a man who is not bald, even though we cannot say which hair this is. Our sensory and conceptual mechanisms are simply not equipped to make the necessary fine-grained discriminations. Sorites arguments, then, rest upon a false premise.

This view is pleasingly straightforward. Unfortunately, it seems counterintuitive. It denies outright the existence of robust or resilient vagueness. Moreover, it seems to misconceive borderline cases. Our ordinary concept of a borderline case is the concept of a case that is neither one thing nor the other. ‘Definitely’ here does not seem to mean ‘known’ or ‘knowably’.

So, whichever way we turn, we quickly become enmeshed in difficulties. Of all the philosophical mysteries, vagueness is one of the most perplexing.

See also: Many-valued logics, philosophical issues in; Logical and mathematical terms, glossary of

MICHAEL TYE

References and further reading

Broome, J. (1984) 'Indefiniteness in Identity', *Analysis* 44: 6-12.(Discussion of whether objects can have vague identities.)
Dummett, M. (1975) 'Wang's Paradox', *Synthè se* 25: 301-24.(Vague observational predicates as logically incoherent.)
Evans, G. (1978) 'Can There Be Vague Objects?', *Analysis* 38: 208.(An argument against the existence of vague objects.)
Fine, K. (1975) 'Vagueness, Truth, and Logic', *Synthèse* 25: 265-300.(Classic early presentation of vagueness as semantic indecision.)
Forbes, G. (1985) *The Metaphysics of Modality*, Oxford: Clarendon Press.(Includes a version of the degrees-of-truth approach.)
Goguen, J. (1969) 'The Logic of Inexact Concepts', *Synthè se* 19: 325-73.(An early statement of the degrees-of-truth theory.)
Heller, M. (1990) *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*, Cambridge: Cambridge University Press.(Sorites arguments deployed in defence of the view that the world is precise.)
Horgan, T. (ed.) (1994a) *Spindel Conference on Vagueness*, supplement to *Southern Journal of Philosophy* 33. (Good collection of new papers on vagueness.)
Horgan, T. (ed.) (1994b) 'Robust Vagueness and the Forced March Sorites Paradox', *Philosophical Perspectives* 8: 159-88.(A new form of sorites reasoning and its consequences.)
Inwagen, P. van (1988) 'How To Reason About Vague Objects', *Philosophical Topics* 16: 255-84.(A non-classical logic for reasoning about vague objects.)
Lewis, D.K. (1986) *The Plurality of Worlds*, Oxford: Blackwell.(Simple statement of the semantic indecision theory.)
Lewis, D.K. (1993) 'Many, But Almost One', in K. Campbell, J. Bacon and L. Reinhardt (eds) *Ontology, Causality, and Mind: Essays on the Philosophy of D.M. Armstrong*, Cambridge: Cambridge University Press. (Arguments against the existence of vague objects.)
Putnam, H. (1983) 'Vagueness and Alternative Logic', *Erkenntnis* 19: 297-314.(A discussion of vagueness and non-classical logic.)
Sainsbury, M. (1991) 'Is There Higher-Order Vagueness?', *Philosophical Quarterly* 41: 167-82.(Discussion of whether there can be borderline borderline cases.)
Sanford, D. (1975) 'Borderline Logic', *American Philosophical Quarterly* 12: 29-39.(Formal development of the degrees-of-truth approach.)
Sorensen, R. (1988) *Blindspots*, Oxford: Clarendon Press.(Vagueness as ignorance.)
Thomason, R. (1982) 'Identity and Vagueness', *Philosophical Studies* 42: 329-32.(Discussion of Evans' argument from identity against the existence of vague objects.)
Tye, M. (1990) 'Vague Objects', *Mind* 99: 535-57.(Vagueness in both the object language and the metalanguage; defence of vagueness in the world.)
Unger, P. (1979a) 'There are No Ordinary Things', *Synthèse* 41: 117-54.(Sorites arguments against the existence of ordinary things.)
Unger, P. (1979b) 'Why There are No People', *Midwest Studies in Philosophy* 4: 177-222.(Sorites arguments against the existence of people.)
Wheeler, S. (1979) 'On That Which is Not', *Synthèse* 41: 155-73.(Sorites arguments against the existence of ordinary things.)
Williamson, T. (1994) *Vagueness*, London: Routledge.(Clear defence of the view that vagueness is ignorance.)
Williamson, T. (1975) 'On the Coherence of Vague Predicates', *Synthèse* 30: 325-63.(Broad examination of the nature of vagueness.)
Zadeh, L. (1965) 'Fuzzy Sets', *Information and Control* 8: 338-53.(Early presentation of the degrees-of-truth approach.)

Vaihinger, Hans (1852-1933)

Hans Vaihinger was a German philosopher and historian of philosophy. Much of his work was a response to Kant's philosophy, and he contributed to the revival of interest in Kant at the end of the nineteenth century both in his published commentaries and in founding a journal and society for the discussion of Kant's thought. He developed his own philosophy, the philosophy of 'as-if', which was derived from the Kantian notion of 'heuristic fictions'.

Vaihinger was born near Tübingen in Germany. He studied theology and philosophy, graduating from the Tübingen seminary. He taught at Strasbourg and later Halle until 1906, when he retired because of failing eyesight.

In 1881 and 1892 he published two volumes of his massive *Kommentar zu Kants Kritik der reinen Vernunft (Commentary on Kant's Critique of Pure Reason)*, a detailed commentary, not only on the first seventy pages of Kant's *Critique of Pure Reason*, but also on previous commentaries. This still valuable work was never completed. However, the numerous articles that he wrote set out his views on the major sections of the *Critique* not covered in the *Kommentar*.

Vaihinger never counted himself a member of any particular school of Neo-Kantianism, but contributed to the renewal of interest in Kant. In 1897 he founded the journal *Kant-Studien* (Kant Studies). In the first issue he claimed that its purpose was to avoid partisan approaches to Kant scholarship that were prevalent in the Neo-Kantian 'school' journals and instead to provide an arena for all Kantian scholars. However, Vaihinger could be as polemical as other Neo-Kantians. In 1904 he formed the *Kant-Gesellschaft*, a society whose membership funded *Kant-Studien*.

With his retirement from teaching,Vaihinger resumed work on the philosophy of 'as-if', begun in 1876. He drew from Lange and Schopenhauer as well as Plato, Darwin and Nietzsche, but above all from Kant. It was primarily the Kantian notion of 'heuristic fictions' that led him to his philosophy of 'as-if'. In the *Critique of Pure Reason* Kant argues that although we cannot prove that the soul is a simple substance, it serves as a regulative idea: we act *as if* the soul is indivisible (see Kant, I. §8). For Vaihinger this counts as a 'fiction' and is to be distinguished from a 'hypothesis'. While we cannot always differentiate between a fiction and a hypothesis, the latter must be capable of being confirmed. Once found to be true a hypothesis loses that status. A fiction, however, can never be confirmed; its validation comes in being justified. Thus the Kantian 'fiction' that we must act as if we are free moral agents is justified by the moral actions that result from that belief. Vaihinger argued that fictions are useful in a wide range of areas. For example, the notions of the atom and the 'thing-in-itself' are important for science and metaphysics. Fictions also have consequences for religion and ethics. Reflecting his early religious training he suggested that the fiction of the 'virgin birth' is particularly 'beautiful'. And he argued that a fiction of major significance in terms of human action is to act *as if* we were morally free agents.

In 1911, Vaihinger and Raymond Schmidt founded *Annalen der Philosophie (Annals of Philosophy)* to further the philosophy of 'as-if'.

CHRISTOPHER ADAIR-TOTEFF

List of works

Vaihinger, H. (1876) *Hartmann, Düring, und Lange*, Iserlohn: J. Baedeker.(An early work, documenting Vaihinger's influences.)

Vaihinger, H. (1881-92) *Kommentar zu Kants Kritik der reinen Vernunft (Commentary on Kant's Critique of Pure Reason)*, Stuttgart: Union deutsche Verlagsgesellschaft; 2nd edn, 1922.(An indispensable but difficult and lengthy account of the *Critique of Pure Reason* which focuses on the first 75 pages.)

Vaihinger, H. (1902a) *Nietzsche als Philosoph (Nietzsche as Philosopher)*, Berlin: Reuther & Reichard.(An early, objective and straightforward account of Nietzsche's thought.)

Vaihinger, H. (1902b) *Die Transcendentale Deduktion der Kategorien (The Transcendental Deduction of the Categories)*, Halle: Niemeyer.(An early but disputed account of the transcendental deduction in which Vaihinger claims that it is a 'patchwork' of ideas from different stages in Kant's development.)

Vaihinger, H. (1911) *Die Philosophie des als-ob*, Berlin: Reuther & Reichard; 10th edn, 1927; trans. C.K. Ogden,

The Philosophy of As-If, London: Harcourt, Brace, 1924.(A lengthy and rather difficult work in which Vaihinger sets out his notion of fiction, based on his reading of Kant and Nietzsche.)

References and further reading

Del Negro, W. (1934) 'Hans Vaihingers philosophisches Werk mit Besonderer Berücksichtigung seiner Kantforschung' (Hans Vaihinger's Philosophical Work, With Special Consideration of His Kant Investigations), *Kant-Studien* 39: 316-27.(A slightly uncritical overview of Vaihinger's scholarly life.)

Handy, R. (1967) 'Hans Vaihinger', in *Encyclopedia of Philosophy*, ed. P. Edwards, New York: Macmillan, vol. 8, 221-4.(A useful overview of Vaihinger's work, with a special emphasis on his notion of 'fictions'.)

Seidel, A. (ed.) (1932) *Die Philosophie des Als Ob und das Leben (The Philosophy of As-If and Life)*, Berlin: Reuther & Reichard; repr. Aalen, Scientia,1986.(A large collection of laudatory articles on Vaihinger's philosophy of 'as-if'.)

Valla, Lorenzo (1407-57)

Unlike most Renaissance humanists, Valla took a special interest in philosophy. However, his most influential writing was a work of grammar, Elegantiae Linguae Latinae (The Fine Points of the Latin Language); he had no comprehensive philosophy, nor did he write mainly on philosophy. Valla considered himself to be a revolutionary overturning received opinions, bragging that through his works he was 'overturning all the wisdom of the ancients'. His preference for Quintilian over Cicero and criticism of classical authors shocked older humanists, and religious authorities were upset by his views on the Trinity and on papal authority, but Valla never sought the overthrow of classical studies - or the papacy for that matter. He sought rather to destroy the Aristotelianism then reigning in the universities. In De Vero Falsoque Bono (On the True and False Good) (1431), he argued for the superiority of Epicureanism over Stoic and Aristotelian ethics. In De Libero Arbitrio (On Free Will) (1439), he corrected Boethius' treatment of free will and predestination. In the Dialectica (1438-9) he set out to reform logic and philosophy because he believed Aristotle had corrupted them. Asserting that Aristotle had falsified thought because he had falsified language, Valla was determined to show how logic rightly conformed to the linguistic usage of the classical literary authors; essentially Valla had aggressively revived the ancient competition between the rhetorical and philosophical traditions. The first great humanist, Francesco Petrarca (better known in English as Petrarch), had attempted something similar in the fourteenth century, but Valla's knowledge of philosophy was greater than Petrarch's and he had access to more sources. Furthermore, Valla knew Greek and could read texts which the medieval Aristotelians knew only in Latin translation.

1 Life and works

Valla was born in Rome. His family came from Piacenza, but had firmly entrenched itself in the papal bureaucracy. Valla's father, grandfather, uncles and brother-in-law all worked in the papal curia as legal and secretarial officials. His father Luca died when Valla was still a boy. His mother Caterina (Scribani) saw eight of her eleven children die young, and buried her two adult sons as well: Paolo (a friar) died in 1439, and Lorenzo on 1 August 1457. Valla never attended university, but he was fortunate enough to learn Greek. He never married but around 1450, in the space of two years, Valla had three children by a servant girl in Rome. He claimed to have done so to prove to his relatives that he was celibate by choice and not incapacity. He gave the children to his sister whose marriage had proven infertile.

In 1430, after failing to gain the position as papal secretary vacated by the death of his uncle, Valla left Rome for northern Italy. He spent time in Venice and Piacenza before becoming professor of rhetoric at the University of Pavia in early 1431. He fled Pavia, however, in the spring of 1433 because of the indignation of the law faculty at his letter attacking the ignorance of the medieval jurist Bartolo of Sassoferrato. He spent the next few years travelling about northern Italy supporting himself as a private teacher. Finally, in 1435, he joined the court of King Alfonso of Aragon, who was engaged in the conquest of the kingdom of Naples. Valla had the title of secretary, but functioned more like a scholar-in-residence. The king was generous, granting Valla temporal benefices and protection during his controversies. Valla's years in southern Italy were the most fruitful of his scholarly career. Once established in Naples (conquered by Alfonso in 1442), Valla was even able to take up teaching rhetoric again. But Valla yearned for Rome. In late 1447, after receiving favourable signs from the papal court, he returned to his native city. Pope Nicholas V (1447-55) made him a papal scribe (in Valla's case this was a sinecure) and appointed him professor of rhetoric at the University of Rome. He also generously rewarded Valla for his scholarly labours, giving him, for instance, 500 gold ducats for translating Thucydides. The next pope, Calixtus III (1455-8), made Valla a papal secretary (another sinecure) and showered him with ecclesiastical benefices, including (although Valla was not a priest) the position of canon in St John Lateran, the cathedral of Rome.

Valla provoked his first controversy in 1428 with a now lost comparison of Quintilian and Cicero, scandalizing fellow humanists by his preference for Quintilian. Valla found in Quintilian's *Training in Oratory* a lifelong guide not only to rhetoric, but also to philosophy, logic and language. Later in life, Valla claimed that he had virtually memorized *Training in Oratory*. Indeed, an awareness of Quintilian's influence is essential for a correct understanding of Valla's work.

Valla completed his first major work, the dialogue *De Voluptate (On Pleasure)*, in the early 1430s; in later

redactions he changed its title to *De Vero Falsoque Bono (On the True and False Good)*, but did not alter its Epicurean slant. The first versions of his other major works were finished while Valla was in the employ of King Alfonso.

To his most ambitious work, Valla gave the shorthand title *Dialectica*. Its full title changed with every redaction: first it was the *Re-tilling…* (*Repastinatio…*), then the *Refabrication…* (*Reconcinnatio…*) and finally the *Restructuring* (*Retractatio*) *of All Dialectic with the Foundations of the Whole of Philosophy*. For all of Valla's grand hopes, the *Dialectica* failed to galvanize Renaissance humanism. It experienced a moderate manuscript circulation and was printed separately only four times before entering Valla's *Opera Omnia* of 1540.

Elegantiae Linguae Latinae (The Fine Points of the Latin Language), on the other hand, was a huge success. With nearly seventy known manuscripts and more than 150 printed editions before 1600, the *Elegantiae* made Valla a major authority on the language. Treating a wide range of grammatical and lexicographical points, Valla took as his standard not the rules of the classical grammarians, but the actual usage of those whom he viewed as the very best practitioners of Latin: the *oratores* (specifically Cicero and Quintilian). With extraordinary erudition, Valla presumed not merely to teach his contemporaries correct classical Latin usage, but also to criticize the Latin of classical writers and grammarians. He conspicuously condemned the late ancient author Boethius as 'a Roman who did not know how to speak Roman'.

In 1440 Valla published his famous *Oratio (Oration)* on the Donation of Constantine, the document with which Emperor Constantine purportedly deeded the western part of the empire to the papacy. Numerous medieval jurists and political thinkers had denied its validity, and some had even denied its authenticity, but no one before Valla had effectively demonstrated (on the basis of linguistic, nuvmismatic and historical evidence) that this was an early medieval forgery.

Valla's *Oratio (Oration)* long remained subject to controversy. Of greater long-term significance, however, was another work Valla completed in these years: the *Collatio Novi Testamenti (Collation of the New Testament)* (called in its last redaction *Annotationes in Novum Testamentum*), where he compared the Greek text of the New Testament with the Vulgate Latin translation. Valla had no sense of the difficulties caused by the Greek tradition and focused narrowly on the peccadillos of the Latin translation instead. None the less, he showed how humanists could use their philological and historical expertise to compete as scriptural theologians with the dogmatic theologians of the medieval tradition. Valla's work had little impact in Italy, but once 'discovered' by Erasmus, the *Annotationes* became an inspiration for humanist scripture scholarship in the Reformation era.

While in southern Italy, Valla also produced works typical of Renaissance humanism. He wrote a history of the deeds of King Alfonso's father, dashed off several philological opuscula, prepared a lost commentary on the pseudo-Ciceronian *Rhetorica ad Herennium*, translated Greek classics (Aesop, Homer, Xenophon, St Basil the Great and Demosthenes), wrote invectives against fellow humanists and composed dialogues. In the first, *De Libero Arbitrio (On Free Will)*, Valla asked how free will related to divine foreknowledge, condemning Boethius for mishandling the issue out of an excessive affection for philosophy. In the second, *De professione religiosorum (On the Professing of Vows by those in Religious Orders)*, he challenged the notion that the members of the religious orders are more religious or meritorious than lay people. *De professione religiosorum* had a minimal circulation, but its boldness had a price. In 1444 Valla was called before the Inquisition. Only royal protection saved him. Subsequently, he wrote a *Defensio quaestionum in philosophia (Defence of Questions in Philosophy)* rebutting the charges of heresy and in 1444 an *Apologia* to Pope Eugenius IV.

The *Apologia* was part of Valla's campaign to return to Rome, which succeeded in late 1447. In the last decade of his life Valla continued to revise his major works, but embarked on no new large project apart from the Latin translations of Herodotus and Thucydides undertaken at the behest of Pope Nicholas V. He also entered into a lengthy literary war with Poggio Bracciolini, a leading humanist in the papal curia. Finally, in these last years, he delivered an encomium on the mystery of the Eucharist and another on Thomas Aquinas. Valla praised Aquinas as the best of the medieval theologians, although he considered Aquinas inferior to the church fathers because they did not mix theology with 'petty dialectical syllogising and metaphysical musing' nor 'make philosophy the basis of their arguments' (*Opera Omnia*, vol. 2: 350).

2 Ethics

The notion that Valla was a neo-pagan libertine is not borne out by his life and writings. The basis of that myth is the Epicureanism of *De Vero Falsoque Bono*; ironically the dialogue was principally an exercise in Christian apologetics. Valla was neither the first nor the last Renaissance intellectual to appreciate Epicureanism. In a 'moral letter' of about 1429, the minor humanist Cosma Raimondi argued that Epicureanism answered the needs of human beings as composites of body and soul much better than Stoicism. In his youth the fifteenth-century Platonist Marsilio Ficino was also much taken with Epicureanism, and in the sixteenth century Thomas More had his Utopians base their lives on the principle that pleasure was the highest good. But no one in the Renaissance argued so extensively and forcefully for Epicurean ethics as Valla (see Epicureanism §§10, 11; Stoicism §16).

In denying that happiness resided in virtue, Valla took the same position towards Stoic ethics as Thomas Aquinas (§13), and observed at the start of the that to argue otherwise would render the Incarnation otiose. By refuting the Stoics, Valla was affirming the insufficiency of human beings to achieve happiness outside the Christian dispensation, but Valla carried the argument much further than was needed to refute Stoicism. This defence of Christianity was not an end in itself. It was rather the means by which Valla was able to attack a highly respected philosophy and specifically a highly respected Christian philosopher, Boethius, the sixth-century Roman author whom Valla seems to have considered the root of medieval philosophical evil. At the end of the *De vero bono* it becomes clear that Valla's aim was to destroy Boethius' assertion in Book 4 of *The Consolation of Philosophy* that virtue was the true path to happiness.

Similarly, Valla's appropriation of Epicureanism can be seen as another means by which to discredit philosophy. Although Valla has been criticized for presenting an inaccurate picture of Epicureanism and for relying only on Latin sources, his intention was not to explain Epicureanism, but simply to use it in his campaign against philosophers. Philosophy, as he says early in the *De vero bono*, is merely a foot-soldier in the army of Queen Oratory. This is not to say that Valla failed to value Epicureanism or Christianity, but rather that neither the exposition of Epicureanism nor the defence of Christianity were his ultimate purpose.

Valla divided the *De vero bono* into three books. In the first, he had the Stoic speaker complain that all nature conspires against virtue. Animate beings spontaneously, that is to say naturally, seek pleasure and flee the pain imposed by virtue, thus refuting by their actions the famous Stoic formula that virtue is living according to nature. Taking his cue from pseudo-Quintilian (*Major Declamations* 2.1), Valla further argued that virtue actually makes us miserable, while pleasure makes us happy. Opponents of Epicureanism had traditionally charged Epicureans with having no reason for rejecting heinous behaviour if it was pleasurable (see Cicero *On the Ends of Good and Evil* 2.27). Valla agreed: at the end of Book 1, Valla had his Epicurean speaker approve of adultery, advocate free love (what Valla called 'Platonic love'), condemn virginity and present for imitation the shameless behaviour of the Olympian gods.

Book 2 presented the other side of the argument. In it, Valla proved that the moral goodness propounded by the philosophers (*honestas philosophorum*) was fraudulent. Even in their most apparently selfless actions people always act out of self-interest. Virtue consists precisely in actions which bring about this selfish advantage (*utilitas*), and actions that are not self-serving are foolish. The moral goodness of the Stoics is nothing more than empty verbiage. At the end of Book 2, the Epicurean speaker denies the immortality of the soul and reaffirms worldly pleasure as the only valid good.

Valla began Book 3 with an attack on the Aristotelian notion of virtue as the mean between two vicious extremes, condemning the artificiality of this scheme. He denied, for example, that liberality was really the mean between avarice and miserliness because the opposite of liberality is in fact a virtue, thriftiness. Earlier, in Book 2, he had the Epicurean speaker refute Aristotle's assertion that the contemplative life was the highest and happiest form of existence. In the *Dialectica* (*Repastinatio Dialectica* 75-7), Valla further attacked Aristotelian ethics by denying that prudence was a virtue, explaining that virtue resided exclusively in the will while prudence resided in the mind (see Aristotle §22-5).

Thus Valla discredited not only Stoicism and pagan Epicureanism, but also Aristotelianism. Having proved the unacceptability of all forms of philosophic ethics, he could 'condemn and damn philosophy' (*De vero bono* 271 §7) and explain that the highest good is found not in philosophy but in religion, which teaches that resistance to vice in this life means torment and death for the sake of eternal happiness in heaven, where we will enjoy the vision of God. In short, having exposed the inanity of the philosophers, Valla changed the terms of the dialogue

and moved from philosophy to theology. In the last part of the dialogue he launched into an exposition of Christian teachings on heaven and hell. In the process, he transformed Christian ethics into a form of Epicureanism, equating the delight of heaven with pleasure, the vision of God with the love of God, and love itself with pleasure (*amatio ipsa delectatio est*). That left the question of why we love God. Christianity traditionally answered that we loved God *propter se*, because of what he is. But if that were so, then the end of Christian ethics would not be pleasure, but God. Consequently, in both the *De vero bono* and the *Dialectica*, Valla insisted that we love God not as an end, but as an efficient cause, as the producer of the pleasure of heaven. Valla's Christian Epicureanism thereby ends up instrumentalizing God.

3 Free will and determinism

In *De Libero Arbitrio*, Valla once again condemned philosophy for corrupting Christian theology and attacked Boethius for leading the way. In the *Consolation of Philosophy*, Boethius had reconciled human free will and divine foreknowledge by arguing that God knows in an instantaneous 'now' what we experience as past, present and future (see Boethius §5). Valla found this explanation inadequate: his own solution was to compare God's foreknowledge to an oracle who had the power to see the future without causing it. At best, this explanation is the same as Boethius'; at worst, it is tautological.

Valla's target, however, was not really Boethius' explanation of divine foreknowledge, but his confidence in having proven human free will. While denying that God has deprived us of liberty and therefore responsibility, Valla did not know how anyone could reconcile free will with the omnipotent will of God. Citing biblical passages which suggest predestination, Valla condemned philosophers who presumed to understand 'the power and will of God'. Instead, he argued that 'we stand by faith, not the probability of reasons' (*De Libero Arbitrio* 180). Thus, when he refuted astrology in the *Dialectica*, he based his argument not on human liberty, but on divine omnipotence: God's will and not the heavens controlled human destiny. He was therefore consistent in the *Dialectica* in insisting that nothing in nature happened by chance. On the subject of divine power, he rebuked Aristotle for denying that God can make not to have happened what has already happened. The past is not immutable because that would be a limitation on divine power. Without feeling any need to develop the implications of his positions, Valla denied contingency, held that God predestined everything and insisted that there was no conceivable limit to God's power (see Free will).

4 Logic

The *Dialectica* was Valla's philosophical *summa*. Into it he poured his thought on logic, philosophy, and theology. It has two main themes. One is that logic is 'a matter indeed short and easy', no more than a subsection of the far more comprehensive discipline of rhetoric. The second and predominant one was that Aristotle and his followers had erred because they had departed from common linguistic usage (*consuetudo loquendi*). By common linguistic usage Valla did not mean what has been called ordinary language: he scorned the 'stupid' popular manner of speaking when it suited him to do so (*Repastinatio Dialectica* 55), and he had little use for the vernacular. Rather Valla meant classical literary Latin. He never stopped to give a theoretical explanation of why linguistic usage is the touchstone of truth.

Valla conceived of logic in terms of the corpus of logical texts used in the twelfth century when medieval Aristotelianism first appeared (Porphyry's *Isagōgē* and Aristotle's *Categories*, *De Interpretatione*, *Prior Analytics*, *Posterior Analytics*, *Topics* and *On Sophistical Refutations*), not in terms of contemporary logic which had moved significantly beyond Aristotle in many areas. Although Valla divided the *Dialectica* into three books, he gave half of it over to the discussion of the categories in Book 1. He did not challenge the value of the categories; his purpose was simply to reduce their number. He kept substance, the first of Aristotle's ten categories, and reduced the remaining nine (the accidental categories) to just two, quality and action, expending a great deal of effort to show how the remaining seven reflected one of these other two.

Valla also reduced six transcendental terms (being, thing, something, one, true and good) to one term, 'thing' (*res*). The transcendentals had emerged among medieval Aristotelians as a set of terms which transcended the limitations of the categories and were convertible with each other. Their number was not universally agreed upon (John Duns Scotus, for instance, recognized more than the six transcendentals listed by Thomas Aquinas and treated by Valla) and they played a minimal role in logic, but they were important in metaphysics. Valla probably seized upon the

transcendentals because the first of them, being (*ens*), provided him with a starting point for his destruction of medieval Aristotelianism. He denied that *ens* could be a transcendental since in classical Latin it was exclusively a participle, not a noun, and therefore resolvable into 'that thing (*res*) which is'. Hence *res*, and not the illegitimate noun *ens*, is the universal term encompassing all terms and all categories.

Valla also denied the validity of a whole group of abstract terms of medieval philosophy formed by the addition of the suffix *itas* to a noun or pronoun, such as *quiditas* (whatness) and *haecceitas* (thisness). These terms, Valla explained, were illegitimate because in Latin only adjectives can properly accept *itas*, an example being *bonitas* (goodness) which is formed from *bonus* (good). Since Valla also denied that Latin permits neuter singular adjectives to constitute entities in themselves (that is, *bonum* alone means not 'the good' but 'good thing'), he has sometimes been portrayed as a radical nominalist. In fact, he quite happily accepted the legitimacy of abstractions and universals if they were properly formed according to the rules of Latin. He simply ignored the metaphysical implications of his positions.

Similarly, he ignored the epistemological implications of his view of truth. He explicitly accepted the traditional correspondence theory of truth, that truth was the correspondence between what one asserts about a thing and what the thing itself is. Yet, in also making God the fount (*fons*) of truth, attributing truth to divine illumination, and asserting without further elaboration that falsity consists in the obstruction of this fount (*Repastinatio Dialectica* 20), Valla left the way open for a different approach to truth based on the absolute power of God. Consistent with his view that God can make not happen what has happened, he seems also to have held that human capacity to attain truth is circumscribed by the fact that God can change truth at any time and that the only people who know truth are those whom God chooses to illumine.

The same indifference to developing the implications of his theories appears in Valla's discussion of language. Treating the category of quality, Valla characterized signification as a quality of human utterance (*Repastinatio Dialectica* 122-4). He further explained that significant speech is a human institution: people apply 'signs' to objects. The first to have done so was Adam, who later taught these significations to posterity. Up to this point Valla hardly differed from the medieval Aristotelians, but then he suddenly dropped this line of argument and took up the relationship between 'word' and 'thing' within his scheme of categories. Though he viewed language as the touchstone of truth, Valla never bothered to develop a linguistic theory (see Language, Renaissance philosophy of §§2, 3).

At several points in Book 1, Valla discussed physics. Having no overarching physical theory of his own (apart from reliance on the criteria of common sense and linguistic usage), he only made small alterations to Aristotelian physical theory. He had no inkling of the use of mathematics in physics and no patience with geometric descriptions of nature, mocking, for instance, the notion that the earth could be treated as a mere point in space. On occasion he appealed to his own experience to correct Aristotle; in one example he cites his observation of modern bombards. He also challenged fundamental Aristotelian tenets, such as the principle that whatever is moved is moved by another. But these criticisms do not come together as a coherent whole and at times amount to very little. For instance, in rejecting the principle that whatever is moved is moved by another Valla merely referred the reader to Macrobius.

After the categories, Valla treated propositions (Book 2) and various forms of argumentation, especially the syllogism (Book 3). Though he rejected much that Aristotle said about modal propositions and modal syllogisms and though he banned outright Aristotle's third figure of the syllogism (see Aristotle §5), Aristotelian logic survived Valla's critique essentially intact. On the subject of topics, examples and enthymemes, Valla incorporated Quintilian's discussions of them from the *Training in Oratory (Institutio Oratoria)*. In sophistics, Valla was much concerned, using literary examples, to defeat the argument from *sorites* (that is, the argument from accumulation: with the subtraction of which grain of sand does a pile cease to be a pile?) and dilemma (see Vagueness §2). To defeat the latter he constructed a fictive oration in which Protagoras answers a dilemma put to him. This section alone disproves the notion that Valla was a sceptic.

The centrepiece of these pages revolves about a point of language. In Greek the marker for an indefinite proposition is *tis*, which can mean the indefinite 'some', but also the particular 'a certain'. Since Boethius, the standard translation in logic of *tis* had been *quidam*, which, Valla pointed out, means 'a certain' and not 'some'. Valla could therefore severely criticize the medieval Latin analysis of propositions and syllogisms wherever

quidam was used as an indefinite. Furthermore, he showed that whereas Greek was limited to *tis*, Latin had many words with nuanced differences to express 'some' and 'a certain', thus proving not only the superiority of Latin (a cause close to Valla's heart), but also that correct logic rested on the correct use of language.

The *Dialectica* failed to overthrow Aristotelian logic, let alone Aristotelian philosophy, and it never really confronted contemporary logic. The vast majority of humanist logicians declined to follow Valla's prescriptions. The *Dialectica* did, however, forcibly argue for the use of a natural language in logic and in this respect Valla was clearly in the vanguard of a broad drive by humanists to simplify logic and make it more amenable to rhetoric and literary practice.

See also: Aristotelianism, Renaissance; Humanism, Renaissance; Language, Renaissance philosophy of; Logic, Renaissance

JOHN MONFASANI

List of works

Valla, L. (1407-57) *Opera Omnia (Complete Works)*, ed. E. Garin, Turin: Bottega d'Erasmo, 1962, 2 vols. (Volume 1 is a photographic reprint of the 1540 *Opera Omnia*, printed by Henricus Petri in Basle; volume 2 collects in photographic reprint opuscula published separately over four centuries.)

Valla, L. (1431) *De Vero Falsoque Bono (On the True and False Good)*, ed. M. Lorch, Bari: Adriatica, 1970; English text and Latin translation in M. Lorch and A.K. Hieatt (trans and eds) *On Pleasure. De Voluptate*, New York: Abaris, 1977.(Valla's Epicurean dialogue was originally called *De Voluptate*; later redactions were retitled *De Vero Falsoque Bono*.)

Valla, L. (1438-9) *Repastinatio Dialecticae et Philosophiae (Re-tilling of Dialectic and Philosophy)*, ed. G. Zippel, Padua: Antenore, 1982.(Edition of all three redactions of Valla's *Dialectica*.)

Valla, L. (1439) *De Libero Arbitrio (On Free Will)*, trans. C. Trinkaus, in E. Cassirer *et al.*, *The Renaissance Philosophy of Man*, Chicago, IL, and London: University of Chicago Press, 1948, 147-82.

Valla, L. (1440) *Oratio de falso credita et ementita Constantini Donatione (Oration on the Falsely Believed and Fabricated Donation of Constantine)*, ed. W. Setz, Munich: Monumenta Germaniae Historica, 1986.(Valla's exposure of the Donation of Constantine as a medieval forgery.)

Valla, L. (1440-1) *De professione religiosorum (On the Professing of Vows by those in Religious Orders)*, ed. M. Cortesi, Padua: Antenore, 1986.

Valla, L. (1443) *Collatio Novi Testamenti (Collation of the New Testament)*, ed. A. Perosa, Florence: Sansoni, 1970.(Comparison of the Greek text of the New Testament with the Vulgate Latin translation. Called *Annotationes in Novum Testamentum* in its last redaction.)

Valla, L. (1444a) *Elegantiae Linguae Latinae (The Fine Points of the Latin Language)*, in E. Garin (ed.) *Opera Omnia*, Turin: Bottega d'Erasmo, 1962, vol. 1, 1-235. (The date of the first redaction is 1444.)

Valla, L. (1444b) *Defensio quaestionum in philosophia (Defence of Questions in Philosophy)*, trans. G. Zippel, 'L'autodifesa di Lorenzo Valla per il processo dell'Inquisizione napoletana (1444)', *Italia medioevale e umanistica* 13 (1970): 59-94.

References and further reading

Besomi, O. and Regoliosi, M. (eds) (1986) *Lorenzo Valla e l'Umanesimo italiano (Lorenzo Valla and Italian Humanism)*, Padua: Antenore.(Acts of a conference of leading Valla scholars.)

Camporeale, S. (1972) *Lorenzo Valla: Umanesimo e teologia (Humanism and Theology)*, Florence: Istituto Nazionale di Studi sul Rinascimento.(An enthusiastic study especially emphasizing the influence of Quintilian.)

Cicero, M.T. (mid-45 BC) *De finibus bonorum et malorum (On the Ends of Good and Evil)*, trans. H. Rackham, Loeb Classical Library, Cambridge, MA: Harvard University Press 1914.(One of Cicero's major works on ethics; presents Epicurean, Stoic and Peripatetic arguments. Referred to in §2.)

Fois, M. (1969) *Il pensiero cristiano di Lorenzo Valla nel quadro storico-culturale del suo ambiente*, Rome: Università Gregoriana.(Sound, comprehensive and containing new factual information.)

Mack, P. (1993) *Renaissance Argument: Valla and Agricola in the Traditions of Rhetoric and Dialectic*, Leiden: Brill.(The best and most thorough study to date of Valla's logic.)

Mancini, G. (1891) *Vita di Lorenzo Valla (The Life of Lorenzo Valla)*, Florence: Sansoni.(The standard biography.)

Monfasani, J. (1994) *Language and Learning in Renaissance Italy: Selected Articles*, London: Variorum.(Articles

on Valla’s logic and philosophical thought.)
Pseudo-Quintilian (late 4th century) *Declamationes Maiores*, trans. L.A. Sussman, *The Major Declamations Ascribed to Quintilian: A Translation*, Studien zur Klassischen Philologie 27, New York: Peter Lang, 1987. (Known to have existed by the late 4th century, the attribution of these rhetorical works to Quintilian is now subject to serious doubt. Referred to in §2.)
Sabbadini, R. (1891) *Cronologia documentata della vita del Panormita e del Valla*, in L. Barozzi and R. Sabbadini, *Studi sul Panormita e sul Valla*, Florence: Istituto di studi superiori practici e di perfezionamento. (Important biographical compendium. Available in the 1962 edition of Valla’s *Opera Omnia*, vol. 2.)
Trinkaus, C. (1970) *In Our Image and Likeness*, Chicago, IL, and London: University of Chicago Press, 2 vols. (Extensive discussion of Valla’s ethics and logic.)

Vallabhācārya (1479-1531)

A pivotal figure in the history of Indian philosophy and religion, Vallabhācārya was the last of the classical Vedānta philosophers, as well as the originator of a religious community which called for the worship of Kṛṣṇa through acts of devotion, in return for grace and deliverance from rebirth. He proposed a modification to Śaōkara's philosophy of nondualism, claiming his 'pure nondualism' better explained the relationship between the Supreme Being and the soul. For the laity, he offered a practical religious regimen called the 'path of fulfilment', through which the devotee is initiated into an individual relationship with Kṛṣṇa before proceeding to fulfil the relationship through specific personal acts of devotional worship.

Vallabhācārya (Vallabha Bhaṭṭa), born into a Tailangana Bhāradvāja Brahman family belonging to the Taittirīya branch of the Yajurveda, was an important religious leader of the later medieval period of North India. He promulgated a religious association (*sampradāya*) called *Puṣṭimārga* ('path of fulfilment'), based on the popular religious notion of selfless devotion (*bhakti*) to God, which has influenced millions and persists to the present day, especially among the mercantile communities of Gujarat, Rajasthan, Maharashtra, Harayana, western Uttar Pradesh and western Madhya Pradesh. Vallabha's life and work must be understood in the context of his times, a context influenced by popular forms of ecstatic religion and the confrontation of Islamic and Hindu thought, as well as threatened by an unstable political atmosphere.

In a sense, Vallabha was a Brahmanic or Sanskritic reaction to the challenge of being Hindu in the early sixteenth century. His birth was the culmination of a prediction that an incarnation of Viṣṇu would appear in his family after it had performed 100 soma sacrifices. His mother gave birth to a stillborn child in a forest while fleeing their home in Benares which was threatened by Muslim armies from Delhi and Jaunpur. Informed by Kṛṣṇa that the child was alive, the parents returned to find Vallabha surrounded and protected by fire. The phenomenon of fire dominates the themes of Vallabha's life. The circumstances of his birth are an obvious connection, but in addition his family were experts in the *Yajurveda*, a book of Vedic rituals and sacrifices; the basis for all Vedic rituals is fire, of which Agni is the god. For his followers, Vallabha speaks as the fire-mouth (*jvālamukha*) of Kṛṣṇa, cleansing the world of impurities. Vallabha called himself 'Vaiṣvānara' ('universal divine fire'), incarnated on earth to be a source of knowledge of the Supreme Being. Upon his death, Vallabha walked into the Ganges at Benares and disappeared in a ball of flame.

Sampradāya sources claim that an image of Kṛṣṇa, which, in 1410, had pushed its hand and arm through the ground on Mount Govardhan in Braj (the region in which Kṛṣṇa was born), suddenly erupted further at the moment of Vallabha's birth, exposing the face and uplifted arm of Kṛṣṇa as Lord Govardhananātha, the form of Kṛṣṇa which protected the people of Braj from the wrath of the rain-god Indra. It is believed that this image was a *svarūpa*, or self-manifestation, of Kṛṣṇa, and that Vallabha was the 'mouth' of the *svarūpa*. Later, Vallabha was instructed by Kṛṣṇa in a dream to proceed to Mathura, identify the half-hidden image and see to its worship. Shortly thereafter, while staying at the village of Gokul, the historic birthplace of Kṛṣṇa, Vallabha was given the *brahmasambandha* (connection to the Supreme Being) by Kṛṣṇa in a dream. The *brahmasambandha mantra*, *śrīkṛṣṇaḥśaraṇam mama* ('Radiant Kṛṣṇa is my refuge'), was an act of initiation, in which Vallabha unconditionally placed himself in the grace of Kṛṣṇa while assuming responsibility for the care and worship of the *svarūpa* of Govardhananātha, later known as Śrīnāthajī.

Vallabha believed that he lived in a time when it had become impossible for the individual to achieve spiritual liberation through study or the use of ascetic practices. Five impurities (*doṣas*) inevitably prevented the soul from realizing its connection with the Supreme Being: innate impurities; impurities resulting from one's habitation; those resulting from one's time of life; those resulting from all kinds of association; and those resulting from direct contact. The notion of impurity, and its absence being an a priori condition for spiritual progress, is an ancient principle of Brahmanism; thus, it is not surprising that Vallabha placed this notion at the centre of his liberation philosophy. Initiation into the *sampradāya* through the *brahmasambandha mantra* results in the immediate cleansing of the five *doṣas* from the soul (*jīva*).

Vallabha belongs to the Vedānta school of Hindu philosophers, which includes Śaōkara, Rāmānuja, Madhva and Nimbārka. Vallabha argued that Śaōkara's form of Advaita (monism) erroneously relied upon the notion of

māyā (illusion) to explain the apparent division or difference between the Supreme Being and the soul. He insisted that no amount of self-study or ascetic austerity could clarify the relative positions of the Supreme Being and the soul. For Vallabha, the soul's ability to discern the relationship correctly is occluded by the presence of impurities accumulated through worldly associations. Only initiation by Vallabha or one of his direct descendants, through taking the *brahmasambandha* and dedicating one's body, mind and wealth to the service (*sevā*) of Kṛṣṇa, can put one in a position to receive the grace of Kṛṣṇa. Such grace, however, is not guaranteed, as it is dispensed by Kṛṣṇa, who may bestow it at any time on whomsoever he pleases. The initiated are given a particular form of Kṛṣṇa, usually drawn from his childhood, to worship as a personal *svarūpa*. Observing this 'path of fulfilment', the initiated create the conditions in which Kṛṣṇa is likely to provide grace.

As a teacher (*ācārya*), Vallabha departed from tradition by marrying and having children. He did so in response to a command from Kṛṣṇa, who insisted that only his direct descendants would be qualified to transmit the *brahmasambandha mantra*. Vallabha's second son, Viṭṭhalanātha, greatly expanded the membership of the *sampradāya*, and elaborated the service (*sevā*) stipulated by Vallabha. He in turn had seven sons, all of whom were given *svarūpas* of Kṛṣṇa. A month before his death, Vallabha took initiation into the fourth stage of life, *sannyāsa*, completing his duty as a Brahman; the lateness of this act is interpreted as a signal to his descendants not to follow the ascetic imperative.

See also: Vedānta

RICHARD J. COHEN

List of works

Vallabhācārya (*c*.1479-1531) *Śrīmadbrahmasūtrānubhāṣyam*, ed. H. Shastri, Bombay: Trustees of Sheth Narayanadasa and Sheth Jethananda Asanamala Trust, 1942.(Commentary on the basic aphorisms of the Vedānta schools of philosophy, in which Vallabhācārya defends his interpretation of Vedānta.)

Vallabhācārya (*c*.1479-1531) *Śrīsubodhinī(Giving a Good Explanation)*, ed. N. Kishora Sharma, Nathadvara: Vidya Vibhaga, 1928.(A commentary on parts of the *Bhāgavata Purāṇa*, a scripture that Vallabhācārya considered of fundamental importance to a correct understanding of Vedānta.)

Vallabhācārya (*c*.1479-1531) *Ṣoḍaśhagranthah: SakārikāRasapañcādhyāyīVeṇugopikayugala-brahmaragitasubodhinīca*, ed. M.G. Shastri, Bombay: Mohanlalbhai Govardhanadas, 1933.(A collection of sixteen separate Sanskrit texts, detailing the main points of Vallabhācārya's philosophy.)

Vallabhācārya (*c*.1479-1531) *The Tattvārtha-Dīpa-Nibandha with Prakāsha (Treatise that is a Lamp onto Reality)*, ed. H. Onkarji Shastri, Bombay: Trustees of Sheth Narayandas, 1943, 2 vols.(Divided into three parts, this text discusses the *Bhagavad Gītā*, the major Hindu philosophical systems and the *Bhāgavata Purāṇa*.)

Vallabhācārya (*c*.1479-1531) *The Tattvadīpanibandha with Prakāsha*, ed. S. Shastri, Bombay: Shridhara Shivalalji, 1905-8, 2 vols.(An older and less easily obtainable edition of the work cited above.)

References and further reading

Barz, R. (1976) *The Bhakti Sect of Vallabhācārya*, Faridabad: Thomson Press (India).(A trustworthy historical treatment of Vallabhācārya's life and work.)

Marfatia, M.I. (1967) *The Philosophy of Vallabhācārya*, Delhi: Munshiram Manoharlal.(Offers an accessible overview of Vallabhācārya's philosophical system, with detailed summaries of his key works. Abundance of Sanskrit terms may be daunting to beginners.)

Parekh, Bhai Manilal C. (1943) *Sri Vallabhācārya: Life, Teachings, and Movement*, Rajkot: Sri Bhāgavata Dharma Mission.(An accessible biographical sketch.)

Vaidya, Chimanlal M. (1959) *Shri Vallabhācārya and His Teachings*, Kapadwanj: Shri Shuddhadwaita Samsada.(A discussion of Vallabhācārya's work and his teachings.)

Value judgments in social science

Leading theorists in the social sciences have insisted that value judgments should be strictly separated from scientific judgments, which should be value-free. Yet these same thinkers recognize that social scientists are often committed to values in carrying out their work and may be motivated by moral goals of removing or remedying social conditions. From this perspective, scientific conclusions (one sort of fact) and moral commitments (one sort of value) are intertwined in scientific practices, and the question arises whether a social scientist qua *scientist makes value judgments or only makes such judgments in a nonscientific capacity. Related questions concern the role played by moral, social, and political values in the pursuit of scientific knowledge and the impact of these values on scientific theories and methods.*

1 The values question in classical social science methodology

Émile Durkheim (§2) (founder of academic sociology in France and a pioneer of the discipline) believed that sociology, as a science, must not be contaminated by value judgments (Durkheim 1895). 'Social facts' alone have a legitimate role in sociological explanation, and the scientific explanation of a social phenomenon seeks only its efficient causes. None the less, Durkheim maintained that a scientific sociology could help determine the appropriateness of moral values and practices for societies by showing which values best promote human happiness and social solidarity. He thought a science that examines society should regard its work as incomplete unless it established a 'science' of morality (comprised of what is 'generally accepted' in a culture), and he therefore moved freely from social facts to embedded moral values and from the scientific study of morality to moral judgments. In this way, a social science delivered practical results.

Max Weber (§2) resisted such linking of science and valuation, arguing that there must be an 'unconditional separation' between empirical facts and the evaluation of those facts (Weber 1949). He viewed facts and values as 'entirely heterogeneous'. He maintained that the social scientist must shun evaluative commitments that might bias scientific work. None the less, Weber agreed with Durkheim that moral and other values are of the highest importance in social life, and Weber recognized that facts and norms are intertwined in several ways in the 'praxis' of social science. For example, scientists choose research projects based on cultural values, and in carrying out science they dedicate themselves to values of honesty and accuracy.

Weber's recommendation of a value-free social science was immensely influential, eventually becoming the orthodox position in the social sciences and philosophy. His recommendation also promoted the view that the judgments of science are demonstrable, that value judgments are not, and that scientific judgments therefore have a superior epistemological status. However, these ideas, which grew out of early sociology, enjoyed less currency in other quarters. For example, many economists and political scientists doubted from the start that a value-free model suits their 'value-impregnated' fields (see Economics, philosophy of §2). Many questions persist about whether the value-free ideal is possible or desirable, and about whether social scientists can or should avoid value judgments in their work.

2 The model of a value-free social science

Weber's distinction between praxis and methodological principle illuminates this dispute: the critics of value-free science tend to point to values present in social scientists' practices and role in society, whereas proponents argue primarily from the methodological principles and ideals of science. Weber himself was not concerned about the evaluative criteria underlying and motivating scientific practice, but with the justification of scientific conclusions. He emphasized the logical independence of facts and values and the validity of scientific methods.

The central premise in the model of a value-free social science is that the truth or falsity of scientific hypotheses and the scientific methods used to determine truth or falsity are independent of value judgments. The model presupposes the defensibility of the distinction between facts and values, but is not dependent upon any particular interpretation of the nature of a fact or the nature of a value. It also does not depend on any particular view about whether value judgments are true or justifiable, though some proponents have argued that value judgments are neither rationally justifiable nor true. Nor need the model incorporate any form of scepticism about either facts or values.

The central premises commonly found in attacks on this value-free model are the following:

1 Science presupposes values such as truthfulness, knowledge, explanatory power, and criteria that distinguish good from bad theories and explanations.
2 In selecting a problem for scientific study, values drive the process of selection.
3 All scientists have evaluative viewpoints that inevitably affect how they interpret and report their findings, including how they construct the evidence and the importance they give to some evidence to the exclusion of other evidence.
4 Whenever scientists accept hypotheses, they make a value judgment that sufficient evidence supports the hypothesis; this judgment does not follow exclusively from the scientific method, because no hypothesis is ever fully verified by that method.
5 A scientist who accepts funding from institutions whose values control the funding tacitly accepts the values and priorities of those institutions.

Value-free conceptions of scientific method and explanation must resolve these problems. Proponents typically argue that the values mentioned are pre-scientific considerations external to the science itself; they believe that observation, explanation, and theory construction are not affected by such contextual matters. However, the intermixture of facts and values in the actual conduct of scientific research makes it difficult to establish clean lines of separation, if indeed they are separate.

3 Problems of motivation and selection

One challenge to a value-free social science comes from the moral and political motivations of scientists, for example, their desire to see protection increased for vulnerable members of society or their ambition to increase the reputation of their institution. These moral and social commitments influence the selection of research topics as well as the way the results of research are interpreted, reported, and applied. Social scientists also report their results selectively, typically in a manner that supports their hypotheses and viewpoints. What will be undertaken in studies, how it will connect to social policy, and what can be accepted as a scientific answer all depend on an evaluative framework.

From Weber to the present, defenders of value-free science have not denied that knowledge is selective, influenced by values, and motivated by social concerns. The question, from the value-free viewpoint, is whether these background or extra-scientific features of work in science are relevant to questions of methodology and whether they infect the science. Underlying motives seem to defenders of value-freedom to have no direct bearing on the science. They maintain that when scientists who are motivated to demonstrate *x* obtain results from their studies demonstrating *not-x*, they report their results accordingly, if they are good scientists. Scientific methodology, from this perspective, is consistent with moral motives and value-driven selection that play a vital role but can be set aside in scientific practice.

The proponents of value-free science acknowledge that selectivity, like standards of accuracy, is a filtering device essential to good science. But they insist that the evaluative character of selectivity does not by itself taint the data or the reporting of data. This contention moves the issues from selectivity in general, to the question of whether selection involves some form of bias or failure to be impartial.

4 Problems of bias and partiality

Rigour in the design of studies to eliminate or minimize bias is a canonical principle. In general, a bias is a deviation of a finding from the truth. Many processes in the collection, interpretation, review, and publication of data can cause bias. However, bias is often difficult to pinpoint, and the term ‘bias’ is used in different senses in these discussions.

One sense of bias is distortion in an alleged finding due to a limitation in the design of the study. A biased sample, for example, may cause a problem in the generalizability of findings. Unless the sampling method ensures that all members of the reference population have a chance of selection in the sample, bias may occur. For example, some early surveys of ‘lesbian behaviour’ were limited to small samples of lesbians encountered in bars. This sample is biased if generalized to all lesbians; it may also convey negative impressions of this diverse community of women. Similarly, an early study of gay men used a sample entirely comprised of gay men in psychiatric therapy.

Many methodologists believe that scientists rarely, if ever, eliminate all biases from their studies. These methodologists propose not the complete elimination of biases, but a reporting of them wherever they can be ascertained (the reverse of so-called ‘reporting bias’, which involves a selective suppression of information). They insist that risks of bias can be minimized in well-designed research, but no method can eliminate bias or ensure that it will not be present. In the ideal, social science is free of bias in this sense, but in practice it almost always falls short.

A second sense of bias is maintaining a personal point of view that distorts the answers scientists give to the questions they investigate. This use of bias refers to a lapse of impartiality or partisanship. Partiality results from a value-directed departure from accuracy, objectivity, and balance, not merely an inadvertent distortion of facts or a limitation in sample selection. Bias of this sort can be a factor in research design, critical analyses, the interpretation and publication of results, and in peer review. Again, a basic methodological principle is that protection against partiality be built into research design, but should this prove impossible, one strategy is to disclose the personal values and how they might affect study results and reports.

Partiality should be distinguished from legitimate forms of opinion and appraisal. Partiality occurs only if there is a value-directed departure from accuracy and objectivity. If, for example, a social scientist fails to notice that an accidental omission has distorted a crucial part of a study, the distortion results from error, not partiality. A bias in the second sense is present only if distorted information is causally connected to the investigator’s values.

5 Problems of advocacy and strength-of-evidence judgments

Charges of bias can arise when a social scientist has adopted an advocacy role or has been employed to represent a particular institutional or special-interest point of view. For example, social scientists who have identified hazardous conditions, criminal activity, government neglect and the like, and who serve as experts on such matters sometimes become advocates who propose interventions and control measures. In practice, these recommendations often develop from patterns of empirical evidence that the advocates may or may not have generated themselves. These scientists apply causal criteria to the findings of studies and then make judgments about causes (of poverty, say) that are inextricably bonded to their recommendations.

Although appropriate to and required of social scientists in some roles, such as those of public health officials, advocacy is a threat to scientific objectivity whenever a group such as a government bureau, a labour union, industrial management, or a public interest group has sponsored the research. Social scientists who view themselves as social architects, using their own research to achieve what they consider a noble goal of eliminating all putative causes of human poverty, distress, disease and the like, are commonly viewed with scepticism by colleagues.

However, advocacy does not necessarily impugn either the underlying science or institutional and social policy recommendations based on it. Being loyal, dedicated, or even partisan is not equivalent to a loss of impartiality about scientific conclusions, although such participation does threaten impartiality. Partisan or institutionally loyal social scientists can, in principle, restrain or even completely eliminate their commitments in conducting a study and reporting its findings. Advocacy is also sometimes distanced from the conduct of research. For example, scientists may be advocates for the funding of research and for particular types of research needed to address what they consider unanswered research questions.

Related to problems of advocacy are problems of strength-of-evidence judgments that depend upon what is morally at stake. The social consequences of making a mistake in assessing evidence are very serious in some cases, minimal in others. These social consequences can and often do affect judgments about the strength of scientific evidence. Judgments that the evidence is sufficient and that a hypothesis should be accepted often depend heavily upon context consequences. For example, consider a study designed to determine whether women who receive state welfare payments can be deterred from becoming pregnant or, if they become pregnant, from giving birth to their child, by depriving them of increased financial aid assistance. The research is to evaluate a deterrent effect by examining the impact of a state-mandated, child-exclusion provision on fertility and birth rates among women. No prior evidence was sufficient to answer this question. Hypothesis acceptance in this study will almost certainly be contingent upon value judgments about what the consequences of acting on the evidence would be for both pregnant women on welfare and state officials.

The idea that degrees of scientific evidence are free of interpretation is challenged by this argument (which was first advanced by Richard Rudner 1953), because 'strength of evidence' is itself a value-impregnated notion. It seems impossible to give responsibility for judgments about strength of evidence exclusively to public officials, contractors, or employers, because it is the business only of a scientist, not someone else, to make a judgment about the strength of scientific evidence and about whether a scientific hypothesis is adequately supported.

6 Professional codes of conduct

Social science research can never be a morally neutral activity as long as it involves human research subjects. Many problems of value judgements in social science concern how these subjects may permissibly be used. The book *A Time to Speak: On Human Values and Social Research* (1968) by Herbert Kelman, combined with a number of controversial cases of abuses of subjects by social scientists, alerted scientists in the late 1960s to a variety of problems in research ethics. Careful attention was subsequently paid to the moral judgments that social scientists do and should make in carrying out their research. Problems stem from practices such as the experimental deception of subjects, the collection of data on persons in an individually identifiable form, and the use of financial incentives to obtain subjects.

Two types of value judgments merit attention in this context. The first type involves moral assessments of the protocols and the behaviour of social scientists. The second type concerns methodological evaluations of research practices (similar to those previously discussed). Methodological criteria govern how a procedure is to function in a research design in order to avoid screening out effects that could render the results ungeneralizable, or perhaps unreliable or of very limited generalizability. A procedure may satisfy criteria of methodological adequacy and be morally unjustified; or the procedure may be morally justified but methodologically inadequate. Here we encounter methodological-moral dilemmas. Sometimes methodological rigour is sacrificed to yield a moral gain, and sometimes moral rigour is sacrificed to yield a methodological gain (for example, by waiving a consent requirement for research involving deception).

Value judgments here intersect in several ways. A design adequate to achieve the objectives of a research protocol is generally considered a necessary condition of the moral justifiability of the research. Many argue both that it is irresponsible to use valuable resources to conduct methodologically flawed research and that it is morally unjustified to conduct research that presents risk to subjects if the research design is defective. Both claims presume criteria for assessing the methodological adequacy of research procedures, and in both cases the reasons for making methodological soundness a condition of moral justifiability are independent of these criteria.

The aforementioned normative rule that study bias must be avoided sometimes contingently conflicts with moral and professional obligations to protect the rights of research subjects. For example, because the validity of a study and the availability of funding to conduct it may be adversely affected by a poor response rate, potential subjects are sometimes encouraged to participate by repeated telephone calls and aggressive follow-ups. This badgering may fail to respect the decisions of subjects and violate their right to privacy. The point is that choices between alternative study populations, control groups, interviewing procedures, and approaches to minimizing the likelihood of bias will involve both ethical and methodological values.

Most professions arguably have an indigenous professional morality with precepts governing research ethics. These precepts specify responsibilities to research subjects, responsibilities to society, responsibilities to employers and funding sources, and responsibilities to professional colleagues. Particular codes developed for professional associations of social scientists are usually attempts to discover, formulate, and develop this inchoate morality rather than attempts to apply some general ethical theory. However, philosophical ethics has been explicitly used in some cases to help in developing this professional morality. Many social scientists have recognized that there are serious problems with attempts to base professional ethical standards entirely in practice standards, that often do not exist within the relevant field, group, or profession. If practice standards are deficient, ethical theory may be a valuable resource for reflection on, criticism of, and reformulation of the standards.

See also: Economics and ethics; Scientific method; Social sciences, philosophy of

TOM L. BEAUCHAMP

References and further reading

Barnes, B. (1974) *Scientific Knowledge and Sociological Theory*, London: Routledge & Kegan Paul.(Defends a 'strong programme' for the sociology of scientific knowledge in which social interests control scientific practice, including the acceptance of scientific hypotheses.)

Beauchamp, T.L., Faden, R.R., Wallace, Jr, R.J and Walters, L. (eds) (1983) *Ethical Issues in Social Science Research*, Baltimore, MD: Johns Hopkins University Press.(An anthology of original essays by leading figures in social science methodology and moral theory.)

Durkheim, É. (1895) *The Rules of Sociological Method*, ed. G.E.G. Catlin, trans. S.A. Solovay and John H. Mueller, Glencoe, IL: The Free Press of Glencoe, 8th edn, 1938. (An influential and difficult work that touches only tangentially on issues of value-free science and the basis of moral norms.)

Gräfrath, B. (1991) 'Forschungsinterese, Tatsachenwissen und praktische Orientierung', *Zeitschrift für philosophische Forschung* 45: 558-70.(Interpretation and defence of Max Weber's concept of value-freedom 'Wertfreiheit'.)

Keat, R. (1989) 'Relativism, Value-Freedom, and the Sociology of Science', in *Relativism: Interpretation and Confrontation*, ed. M. Krausz, Notre Dame, IN: University of Notre Dame Press, 272-98.(Cautious, analytical study of the relationships between issues in value-freedom, value relativism, and value-free sociology of science.)

Kelman, H. (1968) *A Time to Speak: On Human Values and Social Research*, San Francisco, CA: Jossey-Bass. (Powerful and pioneering work on ethical issues in social science research.)

Longino, H.E. (1989) *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*, Princeton, NJ: Princeton University Press.(Wide-ranging interpretation of scientific inquiry as a contextual process in which social values give shape to scientific development and claims to knowledge.)

Myrdal, G. (1958) *Value in Social Theory*, ed. P. Streeten, London: Routledge & Kegan Paul.(A comprehensive treatment by a leading spokesman for the value-impregnation of the social sciences.)

Proctor, R. (1991) *Value Free Science: Purity and Power in Modern Knowledge*, Cambridge, MA: Harvard University Press.(Detailed historical and contemporary treatment of controversies about value judgments in science, but excluding moral problems in the professions and ethics and public policy.)

Rudner, R. (1953) 'The Scientist *Qua* Scientist Makes Value Judgements', *Philosophy of Science* 20: 1-6.(Brief and clear argument that value judgments are essentially involved in the procedures of science.)

Schumpeter, J.A. (1949) 'Science and Ideology', *American Economic Review* 39: 345-59.(Influential address on whether ideological bias is a significant factor in 'scientific economics', with an examination of Adam Smith, Marx, and Keynes.)

Weber, M. (1949) *The Methodology of the Social Sciences*, ed. and trans. E.A. Shils and H.A. Finch, Glencoe, IL: The Free Press of Glencoe.(The most celebrated spokesman for value free science and the most influential work.)

Weed, D. (1994) 'Science, Ethics Guidelines, and Advocacy in Epidemiology', *Annals of Epidemiology* 4: 166-71. (A strong, readable defence of advocacy in social and public health research, especially for scientists in roles requiring value judgments.)

Value, ontological status of

We evaluate persons, characters, mental states, actions, inanimate objects and situations using very abstract terms such as 'good', 'unjust' and 'beautiful', and more concrete terms, such as 'courageous', 'cruel' and 'crass', drawn from fields such as aesthetics, ethics, politics and religion. Do these evaluations ascribe value properties to the entities evaluated? If so, what are these properties like? If not, what are we doing when we evaluate?

The simplest way to understand ethical discourse is by analogy with ordinary fact-stating discourse. An ethical judgment is to be understood as either true or false, being true if the situation it describes obtains, false if it does not. If ethical judgments are fact-stating, then it is natural to take evaluative vocabulary such as 'is good' and 'is courageous' as picking out properties of the entities which are evaluated.

What are these value properties like? First, we need an account of how properties are different from other kinds of entities. Then we need to decide whether values are natural or non-natural properties, where natural properties are the proper subject matter of the various sciences. G.E. Moore in *Principia Ethica* (1903) argued that goodness is a *sui generis* non-natural property. In contrast, many have tried to conceive of value properties as natural, principally by defining evaluative vocabulary in terms of the psychological states of some subject or subjects in certain conditions. By varying the state, subject and conditions, numerous such accounts can be obtained; for example, 'is good' might mean 'is desired by the speaker here and now' or 'is something we would desire to desire….in conditions of full imaginative acquaintance'.

Mackie (1977) thinks that our ethical discourse commits us to the existence of value properties which he calls 'objective values'. But he argues that such values would be non-natural, Moore's property of goodness vividly illustrating their 'queerness'. Our knowledge of them would be unexplained and, since ethical judgments are automatically motivating, these values would have a mysterious action-guiding force, quite unlike natural properties. Thus Mackie adopts an error theory of our evaluative discourse. There are no objective values. Since every ethical judgment commits us to such values, they are all false. His diagnosis of the error draws upon Hume's projectivism (see Projectivism). In truth, there is only the world of natural properties and our affective reactions to it, such as moral approval and disapproval. But we project such reactions back onto the world and speak as if it contained properties, such as goodness and badness, which merit the reactions.

Denying the existence of objective values need not force one to accept the error theory, for one can deny Mackie's conceptual claim that our ethical discourse says that there are objective values. Instead, one can think of this discourse as having an expressive, as opposed to a descriptive, function. Ayer's emotivism is an early version of this idea (see Emotivism). It is too crude, however, because it merely says that ethical judgments serve to express our affective reactions, failing to explain how it is that ethical discourse appears so similar to genuine descriptive discourse. After all, we are happy to say that ethical judgments are true or false, we engage in ethical argument to determine the correct answers to our ethical questions, we presume to persuade others to abandon their mistaken views and we acknowledge that our current ethical opinions may be wrong. Each of these features of our ethical practice tempts us to think that there is a realm of ethical facts which constrain the practice, facts which we aim to describe with our ethical discourse.

Blackburn's quasi-realist project (1984; 1993) attempts to explain the shape of our ethical practice on the basis of its expressive function. He argues that our ethical discourse is all right as it is, for, despite the appearances, it contains no erroneous commitment to objective values. Blackburn insists that we express, rather than describe, our affective reactions when we make ethical judgments. But one should ask whether a sharp line can be drawn between expression and description and, in turn, whether there is much to choose between Blackburn's position and the view which takes values to be natural properties involving our psychological states.

Mackie and Blackburn share the conception of the natural world as the world of science, which exists independently of human sensibility. A real entity belongs to the natural world, and values are not real. But perhaps this is too restrictive. McDowell (1985) has argued that values should be compared with secondary qualities, such as redness (see Secondary qualities). It is a conceptual truth that an object is red if and only if it would look red to appropriately receptive subjects in appropriate conditions. Thus redness is a dispositional property which is

conceptually connected to human sensibility, and so cannot belong to the natural world. But, since an object can possess this disposition independently of any particular experience, we are happy to think of an object as really being red, its looking red to us in the right conditions being perception of the redness which it possesses. Analogously, McDowell urges that a value, such as goodness, is to be thought of as a dispositional property of objects which is conceptually connected to an affective reaction in the right subjects in the right conditions. Again, such a value is not part of the natural world, but it is nevertheless real, the relevant affective reactions detecting its presence. If all this is right, one can argue against Mackie that our ethical discourse only commits us to the reality of values in the weak sense that their existence is independent of any particular experience. The appropriateness of the analogy between values and secondary qualities is fiercely debated. In particular, one ought to question whether any sense can be made of a perceptual route to values, especially when our affective states are often reactions to imagined or described situations, not perceived ones.

See also: Moral realism

ALEX OLIVER

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Victor Gollancz.(Chapter 6 contains his emotivist theory of ethical discourse.)

Blackburn, S. (1984) *Spreading the Word*, Oxford: Clarendon Press.(Chapters 5 and 6 introduce his distinctive combination of projectivism and quasi-realism.)

Blackburn, S. (1993) *Essays in Quasi-Realism*, New York: Oxford University Press.(A collection of essays developing his quasi-realist project; chapter 8 includes criticism of Mackie's error theory and of the analogy between values and secondary qualities.)

Darwall, S., Gibbard, A. and Railton, P. (1992) 'Toward *Fin de siècle* Ethics: Some Trends', *Philosophical Review* 101 (1): 115-89.(An excellent overview of different theories of the metaphysics of value.)

Lewis, D. (1989) 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society* 63: 113-37, supplement.(Argues that it is a conceptual truth that something is a value if and only if we are disposed in conditions of fullest imaginative acquaintance to desire to desire it.)

McDowell, J. (1985) 'Values and Secondary Qualities', in T. Honderich (ed.) *Morality and Objectivity*, London: Routledge & Kegan Paul, 110-29.(Argues that values are real by comparing them with secondary qualities.)

McGinn, C. (1983) *The Subjective View*, Oxford: Clarendon Press.(Chapter 8, sect. iii criticizes the analogy between values and secondary qualities.)

Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin. (Chapter 1 advances the error theory of ethical discourse.)

Moore, G.E. (1903) *Principia Ethica*, ed. T. Baldwin, Cambridge: Cambridge University Press, revised edn, 1993. (Chapter 1 argues for goodness as a *sui generis*, non-natural property.)

Sayre-McCord, G. (ed.) (1988) *Essays on Moral Realism*, Ithaca, NY: Cornell University Press.(An excellent collection of papers both for and against the existence of ethical value, including a reprint of McDowell's paper and the relevant passages from Mackie.)

Values

The theory of value has three main traditions: subjectivism, which holds that the only valuable goods are subjective states of sentient beings; objectivism, which claims that while values must be human-related, they exist independently of us; and Neo-Kantian rationalism, which suggests that value is postulated on the basis of practical reason. Central distinctions in the theory of value are between subjective and objective values, instrumental and final values, intrinsic and extrinsic values, organic unities and the idea of an ultimate or architectonic value. There are also distinctions drawn between different types of value, such as moral and aesthetic value.

The theory of value has been neglected within Anglo-American analytical philosophy. However, recent philosophy has seen a revival in value theory. The *subjectivist* school remains the dominant orthodoxy, but recent philosophy has seen the development of new forms of *objectivism* and an increasing emphasis on value theory within the *Neo-Kantian rationalist* tradition, a tradition which does tie an account of value directly to a theory of practical reasoning.

Subjectivists assert that the only valuable goods are subjective states of sentient organisms. ‘Subjective’ can be given a wide reading to include the essential interests of a subject; on a wide definition of subjective value even the conditions necessary for the flourishing of an organism can be included in the definition. Here, as elsewhere, ‘subjective’ and ‘objective’ accounts of the subject matter shade into each other. Thus, utilitarian theories of value have evolved from Bentham’s simple hedonism, via revealed preference theories, to definitions of the valuable in terms of what is objectively good for people (see Hedonism; Perfectionism; Rationality, practical; Utilitarianism).

Objectivists claim greater degrees of independence between values and human interests and concerns. Moderate objectivists would concede that value is an anthropocentric category, and that their list of the good things in life must relate to human concerns. However, they would insist that these components of a good life are preferable because they are good, and not vice versa (see Good, theories of the §5). Extreme objectivists argue that value is not a category distinctively attuned to human concerns but exists independently of human interests. Is biodiversity, for example, good in itself? Or is it good only relative to our interests (see Environmental ethics §2)?

Neo-Kantian rationalists offer a third approach to the nature of value. They argue that subjective interests give rise to reasons, which are then subjected to formal tests to see if they are sufficient to bring about the ends they prescribe for an agent. If the reason passes these tests, the end is taken to be valuable. This theory views the connection between practical reasoning and value differently from its rivals; practical reasoning does not solely work out the means to pre-given ends but also establishes which ends are worthy of choice.

Something may be valued in itself, or for the contribution it makes to achieving a further state which is valuable. Jogging may be tedious, but it is instrumental in achieving health, which is finally valuable as an end in itself. This distinction between the instrumentally valuable and the finally valuable is properly located within the theory of practical reasoning, since it concerns the place of an ‘end’ or goal within an agent’s system of practical ends. (Thus it is sometimes called the *means/ends* distinction - see Practical reason and ethics.)

It ought not to be confused with a related distinction, properly located within the theory of value, between *intrinsic* and *extrinsic* values. There are philosophically motivated arguments for collapsing these distinctions, but keeping them apart avoids begging the question against those who, for example, believe in final ends which are extrinsically valuable. It has been argued that Kant’s theory of happiness, for example, sees it as an extrinsic value which is none the less our final end (see Kant, I. §11).

The idea of intrinsic value has been explained in many different ways. A central line of thought, found in G.E. Moore, is that intrinsically valuable things are the source of their own value whereas extrinsically valuable things derive their value from another source; intrinsic value is grounded on, but not reducible to, the ‘natural’ intrinsic properties of the object (see Moore, G.E. §1; Supervenience). Another proposal is to explain the grounds of intrinsic value via the idea of *organic unity*. Combinations of valuable things can form ‘organic’ wholes whose value is greater than the sum of the value of their component parts. The same was argued to be true of combinations of valuable with disvaluable or ‘indifferent’ things. For example, harming a sentient creature is bad,

but taking pleasure in doing so much worse.

Many philosophers have argued that values collectively have a structure and that *higher* values can be discerned which organize the entire system of value and are therefore most worthy of choice. Examples would be Aristotle's belief that contemplation was the best and most self-sufficient good that made human life as a whole worthy of choice and the similar role Kant attaches to the good will (see Aristotle §§21, 26; Eudaimonia).

How is *moral value* to be demarcated from other forms of value? One, admittedly controversial, criterion is to relate the nature of value to the test of convergence. If qualified experts converge in the making of a judgment, that is enough to sustain a 'thin' model of objectivity, which may perhaps offer a sufficient account of aesthetic values. There is no more to aesthetic value than being the fit object of such convergent judgment. However, in the moral case one requires the further feature that competent judges converge in judging a case to be, say, cruel in virtue of the fact that it is indeed cruel. In this latter case there is indispensable reference to the property, anthropocentric but real, which merits the response. This method of demarcation would have the advantage of explaining why it is internal to the practice of aesthetic judgment that a plurality of such judgments is inherent in the subject matter and can rationally be tolerated. However, this is not so in the moral case, where divergent judgment invites correction.

See also: Art, value of; Axiology; Fact/value distinction; Good, theories of the; Ideals; Li; Moral realism; Value, ontological status of

ALAN THOMAS

References and further reading

Cavell, S. (1979) *The Claim of Reason*, Oxford: Oxford University Press, part III.(Reflects on the different roles of pluralism and disagreement within moral and aesthetic discourse.)

Grice, P. (1991) *The Conception of Value*, Oxford: Oxford University Press.(Another foundational work in value theory, suggestively parallel to the Korsgaard but drawing inspiration as much from Aristotle as from Kant.)

Griffin, J. (1986) *Well-Being: Its Meaning, Measurement and Moral Importance*, Oxford: Oxford University Press, esp. part I.(A comprehensive and important study of the fundamental idea of an agent's wellbeing which explains many of the different approaches to the issue.)

Korsgaard, C. (1983) 'Two Distinctions in Goodness', *Philosophical Review* 92: 27-49.(Argues against collapsing the means/ends and extrinsic/intrinsic distinctions and suggests that Kant's concept of happiness is as an extrinsic yet final end.)

Korsgaard, C. (1986) 'Aristotle and Kant on the Source of Value', *Ethics* 96: 486-505.(Explains the idea of a 'source' of value in the context of these two foundational theories of morality in the Western tradition.)

Korsgaard, C. (1996) *The Sources of Normativity*, Cambridge: Cambridge University Press.(A classic presentation of a Kantian 'rationalist' approach to practical reasoning and value.)

Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press, part V, chaps 1-6. (Nozick's extended discussion is the closest in recent philosophy to the classical doctrines of the Austro-German school of axiology. These chapters contain discussions of a wide range of issues in value theory, including the doctrine of organic unities.)

Wiggins, D. (1987) *Needs, Values, Truth*, Oxford: Blackwell; revised edn, 1991.(Contains many seminal papers which present the view that moral properties are tied to our evaluative interests and thus anthropocentric, but none the less real.)

Vasubandhu (4th or 5th century AD)

An Indian Buddhist philosopher of the fourth or fifth century, Vasubandhu was a prolific author of treatises and commentaries. Best known for his synthesis of the Sarvāstivāda school of Abhidharma, he was sympathetic with the Sautrāntika school and frequently criticized Sarvāstivāda theory from that perspective. Vasubandhu eventually became an eminent exponent of the Yogācāra school. He also wrote short treatises on logic that influenced Dignāga, traditionally said to have been his disciple.

Probably the most original of Vasubandhu's philosophical works are his two short works in verse, known as the Viṃśatikākārikāvṛtti (Twenty-Verse Treatise) and the Triṃśikākārikāvṛtti (Thirty-Verse Treatise). In these two works, he argues that one can never have direct awareness of external objects, but can be aware only of images within consciousness. Given that some of these images, such as those in dreams and hallucinations, are known to occur without being representations of external objects, one can never be certain whether a given image in awareness corresponds to an external object. Because one can never be sure of what is externally real but can be sure of internal experiences, he concludes, a person seeking nirvāṇa *should focus attention on the workings of the mind rather than on the external world.*

1 Life and works

The earliest and the most complete biography of Vasubandhu was compiled by Paramārtha (499-569), a Chinese Buddhist pilgrim monk. It is preserved in the Chinese Buddhist canon. Some information can be gathered also from Tibetan biographies by Bu ston (1290-1364) and Tāranātha (1575-1634). The Tibetan biographies, however, differ in many points from Paramārtha's work. The available data from the various sources have been collected and studied by Erich Frauwallner.

According to Buddhist tradition, Vasubandhu was born in Peshawar, now part of Pakistan. Early biographers claim that he was initially a follower of the Sarvāstivāda (also known as Vaibhāṣika) school of Buddhism, later switched his allegiance to the Sautrāntika school and finally was converted to Mahāyāna by his elder brother Asaōga, a founder of the Yogācāra school. Modern scholars have been divided on the questions of his date and which of the many works attributed to him are really of his authorship. Frauwallner argued (1951), on the basis of discrepancies within the Chinese and Tibetan biographies concerning Vasubandhu's dates, that there were two influential Buddhist philosophers named Vasubandhu who became conflated by Paramārtha. According to this hypothesis, the earlier Vasubandhu lived in the fourth century and was the younger brother of Asaōga; he wrote numerous extensive commentaries on Mahāyāna scriptures, especially those associated with the Yogācāra school. The later Vasubandhu, according to Frauwallner, was the author of the *Abhidharmakośa* and lived in the late fifth century: it was he who wrote several short Yogācāra treatises and short works on logic and trained Dignāga, who eventually became best known as a logician. Frauwallner's hypothesis was quickly disputed by P.S. Jaini (1958) and has since been disputed by other scholars as well. Although the issue has not been definitively settled, Frauwallner's once-influential hypothesis has become decreasingly accepted by specialists in Vasubandhu's philosophy, although it still has adherents among prominent specialists in the history of Buddhist thought. This controversy has no bearing on the discussions that follow, since all scholars on both sides of the debate agree that the texts discussed below were by the same author.

2 Contributions to Abhidharma theory

One of the most influential texts in all of Indian Buddhism is Vasubandhu's *Abhidharmakośa*. According to the author's own explanation of the title, the text is comparable to a scabbard (*kośa*) from which one extracts the sword of wisdom, which is the means by which one cuts through delusions and attains the highest reality (*abhidharma*), namely *nirvāṇa*. The central argument of the book is that all complex entities are, by the very fact that they are complex, less stable and therefore less capable of providing satisfaction than simple entities. Therefore, one seeking satisfaction should focus the attention on simple objects rather than on complex entities. Examples of simple objects are the basic building blocks of experience, namely the basic qualities such as colour and shape that are apprehended in vision, sounds that are apprehended by hearing, the various qualities apprehended by the other external senses, and the moods and emotions apprehended by the mind. Examples of complex objects are houses, chariots, people and all the various things that one tries to acquire or to influence.

Another example of a complex object is that which we take to be at the centre of all experience, the self.

Much of the *Abhidharmakośa* is dedicated to showing how a world constructed only of causally related qualities can function in the absence of a substantial substratum; the world is a world of properties (*dharma*) but not of property-owning (*dharmin*) substances. The work is divided into nine chapters. The first deals with a catalogue of the *dharmas* that form the content of all experience. The second discusses sense faculties and various other faculties of living beings. The third discusses the cosmos and its various populations. The world, it is argued, derives not from a single source but from the combination of the karma of billions of sentient beings; so karma becomes the subject matter of the fourth chapter. Karma is action that arises from latent patterns of habit; these habitual tendencies are discussed in detail in chapter five. People who eliminate negative, destructive habits become noble persons, who are classified in the sixth chapter. Nobility of character is a function of correct understanding, discussed in chapter seven. Since knowledge arises through skill in meditation, meditation is the subject of the eighth chapter. The ninth chapter, which may originally have been an independent treatise, recapitulates a series of arguments against a substantial basis for our sense of personal identity and individuality.

The *Abhidharmakośa* is a text written in verse, to which the author wrote an extensive prose commentary. The verse text presents a straightforward account of cosmology and psychology as taught by the Vaibhāṣika school of Buddhism. The prose commentary, however, examines these doctrines more critically. Since the author often goes into considerable detail about various positions that Buddhist philosophers of different views presented in the past, the work is an invaluable source of information on the controversies among Buddhists and between Buddhists and non-Buddhists during the first few centuries AD. Vasubandhu often, but by no means always, sides with critics from the Sautrāntika school of Buddhism. One of the salient characteristics of this school was a tendency to regard most Buddhist theories as merely provisional or heuristic doctrines, aimed not at giving a definitive account of the inner and outer worlds but at providing a framework within which the realities of these worlds can be discovered. Because the Sautrāntika school tended to be doctrinally more fluid than the relatively rigid orthodoxy of the Vaibhāṣika school, it is often seen as a conceptual bridge leading from the more conservative schools to the Mahāyāna movement. Owing to its comprehensive nature and to its serving as a propaedeutic for Mahāyāna philosophies, the *Abhidharmakośa* became a cornerstone of the curriculum at Indian Buddhist universities such as the one at Nālandāmonastery. It is still common for Tibetan scholastics to devote five to eight years of their training to a thorough study of the *Abhidharmakośa* before going on to study Mādhyamika and Yogācāra philosophy. The text was also central to Buddhist scholastics in China. All told, it was as central to medieval Buddhist education as Aquinas' *Summa theologiae* to medieval Christian scholastics in Europe.

3 Contributions to Yogācāra

The early Buddhists discussed different worlds of experience, by which was meant different ways of experiencing the world. The 'foolish masses' of unreflective people are preoccupied with pleasures of the senses, so they are said to live in the realm of desire. More reflective people see sensible objects without hankering after them as much as unreflective people, so they are said to live in the world of appearances, by which it is meant that they perceive what appears to the senses without desiring them or despising them. People who abstract their attention away from sensible objects altogether are said to live in the world of nonappearance. While all these beings may live side by side in the same physical world, their subjective experiences of the world are radically different. Taking this commonplace observation as a point of departure, Vasubandhu argues in his *Viṃśatikākārikāvṛtti (Twenty-Verse Treatise)* and *Triṃśikākārikāvṛtti (Thirty-Verse Treatise)* that every sentient being is in a sense locked within its subjective world and has no direct contact with the external physical world. That there is an external world can be known through inference only, not through direct cognition. All experience, in other words, is mediated through one's habitual ways of seeing things. What makes one person's experience of physical circumstances different from another person's experience of those same circumstances is not the circumstances themselves; rather, it is the moods, emotional tendencies, memories and associations that each person uniquely brings to the circumstances.

One of Vasubandhu's arguments for this conclusion is that beings who experience the torments of hell cannot actually be in a physical setting of the sort described in descriptions of the hells. Those reports speak of guardians and beasts that prowl environments with heat of unbearable intensity and torment the inhabitants of those regions. But reason shows that these hell-guardians cannot themselves suffer from the conditions that the victims of hell

find intolerable. If the guardians suffered as much as the inmates, they would be so overcome with pain that they would not be able to perform their task of tormenting the captives. Therefore the guardians and beasts, and indeed all the other torments of hell, must be purely subjective experiences that cannot have an objective counterpart in the external world, no matter how intensely they are felt.

Vasubandhu observes that when our senses are capable of presenting us with experiences of things that are not really as they are experienced, our only recourse to correct the presentations of the senses is reason. But reason also lets us down. Reason, for example, provides us with the notion of atoms, that is, bits of matter that are tiny to the greatest possible extent. Reflection, however, shows that the concept of atoms leads to inconsistencies. If an atom is really ultimately tiny, then it cannot be made up of parts that are physically smaller than the atom itself. If the atom has no parts, then the region of space occupied by the atom can have no subregions; there can be no upper region or lower region or eastern region and so on. If that is true, then the atom can occupy no space at all. If one atom occupies no space, then several atoms added together cannot occupy any space. Therefore what reason presents to us, namely that a macroscopic object is a complex of a very large number of ultimately small component parts, cannot really be the case after all. The atom, then, is a concept to which nothing in the real world can correspond.

If neither the senses nor reason can provide reliable pictures of external reality, we have no reliable means of knowing what the objective basis of our experiences is. When one thoroughly understands that a sentient being can never be aware of anything but its own awareness, says Vasubandhu, full attention can be turned to those subjective factors that give experience its flavour. Distracted neither by the appearances of the senses nor by the conceptual categories of reason, one gives up either hankering for or dreading what experience provides and one cultivates a bare attention to experience. This bare attention becomes an experience that is ‘unthinkable, healthy and stable’. This is the experience of liberation or *nirvāṇa*.

Along with texts by Asaōga and Sthiramati, Vasubandhu’s Yogācāra treatises became the textual foundation for much of scholasticism in both China and Tibet. Given the central importance of both the *Abhidharmakośa* and his Yogācāra writings, Vasubandhu had an influence on Asian Buddhism second only to that of the Buddha himself.

See also: Buddhism, Ābhidharmika schools of; Buddhism, Yogācāra school of

RICHARD P. HAYES
MAREK MEJOR

List of works

Vasubandhu (4th or 5th century AD) *Abhidharmakośa*, ed. and trans. L. de la Vallée Poussin, *L’abhidharmakośa de Vasubandhu*, Paris and Louvain: Paul Geuthner, 1923-31, 6 vols; repr. Brussels: Institut belge des hautes études chinoises, 1971, 6 vols; trans. L.M. Pruden, Berkeley, CA: Asian Humanities Press, 1988.(La Vallée Poussin’s is still the best translation available of this key work from Vasubandhu’s phase as an Abhidharma scholar; Pruden’s translation, from La Vallée Poussin’s French, is to be used with great caution, but is the only one available in English.)

Vasubandhu (4th or 5th century AD) *Karmasiddhiprakaraṇa*, trans. S. Anacker, ‘A Discussion for the Demonstration of Action’, in Anacker, S. 1984; trans. E. Lamotte, ‘Le Traité de l’acte de Vasubandhu. *Karmasiddhiprakaraṇa*’, *Mélanges chinois et bouddiques* 4: 151-288, 1936.(A work from Vasubandhu’s Yogācāra phase in which he criticizes Sarvāstivādin views of the mechanics of ethically charged actions and their fruitions.)

Vasubandhu (4th or 5th century AD) Madhyāntavibhāgaśāstra, trans. S. Anacker, ‘Commentary on the Separation of the Middle from Extremes’, in Anacker 1984; trans. T. Kochumuttom, ‘Discrimination Between Middle and Extremes’, in Kochumuttom, T.A. 1982.(An important work from the author’s Yogācāra phase.)

Vasubandhu (4th or 5th century AD) *Pañcaskandhakaprakaraṇa*, trans. S. Anacker, ‘A Discussion of the Five Aggregates’, in Anacker 1984.(An analysis of the classes of characteristics that constitute the physical body and the mentality of human beings.)

Vasubandhu (4th or 5th century AD) *Trisvabhāvanirdeśa*, trans. S. Anacker, ‘The Teaching of the Three Own-Beings’, in Anacker 1984; trans. T. Kochumuttom, ‘A Treatise on the Three Natures’, in Kochumuttom 1982; trans. F. Tola and C. Cragonetti, ‘The *Trisvabhāvakārikā* of Vasubandhu’, *Journal of Indian Philosophy* 11: 225-66, 1983.(A seminal work in early Yogācāra theory.)

Vasubandhu (4th or 5th century AD) *Triṃśikākārikāvṛtti*, trans. L. de la Vallée Poussin, along with the

Viṃśatikākārikāvṛtti and a Chinese commentary by Xuanzang as *La Siddhi de Hsuan-tsang*, Paris, 1928-48, 3 vols; trans. S. Anacker, 'The Thirty Verses', in Anacker 1984; trans. T. Kochumuttom, 'A Treatise in Thirty Stanzas', in Kochumuttom 1982.(An important Yogācāra treatise; La Vallée Poussin's French translation is excellent and well annotated.)

Vasubandhu (4th or 5th century AD) *Vādavidhi*, trans. S. Anacker, 'A Method for Argumentation', in Anacker 1984.(One of the earliest Buddhist works on debate and logic, later criticized by Dignāga, whose work on logic superseded Vasubandhu's.)

Vasubandhu (4th or 5th century AD) *Viṃśatikākārikāvṛtti*, see *Triṃśikākārikāvṛtti* for French translation; trans. S. Anacker, 'The Twenty Verses and their Commentary', in Anacker 1984; trans. T. Kochumuttom, 'A Treatise in Twenty Stanzas', in Kochumuttom 1982.(Along with the *Triṃśikākārikāvṛtti*, one of the foundational texts for Yogācāra theory.)

References and further reading

Anacker, S. (1984) *Seven Works of Vasubandhu, the Buddhist Psychological Doctor*, Delhi: Motilal Banarsidass. (Contains well-annotated translations, along with good introductions, of seven key works. Also has a general introduction and short biography of Vasubandhu, with discussion of Frauwallner's hypothesis of two Vasubandhus.)

Frauwallner, E. (1951) *On the Date of the Buddhist Master of the Law Vasubandhu*, Rome: Serie Orientale Roma.(The author's first presentation of the hypothesis that there were two prominent Buddhist authors by the same name who were conflated into one by traditional biographers.)

Frauwallner, E. (1957) 'Vasubandhus *Vādavidhiḥ*', *Wiener Zeitschrift für die Kunde Süd- und Ostasiens* 1: 104-46.(A study in German of the *Vādavidhi*.)

Frauwallner, E. (1961) 'Landmarks in the History of Indian Logic', *Wiener Zeitschrift für die Kunde Südasiens* 5: 125-48.(A classical study of the relative chronologies of key Indian philosophers. Further discussion of the two-Vasubandhu hypothesis.)

Jaini, P.S. (1958) 'On the Theory of Two Vasubandhus', *Bulletin of the School of Oriental and African Studies* 21 (1): 48-53.(Disputes Frauwallner's hypothesis.)

Kochumuttom, T.A. (1982) *A Buddhist Doctrine of Experience: A New Translation and Interpretation of the Works of Vasubandhu the Yogācārin*, Delhi: Motilal Banarsidass.(Contains studies and translations of Vasubandhu's Yogācāra writings, disputing the common claim that these texts argue for subjective idealism.)

Nagao, Gadjin (1987) 'Vasubandhu', in M. Eliade (ed.) *Encyclopedia of Religion*, New York: Macmillan, vol. 15, 191a-193b.(A study by one of the world's leading scholars of Yogācāra Buddhism.)

Vātsyāyana (5th century)

Vātsyāyana belonged to the Nyāya school of Indian philosophy, and his Nyāyabhāṣya is the first extant commentary on the Nyāyasūtra, the foundational text of that school. In it, he emphasized the distinctive epistemological and logical character of the topics he deemed appropriate for treatment by Nyāya philosophers. In so doing, he helped both to establish the authority of the Nyāya school in matters related to logical reasoning, and to demarcate the enterprise of the Nyāya school from that of the earlier, more traditional, soteriological approach of the Upaniṣads. His commentary on the Nyāyasūtra set the agenda for succeeding generations of Nyāya commentators and their Buddhist opponents. In particular, Vātsyāyana initiated arguments that were to become crucial in the Nyāya defence of its characteristic brand of realism.

Little is known about Vātsyāyana, who is also known as Pakṣilasvāmin. His precise dates remain a matter of scholarly interpretation of the relevant textual materials, most particularly the dating of the *Nyāyasūtra*, but placing him between 400 and 500 AD is a good working hypothesis. Vātsyāyana's chronological place in the enduring philosophical dialogue between the Nyāya and Buddhist philosophers is, however, much clearer. He directed many of his arguments against his Buddhist predecessor, Nāgārjuna, and, in turn, his work was criticized by the great Buddhist logician Dignāga. The dialogue continued with Uddyotakara, who sought to repair and develop Vātsyāyana's thought in the light of Dignāga's criticisms. In this fashion, the philosophical debate between the Nyāya and Buddhist philosophers continued until the eleventh century, with each side offering an ever more refined defence of its position; it culminated in the great philosophical wrangle between the Nyāya philosopher Udayana and his Buddhist opponent Jñānaśrīmitra.

At stake was the Nyāya defence of realism. Nyāya is distinctive among Indian schools of philosophy in its advocacy of a form of direct realism, in which the world is seen to be made up of everyday objects, such as pots and cows, amenable to both direct perception and to verbal expression. The possibility of forming correct judgments about the objects in the world (*prameya*) rests upon the efficacy of the four accredited instruments of knowledge (*pramāṇa*), listed in the *Nyāyasūtra* as perception, inference, comparison and verbal testimony (see Nyāya-Vaiśeṣika §6). In the *Nyāyabhāṣya*, Vātsyāyana attempted to counter Nāgārjuna's devastating critique of the very notion of a *pramāṇa*, a critique which had endangered the whole epistemological enterprise of the Nyāya school. He also defended, in the face of a developing Buddhist phenomenalism, the Nyāya view of the direct perception of objects. In so doing, he helped to clarify the relationship between perception and language.

Nāgārjuna had charged that the Nyāya dependence on the four *pramāṇas* was itself in need of substantiation. How can we be sure of the reliability of the *pramāṇas* themselves? Does our certitude come from invoking a further set of *pramāṇas*, taking the first step towards an infinite regress? In the *Nyāyasūtra* itself (2.1.19) we find a hint of the problem, when it is suggested rather cryptically that the *pramāṇas* are like lamp light. Since a lamp is self-illuminating, it was suggested that a *pramāṇa* could be self-validating. Vātsyāyana rejected such an interpretation. Not only does it fail to address Nāgārjuna's doubts about the *pramāṇas*, it also contradicts the basic Nyāya view that the *pramāṇas* alone, without any additional category of self-evidence, should be the source of knowledge. In his commentary on this *sūtra*, Vātsyāyana explains that the purpose of the lamplight analogy is to demonstrate that the *pramāṇa*-*prameya* distinction is neither rigid nor mutually exclusive. Since a lamp can fulfil two roles, both revealing other objects with its light and itself being an object of cognition, this demonstrates how one and the same object can function as both *pramāṇa* and *prameya*. Having resisted the idea that any set of *pramāṇas* should be considered axiomatic, Vātsyāyana attempts to show how the justification for any *pramāṇa* can come from the *pramāṇas* themselves. He points out that each *pramāṇa* is a generic term, encompassing different acts of knowing of the same type. Thus if one act of perception is validated by another act of perception, this should not be seen as a *pramāṇa* validating itself. Vātsyāyana's reply to Nāgārjuna may be viewed as somewhat pragmatic. Any individual use of a *pramāṇa* can itself be the object of another cognition. This, he says, is sufficient for the pursuit of everyday business as well as liberation (*mokṣa*), and does not lead to an infinite regress.

In dealing with the individual *pramāṇas*, Vātsyāyana's dispute with the Buddhist phenomenalists over where to draw the line between perception and inference was crucial to Nyāya's realist view of the world. In the

Nyāyasūtra (1.1.4), perception had been characterized as 'that which is not expressible' (*a-vyapadeśyam*). Vātsyāyana uses this *sūtra* to argue that, as with infants, the ability to perceive an object is different from the ability to name it. This distinction between the perception of an object and its linguistic characterization anticipated an important distinction later Nyāya philosophers made between qualified and nonqualified perception. Vātsyāyana, however, strongly opposed the Buddhist view, which sought to drive a wedge between perception and language by making judgments of the form 'this is a cow' inferential rather than perceptual. Nyāya asserted a close accord between language and perception with the argument that perception gives us direct access to substantial, enduring objects like cows and trees. In the course of his arguments for the direct perception of physical objects, Vātsyāyana develops the arguments given in the *Nyāyasūtra* (2.1.31-6) concerning the relationship between a whole and its parts. While accepting that we can never simultaneously perceive all the parts of a tree, Vātsyāyana rejects the assumption that we need to do so in order to see a tree as such. A tree is a unity which is more than a mere conglomeration of its component parts, and which resides in its entirety in the parts of the tree. Vātsyāyana argues that if we are to speak of 'trees' at all, we must maintain the distinction of a whole over and above its parts. Later Nyāya realists, especially Uddyotakara and Udayana, made effective use of this distinction.

See also: Knowledge, Indian views of; Sense perception, Indian views of §§3-6

JOY LAINE

List of works

Vātsyāyana (5th century) *Nyāyabhāṣya*, in G. Jha (trans.) *The Nyāya-Sūtras of Gautama with the Bhāṣya of Vātsyāyana and the Vārttika of Uddyotakara*, Delhi: Motilal Banarsidass, 4 vols, 1984.(This four-volume set gives a translation of Vātsyāyana's commentary (*bhaṣya*) on the foundational text (*sūtras*) of the Nyāya system alongside a later commentary (*vārttika*) by Uddyotakara.)

References and further reading

Matilal, B.K. (1971) *Epistemology, Logic and Grammar in Indian Philosophical Analysis*, The Hague: Mouton.(A detailed account of the differences between the Nyāya and Buddhist schools, with a focus on their respective theories of language and perception. Contains references to Vātsyāyana and the role he played in the early years of this long philosophical dialogue.)

Matilal, B.K. (1977) *A History of Indian Literature*, vol. 6, *Scientific and Technical Literature*, Wiesbaden: Otto Harrassowitz. (Part 3, fasc. 2 deals with Nyāya-Vaiśeṣika. In section 8 of chapter 2, Matilal discusses Vātsyāyana's style of doing philosophy using a short translation of Vātsyāyana's commentary on the *Nyāyasūtra*.)

Matilal, B.K. (1986) *Perception: An Essay on Classical Indian Theories of Knowledge*, Oxford: Clarendon Press. (Chapter 2 contains a good discussion of the *pramāṇas*.)

Mohanty, J.N. (1970) *Phenomenology and Ontology*, The Hague: Martinus Nijhoff. (Chapter 17 gives a good account of Vātsyāyana's arguments concerning the direct perception of physical objects.)

Nyāyasūtra (*c*.400 AD) trans. G. Jha, *The Nyāya-Sūtras of Gautama with the Bhāṣya of Vātsyāyana and the Vārttika of Uddyotakara*, Delhi: Motilal Banarsidass, 4 vols, 1984.(Complete translation with detailed notes.)

Phillips, S.H. (1995) *Classical Indian Metaphysics*, La Salle, IL: Open Court.(A work more specifically about the debate between the New Nyāya, or Navya-Nyāya, school and the Advaitin philosopher Śrīharṣa. It does, however, contain several references to Vātsyāyana in a chapter about earlier Nyāya philosophers.)

Potter, K.H. (ed.) (1977) *Encyclopedia of Indian Philosophies*, vol. 2, *Nyāya Vaiśeṣika*, Delhi: Motilal Banarsidass.(A comprehensive introduction to the Nyāya-Vaiśeṣika system, containing detailed summaries of all known Nyāya-Vaiśeṣika works.)

Sastri, D.N. (1964) *The Philosophy of Nyāya-Vaiśeṣika and its Conflict with the Buddhist Dignāga School*, Delhi: Bharatiya Vidya Prakashan.(A detailed account of the defence of Nyāya realism in the face of Buddhist opposition. There are frequent references to Vātsyāyana and his role in this long philosophical dispute.)

Vedānta

*Indian philosophical speculation burgeoned in texts called Upaniṣads (from 800 BC), where views about a true Self (*ātman*) in relation to Brahman, the supreme reality, the Absolute or God, are propounded and explored. Early Upaniṣads were appended to an even older sacred literature, the Veda ('Knowledge'), and became literally Vedānta, 'the Veda's last portion'. Classical systems of philosophy inspired by Upaniṣadic ideas also came to be known as Vedānta, as well as more recent spiritual thinking. Classical Vedānta is one of the great systems of Indian philosophy, extending almost two thousand years with hundreds of authors and several important subschools. In the modern period, Vedānta in the folk sense of spiritual thought deriving from Upaniṣads is a major cultural phenomenon.*

Understood broadly, Vedānta may even be said to be the philosophy of Hinduism, although in the classical period there are other schools (notably Mīmāṃsā) that purport to articulate right views and conduct for what may be called a Hindu community (the terms 'Hindu' and 'Hinduism' gained currency only after the Muslim invasion of the South Asian subcontinent, beginning rather late in classical times). Swami Vivekananda (1863-1902), the great popularizer of Hindu ideas to the West, spoke of Vedānta as an umbrella philosophy of a Divine revealed diversely in the world's religious traditions. Such inclusivism is an important theme in some classical Vedānta, but there are also virulent disputes about how Brahman should be conceived, in particular Brahman's relation to the individual.

In the twentieth century, philosophers such as Sarvepalli Radhakrishnan, K.C. Bhattacharyya and T.M.P. Mahadevan have articulated idealist worldviews largely inspired by classical and pre-classical Vedānta. The mystic philosopher Sri Aurobindo propounds a theism and evolutionary theory he calls Vedānta, and many others, including political leaders such as Gandhi and spiritual figures as well as academics, have developed or defended Vedāntic views.

1 Pre-classical Vedānta

The most abstract teaching of the earliest Upaniṣads centres on questions of theological metaphysics: what is the ultimate reality, what is its nature and relation to the world, and how can it be known? The ultimate reality is called Brahman, and all Vedānta, from the pre-classical to the modern, is Brahman-centred philosophy. Early Upaniṣads also include much psychological speculation, about a true self (*ātman*), at places declared identical with Brahman. Dimensions or layers of self-experience are identified, for example, a series of states of awareness in dream and sleep, and a series of bodies (*kośa*) ranging from the physical body to a vital 'sheath', then a mental sheath, then a 'supramental' sheath, and finally a most essential sheath said to be 'made of bliss'.

Early Upaniṣads are as much mystical treatises as philosophical texts. A mystical awareness of Brahman (*brahma-vidyā*) is the core teaching concerning human destiny, a 'supreme personal good' (*parama-puruṣārtha*). Upaniṣadic idealist speculation appears to be an attempt to explain the possibility of an exalted consciousness: Brahman as transcendent (timeless, spaceless) but also as world ground is somehow directly experienced as identical with the individual self. Even theistic passages, which suggest that Brahman is transcendent to the world and individuals as their Creator and Lord, teach the value of knowing Brahman directly through assiduous efforts of meditation and yoga (which literally means 'self-discipline').

The psychological thesis that self-awareness is nondual, that awareness is evident to itself without an intermediary or another awareness, is presented with an analogy to light: light is self-luminous; similarly, awareness is self-aware. This thesis is key to the classical Vedāntic subschool known as Advaita ('Nondualism') - as too is a causal doctrine, namely, that an effect is prefigured, or immanent, in its material cause, for example, a gold bracelet in gold (see §3). The Advaita Upaniṣadic reading stresses the sole reality of Brahman, who is nondualistically self-aware. Positions key to theistic subschools, opposed to the Advaita reading, are also adumbrated in various passages. Early Upaniṣads do not speak with a single voice.

However, views are expressed that span not only all the classical Vedāntic subschools but also the full extent of pre-classical, classical and even much modern Indian thought. For example, doctrines of karma and rebirth are veritably pan-Indian. Some Upaniṣadic teachings about moral virtues remain fairly constant through the later

civilization, although different emphases emerge in different religious and ascetic traditions. Social doctrines of caste are not prominent in the earliest Upaniṣads, but there are indications that a hereditary system of social privilege and responsibility was in the process of being formed.

Another pre-classical Vedāntic text is that portion of the *Mahābhārata* (the great Indian epic) known as the *Bhagavad Gītā(Song of God)*, which is probably the most popular religious work in the whole of Indian civilization. The *Bhagavad Gītā* echoes several Upaniṣads and develops Upaniṣadic themes and theses about the self as well as about Brahman. But whereas early Upaniṣads include many passages centred on ideas dear to classical Advaita Vedānta (and there are middle and late Upaniṣads that express only the Advaita point of view), the *Bhagavad Gītā* is to be aligned for the most part with the theistic classical subschools. Brahman is God, according to the *Bhagavad Gītā*, the Creator of the universe that God indwells. God is both transcendent to and immanent in this world (see God, Indian conceptions of §3). The *Bhagavad Gītā* upholds mystical awareness of Brahman - along with the yoga or discipline required for such awareness - as the solution to this-worldly ethical and political crises, a view echoed in modern times by Mahatma Gandhi and other nationalist leaders. In contrast to Upaniṣadic ideas about Brahman as beyond good and evil, in the Vedāntic theism of the *Bhagavad Gītā* and other texts, God upholds *dharma*, right conduct, and in special manifestations or *avatāras* (Kṛṣṇa, Rāma and so on) is born into the world to keep the social order on track.

In this way, by the end of the Upaniṣadic period (200 BC or thereabouts) Vedānta had begun to develop along two lines: Brahman - the Supreme Being by all counts - is conceived as a personal God, the Creator and Sustainer of the universe, manifesting as the several gods and goddesses and incarnate in *avatāras*; or, alternatively, Brahman is seen as an impersonal Ground of Being in the (idealist) sense of the entire phenomenal display of this universe viewed as a dream or illusory projection of a single Self. In all cases, the notion of liberation, that is, mystical knowledge or realization of Brahman (*brahma-vidyā*), is prominent, but the term is not understood in the same way by Advaita Vedāntins and theists. Advaitins come to hold that the self (*ātman*) of everyone is in reality nothing other than Brahman, and that in the mystical knowledge of Brahman there is known only the One, the sole true Existent, whose nature is strictly beyond determination by finite beings such as us; but that it is a homogeneous Being, Consciousness and Bliss is put forth as an idea to guide mystic pursuits. Vedāntic theists, in contrast, come to hold that the individual and God are meaningfully distinct. In the ‘highest knowledge’, the Supreme cannot be known in precisely the fashion that God knows God, for an individual knower is not strictly identical with the Creator, although God indwells everywhere, including, to be sure, the individual self. Advaita Vedāntins come to stress meditation and study of Upaniṣads to attain the *summum bonum*, while Vedāntic theists stress love and devotion to God.

2 The *Brahmasūtra*

Classical Indian philosophy is marked by argument and self-conscious defence of positions, as well as by an acute awareness of systematic interlock among the planks of a worldview or school. Distinct worldviews, represented in the several schools of classical philosophy, had emerged by the first few centuries AD. With exceptions (principally concerning Buddhist schools), each school has an early *Ur*- or root text that lays out positions comprehensively but also in abbreviated form - in *sūtras*, ‘threads’ or aphorisms that are compact and easy to memorize. Examples of basic texts are the *Mīmāṃsāsūtra*, defining the school of Mīmāṃsā, the *Nyāyasūtra*, spelling out Nyāya, and so on. The apparent root text of classical Vedānta is the *Brahmasūtra*. However, although the founders of the prominent Vedāntic subschools (Śaõkara, Bhāskara, Rāmānuja, Nimbārka, Madhva and Vallabhācārya) all write commentaries on the *Brahmasūtra* - each interpreting distinctively at least some of the terse and often vague or ambiguous *sūtras* - the Upaniṣads themselves, and not the *Brahmasūtra*, are the foundational Vedāntic texts. This is unlike the situation with, for example, Nyāya, where the *Nyāyasūtra* is veritably the foundational text. The *Brahmasūtra* should be regarded, then, as the text that first establishes a Vedāntic system, but not as the definitive statement of Vedāntic positions. The fact that there are diverse Vedāntic subschools attests to the diversity of Upaniṣadic teachings and to the vibrancy of Upaniṣadic traditions of study and learning well into classical times.

The *Brahmasūtra* is also known as the *Śārīrakamīmāṃsāsūtra (Aphorisms of (Scriptural) Exegesis Concerning the Embodied)*. The *Brahmasūtra* presupposes the *Mīmāṃsāsūtra (Aphorisms of (Scriptural) Exegesis (Proper))* - that is, an interpretation of the Veda including, according to classical Exegetes (Mīmāṃsakas), Vedānta’s Upaniṣads as well as the Vedas proper. Classical Vedānta is sometimes referred to as Uttara Mīmāṃsā, Later Exegesis, being

focused on the later portions of scripture (Upaniṣads) and contrasting but also continuous with Pūrva Mīmāṃsā, Prior Exegesis, which is focused on the earlier portions of scripture and is the school of Mīmāṃsāproper. A principal task the *Brahmasūtra* sets for itself is to differentiate Vedānta from Mīmāṃsā. This it does by declaring Brahman to be the object of Upaniṣadic statements, whose point is to provide insight and lead to mystical knowledge, as opposed to the Mīmāṃsāunderstanding of scripture as a series of injunctions on right ritual, right practice and in general the right way to live. The *Brahmasūtra* adopts Mīmāṃsā's epistemology of scripture as self-certifying and authoritative within its own sphere (see Knowledge, Indian views of §3; Mīmāṃsā). But it specifies that that sphere is distinct in the case of the Upaniṣads; it is, namely, Brahman and knowledge of Brahman. The rites and actions prescribed by scripture for the embodied soul, the human agent, according to Exegetes, are, according to Vedāntins, nullified by the superior path of knowledge, whereby the embodied will come to know Brahman and be liberated from rebirth and death and future embodiment.

The *Brahmasūtra* has less subtle quarrels with other classical schools, although regarding Sāōkhya, scriptural interpretation remains the major issue. Early Upaniṣads do, in fact, at places resonate with Sāōkhya's understanding of nature or *prakṛti* as prefiguring within her unmanifest form all worldly manifestations. Moreover, Sāōkhya's hierarchical psychology of an embodied soul and its instruments or faculties, from reason to sense organs and organs of action, is rather clearly an Upaniṣadic teaching (see Sāōkhya). The *Brahmasūtra* is intent on integrating these views within a Brahman-centred philosophy, and disputes the contention that Sāōkhya is expressed in the Upaniṣads. But the *Brahmasūtra* also disputes Sāōkhya on metaphysical and other nonscriptural grounds. And Buddhist schools, Nyāya, Vaiśeṣika, Jainas and others are engaged and refuted on a wide range of philosophical issues.

For example, against the Vaiśeṣika notion of inherence (*samavāya*) as ontic glue binding a property to a property-possessor, the *Brahmasūtra* alleges that an infinite series of relators would be required to bind the inherence itself to the property and to the property-possessor. In this way, the notion is shown to be bankrupt. Against a Buddhist understanding of awareness as neither occurring in a self nor revealing external states of affairs, the *Brahmasūtra* alleges that the evidence of perception is to the contrary. The Jaina view that the soul has size is opposed on the grounds of untoward ramifications with regard to liberation. Bādarāyaṇa, the person (otherwise unknown) to whom the *Brahmasūtra* is traditionally ascribed, appears thoroughly knowledgeable about the philosophical landscape of his time (probably the first or second century AD), and the *Brahmasūtra* suggests a similar development of argument as is evinced in other early *sūtra* texts of the philosophical schools.

3 Issues among the subschools

No *Brahmasūtra* commentary has come down us that is earlier than Śaōkara's (early eighth century), but we know that there were earlier commentaries because Śaōkara and others whose works are extant mention a few. After Śaōkara, the great Advaitin, there are commentaries by Bhāskara (*c.*750), Rāmānuja (11th century), Nimbārka (*c.*1250), Madhva (1238?-1317?) and Vallabhācārya (1479-1531), authors who establish subschools rival to Advaita. Other, mostly later (original), commentaries were written as well, but these philosophers are the founders of the major classical subschools. Śaōkara and most of the others have followers who write subcommentaries, and within Advaita a couple of these (by Padmapāda, *c.*725, and Vācaspati Miśra, *c.*950) carve out distinct interpretations that are important even from a distant perspective contemplating classical Indian philosophy as a whole. The writings of classical Vedāntins are not confined to *Brahmasūtra* commentaries and subcommentaries; there are numerous noncommentarial treatises, as well as commentaries written on various Upaniṣads and the *Bhagavad Gītā*.

In these works, the *Brahmasūtra*'s attacks on competing worldviews are extended and refined, with some shift of focus as various non-Vedāntic philosophies become prominent or decline. But increasingly - and as early as Bhāskara - internecine disputes are aired, and by the era of Madhva come to dominate Vedāntic argument. And although it would be an error to ignore differences among non-Advaitic subschools, it is Advaita that the other subschools are keen to combat. With Madhva and his followers, not only is opposition to Advaita vehement, but also an alliance with Nyāya, a pluralist worldview with its own long history and literature, has to be counted as important as any Vedāntic heritage. This contrasts sharply with the polemics of the Advaitin Śrīharṣa (*c.*1150), who devotes enormous energy to refuting Nyāya, dismantling that realist and pluralist philosophy's every position down to apparent minutiae. Thus the nature of Vedāntic unity becomes problematic, and in fact scholars rarely

refer to classical philosophers merely by a 'Vedānta' tag, but at least differentiate Advaitins from the rest. The key issue that divides Advaita from its Vedāntic opponents, and that (loosely) unites the opponents as well, is one that, as noted above (§1), flows out of the Upaniṣads themselves and the *Bhagavad Gītā*: is Brahman the impersonal homogeneous Absolute of Advaita, or the God of the rival camps?

Other issues are tied closely to disputes about Brahman, in particular, controversies about the nature of the individual (*jīva*) in relation to Brahman. In fact, the most commonly employed designations of Vedāntic subschools derive from stances taken on this issue: Advaita (Nondualism: Brahman and the individual - *qua* being conscious or being a self - are identical; otherwise individuality is illusion), Bhedābheda (the Distinctness in Nondistinctness view of Bhāskara and others: Brahman and the individual are in at least one way distinct realities and in another nondistinct), Viśiṣṭādvaita (the Qualified Nondualism view of Rāmānuja and followers: both Brahman and the individual are realities and also identical in a qualified sense), Dvaita (the Dualism view of Madhva and company: Brahman and the individual are fundamentally distinct realities), and so on.

The relation of Brahman and the individual was not considered simply a matter of abstract metaphysics; the soundness of practical disciplines, indeed, of whole ways of life, was viewed as hinging on it. Theistic Vedāntins upheld - and today Hindu theists continue to advocate - practices known as *bhakti*, 'love' or 'devotion', as the way to a mystical *summum bonum* sometimes symbolized as adulterous love-making with God. In contrast, Advaitins advocate more ascetic paths, along with intellectual pursuits - in particular, Upaniṣadic study. And the *summum bonum* is conceived as realization of Brahman where one nondualistically (and timelessly, and so on) knows only oneself.

Exegetical issues dominate the intra-Vedānta crossfire. Advaitins, sensing a tension between Upaniṣads theistic statements and their view of Brahman's transcendence and unitary consciousness, develop exegetical strategies to reinterpret troublesome passages. Foremost among these is a view of scripture as like a wise and patient guru tailoring his teaching to his audience's capacities. Thus theistic views, and *bhakti* practices as well, are regarded as helpful for those incapable of appreciating the highest Upaniṣadic teaching, the utter truth, namely, that there is a single, undifferentiated self. (Some Advaitins make greater concessions to theism, holding that God is the inevitable conception of the Absolute from the finite human point of view. This position is often rehearsed among academic Neo-Vedāntins; see §4.)

Theistic Vedāntins, for their part, insist on a more literal reading of the Upaniṣads. Some also argue that all conveyance of meaning presupposes things as distinct. Moreover, the very act of comprehension of scriptural statements presupposes the reality of an individual cognizer. Advaitins are pushed by some of these points to assert the importance of nonliteral language, and struggle to formulate an appropriate theory of metaphor. They also admit that even a scripturally based understanding belongs to the province of illusion (*māyā*).

Concerning psychological and epistemological stances, there is less disagreement. Vedāntic theists tend to endorse the Advaita view of self-awareness as irreflexive and the Mīmāṃsāepistemology of self-certification (see Epistemology, Indian schools of §3). They also embrace the Upaniṣadic monism of Brahman as all-containing. But they refuse to go the subjectivist route in interpreting this doctrine, and dispute the Advaita view of worldly, finite things as an illusory projection of a single self. Instead, they see the unity of Brahman as a doctrine of material causality. With regard to the 'stuff' of the universe, a kind of Advaita is right: the world is the body of God. Even Madhva, who often seems intent on holding that in no way are God and individuals anything but distinct, says that God's body is the universe. According to Rāmānuja and other theists less radical than Madhva, Brahman has two natures: God as God is in God's self, which is God's necessary and essential nature, and God as God self-manifests in the world, as finite things that are real but contingent and inessential (see God, Indian conceptions of §3).

4 Neo-Vedānta

In modern India, the intellectual landscape has changed dramatically from classical times, reflecting science and influenced by Western thought. Vedānta, however, has proved surprisingly resilient, and not just within religious traditions but within universities as well. Colleges and universities were founded in India during the nineteenth century and academia has expanded enormously since then. Sanskrit, the intellectual language of the ancient and classical civilizations, has been replaced by regional vernaculars or English in most educational settings, and in

higher education English predominates. Philosophy is taught by professionals who are, by and large, at least as knowledgeable about the history of Western philosophy as about ancient or classical Indian thought. Nevertheless, what has come to be called Neo-Vedānta is a major intellectual movement on the campuses. There is also a Neo-Vedānta that is more a folk or religious phenomenon. Both types of Neo-Vedānta have complex roots, reflecting in part contemporary Hinduism and healthy traditions of mysticism, in part survival of classical modes of learning, in part Western idealist philosophy and in part deepened scholarly appreciation of classical Indian thought.

Concerning the folk phenomenon, it is probable that all Indian worldviews propounding a mystic goal grew up within traditions marked by the prominence of a spiritual teacher or guru; similar traditions continue today. Within contemporary Hinduism, each guru fashions a heritage from a wealth of prior Indian religious literature (the *Bhagavad Gītā*, Upaniṣads, Purāṇas, stories of the living enlightened, and so on) and, in some cases, from literature of other traditions as well. The best-known spiritual figures, eminently articulate gurus with national and international audiences, such as Swami Vivekananda (1863-1902) and Sri Aurobindo (1872-1950), are folk Vedāntins in the sense that, without allegiance to a school of classical thought, they draw on the *Bhagavad Gītā* and Upaniṣads and works influenced by classical Vedāntic conceptions. Both of these two great men learned Sanskrit, but, as has become increasingly common, both write in English. Vivekananda defends Vedānta as the plastic philosophy of Hinduism (see Ramakrishna movement); Aurobindo articulates a Brahman-centred worldview which he says is attuned to science but provides a deep explanation of which science, he argues, is incapable.

With regard to the nineteenth and early twentieth centuries, there would be little point in trying to differentiate folk and academic Neo-Vedānta. Although not professional academic teachers of philosophy (a rare thing in India, and indeed the West, at that time), Vivekananda and Aurobindo were well schooled in Western modes of thought. But it is with such broadly important cultural figures as Raja Rammohun Roy (1772-1833), sometimes referred to as the father of modern India, that Vivekananda and Aurobindo have affinity, not with any contemporary professor. Indeed, although Roy was no guru or spiritual leader, he may be taken as the first to articulate in the religious arena a synthesis of Western and Indian views. Among diverse accomplishments, he translated the New Testament into Bengali after learning Greek to read the original text. In ethics, he championed Christianity, urging that Hindus could and should embrace Christian moral teachings without relinquishing their own Vedāntic theological beliefs. In 1828, he founded the Brahmo Samaj (see Brahmo Samaj), a ‘church’ dedicated to ‘the worship and adoration of the Eternal, Unsearchable and Immutable Being, who is the Author and Preserver of the universe’ (Roy 1906, vol. 1: 216).

But although having wide influence, Roy, Vivekananda, Aurobindo and other important cultural figures were and are not professional philosophers. Academic Neo-Vedāntins differ markedly from nonprofessionals in ways reflective of a university setting. Sarvepalli Radhakrishnan (1888-1975), the most distinguished to date of these Neo-Vedāntins - President of India (1962-7), Spaulding Professor at Oxford (1936-9), Vice-Chancellor of Benares Hindu University (1939-48), Chancellor of Delhi University (1953-62) and teacher of philosophy at three other Indian institutions - used books and papers on classical systems, heavily peppered with apt comparisons to Western thought, to develop his own views. These are stated forthrightly in *An Idealist View of Life* (1929), but with Radhakrishnan’s typically careful scholarship and much attention to forerunners, chiefly Western idealists, but also classical Vedāntins. K.C. Bhattacharyya (1875-1949), professor of philosophy at the University of Calcutta and the teacher of many of the most important philosophers of the next generation, supports several Advaita-like views with tightly constructed arguments and insights expressed in austere, difficult prose. T.M.P. Mahadevan (1911-92), longtime professor of philosophy at the University of Madras, authored literally hundreds of papers and a dozen or so books, principally in exposition of the views of classical Advaitins - often defending Advaita in contemporary terms. More conservative than Radhakrishnan or Bhattacharyya, Mahadevan built bridges to folk traditions by discussing such popular saints as Ramakrishna and Ramana Maharshi. The Neo-Vedānta that these three present is a deep-set movement among philosophy professionals. Several others of similar persuasions deserve mentioning, but this is not the place for a long list.

See also: Awareness in Indian thought; Brahman; Hindu philosophy; Monism, Indian

STEPHEN H. PHILLIPS

References and further reading

Bādarāyaṇa (1st-2nd century AD) *Brahmasūtra*, trans. S. Radhakrishnan, *The Brahma Sūtra*, London: Allen & Unwin, 1960.(A translation with extracts from the commentaries of Śaōkara, Rāmānuja, Madhva and others.)

Bhagavad Gītā (200 BC - AD 200, disputed), trans. F. Edgerton, Harvard Oriental Series 38-9, Cambridge, MA: Harvard University Press, 1944, paperback repr. 1972. (Though Edgerton's interpretive essay - orginally volume 39 in the HOS - too much reflects outworn assumptions of nineteenth-century indologists, his translation is excellent, faithful and elegant. There are, however, dozens of acceptable translations into English and other modern languages.)

Bhattacharyya, K.C. (1956, 1958) *Studies in Philosophy*, ed. G. Bhattacharyya, Calcutta: Progressive Publishers, 2 vols.('Studies in Vedantism' in volume 1 is a penetrating analysis by the astute philosopher.)

Bhattacharyya, K.C. (1976) *Search for the Absolute in Neo-Vedanta*, ed. G.B. Burch, Honolulu, HI: University Press of Hawaii.(A collection of three Kantian reflections supporting an Advaita understanding of self; a long introduction by Burch traces influences and themes.)

Dasgupta, S. (1922-55) *A History of Indian Philosophy*, Cambridge: Cambridge University Press, 5 vols. (Volumes 2-5 are dominated by expositions of varieties of classical Vedānta; volume 1 includes a discussion of the thought of early Upaniṣads as well as Śaōkara's Advaita; the five volumes are probably still the premier introduction to ancient and classical Indian philosophies.)

Deutsch, E. and Buitenen, J.A.B. van (eds) (1971) *A Source Book of Advaita Vedānta*, Honolulu, HI: University Press of Hawaii.(The editors' introductions are excellent, explaining the positions and context of pre-classical Vedānta and the *Brahmasūtra*, as well as classical Advaita.)

Lacombe, O. (1937) *Absolu selon le Védānta (The Absolute According to Vedānta)*, Paris: Librairie orientaliste Paul Geuthner.(A classic study of the systems of Śaōkara and Rāmānuja.)

Mahadevan, T.M.P. (1969) *The Philosophy of Beauty*, Bombay: Bharatiya Vidya Bhavan.(A short book that is interesting for Mahadevan's endorsement of the classical Advaita understanding of Brahman on the one hand, and of a criterion of beauty as reflection of Brahman in the terms of temporality and finitude on the other.)

Narain, K. (1962) *An Outline of Madhva Philosophy*, Allahabad: Udayana Publications.(A valuable overview of Madhva's positions, stressing Madhva's opposition to Advaita.)

Potter, K.H. (ed.) (1981) *Encyclopedia of Indian Philosophies*, vol. 3, *Advaita Vedānta*, Princeton, NJ: Princeton University Press.(Contains an excellent introduction to Advaita philosophy, as well as summaries of works in the early Advaita school.)

Radhakrishnan, S. (1929) *An Idealist View of Life*, London: Allen & Unwin, 2nd edn, 1937.(The last chapter, entitled 'Ultimate Reality', shows the influence of classical Vedānta, as well as Western idealism.)

Raju, P.T. (1953) *Idealist Thought of India*, London: Allen & Unwin.(An excellent survey with many apt Western comparisons in the tradition of Radhakrishnan, who among other folk and academic Neo-Vedāntins is himself discussed.)

Roy, Raja Rammohun (1906) *The English Works of Raja Rammohun Roy*, ed. J.C. Ghose, Allahabad: Panini Office, 4 vols.(A collection of some of Roy's administrative, political, religious and philosophical writing; the passage in §4 is from 'The Trust-deed of the Brahmo Samaj'.)

Srinivasachari, P.N. (1934) *The Philosophy of Bhedābheda*, Madras: Adyar Library and Research Centre. (Although focused on Bhāskara, this well-written book traces the development of much classical theistic Vedānta; comparisons with Western thought are drawn at the end.)

Thibaut, G. (trans.) (1890) *The Vedānta Sūtras of Bādarāyaṇa, with the Commentary by Śaōkara*, New York: Dover.(An excellent, highly readable translation of Śaōkara's principal philosophic work.)

Thibaut, G. (trans.) (1904) *The Vedānta Sūtras of Bādarāyaṇa, with the Commentary by Rāmānuja*, New York: Dover.(An excellent, highly readable translation of Rāmānuja's principal philosophic work.)

Upaniṣads (800-300 BC), trans. P. Olivelle, *Upaniṣads*, Oxford: Oxford University Press, 1996.(A new translation that may well be the best; readable and accurate.)

Vallooran, A.D. (1988) *In Search of the Absolute*, Shillong: Vendrame Institute.(A critical study of the Advaita philosophy of T.M.P. Mahadevan.)

Venn, John (1834-1923)

John Venn was a British symbolic logician and methodologist of science. He is known for having invented the method of Venn diagrams for judging the validity of categorical syllogisms and for advocating the 'compartmental' conception of categorical propositions which they display. He strongly defended Boole's algebraic methods in logic by giving them clear logical meanings. He provided the first systematic formulation of the frequency theory of probability, and he showed the uncertainties inherent in the use of J.S. Mill's inductive methods.

Venn was born in Drypool, Hull. A member of Gonville and Caius College, he graduated from Cambridge in 1857. He became college lecturer in moral science at Cambridge in 1862 and worked for thirty years on problems in symbolic logic and the methodology of science. He turned later to biographical and antiquarian topics related to his family and college, of which he was President from 1903 until his death.

Venn is best known as the creator of the method of Venn diagrams for testing categorical syllogisms (see Logic machines and diagrams §3). Two features of Venn's method are especially important. The first is that it is not limited to three classes. He was, in fact, remarkably adept at using Venn diagrams to test arguments involving four or even five class terms. The second feature is that these diagrams graphically represent Venn's analysis of categorical propositions. Venn detected three types of analysis of categorical propositions in the logical tradition: the predicational view ('all *X* are *Y*'), the class inclusional and exclusional view ('the class of *X*s is included in the class of *Y*s'), and the compartmental or existential view ('the class of things that are *X* and non-*Y* is empty'). Venn adopted the latter conception, and this is precisely what is shown by Venn diagrams. His use of the compartmental view was based on its convenience rather than its ultimate truth. While a 'complete Theory of Logic' would require that one give the '*true* account', his own goal was to find the most convenient logical methods.

Venn's most important work, *Symbolic Logic* (1881), was a spirited defence of Boole's algebraic techniques for logic in the face of the modifications which Jevons, Peirce and others had made in the preceding decades; (see Boole, G.; Boolean algebra). Of Boole's use of inverse operations - subtraction as the inverse of class union and division as the inverse of class intersection - the latter caused most controversy. Boole had claimed that the use of division gave rise to uninterpretable formulas, but thought this acceptable as long as the end results of logical calculations were interpretable. Jevons and others, though, claimed that such a mysterious sign should be dropped from logic entirely. Venn strongly defended its retention, holding that by dropping it, 'nearly everything which is most characteristic and attractive in the system is thrown away' (1881: xxviii).

Venn's project, then, was to reinstate Boole's original techniques by allowing division. Contrary to Boole, though, he believed that the formulas which resulted from using division could be given clear logical meanings. On his view, 'x/y' stands for any class which is such that the intersection of x/y and *y* is identical with *x*. This will obtain only when all *x* is *y* (that is, $x = xy$). Furthermore, there will usually be many classes which have this characteristic. Thus 'x/y' stands for a one-many partial function of *x* and *y*. This explanation allowed Venn to give a logical explanation of many of the otherwise puzzling formulas in Boole's logic.

Venn was also concerned with the logic of science. In *The Logic of Chance* (1866), he rejected the commonly held view that probability theory deals with degrees ('gradations') of belief. He presented, instead, the first systematic formulation of the frequency theory of probability, in which probability assertions are empirical assertions about the frequency with which types of events will happen in the long run. Venn realized that the relationship between probability in this sense and other uses of the word 'probability' remained puzzling. The probability of single events seems undefined, as is the probability that an empirical frequency may give to a hypothesis concerning a long-run frequency (see Probability theory and epistemology).

Venn presented a systematic analysis of scientific method in *The Principles of Empirical or Inductive Logic* (1889). While heavily influenced by Mill, he found Mill's inductive methods to be uncertain (see Mill, J.S. §5). Applying Mill's methods to scientific problems requires far more knowledge than scientists normally have. This criticism reflects the strongly empiricist, even sceptical, mode of thought which characterized Venn's work.

See also: Logic in the 19th century §3

DANIEL D. MERRILL

List of works

Venn, J. (1866) *The Logic of Chance*, London: Macmillan, 3rd edn, 1888.(The first systematic formulation of the frequency theory of probability.)

Venn, J. (1881) *Symbolic Logic*, London: Macmillan, 2nd edn, 1894; repr. New York: Chelsea Publishing Company, 1971.(A survey of the symbolic logic of Venn’s time and defence of Boole’s original notation and methods.)

Venn, J. (1889) *The Principles of Empirical or Inductive Logic*, London: Macmillan, 2nd edn, 1907.(A systematic analysis of scientific method, based upon, yet critical of, Mill’s inductive methods.)

References and further reading

Copi, I. and Cohen, C. (1994) *Introduction to Logic*, New York: Macmillan, 9th edn, 235-42, 251-61.(An introduction to Venn diagrams in a standard textbook.)

Keynes, J.M. (1921) *A Treatise on Probability*, London: Macmillan, 92-110.(An incisive critique of Venn’s frequency theory.)

Passmore, J.A. (1957) *A Hundred Years of Philosophy*, London: Duckworth, 134-7. (A brief discussion of Venn’s work.)

Vernia, Nicoletto (d. 1499)

Nicoletto Vernia was a celebrated Aristotelian philosopher during the second half of the fifteenth century. His acquaintances included such personalities as Ermolao Barbaro, Giovanni Pico della Mirandola, Pietro Pomponazzi and Agostino Nifo. His special interests were in natural philosophy and psychology, but he also revealed interests in logic. Although usually characterized as a rigid Averroist, he moved from a clear commitment to Averroes as the true interpreter of Aristotle to a preference for the Greek commentators, especially Themistius and Simplicius. Nonetheless, throughout his career he also maintained a noteworthy interest in Albert the Great. After first attempting to conciliate Albert with Averroes as much as possible, he later attempted to conciliate Albert with the Greek commentators. He was one of the first Renaissance Aristotelians to use the commentary on Aristotle's On the Soul that is attributed to Simplicius, and also to cite Plato, Plotinus and their translator and expositor, Marsilio Ficino.

1 Life and works

Nicoletto Vernia, a native of Chieti, studied at Padua with Cajetan of Thiene (Gaetano da Thiene) and Paul of Pergola. Around 1467-8 he spent a year at Pavia studying the *Calculationes* of Richard Swineshead (see Oxford Calculators), most likely with Giovanni Marliani. After his return to Padua he took the chair of Cajetan of Thiene in 1468. In his early writings he consistently states that Averroes (see Ibn Rushd) gives the true interpretation of Aristotle. What is striking is that he regularly attempts to reconcile Averroes and Albert the Great. In 1489 Pietro Barozzi, Bishop of Padua, issued a decree which forbade lecturing on the unity of the intellect according to Averroes and accused Vernia of having filled Italy with this error. Although Vernia submitted to the decree, it appears from the marginal annotations dating from 1487 and 1489 in his copies of Aristotle and John of Jandun that he already had doubts about the philosophical viability of the doctrine of the unity of the intellect, and had begun to study the Greek commentators on Aristotle, especially Alexander of Aphrodisias, Themistius and Simplicius (see Aristotle Commentators). That interest in the Commentators, which may reflect the influence of Ermolao Barbaro and Giovanni Pico della Mirandola, is obvious in his posthumously published questions against the 'perverse doctrine' of the unity of the intellect (*Contra perversam Averrois opinionem*) (dated 1492, but published posthumously in 1504). In that work, Vernia uses Marsilio Ficino and reconciles Albert the Great with Themistius and Simplicius. Basing himself in particular on Simplicius, he proposes that Plato and Aristotle hold the same doctrine, and differ only in words. Throughout his life Vernia seems to have had an irascible side that on occasion led to the expression of hostility towards others.

2 Logical issues

Vernia was one of the Renaissance Aristotelians who discussed the nature of a so-called most powerful, or absolute, demonstration (*demonstratio potissima*) and the sources of its premises. Such a demonstration was considered to provide perfect causal knowledge of some effect. Vernia explains that after resolution of the information provided by sense experience into a preliminary and imperfect universal knowledge there is a movement of the intellect (*negotiatio intellectus*) that seeks the cause of the effect being studied and that establishes the premises that produce the desired demonstrative or apodictic knowledge (see Zabarella, J. §5).

In his treatise on the division of the sciences (1482a), Vernia states that while the dialectical syllogism concerns the probable, the probable includes necessary things as well as contingent things. Even natural philosophers, mathematicians and metaphysicians can be considered to study the probable. In his question on whether medicine is more excellent than civil law (1482b), he maintains that medicine, subalternated to natural philosophy, demonstrates and knows that it is doing so, whereas law rests on the authority of the learned and is subalternated to politics, a science lower than natural philosophy. At best law achieves the dialectical habit of mind that prepares for demonstration.

3 Natural philosophy

In a question on the subject of natural philosophy (1480), Vernia pays Thomas Aquinas great respect but argues against the view of the Angelic Doctor that mobile being (*ens mobile*), which might include angels, is the subject of natural philosophy. He takes the true interpretation of Aristotle to be that of Averroes, whom he presents as

holding mobile body (*corpus mobile*) to be the subject of natural science (*scientia naturalis*). He attacked such earlier writers as John Canonicus and Antonius Andreas for not speaking in a natural fashion when they appeal to the motion of angels in order to discredit Aristotle's natural philosophy (see Cajetan §3).

In his early *De gravibus et levibus (On Things Heavy and Light)* (before 1476), Vernia examines the problem whether heavy and light inanimate things move locally by themselves or by virtue of something else, once the factor preventing (*prohibens*) motion is removed. He argues that Averroes' position on the question is the correct reading of Aristotle. He emphatically rejects as 'theologizers' (*theologizantes*) those who maintain that the quality of heaviness (*gravitas*) moves objects as a primary and principal agent, arguing from transubstantiation in the Eucharist that the heaviness of the consecrated host moves downwards even though that heaviness is separated from the substance of the bread. In like fashion, he rejects all appeals to God's absolute power (*potentia absoluta*) when discussing questions of natural philosophy.

A particular target of his attack is the position that explains the motion of a projectile not by air pushing behind it but rather by an impetus (*impetus*) imparted to the projectile. Vernia considers this view to deviate from Aristotle and truth. In contrast, he attempts to reconcile Averroes and Albert the Great on the topic, maintaining that when two such excellent philosophers agree, no one of sound mind should doubt that this is the mind of Aristotle.

4 Psychology

In an early work that he calls the *Question on the Unity of the Intellect* but never published, Vernia discusses whether the human soul is eternal and one in number for all humans or whether it is united in existence to the human body as a true substantial form. His announced purpose is threefold: (1) to set forth the doctrine of Plato; (2) to present what Aristotle held according to his most famous Greek and Arab commentators; and (3) to examine the Latin commentators and the truth of the Catholic faith. His brief summary of Plato's doctrine of the soul is culled from Albert the Great and Calcidius' translation of *Timaeus* 41 A-D (see Plato §§4, 13). In the second part, he presents and defends Averroes' doctrine of the unity of the intellect (namely that all human beings share one intellect) as the true interpretation of Aristotle. In the third part he attacks Albert, Aquinas and Duns Scotus for teaching that there are many intellective souls which are created by God, since Aristotle did not admit creation. To attribute to Aristotle the thesis of creation would be to hold that he maintained an infinite number of intellective souls infused in an infinite number of bodies, but this is false. Moreover, he rejects the accounts of individuation held by Aquinas and Duns Scotus, namely individuation through the soul's relation to a body and individuation through 'thisness' (*haecceitas*). He also characterizes Aquinas' and Duns Scotus' insistence that the human being cannot have a direct vision of God during this life as false according to Aristotle.

Subsequently Vernia changed his mind completely and wrote questions, dated 1492 but published posthumously in 1504, against Averroes' doctrine of the unity of the intellect. In the first part he again relies on Albert the Great when presenting Plato. But he now quotes Plato's *Timaeus* 41 A-D in the translation of Marsilio Ficino. In the second part, he rejects Averroes' doctrine of the unity of the intellect, basing himself on what he takes to be the common doctrine of the Greek commentators, namely Theophrastus, Alexander of Aphrodisias, Themistius and Simplicius. His misreading of Alexander as a proponent of the immortality of the soul is also to be found in Giovanni Pico della Mirandola, and was later strenuously attacked by his former student, Agostino Nifo.

Vernia maintained that Aristotle held the intellective soul to be multiplied according to the number of bodies. He even claims, citing Simplicius, that any disagreement between Plato and Aristotle is only verbal, since both maintain that the intellective soul was created by God from all eternity and therefore pre-exists the body. Only subsequently is the soul infused into a body as its substantial form; at death it will return to the dwelling place of a star. Vernia attributes to Aristotle Plato's doctrine that knowledge is recollection from a prior existence. Vernia's psychological thought thus evolved from a strict adherence to Averroes to a commitment to the Greek Commentators, and to an interest in Plato and Plotinus as known through the translations and comments of Ficino.

See also: Aristotelianism, Renaissance; Averroism; Natural philosophy, medieval

EDWARD P. MAHONEY

List of works

Vernia, N. (before 1476) *Quaestio de gravibus et levibus (Question on Things Heavy and Light)*, in Cajetan of

Thiene, *Expositio in libros Aristotelis De coelo et mundo*, ed. N. Vernia, Padua: Magister Bonus.(Vernia considers the question whether inanimate objects move locally by themselves or due to something else to be the most difficult question in physics. An impetus given to air and not to the body is the cause of the body's motion.)

Vernia, N. (1480 or earlier) *Utrum anima intellectiva humano corpore unita tanquam vera forma substantialis dans ei esse specificum substantiale eterna atque unica sit in omnibus hominibus (Whether the Intellective Soul is United to the Human Body as a True Substantial Form or is Eternal and One in All Humans)*.(Vernia's *Question on the Unity of the Intellect*; available in a single manuscript held at the Biblioteca Nazionale San Marco (Marciana) in Venice (see Pagallo 1964). Averroes' doctrine of the unity of the intellect is presented as the true mind of Aristotle. Albert the Great, Aquinas and Duns Scotus are criticized for attempting to prove that each human being has an individual intellective soul.)

Vernia, N. (1480) *Quaestio an ens mobile sit totius naturalis philosophiae subiectum (Question whether Mobile Being is the Subject of All Natural Philosophy)*, in Giles of Rome and Marsilius of Inghen, commentaries on Aristotle's *On Generation and Corruption*, ed. N. Vernia, Padua: Joannes Herbort.(Completed at Padua in 1480. Averroes' position, that mobile body is the subject of natural science, is the true interpretation of Aristotle.)

Vernia, N. (1482 or earlier) *Utrum sint ponendae rationes seminales in materia respectu rerum quae ex ipsa generantur (Whether Seminal Reasons are to be Posited in Matter with Respect to the Things Generated from It)*, in E.P. Mahoney, 'Nicoletto Vernia's Question on Seminal Reasons', *Franciscan Studies* 38 (1978): 299-309.(Vernia presents and then rebuts twenty arguments against seminal reasons.)

Vernia, N. (1482a) *De divisione philosophiae (On the Division of Philosophy)*, in W. Burley, *Expositio in libros octo de physico auditu*, ed. N. Vernia, Venice: Joannes Herbort, 1482.(Vernia shows special interest in Themistius in the translation of Ermolao Barbaro.)

Vernia, N. (1482b) *Quaestio an medicina nobilior atque praestantior sit iure civili (Question whether Medicine is More Noble and Distinguished than Civil Law)*, in E. Garin (ed.) *La disputa delle arti nel Quattrocento*, Florence: Vallecchi, 1947, 111-23.(Completed in 1482 and first published with the *De divisione philosophiae*. Medicine as theory or science consciously uses demonstrative reasoning and is subalternated to natural philosophy; law rests on authority and is subalternated to politics. Medicine is therefore superior.)

Vernia, N. (1492) *Contra perversam Averrois opinionem de unitate intellectus et de animae felicitate quaestiones divinae (Divine Questions Against the Perverse Opinion of Averroes on the Unity of the Intellect and the Felicity of the Soul)*, in Albert of Saxony, *Acutissimae quaestiones super libros de physica auscultatione*, Venice: Alexander Calcedonius and Jacobus Pentius, 1504. (Vernia now takes Averroes to have erred regarding Aristotle on the soul and intellect. The Greek commentators disagree with Averroes on the unity of the intellect. Philosophical arguments can be given for the creation of individual souls by God. Plato and Aristotle can be reconciled.)

References and further reading

Kessler, E. (1994) 'Nicoletto Vernia oder die Rettung eines Averroisten' (Nicoletto Vernia or the Deliverance of an Averroist), in F. Niewöhner and L. Sturlese (eds) *Averroismus im Mittelalter und in der Renaissance*, Zurich: Spur, 269-90.(Accepts that Vernia's abandonment of Averroes' doctrine of the unity of the intellect resulted from studying the Greek commentators and not from fear of Church authorities. Underscores influence of humanism.)

Mahoney, E.P. (1968) 'Nicoletto Vernia and Agostino Nifo on Alexander of Aphrodisias: An Unnoticed Dispute', *Rivista critica di storia della filosofia* 20: 268-96.(Examines in detail Vernia's argument that Alexander taught the immortality of the soul and Nifo's critique of this argument.)

Mahoney, E.P. (1976) 'Nicoletto Vernia on the Soul and Immortality', in E.P. Mahoney (ed.) *Philosophy and Humanism: Renaissance Essays in Honor of Paul Oskar Kristeller*, New York: Columbia University Press and Leiden: Brill, 144-63.(Examines Vernia's early treatise in favour of Averroes' doctrine of the unity of the intellect and his later work attacking that doctrine.)

Mahoney, E.P. (1983) 'Philosophy and Science in Nicoletto Vernia and Agostino Nifo', in A. Poppi (ed.) *Scienza e filosofia all'Università di Padova nel quattrocento*, Contributi alla storia dell'Università di Padova 15, Padua: Lint, 135-202.(Contains details on Vernia's various treatises, their dating and their contents. Impact of the Greek commentators delineated.)

Mahoney, E.P. (1986) 'Marsilio Ficino's Influence on Nicoletto Vernia, Agostino Nifo and Marcantonio Zimara', in G.C. Garfagnini (ed.) *Marsilio Ficino e il ritorno di Platone: Studi e documenti*, vol. 2, Florence: Olschki, 509-31.(Shows how Vernia came to use Ficino and his translations in his late thought.)

Mahoney, E.P. (1991) ‘Nicoletto Vernia’s Annotations on John of Jandun’s *De anima*’, in B. Mojsisch and O. Pluta (eds) *Historia philosophiae medii aevi: Studien zur Geschichte der Philosophie des Mittelalters*, Amsterdam: Grüner, 573-93.(Discusses Vernia’s annotations in his copies of Aristotle and Jandun’s *De anima* that date from the late 1480s and reveal his rejection of the unity of the intellect and his preference for the Greek commentators.)

Nardi, B. (1958) *Saggi sull’aristotelismo padovano dal secolo XIV al XVI (Essays on Paduan Aristotelianism from the 14th to the 16th Century)*, Università degli Studi di Padova: Studi sulla tradizione aristotelica nel Veneto 1, Florence: Sansoni, 95-126, 153-65, 284-7, 323-4.(Stresses the less admirable side of Vernia’s personality but gives some attention to his career and his works and thought.)

Pagallo, G.F. (1964) ‘Sull’autore (Nicoletto Vernia?) di un’anonima e inedita quaestio sull’anima del secolo XV [Venezia, Bibl. Naz., Lat. VI, 105 (=2656)]’ (On the Author of an Anonymous and Unedited Question on the Soul of the 15th Century), in *La filosofia della natura nel medioevo: Atti del Terzo Congresso Internazionale di Filosofia Medioevale*, Milan: Società Editrice Vita e Pensiero, 670-82.(Establishes Vernia’s authorship of an anonymous treatise that presents a strongly Averroist interpretation of Aristotle on the soul.)

Pagallo, G.F. (1983) ‘Di un’inedita "Expositio" di Nicoletto Vernia "In posteriorum librum priorem"’ (On an Unedited ’Expositio’ of Nicoletto Vernia ’On the Prior Book of the *Posterior Analytics*’), in L. Olivieri (ed.) *Aristotelismo veneto e scienza moderna: Atti del 25º anno accademico del Centro per la Storia della Tradizione Aristotelica nel Veneto*, vol. 2, Padua: Antenore.(Important analysis of unpublished lectures by Vernia on Aristotle’s *Posterior Analytics*. Expresses high regard for Themistius and conciliates Plato and Aristotle.)

Ragnisco, P. (1891) *Nicoletto Vernia: Studi storici sulla filosofia padovana nella 2a metà del secolo decimoquinto (Historical Studies on Paduan Philosophy in the Second Half of the 15th Century),* Atti del Reale Istituto Veneto di scienze, lettere ed arti, 7th series, vol. 2.(Only general work on Vernia, it remains a useful study.)

Vasoli, C. (1968) *Studi sulla cultura del Rinascimento (Studies on Renaissance Culture),* Manduria: Lacaita Editore, 241-56.(Discusses Vernia’s conception of nature and physical science.)

Vico, Giambattista (1668-1744)

Vico lived in a period in which the successes of the natural sciences were frequently attributed to the Cartesian method of a priori demonstration. His own first interest, however, was in the cultivation of the humanist values of wisdom and prudence, to which this method was irrelevant. Initially, therefore, he sought a methodology for these values in the techniques of persuasion and argument used in political and legal oratory. But he soon came to believe that the Cartesian method was too limited to explain even the advances in the natural sciences and developed an alternative constructivist theory of knowledge by which to establish the degree of certainty of the different sciences. Wisdom and prudence, however, came low on this scale.

Through certain historical studies in law, he became convinced that, although there were no eternal and universal standards underlying law at all times and places, the law appropriate to any specific historical age was dependent upon an underlying developmental pattern of social consciousness and institutions common to all nations except the Jews after the Fall. His New Science (1725, 1730 and 1744) was a highly original attempt to establish this pattern, originating in a primeval mythic consciousness and concluding in a fully rational, but ultimately corrupt, consciousness. He believed that knowledge of the pattern would enable us to interpret a wide range of historical evidence to provide continuous and coherent accounts of the histories of all actual gentile nations. The primacy of consciousness in the pattern led him to claim that there must be a necessary sequence of ideas upon which institutions rested, which would provide the key to the historical interpretation of meaning in all the different gentile languages. He supported this conception by extensive comparative anthropological, linguistic and historical enquiries, resulting most famously in his interpretation of the Homeric poems. He also advanced a more developed account of his earlier theory of knowledge, in which the work of philosopher and historian were mutually necessary, to show how this conception of 'scientific history' was to be achieved.

Vico believed that the knowledge that wisdom and prudence vary in different historical ages in accordance with an underlying pattern could provide us with a higher insight into those of our own age and enable us to avoid a collapse into barbarism which, in an over rational age in which religious belief must decline, was more or less inevitable. Unfortunately, the metaphysical status of his pattern rendered this impossible. Much of his thought was expressed in a context of theological assumptions which conflict with important aspects of his work. This has given rise to continuous controversy over his personal and theoretical commitment to these assumptions. Despite this, however, his conceptions of the historical development of societies, of the relation between ideas and institutions, of social anthropology, comparative linguistics and of the philosophical and methodological aspects of historical enquiry in general, remain profoundly fruitful.

1 Life

Vico was born in Naples. His erratic early education included grammar, the Latin classics and philosophy. In 1684 he began to study for a practising career in law but he abandoned this in 1686 when he became tutor to the children of a wealthy family at Vatolla near Salerno. The years 1686-95 represent his only prolonged absence from Naples. In his *Autobiography* (1725, 1728, 1731) this is presented as a period of isolation from the lively intellectual life of Naples whereas, in fact, he remained in contact with it through membership of certain private societies and salons which had arisen to counteract the conservative influence of the Church and university. Through an intensive, self-directed course of study in the excellent library at Vatolla, he became acquainted with the thought of Plato and the Neoplatonists, the classical atomists such as Democritus and Lucretius, physicists such as Galileo and Gassendi, English thinkers such as Francis Bacon, Hobbes, Locke and Boyle and the rationalists, Descartes, Malebranche, Spinoza and Leibniz. By the end of this period he had acquired an encyclopedic, if idiosyncratic, understanding of the worlds of ancient and modern learning. The activities of the Inquisition make it difficult to be certain of his religious sentiments. Many of his friends were suspected of being atheists and it is sometimes thought, on the evidence of a Lucretian poem written in 1692, that he may have been an atheist at this time.

In 1694 he took his doctorate in law at Salerno and in 1699 he was appointed to the poorly paid Chair of Rhetoric at the University of Naples. Married in the same year, he subsequently had eight children and was constantly beset with financial problems which forced him to augment his income by the composition of official orations and histories. In 1723 he failed to gain the prestigious Chair of Civil Law. Abandoning hope of personal advancement,

he turned to the construction of the philosophy for which he is most famous, developing it progressively in the various versions of his *New Science* (1725 onwards). Much of our knowledge of his intense intellectual life is derived from the first and second parts of his *Autobiography*, written soon after the production of the first and second editions of the *New Science*. A third part, going up to his death, was later added by the Marquis of Villarosa. He remained a leading figure in Naples for the rest of his life, retiring from his Chair only three years before his death.

2 Early thought

Vico's primary aim was to show how philosophy could contribute to an understanding of the nature of wisdom and prudence in all their manifestations, personal, civil and political. This is evident from his earliest thought, expressed in *Six Inaugural Orations* given between 1699 and 1707, in which he argued that wisdom and prudence could be understood only through a study of the entire world of learning, human and divine. To justify the possibility of success he drew upon a Neoplatonic conception of the human mind as the image of God (see Neoplatonism). Just as God knows the world of his creation because he is active throughout that world, mind is the active incorporeal component in the cognitive, emotional and sensitive abilities required for the acquisition of knowledge. He supported the theological assumptions of this position by a series of Cartesian arguments: the *Cogito* and the causal and ontological proofs of God's existence (see Descartes, R. §§5, 6). But his conclusion, that we have been given all that the discovery of truth and virtue requires, rendered problematic why success should prove so elusive. In the sixth of his *Six Inaugural Orations* he therefore introduced a historical theory, drawn from the Bible, emphasizing our moral and intellectual corruption after the Fall. Dispersed throughout the world, the gentile nations have lost the languages, beliefs and inclinations necessary to reach the truth. What is needed, therefore, is a study of the world of human learning, ancient and modern, in order to establish the correct method to regain it. This is to be achieved by the development of wisdom, eloquence and prudence through an educational programme which respects a natural order of psychological development.

The metaphysical theory of mind played an important part in Vico's next two works. The first, *On the Study Methods of Our Time* (1709), clearly influenced by Francis Bacon, was an ambitious attempt to carry out the programme of the sixth of his *Six Inaugural Orations*. Vico's conclusion is that while there have been major advances in modern geometry, mechanics and physics, not all of these have the certainty claimed for them. They have, however, been influential in a loss of interest in the methods of education required for the successful conduct of social and political life because their advance has mistakenly been attributed to the Cartesian demonstrative method (see Descartes, R.). But the only field of knowledge in which demonstration can produce certainty is geometry, the elements of which are human constructions. This is the first indication of Vico's constructivist theory of knowledge: that we can know fully only what we have made (see Constructivism). Accordingly, although geometrical truths are required for the construction of physical theory, physics cannot achieve similar certainty because of its indispensable need for empirical experiment. At best it can claim verisimilitude. Demonstration is even less appropriate for the wisdom and prudence required in social and political life, because of the ineliminable influence of choice and chance. What is required here is a method for the development of sound judgment in human affairs. This existed in the classical art of topics, which was a body of categories and argumentative procedures used by lawyers to reach judgments in matters of fact and right in everyday life. But since this art is beyond the reach of the masses, who are ruled more by emotion than judgment, politicians must also master the ancient art of eloquence so that they can persuade the masses to adopt the practices recommended by the judgments reached through topics.

3 Early thought (continued)

Vico's next work, *On the Most Ancient Wisdom of the Italians Unearthed from the Origins of the Latin Language* (1710), was intended as a three-part treatise on metaphysics, physics and ethics. Only the first part was completed and published but the constructivist elements in his thought were now made quite explicit. The metaphysical thesis remains in the conception of a God who knows the metaphysical Forms on which the physical world is modelled because he creates them. We, made in the image of God, can also know what we create. This is expressed in Vico's *verum-factum* theory: the true is convertible with the made. The true is thus relativized to the mode of mental construction. To this, however, Vico now added the Aristotelian view that knowledge is of causes, which he treated as meaning that knowledge is possible only where we are responsible for the whole construction,

including its most primitive elements. Since God contains eminently the elements of everything within himself, what he knows is both true and real. As part of God's world, we cannot attain such completeness. Nevertheless, through our ability to abstract from God's creation, we can create such fictive elements as mathematical points and lines and from these construct theorems that are true but fictive. Beyond mathematics, however, the sciences become less certain as their subject matter becomes less amenable to the application of mathematically based theory. Hence, mathematics remains the most certain science, followed by mechanics and physics. Human behaviour, governed by caprice and emotion, remains the least certain of all. Here again, the art of topics is required.

This constructivist theory enabled Vico to reject his earlier Cartesian arguments. We cannot start from *Cogito* since, not knowing how mind is made, we can have certainty - that is, the mere inability to doubt - of our existence, but not full causal knowledge of our mode of existence. Nor can we demonstrate the existence of God since any such demonstration would require us to create him. Possibly realizing the difficulty which this claim created for knowledge of his metaphysical theory of the human mind, Vico left knowledge of God to faith and revelation.

The transition to his mature philosophy was occasioned by a close study of Grotius' *The Law of War and Peace* in 1717 (see Grotius). Grotius' purpose was to justify a system of international law by arguing philosophically and historically that it was an extension of a rationality existent universally in the natural law of nations. This gave Vico an insight into a way in which philosophy, which is concerned with the universal, could contribute to an understanding of the nature of the human world. But he advanced two general criticisms of Grotius. First, he had failed to recognize the metaphysical importance of God and the historical importance of the Fall. Second, his historical treatment of law lacked sound principles of historical interpretation, leading to anachronistic accounts of ancient law which assimilated it to that of his own civilized times. Criticism of the historical accounts of other political theorists, particularly Hobbes, Pufendorf and Selden, confirmed Vico's belief that a sound method was required to show how a rational system of law could have developed historically from the Lucretian-style brutes with whom, it was generally agreed, history began (see Lucretius).

His first attempt at such a system came in two large works, *On the One Beginning and One End of Universal Law* (1720) and *On the Constancy of the Jurisprudent* (1721), the second of which was divided into a philosophical and a philological or historical part. The metaphysical theory of human nature still remained, now as the basis for a deduction of the virtues of prudence, temperance and courage and, thence, the authority of law over property, liberty and wardship. The theory of the Fall was also retained to support the idea that the historical state of law (the 'certain') is never purely contingent and unjustifiable because it incorporates a partial and developing degree of rationality left within us after the Fall, the end product of which will be the full development of the concepts of truth and equity.

Vico first tried to give these claims historical support by showing that the history of Roman law could be read in this way, but he soon realized that their universal character could be demonstrated historically only by showing that they applied to the histories of all nations. His attempt to do this, however, in *On the Constancy of the Jurisprudent*, revealed to him that the law of a nation could not be treated in isolation from other fundamental aspects of its life. Hence, in the various versions of the *New Science*, he engaged in the enormous task of showing that there was a common pattern to the development of all the major facets of the histories of the nations. The one exception which he carefully allowed was Jewish history, in order to respect the Biblical account of the Fall.

4 The *New Science*: structure and contents

The *New Science* exists in three editions, written in 1725, 1730 and 1744. The *First New Science* (1725) and *Second New Science* (1730), as they are called, are structurally very different. The *Third New Science* (1744) was an attempt to incorporate into the text of the *Second New Science* many additions and improvements developed after its publication. Since it would be impossible here to treat each separately, what follows is an overview of Vico's main doctrines.

The *New Science* offers an account both of the philosophical foundations, ontological and epistemological, required for an understanding of history and partial demonstrations of certain actual histories. The central ontological claim is that all (gentile) nations share a common nature which is exhibited to a greater or lesser degree

in their actual histories in so far as they are affected by different external contingencies. This nature has two fundamental features. First, it is essentially sociable, expressing itself in a holistic network of cultural and institutional creations. Second, it is essentially developmental. The seeds of justice and truth left within the gentile nations after the Fall are the causes of a development of culture and society, as we pass from an original, nonrational, mythic consciousness to a rational consciousness which is the fulfilment of human nature. Thereafter, however, civilization descends into a state of intellectual and moral depravity, the 'barbarism of reflection', from which it can be saved only by a return to something like the conditions after the Fall and a recurrence of the whole sequence. In its unactualized form this sequence is an 'ideal eternal history', tracing the career of a nation were it governed only by its internal nature. But since the actual histories of nations are affected by many contingent factors, involving physical causes such as geographical conditions, disease, famine or war, there will be no perfect exemplifications of the 'ideal eternal history'. All actual histories will nevertheless be recognizable variants of it. This conception undoubtedly owes something to the influence of Neoplatonism.

Vico gives two reasons for demonstrating that there is such a universal and necessary pattern. First, that its existence proves that there is a Providence which influences human affairs and, hence, that there is a God in whom we should believe. Second, that since it will end in a state of anarchy, an understanding of its underlying causes will enable us to avert ultimate disaster.

5 The *New Science*: structure and contents (continued)

The sequence offered involves three stages of cultural, social and political development, the 'poetic' or 'theological', the 'heroic' and the 'human' ages. The main structure of this process follows the natural order of psychological development worked out in Vico's earlier writings, now, however, taken as psycho-social principles of development. They issue in forms of a 'common sense', or sets of fundamental beliefs about the world, shared by classes within the nation or by the whole nation, which govern the culture and institutions of their age.

The poetic age is the product of the most primitive kind of human being, a brutish, egoistic creature, lacking reason but endowed with huge ideographic imaginative powers, through which is created the first primeval image of the world as a vast animate being, or God. Fear of this God and his children, who successively come into existence as imagistic ways of conceiving new social necessities, dominates every aspect of the culture and institutions of the age, giving rise to the practices of sacrifice and divination and thence to the political supremacy of the priestly classes. As mediators of the divine will, the latter acquire a semi-divine status through which they become the first kings, while law comes into existence as the means of protecting God's mediated ownership. The basis of all this, however, is our primitive capacity to create through imagination and to believe that which is created as literal truth. This is Vico's theory of myth as *vera narratio*: that what has come down to later ages as imaginative fiction was, in its origin, taken as literal truth. Accordingly, the key to understanding the poetic age depends upon taking myths as literal expressions of the way in which the physical and human world was seen and believed. These basic principles also underlie the morality, politics, geography and history of the age. Much of Vico's own account of the character of this age is dependent upon his highly original interpretation of the Homeric poems as later compilations of these original myths.

The 'heroic' age is that in which the kings, through a sequence of traceable steps, progressively lose their unique status. There are two reasons why this occurs. First, because of their corrupt nature, they abuse their powers to the extent that, in order to defend themselves against their subjects, they form political associations with other kings ('aristocratic republics') which weaken their original monarchical status. Second, their increasingly rational subjects realize the falsity of the claim to semi-divine status upon which their privileged political status depends. Thereafter, it is a question of challenging the legal conceptions which their claim to divine origin can no longer justify until, ultimately, a new form of political organization is required, resting upon new conceptions of humanity and justice.

In the 'human' age these developments reach their '*acme*'. The seeds of truth in our original imperfect nature have now developed into a full understanding of the truth: that equity is the supreme and proper principle of law, that human customs should be informed by our sense of civil duty and that governments should be based upon the principle of equality under the law. The human age is not, however, in all ways superior to the earlier ages. One of Vico's fundamental principles is that imagination weakens as reason strengthens. Accordingly, although theologians and philosophers can now prove the truths of (Christian) religion, they cannot implant in the general

public that degree of instinctive belief which arose when God was a product of the primeval human imagination and which once held together whole communities. Thus, the highest stage of human rationality is also the start of a rapid period of moral corruption which will destroy the customs and institutions necessary for the maintenance of social and political order and lead to a new poetic age. Hence, the course of nations will be succeeded by a recourse such as had occurred in feudal Europe, Vico claimed, after the fall of Rome, in which he descried many parallels with the life of early Rome.

Although the ‘ideal eternal history’ outlines the above sequence, in his more detailed explanations of its character Vico makes extensive reference to the operation of certain ‘natural’ principles or ‘necessities of nature’. These are largely certain psychological and psycho-social propensities which belong to us by nature. They are therefore aspects of a ‘metaphysics of the human mind’ which, he claims, is fundamental to his theory. They are used to explain the development of language from the gestures proper to the first poets to the articulated verbal language of the rational age, the parallel developments of ideas from the ideographic images of the poets to the abstract ideas of rational beings and the development of the institutional systems which presuppose them. The fact that the sequence from imagination to reason is itself classified as a ‘natural necessity’ shows that it is not correct, as is sometimes argued, to see the ‘ideal eternal history’, or its exemplifications, as an account of the self-development and subsequent decay of the imagination alone. Despite its undoubted importance in Vico’s conception of human nature, his ‘imagination’ is not the self-transcending imagination of the great Romantics. Rather, it is an imagination which operates according to certain principles of human nature, some of which are most fully explained in *On the Most Ancient Wisdom of the Italians Unearthed from the Origins of the Latin Language*.

This naturalizing tendency in Vico’s account of the basis of the ‘ideal eternal history’ raises vexed questions about the sincerity of his claim to have proved the influence of providence in history and, thereby, the existence of God. Moreover, his distinction between Jewish and gentile history requires a historical interpretation of the Old Testament which is incompatible with his claim to found the principles of true history upon a *universal* metaphysics of the human mind. Unsurprisingly therefore, he often fails to adhere to the distinction, assimilating events in the Old Testament to others in the gentile nations from which they should differ. Consequently, in the face of contemporary suspicions of heresy, he insisted that the sequence traced in the ‘ideal eternal history’ occurs largely through the unintended, rather than the intended, consequences of human action, thus making it possible to attribute the universally patterned nature of their outcome to the influence of Providence. Some interpreters accept this as evidence of the sincerity of his belief in the theological context in which he set his theory. But since he failed to provide any theory to give his claims about Providence explanatory force, it is difficult to believe that the suspicions he sought to allay were unfounded.

6 Theory of historical knowledge

The theory of knowledge which Vico advanced to support his theory of history is an extension of his early *verum-factum* theory: we can know only what we make or do. In his early works, where he lacked a theory of the causes of social and historical change, he had concluded that, because chance and contingency were ineliminable in the world of everyday affairs, knowledge of it was unattainable. In the *New Science*, however, where such theories are advanced, knowledge becomes possible. If human societies develop through the operation of features of human nature which we share, there seems no reason why we should not use our knowledge of them to recreate the past. Hence, Vico could advance his fundamental claim that, since the world of nations has been made by men, its principles are to be ‘rediscovered within the the modifications of our own human mind’. These ‘modifications’ provide components of his ‘metaphysics of the human mind’.

The crux of his later theory of knowledge lies in a mutually supportive relationship between philosophy and philology, the possibility of which he first saw in Grotius. Philosophy contemplates reason and issues in the true; philology is the study of what people have made and done, their particular languages, institutions and deeds, that is, the ‘certain’. When correctly conceived, philosophy should seek formal and substantive principles for the interpretation and explanation of human history, while philology will reach historical truth through its consistent application of these principles. The central idea, therefore, is that philosophy should produce fundamental theories about human nature and the substantive principles which underlie human historical development, culminating in the ‘ideal eternal history’, in the light of which fragments of historical data can be interpreted and integrated into comprehensive and coherent accounts of the actual historical past. Philosophy does not establish its theories a

priori, however, but through its success in finding a set of universal and necessary principles of which all actual histories are an expression. This conception lies behind the widespread comparative enquiries into the history of thought, languages and institutions in general, which abound in the *New Science*.

Once again, however, Vico's claims about Providence are problematic since if, as he consistently asserts, it affects the process of historical development, its workings will be unintelligible to us. The difficulty is not simply that of understanding a history in which the unintended consequences of individual human actions produce similar patterns of development in different nations. This can be explained by Vico's theory of a common sense, arising from the common nature of nations, which ultimately determines shared social decisions. If Providence is seen as nothing more than common sense, indeed, the problem disappears, but so also does the claim to have demonstrated the existence of God. If, on the other hand, Vico's insistence that Providence is something over and above common sense is genuine, as some commentators claim, there will be something at work in the historical process which is not of human construction and which we therefore cannot know. Vico's claims about Providence are therefore incompatible with his theory of knowledge. This lends support to interpretations which are sceptical of his theistic claims.

A different problem derives from Vico's vacillations about the status of the products of philosophy. In the *First New Science* he refers to them as metaphysical truths but does little to establish them. In the later versions, he describes them as hypotheses which, through confirmation, will attain the status of metaphysical truths. The suggestion here is that they originate as hypotheses and gain their entitlement to truth as a result of their interpretative and explanatory fruitfulness for the historian. In this case it might be questioned whether they are metaphysical at all. But it could equally well be argued that there is no way of establishing the truth of a wide-ranging hypothesis about reality other than by testing its capacity to inject coherence into a fundamental aspect of human experience. This seems to be what Vico finally had in mind.

7 Ultimate doubts

Vico was confident that he had produced a science which would enable us to understand the truth of human history but he became much less certain that this would enable us to diagnose the ills of our society or to remedy them. His suggestions for averting final political and social disaster - the maintenance of law by power of arms under the shield of religion, conquest by another nation or a return to conditions similar to those of the poetic age and a recurrence of the whole cycle - are so unconvincing as to seem little more than expressions of his own perplexity. Nor is this difficult to understand. For if, as he claims, imagination weakens as reason strengthens and if effective religious belief as the main social bond depends upon imagination, then, as soon as the age of reason is reached, religion and social cohesion are bound to decline and there will be no basis for the shield of religion or the other social bonds.

These doubts depend, however, upon a gratuitous inconsistency in his thought. The anarchy in which the life of a nation ends presupposes an incompatibility between imagination and reason. But Vico's general theory requires that they be able to coexist and cooperate, one for the creation of systems of belief, the other for their critical assessment. Without both, indeed, the heroic age would never be superseded. For it is only because the lower classes see the falsity of the imaginary claim to divine origin, which is the basis of the nobility's privileged legal and political status, that the latter becomes challengeable. Thus imagination and reason cannot be incompatible in the way implied by Vico's final doubts. Had he realized this he would have seen no need to postulate either the 'barbarism of reflection' or the recurrence of the life cycle of the nation.

8 Influence

The obscurity of Vico's life and thought has made the extent of his influence upon later thinkers difficult to trace. His emphasis upon the internal coherence of the culture and institutions of different historical societies based upon shared modes of thought and feeling has striking similarities with the thought of Hamann, Herder, Hegel and Marx, of whom only Hegel seems almost wholly ignorant of him. His most direct influence in the nineteenth century was upon Michelet, Comte, Coleridge, Croce and Sorel. In this century he has influenced a number of writers, including Berlin, Collingwood, Gadamer, Joyce and MacIntyre. But there is considerable circumstantial evidence to suggest that he may have influenced many others interested in general theories of the development of civilization from Adam Ferguson and Rousseau onwards.

See also: Anthropology, philosophy of; History, philosophy of; Jurisprudence, historical

LEON POMPA

List of works

Vico, G. (1699-1707) *Orationes I-VI*, trans. G.A. Pinton and A.W. Shippe, *On Humanistic Education (Six Inaugural Orations, 1699-1707)*, Ithaca, NY: Cornell University Press, 1993.(With notes by G.G. Visconti.)

Vico, G. (1709) *De nostri temporis studiorum ratione*, trans. E. Gianturco, *On the Study Methods of Our Time*, Ithaca, NY: Cornell University Press, 1990.(With notes by E. Gianturco.)

Vico, G. (1710) *De antiquissima Italorum sapientia ex linguae latinae originibus eruenda*, trans. L.M. Palmer, *On the Most Ancient Wisdom of the Italians Unearthed from the Origins of the Latin Language*, Ithaca, NY: Cornell University Press, 1988.(With notes by L.M. Palmer.)

Vico, G. (1720) *De universi iuris uno principio et fine uno (On the One Beginning and One End of Universal Law)*, in F. Nicolini (ed.) *Opere di G.B. Vico*, vol. 2, Bari: Laterza, 1936.

Vico, G. (1721) *De constantia iurisprudentis (On the Constancy of the Jurisprudent)*, in F. Nicolini (ed.) *Opere di G.B. Vico*, vol. 2, Bari: Laterza, 1936.

Vico, G. (1725) *Principi Di Una Scienza Nuova Intorno Alla Natura Delle Nazioni Per La Quale Si Ritruovano I Principi Di Altro Sistema Del Diritto Naturale Delle Genti (Principles of a New Science of the Nature of Nations Leading to the Discovery of the Natural Law of the Gentes)*, trans. L. Pompa in *Vico: Selected Writings*, Cambridge: Cambridge University Press, 1982.(With notes by L. Pompa.)

Vico, G. (1725, 1728, 1731) *Vita di Giambattista Vico scritta da se medesimo*, trans. M.H. Fisch and T.G. Bergin, *The Autobiography of Giambattista Vico*, Ithaca, NY: Cornell University Press, 1944, 1963.(This includes an invaluable historical introduction and notes on Vico’s influence, by M.H. Fisch and T.G. Bergin.)

Vico, G. (1730) *Cinque Libri Di Giambattista Vico De Principi D’Una Scienza Nuova D’Intorno Alla Comune Natura Delle Nazioni In Questa Seconda Impressione (Five Books by Giambattista Vico on the Principles of a New Science of the Nature of Nations, Second Edition)*, ed. M. Sanna and F. Tessitore, Morano: Naples, 1991.

Vico, G. (1744) *Principi Di Scienza Nuova Di Giambattista Vico D’intorno Alla Commune Natura Delle Nazioni In Questa Terza Impressione*, trans. T.G. Bergin and M.H. Fisch, *The New Science of Giambattista Vico*, Ithaca, NY: Cornell University Press, 1968; repr. including the ‘Practice of the New Science’, Ithaca, NY: Cornell University Press, 1984.(With revisions and notes by T.G. Bergin and M.H. Fisch.)

References and further reading

Adams, H.P. (1935) *The Life and Writings of Giambattista Vico*, London: Allen & Unwin.(An excellent simple historical introduction to the development of Vico’s thought.)

Bedani, G. (1989) *Vico Revisited*, Oxford: Berg.(A detailed monograph stressing the naturalistic character of Vico’s doctrines and the strategies which he employed to obscure their heretical implications.)

Berlin, I. (1976) *Vico and Herder*, London: The Hogarth Press.(An influential interpretation of Vico’s philosophy, emphasizing the ontological and epistemological role of a self-transcending imagination while accepting that Vico nevertheless believed in historical laws.)

Croce, B. (1911) *La Filosofia di G.B. Vico*, trans. R.G. Collingwood, *The Philosophy of Giambattista Vico*, London: Howard Latimer, 1913.(The initiator of modern Vico studies, involving a much disputed identification of Vico’s ‘Providence’ with Hegel’s concept of the cunning of reason.)

Lilla, M. (1993) *G.B. Vico: The Making of an Anti-Modern*, Cambridge, MA: Harvard University Press.(A very detailed historical account of the whole of Vico’s thought, stressing its dependence upon theological and biblical doctrines.)

Nicolini, F. (1978) *Commento Storico alla Seconda Scienza Nuova (A Historical Commentary on the Second New Science)*, Rome: Edizioni di Storia e Letteratura.(The most scholarly explication of the historical references in what is now known as the *Third New Science* of 1744.)

Piovani, P., Giarrizzo, G. and Tessitore, F. (eds) (1972-) *Bollettino del Centro di Studi Vichiani*, Bibliopolis: Naples.(A journal devoted largely to the historical study of Vico’s life, works and influence.)

Pompa, L. (1990) *Vico: A Study of the ‘New Science’*, Cambridge: Cambridge University Press, 2nd edn.(An analytic account of the main doctrines of the *New Science*, taking as primary Vico’s claim to have produced a science.)

Tagliacozzo, G. (ed.) (1981) *Vico: Past and Present*, Atlantic Highlands, NJ: Humanities Press.(A collection of essays concentrating on philosophical and historical issues.)

Tagliacozzo, G., Mooney, M. and Verene, D.P. (eds) (1980) *Vico and Contemporary Thought*, London:

Macmillan.(A collection of articles focusing on Vico's relevance to contemporary thought.)
Tagliacozzo, G. and Verene, D.P. (eds) (1969) *Giambattista Vico's Science of Humanity*, Baltimore, MD: Johns Hopkins University Press.(A collection of historical, comparative and philosophical articles on aspects of Vico's thought.)
Tagliacozzo, G. and Verene, D.P. (1983-) *New Vico Studies*, A tlantic Highlands, NJ: Humanities Press.(An English language journal, dealing with all aspects of Vico studies but with a particular focus on placing Vico's thought in the tradition of rhetoric.)
Tagliacozzo, G. and White, H.V. (eds) *Giambattista Vico: An International Symposium*, Baltimore, MD: Johns Hopkins University Press.(A wide- ranging collection of historical and philosophical articles on aspects of Vico's thought.)
Verene, D.P. (1981) *Vico's Science of Imagination*, Ithaca, NY: Cornell University Press. (A thorough discussion of Vico's concept of the imagination, arguing that it is both ontologically and epistemologically primary and that Vico did not believe in the existence of historical laws.)
Verene, D.P. (1991) *The New Art of Autobiography: An Essay on the Life of Giambattista Vico Written by Himself*, Oxford: Clarendon Press.(An account of Vico's *Autobiography* presenting it as his reconstruction of his life in the light of the main tenets of the *New Science*.)

Vienna Circle

The Vienna Circle was a group of about three dozen thinkers drawn from the natural and social sciences, logic and mathematics who met regularly in Vienna between the wars to discuss philosophy. The work of this group constitutes one of the most important and most influential philosophical achievements of the twentieth century, especially in the development of analytic philosophy and philosophy of science.

The Vienna Circle made its first public appearance in 1929 with the publication of its manifesto, The Scientific Conception of the World: The Vienna Circle (Carnap, Hahn and Neurath 1929). At the centre of this modernist movement was the so-called 'Schlick Circle', a discussion group organized in 1924 by the physics professor Moritz Schlick. Friedrich Waismann, Herbert Feigl, Rudolf Carnap, Hans Hahn, Philipp Frank, Otto Neurath, Viktor Kraft, Karl Menger, Kurt Gödel and Edgar Zilsel belonged to this inner circle. Their meetings in the Boltzmanngasse were also attended by Olga Taussky-Todd, Olga Hahn-Neurath, Felix Kaufmann, Rose Rand, Gustav Bergmann and Richard von Mises, and on some occasions by visitors from abroad such as Hans Reichenbach, Alfred Ayer, Ernest Nagel, Willard Van Orman Quine and Alfred Tarski. This discussion circle was pluralistic and committed to the ideals of the Enlightenment. It was unified by the aim of making philosophy scientific with the help of modern logic on the basis of scientific and everyday experience. At the periphery of the Schlick Circle, and in a more or less strong osmotic contact with it, there were loose discussion groups around Ludwig Wittgenstein, Heinrich Gomperz, Richard von Mises and Karl Popper. In addition the mathematician Karl Menger established in the years 1926-36 an international mathematical colloquium, which was attended by Kurt Gödel, John von Neumann and Alfred Tarski among others.

Thus the years 1924-36 saw the development of an interdisciplinary movement whose purpose was to transform philosophy. Its public profile was provided by the Ernst Mach Society through which members of the Vienna Circle sought to popularize their ideas in the context of programmes for national education in Vienna. The general programme of the movement was reflected in its publications, such as the journal Erkenntnis ('Knowledge', later called The Journal for Unified Science), and the International Encyclopedia of Unified Science. Given this story of intellectual success, the fate of the Vienna Circle was tragic. The Ernst Mach Society was suspended in 1934 for political reasons, Moritz Schlick was murdered in 1936, and around this time many members of the Vienna Circle left Austria for racial and political reasons; thus soon after Schlick's death the Circle disintegrated. As a result of the emigration of so many of its members, however, the characteristic ideas of the Vienna Circle became more and more widely known, especially in Scandinavia, Britain and North America where they contributed to the emergence of modern philosophy of science. In Germany and Austria, however, the philosophical and mathematical scene was characterized by a prolongation of the break that was caused by the emigration of the members of the Vienna Circle.

1 Scientific philosophy and philosophy of science

Proponents of 'scientific philosophy' think of philosophy not as an autonomous discipline prior to the sciences but as a critical discipline dependent upon the natural and social sciences, logic and mathematics. Changing a motto of Kant, they hold that philosophy without science is empty, science without philosophy is blind. Adoption of this scientific conception of philosophy does not, however, determine the details of one's epistemology, methodology and ontology. As far as epistemology is concerned, the Austrian tradition offers the contrasting examples of the phenomenology of Franz Brentano and the positivism of Ernst Mach. Similarly, there are those who stress the unity of the natural and the social sciences and those who contrast explanation in the natural sciences with the distinctive type of understanding (*verstehen*) characteristic of human affairs. Finally, both idealist and materialist ontological positions are compatible with this understanding of philosophy. Nonetheless all proponents of scientific philosophy demand exact methods and an empirical orientation. They oppose irrational and theological systems of philosophy with an attitude that shows their commitment to the ideals of the Enlightenment and to science.

Historically, the positivism of Mach's scientific philosophy was the most important precondition for the development of the position adopted within the Vienna Circle. The term 'philosophy of science' was used to describe this position, but by this was meant a general scientific conception of philosophy as well as a commitment

to providing a philosophy of the sciences. Thus within the Vienna Circle, philosophy was regarded both as a general analytic and language-oriented activity and as a discipline working on the foundations of the natural and social sciences. At the same time we find within the Vienna Circle those such as Moritz Schlick who defend a methodological dualism of philosophy and science, and those such as Otto Neurath who seek to absorb philosophy altogether within a scientific conception of the world. Independent of this variety of positions, however, empiricism, an orientation towards the sciences, and an exact logical-mathematical methodology remain essential features of the Vienna Circle.

2 Logical positivism

The name ‘Vienna Circle’ was used in public for the first time in 1929 in the programmatic essay *The Scientific Conception of the World: The Vienna Circle* (Carnap, Hahn and Neurath 1929). It was suggested by Neurath and was supposed to have pleasant connotations similar to ‘Vienna Woods’ or ‘Viennese Waltz’. At the same time the term should indicate the origin of this philosophical movement and its collective orientation (Frank 1949), although strictly speaking, it is anachronistic to use it for the period before 1929. In this programmatic essay the position of the ‘radical’ wing around Neurath, Carnap, Hahn, Frank and others was especially prominent. This wing, institutionalized in the Ernst Mach Society, supported the idea of a unified physicalist science as represented in the programme of the *International Encyclopedia of Unified Science*. By contrast the more moderate wing of the Vienna Circle around Schlick, Waismann, Feigl and others - in fact the majority - emphasized their adherence to a dualism of science and philosophy with changing names like ‘consistent empiricism’, ‘logical empiricism’, or ‘logical positivism’.

The widely used term ‘logical positivism’ comes in fact from Albert Blumberg’s and Herbert Feigl’s paper ‘Logical Positivism: a New Movement in European Philosophy’, published in the *Journal of Philosophy* in 1931. Blumberg and Feigl give a concise description of the new synthesis of logical and empirical factors:

> The new logical positivism retains the fundamental principle of empiricism but… feels it has attained in most essentials a unified theory of knowledge in which neither logical nor empirical factors are neglected. From the point of view of logical positivism, the Kantian synthesis concedes too much to rationalism by assuming the existence of synthetic a priori truths. Against Kant the new movement maintains as a fundamental thesis that there are no synthetic a priori propositions. … it holds that factual (empirical) propositions though synthetic are a posteriori, and that logical and mathematical propositions though a priori are analytic.
>
> … By means of the theory of knowledge thus constructed, logical positivism … shows that the propositions of metaphysics, in most senses of the term, are, strictly speaking, meaningless.
>
> (Blumberg and Feigl 1931: 282)

Blumberg and Feigl go on to describe the philosophical transformation from old to new positivism with the adoption of symbolic logic, epistemology, and research into the foundations of science. Finally, they explain, following Wittgenstein, their notion of philosophy: ‘The purpose of philosophy is the clarification of the meaning of propositions and the elimination of… meaningless pseudo-propositions’ (Blumberg and Feigl 1931: 269). Despite its widespread currency, however, the term ‘logical positivism’ has the disadvantage that it associates the Vienna Circle too closely with positivism, and thus, for example, with the ‘positivism dispute’ that runs from Lenin to the Frankfurt School. Hence the term ‘logical empiricism’ is now often preferred: it takes into account the synthesis of rationalism and empiricism, and signals clearly the two most important elements in the philosophy of the Vienna Circle.

In Schlick’s logical empiricism the classical philosophical positions of empiricism and rationalism were integrated with the help of modern logic and mathematics, but a distinction between philosophy and science was still admitted. Neurath’s more radical ‘scientific conception of the world’ aimed at overcoming philosophy itself within his scheme for a unified physicalist science. This divergence in philosophical approach left room for debates within the Circle on such topics as the merits of phenomenalist and physicalist languages, coherence and correspondence theories of truth, logical syntax and semantics, verification and confirmation, and ideal and natural languages. At the same time there was a consensus concerning the merits of a logical analysis of language, a fallibilist epistemology, a scientific attitude to the world and the unity of scientific explanation and knowledge in general. After Schlick’s death, however, his logical empiricist project collapsed following personal and theoretical

disagreements. The project of a unified science, however, continued in the unity-of-science movement (see §4).

An important element of the logical empiricism of the Vienna Circle was the refusal to accept synthetic judgments a priori. Following Russell and Whitehead, symbolic logic and mathematics were regarded as purely analytic (because merely 'conventional') and a priori (and thus independent of any experience). Analytic truths of these kinds were contrasted with empirically true statements of the natural sciences and ordinary experience; these were synthetic judgments a posteriori. But there was no further class of synthetic a priori judgments; instead there was thought to be an important class of 'meaningless' sentences. The elements of this class, being neither analytic nor synthetic a posteriori, are 'metaphysical' in a sense which implies that they are not part of knowledge at all even though they may express some realm of experience. This anti-metaphysical position of the Vienna Circle is most prominently represented by Rudolf Carnap's 'Elimination of Metaphysics Through Logical Analysis of Language' (Carnap 1931). It prepares the logical empiricist programme for a unified reconstruction of science. But the question whether an empirical basis could be a foundation for all knowledge received strongly divergent answers from the coherence theorists around Neurath and the correspondence theorists around Schlick (Hempel 1981; 1993). The apparently strict distinction between analytic and synthetic sentences had also already been questioned (Menger 1979; 1994). Indeed, the ideal of one language of science, logic and mathematics had already been strongly relativized by the Vienna Circle itself, long before Quine put forward his classic critique (Quine 1953). Thus, contrary to its popular reputation, a heterogeneous pluralism of views was in fact characteristic of the Vienna Circle: for example in questions of ethics (Schlick 1930; Menger 1934; Kraft 1937), in regard to 'realism' versus 'positivism' (Schlick 1933; Carnap 1928; Feigl 1929; Kraft 1925), verificationism versus falsificationism (Neurath 1935), and not the least in questions of an ideological and political nature.

3 Logical empiricism and the scientific conception of the world

The relationship between Schlick's logical empiricism and Neurath's distinctive scientific conception of the world is a complex matter. Certain points are of course held in common, such as the view of philosophy as a language-oriented, analytic activity. Again, the principle of verification ('The meaning of a sentence can be given only by giving the rule of its verification.') (Schlick 1938: 341; Hempel 1950), logical atomism (following Russell) and the picture theory of language (following Wittgenstein's *Tractatus*), are constitutive features of the entire movement but are in themselves insufficient for its characterization. Kamitz (1973), mainly referring to Carnap's positions, characterizes the Vienna Circle up to about 1930 by the following principles: the formal character of mathematics (Logicism: that is, subordination of mathematics under logic which itself is purely analytic), verifiability, methodological phenomenalism (Carnap's epistemological position in *The Logical Construction of the World*), and 'scientism', which is a claim to the omnipotence of science when compared with alternative forms of knowledge in philosophy and art. For the period 1930-5 the following principles are listed: the hypothetical character of empirical claims (the criterion of verifiability is replaced by a criterion of confirmability), the conventionalist interpretation of logic, physicalism as the foundation of the unified science (a physicalist, quantitative, empirical language as unifying intersubjective language of science), and the conception of philosophy as logical syntax of the language of science. These principles reflect the growing dominance of Neurath's point of view within the Circle. In particular, the last principle restricts questions of truth to comparisons among sentences along the lines of Neurath's coherence theory in order to avoid the dualism of 'language' and 'world' suggested by the correspondence theory of truth. In line with this, Carnap and Neurath deny any absolute 'foundation of knowledge' of the kind sought by Schlick (Schlick 1934). They hold that in any empirical justification it is not single sentences that are tested, but whole systems of sentences, and science in general. This is a form of relativism that makes Neurath in particular a forerunner of recent holistic approaches in philosophy of science (see Hempel, Popper and others in Skirbekk 1977).

Although this last point of view is distinctive of Neurath, it is important to grasp that the 'scientific conception of the world' which he and others promulgated within the Circle had a much broader cultural goal. It was not simply a neo-positivist anti-metaphysical scientistic programme. Instead he looked to a unified science and a truly scientific conception of the world to make everyday life more humane and democratic. In portrayals of the Vienna Circle written after the Second World War these practical aspirations are often treated as inessential political ambitions when compared with the scientific programmes of logicism and empiricism, whereas in fact internal debates about them were emphasized and regarded as characteristic of the Vienna Circle. We find the clearest presentation of the claim to social reform that is inherent in the scientific conception of the world in the

programmatic essay of the Vienna Circle of 1929:

> The endeavour is to link and harmonise the achievements of individual investigators in their various fields of science. From this aim follows the emphasis on collective efforts, and also the emphasis on what can be grasped intersubjectively; from this springs the search for a neutral system of formulae, for a symbolism freed from the slag of historical languages; and also the search for a total system of concepts. Neatness and clarity are striven for, and dark distances and unfathomable depths rejected. In science there are no 'depths'; all is on the surface. Experience forms a complex network, which cannot always be surveyed and can often be grasped only in parts. Everything is accessible to man; and man is the measure of all things.... The scientific conception of the world knows no unsolvable riddle.
>
> (Carnap, Hahn and Neurath [1929: 15] 1973: 305)

This concluding paraphrase of one of Wittgenstein's claims in the *Tractatus* is the starting point for this late Enlightenment programme of science with its anti-metaphysical orientation. Traditional philosophy with its mannerisms has, in a first step, to be reduced to a critical analysis of language:

> No special 'philosophic assertions' are established, assertions are merely clarified; and at that assertions of empirical science. ... Whichever term may be used to describe such investigations, this much is certain: there is no such thing as philosophy as a basic or universal science alongside or above the various fields of the one empirical science.
>
> (Carnap, Hahn and Neurath [1929: 28] 1973: 316)

The practical impulse behind this therapeutic destruction of a philosophy of metaphysical systems and the rational subject was the desire for a unified and empirical conception of the world on the basis of simple human experience, directed against the 'Zeitgeist' of an increasing number of metaphysical movements whose rise was connected with social and economic factors. Social criticism thus becomes an accompaniment of empirical science and replaces the classical philosophical materialism of the labour movement:

> In previous times, materialism was the expression of this view; in the meantime, however, modern empiricism has shed a number of inadequacies and has taken a strong shape in the scientific conception of the world.
>
> (Carnap, Hahn and Neurath [1929: 29] 1973: 317)

Its closeness to real-life issues and its solidarity with the forces of progress led in the time of emerging fascism to an aggressive determination of its position on social issues:

> We witness the spirit of the scientific conception of the world penetrating in growing measure the forms of personal and public life, in education, upbringing, architecture, and the shaping of economic and social life according to rational principles. The scientific conception of the world serves life, and life receives it.
>
> (Carnap, Hahn and Neurath [1929: 30] 1973: 317)

Social criticism, sociology of knowledge, and philosophic-scientific collective work formed in this conception a programmatic unity in hope of comprehensive progress, which was partly put into practice. But whereas in the natural sciences, Neurath thinks, considerable progress has already been achieved, the situation in the social sciences is less clear ([1930-1: 121] 1983: 44). Neurath therefore attempts in his *Empirical Sociology* (1931) to give a 'physicalist' description of the processes of human social interaction, of the forces that make groups of people cooperate or work against each other and of their influence on the lives of the masses. And his general attitude towards long-term predictions of social sciences is manifested in his cautiously optimistic outlook on possible future developments of society and science (which can nowadays look rather utopian). In this respect it is worth mentioning that after the disintegration of the Vienna Circle (which was also a process of political neutralization) reference to the 'scientific conception of the world' was occasionally used by former members of the Vienna Circle in connection with general ideological questions. For example, Carnap (1963: 81) talks about 'scientific humanism' as a view shared by the majority of the members of the Vienna Circle. By this he means, first, that everyone determines their own life, second, that mankind has the ability to improve their conditions of living and, third, that every liberating action presupposes knowledge about the world, knowledge that is best achieved by scientific means, so that science becomes the most important instrument for an improvement of our lives. According to Carnap, such aims require rational planning which in turn would be best achieved by some

form of socialism and a world-government.

4 *Encyclopedia of Unified Science*

After the dissolution of the Vienna Circle, the forced emigration of most of its members, and the diffusion of the logical empiricist movement from its centres in Vienna, Prague and Budapest, the twin aims of a transformation of philosophy and the establishment of a scientific conception of the world could be envisaged only without reference to their previous cultural context and audience. But even in these difficult times Neurath and his circle still succeeded in organizing well-attended conferences of high standard ('International Congresses for the Unity of Science'), and he also managed to ensure that the unity of science movement continued in the USA (Neurath, Carnap and Morris 1970-1) (see Unity of science). After 1935 Neurath devoted himself to the model of the 'encyclopedia' as a means for furthering this movement. In cooperation with Carnap, Frank and Morris, he planned an international encyclopedia of the unified science and, corresponding to it, worked on a picture language (Isotype) of visual representation. He presented this programme as a development of the ideas of the philosophers of the French Enlightenment. This vision of an unfinished Enlightenment project remains today a striking challenge for the scientific community.

See also: Analytical philosophy; Enlightenment, Continental; Logical positivism; Meaning and verification; Encyclopedists, eighteenth-century

Translated from the German by C. Piller
FRIEDRICH STADLER

List of works

Blumberg, A. and Feigl, H. (1931) 'Logical Positivism: a New Movement in European Philosophy', *Journal of Philosophy* 28: 281-96.(An important presentation of the Vienna Circle in the Anglo-American world.)

Carnap, R. (1928) *Der logische Aufbau der Welt*, Berlin and Schlachtensee: Weltkreis-Verlag; trans. R.A. George, *The Logical Structure of the World*, Berkeley, CA: University of California Press, 1967.(Central book of the phenomenalist period of the early Vienna Circle; methodological solipsism.)

Carnap, R. (1931) 'Überwindung der Metaphysik durch logische Analyse der Sprache', *Erkenntnis* 2: 219-41; trans. A. Pap, 'The Elimination of Metaphysics through Logical Analysis of Language', in A.J. Ayer (ed.) *Logical Positivism*, New York: Free Press, 1959, 60-81.(Classic paper of the physicalist period of the Vienna Circle.)

Carnap, R. (1963) 'Intellectual Autobiography', in P.A. Schilpp (ed.) *The Philosophy of Rudolf Carnap*, La Salle, IL: Open Court, 1-84.(One of the most important reports of the inner life of the Vienna Circle; very well-balanced.)

Carnap, R., Hahn, H. and Neurath, O. (1929) *Wissenschaftliche Weltauffassung: Der Wiener Kreis*, ed. Verein Ernst Mach, Vienna: Artur Wolf Verlag; trans. P. Foulkes and M. Neurath, 'The Scientific Conception of the World: The Vienna Circle', in M. Neurath and R.S. Cohen (eds) *Empiricism and Sociology*, Dordrecht and Boston: Reidel, 1973, 299-318.(The programmatic essay of the Vienna Circle, directed to a general audience, strongly influenced by Neurath; main emphasis is on the concept of a physicalist unified science.)

Feigl, H. (1929) *Theorie und Erfahrung in der Physik ('Theory and Experience in Physics')*, Karlsruhe: G. Braun.

Feigl, H. (1969) 'The Wiener Kreis in America', in D. Fleming and B. Bailyn (eds) *The Intellectual Migration: Europe and America, 1930-1960*, Cambridge, MA: Harvard University Press, 630-73.(Another important autobiographical report on the history of the Vienna Circle and its influence.)

Frank, P. (1949) *Modern Science and Its Philosophy*, Cambridge: Cambridge University Press.(Collected papers by Frank from 1907 to 1947 with an important introduction especially on the background of the Vienna Circle before the First World War.)

Hempel, C.G. (1950) 'Problems and Changes in the Empiricist Criterion of Meaning', *Revue Internationale de Philosophie* 11: 41-63(Authentic description of the transformation of the verification principle.)

Juhos, B. (1971) 'Formen des Positivismus' ('Forms of Positivism'), *Journal for General Philosophy of Science* 2/1: 27-62.(A report of a former Vienna Circle member, with main emphasis on his own position.)

Kraft, V. (1925) *Die Grundformen der wissenschaftlichen Methoden ('The Basic Forms of Scientific Method') Logical Investigations*, Vienna: Sitzungsberichte der Österreichischen Akademie der Wissenschaften. Philos.-historische Klasse 203: 1-104. (Monograph on constructive realism and its hypothetico-deductive justification.)

Kraft, V. (1937) *Die Grundlagen der wissenschaftlichen Wertlehre*, Vienna: Springer; trans. E.H. Schneewind as

Foundations for a Scientific Analysis of Value, ed. H.L. Mulder and E.H. Schneewind, Dordrecht: Reidel, 1981. (A critic of phenomenologist epistemology emphasizing the constructive element in empirical knowledge, and an attempt to justify the norms of accepted morality as means to the satisfaction of universally shared needs.)

Kraft, V. (1950) *Der Wiener Kreis: Der Ursprung des Neopositivismus: Ein Kapitel der jüngsten Philosophiegeschichte*, Vienna and New York: Springer; trans. A. Pap, *The Vienna Circle: The Origin of Neo-Positivism: A Chapter in the History of Recent Philosophy*, New York: Philosophical Library, 1953.(Solid, well-balanced and easily accessible introduction to the main problems, topics, and results of the Vienna Circle.)

Menger, K. (1934) *Moral, Wille, Weltgestaltung: Grundlegung zur Logik der Sitten*, Vienna: Springer; trans. E. van der Schalie, *Morality, Decision, and Social Organization: Towards a Logic of Ethics*, Dordrecht: Reidel, 1974.(A representative externalist treatment of moral values.)

Menger, K. (1979) *Selected Papers in Logic and Foundations, Didactics, Economics*, Dordrecht, Boston and London: Kluwer.(Menger's collected papers with a historical-systematic part I: Papers Introducing Logical Tolerance.)

Menger, K. (1994) *Reminiscences of the Vienna Circle and the Mathematical Colloquium*, eds L. Golland, B. McGuinness and A. Sklar, Dordrecht, Boston and London: Kluwer. (Informative autobiographical account of the ideas of the logical-mathematical wing of the Vienna Circle and of Menger's Mathematical Colloquium.)

Mises, R. von (1968) *Positivism: A Study in Human Understanding*, New York: Dover.(An authentic and well-balanced introduction to the theories of the Vienna Circle, including values, arts and metaphysics.)

Neurath, O. (1930/31) 'Wege der wissenschaftlichen Weltauffassung', *Erkenntnis* 1: 106-25; trans. R.S. Cohen and M. Neurath, 'Ways of the Scientific Conception of the World', in R.S. Cohen and M. Neurath (eds) *Philosophical Papers 1913-1946*, Dordrecht and Boston: Reidel, 1983: 32-47.(An exposition of an anti-metaphysical conception of science in a social and historical context.)

Neurath, O. (1931) *Empirische Soziologie: Der wissenschaftliche Gehalt der Geschichte und Nationalökonomie*, Vienna: Springer; trans. P. Foulkes and M. Neurath, 'Empirical Sociology: The Scientific Content of History and Political Economy' in M. Neurath and R.S. Cohen (eds) *Empiricism and Sociology*, Dordrecht and Boston: Reidel, 1973: 319-421.(Analysis and application of the scientific conception of the world to the social sciences.)

Neurath, O. (1935) 'Pseudorationalismus der Falsifikation', *Erkenntnis* 5: 353-65; trans. R.S. Cohen and M. Neurath, 'Pseudorationalism of Falsification', in R.S. Cohen and M. Neurath (eds) *Philosophical Papers 1913-1946*, Dordrecht and Boston: Reidel, 1983, 121-31.(Fundamental criticism of Popper's Logic of Scientific Discovery as a form of pseudorationalism.)

Neurath, O. (1936) 'L'encyclopédie comme "modèle"', *Revue de Synthèse* 12: 187-201; trans. R.S. Cohen and M. Neurath, 'Encyclopedia as a "model"', in R.S. Cohen and M. Neurath (eds) *Philosophical Papers 1913-1946*, Dordrecht and Boston: Reidel, 1983: 145-58.

Neurath, O. (1936) *International Picture Language*, London: Kegan Paul.(Introduction to education by picture language as a supplement of the encyclopedia project.)

Neurath, O. (1937) *Basic by Isotype*, London: Kegan Paul.(Visualization and illustration of C.K. Ogden's 'Basic English'.)

Neurath, O., Carnap, R. and Morris, C. (eds) (1970/1) *Foundations of the Unity of Science: Towards an International Encyclopedia of Unified Science*, Chicago and London: University of Chicago Press, 2 vols.(A collection of the 19 monographs of the encyclopedia of unified science, which started in 1938).

Schlick, M. (1930) *Fragen der Ethik*, Vienna: Springer; trans. D. Rynin, *Problems of Ethics*, New York: Dover, 1939.(A constructive positivist approach to a number of problems of traditional ethics, such as motivation, freedom and responsibility.)

Schlick, M. (1933) 'Positivismus und Realismus', *Erkenntnis* 3: 1-31; trans. D. Rynin, 'Positivism and Realism', in A.J. Ayer (ed.) *Logical Positivism*, New York: Free Press, 1959: 82-107.(A defence of phenomenalist positivism against Planck's anti-positivism.)

Schlick, M. (1934) 'Über das Fundament der Erkenntnis', *Erkenntnis* 4: 79-99; trans. D. Rynin, 'The Foundation of Knowledge', in A.J. Ayer (ed.) *Logical Positivism*, New York: Free Press, 1979: 209-27.(Essential text for the debate about the foundations of knowledge.)

Schlick, M. (1938) *Gesammelte Aufsätze 1926-1938*, Wien: Gerold and Co.(The main articles by Schlick in German, English and French compiled by Friedrich Waismann.)

References and further reading

Dahms, H.-J. (ed.) (1985) *Philosophie, Wissenschaft, Aufklärung: Beiträge zur Geschichte und Wirkung des Wiener Kreises ('Philosophy, Science, Enlightenment: Studies on the History and Influence of the Vienna Circle')*, Berlin and New York: de Gruyter.(Proceedings of a conference on the development and influence of the Vienna Circle.)

Dahms, H.-J. (1994) *Positivismusstreit: Die Auseinandersetzungen der Frankfurter Schule mit dem logischen Positivismus, dem amerikanischen Pragmatismus und dem kritischen Rationalismus ('The Positivism Dispute: The controversy between the Frankfurt School, American pragmatism and critical rationalism')*, Frankfurt: Suhrkamp.(A study of the so-called positivism-dispute; impressive use of many original sources; well-balanced and not ideological.)

Danneberg, L., Kamlah, A. and Schäfer, L. (eds) (1994) *Reichenbach und die Berliner Gruppe ('Reichenbach and the Berlin Group')*, Braunschweig-Wiesbaden: Vieweg. (First comprehensive documentation of the so-called Berlin Circle and the Berlin Society for Scientific Philosophy.)

Delius, H. (1970) 'Positivismus und Neopositivismus' ('Positivism and NeoPositivism'), in D.A. Diemer and I. Frenzel (eds) *Philosophie*, Frankfurt: Fischer. (Traditional survey article on positivism and neo-positivism.)

Fischer, K.R. (ed) (1995) *Das goldene Zeitalter der Österreichischen Philosophie: Ein Lesebuch ('The Golden Age of Austrian Philosophy: A sourcebook')*, Vienna: WUV Verlag.(Classic and contemporary articles on Austrian philosophy covering Brentano, Mach, Boltzmann, and the Vienna Circle.)

Fischer, K.R. and Wimmer, F.M. (eds) (1993) *Der geistige Anschluß: Philosophie und Politik an der Universität Wien 1930-1950 ('The intellectual Anschluss: Philosophy and Politics at the University of Vienna 1930-50')*, Vienna: WUV Verlag. (Controversial articles and documents on the political and ideological background of the expulsion and Diaspora of the Vienna Circle.)

Geier, M. (1992) *Der Wiener Kreis mit Selbstzeugnissen und Bilddokumenten ('The Vienna Circle with self-descriptions and photographs')*, Reinbeck bei Hamburg: Rowohlt.(Illustrated introduction to the Vienna Circle.)

Giere, R.N. and A.W. Richardson (eds) (1996) *Origins of Logical Empiricism*, Minneapolis, MN and London: University of Minnesota Press.(Contributions to the cultural context, scientific philosophy, logic and mathematics, empirical knowledge of logical empiricism.)

Haller, R. (ed.) (1982) Schlick und Neurath - Ein Symposion ('Schlick and Neurath - A Symposium'), *Grazer Philosophische Studien* 16/17, Amsterdam: Rodopi. (Collection of papers on the occasion of Neurath's and Schlick's centenary.)

Haller, R. (1993) *Neopositivismus: Eine historische Einführung in die Philosophie des Wiener Kreises ('NeoPositivism: A Historical Introduction to the Philosophy of the Vienna Circle')*, Darmstadt: Wissenschaftliche Buchgesellschaft.(Solid, problem-oriented introduction reaching from classic empiricism/positivism to the Vienna Circle; sections on the leading figures of the Vienna Circle as well as on Einstein, Russell and Wittgenstein.)

Hanfling, O. (1981) *Logical Positivism*, Oxford: Blackwell.(A systematic survey of the themes and topics central to the Vienna Circle.)

Hempel, C.G. (1981) 'Der Wiener Kreis und die Metamorphosen seines Empirismus', in N. Leser (ed.) *Das geistige Leben Wiens in der Zwischenkriegszeit ('Intellectual Life in Vienna in the Interwar Years')*, Vienna: ÖBV: 205-15; trans. C.J. Piller as 'The Vienna Circle and the Metamorphosis of its Empiricism', in R. Jeffrey (ed.) *Philosophical Essays: Early and Late*, Cambridge: Cambridge University Press, 1997.(Important historical account by a leading member of the Circle.)

Hempel, C.G. (1993) 'Empiricism in the Vienna Circle and in the Berlin Society', in F. Stadler (ed.) *Scientific Philosophy: Origins and Developments*, Dordrecht, Boston and London: Kluwer, 1-10.(A concise evaluation of logical empiricism from the point of view of contemporary philosophy of science.)

Kamitz, R. (1973) *Positivismus: Befreiung vom Dogma ('Positivism: Liberation from Dogma')*, Munich and Vienna: Langen Müller.(A systematic exposition with main emphasis on Carnap.)

Quine, W.V.O. (1953) *From a Logical Point of View: Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press.(Classic collection of papers containing the influential critique of logical positivism 'Two Dogmas of Empiricism'.)

Skirbekk, G. (ed.) (1977) *Wahrheitstheorien: Eine Auswahl aus den Diskussionen über Wahrheit im 20.Jahrhundert ('Theories of truth: A survey of 20th century discussions of truth')*, Frankfurt: Suhrkamp.(A collection of papers on truth published between 1907 and 1973.)

Stadler, F. (1982) *Vom Positivismus zur "Wissenschaftlichen Weltauffassung": Am Beispiel der Wirkungsgeschichte von Ernst Mach in Österreich von 1895-1934 ('From Positivism to the "Scientific Conception of the World"; an example of the historical influence of Ernst Mach in Austria 1895-1934')*,

Vienna and Munich: Löcker.
Stadler, F. (ed.) (1993) *Scientific Philosophy: Origins and Developments*, Vienna Circle Institute Yearbook 1, Dordrecht, Boston and London: Kluwer.(A contemporary look at scientific philosophy and philosophy of science by former members of the Vienna Circle and their students.)
Stadler, F. (1997) *Studien zum Wiener Kreis: Ursprung, Entwicklung und Wirkung des Logischen Empirismus im Kontext ('Studies on the Vienna Circle: Origin, Development and Influence of Logical Empiricism in Context')*, Frankfurt: Suhrkamp. (Historical studies on the rise of scientific philosophy in its cultural context from the beginning of the century to the Second World War; including so far unpublished protocols of Vienna Circle meetings, an interview with Karl Popper and a bio-bibliography.)
Stadler, F. and Haller, R. (1993) *Wien-Berlin-Prag: Der Aufstieg der wissenschaftlichen Philosophie: Zentenarien Rudolf Carnap, Hans Reichenbach, Edgar Zilsel ('Vienna-Berlin-Prague: The Rise of Scientific Philosophy: Centenary of Rudolf Carnap, Hans Reichenbach, Edgar Zilsel')*, Vienna: Hölder-Pichler-Tempsky. (A collection of essays on the development of scientific philosophy, focusing on three important logical empiricists.)
Uebel, T.E. (ed.) (1991) *Rediscovering the Forgotten Vienna Circle: Austrian Studies on Otto Neurath and the Vienna Circle*, Dordrecht, Boston and London: Kluwer.(A collection of papers aiming at a new interpretation of the history of the Vienna Circle.)
Uebel, T.E. (1992) *Overcoming Logical Positivism from Within: The Emergence of Neurath's Naturalism in the Vienna Circle's Protocol Sentence Debate*, Amsterdam and Atlanta, GA: Rodopi.(Important reconstruction and interpretation of one of the central debates in the Vienna Circle, including an attempt to justify a holistic and naturalistic approach to philosophy of science without foundationalism.)

1 Book series

Studien zur österreichischen Philosophie (1979-), ed. R. Haller, Amsterdam: Rodopi.
Veröffentlichungen des Instituts Wiener Kreis, ed. F. Stadler; vols 1-3, Vienna: Hölder-Pichler-Tempsky, 1991-95; vols 4-, Vienna and New York: Springer, 1996-.
Vienna Circle Collection (1973-), ed. H.L. Mulder, R.S. Cohen, B. McGuinness and R. Haller, Dordrecht, Boston, MA and London: Kluwer.
Vienna Circle Institute Yearbook (1993-), ed. F. Stadler, Dordrecht, Boston, MA and London: Kluwer.

Villey, Michel (1914-88)

Michel Villey was France's leading post-war philosopher of law in the 'natural law' mode. He aimed to rediscover a distinctively philosophical approach to law rooted in the history both of legal ideas and of legal institutions. Legal institutions he considered to have been uniquely a gift of Greco-Roman civilization to the world. They represent a distinctive domain of human activity, concerned with an objectively just ordering of human relationships as these affect external conduct and the possession and use of things. Villey has in common with legal positivism a belief in the differentiation of the legal, concerned with objective interpersonal relations in their 'external' concern with a distribution of things, from morality and from religion with their distinctively 'internal' and 'spiritual' concerns. In opposition to legal positivism, however, he holds that justice is a concept implicit in the legal, and discountenances positivists' tendency to reduce law to a simple aggregation of enacted statutes.

Villey poses the Aristotelian question: what is the defining end of the legal that gives it its distinctive character? His answer is: that of upholding order by settling conflicts of interest among the members of a human society. Law's particular aim is towards objective conflict-resolution in respect of the 'things', corporeal or incorporeal, in which humans take an interest. What law seeks is a just allocation or reallocation of things to persons, in the context of some concrete controversy over their rightful allocation.

This is a creative art, not a science. Guided by the equity of concrete human relations, the judge has to work out creatively the right allocation of things in issue (see Justice, equity and law). This right resolution ('*le droit*', 'the right' in an objective sense) has to be guided by any relevant enacted laws. But in any seriously disputed case, rival rules of statute-law can be advanced by rival parties, and the judge must still choose. Further guidance is to be drawn from contemporary custom and from judicial precedents. These, too, are genuine parts of the law, in opposition to the ideas of merely voluntaristic positivists, who reduce law to being no more than the total body of the statutes in force. But even when custom and precedent are brought in, they only define further the whole context in which the relations of the parties must be considered with a view to reaching the equitable resolution of their conflict, the right of the matter in an objective sense. This objective character of the right Villey regarded as the deepest element implicit in a Thomistic understanding of law. He considered it hostile to the subjectivism implicit in contemporary ideas of human rights, which in his view constituted an ideology inimical to legal order properly conceived.

Villey's version of Thomistic philosophy is an adventurous and highly contestable reading of the texts; but it constitutes an original contribution to natural law thought, and shows in an interesting way the possibility of using natural law premises to cast doubt on contemporary views of human rights.

See also: Law, philosophy of; Legal positivism; Natural law; Roman law; Thomism

NEIL MacCORMICK

List of works

Villey, M. (1962) *Leçons d'histoire de la philosophie du droit (Lectures on the History of Legal Philosophy)*, 2nd edn, Paris: Dalloz.(An introductory work for students, relatively straightforward.)

Villey, M. (1975a) *Philosophie du droit (Philosophy of Law)*, Paris: Dalloz.(A massive two-volume statement of Villey's essential positions, little changed in his later writings.)

Villey, M. (1975b) *La formation de la pensée juridique moderne (The Formation of Modern Legal Thought)*, 4th edn, Paris: Montchrestien.(An account of the history of legal ideas, making clear Villey's view of the centrality of Greco-Roman practice and thought for 'law' properly so-called.)

Villey, M. (1983) *Le droit et les droits de l'homme (Law and the Rights of Man)*, Paris: Presses Universitaires de France.(This work, dedicated to the Pope as well as to his own family, gives Villey's critique of the modern approach to human rights; quite an accessible text, aimed to influence public debate.)

Villey, M. (1991) 'Law in Things', in P. Amselek and N. MacCormick (eds), *Controversies About Law's Ontology*, Edinburgh: Edinburgh University Press, ch. 1.(A relatively short and witty statement of the author's view about the objectivity of right and its essential reference to 'things', in a quite wide sense; one of the regrettably few statements available in English.)

References and further reading

Freund, J. (1992) 'Michel Villey et le renouveau de la philosophie du droit' (Michel Villey and the Rebirth of Legal Philosophy), *Archives de philosophie du droit* 37: 5-14.(A useful and sympathetic short guide to the essential points of Villey's thought.)

Violence

Violence is a central concept for much discussion of moral and political life, but lots of debate employing the concept is confused by the lack of clarity about its meaning and about the moral status it should have in our development of public policy. Wide understandings of the term - for instance, structural violence - not only include too much under the name of violence, but also put an excessively negative moral loading into the concept. This is also a problem for some other definitions of violence, such as legitimist definitions, which treat violence as essentially the illegitimate use of force. It is better to confront directly the important and disturbing claim that violence is sometimes morally permissible than to settle it by definitional fiat.

1 Definitions and paradigms

In discussing violence, it helps to begin with some obvious pre-theoretical examples, such as knife attacks, savage beatings, shootings, bombings and torture. Offering such cases may be insufficient for the definitional clarification of the concept that we seek as philosophers, but these paradigms can fix our attention on at least part of what is to be made clear. If theorists claim that these are not acts of violence, then they have some explaining (or theorizing) to do.

In fact, some theorists seem to claim precisely this, while others claim that, although the above list is correct as far as it goes, there are some more surprising candidates for inclusion as paradigms. Both groups seem to suffer from a certain disdain for communicative clarity, but, more importantly still, from a tendency to overmoralize the concept. Since the idea plays a crucial role in a wide range of significant moral and political debates, it is imperative to be clear about what we mean by the term and what sort of evaluative loading it carries.

2 Two theories and their difficulties

To elaborate on this point, it is significant that, although the tendency to eliminate some of the paradigm cases and the tendency to add surprising ones usually spring from very different political motivations, both draw upon the idea that violence is inherently wrong. We can see this if we examine two theories that embody these different tendencies. Legitimists about violence instantiate the first tendency and proponents of the notion of structural violence embody the second. Legitimist definitions treat violence as the *illegitimate* use of force. Characteristically such an approach is associated with politically conservative outlooks, but this is a psychological rather than a logical connection. A major problem for these thinkers is that they have considerable difficulty explicating a theoretically helpful sense of legitimacy. Most of them employ a norm of political legitimacy (as is clear in Sidney Hook’s definition of violence as ‘the illegal employment of methods of physical coercion for personal or group ends’ (Grundy and Weinstein 1974: 12)), but they also think of this as having some moral underpinning, and hence tend to treat violence as essentially in the category of wrongs. In a legitimate state, properly authorized employment of shooting or savage beating by police will not then count as violence since it is a politically legitimate use of force. But there are obvious problems about distinguishing its various types (*de facto*, *de jure*, internationally recognized though internally contested, and so on) and it is also hard to avoid the awkward consequences that the state of nature, even understood in Hobbes’ extreme terms, cannot be a state of violence (since the ideas of political legitimacy or illegitimacy do not apply), and that warfare between two legitimate states cannot involve violence. In the latter case, both states will usually be engaged in the politically legitimated use of force which, by definition, cannot be violence. It is unlikely that this conceptual apparatus could help in the practical discussion of moral limits to police violence or the conduct of war (see War and peace, philosophy of §§1-2, 6). If, on the other hand, in the face of the problems posed by resort to political legitimacy, moral legitimacy is invoked, then justified killing in self-defence or defence of innocent others will not involve violence and all violence will be morally illegitimate by explicit definition, which represents the serious fudging of a central question.

By contrast, the concept of structural violence, much in vogue with the political left, flaunts intuition, not by excluding what seem to be obvious candidates, but by bringing into the category of violence a wide range of things that would not, at first, be included. This definitional proposal springs from the work of the Norwegian sociologist Johan Galtung (1969) and the basic idea is that we should use the term violence in a very wide and extended way to refer to any form of social injustice whether inflicted by individuals or by institutions or by the workings of

society at large, and whether or not it involves the deliberate infliction of personal injury by episodes of physical or psychological force. Such episodes would be merely one type of violence. Advocates of this wide definitional strategy tend to argue that standard definitions of violence focus too much upon the harms intentionally inflicted by the personal employment of physical force. Since these instantiate only one type of damage that human contrivance can inflict, why not use the term violence to cover all such damage? This style of proposal faces several difficulties: it is confusing, politically unhelpful and evades a central problem about violence.

It is confusing because people do not ordinarily mean by 'violence' any and every form of social injustice, they mean such things as beating people up or torturing them with electrodes. They do not mean to refer to iniquitous taxation proposals or discriminatory housing policies. These things may lead to violence or they may need violence, or the threat of it, to sustain and implement them, but that is another matter. It may be argued that violence is a bad thing, but it is just confusing to treat every bad thing as violence. Moreover, the expansive sense of violence does not help with an agenda of social reform, because it encourages the cosy but ultimately stultifying belief that all social evils are really one and hence will yield to the one solution. The reality seems far more likely to be that the various problems we face are genuinely diverse, although related in various ways, and have their own specific features that call for detailed study with a view to solution. Better gun control legislation might well alleviate problems of violence in the USA while doing nothing to change poverty levels. Furthermore, the proponents of the concept of structural violence assume that violence is essentially morally wrong, and seek broad support for various social reforms by claiming that they will eliminate (structural) violence. But this assumption is contentious. It may be that there is something about violence which makes resort to it always, in certain ways, regrettable, while none the less, at times, it can be argued to be morally legitimate, even morally required. These issues need to be faced, not evaded.

3 The restricted definition

It is therefore preferable to operate with a concept of violence which is both narrower than that of structural violence and less morally loaded than either it or the legitimist definitions. Initial guidance is provided by such dictionary definitions as that of the *Oxford English Dictionary*: 'The exercise of physical force so as to inflict injury on or damage to persons or property'. Yet this is arguably too restrictive in excluding psychological violence since we can give some sense to the idea of damage being inflicted by psychological force. Cases of psychological pressure producing overpowering effects seem close to the paradigms given earlier, as, for example, when a parent cruelly humiliates a vulnerable child to the point of collapse or great distress. There is a tendency in the literature to slide from psychological violence to structural violence, but this embodies a confusion since it rests on the tendency to think of psychological violence as *impalpable* and then to feel that its admission endorses the even more impalpable structural violence. However, the examples which make the category of psychological violence plausible are all very palpable indeed, involving an immediacy and specificity of pressure producing overwhelming effects on the victim. The term 'verbal violence' is sometimes used merely metaphorically, but it can make literal sense in contexts where it is a species of psychological violence, and where the term 'tongue lashing' becomes strictly appropriate.

It is worth noting that a 'restricted' definition of this sort is not entirely 'descriptive' or morally neutral although it is much less morally loaded than the other definitions discussed above. It contains reference to the notions of injury or damage which are at least evaluative with respect to some notion of normal or proper functioning. In addition, these notions have complex relations to such centrally moral concepts as harm and autonomy. The conceptual structure of the definition therefore allows us insight into much that is otherwise puzzling about our attitudes to violence, and helps explain some of the appeal of the structural and legitimist outlooks as well as the nature of the distortions they produce. In particular, it helps show why many people have a strong tendency to think that violence is always wrong and that it is none the less sometimes morally right. This tendency is partly attributable to the fact that violence inherently involves the intentional infliction of injury or damage, and we have a strong tendency to equate this with the infliction of harm or wrong. At the same time, most of us recognize that the police, for instance, are sometimes justified in killing or wounding criminals to prevent violent attacks upon innocent civilians or upon the police themselves. There can be problems about the too-ready resort to violence by police, but there is nothing imaginary or fantastic about the circumstances in which such resort seems morally justifiable. In Melbourne, in December 1994, a man ran amok with a semi-automatic weapon in a suburban street, killed two perfect strangers and kept firing at police and passing motorists until a police marksman shot him dead.

Such criminals are certainly injured, but they are not wronged or, in the morally significant sense, harmed.

4 More benefits of the restricted definition

Another problem that can be clarified by the restricted definition is that of violent sports that involve the consent of their participants. Some have thought that boxing, for instance, cannot be violent because the fighters mutually agree to bombard each other with punches; they are influenced by the idea that some action affecting another cannot constitute a wrong or a harm if the other consents to it. Whatever one thinks of that principle, it cannot affect the question of whether hitting someone very hard on the head with the intention of rendering him unconscious is a violent act, at least if we understand violence in the restrictive manner. Anyone so acting is clearly aiming to inflict an injury.

Moreover, the restricted definition helps elucidate two other features of violence which are otherwise puzzling; namely, the point of having such a concept and the fact that even justified violence is regrettable. On the first issue, it is clear that many of us have a distinctive fear of the intrusion into our lives of those intent upon the forceful infliction of injury upon us. We also fear the effects of natural disasters, diseases, social prejudice and injustice, indifference and accident, but our fear of violence usually has a quite particular significance because we are understandably anxious about damage that is specifically directed against us and that is aimed at immediate and dramatic impairment of our wellbeing. This combines with the fact that even justified resort to violence carries with it problems of misjudgment and escalation, as well as the possibility of disturbing psychological consequences for those who employ it, to make the use of violence regrettable even where it is not immoral.

See also: Civil disobedience

C.A.J. COADY

References and further reading

Arendt, H. (1970) *On Violence*, London: Penguin.(An influential discussion of the significance of violence.)

Coady, C.A.J. (1986) ‘The Idea of Violence’, *Journal of Applied Philosophy* 3 (1): 3-19.(A longer treatment of some of the issues explored here.)

Galtung, J. (1969) ’Violence, Peace and Peace Research’, *The Journal of Peace Research* 6 (2): 167-91, esp. 168 and 173.(A seminal exposition of the concept of structural violence and an argument for its utility.)

Garver, N. (1968) ’What Violence Is’, *The Nation*, 206 (26); repr. in J. Rachels and F.A. Tillman (eds), *Philosophical Issues: A Contemporary Introduction*, New York: Harper & Row, 1972, 223-8.(A defence of a wide understanding of violence in the spirit of the ‘structural’ definition.)

Grundy, K.W. and Weinstein, M.A. (1974) *The Ideologies of Violence*, Columbus, OH: Charles Merrill, 8-13.(A useful introductory discussion of some definitions of violence that bear upon those discussed here. They cite the definition by Sidney Hook mentioned in §1.)

Harris, J. (1980) *Violence and Responsibility*, London: Routledge.(Uses a very wide idea of violence to draw strong moral and political conclusions.)

Honderich, T. (1980) *Violence for Equality: Inquiries in Political Philosophy*, Harmondsworth: Penguin, esp. 96-100.(A partial defence of violence as a political tool, based upon a type of legitimist understanding of violence.)

Lee, S. (1996) ’Poverty and Violence’, *Social Theory and Practice* 22 (1): 67-82.(Thoughtful article exploring the connections between violence and poverty and arguing for a ‘wide’ definition of violence in terms of which violence can be said to be the cause of poverty.)

Shaffer, J.A. (ed) (1971) *Violence*, New York: David McKay.(A good collection of essays, mostly on the conceptual analysis of violence.)

Wilkins, B.T. (1992) *Terrorism and Collective Responsibility*, London and New York: Routledge, esp. ch. 3 ‘Violence and Force’.(A good discussion of some issues to do with violence and terrorism, insisting, in particular, that violence does not always involve the violation of an actual right.)

Wolff, R.P. (1969) ’On Violence’, *Journal of Philosophy* 66 (19): 601-16.(Uses a legitimist definition to devalue the moral importance of resort to violence.)

Virtue epistemology

'Virtue epistemology' is the name of a class of theories that analyse fundamental epistemic concepts such as justification or knowledge in terms of properties of persons rather than properties of beliefs. Some of these theories make the basic concept constitutive of justification or knowledge that of a reliable belief-forming process, or a reliable belief-forming faculty or, alternatively, a properly functioning faculty. Others make the fundamental concept that of an epistemic or intellectual virtue in the sense of virtue used in ethics. In all these theories, epistemic evaluation rests on some virtuous quality of the person that enables them to act in a cognitively effective and commendable way, although not all use the term 'virtue'. The early, simple forms of process reliabilism are best treated as precursors to virtue epistemology since the latter arose out of the former and has added requirements for knowledge intended to capture the idea of epistemic behaviour that is subjectively responsible as well as objectively reliable.

Proponents of virtue epistemology claim a number of advantages. It is said to bypass disputes between foundationalists and coherentists on proper cognitive structure, to avoid sceptical worries, to avoid the impasse between internalism and externalism, and to broaden the range of epistemological inquiry in a way that permits the recovering of such neglected epistemic values as understanding and wisdom.

1 History of virtue epistemology and its varieties

Virtue epistemology focuses on traits of persons or their faculties or psychological processes in the analysis of basic epistemic concepts such as justification or knowledge. It is a reaction to a style of epistemology that makes properties of individual beliefs and evidential relations between beliefs the focus of analysis. The theories in the latter category often handle epistemic evaluation on the model of deontological ethics (see Deontological ethics). To be justified, a belief must not violate any epistemic rules or duties - it must be held for the right reason. In contrast, many forms of virtue epistemology have arisen out of reliabilism. According to reliabilist theories, what makes a true belief an instance of knowledge or, alternatively, what makes a belief justified is that it was formed by a reliable process for obtaining the truth (see Reliabilism). Reliable processes include perception and memory whose reliability need not be cognitively accessible to the believer, and so reliabilism is a form of externalism whereas, typically, deontological theories are internalist (see Internalism and externalism in epistemology). Reliabilism is structurally parallel to consequentialist ethics, in particular rule utilitarianism. In a reliabilist theory, the consequences (a high proportion of true beliefs) make a belief process reliable just as in a rule-utilitarian theory the consequences (a high proportion of utility maximizing actions) make a rule a good one. A belief is epistemically justified (alternatively, warranted) by being the product of a reliable truth-producing process, just as an act is morally justified by being the product of a reliable utility-maximizing rule (see Consequentialism).

An advantage of reliabilism is that its requirements for knowledge are relatively easy to satisfy in ordinary circumstances, even by young children and unsophisticated adults. Hence it is often considered closer to common sense than the more rigorous deontological theories. Some theorists also think it has an advantage with respect to avoiding scepticism. For one has knowledge as long as there is no Cartesian demon and one's cognitive processes do, as a matter of fact, reliably hook up with the truth, whether or not one has an answer to the sceptical challenge (Greco 1992).

Since reliabilism shifts the focus of analysis from properties of beliefs to properties of the believer it sets the stage for a form of epistemology that makes fundamental the virtues and vices of a person. It is not uncommon, then, to classify the more recent forms of reliabilism as virtue epistemology (Kvanvig 1992; Greco 1992), and in one place Alvin Goldman classifies his later version of process reliabilism as a form of virtue epistemology (1992). Simple reliabilism, however, is only distantly related to any kind of virtue ethics. A reliable belief-forming process has little in common with the traditional notion of a virtue and, as already pointed out, reliabilism is structurally parallel to consequentialist ethics, not virtue ethics.

Act-based moral theories such as consequentialist and deontological theories make the central concept that of the right act. Similarly, reliabilist and deontologial theories in epistemology make the central concept that of a belief that is justified or, alternatively, warranted (possessing the property that converts true belief into knowledge). A virtue ethic, in contrast, makes the primary concept that of a virtue. The concept of a right act is derivative and is

often less important. A moral virtue is defined neither in terms of the consequences of the acts to which it gives rise, nor in terms of the rules or principles followed by persons exhibiting the virtues. A virtue includes dispositions to have certain emotions as well as dispositions to act in certain ways, and in its Aristotelian and neo-Aristotelian forms virtuous traits are closely tied to a teleological understanding of human nature (see Virtue ethics).

Given this account of the characteristics of virtue ethics, the epistemic parallel would be one in which intellectual virtue is the fundamental concept, and the concept of a justified belief is derivative and may also be less important. A property is not an intellectual virtue simply because of its propensity to produce true beliefs, nor is it a virtue because it involves following epistemic rules or principles. The parallel with virtue ethics would be even closer if an intellectual virtue includes emotion-dispositions as well as dispositions to act cognitively in certain ways, and a neo-Aristotelian form of virtue epistemology would include a connection between epistemic goods and the good life.

While simple reliabilism is quite a distance from virtue epistemology as just described, recent theories have increasingly moved in that direction. Ernest Sosa (1980) first introduced the concept of intellectual virtue into contemporary literature as a refinement of reliabilism. Sosa proposed that the concept of an intellectual virtue can be used to bypass the dispute between foundationalists and coherentists about the logical or evidential relations needed for proper cognitive structure (see Knowledge, concept of §4). What Sosa meant by an intellectual virtue was a reliable belief-forming faculty like eyesight or memory, and he defined such a virtue in terms of its consequences: 'An intellectual virtue is a quality bound to help maximize one's surplus of truth over error' (Sosa 1985: 227). In subsequent work, Sosa (1991) has distinguished animal knowledge from reflective knowledge. For the former it is sufficient that the belief was caused by a reliable truth-producing faculty. For the latter he adds an internalist component. The believer must have a reliable grasp of the fact that their belief is grounded in a reliable cognitive faculty. Sosa also ties the concept of an intellectual virtue to the 'inner nature' of the knowing subject. These recent moves bring Sosa's concept of an intellectual virtue closer to the traditional notion of virtue as used in ethics.

2 History of virtue epistemology and its varieties (cont.)

A version of virtue epistemology that moves even closer to virtue ethics is Alvin Plantinga's theory of warrant as proper function (1993b). Plantinga uses the term 'warrant' for that property which in sufficient quantity converts true belief into knowledge. A warranted belief, according to Plantinga, is one that is produced by properly functioning faculties in the appropriate environment according to a design plan aimed at truth. Like Sosa, Plantinga makes faculties the focus of analysis, but unlike Sosa he does not identify faculties as virtues because of their consequences. So Plantinga's theory is not modelled on consequentialist ethics, but it is not modelled on virtue ethics either. In fact, Plantinga never uses the term 'virtue' nor does he give any attention to the rich history of the concept of virtue as used in ethics. Plantinga's theory moves closer to traditional accounts of virtue, however, because what counts as properly functioning faculties is determined by a background notion of what is good or desirable for human beings with the nature that they have.

John Greco's version of virtue epistemology modifies Sosa's theory by adding an element that is intended to capture the idea of epistemically responsible as well as reliable belief. To have 'positive epistemic status', Greco says (1990, 1993), not only must a belief be the result of a reliable cognitive faculty (a virtue), but that virtue must have its basis in the believer's conforming to epistemic norms they countenance. Greco's and Sosa's additions of an internalist requirement of belief formation to the reliability condition give their versions of virtue epistemology an ethical component which is missing from simple reliabilism and Plantinga's theory.

Two of the first attempts to link epistemological inquiry with intellectual virtues in the traditional sense of virtue were made by Lorraine Code and James Montmarquet. Code (1987) gives an account of intellectual virtue that 'socializes' epistemology, stresses the place of the knowing subject in the epistemic community, and contextualizes the state of knowing within a background of states that include the non-cognitive (see Feminist epistemology §2-3). She calls her position 'responsibilism' rather than reliabilism because of the knower's degree of choice in cognitive structuring and degree of accountability for these choices. Code does not offer a detailed account of the concept of intellectual virtue, but she is perhaps the first in contemporary epistemology to call attention to the way intellectual virtues in the sense of virtue used in ethics can be useful to the interests of

epistemology.

James Montmarquet (1993) gives an extensive treatment of epistemic virtues in the classical sense of a trait for which we are responsible and which are similar to moral virtues. Montmarquet includes in his list of epistemic virtues such traits as open-mindedness; intellectual carefulness, thoroughness, and impartiality; and intellectual courage and perseverance. He maintains that epistemic virtues are the traits which people who desire the truth would desire to have, but he denies that they are reliably truth-conducive in any straightforward way. So Montmarquet's theory, like Code's, does not evolve out of reliabilism, and his principal concerns are somewhat different from those driving reliabilism.

Linda Zagzebski (1996) has developed a version of virtue epistemology that is explicitly modelled on virtue ethics. Like Code and Montmarquet, what she means by intellectual virtues are such traits as intellectual autonomy and courage, intellectual carefulness and thoroughness, open-mindedness and fair-mindedness, but like Sosa and Greco, she regards reliability as a component of virtue. Zagzebski offers a virtue theory of ethics inclusive enough to handle the intellectual as well as the moral virtues within a single theory, proposing that moral and intellectual virtues have the same basic structure and are acquired in the same way. Both include a motivational component and a component of reliable success in bringing about the aims of the motivational component. In the case of the intellectual virtues this means that reliability in acquiring the truth is a component of intellectual virtue, but there is also a component of intellectual motivation from which the cognitive behaviour characteristic of intellectually virtuous persons arises. Zagzebski then defines the concepts of justified belief and epistemic duty in terms of intellectual virtue in the same way the concepts of right act and of moral duty can be defined in terms of moral virtue in a pure virtue ethics. In both cases the concept of virtue is more basic. She subsequently defines an 'act of virtue', the concept of an act that is praiseworthy in every way - in the act itself, in its motive, in the end it achieves, and which reaches its end because of these other praiseworthy features of the act. An act of intellectual virtue is an act that has an intellectually virtuous motive, is an act that would be characteristically performed by intellectually virtuous persons in the same circumstances, and which reaches truth because of these other features of the act. Knowledge is belief arising out of acts of intellectual virtue. This strategy captures the combination of internal responsibility and epistemic success desired by virtue epistemologists.

3 Advantages and disadvantages of virtue epistemology

When Sosa introduced the concept of intellectual virtue into the contemporary literature, he thought that the shift of focus from properties of beliefs to properties of persons should make it possible to bypass the dispute between foundationalists and coherentists over the logical and evidential relations between beliefs needed for proper epistemic structure. In addition, two possible advantages of reliabilism have already been mentioned. First, its requirements for knowledge are easier to satisfy than internalist theories and, hence, it is said to be closer to common sense than are internalist theories. Second, its requirements for knowledge do not include the possession by the knower of an answer to the sceptical challenge. For example, they need not be able to answer the question 'But what if there is an evil demon?', even though the presence of an evil demon would deprive them of knowledge. Since having an answer to this question is usually thought to be at best difficult, and probably impossible for the average knower, it is an advantage of a theory of knowledge if its formulation permits us to ignore such worries (see Scepticism).

These advantages of reliabilism may be offset by certain disadvantages, however. Consider first the requirement that reliability is a necessary condition for knowledge. Most versions of virtue epistemology have this requirement. A common objection to the reliability requirement for knowledge or justification is the so-called 'generality problem'. This is the problem that a particular belief can be described as an instance of many different belief-forming processes, some broader (believing whatever you read), some narrower (believing what you read in a particular journal), some so narrow as to apply only to a single belief (believing what you read in a certain journal on a certain subject by a certain writer). The degree of truth-conduciveness will vary greatly, depending upon the way in which the belief-forming process is described. This problem requires refinements to the theory rather than outright rejection. Ways of handling the generality problem have been suggested by Goldman (1986), Greco (1993), Sosa (1991) and Zagzebski (1996). Zagzebski suggests that the class of cases against which truth-conduciveness is to be measured should be empirically determined by the way in which people naturally generalize their belief-forming processes. So if a person who believes what they read in a reliable journal then

goes on to believe whatever they read in any periodical, their belief would not count as deriving from a reliable belief-forming process, whereas it would be reliable if they do not tend to generalize their belief-forming tendency to unreliable sources.

Simple reliabilism maintains that reliability is not only necessary, but is sufficient for knowledge. Zagzebski denies that simple reliabilism ought to be considered a form of virtue epistemology, but it is worth mentioning the most serious objection to it since reflection on it was one of the motivations leading to the development of virtue epistemology. The problem is that simple reliabilism does not rule out of the category of knowledge beliefs formed by processes or faculties that are reliable by accident (Plantinga 1993a; Zagzebski 1996). This suggests that any theory that includes reliability as a requirement for knowledge should include an account of what makes a process, trait or faculty reliable in the right way. Virtue epistemologists, then, characteristically add requirements for knowledge other than reliability to eliminate this problem by showing how reliability arises out of something else - in the case of Sosa, the believer’s inner nature; in the case of Plantinga, faculties whose proper function is determined by design; for Greco, the norms the believer countenances; for Zagzebski, intellectual virtues. These conditions involve activity on the part of the knower that exhibits epistemic responsibility or features praiseworthy from an internal perspective.

Requirements for knowledge in addition to reliability are seen by their adherents as improving upon simple reliabilism, but these requirements can also be perceived as detracting from one of the advantages of simple reliabilism. This is its generosity in spreading the domain of knowledge to include true beliefs formed in typical circumstances by young children and unreflective adults. But the more we add responsibility and internal awareness to the conditions for knowledge, the more we shrink the resulting class of knowing states. Can children and unreflective adults have perceptual and memory knowledge on these more demanding theories? Sosa’s way of handling the problem is to distinguish animal knowledge from reflective knowledge. Zagzebski prefers not to make it a condition for knowledge that the believer actually possesses the full virtue but, rather, that the believer has virtuous intellectual motivations and acts cognitively the way virtuous persons do in the same circumstances. Given that virtuous persons formulate perceptual and memory beliefs in many circumstances with little or no reflection or investigation, it is not difficult for children to imitate their behaviour and, hence, to have knowledge in these cases. Greco’s way of handling the problem (1990) is to say that while a belief has positive epistemic status only if it conforms with the norms that the believer countenances, this does not require that the believer should have cognitive access either to the norms themselves or to the fact that their beliefs conform with them.

It can be argued that virtue epistemology has the potential to deal effectively with two other problems of contemporary epistemology. First, there are the problems surrounding the concept of justification which have led to the impasse between internalism and externalism. Since justification is a property of a belief, it is very difficult to adjudicate disputes over this concept if the belief is treated as the bottom-level object of evaluation. If we focus instead on the deeper concept of an intellectual virtue and treat the justifiability of a belief as derivative, it may turn out that justifiability is only one normative property of beliefs among others and the competing intuitions of internalists and externalists require the analysis of more than one property of beliefs, all of which are based in some way on the concept of virtue. Second, focusing epistemology on belief has led to the neglect of two epistemic values that have been important for long periods in the past - understanding and wisdom. An approach focused on intellectual virtues, on the other hand, naturally leads to an account of these values, since understanding and wisdom are either virtues themselves or are very closely connected with virtues (see Wisdom).

See also: Confucian philosophy, Chinese; Justification, epistemic; Normative epistemology

LINDA ZAGZEBSKI

References and further reading

Code, L. (1987) *Epistemic Responsibility*, Hanover, NH: University Press of New England for Brown University Press.(Perhaps the first book to apply virtues in the sense used in ethics to the concerns of epistemology. Particularly good for its examples.)

Goldman, A.I. (1986) *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.(Important example of contemporary reliabilism.)

Goldman, A.I. (1992) ‘Epistemic Folkways and Scientific Epistemology’, in *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge, MA: MIT Press. (The only paper in which Goldman explicitly

associates himself with virtue epistemology.)

Greco, J. (1990) ‘Internalism and Epistemically Responsible Belief’, *Synthèse* 85 (2): 245-77.(Proposes a version of virtue epistemology which combines reliability with following epistemic norms.)

Greco, J. (1992) ‘Virtue Epistemology’, in J. Dancy and E. Sosa (eds) *A Companion to Epistemology*, Oxford: Blackwell.(Good summary of how some of the earlier work in virtue epistemology arose out of reliabilism.)

Greco, J. (1993) ‘Virtues and Vices of Virtue Epistemology’, *Canadian Journal of Philosophy* 23 (3): 413-32. (More detailed discussion of virtue epistemology than his 1992 work.)

Kvanvig, J. (1992) *The Intellectual Virtues and the Life of the Mind*, Lanham, MD: Rowman & Littlefield. (Focuses mostly on reliabilism and its problems. It then proposes a move away from atomistic epistemology.)

Montmarquet, J.A. (1993) *Epistemic Virtue and Doxastic Responsibility*, Lanham, MA: Rowman & Littlefield. (Gives an account of the nature of epistemic virtue and uses it to contribute to the debate on the ethics of belief.)

Plantinga, A. (1993a) *Warrant: The Current Debate*, New York: Oxford University Press.(Important discussion and diagnosis of contemporary theories.)

Plantinga, A. (1993b) *Warrant and Proper Function*, New York: Oxford University Press. (Presents Plantinga’s own influential theory of knowledge.)

Sosa, E. (1980) ‘The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge’, in *Studies in Epistemology. Midwest Studies in Philosophy*, Notre Dame, IN: University of Notre Dame Press, vol. 5.(The paper that introduced the concept of intellectual virtue into the contemporary literature.)

Sosa, E. (1985) ‘Knowledge and Intellectual Virtue’, *Monist* 68 (2) April: 226-45. (Describes Sosa’s important early version of virtue epistemology.)

Sosa, E. (1991) *Knowledge in Perspective*, Cambridge: Cambridge University Press. (This collection of essays is a valuable source on reliabilism and Sosa’s own version of virtue epistemology.)

Zagzebski, L. (1996) *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, Cambridge: Cambridge University Press. (Presents a virtue theory that is designed to handle both moral evaluation and epistemic evaluation within a single theory. It includes a theory of knowledge based on intellectual virtue.)

Virtue ethics

Virtue ethics has its origin in the ancient world, particularly in the writings of Plato and Aristotle. It has been revived following an article by G.E.M. Anscombe critical of modern ethics and advocating a return to the virtues.

Some have argued that virtue ethics constitutes a third option in moral theory additional to utilitarianism and Kantianism. Utilitarians and Kantians have responded vigorously, plausibly claiming that their views already incorporate many of the theses allegedly peculiar to virtue ethics.

Virtue theory*, the study of notions, such as character, related to the virtues, has led to the recultivation of barren areas. These include: What is the good life, and what part does virtue play in it? How stringent are the demands of morality? Are moral reasons independent of agents' particular concerns? Is moral rationality universal? Is morality to be captured in a set of rules, or is the sensitivity of a virtuous person central in ethics?*

From virtue ethics, and the virtue theory of which it is a part, have emerged answers to these questions at once rooted in ancient views and yet distinctively modern.

1 Aristotle and ancient virtue ethics

Modern virtue ethicists often claim Aristotle as an ancestor. Aristotle, however, was himself working through an agenda laid down by Plato and Socrates. Socrates asked the question at the heart of Greek ethics: 'How should one live?' All three of these philosophers believed that the answer to this question is, 'Virtuously' (see Virtues and vices §§1-3).

The ancient philosophical task was to show how living virtuously would be best for the virtuous person. Plato's *Republic* attempts to answer Thrasymachus' challenge that rational people will aim to get the most pleasure, honour and power for themselves. His argument is that justice, broadly construed, is to be identified with a rational ordering of one's soul. Once one sees that one identifies oneself with one's reason, one will realize that being just is in fact best for oneself. Thrasymachus, of course, might respond that he identifies himself with his desires.

Aristotle continued the same project, aiming to show that human *eudaimonia*, happiness, consists in the exercise (not the mere possession of) the virtues (see Eudaimonia; Happiness). The linchpin of his case is his 'function' argument that human nature is perfected through virtue, a standard objection to which is that it confuses the notions of a good man and the good for man. Ultimately, Aristotle's method is similar to Plato's. Much of *Nicomachean Ethics* is taken up with portraits of the virtuous man intended to attract one to a life such as his.

For Aristotle, all of the 'practical' virtues will be possessed by the truly virtuous person, the man of 'practical wisdom' (Aristotle's central 'intellectual' virtue). Socrates believed that virtue was a *unity*, that it consisted in knowledge alone. Aristotle's position is one of *reciprocity*: the possession of one virtue implies the possession of all. At this point he joined Socrates and Plato in their opposition to Greek 'common sense'. This opposition to common sense is not something that characterizes modern virtue ethics.

2 Modern virtue theory

Virtue theory is that general area of philosophical inquiry concerned with or related to the virtues. It includes *virtue ethics*, a theory about how we should act or live. This distinction is a rough one, but it is important to grasp that much of modern virtue theory is by writers not themselves advocating virtue ethics.

Virtue theory has undergone a resurgence since G.E.M. Anscombe's article 'Modern Moral Philosophy' (1958) (see Anscombe, G.E.M. §2). Anscombe believed that it was a mistake to seek a foundation for a morality grounded in legalistic notions such as 'obligation' or 'duty' in the context of general disbelief in the existence of a divine lawgiver as the source of such obligation (see Duty). She recommended that philosophy of psychology should take the place of moral philosophy, until adequate accounts of such central notions as action and intention were available. Then, she suggested, philosophers might return to moral philosophy through an ethics of virtue.

What is virtue ethics? It is tempting to characterize it as a theory advocating acting virtuously, but this is insufficiently precise. Virtue ethics is usually seen as an alternative to utilitarianism or consequentialism in general

(see Consequentialism; Utilitarianism). To put it roughly, utilitarianism says that we should maximize human welfare or utility. A utilitarian, however, may advocate acting virtuously for reasons of utility. Ethical theories are best understood in terms not of what acts they require, but of the reasons offered for acting in whatever way is in fact required.

Which properties of actions, then, according to virtue ethics, constitute our reason for doing them? The properties of kindness, courage and so on. It is worth noting that there is a difference between acting virtuously and doing a virtuous action. One's doing a virtuous action may be seen as doing the action a virtuous person would do in those circumstances, though one may not oneself be a virtuous person. Virtue ethics, then, concerns itself not only with isolated actions but with the character of the agent. There are reasons for doing certain things (such as kind things), and also for being a certain type of person (a kind person).

This account of virtue ethics enables us to distinguish it from its other main opponent, deontology or Kantianism (see Deontological ethics; Kant, I. §§9-10; Kantian ethics). A Kantian, for example, might claim that my reason for telling the truth is that to do so would be in accordance with the categorical imperative. That is a property of the action of telling the truth quite different from its being honest.

3 The good of the agent and the demandingness of morality

A pure form of virtue ethics will suggest that virtuous properties - 'thick' properties as opposed to thin properties such as 'rightness' and 'goodness' - of actions constitute our only reasons for performing them. Aristotle came close to this position, but it is perhaps more plausible to interpret him as claiming that the rationality of virtue lies in its promotion of the agent's *eudaimonia*. Aristotle's view is nevertheless radical. Since my *eudaimonia* consists only in the exercise of the virtues, I have no reason to live a non-virtuous life.

More common than pure forms of virtue ethics are pluralistic views according to which there are other reason-constituting properties, some perhaps of the kind advocated by utilitarians and Kantians. The open-mindedness of virtue ethics contrasts sharply at this point with what Bernard Williams (1985) has identified as the peculiar narrowness of focus in modern ethics (see Williams, B.A.O.). Considerations other than the moral are relevant to the question of how one should live. Modern virtue ethicists can thus adopt a position on the demandingness of morality between the extremes of Aristotle and their modern opponents. For they need claim neither that self-interest is constituted entirely by being moral nor that morality completely overrides self-interest.

Much of virtue theory has been concerned to develop Williams' criticism of utilitarianism and Kantianism that through their impersonality and impartiality, utilitarianism and Kantianism violate the *integrity* of moral agents. Philippa Foot has developed these critical arguments in a direction favourable to virtue ethics. According to both the principle of utility and the Kantian categorical imperative, moral reasons, being universal, are independent of the desires of agents. Foot (1978), impressed by the rationality of fulfilling one's own desires, has argued that moral reasons do depend on the desires of the agent, so that a person who acts consistently ungenerously may be described as ungenerous, but not necessarily as having any reason to act generously, unless they have a desire which would thereby be fulfilled. Foot is here expressing a doubt similar to Anscombe's about the possibility of ungrounded 'ought' judgments.

4 Agency and motivation

Imagine that you are thanking a friend for visiting you in hospital. She replies, 'Oh, it was nothing. It was obvious that morality required me to come.' This case, taken from an influential article by Michael Stocker (1976), and related to the discussion in the previous section concerning the demandingness of morality and the pervasiveness of the moral point of view, serves to illustrate an ideal of agency which lies implicit in much modern ethical theory (see Moral motivation; Moral realism). The unattractiveness of this ideal can be avoided by utilitarians, who may argue that thinking in the way your friend did about morality is likely to be self-defeating in utilitarian terms. Even the utilitarians, however, can be charged with missing the point. What is wrong with your friend is not that moral thinking like hers fails to maximize utility. Nevertheless, the case constitutes a far more serious problem for Kantians, given Kant's insistence on the explicit testing of courses of action using the categorical imperative, and his view that the moral worth of an action lies entirely in its being done out of a sense of duty.

Modern virtue ethicists such as Lawrence Blum (1980) have endeavoured to replace this conception of moral

agency with a virtue-centred ideal allowing agents to be moved directly by emotional concern for others. This ideal can once again be seen as emerging from Anscombe's attack on the notion of duty. A morality of duty is said to pay insufficient attention to the inner life: the dutiful agent is not doing, or feeling, enough (this criticism is an interesting counterpoint to the accusation that Kantianism is excessively demanding). The charge, then, is not only that modern moral theory fails to provide plausible justifying reasons for action, but that the motivational structure of what is clearly moral agency is quite different from what the theories lead us to expect. Moral agency consists at least partly in acting and feeling in ways prompted by bonds of partiality, requiring no further backing from impersonal ethical theory (see Friendship §4).

5 Universality and tradition

We have already seen how some virtue theorists link moral reasons to motivation. This is one route to a narrowing of the scope of moral reasons. Without any motivation to do so, there is no reason, say, to maximize utility or respect the moral law. Another route is followed by writers such as Alasdair MacIntyre who ground moral rationality in traditions.

MacIntyre's critique of modern ethical theory, outlined in *After Virtue* (1981), is the most stringent in virtue theory. He claims that present moral discussion is literal nonsense: we unreflectively use a mix of concepts left over from moribund traditions; since these traditions are incommensurable, arguments using concepts from rival traditions are irresoluble and interminable. MacIntyre does not, however, follow through the implications of this critique into a Nietzschean moral scepticism, advocating instead a return to Aristotelian virtue ethics (see Nietzsche, F.). The question of how he accomplishes this is an example of a puzzle about virtue theory present since Anscombe's original article: after the sustained critique of modern ethical theory by virtue theorists, is the remaining conceptual apparatus sufficiently strong to support an alternative prescriptive ethics?

The relationship between modern virtue ethicists and Aristotle is complex. Most virtue ethicists, including Foot and MacIntyre, combine an Aristotelian emphasis on the virtues with a modern scepticism about the possibility of an objective theory of the good for an individual. Likewise, MacIntyre's stress on the importance of context is quite Aristotelian, but the relativism to which this leads him would be anathema to Aristotle (see Moral relativism). MacIntyre claims that goods are internal to practices, and not assessable from some external point of view, while Aristotle believed that teleological reflection on universal human nature enabled one to identify those practices which are good *haplōs*, 'simply good'.

The relativism of modern virtue ethics has emerged also in political theory in the debate between communitarians, such as MacIntyre, and liberals (see Community and communitarianism; Liberalism). Along with a predilection for virtue-centred over rule-centred ethics go preferences for the local and particular to the universal, the specific to the general, the embedded to the abstracted, the communal to the individual, the inexplicit to the explicit, the traditional to the revised, the partial to the impartial. In MacIntyre's work, the notion of goods internal to diachronic practices and grounded in traditions is tied to a criticism of a free-floating liberal self, choosing goods from some Archimedean standpoint. MacIntyre's narrative conception of a self opens another door for the readmittance of the notion of character into moral philosophy.

The most serious problem for relativism has yet to be resolved in modern virtue ethics. What are we to say of practices and ways of life constituting internally coherent traditions, yet containing undeniably evil components?

6 Practical reason

One reason for the fading of the notion of character in ethical theory is that utilitarianism and Kantianism have commonly been developed as ethics of rules to resolve dilemmas. An argument against a rule-based ethics is found in Aristotle's discussion of the legal virtue of *epieikeia*, 'equity'. Rules will always run out in hard cases, and some sensitivity is required on the part of the judge to fill the gap between the law and the world. Likewise, for Aristotle, the virtuous man possesses *phronēsis*, 'practical wisdom', a sensitivity to the morally salient features of particular situations which goes beyond an ability to apply explicit rules.

This view has been revived in virtue ethics, by among others Iris Murdoch (1970) and John McDowell, in his article, 'Virtue and Reason' (1979). McDowell argues that we cannot postulate a world as seen by both the virtuous and the unvirtuous, and then explain the moral agency of the virtuous through their possessing some

special desire. Since moral rules run out, any object of desire could not be made explicit. McDowell uses Wittgenstein to support his claim that rational action does not have to be rule-governed (see Wittgenstein, L. §§10-12). This has clear implications for moral education: it should consist in enabling the person to see sensitively, not (or at least not only) in inculcating rigid and absolute principles. This is one of the strands in the feminist critique of modern ethical theory, itself closely tied to virtue theory. Writers such as Carol Gilligan (1982) argue that the moral sensibility of women is less rule-governed than that of men, and this has influenced the 'ethics of care' of, for example, Nel Noddings (see Feminist ethics §1; Moral education §§1-2).

The emphasis in virtue ethics on non-rational factors in moral motivation sits well with the notion of moral sensitivity. And this latter notion provides another standpoint from which one might criticize the basing of morality on the categorical imperative. As we have seen, Foot claims that immorality is not necessarily irrational, since moral reasons depend on the agent's desires. Writers such as McDowell who depict practical reason as perceptual can also deny that immorality is irrational. The unvirtuous lack not any capacity of the theoretical or calculative intellect, but moral sensitivity. Unlike Foot, however, McDowell would argue that this is in fact a failure to perceive genuine reasons for action independent of the agent's motivations.

Again, in McDowell, we see the pastiche, characteristic of virtue ethics, of ancient and modern: rationality is made to depend on social practice, and yet, as Socrates thought, virtue turns out to be a kind of knowledge.

See also: Cheng; Duty and virtue, Indian conceptions of; Impartiality; Life, meaning of; Mencius; Moral judgment

ROGER CRISP

References and further reading

Anscombe, G.E.M. (1958) 'Modern Moral Philosophy', *Philosophy* 33: 1-19.(Seminal article, critical of modern ethics and advocating return to the virtues.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. with notes by T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985.(Essential reading. The 'function argument' occurs in I 7. Books II-V concern the virtues of character, while practical wisdom is the topic of VI.)

Blum, L.A. (1980) *Friendship, Altruism and Morality*, London: Routledge & Kegan Paul.(Includes criticism of Kantian ethics, and a virtue-based alternative.)

Crisp, R. (1996) *How Should One Live? Essays on the Virtues*, Oxford: Clarendon Press.(Edited collection of commissioned papers on central issues in virtue theory.)

Crisp, R. and Slote, M. (1997) *Virtue Ethics*, Oxford: Oxford University Press.(A collection of well-known papers on the virtues, with introductory essay. Includes bibliography on various topics.)

Foot, P. (1978) *Virtues and Vices*, Oxford: Blackwell.(A collection of important papers. Chapter 1 is on virtues and vices, while 8-11 concern moral reasons for action.)

Gilligan, C. (1982) *In a Different Voice: Psychological Theory and Women's Development*, Cambridge, MA: Harvard University Press.(Suggests that feminine ethics may be less abstract and rule-governed than masculine ethics.)

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*, trans. L.W. Beck, *Foundations of the Metaphysics of Morals*, Upper Saddle River, NJ: Prentice Hall, 2nd edn, 1995.(Classic defence of the categorical imperative and the centrality of moral motivation by the sense of duty rather than by inclination.)

MacIntyre, A. (1981) *After Virtue*, London: Duckworth.(Influential critique of modern ethics, and advocacy of Thomistic virtue ethics.)

McDowell, J. (1979) 'Virtue and Reason', *The Monist* 62: 331-50.(Important article, combining themes from Aristotle and Wittgenstein. Difficult.)

Murdoch, I. (1970) *The Sovereignty of Good*, London: Routledge & Kegan Paul.(Profound reflections on modernity by a Platonist.)

Noddings, N. (1984) *Caring: A Feminine Approach to Ethics and Moral Education*, Berkeley, CA: University of California Press.(A version of feminist ethics, in which the virtue of caring is made central, as opposed to concern for moral obligation or duty.)

Plato (*c*.380-367 BC) *Republic*, trans. R. Waterfield, Harmondsworth: Penguin, 1994.(First major expression of virtue ethics.)

Stocker, M. (1976) 'The Schizophrenia of Modern Ethical Theory', *Journal of Philosophy* 73: 453-6.(Influential article, capturing much of the unhappiness with modern ethical theory.)

Williams, B. (1973) 'A Critique of Utilitarianism', in J.J. Smart and B. Williams, *Utilitarianism For and Against*,

Cambridge: Cambridge University Press.(A wide-ranging and suggestive series of criticisms of utilitarian ethics. Sections 3-5 concern integrity.)

Williams, B. (1985) *Ethics and the Limits of Philosophy*, Cambridge, MA: Harvard University Press and London: Fontana.(Important discussion and critique of modern ethics.)

Virtues and vices

The concept of a virtue can make an important contribution to a philosophical account of ethics, but virtue theory should not be seen as parallel to other 'ethical theories' in trying to provide a guide to action.

Modern accounts of the virtues typically start from Aristotle, but they need to modify his view substantially, with respect to the grounding of the virtues in human nature; the question of what virtues there are; their unity; and their psychological identity as dispositions of the agent. In particular, one must acknowledge the historical variability of what have been counted as virtues.

Aristotle saw vices as failings, but modern opinion must recognize more radical forms of viciousness or evil. It may also need to accept that the good is more intimately connected with its enemies than traditional views have allowed. Virtue theory helps in the discussion of such questions by offering greater resources of psychological realism than other approaches.

1 Virtues and theory

Ethical theories are standardly presented as falling into three basic types, centring respectively on *consequences*, *rights* and *virtues* (see Consequentialism; Deontological ethics; Rights; Virtue ethics). One way of understanding this division into three is in terms of what each theory sees, at the most basic level, as bearing ethical value. For the first type of theory, it is *good states of affairs*; for the second, it is *right action*; while virtue theory puts most emphasis on the idea of a *good person*, someone who could be described also as an ethically admirable person. The last is an important emphasis, and the notion of a virtue is important in ethics; but its importance cannot be caught in this way, as the focus of a theory which is supposedly parallel to these other types of theory. Consequentialist and rights theories aim to systematize our principles or rules of action in ways that will, supposedly, help us to see what to do or to recommend in particular cases. A theory of the virtues cannot claim to do this: the theory itself says that what one needs in order to do and recommend the right things are virtues, not a theory about virtues. Moreover, the thoughts of a virtuous person do not consist entirely, or even mainly, of thoughts about virtues or about paradigms of virtuous people. Indeed, they will sometimes be thoughts about rights or good consequences, and this makes it clear that thoughts about the good person cannot displace these other ethical concepts, since a good person will have to use some such concepts. 'Virtue theory' cannot be on the same level as the other types of theory.

An emphasis on virtues is important to moral philosophy for other reasons. Although it need not exclude cognitivism, it shifts attention from morality as a system of propositions or truths to its psychological (and hence, eventually, social) embodiment in individual dispositions of action, thought and emotional reaction. It draws attention to the variety of reasons for action and judgment that may play a part in ethical life, beyond the theorists' favourites, duty and utility (see Moral motivation §§1-3; Morality and identity §4; Morality and emotions). Such reasons will not typically embody virtue concepts themselves, or, still less, involve reflection on the agent's own virtues. But virtue theory can help to explain how considerations such as 'she needs it', for instance, or 'he relied on what you said', can function as an agent's reasons. An approach through the virtues also leaves room for the important idea that ethically correct action may be only partly codifiable and may involve an essential appeal to judgment (see Moral judgment §4; Universalism in ethics §3).

2 Beyond Aristotle: ground; content

The first systematic investigation of the virtues was made by Plato, in such works as *Gorgias* and the *Republic*, and it was extremely significant, for instance in setting the problem of the unity of the virtues (see §3). Plato also posed in a particularly challenging form questions about the value of virtues to their possessor. The classical account of the virtues, however, to which all modern treatments refer, is that of Aristotle (*Nicomachean Ethics*) (see Aristotle §22-5). Just because of the power and the influence of this account, it is easy to underestimate the extent to which a modern theory needs to distance itself from Aristotle. A modern account is likely to agree with Aristotle that virtues are dispositions of character, acquired by ethical training, displayed not just in action but in patterns of emotional reaction. It will agree, too, that virtues are not rigid habits, but are flexible under the application of practical reason. But there are at least four matters on which it is likely to disagree with Aristotle,

which may be labelled *ground*; *content*; *unity*; and *reality*.

Ground. Aristotle held that the virtues (for which the word in his language means only ‘excellences’ - see Aretē) had a teleological ground, in the sense that they represented the fullest development of a certain kind of natural creature, a nondefective male human being. No one now is going to agree with Aristotle that there are creatures who are biologically human beings but who are excluded from this full development by their nature as women or as ‘natural slaves’. Having abandoned his views about women and slaves, modern thinkers face the harder question of how far they agree with Aristotle about the natural basis of the virtues. This in turn raises the question of how strongly Aristotle’s own teleological view should be taken. On one interpretation, he had a comprehensive functional conception of the contents of the universe, with each kind of creature fitting into a discoverable overall pattern. On such a conception, substantial parts of the theory of the virtues will be discoverable by top-down systematic inquiry which will tell us what sorts of creatures human beings are, and hence what their best life will be (see Perfectionism; Teleological ethics). Other interpreters give a more moderate account of Aristotle’s enterprise, according to which his intentions will be honoured by a hermeneutical inquiry into what we, now, regard as the most basic and valuable aspects of human beings.

Content. What is undeniably lacking from Aristotle’s thought, as from that of other ancient thinkers, is an historical dimension. Some modern virtue theorists share this weakness. Aristotle’s account is in several respects different from any account of the virtues one would give now, with respect both to what it puts in and what it leaves out. He gives a particularly important place to a quality called *megalopsuchia*, ‘greatness of soul’, which has a lot to do with a grand social manner and which bears even less relation to a contemporary ethic than its name, in itself, might suggest. A modern person, asked for the principal virtues, might well mention kindness and fairness. Fairness bears a relation to an important Aristotelian virtue, justice, but the latter is defined to an important extent in political and civic terms, and gives a fairly restricted account of fairness as a personal characteristic. Kindness is not an Aristotelian virtue at all. Moreover, there is no account of an important modern virtue, truthfulness; what Aristotle calls the virtue of truth is (surprisingly, as it seems to us) concerned exclusively with boasting and modesty.

There has been obvious historical variation in what is seen as the content of the virtues. Aquinas, who notably developed Aristotle’s account, of course modified it to accommodate Christianity, holding in particular that besides the moral virtues, there were ‘theological’ virtues, which have God as their immediate object (see Theological virtues). The pagans were not in a position to display these, but so far as the moral virtues were concerned, they could be truly virtuous in the light of natural reason. However, there was still something imperfect about their virtue even at this level since, Aquinas held, the whole of ethical life is properly grounded in the virtue called charity, which has a divine origin (see Charity).

For Hume (1751), on the other hand, Aristotle’s account and other pagan sources served to support an ethics of the virtues that was precisely designed to discredit and exclude Christianity (see Hume, D. §4). The historical variation, both in philosophical formulations and in cultural realizations of the virtues, raises wider issues of how theories of the virtues are to be understood. The conceptions of human nature and human circumstances that underlie such theories are open to wide reinterpretation in the face of changing values, and the Aristotelian presupposition that an understanding of human nature could yield a determinate account of the virtues - even if that idea is interpreted relatively unambitiously - looks unrealistic. There are of course constants in the psychology and circumstances of human beings that make certain virtues, in some version or other, ubiquitous: in every society people need (something like) courage, (something like) self-control with regard to anger and sexual desire, and some version of prudence. These platitudes, which are stressed by those who look to a substantive universal virtue theory, severely underdetermine the content of such a theory. This is shown by the very simple consideration that the constant features of human life are indeed constant, but the virtues that have been recognized at different times and by different cultures vary considerably.

3 Beyond Aristotle (continued): unity; reality

Unity. Aristotle inherited from Plato, and ultimately from Socrates, an interest in the unity of the virtues (see Socrates §5). Socrates seems to have held that there was basically only one virtue, which he called wisdom or knowledge. The conventional distinctions between the various virtues - justice, self-control, courage and the rest - were taken to mark only different fields of application of this power. Aristotle did think that there were separate

virtues, but nevertheless his view came almost to the same thing as Socrates', since he thought that one could not have one virtue without having them all. One could not properly possess any one virtue unless one had the intellectual virtue which is called in Aristotle's language *phronēsis* (often translated as 'practical wisdom', but better rendered as 'judgment' or 'good sense'); but, Aristotle held, if one had this quality, then one had all the virtues.

It is not hard to see the general idea underlying this position. Generosity is linked to justice - someone who gives only what justice demands is not being generous. Similar points can be made about the interrelations of some other virtues. However, it is important to the theory of the virtues that they provide psychological explanations as well as normative descriptions, and from a realistic psychological point of view it is hard to deny (as many ancient Greeks other than Socrates and Aristotle agreed) that someone can have some virtues while lacking others. In particular, the so-called 'executive virtues' of courage and self-control can be present without other virtues; indeed, they themselves can surely be deployed in the interests of wicked projects. The refusal to acknowledge this may simply represent an ethical reluctance to give moral accolades to bad people.

The fact that the virtues can, to some degree, be separated from one another itself helps to give point to virtue theory. Some modern ethical theories do imply that there is basically only one moral disposition. Utilitarianism, at least in its direct form, places everything on impartial benevolence (see Utilitarianism; Impartiality); and though Kant himself did have a theory of the virtues, Kantianism insists on the primacy of a sense of duty (see Kant, I. §§9-11; Kantian ethics). An advantage of virtue theory is that it allows for a more complex and realistic account of ethical motivation.

Relatedly, it can acknowledge psychological connections between the ethical and other aspects of character, accepting that people's temperaments have something to do with how they conduct themselves ethically. For the same reason, virtue theory is implicitly opposed to sharp boundaries between the 'moral' and the 'nonmoral', and is likely to acknowledge that there is a spectrum of desirable characteristics, and that no firm or helpful line can be drawn round those that are specifically of moral significance. Aristotle did not even try to draw such a line: his own terminology distinguishes only between excellences of character and intellectual excellences, and one of the latter, *phronēsis*, is itself necessary to the excellences of character. Hume, who, unlike Aristotle, was surrounded by moralists who wanted to draw such a line, goes out of his way to mock the attempt to draw it, and his deliberately offensive treatment of the subject is still very instructive.

Reality. Aristotle conceived of the virtues as objective dispositional characteristics of people which they possess in at least as robust a sense as that in which a magnet possesses the power to attract metals, though people, unlike magnets, have of course acquired the dispositions - in the way appropriate to such things - by habituation (see Moral education). Modern scepticism, however, to some extent supported by social and cognitive psychology, questions whether we can take such a naïve view of what it is for someone to have a virtue. There are at least two different sources of doubt. One is the extent to which people's reactions depend on situation: it is claimed that they will act in ways that express a given virtue only within a rather narrow range of recognized contexts, and if the usual expectations are suspended or even, in some cases, slightly shifted, may not act in the approved style.

The other doubt concerns ascription. When we understand people's behaviour in terms of virtues and vices, or indeed other concepts of character, we are selecting in a highly interpretative way from their behaviour as we experience it, and the way in which we do this (as, indeed, we understand many other things) is in terms of stereotypes, scripts, or standard images, which may range from crude 'characters' to sophisticated and more individuated outlines constructed with the help of types drawn, often, from fiction. The available range of such images forms part of the shifting history of the virtues. At different times there have been pattern books of virtue and vice, and one of the first was the *Characters* written by Theophrastus, a pupil of Aristotle (see Theophrastus; Examples in ethics).

Even assuming such ideas to be correct, it is not clear exactly to what extent they have a negative impact on virtue theory. Everyone knows that virtues do not express themselves under all circumstances, and also that agents may be very rigid in their ability to understand how a situation is to be seen in terms of virtues. Again, with regard to ascription, it is very important that if it is true that we construct our interpretations of another person's character in terms of a stock of images, it is equally true that the other person does so as well. The point is not so much that there is a gap between the interpreter and the person interpreted, but rather that all of us, as interpreters of

ourselves and of others, use shared materials that have a history. There are lessons in such ideas for ethics generally and for virtue theory, but they need not be entirely sceptical. The points about the situational character of the virtues and about their ascription serve to remind us that an agent's virtues depend in many different ways on their relations to society: not simply in being acquired from society and reinforced or weakened by social forces, but also in the ways in which they are constructed from socially shared materials.

4 Vices, failings and evil

Aristotle named a variety of vices, each of which was basically constituted by the absence of the restraining or shaping influence of virtue, together with the operation of some natural self-centred motive. Thus cowardice was the disposition, in the absence of courage, to give in to fear; self-indulgence and irascibility the dispositions to give in to bodily pleasure and to anger. In this range of what may be called 'failings', actions that are expressions of vices do have distinctive motives, but those are not in themselves distinctively bad motives: it is rather that natural motives are expressed in ways in which they would not be expressed by a virtuous person. There are other failings in which the agent's motivation is distinctively deplorable, because it is constituted by the exaggeration, parody or perversion of a virtue: an ostentatious disposition to distribute gifts or favours, in place of generosity, or, to take a modern example, sentimentality in place of kindness. Aristotle notices some failings of this type, but, in line with his 'doctrine of the mean', he oversimplifies their psychology under an unexamined category of 'excess'.

A peculiar case, in Aristotle's treatment, is justice. At the level of actions, at least, it might be thought that there were no distinctive motives to injustice; a person can act unjustly from a variety of motives, and indeed Aristotle mentions the possibility that a coward might treat others unjustly, by 'getting an unfair share of safety'. If this is generalized, an unjust person might be understood not as one with some characteristic motive, but rather as one who is simply insensitive to considerations of justice. However, Aristotle does introduce a distinctive motive for injustice - 'greed' or the desire to have more than others. An unjust person, then - as opposed to someone who has some other vice as a result of which he acts unjustly - is, for Aristotle, a particular greedy type, one who might roughly be recognized in modern terms as 'a crook'.

Aristotle also notices another kind of failing or deficiency, a lack of perception or feeling for others, but this is typically registered by him only as an extreme characteristic, lying off the scale of the ethical, in the form of a brutality or beastliness which virtually falls out of the category of the human. The fact that he does not have anything to say about the more domesticated forms of such a failing, very familiar to us, is a corollary to his not recognizing a virtue of kindness.

It follows from Aristotle's holistic and teleological conception of virtue as the fulfilment of the highest human capacities that vices should be basically failings, instances of a lack or an absence. This hardly leaves room for a notion of the *vicious*: the nearest that Aristotle gets to such an idea is the figure of an obsessional and unscrupulous hedonist. We possess, only too obviously, notions of viciousness deeper and more threatening than this. They point to a concept conspicuously lacking from Aristotle (though to a lesser extent, perhaps, lacking from Plato) - the concept of evil (see Evil).

This leads decisively beyond the conception of vices as failings, even very serious failings. Among evil or vicious motivations, a basic type is cruelty, the desire to cause suffering, a disposition which, as Nietzsche pointed out, contrasts markedly with brutality: it has to share, rather than lack, the sensitivity to others' suffering that is displayed by kindness. In the most typical modes of cruelty, agents derive their pleasure from the sense of themselves bringing about the pain or frustration of others, and their cruel behaviour is directly an attempted expression of power. Rather different from this, though close to it, is maliciousness, as it might be called, a motivation in the style of envy, where the desire is merely that other people's happiness should not exist. Persons in this state of mind may be pleased if others come to grief, even though they do not bring it about themselves. Alberich, in Wagner's *Götterdämmerung*, says, 'Hagen, mein Sohn! hasse die Frohen!' - 'hate the happy'; such a hatred can have many expressions, only some of which involve the specifically active pleasures of cruelty.

Sometimes, cruelty may not only share, as it must, the perceptions that kindness uses, but model itself negatively on kindness, calculating what a kind person might want to do, in order to parody or subvert it. It then takes on the character of perversity. This style of reversal can be applied to virtues other than kindness. There is counter-justice, the disposition to frustrate the ends of justice, not simply in one's own interests, or to hurt or

frustrate a particular person whom one hates or envies, but to take pleasure in the frustration of justice as such and the disappointments inflicted on those of good will. At the limit, this can constitute an almost selfless aesthetic of horribleness, one of the less obvious forms that may be taken by the satisfactions of Milton's Satan, with his resolve that evil should be his good.

5 Links between virtue and vice

Unlike the failings recognized by Aristotle, these evil motivations are more than mere negations. It is important, however, that this need not be taken as a metaphysical claim: one need not be committed to a Manichean view (or even the very various compromises with such a view that have been negotiated by orthodox Christianity), to the effect that human nature or the world itself contains some perversely destructive principle (see Manicheism). One might, for instance, hold, as some optimistic programmes of psychotherapy do, that vicious and cruel motivations are, indeed, perversions, produced by a failure of love or other deficiency in the individual's upbringing. This is an encouraging position, inasmuch as it holds out the hope of a world free of such motivations, but it does not think that such motivations, while they exist, are to be understood simply in terms of the lack of a shaping or restraining influence. It would accept that vicious motivations were specially and inventively active.

Other psychological and social views are less hopeful. It is not simply that they see no ground for Utopian hopes that the world could ever be freed from vicious motivations. Some of them detect deeper ways in which virtue, and more generally the good, depend on their opposites. At the most superficial level, there are contemporary versions of the point made in Mandeville's *Fable of the Bees* (1714) (see Mandeville, B. §2): many benefits, including ethical benefits, have come from the development of commercial society, but there is no known way of replacing greed as a means of sustaining it. At another level, there is no doubt that valuable human achievements, for instance in the arts and sciences, have come about only because of a certain indifference to values of justice and benevolence, both at an institutional level and in the lives of those who have brought about these achievements. (Here, as so often, moralists have to face the question whether or not they are relieved that the values which they think should prevail have not always done so.)

At the deepest level, however, it is not a question simply whether nonethical values may often require the neglect or denial of morality, but whether morality itself does not require it. One of the metaphysicians' illusions, Nietzsche said (1886), is 'the belief in opposing values'. In fact, he thought, moral values will always turn out to implicate their opposites - historically (in terms of how new moral values come to exist), socially (in terms of how they sustain themselves), and psychologically (in terms of how they are learned and of how they derive their energy).

Even if we accept the force of the Nietzschean suspicions, this need not damage, but rather encourages, the project of thinking about morality in ways that give an important place to virtues and vices. A theory of virtues, handled in a truthful way, offers better hope of being psychologically realistic than other prominent pictures of the ethical life do. If, further, it extends its realism to the motivations of immorality as well, and does not treat them as mere negations of the moral dispositions, it will better understand morality itself. It will be more successful in this than other theories of morality, which usually pass over in silence the forces that oppose it, or register them simply as objects of moral disapproval, or treat them as the products of a (typically unexplained) cognitive failure.

See also: Help and beneficence; Human Nature; Justice; Mencius; Moral judgment §4; Nietzsche, F.; Prudence; Self-control; Truthfulness

BERNARD WILLIAMS

References and further reading

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. W.D. Ross, revised by J. Urmson, ed. and revised by J. Barnes in *The Complete Works of Aristotle*, vol. 2, Princeton, NJ: Princeton University Press, 1984. (Generally accepted as the most considered statement of his position. For secondary works, see Broadie (1991) and Sherman (1989) below.)

Baier, A.C. (1991) *A Progress of Sentiments: reflections on Hume's Treatise*, Cambridge, MA: Harvard University Press.(A thoughtful consideration of Hume's moral philosophy, useful on the relations between virtue and sentiment.)

Broadie, S. (1991) *Ethics with Aristotle*, New York: Oxford University Press.(A notably subtle and

philosophically helpful commentary.)
Crisp, R. and Slote, M. (eds) (1997) *Virtue Ethics*, Oxford: Oxford University Press.(A helpful collection of papers.)
Flanagan, O. (1991) *Varieties of Moral Personality: Ethics and Psychological Realism*, Cambridge, MA: Harvard University Press.(Relates the moral psychology of various ethical positions to empirical material.)
French, P.A., Uehling, T.E. and Wettstein, H.K. (eds) (1988) *Ethical Theory: Character and Virtue*, Midwest Studies in Philosophy, vol. 13, Notre Dame, IN: University of Notre Dame Press.(A useful collection of papers on contemporary virtue theory.)
Hume, D. (1751) *An Enquiry Concerning the Principles of Morals*, in *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, revised by P.H. Nidditch, Oxford: Clarendon Press, 3rd edn, 1975.(A more compact, though also less searching, account than book III of *A Treatise on Human Nature* (1739-40). Appendix IV discusses the idea of ‘moral’ virtue.)
MacIntyre, A. (1981) *After Virtue*, Notre Dame, IN: University of Notre Dame Press.(An influential, negative, assessment of modern moral ideas in contrast to Aristotelian and medieval virtue theory.)
Mandeville, B. (1714) *The Fable of the Bees: or, Private Vices, Publick Benefits*, ed. F.B. Kaye, Indianapolis, IN: Liberty Fund, 1988.(The work was first intended as a political satire, but has been seen as offering a serious case for private vices turned into public benefits.)
Nietzsche, F. (1886) *Jenseits von Gut und Böse*, trans. R.J. Hollingdale, *Beyond Good and Evil*, Harmondsworth: Penguin, 1990.(Considers, among many other things, the ‘faith in opposing values’.)
Nietzsche, F. (1887) *Zur Genealogie der Moral*, trans. C. Diethe, *On the Genealogy of Morality*, Cambridge: Cambridge University Press, 1994.(A powerfully influential study, phenomenological rather than historical, of moral values - including, importantly, the passion for truthfulness which motivates the work itself.)
Plato (*c*.395-387 BC) *Gorgias*, trans. D.J. Zeyl, Indianapolis, IN: Hackett Publishing Company, 1987.(His most dramatic and radical enquiry into scepticism about the virtues.)
Plato (*c*.380-367 BC) *Republic*, trans. G.M.A. Grube, revised by C.D.C. Reeve, Indianapolis, IN: Hackett Publishing Company, 1992.(His fullest account, political as well as ethical, of the nature and value of the virtues.)
Sherman, N. (1989) *The Fabric of Character: Aristotle’s Theory of Virtue*, Oxford: Clarendon Press.(A useful philosophical and interpretative discussion.)
Westberg, D. (1994) *Right Practical Reason*, Oxford: Clarendon Press.(A discussion of practical reason and virtue in Aquinas.)

Vision

Vision is the most studied sense. It is our richest source of information about the external world, providing us with knowledge of the shape, size, distance, colour and luminosity of objects around us. Vision is fast, automatic and achieved without conscious effort; however, the apparent ease with which we see is deceptive. Since Kepler characterized the formation of the retinal image in the early seventeenth century, vision theorists have known that objects do not look the way they appear on the retina. The retinal image is two-dimensional, yet we see three dimensions; the size and shape of the image that an object casts on the retina varies with the distance and perspective of the observer, yet we experience objects as having constant size and shape. The primary task of a theory of vision is to explain how useful information about the external world is recovered from the changing retinal image.

Theories of vision fall roughly into two classes. Indirect *theories characterize the processes underlying visual perception in psychological terms, as, for example, inference from prior data or construction of complex percepts from basic sensory components.* Direct *theories tend to stress the richness of the information available in the retinal image, but, more importantly, they deny that visual processes can be given any correct psychological or mental characterization. Direct theorists, while not denying that the processing underlying vision may be very complex, claim that the complexity is to be explicated merely by reference to non-psychological, neural processes implemented in the brain.*

The most influential recent work in vision treats it as an information-processing task, hence as indirect. Computational models characterize visual processing as the production and decoding of a series of increasingly useful internal representations of the distal scene. These operations are described in computational accounts by precise algorithms. Computer implementations of possible strategies employed by the visual system contribute to our understanding of the problems inherent in complex visual tasks such as edge detection or shape recognition, and make possible the rigorous testing of proposed solutions.

1 Historical background

The fact that the image on the retina does not correspond in an obvious manner to the way things look suggests that some processing of the stimulus occurs in visual perception. Theorists of vision have proposed various accounts of the nature of the processing responsible for our perception of size, shape and distance. *Geometric* models, popular among optic theorists in the seventeenth century, and suggested in some of René Descartes' work on vision (particularly, his Sixth Set of Replies 1641: §9), construe visual processing as a species of mathematical calculation (see Molyneux problem). Geometric models can therefore be seen as precursors of modern-day computational models of vision (see §§4-7 below). According to one geometric model, the visual system computes the distance of an object in the visual field from the angles at which the light from the object strikes each eye, and the distance between the two eyes. The knowledge required for the calculation, including knowledge of the precise angles formed by the eyes when looking at an object, and knowledge of the relevant mathematical theorems, was thought to be provided by innate mechanisms, rather than acquired from experience (see Innate knowledge). A significant defect of geometric models is that they failed to provide an account of how the requisite knowledge is made available to the visual system, or how it is deployed in calculations that were presumed to be unconscious.

The philosopher George Berkeley, in his influential 'Essay Towards a New Theory of Vision' (1709), questioned the psychological reality of the geometric models, arguing in effect that the information upon which the postulated calculations are based is not available to the visual system. Berkeley agreed with the geometric theorists that retinal information alone is insufficient to account for our perception of distance and size, but, consistent with his more general empiricism, he claimed that the process by which we acquire such knowledge is not a species of calculation based on innately specified information, but rather associative and learned.

According to Berkeley, our ideas of distance and size, unlike our ideas of colour, are not really *visual* ideas at all. Whereas light reflected at different wavelengths affects the retina differentially, and so gives rise to different colour sensations, light reflected from different distances does not. There is no characteristic retinal pattern associated with something's being 10 feet away. As Berkeley put it in his famous 'one point' argument, 'distance being a line directed end-wise to the eye it projects only one point in the fund of the eye, which point

remains invariably the same whether the distance is larger or shorter' (1709: §2). Similarly for size: there is no characteristic retinal pattern produced by our looking at an object that is 6 cubic feet in volume. A larger object placed at a greater distance along the line of sight will have the same geometric effect on the retina. Our ideas of distance and size, Berkeley concluded, derive not from visual experience, but from movement and touch; from our tangible acquaintance with objects. We can tell the distance and size of objects by sight only because we learn to associate visual cues, including sensations caused by the convergence of the eyes and the accommodation of the lens, with ideas originally derived from our tangible sense.

Central to Berkeley's account of distance and size perception is the empiricist doctrine that there are no meaningful abstract ideas, that is, ideas not reducible to sensation (see Empiricism; Sense-data). He rejected the possibility that we might possess abstract spatial ideas that are shared by visual and tangible experience (see Molyneux problem). Later theorists of vision who do not share Berkeley's epistemological and metaphysical assumptions have found his claim that our ideas of distance and size are derived from our sense of touch uncompelling; however, his discussion of the phenomena to be explained by a theory of vision shaped the field well into the twentieth century.

2 Direct v. indirect perception

The claim that visual perception is not direct or immediate involves more than the truism that some processing of the retinal image is necessary to account for what we see. Ideas or perceptions are thought not to be 'direct' if they are produced by *psychological* processes. While the notion of a psychological process admits of no precise definition, examples come readily to mind. Any process that occurs in consciousness, such as the association of ideas, is a psychological process, as is any process that involves learning. Mathematical calculation of distance and size based on the prior *representation* of lines and angles, whether accessible to consciousness or not, is a psychological process (see Intentionality). Since Berkeley's theory and the models of the geometric writers both posit psychological processing of the image (albeit of different sorts), they are considered indirect theories of vision.

The difference between direct and indirect accounts of perception has been characterized as a disagreement over the richness of the stimulus, with direct theorists typically arguing that the stimulus contains more information than indirect theorists have been willing to allow. For example, James J. Gibson (1904-79), a prominent direct theorist, claimed that the input to the visual system is not a series of static 'time slices' of the retinal image, but rather, the smooth transformations of the optic array as the subject moves about its environment (what Gibson (1979) called 'retinal flow'). But to characterize the fundamental difference between direct and indirect theories as a disagreement over the richness of the stimulus is to misplace the dispute. The issue that separates the two camps concerns neither the amount of information contained in the stimulus, nor even the precise character of this information, but, rather, how the information in the stimulus is accessed and used by the visual system to produce knowledge that is useful to the organism. In other words, it concerns the character of the intervening processes. Direct theorists deny that visual processes can be characterized in terms of ideas, beliefs, representations, knowledge or memories. In other words, they deny that visual processes have any true psychological description. A direct theory explicates any intervening or supplementary processing that occurs in perception in terms of neural structures and processes directly implemented in the brain. Indirect theorists, of course, do not deny that perceptual processes are implemented in neural structures, but they argue that such processes should be characterized at a distinct, psychological, level of description.

Direct theories of perception are sometimes explicitly contrasted with accounts that treat perception as a species of inference, akin to the drawing of a conclusion from premises according to a principle or rule. The nineteenth-century German physicist and physiologist Hermann von Helmholtz argued that the processes underlying visual perception are of the same general sort as inductive generalization employed in scientific reasoning (see Inductive inference; Inference to the best explanation). The contemporary psychologist Irvin Rock (1983) advances a view that explicitly treats much of perception as a process of hypothesis generation and testing. But the use of 'inferential' as a blanket term to refer to indirect theories of perception is somewhat misleading. The various processes that can be thought of as psychological (for example, conscious inference, unconscious calculation, habit-based association, and so on) seem too heterogeneous a collection to justify characterizing the entire class in terms of the drawing of conclusions from antecedently established premises.

3 Direct theories of vision

The ‘Gestalt’ movement of the early twentieth century rejected the view, prevalent since Berkeley, that complex percepts can be analysed into simple sensory components (see Gestalt psychology). According to the Gestalt theorists, perception is *holistic*: perceptual wholes are not built up out of more basic sensory elements, in the way, for example, that a painting is just the combination of all the paint-covered segments of the canvas. Gestalt theorists claimed further that perception is direct - perceptual processing is not correctly described in terms of psychological or mental processes. The structure of a visual experience is to be explicated in terms of the structure of the underlying brain states, that is, in neurophysiological terms. The Gestalt psychologist Wolfgang Köhler characterized as a *physical gestalt* any dynamic system that settles into an equilibrium state of minimal energy. A soap bubble forming a perfect sphere is an example of a physical gestalt, as is, Köhler argued, the brain producing an organized percept. Köhler proposed a theory that appealed to electrical fields within the brain to account for perception (and all other mental processes). Gestalt speculative physiology was not borne out by subsequent brain research, which failed to discover evidence of Gestalt mechanisms implicated in perceptual processing.

The psychologist James J. Gibson shared with the Gestalt theorists the belief that visual perception is not mediated by processes characterizable in psychological terms. Gibson argued that indirect theorists have mischaracterized the information in the optical array. If the effective stimulus for the visual system is taken to be retinal flow (the smooth transformations of the optic array as we move about), then, according to Gibson, there are important constancies in the stimulus that indirect theorists have typically missed. There is therefore no need to posit inferences, calculations, memories, association of ideas, or any other intervening psychological process, to explain our perception of size and shape constancy. In addition to brightness and colour, properties directly picked up in the stimulus include, according to Gibson, higher-order properties that remain invariant through movement and changes in orientation. These higher-order invariants specify not only structural properties such as ‘being a cube’, but also what Gibson called ‘affordances’, which are functionally significant aspects of the distal scene, like the fact that an object is edible or could be used for cutting.

Two fundamental assumptions underlie Gibson’s ‘ecological optics’: (1) that functionally significant aspects of the environment structure the ambient light in characteristic ways; and (2) that the organism’s visual system has evolved to detect these characteristic structures in the light. Both assumptions are controversial. With respect to (2), indirect theorists have complained that Gibson provides no account of the mechanism that allegedly detects salient higher-order invariants in the optical array. His claim that the visual system ‘resonates’, like a tuning fork, to these properties is little more than a metaphor. But it should be noted that in claiming that perception of higher-order invariants is direct, Gibson is simply advocating that the mechanism be treated as a black box, from the point of view of psychology, because no inferences, calculations, memories or beliefs mediate the processing. (The physiological account of the mechanism’s operation will no doubt be very complex.) This claim might be plausible if assumption (1) is true - if there is a physically specifiable property of the light corresponding to every affordance. But for all but the simplest organisms it seems unlikely that the light is structured in accordance with the organism’s goals and purposes. More likely, the things that appear to afford eating or cutting or fleeing behaviour structure the light in all kinds of different ways. This likelihood has led indirect theorists to claim that something like categorization - specifically, the bringing of an object identified initially by its shape, colour or texture under a further concept - is at work when an organism sees an object *as* food, *as* a cutting implement, or *as* a predator.

4 Computational models of vision: general approach

The predominant theoretical approach in cognitive psychology in recent years has been computationalism, which treats human cognitive processes, including perceptual processes, as a species of information processing (see Mind, computational theories of). Computational theories of vision attempt to specify the aspects of the external world that are represented by the visual system, and to characterize the operations that derive these representations from the information contained in the retinal image.

One of the most prominent early computational vision theorists was David Marr (1945-80), a researcher in the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology. While the details of Marr’s specific computational model have been challenged by later theorists, his work is of continuing interest to philosophers and psychologists concerned with the foundations of the computational approach to vision.

Marr argued in his book *Vision* (1982) that an information-processing capacity can be analysed at three distinct levels of description. The 'theory of the computation' is a precise specification of the function computed by the mechanism, in other words, what the mechanism does. For example, the theory of the computation for a particular device may tell us that it adds numbers, or computes averages when given a list of numbers as input. The *algorithm* specifies the procedure or rule for computing the function, and the *implementation* level describes how the computation is carried out in neural or computer hardware. The first two levels in the hierarchy - the abstract characterization of the problem and the rule for its solution - exemplify a fundamental commitment of the computational approach: that cognitive processes can be understood in a way that is independent of the particular mechanisms that implement them in the brain.

Computational models treat the visual system as computing from the retinal image a representation of the three-dimensional structure of the distal scene. Marr's theory divides this process into three distinct stages, positing at each stage the construction of a representation that makes explicit (some of) the information contained in the image and represents it in a way that is efficient for later use. Various computational processes, some running in parallel, are defined over these representations. The algorithmic level of description characterizes the procedures the visual system uses to produce increasingly more useful representations of the scene.

Most of the processes that Marr describes are *data driven*, or 'bottom up' - they operate on information contained in the image, without supplementation by high-level information or beliefs about the world that the subject may have. These processes use information about intensity changes across the visual field, or the orientation of surfaces, not such facts as that objects of a particular shape typically make good cutting implements. Marr advocated 'squeezing every ounce of information out of the image' before positing the influx of supplementary knowledge.

Data-driven models of perception have a number of advantages over *hypothesis-driven* models which appeal to high-level knowledge very early in visual processing. Data-driven processes are generally faster - the visual system does not have to retrieve the relevant piece of specialized knowledge before processing the information in the image - and tend to be more reliable. In Marr's model, the point at which high-level information is available to the visual system marks a distinction between *early* and *late* vision. Early visual processes are said to be 'cognitively impenetrable' by the subject's beliefs about the world (see Modularity of mind). As a consequence, they cannot be influenced by learning.

Marr emphasized the importance of the 'topmost' level of description - the theory of the computation - in developing accounts of human cognitive capacities. He noted that there is no point attempting to describe how a mechanism works before knowing what it does. A crucial first step in constructing a theory of a perceptual capacity is discovering constraints on the way the world is structured that enable adapted organisms to solve perceptual problems in their normal environments. An example should make the point clear. Marr's student and co-worker Shimon Ullman (1979) has proved that three distinct orthographic views of four non-coplanar points are sufficient to determine the three-dimensional structure of a rigid body (the 'structure from motion' theorem). If a body is not rigid, much more information is required to compute its shape. In a world such as ours, where most things are relatively rigid, a visual system built (that is, adapted) to assume that the objects in its environment are rigid would be able to compute the structure of those objects more easily and quickly than a visual system that had to consider the many non-rigid interpretations consistent with the data. Accordingly, Marr posited a mechanism that given three views of four non-coplanar points as input selects the unique rigid interpretation consistent with the data.

Recall Berkeley's objection to the geometric theorists' accounts of size and distance perception. He claimed that the information required for the postulated calculations was not generally available to the visual system, nor to the organism. Such a criticism, if true, is devastating for a computational account of a cognitive capacity. Any computational theory that posits processing beyond the computing capabilities of the mechanism, or that relies on information unavailable to the mechanism, is a non-starter as a biological model. An important lesson of Marr's work is that the theorist must attend to the structure of the organism's environment before attempting to characterize computational mechanisms, because the environment determines the nature of the computational problems that the organism's visual system needs to solve. The perceptual systems of adapted organisms can be assumed to 'exploit' very general information about the environment. Consequently, the problems they have to

solve may be simpler and computationally more tractable than might initially be assumed.

5 Computational models of vision: modularity

Another characteristic feature of Marr's theory is that it treats the visual system as comprising a number of individual components or modules that can be analysed independently of the rest of the system. A 'module' is, by definition, cognitively impenetrable: its operation is not influenced by information external to it that may be available to the cognitive system as a whole, for example, information in the system's memory (see Modularity of mind). Marr posited a module responsible for computing three-dimensional structure from apparent motion, another for computing depth from disparity information available in stereo images, a third for computing shape from shading. Each of these modules is designed to exploit general environmental constraints in the manner that the 'structure from motion' module, described above, incorporates the rigidity assumption.

The various modules operate in parallel, and since they yield information about the depth of the distal scene from different input data, they may give inconsistent results. This is an advantage for the organism, because in cases where the general environmental constraints assumed by a processing module do not hold, the output of the module is subject to correction by another module operating on different data, and exploiting different environmental constraints. For example, imagine a non-rigid mass of jelly moving through space. Since the 'structure from motion' module is built to assume rigidity it will probably give an incorrect interpretation of the jelly's structure. But its output is then likely to be inconsistent with, and correctable by, the output of modules operating on shading or disparity information, which, though they exploit other environmental constraints, do not assume rigidity.

The principle of modular design has an evolutionary rationale. Modular processes are typically fast, because a time-consuming search of general memory is avoided. And assuming that the constraints governing a module's operation are generally true, the process will normally be reliable. Commitment to the principle of modular design makes the computational theorist's job easier, since modular processes can be studied and modelled without the theorist knowing how more central reasoning systems work. For all their theoretical advantages, however, modules do pose a general problem. The theorist has to explain how the outputs of various modular processes are combined in a single representation of the structure of the scene. The possibility of inconsistent results from different modules suggests that this is a non-trivial problem.

In general, then, the visual processes posited in Marr's theory have three important features. They are data-driven, adapted to exploit general environmental constraints, and modular. The visual system, according to Marr, computes a series of intermediate representations of distal information, culminating in a representation of the three-dimensional structure of the scene. The input to the system is the image on the retina, in effect, a grey-level intensity array. The initial processing of the image produces what Marr called the 'primal sketch', a representation of the way that light intensities change over the visual field. The primal sketch makes explicit precisely the information that is required for subsequent processing. Discontinuities in intensity tend to be correlated with significant features of the scene, that is, object boundaries, although it is too early at this stage to assume that all sharp intensity changes in the image indicate edges in the world. Some may be produced by changes in illumination or surface reflectance (see Colour and qualia).

The various processing modules described above operate on aspects of the information contained in the primal sketch. The results are encoded in a representation that Marr called the '2.5-dimensional sketch'. It makes explicit the depth and surface orientation of the scene, and is the input representation for later visual processing. The visual system is assumed to be cognitively impenetrable up to the production of the 2.5-dimensional sketch, hence its operation to this point cannot be influenced by learning.

6 Computational models of vision: object recognition

Late or high-level visual processes use the representations of depth and surface orientation produced by early vision for tasks such as object recognition, locomotion and visually guided manipulation. Marr's own account of late visual processing is rather sketchy. His concrete proposals concern the computational level of description, with little or no detail supplied at the algorithmic level. In general, computational models of high-level vision are not as well developed as accounts of early visual processes. The difficulty is due in part to the fact that later

processing is hypothesis- (or goal-)driven, and hence cognitively penetrable. The input to these processes is not limited to information contained in the image. Object recognition, for example, makes use of specific knowledge about objects in the world. This knowledge is usually characterized as a catalogue of object types stored in long-term memory. It is worth noting that only at this rather late stage does the visual system do anything like identify what Gibson calls ‘affordances’, and in computational accounts such identification is typically treated as a process of categorization, in other words, as a psychological process (see Concepts §1).

Various types of computational models of object recognition have been proposed. According to the simplest models, recognizing an object currently in view involves comparing it with previously stored views of objects and selecting the one that most resembles it. A problem with this approach is that it fails to explain our ability to recognize objects from novel views that do not straightforwardly resemble any previously stored views.

More promising are accounts that treat object recognition as associating with the current view of the object a description of the object type, perhaps in addition to previously stored views of representative examples. Here again, different approaches are possible. ‘Invariant-property’ accounts assume that the set of possible retinal projections of objects typically have higher-level invariant properties that are preserved across the various transformations that the object may undergo. Such proposals face the same problem as Gibson’s account of higher-order invariants. For most object types it has proved impossible to find specifiable properties of the image that are common to all possible recognizable views.

The ‘decomposition’ approach to object recognition maintains that objects are identified on the basis of prior recognition of their component parts. An assumption of this approach is that the relevant part-whole relations are invariant and detectable in all possible views where the subject would recognize the object. The most developed proposal is Irving Biederman’s ‘recognition by components’ theory (1990), according to which a given view of an object can be represented as an arrangement of simple primitive volumes called ‘geons’ (for ‘geometric icons’). Geons can themselves be characterized in terms of viewpoint-invariant properties, and, proponents of the theory claim, are recognizable even in the presence of visual noise. In general, though, the decomposition approach to object recognition has proved to be fairly limited in its application. Many objects do not decompose in a natural way into easily characterizable parts; and for many of those that do the decomposition is insufficient to specify the object in question.

A third strategy, known as the ‘alignment’ approach, suggests that the visual system detects the presence of transformations between the current view of an object and a stored model, and can ‘undo’ the transformation to achieve a correspondence between the two. For example, suppose that the current view of the object differs from the model stored in memory because the object has undergone a three-dimensional rotation and moved further away from the viewer. On the current proposal, the visual system first detects the nature of the transformations, and then performs them in reverse on the current view to bring it into ‘alignment’ with the stored model (assuming that the object is rigid). The main problem for this approach, as for the other proposals, is its limited applicability. It is only feasible for a small range of possible transformations that an object can undergo (for example, rotation and scaling) and then only for a limited range of objects. (Imagine detecting and ‘undoing’ the rotation of a crumpled piece of newspaper.)

‘Mixed’ approaches to object recognition attempt to extend the range of applicability of the decomposition and alignment approaches by combining elements of the two, positing separate identification systems that operate in parallel. While mixed accounts appear promising, they face the additional burden of explaining how the outputs of the two recognition systems are combined.

7 Computational models of vision: problems and prospects

The most common criticism of computational models of human cognitive capacities, including accounts of our perceptual abilities, is that they are unable to approximate actual human performance. It is true that many impressive computer models fail miserably in the real world. Sometimes they fail because the information required is not available to the mechanism. As Marr emphasized, the computational theorist can try to avoid this problem by first attempting to characterize the computational problems that perceptual mechanisms, in their natural context, are required to solve, a process that involves discovering general environmental constraints that perceptual mechanisms of adapted organisms can be expected to exploit.

But the study of biological visual systems faces additional hurdles. Even if the information on which a posited process runs is in some abstract sense ‘in the data’, the input may be too ‘noisy’ for the mechanism to make use of it. Computational theorists are of course aware of this problem. Some of the processing posited by computational accounts, especially in early vision, involves the elimination of extraneous or irrelevant information in the image. (For example, the primal sketch, which represents intensity changes in the image, does not preserve the absolute values of intensity gradients at every point in the grey-level array.) Moreover, the theorist must eventually find neural hardware capable of doing the computationally characterized job, before being confident that the model is biologically feasible. Given the difficult nature of the task it is unlikely that a complete computational account of vision is just around the corner. None the less, computational theorists make an important contribution to our understanding of vision by their careful study of the nature of the problems to be solved by visual mechanisms, although the solutions they offer are properly evaluated by their performance in the real world.

An alternative style of computational model may ultimately prove better suited to explicating human vision than models, such as Marr’s, that treat perceptual processing as rule-governed operations defined over representations. In ‘connectionist’ computational architectures information is typically represented by patterns of activation over a connected network of units or nodes. Connectionist processes are explicated at a level distinct from the neurological or implementational. Connectionist cognitive models typically appeal to representations, memory and learning, hence they qualify as indirect; although connectionist accounts of representation, memory and learning differ in significant respects from more traditional computational accounts (see Connectionism). Connectionist theorists have claimed that their models are better able to handle noisy input and ‘multiple simultaneous constraints’ characteristic of real-world processing situations, though traditional computationalists dispute this claim. Whether the best models of human visual processing will be connectionist remains to be seen.

See also: Colour and qualia; Consciousness; Molyneux problem; Perception

FRANCES EGAN

References and further reading

Berkeley, G. (1709) ‘An Essay Towards a New Theory of Vision’, in *The Works of George Berkeley, Bishop of Cloyne*, vol. 1, ed. A.A. Luce and T.E. Jessop, Edinburgh: Thomas Nelson, 9 vols, 1948-57.(Referred to in §1.)

Biederman, I. (1990) ‘Higher-Level Vision’, in D.N. Osherson *et al.* (eds) *Visual Cognition and Action: An Invitation to Cognitive Science*, vol. 2, Cambridge, MA: MIT Press.(An example of the ‘decomposition’ approach to object recognition.)

Descartes, R. (1637) ‘Optics’, in *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1985, vol. 1, 152-75.(Discourses 5 and 6 are particularly relevant.)

Descartes, R. (1641) ‘Author’s Replies to the Sixth Set of Objections’, in *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1984, vol. 2, esp. §9: 294-6.(Referred to in §1 - Descartes’ ‘intellectualist’ theory of vision.)

Fodor, J.A. and Pylyshyn, Z. (1981) ‘How Direct is Visual Perception?: Some Reflections on Gibson’s "Ecological Approach"’, *Cognition* 9: 139-96.(A critical discussion of Gibson’s direct theory of perception. Includes detailed argument but no technicality.)

Gibson, J. (1979) *The Ecological Approach to Visual Perception*, Boston, MA: Houghton Mifflin.(The most developed statement of Gibson’s theory of perception.)

Helmholtz, H. von (1950) *Treatise on Physiological Optics*, ed. J. Southall, New York: Dover, 3 vols.(Influential nineteenth-century account of perceptual processing as a species of inference.)

Hinton, G.E. (1992) ‘How Neural Networks Learn from Experience’, *Scientific American* 267 (3): 144.(Includes a discussion of connectionist models of shape recognition.)

Horn, B. (1986) *Robot Vision*, Cambridge, MA: MIT Press.(A detailed account of work in computer vision, with exercises. Very technical.)

Marr, D. (1982) *Vision*, New York: Freeman Press.(Somewhat technical, but includes a clear account of the rationale behind the computational approach to vision.)

Rock, I. (1983) *The Logic of Perception*, Cambridge, MA: MIT Press.(Account of perceptual processing as a form of hypothesis formation and testing.)

Schwartz, R. (1994) *Vision: Variations on Some Berkelian Themes*, Oxford: Blackwell.(A useful discussion of historical work on the problems of vision. Also includes a chapter on Gibson’s theory.)

Ullman, S. (1979) *The Interpretation of Visual Motion*, Cambridge, MA: MIT Press.(A detailed analysis of the computations involved in visual motion perception.)

Vital du Four (*c.*1260-1327)

A Franciscan philosopher and theologian, Vital du Four was noted for denying the distinction between a thing's essence and its existence, for expounding an Augustinian theory of perception and for emphasizing the absolute power and contingency of God's will in creating the universe. One interpretation of his views holds that created things have no intrinsic goodness, only that which has been conferred upon them by God.

Born in Bazas, in the southwest of France, Vital du Four (or Vitalis de Furno) entered the Franciscan order, was educated in theology at Paris in 1285-91, and taught at Montpellier (1292-6) and Toulouse (1296-1307). He was appointed a cardinal in 1312, becoming Cardinal-Bishop of Albano in 1321. He died in 1327 in Avignon. A contemporary of John Duns Scotus, du Four has been so overshadowed by his Franciscan colleague's philosophical accomplishments that until recently, one of du Four's major works, *De rerum principio (On the Origin of Things)*, was erroneously included in Scotus' *Opera omnia*.

Du Four rejected Thomas Aquinas' thesis that, except for the case of God, there is a distinction between an individual's essence and its existence (see Aquinas, T. §9). According to du Four, the existence of any individual (other than God) just is that individual's essence as related to the individual's efficient cause. Du Four's thesis seems to be that for Socrates to exist is for Socrates' essence - namely, to be a rational animal - to stand in a certain relation to whatever is the efficient cause of Socrates. The relation includes the notion of participation: the relevant activity of Socrates' efficient cause results in an individual, Socrates, who participates in the essence of being a rational animal.

Unlike a theory, popularly associated with Aquinas, that maintained that genuine knowledge must be only of universal, necessary truths, du Four allows knowledge to extend to contingent particulars. Knowledge of physical objects comes about through sensation, which apprehends individual things. Du Four claims, however, that sensation itself necessarily involves the simultaneous operation of the intellect, thus precluding any view that maintains that the intellect operates only after it receives the testimony of the senses. While the senses apprehend individuals, the intellect cognizes the activities of the senses along with the individuals that are the objects of those activities. Augustine's theory of 'vital attention' had maintained that sense perception is possible only when the soul actively attends to the events occurring in the body, because the inferior body cannot in itself cause changes in the superior soul (see Augustine). Du Four's account of the intellect's activity in sense perception is reminiscent of Augustine's theory, although it is not clear that du Four subscribes to Augustine's causal principle.

In contrast to the way in which knowledge of contingent particulars is acquired, du Four introduces a theory of divine illumination (an important expression of which could also be found in Augustine's writings) to account for knowledge of necessary, universal truths. Human intellects are too inconstant to grasp such immutable truths by their own natural means. Knowledge of necessary truth is the result of the soul's union with God's light, and the acquisition of such knowledge is entirely dependent on God's will.

Du Four rejects the opinion, expressed in Plato's *Timaeus* (see Plato §16), amplified by Pseudo-Dionysius the Areopagite and expounded afterwards by Avicenna (see Ibn Sina), that creation is a necessary outcome of God's goodness. God is necessarily infinitely good, but the necessity of God's goodness has its source in God's own immutable will and not in any kind of external compulsion. No aspect of creation is sufficient to move God to create. In particular, the goodness of creatures does not enhance God's goodness, and so their goodness is accidental to God's willing to create.

As du Four puts it, creatures add nothing to God's goodness 'just as a point adds nothing to a line' (*De rerum principio* 4.1.3). He may have intended this analogy only to imply that because God is *infinitely* good intrinsically, no extrinsic good can add to God's goodness; but he may also have intended the analogy to convey the thesis that just as points have no length in themselves by which to augment the length of a line, so created beings have no goodness in themselves. On either interpretation, du Four's view stresses the contingency of God's creating: God could have refrained from creating anything, let alone this world. The second interpretation, however, adds the further thesis that the goodness of the created world is not something intrinsic to it, and suggests that if the world is properly to be regarded as good, its goodness must be conferred on it by a declaration of God's free will. Du

Four's analogy may thus have influenced the voluntaristic views of William of Ockham less than a generation later.

See also: Creation and conservation, religious doctrine of; Goodness, perfect; Voluntarism

WILLIAM E. MANN

List of works

Vital du Four (late 13th-early 14th century) *De rerum principio (On the Origin of Things)*, ed. L. Wadding in *Joannis Duns Scoti Doctoris Subtilis, Ordinem Minorum Opera Omnia*, Paris: L. Vivès, 1891, vol. 4, 267-70. (Question 4 contains du Four's discussion of whether God created of necessity.)

Vital du Four (late 13th-early 14th century) *Quaestiones disputatae (Disputed Questions)*, ed. F.M. Delorme in 'Le Cardinal Vital du Four: Huit questions disputées sur le problème de la connaissance', *Archives d'histoire doctrinal et littéraire du moyen âge*, Paris: Vrin, vol. 2, 1927: 151-337.(Contains du Four's denial of a distinction between a thing's essence and existence and his views on human knowledge.)

References and further reading

Lynch, J.E. (1972) *The Theory of Knowledge of Vital du Four*, St Bonaventure, NY: The Franciscan Institute. (Useful on du Four's theory of knowledge of particulars, along with the theories of other Franciscan thinkers.)

Mann, W.E. (1991) 'The Best of All Possible Worlds', in S. MacDonald (ed.) *Being and Goodness: The Concept of the Good in Metaphysics and Philosophical Theology*, Ithaca, NY: Cornell University Press.(Contains a brief discussion of du Four's possible voluntarism.)

Vitalism

Vitalists hold that living organisms are fundamentally different from non-living entities because they contain some non-physical element or are governed by different principles than are inanimate things. In its simplest form, vitalism holds that living entities contain some fluid, or a distinctive 'spirit'. In more sophisticated forms, the vital spirit becomes a substance infusing bodies and giving life to them; or vitalism becomes the view that there is a distinctive organization among living things. Vitalist positions can be traced back to antiquity. Aristotle's explanations of biological phenomena are sometimes thought of as vitalistic, though this is problematic. In the third century BC, the Greek anatomist Galen held that vital spirits are necessary for life. Vitalism is best understood, however, in the context of the emergence of modern science during the sixteenth and seventeenth centuries. Mechanistic explanations of natural phenomena were extended to biological systems by Descartes and his successors. Descartes maintained that animals, and the human body, are 'automata', mechanical devices differing from artificial devices only in their degree of complexity. Vitalism developed as a contrast to this mechanistic view. Over the next three centuries, numerous figures opposed the extension of Cartesian mechanism to biology, arguing that matter could not explain movement, perception, development or life. Vitalism has fallen out of favour, though it had advocates even into the twentieth century. The most notable is Hans Driesch (1867-1941), an eminent embryologist, who explained the life of an organism in terms of the presence of an entelechy*, a substantial entity controlling organic processes. Likewise, the French philosopher Henri Bergson (1874-1948) posited an* élan vital *to overcome the resistance of inert matter in the formation of living bodies.*

1 Experimental physiology

The role of vitalism in physiology is exemplified in the work of the French anatomist Xavier Bichat (1771-1802). Bichat analysed living systems into parts, identifying twenty-one distinct kinds of tissue, and explaining the behaviour of organisms in terms of the properties of these tissues. He characterized the different tissues in terms of their 'vital properties', as forms of 'sensibility' and 'contractility'. Bichat thought the sensibility and contractility of each tissue type constituted the limit to decomposing living matter into its parts. These vital properties preclude identifying life with any physical or chemical phenomenon because the behaviour of living tissues is irregular and contrary to forces exhibited by their inorganic constituents. Insofar as living matter maintains itself in the face of ordinary physical and chemical processes that would destroy it, Bichat thought it could not be explained in terms of those forces. He therefore allowed that there are additional fundamental forces in nature that are on a par with those Newton ascribed to all matter: 'To create the universe God endowed matter with gravity, elasticity, affinity…and furthermore one portion received as its share sensibility and contractility' (Bichat 1801, vol. 1: xxxvii). These are vital properties of living tissues.

The key to explaining the distinctive properties of living systems is showing how those properties stem from the constitution of the system. Bichat traced the properties of living systems back to their components; when this was done, the vital properties assigned to these components were opposed to their physical properties. François Magendie (1783-1855) provides a useful contrast. Many of Magendie's experiments mirrored those of Bichat, but he interpreted them as revealing the different steps in a physiological process. Magendie's avowed goal was to abolish the vital properties known as sensibility and contractility, and to 'consider them as functions' (1809). Magendie rejected a mechanistic account of those functions, and acknowledged that many physiological phenomena remained beyond experimental reach, so that it was not possible to explain them in more basic physical terms. Because he acknowledged this distance between vital functions in living organisms and what it was possible to explain in physical terms at the time, Magendie was construed by many as a vitalist; if he is a vitalist at all, his vitalism is very different from that of Bichat.

2 Physiological chemistry

Inspired by Lavoisier's new analysis of combustion, and his demonstration with Laplace in 1780 that respiration in animals is 'slow combustion', chemists in the early nineteenth century hoped to explain many of the reactions found in living organisms. Organic compounds are apparently formed only in living organisms, and thus appear to be products of vital activity. The physiological chemists of the early nineteenth century set out to show, contrary to initial appearances, that these products *are* the results of chemical processes. Jacob Berzelius (1779-1848) argued

that chemistry could account for all of the reactions occurring within living organisms, and that organic and inorganic processes differ only in complexity. ‘ There is’, he said, ‘no special force exclusively the property of living matter which may be called a vital force’ (1836).

A vitalistic view of the relationship of chemistry to physiology is found in Justus Liebig’s study of chemical reactions in plants and animals (1842). In animals he was particularly interested in reactions which metabolize foodstuffs, separating the constituents needed for growth. Liebig offered detailed chemical analyses of the sequence of reactions, based upon chemical analysis of the foods taken in, the products absorbed, and the waste products released. Liebig saw a need for some form of regulation of these reactions, and posited a vital force controlling them. Chemical and vital processes operate in opposite ways, and consequently both sorts of process are necessary in order to understand metabolism. Liebig’s vital forces were not meant to undermine a mechanist programme; rather, they are forces comparable to other physical forces such as gravity and chemical affinity that are possessed by matter and would be exhibited under appropriate conditions. ‘ There is nothing to prevent us from considering the vital force as a peculiar property, which is possessed by certain material bodies, and becomes sensible when their elementary particles are combined in a certain arrangement or form’ (Liebig 1842). Vital forces were invoked to explain phenomena which would otherwise lack an explanation.

Though Berzelius and Liebig were divided over what vital forces they would tolerate, they were united in the desire to explain activities in living organisms in chemical terms. They thought vital forces were necessary because some phenomena have no adequate chemical explanation. Their position is evident in the stance they took on fermentation: it is a chemical process and should be interpretable in chemical terms, whether it is occurring within living organisms or in the test tube. They viewed fermentation and putrefaction as the least challenging cases for chemists, since both processes are simply processes of decomposition and thus the result of simple chemical activity of the sort found in inorganic cases. With the development of better microscopes, Theodor Schwann (1810-82) observed in 1838 that single-celled organisms (yeasts) are involved in fermentation, setting the stage for a controversy that continued for the rest of the century. Schwann and Louis Pasteur (1822-95) argued that fermentation was an activity of whole living organisms and not reducible to ordinary chemistry. This seemed to be a vitalist position. Schwann, though, advanced a mechanistic theory of cell formation, claiming cells simply constitute special environments in which ordinary matter appears in different concentrations. This is not vitalistic. Pasteur, by contrast, fitted fermentation into a more general programme describing special reactions that only occur in living organisms. These are irreducibly vital phenomena. Pasteur demonstrated empirically in 1858 that fermentation only occurs when living cells are present and, further, that cells only carry out fermentation in the absence of oxygen, leading him to describe fermentation as ‘life without air’. Finding no support for claims such as those advanced by Berzelius, Liebig, Traube and other chemists that fermentation resulted from chemical agents or catalysts within cells, Pasteur concluded that fermentation was a ‘vital action’.

In addition to their apparent success in showing that fermentation only occurs in living cells, vitalists like Pasteur also appealed to their demonstration that living organisms always originate from living organisms and that there is no spontaneous generation. The idea of spontaneous generation was inspired in part by the observation of small organisms forming in putrefying matter. The controversy is rooted in the conflict between John Needham (1713-81) and Lazarro Spallanzani. Needham heated closed vessels of meat-broth, discovering that when cooled they still yielded micro-organisms. Spallanzani insisted on longer heating, and in his vessels no micro-organisms developed. In this context, Pasteur showed that heated organic matter remained sterile unless contaminated but that, if contaminated, the previously heated material sustained life. This supported the conclusion that new life-forms only emerge from existing ones and provided additional evidence for the vitalist claim that living organisms are inherently different from non-living entities.

3 Developmental biology

Perhaps the greatest challenge for a mechanist, and the context in which vitalism retained its influence most strongly, was development. Beginning with an undifferentiated and singular egg, development results in an organism with a regular and differentiated structure. The problem is to explain how this regular differentiation is possible. Descartes defended an epigenetic view of embryological development; however, Descartes could not explain how a complex living organism could result from matter and motion (see Descartes, R. §8). This led Nicolas Malebranche (1638-1715) to develop a theory of preformation by *emboitement*, according to which the

germ cells contain, fully formed, the organism. During the seventeenth century, preformation offered a way of accommodating the view that mechanistic laws were insufficient as explanations of the construction of living organisms from unorganized matter. Pre-existence of the organism also avoided the atheistic and materialistic implications of a mechanistic epigenesis, by allowing that all organisms were preformed by the creator. Preformation was widely embraced by the beginning of the eighteenth century. Pierre-Louis Maupertuis (1698-1759), the Comte de Buffon (1713-81) and Needham took up the defence of epigenesis in mid-century, challenging preformationism. All three expanded the range of mechanisms available to include attractive forces. Faced with the problem of explaining the emergence of organization, Maupertuis attributed intelligence and memory to the smallest living particles. On the basis of experiments performed with Needham, Buffon proposed that the development of organisms depended on ‘penetrating forces’ analogous to gravity and magnetic attraction. Needham concluded that there was a ‘vegetative force’ which was the source of all the activities of life. These are vitalistic proposals, which make sense only within a mechanistic programme.

Similar problems persisted throughout the eighteenth and nineteenth centuries. Though Berzelius was mechanistic when faced with physiology, the production of organic form seemed to defy chemical explanation. He thus suggested there was a vital force differing from inorganic elements and regulating development. Charles Bonnet (1720-93), on the other side, was an enthusiastic champion of preformationism. He discovered parthenogenesis in the aphid, concluding that the female germ cell contained wholly preformed individuals, though he allowed that it need not be in exactly the form in which it exists in the adult organism. Beyond this he saw no explanation, emphasizing that the current state of physical knowledge does not allow any mechanical explanation of the formation of an animal. Bonnet embraced no vital forces, and therefore needed some primal organization.

At the end of the nineteenth century, analogous controversies resurfaced, though transformed and subject to experimental investigation. In investigating development, Wilhelm Roux (1831-1924) initiated an experimental version of *Entwicklungsmechanik* in support of internal determinants of development. He embraced a ‘mosaic’ theory of development, according to which the hereditary determinants are distributed in a qualitatively uneven way within the fertilized egg. As the cell divides, the daughter cells are genetically differentiated and these differences explain the differentiation of organisms. In 1888, Roux described experiments designed to test the idea of embryonic self-differentiation. At the first cleavage in the development of a frog, he destroyed one blastomere with a hot needle. In about 20 per cent of the cases, the remaining blastomere continued to develop, and it developed into half an embryo. He concluded that blastomeres develop independently, depending primarily on their internal constitution. This supported the view that development was controlled by material that was successively divided among the cells of the organism. This material, he thought, determined the growth of the organism in a fully mechanical form. In 1891, Driesch performed what seemed at first to be a very similar experiment, but with dramatically different results. Using sea urchins, he separated the blastomeres at the two-cell stage. Each blastomere developed into a smaller but complete blastula. He saw this result as inconsistent with Roux’s mechanistic account and, in particular, as inconsistent with the idea that division of the cell involved a division of the ‘germ’ controlling development. Since the blastomeres have the ability to develop into complete organisms, there could not be the kind of internal differentiation and control Roux had observed. Driesch initially sought external epigenetic factors to explain development. He came to see development as the response of a living organism rather than a mechanically predetermined process. He did not deny that physical and chemical processes are manifested in development, but held that the timing of development requires some special explanation. Physical laws thus place constraints on possibilities, but leave the actual outcome underdetermined. The connections were not immediately made, but Driesch was eventually led to a teleological and vitalistic view of development which he thought could explain developmental patterns.

4 Conclusion

Vitalism now has no credibility. This is sometimes credited to the view that vitalism posits an unknowable factor in explaining life; and further, vitalism is often viewed as unfalsifiable, and therefore a pernicious metaphysical doctrine. Ernst Mayr, for example, says that vitalism ‘ virtually leaves the realm of science by falling back on an unknown and presumably unknowable factor’ (1982: 52). C.G. Hempel, by contrast, insists that the fault with vitalism is not that it posits entities which cannot be observed, but that such explanations ‘ render all statements about entelechies inaccessible to empirical test and thus devoid of empirical meaning’ because no methods of test, however indirect, are provided (1965: 257). The central problem is that vitalism offers no definite predictions.

Neither complaint has much historical credibility. Many vitalists were in fact accomplished experimentalists, including most notably Pasteur and Driesch. Moreover, vitalists took great pains to subject their views to experimental test. Magendie, for example, insisted on the importance of precise quantitative laws. Vitalism, as much as mechanistic alternatives, was often deeply embedded in an empirical and experimental programme. Typically, vitalists reacted to perceived inadequacies of mechanistic explanations; in many cases they rightly recognized that the forms of mechanism, materialism or reductionism advocated by their contemporaries were undercut on empirical grounds. In the end, though, their own proposals were supplemented by empirically more adequate mechanistic accounts.

See also: Aristotle; Bergson, H.; Galen; Life, origin of

WILLIAM BECHTEL
ROBERT C. RICHARDSON

References and further reading

Allen, G. (1978) *Life Science in the Twentieth Century*, Cambridge: Cambridge University Press.(A survey of the history of biology in the first part of the twentieth century.)

Berzelius, J.J. (1836) 'Einige Ideen über bei der Bildung organischer Verbindungen in die lebenden Naturwirksame ober bisher nicht bemerkte Kraft', *Jahres-Berkcht über die Fortschritte der Chemie* 15: 237-45. (Organic and inorganic processes differ only in complexity.)

Bichat, X. (1801) *Anatomie générale appliquée à la physiologie et à la médicine*, Paris: Brossom, Gabon et Cie. (See §1.)

Coleman, W. (1977) *Biology in the Nineteenth Century: Problems of Form, Function and Transformation*, Cambridge: Cambridge University Press.(A survey of the history of biology in the nineteenth century.)

Driesch, H. (1914) *The History and Theory of Vitalism*, London: Macmillan.(Driesch's synoptic discussion of vitalism, from its leading twentieth-century proponent.)

Hempel, C.G. (1965) 'Studies in the Logic of Explanation', in C.G. Hempel (ed.) *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science*, New York: Free Press.(See §4.)

Liebig, J. (1842) *Animal Chemistry or Organic Chemistry in its Application to Physiology and Pathology*, trans. W. Gregory, Cambridge: John Owen.(Liebig's statement of the central issues in physiology.)

Magendie, F. (1809) 'Quelques Ideés Générales sur les Phénomènes Particuliers aux Corps Vivens', *Bulletin des Sciences de la Société Médicine d'émulation de Paris* 4. (Considers as functions the vital properties of sensibility and contractility. See §1.)

Maienschein, J. (1991) 'The Origins of *Entwicklungsmechanik*', in S. Gilbert (ed.) *A Conceptual History of Modern Embryology*, New York: Plenum Press, 43-61.(A good synopsis of the principal figures involved in the emergence of developmental biology from the end of the nineteenth century into the early twentieth century.)

Mayr, E. (1982) *The Growth of Biological Thought*, Harvard, NY: Harvard University Press.(An excellent general introduction, covering the span of biological thought from early Greek thought through to the twentieth century.)

Pasteur, L. (1858) 'Mémoire sur la Fermentation Appelée Lactique', *Annales de Chimie Ser.*, 52: 404-18; partially reprinted and translated as 'Pasteur's Memoir on Lactic Fermentation', in J.B. Conant (ed.) *Harvard Case Histories in Experimental Science*, Harvard, NY: Harvard University Press, 1970.(Pasteur's classic work.)

Roe, S. (1981) *Matter, Life and Generation: 18th Century Embryology and the Haller-Wolff Debate*, Cambridge: Cambridge University Press.(A historical examination of debates in embryology and developmental biology at the end of the eighteenth century, focusing on the disputes between Haller and Wolff.)

Roux. W. (1888) 'Beiträge zur Entwicklungsmechanik Des Embryo', *Virchows Archiv für pathologische Anatomie und Physiologie und für klinische Medizin* 114: 113-53; trans. in B. Willier and J.M. Oppenheimer (eds) *Foundations of Experimental Embryology*, New York: Hafner, 1974, 2-37.(Roux's classic experimental work.)

Vitoria, Francisco de (*c.*1486-1546)

Francisco de Vitoria, who spent most of his working life as Prime Professor of Theology at Salamanca, Spain, was one of the most influential political theorists in sixteenth-century Catholic Europe. By profession he was a theologian, but like all theologians of the period he regarded theology as the 'mother of sciences', whose domain covered everything governed by divine or natural, rather than human, law; everything, that is, which belonged to what we would describe as jurisprudence. Vitoria's writings covered a wide variety of topics, from the possibility of magic to the acceptability of suicide. But it is on those which deal with the most contentious juridical issues of the period - the nature of civil power and of kingship, the power of the papacy and, above all, the legitimacy of the Spanish conquest of America - that his fame chiefly rests.

1 Rights and political authority

It was one of Vitoria's central concerns that all rights (*iura*) were natural and the consequence of God's law not of God's grace. The contrasting claim, made first by the fourteenth-century English theologian John Wyclif (§4), then by the fifteenth-century Bohemian reformer Jan Hus. More recently, and far more menacingly, what Vitoria referred to as 'the modern heretics', the Lutherans, made rights, and hence the authority of secular princes, dependent upon God's grace. On this account only a godly ruler could be a just legislator. Thus, if a prince was a heretic or proved, in the eyes of those who chose to judge him, to be in a state of sin, his laws could not be binding in conscience and he might legitimately be deposed. For Vitoria it was vital that such a theory, with all that it implied for the effective right of lesser magistrates to make war upon their ordained princes, should be discredited.

All civil power, for Vitoria, is vested in the commonwealth since, if all societies are natural organisms, it follows that no individual could have held power prior to their formation. It follows too that if, as he said in *De poteste civili (On Civil Power)* (1528), 'legislative power exists in the commonwealth by Divine and Natural Law', the commonwealth may, and indeed if it is to constitute a civil society where there can be only one ruler, *must*, 'delegate its power and offices'. The person or persons to whom it delegates may, of course, be one or many. In the traditional Aristotelian division, with which Vitoria was familiar, there were three types of political rule: monarchy, aristocracy and timocracy (or democracy). Of these Vitoria, like Aristotle, unsurprisingly, considers that 'the greatest and best of all forms of rule and magistracy is monarchy or kingship'. The power exercised by monarchy, apart from being the most common form of rule (or so the historical record would suggest), and hence the most natural, is also of a different order from all others. It is the power to act on behalf of the community - what we today would call the 'high executive prerogative' of the state - and for Vitoria it 'is not from the commonwealth, but from God himself'. Such power, however, is what he calls a 'capability'. It cannot be exercised by an individual monarch until he has received ‘authority or executive power’ from the community. For the king may be the ultimate source of law, but he is also subject to the laws he makes, and these must always be in accordance with the customs of the commonwealth for which they are intended. There is, however, some doubt as to whether he may be coerced in any way if he fails to behave like a subject. Vitoria was willing to admit the possibility of tyrannicide on the grounds that, as he said in *De iure (On Law)* (1533-4) 'even if the commonwealth has given away its authority, it nevertheless keeps its natural right to defend itself'. But, in common with all early-modern political theorists on both sides of the confessional divide, he was also reluctant to ascribe coercive, and thus potentially revolutionary powers, to any authority other than that of the established ruler, whatever his conduct might be.

This conception of power, inevitably, played a central role in Vitoria's observations on the most pressing, and certainly the most intractable, issue in contemporary political theory: the legitimation of the Spanish conquest of America.

2 The Spanish conquest of America

Vitoria composed two treatises on the subject of the Spanish conquest of America: *De Indis (On the American Indians)* (1539) and *De iure belli (On the Law of War)* (1539). Both are answers to the question: By what laws were the barbarians subjected to Spanish rule? The answer to this question, however, turned upon another: Had the Indians, in fact, enjoyed property rights (*dominium*) over their own affairs, and over the territories they occupied, before the arrival of the Spaniards?

There could be, Vitoria argued, only three possible grounds for the Castilian crown's claim that the Indians were not, as were all other men, the subject of natural rights: either that they were sinners, non-Christians, madmen (*insensati*) or simpletons (*amentes*). The first of these invoked, once again, the Lutheran supposition that rights depended not upon God's laws, but upon God's grace. Man, argued Vitoria, is a rational creature and cannot lose that characteristic through sin, any more than he can willingly renounce his natural rights. He therefore concluded 'that, although the Indians were undeniably barbarians, they undoubtedly possessed the same rights, both public and private as any Christians'. This made any a priori claim to rights of conquests, or the claim to have occupied previously unoccupied territory - the so-called 'right of discovery' - invalid.

Vitoria also rejected the widely discussed claim that the Indians might be 'slaves by nature' largely on the empirical grounds that the Indians clearly did have 'some kind of rational order in their affairs'. They lived in cities, had a recognized form of marriage, magistrates, rulers, laws, industry and commerce, 'all of which require the use of reason'.

None of these arguments could provide the Castilian crown with rights to either sovereignty or property in America. Vitoria believed, however, that the Spaniards might legitimately claim a right drawn from the 'law of nations' - the *ius gentium*. This he called 'the right of natural partnership and communication'. There was nothing new in this argument, but Vitoria's formulation of it has won him the reputation of being the 'father of international law'. The seas, shores and harbours are, he argued, necessary to man's survival as a civil being, and must, therefore, be common to all. There was also, under the legal definition given to 'communication', an implied right to trade. If it could be said that the Spaniards had originally come to America as ambassadors, travellers and traders, they had to be treated with respect and be permitted access to all those who wished to trade with them. Vitoria also argued that the laws of nations granted the Spaniards the right to preach their religion without interference (although it did not oblige the Indians to listen) and that it permitted them to wage a just war 'in defence of the innocent against tyranny'.

The Spaniards could enforce these rights if opposed because any attempt to deprive a man of his rights constitutes an injury, and the vindication of injuries provides grounds for a just war. By the terms of a just war the victor acquires the status of a judge and may, therefore, appropriate the property of the vanquished. Similarly prisoners taken in a just war may legitimately be enslaved (see War and peace, philosophy of §§3, 6).

The Castilian crown could claim that its conquests had been 'just wars' only if, in fact, the Indians had 'injured' the Spaniards by denying them access to their lands, or if they had attempted to prevent them from preaching. If, however, as seemed overwhelmingly the case, they had done none of these things, then all that Vitoria was left with was the claim that, since the Spaniards were there already, any attempt to abandon the new American colonies would only result in 'a huge loss to the royal exchequer, which would be intolerable'. However, he argued, there was no evidence that the Portuguese in Africa had gained less by licit trade than the Castilians had gained in America by illicit occupation.

Vitoria's writings on power and the rights of conquest effectively set the agenda for most subsequent discussions on those subjects in Catholic Europe until the late seventeenth century. In Spain his rulings - as they came to be seen - on the legitimation of the colonization of America became something of an orthodoxy. They also provided much of the theoretical underpinning for an extensive body of ethnographical writings on the American Indians. And although it is clearly false to speak of Vitoria as the father of anything so generalized, and modern, as 'International Law', it is the case that his writings became an integral part of later attempts to introduce some regulative principle into international relations (see International relations, philosophy of).

ANTHONY PAGDEN

List of works

Vitoria, F. de (1557) *Relectiones XII in duos tomos diuisae, quarum seriem uersa pagella indicabit, summariis suis ubique locis adietis una cum indice omnium copiosissimo*, Lyon: Jacobus Boyerius, 2 vols in one. (Contains, among others, Vitoria's writings, *De poteste civili*, *De potestate ecclesia prior*, *De potestate ecclesia posterio*, *De iure*, *De Indis* and *De iure belli*.)

Vitoria, F. de (1933-5) *Relecciones teologicas del Maestro Fray Francisco de Vitoria (Theological Reflections of Francisco de Vitoria)*, ed. L.G. Alonso Getino, Madrid: Imprenta La Rafa. (Bilingual edition in Latin and

Spanish.)
Vitoria, F. de (1932-52) *Commentarios á la 'Secunda Secundae' de Santo Tomás (Commentaries on the Secunda Secundae of Thomas Aquinas)*, ed. V. Béltran de Heredia, Salamanca: Universidad de Salamanca, 6 vols.
Vitoria, F. de (1991) *Francisco de Vitoria, Political Writings*, ed. A. Pagden and J. Lawrance, Cambridge: Cambridge University Press.

References and further reading

Barbier, M. (1966) *Introduction* to *Francisco de Vitoria, Leçons sur les indiens et sur le droit de guerre*, Geneva: Droz.(Probably the best general account of Vitoria's view on the laws of war.)
Hamilton, B. (1963) *Political Thought in Sixteenth-Century Spain*, Oxford: Clarendon Press.(Although unsophisticated in many respects, this is still a useful summary of Vitoria's political thought.)
Pagden, A. (1982) *The Fall of Natural Man. The American Indian and the Origins of Comparative Ethnology*, Cambridge: Cambridge University Press.(Offers a detailed analysis of some aspects of Vitoria's writings on the American Indians.)
Pagden, A. (1987) 'Dispossessing the Barbarian: The Language of Spanish Thomism and the Debate over the Property Rights of the American Indians', in *The Language of Political Theory in Early-Modern Europe*, ed. A. Pagden, Cambridge: Cambridge University Press.(An account of the role played by Vitoria and his successors in the struggle over the Castilian crown's claims to *dominium* in America.)
Skinner, Q. (1978) *The Foundations of Modern Political Thought*, vol. 2, *The Age of Reformation*, Cambridge: Cambridge University Press.(Provides the best overall account of the context for Vitoria's political writings.)

Vives, Juan Luis (1493-1540)

Vives, Spanish humanist and educational reformer, was an eclectic but independent thinker, blending Aristotelianism and Stoicism with Christianity. He wrote on philosophy and psychology, religion and social concerns, and a wide range of subjects related to education. He was known by his contemporaries both for his lively attack on scholastic logic and for his practical judgment, or common sense. Familiar with classical, Christian and contemporary literature, he believed, with the Stoics, that learning should be applied for the common good. His original contributions are associated with an empirical approach to the sciences and the observation of nature, and his interest in the practical arts and inventions.

His social concerns included international politics (in which he is always a pacifist), and the relief of the poor in cities. His most scholarly work was an edition, with commentaries, of Augustine's De civitate Dei (The City of God) (1522), but he is best known for his pioneer work on psychology and educational reform. De anima et vita (On the Soul and Life) (1538) offers the first empirical study of the emotions and their relations with the body, based on Galen's theory of humours, and enriched with insights from Vives' lifelong observation of human nature and conduct. De disciplinis (On Instruction) (1531), the outstanding work on education in the sixteenth century, is nothing less than a programme for education from infancy to old age, with due emphasis on moral training and, in the case of the study of nature, reverence for its creator.

1 Life and works

Vives was born in Valencia to Jewish parents who had converted to Christianity. At sixteen he was sent to the University of Paris, and spent the rest of his life in exile. His parents were burnt by the Inquisition as relapsed Jews; he kept their fate to himself. His identification with his native country remained strong, and he proudly used the cognomen Valentinus, but he never saw Spain again. He taught at Louvain and Oxford and associated with the great Northern humanists, especially with his friends Erasmus, More and the French humanist Guillaume Budé. While at Oxford he attended the court of Henry VIII and Catherine of Aragon and was tutor to their daughter, Mary. He was expelled from England for taking Catherine's side in the divorce, and returned to Bruges where he died twelve years later.

At the College of Montaigu in Paris, Vives was drilled for three years in terminist logic and medieval physics. His exasperation at the abstract, sterile speculations, barbarous Latin and combative disputations of the scholastics prompted the book he published in Louvain in 1519, *Adversus pseudodialecticos (Against the False Dialecticians)*, which established him as a prominent figure in the conflict with the scholastic dialecticians (see Logic, Renaissance §2). When he left Paris in 1512 Vives settled among the Spanish colony in Bruges. Here, except for periods in Louvain and England, he was to spend the rest of his life, and here he married, in 1524, Margaret Valdaura of an old Valencian family.

From 1517 to 1523 Vives was at Louvain, teaching privately and at the university. At Erasmus' urging in 1520 he undertook his edition, with expansive commentaries, of Augustine's *De civitate Dei (The City of God)*, which he completed in 1522 and dedicated to Henry VIII. (This is the only major work of Vives not included in the *Opera Omnia* (1555) because it was frowned upon by the church for its dedication and its warm praise of Erasmus, as well as for the heterodoxy of some of Vives' notes; it ended up on the *Index Expurgatorius*.) In 1523, at Cardinal Wolsey's invitation, he went to teach at Oxford. In preparation for his visit to England, he composed and dedicated to his compatriot, Queen Catherine of Aragon, his treatise *Index Expurgatorius (On the Education of a Christian Woman)* (1524). Vives' Spanish austerity towards women gives the book a medieval flavour - for example: 'A woman's greatest ornament is silence' (*De Institutione*: 1 xi.136). More generous is the popular little handbook of morals and manners, blending stoicism and Christianity, which he prepared for the Princess Mary and called *Introductio ad Sapientiam (Introduction to Wisdom)* (1530).

Vives spent the better part of five years in England. He was professor of Latin, Greek and rhetoric at Corpus Christi College, and left his mark on Oxford through humanistic changes and improvements in the curriculum. He complained about the climate, but enjoyed the company of the great English humanists, More, Linacre and Fisher, and began to address himself to problems of a social and political nature. It was during this period that he wrote the original and important work on poor relief which he dedicated to the city of Bruges, *De subventione Pauperum*

(On Assistance to the Poor).

Vives' loyalty to Henry's Spanish queen in the matter of the divorce ended his association with the English court, and he went home to Bruges to devote the rest of his life to writing. During those dozen years of poverty and failing health, racked with gout, he produced his major works. These included *De disciplinis (On Instruction)* (1531) and *De anima et vita* (1538). The eight books *De artibu (On the Arts)*, which were published as Part III of *De disciplinis*, consisted of five books on logic (including 'De censura veri' and 'De instrumento probabilitatis') and three ('De prima philosophia') on Aristotelian physics and metaphysics from a Christian viewpoint. In this period he also produced the most popular of his books on education, *Linguae Latinae Exercitio (Use of the Latin Language)* (1539), which, in twenty-five entertaining dialogues, employs all the Latin words in current usage. A posthumous work, *De veritate fidei Christianae (On the Verity of the Christian Faith)*, intended for the conversion of pagans, Moors, Jews and heretics, was published in Basle in 1543.

2 Moral and political thought

Vives was a moral philosopher of sound and perceptive judgment, admirable for his sincerity and generosity of spirit. Erasmus said of him that he had a 'wonderfully philosophic mind' (*animo mire philosophico*). He was committed to the project of harmonizing Aristotelianism with Christianity and was sympathetic to the Stoic emphasis on personal virtue, inward tranquillity and social engagement. The writers he most often cites are Cicero, Seneca, Pliny, Plutarch and Quintilian. He once said there is no better Christian than a Stoic sage, and his own maxim was 'Without complaint' (*Sine querela*).

Accordingly, he was an *homme engagé*. He agreed with the Stoics that 'having acquired our knowledge we must turn it to usefulness and employ it for the common good' (*Opera Omnia*: vol. 6, 1.423). Deeply troubled by the political turmoil in Europe, he wrote urgently to heads of state and church on behalf of international peace among Christian nations; in *De Europae dissidiis et bello Turcico (On European Discord and the War with the Turks)* (1526), Vives urged that if they must employ their armies, they should unite against the encroaching Turks. His social concern found expression in his enlightened plan for poor relief, which was adopted by numerous Flemish cities. He also wrote a tract against the Anabaptists called *De communione rerum (On the Common Ownership of Goods)* (1535).

He was a staunch Christian. Like Erasmus, he said that if he differed from the church he would submit to it, but his personal religion tended more to ethical and spiritual values than to rite and observance, and it is significant that his ideal institutions, whether for poor relief or education, were intended to be secular. None the less his belief in God and the moral order of the universe was the foundation for all his thought, and piety was the motive and inspiration for his work.

3 Metaphysics and epistemology

Religion apart, his idea of the world and the place of people within it is essentially Aristotelian. Vives was known in his lifetime and after as a critic of Aristotle because of his incisive critique of Aristotle in *De disciplinis*. In fact he had read Aristotle more deeply than most other humanists, and had found him superior to all other philosophers. Vives identifies Aristotle's faults, but the Aristotelian systematization of knowledge based on self-evident first principles forms the basis of Vives' thinking about nature, matter and the mind.

In his metaphysics ('De prima philosophia') he discusses the creation of the world and the innate tendency of all peoples to believe in God. We are sorely hampered because of the Fall, but God has left us one priceless resource - a lively intelligence which is spontaneously active. Vives sometimes identifies this with 'natural light' (*lumen naturalis*), defined as those natural gifts given to us to enable us, through reason, to behold the truth. He accepts the Stoic doctrine of anticipations and common notions in the mind before birth (such as the innate belief among all peoples of the existence of God). From these first truths we gradually infer other truths, as plants grow from seeds. The order of knowledge is: sense perception, imagination, reason (composed of intelligence, memory and will) and judgment. To Vives a trained and enlightened judgment, based on experience, was the single most important quality for learning and life. He called it *prudentia*, that is, practical judgment or common sense. This is the surest guide to conduct. It is also the quality for which Vives was known by his contemporaries.

Vives is well aware of the limits to human reason. Of the three paths we have to knowledge: the senses, faith in

authority and reason based on the senses, none is infallible. Only God has perfect knowledge. Therefore Vives issues a strong caution at the beginning of the metaphysics, that we should not inquire into the intimate workings of nature. God has his own reasons for his creation and it is impious for man to seek, for example, to know about the elements, the forms of living things, and the number, magnitude, dispositions and powers of these things. Here Vives seems to be recalling with distaste the speculative natural philosophy characteristic of late scholasticism.

He is sceptical about the possibility of acquiring certain rational knowledge not only in theology but also about ordinary nature. Even here, he warns us, what knowledge we have gained can only be reckoned as probable, not as absolutely true. His scepticism did not, of course, extend to religious authority. Fundamental doctrines like the Fall were unquestioned, and in fact formed the basis for his doubts about certainty in human knowledge. In contrast with this is the optimism and confidence Vives expresses elsewhere over the progress of science and the open-endedness of learning. He declares that:

> Nature is not yet so effete and exhausted that she can bring forth, in our times, nothing comparable to those of earlier ages. She always remains equal to herself, and not rarely she comes forward stronger and more powerful than in the past…. Much is left for future generations to discover.
>
> (Preface to *De disciplinis*)

Perfect truth may be unattainable, but steady progress in knowledge can be made. By way of proof, Vives likes to point to the recent remarkable discoveries of ‘our [Spanish] countrymen’ beyond the ancient borders of the East and West.

4 Psychology and pedagogy

The new ingredient Vives brought to humanism was an empirical approach to the sciences. Having fled the closed scholastic world and embraced humanism and the classics, he moved a step further into the practical sphere. His original contribution lay in the importance he gave to observation and experiment in psychology, the teaching of natural history, law and medicine, and in his attitude towards the practical arts and inventions. He advocated the use of the vernacular in teaching young pupils, and in the presentation of laws, so that they could be understood by the people. His pragmatic voice is not that of the literary Erasmian humanist, but rather of the Spanish émigré in Bruges, heir to a long Jewish tradition of practical service in law and medicine, friend of merchants (he once held an export licence for his relatives), and especially of doctors and lawyers.

The last precept in Vives’ handbook of morals, The *Introduction to Wisdom*, is first to know oneself, and ultimately to know God. This rule he set out to implement in his two most important works, and to this task he brought his deep learning, critical spirit, and years of experience, observation and reflection on psychology and education.

The first of the two (though published second) was *De anima et vita* (1538), a pioneer work in psychology. Some consider it Vives’ masterpiece. It is an empirical study of the soul or mind and its interaction with the body. He classifies our emotions in relation to the Galenic humours (the four humours: blood, phlegm, yellow bile and black bile, which come to constitute the four classic temperaments: sanguine, phlegmatic, choleric and melancholic), describes how they are affected by internal and external circumstances (such as age, health, climate and events), and how they can be motivated and controlled by thought, judgment and will. Elsewhere Vives mentions the importance of this knowledge in politics and law, and it provides a foundation for his pedagogy (which is based on attention to individual capacities and their development).

The second, *De disciplinis* (1531), is a monumental work, divided into two parts, one critical, *De causis corruptarum artium (The Causes of the Corruption of the Arts)* and one constructive, *De tradendis disciplinis (The Transmission of the Arts)*. *De causis* traces the arts, or disciplines, from their beginning in our primitive human needs, through their growth by an empirical process, to their peak and decline. Decline was brought about sometimes by external causes (such as the barbarian invasions and the destruction by Goths and Vandals which produced the Dark Ages), most often by the weakness and vices of men and the emotions that cloud our reason. For example pride, arrogance and greed have diverted the pure study of astronomy into astrology, which is firmly to be shunned. *De tradendis* is a comprehensive programme for education from infancy to old age. No detail is overlooked, from physical requirements - site of the school, exercise, diet - to the bedside manner of the physician.

Much space is given to nature study, observation and experiment, backed for every subject by a selective and critical reading list ranging from the classics to contemporary sources. Higher studies follow, and professional training. Like Hippocrates, Vives links medicine with dietetics, and includes the composition of medicines, anatomy and dissection. He balances this with moral philosophy (ethics, politics, jurisprudence) and stresses the importance of historical studies. Nor does education stop there. After formal study one will turn one's interest to the arts and inventions that contribute to our daily life: food and shelter, gardening, architecture, travel, navigation. Finally, for the elderly, observing nature offers pleasant recreation, a walking exercise and sauce for the appetite.

Vives' originality as well as his sound judgment attracted many readers. Estelrich mentions nearly 500 editions of his works, 200 for the dialogues alone. In England his influence is evident in Elyot, Ascham and Milton, and he was read and quoted by the dramatist Ben Jonson. Some of his prayers found their way into the Book of Common Prayer. On the Continent his educational theories left their mark on Rabelais, Sturm and Comenius. As a critic of Aristotle he was read by Ramus and Gassendi, who also cited him for his emphasis on experiment. His scepticism about the reliability of knowledge, combined with an optimistic belief in the progress of science, appealed to Francisco Sanches and Montaigne (who also cited his edition of Augustine), as well as Gassendi. His theory of the emotions influenced Descartes' *Les passions de l'âme (The Passions of the Soul)* (1649). Finally, leaders of the Scottish school of Common Sense, Thomas Reid and Dugald Stewart, praised his theory of innate common notions, his emphasis on practical judgment, and his foresight as to the future of science. After a period of neglect he was rediscovered and read by such nineteenth- and twentieth-century scholars as Lange, Renan, Duhem, Dilthey and Cassirer.

See also: Aristotelianism, Renaissance; Humanism, Renaissance; Scepticism, Renaissance

RITA GUERLAC

List of works

Early writings and opuscula of Vives are currently being edited and translated by a group of scholars, notably J. Ijsewijn, C. Matheeusen, C. Fantazzi and E. George, and published by Leiden: Brill (1987-).

Vives, J.L. (1493-1540) *Opera Omnia (Complete Works)*, ed. G. Mayans y Siscar, Valencia, 1782-90, 8 vols; facsimile edn, London: Gregg Press, 1964.(Vives' complete works; following the *Opera Omnia* of 1555, this collection does not include Vives' commentary on Augustine's *De civitate Dei*.)

Vives, J.L. (1493-1540) *Obras Completas (Complete Works)*, ed. L. Riber, Madrid: Aguilar, 1947-8, 2 vols. (Spanish edition of Vives' works.)

Vives, J.L. (1519) *Adversus pseudodialecticos (Against the False Dialecticians)*, Louvain; trans. and ed. R. Guerlac, Synthese Historical Library 18, Dordrecht and Boston, MA: Reidel, 1979; C. Fantazzi (trans. and ed.) *In pseudodialecticos*, Leiden: Brill, 1979.(Critical editions of Vives' attack on scholastic dialecticians. The Guerlac edition also includes Books III and V-VII of *De causis*.)

Vives, J.L. (1522-8) *Literae virorum eruditorum ad Franciscum Craneveldium (Letters of Learned Men to Francis Craneveldt)*, ed. J. de Vocht, Louvain: Libraire universitaire, Uystpruyst, 1928.(An important biographical source, it includes letters between Vives and his friend Craneveldt.)

Vives, J.L. (1524) *De Institutione Feminae Christianae (On the Education of a Christian Woman)*, Antwerp: Hillenius; in G. Mayans y Siscar (ed.) *Opera Omnia*, London: Gregg Press, 1964, vol. 4, 65-301.

Vives, J.L. (1526) *De subventione Pauperum*, Bruges: Hubert de Crouck; in G. Mayans y Siscar (ed.) *Opera Omnia*, London: Gregg Press, 1964, vol. 4, 421-94; trans. A. Tobriner, *A Sixteenth Century Urban Report. Translation of Vives' On Assistance to the Poor*, Chicago, IL: University of Chicago Press, 1971.(One of Vives' important works of practical social philosophy; written during his stay in England.)

Vives, J.L. (1526) *De Europae dissidiis et bello Turcico (On European Discord and the War with the Turks)*, in G. Mayans y Siscar (ed.) *Opera Omnia*, vol. 6, London: Gregg Press, 1964.(Vives' attempt to reconcile European leaders.)

Vives, J.L. (1530) *Introductio ad Sapientiam*, Antwerp: Apud Martinum Caesarem; in G. Mayans y Siscar (ed.) *Opera Omnia*, London: Gregg Press, 1964, vol. 1, 1-48; trans. A. Tobriner, *Vives' Introduction to Wisdom, A Renaissance Textbook*, Chicago, IL: University of Chicago Press, 1968.(The little handbook of manners and morals written for Mary Tudor, daughter of Catherine of Aragon and Henry VIII.)

Vives, J.L. (1531) *De disciplinis (On Instruction)*, in G. Mayans y Siscar (ed.) *Opera Omnia*, London: Gregg Press, 1964, vol. 3, 82-297; vol. 6, 1-437.(Both parts of *De disciplinis* appear in the Gregg Press facsimile edition of Vives' works. The first part of this work, *De causis corruptarum artium*, has been translated into German by Sendnes and Hidalgo Serna, and English by Guerlac; the second part, *De tradendis disciplinis*, by

Watson. The third part, *De artibus*, was published posthumously in 1555.)

Vives, J.L. (1531) *De causis corruptarum artium (On the Causes of the Corruption of the Arts)*, in G. Mayans y Siscar (ed.) *Opera Omnia*, vol. 6, London: Gregg Press, 1964; German trans. W. Sendnes and E. Hidalgo Serna, Munich: Fink, 1990; English trans. R. Guerlac *De causis*, in preparation.(The first part of Vives' *De disciplinis*; traces the history of the arts, or disciplines.)

Vives, J.L. (1531) *De tradendis disciplinis (On the Transmission of the Arts)*, in G. Mayans y Siscar (ed.) *Opera Omnia*, vol. 6, London: Gregg Press, 1964; trans. F. Watson, *On Education, a translation of the 'De Tradendis Disciplinis'*, Cambridge: Cambridge University Press, 1913.(One of Vives' most important works on education; the second part of Vives' *De disciplinis*. The Watson translation provides a useful introduction to Vives's work.)

Vives, J.L. (1531) *De artibus (On the Arts)*, Antwerp: Apud Michaelem Hillenium, 1555; in G. Mayans y Siscar (ed.) *Opera Omnia*, London: Gregg Press, 1964, vol. 3, 82-297.(Published after Vives' death as the third part of *De disciplinis*; includes five books on logic and three on Aristotelian physics and metaphysics, all from a Christian viewpoint.)

Vives, J.L. (1535) *De communione rerum (On the Common Ownership of Goods)*, in G. Mayans y Siscar (ed.) *Opera Omnia*, vol. 5, London: Gregg Press, 1964.(A tract written to oppose the Anabaptists.)

Vives, J.L. (1538) *De anima et vita (On the Soul and Life)*, trans. into Italian and ed. M. Sancipriano, Padua: Gregoriana, 1974.(Critical edition of Vives' groundbreaking work on psychology.)

Vives, J.L. (1539) *Linguae Latinae Exercitio (Use of the Latin Language)*, in G. Mayans y Siscar (ed.) *Opera Omnia*, London: Gregg Press, 1964, vol. 1, 283-408; trans. F. Watson, *Tudor School-boy Life, the Dialogues of Juan Luis Vives*, London: Dent, 1908; repr. London: Frank Cass, 1970.(Twenty-five entertaining dialogues, demonstrating all the Latin words in current usage; most popular of Vives' books on education.)

Vives, J.L. (1543) *De veritate fidei Christianae (On the Verity of the Christian Faith)*, Basle: Oporinus; in G. Mayans y Siscar (ed.) *Opera Omnia*, vol. 8, London: Gregg Press, 1964.(Intended for the conversion of pagans, Moors, Jews and heretics; published posthumously.)

References and further reading

Bonilla y San Martin, A. (1903) *Luis Vives y la filosofia del Rinacimiento (Luis Vives and Renaissance Philosophy)*, Madrid: Imp. del Asilo de Huésfanos.(The first full-scale modern work on Vives.)

Erasmus, D. (1906-47) *Opus Epistolarum Des. Erasmi Roterodami (Collected Letters of Desiderius Erasmus)*, ed. P.S. Allen, Oxford: Clarendon Press, 12 vols.(Includes correspondence with Vives.)

Estelrich, J. (1942) *Vives, Exposition organisée à la Bibliothèque Nationale (Exhibition Organized by the National Library)*, Paris: Darantière.(Referred to in §4. A detailed bibliography and evaluation of Vives' reputation throughout Europe; the exhibition was held in Paris in 1941.)

Guy, A. (1972) *Vivès, ou l'humanisme engagé (Vives, or Committed Humanism)*, Paris: Seghers.(Explores Vives' practical and social concerns.)

Noreña, C.G. (1970) *Juan Luis Vives*, The Hague: Nijhoff.(A general study, but badly edited.)

Noreña, C.G. (1989) *Vives and the Emotions*, Carbondale, IL: Southern Illinois University Press.(A useful study of Vives' *De anima et vita*.)

Noreña, C.G. (1990) *A Vives Bibliography*, Lewiston, NY: Mellen Press.(Supplements Estelrich and brings his work up to date.)

Santamaria, J.A.F. (1990) *Juan Luis Vives. Esceptismo y Prudencia en el Renacimiento (Scepticism and Wisdom in the Renaissance)*, Salamanca: Universidad de Salamanca.

Vlastos, Gregory (1907-91)

A leading figure in the study of ancient Greek philosophy, Vlastos was a pioneer in the application to ancient philosophers of the techniques of analytic philosophy. Concentrating on figures of early Greek philosophy, he made major contributions to the understanding of the Presocratics, Socrates and Plato. He saw the Presocratics as applying ethical concepts to nature which ultimately rendered nature intelligible. He distinguished between the early dialogues of Plato, which represent the philosophy of Plato's master Socrates - a philosophy the early Plato shared - and the middle dialogues in which Plato develops a transcendental metaphysics and rationalist epistemology to ground Socratic ethical concepts. Vlastos's work played a major role in bringing the history of philosophy into the mainstream of philosophical research.

1 Presocratic studies

Born in 1907 in Istanbul to a Greek father and a Scottish mother, Vlastos was brought up as a Protestant and moved to America to study for the ministry at the University of Chicago. An interest in philosophy took him to Harvard, where he completed a Ph.D. under Alfred North Whitehead in 1931. Accepting a teaching position at Queen's University in Ontario, he became heavily involved in causes of economic and social reform. He published articles in Christian journals and wrote several lively and insightful defences of Whiteheadian metaphysics.

In 1938, in what was to be the turning point in his career, he went to Cambridge to study ancient philosophy with F.M. Cornford. His research culminated in a paper, 'The Disorderly Motion in the Timaeus' (1939), which challenged the then almost unanimous view of Plato scholars, including Cornford, that the creation of the world described in the *Timaeus* was meant figuratively rather than literally. His work began a reassessment of the *Timaeus* which has tended to vindicate his literalist reading. After serving as an officer in the Royal Canadian Air Force during the Second World War, Vlastos returned to academia, setting out on an ambitious programme of examining Greek political ideas he saw as providing the foundations of philosophical thought. In 'Equality and Justice in Early Greek Cosmologies' (1947) he explained how democratic ideals of equality before the law shaped conceptions of cosmic justice held by Presocratic philosophers. By ascribing such ideals to nature, early philosophers naturalized justice on the one hand and rendered nature intelligible on the other.

In 1948 Vlastos joined the faculty of Cornell University, where he was able to imbibe from Max Black, Norman Malcolm, and Arthur Murphy the latest techniques of analytic philosophy, which, imported from Britain and Austria, was now sweeping the USA. His 'Theology and Philosophy in Early Greek Thought' (1952), a broad survey of religion in early Greek philosophy, argues for a middle ground between the view that the Presocratics were naturalists without religion (argued by Burnet 1930) and the view that they devised theologies of their own (Jaeger 1947): rather, they assimilate divine attributes to nature, thus naturalizing religious values. On the basis of his research Vlastos was able to argue effectively against two views of Presocratic development (both propounded, at different times, by Cornford (1912; 1952)), first that philosophy developed as a scientific reaction to a religious background, and second that it never provided a scientific alternative to religion, as Hippocratic medicine did. Vlastos argued that Presocratic thought was often scientific in spirit, but that it was most indebted to social-political ideas, such as those of justice and equality, that were projected onto nature. In other studies Vlastos examined the epistemology of Parmenides, defended a traditional reading of Heraclitus against a revisionist anti-flux reading by G.S. Kirk, reconstructed a consistent theory of matter for Anaxagoras and a naturalistic basis for ethics in Democritus. He convincingly refuted claims that Parmenides and Zeno of Elea were reacting to a Pythagorean theory of point-atoms - a widely held view originated by Paul Tannery (1930) and promulgated by Cornford (1939) and J.E. Raven (1948) - by showing that there was no evidence that the alleged theory ever existed (see Presocratic philosophy).

2 Platonic studies

Having returned to the sources of Greek philosophy, Vlastos was now ready to return to Plato. His essay 'The Third Man Argument in the *Parmenides*' (1954) was a watershed event in a number of ways. In the first place it marked not only Vlastos's re-entry into Plato studies, but his application of the tools of analytic philosophy to ancient philosophy. Second, its thesis that Plato's presentation of the argument provided a 'record of honest perplexity' satisfied no one and provoked immediate replies from leading philosophers on both sides of the

Atlantic, initially from Wilfrid Sellars and Peter Geach. But further, the ensuing controversy had the effect of bringing ancient philosophy into the mainstream of philosophical debate - whereas the history of philosophy was in danger of being marginalized by the ahistorical or anti-historical bias of analytic philosophy, particularly in the USA. Indeed, the paper came to serve as a kind of paradigm of a new style of analytic history of philosophy. In the controversy Vlastos more than held his own against clever readings of the Third Man Argument by insisting that any interpretation be grounded in the text. His hermeneutical principle was as follows: 'If history is an empirical discipline, so is the history of philosophy. And no one can practise the empirical method unless he is willing to submit even the most deeply entrenched presumptions - his own or those of others - to the arbitrament of factual data' (1973: 255).

In 1955 Vlastos accepted an appointment as Stuart Professor of Philosophy at Princeton University, where he founded a joint Ph.D. programme in classical philosophy in conjunction with the Classics Department, a programme which quickly became a model for the rigorous training of specialists in ancient philosophy. Students were attracted from around the world and graduates were sent out to the best universities to practise and teach the new methods. Vlastos also published important papers in ethics and social philosophy: 'Justice' (1957) and 'Justice and Equality' (1962), the latter of which was widely anthologized. In the 1960s he began a series of studies on Zeno of Elea which still provide the definitive exposition of that philosopher's thought. He also produced an important series of articles on Plato, many of which were collected in *Platonic Studies* (1973), ranging from discussions of slavery in Plato to technical analyses of predication. His studies combined close readings with broad-ranging scholarship and sophisticated philosophical analyses that often investigated Platonic theories in the light of modern philosophical interests. He did not, however, try to force Plato into a modern mould, but rather to indicate where he stood with respect to modern issues and how and why he differed. For instance, in studying Plato's characterization of degrees of reality, Vlastos argued that what is at stake is not degrees of existence, but degrees of cognitive reliability, linked with moral or aesthetic valuation. By attributing a higher degree of reality to Forms than sensible particulars, Plato was in part exploring the greater degree of cognitive reliability that universals have over particulars - though he should have settled for a theory positing kinds rather than degrees of reality.

One area in which his work was especially fruitful was the study of predication. His earlier paper on the Third Man Argument had laid out the premises of the argument and had focused attention on 'self-predication', that is, on claims of the form 'F-ness is F' (for example, 'Justice is just'), as the most vulnerable step of the argument, which was used against Plato to generate an infinite number of Forms, apparently using only Plato's own principles. In 1965 Vlastos realized that some of Plato's utterances of the form 'F-ness is G' could be taken as 'Pauline predications' (named by reference to St Paul's 'Charity is kind' and so on) and analysed as 'necessarily, whatever is F is G'. Subsequently he found Pauline predications in the *Protagoras* and the *Sophist*, but no evidence that Plato himself recognized the ambiguity in his own use of 'F-ness is G' expressions. In general, Vlastos held, Plato's own view of the Forms as paradigms kept him from seeing the dangers of self-predication.

3 Socratic studies

In 1975 Vlastos published *Plato's Universe*, a series of lectures which demonstrated how Plato's cosmology embodied scientific hypotheses despite its seemingly anti-scientific basis. The following year he retired from Princeton, becoming a permanent 'visiting' professor at the University of California, Berkeley, until his death in 1991, teaching classes, giving summer seminars on Socrates for postdoctoral students, and lecturing widely. He continued to publish on Plato, for instance, defending him against Karl Popper in 'The Theory of Social Justice in the *Polis* in Plato's *Republic*' (1977), where he argued that Plato does have a genuine normative and meta-normative theory of justice. But his work increasingly focused on Socrates. Vlastos had always seen Plato's philosophy as developing over time. Furthermore, he had long been convinced that the philosophy of Socrates could be reconstructed from Plato's early dialogues, which represent a fictionalized but historically faithful portrait of Plato's master. Now he tried to work out to his own satisfaction the details of a coherent philosophy for this enigmatic but, for Vlastos, highly sympathetic figure.

According to Vlastos, the Socrates of Plato's early dialogues is inconsistent in a number of important ways with the Socrates of the middle dialogues. We find that other sources, such as Aristotle and Xenophon, tend to attribute to Socrates the traits we find in the Socrates of the early dialogues, for example, that he is solely a moral

philosopher and claims to have no special knowlege. It is plausible to think that the early Plato agreed with his master and only abandoned the Socratic method as he became immersed in geometry as a result of his contact with the Pythagorean philosopher-mathematician-statesman Archytas on his first voyage to Sicily. Socrates' method is not one of mathematical demonstration, but of refutation, requiring that the interlocutor express his real opinions and have them examined; in pursuing this method Socrates can be ruthless, but he never misleads or cheats his opponent. His *elenchus* or method of refutation is not merely a negative method but yields positive results, for it is based on the (well-confirmed) hypothesis that everyone who holds a false belief holds a true belief which conflicts with it, whereas Socrates' set of beliefs is self-consistent, and hence true.

Socrates' disavowal of knowledge can be understood if we make a distinction between 'certain' (demonstrative) knowledge and 'elenctic' knowledge; the latter, being the fallible result of a dialectical process, does not measure up to the former. Since Socrates' method at best yields elenctic knowledge, he is justified in disavowing certain knowledge. Socrates holds that virtue is sufficient for happiness, though he does not make the stronger claim that virtue is identical to happiness: nonmoral goods can and do contribute to happiness, but only when used in conjunction with virtue.

Vlastos succeeded in reconstructing a philosopher's Socrates who is more theoretically interesting than any previous construct, and his teaching produced a new generation of students of Socrates who are continuing a lively debate. Vlastos's greatest talent was perhaps his ability to see in ancient texts arguments and concepts that had philosophical interest for modern thinkers. No other single individual has contributed more to making ancient philosophy a part of contemporary philosophical study in the Anglo-American tradition.

See also: Owen, G.E.L.

DANIEL W. GRAHAM

List of works

Vlastos, G. (1939) 'The Disorderly Motion in the *Timaeus*', *Classical Quarterly* 33: 71-83; also in Vlastos, 1995, vol. 2, 247-64.(An argument against the prevailing orthodoxy that read the creation story of the *Timaeus* allegorically, that it should be taken literally.)

Vlastos, G. (1947) 'Equality and Justice in Early Greek Cosmologies', *Classical Philology* 42: 156-78; also in Vlastos (1995), vol. 1, 57-88.(Shows the influence of Greek political ideas on the early Greek conception of the cosmos.)

Vlastos, G. (1952) 'Theology and Philosophy in Early Greek Thought', *Philosophical Quarterly* 2: 97-123; also in Vlastos (1995), vol. 1, 3-31.(Argues against Burnet (1930) that Presocratic thinkers were not mere naturalists and against Jaeger (1947) that they were not really theologians either; rather they attributed to nature divine properties.)

Vlastos, G. (1953) 'Review of J.E. Raven, Pythagoreans and Eleatics', *Gnomon* 25: 29-35; also in Vlastos (1995), vol.1, 80-88. (Convincingly criticizes the view shared by Tannery (1930), Cornford (1930), and Raven (1948) that Parmenides and the Eleatics were reacting to a Pythagorean theory of 'number atomism'. It now appears that there never was such a theory. Provides the basis for Vlastos' later studies of Zeno.)

Vlastos, G. (1954) 'The Third Man Argument in the *Parmenides*', *Philosophical Review* 63: 319-49; also in Vlastos (1995), vol. 2, 166-90.(By applying techniques of logical analysis to the Third Man Argument against Plato's Theory of Forms, the author raises discussion of the problem to a new level. His conclusion that the argument was a 'record of honest perplexity' on the part of Plato has been rejected by most scholars. But he challenges his critics to find textual evidence to show that Plato had a reply to the argument.)

Vlastos, G. (1955a) 'Review of F.M. Cornford, *Principium Sapientiae*', *Gnomon* 27: 65-76; also in Vlastos (1995), vol. 1, 112-23.(Uses a superior understanding of scientific method to criticize his mentor's interpretation (Cornford 1952) of the Greek medical writers as scientific, the Ionian Presocratics as unscientific. Also rejects Cornford's earlier theory (1912) that philosophy had grown out of religion, in favour of a view that political ideas were decisive for the development of philosophy.)

Vlastos, G. (1955b) 'On Heraclitus', *American Journal of Philology* 76: 337-68; also in Vlastos (1995), vol. 1, 127-50.(Defence of a traditional reading of Heraclitus as philosopher of flux against Kirk (1952), showing the influence of earlier Ionian thought on Heraclitus.)

Vlastos, G. (1957) 'Justice', *Revue internationale de philosophie* 11: 324-43.(A study in moral and social philosophy arguing that agreement, contribution and need constitute 'irreducible components of the concept of justice'.)

Vlastos, G. (1962) ‘Justice and Equality’, in R. Brandt (ed.) *Social Justice*, Englewood Cliffs, NJ: Prentice Hall, 31-72.(Develops an equalitarian theory of justice based on natural rights which can accommodate unequal distribution of some goods. Reprinted in several anthologies of essays in moral philosophy.)

Vlastos, G. (1973) *Platonic Studies*, Princeton, NJ: Princeton University Press.(Collects many of Vlastos’s essays on Plato.)

Vlastos, G. (1975) *Plato’s Universe*, Seattle, WA: University of Washington Press.(Lectures on Plato’s cosmology and scientific method, arguing that Plato’s appeal to the supernatural paradoxically produced valuable scientific theories.)

Vlastos, G. (1977) ‘The Theory of Social Justice in the *Polis* in Plato’s *Republic*’, in H. North (ed.) *Interpretations of Plato*, Leiden: E.J. Brill, 1-40; also in Vlastos (1995), vol. 2, 69-103.(Arguing against Karl Popper’s criticisms of Plato, this paper maintains that there is a legitimate normative and meta-normative theory of justice in the *Republic*, one based not on inequality but on impartiality.)

Vlastos, G. (1991) *Socrates: Ironist and Moral Philosopher*, Cambridge: Cambridge University Press.(Vlastos’s major study of Socrates.)

Vlastos, G. (1994) *Socratic Studies*, ed. M. Burnyeat, Cambridge: Cambridge Unversity Press.(Essays on Socrates.)

Vlastos, G. (1995) *Studies in Ancient Philosophy*, ed. D.W. Graham, Princeton: Princeton University Press, 2 vols. (Collects those of Vlastos’s major essays which do not appear in *Platonic Studies*, *Socrates: Ironist and Moral Philosopher* and *Socratic Studies*. Volume 1: studies in the Presocratics; volume 2: studies on Socrates, Plato and later figures. Contains a complete bibliography of Vlastos’s work.)

References and further reading

Burnet, J. (1930) *Early Greek Philosophy*, London: Adam & Charles Black, 4th edn.(The standard textbook on the Presocratics of its time, this work presents views on Presocratic naturalism which Vlastos (1952) criticizes.)

Cornford, F.M. (1912) *From Religion to Philosophy*, London.(Presents a view on the origins of philosophy which Vlastos (1955a) criticizes.)

Cornford, F.M. (1939) *Plato and Parmenides*, London: Routledge & Kegan Paul.(Presents a view of Parmenides as reacting to Pythagorean theories which Vlastos (1953) criticizes.)

Cornford, F.M. (1952) *Principium Sapientiae*, Cambridge: Cambridge University Press.(Presents a view on the origins of philosophy different from Cornford (1912), which Vlastos (1955a) criticizes.)

Jaeger, W. (1947) *Theology of the Early Greek Philosophers*, Oxford.(A treatment of the Presocratics as theologians which Vlastos (1952) criticizes.)

Kirk, G.S. (1954) *Heraclitus: The Cosmic Fragments*, Cambridge: Cambridge University Press.(A study of Heraclitus which rejects the traditional interpretation of Heraclitus as a flux theorist. Vlastos (1955b) defends the traditional interpretation against Kirk.)

Lee, E.N., Mourelatos, A.P.D. and Rorty, R. (1976) *Exegesis and Argument*, Assen: Van Gorcum.(Festschrift containing reactions to Vlastos’s work.)

Raven, J.E. (1948) *Pythagoreans and Eleatics*, Cambridge: Cambridge University Press.(Presents a view of Parmenides and the Eleatics as reacting to Pythagorean theories which Vlastos (1953) criticizes.)

Tannery, P. (1930) *Pour l’histoire de la science hellène*, Paris: Gauthier-Villars.(Similar to Raven, holds that Parmenides and the Eleatics reacted to Pythagorean theories.)

Voegelin, Eric (1901-85)

Throughout his career, Voegelin was concerned with modernity; unlike his contemporaries he sought the explanation of its character and deformities (especially totalitarianism) in the restoration of 'political science' as Plato and Aristotle understood it. He therefore explored order in the individual's soul, political society, history and the universe, and its source in God. He did so by studying the representation of order in philosophy (Eastern as well as Western) and in revelation and myth. Voegelin concluded that 'gnosticism', the misinterpretation of the insights of myth, philosophy and revelation as descriptions of some future perfected society, and the wilful denial of transcendence and human limitation, represented the essence of modernity.

1 Life and works

By profession and self-description, Eric Voegelin was a political scientist, but philosopher, political theorist, theologian, mystic and intellectual historian would be equally accurate and equally misleading titles. Expelled by the Nazis from a lectureship at Vienna University, where he had studied, taken his Ph.D. and worked as Hans Kelsen's assistant, Voegelin lectured in political science at Baton Rouge, Louisiana from 1942 until 1958. He then took up Max Weber's former chair at Munich University. In 1969 he returned to the USA as Distinguished Scholar in the Hoover Institute, Stanford, California, and remained there until his death. His intellectual power and astonishing scholarly range have been widely acknowledged. His main endeavour was to reconstitute the study of politics, a project publicly initiated in his *The New Science of Politics* (1952) and partly executed in his *Order and History* (1956-87), as well as many other books and articles.

2 Critique of contemporary social science and philosophy

Voegelin's permanent concern was with order and disorder in the soul and in society, and the nature of modernity. Originally prompted by the rise of the totalitarian movements, of which he had first-hand experience, this concern was widely shared by academics. However, Voegelin's public stance was confrontational, and his insistence that knowledge of Mesopotamian, Israelite, Greek, Roman and medieval history and philology, as well as of myth and non-Western cultures, was indispensable to the restoration of a genuine science of order won him few friends. His condemnation of positivism was comprehensive, but he regarded the 1960s and 1970s debate between positivists and their opponents as merely a theoretically more impoverished re-run of its early twentieth-century predecessor (see Positivism in the social sciences). He interpreted both modernity and totalitarianism as 'gnosticism' (see §4). He made sweeping claims about the 'essence' of modern '-isms', 'climates of opinion', doctrines and movements, and about the mediations between these and the thought of the 'representative' thinkers he identified as their progenitors. His philosophical vocabulary is often difficult (drawing on Greek philosophy), and although bilingual and eloquent, he sometimes relies on a background of untranslatable German terms, most notably when he identifies 'the tension (*Spannung*) to the Ground (*Grund*) of divine Being (*Sein*)' as both the source and the object of the philosopher's quest (1956: 2).

3 Reality and consciousness

The questions Voegelin addresses are familiar to phenomenology and existentialism (Alfred Schütz was his lifelong friend, and he was deeply sympathetic to *L'homme révolté* of Camus), and were formulated in explicit critical confrontation with Kant, Hegel, Comte, Marx, Nietzsche, Weber, Husserl and Heidegger. Of contemporary philosophies, he found something to praise in every contemporary philosophy except positivism and Marxism, which (in his view) yielded virtually nothing valuable enough to compensate for their persistent 'prohibition' of fundamental questions.

At the centre of Voegelin's thinking is the metaphysical question of Leibniz, Schelling and Heidegger: Why is there something? Why not nothing? Why are things as they are? The quest for the answer is 'philosophy', which is not epistemology but the love of knowledge. Knowledge can only be knowledge of truth and reality, and reality is not merely or mainly to be found in external objects, which non-philosophical experience, ordinary language and ideologies suppose to exhaust reality. Reality is experienced in consciousness, which in its exploring never encounters anything in the universe which is not of its own 'substance'. What is encountered at the limit is a transcendent order, or structure, of being which has its Beginning or ground (*archê*) and its End (*telos*) in the

Beyond, or divine. And individuals (there is no consciousness except the consciousness of individuals) apprehend reality in this ultimate sense, not as external observers, but as ‘participant observers’. This reality, the ‘utterly real’ (*realissimum*) of myths, philosophers and mystics, the order of the universe and the ‘divine ground of [all] being’ that generates and sustains it and towards which it tends, is sought and found, the finding being experienced as the divine revealing itself: *theophany*.

There are three kinds of theophany: the wisdom of myth; ‘philosophy’ (paradigmatically Plato’s); and the Revelation of the Old Testament and (paradigmatically) Christ in Paul and St John. These are equally valid explorations of the experience of reality, compact in the case of myth, differentiated in philosophy and revelation. Each such exploration is culturally specific and personal (although the experiences are available to all who do not deliberately close them off), and becomes luminous only through language symbols. Some of these explorations have been epochal, rendering previous symbolizations (for example, the intra-cosmic gods of myth) obsolete and breaking historical time into a ‘before’ and ‘after’ the irruption of the divine. But none can ever be final, or perfectly adequate, or make sense independently of the experience of reality from which it arises.

Voegelin’s exploration of the nature of consciousness becomes ever more complex, since he must transcend solipsism and relativism, but rejects the privileged vantage point of the transcendental ego postulated by Kant, Fichte and Hegel; for Voegelin there is no such vantage point. And neither mystery nor the tensions inherent in the human ‘In-between’ are to be abolished. Nor must the symbols whereby this experience of transcendence and order becomes luminous be broken away from their engendering experience. Conversely, philosophical exploration of reality can never be symbol-free; consciousness and symbols always emerge from reality simultaneously, and language and experience are inseparable. The history of consciousness thus becomes a tracking of the trail of symbols and doctrines to their source in experience.

4 Gnosticism and political order

The experience of reality has profound implications, not only for the order of the philosopher’s or prophet’s own soul, but also for their apperception of true order in society and meaning in history. Voegelin’s historical work always centred on the former. As to the latter, he focused on ‘derailments’ and perversions of insight into order by ‘gnosticism’ (the essence of modernity) from antiquity onwards, culminating in totalitarianism (see Totalitarianism). His two attempts at a historical account of order in society and history foundered. In the 1940s, he abandoned an already well-advanced history of political ideas, because ideas, doctrines and doctrinal disputes of any sort are merely a lifeless (or pernicious) residue, when sundered from the experience which they symbolically articulate, or pervert. In *Order and History* (1956-87), Voegelin intended to present symbolizations of experience and their representation in political order as a meaningful historical sequence. His fidelity to evidence, however, made this impossible, and he came to see the idea of a unilinear history of mankind as itself misguided. ‘Mankind’ is a symbol intelligible only in the context of a common relation to the divine; empirically, mankind is no more a historic unity than ‘cat-kind’. He excoriated all equations of the history of mankind with the history of Western high culture, or what he called Western ecumenical imperialism. Nor, although there is meaning *in* history, is there a meaning *of* history. And any idea of a unilinear history terminating in the self-revelation (*egophany*) of a Condorcet, Hegel, Comte, Marx, Husserl or Heidegger, he saw as the height of the lust for power and a disease of the soul (see Historicism).

He construed such derailments as grounded in a refusal to recognize the inevitably circumscribed and uncertain character of human existence as the divine-human ‘In-between’ (Plato’s *metaxy*); this refusal makes man instead of God the measure of all things. Philosophers and prophets can counteract this evil only to a limited extent. The ground has already been occupied by others (individuals and societies) who claim to know reality and to be able accurately to represent it by the time that they arrive. The philosophers and prophets in turn can offer only symbolizations which are inevitably couched in an object-oriented language and are thus always liable to misrepresentation, especially since philosophy and myth have no certainties or definitive answers of their own to propound; and they cannot abolish empirical reality and its disorder and injustice. In particular they cannot abolish the profound sense of injustice of the subjects of ‘ethnic’ orders (that is, culturally cohesive and territorially limited polities) conquered by ‘ecumenical’ orders (polities aiming at a more or less inclusive empire). To such individuals, ‘ethnic’ orders had represented the order of the universe; ‘ecumenical’ empires represent nothing except the triumph of naked power.

Gnosticism, whether as a pernicious potential within Christianity, or stemming from other sources, insulates itself against reality by prohibiting metaphysical questions (notably Marx's ruling out such questions as irrelevant for 'socialist man'), and by 'immanentizing' (interpreting the symbols of transcendence as descriptions of some present or future perfect mankind or society). It may remain mere speculation, or it may become a political force in times of disorientation and disorder, taking the form of political activism, 'liberation', revolution or nihilism. Voegelin sees a similar, but less resolute and destructive infidelity to reality in doctrinaire religion and progressivism of all kinds.

It would have been absurd for Voegelin to present his own insights as a doctrine, even for the sake of remaking the now vanished link between symbols and their engendering experience provided in the past by myth, religion and authority. And it may be the case that humankind finds reality or uncertainty difficult to bear, especially when it lacks a stable institutional tradition and common sense (which the UK and the USA in Voegelin's view still had). Although he was no cultural pessimist, it did not surprise Voegelin that he did not 'find many, or favourable hearers'.

H.M. HÖPFL

List of works

Voegelin, E. (1990-), *The Collected Works of Eric Voegelin*, series ed. P. Caringella, Baton Rouge, LA: Louisiana State University Press.(An exemplary edition which aims to present all of Voegelin's unpublished, as well as published, writings in thirty-four volumes. Volume 12, *Published Essays 1966-1985*, includes striking critiques of Hegel (as a 'magician') and Heidegger, literary interpretations and late statements of his theology and Christianity.)

Voegelin, E. (1952) *The New Science of Politics*, Chicago, IL, and London: University of Chicago Press.(The most widely known of Voegelin's writings, now superseded by later work but still a useful introduction to Voegelin's themes of representation, modernity and gnosticism. Often reprinted.)

Voegelin, E. (1959) *Wissenschaft, Politik und Gnosis*, Munich: Kösel; trans. W.J. Fitzpatrick, *Science, Politics and Gnosticism*, Chicago, IL: Regnery, 1968.(His most accessible, although polemical presentation in two essays, the former outlining the projected *Order and History*, the latter containing the most succinct exploration of *Gnosis* and the notorious critique of Marx; the translation includes an additional essay on 'Ersatz Religion'.)

Voegelin, E. (1956-87) *Order and History*, vol. I: *Israel and Revelation* (1956), vol. II: *The World of the Polis* (1957), vol. III: *Plato and Aristotle* (1957), vol. IV: *The Ecumenic Age* (1974), vol. V: *In Search of Order* (1987), Baton Rouge, LA: Louisiana State University Press.(The finest of Voegelin's productions. Posthumous and unfinished, it does not cover the Middle Ages, Reformation or Modernity; of especial note are the preface and 'Historiogenesis' in vol. IV.)

Voegelin, E. (1966) *Anamnesis*, Munich: R. Piper; abridged edn, trans. and ed. G. Niemeyer, Notre Dame, IN, and London: University of Notre Dame Press, 1978.(Good translation but omits some important early writings on Husserl; includes the fullest statements of Voegelin's philosophy of consciousness.)

Voegelin, E. (1975) *From Enlightenment to Revolution*, trans. J.H. Hallowell, Durham, NC: Duke University Press.(The only part of Voegelin's projected history of political ideas to be published in his lifetime; a very good introduction to Voegelin's thought; incisive exegesis of Voltaire, Helvétius, Comte and Marx *inter alia*.)

Voegelin, E. (1989) *Autobiographical Reflections*, ed. E. Sandoz, Baton Rouge, LA: Louisiana State University Press.(Fascinating insights into Voegelin's life, times and intellect.)

References and further reading

Cooper, B. (1986) *The Political Theory of Eric Voegelin*, Lewiston, NY, and Queenston, Ont.: Edwin Mellen Press.(A circumspect account.)

Hughes, G. and Lawrence, F. (1995) 'The Challenge of Eric Voegelin', *Political Science Reviewer* 24: 398-452. (Survey of Voegelin's thought in the light of the most recently published works.)

Opitz, P.J. and Sebba, G. (eds) (1981) *The Philosophy of Order: Essays on History, Consciousness and Politics*, Stuttgart: Klett-Cotta Verlag.(Volume commemorating Voegelin's eightieth birthday.)

Sandoz, E. (1971) 'The Foundations of Voegelin's Political Theory', *The Political Science Reviewer* 1: 30-73. (Explores Voegelin's reputation and discusses *Anamnesis*.)

(ed.) (1982) *Eric Voegelin's Thought A Critical Appraisal*, Durham, NC: Duke University Press.(Articles by friends and colleagues, comment by Voegelin; ample references; see particularly contributions by Sebba, Webb, Altizer and Havard.)

Voltaire (François-Marie Arouet) (1694-1778)

Voltaire remains the most celebrated representative of the reformers and free-thinkers whose writings define the movement of ideas in eighteenth-century France known as the Enlightenment. He was not, however, a systematic philosopher with an original, coherently argued world-view, but a philosophe *who translated, interpreted and vulgarized the work of other philosophers. His own writings on philosophical matters were deeply influenced by English empiricism and deism. His thought is marked by a pragmatic rationalism that led him, even in his early years, to view the world of speculative theorizing with a scepticism that was often expressed most effectively in his short stories. As a young man, Voltaire was particularly interested in Locke and Newton, and it was largely through his publications in the 1730s and 1740s that knowledge of Lockean epistemology and Newtonian cosmology entered France and eventually ensured the eclipse of Cartesianism.*

After his stay in England Voltaire became interested in philosophical optimism, and his thinking reflected closely Newton's view of a divinely ordered human condition, to which Alexander Pope gave powerful poetic expression in the Essay on Man (1733-4). This was reinforced for the young Voltaire by Leibnizian optimism, which offered the view that the material world, being necessarily the perfect creation of an omnipotent and beneficent God, was the 'best of all possible worlds', that is to say the form of creation chosen by God as being that in which the optimum amount of good could be enjoyed at the cost of the least amount of evil.

Voltaire's later dissatisfaction with optimistic theory brought with it a similar loss of faith in the notion of a meaningful order of nature, and his earlier acceptance of the reality of human freedom of decision-taking and action was replaced after 1748 with a growing conviction that such freedom was illusory. The 1750s witness Voltaire's final abandonment of optimism and providentialism in favour of a more deterministically orientated position in which a much bleaker view of human life and destiny predominates. Pessimistic fatalism was a temporary phase in his thinking, however, and was replaced in turn by a melioristic view in which he asserted the possibilities of limited human action in the face of a hostile and godless condition.

1 Life

Voltaire (born François-Marie Arouet) was one of the most prolific and controversial writers of *ancien régime* France, and his works encapsulate the spirit and ideology of the French Enlightenment. His writings, which span a broad spectrum of genres, do not readily provide the reader with a single, coherent philosophical system or even a systematically argued world-view. In fact, Voltaire deeply mistrusted systems and system-builders, and he frequently satirized, particularly in his later years, the terminology and theories of metaphysicians such as Spinoza, Descartes, Leibniz and Wolff (see LEIBNIZ, G.W.). He wrote mostly, though not exclusively, as a *philosophe* rather than as a philosopher, that is to say, as a dissident polemicist concerned more with persuasive forms of discourse, and the effective advancement of a programme of moral and political reform, than with the pursuit of abstract speculation and analysis *per se*. Some aspects of his philosophical thinking are conducted at an arguably superficial level in the form of short, satirical tales. The term 'philosopher', in its post-nineteenth-century sense, thus sits uneasily with Voltaire. Had he known about the modern meaning of the term, he would have dissociated himself from it. He remained deeply sceptical about the mission of philosophy, and the absurdity of metaphysics is a striking leitmotif of his writings.

In 1726, as the result of a scandal, Voltaire was exiled from Paris, and decided to go to London where he stayed until 1728. His stay in England enabled him to engage with the Baconian tradition of English empiricism and to familiarize himself with the work of Locke, Hobbes, Newton, Clarke, Berkeley, Collins and others. The impact of England on his thinking resulted in 1734 in the publication of the *Lettres philosophiques*, composed originally in English as the *Letters concerning the English Nation*. In 1734 he also started work on another important fruit of the English experience, the *Traité de métaphysique*. This is one of the few Voltairean treatises to contain sustained, relatively sophisticated, philosophical argument and analysis.

The controversy following the publication of the *Lettres philosophiques* forced Voltaire to flee Paris for the chateau of Mme Du Châtelet-Lomont at Cirey, where he stayed for the rest of the decade working on his most ambitious scientific achievement, the *Éléments de la philosophie de Newton* (1738). After a brief stay at Frederick the Great's court at Potsdam, Voltaire settled in Geneva in 1754-5, and the following decade saw the publication

of major historical, polemical and satirical works, including the anti-Leibnizian *Candide* (1759), the *Philosophie de l'histoire* (1759), the *Dictionnaire philosophique* (1764-9) and *Le Philosophe ignorant* (1766). The 1755 Lisbon earthquake proved to be a catalyst in his long debate with others and with himself on the problem of evil and predetermination, culminating in *Candide*, his most brilliant and enduring philosophical narrative.

Between his earliest writings on philosophical matters in 1734 and those works that appeared after 1755, his position changed from one of qualified deistic providentialism and belief in a logically ordered creation, whose harmony had been convincingly demonstrated by Newton's mathematical principles, to one of uncompromising rejection of providentialism and philosophical optimism. Voltaire has few claims to originality as an abstract thinker. As his commentaries in the *Lettres philosophiques* on the work of Francis Bacon, Locke and above all Newton show, he excels in the art of exposition and vulgarization. As a result, he occupies a uniquely influential position as a mediator and disseminator of English and German seventeenth-century philosophy in a France that was to remain dominated by Cartesianism until well into the 1740s. Newton and Locke, in particular, enter the French consciousness through Voltaire's pen.

2 Metaphysics and ethics: the *Traité de métaphysique*

Voltaire's early adherence to rational deism was reinforced in England, although the problems of theodicy and free will had started to preoccupy him well before his exile (see Deism). The first version of the *Traité de métaphysique* (1734) grew out of Voltaire's reactions to Locke's *Essay concerning Human Understanding* (1689), and its composition was shaped by the controversy aroused in France by the *Lettres philosophiques*. By 1736 the text had evolved to take in the implications of Alexander Pope's *Essay on Man* (1733-4), and included chapters dealing with ethics and with man as a social being, reflecting also the influence of Shaftesbury, Mandeville and Clarke. The philosophical balance of the *Traité* moved between 1734 and 1736 from metaphysics to ethics. The *Traité* consists of a brief introduction dealing with doubts about the nature of man, written in the sceptical tradition as transmitted to the eighteenth century through the writings of Montaigne and Le Mothe Le Vayer, followed by nine chapters. Voltaire's philosophy at this stage owes much to the theodicy of English deism and the optimism of Pope and Shaftesbury. The *Traité* also contains traces of the influence of ancient scepticism and epicureanism, the French Pyrrhonian tradition, and late seventeenth-century libertinism (see Libertins).

Much of the *Traité* is anchored to Lockean problems, the second, third, fifth, sixth, seventh and ninth chapters correlating closely to Locke's programme of epistemological investigation. Voltaire follows Locke in the rejection of innate ideas (attributed to Descartes), and in the location of the source of ideas in the senses and the reflective processes. He does not deal with reflection *per se* except to note that thought is not, *pace* Descartes, necessarily a spiritual attribute. Voltaire fully shared Locke's view of the limitations of human understanding, and epistemological modesty was to be the organizing theme of a later wide-ranging essay, *Le Philosophe ignorant*.

Voltaire's starting-point is an awareness of the relative combined with an insistence on the practical aspects of ethics and the lessons of natural morality. Establishing a significant order of priorities, his argument moves from man to God, bypassing Christian notions of creation and original sin. God's existence is deduced empirically in the second, and longest, chapter in the treatise, although knowledge of God's existence is not seen as being either universal or necessary to man's happiness. God's existence is argued from final causes and the Locke-Clarke propositions concerning necessity. The counter-arguments denying certainty in God's existence, and affirming the reality of evil as a negation of the possibility of belief in divine benevolence and omnipotence, are also expounded, but at this stage in Voltaire's thinking they are not allowed to undermine what was essentially a deistic position.

On the question of the soul, Voltaire's view in the *Traité* was that man was composed of thinking matter. He rejected the notion of thought itself as a material compound, but in 1734 he was not prepared to discount entirely the notion of the spirituality and immortality of the soul. He was to advance that argument even more positively in the article 'Ame' in the *Dictionnaire philosophique*. In spite of a tenacious belief in the existence of the soul, often identified with thought itself as a 'principle' invested in inert matter by God, Voltaire did not allow man to occupy a special, privileged place in nature. Man was just a 'reasonable animal'. Unlike the rest of creation, however, man had a measure of free will, although freedom and the power to act freely in accordance with the dictates of reason, always remained for Voltaire severely limited features of man's condition. But on freedom depended morality, and morality provided the focus for the last two chapters of the *Traité*.

Here Voltaire addressed the question of whether man was a social being, together with the related issue of vice and virtue. Concepts of good and evil were determined by what was useful or harmful to society. Parting company with Locke, Voltaire postulated, reflecting again a measure of inconsistency in his thinking, that man possessed innate moral sentiment and an instinct for justice. The assumption of universal morality followed logically from that, and supported the notion of legality, although Voltaire accepted at the same time that specific, contingent laws could well be diverse and contradictory in different parts of the world. Vice and virtue were seen to be relative to social contexts, as was man's understanding of what was useful. However, like Bayle, Voltaire accepted the proposition that all men concurred in the general notions of what was good or evil inherent in those natural laws on which all men in all societies ultimately agreed. Reflecting Shaftesbury's influence, Voltaire's view of human nature was determined by the premise that man possessed latent moral qualities derived from God, and he was never able to reconcile this successfully with his Lockean rejection of innate ideas in other contexts.

Much as God had given the bees a powerful instinct to organize the communal life of the hive, so he had given man, not moral instructions on how to behave, but certain instincts conducive to social and moral life. These were in addition to the instincts of sexuality and self-preservation common to all animals, and included benevolence, compassion, honour, love of truth, pride and the passions, of which even the unattractive, such as pride and envy, had teleological purpose, and served to promote the public good. In all this, God's role in human affairs remained remote and non-interventionist. Voltaire's moral philosophy was essentially a naturalist code in which his concern was less with logical argument than with a passionate determination to define man as an ethical being whose moral conduct and decisions owed nothing to theology or faith, but a great deal to the will to action.

3 Cosmology: *La Métaphysique de Newton*

When it first appeared in 1738, the *Éléments de la philosophie de Newton* was exclusively concerned with optics and gravitational theory, that is with 'natural philosophy', rather than with philosophy as such, and there is little reference to the 'Scholium generale' with which Newton had concluded his *Principia* (1687). However, the 1741 edition of the *Éléments* contained a nine-chapter essay prefacing the commentary on Newtonian physics. This had been printed separately in 1740 as *La Métaphysique de Newton, ou Parallèle des sentiments de Newton et de Leibniz*.

Much of this work is concerned with an exposition of the problems of free will, morality and the mind-body relationship, many of which had been raised earlier in the *Traité de métaphysique*. The new ground that Voltaire now broke concerned Leibniz, and the treatise represents Voltaire's most sustained analysis of Leibnizian science and cosmology. Voltaire's long engagement with Leibnizianism was to drive much of his thinking on questions relating to theodicy and philosophical optimism between 1740 and 1755, culminating in 1756 in his first major attack on optimistic theory in the *Poème sur le désastre de Lisbonne*, reinforced three years later in his satirical demolition of Leibnizian cosmology in *Candide*. Voltaire's attention had been drawn to Leibniz through the work of Christian von Wolff, who had systemized some of Leibniz's thoughts on logic and metaphysics, and whose merits had been commended to Voltaire by Mme Du Châtelet-Lomont and Frederick the Great. Voltaire's knowledge of Leibniz was also based on a Latin translation of the *Monadology* (1714), and he was familiar with the Leibniz-Clarke correspondence concerning space and time.

In *La Métaphysique de Newton* he resumed Newton's arguments from design as a proof of God's existence as a First Cause, outside time and space, of the contingent phenomena of the universe. Whereas in the *Traité de métaphysique* Voltaire's God had been an absent Prime Mover, in *La Métaphysique* he was more sympathetic to the providentialism that marked Newtonian assumptions of divine paternalism. Voltaire still managed to deflect the paradoxes presented by the spectacle of man's suffering and experience of evil, although in *La Métaphysique* the problem of theodicy was raised again in the context of man's freedom to act within the parameters inherent in a providentially ordered human condition. It was here that Leibniz entered the discussion with reference to sufficient reason, necessity and contingency, matters upon which Voltaire had already exchanged views with Clarke. Like Clarke, he argued that divine omnipotence would be compromised if God's will was made subject to the principle of sufficient reason. Voltaire preferred the Newtonian position in which God is seen to have made many things, the reason for whose existence is God's will alone.

Voltaire avoided determinism at this stage in his thought by preserving, by means of extended implication, the capacity for unmotivated choice arising from 'liberty of indifference'. His position on 'liberty of indifference' was

not consistent or rigorously applied, however, and elsewhere - in the *Discours en vers sur l'homme* (1738-42), for example - this received less emphasis than the Lockean view that man's freedom consisted in the power to act and put conscious choices into effect, the question of motivation or absence of motivation being irrelevant. On the question of freedom, and free will in general, Voltaire's views were to change radically, and by the 1750s his position was much more deterministic.

Despite his acceptance of Newtonian arguments from design as proof of an ordered universe in *La Métaphysique*, signs of contradiction and traces of scepticism, to be more strongly articulated at a later stage, were already surfacing in Voltaire's discussion of Leibniz's arguments relating to sufficient reason and to the existence of simple substances, or monads. Leibnizian monadology underpinned the elaboration of the Great Chain of Being theory and the postulation of a cosmos of active metaphysical atoms, or monads, each following a programme of dynamic evolution established by the Creator, and each in harmony with other monads, one of their more controversial characteristics being perceptivity. This infinite multitude of monads constituted a metaphysical substratum of the world of material reality and conscious experience. Voltaire measured Leibniz's theory of monads by simplistically empirical criteria that he associated with Lockean epistemology and Newtonian mathematics. His approach reflected only a partial understanding of the *Theodicy* (1710) and the *Monadology*, and incoporated a great deal of caricatural commentary.

Voltaire objected that Leibniz's postulation of a cosmos of infinitely divisible, nonmaterial and nonspatial matter as an integral component of the material cosmos was self-contradictory. Any explanation of matter that involved the nonmaterial was fanciful, and the notion of perceptivity in all monads, including those constituting inanimate matter, was against common sense. He countered with arguments not entirely free of speculation themselves, drawn from Newtonian science, and advanced a view of matter whose atomic structure was diverse but immutable. Change was accounted for in terms of the effects of gravitational attraction, the 'active force which sets everything in the universe in motion'. Newtonian physics and cosmology did solve for Voltaire some of the philosophical problems arising from his study of Leibnizian monadology and harmony. After 1741, however, notwithstanding his deep admiration for Newton's empirical rigour, he was to have little more to say about Newton. His preoccupation with Leibnizianism, and with its broader philosophical implications, would continue, on the other hand, for many decades.

4 Optimism and freedom

Voltaire's engagement with the problem of evil arose as part of a general broadening of his philosophical interests during his stay in England. Prior to that, he might have already gained some awareness of the issues through the well-publicized debate on theodicy that had taken place between 1697 and 1716 between Archbishop King, Bayle and Leibniz. It was, however, Pope's *Essay on Man* that fired his interest initially in the philosophical issues generated by optimistic theory.

The first serious discussion of the implications of optimism had occurred in the *Traité de métaphysique* where Voltaire had been concerned almost exclusively, as was Pope, with mounting an effective defence of deism against the objections of the atheists. The central argument in this defence related to the question of human ignorance. Voltaire sought to account for evil in terms of a contingent, rather than transcendental, phenomenon, whose nature and effects were falsely determined by the partial, imperfect nature of human understanding. Evil was relative, and to ascribe to God injustice and cruelty was as meaningless as to call him blue or square. The human mind could not conceive of perfection on a cosmic scale, and thus assertions of imperfection were necessarily flawed. Any moral judgment on God's creation must therefore be suspended for lack of evidence.

While Voltaire never quite shared Pope's confidence in providence as an unambiguously benevolent force in human affairs, his early views can be aligned with Pope's 'Whatever is, is right' formulation, and this was fully reflected in his own *Discours en vers sur l'homme* (see especially Part 6). However, the problem of theodicy had been deflected rather than resolved, and after 1740 it re-emerged in Voltaire's thought as a major issue. By 1740 he was aware of Leibniz's arguments on the *Theodicy* as these related to evil, not as a contingent but as an absolute, universal phenomenon that sought to leave intact the providentialist thesis of creation as the work of a beneficent and omnipotent God. Dogmatic promulgation of Leibnizian optimism by German commentators such as Wolff and Kahle became increasingly a subject for ridicule and caustic satire in Voltaire's letters and works. Leibnizian optimism, immortalized in the figure of Candide's tutor, Dr Pangloss, soon came to exemplify for him

the irrelevance of speculative metaphysics to the real world of human suffering and everyday experience.

The evolution of Voltaire's position with regard to Pope and Leibniz, and the related problems of evil, optimism and freedom, is best illustrated not in treatises but in Voltaire's short philosophical stories, and in particular in *Zadig, ou la destinée* (1748) and *Candide, ou l'optimisme*. In *Zadig* he approached the problem of evil from the standpoint of human destiny. Using dialogic format he distilled philosophical optimism into three main propositions that were not argued, or even endorsed, but asserted arbitrarily by an angelic *persona* as articles of faith: chance does not exist; man's destiny is organized in a beneficent way, immutably determined by providence; every apparent evil is part of a larger good and is impenetrable to limited human logic and understanding. While 'Whatever is, is right' still obtained in theory, by 1748 Voltaire no longer adhered fully to his earlier view that evil had no absolute reality. In *Zadig* the focus of attention is clearly on the realities of human suffering and the illogicality of the human predicament rather than on the metaphysical rationale for its existence. Man's place in the Great Chain of Being might have been alloted to him by God for an unknowable purpose, but Voltaire was now firmly of the view that while transcendental perspectives may help to explain the human condition, they contributed nothing to the alleviation of that condition, or to the resolution of the problems that it posed for individual freedom and action.

In the 1750s Voltaire's thinking was dominated by an awareness of the immediacy of evil in day-to-day human experience, and this gradually undermined his earlier sympathy for the optimistic-providentialist position. Events such as the 1755 Lisbon earthquake and the Seven Years War offered further brutal proof of the sterility of optimistic theory, and the darker tones of Voltaire's world-view emerged clearly in the *Poème sur le désastre de Lisbonne*. It was in this didactic poem that he first gave explicit expression to what had by now become his principal objection to the providentialist axiom that all was for the best in the best of all possible worlds. Such a philosophy was now nothing less than a doctrine of despair and a denial of human freedom and of human capacity for action. Philosophical optimism had nothing to offer the stricken citizens of Lisbon.

On the question of free will, Voltaire's earliest discussion of the problem is to be found in the seventh chapter of the *Traité de métaphysique*, where he had accepted that to will and to act accordingly was to be free. Freedom, however, was relative and, even at this 'optimistic' stage in his thinking, it was almost without constraints. He developed his thoughts on freedom further in 1737 in *De La Liberté*. Here he made use of the Lockean definition of freedom as being the choice of whether to act or to refrain from acting, but he now preferred to concentrate on 'power of self-motion' (Clarke's phrase) as being the only true source of freedom. By the mid-1750s Voltaire's views on freedom had evolved considerably, and in ways that mirrored his change of direction with regard to providentialism and philosophical optimism. His denial of free will, and his scepticism with regard to the notion of freedom in human decision-taking and action, was now complete, and it is set out in the fifty-six 'doubts' that constitute the *Doutes sur la liberté* (1752) (see also the articles 'Destin', 'Franc arbitre' and 'Liberté' in the *Dictionnaire philosophique*). Little trace remained of the arguments against determinism that had characterized the *Traité de métaphysique* and *De La Liberté* some two decades earlier, in which a high value had been placed on man's power of judgment and his power to act on that judgment. Men, like animals, are determined by 'instinct' and by 'ideas' that they receive but over whose reception they have no control, and man is subject to the same laws of necessity that govern the whole of nature. The deep sense of futility that was to permeate *Candide* permeates also his list of doubts (see particularly the fourteenth).

Voltaire's loss of faith in philosophical activity reached its climax in *Candide*, in which he finally turned his back on the 'métaphysico-théologo-cosmolonigologie' of Dr Pangloss. Candide inhabits a bleak, arbitrary and possibly godless universe in which evil is an omnipresent, crushing reality. Man might be the victim of forces beyond his control, but experience (not philosophy) taught Candide that the potential for limited, but effective, action still lay within man's grasp: 'Il faut cultiver le jardin' (We must cultivate the garden). The point of thought was action, not the construction of inconclusive speculative systems.

See also: Education, history of philosophy of; Empiricism; Enlightenment, Continental; Rationalism; Will, the

DAVID WILLIAMS

List of works

Voltaire, F.-M. A. de (1980) *Oeuvres complètes de Voltaire (The Complete Works of Voltaire)*, ed. R. Pomeau,

Oxford: The Voltaire Foundation.(The standard edition.)
Voltaire, F.-M. A. de (1734) *Lettres philosophiques (Philosophical Letters)*, Paris: Jore; repr. G. Lanson and A.M. Rousseau (eds), Paris: Garnier, 1964.(This text was published originally in English in 1733 as the *Letters Concerning the English Nation*: see the edition by N. Cronk, Oxford: Oxford University Press, 1994.)
Voltaire, F.-M. A. de (1737) *De La Liberté (On Freedom)*, in *Oeuvres complètes de Voltaire*, Oxford: The Voltaire Foundation, 1980.(This text was based on the seventh chapter of the *Traité de métaphysique*, and developed in letters to Frederick the Great in 1737-8.)
Voltaire, F.-M. A. de (1738-42) *Discours en vers sur l'homme (Discourse in Verse on Man)*, Paris: Prault; in *Oeuvres complètes de Voltaire*, Oxford: The Voltaire Foundation, 1980, vol. XVII, 453-530.
Voltaire, F.-M. A. de (1738) *Éléments de la philosophie de Newton (Elements of Newton's Philosophy)*, Amsterdam: Ledet and Desbordes; in *Oeuvres complètes de Voltaire*, Oxford: The Voltaire Foundation, 1980, vol. XV, 195-805.
Voltaire, F.-M. A. de (1740) *La Métaphysique de Newton ou Parallèle des sentiments de Newton et de Leibniz (Newton's Metaphysics or Parallel between the Views of Newton and Leibniz)*, Amsterdam: Ledet and Desbordes; in *Oeuvres complètes de Voltaire*, Oxford: The Voltaire Foundation, 1980, vol. XV, 195-252.
Voltaire, F.-M. A. de (1748) *Zadig ou la destinée, histoire orientale (Zadig or Fate, an Oriental Tale)*, Paris: Prault and Machuel; repr. H.T. Mason (ed.), Oxford: Oxford University Press, 1971.(This tale first appeared as *Memnon, histoire orientale*.)
Voltaire, F.-M. A. de (1752) *Dialogue entre un brachmane et un jésuite sur la nécessité et l'enchaînement des choses (Dialogue between a Brahmin and a Jesuit on Necessity and the Chain of Being)*, Dresden: Walther; in *Oeuvres complètes de Voltaire*, Oxford: The Voltaire Foundation, 1980, vol. XV, 646-52.(This text was added to the 1752 edition of the *Éléments* as chapter 6.)
Voltaire, F.-M. A. de (1752) *Doutes sur la liberté (Doubts on Freedom)*, Dresden: Walther; in *Oeuvres complètes de Voltaire*, Oxford: The Voltaire Foundation, 1980, vol. XV, 643-6.(This text was added to the 1752 edition of the *Éléments* as chapter 5.)
Voltaire, F.-M. A. de (1756) *Poèmes sur le désastre de Lisbonne, et sur la loi naturelle (Poems on the Lisbon Disaster, and on Natural Law)*, Geneva: Cramer; repr. F.J. Crowley (ed.), Berkeley, CA: University of California Press, 1938.
Voltaire, F.-M. A. de (1759) *Candide, ou l'optimisme. Traduit de l'allemand de m. le docteur Ralph (Candide or Optimism. Translated from the German by Dr Ralph)*, Geneva: Cramer; trans. R. Pearson, *Candide and Other Stories*, Oxford: Oxford University Press, 1990.(Also appears in *Oeuvres complètes*, vol. XLVIII.)
Voltaire, F.-M. A. de (1764-9) *Dictionnaire philosophique portatif (Pocket Philosophical Dictionary)*, Geneva: Cramer; in *Oeuvres complètes de Voltaire*, Oxford: The Voltaire Foundation, 1980, vols XXXV-XXXVI.
Voltaire, F.-M. A. de (1766) *Le Philosophe ignorant (The Ignorant Philosopher)*, Geneva: Cramer; in *Oeuvres complètes de Voltaire*, Oxford: The Voltaire Foundation, 1980, vol. LXII, 3-105.
Voltaire, F.-M. A. de (1784) *Traité de métaphysique (Treatise on Metaphysics)*, Paris: De l'Imprimerie de Société Littéraire-Typographique; repr. H.T. Patterson (ed.), Manchester: Manchester University Press, 1957. (Also appears in *Oeuvres complètes*, vol. XIV, pages 359-503. The text was composed in 1734.)

References and further reading

Alexander, I. (1944) 'Voltaire and Metaphysics', *Philosophy* 19: 19-48.(This is a well-argued analysis of Voltaire's pantheism and materialism, concentrating particularly on empricism and the influence of Locke. Written for specialists.)
Barber, W. (1955) *Leibniz in France. Arnauld to Voltaire*, Oxford: Clarendon.(A clearly argued general study of Leibniz's impact on France. Readers are referred particularly to pages 174-243 dealing with the linked problems of freedom and optimism. Particularly useful for studying the *Traité de métaphysique* and *De La Liberté*, and aimed mainly at specialists.)
Besterman, T. (1962) 'Voltaire and the Lisbon Earthquake, or the Death of Optimism', *Voltaire Essays and Another*, Oxford: Oxford University Press, ch. 3.(A well-written essay with a biographical focus for the general reader.)
Henry, P. (1977) 'Voltaire as Moralist', *Journal of the History of Ideas* 38: 141-6. (For specialists. Useful background for the *Traité de métaphysique*.)
James, E. (1978) 'Voltaire on the Nature of the Soul', *French Studies* 33: 20-33.(A clear exposition of the evolution of Voltaire's views on this central issue, particularly helpful for an understanding of the *Traité de métaphysique*. Mainly for specialists.)
James, E. (1987) 'Voltaire on Free Will', *Studies on Voltaire and the Eighteenth Century* 249: 1-18.(An

informative study of the impact of the thought of Samuel Clarke, Anthony Collins and John Locke on this central issue in Voltairean metaphysics. For the specialist.)

Kendrick, I. (1956) *The Lisbon Earthquake*, London: Methuen.(A general account of the earthquake and its intellectual impact. Readers are referred to pages 119-31 and 134-9 for information relevant to Voltaire. For the general reader.)

Niklaus, R. (1994) 'Voltaire et l'empirisme anglais', *Revue internationale de philosophie* 48: 9-24.(A detailed study of the debt owed by Voltaire to English empiricism. Particularly relevant to Hume and Locke in the context of the *Traité de métaphysique* and the *Éléments de la philosophie de Newton*. For specialists.)

Spink, J. (1960) *French Free Thought from Gassendi to Voltaire*, London: University of London Press.(A well-documented overview of the key stages in the development of sceptical traditions of thought in France. On Voltaire, see particularly pages 312-24. For the general reader.)

Wade, I. (1969) *The Intellectual Development of Voltaire*, Princeton, NJ: Princeton University Press.(A rich intellectual biography. Part IV deals with Voltaire's philosophical thought and development and has special sections on Newton, Locke, Leibniz, Spinoza and Malebranche relevant to Voltaire's position on cosmology, optimism and freedom. For the general reader.)

Voluntarism

Voluntarism is a theory of action. It traces our actions less to our intellects and natural inclinations than to simple will or free choice. Applied to thinking about God's actions, voluntarism led late medieval philosophers to see the world's causal and moral orders as finally rooted in God's sheer free choice, and to take God's commands as the source of moral obligation. Medieval voluntarism helped pave the way for empiricism, Cartesian doubt about the senses, legal positivism and Reformation theology.

Our wants and our natures (including our rationality) help shape our choices and actions. Intellectualists and voluntarists offer rival views of how great a role these factors play. Aquinas (§§12-13), a major intellectualist, held that our will is a rational appetite, a desire for items we think good for us. Its nature is to choose what we judge good for us; we can choose among actions because different acts can seem good to us in different respects. We can err about what is good for us, or fail to will it. But we cannot knowingly reject anything that seems to us the best in the circumstances. Further, while our wills do influence our intellects, we cannot determine our intellects to accept beliefs in the face of sufficient evidence against them. So Aquinas sees our passions (which shape what seems good to us), our nature as good-seekers and our intellects as moulding our choices.

Duns Scotus (§14) adds a second rational desire to Aquinas' picture. Anselm of Canterbury (§6) held that we have a will for what seems good for us and also a will for moral rectitude. Scotus generalizes Anselm's second 'will': for him, we naturally incline to what seems good for us, but also naturally love things for their objective goodness. So we can reject what seems best for us if we desire more what seems objectively best, or reject what seems objectively best if we desire more what seems best for us: our natures leave us a wider range of choice. Yet both natural will-inclinations are towards good. So we still cannot knowingly reject happiness or whatever else seems best in the circumstances both objectively and for us.

William of Ockham (§10) tends towards voluntarism. For Ockham, it is false that we can choose freely because we have intellects or rational desires: it is just a basic fact about us that we can do so. So our will is independent of reason and natural inclination. Neither restricts our choices. Ockham thinks we can knowingly reject happiness or whatever else seems in all ways best, and will evil because it is evil rather than because we think it in some way good. If we can, that an act seems best does not fully explain our doing it. For we can always reject what reason or our natures counsel.

For both intellectualists and voluntarists, God is 'absolutely' able to do whatever he is powerful enough to do, and 'ordinately' able to do just whatever he *would* do, given his character and covenants. For both, God is absolutely powerful enough to lie, but ordinately cannot do so. That God should lie is no contradiction. So God's omnipotence can effect it. But God would never actually lie. For intellectualists, this is because, as perfectly rational, God sees that lying is wrong and chooses accordingly. For voluntarists, God would not lie just because he has so chosen.

For Aquinas, all ten Commandments are necessary moral truths, which God sees and proclaims. For Scotus, necessarily, loving God is right and hating him wrong: as God is infinitely good, there can be no reason to prefer anything else to him, and so rational agents must love him above all. All other acts, though naturally good or bad (perfecting or destroying our nature), are right or wrong contingently, because God commands or prohibits them. Further, God sometimes revokes his contingent commands. Yet according to Scotus, this does not make morality arbitrary. God commands or prohibits acts due to their natural goodness or badness, which reason sees. God revokes precepts only rarely, where the values a law serves are better served by exception. For Scotus, God cannot command irrationally. So where God makes moral exceptions, these are rational; reason might anticipate his doing so. Even contingent moral laws are 'natural' in the sense of being accessible to reason.

Ockham held that only acts in accord with reason can be virtuous. But reason tells us that if there is one personal God, we should do as he says, and so reason tells us that both reason and a revelation by such a God can teach us what is right. Still, while both reason and revelation can teach right and wrong, only God's command *makes* acts right or wrong. God in fact commands only acts that accord with right reason. He need not have done so. For Ockham, only a superior's command can impose an obligation, and God has no superior: so God has no

obligations. Evil is acting counter to obligation. So according to Ockham, God cannot do evil. God in his absolute power can order fornication or theft. If he did so, these acts would be right. Even if God made Satan hate God, God would do no evil (as he has no obligation not to do this). Nor would Satan, as he would not control his own actions.

For Ockham, God can cause any creaturely effect unaided or block any creaturely cause. So even if a cause reliably yields an effect, the occurrence of one does not entail that of the other: causal connections are contingent. If so, a priori reasoning can tell little about real causal relations. Consequently, Ockhamist voluntarism was a prod to empiricism. Since it could in any case be God rather than a creature who produces an effect, voluntarism led too towards scepticism about created causality (see Nicholas of Autrecourt §2). Among the effects God can produce are apparent sensations of an object. So voluntarism insisted that what we seem to sense need not really be there, thus making God a remote ancestor of Descartes' deceiving demon (see Descartes, R. §4). The idea that only a will's command imposes obligations helped sow the seeds of legal positivism (see Legal positivism §§1-2). The voluntarist claim that God need not have accepted acts consonant with our nature or in our power as meritorious was a source of the Reformers' insistence that human salvation is wholly a matter of divine grace.

See also: Ash'ariyya and mu'tazila §5; Freedom, divine; Occasionalism §§1-2; Religion and morality

BRIAN LEFTOW

References and further reading

Adams, M. (1987) *William Ockham*, Notre Dame, IN: University of Notre Dame Press.(Detailed, careful treatment of Ockham's logic, ontology, natural philosophy and philosophical theology.)

Adams, M. (1987) 'William Ockham: Voluntarist or Naturalist?', in J. Wippel (ed.) *Studies in Medieval Philosophy*, Washington, DC: Catholic University Press.(An overview, with some emphasis on ethics.)

Bourke, V. (1964) *Will in Western Thought*, New York: Sheed & Ward.(Historical survey, more broad than deep.)

Courtenay, W.J. (1984) *Covenant and Causality in Medieval Thought*, London: Variorum.(Discusses voluntarism, the absolute/ordained power distinction, causation and theology.)

Freppert, L. (1988) *The Basis of Morality According to William Ockham*, Chicago, IL: Franciscan Herald Press.(Clear and straightforward.)

Kennedy, L. (1986) *Peter of Ailly and the Harvest of Fourteenth-Century Philosophy*, Queenston, Ont.: Mellen.(Argues that a typical voluntarist's view of God's absolute power led him to scepticism.)

Oakley, F. (1964) *The Political Thought of Pierre D'Ailly*, New Haven, CT: Yale University Press.(Clear, detailed treatment of voluntarist political philosophy.)

Oberman, H. (1963) *The Harvest of Medieval Theology*, Cambridge, MA: Harvard University Press.(The absolute/ordained distinction in action in the voluntarist theology of Gabriel Biel.)

Wolter, A. (1990) *The Philosophical Theology of John Duns Scotus*, Ithaca, NY: Cornell University Press.(Collected articles of the twentieth century's pre-eminent Scotus scholar; several are concerned with will and ethics.)

Voluntarism, Jewish

Voluntarism with respect to humanity and divinity became a powerful current in medieval Jewish philosophy, partly in response to the Neoplatonic doctrine of eternal and necessary emanation, which seemed to rob God of the freedom to create, and partly in response to predestinarianism. Solomon ibn Gabirol and Hasdai Crescas were among the Jewish philosophers whose metaphysics gave pride of place to the divine will over intellect, like medieval Christian voluntarists. For many other Jewish thinkers, the centrality of actions sets voluntarism firmly into the context of moral responsibility rather than of metaphysics. Predestinarian arguments like those of Abner of Burgos seemed to Jewish thinkers to rob human beings of moral responsibility. Among the typical defenders of Jewish voluntarism against these arguments was Abraham Bibago.

1 Will versus intellect

In Western Christianity, voluntarism has long been associated with the precedence of will over intellect in humanity and divinity. An early Jewish voluntarist in this sense, Ibn Gabirol, made the will the intermediary between God and the universal matter and form that anchor his ontology. In classical Neoplatonism, intellect, not will, was the first emanation in the divine production of the world. Crescas reacted against the Aristotelianism of Maimonides, much as Duns Scotus reacted against the Aristotelianism of Aquinas, by making will, not intellect, the basis of the affinity between God and humanity (see Crescas, H.; Maimonides, M.).

Ibn Gabirol and Crescas, however, had a limited impact on later Jewish philosophy. Ibn Gabirol did profoundly influence the Franciscan Augustinians, who knew him as Avicebrol or Avicebron and relied on him in laying the groundworks of later Christian voluntarism. His chief philosophic work, *Mekor Hayyim (Fountain of Life)*, survives intact only in Latin translation, *Fons Vitae*. In Jewish thought, only traces of his approach to foregrounding God’s will are found - in Kabbalah, in the work of Joseph ibn Tzaddik (d. 1149), and in the writings of the thirteenth-century philosopher-Kabbalist, Isaac ibn Latif (see Kabbalah). As for Crescas, his approach to this issue, for many later philosophers, was eclipsed by that of Maimonides.

Medieval Christian philosophers emphasized the will/intellect issue more than Jewish philosophers. Christianity is classically a religion of faith: inner states, beliefs and contemplation are foundational. Greek philosophy, especially Platonism, can place a similar emphasis on the inner life of the soul. This helps explain Augustine’s affinity for Platonism. For the Greeks, however, intellect was the dominant inner faculty: as rational animals, humans differ from other beasts in their powers of rational discrimination. Intellect is determinative for rational choice; will is the dynamic power. Socrates even spoke of knowledge as sufficient to determine choices. For St Paul and Augustine, by contrast, one might know the good and still not do it. Will rather than intellect seemed the dominant inner faculty.

In Judaism actions, not inner states, are the central goal. Religious commitment is expressed not by affirming dogmas but by fulfilling God’s commandments. Ancient Judaism has no creed, and questions about the dominance of will or reason were not central at the outset. But for actions to have moral value and to merit reward or punishment, they must be freely chosen. Freedom of choice, accordingly, was upheld by most Jewish thinkers, typically alongside a strong commitment to divine omnipotence.

In the early fourteenth century, freedom became a particularly vexed doctrinal issue when Abner of Burgos, who had converted to Christianity late in life (changing his name in the process to Alfonso de Valladolid), attacked the idea of human freedom, arguing that God’s omniscience and providence require all human acts to be determined. Seeking to preserve both human freedom and divine providence, Isaac Pollegar (also spelt as Pulgar) responded by proposing that, since the whole universe is one individual, God’s will joins with our own in the production of free human actions, much as the soul’s will joins with the nerves of the hand to move the finger. Moses Narboni also responded to Abner and defended free choice on Maimonidean lines. Over a century later, the determinism of Hasdai Crescas may still have reflected the influence of Abner.

2 Will as freedom

In Jewish philosophy voluntarism was typically a question of the nature of the will and its freedom. Aristotelians generally saw free choice in terms of rationality. But many thinkers put the emphasis on volition, especially in

regard to divine freedom. Maimonides, for instance, demarcated philosophers of the Aristotelian tradition from the adherents of the Torah by the commitment of the latter to the world's creation by God's free act - which we humans can construe only in terms of our idea of will.

Abraham Bibago (d. *c.* 1489), writing two decades before the expulsion of the Jews from Spain, begins *Derekh Emunah (The Way of the Faith)*, his book on divine providence and Jewish faith, with a disquisition on divine will and agency. Starting from an Aristotelian distinction, he describes voluntary agents as being conscious of their actions and being able to perform either of two contrary actions. A natural agent is unconscious of its action and can perform only one of two opposing actions essentially. Divine agency is obviously not natural: God is omniscient. Voluntary action, however, typically aims at a goal which the agent lacks, and lack, or deficiency, is impossible in God. Philosophers consequently posit a third kind of agency for God, conscious but performing only one of two opposing actions, and not in pursuit of a goal.

This doctrine harks back to Plotinus' imagery, in which the world proceeds from the One as light issues from the sun (see Plotinus §5). To many religious philosophers this doctrine seemed to say that God acts involuntarily. But Aristotle's conception of voluntary agency allowed the necessary agency of emanation to be called voluntary, since voluntary agency for Aristotle required only the absence of external compulsion and the presence of knowledge. With this in mind, Avicenna specified that God's agency in emanation was the sheer result of knowledge, and could therefore be called voluntary in Aristotle's sense. Will here becomes God's joy in the overflow of goodness that is in fact an inevitable consequence of that goodness. Maimonides judged such combining of necessity and will to be incoherent, but the philosophers most deeply committed to the Neoplatonic-Aristotelian synthesis continued to call necessary emanation voluntary.

Against these philosophers, Bibago argued that a God who acts by necessity would be just as deficient as a voluntary agent who has needs. Furthermore, a necessary cause must produce an effect similar to itself, as fire produces heat. Thus necessary agency does not explain how an infinite deity can produce a finite universe. Only voluntary agency, in Bibago's view, could solve this problem.

But if a voluntary agent lacks the end it intends, why would a voluntary God not be deficient? Bibago explains that ends can be either internal or external. Pursuit of an external end implies a deficiency in the pursuer. But God acts for an internal end, his own essence, and thus remains without deficiency.

3 Will versus necessity

The intellectualist necessitarianism that Bibago opposed was denounced by traditional philosophers and theologians in Christianity, Judaism and Islam. It was part of the so-called Latin Averroism condemned in 1277 by the Bishop of Paris (see Averroism, Jewish; Averroism). The most 'Averroist' of Jewish philosophers was Isaac Albalag (*fl.* thirteenth century), who translated into Hebrew al-Ghazali's *Maqasid al-falasifah (Intentions of the Philosophers)*, a summary of Avicenna's philosophy. Albalag was routinely attacked by traditional minded Jewish philosophers of the fifteenth century. He seems to be the target of Bibago's arguments against those who considered God an agent by necessity. The Averroist fusion of necessity and freedom forms the background for Spinoza's combination of liberty and necessity and helps explain the hostility of traditional Jews towards his philosophy (see Spinoza §4).

From a Christian perspective, perhaps because of Augustine's Platonic affinities, Neoplatonism has often been viewed as more congenial to religion than the Aristotelianism that came to dominate later medieval philosophy. For Jewish voluntarists, however, the threat to divine and human freedom seemed to come from Neoplatonic necessitarianism. The teleological perspective of Jewish voluntarism and its commitment to human freedom gave it an affinity to Aristotle, although Aristotle assigns no will to God and does not offer a theory about a faculty of free will.

The idea of a voluntary deity resonates with the biblical portrayal of God. For the voluntarists, what God and humans share, the divine image in humanity, is not simply the intellect but the spontaneity of freedom - not just an intellectual link but the possibility of personal dialogue. The spirit of Jewish voluntarism, understood in this vein, persists in the philosophy of dialogue pioneered in modern Judaism by Martin Buber and Franz Rosenzweig.

See also: Crescas, H.; Free will; Ibn Gabirol, S.; Neoplatonism

ALLAN LAZAROFF

References and further reading

Bibago, A. (before 1489) *Derekh Emunah (The Way of the Faith)*, ed. C. Fränkel-Goldschmidt, Jerusalem: Bialik Institute, 1978.(Describes voluntary agents as being conscious of their actions and being able to perform either of two contrary actions; holds that voluntary action typically aims at a goal which the agent lacks.)

Bourke, V.J. (1964) *Will in Western Thought: An Historico-Critical Survey*, New York: Sheed & Ward.(Eight different meanings of will, with introductory and concluding chapters.)

Goodman, L.E. (ed.) (1992) *Neoplatonism and Jewish Thought*, Albany, NY: State University of New York Press.(A collection of eighteen essays, including three on Ibn Gabirol and three on Maimonides, that pay attention to the divine will, as does the editor's introduction.)

Guttmann, J. (1964) *Philosophies of Judaism: The History of Jewish Philosophy from Biblical Times to Franz Rosenzweig*, trans. D.W. Silverman, New York: Holt.(An older but still excellent analysis that includes voluntaristic themes.)

Katz, S.T. (ed.) (1975) *Jewish Philosophers*, New York: Bloch.(Articles from the *Encyclopedia Judaica* on philosophers from the ancient, medieval and modern periods, with an introduction and a fourth part on Jewish thought since 1945 by the editor.)

Lazaroff, A. (1981) *The Theology of Abraham Bibago: A Defense of the Divine Will, Knowledge and Providence in Fifteenth-Century Spanish-Jewish Philosophy*, Tuscaloosa, AL: University of Alabama Press, ch. 4. (Discusses voluntaristic themes.)

Sirat, C. (1985) *A History of Jewish Philosophy in the Middle Ages*, Cambridge: Cambridge University Press. (Comprehensive and detailed survey with numerous references to the themes of free will and divine will.)

Wolfson, H. (1965) *Religious Philosophy: A Group of Essays*, New York: Atheneum.(Contains three essays dealing with the nature of the will and its freedom.)

Von Wright, Georg Henrik (1916-)

G.H. von Wright is one of the most influential analytic philosophers of the twentieth century. Born in Helsinki, Finland, von Wright did his early work on logic, probability and induction under the influence of logical empiricism. In 1948-51 he served as Ludwig Wittgenstein's successor at Cambridge, but returned to his homeland and later became a member of the Academy of Finland. He did pioneering work on the new applications of logic: modal logic, deontic logic, the logic of norms and action, preference logic, tense logic, causality and determinism. In the 1970s his ideas about the explanation and understanding of human action helped to establish new links between the analytic tradition and Continental hermeneutics. Von Wright's later works, which are eloquent books and essays written originally in his two native languages (Swedish and Finnish), deal with issues of humanism and human welfare, history and future, technology and ecology.

Georg Henrik von Wright was born in Helsinki on 16 June 1916. The family, belonging to the Swedish-speaking minority in Finland, had moved from Scotland in around 1650. (Phonetically the name is pronounced as 'rigt' rather than 'rait'.) In 1934 von Wright began to study philosophy at the University of Helsinki under the supervision of Eino Kaila. Kaila, a charismatic teacher and cultural personality, was the Finnish associate to the Vienna Circle (see Vienna Circle). The young von Wright's studies in logic and philosophy of science in the spirit of logical empiricism led to his doctoral dissertation *The Logical Problem of Induction* in 1941. His main thesis is that Hume's problem is logically unsolvable: the demand for a justification of induction is self-contradictory. Von Wright later showed his expertise in probability theory and its history in a number of articles on subjective probability, randomness, paradoxes of confirmation and inductive logic. The most important of his contributions is *A Treatise on Induction and Probability* (1951), which carefully develops a system of eliminative induction in the tradition of Bacon and Mill by using the concepts of necessary and sufficient conditions.

In the spring of 1939 von Wright went to Cambridge to meet C.D. Broad, and was permitted to follow Ludwig Wittgenstein's lectures. Wittgenstein made a very strong personal impression on the young Finn. The contact and friendship was renewed after the war. Regular discussions with G.E. Moore also made an impact on his philosophical views. In 1948 von Wright was invited to become Wittgenstein's successor as Professor at Trinity College, Cambridge. After Wittgenstein's death in 1951 he decided to return to Helsinki, where already in 1946 he had been appointed the Swedish-language Professor of Philosophy. Together with Elizabeth Anscombe and Rush Rhees, von Wright became the executor and editor of Wittgenstein's posthumous works. The study, organization, systematization and publication of this exceptionally rich *Nachlass* was a lifelong task for him.

Von Wright's early work on philosophical logic discusses classical topics like predication, negation and entailment. The characterization of logical truth by means of the constituents of monadic predicate logic led his most renowned Finnish student Jaakko Hintikka to a general theory of distributive normal forms in 1953. In his Cambridge years von Wright became interested in the logical properties of various kinds of modalities: alethic (possible, impossible, necessary), deontic (permission, prohibition, obligation), epistemic (verified, undecided, falsified). In 1951 he published *An Essay in Modal Logic*, which studies in a syntactical way various deductive systems of modal logic. In the same year he published perhaps his most famous article, 'Deontic Logic', in *Mind*, which made him the founder of modern deontic logic. These works of philosophical logic had a profound influence on the direction of analytic philosophy - including action theory and the philosophy of law. Von Wright distinguishes technical oughts (means-ends relationships) from norms issued by a norm authority. The classical book *Norm and Action* (1963a) discusses philosophical problems concerning the existence of norms and the truth of normative statements; on these issues the author has often changed his mind. Von Wright's main work on meta-ethics and value theory is *The Varieties of Goodness* (1963b), which analyses different concepts of goodness. The two major works in 1963 established his position as a leading philosopher of his time.

In *Explanation and Understanding* (1971) von Wright turned to philosophical problems concerning the human and social sciences. He defends a manipulative view of causality, where the concept of action is primary to the concept of cause. He argues that human action cannot be explained causally by Hempel's covering law model, but has to be understood intentionally (see Hempel, C.G.). The basic model of intentionality is practical syllogism, which relates action by a sort of logical connection with wants and beliefs. This work, sometimes characterized as

anti-positivist analytical hermeneutics, has been regarded as an important bridge between analytic philosophy and the Continental tradition. Von Wright's studies in truth, knowledge, modality, lawlikeness, causality, determinism, norms and practical inference were published in 1983-4 as three-volume *Philosophical Papers*. The preparation of a volume in the Library of Living Philosophers was started in the early 1970s, but its publication was delayed until 1989.

In 1961 von Wright inherited Kaila's position as one of the twelve members of the Academy of Finland; this research professorship is the highest honour that Finland has given its most prominent scientists. Von Wright has written eloquent essays in Swedish and Finnish on the history of ideas and the philosophy of culture. In the 1950s he explored the philosophy of history (Tolstoi, Spengler) and myths about knowledge (Genesis, Faust). In his later work on humanism and the ethics of science, he became increasingly critical of the modern scientific-technological civilization, its narrowly instrumental concept of rationality, and its myth of progress. Von Wright's public pleas for peace and human rights, a better society and a more harmonious coexistence between human beings and nature have made him the most esteemed intellectual in the Scandinavian countries.

See also: Action; Deontic logic; Induction, epistemic issues in; Modal logic; Scandinavia, philosophy in

ILKKA NIINILUOTO

List of works

Von Wright, G.H. (1941) *The Logical Problem of Induction*, Helsinki: Societas Philosophica Fennica; 2nd revised edn, Oxford: Blackwell, 1957.(Hume's problem and its attempted solutions.)

Von Wright, G.H. (1951a) *A Treatise on Induction and Probability*, London: Routledge & Kegan Paul. (On eliminative induction.)

Von Wright, G.H. (1951b) *An Essay in Modal Logic*, Amsterdam: North Holland.(Axiomatic systems of modal logic.)

Von Wright, G.H. (1955) *Tanke och förkunnelse (Thought and Prophecy)*, Helsinki: Söderströms; also publ. in Finnish as *Ajatus ja julistus*, Helsinki: Söderström, 1961.(Includes essays on Tolstoi, Dostoevsky, Spengler and Toynbee.)

Von Wright, G.H. (1957a) *Logical Studies*, London: Routledge & Kegan Paul.(Papers on logical truth, quantification, predication, negation, deontic logic and entailment.)

Von Wright, G.H. (1963a) *Norm and Action*, London: Routledge & Kegan Paul; also publ. in German as *Norm und Handlung*, Königstein: Scriptor, 1979.(The truth and existence of norms, and a theory of action.)

Von Wright, G.H. (1963b) *The Varieties of Goodness*,, London: Routledge & Kegan Paul.(A meta-ethical study of different meanings of the concept of goodness.)

Von Wright, G.H. (1971) *Explanation and Understanding*, London: Routledge & Kegan Paul; also publ. in German as *Erklären und Verstehen*, Frankfurt: Atheneum, 1974.(Positivism and hermeneutics in history and the social sciences.)

Von Wright, G.H. (1983a) *Practical Reason: Philosophical Papers, vol. 1*, Oxford: Blackwell.(Papers on practical inference, promises, norms, truth, and logic.)

Von Wright, G.H. (1983b) *Philosophical Logic: Philosophical Papers, vol. 2*, Oxford: Blackwell. (Logical paradoxes, preference, time and change.)

Von Wright, G.H. (1984) *Truth, Knowledge, and Modality: Philosophical Papers, vol. 3*, Oxford: Blackwell. (Truth-logic, predication, determinism, laws of nature.)

Von Wright, G.H. (1993b) *The Tree of Knowledge and Other Essays*, Leiden: Brill; also publ. in German: *Erkenntnis als Lebensform*, Vienna: Böhlau, 1993.(Myths about knowledge and progress, humanism and the good.)

References and further reading

Apel, K.O. (1984) *Understanding and Explanation*, Cambridge, MA: The MIT Press. (Comments on von Wright's relation to Continental hermeneutics.)

Hintikka, J. (ed.) (1976) *Essays on Wittgenstein in Honour of G.H. von Wright*, Amsterdam: North Holland. (Includes material on von Wright's relation to Wittgenstein.)

Manninen, J. and Tuomela, R. (eds.) (1976) *Essays on Explanation and Understanding*, Dordrecht: Reidel. (Critical essays and replies.)

Niiniluoto, I., Sintonen, M., and von Wright, G.H. (eds) (1992) *Eino Kaila and Logical Empiricism*, Helsinki: Societas Philosophica Fennica.(On von Wright's teacher.)

Schilpp, P.A. and Hahn, L.E. (eds.) (1989) *The Philosophy of Georg Henrik von Wright*, La Salle, IL: Open Court.(Includes an intellectual autobiography, thirty-two critical essays and replies, and a complete bibliography up to 1988.)

Vulnerability and finitude

Power has always been a central category of political thought and theory; its counterparts, powerlessness or vulnerability, and more generally finitude, have seemingly been much less discussed. Yet finitude has been a theme for many writers, particularly on metaphysical and epistemological topics, who emphasize that claims about human knowledge cannot presuppose that we command a God's-eye view of ourselves or the world; while vulnerability has been a theme in ethical and political philosophy, challenging idealized 'models of man' that take exaggerated views of human capacities and autonomy, and which overlook the mundane realities of dependence, poverty and frailty.

1 Finitude

Writing on finitude emphasizes that we can look at matters only from the vantage point of finite and fallible beings, of men rather than of angels. It criticizes some traditional metaphysical writing for taking divine omniscience and omnipotence, and other divine 'perfections', as models or standards of power, goodness and knowledge, for measuring human achievements by these yardsticks, and for concluding that human finitude is to be seen as (ineliminable) deficiency, lack or privation.

Such metaphysical writers - (Plato, Descartes and other rationalists are often cited as paradigms - see Rationalism) - assume that, finite as they are, human beings can form a conception of unlimited power and knowledge, and can understand their own limitations from that perspective. In doing so these writers take divine or infinite power, reason, goodness and knowledge as models for human performance - or rather lack of performance. Assumptions that human beings can appeal to powers of understanding and standards of perfection appropriate to an infinite being or to a transcendent vantage point have repeatedly been criticized. Yet many writers who have thought that they had successfully put aside appeals to transcendent, ideal standards have been accused by their successors of covertly reintroducing one or another variant of those very standards. For example, Kant (1781/1787) insists that he starts from the standpoint of *human* reason, but many of his critics have thought that he reinstates a God's-eye view of the world and human life, and in particular that his accounts of freedom and of ethics require him to do so (see Kant, I. §§9-11; Kantian ethics). Nietzsche and Heidegger both criticized Kant for this failing, and both aimed to put aside any appeal to a transcendent standpoint or to the knowledge and goodness of an infinite being; both have in their turn been criticized for tacit reliance on such a standpoint.

2 Vulnerability

In one way, writing in ethics must assume the possibility of human limitations: it would be pointless to propose norms or standards unless the possibility of failure to live up to them was acknowledged. However, much writing on ethics, while allowing for human failure, has invoked idealized conceptions of human rationality, knowledge and autonomy, or alternatively of human intimacy and affections, which are not achieved by many or even by any human beings (see Autonomy, ethical). Writing that is unrealistic in these ways can overlook the vulnerability of human lives and may then recommend action that takes too little account of others' vulnerabilities. Critics of various idealized conceptions of the human agent in ethics and political philosophy have emphasized a variety of ways in which people may be vulnerable. Some have stressed ways in which poverty limits action and capacities for action; others have pointed more generally to the fragility of human life, action and achievement; yet others have pointed to the fact that dependence on others is normal and far from being a form of moral failure may be a form of moral achievement.

There are then many ways in which writing on ethics can underestimate the significance of human vulnerability. For example, certain models of rational choice, and with them much utilitarian writing, assume idealized abilities to choose rationally, which are not in fact achievable by human beings: ordinary people may find such reasoning and the prescriptions it points to unconvincing (see Utilitarianism; Rationality, practical).

Some accounts of justice, in particular those favoured by libertarians, see justice as a matter only of according others like liberties (see Libertarianism). They deploy idealized views of human autonomy, by which others' needs and vulnerability are irrelevant to justice. Many other accounts of justice take some account of poverty and need (see Liberalism). They may, for example, argue for welfare systems that could protect people from these sources of

vulnerability, or for certain sorts of equality. Yet theories of justice often say little about vulnerability that arises from subordination or dependence on others, in particular about those that can arise from quite ordinary familial or gender relationships (see Family, ethics and the).

Some of these misleading idealizations have been challenged in feminist writing, in discussions of friendship, and in writing on the virtues (see Feminist ethics; Feminist political philosophy; Virtue ethics; Virtues and vices). A common criticism of all writing on justice is that it takes an idealized view of the autonomy of individuals or of citizens, forgetting that people are also linked in friendship and family, and that family relationships are often relationships of dependence, and so of power and of vulnerability. On the other hand, some of the proponents of friendship and virtue ethics, who acknowledge the moral importance of relationships and attachments, tend to take a rosy and even a falsely idealized view of the quality of close relationships, and thereby ignore the fact that vulnerabilities can be created and deepened by dependence. Feminist writers have often been more realistic about this range of vulnerabilities.

See also: Friendship; God, concepts of; Human nature; Ideals; Levinas, E.; Recognition

ONORA O'NEILL

References and further reading

Blum, L.A. (1980) *Friendship, Altruism and Morality*, London: Routledge & Kegan Paul.(Argues for the moral importance of friendship and sympathy; criticizes the 'models of man' used in moral philosophy which neglects these themes.)

Drèze, J. and Sen, A. (1990) *Hunger and Public Action*, Oxford: Clarendon Press.(Discusses the realities of poverty and hunger. Rather than assuming that basic capabilities are given, we should take action to secure them.)

Heidegger, M. (1927) *Sein und Zeit*, trans. J. Macquarrie and E. Robinson, *Being and Time*, Oxford: Blackwell, 1962.(Depicts human beings as finite, worldly, historical, oriented to care for other beings.)

Heidegger, M. (1943) 'The Word of Nietzsche: "God is Dead"', in W. Lovitt (trans. and ed.) *The Question of Technology and Other Essays*, New York: Harper & Row, 1977.(Heidegger on Nietzsche's writings on the 'death of God'.)

Kant, I. (1781/1787) *Critik der Reinen Vernunft*, trans. N. Kemp Smith, *Critique of Pure Reason*, London: Macmillan, 1973.(The limits of human reason and its incapacity to sustain metaphysical claims which assume a God's-eye view are basic themes.)

Levinas, E. (1982) *Éthique et infini*, Paris: Fayard; trans. R. Cohen, *Ethics and Infinity*, Pittsburgh, PA: Duquesne University Press, 1985.(Introductory discussion of themes such as the ethical primacy of the face-to-face relationship with the other and its relation to the infinite.)

Nietzsche, F. (1882) *Die fröhliche Wissenschaft*, trans. W. Kaufmann, *The Gay Science*, New York: Random House, 1974, esp. no. 125.(Nietzsche's most famous discussion of the 'death of God', the human realization of our lack of an infinite vantage point.)

Nussbaum, M. (1986) *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*, Cambridge: Cambridge University Press.(Explores Greek ethics and its awareness of the vulnerability of human life to contingency.)

O'Neill, O. (1996) *Towards Justice and Virtue: A Constructive Account of Practical Reasoning*, Cambridge: Cambridge University Press.(Argues for a conception of practical reasoning which does not presuppose a transcendent vantage point and can support linked accounts of justice and virtue which take account of human vulnerabilities.)

Pateman, C. (1988) *The Sexual Contract*, Cambridge: Polity Press.(Considers the justice of the traditional marriage contract and of the dependence it created.)

Vygotskii, Lev Semënovich (1896-1934)

Vygotskii was a Soviet psychologist, the most comprehensive in creative reach and the most influential. Trained in literary studies and originally active as a critic, he took a post in a pedagogical institute and came thus to psychological science, with a special interest in child development. That was the period of foundational debates between rival schools of psychology, intensified in the Russian case by the Revolution of 1917 and the subsequent campaign for a Marxist school. Vygotskii became the major theorist at the central Institute of Psychology. While dying of tuberculosis he worked his intensive way through contested claims to know what mind is and how it acts. His profuse reflections on that large contest remained largely unpublished for decades, while disciples echoed his call for a 'cultural-historical' approach to a unified science of the mind, and actually worked on the mental development of children and the neuropsychology of brain damage. The concept of 'activity', which was supposed to resolve philosophical issues, served largely to evade them, while harmonizing with Marxist-Leninist slogans on 'practice'. Among Western cognitive psychologists Vygotskii acquired a tardy reputation as a pioneer who emphasized social interaction in the mental development of children. The publication of his major works in the 1970s and 1980s revealed a much broader theorist. His central theme was the obvious truth at the basis of each artistic and psychological school, the lure of an effort to unify all of them, and the present impossibility of achieving such unification within science, outside philosophical speculation.

1 Life

Son of a bank officer in a provincial town of tsarist Russia, Vygotskii was given a secularized Jewish education until he entered Moscow University, where he studied language and literature. He was already publishing critical essays when he graduated in 1917, the year of the communist Revolution. Vygotskii returned to his home town, where he spent the years of civil war and economic collapse teaching in a secondary school and a pedagogical institute, adding child development and mental defect to literary art as crucial problems for claims to know what mind is and how it acts. A Marxist approach, increasingly demanded by the new government, provoked hostility among some psychologists, varied mixtures of Marxisms and psychologies among others, and an inclination in the ideological establishment to favour the reduction of mind to neural reflexes, as Pavlov and Bekhterev claimed to be doing. In 1924, at a conference devoted to the campaign for Marxism, Vygotskii's presentations struck the audience as a major event, the illumination of a way out of an impasse.

Bukharin, the Party's chief of ideology, had been commending 'reflexology', but the central Institute of Psychology had continued to set it aside as physiology, not psychology, which seeks to understand consciousness, not to explain it away. The founder of the Institute, G.I. Chelpanov, who rebuked the Party ideologists for misunderstanding both psychology and Marxism, had been replaced as director by K.N. Kornilov, who enthusiastically endorsed the Party's campaign, while continuing the kinds of research that he had learned from Chelpanov. The impending clash between the ideological establishment and the specialists in psychology was brilliantly finessed by Vygotskii. He praised Bukharin while bringing out the underlying clash between Pavlov and Marx, between the reduction of mind to conditioned reflexes and the analysis of mind as it expresses itself in 'cultural-historical' development.

He was appointed to the Institute of Psychology, where he became the most influential theorist by virtue of his astonishing breadth of knowledge, acute insight and critical appreciation for all the approaches to a contested discipline. That eclecticism, at odds with the drift in the ideological establishment towards an exclusive Marxism, was the probable reason why Vygotskii's major writings of the 1920s went unpublished. They circulated in manuscript, provoking discussion, attracting disciples in part, one supposes, by the sense of daring in his quiet dismissal of projects for a uniquely Marxist science. *Psikhologiia iskusstva (The Psychology of Art)* - awarded a doctorate in 1925, first published in 1965, heavily abridged - emphasized Marx's confession that enduring works of art did not fit within his scheme of evolving base and superstructure, and drew on 'formalist' aesthetics and on Freudian psychology, though both were disapproved of by the ideological establishment. *The Historical Meaning of the Psychological Crisis* - completed in 1927, unpublished until 1982 - used Marxism once again as an expression of the crisis more than a solution for it: each major school of psychology was founded on an obvious truth, and all together were an incoherent jumble. Marxism suggested a way to coherence only at the most abstract level of methodology, by proposing a 'cultural-historical' approach.

Condensed bits of these views were published during Vygotskii's lifetime in special studies and brief surveys, which can be confusing, for Vygotskii was conciliatory towards the ideological establishment. If it believed, for instance, that Pavlov had discovered the neural mechanism of mental activity, Vygotskii would endorse that view in some publications - such as a textbook or a celebratory survey of Soviet psychology - while elsewhere he showed that Pavlov's claim was at odds with neuroscience and with the *Gestalten* of mental activity. Such amiable segregation of viewpoints became extremely difficult after 1930, when vehement crude young men brought Stalin's 'great break' (*velikii perelom*) to higher learning as well as industry and agriculture. They required an aggressively Soviet philosophy and psychology, thoroughly hostile to 'bourgeois' or Western scholarship. Vygotskii still tried to be nice. He did not perform the odious ritual of 'self-criticism', for instance, for praising Piaget as the main source of his own work on child development. He published a translation of a major work by Piaget, prefaced by exaggerated disagreement - a common technique in the Stalin era for keeping foreign ideas before the public in spite of official disapproval.

The 'great break' brought insistence on practicality as well as exclusivity: science must prove its difference from 'bourgeois' pseudo-science by immediate service to 'the construction of socialism'. Vygotskii and his disciples responded in three ways: by emphasizing the pedagogical utility of their studies in child development, the therapeutic promise of their work in mental illness and brain damage, and most broadly the usefulness of their evolutionary vision in the 'cultural revolution', which they interpreted as a drive to modernize the primitive mentality of the masses. *Etiudy po istorii poredeniia: obez'iana, primitiv, rebënok (Studies in the History of Behavior: Ape, Primitive, Child),* a little survey that Vygotskii and Luria brought out in 1930, presented their version of 'the biogenetic law'. Ontogeny recapitulates phylogeny; the child's mental development repeats in brief the evolution from ape through primitive to civilized mentality. That was offered as an orienting vision for organizers of collective farms as well as teachers of lower-class children, but the establishment declared it an insufferable insult to the masses.

A field study of 1931 was especially offensive. The peasant mentality in Kazakhstan was found to be mired in 'situational thinking', ill prepared for universal rationality. Though unpublished, the study was angrily attacked in the press. Any effort at comparative analysis of different mentalities was implicitly disapproved, even one that might try, as Vygotskii and his colleagues did not, to take into account the shaping influence of class or status in the interaction of researcher and subjects. Group psychology was a field of intuitive expertise for master politicians and managers; they were not to be instructed by academics, not even psychologists preaching visions of a 'cultural-historical' Marxist discipline founded on 'the biogenetic law' of mental development.

Tuberculosis killed Vygotskii in 1934, while he and his school were being denounced as pseudo-Marxist imitators of 'bourgeois' or Western scholarship, along with all other schools of would-be Soviet psychology. Even in that time of the triumphant yahoo, Vygotskii had exceptional influence. V.N. Kolbanovskii, the militant communist who was installed as director of the Institute of Psychology with a mandate to 'Bolshevize' it, published a brief précis of Vygotskii's *Istoricheskii smysl psikhologicheskogo krizisa (The Historical Meaning of the Psychological Crisis)*, expressing the hope that this landmark book would soon appear. Disappointed in that hope, Kolbanovskii managed to bring out a little collection of Vygotskii's essays called *Myshlenie i rech' (Thinking and Speech)* (1934), called *Thought and Language* in the 1962 English translation.

In Russia and abroad that text shaped understanding of Vygotskii's psychology for so long that the limitations and misperceptions built into it may be incorrigible. The dying author, with some help from Kolbanovskii, fashioned the book to appease the establishment in a time of rampant Stalinism - for instance, by omitting his arguments against any claim of a uniquely Marxist science, and by including his exaggerated criticism of Piaget. Moreover, the surviving disciples shaped the dead man's legacy to fit their own agendas, avoiding a 'cultural-historical' approach as anything but a slogan for studies of child development. The recovery of Vygotskii's actual thought became a historical project, of little interest to most specialists in psychology, who prefer to work in their separate trends and ignore the rivals. Foundational debates over the opposed assumptions of diverse trends, which preoccupied the emergent discipline in the late nineteenth and early twentieth centuries, have come to be considered pointless excursions into useless philosophy. In academic practice that is what the separation of psychology from philosophy means: studied disregard for the implicit incoherence of rival trends in psychology. The transformation of Vygotskii's legacy is an especially ironic case in point.

2 Thought

Specialists in the cognitive development of children often distinguish Vygotskii's approach from Piaget's as a contrast between exogeny and endogeny, acquisition from without versus unfolding from within (see Piaget, J.; Cognitive development). That is, as most acknowledge, an oversimplified scheme, for interaction between the child and its external world, mediated by adults and other children, is essential to Piaget's approach as it is to Vygotskii's, or indeed to any analysis of action, thought and speech in their intertwined emergence, whether in phylogeny or ontogeny or both together. Piaget's landmark studies in that large field were fundamental to Vygotskii, as he repeatedly acknowledged. The correction that he offered was originally limited to the special problem of 'egocentric' speech. Very young children talk to themselves a lot, gradually ceasing by school age. Piaget interpreted this by analogy with autism; Vygotskii demurred. He argued that talking to oneself is the child's way of trying out patterns of speech - and thought and appropriate action - as learned from other people. When Piaget learned of that correction, many years later, he gratefully accepted it as 'a new hypothesis: that egocentric speech is the point of departure for the development of inner speech, which is found at a later stage of development, and that this interiorized language can serve both autistic ends and logical thinking' (1962: 7).

The important distinctiveness of Vygotskii's approach is implicit in Piaget's final phrase, his simple contrast between autism and 'logical thinking', his neglect of expressive thinking in the mind's connections with the world. Piaget came to the problem of childhood mental development from an original interest in biological science and in the standardization of intelligence tests, and wound up with a theory of evolutionary epistemology, oriented most of all towards the scientific or logical and analytical functions of thought and language. Vygotskii came to the would-be science of mind through literary studies, trying to understand aesthetic experience as a mode of cognition at odds with the scientific mode.

His first major work was a lengthy analysis of *Hamlet* that focused on the meaning of the play, as established by a cultural tradition endlessly absorbed and reshaped in the interacting minds of audiences, actors and writers, including critics as well as original authors. Set within that context Vygotskii's interpretation of *Hamlet* came down to a sense of tension between a traditional formula and a creative alteration of it. What the audience expects to see enacted on the stage - criminal outrage corrected by vengeful justice - is contradicted by what the author and actors actually present - random violence, insanity, vengeance dissolving into pointless deaths. Melodrama, a traditional formula for anxiety aroused and set to rest, is thus evoked and altered to create tragedy, an unsettling sense of life and death as a process that cannot be justified. Thus Vygotskii sets a problem for the would-be science of mind: how to explain the satisfaction that author-actors-audience jointly derive from artistic expression of a tragic meaning of life.

Vygotskii recalls the ancient explanation - catharsis - and rehearses some modern claims such as Freud's of scientific grounding for it. He criticizes Freud for a primitive understanding of 'the pleasure principle' at odds with 'the reality principle', and for other sloppinesses in mixing aesthetic understanding with claims of scientific explanation. But Vygotskii's deepest disagreement is with all contemporary schools of psychological science which rest on a dream of reducing subjective experience and cultural expression to nerve energies accumulating and discharging. Such explanatory reduction of mind to body would be irrelevant to knowledge of the mind in the sense of understanding, in this case, the meaning of *Hamlet*.

Vygotskii quotes Spinoza, confronting the fundamental anomaly as follows: 'Solely from the laws of nature considered as extended substance' one cannot 'deduce the causes of buildings, pictures, and things of that kind which are produced only by human art; nor would the human body, unless it were determined and led by the mind, be capable of building a single temple'. Since Vygotskii cannot find in modern science any way out of that dualism, *The Psychology of Art* begins and ends with Spinoza's metaphysical response: 'The objectors [to monism] cannot fix the limits of the body's power, or say what can be concluded from a consideration of its sole nature'. Intent on a monistic fusion of aesthetic understanding and psychological explanation, Vygotskii was sufficiently rigorous to acknowledge that the two types of inquiry are incoherent.

Specialists in cognitive psychology have paid scant attention to Vygotskii's emphasis on the problems that expressive thought presents for their discipline, though *Thought and Language*, the little book that most shaped their understanding of Vygotskii, comes back to such problems repeatedly. A crucial case in point is his distinction between *znachenie* and *smysl*, usually translated as 'meaning' and 'sense'. It might be less bewildering to say

logical meaning versus expressive meaning, the former acknowledging the law of identity, the latter deliberately violating it. Vygotskii offers a vivid example from a conversation that Dostoevskii overheard among six drunken men, who took delight in using a one-syllable obscene word for six different meanings, expressed by changing tone and gesture. (The fact that the word is still taboo in a respectable encyclopedia is further evidence of the difference between logical and expressive meaning.)

Vygotskii liked to recall Franz Brentano’s *Psychology from the Empirical Viewpoint* (1874), which boldly confronted the unscientific clutter of psycholog*ies*, and started two more schools (Gestalt and phenomenology) with its arguments for unification (see Brentano, F.C.; Gestalt psychology; Phenomenological movement). The gap that Brentano came to acknowledge between ’genetic’ and ’descriptive’ psychology, and the analogous distinction that several thinkers made between ’explanation’ and ’understanding’ as different modes of knowing ourselves, were major influences in Vygotskii’s brooding over psychology. So was William James’ review of 1892, ending with confidence in the future, though now ’there is no science,… only the hope of a science…. Something definite happens when to a certain brain-state a certain "sciousness" corresponds’ (see James, W.). By the 1920s, among psychologists who were unwilling to put on blinkers, to ignore the fragmentation of the would-be science and its separation from expressive understanding of the mind, James’ witty optimism had given way to gloomy talk of ’crisis’. Karl Bühler, most notably, was writing *Die Krise der Psychologie* about the same time that Vygotskii wrote *The Historical Meaning [smysl] of the Psychological Crisis*, which shares with other such ruminations the ironic fate of marking the end of an era by dreaming of a new one. Vygotskii’s dream was unique in its special attention to Marxism as well as the ’meaning [*smysl*] of the psychological crisis’ - in the sense of a crisis in ’the meaning of life’ as well as the would-be science of mind.

He disapproved of claims of a supposedly distinctive Marxist psychology, arguing that they confused three different realms of discourse: the fragmented discipline called psychology, the methodology that might transform it into a genuine science, and the most general methodology of knowledge as a whole. Marxism, in his view, was relevant only in discussion of that broadest realm. It could be transformed into concrete methodological principles for psychology only through protracted labour within the discipline, striving to overcome the war of schools. ’Our science will become Marxist to the degree that it will become true, scientific; and we will work precisely on that, its transformation into a true science, not on its agreement with Marx’s theory’.

His reading of ’Marx’s theory’ emphasized its ’cultural-historical’ vision of human development, including its insistence on a reciprocal relationship between understanding ourselves and changing ourselves, progressing thus towards authentic humanity in activity-cum-consciousness. ’Activity’ in this very long-term historical usage is so much broader than in psychological studies of a child’s ’activity’ that the word becomes a different concept. Inattention to that difference, which marked Soviet psychologists’ talk of ’activity’ after Vygotskii, deprived such talk of any deeper function than soothing the ideological establishment, which glorified ’practice’ in even looser, less meaningful ways.

When Vygotskii argued that the science of mind is inchoate, he recalled Trotsky’s observation that ’man is himself inchoate’ (*stikhiia*, elemental chaos, from Greek *stoicheia*). The socialist revolution would create a new authentic human being along with a new authentic science of psychology. When Vygotskii indulged in such talk, he was far closer than Trotsky or any other political leader of ’the Soviet experiment’ to explicit utopianism, with overtones of modernist absurdity:

> In the future society psychology will be in actuality the science of a new person [*chelovek*]. Without that perspective Marxism and the history of science would not be complete. But that science of the new person will still be psychology; we now hold in our hands the thread leading to it. Never mind that that psychology will resemble present-day psychology as little as - in Spinoza’s words - the constellation Canis resembles the dog, a barking animal.

(1982-4: vol. 1: 436)

That image - the would-be science of mind as a dog barking at the orderly splendour of the night sky - concluded Vygotskii’s *Historical Meaning of the Psychological Crisis*, projecting Spinozist faith in monism upon the incoherence of contemporary culture.

See also: Piaget, J.

DAVID JORAVSKY

List of works

Vygotskii, L.S. (1982-4) *Sobranie sochinenii (Collected Works)*, Moscow: Pedagogika, 6 vols.(The most comprehensive collection of his works. Vol. 1 contains *The Historical Meaning of the Psychological Crisis*.)

Vygotskii, L.S. (1987-3) *The Collected Works*, New York: Plenum, 3 vols, trans. and ed. R.W. Rieber, A.S. Carton, J.S. Bruner and N. Minnick.(Translation of vols 2 and 5 of *Sobranie sochinenii*, with extensive notes and commentary.)

Vygotskii, L.S. (1925) *Psikhologiia iskusstva; The Psychology of Art*, Cambridge, MA: MIT Press, introduction and commentary by A.N. Leontiev and V.V. Ivanov, 1971.(This is a translation of the Russian edition of 1968.)

Vygotskii, L.S. and Luria, A.R. (1930) *Etiudy po istorii poredeniia: obez'iana, primitiv, rebenok*, trans., ed., and introduced by V.I. Golod and J.E. Knox, *Studies in the History of Behavior: Ape, Primitive and Child*, Hillsdale, NJ: Lawrence Erlbaum, 1993.(Reproduces the 1930 survey with a substantial introduction.)

Vygotskii, L.S. (1934) *Myshlenie i rech' (Thinking and Speech)*; trans. *Thought and Language*, Cambridge, MA: MIT Press, 1962; revised and ed. by A. Kozulin, 1986.(An expanded translation, with an illuminating introduction.)

References and further reading

Iaroshevskii, M. (1989) *Lev Vygotskii*, Moscow: Progress.(Vygostkii's achievement, by a Soviet historian of psychology.)

Joravsky, D. (1989) *Russian Psychology: A Critical History*, Oxford: Blackwell.(Vygotskii and his school within a large historical context.)

Kozulin, A. (1990) *Vygotskii's Psychology: A Biography of Ideas*, Cambridge, MA: Harvard University Press. (Vygotskii's development, by a thoughtful, well-informed psychologist.)

Piaget, J. (1962) 'Comments', in Vygotskii, *Thought and Language*, Cambridge, MA: MIT Press, 1962.(Referred to in §2.)

Valsiner, J. (1988) *Developmental Psychology in the Soviet Union*, Bloomington, IN: Indiana University Press. (The history of the Vygotskian school.)

van der Veer, R. and Valsiner, J. (1991) *Understanding Vygotskii: A Quest for Synthesis*, Oxford: Blackwell. (The richest biography, using archives and published matter; with the most extensive bibliography of Vygotskii's writings and unpublished papers.)

Wertsch, J.V. (1985) *Vygotskii and the Social Formation of Mind*, Cambridge, MA: Harvard University Press. (The Vygotskian legacy in developmental psychology as perceived by an American admirer.)

Vysheslavtsev, Boris Petrovich (1877-1954)

Boris Petrovich Vysheslavtsev, Russian idealist philosopher and religious thinker, was exiled from his homeland in 1922 because of his anti-Marxism (which he later elaborated in a full-fledged philosophical critique). In western Europe he became a leading figure in the Russian émigré philosophical community, lecturing and writing on questions of metaphysics, ethics, philosophical psychology and social philosophy. Vysheslavtsev was particularly noted for his study (begun in an early work on the ethics of Fichte) of the irrational as the sphere of human contact with the Absolute. Subsequently he developed this theme through the application of concepts of depth psychology to ethics and to the interpretation of Christian doctrine.

The son of a Moscow lawyer, Vysheslavtsev matriculated in 1895 in the Law Faculty of the University of Moscow, where his mentor was the eminent legal philosopher Pavel Novgorodtsev. After graduating in 1889 he worked for some years as an attorney before beginning postgraduate studies at the university. In 1908 he was sent abroad for two years to prepare himself for a professorship; he worked principally at Marburg, under the Neo-Kantians Hermann Cohen and Paul Natorp. Vysheslavtsev was awarded the doctorate by the University of Moscow and appointed to a position as docent there in 1914; three years later he was named professor in the Law Faculty. After his expulsion from Russia he went first to Berlin and in 1924 settled in Paris, where he became a professor at the Orthodox Theological Institute. During the Second World War he moved to Switzerland. He died in Geneva in October 1954.

Vysheslavtsev’s first major philosophical work was his doctoral dissertation, *Ėtika Fikhte (Fichte’s Ethics)* (1914). Although he was initially an orthodox proponent of Fichte’s transcendental idealism, in this book Vysheslavtsev presented an original interpretation of Fichte by concentrating on the role of the irrational and the concept of ‘the Absolute’ in the German thinker’s outlook. Arguing that philosophy is confronted with a basic antinomy between what he calls ‘system’ (rationality) and ‘infinity’ (irrationality), Vysheslavtsev affirmed that beyond every rational concept there is ‘the mysterious limitless’ - the infinity or Absolute that cannot be dealt with conceptually but is accessible only to intuition. Vysheslavtsev attempted to describe this irrational infinity by the use of concepts in the modern philosophy of mathematics, such as Cantor’s distinction between potential and actual infinities (see Cantor §1-2). But he concluded that no form of discursive, conceptual thinking is adequate to the task; although philosophy can show us the necessity of acknowledging the Absolute, philosophy cannot analyse it as Hegel sought to do in his metaphysics of Absolute Spirit.

In his efforts to explore this elusive realm, Vysheslavtsev in later works turned towards the study of religion, believing that the quest for the Absolute is the core of the religious attitude. In *Serdtse v khristianskoi i indiiskoi mistike (The Heart in Christian and Indian Mysticism)* (1929), he examined the biblical notion of ‘the heart’ as a fundamental non-rational human capacity; he found that ‘the heart’, in addition to being a source of mystical knowledge distinct from the knowledge gained by the intellect, is the fundamental ontological principle at the root of the human self as a loving and free being. He noted that both Christian and Hindu mystics affirm such an ontological principle, although they differ in their understanding of its relation to human individuality.

In his most important philosophical work, *Ėtika preobrazhennogo Ėrosa (The Ethics of Transfigured Eros)* (1931), Vysheslavtsev employed concepts of depth psychology to construct an ethical theory grounded in the irrational nature of human beings and their relation to the Absolute. Ethical ‘laws’ or imperatives as traditionally understood are ineffectual, Vysheslavtsev believed, because they are merely addressed to consciousness and rationality and do not engage the unconscious depths of instinctual human life; indeed, drawing on the work of the French psychologists Émile Coué and Charles Baudouin, he argued that such imperatives generate subconscious resistance in accordance with what he calls ‘the law of irrational opposition’. Christian ethics, in Vysheslavtsev’s view, pursues the correct approach to morality by presenting not legalistic commands but beautiful images of moral virtue in the persons of Christ and the saints; these images engender deep-seated responses of admiration and love, thereby transfiguring the irrational instincts and impulses (‘Eros’) that resist conscious control. In the spirit of Freud and Jung (he knew their works well), Vysheslavtsev called this moral transformation of basic instincts ‘sublimation’, and he saw in it the key to the understanding of the Christian moral life.

Sublimation, for Vysheslavtsev, is not a purely deterministic process subject to psychologistic explanation. It occurs only if the individual will has freely chosen the path of elevating and refining Eros and has enlisted the imagination in creating the needed images of moral virtue. Sublimation takes place, furthermore, only with the aid of divine grace; the individual spirit must be inspired from on high, by its vision of the Absolute as the fullness and perfection of being. This orientation towards the Absolute makes sublimation possible, but the divine grace that flows from it must still be freely accepted in order to be morally significant in the life of the individual. Thus in Vysheslavtsev's ethics of sublimation, both free will and grace are necessary but not sufficient conditions for the moral life. In their interaction free will is itself sublimated, being raised from its primitive state of arbitrariness to the condition of moral freedom.

The two books published by Vysheslavtsev in the last years of his life reflect the heightened interest in social issues that marked his thinking from the late 1930s, when he was also active in the ecumenical movement. *Krizis industrial'noi kul'tury (The Crisis of Industrial Culture)* (1953) is a work of social philosophy in which he rejects Marxism (and socialism generally) as a socioeconomic system and elaborates an alternative he calls 'neo-liberalism', which stresses the humanization of industrial culture through the widespread application of economic democracy. In *Filosofskaia nishcheta marksizma (The Philosophical Poverty of Marxism)* (written in the mid-1940s but not published until 1952), he presented a searching and philosophically sophisticated critique of the self-styled 'dialectical and historical materialism' then prevailing in the Soviet Union.

JAMES P. SCANLAN

List of works

Vysheslavtsev, B.P. (1914) *Ėtika Fikhte: osnovy prava i nravstvennosti v sisteme transtsendental'noi filosofii (Fichte's Ethics: The Foundations of Law and Morality in the System of Transcendental Philosophy)*, Moscow: A. Snigirov.

Vysheslavtsev, B.P. (1929) *Serdtse v khristianskoi i indiiskoi mistike (The Heart in Christian and Indian Mysticism)*, Paris: YMCA-Press.

Vysheslavtsev, B.P. (1931) *Ėtika preobrazhennogo Ėrosa: problemy zakona i blagodati (The Ethics of Transfigured Eros: Problems of Law and Grace)*, Paris: YMCA-Press; repr. together with *Vechnoe v russkoi filosofii* (see below) in a single volume, with an introduction and notes by V.V. Sapov, Moscow: Respublika, 1994.

Vysheslavtsev, B.P. (1952) *Filosofskaia nishcheta marksizma (The Philosophical Poverty of Marxism)*, Frankfurt am Main: Posev.(This first edition was published under the pseudonym 'B. Petrov'; a second edition was published under Vysheslavtsev's own name in 1957.)

Vysheslavtsev, B.P. (1953) *Krizis industrial'noi kul'tury: marksizm, neosotsializm, neoliberalizm (The Crisis of Industrial Culture: Marxism, Neo-Socialism, Neo-Liberalism)*, New York, NY: Chekhov Press.

Vysheslavtsev, B.P. (1955) *Vechnoe v russkoi filosofii (The Permanent in Russian Philosophy)*, New York: Chekhov Press.(A posthumous collection of essays; see 1931 title above.)

References and further reading

Kline, G.L. (1955) 'A Philosophical Critique of Soviet Marxism', in *The Review of Metaphysics* 9 (1): 90-105. (Formally a review of Vysheslavtsev's 1952 book on Marxism, this article contains much information about Vysheslavtsev's thought.)

Zenkovsky, V.V. (1948-50) *Istoriia russkoi filosofii*, Paris: YMCA-Press, 2 vols; 2nd edn 1989; trans. G.L. Kline, *A History of Russian Philosophy*, 2 vols, London: Routledge & Kegan Paul and New York: Columbia University, 1953, vol. 2, 814-19.(Summary of Vysheslavtsev's thought by the authoritative historian of Russian philosophy.)

Zenkovsky, V.V. (1955) 'B.P. Vysheslavtsev, kak filosof' (B.P. Vysheslavtsev as a Philosopher), in *Novyi zhurnal* 15: 249-61.(A more detailed treatment of Vysheslavtsev's work.)

Wallace, Alfred Russel (1823-1913)

Co-discoverer with Charles Darwin of the theory of natural selection, Wallace travelled to the Amazon in 1848. Four years of collecting specimens there for sale in Europe revealed patterns of geographical distribution among animals. Unfortunately, much of his South American collection was lost in a fire at sea during the voyage home, which forced him to begin his collecting anew. This led to eight more years of travel (1854-62), this time in the Malay Archipelago, where he made his own momentous discovery of the theory of natural selection in 1858. An exceptionally clear thinker, he made many valuable contributions to evolutionary thought.

Wallace was a self-taught naturalist, as was his friend H.W. Bates, who accompanied him to South America. They read many of the books that Darwin read, thus sharing to some degree the same intellectual climate despite their plebeian background. Indeed, both Wallace and Darwin read T.R. Malthus on population, and both cited him as a major factor in their independent discoveries of the agency of natural selection. Wallace also read Darwin and C. Lyell, among others, Lyell being of special interest because Wallace cast him in the role of chief opponent as he developed his own thoughts during his travels in the Malay Archipelago, as recorded in his (unpublished) ‘Notebook’ (see Geology, philosophy of §1).

It was broadly believed in the 1850s that species were permanent, while varieties were not; rather, they were thought to be produced by ordinary generation and to vary only within strict limits. Species were somehow ’created’ in one or more centres of ‘special creation’, a term widely used, even by Wallace and Darwin, meaning to a zoologist, as Robert Owen remarked, ’a process he knows not what’. Wallace maintained that ’ every species has come into existence coincident both in time and space with a pre-existing closely allied species’ (1855). In his ’Note on the Theory of Permanent and Geographical Varieties’ (1858a), he asked, ’ why should a special act of creation be required to call into existence an organism differing only in degree from another which has been produced by existing laws?’. And, finally, there was the famous manuscript sent to Darwin in the spring of 1858. These three papers formed the foundation of Wallace’s theory of evolution by means of natural selection, his explanation of ’creation’, with natural selection as the agent of change. H. Spencer later suggested ’survival of the fittest’ as an alternative term for ’natural selection’, an idea accepted by both Wallace and Darwin.

Wallace’s manuscript reached Darwin on 18 June 1858. In an effort to protect Darwin’s priority, Lyell and J.D. Hooker worked out the immediate reading of Wallace’s paper before the Linnaean Society of London on 1 July, preceded by two contributions from Darwin. These three taken together are known as the ’joint papers’. Neither author was present. Darwin always maintained that their two theories were exactly alike, when in fact they differed markedly in detail. For example, Darwin relied heavily on data from domestic animals, while Wallace based his theory largely on data from wild animals. Only the concept was the same. Furthermore, over the years various errors regarding this event have crept into the record. Wallace and Darwin continued their fruitful correspondence, initiated by Wallace in October 1856, until just a year before Darwin’s death in 1882, as often as not disagreeing with one another in lively argument.

Attracted to the subject of geographical distribution during his travels in South America, Wallace became its leading authority, his work culminating in his *Geographical Distribution of Animals* in 1876: ’ The present work is an attempt to collect and summarize the existing information on the Distribution of Land Animals; and to explain the more remarkable and interesting of the facts, by means of established laws of physical and organic change’. This was supplemented four years later by *Island Life*. Basic to this development was the recognition of zoological regions, Wallace supporting the scheme proposed by P.L. Sclater in 1858. The sharp division between the Asian and Australian faunas found on either side of the strait separating the Indonesian islands of Bali and Lombok was named ’Wallace’s Line’ in his honour by T.H. Huxley.

When it came to human beings, Wallace partially abandoned natural selection. In 1864, in the first of several papers, he tried to show how humans could have been affected by natural selection, first in regard to physical form and then to the brain (also subject to natural selection) which took over control once sufficiently advanced, while the body remained essentially unchanged. Humans could easily make a warmer coat if necessary, once the mind had been developed, while an animal would have to go through the slow process of growing one, thus changing physically under pressure of natural selection. In *Darwinism* (1889), Wallace recognized a number of human

faculties - mathematical, musical, and so on, as well as knowledge of such fields as gravitation, chemical force, electricity - which he felt could not have been the results of natural selection but to belong rather to some spiritual world. This explains in part his interest in spiritualism, then a very popular but non-scientific pastime. In fact, a century later questions he raised in that connection have yet to be fully resolved. He disagreed with Darwin on the importance of sexual selection as the principal agent in human evolution, holding that sexual selection was merely a form of natural selection.

See also: Evolution, theory of; Huxley, T.H.; Species; Taxonomy

BARBARA G. BEDDALL

List of works

Wallace, A.R. (1855) ‘On the Law Which has Regulated the Introduction of New Species’,*Annals and Magazine of Natural History* 16: 184-96.(Darwin asked, ’but why does his law hold good?’.)

Wallace, A.R. (1858a) ‘Note on the Theory of Permanent and Geographical Varieties’, *Zoologist* 16: 5,887-8.(An important step in the development of his theory.)

Wallace, A.R. (1858b) ’On the Tendency of Varieties to Depart Indefinitely from the Original Type’, *Journal of the Linnaean Society of London (Zoology)* 3: 53-62.(Wallace’s part of the ’joint papers’.)

Wallace, A.R. (1858c) ‘Charles Darwin and Alfred Russel Wallace, On the Tendency of Species to Form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection’, *Journal of the Linnean Society of London (Zoology)* 3: 45-62.(This is the collective title of the ’joint papers’.)

Wallace, A.R. (1864) ‘The Origin of Human Races and the Antiquity of Man Deduced from the Theory of "Natural Selection"’, *Journal of the Anthropological Society of London* 2: clviii-clxx.(A caveat on natural selection.)

Wallace, A.R. (1876) *The Geographical Distribution of Animals: With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth’s Surface*, London: Macmillan, 2 vols.(A typical Wallace book, filled with facts and insights.)

Wallace, A.R. (1880) *Island Life*, London: Macmillan.(A further elaboration of Wallace’s contribution to geographical distribution.)

Wallace, A.R. (1889) *Darwinism: An Exposition of the Theory of Natural Selection, with Some of Its Applications*, London: Macmillan.(A comprehensive and well-organized survey.)

References and further reading

Beddall, B.G. (ed.) (1969) *Wallace and Bates in the Tropics: An Introduction to the Theory of Natural Selection*, London: Macmillan.(Aimed at the nonprofessional.)

Beddall, B.G. (1988) ‘Darwin and Divergence: The Wallace Connection’, *Journal of the History of Biology* 21: 1-68.(Continues an examination, begun in 1968, of circumstances surrounding the discovery of the agency of natural selection.)

George, W. (1964) *Biologist Philosopher: A Study of the Life and Writings of Alfred Russel Wallace*, London: Abaelard-Schuman.(Contains a lengthy list of references.)

Marchant, J. (1916) *Alfred Russel Wallace: Letters and Reminiscences*, New York and London: Harper.(The Wallace-Darwin correspondence, as well as Wallace’s correspondence with others, and an extensive bibliography of his works.)

McKinney, H. (1980) *Wallace and Natural Selection*, New Haven, CT: Yale University Press.(Interesting documents included, as well as a lengthy bibliography.)

Wang Chong (AD 27-*c*.90)

The Han philosopher Wang Chong wrote a text called Lunheng (Disquisitions or Discourses Weighed in the Balance), one of the most exceptional and original documents in Chinese thought, compiled as it was during a crucial transitional period. Wang's main approach can be defined as a rational scepticism, questioning accepted history, the contents of canonical texts and philosophical claims about reality. His writings on fate develop ideas such as necessity, cause and uncertainty.

Wang Chong was born in Shangyu in the district of Kuaiji. His work, the *Lunheng (Disquisitions or Discourses Weighed in the Balance)*, composed of eighty-five chapters and more than two hundred thousand characters. Three themes pervade the text: a rational discussion on the nature of the classics (*jing*) (see Chinese Classics), his rational approach to nature and his original conception of fate (*ming*).

Wang Chong's intellectual context was the complex debate between the so-called New Text and Old Text schools, a discussion concerning the authenticity and interpretation of certain early texts. An example of his approach may be seen in Chapter 45 (On the Rain Sacrifice) of the *Lunheng*. Here, Wang asks on what do the texts base their affirmation that the rain sacrifice is necessary. In another example, in Chapter 23 (On Thunder), he asks if the destructive properties of thunderstorms actually reflect the 'anger of heaven'. He responds to the question by concluding that thunder is fire.

In questioning the ideal type of sage, the authority of the canonical texts and the practice of divination, Wang Chong adheres to the Old Text tradition, which emphasizes the creative nature of interpretation rather than a narrow scholasticism. In Chapter 14 (The Nature of Things), he states that Heaven did not create man on purpose. He makes it clear that there is no link between *ming* and virtue. One example he gives cites the careers of the favourite disciples of Confucius, Ran Boniu and Yan Hui, who were virtuous and scholarly but nevertheless died young. He believed in the injustice of Heaven, and gives many examples of the incapacity of Heaven to punish the guilty.

Wang's concept of fate is complex because it is linked with divination, cosmology and politics. He proposed three kinds of fate: natural destiny (*zhengming*), concomitant destiny (*suiming*) and adverse destiny (*zaoming*). The first refers to luck that results from original physical conditions; the second results from a combination of unpredictable coincidences; and the third signifies events that are contrary to expectations and opens up the possibility of contingency and the random factor. These approaches can be found in Chapter 3 of the *Lunheng*. This seems to be the most crucial and interesting chapter, especially as it concerns the ideas of necessity, cause and uncertainty.

The *Lunheng* is a text which belongs to the critical sources of Chinese culture, and it demonstrates that such sources are able to bring up original ideas and a new point of view concerning the Chinese world. Because of this text, we are aware that the Chinese world has many contradictory voices. There is not a static consensus in this debate, but a lively interplay of views among Chinese thinkers.

See also: Chinese Classics; Tian

AGNES CHALIER

List of works

Wang Chong (AD 27-*c*.90) *Lunheng (Disquisitions or Discourses Weighed in the Balance)*, ed. Huang Hui, 1938; repr. Taiwan: Shangwu yinshuguan, 1983; trans. A. Forke, *Lun-heng: Philosophical Essays of Wang Ch'ung*, Berlin, 1907; repr. New York: Paragon, 1962, 2 vols.(The standard translation of most but not all of the text.)

References and further reading

Loewe, M. (1974) *Crisis and Conflict in Han 104 BC-9 AD*, London: George Allen & Unwin.(Concerns the social and historical conflict in China prior to Wang Chong's time.)

Loewe, M. (1982) *Chinese Ideas of Life and Death*, London: George Allen & Unwin.(An account of the mainstream of Chinese thought and its internal contradictions during the Han period.)

Pokora, T. (1957) 'How Many Works Were Written by Wang Ch'ung?', *Archiv Orientalni* 25.(An examination of the writings of Wang Chong.)

Pokora, T. (1962) 'The Necessity of More Thorough Study of the Philosopher Wang Chong and His Predecessors', *Archiv Orientalni* 30.(An invitation to read classic Chinese texts and Wang Chong in particular, so as to better understand Chinese antiquity.)

Wang Fuzhi (1619-92)

A seventeenth-century neo-Confucian and Ming loyalist, Wang Fuzhi is best known for his nationalism and his theories of historical and metaphysical change. His classical commentaries and other writings, not published until the nineteenth and twentieth centuries, present an exceptionally comprehensive and vigorously argued synthesis and critique of China's intellectual tradition. His ideas on topics such as politics, cosmology and knowledge have fascinated readers of widely differing philosophic persuasions.

Having but recently passed his first civil service examination when the Manchus invaded the Ming capital of Beijing in 1644, Wang Fuzhi (Wang Chuanshan) defied the new Qing dynasty, at first militarily and politically, then with his pen through more than forty years of relative reclusion. The secrecy of his writings accounts for their preservation until their publication in the nineteenth and twentieth centuries. It also explains his overdue recognition as one of China's best thinkers. The sense of political crisis underlying his writing attracts the interest of Chinese Marxists and non-Marxists alike.

More philosophically significant in this regard is Wang's cosmology and metaphysics, traced by him to the Song philosopher Zhang Zai but also historically connected to late Ming trends. Wang's work is based on the dynamic concept of *qi* (material force, literally breath) (see Qi). His practical emphasis on *qi* as opposed to *li* (principle) lends itself to a materialist interpretation particularly when linked with Wang's claim that the world consists of concrete things (see Li). However, the importance of *qi* is that it is self-moving and not predetermined from without, yet so constituted that its activity necessarily tends to produce order. This, together with mentalistic descriptions of *qi*'s activity, suggests an expressive cosmology akin to Wang's view of poetic expression. His mode of thought is organicist, which is partly why it is difficult to classify in idealist-materialist terms.

Wang's epistemology and ethical psychology are shaped by an awareness of non-Confucian systems. He opposes the extreme scepticism of the Daoists and Buddhists, but a dash of Chinese empiricism, together with his belief in the ultimate unfathomability of the non-human cosmos, makes him sharply critical of tidy, predictive systems such as those offered by numerology or Western science. Mind and senses work together to produce knowledge, but the mind comes into its own in the development of a 'synthetic a priori' type of moral knowledge through a mixture of intuition and reflection (also represented as the unfolding of *qi*). This kind of knowledge is based on moral experience and leads to moral practice (see Knowledge, concept of). Supporting Wang's ethical epistemology is an original and dynamic view of human nature as a perpetually renewed source and product of moral cultivation (see Xing).

Wang's organic conception of *qi* lends originality to his philosophy of history. In place of the common Confucian nostalgia for the 'feudal' past, Wang credits the historical process with its own natural trends and rationale (*li*). He even sees the possibility of other civilizations replacing that of the Chinese. However, this theoretic detachment does not extend to tolerance of foreign conquest in his own time. Wang Fuzhi's nationalism comes closer to racism than does normal Confucian ethnocentrism. Though passionately felt, it stands in uneasy relation to the philosophy which it ironically helped to inspire.

See also: Neo-Confucian philosophy; Qi

ALISON H. BLACK

List of works

Wang Fuzhi (1619-92) *Chuanshan quanshu (Complete Works of Wang Fuzhi)*, Changsha: Yuelu shushe, 1988-, 16 vols.(The most comprehensive edition of Wang's writings, including all extant titles. Vol. 16, consisting of miscellaneous supporting materials, has so far not been published.)

Wang Fuzhi (1619-92) *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(Based on earlier, less complete editions, the most important of which is that edited by Zeng Guofan in 1865. Various facsimile reprints of the 1933 edition are available under title *Chuanshan quanji (Complete Works of Wang Fuzhi)* or *Chuanshan yishu quanji (Complete Surviving Works of Wang Fuzhi)*.)

Wang Fuzhi (begun 1655) *Zhouyi waizhuan (Outer Commentary on the Book of Changes)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(Along with *Zhangzi Zhengmeng zhu*,

provides the metaphysical and cosmological foundation for Wang Fuzhi's philosophy of change, including criticism of Daoist and Buddhist ideas.)

Wang Fuzhi (1656) *Huangshu (Yellow Book)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(A short but important text for Wang's nationalistically oriented political thought.)

Wang Fuzhi (1655-65) *Shangshu yinyi (Elaboration on the Meaning of the Book of Documents)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(Wide-ranging commentary on the Confucian classic, embracing metaphysics, ethics and epistemology.)

Wang Fuzhi (*c.*1660?) *Shiyi (Interpretation of the Odes)*, trans. as *Jiangzhai shihua* in Wong Siu-kit, *Notes on Poetry from the Ginger Studio*, Hong Kong, 1987.(Constitutes an accessible source for Wang's theory of poetry.)

Wang Fuzhi (after 1660-5) *Shi guangzhan (Extended Commentary on the Odes)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(Important for Wang's theories of mind and human nature as well as of poetry.)

Wang Fuzhi (revised 1655) *Du Sishu Daquan shuo (Revised Commentary on Reading the Compendium of the Four Books)* (revised 1665), in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(An important critique of the neo-Confucian handling of the four books at the heart of its canon, this text is essential for Wang's ethical psychology.)

Wang Fuzhi (after 1678) *Siwen lu (Record of Thoughts and Questions)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(A series of significant 'jottings' on miscellaneous topics such as Chinese and Western cosmology.)

Wang Fuzhi (after 1678) *Zhangzi Zhengmeng zhu (Zhang Zai's Correction of Youthful Ignorance Annotated)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(In the form of a commentary on Zhang Zai, this is the mature expression of Wang's *qi*-based metaphysics.)

Wang Fuzhi (before 1682?) *Zhuangzi jie (Zhuangzi explained)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(Reckoned one of the best commentaries on the Daoist philosopher.)

Wang Fuzhi (1682) *Emeng (Nightmare)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(A late reiteration of Wang's political nationalism.)

Wang Fuzhi (1685) *Zhouyi neizhuan (Inner Commentary on the Book of Changes)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(More strictly focussed and esoteric than the *Waizhuan*.)

Wang Fuzhi (1687) *Du Tongjian lun (On reading the Comprehensive Mirror)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(An examination of Chinese imperial history as treated by the Song historian Sima Guang, this is an essential source for Wang's philosophy of history, state and society.)

Wang Fuzhi (1690) *Xitang Yongri xulun (Jottings from Summer Days in the Western Chamber)*, trans. as *Jiangzhai shihua* in Wong Siu-kit, *Notes on Poetry from the Ginger Studio*, Hong Kong, 1987.(Constitutes an accessible source for Wang's theory of poetry.)

Wang Fuzhi (1691) *Song lun (On the Song dynasty)*, in *Chuanshan yishu (Surviving Works of Wang Fuzhi)*, Shanghai: Taipingyang, 1933.(A continuation of Wang's efforts to draw political lessons from history. An annotated translation of the first chapter can be found in A.H. Black, 'The Theme of the Mandate of Heaven in the Political Thought of Wang Fu-chih (1619-92)', M.Litt. thesis, University of Glasgow, 1970.)

References and further reading

Black, A.H. (1989) *Man and Nature in the Philosophical Thought of Wang Fu-chih*, Seattle, WA: University of Washington Press.(Develops expressionist interpretation of Wang's philosophy-cosmology, epistemology and theory of poetry - against Western creationist background.)

Gernet, J. (1978-90) 'Cours sur Wang Fuzhi' (Lectures on Wang Fuzhi), in *Annuaire du Collège de France*, Paris.(Studies by an eminent Sinologist with a particular interest in Wang's view of history.)

Jullien, F. (1989) *Procès ou Création: Une introduction à la pensée des lettrés chinois: Essai de problématique interculturelle (Process or Creation: An Introduction to Chinese Literati Thought: Essay on an Intercultural Problematic)*, Paris: Éditions du Seuil.(Comparative study of Wang as a process philosopher in contrast to Western creationism. Good presentation of the relevant ideas and terminology.)

McMorran, I. (1992) *The Passionate Realist: An Introduction to the Life and Political Thought of Wang Fuzhi*, Hong Kong: Sunshine Book Company.(Substantially based on his doctoral thesis, McMorran's study forms an excellent introduction to the subject and includes a useful bibliography.)

Vierheller, E.J. (1968) *Nation und Elite im Denken von Wang Fu-chih (1619-92) (Nation and Elitism in the*

Philosophy of Wang Fuzhi), Hamburg: Gesellschaft für Natur- und Völkerkunde Ostasiens.(A critical interpretation of Wang's nationalism and its motivation; reviewed at length by J.-F. Billeter in *T'oung Pao* 57, 1970.)

Yan Shoucheng (1994) 'Coherence and Contradiction in the World View of Wang Fuzhi', Ph.D. dissertation, University of Indiana.(On how to reconcile apparent contradictions in Wang's philosophy. Through detailed discussion of the main terms and concepts, the author argues effectively for Wang's role as the greatest synthesizer of Chinese intellectual tradition since Zhu Xi. Good bibliography.)

Wang Yangming (1472-1529)

Wang Yangming was an influential Confucian thinker in sixteenth-century China who, like other Confucian thinkers, emphasized social and political responsibilities and regarded cultivation of the self as the basis for fulfilling such responsibilities. While sometimes drawing on ideas and metaphors from Daoism and Chan Buddhism, he criticized these schools for their neglect of family ties and social relations. And, in opposition to a version of Confucianism which emphasized learning, he advocated directly attending to the mind in the process of self-cultivation.

Having gone through a series of political ups and downs, Wang Yangming was acutely aware of the political degeneration and moral decay of his times. The influential teachings of Zhu Xi no longer offered a solution to such problems, the study of classics emphasized by Zhu having degenerated into a tool for social advancement, resulting in habits of memorization and fragmentary studies. Wang, who in his youth studied Zhu's teachings, developed a new version of Confucian thought, the main ideas of which are contained in the *Quanxilu (Instructions for Practical Living)* a collection of conversations and letters compiled by Wang's disciples, as well as in other essays and poems.

According to Zhu Xi, the mind already knows principle (or pattern); in relation to human lives, principle constitutes the realm of the ethical (see Li). In the original state, the mind's knowledge of principle guides and is sufficient for ethical action, but such knowledge can be obscured by distortive factors of the mind often referred to as selfish desires. 'Investigation of things' and 'extension of knowledge', two steps in a self-cultivation process described in a Confucian classic, are interpreted to mean arriving at the principle in things (via the study of classics and examination of daily affairs) so as to expand one's knowledge of principle, thereby regaining the knowledge one originally has (see Self-cultivation in Chinese philosophy).

While Wang also believed that everyone has 'innate knowledge' which can be obscured by selfish desires, he regarded innate knowledge not as prior knowledge of principle but as a disposition to respond to particular situations in appropriate ways without guidance from such prior knowledge. Principle is revealed in the responses of the mind when unobscured, and 'mind is principle' in the sense that principle does not exist independently of such responses. Furthermore, while such responses may involve some kind of knowing, there is a 'unity of knowledge and action' in the sense that the responses also involve motivations to act which are not explained by the knowing and which are sufficient for action (see Action).

Given his view that principle resides just in the responses of the mind when unobscured, Wang regarded Zhu Xi's emphasis on learning as misguided. He interpreted 'investigation of things' and 'extension of knowledge' to mean correcting what is incorrect in the activity of the mind so as to allow innate knowledge to reach out. That is, a main task of self-cultivation is constantly to watch for and eliminate selfish desires, until the mind is purified and selfish desires no longer arise. While de-emphasizing learning, he did not totally reject it since studying the classics can help one understand one's mind, and since there are details such as institutional arrangements that innate knowledge may tell one to learn.

Other interesting ideas of Wang's include his conception of the myriad things as forming 'one body'; that different things are linked is seen from the way plants and animals nourish human beings or the way herbs cure disease. Thus, ideally, one should be sensitive to the well-being of all things. Another idea is his view that things do not exist outside of the mind, an idea sometimes interpreted as a form of idealism (see Idealism). It is possible, though, that 'things' in this context refers to what forms a meaningful structure, so that what is at issue is not an ontological thesis but an observation about how the mind's awareness contributes to objects forming a meaningful structure for humans (see Awareness in Indian thought).

Another well-known idea of Wang's is his 'Four Sentence Teaching', which describes (1) what is without good and evil as the substance of the mind, (2) what is with good and evil as the movement of intent (the starting point of the mind's activity), (3) what knows good and evil as innate knowledge, and (4) what does good and eliminates evil as the investigation of things. Since principle, and correspondingly the distinction between good and evil, resides ultimately in the responses of innate knowledge ('mind is principle'), there is a sense in which innate

knowledge, which Wang identified with the substance of the mind, cannot itself be described as good or evil. Furthermore, innate knowledge is without good and evil in the further sense that, when responding to situations, it is not guided by thoughts of doing good or avoiding evil (‘unity of knowledge and action’). As for the second sentence, it describes how, when the mind is activated, intent arises and can point in a good or evil direction because of the distortive effects of selfish desires. The third sentence concerns how, despite such distortive effects, innate knowledge can still distinguish between good and evil, while the fourth sentence explains the investigation of things in terms of doing good and eliminating selfish desires so as to restore the original state of the mind. On this interpretation of the teaching, self-cultivation is a gradual process of persistently doing good and eliminating evil.

However, if it is innate knowledge that enables one to recognize selfish desires as problematic, then, since innate knowledge operates without guidance from a conception of good and evil, it seems that the elimination of selfish desires should also be a spontaneous process not requiring effort. That is, as soon as innate knowledge is brought to bear on things, all incorrect elements get corrected and one achieves sagehood without any need for persistent efforts to do good and eliminate evil. These two views represent a disagreement among disciples which Wang addressed by proposing that the two views capture two teachings, directed respectively to the ordinary people and to those of sharp intelligence. Subsequently, such disagreement led to divergent developments of Wang’s teachings by later followers.

See also: Confucian philosophy, Chinese; Legalist philosophy, Chinese; Neo-Confucian Philosophy; Self-Cultivation in Chinese philosophy; Zhu Xi

SHUN KWONG-LOI

List of works

Wang Yangming (1472-1529) *Quanxilu (Instructions for Practical Living)*, trans. Chan Wing-tsit, *Instructions for Practical Living and Other Neo-Confucian Writings by Wang Yang-ming*, New York: Columbia University Press, 1963.(Translation of main work containing Wang’s philosophical ideas, with introduction by translator giving summary of Wang’s biography and philosophy.)

Wang Yangming (1503-27) Philosophical Letters, ed. and trans. J. Ching, *The Philosophical Letters of Wang Yang-ming*, Columbia, SC: University of South Carolina Press, 1972.(Translation of Wang’s philosophical letters between 1503 and 1527.)

Wang Yangming (1472-1529) Essays and Poems, trans. J. Ching in *To Acquire Wisdom: The Way of Wang Yang-ming*, New York: Columbia University Press.(Translation of selected essays and poems by Wang, with introduction by translator giving summary of Wang’s philosophy and analysis of key terms.)

References and further reading

Chang, C. (1962) *Wang Yang-ming: Idealist Philosopher of Sixteenth-Century China*, New York: St John’s University Press.(An introduction to Wang’s life and philosophy.)

Ching, J. (1976) *To Acquire Wisdom: The Way of Wang Yang-ming*, New York: Columbia University Press. (Translation of selected essays and poems by Wang, with introduction by translator giving summary of Wang’s philosophy and analysis of key terms.)

Cua, A.S. (1982) *The Unity of Knowledge and Action: A Study in Wang Yang-ming’s Moral Psychology*, Honolulu, HI: University Press of Hawaii.(Explication of Wang’s doctrine of the unity of knowledge and action in light of modern ethical theories, and reconstruction of Wang’s moral psychology.)

Tu Wei-ming (1976) *Neo-Confucian Thought in Action: Wang Yang-ming’s Youth (1472-1509)*, Los Angeles, CA: University of California Press.(Discussion of the formative years of Wang that led to his advocacy of the unity of knowledge and action.)

War and peace, philosophy of

The war/peace dichotomy is a recurrent one in human thought and the range of experience it interprets is vast. Images of war and peace permeate religion, literature and art. Wars, battles, pacts and covenants appear as outcomes and antecedents in historical narratives. Recurrent patterns of warlike and pacific behaviour invite scientific explanations in terms of underlying biological, psychological or economic processes. War and peace are also often matters of practical concern, predicaments or opportunities that call for individual or collective action. While philosophers have explored all these ways of looking at war and peace, they have paid most attention to the practical aspects of the subject, making it part of moral and political philosophy.

Practical concern with war and peace can go in either of two main directions, one focusing on war and the other on peace. Those who doubt that war can be abolished naturally worry about how it can be regulated. So long as war is possible, there will be principles for waging it. Whether such principles should limit war-making to ends like self-defence or leave the choice to the discretion of political and military leaders is a matter of continuing dispute. Nor is there agreement regarding restrictions on the conduct of war, some holding that belligerents need only avoid disproportionate damage, others that it is morally wrong to harm innocents (for example, noncombatants). In situations of emergency both limits may give way, and moralists have debated whether this relaxation of standards is defensible. Disputes over the principles governing war raise difficult questions about action, intention and the character of morality itself.

If we think that wars can be prevented, it becomes important to focus on the conditions of permanent peace. Some who do this conclude that peace depends on the conversion of individuals to an ethic of nonviolence, others that it requires strengthening the rule of law. According to a powerful version of the latter argument, the absence of law creates a condition in which persons and communities are at liberty to invade one another: a condition that, in Hobbes's classic metaphor, is a 'state of nature' which is also a perpetual state of war. While treaties of peace may terminate particular wars, only political institutions that establish the rule of law within and between communities can provide security and guarantee peace.

1 War and peace as ideas

War can be conceived in many ways. For some it is a metaphor for all that is violent, destructive and lawless. The realm outside the world of Islam, for example, is thought of by Muslims as 'the house of war', a realm of unbelief and strife. General Sherman's idea that 'war is hell' exploits the Christian idea of hell as a domain of radical disorder (Pandemonium) as well as of fire and torment. War can also stand for the clash of opposites from which (as Heraclitus taught) all things spring. For Clausewitz, war is a self-driven process having its own principles of order. And war may be understood as a social practice or institution that takes different forms over time or in different cultures. Understood as organized armed combat, war may be distinguished both from lawless violence and from other kinds of force and coercion (see Coercion; Violence). 'War', in short, is a kind of order as well as disorder.

Similar dichotomies mark conceptions of peace. 'Peace' can mean either a temporary suspension of hostilities (a truce) or a permanent condition of tranquillity in which war is no longer feared. Where war is identified with lawlessness, peace becomes not the mere absence of fighting but a secure condition of mutual cooperation between individuals within a community or between communities. The idea that peace depends on legal order is a recurrent theme of Western political philosophy, one stated eloquently by Thomas Hobbes in *Leviathan* (1651) (see Hobbes, T. §7).

Hobbes identifies war with the condition of humanity in the absence of civil order - a condition of nature in which each human being has a right to all things. The state of war is the opposite of the civil condition, and avoiding its inconveniences is the chief motive for seeking peace. War is the natural consequence of a disposition to fight that exists wherever there is no guarantee of peace. Peace requires assurance that one will not be forced to fight, and this is possible only in the civil condition. Peace is a state of affairs that emerges not from the kind of agreement that terminates a particular war, but from an agreement creating a permanent legal order with civil institutions to declare and administer its laws (see §5).

On this view, peace is disrupted not by the lawful use of armed force, but only by the use of force contrary to law. The postulated antithesis between war and law must be seen not only in causal terms but conceptually: not only is law needed to secure the conditions that lead to war, but ‘war’ *means* the abrogation of lawfulness. One cannot hold that people are obliged by law to respect one another’s rights and at the same time that they are free to attack one another at will.

The premise of most political philosophy is that war is an evil to be escaped or at least regulated. But some philosophers, like Machiavelli (§4) and Nietzsche, have considered war a good rather than an evil. It may be a field for the display of virtues like courage, virtues which are quite different from those urged by Jesus in the Sermon on the Mount, or a way of cleansing the corruption of a decadent civilization, thereby recovering the greatness of former times or developing unrealized human potentialities.

For other philosophers, war must be understood before it is judged. It calls for philosophical investigation because it is an aspect of experience, a part of reality. In one version of this approach (illustrated by J. Glenn Gray in *The Warriors* (1959)), war is understood phenomenologically by considering how it appears to those who experience it. In another, the study of war takes a scientific turn as the philosopher seeks to identify the ‘causes of war’ in human nature, in the character of different states or in the structure of the international system. The concept of innate human aggression, the hypothesis that ‘liberal’ states do not fight wars with one another and the argument that war is an inevitable consequence of a decentralized international order illustrate explanatory theories at each of these levels.

2 Grounds for war

Although the philosophical justification of war predates the modern states system, the problem of justification is most easily introduced in this setting. According to what might be called the ‘political realist’ point of view, there is in the international system no authority superior to that of the separate states. Force may be used by governments to uphold their own internal law, but there is no international law that it properly upholds (see International relations, philosophy of §2). War is an instrument of foreign policy and its use is restrained only by prudence, not justice. It follows that there can be no distinction between just and unjust wars. States may agree for prudential reasons to restrict the use of force but obligations thus incurred can be overridden by ‘reasons of state’; that is, by considerations of national security. Where vital interests are at stake, states remain free to violate their agreements, acting in ways that non-realists might deem ‘unjust’.

The alternative view is that one *can* distinguish lawful and unlawful uses of force between states, that is, distinguish just from unjust wars. Although this has long been a part of the tradition of Catholic natural law, this view has had many non-Catholic interpreters, from Grotius to Michael Walzer, and is thoroughly embedded in international law. An unjust war does not merely interrupt peace (understood as a condition of calm marked by the absence of fighting), it violates the lawful order that guarantees states their independence and territory. An unjust war is not a brute fact but a wrong, and as such it justifies the use of force by other states to resist and punish it.

Underlying the distinction between just and unjust wars is the assumption that there is a society of states comparable to the community of individuals that comprise a single state. In place of laws made and applied by sovereign authority, the laws of the international community rest on international custom and natural justice. Although the international order is more uncertain than that which prevails within a peaceful state, only if we presume such an order can we distinguish just (lawful) and unjust (unlawful) wars (see Justice, international §1-2). In the same way that the rule of law in civil society protects the liberty and security of each citizen, the rule of law in international society protects the political independence and territorial integrity of each state. A state that violates these principles makes itself liable to action by other states, acting singly or together, to protect their rights and defend the principles on which international order rests. Unlike a citizen, a state may fall back on self-help because there is no superior power on which it can rely.

The idea of just war may be seen as a middle position between realist and holy war conceptions. For political realists, any war a state chooses to fight to protect its security is justified. But because security can be threatened by the mere power of other states as well as their conduct, realist reasoning can be used to rationalize aggression. For holy warriors (and their secular imitators), in contrast, it is the imperative to disseminate the faith that justifies

attacking others and, often, subjugating, converting, expelling or exterminating them. If we understand war to be an instrument of human will to secure justice, a just war must be distinguished both from self-preservation and from the promotion of a religious or ideological cause.

Although it is agreed in the just-war tradition that a war must be fought for 'just cause', there is dispute about which grounds for war meet this criterion. According to Aquinas (§13), a just war is neither a war of conquest nor a war of self-defence. It is a war in which another community is attacked for refusing to redress wrongs inflicted by its members or to restore goods or territory it has unjustly seized (Aquinas 1266-73: II, 2, q.40, a.1). The United Nations Charter, in contrast, limits the grounds for war by states to self-defence. But such a limitation, taken literally, would forbid assisting another state to resist attack. It would also forbid armed intervention within another state to end abuses by its government, for a government can mistreat its own subjects without attacking other states - a fact international law now recognizes and tries to prevent by distinguishing crimes against humanity from the crime of aggressive war. If a just war is one that redresses wrongs, the crimes of a government against its own subjects (today understood as violations of their 'human rights') may be included among these wrongs and therefore among the lawful grounds for war.

A much-debated question is whether war can justly punish, as well as resist, wrongs committed by other states. Some argue that states may resist aggression but not punish it. Punishment must be licensed by an international authority, and there is no such authority. Others reply that if states are not authorized to punish wrongdoing, it is hard to see how they can be authorized to resist it, either. Yet both might agree that waging punitive war against entire communities, long sanctioned by moralists as well as by state practice, rests on unacceptable views of collective responsibility. The practice of reprisals, however, continues to be justified by some moralists as a deterrent to further wrongdoing.

Other considerations besides 'just cause' may be offered for judging decisions to initiate a war. Aquinas, for example, argues that war cannot be justly chosen, even by those whose cause is just, unless that choice has been authorized by a ruler: 'private wars' cannot be just. This constraint does not limit the wars of sovereigns against one another, however, unless there is a higher authority to whom they are subject. Today the United Nations claims such authority. Although the Charter prohibits the unilateral use of force by states except in self-defence against armed attack, the Security Council may authorize collective military action to rectify a variety of other wrongs.

Aquinas also repeats the judgment of Augustine, that those who choose war must do so to resist wrongdoing and secure peace, not for the sake of vengeance or to gain power. In other words, the intention with which one fights, like the act of fighting itself, must meet a moral test. On one view of intention, to intend is to choose and thus to act. But we can further distinguish between intentions (what is aimed at) and motives (the state of mind or 'spirit' in which one acts). Some in the just-war tradition argue that intentions are always morally relevant, even if motives are not. For them, a war justified by the 'cause' of reversing an unjust conquest (like the Iraqi conquest of Kuwait) but one which is really intended to achieve some less defensible end (like permanently weakening Iraq) must be deemed unjust, as must one carried out in a spirit of hatred or cynicism. For political realists, in contrast, neither intentions nor motives count for much.

Because wars can occur within as well as between communities, moralists have asked whether there can be just civil wars. Under what circumstances are revolutions or wars of national liberation justified? When is it proper to overthrow a tyrant or usurper, or to resist a government when it demands the performance of immoral acts like participation in genocide or service in an unjust war? How, if at all, can terrorist acts against a regime be defended? Is morally-justified resistance to authority limited to various forms of noncooperation, including conscientious objection? Although just-war and pacifist ideas are clearly relevant to these questions, answering them also raises questions about political authority and obligation (see Civil disobedience; Revolution).

3 Conduct of war

Just wars are defined not only by the ends for which they can be fought but by the means used to fight them. If war is a practice, it has rules. Of these, rules restricting the targets that can be attacked are more basic, morally speaking, than those restricting the use of particular weapons or techniques like napalm or area bombing. Some moralists argue that the rules of war are adequately summed up in the principle that the harm inflicted by military

operations should be ‘necessary’ for and ‘proportionate’ to the ends sought, others insist that attacks on persons or objects not involved in fighting are wrong for reasons unrelated to these expediential considerations. Such attacks are wrong because they violate a fundamental prohibition against harming the innocent or their possessions.

This further restriction on the conduct of war is often stated as the proposition that persons who fight must not intend, either as an end or a means, the deaths of those not directly involved in the fighting. Waging war in ways that fail to discriminate between combatants and noncombatants is morally forbidden. In the classic vocabulary of natural law, those who are not engaged in fighting are ‘innocent’ (from Latin *nocens*, engaged in harmful activity). Innocents may not be attacked ‘intentionally’, although no wrong is done if they are harmed incidentally, as a foreseeable but unavoidable side-effect of permissible military action. According to the principle of ‘double effect’, acts that harm innocents may be justified if such harm cannot be avoided in pursuing a morally permissible end and the level of such harm is not disproportionate. One question in calculating proportionality is whether one should consider impartially the interests of everyone affected by an act, as both natural lawyers and many consequentialists argue, or should favour the interests of one’s own soldiers and civilians, as political realists hold. A more difficult issue is whether the distinction between the intended and unintended effects of an act is philosophically coherent (see Double effect, principle of).

An even stricter position is that one is barred from intending the deaths of combatants as well as noncombatants. One may seek to thwart an attack, but in doing so one must use no more than the minimum force needed. Here, too, the double effect principle may be invoked to distinguish permissible from impermissible harm. That one can foresee the deaths of enemy soldiers whose attack one is resisting does not mean that one intends their deaths, provided that in resisting one is choosing only to repel the attack, and that the foreseen deaths are the unavoidable consequence of using the only means of resistance available.

Should moral limits on the conduct of war be ignored if the stakes are high enough? If a country is about to be overrun by a genocidal adversary, many think that the victim may use any means to defeat the conqueror. Moral constraints that hold in ordinary situations are not binding in situations of extremity. This view reverses the usual structure of just-war argument: thus, although under normal circumstances moral principles constrain consequentialist calculations, when catastrophe threatens such calculations may be substituted for reliance on moral principle. There is an escape clause in cases of authentic necessity.

Most natural law and Kantian theories of morality reject the argument from extremity. In situations of communal emergency one might justifiably waive rules based on mere custom or utility. But one may not violate the exceptionless precepts of the moral law, like those forbidding aggression or the deliberate killing of innocents. For these reflect the basic principle (sometimes called ‘St Paul’s principle’) that evil may not be done for the sake of good. To permit such acts is to hold that a moral community can survive by measures that its own traditions condemn as immoral. In abandoning its principles, the community would be trying to preserve the lives of its members while denying the principles that define its integrity as a community. This problem does not arise for political realism, which treats all rules as merely conventional and instrumental. Realism assumes that a community can temporarily transgress its moral principles without changing its character as that community.

The argument from extremity plays an important part in the debate over nuclear deterrence. In so far as it targets civilians, and thus intentionally threatens innocents, a nuclear retaliatory strategy is not easily reconciled with ordinary moral ideas. It is sometimes argued that the deterrent is not immoral because it only threatens, but does not actually destroy, innocent lives. But if (see §2 above) intentions are themselves actions - if they involve choices and commitments, can determine the character of the chooser and can have a variety of consequences - it can be wrong to intend what it is wrong to do (the ‘wrongful intentions principle’). Unless it is just a bluff, the deterrent threat expresses an intention to do wrong under certain conditions and is therefore morally objectionable. It is doubtful whether such stratagems as counterforce targeting or the mere possession of nuclear weapons can justify nuclear deterrence under ordinary moral principles. The claim that the nuclear predicament creates a (permanent) state of emergency, which therefore necessitates the deterrent threat, is yet another way of attempting such a justification.

4 Pacifism and nonviolence

Because those who condemn war do so on many grounds, there are many different kinds of ‘pacifism’. Some

pacifisms are personal (it is morally wrong for me to fight), others general (no-one may fight). Some object to organized war, others to killing human beings, still others to killing any living creature or to any use of force whatsoever. In each case, war is opposed as a matter of moral principle.

But the label ‘pacifist’ is also applied to the position of those who, like Mahatma Gandhi or Martin Luther King, Jr, urge nonviolent resistance on pragmatic as well as moral grounds. Arguing that power rests on acquiescence, its advocates contend that tyranny and aggression can be resisted without armed force (although not without a kind of nonviolent coercion). Nonviolent resistance can range from militant noncooperation (like Sorel’s general strike) to the practice of ‘living in truth’ followed by dissidents in communist eastern Europe during the 1980s.

So-called nuclear pacifism, which objects to war involving nuclear but not conventional weapons, would seem to follow from just-war or perhaps even realist premises rather than from a principled objection to war as such. What is to be abolished is not war itself but a weapon that makes war too terrible to fight. More radical are movements that seek to abolish war by getting rid of war-causing social or political institutions (capitalism, the military-industrial complex, patriarchy) or by reforming the international order (through promoting free trade, collective security or world government).

Christian pacifism rejects war as an instrument of human purposes while recognizing it as a recurrent situation in which Christians must bear witness to their faith, serve others and search for nonviolent ways of dealing with a hostile world. Pacifism of this kind is an interpretation of the meaning of a religious faith. Like the early Christianity from which it takes its guidance, it focuses on creating a community of believers committed to living their lives together according to Christian principles. There are Christian pacifists like Tolstoi who see government as intrinsically violent and who therefore incline to anarchism: to govern is necessarily to act evilly, like Caiaphas, for the common good. Although often branded as ‘sectarian’ for turning away from the world, Christian pacifism can be missionary, concerned with transforming the world.

Pacifists attempt to seize the moral high ground, charging not only realists but just-war moralists with a myopic acceptance of the existing world and lack of faith in the possibility of social transformation. They also accuse those who believe war can be justified of presuming, against the evidence, that force is more effective than nonviolent action. In so far as the issue is whether morality permits war in some circumstances, pacifists are engaged in debate with those committed to the just-war idea. But if the issue is whether morality applies to war at all, pacifists find themselves allied with just warriors against political realists and others for whom morality is circumscribed by prudence. And in invoking necessity, even realists make clear their aversion to war, thereby taking sides with the just warrior and the pacifist against those who celebrate war as a good in itself.

5 Peace as an institution

For many philosophers, peace is an aspect of legal order. But what kind of legal order? For law can be a tool of domination and oppression, a form of war by other means - as Augustine was not the first nor Foucault the last to observe. Any legal order that can establish and guarantee peace acknowledges the subjectivity (and therefore the personality, freedom and human rights) of those who belong to it. So-called ‘martial law’, which is really a suspension of law, is an expedient rationalized as necessary to restore legal order. It is not the foundation on which permanent peace, an ordered tranquillity, can be constructed.

It is sometimes argued that although peace based on law can be enjoyed by portions of humanity (if states can be established), such peace is not possible for the world as a whole. Hegel argued that a world state is inconceivable because the existence of one state implies the existence of others: ‘without relations with other states, the state can no more be an actual individual than an individual can be an actual person without a relationship with other persons’ (1821: §331). The state therefore cannot exist alone; it needs other states to accord it recognition and give it an identity. But a state might acquire an identity in resisting other associations: pirates, terrorists or insurgents, for example. The idea of a world state is different from that of a state among states, but it does not follow that a single law-making authority for all humanity is inconceivable.

Others perceive not logical but practical difficulties: a world state would be ungovernable or likely to suppress human diversity and freedom. For Kant (§10), a world state is conceivable but would pass through despotism back to anarchy. For Hobbes, a world state is conceivable but unnecessary because adequate security can be enjoyed in much smaller associations: the state of nature that exists between sovereign states is not nearly as bad as that

which exists among individuals in the absence of civil order.

Both these conclusions may seem doubtful, however. If it is the fear of war that motivates individuals to accept political authority, one can imagine circumstances in which this fear is better escaped by creating a single world state, which would abolish war between states, than by creating many states, which can only perpetuate war. This argument becomes more persuasive as war becomes more dangerous. If, in contrast, states can coexist without threatening the lives or prosperity of their members, then individuals cannot be rationally motivated to establish a world state. But this second line of argument seems to undermine the case for any state. If states can coexist in a condition of nature, why cannot individuals or voluntary associations of individuals do the same? If the state of nature is tolerable for sovereigns then, under some imaginable circumstances, it may be tolerable for others. Even though its inconveniences may be greater, it does not follow that these must be so great as to necessitate sovereignty.

Although he thinks a world state would endanger rather than secure perpetual peace, Kant (1795) insists on the moral necessity of establishing a global legal order by means of a contract among independent states, for the rule of law among states both depends upon and reinforces the rule of law within each state. If peace depends on legal order, it must be ‘formally instituted’ by means of a constitution, internationally as well as internally. But the order established by the international constitution is confederal rather than unitary. The crucial elements of this constitution are: (1) that every state shall have a ‘republican’ constitution - that is, one that makes government the common business (*res publica*) of the people, and ensures that it is conducted according to law; and (2) that the international order must be a federation of republican states, not a world state. The first condition means not that each state must be a democracy, for democracies can be despotic, but that each should be governed by the rule of law; for Kant, this means separating the legislative and executive powers. If there is to be an international order based on just laws, each state must conduct its internal as well as external affairs on the basis of such laws. The second condition (federation) demands that even republican states give up their liberty to make war. If states do not yield this sovereign prerogative, they remain in a state of nature. To establish permanent peace, they must renounce not only the means of war (for example, standing armies) but the right of war. The new power to which they surrender this right must be a federation of republics.

Kant’s argument, then, is that to secure permanent peace, states must first agree to acknowledge one another’s subjectivity by observing the rules of civilized international relations. Those that are republics must then establish a federation. If a state is not a republic, it must become one before it can join this federation. There will be a universal, permanent peace only when all states are republics and have entered the federation. This argument seems open to the objection that a federation in which states can remain sovereign would not succeed in bringing them out of the state of nature. Either the members retain their sovereignty, in which case the state of nature continues, or they do not, in which case we have a world state, not a confederation of independent states. It is an open question whether this contradiction can be resolved (see State §3).

6 War and peace within and between communities

For Hobbes, the rule of law is possible within but not between political communities. For Kant, permanent peace based on the rule of law is possible at the international level, but it holds only among republics - that is, states governed according to the rule of law. War remains a real possibility in relations with other kinds of states. The world is bifurcated into two realms, each ruled by its own principles. The divided structure of international order implicit in these arguments appears in many other theories of international relations. Socrates suggests that Greeks and barbarians are natural enemies who may ravage one another in war in ways that are entirely inappropriate among Greeks (Plato: 469b-471c), a sentiment later echoed by J.S. Mill in the context of European imperialism. Muslims distinguish between the house of war and the house of Islam, Christian pacifists between the house of fear and the house of love. Liberal thinkers divide the world into a ‘free world’ or ‘zone of peace’ and a world inhabited by the enemies of freedom, states which, because of the character of their regimes, are not ‘peace-loving’. In the international law theory of the former Soviet Union, the rules governing cooperation within the family of communist states differ from those regulating the coexistence of communist and non-communist states.

This multilevel conception of international order continues to be influential. John Rawls’ ‘law of peoples’, for example, invokes the idea of a Kantian federation embracing liberal (what Kant calls ‘republican’) states, based on

principles of non-aggression and non-intervention. However, Rawls' law permits liberal states to intervene in the internal affairs of non-liberal societies (so-called 'outlaw states') (Rawls 1993).

Implicit in all these views are actually three rather than two levels of international order. The first is a condition of war among communities forced to take each other's power into account but unwilling to grant each other recognition. The second level is a pluralist society composed of states that recognize one another's rights as states, no matter what the character of their internal regimes; its rules are those of a minimum order of coexistence based on mutual restraint but not on shared values or a common way of life. Here, although not a constant condition, war is always possible and sometimes justified. The third level is a peaceful international society of rule-of-law states in which each conducts both its internal and external affairs according to shared standards of justice.

See also: Bushi philosophy; International relations, philosophy of; Sunzi

TERRY NARDIN

References and further reading

Airaksinen, T. and Bertman, M.S. (eds) (1989) *Hobbes: War among Nations*, Aldershot: Avebury Press. (Considers many aspects of the Hobbesian analysis of war and peace.)

Aquinas, T. (1266-73) *Summa theologiae (Synopsis of Theology)*, selections translated in *On Law, Morality and Politics*, ed. W.P. Baumgarth and R.J. Regan, Indianapolis, IN: Hackett Publishing Company, 1988.(Includes Aquinas' brief but classic essay on war)

Barnes, J. (1982) 'The Just War', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press. (Provides a context for understanding Aquinas on war.)

Ceadel, M. (1987) *Thinking about War and Peace*, Oxford: Oxford University Press.(A clear introduction to a wide range of positions in twentieth-century debates about war and peace.)

Finnis, J., Boyle, J.M., Jr and Grisez, G. (1987) *Nuclear Deterrence, Morality and Realism*, Oxford: Clarendon Press.(Natural law treatment of deterrence that covers much of the literature and debate; excellent bibliography.)

Forsyth, M. (1981) *Unions of States: The Theory and Practice of Confederation*, Leicester: Leicester University Press.(Clear, comprehensive and reliable study of federalism as an approach to peace.)

Gray, J.G. (1959) *The Warriors: Reflections on Men in Battle*, New York: Harcourt Brace.(Readable war memoir by a philosopher.)

Hegel, G.F.W. (1821) *Elements of the Philosophy of Right*, ed. A.W. Wood, Cambridge: Cambridge University Press, 1991.(Discusses international relations and war in §§330-40.)

Hobbes, T. (1651) *Leviathan*, ed. E. Curley, Indianapolis, IN: Hackett Publishing Company, 1994.(Presents law and war as conceptual alternatives.)

Kant, I. (1795) *Perpetual Peace*, trans. H. Nisbet in H. Reiss (ed.) *Political Writings*, Cambridge: Cambridge University Press, 2nd edn, 1991.(Includes several essays on war, peace and international relations.)

Kenny, A. (1985) *The Logic of Deterrence*, London: Firethorn Press.(An accessible investigation of arguments for and against nuclear deterrence and nuclear disarmament.)

Lackey, D.P. (1989) *The Ethics of War and Peace*, Englewood Cliffs, NJ: Prentice Hall.(An introductory textbook covering pacifism, just-war theory and nuclear strategy, with many examples and extensive suggestions for further reading.)

Nardin, T. (ed.) (1996) *The Ethics of War and Peace: Religious and Secular Perspectives*, Princeton, NJ: Princeton University Press.(Compares natural law, realist, Jewish, Islamic, pacifist and feminist viewpoints.)

Norman, R. (1995) *Ethics, Killing and War*, Cambridge: Cambridge University Press. (Rejects the just-war theory in favour of a consequentialist ethic of war.)

Pangle, T. (1976) 'The Moral Basis of National Security: Four Historical Perspectives', in K. Knorr (ed.) *Historical Dimensions of National Security Problems*, Lawrence, KS: University Press of Kansas.(Surveys the ideas on war and peace of the major philosophers.)

Plato (*c*.380-367 BC) *Republic*, trans. R. Waterfield, Oxford: Oxford University Press, 1994, 416d-324a.(Implies but does not develop a theory of international relations.)

Rawls, J. (1993) 'The Law of Peoples', in S. Shute and S. Hurley (eds) *On Human Rights*, New York: Basic Books.(Extends Rawls' theory of justice to international society.)

Waltz, K.N. (1959) *Man, the State and War*, New York: Columbia University Press.(Treats philosophers as seeking causal explanations of war in the individual, the society and the international system.)

Walzer, M. (1977) *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, New York: Basic Books.(The best introduction to its topic.)

Yoder, J.H. (1992) *Nevertheless: Varieties of Religious Pacifism*, revised, Scottdale, PA: Herald Press.(Discusses alternative conceptions of pacifism from an evangelical perspective.)

Watsuji Tetsurō (1890-1960)

Watsuji Tetsurō stands out as the leading thinker on ethics in twentieth century Japanese philosophy. He is regarded as a peripheral member of the 'Kyoto School' of philosophers centring around the thought of Nishida Kitarō. Like Nishida and the Kyoto School, the thought of Watsuji can be characterized by the effort to formulate a syncretic East-West philosophy developed within the framework of a Buddhist metaphysic of 'emptiness'. At the same time, Watsuji established his own highly distinctive system of ethics. He must rank as one of the most creative and profound thinkers in modern Japanese philosophy.

Watsuji was born in Himeji, Hyōgo prefecture, and entered the philosophy department at Tokyo University in 1909. From 1925 to 1934 he was a professor of ethics in the philosophy department at Kyoto University, and was a colleague of Nishida Kitarō, Tanabe Hajime, Kuki Shūzō and others associated with the 'Kyoto School' of modern Japanese philosophy (see Kyoto School). During this period Watsuji also spent two years studying abroad in Germany under the direct tutelage of Martin Heidegger. In his later years Watsuji moved to Tokyo University, where he taught from 1934 until his retirement in 1949.

The *Watsuji Tetsurō zenshū (Complete Works of Watsuji Tetsurō)* fill twenty volumes and cover a vast range of topics in philosophy, art, religion and cultural history. Whereas his earliest works focussed on Western philosophers, notably studies of Nietzsche and Kierkegaard, in his book entitled *Guzōsaikō (Revival of Idols)* (1918) he announced a 'turning point', a shift from Western individualism to a social and ethical standpoint represented by Sino-Japanese modes of thought. At this point, Watsuji's interest in Asian and Japanese moral thought was influenced especially by the *samurai* ethic of group-loyalty and self-sacrifice (see Bushi philosophy) as developed by his teacher Nitobe Inazō, in a book entitled *Bushidō, The Soul of Japan* (1905). He was also indebted to another of his teachers, the writer Natsume Sōseki, who describes the turn from egocentrism to a Buddhistic social concept of existence in his various letters, literary essays and novels. Among Watsuji's investigations into Japanese Buddhist philosophy, art and culture, of special importance is his pioneering study of Dōgen; his study *Shamon Dōgen (Dōgen, The Novice)* (1920-3) is regarded as the work responsible for rediscovering the significance of Dōgen as the foremost original thinker in classical Japanese Buddhist philosophy.

Watsuji is above all recognized for his ethics and moral philosophy, as developed in a three-volume work entitled *Rinrigaku (Ethics)*. For Watsuji a system of ethics requires an adequate definition of the person. From this perspective, he criticizes Western liberal ethics based on individualism, egocentrism and atomism as the departure point for working out a Japanese communitarian ethics of the family system based on a social, relational and contextual model of self. According to Watsuji's philosophical anthropology, the Japanese word for 'person' (ningen) clarifies the twofold structure of the self as both individual and social in character. This social understanding of the self in turn provides the basis for a communitarian system of ethics based on the 'betweenness' (*aidagara*) of persons in relation. In Watsuji's ethics, this social concept of the person as an individual-society interaction implies not only a moral but also a deeply aesthetic mode of human existence as manifest through Japanese social art forms such as the tea ceremony (*chanoyu*) and linked verse (*renga*) (see Aesthetics, Japanese).

Watsuji's social concept of Japanese personhood represents an original synthesis of Confucian and Buddhist ideals of the self (see Confucian philosophy, Japanese; Buddhist philosophy, Japanese). He articulates a Confucian model of self as a network of familial and social relationships normatively governed by rites of propriety which embodies increasingly larger social groups in the process of becoming a person. At the same time, his social concept of the person is ultimately grounded in a Buddhist metaphysic of 'emptiness' (*kū*) whereby there is a dependent co-arising of the individual and society (see Buddhist concept of emptiness). There is thus a systematic character underlying Watsuji's framework in that his aesthetics and ethics are both developed in connection to his social concept of the person, which is in turn formulated in the context of a Buddhist notion of emptiness as interrelational existence.

In his most popular book, *Fūdo (Climate)*, first published in 1934, Watsuji criticizes Heidegger's ethics as being grounded in an individualistic concept of self based on a one-sided phenomenological description of authentic human existence as a being-in-time. According to Watsuji the Japanese concept of 'person' (*ningen*) has a twofold

structure as both an individual existence in time and a social existence in a 'climate' (*fūdo*) of space. Watsuji thus establishes the basis for an environmental ethics arguing that the self arises through a relationship between the individual and its entire spatial climate including both human society and living nature (see Society, concept of).

Scathing criticisms have often been levelled against the ultra-nationalism, totalitarianism and authoritarianism of Watsuji's socio-political framework, by both Japanese and Western scholars. Critics point out that in Watsuji's ethics of the Japanese family system and its underlying metaphysic of emptiness, the result is a totalitarian ethics wherein the individual is 'emptied' into progressively larger household groups, finally emptying into the Imperial household of the Japanese nation-family through an ideal of loyal self-sacrifice to the Emperor. Although Watsuji's philosophical anthropology starts with an effort to formulate a balanced notion of self as a twofold individual-society relation, thereby to establish a middle way between individualism and collectivism, it ends by falling into a mode of totalitarian collectivism wherein the individual is emptied into the whole of the nation through the absolute negativity of Buddhist emptiness. Moreover, Watsuji's ethical framework has been charged with falling into a dangerous form of moral relativism which defines good and evil or right and wrong in terms of the particular familial group as the frame of reference, not as universal principles which transcend the group.

See also: Buddhist philosophy, Japanese; Confucian philosophy, Japanese; Ethics; Japanese philosophy; Kyoto School; Totalitarianism

STEVE ODIN

List of works

Watsuji Tetsurō (1913-49) *Watsuji Tetsurō zenshū (Complete Works of Watsuji Tetsurō)*, Tokyo: Iwanami Shoten, 1961-3, 20 vols.(Edition of Watsuji's complete works.)

Watsuji Tetsurō (1934) *Fūdo (Climate)*, trans. G. Bownas, *A Climate*, Tokyo: Unesco, 1961.(Work on climate and nature.)

Watsuji Tetsurō (1934) *Ningenaku to shite no tetsugaku (The Significance of Ethics as the Study of Man)*, trans. D.A Dilworth, *Monumenta Nipponica* 26, 1971 (3-4).(Work on ethics.)

Watsuji Tetsurō (1934) *Nihon no bungei to bukkyō shishō (Japanese Literary Arts and Buddhist Philosophy)*, trans. Hirano Umeyo, *The Eastern Buddhist*, 4, 1971 (1): 88-115.(Article on Buddhism and literary aesthetics.)

References and further reading

Bellah, R.N. (1965) 'Japan's Cultural Identity: Some Reflections on the Work of Watsuji Tetsurō', *Journal of Asian Studies* 24 (4): 573-94.(Looks at Watsuji's importance in the study of Japanese culture.)

Dilworth, D.A. (1974) 'Watsuji Tetsurō: Cultural Phenomenologist and Ethician', *Philosophy East and West* 24 (1).(Looks at his ethical and cultural works.)

Kodera, T.J. (1987) 'The Romantic Humanism of Watsuji Tetsurō', *Dialogue and Alliance* 1 (3): 4-11. (Discussion of Watsuji's views.)

LaFleur, W.R. (1978) 'Buddhist Emptiness in the Ethics and Aesthetics of Watsuji Tetsurō', *Religious Studies* 14 (June): 237-50.(The impact of Buddhism on Watsuji's thought.)

Nagami Isamu (1981) 'The Ontological Foundations in Tetsurō Watsuji's Philosophy: *Kū* and Human Existence', *Philosophy East and West* 31 (3): 279-96.(Looks at Watsuji's ontology.)

Odin, S. (1992) 'The Social Self in Japanese Philosophy and American Pragmatism: A Comparative Study of Watsuji Tetsurō and G.H. Mead', *Philosophy East and West* 42 (3).(Cross-cultural study of Watsuji and Mead.)

Weber, Max (1864-1920)

Max Weber, German economist, historian, sociologist, methodologist, and political thinker, is of philosophical significance for his attempted reconciliation of historical relativism with the possibility of a causal social science; his notion of a verstehende *(understanding) sociology; his formulation, use and epistemic account of the concept of 'ideal types'; his views on the rational irreconcilability of ultimate value choices, and particularly his formulation of the implications for ethical political action of the conflict between ethics of conviction and ethics of responsibility; and his sociological account of the causes and uniqueness of the western rationalization of life.*

These topics are closely related: Weber argued that the explanatory interests of the historian and social scientist vary historically and that the objects of their interest were constituted in terms of cultural points of view, and that consequently their categories are ultimately rooted in evaluations, and hence subjective. But he also argued that social science cannot dispense with causality, and that once the categories were chosen, judgments of causality were objective. The explanatory interests of the sociologist, as he defined sociology, were in understanding intentional action causally, but in terms of categories that were culturally significant, such as 'rational action'. Much of his influence flowed from his formulation of the cultural situation of the day, especially the idea that the fate of the time was to recognize that evaluations were inescapably subjective and that the world had no inherent 'meaning'. The existential implications of this novel situation for politics and learning were strikingly formulated by him: science could not tell us how to live; politics was as a choice between warring Gods. Weber's scholarly work and his politics served as a model for Karl Jaspers, and a subject of criticism and analysis for other philosophers, such as Karl Löwith, Max Scheler, and the Frankfurt School.

1 Life

Weber was born in Erfurt in Thuringia 12 April 1864. His father was a businessman and prominent National Liberal politician, who served in the Prussian House of Deputies (1868-97) and the German Reichstag (1872-84). The Weber household was a meeting place for prominent academics and political figures. Weber began his university training in Heidelberg in 1882 and continued it in Berlin, Strasbourg and at Göttingen to 1886. He attended lectures on law and history, passing his bar exam in 1887. He received a Berlin doctorate in 1889 for a study of medieval trading companies: his habilitation, in 1891, was for a study of legal aspects of Roman agrarian history. While preparing for his habilitation, he participated in a study under the sponsorship of the Verein für Sozialpolitik (Social Policy Association) on the subject of agricultural labourers in the East Elbian estates. He began a legal career, but soon returned to academic life. From 1894 to 1897 he was a professor of Nationalökonomie at Freiburg; from there he was appointed to a similar position at Heidelberg. A mental breakdown in 1902 led him to relinquish his teaching duties, but he soon returned to an active role as an editor and producing scholar. His sociological works began to appear after the 1904-5 publication of *The Protestant Ethic and the Spirit of Capitalism* and included a series of works on the sociology of religion dealing primarily with the theme of the economic ethics of the world religions. It is controversial whether these works or his *Wirtschaft und Gesellschaft (Economy and Society)*, posthumously published in 1922, represented the core of his thought.

His major 'methodological' essays (posthumously published as his *Gesammelte Aufsätze zur Wissenschaftslehre (Collected Essays in Scientific Theory)* in 1922) included part one of 'Roscher and Knies', published in 1903, '"Objectivity" in Social Science and Social Policy', in 1904, part two of 'Roscher and Knies' and 'Critical Studies in the Logic of the Cultural Sciences', in 1905 and 1906, and, in 1907, 'Critique of Stammler'. Among the other essays collected in the *Wissenschaftslehre* were 'The Meaning of "Ethical Neutrality"', published in 1918, which grew out of an earlier dispute with the Social Policy Association on the possibility of making policy without making value-choices that cannot be grounded scientifically, and a widely distributed speech, 'Science as a Calling', published in 1919. This and another speech, 'Politics as a Calling', given and published in 1919, were actively debated by the younger generation of intellectuals.

After the war Weber served as a member of the constitutional commission of the Weimar republic, became a professor of economics at Vienna in 1918 and Munich in 1919, and unsuccessfully pursued a career in politics. His early death came when his influence among intellectuals was at its peak. The 'Calling' speeches were an aggressive and telling formulation of a world-view that challenged the cultural and political optimism of the time,

and remained in the Weimar era as a central reference in the long-running discussion of the ‘Crisis of the Sciences’. Much of his work was published posthumously in editions that did not reflect his original intentions, but his ‘methodological’ writings were intended to be collected, and the collection largely follows his intentions. The major work published as *Economy and Society*, however, was unfinished, and his intentions for the texts that compose the extant edition were unclear.

Weber in his political writings and methodological writings rigorously separated factual and evaluative considerations, especially in connection with the consideration of policy means to political ends, in order to demonstrate the concealed and confused value premises of the teaching of the ‘Socialists of the Chair’ and the factual naïveté of ‘ethical’ political postures, and he promoted the idea that intellectual integrity demanded the making of explicit ultimate value choices. He believed that in a democratic age the only political leaders who could make free value choices and rise above mere interest-politics to greatness were those who had the charisma necessary to build a personal following, and he devoted his constitutional efforts to the task of creating means by which such leaders could rise.

2 The methodological essays of 1903-7

Weber’s major methodological article, ‘Objectivity in Social Science and Social Policy’, published in 1904, brought together the key themes of his thinking as a methodologist of social science. Weber argued that the results of the cultural sciences could be valid in an objective sense, and that indeed, ‘scientific truth is precisely what is *valid* for all who seek the *truth*’ (1904-17: 84). Yet he also argued that the objects of explanation in social science are constituted by culture-specific and discipline-specific cognitive interests and consequently are historically relative and in this sense subjective (see Explanation in history and social science; Value judgments in social science). He resolved the conflict between these two claims by a complex argument. The premises were epistemic: exhaustive description and hence explanation of concrete reality was impossible; all science is necessarily selective and concerned with a finite portion of reality. The natural sciences are, however, selective in a different way than the cultural sciences. The natural sciences select as important the segment that can be constructed as casual regularities, and take it to be the essence of reality. The cultural sciences, in contrast, begin with the finite segment that is significant to us as cultural beings with a specific culture, and then construct within this segment objects of explanation, which he called, following Heinrich Rickert, ‘historical individuals’ that are significant or meaningful to us as cultural beings. This enterprise is necessarily ‘evaluative’ in the sense that only as cultural beings do we regard these objects as significant, and culture is conditioned by valuative ideas. Accepting as our starting point this distinctive segment of reality has the effect of excluding the possibility of a social science that is based on principles like those of physics or chemistry operating at the level of an atomistic psychology, because a physics-like social science that reduced social life to elementary factors, if such a thing were indeed possible, would simply fail to apply to objects that are configured to be culturally significant to us, and would apply to a different or differently configured finite segment of reality.

One difficulty this line of argument poses arises over the question of whether causation itself is relative to our cultural starting point and interests, and is thus not a truth there for all who seek it? Weber’s response to this question is elaborated in ‘The Logic of the Cultural Sciences’ (1904-17), in which he argues that historical explanation is logically equivalent to establishing legal responsibility, and applies a contemporary legal theory of probabilistic causation, the theory of ‘objective possibility’ and ‘adequate cause’, to the problem of the objectivity of causal explanation. Legal determinations of responsibility perforce employ categories that arise from a particular and distinctive interest, the juridical interest; the cultural scientist simply has different interests and different categories.

The criteria for ‘adequate’ causation in the theory is the existence of a significant difference between the probability of a given result if a particular set of conditions held (the background probability), and the probability of the result if these conditions absent the condition in question. If purely classificatory concepts could be employed and probabilities for the two cases empirically derived, these judgments would be ‘empirical’. Thus the core of causal judgment in this theory is an objective fact. But there are two kinds of selection that are ordinarily involved in historical explanation that are not objective. Whether a difference in probabilities is significant, and hence whether a cause is ‘adequate’, depends on the set of conditions selected to provide the background probability. This choice cannot be made objectively, but must be made on the basis of the historian’s interest. A

second kind of selection occurs because 'historical individuals', the objects of the concepts of interest to historians, cannot be reduced to the sum of the ways they can be classified. So the historical analyst proceeds by constructing abstract ideal types and comparing them to other causal situations that are also known to be irreducible to the ways they can be classified. Judging the 'adequacy' of a historical causal relation is thus, in practice, largely abstract or conceptual, and unavoidably so.

Weber used these considerations to support the claim that explicitly constructed conceptual schemes were necessary in the cultural sciences, for example in economics. But he argued for a particular understanding of these constructions: as ideal types that diverged from empirical reality in its full richness. The constructed types of the cultural sciences resemble those of the natural sciences with respect to selectivity; they differ in the cognitive purposes they serve: the cultural scientist seeks conceptual clarity in relation to the particular intellectual aims of particular disciplines, each of which arises from practical purposes (see Value judgments in social science §§1-2). Weber also used this understanding of ideal-types critically, to argue against the taking of ideal type constructions for a deeper or essential reality. Crude Marxism, he thought, traded on this error, as did German theories of the state.

Weber's ideal types of legitimate rule (*Herrschaft*) were perhaps the most influential of his constructions. Weber distinguished between three kinds of belief in the legitimacy of orders or rulers: traditional, based on unwritten rules; rational-legal, based on written rules; and charismatic, where beliefs about the extraordinary character of a leader justified command. He showed how these beliefs were associated with particular kinds of social organization, and how particular kinds of belief, such as 'traditional' patrimonial beliefs were associated with orders that developed in history in particular ways. Charismatic authority, which in its pure form is personal, is unstable and short-lived, though it can be combined with other forms to ground a routinized order, such as a monarchy, which retains some charismatic elements, such as the notion of royal blood. Rational-legal orders make belief in the legality of the order the basis of legitimacy. This form characteristically results in the bureaucratization of administration. Weber explicitly intended the concept of bureaucracy to be a timeless and value free ideal type rather than a historical teleological construct. But the vision of a world of men who are little cogs trying to become bigger cogs in the huge machinery of modern bureaucracies horrified him and represented the end of the practical possibility of an individual making a choice to devote his life to some higher self-chosen goal.

3 The idea of an 'understanding' sociology

Weber's essay on 'Some Categories of Interpretive Sociology' (1913) and his introduction to *Economy and Society* stipulated a definition of sociology as the science concerned with the interpretive understanding (*Verstehen*) of social action, that is, actions that take account of others, and the causal understanding of the course and consequences of actions (see Social action). The intended meaning of an act for its agent is causally essential to an act. Accordingly, a sociological explanation of an action should be correct both on the level of cause and meaning. The understanding of meanings requires the application of type-concepts rather than empathy. Meanings are ascertained by comparing the course of action to an appropriate ideal type of a meaningful act, such as obedience to a command. Many of the causes bearing on social life, such as psychophysical or biological effects, are not intentional in character, and thus are not part of sociology as Weber defines it. 'Collective' concepts, such as the state, understood as real entities, are also excluded by Weber's definition, but the social phenomenon they refer to may be understood sociologically in terms of individual actions: the concepts are ideal-typifications of patterns of action that in turn result from the beliefs and expectations of individuals.

Subsequent commentators on Weber have made a great deal of his use of the notion of *Verstehen*, and have argued that a social science based on *Verstehen* or understanding must be radically unlike one based on any analogy to the natural sciences. The distinction between understanding and explanation was treated as a fundamental epistemological divide between two kinds of science. When Weber's works were imported to the Anglo-American world, especially during the 1930s, primarily to an audience of sociologists, this distinction was often stated as methodological distinction, and *Verstehen* was understood as a 'method'. Social phenomenologists, such as Alfred Schütz, attempted to provide a basis for an alternative social science in the problem of understanding and particularly in the problem of the attainment of 'adequacy' in the individual's understanding of another.

Weber's own statements about *Verstehen* are characteristically directed toward other issues. *Verstehen* entered

primarily in connection with concepts used by the historical figure whose actions were being explained as well as by those employed by the historical interpreter. Since the historical interpreter often was part of a culture which employed different concepts, there is a positive task of interpreting the purposes and therefore the concepts of the agent whose actions are to be explained in terms of broader and more permanent categories of action or in terms of the concepts of the historian and his own historical audience.

The treatment of intention at the beginning of *Economy and Society* distinguishes between direct and indirect evidence of intention. This is a distinction borrowed from the law, and reflects Weber's utterly unmystical attitude to assessing an agent's purposes and motives. Weber specifically rejected appeals to historical intuition. His discussion of interpretation in the beginning of *Economy and Society* treats the attribution of motivation and intent as a practical necessity dictated by the task of the inquiry, which could ordinarily be done by examining the course of action and the information available to the agent. Assessing such things requires, to some extent, thinking through a problem from the point of view of the agent, but Weber stressed that one did not need to be Caesar in order to 'understand' Caesar, but rather to assimilate the acts of a Caesar to an already 'understood' model or ideal type of action. Models of perfect rational action are, in practice, a starting point for reconstructions of intentions in complex situations, such as the battlefield decisions of a general, for which there is no alternative.

4 The conflict of values and world-view

Weber's writings on the subject of values, value-orientations, and value conflicts were directed at specific topics, such as the ethics of advocacy in the classroom, but had a well-defined philosophical core. Weber held that there were rational grounds for the selection of means to ends, and for assessing the consistency of sets of choices of ends, but that ultimate value choices could not be rationally justified. He also held that hidden conflicts became apparent when means and the subsidiary consequences of the use of the means to obtain given ends were considered, and that such conflicts were common.

In the sphere of politics, Weber argued, the decisive means is violence, and this inevitably produces conflicts of a particular kind. Consider the political leader who is an adherent to what Weber calls an 'ethic of uncompromising conviction' with respect to truth-telling in politics, that is, who upholds the value of truth-telling come what may. A foreseeable subsidiary effect of an instance of truth-telling might be to unleash uncontrollable passions. Weber argues that the only way an ethic of uncompromising conviction may be consistently upheld in the political sphere would be for its adherents to disclaim responsibility for such results of their actions, as do religious ethics which counsel their followers to 'do rightly and leave the results with God'. Weber contrasted such ethics with what he called 'ethics of responsibility', which hold their adherents to responsibility for the foreseeable consequences, both good and evil, of their actions. He rejected political moralists who obscured the tragic character of politics through wishful thinking, such as the belief that only good comes from good and only evil from evil.

Weber's characterization of the modern situation, and his explanation of it, are the subject of continuing controversies of interpretation. Weber posed the problem of explaining the distinctive 'rationality' of Western society, and particularly the dynamic process of continuing rationalization. The most famous portion of Weber's account of this process takes the form of an ideal-typical construction of the genetic development of the psychological effect of particular religious ideas. Weber argued that the Calvinist doctrine of predestination was a source of deep anxiety to believers, who sought signs that they were among the elect. He argued that the pastoral practice of responding to this anxiety through the idea that diligence in the pursuit of worldly callings and the living of a self-disciplined life were signs of election, channelled this psychological force in a direction that produced a new type of person. The presence of this new type - rationalistic, individualistic, self-examining, conscience-driven, and work-sanctifying - facilitated the process by which capitalist economic conduct and the rational organization of work displaced traditional modes of economic conduct. In the later stages of this development, however, the routines of capitalistic rationalization become inescapable, and these routines undermined religious belief. In response, religion became more 'other-worldly', and culture became the increasingly senseless pursuit of mutually antagonistic self-chosen ends in a world that had itself become 'disenchanted' or 'meaningless'.

Weber's Protestant ethic essay could be construed as an alternative to the Marxian account of the rise of capitalism. The idea that religious ideas were an extra-economic cause of economic development is in conflict with Marx's *Überbau* or superstructure theory of ideology. Weber greatly expanded on his insight that different

religious traditions create characteristically different ‘economic ethics’ or modes of thinking morally about the economic world. Most of these modes precluded the development from within the society of modern capitalism of the Western kind, which Weber took to be based on the rational organization of labour. But Weber did not address Marx’s history of capitalism directly until his late lectures on economic history, in which he provided an overall account of the history of capitalism that incorporates the few elements of Marx’s account of the development of factory production that Weber accepted. The differences in their views of this topic are characteristic: where Marx saw the organization of work into factories simply as a means of increasing exploitation, Weber argued that production was itself rationalized, and that factory work required the new type of worker that the rationalizing discipline of Protestantism was producing.

In the Weimar era, during which Marx was taken more seriously as a thinker, interpreters such as Karl Löwith, compared the two more systematically. Weber came to be regarded as the bourgeois alternative to Marx, and former members of Weber’s circle, such as Georg Lukács and Karl Jaspers, also construed Weber in this way, each for their own purposes. Jaspers made Weber into a heroic figure whose radical liberal leadership Germany had tragically failed to accept. In a work by Lukács designed to discern and root out the sources of his own ideological deviations from Communist orthodoxy, Weber was classed as a powerful exponent of irrationalism, to which dialectical materialism was the alternative. Weber later became both subject and hidden interlocutor for the critical theorists of the Frankfurt School, and as they and their successors became concerned with the problem of modernity, Weber came to be regarded as a major theorist of modernity (see Frankfurt School).

STEPHEN P. TURNER
REGIS A. FACTOR

List of works

Weber, M. (1903-5) *Roscher und Knies: The Logical Problems of Historical Economics*, trans. G. Oakes, New York: Free Press, 1975.(Contains a useful introductory essay by G. Oakes.)

Weber, M. (1904-5) *The Protestant Ethic and the Spirit of Capitalism*, trans. and ed. T. Parsons, New York: Charles Scribner’s Sons, 1958.(Presents Weber’s account of the role of the Protestant theology of salvation in producing the rationalistic individuals who could bring about the capitalistic rationalization of labour.)

Weber, M. (1904-17) *The Methodology of the Social Sciences*, trans. and eds E.A. Shils and H.A. Finch, New York: Free Press, 1949. (Includes Weber’s three most important methodological essays, two of which focus on causality.)

Weber, M. (1907) *Critique of Stammler*, trans. and ed. G. Oakes, New York: Free Press, 1977.(Contains a valuable introductory essay by G. Oakes.)

Weber, M. (1913) ‘Some Categories of Interpretive Sociology’, trans. E. Graber, *The Sociological Quarterly* 22: 151-80, 1981.(Formulates an individualist analysis of collective social phenomenon within the framework of an account of action explanation.)

Weber, M. (1916a) *The Religion of China: Confucianism and Taoism*, trans. and ed. H.H. Gerth, Glencoe, IL: Free Press, 1951.(Companion study to the *Protestant Ethic* and part of a larger project on the economic ethics of the world religions; deals with Chinese religion and capitalism, and especially with the inhibitory effects on capitalism of Confucianism.)

Weber, M. (1916b) *The Religion of India: The Sociology of Hinduism and Buddhism*, trans. and ed. H.H. Gerth and D. Martindale, Glencoe, IL: Free Press, 1958.(Similar study of Indian religion as part of same project as 1916a.)

Weber, M. (1917-18) *Ancient Judaism*, trans. and ed. H.H. Gerth and D. Martindale, Glencoe, IL: Free Press, 1952.(Part of same 1916a project, but important for other themes, such as the role of the prophets in rationalizing theodicy.)

Weber, M. (1922a) *Wirtschaft und Gesellschaft* (*Economy and Society*), trans. G. Roth and C. Wittich, Berkeley, CA: University of California Press, 1978.(Detailed examination of the basic concepts of social sciences and economics, as applied to history.)

Weber, M. (1922b) *Gesammelte Aufsätze zur Wissenschaftslehre (Collected Essays in Scientific Theory)*, Tübingen: Mohr, expanded edn, 1951.(This text is a collection of essays, most of which have been translated into English, but not in one place. Many of the other works cited in this entry contain parts of this text, some include other texts as well.)

Weber, M. (1923) *General Economic History*, trans. F.H. Knight, Glencoe, IL: Free Press, 1950.(Posthumously published lecture notes providing a summary of Weber’s account of the origin and history of capitalism from

the ancient world to modern Europe.)

Weber, M. (1984-) *Max Weber-Gesamtausgabe*, H. Baier, M.R. Lepsius, W.J. Mommsen, W. Schluchter and J. Winckelmann (eds), Tübingen: J.C.B. Mohr (Paul Siebeck).(Weber's works, presently being published in a full scholarly edition.)

Weber, M. (1994) *Weber: Political Writings*, P. Lassman and R. Speirs (eds), Cambridge: Cambridge University Press.(Includes major political writings, most of which apply themes of his scientific writing.)

References and further reading

Bruun, H.H. (1972) *Science, Values and Politics in Max Weber's Methodology*, Copenhagen: Munksgaard.(A comprehensive discussion of Weber's use of ethical terms.)

Gerth, H.H. and Mills, C.W. (1946) *From Max Weber: Essays in Sociology*, New York: Oxford University Press.(The standard translation of Weber's major essays.)

Mommsen, W. (1959) *Max Weber und die deutsche Politik, 1890-1920* (*Max Weber and German Politics, 1890-1920*), trans. M.S. Steinberg, Chicago, IL: University of Chicago Press, 1984.(Historical overview of Weber's political activities and writings.)

Oakes, G. (1988) *Weber and Rickert: Concept Formation in the Cultural Sciences*, Cambridge, MA: MIT Press.(An analysis of the relation between the Neo-Kantianism of Rickert and Weber.)

Turner, S.P. (1986) *The Search for a Methodology of Social Science: Durkheim, Weber and the Nineteenth Century Problem of Cause, Probability, and Action*, Dordrecht: Reidel.(Contains an explication of Weber's concept of probabilistic causality and its epistemic presuppositions.)

Wegener, W. (1962) *Die Quellen der Wissenschaftsauffassung. Max Weber und die Problematik der Werturteilsfreiheit der Nationalökonomie: ein wissenschaftssoziologischer Beitrag* (The Sources of Scientific Views. The Set of Problems Relating to Freeing the National Economy from Value Judgments: a Sociological Contribution), Berlin: Duncker & Humblot. (Contends, following a long tradition of German criticism, that Weber's social science is reflexively inconsistent.)

Weil, Simone (1909-43)

Simone Weil's life and work represent an unusual mixture of political activism, religious mysticism and intense speculative work on a wide range of topics, including epistemology, ethics and social theory. Much of her most important writing survives in fragmentary form, in notebooks published after her untimely death. Though Jewish by family, her attitude to Judaism was largely hostile; and despite a deep commitment in the later part of her life to Christian ideas and symbols, she consistently refused to be baptized. Her religious views are eclectic in many ways, drawing on Plato and on Hindu sources. In everything she wrote, she was preoccupied with the dehumanizing effects of economic unfreedom and the servile labour required by industrial capitalism; but this is only one instance, for her, of the experience of 'necessity' or 'gravity' that dominates material transactions. The essence of moral and spiritual action is the complete renunciation of any privileged position for an ego outside the world of 'necessity'. Such renunciation is the only escape from necessity, in fact: what she calls 'decreation' becomes our supremely creative act, since only in the ego's absence is love, or an apprehension of non-self-oriented goods, possible. Marx, Kant and the gospels are all in evidence here.

1 Being in the world

Born in Paris in 1909, Weil studied at the Lycée Henri IV and the École Normale Supérieure. Her career as a schoolteacher was interrupted by a year of manual work in a factory and a brief and disastrous spell as a volunteer in the Spanish Civil War. Ill-health obliged her to resign from full-time work, and, after living with her parents for a time (and working on the land in Southern France after the German occupation), she joined them in emigrating to New York in 1942. By the end of that year, however, she was in London, working for the Free French. Overwork, a lung infection and the refusal to eat more than the current level of rationing in France led to a collapse in the spring of 1943. In hospital, she refused food and resisted medical help, and, because of this, her death in August 1943 was recorded as suicide. Some of the most significant of her notebooks are from the period in the South of France (1940-2) and the months in New York and London: they show very clearly the effects upon her thinking of a profound experience of religious conversion in 1938, but continue to develop themes that she had begun to explore almost a decade earlier.

Her earliest published work, from her time at the École Normale, was much influenced by her lycée teacher Alain (Émile Chartier), an eccentric and charismatic thinker who stressed the radical freedom of the will and the intimate connection of will and intelligence: mental action is an indivisible whole, a moral self-location in response to the experience of the body's life in the world. Weil's early work applies this perspective to the question of labour: work is the antithesis to the immediacy of thought, because it demands that we engage in actions we do not want to perform in order to realize the goal we *do* want. Such actions have no intrinsic relation to the mind's movement towards its goal; but the pure freedom of thought and will has to be activated in the concrete world by mediation and indirectness. Work is thus a paradigm of all she would later designate as necessity - that which imposes itself on mind. Yet without the primitive mental act meeting frustration in this way, there would be no *conceptuality* of the world as extended in time and space, no way of talking about objects or about duration (what intervenes between desire and realization).

We form concepts as we form 'strategies' for getting what we want. Our concepts are neither Platonic givens nor abstractions from sensation, but more like compact sets of instructions for negotiating obstacles. They are - as Weil's early *Lectures on Philosophy* put it - the product of 'a sort of dance', the body learning its way around in a world where realization is *not* contemporary with desire. And what we call 'reality' is thus neither purely given nor self-constructed but, as she says in a late fragment, 'a certain relationship to what is given'. This relationship is charged with ambiguity. At the simplest level, the body learns to interpret situations in a way that is more or less instantaneous: it 'reads' what has to be done (as promptly as we interpret marks on paper as words). But reading becomes more complex as we reflect on it, and the imagination intrudes - a faculty which for Weil is always morally shady. Imagination pulls apart the direct relationship of embodied thought to the surrounding obstacles, denying the character of these obstacles, seeking to avoid the suffering of real work, offering *ersatz* gratifications for our desires. This is specially acute in our relations to other subjects: slavery is the imposition of my (imaginative) reading of another person upon their reality (their actual relation to me). So being in the world as a moral subject, wanting the truth, wanting to live in appropriate relation to the given, requires a fierce assault on the

imagination, so that we may be returned to a direct confrontation with necessity.

2 Decreation and grace

There are two implications to this programme. First, all our *specific* goals or desires are likely to be flawed, because they invite us to form images of what we want, and so to evade necessity. Second, the root of corruption in our seeing and acting is the desire not to lose control, the longing to pretend that the immediacy of thought can be translated into our negotiations with what is given. The moral and spiritual task is to purify our desire by the rigorous renunciation of the ego, its position, its ‘rights’, its account of its individual wants. Pure desire is absolute consent to what is given, a submission to the otherness of the world that abandons any claim to a place, over against the world, from which I can manipulate the given. In the most radical expressions of this in the late notebooks, Weil will speak of the need for the ‘I’ to disappear: ‘To say "I" is to lie’, to defend an illusory self against necessity. The residual Cartesian and Kantian elements in her earlier thought are here subjected to their most ruthless purgation - though not without leaving strong traces. Here too her religious vision comes most clearly in view. The total acceptance of the given, the ‘desire’ for reality, *is* the desire for God, the only proper object of desire, since God is no particular thing. The Christian narrative shows us a God who wills the universe into being, a universe that is devoid of God, in the sense that God is not identical with any object or state of affairs in it. Thus God’s supreme liberty is expressed in creation as consent to an order of God-less necessity. Creation is divine renunciation; and the further narrative of the incarnation shows God becoming present in the Godless world in the only way possible - as a person without power or rights, becoming finally a dead body on the cross and a passive, mute object in the bread of the Mass. The Christian God is thus absent from the world but also, consequently, ‘present’ in the very character of the world as God-forsaken.

To accept the world is to love God. But the introduction of God into the discourse here has its paradoxical side. While we can deploy the theological narrative of creation and incarnation, we have no possibility of constructing a religious metaphysic in the usual sense. As soon as we conceptualize and represent God and an existing reality over against us, we lose sight of the ‘real’ God; and, on the basis of what is said elsewhere about reality, we should conclude that the reality of God is a particular relation between the self and what is given *in its entirety*. Weil’s kinship is with the tradition, inside and outside Christianity, of regarding with suspicion any suggestion that God is a kind of object among other possible objects of experience. She is not straightforwardly aligned with those radical theologians who deny the extra-mental reality of God.

When the ego is displaced in this complete acceptance, what follows is ‘grace’ or ‘the supernatural’, a relation to persons and things devoid of private interest and desire. Only this can properly be called love. The connection between such love and any particular person or object is complex, and Weil’s interpreters are not at one in analysing it. On the one hand, the sheer otherness of the world in its concrete variety is precisely what resists and checks my ego; and I begin to learn something of grace by submission to this concrete otherness, by what Weil calls ‘attention’. One of her most deservedly famous essays (‘Reflections on the Right Use of School Studies with a View to the Love of God’, written in 1942 and published in *Attente de Dieu* (1950a)) elaborates the spiritual significance of intellectual discipline. But, on the other hand, love is not properly bestowed on any individual object in its historicity, because this would be to make it serve transient needs or projections of the ego. Only what is timeless can be loved without corruption. There is no easy resolution of this tension, which points up central difficulties in Weil’s conception of the self: the self’s material and historical identity is both axiomatic and the source of ‘infinite error’.

3 Violence and justice

This creates some problems in her social thinking. Much of what she writes suggests that social life and cooperative labour is always ‘primitively’ a common sacrifice of individual desires before the given imperatives of collaboration. The struggle for social dominance is the struggle for the position of least sacrifice. She writes with great eloquence about the distortions of awareness produced by the exercise of force, both in the powerful and the powerless, and she is deeply sceptical about prevailing models of political legitimacy, resting as they do upon uneasy truces between different centres of power and interest. In her very substantial late (and unfinished) essay *L’enracinement* (1949) (*The Need for Roots* (1952)), designed in part as a vision for post-war France, she pleads for a society in which the impersonal instrumentalism of labour in a capitalist economy is replaced by work that can more readily be grasped as direct confrontation with necessity. This implies a radical dismantling of

international capitalist structures, including colonialism, and a reorientation towards local (regional rather than national) sociopolitical units. It also means the dismantling of 'modern' political structures, such as political parties and trade unions, since these take for granted the 'violence' of competitive interests. Justice will emerge not from bargaining but from the common reality of encounter with necessity; and this spiritual anchor for justice will be nourished by the corporate appropriation of a local past, a specific and harmonious culture. Weil did not shrink from the further implication that there could be no long-term future in such a society for religious or cultural minorities: they would be gently but firmly educated out of existence. Nothing could more clearly underline her own passionate rejection of her Jewishness, her own 'roots'. It is a final, and revealing, paradox in an *oeuvre* full of contradictions, yet so often profound in its critiques of modernity, of moral illusion and of cheap religious consolation. Only gradually has she come to attract serious philosophical study, rather than either mere abuse or hagiography; but she deserves just the unsparing 'attention' she herself so forcefully commends.

ROWAN WILLIAMS

List of works

Weil, S. (1988-) *Oeuvres complètes*, Paris: Gallimard.(The collected works.)

Weil, S. (1947) *La Pesanteur et la grace*, Paris: Plon; trans. E. Crauford, *Gravity and Grace*, London: Routledge, 1952.(Selected aphorisms from the *Cahiers*, often, regrettably, quoted or discussed without reference to their context in these notebooks.)

Weil, S. (1949) *L'enracinement*, Paris: Gallimard; trans. A. Wills with a preface by T.S. Eliot, *The Need for Roots*, London: Routledge, 1952.(Written in 1943, shortly before her death, this was intended both as an examination of the causes of alienation (especially among workers) in modern society and also as a blueprint for a post-war future in which individuals might enjoy a fruitful, rooted relation with society as a whole.)

Weil, S. (1950a) *Attente de Dieu*, Paris: La Colombe; trans. E. Crauford, *Waiting on God*, London: Routledge, 1951.(A collection of essays and meditations, largely but not exclusively religious. It has been the most widely read introduction to Weil's thought.)

Weil, S. (1950b) *La Connaissance surnaturelle*, Paris: Gallimard.(Contains Weil's *Notebooks* written in New York in 1942.)

Weil, S. (1951a) *La Condition Ouvrière*, Paris: Gallimard.(A painstaking diary of the time she spent labouring in factories (Alsthom and Renault) in 1934-5, with a series of texts reflecting on the experience in political and moral terms.)

Weil, S. (1951b) *Intuitions pré-chrétiennes*, Paris: La Colombe.(Material in Greek philosophy and its relation to Christianity.)

Weil, S. (1951c) *Lettre à un religieux*, Paris: Gallimard; trans. A. Wills, *Letter to a Priest*, London: Routledge, 1953.(A long letter addressed to a Dominican, le Pére Couturier, outlining her doubts about, and spiritual affinities with, Catholicism. Written in the autumn of 1942, ten months before her death.)

Weil, S. (1951d) *Cahiers*, Paris: Plon; trans. A. Wills, *The Notebooks of Simone Weil*, London: Routledge, 1956; also trans. R. Rees, *First and Last Notebooks*, London: Oxford University Press, 1970.(Fundamental for her developed religious metaphysics, though often very difficult for the interpreter.)

Weil, S. (1953) *La source grecque*, Paris: Gallimard.(Essays on Greek literature, also on Plato.)

Weil, S. (1955) *Oppression et liberté*, Paris: Gallimard; trans. A. Wills and J. Petrie, *Oppression and Liberty*, London: Routledge, 1958.(Contains the essay 'Reflections Concerning the Causes of Liberty and Social Oppression'. This considers the manner in which our social development has put limitations upon, rather than enhancing, human liberty.)

Weil, S. (1957) *Intimations of Christianity among the Ancient Greeks*, chapters from 1951b and 1953, trans. E. Geissbuhler, London: Routledge.(Finished reviews, and the texts of talks given in the Dominican Monastery at Marseilles. Discusses Greek philosophy as a harbinger of the Christian message.)

Weil, S. (1959) *Leçons de philosophie, Roanne 1933-4*, Paris: Plon; trans. H. Price with an introduction by R. Rees, *Lectures on Philosophy*, Cambridge: Cambridge University Press, 1978.(Provides insights into her earliest philosophical positions.)

Weil, S. (1962) *Selected Essays: 1934-1943*, trans. R. Rees, Oxford: Clarendon Press.(Mostly political essays.)

Weil, S. (1966) *Sur la Science*, Paris: Gallimard.(Early work on Descartes, and later essays on science.)

Weil, S. (1970) *Science, Necessity, and the Love of God*, trans. R. Rees, London: Oxford University Press.(Essays on science and on Plato, from the 1953 work.)

Further reading

There are several other collections and anthologies; see Bell (1993) for a good list.

Bell, R.H. (ed.) (1993) *Simone Weil's Philosophy of Culture: Readings Toward a Divine Humanity*, Cambridge: Cambridge University Press.(Essays on a wide range of aspects of Weil's thought.)

Cabaud, J. (1965) *Simone Weil: A Fellowship in Love*, New York: Harvill.(A fundamental survey of life and work together.)

McLellan, D. (1989) *Simone Weil: Utopian Pessimist*, London: Macmillan.(Oriented towards Weil's social thought.)

Petrement, S. (1976) *Simone Weil, A Life*, trans. R. Rosenthal, New York: Pantheon, and London/Oxford: Mowbray.(The most comprehensive biography, by a former fellow student of Weil's.)

Vetö, M. (1994) *The Religious Metaphysics of Simone Weil*, Albany, NY: State University of New York Press.

Winch, P. (1989) *Simone Weil: 'The Just Balance'*, Cambridge: Cambridge University Press.(Incomparably the best philosophical study, drawing out parallels with Wittgenstein.)

Weinberger, Ota (1919-)

Weinberger is noted as a proponent of 'institutionalist positivism' in legal theory. By contrast with earlier forms of so-called 'institutionalism' in law, Weinberger advances a theory in which norms are ideal entities linked by logical relations inter se*, while being at the same time social realities identifiable in terms of the effect they exercise in guiding human social behaviour. The institutions which make possible this duality of ideal entity and social reality have themselves to be understood as structured by norms. Hence, in contrast with earlier proponents of institutionalism, who denied the foundation of law in norms, Weinberger is normativist in his approach; and for the metaphysical vitalism of precursors, he substitutes a social realism.*

Born in Brno, Czechoslovakia, Weinberger trained in law and philosophy in Brno under František Weyr in the 'Brno School' of the 'Pure Theory of Law' (see Kelsen, H.), developing an early interest in the logic of norms. Imprisoned in a wartime concentration camp, he was able to resume his career as a scholar after 1945 in the difficult circumstances of post-war Czechoslovakia. During the 'Prague Spring' of 1968, he was able to leave for Austria, where he became Professor in Philosophy of Law in Graz, continuing actively as professor emeritus after retirement in 1988.

While rejecting ethical cognitivism, and with it natural law, Weinberger sees his position as promising some *rapprochement* between positivism and natural law theory (see Legal positivism §5; Natural law §1). For he acknowledges that concerns with justice are inherent in attitudes to law. He also fully acknowledges a place for reason - for rational analysis - in practical matters, but denies that reason ever finally determines the right, this being dependent on a taking-up of practical attitudes in a non-rational or pre-rational way. Reason and justice have a critical role to play in reflection upon action. But justice is never itself a primary motive to action; rather, it sets critical constraints on conduct whose motivation is fundamentally utilitarian (see Justice, equity and law).

Weinberger's early work was in spheres of logic which eluded even a stringent communist censorship, and already in his pre-Graz period he made significant contributions to the study of norm-logic, a field still in its infancy. He sees two spheres for this, one concerning the logical relations of established norms, the other concerning their establishment, in particular the question what logical account can be given of the way in which a power-exercising act under a norm of competence is to be understood as validly establishing a lower-order norm (see Norms, legal). In parallel to his differentiation of justice and utility, he has distinguished the logic of action from norm-logic, producing what he calls a 'formal-finalistic' or 'formal-teleological' theory of action, in which action is seen as activity guided by a processing of information. In this context, he insists upon the ultimate non-derivability of practical from theoretical information. This is his version of the non-derivability of 'ought' from 'is'. He rejects John Searle's purported derivation, reflection on which contributed greatly to development of his own conception of institutions such as promising or law-making.

See also: Institutionalism in law; Law, philosophy of

NEIL MacCORMICK

List of works

Weinberger, O. with MacCormick, N. (1986) *An Institutional Theory of Law*, Dordrecht: Reidel.(A collection of essays by two authors, with a joint introductory essay, explaining the development and implications of an 'institutionalist' theory of law, and its relationship to legal positivism.)

Weinberger, O. (1991) *Law, Institution, and Legal Politics*, Dordrecht: Kluwer.(A selection of the author's most definitive papers on fundamental problems of legal theory and social philosophy, including also important essays on the logic of norms and the theory of action.)

References and further reading

Koller, P. *et al.* (1994) *Institution und Recht: Grazer Internationales Symposion zu Ehren von Ota Weinberger (Institution and Law: an International Symposium at Graz in Honour of Ota Weinberger)*, Berlin: Duncker und Humblot.(This volume, which appears as Beiheft 14 of the journal *Rechtstheorie*, contains a series of tributes to and arguments with Ota Weinberger by a significant group of colleagues and friendly critics; it also contains a complete bibliography of Weinberger's vast output of writings in languages other than in English, and of the

smaller body of his work, including journal articles, published in English.)

Welfare

Notions of welfare occur widely in political philosophy and political argument. For example, utilitarianism is a social ethic that may be interpreted as giving a pre-eminent place to the idea that the welfare of society should be the overriding goal of public policy. Discussion of the ethics of redistribution focuses upon the institutions and practices of the so-called welfare state. Even those not convinced that we can validly speak of animal rights will often accept that considerations of animal welfare should play a part in legislation and morals. Moreover, the concept of welfare is clearly related to, and indeed overlaps with, concepts like 'needs' or 'interests', which are also central to public decision making and action.

Welfare can be thought of in three ways. Firstly, there is a subjective sense, in which to say that something contributes to a person's welfare is to say that it makes for the satisfaction of a preference. However, people can adapt their preferences to their circumstances, and happy slaves might be better off changing their preferences than having them satisfied. This thought leads on to the second sense of welfare as doing well according to some objective measure, like the possession of property. However, this conception can ignore subjective differences between people and fail to account for their capacity to take advantage of their objective circumstances. Hence, a third conception of welfare would make the capacity to take advantage of one's possessions an essential element of welfare. A satisfactory overall conception will have to bring these ideas together.

1 Welfare as preference

A common, perhaps the predominant modern, approach to the idea of welfare is to see it as the satisfaction of individual preferences or wants. This idea derives from the Benthamite felicific calculus (Bentham 1789) as mediated through modern welfare economics and the idea of consumers' utility. In this conception, persons are regarded as having preferences over various commodities (apples versus oranges) or states of the world (a world with or without the whale), and peoples' welfare, or utility, depends upon the extent to which their most highly ranked preferences are satisfied.

The notion of preference in this conception hangs uneasily between the notion of wanting and that of choice. Preferring commodity A to commodity B can seem to be like wanting or desiring to possess A rather than B, where the want is either a mental event or a disposition. On the other hand, some accounts of preference identify the notion simply with choice, asserting that all that is involved in preferring commodity A to commodity B is that one would choose to have A rather than B, without any assumptions being made that a choice reflected any inner feelings.

Whichever of these two views of preference is chosen, the main thrust of the preference-based conception of welfare is that the welfare of a person is to be judged from their point of view. This conception is particularly suited to those political theories that stress the diversity of individuals and societies, thereby implying that it is impossible to construct generalizations of what makes life go well for people.

Despite its appeal, the preference-based conception of welfare has difficulties. Some have urged that there must be limits on the desires or wants that we can sensibly ascribe to individuals, so that we should simply find unintelligible, not just unusual, someone who said that they wanted, say, a saucer of mud, without specifying any recognizable purpose for it. From this point of view the subjectivism implicit in the preference-based conception of welfare would be bounded by an understanding of desires that it would be intelligible to ascribe to individuals.

However, even within these limits, there are problems with treating welfare as preference. Slaves and harassed tenants may find that they adapt their preferences to their lot, coming to believe that it is in the nature of things. But if, after a while, such persons find that they prefer secure bondage to previously yearned for freedom, has their welfare increased? In terms of the preference-based account it has, but to many this has seemed counterintuitive. Limiting one's wants in this way seems instead to be the model of a frustrated life.

For these sorts of reasons theorists who favour a preference approach have sought to modify its simple form to take into account not raw preferences but informed ones. If the previously resigned come to see that things could be other than they are, then their preferences may change. Taken to the limit, the thought is that it is only the preferences that persons have in the full knowledge of their circumstances and the possibilities open to them that

could form the basis for a judgement about their welfare. If I know all there is to know about a commodity or a state of the world and I still choose it in preference to something else, then we might want to say that I have thereby established what makes for my welfare.

One advantage of this approach is that it seems to offer a coherent analysis of the idea of human dignity. For example, when J.S. Mill (§9) argues that it is better to be Socrates dissatisfied than a fool satisfied, because the fool knows only one side of the story but the dissatisfied knows both, he is implicitly appealing to a hypothetical choice or preference criterion, in order to make the idea of welfare consistent with some idea of human dignity.

Despite the appeal of the account, it raises a number of questions. To say that someone currently prefers A to B, in their present state of knowledge, but would prefer B to A if they were fully informed invites the question of how we might know what the hypothetical preference was. In the absence of any theory about how people choose, it is impossible to say. Moreover, if there were such a theory and if it were based on accounts of how in fact the bulk of persons do choose, it looks as though the account is merely slipping into an objectivist account of welfare reflecting views about human nature and what makes things go well for people in general (see Needs and interests).

A final difficulty with the preference-based account, in either its raw or informed preference view, is the difficulty of making interpersonal comparisons of welfare. The problem here is that interpersonal comparisons seem plausible (we think we know that the prince is better off than the pauper), but it is difficult to see how such judgements could be intellectually justified (see Economics and ethics; Utilitarianism).

Within the preference-based account these difficulties appear to be insurmountable. If the welfare of persons is simply constituted by their preferences over alternatives, what evidence could be employed to make comparisons? Ingenious attempts have been made to employ the idea of extended sympathy in this context, in which a judgement of interpersonal comparisons is made by someone saying that they would rather be person A, with all that A has, than person B, with all that B has, but it has been pointed out that this merely shows that people think they can say what they find attractive or otherwise in the position of others, not what is comparatively of value in the position of others.

2 Objective accounts of welfare

Objective accounts of welfare appeal to the thought that there are features of the circumstances, position or characteristics of persons that enable us to judge how well off they are.

There are in principle two forms of objective account of welfare; the criterial and the indexical. Criterial accounts state that welfare is made up of those things that are necessary to achieve some end state or goal, such as human flourishing or agency. To enjoy a certain level of welfare, on such accounts, is simply to be in possession of just those goods that make the achievement of the goal or end state possible. Special mention should be made in this context of Rawls' notion of 'primary goods' (Rawls 1971) (see Rawls, J. §1). These are defined as those things it would be rational to want if one were to want anything at all. Although this seems like a subjectivist notion, it is more plausibly construed as objectivist, since the wanting is grounded in external reasons, and the criterion is normally interpreted by appeal to standard empirical evidence about what is necessary to lead a life in society.

Indexical accounts simply list goods thought to be components of welfare, such as health, possessions and good fortune. In practice these two distinct logical conceptions, the criterial and the indexical, turn out to have similar implications, since the criterion of welfare will typically end in a list of some sort. Thus, Rawls' primary goods are listed as income and wealth, opportunities and powers and the bases of self-respect, and subsequent commentators (for example, Pogge (1989)) simply cash them out as familiar lists of rights.

One appeal of the objective approach is that it seems consistent with much that is ordinarily said and thought about welfare. It seems to make sense to say that we are well off when our health is good, our careers go well, our possessions and prospects are secure, our political and civil rights protected and our personal relationships flourishing. Moreover, such an objective approach would solve the problems associated with the interpersonal comparison of welfare. Although it might be difficult to combine the various dimensions of wellbeing into a single index, we could at least measure how well things were going for people by reference to such criteria as have been listed. Indeed, there is a branch of social science - the construction of social indicators - that is devoted to

measuring the welfare of people in society by reference to such criteria.

At this point the subjectivist will object and say that just as possessions do not make for happiness, so doing well on some objective list does not make for welfare. There are a number of reasons why this might be so. Certain spiritual aspirations may mean that people have to forgo their possessions in order to discover their soul, and Brentano is said to have treated the blindness of his old age as a blessing because it meant that he concentrated on his philosophy. Moreover, how in the absence of an appeal to subjective assessment do we evaluate conflicts among goods on the list? For example, certain risky occupations involve a trade-off between income and the chance of accidents leading to ill health. How people balance the conflict between income and additional health risks seems properly a matter for their own judgement, for which the most obvious rationalization is the thought that the assessment of welfare must be done from the point of view of the person experiencing it.

Yet these anti-objectivist arguments are not without problems. Cases of giving up one's possessions to save one's soul or the example of Brentano's blindness, although revealing, may be treated as special in the sense that they give extended meaning to a concept but do not affect its central application. The argument from the trade-off of benefits may be reinterpreted to yield objectivist conclusions, if we admit that it makes sense to have an external evaluation of the choices that people make. It seems entirely plausible to think that we can know in some cases that people make the wrong choice, from the point of view of their own welfare, between income and health, and we may also be critical of social arrangements that give the poor hard choices between earning a living and avoiding an unsafe working environment. All such judgements presuppose a certain objective conception of welfare.

The appeal to diversity does have a point, however. Suppose we observe a community of persons with roughly similar material circumstances and similar educational opportunities pursuing different occupations, different lifestyles and taking different attitudes to risk. Each person may enjoy quite distinct bundles of possessions and characteristics. Yet we may judge the members of the community to have similar levels of welfare, in the sense that each person has used similar resources to different effect according to their own conception of what is good for them. This line of argument might suggest that we can reconcile objective and subjective accounts of welfare provided we take into account the ability of people to use the resources at their disposal. This thought brings us to the third conception of welfare, namely that of welfare as capabilities.

3 Welfare as capabilities

Amartya Sen has argued that to treat welfare simply as a function of the commodities that people have is misleading, since it ignores the capacity or capability that people have to use those commodities to their own advantage (Nussbaum and Sen 1993). An obvious example would be the contrast between the physically handicapped and the able-bodied. The former may need more resources to bring them up to the same standard of living as the latter. The implication of this view is that we need to take into account the ability of people to use their objectively given level of resources to achieve welfare in their own terms.

Sen suggests that we need to base our conception of welfare on capabilities, rather than commodities, in order to do justice to this insight. However, it is doubtful if this implication can be drawn. Capabilities are difficult to measure directly, and the way in which the idea has in practice been developed is in terms of social indicators of the rather conventional objectivist sort. However, the value of the idea of capabilities is that we need to take into account this dimension of welfare if we are to make adequate interpersonal comparisons.

4 Conclusion

Can we extract a coherent conception of welfare from these competing views? If we accept that human diversity means that people will find their welfare in diverse combinations of goods, then there has to be an irreducibly subjective element to our concept of welfare. One way to avoid the problems with such an account is to modify our assessment of peoples' choices and preferences not in terms of a hypothetical choice that they would make, but in terms of our knowledge of preferences generally exhibited by persons in circumstances we understand. Moreover, we should further modify this assessment by a calculation of how far people are able to take advantage of the resources at their disposal, which is in large part an empirical matter.

If we take the notion of welfare to have both these subjective and these objective elements, allowing the subjective to operate within the limits of our understanding of what in general makes life go well for people, then we are

faced with the problem of how broadly we are to construe the notion of welfare. In particular, is individual freedom or autonomy a component of welfare or a distinct characteristic of humans? There is probably only a conventional, rather than essentialist, answer to this question, but if it is sensible to ask how valuable it is for people to secure their autonomy, then it would seem to be more plausible to make freedom an element of welfare rather than distinct from it. As such, the concept of welfare will always remain central to questions of political philosophy.

See also: Happiness

ALBERT WEALE

References and further reading

Arrow, K.J. (1963) *Social Choice and Individual Values*, New Haven, CT: Yale University Press, 2nd edn.(States the idea that interpersonal comparisons in the preference view of welfare might be based on extended sympathy.)

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, ed. J. Burns and H.L.A. Hart, London: Athlone Press, 1970.(The best statement of the traditional utilitarian view.)

Doyal, L. and Gough, I. (1991) *A Theory of Human Need*, Basingstoke: Macmillan. (Provides a good foundation for the understanding of social indicators and their philosophical significance.)

Griffin, J. (1986) *Well-being*, Oxford: Clarendon Press.(A clear statement of the informed desire view of welfare.)

Hargreaves Heap, S. Hollis, M., Lyons, B., Sugden, R. and Weale, A. (1992) *The Theory of Choice*, Oxford: Blackwell.(An introduction to economic theories of welfare.)

Nussbaum, M. and Sen, A. (1993) *The Quality of Life*, Oxford: Clarendon Press. (Contains a set of essays exploring the view of welfare as capabilities, including a review essay by Sen.)

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press.(Has a valuable appendix on various theories of welfare.)

Plant, R. (1991) *Modern Political Thought*, Oxford: Blackwell.(Has a good discussion of various approaches to welfare, including their implications for social policy.)

Pogge, T. (1989) *Realizing Rawls*, Ithaca, NY: Cornell University Press.(Elaborates the implications of Rawls' notion of primary goods.)

Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.(States the idea of primary goods.)

Weyl, Hermann (1885-1955)

A leading mathematician of the twentieth century, Weyl made fundamental contributions to theoretical physics, to philosophy of mathematics, and to philosophy of science. Weyl wrote authoritative works on the theory of relativity and quantum mechanics, as well as a classic philosophical examination of mathematics and science. He was briefly a follower of Brouwer's intuitionism in philosophy of mathematics. Upon moving closer to Hilbert's finitism, he articulated a conception of mathematics and physics as related species of 'symbolic construction'.

Hermann Weyl was born on 9 November, 1885 in Elmshorn, near Hamburg, Germany. He received his doctorate in mathematics under the direction of David Hilbert at Göttingen University in 1908. From 1913 until 1930, when he returned to Göttingen as Hilbert's successor, Weyl was Professor of mathematics at the Federal Institute of Technology in Zurich. As head of the world-renowned Mathematical Institute in Göttingen when the Nazi regime came to power in 1933, Weyl's unsuccessful attempts to protect his colleagues from arbitrary dismissal prompted him to accept, later that year and at Einstein's urging, an appointment as Professor at the newly formed Institute for Advanced Study in Princeton. The recipient of many prizes and awards for his mathematical work, including the Lobachevskii prize in 1925 for his geometrical research, Weyl was elected a contributing member of the Royal Society of London in 1936. Upon his retirement from the Institute in 1951, Weyl divided his time between Princeton and Zurich where he died of a heart attack on 9 December 1955.

A mathematician of extraordinary depth and breadth, Weyl worked in nearly all parts of mathematics: analysis, geometry, number theory, algebra and mathematical logic. He also continued the Göttingen tradition of Felix Klein and David Hilbert of carrying out mathematical investigations of fundamental questions in theoretical physics. In the decade 1917-27 he was at the height of his powers, producing a succession of contributions of the highest order to mathematics and to physics. In 1918 he published *Raum-Zeit-Materie*, the first self-contained treatment of Einstein's new general theory of relativity, which Einstein termed 'a masterful symphony'. Running through five editions by 1923, the work was no mere exposition but considerably illuminated the mathematical and philosophical foundations of the theory. For example, Weyl was first to show that there is a natural geometric but non-metrical expression ('manifold with an affine connection') for the combined gravitational-inertial field, underscoring that in general relativity gravitation is but counterpart to inertia.

Still in 1918, Weyl proposed a generalization of the Riemannian geometry underlying the Einstein theory which contained an additional metrical term formally analogous to the tensor expression for the electromagnetic field. As gravitation and electromagnetism were then the only known physical forces, Weyl's geometry appeared to account for all known physical phenomena in terms of the metric of spacetime. Though his theory did not survive the advent of quantum mechanics in 1925-6, Weyl's approach considerably influenced Einstein's first attempts at a unified field theory, and its central principle, 'local gauge [scale] invariance', now reinterpreted in the context of quantum theory, remains a leading principle of the physics of elementary particles. As an outcome of his research on group representations in 1925-6, which Weyl himself viewed as his greatest mathematical achievement, he pioneered investigation of the connections between group theory and the new quantum mechanics in 1927.

Despite his then deep involvement with the theory of relativity, Weyl also published in 1918 *Das Kontinuum*, a controversial monograph on the foundations of analysis. In agreement with the diagnosis of Poincaré and Russell that the root of the crisis in the foundations of mathematics concerned the freedom in unrestricted set theory to form 'impredicative' sets or classes, Weyl sketched out an iterative method of set formation (essentially a predicative set theory) in whose terms some, but not all, of classical analysis might be reconstructed. The remainder, Weyl proclaimed, would simply have to be jettisoned as illicit extensions of mathematical thinking. Thus would be surrendered such a fundamental theorem as that every bounded set of real numbers contains a unique least upper bound. At the time predisposed toward Husserlian phenomenology, by 1920 Weyl had abandoned his own attempts to resolve the crisis in foundations and joined the Dutch mathematician Brouwer in intuitionism. Weyl proved to be a wayward disciple. While endorsing the conception of the continuum as an *open* and not a *completed* totality, Weyl differed fundamentally from Brouwer in linking the grounding of analysis (for example, the concepts of real number and function) with its applications, especially in physics.

In the late 1920s, Weyl argued that the interconnected character of justification of individual mathematical

assertions, as shown in Hilbert's finitism and paralleling the epistemological situation in physics where theories as a whole confront the evidence of experience, posed 'a decisive defeat for the program of pure phenomenology' in the foundations of mathematics. In later writings, Weyl emphasized the *essentially* symbolic character of mathematics and physics. As species of 'constructive cognition', both exhibit the freedom of mental operations to idealize from the given and to introduce symbols, which may then be freely manipulated according to certain rules. In this way symbols acquire a relative independence. Agreeing with idealism that scientific objectivity is not given but is constructed in symbols, Weyl nonetheless posited an unbridgeable chasm between the objective world of symbols and the ego, the 'light of consciousness'.

See also: General relativity, philosophical responses to; Hilbert's programme and formalism; Intuitionism; Phenomenological movement

T.A. RYCKMAN

List of works

Weyl, H. (1918a) *Raum-Zeit-Materie*, Berlin: Springer; 4th edn, 1921, trans. H. Brose, *Space-Time-Matter*, London: Methuen, 1922; repr. New York: Dover, 1952.(The third and later editions present Weyl's generalization of Reimannian geometry.)

Weyl, H (1918b) *Das Kontinuum; Kritische Untersuchungen ueber die Grundlagen der Analysis*, Leipzig: Veit; trans. S. Pollard and T. Bole, *The Continuum; A Critical Examination of the Foundation of Analysis*, Kirksville, MO: Thomas Jefferson University Press, 1987; repr. New York: Dover, 1994.(A controversial monograph on the foundations of analysis.)

Weyl, H (1927) *Philosophie der Mathematik und Naturwissenschaft*, Munich and Berlin: R. Oldenbourg; repr. *Philosophy of Mathematics and Natural Science*, Princeton, NJ: Princeton University Press, 1949.(The revised and augmented edition of 1949 was based on a translation by Olaf Helmer, and included material Weyl added in English. Chapter II 'Methodology' contains an account of 'constructive cognition'.)

Weyl, H (1928) *Gruppentheorie und Quantenmechanik*, Leipzig: S. Hirzel; trans. H.P. Robertson, *The Theory of Groups and Quantum Mechanics*, London: Methuen, 1931; repr. New York: Dover, 1950.(The 1931 translation is based on the second, 1931, German edition. Explores the connection between group theoretic symmetry principles and fundamental relationships in quantum mechanics.)

Weyl, H (1952) *Symmetry*, Princeton, NJ: Princeton University Press.(Lectures on the role of symmetry in mathematics, science and the arts.)

Weyl, H (1968) *Gesammelte Abhandlungen*, ed. K. Chandrasekharan, Berlin, Heidelberg and New York: Springer, 4 vols.(The collected scientific papers, many also of philosophical interest.)

References and further reading

Chandrasekharan, K. (ed.) (1986) *Hermann Weyl 1885-1985*, Centenary Lectures at the Eidgenössische Technische Hochschule in Zurich, Berlin, Heidelberg, and New York: Springer.(Appreciations by C.N. Yang, R. Penrose, and A. Borel of Weyl's contributions to physics, geometry and group theory.)

Deppert, W. and Hübner, K. (eds) (1988) *Exact Sciences and their Philosophical Foundations*, Proceedings of the International Hermann Weyl Congress, Kiel, 1985, Frankfurt, Bern: Peter Lang.(Papers on many aspects of Weyl's scientific and philosophical work.)

Feferman, S. (1988) 'Weyl Vindicated: ''Das Kontinuum'' 70 Years Later', Atti del Congresso *Temi e prospettive della logica e della filosofia della scienza contemporanee*, Cesena, 1987, Bologna: CLUEB, vol. I, 59-93.(A recasting, in modern terms, and a defence of the essentials of Weyl's monograph by a leading proof theorist.)

Newman, M.H.A. (1957) 'Hermann Weyl', *Biographical Memoirs of the Royal Society of London* 3: 305-28. (Obituary summarizing Weyl's scientific accomplishments.)

Weyr, František (1879-1951)

František (Franz) Weyr was Professor in Legal Philosophy and Public Law in Brno, Czechoslovakia, and a main author of the Czechoslovakian Constitution of 1920. His influence on Czechoslovakian jurisprudence was exceptional. He advocated the 'Pure Theory of Law', demanding that law be studied in a methodologically distinct way, pure of natural-scientific or ideological inputs. He was founder and leader of the 'Brno School' of Pure Theory (or 'Normative Theory', in his preferred terminology). This school stands close to the Vienna School of Hans Kelsen.

Weyr and Kelsen were also close in political stance. Both fought for democracy and against the Austrian monarchy, with its elements of a police state. Weyr's analysis of the relation between private and public law is particularly pro-democratic in tendency; according to him, procedures in public law no less than in private law are themselves bound by law.

In Weyr's conception, law belongs to a legal order which is: (1) a logically unitary system of norms ascribed to the state; (2) a dynamic system with hierarchical structure; (3) valid under a presupposed basic norm; (4) subject to cognition of its content understood as *sui generis* and studied in abstraction from social and evaluative considerations (thus jurisprudence as cognition differs from legal politics as volition); (5) identifiable through attribution to the state as their ideal subject. (Hence Weyr sharply rejects Kelsen's sanction theory of legal norms; Norms, legal §2.)

Weyr's philosophical background is Kantian, grounded in the opposition of 'is' and 'ought', cognition and volition. Thus his theory of law and of legal processes explains cognition of them in terms of their formal and general validity. He accepts Kant's theory of the nature of cognition so far as it stresses the distinction of 'is' and 'ought', but rejects Kant's commitment to a form of natural law theory (see Legal idealism §2; Natural law). The dualism of 'is' and 'ought' is not only semantic, but rooted in different rational relations in the respective fields. Weyr is a strict positivist and value-relativist, and founded legal methodology on the difference of argumentations about the existence of law and about its merit as such. The realm of jurisprudence is a realm clearly distinct from that of nature: in law (and morality) free action is essential, whereas nature is governed by causality. Latterly, he tended towards Karel Engliš' triadic view of the forms in which we apprehend reality, namely causality, teleology and normativity, but underlined the nearer relation between teleology and normativity. He articulated explicitly the problem of the specificity of norm-logical relations, and acknowledged that the logic of descriptive language is inapplicable to norms. He rejected the norm-logical scepticism of Engliš and Kelsen.

Elements of Weyr's theory that remain of importance are: the view that structure-theory is the basis of legal theory, and that it should be linked with an appropriate theory of legal argumentation; and the dynamic theory of law as foundational for investigations of legal processes and legality. Weyr's and Kelsen's teachings belong to the first phase of analytical jurisprudence, seeking a universal structure of all law. The present second phase looks for differentiations of legal structures, thus transcending the Brno and Vienna schools alike. The postulate of strict purity of jurisprudence should be given up; what survives is Weyr's postulate that methods of inquiry must not be mixed up, but must match objects of inquiry.

See also: Law, philosophy of; Legal positivism

OTA WEINBERGER

List of works

Weyr, F. (1908) 'Zum Problem eines einheitlichen Rechtssystems' ('On the Problem of a Unified System of Law'), *Archiv für öffentliches Recht* 23: 529-80.(An essay on the problem of the unitary character of a legal system.)

Weyr, F. (1914) 'Zum Unterschiede zwischen Öffentlichem und privatem Recht' ('On the Difference between Public and Private Law'), *Österreichische Zeitschrift für öffentliches Recht* 40: 439-41.(This essay discusses the distinction between private and public law, insisting that both are subject to the demands of the rule of law.)

Weyr, F. (1920) *Základy filosofie právní. Nauka o poznávání právnickém (Foundations of the Philosophy of Law:*

the Theory of Legal Cognition), Brno: A. Píša.
Weyr, F. (1922) *Československé právo správní. Část obecná (Czechoslovakian Administrative Law: the General Part)*, Brno: Právník.
Weyr, F. (1930) *Správní řád (The Administrative System)*, Brno: Barviča Novotný.
Weyr, F. (1936) *Teorie práva (The Theory of Law)*, Brno and Prague: Orbis.
Weyr, F. (1937) *Československé právo ústavní (Czechoslovakian Constitutional Law)*, Prague: Melantrich.

References and further reading

Kubeš, V. and Weinberger, O. (1980) *Die Brünner rechtstheoretische Schule (The Brno School of Legal Theory)*, Vienna: Manz Verlag.(Weyr’s work is almost inaccessible in English or even in French; but this collection of essays in German, by Czech authors, gives a rounded account of the work of Weyr and his school.)

Whewell, William (1794-1886)

William Whewell's two seminal works, History of the Inductive Science, from the Earliest to the Present Time (1837) and The Philosophy of the Inductive Sciences, Founded upon their History (1840), began a new era in the philosophy of science. Equally critical of the British 'sensationalist' school, which founded all knowledge on experience, and the German Idealists, who based science on a priori ideas, Whewell undertook to survey the history of all known sciences in search of a better explanation of scientific discovery. His conclusions were as bold as his undertaking. All real knowledge, he argued, is 'antithetical', requiring mutually irreducible, ever-present, and yet inseparable empirical and conceptual components. Scientific progress is achieved not by induction, or reading-out theories from previously collected data, but by the imaginative 'superinduction' of novel hypotheses upon known but seemingly unrelated facts. He thus broke radically with traditional inductivism - and for nearly a century was all but ignored. In the Philosophy the antithetical structure of scientific theories and the hypothetico-deductive account of scientific discovery form the basis for novel analyses of scientific and mathematical truth and scientific methodology, critiques of rival philosophies of science, and an account of the emergence and refinement of scientific ideas.

1 Life and works

Whewell was born on 25 May 1794, son of a master carpenter in Lancaster. He died on 6 March 1866, Master of Trinity College, Cambridge, and one of the most eminent and prolific individuals of the era. His works include scientific research, physics and mathematics textbooks, books on ethics and law, university education, natural theology, church architecture, scientific nomenclature, political economy and the history and philosophy of science. He wrote poetry, translated Greek and German classics and edited the works of Hugo Grotius and Isaac Barrow. He was an active member of more than a dozen learned societies, a force in the founding and running of the British Association for the Advancement of Science, an influential Master of Trinity (1841-66), and dominant Vice Chancellor of Cambridge University during two terms of office (1842 and 1856). Whewell, however, was not just a polymath. He wrote with the explicit view not merely of encompassing the world of early Victorian learning, but of synthesizing it and rendering it a unified intellectual whole.

Drawing on the *History* that endeavoured, for the first time ever, to chart the development of all known 'inductive sciences', the *Philosophy* proposes a novel theory and methodology of science that describe the nature, structure and aims of a well-formed natural science, explain how scientific discoveries are made and propose ways of retrospectively assessing their truth. The *Philosophy*, however, deviated too sharply from the prevailing empiricism, advocated, for example, by John Herschel and John Stuart Mill, to obtain acceptance in its time (see Mill, J.S. §5; Empiricism; Inductive inference §4). Only recently, especially in the light of Karl Popper's critique of inductivism, is Whewell's philosophy beginning to receive the hearing it deserves (see Popper, K. §§2-3).

2 Theory of science

Whewell's theory of science is presented in the *Philosophy* as an argument from a theory of the sources of knowledge to a detailed description of well-developed science. Somewhat systematized, it comprises the following four theses:

(1) *The Fundamental Antithesis*. Convinced that neither the empiricists, who sought to base all knowledge on experience alone, nor the intellectualists, who based science on a priori ideas, could adequately account for the development and accomplishments of science, Whewell proposed a bold alternative: all knowledge necessarily incorporates an objective element given by the world, informed and structured by conceptions furnished by the mind (which are refined during scientific research to describe the world as it is). Even the most commonplace perception must go beyond mere sensation. To perceive requires a singling out of sensation, a retention of it by the mind, and some awareness of its relations to other sensations past and present - none of which can be given by sensation alone. All knowledge thus embraces an antithesis of irreducible opposites: things as opposed to ideas, sensation as opposed to conception, fact as opposed to theory. And yet, Whewell notes, it is a 'Fundamental Antithesis of Philosophy', one that can only be reflected upon theoretically. In practice no item of real knowledge can be fully resolved into its primordial factual and conceptual ingredients; so Kant's conclusion is unavoidable: the purely substantial is unknowable in itself (see Kant, I. §5).

(2) *Theory of 'Induction'*. From the near-Kantian epistemology of the Fundamental Antithesis, Whewell goes on to argue for a theory of the acquisition of knowledge thus construed. He termed the process 'induction', but meant something quite different from all current uses of the word. Whewell envisaged induction as a corrective hermeneutic process of reading meaning, structure, regularity and law *into* the facts, rather than gleaning such information *from* the empirical data. If all items of real knowledge comprise facts, unconnected in themselves and 'colligated' by concepts furnished by the mind, it follows that new knowledge is obtained exactly when a new concept, suggested by the mind, is successfully 'superimposed' upon formerly known but seemingly unrelated facts. His language notwithstanding, Whewell was an outspoken anti-inductivist.

(3) *Theory of Excellent Science*. A well-formed science takes shape in the course of repeated 'inductions' of ever higher generality that yield a series of nested, hence deductively related, conceptual colligators. According to Whewell, the conceptual component of each such science is governed by a 'Fundamental Idea', such as that of space for geometry, force and matter for dynamics, or life for biology. As work progresses certain features of the science's Fundamental Idea gradually come to light and are formulated as axioms considered to be true *of* the Fundamental Idea. Ideally, a science will consist of an entire corpus of fact perfectly colligated by such a fully axiomatized conceptual scheme.

(4) *Theory of Truth*. Lastly, such a view of science inevitably yields a two-pronged theory of scientific truth. This is because the propositions comprising a scientific theory strive to attain two different, and *prima facie* independent forms of perfection: to be on the one hand well-formed theorems of a well-formed system of ideas, and at the same time perfect colligators of fact. In the role of colligator a proposition may at best be considered empirically true, while as a theorem of a well formed system of ideas, it may be considered necessarily true *of the science's Fundamental Idea* - acknowledged to the point that *within the system* one cannot distinctly intuit its negation. Whewell's contention that a proposition may thus be considered necessarily true (of a Fundamental Idea) and yet eventually prove empirically false, has baffled many of his critics. Truth within a conceptual scheme only implies that the proposition in question is conceived to be true of a Fundamental Idea (and hence must necessarily hold each time the Idea is applied to facts). But whether or not it is also true of the facts is forever contingent upon empirical test. Whewell's notion of truth within a conceptual scheme thus defies Kant's celebrated taxonomy. Since he insisted that a truth relation (other than mere coherence and resting on more than formal definitions) obtains within pure systems of ideas (as in pure mathematics) it cannot be considered merely analytic. On the other hand, although a priori with respect to the facts, such truths cannot be considered synthetically true prior to independent empirical test.

3 Methodology

Whewell's methodology falls under three major headings corresponding to the stages preceding, during and following scientific discovery. This part of his philosophy is largely anticipated in the *History*, in which major scientific breakthroughs are broken down into 'inductive epochs', their 'preludes' and their 'sequels'.

Whewell considered discovery itself reducible to no more than an imaginative 'happy guess', a 'felicitous and inexplicable stroke of inventive talent' defying methodological prescription. Like Popper almost a century later, Whewell called upon scientists to hypothesize freely, fancifully and frequently. Bacon's condemnation of 'anticipation', and Newton's '*Hypotheses non fingo*', he urged, expressed the wrong attitude. He refused, however, to view scientific conjecture as entirely capricious. By 'decomposing' the facts in hand as far as possible into their conceptual and factual ingredients, he argued, a part of the science's conceptual scheme, at least that to which the facts in hand directly pertain, may be (re)constructed. Such (partial) knowledge of a science's conceptual history can provide discoverers with the appropriate setting and idiom for their new conjectures.

But bold speculation will lead to real knowledge, he argued further, only if firmly coupled to an 'ingenuity and skill which devises means for rapidly testing false suppositions as they offer themselves'. The 'Popperian' ring is again unmistakable: the worth of a conjecture is measured not by reports of favourable data, but by its proven resistance to falsification.

Resilience in the face of prudent testing, however persistent, Whewell observed, cannot be considered conclusive proof of the actual (empirical) truth of a hypothesis. And yet he maintained that truth lies within the reach of science. True theories distinguish themselves from merely good ones on two related accounts. First in repeatedly

exhibiting what he termed ‘consilience of inductions’, or explanatory surprise. And second, in generating ‘progressive simplification’ as they are repeatedly applied to new and different kinds of phenomena. In Whewell’s opinion both gravitation theory and the undulatory theory of light had proved themselves true in this manner. Both criteria are as questionable (as criteria of truth) as they are intriguing (as criteria of acceptability), and have attracted considerable attention.

See also: Conventionalism; Discovery, logic of; Science, 19th century philosophy of §§3-6; Scientific method

MENACHEM FISCH

List of works

Whewell, W. (1833) *Astronomy and General Physics Considered with Reference to Natural Theology* (Bridgewater Treatise), London: William Pickering.(Whewell’s first major reflections on science. Contains an interesting chapter on the Inductive and Deductive ‘habits’.)

Whewell, W. (1835) *Thoughts on the Study of Mathematics as Part of a Liberal Education*, Cambridge: J.J. Deighton.(Mathematics presented for the first time by Whewell as governed by Fundamental Ideas of which its axioms are true.)

Whewell, W. (1837) *History of the Inductive Sciences, from the Earliest to the Present Time*, London: J.W. Parker, 3 vols.(Whewell’s history of science.)

Whewell, W. (1840) *The Philosophy of the Inductive Sciences, Founded upon their History*, London: J.W. Parker, 2 vols.(The third, enlarged edition of 1858-60 was republished by J.W. Parker, in three separate works: a conceptual history of the sciences, *The History of Scientific Ideas* (1858); a critical history of the philosophy of science, *On the Philosophy of Discovery* (1860); and Whewell’s own philosophical system, dubbed *Novum Organon Renovatum* (1858).)

Whewell, W. (1845) *The Elements of Morality, Including Polity*, London: J.W. Parker, 2 vols. (A system of ethics ranging from jurisprudence to international law, developed in close analogy to science and mathematics.)

Whewell, W. (1849) *Of Induction, with Special Reference to Mr J. Stuart Mill’s System of Logic*, London: J.W. Parker.(Whewell’s systematic reply to J.S. Mill’s criticism of the *Philosophy*.)

Buchdahl, G. and Laudan, L.L. (eds) (1987) *The Historical and Philosophical Works of William Whewell*, London: Frank Cass, 10 vols.(An annotated edition of all of Whewell’s published work in history and philosophy of science.)

Butts, R.E. (ed.) (1968) *William Whewell’s Theory of Scientific Method*, Pittsburgh, IL: University of Pittsburgh Press.(A short selection of Whewell’s philosophical writings.)

Elkana, Y. (ed.) (1984) *William Whewell: Selected Writings on the History of Science*, Chicago, IL, and London: University of Chicago Press.(A selection of Whewell’s philosophical and historiographical texts that bear on his history of the sciences.)

References and further reading

Blanché, R. (1935) *Le Rationalisme de Whewell* (Whewell’s Rationalism), Paris: Librairie Felix Alcan.(A close study of the ultimately unreconciled Kantian and Platonist aspects of Whewell’s philosophy.)

Fisch, M. (1991) *William Whewell: Philosopher of Science*, Oxford: Clarendon Press. (A detailed account of the formation of Whewell’s philosophy of science from his early involvement in the reform of Cambridge mathematics, and a critical appraisal of his mature system.)

Fisch, M. and Schaffer, S. (eds) (1991) *William Whewell: A Composite Portrait*, Oxford: Clarendon Press.(A collection of 13 authoritative studies of the content and context of Whewell’s life, work and impact. Contains a useful bibliography.)

Marcucci, S. (1963) *L’idealismo scientifico di William Whewell* (William Whewell’s Scientific Idealism), Pisa: Instituto di filosofia.(A study of Whewell’s philosophy of science that places him within the idealist school.)

Stair-Douglas, M. (1881) *The Life and Selections from the Correspondence of William Whewell D.D.*, London: Kegan Paul.(A brief biography and selection of correspondence pertaining mainly to Whewell’s personal life.)

Todhunter, I. (1876) *William Whewell, D.D. Master of Trinity College Cambridge: An Account of his Writings with Selections from his Literary and Scientific Correspondence*, London: Macmillan, 2 vols.(Contains an intellectual biography structured around Whewell’s published works, and a wide selection of his professional correspondence.)

Whitehead, Alfred North (1861-1947)

Whitehead made fundamental contributions to modern logic and created one of the most controversial metaphysical systems of the twentieth century. He drew out what he took to be the revolutionary consequences for philosophy of the new discoveries in mathematics, logic and physics, developing these consequences first in logic and then in the philosophy of science and speculative metaphysics. His work constantly returns to the question: what is the place of the constructions of mathematics, science and philosophy in the nature of things?

Whitehead collaborated with Bertrand Russell on Principia Mathematica (1910-13), which argues that all pure mathematics is derivable from a small number of logical principles. He went on in his philosophy of science to describe nature in terms of overlapping series of events and to argue that scientific explanations are constructed on that basis. He finally expanded and redefined his work by developing a new kind of speculative metaphysics. Stated chiefly in Process and Reality (1929), his metaphysics is both an extended reflection on the character of philosophical inquiry and an account of the nature of all things as a self-constructing 'process'. On this view, reality is incomplete, a matter of the becoming of 'occasions' which are centres of activity in a multiplicity of serial processes whereby the antecedent occasions are taken up in the activities of successor occasions.

1 Life

Alfred North Whitehead studied mathematics at Trinity College, Cambridge. In 1884 he became a Fellow of Trinity, where he taught mathematics, Bertrand Russell and J.M. Keynes being among his pupils. He was a liberal in politics and an advocate of women's rights. He wrote*Principia Mathematica* (1910-13) with Russell, moving to London in 1910, where he taught mathematics at University College and was active in educational reform. He became Professor of Applied Mathematics at the Imperial College of Science and Technology in 1914, thereafter writing on the philosophy of science. In 1924 he took up a chair in philosophy at Harvard, where he produced a series of works in speculative metaphysics.

Whitehead's work falls into three periods: the early period of mathematics and logic, the middle period concerned with the philosophy of science; and the late period of speculative metaphysics. Although a matter of debate, his philosophical development is best seen as a shifting, ever-widening analysis of the concept of construction (see Constructivism §1; Constructivism in mathematics §6).

2 Mathematics and logic

Whitehead's three main early works indicate the central role played by mathematics, logic and science in shaping the themes and models which inform his philosophical thought. His first book, *A Treatise on Universal Algebra: With Applications* (1898), develops Grassmann's work in what was then the new field of abstract algebra. Although it had little influence on the subsequent development of mathematics, the *Treatise* has at least three philosophically significant features.

First, it is concerned with the 'universalization' or 'generalization' of variables beyond their traditional restriction as symbols for numbers. Generalization takes the form of 'substitutive schemes', which at this stage in Whitehead's career are non-axiomatized algebraic formulae having the status of a calculus which symbolizes the operations of addition and multiplication. He holds that the construction of substitutive schemes involves some (here unspecified) relation to the empirical world and that, to be significant, such schemes must have applications or substitutions in some field. In turn, application assists in the investigation of the schemes themselves, which have heuristic value even if they are only partially interpretable at a given stage of knowledge.

Second, alongside the emphasis on empirical connection and a realist concern with the significant application of schematized structures, a strong formalist tendency is evident in the *Treatise*: while consistency is repudiated as the sole ground for existence-theorems, mathematical schemes are defined as conventional idealizations independent of perceptual content (see Realism in the philosophy of mathematics §2). Third, however, Whitehead also emphasizes the synthetic processes of intellectual construction involved in mathematical inference. Like the mathematical intuitionists later on, he regards '$2+3$' and '$3+2$' as nonidentical (see Intuitionism §1); he holds that the difference of order directs different processes of thought and that equivalence is a matter of identity-in-difference.

The development of a philosophy which would coherently relate the different orientations - realist, formalist and intuitionist - evident in the treatment of construction in his mathematical writings is a central concern of Whitehead's subsequent thought.

Whitehead's Royal Society memoir On Mathematical Concepts of the Material World (1906) offers different logical schemes or models for different theories of the structure of the physical world. As later in *Process and Reality* (1929), Whitehead employs the axiomatic method in constructing a scheme and defines a scheme as a hypothesis to be assessed on the principle of Ockham's razor (see William of Ockham §2): the preferred scheme in the memoir is that which posits only one class of ultimate entities. He anticipates his later thought by criticizing the Newtonian account of nature as composed of externally related atoms each occupying a position in absolute space at an absolute time (see Mechanics, classical §§1, 4-5; Newton, I. §§3-4). His preferred scheme is constructed on the model of electromagnetic theory in terms of the flux of energy. The ultimate entities are complexes of relations: points are classes of linear relations which can be given an empirical interpretation in terms of vectors or lines of force. The emphasis is on a logic of relations as a way of uniting permanent structure and change in one schematized serial order - an issue which occupied Whitehead for the rest of his career.

Principia Mathematica (1910-13), written in collaboration with Russell, (see Russell, B. §3) is best seen as an account of the logic of relations and an attempt to develop as far as possible the hypothesis of 'logicism' - the claim that all pure mathematics can be derived from an axiomatic scheme of logical concepts taken as primitive (see Logicism §4). The effect of the difficulties which subsequently emerged in the logicist project was to strengthen the mathematical-intuitionist tendencies in Whitehead's thought: he later treats Gödel's incompleteness theorem as indicating that mathematical logic is an instance of the finite character of all constructions (see Gödel's theorems §§3-5).

3 Philosophy of science

Whitehead's philosophy of science analyses the ontological status of scientific and mathematical concepts in terms of their derivation from the elements and relations of nature as disclosed in sense-experience. He abandons the Newtonian concept of 'nature at an instant' and rejects the 'bifurcation' of nature into perceived qualities and the theoretical entities of science (such as electrons), in particular Russell's phenomenalist view of theoretical entities as constructions out of atomic sense-data. The new theories of physics, the new logic of relations and Bergson's account of the fluidity of nature are seen as making possible an empiricism which is neither representationalist nor phenomenalist. Logical constructions and perceived qualities are to be analysed as features of 'one system' of multiple relations by means of a redefinition of sense-experience as the disclosure of the 'passage of nature', that is, of occurrences or events with spatiotemporal spread.

'Events' are one of the two basic constituents of nature that Whitehead holds to be disclosed in sense-experience. Taking the place of traditional concepts of substance, events are unrepeatable, relational entities which overlap or extend over one another (see Substance §3). Their nature as spatiotemporal regions, however, expresses only some of the features of this fundamental relation of extension, for Whitehead also terms it 'process' and 'creative advance'. Events thus imply a special kind of activity pointing towards the later, speculative account of reality as 'creativity' (see Events).

'Objects' are the second basic constituent of nature disclosed in sense-experience. Objects take the place of universals: they are repeatable characters or properties which are neither Platonic forms nor nominalist resemblances but, as 'ingredients' in events, are like Aristotelian *universalia in rebus* (see Aristotle §15; Universals §1). However, objects do not act and do not stand to events in an invariable two-termed relation of predication; they are terms in a multiple relation of 'ingression'. Thus a 'sense-object' - for example, a colour which is perceived as situated in an event - involves multiple relations between the percipient event (the relevant bodily state of the observer), the event which is the situation of the object, and the conditioning events relating the percipient event to the situation. As this multiple relation is a facet of nature, the perspective relative to the percipient event is not 'outside' the world but belongs to the world in that relation, thereby dissolving the duality of 'subjectivity' and 'objectivity'. Scepticism is never an issue for Whitehead, as scepticism assumes that duality has fundamental metaphysical status (see Scepticism §1). His complex account of different types of objects is replaced in his later work by an account of the relation of what he calls 'eternal objects' to the becoming of

‘occasions’.

What Whitehead called his ‘Method of Extensive Abstraction’ is primarily an application of a constructivist logic of classes to the question of the status of such theoretical entities as instants of time and points in space. Instants and points are routes of approximation across, respectively, the durations of events and spatial volumes, which are arranged in a continuum of whole and part like a nest of Chinese boxes. As we descend in the series we progressively reach durations and volumes of ever smaller extension. Instants and points are sets or classes of the whole - part relations of the durations and volumes which enclose them.

Whitehead’s attempt to unify space, time and matter in a single system of relations led him to address Einstein’s theory of relativity critically. He agrees with Einstein that space and time are abstractions from spacetime events and that there is an infinite plurality of different ‘time-series’. However, Whitehead’s analysis of the perceptual relations between events allows him to maintain the objectivity both of the distinction between space and time for the observer and of the relative position of the observer. Further, the problems of rotation, incongruent counterparts and the application of geometries lead him to give the passage of events its own internal, uniform (homoloidal) spatial structure, independent of any relation to the objects of which they are the situations, and he treats that structure as actual. This brings him into conflict with the general theory of relativity in which spacetime varies with its (material) contents (see Relativity theory, philosophical significance of §2).

4 Speculative metaphysics

Whitehead’s speculative thought challenges the critique of metaphysics characteristic of twentieth-century Anglo-American and European philosophy. In contrast to Russell and his successors who interpret modern mathematical and logical developments in the context of a weak theory of being or existence as quantification, Whitehead sees these developments as reopening the possibility of a strong theory of being or existence as ‘act’ (see Being §4; Existence §2). Like Bergson and the later Heidegger (see Heidegger, M. §2), Whitehead defines the act of being in terms of finite self-actualization, independent of any metaphysically complete cause or ground; but unlike them, he regards self-actualization as rationally analysable.

There is no longer any foundationalist appeal to sense-experience in Whitehead’s metaphysics (see Foundationalism §1). The subject matter of his ‘speculative scheme’ is the ‘empirical side’ of the analysis, defined as everything of which we are conscious as historically situated beings. The construction of an axiomatized scheme of categories is a matter of ‘imaginative generalization’: significant features of the historical world are employed as analogues for the analysis of the nature of all things. The cogency of the categories depends upon the range of their ‘substitutions’ or ‘applications’ on the ‘empirical side’. Categories and applications stand in a relation of proportional analogy (as A is to B, so C is to D).

As a ‘generalized mathematics’, Whitehead’s speculative scheme of categories can be epitomized as universalizing the mathematical concept of series and as closely akin to the mathematical intuitionists’ theory of number construction. The self-actualization of all things is interpreted as a finite series of acts of self-construction which are asymmetrical, transitive and irreflexive, constituting an iterative, infinitely proceeding multiplicity of sequences. This is the ‘process’ of ‘actual [that is, actualizing] entities’, or ‘actual occasions’ which constitute the ‘one genus’ by which the scheme aspires to describe everything. Whitehead terms the scheme ‘the philosophy of organism’ to indicate that it embraces biology as much as physics and constitutes a theory of the ‘experience’ of all things as self-constructing centres of activity (see Processes §4).

Whitehead’s complex analysis of the serial process of the becoming of occasions employs a variety of analogies with ‘mentality’, ‘feeling’ and the vectorial transfer of energy. Although they do not overlap, there are no ‘single’ occasions, for occasions are relational or serially constituted entities in two senses. First, the completion of an occasion in a series of occasions is its ‘objectification’ by successor occasions. Second, a successor occasion has its own internal serial structure as a self-constructing synthesis or ‘concrescence’. The relation of freedom and determination simply depends on an occasion’s complexity of response, allowing it to modify the extent to which it is determined by its predecessors and environment.

Whitehead’s theory of eternal objects or ‘pure potentials’ is modelled on that of the propositional function. However, the theory does not grant the propositional function any metaphysically primitive status but explains what makes the propositional function possible. In the state of ‘general potentiality’ eternal objects do not form an

infinite class of possibilities, but a ‘matrix’ of the ‘multiplicity’ of the endlessly nested possibilities which finite actualizations afford. As components of predecessor occasions which are objectifiable by successor occasions, eternal objects have the status of ‘real potentiality’. Potentiality and actuality are defined as relations of seriality, states or stages of actualization, which stand in a strictly functional relation of satisfaction and not in any of the traditional relations of agency, resemblance or imitation.

Eternal objects of a special kind, ‘the objective species’, constitute the ‘extensive continuum’ or abstract system of part and whole relations. The key philosophical difference with Whitehead’s middle-period extension is that the extensive continuum is potential, not actual, and is actualized by the becoming of occasions. Mathematical relations can thus be intuitionistically defined as equivalently ideal and real, potential and actual: that is to say, they are definable as states or stages of serial actualization.

Whitehead’s scheme of categories has a number of significant applications, some of which can be summarized here.

Epistemological Realism. By serializing occasions, Whitehead would reconcile the claim that knowledge is perspectival or ‘situated’ with the claim that knowledge is to be defined in terms of objects distinct from and independent of the subject or concrescent occasion.

Relations. Whitehead would dissolve the opposition between F.H. Bradley and Russell on relations by interpreting the doctrines of internal and external relations in the context of his serial pluralism (see Bradley, F.H. §§4-5): a successor occasion is internally related to its antecedent occasions, but its antecedent occasions (as completed or ‘perished’) are not internally related to it. Internal and external relations are thus states or stages of seriality.

Perception. The primary mode of perception is not a matter of distinct impressions or sense-data but, under the principle of serial connection, is the direct experience of the ‘causal efficacy’ of the antecedent world. The other mode of perception is ‘presentational immediacy’, which is the presented locus or spacetime region of the percipient occasion. The two are combined in an adverbial theory of sensation.

God. Whitehead does not attempt to prove God’s existence but to show that, contrary to the views of many theorists of finite self-actualization, there is nothing to prevent the redefinition of the concept of God in that context. Thus eternal objects in their state of ‘general potentiality’ are ‘termed’ God because of the ultimate, underivable character of possibilities of structure: as the final principle of determination, the concept of God is their nearest conceptual correlate. Similarly, ‘real potentiality’ can be said to be derived from God because the experience of potentiality from an occasion’s standpoint is nothing other than the ultimacy of ‘appetition’ for actualization. Interpreted as God, these aspects are termed his ‘primordial nature’ and revise the concept of God as creator: he does not create occasions but provides them with potentiality. In his ‘consequent nature’ God objectifies and transforms occasions in the eternal harmony of his nature. As a nontemporal being, God is an ‘actual entity’, not an occasion, but his primordial and consequent natures define him as a component of serial process. This ‘dipolar’ nature expresses the way in which the scheme has a place for the concept of God as both first cause and redeemer (see Process theoism).

Speculative Analysis. Whitehead’s general strategy is to define ideality and reality, not as fundamental metaphysical opposites, but as states or stages of the serial process of self-construction. While his universalization of the concept of finite or serial construction makes him a speculative realist, his metaphysics is also self-referentially inclusive, that is, the metaphysical scheme is an instance of finite or serial intellectual construction which as such is not subject to the law of the excluded middle (see Intuitionistic logic and antirealism §1). The scheme can thus be regarded as real inasmuch as it is constructed or applied, and ideal or hypothetical inasmuch as it is finite and revisable and consequently does not exclude alternative analyses. While most readers of Whitehead interpret his speculative realism to mean that his categories are accounts of the metaphysical contents of the world, one can also regard them as conditions of the contents of the world, that is, as transcendental categories of a new kind which are neither real in the medieval sense nor ideal in the Kantian sense, but fallible, historically situated constructions, formulating the ultimate generalities which constitute the nature of things.

JAMES BRADLEY

List of works

Whitehead, A.N. (1898) *A Treatise on Universal Algebra: With Applications*, Cambridge: Cambridge University Press.(A pioneer work in the field of abstract algebra.)

Whitehead, A.N. (1906) On Mathematical Concepts of the Material World. Philosophical Transactions of the Royal Society of London, Series A, in F.S.C. Northrop and Mason W. Gross, *Alfred North Whitehead: An Anthology*, Cambridge: Cambridge University Press, 1953, 11-82.(This book contains extensive extracts from Whitehead's main works.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1927.(One of the central works in modern logic and the philosophy of mathematics.)

Whitehead, A.N. (1911) *An Introduction to Mathematics*, London: Williams & Norgate; repr. Oxford: Oxford University Press, 1948, 1958, 1969.(A clear, elegant account of the basic elements of modern mathematics.)

Whitehead, A.N. (1917) *The Organization of Thought*, London: Williams & Norgate, and Philadelphia, PA: J.B. Lippincott.(This collection of papers, with others presented to the Aristotelian Society between 1916 and 1923, notably 'Uniformity and Contingency' (1922), is reprinted in A.H. Johnson (ed.) *Alfred North Whitehead: The Interpretation of Science*, New York: Bobbs-Merrill, 1961.)

Whitehead, A.N. (1919, 1925) *An Enquiry Concerning the Principles of Natural Knowledge*, Cambridge: Cambridge University Press.(The second edition of 1925, with additional notes, should be consulted.)

Whitehead, A.N. (1920) *The Concept of Nature*, Cambridge: Cambridge University Press, 1993. (A nontechnical analysis of the philosophy of science and nature which Whitehead sees as required by physics.)

Whitehead, A.N. (1922) *The Principle of Relativity*, Cambridge: Cambridge University Press; philosophical sections repr. in Whitehead (1906).(A critical analysis of the theory of relativity from the point of view of Whitehead's philosophy of science.)

Whitehead, A.N. (1925) *Science and the Modern World*, New York: Macmillan, and Cambridge: Cambridge University Press, 1926; New York: The Free Press, 1967.(A critical history of the development of modern science, concentrating on the concept of organism as the basic principle of modern scientific thought.)

Whitehead, A.N. (1926) *Religion in the Making*, New York: Macmillan, and Cambridge: Cambridge University Press, 1926; New York: Meridan, 1974.(A historical and critical reflection on the nature of religion in the light of Whitehead's metaphysics. References are usually to this or the Cambridge edition.)

Whitehead, A.N. (1926) *Symbolism, Its Meaning and Effect*, Cambridge: Cambridge University Press; New York: Macmillan, 1927; London: G.P. Putnam's Sons, 1959.(A clear statement of Whitehead's theory of perception, later elaborated in *Process and Reality*.)

Whitehead, A.N. (1929) *The Aims of Education and Other Essays*, New York: Macmillan, and London: Ernest Benn, 1929; New York: The Free Press, 1967.(This work is essentially a shorter version of *The Organization of Thought*.)

Whitehead, A.N. (1929) *The Function of Reason*, Princeton, NJ: Princeton University Press, 1929. Boston, MA: Beacon Press, 1958.(A prolegomenon to *Process and Reality*, this work presents philosophy as a matter of the creation of novel concepts.)

Whitehead, A.N. (1929) *Process and Reality: An Essay in Cosmology*, New York: Macmillan, and Cambridge: Cambridge University Press; corrected edn, with comparative readings and detailed index, by D.R. Griffin and D.W. Sherburne, New York: The Free Press, 1978. (Whitehead's central work in speculative metaphysics, presented in axiomatized form with extensive applications.)

Whitehead, A.N. (1933) *Adventures of Ideas*, New York: Macmillan, and Cambridge: Cambridge University Press; New York: The Free Press, 1967.(An analysis of the concepts of civilization and history from the standpoint of Whitehead's metaphysics, this work includes a clear statement of his position and a redefinition of the 'transcendentals' of medieval philosophy.)

Whitehead, A.N. (1938) *Modes of Thought*, New York: Macmillan, and Cambridge: Cambridge University Press; New York: The Free Press, 1968.(A rich and suggestive restatement of Whitehead's metaphysics, clearly underlining his mathematical intuitionist orientation.)

Whitehead, A.N. (1947) *Essays in Science and Philosophy*, New York: The Philosophical Library. (An important collection of papers, including some of Whitehead's clearest programmatic and methodological statements, notably 'Uniformity and Contingency', 'Analysis of Meaning', 'Process and Reality' and 'Mathematics and the Good'.)

Woodbridge B.A. (ed.) (1977) *Alfred North Whitehead: A Primary-Secondary Bibliography*, Bowling Green, OH: Philosophy Documentation Center, Bowling Green State University.(A complete bibliography of Whitehead's works and the secondary literature up to 1977, totalling 1,868 entries.)

References and further reading

Braithwaite, R.B. (1926) Review of Science and the Modern World, *Mind* 35: 489-500.(A clear, critical analysis of Whitehead's metaphysics in the light of his earlier work.)

Braithwaite, R.B. (1927) Contribution to 'Symposium: Is The "Fallacy of Simple Location" A Fallacy?', *Aristotelian Society* supplementary vol. 7: 224-36.(A clear, critical analysis of Whitehead's philosophy of science.)

Christian, W.A. (1959) *An Interpretation of Whitehead's Metaphysics*, New Haven, CT: Yale University Press.(A standard, detailed analysis.)

Deleuze, G. (1988) *Le Pli: Leibnix et le Baroque*, Paris: Les Éditions de Minuit; trans. T. Conley, *The Fold: Leibniz and the Baroque*, Minneapolis, MN: University of Minnesota Press, 1992; London: The Athlone Press, 1993.(The importance of Whitehead's metaphysics in the modern philosophical tradition.)

Dewey, J. (1937) 'Whitehead's Philosophy', *Philosophical Review* 46: 170-7; also in *Problems of Men*, New York: Philosophical Library, 1946, 410-18.(Whitehead's response is 'Analysis of Meaning' in Whitehead (1947): 122-31.)

Emmet, D. (1946) *The Nature of Metaphysical Thinking*, London: Macmillan; repr. New York: St Martin's Press, 1966.(The role of analogy in metaphysical thinking, with valuable analyses of Whitehead's theory of perception and use of analogy).

Fitch, F.B. (1957) 'Combinatory Logic and Whitehead's Theory of Prehensions', *Philosophy of Science* 24: 331-5.(A partly technical analysis, presenting combinatory logic as the basis of the theory of prehensions and eternal objects.)

Ford, L.S. (1984) *The Development of Whitehead's Metaphysics*, *1925-1929*, Albany: State University of New York Press, 1984.(A detailed discussion of the development of Whitehead's central metaphysical concepts).

Ford, L.S. (ed.) (1971) *Process Studies*, California: Claremont.(A journal devoted to Whitehead and 'process' philosophies.)

Ford, L.S. and Kline, G.L. (1983) *Explorations in Whitehead's Philosophy*, New York: Fordham University Press.(A useful collection which includes essays by Buchler, Christian, Cobb, Henry, Kline, Ford and Rorty.)

Haack, S (1979) 'Descriptive and Revisionary Metaphysics', *Philosophical Studies* 35: 361-71.(An instructive comparison of *The Concept of Nature* with P.F. Strawson's *Individuals*.)

Hampe, M. and Maassen, H. (eds) (1991) *Prozess, Gefühl und Raum-Zeit: Materialen zu Whitehead's Prozess und Realität*, vol. 1; *Die Gifford Lectures und Ihre Deutung: Materialen zu Whitehead's Prozess und Realität*, vol. 2, Frankfurt: Suhrkamp.(A valuable collection, historical and contemporary. Some of the material translated into German has not been published in English.)

Harris, E.E. (1954) *Nature, Mind and Modern Science*, New York: The Macmillan Company.(A valuable analysis and critique of Whitehead's metaphysics from a British Idealist viewpoint, with comparisons to Hegel).

Holz, H. and Wolf-Gazo, E. (eds) (1981) *Whitehead and the Idea of Process*, Proceedings of the First International Whitehead Symposium, Freiburg and Munich: Karl Alber.(A wide-ranging collection of essays by American, British and Continental philosophers.)

Kline, G.L. (ed.) (1963) *Alfred North Whitehead: Essays on His Philosophy*, Englewood Cliffs, NJ: Prentice Hall. (A useful collection which includes essays by Hall, Hartshorne, Leclerc and Rorty.)

Lawrence, N. (1956) *Whitehead's Philosophical Development: A Critical History of the Background of Process and Reality*, Berkeley, CA: University of California Press. (A useful account, which reads Whitehead's development in terms of a tension between 'realism' and 'conceptualism' - a view strongly challenged by Schmidt (1967).)

Leclerc, I. (1958) *Whitehead's Metaphysics: An Introductory Exposition*, London: Allen & Unwin; 2nd edn, 1965. (A standard introductory text to Whitehead's metaphysics.)

Leclerc, I. (ed.) (1961) *The Relevance of Whitehead*, New York: Humanities Press and London: Allen & Unwin. (A collection which includes indispensable essays by Leclerc on Aristotle and Whitehead, and Mays on the Royal Society Memoir of 1906.)

Lowe, V. (1962) *Understanding Whitehead*, Baltimore, MD: The Johns Hopkins University Press; 2nd edn, 1966. (The standard introductory text to Whitehead's entire *oeuvre*.)

Lowe, V. (1985, 1990) *Alfred North Whitehead: The Man and His Work*, Baltimore, MD: The Johns Hopkins University Press, 2 vols.(A detailed biography of Whitehead, with analyses of his works.)

Mays, W. (1959) *The Philosophy of Whitehead*, New York: The Macmillan Company and London: Allen & Unwin.(An indispensable, nontechnical analysis of Whitehead's metaphysics, relating it to his logical and mathematical work.)

Murphy, A.W. (1927) 'Objective Relativism in Dewey and Whitehead', *Philosophical Review* 36: 121-44; also in W. Hay and M.G. Singer (eds) *Reason and the Common Good: Selected Essays of A.W. Murphy*, Englewood

Cliffs, NJ: Prentice Hall Inc., 1963, 163-77.(A striking analysis of the significance and relation of the philosophies of Dewey and Whitehead.)

Nobo, J.L. (1986) *Whitehead's Metaphysics of Extension and Solidarity*, Albany, NY: State University of New York Press, 1986.(This work contains very useful analyses and clarifications of Whitehead's central concepts.)

Palter, R.M. (1960) *Whitehead's Philosophy of Science*, Chicago, IL: University of Chicago Press; 2nd edn, 1970. (A technical and detailed analysis of Whitehead's theories of extension, including an account of his relativity theory.)

Quine, W.V. (1941) 'Whitehead and the Rise of Modern Logic' in P.A. Schilpp (ed.) *The Philosophy of A.N. Whitehead*, The Library Of Living Philosophers, La Salle, IL: Open Court, 127-63.(An indispensable analysis, partly technical, which presents Whitehead as tending to 'platonic' realism.)

Rapp, F. and Wiehl R. (eds) (1983) *Whitehead's Metaphysics of Creativity*, Proceedings of the International Whitehead Symposium, Freiburg and Munich: Verlag Karl Alber, 1986; trans. G. Treash *et al*. New York: State University of New York Press, 1990. (A useful collection of essays mainly by German philosophers and theologians.)

Russell, B. (1948) 'Whitehead and Principia Mathematica', *Mind* 57, 137-8. (An account of the co-authorship of *Principia*, emphasizing Whitehead's role.)

Russell, B. (1956) *Portraits From Memory*, London: Allen & Unwin, 1956, ch. 5.(An historically important and amusing account of the influence on Russell of Whitehead's early thoughts on 'construction'.)

Schmidt, P.F. (1967) *Perception and Cosmology in A.N. Whitehead's Philosophy*, New Brunswick, NJ: Rutgers University Press, 1967.(A clear and detailed analysis of Whitehead's philosophical development.)

Wahl, J. (1932) *Vers le concret*, Paris: Vrin.(An influential analysis of Whitehead's place in twentieth-century philosophy, relating him to Heidegger and others.)

Wilson, E. (1931) *Axel's Castle*, New York: Charles Scribner's Sons. (The conceptual connections between Whitehead's thought and twentieth-century literature, Proust in particular.)

Will, the

As traditionally conceived, the will is the faculty of choice or decision, by which we determine which actions we shall perform. As a faculty of decision, the will is naturally seen as the point at which we exercise our freedom of action - our control of how we act. It is within our control or up to us which actions we perform only because we have a capacity to decide which actions we shall perform, and it is up to us which such decisions we take. We exercise our freedom of action through freely taken decisions about how we shall act.

From late antiquity onwards, many philosophers took this traditional conception of the will very seriously, and developed it as part of a general theory of specifically human action. Human action, on this theory, is importantly different from animal action. Not only do humans have a freedom of or control over their action which animals lack; but this freedom supposedly arises because humans can act on the basis of reason, while animal action is driven by appetite and instinct. Both this freedom and rationality involve humans possessing what animals are supposed to lack: a will or rational appetite - a genuine decision making capacity.

From the sixteenth century on, this conception of the will and its role in human action met with increasing scepticism. There was no longer a consensus that human action involved mental capacities radically unlike those found in animals. And the idea that free actions are explained by free decisions of the will came to be seen as viciously regressive: if our freedom of action has to come from a prior freedom of will, why shouldn't that freedom of will have to come from some yet further, will-generating form of freedom - and so on ad infinitum?

Yet it is very natural to believe that we do have a decision making capacity, and that it is up to us how we exercise that capacity - that it is indeed up to us which actions we decide to perform. The will-scepticism of early modern Europe, which persists in much modern Anglophone philosophy of action, may then have involved abandoning a model of human action and human rationality that is deeply part of common sense. We need to understand this model far better before we can conclude that its abandonment by so many philosophers really was warranted.

1 The will and rational action

Talk of an individual's will may mean simply what actions they are motivated to perform. In this sense any performer of actions, or *agent*, has a will. But talk of the will has served more specifically to pick out the motivations of those action performers who are *rational agents*.

Suppose that there are rational justifications for performing some actions rather than others - justifications which come from a practical or action-governing reason. Then a rational agent is an agent who has the capacity to come to recognize these requirements, and then apply them.

Not all agents need be rational agents. To be a rational agent, one needs, first, a practical intellect or deliberative capacity to form practical judgments about what actions are rationally justified; and second, one needs an executive capacity to apply these judgments in one's actions. Animals, in particular, have often been seen (by Aquinas, for instance) to lack practical intellect. They supposedly have no conception of a justification for action, and so act simply on the basis of appetite.

Notice that rational agents, on our definition, can act irrationally, through defective use of their deliberative and executive capacities. They might make a deliberative mistake as to what actions are justified, or (though this possibility is controversial; see Akrasia) the defect might be purely executive, in their deliberate performance of an action despite their judgment of this action as unjustified.

Someone who does count as a rational agent can be said to possess a *rational appetite* - a capacity to become motivated to act through rational decision making, thereby forming intentions to act, decisions and intentions based in turn on their deliberations about what actions would be justified. Talk of a faculty of will has typically served to pick out this rational appetite or decision making capacity. *Intellectualists*, most famously Thomas Aquinas, see the will as tied to the practical intellect, decisions to perform a particular action occurring through a practical judgment that that action should be performed. Accordingly, any freedom of decision must derive from a freedom of the practical intellect (see Voluntarism). Whereas, flourishing after Aquinas, *voluntarists* maintain that the will has a freedom independent of the practical intellect. One might even decide, irrationally, to perform an

action despite judging its performance to be unjustified.

Medieval and early modern scholasticism produced highly developed theories of human agency as involving the will. Thus Aquinas saw intentional or deliberate human action as generally involving the exercise of the will: when I intentionally move my hand, that action arises out of my willing my hand to move. And Aquinas thought this because, in his view, a truly deliberate action must arise from a capacity to respond motivationally to rational requirements attendant upon action - in other words, from an exercise of the rational appetite. Animal action could not be intentional as ours is, precisely because animals lack this rational motivational capacity.

Furthermore, Aquinas conceived the exercise of the will to be itself a case of action. The *first-order* agency of ordinary actions such as deliberately raising one's hand, was to be explained in terms of a *second-order*, action-generating agency of the will. Why see the willings or decisions to act which, allegedly, explain our actions as being actions themselves? A key assumption is that action in its distinctively human, fully intentional form, consists in agents conforming to or violating requirements of practical reason. 'An act is [fully] voluntary when it is an operation of reason' (*Summa theologiae* Ia IIae, q.6 a.1). On this *practical reason-based* view of human agency, to act is to exercise our rationality, whether competently or defectively, in a distinctively practical way. And this we do, in particular, in the decisions by which we motivate ourselves to perform this action rather than that. For whether we count as conforming to or violating practical reason does indeed plausibly depend on which actions we decide to perform (see Aquinas, T. §12).

Wherever a conception of human agency as essentially the exercise of practical rationality has remained compelling, so theorists have been inclined to characterize ordinary human actions, such as intentional hand-raisings, as involving an inherently intentional, action-motivating agency of the will. How more precisely the will is conceived then greatly depends on the accompanying conception of our practical rationality. Thus Kant, holding practical reason to be the legislation of all rational agents, conceived the will partly as *Wille*, a capacity for autonomous legislation, and partly as *Willkür*, an executive, action generating capacity that was itself exercised freely. Deliberate human agency, whether rational or irrational - whether conforming to *Wille* or not - generally involved the operation of *Willkür*.

Did the Greeks entertain analogously will-based theories of human agency? Only rational humans, according to Aristotle, can act on the basis of *prohairesis*, or a rational choice based on deliberation about how it would be good to act (see *Nichomachean Ethics*, Book III). Aquinas took Aristotle's prohairesis to be equivalent to his own *electio* or decision of the will; but mistakenly, since Aquinas' decision of the will was supposed to be a general feature of intentional human action, whereas for Aristotle only some such human action was prohairetic: intentional human agency could also occur without involving a distinctively rational appetite.

Stoicism did develop a more truly will-involving conception of human agency. While animal action was seen as simply driven by impulse, humans were considered to have a rational capacity for *sunkatathēsis* - a capacity for judgmental assent to, or dissent from, a given action - which capacity was believed essential to our deliberate agency. Thus humans had control over what actions they assented to. The voluntary was thought to begin then at the point where, through *sunkatathēsis*, we motivated ourselves to perform a given action (see Stoicism §12).

2 The will and the free will problem

The free will problem can arise as a strictly ethical problem about our moral responsibility for our actions: under what conditions can we genuinely deserve morally based reward or blame for what we do? (See §3.) But the free will problem can also arise as a problem, not immediately ethical, about freedom of action - about whether we possess genuine control of how we act, in the sense of possessing a freedom to act otherwise than as we actually do: is it really up to us which actions we perform?

It is nowadays sometimes assumed that, as a problem about action, the free will problem has nothing much to do with the will. But this is a mistake, as our ordinary intuition about freedom of action suggests. It is up to us which actions we perform only because we are capable of taking decisions about how we shall act, and it is up to us which actions we decide to perform. Is not that a claim about our freedom to which we naturally assent? If so, then our ordinary conception of our freedom is *psychologizing*: that is, the freedom of our first-order action is held to depend on our possession of a prior, psychological analogue of that freedom located in a second-order agency of the will.

We have seen that the idea that we have a capacity for a second-order agency of the will at all - that the decisions which explain our actions are actions themselves - comes with a practical reason-based conception of our agency: human action consists in the exercise of practical rationality. It is plausible then that our psychologizing conception of our freedom of action is likewise to be explained in terms of some connection between freedom of action and practical rationality. A widely believed condition of possessing freedom of action, after all, is that one have a capacity for practical rationality. Free agents must be able to exercise their control over their actions in a reason-applying way. Animals such as sharks or mice are not free to act otherwise than as they do, precisely because they are not plausibly rational agents. Perhaps this connection between freedom and practical rationality can explain why freedom of action should depend on a freedom of the will in particular.

One way this explanation might work is if we suppose that free agents must indeed be able to exercise their control over their actions in a reason-applying way. How then does practical reason recommend actions to us? Plausibly it does so in the form of plans, or sequences extending through time. When we deliberate about how to act, we are almost invariably deliberating between rival *plans* or sequences of action extending into the future, our performance of which as a whole depends on our continuing motivation into the future. And that is because how we act in the present generally only has value for us on the assumption that we do go on to perform further appropriate actions in the future. Saying a given word now generally only has value if we do go on to utter the other words composing our message. Stepping out of the door now only has value if we do continue walking to the supermarket thereafter. The unit of practical reason - the form in which actions have value for us, and so the form in which we deliberate about them - is the plan.

So to exercise action control in a reason-applying way, we need to exercise control over whole plans, and so over sequences of action extending into the future. Without plan control, as we just saw, we should at any given time lack control over such elementary matters as what we communicated, or where we shopped: for such depend on which plans we execute. Now such plan control can only come from some controlled doing in the present, available to us as rational deliberators, by which we can determine which future actions we shall be performing. And deciding to act looks as though it is precisely the doing required. By deciding on a given plan of action, we can ensure that we thereafter remain motivated to execute it, so that we do act as we have deliberated we should. That is why our decisions can often leave us and others knowing in advance that we shall indeed be acting as we have decided. Decision making can give us plan control, provided of course that taking a particular decision to act is indeed something that we deliberately do, so that we can have control over which actions we decide to perform. Given free will, decision control gives us a capacity for rational self-determination - for controlling our actions in the form in which they matter to us as rational agents. And that is a capacity which free agents plausibly need.

3 The will and ethical theory

Does being a morally responsible agent require any special capacity of will, beyond even that required for being a rational agent?

Thomas Nagel has claimed that there is an aspect of common sense morality which excludes luck: someone's moral goodness or conformity to moral requirements must not depend on any factors which are outside their control, or for which they are not morally responsible. Yet people's ordinary first-order actions and their consequences do seem to depend on luck. Whether, for example, a given person ever manages to help others must depend on matters for which they may well not be responsible, such as the presence of others needing help, and their own possession of the skill and resources to provide the help (see Moral luck).

The only capacity for action that could ever be independent of luck seems to be the second-order action of one's will - of one's decisions to perform first-order actions. For whether or not others are around to be helped, and whether or not one could help others if one tried, one can still take an advance *decision* to help others should the need arise and should one be able to help. It is in such inner decisions of the will, then, that the moral life must be lived: moral obligations are obligations on the will; and it is solely through the agency of one's will that one can count as virtuous or morally good.

However, perhaps even the actions of the will are not really independent of luck either. What decisions I take depends on my education and upbringing, matters of luck too. The doctrine that morality excludes luck is arguably a sceptical view, that prevents morality applying to any empirical agents like us.

But there is a quite different view which also implies moral obligations on the will. *Virtue legalism* is the view that the moral life involves conformity to a moral law containing obligations; and that, in particular, we are under an obligation to be virtuous or morally good.

Now moral goodness does seem to depend, not on our ordinary first-order actions and their consequences, but on the motivation behind those actions. The morally good person is not the person who actually helps others, but rather the person who is motivated to help others if they can. One's moral goodness reflects, after all, how far one is swayed to appropriate first-order action by morally relevant considerations, such as the needs of others. It follows then, that an obligation to be morally good is an obligation to be motivated to perform certain first-order actions.

Now one very natural view is that the only proper object of obligation is action, or refraining from action. We can only be obliged, just as we can only be commanded, to do things or refrain from doing them. We cannot be under an obligation to have things simply happen to us, independently of any action or omission on our part. Suppose then that any moral obligation must be an obligation on action; then, if moral virtue lies in being motivated to perform certain first-order actions, and this moral virtue is obligatory, it follows that such motivation must itself arise from an action-motivating, second-order agency of the will.

Virtue legalism has a long history. Writers as diverse as Abelard (1136-9) and Leibniz (1706), the Stoics and Richard Cumberland (1672), teach that the moral law is distinguished from merely human laws precisely by virtue of its demanding, not simply conformity to rules governing our external behaviour, but the right disposition of will.

Recent moral philosophy has seen a revival of the view that morality presupposes an agency of the will - but this time outside virtue legalism. David Gauthier (1986) has argued that conformity to moral obligations, say to altruism, can be rationally justified in terms of self interest. While altruistic first-order actions may not be in one's interest, the disposition to perform them may be, in that holding it secures the cooperation of others. In which case self interest makes it rational to secure that cooperation by disposing oneself - through a decision of the will - to altruism. The rationality of the disposition formation *qua* decision to act is then supposed to imply the rationality of the altruistic action decided on, should the occasion for performing it ever arise.

Gauthier's theory reveals a tension in our ordinary thinking about rationality and the will. If taking a particular decision is itself an action, then can it not be made rational, like any action, by its expected good consequences? But these expected good consequences may arise, as in Gauthier's example, independently of performance of the action decided upon. So the thought that decisions to act are actions themselves seems to imply that the rationality of a decision to act is to be determined as Gauthier determines it - in terms of the decision's consequences, and not those of the action decided upon, and so other than by reference to a prior and independent rationality attaching to the action decided upon.

But we also think that the rationality of deciding to do A *does* depend wholly on a prior and independent rationality attaching to doing A: since the function of decisions to act is to apply reason as it concerns the actions decided upon, what makes it rational to decide to do A in particular can only be the prior rationality of doing A - in which case, of course, the rationality of deciding to do A would indeed (as Gauthier also wants to maintain) imply the rationality of doing A thereafter.

The natural idea that decisions serve to apply reason as it concerns the actions decided upon is not obviously consistent with the equally natural idea that decisions are themselves deliberate actions. This is a deep and unresolved problem in our conception of rational agency and the will.

4 Scepticism about actions of the will

Scepticism about actions of the will developed when, in the early modern period, theorists abandoned a practical reason-based theory of human action, characterizing human action instead as, hitherto, animal action had been characterized - in purely motivational terms, as a product of appetite.

Something at least analogous to human action occurs in animals. We talk of animals performing actions for purposes, such as stalking in order to catch prey. But if animals are not rational agents, their actions cannot be

characterized as exercises of practical rationality. And so the same philosophers who proposed practical reason-based theories of human action, also proposed importantly different, motivation-based accounts of animal action. Animal action was merely the purposive product of appetite: it was performed simply as a believed or perceived means to desired ends, whether for its own sake or as a means to further ends. A non-rational animal's A-ing counted as its action in virtue of its being moved to A by some appetite for a given end, and a belief or perception that doing A would or might further that end.

We have here an analogue of human action, simply because so much human action is purposive in its motivation too. It is just that this motivation in humans was supposed to come from an exercise of rationality. Whereas in animals, the appetites that motivated were supposed non-rational in origin.

Notice, however, that not all human actions are purposive. Particular decisions to act, even if actions themselves, still are not taken as means to ends. My decision to do A rather than B will characteristically be based on beliefs about what ends would be furthered by *doing* A rather than B - not on the basis of beliefs about what ends would be furthered by *deciding to do* A rather than B. In deliberation we consider as means to ends, not our decisions to act themselves but the actions between which we must decide. If deciding to perform a particular action counts as a deliberate action itself, that must be because the decision counts as an exercise of practical rationality, and not because of its purposiveness.

From the sixteenth century on, we find increasing numbers of theorists characterizing all action, human and animal, in terms of a purposive motivation. And the will itself was reconceived, no longer as a distinctively rational appetite peculiar to humans, but as a simple locus of action-motivating passion found in animal and human alike. In other words, all action, human and animal alike, was characterized as a product of a will to perform it. As Hobbes put it in *Leviathan* chapter 6, 'a Voluntary Act is that, which proceedeth from the Will, and no other'. Since the willing of a particular action is not plausibly purposive itself, or willed - on any view, we do not generally take particular decisions to act on the basis of having decided so to decide - the will itself ceased to appear to be a locus of agency or freedom.

What encouraged the abandonment of a practical reason-based conception of human agency? Relevant developments in early modern thought are: (a) scepticism about human practical rationality; (b) projects of reconceiving rational human action in the same terms as animal action; and (c) scepticism about practical reason itself.

(a) Thanks to Augustine, it was generally agreed by Western Christians that Adam's Fall had seriously damaged our practical rationality. Now, on many views, fallen humans still retained some capacity for practical rationality, the will remaining a rational appetite. But the Reformation popularized a radical conception of the Fall that denied the practical rationality of fallen humanity outright. John Calvin, for example, admitted that as fallen we retained an intellect which distinguished us from the animals. We still formed rational judgments about which actions were justified. But the Fall had wholly removed our executive capacity to apply these rational judgments, which no longer affected what actions we were motivated to perform. Human action now took the same appetite-driven form as animal action, and the human will was no longer a rational appetite or locus of free agency (see Calvin 1559 Book II, Chapter 2).

(b) Unlike Calvin, Thomas Hobbes allowed that humans could still be practically rational as animals were not. However, Hobbes also wanted to develop an economical action-explanatory theory in materialist terms, characterizing rational human action in the same terms as non-rational animal action: so not as an exercise of a distinctively human practical rationality, but as a will-motivated, purposive doing. For Hobbes, human practical rationality had done nothing to transform human action into anything different in nature from animal action. Our practical rationality came with language - reasoning, in Hobbes' view, being just the reckoning of the implications of the general terms of a language. Language on this view was simply an invented tool which served to record and express action-motivating beliefs and appetites that were themselves language-independent. As language users, reasoning beings did not have either a psychology or a capacity for agency different in kind from that found in non-reasoning beings, any more than the actions of those who enjoy the invention of another communicative tool, writing, are different in kind from the actions of those who lack that tool.

(c) A final source of support for a motivation-based account of human agency is, of course, scepticism about

practical reason itself. There may seem little point in characterizing our agency as an exercise of rationality if, in fact, as Hume expressly claimed, our agency is not governed by reason at all. It is no surprise then that for David Hume, far from being a locus of rational appetition and agency, the will was simply a phenomenological marker for passion-motivated action: as he put it in the *Treatise* in 'Of liberty and necessity': 'By the *will*, I mean nothing but *the internal impression we feel and are conscious of, when we knowingly give rise to any new motion of our body, or new perception of our mind*'.

See also: Action; Free will; Intention; Practical Reason and Ethics; Virtue ethics

THOMAS PINK

References and further reading

Abelard, P. (1136-9) *Ethics*, trans. and ed. D. Luscombe, Oxford: Clarendon Press, 1971.(Referred to in §3; states a legalism of merit or moral worth in the form of an ethics of obligations on the will.)

Aquinas, T. (1265-74) *Summa theologiae*, ed. T. Gilbey, with Latin and English texts and notes, London and New York: Blackfriars, 60 vols, 1962-.(Referred to in §1; Aquinas's theory of action is given in volume 17.)

Aristotle (*c*. mid 4th century BC) *Nicomachean Ethics*, trans. and notes T. Irwin, Indianapolis, IN: Hackett Publishing Company, 1985.(Referred to in §1; *prohairesis* is characterized in books 3, 6 and 7.)

Calvin, J. (1559) *Institutes of the Christian Religion*, trans. F.L. Battles, ed. J.T. McNeill, Philadelphia, PA: Library of Christian Classics, 1960.(Referred to in §4; Calvin's account of fallen human action and the will is given in volume 1, book II, chapter 2.)

Cumberland, R. (1672) *A Treatise of the Laws of Nature*, selections in J.B. Schneewind (ed.) *Moral Philosophy from Montaigne to Kant*, Cambridge: Cambridge University Press, vol. 1, 1990.(Referred to in §3; chapter 8 contains a statement of virtue legalism as implying obligations on the will.)

Gauthier, D. (1986) *Morals by Agreement*, Oxford: Oxford University Press.(Chapter 6 contains the argument for the rationality of altruistic actions from the rationality, in self interested terms, of deciding to act altruistically. This argument is referred to in §3.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.(Referred to in §4. Chapters 4-5 provide Hobbes' account of human rationality as language-involving. Chapter 6 his account of human and animal action as uniformly language and reason independent.)

Hume, D. (1739-40) *Treatise of Human Nature*, ed. L.A. Selby-Bigge, revised P.H. Nidditch, Oxford: Oxford University Press, 1978.(Referred to in §4; see Book 2, Part 3, Section 3 'Of liberty and necessity' where the will is characterized merely as a phenomenological marker for subsequent action.)

Kahn, C. (1988) 'Discovering the Will', in J. Dillon and A. Long (eds) *The Question of 'Eclecticism'*, Berkeley, CA: University of California Press.(Compares conceptions of motivation and will in Aristotle, the Stoics and Aquinas.)

Kent, B. (1996) *Virtues of the Will: The Transformation of Ethics in the Late Thirteenth Century*, Washington, DC: Catholic University of America Press.(Describes the medieval conflict between intellectualist and voluntarist views of the will.)

Leibniz, G.W. (1706) *Opinion on the Principles of Pufendorf*, in *Leibniz: Political Writings*, ed. P. Riley, Cambridge: Cambridge University Press, 1972.(Referred to in §3; in Section 3 Leibniz defends, against Pufendorf, the virtue legalistic view that natural law addresses the internal movements of the soul, and not merely external behaviour.)

Nagel, T. (1976) 'Moral Luck', in *Mortal Questions*, Cambridge: Cambridge University Press, 1979.(Referred to in §3. Nagel considers the view that morality is concerned with the will, rather than external action, because external action depends on luck, while morality excludes luck.)

Pink, T. (1996) *The Psychology of Freedom*, Cambridge: Cambridge University Press.(Expansion of the material in §§2-3. The book explains how common sense psychology conceives of decision making as a second-order, action-generating action; and how our freedom of action depends on a freedom specifically of decision-making.)

Pink, T. (1997) 'Reason and Agency', *Proceedings of the Aristotelian Society*, 263-80.(Expansion of the material in §§1 and 4. The article compares a practical reason-based conception of action, which allows for an agency of the will, with motivation-based conception, which does not.)

William of Auvergne (*c.*1180-1249)

Active in Paris during the third and fourth decades of the thirteenth century, when universities were emerging as centres of Western European intellectual life, William played a decisive role in the early development of high medieval philosophy. His writing reveals a familiarity with Aristotle, all of whose major works except the Metaphysics were readily available in Latin translation, and with the Islamic philosophers, most especially Avicenna but also Averroes, whose commentaries on Aristotle were just beginning to circulate. William looked back to the Neoplatonic traditions of the twelfth century, but he also looked ahead to the late-thirteenth-century Aristotelianizing that he and his contemporary, Robert Grosseteste, did so much to promote.

Born in Aurillac, France, around 1180, William of Auvergne went to Paris to study arts and then theology, in which field he became a university master in the 1220s. In 1228 he was named bishop of Paris and ex officio the ecclesiastical overseer of the university. He died in that post in 1249. His major philosophical work is a massive and sometimes rambling collection of treatises called the *Magisterium divinale ac sapientiale (Teachings on Divinity and Wisdom)*, the exact contents of which are debated but which surely included the important *De universo (On the Universe)* and possibly *De anima (On the Soul)*. Parts of the *Magisterium* date from the 1220s, but the finishing touches were applied in the 1230s and perhaps, in the case of *De anima*, shortly after 1240.

On the surface, William presented himself as a resolute defender of the largely Augustinian principles of Christian tradition (see Augustinianism) against what he saw as the alternately dangerous and absurd presuppositions of the pagan philosophers, most especially Plato and Aristotle. His metaphysics was consciously anti-Platonic, for although he accepted the exemplarist notion that the world was fashioned after archetypal ideals - which he, like Augustine, interpreted as ideas in the mind of God - he resolutely opposed Plato's preoccupation with the tenuousness of material reality. His position was manifested most clearly in semantics, where he rejected the contention that the real reference of the terms of normal speech was more directly to the archetypes than to the concrete world around us. Like Aristotle, he held that simple terms referred to real essences in the actual world (see Language, medieval theories of). Yet William's concrete realism likewise spurned the Aristotelian notion that individuals were the instantiation of universal essences particularized by material accidents. Instead, he maintained that real essences were in themselves completely particular and fully specific, so that individuals differed not just by their material accidents but also by their own individual essential cores. The idea remained an important subcurrent in medieval thought, emerging again with particular force in the post-Aristotelian metaphysics of the fourteenth century.

In his philosophy of mind, William also claimed to be setting himself against the Greek models. His metaphysics of essence ruled out Platonic reminiscence, for knowledge was of and from essences in the actual, material world. Yet he also attacked the Aristotelian notion of an agent intellect providing the mind with the forms of objects to be known. On this point he was reading Aristotle through the eyes of Avicenna (see Ibn Sina), who posited a separate agent intellect, the tenth of the celestial intelligences and immediate creator of the human mind. In contrast, William proposed an Augustine-inspired intellectual occasionalism by which the mind, naturally capable of perceiving essences directly but handicapped in the world of sin, drew from sensation the material for an almost discursive penetration beyond accidents to the essential nature. For this the mind needed no separation into active and passive functions and did not have to be made subject to material impressions from the level of sensation, but merely relied on its own dynamic capacity to grasp the objects of intelligibility. His explanation of the procedure whereby it did so depended on the Aristotelian notion of abstraction of the universal from the particular and provided succeeding thinkers with a theoretical base from which to develop the idea at length.

For all his rhetorical anti-peripateticism, therefore, William drew extensively upon Aristotelian perspectives. However, one last aspect of his philosophy of mind provided later Neoplatonizing scholastics such as Roger Bacon with reason to claim him as a defender of their view that God was the mind's agent intellect and thus its source of knowledge. In the case of knowledge of selected terms, most significantly those that entered into the first, self-evident principles of science and truth - terms such as 'true', 'good', 'whole' and 'part' - William maintained that the human intellect was indeed required to turn to God, who poured into it the formal elements of such cognition by a kind of intelligible illumination. In this restricted sense, God was the agent intellect William

otherwise so vehemently opposed; and to this degree he sowed the seeds of Augustinian illuminationism of the latter part of the thirteenth century.

In epistemology, William was more completely faithful to the emergent Aristotelianism of his day (see Aristotelianism, medieval). Knowledge of the truth was obtained by comparing a proposition to actual conditions in the external world, due account being taken for temporal determination and proper quantification. A true proposition states what actually has been, is or will be the case. He furthermore adopted the basic elements of Aristotle's apodictic theory of science, referring his readers to the classic expositions of scientific knowledge in the *Physics* and the *Posterior Analytics*. His account served along with that of Grosseteste as a primary conduit for the reception of such views in the Latin West. 'Science' was certain knowledge of universal conclusions demonstrated from even more certain premises, the whole structure being ultimately reducible to indemonstrable, sometimes self-evident, first principles. This foundationalist model was dominant throughout the rest of the Middle Ages.

However, it is important not to overlook an empiricist side to William's views on science. Like most Aristotelians of the thirteenth century, he thought science - or, in his case, that part of science not traceable to God's illuminated principles - was derived by induction from experience in the world. There is moreover in his works a dramatically non-Aristotelian notion of experience that harked back to Arabic and Hebrew lore on magic and the occult. The real 'experimenters' were for him natural magicians, whose ability to manipulate hidden forces in the natural world held out the promise of marvellous accomplishments. Here again was a theme of major importance for later scholastic thought, picked up with fervour by Roger Bacon and many of the so-called Perspectivists.

See also: Aristotelianism, medieval; Augustinianism; Grosseteste, R.; Natural philosophy, medieval

STEVEN P. MARRONE

List of works

William of Auvergne (*c*.1220-after 1240) *Guilemus Alverni Opera omnia (Complete Works of William of Auvergne)*, Orleans: Hotot, 1674; reprinted Frankfurt: Minerva, 1963, 2 vols with supplement.(Contains the treatises variously assigned to the *Magisterium*, including *De anima*.)

William of Auvergne (*c*.1223) *De trinitate (On the Trinity)*, ed. B. Switalski, Toronto, Ont.: Pontifical Institute of Mediaeval Studies, 1976; trans. R.J. Teske and F.C. Wade, *The Trinity*, Milwaukee, WI: Marquette University, 1989.(Theological treatise, also important for psychology and theory of mind.)

William of Auvergne (*c*.1225-*c*.1235) *De immortalitate animae (On the Immortality of the Soul)*, ed. G. Bülow, *Des Dominicus Gundissalinus Schrift von der Unsterblichkeit der Seele*, Beiträge zur Geschichte der Philosophie des Mittelalters II, 3, Münster: Aschendorff, 1897; trans. R.J. Teske, *The Immortality of the Soul*, Milwaukee, WI: Marquette University, 1991. (This treatise, attributed by Bülow to Gundisalvi, is now generally held to be by William and is extant in two versions.)

William of Auvergne (*c*.1225) *De bono et malo (On Good and Evil)*, ed. J.R. O'Donnell, *Mediaeval Studies* 8, 1946: 245-99.(Besides ethics, contains much on theory of mind.)

William of Auvergne (after 1228) *Tractatus secundus de bono et malo (Second treatise on Good and Evil)*, ed. J.R. O'Donnell, *Mediaeval Studies* 16, 1954: 219-71.(Similar to the earlier treatise of the same name.)

References and further reading

Baumgartner, M. (1893) *Die Erkenntnislehre des Wilhelm von Auvergne (The Epistemology of William of Auxerre)*, Münster: Aschendorff.(Still reliable despite its early date.)

Jüssen, G. (1987) 'Wilhelm von Auvergne und die Transformation der scholastischen Philosophie im 13. Jahrhundert' (William of Auvergne and the Transformation of Scholastic Philosophy in the Thirteenth Century), in J.P. Beckmann *et al.* (eds) *Philosophie im Mittelalter*, Hamburg: Meiner, 141-64.(The best overview of William's place in the history of scholasticism.)

Marrone, S.P. (1983) *William of Auvergne and Robert Grosseteste: New Ideas of Truth in the Early Thirteenth Century*, Princeton, NJ: Princeton University Press. (Emphasizes William's epistemology and relation to Aristotelian 'science'.)

Masnovo, A. (1945-6) *Da Guglielmo d'Auvergne a s. Tommaso d'Aquino (From William of Auvergne to St Thomas Aquinas)*, Milan: Vita e Pensiero, 3 vols.(Pathbreaking but sometimes unreliable.)

Moody, E.A. (1975) 'William of Auvergne and his Treatise *De anima*', in *Studies in Medieval Philosophy, Science, and Logic: Collected Papers, 1933-69*, Berkeley, CA: University of California Press, 1-109.(Excellent

analysis of William's psychology.)
Rohls, J. (1980) *Wilhelm von Auvergne und der mittelalterliche Aristotelismus (William of Auvergne and Medieval Aristotelianism)*, Munich: Kaiser.(Places William in the context of thirteenth-century theology.)
Teske, R.J. (1994) 'William of Auvergne on the Individuation of Human Souls', *Traditio* 49: 77-93.(Locates William's arguments concerning the nature of the soul precisely in historical context.)

William of Auxerre (1140/50-1231)

William's career spans the decades at the end of the twelfth century and the beginning of the thirteenth century during which the newly recovered Aristotelian natural philosophy, metaphysics and ethics and the newly available works of great Muslim thinkers such as Avicenna and Averroes brought enormous energy and upheaval to intellectual culture. William's own views are traditional, owing their largest debts to Augustine, Boethius and Anselm. However, his major work, Summa aurea, is an influential precursor of the monumental systematic theological treatises that followed half a century later.

William of Auxerre (Guillelmus Altissiodorensis) was a master of theology at Paris. In 1231 Pope Gregory IX appointed him to a commission charged with correcting the works of Aristotle that had been proscribed at Paris in 1210, but William died a few months later and the commission seems never to have carried out its assignment. Of William's work, we possess only *Summa aurea (The Golden Synopsis [of Theology])* composed sometime between 1215 and 1225, and *Summa de officiis ecclesiasticis (Synopsis of Ecclesiastical Offices)*, a detailed description and explanation of the offices and liturgical practices of the church.

Summa aurea is modelled generally on Peter Lombard's *Sentences* (see Lombard, P.) Like the *Sentences*, the *Summa* consists of four books devoted, respectively, to the unity and trinity of God, the created realm (especially human beings and humanity's fall through sin), the incarnation of the Son and the redemption and reparation of fallen humanity, and the sacraments. However, the *Summa* diverges from the *Sentences* significantly in ways that point toward future developments in medieval philosophical thought and literature.

William has broad systematic and philosophical interests reflective of and sensitive to the rapidly expanding intellectual horizons of the early thirteenth century. In the prologue to *Summa aurea*, for example, William reflects on the nature of his project and the methodology appropriate to it, raising the question of the relation between faith and what he calls 'natural reasoning'. His Augustinian position is optimistic about the usefulness of what we would characterize as broadly philosophical methods in theological inquiry: one ought not to believe only what can be established by natural reasoning or expect natural reasoning to be entirely adequate for understanding divine matters, but natural reasoning can support and confirm faith, defend it against heresy and move simple-minded people toward the true faith.

The openness to philosophical reasoning and investigation expressed in the prologue goes some way toward explaining William's willingness to depart from the narrowly doctrinal and theological interests of Lombard's *Sentences* to develop systematic philosophical foundations for the theological issues he wants to address. Whereas Lombard's treatment of God (in Book I of the *Sentences*) begins with a discussion of the mystery of God's trinity-and-unity, William takes up the doctrine of the Trinity only after independent discussion of four proofs for God's existence (in the first treatise of Book I) and arguments for God's unity, simplicity and eternality (in the second treatise).

In Book III, concern for philosophical depth and systematic procedure lead William to replace Lombard's collection of assorted matters associated with the theological virtues with an enormous systematic discussion of moral issues. William begins (in treatise 10) by raising issues prefatory to a discussion of the virtues (*quaestiones praeambulae ad virtutes*). Among these prefatory issues is one concerning the nature of the good in general: 'The fourth question prefatory to the discussion of virtue is a question about the good. This must be dealt with prior to dealing with virtue because good is placed in the definition of virtue. Moreover, the good is the end of virtue, and the end is conceptually prior to what is directed toward the end' (Book III, treatise 10, ch. 4). William's discussion of the nature of the good is the seed from which grow the famous thirteenth-century treatises on the transcendentals by Philip the Chancellor, Alexander of Hales, Albert the Great, Thomas Aquinas and Bonaventure. William goes on to discuss virtue in general (treatise 11), the theological virtues (12-6), natural law (18), the political virtues (19), cardinal and other moral virtues (20-9), the gifts of the Holy Spirit (30-4) and the beatitudes (35). In all, William's treatment of these matters takes up nearly the whole of Book III (46 of its 55 treatises, 900 of the 1068 pages in the critical edition), and raises moral theory in the Middle Ages to a new level of philosophical sophistication. The immediate influence of this part of *Summa aurea* can be clearly seen in both Philip the Chancellor's *Summa de bono* and Albert the Great's *Summa de bono*.

See also: Albert the Great; Philip the Chancellor

SCOTT MacDONALD

List of works

William of Auxerre (1215-25) *Summa aurea (The Golden Synopsis)*, ed. J. Ribaillier, Grottaferrata: Editiones Collegii S. Bonaventurae, 1980-7.(William's broad summary of theological and philosophical doctrine, modelled on Lombard's *Sentences*.)

References and further reading

Lottin, O. (1942-60) *Psychologie et Morale aux XIIe et XIIIe siècles*, Gembloux: Duculot.(Several important studies in the first four of these six volumes discuss William's contributions to medieval thought.)

MacDonald, S. (1992) 'Goodness as Transcendental: The Early Thirteenth-Century Recovery of an Aristotelian Idea', *Topoi* 11: 173-86.(An examination of William's and Philip the Chancellor's discussions of the metaphysics of goodness.)

Ribaillier, J. (1967) 'Guillaume d'Auxerre', *Dictionnaire de Spiritualité* vol. 6, ed. Gabriel-Guzman, Paris: Beauchesne, cols. 1192-9.(A fuller account of William's life and work than that provided by the editors in the introductory material in Ribaillier's edition of *Summa aurea*.)

St. Pierre, J.A. (1966) 'The Theological Thought of William of Auxerre: An Introductory Bibliography', *Recherches de théologie ancienne et médiévale* 33: 147-55.(A useful bibliographical essay.)

William of Champeaux (*c.*1070-*c.*1120)

William studied under Anselm of Laon and became one of a number of famous teachers of logic, rhetoric, grammar and theology in early twelfth-century France, teachers who helped to establish the schools which eventually turned into the University of Paris. He is perhaps best known for his dispute with his young pupil Peter Abelard over the reality of universals, a debate which William lost so badly that most of his students elected to be taught by Abelard instead.

William's position on the nature of universals (genera and species) was that they were real things that were determined by accidents in much the way that some amorphous material is determined by shapes accidentally accruing to it (see Universals). Thus a highest genus, such as substance, becomes determined to its various species, such as body and spirit, by different, opposed accidents which it possesses simultaneously. Like many thinkers at this time, especially among the theologians, William accepted the Porphyrian interpretation of Aristotle's *Categories*, according to which the categories provide a 'tree-structured' classification scheme for all entities (see Porphyry). Actually there are ten different 'trees' corresponding to each of Aristotle's ten categories, and each successive 'branching' of each tree leads to more and more specific species of entities in the category in question. Each tree culminates in the individuals within the category, that is, the things that do not undergo any further 'branching' (see Categories §1).

William treated the mutually opposed differences, which create the species that immediately branch off from a genus, as accidents of that genus, and then went on to allow for individuating differences, which create the individuals that branch off from a final species and are mutually opposed accidents of that species. Just as one individual is distinguished from another individual of the same species only by characteristics that are accidental to that species, so each species under the same genus is distinguished by forms accidental to that genus. William's view encourages one to see each individual as ontologically constructed of successive layers of forms attaching themselves to a core that is like the highest genus under which the individual is subsumed.

Abelard quickly pointed out that this view defied the principle that opposites cannot simultaneously belong to one and the same thing, and that it also entailed that there were really only ten distinct things in the world, the ten highest genera, or categories. In William's view, everything turns out to be either one of those highest genera without any accidents or one of them with layered accumulations of accidental forms. Naturally, few were willing to accept such consequences, and William's form of realism ceased to be a live option among medieval thinkers once Abelard's critique became known.

According to Abelard, William also held that simple categorical sentences of the form of 'Socrates is white' had two senses, a grammatical sense and a dialectical (that is, logical) sense. The former treats the sentence as identifying Socrates with some white thing, while the latter treats it as asserting that whiteness inheres in Socrates. Although, according to William, the dialectical sense is 'higher', it is in its grammatical sense that the sentence is true or false. William's realist approach to universals is again apparent in the talk of the inherence of some property.

William's realism surfaces again in connection with the doctrine of topics. The topic or *locus* on which some argument depends for its logical force is, according to him, a thing in the world the argument is talking about. For example, the argument 'Socrates is a man; therefore he is an animal' has as its *locus* the thing 'man'. The general principle behind the argument, the *maxima*, is, 'What the species is predicated of, so also is the genus', and William interprets this as multiply ambiguous. Each conditional proposition that exemplifies the *maxima*, such as, 'If Socrates is a man, Socrates is an animal', expresses one of the *maxima*'s meanings. The *maxima* is then a kind of schema for propositions rather than a proposition itself. According to Abelard, one reason William held this was that the *maxima* could prove inferences only if it were about the things those inferences concerned. Thus again realism shapes William's logical doctrines.

This controversy over universals is the result of very divergent views of how the genera, species and differences, mentioned so frequently in the Aristotelian writings, should be interpreted. Among logicians in the late eleventh and early twelfth centuries the view that universality is consequent on linguistic signification had been gaining

converts, while among theologians there was a pronounced tendency to see universality as a feature ontologically independent of thought and language. Abelard represents the former interpretation, and William the latter. William went on from his disastrous confrontation with Abelard to become a teacher of theology and eventually Bishop of Chalons-sur-Marne.

See also: Abelard, P.; Aristotelianism, medieval; Language, medieval theories of; Realism and antirealism; Universals

MARTIN M. TWEEDALE

List of works

Fragments of works probably by William of Champeaux can be found in Fredborg (1976) and Green-Pedersen (1974).

William of Champeaux (*c.*1070-*c.*1120) Commentary on *Rhetorica ad Herennium*, ed. K.M. Fredborg, 'The Commentaries on Cicero's *De Inventione* and *Rhetorica ad Herennium* by William of Champeaux', *Cahiers de l'Institut du Moyen-Age Grec et Latin* XVII, 1976, 1-39.(A recently discovered text by William; for advanced scholars.)

William of Champeaux (*c.*1070-*c.*1120) Commentary on Boethius' *De topicis differentiis*, in N.J. Green-Pedersen (ed.), 'William of Champeaux on Boethius' Topics according to Orleans Bibl. Mun. 266', *Cahiers de l'Institut du Moyen Age Grec et Latin* XIII, 1974: 13-30.(For specialists on the 'topics' literature.)

References and further reading

Fredborg, K.M. (ed.) (1976) 'The Commentaries on Cicero's *De Inventione* and *Rhetorica ad Herennium* by William of Champeaux', *Cahiers de l'Institut du Moyen-Age Grec et Latin* XVII: 1-39.(Some significant, recently discovered texts, evidently by William himself; for advanced scholars.)

Green-Pedersen, N.J. (ed.) (1974) 'William of Champeaux on Boethius' *Topics* according to Orleans Bibl. Mun. 266', *Cahiers de l'Institut du Moyen Age Grec et Latin* XIII: 13-30.(For specialists on the 'topics' literature.)

Green-Pedersen, N.J. (1984) *The Tradition of the Topics in the Middle Ages; the Commentaries on Aristotle's and Boethius' Topics*, Munich: Philosophia Verlag.(See pages 165- for a discussion of William's contribution to the medieval devlopment of 'topics', that is, non-formal logical inference.)

Stump, E. (1988) 'Logic in the Early Twelfth Century', in N. Kretzmann (ed.) *Meaning and Inference in Medieval Philosophy*, Dordrecht: Kluwer, 31-55. (Discusses a treatise composed in William's school as well as other known logical works of the period.)

Tweedale, M. (1988) 'Logic (i): From the Late Eleventh Century to the Time of Abelard', in P. Dronke (ed.) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press, 196-226.(Places William in the general context of logical and grammatical thought of his period.)

Tweedale, M. (1976) *Abailard on Universals*, Amsterdam: North Holland, 95-111.(An analysis of William's view of universals and Abelard's critique.)

William of Conches (*fl. c.*1130)

William of Conches - whom many historians have attached to the School of Chartres - was one of the early twelfth century's keenest commentators on Platonic texts, and wrote also on natural science. He believed in the harmony of Platonism and Christianity. He thought that the ostensibly pagan texts of Plato and his followers contained Christian truths which the interpreter needed to uncover, while Platonic (and more recent) science could help towards an understanding of the account of creation in Genesis.

Born in Normandy, William of Conches was already teaching and writing in the early 1120s and appears to have remained active until the 1150s. Where he taught is much disputed; he may have taught at Paris, but there is a little more evidence to link him with Chartres. Certainly his pupil John of Salisbury regarded him as at least the spiritual successor of Bernard of Chartres (see Chartres, School of). Like Bernard, William was a grammar teacher as well as an enthusiast of Plato, and it is around the twelfth-century conception of grammar that William's apparently diverse interests and writings cohere. Grammar involved detailed technical study of Latin (including the theoretical questions about semantics it raised) (see Language, medieval theories of). William's commentary on Priscian's *Institutiones grammaticae (Principles of Grammar)*, written when he was young, and revised by him as an old man, fulfilled this task, drawing extensively on the anonymous eleventh-century *Glosule* to Priscian.

The grammarian's work also included the detailed explanation of classical texts, and there was no clear distinction between literary and philosophical material. William is known to have glossed Juvenal, but he concentrated on Platonic works including Plato's own *Timaeus* (in Calcidius' translation), Boethius' *De consolatione philosophiae (On the Consolation of Philosophy)*, Macrobius' commentary on Cicero's *Somnium Scipionis (Scipio's Dream)* and probably Martianus Capella. The *Timaeus* and Macrobius were regarded as important sources for natural science, so it is not unexpected that William should have developed this interest independently, in his *Philosophia* and the dialogue *Dragmaticon* (between 1144 and 1149). In these works he also made use of - and sometimes combined or developed in an original way - medical and scientific sources translated from the Arabic and Greek.

By William's time, there was already a long tradition of assimilating Platonic texts by interpreting them allegorically in explicitly Christian terms (see Platonism, medieval). Indeed, at times he could simply borrow from the late ninth-century commentary on Boethius by Remigius of Auxerre. William differed from his predecessors in the thoroughness with which he applied the idea that the philosophers were using what he called *integumenta* (extended metaphors) to cloak the Christian truth. When pressed, he would admit that the pagan authors were indeed pagans and could not be trusted in everything they said, but usually he discovered a satisfactory reading.

His most daring interpretation was that which made Plato's World Soul an *integumentum* for the Holy Spirit. He shared this interpretation with Abelard, who may even have been his source. However, there is a revealing difference between the two thinkers' approaches. The identification was controversial because it could be taken to suggest that the Holy Spirit is lower than the other two persons of the Trinity. For Abelard, this had important implications about the knowability of God and the soundness of pagan virtues, which made it impossible for him to give up this identification. By contrast, William is willing to qualify, question and finally, after Abelard's condemnation in 1141, to drop the idea altogether. For him, the identification is no more than a reading of an ancient text, convenient but dispensable.

It was when classical texts provided him with scientific, rather than theological or philosophical ideas, that William took them most seriously and elaborated them for his own use. The account of creation which he gives in the *Philosophia* and the *Dragmaticon*, and also in the glosses to the *Timaeus*, shows how all things - not merely the heavens and earth, but also birds, fishes, animal and humans - came into being through the natural interaction of the four elements (fire, air, water and earth). Only the creation of the human soul required separate divine intervention. To potential critics who objected on religious grounds to such discussions, William argued that he was merely extending the account of creation in Genesis. The Bible says certain things were made: William is explaining *how* they were made. While remaining in the Catholic faith, he says, we should always look for a reason for things and only fall back on authority as a last resort. It is not enough to say 'God made it', because although God can act directly - as he does in miracles - he does not usually do so.

See also: Abelard, P.; Chartres, School of; Natural philosophy, medieval; Platonism, medieval

JOHN MARENBON

List of works

William of Conches (?1120-30) Commentary on Boethius' *De consolatione philosophiae*, in C. Jourdain, 'Des commentaires inédits de la Guillaume de Conches et Nicholas Trivet sur la Consolation de la philosophie de Boèce', *Notices et extraits des manuscrits de la bibliothèque impériale* 20, 2, Paris: Imprimerie impériale, 1865.(Contains extracts from William's commentary on Boethius' *De consolatione philosophiae*.)

William of Conches (?1120-30) *Philosophia mundi (Natural Philosophy)*, ed. G. Maurach, Pretoria: University of South Africa Press, 1980.(A treatise on the universe, nature, physiology and the soul.)

William of Conches (*c.*1130) *Glosae super Platonem (Glosses on Plato)*, ed. E. Jeauneau, Paris: Vrin, 1965.(Commentary on *Timaeus*.)

William of Conches (1144-9) *Dragmaticon*, ed. W. Gratarolus, *Dialogus de substantiis physicis*, Strasbourg, 1567; repr. Frankfurt: Minerva, 1967.(A version of the *Philosophia mundi* in dialogue form.)

References and further reading

Elford, E. (1988) 'William of Conches', in P. Dronke (ed.) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press, 308-27.(Mainly on William's scientific thought.)

Gregory, T. (1955) *Anima mundi. La filosofia di Guglielmo di Conches e la Scuola di Chartres (The World Soul: The Philosophy of William of Conches and the School of Chartres)*, Florence: Sansoni.(General study of William's thought and its context.)

Jeauneau, E. (1973) *Lectio philosophorum (Reading the Philosophers)*, Amsterdam: Hakkert.(Includes essays on William's Priscian and Macrobius commentaries.)

William of Ockham (*c*.1287-1347)

William of Ockham is a major figure in late medieval thought. Many of his ideas were actively - sometimes passionately - discussed in universities all across Europe from the 1320s up to the sixteenth century and even later. Against the background of the extraordinarily creative English intellectual milieu of the early fourteenth century, in which new varieties of logical, mathematical and physical speculation were being explored, Ockham stands out as the main initiator of late scholastic nominalism, a current of thought further exemplified - with important variants - by a host of authors after him, from Adam Wodeham, John Buridan and Albert of Saxony to the school of John Mair far into the sixteenth century.

As a Franciscan friar, Ockham taught theology and Aristotelian logic and physics from approximately 1317 to 1324, probably in Oxford and London. He managed to develop in this short period an original and impressive theological and philosophical system. However, his academic career was interrupted by a summons to the Papal Court at Avignon for theological scrutiny of his teachings. Once there, he became involved in the raging quarrel between Pope John XXII and the Minister General of the Franciscan Order, Michael of Cesena, over the poverty of the church. Ockham was eventually excommunicated in 1328. Having fled to Munich, where he put himself under the protection of the Emperor Ludwig of Bavaria, he fiercely continued the antipapal struggle, devoting the rest of his life to the writing of polemical and politically-oriented treatises.

Because he never was officially awarded the title of Doctor in Theology, Ockham has been traditionally known as the venerabilis inceptor*, the 'venerable beginner', a nickname which at the same time draws attention to the seminal character of his thought. As a tribute to the rigour and strength of his arguments, he has also been called the 'Invincible Doctor'.*

The core of his thought lies in his qualified approach to the old problem of universals, inherited by the Christian world from the Greeks through Porphyry and Boethius. Ockham's stand is that only individuals exist, generality being but a matter of signification. This is what we call his nominalism. In the mature version of his theory, species and genera are identified with certain mental qualities called concepts or intentions of the mind. Ontologically, these are individuals too, like everything else: each individual mind has its own individual concepts. Their peculiarity, for Ockham, lies in their representative function: a general concept naturally signifies *many different individuals. The concept 'horse', for instance, naturally signifies all singular horses and the concept 'white' all singular white things. They are not arbitrary or illusory for all that: specific and generic concepts, Ockham thought, are the results of purely natural processes safely grounded in the intuitive acquaintance of individual minds with real singular objects; and these concepts do cut the world at its joints. The upshot of Ockham's doctrine of universals is that it purports to validate science as objective knowledge of necessary connections, without postulating mysterious universal entities 'out there'.*

Thought, in this approach, is treated as a mental language. Not only is it composed of signs, but these mental signs, natural as they are, are also said to combine with each other into propositions, true or false, just as extra-mental linguistic signs do; and in so doing, to follow rules of construction very similar to those of spoken languages. Ockham thus endowed mental discourse with grammatical categories. However, his main innovation in this respect is that he also adapted and transposed to the fine-grained analysis of mental language a relatively new theoretical apparatus that had been emerging in Europe since the twelfth century: the theory of the 'properties of terms' - the most important part of the logica modernorum*, the 'logic of the moderns' - which was originally intended for the semantical analysis of spoken languages. Ockham, in effect (along with some of his contemporaries, such as Walter Burley) promoted this new brand of semantical analysis to the rank of philosophical method* par excellence*. In a wide variety of philosophical and theological discussions, he made sustained use of the technical notions of 'signification', 'connotation' and, above all, 'supposition' (or reference) and all their cognates. His distinctive contribution to physics, for example, consists mainly in semantical analyses of problematic terms such as 'void', 'space' or 'time', in order to show how, in the end, they refer to nothing but singular substances and qualities.*

Ockham's rejection of universals also had a theological aspect: universals, if they existed, would unduly limit God's omnipotence. On the other hand, he was convinced that pure philosophical reasoning suffices anyway for decisively refuting realism regarding universals, since all its variants turn out to be ultimately self-contradictory,

as he endeavoured to show by detailed criticism.

On the whole, Ockham traced a sharper dividing line than most Christian scholastics before him between theological speculation based on revealed premises and natural sciences in the Aristotelian sense, which are based on empirical evidence and self-evident principles. He wanted to maintain this clear-cut distinction in principle through all theoretical and practical knowledge, including ethics and political reasoning. In this last field, in particular, to which Ockham devoted thousands of pages in the last decades of his life, he strenuously defended the independence of secular power from ecclesiastical power, stressing whenever he could the autonomy of right reason in human affairs.

1 Life

William of Ockham was born around 1287, probably in the village of Ockham in Surrey, near London. Of his youth we know next to nothing, only that he joined the Franciscan Order (perhaps as a child), that he was ordained subdeacon in London on 27 February 1306 (an appointment for which he would normally have had to be at least eighteen), and that he was licensed to hear confessions on 19 June 1318 (for which he would have had to be at least 30). He must have studied Aristotelian philosophy - and especially logic - quite thoroughly before turning to theology in Oxford around 1310. Following the normal curriculum, he would then have spent the next seven years as a theology student and have lectured for two years (probably from 1317 to 1319) on Peter Lombard's *Sentences* (see Lombard, P.) The commentary that resulted is Ockham's first major work. It is a huge treatise in four books, the first of which was later revised by the author for publication and is known, for that reason, as the *Ordinatio*, while the other three, jointly known as the *Reportatio*, circulated merely in the form of a detailed, authorized set of students' notes.

This early, and strikingly original, work might already have raised misgivings in certain quarters at Oxford. Whether for that reason or some other, Ockham was not immediately granted the degree of master of theology, which would have entitled him to continue teaching in the field. He was sent instead to a Franciscan convent, possibly in London, to teach logic and physics, which he did from 1320 to 1324. This turned out to be a very productive period, during which he wrote his commentaries on Porphyry's *Isagoge*, Aristotle's *Categories*, *On Interpretation* and *Sophistical Refutations*, and at least part of his commentary on Aristotle's *Physics*. He also revised his *Ordinatio*, engaged in public disputations on various matters theological and philosophical, and wrote a few more treatises on especially delicate questions such as transsubstantiation and God's foreknowledge. In the same years he may also have begun (or even completed, according to some) his great *Summa logicae (Summa of Logic)*; originally intended for beginners, it eventually became one of the major reference books for the new nominalistic semantics and logic which was to be so influential in the fourteenth and fifteenth centuries.

In the summer of 1324, Ockham left England for Avignon. Most scholars think he was summoned to the papal court there by Pope John XXII for investigation of his teachings. In the previous year a Provincial Franciscan Chapter held in Cambridge had already publicly raised questions about some of his philosophical theses that could be deemed theologically dangerous. In 1325, the Pope established an advisory commission of six renowned theologians to examine fifty-one articles extracted from Ockham's commentary on the *Sentences*. The group included the French Dominican Durandus of St Pourçain, otherwise known for his nominalistic sympathies, and John Lutterell, the ex-Chancellor of Oxford and one of Ockham's fiercest opponents. In 1326 the commission gave a preliminary report, which, while identifying many of Ockham's opinions as erroneous, did not condemn any as clearly heretical. Apparently this did not satisfy the Pope, and a second report was prepared, much more along Lutterell's lines. Ockham became officially suspected of heresy, especially with regard to his theory of moral merits, which was accused of Pelagianism (see Pelagianism), and in 1327 Pope John XXII launched a formal inquisitorial process against him.

This process was never brought to a conclusion. However, Ockham in the meantime had got involved in other sorts of troubles. It is probable that during his first years at the Franciscan convent in Avignon, he continued to work on some of his philosophical or theological treatises, such as the *Summa logicae*, the *Quodlibeta septem (Seven Quodlibetal Questions)* and the *Expositio in libros Physicorum Aristotelis (Commentary on Aristotle's Physics)*. Towards the end of 1327 the Franciscan Minister General, Michael of Cesena, who had himself been recently summoned to Avignon by the Pope, ordered Ockham to study John XXII's bulls of 1322-4 on whether Christ and the Apostles had owned anything and on the poverty of the Franciscan Order in particular. The question was

delicate. The papal court was by then scandalously opulent, and John XXII was very much aware of the challenge to it by the advocacy of poverty that had been spreading in many Franciscan quarters in the previous decades. His series of bulls on the subject had been intended to put a stop to this movement. Comparing these with the Gospels and the writings of previous popes, Ockham boldly concluded that John XXII was himself heretical. Matters eventually got worse between Michael of Cesena and the Pope, and on the evening of 26 May 1328 five Franciscans, including Michael and William, fled from Avignon, fearing for their lives. Having narrowly escaped the Pope's soldiers, they reached Pisa by boat and joined another enemy of John XXII, the Emperor Ludwig of Bavaria. All five were immediately excommunicated.

From 1330 on, Ockham stayed at the Franciscan convent in Munich under the protection of Ludwig of Bavaria and devoted himself to the writing of political treatises. At first, the goal of Michael of Cesena's supporters was to convince the Christian community that John XXII ought to be deposed. Such was the context for Ockham's *Opus nonaginta dierum (Work of Ninety Days)*, written in three months in 1332 or 1333, and for the first part of his important unfinished dialogue on the distribution of powers within and outside the church, known simply as the *Dialogus*. In December 1334, John XXII died. Political discussions were opened between Ludwig of Bavaria and the new Pope, Benedict XII, during which Ockham's literary activity was brought to a halt. However, no agreement was reached, and from 1337 to 1342 he produced an impressive number of new writings in which he defended more and more explicitly the independence of the imperial and royal powers from those of the Pope. Especially worth mentioning are the *Tractatus contra Benedictum (Treatise Against Benedict)*, the *Octo quaestiones de potestate papae (Eight Questions on the Power of the Pope)* and treatises 1 and 2 (the only ones extant) of *Dialogus* III. After another hiatus at the time of Clement VI's accession to the papal throne, Ockham finally synthetised his political thought in the treatise *De Imperatorum et Pontificum potestate (On the Power of Emperors and Popes)* in 1346 or 1347. He died in April 1347, a few months before Ludwig of Bavaria. Contrary to an old legend, he had taken no steps toward submitting to the Avignon church.

2 Ontology: antirealism

The core of Ockham's philosophical system lies in the idea that only individual beings exist and that universals are nothing but signs, spoken, written or mental. This is what has been called his *nominalism*, a label that came to be associated with his thought towards the end of the fourteenth century (see Nominalism).

The theoretical problem it was intended to solve was provided by Porphyry's old questions about universals (see Porphyry): do species (such as 'horse') and genera (such as 'animal') exist by themselves outside the mind or not? Are they corporeal or incorporeal? And are they located in the individual beings or do they exist apart from them? By Ockham's time, Platonism, identified as the position according to which universals do exist outside the mind as incorporeal beings over and above the individuals that participate in them, had fallen into disrepute (see Platonism, medieval). Aristotle was credited with having definitively refuted it. Most philosophers, then, would follow Thomas Aquinas or John Duns Scotus in adopting some form or other of what we today call 'moderate realism', the thesis according to which universals exist outside the mind but only within the individuals that exemplify them. But this, in the eyes of Ockham, was 'the worst error in philosophy'; and his defence of his own ontology - which he always presented as orthodox Aristotelianism - consisted largely in a detailed, ruthless criticism of the different variants of moderate realism (see Aristotelianism, medieval).

For one thing, he deemed them all incompatible with God's omnipotence. If the universal 'horse' was a common part of each individual horse, then God could not completely destroy any single horse - Bucephalus, for example - without destroying all the others; for in order to destroy the universal 'horse' that was part of Bucephalus, God would have to destroy all the other horses as well. This overtly theological argument reveals one of Ockham's deepest philosophical intuitions: everything that exists, he insisted, should be logically independent of any other thing. This ontological atomism is what he frequently expresses by means of thought experiments involving God's omnipotence. Since God is credited with the power to do anything which is not self-contradictory - a fundamental theological principle, according to Ockham - the realm of what he can do corresponds to the whole array of logical possibilities, and he should, in particular, be able to destroy any single contingent being while keeping any of the others in existence.

However, Ockham's main accusation against the varieties of moderate realism he identified among his contemporaries is that in one way or another they all fall short of internal logical consistency. His chief weapon in

ontological disputes was the principle of non-contradiction. The best example of this line of attack lies in his famous critique of Duns Scotus' 'formal distinction'. Scotus, also a Franciscan, whose work Ockham deeply respected and often discussed at great length, had held that the universal natures such as equinity or animality did have some sort of existence outside the mind, but that they were not *in reality* distinct from the individuals that exemplified them. They were merely, Scotus contended, *formally* distinct from them. By this distinction he meant that while they could not exist in reality apart from their individual exemplifications - even by God's omnipotence - there nevertheless was a 'foundation in the thing' for thinking of them as separate. Scotus, like Ockham after him, was dissatisfied with any version of moderate realism that posited the universal nature as being both internal to the individual thing and really distinct from it, because this would entail that the individual thing and the universal nature should logically be able to exist apart from one another, which leads to absurdities. On the other hand, Scotus thought that the validity of the conceptual distinction between the universal and the individual did require that they should not be wholly identical with one another: hence his idea of a formal distinction within real identity.

According to Ockham, this leads directly to contradiction. If the universal nature and the individual thing are really identical to one another, then whatever is true of one of them should also be true of the other, and thus it will be true of the same thing that it is both universal and not universal, individual and not individual. Conversely, if incompatible predicates such as universality and individuality are attributed respectively to the common nature and the individual thing, then this in itself suffices to show that they are distinct from one another *in reality*, for this is precisely the paradigmatic way of proving that two things are really distinct from one another. Scotus' formal distinction thus collapses into the real distinction he wanted to reject in the first place. It is easy to see that this argument, like many others Ockham adduces in similar contexts (against Aquinas, for example), makes a crucial use of what we nowadays call the principle of the indiscernibility of identicals: if *a* and *b* are identical to one another, then whatever is true of *a* is true of *b*; conversely, if something is true of *a* which is not true of *b*, then *a* and *b* are really distinct from one another.

To avoid an old misunderstanding, it must be stressed that Ockham's critique of the ontology of universals does not explicitly rely on the methodological principle of economy that Sir William Hamilton, in the nineteenth century, labelled 'Ockham's razor': entities are not to be multiplied without necessity. It is true that Ockham's thought consistently shows a strong drive towards ontological economy and that he did on many occasions use the razor (which he himself formulated either as 'a plurality should never be posited without necessity' or as 'it is pointless to do with more what can be done with fewer'). However, this principle was commonly accepted before him, even by realist philosophers. It was thought of - by Ockham as well as by others - as a mere methodological rule which, taken by itself, yielded only 'probable' conclusions. By contrast, the refutation of the external reality of universals was considered by Ockham as absolutely conclusive since it rested on the principle of non-contradiction.

3 Ontology: substance, quality, form, matter

Although Ockham's nominalism posits only individual things in the external world, it nevertheless admits of ontological differences among them. First of all, he accepted (for a time) a fundamental difference of status between extramental and mental existence. While universals were refused the former, they were granted the latter. They were thought of, in the first version of Ockham's *Ordinatio*, as enjoying a special sort of intentional existence in the mind. This was the existence appropriate to a *fictum*, a mind-made intelligible object which existed only while consciously produced by the mind as an abstract correlate for acts of intellection, and which, for the duration of its existence, could serve as a natural sign representing certain individual beings. Admittedly, there was no room for this sort of existence within the framework of Aristotle's ten categories, but Ockham at first explained away this lacuna by restricting Aristotle's taxonomy to a limited theory of predicates for extramental entities. Later on, however - probably in the early 1320s, under the incisive criticism of his confrère Walter Chatton - Ockham realized that he could achieve an important ontological economy by letting the act of intellection itself play the role of the natural sign. Since internal acts, interpreted as actualized mental potentialities (*actus*), were commonly classified as qualities of the mind, no special ontological category was needed any longer for intramental existence. The intentional *fictum* could simply be dropped as soon as one recognized that intellectual acts themselves can serve as semantic contents without any need for internal correlates after all. This is a salient instance of Ockham's actually using the razor principle: after a period of hesitation, he finally favoured the *actus*

theory over the *fictum* theory as more probable because it could do with less what the other did with more.

However, even with this simplification, ontology was not utterly stripped of categorial plurality. For one thing, Ockham always maintained a fundamental distinction between substances and qualities. Although he interpreted Aristotle's theory of the ten categories as a classification of signs rather than things, he would nevertheless acknowledge, corresponding to it, a basic duality amongst individual entities themselves. Substances, whether corporeal or spiritual, were thought of along the Aristotelian line as autonomous beings, while qualities were held to depend upon them for their own existence. Qualities, Ockham would say, 'inhere' in the individual substance which displays them. They are not essential parts of it, however, and any one of them can be removed (by natural causes or by God's omnipotence) without the substance losing its identity. Each substance, on the other hand, has many qualities, simultaneously and successively, and each one has its own: the whiteness of Bucephalus is a particular thing, numerically distinct from Bucephalus itself and from all the whitenesses of any other beings in the world as well, however similar they might be. Ockham endorses particularism with respect to qualities.

Moreover, this nominalistic ontology is further complicated by the incorporation of Aristotelian hylomorphism, the theory of matter and form. Each singular composite corporeal substance is endowed with a variety of internal parts without which it would not be itself: a certain parcel of prime matter on the one hand, and one or more substantial forms on the other hand. Each human being, for example, is a compound of a determinate piece of matter and a number of substantial forms, such as a nutritive form, a sensory form and an intellectual form. Although these *partes essentiales* are naturally incapable of existing by themselves outside the hylomorphic compound, they are nevertheless counted by Ockham as real, and therefore singular, beings. Matter is thus granted a certain actuality of its own (in opposition to Aquinas). Forms, on the other hand, are stripped of intrinsic generality, each substance having its own.

Ockham, in summary, admits of four basic sorts of singular beings: substances, qualities, substantial forms and pieces of matter. Together they make up a still comparatively simple ontological system, of which singular substances constitute the core, qualities being attached to them as external properties, and forms and matter incorporated as internal parts.

What struck Ockham's contemporaries in this picture - and many commentators after them - was that it left out not only universals, but relations and quantities as well. These last exclusions, in fact, were a major source of Ockham's early troubles with religious authorities. The Catholic church was at the time imposing severe pressure on philosophical theories, for fear that philosophers might rule out any of its doctrines as impossible. The doctrines of the Trinity, the Incarnation and the Eucharist in particular, were commonly considered as having important implications for theories about relational connections (such as parenthood or ownership) and quantitative dimensions (such as weight, height or length). Ockham held that neither those connections nor those dimensions had any distinct existence of their own, and this aroused suspicion. However, Ockham had not meant to say that sentences making relational or quantitative claims were neither true nor false, or arbitrarily so. Most of them, he thought, are indeed true or false independently of the human mind: it is not the intellect, for example, that brings it about that Plato is taller or heavier than Socrates. What makes such sentences true or false, Ockham strenuously tried to show, is nothing but substances and qualities themselves. What characterizes a relational or a quantitative sentence, he would typically say, is not the sort of things needed to make it true, but its terms' modes of signifying. This last point he developed indefatigably and argued for in great detail. To the semantics of quantitative language, in particular he devoted many lengthy passages in his major works, as well as the bulk of two special treatises on the Eucharist, *De corpore Christi (On the Body of Christ)* and *De quantitate (On Quantity)*, which were obviously written as defences against theological attacks.

4 Epistemology: intuitive and abstractive cognition

Apart from pointing out the inconsistency and imprecision of his opponents, what Ockham had to do in order to accredit his simplified ontology was also to explain as clearly as possible how it fitted with Aristotelian science and with Christian doctrine. For this he had especially to account for the adequacy of knowledge in general. It was a respected Aristotelian adage that science had essentially to do with universals. How could it be salvaged if there are no such universals in reality? Ockham answered the challenge by providing a detailed theory of how mental universals are produced, a theory that was meant as the basis for any validation of conceptual knowledge.

The epistemological process, in this picture, always starts as a direct encounter between singular beings. When a knowing mind equipped with both sensory and intellective substantial forms is in the presence of a perceptible singular object such as a table or a horse, a causal chain naturally ensues, according to Ockham. First, there occurs an apprehension, or *intuition*, within the sensitive part of the mind. This in turn, together with the object itself, causes certain actual states, or *actus*, in the intellectual part, states which Ockham labels ‘intellectual intuitions’. These are characterized by the fact that they naturally elicit within the intellect true contingent beliefs about the world, especially judgments regarding existence or non-existence. If Socrates has the right relation to Bucephalus (perceiving it from the right distance, awake, sane and so on), there naturally takes place within Socrates a simple intellectual act, his ‘intuitive cognition’ of Bucephalus, which in turn causes his judgments that Bucephalus exists, is white, is standing there and so on.

Such judgments, of course, even though they are singular, require the availability of certain general concepts such as those of whiteness and existence. These, in Ockham’s view, are also brought about by the same concrete intuitions through an equally natural course. Every time Socrates forms an intuitive act caused by a certain individual in the world, there occur within Socrates’ mind other derivative simple intellectual acts, called *abstractive* cognitions. These new acts in themselves will not induce empirical judgments of existence; and for that reason they are said to ‘abstract’ from existence and non-existence. However, in virtue of a certain isomorphism (*similitudo*) they have with the original external thing, they may be used by the mind to represent or stand for this very same thing in its absence, and for any others as well that resemble it sufficiently for the same isomorphism to hold. General signs are thus naturally generated, either as mind-made objects of abstractive acts (according to the *fictum* theory) or as abstractive acts themselves (according to the *actus* theory). In either case, the result is what Ockham calls the mind’s concepts (*conceptus* or *intentiones*). General concepts, for him, are always causally derivative: ‘every naturally acquired abstractive cognition of a thing presupposes an intuitive cognition of the same thing’ (*Ordinatio*).

Intellectual intuitions and abstractive concepts, then, are simple mental terms or signs, each naturally representing one or more singular beings in the world. Once produced, these simple terms can be combined with each other, by natural causes or by the will of their possessors, to form cognitive complexes, such as mental propositions. A mental proposition is a structured sequence of mental terms that is either true or false with respect to the singular beings that are referred to by its terms. Such a propositional complex is formed by an intellect’s composite act of apprehension (*actus apprehensivus*). In the *fictum* theory, Ockham saw the mental proposition as the intentional correlate of this complex apprehensive act, but later he simply identified the two in the *actus* theory. Once the mind apprehends a proposition, it can go further and commit itself with respect to its truth or falsity; this is the act of assent, the judgment (*actus judicativus*). Mental propositions, finally, can be compounded with each other in chains of reasonings, theoretical or practical.

The detailed theory of exactly which structural features and relations with the world render a mental proposition true or false was part of logic for Ockham (see §7 below). Supposing that this semantical part of the programme can be achieved, the general picture we have of knowledge in Ockham’s final theory is that of a natural causal process, involving, apart from external objects themselves, only real qualities of the mind, and leading in stages from intuitive encounters with external singular beings to judgments and inferences involving complex propositional contents, each basic signifying unit in the process being a singular act of a singular mind. There is no longer any such thing as *the* concept ‘horse’, for example; only token concepts exist. Every mind forms its own single, momentary intellectual acts, which owe their generality, if any, to their signification rather than to a special mode of existence: ‘Nothing is universal save by signification’ (*Summa logicae* I, 14).

Such mental tokens normally will not subsist very long, but they leave traces when they disappear. We keep intellectual memories of them. Abstractive acts, Ockham contends, causally generate within the mind certain corresponding dispositions (*habitus*) to form similar acts in the future, and these cognitive dispositions are themselves identified with real qualities of the mind. Thus a general concept comes to be available to a particular mind for future use. When Socrates met Bucephalus, there naturally occurred within Socrates an intellectual intuitive cognition of Bucephalus; and then, as a causal result of this, abstractive cognitions were also produced, which turned out to be capable of simultaneously representing all beings in the world that are sufficiently similar to Bucephalus. In this way Socrates acquires the concept of a horse, the concept of a white thing and so on. Once formed, these general concepts elicit in turn mental dispositions within Socrates to produce replicas of them under

appropriate circumstances, new mental tokens representing all horses or all white things, just as the original acts did.

Accordingly, sciences - organized bodies of knowledge - are described by Ockham as 'collections of intellectual *habitus*' in the mind of particular knowers. To say of someone that they have a certain science - that of grammar, for example - is to say that they have internalized, as real qualities of their mind, a certain collection of intellectual dispositions which enable them to form general propositions of grammar when required and which adequately incline them to assent with certainty to those which are true: sciences are ultimately made up of apprehensive and judicative psychological dispositions. Ockham has often been taken to task for having taught that the objects of a science, like the objects of all knowledge or belief, are mental propositions rather than things themselves; but by this he only meant that mental propositions - corresponding to complex apprehensive acts - are precisely what knowers are inclined to assent to by the sciences they have internalized. The contents of knowledge are thus seen as sequences of mental propositions, of which certain terms naturally signify certain singular beings in the world. Thought is mental discourse.

5 Epistemology: direct realism

Ockhamism has often been suspected of leading to radical scepticism in epistemology. If general concepts represent nothing but independent individual things, how can such groupings fail to be arbitrary or misleading? But radical scepticism was, in fact, very far from Ockham's intentions and spirit, and his theory of universals and knowledge has been more accurately characterized by the best commentators as 'realistic conceptualism' (Boehner 1958) or 'direct realism' (Adams 1987) (see Epistemology, history of).

For one thing, as was seen in the previous section, the process of concept formation is completely natural. It is the same in every human being and leaves no room for idiosyncratic vagaries. Secondly - and most importantly - this naturally acquired stock of concepts does cut the world at its joints, according to Ockham. Species and genera, of course, are denied any outside reality as distinct beings, but Ockham readily grants that whether an individual thing is or is not of the same species or of the same genus as another is not for the human mind to decide. The conditions of being in the same genus or of the same species are not subjective or arbitrary. They are determined, on the contrary, by what the individual things are in themselves and how they in fact stand to each other. Whether something resembles Bucephalus enough to be naturally represented by the same general concepts does not, Ockham insists, depend upon the human mental apparatus at all. It simply is a rock-bottom fact of the universe. Concept formation, being a natural process, derivatively mirrors the natural distribution of causal powers among individual things, which powers depend in turn on real singular essences and qualities. In this lies the adequacy of abstractive cognition.

Different kinds of accurate, simple, general concepts are thus formed on the basis of direct experience, according to this theory. First, there are more or less general substantial concepts, such as 'horse' or 'animal', which are identified with species and genera and which are taken to represent in the mind certain groups of individual substances essentially similar to each other independently of their qualities. Second, there are concepts such as 'whiteness' or 'colour', which represent immediately perceptible singular qualities of substances but not substances themselves. Third, there is another sort of qualitative concepts, such as 'white' or 'coloured', which represent substances - white things or coloured things in the chosen examples - insofar as they possess certain perceptible qualities. Finally, there are some elementary relational concepts such as 'darker' or 'taller', which represent individual beings - substances or qualities - insofar as they are related to some others in certain perceptible ways. The first two groups are called 'absolute concepts', while the last two are classified among 'connotative concepts' (see §6), but all of these are simple mental terms naturally generated as causal results of empirical encounters with singular objects, without any combinatorial activity from the intellect. Together, they are held to provide an adequate basis for knowledge.

Is this empiricism? It would appear so, insofar as it holds that the formation of simple representative concepts in human minds is always triggered by direct singular experiences (of external objects or of mental states themselves). However, it must be stressed that the cognitive process described by the theory presupposes a very strong innate apparatus of rational capacities. A well-formed basic species-concept such as 'horse', for example, is automatically generated, according to Ockham, as the result of a single encounter with a horse. More general genus-concepts such as 'animal' are formed, in turn, as the results of a minimal number of encounters with

individuals of different species. Moreover, certain true contingent beliefs, such as ‘a horse exists’ or ‘this horse is larger than this dog’ are naturally triggered in the mind, as has been seen, as the results of causal processes which also involve highly determinate mental machinery. Certain universal truths are even supposed to be immediately known as such by the mind once the relevant concepts are made available to it, for example, that all horses are animals. The mind, according to Ockham, is equipped from the start with a rich apparatus of cognitive capacities which enable it, under the right circumstances, to reach true scientific knowledge.

One independent reason for linking Ockhamism with scepticism has sometimes been found in its acceptance of the possibility of an intuitive or pseudo-intuitive cognition of nonexistent beings (see Scepticism). However, two different theses must be distinguished in this respect, and neither of them, in Ockham’s view, leads to radical scepticism. First, Ockham did admit the theoretical possibility of an intuitive cognition of nonexistent beings, realizable only, he thought, by the supernatural intervention of God. Since an intuitive act is a real quality of the mind, he reasoned, it is itself a singular being distinct from its own external object; and hence, it is not self-contradictory that it should exist without the object. If this is not self-contradictory, then God can induce in a prophet’s mind, for example, a certain singular cognition of a presently nonexistent being, in virtue of which the prophet would know that this particular being - a future one, or even a merely possible one - does not presently exist, but that it will exist in the future or that it would exist under appropriate circumstances. This singular cognition would then be an adequate intuitive cognition since it would cause within the prophet’s mind true existential judgments about the thing in question. Accordingly, there is no hold for scepticism here: such a supernaturally induced intuition, although abnormal, would not be misleading.

The second possibility seems more disturbing. Ockham admits, since this is not self-contradictory either, that God could directly cause in any human being the *false* belief that there is a horse in front of them while in fact there is none. Strictly speaking, this would not be an intuitive cognition (it would not cause a true judgment of existence or non-existence), but it could certainly be an undetectable source of error. Such a possibility, however, does not belong to the range of natural processes; and, although it cannot be ruled out by the theory, it can in practice be disregarded by the epistemologist who wishes to assess the scope of natural knowledge. In the normal course of things, intuitive cognition is a totally reliable starting point for science.

6 Logic and philosophy of language: levels of discourse, connotation, intentions

Once concepts are available as natural signs, mental propositions can be assembled. Their truth and falsity will depend upon the semantical properties of their component terms and their syntactical arrangements. Thought is seen as an internal discourse with a compositional structure very similar to that of spoken sentences. Ockham took the ancient idea of mental discourse (found in Boethius, for example) more seriously than any other philosopher before him, and his logic systematically exploits, for the analysis of mental propositions and inferences, the theory of the properties of terms (*proprietates terminorum*) to which he had been introduced as a student. This theory was an original medieval contribution, which had been progressively developed since the twelfth century by a host of logicians such as Peter of Spain and William of Sherwood, for the analysis of spoken sentences. Ockham adapted it to his nominalistic ontology on the one hand, and enlisted its rich apparatus of concepts and distinctions on the other hand - the theory of *suppositio* in particular - in the service of a fine-grained understanding of mental computation. This is to a large extent what his logic is about (see Logic, medieval).

At the very outset of *Summa logicae*, Ockham states that there are three levels of propositions: spoken, written and mental. Those of the first two groups, he says, are composed of publicly perceptible signs, visible or audible; but mental propositions are sequences of concepts - or intentions of the mind - existing only in the privacy of the intellect. Each proposition is composed of terms, and the basic semantical property in this picture is the signification (*significatio*) of the component terms. The signification of a concept is natural, as has been seen, and that of the other two sorts of terms is conventionally derived from it: spoken terms are conventionally subordinated to concepts, and written terms are conventionally subordinated to spoken ones. A spoken or written term inherits the signification of the previously existing sign to which it is subordinated. When the Latin spoken word ‘*equus*’, for example, is conventionally subordinated to the mental concept ‘horse’, it thereby acquires the signification that belonged to that concept naturally; and the very same signification is afterwards transmitted in the same way to the corresponding written term. Ockham, then, refuses to say - as did Boethius and Aquinas, among others - that a spoken word signifies the underlying concept. It will instead be said to signify conventionally the individual things

themselves that the concept was a natural sign of in the first place: real singular horses in the case of the spoken words ‘horse’, ‘*cheval*’ or ‘*equus*’, for example. The same will hold, *mutatis mutandis*, for the corresponding written words.

At each of the three levels, many distinctions are to be drawn among terms with respect to their signification. First, categorematic terms are distinguished from syncategorematic ones by the fact that they do have a signification of their own before being combined with others into propositions. Proper names such as ‘Socrates’ and common nouns or adjectives such as ‘horse’ or ‘white’ are salient examples of categorematic terms. *Syncategoremata*, on the contrary, are particles such as prepositions, logical connectives and quantifiers, which do not when taken alone direct the mind towards particular individual beings in the world. Their semantical roles are ancillary: they do not signify anything by themselves but, when combined with categorematic terms, they can affect the modes of reference of those terms and determine the truth-conditions of propositions.

Another important distinction found in mental language as well as in spoken and written discourse, according to Ockham, is between absolute and connotative terms. The former correspond to what we today call natural kind terms, such as ‘horse’ or ‘animal’. These are characterized by the fact that they simultaneously and equally signify all the individuals they are true of, and nothing else: ‘horse’ signifies all horses and nothing but horses. Each connotative term, on the other hand, has at least two different groups of significates: its primary significates are the individuals the term is true of; and its secondary significates are individuals also called to mind by the term, but in an ‘oblique’ way. ‘White’, for example, is a salient case of a connotative term: it primarily signifies all white things and secondarily signifies - or connotes - their whitenesses; although it cannot be said that whitenesses are white, they are nevertheless obliquely called to mind by the term ‘white’. All relational terms, Ockham insists, are connotative: ‘owner’ primarily signifies all owners and connotes their possessions, ‘father’ primarily signifies the fathers and connotes their children, and so on. This idea of connotation thus turns out to be a crucial device for the simplification of ontology. All significates, whether primary or secondary, are singular substances or qualities, and nothing else. Relations, in particular, are no longer needed among real beings. The relevant complexity, here, is located in the semantical structure of certain terms rather than in the ontology.

Finally, some terms signify nonlinguistic things, while others signify signs. When restricted to mental terms, this idea is cashed out, in Ockham’s vocabulary, as a distinction between ‘first and second intentions’. First intentions are mental terms, the significates of which are external beings. Aristotle’s theory of the ten categories is precisely interpreted in this framework as a classification of first intentions according to their modes of signification: the category of relation, for example, is identified with the group of relational first intentions such as ‘owner’ or ‘mother’, the category of quality with that of qualitative first intentions such as ‘white’ or ‘whiteness’, and so on. Second intentions, on the other hand, are concepts signifying concepts. Philosophically, in Ockham’s eyes, the most interesting instances of these are the mental terms ‘species’ and ‘genus’ themselves. ‘Species’ is seen as a metalinguistic concept signifying certain general terms such as ‘horse’ or ‘dog’, while ‘genus’ amounts to approximately the same except that it signifies terms even more general, such as ‘animal’ or ‘flower’. This is exactly what realism about universals crucially misses: it fails to see that concepts referring *to* universals are fundamentally metalinguistic (see Language, medieval theories of).

7 Logic and philosophy of language: supposition and inference

Signification, then, in Ockham’s logic, is a pre-propositional property of categorematic terms taken in themselves. When such a term, whether spoken, written or mental, is inserted into a proposition as subject or predicate, it is said to acquire a new semantical property called *supposition*, which is derivative with respect to its signification but which, contrary to signification, can vary from one proposition to another according to contextual factors. Supposition, in modern terms, is the referential function fulfilled by the term in a given propositional context: for a term to ‘supposit for’ certain things, Ockham says, is just for it to stand for these things in a given proposition.

Normally a subject or predicate term supposits for its primary significates. It is then said to be taken in personal supposition (*suppositio personalis*): in ‘some horses are white’, for example, the subject term ‘horse’ personally supposits for - or refers to - all horses, whether white or not, while the predicate term ‘white’ personally supposits for all white things (but not for whitenesses, which are its secondary significates). Special cases can occur, however: in a sentence such as ‘horse is a five-letter word’, the subject term ‘horse’ does not stand for its primary significates - real horses - but for written tokens of the English word ‘horse’. When a term thus supposits for

spoken or written tokens of itself, it is said to be taken in material supposition (*suppositio materialis*). When it stands for tokens of the corresponding mental sign, such as ‘horse’ in ‘horse is a concept’, it is said to be taken in simple supposition (*suppositio simplex*). This last case is especially important in Ockham’s discussion of the problem of universals, since the subject terms of sentences such as ‘horse is a species’ or ‘animal is a genus’ are typically considered by him as being taken in simple supposition, and interpreted accordingly as referring in a special way to mental tokens rather than to mysterious external entities such as equinity or animality.

Which individual beings a term refers to in a certain propositional context is thus determined in the first place by the signification of the term and by whether it is taken in personal, simple or material supposition. However, this is still incomplete. It is also determined, Ockham insists, by the tense and the modality of the main verb in the proposition, according to precise rules. A present-tense verb, in particular, is said to restrict both the subject and the predicate of the proposition to supposit for individual beings which exist at the time of utterance. A past-tense verb restricts the predicate to supposit for past individuals, but allows the subject to supposit both for past individuals and for individuals existing at the time of utterance, and so on.

In this way, propositions are linked to reality. Discourse, whether spoken, written or mental, refers to nothing but singular entities, some of which are themselves significant tokens, and it does so through the referential functions of the terms in specific contexts. Propositions considered as complex semantical units do not, according to this view, signify special entities over and above the *supposita* of their terms. Their peculiarity instead is to have truth-conditions, a detailed theory of which Ockham gives in the second part of his *Summa logicae*. For elementary statements of the form ‘subject + copula + predicate’, these truth-conditions are entirely formulated in terms of relations between the supposition of the subject-term and that of the predicate. For the truth of a universal affirmative proposition, for example, it is both necessary and sufficient that the predicate supposits for all the individuals the subject supposits for. Similarly, the truth of a universal negative proposition requires that the predicate have no common *supposita* with the subject, and so on. The theory gets more complicated when it reaches modal propositions, non-elementary propositions (such as conjunctions, disjunctions or conditionals) or other difficult cases such as propositions involving special verbs like ‘cease’ or ‘begin’, which fascinated logicians in the fourteenth century. However, the basis is always provided by the referential link established through supposition between the terms in the proposition and the singular beings out there in the world.

The same is true generally of Ockham’s theory of inferences developed in the very long third part of *Summa logicae*. This is not formal logic in the modern sense. The validity of inferences is always made to rest in the last analysis upon the referential properties of the component terms. Syllogisms, notably, are described as special arrangements of terms, and their validity is said to depend on the connections between the purported suppositions of the major, the minor and the middle terms. The whole syllogistic theory, in Ockham’s view, ultimately rests, directly or indirectly, on only two relevant connections between terms: the *dici de omni* (being said of all) and the *dici de nullo* (being said of none). The former holds when a certain term *A* supposits for everything another term *B* supposits for, and the latter when *A* supposits for none of the *supposita* of *B*. ‘All horses are mammals, all mammals are animals, therefore all horses are animals’, for example, is based on the transitivity of the *dici de omni*. ‘No stone is an animal, all horses are animals, therefore no horse is a stone’ is based on an interplay between the *dici de omni* and the *dici de nullo*. All other syllogistic figures are reduced in one way or another to these two connections. In the same vein, nonsyllogistic inferences - which had come to be a major new field of interest in medieval logic under the heading of *consequentiae* - are also systematically analysed by Ockham as depending upon interconnections between the suppositions of terms. This is how he is led to accept the soundness of such direct inferences as ‘all animals run, therefore all horses run’ or ‘no material body exists, therefore no white thing exists’, the key to which lies not in the formal structure, but in the semantical relations between terms, and ultimately on their reference to singular beings (see Language, medieval theories of; Logic, medieval).

8 Physics

The objects of science for Ockham, as was seen in §4, were universal mental propositions. This controversial thesis, however, did not lead him anywhere near idealism or scepticism, precisely because the supposition of terms was called upon to secure the required links between such mental propositions and real external individual beings. Natural science in particular was recognized as adequate general knowledge of sensible substances. Ockham in fact devoted several treatises to it, the most important of which are his unfinished *Expositio in libros Physicorum*

Aristotelis, his *Summa philosophiae naturalis (Synopsis of Natural Philosophy)* and his *Quaestiones in libros Physicorum Aristotelis (Questions on the Physics of Aristotle)*. Taken together, these three works occupy more than 2,000 pages in the Franciscan Institute edition and contain some of Ockham's most intriguing ideas.

The main goal of natural philosophy, as Ockham sees it, is to provide a theory of the general principles of change among sensible things. Following Aristotle, Ockham acknowledges different types of change in nature: generation and corruption of substances, intensification and diminution of qualities, rarefaction and condensation of matter and, finally, local motion. About each one of these he has detailed - and often provocative - positions to defend. However, in the last analysis all change is accounted for in his system on the sole basis of the four kinds of beings admitted by his ontology: substances, qualities, substantial forms and prime matter. Much of his natural philosophy is accordingly devoted to showing that space, void, time, motion and the like are not distinct, absolute things. Prime matter in particular is considered as being actually extended by itself, and its spatial dimensions are not - as in Aquinas and many other medieval authors - given the status of special quantitative forms enjoying an existence of their own.

The core of the method in this peculiar variety of physics is neither empirical nor mathematical but semantical, the point being to show that general sentences which appear to be about space, void, time or motion are in fact about substances, qualities, forms and pieces of matter. Thus reinterpreted, such sentences can very well convey truths, even necessary truths, about physical reality. The crucial device for this type of reductive analysis is the idea that a single word in spoken or written language, such as 'motion' or 'time', is often an abbreviation for a complex mental expression including not only absolute categorematic terms, but also conjunctions, adverbs, verbs and, above all, connotative terms. Moreover, these abbreviative words are often to be interpreted contextually - for example, the analyst should not systematically try to replace the relevant word by one and the same definition in all its occurrences, but should instead try to reformulate, on the basis of a few rules of thumb, the whole sentences in which the problematic word occurs.

An abstract noun such as 'change' (*mutatio*), for example, does not designate a special entity of its own. Rather, its uses are derivative with respect to those of the corresponding verb 'to change' (*mutare*). The meaning of the verb in such a case provides the key for the interpretation of the various sentences in which the derivative noun occurs. The verb 'to change' in effect normally applies to something (for example, a substance) acquiring something which it did not have before (for example, a new form or a new quality) or entering into relations with new things (for example, new substances). This is what should be kept in mind when considering spoken or written sentences involving the derivative noun, such as 'every change is by an agent' (which is to be interpreted as 'everything that changes is changed by an agent') or 'change moves from the prior to the posterior' (which means that 'when something changes, it moves from a prior state to a posterior state'). The term 'motion', in the same vein, normally supposits for the things which are moved, but it may also, if the context favours it, supposit for the things that cause the others to move; or it may even in special cases receive a metalinguistic interpretation and supposit for linguistic or conceptual units such as the verb 'to move'. The word 'time', to take a last example, does not signify any special extramental entity either, but supposits, in the sentences in which it occurs, for the very same things the word 'motion' would supposit for in similar contexts, and connotes in addition a certain numbering activity of the human soul.

Sentences involving such words, then, can be 'saved' without enriching the ontology, and when they are so reinterpreted, it can be admitted that their truth or falsity is neither illusory nor subjective. Nominalism here does not preclude the objectivity, the mind-independence, or even the necessity of physical phenomena. Ockham in particular is no precursor of Hume in being suspicious about natural causality. Efficient causality in his view is a natural necessary connection holding between singular causes and singular effects in virtue of their respective intrinsic natures; and it can be safely ascertained as such in appropriate circumstances. Ockham readily grants a causal version of the principle of the uniformity of Nature: 'causes of the same kind have effects of the same kind' (*Ordinatio*, prologue, q. 1). The fact that a certain individual *A* is, in virtue of its intrinsic substantial forms, of the same species as another individual *B* necessarily implies a strong similarity between *A* and *B* with respect to their causal powers. This is true to such an extent that a single observation, in appropriate circumstances, suffices in principle to justify an induction about causal connections: if it could be safely established in a single case that a given singular plant has a certain curative power, it could be correctly concluded by induction that all plants of the same species as this one do have similar curative powers. In conformity with the Aristotelian model, natural

philosophy in Ockham's view is an authentic science of necessary connections. Although the ontology is simplified with the help of sophisticated semantical analysis, traditional metaphysics is not abandoned (see Natural philosophy, medieval).

9 Natural theology

Theology, on the contrary, is not in the normal course of things properly called a science, according to Ockham, because it does not rest on the required sort of natural evidence. In his view a science, strictly speaking, is a set of cognitive *habitus* (see §4) having to do with necessary conclusions that were formerly dubitable for the agent but that became ascertained through syllogistic reasoning from self-evident general premises or well-established empirical generalizations. It is not so with theology, most of which rests on faith and revelation; the parts that can be demonstrated by natural reason alone are very thin indeed. Compared to most of his medieval predecessors, Ockham is quite sceptical with respect to a purely natural theology (see Natural theology).

First, no simple and proper concept of God, he holds, is accessible to human beings in this life. The human person in this life - frequently called the 'traveller' (*viator*) by medieval theologians - is characterized precisely by the fact of having no direct intuitive cognition of God. However, intuitive cognition of something is a prerequisite condition for the formation of any proper and simple abstract cognition of that thing (see §4), and therefore neither is any such cognition naturally possible in this life with respect to God. The only simple concepts applicable to God that a human being can naturally form are general concepts that will also signify other beings in one way or another. Some will be absolute (or quidditative) concepts (see §6), such as the general concept of 'being' which, according to Ockham, univocally applies to God and to creatures. However, most will be connotative or negative terms, such as 'creator' (which connotes the creatures) or 'immortal' (which negatively signifies death). A proper concept of God, for Ockham, is naturally constructible in this life only as a composite bundle of such general quidditative, connotative and negative concepts.

In the second place, even if some general or composite concepts applicable to God can naturally be formed, this will not suffice for the 'traveller' to *know* that these concepts apply to the Supreme Being. What would have to be known in addition is at least that such a being exists. One can easily form, for instance, the concept of an immortal being, simply by negatively referring to death, but this possibility in itself does not suffice to warrant the belief that such a concept does apply to anything real.

Can the existence of God be proven by natural reason? Ockham criticizes and rejects most of the alleged proofs of his predecessors. Anselm's celebrated argument, for one, is considered by him as valid only insofar as it proves that among actually existing entities there is at least one with respect to which none is greater (see Anselm of Canterbury). This is far from enough, in Ockham's eyes, to conclude that nothing could possibly be greater, if it existed. Consequently, the argument does not establish that the greatest possible being does exist in fact. Proofs from final causality, such as Aquinas' fifth way, do not work either because, Ockham says, it cannot be demonstrated that material things that lack cognition do act according to final causes pre-established by a will (see God, arguments for the existence of).

The proofs of God's existence that Ockham considers strongest are the arguments from efficient causality. However, these are not all equally conclusive, and even in the best cases they do not prove as much as most of his predecessors would have hoped for. Ockham admits as a self-evident principle that being cannot come from non-being, and concludes accordingly that every natural thing does indeed need to have been brought into existence by some external cause. But nothing, he adds, impedes in principle the possibility of an infinite regress in time in the series of such productive causes. Since no actual infinity would be involved, the past eternity of the material world is not philosophically impossible. What Ockham does deem impossible, on the other hand, is the existence of an actual infinity of simultaneous *conserving* causes; and this is the basis for the only relevant scientific proof he is ready to grant on this subject. Each caused thing, he believes, needs not only to be brought into existence, but also to be conserved in existence by some external cause (see Creation and conservation, religious doctrine of). Such a conserving cause must exist simultaneously with its effect as long as the effect exists, and the infinite regress in this case is consequently ruled out in principle since it would entail the existence of an actual infinity of distinct beings. This argument, then, does establish that there actually exists at any moment a first efficient conserving cause. What it does not prove, however, is that there should be only one such cause at any moment and that this cause should be identified with God: a celestial sphere or an angel could do equally well.

A fortiori, none of the other traditionally accepted characteristics of the Supreme Being can be proven by natural reason alone to be actually exemplified. In particular, it cannot be philosophically demonstrated, Ockham argues, that God is the first efficient cause of everything, that he is the final cause of something, that he is infinite or even that he knows anything. Articles of faith about God are not in general a matter for reason to settle. Rational argumentation is not ruled out of theology, of course - far from it - but its most crucial premises are accepted solely on the basis of religious faith. Some of the theological truths indeed even run counter to what natural reason by itself would favour: Ockham admits, for instance, that the doctrine of the Trinity involves the actual existence of distinct relational entities, something which, as a philosopher, he firmly rejects. In such cases, Ockham always clearly asserts the priority of the authority of the church over the pronouncements of reason and experience. Philosophy and theology nevertheless tend to be much less intertwined in his thought than they were in the thought of Aquinas or Scotus, for example.

10 Ethics

Ethics is especially interesting in this regard insofar as it simultaneously displays both the intimate interconnection and the relative independence of the philosophical and the theological outlooks. Ockham devoted no special elaborate treatise to ethics and nowhere systematically collected his views on the matter; but many scattered relevant developments in the commentary on the *Sentences*, the *Quodlibeta septem* and the collection known as the *Quaestiones variae (Various Questions)* do reveal, when considered together, an original two-sided theory of the object and foundations of moral knowledge (see Ethics).

Taken as a whole, moral science in this view is the practical normative knowledge of what is to be done, in general or under particular circumstances. It includes two distinct and independent parts. Positive moral science, on the one hand, is knowledge about the legal obligations one is formally subject to; it subdivides in turn into knowledge about human laws and knowledge about divine laws. Non-positive moral science, on the other hand, is directive knowledge about human action (or *praxis*) insofar as it is based on natural reason and experience. The former, says Ockham, is not a demonstrative science properly speaking, since it depends on human and divine precepts which are not evidently known *per se* or empirically. The latter, which corresponds to what Aristotle discusses in his own ethical books, is indeed a proper part of demonstrative philosophy, and it is even more certain and useful, Ockham insists, than many other sciences. It rests jointly on natural experience and on a number of general principles evidently known *per se*, such as: ‘what is honest should be done and what is dishonest avoided’, ‘benefactors deserve gratitude’, ‘the will should conform to the dictates of right reason’, and so on.

The philosophical part of ethics will shed light, in particular, on the general criteria of what is humanly laudable or blameworthy, even for a pagan. Ockham’s theses on this are characteristic. Only acts, according to him, can be morally good or virtuous; and among acts, only internal acts of the will can be rightly considered as *intrinsically* virtuous. Other sorts of acts, whether bodily or mental, are sometimes called ‘virtuous’, of course, but this is only ‘by extrinsic denomination’, exactly as a beverage is sometimes called ‘healthy’ not because it is itself in good health but because it has some salient causal relation with something else which is (or could be) intrinsically in good health. Ockham’s ethics, like Abelard’s, consistently stresses the basic moral character of intentions rather than that of outward behaviour (see Abelard, P.). His point is that for an act to be intrinsically virtuous, some necessary conditions need to be met that can be fulfilled only by internal acts of the will. First, it has to be free of internal necessity or external compulsion; it has, in other words, to be under the direct power of the will itself. Second, it has to conform to the dictates of what he repeatedly calls ‘right reason’ (*recta ratio*), which is the rational and well-informed use of prudence and moral conscience.

This notion of ‘right reason’ is central to Ockham’s philosophical ethics. The highest degree of moral virtue pagans can reach, according to him, precisely requires that they should will not merely to fulfil their moral obligations, but also that they should so will precisely in view of the fact that these obligations are dictated by right reason. Right reason, on the other hand, does not reduce to natural reason alone. It involves correct and prudential reasoning enlightened by whatever relevant true knowledge is available to the agent. In the case of the Christian believer, then, the dictates of right reason cannot be independent of the content of Christian faith. This is the exact point where non-positive moral demonstrative science yields, for Ockham, to a higher form of ethics: positive Christian moral science, as enlightened by religious faith.

This is not to say that philosophical moral science by itself is completely unreligious. Ockham explicitly holds, on the contrary, that reason suffices to demonstrate that there is at any time an actually existing *summum bonum*, and right reason, then, dictates, independently of any divine revelation, that nothing else ought to be loved more than this *summum bonum*. However, Christian revelation goes much further. It tells the believer, among other things, that this *summum bonum* is the unique God, that he is not only the highest good in fact, but also the highest possible good, that he is our creator and benefactor, that he is a personal knower and so on. Once these data from revelation are taken into account, right reason should lead the believer towards an even higher degree of moral virtue, which will correspond to Christian morality proper. It is reached when agents will to fulfil their obligations not merely because they conform to the dictates of right reason anymore, but primarily for the love of God himself and because these obligations derive from God's will. This alone is the level of perfect moral virtue.

The real ultimate foundation of ethical norms, then, turns out to be God's will and nothing else. The sole fact that God commands something to be done is what makes that thing morally good. God could at any time change his precepts and our moral obligations would then change accordingly. He could command, for example, that we should from now on steal whenever we can and it would *ipso facto* become morally good to do so. However, this is no longer demonstrative ethics; it is Christian theology based on revelation and on the indemonstrable principle of God's absolute supremacy. Even though religious faith leads him beyond philosophical ethics into what has often been called a 'divine command ethics', which ultimately plays a foundational role in his overall system, Ockham in the last analysis does not blur the distinction he had drawn between the two sorts of moral knowledge, and he never expresses serious doubts about the soundness of secular moral reasoning in the actual world.

11 Political thought

While Ockham had not ventured into political philosophy during his academic career in England, almost all the works he is taken to have written in Munich from 1330 to his death in 1347 are directly related to his conflicts with the Avignon papacy (with the possible exception of one short logical tract, the *Elementarium logicae (Elements of Logic)*). His goal at first was to expose the errors and heresies of Pope John XXII, especially about poverty within the church and about the beatific vision. This raised in effect the delicate question of papal heresy: who, in such a case, is entitled to denounce it? Who has the power to depose the Pope? Reflections on such matters progressively led Ockham to be more and more interested in political theory proper, and most of his writings after 1337 were directly devoted to it, notably the *Octo quaestiones de potestate papae*, Book III of the *Dialogus* and the *De Imperatorum et Pontificum potestate*.

In this endeavour, Ockham primarily thought of himself as a theologian. Many of his arguments against the Avignon curialists were designed for the ecclesiastical community and based on the Bible, the writings of the Church Fathers or canon law. However, he also liberally used Aristotle's *Politics* and much of his thought in this field can be seen, in fact, as being in direct continuation with his philosophical ethics of right reason: politics, after all, is but a province of human *praxis*.

Indeed, Ockham's basic political idea is that the spiritual and the temporal powers should normally be kept apart and not be allowed to overlap, except in special cases of crisis such as papal heresy or tyrannical abuse. On the whole, he found much to say in favour of the institution of papacy, but he energetically opposed the papalist doctrine of the *plenitudo potestatis* of the Pope and emphatically denied the dependence in principle of secular on ecclesiastical power. The church, he thought, should as a general rule disentangle itself from worldly politics. Consecration by the Pope, for example, was not in his view a necessary condition for the legitimacy of the Emperor.

What then is the basis for the legitimacy of secular political power? Popular consent, according to Ockham, is the normal original source of legitimacy; but beyond this, the rightness of a certain political regime ultimately depends on how well it fulfils its functions. Government is instituted for the good of a community of individual subjects, and if a regime adequately performs this task, it is not legitimate, for the Pope or even for the majority of subjects, to try to overthrow it by force. Ockham's approach to politics is basically functional. What he thinks should be expected from a government is mainly that it should prevent and punish injustice and so preserve in this way the rights and freedom of the individual members of the community. Monarchy, for him, is usually the best regime insofar as it is, in most circumstances, best suited to such a role; but there is nothing absolute - or divinely founded - in this judgment in Ockham's eyes. A political regime is a man-made institution to be judged by its expediency in

achieving the goal for which it should ideally be designed: and common welfare or public security can in principle be enhanced by different regimes in different circumstances.

Natural rights, in this perspective, are what a government is expected to protect (see Rights). Ockham never gave a systematic account of these rights, but examples of them in his writings include the individual's right to the necessities of life, the collective right for a people lacking a ruler to choose one, the right to resist an unjust government, and so on. As a general rule, the rights and obligations of rulers and subjects, he thought, can be determined in particular circumstances by the normal prudential use of human right reason. Ultimately, what the political and legal institutions should be aiming at preserving is the individual freedom of the members of the community in their own private spheres.

At the same time this respect for freedom sets a limit in principle to the exercise of any political power, whether ecclesiastical or secular: 'free persons, who are not slaves', Ockham writes, 'should never, without some fault, be compelled by new laws to do things which are not necessary to the commonwealth or to their neighbour' (*Breviloquium de potestate Papae* II, 17). Political rule should be kept to the minimum which is required in any circumstances for the fulfilment of its social functions.

It has often been wondered whether Ockham's political thought is dependent on the rest of his philosophy, and on his nominalism in particular. Certainly his political theories were not deduced from nominalistic premises in ontology, theory of knowledge or philosophy of language. He never defended his political stands by arguments drawn from these fields, and his conceptions of ecclesiastical and political power are compatible with different ontological or epistemological doctrines. Moreover, probably because of the new audience he was trying to reach, he did not make much use in his political writings of the technical tools of semantical analysis which he had so extensively exploited in theology and philosophy. Nevertheless, the break between the two parts of his work is far from complete. Ockham's approach to politics and natural rights is in direct continuation with his philosophical ethics of right reason. And, more importantly, the emphasis of his later writings on personal freedom can certainly be seen as the political counterpart of the accent on singularity which was the trademark of his nominalism. Even though there is no logical connection here, both aspects of Ockham's thought harmoniously fall in line with each other in consistently stressing the primacy of individuals.

See also: Aristotelianism, medieval; Biel, G. §2; Buridan, J.; Chatton, W.; Duns Scotus, J.; Empiricism; Gregory of Rimini; Language, medieval theories of; Logic, medieval; Natural philosophy, medieval §9; Nominalism; Universals; Wodeham, A.

CLAUDE PANACCIO

List of works

1 Critical editions

William of Ockham (1317-47) *Opera Philosophica et Theologica (Philosophical and Theological Works)*, eds G. Gál *et al.*, St Bonaventure, NY: The Franciscan Institute, 1967-88, 17 vols.(Excellent critical edition of all of Ockham's non-political writings, subdivided into two series: *Opera Philosophica*, 7 vols, and *Opera Theologica*, 10 vols. A selection of basic philosophical texts can be found in English translation in P. Boehner, *Ockham: Philosophical Writings*, London: Nelson, 1957.)

William of Ockham (1332-47) *Opera Politica (Political Works)*, eds H.S. Offler *et al.*, Manchester: The University Press, 1956-74, vols I-III; Oxford: Oxford University Press, 1997, vol.IV.(Still incomplete; it lacks in particular the very important *Dialogus*. A selection of political texts, most of which come from the *Dialogus*, can be found in English translation in A.S. McGrade and J. Kilcullen, *William of Ockham: A Letter to the Friars Minor and Other Writings*, Cambridge: Cambridge University Press, 1995.)

2 Major works

William of Ockham (1317-19) *Ordinatio. Scriptum in librum primum Sententiarum (Commentary on the First Book of the Sentences)*, in *Opera Theologica*, St Bonaventure, NY: The Franciscan Institute, vols I-IV.(The basic work for Ockham's epistemology, ontology and theology; partly revised 1320-6.)

William of Ockham (1317-19) *Reportatio. Quaestiones in libros Sententiarum II-IV (Questions on Books Two, Three and Four of the Sentences)*, in *Opera Theologica*, St Bonaventure, NY: The Franciscan Institute, vols

V-VII.(Authorized students' notes; a useful complement to the *Ordinatio*.)
William of Ockham (1317-24) *Quaestiones variae (Various Questions)*, in *Opera Theologica*, St Bonaventure, NY: The Franciscan Institute, vol. VIII.(Eight developments on various points such as the language of angels, the interrelations between virtues and so on.)
William of Ockham (1319-21) *Summula philosophiae naturalis (Synopsis of Natural Philosophy)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vol. VI.(Unfinished work: Ockham's first attempt at physics.)
William of Ockham (1320-3) *Expositio in librum Porphyrii de praedicabilibus (Commentary on Porphyry's Treatise on the Predicables)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vol. II.(Mainly on universals.)
William of Ockham (1320-3) *Expositio in librum praedicamentorum Aristotelis (Commentary on Aristotle's Categories)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vol. II.(Introduce's Ockham's nominalistic interpretation of Aristotle's ten categories.)
William of Ockham (1320-3) *Expositio in librum Perihermenias Aristotelis (Commentary on Aristotle's De interpretatione)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vol. II.(Contains one of Ockham's most detailed discussions on the mental existence of concepts.)
William of Ockham (1320-3) *Tractatus de praedestinatione et de praescientia Dei respectu futurorum contingentium (Treatise on Predestination and God's Foreknowledge of Future Contingents)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vol. II; trans. M.M. Adams and N. Kretzmann, *William of Ockham: Predestination, God's Foreknowledge, and Future Contingents*, Indianapolis, IN: Hackett, 2nd edn, 1983.(How to reconcile God's infallible knowledge with the contingency of future events.)
William of Ockham (1322-4) *Expositio super libros Elenchorum (Commentary on Aristotle's Sophistical Refutations)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vol. III.(Ockham's theory of logical fallacies.)
William of Ockham (1322-4) *Tractatus de quantitate et Tractatus de corpore Christi (Treatise on Quantity and Treatise on the Body of Christ)*, in *Opera Theologica*, St Bonaventure, NY: The Franciscan Institute, vol. X; trans. B.T. Birch, *The De Sacramento Altaris of William of Ockham*, Burlington, IA: The Lutheran Literary Board, 1930.(Mainly on the semantics of quantitative language; Birch's translation should be used with caution.)
William of Ockham (1322-4) *Brevis summa libri Physicorum (A Short Summary of the Book of Physics)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vol. VI; trans. J. Davies, *Ockham on Aristotle's Physics*, St Bonaventure, NY: The Franciscan Institute, 1989.(Ockham's shortest physical work; a useful synthesis of his views on motion.)
William of Ockham (1322-4) *v(Questions on the Physics of Aristotle)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vol. VI.(151 Questions on concepts, change, motion, time, place and related subjects.)
William of Ockham (1322-7) *Expositio in libros Physicorum Aristotelis (Commentary on the Physics of Aristotle)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vols IV-V.(Ockham's largest and most important work on physics.)
William of Ockham (1322-7) *Quodlibeta septem* (Quodlibetal Questions), in *Opera Theologica*, St Bonaventure, NY: The Franciscan Institute, vol. IX; trans. A.J. Freddoso and F.E. Kelley, New Haven, CT: Yale University Press, 1991, 2 vols. (A set of 170 short questions giving Ockham's latest views on important philosophical and theological subjects.)
William of Ockham (1322-7) *Summa logicae (Summa of Logic)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vol. I; Part I trans. M.J. Loux (1974) *Ockham's Theory of Terms*, Notre Dame, IN: University of Notre Dame Press; Part II trans. A.J. Freddoso and H. Schuurman (1980) *Ockham's Theory of Propositions*, Notre Dame, IN: University of Notre Dame Press. (The basic work for Ockham's semantics and logic.)
William of Ockham (1332-4) *Opus nonaginta dierum (The Work of Ninety Days)*, in *Opera Politica*, Manchester: The University Press, vols I-II. (A detailed discussion of Pope John XXII's ideas on poverty.)
William of Ockham (1334-47) *Dialogus (The Dialogue)*, Lyons, 1494-6; repr. London: Gregg Press, 1962. (Ockham's major political synthesis; unfinished.)
William of Ockham (1334-5) *Tractatus contra Ioannem (Treatise Against John)*, in *Opera Politica*, Manchester: The University Press, vol. III.(Argues that Pope John XXII, who had just died, was a heretic despite his supposed final 'retractation'.)
William of Ockham (1337-8) *Tractatus contra Benedictum (Treatise Against Benedict)*, in *Opera Politica*,

Manchester: The University Press, vol. III.(A discussion on papal heresy and the nature of the church; attacks Pope Benedict XII for not having reproved the errors of his predecessor.)

William of Ockham (1338-9) *Compendium errorum Papae Ioannis XXII*, Lyons, 1495; repr. London: Gregg Press, 1962; in *Opera Politica*, Oxford: Oxford University Press, 1997, vol. IV.(Synthesizes Ockham's doctrinal critique of John XXII.)

William of Ockham (1339-40) *Breviloquium de potestate Papae (A Short Discourse on the Power of the Pope)*, ed. L. Baudry, Paris: Vrin, 1937; trans. A.S. McGrade and J. Kilcullen, *William of Ockham: A Short Discourse on Tyrannical Government*, Cambridge: Cambridge University Press, 1992.(On the relations between the power of the Pope and that of the Emperor.)

William of Ockham (1340-1) *Octo quaestiones de potestate Papae*, in *Opera Politica*, Manchester: The University Press, vol. I.(On various questions concerning the alleged authority of the Pope over the Emperor.)

William of Ockham (1340-7) *Elementarium logicae (Elements of Logic)*, in *Opera Philosophica*, St Bonaventure, NY: The Franciscan Institute, vol. VII.(An interesting logical treatise, but its authenticity is still doubtful.)

William of Ockham (1346-7) *De imperatorum et pontificum potestate (On the Power of Emperors and Popes)*, ed. C.K. Brampton, Oxford: Clarendon Press, 1927; part ed. W. Mulder in *Archivum Franciscanum Historicum* 16, 1923: 469-92, and 17, 1924: 72-97; in *Opera Politica*, Oxford: Oxford University Press, 1997, col. IV.(A final discussion of civil and papal powers.)

References and further reading

Adams, M.M. (1987) *William Ockham*, Notre Dame, IN: University of Notre Dame Press, 2 vols.(The best overall book on Ockham's philosophy and theology; a must for any serious study. Leaves out the political works.)

Baudry, L. (1950) *Guillaume d'Occam. Sa vie, ses oeuvres, ses idées sociales et politiques (William of Ockham: His Life, Works and Political Ideas)*, vol. I: *L'homme et les oeuvres*, Paris: Vrin.(Outdated on many points, it remains the only detailed biography available; more up-to-date biographical information is to be found in the introductions to the various volumes of the critical edition.)

Beckmann, J.P. (1992) *Ockham - Bibliographie 1900-1990*, Hamburg: Felix Meiner. (A useful tool for further research.)

Boehner, P. (1958) *Collected Articles on Ockham*, St Bonaventure, NY: The Franciscan Institute.(A collection of essays by a leading pioneer; still an indispensable reference work in the field.)

Freppert, L. (1988) *The Basis of Morality According to William Ockham*, Chicago, IL: Franciscan Herald Press. (A clear and well-informed presentation of Ockham's ethics.)

Goddu, A. (1984) *The Physics of William of Ockham*, Leiden: Brill.(The authoritative monograph on the subject.)

McGrade, A.S. (1974) *The Political Thought of William of Ockham*, Cambridge: Cambridge University Press.(A very useful and illuminating synthesis of a large body of material.)

Moody, E.A. (1935) *The Logic of William of Ockham*, New York: Sheed & Ward.(A seminal study, and a good introduction to Ockham's semantics.)

Panaccio, C. (1992) *Les mots, les concepts et les choses. La sémantique de Guillaume d'Occam et le nominalisme d'aujourd'hui (Words, Concepts and Things: The Semantics of William of Ockham and the Nominalism of Today)*, Montreal/Paris: Bellarmin/Vrin.(A philosophical discussion of Ockham's nominalism with reference to contemporary analytical philosophy.)

Pasnau, R. (1997) *Theories of Cognition in the Later Middle Ages*, Cambridge: Cambridge University Press.(An analysis of the main issues in epistemology from Aquinas through Ockham.)

Tachau, K.H. (1988) *Vision and Certitude in the Age of Ockham: Optics, Epistemology and the Foundations of Semantics 1250-1345*, Leiden: Brill.(Provides a rich intellectual context for Ockham's epistemology.)

William of Sherwood (*c.*1200/5-*c.*1266/75)

William of Sherwood, an English logician of the mid-thirteenth century, is most noted for his theories of supposition and syncategorematic terms. In application, these theories enable us to express the true logical form of sentences with misleading grammatical forms. William's Insolubilia (Insolubles) deals with paradoxes of self-reference, such as 'I am now uttering a falsehood'.

William, who was born probably in Nottinghamshire, studied at Oxford and became a master there. Roger Bacon praised him as a greater logician than Albert the Great. His *Introductiones in logicam (Introduction to Logic)* includes five chapters on traditional logic, corresponding to Aristotle's *On Interpretation*, *Prior Analytics*, *Topics* and *Sophistical Refutations* (see Aristotle) and Porphyry's *Isagōgē* (see Porphyry), and it is typical of its time in its reliance on Boethius for much of what it says about the traditional topics.

The most interesting chapter of *Introductiones in logicam* is the one on the 'properties of terms', a logical doctrine that originated in the twelfth century. In William's account, the first of these properties is signification, the 'presentation of the form of something to the understanding'. This is the meaning of a term taken in isolation from its context, and specified extra-linguistically. Sometimes the signification of a term has nothing to do with its use in a sentence and the term 'supposits for' itself, as in 'man is a monosyllable'; this is material supposition. A term also supposits a form in a sentential context if its use there reflects its signification, so that it supposits *for* the form itself or something subordinate to it; this is formal supposition, in which a term supposits for an 'understanding of something' or, as William of Ockham might have said, for a mental term.

'Copulation' is parallel to supposition but belongs to adjectives and depends on a signification of something not as existing, but as adjoined to what exists. A term's 'appellation' is the set of those things of which the term can be predicated truly in the present tense. In an ordinary present-tense sentence a term's supposition will not extend beyond its appellation, for its subject can only supposit for actually existing things. The form that a term signifies does not of itself provide reference, then, unless the term supposits for the form itself in simple supposition, as in 'man is a species'. Otherwise, the term depends on existing individuals for its reference, a view consistent with nominalism, permitting us to take the form supposited for in simple supposition as an 'understanding of something', or a mental term, acquired through acquaintance with individuals, which are the term's primary referents. Thus William's views contrast with those of Peter of Spain, which reflect a realist stance.

Usually a term supposits for individuals (or, to be precise, concepts of individuals) falling under the form it signifies, which is personal supposition. Personal supposition is 'determinate' when a word supposits for a single unspecified individual, as in 'a man is running', and 'confused' when a word supposits for more than one thing. Confused personal supposition is 'mobile' if one is entitled to attach its predicate to each of the individuals for which the term supposits, and is otherwise 'immobile'. Rules concerning personal supposition were used to assess the validity of arguments treated nowadays with quantification theory, but supposition occurs within a natural language, and no artificial language adapted to the uses of logic had been constructed. As a result, an open-ended set of rules governing the different types of supposition was introduced to handle all sorts of special contexts in scholastic Latin.

William's *Syncategoremata (Treatise on Syncategorematic Words)* extends supposition theory to sentences containing words that produce a misleading grammatical form, sometimes by altering the supposition of associated terms or concealing the true logical subject, and sometimes by abbreviating composite statements and making them appear categorical. Such words include 'every', 'whole', 'all', 'both', 'nothing', 'but', 'only', 'necessarily', 'begins', 'unless' and many others.

Also attributed to William, though doubtfully, are the treatises *De petitionibus contrarium (On Contrary Assumptions)* and *Insolubilia (Insolubles)*, and an *Obligationes* which is probably by Walter Burley. The *Insolubilia* deals with semantic paradoxes of self-reference. For instance, when one utters 'I am now saying a falsehood', William insists that the sentence is meaningful and rejects the general rule excluding any reference within a sentence to itself, since some self-referential sentences make perfect sense and are true. Rather, he appeals to the principle of charity and takes it that the sentence refers outside itself in this case, since it is otherwise absurd.

Thus, those who utter ‘I am now saying a falsehood’ assert that they are saying some falsehood other than the sentence they utter, which is, of course, false (see Semantic paradoxes and theories of truth).

See also: Language, medieval theories of

JOHN LONGEWAY

List of works

William of Sherwood (*c.*1250?) *Introductiones in logicam (Introduction to Logic)*, ed. M. Grabmann, Munich: Verlag der Bayerischen Akademie der Wissenschaften, 1937; trans. N. Kretzmann, *William of Sherwood’s Introduction to Logic*, Minneapolis, MN: University of Minnesota Press, 1966; ed. C.H. Lohr with P. Kunze and B. Mussler, ‘William of Sherwood, "Introductiones in logicam", Critical Text’, *Traditio* 39, 1983: 219-99; ed. and trans. H. Brands and C. Kann, *William of Sherwood. Introductiones in logicam: Einführung in die Logic*, Hamburg: Meiner, 1995.(Kretzmann’s translation contains extensive analytical notes. The Lohr edition supersedes the older Grabmann edition; Brands and Kann’s new German translation with full explanatory notes supersedes both earlier editions.)

William of Sherwood(*c.*1250?) *Syncategoremata (Treatise on Syncategorematic Words)*, ed. R. O’Donnell, *Medieval Studies* 3 (1941): 46-93; trans. N. Kretzmann, *Treatise on Syncategorematic Words*, Minneapolis, MN: University of Minnesota Press, 1968.(Kretzmann’s translation contains extensive analytical notes.)

William of Sherwood(*c.*1250?) *Insolubilia (Insolubles)*, ed. M.L. Roure, ‘Insolubilia Guillelmi Shyreswood (?)’, *Archives d’histoire doctrinale et littéraire du moyen âge* 37, 1970: 248-61.(No translation of this work is currently available.)

References and further reading

Braakhuis, H.A.G. (1977) ‘The Views of William of Sherwood on Some Semantical Topics and Their Relation to Those of Roger Bacon’, *Vivarium* 15: 111-42.(An important article on William’s semantics.)

De Libera, A. (1982) ‘The Oxford and Paris traditions in Logic’, in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(Places William’s work in the context of his time.)

Williams, Bernard Arthur Owen (1929-)

Bernard Williams has written on the philosophy of mind, especially personal identity, and political philosophy; but the larger and later part of his published work is on ethics. He is hostile to utilitarianism, and also attacks a view of morality associated in particular with Kant: people may only be properly blamed for what they do voluntarily, and what we should do is the same for all of us, and discoverable by reason. By contrast Williams holds that luck has an important role in our evaluation of ourselves and others; in the proper attribution of responsibility the voluntary is less central than the Kantian picture implies. Williams thinks shame a more important moral emotion than blame. Instead of there being an independent set of consistent moral truths, discoverable by reason, how we should live depends on the emotions and desires that we happen to have. These vary between people, and are typically plural and conflicting. Hence for Williams ethical judgment could not describe independent or real values - by contrast with the way in which he thinks that scientific judgment may describe a real independent world.

1 General

Bernard Williams studied philosophy and classics at the University of Oxford. He was later Professor of Philosophy in London, Cambridge, Oxford and Berkeley, California, as well as chairing the commission set up by the British Government in 1977 to consider the legal control of obscene material.

Williams has written about many different areas of philosophy. His books include full-length studies of Descartes and of early Greek ethical thought, and he has written extensively on the philosophy of mind. His most influential writing has been in ethics, where he wrote *Morality* (1972), a book which pretends to be introductory but is in fact much more, *Ethics and the Limits of Philosophy* (1985), a criticism of utilitarianism (the ‘against’ half of *Utilitarianism: For and Against* (1973a)), and many occasional papers, the most important of which are collected in *Moral Luck* (1981b).

Before turning to this ethical work two important early papers on other areas should be mentioned. His first collection, *Problems of the Self* (1973b), starts with a paper on personal identity. Williams takes an imaginary case in which twin brothers both claim to remember being the same historical figure. This shows that memory claims are not sufficient to establish personal identity (since the twins, being different from each other, cannot both be the earlier person). Hence, thinks Williams (although this would involve another argumentative step), the necessity of bodily identity for personal identity. This paper’s use of duplication is the basis of much of the subsequent discussion (see Personal identity).

Another influential early paper in this collection is ‘The Idea of Equality’. Williams here argues that some activities have an internal object, so that, for example, the internal object of the provision of medical services is medical need. Hence it is not optional whether health care should be distributed according to health need or according to ability to pay. By solving the problem of relevance in this way, Williams solves the problem of which property to consider in distribution, and hence a central problem of making egalitarian proposals practical (see Equality).

2 Morality, the peculiar institution

Williams, like other philosophers, often works by critique and contrast. The centre of his ethical position can best be brought out by considering what he takes himself to be opposing. This is a view of morality principally associated with Kant. In this view, morality imposes on all people a set of duties discoverable by reason alone. These obligations are taken to be objectively true, untouched by emotion and unconstrained by interest. The area of morality is meant to be sharply distinguishable from other practical concerns and to be particularly connected to the voluntary. For Kant, we can all equally will and are equally responsible for, and only for, what we will (see Kantian ethics).

All this Williams criticizes. He resists even the word ‘morality’, preferring the term ‘ethics’ for the study of how we should live. The voluntary he regards as too weak a notion to found obligation. He holds that our moral psychology and ethical intuitions reveal our values to be more tangled, pluralist and impure than the Kantian model allows. Kantianism is centred on obligation and blame. Instead of blame, Williams proposes the centrality

of shame. He discusses the nature of shame by considering the thought of the early Greeks, which he thinks is more applicable to the actuality of our current practical thinking than Kant's idealization. For Williams, Kant is wrong on the psychological facts; on the nature of the person; and on motivation. Kant is wrong to exclude the emotions. By contrast, Williams uses emotions, particularly regret, as a tool to uncover ethical requirements (see Morality and emotions; Praise and blame).

Perhaps the topic that focuses these thoughts most fully is that of moral luck. In the Kantian model, moral evaluation of people cannot depend upon mere luck (hence the importance of intention: there may be luck whether people succeed; there is no luck about whether they try). Williams wishes to show instead that our central evaluations may depend upon luck. For example, if I drive my car carefully and through no fault of my own run over a child, I may still, according to Williams, feel what he calls 'agent-centred regret' (that is, not just regret that it happened, but also regret that I did it). I may feel it proper to try and make some sort of recompense or acknowledgement of my responsibility. Yet my intentions were impeccable; if nevertheless regret is an appropriate emotion, it is a product of pure bad luck in a way that the Kantian model would preclude.

Williams' central example is the case of Gauguin (and a picture of Gauguin is used to illustrate the cover of *Moral Luck*). Gauguin left his wife and family to go to the South Seas to paint great pictures. If he succeeds he is justified; if he doesn't he is reprehensible. Again, mere intention is not enough; proper evaluation depends upon the luck of success (see Moral luck).

3 Internal and external reasons

For Kant, once I know that something is the right thing to do (which I am supposed to be able to do by pure reason alone, uninfluenced by desire), this alone supplies me with a reason for action. This Williams resists. He calls such reasons 'external' reasons, and wants to argue, by contrast, that reason can only motivate me if it is connected to some already existing motivational state of myself (such as a desire). Such reasons Williams calls 'internal' reasons, because they follow from something internal to the agent. Williams' terminology here needs care and is liable to be confusing as he uses these terms in a different manner and with opposite effect than most philosophers (for whom 'internal' is taken to mean internal not to the person, but to morality; in the standard use, a reason based on antecedent desire is hence 'external').

In Williams' terminology, all reasons (that is, all real reasons) are internal reasons. That something, for example, follows from the rationally apprehended supposed requirements of justice gives me in itself no reason to do it. I only have a reason if I also happen to want to be just. For Williams it is the same with prudence. Some philosophers, such as Sidgwick, argue that I am rationally required to regard all times to be of equal value; Williams, by contrast, argues that I only have a reason to consider the distant future as important as the near future if I happen to have a particular kind of prudential psychology. Again, what counts as a reason for me depends upon my own idiosyncratic psychological state; pure reason alone is unable to tell me what to do (see Moral motivation).

4 Utilitarianism

Williams also attacks utilitarianism, which shares with Kantianism the claim that morality has an impartial and objective foundation. Williams attacks the idea that the plurality of values can be reduced to a single metric. He probes the question of whether utilitarians can properly think that they are utilitarians. However, his central objection is again in terms of the impossible psychology that the moral theory demands.

In utilitarianism everything is justified by the consequences. In criticism, Williams wants to show that it is important not just what happens but also who does it. He takes the example of someone having the choice of doing a terrible thing, such as killing someone, in order to prevent more deaths being caused by someone else. For utilitarianism, which just counts the number of deaths, the answer is simple. But for Williams the answer is not simple; which shows that we are also concerned with who does the killing. He puts this in terms of people's integrity: there are things which people cannot be expected to do, whatever the consequences, if they are to be able to preserve any sense of themselves as people, or agents, at all (see Utilitarianism).

5 Consistency and realism

In two early papers Williams attempts to show that moral judgments are much more analogous to desires than to beliefs. Where there is a conflict of beliefs, he argues, this must in itself tend to weaken at least one of the beliefs, since beliefs are aimed at truth and two contradictory beliefs cannot both be true. Yet inconsistent desires are not like this. Although they cannot both be satisfied, this has no tendency to weaken at least one of them.

Williams shows that in tragic cases someone may be under two incompatible obligations, as Agamemnon in classical Greek tragedy was under incompatible obligations to his daughter and to his fleet. Whatever he does is wrong. On a Kantian view such conflict would not be possible; and it would not be rational to feel guilt about the thing not done. Yet Williams, again using moral psychology, shows that we both do and should feel regret in this sort of case. Hence what we did was wrong; hence there were incompatible obligations.

He derives from this conclusions about moral realism. The possibility of contradictory obligations shows that they cannot be taken to be descriptions of some real, independent, moral state of affairs. Our evaluative beliefs are plural and conflicting. They are therefore not descriptions of independent values which would provide objective and impersonal reasons for action (see Berlin, I.; Moral pluralism).

So far this is similar to other twentieth-century departures from the abstract impersonal reason of the Enlightenment. However, Williams disagrees with other thinkers in making a sharp distinction between ethical and other beliefs. For him, ethics is strongly contrasted with science. Both can be called kinds of knowledge, in that there can be agreement of belief about them. However, for Williams, this does not survive reflection in ethics, whereas in science it can. Scientific knowledge converges and can also explain how our scientific beliefs are caused by a real, independent world. Moral beliefs cannot in the same way survive explication of their causes.

An idea which Williams originally discussed in his book about Descartes, and which can be used to underline this distinction, is that of the ‘absolute conception’ of reality. Our knowledge and belief is perspectival, or from a particular position. Yet we have also the idea of things being true absolutely; that is, not from any particular position. This applies, thinks Williams, in science; at least as a regulative ideal. Things are just true; even if we do not know them. However, in ethics, it is all from a position. It is not just that it is psychologically impossible to think of ourselves being motivated by supposed moral truths which are true for all peoples and for all times; it is not just that our variable ideas show that we have no knowledge of any such truths; it is also that there are no such truths to be known.

See also: Morality and identity §4

ROSS HARRISON

List of works

Williams, B.A.O. (1972) *Morality*, Harmondsworth: Penguin; repr. Cambridge: Cambridge University Press, 1976.(Simple, introductory account, with useful treatments of subjectivism and amoralism.)

Smart, J.J.C. and Williams, B.A.O. (1973a) *Utilitarianism: For and Against*, Cambridge: Cambridge University Press.(Smart proposes and then Williams disposes of utilitarianism; fairly easily accessible; discussed in §4.)

Williams, B.A.O. (1973b) *Problems of the Self*, Cambridge: Cambridge University Press.(A selection of early papers, up to 1972; which contains the two papers discussed in §1 and the two discussed in §5.)

Williams, B.A.O.(1978) *Descartes: The Project of Pure Enquiry*, Harmondsworth: Penguin. (Critical study of Descartes; moderately hard in places; the ‘absolute conception’ discussed in §5 appears on page 245 and elsewhere.)

Williams, B.A.O. (1981a) *Obscenity and Film Censorship*, Cambridge: Cambridge University Press.(Abridgement of the report mentioned in §1.)

Williams, B.A.O. (1981b) *Moral Luck*, Cambridge: Cambridge University Press.(Papers of 1973-80; collecting the most important of Williams’ papers on ethics, including the papers ‘Moral Luck’ and ‘Internal and External Reasons’ discussed in §§2 and 3.)

Williams, B.A.O. (1985) *Ethics and the Limits of Philosophy*, London: Fontana.(Fairly accessible central account of ethical position, including contrast between ethics and science discussed in §5.)

Williams, B.A.O. (1993) *Shame and Necessity*, Berkeley, CA: University of California Press.(Sather Classical lectures; on early Greek ethical thought and its modern resonances; fairly accessible; and says more on shame and on the voluntary.)

Williams, B.A.O. (1995) *Making Sense of Humanity and Other Philosophical Papers*, Cambridge: Cambridge

University Press.(Papers from 1982-93; contains the discussion of Sidgwick mentioned in §3.)

References and further reading

Altham, J.E.J. and Harrison, R. (eds) (1995) *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, Cambridge: Cambridge University Press. (Critical essays of varying difficulty.)

Wisdom

*In ancient times, wisdom was thought of as the type of knowledge needed to discern the good and live the good life. Philosophy takes its name from it (*philosophía *means love of wisdom). But wisdom is little evident as a subject of contemporary philosophical discussion. It is interesting to ask how the concept of wisdom has come to vanish almost entirely from the philosophical map.*

1 Classical views of wisdom

Among the classical Greeks, wisdom (*sophia*) was one of the principal four virtues or human excellences (*aretai*), along with justice, moderation and courage, and was the primary focus of all epistemological discussion. In Plato's early dialogues, we find Socrates portrayed as a man on a mission given him by the god of Delphi (Apollo). Socrates tells the story (in Plato's *Apology*) that his friend Chairephon approached the oracle and asked if anyone was wiser than Socrates. The oracle answered 'no'. When Chairephon reported this encounter to Socrates, the latter found it deeply puzzling, for he regarded himself as not at all wise. In seeking to unravel the riddle of the oracle, Socrates questioned others who had the reputation for wisdom, and learned that he was indeed the wisest of men, for he alone recognized that he lacked wisdom. The mission of Socrates on behalf of the god, therefore, was to examine anyone willing to submit to his questioning, to see if they were wise and, if not, to 'assist the god' (*Apology* 23B) by showing that they are not. The pursuit of wisdom, then, and the exposure of the pretence of wisdom, was at the heart of Socratic philosophizing.

If Plato's early dialogues tell us anything about the historical Socrates, they show that, for him, wisdom would provide its possessor not only with all the other virtues (through Socrates' commitment to the 'unity of the virtues'), but also with infallible judgment in one's pursuit of a good or happy (*eudaimōn*) life. Some scholars even suppose that Socrates regarded the possession of wisdom as both necessary and sufficient to having a good or happy life, but this is controversial and seems to yield the result that no one could have a good or happy life, despite the fact that at least Socrates himself seems to have had one. Plato's early-period Socrates also recognizes minor and particularized forms of wisdom (each true craftsman would be wise in his craft, for example - see *Apology* 22C-D), but seems to regard these as philosophically insignificant, except as providing small-scale models for the really important form, which one could use in pursuit of a good and happy life. Wisdom, for Socrates, invariably requires the possession of knowledge, and not just any knowledge, but that which provides the basis for infallibly good judgment in decisions pertinent to how one should live. Wisdom, then, is the possession of such knowledge plus the disposition and skill to use this knowledge in the right ways.

In later works, Plato increasingly neglects the specific, craft-related wisdoms, coming eventually to identify wisdom only with the kind of knowledge that permits one to be infallible in judging the good. In the *Republic*, for example, Plato envisages a class of philosophical rulers who, through extensive training in mathematics and dialectic, will come to know the Form of the Good. This knowledge, we are told, will give the philosopher-rulers the fully actualized power to make infallible judgments in the state, seeing clearly, as it were, which institutions and political decisions were most just.

Aristotle is critical of Plato's theory of Forms in general, and of Plato's conception of the Good in particular. Aristotle's conception of the 'Four Causes' (*aitiai*) and his views about the categories of being require that there is a specific good for each thing - Plato's idea that there must be one Good for all things seems to Aristotle to be an absurdity (*Nicomachean Ethics* I.6; *Eudemian Ethics* I.8). Accordingly, Aristotle cannot see wisdom as requiring some vision of the Good. Aristotle accepts the infallibilism inherent in the Socratic and Platonic accounts, but divides wisdom into two sorts: theoretical and practical. Theoretical wisdom (*sophia*) is *nous* (the ability to grasp first principles) plus *epistēmē* (scientific knowledge or understanding - see *Nicomachean Ethics* VI.7), which involves knowledge of the 'causes' - the 'why' of things (*Metaphysics* A.1-2). In so far as there can be different sciences with different first principles, there can be different examples of theoretical wisdom - one for each science. Practical wisdom (*phronêsis*), for Aristotle, is knowledge of means and ends. Some have criticized this view as being insufficiently attentive to which ends are, and which could not be, the right ends for a wise person. This criticism, however, is controversial, and depends upon the degree to which Aristotle thought the study of ends could itself be scientific.

2 Later conceptions of wisdom

After Aristotle, wisdom begins to lose its place of primacy within philosophical discourse, no doubt partly due to Sceptical attacks, and partly due (later) to Christianity. The Sceptics rarely discussed wisdom. They appear to have accepted something like the Aristotelian account of theoretical wisdom, but then cast doubt on our ability to grasp first principles by arguing that even the most rudimentary and particular forms of knowledge - such as grasping some particular fact - are problematical. Sceptical practical wisdom, such as it was, became something like the ability to withhold assent from any particular proposition, thus avoiding dogmatism and achieving quietude (*ataraxia*). The connection with knowledge was thus either broken (for practical wisdom) or required a vast and difficult defence of knowledge (for theoretical wisdom) before philosophy could even return to wisdom as a subject of study. To a substantial degree, contemporary philosophy, and its preoccupations with justification and information, continues to focus its energies on the areas problematized by the ancient Sceptics.

Early Church Fathers, on the other hand, were generally suspicious of what the Greeks had identified as wisdom. They generally agreed that wisdom at best subserved Christian faith (see, for example, Augustine’s *City of God* 8.8); at worst they simply identified it with useless and deceitful book-learning and arrogant intellectual presumption, perhaps following St Paul’s warning against philosophy in *Collossians* 2: 2-8 (see, for examples, Justin Martyr *Dialogue with Trypho* 2.4-5, 7.1; Tertullian *Apology* 46.5-8). Subsequent epistemological disputes within religious contexts have also largely failed to revive wisdom as an object of rational pursuit, instead attending mainly to the degree to which various religious claims can or cannot be justified.

Wisdom can be found among medieval conceptions of the virtues, but it no longer held the central place it had been afforded by the Greeks. By the time we reach the early modern period, we find wisdom mentioned only rarely, and then only at the periphery of greater works (as in the preface to the French edition of Descartes’ *Principles of Philosophy*), or in revivals of ancient Greek points of view (as in the neo-Sceptical work, *De la sagesse* by Pierre Charron). From this period on, wisdom is mentioned only in passing, or simply passed by altogether by philosophers.

Given its disappearance from our discussions, none of the claims by the early Greeks has been sufficiently well scrutinized by philosophers. It is interesting, for example, that the development of fallibilism in epistemology has not served to liberate wisdom from its early association with infallibility, which was guaranteed to make wisdom seem an impossible achievement for human beings (see Fallibilism). On the other hand, the early connection between wisdom and some notion of a good life seems exactly right, but contemporary philosophy is deeply divided over what makes a life good. Perhaps as the criteria for human knowledge relax and conceptions of the good life (or good lives) become clearer, philosophy will find that it can return to the study of wisdom.

See also: Aretē; Eudaimonia; Nirvāna; Nous; Virtue epistemology; Zhi

NICHOLAS D. SMITH

References and further reading

Aristotle (*c.* mid 4th century BC) *Eudemian Ethics*, ed. J. Barnes in *The Complete Works of Aristotle*, vol. 2, 1922-81, Princeton, NJ: Princeton University Press, 1984.(Scholars have disputed whether this or the *Nicomachean Ethics* was written first. This work seems to be more interested in practical wisdom.)

Aristotle (*c.* mid 4th century BC) *Metaphysics*, ed. J. Barnes in *The Complete Works of Aristotle*, vol. 2, 1552-728, Princeton, NJ: Princeton University Press, 1984.(In a certain sense, this is Aristotle’s attempt to define and describe philosophy, and hence what the ‘love of wisdom’ consists in, in its purest sense. In many ways the *Metaphysics* is at the centre of nearly all the rest of Aristotle’s philosophy. Definitely not for beginners.)

Aristotle (*c.* mid 4th century BC) *Nicomachean Ethics*, ed. J. Barnes in *The Complete Works of Aristotle*, vol. 2, 1729-867, Princeton, NJ: Princeton University Press, 1984. (Compared to the *Eudemian Ethics*, this work is more interested in theoretical wisdom, especially in Book X.)

Augustine (413-27) *City of God*, trans. D.S. Wiesen, Loeb Classical Library, vol. 3, Cambridge, MA: Harvard University Press, 1968.(One of Augustine’s three major works - his most detailed review and critique of ancient Greek ‘wisdom’ may be found here.)

Blanshard, B. (1967) ‘Wisdom’, in P. Edwards (ed.) *The Encyclopedia of Philosophy*, vol. 7, New York and London: Macmillan, 322-4.(Discussion of the classical accounts of wisdom.)

Charron, P. (1601/1604) *De la sagesse (On wisdom)*, ed. B. de Negroni, Paris: Fayard, 1986.(A neo-sceptical

work, which was influential in the development of libertinism and fideism in the 17th century. The main argument is that human beings can discover truth only by revelation.)

Descartes, R. (1647) *Principles of Philosophy* (Preface to the French edition), trans. J. Cottingham in *The Philosophical Writings of Descartes*, vol. 1, 179-90, Cambridge: Cambridge University Press, 1985. (Descartes' penultimate work. The topic of wisdom appears only in the preface to the French edition, and the connection of this topic to Descartes' other epistemological preoccupations is only barely made.)

Justin Martyr (*c*.150) *Dialogue with Trypho*, in J.C.M. Van Winden, *An Early Christian Philosopher*, Leiden: E.J. Brill, 1971, 7-15.(Belittles the 'wisdom' of all the philosophical schools of his day, except that of the Platonists, claiming that the end of Plato's philosophy was to 'look upon God'.)

Kekes, J. (1983) 'Wisdom', *American Philosophical Quarterly* 20(3): 277-86. (One of the few discussions of wisdom in contemporary philosophy.)

Plato (*c*.395-387 BC) *Apology*, trans. H.N. Fowler, Loeb Classical Library, Cambridge, MA: Harvard University Press, 1914.(Plato's version of Socrates' three speeches at his trial: the defence proper, the counter-penalty proposal, and some final words to his jurors after his condemnation. Plato's Socrates describes himself as wisest among men only because he alone recognizes that he wholly lacks wisdom.)

Plato (*c*.380-367 BC) *Republic*, trans. P. Shorey, Loeb Classical Library, Cambridge, MA: Harvard University Press, 1914.(According to Plato, political troubles can only be brought to an end in the cities 'if philosophers become rulers or rulers become philosophers'. Books V-VII give Plato's account of the kind of knowledge that makes philosophers wise as a power (dunamis) to be developed through rigorous education.)

Sternberg, R.J. (ed.) (1990) *Wisdom. Its Nature, Origins, and Development*, Cambridge: Cambridge University Press.(Collection of essays on wisdom by psychologists.)

Tertullian (*c*.197) *Apology*, trans. T.R. Glover in *Tertullian: Apology, De Spectaculis; Minucius Felix*, Loeb Classical Library, Cambridge, MA: Harvard University Press, 1931.(Disparages earlier Greek philosophers, even to the point of accusing Socrates of having been influenced by the devil.)

Witherspoon, John (1723-94)

John Witherspoon, Scottish-American clergyman, political leader and educator, was born at Gifford, East Lothian, educated at Edinburgh University and ordained Presbyterian minister. In his mid-forties he went to America as president of the College of New Jersey, later Princeton University. He held political office for New Jersey and played a major role in organizing the Presbyterian Church in America and improving the College at Princeton. Witherspoon was representative of eighteenth-century Scottish and American Calvinists who tried to reconcile their orthodox theological doctrines with the Enlightenment's philosophical currents of empiricism, scepticism, and utilitarianism by harmonizing reason and revelation. Although Witherspoon was a philosophical eclectic, Francis Hutcheson's moral sense philosophy was the major source of his utilitarian ethics and republican politics. Witherspoon was not an original thinker, but his popularization of Scottish common sense and moral sense philosophy through his forceful personality and effective teaching laid the foundation for its dominance of nineteenth-century American academic philosophy.

John Witherspoon was born at Gifford, East Lothian, Scotland on 5 February 1723 and died at Princeton, New Jersey, United States of America on 15 November 1794. A man of action but also of reflection, Witherspoon combined a strong interest in philosophical and theological questions with clerical, educational, and political activism. In Scotland (1723-68), he was a notable parish minister, theologian, and ecclesiastical politician in the Church of Scotland. Witherspoon gained fame in America (1768-94) as a cleric, educator, philosopher, and national leader in Church and state as president of the College of New Jersey, later Princeton University.

Witherspoon wrote numerous works on a wide variety of subjects, most notably moral philosophy, rhetoric, and theology. His divinity lectures, sermons and theological treatises stressed the practical ethics of orthodox Calvinist dogma: God's sovereignty; the individual's sinful nature; God's justification of the sinful through Christ; scriptural authority; and regeneration (see Calvin, J.).

Witherspoon's introduction to philosophy began as an Edinburgh University student with the epistemological empiricism of John Locke and the logic and moral philosophy of Johann Gottlieb Heineccius. He continued to read widely in philosophy during his Scottish years, but he only wrote two brief works on philosophy of religion. His 1739 university graduation thesis on the immortality of the mind relied on traditional philosophical and theological arguments, while his 'Remarks on an Essay on Human Liberty' (1753) attempted to refute Kames' position on the freedom of the will (see Home, H. (Lord Kames)) using a naïve version of Thomas Reid's common sense realism (see Reid, T.; Common sense school). Both works attempted to reconcile the conflicting claims of reason and revelation in defence of Calvinism against the attacks of philosophical sceptics.

Witherspoon's Scottish intellectual background provided the basis for his outstanding American career as churchman, educator, philosopher and politician. As Princeton's president (1768-94), Witherspoon stressed the unity of learning and piety in its curriculum. Required to teach philosophy for the first time, Witherspoon's most important contribution was his course on moral philosophy. His *Lectures on Moral Philosophy* (1768) derived from a wide variety of modern philosophers, but their eclectic nature caused him to make numerous philosophical errors. They were based primarily on Scottish thought, especially Hutcheson's moral sense philosophy, interpreted in the context of Witherspoon's Calvinism (see Hutcheson, F.; Moral sense theories). Although the organization and subjects of the *Lectures* closely followed Hutcheson's works, they did not simply mimic Hutcheson's ideas, for on critical points Witherspoon's arguments expanded on and took different positions from Hutcheson's.

Throughout his nine brief lectures on ethics Witherspoon's epistemology rested on his claim that reason and revelation combined to provide the foundations of human knowledge. But like Hutcheson, he was also a social Newtonian who conceived of moral philosophy an as empirical science of human beings and society, based on reason and independent of revelation. Ethics was 'that branch of Science which treats the principles and laws of Duty or Morals' (Witherspoon 1768: 269). Individuals acquired their knowledge of the world through external and internal senses. Witherspoon attacked Berkeley's idealism, which he found at Princeton (see Berkeley, G.) and replaced it with Reid's realism. In his faculty psychology he agreed with Hutcheson that the operations of the affections or passions were more important mental processes than the understanding or the will, but added religious affections to Hutcheson's incomplete catalogue of feelings.

The critical faculty was the moral sense, because it allowed persons to judge actions as evil or good. Joseph Butler’s concept of conscience influenced both Hutcheson and Witherspoon, although Witherspoon mistakenly equated the moral sense ’with what, in scripture and common language, we call conscience’ (1768: 281) (see Butler, J. §3; Conscience). Contrary to Hutcheson, Witherspoon followed Reid in arguing that reason played a critical role in aiding the moral sense to discover moral truth. The moral sense provided the philosophical linkage between Witherspoon’s epistemological, psychological and moral ideas and his theological doctrines. Since it originated in God, it bound all persons to the creator in a moral relationship. Thus, the moral sense harmonized reason and revelation. Witherspoon adopted Hutcheson’s utilitarian position ‘that benevolence or public affection is virtue, and that a regard to the good of the whole is the standard of virtue’ (1768: 285). As to the obligation of virtue, he favoured duty over interest, because the moral sense intimates ’that it is duty independent of happiness’ (1768: 291). Benevolence was the central virtue in motivating persons to the duties they owed God, their fellow human beings, and themselves. Witherspoon’s discussion of rights emphasized the inalienable rights of intellectual, political, and religious liberty as well as self-preservation.

Witherspoon’s political thought, explicated in three short lectures on politics, domestic society and civil society, was a logical extension of his ethics as a science of society. His republican political ideas, like his ethics, were eclectic, generally followed Hutcheson’s, and sought always to harmonize his philosophical and theological ideas. A political realist unlike Hutcheson, Witherspoon argued that civil society originated both in Locke’s social contract and in Calvin’s religious covenant. He thought that a wide distribution of property was necessary for the public good; that a republic was the best form of government; and that the right of revolution was fundamental. The value and advantage of civil liberty was not that it caused virtuous behaviour or insured individual freedom, but rather that it was useful. Witherspoon elaborated his republicanism in pamphlets, sermons, speeches and other works during the American Revolution. His moral philosophy and theology led him to conclude that the Revolution resulted in the moral regeneration of the American people and the creation of a virtuous, Christian republic.

See also: American philosophy in the 18th and 19th centuries; Rights; Utilitarianism

R.J. FECHNER

List of works

Witherspoon, J. (1800-1) *The Works of the Rev. John Witherspoon*, Philadelphia, PA: Woodward.(This first edition contains all of Witherspoon’s major works, and several, but not all, of his minor writings.)

Witherspoon, J. (1753) ’Remarks on an Essay on Human Liberty’, *Scots Magazine* XV: 165-70.(Witherspoon’s attempt to refute Kames’ scepticism.)

Witherspoon, J. (1768) *Lectures on Moral Philosophy*, in *The Works of the Rev. John Witherspoon*, Philadelphia, PA: Woodward, vol. 3.(Not published until 1800.)

Witherspoon, J. (1982) *An Annotated Edition of Lectures on Moral Philosophy by John Witherspoon*, ed. J. Scott, Newark, DE: University of Delaware Press.(The most recent and accessible edition of the *Lectures* with excellent annotations, bibliography and introduction.)

References and further reading

Collins, V. (1925) *President Witherspoon: A Biography*, Princeton, NJ: Princeton University Press, 2 vols.(The standard biography, portraying Witherspoon in great detail as a man of action.)

Fechner, R. (1982) ’The Godly and Virtuous Republic of John Witherspoon’, in H. Craven (ed.) *Ideas in America’s Cultures: From Republic To Mass Society*, Ames, IA: Iowa State University Press, 7-25.(A succinct summary of the relationship between Witherspoon’s moral philosophy and theology relative to his political ideas.)

Fechner, R. (1974) *The Moral Philosophy of John Witherspoon and the Scottish-American Enlightenment*, unpublished Ph.D. dissertation, University of Iowa, Ann Arbor, MI: Xerox-University of Michigan Microfilms, order no. 75-13, 747.(The fullest exposition of Witherspoon’s moral philosophy in its Scottish and American cultural, intellectual and religious contexts.)

Flower, E. and Murphey, M.G. (1977) ’Philosophy in Academia Revisited - Mainly Princeton’, *A History Of Philosophy In America*, New York: Putnam, vol. 1, 203-73.(A sophisticated analysis of the place of Scottish philosophy in the history of American philosophy with special focus on Witherspoon’s role, 226-38.)

Landsman, N. (1990) ’Witherspoon and the Problem of Provincial Identity in Scottish Evangelical Culture’, in R.

Sher and J. Smitten (eds) *Scotland and America in the Age of Enlightenment*, Edinburgh: Edinburgh University Press, 29-45.(A subtle discussion of Witherspoon as a Scottish and American provincial in the wider cultural and religious world of eighteenth-century Britain.)

Meyer, D. (1976) *The Democratic Enlightenment*, New York: Putnam, 182-209.(A sensitive characterization of Witherspoon as a public intellectual in the American Enlightenment through his role as an educator.)

Noll, M. (1989) *Princeton and the Republic, 1768-1822*, Princeton, NJ: Princeton University Press, 16-58.(A skilful weaving together of several strands of Witherspoon's thought in its academic and religious contexts.)

Sloan, D. (1971) *The Scottish Enlightenment and the American College Ideal*, New York: Columbia University Teachers College Press, 101-45.(A thoughtful study of Witherspoon's role in the transmission of the Scottish Enlightenment to American Presbyterianism and to Princeton.)

Wittgenstein, Ludwig Josef Johann (1889-1951)

Ludwig Wittgenstein was born in Vienna on 26 April 1889 and died in Cambridge on 29 April 1951. He spent his childhood and youth in Austria and Germany, studied with Russell in Cambridge from 1911 to 1914 and worked again in Cambridge (with some interruptions) from 1929 to 1947.

His first book, the Tractatus Logico-Philosophicus, was published in 1921. It presents a logical atomist picture of reality and language. The world consists of a vast number of independent facts, each of which is in turn composed of some combination of simple objects. Each object has a distinctive logical shape which fits it to combine only with certain other objects. These objects are named by the basic elements of language. Each name has the same logical shape, and so the same sort of possibilities of combination, as the object it names. An elementary sentence is a combination of names and if it is true it will be a picture of the isomorphic fact formed by the combination of the named objects. Ordinary sentences, however, are misleading in their surface form and need to be analysed before we can see the real complexity implicit in them.

Other important ideas in the Tractatus are that these deep truths about the nature of reality and representation cannot properly be said *but can only be* shown*. Indeed Wittgenstein claimed that pointing to this distinction was central to his book. And he embraced the paradoxical conclusion that most of the Tractatus itself is, strictly, nonsense. He also held that other important things can also be shown but not said, for example, about there being a certain truth in solipsism and about the nature of value. The book is brief and written in a simple and elegant way. It has inspired writers and musicians as well as being a significant influence on logical positivism.*

After the Tractatus Wittgenstein abandoned philosophy until 1929, and when he returned to it he came to think that parts of his earlier thought had been radically mistaken. His later ideas are worked out most fully in the Philosophical Investigations, published in 1953.

One central change is from presenting language as a fixed and timeless framework to presenting it as an aspect of vulnerable and changeable human life. Wittgenstein came to think that the idea that words name simple objects was incoherent, and instead introduced the idea of 'language games'. We teach language to children by training them in practices in which words and actions are interwoven. To understand a word is to know how to use it in the course of the projects of everyday life. We find our ways of classifying things and interacting with them so natural that it may seem to us that they are necessary and that in adopting them we are recognizing the one and only possible conceptual scheme. But if we reflect we discover that we can at least begin to describe alternatives which might be appropriate if certain very general facts about the world were different or if we had different interests.

A further aspect of the change in Wittgenstein's views is the abandonment of solipsism. On the later view there are many selves, aware of and co-operating with each other in their shared world. Wittgenstein explores extensively the nature of our psychological concepts in order to undermine that picture of 'inner' and 'outer' which makes it so difficult for us to get a satisfactory solution to the so-called 'mind-body problem'.

Although there are striking contrasts between the earlier and later views, and Wittgenstein is rightly famous for having developed two markedly different philosophical outlooks, there are also continuities. One of them is Wittgenstein's belief that traditional philosophical puzzles often arise from deeply gripping but misleading pictures of the workings of language. Another is his conviction that philosophical insight is not to be gained by constructing quasi-scientific theories of puzzling phenomena. Rather it is to be achieved, if at all, by seeking to be intellectually honest and so to neutralize the sources of confusion.

1 Life

Wittgenstein was the eighth and last child of a wealthy Austrian industrialist. From 1903-8 he was educated on the assumption that he would be an engineer and in 1908 he came to Manchester to study aeronautics. He continued with this for three years, but at the same time developed his interest in philosophy. He was particularly engaged with logic and the foundations of mathematics, in connection with which he read Frege (§§6-10) and Russell (§§4-11). In October 1911 he gave up engineering and, on Frege's advice, came to Cambridge to study with Russell. In the 1914-18 war he served in the Austro-Hungarian army and during this time completed the *Tractatus Logico-Philosophicus* (1922).

For a time Wittgenstein thought that the *Tractatus* said everything which could be said in philosophy, and so he turned to other things. From 1920 to 1926 he was a schoolteacher in Austria, though this was not a success, since he was severe and demanded too much of his pupils. In 1926-8 he helped to design a house for his sister. In 1927 he resumed philosophical discussion with some members of the Vienna Circle, and in 1929 he returned to Cambridge, lecturing there from 1930 to 1936. From 1936 to 1938 he visited Norway and Ireland, returning to Cambridge in 1938 and being appointed professor there in 1939. He held the chair until 1947, although from 1941 to 1944 he was given leave of absence to work first at Guy's Hospital, London, then at the Royal Victoria Infirmary, Newcastle. After resigning his chair in 1947 he lived in various places, Ireland chief among them, and also visited America.

Wittgenstein impressed those who met him with the power of both his intellect and personality. He had an intense concern for truth and integrity which exerted great attraction, but which also made him difficult to deal with, since he was liable to accuse others of superficiality or dishonesty. He greatly disliked what he perceived as the artificiality and pretentiousness of academic life. His later ideas became known in the 1930s and 1940s through the circulation of copies of *The Blue and Brown Books* (1958) and reports of his lectures, and they acquired considerable influence.

2 Works and method of writing

Throughout his life Wittgenstein wrote down his thoughts in notebooks, returning to the same topics many times, trying to get the most direct and compelling formulation of the ideas. He then made selections and arrangements from these remarks, followed by yet further selection, reworking and rearrangement. The *Tractatus* was the only book published during his lifetime. In 1930 he assembled what we now know as the *Philosophical Remarks*, a work still having much of the outlook of the *Tractatus* and also showing considerable sympathies with verificationism, and in 1932-4 he wrote the *Philosophical Grammar*, in which some central themes of the later philosophy are foreshadowed. But he was not satisfied with either of these, and from 1936 onwards worked on various versions of what we now know as the *Philosophical Investigations* (1953), which he hoped would provide a definitive presentation of his thought. The earlier half of the volume is the part of his work with which he was most nearly satisfied, but he was never fully content with any of it, and in 1949 he abandoned the project of completing it.

The other books we have under his name are all early or intermediate versions of material, left in his papers and edited and published after his death. The *Notebooks* are preliminary versions of ideas which later became the *Tractatus*. *The Blue and Brown Books* were prepared so as to help his students in 1932 and 1933. *Remarks on the Foundations of Mathematics* (1956) contain ideas he worked on from 1937 to 1944 and which he intended at that time to form the second part of the *Investigations* (rather than the psychological topics we now have). From 1944 onwards he worked mainly on philosophical psychology: *Zettel* (1967), *Remarks on the Philosophy of Psychology I and II* (1980) and *Last Writings on Philosophical Psychology I and II* (1982) are from these years. From 1950 to 1951 we also have *On Certainty* (1969) and *Remarks on Colour* (1977). Another source for his views is records of his conversations and lectures taken by friends and pupils.

3 The picture theory of meaning

The *Tractatus* consists of nearly eighty printed pages of numbered remarks. The numbers do not run consecutively but are designed to indicate the relative importance and role of the remarks. There are seven major sentences and each of them (except 7) has subordinate and clarificatory remarks following it, labelled '2.2', '5.4' and so on, down as far as such numbers as '4.0312'.

The topics that preoccupied Wittgenstein when he arrived in Cambridge included the nature of logical truth and Russell's Theory of Types. On both of these matters Russell held that we need an account of very general features of the world and of the kinds of things in it. But Wittgenstein soon came to think that the route to insight was through the contemplation of the nature and presuppositions of individual meaningful sentences such as 'Socrates is wise' and that this contemplation showed Russell's approach to be misguided.

The central fact about such individual sentences is that each says one thing - that Socrates is wise, for example - but is essentially such that it may be either true or false. A false sentence is both out of touch with the world,

inasmuch as it is false, but also in touch with the world, inasmuch as it succeeds in specifying a way that things might be. Wittgenstein holds that all this is possible only because the sentence is complex and has components which represent elements of reality, which exist whether the sentence is true or false and are (potentially) constituents of states of affairs. So, in rough illustration, 'Socrates' represents Socrates and 'is wise' represents wisdom. The truth or falsity of the sentence then depends on whether these elements are or are not assembled into a fact.

Not all sequences of sentence components are acceptable. A mere list of names (for example, 'Socrates Plato') does not hang together as a sentence. And although it looks as if we may apply a predicate to itself (as in '''is in English'' is in English' or '''is wise'' is wise') it seems important to disallow such sequences as truth-evaluable sentences, on pain of falling into Russell's paradox. Russell's account of these matters is that elements of reality come divided into different types - individuals, properties and so on - and that a sentence is to be allowed as meaningful, and so truth-evaluable, only if the elements picked out by the components are of suitably related types.

Wittgenstein maintains, against this, that we do not need rules to bar sentences which would lead to paradox, because when we properly understand the nature of our language we see that we cannot formulate the supposed sentences in the first place. We think we can only because we have misidentified that component in 'Socrates is wise' whose presence in the sentence attributes wisdom to Socrates. This component is not the phrase 'is wise' but the property which the word 'Socrates' has when the words 'is wise' are written to the right of it. To see why this is plausible, consider the fact that there could be a language in which properties are attributed to people by writing their names in different colours. If, for example, we could claim that Socrates is wise only by writing 'Socrates' in red letters then we could never formulate any analogue to '''is wise'' is wise' because we could not take the redness of 'Socrates' and make it red. And although we have given a linguistic role to 'Plato' (as representing Plato) we have not given any role to the property of a name which it acquires when we write 'Plato' to the right of it. That is why a mere list of names does not hang together to make a statement.

Wittgenstein generalizes this idea to claim that the formal properties of any element of the world, that is, the properties which fix its potential for combining into facts with other elements, must be mirrored in the formal properties of the linguistic component which represents it. So the kind of item we call a sentence, and which we often wrongly think to be a complex object, is really a fact. In a sentence certain linguistic components are put together experimentally in a structure which mirrors the formal structure of some possible state of affairs (see Theory of types §1).

4 Negation and tautology

This account does away with the need for a theory of types and Wittgenstein holds that the ideas invoked by Russell to explain the nature of logical truth are similarly unnecessary. His view was that logical truths, such as that all sentences of the form 'p or not p' are true, should be explained by pointing to relations holding between some very abstract kind of logical items - negation, disjunction and the like.

Wittgenstein maintains that the negation sign is not a component of a sentence and so does not represent any element of a possible fact (in the quasi-technical sense of 'component', 'element' and 'represent' introduced above). Rather it is the visible mark of an operation one can perform on a meaningful sentence to produce another sentence. The role of the second sentence is to deny that things are as they would need to be to make the first sentence true.

To see the force of this, we must look again at the account of truth given above. What makes a negative sentence true is not the presence of some 'not-ness' in a fact but rather the absence of that (the combination of elements) which would have made the unnegated sentence true. This is similarly the case with the other so-called logical constants. Thus 'or' does not stand for a possible element in a fact but is a sign by which one can correctly link two sentences if the components of either are so combined as to yield a truth.

Logical tautologies thus do not reveal the nature of special logical objects. Such sentences as 'It is raining or it is not raining' do not say anything. But their possibility is a corollary of the existence of a language adequate to say the kind of thing which can be said. So contemplating them can draw our attention to the logical structure of the world (see Logical constants; Logical form; Logical laws).

5 Simples

The ideas outlined in the last two sections are already present in the sets of notes which Wittgenstein wrote in late-1913 and 1914. But the *Tractatus* in its complete form incorporated several further important ideas. One of these, perhaps adopted from Frege, is the view that sense must be determinate, that is, that every meaningful sentence must be either true or false in every possible state of affairs. But Wittgenstein differs from Frege in thinking that ordinary language, although misleading in surface form, is in order, and so already fulfils this condition of determinacy.

Determinacy entails that there must be 'objects', that is, utterly simple, eternal and unalterable elements, out of which all facts are composed. Moreover the links between our language system and reality must be set in place at this basic level. Suppose that language-world links were set up so as to connect a basic linguistic component with some element of reality which was not basic. The existence of this element would be contingent and would depend upon some simpler elements being suitably combined in a fact. A sentence containing this imagined basic component is clearly not true in a world where the simpler elements are not suitably combined. But it is equally unhappy to say that it is false, because the component itself does not specify what the simpler elements are or that they must be combined, and so it is no part of its meaning that their failure to be combined is relevant to its falsehood. To insist on the undefinability of this imagined basic element of language is to insist that it has meaning only through its connection with the item it represents. So in a world lacking that item it has no meaning, and sentences containing it are neither true nor false. But this, given the assumption of determinacy of sense, is a *reductio ad absurdum* of the idea that meaning can be conferred in this imagined way.

We may have an apparently basic sentence component which is linked to a contingently existing item (for example 'Socrates' as a name of Socrates). But this is only possible because a definition of that component can be given in our language system. The link between Socrates and his name is thus not a basic point of attachment between language and the world (contrary to the impression given in our earlier rough and ready example). It is a consequence of these ideas that there must be a complete analysis of every sentence of our language into a truth-functional combination of elementary propositions, the components of which are simple signs representing objects.

But what are these simple objects? Wittgenstein, like Russell but unlike Frege, does not allow for any contrast between sense and reference within the meaning possessed by names of simples. This puts one demand on simples: they cannot be items with distinguishable aspects, that is, items which can be conceived of in several logically distinct ways. If they were, then there would also be the possibility of one name for a simple as conceived one way and another non-synonymous name for it as conceived another way - contrary to the denial of the sense-reference distinction. So a simple is the kind of thing which, if apprehended in such a way that it can be named, is apprehended exactly as it is in its entirety.

In so far as Wittgenstein drops any hints, it is that simples are phenomenally presented items, such as points in the sensory field and the properties they have, for example, shades of experienced colour. But he cannot give this answer officially because to do so would clash with another of the themes brought to prominence in the later development of the *Tractatus*. This is the claim that all necessity is logical necessity, and hence would be revealed as tautological in a complete analysis (see §7 below). A corollary of this is that all atomic facts are independent of each other and no elementary sentence can entail or be the contrary of any other. Such things as colours cannot then be 'objects' because attribution of different colours to one thing, as in 'a is red' and 'a is green', produces sentences which are contraries.

The topics so far discussed are treated primarily in the remarks following the main sentences numbered 2, 3 and 4. The remarks following 5 and 6 deal mainly with implications of this atomist conception for certain issues in logic (generality and identity for example) and for the nature of science, mathematics and statements of probability. On the last mentioned subject, Wittgenstein's brief remarks are one important source for the approach later developed by Rudolf Carnap (see Carnap, R. §5; Logical atomism §1).

6 Thought, self and value

Wittgenstein writes, 'There is no such thing as the subject that thinks or entertains ideas' (5.631). His grounds for

this are similar to those of Hume, namely that a unified, conscious self cannot be an element in any encountered fact. Hence no such item can be among those objects represented in thought. So reports of the form ‘A believes that p’, which seem to mention such a subject, are really of the form ‘"p’’ says that p’. They report the existence of a sentential complex, the components of which are correlated with the elements of the potential fact that p. There are then no selves in the world but, at best, bundles of sentence-like items.

But Wittgenstein also says ‘What the solipsist means is quite correct; only it cannot be said but makes itself manifest’ (5.62), and ‘The subject does not belong to the world: rather, it is a limit of the world’ (5.632). One interpretation connects these remarks with the idea of projection - he speaks of using a propositional sign as a projection of a possible situation by thinking out its sense. The picture suggested is that of a subject who is the origin of the lines of projection which link representing items with what they represent and whose existence is thus presupposed by their meaningfulness. ‘The world is *my* world: this is manifest in the fact that the limits of *language* (of that language which alone I understand) mean the limit of *my* world’ (5.62). So perhaps Wittgenstein’s idea is that the existence of a unique self (me) at the limit of the world is shown by the existence of representations which are meaningful to me.

Wittgenstein also offers, in the closing pages of the *Tractatus*, a number of gnomic remarks about value, death and the mystical, among them that no value exists in the world, that ethics cannot be put into words, that the will as a subject of ethical attributes cannot alter facts but only the limits of the world, that at death the world does not alter but comes to an end, that feeling the world as a limited whole is the mystical and that the solution of the problem of life is seen in the vanishing of the problem. These claims are to some extent intelligibly grounded in ideas concerning the self and what can be said. But they also represent a leap of development beyond those, a leap which comes in part from Wittgenstein’s experiences in the First World War and the religious convictions to which his always intense and serious outlook then led him.

7 Saying and showing

Wittgenstein thought that by reading the *Tractatus* one might come to grasp certain things about the nature of meaning, reality and value, among them the kinds of things outlined above. But he also held that these things could not be *said* but only *shown*, and the attempt to say them ends up producing nonsense. Most of the *Tractatus* itself is thus nonsense. This claim is highly paradoxical and may seem to be unnecessary and grandiose mysticism, especially when viewed as part of the same package as the difficult remarks about the will and solipsism. But this is unfair. In many of its applications, the claim is well motivated, given the picture theory of meaning.

Most of the linguistic manoeuvres which Wittgenstein condemns as nonsensical are attempts to say things which are both necessary but also substantive - that is, not mere tautologies. Thus they include moves to assign elements of reality or language to their logical types (‘Socrates is a particular’), related attempts to describe the logical forms of sentences or facts and also efforts to list the simples. (Claims about what is valuable could also have this status of seeming to be both substantial and necessary.) But if the picture theory of meaning holds in complete generality there cannot be such statable necessary truths. To say, for example, that object b is F we require that there be a linguistic representative of b (‘b’) and one of F-ness (the property of having ‘F’ to the right) which can be combined or not, just as b and F-ness can be combined or not. There must be complexity and there must therefore be the possibility of dissociation as well as association. But if b and F-ness necessarily go together (for example, Socrates must be a particular) then b cannot be dissociated from F-ness and it is a confusion to imagine that its being F is a fact with a composite structure. Hence it is also a confusion to imagine that it is something of the kind which can be said, on the account of saying which is offered by the picture theory.

8 Transition

In the *Tractatus* Wittgenstein took a lofty tone about simple objects; he had proved that they existed and it was of no importance that we cannot say what they are. But in the first years of his return to philosophy in 1927-31 one thing which occupied his attention was the detailed workings of various parts of our language, notably those involved in talking of shape, length and colour and other observable properties of items around us. His aim in considering them was to fill in that earlier gap by giving an account of the fundamental features of both language and the world.

He soon became convinced that the idea of independent elementary propositions was indefensible. For example, the incompatibility of 'a is red' and 'a is green' cannot be explained (contrary to what he had urged in the *Tractatus*) by analysing the two propositions and showing that one contains some elementary proposition which contradicts an elementary proposition in the analysis of the other. Rather the whole collection of colour judgments come as one set, as the marks along the edge of a ruler come as a set. To measure an object we hold a ruler with its marks against the object, that is, in effect we hold up a whole set of possible judgments of length, and we read off which is correct; to see that one judgment is correct is to see at the same time that all the others are incorrect. Something similar holds for colour and for many other concepts, except that in these cases the 'ruler' is not physically present. The differences between concepts have to do with the logical shapes of their 'rulers' and with the different methods by which they are compared to reality.

Other topics which occupied Wittgenstein at this time were those of psychological phenomena and the use of the word 'I'. And he also worked extensively on the nature of mathematics. Ideas in common with those of the logical positivists are apparent in some of the writings of this time. Indeed the slogan known as the verification principle - 'the meaning of a proposition is its method of verification' - may have originated with Wittgenstein. But he found it impossible to accept this as a clear statement which could provide one of the starting points for elaborating a philosophical system. He was aware of further puzzles and was temperamentally incapable of putting them on one side for the purpose of building an intellectual construct which might be based on misapprehension and which failed to address questions which still perplexed him (see Logical positivism §§3-4; Vienna Circle §2).

9 Dismantling the *Tractatus* picture

Paragraphs 1-242 of the *Philosophical Investigations* (roughly the first third of the book) are generally agreed to provide the most focused presentation of some of the central ideas of Wittgenstein's later outlook, in the context of which his views on philosophy of mind, mathematics and epistemology can helpfully be seen. We may divide the paragraphs into three groups: §§1-88 raise a variety of interrelated difficulties for the outlook of the *Tractatus*, §§89-142 discuss the nature of logic, philosophy and truth, and §§143-242 contain the so-called rule-following considerations.

The first group has two main targets: first, the idea that most words have meaning in virtue of naming something and, second, the idea that meaning requires determinacy and so exactness. Against the former Wittgenstein points out that different words function in different ways. To understand 'five', for example, a person needs to be able to count and behave appropriately on the result of a counting; to understand a colour term might, by contrast, involve knowing how to compare the specimen to be judged with a sheet of samples. To teach language one must train a person to produce and respond to words in the context of everyday activities such as fetching things, measuring, building, buying and selling. We can throw light on meaning by reflecting on simple 'language games', involving such integration of speech and action. To say that every word names something is like saying that every tool in the tool box modifies something. We can describe things this way if we insist: 'The saw modifies the shape of the board; the ruler modifies our knowledge of a thing's length'. But such assimilation may lead us to overlook important variety rather than representing a useful insight. To get someone to understand a word it is not enough to bring them face-to-face with the supposed referent while repeating the word. In order to profit from the confrontation, the learner must know what kind of word is being taught (for a number, shape, colour, and so on). And this in turn involves already being at home with the everyday activities into which remarks using the word are woven. 'For a *large* class of cases - though not for all - in which we employ the word ''meaning'' it can be defined thus: the meaning of a word is its use in the language' (1953: §43).

On the second topic Wittgenstein remarks that drawing a contrast between 'simple' and 'complex' depends upon context and interest. Items which might be seen as complex in one context could be taken as simple in another and vice versa. So the related notion of exactness is also context-relative. A word does not become unusable and hence meaningless because its use is not everywhere bounded by rules. That we can imagine circumstances in which a given description would seem inappropriately vague or in which we would not know whether to say that it was true or false is no criticism of its current use and hence no argument that it does not have meaning. Hence the idea that every meaningful sentence must have some underlying analysis in terms of simples is mistaken.

Each sort of word is at home in its own language game. But there are not always clear-cut relations of subordination or dependence between different language games. There are many predicates (for example,

‘intention’, ‘thought’, ‘statement’, ‘number’, ‘game’) which clearly do not name simples, because they have interesting richness and apply to complex items. But it is not the case that such predicates must have an analysis in terms of ‘simpler’ predicates. Search for such an analysis may reveal instead a ‘family resemblance’. Persons who recognizably belong to the same family may have various resemblances, of build, features, colour of eyes, gait, temperament and so on, which ‘form a complicated network of similarities, overlapping and criss-crossing’ (1953: §66), without one set of such resemblances being necessary and sufficient for having the appearance of a member of that family.

On the basis of this survey of the actual workings of language, Wittgenstein then concludes that the *Tractatus* picture of a detailed crystalline structure present in both world and language is an illusion.

But perhaps the considerations outlined require only minor tinkering with the *Tractatus* picture? We might say: ‘Certainly linguistic representations of states of affairs are put to various uses, in commands, jokes, stories etc., as well as in straight reports; also what degree of exactness we need is fixed by context; hence we should accept that for many practical purposes vague remarks are adequate. But none of this shows that we must discard the *Tractatus* picture of a fully determinate world, structured by simples; nor does it show that the idea of constructing a complete and exhaustive description of it need be abandoned. All it shows is that our everyday language mirrors the world less accurately than an imaginable ideal scientific language.’ That such a reading is inadequate is shown by considering the remarks on rule-following found in paragraphs 143-242.

10 Rule-following

By ‘rule’ Wittgenstein does not mean an abstract standard according to which some act may be judged right or wrong. Rather he means a concrete item, such as a noise, mark or gesture, which is presented to a person and by attending to which they direct their behaviour, the link between rule and response being learned and conventional. An enormous number of human activities can be seen as instances of rule-following. The activities include imitating the gestures and noises which others make, copying shapes, converting marks into noises as in reading music, chanting the number sounds in sequence, and so on. More generally, both nonverbal behaviour in response to verbal instruction (fetching a book when told to do so) and also producing linguistic reports (where the world itself is the guide and the utterance is the response) may be described as rule-following. Rule-following is thus at the heart of linguistic competence. If we further accept that coming to use a rich and expressive language is an indispensable part of coming to grasp complex concepts and to make reflective judgments, then rule-following is also at the heart of our lives as thinking creatures.

It is generally agreed that Wittgenstein has telling negative points to make about one attractive but misleading picture of rule-following. On this picture, to understand a rule, for example, to grasp what is meant by ‘Add two’, it is necessary and sufficient to have a certain sort of item, an image, feeling or formula, occur in the mind when the instruction is heard. For example, having a mental image of two blocks appearing at the end of a line of blocks is the sort of thing which might be imagined to constitute understanding ‘Add two’. This image is supposed to do two things. First, it helps bring about that the person goes on to produce a particular response, for example, saying ‘Eight’ if the previously given number was six; second, it sets a standard by which that response can be judged correct or incorrect.

But the picture will not do. A person might have such an image while responding to ‘Add two’ as if it meant ‘Multiply by two’. Moreover the person’s behaviour (the regular patterns of action, what seems to be regarded as a mistake, and so on) could show that for them, ‘Add two’ actually means ‘Multiply by two’. So images guarantee neither subsequent behaviour nor the appropriateness of a particular standard of assessment.

What the case makes us see is that an image, feeling or formula is merely another rule-like object (that is, a potential vehicle of meaning) rather than the meaning itself. An item is not automatically a self-interpreting sign, that is, one which fixes and enforces a certain reading of itself, simply in virtue of existing in the mind rather than in the outer public world. So images and the like are not sufficient for understanding; but neither are they necessary, since in many cases they do not occur. Typically when someone responds to everyday and familiar language they just act unhesitatingly and spontaneously, without consulting any inner item.

To teach someone to follow a rule, for example, to understand ‘Add two’, we put them through a finite amount of training, primarily by working through examples of adding two. These examples may appear to be another

resource for pinning down meaning. But being only finite in number, they are bound to have more than one feature in common. Thus they do not themselves determine a unique interpretation for the sign we associate with them. A learner might exhibit a future bizarre divergence from what is expected, for instance by saying that adding two to 1000 yields 1004. And if this occurred it would suggest that they had all along been struck by some feature other than the one intended.

The central point here is that, for there to be meaning, the rule-followers must have fixed on one rather than another of the various similarities between the teaching examples and have associated it with the rule, that is, with the mark or sign to which they respond. ‘The use of the word ‘‘rule’’ and the use of the word ‘‘same’’ are interwoven’ (1953: §224). But neither the examples nor the rule itself determine which similarity this is; and imagined inner surrogates, in which we would like to see the relevant resemblance encapsulated, turn out to be equally inefficacious.

These reflections do not just undermine one picture of the psychology of understanding. They are also relevant to the metaphysics of the *Tractatus*. If the world consisted of intrinsically articulated facts as envisaged in the *Tractatus*, then there would exist items, namely the simple objects, which would fix the one and only absolute standard of similarity. If there is a simple which is a common element in two separate facts then there is a basic real resemblance between those facts; if not, not. Every other real resemblance which can be meaningfully labelled, for example, by the predicates of everyday language or science, must be founded in simples. A putative linguistic expression which is not tied to some definite combination of simples is, on the *Tractatus* view, an expression without meaning which is merely randomly applied. Further, as we saw earlier, a simple is the kind of thing which, when apprehended, must be apprehended as it is. So representing a simple, whether by a direct cognition of it or by having in mind something which encapsulates its nature, is to be aware of a self-interpreting item, something which dictates what is to count as ‘the same’. But this sort of confrontation is what the rule-following considerations suggest to be unintelligible.

Thus the discussions of §§143-242 can be seen as interweaving with and reinforcing those of §§1-88. The whole undermines not only the idea of closeness of fit between a *Tractatus* world and everyday language, but the underlying conception of that world itself, namely as already determinately articulated into facts by simples which we can apprehend (see Meaning and rule-following §§1-2).

11 The later picture of meaning

The *Tractatus* view offers us a world articulated of its own nature into value-free facts which are the subject matter of the natural sciences together with a mind confronting that world and attempting to mirror it in its thoughts. It also tells us that there is only one self and that it is an item at the limit of the world which cannot act responsibly in the world.

However attractive the first element here, everyone would agree that there is something seriously wrong with the solipsism of the second. So one essential move in amendment is to make the self responsibly active, to bring it in from the limit and to locate it firmly in the world, together with other selves. We may do this while leaving in place the idea of the world as the totality of value-free facts. Then the self which appears in the world must be some subset of such facts. This yields an extremely powerful and attractive overall picture. But it also generates many philosophical puzzles, those to do, for example, with giving naturalistic accounts of consciousness, free will, rationality and so on.

This overall picture cannot be Wittgenstein’s, however, if §10 is right in its reading of the rule-following considerations. The idea of an independently articulated world is not acceptable to him. We cannot understand our concepts by pointing to simples in the world which force them on us. To understand meaning we must look to use, at how our actions and concepts are interwoven. The fact that makes a sentence true is grasped through seeing when the sentence is correctly used, and that in turn is grasped only by seeing the full shape of the language games in which it is used. For a concept to be truly applicable to the world, and so for its corresponding property to have instances, is not for it to pick out some simple which is among the timelessly given building-blocks of all worlds. Rather it is for the life of which use of that concept is a part to be liveable in this world. Wittgenstein thus moves from a bold and simple form of the correspondence theory of truth in the *Tractatus* to a redundancy theory in the *Investigations*.

The self need not, on this view, be an assemblage of value-free facts. It is rather a locus of abilities, a person who can be trained to follow rules, to use and respond to language, in the way normal humans can. And since concepts are aspects of our way of life rather than items forced on us by the world, understanding what it is to have a particular concept involves 'assembling reminders' about how it works for us and how our various activities and ways of talking build together into our way of life.

If it is correct to conceive of understanding as an ability, then the exercise of this ability in everyday situations will often be just some confident, spontaneous action or utterance, which the subject will not be able to justify by pointing to something, other than the situation or words responded to, which guided them.

> 'How am I able to obey a rule?' - if this is not a question about causes, then it is about the justification for my following the rule in the way I do. If I have exhausted the justification I have reached bedrock and my spade is turned. Then I am inclined to say: 'This is simply what I do'.
>
> (1953: §217)

But this need not worry us. 'To use a word without justification does not mean to use it without right' (1953: §289). The fact is that we do find such confident and unhesitating responses in ourselves. Also we (usually) agree with others; and where we do not we (usually) agree on how to settle the dispute. So we have no reason to doubt that in general we do indeed mean what we take ourselves to mean.

Indeed we can put things more strongly than this. It is not just that it is sensible, practically speaking, for me to make a leap of faith and decide to carry on as if I and others mean what it seems we mean. We have no more choice about this than we do about taking 'Eight' to be the right response to 'Add two to six'. The language game of ascribing meanings to the remarks of ourselves and others is as central and indispensable to a recognizably human life as anything in our linguistic repertoire. Moreover the rich and complex social world in which we find ourselves sustains our practice of so doing. So we and our meanings are just as much part of the world as the stars, rocks and trees around us. And since we are no longer committed to the idea of one totality of facts, those of value-free natural science, this recognition does not now produce cramps or pressures to reductive manoeuvres (see Private states and language §4).

12 Alternative readings

The account of §§9-11 presents Wittgenstein as inviting us to abandon the idea of our meanings and judgments being securely moored to something outside us which imposes itself on us and keeps us in line. We are to become aware of our involvement in and responsibility for our own judgments and the way of life of which they are part. We are also to acknowledge that we cannot prove the unique correctness of our way of life and its associated concepts. (The arguments of the *Investigations* against the position of the *Tractatus* thus have much in common with themes explored by other late-nineteenth-century and twentieth-century thinkers, such as Nietzsche, William James, Heidegger, Quine and Derrida (see Nietzsche, F. §6; Heidegger, M. §§2-4; Quine, W.V. §5).) But, the reading given in §11 implicitly suggests, this need not lead us to scepticism about the notions of meaning, fact, objectivity or truth.

This interpretation, although not idiosyncratic, is by no means generally accepted. There are a large number of differing construals of Wittgenstein's overall intention, many of which have in common that they present the consequences of abandoning the *Tractatus* view as more radical and/or more deflationary, than is suggested in §11.

One interpretation stresses a contrast between the *Tractatus* and the later writings which is different from any highlighted earlier. It takes the rule-following considerations to show that we cannot make sense of grasp of meaning which fixes truth conditions independent of our ability to verify that they obtain. The later Wittgenstein is thought to insist (as against his earlier self) that all meaning be explicated by appeal to assertibility conditions rather than such possibly verification-transcendent truth conditions and he is recruited onto the antirealist side of the debate in the dispute between realism and antirealism.

Another much discussed view is presented by Saul Kripke. If there were facts about meaning, he argues, they would have to be constituted by something about past behaviour or present occurrences in the mind. So §§143-242 can be read as showing that there are no facts about meaning. Our practice of labelling remarks 'correct' or

'incorrect' and ascribing 'meanings' to them has a role in our social life. But such linguistic moves do not have truth conditions. Instead they have only appropriateness conditions. We are licensed to make them when others in our community keep in step with us in certain ways in their patterns of utterance.

Yet a third interpretation takes it that Wittgenstein espouses relativism. One response to the idea that there are no simples is to take it that the world is a featureless mush or unknowable something. Any apparent structure in it is then imposed by us. Hence the familiar physical and social world we experience is a creation of ours. But there are several possible but incompatible ways of imposing structure, one of which we are physiologically and/or socially caused to adopt. So no judgment can claim to be 'true' in a non-relative sense; at best it can be 'true for us'.

One interesting issue in assessing this view is what status Wittgenstein intended for the sketched alternative ways of responding to language teaching. Certainly they need enough feasibility to dislodge the idea of the one and only enforced way of dividing up the world. On the other hand it is not clear that he takes us to be entitled to assert that there are conceptual schemes which are both incompatible with ours and also fully possible.

Many other readings are also possible, detecting in his writings elements of pragmatism, behaviourism and even deconstructionism (see Behaviourism, analytic; Deconstruction; Pragmatism §2; Realism and antirealism §4; Relativism).

13 Philosophy of mind

The *Tractatus* picture of the relation of language to its subject matter is especially attractive in the case of psychological notions. A sensation such as pain is easily conceived as a phenomenon which impresses its nature and identity conditions on one who has it, independent of external circumstances or bodily behaviour. The private language argument (§§243-71) examines this idea in the light of the earlier discussion of meaning. One aim is to show that our actual use of terms for sensations does not and could not conform to the pattern suggested.

The rule-following considerations suggest that no standard for what is to count as 'the same' can be fixed merely by uttering a word to oneself while being vividly aware of what one experiences. For one kind of item rather than another to come into focus out of the indefinite variety potentially presented in an experience, that experience must be embedded in one kind of life rather than another. Relatedly, for a word to have meaning there must be some extended practice in which its use has a point. This is as true of sensation words as of any others. We teach and use them in a complex setting of physical circumstances and expressive bodily behaviour. This setting, says Wittgenstein, is not externally and contingently linked to sensation but is an integral part of the sort of life in which the general category 'sensation' makes sense and in which particular sensations can be individuated.

Wittgenstein considers many other topics in philosophical psychology, among them intention, expectation, calculating in the head, belief, dreaming and aspect perception. A constant theme is the need to counter the attraction of the model of name and object, which (together with such things as the special authority which each person has to pronounce on their own psychological states) leads us to conceive of the 'inner' as a special mysterious realm, distinct from the 'outer' or physical. He offers such general remarks as 'An ''inner process'' stands in need of outward criteria' (1953: §580). He also returns repeatedly to the idea that authoritative first-person psychological claims should be seen as expressions or avowals of those states which we are inclined to insist that they describe. These sorts of moves have led to the idea that he denies the existence of the 'inner' and is really a behaviourist.

He was aware of the risk of this reading:

> 'But you will surely admit that there is a difference between pain-behaviour accompanied by pain and pain-behaviour without any pain?' - Admit it? What greater difference could there be? - 'And yet you again and again reach the conclusion that the sensation itself is a *nothing*.' Not at all. It is not a *something*, but not a *nothing* either!
>
> (1953: §304)

Thoughts and experiences are, on his view, necessarily linked to expressive behaviour. 'Only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious' (1953: §281). But this does not mean that any reduction of the mental

to the behavioural is possible or that the psychological is not real. To see Wittgenstein's view sympathetically it is important to keep in mind the upshot of §§1-242. There is no a priori guarantee of some privileged set of classifications (for instance, those of natural science) in terms of which all others must be explained. To understand any phenomenon we must get a clear view of the language games in which terms for it are used; and the logical shapes of these may be very different from those which are initially suggested by the pictures which grip us (see Private language argument §§1-3).

14 Philosophy of mathematics

Platonism in mathematics involves two claims, that there is a realm of necessary facts independent of human thought and that these facts may outrun our ability to get access to them by proofs. Platonism is attractive because it accounts for several striking features of mathematical experience: first that proofs are compelling and yet may have conclusions which are surprising, and second that we seem to be able to understand some mathematical propositions without having any guarantee that proofs of them exist.

Wittgenstein never accepted Platonism because he always took the view that making substantive statements is one thing, while articulating the rules for making them is another. So-called necessary truths clearly do present rules of language, inasmuch as accepting them commits one to allowing and disallowing certain linguistic moves. Wittgenstein holds that it is therefore a muddle to think that such formulations describe some particularly hard and immovable states of affairs. Thus in the *Tractatus* mathematical propositions are treated together with tautologies as sets of signs which say nothing, but show the logic of the world.

Nevertheless the *Tractatus* view has some kind of affinity with at least the first claim in Platonism, inasmuch as the rules of our language are taken to reflect some logical structure independent of us. But when Wittgenstein comes to see linguistic rules as features internal to our practices, the resulting picture is unwelcoming even to this. There is now no such thing as 'the logic of the world', whether to be shown or said. Instead, in *Remarks on the Foundations of Mathematics*, he explores ideas of the following kinds.

At a given time we have linguistic practices directed by certain rules. Someone may now produce a proof of a formula which if accepted would be a new rule - for example, '$14 + 3 = 17$'. It is natural to think that to accept this is to unpack what we were already committed to by our understanding of '17', '+', and so on. But the rule-following considerations unsettle this assumption because they undermine the idea of an intellectual confrontation with an abstract item which forces awareness of its nature upon us and they also bring to our attention the element of spontaneity in any new application of a given term. Rather to accept the proof and its outcome is to change our practices of applying signs like '17', because it is to adopt a new criterion for judging that seventeen things are present, namely that there are two groups of fourteen and three. Hence to accept the proof is to alter our concepts. What makes mathematics possible is that we nearly all agree in our reaction to proofs, and in finding them compelling. But to seek to explain this by pointing to Platonic structures is to fall back into incoherent mythology.

My own view is that it is persistent uneasiness with the first claim in Platonism which primarily motivates Wittgenstein's reflections on mathematics. But those who see him as an antirealist will put more stress on hostility to the second claim (the idea of verification transcendence) and certainly some of Wittgenstein's remarks (for example, his suspicion of the application of the law of excluded middle to mathematical propositions) have affinities with ideas in intuitionistic logic. A third reading will bring out the conventionalist-sounding elements, on which we choose what linguistic rules to adopt on pragmatic grounds.

In addition to reflections on the nature and use of elementary arithmetical claims, Wittgenstein also applies his ideas to some more complex constructs in mathematical logic, such as the Frege-Russell project of deriving mathematics from logic, Cantor's diagonal argument to the non-denumerability of the real numbers, consistency proofs and Gödel's theorem. His general line here is not that there is anything wrong with the mathematics but that the results have been misconstrued, because they have been interpreted against a mistaken background Platonism. Some mathematical logicians claim that Wittgenstein has not understood properly what he is discussing. His views on consistency and Gödel in particular have aroused annoyance (see Antirealism in the philosophy of mathematics §2; Intuitionism; Realism in the philosophy of mathematics §2).

15 Ethics, aesthetics and philosophy of religion

In the *Tractatus* Wittgenstein consigns ethics to the realm of the unsayable, and he takes the same line in his ‘Lecture on Ethics’ ([1929] 1993). Here he says that ethics (which he links to aesthetics and religion) arises from a tendency in the human mind to try to express in words something - roughly the existence and nature of absolute value - which seems to manifest itself to us in certain experiences. (He gives as an instance the experience of finding the existence of the world miraculous.) It is essential to this impulse that it seeks to go beyond the world and significant language; so it is bound to issue in utterances which are nonsensical. Nevertheless, he says, he has the greatest respect for this impulse and would not for his life ridicule it.

This position resembles the emotivism associated with logical positivism in distinguishing ethical utterances sharply from those of science (that is, those which are capable of rational assessment, and can be true or false). But it also differs from it in being, in spirit, an ethical realism, albeit of a mystical kind.

In his later writing he rethought his views on meaning, mathematics and the mind but did not return to any sustained discussion of ethics or aesthetics (although there are scattered remarks, particularly on the aesthetics of music, in *Culture and Value* (1980)). One interpretation of the later outlook, however, provides a hospitable setting for an ethical realism of a less mysterious kind, one which allows for the statement and rational discussion of truth-evaluable ethical claims. Philosophers of meta-ethics taking themselves to be working within a Wittgensteinian outlook have urged that our inclination to insist on a dichotomy between fact and value, or between cognition and feeling, should be resisted, as the outcome of the grip on us of some misapplied picture. Moreover Wittgenstein’s emphasis on attention to the actual workings of language could encourage a distinctive approach to first order ethical questions (see Wittgensteinian ethics §2). But he himself never developed this, nor does he engage with issues in political philosophy.

The later outlook enjoins us to study each distinctive area of language as far as possible without preconceptions. If we do this for religious language, Wittgenstein holds, we shall see that religion is not a kind of science and hence is not open to criticism on the grounds that, as science, it is unconvincing (see, for example, ‘Remarks on Frazer’s *Golden Bough*’ ([1931] 1993)). Some take it that this implies that no religious utterance can be properly subject to any criticism other than that coming from inside the same religious community or tradition.

16 Epistemology

One familiar traditional philosophical problem is that of scepticism, that is, whether we can rightly claim to know such things as that physical objects exist independent of our perception, that the world was not created five minutes ago and so forth.

Wittgenstein’s most extended discussion of these issues is in *On Certainty* (1969). He starts from the kinds of examples invoked by G.E. Moore in his attempt to combat scepticism, such as ‘Here is a hand’ and ‘The Earth has existed for a long time before my birth’ (see Moore, G.E. §§3-4). Moore is wrong, Wittgenstein thinks, in taking it that we are plainly entitled to assert that we know these things. But Moore is right in thinking that the claims form an interesting class. It is impossible to conduct life and thought without taking some things entirely for granted, and the propositions Moore identifies are the articulated forms of things which play this role for us. They help to define our world picture and underpin the procedures by which other claims (ones that are in fact doubted and tested) can be assessed. But they cannot themselves be assessed because there is nothing relatively more certain by which we can get leverage on them. Someone who seems to doubt them is thought mad and, from a first-person point of view, when I imagine doubting such things I contemplate a situation in which I would no longer know how to reason about anything. There are close links between these themes and the idea that the workability of any language game presupposes certain very general facts of nature.

The relevance of this for the traditional question of scepticism is that it is, in its form, misconceived. The central use of ‘know’ is in connection with propositions where testing is possible. Hence one who uses it in connection with the propositions which help define our worldview (as is in fact done only in philosophy and not in ordinary life) has extended the word to a situation where procedures do not exist for assessing either the first-order claim or the claim to knowledge of it. This is not to say that the word ‘know’ is unintelligibly and wrongly used in the philosophical debate. We can sympathize with the sceptical impulse, which springs from awareness of the fact that our language games are not based on grounds which compel us to them or guarantee their continued success. But

we can also sympathize with the anti-sceptical position which insists that acceptance of these central propositions underpins our being able to do any thinking at all, so that claims to doubt them are empty (see Scepticism).

17 Wittgenstein's conception of philosophy

In two central respects Wittgenstein stands squarely within the main historical tradition of philosophy, first in the nature of the issues which excited and intrigued him intellectually - meaning, the self, consciousness, necessity - and second (going back to the roots of the tradition) in his being a 'lover of wisdom', that is, one who is seriously concerned about having a right stance to the world both intellectually and practically and who is committed to the use of the intellect (among other things) in helping to achieve this.

But he differs from many philosophers in his conviction that a great number of traditional philosophical problems are the result of some deep kind of muddle, and in his belief that the answers given and the way they are debated hinder rather than help us in achieving wisdom. This conviction gripped him from very early on and philosophical thought therefore presented itself to him as a tormentingly difficult struggle to be honest and to free himself from misleading preconceptions.

So the word 'philosophy' has, in all his writings, two uses. On one it describes a body of confused utterances and arguments, arising largely from misunderstanding of the workings of language, and on the other it describes an activity of helping people to get free of the muddles. Another important continuity is his insistence that there cannot be philosophical theories and that the helpful activity of philosophy ought only to consist of making uncontentious statements, of describing and assembling reminders. In the context of the picture theory of meaning, this is comprehensible (see §7). But it is less clear that it is required by the later view.

In part Wittgenstein is here stressing that we cannot have the kind of explanation of our concepts which the *Tractatus* picture seemed to promise. Our form of life cannot be grounded but only described and lived. In part he is questioning the impulse to look for quasi-scientific theories of the nature of philosophically puzzling phenomena. But these two interrelated points do not obviously add up to a complete embargo on anything which could be called 'philosophical theory'. It is in the spirit of the later philosophy to point out that there are many different kinds of things which can be called 'theories'. Everyone engaged in reflection on the topics Wittgenstein considers (including Wittgenstein himself) finds it natural to articulate in words the states they arrive at and to engage with these words and those of others in the mode of further comment and assessment.

We become aware here, and at many other places, of the open-ended and unfinished nature of Wittgenstein's reflections. His writings have aroused great devotion because of the honesty and depth which many find in them. But it is important not to treat them with superstitious reverence. Rather they should be read in the spirit in which he intended, namely as an invitation to explore with as much integrity as possible one's own perplexities and what would resolve them.

See also: Criteria

JANE HEAL

List of works

Wittgenstein, L.J.J. (1922) *Tractatus Logico-Philosophicus*, trans. C.K. Ogden and F.P. Ramsey, London: Routledge; trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1961.(The major work of Wittgenstein's early period and the only book published during his lifetime. The first English translation was revised and approved by Wittgenstein himself, though the later version is now standard. The German version was published in 1921 in *Annalen der Naturphilosophie*.)

Wittgenstein, L.J.J. (1953) *Philosophical Investigations*, ed. G.E.M. Anscombe and R. Rhees, trans. G.E.M. Anscombe, Oxford: Blackwell.(The most polished and worked over of all Wittgenstein's later work; it contains the presentation of his ideas on meaning and philosophical psychology with which he was most nearly satisfied.)

Wittgenstein, L.J.J. (1956) *Remarks on the Foundations of Mathematics*, ed. G.H. von Wright, R. Rhees and G.E.M. Anscombe, trans. G.E.M. Anscombe, Oxford: Blackwell, 3rd edn, 1978.(Selections from Wittgenstein's notebooks and manuscripts from 1937 to 1944. The third edition contains a larger selection of material than the first edition.)

Wittgenstein, L.J.J. (1958) *The Blue and Brown Books*, Oxford: Blackwell.(These were dictated to his pupils in

1933-5 and are among the few works composed by Wittgenstein in English. A good way of approaching his later thought.)

Wittgenstein, L.J.J. (1961) *Notebooks 1914-16*, ed. G.H. von Wright and G.E.M. Anscombe, trans. G.E.M. Anscombe, Oxford: Blackwell.(Helpful for understanding the *Tractatus*, since it contains what is left of the preliminary writings of that period.)

Wittgenstein, L.J.J. (1967) *Zettel*, ed. G.E.M. Anscombe and G.H. von Wright, trans. G.E.M. Anscombe, Oxford: Blackwell.(A selection, made by Wittgenstein himself, of remarks that he wrote mainly between 1945 and 1948.)

Wittgenstein, L.J.J. (1969) *On Certainty*, ed. G.E.M. Anscombe and G.H. von Wright, trans. D. Paul and G.E.M. Anscombe, Oxford: Blackwell.(A collection of all the material on knowledge and certainty from the last year and a half of Wittgenstein's life, where he treated the topics at length at several points in his notebooks.)

Wittgenstein, L.J.J. (1974) *Philosophical Grammar*, ed. R. Rhees, trans. A. Kenny, Oxford: Blackwell.(Written in 1931-4, this deals extensively with logic and mathematics, as well as topics such as language and meaning.)

Wittgenstein, L.J.J. (1975) *Philosophical Remarks*, ed. R. Rhees, trans. R. Hargreaves and R. White, Oxford: Blackwell.(Written in 1929-30 and interesting in that it shows the kinds of reflection on the *Tractatus* which drove Wittgenstein from his earlier to his later outlook.)

Wittgenstein, L.J.J. (1977a) *Remarks on Colour*, ed. G.E.M. Anscombe, trans. L.L. McAlister and M. Schättle, Oxford: Blackwell.(All the material on this topic from the writings of 1950-1, in which Wittgenstein worked on it extensively.)

Wittgenstein, L.J.J. (1977b) *Vermischte Bemerkungen*, Suhrkamp Verlag: Frankfurt am Main; trans. P. Winch, ed. G.H. von Wright and H. Nyman, *Culture and Value*, Oxford: Blackwell, 1980.(Wittgenstein's notebooks and typescripts often contain remarks on topics which are not directly philosophical. This collection assembles all of them from 1914 to 1951.)

Wittgenstein, L.J.J. (1980) *Remarks on the Philosophy of Psychology*, vols 1 and 2, ed. G.E.M. Anscombe and G.H. von Wright, trans. G.E.M. Anscombe, Oxford: Blackwell. (Selections made by Wittgenstein in 1947 and 1948 from material written in 1946-8. Contains much of interest on the topics treated in Part II of the *Philosophical Investigations*.)

Wittgenstein, L.J.J. (1982; 1992) *Last Writings on the Philosophy of Psychology*, ed. G.E.M. Anscombe, G.H. von Wright and H. Nyman, trans. C.G. Luckhardt and M.A.E. Aue, Oxford: Blackwell, 2 vols.(Writings from 1948-9, from which selections were made by Wittgenstein for Part II of the *Philosophical Investigations*.)

Wittgenstein, L.J.J. (1993) *Philosophical Occasions*, ed. J.C. Klagge and A. Nordmann, Indianapolis, IN: Hackett.(Usefully anthologizes several short pieces, including the 'Lecture on Ethics' (1929) and 'Remarks on Frazer's *Golden Bough*' (1931).)

References and further reading

Anscombe, G.E.M. (1959) *An Introduction to Wittgenstein's Tractatus*, London: Hutchinson.(A stimulating, although in parts quite difficult, discussion of many of the central ideas of the *Tractatus*.)

Baker, G.P. and Hacker, P.M.S. (1980, 1988, 1990) *An Analytical Commentary on the Philosophical Investigations*, Oxford: Blackwell, 3 vols.(An immensely detailed commentary, informed by extensive knowledge of the Wittgenstein papers. The third volume is by P.M.S. Hacker alone.)

Budd, M. (1989) *Wittgenstein's Philosophy of Psychology*, London: Routledge.(A clear survey of Wittgenstein's views on a variety of psychological topics.)

Canfield, J.V. (ed.) (1986-8) *The Philosophy of Wittgenstein*, New York and London: Garland Publishing Company, 15 vols.(A useful collection of articles on all aspects of Wittgenstein's philosophy.)

Carruthers, P. (1990) *The Metaphysics of the Tractatus*, Cambridge: Cambridge University Press.(Contains extensive consideration of determinacy of sense and the related topic of simple objects.)

Cavell, S. (1979) *The Claim of Reason*, Oxford: Oxford University Press.(Subtle and stimulating book on Wittgenstein's later philosophy. The elaborate style takes some getting used to.)

Fogelin, R.J. (1987) *Wittgenstein*, London: Routledge, 2nd edn.(A good general book, particularly helpful as an introduction to Wittgenstein's treatment of logic and mathematics. Also contains a helpful bibliography.)

Hacker, P.M.S. (1986) *Insight and Illusion*, Oxford: Oxford University Press, 2nd edn.(Another good general book, particularly strong on Wittgenstein's treatment of the self.)

Heal, J. (1989) *Fact and Meaning*, Oxford: Blackwell.(Further discussion of the themes of §§9-12.)

Johnston, P. (1993) *Wittgenstein: Rethinking the Inner*, London: Routledge.(A reading of Wittgenstein's later views on mind.)

Kenny, A. (1973) *Wittgenstein*, London: Allen Lane.(The best general introduction.)

Kripke, S. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell. (An exceptionally clear and gripping exploration of the idea that the rule-following considerations lead to scepticism about meaning.)
McDowell, J. (1984) 'Wittgenstein on Following a Rule', *Synthèse* 58: 325-63.(In part a criticism of Kripke (1982), in part an interesting exposition of Wittgenstein.)
McGinn, M. (1989) *Sense and Certainty: A Dissolution of Scepticism*, Oxford: Blackwell.(A clear account of Wittgenstein's views on knowledge and certainty, expanding on the points in §16 above.)
McGuinness, B. (1988) *Wittgenstein: A Life, Young Ludwig 1889-1921*, London: Duckworth.(Biographical study, containing much on Wittgenstein's intellectual background which is not in Monk's book and culminating on a chapter on the *Tractatus*.)
Monk, R. (1990) *Ludwig Wittgenstein*, London: Jonathan Cape.(A full and illuminating biography.)
Pears, D. (1987) *The False Prison*, Oxford: Oxford University Press.(A two-volume study containing much interesting discussion of the transition period and the private language argument.)
Wright, C. (1980) *Wittgenstein on the Foundations of Mathematics*, London: Duckworth.(Extended exploration of the idea that Wittgenstein should be seen as an antirealist or as a conventionalist.)

Wittgensteinian ethics

Although the strict 'fact-value distinction' of Wittgenstein's early period has shaped much subsequent work on ethics, his most profound influence on the subject stems from the later Philosophical Investigations and associated writings. Of particular significance have been, first, the concept of a 'language game', and second, the discussion of following a rule. The vision of morality itself as a language game - a complex of speech and action ordered in a way that makes sense to the participants - has seemed to diminish the urgency of traditional questions about the 'foundations' of ethics, and has promoted acceptance of moral experience and consciousness as natural (human) phenomena. More recently there has been a growing interest in how Wittgenstein's general reflections on rule-governed practices might apply to the specific case of moral understanding.

1 The early Wittgenstein

Wittgenstein once told a friend that, although he was not a religious man, he could not help seeing every problem from a religious point of view (Rhees (ed.) 1984: 79). Given the austerity of much of his philosophical subject matter, the statement is a curious one. However, it is a useful reminder that throughout his career he was in search not just of intellectual insight but of a correct attitude to life. This is already apparent in the first phase of his work.

The main ethical theme of Wittgenstein's *Tractatus Logico-Philosophicus* (1921), and the subject of his 'Lecture on Ethics' (1929), is the inexpressibility of (absolute) value. He bases this idea on a distinction between the world as a whole, or the 'totality of facts', and the particular facts that obtain within the world. Whereas the latter can be represented linguistically, anything we may say about the former will be 'nonsense' (even if it is the special kind of nonsense exemplified by all the propositions of the *Tractatus*). But since any given fact might have been otherwise, and is thus a mere accident, Wittgenstein holds that the entire domain of fact - and hence of what can be put into words at all - is, from an absolute standpoint, devoid of value. Ethics therefore - along with aesthetics - has to do not with particular facts but with the totality, or rather with the way in which the totality presents itself to us; and this in turn depends on the quality of our will - not the empirical 'will' recognized by psychology, but one whose good or bad exercise makes the world 'wax and wane as a whole'.

Wittgenstein mentions in the 'Lecture' some states of mind that seem to possess this transfigurative power: 'wondering at the existence of the world', 'feeling absolutely safe' and 'feeling guilty'. However, he stresses that one can speak only for oneself about such experiences and that language can in any case do no more than gesture towards what is important in them. We rely here on (apparent) allegory or simile of a kind that is essentially unconvertible into literal factual statement.

These views reflect Wittgenstein's desire to safeguard the private, spiritual nature of ethics and not to let it be demeaned by philosophical 'chatter'. Even at a more practical level, though, he sees little scope for discursive argument about moral questions. An individual may or may not 'have an ethics'; if so, then they will have a basis (be it Christian, Nietzschean, or whatever) for their life choices; but the question whether one *ought* to be a Christian or a Nietzschean is senseless, since there is no suitably authoritative standard by which to settle it.

2 Describing language games

One offshoot of this conception of value was the 'noncognitivist' theory of ethics, which classifies the functions of language under the two main headings of description and expression (see Analytic ethics). Wittgenstein himself, by contrast, moved during the 1930s towards an appreciation of the indefinite variety of 'language games' or types of rule-governed linguistic activity. In *Philosophische Untersuchungen (Philosophical Investigations)* (1953), our main source for his later thought, he sees language-use as part of our 'natural history' - one of those elements in human life that are given rather than chosen, and so 'have to be accepted'. This does not mean that 'good', in the ethically interesting sense, can after all be analysed ('naturalistically') in terms of a value-free notion of human welfare (see Naturalism in ethics). What it means is that the proper aim of moral philosophy (like that of philosophy in general) is not justification but a certain kind of clarity, and that its proper method is one of reflection on what we might do or say in various concretely imagined contexts - not of trying to formulate universal normative principles that would validate our behaviour.

This approach was seen by its exponents in the 1950s and 60s partly as a check on the pretensions of ‘theory’, but it also yielded illuminating discussions of familiar moral practices, which could now be accorded the same respect as the ‘alien’ anthropological phenomena conjured up in some of Wittgenstein’s thought experiments. A practice, it was argued, generates its own characteristic patterns of reasoning, and theorists are wrong to suppose that the behaviour internal to it admits of any better explanation than can be gathered from the participants themselves. Thus if someone (presumably a child) does not understand punishment, what we actually provide is help in seeing how the concept of punishment relates to others within the same region of social experience, such as wrongdoing, responsibility, forgiveness, making amends. And this is all the ‘explanation’ needed.

The view that philosophy should limit itself to ‘describing language games’ without passing judgment on them may seem to lead in the direction of relativism, a tendency already discernible at the end of §1 (see Moral relativism). It should be remembered, though, that Wittgenstein officially holds ‘ordinary’ (as opposed to philosophical) language to be in perfect order as it is, so that if our (ordinary) ethical language game happens to incorporate activities of critical observation and reflection, there would seem to be no (philosophical) reason why we should renounce these. Such activities might well go against Wittgenstein’s own anti-modern, even primitivist taste; but in keeping with the nonjudgmental attitude, he states no principle by reference to which they could be condemned. (Relatedly, one might ask about the ‘we’ that figures in this section: who exactly counts as being located outside ‘our’ moral practices, and so beyond the reach of ‘our’ judgment? (see Moral agents; Universalism in ethics §1).)

3 Rule-following and ethics

During the 1970s and 80s Wittgenstein studies took a new turn, the effect of which was felt in ethics. Partly in response to Saul Kripke’s reappraisal of the ‘rule-following considerations’, attention began to focus less on the specificity of particular language games and more on the idea of a rule-governed practice as such - the genus that includes all ‘games’ involving conventional signs (see Meaning and rule-following §3). Above all, how should we understand Wittgenstein’s account of the way a rule or order can ‘reach into the future’ and determine in advance what will and will not comply with it?

Because of the prominence it gives to the notion of criteria and to communal standards of correctness, Wittgenstein’s later philosophy has often been regarded as ‘anti-realist’. However, the rule-following considerations can be interpreted in a way that challenges this view and gives more weight to Wittgenstein’s own metaphysical ‘quietism’ - his disavowal of any substantive philosophical position. This different interpretation is unfavourable, in particular, to theories that refuse to take at face value the ostensibly factual character of certain regions of discourse. For example, it is opposed to the neo-Humean view that evaluative judgment cannot be genuinely factual because it involves a projection of our subjective attitudes onto the world (see Projectivism).

In the spirit of Wittgenstein’s remark that ‘words are also deeds’ (1953: part I, 546), and of the link he sets up in *Philosophische Untersuchungen* between linguistic understanding and socialization, the ‘quietist’ reading gives a central place to the ethical in reconstructing his overall account of rules. Wittgenstein notes that interpretation, explanation or justification necessarily ‘comes to an end’, giving way - if successful - to a moment when we see the point (or grasp the message) without the mediation of any further signs. The possibility of such immediate, tacit understanding is ultimately due to our common nature as members of an animal species, but the human species is distinguished from others by the fact that it exploits this natural ‘likemindedness’ in order to transmit to successive generations the various culturally distinct forms of likemindedness realized in particular languages. The work of socialization carried on within human cultures maintains in existence an array of shared sensibilities - the ‘feel’ speakers have for what a rule-governed practice requires in this or that particular case - and it is these sensibilities that support all our resources of normative judgment, whether about overtly ethical ‘right’ and ‘wrong’ or about right and wrong ways of describing things, right and wrong conclusions to draw from given premises, and so on.

SABINA LOVIBOND

References and further reading

Johnston, P. (1989) *Wittgenstein and Moral Philosophy*, London: Routledge.(Monograph providing useful orientation and guide to sources.)

Lovibond, S. (1983) *Realism and Imagination in Ethics*, Oxford: Blackwell.(Examines the relation of the later Wittgenstein to a possible ‘moral realism’ emerging from truth-conditional semantics.)
McDowell, J. (1979) ‘Virtue and Reason’, *Monist* 62: 331-50.(Connects Aristotle’s ethics with Wittgenstein’s rule-following considerations.)
McDowell, J. (1981) ‘Non-cognitivism and rule-following’, in S. Holtzman and C. Leich (eds) *Wittgenstein: To Follow a Rule*, London: Routledge & Kegan Paul.(A Wittgensteinian critique of noncognitivist theories of ethics.)
McDowell, J. (1992) ‘Meaning and Intentionality in Wittgenstein’s Later Philosophy’, *Midwest Studies in Philosophy* 17: 40-52.(Defends a reading of Wittgenstein as ‘quietist’ - one who disavows any substantive philosophical position.)
Phillips, D.Z. and Mounce, H.O. (1969) *Moral Practices*, London: Routledge & Kegan Paul.(Wittgensteinian argument for the ineliminable diversity of moral life.)
Pitkin, H.F. (1972) *Wittgenstein and Justice*,, Berkeley, CA: University of California Press.(Extended discussion of the significance of the later Wittgenstein for political theory.)
Rhees, R. (ed.) (1984) *Recollections of Wittgenstein*, Oxford: Oxford University Press.(For insight into Wittgenstein’s personality and moral concerns.)
Wittgenstein, L. (1921) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London: Routledge & Kegan Paul, 1961.(The only philosophical work by Wittgenstein published during his lifetime, and the source for his ‘early’ view of logic and language; sections 6.4 onwards relate to ethics.)
Wittgenstein, L. (1929) ‘Lecture on Ethics’, first published posthumously in 1965 as ‘Wittgenstein’s Lecture on Ethics’, in *Philosophical Review* 74: 3-26.(Short statement influenced by the general philosophical position of the *Tractatus*; with valuable background material by F. Waismann and R. Rhees.)
Wittgenstein, L. (1953) *Philosophische Untersuchungen*, trans. G.E.M. Anscombe, ed. G.E.M. Anscombe and R. Rhees, *Philosophical Investigations*, Oxford: Blackwell; 3rd edn, 1967.(Little explicitly on ethics; should be studied as a whole.)
Wittgenstein, L. (1961) *Notebooks 1914-1916*, ed. G.E.M. Anscombe and G.H. von Wright, trans. G.E.M. Anscombe, Oxford: Blackwell; 2nd edn, 1979.(Contains interesting pre-Tractarian material on ethics and the will.)
Wittgenstein, L. (1980) *Vermischte Bermerkungen*, trans. P. Winch, ed. G.H. von Wright and H. Nyman, *Culture and Value*, Oxford: Blackwell.(Collected remarks, mainly from the 1930s and 40s.)

Wodeham, Adam (*c*.1298-1358)

An English Franciscan theologian, Wodeham was preoccupied with logical and semantic questions. He lectured for about a decade on Peter Lombard's Sentences, *first at London, then at Norwich and finally at Oxford. His lectures emphasized the dependence of the created world on God and the contingency of nature and salvation.*

John Duns Scotus and William of Ockham exerted the most important influences on Wodeham. He regarded Scotus as a vigorous thinker and respected him enough to accept his opinion in case of doubt. Proud to have learned logic from Ockham, Wodeham devoted considerable time to defending Ockham's views from Walter Chatton, whom he saw as someone whose errors in logic arose from malice as well as ignorance. However, despite Wodeham's reservations about Chatton, he was considerably influenced by him. Similarly, Wodeham modified his own opinion about sensory illusions in response to Peter Aureol, whom he saw as skilled and prudent but often mistaken, sometimes as a result of faulty logic.

1 Psychology

Appealing to experience, Wodeham denied the distinction between the sensory and intellective souls; a single soul suffices to explain all the cognitive acts we perform. On this point, Wodeham opposed both Duns Scotus (who proposed formally distinct sensory and intellective souls) and Ockham (who proposed souls that were really distinct). Against Ockham's opinion that sensory and intellective souls must be distinct since contraries could not coexist in the same subject, Wodeham argued that sensory inclination and intellectual appetite in regard to the same external object were only virtually, not formally, contraries. The same soul apprehends sensible particulars and universals. These acts are sensations when they are partially caused by external objects; intellections, when they abstract from singulars (*Lectura secunda* I, 9-33).

A similar predilection for reductionism reveals itself in Wodeham's discussion of enjoyment. He claims that all appetitive acts are cognitive acts, since we cannot will or desire an object without apprehending it; but cognition does not necessitate volition. Like Ockham, Wodeham holds that clear knowledge of God without enjoyment is possible, at least at first. Conversely, loving God necessarily includes the implied judgment that God is lovable. This leads Wodeham to conclude that acts of volition can be described as true or false. Rejoicing about being a Franciscan, for example, includes the correct apprehension of a practical truth as well as the act of enjoyment (*Lectura secunda* I, 253-85).

2 Epistemology

Scepticism is a problem for Wodeham, who sees Ockham's response to it as inadequate. For Ockham, intuitive cognition, the basis of all other cognition, is reliable by definition. By contrast, Wodeham holds that whether or not the object of intuition exists, intuition will always incline us to believe that its object exists. He was among the first to recognize that sceptical consequences could be drawn from the long list of sensory illusions adduced by Peter Aureol against Scotus' account of intuitive cognition. In response, Wodeham defined three degrees of certainty. The greatest degree, which compels the intellect, is not available regarding contingent propositions, since the intellect is aware of the possibility of deception. The least degree of certainty is compatible with error. It is in this degree that I may be certain of a mistaken proposition - as for example, when I judge that a straight stick half submerged in water is bent.

Despite his preoccupation with possible natural and supernatural obstructions in the perceptual process and the concessions he made to them, Wodeham remained a reliabilist, maintaining that cognition is reliably though not infallibly caused by its object. His basic reply to the sensory illusions adduced by Aureol was that although illusions will continue to incline us to make false judgments, reason and experience enable us to correct our judgments and to avoid being systematically misled by illusions (*Lectura secunda* I, 65-111, 163-79) (see Epistemology, history of).

3 Semantics

In semantics, Wodeham's most important contribution was the concept of the *complexe significabile* (complexly signifiable) (see Language, medieval theories of). Wodeham developed his views as a compromise between

Chatton and Ockham on the question of the object of scientific knowledge. What is the object of our assent when we assert that man is a rational animal? Is it an external object in the real world (Chatton's *res*) or a mental object (Ockham's *complexum*)? Are we assenting to man considered as a rational animal, or to the complex mental entity that-man-is-a-rational-animal? Wodeham argues against both solutions. If Chatton is right, then there is no difference between a Catholic who asserts that Christ was incarnate and a heretic who asserts that Christ will be incarnate. Both assent to Christ as incarnate; the difference between their creeds can only be expressed by the tense of the verbs which connect Christ and incarnation. If Ockham is right, then sciences of real things deal only with mental objects and not with the external world; the objects of the 'real' sciences do not differ from the objects of logic and semantics. Such considerations lead Wodeham to postulate a third way: the true object of assent is what is complexly signifiable by a phrase, such as that-man is-a-rational-animal, the *complexe significabile* (Nuchelmans 1980).

The *complexe significabile* itself is neither something in the external world nor a mental object. It is neither a substance nor an accident and thus is not included in any Aristotelian category. It is not something, but neither is it nothing. Rather, the question 'What is it?' is ill-formed; it makes no more sense than the question 'Is a people a man or a non-man?' When we assent to a *complexe significabile*, we are not assenting to some thing; instead we are affirming that something is the case (*Lectura secunda* I, 180-208).

4 Logic

A terminist logician and a conceptualist, Wodeham employs three different kinds of supposition: personal, simple and material. Personal supposition is reference to things in the external world, simple supposition to concepts, and material supposition to written or spoken words (see Language, medieval theories of; Logic, medieval).

Wodeham's analysis of the distinction between abstract and concrete predication was indebted to but different from that of Ockham. In defining concrete predication, Wodeham aims to avoid generating nonsense expressions. Thus for Wodeham the nominal definition of the term *albus* is 'having whiteness', not 'body having whiteness'. Eliminating the bearer from the definition of concrete terms prevents the generation of the nonsense-sentence 'Socrates is a body having whiteness body', which would otherwise result from replacing 'white' with its definition in the sentence 'Socrates is a white body' (*Lectura secunda* II, 244).

Wodeham's denial that the bearer is signified *per se* in concrete predication is reminiscent of Anselm's distinction between *per se* and *per aliud* predication (see Anselm of Canterbury). For Wodeham as for Anselm, reference is signification only in a secondary sense. What we think of when we hear a term is not necessarily the objects to which it refers (its *supposita* or *appellata*). There is no difference between the *per se* signification of abstract and concrete terms; and in the case of substance terms, supposition and signification coincide. Thus both 'man' and 'humanity' signify and supposit for a substance composed of body and soul. The sentence 'A man is a humanity' is false only in the case of Jesus, who has a divine as well as a human nature and hence cannot be identified with his humanity (Spade 1988).

Wodeham relies on his theory of *per se* predication in dealing with problems of Trinitarian theology. The chief problem was the apparent conflict between the Aristotelian principle of non-contradiction and the truths of the faith. Thus Aristotelian logic seems to require that if 'the Father is not the Son' and 'the Father is the deity', then 'the deity is not the Son'. Confronting this problem, Wodeham rejects *ad hoc* theological modifications of logic; such approaches, he says, deserve the ridicule of heathens. Wodeham also criticizes the solutions of Ockham and Scotus. His own solution is to treat identity as a special case of predication; identity is not necessarily a symmetrical relationship (see Identity). Following Scotus, Wodeham distinguishes between identifying or formal predication and denominative predication. Unlike denominative predication, where the subject and predicate have the same supposition, in formal predication the predicate has broader supposition than the subject. Thus deity can be formally predicated of the Father, but the converse is not true (*Lectura Oxoniensis* I d.33; Gelber 1974).

5 Metaphysics

Wodeham was a conceptualist who identified universals with mental acts, and hence maintained that only individuals exist (*Lectura secunda* I, 21). In this respect, as in his emphasis on ontological parsimony, Wodeham followed Ockham (see Nominalism; Universals).

Considering proofs for God's existence, Wodeham analyses fourteenth-century Franciscan theories of causality. He defends Ockham's challenge to Scotus' inference: 'since the universe of essentially ordered effects is caused, the universe must be caused'. Exploiting his understanding of the logic of infinity, Wodeham argues that Chatton's defence of Scotus' theory of essentially ordered causes and effects is absurd. Chatton errs by inferring categorematic conclusions from premises which are true only if understood syncategorematically (*Lectura secunda* II, 117-21; Adams 1993).

6 Natural philosophy

In opposing the atomism or indivisibilism of his day, Wodeham repeatedly applied his understanding of the logic of infinity, continuity and infinitesimal change (see Natural philosophy, medieval). Like Ockham he believed that limits of place, time and motion - such as points, lines, surfaces, temporal instants and mutations (instants of change) - had no independent ontological status. Indeed, Wodeham says he anticipated Ockham's arguments for non-entitism, the claim that terms such as 'point' are non-referring (*Tractatus de indivisibilibus* q.2 a.1 n.20).

Wodeham presents many arguments against indivisibilism. Some depend on the distinction between the categorematic (perfectively actualized) and syncategorematic (processively actualized) meaning of the phrase 'infinitely divide'. Others, based on an analysis of the compounded and divided senses, claim that strictly speaking not the continuum, but its parts are divided. Most instructive in his view were arguments relying on the impossibility of indivisibles in contact with one another.

Wodeham accepted Ockham's claim that the infinity of parts in a continuum exists not just potentially but actually. Acceptance of this claim led Wodeham to argue for the possibility of unequal infinities (*Tractatus de indivisibilibus* q.5).

7 Influence

Although Ockham died in 1347, less than forty years after Scotus, Wodeham saw Scotus as an 'ancient' author and Ockham as a modern contemporary. Respectful but seldom vitally concerned with his views, Wodeham calls Thomas Aquinas the *Doctor communis* (Universal Doctor). Many subsequent authors, even in the fifteenth century, saw the history of philosophical theology in the same way (Courtenay 1978). Wodeham's view of the history of thought was based less on the dating of authors than on the extent to which their theology was characterized by explicit appeals to logical arguments. The continued currency of his views is a measure of his influence.

Huge crowds flocked to Wodeham's lectures in the 1330s, and John Major held that Wodeham would be reputed a greater philosopher than Ockham were it not for Ockham's political writings. However, almost nothing by or about Wodeham was published in the four centuries between 1512 and 1976. Despite a spate of recent publications on Wodeham, comparatively little is or can be known about his views until a critical edition of his most important work is prepared.

See also: Chatton, W.; Language, medieval theories of; Natural philosophy, medieval; William of Ockham

REGA WOOD

List of works

Wodeham's most significant works were three sets of theology lectures based on Peter Lombard's *Sentences*. His first lectures, delivered in London, have not been independently preserved, though portions are included in his second lectures. His second lectures (*Lectura secunda*), intended for a Franciscan audience in Norwich, were first published in 1990. The final lectures, written for a larger Oxford audience, are his most influential work but still await publication. In the Renaissance, John Major chose to print Henry Totting von Oyta's abbreviation of the Oxford lectures, rather than the work itself. Lost works by Wodeham include two Bible commentaries (on the *Canticum Canticorum* and on *Ecclesiasticus*), a set of *Determinationes* and a treatise against Richard of Wetherset, defending the right of friars to hear confessions.

Wodeham, Adam (after 1323, before 1331) *Tractatus de indivisibilibus*, ed. R. Wood, Dordrecht: Kluwer, 1988. (Wodeham's longest sustained discussion of issues in natural philosophy, which quotes frequently from Ockham's *Physics*. Also printed here is Wodeham's *Quaestione de divsione et compositione continua*, dated 1322-31.)

Wodeham, Adam (before 1333) *Lectura Oxoniensis*, ed. J. Major, *Abbreviatio Henrici Totting de Oyta, seu Adam goddam super quattuor libros sententiarum*, Paris: P. le Preux, 1512.(This Renaissance edition prints Henry Totting von Oyta's abbreviation of the Oxford lectures, not the work itself. In addition to the bulk of *Lectura Oxoniensis*, this volume also includes Major's life of Wodeham.)

Wodeham, Adam (before 1339) *Lectura secunda*, ed. R. Wood and G. Gál, St Bonaventure, NY: St Bonaventure University, 1990.(Though limited to the first book of the *Sentences*, these three volumes include more than 1260 pages of exposition. The introduction discusses Wodeham's life and influence and summarizes some of his views. There is a bibliography in Vol. III.)

References and further reading

Adams, M.M. (1993) 'Review of Wodeham's *Lectura secunda*', *Philosophical Review* 102: 588-94.(An able summary of about 800 pages of text.)

Adams, M.M. and Wood, R. (1981) 'Is to Will It as Bad as to Do It?', *Franciscan Studies* 41: 5-60.(Compares Wodeham with Scotus and Ockham on the morality of external acts.)

Courtenay, W. (1975) 'Ockhamism among the Augustinians: the Case of Adam Wodeham', *Scientia augustiniana*, ed. C. Mayer and W. Eckermann, *Cassiciacum* 30: 267-75.(Illustrates Wodeham's influence.)

Courtenay, W. (1978) *Adam Wodeham*, Leiden: Brill.(The most complete introduction to Wodeham's life and times.)

Gál, G. (1977) 'Adam Wodeham's Question on the *Complexe Significabile*', *Franciscan Studies* 37: 66-102. (Shows that Wodeham, not Gregory of Rimini, should be credited with the view that the immediate object of an act of scientific knowledge is the *complexe significabile*.)

Gelber, H.G. (1974) 'Logic and the Trinity: A Clash of Values in Scholastic Thought, 1300-1335', Ph.D. dissertation, University of Wisconsin.(See especially 235-64 and 615-28, where extended excerpts from Wodeham's *Lectura Oxoniensis* I d.33 are printed and discussed.)

Grassi, O. (1986) *Intuizione e Significato: Adam Wodeham e il problema della conoscenza nel XIV secolo (Intuition and its Significate: Adam Wodeham and the Problem of Knowledge in the Fourteenth Century)*, Milan: Editoriale Jaca.(Epistemological study comparing Aureol, Ockham, Chatton and Wodeham; based on texts from Wodeham's *Lectura Oxoniensis* as well as the *Lectura secunda*.)

Kretzmann, N. (1984) 'Adam Wodeham's Anti-Aristotelian Anti-Atomism', *History of Philosophy Quarterly* 1: 381-98.(Characterizes Wodeham as a macro-indivisibilist.)

Maierù, A. (1984) 'Logique et Théologie Trinitaire: Pierre d'Ailly' (Logic and Trinitarian Theology: Pierre d'Ailly), in A. Kaluz and P. Vignaux (eds) *Preuve et Raisons. l'Université de Paris*, Paris: Vrin, 253-68.(A study of Pierre d'Ailly's response to Wodeham and Gregory of Rimini in the field of Trinitarian theology.)

McGrade, A.S. (1987) 'Enjoyment after Ockham: Philosophy, Psychology and the Love of God', in A. Hudson and M. Wilks (eds) *From Ockham to Wyclif*, Oxford: Blackwell, 63-88.(Wodeham on fruition.)

Nuchelmans, G. (1980) 'Adam Wodeham on the Meaning of Declarative Sentences', *Historiographia Linguistica* 7: 177-86.(Insightful analysis suggesting that Wodeham's theory is more coherent than that of Rimini.)

Reina, M.E. (1986) 'Cognizione intuitiva ed esperienza interiore in Adamo Wodeham' (Intuitive Cognition and Internal Experience in Adam Wodeham), *Rivista di storia della filosofia* 41: 19-49, 211-44.(A detailed analysis of Wodeham's views on the soul and introspective cognition.)

Spade, P.V. (1988) 'Anselm and the Background to Adam Wodeham's Theory of Abstract and Concrete Terms', *Rivista di storia della filosofia* 43: 261-71.(Clear and concise.)

Tachau, K. (1988) *Vision and Certitude in the Age of Ockham*, Leiden: Brill, 275-99. (Wodeham viewed as an influential interpreter of Peter Aureol.)

Wood, R. (1982) 'Adam Wodeham on Sensory Illusions', *Traditio* 38: 214-52.(Discusses Wodeham's refutation of Peter Aureol and Walter Chatton on the topic of sensory illusions.)

Wood, R. (1989) 'Epistemology and Omnipotence', in S. Chodorow and J. Sweeney (eds) *Popes, Teachers and Canon Law in the Middle Ages*, Ithaca, NY: Cornell University Press, 160-78.(Establishes Wodeham's influence on John of Mirecourt, and compares Wodeham's views on certainty and intuitive cognition of nonexistents with his contemporaries.)

Zupko, J. (1994) 'Nominalism Meets Indivisibilism', *Medieval Philosophy and Theology* 3: 158-85.(Summarizes Wodeham's discussion of contiguity and explores his influence on Buridan.)

Wolff, Christian (1679-1754)

Christian Wolff was a rationalistic school philosopher in the German Enlightenment. During the period between the death of Leibniz (1714) and the publication of Kant's critical writings (1780s), Wolff was perhaps the most influential philosopher in Germany.

There are many reasons for this, including Wolff's voluminous writings in both German and Latin in nearly every field of philosophy known to his time, their unvarying employment of a strict rationalistic method to establish their conclusions, the attention directed to Wolff and his views as a result of bitter controversies with some theological colleagues, his banishment from Prussia by King Frederick Wilhelm I in 1723 and triumphant return from Hesse-Cassel in 1740 after Frederick the Great assumed the throne, and his active teaching at the Universities of Halle and Marburg for nearly 50 years. Through his work as a university professor, his prolific writings, and the rigour and comprehensiveness of his philosophy, Wolff influenced a very large group of followers, educators and other writers. Even after his influence had begun to wane, Kant still referred to 'the celebrated Wolff' and spoke of 'the strict method of the celebrated Wolff, the greatest of all dogmatic philosophers'.

Wolff thought of philosophy as that discipline which provides reasons to explain why things exist or occur and why they are even possible. Thus, he included within philosophy a much broader range of subjects than might now be recognized as 'philosophical'. Indeed, for Wolff all human knowledge consists of only three disciplines: history, mathematics and philosophy.

The reasons provided by Wolff's philosophy were to be established through unfailing adherence to a strict demonstrative method. Like Descartes, Wolff first discovered this method in mathematics, but he concluded that both mathematical and philosophical methods had their ultimate origins in a 'natural logic' prescribed to the human mind by God. In fact, the heart of Wolff's philosophical method is a deductive logic making use of syllogistic arguments.

For Wolff, the immediate objective of philosophical method is to achieve certitude by establishing an order of truths within each discipline and a system within human knowledge as a whole. The ultimate goal is to establish a reliable foundation for the conduct of human affairs and the enlargement of knowledge.

Wolff applied his philosophical method unfailingly in each of the three principal parts of philosophy: metaphysics - knowledge of those things which are possible through being in general, the world in general, human souls, and God; physics - knowledge of those things which are possible through bodies; and practical philosophy - knowledge of those things which are concerned with human actions. Wolff's philosophical system also includes logic, an art of discovery (to guide the investigation of hidden truth and the production of new insights), some experiential disciplines (for example, empirical psychology) and several bodies of philosophical knowledge that were not well developed in Wolff's time concerning law, medicine, and both the practical and liberal arts.

1 Life and works

Christian Wolff was born in Breslau, where he received his early education. He studied at the University of Jena and received a master's degree in 1703 at the University of Leipzig. In these settings, Wolff became familiar with Lutheran, Calvinist and Roman Catholic viewpoints in theology; Aristotelian, Scholastic and Cartesian school traditions in philosophy; and the emerging empirical methods of Newtonian science. The most important single influence on Wolff's thought was Leibniz. But Wolff resisted characterizations of himself as merely a follower or systematizer of Leibniz's thought. This unflattering stereotype first arose during Wolff's lifetime and is commonly heard today.

Leibniz provided a letter of recommendation which helped Wolff obtain his first teaching position as Professor of Mathematics and the Natural Sciences at the new University of Halle in 1706 where he taught and wrote about the subjects indicated, only gradually taking an interest in selected philosophical topics.

In 1713, Wolff published his *Vernünftige Gedanken von den Kräften des menschlichen Verstandes (Rational Thoughts on the Powers of the Human Understanding; referred to as the German Logic)*, an introductory handbook on logic. This became the most popular of all of Wolff's writings, appearing in fourteen different

editions during his lifetime. It was his only major philosophical work published during Leibniz's lifetime.

During the early 1720s, Wolff published a series of philosophical works in German. These included books on metaphysics (1720), ethics (1720), politics (1721), physics (1723), teleology (1724), and physiology (1725), many of which later appeared in numerous reprintings or new editions.

The titles of all of Wolff's central philosophical writings in German begin with the common phrase, 'Rational Thoughts on… '. For example, the *German Metaphysics* (published in twelve editions during Wolff's lifetime) bears the ambitious title, 'Rational Thoughts on God, the World, and the Human Soul, and All Things in General', indicating both the methodic rigour which the author prized and the ambitious scope of the work.

Together with a volume of annotations on the metaphysics (1724) and another (1726) offering a review and defence of all of Wolff's previous writings on the various parts of philosophy ('Weltweisheit'), these German publications contributed significantly to the establishment of the German language as a vehicle for scientific communication (Blackall 1959). In addition, during Wolff's lifetime Latin remained in wide use by scholars and other vernacular languages (for example, French) were far more advanced in their development than German.

Unfortunately, Wolff's writings and lectures entangled him in bitter quarrels with critics led by Pietist members of the theology faculty at Halle, such as Joachim Lange, and Johann Franz Budde at Jena (see Pietism). The issues had to do with the proper roles of reason and faith in human life, along with charges that Wolff's views implied a fatalistic and necessitarian view of the world, human beings and God. For example, it seemed heretical to Wolff's opponents for him to have said in a public lecture in 1721 that the pagan Chinese could develop a satisfactory ethical philosophy on the basis of reason alone without the assistance of the Christian revelation. Similarly, Wolff's acceptance of the Leibnizian doctrine of pre-established harmony and his emphasis on God's intellect as the controlling framework for divine freedom and power seemed to his critics to lead directly to a deterministic universe.

When Wolff did not give in to his critics and could not otherwise be bested, his opponents appealed to external authority. Without holding a hearing on the charges against him, King Frederick William I issued an order on 8 November 1723 that removed Wolff from his professorship and banned him from Prussia within forty-eight hours on pain of death.

Wolff was able to take up a teaching post previously offered to him at the University of Marburg in Hesse-Cassel. Official censure and exile only heightened his popularity. Nevertheless, Wolff decided to write no more in German and to turn to the wider intellectual world by composing a new series of philosophical publications in Latin, the dominant scientific language of the time. That series occupied the rest of his life.

Wolff's Latin works expanded and restated in a more thoroughgoing and systematic way his views on philosophy and other subjects. They began with a Preliminary Discourse on Philosophy in General (1728). Most of Wolff's Latin writings share a common subtitle, *methodo scientifica pertractata* ('treated according to the scientific method'). Thus, the Latin ontology volume is entitled, *Philosophia prima, sive Ontologia, methodo scientifica pertractata, qua omnis cognitionis humanae principia continentur (First Philosophy, or Ontology, Treated According to the Scientific Method, in Which Are Contained All of the Principles of Human Knowledge; referred to henceforth as the Latin Ontology).*

Wolff's Latin works include treatises on logic (1728), ontology (1729), general cosmology (1731), empirical (1732) and rational (1734) psychology, natural theology (2 vols, 1736-7), practical philosophy (2 vols, 1738-9), natural law (8 vols, 1740-8), international law (1749), ethics (5 vols, 1750-3) and economics (2 vols, 1754-5). In 1740, Wolff was recalled to the University of Halle where he became Chancellor in 1743. After a period of fame and influence, he died on 9 April 1754, the author of some seventy volumes of published work in mathematics and philosophy.

Wolff's distinctive contributions involve his description of the scope of human knowledge (comprising history, philosophy and mathematics), his definition of philosophy and its goals, his systematic description of the parts of philosophy, his views on philosophical method and some of his central metaphysical doctrines.

2 Human knowledge: history, mathematics and philosophy

Wolff divided all human knowledge into three basic types: (1) history or ’knowledge of those things which are and occur either in the material world or in immaterial substances’; (2) philosophy or ’knowledge of the reason of things which are or occur’; and (3) mathematics or ’knowledge of the quantity of things’ (*Preliminary Discourse* §§3, 6 and 14). This division is intended to organize human knowledge by finding the proper place for each division and identifying its role in the schema of human understanding.

According to Wolff, philosophy originates in experience or historical knowledge. Some historical facts are common to all, requiring only human attention and some acumen. Other facts ’do not spontaneously present themselves to one who is attentive’ (*Preliminary Discourse* §20). Such facts are hidden and must be sought out by observation, investigation, and experiments. Mathematics assists the search for such facts, as well as their philosophical explanation, by perfecting the evidence, guiding reason and confirming demonstrations.

3 Philosophy: definition and goals

Wolff defined philosophy more precisely as ’the science of the possibles in so far as they can be’ (*Preliminary Discourse* §29; École 1978). A key element here is the meaning of the term ’science’. ‘By science here I mean the habit of demonstrating propositions, that is, the habit of inferring conclusions by legitimate sequence from certain and immutable principles’ (*Preliminary Discourse* §30). Scientific knowledge is characterized by its method and order, that is, certain and immutable principles, methodic inferences, and orderly conclusions.

Wolff’s proximate goal in philosophy and the trait which is most typical of his writings was demonstration, the ’showing forth’ of conclusions distinguished by their certitude. Thus, he wrote that ‘nothing is more important than certitude’ (*Preliminary Discourse* §28). In fact, however, Wolff pursued certitude as the foundation upon which philosophy can best serve human interests and needs. Thus, he observed that ‘the fruit of philosophy… [is] its usefulness in knowledge and in life’ (*Preliminary Discourse* §116; compare Corr 1970).

4 The parts of philosophy

Wolff’s mature description of the parts of philosophy appears in Table 1 (compare Blackwell 1961b). The basic division is between those parts of philosophy that deal with things and those concerned with human actions. Within that, philosophical discussions of things are divided into metaphysics and physics, while practical philosophy is divided into ethics, economics and politics, together with some mention of philosophy of the arts and with an important role for logic or the philosophy of human cognition.

Table 1. Wolff's division of philosophy

Of things	**Of human actions**			
Metaphysics (being in general) (spirits) (the world in general)		Ontology (being in general)	Cognitive philosophy	Logic (directing the cognitive faculty in knowing truth and avoiding error
		General cosmology (the world in general)		Ars inveniendi (art of investigating or discovering hidden truth)
	Pneumatics (spirits)	Psychology (human souls)	Practical philosophy (directing the appetitive faculty in choosing good and avoiding evil)	Jus naturae or Law of nature (knowledge of good and evil actions)
		Natural theology		Universal practical

		(God)		philosophy (general theory and practice of practical philosophy
Physics (bodies)	General physics (general affections of all bodies or of many diverse species)			Ethics (directing man's free actions in the natural state or individual morality)
	Cosmology (world as such)			Economics (directing free actions in smaller societies distinct from the state)
	Meteorology (atmospheric phenomena)			Politics (directing free actions in a civil society or state)
	Oryctology (minerals)		Philosophical knowledge of the arts	(reasons explaining the arts and works of art, whether medicine or law, practical arts such as wood cutting, agriculture, or civil architecture [technology], or the liberal arts (grammar, rhetoric, poetry)
	Hydrology (fluids)			
	Phytology (plants or vegetables)			
	Physiology (animated bodies)			
	Teleology (final cause of natural things)			

According to Wolff, ‘the beings which we know by examining ourselves before we philosophize are God, human souls and bodies’ (*Preliminary Discourse* §55). Natural theology is ’the science of those things which are known to be possible through God’; psychology is ’the science of those things which are possible through human souls’

(*Preliminary Discourse* §§57-8). In addition, ontology or first philosophy is ’the science of being in general, or in so far as it is being’, and general cosmology is ’the science of the world in general’ or of that which is common to the existing world and to any other possible world (*Preliminary Discourse* §§73 and 78).

Together, natural theology and psychology make up ’pneumatics’ or ’the science of spirits’; along with ontology and general cosmology, these four bodies of philosophical knowledge constitute ’metaphysics’ (*Preliminary Discourse* §79).

Physics is ’the science of those things which are possible through bodies’ (*Preliminary Discourse* §59). It includes general physics (the science of those things which are common to all bodies or to many diverse kinds of bodies), together with specific physical sciences of particular kinds of bodies.

Practical philosophy is ’the science of directing the appetitive faculty in choosing good and avoiding evil’ (*Preliminary Discourse* §62). Its components address free human actions in different situations: ethics examines free actions in the natural state or when humans act as individuals; economics studies free actions within smaller societies distinct from the state; and politics explores free actions in a civil society or state. Practical philosophy also includes a general science or basic theory about good and evil actions (the law of nature or *jus naturae*), and a body of knowledge about the general theory and practice of practical philosophy (universal practical philosophy). Wolff acknowledged that these last two sciences (presumably, along with *jus gentium* or international law which he developed later) could in principle be treated as subordinate parts of ethics, economics, and politics, but he preferred to keep them separate for the sake of clarity and demonstrative rigour.

Within his discussion of human action, Wolff admitted the possibility of philosophical knowledge about the arts, whether these are the practical arts like wood cutting, agriculture, or civil architecture, bodies of knowledge like law and medicine, or the liberal arts like grammar, rhetoric and poetry. Nevertheless, he acknowledged that philosophers had not previously given much attention to the reasons behind these types of human arts.

Logic is ’the science of directing the cognitive faculty in the knowing of truth’ and avoiding error (*Preliminary Discourse* §61). Because logic guides all of philosophy, Wolff recommended that it be given first place among philosophical disciplines. However, to demonstrate its own rules, logic must follow ontology and psychology from which its principles are derived. In addition to logic, Wolff also commended a new body of knowledge called the ’art of discovery’ (*ars inveniendi*) (*Preliminary Discourse* §74) intended to guide the investigation of hidden truth by following the example of algebra and the analytic arts of the mathematicians (Corr 1972).

Wolff believed ‘there are many philosophical disciplines which still lie hidden’ (*Preliminary Discourse* §86). His review of the parts of philosophy was intended to organize those bodies of philosophical knowledge known to his time and to show that the various parts of philosophy must be ’ordered in such a way that those parts should come first which provide principles for the other parts’ (*Preliminary Discourse* §87; see Table 1).

Wolff chose the method of learning over the method of demonstration by beginning with logic because of its importance in leading to a knowledge of all other parts of philosophy. Also, Wolff believed that logic could easily explain the ontological and psychological principles which it needed.

Similarly, because ‘the principles of philosophy must be derived from experience’ (*Preliminary Discourse* §34), Wolff acknowledged the importance of observations and experiments in philosophy. In addition, he hoped to develop an art of discovery, and he distinguished between empirical and rational psychology (*Preliminary Discourse* §§111-12; see Blackwell 1961a; Corr 1975b). All of this was intended to contribute to ‘a marriage of reason and experience’ (*Latin Logic* §1232) within philosophy in particular and human knowledge in general.

Nevertheless, it is the requirements of demonstration that govern Wolff’s analysis of the parts of philosophy. Accordingly, metaphysics must precede physics and practical philosophy, and within each of these three primary philosophical disciplines the order of their subordinate parts should be that given in Table 1. Furthermore, demonstrative method is the most notable feature in each part of Wolff’s philosophy.

5 Philosophical method

‘Philosophical method is the order which the philosopher ought to use in treating dogmas’ (*Preliminary Discourse* §115). Furthermore, ‘the supreme law of philosophical method is that those things must come first through which

later things are understood and established' (*Preliminary Discourse* §133). In other words, method involves a certain kind of ordering of dogmas or propositions. It proceeds from accurate definitions and precise terminology to establish certain and immutable principles, legitimate deductions, and demonstrated conclusions. Initial conclusions become the bases for further inferences and deductions, all of which yields certitude and utility.

Like some contemporaries, Wolff believed that philosophical and mathematical method are fundamentally identical because they derive from a common source, the notion of certitude. Thus, he observed that 'philosophy does not borrow its method from mathematics; rather, both philosophy and mathematics derive their methods from true logic' (*Preliminary Discourse* §139). In his *Eigene Lebensbeschreibung (Autobiography)* (pp. 121-2, 143), Wolff wrote that he had studied mathematics in order to discover the key to its success and to apply this discovery to the improvement of philosophy, theology, and the art of discovery. This reflects Wolff's commitment to a 'natural logic' or set of 'rules prescribed by God to the understanding and the natural ability to act accordingly' (*German Logic*, ch. 16 §3; compare *Latin Logic* §6; also Corr 1970).

Unlike others, Wolff found in the categorical syllogism what he understood to be the basic principles and the best expression of true logic and philosophical method. 'By these syllogisms everything is found out that can be discovered by human understanding, and everything is proven to others that they want to be convinced of, although neither in discovering nor in demonstrating do we always have the form of the syllogisms distinctly before our eyes' (*German Logic*, ch. 4 §20; compare *Latin Logic* §50). Thus, 'a syllogism is a speech in which reasoning or discourse is distinctly stated' (*Latin Logic* §332).

Whenever certitude cannot be established, a principle of tried usefulness can be admitted into philosophy as a probability or hypothesis, that is, 'an assumption which cannot yet be demonstrated, but which provides a reason' (*Preliminary Discourse* §126). This recognizes that, 'probability, in so far as it is clearly distinguished from certitude, is not to be eliminated from philosophy' (*Preliminary Discourse* §132).

6 Selected metaphysical doctrines

As noted above, Wolff defined philosophy as 'the science of the possibles in so far as they can be' (*Preliminary Discourse* §29). Accordingly, natural theology is the science of that which is possible through God, physics the science of that which is possible through bodies, and so forth.

This means that philosophy must provide scientific or demonstrative explanations as to why anything is or is not intrinsically possible (that is, not inherently self-contradictory). Further, philosophy must also explain why any possible being actually does or does not exist. Wolff defined existence as the 'fulfilment of possibility' (*German Metaphysics* §14) or the 'complement of possibility' (*Latin Ontology* §174; Van Peursen 1987).

Within Wolff's system, 'ontology or first philosophy is defined as the science of being in general, or in so far as it is being' (*Latin Ontology* §1; École 1990). Ontology is rooted in two central axioms which apply equally to the orders of being and of knowing (there is no separate, explicit discussion of theory of knowledge within Wolff's philosophy; see Gracia 1994). These fundamental axioms are the principle of contradiction and the principle of sufficient reason. The first and most basic is the principle of contradiction: 'it cannot happen that the same thing is and is not' (*Latin Ontology* §28). The principle of sufficient reason ('nothing exists without a sufficient why it is able to be rather than is not'), is explicitly subordinated to and deduced from the principle of contradiction (*Latin Ontology* §70). Following these two fundamental principles, every philosophical explanation must demonstrate two things about the subject matter in question: (1) that it is not impossible, that is, not inherently self-contradictory; and (2) that there is a sufficient reason why it actually exists or occurs.

'Being' is defined by Wolff as 'that which is able to exist, that is, to which existence is not repugnant' (*Latin Ontology* §134). The structure of being depends on three principal components: (1) 'essentialia' or ultimate factors within a being characterized by their lack of mutual determination and contradiction; (2) 'attributes' or constant factors which depend on 'essentialia' to explain their possibility and actuality; and (3) 'modes' or variable elements which must be compatible with its 'essentialia', but whose actual presence depends on external factors such as circumstances and other agents (*Latin Ontology* §§143, 146 and 148).

Both Wolff's ontology and general cosmology are notable for not including the Leibnizian doctrine of monads (see Leibniz, G.W. §§4-5). Wolff described the world and the bodies within it as complex or compound entities

constituted by the interconnection of simpler entities (*General Cosmology* §§48 and 119). Ultimately, the most fundamental of these simple substances or 'elements' are indivisible, physical points or 'atoms of nature' (*General Cosmology* §187) which are imperceptible and simple. They lack extension or figure, but are not characterized by a spiritual force or the perceptual activity of Leibniz's monads.

As noted above, Wolff divided psychology into 'empirical' and 'rational' parts (Blackwell 1961a; Corr 1975b). This division respects the need to incorporate observation and experience within philosophy, although the way in which the two parts of psychology are implemented reflects Wolff's search for demonstrative conclusions. Thus, empirical psychology organizes the facts of our psychological experience through an extended review of human cognitive and appetitive operations, and an attempt to define the basic vocabulary of psychology. It also offers a pivotal argument for the existence of the human soul:

> Whatever being is actually conscious of itself and of other things outside itself, that being exists. But we are actually conscious of ourselves and of other things outside ourselves. Therefore we exist.
>
> (*Empirical Psychology* §16)

On this basis, rational psychology provides a definition of the essence of the human soul: 'The essence of the soul consists in the force of representing the universe limited materially by the placement of the organic body in the universe and formally by the constitution of the sensory organs' (*Rational Psychology* §66).

Wolff's psychology accepted a classic dualism between soul and body (*Rational Psychology* §§530-642). Ideally, empirical psychology would disclose the simultaneous occurrence of mental and bodily actions, while rational psychology would systematically explain their interdependence. However, Wolff conceded that he had not been able to find any reason in the nature of the soul from which the necessity of its 'commerce' or interaction with the body could be deductively demonstrated. As a result, he resorted to a hypothetical explanation. Among possible candidates for this, both direct physical influx or interaction and occasionalism were rejected as unlikely (see Occasionalism), while Leibniz's theory of pre-established harmony (see Leibniz, G.W. §6) was accepted as the most likely account (*Rational Psychology* §639). However, in Wolff's view pre-established harmony was not without its own difficulties and was admitted only as a probable explanation limited to the relationship between the human soul and body.

Natural theology is 'the science of those things which are possible through God' (*Natural Theology* I §1). By contrast with a supernatural or revealed theology originating in a divine revelation or sacred scripture, natural theology proceeds 'solely by the light of nature' (*Natural Theology* I §1n). Nevertheless, for Wolff 'there is no more sublime philosophy than that which clearly demonstrates the highest perfections of the supreme divinity and how all things are in it, from it, and through it, and which completely roots out the profane thoughts of men concerning God and religion' (*Natural Theology*, dedication to Part I).

Wolff divided natural theology into two major parts: Part I contains an a posteriori demonstration of the existence and attributes of God drawn from the characteristics of the visible world; Part II treats the same subjects from an a priori standpoint based on the notion of a most perfect being and the nature of the human soul (Bissinger 1970; Corr 1973). The a posteriori argument for the existence of God is complex (compare, *Natural Theology* I §§24 onwards): the human soul exists or we exist; there must be a sufficient reason why this is so; that reason must be contained either in ourselves or in some other being different from us; in either case, the reason must lie in a necessary being if it is to be truly sufficient; therefore, a necessary being exists; a necessary being is an *ens a se* (being in itself) or one which contains the sufficient reason of its existence in its own essence, that is, an '*ens a se* exists therefore, because it is possible' (*Natural Theology* I §34); neither the visible world, its elements, or the human soul can be *ens a se*; therefore, God exists; 'God' being defined nominally as 'the *ens a se* in which is contained the sufficient reason of the existence of this visible world and of our souls' (*Natural Theology* I §67).

Wolff noted that an *ens a se* (in contrast to an *ens ab alio* or being from another) is possible because its essential constituents involve no contradiction; it exists because it possesses these particular essential determinations. The fact that its essential constituents explain both its possibility and its actual existence is the special 'privilege' of an *ens a se*. 'And thus it is that existence is not inferred from possibility in general, nor is existence determined from possibility considered in general or in itself' (*Natural Theology* I §34n).

Wolff's a priori proof for the existence of God was based on the notion of an *ens perfectissimum* or 'most perfect

being' understood as that being 'to which belong all compossible realities in the absolutely highest degree' (*Natural Theology* II §6). The *ens perfectissimum*, nominally defined as God, contains in the highest degree all of the real or ontological components of the essence of a being which are both possible in themselves and in relation to each other. On this basis, the a priori demonstration of God's existence is this:

> God contains all compossible realities in the absolutely highest degree. But He is possible. Wherefore, since the possible can exist, existence can belong to it. Consequently, since existence is a reality, and since realities are compossible which can belong to a being, existence is in the class of compossible realities. Moreover, necessary existence is the absolutely highest degree. Therefore, necessary existence belongs to God or, what is the same, God necessarily exists.
>
> (*Natural Theology* II §21)

In other words, 'God exists through His essence or His existence is essential' (*Natural Theology* II §27).

The principal attributes of God are intellect, power, and will. The divine intellect explains the intrinsic possibility of the world (its lack of internal self-contradiction), the divine power explains its extrinsic possibility (or compatibility with the nexus of other existents that together constitute a particular universe), and the divine will accounts for its actual existence (by providing the sufficient reason for or final determination of that existence). In addition, 'God wills whatever He represents to Himself as the best, both in itself and in relation to Himself' (*Natural Theology* I §390). Thus, the visible universe is the best of all possible worlds. But Wolff believed God may bring other possible worlds - one or all of them - into actual existence along with the best one. Therefore, Wolff maintained that the existence of any world is only a contingent fact. God acts reasonably, but the divine creative act is not coerced or necessitated by the objective, existential dynamism of a world of possible essences.

7 Practical philosophy

Wolff wrote extensively on various areas of practical philosophy, but that portion of his philosophy has had less influence than his views on method and metaphysics. Contrary to radical fideists, Wolff argued for a moral philosophy based on the concept of perfection (*German Ethics* §2). The good is that which makes us and our condition more perfect; the bad is that which makes us and our condition more imperfect. These notions arise from the structure of being and can be known by human beings through natural reason even without the assistance of Christian revelation. In short, neither the good nor our knowledge of the good are directly dependent upon God's will or influence.

Perfection involves self improvement, both in internal and external conditions, as well as the improvement of other human beings and the promotion of God's honour. In striving for this extended concept of perfection, human beings are acting according to a natural moral law, one which applies to individuals, small communities, nation states, and the international human community. The idea of perfection is the fundamental motivating force in the practical sphere. Thus, intellect or reason directs the will in the search for an objective moral good. And education or the formation of clear and distinct ideas associated with perfection is the chief spur to moral living. Here is the ultimate outcome of Wolff's conviction that clear and certain ideas are the best foundation for both the advancement of theoretical knowledge and practical living.

8 An assessment

Christian Wolff is not a first-rank, original thinker. He was important in particular as an organizer and systematizer of philosophical thought during the eighteenth century. Wolff's thought emphasized clarity, rigour and comprehensiveness. In the history of German philosophy, Wolff was the first major systematic thinker. His philosophical system helped to define important philosophical terms and advance the cause of the German language as a tool for scholarly work.

Leibniz was the single most important influence on Wolff's thought. But Wolff did not like to be regarded merely as someone who gave a new and more orderly presentation to the views of his great predecessor (*Autobiography*: 142-3). In fact, there were important limitations in what the two men knew of each other's writings and Wolff departed from Leibniz's views on several key points (Corr 1975a). In addition, Wolff drew upon other sources as well (Corr 1983). And Wolff helped to make room for the emerging sciences of his day, even though his own deductive methodology in philosophy was not wholly compatible with the methods of those sciences.

Through his lengthy teaching career, extensive writings, and many followers, Wolff was extremely influential in shaping German philosophical thought for a period of fifty years and more during the middle of the eighteenth century. Against the Pietists who became his opponents, Wolff spoke on behalf of what he regarded as the legitimate claims of natural human reason, as well as the compatibility of faith and reason (École 1983). In defending his own views, he composed extended refutations of atheism, deism, fatalism, naturalism and Spinozism (Morrison 1993). In reaction to censure from his theological critics and political opponents, Wolff also wrote eloquently on behalf of the freedom to philosophize (*Preliminary Discourse* §§151-71), arguing that 'if one is to use philosophical method in developing philosophy, then the yoke of philosophical servitude cannot be imposed' (§153). Praised by Kant for the rigour of his thought, Wolff helped to frame many of the issues addressed by his great successor.

See also: Enlightenment, Continental; Rationalism

CHARLES A. CORR

List of works

Wolff, C. (1964-) *Christian Wolff's Gesammelte Werke (Collected Works)*, Hildesheim and New York: Georg Olms.(Wolff's collected works are being reprinted in three series: German writings; Latin writings; and a supplementary series of materials and documents relevant to Wolff's life, works and thought. This very large publishing project is now almost complete. It consists of a total of 109 separate volumes: 29 German, 41 Latin and 39 supplementary.)

1 German series

The principal works of philosophical importance in the German series are given here.

Wolff, C. (1713) *Vernünfftige Gedancken von den Kräfften des menschlichen Verstandes und ihrem richtigen Gebrauche in Erkänntnis der Wahrheit*, trans. as *Logic, or Rational Thoughts on the Powers of the Human Understanding, with Their Use and Application in the Knowledge and Search of Truth*, London: L. Hawes, W. Clarke & R. Collins, 1770.(Referred to as the *German Logic*; translated anonymously into English.)

Wolff, C. (1720) *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt (Rational Thoughts on God, the World and the Human Soul, and All Things in General)*.(Referred to as the *German Metaphysics*.)

Wolff, C. (1720) *Vernünfftige Gedancken von der Menschen Thun und Lassen, zu Beförderung ihrer Glückseeligkeit, den Liebhabern der Wahrheit mitgetheilet (Rational Thoughts on Human Conduct)*.(Referred to as the *German Ethics*.)

Wolff, C. (1721) *Vernünfftige Gedancken von dem gesellschaftlichen Leben der Menschen und insonderheit dem gemeinen Wesen zu Beförderung der Glückseeligkeit des menschlichen Geschlechts (Rational Thoughts on the Social Life of Humans and Especially the Commonwealth for the Promotion of the Happiness of the Human Race)*.(Referred to as the *German Politics*.)

Wolff, C. (1723) *Vernünfftige Gedancken von den Wirkungen der Natur, den Liebhabern der Wahrheit mitgetheilet (Rational Thoughts on the Operations of Nature*.(Referred to as the *German Physics*.)

Wolff, C. (1724) *Vernünfftige Gedancken von den Absichten der natürlichen Dinge, den Liebhabern der Wahrheit mitgetheilet (Rational Thoughts on the Purposes of Natural Things)*.(Referred to as the *German Teleology*.)

Wolff, C. (1724) *Anmerckungen über die vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt, zu besserem Verstande und bequemeren Gebrauche derselben (Annotations to the Rational Thoughts on God, the World and the Human Soul)*.

Wolff, C. (1725) *Vernünfftige Gedancken von dem Gebrauche der Theile in Menschen, Thieren und Pflantzen, den Liebhabern der Wahrheit mitgetheilet (Rational Thoughts on the Use of Parts in Humans, Animals and Plants)*.(Referred to as the *German Physiology*.)

Wolff, C. (1726) *Ausführliche Nachricht von seinen eigenen Schrifften die er in deutscher Sprache von den verschiedenen Theilen der Welt-Weiszheit herausgegeben, auf Verlangen ans Licht gestellet (Detailed Account of His Own Works Which He Published in the German Language on the Different Parts of Philosophy)*.

Wolff, C. (*c*.1740) *Christian Wolff's eigene Lebensbeschreibung (Autobiography)*.

2 Latin series

The principal works of philosophical importance in the Latin series.

Wolff, C. (1718) *Ratio praelectionum Wolffianarum in mathesin et philosophiam universam (Plan of the Wolffian Lectures in Mathematics and Universal Philosophy)*.

Wolff, C. (1728) *Discursus praeliminaris de philosophia in genere*, trans. R.J. Blackwell, *Preliminary Discourse on Philosophy in General*, Indianapolis, IN: Bobbs-Merrill, 1963.(Originally published together with the *Philosophia rationalis sive Logica*.)

Wolff, C. (1728) *Philosophia rationalis sive Logica, methodo scientifica pertractata et ad usum scientiarum atque vitae aptata (Rational Philosophy or Logic, Treated According to the Scientific Method)*.(Referred to as the *Latin Logic*.)

Wolff, C. (1729) *Philosophia prima, sive Ontologia, methodo scientifica pertractata, qua omnis cognitionis humanae principia continentur (First Philosophy, or Ontology, Treated According to the Scientific Method)*.(Referred to as the *Latin Ontology*.)

Wolff, C. (1731) *Cosmologia generalis, methodo scientifica pertractata, qua ad solidam, inprimis Dei atque naturae, cognitionem via sternitur (General Cosmology, Treated According to the Scientific Method)*.(Referred to as the *General Cosmology*.)

Wolff, C. (1732) *Psychologia empirica, methodo scientifica pertractata, qua ea, quae de anima humana indubia experientiae fide constant, continentur et ad solidam universae philosophiae practicae ac theologiae naturalis tractationem via sternitur (Empirical Psychology, Treated According to the Scientific Method)*.(Referred to as the *Empirical Psychology*.)

Wolff, C. (1734) *Psychologia rationalis, methodo scientifica pertractata, qua ea, quae de anima humana indubia experientiae fide innotescunt, per essentiam et naturam animae explicantur, et ad intimiorem naturae ejusque autoris cognitionem profutura proponuntur (Rational Psychology, Treated According to the Scientific Method)*.(Referred to as the *Rational Psychology*.)

Wolff, C. (1736) *Theologia naturalis, methodo scientifica pertractata. Pars Prior, integrum Systema complectens, qua existentia et attributa Dei a posteriori demonstrantur (Natural Theology, Treated According to the Scientific Method. First Part of the Complete System, in Which the Existence and Attributes of God are Demonstrated a posteriori)*.(The first part of the *Natural Theology*.)

Wolff, C. (1737) *Theologia naturalis, methodo scientifica pertractata. Pars Posterior, qua existentia et attributa Dei ex notione entis perfectissimi et natura animae demonstrantur, et Atheismi, Deismi, Fatalismi, Naturalismi, Spinosismi, aliorumque de Deo errorum fundamenta subvertuntur (Natural Theology, Treated According to the Scientific Method. Second Part, in Which the Existence and Attributes of God are Demonstrated from the Notion of a Most Perfect Being and the Nature of the Soul, and the Foundations of Atheism, Deism, Fatalism, Naturalism, Spinozism and Other Errors Concerning God are Overthrown)*.(The second part of the *Natural Theology*.)

Wolff, C. (1738-9) *Philosophia practica universalis, methodo scientifica pertractata. Partes I-II (Universal Practical Philosophy, Treated According to the Scientific Method).*

Wolff, C. (1740-8) *Jus Naturae, methodo scientifica pertractatum. Partes I-VIII (Natural Law, Treated According to the Scientific Method)*.

Wolff, C. (1749) *Jus Gentium, methodo scientifica pertractatum, in quo jus gentium naturale ab eo, quod voluntarii, pactitii et consuetudinarii est, accurate distinguitur (Law Binding the World at Large, Treated According to the Scientific Method)*.

Wolff, C. (1750) *Institutiones juris naturae et gentium, in quibus ex ipsa hominis natura continuo nexu omnes obligationes et jura omnia deducuntur (Precepts of Natural Law and Law Binding the World at Large)*.

Wolff, C. (1750-3) *Philosophia moralis sive Ethica, methodo scientifica pertractata. Partes I-V (Moral Philosophy or Ethics, Treated According to the Scientific Method).*

Wolff, C. (1754-5) *Oeconomica, methodo scientifica pertractata. Partes I-II (Economics, Treated According to the Scientific Method)*.

3 Supplementary series

The supplementary series (39 separate volumes) contains: contemporaneous works on the history and influence of Wolff’s philosophy; some textbooks and other titles by Wolff’s students or followers (including Baumeister, Baumgarten, Frobesius, Gottsched, Stiebritz and Thümmig) who sought to popularize various aspects of Wolff’s thought; tracts by some of Wolff’s critics (such as Lange); and selected secondary works by modern writers (for example, Campo, École and Thomann) on Wolff’s thought.

References and further reading

Bissinger, A. (1970) *Die Struktur der Gotteserkenntnis: Studien zur Philosophie Christian Wolffs (The Structure of the Recognition of God: Studies to accompany the philosophy of Christian Wolff)*, Bonn: H. Bouvier & Co. (A careful study of Wolff's natural theology.)

Blackall, E.A. (1959) *The Emergence of German as a Literary Language 1700-1775*, Cambridge: Cambridge University Press.

Blackwell, R.J. (1961a) 'Christian Wolff's Doctrine of the Soul', *Journal of the History of Ideas* 22 (3): 339-54. (On Wolff's psychology.)

Blackwell, R.J. (1961b) 'The Structure of Wolffian Philosophy', *The Modern Schoolman* 38 (3): 203-18.(A helpful schematic exposition and analysis by the translator of Wolff's *Preliminary Discourse on Philosophy in General*.)

Corr, C.A. (1970) 'Certitude and Utility in the Philosophy of Christian Wolff', *The Southwestern Journal of Philosophy* 1 (1): 133-42.(Discussion of the basic goals of Wolff's philosophy.)

Corr, C.A. (1972) 'Christian Wolff's Treatment of Scientific Discovery', *Journal of the History of Philosophy* 10 (3): 323-34.(Exploration of Wolff's undeveloped notion of an *ars inveniendi* or art of discovery.)

Corr, C.A. (1973) 'The Existence of God, Natural Theology, and Christian Wolff', *International Journal for Philosophy of Religion* 4 (2): 105-18.

Corr, C.A. (1975a) 'Christian Wolff and Leibniz', *Journal of the History of Ideas* 36 (2): 241-62.(Detailed examination of direct testimony, circumstantial data and doctrinal comparisons bearing on the links between Wolff and Leibniz; questions the stereotype of Wolff as little more than a systematizer and vulgarizer of his predecessor's thought.)

Corr, C.A. (1975b) 'Christian Wolff's Distinction Between Empirical and Rational Psychology', *Studia Leibnitiana*, Supplementum 14: 195-215.

Corr, C.A. (1983) 'Cartesian Themes in Christian Wolff's German Metaphysics Volumes', in *Christian Wolff (1679-1754)*, ed. W. Schneiders, Hamburg: Felix Meiner, 113-20.

École, J. (1978) 'La conception wolffienne de la philosophie d'après le Discursus praeliminaris de philosophia in genere' (The Wolffian Conception of Philosophy according to the *Preliminary Discourse on Philosophy in General*), *Filosofia Oggi* 4 (4): 403-28. (Thorough review of Wolff's conception of philosophy as set forth in the *Preliminary Discourse on Philosophy in General*.)

École, J. (1983) 'Les rapports de la raison et de la foi selon Christian Wolff' (The Relationship between Reason and Faith according to Christian Wolff), *Studia Leibnitiana* 15 (2): 205-14.(Knowledgeable reflection on the relationship between reason and faith in Wolff's philosophy.)

École, J. (1990) *La métaphysique de Christian Wolff*, Hildesheim: Georg Olms; in C. Wolff, *Gesammelte Werke*, vol. 3, 12.1-12.2. (Definitive study of its subject; clearly superior to the only earlier book-length treatment (Campo's *Cristiano Wolff e il razionalismo precritico*, 2 vols, 1939; reprinted in *Gesammelte Werke*, vol. 3, section 9). École also has an extensive series of articles published between 1961-80 in journals such as *Giornale di Metafisica*, *Les études philosophiques*, *Filosofia Oggi*, *Archives de Philosophie* and *Studia Leibnitiana*, many of which arose from editorial work for the Latin philosophical texts in Wolff's *Gesammelte Werke*.)

Gerlach, H.-M., Schenk, G. and Thaler, B. (eds) (1980) *Christian Wolff als Philosoph der Aufklärung in Deutschland (Christian Wolff as Philosopher of the Enlightenment in Germany)*, Halle: Wissenschaftliche Beiträge der Martin-Luther-Universität Halle-Wittenberg.(Proceedings of a colloquium held at Halle in 1979 on the 300th anniversary of Wolff's birth; contains some two dozen contributions.)

Gracia, J.J.E. (1994) 'Christian Wolff on Individuation', in *Individuation and Identity in Early Modern Philosophy: Descartes to Kant*, ed. K.F. Barber and J.J.E. Gracia, Albany, NY: State University of New York Press, 219-43.(An expanded version of a 1993 article by the same title in *History of Philosophy Quarterly* 10 (2): 147-64 examining Wolff's views on individuation as an example of a conjunction of metaphysical and epistemological issues that the author believes is characteristic of modern philosophy.)

Morrison, J.C. (1993) 'Christian Wolff's Criticisms of Spinoza', *Journal of the History of Philosophy* 31 (3): 405-20.(Addresses issues at the heart of Wolff's philosophy and his disputes with theologians. Can usefully be compared with an earlier article by École from 1983: 'La critique wolffienne du Spinozism' (Interpretations in his philosophy and its influence, with a bibliography of literature on Wolff), *Archives de Philosophie* 46 (4): 553-67.)

Schneiders, W. (ed.) (1983) *Christian Wolff (1679-1754): Interpretationen zu seiner Philosophie und deren Wirkung, mit einer Bibliographie der Wolff-Literatur (trans?)*, Hamburg: Felix Meiner.(Proceedings of a colloquium held at Wolfenbuttel on November 21-3, 1979; contains twenty articles and a good bibliography of secondary literature published between 1800-1982.)

Van Peursen, C.A. (1987) 'Christian Wolff's Philosophy of Contingent Reality', *Journal of the History of Philosophy* 25 (1): 69-82.(Sensitive and thoughtful re-examination of this important issue in Wolff's thought.)

Wollaston, William (1660-1724)

William Wollaston, a popular eighteenth-century English moral philosopher, is often grouped with Samuel Clarke as a staunch defender of the kind of moral rationalism that David Hume later opposed. Wollaston's project, as he describes it, is to find a rule to distinguish right actions from wrong. He complains that previous philosophers have either overlooked this task or proposed rules which are imprecise, incomplete or misleading. The rule he proposes is fidelity to truth. Actions, he argues, express propositions and so may be true or false. Moral actions express truths and immoral actions express falsehoods. He thinks this rule explains other widely held views about morality, for example, that we should live in accordance with nature, right reason or the will of God. His most remembered (and most misunderstood) claim is that an evildoer 'lives a lie'.

Wollaston was born at Coton-Clanford, England. He received an MA from Sidney Sussex College, Cambridge, in 1681, and was then assistant master in Birmingham Grammar School. In his late twenties he came into a large inheritance, married an heiress and retired to a life of study. His *The Religion of Nature Delineated* was first published privately in 1722.

More than a decade before *The Religion of Nature Delineated* appeared, Samuel Clarke had argued that right actions are 'fit' or appropriate to the real nature and relations of things. Evildoers, Clarke insists, by opposing the nature and reason of things, 'endeavour (as much in them lies) to make things be what they are not and cannot be' (1706: 613). This is as absurd as attempting to alter a mathematical truth. Wollaston constructs his whole moral theory around this idea. But, unlike Clarke, for whom the basic moral notions are fitness and unfitness, Wollaston seeks to reduce moral notions to truth and falsehood. True propositions 'express things as they are'; actions in general express propositions and so may be true or false. Whatever can be understood has meaning and, since actions can be understood, they have meaning. Actions, as Wollaston conceives of them, are simply our thoughts translated into deeds. Since we represent things with our actions, what we represent may be true or false. Immoral actions deny things to be what they are and so express falsehoods. If I take your property, I falsely declare it is mine. If I am ungrateful, I falsely declare I never received favours from you. The evildoer 'lives a lie' because his conduct, as Wollaston puts it, breathes untruths. As moral creatures, our aim should be to mirror truth, representing the way things are in our conduct. But Wollaston thinks that the reality we should strive to represent is our ordinary, everyday natural and social world, not a metaphysically independent moral order, as Clarke had maintained.

Wollaston thinks that the Stoic formula - follow nature - is correct if interpreted as enjoining us to treat things as being what they are and so according to truth (see Stoicism §15). He also argues that 'the way to happiness, and the practice of reason' are the same as 'a careful observation of truth'. Our nature is such that we aim at our own happiness. Not only is happiness our natural good, but we also have a duty to strive for our own and other people's happiness. Anticipating Jeremy Bentham, Wollaston defines happiness as the 'true quantity of pleasure': pleasures and pains may be measured in terms of their intensity and duration, and we are happy when the sum total of pleasures exceeds the sum total of pains. He argues that just as happiness cannot be achieved by anything that interferes with morality (truth), so the practice of truth (acting morally) cannot make a person unhappy. Morality and happiness are congruent - if not in this world, then in the afterlife.

Wollaston thinks his theory explains what we mean when we say that morality is a matter of following right reason. He characterizes reason as the power to survey and compare ideas and to make inferences from these. We discover truth by means of reason or, more precisely, right reason. So when our actions are in accord with right reason, they express truths. Immoral actions are contrary to reason because they assert falsehoods. The reports of our senses are true and may be acted upon as long as there is no reason against them. It is the nature of reason to command, Wollaston says; it enjoins, forbids or permits, and as rational creatures it should govern us. True happiness can be achieved only by pursuing means consistent with our rational nature; false pleasures are inconsistent with or destructive of it.

Wollaston concludes that the 'truest' definition of morality is 'the pursuit of happiness by the practice of truth and reason'. And when we act morally, he argues, we obey God's will. In rejecting the way things are, the evildoer is implicitly rebelling against God's will. The truths we should aim to mirror in our actions are God's truths. They

are natural, however, because we are able to grasp them by reason unaided by divine revelation. So there is, Wollaston says, such a thing as natural religion (see Natural theology).

Wollaston's attempt to reduce moral categories to truth and falsehood was criticized by subsequent philosophers, starting with David Hume (1739/40) and continuing today with John Mackie (1980) and Joel Feinberg (1977) (see Hume, D.). The problem these philosophers found with Wollaston's idea that immoral actions express falsehoods is that as a criterion of wrong actions it is circular. It is wrong for me to take your property, Wollaston says, because I falsely declare it to be mine, not yours. But if we ask why this is what my action means, the answer is that the fact that it is yours *means* that I should not steal it. In every case, the truth that is supposedly denied by a wrong action already has moral or normative content.

In the latter part of *The Religion*, Wollaston details our specific duties to God, others and ourselves. Anticipating Price and Kant, he thinks many duties command us to respect our own and other people's rational nature. He argues against attempts to persuade others without giving them reasons. To do so is to attempt to treat others as if they did not have reason.

CHARLOTTE R. BROWN

List of works

Wollaston, W. (1691) *On the Design of Part of the Book of Ecclesiastes, or the Unreasonableness of Men's Restless Contention for the Present Enjoyments, Represented in an English Poem*, printed privately.(Contains Wollaston's poem in heroic couplets inspired by chapters 1-4 of the Ecclesiastes, plus a paraphrase of these chapters, with notes on the paraphrase.)

Wollaston, W. (1722) *The Religion of Nature Delineated*, printed privately; repr. London: Samuel Palmer, 1724. The 'Life of Wollaston' is prefixed to the 6th edn, London, 1738. 10 more editions by 1750.(Wollaston's philosophically important work; argues that as moral creatures we should aim to mirror truth, representing the way things are in our conduct.)

References and further reading

Clarke, S. (1706) *A Discourse concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation*, in *The Works of Samuel Clarke*, vol. 2, New York: Garland, 1978. (Clarke defends a rationalist view of ethics. He argues that we come to see that actions are fit or unfit by rationally grasping the eternal and necessary relations holding between actions and situations. He also claims that reason by itself both obliges and motivates us to do what is fit.)

Feinberg, J. (1977) 'Wollaston and his Critics', *Journal of the History of Ideas* 38: 345-52.(A defence of Wollaston against his critics and their misinterpretations, although he agrees to some extent with Hume's main objections.)

Hume, D. (1739/40) *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1960, esp. book 3, part I, section 1, 461-3, and footnote.(Hume opposes Wollaston's moral theory, although at points he misinterprets it.)

Joynton, O. (1984) 'The Problem of Circularity in Wollaston's Moral Philosophy', *Journal of the History of Philosophy* 22: 435-43.(A recent attempt to defend Wollaston from the charge that his criterion of right action is circular or empty.)

Mackie, J.L. (1980) *Hume's Moral Theory*, London: Routledge & Kegan Paul, ch. 2.(Contains a brief but clear explanation of Wollaston's moral theory.)

Tweyman, S. (1976) 'Truth, Happiness and Obligation: The Moral Philosophy of William Wollaston', *Philosophy* 51: 35-46.(A sympathetic explanation of Wollaston's moral theory.)

Wollstonecraft, Mary (1759-97)

Wollstonecraft used the rationalist and egalitarian ideas of late eighteenth-century radical liberalism to attack the subjugation of women and to display its roots in the social construction of gender. Her political philosophy draws on Rousseau's philosophical anthropology, rational religion, and an original moral psychology which integrates reason and feeling in the production of virtue. Relations between men and women are corrupted by artificial gender distinctions, just as political relations are corrupted by artificial distinctions of rank, wealth and power. Conventional, artificial morality distinguishes between male and female virtue; true virtue is gender-neutral, consists in the imitation of God, and depends on the unimpeded development of natural faculties common to both sexes, including both reason and passion. Political justice and private virtue are interdependent: neither can advance without an advance in the other.

1 Life and influences

Mary Wollstonecraft was born into a declining middle-class family. Her father became a heavy drinker who beat his wife, and possibly his daughter too. Wollstonecraft had little formal education and early sought independence as a lady's companion, a schoolkeeper, and a governess, before rejecting such conventionally female occupations for the usually male one of translator and reviewer for *The Analytical Review*, a periodical founded by the dissenting publisher Joseph Johnson. She achieved fame with two political tracts: *A Vindication of the Rights of Men* (1790), attacking Edmund Burke's *Reflections on the Revolution in France* (1790) and*A Vindication of the Rights of Woman* (1792), her most substantial work (see Burke, E.). She also published educational works, a novel, an account of the French Revolution, and a travel book; a second novel was unfinished at her death. All the work published in her lifetime was written for money, the two *Vindications* extremely rapidly, without revision. While living in France to study the Revolution of 1789 at first hand she had an illegitimate daughter; after return to London and rejection by the child's father she made two attempts at suicide. She then resumed her journalistic career and formed a liaison with the political philosopher William Godwin, whom she married on becoming pregnant with a second daughter (later to become Mary Shelley). She died of complications following childbirth. Wollstonecraft seems to have been influenced chiefly by the eighteenth-century culture of sensibility, the educational theories of Locke and Catherine Macaulay, the dissenting, 'enlightened', politically radical circles round Joseph Johnson and Richard Price, who lived near the school she kept in Newington Green, and by wide and miscellaneous reading, particularly of Rousseau and Burke, opposition to whom formed her understanding of her own experience.

2 Wollstonecraft and philosophy

Wollstonecraft was not a systematic political philosopher aiming primarily at theoretical rigour, but something more like a *philosophe* as defined in Diderot's *Encyclopedie* (see Diderot, D. §1) For the *philosophe*, reasoning consisted in speculative generalization from experience, applied to the social issues of the day with an eye to practical improvement. Wollstonecraft combined ideas drawn from a variety of Enlightenment philosophies, from contemporary science, and from prevailing political, cultural and social movements into an explanation and evaluation of the current condition of women, of the state of society, and of her personal and professional experience within it. Her object was a rational programme of reform; her political works use philosophy only incidentally, to support political polemic.

Though this political polemic uses the language of liberalism, its philosophical foundations are, first, a speculative anthropology which distinguishes between natural and artificial human attributes, assigns most of human development to environmental influences, and expects progressive human improvement in morality and civilization; and second, a moral philosophy and psychology inspired by personal experience, current theories of the nervous system, and the suggestions of rational religion about the purpose of human life. This philosophical moral psychology achieved only fragmentary expression in Wollstonecraft's political works, but its outlines are fairly clear from her educational, fictional and travel writings.

3 Political philosophy

For Wollstonecraft, human beings were created to perfect their nature as rational and moral beings (see

Perfectionism). Natural humanity is not unsocialized humanity, as in Rousseau, but humanity freely developing its capacity for self-improvement, which includes the capacity for socialization. The function of society is further improvement of humankind. Social and political arrangements which thwart or fail to recognize this corrupt human nature.

Eighteenth-century society and politics are corrupted by artificial distinctions of rank (such as the aristocracy), the products of historical accident, which impede the development of all ranks towards human perfection; relations between the sexes are similarly corrupted by artificial distinctions of character and capability, such as the association of reason and moral strength with men, and feeling or sensibility and moral weakness with women. These artificial distinctions impede the development of both sexes towards full flowering of their human faculties.

Such corrupted environmental influences produce social injustice, along with vice in the aristocracy and brutish insensibility in the poor; the best hope of virtue lies with the middle classes. But even among these, artificial rank and gender distinctions are maintained by factitious interests such as social snobbery, male interest in the reduction of women to docile sexual objects, and female enjoyment of the opportunity this affords for sexual tyranny over men.

Thus the effect of artificial distinctions on social relations and public and private morality is, in the public sphere, political tyranny, social enmity, and a distorted conventional morality assigning different virtues to the two sexes; in the private sphere, domestic tyranny and the degradation of both male and female nature, to the detriment of physical health, moral development and parenthood.

The remedy is the abolition of artificial distinctions through political and social reform and the education of potential future citizens of both sexes into a radically revised, gender-neutral morality, based on religious but rationally derived principles. This new morality will redescribe conventionally gendered and sexualized virtues such as courage and modesty in forms applicable to both sexes, enabling women to become independent moral agents and rational wives and mothers.

Moral re-education, however, will be effective only in a context of gradual institutional, cultural and political evolution towards a republican meritocracy. Political revolution without moral evolution is dangerous (though it may sometimes be necessary as the lesser evil), since individual and social moral development are interdependent; private virtue will be difficult and therefore rare unless supported by the appropriate social and political structures, which will in turn be unstable without private virtue. For instance, the destructive passions aroused by oppression are likely to wreak havoc if released from social and political control before development of the capacity for rational management of the passions in the oppressed. Hence the later excesses of the initially benign and rational French Revolution of 1789.

4 Moral philosophy

The purpose of human life is perfection of human faculties in the imitation of God, who is moral perfection; however human perfection can be achieved only beyond the grave. In this life imitation of God consists in virtue, that is, acting autonomously from moral principle derived by reason from the attributes of God. Moral principles are not specified, but the supreme principle seems to be universal benevolence, conjoined with or entailing a principle of justice; God’s benevolence can be inferred from the convenient arrangements for his creatures manifested in nature. Benevolence prompted by sympathy for particular cases is a juvenile precursor, perhaps a precondition, of rational, adult morality motivated by principled universal benevolence.

5 Moral psychology

Moral development towards principled action depends on adequate development of the appropriate faculties, namely imagination, passion and reason, whose joint function is to reveal the true end of human life. Human beings come into the world endowed with appetite and the capacity to develop these higher faculties, though capacities vary with sensibility and development may be impeded by mistaken upbringing. Appetite, whose function is to preserve the body, proposes merely sensual, achievable objects of desire; imagination proposes ideal objects which make possible desire for something beyond the physically attainable. Such desires spring from the passions (fear, anger, love and so on), whose function is to elevate the mind. Sensibility (nervous sensitivity) is, or is linked to, capacity for imagination and passion.

Reason is developed by reflecting on efforts, especially frustrated efforts, to satisfy the passions and appetites. The reason which recognizes the function of the appetites, and tempers their satisfaction accordingly, is prudence; temperately satisfied appetite is earthly or animal happiness. The reason which recognizes the function of the passions, whose natural outcome is discontent, identifies the true end of human life, namely moral perfection or likeness to God; it produces virtue by transforming passionate desire into desire for perfection as the only object which can satisfy it. Attainment of perfection is true human happiness, which is impossible in this life.

Since the development of virtue depends on reason's interaction with the appetites and passions, virtue is not served by premature or inappropriate attempts to curb or stifle these. Rather, social conventions and education which restrict the activities and aspirations of women prevent the development of both reason and virtue, thereby justifying and reinforcing conventional gender roles. Conventional female education, by encouraging attention to sensory attractions, and instilling outward propriety instead of virtue, discourages development of the higher functions of imagination and passion, thus depriving an already weakened reason of opportunities to acquire knowledge of the true end of human life, and hence adult virtue.

See also: Feminism; Feminist ethics; Morality and emotions

SUSAN KHIN ZAW

List of works

Wollstonecraft, M. (1989) *The Works of Mary Wollstonecraft*, ed. J. Todd and M. Butler, London: William Pickering, 7 vols.(Contains all of Wollstonecraft's works, which are listed individually below.)

Wollstonecraft, M. (1787) *Thoughts on the Education of Daughters: with Reflections on Female Conduct, in the More Important Duties of Life*, in *The Works of Mary Wollstonecraft*, London: William Pickering, vol. 4, 1989.(Wollestonecraft's first published work; a handbook on female education probably based on her own practical experience.)

Wollstonecraft, M.(1788) *Mary: a Fiction*, in *The Works of Mary Wollstonecraft*, London: William Pickering, vol. 1, 1989.(A novel of ideas written to show 'that a genius will educate itself'; illustrates Wollstonecraft's early view of the course of ideal moral development, of the disadvantages of marriage for women, and of what would constitute an acceptable female social role.)

Wollstonecraft, M. (1788) *Original Stories from Real Life: With Conversations Calculated to Regulate the Affections and Form the Mind to Truth and Goodness*, in *The Works of Mary Wollstonecraft*, London: William Pickering, vol. 4, 1989.(A children's storybook carefully designed to promote moral development; the most detailed working-out of Wollstonecraft's views on moral education.)

Wollstonecraft, M. (1790) *A Vindication of the Rights of Men, in a Letter to the Right Honourable Edmund Burke*, in *The Works of Mary Wollstonecraft*, London: William Pickering, vol. 5, 1989.(An impassioned defence, by appeal to a rationalist as opposed to a sentimentalist moral philosophy, of Price's welcome of the French Revolution against Burke's attack; lays the foundation for Wollstonecraft's feminist politics.)

Wollstonecraft, M. (1792) *A Vindication of the Rights of Woman with Strictures on Moral and Political Subjects*, in *The Works of Mary Wollstonecraft*, London: William Pickering, vol. 5, 1989.(Attacks conventional female education, morality and gender role as presented by various contemporary authors of educational and conduct books, principally by Rousseau in Émile.)

Wollstonecraft, M. (1794) *An Historical and Moral View of the Origin and Progress of the French Revolution; and the Effect it has produced within Europe*, in *The Works of Mary Wollstonecraft*, London: William Pickering, vol. 6, 1989.(Wollstonecraft's reaction to the increasing violence of the French Revolution; historically derivative but useful for signs of development in her political thought.)

Wollstonecraft, M. (1796) *Letters Written During a Short Residence in Sweden, Norway, and Denmark*, in *The Works of Mary Wollstonecraft*, London: William Pickering, vol. 6, 1989.(Provide probably the most accessible introduction to Wollstonecraft's thought; they suggest a movement away from rationalism towards a greater interest in imagination and sensibility under the impact of personal experience.)

Wollstonecraft, M. (1798) *The Wrongs of Woman, or Maria*, in W. Godwin (ed.) *Posthumous Works of the Author of A Vindication of the Rights of Woman*, in *The Works of Mary Wollstonecraft*, London: William Pickering, vol. 1, 1989.(Unfinished; her first engagement with the wrongs of lower-class women; suggests that her earlier political and moral ideas, particularly on sensibility, were in a state of constant flux.)

Wollstonecraft, M. (1979) *Collected Letters of Mary Wollstonecraft*, ed. R. Wardle, Ithaca, NY: Cornell University Press.(Useful for following the development of Wollstonecraft's self-conception; also one of the few sources of knowledge of what she read.)

References and further reading

Barker-Benfield, G.J. (1992) *The Culture of Sensibility: Sex and Society in Eighteenth-Century Britain*, Chicago, IL and London: University of Chicago Press.(A sociocultural study of a key concept in Wollstonecraft's moral psychology.)

Blair, H. (1783) *Lectures on Rhetoric and Belles-Lettres*, London: A. Strahan, T. Cadell, W. Creech, 1787.(A handbook of style summarizing Enlightened sentimentalist views on discourse; described by Wollstonecraft as 'an intellectual feast' (Letters: 138).)

Burke, E. (1757) *A Philosophical Enquiry into the Origin of our Ideas of the Sublime and Beautiful*, ed. J. Boulton, London: Routledge & Kegan Paul, 1958.(Burke's gendered aesthetic theory, partly an influence on Wollstonecraft and partly her target.)

Burke, E. (1790) *Reflections on the Revolution in France*, ed. C.C. O'Brien, Harmondsworth: Penguin, 1968.(A furious response to Richard Price's A Discourse on the Love of our Country (1790, 3rd edn); defends British political traditions and correctly predicts some of the disastrous developments which followed. Criticized in Wollstonecraft's first Vindication.)

Coole, D.H. (1988) *Women in Political Theory*, Sussex: Wheatsheaf Books, Boulder, CO: Lynne Rienner Publishers, ch. 5.(Places Wollstonecraft as a political theorist; good on the relationship with Rousseau.)

Khin Zaw, S. (1994) '''Appealing to the Head *and* Heart": Wollstonecraft and Burke on Taste, Morals and Human Nature', in G. Perry and M. Rossington (eds) *Femininity and Masculinity in Eighteenth Century Art and Culture*, Manchester and New York: Manchester University Press.(Analyses Wollstonecraft's philosophical debt to Burke in her Rights of Men.)

Locke, J. (1693) *Thoughts on Education, in The Educational Writings of John Locke*, ed. J.L. Axtell, Cambridge: Cambridge University Press, 1968.

Macaulay, C. (1790) *Letters on Education, with Observations on Religious and Metaphysical Subjects*, London. (Conduct literature grounded in contemporary philosophy; generously credited in Wollstonecraft's second Vindication.)

Rousseau, J.-J. (1762a) *Du Contrat social*, trans. and ed. D.A. Cress, with introduction by P. Gay, *On the Social Contract*, Indianapolis, IN: Hackett Publishing Company, 1987.(Prime source of French Revolutionary political theory.)

Rousseau, J.-J. (1762b) *Émile: ou, de l'éducation*, trans. A. Bloom, *Emile: or, On Education*, Harmondsworth: Penguin, 1991.(Rousseau's theory of gendered education; praised and attacked by Wollstonecraft in her second Vindication.)

Sapiro, V. (1992) *A Vindication of Political Virtue: the Political Theory of Mary Wollstonecraft*, Chicago, IL and London: University of Chicago Press.(The first book-length study of Wollstonecraft's philosophy; generally reliable.)

Wônch’ūk (613-96)

Wônch’ūk, a Korean monk-scholar, was head of the Ximing Monastery in Tang China. Neglected by history, research has now recovered this prolific writer, whose commentaries on Yogācāra texts influenced later Buddhist scholars in China and also in Tibet, notably Tsong kha pa.

The *Prajñāpāramitā-sūtras* and the *Samdhinirmocana-sūtra* are scriptures of Mādhyamika and Yogācāra Buddhism respectively. Xuanzang (596-664) introduced Yogācāra to China after having studied at NālandāUniversity, the bastion of Yogācāra in India. Returning to China with voluminous Sanskrit manuscripts, he presided over the Imperial translation project at Changan.

Wônch’ūk, Xuanzang’s colleague and critic, wrote an extensive commentary on the *Samdhinirmocana-sūtra*, ‘Explication of the Underlying Meaning’ which is extant today. One of the main points at issue is that while Mādhyamika declares that ‘dependently co-originated things’ do not have an unchanging intrinsic nature at all, Yogācāra insists on some real basis for these occurrences. Citing the theory of ‘three natures’ (*svabhāvatraya*), Xuanzang’s disciple Ji (632-82) simply asserts the superiority of the *Samdhinirmocana-sūtra* over the *Prajñāpāramitā-sūtras*’ ‘two truths’ (*satya dvaya*) theory. Wônch’ūk, however, unfolds a more careful investigation by claiming that Mādhyamika and Yogācāra are not competitors, but are in fact complementary schools. The *Prajñāpāramitā-sūtras*, for example, also contain the ‘three natures’ theory.

Wônch’ūk had an influence on Tibetan Buddhism as well. In the ninth century, Chos grub was ordered by a Tibetan king to translate Wônch’ūk’s commentary on the *Samdhinirmocana-sūtra* into Tibetan. Later, Tsong kha pa (1357-1419) followed Wônch’ūk’s methodology in his magnum opus *Drang nges legs bshad snying po (The Essence of True Eloquence)*. Here he quotes Wônch’ūk’s commentary on the *Samdhinirmocana-sūtra* extensively.

Yogācāra provides a religio-psychological analysis of the structure of mental activity by charting the path between consciousness and the dynamic unconscious, that is, between *manas* and *ālaya-vijñāna* (seventh and eighth consciousness) (see Buddhism, Yogācāra School of). Paramārtha (499-569) had posited the ninth consciousness as pure cognition. Wônch’ūk refutes this view in detail in his commentary on the *Samdhinirmocana-sūtra*. Tsong kha pa in turn uses the same argument critically in his work, *Yid dang kun Gzhi*.

Ji forced Xuanzang to make a selective translation entitled *Ren weishi lun (Treatise on the Proof of Representation Only)*, gleaned from the Indian commentaries of the *Trim’sikā (Thirty Verses)*. Although Wônch’ūk, along with the other disciples, was excluded from this project, he wrote his own commentary on this work, expressing many views divergent from those of Ji. Some pages of this lost work can be recovered from the rebuttals of Wônch’ūk’s commentaries made by Ji’s disciples.

While Ji became the systematizer of the Faxiang or *Yuishiki Hossō* school in Japan, Wônch’ūk’s fate was bleak. Even so, he was revered ‘as a living Buddha’ by the Empress Wu, who refused the request of a Silla king to send him home. But in the *Biographies of the Eminent Monks of the Song*, Wônch’ūk is described as being of obscure origin and of clever but dubious character. The truth is suggested by the following eulogy by his country man, Ch’oe:

Which ever translation of a *sūtra* it might be, he always
presided over that endeavour:
What ever commentary it might be, he always had something
unique to say:
Whenever he sought seclusion, he always attained inspiration:
To whomever he gave lectures on the Dharma,
he always reached everyone.

See also: Buddhism, Mādhyamika: India and Tibet; Buddhism, Yogācāra School of; Buddhist philosophy, Chinese; Tsong kha pa

SHOTARO IIDA

References and further reading

Iida Shotaro (1986) ‘Who best can turn the Dharma-cakra? A controversy between Wônch’ūk and Kuei-chi’, *Journal of Indian and Buddhist Studies* 34 (2): 11-18.(On Wônch’ūk ’s thought.)

Lancaster, L.R. and Yu, C.S. (eds) (1991) *Assimilation of Buddhism in Korea*, Berkeley, CA: Asian Humanities Press.(On the arrival of Buddhism in Korea.)

Paul, D.V. (1984) *Philosophy of Mind in Sixth-century China: Paramārtha’s ‘Evolution of Consciousness’*, Stanford, CA: Stanford University Press.(Looks at the background of Korean Buddhism.)

Thurman, R.A.F. (1984) *Tsong Khapa’s Speech of Gold in the Essence of True Eloquences: Reason and Enlightenment in the Central Philosophy of Tibet*, Princeton, NJ: Princeton University Press.(Translation with an introduction.)

Wônhyo (617-86)

Wônhyo is one of the most important figures in Korean Buddhism, and a significant influence on the development of East Asian Buddhism in general. His lifework was the reconciliation of ideological conflicts among the various Buddhist schools. His goal was to create an all-inclusive, non-sectarian Buddhist doctrine. To do this he utilized the all-embracing, systematic metaphysics of the Hwaôm school of Buddhism, deriving both a guiding theoretical principle - hwajaeng *or 'the harmonization of all disputes' - and a powerful dialectical method for the examination of Buddhist doctrinal conflicts.*

Born in the Korean village of Puljich'on, not far from present-day Kyôngju, Wônhyo became a Buddhist monk at an early age and was a prolific writer, composing over one hundred books in 240 rolls. However, only twenty-three of these works survive, complete or in fragments. The most well-known of these are *Shimmun hwajaengnon (The Treatise on the Harmonization of all Disputes in Ten Chapters)*, *Daesûng kishillonso (The Commentaries on the Awakening of Faith in the Mahāyāna)*, *Kûmgang sammaegyông (The Treatise on the Vajrasamādhi-sūtra)* and *Yôlban'gyông jong'yo (The Essentials on the Nirvāna-sūtra).*

Wônhyo's dialectical method is based on the interrelation of two pairs of concepts, 'doctrine' (*chong*) and 'essence' (*yo*), as well as what he called the 'opening' (*kae*) and 'sealing' (*hap*) of Buddhist truth. Using these four key terms, he was able to place the various Buddhist ideologies of his day within a larger conceptual structure, demonstrating how points of conflict on the level of doctrine could be reconciled on the level of essence.

To Wônhyo, the term 'doctrine' referred to the expression of Buddhism's essentially ineffable, unitary truth through the differing conceptual vocabularies of the various Buddhist schools. When this unitary truth was 'opened' through philosophical discourse it became doctrine, accessible to the intellect for study and contemplation. Yet this 'opening of the one into the many', as Wônhyo termed it, was also the source of conflict and disagreement, for doctrinal truth was inherently self-limiting, defined by its own means of expression and therefore partial and biased.

In contrast, Wônhyo used the term 'essence' to refer back to the ineffable truth underlying all Buddhist ideology. He believed that doctrinal truth could be 'sealed' - that is, converted back to its unitary essence - beyond the scope of language or reason and therefore free of the limited perspective imposed by doctrine. He called this return to essence 'the sealing of the many into the one', and through it he sought to reconcile the various schools of Buddhist thought into a single, syncretic vision: *t'ong pulgyo*, 'the Buddhism of total interpenetration'.

The search for *t'ong pulgyo* has continued among Buddhist thinkers ever since, but Wônhyo's influence remains indelible. By combining the concepts of doctrine and essence with those of opening and sealing, he created a powerful interpretive tool for the examination of Buddhist texts, one that allowed him to apply the principle of *hwajaeng*, or 'harmonization', to a wide variety of theoretical conflicts. In so doing, he created the conceptual framework within which all subsequent discussion of *t'ong pulgyo* has taken place, up to and including the present day.

See also: Buddhist philosophy, Korean

SUNG BAE PARK

List of works

Wônhyo (617-86) *Collected works*, in *Han'guk pulgyo chônsô (The Collected works of Korean Buddhism)*, Seoul: Dongguk University Press, 1979, vol. 1, 480-843. (Contains all the extant works of Wônhyo, a total of twenty-three texts in Sino-Korean.)

Wônhyo (617-86) *Shimmun hwajaengnon (The Treatise on the Harmonization of all Disputes in Ten Chapters)*, in *Han'guk pulgyo chônsô*, Seoul: Dongguk University Press, 1979, vol. 1, 838-40.(The most well-known of Wônhyo's works, the theme of which is the harmonization of all scholastic disputes among different schools and texts. Only a few fragments remain.)

Wônhyo (617-86) *Daesûng kishillonso (The Commentaries on the Awakening of Faith in the Mahāyāna)*, in *Han'guk pulgyo chônsô*, Seoul: Dongguk University Press, 1979, vol. 1, 698-732.(Wônhyo's philosophical introduction to the *Awakening of Faith* along with a running commentary to the text.)

Wônhyo (617-86) *Kûmgang sammaegyông (The Treatise on the Vajrasamādhi-sūtra)*, in *Han'guk pulgyo chônsô*, Seoul: Dongguk University Press, 1979, vol. 1, 604-77.(Wônhyo's philosophical introduction to the text and a running commentary. The relationship between original enlightenment and and initial enlightenment is seriously discussed. The text is seen to have directly influenced Chan Buddhism in East Asia.)

Wônhyo (617-86) *Yôlban'gyông jong'yo (The Essentials on the Nirvāna-sūtra)*, in *Han'guk pulgyo chônsô*, Seoul: Dongguk University Press, 1979, vol. 1, 524-47.(Comments on the *sūtra* with comprehensive overall analysis; there is an extensive discussion of the meaning of *nirvāna*.)

References and further reading

Koh Ikchim (1989) *Han'guk Kodae Pulkyo sasangsa (A History of Buddhist Thought Before the Koryo Dynasty)*, Seoul: Dongguk University Press.(A collection of essays discussing pre-Koryo Dynasty Buddhist philosophy.)

Kim Young-tae (1986) *Introduction to the Buddhist History of Korea*, Seoul: Kyôngsô wôn.(A chronological history of Korean Buddhism.)

Lee, P.H. (ed.) (1993) *Sourcebook of Korean Civilization*, vol. 1, *From Early Times to the Sixteenth Century*, New York: Columbia University Press.(Contains selected English translations of important Korean texts reflecting the culture and civilization of Korea, including excerpts from Wônhyo on pages 145-59.)

Park Chong-hong (1976) 'Wônhyo ûi chôrhak sasang', in *Han'guk sasang sa (A History of Korean Thought)*, edited by the Han'guk sasang yôn'guhoe (Association for the Study of Korean Thought), Seoul: Ilshinsa, 59-88.(A philosophical approach to the logical structure of Wônhyo's soteriological system.)

Park Sung Bae (1983) *Buddhist Faith and Sudden Enlightenment*, Albany, NY: State University of New York Press.(A thematic approach to the relationship between patriarchal faith and sudden enlightenment in the actual practice of Buddhist meditation.)

Park Sung Bae (forthcoming) *Wônhyo's Commentary on the 'Treatise on the Awakening of Mahāyāna Faith'*, Albany, NY: State University of New York Press.(A translation and exegesis of Wônhyo's commentaries on the *Awakening of Faith in the Mahāyāna*.)

Work, philosophy of

Unlike play, work is activity that has to involve significant expenditure of effort and be directed toward some goal beyond enjoyment. The term 'work' is also used to signify an individual's occupation, the means whereby they gain their livelihood. In modern market economies individuals contract to work for other individuals on specified terms. Beyond noting this formal freedom to choose how one shall work, critics of market economies have maintained that one's occupation should be a realm of substantive freedom, in which work is freely chosen self-expression. Against this unalienated labour norm, others have held that the freedom of self-expression is one good among others that work can provide, such as lucrative pay, friendly social contact and the satisfaction of the self-support norm, and that none of these various work-related goods necessarily should have priority over others. Some philosophers place responsibility on society for providing opportunities for good work for all members of society; others hold that the responsibility for the quality of one's occupational life appropriately falls on each individual alone. Finally, some theorists of work emphasize that performance of hard work renders one deserving of property ownership (John Locke) or enhances one's spiritual development (Mahatma Gandhi).

1 Work versus play; work versus leisure

Work is effort directed at some goal other than enjoyment taken in the experience of putting forth the effort. In contrast, play is effort that is not aimed at any goal beyond the enjoyable experience of that very effort. Work and play seem to differ further in that a threshold level of effort must be put forth before an activity is properly called work, whereas play either involves a much lower threshold of effort or none at all. Tying my shoelaces is not typically work, even though the activity is goal-directed, because the effort involved is too slight. Effortlessly twiddling my shoelaces for fun qualifies as play even though hardly any effort is expended. Another example would be a spontaneous expression of joy by dancing a jig or of sorrow by singing a sad song. These expressive activities are not aimed at enjoyment, rather at expression of emotion, hence they do not qualify as play. But if effortless or close to effortless, they do not qualify as work either. Another example of an activity that is neither work nor play is lying on a beach in order to get a suntan. Again the effort constraint seems to be operative. Of course, if sunbathing can be described as involving expenditure of effort, say, because one's body is becoming hot or one is bored and one must actively overcome the desire to leave the beach by an effort of will, then in these circumstances sunbathing can be characterized as work.

In the Book of Genesis, God punishes Adam for sin by condemning him to unpleasant labour for his daily bread: 'By the sweat of your brow shall you live'. In order to meet their basic needs many, perhaps most, people throughout history have had to work in ways that are onerous and unpleasant, recalling Adam's curse. But work can give intrinsic satisfaction without ceasing to be work. Of persons engaged in work one can ask: (1) would they prefer to gain the goal they are seeking without working for it (if that is possible)?; and (2) would they wish to engage in the work activity for its enjoyment even if the goal they are seeking could be attained without the effort of work? An affirmative answer to the second question indicates that the work has the quality of a fusion of work and play. A negative answer to the first question indicates working is partly valued for its own sake, not just as a means to a further goal.

Work in the sense of goal-directed activity is opposed to play. In another sense of the term, work refers to what persons do in order to gain their livelihood. Time spent at work in this sense - work as occupation - is distinct from time spent at leisure - the free time in which individuals may do as they please and which is not spent at one's occupation. A day in the life of an individual may be divided into time spent at one's occupation, time spent in nightly rest, time spent in nonoccupational necessary drudgery such as household chores, and leisure or free time. This division is not exhaustive, however, because hours spent at childrearing do not comfortably fit into any of the categories just listed.

2 Work as occupation

With the rise of market economies work as occupation has become organized primarily by means of voluntary contracts among individuals who are not deemed to be under prior social obligation to engage in economic cooperation. Owners and managers of economic firms hire individuals to work as employees on specified terms. The tendency of market economies has been to generate greatly increased specialization of economic function and

task segmentation compared to earlier modes of economic organization. In the late eighteenth century Adam Smith analysed this process of division of labour and its economic consequences. He described a pin factory in which the making of pins was broken down into many small separate tasks, each small task performed over and over by an individual worker. This division of labour increases the productivity of labour; more is produced with less human labour input. Smith expressed the concern that repeated performance of a small boring task would ultimately dull the mental faculties of the workers.

Towards the middle of the nineteenth century Karl Marx (§4) propounded an ideal of unalienated labour and criticized the capitalist market economy of the day for subjecting ordinary workers, the bulk of the population, to grim lives that revolve around alienated labour. A capitalist market economy is one in which owners of capital establish firms that hire propertyless workers to produce goods for sale. In Marx's conception there are four aspects to alienated labour. One can be alienated from one's own activity of working, from the product one creates, from one's fellow human beings at work, and from one's true human nature. According to Marx, the worker in a capitalist factory is the epitome of an alienated labourer. A paradigm example of unalienated labour would be that of a creative artist, one who enjoys the freedom to produce art in one's own way and exercises this freedom in ways that are satisfying, controls the disposition of the art works so created, produces in such a way as to establish community with those directly affected by these economic transactions, and is motivated to create art by the desire to serve the human community (Alienation §§3-5).

The unalienated labour ideal raises many questions. One is whether fulfilment at work should be made available to all members of society as a matter of justice, or is better seen as a prize for which all may compete but which only especially meritorious individuals will win. If the norm of reducing alienated labour is deployed to criticize the institutions of society, the question arises how one should measure the progress of society in this respect. (For simplicity, this discussion assumes the possibility of precise measurement.) One might rate society by the average level of unalienated labour among those in the workforce, or one might rate society by the level of unalienated labour enjoyed by the member of society who is worst off in this respect, or by some other standard. In societies as we know them, access to desirable conditions of work is unevenly distributed between men and women and between members of favoured and disfavoured racial, ethnic and religious groups. There are the questions of how to characterize the good of desirable working conditions and what would count as a fair distribution of this good.

Jobs can evidently score higher or lower on the different dimensions of the unalienated labour norm, and one may wonder which dimensions are ethically more significant. If my occupation consists of menial work such as licking stamps for a worthy cause to which I am dedicated, my work life may be on the whole very meaningful, even if the work activity itself is unsatisfying. Alternatively, an individual's occupation might consist of challenging and interesting tasks performed for an enterprise that manufactures shoddy products of little true social utility. If one's workplace is a cooperative owned and managed by its workers, employment provides the opportunity for democratic participation in policy formation. A job that is regimented and unpleasant might provide high pay, which gives the jobholder freedom to carry out any of a wide variety of richly fulfilling plans of life. Jobs provide varied packages of benefits and burdens.

Work as occupation can provide benefits other than those that figure in the unalienated labour ideal. Lucrative pay, mentioned above, is one. Where most people subscribe to the social norm that each able-bodied adult should be self-supporting, supporting oneself by holding a job, even an undesirable job, earns the respect of other people and bolsters one's own self-esteem. There is a flip side to the coin of social recognition through jobholding, however. Since jobs vary greatly in prestige, holding an unprestigious job confers low status that can erode the jobholder's self-esteem. Another potential benefit from jobholding is that a regular job gives structure to one's life and organizes one's daily schedule. No doubt fully autonomous individuals could give a sensible shape to their lives without this external prop, but many of us are less than fully autonomous, and benefit from some degree of imposed order. Also, most jobs provide regular social contacts with workmates and perhaps with customers of one's firm. These contacts can be pleasant or abrasive, but usually there is at least the opportunity for friendly human interchange, and sometimes opportunities develop for deeper bonds of solidarity, friendship and mutual aid.

Regarding work as occupation, philosophers have argued for opposed views of the obligation of society. One is that the responsibility for each individual's work life belongs to that person alone; there is no positive obligation of society to ensure good work for all. A second view is that opportunities for decent employment should be made

available to all able-bodied adults. A third view is that the obligation of society is to provide each individual, so far as is feasible, freedom from the necessity to work by guaranteeing a basic minimum income to all individuals whether or not they participate in the workforce. The contrast between the second and third views reflects disagreement as to whether the necessity of working is a benefit or a burden and disagreement as to the validity of the social norm that each able-bodied adult is obligated to work for one's own livelihood.

3 Other perspectives

Philosophers have espoused perspectives on work different in kind from any of the views canvassed to this point. The seventeenth-century English philosopher John Locke emphasizes work as conferring entitlement to ownership of natural resources, particularly land, under appropriate conditions. According to Locke, an individual acquires ownership of nonscarce unowned land by labouring on it with the intent to establish ownership, and one sustains the right of ownership by continuing to work the land productively or by arranging for its productive use. Hard work renders an individual deserving of good fortune. Locke maintains that a more complex version of this story still holds true when land and other natural resources become scarce. The twentieth-century leader of the movement for the emancipation of India from colonial rule, Mahatma Gandhi, similarly holds that hard work for one's livelihood is a moral obligation and that working hard renders one deserving and morally virtuous. But in contrast to Locke, Gandhi upholds a norm of voluntary poverty and asceticism and recommends the discipline of hard work for its enhancement of the spiritual growth of the worker. In this spirit Gandhi urges the Indian people, and by implication all of us, to eschew modern industrial civilization and to earn our livelihood by hard work at pre-industrial, traditional forms of craft work.

RICHARD ARNESON

References and further reading

Arneson, R.J. (1990) 'Is Work Special? Justice and the Distribution of Employment', *American Political Science Review* 84 (4): 1127-47.(Argues that society ought to provide opportunities for decent employment to all its members.)

Elster, J. (1988) 'Is There (or Should There Be) a Right to Work?', in Amy Gutmann (ed.) *Democracy and the Welfare State*, Princeton, NJ: Princeton University Press.(Argues against the claim that there should be a right to work in the sense of a right on the part of each individual in society to guaranteed employment opportunities.)

Gandhi, M. (1966) *An Autobiography: The Story of My Experiments with Truth*, trans. M. Desai, Boston, MA: Beacon Press.(Outlines Gandhi's ascetic philosophy and celebrates pre-industrial craft work.)

Locke, J. (1690) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1960.(The chapter 'Of Property' in the *Second Treatise* is the classical statement of a liberal view of the connection between work and entitlement to property ownership in the context of a defence of the market economy.)

Marx, K. (1844) 'Economic and Philosophic Manuscripts of 1844', in Robert Tucker (ed.) *The Marx-Engels Reader*, New York: W.W. Norton & Co.(Outlines the ideal of unalienated labour and criticizes capitalist market economies for subjecting workers to alienated labour.)

Nozick, R. (1974) *Anarchy, State, and Utopia*, New York: Basic Books.(Expresses a libertarian view that repudiates any positive obligations on the part of society to provide desirable work opportunities for members of society.)

Parijs, P. van (1995) *Real Freedom for All*, Oxford: Oxford University Press.(Argues that so far as is feasible, society is obligated to guarantee freedom from the necessity of working to all its members.)

Walzer, M. (1983) *Spheres of Justice: A Defense of Pluralism and Equality*, New York: Basic Books.(Chapter 6 considers how society ought to distribute hard - in the sense of oppressive - work.)

Wróblewski, Jerzy (1926-90)

Jerzy Wróblewski was a leading representative of analytical legal theory in Poland in the second half of the twentieth century. Leon Petrażycki and the school of logical thought of Lwów and Warsaw provided his background inspiration. His approach to legal theory and legal science belongs, in philosophical terms, to minimalism, relativism and moderate reconstructivism.

Born in Vilnius, Wróblewski became involved as a partisan in guerrilla warfare during the German occupation of Poland, 1939-45, and was able to complete his education, taking a doctorate at the Jagiellonian University of Cracow, only after the defeat of Germany and the restoration of government in Poland. Immediately on graduation, he moved to Lódz to take up the chair in the theory of state and law there, remaining at that university (and serving a term as its Rector in the troubled years around 1980) until his death.

Wróblewski subdivides the tasks of legal theory into seven problem areas: the theory of the legal norm, the theory of interpretation of law, the theory of legal system, the theory of legal interpretation, the theory of the application of law, the theory of law-making, and the methodology of legal sciences. Norms he considered in linguistic terms, as part of natural language, giving a pragmatic analysis in terms of the speech acts from which they originate and through which they are interpreted and applied (see Norms, legal). His theory of legal system was Kelsenian in background, though he rejected transcendental reasoning and disputed the conception of legal order as purely dynamic, involving no possibility of logical derivation of conclusions from normative statements.

His most distinctive work was on the application of law, understood as the process of implementation of a legal norm by a state organ. His approach is marked by a painstaking and detailed analysis. He proposed what he called a ‘decisional’, as opposed to an informational or a functional, model of law-application. He analysed the problem of truth in judicial fact-finding on the basis of evidence, paying special attention to the variety of evidentiary facts presented to courts. In relation to the application of law, he disputed a narrow positivism that treats enacted law as sufficient in itself to control decision-making, and he argued for the necessity of referring to evaluations and values in this process; relevant sets of values he called the ‘ideology of application of the law’.

His interest in the rationality of legal processes extended to a concern with rationality in law-making and in the application of principles of legislative policy, using an instrumental conception of rationality based on the relationship between means and ends. As for methodology, he differentiated one-dimensional from multi-dimensional models of legal science. The former treat law exclusively as norm or as psychological fact or as social fact; the latter (favoured by Wróblewski) treat law as an ontologically and epistemologically complex phenomenon, and therefore approach the study of law using a variety of methods, techniques and conceptual apparatus derived from different social and mental sciences.

See also: Kelsen, H.; Law, philosophy of; Legal reasoning and interpretation; Petrażycki, L.

MAREK ZIRK-SADOWSKI

List of works

Wróblewski, J. (1992) *The Judicial Application of Law*, ed. Z. Bankowski and N. MacCormick, Dordrecht, Boston, MA, and London: Kluwer.(This is Wróblewski’s major work available in English, giving the main elements of the points discussed above, and exhibiting the author’s logical as well as analytical proclivities. It contains a substantial bibliography of Wróblewski’s main works, in the various languages in which he was competent.)

References and further reading

MacCormick, D.N. and Summers, R.S. (1991) *Interpreting Statutes: A Comparative Study*, Aldershot: Dartmouth.(This is a collection of papers by several authors on interpretation of statutes in a number of legal systems. Wróblewski contributed the chapter on interpretation in Poland, and had a decisive hand in helping to establish a common framework and methodology for the comparative study of interpretation. See in particular chapter 2, which he co-authored but chiefly inspired.)

Wundt, Wilhelm (1832-1920)

The German philosopher, psychologist and physician Wilhelm Wundt founded the world's first psychological laboratory in Leipzig in 1879 - at a time when psychology was still generally regarded as a theoretical and institutional part of philosophy. This event typified his life's work and its reception in many respects. On the one hand Wundt tried to develop psychology as an independent science by defining its subject matter and methodology; on the other, he wanted to integrate psychology into the context of philosophy, cultural theory and history. With both attempts he acquired world fame and at the same time became a most controversial figure. Systematizing his approach, Wundt worked on a great amount of material in very different disciplines. He has been called the last philosophical 'polyhistor' in the tradition of Leibniz and Hegel, as well as the first modern scientist in psychology.

1 Life

Wundt was born in 1832 in Neckarau, near Mannheim, and died in 1920 in Großboten, near Leipzig. He studied medicine at the universities of Tübingen, Heidelberg and Berlin, and began his academic career as a research assistant of Johannes Müller and Emil du Bois-Reymond in Berlin. From 1857 to 1864 he was *Privatdozent* at the Physiological Institute in Heidelberg directed by Hermann von Helmholtz. Rejecting Helmholtz's materialism, Wundt's own physiological studies shifted to the theory of sense perception, which he conceived of as a deeply philosophical question. Furthermore, he began to lecture on anthropology, ethnography and natural history. During these Heidelberg years, Wundt changed from physiologist to psychologist, yet always orientated himself within a philosophical framework. After a professorship in inductive philosophy at Zürich, in 1875 he accepted a professorship in Leipzig, where he founded his famous *Institut für Experimentelle Psychologie*. Beginning as a purely private institute, it later became the point of departure for many international students who returned home to establish similar institutions. Wundt taught at Leipzig until his death, continuously developing and changing his system in a spirit of encyclopedic scholarship. Over the last twenty years of his life he worked out his *Völkerpsychologie*. As an old man he became the rector of Leipzig University and was accorded tremendous national and international recognition.

2 Philosophy, psychology and science

With Wundt, the psychological zeitgeist shifted from philosophy to science, yet, even as a scientist, Wundt never left philosophical ground. Beside his seminal studies in physiological psychology, he lectured and wrote on logic, ethics, epistemology and other traditional philosophical topics, drawing especially on Leibniz and Hegel, but also on Mill's methodology (see Mill, J.S.). It has often been overlooked that for Wundt himself 'experimental psychology' was only a particular way to tackle traditional philosophical issues - and to engender new ones. If he was himself an experimenter at heart, as Robinson notes, then he was surely the 'first philosophical psychologist spawned by experimental science' (1982: 129).

Although his *Physiological Psychology* was taken above all as a programme for experimental research, it also represented an attempt to resolve the mind-body problem. Investigating both physiological and mental phenomena, Wundt's basic intention was twofold: he wanted to overcome the traditional Cartesian dualism of *res extensa* and *res cogitans* without falling into materialistic monism. As he saw it, psychology studies particular events which cannot be reduced to a class of mere bodily events. Both physical and psychical processes exist 'side by side' but must conform to their own epistemological principles. Rather, historically and systematically, psychology draws upon these different approaches; its subject matter emerges at the point of contact of distinct epistemic spheres. Thus, Wundt's idea of psychophysical parallelism was a heuristic principle, not an objective law. The main distinction was not between mind and body, but between reasons and causes, that is, between forms of explanation adequate for different categories of events.

In calling the new discipline of this border region 'experimental psychology', Wundt wanted to place a major emphasis on the methodological characteristics by which he distinguished his approach from the traditional science of mind, which was based above all on introspection. Rejecting any idea of the mind as a separate spiritual or mental substance, Wundt believed it to be the totality of conscious experiences (and its partly unconscious emergence) at a given moment. He called this the 'actuality' of the mind.

Experiences, ideas and thoughts are active events - neither simple reactions nor things or impressions appearing in and disappearing from consciousness (as associationists and structuralists claimed). Will is the fundamental principle of mental activities. But there is no will without emotion. Inextricably interwoven, both intention and feeling organize the meaningful syntheses of all psychological events. Thus, Wundt called his conception of the mind 'voluntaristic', underlining the peculiar human quality of actively creating meanings.

Wundt designated the idea of a conscious focus of attention, which synthesizes the elements of experience and thought into larger units (*Gesamtvorstellungen*), the 'principle of creative resultants or synthesis'. This is closely linked to another key concept in Wundt's system: the concept of 'apperception'. Deriving from Leibniz and Kant, 'apperception' referred to the active synthesizing function of consciousness. Hence, it represented a central feature of higher mental activities, as it is responsible for all conscious constructions or wholes, such as thought, language, reasoning and judgment.

In adapting the methods of empirical, experimental and historical sciences to the questions of philosophy, Wundt wanted to create an integrated discipline of the mind. This was the content of his 'Heidelberg programme' of psychology which he only partly realized in the second half of his life in Leipzig (*Graumann 1980*). This programme consisted of: (1) the experimental study of human consciousness or mind; (2) the evolutionary history of mind; (3) comparative psychology (particularly dedicated to the analysis of ontogenetic and phylogenetic development).

Wundt's theory of the mind was a universal developmental theory. To understand the strong historic-genetical dimension of this conception (which for many decades had been widely ignored), one must take into account the influence of the two cultures Wundt wanted to fuse together: for the philosopher Wundt, the emphasis on cultural-historical development was a typical legacy of most German nineteenth-century philosophy since Hegel, Herder and the Enlightenment; for the physiologist and physician, the crucial event of those years - as it was for all life scientists - was the publication of Darwin's *Origin of Species* in 1859.

3 *Völkerpsychologie*

Wundt considered experimental psychology only as an elementary approach to the study of the mind. He rejected the idea that higher mental processes could be represented by a kind of mechanics of the mind parallel to physical mechanics. Physical causality (as investigated, for example, in the psychophysics of Gustav Fechner (§2)) is only a quite limited aspect of human psychology which, as a whole, is characterized by a qualitatively different causality: 'psychical causality'.

To understand fully the psychological causality of the mind, it is not enough to examine the isolated consciousness of an individual, because in reality this is always embedded in a cultural web. These cultural mediations between the individual and the world determine the characteristics of the human mind. Therefore, we must expand the merely analytical framework of the individual mind and conceive of it as the product of a multilayered development. Ontogeny, natural history and social history are interwoven processes which demand multidisciplinary study. Wundt called this study of cultural development '*Völkerpsychologie*'. (The translation 'folk psychology' is misleading; better would be 'cultural psychology' or 'historical psychology' - in fact, Wundt also used both these terms.)

Völkerpsychologie is the comparative study of the products of social life in different historical periods and cultures. Wundt focused on language, custom, ritual, institutions and collective representations such as myth, art and religion. The great importance given to language as the central mediation between the individual mind and a culturally meaningful context was reflected in his psycholinguistic theory.

Like the 'Heidelberg programme' of Wundt's early years, the cultural-historical programme of the Leipzig *Völkerpsychologie* must be seen against the background of the tradition of philosophy of mind in German Idealism (see German idealism). Wundt's thought was a Hegelian variant, as was the conception of *Geisteswissenschaften* of his contemporary Wilhelm Dilthey; while a few years later, Ernst Cassirer presented a Kantian variant in his *Philosophie der symbolischen Formen*.

Wundt believed that in order to complete individual psychology, the methodological spectrum had to be broadened. Like history, anthropology or psychopathology, *Völkerpsychologie* is based on the ability to immerse

oneself in an alien state of mind and various stages of the cultural evolution of human thought. In the ten volumes of *Völkerpsychologie*, Wundt outlined an initial methodology of 'participant observation' which he thought essential to all cultural psychology.

While many elements of *Völkerpsychologie* were developed in disciplines such as sociology, anthropology and cultural history, Wundt's approach did not play a major role in mainstream psychology or philosophy of mind. In the behaviourist movement especially, it was completely discarded. Only since the 1970s has a new orientation in philosophy and psychology towards the mind yielded a growing interest in the cultural and historical mediations of mental development and thus in Wundt's work (see Bruner 1990; Jahoda 1992).

See also: Dualism; Introspection, psychology of; Psychology, theories of §2

JENS BROCKMEIER

List of works

Wundt, W. (1862) *Beiträge zur Theorie der Sinneswahrnehmung (Contributions to the Theory of Sense Perception)*, Leipzig and Heidelberg: Winter.(An exposition of Wundt's starting point: experimental analysis of the elementary psychological (not physiological) processes of perception.)

Wundt, W. (1863-) *Vorlesung über die Menschen- und Thierseele*, Leipzig: Voß; trans. J.E. Creighton and E.B. Titchener, *Lectures on Human and Animal Psychology*, New York: Macmillan, 1894.(Wundt's first systematic treatment of the traditional philosophical field of 'mind', 'soul' and 'thought' from the point of view of an experimental and quantitative psychology.)

Wundt, W. (1874) *Grundzüge der physiologischen Psychologie*, Leipzig: Engelmann, 6th edn, 3 vols, 1908-11; Part 1 of 5th edn, trans. E.B. Titchener, *Principles of Physiological Psychology*, New York: Macmillan, 1904. (A systematic account of the new scientific (that is, psychological) psychology, which deals with all areas of psychological life from the nervous system to thought, affect and intention, to anomalies of the mind. Volume 3 includes the exposition of a specific epistemology of psychology.)

Wundt, W. (1880-3) *Logik. Eine Untersuchung der Prinzipien der Erkenntnis und der Methoden der wissenschaftlichen Forschung (Logic: An Investigation of the Principles of Knowledge and on the Methods of Scientific Research)* Stuttgart: Enke, 2 vols; 4th and 5th edn, Leipzig: Engelmann, 1919-24, 3 vols.(Wundt's systematic account of methodology, epistemology and philosophy of science.)

Wundt, W. (ed.) (1883-1902) *Philosophische Studien*, 20 vols.(A periodical to which Wundt also contributed numerous articles.)

Wundt, W. (1886) *Ethik. Eine Untersuchung der Tatsachen und Gesetze des sittlichen Lebens*, Stuttgart: Enke; 5th edn, 1923-4, 3 vols; trans. E.B. Titchener, J.H. Gulliver and M.F. Washburn, *Ethics: An Investigation of the Facts and Laws of the Moral Life*, New York: Macmillan, 1897.(An empirical and scientific analysis of 'moral life'; for Wundt the basis of all moral philosophy and metaphysics.)

Wundt, W. (1889) *System der Philosophie (System of Philosophy)*, Leipzig: Engelmann; 4th edn, 1919, 2 vols. (Wundt's systematic treatment of metaphysics on an empirical basis; at the same time it is shown that science always presupposes philosophical thought.)

Wundt, W. (1896) *Grundriß der Psychologie*, Leipzig: Engelmann; 15th edn,1922; trans. C.H. Judd, *Outlines of Psychology*, Leipzig: Engelmann; 3rd revised edn, 1907.(One of Wundt's most published and read works, summarizing his view of scientific psychology and its relevance to philosophy.)

Wundt, W. (1901) *Einleitung in die Philosophie* (An Introduction to Philosophy), Leipzig: Kröner; 9th edn, 1922. (A historical presentation of philosophy and its development from Greek antiquity to the twentieth century, based on a series of lecture courses.)

Wundt, W. (1900-4) *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus und Sitte (Folk Psychology: An Investigation of the Laws of Development of Language, Myth and Custom)*, Leipzig: Engelmann, 2 vols; 3rd edn, Stuttgart, Kröner, 10 vols, 1911-20.(The monumental attempt to develop an all-encompassing cultural-historical psychology focusing on language (volumes 1 and 2), myth and religion (volumes 4 to 6), society (volumes 7 and 8), law (volume 9) and history of culture (volume 10).)

Wundt, W. (1911) *Einführung in die Psychologie*, Leipzig: Voigtländer; trans. R. Pinter, *An Introduction to Psychology*, London, 1912.(A brief introduction to the psychology of

Wundt, W. (1914) *Leibniz*, Leipzig: Kröner.(Wundt's only monographic work on a philosopher, emphasizing his importance for the modern conception of scientific psychology and philosophy.)

Wundt, W. (1912) *Elemente der Völkerpsychologie. Grundlinien einer psychologischen Entwicklungsgeschichte der Menschheit*, Leipzig: Kröner; trans. E.L. Schaub, *Elements of Folk Psychology: Outlines of a*

Psychological History of Mankind, London, 1916.(A synthesis of the major themes of *Völkerpsychologie*.)
Wundt, W. (1920) *Erlebtes und Erkanntes*, Stuttgart: Kröner.(Wundt's autobiography: the attempt to describe his life as a 'resultant' of cultural history.)
Wundt, W. (1927) *Wilhelm Wundts Werke. Ein Verzeichnis seiner sämtlichen Schriften*, Munich: Beck.(A bibliography of Wundt's publications.)

References and further reading

Beuchelt, E. (1974) *Ideengeschichte der Völkerpsychologie (A History of 'Völkerpsychologie')*, Meisenheim am Glan: Hain.(The history of *Völkerpsychologie* before and after Wundt, in Germany, France, Britain and North America.)
Boring, E.G. (1957) *A History of Experimental Psychology*, New York: Appleton Century Crofts, 2nd edn, 318-47.(Boring, a student of Titchener, highlighted the significance of Wundt as an experimental scientist but dismissed the philosophical and cultural-historical aspects of his work. This was for a long time the dominant reading of Wundt.)
Bringmann, W.G. and Scherer, E. (eds) (1980) *Psychological Research*, *Wundt Centennial Issue* 42 (1-2).(The articles of these volumes offer a rather differentiated overview of many aspects of Wundt's life and work, pointing out its impact especially on American philosophy and psychology, for instance on William James.)
Bringmann, W.G. and Tweney, R.D. (eds) (1980) *Wundt Studies*, Toronto: Hogrefe. (A selection of essays on Wundt's life and work.)
Bringmann, W.G. and Ungerer, G.A. (1980) 'An archival journey in search of Wilhelm Wundt', in J. Brozek and L.J. Pongratz (eds) *Historiography of Modern Psychology*, Toronto: Hogrefe, 201-40.(An overview of the numerous primary sources by and about Wundt in archives and libraries, and of major research projects on Wundt.)
Bruner, J.S. (1990) *Acts of Meaning*, Cambridge, MA, and London: Harvard University Press.(Bruner argues that the study of human nature must be based on cultural psychology: developing the Wundtian approach after cognitive revolution and linguistic turn.)
Diamond, S. (1980) 'Wundt before Leipzig', in R.W. Rieber (ed.) *Wilhelm Wundt and the Making of a Scientific Psychology*, New York and London: Plenum, 3-70. (Diamond sketches the first forty years of Wundt's life as a period of scientific and philosophical formation; at the end of this period the appearance of *Physiological Psychology* suddenly made Wundt the most prominent figure in an emerging science.)
Graumann, C.F. (1980) 'Experiments, statistics, history: Wundt's first programme of psychology', in W.G. Bringmann and R.D. Tweney (eds) *Wundt Studies*, Toronto: Hogrefe.(An account of the 'Heidelberg programme' of psychology developed in Wundt's seventeen years in Heidelberg. This programme included the attempt to systematically unite the two disparate branches of the study of the mind - experimental individual psychology and non-experimental *Völkerpsychologie* - in an evolutionary and historical perspective.)
Jahoda, G. (1992) *Crossroads between Culture and Mind: Continuities and Change in Theories of Human Nature*, New York and London: Harvester Wheatsheaf.(Outlines the history of cultural psychology, focusing on Enlightenment, German Idealism and - against this background - Wundt's conception of mind, *Kultur* and *Völkerpsychologie*.)
Robinson, D.N. (1982) *Toward a Science of Human Nature. Essays on the Psychologies of Mill, Hegel, Wundt, and James*, New York: Columbia University Press.(Robinson locates Wundt along the intellectual coordinates of nineteenth-century philosophy. He presents the philosophical psychologist Wundt as a psychological philosopher who developed his system - 'a kind of applied Hegelian psychology' - at a time when the tensions between *Geisteswissenschaften* (humanities) and *Naturwissenschaften* (sciences) centred on problems of psychology.)
Wundt, E. (1927) *Wilhelm Wundts Werke. Ein Verzeichnis seiner sämtlichen Schriften*, München: Beck.(A complete catalogue of Wundt's writings.)

Wyclif, John (*c.*1330-84)

John Wyclif was a logician, theologian and religious reformer. A Yorkshireman educated at Oxford, he was first prominent as a logician; he developed some technical notions of the Oxford Calculators, but reacted against their logic of terms to embrace with fervour the idea of the real existence of universal ideas. He expounded his view as a theologian, rejecting the notion of the annihilation of substance (including the eucharistic elements) and treating time as merely contingent. The proper understanding of universals became his touchstone of moral progress; treating scripture as a universal idea, he measured the value of human institutions, including the Church and its temporal property, by their conformity with its absolute truth. These views, though temporarily favoured by King Edward III, were condemned by Pope Gregory XI in 1377 and by the English ecclesiastical hierarchy in 1382, forcing him into retirement but leaving him the inspirer of a clandestine group of scholarly reformers, the Lollards.

1 Life

John Wyclif was born to a Richmondshire (Yorkshire) family about 1330 and ordained in 1351. He made his reputation as a logician and natural philosopher at Oxford. Though his earliest logical works and *De actibus animae (The Actions of the Soul)* were unremarkable, his distinctive ideas emerged from theological study, in the several tracts of his *Summa de ente (Summa on Being)* and especially in *De universalibus (On Universals)*, where he developed his distinctive theory of universals in logical terms, though not without theological opposition. Seeing the Bible as one such eternal idea in the mind of God and therefore an unchanging moral yardstick, he contrasted the ephemeral institutions of the Church with the absolute values instilled by scripture in the individual conscience, in *De civili dominio (On Civil Dominion)* and *De ecclesia (On the Church)*. These opinions made him numerous ecclesiastical enemies, while gaining him the temporary patronage of government and the lasting loyalty of a group of scholars. He went on to dismiss the doctrine of transubstantiation on logical grounds, an unorthodox view which forced him to leave Oxford for his Leicestershire living. In retirement, he produced a mass of repetitive controversial pamphlets and summaries. He inspired the Lollard scholars and preachers who supported him in translating the scriptures into English and in disseminating his ideas throughout England.

2 Contribution to logic

Wyclif's philosophical ideas were those of an accomplished logician inspired by a Platonizing religious vision. His contribution to logic is primarily found in the tracts published as *De logica (On Logic)*, written for theologians with a defective logical training. He expounded the logic of the Oxford Calculators, of whom Thomas Bradwardine, William Heytesbury and Richard Billingham were the most influential (see Oxford Calculators). Billingham in particular, whose lectures Wyclif may have attended in the 1350s, provided the basis for Wyclif's ideas of possible proofs of propositions. Wyclif proposed a fourfold division of proof: a priori, a posteriori, proof by reduction and proof *ab aequo*, that is, by definition and exposition or resolution of terms. This division was adopted and refined by later Oxford logicians (see Logic, medieval).

Wyclif also made a contribution to the understanding of insoluble propositions which approached the modern notion of semantic groundedness. Furthermore, he rejected the nearly unanimous view of contemporary logicians that there were real *continua*, such as time, maintaining instead that these consisted of discrete instants, points and so on (see Natural philosophy, medieval). He saw such abstract concepts, or universals, as real entities, indestructible ideas in the mind of God, rejecting the linguistic thinking of contemporary logicians, whom he came to dismiss in later works as 'sign-doctors' (see Language, medieval theories of).

3 Philosophical ideas

Wyclif took a distinct view in his most important philosophical work, *De universalibus*, on the real existence of universals, to which his approach was again that of a logician. He set out his view primarily as an analysis of predication: how could Socrates *be* a man, and how could man *be* a species? Predication involved the notions of identity and distinction; besides the traditional 'essential' distinctions of genus and species, he posited the 'real' distinction of entities with the same essence, like the persons of the Trinity, and the 'formal' distinction between the matter and substantial form of an individual, such as between God the Father and the divine nature. An

individual 'participated' in a number of universals, being the end of a chain of inherence. The predication of a universal could presuppose a real and not merely a linguistic relation: as smoke signifies fire, and fire is thus predicated of smoke, so a real universal could be predicated of an individual (see Universals).

It followed that if a real universal outside the accidents of time and space inhered in all substances, no substance could be destructible: God 'could not annihilate a creature without annihilating the whole created universe' (*De universalibus* 307), as such destruction would include the destruction of eternal intelligible being, a universal in which it inhered. Creatures subject to change and confined in place and time, being also ideas, must equally exist outside time, in the state Wyclif called *duratio*. Wyclif resolved the question of what distinguished two individuals in which the same universal inhered by positing a distinction in their different positions in time.

Wyclif's understanding of real universals seemed to him to resolve in an original way a great question of the schools, the relation of human free will (which he acknowledged like his contemporaries) to the perfection of divine foreknowledge. Though all human actions follow upon the will of God, still the unchanging will of God as to future contingent events may be determined by human volition at any moment. Since God's will is outside time, operating at once in the past and the future, events now may constitute causes determining God's will from the beginning of time. Human merit, therefore, may temporally, but not in principle, be prior to divine grace; grace must be necessary and eternal as it proceeded from God, and only contingent as it inhered in essences (see Eternity; Free will; Grace).

Some further, unorthodox theological consequences of real universals were more difficult to adopt, and Wyclif hesitated before embracing them. One clear consequence of his belief that substances could not be annihilated touched the eucharist. The doctrine of transubstantiation required that the elements, bread and wine, were annihilated and replaced by the body of Christ. Eventually, after attempting to reconcile orthodox doctrine with real universals, Wyclif rejected transubstantiation, maintaining that the sacramental elements remained in the consecrated host. This was a principal reason for the condemnation of his theology.

His view of universals, albeit couched in contemporary logical language, was at root a religious conviction formed under the Platonizing influence of Robert Grosseteste. Wyclif considered a proper understanding of universals the firm ground of moral progress, a prophylactic against moral relativism. Awareness of this marked a quantum leap (*gradus percipuus*) up the ladder of wisdom: 'error about universals is the cause of all the sin that reigns in the world' *De universalibus* 175-7). His convictions gave him the strength to pursue their moral and political implications.

4 Scripture and human dominion

From about 1374, while Wyclif worked out what he saw as the moral and practical consequences of his ideas, he propagated them in countless lectures, pamphlets and sermons, and gathered disciples who carried out a scheme for an English text of the scriptures. He had already identified the Bible as a real universal. Faced with the problem of a corrupt text, he explained that 'intellection in the mind is more truly scripture than lines on a membrane' (*De veritate sacrae scripturae* (1377-8), i. 189). Scripture was a moral absolute, but operated internally as a guide, allowing for no earthly authority to interpret it.

On this basis, he approached the currently discussed question of divine and human dominion. The latter was merely contingent and relative, depending on its conformity with God's will in each individual case; there was no independent right of dominion on the part of ecclesiastical or secular authorities. Lordship was merely stewardship, and depended upon the state of grace of its wielder. Wyclif strongly emphasized the implications for the temporal possessions of the clergy; the consequences for civil government were only gradually recognized. His subjective political philosophy thus attracted some superficial support from the English government, but was explicitly condemned by successive church authorities. Together with his views on transubstantiation, it ensured his removal from Oxford and confinement to his role as the inspirer of a body of evangelizers.

5 Influence

As the inspirer of the Wycliffite English bible and of a programme of disendowment and purification of devotion, Wyclif had great influence in England, in Bohemia through the Hussite reformers (see Hus, J.) and throughout Protestant Europe in the sixteenth century. His logic circulated with that of other Oxford logicians in

fifteenth-century Italy, as a set of texts for instruction. His more general philosophical works were much read in England up to the mid-fifteenth century, and played a subordinate role in the corpus of his writings copied in Bohemia; but they were forgotten in the general disfavour which the logic of the schools attracted after 1500, and when Wyclif's works began to be republished in the late nineteenth century, their importance and originality was not recognized. Substantial sections are still unpublished, including the important tract *De tempore*. However the significance of their contents was appreciated by several scholars, beginning with M.H. Dziewicki and S. Harrison Thomson (1931), who each edited some of his philosophical works, and by J.A. Robson (1961), Anthony Kenny (1985, 1986) and Paul V. Spade (1975), who all noted original features of his thought. Further study and editions of the remaining philosophical works are still required.

See also: Language, medieval theories of; Trinity; Universals

JEREMY CATTO

List of works

Wyclif, John (1368-9) *De actibus animae (On The Actions of the Soul)*, ed. M.H. Dziewicki, London: Trubner, for the Wyclif Society, 1902.(The only surviving account of Wyclif's earliest philosophical thought.)

Wyclif, John (1368-9) *De ente praedicamentali (On Categorial Being)*, ed. R. Beer, London: Trubner, for the Wyclif Society, 1901.(Develop's Wyclif's views of quality and quantity.)

Wyclif, John (1368-9) *Summa insolubilium (Summary of Insoluble Propositions)*, ed. P.V. Spade and G.A. Wilson, Binghamton, NY: Medieval and Renaissance Texts and Studies, 1986.(Contains the fullest discussion of Wyclif's logic of insolubles.)

Wyclif, John (1371-2) *Summa de ente libri primi tractatus primus et secundus (Summary on Being, Book I, Tracts 1 and 2)*, ed. S.H. Harrison Thomson, Oxford, 1930.(Wyclif's first ideas on universals.)

Wyclif, John (1371-3) *Tractatus de logica (On Logic)*, ed. M.H. Dziewicki, London: Trubner, for the Wyclif Society, 1893.(An extended treatment of Wyclif's logic for the use of theologians.)

Wyclif, John (1371-4) *De ente (On Being)*, ed. M.H. Dziewicki, London: C.K. Paul, for the Wyclif Society, 1909.(Further development of Wyclif's ideas on being and universals.)

Wyclif, John (1371-4) *De trinitate (On The Trinity)*, ed. A. du Pont Breck, Boulder, CO: University of Colorado Press, 1962.(Wyclif's idea of God in the light of his theory of universals.)

Wyclif, John (1374) *Tractatus de universalibus (Treatise on Universals)*, ed. I.J. Mueller, Oxford: Clarendon Press, 1985; trans. A. Kenny, *On Universals*, Oxford: Clarendon Press, 1985.(Wyclif's final and full ideas on universals.)

Wyclif, John (1375-7) *De civili dominio (On Civil Dominion)*, ed. R.L. Poole and J. Loserth, London: Trubner, for the Wyclif Society, 1885-1904.(Wyclif's theory of dominion and his answer to his critics.)

Wyclif, John (1377-8) *De veritate sacrae scripturae (The Truth of Holy Scripture)*, ed. R. Buddensieg, London: Trubner, for the Wyclif Society, 1905-7.(Wyclif's application of his idea of universals to scripture.)

Wyclif, John (1378) *De ecclesia (On The Church)*, ed. J. Loserth, London: Trubner, for the Wyclif Society, 1886.(Wyclif's theory of the visible and invisible church.)

Wyclif, John (1380-3) *De eucharistia (On The Eucharist)*, ed. J. Loserth, London: Trubner, for the Wyclif Society, 1892.(Wyclif's final views on transubstantiation.)

References and further reading

Ashworth, E.J. and Spade, P.V. (1992) 'Logic in Late Mediaeval Oxford', in J.I. Catto and T.A.R. Evans (eds) *The History of the University of Oxford*, Oxford: Clarendon Press, vol. 2, 35-64.(Discusses Wyclif's logic in historical context.)

Catto, J.I. (1992) 'Wyclif and Wycliffism in Oxford, 1356-1430', in J.I. Catto and T.A.R. Evans (eds) *The History of the University of Oxford*, Oxford: Clarendon Press, vol. 2, 175-261.(A comprehensive survey of Wyclif's and his academic contemporaries' ideas and activities.)

Hudson, A. (1988) *The Premature Reformation*, Oxford: Clarendon Press.(A full study of Wycliffite popular thought and its influence.)

Kenny, A. (1985) *Wyclif*, Oxford: Clarendon Press.(An admirably clear exposition of Wyclif's philosophy.)

Kenny, A. (ed.) (1986) *Wyclif in his Times*, Oxford: Clarendon Press. bib_not>(Contains excellent essays on Wyclif's thought and influence by Anne Hudson, Maurice Keen, Anthony Kenny, Norman Kretzmann and Gordon Leff.)

Mallard, W. (1961) 'John Wyclif and the Tradition of Biblical Authority', *Church History* 30: 50-60.(A clear

summary of Wyclif’s view on scripture.)
Robson, J.A. (1961) *Wyclif and the Oxford Schools*, Cambridge: Cambridge University Press.(The best description and analysis in context of Wyclif’s philosophy.)
Stalder, R. (1962) ‘Le concept de l’église selon le "de Ecclesia" de Wiclif’ (The Concept of the Church in Wyclif’s *De ecclesia*), *Bijdragen. Tijdschrift voor Filosofie en Theologie* 22: 38-81, 287-302.(The best exposition of Wyclif’s notion of church authority.)
Spade, P.V. (1975) *The Mediaeval Liar*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies.(Discusses the literature, including that of Wyclif, on insoluble propositions.)
Thomson, S.H. (1931) ‘The Philosophical Basis of Wyclif’s Theology’, *Journal of Religion* 11: 86-116.(The pioneering essay on Wyclif’s philosophical ideas.)
Thomson, W.R. (1983) *The Latin Writings of John Wyclyf: an Annotated Catalog*, Toronto: Pontifical Institute of Medieval Studies.(A full bibliography of Wyclif’s Latin works.)
Workman, H.B. (1926) *John Wyclif*, Oxford: Clarendon Press, 2 vols.(The most modern biography and a comprehensive study.)

Xenocrates (396-314 BC)

The Greek philosopher Xenocrates was the third head of the Platonic Academy. Like his predecessor Speusippus, he further developed Plato's philosophy, but along more orthodox lines. Indeed, Xenocrates contributed much to the formalization of Plato's philosophy into dogma. Starting from a metaphysical system of Monad and Dyad, the former being a self-contemplating intellect on the Aristotelian model and the latter a material principle, he systematically derived the rest of creation, postulating first the generation of number, and then soul, defined as 'self-moving number'. He is notable for a tendency towards triadic divisions of the universe, and a developed theory of daemons. He was probably responsible for the first definitive edition of Plato's works.

1 Life and works

Xenocrates was a native of Chalcedon, on the Bosporus. He succeeded to the headship of the Platonic Academy in 339 BC, on the death of Speusippus (see Academy). He had been a member of the Academy since early youth, and had accompanied Plato to Sicily, presumably on Plato's second or third visit (see Plato §1). He continued in the headship to his death, when he was succeeded by Polemo.

Xenocrates is credited with being the first to distinguish formally between the three branches of philosophy that comprise physics (including metaphysics), ethics (including politics) and logic (fr. 1 Heinze, fr. 82 Isnardi Parente; Xenocrates' fragments appear in two separate editions, cited below as 'H' and 'IP', respectively) - although Aristotle does seem to make the distinction in *Topics* I 14. Seventy-six works of his are listed by the biographer Diogenes Laertius, covering all of these fields, the most important being *On Nature* in six volumes, *On Wisdom* (*sophia*) also in six books, *On Being*, *On Fate*, *On Virtue*, *On Forms*, *On the Gods*, *On the Soul*, *On the Good*, *Solution of Logical Problems* (ten volumes) and *On Genera and Species*. He also wrote on Pythagoreans and on numbers, which indicates a continuation of Speusippus' interest in these areas. All of his works have perished, apart from a short extract from his *Life of Plato* (fr. 53 H, frs 264-6 IP), attributing to Plato a doctrine of five (instead of the usual four) elements, on the basis of an interpretation of the *Timaeus*.

2 Metaphysics and cosmology

As first principles, like Plato and Speusippus, Xenocrates postulated a pair, which he seems to have termed Monad and Dyad (fr. 15 H, fr. 213 IP), the latter a principle of multiplicity and unlimitedness (fr. 68 H, fr. 188 IP). Xenocrates also identifies his second principle as matter (fr. 28 H, fr. 101 IP). What we seem to have here is a 'female' principle manifesting itself sequentially at various levels of reality, rather like that of Speusippus. As regards the Monad, however, Xenocrates differs significantly from Speusippus in declaring it to be an intellect, or *nous* (fr. 16 H, fr. 214 IP) - possibly in reaction to Aristotle's criticisms of Speusippus. At any rate, the essentially Aristotelian concept of a self-contemplating divine intellect is that which is dominant in all subsequent 'official' Platonism up to Plotinus.

Another point of difference between Xenocrates and his predecessors seems, if we accept the view of Theophrastus (*Metaphysics* 6a23-), to have been the comprehensiveness of his attempt to derive the totality of existence from his first principles. The doxographic report in fragment 15 H (fr. 213 IP) tends to confirm this, although it contains problems: it is not even clear whether Xenocrates intends his Monad to be transcendent or immanent, while the Dyad, unless the text is to be amended, is presented as identical with the world-soul, which it should not be, since we learn from Plutarch (*On the Creation of Soul* 1012d-e) that for Xenocrates the world-soul is the product of Monad and Dyad.

Xenocrates first derived number from the Monad and Dyad, then point, line, plane and solid from number - or perhaps from the first four numbers, the Pythagorean *tetraktys* (see Pythagoreanism §2). This involved him in the troublesome doctrine of indivisible lines (and planes and solids), which is criticized in the pseudo-Aristotelian work *On Indivisible Lines* (frs 41-9 H, frs 123-47 IP) (see Atomism, ancient).

As regards the Platonic theory of Forms, Xenocrates is reported as going the opposite way to Speusippus, abandoning the 'mathematicals' while Speusippus dismissed the Platonic ideal numbers, but this is doubtless an oversimplification. Aristotle asserts that Xenocrates postulated only Forms, which he identified with numbers (fr. 34 H, frs 103-12 IP). Although mathematical entities, the Forms are still causes of sensible particulars. Xenocrates

is on record (fr. 30 H, fr. 94 IP) as defining a Form as 'the paradigmatic cause of regular natural phenomena', a definition intended to rule out forms of artificial objects and things contrary to nature (*ta para physin*). This became the standard definition in later Platonism. As to the location of the Forms, since the Monad is an intellect, and an intellect must think, it seems not unreasonable to trace the later Platonic doctrine of the Forms as the thoughts of god back to Xenocrates, but we have no evidence on this point (see Platonism, Early and Middle (§3) for a more detailed discussion of Form-numbers).

Soul he defined as 'a self-moving number' (fr. 60 H, frs. 176-87 IP). We learn from Plutarch's *On the Creation of Soul* that Xenocrates gave a description of the formation of the soul that was simply a development of the creation of number (and thus of Forms) (fr. 68 H, fr. 188 IP). Basing himself on Plato's account of the soul's components in *Timaeus*, he takes the 'indivisible substance' as the Monad and the 'divisible substance' as the Dyad. Their mixture results in number, but not yet soul. It lacks 'motion', the defining characteristic of soul. This is supplied by the elements of sameness and otherness, the former representing stability, the latter motion and change.

A notable feature of Xenocrates' physics is his triadic division of reality into intelligible, sensible and 'opinable' realms, this last comprising the heavens, the contents of which are visible, but their nature only cognizable by reasoning (fr. 5 H, fr. 83 IP). A rather different, and more peculiar, triadic division is reported by Plutarch (fr. 56 H, fr. 61 IP), involving different degrees of 'density', combined with various of the four elements.

Xenocrates had a developed doctrine of daemons (see Platonism, Early and Middle §7), about which we learn much, again, from Plutarch (fr. 23 H, frs 222-3 IP). Again, mathematics is involved, since daemons are declared analogous to isosceles triangles, intermediate between the equilateral triangles of the gods and the scalene triangles proper to humanity. He postulated evil daemons as well as good, as an explanation for all the nasty aspects of religion such as ritual obscenity and human sacrifice (fr. 25 H, frs 229-30 IP).

3 Ethics and logic

On ethics Xenocrates wrote a good deal, but not much that is distinctive survives. He defined happiness (see Eudaimonia) as 'the possession of the excellence (*arēte*) proper to us, and of the power subservient to it' (fr. 77 H, fr. 232 IP). He agreed with Aristotle about the necessity of a modicum of bodily and external goods (this constituting the 'power' of the definition). He anticipated the Stoics by identifying the first principles of happiness as 'the primary natural instincts' (*ta prōta kata physin*) (frs 78-9 H, frs 233-4 IP) (see Stoicism §15-17).

In logic, he still maintained the Platonic system of 'division', together with the basic categories of absolute and relative (fr. 12 H, fr. 95 IP), rejecting the Aristotelian ten categories (see Aristotle §7). In epistemology he recognized three levels of cognition, corresponding to the three levels of the universe (fr. 5 H, fr. 83 IP).

JOHN DILLON

References and further reading

Cherniss, H. (1945) *The Riddle of the Early Academy*, Berkeley and Los Angeles, CA: University of California Press.(A useful discussion of Form-numbers, especially in chapter 2.)

Dillon, J. (1977) *The Middle Platonists*, London: Duckworth.(See pages 22-39 for an introductory account of Xenocrates.)

Krämer, H.-J. (1967) *Der Ursprung der Geistmetaphysik (The Origin of the Metaphysic of Mind)*, Amsterdam: Grüner.(Pages 21-191 deal with Xenocrates, with a useful presentation of the evidence, although on a rather speculative level.)

Xenocrates (396-314 BC) Fragments, in R. Heinze, *Xenokrates*, Leipzig: Teubner, 1892; repr. Hildesheim: Olms, 1965; M. Isnardi Parente, *Senocrate - Ermodoro: frammenti*, Naples: Bibliopolis, 1982.(Two collections of fragments with commentary, the latter including an Italian translation; textual references to fragments from these texts are H and IP, respectively.)

Xenophanes (*c*.570-*c*.478 BC)

Xenophanes was a philosophically minded poet who lived in various cities of ancient Greece. He is best remembered for an early comment on the limits of knowledge, a critique of anthropomorphism in religion and an advance towards monotheism. The surviving fragments of his poems span a wide range of topics, from proper behaviour at symposia and the measures of personal excellence to the nature of the divine, the forces that rule nature and how much can be discovered by mortals concerning matters in either realm. Both Plato and Aristotle characterized him as the founder of Eleatic philosophy, a view echoed in the pseudo-Aristotelian treatise, On Melissus, Xenophanes and Gorgias*, and in ancient doxographical summaries. But in many of his poems Xenophanes speaks as a civic counsellor and inquirer into nature in the tradition of the philosopher-scientists of Miletus. While his one, unmoving, whole and eternal divinity bears some resemblance to Parmenides' 'being', in other teachings he anticipates the views of Heraclitus and Empedocles. His comments on divine perfection, the limited utility of the victorious athlete and the need to restrict poetic expression all foreshadow views expressed by Plato in the* Republic*.*

1 Life and writings

Xenophanes is reported to have been the son of Dexios (or Dexinos or Orthomenes) of Colophon. In fragment 8 he claims to have lived in Colophon for twenty-five years before leaving to spend the next sixty-seven as a travelling bard 'tossing about the Greek land'. If we tie his departure from Colophon to the conquest of Ionia by Harpagus the Mede (in 546/5 BC) and accept the report of the historian Timaeus that Xenophanes lived in the reign of Hieron of Syracuse (478-467 BC), we may date his life to the years 570-478 BC.

Ancient reports of his contacts with other philosophers are few and inconclusive, but the reference in fragment 7 to a puppy which possessed the soul of a friend shows some awareness of Pythagorean teachings (see Pythagoreanism §3) and the remarks about earth, sun, sea, waters and rainbow in fragments 27-32 display an interest in matters investigated by the Milesians. Herodotus (I 74) reports Xenophanes' 'admiration' or 'wonder' (or perhaps 'amazement') regarding Thales' successful prediction of a solar eclipse (see Thales §1); Heraclitus' fragment 40 disparages him as a (mere) polymath; and Empedocles' fragment 39 challenges the view of the earth's 'unlimited' depths expressed in Xenophanes' fragment 28.

Xenophanes wrote in verse, a fact which reflects both his chosen profession and an age that drew no sharp distinction between poet, sage and teacher. Most of his poems eschew complex argument in favour of simple and emphatic dissent (the Greek words *de* and *alla* - both meaning 'but' - appear often), but fragments 2, 15, 30, 34 and 38 employ hypothetical suppositions or 'thought experiments' to establish unnoticed contrasts and connections, and the different aspects of his teachings can be placed within a coherent overall scheme. He is the first Greek philosopher for whom an appreciable body of work has survived, but his difficult terminology and phrasing, along with the conflicting ancient testimonia, have spawned radically different interpretations of his teachings and sharply divergent appraisals of his importance as a thinker.

2 Social criticism

The material contained in five fragments (1, 5, 8, 22 and 45) appears to have been composed for performance at symposia; seven others (2, 3, 4, 6, 7, 20 and 21) mention individuals or issues suitable for discussion in such a setting. Although often considered of little philosophical interest, these verses establish Xenophanes in a role subsequently assumed by Heraclitus and Socrates - the philosopher as civic gadfly.

The detailed description of a wholesome banquet scene with which fragment 1 opens prepares the way for a concluding injunction to a parallel purity in speech and conduct, particularly in the songs sung on such festive occasions. Xenophanes' disapproval of the old stories of divine warfare and factional strife reflects the concern for the wellbeing of the city evident in his criticisms of the honours lavished on athletes (fr. 2), the luxurious lifestyle of the citizens of old Colophon (fr. 3), the stories told about the gods by Homer and Hesiod (frs 11-12) and Homer's exalted status as the educator of Greece (fr. 10). His description of himself as a travelling counsellor (frs 8 and 45), a reference to the Lydians as the inventors of coinage (fr. 4), a preferred method of wine-mixing (fr. 5) and a listing of topics suitable for discussion at symposia (fr. 22) reflect similar concerns.

Since Xenophanes lived before the time of the Sophists he had nothing to say on the many questions their sceptical and relativist views generated. Nevertheless, a number of his teachings - that the divine is incapable of immoral conduct, that wealth and luxury pose a threat to the survival of the city, that a civic good outweighs any individual good, and that the songs sung by poets must be censored for the sake of the city’s welfare - survive in the work of Plato and Aristotle.

3 Religious views

Since ’greatness’ for a Greek god typically meant ’greatness in power and honour’ (Homer, *Iliad* II 412 and *Odyssey* V 4), Xenophanes’ assertion that ’one god is greatest among gods and men’ (fr. 23) helps to explain a number of his remarks about god’s powers and perfections: god’s capacity to shake ’all things’ simply by thinking (fr. 25); the ’whole’ character of god’s thinking and perceiving (fr. 24); the degree of seemliness appropriate to god (fr. 26); various denials of divine births, bodies and voices (frs 14 and 23); the call to honour the gods (fr. 1); and the repudiation of the poets’ tales (frs 1, 11 and 12). While the ’one greatest god’ of Xenophanes embodies some qualities popularly ascribed to Zeus, his unchanging and isolated form of existence (according to frs 25 and 26) rules out many other popular beliefs - for example, that a god might reside in pine branches (fr. 17), that gods intimate their intentions to mortals (fr. 18, A52); and that the sun, sea, earth, moon and rainbow are themselves gods or goddesses (frs 30-32, A39, 41-6).

Xenophanes’ repeated use of the plural ’gods’ has sparked debate concerning the depth of his monotheism. Fragment 23 speaks merely of one *greatest* god, and it was the immoral conduct of Homer’s and Hesiod’s gods that prompted Xenophanes’ criticism, not their plurality. We may also reasonably doubt ’whether a convinced monotheist in an unreceptive polytheistic society would cloud the issue by a mention of plural gods which is at best ambiguous in the very context in which he is stating his revolutionary view’ (Stokes 1971: 76).

Fragment 16 asserts that ’Ethiopians [say that their gods are] snub-nosed and black/Thracians [that theirs are] blue-eyed and red-haired’ (similarly, fr. 15: ’… if horses or oxen or lions had hands… horses would draw the figures of the gods as similar to horses and the oxen as similar to oxen… ’). It is not clear whether these famous remarks were meant to refute all anthropomorphic conceptions of the gods, or merely to explain how the gods came to have the qualities popularly ascribed to them (that is, rightly or wrongly, people always assume that the gods are like themselves). Fragment 34 concedes that even unintelligent opinion occasionally enshrines the truth (see §5) and, since a different bodily feature is mentioned in each case, fragment 16 does not actually point to inconsistencies in popular views of the gods.

Later doxographical summaries report that Xenophanes identified god with the entire (spherical) physical universe, but many scholars consider the doxographers’ accounts to have been inspired mainly by Plato’s and Aristotle’s loose identifications of Xenophanes as the founder of Eleatic thought, and the practice of crediting the putative founder of a school with discoveries made by his followers.

4 Natural science

The seriousness of Xenophanes’ interest in natural science has been generally discounted, in part as a consequence of the bizarre character of some of the views ascribed to him and also (as in the case of Aristotle) because his real interests were assumed to lie elsewhere. But neither of these lines of thinking can withstand scrutiny.

Fragment 29 gives earth and water as his dualist response to the ’basic substance’ question raised by the Milesians. Fragments 27-33 and testimonia A1, 32, 33 and 36-48 all address standard Ionian scientific topics. Fragment 18 asserts in part that ’as mortals search they discover (a) better’, and A33 (from Hippolytus) reports that Xenophanes based his theory of periodic flooding on the discovery of fossilized sea creatures at inland locations. The view of all atmospheric and celestial phenomena as essentially ’cloud’ (see A38-41 and 43-6) fits nicely with the description of ’great sea’ in fragment 30 as the source of all winds, clouds and waters, together with the view of fragment 27 that ’all things come from and return to (the) earth’. The breadth of Xenophanes’ scientific interests, his evident appreciation of the value of inquiry and observation, and the internal coherence of his explanations make it unlikely that he pursued scientific inquiry merely to acquire a weapon to wield against popular religion.

To charge Xenophanes with a belief in the earth’s ’infinite’ depth (fr. 28), ’infinite’ worlds (A1), and the sun’s ’infinite’ travels (A 41a) we would have to translate *apeiron* by ’infinite’ rather than the more plausible

’indefinite’. His eccentric view that the sun cannot shine without someone present to observe it (A41a) has been plausibly reinterpreted as the argument that since the sun is generated from rising moist vapours it cannot continue to exist in arid (hence unpopulated) regions. While Xenophanes’ scientific understanding was deficient in a number of respects, he remains a central figure in the Ionian scientific revolution.

5 Theory of knowledge

Xenophanes’ brief remarks about human knowledge were the subject of competing interpretations as early as the fourth century BC. Although sometimes regarded as an early statement of philosophical scepticism, they are best understood in the light of a traditional poetic contrast between the narrow scope of human experience and the synoptic view enjoyed by the gods (Homer, *Iliad* II 484-7; Semonides, fr. 1; Archilochus, fr. 70). Fragment 37 speaks in this vein when it links what mortals think to what they have experienced: ’If god had not made yellow honey, they would think that figs were much sweeter’. Fragment 36, ’… however many they have made evident for mortals to look upon’, suggests that what we can experience is neither unlimited nor entirely within our control. Fragment 35, ’let these be accepted, of course, as like the realities, but… ’, presents a similarly divided outlook: one should accept these (perhaps what Xenophanes presents as his account of the nature of all things) as true, but… (that is, certain knowledge exceeds our grasp).

In fragment 34 Xenophanes denies that anyone has known or ever will know ’the clear and certain truth (*to saphes*) about the gods and what I say about all things’, adding (in yet another ‘thought experiment’) that ’even if one were to say precisely what is brought to pass, he himself would still not know’. Since other ancient writers linked the possession of knowledge of *to saphes* with direct observation (Herodotus, II 44; Thucydides, I 22.4; Alcmaeon, fr. 1), Xenophanes’ argument can be analysed as a combination of a common view of the requirements for knowledge with the recognition that no account of the nature of the divine or the forces controlling the cosmos (including his own novel proposals on both topics) could possibly be confirmed on the basis of direct observation.

J.H. LESHER

References and further reading

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge & Kegan Paul, 2 vols.(Includes an idiosyncratic but philosophically stimulating account of Xenophanes’ views and supporting arguments.)

Classen, C.J. (1989) ’Xenophanes and the Tradition of Epic Poetry’, in K.J. Boudouris (ed.) *Ionian Philosophy*, Athens: International Centre for Greek Philosophy and Culture, 91-103.(An informative study of Xenophanes’ diction and poetic style.)

Fränkel, H. (1925) ’Xenophanesstudien’, *Hermes* 60: 174-92; repr. in *Wege und Formen frühgriechischen Denkens*, Munich: C.H. Beck Verlag, 3rd edn, 1968; a portion has been translated into English by M.R. Cosgrove, ’Xenophanes’ Empiricism and his Critique of Knowledge’, in A.P.D. Mourelatos (ed.) *The Pre-Socratics*, Garden City, NY: Doubleday, 1974.(An important study of Xenophanes’ views on knowledge.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols. (Includes the most detailed and comprehensive English-language history of early Greek thought; Xenophanes is discussed in volume 1 pages 360-402.)

Heitsch, E. (1983) *Xenophanes: Die Fragmente*, Munich and Zurich: Artemis Verlag.(A full-length study of Xenophanes; contains Greek texts of the fragments with translations and commentary in German.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(A valuable survey of Presocratic philosophy; selected fragments appear in Greek with English translations; Xenophanes is discussed in chapter 5 pages 163-80.)

Lesher, J. H. (1992) *Xenophanes of Colophon: Fragments*, Toronto, Ont.: University of Toronto Press.(A full-length study of Xenophanes in English; contains Greek texts of fragments with English translations, notes and commentaries, translations of ancient testimonia and essays on various aspects of Xenophanes’ philosophy.)

Stokes, M.C. (1971) *One and Many in Presocratic Philosophy*, Washington, DC: Center for Hellenic Studies. (Contains a lucid discussion of the doxographers’ accounts of Xenophanes’ theological views.)

Xenophanes (*c*.570-*c*.478 BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 1, 126-38.(The standard collection of the ancient sources, both fragments and testimonia, the latter designated by ’A’; includes Greek texts with translations in German.)

Xenophon (*c.*427-355/50 BC)

The Greek historian and philosophical writer Xenophon was a companion of Socrates, and is second in importance only to Plato as a source for our knowledge of him. He was also a penetrating and influential political thinker in his own right. He left Athens to embark on a spectacular military career in 401 BC, two years before Socrates' execution, and his military experiences had a deep impact on his thought and writings. An important historian and an innovator in literary forms, Xenophon limited his philosophical interests to political and ethical themes. Two questions are especially prominent: (1) What are the psychological roots of human virtue, and how can it be taught? (2) What are the limits of and the prospects for human attainment of self-sufficiency? He develops these themes most fully in two major works that present two competing models of the best human being. His Memoirs (usually referred to by the Latin title Memorabilia) presents the model of the philosophical life, mainly by recounting conversations between Socrates and a wide variety of human types. His The Education of Cyrus (often referred to by its Latin title Cyropaedia) presents the model of the political life, mainly by giving a fictionalized account of the rise to power of Cyrus the Great, founder of the Persian empire. Generally speaking, Xenophon seems in these works less willing than Plato and Aristotle to privilege the claims of philosophy over the claims of politics, and less optimistic about the power of reason to produce happiness. His works were highly esteemed by the Romans, as well as by such moral thinkers as Machiavelli, Montaigne and Rousseau.

1 Life

Born in Athens, Xenophon was in his youth a companion of Socrates. While still a young man he became an accomplished military leader, then in later life a distinguished prose author. He wrote two major historical works. The *Hellenica* covers political and military events from the late 400s down to the 360s. The autobiographical *Anabasis* is a pseudonymous account of his own greatest military success. It tells the story of how, although Socrates warned him of trouble ahead, Xenophon left Athens in 401 to join a Greek mercenary force hired by Cyrus, the younger brother of the king of Persia, for a campaign to take over the throne. (This Cyrus must not be confused with Cyrus the Great, founder of the Persian empire.) When Cyrus and the Greek generals were defeated and killed, the still youthful and inexperienced Xenophon was elected to take over leadership of the desperate Greek remnant, trapped in the middle of Asia Minor. He proved a brilliant commander, and saved the mercenaries by leading them to the Black Sea, where they met up with a Spartan force. His account of this expedition shows him a great tactician and skilful rhetorician. For obscure reasons, perhaps involving his Spartan connections, Xenophon was banished from Athens in the 390s, and subsequently served with distinction under the Spartan king Agesilaus. He did not return to Athens until the banishment was lifted shortly before his death in the later 350s.

Xenophon's thought and writings were deeply influenced by his military experiences. He was also a devotee of the vigorous outdoor entertainments of Greek gentlemen, and wrote influential practical treatises on hunting and horsemanship. Perhaps because of these influences and interests, scholarly opinion in the twentieth century tended to cast him as rather conventional and simple, lacking in the gifts of intellect and imagination necessary for original philosophical thought, or for a real appreciation of Socrates. But Xenophon was clearly a very brilliant and charismatic young man, in talents if not in tastes a rival of Alcibiades, to judge from their achievements. His moral wisdom was esteemed by the Romans (Cicero recommended Xenophon to his brother as the best writer on effective political leadership), as well as by moral thinkers such as Montaigne and Rousseau, who shared his lack of interest in more theoretical areas of philosophy. His intellectual interests and achievements are perhaps best compared to those of Machiavelli, who made much use of Xenophon. Both excel at thoughtful observation rather than theory construction, and both are model writers of clear, vivid and discreet prose, which often leaves their readers to draw out the lessons left implicit in their presentations of conversations and events.

2 Philosophical writings

After Socrates' execution in 399 BC, a number of his companions began to write reflections and remembrances about him, usually in the form of dialogues, and the style and approach of these 'Socratic discourses' defined them as a literary genre (see Socrates §§1-2, 7; Socratic dialogues). The genre came to include works in the dialogue style with protagonists other than Socrates, such as Cyrus the Great. Plato was of course far the greatest of the writers in this genre, but Xenophon was an honourable second. This is the literary context within which Xenophon

wrote his most directly philosophical works.

Many scholars in the nineteenth and twentieth centuries have tried to locate the true, historical Socrates by comparing and contrasting the Socratic dialogues of Plato and Xenophon. This search has reached no certain results. Some scholars have claimed that because Xenophon was a historian he is likely to be a better witness than Plato; others declare that because Plato was a ’real’ philosopher he could appreciate Socrates with a depth allegedly unavailable to the soldierly Xenophon. Neither position takes into account the peculiar nature of the genre of ’Socratic discourses’, best summarized by Arnaldo Momigliano: ’[B]iography acquired a new meaning when the Socratics moved to that zone between truth and fiction which is so bewildering to the professional historian. We shall not understand what biography was in the fourth century if we do not recognize that it came to occupy an ambiguous position between fact and the imagination…. With a man like Plato, and even with a smaller but by no means simpler man like Xenophon, this is a consciously chosen ambiguity’ (1993: 46).

Xenophon constructed his Socratic writings with a view to thinking about the moral and political issues raised by Socrates’ way of life. He emphasizes primarily what furthers his own thinking, not what gives a full or historical portrait of Socrates. Where he differs from Plato, the difference is more likely to reflect their different estimates of Socrates’ significance than their different recollections of the historical Socrates. Both bend historical fact to the service of moral analysis.

Xenophon wrote six works in this genre of Socratic discourses. His *Apology of Socrates to the Jury* and *Symposium* appear to be responses to Plato’s works of the same titles. The *Oeconomicus* is in part a response to the comic attack on Socrates in Aristophanes’ *Clouds*. The *Hiero* is a conversation about political power between Hiero the tyrant and Simonides the poet. His two major philosophical works, the *Memorabilia* and *Cyropaedia*, focus on two philosophical issues: (1) What are the psychological roots of human virtue, and how can it be taught? (2) What are the limits of and the prospects for human attainment of self-sufficiency? Generally speaking, Xenophon seems in these works less willing than Plato and Aristotle to privilege the claims of philosophy over the claims of politics, and less optimistic about the power of reason to produce happiness.

3 Virtue and education

Xenophon presents both Socrates and Cyrus the Great as educators, and has much to say about their influence on others. His Socrates is presented in the *Memorabilia* primarily through the educational effect he has on his companions, while the *Cyropaedia* is as much about the education that Cyrus *gives* to those he commands as it is about the education he *receives* in his youth. In these works, Xenophon presents no general theory or scheme of education comparable to, for example, Plato’s in the *Republic* (see Plato §14). Instead, he emphasizes the many ways in which moral formation must accommodate the particular aspirations and even prejudices and vices of those to be educated. Through the contrast between Socrates and Cyrus as educators, Xenophon reveals a stark choice between two competing possibilities for human virtue (see Aretē), but without clearly privileging one possibility over the other. He seems more interested in exploring the ambiguities of virtue than in making a case for one type over the other.

Xenophon’s Socrates influences others partly through precepts he teaches in conversation, partly simply through the force of his own example. The moral advice and example he imparts amount to commentary on the two great maxims inscribed over the portico of the oracle of Apollo at Delphi: ’Know thyself’ and ’Nothing in excess’. In conversation, Xenophon’s Socrates is a master of refutation (elenchus), like Plato’s Socrates (see Socrates §§3-4). These conversations often demonstrate to Socrates’ interlocutors that they are ignorant of what they must know to live up to their own aspirations. But Xenophon’s Socrates is more willing than Plato’s to offer explicit moral advice, at least to receptive young men whose pride has been humbled by refutation, and quicker to drop the pose of being a mere questioner who has no knowledge of his own. He teaches his companions to consider how much knowledge they must really have to prosecute successfully the projects they want to undertake, especially when these projects involve political power. The knowledge required turns out in part to be quite practical, such as how to guarantee Athens’ grain supply or the productivity of the silver mines. But it also requires answers to more philosophical questions, such as whether a young man will benefit from wielding great power rather than being destroyed by it. Xenophon has his Socrates show how ambition often would be made less rash if forced to acknowledge such considerations. In addition, Socrates exhorts his companions and inspires them through his own example to an ascetic control over passion and desire. He presents his own ascetic way of life as the foundation of

his freedom, and Xenophon makes this freedom the heart of people's fascination with Socrates. Indeed, Xenophon's portrait seems to give Socrates' freedom precedence over his wisdom, probably in conscious competition with Plato's portrait.

Xenophon's Cyrus is also a paragon of knowledge and self-control, but his example and his exhortations have an essentially different emphasis from those of Xenophon's Socrates. In general, Socrates improved the virtue of his companions by deflating their ambitions and reining in their desires. Cyrus instead convinces his followers that virtue is an indispensable means if they are to reach the ends of glory and wealth that they seek. The crucial moment in Cyrus' rise to power, as Xenophon presents it, comes when he convinces his Persian troops to pursue virtue, not for its own sake, but for the rewards that come from it. For this sort of education in virtue to take hold, it is absolutely necessary that the 'students' have vigorous desires for wealth and passions for power; such a psychological endowment is the fertile ground out of which Cyrus grows virtue, whether his own or other people's. Xenophon shows that this view of virtue creates a difficult dilemma for the leader: the very desires and passions that make talented people virtuous and therefore manageable by the effective leader also motivate them to become rivals for power. Cyrus is a master at using others' talents without undermining his own position of authority. The importance of this theme in the *Cyropaedia* explains why Xenophon's Cyrus is such an important point of reference in Machiavelli's *Prince* (see Machiavelli, N.).

This tension between Socrates and Cyrus as moral educators reveals an ambiguity about the nature and preconditions of virtue that is also prominent in two of Xenophon's smaller contributions to the genre of 'Socratic discourses'. His *Oeconomicus* shows Socrates trying to convince the young and rather irresponsible Critobulus to take more care of his family's wealth. This treatise on the household was a popular classic with the Romans, but it might seem odd that it should also be a vehicle for exploring political themes about virtue. In the conversation with Critobulus Socrates makes explicit the connection between the themes: he claims that there is in principle no difference between governing a city and governing a household, a claim also found in Plato's *Statesman* (see Plato §16). Most of the dialogue is taken up with Socrates' account of the teaching about household management of a gentleman farmer named Ischomachus. A critical component of household management as Ischomachus understands it is knowing how to motivate one's wife, servants and slaves to promote one's own good. In effect, Ischomachus looks for the same psychological endowment in these domestic subordinates that Cyrus sought in good soldiers: strong desire mixed with prudent calculation. Xenophon presents a politically more provocative version of essentially this same view of leadership and moral education in his *Hiero*, a dialogue between Hiero the tyrant and Simonides the poet about the costs and benefits of tyranny. Hiero complains that the best men in his city are exactly the ones most irked by subjection to his rule, and especially by the presence of his personal bodyguard of mercenaries. Simonides suggests that the tyrant can make himself more acceptable to these best men if he deploys the mercenaries to aid his subjects in controlling their own slaves. In other words, appeal not just to their desire for wealth or glory, but directly to their own interests in despotism. This will make the mercenary bodyguard palatable to the best men, even although it is the guarantee of their own subjection.

4 Self-sufficiency and the divine

Xenophon also considers Socrates and Cyrus in themselves rather than in their influence on others. This allows him to compare the highest achievement possible within the ways of life of philosophy and politics. For Xenophon, this comparison focuses on the extent to which the two ways of life achieve self-sufficiency (*autarkeia*). He conceives of self-sufficiency as a defining attribute of the divine, so that it is no mere figure of speech to say that Xenophon's two heroes represent two different paths to apotheosis. Here again, Xenophon is engaged in an examination that also engaged Plato and Aristotle, as well as the tragedians and Herodotus. And here too Xenophon refuses simply to subject the aspirations of the political life to a philosophical critique. Instead, he brings out the limitations of *either* political power *or* philosophy for living up to the aspiration to self-sufficiency. To a large extent, Xenophon opposes the position of Aristotle (§26), who said that humans should not merely think human thoughts, but should try to be as divine as possible.

Xenophon's Socrates seems to believe that both paths to divinization, the philosophical and the political, are impious and based on a failure to acknowledge divine constraints on human ambition. He rejects outright the claims of natural philosophers that the study of the cosmos is humanity's highest accomplishment. He denies that

they in fact have the knowledge of divine things they claim, and defends the Delphic view that 'Know thyself' takes precedence over any other investigation. To political ambition his response is more complicated. He fascinates his politically ambitious young companions because they see in him an image of the self-sufficient control over happiness to which they aspire. They are not sure how he does it, but they think that spending time with him will surely let them in on the secret. Socrates lets them believe this, and the tempting and teasing that result are at the core of Socrates' claim that his life is filled with *erōs* (love). Ambitious young men conceive an erotic passion for Socrates because they see him as the means to the knowledge, especially the self-knowledge, required by their political projects. Xenophon exposes the vanity of this passion by showing that, while in a sense Socrates does have the divinizing wisdom such men seek, it cannot be communicated or exchanged. Socrates' secure possession of the knowledge of what makes humans happy is ultimately supernatural: his wisdom is grounded in his 'divine sign', which is his unique prophetic endowment. This 'divine sign', a voice that occasionally gave Socrates advice and which is also reported by Plato, is crucial to Xenophon's understanding of how Socrates chastened the political ambitions of his companions. Socrates forced them to acknowledge that their own lack of this supernatural wisdom was an insuperable limitation to their pursuit of self-sufficient control over their own political success.

On this issue, too, Cyrus is the vehicle Xenophon employs to explore a less uncompromising view of the potentialities of the political life. His Cyrus does not have Socrates' sure access to divine guidance about which choices will prove beneficial or harmful. But he has the best 'political' approximation to the 'supernatural' divine sign of Socrates: he knows how to be his own diviner, so that he will never be at the mercy of anyone else's interpretation of sacrifices, omens or other signs of the divine purpose. Besides the useful information this pious dependence on divination provides to Cyrus, it is also a powerful rhetorical tool, for people are happy to follow a leader who seems to operate with divine favour. (The *Anabasis* shows that Xenophon was fond of such rhetoric himself.) Cyrus can then achieve a sort of self-sufficient control within politics that avoids the demands of knowledge that Socrates impressed upon his interlocutors. Thus Cyrus shows the farthest extreme of merely human control over success, a control whose limitations are brought into focus by comparison to Socrates.

DAVID K. O'CONNOR

List of works

Xenophon (*c*.390 BC) *Anabasis*, trans. C.L. Brownson, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1922.(Parallel Greek text and English translation.)

Xenophon (*c*.385 BC) *Apology of Socrates to the Jury*, trans. O.J. Todd, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1923.(Parallel Greek text and English translation.)

Xenophon (*c*.370s BC) *Symposium*, trans. O.J. Todd, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1923.(Parallel Greek text and English translation.)

Xenophon (perhaps *c*.370 BC) *Hiero*, trans. E.C. Marchant, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925.(Parallel Greek text and English translation.)

Xenophon (*c*. 360s BC) *Memorabilia*, trans. E.C. Marchant, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1923.(Parallel Greek text and English translation.)

Xenophon (*c*. 360s BC) *Oeconomicus*, trans. E.C. Marchant, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1923.(Parallel Greek text and English translation.)

Xenophon (*c*.361 BC) *Cyropaedia*, trans. W. Miller, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1914.(Parallel Greek text and English translation.)

Xenophon (*c*.358 BC) *Hellenica*, trans. C.L. Brownson, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1918.(Parallel Greek text and English translation.)

References and further reading

Anderson, J.K. (1974) *Xenophon*, London: Duckworth.(A useful account of Xenophon's life and times, with a typical twentieth-century prejudice against him as a serious thinker.)

Gera, D.L. (1993) *Xenophon's Cyropaedia*, Oxford: Oxford University Press.(A detailed study of the literary context and accomplishment of the *Cyropaedia*.)

Joel, K. (1893-1901) *Der Echte und der Xenophontische Sokrates*, Berlin: Gaertner.(In German, a massively documented argument against the historicity of Xenophon's portrait of Socrates.)

Momigliano, A. (1993) *The Development of Greek Biography*, Cambridge, MA: Harvard University Press, 1971.

(A short but pithy discussion of the literary genres within which Xenophon worked; chapter 3 is especially relevant.)

Morrison, D. (1987) 'On Professor Vlastos' Xenophon', *Ancient Philosophy* 7: 9-22.(A useful corrective to Gregory Vlastos' influential and dismissive attitude to Xenophon's Socratic writings.)

Morrison, D. (1988) *Bibliography of Editions, Translations, and Commentaries on Xenophon's Socratic Writings 1600-Present*, Pittsburgh, PA: Mathesis Publications.(An excellent reference tool with a full listing of scholarly work on Xenophon's Socratic writings.)

Strauss, L. (1991) *On Tyranny*, New York: Free Press.(This seminal work contains Strauss' close and provocative analysis of the *Hiero*, followed by a brilliant exchange between Alexandre Kojève and Strauss over the contemporary relevance of Xenophon's thought; also contains a translation of *Hiero* more careful than the 1925 Loeb edition.)

Tatum, J. (1989) *Xenophon's Imperial Fiction*, Princeton, NJ: Princeton University Press.(A stimulating literary examination of the *Cyropaedia*.)

Vander Waerdt, P.A. (ed.) (1994) *The Socratic Movement*, Ithaca, NY: Cornell University Press.(A collection of essays on Socrates' influence, the following of which are especially relevant: P.A. Vander Waerdt's Introduction and D. Clay for the issues on Xenophon's literary and philosophical context discussed in §2 of this entry; D.R. Morrison and J.A. Stevens on the themes of §3 on virtue and education; and D.K. O'Connor and T.L. Pangle on the themes of §4 on Socrates and Cyrus as exemplars of self-sufficiency.)

Vlastos, G. (1971) 'The Paradox of Socrates', in *The Philosophy of Socrates*, Garden City, NY: Anchor Books, 1-21.(Vlastos' widely influential dismissal of Xenophon's Socratic writings occurs in pages 1-3, which are the focus of the rebuttal in Morrison (1987).)

Xin (heart-and-mind)

In the West, questions of the distinguishability of mind and matter and of rationality and emotion or sentiment are central issues within the philosophy of mind. Neither of these topics is of much interest, however, to the mainstream of Chinese thought. On the one hand, the notion of qi*, the vital energizing field that constitutes all natural processes, renders discussions of the relevance of any psychophysical dualism moot. On the other hand,* xin*, normally translated as 'heart-and-mind', preludes the assumption of distinctions between thinking and feeling, or idea and affect.* Xin *is often translated simply as 'heart', but since it is the seat of thinking and judgment, the notion of mind must be included in its characterization if the term is to be properly understood. Indeed, what we often think of as 'will' or 'intention' is likewise included in the notion of* xin*.*

In the classical period, the heart (*xin*) as the seat of thinking is considered to be an organ similar to the other sensing organs, but with the advantage of being able to think: 'Organs such as those of hearing and of sight, being unable to think, can be misled by external things.… But the heart does think. Only by thinking will the answer be found' (*Mengzi* 6A15). Such a characterization is not, of course, unique to the Chinese; the classical Hebrews also believed the heart to be a seat of thought and action.

The interpenetration of idea, intention and affect expressed in the notion of *xin* entails the conclusion that thinking is never a dispassionate speculative enterprise but involves normative judgments which assess the relative merit of the sensations, inclinations and appetites that interpenetrate our experience of the world and ourselves. Since appetites and ideas are always clothed with emotion, they are to be understood, more often than not, as dispositions to act. (We have here some basis for understanding what is often thought to be the paternalistic desire of Chinese governments, from the classical period to the present, to protect the people from the 'disruptive' consequences of ideas.)

Another implication of the unity of feeling and thinking is the practical orientation of most of Chinese thought. If ideas are dispositions to act, what might be thought of as theories are little more than wholesale practical recommendations. Thus it is most difficult in Chinese cultures to find contexts within which the separation of theoretical and practical activities would prevail. When, for example, Confucius said, 'at fifteen my heart-and-mind was set upon learning' (*Analects* 2.4), he was indicating his commitment to an ethical regimen aimed at self-realization. Thinking and learning are, within the Chinese tradition, oriented to the practical ends of the moral life (see Self-cultivation in Chinese philosophy). As Mencius observes: 'For a person to realize fully one's heart-and-mind is to realize fully one's nature and character' (*Mencius* 7A1).

Western people are accustomed to think of efforts aimed at moral perfection (see Perfectionism) as involving a struggle between reason and passion, or between what we believe we ought to do and an obstreperous will that frustrates the enactment of that belief, or in the words of St Paul: 'The good that I would do I do not do, and the evil that I would not do, that I do.' In the Chinese tradition there is little such internal conflict involved in ethical development. The unpartitioned self characterized by *xin* means that it is unlikely that we should find Hamlets or St Pauls prominent among the Chinese.

If, however, the conflict associated with self-realization is not between heart and mind, what are the dynamics of moral development? If the problematic of unrealized selfhood does not entail the self divided against itself, what is the source and nature of the disturbance that the moral discipline is meant to overcome? If it is not located primarily within the soul, it can only be a disturbance in the relationships which constitute the self in its interactions with external things. 'The stillness of the sage is not a matter of his saying: "It is good to be still!" and thus he is still. He is still because none of the myriad things are able to agitate his heart and mind' (*Zhuangzi* 13).

It is precisely *not* through an internal struggle of reason against the passions, but through mirroring the things of the world as they are in their relatedness to us, that we reach a state in which 'none of the myriad things is able to agitate' our hearts-and-minds. In other words, we defer to the integrity of those things which contextualize us, thus establishing a frictionless relationship with them.

See also: Chinese philosophy; Confucian philosophy, Chinese; Intention; Mencius; Self-cultivation in Chinese philosophy; Self-realization; Zhuangzi

DAVID L. HALL
ROGER T. AMES

References and further reading

Analects (*c*. 4th-2nd century BC?), trans. D.C. Lau, *Confucius: The Analects*, Harmondsworth: Penguin Classics, 1979. (An authoritative translation.)

Mengzi (*c*. 2nd century BC-2nd century AD?), trans. D.C. Lau, Harmondsworth: Penguin Classics, 1970. (An authoritative translation.)

Zhuangzi (*c*. 4th century BC?), trans. A.C. Graham, *Chuang-tzu: The Seven Inner Chapters and Other Writings From the Book 'Chuang-tzu'*, London: Allen & Unwin, 1981. (The most philosophically sensitive translation.)

Xin (trustworthiness)

The earliest and basic sense of xin *is 'being true to one's word'. While one's words can be* xin *(that is, worthy of trust), in most cases* xin *indicates an excellence of character; it is thought to be the central virtue governing the relationship between friends. Since* xin *is primarily a virtue, its exercise involves practical reasoning and not a mechanical adherence to one's promises.* Xin *later was added to an original list of four cardinal Confucian virtues, though its status as a distinct disposition remained controversial. Buddhist thinkers broadened the sense of the term to include religious faith. This innovation in turn influenced certain neo-Confucian thinkers who then talked about the need to* xin *(have faith in) one's innate moral faculty.*

The basic sense of being trustworthy and reliable is well-attested in some of the earliest appearances of the character *xin*. In the *Zuozhuan* (Zuo Annals), a true and reliable worshipper is described as *xin* (Duke Zhuang, 10th year), as is a son who can be counted on to uphold and maintain his father's orders (Duke Xi, 7th year). In both cases, *xin* is used to describe someone who can be trusted to fulfil and not overstep their proper religious and ethical obligations.

In a variety of early texts, *xin* is linked with another character to form the pair *zhong xin* (loyalty and trustworthiness). These two virtues are seen as cognates: those who are *zhong* (loyal) steadfastly maintain their obligations, which makes them *xin* (trustworthy). Again, however, they are not 'loyal' and 'trustworthy' in a blind or mechanistic fashion. Their allegiance is to the moral way.

During the Han dynasty, *xin* was added to the four cardinal virtues of humanity, righteousness, propriety and wisdom (see Zhi; Confucian philosophy, Chinese §§5-6) first put forth by Mencius (*Mengzi* 2A6). This addition posed a problem for later Confucian philosophers, as the virtue *xin*, unlike the other four, was not associated with a corresponding nascent tendency of the human heart-mind. Cheng Yi accounted for this asymmetry by regarding *xin* as a kind of meta-virtue which supervenes upon and emerges from the original four. Zhu Xi believed that it was the ground or necessary constituent of the other virtues.

Xin was used as a verb to mean 'to believe or trust' (for example, in the *Daodejing* 49) and hence could mean faith in the sense of mental assent. However, since indigenous Chinese religion and ethics were strongly naturalistic, there was little room for a notion of 'faith' in the sense of *fiducia*. However, with the arrival of Buddhism (see Buddhist philosophy, Chinese), it came to mean 'belief' or 'faith'; in this latter, voluntaristic sense, in a teaching, text, order or deity. This is clearly its meaning in the title of Buddhist works such as the *Dasheng qixinlun (The Awakening of Faith in Mahāyāna)* (see Awakening of Faith in Mahāyāna).

This sense of *xin* as the attitude or spiritual virtue which provides one with the motivation to remain true to what is right and good is seen in later Confucian thinkers such as Wang Yangming. Wang believed that all people are endowed with a complete and perfect moral faculty called *liangzhi* (pure knowing) which, given full rein, had the power to eliminate all misguided tendencies and steer one unerringly to what is right and good. The challenge was to generate a full and active commitment to follow one's *liangzhi*, a task which Wang once described as 'creating something out of nothing'. His advice to simply *xin* (have faith in) *liangzhi* is similar to the kind of religious faith familiar in Christian thinkers such as Luther.

See also: Awakening of Faith in Mahāyāna; Neo-Confucian philosophy; Confucius; Daxue; Trust; Wang Yangming

PHILIP J. IVANHOE

References and further reading

Zuozhuan (Zuo Annals), trans. J. Legge, *The Ch'un Ts'ew with the Tso Chuen*, in *The Chinese Classics*, vol. 5, Hong Kong: Hong Kong University Press, 1970. (A translation of the *Zuozhuan*.)

Xing

Xing *is conventionally translated as ‘nature’ or ‘human nature’. Some read* xing *as meaning a heavenly endowed tendency, directionality, or potentiality of growth in the individual. On this essentialistic reading,* xing *is an innate and unchanging ‘given’, a defining condition of all human beings. Others have given a historicist interpretation of* xing*, reading it as an achievement concept rather than as a given. In this view,* xing *is derived from, and is a refinement on,* sheng*, denoting the entire process of birth, growth and ultimate demise that constitutes the life of a living creature.*

There is some controversy in the published literature on the meaning of *xing*, conventionally translated as ‘nature’, or ‘human nature’. On one side, there are those who read *xing* as a heavenly endowed tendency, directionality, or potentiality of growth in the individual, and see it as similar to the Greek *phuo* (to grow) and the Latin *nascor* (to be born) (see Human Nature; Essentialism). On the other side, there are those who have an historicist interpretation of *xing*, reading it as an achievement concept rather than as a given. A.C. Graham (1990) takes issue with the simple identification of the classical Chinese concept of *xing* with the familiar conception of ‘nature’ as something ‘inborn and innate’, those qualities which a thing has to start with. His claim is that the dynamic thrust of *xing* has not been adequately noticed.

As a corrective on his own earlier work, Graham argues that early Chinese thinkers who discuss *xing* seldom seem to be thinking of fixed qualities going back to a thing’s origin, but rather to the process of its maturation within a specific context. *Xing* thus understood covers the career of a person’s existence, denoting the entire process of becoming human. Strictly speaking, a person is not a sort of *being* but first and foremost a *doing* or *making* and, only derivatively and retrospectively, something done.

If we were going to speculate on why terms such as *xing* tend generally to be more dynamic in meaning than their Western equivalents (which indeed seems to be the case), we might want to reflect on the implications of cosmogonic speculation, a signal feature of the Western tradition that is made important in this analysis by its absence in classical Chinese cosmology. Where something within a single-ordered *kosmos* is shaped and invested by an external originative principle, the most fully creative act lies in the creator’s endowment of a given potential: the creature’s subsequent actualization of that potential is derivative. Where a phenomenon is initiated by, and dependent upon, some externally derived or ‘given’ creative principle for the ‘nature’ of its existence and, put another way, where it is other than self-generative (*ziran*), the creative contribution of that phenomenon tends to be diminished. In the absence of cosmogonic beginning, on the other hand, the power of creativity and the responsibility for creative product reside more broadly in the phenomena themselves in their ongoing interactive processes of becoming (see Tian; Dao).

The difference between the ‘nature’ of a thing in a cosmogonic tradition and its *xing* in a non-cosmogonic cosmology is suggested by the kinds of questions that each culture’s philosophers ask. Cosmogonic concern generates metaphysical questions, a search for essential principles. How did the cosmos begin? What are its first principles? What are the fundamental elements from out of which it was constructed? What is the origin of the existence and growth of natural phenomena? The search is for the One behind the many.

Genealogical cosmology, on the other hand, will generate primarily historical and rhetorical questions: who and what are our historical antecedents that have given us our present definition? What are their achievements that we can appropriate to enculturate ourselves? How can we further cultivate ourselves so as to contribute to the appropriated tradition as it is embodied in our contemporary exemplars? How can we turn this historical and cultural interdependence to maximum benefit? The thinker’s role in the non-cosmogonic tradition, then, will not be as much to discover an answer as to create a model of humanity that is persuasive, and that evokes emulation.

A related implication of this distinction between a cosmogonic and non-cosmogonic worldview is that in the absence of some overarching *archē* (beginning) as an explanation of the creative process, and under conditions which are thus ‘anarchic’ in the philosophic sense of this term (see Archē), although *xing* might indeed refer to ‘kinds’, genus and species as categories would be dependent upon generalizations made by analogy among *sui generis* phenomena. Difference is prior to identified similarities. Certain things to which *xing* is applied - water

and rocks, for example - are not over their respective careers marked by growth and cultivation, and hence it makes little sense to speak of them in terms of starting conditions and mature state. The *xing* of such things remains relatively constant. However, the human being - that phenomenon most given to cultivation and refinement - is a different case.

While the human *xing* might include certain generalizable conditions that define it at birth, in its more important aspects it seems to refer to what is existentially achieved. In defining the human *xing*, the relatively constant and uninteresting tendencies which constrain the creative project of personal development are outweighed by the massive transformative process that occurs between the ‘stirring’ or ‘germination’ of the initial fundament and the full-blown creative achievement. What is ‘innate’ in the *xing* of persons is simply the propensity for growth, cultivation and refinement. *Xing*, then, denotes a human capacity for radical changeability that is qualitatively productive.

Further, *xing* is realized *in situ*. It is a dynamic process conditioned by its particular context. Stated more explicitly, the human *xing* is a creative process that can only be understood situationally as the outcome of specific interdependent relationships. It at once refers to the continuing existence of a particular thing itself, and also to that in one thing which continues the life and culture of other things.

See also: Archē; Chinese philosophy; Cosmology; Dao; De; Human nature; Tian

DAVID L. HALL
ROGER T. AMES

References and further reading

Ames, R.T. (1991) ‘The Mencian Concept of *Ren xing*: Does It Mean "Human Nature?"’ in H. Rosemont, Jr (ed.) *Chinese Texts and Philosophical Contexts*, LaSalle, IL: Open Court. (A further development of A.C. Graham’s analysis.)

Bloom, I. (1994) ‘Mencian Arguments on Human Nature (*jen-hsing*)’, *Philosophy East and West* 44 (1): 19-53. (A summary statement of the innatist position.)

Graham, A.C. (1990) ‘The Background of the Mencian Theory of Human Nature’, in A.C. Graham (ed.) *Studies in Chinese Philosophy and Philosophical Literature*, Albany, NY: State University of New York Press. (A rejection of the innatist position.)

Xunzi (*fl.* 298-238 BC)

Xunzi is one of the most brilliant Confucian thinkers of ancient China. His works display wide-ranging interest in such topics as the relation between morality and human nature, the ideal of the good human life, the nature of ethical discourse and argumentation, the ethical uses of history, moral education and personal cultivation. Because of the comprehensive and systematic character of his philosophical concerns, Xunzi is sometimes compared to Aristotle. Noteworthy is his emphasis on li, *or rules of proper conduct, and the holistic character of* dao, *the Confucian ideal of the good human life. He criticized other philosophers not because of their mistakes, but because of their preoccupation with one aspect of dao to the exclusion of others.*

Xunzi (Hsün Tzu) is also known as Xun Kuang and Xun Qing. His fundamental doctrines are contained in the *Xunzi*. Like his predecessors, Confucius and Mencius, Xunzi focuses on developing a conception of a well-ordered society governed by an enlightened, virtuous and sagely ruler, a ruler who cares for people's welfare and moral character based on the cultivation of *ren* (benevolence), *li* (propriety) and *yi* (rightness, righteousness). The enlightened ruler is one who is good at organizing the people in society in accordance with *ren*, *li* and *yi*.

The *li*, as formal prescriptions or rules of proper conduct, are especially emphasized by Xunzi. While they represent an inherited ethical tradition, they do not always provide adequate guidance in the perplexing, exigent situations of human life. As markers of *dao* (the Way), 'the *li* provide models, but no explanations' (*Xunzi* 1, in Watson 1963: 20); basically they provide general guidelines delimiting the boundary of proper behaviour and individual responsibility. Perhaps even more important, the *li* 'provide for satisfaction of people's pursuit' of their individual desires or interest (*Xunzi* 19). In this respect, the *li* may be compared to the laws of contracts or of wills, enabling the agents to fulfil their desires effectively. Of course the key question, whether the existing *li* in fact perform their proper functions, depends on whether they satisfy the requirements of *yi* (that is, whether they are the right sort of rules) and the extent to which they ennoble human character in terms of *ren* (benevolence) and cultural refinement (*wen*) (Cua 1989) (see Confucian philosophy, Chinese).

For Xunzi, the regulative or delimiting function of *li* is important in ensuring a good, stable social and political order. The necessity of the *li* is due to the problematic motivational structure of human nature. This is the force of his rather misleading remark that 'the human's nature is evil; his goodness is the product of conscious activity' (*Xunzi* 23, in Watson 1963: 157). In terms of its inborn motivational structure, consisting of feelings and desires, every human is a self-seeking animal actuated by fondness for personal gain. In the absence of regulative rules, this propensity may be viewed as bad or evil because it tends to lead to conflict and disorder in society. What is good is that which is 'upright, reasonable and orderly' and what is bad is 'that which is partial, irresponsible and chaotic' (*Xunzi* 23, in Watson 1963: 162). In addition, scarcity of resources to satisfy everyone's desires will lead to competition which requires *li*-regulation as well as a sense of *yi* (rightness).

In this light, inborn human nature is value-neutral, not intrinsically bad or evil but requiring regulation and moulding. This nature provides the basic materials for ethical transformation as does clay for the making of pottery. Indeed, in respect to human nature, the sage and ordinary humans are alike. Every man can become a sage: 'What makes the sage emperor Yu a Yu was the fact that he practised *ren* and *yi* and complied with proper rules and standards of conduct.' Moreover, the rationales of *ren*, *yi* and proper rules and standards can be known and practised by ordinary persons, because they possess the capability of obtaining such knowledge and acting accordingly. It must be admitted that not every person can actually become a sage, for they are not willing to make the effort, especially in moral learning (*Xunzi* 23).

Xunzi's essay 'Encouraging Learning' presents a sketch of his conception of moral learning as a continuing process that ceases only with death. Principally, the learning is devoted to acquiring knowledge of various classics such as the *Shijing (Book of Songs)*, the *Shujing (Book of Documents)*, the *Lijing (Classic of Rites)* and the *Spring and Autumn Annals*, as well as the established musical text. While recitation is stressed, Xunzi calls attention to understanding and insight concerning the concrete significance of these works in human life. Initially, learning to be an ethical person depends on teachers of exemplary ethical character, but this learning continues throughout one's life. The proximate objective is to become a scholar-official; the ultimate objective is to become a sage. Notably, the classics provide an introduction to what Xunzi considers as the best Ru or Confucian tradition.

Xunzi's own writings attest to the importance of acquiring historical knowledge (Cua 1985b). Special stress is given to four different functions of the ethical uses of history in discourse: pedagogical, rhetorical, elucidative and justificatory uses (see History, philosophy of).

These uses of history provide a better guidance than religious beliefs and superstition about *tian* (Heaven) as dispensing fortunes and misfortunes (see Tian). *Tian* is a natural order of regularities, not an order governed by a supreme, personal being. The proper attitude toward strange events or phenomena is wonder and awe (*Xunzi* 17). Practices such as prayer and divination are based on superstitious beliefs. However, the *li* or rites of mourning and sacrifices are ethically acceptable and, in fact, to be encouraged. Their significance lies in their function of satisfying human yearnings and providing an avenue for the proper expression of reverence, honour and affection for the dead. Moreover, these rites contribute to the Confucian conception of cultural refinement (*wen*) (*Xunzi* 19). Moral learning culminates in the attainment of completeness and purity, that is, a state of ethical integrity (*Xunzi* I) (see Moral development).

Xunzi frequently uses the word *dao* to express his ideal of the good human life as a whole. The significance of *dao* as a generic term lies in a practical understanding of *ren*, *li* and *yi*. However, knowing *dao* requires clarity of mind (*ming*). When a clear mind is guided by reason, it will be free from all sorts of factors that obscure (*bi*) the clarity of vision. One must be especially watchful of the contextual significance of such distinctions as between desires and aversions, immediate and distant consequences, the past and the present. Failure to weigh the relevance of one item of the distinction is liable to blind the agent to consideration of the significance of the other item. Basically, distinctions are relative to the purpose and context of discourse; they are not exclusive disjunctions or absolute dichotomies. Since distinctions in general are the products of the mind's intellectual function at the service of cognition and action, their utility in a particular situation is also relative to purpose and context. Even in cases where the utility of the distinction is not in question, the agent must render a reasoned judgement concerning the significance of each item in the distinction. Obscuration (*bi*) of the mind arises when the mind exclusively attends to the significance of one item without proper regard for that of the other (see Perception, epistemic issues in).

Even philosophers are susceptible to *bi*. Mozi, for example, exaggerates the importance of utility without understanding the importance of culture or the beauty of forms; Zhuangzi is preoccupied with *tian* (Heaven) without regard for humanity. In general, ordinary people as well as philosophers tend to see only 'one corner' of *dao* and thus fail to appreciate its holistic character. Only an enlightened sage 'confronts all things and weighs them impartially on a balance' (*Xunzi* 21; Cua 1993).

Today, Xunzi is widely acknowledged to be one of the greatest Chinese philosophers. Apart from his critical and coherent exposition of ancient Confucianism, he is also distinguished for his insights into the nature of ethical knowledge and argumentation (Cua 1985a).

See also: Confucian philosophy, Chinese; Dao; Ethics; Law and ritual in Chinese philosophy; Moral development; Self-cultivation in Chinese philosophy

A.S. CUA

List of works

Xunzi (*c*.298-238 BC) *Xunzi*, ed. Li Disheng, *Xunzi chishi*, Taibei: Xuesheng, 1979; trans. B. Watson, *Hsün Tzu [Xunzi]: Basic Writings*, New York: Columbia University Press, 1963; trans. H.H. Dubs, *The Works of Hsüntze [Xunzi]*, Taibei: Chengwen, 1966; ed. and trans. J. Knoblock, *Xunzi: A Translation and Study of the Complete Works*, 3 vols, Stanford, CA: Stanford University Press, 1988, 1990, 1994.(Li is a good modern Chinese annotated edition. The Knoblock edition contains long, valuable introductions in volumes 1 and 2, and extensive bibliographies in volumes 1 and 3.)

References and further reading

Chan Wing-tsit. and Fu, C. (1978) *Guide to Chinese Philosophy*, Boston, MA: C.K. Hall.(A contemporary bibliography.)

Cua, A.S. (1985a) *Ethical Argumentation: A Study in Hsün Tzu's [Xunzi's] Moral Epistemology*, Honolulu, HI: University of Hawaii Press.(A Xunzi-based philosophical study of Confucian ethical argumentation with emphasis on the character of the participants, standards of competence, uses of definition and diagnosis of erroneous ethical beliefs.)

Cua, A.S. (1985b) ‘Ethical Uses of the Past in Early Confucianism: The Case of Hsün Tzu [Xunzi]’, *Philosophy East and West* 35: 133-56.(A study of Xunzi’s uses of the historical appeal from the point of view of Confucian discource, focusing its pedagogical, rhetorical, elucidative and evaluative functions.)

Cua, A.S. (1989) ‘The Concept of *Li* in Confucian Moral Theory’, in R.E. Allinson (ed.) *Understanding the Chinese Mind: The Philosophical Roots*, Hong Kong: Oxford University Press.(A study of three basic functions of *li* and their ethical roots.)

Cua, A.S. (1993) ‘The Possibility of Ethical Knowledge: Reflections on a Theme in the *Hsün Tzu [Xunzi]*’, in H. Lenk and G. Paul (eds) *Epistemological Issues in Ancient Chinese Philosophy*, Albany, NY: State University of New York Press.(A study of possibility of knowing dao and the problem of cognitive blindness.)

Dubs, H.H. (1927) *Hsüntze [Xunzi]: The Moulder of Ancient Confucianism*, London: Arthur Probsthain.(The first comprehensive English study of Xunzi.)

Yang Xiong (53 BC-AD 18)

Master Yang Xiong, the first Confucian classicist and the greatest of the pre-Song metaphysicians, is best known for two major philosophical works, the Taixuanjing (Canon of Supreme Mystery) and the Fayan (Model Sayings). Both works explore the interaction between significant cosmic and social patterns by explicit reference to earlier canonical traditions.

Master Yang came to Changan (modern Xian), then the capital of the Western Han dynasty, in his early thirties. As the foremost poet of his age, Yang Xiong was in effect appointed poet laureate to the court in 10 BC. In middle age, after the death of his beloved son, Yang experienced a sense of profound revulsion for his earlier poetic efforts. Condemning court poetry as inherently frivolous, if not immoral, Yang turned to composing works of philosophy.

His first such work was the *Taixuanjing (Canon of Supreme Mystery)*, which like its prototype, the *Yijing (Book of Changes)*, correlates the significant patterns of the universe with different combinations of solid and broken lines accompanied by brief texts and ten autocommentaries. However, the *Taixuanjing* adjusts the structure and imagery of the *Yijing* to better address Han preoccupations with correlative thinking, fate and time. As the first grand synthesis of Chinese thought, skilfully weaving together elements of early Confucianism, classical Daoism, *yin-yang* Five Phases theory (see Yin-yang), alchemy, and astrology into an organic whole, the *Taixuanjing* came to occupy a place in Chinese intellectual history roughly comparable to that of the *Summa theologiae* of Thomas Aquinas in the West. Its most important contribution is its account of individual human destiny, which is said to result from the interplay of four major factors: time (meaning 'present opportunity'), virtue (or 'character'), tools (including training) and position (societal and physical). Of these four factors, only virtue is completely subject to human control, so Yang identifies adherence to the Good as defined in the Confucian 'Way' as the single most reliable method of improving individual fate. In the course of outlining this theory of fate, the *Taixuanjing* pinpoints the logical fallacies inherent in various philosophical positions that competed with Confucianism for general approval during Yang's time.

Following a draft of the *Taixuanjing* before 2 BC, there came at least four lengthy philosophical poems whose style parodies the prose-poems of Yang's youth. Two of these *fu* are dedicated to the defence of his provocative masterwork, the *Taixuanjing*. Around AD 12, Yang finished the *Fayan (Model Sayings)*, which adopts both the format and style of the Confucian *Analects* to provide a relatively straightforward catechism for the would-be sage (see Confucius). The *Fayan* evaluates the conflicting goals of immortality, fame, power and scholarship to disentangle prevailing notions that confused the 'good life' with ideas of the Good. As its main theme is the need to immerse oneself in the model of the former sages, the *Fayan*, like the *Analects*, devotes considerable attention to the reassessment of historical figures whose conduct or writings were considered worthy of emulation (see History, Chinese theories of).

When Yang died, he left unfinished a third remarkable work on significant pattern, the *Fangyan (Regional Expressions)*. As the first Chinese dialect dictionary, the *Fangyan* is far more than a sourcebook of philological glosses culled from disparate sources; it is a synchronical word list that focuses on rare expressions in an attempt to trace linguistic boundaries and connections. In a surviving letter, Yang justified the enormous task of compiling this text in fairly standard Confucian terms: Since regional cultures evolve from complex factors, the wise ruler must have such texts at his disposal so that government policies may be suitably adjusted to varying locales.

Thanks to the volume and quality of his output, Yang's stature as Confucian master remained unchallenged until the Song period. At that time a few prominent neo-Confucian thinkers (including Zhu Xi), in an effort to disengage themselves from certain pre-Song moral constructs, objected to Yang's considerable influence and denounced his character, cosmological vision and eclecticism. Nonetheless, as most Chinese philosophical works dating to the first millennium have been lost, the *Taixuanjing* and *Fayan*, two undisputed classics, together provide one of the best keys to understanding early Confucian ideology in its formative stage.

See also: Confucius; Confucian philosophy, Chinese; Daoist philosophy; Yijing; Yin-yang; Zhu Xi

MICHAEL NYLAN

List of works

Yang Xiong (*c.* 4 BC) *Taixuanjing (Canon of Supreme Mystery)*, trans. M. Nylan, Albany, NY: State University of New York Press, 1993; ed. Zheng Wangeng, *Taixuan jiaoshi (Collated Explications for the Supreme Mystery)*, Beijing: Teacher's Normal University, 1989.(An inquiry into the relation of fate and virtue. Zheng is the best annotated edition of the classic text, with notes in modern Chinese.)

Yang Xiong (*c.* AD 12) *Fayan (Model Sayings)*, trans. E. von Zach, *Yang Hsiung's Fa yen: Worte strenger Ermahnung*, in Sinologische Beiträge 4, Batavia: Drukkerji Lux, 1939; B. Belpaire, *Le catechisme philosophique de Yang-Hiong-tsé*, Brussels: Éditions de L'Occident, 1960.(Analecta devoted to historical figures and philosophical concepts.)

Yang Xiong (before AD 18) *Fangyan (Regional Expressions)*, in P.L.-M. Serruys, *The Chinese Dialects of Han Time According to Fang Yen*, Berkeley, CA: University of California Press, 1959.(There are several editions of this work, all of which coincide.)

References and further reading

Knechtges, D. (1981) *The Han shu Biography of Yang Xiong (53 B.C.-A.D. 18)*, Tempe, AZ: Arizona State University.(Translates Yang's *Hanshu* biography, based on his autobiography.)

Mitarai Masaru (1957) 'YōYūto Taigen' (Yang Hsiung and the Supreme Mystery), *Shinagaku kenkyū* 18: 22-32. (Emphasizes the originality of the *Tai xuan jing*.)

Suzuki Yoshijirō (1964) *Taigen no kenkyū(Research on the Supreme Mystery)*, Tokyo: Meitokusha.(Explains the *Tai xuan jing* in light of Han cosmological theories.)

Yangzhu (5th-4th century BC)

Yangzhu, detested by the Confucians, is important in the Chinese tradition for initiating the explicit discussion of human nature. He focuses on the thesis that human nature has no inherent ethical or mystical qualities; instead, there is simply an innate tendency to live a long life, a tendency that must be carefully nurtured by a rational regulation of sense stimulation and by the avoidance of any of the entanglements incumbent in a life of working for the good of human society.

Yangzhu is commonly dated to a generation or two before Mencius, who said of him that he was so egotistical he would not even pull out one hair to benefit human society. There are no extant writings of Yangzhu; we know of his ideas through the writings of his followers included in five essays in the *Lushi chunqiu (Lu's Spring and Autumn Annals)* in the third century BC, and in four chapters from the *Zhuangzi* that are dated to about 205 BC (see Lushi Chunqiu; Zhuangzi). There is no evidence of his intellectual lineage continuing after this time.

Yangzhu is known for initiating in Chinese thought the explicit discussion of human nature (see Xing), and also for three related teachings: 'keeping one's nature intact', 'protecting one's genuineness' and 'not letting the body be tied by external things'. For Yang, human nature at its most basic level consists in the spontaneous tendency for humans to live a long life if undisturbed by external things, since the various 'physiological' systems of *qi* that constitute the human organism tend to function harmoniously provided they are unimpaired and sufficiently nourished by the proper degree of sensory stimulation (see Qi).

A related aspect of human nature is the desire of the various sense faculties for their objects. It is this that helps to maintain the health and development of the organism, enabling it to realize its inherent tendency for longevity. However the senses, themselves, need to be regulated and limited to only a suitable amount of stimulation; there is a naturally suitable degree of sense-stimulation which is conducive to health and development. This degree must be rationally determined by the Sage; the senses on their own cannot do this. Excessive stimulation, say the Yangists, reduces the individual's finite supply of the refined form of *qi* called *jing* (vital essence) needed to maintain health and vitality.

Yangist political philosophy is meagre, but straightforward. One should not seek after fame, wealth and power, all of which are far beyond essential needs. One should never do anything to impair one's inherent tendency to live a long and satisfied life. Only when a person knows this do they understand 'the essentials of our nature and destiny' and are then fit to rule over others.

Yangzhu's challenge to the social interpretation of human nature of the Confucians consisted in the primacy he placed on the maintenance of the individual. In his advocacy of personal cultivation over social concern, he initiated an individualistic counteractive to the often overwhelming pressure toward social conformity in China that was later continued by Daoist and Buddhist thought. However unlike these later schools, there was absolutely nothing mystical in the teachings of Yang and his followers, even though some think them related to the early Daoists.

See also: Confucian philosophy, Chinese; Daoist philosophy; Self-cultivation in Chinese philosophy; Qi

H.D. ROTH

References and further reading

Fung Yu-lan (1952) *A History of Chinese Philosophy*, trans. D. Bodde, Princeton, NJ: Princeton University Press, 132-43.(Presents a detailed but rather difficult analysis of the ideas of Yangzhu, valuable for its discussion of possible links between him and early Daoists.)

Graham, A.C. (1967) 'The Background of the Mencian Theory of Human Nature', *Tsing Hua Journal of Chinese Studies* 6 (1-2): 215-74; repr. in A.C. Graham, *Studies in Chinese Philosophy and Philosophical Literature*, Albany, NY: State University of New York Press, 1986, 7-66.(Contains a useful summary of Yangzhu's ideas on human nature as an important part of the intellectual context for Mencius' development of the concept.)

Graham, A.C. (1981) *Chuang Tzu: The Inner Chapters*, London: George Allen & Unwin. (Provides the authoritative English translation of the Yangist essays in the *Zhuangzi*.)

Graham, A.C. (1989) *Disputers of the Tao: Philosophical Argumentation in Ancient China*, LaSalle, IL: Open

Court, 53-64.(This is the best introduction available to Yangzhu’s thought.)

Knoblock, J. and Riegel, J. (trans.) (1997) *The Almanac of Lü Buwei*, Stanford, CA: Stanford University Press. (Translation of the *Lushi chunqiu*.)

Yi Hwang (1501-70)

Yi Hwang, also known by his honorific name T'oegye, is one of the two most honoured thinkers of the Korean neo-Confucian tradition. His fully balanced and integral grasp of the complex philosophical neo-Confucian synthesis spun by Zhu Xi during China's Song dynasty marks the tradition's arrival at full maturity in Korea. His 'Four-Seven Debate' with Ki Taesûng established a distinctive problematic that strongly oriented Korean neo-Confucian thought towards exacting investigation of critical issues regarding the juncture of metaphysics and their all-important application in describing the inner life of the human heart-and-mind.

Yi Hwang, also known by his honorific name T'oegye, was born of a relatively modest aristocratic lineage in the village of Ongyeri, near Andong in Kyongsan province, about 200 kilometres southwest of Seoul. He took the civil service examinations, and served in government for a number of years, but his true longing was for a life of quiet study, reflection, and self-cultivation. He retired from office in his late forties to pursue his dream, and the following two decades were a period of tremendous productivity in spite of frequent recalls to office as his fame as a scholar and teacher grew.

Neo-Confucianism was adopted as the official orthodoxy at the foundation of the Chosôn dynasty in 1392. It was a rich synthesis, containing a metaphysical system of Daoist proportions and a meditative cultivation of consciousness reminiscent of Buddhist practice which Zhu Xi and other early neo-Confucians had woven about the core of traditional Confucian concerns for government and proper social ethics (see Neo-Confucian philosophy). It provided wide scope for varied and uneven development. During the first century of the Chosôn dynasty, activists in government focused on institutional reform while, far from the capital, scholars in the countryside concentrated on the more meditative features of neo-Confucian learning and self-cultivation. The differing orientations crystallized into bloody clashes and purges by the end of the fourteenth century as young men steeped in moral rigorism began to move from the countryside into government.

Yi Hwang's comprehensive grasp of Zhu Xi's thought clarified the balance between activity and quietism, government and retired self-cultivation, and by the end of his life it was his disciples who were moving into high government positions. A year before his death he crystallized and presented to the king his understanding of the way in which metaphysics and psychological structures inform ascetical theory and eventuate in the conduct of daily life. This work, the *Sônghak sipdo (Ten Diagrams on Sage Learning)*, became one of the most famous and influential works of Korean neo-Confucianism. After Yi Hwang, it was no longer possible to deny the legitimacy of intensive, almost monastic devotion to study and meditative self-cultivation when the situation permitted, nor to ignore the expectation that the proper fruition of such formation should be the proper conduct of government and the ordering of society.

On the level of philosophical theory, Yi Hwang left a lasting imprint on Korean neo-Confucianism, for his 'Four-Seven debate', carried on in correspondence with a younger scholar, Ki Taesûng (1527-72), established the problematic for Korean thinkers for centuries. In particular, it centred Korean neo-Confucian reflection on questions relating to the interface of metaphysics and psychological theory. For Yi Hwang and other neo-Confucians committed to self-cultivation, this was a topic of intense concern: in the framework of neo-Confucian thought, a proper, metaphysically-grounded understanding of the structure and functioning of the psyche explains human perfection and imperfection; it is thus the foundation for any theory behind the practice of spiritual cultivation.

The Four-Seven Debate centred on an issue directly related to one of the most important advances of neo-Confucian thought, its explanation of a cause within human beings that could disrupt our natural responsiveness to situations and cause us to err. This was accomplished by the elaboration of a dualistic monism. The single *dao* or *li* (in Korean, *to* and *i*), both formative and normative, runs through all things, constituting the diverse nature of each one without losing its ultimate unity. The diversity is accomplished by pairing it with a complement, *qi* (in Korean, *ki*), a vital, active, concretizing and hence particularizing component. The unity of *i*, the substance of mind as well as the nature of all things in existence, supported the possibility of cultivating a condition of perfect responsive interrelation with all things, a Confucian sagehood equivalent to Buddhist enlightenment. *Ki*, on the other hand, provided an explanation for imperfection: its various degrees of turbidity or

purity accounted not only for the differentiation of *i* into various levels of nature, but also distorted to varying degrees the otherwise perfect responsiveness of the human mind to its surroundings (see Qi).

The Four-Seven Debate took up the question of the role of this dualism in the issuance of the feelings (see Confucian philosophy, Korean). Spontaneously good responses, such as compassion or the tendency to save a child from danger, seem different from feelings such as anger or desire, which may or may not be appropriate and always need watching. Since the nature (*i*) is normative and *ki* the explanation of disruption, an easy conclusion might be that the reliable feelings (the 'four') arise from *i* while the questionable ones (the 'seven') are rooted in *ki*. This, however, is an extremely dualistic interpretation that ignores the necessary and constant interdependence of *li* and *ki*. Good ethics becomes bad metaphysics. Emphasizing the interdependence, on the other hand, puts all our responsive feelings on exactly the same footing. Neo-Confucian cultivation, and especially meditative practice, set great store on somehow establishing a quiet contact with our purely good deep nature and nurturing the sound tendencies stemming from it. Yi Hwang conceived an approach that recognized interdependence but carved out a distinction between the two sorts of feelings, though this could not settle the question, for there were those who wanted such consistent interdependence that no such variation was satisfactory. Yi Hwang's theory, entirely within the framework of Zhu Xi's thought, was a sophisticated alternative to the radical cleavage that arose in China between the schools of Zhu Xi and Wang Yangming in connection with similar issues.

See also: Confucian philosophy, Korean; Neo-Confucian philosophy; Yi Yulgok

MICHAEL C. KALTON

List of works

Yi Hwang [T'oegye] (1501-70) *T'oegye Chônsô (The Complete Works of Yi Hwang)*, Seoul: Sônggyun'gwan Taehakkyo Taedong Munhwa Yônguwôn, 1958. (Photo reprint.)

Yi Hwang [T'oegye] (1501-70) *Chu Sô Chôryo (The Essentials of Zhu Xi's Letters)*, in Abe Yoshio (ed.) *Ilbon kakp'an Yi Toegye chônjip* (Japanese edition of the works of Yi T'oegye), Seoul: Taeil Chongp'ansa, 1975, 2 vols.(Yi Hwang's comments on Zhu Xi.)

Yi Hwang [T'oegye] (1569) *Sônghak sipdo (Ten Diagrams on Sage Learning)*, ed. and trans. M.C. Kalton, *To Become a Sage: The Ten Diagrams on Sage Learning by Yi T'oegye*, New York: Columbia University Press, 1988.(A translation of the *Sônghak sipdo*, his most famous work. Includes a long introduction dealing with Yi Hwang's life, times and work.)

References and further reading

de Bary, W.T. and Haboush, J.K. (eds) (1985) *The Rise of Neo-Confucianism in Korea*, New York: Columbia University Press.(This conference volume includes three chapters dealing with various aspects of Yi Hwang's thought.)

Kalton, M.C. (ed. and trans.) (1994) *The Four-Seven Debate: An Annotated Translation of the Most Famous Controversy in Korean Neo-Confucian Thought*, New York: State University of New York Press.(A translation of the debate correspondence between Yi Hwang and Ki Taesûng, and between Yi Yulgok and Sông Hon as well. Includes a long introduction dealing with the background, content and significance of the debate.)

Lee, P.H. *et al.* (eds) (1993) *Sources of Korean Tradition*, New York: Columbia University Press.(Includes translations from a variety of Yi Hwang's works.)

Yun Sasoon (1991) *Critical Issues in Neo-Confucian Thought: The Philosophy of Yi T'oegye*, trans. M.C. Kalton, Seoul: Korea University Press.(A sophisticated analysis of Yi Hwang's philosophy by one of Korea's leading experts on neo-Confucianism.)

Yi Kan (1677-1727)

Yi Kan was a major Korean neo-Confucian thinker. He is best remembered as a major protagonist in the Horak controversy where he opposed Han Wônjin, championing the position of the school of Yi Yulgok against Han's novel tri-level theory of nature.

The Horak controversy originated when Han Wônjin stated that even in the state of meditative quiescence, *ki* (in Chinese, *qi*) or 'material force' is present. Neo-Confucian metaphysics is a dualistic monism in which the single *dao* or pattern running through all things is concretized and individualized in real beings only through *ki*. The patterning element, known as *i* (in Chinese, *li*) and often translated as 'principle', also constitutes the innermost nature of all things; its presence as the substance of the human heart-and-mind accounts for our innate potential to respond appropriately to all things and situations (see Li; Qi).

Yi objected strongly to Han's proposition. Metaphysically, it seems the obvious consequence of the interdependence of *i* and *ki*, a theme especially emphasized in the school of Yi Yulgok, to which both Han and Yi Kan belonged. However, in addition to individuation, imperfection or turbidity in *ki* also is a distorting element that accounts for our often imperfect response to things. Han's observation seemed to imply that a quiet criminal is still just a criminal. But in the ascetical life the meditative practice of 'quiet-sitting' was commonly held to obviate the distortion of *ki* and nurture the responsiveness of the heart-and-mind in direct union with its pure, unsullied inner nature. Yi accused Han of violating the authoritative tradition regarding the pure 'original nature' and its role in spiritual cultivation.

Metaphysical theory might favour Han's position, but its implications were intolerable, especially insofar as the major rationale for neo-Confucian metaphysics is to serve as a framework for ascetical theory. In defending conventional ascetical theory and practice, Yi Kan was forced into an unconventional metaphysical position: to make the pure nature accessible without denying the ever-present *ki*, he was forced to introduce the notion that the *ki* of the heart-and-mind is naturally pure; it is only the *ki* that constitutes our physical bodies that is subject to turbidity. In quieting the body and returning to pure objectless consciousness, we enter a level in which *ki* offers no distortion to the nature, which is also the substance or ground of the heart-and-mind. Instead of a dualism of *li* and *ki*, we thus end with a dualism of *ki*. This illustrates the difficulty of accommodating realistically the moral tensions of spiritual practice in the consistently monistic interpretation of Zhu Xi as advocated in Yulgok's school.

In the closely related aspect of the controversy regarding whether the normative nature of all creatures is the same or differentiated according to species, Yi Kan championed the more conventional position against Han's novel tri-level theory of the nature. According to Yi, *i* is normative precisely as considered apart from any (distorting) limitation from *ki*, but *ki* is likewise the source of all differentiation. Hence there is only the unitary *i* of 'the single origin,' and the multitude of concrete individuals, each with their unique *ki* constitution.

See also: Confucian philosophy, Korean; Neo-Confucian philosophy; Han Wônjin

MICHAEL C. KALTON

List of works

Yi Kan (1677-1727) *Woeam Yugo (The Writings of Yi Kan)*, Seoul: Cho Yongsung, 1976, photo reprint.(Contains the collected writings of Yi Kan.)

References and further readings

There is at present no further information on Yi in western languages.

Yi Yulgok (1536-84)

Yulgok was one of the foremost neo-Confucian scholars in Korea during the Yi (Chosôn) dynasty. He is considered one of two pillars, along with Yi T'oegye, of the Korean neo-Confucian tradition. Yulgok, an active statesman and educator as well as scholar, not only compiled the theories of previous Confucian scholars of China and Korea but, more importantly, developed his own interpretations of them.

Yi Yulgok (real name Yi I; Yulgok was the village to which he retired) was known as a non-dualistic thinker in interpreting the relationship between *i* and *ki*, or 'principle' and 'material force', 'the mind of *dao*' and the 'human mind', 'the four beginnings' and 'the seven feelings'. Yulgok was neither a dualist nor a monist. For example, while he rejected the division between 'principle' and 'material force' on an ontological level, he made a clear distinction between these two concepts on a phenomenological level.

Yulgok's family was intellectually very distinguished: his father came from a long line of Confucian scholar-officials, and his mother, Sin Saimdang, was famous for her intellectual and artistic brilliance. She took Yulgok's education personally in hand and became his first teacher. Yulgok himself was a prodigy. He underwent a long series of civil service examinations, taking the first at the age of thirteen and the last when he was twenty-nine. These examinations opened opportunities to various important functions in the government, and Yulgok performed brilliantly, reaching the very highest official ranks. One examination essay written at the age of twenty-three, *Ch'ôndo ch'aek (A Treatise on the Way of Heaven)*, offered a compelling theory of the unity of heaven and humanity, and it became so important among Confucian scholars of the time that it was eventually carried as far as Ming dynasty China.

Yulgok passed his last civil service examination in 1565. In 1570, he wrote *Tongho mundap (Questions and Answers at Eastern Lake)*, dedicated to the new king Sônjo, with the purpose of instructing him in methods for realizing the ideal Confucian state. An imaginary dialogue between host and guest, cast in a question and answer format, it discusses ways to deal with disordered government, explains 'the way of the ruler' as opposed to 'the way of the subject', and juxtaposes 'the way of the king' with that of the despot. It stresses the importance of the moral cultivation of the ruler himself as a necessary first step towards good government in general. *Tongho mundap* reflects Yulgok's strong social and political concerns, and his horror of the petty political factionalism he encountered at court.

In 1571, while serving as the governor of Ch'ôngju, Yulgok wrote *Sôwôn hyangyak (Community Treaty of Sôwôn)*, a manual on local self-government aimed at teaching local people how to govern themselves. The following year he resigned from the governorship due to illness and retired. At this time he started his famous debate with the Confucian thinker Sông Hon (Ugye) on the interconnected concepts of *i* (principle) and *ki* (material force); 'the four beginnings' and 'the seven feelings'. Yulgok argued that 'the four beginnings' do not contain 'the seven feelings' (or emotions) but the seven feelings (or emotions) do contain the four beginnings. Throughout this debate Yulgok maintained a non-dualistic way of understanding the nature of the relationship between *i* and *ki*; *tosim* (the mind of *dao*) and *insim* (the human mind); and the four beginnings and the seven feelings.

Over the next four years Yulgok also wrote a number of works, addressing issues from practical government to the individual moral cultivation. He compiled *Sônghak chibyo (A Compendium of Sagely Learning)*, selections from the Four Books and the Five Classics - central works of the Confucian canon - grouped around the two themes of moral self-cultivation and proper governance of the people. The choice of selections and their presentation were uniquely Yulgok's own, based on his by then mature grasp of the Confucian and neo-Confucian traditions.

In 1582, Yulgok finished his famous book, *Kyôngmong yogyôl (A Key to Annihilating Ignorance)*, a work which had an enormous and long-lasting influence on Confucian education in Korea. An introduction delineating all the basic elements of Confucian learning and sagely discipline, it was widely used as a textbook for both beginning and advanced students in the Confucian schools. *Yukchokye (The Memorials of Six Articles)*, his last work, was a practical treatise on defence policy, outlining methods for strengthening the kingdom's defences. This work became a central work on the subject, commonly referred to by Yulgok's own slogan, 'nourish and train hundreds of thousands of soldiers'. It was while serving as Defence Minister that Yulgok died of illness in 1584.

As mirrored in his life, Yulgok's thought was divided and moved in two directions. First, there was the study of 'human nature and principle' (*sôngnihak*; in Chinese, *xinglixue*); second, there were the practical considerations of good government. In attempting to relate these two areas of concern - individual self-cultivation and sociopolitical application of the Confucian ideal - Yulgok had to confront on a practical level many of the most pressing political issues of his time: the sixteenth-century Yi dynasty was plagued by countless problems ranging from economics to defence. Yulgok tried to establish a Confucian state by realizing the Confucian social and political ideals through moral self-cultivation.

See also: Confucian philosophy, Korean; Li; Qi; Self-cultivation in Chinese philosophy; Yi Hwang

YOUNG-CHAN RO

List of works

Yi I [Yi Yulgok] (1564-84) *Yulgok chônsô (The Complete Works of Yi Yulgok),* Seoul: Sônggyun'gwan Taehakkyo Taedong Munhwa Yôn'guwon, 1958, 2 vols. (Collection of Yulgok's original writings in Chinese.)

Yi I [Yi Yulgok] (1564) *Ch'ôndo ch'aek (A Treatise on the Way of Heaven),* in *Yulgok chônsô,* Seoul: Sônggyun'gwan Taehakkyo Taedong Munhwa Yôn'guwon, 1958, 14: 54a-60b.(This is a concise and yet broad treatise on neo-Confucian metaphysics, cosmology, ethics and political thought, written for a civil service examination when Yulgok was twenty-two.)

Yi I [Yi Yulgok] (1570) *Tongho mundap (Questions and Answers at Eastern Lake)*, in *Yulgok chônsô*, Seoul: Sônggyun'gwan Taehakkyo Taedong Munhwa Yôn'guwon, 1958, 15: 2a-33b.(A major political writing on the way of righteous government, written while Yulgok was minister of justice. It deals with the Confucian idea of sagely and kingly government and how to reform government administration.)

Yi I [Yi Yulgok] (1571) *Sôwôn hyangyak (Community Treaty of Sôwôn),* in *Yulgok chônsô,* Seoul: Sônggyun'gwan Taehakkyo Taedong Munhwa Yôn'guwon, 1958, 16: 2a-7a.(A systematic political and social guideline for local community and government addressing social order, Confucian moral norm and mutual cooperation for the community.)

Yi I [Yi Yulgok] (1572) *Sônghak chibyo (A Compendium of Sagely Learning),* in *Yulgok chônsô,* Seoul: Sônggyun'gwan Taehakkyo Taedong Munhwa Yôn'guwon, 1958, 20: 1a-38b.(Yulgok's unique way of summarizing and interpreting Cheng-Zhu neo-Confucian thought. It also includes the moral and practical aspects of government administration, and Yulgok's concern for practical learning.)

Yi I [Yi Yulgok] (1582) *Kyôngmong yogyôl (A Key to Annihilating Ignorance),* in *Yulgok chônsô,* Seoul: Sônggyun'gwan Taehakkyo Taedong Munhwa Yôn'guwon, 1958, 27: 3a-21b.(Deals with various themes of Confucian way in practice, including not only Confucian learning and self-cultivation but also practical guidelines for Confucian family regulation, human relationships and rituals.)

Yi I [Yi Yulgok] (1583) *Yukchokye (The Manifestation of Six Articles),* in *Yulgok chônsô*, Seoul: Sônggyun'gwan Taehakkyo Taedong Munhwa Yôn'guwon, 1958, 8: 18a-23b.(Six specific reform proposals for economy, government, defence policy, nourishing people and army, education, society and people; written while Yulgok was defence minister and submitted to the king.)

References and further readings

Chung, E.Y.J. (1995) *The Korean Neo-Confucianism of Yi T'oegye and Ti Yulgok,* Albany, NY: State University of New York Press.(A comparative study of Yi T'oegye (Yi Hwang) and Yi Yulgok; a very good introduction to Korean Confucianism.)

Hyun Sang-yun (1971) *Chosôn yuhaksa (History of Korean Confucianism),* Seoul: Minjungsokwan.(A comprehensive view of Korean Confucianism, written in Korean.)

Kalton, M.C., Kim Oaksook, Park Sung-Bae, Ro Youngchan, Tu Wei-ming and Yamashita, S. (eds) (1994) *The Four-Seven Debate: An Annotated Translation of the Most Famous Controversy in Korean Neo-Confucian Thought*, Albany, NY: State University of New York Press.(A most thorough and comprehensive discussion on the issues regarding the Four Beginnings and the Seven Feelings among the Korean neo-Confucian scholars, including Yulgok.)

Kim Kyung Tak (1960) *Yulgok eui Yôn'gu (The Study of Yulgok),* Korean Studies Series 7, Seoul: The Korean Research Center.(An in-depth study of Yulgok's neo-Confucian philosophy. This book has an English summary at the back.)

Lee, P., Baker, D., Ch'oe Yongho, Kang, H.H. and Kim Han-kyo (eds) (1992) *Sources of Korean Tradition,* New York: Columbia University Press.(For a general understanding of Korean cultural and intellectual understanding including Korean Confucianism, this book is a good resource.)

Pae Chong-ho (1974) *Han'guk yuhaksa (History of Korean Confucianism)*, Seoul: Yonsei Taehakkyo Ch'ulp'anbu.(Korean neo-Confucian thought is well discussed in this book, including a comparison between the two most significant Korean neo-Confucian thinkers, Yi T'oegye and Yi Yulgok.)
Pak Chong-hong (1976) *Han'guk sasangsa (History of Korean Thought)*, Seoul: Somun-dang. (Although this book is a history of Korean philosophy, it includes an in-depth investigation of Yulgok's thinking.)
Ro Young-chan (1989) *The Korean Neo-Confucianism of Yi Yulgok*, Albany, NY: State University of New York Press.(An interpretation of Yulgok's non-dualistic way of thinking in his cosmology, ontology and anthropology.)
Son In Soo (1976) *Yulgok eui Kyoyuksasag (Yulgok's Philosophy of Education)*, Seoul: Pakyoungsa. (A good summary introduction to Yulgok's idea of education.)
Yi Byung Do (1973) *Yulgok eui Saengae wa Sasang (Life and Thought of Yulgok)*, Seoul: Seomoon Dang.(A good survey of Yulgok's life and thought.)
Yi Joon Ho (1973) *Yulgok eui Sasang (Yulgok's Thought)*, Seoul: Hyunamsa.(Yulgok's political, philosophical and educational thought are well discussed.)

Yijing

*The Yijing (Book of Changes) or Zhouyi (Changes of the Zhou) was originally a divination manual, which later gradually acquired the status of a book of wisdom. It consists of sixty-four hexagrams (*gua*) and related texts. By the time the Yijing became a coherent text in the ninth century* BC*, hexagram divination had changed from a means of consulting and influencing gods and spirits to a method of penetrating moments of the cosmic order to learn the shape and flow of the* dao *and determine one's own place in it. By doing so, one avoids wrong decisions, failure and misfortune and achieves their contrary. Tradition has it that the Yijing can only be successfully approached through humility, honesty and an open mind. Through interaction with it, one gains ever increasing self-knowledge and sensitivity to one's relations to others and to one's situation in life. 'Good fortune', 'happiness' and 'success' are but by-products of such self-knowledge and sensitivity.*

1 Organization of the *Yijing*

The *Yijing (Book of Changes)* consists of sixty-four hexagrams (*gua*) and related texts. The hexagrams are formed by combinations of two trigrams and result in six solid or unbroken (yang +) and/or broken (yin -lines (*yao*), arranged one above the other in vertical sequence. The combinations are determined by the numerical manipulation of divining sticks, originally yarrow stalks (*achillea millefolium*), or later, by the casting of coins. Each hexagram is accompanied by a hexagram name (*guaming*), a hexagram statement (*guaci*) or 'judgment' (*tuan*), and line statements (*yaoci*) for each of the six lines. The line statements have a sequential organization related to the judgment; each states a specific instance of the judgment, which in complete line statements (many are fragmentary) is followed by an injunction - that one should take some action or refrain from it - and a final determination - 'misfortune', 'good fortune' and so on.

The hexagrams, hexagram statements or judgments and the line statements are the oldest parts of the *Yijing*. Whereas the hexagrams themselves may be much older, the names and statements probably date from the ninth century BC and constitute the first layer in a three-layered text. The second layer consists of another two parts: commentaries on the hexagram statements or judgments called *tuanzhuan* (commentary on the judgments) and commentaries on the abstract meanings or 'images' (*xiang*) of the judgments and the line statements called *xiangzhuan* (commentary on the images). The judgments have 'great images' (*daxiang*) - the abstract meanings of whole hexagrams - and the line statements have 'little images' (*xiaoxiang*) - the abstract meanings of individual lines. These commentaries seem to date from the sixth or fifth century BC and are the first of the exegetical materials in the *Yijing* traditionally attributed to Confucius.

The traditional format of the *Yijing* divides the *tuanzhuan* and the *xiangzhuan* each into two sections. They thus form the first four of the so-called 'Ten Wings' (*shiyi*) of the exegetical material included in the classic. Although all of the 'Ten Wings' are attributed to Confucius, individual 'wings' actually date from different periods, with some earlier and others from as late as the third century BC. However, except for the commentaries on the judgments and the commentaries on the images, which do seem the direct product of Confucius' school if not of Confucius himself, the rest of the 'Ten Wings' consists of later materials, which probably contain reworking of earlier texts. These constitute the third layer of the *Yijing* and include two fragments of an apparently lost commentary on the hexagrams as a whole called the *Wenyan (Commentary on the Words of the Text)*, but only those parts attached to the first two hexagrams - *qian* (pure *yang*) and *kun* (pure *yin*) - survived into the period of textual redaction (early Han era, in the third century BC). The *Wenyan* seems to be a borderline text that contains elements of both the second and third layers. It constitutes the fifth of the 'Ten Wings' and deals with the philosophical and ethical implications of the judgments, line statements and images, all given a Confucian slant.

The sixth and seventh 'wings' are formed by the two sections of the so-called *Xicizhuan (Commentary on the Appended Phrases)* or *Dazhuan (Great Commentary)*. This commentary contains two kinds of material; one deals with the nature and meaning of the *Yijing* in general, and the other is concerned with the meaning of the judgments and line statements of individual hexagrams. The *Xicizhuan* seems to consist of fragments of two different texts, one a general essay or group of essays and the other a collection of specific remarks on individual hexagrams. The eighth of the 'Ten Wings,' *Shuogua (Explaining the Trigrams)* consists of remarks on the nature and meaning of the eight trigrams (*bagua*), the permutations of which form the sixty-four hexagrams. Much here is couched in

terms of *yin-yang* dualism and the theory of the *wuxing* (five elements), and so dates probably from the early Han era (third century BC) (see Yin-yang). It is among the latest of the exegetical materials included. The ninth 'wing' is the *Xugua (Providing the Sequence of the Hexagrams)*, a collection of remarks on each of the hexagrams which attempts to justify their order in terms of etymological and pseudo-rational considerations. This also seems to be quite late material. The tenth 'wing', the *Zagua (Hexagrams in Irregular Order)*, is a collection of brief remarks on the meanings of individual hexagrams, often in terms of contrasting pairs; this is another late addition to the text.

2 History of the text

Traditionally, the hexagrams are thought to have been developed by King Wen of the Zhou in the twelfth century BC, out of the eight trigrams invented by the legendary cultural hero and sage Fu Xi in remotest antiquity. King Wen is also supposed to have composed the hexagram statements or judgements. The line statements are attributed to the Duke of Zhou (d. 1094 BC). However, the assertion that historically identifiable sages are responsible for the origins of the hexagrams and the composition of the first layer of the material is most unlikely. Recent advances in archaeology, paleography and textual studies, which compare the earliest textual layer of the *Yijing* with roughly contemporary inscriptions on bone, shell, metal, stone and other ancient writings that exhibit similar syntax and vocabulary have thoroughly discredited the myth of its sagely authorship. Modern scholarship has also discovered that the original meaning of the judgments and line statements is radically different from what the earliest layer of exegesis took it to be, and often bears little relation to the values of Confucian morality and ethics. Either the writers of the *tuanzhuan* and the *xiangzhuan* were ignorant of this original meaning - being concerned largely with the mechanics of divination and (often) its amoral consequences - or they knowingly suppressed it in order to replace it with a Confucian (or proto-Confucian) reading. However, with this first layer of exegesis the collection of texts, which eventually developed into the *Yijing*, was given a Confucian slant that shaped all subsequent interpretation right up to modern times.

3 The commentary tradition

Although many commentaries on the *Yijing* were written during the Han period (206 BC- AD 220), these were mostly prognostication recipes based on numerological correspondences between the cosmic, natural and human realms (including calendrical considerations) - a so-called *xiangshu* (image and number) approach in which interpreters identified the trigrams and hexagrams with certain standardized images (symbols with fixed meanings). Wang Bi (226-49) broke completely with this and may be said to have written the first philosophical commentary on the *Yijing*; that is, apart from those sections of the work which are themselves commentaries. His *Zhouyizhu (Commentary on the Changes of the Zhou)*, a synthesis of Confucian, Legalist and Daoist views with Confucian views predominant, was extremely influential during the pre-Tang and Tang eras (fourth to tenth centuries AD) and was canonized in the *Zhouyi zhengyi (Correct Meaning of the Changes of the Zhou)* of Kong Yingda (574-648) as the orthodox interpretation of the *Yijing*. Although the commentaries of the later neo-Confucians largely eclipsed Wang's interpretation, much of his commentary was incorporated into the official neo-Confucian orthodox view of the *Yijing*, and what they rejected also helped to shape that view.

A comparison of Wang's commentary with those of Cheng Yi (1033-1107) and Zhu Xi (1130-1200), the most important of neo-Confucian interpretations, reveals how carefully Cheng and Zhu read Wang's remarks and, whether they agreed with him or not (disagreements mostly involve rejection of elements of Legalism and Daoism perceived in Wang's thought), how his arguments tended to shape theirs. Cheng's *Yichuan Yizhuan (Yichuan's Commentary on the Yijing)*, a commentary of textual interpretation (*yili*), explains the work in terms of how the interplay of *yin* and *yang* defines the moral ideals and behaviour of the noble man (*junzi*). Zhu's *Zhouyi benyi (Original Meaning of the Changes of the Zhou)* attempts a reform of the *xiangshu* (image and number) approach combined with neo-Confucian moral interpretation (*yili*): here the *Yijing* is interpreted both as a manual of divination and a book of moral wisdom. Many other commentaries on the *Yijing* were written throughout traditional times, and not only Confucian/neo-Confucian but also Daoist and even Buddhist versions exist. Nevertheless, the most important commentaries were those noted above by Wang Bi, Cheng Yi and Zhu Xi, and most modern annotated editions and translations in Chinese, Japanese, English and other Western languages are based on one or more of these commentaries.

4 The hexagram as discourse

Although later neo-Confucians often disagreed with Wang Bi, they agreed with his general approach to the meaning of whole hexagrams, individual lines and line positions, and the interaction of lines and positions: each hexagram is a unified entity whose 'controlling principle' is expressed in its name and amplified in the hexagram judgement. The controlling principle usually resides in the ruler of the hexagram, one line that is sovereign over all the others. Change occurs because of the interaction between the innate tendency of things and their counter-tendencies to behave in ways opposed to their natures. 'Things' include individual human beings, and the lines of a hexagram represent different kinds of people in different positions and different situations. The action and interaction of lines occur as if people were involved in a particular set of circumstances. Some lines respond to each other and resonate together, signifying harmonious relationships, and some lines repel and clash, signifying opposition and divergence of interests. This resonance or clash produces movement and change, the understanding of which leads to insight into the innate tendencies of things - 'how things are going' - and adjusting one's behaviour accordingly leads to success. Certain hexagrams signify moments of either obstruction or facility and are indicators that one should either refrain from action or engage in it. *Yin* lines are soft and weak; *yang* lines are hard and strong.

The positions of a hexagram are calculated from bottom to top. The odd number places - first (bottom), third and fifth - are strong *yang* positions; the even number places - second, fourth and sixth (top) - are weak *yin* positions. *Yin* and *yang* lines form resonating pairs, and *yin* and *yin* or *yang* and *yang* lines form discordant pairs: the unlike attract and the like repel. Proper resonate relationships can take place between lines of the lower and upper trigrams: one with four, two with five, three with six, but each must pair with its opposite, *yin* with *yang* or *yang* with *yin*. Secondary harmonious relationships can also occur between contiguous lines when '*yang* rides atop *yin*' or '*yin* carries *yang*' but never the reverse, for this is an unnatural, discordant relationship - as when a superior supports or 'carries' his subordinate.

These sixty-four combinations of *yin* and *yang* lines and *yin* and *yang* positions schematically represent all major kinds of situations found in life. One must know how to cast the hexagrams and how to understand the texts of the *Yijing*, for if one can determine which situation prevails at any given moment, what one's place is in that moment and situation and how one relates to the other participants, 'change will yield its all'. As first (bottom) and sixth (top) lines are at the beginning and end points of hexagrams respectively (and thus signify the beginning and ending of situations), they are at the junctures of what precedes and what follows a given situation, and so 'neither of these positions has a constant status'. The other four line positions are either *yang* and 'noble' (three and five) or *yin* and 'humble' (two and four). *Yang* lines 'should' be in *yang* positions and *yin* lines 'should' be in *yin* positions, for this results in hexagrams that generally indicate facility and harmony. Lines 'out of position' (*yin* in *yang* positions, *yang* in *yin* positions) result in hexagrams that generally indicate obstruction and disharmony.

A special role is assigned to the middle positions in the trigrams, positions two and five. These middle positions indicate 'centrality' and 'the mean' (*zhong*), the territory of proper and balanced behaviour and action. The middle position in the lower trigram is a *yin* position (two), and the one in the upper trigram is a *yang* position (five), so very often, regardless of other considerations, the line in the fifth position - whether *yin* or *yang* - is the ruler of the hexagram as a whole, for it is the 'most noble' place, the 'exalted position'.

This approach to the *Yijing* assumes that the casting of hexagrams is an absolutely sure and accurate method of determining the character of moments of time. This is true of all traditional commentators on the *Yijing*: the way that the yarrow stalks or coins fall is indicative of the shape of that particular moment. Everything that occurs in a given moment is interrelated, and all such events share in the same basic character.

See also: Chinese Classics; Confucian philosophy, Chinese; Daoist philosophy; Neo-Confucian philosophy; Yin-yang

RICHARD JOHN LYNN

References and further reading

Kong Yingda (574-648) *Zhouyi zhengyi (Correct Meaning of the Changes of the Zhou)*, in Ruan Yuan (ed.) *Shisanjing zhushu (Commentaries and Subcommentaries on the Thirteen Classics)*, 1815 woodblock edn; repr. Taibei: Yiwen yinshuguan, 1955. (This edition contains texts of the *Changes*, with the entire commentaries of

Wang Bi and Kong Yingda.)

Li Guangdi (ed.) (1715) *Zhouyi zhezhong (Equitable Judgments on Interpretations of the Changes of the Zhou)*, woodblock edn; in Yan Lingfeng (ed.) *Yijing jicheng (Collectanea of Works on the Book of Changes)*, Taibei: Chengwen, 1975. (Contains texts of the *Yijing*, with the entire commentaries of Cheng Yi and Zhu Xi as well as excerpts from other significant commentaries.)

Smith, K., Bol, P.K., Adler, J.A. and Wyatt, D.J. (1990) *Sung Dynasty Uses of the I Ching*, Princeton, NJ: Princeton University Press. (Collection of essays on the history, interpretation and role of the *Yijing* in Song era (960-1279) intellectual life, the formative period of neo-Confucianism.)

Wang Bi (226-49) *Zhouyizhu (Commentary on the Changes of the Zhou)*, in vols 1- 2 of Lou Yulie (ed.) *Wang Bi ji jiaoshi (Critical Edition of the Works of Wang Bi with Explanatory Notes)*, Beijing: Zhonghua shuju, 1980. (The best edition of Wang Bi's commentary, with annotations in modern Chinese.)

Yijing (*c*.9th-2nd century BC), trans. R.J. Lynn, *The Classic of Changes: A New Translation of the I Ching as Interpreted by Wang Bi*, New York: Columbia University Press, 1994. (An integral translation of the *Yijing* along with the commentary by Wang Bi. Where Wang's interpretation differs significantly from the later readings of Cheng Yi and Zhu Xi, references to and translated excerpts from their commentaries are provided for purposes of comparison.)

Yin-yang

Yin *and* yang *always describe the relationships that obtain among unique particulars. Originally these terms designated the shady side and the sunny side of a hill, and gradually came to suggest the way in which one thing 'overshadows' another in some particular aspect of their relationship. Any comparison between two or more unique particulars on any given topic is necessarily hierarchical: one side is* yang *and the other is* yin. *The nature of the opposition captured in this pairing expresses the mutuality, interdependence, hierarchical relationship, diversity and creative efficacy of the dynamic relationships that are immanent in and give value to the world. The full range of difference in the world is deemed explicable through this pairing.*

Yin and *yang* are elements of a correlative pairing which are pragmatically useful in sorting out 'this' and 'that', and are not, as often claimed, dualistic principles of light and dark, male and female, action and passivity where light and dark both exclude each other and logically entail each other and, in their complementarity, constitute a totality. Rather, *yin* and *yang* are first and foremost a vocabulary of qualitative contrasts which have application to specific situations and enable us to make specific distinctions.

To bring this observation to bear on our understanding of *yin* and *yang*, we must start with the relationship that obtains between any two particular things or events. For example, in a given relationship, 'this' older woman might by virtue of her wisdom be regarded as *yang* in contrast to 'that' younger woman who is *yin*. However, if we were to focus on their fecundity or physical strength, the correlation would likely be the opposite. Important here is the primacy of the particular and the fluidity of the relationship. Although things in this world - that is, particular things - are resolutely hierarchical, no one thing excels in all respects, making this same hierarchy the basis for their complementarity.

In the classical Chinese world, things of the same 'kind' are not defined in terms of their essences as natural kinds but by virtue of their affinity or 'kinship' resemblances that associate them, their 'family resemblances'. Important here is the primacy of particular difference and the absence of any assumed sameness or strict identity. Things are deemed to have resemblances based upon analogous roles or functions. Thus one thing, by virtue of its relationships, evokes many. The suggestiveness of each phenomenon in calling up other similar phenomena is comparable to the multivalence of poetic images. Describing a particular phenomenon does not require the discovery of some underlying determinative and originative principle - a basis for making 'many' one - but a mapping out and an unravelling of the phenomenon's multiple correlations and of the relationships and conditions that make up its context. *Yin* and *yang* define the tension between multiple perspectives on phenomena, and they enable us to interpret and bring coherence to our circumstances by allowing us to discern the patterns in relationships within particular contexts. They provide a vocabulary for sorting out the relationships that obtain among things as they come together and constitute themselves in unique compositions.

Thus, *yin* and *yang* as correlatives are not universal principles that define some essential feature of phenomena, but are explanatory categories that register a creative tension in specific differences and thus make the immediate concrete things of the world intelligible. It is only through a process of generalization that feminine and male gender traits are construed as predominantly *yin* and *yang* respectively, and vocabulary such as vaginal orifice (*yinmen*) and virility (*yangdao*) emerges to essentialize the *yin* and *yang* contrast.

The *yin-yang* vocabulary is functional. In order to evaluate the propensity of a situation and manipulate it in advance, we must make distinctions. This is where the vocabulary of *yin* and *yang* comes into play. We begin from the assumed uniqueness of each situation and the uniqueness of the components that constitute it. The *yin-yang* contrast provides a line, enabling us to divide a continuous situation into distinct yet interdependent particulars. *Yin* is a becoming-*yang*; *yang* is a becoming-*yin*. We must come to know the particular conditions that govern a situation so that we can manipulate them to advantage. This requires that one translate the situation into the *yin-yang* vocabulary of complementary opposites: strong-weak, fast-slow, many-few, regular-irregular and so on. *Yin-yang* is a vocabulary that enables us to discriminate among the many factors which together constitute the force of circumstances, and that allows us to control this force through the strategic adjustments. Once we have arrived at an understanding of the circumstances, we must identify those critical factors which will enable us to turn the configuration of an unfolding situation into an opportunity.

See also: Categories §4; Chinese philosophy; Dao; De; Particulars; Qi; You-wu

ROGER T. AMES

References and further reading

Graham, A.C. (1989) *Disputers of the Tao: Philosophical Argument in Ancient China*, LaSalle, IL: Open Court. (A survey of ancient Chinese thought.)

Hall, D.L. and Ames, R.T. (1995) *Anticipating China: Thinking Through the Narratives of Chinese and Western Culture*, Albany, NY: State University of New York Press. (A comparative study of the uncommon assumptions that ground the Chinese and Western philosophical traditions.)

Yoruba epistemology

*The Yoruba of west Africa have articulated systematic criteria that are used to assign varying degrees of epistemic certainty to experience. What one views with one's own eyes and experiences at first-hand (*ìmò*) are judged as reliable ways of knowing the truth, providing there is conscious comprehension of what one is perceiving. Only propositions describing such experiences are regarded as true, or *òótó*. Less reliable is information received via books, other people, the media and the oral tradition. If such comparatively second-hand information, or *ìgbàgbó*, can be experimentally tested and accordingly verified, it has the potential to become *ìmò*. If verification cannot be attested, discussion, analysis and good judgment are essential tools for distinguishing the more reliable information from the less reliable.*

The semantic distinction the Yoruba make between *ìmò* and *ìgbàgbó* serves as a basis for comparison with that between knowledge and belief in traditional Western epistemological theory (see Belief and knowledge §1). Although the Yoruba affirm the epistemic importance of all sensory perception, they single out visual perception, or *ìrirn*, as clearer and more reliable. When conjoined with cognitive comprehension, or *èrí okòn* (literally meaning 'the witnessing of the mind', connoting the conscious processes that arise from both the act of seeing and the sense that one understands what one sees), the two satisfactory conditions for *ìmò*, or knowledge, have been achieved. Experience that satisfies these conditions and propositions that relate such experience, are characterized as being true, or *òótó*. The emphasis placed upon propositions being true because they accurately report experience indicates elements of a correspondence theory of truth (see Truth, correspondence theory of §1).

Ìgbàgbó encompasses all the events and information which do not constitute first-hand experience. This applies to most of what one is taught in the course of a formal education, what one learns from books, the media, other people and, of particular interest in the case of the Yoruba, from the oral tradition. *Ìgbàgbó* is the conflation of 'agree' and 'accept', or *gbà*, and 'hear' and 'understand', or *gbó*. The independent application of negation to its component parts produces four distinct logical possibilities, each of which may be said to correspond to a likely speaker-hearer relationship. These four levels of communication are identified by the Yoruba in the following concrete terms:

(1) *Hearing and agreeing* (p.q): where a hearer, H, understands and accepts as credible what is being reported and therefore, in principle, has no objection to incorporating it into their own store of information. This is optimum communication.
(2) *Hearing and not-agreeing* (p.[thksim]q): H understands what is being offered but declines to accept it as authoritative or correct. Disagreements or lies are offered as standard examples.
(3) *Not-hearing and not-agreeing* ([thksim].[thksim]q): failure to communicate. H does not understand what is being said and therefore cannot agree.
(4) *Not-hearing and agreeing* ([thksim]p.q): in effect, to agree with what one does not understand. Either a case of error or H is described as a fool.

When *ìmò* and *ìgbàgbó* are challenged and an argument or *àríyàn jiyàn* results, the tactics recommended for resolving the dispute are complex and interwoven. Often the disputants do not share the same (relevant) first-hand experiences. This may be corrected by a process of verification so long as it is possible to test the claim empirically and thereby enable people to see the results for themselves: counterclaims about the effectiveness of a certain medicine, for example, can be checked by testing the treatment. When such testing is impossible, there is no way in which the *ìmò* of one person can become the *ìmò* of another. In such a case, the only recourse is to a process of justification. The parties concerned should each explain (*àlàyé*) their own position by giving a full account of relevant first- or second-hand experiences. Any witnesses whose experiences may be relevant should also be called in the hope that their testimony will help decide which account, if any, should be favoured. Such things as the moral character, or *ìwà*, of each participant must be considered, as this might affect the reliability of information.

The aim in this extensive process of justification and evaluation is to lead the various parties to reflect upon and perhaps modify their respective positions, so that agreement may be reached on an accurate account of events, or *papò*, meaning literally 'the word has come together'. Of course it is recognized that, even if the parties concerned

supply full explanations, their moral characters are found to be intact and the testimonies of witnesses are adjudged satisfactory, the dispute may remain unsettled and the argument may continue. In such cases the only realistic alternative is to begin again in the hope that further discussion and reflection may achieve a resolution.

The system that emerges from these criteria is three-tiered. *Imò* is the sole category of experience or propositions permitted to be regarded as both certain and true (*òótó*). *Ìgbàgbó*, which in principle is open to empirical testing, verification and thereby transformation into *ìmò*, is the next most reliable category.*Ìgbàgbó* that can never be verified and can only be evaluated on the basis of explanation, discussion and reflection is the least certain.

Prevalent paradigms describe African traditional thought systems as insignificantly critical or reflective (particularly of the oral traditions, which are said to be the locus of knowledge in such societies), and only marginally empirical because of their emphasis on magic, superstition and secrecy. This account of the ways in which the Yoruba employ and explain key epistemological terms in their language demonstrates that they are more critically- and empirically-minded than had been supposed (see Epistemic relativism §1). The oral tradition, the wealth of knowledge passed down through one's forbears, has not been regarded as an authoritative system of truths. As *ìgbàgbó*, the status of its truth is hypothetical and conditional on being tested or proven practically effective. When proof cannot be obtained, critical discussion and reflection are vital.

BARRY HALLEN

References and further reading

Abimbola, W. and Hallen, B. (1993) 'Secrecy and Objectivity in the Methodology and Literature of *Ifá* Divination', in M.H. Nooter (ed.) *Secrecy: African Art that Conceals and Reveals*, New York, NY: Museum for African Art.(The instrumental functions of secrecy in an African system of divination.)

Appiah, K.A. (1982) 'Old Gods, New Worlds: Some Recent Work in the Philosophy of African Traditional Religion', in *In My Father's House: Africa in the Philosophy of Culture*, London: Methuen; 2nd edn, Oxford: Oxford University Press, 1992.(Incisive discussion and analysis of various characteristics said to distinguish 'traditional' from 'modern' societies, particularly with respect to the formulation and criticism of theoretical explanations.)

Hallen, B. and Sodipo, J.O. (1986) *Knowledge, Belief and Witchcraft: Analytic Experiments in African Philosophy*, London: Ethnographica.(Chapter two expands on Yoruba epistemology and draws comparisons with English-language epistemological equivalents.)

Wiredu, K. (1996) 'The Concept of Truth in the Akan Language', *Cultural Universals and Particulars: an African Perspective*, Bloomington and Indianapolis: Indiana University Press.(Analysis of epistemological variations between the Akan language of Ghana and English. Further evidence of the epistemological sophistication of traditional African systems of thought.)

You-wu

In the Western metaphysical tradition, 'being' has most generally been thought to denote either a common property of things or a container which relates things by placing them within its own structure. Metaphysical notions of being are generally associated with the concept of ground. By contrast, the Chinese existential verb you *(being) overlaps with the sense of 'having' rather than the copula, and therefore* you *(to be) means 'to be present' or 'to be around' while* wu *(not to be) means 'not to be present' or 'not to be around'. This means that* wu *does not indicate strict opposition or contradiction, but absence. Thus, the* you-wu *distinction suggests mere contrast in the sense of either the presence or absence of* x*, rather than an assertion of the existence or nonexistence of* x*.*

Not a little misunderstanding of classical Chinese thought has been occasioned by the fact that uses of *you* (being) and *wu* (not-being) often seem to echo the speculative understandings of classical Western metaphysics. Thus the Daoist sentiment, often rendered in English as 'Not-being is superior to being,' has given rise to a host of mystical and speculative reflections inspired by the Parmenidean or Heideggerian projects (see Parmenides; Heidegger, M.). In fact, a better translation is simply, 'Nothing is superior to something'. However, even this might suggest presence and absence in some absolute sense. For the Western thinker, an even more accurate translation would be 'Not-having is superior to having': or as a contemporary Chinese Marxist translation reads, 'Not-owning is superior to owning'.

It may be said that the proper understanding of 'being' in the Chinese tradition, more than any other philosophical concept, depends upon the ability to suspend predispositions deriving from Western philosophical sources. For, without recourse to the senses of being associated with Western speculative philosophies (see Being), and in the absence as well of the sophisticated logical apparatus employed by the nominalistic traditions in contemporary thought, the interpretative contexts familiar to Western trained thinkers are simply not relevant. Thus the Chinese discussions of the *you-wu* relation must not be thought to be 'metaphysical', if we mean by metaphysics anything like a universal science of principles or of being-itself. Nor may we be allowed to interpret 'being' and 'not-being' by recourse to logical operators associated with modal or truth-functional logics (see Modal logic).

Thus when the Chinese discuss cosmological issues of the sort that require an understanding of kinds of existing things, or the basic categories that make up the world as we know it, we must proceed with caution. While it is true that Chinese thinkers, particularly the Daoists, ask about things, they do not ask about 'categories' or 'kinds' in any manner that would suggest that things have 'logical essences' or constitute 'natural kinds' (see Essentialism).

The principal reason Chinese thinkers are not apt to ask after the *logos* of the cosmos is that they lack an operative sense of cosmos or 'world' as a coherent whole. Their sense of *logos* - *li* (pattern) - is a radically situated notion (see Li). In fact, the Chinese understanding of 'cosmos' as 'the ten thousand things' means that, strictly speaking, they have no concept of cosmos at all insofar as that notion entails either a coherent, single-ordered world or a congeries of entities with essential features or essential modes of connectedness. The Chinese are therefore primarily 'acosmotic' thinkers.

In the absence of a sense of being as the ground of things, the act of understanding and articulating the sense of things - *daoli* (the grasp of the patternings (*li*) of *dao* (the ways of things)) - cannot have ontological reference (see Dao). The closest to our typical understanding of 'reason' or 'reasoning' is to be found in 'seeking the *li*'. *Wulixue* in modern Chinese is 'the investigation of the patterns of things and events', what we term 'physics'. 'Psychology' is *xinlixue*, 'the investigation of the patterns of the heart-mind'. In general, to be *heli*, to be in accord with *li*, is to be reasonable or rational (see Li; Confucian philosophy, Chinese).

The translation of *li* as 'principle' has created some confusion in the minds of Western interpreters of Chinese, many of whom just assume that *li* must, at some level, be transcendent. But *li*, as a 'making sense of things', cannot be understood as a process of seeking principles as determining sources of order, or of discovering essential categories inclusive of particular things. It is, rather, an activity which constructs categories (*lei*) analogically, then traces, again by analogical means, correlated details which manifest patterns of relationships immanent within things and events. This sort of reasoning depends upon noninferential access to enlarged and deepened patternings.

Inclusion or exclusion of items within a ‘category’ are never associated with notions of logical ‘type’ or ‘class’. The sortings are analogical, not logical. This understanding of ‘reason’ and ‘reasoning’ is one more implication of the fact that the Chinese never asked ‘the question of being’.

See also: Being; Chinese philosophy; Confucian philosophy, Chinese; Dao; Daoist philosophy; Existence; Li

DAVID L. HALL
ROGER T. AMES

References and further reading

Daodejing (*c*.350-250 BC?), trans. D.C. Lau, *Tao te Ching*, Harmondsworth: Penguin, 1963. (The being-nonbeing relation, expressed through the *yin-yang* pairing, is a principal theme of the *Daodejing*.)

Graham, A.C. (1989) *Disputers of the Tao: Philosophical Argument in Ancient China*, LaSalle, IL: Open Court, 1989. (See especially Parts II: 3, III: 1 and Appendix 2.)

Graham, A.C. (1990) ‘Being in Western Philosophy Compared with Shih/Fei and Yu/Wu in Chinese Philosophy’ in A.C. Graham (ed.) *Studies in Chinese Philosophy and Philosophical Literature*, Albany, NY: State University of New York Press. (Graham’s discussions in this essay and the work immediately cited above, constitute the most nuanced comparative treatment of the senses of ‘being’ and ‘nonbeing’ in Chinese and Western traditions.)

Hall, D.L. and Ames, R.T. (1995) *Anticipating China: Thinking Through the Narratives of Chinese and Western Culture*, Albany, NY: State University of New York Press. (A comparative study of the uncommon assumptions that ground the Chinese and Western philosophical traditions.)

Zabarella, Jacopo (1533-89)

Jacopo Zabarella was a professor of philosophy at the University of Padua. His work shows conclusively not only that it was possible to philosophize creatively within the limits of the Aristotelian tradition but also that this was still being done towards the end of the Renaissance period. Zabarella's aim was not to overthrow Aristotle's doctrines, but to expound them as clearly as possible. He produced an extensive body of work on the nature of logic, arguing that it was neither an art nor a science, but rather an instrumental intellectual discipline which arose from the philosopher's practice of philosophizing or forming secondary notions. He also worked extensively on scientific method. He gives an account of order as disposing what we come to know through method, and he divides method into the method of composition, which moves from cause to effect, and the method of resolution, which moves from effect to cause. He also discussed regressus *(a method for uniting composition and resolution) and thought that it would enable the scientist to discover new causal relations at the same time as proving conclusions with absolute necessity. Zabarella's work was instrumental in a renewal of natural philosophy, methodology and the theory of knowledge; and it had a major impact on seventeenth-century philosophy textbooks, especially in the Protestant countries of northern Europe.*

1 Life

Count Jacopo (or Giacomo) Zabarella seems never to have left his native city, Padua. He enjoyed a thorough education, receiving his doctorate in philosophy and the liberal arts in 1553, and began his teaching career in 1564 when he obtained the first chair of logic. In 1568 he obtained the second extraordinary chair of natural philosophy, and was promoted to the first extraordinary chair of natural philosophy in 1577. Finally, in 1585, Zabarella obtained the second ordinary chair of natural philosophy, which he held until his death. Padua was a leading university, and Zabarella was taught by some of the best teachers there: Francesco Robortello in the humanities, Bernardino Tomitano in logic, Marcantonio Genua in physics and metaphysics, and Pietro Catena in mathematics. His academic career was exemplary for a Paduan professor of philosophy, with two exceptions. First, the statutes of the university prevented him, as a native Paduan, from obtaining the first ordinary chair in natural philosophy; second, he never took a degree in medicine, perhaps through his own choice. His publications reflect his teaching in the Aristotelian tradition. From 1578 to 1584, he published on logical problems, including a commentary on the *Posterior Analytics*. In 1590 he published a volume of questions in the contemporary Aristotelian manner, which contained 30 books and followed the traditional order of Aristotle's works on nature. These were supplemented by commentaries on the *Physics*, *On Generation and Corruption*, *On Meteors* and *On the Soul*.

2 The nature of logic

Zabarella's introductory treatise on the nature of logic, *De natura logicae (On the Nature of Logic)*, is basic to his teaching in logic (see Logic, Renaissance §1). He defines logic as being neither a science nor an art, but, in keeping with the traditional meaning of the word 'organon', just an instrument of the arts and sciences. Logic does not have a real subject of its own, but deals with concepts, which stand for real beings. In this it is comparable to grammar. The difference between grammar and logic is that grammar is concerned with the perfect verbal expression of concepts, and hence is a linguistic discipline (*disciplina sermocinalis*), while logic invents second notions (*notiones secundae*, often called second intentions), that are able to create order among concepts, and serve to recognize the truth and distinguish it from falsehood in every instance. Logic is thus a rational discipline (*disciplina rationalis*). It is not itself philosophy, but springs from philosophy and is devoted to philosophical ends.

Zabarella followed Averroes in dividing logic into two parts: universal logic, which is common to all subjects; and particular logic, which is specific to particular subjects. The first three books of Aristotle's Organon, the *Categories*, *De Interpretatione* and the *Prior Analytics*, constitute universal logic. Aristotle's *Posterior Analytics*, *Topics* and *Sophistical Refutations* are said to deal with particular logic, in that they deal respectively with the demonstrative syllogism, the dialectical syllogism and the sophistical syllogism (see Aristotle §1). Following the Neoplatonic commentators Simplicius and Ammonius, Zabarella included Aristotle's *Rhetoric* and *Poetics* in logic. The former is included because it teaches the use of the rhetorical syllogism or enthymeme, and rhetorical induction or example; the latter because it also teaches the use of example, not to persuasive ends, but for imitation. It has been argued that Zabarella here paved the way for modern aesthetics.

3 Arts and sciences

Since logic, viewed as the universal instrument for distinguishing between the true and the false, differs according to the objects to which it is applied and the ends for which it is used, its nature depends on the realm of possible objects and ends. Rhetoric and poetics are special cases because they deal not with knowledge but with the political disciplines (*disciplinae civiles*) in so far as they are concerned with the good of the people. Sophistical syllogistic is another special case, because it is directed towards deception and prefers to use falsehoods as its material. Dialectic and demonstration, however, are directed towards the expression of truth. Dialectic is aimed at the production of opinion, and deals with probable and contingent material; demonstration is dedicated to the acquisition of truth, and so is exclusively occupied with necessary, true objects. As a result the sciences in the proper sense of that term, as pertaining to demonstrative knowledge, are limited to those disciplines that deal with the necessary and eternal or with what can be deduced from necessary principles. These disciplines are the contemplative sciences, metaphysics, mathematics and physics. The necessity involved is both ontological, with respect to the objects known, and cognitive, with respect to the knowing subjects.

All the other so-called arts and sciences deal with contingent objects, that is, with objects which depend on our volition and are therefore either actions, as in the case of the moral disciplines, or artificial beings generated by our productive faculties, as in the case of the arts in the proper sense of that term. The arts as arts are unable to generate true knowledge rather than mere opinion, except where inferences from the desired ends to the necessary means are concerned.

Given this distinction, Zabarella presents the contemplative sciences as being the only defenders of true knowledge. However, by describing the arts in terms of their specifically productive character, he secures for them the autonomy of modern technical disciplines. The use made of Zabarella’s discussion in the textbooks of seventeenth-century Protestant countries prove this seemingly traditional Aristotelian division to have been fruitful.

4 Method

The instrument for acquiring knowledge in the broadest sense is method. Since knowledge can be acquired either through teaching or through invention, Zabarella distinguished between two kinds of method: *ordo doctrinae*, which is the order of teaching, and *methodus* in the narrow sense, which is the method of finding new knowledge. Both involve leading the way from something known to something hitherto unknown. *Methodus* does this by the inference of new items of knowledge; *ordo* does it by the disposition of items already known to the teacher. Neither procedure is governed by the nature of the objects known, but by the nature of human knowledge. *Methodus* starts with what is better known to us, which can be the cause for further cognition, rather than with the better known in itself, which can be the cause of a known object. *Ordo* starts with what is either necessary or useful for teaching and learning. In the contemplative sciences, which aim at perfect knowledge, *ordo* follows the so-called way of composition (*compositio*) from general principles to particular beings; in moral philosophy and the arts, which aim at action or production, *ordo* follows the so-called way of resolution (*resolutio*) from the desired end to its first principles. The bitterly polemical reaction of Zabarella’s contemporaries, especially his colleague Francesco Piccolomini, to this grounding of method and order on cognitive rather than ontological principles, shows how revolutionary his approach seemed at the time.

5 The method of *regressus*

Although, given their concern with perfect knowledge, the contemplative sciences cannot do without the deductive procedures of composition, they also depend on the inductive procedures of resolution, at least in the case of physics, where the particular beings or effects are better known to us.

The so-called method of *regressus* is a model for combining both procedures that is found in the Aristotelian tradition from Averroes on, and was revived during the Italian Renaissance. According to this method, the natural philosopher should first infer from the known effect the existence of the cause of the effect. This is the so-called *demonstratio quia* or demonstration of the fact. Then in a second step, in the so-called *demonstratio propter quid* or demonstration of the reasoned fact, the natural philosopher should infer from the cause to the effect, which is now known through its cause, and hence in a scientific manner. The crucial problem with this procedure is how to avoid mere circular reasoning, or rather, how to make sure that the cause, whose existence is demonstrated in the

first step, is indeed the cause of that effect. From the beginning of the sixteenth century, it had become clear that it was necessary to introduce a third, intermediary step, which involved some kind of *negotiatio intellectus* or intellectual operation. How the intellect did this was a matter of dispute.

Zabarella himself solves the problem of the *negotiatio intellectus,* which he calls a mental examination (*examen mentale*) in terms of his psychology of knowledge. Since for him the task of this intermediary step is to make distinct the confused knowledge of the cause that was acquired through the first step, he refers to his work on the agent mind in which he develops an account of the transformation of confused into distinct knowledge through the analysis of a given whole in terms of its parts. He presents this process as the specific ability of the human mind. Thus once more method as a means for acquiring knowledge is based on the cognitive structure of the knowing subject rather than on the ontological structure of the object of knowledge. This foreshadows the substitution of epistemology for metaphysics in much modern philosophy.

6 Natural philosophy

Natural philosophy, which deals with corporeal beings that have an inner principle of movement, differs from metaphysics (which contemplates being as being) and from mathematics (which deals with abstracted beings), both with respect to the object considered (*res considerata*) and with respect to the way of considering (*modus considerandi*). As a result, it is autonomous and independent of both the other contemplative sciences. It has to know and teach the very essence of natural beings. First, it has to deal with their basic principles, such as matter and motion, which are not natural beings themselves. They are discussed in Aristotle's *Physics*. Second, it has to deal with the accidents of natural beings understood through their causes. These are the subject of Aristotle's other writings on nature, from *On the Heavens* to *On the Soul*. Generally speaking, Zabarella saw natural philosophy as developed by Aristotle as complete according to the scope of its subject matter as well as in its manner of consideration. His main emphasis is therefore on teaching, for invention is only needed in the rare case of a gap in the Aristotelian corpus (see Natural philosophy, medieval).

In spite of its empirical basis, Zabarella's natural philosophy is not concerned with anything akin to experiment. Indeed, if experiments were to be developed, they would find their place in the productive arts rather than in natural philosophy. Nor was he concerned with the role of mathematics in scientific procedure. However, attempts have been made to link him with Galileo. For instance, Randall (1961), following Cassirer, referred to the Renaissance discussion of *regressus* up to Zabarella as a preparation for Galileo's new method of natural science. This hypothesis has been widely debated but remains problematic.

See also: Aristotelianism, Renaissance; Logic, Renaissance

ECKHARD KESSLER

List of works

Zabarella, J. (before 1578) *De natura logicae (On the Nature of Logic)*, in J. Zabarella, *Opera Logica*, Hildesheim: Olms, 1966, 1-102.(Zabarella's fundamental work on logic.)

Zabarella, J. (1582) *In duos Aristotelis libros Posteriores Analyticos commentarii (Commentary on the Two Books of Aristotle's Posterior Analytics)*, Venice.(Exposition of Aristotle's theory of science.)

Zabarella, J. (1585) 'Una "oratio" programmatica di G. Zabarella' (A Programmatical Lecture by Jacopo Zabarella), ed. M. Dal Pra, *Rivista critica di storia della filosofia* 21 (1966): 286-91.(Inaugural lecture on the relation between logic and philosophy.)

Zabarella, J. (1597) *Opera logica (Logical Works)*, Cologne: Lazarus Zetznerus; repr. Hildesheim: Olms, 1966. (The basic collection of Zabarella's writings on logic, method and theory of science. Contains editions from 1578, 1580, 1582 and 1584.)

Zabarella, J. (1602) *Commentarii in Aristotelis libros Physicorum, In libros De generatione et corruptione, In Meteora (Commentary on Aristotle's Physics, On Generation and Corruption and On Meteors)*, Frankfurt: Lazarus Zetznerus.(Expositions of the three most general Aristotelian writings on nature.)

Zabarella, J. (1606) *In tres libros Aristotelis De anima commentarii (Commentary on Aristotle's Three Books On the Soul)*, Frankfurt: Lazarus Zetznerus; repr. Frankfurt: Minerva, 1966.(Exposition of Aristotle's psychology, including his theory of knowledge.)

Zabarella, J. (1607) *De rebus naturalibus (On Natural Things)*, Frankfurt: Lazarus Zetznerus; repr. Frankfurt: Minerva 1966.(The basic collection of his treatises in natural philosophy.)

References and further reading

Edwards, W.F. (1960) 'The logic of Jacopo Zabarella (1533-1589)', unpublished dissertation, Columbia University, NY.(Logic and methodology presented from Randall's point of view; vast biography and bibliography.)

Edwards, W.F. (1969) 'Jacopo Zabarella: A Renaissance Aristotelian's View of Rhetoric and Poetry and their Relation to Philosophy', in *Arts libéraux et philosophie au moyen âge*, Paris and Montreal, Que.: Vrin.(On the view that rhetoric and poetics are parts of logic.)

Kessler, E. (1988) 'The intellective soul', in C.B. Schmitt *et al.* (eds) *The Cambridge History of Renaissance Philosophy*, Cambridge: Cambridge University Press, 485-534.(Zabarella's psychology in the context of the Renaissance.)

Lohr, C.H. (1988) *Latin Aristotle Commentaries II: Renaissance Authors*, Florence: Olschki, 497-503. (Bibliography including Zabarella's unpublished works and works available only in Renaissance editions.)

Mikkeli, H. (1992) *An Aristotelian Response to Renaissance Humanism. Jacopo Zabarella on the Nature of Arts and Sciences*, Studia Historica 41, Helsinki: Societas Historica Finlandiae.(Basic exposition of Zabarella's conception of the arts and the sciences; good bibliography.)

Randall, J.H. (1961) *The School of Padua and the Emergence of Modern Science*, Padua: Antenore.(Randall's revolutionary hypothesis about the significance of the discussion of *regressus*. Referred to in §6.)

Schmitt, C.B. (1969) 'Experience and Experiment: A Comparison of Zabarella's View with Galileo's in *De motu*', *Studies in the Renaissance* 16: 80-138.(Explains why Zabarella is unacquainted with experiments in the modern sense.)

Wallace, W.A. (1988) 'Randall Redivivus: Galileo and the Paduan Aristotelians', *Journal of the History of Ideas* 49: 133-49.(Defence of Randall's hypothesis supported by new documents.)

Zeami (1363-1443)

Zeami was one of the leading innovators in the art of Nō*, at a time when Zen Buddhism dominated the Japanese intellectual and cultural order. He practised Zen Buddhism (Sōtō branch), and found in Zen teachings the epistemology that gave* Nō *its aesthetic foundations. While his* Nō *treatises are buttressed by his observation of the nature and the workings of the mind, they also reveal Shintō sensibility in their view that the origin of entertainment is sacred.*

Nō (or *noh*) is a theatrical entertainment which mixes singing and dance movements by actors (the main actor usually wears a mask), accompanied by a chorus and musicians. Zeami (real name Kanze Motokiyo) became the head of the Kanze-za, a prominent *Nō* troupe. He composed or reworked at least sixty *Nō* plays, directed and choreographed them, trained his sons as *Nō* actors and playwrights, wrote *Nō* treatises for them, toured with his troupe and entered *Nō* competitions on various occasions.

A passage in the treatise *Kakyō (A Mirror Held to the Flower)* illustrates how Zeami gave his attention to the mental and psychological aspects of *Nō* performance. He discusses several essential practices for the *Nō* actor. For instance, in order to move the audience, he advises that an actor first enunciate his line and then follow with a physical movement. It is more effective for an actor to sing, ‘I cry’, and then move his arm to cover his face to express that gesture, than it is for him to cover his face and then sing ‘I cry’. Zeami explains that in the latter case the actor’s words would have no lingering effect. He holds that this slight staggering of the auditory and visual inputs is crucial in creating an artistic effect.

Zeami places at the foundation of his art the same development of mental awareness that is sought in Zen Buddhism (see Buddhist philosophy, Japanese). In *Kakyō*, he notes that to achieve a graceful appearance, the actor must see what is at his right, left and in front of him with his physical eyes, but he must also train himself to see himself from behind with his mind. Zeami terms this ability to see what the physical eyes cannot see *riken no ken* (seeing separate [from an ego-bound mode of] seeing). The cultivation of this mental capacity, which transcends our ego-centred perspective, requires of the actor a conscious training of his mind. Zeami does not say outright that a *Nō* actor must go through meditation practice as a Zen monk, but he seems to maintain that the mental practice involved in *Nō* training and Zen practice is one and the same. Through his assiduous practice and training, the *Nō* actor will eventually have to attain awakening (*satori*), just as a Zen monk must, and he then must further incorporate his awakening into his performance. Zeami warns his sons that expert spectators also see into the accomplishment of the actor and his performance with their *riken no ken*. Thus, it is essential for the actor to develop his *riken no ken* to take account of even the critical eyes of the spectators.

Because of the homogeneity of the mindset required of the *Nō* actor and Zen practitioner, the aesthetic hierarchy of *Nō*, at least at the time of Zeami, was heavily informed by Zen taste. Accordingly, what Zeami terms the higher sort of plays or performances do not consist of surface beauty that merely please eyes and ears. Rather, they comprise plays and performances that possess an ethereal beauty like that of the heavenly maiden, or which can be compared to ‘white plum blossoms’ or the ‘icy quality’ of ‘snow piled up in a silver container’. According to Zeami, this sort of beauty is pure, beyond any conscious manipulation of mere technique, and evokes deep emotional resonance in the minds of the audience (see the *Kakyō*, *Kyūi (Notes on the Nine Levels)* and *Shūgyoku tokka (Finding Gems and Gaining the Flower)*).

Zeami could not neglect Zen partly because the Shogun Yoshimochi (reigned 1408-28) patronized it. However, Shintō sensibility was also inherently present in his formation of *Nō* as a distinct genre of art. For instance, Zeami ascribes one of the origins of *Nō* to the Shintō account of the ‘incident of the heavenly rock cave’, in which the Sun Goddess Amaterasu hid herself inside a rock cave and the entire world became dark. Ama no Uzume, a goddess, performed a shamanic dance in a trance and aroused the other deities present. The uproarious noise these deities made brought Amaterasu out of the cave; the world became bright again, and the gods’ faces (*omo*) shone ‘white’ (*shiro*) in the bright light. In the *Fūshi kaden (Teachings on Style and the Flower)* Zeami subscribes to this traditional etymology of the word *omoshiro* (delightful or interesting).

We find in Zeami’s treatises penetrating psychological insights, an applied Zen epistemology and vibrant religious

sensibility. All of these contributed to his articulation of the art of *Nō* as a path of self-awakening and a god-inspired entertainment that gives pleasure to the people.

See also: Aesthetics, Japanese; Buddhist philosophy, Japanese; Shintō

MICHIKO YUSA

List of works

Zeami (1400-33) *Zeami-shū (Works of Zeami)*, ed. Konishi J., Tokyo: Chikuma Shobō, 1970; ed. Omote A. and Katō S., *Zeami. Zenchiku*, Nihon Shisō Taikei 24, Tokyo: Iwanami Shoten, 1974.(Konishi contains the original text and translation into modern Japanese; helpful. Omote and Kato contains the original texts of Zeami's and Zenchiku's *Nō* treatises with extensive notes.)

Zeami (1400-6) *Fūshi kaden (Teachings on Style and the Flower)*, ed. and trans. J.T. Rimer and Yamazaki Masakazu, *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press, 1984.(Zeami's earliest and most complete work. As well as instructions for actors, the last part of the work contains Zeami's own philosophy of the art of *Nō* performance.)

Zeami (1420) *Shikadō (The True Path to the Flower)* ed. and trans. J.T. Rimer and M. Yamazaki, *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press, 1984.(An instruction on how to achieve the 'free rank' of the art of performance, going beyond the ordinary level; the importance of mental attitude is introduced.)

Zeami (*c*.1420) *Yūgaku shūdō fūken (Disciplines for the Joy of Art)*, trans. Izutsu Toshihiko and Izutsu Toyo, *The Theory of Beauty in the Classical Aesthetics of Japan*, The Hague: Martinus Nijhoff, 1981; ed. and trans. J.T. Rimer and M. Yamazaki, *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press, 1984.(An instruction on how to raise a young actor into a mature actor.)

Zeami (1423) *Nōsakusho (also known as Sandō, The Three Elements for Composing a Play)*, ed. and trans. J.T. Rimer and M. Yamazaki, *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press, 1984.(An instruction by Zeami to his son on how to write good *Nō* plays.)

Zeami (1424) *Kakyō (A Mirror Held to the Flower)*, trans. M. Nearman, '*Kakyō*, Zeami's Fundamental Principles of Acting', *Monumenta Nipponica* 1982, 37 (3): 333-74 and 37 (4): 459-96; ed. and trans. J.T. Rimer and M. Yamazaki, *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press, 1984.(Zeami's secrets of how to perform well, contains the gist of his understanding that leads to successful *Nō* performance; very important work.)

Zeami (*c*.1424) *Ongyoku kowadashi kuden (The Art of Singing and Articulation)*, ed. Konishi J., Tokyo: Chikuma Shobō, 1970.(Instructions on how to use the voice in performance.)

Zeami (1428) *Kyūi (Notes on the Nine Levels)*; trans. Izutsu Toshihiko and Izutsu Toyo, *The Theory of Beauty in the Classical Aesthetics of Japan*, The Hague: Martinus Nijhoff, 1981; ed. and trans. J.T. Rimer and M. Yamazaki, *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press, 1984.(Divides *Nō* plays into nine levels and describes how actors learn to play different levels as they reach different degrees of accomplishment.)

Zeami (1428) *Shūgyoku tokka (Finding Gems and Gaining the Flower)*, trans. Izutsu Toshihiko and Izutsu Toyo, *The Theory of Beauty in the Classical Aesthetics of Japan*, The Hague: Martinus Nijhoff, 1981; ed. and trans. J.T. Rimer and M. Yamazaki, *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press, 1984.(An instruction handed down to his son-in-law, containing a summary of Zeami's secret of successful performance.)

Zeami (1430) *Shūdōsho (Learning the Way)*, ed. and trans. J.T. Rimer and M. Yamazaki, *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press, 1984.(An instruction to the members of the Kanze troop, regarding the responsibilities of each.)

Zeami (1430) *Sarugaku dangi (An Account of Zeami's Reflection on Art)*, ed. and trans. J.T. Rimer and M. Yamazaki, *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press, 1984.(A record of Zeami's words and comments, recorded and handed down by his second son, Kanze Motoyoshi.)

Zeami (1433) *Kyakuraika (The Flower of Returning to the Beginning)*, ed. Konishi J. Tokyo: Chikuma Shobō, 1970.(Written for his son, this short work intimates but does not spell out the highest stage in the art of *Nō* performance.)

References and further reading

Keene, D. (1966) *Nō: The Classical Theatre of Japan*, Tokyo: Kodansha International.(Includes a comprehensive

‘List of Nō Plays Currently Performed’.)
Keene, D. (ed.) (1970) *Twenty Plays of the Nō Theatre*, New York: Columbia University Press.(An English translation of some often played *Nō* texts.)
Izutsu, T. and Izutsu, T. (1981) *The Theory of Beauty in the Classical Aesthetics of Japan*, The Hague: Martinus Nijhoff.(Includes a discussion of the metaphysical background of the theory of *Nō*, and English translations of *Kyūi*, *Yūgaku shūdō fūken* and *Shūgyoku tokka*.)
Hare, T.B. (1986) *Zeami’s Style*, Stanford, CA: Stanford University Press.(A fine and detailed study of Zeami’s life and works in English, well-supported by the wealth of Japanese scholarship on the history and performance of *Nō*.)
Kōsai, T. (1962) *Zeami Shinkō (New Thoughts About Zeami)*, Tokyo: Wanya Shoten, 3-8.(Articles in this work cover subjects such as Zeami’s possible affiliation with the Sōtō branch of Zen Buddhism, his affiliation with the Fuganji Temple and his knowledge of Zen literature.)
Nose, A. (1930) *Zeami Jūrokubushū Hyōshaku (Exposition of Sixteen Works by Zeami)*, Tokyo: Iwanami Shoten. (A classic study of Zeami’s *Nō* treatises with detailed commentaries.)
Rimer, J.T. and Yamazaki, M. (ed. and trans.) (1984) *On the Art of the Nō Drama: The Major Treatises of Zeami*, Princeton, NJ: Princeton University Press.(English translation of eight treatises written by Zeami plus the *Sarugaku dangi*. The volume includes a helpful glossary of important persons mentioned in the text and a list of plays and entertainments mentioned in the text.)
Sanari, K. (ed.) (1964) *Yōkyoku Taikan (Compendium of Nō Plays)*, Tokyo: Meiji Shoin, 5 vols.(A total of 235 *Nō* plays are compiled; essential for the study of *Nō* plays.)
Smethurst, M. (1989) *The Artistry of Aeschylus and Zeami: A Comparative Study of Greek Tragedy and Nō*, Princeton, NJ: Princeton University Press.(An eye-opening application of Zeami’s dramaturgy to reconstructing Aeschylus’s works.)
Yasuda, K. (1989) *Masterworks of the Nō Theater*, Bloomington, IN: Indiana University Press.(Detailed English translation of 17 *Nō* plays, each accompanied by an informative introduction.)
Yusa, M. (1987) ‘*Riken no Ken*, Zeami’s Theory of Acting and Theatrical Appreciation’, *Monumenta Nipponica* 42 (3): 331-45.(A closer study of Zeami’s theory of *riken no ken* reveals its affinity with the teaching of ‘the original mind’ and ‘awakening’ espoused in Zen Buddhism.)

Zeno of Citium (334-262 BC)

Zeno of Citium, a Greek philosopher from Cyprus, founded the Stoic school in Athens c.*300 BC. His background and training lay in various branches of the Socratic tradition, including the Platonic Academy, but especially Cynicism. His controversial Republic was a utopian treatise, founded on the abolition of most civic norms and institutions. He laid the main foundations of Stoic doctrine in all areas except perhaps logic.*

Zeno was born at Citium, a town in Cyprus of mixed Greek and Phoenician culture. His father, a merchant, used to bring copies of Socratic dialogues back from his visits to Athens. It was perhaps these that inspired Zeno, at the age of 22, to move to Athens and study philosophy in the milieu in which Socrates himself had lived.

Zeno chose a series of teachers representing different branches of the Socratic tradition, starting with the Cynic Crates. Cynic ethics (see Cynics) remained a dominant influence on his thought. Polemo, the head of the Platonic Academy, (see Platonism, Early and Middle §§1,8) and the Megarian philosopher Stilpo (see Megarian School) were among his other teachers, and both will have supplied a primarily ethical orientation as well, although with some input of metaphysics. Stilpo's most celebrated doctrine was the self-sufficiency of the wise. Finally Diodorus Cronus, whose classes he attended along with the future logician Philo, represented the dialectical side of the Socratic tradition, offering Zeno a training in logic as well as in the study of sophisms (see Philo the Dialectician).

In time - around the turn of the century - Zeno formed his own philosophical group, at first known as Zenonians but eventually dubbed 'Stoics' after the Painted Stoa in which they used to congregate. Zeno remained in Athens until his death in 262 BC, and the school he had founded became the dominant school of the Hellenistic age.

Zeno was more an inspirational than a systematic philosophical writer, and later generations took on the task of formalizing and developing his philosophy. The most provocative of his twenty-seven recorded works - reported also to be his first - was a utopian political tract, the *Republic*, whose title echoes those of like-named works not just by Plato but also by the Cynic Diogenes of Sinope. Its exact character is disputed. Some fragments suggest a cosmopolitan theme, anticipating and no doubt inspiring the later Stoic conception of the cosmic city (see Stoicism §18). But many others make it clear that its main proposals were for the radical reform of a single political community. In characteristically Cynic fashion, most civic institutions - temples, law courts, coinage, differential dress for the sexes, conventional education, marriage, and so on - were to be abolished. The city was to be a community of the wise, united by concord and, it seems, a sublimated form of erotic love. Some later Stoics found this work so deeply embarrassing that they branded it, or parts of it, as spurious. However Chrysippus, his second successor, wholeheartedly defended its proposals and tenets.

Despite the controversy surrounding his *Republic*, Zeno undoubtedly established, often against tough opposition, the main canonical positions of Stoic ethics, epistemology and physics. His most distinctive ethical innovation was the insistence that moral indifferents like wealth and poverty could be placed in a natural ranking order of preferability. In developing it he faced resistance on the one side from his former Platonist teacher Polemo, who insisted that such a ranking must be a moral one, in terms of relative goodness, on the other from his colleague Ariston of Chios (§2) who denied that any differentiation between 'indifferents' was possible. His epistemological thesis that certainty can be attained through the senses was persistently challenged by the New Academy (see Arcesilaus §2) and adjusted to meet the criticisms. He approved of dialectic mainly as a defence against the sophistries of others, and is not known to have made any major contribution of his own to Stoic logic.

In some areas it was less his doctrines than his daring style of argument that forced later Stoics to rally to his defence. The flavour can be glimpsed from two of his more controversial syllogisms. 'It would be reasonable to honour the gods; it would not be reasonable to honour the non-existent; therefore the gods exist.' 'Voice comes through the windpipe; if it were coming from the brain, it would not come through the windpipe; where speech comes from is where voice also comes from; but speech comes from the mind; therefore the mind is not in the brain.' (This latter argument obliged Chrysippus and other Stoics to fight a rearguard action against the latest anatomical discoveries.)

See also: Cleanthes

DAVID SEDLEY

References and further reading

Arnim, H. von (1903-5) *Stoicorum Veterum Fragmenta (Fragments of the Early Stoics)*, Leipzig: Teubner, with vol. 4, indexes, by M. Adler, 1924, vol. 1, 3-72.(The standard collection of early Stoic fragments, in Greek and Latin, commonly abbreviated as *SVF*; Zeno is in volume 1, pages 1-72.)

Baldry, H.C. (1959) ‘Zero’s Ideal State’, *Journal of Hellenic Studies* 79: 3-15.(Reconstruction of Zeno’s *Republic*.)

Erskine, A. (1990) *The Hellenistic Stoa*, London: Duckworth.(A challenging reinterpretation of Stoic political thought and practice, including Zeno’s *Republic*.)

Graeser, A. (1975) *Zenon von Kitium: Positionen und Probleme (Zeno of Citium: Positions and Problems)*, Berlin: de Gruyter.(Too important to ignore, but a very difficult read, made even harder by the ubiquitous misprints.)

Schofield, M. (1983) ‘The syllogisms of Zeno of Citium’, *Phronesis* 28: 31-58.(Explains Zeno’s philosophical style and the problems it created for later generations of Stoics.)

Schofield, M. (1991) *The Stoic Idea of the City*, Cambridge: Cambridge University Press. (Outstanding reconstruction of Zeno’s *Republic* and its place in Stoic political thought.)

Zeno of Citium (334-262 BC) Fragments, in A.C. Pearson (ed.) *The Fragments of Zeno and Cleanthes*, Cambridge: Cambridge University Press, 1891.(Still useful, but mainly for specialists.)

Zeno of Elea (*fl. c.*450 BC)

The Greek philosopher Zeno of Elea was celebrated for his paradoxes. Aristotle called him the 'founder of dialectic'. He wrote in order to defend the Eleatic metaphysics of his fellow citizen and friend Parmenides, according to whom reality is single, changeless and homogeneous. Zeno's strength was the production of intriguing arguments which seem to show that apparently straightforward features of the world - most notably plurality and motion - are riddled with contradiction. At the very least he succeeded in establishing that hard thought is required to make sense of plurality and motion. His paradoxes stimulated the atomists, Aristotle and numerous philosophers since to reflect on unity, infinity, continuity and the structure of space and time. Although Zeno wrote a book full of arguments, very few of his actual words have survived. Secondary reports (some from Plato and Aristotle) probably preserve accurately the essence of Zeno's arguments. Even so, we know only a fraction of the total.

According to Plato the arguments in Zeno's book were of this form: if there are many things, then the same things are both F and not-F; since the same things cannot be both F and not-F, there cannot be many things. Two instances of this form have been preserved: if there were many things, then the same things would be both limited and unlimited; and the same things would be both large (that is, of infinite size) and small (that is, of no size). Quite how the components of these arguments work is not clear. Things are limited (in number), Zeno says, because they are just so many, rather than more or less, while they are unlimited (in number) because any two of them must have a third between them, which separates them and makes them two. Things are of infinite size because anything that exists must have some size: yet anything that has size is divisible into parts which themselves have some size, so that each and every thing will contain an infinite number of extended parts. On the other hand, each thing has no size: for if there are to be many things there have to be some things which are single, unitary things, and these will have no size since anything with size would be a collection of parts.

Zeno's arguments concerning motion have a different form. Aristotle reports four arguments. According to the Dichotomy, motion is impossible because in order to cover any distance it is necessary first to cover half the distance, then half the remainder, and so on without limit. The Achilles is a variant of this: the speedy Achilles will never overtake a tortoise once he has allowed it a head start because Achilles has an endless series of tasks to perform, and each time Achilles sets off to catch up with the tortoise it will turn out that, by the time Achilles arrives at where the tortoise was when he set off, the tortoise has moved on slightly. Another argument, the Arrow, purports to show that an arrow apparently in motion is in fact stationary at each instant of its 'flight', since at each instant it occupies a region of space equal in size to itself. The Moving Rows describes three rows (or streams) of equal-sized bodies, one stationary and the other two moving at equal speeds in opposite directions. If each body is one metre long, then the time taken for a body to cover two metres equals the time taken for it to cover four metres (since a moving body will pass two stationary bodies while passing four bodies moving in the opposite direction), and that might be thought impossible.

Zeno's arguments must be resolvable, since the world obviously does contain a plurality of things in motion. There is little agreement, however, on how they should be resolved. Some points can be identified which may have misled Zeno. It is not true, for example, that the sum of an infinite collection of parts, each of which has size, must itself be of an infinite size (it will be false if the parts are of proportionally decreasing size); and something in motion will pass stationary bodies and moving bodies at different velocities. In many other cases, however, there is no general agreement as to the fallacy, if any exists, of Zeno's argument.

1 Life and works

Information about Zeno's life is scarce. Plato (§15), in his *Parmenides*, reports a visit to Athens by Zeno, in the company of Parmenides, usually dated around 450 *bc*. Zeno was then in his late thirties, a younger contemporary of Parmenides who was then in his mid-sixties. He had associated with Parmenides for some time. By the time of this visit Zeno had already produced a book, the reputation of which preceded him to Athens. Diogenes Laertius writes about Zeno's courage and his political activities, detailing some unreliable stories about his heroic death at the hands of a tyrant.

Reliable information about Zeno's writings is equally paltry. Some of his arguments have been saved verbatim

(mainly those concerning plurality); others have survived only via brief reports (mainly those concerning motion). In most cases, the arguments are baldly stated, and some effort is required to reveal their strength and structure.

2 The arguments against plurality

There is no universal agreement on what Zeno hoped to achieve with his arguments. The orthodox view, stemming from Plato, is that Zeno's purpose was to defend Parmenides' monism, which denied the reality of both plurality and motion. Zeno defended Parmenides by attacking those opponents who mocked his position. This view of Zeno has two consequences. First, Zeno's arguments are *ad hominem*: so, while Zeno's opponents should be committed to any assumptions they involve, those assumptions need not be endorsed by Zeno himself. Second, the Parmenidean position should, ideally, be immune to the destructive force of Zeno's arguments.

Monism, the view that only one thing exists, is not very plausible. An intuitively more attractive position is pluralism:

(1) The world contains more than one thing.

Pluralists differ from one another as regards *which* things the world contains. We do not know which pluralists Zeno was attacking. Some have argued for a philosophically sophisticated target - for example, Pythagoreans - but this is now widely rejected. A more appealing denial of monism is an unsophisticated pluralism. 'Simple pluralism' is the common-sense view which holds both (1) and

(2) These things are ordinary objects such as horses, human beings, oaks,…

According to the sixth-century philosopher Philoponus, Zeno's target was simple pluralism:

> Those who introduce plurality put their confidence in its self-evidence; for there exist horses and people and a variety of individual things, and the aggregation of these produces plurality. This self-evidence therefore Zeno tried to overthrow…
>
> (*On Aristotle's Physics* 42.18-21 in Lee 1936: §8)

Zeno's challenge to simple pluralism is successful, in that he forces anti-Parmenideans to go beyond common sense. Post-Zenonian atomism (see Atomism, Ancient; Democritus §§2-3), for example, conjoins (1) not with (2) but with

(3) These things are indivisible masses of atomic stuff moving in empty space.

Aristotelian pluralism retains (2), but only by working out a far more sophisticated understanding of what constitutes a unity than is provided by common sense (see Aristotle §§11-14)).

Several of Zeno's arguments rely on a pluralist's accepting that any magnitude can be divided infinitely many times. However, Parmenides' One would not be vulnerable to these arguments, because the only premise concerning infinite divisibility that the arguments require is

(ID) If anything at all is divisible, then whatever is extended is divisible.

Since Parmenides (§5) held that there is nothing which is divisible, the conditional (ID) would pose no threat for him. Simple pluralists, by contrast, think that the whole is divisible - that is what makes them pluralists. (ID) relies on the thought that there is something objectionably arbitrary in a world which is a homogeneous and continuous mass of existing things being divisible, but only up to a certain point: being divisible just this far (into horses and oaks,…) but no further (not, for example, into horse bits and oak bits). Subsequent pluralists could nevertheless evade Zeno's arguments by denying (ID), as the atomists did; Democritus denied the homogeneity of the world, by introducing two quite distinct types of real entity - atoms and void. In such a case (ID) is no longer plausible, since an explanation can be provided of why the world should be divisible into atoms and no further (see Democritus §2).

It should be noted, however, that Plato's interpretation of Zeno as supporting Parmenides is not mandatory. Eudemus, the pupil of Aristotle, read him as attacking not only pluralities but also Parmenides' One, and this alternative view of Zeno as a nihilist, rather than a monist, gained some currency (Seneca, *Letters* 88.44-5).

3 The ‘limited and unlimited’ argument

The ‘limited and unlimited’ argument is fragment 3. It is preserved by Simplicius, and appears to be a direct quotation from Zeno.

[(A)] If there are many things, it is necessary that they are just as many as they are, and neither more nor less than that. But if they are as many as they are they will be limited.
[(B)] If there are many things, the things that are are unlimited; for there are always others between the things that are, and again others between those. And thus the things that are are unlimited.

(A) is a strong argument, granted a false, although very appealing, assumption. Suppose

(A1) There are many things [call them the totality T].

Then

(A2) It is necessary that they are just as many as they are [T contains exactly as many items as T]
(A3) and neither more [T does not contain as many items as T^{+}, which contains everything T contains, along with some extra items] nor less [T does not contain as many items as T^{-}, which contains almost everything T contains, apart from some items].

So

(A4) If they are as many as they are they will be limited [the content of T is limited to just so many items, in contrast to the number in T^{+} or the number in T^{-}].

This argument requires the assumption

(A5) If T^{*} contains everything that T contains, along with some extra items, then T^{*} contains more items than T and T contains fewer items than T^{*}.

(A5) is extremely appealing: for it can seem that T^{*} *must* contain more items than T, just because it is T with some extra items added. But (A5) is false. It claims that if T is a subset of T^{*} then T and T^{*} do not contain the same number of items, and that is not unrestrictedly true: it is false if T and T^{*} each contain infinitely many things (see Infinity §6).

The other half of Zeno’s argument, (B), shows in contrast that (A4) fails for any totality of units comprising the simple pluralists’ world. (B) provides a method for adding extra items to any specification of T, and so expanding it into a totality T^{+}, which according to (A5) contains more items than T.

The method Zeno provides for expanding T into T^{+} is this: *there are always others between the things that are, and again others between those*. Suppose T is specified as this horse, that human being, the oak over there… (B) concentrates attention on the claim that this horse and that human being, for example, are distinct units. A simple intuition lies behind including this horse and that human being within T: namely, that they are distinct units because there is something else between them - for example, this saddle. But an obvious question then arises: what makes this saddle and this horse distinct units? Zeno envisages two possible answers. First, that they too are distinct because there is something else between them - for example, this horse blanket. Second, that they are distinct even though they are in contact. If the simple pluralist offers the first answer, a method of expanding any specification of T follows immediately (since the next question to arise will be: what makes this saddle and this horse blanket distinct items?). If the second answer is offered, another expansion method suggests itself. If this horse and this saddle can be distinct units, even though they are in contact, then why not say, by parity of reasoning, that this leftmost horse chunk and that rightmost horse chunk are distinct units? In either case (A4) fails to hold: any specification of T can be expanded into a totality containing extra members.

It is necessary to go beyond the resources of simple pluralism to respond to this argument. Zeno assumes that his opponents view the world as a homogeneous and continuous mass of existing things. In the absence of absolutely basic differences between things, the only way to pick out units will be by their relations to one another - for

example, by their being spatially discrete. (B) is a cogent argument against that position. One response would deny the assumption of homogeneity. Democritus was able to identify units because his world contained atomic stuff and void as two fundamentally different types of entity. Any two atoms are distinct because they are separated from one another by void, while an atom and the void surrounding it are distinct because they are intrinsically different. But to resist Zeno's argument by making that move is to abandon simple pluralism. Democritus' world is not a plurality of horses and human beings, but of atoms. A deeper response would involve a more sophisticated view of units and pluralities: for example, the Aristotelian ontology of substantial unities rehabilitates entities like this horse and that human being and picks them out as units in terms of their functional organization (see Aristotle §13).

4 The 'large and small' argument

The 'large and small' argument can be found in fragments 1 and 2, which consist of material quoted by Simplicius in his commentary on Aristotle's *Physics* (he reports Zeno's 'limited and unlimited' argument in the same pages). Simplicius reports three different sub-arguments, contributing to a single Zenonian argument: that if there is a plurality, the same things will be 'large' (of infinite magnitude) and 'small' (of no magnitude).

According to Simplicius (in fr. 2.17-18), Zeno argued that the units of a plurality would have *no* magnitude by means of the following argument:

[(C)] It has no magnitude since each of the many is the same as itself and one.

Argument (C) requires three premises, including (ID) (see above):

(C1) If there is a plurality there must be units *which are not themselves pluralities*.
(C2) Whatever is divisible is *ipso facto* a plurality.

According to the simple pluralist there is a plurality. So given (C1) there must be non-plural units. But those units cannot have magnitude. For according to (ID), if they had magnitude, they would be divisible, in which case according to (C2) they would themselves be pluralities, contrary to (C1). So (ID), (C1) and (C2) imply that any unit (anything that is *the same as itself and one*) will have no magnitude. Argument (C) is valid. The way to evade the argument is to deny one of its premises. The atomists deny (ID), while denial of either (C1) or (C2) augurs a more sophisticated understanding of the relation between unity and plurality.

Before giving Zeno's argument for the conclusion that the units of a plurality must have unlimited magnitude, Simplicius (fr. 2.8-14) reports a Zenonian argument that what has no magnitude does not exist:

[(D)] For if it were added to something else that exists, it would make it no larger; for if it were of no magnitude, but were added, the thing it was added to could not increase in magnitude. And thus what was added would in fact be nothing. If when it is taken away the other thing is no smaller, and again when it is added will not increase, it is clear that what was added was nothing nor again what was taken away.

Adding what has no magnitude (for example, a point) to something would not increase its magnitude, and so would be tantamount to adding nothing. Zeno concludes that what has no magnitude would *be* nothing, and so would not exist. It seems, however, that Zeno's argument will inevitably be question-begging. Imagine adding an electric charge to a metal sphere: the sphere's *magnitude* will not be increased, but adding an electric charge is not tantamount to adding nothing, since the addition may greatly affect the sphere's behaviour. Zeno's conclusion would follow from the assumption that the only significant differences are differences in magnitude: but that assumption is worryingly close to the conclusion Zeno wants to establish by argument (D), namely:

(D1) Whatever exists has magnitude.

Still Zeno's endorsement of (D1) would be consistent with his defending Parmenides, according to whom what exists is apparently spherical in shape (see Parmenides §8). What is more interesting than the argument for (D1), however, is the way in which Zeno uses argument (D) in order to show that the units of a plurality will be of unlimited magnitude.

Fragment 1 gives the argument for this conclusion, which Simplicius puts in Zeno's own words.

For having first proved [(D) above] that if what is had no magnitude it would not even exist, [Zeno] goes on

[(E)] But if it exists, it is necessary for each to have some magnitude and thickness, and for the one part of it to be away from the other. And the same argument holds about the part out in front; for that too will have magnitude and a part of it will be out in front. Indeed it is the same thing to say this once and to go on saying it always; for no such part of it will be last, nor will there not be one part related to another.

Thus if there are many things, it is necessary that they are both small and large: so small as not to have magnitude [the conclusion of argument (C)], so large as to be unlimited [the conclusion of argument (E)].

In argument (E) Zeno relies on the earlier argument (D), and the claim (ID) that if anything is divisible, then whatever is extended is divisible.

Each unit of a plurality must have some magnitude: Zeno argued in (D) that what has no magnitude does not exist, so units without magnitude would not exist and could not compose a plurality. But according to (ID) any product of division which has magnitude is divisible, and the products of that further division would have magnitude (since according to argument (D) they would not exist otherwise). This line of argument can be repeated endlessly. Zeno draws two related conclusions in argument (E), deriving the second from the first:

(E1) Any unit will contain an unlimited *number* of parts.
(E2) unit will be of an unlimited *size*.

Argument (E) can be blocked at different stages. First (E1) might be denied, on the ground that divisions can be carried on indefinitely but cannot be completed. (ID) entitles Zeno to argue that

(4) Given any finite *N* a magnitude admits of more than *N* divisions and contains more than *N* (non-overlapping) parts.

But (4) is not equivalent to, and does not entail

(5) A magnitude admits of an infinite number of divisions, and contains an infinite number of (non-overlapping) parts.

and it is (5) that is required for (E1).

An alternative response would be to allow that (E1) is true, although harmless, and to deny (E2). This response involves a distinction between two types of division (and part). A ‘partial’ division (P-division) of a magnitude would be a division into halves, followed by a division of one of those halves into two quarters, followed by a division of one of those quarters into two eighths,… At any stage only some of the possible divisions of equal-sized parts are made; the products of a completed P-division would include some P-parts that could be further divided. A ‘total’ division (T-division) would be a division into halves, followed by a division of both halves into four quarters, followed by a division of all those quarters into eight eighths,… At any stage all possible divisions of equal sized parts are made; if a T-division could be completed, all the divisions that could have been made would have been made. At any stage of a division the T-parts produced will be of equal size, whereas the P-parts will be of different size.

Zeno probably gets from (E1) to (E2) via an assumption about addition (ADD):

The sum of an infinite number of parts, each having a positive magnitude, is an infinite magnitude.

But (ADD) is not unrestrictedly true. In particular, it would be false of an infinite number of P-parts whose magnitudes decreased in a regular progression. Replacing (ADD) with (ADD′)

The sum of an infinite number of parts, each having an *equal* positive magnitude is an infinite magnitude.

gives something which is unrestrictedly true, but (ADD′) will not license the move from (E1) to (E2) if the parts to which (E1) refers are P-parts, since P-parts are not of equal magnitude.

The orthodoxy is that Zeno’s argument (E) does introduce P-parts. However, it is easy enough to reformulate

Zeno's argument in terms of T-divisions and T-parts (Simplicius, *On Aristotle's Physics* 139.26-140.6 in Lee 1936: §2). There are two ways of responding to argument (E), if it is thus reformulated, although neither of these appealed to ancient thinkers.

First, it could be claimed that even if a P-division could be completed, a T-division could *not* be. The completability of a T-division cannot be inferred from the completability of a P-division, since there would be more T-parts produced by a completed T-division than there are P-parts produced by a completed P-division: there would be a denumerable infinity ($\aleph_0$) of P-parts, but a super-denumerable infinity($2^{\aleph_0}$) of T-parts (since at the first (pre-division) stage there would be 1 P-part and 1 T-part, at the second stage 2 P-parts and 2 T-parts, at the third stage 3 P-parts and 4 T-parts… *n*th stage *n* P-parts and 2^{n-1} T-parts). A proof that not all infinities are of the same size awaited the work of Georg Cantor in the nineteenth century (see Cantor's Theorem; Continuum Hypothesis, the §2; Infinity §1).

An alternative response would be to allow that a T-division can be completed, and accept that the T-parts that result have equal, although *zero* magnitude, but to deny that it follows that the original magnitude composed of those T-parts would have no magnitude. None of those in the ancient world who considered Zenonian arguments in terms of T-divisions found this solution attractive, and the exposition by its modern champions requires sophisticated mathematical resources.

5 The Dichotomy

We now move from Zeno's arguments against plurality to those concerning motion. The Dichotomy is probably the best known of these. It is reported by Aristotle (*Physics* 239b11-13).

(F) [The Dichotomy] asserts the non-existence of motion on the ground that that which is in locomotion must arrive at the half-way stage before it reaches the goal.

Zeno's aim is to redescribe motion in such a way that it can be shown to be impossible. It is impossible to complete an arbitrary journey from A to B - to start at A, move to B, and then stop, for

(F1) A journey from A to B is a series of sub-journeys with no last member: from A to $\frac{1}{2}\mathrm{AB}$, from $\frac{1}{2}\mathrm{AB}$ to $\frac{3}{4}\mathrm{AB}, \ldots.$

and

(F2) It is impossible to complete a series of sub-journeys with no last member.

It would equally be possible to specify instead a series of sub-journeys with no *first* member, from A to $\frac{1}{2}\mathrm{AB}$, preceded by A to $\frac{1}{4}\mathrm{AB}$,…, allege a corresponding impossibility.

The form of argument is valid: if doing *X* is doing *Y*, and doing *Y* is impossible, then doing *X* is impossible. Zeno needs to establish both that some redescription of a journey from A to B, such as (F1) provides, is acceptable; and that a journey so redescribed is impossible, as (F2) alleges. There are, correspondingly, two main strategies for avoiding Zeno's conclusion: rejection of (F1), and rejection of (F2).

A particularly brusque rejection of (F1) is the view that there are minimal distances (*d*) which do not admit of further division, so that there is a last sub-journey, from $\mathrm{B}-d$ to B (and a first, from A to $\mathrm{A}+d$). Although there is little evidence of spatial atomism among Presocratics, a thorough-going spatial atomism is found later in Diodorus Cronus (§2). Plato's successor Xenocrates (§2) may also have adopted this position. But the costs of this strategy are high: for example, as Aristotle pointed out, the mathematical consequences are severe (all lines, for instance, will be commeasurable).

If a denial of (F1) is not feasible, there are further alternatives. One is a straightforward acceptance of (F1) and denial of (F2). Another is to distinguish different readings of (F1) and (F2), and deny that there is any single reading on which both are true together.

An example of the first strategy is Aristotle's claim that the distance between A and B and the time available to cover it are infinite in the same way, since both are divisible into smaller intervals without limit (*Physics* 233a24-31). This disposes of an intuition which may lie behind (F2); a runner would have insufficient time in

which to complete a series of sub-journeys with no last member. But that response will not apply to the following variant of the Dichotomy (*Physics* 263a11-23): passage of an arbitrary time interval $t_1 - t_2$ is impossible, since

(F1′) A time interval $t_1 - t_2$ is a series of sub-intervals with no last member: from t_1 to $\frac{1}{2}(t_1 - t_2)$, from $\frac{1}{2}(t_1 - t_2)$ to $\frac{3}{4}(t_1 - t_2)$,…
(F2′) Passage of a series of sub-intervals with no last member is impossible.

Aristotle's preferred response to the Dichotomy is an instance of the second strategy. He distinguishes between *actual* and *potential* sub-intervals (*Physics* 263a28-b8). An actual interval is an interval made actual by being marked out in some way from its surroundings - for example, by a runner starting and stopping, or by a division being made, at its termini. Intervals which could be, but are not, made actual are merely potential. Actual sub-journeys are journeys over actual sub-intervals. With Aristotle's distinction to appeal to, the premises of the Dichotomy can be taken in two ways: either as

(F1a) A journey from A to B is a series of actual sub-journeys with no last member.
(F2a) It is impossible to complete a series of actual sub-journeys with no last member.

or as

(F1b) A journey from A to B is a series of potential sub-journeys with no last member.
(F2b) It is impossible to complete a series of potential sub-journeys with no last member.

Aristotle's view is that Zeno fails to establish his conclusion because with neither (F1a)/(F2a) nor (F1b)/(F2b) do we have a pair of jointly true premises. (F1b) is a true anti-atomist claim: there is no sub-journey across an interval within AB that *has* to be the last, and there is no sub-journey that *could not* be broken into two further discrete sub-journeys. But (F2b) is false: there is no difficulty in completing a series of sub-journeys that is endless only in the sense that it contains no member that *has* to be the last.

On the other hand, according to Aristotle, (F2a) is true. It would be impossible, he claims, to complete a journey by marking off each of an endless series of sub-journeys: a runner stopping for $\frac{1}{2}$ minute after covering the first half of AB in $\frac{1}{2}$ minute, for $\frac{1}{4}$ minute after covering the next quarter in $\frac{1}{4}$ minute, … and for $(\frac{1}{2})^n$ minutes after covering the next $(\frac{1}{2})^n$ in $(\frac{1}{2})^n$ minutes would *not* cover AB in a staccato fashion in 2 minutes. Such cases have generated considerable discussion. But reaching agreement about these cases, and hence about (F2a), is less urgent if Aristotle's claim that (F1a) is false is both defensible and important: for in that case someone running from A to B does not have to complete a series of actual sub-journeys with no last member, and so would not be trying to do what (F2a) alleges cannot be done.

However, concentrating further on (F1a) would not be a good strategy for dealing with the Dichotomy. For Zeno's problem persists so long as there is *some* series which would have to be completed in order to get from A to B, but which cannot be completed. Suppose Zeno allows that (F1a) is false. Still, consider a series of points at which AB could be divided: $\frac{1}{2}\mathrm{AB}$, $\frac{3}{4}\mathrm{AB}$,… To get from A to B a runner will have to be next to each of those points in turn, although the runner will not stop at each point; and, given any one of those points, there is some other point the runner has to be next to before arriving at B. Here, Zeno will say, we have a series that cannot be completed because it has no last member.

Another distinction is now relevant. There are two series of points that Zeno might have in mind: either the (half-open) series

$$to1 \frac{}{2\mathrm{AB}, \frac{3}{4}\mathrm{AB}, \ldots, \frac{2^n - 1}{2^n}\mathrm{AB}, \ldots (\mathrm{S}) (\text{that is, B not included})} \mathrm{x}$$

or the (closed) series

$$to1 \frac{}{2\mathrm{AB}, \frac{3}{4}\mathrm{AB}, \ldots, \frac{2^n - 1}{2^n}\mathrm{AB}, \ldots, \mathrm{B}(\mathrm{S}') (\text{that is, B included}).} \mathrm{x}$$

Zeno alleges that it is impossible for a runner to start at A, move to B and then stop. The task which Zeno alleges to be impossible might be

(TS) Being adjacent to each member of (S), not being adjacent to any point not in (S), and then after that resting.

It is indeed impossible to complete task (TS): where would a runner who completed that task be? Furthermore, the reason it is impossible to complete (TS) is that the series (S) is a series with no last member. So if moving from A to B and then resting were a matter of completing (TS), motion from A to B would be impossible. But travelling from A to B and then resting is not a matter of performing task (TS) and doing nothing else. It is rather a matter of completing the task

(TS′) Being adjacent to each member of S′, not being adjacent to any point not in S′, and then resting.

and nothing Zeno has said shows that it is impossible to complete (TS′). For the series (S′) does have a last member (namely B), and that is just where a runner who finishes (TS′) will be. Of course Zeno may now try to introduce some other series with no last member that would, although impossible, have to be traversed in order to complete the task (TS′): for example, how can a runner complete (TS′) and get to B, since there is no last point at which the runner has to arrive just before arriving at B? But that question assumes that completing (TS′) would require completing TS and then doing one more thing (arriving at B), and that is a false assumption.

It is plain that Zeno has a strategy for keeping the Dichotomy argument going: he can look for another series which would have to be traversed in order to get from A to B, but which cannot be traversed. Neither the series of actual sub-journeys, nor the series of potential sub-journeys, nor the series of points between A and B provides him with such a series. But Zeno's strategy is open-ended, and two consequences follow from that. First, there is unlikely to be a single resolution of the Dichotomy argument: given different choices of series, it will sometimes be relevant to deny that the series chosen cannot be traversed, and sometimes to deny that it has to be traversed. Second, given the ingenuity of philosophers, there are likely to be ever more subtle redescriptions of a journey from A to B which call for ever more subtle responses.

6 The Achilles

Zeno's Achilles argument is very closely related to that of the Dichotomy. According to Aristotle (*Physics* 239b14-16),

[(G)] The second [argument concerning motion] is the so-called Achilles, and it amounts to this, that in a race the quickest runner can never overtake the slowest, since the pursuer must first reach the point where the pursued started, so that the slower must always hold a lead.

Suppose Achilles is running faster than the tortoise. Then as their runs proceed, the distance between the two will decrease, at a rate determined by their relative velocities. Let the distance between them initially be AB, and suppose their relative velocities are such that after $\frac{1}{2}$ minute it has shrunk to $\frac{1}{2}\mathrm{AB}$, after a further $\frac{1}{4}$ minute to $\frac{1}{4}\mathrm{AB}$,… and so on. Achilles' task is to complete an infinite series of lead-narrowings ($\frac{1}{2}\mathrm{AB}$, $\frac{1}{4}\mathrm{AB}$,…). This series of lead-narrowings will sum to a total distance no more than AB, and thus it is plain that as long as Achilles can traverse a distance AB in a finite time then he can close the distance between himself and the tortoise in a finite time. That is to say that as long as the Dichotomy can be resolved, the Achilles can be resolved. The resolutions will be parallel.

7 The Arrow

Aristotle (*Physics* 239b5-9, 239b30-3) reports the Arrow along with a summary response.

[(H)] Zeno argues fallaciously; for if, he says, everything always rests when it is against what is equal, and what is in locomotion is always in the now, the arrow in locomotion is motionless. But this is false, for time is not composed of indivisible 'nows', no more than is any other magnitude.

The third [argument concerning motion] is the one just mentioned, that the arrow in locomotion is at rest. This follows from assuming that time is composed of 'nows'; for if that is not granted, the conclusion will not follow.

The argument as Aristotle sets it out requires expansion:

(H1) Everything always rests when it occupies a space equal to itself.

(H2) What is moving is always in the now.
(H3) What is in the now occupies a space equal to itself.

Therefore

(H4) What is moving always occupies a space equal to itself [from (H2) and (H3)].

Therefore

(H5) What is moving is always at rest [from (H4) and (H1)].
(H6) Time is composed of indivisible nows.

Therefore

(H7) The ‘flying arrow’ does not move.

(H1) and (H2) are from Aristotle’s text. (H1) glosses the Aristotelian expression ‘is against what is equal’. (H3) is required to get from (H1) and (H2) to (H4). (H6) is included because it is mentioned by Aristotle. Some have suggested that the Aristotelian report gives only a fragment of a wider Zenonian argument; whether or not that is so, (H1)-(H7) present the core of the problem. The orthodoxy is that the phrase ‘the now’ in (H2) and (H3) is an Aristotelian way of talking about unextended temporal instants (or possibly indivisible temporal atoms); according to a less common interpretation the argument raises problems concerning the present.

One main line of approach, endorsed by Aristotle among others, is denial of (H1). Another is to deny an appropriate interpretation of (H3). Rejection of (H6) is also recommended by Aristotle: however, while (H6) may well be false, its relevance to Zeno’s argument is disputed. (H2) seems unassailable.

Aristotle denies (H1) on the grounds that ‘being at rest’ and ‘being in motion’ are defined only over temporal periods, and nothing can move or be at rest at an unextended instant of time: to be at rest is to occupy the same space for some period of time (*Physics* 239a23-b4). In that case (H1) is false since it admits of counter-instances: any object considered at an *instant* will occupy a space equal to itself at that instant, but will not be at rest at that instant. However, Zeno would accommodate this point by amending (H1) to

(H1′) Nothing moves when it occupies a space equal to itself.

and (H5) correspondingly to

(H5′) What is moving never moves.

(H1′) should be unobjectionable to Aristotle, yet (H5′) still seems dangerous.

Another way of denying (H1) would be to introduce definitions of instantaneous motion (and instantaneous rest), and consequently to reject both (H1) and (H1′). Zeno has a response. Someone who tries in this way to deny (H1) and (H1′has two options.

(6) To define motion at an instant derivatively from *motion over a period*: *A* is moving at an instant *t* as long as *A* is moving over a period containing *t*; *A* is moving over a period as long as *A* covers a distance during that period.
(7) To define motion at an instant derivatively from *location at other instants*: *A* is moving at an instant *t* as long as *A* occupies other places at all other instants within some period containing *t*.

Rejecting (H1) and (H1′) on the basis of (6) begs the question against Zeno: it is to say that an arrow is flying at *t* because *t* occurs within a period in which it is flying, and Zeno will not be impressed by such a blunt rejection of the Parmenidean world view. On the other hand, (7) allows the possibility of a quasi-Parmenidean view on which temporal process is replaced by a series of static distributions of objects to places (see Tense and temporal logic); furthermore, if instants were replaced by temporal atoms, (7) would lend support to the view of Diodorus Cronus (§2), who was widely regarded in antiquity as denying the existence of motion.

An alternative line is to challenge (H3). If read as a claim about the *present*, (H3) relies for its appeal on the assumption that the present is an unextended instant separating past and future. That assumption could be denied,

and replaced with an account of the present as a period of indeterminate extension - although Zeno could reasonably demand some independent support for such a revisionary view of the present. Still, if an instantaneous present is rejected, (H3) would no longer be guaranteed. Zeno’s Arrow argument would be no more persuasive than the following: no object can have a size; for every object is, for some place, located *here*; any extended place can be split into part *there* and part *there*; so *here* must be extensionless; so any object located *here* must be extensionless. That argument rests on the false assumption that ‘here’ must be an extensionless location.

Aristotle has another response to Zeno’s argument, which is to reject (H6). Some commentators doubt the relevance of (H6) to Zeno’s argument; and (H6) embodies something which Aristotle is keen to emphasize independently of his concern with Zeno’s arguments. Aristotle’s point is that motion and rest are characteristics which apply irreducibly over periods of time. In much the same way, colour is a feature of extended surfaces which cannot be reduced to features of extensionless points: it would not be legitimate to argue that, since no body is coloured at an extensionless point, there are no coloured bodies.

8 The Moving Rows

Aristotle’s report of the Moving Rows is condensed, and a diagram, such as is provided by Alexander of Aphrodisias, is useful (Aristotle, *Physics* 239b33-40a1).

[(J)] The fourth [argument about motion] is the one about equal bodies which move in opposite directions past equal bodies in a stadium at equal speed, the one row from the end of the stadium and the other from the middle - in which he thinks it follows that half the time is equal to double.

Figure 1			$A1$	$A2$	$A3$	$A4$		
	$B4$	$B3$	$B2$	$B1$				
					$C1$	$C2$	$C3$	$C4$

Figure 2	$A1$	$A2$	$A3$	$A4$
	$B4$	$B3$	$B2$	$B1$
	$C1$	$C2$	$C3$	$C4$

The *A*s, *B*s and *C*s are of the same size; the *A*s are stationary, and the *B*s and *C*s are moving at equal speeds in opposite directions. The two examples above show how the positions of the bodies change over the time taken for *B*1 to go past two of the *A*s. Neither the significance of the conclusion, that ‘half the time is equal to double’, nor Zeno’s route to that conclusion are altogether clear. Suppose Zeno assumed that

(J1) A body moving at a certain velocity will take the same time to pass two bodies of the same size, even if one is stationary and one moving.

He could then argue in this way. Suppose the time it takes *B*1 to move from its position in the first example to its position in the second example is *t*. In the second example *B*1 has arrived at *C*4, and so

(J2) The time it takes *B*1 to move from its starting position to *C*4 is *t*.

But it is equally true that *t* is the time taken for *B*1 to get from its starting position to *A*2; then by assumption (J1) it will take *B*1 the same time (*t*) to get from its starting position to *C*2 (since *C*1 and *C*2 are together equal in size to *A*1 and *A*2); and therefore it will take *B*1 twice as long to get to *C*4 (since *C*3 and *C*4 together represent an extra distance to be covered which is equal in size to *C*1 and *C*2, and *B*1 continues at a constant speed); therefore

(J3) The time it takes *B*1 to move from its starting position to *C*4 is 2*t*.

(J2) and (J3) together entail

(J4) $t = 2t$

If Aristotle’s diagnosis is based on an accurate presentation of Zeno’s argument, then Zeno’s argument is weak. For (J1) is plainly false, and in the absence of (J1) no contradiction threatens: *B*1 passes twice as many *C*s as *A*s in a given period of time, but this is because the *C*s but not the *A*s are moving relative to *B*1. Some have found a

better argument for Zeno by supposing that the bodies in question are atomic, moving in an atomically structured time. It is not likely, though, that Zeno had such a target in mind; and it is of course possible that Zeno offered the occasional weak argument. It is also possible that Zeno sought to direct attention to the relativity of velocity and motion, and by doing so to impugn their status: a similar suspicion of what is relative is found slightly later in the atomist Democritus, who downgraded sensible properties such as sweet, bitter, hot and cold because they depended as much on the perceiver as on what is perceived (see Democritus §5).

9 Two other arguments

Zeno's reputation is founded on his arguments concerning plurality and motion. However, there are ancient reports (of doubtful reliability) which identify two further arguments as Zenonian. Aristotle twice mentions a Zenonian argument against the existence of place (*Physics* 209a23-5, 210b22-4, Lee §38), resting on the thought that if place exists it would itself have to be in a place. Another Zenonian argument mentioned by Aristotle poses this problem (250a19-24; expanded by Simplicius, *On Aristotle's Physics* 1108.18-28, Lee §38, who identifies Protagoras as its target): how can smooth variations give rise to radical and discontinuous differences? When a large mass of corn hits the ground it makes a noise, so that a fraction of that mass (for example, one ten-thousandth of a single grain) should make a fraction of the noise - but in fact it makes no noise at all. The second of these arguments in particular raises large issues, which cannot be adequately considered here.

STEPHEN MAKIN

References and further reading

Allen, R.E. and Furley, D.J. (eds) (1975) *Studies in Presocratic Philosophy*, vol. 2, London: Routledge & Kegan Paul.(Contains important papers by Fränkel, Owen and Vlastos on a wide range of Zeno's arguments.)

Aristotle (*c*. mid 4th century BC) *Physics*, ed. W.D. Ross, Oxford: Oxford University Press, 1949; trans. in J. Barnes (ed.) *The Complete Works of Aristotle*, Princeton, NJ: Princeton University Press, 1985.(Much of book VI is concerned with Zenonian issues.)

Barnes, J. (1979) *The Presocratic Philosophers*, London: Routledge.(Chapters 12 and 13 contain provocative discussions of all Zeno's arguments, including fuller accounts of the two arguments mentioned in §9 above.)

Barnes, J. (1987) *Early Greek Philosophy*, Harmondsworth: Penguin.(Chapter 11 translates Plato's account of Zeno, and the passages relevant to all Zeno's extant arguments, along with their original contexts.)

Cajori, F. (1915) 'The History of Zeno's Arguments on Motion' *American Mathematical Monthly* 22: 1-6, 77-82, 109-15, 143-9, 179-86, 215-20, 253-8, 292-7.(An account of the influence of Zeno's arguments up to Russell and Bergson in the twentieth century.)

Grünbaum, A. (1968) *Modern Science and Zeno's Paradoxes*, London: Allen & Unwin.(Concentrates on the problems Zeno raises for modern mathematical physics; the book requires some mathematical ability.)

Guthrie, W.K.C. (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols. (Volume 2 chapter 1B provides a clear and detailed English-language survey of work on Zeno.)

Kirk, G.S., Raven, J.E. and Schofield, M. (1983) *The Presocratic Philosophers*, Cambridge: Cambridge University Press, 2nd edn.(Chapter 11 provides the Greek text, with English translation, of Zeno's arguments against plurality and motion, with accompanying exposition; the translations in this entry mainly follow Kirk, Raven and Schofield.)

Lear, J. (1981) 'A Note on Zeno's Arrow', *Phronesis*, 26: 91-104.(Avoids the Arrow argument by denying that the present is unextended; also discusses the appropriate way to respond to 'sceptical' puzzles such as Zeno's.)

Lee, H.D.P. (1936) *Zeno of Elea*, Cambridge: Cambridge University Press.(Greek text and English translation of the majority of fragments and passages relevant to Zeno's arguments; Lee's comments have now been somewhat superseded.)

Moore, A.W. (1990) *The Infinite*, London: Routledge.(A relatively non-technical account of mathematical and philosophical problems concerning infinity; chapter 1 on early Greek thought, and chapters 8 and 10 on Cantor and transfinite mathematics, are particularly useful.)

Owen, G.E.L (1957) 'Zeno and the Mathematicians', *Proceedings of the Aristotelian Society* 58: 199-222; repr. in R.E. Allen and D.J. Furley (eds) *Studies in Presocratic Philosophy*, vol. 2, London: Routledge & Kegan Paul, 1975; repr. in W.C. Salmon, *Zeno's Paradoxes*, Indianapolis, IN: Bobbs-Merrill, 1970; repr. in G.E.L. Owen, *Logic, Science and Dialectic*, Cambridge: Cambridge University Press, 1986.(Offers an account of Zeno's arguments as an organized whole; includes a lengthy discussion of the Arrow.)

Plato (*c*.380-67BC) *Parmenides*, trans. F.M. Cornford, *Plato and Parmenides*, London: Routledge & Kegan Paul,

1939; trans. M.L. Gill and P. Ryan, *Plato: Parmenides*, Indianapolis, IN: Hackett, 1996.(Presents Zeno in conversation with Socrates.)

Sainsbury, M. (1988) *Paradoxes*, Cambridge: Cambridge University Press.(Chapter 1 is a philosophically lively discussion of a selection of Zeno's arguments; and includes a number of footnoted questions, designed to draw the reader into the problems Zeno raises.)

Salmon, W.C. (1970) *Zeno's Paradoxes*, Indianapolis, IN: Bobbs-Merrill.(A collection of previously published papers, mostly concerning the arguments about motion; includes influential papers by Black, Thomson and Benacerraf on the possibility of completing infinite series of tasks in staccato fashion.)

Simplicius (AD 538-or after) *On Aristotle's Physics 6*, trans. D. Konstan, London: Duckworth, 1989.(Contains commentary on Aristotle's report and discussion of Zeno's arguments against motion.)

Sorabji, R. (1983) *Time, Creation and the Continuum*, London: Duckworth.(Chapter 21 concerns Zeno's arguments; chapters 22-5 trace out the development of atomist accounts of time, space and motion.)

Stokes, M.C. (1971) *One and Many in Presocratic Philosophy*, Washington, DC: Center for Hellenic Studies.(Relates Zeno to other Presocratic philosophers, including Parmenides, Melissus and Democritus.)

Vlastos, G. (1971) 'A Zenonian Argument against Plurality', in *Essays in Ancient Greek Philosophy*, ed. J.P. Anton and G.L. Kustas, Albany, NY: State University of New York Press.(Reconstruction and discussion of the 'large and small' argument in Zeno's fragments 1 and 2.)

White, M.J. (1992) *The Continuous and the Discrete: Ancient Physical Theories from a Contemporary Perspective*, Oxford: Clarendon Press.(An extended account of the conflict between different ancient accounts of time and space, with some reference to the problems raised by Zeno's arguments.)

Zeno of Elea (*fl. c*.450 BC) Fragments, in H. Diels and W. Kranz (eds) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidemann, 6th edn, 1952, vol. 1, 247-58.(The standard collection of the ancient sources; includes Greek texts with translations in German.)

Zermelo, Ernst (1871-1953)

The German mathematician Ernst Zermelo is today best known for his axiomatization of set theory (1908b), presented in the spirit of Hilbert's early axiomatic programme. Originally working in the calculus of variations and mathematical physics, Zermelo concentrated on set theory after proving that every set can be well-ordered (1904). His proof, based on the axiom of choice, provoked a lively controversy. In the 1930s Zermelo worked on infinitary logic, trying to overcome Gödel's incompleteness results.

1 Life

Ernst Zermelo studied mathematics, physics and philosophy at the universities of Berlin, Halle and Freiburg. In 1894 he obtained his doctorate from Berlin with a dissertation taking a Weierstrassian approach to the calculus of variations (see Analysis, philosophical issues in). He became an assistant to Max Planck, but in 1897 he went to the University of Göttingen where he finished his *Habilitation* thesis on hydrodynamics in 1899. He began to teach as a *Privatdozent* at Göttingen and was made a titular professor in 1905. In 1910 he accepted a chair of mathematics at the University of Zurich, but resigned in 1916 because of poor health. After 1921 he lived in retirement in Freiburg, where he became honorary professor of mathematics in 1926. After National Socialistic students had denounced him, he resigned to avoid being dismissed. Following the war he was reinstated but, then almost blind, was never again able to teach.

2 Work

After working in the calculus of variations, thermodynamics and statistical mechanics, Zermelo published his first paper on set theory, concerning the addition of transfinite cardinals (see Set theory §3), in 1902. His later work on set theory can be seen as a contribution to the lively discussion of Cantor's continuum hypothesis. Cantor had conjectured that every infinite subset of the continuum is either denumerable or has the power of the continuum (1878: 257; see Continuum hypothesis). In 1900, in his seminal Paris lecture, 'Mathematische Probleme', Hilbert had emphasized the importance of proving Cantor's 'remarkable assertion' that every set could be well-ordered, and that this fact could be used to prove the continuum hypothesis. In 1904 Zermelo succeeded in proving the well-ordering theorem using a construction of sets by transfinite recursion justified by appeal to what would later be called the 'axiom of choice' (see Axiom of choice). This proof, together with the axiom of choice, was rejected by many mathematicians working in foundations. Zermelo responded to his critics with a new proof, retaining the axiom of choice and emphasizing the purely formal character of the well-ordering, which has nothing to do with 'spatiotemporal arrangements' (1908a: 107). He accompanied his proof by a polemical discussion of objections, arguing that his critics also used infinite numbers of arbitrary choices and that he was by no means the first to use the axiom.

Zermelo shifted his emphasis to set theory both because of the heavy reactions to his former proof and also because of the fact that set theory began to move into the focus of foundational research in Göttingen. The sensation caused by the publication of the set-theoretical paradoxes by Bertrand Russell in 1903 led to a revision of David Hilbert's early axiomatic programme (see Paradoxes of set and property; Hilbert's programme and formalism). Although Zermelo had independently found a similar paradox in 1899 or 1900, it was only with Russell's publications that it became evident that Hilbert's programme, which centred on the problem of proving the consistency of arithmetic, had to take into account the consistency of logic itself; and that set theory was necessary for an axiomatization of arithmetic. In this vein Zermelo began working on a new axiomatization of set theory by 1905 (see Moore 1982: 152-7), though, unable to find a consistency proof for it, he delayed publication until 1908. In the published version (1908b: 262-7) Zermelo stated seven axioms: the axiom of choice, the power set axiom and the axioms of extensionality, elementary sets, separation, union and infinity (see Set theory §1; Set theory, different systems of). He mentioned the lack of a consistency proof for this set of axioms, but remarked that it was not subject to Russell's paradox since from his axiom of separation (if a propositional function $\mathrm{E}(x)$ is definite for a set *M*, then there is a subset M_{E} containing precisely those elements *x* of *M* for which $\mathrm{E}(x)$ is true) it follows that a universal set cannot be derived.

Zermelo's axiomatization underwent several revisions, the most important of which was the addition of the axiom of replacement, formulated by Dmitry Mirimanoff (1917), Thoralf Skolem (1923) and Abraham Fraenkel (1922),

and adopted by Zermelo in Fraenkel's version. Skolem gave the first first-order exposition of Zermelo's set theory, which grew into the Zermelo-Fraenkel set theory (ZF) of today.

Subsequent work was spawned by Zermelo's opposition to the 'Skolemist' interpretation of the Löwenheim-Skolem theorem, according to which '*every* mathematical theory, even set theory, can be realized in a denumerable model' (Zermelo 1931: 85; original emphasis). In 1929 Zermelo gave a second-order axiomatization of the property of definiteness which was presupposed by the axiom of separation and which was the feature of his set theory at which most of the criticism had been directed. In response to his critics (among them especially Thoralf Skolem), Zermelo created in 1930 a new (supplemented) version of ZF - ZF' - which was formed by dropping the axiom of infinity and reclassifying the axiom of choice as a principle of the background logic. He also introduced the axiom of foundation, which prohibits infinitely descending chains of membership (1930: 31). This led to the introduction of well-founded sets which were used to define a cumulative type hierarchy, consisting of layers (indexed by ordinals) which are sets made up of objects from previous layers. At the annual meeting of the Deutsche Mathematiker-Vereinigung at Bad Elster on 15 September 1931, Zermelo used this hierarchy to propose a new infinitary mathematical logic without any quantifiers (see Moore 1980: 124-8). He claimed, contrary to what was suggested by Gödel's then just recently announced work, that this logic would decide every proposition. These ideas did not find the audience that Zermelo had hoped for, although there are several similarities to today's systems of set theory (see Lavine 1994).

See also: Infinitary logics; Logical and mathematical terms, glossary of; Löwenheim-Skolem theorems and nonstandard models; Set theory; Set theory, different systems of

VOLKER PECKHAUS

List of works

Zermelo, E. (1894) *Untersuchungen zur Variationsrechnung (Investigations on the Calculus of Variations)*, Berlin: Diss.(Zermelo's doctoral dissertation on the calculus of variations.)

Zermelo, E. (1897) 'Hydrodynamische Untersuchungen über die Wirbelbewegungen in einer Kugelfläche' ('Hydrodynamic Investigations into the Spin of a Spherical Surface'), *Zeitschrift für Mathematik und Physik* 47: 201-37, 1902.(Zermelo's *Habilitation* thesis, published several years after its submission.)

Zermelo, E. (1902) 'Über die Addition transfiniter Cardinalzahlen' ('On the Addition of Transfinite Cardinals'), *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse aus dem Jahre 1901*: 34-8.(Zermelo's first paper on set theory.)

Zermelo, E. (1904) 'Beweis, daß jede Menge wohlgeordnet werden kann', *Mathematische Annalen* 59: 514-16; trans. S. Bauer-Mengelberg, 'Proof that Every Set can be Well-Ordered', in van Heijenoort (1967), 139-41. (Includes the heavily debated proof of the well-ordering theorem.)

Zermelo, E. (1908a) 'Neuer Beweis für die Möglichkeit einer Wohlordnung', *Mathematische Annalen* 65: 107-28; trans. S. Bauer-Mengelberg, 'A New Proof of the Possibility of a Well-Ordering', in van Heijenoort (1967), 183-98.(Zermelo's reply to his critics, and a new proof of the well-ordering theorem.)

Zermelo, E. (1908b) 'Untersuchungen über die Grundlagen der Mengenlehre. I', *Mathematische Annalen* 65: 261-81; trans. S. Bauer-Mengelberg, 'Investigations in the Foundations of Set Theory I', in van Heijenoort (1967), 200-15.(The first axiomatization of set theory.)

Zermelo, E. (1929) 'Über den Begriff der Definitheit in der Axiomatik' ('On the Notion of Definiteness in Axiomatics'), *Fundamenta Mathematicae* 14: 339-44.(Second-order axiomatization of the property of definiteness.)

Zermelo, E. (1930) 'Über Grenzzahlen und Mengenbereiche. Neue Untersuchungen über die Grundlagen der Mengenlehre', *Fundamenta Mathematicae* 16: 29-47; trans. M. Hallett, 'On Boundary Numbers and Domains of Sets: New Investigations in the Foundations of Set Theory', in W. Ewald (ed.) *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Oxford: Clarendon Press, 1996, vol. 2, 1,219-33.(Revised version of the axiomatization of set theory.)

Zermelo, E. (1931) 'Über Stufen der Quantifikation und die Logik des Unendlichen' ('On Levels of Quantification and the Logic of the Infinite'), *Jahresbericht der Deutschen Mathematiker-Vereinigung* 40 (2nd section): 85-8.(The introduction of an infinitary logic without quantifiers.)

Zermelo, E. (ed.) (1932) G. Cantor, *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts (Collected Papers with Mathematical and Philosophical Contents)*, Berlin: Springer, 1980.(Still the edition of choice of Cantor's works.)

Zermelo, E. (1935) 'Grundlagen einer allgemeinen Theorie der mathematischen Satzsysteme (erste Mitteilung)'

('Foundations of a General Theory of Systems of Mathematical Sentences (First Communication)'), *Fundamenta Mathematicae* 25: 136-46.(Elaboration of the 1931 paper.)

References and further reading

Cantor, G. (1878) 'Ein Beitrag zur Mannigfaltigkeitslehre' (A Contribution to the Doctrine of Manifolds), *Journal für die reine und angewandte Mathematik* 84: 242-58; repr. in *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, ed. E. Zermelo, Berlin: Springer, 1932, 118-33.(Includes Cantor's continuum hypothesis.)

Fraenkel, A. (1922) 'Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre' ('On the Foundations of Cantor-Zermelo Set Theory'), *Mathematische Annalen* 86: 230-7.(Revision of Zermelo's axiomatization of set theory.)

Heijenoort, J. van (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press.(Important omnibus version with key texts in the history of logic and foundations of mathematics.)

Hilbert, D. (1900) 'Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris 1900', *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse*: 253-97; trans. M.W. Newson, 'Mathematical Problems', *Bulletin of the American Mathematical Society* 8: 437-79, 1902.(Seminal lecture which influenced the direction of mathematical research in several branches.)

Hilbert, D. (1905) 'Über die Grundlagen der Logik und der Arithmetik', in A. Krazer (ed.) *Verhandlungen des dritten internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, Leipzig: Teubner, 174-85; trans. B. Woodward, 'On the Foundations of Logic and Arithmetic', in van Heijenoort (1967), 130-8.(Hilbert's first step towards an axiomatization of arithmetic as a reaction to the publication of Russell's paradox.)

Lavine, S. (1994) *Understanding the Infinite*, Cambridge, MA, and London: Harvard University Press, 134-41.(An account of Zermelo's late second-order set theories.)

Mirimanoff, D. (1917) 'Les antinomies de Russell et de Burali-Forti et le problème fondamental de la théorie des ensembles' ('The Antinomies of Russell and Burali-Forti and the Foundational Problem of Set Theory'), *L'Enseignement mathématique* 19: 37-52.(Revision of Zermelo's axiomatization of set theory.)

Moore, G.H. (1980) 'Beyond First-Order Logic: The Historical Interplay Between Mathematical Logic and Axiomatic Set Theory', *History and Philosophy of Logic* 1: 95-137.(A nontechnical historical overview of the relations between mathematical logic and set theory.)

Moore, G.H. (1982) *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*, Studies in the History of Mathematics and Physical Sciences, 8, New York, Heidelberg and Berlin: Springer. (Despite its title, a full account of Zermelo's writings on set theory, their historical context and their influences in contemporary set theory.)

Peckhaus, V. (1990) '"Ich habe mich wohl gehütet, alle Patronen auf einmal zu verschießen". Ernst Zermelo in Göttingen' ('"I Made Sure Not To Use Up All My Bullets At Once". Ernst Zermelo in Göttingen'), *History and Philosophy of Logic* 11: 19-58.(Focuses on Zermelo's collaboration with Hilbert during his time at Göttingen.)

Skolem, T. (1923) 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre', in *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, Helsinki: Akademiska Bokhandeln, 217-32; repr. in *Selected Works in Logic*, ed. J.E. Fenstad, Oslo, Bergen and Tromsö: Universitetsforlaget, 1970, 137-52; trans. S. Bauer-Mengelberg, 'Some Remarks on Axiomatized Set Theory', in van Heijenoort (1967), 291-301.(First first-order exposition of Zermelo's set theory.)

Zhang Zai (1020-77)

Zhang Zai was a seminal neo-Confucian cosmologist and ethical thinker. Like Zhou Dunyi and Shao Yong, he was inspired by the Yijing (Book of Changes) and its commentaries; unlike them, he worked out a conception based solely on the concept of qi *(cosmic vapour). He espoused an ethical vision, global in spirit, that greatly enhanced the moral significance of Confucianism.*

Zhang Zai was born in Changan, China, and grew up in an age of intellectual ferment. He studied Buddhism, Daoism and even military strategy before embarking on a serious study of the Confucian classics and joining the ranks of the neo-Confucians. He focused his learning on the *Zhongyong (Doctrine of the Mean)* and the *Yijing (Book of Changes)* (see Zhongyong; Yijing). Later, when he lectured on the *Yijing* in the capital, his students included such luminaries as the historian Sima Guang and the philosopher brothers Cheng Hao and Cheng Yi. Zhang held occasional government positions but, since he opposed the reforms of Wang Anshi (1021-86), he balanced his government service with teaching at home. He died of illness while travelling home from the capital.

In 1070, Zhang had an uncanny dream in which he envisaged the formation of the cosmos. He articulated this vision in the book *Zhengmeng (Correcting Youthful Ignorance)*, an attempt to explain the cosmos and man, and even such fundamental terms as the supreme polarity (*taiji*) and the Way (*dao*), in terms of *qi* states, permutations and processes. To many interpreters, this approach resembles reductive materialism, but *qi* in the Chinese view is a fluid proto-material which gives rise to phenomena of all sorts, mental and spiritual as well as physical (see Qi). Hence, unlike reductive materialism, there is no attempt to deny any realms of phenomena such as the mental. However, because Zhang's *qi*-based theory did not seem to provide an adequate account of order in the cosmos and the constancy of human values, it was long overshadowed by the *li* (pattern) based thought of the Cheng brothers and Zhu Xi (see Li). Centuries later, Wang Fuzhi, believing that the Cheng-Zhu idea of *li* was arbitrary and inflexible and thus lent itself to rigid conservatism, strongly advocated a revival of Zhang Zai's teachings. Chinese scholars today tend to praise Zhang's thought as anti-metaphysical and proto-scientific.

Cheng Yi accused Zhang's book of a false objectivism that diluted the Confucian moral message, which Cheng deemed purely a matter of subjective intuition. On his part, Zhang postulated *qi* as the common substance to ground a universal morality of sympathy and benevolence. Whatever the merits of this idea, he wrote the noblest and most compelling moral tract to come out of the neo-Confucian movement, the *Ximing (Western Inscription)*, which reads, in part:

> Heaven is my father and Earth is my mother, and even such a small creature as I find an intimate place in their midst.
>
> Therefore that which fills the universe I regard as my body and that which directs the universe I consider as my nature.
>
> All people are my brothers and sisters, and all things are my companions.
>
> … Wealth, honour, blessings, and benefits are meant for the enrichment of my life, while poverty, humble station, and sorrow are meant to help my fulfillment.
>
> In life I follow and serve [Heaven and Earth]. In death I will be at peace.

(*Ximing* in Chan (1963): 497-8)

See also: Neo-Confucian philosophy; Qi; Zhu Xi

KIRILL OLE THOMPSON

List of works

Zhang Zai (1020-77) *Ximing (The Western Inscription)*, trans. Chan Wing-tsit, *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press, 1963.(This inspiring poetic tract placed Confucian ethics in cosmological perspective. Moreover, it broadened the scope of Confucian 'benevolence' to all of mankind.)

Zhang Zai (1020-77) *Zhengmeng (Correcting Youthful Ignorance)*, trans. Chan Wing-tsit, *A Source Book in*

Chinese Philosophy, Princeton, NJ: Princeton University Press, 1963.(This work sets forth a philosophy of nature, humanity and ethics, based on the concept of *qi*. Notably, this book contains a serious attempt to solve the problem of evil in Confucianism.)

References and further reading

Kasoff, I. (1984) *The Thought of Chang Tsai (1020-1070)*, Cambridge: Cambridge University Press.(Systematic overview of Zhang's philosophy; clear and accessible.)

Huang, S. (1968) 'Chang Tsai's Concept of *Ch'i*', *Philosophy East and West* 18 (3): 247-59.(Analytic account of Zhang's fundamental cosmological concept; somewhat technical.)

Huang, S. (1971) 'The Moral Point of View of Chang Tsai', *Philosophy East and West* 21 (2): 141-56.(Study of the implications of Zhang's *Ximing (Western Inscription)*; clear, but somewhat technical.)

Tang, C. (1956) 'Chang Tsai's Theory of Mind and Its Metaphysical Basis', *Philosophy East and West* 6 (2): 113-36.(Sophisticated analysis of Zhang's moral psychology; rather difficult, but rewarding.)

Zheng Xuan (AD 127-200)

Zheng Xuan, perhaps the most influential commentator on the Confucian classics, is widely credited with constructing both a compelling unitary vision of Chinese civilization and a hierarchy of written authorities which upheld the supremacy of the five Confucian classics as infallible guides to morality and history. Inevitably, scholars (including Zheng's many critics) have 'read' early Chinese society through the filter of Zheng's surviving commentaries, which ultimately superseded earlier divergent scholastic traditions.

Of the nearly sixty texts attributed to Zheng, only four complete works remain: commentaries on each of the three Rites Classics - the *Liji (Book of Rites)*, *Zhouli (Rites of Zhou)* and *Yili (Ceremony and Rites)* - and a subcommentary appended to the Mao *Shijing (Book of Songs)*. Judging from these works, Zheng Xuan's genius lay in his ability to embody a great deal of information in succinct comments; in imitation of the heroic sages of antiquity, Zheng aimed to 'raise a single principle that applies to a myriad cases… so as to minimize [the need for independent] thought'. Thus, Zheng's commentaries frequently include general statements recapitulating the main principles that underpin the entire Confucian tradition, which he tied to philological explanations based on archaic or regional pronunciations (complementing the *Shuowen* dictionary's emphasis on the visual components of Chinese characters). Zheng's commentaries also relate material from different classical authorities, so that puzzling passages may be solved by 'internal evidence' provided by the sages; in addition, Han customs, popular expressions and even artefacts are described whenever they shed light on ancient practices. As a result, Zheng's annotations go far beyond the standard pre-Han commentarial format (which usually presented glosses in the form '*X* means *Y*'), without lapsing into the prolixity of many Han scholarly explications.

Collected fragments of early polemical texts offer tantalizing clues about three famous disputes (with Xu Shen, He Xiu and Wang Su) involving Zheng and his disciples. Often misinterpreted, these disputes did not centre on either the authenticity of the apocryphal (*wei*) texts or allegiances to New Script/Old Script (*jinwen/guwen*) 'schools'. As Zheng cited the apocrypha only when other early sources corroborate their accounts, his reliance on the apocrypha was quite conservative by contemporary standards. Further, after twenty years devoted to the study of both Old and New Script texts, Zheng was determined to find common ground in competing theories about the classics. Basically, Zheng deplored the literalist approach of Xu and He, which would require the rejection of parts of the sacred canon. Zheng further objected to He's elevation of the *Gongyangzhuan (Gongyang Commentary)* to the *Chunqiu (Spring and Autumn Annals)* to virtual canonical status (see Chinese Classics); in Zheng's view, none of the three major commentaries to the *Chunqiu* could rival the 'subtle message' (*weiyan*) of Confucius' own masterwork. A third dispute, continued over centuries by the disciples of Zheng and Wang Su, was apparently prompted by Zheng's failure to preserve *jiafa* (discrete scholastic lines) and by Wang's pronounced royalist sentiments. The very virulence of the attacks launched against Zheng only attests to his seminal influence on classical studies.

Although Zheng established the basic conventions for critical scholarship in China, modern scholarship discounts his famous 'proofs' for the single, divine origin for the classics and their apocrypha, and also questions Zheng's dating of important works, including the *Zhouli (Rites of Zhou)*. Such empirical errors clearly stem from Zheng's fervent desire to reconcile contradictory passages. Zheng's attempts to shore up the political system of late Han dynasty certainly coloured his commentaries; hence his continual reiteration of ritual's function in preserving the *status quo* and his rejection of the usual ethical dimensions of key status terms (such as 'ruler' and 'gentleman'). However, researchers readily acknowledge Zheng's remarkable erudition with regard to scribal errors, loan characters, dialect pronunciations, script variants, authorial mood and literary context. The ease with which Zheng moves between different disciplines and interpretations confirms his place as a *tongru* (polymath), though it surely also signals a loss of faith in the earlier formulations of Confucian orthodoxy.

See also: Chinese Classics; Chinese philosophy; Confucius

MICHAEL NYLAN

List of works

Zheng Xuan (AD 127-200) *Bowujingyiyi (Contra Variant Interpretations of the Five Classics)*, in *Wenjingtang*

congshu, compiled by Sun Fengyi in *Baibucongshujicheng*, Taibei: Yiwen yinshuguan, 1968.(The most important polemic on the classics by Zheng, now in fragments.)

References and further reading

Chang, S. (1984) *Zheng Xuan congshu (Collected Writings on Zheng's Scholarship)*, Jinan: Qilushushe.(Covers all aspects of Zheng's work, with particular emphasis on his distinctive style in writing commentaries.)

Kramers, R.P. (1949) *K'ung tzu chia yü: The School Sayings of Confucius*, Leiden: Brill, 77-98.(Outlines the major controversy between Zheng Xuan and Wang Su.)

Ozawa, B. (1970) *Kandai Ekigaku no kenkyū(Research on Yijing Studies in the Han)*, Tokyo: Meitoku Publishers. (Limits his discussion to Zheng's theories on the *Yijing (Book of Changes)*.)

Wang, L. (1983) *Zheng Kangcheng nianpu (Chronology of Zheng Xuan)*, Jinan: Qilushushe.(The standard chronological biography (*nianpu*) of Zheng.)

Zhi

In classical Chinese philosophy, zhi*, conventionally translated as 'knowing', is not so much a knowing 'what', which provides some understanding of the natural world, as it is a knowing 'how' to be adept in relationships, and 'how', in optimizing the possibilities that these relations provide, to develop trust in their viability. The cluster of terms that define knowing are thus programmatic and exhortative, encouraging the quality of the roles and associations that define us. Propositions may be true, but it is more important that husbands and friends be so.*

If the major concern in the Chinese tradition is for *how* rather than *what* - finding a way (*dao*) rather than seeking the Truth - this can only mean that most of the theoretical and practical correlates of the search for truth will themselves be absent, with the consequence that the shape of the Chinese cultural sensibility may be expected to be radically distinct from that of Western culture. Rather than a vocabulary of truth and falsity, right and wrong, good and evil - terms that speak to the 'whatness' of things - we find pervasively the language of harmony and disorder, genuineness and hypocrisy, trust and dissimulation, adeptness and ineptness; terms which reflect the priority of the continuity that obtains among things, and the quality of that continuity. That is, 'how well' do things hang together?

In contrast to Western philosophy that began from the decontextualizing metaphysical sensibilities of the classical Greeks, the Chinese tradition is resolutely historicist. Reason is a series of historical instances of reasonableness; culture is a specific historical pattern of human flourishing; logic is the internal coherence of this particular human narrative; knowledge, always local and site-specific, is knowing how to make one's way smoothly and without obstruction. In fact, the performative and perlocutionary entailments of 'knowing' often make 'realizing', in the sense of 'making real', a better translation of this term. Truth is proximate: it is the capacity to foster productive relationships that begin with the maintenance of one's own integrity and extend to the enhancement of one's own natural, social and cultural contexts.

Truth-seekers want finally to get to the bottom line, to establish facts, principles and theories which characterize the way things are. Truth-seekers see a world divided into the way things appear and the way they truly are, and continually try to come up with propositions - true sentences - that describe the reality of things in an accurate manner.

Way-seekers have dramatically different interests. They accept that they live in a human world and seek those forms of action which promote harmonious social existence. For the Way-seekers, truth is most importantly a quality of persons rather than of propositions. Truth as 'way' refers to the genuineness and integrity of a fully functioning person - a *true* human being. To 'know' the way is to walk it, and to become wise in doing so.

See also: Chinese philosophy; Dao; Daoist philosophy; Knowledge, concept of; Truth, correspondence theory of

DAVID L. HALL
ROGER T. AMES

References and further reading

Graham, A.C. (1989) *Disputers of the Tao: Philosophical Argument in Ancient China*, LaSalle, IL: Open Court. (A survey of ancient Chinese thought.)

Hall, D.L. and Ames, R.T. (1995) *Anticipating China: Thinking Through the Narratives of Chinese and Western Culture*, Albany, NY: State University of New York Press. (A comparative study of the uncommon assumptions that ground the Chinese and Western traditions.)

Zhi Dun (AD 314-66)

Buddhist monk and specialist on the Zhuangzi, *Zhi Dun was active in the* xuanxue *or 'learning of the mysterious' salons of the Eastern Jin regime in southeastern China. For him, the sage described by the* Zhuangzi *and Buddhist texts alike was a great man who knows the ways of heaven triumphantly and responds to beings in perfect freedom. Zhi Dun was also known for his interpretation of emptiness and his expansion of the concept of li ('order' or 'pattern') into an underlying metaphysical principle.*

Zhi Dun is the primary representative of a new social development in Chinese history, when numbers of men sought purely spiritual learning in monasteries apart from the secular realm, and when the religion became widely accepted at the highest levels of Chinese society, among both men and women. He was ordained in AD 338 after studying the Perfection of Wisdom (*prajñā pāramitā*) so popular at the time, and died after spending several years teaching at the imperial court. At some point he wrote a commentary on the 'Free and Easy Wandering' chapter of the *Zhuangzi*, now lost. A homely man known for his waspish tongue and trenchant characterizations of people, he was friendly with some of the most prominent men of his day, including the Daoists Wang Xizhi (a famous calligrapher) and Sun Chuo. Xi Chao, a layman whose summary of Buddhist doctrine survives, was his student. Zhi Dun was a devotee of Amitābha, the Buddha of the Pure Land, and wrote a commentary on a scripture devoted to meditation on breathing.

The most significant element in Zhi Dun's thought was his understanding of the sage or perfect man, which he described on the basis of both the *Zhuangzi* and Buddhist texts (see Zhuangzi). He disagreed with the contemporary interpretation of the *Zhuangzi* which held that everyone should follow one's own nature and be happy, which was too amoral and deterministic for his Buddhist inclinations. Instead, the sage was a supramundane embodiment of truth, able to respond to the myriad demands of sentient beings without moving himself. Just as the myriad sounds of the world would cause a bell to reverberate, so would the sage resonate according to the needs of beings, yet the various words and teachings spoken by the sage would not be identical to his own inner wisdom. Implicit in this concept of the sage is the sense that there is an absolute (*tathatā*, but often rendered as *li*) underlying all things (see Li). Zhi Dun was one of the earliest figures to explain Buddhism in terms of being and nonbeing (*you* and *wu*) (see You- wu).

Zhi Dun's understanding of emptiness is described as 'form as such'; known only through very sketchy fragments, it apparently refers to the existence of form through a plethora of co-dependent causes. He denied that there was any permanent substrate underlying form, and he realized that form was both equivalent to and different from emptiness. (Here form represents *se*, or *rūpa* in Sanskrit, the first of the five *skandhas* or constituents of the human being; it thus stands as part for the whole of conditioned phenomena.) Based on the scanty evidence extant, Zhi Dun's theory of emptiness seems to contradict his redefinition of *li* as a metaphysical substrate underlying phenomena.

See also: Buddhist philosophy, Chinese; Buddhist concept of Emptiness; Li; Zhuangzi

JOHN R. McRAE

References and further reading

Zürcher, E. (1959) *The Buddhist Conquest of China: The Spread and Adaptation of Buddhism in Early Medieval China*, Leiden: Brill.(The best treatment of the earliest phase of Chinese Buddhism in any language.)

Zhiyi (538-97)

The Chinese Buddhist monk Zhiyi is revered as the chief architect of the Tiantai school of Buddhism, one of the most distinctive and influential systems of Mahāyāna Buddhist thought and practice to take shape on East Asian soil. His systematization of the Tientai teachings marked the emergence of the first major indigenous articulation of Buddhist thought and practice in China. This is considered an important watershed in the development of the East Asian Buddhist tradition, for it effectively brought to a close some five centuries of Chinese dependence on Indian Buddhist traditions of exegesis and opened the way to creation of the distinctive forms of scriptural hermeneutics and motifs of religious life that we regard as representative of East Asian Buddhism today.

Historically, Zhiyi's religious career spans the key transition between the last years of the strife-torn North-South Dynasties Period (317-581) and the lengthy era of political unification ushered in by the Sui (589-618) and Tang dynasties (620-906). Zhiyi himself is often credited with producing a synthesis of different regional currents of medieval Buddhism that is the religious analogue to the reintegration achieved by the Tang ruling house in the social and cultural spheres.

Born of a well-to-do family distantly connected with the Liang ruling house, Zhiyi's life was abruptly shattered at the age of 17 with the death of his parents and the sack of his home city, Jiangling, at the hands of the Western Wei army. Deeply shaken, he resolved shortly thereafter to become a Buddhist monk. Over the next few years he wandered through Hubei and Henan in search of a suitable teacher, eventually settling down with Nanyue Huisi (515-77), the spirited master of meditation who would later be listed as Zhiyi's predecessor in the Tiantai patriarchal lineage. Under Huisi's guidance, Zhiyi achieved his first major spiritual awakening and was exposed to the rudimentary system of doctrinal classification around which he would fashion his own mature scheme of Tiantai thought and practice.

In 568, at his teacher's urging, Zhiyi proceeded to the southern capital of Jinling to begin his own teaching career. He was received with enthusiasm and over the next seven or eight years acquired a large following among the aristocracy and Buddhist clergy of the capital, including members of the Chen ruling house. However, despite his growing popularity, Zhiyi became increasingly troubled by the limitations that such a bustling environment placed on his ability to train disciples and pursue his own religious practice. This led him in 575 to leave the capital and retire with a small group of students to the isolation of Mount Tiantai in central Zhejiang province. The decade of quiet religious pursuit that followed is regarded as a turning point in Zhiyi's spiritual and intellectual development, marked above all by the maturation of the distinctive system of doctrine and practice that we identify with Tiantai.

Between 585 and his death in 597, Zhiyi again threw himself into the bustling life of court and capital. During this period he enjoyed not only the support of the last Chen ruler but, when the Chen fell to the Sui in 581, also gained the enthusiastic patronage of prince Yang Guang, the man who would ultimately succeed the first Sui emperor, Wen (or Yang Jian), as the ill-fated Emperor Yang. Through the influence of Yang Guang and the Sui court, Zhiyi built Yuquan Monastery in Hubei and laid plans for the expansion of monastic facilities on Mount Tiantai (the construction of the famous Guoqing Monastery, realized several years after Zhiyi's death). With the death of Yang Guang and the collapse of the Sui in 617, Tiantai access to imperial patronage was severed. Nevertheless, over the two centuries that followed the Guoqing and Yuquan communities continued to function as vibrant centres of Tiantai learning. Their traditions influenced not only the formation of Tang Buddhism at large but also found their way to Korea and Japan, giving rise to the powerful Chôntae and Tendai orders in those two countries (see Buddhist philosophy, Japanese; Buddhist philosophy, Korean).

Despite the significance that we attribute to Zhiyi as *de facto* founder of the Tiantai tradition, it would be misleading to say that he intentionally set out to create a new Buddhist school. Like Confucius many centuries before him, Zhiyi thought of himself more as a reclaimer of the old than a creator of the new. Throughout Zhiyi's writings one finds a deep sense of anxiety over the petty rivalries and clerical malaise that he perceived among his Buddhist compatriots. As one who believed profoundly in the inherent unity of the Buddha's enlightened vision and received word, his synthesis was nothing short of an effort to rescue the 'true Dharma' or 'true intent of the Buddha' from the fragmentation and moral ossification that he saw around him. Against this backdrop, Zhiyi

develops a distinctive vision of the Buddha as master of both the transcendent and mundane worlds, whose special genius lies not so much in his grasp of the ineffable reality of enlightenment as in the compassion and pedagogical skill to reach out and guide beings effectively to this end. Zhiyi defines the Tiantai religious path as a twofold discipline involving the seamless integration of meditative experience (*guan*) with comprehensive study of received scriptural tradition (*jiao*). To serve these ends, he developed detailed schemes for meditative practice as well as classification of scripture and its contents (see Buddhist philosophy, Chinese).

Tiantai tradition credits Zhiyi with a prodigious literary output - over fifty works in nearly two hundred fascicles, according to the most fanciful count. However, critical research has significantly reduced this number. Of the works that we may safely connect with Zhiyi, only a handful were written by the man himself; the majority were taken down and edited from his lectures by various disciples, notably his attendant Guanding (561-632). Three treatises in particular have been singled out by later tradition as offering the definitive statement of Tiantai teaching: the *Fahua xuanyi (Profound Meaning of the Lotus Sutra)*, *Fahua wenju (Words and Phrases of the Lotus Sutra)*, and *Mohe zhiguan (Great Calming and Contemplation)*. The first two are synthetic and interlinear commentaries to the *Lotus Sutra*, the scripture hailed by Tiantai tradition as the culmination of the Buddha's career, in which the Buddha professed with the greatest clarity and finality the harmonious unity of his *dharma* (see Buddha). The *Mohe zhiguan* is a work of meditation dedicated to the 'perfect and sudden' contemplation. Although no explicit connection with the *Lotus Sutra* is developed in this text, Tiantai tradition regards it as the practical counterpart to the sublime doctrinal vision spelled out in the two *Lotus* commentaries. All three treatises bear the mark of Guanding's editorial hand. Among contemporary scholars there has developed considerable debate as to how much of their final form is the work of Zhiyi or of Guanding.

See also: Buddhist philosophy, Chinese; Buddhist philosophy, Japanese; Buddhist philosophy, Korean

DANIEL B. STEVENSON

List of works

Zhiyi (538-97) *Fahua xuanyi (Profound Meaning of the Lotus Sutra)*, ed. Takakusu Junjirō and Watanabe Kaigyoku, Taishō daizōkyō 1716, vol. 33, Tokyo: Taishō issaikyō kankōkai, 1924-35.(One of Zhiyi's two commentaries on the *Lotus Sutra*.)

Zhiyi (538-97) *Fahua wenju (Words and Phrases of the Lotus Sutra)*, ed. Takakusu Junjirō and Watanabe Kaigyoku, Taishō daizōkyō 1718, vol. 34, Tokyo: Taishō issaikyō kankōkai, 1924-35.(One of Zhiyi's two commentaries on the *Lotus Sutra*.)

Zhiyi (538-97) *Mohe zhiguan (Great Calming and Contemplation)*, ed. Takakusu Junjirō and Watanabe Kaigyoku, Taishō daizōkyō 1716, vol. 33, Tokyo: Taishō issaikyō kankōkai, 1924-35; ed. and trans. N. Donner and D. Stevenson, *The Great Calming and Contemplation: A Study and Annotated Translation of the First Chapter of Chi-i's Mo-ho chih-kuan*, Kuroda Institute Classics in East Asian Buddhism, Honolulu, HI: University of Hawaii Press, 1993.(A work of meditation dedicated to 'perfect and sudden' contemplation.)

References and further reading

Chappell, D. (ed.) (1983) *T'ien-t'ai Buddhism: An Outline of the Fourfold Teaching*, Tokyo: Daiichi-shobō.

Hurvitz, L. (1962) 'Chih-i (538-597): An Introduction to the Life and Ideas of a Chinese Monk', *Mélanges Chinoises et Bouddhiques* 12: 1-372.

Swanson, P. (1989) *Foundations of T'ien-t'ai Philosophy*, Berkeley, CA: Asian Humanities Press.

Weinstein, S. (1973) 'Imperial Patronage in the Formation of T'ang Buddhism', in A. Wright and D. Twitchett (eds) *Perspectives on the T'ang*, New Haven, CN: Yale University Press.

Zhongyong

The Zhongyong (Doctrine of the Mean) has traditionally been ascribed to Zisi, the grandson of Confucius and the indirect teacher of Mencius. Although this ascription has been challenged by modern critical scholarship since the turn of the twentieth century, recent archaeological finds indicate that the traditional view is not without textual base. If the Zhongyong actually predated the Mengzi, it seems that a significant portion of the Liji (Book of Rites), of which the Daxue (Great Learning) and Zhongyong are chapters, contains documents of the fifth century BC*. This fact alone merits a fundamental restructuring of classical Confucian chronology and reinterpretation of the Mencian line of the Confucian tradition.*

The Chinese title of the essay is composed of the characters ‘centrality’ (*zhong*) and ‘normality’ or ‘commonality’ (*yong*). *Zhongyong (Doctrine of the Mean)*, suggesting the basic moral idea of moderation, balance and suitability, conveys a profoundly meaningful notion of humanity as both an anthropological and cosmological ideal. Centrality and commonality in the *Zhongyong* connotes that ordinary human virtues such as filial piety, honesty, reciprocity, and benevolence, if thoroughly comprehended and fully embodied in our action, can help us to understand not only ourselves but also nature and Heaven. This faith in the improvability of the human condition through self-effort and in the human capacity to know and realize the truth and reality of Heaven through self-knowledge leads to the theory and practice that philosophy as a way of life addresses the ontological issues of who we are, what we know, how we should act and what we can hope (see Self-cultivation in Chinese philosophy).

Intent on relating what is most essential in humanity to the underlying reality of the cosmos, the *Zhongyong* is organized around three integrated visions: the profound person, the fiduciary community and moral metaphysics. Since human nature is endowed by the decree of Heaven, there is a wellspring of creativity inherent in the structure of human existence here and now. If we allow that wellspring to flow naturally, we will find the most authentic way of learning to be human. Indeed, the ultimate concern of education is none other than the cultivation of the human way. Implicit in this seemingly optimistic vision of self-realization are four inseparable dimensions of human flourishing: self, community, nature and Heaven. Learning to be human entails the fruitful interplay between self and community (family, society, nation and the world), the dynamic equilibrium between the human species and nature and the creative fidelity between the human heart-and-mind and the way of Heaven (see Tian; Xin). This requires constant attention to even the minute details in our daily living and ceaseless effort in elevating our levels of consciousness and self-reflexivity.

Profound people optimize their capacity to realize the full potential inherent in human nature. Such a project of self-realization, if shared as a joint venture for mutual exhortation, will form a discourse fellowship which may serve as a standard of inspiration for the creation of a fiduciary community. A community, based upon mutual trust of its members for the sake of completing the human project as co-creator of the cosmic process, is a ‘holy rite’ in which the sanctity of the earth, the sacredness of the body and the beatitude of the family enable a concrete living person to become not only a moral agent but also an anthropocosmic presence:

> Only those who thoroughly embody sincerity can fully realize their own nature. If they can fully realize their own nature, they can fully realize human nature. If they can fully realize human nature, they can fully realize the nature of all things. If they can fully realize the nature of all things, they can participate in the transforming and nourishing processes of Heaven and Earth. If they can participate in the transforming processes of Heaven and Earth, they can form a trinity with Heaven and Earth.
>
> (*Zhongyong* XXII)

See also: Confucian philosophy, Chinese; Chinese philosophy; Self-cultivation in Chinese theories of; Tian

TU WEI-MING

References and further reading

Chan Wing-tsit (1963) *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press.(A translation and commentary on this text translated as *The Doctrine of the Mean*.)

Schwartz, B.I. (1985) *The World of Thought in Ancient China*, Cambridge, MA: Harvard University Press.(A

discussion of several of the central concepts in the *Zhongyong*.)
Tu Wei-ming (1989) *Centrality and Commonality: An Essay on Confucian Religiousness*, Albany, NY: State University of New York Press.(A monograph dedicated to an analysis and exegesis of the *Zhongyong*, which uses the standard Chan (1963) translation.)

Zhou Dunyi (1017-73)

*Zhou Dunyi was the father of Chinese neo-Confucianism. His oracular presentation of the notions of supreme polarity (*taiji*),* yin *and* yang*, and the five phases to explain the formation of the cosmos and sagehood became enshrined by Zhu Xi as the authoritative neo-Confucian view.*

Born in Daozhou, China, Zhou served with distinction in a series of official posts from about 1040 to 1072 but is famed for his love for nature and his secluded cottage near a waterfall on Lu Mountain. There he wrote two visionary tracts on metaphysics, sagehood and ethics. His contemporaries Shao Yong and Zhang Zai pursued similar themes, but lacked Zhou's clear synthetic vision and subtle articulation. Cheng Hao and Cheng Yi, who went on to develop neo-Confucianism into an intellectual movement, studied under Zhou but later concealed his formative influence.

Zhou's synoptic essay *Taiji tushuo (Explanation of the Diagram of the Supreme Polarity)* presents a dynamic conception of reality. It begins with an account of the formation of the cosmos in levels from elementary forces to complex phenomena, and ends with an account of the sage as a reflection of the cosmos. The text opens with an intriguing dialectical proposition that even now arouses discussion and debate: 'Free from any polarity (*wuji*), and yet the supreme polarity (*taiji*)!' Taking *wuji* as the Daoist 'utmost of non-being', Lu Xiangshan (1139-93) saw this proposition as an illicit attempt to bifurcate basic reality into two distinct categories. His contemporary Zhu Xi countered that *wuji* here does not refer to a positive concept. It simply indicates that *taiji* (supreme polarity) is not a discrete, locatable entity; it is the primal pattern of change, that which describes the intercourse between *yin* and *yang* and thus pervades reality. Next, the *Taiji tushuo* goes on to describe how *yin-yang* intercourse gives rise to the five phases - earth, wood, fire, water, metal - which in turn engender 'heaven, earth and the myriad things', the empirical world (see Yin-yang).

Humanity reflects this fecund cosmos, but most consummately in the person of the sage, 'whose virtue matches that of heaven and earth, whose brilliance matches that of the sun and moon, whose order matches that of the four seasons, and whose fortunes match those of the gods'. In the *Tongshu (Penetrating the Book of Changes)*, Zhou describes sagehood, self-cultivation and related concepts. The distinguishing virtue of sagehood is cheng (sincerity, authenticity). The sage authentically manifests the creative virtues of the cosmos. Through this sincerity, he abides by the mean and acts according to the cardinal virtues adumbrated by Confucius. The sage is thus one who acts in an upright manner yet spontaneously and without deliberation. Those who aspire to this sphere must, above all, cultivate a pure, tranquil mind (see Confucian philosophy, Chinese).

Dedicated to this view, Zhou impressed others with his distinctive manner. The poet Huang Tingjian (1045-1105) lauded Zhou as 'a man of lofty character, fresh and vigorous as a breeze in the sunlight…. Modest in striving for renown yet ardent in seeking self-realization, he is indifferent to worldly success yet steadfast in friendship' (*Songshi* (History of the Song Dynasty) 427: 3).

See also: Neo-Confucian philosophy; Self-cultivation in Chinese philosophy; Zhu Xi

KIRILL OLE THOMPSON

List of works

Zhou Dunyi (1017-73) *Taiji tushuo (Diagram Explaining the Supreme Ultimate)*, trans. Wing-tsit Chan, *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press, 1963, 463-80.(This essay is a rather oracular presentation of Zhou's metaphysics and ethics; thus, the supporting notes and discussion are required reading.)

Zhou Dunyi (1017-73) *Tongshu (Penetrating the Book of Changes)*, Princeton, NJ: Princeton University Press, 1963, 463-80.(This essay is a rather oracular presentation of Zhou's metaphysics and ethics; thus, the supporting notes and discussions are required reading.)

References and further reading

Chan Wing-tsit (ed. and trans.) (1966) *Reflections on Things at Hand: the Neo-Confucian anthology compiled by Chu Hsi and Lu Tsu-ch'ien*, New York: Columbia University Press.(Presents passages from Zhou's works with

commentaries by traditional scholars; abstruse and technical yet interesting.)

Fung Yu-lan (1953) ‘Chou Tun-i’, trans. D. Bodde in *A History of Chinese Philosophy*, Princeton, NJ: Princeton University Press, 434-50.(Systematic overview of Zhou’s metaphysics and ethics; somewhat technical, but generally understandable.)

Needham, J. (1956) ‘The "Supreme Pole"’, *Science and Civilisation in China*, vol. 2, *History of Scientific Thought*, Cambridge: Cambridge University Press, 460-72. (Influential attempt to understand Zhou’s thought according to the philosophy of organism; abstruse yet lucid and fascinating.)

Huang, S. (1974) ‘The Concept of *T’ai-chi* (Supreme Ultimate) in Sung Neo-Confucianism’, *Journal of Chinese Philosophy* 1 (3): 275-94.(Review of various Song Chinese understandings of Zhou’s guiding concept of *taiji*; technical.)

Zhu Xi (1130-1200)

The Chinese neo-Confucian philosopher Zhu Xi was a consummate scholar and classicist as well as a superb critical and synthetic thinker. He fused the ideas of the seminal eleventh-century thinkers Shao Yong, Zhou Tunyi, Zhang Zai, Cheng Hao and Cheng Yi into a grand philosophical synthesis. In addition, by effectively editing and annotating the essential classical Confucian texts - the Analects of Confucius - the Mengzi of Mencius, the Daxue (Great Learning) and the Zhongyong (Doctrine of the Mean) - as the Four Books, Zhu worked out a lasting renewal of the Confucian project.

Zhu Xi was born in Youqi in Fujian province, China. He was a precocious child who at five asked what lay beyond heaven and at eight grasped the significance of the *Xiaojing (Book of Filial Piety)*. Reared in the company of scholars, Zhu obtained the *jinshi* degree (presented scholar degree, the highest level qualifying degree in the examination system) at nineteen. He nurtured an eclectic interest in Daoism and Chan Buddhism until becoming a student of Li Tong (1093-1163) in 1160. Li persuaded him of the superiority of Confucianism, to which he devoted himself for the rest of his life. He held various prefectural administrative posts but, disapproving of court policy, preferred temple guardianships, sinecures that gave him leisure to read, write and teach. He thus became a highly productive scholar who made lasting contributions to classical studies, historiography and literary criticism as well as to philosophy. He also was a master prose stylist and poet.

As a renowned master, Zhu taught the classics and neo-Confucianism. His teachings are preserved in the *Zhuzi yulei (Classified Dialogues of Master Zhu)*. An assiduous teacher and scholar, Zhu continued to work right up to his death. He set rigorous new standards for textual criticism and philosophic thought; his idealistic spirit and productive life remain an inspiration to this day in East Asia.

Zhu Xi created an impressive metaphysical synthesis that has been compared broadly to the systems of such cardinal Western thinkers as Plato, Aristotle, Thomas Aquinas and A.N. Whitehead. Zhu developed Zhou Dunyi's dynamic conception of reality, as depicted in the *Taiji tu (Diagram of the Supreme Polarity)* (see Zhou Dunyi), in order to conceive the Cheng brothers' concept of *li* (pattern) and Zhang Zai's notion of *qi* (cosmic vapour) as organically integrated (see Cheng Hao; Cheng Yi; Zhang Zai). In Zhou's treatise *Taiji tushuo (Diagram Explaining the Supreme Ultimate)*, Zhu found a viable account of the formation of the cosmos in levels: from the original unformed *qi* and the supreme polarity (*taiji*), to *yin, yang* and the five phases - earth, wood, fire, water and metal - down to 'heaven, earth and the myriad things', the empirical world. Zhu blended this conception with ideas from the *Yijing (Book of Changes)* and its commentaries (see Yijing) to set forth a comprehensive philosophy of cosmic creativity and to ground the Confucian notions of human nature, self-cultivation and self-realization. Zhu's penchant for thinking in polarities, however, has stirred critics to regard him as a dualist.

On the basis of this complex conception of reality, Zhu set forth a sophisticated theory of human nature that accounts for the possibility of evil as well as of sagehood. In this view, whereas people are fundamentally good - that is, sensitive and well-disposed - how one manifests this original nature will be conditioned by one's specific *qi* make-up, one's personality and intelligence, and level of attainment. The differences in people's disposition, character and aptitude for self-realization are regarded as due to variations in their *qi*-endowments.

Not registering the complexity of human nature and the wide variation of individual differences, the preceding generations of neo-Confucian scholars had advocated straightforward methods of self-cultivation to purify the mind and elicit the original nature; they understood this to constitute self-realization (see Xing; Self-cultivation in Chinese philosophy). Zhu's teacher Li Tong had advocated a form of meditation called 'quiet-sitting', the efficacy of which Zhu doubted from the outset. Later Zhu held discussions with Zhang Shi, a disciple of Hu Hong, who advocated introspection in action. At first, Zhu accepted this approach, but he soon found that introspection in action is not viable, for such introspection cannot inform action in the desired way; it in fact impedes effective action.

Zhu Xi's solution was a two-pronged approach to cultivation: to nurture a sense of reverence (*jing*) while investigating things to explore their patterns (*li*). Reverence purifies the mind, attunes one to the promptings of the original nature and impels one to act with appropriateness. At the same time, by grasping the patterns which

constitute the world and upright interpersonal relations, one learns more about appropriate conduct. The mind which is imbued with reverence and comprehends these patterns will develop into a good will.

In later years, Zhu regarded this conception of cultivation as insufficiently motivational. He then argued that, like Confucius, one must at the outset establish the resolve to realize the Confucian ideals (see Confucian philosophy, Chinese). Zhu believed that such resolve lends an overriding purpose to one's cultivation effort and thus preserves one's motivation. This was a theme which Wang Yangming developed effectively some three centuries later.

In *Renshuo (A Treatise on Humanity)*, Zhu articulates the classical Confucian ideal of humanity (*ren*) in cosmic and human perspective. At the same time, he effectively criticizes competing accounts of *ren* on logical and semantic grounds. Zhu associates *ren* with cosmic creativity; at root, *ren* is the impulse of 'heaven and earth' to produce things. It is manifested in the cycle of seasons and the fecundity of nature. This impulse is instilled in all creatures, but in man it is sublimated into the virtue of *ren*, which when fully realized involves being diligent and responsible to others. Zhu similarly correlated the four stages of production in nature - origination, growth, flourishing and firmness - mentioned in the *Yijing* with the four cardinal virtues enunciated by Confucius - humanity, appropriateness, ritual action and wisdom - in order to portray the realized person as a vital participant in cosmic creativity as well as a positive catalyst for the well-being and self-realization of others. On this basis, Zhu went on to formulate the authoritative definition of *ren* for the tradition as 'the character of mind' and 'the pattern of love'. *Ren* thus grounds the fundamental disposition of mind as commiserative and describes the core of moral self-realization as love for others, appropriately manifested.

See also: Cheng Hao; Cheng Yi; Confucian philosophy, Chinese; Li; Neo-Confucian philosophy; Qi; Self-cultivation in Chinese philosophy; Zhang Zai; Zhou Dunyi

KIRILL OLE THOMPSON

List of works

Zhu Xi (1130-1200) *Zhuzi yulei (Classified Dialogues of Master Zhu)*, trans. J.P. Bruce, *The Philosophy of Human Nature*, London, 1922.(Compendium of Zhu's moral psychology drawn from the *Zhuzi quanshu* ('Complete' works of Master Zhu), abstruse. Translated selections can be found in Chan 1963, 1966; Gardner 1986, 1990; Wittenborn 1991.)

References and further reading

Chan Wing-tsit (1963) 'The Great Synthesis in Chu Hsi', in *A Source Book in Chinese Philosophy*, Princeton, NJ: Princeton University Press, 605-53.(Detailed overview of Zhu's thought drawn primarily from *Zhuzi quanshu*; clear and thoroughly annotated.)

Chan Wing-tsit (1966) *Reflections on Things at Hand: The Neo-Confucian Anthology Compiled by Chu Hsi and Lu Tsu-ch'ien*, New York: Columbia University Press.(Zhu's comprehensive compendium of seminal neo-Confucian pronouncements; clear and well-annotated.)

Chan Wing-tsit (ed.) (1986) *Chu Hsi and Neo-Confucianism*, Honolulu, HI: University of Hawaii Press. (Anthology of essays by prominent Zhu Xi scholars; contains Zhu's biography and authoritative articles on all aspects of Zhu's thought and scholarship.)

Chan Wing-tsit (1987) *Chu Hsi: Life and Thought*, Hong Kong: Hong Kong University Press.(General essays; clear and accessible.)

Chan Wing-tsit (1989) *Chu Hsi: New Studies*, Honolulu, HI: University of Hawaii Press.(Detailed studies of arcane issues in Zhu Xi scholarship; for the specialist.)

Fung Yu-lan (1953) 'Chu Hsi', trans. D. Bodde in *A History of Chinese Philosophy*, Princeton, NJ: Princeton University Press, 533-71.(Pioneering English presentation of Zhu's thought; technical but clearly presented.)

Gardner, D. (1986) *Chu Hsi and the Ta-hsueh: Neo-Confucian Reflection on the Confucian Canon*, Cambridge, MA: Harvard University Press.(Contains Zhu's commentary on the *Daxue (Great Learning)*, a classical cultivation text; commentary with supporting essays.)

Gardner, D. (1990) *Learning to Be a Sage: Selections from the Conversations of Master Chu, Arranged Topically*, Berkeley, CA: University of California Press.(Zhu's discussions on learning as a method of self-cultivation; clear and accessible.)

Needham, J. (1956a) 'The Neo-Confucians', in *Science and Civilisation in China*, vol. 2, *History of Scientific Thought*, Cambridge: Cambridge University Press, 455-84.(Innovative organismic account of Zhu's thought; lucid and fascinating.)

Needham, J. (1956b) ‘Chu Hsi, Leibniz, and the Philosophy of Organism’, in *Science and Civilisation in China*, vol. 2, *History of Scientific Thought*, Cambridge: Cambridge University Press, 455-84.(Innovative organismic account of Zhu’s thought; lucid and fascinating.)

Schirokauer, C. (1962) ‘Chu Hsi’s Political Career: A Study in Ambivalence’, in A. Wright and D. Twitchett (eds) *Confucian Personalities*, Stanford, CA: Stanford University Press, 162-88.(Detailed account of Zhu’s official career; quite readable and engaging.)

Thompson, K. (1988) ‘*Li* and *Yi* as Immanent: Chu Hsi’s Thought in Practical Perspective’, *Philosophy East and West* 38 (1): 30-46.(Reinterpretation of Zhu’s ontology and ethical theory; lucid and informative.)

Wittenborn, A. (1991) *Further Reflections at Hand: A Reader: Chu Hsi*, New York: University Press of America. (A comprehensive compendium of Zhu’s philosophic pronouncements; clear translation with detailed commentary.)

Zhuangzi

The Zhuangzi is a Daoist text usually associated with 'Master Zhuang' (fourth century BC), also known as Zhuang Zhou. Scholarly consensus regards the thirty-three chapters of this text to be composite, containing passages that offer different and sometimes even contradictory interpretations of basic Daoist tenets. The opening seven 'inner chapters' are traditionally thought to be from the hand of Master Zhuang himself, while the remaining 'outer' and 'miscellaneous' chapters are taken to be later elaborations and commentary by members of what retrospectively can be called a Master Zhuang school, or better, lineage. As a philosophical text, the Zhuangzi is for the most part addressed to the project of personal realization, and only derivatively concerned about social and political order. As one of the finest pieces of literature in the classical Chinese corpus, the Zhuangzi is itself an object lesson in marshalling every trope and literary device available to provide rhetorically charged flashes of insight into the most creative way to live one's life in the world.

The consummate human being in *Zhuangzi* is called the 'Authentic Person' (*zhenren*), and consummate 'ethical' concerns have to do with the contributions one is able to make to the *ethos* or total character of one's world. Thus, 'ethics' in a Daoist sense would involve not simply moral judgments, but the total character of the person as it is constructed in their social and natural environments (see Daoist philosophy). This total character is the changing lineaments of habit and disposition that include not only what might be deemed 'good' or 'moral' but the full complement of the qualities which constitute the personal, social and natural fabric of one's existence. Having said this, it must be clear that any 'ethical' judgments in the narrow sense are going to be derived from aesthetic sensibilities - the intensity, integrity and appropriateness that one detail has for its environing elements as interpreted from some particular perspective. Thus, no clear direction into ethics, social, political, and environmental philosophy, exists in this tradition.

To the extent that Daoism is normative and prescriptive (and it certainly is), it is so not by articulating rules to follow or asserting the existence of some underlying moral principles, but by describing the conduct of an achieved human being - the Authentic Person - as a recommended object of emulation. The model for this human ideal, in turn, is the orderly, elegant and harmonious processes of nature. Throughout the philosophical Daoist corpus, there is a 'grand' analogy established in the shared vocabulary used to describe the conduct of the achieved human being on the one hand, and the productive harmony achieved in the mutual accommodations of natural phenomena on the other.

Among the *Zhuangzi*'s central concerns is human creativity. To be fully integrative, inchoate individuals must overcome any sense of discreteness and disjunction with their environments to contribute personally to the emerging pattern and regularity of the world around them. This creativity can be compromised, however, where one attempts to express one's unique particularity in a 'dis-integrative' way that fails to accommodate the mutuality and interdependence of other things. This diminution in creative possibilities can be brought about either by interpreting one's environs reductionistically through one's own fixed conceptual structures and values, thereby impoverishing context in service to oneself, or by allowing oneself to be shaped wholly by context without contributing one's own uniqueness, thereby abnegating oneself in service to context.

The choice of 'authentic' rather than 'true' or 'real' as a translation for '*zhen*' in *zhenren* is calculated. With the same root as 'author', 'authentic' captures the primacy given to the creative contribution of the particular person in authoring the world. It further registers this contribution as what is most fundamentally 'real' and 'true'. It is because of the primacy of the 'authorship' of the 'authentic person' in creating human order that the *Zhuangzi* insists: 'there must be the Authentic Person before there can be authentic knowledge'. To 'know' a world is to 'realize' it in the sense of making it real. This claim made at the beginning of *Zhuangzi* 6 means that 'knowing' is not simply a cognitive exercise, the grasping of some pre-existing reality. Rather, 'knowing' is also performative: one must participate actively in the 'realization' of one's world through one's own 'self-disclosure' (*ziran*). 'Knowing' for the Daoist, in addition to being cognitive, is profoundly experiential and collaborative.

In translating (*ziran*) as 'self-disclosure', we must bear in mind that 'self' is always 'in context'. For the Daoist, like the classical Chinese tradition broadly, there is a priority of situation over agency, and a priority of change over form. The particular self is thus neither discrete nor superordinate, but is always continuous with its ever

changing world. It is sponsored by, and ultimately reflects in itself, the full consequence of existence. 'Self' is an individuated focus (*de*) in the ongoing field of the social, natural and cultural processes (*dao*) that articulate our world.

The Authentic Person's activity, characterized as *wuwei* (non-assertive action) is defined by flexibility, efficacy and non-contentiousness. One collaborates with one's social and natural environments, serving as frictionless ground for their self-disclosure, and they reciprocally for one's own. Because one develops a sense of identity and continuity with the process as a whole, one is calm and imperturbable. This sense of continuity enables one to exist beyond the plethora of disintegrative dualisms - self and other, creator and creature, life and death, reality and appearance - and to overcome death not by escaping to some more 'real' existence, but by coming to realize and to celebrate the mutually entailing identity of oneself and one's world.

Early Daoism has had an incalculable influence on the development of Chinese philosophy and culture. Second in influence only to the Confucian school, the classical Daoist philosophers in many ways have been construed as both a critique on and a complement to the more conservative, regulatory precepts of their Confucian rivals. When Buddhism entered China in the second century AD, Buddhist adepts appropriated many of the ideas and vocabularies of indigenous Daoism as a sympathetic language through which to express a fundamentally foreign system of thought. In so doing, they made Buddhist ideas more palatable to the Chinese audience, but at the same time, they also set Buddhism on a course of irrevocable Sinicization.

See also: Chinese philosophy; Dao; Daodejing; Daoist philosophy; De; Huainanzi; Qi; Yin-yang; You-wu

ROGER T. AMES

References and further readings

Hall, D.L. and Ames, R.T. (1997) *Thinking from the Han: Self, Truth, and Transcendence in China and the West*, Albany, NY: State University of New York Press.(An interpretative study of some of the presuppositions underlying the Daoist world view.)

Jullien, F. (1995) *The Propensity of Things: Toward a History of Efficacy in China*, New York: Zone Books.(An examination of the way in which underlying assumptions have influenced the evolution of Chinese culture broadly).

LaFargue, M. (1994) *Tao and Method*, Albany, NY: State University of New York Press.(A primer on how to read the Daoist texts.)

Roth, H.D. (1991) 'Who Compiled the *Chuang Tzu*?', in H. Rosemont (ed.) *Chinese Text and Philosophical Contexts: Essays Dedicated to Angus C. Graham*, La Salle, IL: Open Court.(A textual study of the *Zhuangzi* that continues the earlier work of A.C. Graham in sorting the text into different philosophical strata.)

Wu Kuang-ming (1990) *The Butterfly as Companion: Meditations on the First Three Chapters of the Chuang Tzu*, Albany, NY: State University of New York Press.(A detailed analysis and exegesis of the first three chapters.)

Zhuangzi (*c.*4th century BC?), trans. B. Watson, *The Complete Works of Chuang Tzu*, New York: Columbia University Press, 1968; trans. A.C. Graham, *Chuang-tzu: The Inner Chapters*, London: George Allen & Unwin, 1981.(Watson is the standard translation of the *Zhuangzi* which takes advantage of the Japanese translation by Fukunaga Mitsuji. Graham is a philosophically sophisticated translation of about 85 per cent of the original text.)

Zionism

Zionism, the idea of Jewish nationality in its modern form, emerged towards the end of the nineteenth century, several decades after nationalism had taken hold among most European peoples. The term denotes the ideology as well as the movement(s) whose goal is re-establishment of the Jewish people as a nation in its homeland in Palestine. Unlike many other movements of national liberation, Zionism seeks not the removal of a colonial regime but the return of a people to its land. The most ancient roots of this aspiration were religious, but many forms of modern Zionism shed the messianic and eschatological elements of the two-thousand-year-old hope in favour of more immediate political and social aims that were often philosophically shaped.

1 Moses Hess

The most important philosophical forerunner of Zionism was Moses Hess (1812-75). Against the background of the unification of Italy in 1859, Hess saw the quest of the Jewish people for renewed nationhood as ‘the last national problem’. Philosophically, his vision was shaped by Hegelian philosophy (see Hegelianism), an enthusiastic response to the philosophy of Spinoza, and the Fichtean idea of *Bestimmung*, destiny, an idea that conveyed at once a sense of mission and of objective historical laws setting the parameters of that mission’s fulfilment. The historic mission of the Jewish people was to spread the idea of universal human harmony throughout the world and so to contribute to the global struggle for social and national liberation. But that ancient mission could be achieved only if the people of Israel could re-establish its own state.

Hess’s metaphysical conceptualization of Israel’s mission among the nations did not diminish the concreteness of his attention to the anomalies of diaspora Jewish life. His prescriptions for renewal and normalization in the historic Jewish homeland anticipated many of the ideas proclaimed half a century later by the labour Zionist movement. A friend and co-worker of the young Karl Marx, Hess was a pioneer both of socialism and of Zionism, and particularly of socialist Zionism. His *Rom und Jerusalem (Rome and Jerusalem)* (1862) evoked no positive responses when first published. Only some thirty years later, with the founding of the Zionist organization, was the book retrieved from oblivion and made integral to Zionist ideology. Although unsystematic and often favouring intuition and sentiment over rigorous argument as a guide to its speculation, the work endures as a monument and anchor of philosophical Zionism.

2 Hibbat-Zion

Zionist ideas were spread through eastern Europe by the Hibbat-Zion movement, which drew its inspiration from Jewish tradition. Religious longing for ‘the Land’ joined with disappointment in the fruits of the emancipation promised by Enlightenment rhetoric. These emotions far outweighed any philosophical deliberations in the rise of Zionism as a popular ideology. But Leopold Pinsker and Ahad Ha’am did give a philosophical bent to Hibbat-Zion.

Pinsker (1821-91) began his celebrated pamphlet *Auto-Emancipation* (1882) with an analysis of anti-Semitism. Like Hess, he argued that the Jewish people is a distinct ethnicity that cannot be assimilated - not that Jews will not integrate into the societies in which they live but that in the long run they are not tolerated, not allowed to do so. Hatred towards Jews, ‘Judophobia’, Pinsker argued, was the chief cause of Jewish separateness (see Anti-Semitism), and the perennial isolation of Jews as outsiders can be overcome only when Jews become once again a proper nation with a state of their own, regaining their own homeland and the chance to live in peace and dignity like any other nation. Among his chief goads in writing *Auto-Emancipation*, Pinsker confesses, were Spinoza’s parallel but more negatively couched remarks at the end of the third chapter of the *Tractatus Theologico-politicus*.

Theodor Herzl, the founder of political Zionism, devoted his *Der Judenstaat (The Jewish State)* (1896) to elaborating Pinsker’s thesis that ‘the Jewish problem’ could be solved only if the Jews were no longer a national anomaly. Yet Herzl’s role as a visionary statesman and powerful journalist did not make him a philosopher.

Ahad Ha’am (Asher Ginzberg, 1856-1927) is a thinker of a different stamp. Although no systematic philosopher, he was deeply influenced by nineteenth-century evolutionist ideas, especially those of Herbert Spencer. His outlook never became the main road of Zionist thinking, but it had a powerful impact on Zionist theory. Jewish

particularity, he argued, is grounded in a spontaneous sentiment of national belonging. The sentiment, akin to family ties, needs no theoretical justification or argument. For such feelings are prior to consciousness itself. A nation's 'will to live' is a resultant of every individual's will to live. But the Jewish situation engenders problems of special complexity - most portentously, that of assimilation. A 'spiritual centre' in Palestine, Ahad Ha'am argued, would constitute an effective barrier against assimilation: here Judaism could be preserved, and the identity of the Jews of the diaspora would be assured by the ideas radiating from this centre. Tragically, however, Ahad Ha'am neglected the underlying problem of anti-Semitism and its pressures towards assimilation, isolation and, in the extreme, annihilation of the Jews.

3 Land and language

Jacob Klatzkin (1882-1948), whose intellectual work dealt mainly with medieval Jewish philosophy and Spinoza, also elaborated a philosophical Jewish nationalism. He castigated the tendency of Reform Jews to transform Judaism into a merely spiritual idea. This 'trying to be a Jew without being Jewish' (*Jude-Sein ohne Jüdisch-Sein*) only paves the way to assimilation. But Orthodoxy is equally flawed. Where the reformers seemed to reduce Judaism to a spiritual mission, Orthodoxy reduced Jewishness to mere norms of practice - just another subjective criterion of Jewishness. If Jewishness is to be a durable fact, Klatzkin argued, it cannot be a matter of religion or morality. For to be a Jew no longer entails a particular religious or spiritual commitment. Rather, the subjective historical will to belong to the Jewish people finds its objective expression in the fact of national belonging. The spiritual influence of Judaism on other religions and on Western culture was undeniable, and Klatzkin did not deny the distinctive spirit of Judaism and of Jewish ethics and values. But he bitterly opposed efforts to make these the national paradigms or foci. Jewish nationalism must be rooted in the objective ground of land and language. The very survival of the Jewish people demanded that the Jewish people abandon its inveterate intellectualizing and spiritualizing and resume national life in its own homeland, with its own language.

Klatzkin's arguments were fraught with conceptual shortcomings. His thesis that 'only forms can serve as national criteria', for example, led him to define 'land' and 'language' as national forms, denying them material content! But even such philosophically dubious moves do not diminish the importance of his recognition that the will to settle on one's own land and speak one's own language is vital to Jewish peoplehood. Nor did Klatzkin underrate such external factors as hatred directed against Jews as such, the persecution and discrimination that Ahad Ha'am had underrated. A mere spiritual centre in Palestine would not be enough. Only a national home would resolve the Jewish question.

The inference was clear: the modern definition of a Jew is secular. Only those whose homeland is Palestine and whose language is Hebrew (or who aspire to achieve this) can be Jews. Klatzkin's arbitrary definition of Jewishness in effect excludes all Jews who continue to live outside Israel and do not intend to settle there. It was perhaps his exclusion of the bulk of diaspora Jewry that in the end marginalized Klatzkin as a Zionist thinker and led to his playing only a minor role in the unfolding of Zionist thought.

4 Socialist Zionism

Among the Zionist thinkers who deserve philosophical attention, Nahman Syrkin (1868-1924) developed an early synthesis of socialism and Zionism, and Ber Norochov (1881-1917) went further in the same direction. Borochov grounded his reflections in a distinctive version of the Marxist idea of class struggle. Without a country of their own, he reasoned, Jews remain a powerless minority, confined to 'unproductive' and peripheral pursuits, petty and secondary trades. Barred from heavy industry, 'the axle of the historical wheel', they cannot become true proletarians and so lack a 'strategic base' for carrying forward the class struggle. Jewish 'proletarization' will occur only when the 'inverted pyramid' of Jewish economic life is returned to its broad base; and only in their own country will Jewish workers be able to struggle successfully against the bourgeoisie and win their way to a healthy and independent economic life. Socialism and Zionism are thus necessarily intertwined. Borochov, accordingly, unlike many a Marxist and bourgeois Jewish thinker, condemned assimilation: not only will it never succeed but even its pursuit is a dangerous illusion, since it turns Jews away from the struggle for national emancipation. Efforts to emigrate from Europe to other countries such as those of North and South America, similarly, will only perpetuate the anomaly of the diaspora. What Jews must do is return to their own country, Palestine. Many of Borochov's Marxist notions about Jewish nationhood now require critical re-examination. But his thinking had a powerful impact on labour Zionism and endures in the recognition that labour must be the foundation of the

restored Jewish nation.

Among the non-Marxist socialist Zionists, A.D. Gordon, Berl Katznelson and their disciples were especially important. For them Zionism begins in a voluntary act of the individual, affirming the dignity of physical labour and of ties to the soil. These rather agrarian thinkers aspired to create a new Jew to replace the dispossessed Jews of the diaspora. J.H. Brenner and others, influenced by Nietzsche, similarly denounced what they saw as a diaspora mentality and demanded a radical break with the Jewish spiritual heritage. But here, as with Klatzkin, the exclusionary tendency in the end proved more limiting to the influence of its authors than to the scope and appeal of the Zionist idea.

5 Martin Buber

Finally, Martin Buber played an important role in shaping Zionist consciousness among young Jews in western Europe, especially through his famous *Drei Reden über das Judentum (Three Addresses on Judaism)* (1909-11). In time his Zionist thought was overshadowed by his studies of Hasidism and the Bible and above all by his dialogical philosophy and social thinking. But Buber's Zionism carried forward some of the utopian socialist ideas developed by A.D. Gordon and the anarchist G. Landauer. It grew into what he called Hebrew Humanism, which found the differentia of Zionism from other national movements in its idealism. In keeping with his philosophical vision and with the universalism that he drew from the wellsprings of Jewish moral thought, Buber was among the first and most important Zionist theorists seriously to address the issue of Jewish-Arab relations. His philosophical humanism led him to insist that Jews and Arabs should live together in peace and harmony in a common homeland. But at the time of their enunciation these claims were not taken up by most Zionists as integral to the Zionist programme.

The foregoing examples of philosophical Zionism, exemplified in the writings of its 'founding fathers' are hardly a comprehensive listing. Many other philosophers have engaged in theoretical exploration of the promise and problems of Zionism - among them, J. Kaufmann, J. Klausner, F. Weltsch, M. Brod, N. Rotenstreich, Y. Leibowitz, M.M. Kaplan, A.J. Heschel, A. Neher and E. Levinas. Still other names would need to be mentioned if one were to address the politically divergent views within Zionism, such as Zionist Revisionism. But despite the importance and originality of the living exponents of Zionist thinking, they all belong to a generation for whom Zionism is a *fait accompli*, a living reality, whose theorists, perforce, must now regard not only its 'roots' but - to continue the metaphor - the cultivation of its varied and multi-coloured 'flowers'.

On the theoretico-philosophical plane, a controversy has arisen recently concerning whether contemporary Judaism and the State of Israel belong already to a post-Zionist era, characterized by coexistence between the State of Israel and diaspora Judaism. The majority of the present-day immigrants to Israel (mainly from Russia and Ethiopia) do not migrate out of Zionist conviction but for pragmatic reasons - to exchange their adverse living conditions for better ones. This, however, reaffirms the Zionist idea that a Jewish homeland is the solution to their existential problems in the diaspora.

See also: Anti-Semitism; Buber, M.; Ha'am, Ahad; Jewish philosophy, contemporary; Leibowitz, Y.; Nation and nationalism; Political philosophy, history of

ZE'EV LEVY

References and further reading

Avineri, S. (1981) *The Making of Modern Zionism*, London: Weidenfeld & Nicolson.(A historical survey of the unfolding of Zionist theory.)

Buber, M. (1909-11) *Drei Reden über das Judentum (Three Addresses on Judaism)*, Frankfurt: Rütten & Löning; trans. E. Jospe, 'Three Addresses on Judaism', in N. Glatzer (ed.) *On Judaism*, New York: Schocken, 1967.(On Judaism and Jews, Judaism and humanity, and the renewal of Judaism by Zionism.)

Borochow, B. (1937) *Nationalism and the Class Struggle*, selections ed. M. Cohen, New Brunswick, NJ: Transaction Books.(A Marxian view of Zionism as the solution to the 'Jewish Problem'.)

Cohen, I. (1946) *The Zionist Movement*, New York: The Zionist Organization of America.(History and issues of Zionism, with a supplementary chapter on the Zionist movement in America.)

Halpern, B. (1969) *The Idea of the Jewish State*, Cambridge, MA: Harvard University Press.(Historical survey.)

Heller, J. (1949) *The Zionist Idea*, New York: Harper & Row.(Ideological analysis and survey.)

Hertzberg, A. (ed.) (1970) *The Zionist Idea*, Westport, CT: Greenwood Press.(Anthology about the intellectual history of the Zionist movement.)

Herzl, T. (1896) *Der Judenstaat. Versuch einer modernen Loesung der juedischen Frage (The Jewish State: Attempt at a Modern Solution to the Jewish Question)*, New York: American Zionist Emergency Council, 1946. (Proposes the creation of a Jewish state in Palestine (or Argentina) as a prerequisite for organized migration and settlement, in order to solve the Jewish question.)

Hess, M. (1862) *Rom und Jerusalem: Die letzte Nationalitätenfrage (Rome and Jerusalem: The Last Questions of Nationality)*, Cologne, Vienna: R. Löwit, 1918; trans. M.J. Bloom, *Rome and Jerusalem*, New York: Philosophical Library, 1958.(Protozionist classic.)

Klatzkin, J. (1921) *Krisis und Entscheidung im Judentum (Crisis and Decision for Jewry)*, Berlin: Jüdischer Verlag.(An analysis of Jewish existence and nationality, of diaspora versus Palestine as a national centre.)

Pinsker, L. (1882) ‘"Autoemancipation": Mahnruf an seine Stammesgenossen von einem russischen Juden’; *Auto-Emancipation*, London: Federation of Zionist Youth, 1936.(An analysis of the psychological and social roots of antisemitism and call for the establishment of a Jewish homeland as a national centre.)

Sokolow, N. (1919) *History of Zionism*, London: Longman, Green, 2 vols.(Traces the origins of the Zionist movement.)

Zongmi (780-841)

Zongmi was a Chinese Buddhist Chan (Zen) and Huayan scholar, traditionally reckoned as the fifth 'patriarch' both in the Heze line of Southern Chan and in the Huayan scholastic tradition. He is important for his revision of Huayan doctrine, his commentaries to the Yuanjuejing (Scripture of Perfect Enlightenment), his accounts of Chan teachings and his contribution to the theory of the essential unity of the three teachings of Buddhism, Confucianism and Daoism.

Zongmi (more fully, Guifeng Zongmi) was born into a family of local prominence in a town in the southwestern province of Sichuan. As a youth he received a traditional education in the Confucian classics; he became interested in Buddhism as an adolescent, but continued his Confucian studies at a local academy in preparation for the imperial examinations that would open the door to an official career. In 804, however, he met the Chan master Daoyuan and was so impressed that he abandoned his worldly ambitions to become the latter's disciple. Later, while still a novice monk, he experienced his first awakening after reading a few pages of the *Yuanjuejing (Scripture of Perfect Enlightenment)*. Zongmi continued his Chan study under Daoyuan, receiving full ordination in 807.

In 810, Zongmi was introduced to the commentary and subcommentary to the *Avatamsaka Sūtra* written by the eminent Huayan scholar Chengguan (738-839). These works so impressed him that in 812 he became a disciple of Chengguan and studied intensively with him for two years in the imperial capital of Changan. He subsequently embarked on an extensive reading of the Buddhist canon. In 816 he left the capital to continue his research at various temples on Mount Zhongnan, where he spent much of the remainder of his life. This period of prolonged study bore fruit in a series of definitive commentaries and subcommentaries to the *Yuanjuejing* completed during 823 and 824.

In 828 Zongmi was invited to court, where he was honoured with a purple robe and the title of 'Great Worthy'. He returned to Mount Zhongnan a year or two later, where he carried on an active correspondence with a number of the leading scholar-officials he had met in Changan. He also wrote two major works on Chan, *Chanyuan zhuquanji duxu (Preface to the Collected Writings on the Source of Chan)* and *Zhonghua chuanxindi chanmen shizi chengxi tu (Chart of the Master Disciple Succession within the Chan School That Transmits the Mind Ground in China)*. His scholarly activity came to an abrupt end in 835, when he became implicated in an abortive attempt to oust the eunuchs from power at court. Afterwards he withdrew from public life, and nothing further is known about him until his death in 841.

Zongmi's work can be seen as the culmination of a number of themes important in the evolution of the medieval Chinese Buddhist scholastic enterprise. By simplifying the doctrinal complexities of earlier scholasticism, it achieved a greater coherence and systematic unity. Philosophically, it illustrates the tendency within Chinese Buddhism to de-emphasize the teaching of emptiness, which undermines ontological speculation (see Buddhist concept of emptiness), by elaborating the ontological implications within the Indian Buddhist doctrine of the *tathāgatagarbha* (the potentiality for Buddhahood inherent in all sentient beings, more popularly referred to as Buddha-nature) (see Buddhist philosophy, Chinese).

In the context of the Huayan tradition, Zongmi's systematic classification of Buddhist teachings marks a significant revision of the tradition from its classical formulation by Fazang (643-712). Most significantly, Zongmi omitted the Perfect Teaching, which Fazang had ranked highest in his fivefold scheme, and, in its stead he raised to pre-eminence the teaching that Fazang had merely ranked third. The teaching that was displaced had elaborated the unimpeded interrelation of all phenomena, and was precisely that which Fazang had taken to express the most profound insight of the Buddha, which was revealed exclusively in the *Avataṃsaka Sūtra* and was therefore the basis of the Huayan tradition's claim to represent the most exalted teaching of Buddhism. Textually, Zongmi's rearrangement of Buddhist teachings signals his displacement of the *Avataṃsaka Sūtra* in favour of the *Yuanjuejing* as the ultimate basis of Huayan thought; doctrinally, it reveals his emphasis on the *tathāgatagarbha* as the most fundamental teaching of the Buddha.

The revision that Zongmi made in Huayan doctrinal classification can be understood as part of his effort to provide

an ontological basis for Buddhist practice. His revalorization of Huayan thought in turn points back to his involvement with, and reaction to, various developments that had taken place within Chan. The iconoclastic rhetoric of the radical movements that had gained currency within Chinese Chan during the later part of the Tang dynasty (618-907) could easily be misinterpreted in antinomian ways that denied the need for spiritual cultivation and moral discipline. Having grown up and received his early Chan training in Sichuan, an area in which the most extreme of these movements flourished in the late eighth and early ninth centuries, Zongmi was particularly sensitive to such ethical dangers. Zongmi's concern with the ethical implications of Chan teachings, moreover, reflects the lasting influence of his early study of the Confucian classics.

The importance of Confucianism for Zongmi can be seen in his attempt to develop a syncretic framework in which Confucian moral teachings could be integrated within Buddhism. He thus claimed that the differences among the three teachings of Buddhism, Confucianism and Daoism were a matter of expedients, being a function of the particular historical circumstances in which they were taught and having nothing to do with level of understanding attained by each of the three sages (Buddha, Confucius and Laozi). Zongmi's writings demonstrate his life-long effort to justify the values that he had learned as a youth in terms of the discrepant claims of the religion to which he had converted as an adult.

See also: Buddhist philosophy, Chinese; Confucian philosophy, Chinese; Fazang

PETER N. GREGORY

List of works

Zongmi (780-841) *Chanyuan zhuquanji duxu (Preface to the Collected Writings on the Source of Chan)*, ed. and trans. J.L. Broughton, 'Kuei-feng Tsung-mi: The Convergence of Ch'an and the Teachings', Ph.D. dissertation, Columbia University, 1975.(The thesis contains an annotated translation of Zongmi's *Chanyuan zhuquanji duxu*.)

Zongmi (780-841) *Zhonghua chuanxindi chanmen shizi chengxi tu (Chart of the Master Disciple Succession within the Chan School That Transmits the Mind Ground in China).*

Zongmi (780-841) *Yuanrenlun (Inquiry into the Origin of Humanity)*, ed. and trans. P.N. Gregory, *Inquiry into the Origin of Humanity: An Annotated Translation of Tsung-mi's Yüan jen lun with a Modern Commentary*, Honolulu, HI: University of Hawaii Press, 1995.(Uses a translation and commentary to Zongmi's *Yuanrenlun* as a means of presenting medieval Chinese Buddhist thought to the non-specialist.)

References and further reading

Gregory, P.N. (1991) *Tsung-mi and the Sinification of Buddhism*, Princeton, NJ: Princeton University Press.(The most comprehensive treatment of Zongmi's life and thought within the context of Chinese Buddhism and Chinese intellectual history.)

Jan Yün-hua (1972) 'Tsung-mi: His Analysis of Ch'an Buddhism', *T'oung Pao* 58: 1-54.(Good introduction to Zongmi's life and understanding of Chan.)

Zoroastrianism

Zarathushtra, better known to the Classical and modern world in the Greek form of his name 'Zoroaster', revealed his vision of truth, wisdom and justice in the verse texts known as the Gāthās (c.1200-1000 BC) and is revered by Zoroastrians as their holy prophet. The religion is correctly described as mazdāyasna, *'the worship of Ahura ('Lord') Mazdā, creator of the world and source of all goodness. Since the Avestan word* mazdā *means 'wise, wisdom', Zoroastrians see their prophet as the original* philosophos, *'lover of wisdom'. Zarathushtra's message is primarily ethical and rationalistic. Zoroastrianism teaches a life based on (1) the avoidance of evil, through rigorous discrimination between good and evil, and (2) the service of wisdom through the cherishing of seven ideals. These latter are personified as seven immortal, beneficent spirits: Ahura Mazdāhimself, conceived as the creative 'holy' spirit; Sublime Truth; Virtuous Power; Good Purpose/Mind; Beneficent Piety; Wholeness/Health and, finally, Immortality. Evil originates neither from God nor from his creatures, but from a wholly other source, personified as Angra Mainyu, the 'Hostile Spirit', whose existence is ritually and doctrinally rejected as being pretended and parasitic. Real existence is solely the domain of Ahura Mazdāand his creation; Angra Mainyu and his demons are actually states of negativity, denial or, as the religion puts it, 'the Lie'. Thus the charge that the religion is ontologically dualistic is no more true than it is of other systems which conceive of good and evil as being in fundamental opposition. Equally, the allegation that its theology is ditheistic or polytheistic is a misunderstanding of the Zoroastrian theological and ritual tradition. The influence which this religion has exerted on classical philosophy and the thought and practice of Judaism, Christianity and Islam is being reappraised by scholars in modern times.*

1 Prophet and community

The Iranian prophet Zarathushtra founded one of the world's oldest and most enduring religions many centuries before Jesus of Nazareth and several millennia before Nietzsche's usurpation of the name (see Nietzsche, F. §10). Zoroastrianism was the religion of pre-Islamic Iran and survives to this day in one of the world's smallest and most widely dispersed communities, the Iranian Zardushtis and their Indian 'Parsi' co-religionists. In the classical world and in pre-modern Europe, the figure of Zoroaster has been associated, in esoteric and popular circles, with a mysterious and arcane wisdom, but the actual religious tradition which bears his name is generally unknown, or underestimated as to its historical importance.

Zoroastrianism was the principal religion of the Iranians under the Achaemenian, Parthian and Sasanian empires for a thousand years from the sixth century BC, and as the state church of the Sasanian kings in their long wars against Rome it clashed directly with Christianity. It was disestablished by the Islamic conquest of Iran in the mid-seventh century AD, and since then Zoroastrians have lived as a dispossessed religious minority in their homeland. As a result of religious persecution, a number of them migrated to northwest India in the tenth century AD and established the Parsi ('Persian') community. Their descendants flourished in the Indian milieu, first as agriculturalists, then as traders, merchants and entrepreneurs in the colonial and postcolonial periods. Zoroastrianism is an ethnic religion which generally does not accept converts; in modern times the problems of the community's demographic decline have worsened as a result of inter-marriage, secularization and emigration to Western countries from India, Pakistan and post-revolutionary Iran. The total number of Zoroastrians worldwide is reckoned to be less than 120,000.

Zarathushtra's revelation is preserved in the *Gāthās*, a small body of very ancient hymns dated by scholars as originating from the latter part of the second millennium BC and written in a northeastern Iranian language, Gāthic Avestan. The *Gāthās* are the oldest stratum of a larger scriptural canon, known collectively as the *Avesta*, which were originally transmitted orally and subsequently written down in Avestan (a language related closely to Sanskrit). As a result of the destruction of the priesthood and their lore at the hands of Alexander the Great (forever cursed by the tradition) and the later Muslim conquerors, only one fifth of the original *Avesta* survives. The *Avesta* was from ancient times accompanied by an exegetical body of texts known as the *Zand*. This *Zand* is extant now only in a later, Middle Persian language, Pahlavi, written down in its present form in the Sasanian and early Islamic periods (4th-10th centuries AD). Taken on their own, without the evidence of these later books, the *Gāthās* are difficult to interpret because of their prehistoric context and archaic symbolism: the study of the *Gāthās* continues to divide modern scholars, with regard to dating, geography and the precise nature of Zarathushtra's

teachings. However, the Zoroastrian tradition as preserved in the Avestan and Pahlavi texts gives a clearer picture of the prophet and his mission. Zarathushtra refers to himself as a priest of his ancestral religion, skilled in its ritual and teachings. Dissatisfied, however, he went in search of a nobler doctrine to deliver the world from the tyranny of oppressors and, after years spent wandering, at the age of thirty he received the first of many visions from Ahura Mazdā. His prophetic mission lasted the rest of his life, receiving and delivering the divine message to those who would listen until, as tradition says, he was killed by an assassin in his old age.

2 Cosmology and doctrine

According to Zoroastrian doctrine, before all existence there were two spirits, one good, creative of life, and beneficent; the other one evil, inimical to life, and malevolent. The two spirits are eternally opposed to one another. Ahura Mazdāis identified in the tradition with the spirit of goodness, called Spenta Mainyu, translated as ‘Holy or Beneficent Spirit’; he created first the spiritual then the physical worlds as a means of combating and finally annihilating Angra Mainyu, the ‘Hostile Spirit’. The physical world was originally perfect when created, but was invaded and is now afflicted by the Hostile Spirit. Zarathushtra urges those who will listen to choose goodness and smite evil by aligning themselves with Ahura Mazdāand his spiritual helpers, the *amesha spentas*, ‘blessed immortals’ and *yazatas* ‘[beings] worthy of worship’. Human beings have a soul, embodied in this physical world, which is part of an eternal nature that remains in the spiritual world. The human soul records a person’s acts of virtue and wickedness which will be respectively rewarded and atoned for in a spiritual state after one life in the physical state. The religion advocates purity in thought, word and deed, as Angra Mainyu and his demons benefit from all acts of wickedness and defilement of Ahura Mazdā’s creation. Human life is therefore a constant struggle for both goodness and purity. Towards the end of human history a *saoshyant*, ‘one who brings benefit’, will come to champion the victory of the righteous over the forces of evil. Time will end, there will be a resurrection of all the dead and a last judgment of all souls. A new, spiritually and physically perfect world will be made here on earth, and evil will be banished for eternity from the universe. There will be a return to the pristine perfection of the original creation, but with the fullness of physical and spiritual multiplicity of forms unafflicted by evil in a state called *Frashokereti* (‘Renovation’). The religion is thus based on a vision which is primarily soteriological and eschatological, promising deliverance from imperfection, through human effort aligned with the divine will. There is also a sophisticated philosophical theology which reflects on Zoroastrian salvation history, cosmology and ethics, contained in some of the Pahlavi books, notably the *Dēnkard*, based upon older Avestan material which is now lost.

3 Moral and physical purification

Zarathushtra’s vision became embedded in a sacramental liturgical tradition which is a continuation and elaboration of the pre-Zoroastrian cult. Scholars believe that in remote antiquity an Indo-Iranian civilization of Central Asia comprised one great cultural and linguistic group, living a semi-nomadic pastoral existence on the inner Asian steppes. Their common religious tradition would have been formed in the fourth to third millennia BC, and thus when, around 1800 BC, the Indians and Iranians separated and began to move to what are now their respective homelands, they took with them the basis of what became the related religious cultures of Brahmanism and Zoroastrianism (see Hindu philosophy). Over time they articulated their religious cult and beliefs in distinctive ways, but both share a fundamental regard for the world as *alive with divinity*, needing maintenance, through spiritual and physical acts, which refresh and reconsecrate the elements and of the cosmos as a whole. The central religious act was the rite called *yasna* in Avestan, *yajña* in Vedic Sanskrit, literally ‘act of worship’, performed in order to re-enact the primordial cosmic sacrifice by which the universe was begun. In Zoroastrianism, although it is an elaborate ceremony, it is essentially an act of reconsecration of the elements of the world: metal, water, earth, plant, animal, human and fire.

The common basis of the Indian and Iranian religious traditions is also exemplified by the cognate words of Vedic Sanskrit *ṛta* and Zoroastrian Avestan *asha*, both of which mean ‘order, law, truth’, the highest principle in the universe. *Asha* also means spiritual and moral righteousness: the righteous man is called *ashavan*, as distinct from the *dregvant*, the one who adheres to evil (Avestan *drug*). Zarathushtra’s message is a transformation of the old Indo-Iranian religion, which emphasizes as paramount the *human* (as opposed to superhuman) ethical responsibility for the maintenance of the world through the fight against the forces of disorder. The beliefs and practices of Zoroastrianism all aim at achieving a harmony of the human world with that of nature, physical,

spiritual and divine. The Blessed Immortals named in the *Gāthās* were associated with the elements of the physical world, not as merely symbolic correspondences but as the potent invisible realities which protect and inhabit the elements of the physical world. The living world of plants and animals, and the rest of the natural world, is felt to be alive with divinity, not in a pantheistic sense of a 'nature-cult', but because the physical and spiritual realms of existence are thought to be mutually sustained and protected. This is the basis of the moral and physical code of purity in Zoroastrianism. Fire, for example, is revered not in itself but as the manifestation of the Blessed Immortal Asha Vahishta 'Best Righteousness/Law', and has the epithet 'truth-strong fire'. The dubbing of Zoroastrians as 'Fire-worshippers' is an old and obvious misunderstanding of the religion.

Zoroastrians understand purity as both moral and physical: physical purity is an essential requirement for any religious or moral action. Although the physical world is vulnerable to pollution, it must be emphasized that it is not intrinsically polluted or inferior to the spiritual state. The traditional daily regime of purification for the laity and particularly the priesthood might be thought strenuous by non-Zoroastrians, but this is seen as enhancing the state of persons in physicality, not diminishing their condition of embodiment. As far as is known there have never been any tendencies towards the ascetic practices of self-mortification found in other religions. Little has been said in this entry of the complex ritual and liturgy of the traditional practice of priesthood and laity. Modernist tendencies in the contemporary community question the value of the traditional forms of practice, while conservative elements adhere to them as unchangeable. Underlying all forms of Zoroastrianism, however, is the belief that the example of Zarathushtra and his *spenta daēna* ('virtuous conception') of humanity express a true human potentiality to reflect the divine nature.

See also: Illumination; Neoplatonism; Religion, philosophy of

ALAN WILLIAMS

References and further reading

There is no complete English translation of the Avestan and Pahlavi scriptures. However, Boyce (1984: 161-5) provides a bibliography of sources.

Boyce, M. (1975-) *A History of Zoroastrianism*, Leiden: Brill.(The most authoritative scholarly treatment of the subject to date. Volumes 1, 2 and 3 have been published.)

Boyce, M. (1977) *A Persian Stronghold of Zoroastrianism*, Oxford: Universities of America Press, 1989. (Ethnographic study of Zoroastrians in pre-revolutionary Iran.)

Boyce, M. (1979) *Zoroastrians, their Beliefs and Practices*, London: Routledge & Kegan Paul; 3rd revised repr. 1988.(Succinct account of the religion.)

Boyce, M. (1984) *Textual Sources for the Study of Zoroastrianism*, Manchester: Manchester University Press. (New, reliable translations of major texts.)

de Menasce, J. (1973) *Le troisième livre du Dēnkart*, Paris: Travaux de l'Institut d'Etudes Iraniennes de l'Université de Paris III, 5, Bibliothèque des Oeuvres Classiques Persanes, 4.(The only European translation of this important philosophical text.)

Insler, S.H. (1975) *The Gāthās of Zarathustra*, Acta Iranica 8, Leiden: Brill.(Scholarly translation of the oldest Zoroastrian scripture.)

Kulke, E. (1974) *The Parsis in India, A Minority as Agent of Social Change*, Munich: Verlag.(On Parsis in the modern world.)

Williams, A.V. (1989) 'The Body and the Boundaries of Zoroastrian Spirituality', *Religion* 19: 227-39.(On purity and pollution in the scheme of cosmology.)

Williams, A.V. (1990) *The Pahlavi Rivāyat Accompanying the Dādestān īDēnīg*, Copenhagen: Royal Danish Academy of Sciences and Letters, 2 vols.(Representative miscellany of ninth-century Pahlavi religious genres.)

Williams, A.V. (1997) 'Later Zoroastrianism', in I. Mahalingam and B. Carr (eds) *Companion Encyclopedia of Asian Philosophy*, London, Routledge, 24-45.(On the philosophy and theology of the Pahlavi books.)